



Σχολή Θετικών Επιστημών και Τεχνολογίας
Μεταπτυχιακή Εξειδίκευση στα Πληροφοριακά Συστήματα

Διπλωματική Εργασία
Πρόβλεψη της ηλεκτρικής κατανάλωσης κτηρίων με χρήση
τεχνικών μηχανικής μάθησης

Γεώργιος Γερασιμόπουλος

Επιβλέπων καθηγητής: Βασίλειος Καψάλης

Αμβούργο, Σεπτέμβριος 2026

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία του φοιτητή Γερασιμόπουλου Γεώργιου που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης ο συγγραφέας/δημιουργός εκχωρεί στο ΕΑΠ, μη αποκλειστική άδεια χρήσης του δικαιώματος αναπαραγωγής, προσαρμογής, δημόσιου δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσής τους διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος και για όλο το χρόνο διάρκειας των δικαιωμάτων πνευματικής ιδιοκτησίας. Η ανοικτή πρόσβαση στο πλήρες κείμενο για μελέτη και ανάγνωση δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, αποθήκευση, πώληση, εμπορική χρήση, μετάδοση, διανομή, έκδοση, εκτέλεση, «μεταφόρτωση» (downloading), «ανάρτηση» (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού. Ο συγγραφέας/δημιουργός διατηρεί το σύνολο των ηθικών και περιουσιακών του δικαιωμάτων.



Πρόβλεψη της ηλεκτρικής κατανάλωσης κτηρίων με χρήση τεχνικών μηχανικής μάθησης

Γεώργιος Γερασιμόπουλος

Επιτροπή Επίβλεψης Διπλωματικής Εργασίας

Επιβλέπων Καθηγητής:

Βασίλειος Καψάλης

Καθηγητής Τμήματος Ηλεκτρολόγων
Μηχανικών και Μηχανικών Υπολογιστών
Πανεπιστημίου Πελοποννήσου

Συν-Επιβλέπων Καθηγητής:

Γεώργιος Ρηγόπουλος

Μέλος ΣΕΠ ΕΑΠ

Αμβούργο, Σεπτέμβριος 2026

Ευχαριστίες

Με την ολοκλήρωση της παρούσας διπλωματικής εργασίας, θα ήθελα αρχικά να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ. Βασίλειο Καψάλη, για τις συμβουλές και την καθοδήγηση του σε κρίσιμα σημεία κατά τη διαδικασία της εκπόνησης. Ένα μεγάλο ευχαριστώ στη σύζυγό μου, για την ενθάρρυνση, την υποστήριξη και τη συμβολή της σε όλο το διάστημα των σπουδών μου στο μεταπτυχιακό αυτό πρόγραμμα, καθώς και σε όλη την οικογένειά μου.

«Στη σύζυγό μου Μαρίνα»

Περίληψη

Τα ενεργειακά αποδοτικά και βιώσιμα κτήρια διαδραματίζουν καθοριστικό ρόλο στην προστασία του περιβάλλοντος και στον περιορισμό της κλιματικής αλλαγής, καθώς ο κτηριακός τομέας ευθύνεται για σημαντικό ποσοστό των παγκόσμιων εκπομπών αερίων του θερμοκηπίου. Στο πλαίσιο αυτό, η πρόβλεψη της ηλεκτρικής κατανάλωσης αποτελεί βασικό εργαλείο για τη βελτιστοποίηση της ενεργειακής διαχείρισης και τη λήψη αποδοτικότερων αποφάσεων.

Η πρόβλεψη της κατανάλωσης ενέργειας αποτελεί ένα σύνθετο πρόβλημα, καθώς επηρεάζεται από πλήθος παραγόντων, όπως οι καιρικές συνθήκες, τα χαρακτηριστικά του κτηρίου και η συμπεριφορά των χρηστών. Στην παρούσα εργασία εξετάζεται η βραχυπρόθεσμη πρόβλεψη ηλεκτρικής κατανάλωσης σε οικιακό περιβάλλον, με χρήση τεχνικών μηχανικής μάθησης, αξιοποιώντας πραγματικά δεδομένα κατανάλωσης και περιβαλλοντικών μεταβλητών.

Συγκεκριμένα, αναπτύσσονται και αξιολογούνται μοντέλα πρόβλεψης σε δύο διαφορετικά σύνολα δεδομένων, τόσο χωρίς χρήση ιστορικών δεδομένων όσο και με ενσωμάτωση ιστορικών χαρακτηριστικών, αναδεικνύοντας τη σημασία της χρονικής εξάρτησης στην κατανάλωση ενέργειας. Παράλληλα, υλοποιείται ένα ολοκληρωμένο σύστημα πρόβλεψης που περιλαμβάνει διεπαφή προγραμματισμού εφαρμογών και γραφική διεπαφή χρήστη, επιτρέποντας την πρακτική αξιοποίηση των μοντέλων.

Η εργασία καταλήγει στην ανάπτυξη ενός λειτουργικού συστήματος πρόβλεψης κατανάλωσης επόμενης ημέρας σε επίπεδο κατοικίας, το οποίο μπορεί να υποστηρίξει εφαρμογές ενεργειακής διαχείρισης.

Λέξεις – Κλειδιά

Πρόβλεψη ηλεκτρικής κατανάλωσης, μηχανική μάθηση, οικιακά κτήρια, βραχυπρόθεσμη πρόβλεψη φορτίου, ιστορικά χαρακτηριστικά, συστήματα πρόβλεψης.

Forecasting the electricity consumption of buildings using machine learning techniques

Georgios Gerasimopoulos

Abstract

Energy-efficient and sustainable buildings play a crucial role in environmental protection and in mitigating climate change, as the building sector is responsible for a significant share of global greenhouse gas emissions. In this context, electricity consumption forecasting is an essential tool for optimizing energy management and supporting more efficient decision-making.

Forecasting energy consumption is a complex problem, as it is influenced by multiple factors, such as weather conditions, building characteristics, and user behavior. In this thesis, short-term electricity consumption forecasting in a residential setting is examined using machine learning techniques, based on real consumption data and environmental variables.

More specifically, forecasting models are developed and evaluated on two different datasets, both without the use of historical data and with the incorporation of historical features, highlighting the importance of temporal dependence in energy consumption. At the same time, a complete forecasting system is implemented, including an application programming interface and a graphical user interface, enabling the practical use of the models. The thesis leads to the development of a functional system for predicting next-day household consumption, which can support energy management applications.

Keywords

Electricity consumption forecasting, machine learning, residential buildings, short-term load forecasting, historical features, forecasting systems.

Περιεχόμενα

Περίληψη.....	v
Abstract	vi
Περιεχόμενα	vii
Κατάλογος Εικόνων / Σχημάτων	xi
Κατάλογος Πινάκων	xiv
Συντομογραφίες & Ακρωνύμια.....	xvii
1. Εισαγωγή.....	1
1.1 Υπόβαθρο και κίνητρο	1
1.2 Σκοπός και αντικείμενο της εργασίας	2
1.3 Ερευνητικά ερωτήματα	3
1.4 Μεθοδολογική προσέγγιση	4
1.5 Δομή της διπλωματικής εργασίας	5
2. Θεωρητικό υπόβαθρο.....	8
2.1 Πρόβλεψη ενεργειακής κατανάλωσης κτηρίων.....	8
2.2 Κατηγορίες μοντέλων πρόβλεψης.....	9
2.3 Δεδομενοκεντρική προσέγγιση πρόβλεψης σε οικιακό περιβάλλον	9
2.4 Χαρακτηριστικά εισόδου και διερευνητική ανάλυση (EDA).....	11
2.5 Βασικά μοντέλα πρόβλεψης.....	12
2.6 Παραγωγικά συστήματα και APIs για πρόβλεψη φορτίου	12
3. Δεδομένα και Προεπεξεργασία.....	14
3.1 Το σύνολο δεδομένων IDEAL	14
3.2 Επιλογή και φιλτράρισμα κατοικιών	15
3.3 Ροή προεπεξεργασίας και ενοποίησης δεδομένων.....	15
3.4 Ημερολογιακά χαρακτηριστικά και τελικό σύνολο δεδομένων	17
4. Διερευνητική Ανάλυση Δεδομένων και Χαρακτηριστικά	18
4.1 Προσέγγιση διερευνητικής ανάλυσης δεδομένων	18
4.2 Ημερήσια και ωριαία προφίλ κατανάλωσης	18
4.3 Επίδραση εξωτερικής θερμοκρασίας	19
4.4 Διαφορές ανά τύπο κατοικίας (hometype).....	20
4.5 Διαφορές ανά κατηγορίες αστικότητας, αριθμού ενοίκων και εμβαδού κατοικίας.....	21
4.6 Επιλογή και τελικός ορισμός χαρακτηριστικών	24
4.7 Βασικά μοντέλα αναφοράς	27
5. Μεθοδολογία Μοντελοποίησης	29
5.1 Ορισμός προβλήματος και ορίζοντας πρόβλεψης.....	29
5.2 Κατηγοριοποίηση κατοικιών μέσω χαρακτηριστικών	29
5.3 Σενάρια πρόβλεψης.....	30
5.4 Διαχωρισμός σε σύνολο εκπαίδευσης και σύνολο ελέγχου (train/test)	31
5.5 Μετασχηματισμοί στόχου και προεπεξεργασία.....	32
5.6 Επιλογή αλγορίθμων μηχανικής μάθησης	33
5.7 Random Forest	35
5.8 XGBoost.....	36
5.9 LightGBM.....	37
6. Αποτελέσματα και αξιολόγηση βασικών μοντέλων αναφοράς και μοντέλων μηχανικής μάθησης.....	39
6.1 Μετρικές αξιολόγησης και πειραματική διαδικασία	39
6.2 Απόδοση των βασικών μοντέλων αναφοράς	40
6.3 Απόδοση των μοντέλων στο σενάριο χωρίς ιστορικά δεδομένα	42

6.4 Απόδοση των μοντέλων στο σενάριο με χρήση ιστορικών δεδομένων.....	43
6.5 Συνολική σύγκριση και αξιολόγηση	44
6.6 Πρόσθετα πειράματα βελτίωσης των προβλέψεων	45
6.6.1 Προσαρμογή της ημερήσιας καμπύλης φορτίου με βάση το πρόσφατο ιστορικό κατανάλωσης.....	45
6.6.2 Τεχνικές μεταγενέστερης διόρθωσης των προβλέψεων	46
6.6.3 Ομαδοποίηση κατοικιών και εναλλακτικές προσεγγίσεις	47
7. Ανάπτυξη και Υλοποίηση του Συστήματος Πρόβλεψης	49
7.1 Αρχιτεκτονική του συστήματος πρόβλεψης	49
7.2 Επίπεδο μοντέλων (Model Layer).....	50
7.3 Επίπεδο υπηρεσιών και προγραμματιστική διεπαφή (API Layer).....	51
7.4 Διαχείριση ιστορικών δεδομένων κατανάλωσης	52
7.5 Διεπαφή χρήστη (User Interface Layer)	53
7.6 Εγκατάσταση και περιβάλλον εκτέλεσης	55
8. Εφαρμογή, διεπαφή χρήστη και σενάρια χρήσης	56
8.1 Αρχιτεκτονική εκτέλεσης και μηχανισμός εκκίνησης	56
8.2 Επισκόπηση γραφικής διεπαφής χρήστη	58
8.3 Βασική ροή πρόβλεψης χωρίς ιστορικά δεδομένα	59
8.4 Πρόβλεψη με ιστορικά δεδομένα και ανάλυση ιστορικού	62
8.5 Σύγκριση μοντέλων και συνδυαστικές προβλέψεις	67
8.6 Πρόσθετες λειτουργίες προσαρμογής των προβλέψεων.....	68
8.7 Χειρισμός εξωτερικής θερμοκρασίας	69
9. Αξιολόγηση του συστήματος πρόβλεψης	71
9.1 Μεθοδολογία αξιολόγησης	71
9.2 Αξιολόγηση επιλεγμένων κατοικιών του συνόλου ελέγχου	73
9.2.1 Αποτελέσματα για την κατοικία home72.....	73
9.2.2 Αποτελέσματα για την κατοικία home100.....	77
9.2.3 Αποτελέσματα για την κατοικία home268.....	81
9.2.4 Αποτελέσματα για την κατοικία home316.....	85
9.3 Εποχικές μελέτες περίπτωσης	90
9.3.1 Φθινοπωρινή αξιολόγηση για την κατοικία home72	90
9.3.2 Θερινή αξιολόγηση για την κατοικία home100	93
9.3.3 Χειμερινή αξιολόγηση για την κατοικία home268	96
9.4 Συνολική αποτίμηση κύριας και εποχικής αξιολόγησης.....	102
9.5 Συνολικά συμπεράσματα κεφαλαίου	106
10. Το σύνολο δεδομένων PLEGMA.....	110
10.1 Εισαγωγή.....	110
10.2 Περιγραφή του συνόλου δεδομένων PLEGMA	110
10.3 Ρόλος του PLEGMA dataset στην παρούσα εργασία	112
11. Προεπεξεργασία και συγχρονισμός δεδομένων PLEGMA	113
11.1 Προετοιμασία και καθαρισμός δεδομένων	113
11.2 Συγχρονισμός και χρονική ευθυγράμμιση δεδομένων.....	114
11.3 Μετασχηματισμός χαρακτηριστικών και δημιουργία τελικού συνόλου δεδομένων	116
11.3.1 Χρονικά χαρακτηριστικά	116
11.3.2 Περιβαλλοντικά χαρακτηριστικά	117
11.3.3 Στατικά χαρακτηριστικά κατοικίας και νοικοκυριού.....	117
11.3.4 Διαχείριση ανωμαλιών και μη ρεαλιστικών τιμών	118
11.3.5 Δημιουργία τελικού συνόλου δεδομένων	118

12. Διερευνητική ανάλυση δεδομένων (EDA) στο PLEGMA	119
12.1 Προσέγγιση διερευνητικής ανάλυσης	119
12.2 Ημερήσια και ωριαία προφίλ κατανάλωσης	120
12.3 Σύγκριση καθημερινών και Σαββατοκύριακου	121
12.4 Επίδραση εξωτερικής θερμοκρασίας στην κατανάλωση	122
12.5 Κατανομή κατανάλωσης και βασικά στατιστικά χαρακτηριστικά	123
13. Μοντελοποίηση στο σύνολο δεδομένων PLEGMA	125
13.1 Μεθοδολογική προσέγγιση και προσαρμογή στο PLEGMA	125
13.2 Σενάρια πρόβλεψης	126
13.3 Δημιουργία χαρακτηριστικών για τα σενάρια πρόβλεψης	127
13.4 Μοντέλα μηχανικής μάθησης	127
13.5 Βασικά μοντέλα αναφοράς	128
13.6 Διαδικασία εκπαίδευσης και χρονικός διαχωρισμός δεδομένων	129
13.7 Αποθήκευση μοντέλων και σύνδεση με την εφαρμογή πρόβλεψης	130
14. Αξιολόγηση βασικών μοντέλων αναφοράς και μοντέλων μηχανικής μάθησης	131
14.1 Ρυθμίσεις αξιολόγησης και πειραματικό πλαίσιο	131
14.2 Αποτελέσματα βασικών μοντέλων αναφοράς	132
14.3 Απόδοση μοντέλων μηχανικής μάθησης στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης	133
14.4 Απόδοση μοντέλων μηχανικής μάθησης στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης	134
14.5 Ερμηνεία της σημασίας των χαρακτηριστικών	135
14.6 Ερμηνεία αποτελεσμάτων	136
15. Προσαρμογή της εφαρμογής πρόβλεψης στο σύνολο δεδομένων PLEGMA	138
15.1 Σκοπός της εφαρμογής και σχέση με την υλοποίηση του IDEAL	138
15.2 Προσαρμογή της αρχιτεκτονικής API/UI στο PLEGMA	139
15.3 Διαχείριση μοντέλων και σεναρίων πρόβλεψης	140
15.4 Διαχείριση και ανάλυση ιστορικών δεδομένων κατανάλωσης	141
15.5 Γραφική διεπαφή χρήστη και βασική ροή πρόβλεψης	142
15.6 Διαχείριση μετεωρολογικών και περιβαλλοντικών δεδομένων	145
15.7 Περιορισμοί και λειτουργικές παραδοχές της εφαρμογής	146
16. Αξιολόγηση προβλέψεων σε μεμονωμένες κατοικίες (PLEGMA)	148
16.1 Μεθοδολογία αξιολόγησης	148
16.2 Αξιολόγηση επιλεγμένων κατοικιών του συνόλου ελέγχου	149
16.2.1 Αποτελέσματα για την κατοικία home01	149
16.2.2 Αποτελέσματα για την κατοικία home04	152
16.2.3 Αποτελέσματα για την κατοικία home05	154
16.2.4 Αποτελέσματα για την κατοικία home07	157
16.3 Εποχικές μελέτες περίπτωσης	160
16.3.1 Θερινή αξιολόγηση για την κατοικία home01	160
16.3.2 Φθινοπωρινή αξιολόγηση για την κατοικία home04	163
16.3.3 Χειμερινή αξιολόγηση για την κατοικία home05	166
16.3.4 Εαρινή αξιολόγηση για την κατοικία home07	169
16.4 Συγκεντρωτική αποτίμηση κύριας και εποχικής αξιολόγησης	171
16.5 Συμπεράσματα κεφαλαίου	174
17. Συμπεράσματα και μελλοντικές επεκτάσεις	176
17.1 Συνοπτική ανακεφαλαίωση της εργασίας	176
17.2 Συγκριτική αποτίμηση IDEAL και PLEGMA	177
17.3 Κύρια συμπεράσματα και σύνδεση με τη βιβλιογραφία	179

17.4 Σύνδεση με τα ερευνητικά ερωτήματα	181
17.5 Περιορισμοί της μελέτης.....	181
17.6 Προτάσεις για μελλοντικές επεκτάσεις.....	183
Βιβλιογραφία.....	185
Παράρτημα Α: Βιβλιοθήκες και Αρχεία με κώδικα σε γλώσσα Python.....	189
Δομή αποθετηρίου GitHub.....	189
Βιβλιοθήκες Python	189

Κατάλογος Εικόνων / Σχημάτων

Εικόνα 1 Συνολική αρχιτεκτονική του συστήματος πρόβλεψης	50
Εικόνα 2 Γραφικό περιβάλλον του launcher_ui του IDEAL Forecasting System	56
Εικόνα 3 IDEAL Load Forecasting API μέσω Swagger	57
Εικόνα 4 Κύρια διεπαφή χρήστη του IDEAL Load Forecasting System	58
Εικόνα 5 Παράδειγμα αποτελεσμάτων πρόβλεψης στο σενάριο χωρίς χρήση ιστορικών δεδομένων	61
Εικόνα 6 Παράδειγμα ανάλυσης ιστορικού κατανάλωσης	62
Εικόνα 7 Παράδειγμα αποτελεσμάτων πρόβλεψης στο σενάριο με χρήση ιστορικών δεδομένων	66
Εικόνα 8 Παράδειγμα συνδυαστικής πρόβλεψης (Combined Prediction) και σύγκρισης μοντέλων	68
Εικόνα 9 Γραφική διεπαφή χρήστη της εφαρμογής πρόβλεψης για το PLEGMA	143
Εικόνα 10 Γραφική διεπαφή χρήστη της εφαρμογής πρόβλεψης για το PLEGMA	144

Σχήμα 1 Μέση ωριαία κατανάλωση (home-weighted) για καθημερινές και Σαββατοκύριακα.	19
Σχήμα 2 Μέση και διάμεση ωριαία κατανάλωση σε συνάρτηση με την εξωτερική θερμοκρασία.....	20
Σχήμα 3 Μέση ωριαία κατανάλωση (home-weighted) ανά τύπο κατοικίας.....	21
Σχήμα 4 Μέση ωριαία κατανάλωση (home-weighted) ανά κατηγορία urban_rural_class.....	22
Σχήμα 5 Κατανομή μέσης ημερήσιας κατανάλωσης ανά κατοικία και περιοχή	22
Σχήμα 6 Μέση ωριαία κατανάλωση (home-weighted) ανά κατηγορία εμβαδού κατοικίας	23
Σχήμα 7 Σημαντικότητα χαρακτηριστικών για το μοντέλο Random Forest στο σενάριο χωρίς χρήση ιστορικών δεδομένων.....	25
Σχήμα 8 Σημαντικότητα χαρακτηριστικών για το μοντέλο Random Forest στο σενάριο με χρήση ιστορικών δεδομένων.....	26
Σχήμα 9 Τα 5 σημαντικότερα χαρακτηριστικά ανά μοντέλο (RF, XGB, LGBM) και σενάριο	26
Σχήμα 10 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	74
Σχήμα 11 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στο απλό σενάριο με χρήση ιστορικών δεδομένων	75
Σχήμα 12 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	76
Σχήμα 13 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	78
Σχήμα 14 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στο απλό σενάριο με χρήση ιστορικών δεδομένων	79
Σχήμα 15 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	80
Σχήμα 16 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	82
Σχήμα 17 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στο απλό σενάριο με χρήση ιστορικών δεδομένων	83

Σχήμα 18 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	84
Σχήμα 19 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	86
Σχήμα 20 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στο απλό σενάριο με χρήση ιστορικών δεδομένων	87
Σχήμα 21 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	88
Σχήμα 22 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στη φθινοπωρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων	91
Σχήμα 23 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στη φθινοπωρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	92
Σχήμα 24 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στη θερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων	94
Σχήμα 25 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στη θερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	95
Σχήμα 26 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στη χειμερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων	97
Σχήμα 27 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στη χειμερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	98
Σχήμα 28 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στην εαρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων	100
Σχήμα 29 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στην εαρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	101
Σχήμα 30 Μέση ωριαία κατανάλωση ηλεκτρικής ενέργειας στο σύνολο δεδομένων PLEGMA	120
Σχήμα 31 Μέση ωριαία κατανάλωση για καθημερινές και Σαββατοκύριακα στο σύνολο δεδομένων PLEGMA	121
Σχήμα 32 Μέση κατανάλωση ηλεκτρικής ενέργειας σε συνάρτηση με την εξωτερική θερμοκρασία στο σύνολο δεδομένων PLEGMA	122
Σχήμα 33 Κατανομή της κατανάλωσης ηλεκτρικής ενέργειας στο σύνολο δεδομένων PLEGMA	123
Σχήμα 34 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home01 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	150
Σχήμα 35 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home01 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	151
Σχήμα 36 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home04 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	153
Σχήμα 37 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home04 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	154
Σχήμα 38 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home05 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	155
Σχήμα 39 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home05 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	156

Σχήμα 40 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home07 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	158
Σχήμα 41 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home07 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	159
Σχήμα 42 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home01 στη θερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων	161
Σχήμα 43 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home01 στη θερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	162
Σχήμα 44 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home04 στη φθινοπωρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων	164
Σχήμα 45 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home04 στη φθινοπωρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	165
Σχήμα 46 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home05 στη χειμερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων	167
Σχήμα 47 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home05 στη χειμερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	168
Σχήμα 48 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home07 στην εαρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων	169
Σχήμα 49 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home07 στην εαρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	170

Κατάλογος Πινάκων

Πίνακας 1 Απόδοση βασικών μοντέλων αναφοράς στο σύνολο ελέγχου του IDEAL.....	41
Πίνακας 2 Απόδοση των μοντέλων μηχανικής μάθησης στο σενάριο χωρίς ιστορικά δεδομένα.....	42
Πίνακας 3 Απόδοση των μοντέλων μηχανικής μάθησης στο σενάριο με χρήση ιστορικών δεδομένων	43
Πίνακας 4 Συγκεντρωτική σύγκριση βασικού μοντέλου αναφοράς και μοντέλων μηχανικής μάθησης.....	44
Πίνακας 5 Σύγκριση balanced και daily-calibrated εκδόσεων των μοντέλων.....	47
Πίνακας 6 Αντιστοίχιση κατηγορίας σταθερότητας ιστορικού και προτεινόμενου max_alpha	64
Πίνακας 7 Απόδοση μοντέλων για την κατοικία home72 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	74
Πίνακας 8 Απόδοση μοντέλων για την κατοικία home72 στο απλό σενάριο με χρήση ιστορικών δεδομένων	75
Πίνακας 9 Απόδοση μοντέλων για την κατοικία home72 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	76
Πίνακας 10 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home72.....	77
Πίνακας 11 Απόδοση μοντέλων για την κατοικία home100 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	78
Πίνακας 12 Απόδοση μοντέλων για την κατοικία home100 στο απλό σενάριο με χρήση ιστορικών δεδομένων	79
Πίνακας 13 Απόδοση μοντέλων για την κατοικία home100 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	80
Πίνακας 14 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home100.....	81
Πίνακας 15 Απόδοση μοντέλων για την κατοικία home268 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	82
Πίνακας 16 Απόδοση μοντέλων για την κατοικία home268 στο απλό σενάριο με χρήση ιστορικών δεδομένων	83
Πίνακας 17 Απόδοση μοντέλων για την κατοικία home268 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	84
Πίνακας 18 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home268.....	85
Πίνακας 19 Απόδοση μοντέλων για την κατοικία home316 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	86
Πίνακας 20 Απόδοση μοντέλων για την κατοικία home316 στο απλό σενάριο με χρήση ιστορικών δεδομένων	87
Πίνακας 21 Απόδοση μοντέλων για την κατοικία home316 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	88
Πίνακας 22 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home316.....	89
Πίνακας 23 Απόδοση μοντέλων για την κατοικία home72 στη φθινοπωρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων.....	91
Πίνακας 24 Απόδοση μοντέλων για την κατοικία home72 στη φθινοπωρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	92
Πίνακας 25 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home72 στη φθινοπωρινή περίπτωση.....	93
Πίνακας 26 Απόδοση μοντέλων για την κατοικία home100 στη θερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων.....	94

Πίνακας 27 Απόδοση μοντέλων για την κατοικία home100 στη θερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση.....	95
Πίνακας 28 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home100 στη θερινή περίπτωση.....	96
Πίνακας 29 Απόδοση μοντέλων για την κατοικία home268 στη χειμερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων.....	97
Πίνακας 30 Απόδοση μοντέλων για την κατοικία home268 στη χειμερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	98
Πίνακας 31 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home268 στη χειμερινή περίπτωση	99
Πίνακας 32 Απόδοση μοντέλων για την κατοικία home316 στην εαρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων.....	100
Πίνακας 33 Απόδοση μοντέλων για την κατοικία home316 στην εαρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	101
Πίνακας 34 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home316 στην εαρινή περίπτωση.....	102
Πίνακας 35 Συγκεντρωτική σύγκριση καλύτερων σεναρίων στην κύρια αξιολόγηση.....	103
Πίνακας 36 Συγκεντρωτική σύγκριση καλύτερων σεναρίων στις εποχικές μελέτες περίπτωσης.....	104
Πίνακας 37 Απόδοση βασικών μοντέλων αναφοράς στο σύνολο δεδομένων PLEGMMA	132
Πίνακας 38 Απόδοση μοντέλων μηχανικής μάθησης στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης	133
Πίνακας 39 Απόδοση μοντέλων μηχανικής μάθησης στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης	134
Πίνακας 40 Απόδοση μοντέλων για την κατοικία home01 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	150
Πίνακας 41 Απόδοση μοντέλων για την κατοικία home01 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	150
Πίνακας 42 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home01.....	151
Πίνακας 43 Απόδοση μοντέλων για την κατοικία home04 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	152
Πίνακας 44 Απόδοση μοντέλων για την κατοικία home04 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	153
Πίνακας 45 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home04.....	154
Πίνακας 46 Απόδοση μοντέλων για την κατοικία home05 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	155
Πίνακας 47 Απόδοση μοντέλων για την κατοικία home05 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	156
Πίνακας 48 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home05.....	157
Πίνακας 49 Απόδοση μοντέλων για την κατοικία home07 στο σενάριο χωρίς χρήση ιστορικών δεδομένων	157
Πίνακας 50 Απόδοση μοντέλων για την κατοικία home07 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	158
Πίνακας 51 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home07.....	159
Πίνακας 52 Απόδοση μοντέλων για την κατοικία home01 στη θερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων.....	161
Πίνακας 53 Απόδοση μοντέλων για την κατοικία home01 στη θερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση.....	161

Πίνακας 54 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home01 στη θερινή περίπτωση.....	162
Πίνακας 55 Απόδοση μοντέλων για την κατοικία home04 στη φθινοπωρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων.....	163
Πίνακας 56 Απόδοση μοντέλων για την κατοικία home04 στη φθινοπωρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	164
Πίνακας 57 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home04 στη φθινοπωρινή περίπτωση.....	165
Πίνακας 58 Απόδοση μοντέλων για την κατοικία home05 στη χειμερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων.....	166
Πίνακας 59 Απόδοση μοντέλων για την κατοικία home05 στη χειμερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση	167
Πίνακας 60 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home05 στη χειμερινή περίπτωση	168
Πίνακας 61 Απόδοση μοντέλων για την κατοικία home07 στην εαρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων.....	169
Πίνακας 62 Απόδοση μοντέλων για την κατοικία home07 στην εαρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση.....	170
Πίνακας 63 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home07 στην εαρινή περίπτωση.....	171
Πίνακας 64 Συγκεντρωτική σύγκριση καλύτερων σεναρίων στην κύρια αξιολόγηση του PLEGMA	172
Πίνακας 65 Συγκεντρωτική σύγκριση καλύτερων σεναρίων στις εποχικές μελέτες περίπτωσης του PLEGMA.....	172

Συντομογραφίες & Ακρωνύμια

ΔΕ	Διπλωματική Εργασία
ΕΑΠ	Ελληνικό Ανοικτό Πανεπιστήμιο
API	Application Programming Interface.
ARIMA	Autoregressive Integrated Moving Average.
BEMS	Building Energy Management System(s).
CNN	Convolutional Neural Network.
DL	Deep Learning.
EDA	Exploratory Data Analysis.
EMS	Energy Management System(s).
GRU	Gated Recurrent Unit.
HEMS	Home Energy Management System(s).
IDEAL	Improving Domestic Energy Advice through Listening.
JSON	JavaScript Object Notation.
LGBM	Light Gradient Boosting Machine.
LSTM	Long Short-Term Memory.
MAE	Mean Absolute Error.
MAPE	Mean Absolute Percentage Error.
ML	Machine Learning.
N-BEATS	Neural Basis Expansion Analysis for Time Series.
NILM	Non-Intrusive Load Monitoring.
RMSE	Root Mean Squared Error.
RF	Random Forest.
SHAP	Shapley Additive Explanations.
STLF	Short-Term Load Forecasting.
UI	User Interface.
XGB	XGBoost.

1. Εισαγωγή

Σε αυτό το εισαγωγικό κεφάλαιο, παρουσιάζεται το γενικό πλαίσιο μέσα στο οποίο εντάσσεται η διπλωματική εργασία και αποσαφηνίζεται η σημασία της βραχυπρόθεσμης πρόβλεψης ηλεκτρικής κατανάλωσης σε κτηριακό και οικιακό επίπεδο. Αρχικά, αναλύεται το ενεργειακό και ερευνητικό υπόβαθρο του προβλήματος, με έμφαση στον ρόλο του κτηριακού τομέα στην κατανάλωση ενέργειας και στις δυνατότητες που δημιουργεί η αυξανόμενη διαθεσιμότητα δεδομένων από έξυπνους μετρητές και συστήματα παρακολούθησης. Στη συνέχεια, ορίζονται ο σκοπός και το αντικείμενο της εργασίας, διατυπώνονται τα βασικά ερευνητικά ερωτήματα και περιγράφεται συνοπτικά η μεθοδολογική προσέγγιση που ακολουθείται. Τέλος, παρουσιάζεται η συνολική δομή της διπλωματικής εργασίας, ώστε να αποτυπωθεί η λογική αλληλουχία των κεφαλαίων, από τη θεωρητική θεμελίωση και την επεξεργασία των δεδομένων έως την ανάπτυξη και διερεύνηση διαφορετικών προσεγγίσεων πρόβλεψης, την υλοποίηση ενός ολοκληρωμένου συστήματος πρόβλεψης και την αξιολόγηση της απόδοσης των μοντέλων σε ρεαλιστικά σενάρια χρήσης.

1.1 Υπόβαθρο και κίνητρο

Ο κτηριακός τομέας αποτελεί έναν από τους σημαντικότερους καταναλωτές ενέργειας παγκοσμίως, συμβάλλοντας σημαντικά τόσο στη ζήτηση τελικής ενέργειας όσο και στις εκπομπές αερίων του θερμοκηπίου. Διεθνείς εκθέσεις για τα κτήρια και τις κατασκευές δείχνουν ότι τα κτήρια ευθύνονται για περίπου το ένα τρίτο της παγκόσμιας τελικής κατανάλωσης ενέργειας και για σημαντικό μερίδιο των ενεργειακών εκπομπών, ιδιαίτερα όταν συμπεριλαμβάνονται οι έμμεσες εκπομπές από την παραγωγή της ηλεκτρικής ενέργειας που καταναλώνεται στη λειτουργία τους (UN Environment Programme, 2023). Παρά τις πολιτικές εξοικονόμησης και τη βελτίωση της ενεργειακής απόδοσης, οι ενεργειακές απαιτήσεις στον κτηριακό τομέα συνεχίζουν να αυξάνονται λόγω της πληθυσμιακής αύξησης, της αστικοποίησης και της βελτίωσης του βιοτικού επιπέδου (Amasyali & El-Gohary, 2018; Bourdeau, Zhai, Nefzaoui, Guo, & Chatellier, 2019).

Τα τελευταία χρόνια, η ευρεία διάδοση έξυπνων μετρητών και συστημάτων προηγμένης μέτρησης (smart meters, advanced metering infrastructure) έχει αλλάξει ριζικά τις δυνατότητες παρακολούθησης και διαχείρισης της κατανάλωσης (Hong & Fan, 2016). Σε χώρες όπως η Γερμανία, ο ρυθμός εγκατάστασης έξυπνων μετρητών επιταχύνεται, με ολοένα μεγαλύτερο ποσοστό νοικοκυριών και επιχειρήσεων να αποκτούν πρόσβαση σε μετρήσεις υψηλής χρονικής ανάλυσης (Forschungsgesellschaft für Energiewirtschaft, 2025; Clean Energy Wire, 2025). Η αυξημένη διαθεσιμότητα δεδομένων ανοίγει τον δρόμο για προηγμένες υπηρεσίες διαχείρισης ζήτησης, βελτιστοποίηση λειτουργίας κτηρίων και στοχευμένα μέτρα εξοικονόμησης (Olu-Ajayi, Alaka, Owolabi, Akanbi, & Ganiyu, 2023). Σε αυτό το πλαίσιο, η βραχυπρόθεσμη πρόβλεψη ηλεκτρικού φορτίου (short-term load forecasting, STLF) αποκτά ιδιαίτερη σημασία σε κτηριακό επίπεδο. Σε επίπεδο συστήματος ή υποσταθμού, η STLF αποτελεί ώριμο πεδίο με μακρά χρήση στατιστικών και στοχαστικών μοντέλων (Hong & Fan, 2016), ενώ σε κτηριακό επίπεδο –και ειδικά σε μεμονωμένες κατοικίες– παραμένει πιο απαιτητική λόγω της υψηλής μεταβλητότητας και της έντονης εξάρτησης από τη συμπεριφορά των ενοίκων (Runge & Zmeureanu, 2019; Olu-Ajayi, Alaka, Sulaimon, Sunmola, & Ajayi, 2022).

Επιπλέον, σε επίπεδο μεμονωμένης κατοικίας, η κατανάλωση χαρακτηρίζεται συχνά από απότομες αιχμές φορτίου, μη επαναλαμβανόμενα πρότυπα χρήσης και υψηλή στοχαστικότητα, που σχετίζονται με τη συμπεριφορά των χρηστών και τη λειτουργία επιμέρους συσκευών. Η ιδιαιτερότητα αυτή καθιστά ιδιαίτερα δύσκολη όχι μόνο τη μείωση των συνολικών σφαλμάτων πρόβλεψης, αλλά και την ακριβή αποτύπωση της χρονικής κατανομής της κατανάλωσης και των βραχυχρόνιων αιχμών φορτίου, ακόμη και όταν χρησιμοποιούνται προηγμένες τεχνικές μηχανικής μάθησης και βαθιάς μάθησης.

Παρ' όλα αυτά, οι προβλέψεις σε οικιακό επίπεδο μπορούν να τροφοδοτήσουν κρίσιμες εφαρμογές, όπως διαχείριση ζήτησης σε επίπεδο νοικοκυριού ή μικροδικτύου (Li, Tu, Feng, & Zou, 2025), σχεδιασμό και λειτουργία συστημάτων αποθήκευσης και μικρής κλίμακας ανανεώσιμων πηγών (Kasprzyk, Pekała, Oreshkin, & Dudek, 2025), καθώς και εφαρμογές ανατροφοδότησης και συμβουλών στους χρήστες που συνδυάζουν λεπτομερείς μετρήσεις με ψηφιακά εργαλεία ώστε να υποστηρίξουν την κατανόηση και βελτιστοποίηση της κατανάλωσης (Pullinger, Kilgour, Skatova, Frontuto, Krüger, Shi, Dunlop, & Staddon, 2021).

Η ανάπτυξη ακριβών και αποδοτικών μοντέλων πρόβλεψης σε αυτό το επίπεδο απαιτεί την αξιοποίηση τόσο των χρονοσειρών κατανάλωσης όσο και των μεταδεδομένων κτηρίου/νοικοκυριού (τύπος κατοικίας, εμβαδόν, αριθμός ενοίκων, βαθμός αστικοποίησης), καθώς και εξωγενών παραγόντων όπως η εξωτερική θερμοκρασία και τα ημερολογιακά χαρακτηριστικά (ημέρα εβδομάδας, Σαββατοκύριακα, αργίες). Παράλληλα, η ταχεία εξέλιξη της μηχανικής μάθησης και της βαθιάς μάθησης έχει οδηγήσει σε νέες προσεγγίσεις πρόβλεψης φορτίου, όπως συνδυαστικών δένδρων αλγορίθμων, υβριδικά σχήματα στατιστικών μεθόδων και μηχανικής μάθησης, επαναληπτικά νευρωνικά δίκτυα τύπου LSTM, συνελκτικά δίκτυα (CNN) και πιο πρόσφατα αρχιτεκτονικές τύπου Transformers και N-BEATS (Hong & Fan, 2016; Kasprzyk et al., 2025; Kim et al., 2025). Ειδικότερα για το οικιακό επίπεδο, ανασκοπήσεις στη διεθνή βιβλιογραφία καταγράφουν σαφή μετατόπιση από κλασικά στατιστικά και απλά νευρωνικά δίκτυα προς πιο εξελιγμένα μοντέλα μηχανικής μάθησης και βαθιάς μάθησης (ML/DL), με έμφαση στην ενσωμάτωση αβεβαιοτήτων, ειδικών χαρακτηριστικών καταναλωτή και εξωγενών παραγόντων (Ahmad, Mourshed, & Rezgui, 2018; Amasyali & El-Gohary, 2018; Runge & Zmeureanu, 2019; Olu-Ajayi et al., 2023).

Μέσα σε αυτό το πλαίσιο, η παρούσα διπλωματική εργασία αξιοποιεί δύο πραγματικά σύνολα δεδομένων οικιακής κατανάλωσης ενέργειας: το IDEAL household energy dataset, το οποίο περιλαμβάνει εκτεταμένες μετρήσεις από κατοικίες στο Ηνωμένο Βασίλειο και το σύνολο δεδομένων PLEGMA, το οποίο αφορά ελληνικές κατοικίες με διαφορετικά χαρακτηριστικά και συνθήκες χρήσης. Η παράλληλη αξιοποίηση των δύο συνόλων δεδομένων επιτρέπει τη διερεύνηση της απόδοσης των μοντέλων σε διαφορετικά ενεργειακά και γεωγραφικά πλαίσια και ενισχύει τη συγκριτική αξία των συμπερασμάτων.

1.2 Σκοπός και αντικείμενο της εργασίας

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη και η αξιολόγηση μοντέλων πρόβλεψης της βραχυπρόθεσμης ηλεκτρικής κατανάλωσης σε επίπεδο κατοικίας, αξιοποιώντας τεχνικές μηχανικής μάθησης και πραγματικά δεδομένα υψηλής χρονικής ανάλυσης.

Η εργασία εστιάζει στην πρόβλεψη της κατανάλωσης για χρονικό ορίζοντα 24 ωρών (day-ahead forecasting), με χρήση συνδυασμού χρονικών, περιβαλλοντικών και στατικών

χαρακτηριστικών των κατοικιών. Ιδιαίτερη έμφαση δίνεται στη διερεύνηση της συμβολής των ιστορικών δεδομένων κατανάλωσης, μέσω της σύγκρισης δύο βασικών σεναρίων πρόβλεψης: του σεναρίου χωρίς χρήση ιστορικών δεδομένων (no-history) και του σεναρίου με ενσωμάτωση ιστορικών χαρακτηριστικών (with-history).

Παράλληλα, διερευνώνται διαφορετικές στρατηγικές βελτίωσης των προβλέψεων, όπως τεχνικές μετα-επεξεργασίας (post-processing), προσεγγίσεις διόρθωσης των σφαλμάτων πρόβλεψης (residual correction), τεχνικές που δίνουν ιδιαίτερη έμφαση στην αποτύπωση των αιχμών κατανάλωσης (peak-aware techniques) και πιο σύνθετες αρχιτεκτονικές βαθιάς μάθησης. Οι στρατηγικές αυτές έχουν στόχο τη διερεύνηση των περιορισμών και των συμβιβασμών που εμφανίζονται μεταξύ διαφορετικών στόχων βελτιστοποίησης, όπως η μείωση των συνολικών σφαλμάτων, η ακριβής αποτύπωση αιχμών κατανάλωσης και η διατήρηση ρεαλιστικών ημερήσιων προφίλ φορτίου.

Η μεθοδολογική προσέγγιση βασίζεται κυρίως σε συνδυαστικά δένδροειδή μοντέλα μηχανικής μάθησης (tree-based ensembles) και συγκεκριμένα στα μοντέλα Random Forest, XGBoost και LightGBM, τα οποία έχουν αποδειχθεί ιδιαίτερα αποδοτικά σε προβλήματα πρόβλεψης με δομημένα δεδομένα (tabular data).

Για την υλοποίηση των παραπάνω, η εργασία αξιοποιεί δύο διαφορετικά σύνολα δεδομένων πραγματικών μετρήσεων: το σύνολο δεδομένων IDEAL, το οποίο περιλαμβάνει εκτεταμένες καταγραφές από κατοικίες στο Ηνωμένο Βασίλειο και το σύνολο δεδομένων PLEGMA, το οποίο αφορά ελληνικές κατοικίες με διαφορετικά χαρακτηριστικά και ενεργειακές συνθήκες. Η χρήση δύο συνόλων δεδομένων επιτρέπει τη διερεύνηση της απόδοσης των μοντέλων σε διαφορετικά περιβάλλοντα και τη σύγκριση της συμπεριφοράς τους υπό διαφορετικά επίπεδα πολυπλοκότητας και διαθεσιμότητας δεδομένων.

Στο πλαίσιο αυτό, η εργασία δεν περιορίζεται μόνο στην ανάπτυξη και αξιολόγηση των μοντέλων, αλλά επεκτείνεται και στην υλοποίηση μιας λειτουργικής εφαρμογής πρόβλεψης, η οποία περιλαμβάνει διεπαφή χρήστη (UI) και προγραμματιστική διεπαφή (API), επιτρέποντας την πρακτική αξιοποίηση των αποτελεσμάτων σε περιβάλλον που προσεγγίζει πραγματικές συνθήκες χρήσης.

1.3 Ερευνητικά ερωτήματα

Με βάση τον σκοπό της εργασίας, διατυπώνονται τα ακόλουθα ερευνητικά ερωτήματα, τα οποία καθοδηγούν τη μεθοδολογική προσέγγιση και την αξιολόγηση των αποτελεσμάτων:

- Σε ποιο βαθμό μπορούν τα μοντέλα μηχανικής μάθησης να προβλέψουν με ακρίβεια τη βραχυπρόθεσμη ηλεκτρική κατανάλωση σε επίπεδο κατοικίας;

Το ερώτημα αυτό εξετάζει τη συνολική ικανότητα των μοντέλων να αποτυπώνουν τόσο το επίπεδο όσο και τη χρονική κατανομή της κατανάλωσης, με χρήση πραγματικών δεδομένων.

- Ποια είναι η επίδραση της χρήσης ιστορικών δεδομένων στην απόδοση των μοντέλων πρόβλεψης;

Στο πλαίσιο αυτό, διερευνάται η διαφορά μεταξύ των σεναρίων σε σχέση με τη χρήση ιστορικών δεδομένων, καθώς και ο βαθμός στον οποίο τα ιστορικά χαρακτηριστικά συμβάλλουν στη βελτίωση της πρόβλεψης.

- Ποια χαρακτηριστικά εισόδου (χρονικά, περιβαλλοντικά, στατικά και ιστορικά) έχουν τη μεγαλύτερη συμβολή στην πρόβλεψη της κατανάλωσης;

Το ερώτημα αυτό εξετάζεται μέσω ανάλυσης σημασίας χαρακτηριστικών, με στόχο την κατανόηση των βασικών παραγόντων που επηρεάζουν τη συμπεριφορά της κατανάλωσης.

- Πώς διαφοροποιείται η απόδοση των μοντέλων μεταξύ διαφορετικών κατοικιών και ενεργειακών προφίλ;

Η διερεύνηση αυτή επιτρέπει την κατανόηση της επίδρασης της πολυπλοκότητας, του μεγέθους του νοικοκυριού και της ενεργειακής υποδομής στη συμπεριφορά των μοντέλων.

- Ποιοι περιορισμοί και ποιοι συμβιβασμοί εμφανίζονται κατά τη βελτίωση των προβλέψεων σε επίπεδο μεμονωμένης κατοικίας;

Το ερώτημα αυτό εξετάζει κατά πόσο η βελτίωση συγκεκριμένων χαρακτηριστικών της πρόβλεψης, όπως η αποτύπωση αιχμών φορτίου ή η μείωση της ρίζας του μέσου τετραγωνικού σφάλματος (RMSE), μπορεί να επηρεάζει αρνητικά άλλες μετρικές αξιολόγησης ή τη συνολική σταθερότητα των προβλέψεων.

- Σε ποιο βαθμό τα αποτελέσματα που προκύπτουν από ένα σύνολο δεδομένων μπορούν να μεταφερθούν ή να συγκριθούν με ένα διαφορετικό σύνολο δεδομένων;

Το ερώτημα αυτό αφορά τη συγκριτική ανάλυση μεταξύ των συνόλων δεδομένων IDEAL και PLEGMA και τη διερεύνηση της γενικευσιμότητας της προτεινόμενης μεθοδολογίας.

1.4 Μεθοδολογική προσέγγιση

Για να απαντηθούν τα ερευνητικά ερωτήματα, η εργασία ακολουθεί μία ολοκληρωμένη μεθοδολογία, η οποία εφαρμόζεται αρχικά στο σύνολο δεδομένων IDEAL και στη συνέχεια προσαρμόζεται και επαναλαμβάνεται στο σύνολο δεδομένων PLEGMA, επιτρέποντας τη συγκριτική αξιολόγηση των αποτελεσμάτων.

Η μεθοδολογία μπορεί να συνοψιστεί στα παρακάτω στάδια:

1. Συλλογή και προεπεξεργασία δεδομένων

- Εξαγωγή και επεξεργασία χρονοσειρών κατανάλωσης και περιβαλλοντικών μεταβλητών από τα σύνολα δεδομένων IDEAL και PLEGMA.
- Στο IDEAL, δημιουργία ωριαίων χρονοσειρών κατανάλωσης και θερμοκρασίας και ενοποίηση με μεταδεδομένα κατοικιών.
- Στο PLEGMA, συγχρονισμός δεδομένων υψηλής ανάλυσης (ηλεκτρικών και περιβαλλοντικών μετρήσεων), μετατροπή ισχύος σε ενέργεια και δημιουργία του τελικού συνόλου δεδομένων.

2. Διερευνητική ανάλυση δεδομένων και ανάπτυξη χαρακτηριστικών εισόδου

- Ανάλυση ημερήσιων και ωριαίων προφίλ κατανάλωσης, διαφορών μεταξύ τύπων κατοικίας και συσχέτισης με περιβαλλοντικούς παράγοντες.
- Σχεδιασμός και επιλογή χρονικών, περιβαλλοντικών, στατικών και ιστορικών χαρακτηριστικών.

3. Ανάπτυξη στατιστικών μοντέλων και μοντέλων μηχανικής μάθησης

- Υλοποίηση βασικών (baseline) μοντέλων αναφοράς.
- Εκπαίδευση μοντέλων μηχανικής μάθησης στα δύο σενάρια, δηλαδή τόσο χωρίς ιστορικά δεδομένα όσο και με χρήση ιστορικών χαρακτηριστικών

- Χρήση χρονικού διαχωρισμού των δεδομένων σε σύνολο εκπαίδευσης και ελέγχου.
- Αποθήκευση των εκπαιδευμένων μοντέλων και των συνοδευτικών αρχείων προεπεξεργασίας (artifacts) για χρήση στο στάδιο της υλοποίησης.
- Διερεύνηση εναλλακτικών προσεγγίσεων πρόβλεψης, όπως μοντέλα βαθιάς μάθησης, εξατομικευμένες στρατηγικές εκπαίδευσης και τεχνικές μεταγενέστερης διόρθωσης των προβλέψεων.

4. Συγκριτική αξιολόγηση και ανάλυση σφαλμάτων

- Αξιολόγηση των μοντέλων.
- Ανάλυση της απόδοσης σε διαφορετικά σενάρια και επίπεδα πληροφορίας.

5. Σχεδίαση και υλοποίηση συστήματος πρόβλεψης

- Ανάπτυξη FastAPI εξυπηρετητή που φορτώνει τα εκπαιδευμένα μοντέλα και υλοποιεί σημείο πρόσβασης για την εκτέλεση προβλέψεων επόμενης ημέρας.
- Δημιουργία γραφικής διεπαφής χρήστη για τη δοκιμή και αξιοποίηση των προβλέψεων.
- Υλοποίηση μηχανισμών διαχείρισης ιστορικών δεδομένων και περιβαλλοντικών εισόδων.

6. Αξιολόγηση προβλέψεων σε συνθήκες χρήσης

- Αξιολόγηση των προβλέψεων που παράγονται από το σύστημα (API και UI) μέσω σύγκρισης με πραγματικές μετρήσεις κατανάλωσης.
- Ανάλυση της απόδοσης σε ρεαλιστικά σενάρια χρήσης.
- Διερεύνηση της ακρίβειας των προβλέψεων τόσο ως προς τη χρονική κατανομή της κατανάλωσης όσο και ως προς τη συνολική ημερήσια κατανάλωση.
- Αξιολόγηση της συμπεριφοράς των μοντέλων σε επιλεγμένες αντιπροσωπευτικές κατοικίες και σενάρια.

1.5 Δομή της διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία οργανώνεται σε επιμέρους κεφάλαια, τα οποία καλύπτουν διαδοχικά το θεωρητικό υπόβαθρο, την ανάπτυξη της μεθοδολογίας, την εφαρμογή σε πραγματικά δεδομένα και την αξιολόγηση των αποτελεσμάτων.

Η δομή της εργασίας διαχωρίζεται ουσιαστικά σε δύο κύρια μέρη. Το πρώτο μέρος, το οποίο εκτείνεται από το Κεφάλαιο 3 έως το Κεφάλαιο 9, αφορά την ανάπτυξη και αξιολόγηση της προτεινόμενης μεθοδολογίας στο σύνολο δεδομένων IDEAL. Το δεύτερο μέρος, από το Κεφάλαιο 10 έως το Κεφάλαιο 16, αφορά την εφαρμογή της ίδιας μεθοδολογίας στο σύνολο δεδομένων PLEGMA, επιτρέποντας τη συγκριτική ανάλυση των αποτελεσμάτων μεταξύ διαφορετικών ενεργειακών πλαισίων.

Έτσι η υπόλοιπη εργασία οργανώνεται ως εξής:

- Στο Κεφάλαιο 2 παρουσιάζεται το θεωρητικό υπόβαθρο και η σχετική βιβλιογραφία γύρω από την πρόβλεψη φορτίου σε κτηριακό επίπεδο, τις τεχνικές μηχανικής

- μάθησης που εφαρμόζονται, καθώς και προηγούμενες εργασίες σε οικιακό περιβάλλον.
- Στο Κεφάλαιο 3 περιγράφονται αναλυτικά τα δεδομένα του IDEAL και η πλήρης διαδικασία προεπεξεργασίας και ενοποίησης.
 - Στο Κεφάλαιο 4 παρουσιάζεται η διερευνητική ανάλυση δεδομένων (EDA), η επιλογή χαρακτηριστικών και τα στατιστικά μοντέλα αναφοράς.
 - Στο Κεφάλαιο 5 αναλύεται η μεθοδολογία μοντελοποίησης, οι διαφορετικές στρατηγικές πρόβλεψης και οι προσεγγίσεις βελτίωσης της ακρίβειας των μοντέλων.
 - Στο Κεφάλαιο 6 παρουσιάζονται τα πειραματικά αποτελέσματα, οι συγκριτικές αξιολογήσεις μεταξύ διαφορετικών προσεγγίσεων πρόβλεψης καθώς και η ανάλυση των περιορισμών και των συμβιβασμών που εμφανίζονται στη βραχυπρόθεσμη πρόβλεψη οικιακής κατανάλωσης.
 - Στο Κεφάλαιο 7 περιγράφεται η υλοποίηση του παραγωγικού API και της συνοδευτικής γραφικής διεπαφής χρήστη.
 - Στο Κεφάλαιο 8 παρουσιάζεται η εφαρμογή, η διεπαφή χρήστη και τα σενάρια χρήσης του συστήματος πρόβλεψης.
 - Στο Κεφάλαιο 9 πραγματοποιείται αξιολόγηση των προβλέψεων σε περιβάλλον εφαρμογής για το IDEAL.
 - Στο Κεφάλαιο 10 παρουσιάζεται το σύνολο δεδομένων PLEGMA και ο ρόλος του στην εργασία.
 - Στο Κεφάλαιο 11 περιγράφεται η διαδικασία προεπεξεργασίας, συγχρονισμού και ενοποίησης των δεδομένων του PLEGMA.
 - Στο Κεφάλαιο 12 παρουσιάζεται η διερευνητική ανάλυση δεδομένων (EDA) για το PLEGMA dataset.
 - Στο Κεφάλαιο 13 περιγράφεται η διαδικασία μοντελοποίησης για το σύνολο δεδομένων PLEGMA.
 - Στο Κεφάλαιο 14 παρουσιάζονται τα αποτελέσματα των μοντέλων και η αξιολόγησή τους.
 - Στο Κεφάλαιο 15 περιγράφεται η υλοποίηση του συστήματος πρόβλεψης για το PLEGMA.
 - Στο Κεφάλαιο 16 πραγματοποιείται αναλυτική αξιολόγηση σε επίπεδο μεμονωμένων κατοικιών.
 - Τέλος, στο Κεφάλαιο 17 παρουσιάζονται η συζήτηση των αποτελεσμάτων, τα βασικά συμπεράσματα της εργασίας, καθώς και προτάσεις για μελλοντική έρευνα.

Ο πηγαίος κώδικας που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας, συμπεριλαμβανομένων των αρχείων Python και των Jupyter notebooks που χρησιμοποιήθηκαν για την εκπαίδευση, αξιολόγηση και υλοποίηση του συστήματος, έχει αναρτηθεί σε δημόσιο αποθετήριο GitHub στον ακόλουθο σύνδεσμο: [GitHub - stdGeorgios/Forecasting-the-electricity-consumption-of-buildings-: Thesis](https://github.com/stdGeorgios/Forecasting-the-electricity-consumption-of-buildings-:Thesis).

Στο ίδιο αποθετήριο διατίθεται επίσης, ως ξεχωριστό συνοδευτικό αρχείο, η αναλυτική τεχνική περιγραφή της υλοποίησης του συστήματος πρόβλεψης. Η περιγραφή αυτή περιλαμβάνει λεπτομέρειες σχετικά με την αρχιτεκτονική των μοντέλων, τη διαδικασία εκπαίδευσης, τη δημιουργία των επιμέρους αρχείων και αντικειμένων που απαιτούνται για την εκτέλεση του συστήματος, καθώς και τη λειτουργία του API και της γραφικής διεπαφής χρήστη. Η τεχνική ανάλυση βασίζεται κυρίως στην υλοποίηση για το σύνολο δεδομένων IDEAL, ενώ αντίστοιχη λογική ακολουθήθηκε και για την προσαρμογή του συστήματος στο σύνολο δεδομένων PLEGMA.

Στο **Παράρτημα Α** παρατίθενται οι βασικές βιβλιοθήκες της Python που χρησιμοποιήθηκαν, καθώς και σύντομη περιγραφή της δομής του αποθετηρίου.

2. Θεωρητικό υπόβαθρο

Το κεφάλαιο αυτό παρουσιάζει το θεωρητικό και βιβλιογραφικό υπόβαθρο στο οποίο βασίζεται η διπλωματική εργασία. Αρχικά, αναλύεται η έννοια της πρόβλεψης ενεργειακής κατανάλωσης σε κτηριακό επίπεδο και διευκρινίζεται η σημασία της βραχυπρόθεσμης πρόβλεψης ηλεκτρικού φορτίου σε οικιακό περιβάλλον. Στη συνέχεια, παρουσιάζονται οι βασικές κατηγορίες μοντέλων πρόβλεψης, με έμφαση στη διάκριση μεταξύ φυσικών, υβριδικών και δεδομενοκεντρικών προσεγγίσεων. Ιδιαίτερη έμφαση δίνεται στη δεδομενοκεντρική προσέγγιση, καθώς αυτή αποτελεί τη μεθοδολογική βάση της παρούσας εργασίας. Επιπλέον, εξετάζονται τα βασικά χαρακτηριστικά εισόδου που χρησιμοποιούνται σε προβλήματα πρόβλεψης κατανάλωσης, όπως τα χρονικά, περιβαλλοντικά, στατικά και ιστορικά χαρακτηριστικά, καθώς και ο ρόλος της διερευνητικής ανάλυσης δεδομένων στην επιλογή τους. Τέλος, παρουσιάζεται η σημασία των βασικών μοντέλων αναφοράς και η ανάγκη ενσωμάτωσης των μοντέλων πρόβλεψης σε λειτουργικά συστήματα και διεπαφές εφαρμογών, ώστε τα ερευνητικά αποτελέσματα να μπορούν να αξιοποιηθούν σε πρακτικά περιβάλλοντα ενεργειακής διαχείρισης.

2.1 Πρόβλεψη ενεργειακής κατανάλωσης κτηρίων

Η πρόβλεψη της ενεργειακής κατανάλωσης σε κτηριακό επίπεδο αποτελεί κρίσιμο εργαλείο για τον σχεδιασμό πολιτικών εξοικονόμησης, την ενεργειακή αναβάθμιση κτηρίων και τη λειτουργία συστημάτων διαχείρισης ενέργειας (Energy Management Systems, EMS) (Ahmad, Mourshed, & Rezgui, 2018). Στη διεθνή βιβλιογραφία γίνεται συνήθως διάκριση μεταξύ βραχυπρόθεσμης (short-term), μεσοπρόθεσμης (medium-term) και μακροπρόθεσμης (long-term) πρόβλεψης φορτίου, ανάλογα με τον χρονικό ορίζοντα και τη χρήση των αποτελεσμάτων (Hong & Fan, 2016). Η βραχυπρόθεσμη πρόβλεψη (από λίγα λεπτά έως λίγες ημέρες μπροστά) χρησιμοποιείται κυρίως για λειτουργικές αποφάσεις, όπως ο προγραμματισμός μονάδων, η διαχείριση ζήτησης και η βελτιστοποίηση αποθήκευσης, ενώ η μεσοπρόθεσμη και μακροπρόθεσμη πρόβλεψη συνδέονται περισσότερο με χωροταξικό σχεδιασμό, επενδυτικές αποφάσεις και πολιτικές ενεργειακής αποδοτικότητας (Hong & Fan, 2016; Olu-Ajayi, Alaka, Owolabi, Akanbi, & Ganiyu, 2023). Σε επίπεδο κτηρίου, η πρόβλεψη μπορεί να στοχεύει είτε στην ολική ενεργειακή κατανάλωση (π.χ. ηλεκτρική ενέργεια ή συνδυασμός ηλεκτρισμού και θέρμανσης) είτε σε επιμέρους ενεργειακές χρήσεις (θέρμανση, ψύξη, φωτισμός κ.λπ.), ανάλογα με τη διαθεσιμότητα μετρήσεων και τις ανάγκες της εφαρμογής (Ahmad et al., 2018; Hong & Fan, 2016). Επιπλέον, διακρίνονται διαφορετικά επίπεδα χωρικής ανάλυσης, από ολόκληρα δίκτυα διανομής και υποσταθμούς μέχρι επιμέρους κτήρια και μεμονωμένα νοικοκυριά. Όσο μικρότερη είναι η κλίμακα, τόσο μεγαλύτερη είναι η σχετική τυχαία μεταβλητότητα και τόσο μεγαλύτερη η σημασία της ενσωμάτωσης πληροφοριών για τα χαρακτηριστικά του κτηρίου και των ενοίκων (Ahmad et al., 2018; Olu-Ajayi et al., 2023).

Η παρούσα εργασία εστιάζει στη βραχυπρόθεσμη πρόβλεψη ηλεκτρικής κατανάλωσης σε επίπεδο μεμονωμένης κατοικίας, με ορίζοντα επόμενης ημέρας σε ωριαία ανάλυση. Αυτό το πρόβλημα βρίσκεται στο πιο απαιτητικό άκρο του φάσματος, καθώς τα οικιακά φορτία επηρεάζονται έντονα από την τυχαία και μη τυποποιημένη συμπεριφορά των νοικοκυριών, από το μικρό μέγεθος δείγματος και από υψηλά επίπεδα θορύβου στα δεδομένα (Kong, Dong, Hill, Moy, & Chapman, 2019).

Η αυξημένη δυσκολία της πρόβλεψης σε επίπεδο μεμονωμένης κατοικίας οφείλεται όχι μόνο στη μικρότερη κλίμακα συγκέντρωσης του φορτίου, αλλά και στην παρουσία βραχυχρόνιων μεταβολών που σχετίζονται με τη λειτουργία επιμέρους συσκευών και με μη προβλέψιμες ανθρώπινες δραστηριότητες. Ως αποτέλεσμα, η επίτευξη χαμηλών σφαλμάτων πρόβλεψης δεν συνεπάγεται απαραίτητα και ακριβή αποτύπωση των χρονικών αιχμών κατανάλωσης ή της ημερήσιας μορφής του φορτίου.

2.2 Κατηγορίες μοντέλων πρόβλεψης

Τα μοντέλα πρόβλεψης ενεργειακής κατανάλωσης σε κτήρια ταξινομούνται συχνά σε τρεις βασικές κατηγορίες: φυσικά ή «λευκού κουτιού» (white-box), δομικά/υβριδικά ή «γκρι κουτιού» (grey-box) και δεδομενοκεντρικά ή «μαύρου κουτιού» (black-box) (Ahmad et al., 2018).

Τα μοντέλα «λευκού κουτιού» βασίζονται σε φυσικές εξισώσεις θερμικής και ηλεκτρικής συμπεριφοράς του κτηρίου και συχνά υλοποιούνται σε περιβάλλοντα προσομοίωσης, όπως EnergyPlus ή TRNSYS (Ahmad et al., 2018). Παρέχουν υψηλό επίπεδο ερμηνευσιμότητας και μπορούν να χρησιμοποιηθούν για ανάλυση σεναρίων (what-if), αλλά απαιτούν λεπτομερή δεδομένα γεωμετρίας, δομικών στοιχείων και συστημάτων, καθώς και εξειδικευμένη γνώση για τη βαθμονόμησή τους.

Τα μοντέλα «γκρι κουτιού» συνδυάζουν απλοποιημένες φυσικές εξισώσεις με στοχαστικούς ή δεδομενοκεντρικούς όρους με παραμέτρους που εκτιμώνται από δεδομένα. Προσφέρουν έναν συμβιβασμό μεταξύ ερμηνευσιμότητας και ευελιξίας, αλλά παραμένουν σχετικά πολύπλοκα στην υλοποίηση.

Τέλος, τα μοντέλα «μαύρου κουτιού» ή δεδομενοκεντρικά (data-driven) μοντέλα βασίζονται αποκλειστικά σε ιστορικά δεδομένα εισόδου-εξόδου, χωρίς ρητή ενσωμάτωση φυσικών εξισώσεων. Στην κατηγορία αυτή ανήκουν τα κλασικά στατιστικά μοντέλα (γραμμική παλινδρόμηση, ARIMA), καθώς και οι μέθοδοι μηχανικής μάθησης και βαθιάς μάθησης, όπως Support Vector Regression, Random Forest, Gradient Boosting και τεχνητά νευρωνικά δίκτυα (Olu-Ajayi et al., 2023; Hong & Fan, 2016). Σύγχρονες ανασκοπήσεις δείχνουν ότι, για την πρόβλεψη ηλεκτρικής κατανάλωσης σε κτήρια, τα δεδομενοκεντρικά μοντέλα επιτυγχάνουν συχνά καλύτερη προγνωστική απόδοση από τα κλασικά στατιστικά μοντέλα, ιδίως όταν είναι διαθέσιμα επαρκή δεδομένα και κατάλληλα χαρακτηριστικά (Ahmad et al., 2018; Olu-Ajayi et al., 2023).

Στο πλαίσιο της παρούσας εργασίας υιοθετείται καθαρά δεδομενοκεντρική προσέγγιση, με χρήση μοντέλων μηχανικής μάθησης που εκπαιδεύονται πάνω σε ωριαίες χρονοσειρές κατανάλωσης και σε σύνολο στατικών, χρονικών και κλιματικών χαρακτηριστικών. Η επιλογή αυτή ευθυγραμμίζεται με τη σύγχρονη τάση στη βιβλιογραφία, όπου τα συνδυαστικά δένδροειδή μοντέλα και τα νευρωνικά δίκτυα αξιοποιούνται ευρέως για τη βραχυπρόθεσμη πρόβλεψη φορτίου (Short-Term Load Forecasting – STLF) σε κτηριακό και οικιακό επίπεδο (Kong et al., 2019; Ahmad et al., 2018; Olu-Ajayi et al., 2023).

2.3 Δεδομενοκεντρική προσέγγιση πρόβλεψης σε οικιακό περιβάλλον

Η πρόβλεψη φορτίου σε επίπεδο μεμονωμένων κατοικιών διαφέρει ουσιαστικά από την πρόβλεψη σε επίπεδο συστήματος ή μεγάλων κτηρίων, κυρίως λόγω της υψηλής

ετερογένειας μεταξύ νοικοκυριών και της ισχυρής επίδρασης του ανθρώπινου παράγοντα (Kong et al., 2019). Σε αντίθεση με τα μεγάλα εμπορικά κτήρια, όπου τα προγράμματα λειτουργίας και τα φορτία είναι συχνά πιο προβλέψιμα, τα οικιακά φορτία χαρακτηρίζονται από αιχμές μικρής διάρκειας, έντονη εποχικότητα και συχνές μεταβολές στη σύνθεση και συμπεριφορά των νοικοκυριών (Ahmad et al., 2018; Olu-Ajayi et al., 2022).

Πρόσφατες μελέτες στη βραχυπρόθεσμη πρόβλεψη φορτίου σε κατοικίες (residential STLF) αξιοποιούν κυρίως τρεις κατηγορίες εισόδων (Ahmad et al., 2018; Olu-Ajayi et al., 2023). Πρώτον, χρονικά χαρακτηριστικά, όπως ώρα, ημέρα εβδομάδας, ένδειξη Σαββατοκύριακου και αργιών, καθώς και εποχικοί δείκτες (μήνας, εποχή), ώστε τα μοντέλα να αποτυπώνουν ημερήσια και εβδομαδιαία μοτίβα (Kong et al., 2019). Δεύτερον, κλιματικά χαρακτηριστικά, κυρίως εξωτερική θερμοκρασία και, σε ορισμένες περιπτώσεις, υγρασία ή ηλιακή ακτινοβολία, τα οποία επηρεάζουν άμεσα τις ανάγκες θέρμανσης/ψύξης (Ahmad et al., 2018; Li, Tu, Feng, & Zou, 2025). Τρίτον, στατικά χαρακτηριστικά κτηρίου/νοικοκυριού, όπως τύπος κατοικίας, εμβαδόν, αριθμός ενοίκων και δείκτες αστικότητας, που συσχετίζονται με το μέσο επίπεδο κατανάλωσης και τη μορφή των προφίλ φορτίου (Olu-Ajayi et al., 2023; Pullinger et al., 2021).

Σε ό,τι αφορά τα μοντέλα, οι Kong et al. (2019) έδειξαν ότι ένα LSTM μοντέλο για μεμονωμένα νοικοκυριά μπορεί να υπερτερεί σημαντικά κλασικών μοντέλων όπως ARIMA, προωθητικά νευρωνικά δίκτυα (feed-forward neural networks) και η μέθοδος των k-πλησιέστερων γειτόνων (k-nearest neighbours), ιδιαίτερα σε ασταθή προφίλ κατανάλωσης. Άλλες εργασίες αξιοποιούν συνδυαστικά σύνολα δένδρων μοντέλων, όπως Random Forest και Gradient Boosting, ιδίως όταν το διαθέσιμο δείγμα σε επίπεδο κατοικιών είναι σχετικά μικρό αλλά οι εξηγητικές μεταβλητές είναι πολλές (Ahmad et al., 2018; Olu-Ajayi et al., 2023).

Τα τελευταία χρόνια εμφανίζονται επίσης πιο σύνθετες αρχιτεκτονικές βαθιάς μάθησης για χρονοσειρές, όπως συνελκτικά χρονικά δίκτυα, αρχιτεκτονικές βασισμένες σε μηχανισμούς προσοχής και μοντέλα τύπου Transformers, τα οποία επιχειρούν να αποτυπώσουν πιο πολύπλοκες χρονικές εξαρτήσεις και αλληλεπιδράσεις μεταξύ μεταβλητών (Kim et al., 2025; Kasprzyk et al., 2025). Παρά τη σημαντική πρόοδο των μοντέλων αυτών, η εφαρμογή τους σε επίπεδο μεμονωμένων κατοικιών παραμένει απαιτητική λόγω της υψηλής μεταβλητότητας και της περιορισμένης διαθεσιμότητας δεδομένων ανά νοικοκυριό.

Παρότι οι σύγχρονες αρχιτεκτονικές βαθιάς μάθησης έχουν επιτύχει εντυπωσιακά αποτελέσματα σε διάφορα προβλήματα χρονοσειρών, αρκετές μελέτες επισημαίνουν ότι η υπεροχή τους δεν είναι καθολική. Σε προβλήματα με περιορισμένο αριθμό παρατηρήσεων ανά οντότητα, έντονο θόρυβο και πλούσιο σύνολο δομημένων χαρακτηριστικών, τα συνδυαστικά δένδρα μοντέλα εξακολουθούν να αποτελούν ιδιαίτερα ανταγωνιστικές λύσεις, προσφέροντας συχνά καλύτερη ισορροπία μεταξύ ακρίβειας, ερμηνευσιμότητας και υπολογιστικού κόστους.

Ένα επαναλαμβανόμενο συμπέρασμα των παραπάνω μελετών είναι ότι δεν υπάρχει ένα καθολικά βέλτιστο εργαλείο, αλλά η απόδοση κάθε μοντέλου εξαρτάται από τα χαρακτηριστικά των δεδομένων, τον χρονικό ορίζοντα και τον συνδυασμό χαρακτηριστικών που χρησιμοποιούνται (Olu-Ajayi et al., 2023). Γι' αυτόν τον λόγο, πολλές πρόσφατες εργασίες προτείνουν την παράλληλη αξιολόγηση διαφορετικών δεδομενοκεντρικών εργαλείων (π.χ. SVM, RF, GBT, LSTM) σε συγκεκριμένο σύνολο δεδομένων, με χρήση τυποποιημένων μετρικών σφάλματος (MAE, RMSE, MAPE) και με έμφαση στη δυνατότητα γενίκευσης σε νέα κτήρια ή νοικοκυριά (Ahmad et al., 2018; Hong & Fan, 2016; Runge & Zmeureanu, 2019).

2.4 Χαρακτηριστικά εισόδου και διερευνητική ανάλυση (EDA)

Η επιλογή και ο κατάλληλος σχεδιασμός των χαρακτηριστικών εισόδου (feature engineering) αποτελούν κρίσιμο στάδιο στην πρόβλεψη ενεργειακής κατανάλωσης, ιδίως όταν χρησιμοποιούνται δεδομενοκεντρικά μοντέλα (Ahmad et al., 2018; Olu-Ajayi et al., 2023). Στη βιβλιογραφία οι είσοδοι των μοντέλων χωρίζονται συνήθως σε τρεις βασικές κατηγορίες: χρονικά, κλιματικά και στατικά χαρακτηριστικά.

Τα χρονικά χαρακτηριστικά περιλαμβάνουν την ώρα της ημέρας, την ημέρα της εβδομάδας, δείκτες Σαββατοκύριακου και αργιών, καθώς και εποχικούς δείκτες (μήνας, εποχή). Αυτά τα χαρακτηριστικά επιτρέπουν στα μοντέλα να αποτυπώσουν ημερήσια και εβδομαδιαία περιοδικότητα, όπως πρωινές και βραδινές αιχμές ή διαφορές μεταξύ καθημερινών και Σαββατοκύριακου (Kong et al., 2019).

Ιδιαίτερη σημασία στη βραχυπρόθεσμη πρόβλεψη φορτίου έχουν επίσης τα ιστορικά χαρακτηριστικά κατανάλωσης, όπως προηγούμενες τιμές φορτίου και στατιστικά χαρακτηριστικά κινούμενων παραθύρων. Τα χαρακτηριστικά αυτά επιτρέπουν στα μοντέλα να αξιοποιούν τη χρονική συνέχεια και την αυτοσυσχέτιση της κατανάλωσης και συχνά αποτελούν τους σημαντικότερους παράγοντες πρόβλεψης σε προβλήματα οικιακής κατανάλωσης (Hong & Fan, 2016; Ahmad et al., 2018; Kong et al., 2019).

Τα κλιματικά χαρακτηριστικά, όπως η εξωτερική και εσωτερική θερμοκρασία όπου είναι διαθέσιμη, χρησιμοποιούνται ευρέως για την πρόβλεψη ηλεκτρικής κατανάλωσης σε κτήρια, καθώς συνδέονται άμεσα με τις ανάγκες θέρμανσης, ψύξης και αερισμού. Πολλές μελέτες δείχνουν ότι η ενσωμάτωση της εξωτερικής θερμοκρασίας βελτιώνει σημαντικά την πιστότητα των προβλέψεων, ιδίως σε κλίμακα κατοικιών, όπου η χρήση θερμαντικών/ψυκτικών συστημάτων προκαλεί έντονες εποχικές διακυμάνσεις (Ahmad et al., 2018).

Τα στατικά χαρακτηριστικά κτηρίου και νοικοκυριού περιλαμβάνουν τον τύπο κατοικίας, το εμβαδόν, τον αριθμό ενοίκων και δείκτες αστικότητας. Αυτά τα χαρακτηριστικά συσχετίζονται με το μέσο επίπεδο κατανάλωσης και τη μορφή των προφίλ φορτίου. Για παράδειγμα, μεγαλύτερες κατοικίες ή σπίτια σε αγροτικές περιοχές τείνουν να εμφανίζουν υψηλότερη κατανάλωση και πιο έντονη εξάρτηση από τις καιρικές συνθήκες (Olu-Ajayi et al., 2023; Pullinger et al., 2021).

Σημαντική παράμετρος για την επιλογή χαρακτηριστικών είναι η διερευνητική ανάλυση δεδομένων (EDA). Μέσω διερευνητικής ανάλυσης δεδομένων, οι ερευνητές εξετάζουν ημερήσια και ωριαία προφίλ κατανάλωσης, διαφορές μεταξύ κατηγοριών κατοικιών και σχέσεις με τη θερμοκρασία, προκειμένου να αναγνωρίσουν πρότυπα και να αποφασίσουν ποια χαρακτηριστικά έχουν ουσιαστική προγνωστική αξία (Hong & Fan, 2016; Ahmad et al., 2018). Ανασκοπήσεις σε προβλέψεις ενεργειακής κατανάλωσης κτηρίων δείχνουν ότι ο προσεκτικός συνδυασμός διερευνητικής ανάλυσης δεδομένων και τεχνικών επιλογής χαρακτηριστικών (filter, wrapper, embedded methods), μπορεί να βελτιώσει αισθητά την απόδοση των μοντέλων, αν και η υπερβολικά πολύπλοκη διαδικασία επιλογής χαρακτηριστικών (feature selection), δεν αποδίδει πάντα ανάλογα οφέλη σε σχέση με το υπολογιστικό κόστος (Ahmad et al., 2018; Olu-Ajayi et al., 2023).

Στην παρούσα εργασία, η επιλογή χαρακτηριστικών ευθυγραμμίζεται με τις παραπάνω κατευθύνσεις: αξιοποιούνται συστηματικά χρονικές, κλιματικές και στατικές μεταβλητές, ενώ η χρησιμότητά τους τεκμηριώνεται μέσω διερευνητικής ανάλυσης δεδομένων πάνω στα σύνολα δεδομένων και μέσω της σύγκρισης διαφορετικών μοντέλων πρόβλεψης.

2.5 Βασικά μοντέλα πρόβλεψης

Πριν την εφαρμογή σύνθετων μοντέλων μηχανικής μάθησης, θεωρείται καλό πρακτικό βήμα η καθιέρωση απλών στατιστικών μοντέλων, τα οποία παρέχουν ένα σημείο αναφοράς για την αξιολόγηση της προστιθέμενης αξίας των πιο περίπλοκων προσεγγίσεων (Hong & Fan, 2016). Τέτοια βασικά μοντέλα αναφοράς χρησιμοποιούνται ευρέως σε μελέτες πρόβλεψης φορτίου, αλλά και σε γενικότερες εφαρμογές χρονοσειρών (Hyndman & Athanasopoulos, 2018).

Στο πλαίσιο της πρόβλεψης ηλεκτρικού φορτίου, κλασικά μοντέλα αναφοράς είναι το απλό μοντέλο (naive), όπου η πρόβλεψη για την επόμενη χρονική στιγμή ισούται με την τελευταία παρατήρηση, το εποχικό απλό μοντέλο (seasonal naive), όπου η πρόβλεψη για μία χρονική στιγμή ισούται με την παρατήρηση της αντίστοιχης στιγμής της προηγούμενης ημέρας ή εβδομάδας, καθώς και κινούμενοι μέσοι όροι (moving averages) σε σταθερά παράθυρα (π.χ. 24 ή 168 ωρών) (Hong & Fan, 2016; Hyndman & Athanasopoulos, 2018). Σε πιο εξειδικευμένες εφαρμογές, βασικά μοντέλα αναφοράς μπορούν επίσης να θεωρηθούν γραμμικά μοντέλα παλινδρόμησης με ελάχιστα χαρακτηριστικά (π.χ. μόνο θερμοκρασία και ώρα) ή απλές εκδόσεις μοντέλων όπως ARIMA, τα οποία χαρακτηρίζονται από χαμηλή πολυπλοκότητα, αλλά παρουσιάζουν περιορισμένη ικανότητα αποτύπωσης μη γραμμικών σχέσεων (Ahmad et al., 2018; Olu-Ajayi et al., 2023).

Ο ρόλος των βασικών μοντέλων αναφοράς δεν είναι να επιτύχουν την καλύτερη δυνατή απόδοση, αλλά να λειτουργήσουν ως σημείο αναφοράς. Ένα μοντέλο μηχανικής μάθησης που δεν μπορεί να ξεπεράσει σαφώς τα βασικά μοντέλα αναφοράς σε σταθερές μετρικές (MAE, RMSE, MAPE) δύσκολα δικαιολογεί την πρόσθετη πολυπλοκότητα και υπολογιστική επιβάρυνση (Hong & Fan, 2016). Για τον λόγο αυτό, η παρούσα εργασία ορίζει και υλοποιεί βασικά στατιστικά και εμπειρικά μοντέλα αναφοράς σε ωριαία κλίμακα, τα οποία χρησιμοποιούνται συστηματικά στη σύγκριση με τα πιο σύνθετα μοντέλα μηχανικής μάθησης.

Η αξιολόγηση μοντέλων σε επίπεδο μεμονωμένης κατοικίας απαιτεί συνδυαστική εξέταση διαφορετικών μετρικών σφάλματος, καθώς κάθε μετρική αποτυπώνει διαφορετικές πτυχές της ποιότητας της πρόβλεψης. Για παράδειγμα, ορισμένες μετρικές είναι περισσότερο ευαίσθητες σε μεγάλες αποκλίσεις ή αιχμές κατανάλωσης, ενώ άλλες αποτυπώνουν καλύτερα τη μέση συμπεριφορά του μοντέλου (Hong & Fan, 2016; Ahmad et al., 2018).

Επιπλέον, η χρήση πολλαπλών μετρικών αξιολόγησης είναι ιδιαίτερα σημαντική σε προβλήματα οικιακής κατανάλωσης, καθώς διαφορετικά μοντέλα μπορεί να υπερέχουν σε διαφορετικές πτυχές της πρόβλεψης, όπως η μέση ακρίβεια, η αποτύπωση αιχμών φορτίου ή η εκτίμηση της συνολικής ημερήσιας κατανάλωσης.

2.6 Παραγωγικά συστήματα και APIs για πρόβλεψη φορτίου

Η βιβλιογραφία για την πρόβλεψη φορτίου εστιάζει παραδοσιακά στην ακρίβεια των μοντέλων, ωστόσο τα τελευταία χρόνια αυξάνεται το ενδιαφέρον για την παραγωγική ενσωμάτωση των μοντέλων σε πραγματικά συστήματα, μέσω διαδικτυακών υπηρεσιών (web services) και διεπαφών προγραμματισμού εφαρμογών ή εν συντομία APIs (Olu-Ajayi et al., 2023). Στο πλαίσιο έξυπνων δικτύων και συστημάτων διαχείρισης ενέργειας, οι προβλέψεις φορτίου χρησιμοποιούνται από άλλα υποσυστήματα – όπως αλγορίθμους

βελτιστοποίησης, εργαλεία σχεδιασμού υποδομών ή πλατφόρμες αλληλεπίδρασης με τον χρήστη – τα οποία χρειάζονται μια τυποποιημένη διεπαφή για να ζητούν και να λαμβάνουν προβλέψεις.

Τα RESTful APIs αποτελούν την επικρατούσα τεχνολογία για τέτοιες διεπαφές: ένα API μπορεί να δέχεται αιτήματα (π.χ. σε μορφή JSON) που περιλαμβάνουν ιστορικά δεδομένα κατανάλωσης και εξωγενή χαρακτηριστικά (π.χ. πρόγνωση καιρού, ημερολογιακές πληροφορίες) και να επιστρέφει προβλέψεις φορτίου για συγκεκριμένο χρονικό ορίζοντα. Με τον τρόπο αυτό, το μοντέλο πρόβλεψης, ενσωματωμένο μέσα σε μια υπηρεσία, μπορεί να χρησιμοποιηθεί από πολλές διαφορετικές εφαρμογές, χωρίς να χρειάζεται κάθε φορά να επανυλοποιείται η λογική του.

Παράλληλα, αρκετά πρόσφατα έργα επισημαίνουν τη σημασία της αναπαραγωγιμότητας και της συντηρησιμότητας των μοντέλων, μέσω της αποθήκευσης όχι μόνο των βαρών του μοντέλου αλλά και των συνοδευτικών αρχείων (artifacts), όπως οι στήλες χαρακτηριστικών, οι τιμές κανονικοποίησης, οι παράμετροι συμπλήρωσης ελλিপών τιμών (imputation) και τα αρχεία ρυθμίσεων εκπαίδευσης (Hong & Fan, 2016; Ahmad et al., 2018). Η χρήση τέτοιων συνοδευτικών αρχείων είναι απαραίτητη ώστε το API να εφαρμόζει ακριβώς την ίδια διαδικασία προεπεξεργασίας που χρησιμοποιήθηκε κατά την εκπαίδευση, εξασφαλίζοντας συνέπεια μεταξύ συμπεριφοράς κατά την εκπαίδευση και κατά τη χρήση του συστήματος.

Η παρούσα εργασία ακολουθεί αυτή τη λογική, υλοποιώντας έναν διακομιστή εφαρμογής με χρήση του FastAPI, ο οποίος αξιοποιεί τα εκπαιδευμένα μοντέλα και τα συνοδευτικά αρχεία τους και παρέχει σημείο πρόσβασης για την εκτέλεση πρόβλεψης επόμενης ημέρας σε επίπεδο κατοικίας. Η ύπαρξη μιας απλής γραφικής διεπαφής επιτρέπει επιπλέον τη χρήση της υπηρεσίας από μη τεχνικούς χρήστες, διευκολύνοντας την αξιολόγηση και την πιθανή ενσωμάτωση σε πραγματικά πληροφοριακά συστήματα διαχείρισης ενέργειας.

3. Δεδομένα και Προεπεξεργασία

Στο κεφάλαιο αυτό παρουσιάζεται το σύνολο δεδομένων IDEAL και η διαδικασία προετοιμασίας του για τις ανάγκες της εργασίας. Αρχικά, περιγράφονται τα βασικά χαρακτηριστικά του συνόλου δεδομένων, το είδος των διαθέσιμων μετρήσεων και τα μεταδεδομένα που το συνοδεύουν. Στη συνέχεια, αναλύεται η διαδικασία επιλογής και φιλτραρίσματος των κατοικιών, με στόχο τη διατήρηση αξιόπιστων χρονοσειρών κατάλληλων για εκπαίδευση και αξιολόγηση μοντέλων πρόβλεψης. Ακολουθεί η παρουσίαση της ροής προεπεξεργασίας, η οποία περιλαμβάνει τη δημιουργία ωριαίων χρονοσειρών, τον έλεγχο και τη συμπλήρωση ελλειπόν τιμών, την ενσωμάτωση εσωτερικών και εξωτερικών περιβαλλοντικών μεταβλητών, καθώς και τη σύνδεση των χρονοσειρών με τα στατικά χαρακτηριστικά των κατοικιών. Τέλος, παρουσιάζεται το τελικό ενοποιημένο σύνολο δεδομένων, εμπλουτισμένο με ημερολογιακά χαρακτηριστικά, το οποίο αποτελεί τη βάση για τη διερευνητική ανάλυση και τη μοντελοποίηση στα επόμενα κεφάλαια.

3.1 Το σύνολο δεδομένων IDEAL

Το IDEAL household energy dataset είναι ένα πλούσιο σύνολο δεδομένων που συγκεντρώθηκε στο πλαίσιο του ερευνητικού έργου IDEAL (Improving Domestic Energy Advice through Listening) από 255 κατοικίες στην περιοχή του Εδιμβούργου και των γειτονικών περιοχών Lothians και south Fife στη Σκωτία (IDEAL Household Energy Project, 2021; Pullinger et al., 2021). Το σύνολο δεδομένων περιλαμβάνει υψηλής ανάλυσης χρονοσειρές κατανάλωσης ηλεκτρικής ενέργειας και φυσικού αερίου, περιβαλλοντικές μετρήσεις ανά δωμάτιο, μεταδεδομένα κατοικιών και συσκευών, καθώς και εκτεταμένα δημογραφικά και συμπεριφορικά δεδομένα από ερωτηματολόγια (Pullinger et al., 2021).

Η συλλογή δεδομένων διήρκεσε περίπου 23 μήνες, μέχρι τις 30 Ιουνίου 2018, με διάρκεια συμμετοχής ανά κατοικία από μερικές δεκάδες έως εκατοντάδες ημέρες (μέση διάρκεια περίπου 9–10 μήνες). Οι χρονοσειρές πλαισιώνονται από πλούσια μεταδεδομένα, αποθηκευμένα σε αρχεία CSV που περιέχουν βασικά χαρακτηριστικά κάθε κατοικίας (τύπος κτηρίου, έτος κατασκευής, σύστημα θέρμανσης κ.ά.), χαρακτηριστικά δωματίων (εμβαδόν, αριθμός παραθύρων, προσανατολισμός), πληροφορίες για συσκευές και δημογραφικά στοιχεία (αριθμός ενοίκων, ηλικιακή σύνθεση, κοινωνικοοικονομικά χαρακτηριστικά). Επιπλέον, συνοδευτικά αρχεία και τεκμηρίωση παρέχουν εξωτερικά καιρικά δεδομένα και δείκτες αστικότητας, γεγονός που καθιστά το IDEAL ιδιαίτερα κατάλληλο για μελέτες που συνδυάζουν κτηριακά, περιβαλλοντικά και συμπεριφορικά χαρακτηριστικά.

Στην παρούσα εργασία αξιοποιούνται κυρίως τα δεδομένα κύριας ηλεκτρικής κατανάλωσης (mains electricity), οι μετρήσεις θερμοκρασίας εσωτερικού χώρου και τα βασικά μεταδεδομένα κατοικίας, τα οποία μετατρέπονται σε ωριαίο σύνολο δεδομένων, κατάλληλο για εκπαίδευση μοντέλων πρόβλεψης επόμενης ημέρας.

3.2 Επιλογή και φιλτράρισμα κατοικιών

Παρότι το IDEAL περιλαμβάνει δεδομένα από 255 κατοικίες, δεν είναι όλες ισοδύναμες ως προς την ποιότητα και πληρότητα των χρονοσειρών. Η σχετική βιβλιογραφία υπογραμμίζει ότι ακόμη και μικρά ποσοστά σφαλμάτων σε αισθητήρες ή διακοπές επικοινωνίας μπορεί να επηρεάσουν ουσιαστικά την αξιοπιστία των μοντέλων που εκπαιδεύονται πάνω σε αυτά (Ahmad et al., 2018; Olu-Ajayi et al., 2023).

Για τον λόγο αυτό, πραγματοποιήθηκε αρχικός ποσοτικός έλεγχος ποιότητας δεδομένων, με στόχο τον εντοπισμό κατοικιών με ιδιαίτερα προβληματικές χρονοσειρές κατανάλωσης ή θερμοκρασίας. Ο έλεγχος επικεντρώθηκε στο ποσοστό ελλিপών τιμών ανά κατοικία και μεταβλητή, στην ύπαρξη εκτεταμένων χρονικών κενών και στην παρουσία ασυνεπειών στις χρονικές σφραγίδες, όπως μεγάλα χρονικά άλματα ή επικαλύψεις.

Αρχικά, αποκλείστηκε μία κατοικία που εμφάνιζε εκτεταμένα κενά και ασυνέχειες στις ωριαίες χρονοσειρές κατανάλωσης, οι οποίες δεν μπορούσαν να αποκατασταθούν αξιόπιστα μέσω συμπλήρωσης. Όλες οι υπόλοιπες κατοικίες διατήρησαν επαρκή ποιότητα δεδομένων και εντάχθηκαν στο αρχικό ενοποιημένο σύνολο δεδομένων.

Σε επόμενο στάδιο και ειδικότερα κατά την προετοιμασία των τελικών συνόλων εκπαίδευσης και ελέγχου για τα μοντέλα μηχανικής μάθησης, εντοπίστηκαν επιπλέον τέσσερις κατοικίες με μη ρεαλιστικά πρότυπα κατανάλωσης, όπως μηδενικό δηλωμένο αριθμό ενοίκων, ή εκτεταμένες περιόδους μηδενικής κατανάλωσης. Οι κατοικίες αυτές αποκλείστηκαν από τα τελικά πειράματα, ώστε να διασφαλιστεί η συνέπεια και η αξιοπιστία των χαρακτηριστικών εισόδου και των αποτελεσμάτων αξιολόγησης.

3.3 Ροή προεπεξεργασίας και ενοποίησης δεδομένων

Η παρούσα εργασία βασίζεται στα ωριαία δεδομένα του συνόλου IDEAL, όπως αυτά παρέχονται στα συμπληρωματικά αρχεία της πλατφόρμας Edinburgh DataShare. Στο πρώτο στάδιο υλοποιήθηκε μία ολοκληρωμένη ροή επεξεργασίας δεδομένων, με χρήση αυτόνομων σεναρίων σε γλώσσα Python, τα οποία είναι εκτελέσιμα τόσο μέσω περιβάλλοντος Jupyter Notebook όσο και από τη γραμμή εντολών. Η ροή αυτή μετασχηματίζει τα αρχικά αρχεία σε ένα ενιαίο, ενοποιημένο ωριαίο σύνολο δεδομένων ανά κατοικία και στη συνέχεια, σε ένα συγκεντρωτικό σύνολο δεδομένων (master dataset) που περιλαμβάνει το σύνολο των κατοικιών

Τα βασικά βήματα είναι τα ακόλουθα:

Δημιουργία ωριαίου αρχείου ανά κατοικία

Τα αρχικά συμπληρωματικά αρχεία φορτώνονται και για κάθε κατοικία, δημιουργείται ένα ενιαίο ωριαίο αρχείο που περιέχει την ωριαία κατανάλωση ηλεκτρικής ενέργειας (consumption_Wh) από τον φάκελο mains και τη μέση εσωτερική θερμοκρασία, υπολογισμένη από όλους τους αισθητήρες του σπιτιού. Η διαδικασία αυτή είναι συνεπής με τις πρακτικές που περιγράφονται στην τεκμηρίωση του IDEAL για παραγωγή ωριαίων χρονοσειρών (Pullinger et al., 2021).

Έλεγχος ελλিপών τιμών

Για κάθε ωριαίο αρχείο πραγματοποιείται ποσοτικός έλεγχος των ελλিপών τιμών (NaN) σε όλες τις βασικές μεταβλητές, παράγοντας συγκεντρωτικά reports ανά κατοικία. Ο έλεγχος αυτός επιτρέπει τον εντοπισμό σπιτιών με μεγάλα ποσοστά ελλείψεων και τον σχεδιασμό κατάλληλων στρατηγικών συμπλήρωσης.

Εποχική και χρονική συμπλήρωση ελλείψεων

Ανάλογα με το ποσοστό και τη δομή των κενών, εφαρμόζεται συνδυασμός τεχνικών. Για κατοικίες με αρκετά κενά, υιοθετείται αρχικά εποχική συμπλήρωση, χρησιμοποιώντας τιμές από την ίδια χρονική στιγμή σε γειτονικές ημέρες ή εβδομάδες, ώστε να διατηρηθεί η εποχικότητα. Για τα εναπομείναντα κενά εφαρμόζεται χρονική παρεμβολή στις χρονοσειρές (time interpolation). Παρόμοιοι συνδυασμοί εποχικής και χρονικής συμπλήρωσης χρησιμοποιούνται και σε προηγούμενες εργασίες πρόβλεψης φορτίου, όπου η σταθερότητα των μοντέλων επηρεάζεται σημαντικά από την ποιότητα της συμπλήρωσης (Hong & Fan, 2016; Ahmad et al., 2018).

Υπολογισμός εμβαδού και άλλων μεταδεδομένων

Τα μεταδεδομένα σε επίπεδο δωματίου (room-level metadata) του συνόλου IDEAL συγκεντρώνονται σε χαρακτηριστικά σε επίπεδο κατοικίας, όπως το συνολικό και το μέσο εμβαδόν, ο αριθμός δωματίων, καθώς και ο αριθμός εξωτερικών παραθύρων και εξωτερικών τοίχων. Για κάθε κατοικία, τα εμβαδά των επιμέρους δωματίων και χώρων αθροίζονται ώστε να προκύψει το συνολικό εμβαδόν, ενώ παράλληλα κανονικοποιούνται στη μονάδα μέτρησης τετραγωνικών μέτρων (m²) και αποθηκεύονται σε ένα ενιαίο αρχείο μεταδεδομένων. Η διαδικασία αυτή ακολουθεί τη λογική της συγκέντρωσης δεδομένων (aggregation) σε επίπεδο κατοικίας, η οποία χρησιμοποιείται ευρέως σε μελέτες ενεργειακής απόδοσης κτηρίων (Ahmad et al., 2018; Olu Ajayi et al., 2023).

Ενσωμάτωση εξωτερικής θερμοκρασίας

Για τις ανάγκες της παρούσας εργασίας, οι εξωτερικές θερμοκρασίες αντλήθηκαν από ανοιχτή υπηρεσία μετεωρολογικών δεδομένων (Open Meteo) για την περιοχή του Εδιμβούργου και αποθηκεύτηκαν σε ενιαίο ωριαίο αρχείο. Λαμβάνοντας υπόψη ότι η συντριπτική πλειοψηφία των κτηρίων του IDEAL ανήκει στην πόλη του Εδιμβούργου και ότι οι διαφορές στις τιμές της εξωτερικής θερμοκρασίας ανάμεσα στο Εδιμβούργο και στις υπόλοιπες πόλεις του δείγματος είναι πολύ μικρές, υιοθετείται η απλοποιητική παραδοχή ότι και αυτές οι κατοικίες έχουν ίδια εξωτερική θερμοκρασία με τα κτήρια της πόλης του Εδιμβούργου. Στη συνέχεια οι τιμές εξωτερικής θερμοκρασίας συγχωνεύονται με τα ωριαία αρχεία κάθε κατοικίας.

Δημιουργία κύριου συνόλου δεδομένων (master dataset)

Στο τελικό στάδιο, τα ωριαία αρχεία όλων των κατοικιών συνενώνονται σε ένα ενιαίο κύριο σύνολο δεδομένων, με προσθήκη αναγνωριστικού κατοικίας (home_id) και διατήρηση της χρονικής σειράς. Το κύριο σύνολο δεδομένων εμπλουτίζεται με τα μεταδεδομένα από τα αρχεία κατοικιών και παράγονται ημερολογιακά χαρακτηριστικά (ώρα, ημέρα εβδομάδας, δείκτης Σαββατοκύριακου, δείκτης αργιών Σκωτίας), ακολουθώντας πρακτικές που προτείνονται για βραχυπρόθεσμη πρόβλεψη ηλεκτρικού φορτίου (Hong & Fan, 2016; Ahmad et al., 2018).

3.4 Ημερολογιακά χαρακτηριστικά και τελικό σύνολο δεδομένων

Για να αποτυπωθούν τα σαφή ημερήσια και εβδομαδιαία μοτίβα της ηλεκτρικής κατανάλωσης, το ενοποιημένο κύριο σύνολο δεδομένων εμπλουτίζεται με ένα σύνολο ημερολογιακών χαρακτηριστικών. Συγκεκριμένα, για κάθε ωριαία εγγραφή υπολογίζονται:

- `hour`: ακέραιος 0–23, που δηλώνει την ώρα της ημέρας
- `day_of_week`: ακέραιος 0–6 (0 = Δευτέρα), που αποτυπώνει την εβδομαδιαία περιοδικότητα
- `is_weekend`: δυαδική μεταβλητή (0 = καθημερινή, 1 = Σαββατοκύριακο)
- `is_holiday`: δυαδικός δείκτης επίσημης αργίας στη Σκωτία

Η χρήση τέτοιων χαρακτηριστικών επιτρέπει στα μοντέλα να διαφοροποιούν τη συμπεριφορά τους ανάλογα με την ώρα και τον τύπο ημέρας, χωρίς να απαιτείται ξεχωριστή εκπαίδευση μοντέλων για κάθε υποομάδα δεδομένων (Hong & Fan, 2016).

Το τελικό κύριο σύνολο δεδομένων, `IDEAL_final_hourly_dataset`, περιλαμβάνει για κάθε κατοικία και ωριαία χρονική στιγμή την ωριαία κατανάλωση ηλεκτρικής ενέργειας (`consumption_Wh`), τη μέση εσωτερική θερμοκρασία, την εξωτερική θερμοκρασία, στατικά χαρακτηριστικά κατοικίας και νοικοκυριού, καθώς και ημερολογιακά χαρακτηριστικά, όπως `hour`, `day_of_week`, `is_weekend` και `is_holiday`.

Το κύριο σύνολο δεδομένων αποθηκεύεται ως `IDEAL_final_hourly_dataset` και αποτελεί τη βάση για τη διερευνητική ανάλυση (Κεφάλαιο 4) και την ανάπτυξη βασικών μοντέλων αναφοράς και μοντέλων μηχανικής μάθησης που παρουσιάζονται στα Κεφάλαια 5 και 6.

4. Διερευνητική Ανάλυση Δεδομένων και Χαρακτηριστικά

Το παρόν κεφάλαιο παρουσιάζει τη διερευνητική ανάλυση του τελικού ωριαίου συνόλου δεδομένων IDEAL, με στόχο την κατανόηση των βασικών μοτίβων κατανάλωσης και τη θεμελίωση των χαρακτηριστικών που χρησιμοποιούνται στη μοντελοποίηση. Αρχικά, περιγράφεται η προσέγγιση της ανάλυσης και η χρήση συγκεντρώσεων που λαμβάνουν υπόψη την ανομοιόμορφη διάρκεια συμμετοχής των κατοικιών στο σύνολο δεδομένων. Στη συνέχεια, εξετάζονται τα ημερήσια και ωριαία προφίλ κατανάλωσης, οι διαφοροποιήσεις μεταξύ καθημερινών και Σαββατοκύριακου, καθώς και η σχέση της ηλεκτρικής κατανάλωσης με την εξωτερική θερμοκρασία. Επιπλέον, αναλύονται οι διαφορές που προκύπτουν από στατικά χαρακτηριστικά των κατοικιών, όπως ο τύπος κατοικίας, ο βαθμός αστικότητας, ο αριθμός ενοίκων και το συνολικό εμβαδόν. Με βάση τα ευρήματα αυτά, ορίζονται οι βασικές κατηγορίες χαρακτηριστικών που θα αξιοποιηθούν στα μοντέλα πρόβλεψης, ενώ συμπληρωματικά εξετάζεται η σημαντικότητα των χαρακτηριστικών στα σενάρια χωρίς και με χρήση ιστορικών δεδομένων. Τέλος, παρουσιάζονται τα βασικά μοντέλα αναφοράς, τα οποία χρησιμοποιούνται ως αρχικό μέτρο σύγκρισης για την αξιολόγηση των μοντέλων μηχανικής μάθησης στα επόμενα κεφάλαια.

4.1 Προσέγγιση διερευνητικής ανάλυσης δεδομένων

Η διερευνητική ανάλυση δεδομένων (Exploratory Data Analysis, EDA) αποτελεί βασικό βήμα για την κατανόηση της δομής και της μεταβλητότητας του κύριου συνόλου δεδομένων IDEAL_final_hourly_dataset και για τη θεμελίωση των επιλογών χαρακτηριστικών που θα χρησιμοποιηθούν στα μοντέλα πρόβλεψης.

Στο πλαίσιο αυτό εφαρμόζονται δύο τρόποι ομαδοποίησης και υπολογισμού των μέσων τιμών. Ο πρώτος βασίζεται στο σύνολο των ωριαίων παρατηρήσεων όλων των κατοικιών (pooled aggregations), οι οποίες εξετάζονται ως ένα ενιαίο σύνολο δεδομένων. Ο δεύτερος βασίζεται στον υπολογισμό του μέσου προφίλ κάθε κατοικίας ξεχωριστά (home-weighted aggregations) και στη συνέχεια, στον υπολογισμό του μέσου όρου μεταξύ των κατοικιών.

Η δεύτερη προσέγγιση θεωρείται πιο ισορροπημένη, καθώς περιορίζει την υπεραντιπροσώπευση κατοικιών που διαθέτουν μεγαλύτερη διάρκεια μετρήσεων και συνεπώς περισσότερες ωριαίες εγγραφές. Με τον τρόπο αυτό, η ανάλυση δεν επηρεάζεται δυσανάλογα από λίγες κατοικίες με εκτενέστερες χρονοσειρές και αποτυπώνει πιο αντιπροσωπευτικά τη μέση συμπεριφορά ενός τυπικού νοικοκυριού.

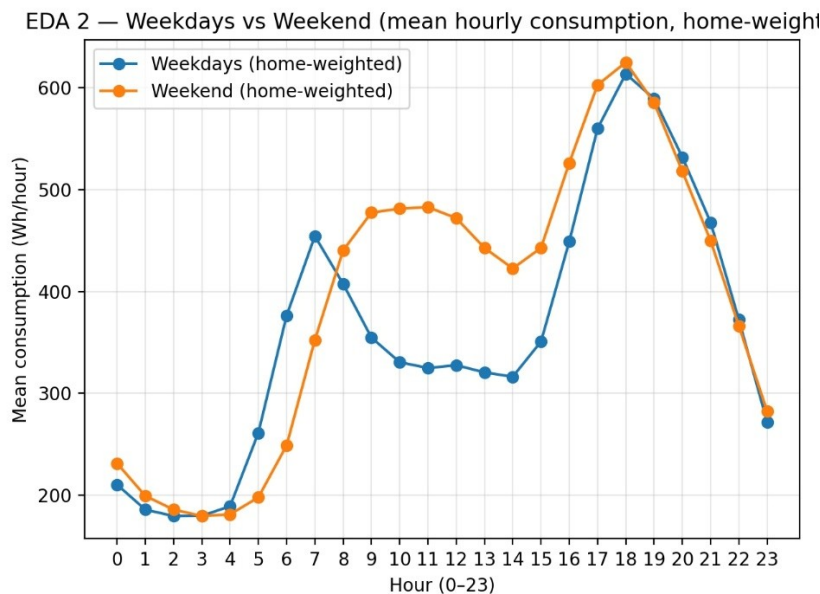
Βασικός στόχος της διερευνητικής ανάλυσης δεδομένων είναι η κατανόηση των ημερήσιων και εβδομαδιαίων μοτίβων κατανάλωσης, η διερεύνηση της επίδρασης των χαρακτηριστικών και η ανάδειξη διαφοροποιήσεων ανά τύπο κατοικίας και δημογραφικά χαρακτηριστικά, όπως προτείνεται και σε αντίστοιχες μελέτες ενεργειακής κατανάλωσης κτηρίων (Ahmad et al., 2018; Olu-Ajayi et al., 2023).

4.2 Ημερήσια και ωριαία προφίλ κατανάλωσης

Πρώτο βήμα της ανάλυσης είναι η διερεύνηση των ημερήσιων και ωριαίων προφίλ κατανάλωσης ηλεκτρικής ενέργειας. Υπολογίζεται το μέσο ωριαίο προφίλ για όλες τις

κατοικίες, τόσο σε επίπεδο όλης της εβδομάδας όσο και χωριστά για καθημερινές και Σαββατοκύριακα. Συγκεκριμένα, το μέσο ωριαίο προφίλ χαρακτηρίζεται από χαμηλή κατανάλωση τις μεσημεριανές ώρες, μια πρωινή αύξηση που κορυφώνεται γύρω στις 7–8 π.μ., πιο ήπια επίπεδα κατανάλωσης κατά τη διάρκεια της ημέρας και μία δεύτερη, εντονότερη αιχμή το βράδυ περίπου στις 18:00–20:00 που αντιστοιχεί σε ώρες επιστροφής στο σπίτι και χρήσης συσκευών (μαγείρεμα, ψυχαγωγία κ.λπ.), εύρημα που συμφωνεί με αντίστοιχες μελέτες οικιακής κατανάλωσης ενέργειας (Kong et al., 2019; Pullinger et al., 2021).

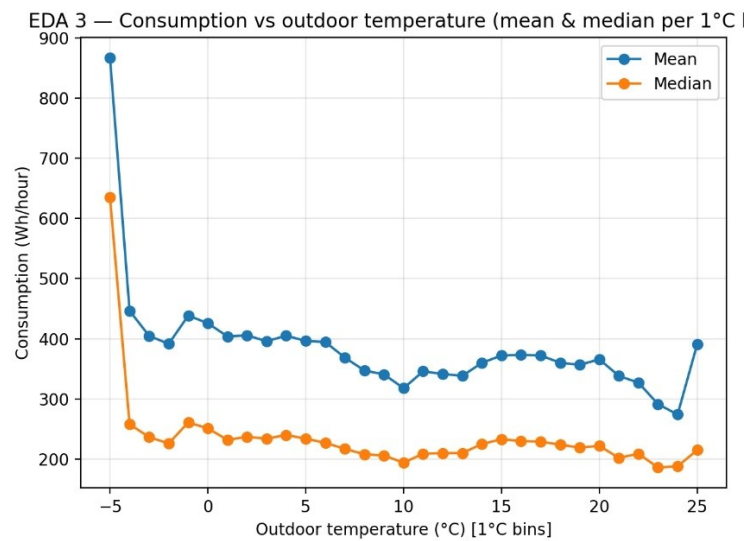
Η διάκριση μεταξύ καθημερινών και Σαββατοκύριακου αποκαλύπτει πρόσθετες διαφοροποιήσεις: στο Σχήμα 1 διακρίνεται ότι τα Σαββατοκύριακα οι αιχμές μετατοπίζονται σε ελαφρώς αργότερες ώρες και τα επίπεδα κατανάλωσης κατά τη διάρκεια της ημέρας είναι αυξημένα, αντανakλώντας την παρατεταμένη παρουσία των ενοίκων στο σπίτι. Τα ευρήματα αυτά επιβεβαιώνουν τη σημασία των χρονικών χαρακτηριστικών (hour, day_of_week, is_weekend) για την αποτύπωση της ημερήσιας και εβδομαδιαίας περιοδικότητας στα μοντέλα πρόβλεψης.



Σχήμα 1 Μέση ωριαία κατανάλωση (home-weighted) για καθημερινές και Σαββατοκύριακα.

4.3 Επίδραση εξωτερικής θερμοκρασίας

Η εξωτερική θερμοκρασία αποτελεί έναν από τους σημαντικότερους εξωγενείς παράγοντες που επηρεάζουν την ηλεκτρική κατανάλωση, ιδίως όταν μέρος του φορτίου συνδέεται με ηλεκτρική θέρμανση ή ψύξη (Ahmad et al., 2018; Li et al., 2025). Για να διερευνηθεί η σχέση κατανάλωσης–θερμοκρασίας, η εξωτερική θερμοκρασία ομαδοποιείται σε διαστήματα (bins) και για κάθε διάστημα υπολογίζεται η μέση ή διάμεση ωριαία κατανάλωση ανά κατοικία. Το αποτέλεσμα απεικονίζεται στο Σχήμα 2, όπου φαίνεται ότι σε χαμηλές θερμοκρασίες (κάτω από περίπου -5°C) η κατανάλωση αυξάνεται απότομα, ενώ σε ηπιότερες θερμοκρασίες παραμένει σχετικά επίπεδη.

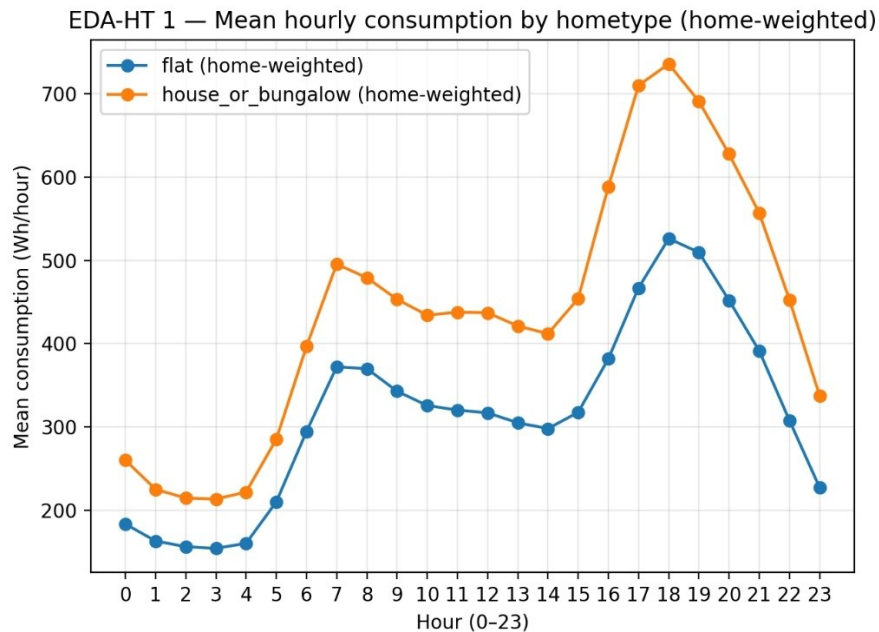


Σχήμα 2 Μέση και διάμεση ωριαία κατανάλωση σε συνάρτηση με την εξωτερική θερμοκρασία

Η διαφορά μεταξύ μέσου και διάμεσης τιμής υποδεικνύει την ύπαρξη κατοικιών με ιδιαίτερα υψηλή κατανάλωση στις πολύ χαμηλές θερμοκρασίες, πιθανώς λόγω ηλεκτρικής θέρμανσης ή κακής μόνωσης, που τραβούν προς τα πάνω τον μέσο όρο. Η μορφή της κατανάλωσης ως προς τη θερμοκρασία και ειδικά η έντονη αύξηση σε χαμηλές τιμές, συμφωνεί με ευρήματα της διεθνούς βιβλιογραφίας για την πρόβλεψη φορτίου σε κατοικίες και τη χρήση ηλεκτρικής ενέργειας στα νοικοκυριά, όπου συχνά παρατηρούνται σχέσεις τύπου U ή V ανάλογα με το κλίμα και το μείγμα των συστημάτων θέρμανσης και ψύξης (π.χ. Ahmad et al., 2018; Yao, 2021). Τα αποτελέσματα αυτά δικαιολογούν την ένταξη της εξωτερικής θερμοκρασίας ως κλιματικού χαρακτηριστικού των μοντέλων.

4.4 Διαφορές ανά τύπο κατοικίας (hometype)

Σημαντικές διαφοροποιήσεις προκύπτουν όταν οι κατοικίες ομαδοποιηθούν ως προς τον τύπο τους, π.χ. διαμερίσματα (flats) έναντι μονοκατοικιών (houses/bungalows). Για κάθε κατηγορία υπολογίζονται μέση ημερήσια και ωριαία κατανάλωση, καθώς και η σχέση κατανάλωσης–εξωτερικής θερμοκρασίας. Στο Σχήμα 3 παρουσιάζεται το μέσο ωριαίο προφίλ κατανάλωσης ανά τύπο κατοικίας, σταθμισμένο ανά κατοικία:



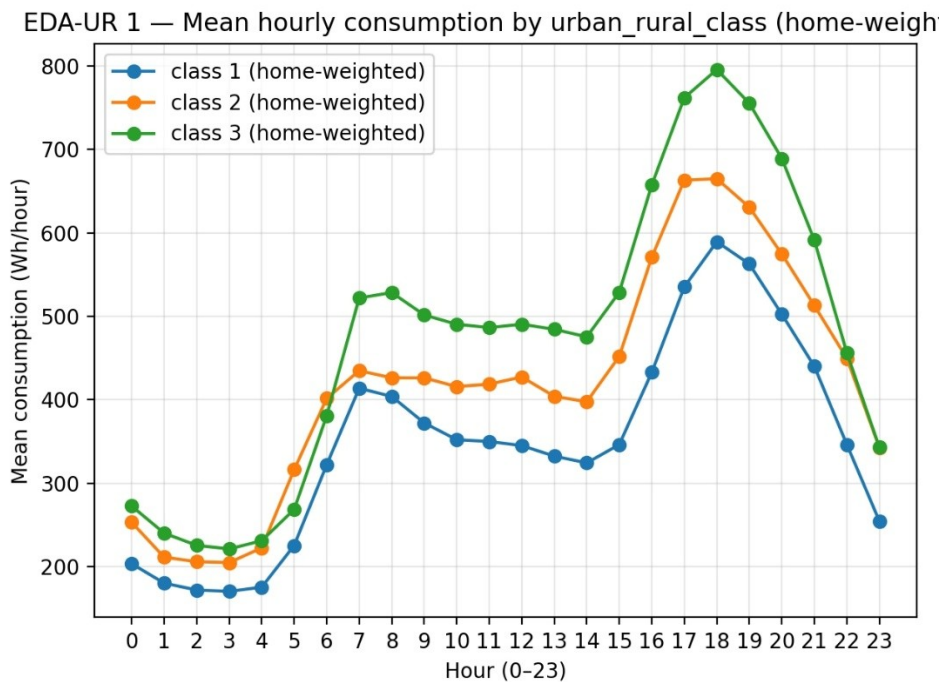
Σχήμα 3 Μέση ωριαία κατανάλωση (home-weighted) ανά τύπο κατοικίας

Τα διαμερίσματα τείνουν, κατά μέσο όρο, να παρουσιάζουν χαμηλότερη συνολική κατανάλωση και μικρότερη ευαισθησία σε χαμηλές θερμοκρασίες, πιθανώς λόγω μικρότερου εμβαδού και καλύτερης θερμικής προστασίας από γειτονικά διαμερίσματα. Αντίθετα, η κατηγορία μονοκατοικιών εμφανίζει υψηλότερη μέση κατανάλωση και πιο έντονη αύξηση κατανάλωσης σε χαμηλές θερμοκρασίες. Η διαφορά είναι ιδιαίτερα εμφανής στις ώρες 17:00–21:00, όπου οι μονοκατοικίες παρουσιάζουν σημαντικά υψηλότερες τιμές από τα διαμερίσματα, κάτι που μπορεί να σχετίζεται με μεγαλύτερα εμβαδά, περισσότερους χώρους θέρμανσης και διαφορετικές συνήθειες χρήσης συσκευών, κάτι που είναι συνεπές με τα ευρήματα μελετών που συσχετίζουν τον τύπο κατοικίας με ενεργειακή ζήτηση (Olu-Ajayi et al., 2023).

Η ανάλυση αυτή υπογραμμίζει τη σημασία της συμπερίληψης του τύπου κατοικίας (hometype) ως στατικού χαρακτηριστικού στα μοντέλα πρόβλεψης, ώστε να ληφθούν υπόψη οι δομικές και λειτουργικές διαφορές μεταξύ κατηγοριών κτηρίων.

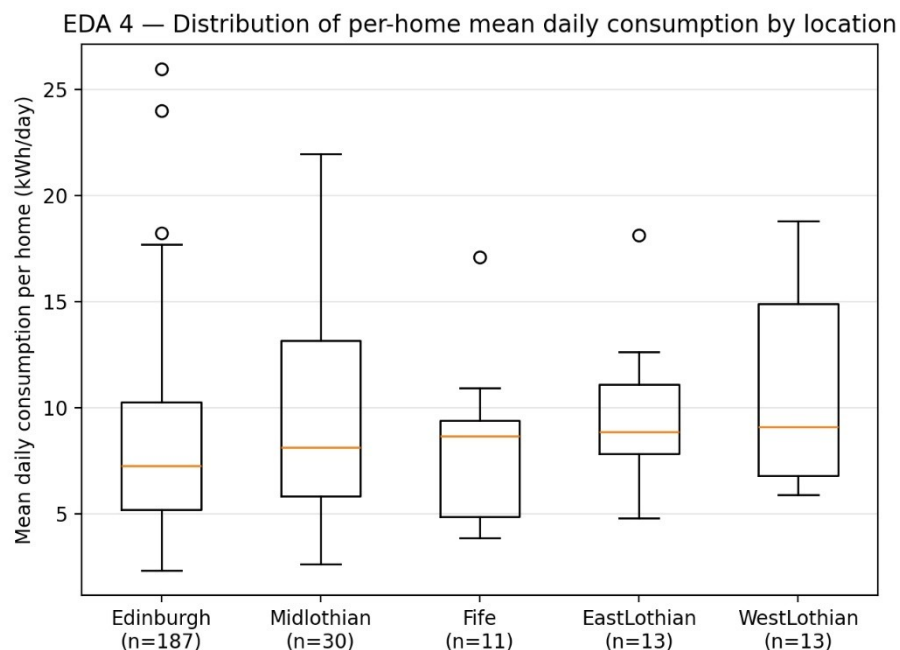
4.5 Διαφορές ανά κατηγορίες αστικότητας, αριθμού ενοίκων και εμβαδού κατοικίας

Πέρα από τον τύπο κατοικίας, σημαντικό ρόλο στη διαμόρφωση του επιπέδου και του προφίλ κατανάλωσης διαδραματίζουν τα χωρικά και δημογραφικά χαρακτηριστικά. Η μεταβλητή `urban_rural_class` χρησιμοποιείται για τη διάκριση των κατοικιών ως προς τον βαθμό αστικότητας. Οι χαμηλότερες τιμές της μεταβλητής αντιστοιχούν σε περισσότερο αστικές περιοχές, ενώ οι υψηλότερες τιμές αντιστοιχούν σε πιο αγροτικές, απομακρυσμένες ή λιγότερο πυκνοκατοικημένες περιοχές. Η ανάλυση δείχνει συστηματικές διαφορές στη μέση ωριαία κατανάλωση μεταξύ των κατηγοριών αυτών, όπως φαίνεται στο Σχήμα 4.



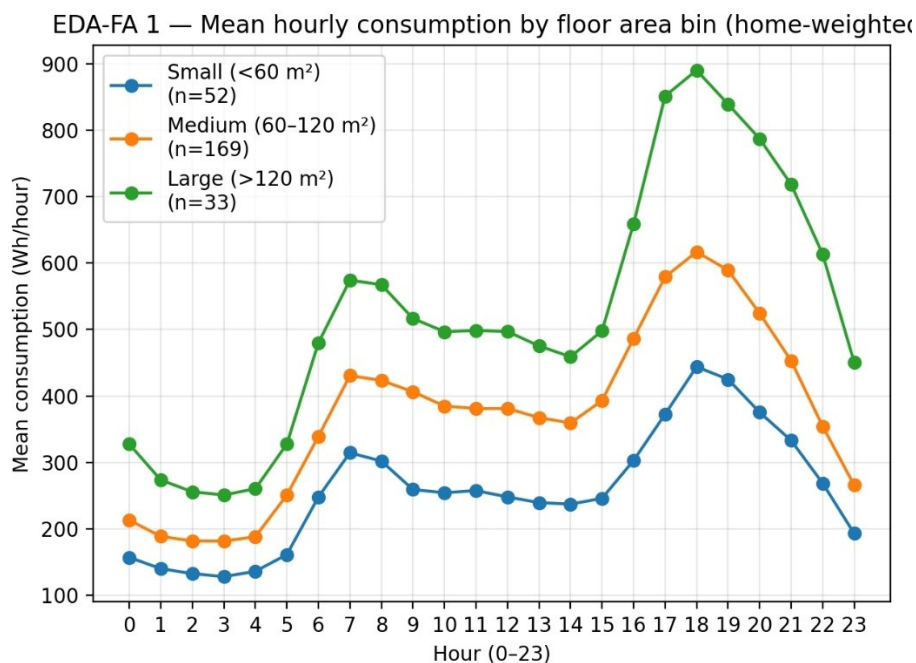
Σχήμα 4 Μέση ωριαία κατανάλωση (home-weighted) ανά κατηγορία urban_rural_class

Οι πιο αγροτικές κατηγορίες παρουσιάζουν υψηλότερο φορτίο σε όλη τη διάρκεια της ημέρας και ιδιαίτερα έντονες απογευματινές και βραδινές αιχμές, κάτι που μπορεί να αποδοθεί σε μεγαλύτερα κτήρια, διαφορετικά συστήματα θέρμανσης και ενδεχομένως διαφορετικό τρόπο ζωής. Η κατανομή της μέσης ημερήσιας κατανάλωσης ανά περιοχή, όπως παρουσιάζεται στο Σχήμα 5, επιβεβαιώνει την ύπαρξη χωρικής ετερογένειας, με ορισμένες περιοχές να εμφανίζουν υψηλότερη διάμεση κατανάλωση και μεγαλύτερη διασπορά.



Σχήμα 5 Κατανομή μέσης ημερήσιας κατανάλωσης ανά κατοικία και περιοχή

Όσον αφορά το συνολικό εμβαδόν της κατοικίας, το Σχήμα 6 παρουσιάζει το μέσο ωριαίο προφίλ κατανάλωσης για τρεις κατηγορίες συνολικού εμβαδού δηλαδή μικρές, μεσαίες και μεγάλες κατοικίες, σε σταθμισμένη βάση ως προς τις κατοικίες.



Σχήμα 6 Μέση ωριαία κατανάλωση (home-weighted) ανά κατηγορία εμβαδού κατοικίας

Παρατηρείται ότι η αύξηση του εμβαδού συνοδεύεται από σταθερή αύξηση της κατανάλωσης σε όλες τις ώρες της ημέρας, με τα μεγαλύτερα σπίτια (>120 m²) να παρουσιάζουν ιδιαίτερα τονισμένες βραδινές αιχμές. Αντίστοιχα, η κατανομή της μέσης ημερήσιας κατανάλωσης ανά κατηγορία εμβαδού, δείχνει υψηλότερες διάμεσες τιμές και μεγαλύτερη διασπορά για τα μεγαλύτερα σπίτια, υποδηλώνοντας ότι το εμβαδόν είναι ισχυρός προσδιοριστικός παράγοντας της ενεργειακής ζήτησης.

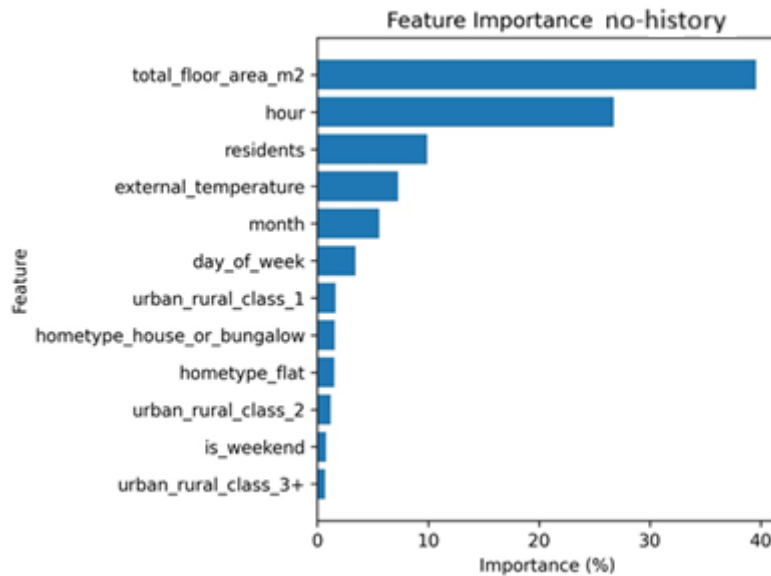
Τέλος, όπως είναι λογικό ο αριθμός των ενοίκων συνδέεται θετικά με το επίπεδο κατανάλωσης, όπως αναμένεται για μεγαλύτερα νοικοκυριά. Τα σπίτια με περισσότερους κατοίκους εμφανίζουν αυξημένη μέση κατανάλωση και έντονες απογευματινές/βραδινές αιχμές, γεγονός που αντανάκλα την πιο συχνή και ταυτόχρονη χρήση οικιακών συσκευών. Συνολικά, η ανάλυση των στατικών χαρακτηριστικών αστικότητας (urban_rural_class), αριθμού ενοίκων (residents) και συνολικού εμβαδού (total_floor_area), επιβεβαιώνει ότι φέρουν σημαντική πληροφορία για το επίπεδο και τη μεταβλητότητα της κατανάλωσης, σε συμφωνία με τη σχετική βιβλιογραφία για ενεργειακή κατανάλωση κτηρίων (Ahmad et al., 2018; Olu-Ajayi et al., 2023). Η συμπερίληψή τους ως εισόδων στα μοντέλα πρόβλεψης κρίνεται επομένως απαραίτητη, καθώς συμβάλλουν στη διακριτοποίηση διαφορετικών τύπων νοικοκυριών και στη βελτίωση της ακρίβειας των προβλέψεων.

4.6 Επιλογή και τελικός ορισμός χαρακτηριστικών

Συνδυάζοντας τα ευρήματα της διερευνητικής ανάλυσης με τη βιβλιογραφία για πρόβλεψη φορτίου κτηρίων (Ahmad et al., 2018; Hong & Fan, 2016; Olu-Ajayi et al., 2023), η παρούσα εργασία ορίζει τέσσερις βασικές κατηγορίες χαρακτηριστικών.

- Χρονικά χαρακτηριστικά, τα οποία περιλαμβάνουν τις μεταβλητές `hour`, `day_of_week`, `is_weekend`, `is_holiday` και εντάσσοντας αργότερα `month`, με στόχο την αποτύπωση ημερήσιας, εβδομαδιαίας και ετήσιας περιοδικότητας.
- Κλιματικά χαρακτηριστικά, τα οποία περιλαμβάνουν την εξωτερική θερμοκρασία και τη μέση εσωτερική θερμοκρασία, οι οποίες συνδέονται άμεσα με τις ανάγκες θέρμανσης/ψύξης και βελτιώνουν την ακρίβεια πρόβλεψης.
- Στατικά χαρακτηριστικά, τα οποία περιλαμβάνουν τις μεταβλητές `hometype`, `total_floor_area_m2`, `residents`, `urban_rural_class`, που αποτυπώνουν σταθερές ιδιότητες του νοικοκυριού και του κτηρίου.
- Ιστορικά χαρακτηριστικά, τα οποία περιλαμβάνουν υστερήσεις της κατανάλωσης, όπως `lag_1h` και `lag_24h`, καθώς και κυλιόμενους μέσους όρους, όπως `roll_mean_24h` και `roll_mean_168h`, τα οποία χρησιμοποιούνται στα σενάρια με χρήση ιστορικών δεδομένων, για να ενσωματώσουν την τοπική χρονική δυναμική του φορτίου.

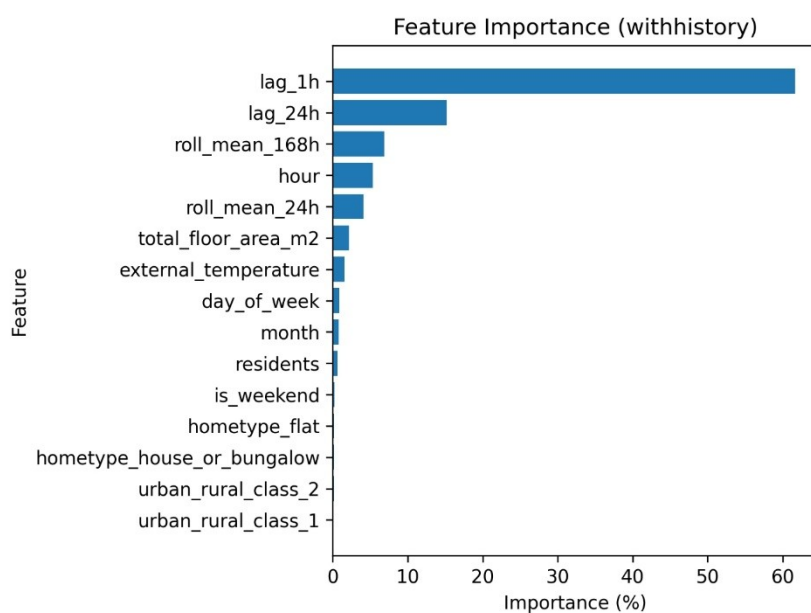
Συμπληρωματικά προς τη διερευνητική ανάλυση, πραγματοποιήθηκε προκαταρκτική ανάλυση σημαντικότητας χαρακτηριστικών, με χρήση των δένδροειδών μοντέλων Random Forest, XGBoost και LightGBM, τα οποία αξιοποιούνται στη συνέχεια της εργασίας. Η ανάλυση αυτή εφαρμόστηκε αρχικά στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, ώστε να εξεταστεί ποια χρονικά, κλιματικά και στατικά χαρακτηριστικά συμβάλλουν περισσότερο στην πρόβλεψη όταν δεν υπάρχει διαθέσιμο πρόσφατο ιστορικό φορτίου. Στο Σχήμα 7 παρουσιάζονται τα αποτελέσματα της ανάλυσης σημαντικότητας χαρακτηριστικών για το μοντέλο Random Forest στο σενάριο χωρίς χρήση ιστορικών δεδομένων.



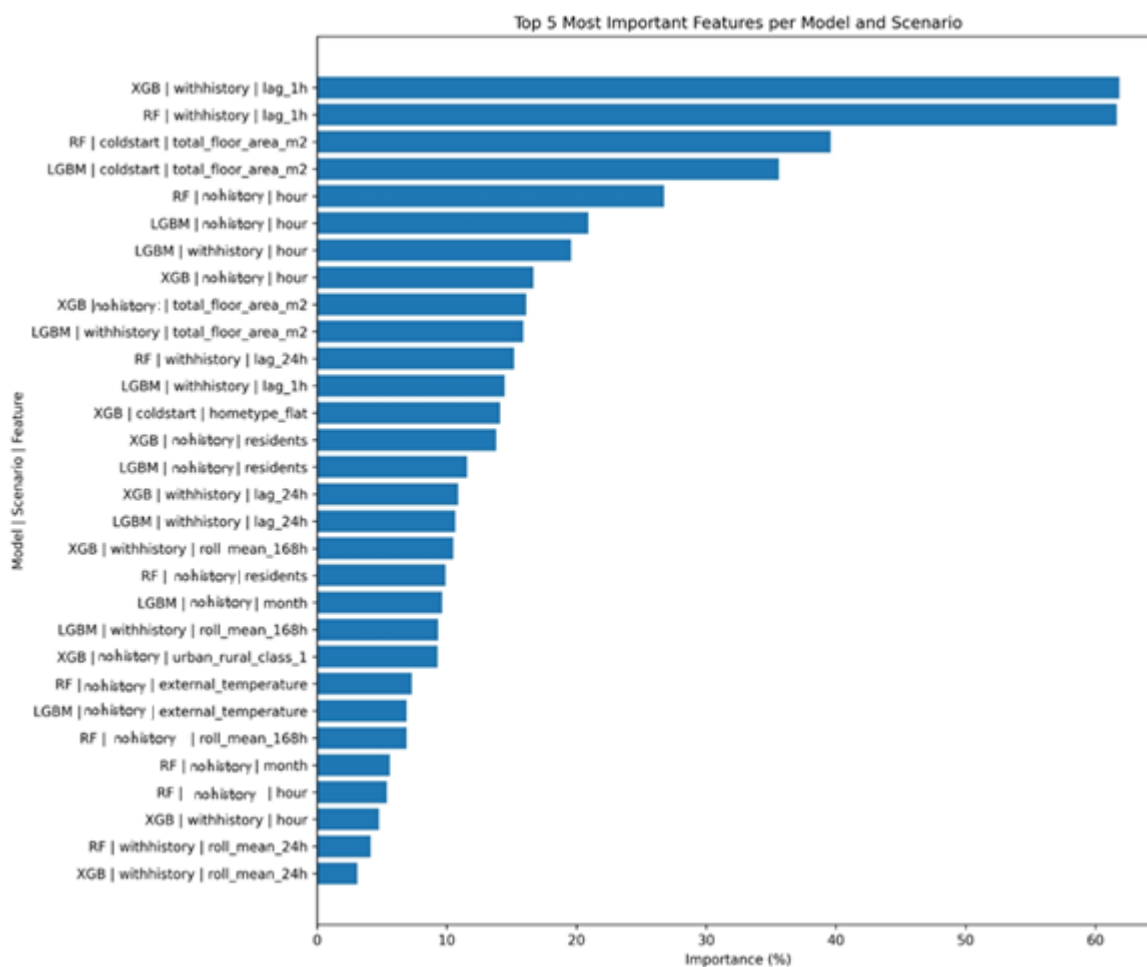
Σχήμα 7 Σημαντικότητα χαρακτηριστικών για το μοντέλο Random Forest στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Για το μοντέλο αυτό, το σημαντικότερο χαρακτηριστικό είναι η συνολική επιφάνεια της κατοικίας (*total_floor_area_m2*), η οποία συγκεντρώνει περίπου 40% της συνολικής σημαντικότητας, ακολουθούμενη από την ώρα της ημέρας (*hour*) με περίπου 27%. Στη συνέχεια έρχονται ο αριθμός ενοίκων (*residents*), η εξωτερική θερμοκρασία (*external_temperature*) και οι μεταβλητές *month* και *day_of_week*, με μικρότερες αλλά μη αμελητέες συνεισφορές. Αντίστοιχη εικόνα παρατηρείται και στα μοντέλα XGBoost και LightGBM. Στα σενάρια χωρίς χρήση ιστορικών, η σημαντικότητα κατανέμεται κυρίως μεταξύ δομικών χαρακτηριστικών (*total_floor_area_m2*, *hometype*, *residents*) και χρονικών μεταβλητών (*hour*, *month*, *day_of_week*), με τους περιβαλλοντικούς παράγοντες να παίζουν δευτερεύον αλλά σταθερά θετικό ρόλο. Τα ευρήματα αυτά υποδηλώνουν ότι, χωρίς ιστορική πληροφορία κατανάλωσης, η πρόβλεψη βασίζεται κυρίως σε στατικά και ημερολογιακά χαρακτηριστικά, κάτι που είναι ιδιαίτερα σημαντικό για σενάρια όπου δεν υπάρχουν διαθέσιμα δεδομένα από έξυπνο μετρητή.

Για τα σενάρια με χρήση ιστορικών δεδομένων, η ίδια διαδικασία εφαρμόστηκε σε σύνολο δεδομένων που περιλαμβάνει ρητά ιστορικά χαρακτηριστικά κατανάλωσης (*lag_1h*, *lag_24h*, *roll_mean_24h*, *roll_mean_168h*). Στο Σχήμα 8 παρουσιάζονται τα αποτελέσματα της ανάλυσης σημαντικότητας χαρακτηριστικών για το μοντέλο Random Forest with-history, ενώ στο συνοπτικό Σχήμα 9 απεικονίζονται τα πέντε σημαντικότερα χαρακτηριστικά ανά μοντέλο και σενάριο.



Σχήμα 8 Σημαντικότητα χαρακτηριστικών για το μοντέλο Random Forest στο σενάριο με χρήση ιστορικών δεδομένων



Σχήμα 9 Τα 5 σημαντικότερα χαρακτηριστικά ανά μοντέλο (RF, XGB, LGBM) και σενάριο

Στο σενάριο με χρήση ιστορικών δεδομένων για το μοντέλο Random Forest, η μεταβλητή lag_1h κυριαρχεί σαφώς, συγκεντρώνοντας περίπου 62% της συνολικής σημαντικότητας, με τη lag_24h να ακολουθεί με περίπου 15% και τον κυλιόμενο μέσο 168 ωρών (roll_mean_168h) με περίπου 7%. Η ώρα της ημέρας (hour) και ο κυλιόμενος μέσος 24 ωρών (roll_mean_24h) διατηρούν μικρότερη αλλά αξιοσημείωτη συνεισφορά, ενώ τα στατικά χαρακτηριστικά (total_floor_area_m2, residents) και η εξωτερική θερμοκρασία υποχωρούν σε σαφώς δεύτερο ρόλο. Παρόμοιο μοτίβο εμφανίζει και το XGBoost, όπου το lag_1h παραμένει το σημαντικότερο χαρακτηριστικό με ποσοστό άνω του 60%, ενώ ακολουθούν τα lag_24h και roll_mean_168h. Το LightGBM εμφανίζει πιο ισορροπημένη κατανομή, όμως και πάλι οι όροι ιστορικής κατανάλωσης (lag_1h, lag_24h, roll_mean_168h, roll_mean_24h) καταλαμβάνουν μεγάλο μέρος της συνολικής σημαντικότητας σε σχέση με τα υπόλοιπα χαρακτηριστικά.

Η συνέπεια αυτών των αποτελεσμάτων, παρά τις διαφορές μεταξύ των αλγορίθμων, υποδεικνύει ότι η ωριαία κατανάλωση της προηγούμενης ώρας (lag-1), λειτουργεί ως κυρίαρχος αυτοπαλινδρομικός όρος, αποτυπώνοντας τη βραχυχρόνια χρονική εξάρτηση του φορτίου. Οι όροι lag_24h και οι κυλιόμενοι μέσοι 24 και 168 ωρών συλλαμβάνουν συμπληρωματικά ημερήσια και εβδομαδιαία μοτίβα κατανάλωσης, με αποτέλεσμα τα υπόλοιπα χρονικά, κλιματικά και κτηριακά χαρακτηριστικά να αποκτούν δευτερεύουσα σημασία στα μοντέλα με χρήση ιστορικών δεδομένων. Τα ευρήματα επιβεβαιώνουν ότι η βραχυπρόθεσμη πρόβλεψη ηλεκτρικής κατανάλωσης σε ωριαία βάση αποτελεί κατεξοχήν πρόβλημα χρονικών σειρών, όπου η πρόσφατη ιστορική συμπεριφορά του συστήματος έχει καθοριστικό ρόλο και δικαιολογεί την ανάπτυξη ξεχωριστών μοντέλων με χρήση ιστορικών δεδομένων όταν υπάρχει διαθέσιμο ωριαίο ιστορικό π.χ. από έξυπνο μετρητή.

4.7 Βασικά μοντέλα αναφοράς

Πριν από τη χρήση σύνθετων μοντέλων μηχανικής μάθησης, είναι χρήσιμο να οριστούν απλά στατιστικά μοντέλα, τα οποία λειτουργούν ως σημείο αναφοράς για την αξιολόγηση της προστιθέμενης αξίας πιο πολύπλοκων προσεγγίσεων (Olu-Ajayi et al., 2023). Στη βιβλιογραφία για πρόβλεψη φορτίου και κατανάλωσης ενέργειας σε κτήρια χρησιμοποιούνται συχνά μοντέλα εμμονής και μοντέλα μέσων τιμών ως στοιχειώδεις γραμμές βάσης απόδοσης (Al-Shargabi et al., 2022; Olu-Ajayi et al., 2023).

Στην παρούσα εργασία υλοποιούνται τέσσερα βασικά μοντέλα αναφοράς σε ωριαία κλίμακα:

Το πρώτο μοντέλο αναφοράς (Previous day – same hour), προβλέπει την κατανάλωση της ώρας t χρησιμοποιώντας την κατανάλωση της ίδιας ώρας της προηγούμενης ημέρας.

Το δεύτερο μοντέλο αναφοράς (Global mean by hour), χρησιμοποιεί ως πρόβλεψη τη μέση κατανάλωση της αντίστοιχης ώρας της ημέρας, όπως αυτή υπολογίζεται συνολικά από όλες τις κατοικίες και όλες τις διαθέσιμες ημέρες.

Το τρίτο μοντέλο αναφοράς (Mean by hour with weekday/weekend distinction), αποτελεί επέκταση του προηγούμενου μοντέλου, καθώς υπολογίζει ξεχωριστούς μέσους όρους για εργάσιμες ημέρες και Σαββατοκύριακα, αξιοποιώντας το χαρακτηριστικό is_weekend.

Το τέταρτο μοντέλο αναφοράς (Baseline 04 Per-home mean by hour), υπολογίζει ξεχωριστά για κάθε κατοικία τη μέση κατανάλωση κάθε ώρας της ημέρας. Με τον τρόπο αυτό αποτυπώνεται το χαρακτηριστικό ημερήσιο προφίλ κάθε νοικοκυριού και εισάγεται ένας απλός βαθμός εξατομίκευσης στη γραμμή βάσης.

Απλά γραμμικά ή δένδροειδή μοντέλα μπορούν επίσης να χρησιμοποιηθούν ως γραμμές βάσης σε μελέτες φόρτισης κτηρίων, ιδιαίτερα όταν αξιοποιούν περιορισμένα, εύκολα διαθέσιμα χαρακτηριστικά (Olu-Ajayi et al., 2023). Ωστόσο, στη συγκεκριμένη ανάλυση ο ρόλος της κατώτερης γραμμής απόδοσης καλύπτεται από τα τέσσερα ανωτέρω στατιστικά μοντέλα αναφοράς, τα οποία ευθυγραμμίζονται με την πρακτική που περιγράφεται σε πρόσφατες ανασκοπήσεις για δεδομενοκεντρική πρόβλεψη κατανάλωσης ενέργειας σε κτήρια (Al-Shargabi, Almhafdy, Ibrahim, Alghieth, & Chiclana, 2022; Olu-Ajayi et al., 2023).

Ο ρόλος των στατιστικών αυτών μοντέλων είναι να παρέχουν ένα σαφές σημείο αναφοράς. Ένα μοντέλο μηχανικής μάθησης που δεν μπορεί να υπερκεράσει συστηματικά τα βασικά μοντέλα αναφοράς σε μετρικές όπως MAE και RMSE δύσκολα δικαιολογεί την πρόσθετη πολυπλοκότητα και υπολογιστική επιβάρυνση. Στο πλαίσιο της παρούσας εργασίας, ιδιαίτερη έμφαση δίνεται στο τέταρτο μοντέλο αναφοράς, καθώς ενσωματώνει εξατομίκευση ανά κατοικία και αποτελεί το ισχυρότερο από τα στατιστικά μοντέλα αναφοράς. Τα μοντέλα μηχανικής μάθησης που παρουσιάζονται στα επόμενα κεφάλαια θεωρείται ότι προσφέρουν ουσιαστική πρόοδο μόνο εφόσον επιτυγχάνουν καλύτερη απόδοση από αυτή τη γραμμή βάσης.

5. Μεθοδολογία Μοντελοποίησης

Σε αυτό το κεφάλαιο παρουσιάζεται η μεθοδολογία μοντελοποίησης που ακολουθείται για την πρόβλεψη της ωριαίας ηλεκτρικής κατανάλωσης στο σύνολο δεδομένων IDEAL. Αρχικά, ορίζεται το προβλεπτικό πρόβλημα και ο χρονικός ορίζοντας της πρόβλεψης, με έμφαση στην παραγωγή προβλέψεων επόμενης ημέρας σε επίπεδο κατοικίας. Στη συνέχεια, περιγράφεται ο τρόπος με τον οποίο τα χαρακτηριστικά των κατοικιών αξιοποιούνται για την έμμεση κατηγοριοποίησή τους και τη διαφοροποίηση της προβλεπτικής συμπεριφοράς των μοντέλων. Ακολουθεί η παρουσίαση των δύο βασικών σεναρίων πρόβλεψης, χωρίς χρήση ιστορικών δεδομένων και με ενσωμάτωση ιστορικών χαρακτηριστικών, καθώς και η διαδικασία χρονικού διαχωρισμού των δεδομένων σε σύνολο εκπαίδευσης και ελέγχου. Επιπλέον, αναλύονται τα βασικά βήματα προεπεξεργασίας, οι μετασχηματισμοί της μεταβλητής στόχου και η διαχείριση αριθμητικών και κατηγορικών χαρακτηριστικών. Τέλος, τεκμηριώνεται η επιλογή των αλγορίθμων Random Forest, XGBoost και LightGBM και παρουσιάζονται οι βασικές ρυθμίσεις των μοντέλων που χρησιμοποιούνται στα πειράματα.

5.1 Ορισμός προβλήματος και ορίζοντας πρόβλεψης

Η παρούσα εργασία εξετάζει την πρόβλεψη της ωριαίας κατανάλωσης ηλεκτρικής ενέργειας σε επίπεδο κατοικίας, με ορίζοντα πρόβλεψης μία ημέρα μπροστά. Ειδικότερα, για κάθε κατοικία και για κάθε ώρα της επόμενης ημέρας εκτιμάται η αντίστοιχη τιμή κατανάλωσης, με βάση όλες τις διαθέσιμες πληροφορίες έως και την τελευταία γνωστή χρονική στιγμή. Κατά αυτόν τον τρόπο, η μονάδα πρόβλεψης ορίζεται από τον συνδυασμό μιας συγκεκριμένης κατοικίας και μιας συγκεκριμένης χρονικής στιγμής, επιτρέποντας την αποτίμηση εξατομικευμένων ημερήσιων προφίλ φορτίου για μεγάλο πλήθος νοικοκυριών. Η επιλογή ορίζοντα επόμενης ημέρας είναι ευθυγραμμισμένη με πλήθος εφαρμογών στη διαχείριση ηλεκτρικής ενέργειας, όπως ο προγραμματισμός παραγωγής και αγοράς, η βελτιστοποίηση λειτουργίας μικροδικτύων και η υλοποίηση προγραμμάτων διαχείρισης ζήτησης (demand response) σε οικιακό επίπεδο. Σε τέτοια σενάρια, η διαθεσιμότητα αξιόπιστων ημερήσιων προβλέψεων επιτρέπει τον καλύτερο προγραμματισμό ελέικτων φορτίων και αποθηκευτικών μέσων, καθώς και τον έγκαιρο εντοπισμό περιόδων αιχμής κατανάλωσης.

5.2 Κατηγοριοποίηση κατοικιών μέσω χαρακτηριστικών

Για τις ανάγκες της παρούσας εργασίας δεν εφαρμόζεται ξεχωριστό στάδιο ομαδοποίησης των κατοικιών. Αντίθετα, οι διαφορές μεταξύ των κατοικιών ενσωματώνονται άμεσα μέσω των χαρακτηριστικών εισόδου που χρησιμοποιούνται κατά την εκπαίδευση των μοντέλων. Με τον τρόπο αυτό, η περιγραφή κάθε κατοικίας βασίζεται σε πραγματικά διαθέσιμα μεταδεδομένα και είναι πλήρως συμβατή με το προβλεπτικό σενάριο χρήσης του συστήματος.

Κάθε κατοικία περιγράφεται από ένα σύνολο δομικών, κοινωνικών και χωρικών χαρακτηριστικών, όπως ο τύπος κατοικίας (flat ή house_or_bungalow), η κατηγορία

αστικότητας (urban_rural_class), το συνολικό εμβαδόν (total_floor_area_m2) και ο αριθμός ενοίκων (residents). Τα χαρακτηριστικά αυτά συνδυάζονται με χρονικές μεταβλητές, όπως η ώρα, η ημέρα της εβδομάδας, οι ενδείξεις Σαββατοκύριακου και αργιών, καθώς και με περιβαλλοντικά και ιστορικά χαρακτηριστικά όπου αυτά είναι διαθέσιμα.

Η προσέγγιση αυτή μπορεί να χαρακτηριστεί ως έμμεση κατηγοριοποίηση μέσω χαρακτηριστικών εισόδου. Δεν αποδίδονται προκαθορισμένες ετικέτες κατηγορίας στις κατοικίες, αλλά τα μοντέλα μαθαίνουν αυτόματα διαφορετικούς κανόνες πρόβλεψης για διαφορετικούς συνδυασμούς χαρακτηριστικών. Για παράδειγμα, μία μικρή αστική κατοικία με έναν ένοικο αναμένεται να παρουσιάζει διαφορετικό πρότυπο κατανάλωσης από μία μεγαλύτερη μονοκατοικία με περισσότερους ενοίκους. Οι διαφορές αυτές αποτυπώνονται μέσω των χαρακτηριστικών εισόδου και αξιοποιούνται κατά τη διαδικασία εκπαίδευσης.

Η συγκεκριμένη προσέγγιση είναι ιδιαίτερα κατάλληλη για συνδυαστικά δένδροειδή μοντέλα, όπως τα Random Forest, XGBoost και LightGBM, τα οποία μπορούν να αναγνωρίζουν μη γραμμικές αλληλεπιδράσεις μεταξύ των χαρακτηριστικών και να εφαρμόζουν διαφορετικούς κανόνες πρόβλεψης σε διαφορετικούς τύπους κατοικιών χωρίς να απαιτείται προηγούμενο στάδιο ομαδοποίησης (Al Shargabi et al., 2022; Olu Ajayi et al., 2023).

Επιπλέον, η προσέγγιση αυτή συνδέεται άμεσα με τον σχεδιασμό του συστήματος πρόβλεψης που αναπτύσσεται στην παρούσα εργασία. Ο χρήστης δεν χρειάζεται να επιλέξει κάποια προκαθορισμένη κατηγορία κατοικίας, αλλά αρκεί να παρέχει βασικά στοιχεία, όπως το εμβαδόν, τον τύπο κατοικίας, τον αριθμό ενοίκων και την κατηγορία αστικότητας. Τα στοιχεία αυτά μετατρέπονται σε χαρακτηριστικά εισόδου και αξιοποιούνται από τα μοντέλα για την παραγωγή εξατομικευμένων προβλέψεων κατανάλωσης.

5.3 Σενάρια πρόβλεψης

Για να καλυφθούν διαφορετικά επίπεδα διαθεσιμότητας δεδομένων, η παρούσα εργασία ορίζει δύο διακριτά σενάρια πρόβλεψης: το σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης (no-history) και το σενάριο με χρήση ιστορικών δεδομένων (with-history).

Στο σενάριο χωρίς ιστορικά δεδομένα θεωρείται ότι για μια κατοικία δεν είναι διαθέσιμες προηγούμενες μετρήσεις κατανάλωσης, αλλά υπάρχουν χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά, όπως αυτά που παρουσιάστηκαν στο Κεφάλαιο 4. Τα μοντέλα αξιοποιούν μεταβλητές όπως η ώρα της ημέρας, η ημέρα της εβδομάδας, η ένδειξη Σαββατοκύριακου, ο μήνας, η εξωτερική θερμοκρασία, το συνολικό εμβαδόν της κατοικίας και ο αριθμός ενοίκων, καθώς και κατάλληλα κωδικοποιημένες κατηγορικές μεταβλητές, όπως ο τύπος κατοικίας και η κατηγορία αστικότητας.

Επιπλέον, στο σενάριο αυτό αξιοποιούνται βοηθητικά προφίλ συμπεριφοράς που δημιουργούνται από το σύνολο εκπαίδευσης και αποτυπώνουν τη μέση καταναλωτική συμπεριφορά κατοικιών με παρόμοια χαρακτηριστικά. Παράλληλα, χρησιμοποιείται μηχανισμός εντοπισμού παρόμοιων κατοικιών μέσω του αλγορίθμου k-πλησιέστερων γειτόνων (k-nearest neighbors), ο οποίος βασίζεται σε στατικά χαρακτηριστικά, όπως ο τύπος κατοικίας, το εμβαδόν, ο αριθμός ενοίκων και ο αριθμός ηλεκτρικών συσκευών. Με τον τρόπο αυτό, τα μοντέλα μπορούν να αξιοποιούν πληροφορία από παρόμοιες κατοικίες χωρίς να απαιτείται ιστορικό κατανάλωσης της ίδιας της υπό εξέταση κατοικίας.

Στόχος του σεναρίου αυτού είναι η παραγωγή εύλογων προβλέψεων ακόμη και για νέες κατοικίες ή χρήστες που εντάσσονται στο σύστημα χωρίς διαθέσιμο ιστορικό κατανάλωσης.

Στο σενάριο με χρήση ιστορικών δεδομένων θεωρείται ότι για κάθε κατοικία υπάρχει διαθέσιμο πρόσφατο ιστορικό ωριαίων μετρήσεων. Στην περίπτωση αυτή, τα μοντέλα αξιοποιούν και χαρακτηριστικά που προκύπτουν από ένα ιστορικό παράθυρο διάρκειας 168 ωρών (μίας εβδομάδας). Από το παράθυρο αυτό εξάγονται χρονικά υστερούσες μεταβλητές και στατιστικά χαρακτηριστικά κινούμενων παραθύρων. Ενδεικτικά χρησιμοποιούνται χαρακτηριστικά όπως η κατανάλωση της προηγούμενης ώρας (lag_1h), της αντίστοιχης ώρας της προηγούμενης ημέρας (lag_24h), καθώς και μέσοι όροι κατανάλωσης σε διαφορετικά χρονικά παράθυρα.

Η ενσωμάτωση των ιστορικών χαρακτηριστικών επιτρέπει στα μοντέλα να αξιοποιούν τη χρονική εξάρτηση της κατανάλωσης και να αποτυπώνουν τόσο τη βραχυπρόθεσμη δυναμική του φορτίου όσο και επαναλαμβανόμενα ημερήσια και εβδομαδιαία πρότυπα συμπεριφοράς. Ιδιαίτερα σημαντικές θεωρούνται οι πρόσφατες μετρήσεις κατανάλωσης, καθώς η κατανάλωση των προηγούμενων ωρών ή ημερών αποτελεί συχνά ισχυρό προγνωστικό παράγοντα στη βραχυπρόθεσμη πρόβλεψη φορτίου.

Τα δύο αυτά σενάρια εξυπηρετούν διαφορετικές πρακτικές ανάγκες. Το σενάριο χωρίς ιστορικά δεδομένα προσομοιώνει περιπτώσεις νέων κατοικιών ή χρηστών για τους οποίους δεν υπάρχουν προηγούμενες μετρήσεις κατανάλωσης. Αντίθετα, το σενάριο με χρήση ιστορικών δεδομένων προσομοιώνει κατοικίες που διαθέτουν πρόσφατο ιστορικό κατανάλωσης, επιτρέποντας την αξιοποίηση της χρονικής συνέχειας της κατανάλωσης. Για τον λόγο αυτό, τα μοντέλα με χρήση ιστορικών δεδομένων αναμένεται να επιτυγχάνουν υψηλότερη ακρίβεια πρόβλεψης.

5.4 Διαχωρισμός σε σύνολο εκπαίδευσης και σύνολο ελέγχου (train/test)

Για την αξιολόγηση των μοντέλων εφαρμόζεται χρονικός διαχωρισμός των δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Αρχικά, όλες οι παρατηρήσεις ταξινομούνται χρονολογικά και στη συνέχεια διαχωρίζονται σε διαδοχικές χρονικές περιόδους, ώστε κάθε μοντέλο να εκπαιδεύεται αποκλειστικά σε δεδομένα προηγούμενων χρονικών περιόδων και να αξιολογείται σε μεταγενέστερες χρονικές περιόδους.

Η χρήση χρονικού, αντί τυχαίου, διαχωρισμού αποτρέπει τη διαρροή πληροφορίας από το μέλλον προς το παρελθόν και προσομοιώνει ρεαλιστικά τη λειτουργία ενός πραγματικού συστήματος πρόβλεψης, στο οποίο οι προβλέψεις παράγονται για χρονικές στιγμές που δεν έχουν ακόμη παρατηρηθεί (Hyndman & Athanasopoulos, 2018).

Για τα τελικά μοντέλα XGBoost και LightGBM χρησιμοποιήθηκε ο ακόλουθος χρονικός διαχωρισμός:

- Σύνολο εκπαίδευσης (train): από την αρχή του διαθέσιμου ιστορικού έως την 1η Μαΐου 2018.
- Σύνολο επικύρωσης για τον έλεγχο υπερπροσαρμογής των μοντέλων (early stopping): από 2 Μαΐου 2018 έως 16 Μαΐου 2018.
- Σύνολο επικύρωσης για βαθμονόμηση των προβλέψεων και αξιολόγηση πρόσθετων ρυθμίσεων: από 17 Μαΐου 2018 έως 31 Μαΐου 2018.
- Σύνολο ελέγχου (test): από 1 Ιουνίου 2018 έως 30 Ιουνίου 2018.

Η διάκριση των δύο συνόλων επικύρωσης επιτρέπει αφενός τον έλεγχο υπερπροσαρμογής κατά την εκπαίδευση των μοντέλων διαδοχικής ενίσχυσης δέντρων αποφάσεων (gradient boosting) και αφετέρου τη δίκαιη αξιολόγηση πιθανών προσαρμογών και ρυθμίσεων χωρίς χρήση του τελικού συνόλου ελέγχου.

Για τα μοντέλα Random Forest δεν εφαρμόζεται διαδικασία πρόωρης διακοπής της εκπαίδευσης (early stopping), καθώς ο αλγόριθμος δεν υποστηρίζει αντίστοιχο μηχανισμό. Ωστόσο, διατηρείται η ίδια χρονική λογική διαχωρισμού και το τελικό μοντέλο εκπαιδεύεται εκ νέου χρησιμοποιώντας όλα τα διαθέσιμα δεδομένα έως τις 31 Μαΐου 2018, ενώ ο Ιούνιος 2018 παραμένει αποκλειστικά ως ανεξάρτητο σύνολο ελέγχου.

Το τελικό σύνολο ελέγχου είναι κοινό για όλα τα μοντέλα και για τα δύο σενάρια πρόβλεψης (χωρίς ιστορικά δεδομένα και με χρήση ιστορικών δεδομένων), αντιστοιχώντας αποκλειστικά στον Ιούνιο του 2018. Με τον τρόπο αυτό εξασφαλίζεται η άμεση συγκρισιμότητα των αποτελεσμάτων μεταξύ διαφορετικών αλγορίθμων και διαφορετικών σεναρίων πρόβλεψης, καθώς όλα τα μοντέλα αξιολογούνται στην ίδια χρονική περίοδο και υπό τις ίδιες συνθήκες.

5.5 Μετασχηματισμοί στόχου και προεπεξεργασία

Πριν από την εκπαίδευση των μοντέλων εφαρμόζονται συγκεκριμένα βήματα προεπεξεργασίας τόσο στα δεδομένα εισόδου όσο και στη μεταβλητή στόχο. Αρχικά, από τη χρονική μεταβλητή *timestamp* εξάγονται χαρακτηριστικά που περιγράφουν τη χρονική θέση κάθε παρατήρησης, όπως η ώρα της ημέρας (*hour*), η ημέρα της εβδομάδας (*day_of_week*), ο μήνας (*month*) και η ένδειξη Σαββατοκύριακου (*is_weekend*). Επιπλέον, σε ορισμένα μοντέλα χρησιμοποιούνται και κυκλικές αναπαραστάσεις των χρονικών μεταβλητών μέσω ημιτονοειδών και συνημιτονοειδών μετασχηματισμών, ώστε να αποτυπώνεται ο κυκλικός χαρακτήρας της ώρας, της ημέρας και του μήνα. Στη συνέχεια, οι αριθμητικές μεταβλητές μετατρέπονται σε κατάλληλη αριθμητική μορφή και η μεταβλητή στόχος *consumption_Wh* καθαρίζεται από μη έγκυρες ή αρνητικές τιμές. Για τη βελτίωση της εκπαίδευσης των μοντέλων εφαρμόζεται λογαριθμικός μετασχηματισμός στη μεταβλητή στόχο. Συγκεκριμένα, αντί της αρχικής κατανάλωσης (y), τα μοντέλα εκπαιδεύονται στην ποσότητα:

$$z = \ln(1 + y) \quad (1)$$

Ο μετασχηματισμός αυτός μειώνει την επίδραση των πολύ μεγάλων τιμών κατανάλωσης, περιορίζει την ασυμμετρία της κατανομής και συμβάλλει στη σταθερότερη εκπαίδευση των μοντέλων (Ahmad et al., 2018).

Μετά την ολοκλήρωση της εκπαίδευσης, οι προβλέψεις μετατρέπονται ξανά στην αρχική κλίμακα κατανάλωσης μέσω της αντίστροφης σχέσης:

$$y = e^z - 1 \quad (2)$$

ώστε τα τελικά αποτελέσματα να εκφράζονται σε Wh, δηλαδή στην πραγματική μονάδα μέτρησης της κατανάλωσης.

Κατά τη διάρκεια της έρευνας εξετάστηκε επίσης η δυνατότητα περιορισμού ακραίων τιμών της μεταβλητής στόχου μέσω κατάλληλων ποσοστημοριών. Ωστόσο, η τελική διαμόρφωση των μοντέλων βασίστηκε κυρίως στον λογαριθμικό μετασχηματισμό, χωρίς να απαιτείται συστηματική εφαρμογή περιορισμού ακραίων τιμών.

Οι ελλειπείς τιμές των συνεχών μεταβλητών συμπληρώνονται με τη διάμεσο τιμή της αντίστοιχης μεταβλητής. Η διάμεσος υπολογίζεται αποκλειστικά από το σύνολο εκπαίδευσης και στη συνέχεια εφαρμόζεται στα σύνολα επικύρωσης και ελέγχου, ώστε να

αποφεύγεται οποιαδήποτε διαρροή πληροφορίας από τα δεδομένα αξιολόγησης προς τη διαδικασία εκπαίδευσης.

Για τις κατηγορικές μεταβλητές, όπως ο τύπος κατοικίας (*hometype*) και η κατηγορία αστικότητας (*urban_rural_class*), χρησιμοποιείται δυαδική κωδικοποίηση (*one-hot encoding*). Η κωδικοποίηση βασίζεται αποκλειστικά στις κατηγορίες που εμφανίζονται στο σύνολο εκπαίδευσης, ενώ οι αντίστοιχες μεταβλητές των συνόλων επικύρωσης και ελέγχου ευθυγραμμίζονται ώστε να περιλαμβάνουν ακριβώς τα ίδια πεδία.

Στο σενάριο με χρήση ιστορικών δεδομένων δημιουργούνται επιπλέον χαρακτηριστικά χρονικής εξάρτησης, όπως χρονικές υστερήσεις και στατιστικά χαρακτηριστικά κινούμενων παραθύρων. Τα χαρακτηριστικά αυτά περιλαμβάνουν μεταβλητές που αποτυπώνουν την κατανάλωση προηγούμενων ωρών και ημερών, καθώς και στατιστικά μεγέθη που υπολογίζονται σε διαφορετικά χρονικά παράθυρα. Ενδεικτικά χρησιμοποιούνται χρονικές υστερήσεις 1, 2, 3, 6, 12, 24, 48, 72 και 168 ωρών, καθώς και στατιστικά χαρακτηριστικά κινούμενων παραθύρων διάρκειας από 3 έως 168 ώρες. Όλα τα χαρακτηριστικά υπολογίζονται αποκλειστικά από προηγούμενες χρονικές στιγμές της ίδιας κατοικίας, ώστε να διασφαλίζεται ότι καμία πληροφορία από το μέλλον δεν χρησιμοποιείται κατά την εκπαίδευση ή την πρόβλεψη.

Στο σενάριο χωρίς ιστορικά δεδομένα χρησιμοποιείται επιπλέον αλγόριθμος *k*-πλησιέστερων γειτόνων για τον εντοπισμό κατοικιών με παρόμοια στατικά χαρακτηριστικά. Για τον υπολογισμό των αποστάσεων μεταξύ κατοικιών εφαρμόζεται τυποποίηση των σχετικών μεταβλητών. Η διαδικασία αυτή χρησιμοποιείται αποκλειστικά για τον υπολογισμό ομοιότητας μεταξύ κατοικιών και όχι ως γενικό βήμα προεπεξεργασίας των χαρακτηριστικών που χρησιμοποιούνται από τα μοντέλα πρόβλεψης.

Τέλος, όλες οι παράμετροι προεπεξεργασίας (τιμές συμπλήρωσης ελλিপών δεδομένων, κωδικοποιήσεις κατηγορικών μεταβλητών και λοιπές ρυθμίσεις μετασχηματισμού) αποθηκεύονται μαζί με τα εκπαιδευμένα μοντέλα, ώστε η ίδια ακριβώς διαδικασία να εφαρμόζεται κατά τη χρήση του συστήματος πρόβλεψης μέσω του προγραμματιστικού περιβάλλοντος και της διεπαφής χρήστη.

5.6 Επιλογή αλγορίθμων μηχανικής μάθησης

Τα δεδομένα που χρησιμοποιούνται στην παρούσα εργασία έχουν σαφώς πινακοποιημένη δομή. Για κάθε ωριαία εγγραφή κατοικίας υπάρχουν πολλαπλά αριθμητικά χαρακτηριστικά (χρονικά, περιβαλλοντικά και ιστορικά), καθώς και κατηγορικές μεταβλητές, όπως ο τύπος κατοικίας και η κατηγορία αστικότητας. Σύμφωνα με τη σχετική βιβλιογραφία για τη μοντελοποίηση κτηρίων και την πρόβλεψη ενεργειακής κατανάλωσης, τα μοντέλα που βασίζονται σε δέντρα αποφάσεων και τα συνδυαστικά σύνολα αυτών παρουσιάζουν ιδιαίτερα καλή απόδοση σε τέτοιου είδους δεδομένα, καθώς μπορούν να αποτυπώσουν μη γραμμικές σχέσεις, αλληλεπιδράσεις μεταξύ χαρακτηριστικών και συνδυασμούς αριθμητικών και κατηγορικών μεταβλητών χωρίς να απαιτούν εκτεταμένους μετασχηματισμούς των εισόδων (Al Shargabi et al., 2022; Olu Ajayi et al., 2023).

Στη μορφή που χρησιμοποιείται στην παρούσα εργασία, το σύνολο δεδομένων IDEAL αποτελεί ένα μεγάλο και ετερογενές σύνολο ωριαίων παρατηρήσεων, το οποίο περιλαμβάνει χρονικά χαρακτηριστικά, περιβαλλοντικές μεταβλητές, μεταδεδομένα κατοικίας και νοικοκυριού, καθώς και ιστορικά χαρακτηριστικά κατανάλωσης στα σενάρια όπου αυτά είναι διαθέσιμα. Συνεπώς, το πρόβλημα διαμορφώνεται ως πρόβλημα πολυπαραγοντικής παλινδρόμησης σε πινακοποιημένα δεδομένα.

Με βάση τα παραπάνω κριτήρια επιλέχθηκαν τρία μοντέλα μηχανικής μάθησης από την οικογένεια των συνδυαστικών δένδροειδών μεθόδων: το Random Forest, ως σχετικά απλή και ανθεκτική μέθοδος αναφοράς και τα XGBoost και LightGBM, ως σύγχρονες υλοποιήσεις ενισχυτικής μάθησης με δέντρα αποφάσεων που έχουν χρησιμοποιηθεί με επιτυχία σε προβλήματα βραχυπρόθεσμης πρόβλεψης φορτίου και ενεργειακής ζήτησης (Zhang et al., 2018; Olu Ajayi et al., 2023).

Παρά τη συγγενειά τους, τα μοντέλα αυτά ακολουθούν διαφορετική φιλοσοφία εκπαίδευσης. Το Random Forest βασίζεται στη μέθοδο της δειγματοληψίας με επανατοποθέτηση (bagging), δηλαδή στην παράλληλη εκπαίδευση πολλών ανεξάρτητων δέντρων με στοιχεία τυχαιοποίησης, γεγονός που προσφέρει υψηλή ανθεκτικότητα στην υπερπροσαρμογή και περιορισμένη ανάγκη για λεπτομερή ρύθμιση υπερπαραμέτρων. Αντίθετα, τα XGBoost και LightGBM βασίζονται σε μεθόδους διαδοχικής ενίσχυσης δέντρων αποφάσεων (boosting), κατά την οποία νέα δέντρα προστίθενται διαδοχικά με στόχο τη διόρθωση των σφαλμάτων των προηγούμενων μοντέλων, επιτρέποντας την αποτύπωση πιο σύνθετων μη γραμμικών σχέσεων μεταξύ των χαρακτηριστικών και της μεταβλητής στόχου (Al Shargabi et al., 2022; Olu Ajayi et al., 2023).

Η πρόσφατη βιβλιογραφία στη βραχυπρόθεσμη πρόβλεψη φορτίου συγκλίνει στο ότι οι συνδυαστικές δένδροειδείς μέθοδοι αποτελούν ιδιαίτερα ανταγωνιστικές επιλογές για δεδομένα αυτού του τύπου, καθώς μπορούν να διαχειριστούν αποτελεσματικά μεγάλους αριθμούς χαρακτηριστικών, πολύπλοκες αλληλεπιδράσεις και μη γραμμικές σχέσεις, ενώ συχνά υπερέχουν έναντι απλούστερων στατιστικών προσεγγίσεων (Al Shargabi et al., 2022; Olu Ajayi et al., 2023).

Παράλληλα, κατά τη διάρκεια της έρευνας εξετάστηκαν και εναλλακτικές προσεγγίσεις πρόβλεψης, όπως τεχνικές βαθιάς μάθησης για χρονοσειρές και στρατηγικές μεταγενέστερης διόρθωσης των προβλέψεων. Ειδικότερα, μελετήθηκαν σύγχρονες αρχιτεκτονικές βαθιάς μάθησης, όπως τα μοντέλα Temporal Fusion Transformer (TFT) και N-HiTS, οι οποίες έχουν παρουσιάσει αξιολογικά αποτελέσματα σε προβλήματα πρόβλεψης χρονοσειρών. Ωστόσο, τα συνδυαστικά δένδροειδή μοντέλα παρουσίασαν σταθερά υψηλή απόδοση στα δεδομένα της παρούσας εργασίας, καλή δυνατότητα γενίκευσης, μικρότερη υπολογιστική πολυπλοκότητα και ευκολότερη ενσωμάτωση στο τελικό σύστημα πρόβλεψης. Για τον λόγο αυτό επιλέχθηκαν ως ο βασικός άξονας της μεθοδολογίας και ως τα μοντέλα που ενσωματώθηκαν στην τελική εφαρμογή. Η επιλογή αυτή ευθυγραμμίζεται και με τον πρακτικό στόχο της εργασίας για την ανάπτυξη ενός ρεαλιστικού, επεκτάσιμου και λειτουργικού συστήματος πρόβλεψης μέσω διεπαφής προγραμματισμού εφαρμογών (API) και γραφικής διεπαφής χρήστη.

Τα συγκεκριμένα μοντέλα χαρακτηρίζονται από σχετικά γρήγορη εκπαίδευση και εκτέλεση, περιορισμένες απαιτήσεις προεπεξεργασίας και δυνατότητα ερμηνείας μέσω της ανάλυσης σημαντικότητας χαρακτηριστικών, γεγονός που τα καθιστά ιδιαίτερα κατάλληλα για πρακτικές εφαρμογές πρόβλεψης ηλεκτρικής κατανάλωσης επόμενης ημέρας.

Για την εκπαίδευση των μοντέλων πραγματοποιήθηκε διαδικασία επιλογής υπερπαραμέτρων μέσω συνδυασμού περιορισμένης αναζήτησης σε πλέγμα τιμών και επαναλαμβανόμενων πειραματικών δοκιμών. Η διαδικασία αυτή εφαρμόστηκε στα βασικά μοντέλα μηχανικής μάθησης που χρησιμοποιήθηκαν στην εργασία, δηλαδή στο Random Forest, στο XGBoost και στο LightGBM, με στόχο την εύρεση ρυθμίσεων που προσφέρουν ικανοποιητική ισορροπία μεταξύ ακρίβειας πρόβλεψης, δυνατότητας γενίκευσης και υπολογιστικής αποδοτικότητας.

Η αναζήτηση επικεντρώθηκε σε βασικές υπερπαραμέτρους που επηρεάζουν σημαντικά τη συμπεριφορά των μοντέλων, όπως το πλήθος των εκτιμητών, το μέγιστο βάθος ή η πολυπλοκότητα των δέντρων, ο ρυθμός μάθησης στα ενισχυτικά μοντέλα και οι παράμετροι

κανονικοποίησης, ανάλογα με τον εκάστοτε αλγόριθμο. Η διαδικασία δεν είχε τον χαρακτήρα πλήρους εξαντλητικής αναζήτησης όλων των πιθανών συνδυασμών, καθώς κάτι τέτοιο θα αύξανε σημαντικά το υπολογιστικό κόστος, ιδιαίτερα λόγω του μεγάλου όγκου των δεδομένων, των δύο σεναρίων πρόβλεψης και της ανάγκης εκπαίδευσης πολλαπλών μοντέλων. Οι τελικές ρυθμίσεις επιλέχθηκαν με βάση την απόδοση στο σύνολο επικύρωσης, τη σταθερότητα των προβλέψεων και τη δυνατότητα ενσωμάτωσης των μοντέλων στο τελικό σύστημα πρόβλεψης.

Τα τελικά μοντέλα, μαζί με τις αντίστοιχες παραμέτρους προεπεξεργασίας και τις ρυθμίσεις χαρακτηριστικών, αποθηκεύτηκαν ως συνοδευτικά αρχεία υλοποίησης, ώστε να μπορούν να χρησιμοποιηθούν από την τελική εφαρμογή πρόβλεψης. Με τον τρόπο αυτό διασφαλίστηκε ότι η διαδικασία παραγωγής προβλέψεων στο στάδιο της εφαρμογής ακολουθεί την ίδια λογική με τη διαδικασία εκπαίδευσης και αξιολόγησης.

5.7 Random Forest

Το μοντέλο Random Forest αποτελεί μία από τις πιο διαδεδομένες συνδυαστικές δενδροειδείς μεθόδους παλινδρόμησης. Αντί να εκπαιδεύεται ένα μόνο δέντρο απόφασης, δημιουργείται ένα σύνολο πολλών δέντρων, τα οποία εκπαιδεύονται σε διαφορετικά τυχαία υποσύνολα των δεδομένων και των χαρακτηριστικών. Η τελική πρόβλεψη προκύπτει ως ο μέσος όρος των προβλέψεων όλων των δέντρων, γεγονός που μειώνει τη διακύμανση των αποτελεσμάτων και περιορίζει την υπερπροσαρμογή που συχνά εμφανίζουν τα μεμονωμένα δέντρα απόφασης. Για τον λόγο αυτό, η μέθοδος χρησιμοποιείται ευρέως σε προβλήματα πρόβλεψης ενεργειακής κατανάλωσης και βραχυπρόθεσμης πρόβλεψης φορτίου (Al Shargabi et al., 2022; Olu Ajayi et al., 2023).

Στην παρούσα εργασία χρησιμοποιείται ο αλγόριθμος Random Forest Regressor ως μία ισχυρή μέθοδος αναφοράς για τη σύγκριση με τα μοντέλα ενισχυτικής μάθησης. Υλοποιούνται δύο διαφορετικές εκδοχές του μοντέλου, οι οποίες αντιστοιχούν στα δύο σενάρια πρόβλεψης που παρουσιάστηκαν στην Ενότητα 5.3.

Η πρώτη εκδοχή αφορά το σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης (RF no-history). Στην περίπτωση αυτή το μοντέλο εκπαιδεύεται χρησιμοποιώντας αποκλειστικά χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά της κατοικίας, όπως η ώρα της ημέρας, η ημέρα της εβδομάδας, η ένδειξη Σαββατοκύριακου, ο μήνας, η εξωτερική θερμοκρασία, το εμβαδόν της κατοικίας, ο αριθμός ενοίκων, ο τύπος κατοικίας και η κατηγορία αστικότητας. Η τελική διαμόρφωση περιλαμβάνει 200 δέντρα απόφασης ($n_estimators = 200$), μέγιστο βάθος 20 επιπέδων ($max_depth = 20$), ελάχιστο αριθμό 20 παρατηρήσεων για τη δημιουργία νέου διαχωρισμού ($min_samples_split = 20$), ελάχιστο μέγεθος φύλλου 10 παρατηρήσεων ($min_samples_leaf = 10$) και χρήση έως του 70% των διαθέσιμων χαρακτηριστικών σε κάθε διαχωρισμό ($max_features = 0,7$).

Η δεύτερη εκδοχή αφορά το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης (RF with-history). Εκτός από τα χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά, το μοντέλο αξιοποιεί εκτεταμένο σύνολο ιστορικών χαρακτηριστικών, όπως χρονικές υστερήσεις πολλαπλών χρονικών οριζόντων, στατιστικά χαρακτηριστικά κινούμενων παραθύρων, καθώς και πρόσθετα χαρακτηριστικά που περιγράφουν τη μακροχρόνια καταναλωτική συμπεριφορά κάθε κατοικίας.

Σε ορισμένες πειραματικές εκδοχές, η ταυτότητα της κατοικίας χρησιμοποιήθηκε για τη διερεύνηση της επίδρασης της εξατομικευμένης πληροφορίας. Ωστόσο, στα τελικά μοντέλα που συνδέθηκαν με την εφαρμογή πρόβλεψης δεν χρησιμοποιείται το αναγνωριστικό της

κατοικίας ως χαρακτηριστικό εισόδου, ώστε το σύστημα να μπορεί να εφαρμοστεί και σε νέες κατοικίες. Η τελική διαμόρφωση περιλαμβάνει 120 δέντρα απόφασης ($n_estimators = 120$), μέγιστο βάθος 20 επιπέδων ($max_depth = 20$), ελάχιστο αριθμό 30 παρατηρήσεων για νέο διαχωρισμό ($min_samples_split = 30$), ελάχιστο μέγεθος φύλλου 15 παρατηρήσεων ($min_samples_leaf = 15$), χρήση έως του 55% των διαθέσιμων χαρακτηριστικών σε κάθε διαχωρισμό ($max_features = 0,55$) και δειγματοληψία του 80% των διαθέσιμων παρατηρήσεων για κάθε δέντρο.

Οι τελικές ρυθμίσεις των δύο εκδόσεων επιλέχθηκαν έπειτα από επαναλαμβανόμενες πειραματικές δοκιμές και αξιολόγηση της απόδοσης των μοντέλων, με στόχο την επίτευξη ισορροπίας μεταξύ ακρίβειας πρόβλεψης, δυνατότητας γενίκευσης και υπολογιστικού κόστους.

Η παράλληλη αξιολόγηση των δύο εκδοχών επιτρέπει τη διερεύνηση της συμβολής των ιστορικών δεδομένων στην απόδοση του ίδιου αλγορίθμου, διατηρώντας σταθερό το υποκείμενο μοντέλο μηχανικής μάθησης. Με τον τρόπο αυτό καθίσταται δυνατή η άμεση σύγκριση μεταξύ των δύο σεναρίων πρόβλεψης και η ποσοτική αποτίμηση της προστιθέμενης αξίας που προσφέρει η ύπαρξη ιστορικού κατανάλωσης.

Για λόγους υπολογιστικής εφικτότητας, η εκπαίδευση της εκδοχής με ιστορικά δεδομένα πραγματοποιείται σε αντιπροσωπευτικό δείγμα του συνόλου εκπαίδευσης, διατηρώντας κατοικίες από όλο το διαθέσιμο χρονικό εύρος.

5.8 XGBoost

Το μοντέλο XGBoost (Extreme Gradient Boosting) ανήκει στην οικογένεια των μεθόδων ενισχυτικής μάθησης με δέντρα αποφάσεων και αποτελεί μία από τις πιο επιτυχημένες προσεγγίσεις για προβλήματα παλινδρόμησης σε δεδομένα πινακοποιημένης μορφής. Σε αντίθεση με τις μεθόδους τύπου bagging, όπως το Random Forest, το XGBoost δημιουργεί διαδοχικά μία ακολουθία δέντρων, όπου κάθε νέο δέντρο εκπαιδεύεται με στόχο τη διόρθωση των σφαλμάτων που παραμένουν από τα προηγούμενα. Με τον τρόπο αυτό μπορεί να αποτυπώσει σύνθετες μη γραμμικές σχέσεις μεταξύ των χαρακτηριστικών εισόδου και της μεταβλητής στόχου, ενώ παράλληλα ενσωματώνει μηχανισμούς τακτοποίησης (regularization) που συμβάλλουν στον περιορισμό της υπερπροσαρμογής (Chen & Guestrin, 2016).

Στην παρούσα εργασία χρησιμοποιείται ο αλγόριθμος XGBRegressor ως μία από τις βασικές μεθόδους πρόβλεψης. Για το σενάριο με χρήση ιστορικών δεδομένων η τελική διαμόρφωση περιλαμβάνει 5000 δέντρα απόφασης, ρυθμό μάθησης 0,015, μέγιστο βάθος οκτώ επιπέδων, ελάχιστο βάρος κόμβου 20, ποσοστό δειγματοληψίας παρατηρήσεων 90% και ποσοστό δειγματοληψίας χαρακτηριστικών 90%. Για το σενάριο χωρίς ιστορικά δεδομένα χρησιμοποιείται αντίστοιχη διαμόρφωση με 5500 δέντρα, μέγιστο βάθος επτά επιπέδων και ελάχιστο βάρος κόμβου 25. Επιπλέον εφαρμόζονται μηχανισμοί τακτοποίησης μέσω των παραμέτρων reg_alpha και reg_lambda , ενώ η εκπαίδευση πραγματοποιείται διαδοχικά με την προσθήκη νέων δέντρων που διορθώνουν τα σφάλματα των προηγούμενων. Για τον περιορισμό της υπερπροσαρμογής χρησιμοποιείται διαδικασία πρόωρης διακοπής της εκπαίδευσης με βάση ανεξάρτητο σύνολο επικύρωσης. Ως συνάρτηση απώλειας χρησιμοποιείται το τετραγωνικό σφάλμα παλινδρόμησης. Οι συγκεκριμένες ρυθμίσεις προέκυψαν έπειτα από επαναλαμβανόμενες δοκιμές και αξιολόγηση της απόδοσης του μοντέλου, επιδιώκοντας ισορροπία μεταξύ ακρίβειας πρόβλεψης, δυνατότητας γενίκευσης και υπολογιστικού κόστους.

Για τα μοντέλα XGBoost εφαρμόζεται διαδικασία ελέγχου υπερπροσαρμογής μέσω συνόλου επικύρωσης, όπως περιγράφηκε στην Ενότητα 5.4. Με τον τρόπο αυτό η εκπαίδευση διακόπτεται όταν η απόδοση παύει να βελτιώνεται ουσιαστικά, συμβάλλοντας στη δημιουργία μοντέλων με καλύτερη δυνατότητα γενίκευσης σε νέα δεδομένα.

Η παράλληλη αξιολόγηση των δύο εκδοχών του XGBoost επιτρέπει τη διερεύνηση της επίδρασης των ιστορικών δεδομένων στην απόδοση του ίδιου αλγορίθμου, διατηρώντας σταθερό τον μηχανισμό μάθησης. Έτσι καθίσταται δυνατή η άμεση σύγκριση μεταξύ των δύο σεναρίων πρόβλεψης και η ποσοτική εκτίμηση του οφέλους που προσφέρει η αξιοποίηση ιστορικού κατανάλωσης.

5.9 LightGBM

Το LightGBM (Light Gradient Boosting Machine) αποτελεί μία σύγχρονη υλοποίηση αλγορίθμων gradient boosting με δέντρα απόφασης και έχει σχεδιαστεί με έμφαση στην υπολογιστική αποδοτικότητα και την ικανότητα διαχείρισης μεγάλων συνόλων δεδομένων. Σε αντίθεση με τις κλασικές υλοποιήσεις boosting, χρησιμοποιεί τεχνικές όπως η ανάπτυξη δέντρων με βάση τα φύλλα (leaf-wise tree growth) και η ομαδοποίηση τιμών μέσω ιστογραμμάτων (histogram-based learning), επιτυγχάνοντας υψηλή ακρίβεια πρόβλεψης με μειωμένο υπολογιστικό κόστος. Για τον λόγο αυτό χρησιμοποιείται ευρέως σε προβλήματα πρόβλεψης ενεργειακής κατανάλωσης και βραχυπρόθεσμης πρόβλεψης φορτίου (Zhang et al., 2018; Olu Ajayi et al., 2023).

Στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης (LGBM with-history), η τελική διαμόρφωση περιλαμβάνει 9000 δέντρα απόφασης ($n_estimators = 9000$), ρυθμό μάθησης 0,012 ($learning_rate = 0,012$), μέγιστο αριθμό φύλλων 127 ($num_leaves = 127$), ελάχιστο πλήθος δειγμάτων ανά φύλλο 80 ($min_child_samples = 80$), ποσοστό δειγματοληψίας παρατηρήσεων 90% ($subsample = 0,9$) και ποσοστό δειγματοληψίας χαρακτηριστικών 90% ($colsample_bytree = 0,9$). Επιπλέον εφαρμόζονται μηχανισμοί τακτοποίησης μέσω των παραμέτρων $reg_alpha = 0,05$ και $reg_lambda = 0,20$. Στην τελική αυτή διαμόρφωση η ταυτότητα της κατοικίας ($home_id$) δεν χρησιμοποιείται ως χαρακτηριστικό εισόδου, ώστε το μοντέλο να μπορεί να χρησιμοποιηθεί σε γενικευμένα σενάρια πρόβλεψης μέσω της διεπαφής χρήστη.

Στο σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης (LGBM no-history), το μοντέλο χρησιμοποιεί χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά της κατοικίας, καθώς και πρόσθετα χαρακτηριστικά που προκύπτουν από προφίλ συμπεριφοράς (behavioral profiles) και από προφίλ παρόμοιων κατοικιών μέσω μεθόδων πλησιέστερων γειτόνων (KNN similar-home profiles). Η τελική διαμόρφωση περιλαμβάνει 7000 δέντρα απόφασης ($n_estimators = 7000$), ρυθμό μάθησης 0,015 ($learning_rate = 0,015$), μέγιστο αριθμό φύλλων 63 ($num_leaves = 63$), ελάχιστο πλήθος δειγμάτων ανά φύλλο 80 ($min_child_samples = 80$), ποσοστό δειγματοληψίας παρατηρήσεων 90% ($subsample = 0,9$) και ποσοστό δειγματοληψίας χαρακτηριστικών 90% ($colsample_bytree = 0,9$), ενώ εφαρμόζονται οι ίδιες παράμετροι τακτοποίησης με το μοντέλο με ιστορικά δεδομένα.

Για την εκπαίδευση και των δύο εκδοχών χρησιμοποιείται διαδικασία πρόωρης διακοπής της εκπαίδευσης με βάση ανεξάρτητο σύνολο επικύρωσης. Με τον τρόπο αυτό η εκπαίδευση σταματά όταν η απόδοση παύει να βελτιώνεται ουσιαστικά, περιορίζοντας τον κίνδυνο υπερπροσαρμογής και βελτιώνοντας τη δυνατότητα γενίκευσης του μοντέλου σε νέα δεδομένα.

Οι τελικές ρυθμίσεις επιλέχθηκαν έπειτα από εκτεταμένες πειραματικές δοκιμές και συγκριτική αξιολόγηση διαφορετικών διαμορφώσεων υπερπαραμέτρων, με στόχο την επίτευξη υψηλής ακρίβειας πρόβλεψης, καλής δυνατότητας γενίκευσης και αποδεκτού υπολογιστικού κόστους. Το LightGBM αποτελεί ένα από τα βασικά μοντέλα που χρησιμοποιούνται στην παρούσα εργασία τόσο για την αξιολόγηση των διαφορετικών σεναρίων πρόβλεψης όσο και για την υλοποίηση της τελικής εφαρμογής πρόβλεψης επόμενης ημέρας.

6. Αποτελέσματα και αξιολόγηση βασικών μοντέλων αναφοράς και μοντέλων μηχανικής μάθησης

Στο παρόν κεφάλαιο παρουσιάζονται και αξιολογούνται τα πειραματικά αποτελέσματα των βασικών μοντέλων αναφοράς και των μοντέλων μηχανικής μάθησης στο σύνολο δεδομένων IDEAL. Αρχικά, περιγράφονται οι ρυθμίσεις των πειραμάτων, ο χρονικός διαχωρισμός των δεδομένων και οι μετρικές αξιολόγησης που χρησιμοποιούνται για την αποτίμηση της ακρίβειας των προβλέψεων. Στη συνέχεια, παρουσιάζεται η απόδοση των απλών στατιστικών μοντέλων αναφοράς, τα οποία λειτουργούν ως αρχική βάση σύγκρισης. Ακολουθεί η αξιολόγηση των μοντέλων Random Forest, XGBoost και LightGBM στα δύο βασικά σενάρια πρόβλεψης: στο σενάριο χωρίς χρήση ιστορικών δεδομένων και στο σενάριο με ενσωμάτωση ιστορικών χαρακτηριστικών. Τέλος, πραγματοποιείται συνολική σύγκριση των αποτελεσμάτων, με στόχο να αποτιμηθεί η προστιθέμενη αξία των μοντέλων μηχανικής μάθησης έναντι των βασικών μοντέλων αναφοράς, καθώς και η επίδραση της πρόσφατης ιστορικής κατανάλωσης στη βελτίωση της προβλεπτικής απόδοσης.

6.1 Μετρικές αξιολόγησης και πειραματική διαδικασία

Όλα τα μοντέλα αξιολογούνται στο ίδιο ανεξάρτητο σύνολο ελέγχου, το οποίο αντιστοιχεί στον Ιούνιο του 2018. Με τον τρόπο αυτό εξασφαλίζεται η άμεση συγκρισιμότητα των αποτελεσμάτων μεταξύ διαφορετικών αλγορίθμων και διαφορετικών σεναρίων πρόβλεψης. Για την αξιολόγηση της απόδοσης χρησιμοποιούνται τέσσερις βασικές μετρικές σφάλματος: το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error – MAE), η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error – RMSE), το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error – MAPE) και το Απόλυτο Ημερήσιο Σφάλμα Κατανάλωσης (Daily Absolute Error).

Το Μέσο Απόλυτο Σφάλμα (MAE) εκφράζει τη μέση απόλυτη απόκλιση μεταξύ πραγματικών και προβλεπόμενων τιμών κατανάλωσης και υπολογίζεται ως:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (3)$$

όπου y_i είναι η πραγματική τιμή κατανάλωσης, $\hat{y}_i$ η αντίστοιχη πρόβλεψη και N το πλήθος των παρατηρήσεων.

Η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) δίνει μεγαλύτερη βαρύτητα στα μεγάλα σφάλματα και υπολογίζεται ως:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (4)$$

Η μετρική αυτή είναι ιδιαίτερα χρήσιμη σε προβλήματα πρόβλεψης φορτίου, καθώς οι μεγάλες αποκλίσεις στις ώρες αιχμής επηρεάζουν περισσότερο το τελικό αποτέλεσμα.

Για την αξιολόγηση της σχετικής ακρίβειας χρησιμοποιείται το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (MAPE), το οποίο εκφράζει το σφάλμα ως ποσοστό της πραγματικής κατανάλωσης. Στην τυπική του μορφή, το MAPE υπολογίζεται ως:

$$MAPE = \frac{100}{N} \sum_{i=1}^N \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (5)$$

Για την εργασία η μετρική MAPE εφαρμόζεται σε «ασφαλή» μορφή (*safe MAPE*), ώστε να αποφεύγεται η αστάθεια που προκαλείται από παρατηρήσεις με μηδενική ή πολύ μικρή πραγματική κατανάλωση. Συγκεκριμένα, από τον υπολογισμό εξαιρούνται οι παρατηρήσεις στις οποίες η πραγματική τιμή κατανάλωσης είναι σχεδόν μηδενική, δηλαδή μικρότερη από ένα πολύ μικρό θετικό κατώφλι. Με τον τρόπο αυτό, η μετρική δεν επηρεάζεται δυσανάλογα από τιμές κοντά στο μηδέν και παραμένει πιο αντιπροσωπευτική της σχετικής ακρίβειας των μοντέλων.

Εκτός από τις ωριαίες μετρικές, αξιολογείται και η ακρίβεια πρόβλεψης της συνολικής ημερήσιας κατανάλωσης μέσω του Απόλυτου Ημερήσιου Σφάλματος Κατανάλωσης (*Daily Absolute Error*). Η μετρική αυτή υπολογίζει την απόλυτη διαφορά μεταξύ της πραγματικής και της προβλεπόμενης ημερήσιας κατανάλωσης:

$$\text{Daily Absolute Error} = |E_{real} - E_{pred}| \quad (6)$$

όπου E_{real} είναι η πραγματική συνολική ημερήσια κατανάλωση και E_{pred} η αντίστοιχη προβλεπόμενη ημερήσια κατανάλωση. Η μετρική εκφράζεται σε kWh και επιτρέπει την αξιολόγηση της συνολικής ενεργειακής ακρίβειας των μοντέλων.

Η χρήση πολλαπλών μετρικών κρίνεται απαραίτητη, καθώς κάθε μετρική αποτυπώνει διαφορετικές πτυχές της ποιότητας των προβλέψεων. Το MAE αποτυπώνει το μέσο μέγεθος του σφάλματος, το RMSE είναι περισσότερο ευαίσθητο σε μεγάλες αποκλίσεις και σφάλματα αιχμών κατανάλωσης, το MAPE εκφράζει το σφάλμα σε ποσοστιαία μορφή ανεξάρτητα από την κλίμακα της κατανάλωσης, ενώ το Απόλυτο Ημερήσιο Σφάλμα Κατανάλωσης αξιολογεί την ικανότητα του μοντέλου να εκτιμά σωστά τη συνολική ημερήσια ενεργειακή ζήτηση. Η συνδυαστική χρήση των παραπάνω μετρικών επιτρέπει μια πιο ολοκληρωμένη και αξιόπιστη αξιολόγηση της απόδοσης των μοντέλων.

Στις ενότητες που ακολουθούν, οι τιμές που παρουσιάζονται στους πίνακες αποτελεσμάτων αποτελούν μέσους όρους των αντίστοιχων μετρικών για όλες τις κατοικίες και όλες τις διαθέσιμες χρονικές εγγραφές του συνόλου ελέγχου, αποτυπώνοντας έτσι τη συνολική μέση απόδοση κάθε μοντέλου στο ανεξάρτητο σύνολο αξιολόγησης.

6.2 Απόδοση των βασικών μοντέλων αναφοράς

Όπως αναφέρθηκε στην Ενότητα 4.7, πριν από την εφαρμογή των μοντέλων μηχανικής μάθησης υλοποιήθηκαν τέσσερα απλά στατιστικά μοντέλα πρόβλεψης σε ωριαία κλίμακα, με σκοπό να αποτελέσουν σημείο αναφοράς για την αξιολόγηση της προστιθέμενης αξίας των πιο σύνθετων προσεγγίσεων. Τα βασικά μοντέλα αναφοράς δεν αποσκοπούν στην επίτευξη της βέλτιστης δυνατής ακρίβειας, αλλά στη δημιουργία ενός ελάχιστου επιπέδου απόδοσης, έναντι του οποίου μπορούν να συγκριθούν τα μοντέλα μηχανικής μάθησης. Τα μοντέλα αυτά και οι επιδόσεις τους παρουσιάζονται στον Πίνακα 1. Για τα τρία πρώτα βασικά μοντέλα αναφοράς, δεν αναφέρεται τιμή MAPE, καθώς τα μοντέλα αυτά χρησιμοποιήθηκαν μόνο ως αρχικές γραμμές βάσης και εμφάνισαν σαφώς ασθενέστερη απόδοση σε όρους MAE και RMSE. Η τιμή MAPE παρουσιάζεται για το τέταρτο μοντέλο, το οποίο αποτέλεσε το ισχυρότερο βασικό μοντέλο αναφοράς και χρησιμοποιήθηκε ως κύριο σημείο σύγκρισης με τα μοντέλα μηχανικής μάθησης.

Μοντέλο	Περιγραφή	MAE (Wh/h)	RMSE (Wh/h)	MAPE (%)
Baseline 01 – Previous day, same hour	Κατανάλωση ώρας t ίση με κατανάλωση ίδιας ώρας προηγούμενης μέρας ($t - 24$).	224,0	462,0	-
Baseline 02 – Global mean by hour	Μέση κατανάλωση ανά ώρα της ημέρας, σε όλες τις κατοικίες/ημέρες.	254,0	421,0	-
Baseline 03 – Mean by hour + weekend	Όπως το 02, με διάκριση εργάσιμων/Σαββατοκύριακου (is_weekend).	253,0	419,0	-
Baseline 04 – Per-home mean by hour	Μέση κατανάλωση ανά ώρα, ξεχωριστά για κάθε κατοικία.	198,6	363,4	76

Πίνακας 1 Απόδοση βασικών μοντέλων αναφοράς στο σύνολο ελέγχου του IDEAL.

Το μοντέλο **Baseline 01** εκμεταλλεύεται άμεσα τη χρονική αυτοσυσχέτιση της κατανάλωσης, προβλέποντας ότι η κατανάλωση την ώρα t θα είναι ίδια με την κατανάλωση την ίδια ώρα της προηγούμενης ημέρας. Παρότι αποτελεί μια λογική γραμμή βάσης, τα σφάλματα $MAE \approx 224$ Wh και $RMSE \approx 462$ Wh δείχνουν ότι η προσέγγιση αυτή αδυνατεί να αποδώσει καλά σε ημέρες με διαφορετικές συνθήκες (π.χ. διαφορετικές καιρικές συνθήκες ή αλλαγές στη συμπεριφορά των νοικοκυριών), ιδίως στις αιχμές.

Τα **Baseline 02 και 03** βασίζονται σε μέσους όρους: το δεύτερο μοντέλο χρησιμοποιεί τη μέση κατανάλωση ανά ώρα της ημέρας σε όλο το δείγμα, ενώ το τρίτο προσθέτει και διάκριση μεταξύ εργάσιμων ημερών και Σαββατοκύριακου. Τα αποτελέσματα ($MAE \approx 254$ Wh, $RMSE \approx 421$ Wh για το 02 και $MAE \approx 253$ Wh, $RMSE \approx 419$ Wh για το 03) δείχνουν ότι οι καθολικοί μέσοι όροι (global averages) δεν επαρκούν για ακριβή πρόβλεψη και ότι η απλή διάκριση μεταξύ καθημερινών και Σαββατοκύριακου βελτιώνει ελάχιστα την εικόνα. Αυτό υποδηλώνει ότι η ετερογένεια μεταξύ κατοικιών και ημερών είναι σημαντική και δεν μπορεί να συλληφθεί μόνο με μια γενική καμπύλη μέσου ωριαίου φορτίου.

Το **Baseline 04** υπολογίζει ξεχωριστή μέση κατανάλωση ανά ώρα για κάθε κατοικία, αναγνωρίζοντας ότι κάθε νοικοκυριό έχει σχετικά σταθερό, χαρακτηριστικό ημερήσιο προφίλ. Το μοντέλο αυτό επιτυγχάνει σαφώς χαμηλότερα σφάλματα ($MAE \approx 198,6$ Wh και $RMSE \approx 363,4$ Wh), υπερτερώντας όλων των καθολικών βασικών μοντέλων αναφοράς και επιβεβαιώνοντας ότι η εξατομίκευση σε επίπεδο κατοικίας είναι κρίσιμη για τη βελτίωση της ακρίβειας. Τα στατιστικά αυτά υποδεικνύουν ότι τα μοντέλα μηχανικής μάθησης που εξετάζονται στα επόμενα τμήματα θα πρέπει, τουλάχιστον, να ξεπερνούν την απόδοση του Baseline 04, προκειμένου να θεωρηθούν ουσιαστική πρόοδος σε σχέση με τις απλές στατιστικές προσεγγίσεις.

Τα αποτελέσματα των βασικών μοντέλων αναφοράς δείχνουν ότι η απλή αξιοποίηση ιστορικών μέσων τιμών ή γενικών προτύπων κατανάλωσης δεν επαρκεί για την ακριβή πρόβλεψη της οικιακής ηλεκτρικής κατανάλωσης. Παράλληλα, η επίδοση του Baseline 04

θέτει ένα απαιτητικό σημείο αναφοράς για τα μοντέλα μηχανικής μάθησης που παρουσιάζονται στις επόμενες ενότητες. Τα μοντέλα αυτά θα πρέπει να επιτυγχάνουν σημαντικά χαμηλότερα σφάλματα, ώστε να δικαιολογείται η αυξημένη πολυπλοκότητα και ο πρόσθετος υπολογιστικός φόρτος που συνεπάγεται η χρήση τους.

6.3 Απόδοση των μοντέλων στο σενάριο χωρίς ιστορικά δεδομένα

Στο σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης, τα μοντέλα εκπαιδεύονται και αξιολογούνται χωρίς τη χρήση προηγούμενων μετρήσεων κατανάλωσης της κατοικίας. Αντί των ιστορικών χαρακτηριστικών, αξιοποιούνται χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά της κατοικίας, καθώς και βοηθητικά χαρακτηριστικά που περιγράφουν γενικά πρότυπα κατανάλωσης και ομοιότητες μεταξύ κατοικιών. Ο Πίνακας 2 παρουσιάζει τα αποτελέσματα των τριών μοντέλων μηχανικής μάθησης στο ανεξάρτητο σύνολο ελέγχου του Ιουνίου 2018.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	Mean Daily Error (kWh)
Random Forest	147,57	319,40	48,28	1,95
XGBoost	153,03	320,51	50,30	1,93
LightGBM	152,85	320,96	50,80	1,96

Πίνακας 2 Απόδοση των μοντέλων μηχανικής μάθησης στο σενάριο χωρίς ιστορικά δεδομένα

Παρατηρείται ότι το Random Forest παρουσιάζει τη συνολικά καλύτερη απόδοση στο σενάριο χωρίς ιστορικά δεδομένα, επιτυγχάνοντας το χαμηλότερο Μέσο Απόλυτο Σφάλμα ($MAE \approx 147,6$ Wh), τη χαμηλότερη τιμή RMSE ($\approx 319,4$ Wh) και το χαμηλότερο ποσοστιαίο σφάλμα ($MAPE \approx 48,3\%$). Τα αποτελέσματα αυτά δείχνουν ότι η μέθοδος γενικεύει αποτελεσματικά όταν δεν υπάρχει διαθέσιμη πληροφορία προηγούμενης κατανάλωσης.

Τα μοντέλα XGBoost και LightGBM εμφανίζουν παρόμοια συμπεριφορά, με ελαφρώς υψηλότερα σφάλματα σε όλες τις βασικές μετρικές. Αν και οι διαφορές δεν είναι ιδιαίτερα μεγάλες, δεν παρατηρείται κάποιο σαφές πλεονέκτημα των μεθόδων ενισχυτικής μάθησης έναντι του Random Forest στο συγκεκριμένο σενάριο. Το γεγονός αυτό δείχνει ότι, όταν η διαθέσιμη πληροφορία περιορίζεται σε στατικά, χρονικά και περιβαλλοντικά χαρακτηριστικά, η χρήση πιο σύνθετων δένδρων μοντέλων δεν οδηγεί απαραίτητα σε καλύτερη ακρίβεια πρόβλεψης.

Σε σύγκριση με τα βασικά μοντέλα αναφοράς, τα δένδρων μοντέλα στο σενάριο χωρίς ιστορικά δεδομένα επιτυγχάνουν αισθητή μείωση σφάλματος, τόσο σε όρους MAE όσο και RMSE, γεγονός που υποδηλώνει ότι η ενσωμάτωση κλιματικών και στατικών χαρακτηριστικών παρέχει σημαντική πρόσθετη πληροφορία πέρα από την απλή χρονική αυτοσυσχέτιση. Ωστόσο, τα σχετικά υψηλά ποσοστιαία σφάλματα καταδεικνύουν ότι για υψηλότερη ακρίβεια απαιτείται η αξιοποίηση της πρόσφατης ιστορίας κατανάλωσης, όπως προτείνεται και στη σχετική βιβλιογραφία για τη βραχυπρόθεσμη πρόβλεψη φορτίου (Zhang, Wei, Li, Tan, & Zhou, 2018; Ahmad & Chen, 2020; Wang & Li, 2025).

6.4 Απόδοση των μοντέλων στο σενάριο με χρήση ιστορικών δεδομένων

Στο σενάριο με χρήση ιστορικών δεδομένων, τα μοντέλα αξιοποιούν επιπλέον πληροφορία από πρόσφατες μετρήσεις κατανάλωσης της ίδιας κατοικίας. Συγκεκριμένα, χρησιμοποιούνται χαρακτηριστικά χρονικής εξάρτησης, όπως χρονικές υστερήσεις (lag features) και στατιστικά χαρακτηριστικά κινούμενων παραθύρων (rolling features), τα οποία προκύπτουν από ιστορικό παράθυρο διάρκειας 168 ωρών. Με τον τρόπο αυτό, τα μοντέλα αποκτούν τη δυνατότητα να αξιοποιούν τη χρονική συνέχεια της κατανάλωσης και να αναγνωρίζουν επαναλαμβανόμενα ημερήσια και εβδομαδιαία πρότυπα συμπεριφοράς. Ο Πίνακας 3 παρουσιάζει τα αποτελέσματα των τριών μοντέλων μηχανικής μάθησης στο ανεξάρτητο σύνολο ελέγχου του Ιουνίου 2018.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	Mean Daily Error (kWh)
Random Forest	128,09	305,61	35,38	1,26
XGBoost	125,68	299,25	34,93	1,27
LightGBM	125,42	299,40	34,58	1,27

Πίνακας 3 Απόδοση των μοντέλων μηχανικής μάθησης στο σενάριο με χρήση ιστορικών δεδομένων

Σε σύγκριση με το σενάριο χωρίς ιστορικά δεδομένα, παρατηρείται σημαντική βελτίωση για όλα τα μοντέλα. Το Μέσο Απόλυτο Σφάλμα μειώνεται κατά περίπου 20 έως 28 Wh ανά ώρα, ενώ το ποσοστιαίο σφάλμα μειώνεται από επίπεδα κοντά στο 50% σε επίπεδα περίπου 35%. Παρόμοια βελτίωση παρατηρείται και στο RMSE, γεγονός που υποδηλώνει καλύτερη αποτύπωση τόσο της μέσης συμπεριφοράς όσο και των περιόδων αυξημένης κατανάλωσης. Τα αποτελέσματα αυτά επιβεβαιώνουν ότι η πρόσφατη ιστορία κατανάλωσης αποτελεί έναν από τους σημαντικότερους προγνωστικούς παράγοντες στη βραχυπρόθεσμη πρόβλεψη φορτίου. Χαρακτηριστικά όπως η κατανάλωση της προηγούμενης ώρας, της προηγούμενης ημέρας και οι κυλίσμενοι μέσοι όροι επιτρέπουν στα μοντέλα να εκμεταλλεύονται τη χρονική αυτοσυσχέτιση της κατανάλωσης και να προσαρμόζονται καλύτερα στις ιδιαιτερότητες κάθε κατοικίας.

Όσον αφορά τη σύγκριση μεταξύ των αλγορίθμων, οι διαφορές παραμένουν σχετικά μικρές. Το LightGBM επιτυγχάνει το χαμηλότερο MAE και το χαμηλότερο MAPE, ενώ το XGBoost παρουσιάζει οριακά χαμηλότερο RMSE. Το Random Forest εμφανίζει ελαφρώς υψηλότερα ωριαία σφάλματα, αλλά παρουσιάζει το μικρότερο ημερήσιο σφάλμα κατανάλωσης. Συνολικά και τα τρία μοντέλα εμφανίζουν παρόμοιο επίπεδο απόδοσης, γεγονός που υποδηλώνει ότι η επιλογή του αλγορίθμου επηρεάζει λιγότερο το τελικό αποτέλεσμα από ό,τι η ύπαρξη ιστορικών δεδομένων.

Ένα ιδιαίτερα σημαντικό συμπέρασμα που προκύπτει από τη σύγκριση των Ενοτήτων 6.3 και 6.4 είναι ότι η διαθεσιμότητα ιστορικών δεδομένων έχει μεγαλύτερη επίδραση στην ακρίβεια των προβλέψεων από ό,τι η επιλογή μεταξύ διαφορετικών αλγορίθμων μηχανικής μάθησης. Η μετάβαση από το σενάριο χωρίς ιστορικά δεδομένα στο σενάριο με χρήση ιστορικών δεδομένων οδηγεί σε σημαντική μείωση των σφαλμάτων για όλα τα μοντέλα, ενώ οι διαφορές μεταξύ Random Forest, XGBoost και LightGBM παραμένουν συγκριτικά μικρές. Το εύρημα αυτό συμφωνεί με τη διεθνή βιβλιογραφία της βραχυπρόθεσμης πρόβλεψης φορτίου, σύμφωνα με την οποία η ποιότητα και η πληρότητα των διαθέσιμων

χαρακτηριστικών εισόδου επηρεάζουν συχνά περισσότερο την τελική απόδοση από την ίδια την επιλογή του αλγορίθμου (Hong & Fan, 2016; Ahmad et al., 2018; Olu Ajayi et al., 2023).

6.5 Συνολική σύγκριση και αξιολόγηση

Ο Πίνακας 4 συνοψίζει τις επιδόσεις του ισχυρότερου βασικού μοντέλου αναφοράς (Baseline 04) και των τριών μοντέλων μηχανικής μάθησης στα δύο σενάρια πρόβλεψης που εξετάστηκαν: χωρίς ιστορικά δεδομένα κατανάλωσης και με χρήση ιστορικών δεδομένων.

Μοντέλο / Σενάριο	MAE (Wh)	RMSE (Wh)	MAPE (%)
Baseline 04 – Per home mean by hour	198,6	363,4	76,0
Random Forest (χωρίς ιστορικά δεδομένα)	147,57	319,40	48,28
XGBoost (χωρίς ιστορικά δεδομένα)	153,03	320,51	50,30
LightGBM (χωρίς ιστορικά δεδομένα)	152,85	320,96	50,80
Random Forest (με ιστορικά δεδομένα)	128,09	305,61	35,38
XGBoost (με ιστορικά δεδομένα)	125,68	299,25	34,93
LightGBM (με ιστορικά δεδομένα)	125,42	299,40	34,58

Πίνακας 4 Συγκεντρωτική σύγκριση βασικού μοντέλου αναφοράς και μοντέλων μηχανικής μάθησης

Τα αποτελέσματα δείχνουν ότι όλα τα μοντέλα μηχανικής μάθησης υπερτερούν σημαντικά του καλύτερου στατιστικού μοντέλου αναφοράς. Ακόμη και στο σενάριο χωρίς ιστορικά δεδομένα, τα δένδροειδή μοντέλα επιτυγχάνουν μείωση του MAE κατά περίπου 45–50 Wh ανά ώρα και μείωση του MAPE κατά περίπου 25 ποσοστιαίες μονάδες σε σχέση με το Baseline 04. Το εύρημα αυτό επιβεβαιώνει ότι ο συνδυασμός χρονικών, περιβαλλοντικών και στατικών χαρακτηριστικών παρέχει σημαντική πρόσθετη πληροφορία πέρα από τις απλές στατιστικές προσεγγίσεις.

Ακόμη μεγαλύτερη βελτίωση παρατηρείται όταν τα μοντέλα εμπλουτίζονται με ιστορικά χαρακτηριστικά κατανάλωσης. Η μετάβαση από το σενάριο χωρίς ιστορικά δεδομένα στο σενάριο με χρήση ιστορικών δεδομένων οδηγεί σε αισθητή μείωση όλων των μετρικών σφάλματος. Ενδεικτικά, το MAE μειώνεται κατά περίπου 20–28 Wh ανά ώρα, ενώ το MAPE μειώνεται από επίπεδα κοντά στο 50% σε επίπεδα περίπου 35%. Η βελτίωση αυτή καταδεικνύει την υψηλή προγνωστική αξία της πρόσφατης ιστορίας κατανάλωσης και επιβεβαιώνει ότι η χρονική αυτοσυσχέτιση αποτελεί βασικό χαρακτηριστικό της οικιακής ηλεκτρικής κατανάλωσης.

Ιδιαίτερο ενδιαφέρον παρουσιάζει το γεγονός ότι οι διαφορές μεταξύ των τριών αλγορίθμων είναι σχετικά μικρές σε σύγκριση με τη διαφορά μεταξύ των δύο σεναρίων πρόβλεψης. Για παράδειγμα, η διαφορά MAE μεταξύ του καλύτερου και του χειρότερου μοντέλου στο σενάριο με ιστορικά δεδομένα είναι μικρότερη από 3 Wh ανά ώρα, ενώ η διαφορά μεταξύ

των σεναρίων χωρίς και με ιστορικά δεδομένα υπερβαίνει τις 20 Wh ανά ώρα για όλα τα μοντέλα. Το αποτέλεσμα αυτό υποδηλώνει ότι η ποιότητα και η πληρότητα των διαθέσιμων χαρακτηριστικών εισόδου επηρεάζουν περισσότερο την τελική ακρίβεια από ό,τι η επιλογή μεταξύ διαφορετικών συνδυαστικών δένδρων αλγορίθμων.

Τα αποτελέσματα δείχνουν ότι τα μοντέλα μηχανικής μάθησης αποτελούν σαφή βελτίωση σε σχέση με τις απλές στατιστικές προσεγγίσεις, ενώ η αξιοποίηση ιστορικών δεδομένων κατανάλωσης αποτελεί τον σημαντικότερο παράγοντα αύξησης της προγνωστικής ακρίβειας. Παράλληλα, τα μοντέλα XGBoost και LightGBM παρουσιάζουν οριακά καλύτερη συνολική απόδοση στο σενάριο με ιστορικά δεδομένα, ενώ το Random Forest εμφανίζει ιδιαίτερα ανταγωνιστική συμπεριφορά παρά τη μικρότερη πολυπλοκότητά του.

6.6 Πρόσθετα πειράματα βελτίωσης των προβλέψεων

Πέρα από τη βασική συγκριτική αξιολόγηση των μοντέλων Random Forest, XGBoost και LightGBM, πραγματοποιήθηκαν επιπλέον πειράματα με στόχο τη διερεύνηση πιθανών τρόπων περαιτέρω βελτίωσης της ακρίβειας των προβλέψεων. Τα πειράματα αυτά επικεντρώθηκαν κυρίως στο σενάριο με χρήση ιστορικών δεδομένων, καθώς αυτό παρουσίασε τις καλύτερες επιδόσεις κατά την αρχική αξιολόγηση.

Οι πρόσθετες προσεγγίσεις που εξετάστηκαν περιλάμβαναν τεχνικές μεταγενέστερης διόρθωσης των προβλέψεων (post-processing), μοντέλα διόρθωσης υπολειμμάτων (residual correction), τεχνικές βελτίωσης της απόδοσης κατά τις περιόδους αιχμής κατανάλωσης, προσεγγίσεις ομαδοποίησης κατοικιών ανά επίπεδο κατανάλωσης, καθώς και εναλλακτικές αρχιτεκτονικές βαθιάς μάθησης για χρονοσειρές. Παρόμοιες κατευθύνσεις έχουν προταθεί στη βιβλιογραφία της βραχυπρόθεσμης πρόβλεψης φορτίου με στόχο τη βελτίωση της ακρίβειας των προβλέψεων και την καλύτερη προσαρμογή σε διαφορετικά πρότυπα ενεργειακής συμπεριφοράς (Hong & Fan, 2016; Ahmad et al., 2018).

6.6.1 Προσαρμογή της ημερήσιας καμπύλης φορτίου με βάση το πρόσφατο ιστορικό κατανάλωσης

Μετά την ανάπτυξη των βασικών μοντέλων με χρήση ιστορικών δεδομένων εξετάστηκε μία επιπλέον τεχνική μεταγενέστερης διόρθωσης των προβλέψεων, η οποία αξιοποιεί πρόσφατο ιστορικό κατανάλωσης της ίδιας κατοικίας. Η βασική ιδέα της προσέγγισης είναι ότι η πρόσφατη πραγματική συμπεριφορά μιας κατοικίας μπορεί να περιέχει χρήσιμη πληροφορία για το αναμενόμενο ημερήσιο προφίλ κατανάλωσης, συμπληρώνοντας την πρόβλεψη που παράγεται από το μοντέλο μηχανικής μάθησης.

Για τον σκοπό αυτό υπολογίζεται αρχικά ένα ιστορικό προφίλ αναφοράς από τις τελευταίες επτά ημέρες διαθέσιμων μετρήσεων. Αντί της χρήσης μέσου όρου, χρησιμοποιείται η διάμεσος ωριαία κατανάλωση για κάθε ώρα της ημέρας, ώστε να περιορίζεται η επίδραση ακραίων τιμών ή ασυνήθιστων ημερών. Στη συνέχεια το ιστορικό προφίλ προσαρμόζεται στο συνολικό ενεργειακό επίπεδο που προβλέπει το μοντέλο για την επόμενη ημέρα.

Η τελική πρόβλεψη προκύπτει ως συνδυασμός της αρχικής πρόβλεψης του μοντέλου και του προσαρμοσμένου ιστορικού προφίλ:

$$\hat{y}_t^{final} = (1 - \alpha_t)\hat{y}_t + \alpha_t H_t \quad (7)$$

Όπου $\hat{y}_t$ είναι η αρχική πρόβλεψη του μοντέλου για την ώρα t , το H_t είναι το ιστορικό ωριαίο προφίλ κατανάλωσης που προκύπτει από τη διάμεσο των τελευταίων επτά ημερών και α_t είναι ο συντελεστής διόρθωσης για την αντίστοιχη ώρα.

Ο συντελεστής α_t δεν είναι σταθερός αλλά προσαρμόζεται δυναμικά ανάλογα με τη σταθερότητα του πρόσφατου ιστορικού κατανάλωσης. Με τον τρόπο αυτό μεγαλύτερη εμπιστοσύνη δίνεται στο ιστορικό προφίλ όταν η κατανάλωση εμφανίζει επαναλαμβανόμενα και σταθερά πρότυπα, ενώ σε περιπτώσεις υψηλής μεταβλητότητας η τελική πρόβλεψη βασίζεται κυρίως στο μοντέλο μηχανικής μάθησης. Για λόγους σταθερότητας εφαρμόζεται ανώτατο όριο στη συνεισφορά του ιστορικού προφίλ, ώστε η διόρθωση να λειτουργεί συμπληρωματικά και να μην αντικαθιστά πλήρως την πρόβλεψη του μοντέλου.

Η προσέγγιση αυτή αξιολογήθηκε σε ξεχωριστά πειράματα και παρουσίασε βελτιώσεις σε συγκεκριμένες κατοικίες και ημέρες με σχετικά σταθερή συμπεριφορά κατανάλωσης. Ωστόσο, τα αποτελέσματα δεν ήταν ομοιόμορφα για όλες τις κατοικίες και δεν προέκυψε σταθερή βελτίωση όλων των μετρικών αξιολόγησης στο συνολικό σύνολο δοκιμών. Για τον λόγο αυτό η τεχνική δεν χρησιμοποιήθηκε ως κύρια μορφή πρόβλεψης στα αποτελέσματα του παρόντος κεφαλαίου, αλλά ενσωματώθηκε ως προαιρετική λειτουργία στο τελικό σύστημα πρόβλεψης.

6.6.2 Τεχνικές μεταγενέστερης διόρθωσης των προβλέψεων

Αρχικά εξετάστηκαν διάφορες τεχνικές μεταγενέστερης διόρθωσης των προβλέψεων, οι οποίες εφαρμόζονταν μετά την παραγωγή της αρχικής πρόβλεψης από τα βασικά μοντέλα. Στόχος των τεχνικών αυτών ήταν η μείωση της μεροληψίας των προβλέψεων, η βελτίωση της εκτίμησης των ωρών αιχμής και η ακριβέστερη αποτύπωση της συνολικής ημερήσιας κατανάλωσης.

Παράλληλα δοκιμάστηκαν μοντέλα διόρθωσης υπολειμμάτων (residual correction), στα οποία ένα δεύτερο μοντέλο εκπαιδευόταν για να προβλέπει και να διορθώνει το σφάλμα του αρχικού μοντέλου. Η βασική ιδέα των προσεγγίσεων αυτών είναι ότι ένα δευτερογενές μοντέλο μπορεί να μάθει συστηματικά πρότυπα σφάλματος τα οποία δεν έχουν αποτυπωθεί από το αρχικό μοντέλο και να τα διορθώσει εκ των υστέρων. Παρόμοιες στρατηγικές έχουν εφαρμοστεί σε προβλήματα πρόβλεψης ενεργειακής κατανάλωσης και ηλεκτρικού φορτίου με στόχο τη βελτίωση της συνολικής ακρίβειας των προβλέψεων (Ahmad et al., 2018; Wang & Li, 2025).

Επιπλέον εξετάστηκαν διαφορετικοί συνδυασμοί βασικών και δευτερογενών μοντέλων, καθώς και διαφορετικές στρατηγικές έμφασης στις ώρες αιχμής κατανάλωσης. Αν και σε ορισμένες περιπτώσεις παρατηρήθηκαν βελτιώσεις συγκεκριμένων μετρικών, δεν προέκυψε κάποια προσέγγιση που να βελτιώνει σταθερά όλες τις μετρικές αξιολόγησης ταυτόχρονα.

Ιδιαίτερο ενδιαφέρον παρουσίασαν οι τεχνικές ημερήσιας βαθμονόμησης (daily calibration), οι οποίες στόχευαν στη βελτίωση της εκτίμησης της συνολικής ημερήσιας κατανάλωσης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 5.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	Daily Absolute Error (kWh)
LightGBM (balanced)	125,42	299,40	34,58	1,2737
XGBoost (balanced)	125,68	299,25	34,93	1,2680
Random Forest (balanced)	128,09	305,61	35,38	1,2647
LightGBM (daily-calibrated)	130,84	295,66	39,66	1,0938
XGBoost (daily-calibrated)	131,27	295,58	40,11	1,0922
Random Forest (daily-calibrated)	133,63	302,18	40,53	1,0699

Πίνακας 5 Σύγκριση balanced και daily-calibrated εκδόσεων των μοντέλων

Παρατηρείται ότι οι τεχνικές ημερήσιας βαθμονόμησης μειώνουν σημαντικά το ημερήσιο σφάλμα κατανάλωσης, με το Random Forest να επιτυγχάνει τη χαμηλότερη τιμή (1,07 kWh). Ωστόσο, η βελτίωση αυτή συνοδεύεται από επιδείνωση των ωριαίων μετρικών σφάλματος (MAE και MAPE). Το εύρημα αυτό αναδεικνύει έναν σαφή συμβιβασμό μεταξύ της ακριβούς εκτίμησης της συνολικής ημερήσιας ενέργειας και της ακριβούς πρόβλεψης της ωριαίας καμπύλης φορτίου.

Δεδομένου ότι βασικός στόχος της παρούσας εργασίας είναι η όσο το δυνατόν ακριβέστερη πρόβλεψη της ωριαίας κατανάλωσης ηλεκτρικής ενέργειας και όχι αποκλειστικά της συνολικής ημερήσιας κατανάλωσης, προτιμήθηκαν οι εκδόσεις των μοντέλων τύπου *balanced*, οι οποίες επιτυγχάνουν καλύτερες ωριαίες μετρικές σφάλματος. Αν και οι εκδόσεις *daily-calibrated* παρουσιάζουν μικρότερο ημερήσιο σφάλμα, η επιδείνωση της ακρίβειας σε ωριαίο επίπεδο θεωρήθηκε σημαντικότερη για τους σκοπούς της παρούσας μελέτης.

Συνεπώς, οι τεχνικές ημερήσιας βαθμονόμησης κρίθηκαν χρήσιμες ως προαιρετικές επεκτάσεις του συστήματος, ιδιαίτερα σε εφαρμογές όπου ενδιαφέρει πρωτίστως η εκτίμηση της συνολικής ημερήσιας κατανάλωσης. Ωστόσο, δεν επιλέχθηκαν ως βασική μορφή λειτουργίας των μοντέλων που χρησιμοποιούνται στην κύρια αξιολόγηση και στο τελικό σύστημα πρόβλεψης.

6.6.3 Ομαδοποίηση κατοικιών και εναλλακτικές προσεγγίσεις

Μία δεύτερη κατηγορία πειραμάτων αφορούσε την ομαδοποίηση των κατοικιών σε κατηγορίες κατανάλωσης (*consumption regimes*). Η βασική ιδέα ήταν ότι κατοικίες με χαμηλή, μεσαία και υψηλή κατανάλωση ενδέχεται να παρουσιάζουν διαφορετικά πρότυπα

συμπεριφοράς και, επομένως, να ωφελούνται από εξειδικευμένα μοντέλα ή διαφορετικές στρατηγικές διόρθωσης.

Η προσέγγιση αυτή βασίζεται στην παρατήρηση ότι διαφορετικές κατηγορίες καταναλωτών εμφανίζουν συχνά διαφορετικά ενεργειακά προφίλ και διαφορετικό βαθμό μεταβλητότητας της κατανάλωσης. Για τον λόγο αυτό έχουν προταθεί στη βιβλιογραφία εξειδικευμένα μοντέλα ή ξεχωριστές στρατηγικές πρόβλεψης ανά κατηγορία καταναλωτών (Hong & Fan, 2016; Olu Ajayi et al., 2023).

Για τον σκοπό αυτό εξετάστηκαν διαφορετικές μορφές κατηγοριοποίησης και αναπτύχθηκαν ξεχωριστές στρατηγικές πρόβλεψης για κάθε ομάδα κατοικιών. Παρότι σε ορισμένες περιπτώσεις παρατηρήθηκαν βελτιώσεις συγκεκριμένων μετρικών, τα οφέλη δεν ήταν σταθερά για όλες τις κατηγορίες και όλα τα μοντέλα. Επιπλέον, η ανάγκη διατήρησης πολλαπλών μοντέλων αύξανε σημαντικά την πολυπλοκότητα της υλοποίησης και της συντήρησης του συστήματος.

Παράλληλα διερευνήθηκε η δυνατότητα αξιοποίησης σύγχρονων αρχιτεκτονικών βαθιάς μάθησης, όπως το Temporal Fusion Transformer (TFT) (Lim et al., 2021) και το N-HiTS (Challu et al., 2023). Η διερεύνηση αυτή πραγματοποιήθηκε σε προκαταρκτικό επίπεδο και δεν αποτέλεσε τον κύριο άξονα της παρούσας εργασίας, η οποία επικεντρώθηκε κυρίως στις δένδροειδείς μεθόδους για πινακοποιημένα δεδομένα. Οι αρχιτεκτονικές αυτές απαιτούν σημαντικά μεγαλύτερους υπολογιστικούς πόρους, μεγαλύτερο χρόνο εκπαίδευσης και εκτεταμένη διαδικασία ρύθμισης υπερπαραμέτρων σε σχέση με τα δένδροειδή μοντέλα.

Τα πρόσθετα πειράματα επιβεβαίωσαν ότι είναι δυνατόν να βελτιωθούν επιμέρους πτυχές της απόδοσης των μοντέλων, όπως η ακρίβεια της συνολικής ημερήσιας κατανάλωσης ή η απόδοση σε συγκεκριμένες κατηγορίες κατοικιών. Ωστόσο, δεν προέκυψε κάποια εναλλακτική προσέγγιση που να υπερτερεί σταθερά των επιλεγμένων δένδροειδών μοντέλων σε όλες τις μετρικές αξιολόγησης. Για τον λόγο αυτό, η τελική υλοποίηση του συστήματος πρόβλεψης βασίστηκε στα μοντέλα Random Forest, XGBoost και LightGBM που παρουσιάστηκαν στις προηγούμενες ενότητες, τα οποία προσφέρουν την καλύτερη ισορροπία μεταξύ ακρίβειας, υπολογιστικού κόστους, ερμηνευσιμότητας και ευκολίας ενσωμάτωσης σε πραγματικό περιβάλλον λειτουργίας.

Οι πρόσθετες τεχνικές που παρουσιάστηκαν στην ενότητα αυτή είχαν κυρίως διερευνητικό χαρακτήρα και χρησιμοποιήθηκαν με στόχο την κατανόηση των πιθανών ορίων βελτίωσης των βασικών μοντέλων. Για τον λόγο αυτό, δεν ενσωματώθηκαν όλες σε πλήρη συγκριτική αξιολόγηση αντίστοιχη με τα βασικά μοντέλα Random Forest, XGBoost και LightGBM. Η επιλογή αυτή έγινε ώστε η τελική αξιολόγηση να παραμείνει επικεντρωμένη στις προσεγγίσεις που μπορούν να ενσωματωθούν με συνέπεια και σταθερότητα στο τελικό σύστημα πρόβλεψης. Τα αποτελέσματα των διερευνητικών πειραμάτων αξιοποιήθηκαν κυρίως για την ερμηνεία των περιορισμών των προβλέψεων και για τη διαμόρφωση των προτάσεων μελλοντικής επέκτασης.

7. Ανάπτυξη και Υλοποίηση του Συστήματος Πρόβλεψης

Μετά την εκπαίδευση και αξιολόγηση των μοντέλων μηχανικής μάθησης, αναπτύχθηκε ένα ολοκληρωμένο σύστημα πρόβλεψης ηλεκτρικής κατανάλωσης, το οποίο επιτρέπει τη χρήση των μοντέλων σε πραγματικές συνθήκες. Στόχος του συστήματος είναι η παραγωγή προβλέψεων επόμενης ημέρας μέσω μιας εύχρηστης διεπαφής χρήστη, χωρίς να απαιτούνται εξειδικευμένες γνώσεις μηχανικής μάθησης από τον τελικό χρήστη.

Η υλοποίηση βασίζεται σε αρχιτεκτονική τριών επιπέδων (three-tier architecture), η οποία διαχωρίζει τη διαχείριση των εκπαιδευμένων μοντέλων, την επιχειρησιακή λογική των προβλέψεων και τη διεπαφή αλληλεπίδρασης με τον χρήστη. Συγκεκριμένα, το σύστημα αποτελείται από: (α) το επίπεδο των μοντέλων μηχανικής μάθησης, όπου αποθηκεύονται και διαχειρίζονται τα εκπαιδευμένα μοντέλα Random Forest, XGBoost και LightGBM, (β) ένα επίπεδο υπηρεσιών βασισμένο σε FastAPI, το οποίο αναλαμβάνει την προετοιμασία των δεδομένων εισόδου και την παραγωγή προβλέψεων και (γ) μια διαδικτυακή διεπαφή χρήστη υλοποιημένη με Gradio, μέσω της οποίας ο χρήστης μπορεί να εισάγει τα απαιτούμενα δεδομένα και να λαμβάνει τα αποτελέσματα των προβλέψεων.

Επιπλέον, το σύστημα υποστηρίζει τόσο το σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης όσο και το σενάριο με χρήση ιστορικών δεδομένων. Παράλληλα, παρέχει τη δυνατότητα σύγκρισης διαφορετικών μοντέλων, συνδυασμού των προβλέψεών τους και αξιοποίησης πρόσφατων δεδομένων κατανάλωσης, με στόχο τη βελτίωση της ακρίβειας των τελικών προβλέψεων.

Στις επόμενες ενότητες παρουσιάζονται αναλυτικά η αρχιτεκτονική του συστήματος, η οργάνωση των μοντέλων, η υλοποίηση του προγραμματιστικού περιβάλλοντος (API), η διαχείριση ιστορικών δεδομένων, η διεπαφή χρήστη και το περιβάλλον εγκατάστασης και εκτέλεσης της εφαρμογής.

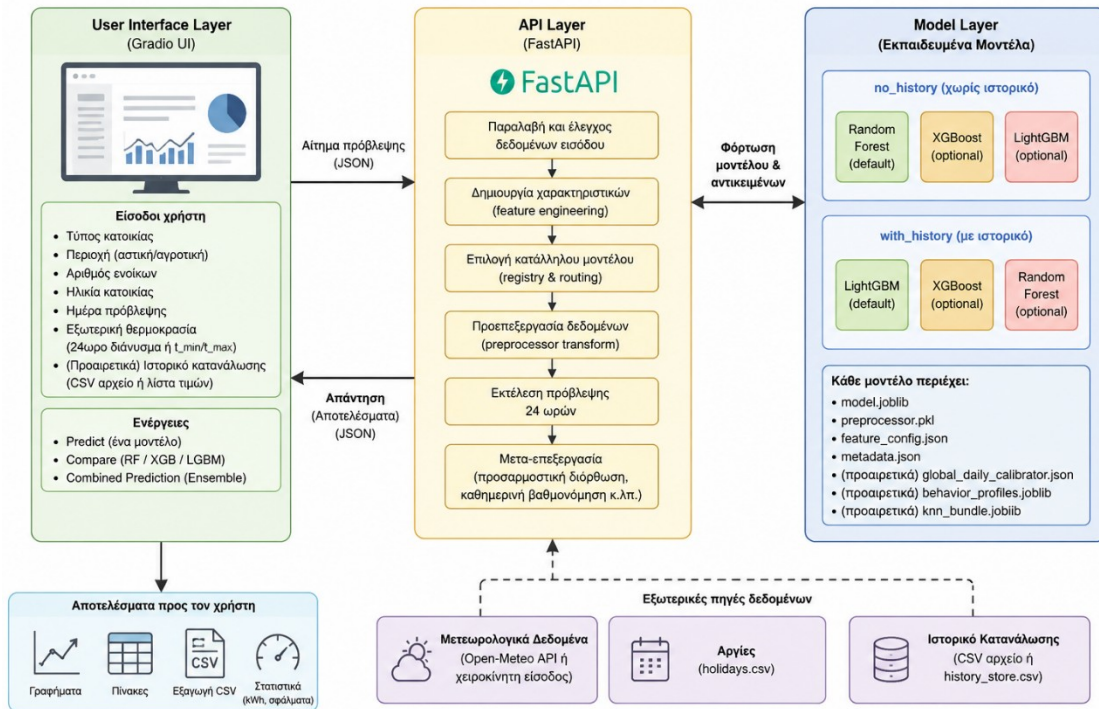
7.1 Αρχιτεκτονική του συστήματος πρόβλεψης

Το σύστημα πρόβλεψης που αναπτύχθηκε, σχεδιάστηκε με στόχο τον σαφή διαχωρισμό μεταξύ των εκπαιδευμένων μοντέλων μηχανικής μάθησης, της επιχειρησιακής λογικής των προβλέψεων και της αλληλεπίδρασης με τον τελικό χρήστη. Για τον λόγο αυτό υιοθετήθηκε αρχιτεκτονική τριών επιπέδων (three-tier architecture), η οποία διευκολύνει την επεκτασιμότητα, τη συντήρηση και τη μελλοντική εξέλιξη του συστήματος.

Η αρχιτεκτονική αποτελείται από τρία βασικά επίπεδα:

- **Επίπεδο μοντέλων (Model Layer)**, στο οποίο αποθηκεύονται τα εκπαιδευμένα μοντέλα Random Forest, XGBoost και LightGBM, καθώς και όλα τα απαραίτητα αντικείμενα προεπεξεργασίας και μετασχηματισμού δεδομένων.
- **Επίπεδο υπηρεσιών (API Layer)**, το οποίο υλοποιείται με το πλαίσιο FastAPI και αναλαμβάνει την παραλαβή των δεδομένων εισόδου, την εφαρμογή των απαραίτητων μετασχηματισμών, την εκτέλεση των προβλέψεων και την επιστροφή των αποτελεσμάτων.
- **Επίπεδο διεπαφής χρήστη (User Interface Layer)**, το οποίο υλοποιείται με τη βιβλιοθήκη Gradio και παρέχει ένα φιλικό περιβάλλον αλληλεπίδρασης, επιτρέποντας στον χρήστη να εισάγει τα χαρακτηριστικά της κατοικίας, τα μετεωρολογικά δεδομένα και, όπου απαιτείται, ιστορικά δεδομένα κατανάλωσης.

Η συνολική ροή λειτουργίας του συστήματος παρουσιάζεται στην Εικόνα 1.



Εικόνα 1 Συνολική αρχιτεκτονική του συστήματος πρόβλεψης

Αρχικά ο χρήστης εισάγει τα απαιτούμενα δεδομένα μέσω της διεπαφής χρήστη. Τα δεδομένα αυτά αποστέλλονται στο επίπεδο υπηρεσιών, όπου πραγματοποιούνται οι απαραίτητοι έλεγχοι εγκυρότητας, η δημιουργία των χαρακτηριστικών εισόδου και η επιλογή του κατάλληλου μοντέλου. Στη συνέχεια το επιλεγμένο μοντέλο παράγει την πρόβλεψη της επόμενης ημέρας, η οποία επιστρέφεται στη διεπαφή χρήστη μαζί με συνοπτικά στατιστικά στοιχεία, γραφήματα και δυνατότητες εξαγωγής των αποτελεσμάτων.

Ιδιαίτερη έμφαση δόθηκε στην υποστήριξη δύο διαφορετικών σεναρίων πρόβλεψης. Το πρώτο αφορά περιπτώσεις στις οποίες δεν είναι διαθέσιμο ιστορικό κατανάλωσης ενώ το δεύτερο αξιοποιεί πρόσφατο ιστορικό δεδομένων κατανάλωσης. Η κοινή αρχιτεκτονική επιτρέπει την υποστήριξη και των δύο σεναρίων χωρίς αλλαγές στη διεπαφή χρήστη, ενώ παράλληλα παρέχει δυνατότητες σύγκρισης διαφορετικών μοντέλων και εφαρμογής προαιρετικών τεχνικών βελτίωσης των προβλέψεων.

Το συγκεκριμένο αρχιτεκτονικό σχήμα επιλέχθηκε επειδή επιτρέπει την ανεξάρτητη ανάπτυξη και συντήρηση κάθε επιπέδου του συστήματος. Επιπλέον, καθιστά δυνατή τη μελλοντική προσθήκη νέων μοντέλων πρόβλεψης ή νέων λειτουργιών στη διεπαφή χρήστη χωρίς να απαιτούνται σημαντικές αλλαγές στον πυρήνα του συστήματος.

7.2 Επίπεδο μοντέλων (Model Layer)

Το επίπεδο μοντέλων αποτελεί τον πυρήνα του συστήματος πρόβλεψης και είναι υπεύθυνο για την αποθήκευση, διαχείριση και αξιοποίηση των εκπαιδευμένων μοντέλων μηχανικής μάθησης. Στο επίπεδο αυτό φιλοξενούνται τα τελικά μοντέλα Random Forest, XGBoost και LightGBM που παρουσιάστηκαν και αξιολογήθηκαν στα προηγούμενα κεφάλαια.

Η οργάνωση των μοντέλων ακολουθεί τη διάκριση μεταξύ των δύο βασικών σεναρίων πρόβλεψης. Για κάθε σενάριο διατηρούνται πολλαπλά εκπαιδευμένα μοντέλα, επιτρέποντας στο σύστημα είτε να χρησιμοποιήσει ένα προκαθορισμένο μοντέλο είτε να πραγματοποιήσει συγκρίσεις μεταξύ διαφορετικών αλγορίθμων. Στην τελική υλοποίηση επιλέχθηκαν ως προεπιλεγμένα μοντέλα το Random Forest για το σενάριο χωρίς χρήση ιστορικών δεδομένων και το LightGBM για το σενάριο με χρήση ιστορικών δεδομένων, καθώς παρουσίασαν την καλύτερη συνολική ισορροπία μεταξύ ακρίβειας πρόβλεψης και δυνατότητας γενίκευσης κατά την αξιολόγηση του Κεφαλαίου 6.

Κάθε εκπαιδευμένο μοντέλο αποθηκεύεται μαζί με όλα τα απαραίτητα αντικείμενα προεπεξεργασίας που απαιτούνται για την παραγωγή προβλέψεων. Συγκεκριμένα, εκτός από το ίδιο το μοντέλο, αποθηκεύονται οι παράμετροι μετασχηματισμού των δεδομένων, οι πληροφορίες των χαρακτηριστικών εισόδου, οι ρυθμίσεις εκπαίδευσης και τα σχετικά μεταδεδομένα. Με τον τρόπο αυτό εξασφαλίζεται ότι κατά τη χρήση του συστήματος εφαρμόζεται ακριβώς η ίδια διαδικασία προεπεξεργασίας που χρησιμοποιήθηκε κατά την εκπαίδευση.

Η επιλογή του κατάλληλου μοντέλου πραγματοποιείται δυναμικά μέσω ειδικού μηχανισμού καταχώρισης μοντέλων (model registry). Ο μηχανισμός αυτός λειτουργεί ως ενδιάμεσο επίπεδο μεταξύ του API και των αποθηκευμένων μοντέλων, παρέχοντας πληροφορίες σχετικά με τα διαθέσιμα μοντέλα, τα υποστηριζόμενα σενάρια πρόβλεψης και τις αντίστοιχες απαιτήσεις εισόδου. Με τον τρόπο αυτό αποφεύγεται η άμεση εξάρτηση του υπόλοιπου συστήματος από συγκεκριμένα αρχεία ή διαδρομές αποθήκευσης.

Επιπλέον, η συγκεκριμένη αρχιτεκτονική επιτρέπει τη μελλοντική ενσωμάτωση νέων μοντέλων χωρίς να απαιτούνται αλλαγές στη διεπαφή χρήστη ή στο επίπεδο υπηρεσιών. Η προσθήκη ενός νέου μοντέλου περιορίζεται στην εκπαίδευση, αποθήκευση και καταχώρισή του στο μητρώο μοντέλων, διατηρώντας αμετάβλητη τη συνολική λειτουργία του συστήματος.

7.3 Επίπεδο υπηρεσιών και προγραμματιστική διεπαφή (API Layer)

Το επίπεδο υπηρεσιών αποτελεί τον συνδετικό κρίκο μεταξύ της διεπαφής χρήστη και των εκπαιδευμένων μοντέλων μηχανικής μάθησης. Υλοποιήθηκε με τη χρήση του πλαισίου FastAPI και είναι υπεύθυνο για την παραλαβή των δεδομένων εισόδου, την προεπεξεργασία τους, την επιλογή του κατάλληλου μοντέλου πρόβλεψης και την επιστροφή των αποτελεσμάτων στον χρήστη.

Η επιλογή του FastAPI βασίστηκε στην υψηλή απόδοσή του, στην ευκολία ανάπτυξης διαδικτυακών υπηρεσιών και στη δυνατότητα αυτόματης τεκμηρίωσης των διαθέσιμων λειτουργιών. Μέσω του API η λειτουργικότητα του συστήματος καθίσταται ανεξάρτητη από τη διεπαφή χρήστη, επιτρέποντας τη μελλοντική αξιοποίηση των μοντέλων και από άλλες εφαρμογές ή υπηρεσίες χωρίς να απαιτούνται αλλαγές στον πυρήνα του συστήματος. Το API παρέχει τέσσερις βασικές υπηρεσίες. Η σημαντικότερη είναι η υπηρεσία πρόβλεψης, μέσω της οποίας ο χρήστης μπορεί να ζητήσει πρόβλεψη της ηλεκτρικής κατανάλωσης για την επόμενη ημέρα. Παράλληλα παρέχονται υπηρεσίες ενημέρωσης σχετικά με τα διαθέσιμα μοντέλα, ελέγχου της κατάστασης λειτουργίας του συστήματος και διαχείρισης πληροφοριών που αφορούν τα ιστορικά δεδομένα κατανάλωσης.

Κατά την υποβολή ενός αιτήματος πρόβλεψης, ο χρήστης επιλέγει αρχικά το επιθυμητό σενάριο λειτουργίας. Το σύστημα υποστηρίζει δύο διαφορετικά σενάρια: το σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης και το σενάριο με χρήση ιστορικών δεδομένων. Με βάση

την επιλογή αυτή, το API δρομολογεί αυτόματα το αίτημα προς το κατάλληλο μοντέλο και εκτελεί τα απαιτούμενα βήματα προεπεξεργασίας.

Στο σενάριο χωρίς ιστορικά δεδομένα, οι προβλέψεις βασίζονται σε χρονικά χαρακτηριστικά, χαρακτηριστικά της κατοικίας και περιβαλλοντικές μεταβλητές, όπως η εξωτερική θερμοκρασία. Στο σενάριο με χρήση ιστορικών δεδομένων απαιτείται επιπλέον πρόσφατο ιστορικό κατανάλωσης διάρκειας τουλάχιστον επτά ημερών (168 ωρών), από το οποίο παράγονται τα χαρακτηριστικά χρονικής εξάρτησης που χρησιμοποιούνται από τα μοντέλα.

Η παραγωγή της 24ωρης πρόβλεψης πραγματοποιείται με αναδρομικό τρόπο. Αρχικά παράγεται η πρόβλεψη της πρώτης ώρας της επόμενης ημέρας. Η τιμή αυτή ενσωματώνεται προσωρινά στο διαθέσιμο ιστορικό και χρησιμοποιείται για τον υπολογισμό των χαρακτηριστικών χρονικής υστέρησης που απαιτούνται για την πρόβλεψη της επόμενης ώρας. Η διαδικασία επαναλαμβάνεται διαδοχικά μέχρι να παραχθεί πλήρης πρόβλεψη για όλες τις ώρες της ημέρας. Με τον τρόπο αυτό είναι δυνατή η αξιοποίηση των χαρακτηριστικών ιστορικού χωρίς να απαιτείται γνώση μελλοντικών πραγματικών μετρήσεων.

Ιδιαίτερη έμφαση δόθηκε και στη διαχείριση των μετεωρολογικών δεδομένων, καθώς η εξωτερική θερμοκρασία αποτελεί σημαντικό παράγοντα που επηρεάζει την ηλεκτρική κατανάλωση των κατοικιών. Το σύστημα μπορεί είτε να χρησιμοποιήσει ωριαίες τιμές θερμοκρασίας που παρέχονται από τον χρήστη είτε να ανακτήσει αυτόματα μετεωρολογικά δεδομένα από εξωτερικές διαδικτυακές υπηρεσίες. Σε περιπτώσεις όπου δεν είναι διαθέσιμο πλήρες ωριαίο προφίλ θερμοκρασίας, το API μπορεί να δημιουργήσει μία προσεγγιστική ωριαία καμπύλη με βάση τις ελάχιστες και μέγιστες ημερήσιες θερμοκρασίες.

Μετά την ολοκλήρωση της πρόβλεψης, το API επιστρέφει τα αποτελέσματα σε μορφή JSON. Η απόκριση περιλαμβάνει την προβλεπόμενη κατανάλωση για κάθε μία από τις επόμενες 24 ώρες, τη συνολική ημερήσια κατανάλωση, πληροφορίες σχετικά με το μοντέλο που χρησιμοποιήθηκε, καθώς και πρόσθετα στοιχεία που αφορούν τη διαδικασία πρόβλεψης. Τα δεδομένα αυτά αξιοποιούνται από τη διεπαφή χρήστη για την παραγωγή γραφημάτων, πινάκων και στατιστικών στοιχείων που διευκολύνουν την ερμηνεία των αποτελεσμάτων.

Ο διαχωρισμός της επιχειρησιακής λογικής από τη διεπαφή χρήστη μέσω του API προσφέρει σημαντικά πλεονεκτήματα ως προς τη συντήρηση και την επεκτασιμότητα του συστήματος. Νέα μοντέλα πρόβλεψης μπορούν να ενσωματωθούν χωρίς να απαιτούνται αλλαγές στη διεπαφή χρήστη, ενώ παράλληλα η ίδια υπηρεσία μπορεί να χρησιμοποιηθεί από διαφορετικές εφαρμογές ή περιβάλλοντα στο μέλλον.

7.4 Διαχείριση ιστορικών δεδομένων κατανάλωσης

Η διαχείριση των ιστορικών δεδομένων κατανάλωσης αποτελεί βασικό τμήμα του συστήματος πρόβλεψης, καθώς το σενάριο με χρήση ιστορικών δεδομένων βασίζεται στην αξιοποίηση πρόσφατων μετρήσεων της ίδιας κατοικίας. Για τον λόγο αυτό αναπτύχθηκε μηχανισμός εισαγωγής, αποθήκευσης και επεξεργασίας ιστορικών δεδομένων, ο οποίος επιτρέπει στο σύστημα να λειτουργεί ανεξάρτητα από συγκεκριμένα αναγνωριστικά κατοικιών και να προσαρμόζεται σε διαφορετικούς χρήστες.

Στην τελική υλοποίηση κάθε αρχείο ιστορικού αντιστοιχεί σε μία μόνο κατοικία και περιλαμβάνει δύο βασικές στήλες: τη χρονική σήμανση της μέτρησης (timestamp) και την αντίστοιχη κατανάλωση ηλεκτρικής ενέργειας (consumption_Wh). Με τον τρόπο αυτό

αποφεύγεται η ανάγκη χρήσης αναγνωριστικών κατοικιών (home_id) κατά τη διαδικασία πρόβλεψης και το σύστημα μπορεί να χρησιμοποιηθεί από οποιονδήποτε χρήστη διαθέτει ιστορικά δεδομένα κατανάλωσης.

Για την παραγωγή προβλέψεων στο σενάριο με χρήση ιστορικών δεδομένων απαιτείται η ύπαρξη τουλάχιστον 168 ωρών ιστορικού κατανάλωσης, δηλαδή δεδομένων μίας πλήρους εβδομάδας. Η απαίτηση αυτή προκύπτει από τη χρήση χαρακτηριστικών χρονικής υστέρησης και κινούμενων παραθύρων, τα οποία αξιοποιούν πληροφορία από προηγούμενες ώρες και ημέρες. Εφόσον δεν υπάρχει επαρκές ιστορικό και ζητηθεί πρόβλεψη με χρήση ιστορικών δεδομένων, το σύστημα ενημερώνει τον χρήστη για την έλλειψη διαθέσιμων δεδομένων και δεν επιτρέπει την εκτέλεση του συγκεκριμένου σεναρίου.

Η εισαγωγή ιστορικών δεδομένων μπορεί να πραγματοποιηθεί είτε μέσω αρχείου CSV είτε μέσω προκαθορισμένου αποθηκευμένου αρχείου ιστορικού. Κατά την επιλογή ενός νέου αρχείου πραγματοποιούνται βασικοί έλεγχοι εγκυρότητας και συμβατότητας των δεδομένων με το σύστημα πρόβλεψης. Μετά την επιτυχή εισαγωγή, το αρχείο αποθηκεύεται σε ειδικό χώρο του συστήματος ώστε να είναι διαθέσιμο τόσο στη διεπαφή χρήστη όσο και στο επίπεδο υπηρεσιών.

Τα ιστορικά δεδομένα χρησιμοποιούνται αρχικά για τη δημιουργία των χαρακτηριστικών χρονικής εξάρτησης (lag features και rolling features) που απαιτούνται από τα μοντέλα πρόβλεψης. Παράλληλα, αξιοποιούνται και από προαιρετικούς μηχανισμούς βελτίωσης των προβλέψεων, όπως η προσαρμοστική διόρθωση με βάση το πρόσφατο ιστορικό κατανάλωσης, η οποία παρουσιάστηκε στην Ενότητα 6.6.

Η υιοθέτηση της συγκεκριμένης προσέγγισης καθιστά το σύστημα περισσότερο ευέλικτο και ανεξάρτητο από το σύνολο δεδομένων που χρησιμοποιήθηκε κατά την εκπαίδευση. Έτσι, το ίδιο περιβάλλον πρόβλεψης μπορεί να αξιοποιηθεί τόσο για ερευνητικούς σκοπούς όσο και για πραγματικά δεδομένα χρηστών, χωρίς να απαιτείται γνώση ή χρήση συγκεκριμένων κατοικιών του αρχικού συνόλου δεδομένων.

7.5 Διεπαφή χρήστη (User Interface Layer)

Για τη διευκόλυνση της χρήσης του συστήματος αναπτύχθηκε γραφική διεπαφή χρήστη με χρήση της βιβλιοθήκης Gradio. Στόχος της διεπαφής είναι να επιτρέπει την αξιοποίηση των εκπαιδευμένων μοντέλων πρόβλεψης χωρίς να απαιτούνται γνώσεις προγραμματισμού ή μηχανικής μάθησης από τον τελικό χρήστη. Μέσω ενός ενιαίου περιβάλλοντος, ο χρήστης μπορεί να εισαγάγει τα απαραίτητα δεδομένα, να επιλέξει το επιθυμητό σενάριο πρόβλεψης και να λάβει αναλυτικά αποτελέσματα για την ηλεκτρική κατανάλωση της επόμενης ημέρας.

Η διεπαφή υποστηρίζει τα δύο βασικά σενάρια πρόβλεψης που παρουσιάστηκαν στα προηγούμενα κεφάλαια. Στο σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης, ο χρήστης εισάγει χαρακτηριστικά της κατοικίας, όπως ο τύπος κατοικίας, το εμβαδόν, ο αριθμός ενοίκων και η τοποθεσία, καθώς και την ημερομηνία για την οποία επιθυμεί να παραχθεί πρόβλεψη. Στο σενάριο με χρήση ιστορικών δεδομένων, παρέχεται επιπλέον η δυνατότητα αξιοποίησης πρόσφατου ιστορικού κατανάλωσης για τη βελτίωση της ακρίβειας των προβλέψεων.

Η εισαγωγή ιστορικών δεδομένων πραγματοποιείται μέσω αρχείων CSV, τα οποία μπορούν να επιλεγούν απευθείας από τον χρήστη μέσω της διεπαφής. Τα αρχεία αυτά περιέχουν χρονοσειρές κατανάλωσης της κατοικίας και χρησιμοποιούνται για την παραγωγή των

χαρακτηριστικών ιστορικού που απαιτούνται από τα μοντέλα του σεναρίου με χρήση ιστορικών δεδομένων. Επιπλέον, παρέχεται ξεχωριστή λειτουργία ανάλυσης ιστορικού, μέσω της οποίας ο χρήστης μπορεί να εξετάσει πληροφορίες σχετικά με το χρονικό εύρος, την επάρκεια και τη σταθερότητα των διαθέσιμων δεδομένων κατανάλωσης.

Η εφαρμογή παρέχει πολλαπλές λειτουργίες πρόβλεψης. Ο χρήστης μπορεί να επιλέξει την παραγωγή πρόβλεψης από ένα συγκεκριμένο μοντέλο ή να ζητήσει ταυτόχρονη σύγκριση των διαθέσιμων μοντέλων Random Forest, XGBoost και LightGBM.

Παράλληλα υποστηρίζεται η δημιουργία συνδυαστικών προβλέψεων (ensemble predictions), στις οποίες η τελική πρόβλεψη προκύπτει από σταθμισμένο συνδυασμό των προβλέψεων των επιμέρους μοντέλων. Ο χρήστης μπορεί να καθορίζει τα βάρη συμμετοχής κάθε μοντέλου, επιτρέποντας τη διερεύνηση διαφορετικών στρατηγικών συνδυασμού και την αξιολόγηση της επίδρασής τους στα τελικά αποτελέσματα.

Στο σενάριο με χρήση ιστορικών δεδομένων παρέχεται επίσης η δυνατότητα ενεργοποίησης της προσαρμοστικής διόρθωσης με βάση το πρόσφατο ιστορικό κατανάλωσης, η οποία παρουσιάστηκε στην Ενότητα 6.6. Η λειτουργία αυτή εφαρμόζεται μετά την παραγωγή της αρχικής πρόβλεψης και αξιοποιεί πληροφορίες από το πρόσφατο ιστορικό της κατοικίας για την προσαρμογή της τελικής καμπύλης φορτίου.

Επιπλέον, η διεπαφή υποστηρίζει διαφορετικές στρατηγικές βελτιστοποίησης των προβλέψεων μέσω της παραμέτρου Optimization Mode. Ο χρήστης μπορεί να επιλέξει μεταξύ της λειτουργίας balanced και της λειτουργίας daily. Η λειτουργία balanced αποτελεί την προεπιλεγμένη επιλογή του συστήματος και δίνει προτεραιότητα στη συνολική ακρίβεια των ωριαίων προβλέψεων, καθώς ο βασικός στόχος της παρούσας εργασίας είναι η όσο το δυνατόν ακριβέστερη αναπαράσταση της ωριαίας καμπύλης κατανάλωσης. Αντίθετα, η λειτουργία daily εφαρμόζει πρόσθετες τεχνικές βαθμονόμησης με στόχο τη βελτίωση της εκτίμησης της συνολικής ημερήσιας κατανάλωσης. Όπως παρουσιάστηκε στην Ενότητα 6.6, η προσέγγιση αυτή μπορεί να μειώσει το ημερήσιο ενεργειακό σφάλμα, συνήθως όμως με κόστος μια μικρή επιδείνωση της ακρίβειας των επιμέρους ωριαίων προβλέψεων. Με τον τρόπο αυτό ο χρήστης μπορεί να επιλέξει τη στρατηγική που ανταποκρίνεται καλύτερα στις απαιτήσεις του εκάστοτε σεναρίου χρήσης.

Επιπροσθέτως, η διεπαφή περιλαμβάνει μηχανισμό προσαρμογής της αναμενόμενης συμπεριφοράς κατανάλωσης (behavioral adjustment). Μέσω της λειτουργίας αυτής ο χρήστης μπορεί να δηλώσει αναμενόμενες αυξήσεις ή μειώσεις της κατανάλωσης κατά συγκεκριμένες ώρες της ημέρας, ενσωματώνοντας πληροφορίες που δεν είναι διαθέσιμες στα ιστορικά δεδομένα ή στα χαρακτηριστικά εισόδου των μοντέλων.

Τα αποτελέσματα παρουσιάζονται με τη μορφή πινάκων και γραφημάτων, διευκολύνοντας την ερμηνεία των προβλέψεων. Για κάθε πρόβλεψη εμφανίζεται η αναμενόμενη κατανάλωση ανά ώρα, η συνολική ημερήσια κατανάλωση και συγκεντρωτικά στατιστικά στοιχεία που βοηθούν τον χρήστη να αξιολογήσει το παραγόμενο αποτέλεσμα. Παράλληλα παρέχεται δυνατότητα εξαγωγής των αποτελεσμάτων σε μορφή CSV για περαιτέρω ανάλυση ή αποθήκευση.

Η ανάπτυξη της διεπαφής χρήστη αποτελεί βασικό στοιχείο της παρούσας εργασίας, καθώς μετατρέπει τα αποτελέσματα των μοντέλων μηχανικής μάθησης σε ένα πρακτικό και εύχρηστο εργαλείο πρόβλεψης. Με τον τρόπο αυτό η προτεινόμενη μεθοδολογία δεν περιορίζεται σε πειραματικό επίπεδο, αλλά μπορεί να αξιοποιηθεί ως βάση για πραγματικές εφαρμογές πρόβλεψης και διαχείρισης ηλεκτρικής κατανάλωσης

7.6 Εγκατάσταση και περιβάλλον εκτέλεσης

Για την ορθή λειτουργία και την αναπαραγωγικότητα του συστήματος πρόβλεψης δημιουργήθηκε απομονωμένο περιβάλλον εκτέλεσης Python (virtual environment). Η χρήση εικονικού περιβάλλοντος επιτρέπει την ανεξάρτητη εγκατάσταση των απαραίτητων βιβλιοθηκών και περιορίζει τον κίνδυνο ασυμβατοτήτων μεταξύ διαφορετικών εκδόσεων λογισμικού ή ήδη εγκατεστημένων πακέτων στο σύστημα του χρήστη.

Για την υλοποίηση της εργασίας, η εγκατάσταση του συστήματος υποστηρίζεται από αυτοματοποιημένο σενάριο τύπου batch (*install_requirements.bat*), το οποίο εκτελεί διαδοχικά όλα τα απαραίτητα βήματα προετοιμασίας του περιβάλλοντος. Ειδικότερα, το σενάριο ελέγχει αρχικά τη διαθεσιμότητα συμβατής έκδοσης Python (3.11 ή 3.12), δημιουργεί νέο εικονικό περιβάλλον *venv*, αναβαθμίζει το εργαλείο διαχείρισης πακέτων *pip* και εγκαθιστά τις απαιτούμενες βιβλιοθήκες που καταγράφονται σε ένα συγκεκριμένο αρχείο (*requirements.txt*).

Επιπλέον, μετά την ολοκλήρωση της εγκατάστασης εκτελείται αυτοματοποιημένος έλεγχος ορθής ρύθμισης του περιβάλλοντος (*self-test*), μέσω δοκιμαστικής εισαγωγής των βασικών βιβλιοθηκών του συστήματος, όπως *FastAPI*, *Uvicorn*, *pandas*, *NumPy*, *scikit-learn*, *Gradio*, *requests* και *joblib*. Με τον τρόπο αυτό επιβεβαιώνεται ότι το περιβάλλον εκτέλεσης έχει διαμορφωθεί σωστά και ότι το σύστημα είναι έτοιμο προς χρήση.

Η προσέγγιση αυτή διευκολύνει σημαντικά τη μεταφορά και την εγκατάσταση της εφαρμογής σε άλλο υπολογιστικό σύστημα, καθώς μειώνει την ανάγκη χειροκίνητης παραμετροποίησης και επιτρέπει την επαναχρησιμοποίηση της ίδιας ακριβώς στοίβας λογισμικού. Κατά συνέπεια, ενισχύεται η αναπαραγωγικότητα της υλοποίησης και καθίσταται ευκολότερη η αξιολόγηση του συστήματος από τρίτους χρήστες.

Τέλος, μετά την επιτυχή εγκατάσταση των εξαρτήσεων, η εκκίνηση της εφαρμογής πραγματοποιείται μέσω ξεχωριστού σεναρίου εκκίνησης, το οποίο ενεργοποιεί το περιβάλλον εκτέλεσης και θέτει σε λειτουργία τόσο τον διακομιστή *FastAPI* όσο και τη διεπαφή χρήστη *Gradio*. Οι δύο υπηρεσίες λειτουργούν ανεξάρτητα και επικοινωνούν μεταξύ τους μέσω της προγραμματιστικής διεπαφής εφαρμογής (API), σύμφωνα με την αρχιτεκτονική που παρουσιάστηκε στις προηγούμενες ενότητες.

8. Εφαρμογή, διεπαφή χρήστη και σενάρια χρήσης

Στο προηγούμενο κεφάλαιο παρουσιάστηκε η αρχιτεκτονική και η υλοποίηση του συστήματος πρόβλεψης ηλεκτρικής κατανάλωσης, συμπεριλαμβανομένων των επιμέρους επιπέδων λογισμικού, της διαχείρισης ιστορικών δεδομένων και της διεπαφής χρήστη. Στο παρόν κεφάλαιο η έμφαση μετατοπίζεται από την εσωτερική υλοποίηση στη λειτουργική αξιοποίηση της εφαρμογής από την πλευρά του τελικού χρήστη.

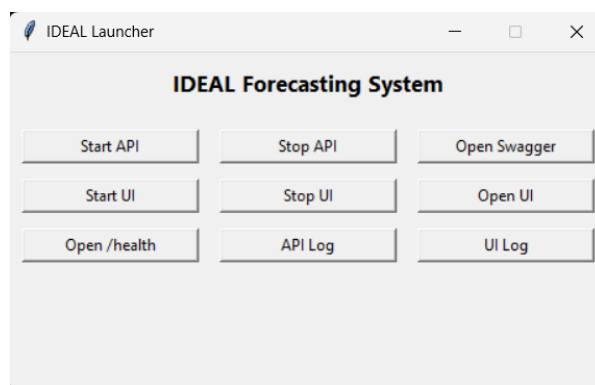
Αρχικά παρουσιάζεται η διαδικασία εκκίνησης της εφαρμογής και η πρόσβαση στις διαθέσιμες υπηρεσίες. Στη συνέχεια περιγράφεται η βασική διεπαφή χρήστη και η λειτουργία των κύριων στοιχείων της. Ακολουθούν ενδεικτικά σενάρια χρήσης που καλύπτουν τόσο την πρόβλεψη χωρίς ιστορικά δεδομένα κατανάλωσης όσο και την πρόβλεψη με χρήση ιστορικού, αξιοποιώντας τις δυνατότητες που παρέχει το σύστημα για ανάλυση ιστορικών δεδομένων και βελτίωση των προβλέψεων.

Επιπλέον, παρουσιάζονται οι λειτουργίες σύγκρισης μοντέλων και δημιουργίας συνδυαστικών προβλέψεων (*ensemble predictions*), καθώς και οι πρόσθετες δυνατότητες προσαρμογής των αποτελεσμάτων που προσφέρονται μέσω της διεπαφής. Τέλος, εξετάζεται ο τρόπος διαχείρισης των μετεωρολογικών δεδομένων και η αξιοποίησή τους κατά τη διαδικασία πρόβλεψης.

Στόχος του κεφαλαίου είναι να καταδείξει τον τρόπο με τον οποίο τα μοντέλα μηχανικής μάθησης και οι τεχνικές που παρουσιάστηκαν στα προηγούμενα κεφάλαια ενσωματώνονται σε ένα ολοκληρωμένο και λειτουργικό σύστημα πρόβλεψης, το οποίο μπορεί να χρησιμοποιηθεί από μη εξειδικευμένους χρήστες χωρίς την ανάγκη άμεσης αλληλεπίδρασης με τον υποκείμενο κώδικα ή τα επιμέρους μοντέλα.

8.1 Αρχιτεκτονική εκτέλεσης και μηχανισμός εκκίνησης

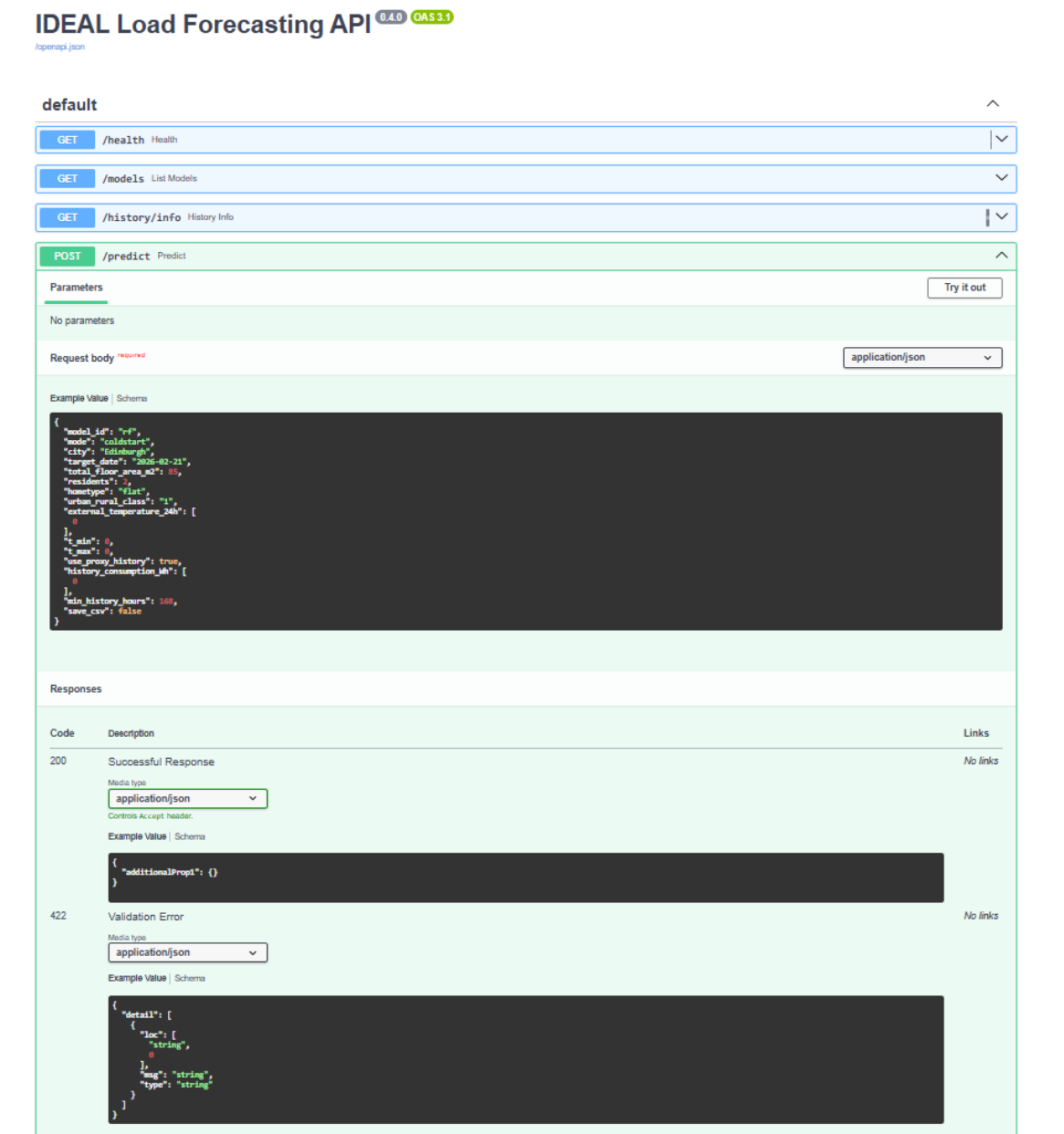
Η εκτέλεση του συστήματος οργανώνεται γύρω από το αρχείο *launcher_ui.py*, το οποίο λειτουργεί ως κεντρικός μηχανισμός εκκίνησης και διαχείρισης των επιμέρους συνιστωσών του συστήματος. Ο μηχανισμός αυτός παρέχει ένα απλό γραφικό περιβάλλον ελέγχου, το οποίο, όπως φαίνεται στην Εικόνα 2, περιλαμβάνει επιλογές όπως η εκκίνηση και διακοπή του επιπέδου υπηρεσιών (*Start API*, *Stop API*), η εκκίνηση και διακοπή της γραφικής διεπαφής χρήστη (*Start UI*, *Stop UI*), καθώς και η πρόσβαση στη διεπαφή τεκμηρίωσης (*Open Swagger*) και στη γραφική διεπαφή (*Open UI*).



Εικόνα 2 Γραφικό περιβάλλον του *launcher_ui* του IDEAL Forecasting System

Η τυπική διαδικασία εκτέλεσης ξεκινά με την ενεργοποίηση του API μέσω της επιλογής *Start API*. Αφού ολοκληρωθεί επιτυχώς η εκκίνηση του επιπέδου υπηρεσιών, ο χρήστης μπορεί να ενεργοποιήσει τη γραφική διεπαφή μέσω της επιλογής *Start UI* και να αποκτήσει πρόσβαση σε αυτή είτε μέσω του κουμπιού *Open UI* είτε απευθείας μέσω της τοπικής διεύθυνσης που παρέχει η εφαρμογή.

Παράλληλα, η επιλογή *Open Swagger* παρέχει πρόσβαση στην αυτόματα παραγόμενη τεκμηρίωση του API. Μέσω της διεπαφής αυτής είναι δυνατή η προβολή και η δοκιμή των διαθέσιμων υπηρεσιών πρόβλεψης, καθώς και η εξέταση των δομών εισόδου και εξόδου που χρησιμοποιούνται από το σύστημα (Εικόνα 3).



Εικόνα 3 IDEAL Load Forecasting API μέσω Swagger

Επιπλέον, η επιλογή *Open Health* επιτρέπει τον άμεσο έλεγχο της κατάστασης λειτουργίας του API, εμφανίζοντας πληροφορίες σχετικά με τη διαθεσιμότητα της υπηρεσίας. Παράλληλα, οι επιλογές *API Log* και *UI Log* παρέχουν πρόσβαση στα αρχεία καταγραφής των δύο υποσυστημάτων, επιτρέποντας την παρακολούθηση της λειτουργίας τους και τη διάγνωση πιθανών προβλημάτων κατά την εκκίνηση ή τη χρήση της εφαρμογής.

Η συγκεκριμένη αρχιτεκτονική υποστηρίζει δύο διαφορετικούς τρόπους αλληλεπίδρασης με το σύστημα. Ο πρώτος βασίζεται στην άμεση χρήση του API μέσω της διεπαφής Swagger, ενώ ο δεύτερος αξιοποιεί τη γραφική διεπαφή χρήστη που αναπτύχθηκε με τη βιβλιοθήκη Gradio. Στην πράξη, η χρήση της γραφικής διεπαφής είναι σημαντικά πιο φιλική για τον τελικό χρήστη, καθώς επιτρέπει την εκτέλεση προβλέψεων, την επιλογή μοντέλων και την επισκόπηση των αποτελεσμάτων χωρίς να απαιτείται γνώση της δομής των αιτημάτων ή διαχείριση δεδομένων σε μορφή JSON.

Η επόμενη ενότητα παρουσιάζει αναλυτικά τη γραφική διεπαφή χρήστη και τον τρόπο με τον οποίο υποστηρίζει τα διαφορετικά σενάρια πρόβλεψης που υλοποιήθηκαν στο πλαίσιο της παρούσας εργασίας.

8.2 Επισκόπηση γραφικής διεπαφής χρήστη

Μετά την εκκίνηση της εφαρμογής, ο χρήστης αποκτά πρόσβαση στην κύρια διεπαφή χρήστη του συστήματος πρόβλεψης. Η διεπαφή έχει σχεδιαστεί ώστε να συγκεντρώνει όλες τις απαραίτητες λειτουργίες σε μία ενιαία οθόνη, επιτρέποντας την εύκολη διαμόρφωση των παραμέτρων πρόβλεψης και την άμεση πρόσβαση στα αποτελέσματα (Εικόνα 4).

IDEAL Load Forecasting — Generic API/UI Demo

Συμπλήρωσε τα χαρακτηριστικά της κατοικίας και πάτα Predict για ένα μοντέλο, Compare για σύγκριση RF / XGB / LGBM ή Combined Prediction για weighted ensemble.

API: OK | Ready models: 6 | No-history: rf_coldstart_default | With-history: lgbm_withhistory_default

Refresh API/models

Αν ενεργοποιηθεί, το API χρησιμοποιεί mode="with_history" και απαιτεί πρόσφατο ιστορικό κατανάλωσης.

☐ Πρόβλεψη με χρήση ιστορικών δεδομένων

Μοντέλο
Auto default: RF για no-history, LGBM για with-history.

auto — Auto default

Optimization
daily χρησιμοποιεί daily-calibrated output όπου υπάρχει διαθέσιμο.

balanced

City

Edinburgh

Target date (YYYY-MM-DD)

2026-06-04

Total floor area (m²)

85

Residents

2

Number of electric appliances [optional]

0

Home type

flat

Urban/Rural class

1 — Large urban area

Weather / environmental inputs

Το API προσπαθεί να πάρει αυτόματα εξωτερική θερμοκρασία από Open-Meteo. Αν δεν βρεθεί, συμπληρώσε t_min / t_max ή δώσε 24 ωριαίες τιμές.

external_temperature_24h (24 τιμές, comma-separated) [optional]

π.χ. 3.1, 2.9, 2.7, ... (24 συνολικά)

internal_temperature_24h (24 τιμές, comma-separated) [optional]

π.χ. 19.5, 19.4, 19.3, ... (24 συνολικά)

Behavioral adjustment (optional)

Ensemble weights (optional)

Εικόνα 4 Κύρια διεπαφή χρήστη του IDEAL Load Forecasting System

Στο ανώτερο τμήμα της οθόνης εμφανίζονται πληροφορίες σχετικά με την κατάσταση του συστήματος και τη διαθεσιμότητα των εκπαιδευμένων μοντέλων. Παράλληλα παρέχεται η δυνατότητα ανανέωσης των διαθέσιμων μοντέλων μέσω της επιλογής *Refresh API/models*, η οποία επικοινωνεί με το επίπεδο υπηρεσιών και ενημερώνει τη διεπαφή για τα μοντέλα που είναι διαθέσιμα προς χρήση.

Αμέσως κάτω από την περιοχή αυτή βρίσκεται η επιλογή ενεργοποίησης του σεναρίου πρόβλεψης με ιστορικά δεδομένα κατανάλωσης (*with-history mode*). Όταν η επιλογή παραμένει απενεργοποιημένη, το σύστημα λειτουργεί στο σενάριο χωρίς χρήση ιστορικών δεδομένων, αξιοποιώντας αποκλειστικά χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά της κατοικίας. Αντίθετα, όταν ενεργοποιηθεί, το σύστημα χρησιμοποιεί και πρόσφατο ιστορικό κατανάλωσης, επιτρέποντας την παραγωγή ακριβέστερων προβλέψεων.

Στη συνέχεια παρέχονται επιλογές για την επιλογή του μοντέλου πρόβλεψης και του τρόπου βελτιστοποίησης των αποτελεσμάτων. Ο χρήστης μπορεί να επιλέξει συγκεκριμένο μοντέλο (Random Forest, XGBoost ή LightGBM) ή να αφήσει το σύστημα να χρησιμοποιήσει το προεπιλεγμένο μοντέλο για το αντίστοιχο σενάριο λειτουργίας. Παράλληλα υποστηρίζονται διαφορετικές στρατηγικές βελτιστοποίησης, όπως η λειτουργία *balanced*, η οποία δίνει έμφαση στη συνολική ακρίβεια των ωριαίων προβλέψεων και η λειτουργία *daily*, η οποία στοχεύει στη βελτίωση της εκτίμησης της συνολικής ημερήσιας κατανάλωσης.

Η κύρια περιοχή εισαγωγής δεδομένων περιλαμβάνει τα χαρακτηριστικά της κατοικίας και του νοικοκυριού. Συγκεκριμένα, ο χρήστης μπορεί να εισαγάγει την πόλη, την ημερομηνία πρόβλεψης, το συνολικό εμβαδόν της κατοικίας, τον αριθμό των ενοίκων, τον αριθμό των ηλεκτρικών συσκευών, τον τύπο κατοικίας και την κατηγορία αστικότητας. Τα στοιχεία αυτά χρησιμοποιούνται για τη δημιουργία των χαρακτηριστικών εισόδου που απαιτούνται από τα μοντέλα μηχανικής μάθησης.

Η διεπαφή παρέχει επίσης ξεχωριστή ενότητα για την εισαγωγή περιβαλλοντικών δεδομένων. Το σύστημα επιχειρεί αρχικά να ανακτήσει αυτόματα μετεωρολογικά δεδομένα μέσω εξωτερικών υπηρεσιών. Εναλλακτικά, ο χρήστης μπορεί να εισαγάγει χειροκίνητα ωριαίες τιμές εξωτερικής ή εσωτερικής θερμοκρασίας, οι οποίες θα χρησιμοποιηθούν κατά τη διαδικασία πρόβλεψης.

Στο κάτω τμήμα της οθόνης βρίσκονται επιπλέον προαιρετικές λειτουργίες, όπως οι προσαρμογές συμπεριφοράς (*behavioral adjustments*) και οι συνδυαστικές προβλέψεις (*ensemble predictions*). Οι λειτουργίες αυτές επιτρέπουν την περαιτέρω προσαρμογή των προβλέψεων στις ανάγκες του χρήστη και παρουσιάζονται αναλυτικότερα στις επόμενες ενότητες.

Η συγκέντρωση όλων των παραμέτρων σε μία ενιαία οθόνη επιτρέπει την ταχεία διαμόρφωση σεναρίων πρόβλεψης και μειώνει σημαντικά την πολυπλοκότητα χρήσης του συστήματος.

8.3 Βασική ροή πρόβλεψης χωρίς ιστορικά δεδομένα

Το σενάριο πρόβλεψης χωρίς ιστορικά δεδομένα κατανάλωσης απευθύνεται σε περιπτώσεις όπου δεν είναι διαθέσιμες προηγούμενες μετρήσεις ηλεκτρικής κατανάλωσης της κατοικίας. Το συγκεκριμένο σενάριο είναι ιδιαίτερα χρήσιμο για νέες κατοικίες ή για

χρήστες που επιθυμούν να λάβουν πρόβλεψη χωρίς να διαθέτουν καταγεγραμμένο ιστορικό κατανάλωσης.

Για την παραγωγή πρόβλεψης ο χρήστης αφήνει απενεργοποιημένη την επιλογή «Πρόβλεψη με χρήση ιστορικών δεδομένων» και συμπληρώνει τα διαθέσιμα χαρακτηριστικά της κατοικίας. Τα βασικά στοιχεία εισόδου περιλαμβάνουν την πόλη, την ημερομηνία πρόβλεψης, το συνολικό εμβαδόν της κατοικίας, τον αριθμό των ενοίκων, τον τύπο κατοικίας και την κατηγορία αστικότητας. Προαιρετικά μπορούν να δοθούν επιπλέον πληροφορίες, όπως ο αριθμός των ηλεκτρικών συσκευών ή ωριαίες τιμές εσωτερικής θερμοκρασίας.

Στη συνέχεια ο χρήστης επιλέγει το μοντέλο πρόβλεψης που επιθυμεί να χρησιμοποιήσει. Η εφαρμογή παρέχει τη δυνατότητα επιλογής των μοντέλων Random Forest, XGBoost και LightGBM, καθώς και την επιλογή *Auto*, η οποία χρησιμοποιεί το προτεινόμενο προεπιλεγμένο μοντέλο του συστήματος για το συγκεκριμένο σενάριο. Σύμφωνα με τα αποτελέσματα που παρουσιάστηκαν στο Κεφάλαιο 6, ως προεπιλεγμένο μοντέλο για το σενάριο χωρίς χρήση ιστορικών δεδομένων έχει επιλεγεί το Random Forest, καθώς παρουσίασε τη συνολικά καλύτερη απόδοση στις βασικές μετρικές αξιολόγησης.

Παράλληλα, ο χρήστης μπορεί να επιλέξει τον τρόπο βελτιστοποίησης της πρόβλεψης μέσω της παραμέτρου *Optimization*. Η λειτουργία *balanced* δίνει προτεραιότητα στη συνολική ακρίβεια των ωριαίων προβλέψεων, ενώ η λειτουργία *daily* εστιάζει στη βελτίωση της εκτίμησης της συνολικής ημερήσιας κατανάλωσης.

Αφού ολοκληρωθεί η συμπλήρωση των απαραίτητων στοιχείων, η πρόβλεψη εκτελείται μέσω της επιλογής *Predict*. Η διεπαφή αποστέλλει τα δεδομένα στο επίπεδο υπηρεσιών FastAPI, το οποίο αναλαμβάνει τη δημιουργία των κατάλληλων χαρακτηριστικών εισόδου, την εφαρμογή των αντίστοιχων βημάτων προεπεξεργασίας, την εκτέλεση του επιλεγμένου μοντέλου και την επιστροφή των αποτελεσμάτων.

Τα αποτελέσματα παρουσιάζονται σε μορφή πίνακα, όπου εμφανίζεται η προβλεπόμενη κατανάλωση για κάθε ώρα της επόμενης ημέρας. Παράλληλα εμφανίζονται συνοπτικές πληροφορίες σχετικά με το μοντέλο που χρησιμοποιήθηκε, τη συνολική προβλεπόμενη ημερήσια κατανάλωση και τις βασικές παραμέτρους του σεναρίου πρόβλεψης.

Η λειτουργία αυτή επιτρέπει την παραγωγή 24ωρων προβλέψεων ακόμη και όταν δεν υπάρχει διαθέσιμο ιστορικό κατανάλωσης, αξιοποιώντας αποκλειστικά τα χαρακτηριστικά της κατοικίας, τα χρονικά χαρακτηριστικά και τις διαθέσιμες περιβαλλοντικές πληροφορίες. Με τον τρόπο αυτό το σύστημα μπορεί να υποστηρίξει σενάρια αρχικής εκτίμησης κατανάλωσης ή νέες εγκαταστάσεις όπου δεν έχει ακόμη συγκεντρωθεί επαρκές ιστορικό δεδομένων.

Ένα ενδεικτικό παράδειγμα εκτέλεσης πρόβλεψης, στο σενάριο χωρίς χρήση ιστορικών δεδομένων, παρουσιάζεται στην Εικόνα 5.



Εικόνα 5 Παράδειγμα αποτελεσμάτων πρόβλεψης στο σενάριο χωρίς χρήση ιστορικών δεδομένων

8.4 Πρόβλεψη με ιστορικά δεδομένα και ανάλυση ιστορικού

Εκτός από το σενάριο πρόβλεψης χωρίς ιστορικά δεδομένα, το σύστημα υποστηρίζει και πρόβλεψη με αξιοποίηση πρόσφατου ιστορικού κατανάλωσης. Για την ενεργοποίηση του σεναρίου αυτού ο χρήστης επιλέγει την επιλογή «*Πρόβλεψη με χρήση ιστορικών δεδομένων*» και στη συνέχεια εισάγει αρχείο CSV που περιέχει ιστορικές μετρήσεις κατανάλωσης της κατοικίας. Η επιλογή του αρχείου πραγματοποιείται μέσω της λειτουργίας *Select History CSV*, η οποία επιτρέπει την εύκολη φόρτωση των δεδομένων από το τοπικό σύστημα αρχείων.

Πριν από την παραγωγή πρόβλεψης, το σύστημα παρέχει τη δυνατότητα ανάλυσης του διαθέσιμου ιστορικού μέσω της επιλογής *Analyze History Stability*. Η λειτουργία αυτή εξετάζει τα πρόσφατα δεδομένα κατανάλωσης και υπολογίζει μια σειρά από δείκτες που σχετίζονται με τη σταθερότητα και την επαναληψιμότητα της ενεργειακής συμπεριφοράς της κατοικίας. Επιπλέον, εμφανίζει περιγραφικά στοιχεία για το ημερήσιο επίπεδο κατανάλωσης, όπως τη μέση ημερήσια κατανάλωση, τη διάμεση ημερήσια κατανάλωση και το εύρος των ημερήσιων τιμών στο εξεταζόμενο ιστορικό.

History options — μόνο για with-history mode
Το ιστορικό αρχείο θεωρείται ότι αφορά μία μόνο κατοικία. Αν θέλεις άλλη κατοικία, επίλεξε άλλο CSV. Απαιτούμενες στήλες: timestamp, consumption_wh.

Select history CSV

Selected history CSV path [optional]
C:\IDEAL_Programming\processed\stores\selected_history_files\history_store_20260610_103249.csv

Selected history CSV: C:\IDEAL_Programming\processed\stores\selected_history_files\history_store_20260610_103249.csv
Fallback μόνο αν δεν βολέει upload/path/manual history.

☒ Use legacy default history_store.csv if no file/path is selected

min_history_hours
168

history_consumption_wh (comma-separated) [optional]
π.χ. 120, 98, 105, ... (≈168 τιμές αν το δώσεις εσύ)

Adaptive history-aware correction
Εφαρμόζεται μόνο στο with-history mode, αν υπάρχει επαρκές πρόσφατο ιστορικό.

☒ Apply adaptive 7-day median shape correction

history_correction_days
7

history_correction_max_alpha
Manual value. If 'Use recommended max alpha' is active, the recommended value below is used instead.
0.2

☐ Use recommended max alpha

Analyze history stability / Recommend max alpha

History stability score
0.490

History stability category
Medium

Recommended max_alpha
Computed from history stability. Excellent stability can recommend up to 0.80; values above 0.65 should remain manual/experimental.
0.1

History stability explanation

History source: C:\IDEAL_Programming\processed\stores\selected_history_files\history_store_20260610_103249.csv
Target date: 2026-06-10
History window checked: 2026-06-03 00:00:00 έως 2026-06-10 00:00:00 (exclusive)
Expected hours: 168 | Observed hours: 168 | Missing hours: 0
Completeness score: 1.000
Mean daily consumption: 6.237 kWh/day | Median: 5.989 kWh/day
Daily consumption range: 4.009-9.028 kWh/day
Daily total CV: 0.229 | Daily stability score: 0.346
Hourly profile: median CV: 0.274 | Hourly stability score: 0.634
Peak ratio CV: 0.463 | Peak stability score: 0.673
Same day type days in history: 5 | Day-type score: 1.000
Recommended max_alpha range for category: 0.10-0.15
Recommendation reason: Το ιστορικό έχει μέτρια σταθερότητα. Προτείνεται χαμηλό max_alpha ώστε να περιοριστεί ο κίνδυνος υπερπροσαρμογής στο πρόσφατο history shape.

☒ History stability analyzed. Category=Medium. Recommended max_alpha=0.10 (range 0.10-0.15). Tick 'Use recommended max alpha' to use it in Predict/Compare/Combined.

Εικόνα 6 Παράδειγμα ανάλυσης ιστορικού κατανάλωσης

Η ανάγκη για τον συγκεκριμένο έλεγχο προκύπτει από το γεγονός ότι η ύπαρξη ιστορικών δεδομένων δεν συνεπάγεται απαραίτητα ότι αυτά είναι αντιπροσωπευτικά της ημέρας πρόβλεψης. Ένα ιστορικό μπορεί να είναι πλήρες ως προς τον αριθμό των ωριαίων

μετρήσεων, αλλά να παρουσιάζει μεγάλη μεταβλητότητα ως προς το ημερήσιο επίπεδο κατανάλωσης, το ωριαίο προφίλ ή τις αιχμές. Σε τέτοιες περιπτώσεις, η υπερβολική προσαρμογή της πρόβλεψης στο πρόσφατο ιστορικό μπορεί να οδηγήσει σε χειρότερα αποτελέσματα. Οι υπολογισμοί που αφορούν τη σταθερότητα του ιστορικού, δεν αποτελούν μέρος της εκπαίδευσης των μοντέλων, αλλά πραγματοποιούνται κατά τη χρήση της εφαρμογής, στο πλαίσιο της μεταγενέστερης προσαρμογής των προβλέψεων.

Η ανάλυση πραγματοποιείται στο χρονικό παράθυρο που προηγείται της ημέρας πρόβλεψης. Στην υλοποίηση της εφαρμογής χρησιμοποιείται, κατά προεπιλογή, παράθυρο 168 ωρών, δηλαδή οι επτά ημέρες πριν από την ημέρα στόχο. Το σύστημα ελέγχει αρχικά αν υπάρχουν όλες οι αναμενόμενες ωριαίες εγγραφές στο διάστημα αυτό και υπολογίζει δείκτη πληρότητας (completeness score). Η πληρότητα είναι σημαντική, επειδή τα μοντέλα με ιστορικά χαρακτηριστικά απαιτούν επαρκείς προηγούμενες τιμές για τη δημιουργία χρονικών υστερήσεων και κυλιόμενων στατιστικών.

Παράλληλα, το σύστημα υπολογίζει τα συνολικά ημερήσια ενεργειακά αθροίσματα του ιστορικού και εμφανίζει τη μέση ημερήσια κατανάλωση, τη διάμεση ημερήσια κατανάλωση και το εύρος των ημερήσιων τιμών. Οι πληροφορίες αυτές δεν χρησιμοποιούνται ως άμεσος περιορισμός της πρόβλεψης, αλλά λειτουργούν ως διαγνωστικό και ερμηνευτικό στοιχείο. Με αυτόν τον τρόπο ο χρήστης μπορεί να εκτιμήσει αν το πρόσφατο ιστορικό υποδηλώνει χαμηλό, μέτριο ή αυξημένο ημερήσιο επίπεδο κατανάλωσης και αν η τελική πρόβλεψη κινείται σε εύλογο εύρος σε σχέση με τη συμπεριφορά των προηγούμενων ημερών.

Στη συνέχεια υπολογίζονται επιμέρους δείκτες σταθερότητας. Ο πρώτος αφορά τη μεταβλητότητα της συνολικής ημερήσιας κατανάλωσης στο πρόσφατο ιστορικό. Για τον σκοπό αυτό υπολογίζεται ο συντελεστής μεταβλητότητας των ημερήσιων ενεργειακών συνόλων. Όσο μικρότερη είναι η μεταβλητότητα, τόσο πιο σταθερό θεωρείται το ημερήσιο επίπεδο κατανάλωσης της κατοικίας.

Ο δεύτερος δείκτης αφορά τη σταθερότητα του ωριαίου προφίλ. Το σύστημα εξετάζει κατά πόσο η κατανάλωση σε κάθε ώρα της ημέρας παρουσιάζει παρόμοια συμπεριφορά στις ημέρες του ιστορικού. Η πληροφορία αυτή είναι ιδιαίτερα σημαντική για την προσαρμοστική διόρθωση, επειδή η διόρθωση επιχειρεί να αξιοποιήσει το πρόσφατο σχήμα της ημερήσιας καμπύλης κατανάλωσης.

Ο τρίτος δείκτης αφορά τη σταθερότητα των αιχμών. Για κάθε ημέρα του ιστορικού υπολογίζεται η σχέση της μέγιστης ωριαίας κατανάλωσης προς τη μέση ημερήσια κατανάλωση και στη συνέχεια εκτιμάται η μεταβλητότητα αυτού του λόγου. Με τον τρόπο αυτό εξετάζεται αν οι αιχμές εμφανίζονται με σχετικά σταθερό τρόπο ή αν παρουσιάζουν έντονη αστάθεια.

Επιπλέον, λαμβάνεται υπόψη και ο τύπος της ημέρας. Το σύστημα ελέγχει πόσες ημέρες του ιστορικού έχουν τον ίδιο τύπο με την ημέρα πρόβλεψης, δηλαδή αν πρόκειται για εργάσιμη ημέρα ή για Σαββατοκύριακο. Όσο περισσότερες ημέρες ίδιου τύπου υπάρχουν στο ιστορικό, τόσο μεγαλύτερη θεωρείται η συνάφεια του ιστορικού με την ημέρα στόχο.

Ο βασικός υπολογισμός των επιμέρους δεικτών βασίζεται στον συντελεστή μεταβλητότητας (Coefficient of Variation – CV), ο οποίος εκφράζει τη σχετική διακύμανση ενός μεγέθους ως προς τη μέση τιμή του:

$$CV(x) = \sigma(x) / \bar{x} \quad (8)$$

όπου $\sigma(x)$ είναι η τυπική απόκλιση και $\bar{x}$ η μέση τιμή του μεγέθους x . Για τη μετατροπή του συντελεστή μεταβλητότητας σε δείκτη σταθερότητας χρησιμοποιείται κανονικοποίηση στο διάστημα $[0,1]$, σύμφωνα με τη σχέση:

$$S_CV = \max(0, \min(1, 1 - CV / CV_bad)) \quad (9)$$

όπου CV_bad είναι τιμή αναφοράς πάνω από την οποία το αντίστοιχο μέγεθος θεωρείται ιδιαίτερα ασταθές. Με αυτόν τον τρόπο, μικρές τιμές μεταβλητότητας οδηγούν σε υψηλό δείκτη σταθερότητας, ενώ μεγάλες τιμές μεταβλητότητας οδηγούν σε χαμηλό δείκτη. Στην υλοποίηση της εφαρμογής χρησιμοποιούνται διαφορετικές τιμές αναφοράς ανάλογα με τον δείκτη που υπολογίζεται: 0,35 για τη σταθερότητα της ημερήσιας κατανάλωσης, 0,75 για τη σταθερότητα του ωριαίου προφίλ και 0,50 για τη σταθερότητα των αιχμών.

Συγκεκριμένα, υπολογίζονται τρεις βασικοί δείκτες σταθερότητας: ο δείκτης ημερήσιας σταθερότητας S_daily , ο δείκτης σταθερότητας ωριαίου προφίλ S_hourly και ο δείκτης σταθερότητας αιχμών S_peak . Οι δείκτες αυτοί συνδυάζονται με τον δείκτη πληρότητας S_comp σε έναν ενιαίο βασικό δείκτη σταθερότητας:

$$S_base = 0,30 \cdot S_daily + 0,35 \cdot S_hourly + 0,20 \cdot S_peak + 0,15 \cdot S_comp \quad (10)$$

Στη συνέχεια λαμβάνεται υπόψη και ο δείκτης τύπου ημέρας $S_daytype$, ο οποίος εκφράζει κατά πόσο το διαθέσιμο ιστορικό περιλαμβάνει επαρκείς ημέρες ίδιου τύπου με την ημέρα πρόβλεψης. Ο τελικός δείκτης σταθερότητας του ιστορικού υπολογίζεται ως:

$$S_hist = \max(0, \min(1, S_base \cdot (0,85 + 0,15 \cdot S_daytype))) \quad (11)$$

όπου S_hist είναι ο συνολικός δείκτης σταθερότητας του ιστορικού. Η τιμή αυτή χρησιμοποιείται για την κατάταξη του διαθέσιμου ιστορικού σε κατηγορία σταθερότητας και για την επιλογή κατάλληλου μέγιστου συντελεστή προσαρμογής max_alpha .

Για τον υπολογισμό των διαγνωστικών δεικτών σταθερότητας, τυχόν ελλείπουσες τιμές στο παράθυρο ιστορικού συμπληρώνονται προσωρινά με παρεμβολή και εμπροσθο/οπισθοσυμπλήρωση. Η διαδικασία αυτή χρησιμοποιείται μόνο για την εκτίμηση της σταθερότητας του ιστορικού και δεν τροποποιεί το αρχικό αρχείο ιστορικών μετρήσεων.

Η αντιστοίχιση μεταξύ κατηγορίας σταθερότητας και προτεινόμενης τιμής max_alpha πραγματοποιείται με συντηρητικό τρόπο, ώστε το ιστορικό να επηρεάζει περισσότερο την πρόβλεψη μόνο όταν είναι αρκετά σταθερό και επαναλαμβανόμενο. Ο Πίνακας 6 δείχνει ότι η προτεινόμενη τιμή του max_alpha αυξάνεται όσο αυξάνεται η σταθερότητα του ιστορικού. Έτσι, σε χαμηλή ή μέτρια σταθερότητα η διόρθωση παραμένει ήπια, ενώ σε περιπτώσεις πολύ υψηλής ή εξαιρετικής σταθερότητας επιτρέπεται μεγαλύτερη επίδραση του πρόσφατου ιστορικού προφίλ.

Κατηγορία	Συνθήκη	Προτεινόμενο max_alpha	Εύρος
Low	ελλιπές ιστορικό ή $S_{hist} < 0,40$	0,05	0,00–0,10
Medium	$0,40 \leq S_{hist} < 0,60$	0,10	0,10–0,15
High	$0,60 \leq S_{hist} < 0,80$	0,25	0,20–0,30
Very High	$0,80 \leq S_{hist} < 0,93$	0,40	0,35–0,50
Excellent	$S_{hist} \geq 0,93$	0,60	0,55–0,65

Πίνακας 6 Αντιστοίχιση κατηγορίας σταθερότητας ιστορικού και προτεινόμενου max_alpha

Επιπλέον, εφαρμόζεται κανόνας ασφαλείας που προηγείται της τελικής κατάταξης. Αν η πληρότητα του ιστορικού είναι μικρότερη από 0,85 ή αν οι διαθέσιμες ωριαίες μετρήσεις είναι ανεπαρκείς σε σχέση με το ελάχιστο απαιτούμενο ιστορικό, τότε το ιστορικό χαρακτηρίζεται ως χαμηλής σταθερότητας και προτείνεται πολύ χαμηλή τιμή max_alpha .

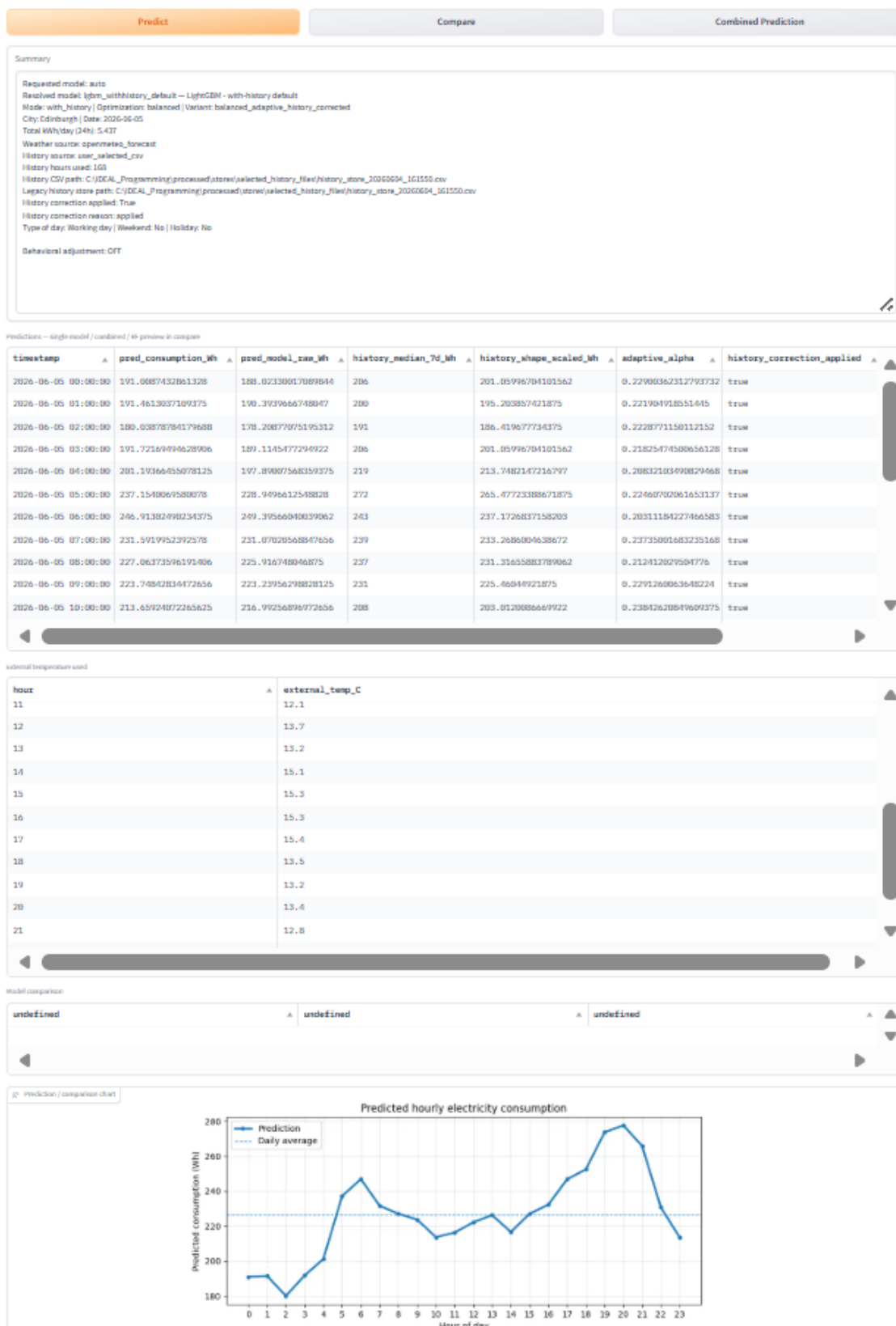
Με αυτόν τον τρόπο αποφεύγεται η έντονη επίδραση ενός ελλιπούς ή μη αξιόπιστου ιστορικού στην τελική πρόβλεψη.

Η λειτουργία του `max_alpha` συνδέεται άμεσα με τον συντελεστή προσαρμογής α που παρουσιάστηκε στην Ενότητα 6.6 κατά την περιγραφή της προσαρμοστικής διόρθωσης. Στο πλαίσιο εκείνης της ανάλυσης, ο συντελεστής α εκφράζει τον βαθμό με τον οποίο η τελική πρόβλεψη μετατοπίζεται από την αρχική πρόβλεψη του μοντέλου προς το προσαρμοσμένο ιστορικό προφίλ της κατοικίας. Στην υλοποίηση του UI/API, η τιμή `max_alpha` δεν εφαρμόζεται απαραίτητα αυτούσια σε κάθε ώρα, αλλά χρησιμοποιείται ως μέγιστο επιτρεπτό όριο του προσαρμοστικού συντελεστή. Με τον τρόπο αυτό, η εφαρμογή διατηρεί τη λογική της μεθόδου που παρουσιάστηκε στο Κεφάλαιο 6, αλλά ταυτόχρονα περιορίζει δυναμικά την επίδραση του ιστορικού όταν η ανάλυση σταθερότητας δείχνει ότι το πρόσφατο ιστορικό δεν είναι αρκετά αξιόπιστο.

Μετά την ολοκλήρωση της ανάλυσης, ο χρήστης μπορεί να προχωρήσει στην παραγωγή πρόβλεψης μέσω της επιλογής `Predict`. Το σύστημα δημιουργεί τα απαιτούμενα χαρακτηριστικά ιστορικού, εφαρμόζει τα αντίστοιχα βήματα προεπεξεργασίας και εκτελεί το επιλεγμένο μοντέλο πρόβλεψης. Εφόσον έχει ενεργοποιηθεί η επιλογή `Adaptive History Correction`, εφαρμόζεται επιπλέον η διαδικασία προσαρμοστικής διόρθωσης που παρουσιάστηκε στην Ενότητα 6.6.

Η λειτουργία αυτή επιτρέπει την παραγωγή εξατομικευμένων προβλέψεων που αξιοποιούν όχι μόνο τα στατικά χαρακτηριστικά της κατοικίας και τις περιβαλλοντικές συνθήκες, αλλά και τη συμπεριφορά κατανάλωσης που καταγράφεται στο πρόσφατο ιστορικό της. Παράλληλα, η ανάλυση σταθερότητας λειτουργεί ως διαγνωστικό εργαλείο, καθώς επιτρέπει στον χρήστη να εκτιμήσει αν το διαθέσιμο ιστορικό είναι αρκετά αξιόπιστο ώστε να επηρεάσει την τελική πρόβλεψη. Η εμφάνιση της μέσης, της διάμεσης και του εύρους ημερήσιας κατανάλωσης ενισχύει περαιτέρω την ερμηνευσιμότητα της πρόβλεψης, καθώς βοηθά να διαπιστωθεί αν η προβλεπόμενη ημέρα κινείται κοντά ή μακριά από το πρόσφατο ημερήσιο επίπεδο κατανάλωσης της κατοικίας.

Ένα ενδεικτικό παράδειγμα εκτέλεσης πρόβλεψης, στο σενάριο με χρήση ιστορικών δεδομένων, παρουσιάζεται στην Εικόνα 7.



Εικόνα 7 Παράδειγμα αποτελεσμάτων πρόβλεψης στο σενάριο με χρήση ιστορικών δεδομένων

8.5 Σύγκριση μοντέλων και συνδυαστικές προβλέψεις

Πέρα από την εκτέλεση πρόβλεψης με ένα μόνο μοντέλο, η εφαρμογή παρέχει τη δυνατότητα ταυτόχρονης σύγκρισης πολλαπλών μοντέλων, καθώς και τη δημιουργία συνδυαστικών προβλέψεων (ensemble predictions). Οι λειτουργίες αυτές επιτρέπουν στον χρήστη να αξιολογήσει τις διαφορές μεταξύ των διαθέσιμων αλγορίθμων και να εξετάσει εναλλακτικές στρατηγικές πρόβλεψης.

Η λειτουργία *Compare Models* εκτελεί διαδοχικά τα διαθέσιμα μοντέλα Random Forest, XGBoost και LightGBM χρησιμοποιώντας ακριβώς τα ίδια δεδομένα εισόδου. Τα αποτελέσματα παρουσιάζονται συγκεντρωτικά, επιτρέποντας την άμεση σύγκριση των προβλεπόμενων καμπυλών κατανάλωσης. Με τον τρόπο αυτό ο χρήστης μπορεί να παρατηρήσει τις διαφορές που προκύπτουν από τις διαφορετικές μεθοδολογίες μάθησης και να αξιολογήσει ποιο μοντέλο ανταποκρίνεται καλύτερα στις απαιτήσεις του συγκεκριμένου σεναρίου.

Παράλληλα, η εφαρμογή υποστηρίζει τη λειτουργία *Combined Prediction*, μέσω της οποίας η τελική πρόβλεψη προκύπτει από συνδυασμό των επιμέρους προβλέψεων των μοντέλων. Η προσέγγιση αυτή βασίζεται στη λογική των συνδυαστικών μοντέλων (ensemble learning), σύμφωνα με την οποία διαφορετικοί αλγόριθμοι μπορούν να συνεισφέρουν συμπληρωματική πληροφορία στη διαδικασία πρόβλεψης.

Ο χρήστης έχει τη δυνατότητα να καθορίζει τα βάρη συμμετοχής των μοντέλων Random Forest, XGBoost και LightGBM. Η τελική πρόβλεψη προκύπτει ως σταθμισμένος μέσος όρος των επιμέρους προβλέψεων:

$$\hat{y}_{ensemble} = w_{RF}\hat{y}_{RF} + w_{XGB}\hat{y}_{XGB} + w_{LGBM}\hat{y}_{LGBM} \quad (12)$$

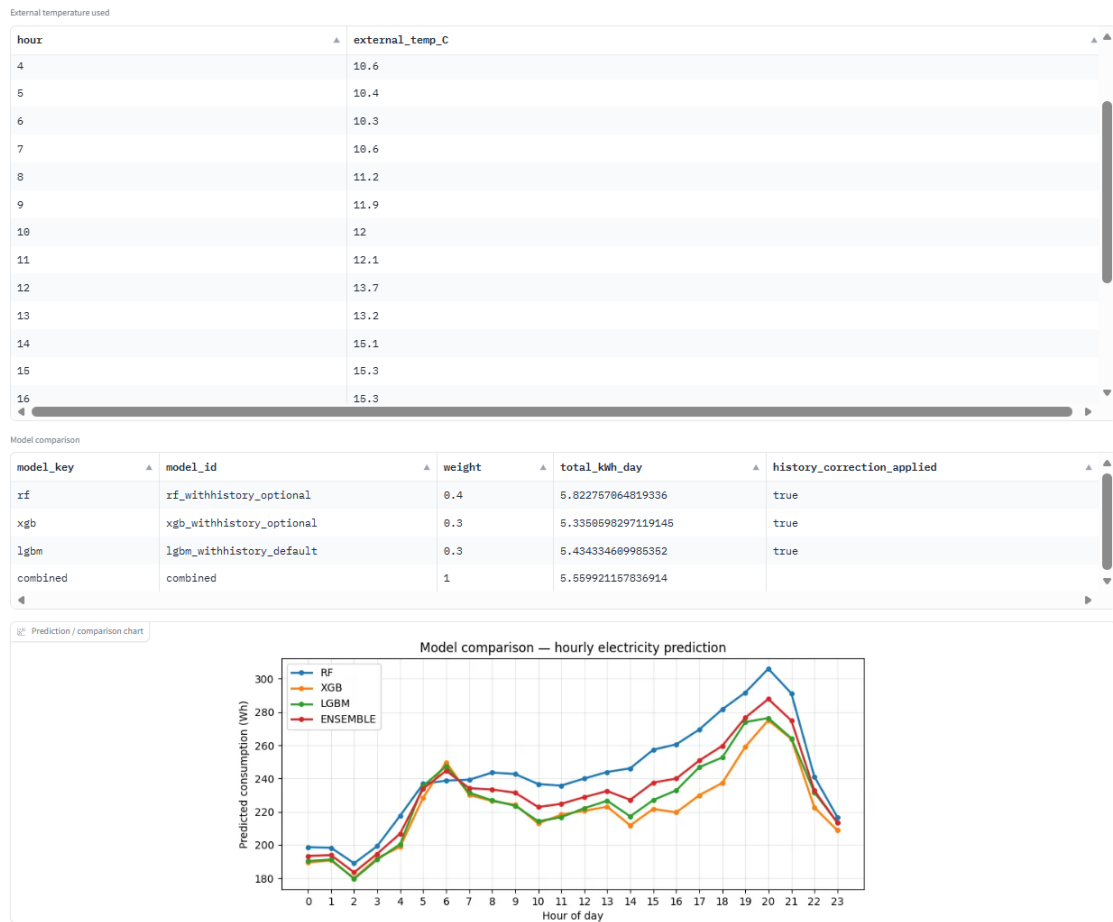
όπου $\hat{y}_{ensemble}$ είναι η τελική πρόβλεψη του συνδυαστικού μοντέλου, $\hat{y}_{RF}$ η πρόβλεψη του μοντέλου Random Forest, $\hat{y}_{XGB}$ η πρόβλεψη του μοντέλου XGBoost και $\hat{y}_{LGBM}$ η πρόβλεψη του μοντέλου LightGBM. Τα w_{RF} , w_{XGB} και w_{LGBM} είναι τα βάρη συμμετοχής των τριών μοντέλων και ισχύει:

$$w_{RF} + w_{XGB} + w_{LGBM} = 1 \quad (13)$$

Ο περιορισμός αυτός εξασφαλίζει ότι το σύνολο των βαρών ισούται με 1, ώστε η τελική πρόβλεψη να αποτελεί σταθμισμένο μέσο όρο των επιμέρους προβλέψεων και να παραμένει στην ίδια κλίμακα με τις αρχικές τιμές κατανάλωσης.

Στην τελική έκδοση της εφαρμογής χρησιμοποιούνται προεπιλεγμένα βάρη τα οποία βασίζονται στα αποτελέσματα αξιολόγησης του Κεφαλαίου 6. Ωστόσο, ο χρήστης μπορεί να τα τροποποιήσει ώστε να εξετάσει διαφορετικούς συνδυασμούς μοντέλων και να προσαρμόσει τη συμπεριφορά του συστήματος στις ανάγκες του.

Τα αποτελέσματα της συνδυαστικής πρόβλεψης παρουσιάζονται τόσο σε μορφή πίνακα όσο και σε μορφή γραφήματος. Στο γράφημα απεικονίζονται ταυτόχρονα οι καμπύλες των επιμέρους μοντέλων και η τελική καμπύλη ensemble, διευκολύνοντας την οπτική σύγκριση των αποτελεσμάτων.



Εικόνα 8 Παράδειγμα συνδυαστικής πρόβλεψης (Combined Prediction) και σύγκρισης μοντέλων

Η δυνατότητα σύγκρισης και συνδυασμού μοντέλων προσφέρει αυξημένη ευελιξία στη χρήση του συστήματος και επιτρέπει την αξιοποίηση των πλεονεκτημάτων διαφορετικών αλγορίθμων μέσα από ένα ενιαίο περιβάλλον. Παράλληλα, παρέχει στον χρήστη τη δυνατότητα διερεύνησης της επίδρασης των επιμέρους μοντέλων στην τελική πρόβλεψη, χωρίς να απαιτείται επανεκπαίδευση ή τροποποίηση του υποκείμενου συστήματος μηχανικής μάθησης.

8.6 Πρόσθετες λειτουργίες προσαρμογής των προβλέψεων

Ορισμένες από τις λειτουργίες που παρουσιάζονται στην παρούσα ενότητα έχουν ήδη αναφερθεί συνοπτικά στις προηγούμενες ενότητες, στο πλαίσιο της βασικής ροής χρήσης της εφαρμογής. Ωστόσο, λόγω της σημασίας τους για την τελική διαμόρφωση των προβλέψεων, κρίνεται σκόπιμο να παρουσιαστούν αναλυτικότερα οι μηχανισμοί λειτουργίας και οι δυνατότητες προσαρμογής που παρέχουν στον χρήστη.

Οι λειτουργίες αυτές δεν απαιτούν επανεκπαίδευση των μοντέλων μηχανικής μάθησης, αλλά εφαρμόζονται κατά τη διαδικασία παραγωγής της πρόβλεψης ή ως στάδιο μεταγενέστερης επεξεργασίας (post-processing) των αποτελεσμάτων. Με τον τρόπο αυτό επιτρέπουν την προσαρμογή των προβλέψεων στις ιδιαίτερες συνθήκες κάθε κατοικίας και στις ανάγκες του εκάστοτε χρήστη.

Μία από τις σημαντικότερες δυνατότητες της εφαρμογής είναι η επιλογή διαφορετικού στόχου βελτιστοποίησης μέσω της παραμέτρου *Optimization Mode*. Ο χρήστης μπορεί να επιλέξει μεταξύ της λειτουργίας *balanced* και της λειτουργίας *daily*.

Η λειτουργία *balanced* αντιστοιχεί στη βασική μορφή λειτουργίας του συστήματος και δίνει προτεραιότητα στη συνολική ακρίβεια των ωριαίων προβλέψεων. Η συγκεκριμένη επιλογή χρησιμοποιεί τα μοντέλα που παρουσίασαν τις καλύτερες συνολικές επιδόσεις στις μετρικές MAE, RMSE, MAPE και sMAPE και αποτελεί την προτεινόμενη επιλογή για τις περισσότερες περιπτώσεις χρήσης.

Αντίθετα, η λειτουργία *daily* δίνει έμφαση στην ακριβέστερη εκτίμηση της συνολικής ημερήσιας κατανάλωσης. Όπως παρουσιάστηκε στα αποτελέσματα του Κεφαλαίου 6, η προσέγγιση αυτή μπορεί να μειώσει σημαντικά το ημερήσιο ενεργειακό σφάλμα, αν και συνήθως συνοδεύεται από ελαφρώς αυξημένα σφάλματα στις επιμέρους ωριαίες προβλέψεις. Ωστόσο, δεδομένου ότι βασικός στόχος της παρούσας εργασίας είναι η όσο το δυνατόν ακριβέστερη πρόβλεψη της ωριαίας καμπύλης κατανάλωσης, η λειτουργία *balanced* έχει επιλεγεί ως προεπιλεγμένη επιλογή του συστήματος.

Για τις προβλέψεις με χρήση ιστορικών δεδομένων παρέχεται επιπλέον η δυνατότητα ενεργοποίησης της λειτουργίας *Adaptive History Correction*. Η λειτουργία αυτή αξιοποιεί το πρόσφατο ιστορικό κατανάλωσης και εφαρμόζει προσαρμοστική διόρθωση της προβλεπόμενης καμπύλης φορτίου, ώστε η τελική πρόβλεψη να αντανακλά καλύτερα τα πιο πρόσφατα πρότυπα κατανάλωσης της συγκεκριμένης κατοικίας.

Πριν από την εφαρμογή της διόρθωσης, το σύστημα μπορεί να αναλύσει το διαθέσιμο ιστορικό κατανάλωσης και να εκτιμήσει τη σταθερότητα των καταναλωτικών προτύπων. Με βάση τα αποτελέσματα της ανάλυσης αυτής υπολογίζεται προτεινόμενη τιμή του συντελεστή προσαρμογής, η οποία καθορίζει τον βαθμό συμμετοχής του ιστορικού προφίλ στην τελική πρόβλεψη.

Παράλληλα, η εφαρμογή παρέχει τη δυνατότητα προσαρμογής της προβλεπόμενης κατανάλωσης μέσω της λειτουργίας *Behavioral Adjustment*. Μέσω της λειτουργίας αυτής ο χρήστης μπορεί να δηλώσει αναμενόμενες μεταβολές στη γενικότερη συμπεριφορά κατανάλωσης κατά συγκεκριμένες χρονικές περιόδους της ημέρας. Η δυνατότητα αυτή είναι ιδιαίτερα χρήσιμη όταν ο χρήστης γνωρίζει εκ των προτέρων ότι η επόμενη ημέρα θα διαφέρει από τη συνήθη λειτουργία της κατοικίας, για παράδειγμα λόγω προγραμματισμένης χρήσης ενεργοβόρων συσκευών ή αυξημένης παρουσίας ενοίκων.

Οι παραπάνω λειτουργίες αυξάνουν σημαντικά την ευελιξία του συστήματος και επιτρέπουν την προσαρμογή των προβλέψεων σε διαφορετικά σενάρια χρήσης. Με τον τρόπο αυτό η εφαρμογή δεν λειτουργεί μόνο ως ένα στατικό εργαλείο πρόβλεψης, αλλά ως ένα διαδραστικό σύστημα υποστήριξης αποφάσεων που μπορεί να προσαρμόζεται στις ιδιαίτερες απαιτήσεις κάθε σεναρίου πρόβλεψης.

8.7 Χειρισμός εξωτερικής θερμοκρασίας

Η εξωτερική θερμοκρασία αποτελεί κρίσιμη είσοδο για τα μοντέλα πρόβλεψης επόμενης ημέρας, καθώς επηρεάζει άμεσα τη ζήτηση για θέρμανση ή ψύξη και, κατά συνέπεια, τη συνολική ηλεκτρική κατανάλωση μιας κατοικίας. Για τον λόγο αυτό, η εφαρμογή παρέχει τη δυνατότητα ενσωμάτωσης μετεωρολογικών δεδομένων είτε μέσω αυτόματης ανάκτησης είτε μέσω εισαγωγής δεδομένων από τον χρήστη.

Κατά την εκτέλεση μιας πρόβλεψης, το σύστημα επιχειρεί αρχικά να ανακτήσει αυτόματα ωριαίες προβλέψεις εξωτερικής θερμοκρασίας από την υπηρεσία Open-Meteo (open-meteo.com), για την επιλεγμένη πόλη και την ημέρα στόχο. Οι τιμές αυτές χρησιμοποιούνται απευθείας ως περιβαλλοντικά χαρακτηριστικά εισόδου στα μοντέλα μηχανικής μάθησης.

Σε περιπτώσεις όπου η αυτόματη ανάκτηση των δεδομένων δεν είναι εφικτή, ενεργοποιείται μηχανισμός εφεδρείας (fallback) που βασίζεται στην ελάχιστη και μέγιστη ημερήσια θερμοκρασία που εισάγει ο χρήστης. Στην περίπτωση αυτή, το σύστημα δημιουργεί συνθετική ωριαία καμπύλη θερμοκρασίας 24 σημείων, η οποία προσεγγίζει τη φυσική ημερήσια διακύμανση της θερμοκρασίας.

Η καμπύλη αυτή παράγεται με χρήση ημιτονοειδούς συνάρτησης της μορφής:

$$T(h) = T_{\min} + (T_{\max} - T_{\min}) \cdot \frac{1 + \sin\left(\frac{2\pi(h - \phi)}{24}\right)}{2}, h = 0, \dots, 23 \quad (14)$$

όπου $T(h)$ είναι η εκτιμώμενη θερμοκρασία για την ώρα h , $T_{\min}$ η ελάχιστη ημερήσια θερμοκρασία, $T_{\max}$ η μέγιστη ημερήσια θερμοκρασία και ϕ παράμετρος χρονικής μετατόπισης που καθορίζει τη χρονική στιγμή εμφάνισης της μέγιστης θερμοκρασίας κατά τη διάρκεια της ημέρας (συνήθως μεταξύ 14:00 και 16:00).

Η συγκεκριμένη προσέγγιση εξασφαλίζει ομαλή μεταβολή της θερμοκρασίας κατά τη διάρκεια του εικοσιτετραώρου, χωρίς απότομα άλματα ή μη ρεαλιστικές μεταβολές. Επιπλέον, όλες οι παραγόμενες τιμές παραμένουν εντός του διαστήματος ($T_{\min}$, $T_{\max}$), γεγονός που καθιστά τη συνθετική καμπύλη κατάλληλη για χρήση από τα μοντέλα μηχανικής μάθησης.

Ωστόσο, πρέπει να επισημανθεί ότι η μέθοδος αυτή παράγει συνθετικές και όχι πραγματικές ωριαίες θερμοκρασίες, καθώς δεν λαμβάνει υπόψη μετεωρολογικά φαινόμενα όπως η νεφοκάλυψη, οι βροχοπτώσεις ή οι απότομες μεταβολές των καιρικών συνθηκών. Για τον λόγο αυτό, η χρήση πραγματικών ωριαίων προβλέψεων από την υπηρεσία Open-Meteo αποτελεί την προτιμώμενη επιλογή όταν αυτές είναι διαθέσιμες, ενώ η συνθετική καμπύλη λειτουργεί ως πρακτική εναλλακτική λύση για τη διασφάλιση της λειτουργικότητας του συστήματος.

Η διαδικασία παραγωγής της συνθετικής καμπύλης εκτελείται αυτόματα από το API και δεν απαιτεί καμία επιπλέον ενέργεια από τον χρήστη πέρα από την εισαγωγή των τιμών ($T_{\min}$) και ($T_{\max}$).

Τέλος, μετά από κάθε επιτυχή εκτέλεση πρόβλεψης, η εφαρμογή εμφανίζει στη σύνοψη αποτελεσμάτων την πηγή των μετεωρολογικών δεδομένων (weather source), ώστε να είναι σαφές αν χρησιμοποιήθηκαν πραγματικές ωριαίες προβλέψεις από το Open-Meteo ή αν ενεργοποιήθηκε ο μηχανισμός εναλλακτικής εκτίμησης μέσω της συνθετικής καμπύλης θερμοκρασίας.

9. Αξιολόγηση του συστήματος πρόβλεψης

Στο παρόν κεφάλαιο παρουσιάζεται η αξιολόγηση του συστήματος πρόβλεψης ηλεκτρικής κατανάλωσης που αναπτύχθηκε στο πλαίσιο της εργασίας. Σε αντίθεση με την αξιολόγηση των βασικών μοντέλων του Κεφαλαίου 6, η οποία πραγματοποιήθηκε σε επίπεδο συνολικού συνόλου δεδομένων, η παρούσα ανάλυση επικεντρώνεται σε επιλεγμένες κατοικίες και συγκεκριμένες ημέρες πρόβλεψης.

Στόχος της αξιολόγησης είναι να εξεταστεί η συμπεριφορά του συστήματος σε ρεαλιστικά σενάρια χρήσης, όπως αυτά υποστηρίζονται από τη διεπαφή χρήστη και το API. Για τον σκοπό αυτό αξιολογούνται αρχικά τρεις βασικές εκδοχές λειτουργίας: το σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης, το σενάριο με χρήση ιστορικών δεδομένων χωρίς πρόσθετη διόρθωση και το σενάριο με χρήση ιστορικών δεδομένων μετά την εφαρμογή προσαρμοστικής διόρθωσης βάσει ιστορικού κατανάλωσης (with-history adaptive).

Η αξιολόγηση πραγματοποιείται σε δύο επίπεδα. Αρχικά εξετάζονται τέσσερις επιλεγμένες κατοικίες από το ανεξάρτητο σύνολο ελέγχου του IDEAL dataset, ώστε να διερευνηθεί η απόδοση του συστήματος σε διαφορετικά προφίλ κατανάλωσης. Στη συνέχεια παρουσιάζονται τέσσερις πρόσθετες εποχικές μελέτες περίπτωσης, μία για κάθε επιλεγμένη κατοικία. Οι περιπτώσεις αυτές επιτρέπουν την αξιολόγηση του συστήματος σε διαφορετικές χρονικές περιόδους του έτους και την εξέταση της επίδρασης της σταθερότητας του ιστορικού κατανάλωσης στην ποιότητα των προβλέψεων. Στις εποχικές μελέτες περίπτωσης παραλείπεται στην αξιολόγηση το σενάριο με χρήση ιστορικών δεδομένων χωρίς πρόσθετη διόρθωση και εξετάζονται τα υπόλοιπα δύο σενάρια.

Τα αποτελέσματα που ακολουθούν επιτρέπουν την πληρέστερη κατανόηση της πρακτικής συμπεριφοράς του συστήματος και αναδεικνύουν τις συνθήκες υπό τις οποίες η αξιοποίηση ιστορικών δεδομένων μπορεί να βελτιώσει ή να περιορίσει την ακρίβεια των προβλέψεων. Ιδιαίτερη έμφαση δίνεται στη διάκριση μεταξύ της απλής χρήσης ιστορικών χαρακτηριστικών και της μεταγενέστερης προσαρμοστικής διόρθωσης της προβλεπόμενης καμπύλης.

Οι κώδικες σε γλώσσα Python που χρησιμοποιήθηκαν για την αξιολόγηση των αποτελεσμάτων βρίσκονται στον φάκελο IDEAL/Evaluation του δημόσιου αποθετηρίου GitHub, στον ακόλουθο σύνδεσμο: <https://github.com/stdGeorgios/Forecasting-the-electricity-consumption-of-buildings->.

9.1 Μεθοδολογία αξιολόγησης

Η αξιολόγηση του συστήματος πραγματοποιήθηκε με στόχο να προσομοιωθεί όσο το δυνατόν πιστότερα η πραγματική λειτουργία της εφαρμογής που παρουσιάστηκε στα Κεφάλαια 7 και 8. Για τον λόγο αυτό, οι προβλέψεις δεν αξιολογήθηκαν μόνο σε επίπεδο ολόκληρου συνόλου δεδομένων, αλλά και σε επίπεδο μεμονωμένων κατοικιών, ακολουθώντας τη διαδικασία που θα ακολουθούσε ένας πραγματικός χρήστης μέσω της διεπαφής χρήστη ή του API.

Η αξιολόγηση βασίστηκε στο IDEAL dataset και πραγματοποιήθηκε σε δύο διακριτές φάσεις. Στην πρώτη φάση επιλέχθηκαν τέσσερις κατοικίες από το ανεξάρτητο σύνολο ελέγχου του Ιουνίου 2018. Οι κατοικίες αυτές χρησιμοποιήθηκαν για τη σύγκριση των διαφορετικών σεναρίων πρόβλεψης σε δεδομένα που δεν είχαν χρησιμοποιηθεί κατά την εκπαίδευση των μοντέλων. Στη δεύτερη φάση πραγματοποιήθηκαν τέσσερις επιπλέον

εποχικές μελέτες περίπτωσης, οι οποίες αφορούσαν διαφορετικές εποχές του έτους και χρησιμοποιήθηκαν για τη διερεύνηση της επίδρασης της σταθερότητας του ιστορικού κατανάλωσης στην απόδοση των προβλέψεων.

Για κάθε κατοικία εξετάστηκαν τρεις εκδοχές πρόβλεψης. Η πρώτη εκδοχή αφορά το σενάριο χωρίς χρήση ιστορικών δεδομένων, στο οποίο τα μοντέλα αξιοποιούν αποκλειστικά τα χαρακτηριστικά της κατοικίας, τα χρονικά χαρακτηριστικά και τις διαθέσιμες περιβαλλοντικές πληροφορίες. Η δεύτερη εκδοχή αφορά το απλό σενάριο με χρήση ιστορικών δεδομένων, στο οποίο χρησιμοποιούνται επιπλέον πρόσφατες μετρήσεις κατανάλωσης της ίδιας κατοικίας για τη δημιουργία χαρακτηριστικών χρονικής εξάρτησης. Η τρίτη εκδοχή αφορά το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, όπου μετά την αρχική πρόβλεψη εφαρμόζεται πρόσθετο στάδιο μεταγενέστερης επεξεργασίας της καμπύλης με βάση τη σταθερότητα του πρόσφατου ιστορικού.

Για την υποστήριξη του σεναρίου με χρήση ιστορικών δεδομένων δημιουργήθηκαν ειδικά αρχεία ιστορικού κατανάλωσης για κάθε κατοικία. Τα αρχεία αυτά περιείχαν πραγματικές μετρήσεις κατανάλωσης που προηγούνταν χρονικά της ημέρας πρόβλεψης και χρησιμοποιήθηκαν ως είσοδος στη διεπαφή χρήστη και στο API. Με τον τρόπο αυτό εξασφαλίστηκε ότι δεν υπήρχε διαρροή πληροφορίας από την ημέρα στόχο προς τη διαδικασία πρόβλεψης.

Πριν από την εκτέλεση των σεναρίων με χρήση ιστορικών δεδομένων πραγματοποιήθηκε, για κάθε κατοικία, ανάλυση σταθερότητας του διαθέσιμου ιστορικού μέσω της λειτουργίας Analyze History Stability της διεπαφής. Η διαδικασία αυτή, η οποία παρουσιάστηκε αναλυτικά στην Ενότητα 8.4, παρείχε πληροφορίες σχετικά με την πληρότητα και τη σταθερότητα του πρόσφατου ιστορικού, καθώς και προτεινόμενη τιμή του μέγιστου συντελεστή προσαρμογής $\max_alpha$. Επιπλέον, εμφανίζονταν η μέση ημερήσια κατανάλωση, η διάμεση ημερήσια κατανάλωση και το εύρος των ημερήσιων τιμών στο εξεταζόμενο ιστορικό. Η πληροφορία αυτή χρησιμοποιήθηκε τόσο για την εφαρμογή της προσαρμοστικής διόρθωσης όσο και για την ερμηνεία των αποτελεσμάτων ανά κατοικία.

Για την περιβαλλοντική είσοδο των μοντέλων χρησιμοποιήθηκαν οι πραγματικές 24 ωριαίες τιμές εξωτερικής θερμοκρασίας της ημέρας στόχου, όπως αυτές ήταν διαθέσιμες στο IDEAL dataset. Η επιλογή αυτή έγινε ώστε η αξιολόγηση να βασίζεται όσο το δυνατόν περισσότερο σε ορθές περιβαλλοντικές πληροφορίες και να μην επηρεάζεται από σφάλματα εξωτερικών μετεωρολογικών προβλέψεων ή από τη συνθετική καμπύλη θερμοκρασίας του μηχανισμού fallback. Οι τιμές αυτές δεν περιλαμβάνουν πληροφορία κατανάλωσης και επομένως δεν προκαλούν διαρροή της μεταβλητής στόχου, αλλά χρησιμοποιούνται ως εξωγενής μεταβλητή εισόδου. Σε πραγματική μελλοντική χρήση της εφαρμογής, οι αντίστοιχες τιμές θα προέρχονται από μετεωρολογική πρόβλεψη ή από τις διαθέσιμες τιμές που εισάγει ο χρήστης.

Οι προβλέψεις πραγματοποιήθηκαν με χρήση των μοντέλων Random Forest, XGBoost και LightGBM, καθώς και με χρήση συνδυαστικών προβλέψεων. Για το σενάριο χωρίς χρήση ιστορικών δεδομένων χρησιμοποιήθηκαν βάρη 0,40 για το Random Forest και 0,30 για τα XGBoost και LightGBM. Για τα σενάρια με χρήση ιστορικών δεδομένων χρησιμοποιήθηκαν βάρη 0,10 για το Random Forest και 0,45 για τα XGBoost και LightGBM.

Η αξιολόγηση πραγματοποιήθηκε με τις μετρικές που παρουσιάστηκαν στο Κεφάλαιο 6, δηλαδή MAE, RMSE, MAPE και Ημερήσιο Σφάλμα (Daily Absolute Error). Οι μετρικές αυτές επιτρέπουν την ταυτόχρονη αξιολόγηση τόσο της ακρίβειας των ωριαίων προβλέψεων όσο και της ικανότητας εκτίμησης της συνολικής ημερήσιας κατανάλωσης.

Στις επόμενες ενότητες παρουσιάζονται αναλυτικά τα αποτελέσματα των επιλεγμένων κατοικιών και των εποχικών μελετών περίπτωσης, καθώς και τα βασικά συμπεράσματα που

προκύπτουν σχετικά με τη χρήση ιστορικών δεδομένων στην πρόβλεψη ηλεκτρικής κατανάλωσης κατοικιών.

9.2 Αξιολόγηση επιλεγμένων κατοικιών του συνόλου ελέγχου

Η κύρια αξιολόγηση του συστήματος πραγματοποιήθηκε σε τέσσερις επιλεγμένες κατοικίες του ανεξάρτητου συνόλου ελέγχου του IDEAL dataset. Οι ημερομηνίες αξιολόγησης ανήκουν στον Ιούνιο του 2018, δηλαδή στην περίοδο που χρησιμοποιήθηκε ως test set κατά την εκπαίδευση και αξιολόγηση των μοντέλων. Με τον τρόπο αυτό, οι προβλέψεις εξετάζονται σε δεδομένα που δεν είχαν χρησιμοποιηθεί κατά την εκπαίδευση των μοντέλων.

Στην κύρια αξιολόγηση των κατοικιών του συνόλου ελέγχου εξετάστηκαν τρεις εκδοχές πρόβλεψης. Η πρώτη είναι το σενάριο χωρίς ιστορικά δεδομένα κατανάλωσης, στο οποίο τα μοντέλα χρησιμοποιούν μόνο χρονικά, στατικά και περιβαλλοντικά χαρακτηριστικά. Η δεύτερη είναι το απλό σενάριο με χρήση ιστορικών δεδομένων, στο οποίο αξιοποιείται πρόσφατο ιστορικό κατανάλωσης 168 ωρών πριν από την ημέρα πρόβλεψης, χωρίς την εφαρμογή πρόσθετης προσαρμοστικής διόρθωσης. Η τρίτη είναι το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, στο οποίο εφαρμόζεται επιπλέον η τελική μορφή της προσαρμοστικής διόρθωσης βάσει ιστορικού κατανάλωσης, όπως αυτή ενσωματώθηκε στο σύστημα.

Η αξιολόγηση πραγματοποιήθηκε για τα μοντέλα Random Forest, XGBoost, LightGBM και για το αντίστοιχο ensemble. Για κάθε περίπτωση υπολογίστηκαν οι μετρικές MAE, RMSE και MAPE, ενώ επιπλέον εξετάστηκε η συνολική προβλεπόμενη ημερήσια κατανάλωση σε σχέση με την πραγματική ημερήσια κατανάλωση. Στους πίνακες που ακολουθούν παρουσιάζονται τα αποτελέσματα ανά κατοικία και ανά σενάριο πρόβλεψης. Στους πίνακες αυτούς, η στήλη pred_daily αναφέρεται στην προβλεπόμενη συνολική ημερήσια κατανάλωση σε kWh, ενώ η στήλη real_daily αναφέρεται στην πραγματική συνολική ημερήσια κατανάλωση σε kWh.

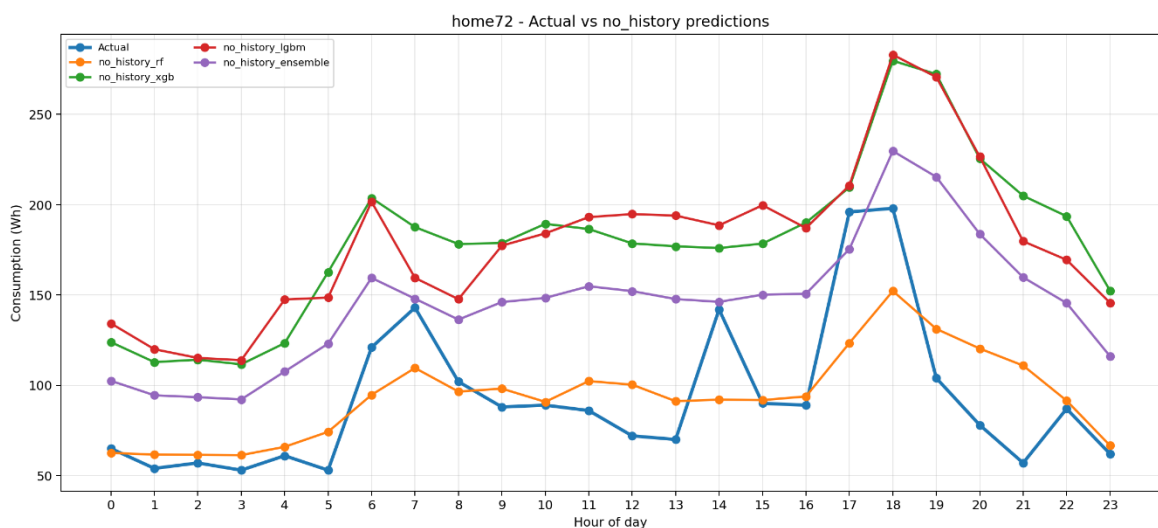
9.2.1 Αποτελέσματα για την κατοικία home72

Η πρώτη κατοικία που εξετάστηκε είναι η home72, με ημέρα αξιολόγησης την 15η Ιουνίου 2018. Η κατοικία αντιστοιχεί σε διαμέρισμα με δύο ενοίκους, συνολικό εμβαδόν 80 m² και 9 καταγεγραμμένες ηλεκτρικές συσκευές. Ως προς την κατηγορία αστικότητας, η κατοικία ανήκει στην κατηγορία 1, η οποία αντιστοιχεί σε μεγάλη αστική περιοχή. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 2,217 kWh. Η κατοικία αυτή αποτελεί περίπτωση χαμηλής ημερήσιας κατανάλωσης, γεγονός που καθιστά την ποσοστιαία αξιολόγηση πιο ευαίσθητη σε μικρές απόλυτες αποκλίσεις. Στον Πίνακα 7 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	20,82	28,48	22,20	2,245	2,217	0,028
XGBoost	87,24	94,20	112,09	4,311	2,217	2,094
LightGBM	86,48	93,68	111,76	4,292	2,217	2,075
Ensemble	54,29	60,95	71,92	3,479	2,217	1,262

Πίνακας 7 Απόδοση μοντέλων για την κατοικία home72 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το Random Forest εμφανίζει καθαρά την καλύτερη απόδοση για την κατοικία home72, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Επιπλέον, η συνολική ημερήσια κατανάλωση που προβλέπει το Random Forest είναι πολύ κοντά στην πραγματική τιμή, με ημερήσιο σφάλμα μόλις 0,028 kWh. Αντίθετα, τα XGBoost και LightGBM υπερεκτιμούν σημαντικά την κατανάλωση, γεγονός που αποτυπώνεται τόσο στις ωριαίες μετρικές όσο και στη συνολική ημερήσια πρόβλεψη.



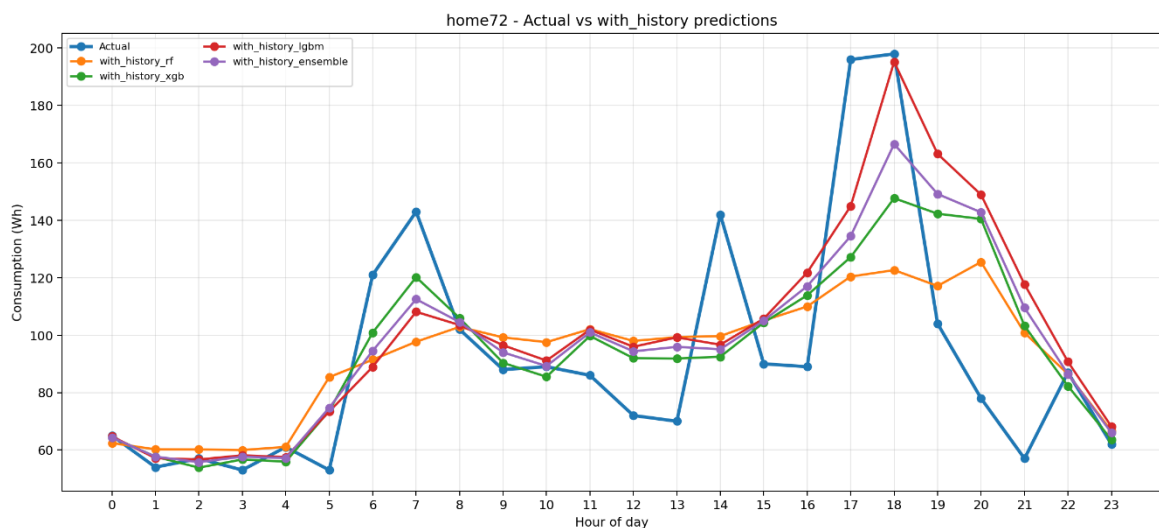
Σχήμα 10 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων, χωρίς την εφαρμογή της προσαρμοστικής διόρθωσης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 8.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	23,22	31,81	24,03	2,232	2,217	0,015
XGBoost	21,10	29,29	22,12	2,266	2,217	0,049
LightGBM	22,05	30,78	25,03	2,406	2,217	0,189
Ensemble	21,44	29,29	23,27	2,326	2,217	0,109

Πίνακας 8 Απόδοση μοντέλων για την κατοικία home72 στο απλό σενάριο με χρήση ιστορικών δεδομένων

Στο απλό σενάριο with-history, τα μοντέλα XGBoost και Ensemble παρουσιάζουν σχεδόν ταυτόσημη απόδοση ως προς το RMSE. Ωστόσο, το XGBoost εμφανίζει χαμηλότερο MAE και MAPE, ενώ παράλληλα πετυχαίνει και μικρότερο ημερήσιο σφάλμα σε σχέση με το Ensemble. Για τον λόγο αυτό, το XGBoost θεωρείται το καλύτερο συνολικά μοντέλο στο απλό σενάριο με χρήση ιστορικών δεδομένων για την κατοικία home72. Το Random Forest, αν και δεν πετυχαίνει το χαμηλότερο ωριαίο σφάλμα, παρουσιάζει το μικρότερο ημερήσιο σφάλμα, με απόκλιση μόλις 0,015 kWh από την πραγματική ημερήσια κατανάλωση.



Σχήμα 11 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στο απλό σενάριο με χρήση ιστορικών δεδομένων

Η ανάλυση σταθερότητας έδειξε μέτρια σταθερότητα ιστορικού. Το διαθέσιμο ιστορικό ήταν πλήρες, καθώς περιλάμβανε και τις 168 αναμενόμενες ωριαίες εγγραφές χωρίς ελλείψεις. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 2,680 kWh, η διάμεση τιμή ήταν 2,397 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 1,819–4,093 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 2,217 kWh, βρισκόταν εντός του ιστορικού εύρους και σχετικά κοντά στη διάμεση τιμή. Ωστόσο, οι δείκτες μεταβλητότητας του ημερήσιου συνόλου, του ωριαίου προφίλ και των αιχμών έδειξαν ότι το πρόσφατο ιστορικό δεν παρουσίαζε πολύ υψηλή επαναληψιμότητα. Με βάση την ανάλυση αυτή, το σύστημα πρότεινε χαμηλό εύρος τιμών για τον συντελεστή προσαρμογής (max_alpha),

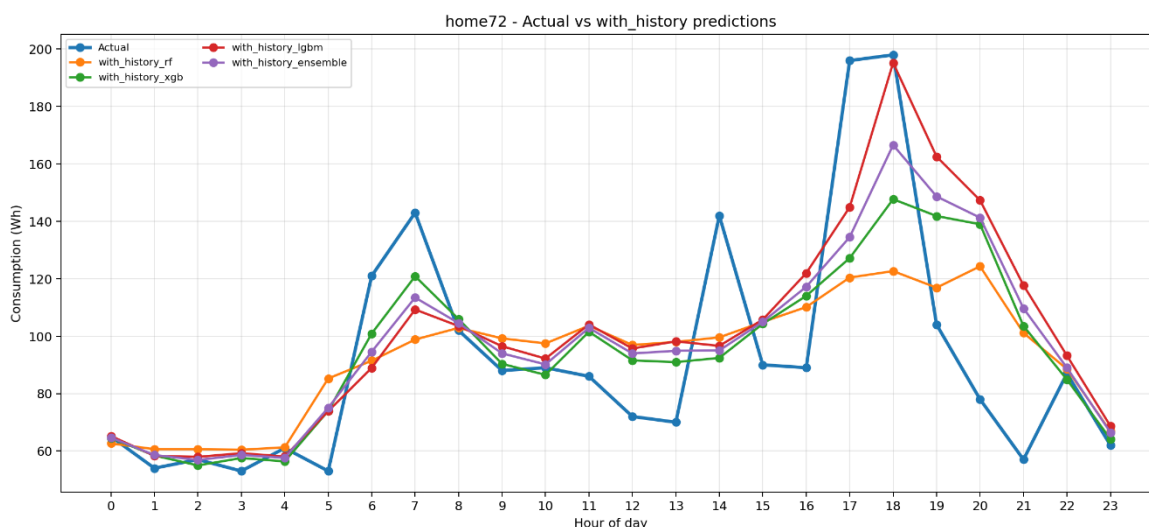
συγκεκριμένα 0,1 ώστε να περιοριστεί ο κίνδυνος υπερπροσαρμογής στο πρόσφατο ιστορικό σχήμα κατανάλωσης.

Στον Πίνακα 9 παρουσιάζονται τα αποτελέσματα του σεναρίου με χρήση ιστορικών δεδομένων μετά την εφαρμογή της προσαρμοστικής διόρθωσης βάσει ιστορικού κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	23,19	31,65	24,03	2,235	2,217	0,018
XGBoost	20,91	29,13	21,93	2,273	2,217	0,056
LightGBM	22,24	30,58	25,35	2,415	2,217	0,198
Ensemble	21,49	29,12	23,35	2,333	2,217	0,116

Πίνακας 9 Απόδοση μοντέλων για την κατοικία home72 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Μετά την εφαρμογή της προσαρμοστικής διόρθωσης, τα μοντέλα XGBoost και Ensemble εξακολουθούν να παρουσιάζουν πολύ κοντινή απόδοση ως προς το RMSE. Ωστόσο, το XGBoost εμφανίζει χαμηλότερο MAE και MAPE, ενώ παράλληλα διατηρεί μικρότερο ημερήσιο σφάλμα σε σχέση με το Ensemble. Για τον λόγο αυτό, το XGBoost θεωρείται το καλύτερο συνολικά μοντέλο στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση για την κατοικία home72. Η επίδραση της προσαρμοστικής διόρθωσης είναι περιορισμένη, καθώς το RMSE του XGBoost μειώνεται οριακά από 29,29 Wh στο απλό σενάριο με χρήση ιστορικών δεδομένων, σε 29,13 Wh μετά την εφαρμογή της προσαρμοστικής διόρθωσης. Παράλληλα, το ημερήσιο σφάλμα αυξάνεται ελαφρά από 0,049 kWh σε 0,056 kWh, γεγονός που δείχνει ότι η διόρθωση βελτίωσε ελαφρά την ωριαία ακρίβεια, χωρίς όμως να βελτιώσει την εκτίμηση της συνολικής ημερήσιας κατανάλωσης.



Σχήμα 12 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των τριών εκδοχών παρουσιάζεται στον Πίνακα 10, λαμβάνοντας υπόψη το μοντέλο με την καλύτερη συνολική ισορροπία μεταξύ RMSE, MAE, MAPE και ημερήσιου σφάλματος.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	RF	20,82	28,48	22,20	2,245	2,217	0,028
with-history	XGB	21,10	29,29	22,12	2,266	2,217	0,049
with-history adaptive	XGB	20,91	29,13	21,93	2,273	2,217	0,056

Πίνακας 10 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home72

Για την κατοικία home72, το σενάριο χωρίς χρήση ιστορικών δεδομένων εμφανίζει οριακά καλύτερη απόδοση από τα σενάρια με χρήση ιστορικών δεδομένων. Η διαφορά, ωστόσο, είναι πολύ μικρή, καθώς το RMSE του Random Forest στο σενάριο χωρίς χρήση ιστορικών δεδομένων είναι 28,48 Wh, ενώ το RMSE του XGBoost στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση είναι 29,13 Wh. Επομένως, η χρήση ιστορικών δεδομένων δεν προσέφερε ουσιαστική βελτίωση στη συγκεκριμένη περίπτωση, χωρίς όμως να προκαλέσει σημαντική επιδείνωση της πρόβλεψης.

Η αξιολόγηση της συγκεκριμένης κατοικίας δείχνει ότι η χρήση πρόσφατου ιστορικού δεν βελτιώνει απαραίτητα την πρόβλεψη όταν το ιστορικό παρουσιάζει μόνο μέτρια σταθερότητα. Σε τέτοιες περιπτώσεις, η πρόβλεψη χωρίς ιστορικά δεδομένα μπορεί να παραμείνει ανταγωνιστική ή ακόμη και οριακά καλύτερη, ιδιαίτερα όταν τα στατικά, χρονικά και περιβαλλοντικά χαρακτηριστικά αποτυπώνουν επαρκώς το επίπεδο κατανάλωσης της ημέρας.

Παράλληλα, η πολύ χαμηλή τιμή του ημερήσιου σφάλματος στο σενάριο χωρίς χρήση ιστορικών δεδομένων δείχνει ότι το Random Forest κατάφερε να εκτιμήσει με μεγάλη ακρίβεια το συνολικό επίπεδο κατανάλωσης της ημέρας, ακόμη και χωρίς πρόσφατο ιστορικό κατανάλωσης. Η ανάλυση ιστορικού έδειξε ότι η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν εντός του εύρους των πρόσφατων ημερήσιων τιμών και κοντά στη διάμεση ημερήσια κατανάλωση. Ωστόσο, το ιστορικό χαρακτηρίστηκε μόνο μέτριας σταθερότητας. Έτσι, η περιορισμένη επίδραση της προσαρμοστικής διόρθωσης είναι αναμενόμενη, καθώς το σύστημα απέφυγε να μεταβάλει έντονα την αρχική πρόβλεψη βάσει ενός ιστορικού που δεν παρουσίαζε πολύ υψηλή επαναληψιμότητα.

9.2.2 Αποτελέσματα για την κατοικία home100

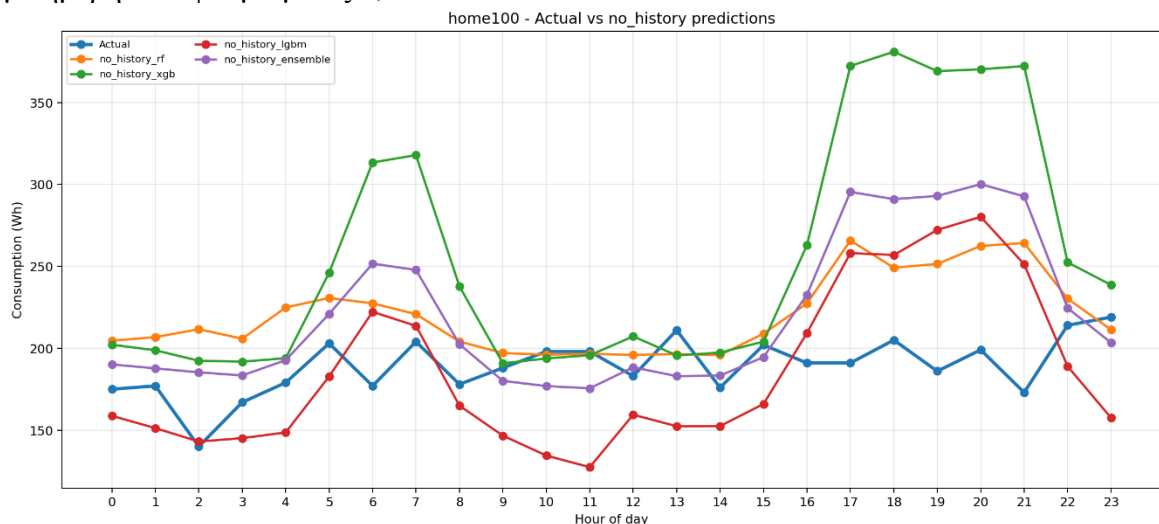
Η δεύτερη κατοικία που εξετάστηκε είναι η home100, με ημέρα αξιολόγησης την 28η Ιουνίου 2018. Η κατοικία αντιστοιχεί σε διαμέρισμα με έναν ένοικο, συνολικό εμβαδόν 72,5 m² και 11 καταγεγραμμένες ηλεκτρικές συσκευές. Ως προς την κατηγορία αστικότητας, η κατοικία ανήκει στην κατηγορία 3+, η οποία αντιστοιχεί σε πιο απομακρυσμένη μικρή πόλη ή αγροτική περιοχή. Η πραγματική συνολική ημερήσια

κατανάλωση της συγκεκριμένης ημέρας ήταν 4,534 kWh. Στον Πίνακα 11 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	33,46	41,58	18,53	5,285	4,534	0,751
XGBoost	66,98	94,99	35,82	6,098	4,534	1,564
LightGBM	40,50	47,31	21,14	4,444	4,534	0,090
Ensemble	39,49	53,85	21,21	5,277	4,534	0,743

Πίνακας 11 Απόδοση μοντέλων για την κατοικία home100 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το Random Forest εμφανίζει την καλύτερη απόδοση ως προς τις βασικές ωριαίες μετρικές, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του Random Forest είναι 41,58 Wh, ενώ το MAPE ανέρχεται σε 18,53%. Ωστόσο, το μοντέλο υπερεκτιμά τη συνολική ημερήσια κατανάλωση, καθώς προβλέπει 5,285 kWh έναντι πραγματικής τιμής 4,534 kWh, με ημερήσιο σφάλμα 0,751 kWh. Αντίθετα, το LightGBM, παρότι δεν παρουσιάζει το χαμηλότερο RMSE, εκτιμά με μεγαλύτερη ακρίβεια τη συνολική ημερήσια κατανάλωση, με ημερήσιο σφάλμα μόλις 0,090 kWh.



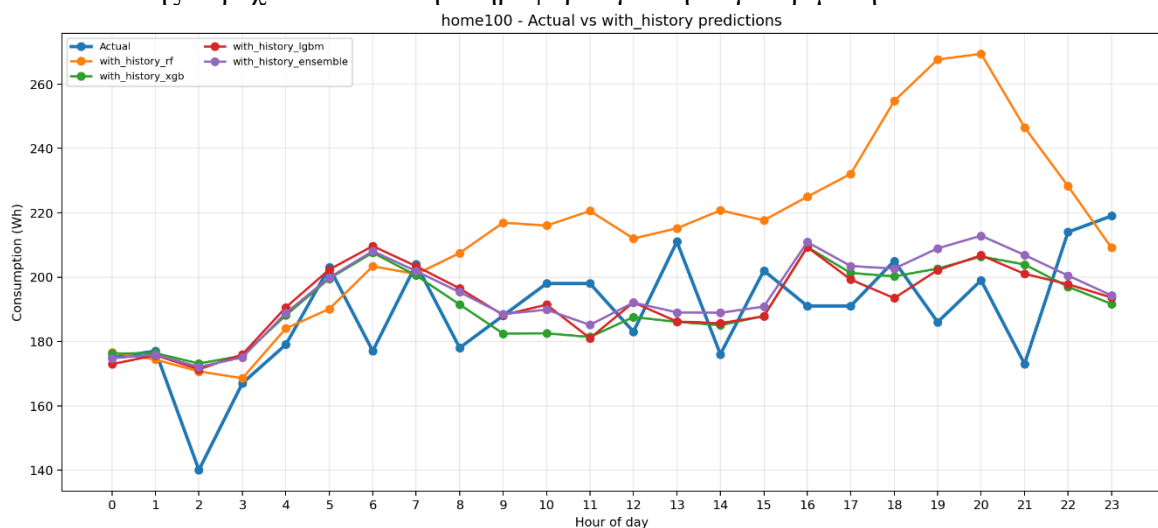
Σχήμα 13 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων, χωρίς την εφαρμογή της προσαρμοστικής διόρθωσης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 12.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	27,12	35,44	14,62	5,128	4,534	0,594
XGBoost	13,61	16,65	7,38	4,594	4,534	0,060
LightGBM	13,37	16,44	7,26	4,614	4,534	0,080
Ensemble	13,55	16,80	7,40	4,656	4,534	0,122

Πίνακας 12 Απόδοση μοντέλων για την κατοικία home100 στο απλό σενάριο με χρήση ιστορικών δεδομένων

Στο απλό σενάριο με χρήση ιστορικών δεδομένων, η εισαγωγή ιστορικών δεδομένων οδηγεί σε σαφή βελτίωση της απόδοσης όλων των μοντέλων σε σχέση με το σενάριο χωρίς χρήση ιστορικών δεδομένων. Το LightGBM παρουσιάζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE μειώνεται σε 16,44 Wh και το MAPE σε 7,26%. Παράλληλα, το XGBoost παρουσιάζει το μικρότερο ημερήσιο σφάλμα, με απόκλιση μόλις 0,060 kWh από την πραγματική ημερήσια κατανάλωση. Η σημαντική μείωση των σφαλμάτων δείχνει ότι το πρόσφατο ιστορικό κατανάλωσης παρέχει ουσιαστική πληροφορία για τη συγκεκριμένη κατοικία.



Σχήμα 14 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στο απλό σενάριο με χρήση ιστορικών δεδομένων

Η ανάλυση σταθερότητας έδειξε υψηλή σταθερότητα ιστορικού. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 5,031 kWh, η διάμεση τιμή ήταν 4,949 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 4,503–6,446 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 4,534 kWh, βρισκόταν εντός του ιστορικού εύρους και πολύ κοντά στη διάμεση τιμή. Επιπλέον, οι δείκτες μεταβλητότητας έδειξαν χαμηλή μεταβλητότητα τόσο στη συνολική ημερήσια κατανάλωση όσο και στο ωριαίο προφίλ, ενώ η σταθερότητα των αιχμών ήταν πιο περιορισμένη. Με βάση την ανάλυση αυτή, το σύστημα πρότεινε μέτριο εύρος τιμών για τον συντελεστή προσαρμογής max_alpha, συγκεκριμένα 0,25 καθώς το ιστορικό χαρακτηρίστηκε σταθερό και κατάλληλο για μέτρια προσαρμοστική διόρθωση. Στον Πίνακα 13 παρουσιάζονται τα αποτελέσματα του σεναρίου με χρήση ιστορικών

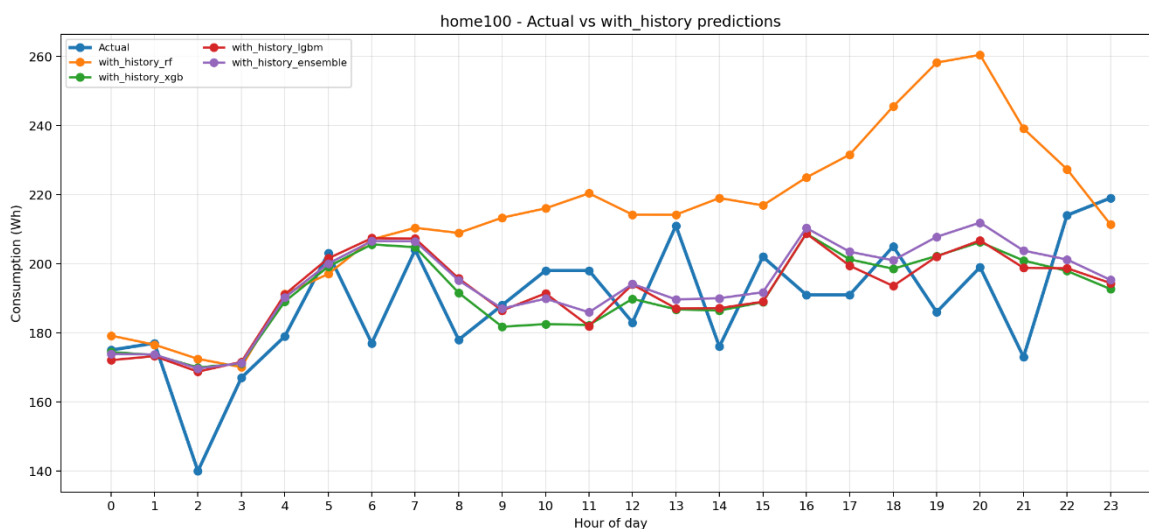
δεδομένων μετά την εφαρμογή της προσαρμοστικής διόρθωσης βάσει ιστορικού κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	25,80	32,72	13,98	5,125	4,534	0,591
XGBoost	13,13	15,80	7,10	4,586	4,534	0,052
LightGBM	13,14	15,67	7,11	4,608	4,534	0,074
Ensemble	13,22	16,03	7,20	4,650	4,534	0,116

Πίνακας 13 Απόδοση μοντέλων για την κατοικία home100 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Μετά την εφαρμογή της προσαρμοστικής διόρθωσης, παρατηρείται περαιτέρω βελτίωση της απόδοσης των μοντέλων. Το LightGBM παρουσιάζει το χαμηλότερο RMSE, ίσο με 15,67 Wh, ενώ το XGBoost εμφανίζει οριακά χαμηλότερο MAE και MAPE, καθώς και το μικρότερο ημερήσιο σφάλμα, ίσο με 0,052 kWh. Για τον λόγο αυτό, το LightGBM θεωρείται το καλύτερο μοντέλο με βάση την ωριαία ακρίβεια, ενώ το XGBoost παρουσιάζει την καλύτερη εκτίμηση της συνολικής ημερήσιας κατανάλωσης.

Η επίδραση της προσαρμοστικής διόρθωσης είναι θετική αλλά περιορισμένη, καθώς το RMSE του LightGBM μειώνεται από 16,44 Wh στο απλό σενάριο με χρήση ιστορικών δεδομένων, σε 15,67 Wh μετά την εφαρμογή της διόρθωσης. Παράλληλα, το ημερήσιο σφάλμα του LightGBM μειώνεται από 0,080 kWh σε 0,074 kWh. Το αποτέλεσμα αυτό είναι συνεπές με την ανάλυση ιστορικού, η οποία έδειξε ότι το πρόσφατο ιστορικό της home100 είναι αρκετά σταθερό και επομένως μπορεί να αξιοποιηθεί για μέτρια προσαρμογή της τελικής καμπύλης πρόβλεψης.



Σχήμα 15 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των τριών εκδοχών παρουσιάζεται στον Πίνακα 14, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE, καθώς βασικός στόχος της αξιολόγησης είναι η ακρίβεια της ωριαίας καμπύλης πρόβλεψης.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	RF	33,46	41,58	18,53	5,285	4,534	0,751
with-history	LGBM	13,37	16,44	7,26	4,614	4,534	0,080
with-history adaptive	LGBM	13,14	15,67	7,11	4,608	4,534	0,074

Πίνακας 14 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home100

Για την κατοικία home100, το σενάριο με χρήση ιστορικών δεδομένων εμφανίζει σαφώς καλύτερη απόδοση από το σενάριο χωρίς χρήση ιστορικών δεδομένων. Το RMSE μειώνεται από 41,58 Wh στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων, σε 16,44 Wh στο απλό σενάριο με χρήση ιστορικών δεδομένων και σε 15,67 Wh μετά την εφαρμογή της προσαρμοστικής διόρθωσης. Η συνολική μείωση του RMSE σε σχέση με το σενάριο χωρίς χρήση ιστορικών δεδομένων είναι περίπου 62%, γεγονός που δείχνει ότι η χρήση πρόσφατου ιστορικού κατανάλωσης βελτίωσε ουσιαστικά την πρόβλεψη.

Παράλληλα, η συνολική ημερήσια κατανάλωση εκτιμάται με πολύ μεγαλύτερη ακρίβεια στο σενάριο με χρήση ιστορικών δεδομένων. Το ημερήσιο σφάλμα στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων είναι 0,751 kWh, ενώ στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση μειώνεται σε 0,074 kWh για το LightGBM. Το αποτέλεσμα αυτό δείχνει ότι το ιστορικό κατανάλωσης συνέβαλε όχι μόνο στη βελτίωση της ωριαίας καμπύλης, αλλά και στην καλύτερη εκτίμηση του συνολικού ημερήσιου επιπέδου κατανάλωσης.

Η κατοικία home100 αποτελεί χαρακτηριστικό παράδειγμα κατοικίας όπου το πρόσφατο ιστορικό είναι πλήρες, σταθερό και αντιπροσωπευτικό της ημέρας πρόβλεψης. Η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν εντός του πρόσφατου ιστορικού εύρους και κοντά στη διάμεση ημερήσια κατανάλωση του ιστορικού. Για τον λόγο αυτό, τα μοντέλα με χρήση ιστορικών δεδομένων αποδίδουν σημαντικά καλύτερα από τα αντίστοιχα μοντέλα χωρίς χρήση ιστορικών δεδομένων, ενώ η προσαρμοστική διόρθωση προσφέρει επιπλέον μικρή αλλά συνεπή βελτίωση της ωριαίας ακρίβειας.

9.2.3 Αποτελέσματα για την κατοικία home268

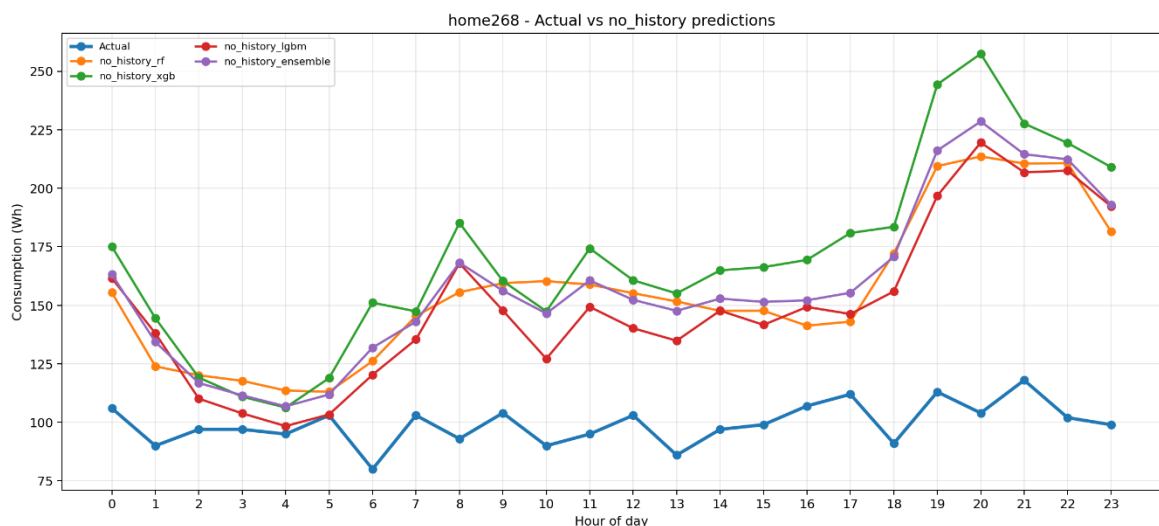
Η τρίτη κατοικία που εξετάστηκε είναι η home268, με ημέρα αξιολόγησης την 29η Ιουνίου 2018. Η κατοικία αντιστοιχεί σε μονοκατοικία, με έναν ένοικο, συνολικό εμβαδόν 89,5 m² και 13 καταγεγραμμένες ηλεκτρικές συσκευές. Ως προς την κατηγορία αστικότητας, η κατοικία ανήκει στην κατηγορία 3+, η οποία αντιστοιχεί σε πιο απομακρυσμένη μικρή πόλη ή αγροτική περιοχή. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης

ημέρας ήταν 2,384 kWh. Στον Πίνακα 15 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	56,26	62,64	56,68	3,734	2,384	1,350
XGBoost	70,64	79,13	70,91	4,079	2,384	1,695
LightGBM	50,76	59,02	50,89	3,602	2,384	1,218
Ensemble	58,92	66,17	59,21	3,798	2,384	1,414

Πίνακας 15 Απόδοση μοντέλων για την κατοικία home268 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το LightGBM εμφανίζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του LightGBM είναι 59,02 Wh, ενώ το MAPE ανέρχεται σε 50,89%. Παρ' όλα αυτά, όλα τα μοντέλα υπερεκτιμούν σημαντικά τη συνολική ημερήσια κατανάλωση. Ακόμη και το καλύτερο μοντέλο, το LightGBM, προβλέπει 3,602 kWh έναντι πραγματικής τιμής 2,384 kWh, με ημερήσιο σφάλμα 1,218 kWh. Το αποτέλεσμα αυτό δείχνει ότι, χωρίς πρόσφατο ιστορικό κατανάλωσης, το σύστημα δυσκολεύεται να εκτιμήσει σωστά το πραγματικό επίπεδο κατανάλωσης της συγκεκριμένης κατοικίας.



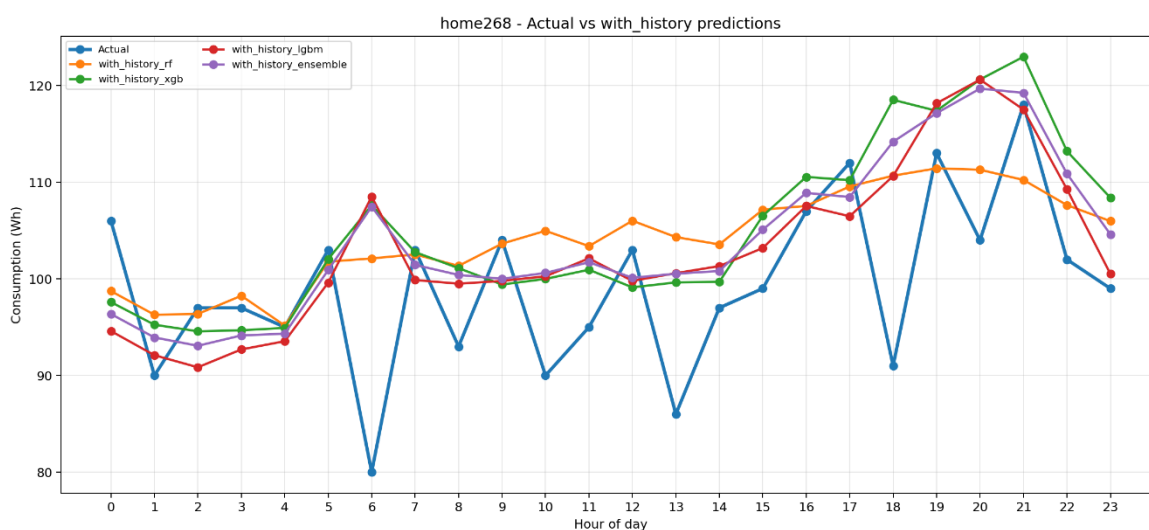
Σχήμα 16 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων, χωρίς την εφαρμογή της προσαρμοστικής διόρθωσης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 16.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	6,64	9,11	7,13	2,500	2,384	0,116
XGBoost	7,63	10,54	8,11	2,518	2,384	0,134
LightGBM	7,15	9,74	7,61	2,469	2,384	0,085
Ensemble	7,18	9,84	7,67	2,494	2,384	0,110

Πίνακας 16 Απόδοση μοντέλων για την κατοικία home268 στο απλό σενάριο με χρήση ιστορικών δεδομένων

Στο απλό σενάριο με χρήση ιστορικών δεδομένων, η βελτίωση σε σχέση με το σενάριο χωρίς χρήση ιστορικών δεδομένων είναι ιδιαίτερα έντονη. Το Random Forest παρουσιάζει την καλύτερη απόδοση ως προς το RMSE, με τιμή 9,11 Wh, ενώ το LightGBM εμφανίζει το μικρότερο ημερήσιο σφάλμα, με απόκλιση μόλις 0,085 kWh από την πραγματική ημερήσια κατανάλωση. Η συνολική εικόνα δείχνει ότι η εισαγωγή πρόσφατου ιστορικού κατανάλωσης βελτιώνει σημαντικά τόσο την ωριαία ακρίβεια όσο και την εκτίμηση του ημερήσιου επιπέδου κατανάλωσης. Ενδεικτικά, το RMSE μειώνεται από 59,02 Wh στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων, σε 9,11 Wh στο καλύτερο μοντέλο του απλού σεναρίου με χρήση ιστορικών δεδομένων.



Σχήμα 17 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στο απλό σενάριο με χρήση ιστορικών δεδομένων

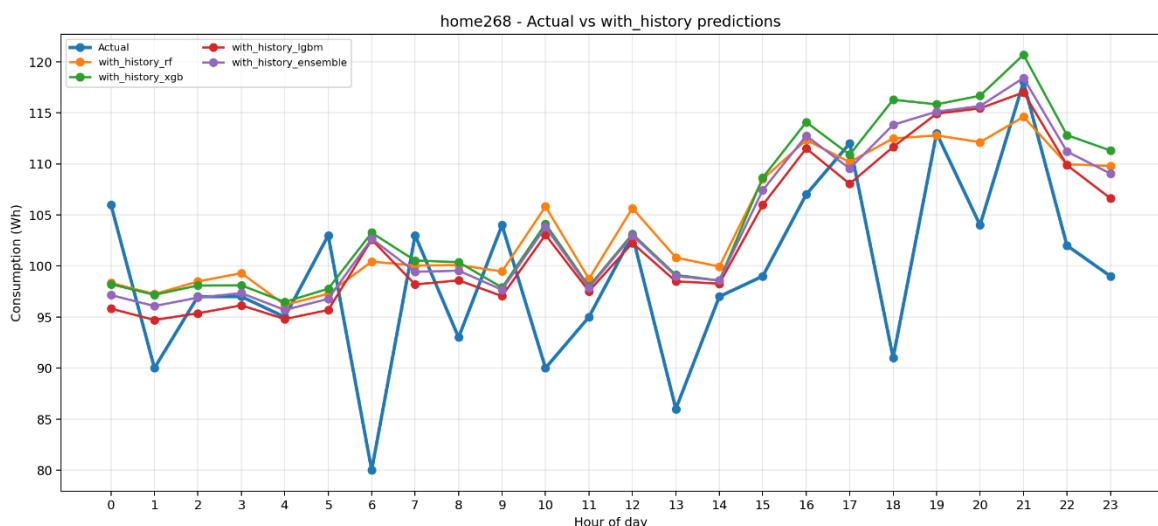
Η ανάλυση σταθερότητας έδειξε εξαιρετικά υψηλή σταθερότητα ιστορικού. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 2,364 kWh, η διάμεση τιμή ήταν 2,354 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν ιδιαίτερα περιορισμένο, από 2,311 έως 2,423 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 2,384 kWh, βρισκόταν εντός του ιστορικού εύρους και πολύ κοντά τόσο στη μέση όσο και στη διάμεση τιμή. Παράλληλα, η μεταβλητότητα του ημερήσιου επιπέδου κατανάλωσης, του ωριαίου προφίλ και των αιχμών ήταν ιδιαίτερα χαμηλή, γεγονός που δείχνει ότι το πρόσφατο ιστορικό παρουσίαζε πολύ υψηλή επαναληψιμότητα. Με βάση την ανάλυση αυτή, το σύστημα

πρότεινε υψηλό εύρος τιμών για τον συντελεστή προσαρμογής $\max_alpha$, συγκεκριμένα 0,65 καθώς το ιστορικό χαρακτηρίστηκε εξαιρετικά σταθερό και κατάλληλο για ισχυρότερη προσαρμοστική διόρθωση. Στον Πίνακα 17 παρουσιάζονται τα αποτελέσματα του σεναρίου with-history μετά την εφαρμογή της προσαρμοστικής διόρθωσης βάσει ιστορικού κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	7,05	9,11	7,50	2,501	2,384	0,117
XGBoost	7,48	9,99	7,93	2,518	2,384	0,134
LightGBM	6,70	8,89	7,10	2,470	2,384	0,086
Ensemble	6,90	9,34	7,34	2,494	2,384	0,110

Πίνακας 17 Απόδοση μοντέλων για την κατοικία home268 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Μετά την εφαρμογή της προσαρμοστικής διόρθωσης, το LightGBM παρουσιάζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE μειώνεται σε 8,89 Wh και το MAPE σε 7,10%. Η προσαρμοστική διόρθωση βελτιώνει ελαφρά την ωριαία ακρίβεια σε σχέση με το απλό σενάριο με χρήση ιστορικών δεδομένων, καθώς το καλύτερο RMSE μειώνεται από 9,11 Wh σε 8,89 Wh. Παράλληλα, το ημερήσιο σφάλμα παραμένει πολύ χαμηλό, περίπου 0,086 kWh, γεγονός που δείχνει ότι η διόρθωση δεν αλλοίωσε ουσιαστικά το συνολικό ημερήσιο επίπεδο της πρόβλεψης.



Σχήμα 18 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των τριών εκδοχών παρουσιάζεται στον Πίνακα 18, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE, καθώς βασικός στόχος της αξιολόγησης είναι η ακρίβεια της ωριαίας καμπύλης πρόβλεψης.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	LGBM	50,76	59,02	50,89	3,602	2,384	1,218
with-history	RF	6,64	9,11	7,13	2,500	2,384	0,116
with-history adaptive	LGBM	6,70	8,89	7,10	2,470	2,384	0,086

Πίνακας 18 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home268

Για την κατοικία home268, το σενάριο με χρήση ιστορικών δεδομένων εμφανίζει πολύ μεγάλη βελτίωση σε σχέση με το σενάριο χωρίς χρήση ιστορικών δεδομένων. Το RMSE μειώνεται από 59,02 Wh στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων σε 9,11 Wh στο απλό σενάριο με χρήση ιστορικών δεδομένων και σε 8,89 Wh μετά την εφαρμογή της προσαρμοστικής διόρθωσης. Η συνολική μείωση του RMSE σε σχέση με το σενάριο χωρίς χρήση ιστορικών δεδομένων είναι περίπου 85%, γεγονός που καθιστά την κατοικία home268 ένα από τα πιο χαρακτηριστικά παραδείγματα επιτυχούς αξιοποίησης πρόσφατου ιστορικού κατανάλωσης.

Παράλληλα, η εκτίμηση της συνολικής ημερήσιας κατανάλωσης βελτιώνεται σημαντικά. Στο σενάριο χωρίς χρήση ιστορικών δεδομένων το καλύτερο μοντέλο υπερεκτιμά την ημερήσια κατανάλωση κατά 1,218 kWh, ενώ στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, το αντίστοιχο ημερήσιο σφάλμα μειώνεται σε 0,086 kWh. Το αποτέλεσμα αυτό δείχνει ότι το πρόσφατο ιστορικό συνέβαλε τόσο στη διόρθωση του επιπέδου κατανάλωσης όσο και στην καλύτερη αναπαραγωγή της ωριαίας καμπύλης.

Η κατοικία home268 αποτελεί χαρακτηριστική περίπτωση όπου το πρόσφατο ιστορικό είναι πλήρες, εξαιρετικά σταθερό και ιδιαίτερα αντιπροσωπευτικό της ημέρας πρόβλεψης. Η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν εντός ενός πολύ στενού ιστορικού εύρους και πολύ κοντά στη μέση και στη διάμεση ημερήσια κατανάλωση του ιστορικού. Για τον λόγο αυτό, η εισαγωγή ιστορικών δεδομένων οδηγεί σε πολύ μεγάλη βελτίωση της πρόβλεψης σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Η προσαρμοστική διόρθωση προσφέρει επιπλέον μικρή βελτίωση της ωριαίας ακρίβειας, διατηρώντας παράλληλα πολύ χαμηλό ημερήσιο σφάλμα.

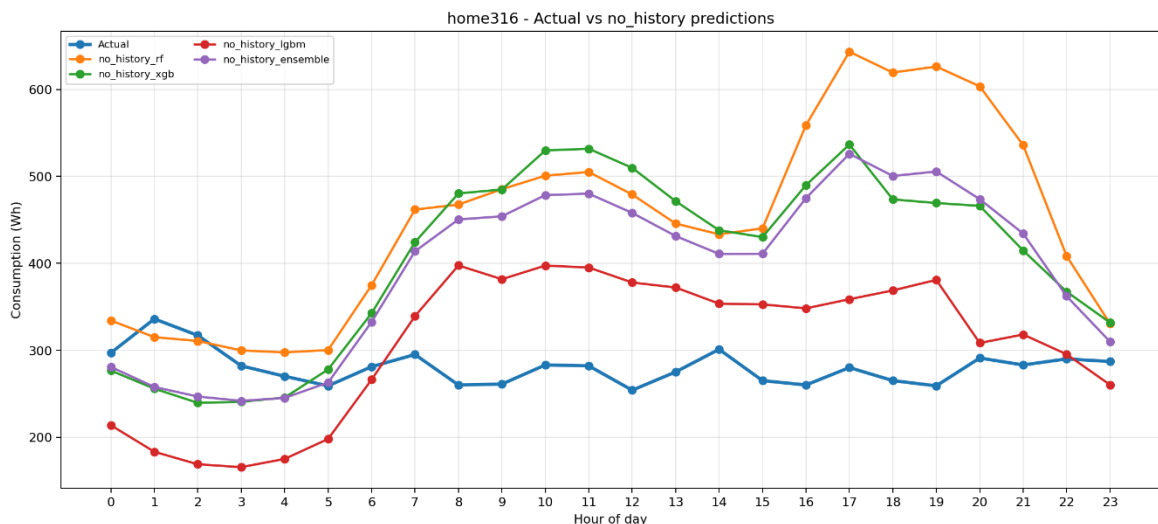
9.2.4 Αποτελέσματα για την κατοικία home316

Η τέταρτη κατοικία που εξετάστηκε είναι η home316, με ημέρα αξιολόγησης την 30ή Ιουνίου 2018. Η κατοικία αντιστοιχεί σε μονοκατοικία, με τέσσερις ενοίκους, συνολικό εμβαδόν 143,5 m² και 13 καταγεγραμμένες ηλεκτρικές συσκευές. Ως προς την κατηγορία αστικότητας, η κατοικία ανήκει στην κατηγορία 3+, η οποία αντιστοιχεί σε πιο απομακρυσμένη μικρή πόλη ή αγροτική περιοχή. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 6,733 kWh. Στον Πίνακα 19 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	170,74	206,03	62,43	10,776	6,733	4,043
XGBoost	145,10	167,10	52,79	9,727	6,733	2,994
LightGBM	84,99	95,12	30,53	7,375	6,733	0,642
Ensemble	132,01	153,48	48,04	9,441	6,733	2,708

Πίνακας 19 Απόδοση μοντέλων για την κατοικία home316 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το LightGBM εμφανίζει καθαρά την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του LightGBM είναι 95,12 Wh, ενώ το MAPE ανέρχεται σε 30,53%. Επιπλέον, το LightGBM παρουσιάζει και το μικρότερο ημερήσιο σφάλμα, καθώς προβλέπει συνολική ημερήσια κατανάλωση 7,375 kWh έναντι πραγματικής τιμής 6,733 kWh, με απόκλιση 0,642 kWh. Αντίθετα, τα υπόλοιπα μοντέλα υπερεκτιμούν εντονότερα τη συνολική ημερήσια κατανάλωση, γεγονός που αποτυπώνεται τόσο στις ωριαίες μετρικές όσο και στο ημερήσιο σφάλμα.



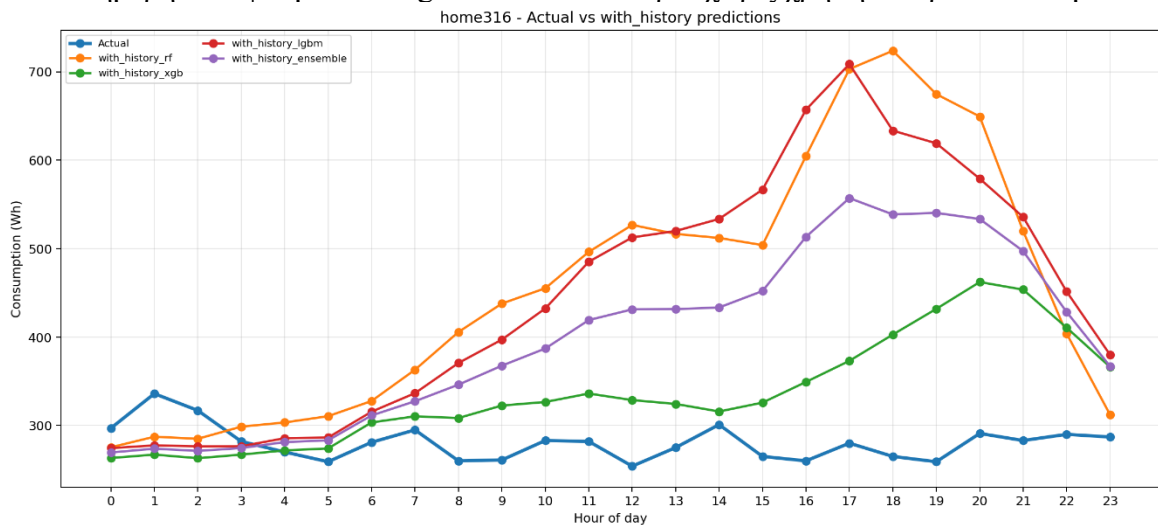
Σχήμα 19 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων, χωρίς την εφαρμογή της προσαρμοστικής διόρθωσης. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 20.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	182,02	228,79	66,59	10,897	6,733	4,164
XGBoost	69,52	86,03	25,02	8,058	6,733	1,325
LightGBM	176,36	220,24	64,34	10,711	6,733	3,978
Ensemble	128,71	157,00	46,82	9,536	6,733	2,803

Πίνακας 20 Απόδοση μοντέλων για την κατοικία home316 στο απλό σενάριο με χρήση ιστορικών δεδομένων

Στο απλό σενάριο με χρήση ιστορικών δεδομένων, το XGBoost παρουσιάζει την καλύτερη απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Το RMSE μειώνεται σε 86,03 Wh και το MAPE σε 25,02%, γεγονός που δείχνει ότι η χρήση ιστορικών δεδομένων βελτιώνει την ωριαία ακρίβεια σε σχέση με το καλύτερο αποτέλεσμα στο σενάριο χωρίς χρήση ιστορικών δεδομένων. Ωστόσο, η βελτίωση αυτή δεν συνοδεύεται από αντίστοιχη βελτίωση στη συνολική ημερήσια κατανάλωση. Το XGBoost προβλέπει 8,058 kWh έναντι πραγματικής τιμής 6,733 kWh, με ημερήσιο σφάλμα 1,325 kWh, το οποίο είναι υψηλότερο από το ημερήσιο σφάλμα του LightGBM στο σενάριο χωρίς χρήση ιστορικών δεδομένων.



Σχήμα 20 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στο απλό σενάριο με χρήση ιστορικών δεδομένων

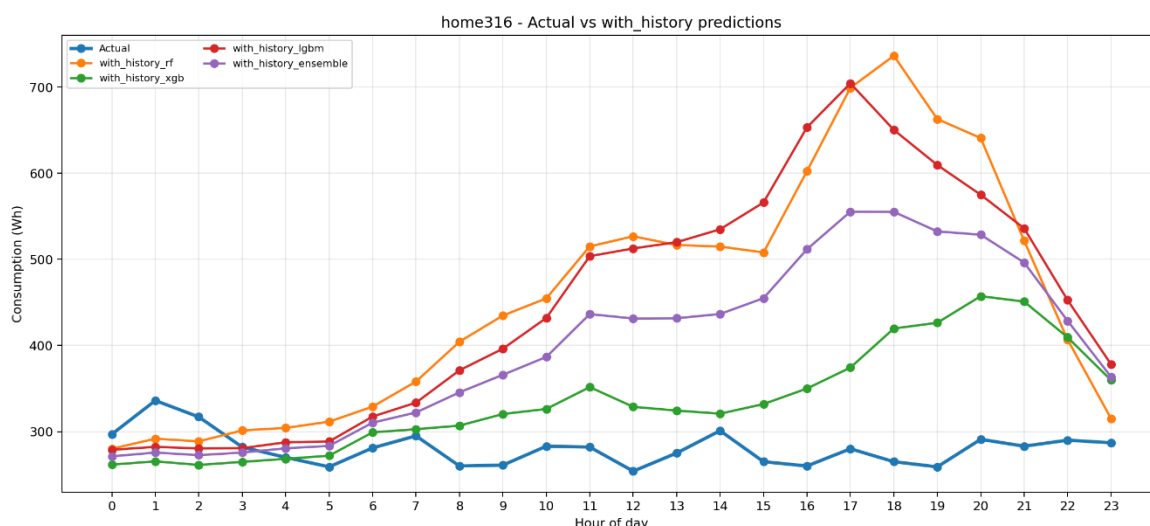
Η ανάλυση σταθερότητας έδειξε μέτρια σταθερότητα ιστορικού. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 12,389 kWh, η διάμεση τιμή ήταν 13,639 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 6,865–15,362 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 6,733 kWh, ήταν ελαφρώς χαμηλότερη ακόμη και από την ελάχιστη τιμή του πρόσφατου ιστορικού. Το στοιχείο αυτό δείχνει ότι το διαθέσιμο ιστορικό υποδήλωνε αυξημένο επίπεδο κατανάλωσης σε σχέση με την ημέρα στόχο. Παράλληλα, οι δείκτες μεταβλητότητας έδειξαν ότι το πρόσφατο ιστορικό δεν παρουσίαζε πολύ υψηλή επαναληψιμότητα, ιδιαίτερα ως προς το ημερήσιο επίπεδο κατανάλωσης και το ωριαίο προφίλ. Με βάση την ανάλυση αυτή, το σύστημα πρότεινε χαμηλό εύρος τιμών για τον συντελεστή προσαρμογής $\max_alpha$, συγκεκριμένα 0,1 ώστε να περιοριστεί ο

κίνδυνος υπερπροσαρμογής στο πρόσφατο ιστορικό σχήμα κατανάλωσης. Στον Πίνακα 21 παρουσιάζονται τα αποτελέσματα του σεναρίου με χρήση ιστορικών δεδομένων μετά την εφαρμογή της προσαρμοστικής διόρθωσης βάσει ιστορικού κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	182,12	228,96	66,64	10,925	6,733	4,192
XGBoost	70,08	86,56	25,23	8,054	6,733	1,321
LightGBM	176,32	220,63	64,35	10,745	6,733	4,012
Ensemble	128,87	157,72	46,89	9,552	6,733	2,819

Πίνακας 21 Απόδοση μοντέλων για την κατοικία home316 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Μετά την εφαρμογή της προσαρμοστικής διόρθωσης, το XGBoost παραμένει το καλύτερο μοντέλο ως προς τις βασικές ωριαίες μετρικές. Ωστόσο, η προσαρμοστική διόρθωση δεν οδηγεί σε ουσιαστική βελτίωση σε σχέση με το απλό σενάριο με χρήση ιστορικών δεδομένων. Το RMSE του XGBoost αυξάνεται ελαφρά από 86,03 Wh σε 86,56 Wh, ενώ το MAPE αυξάνεται από 25,02% σε 25,23%. Το ημερήσιο σφάλμα μειώνεται οριακά από 1,325 kWh σε 1,321 kWh, διαφορά η οποία είναι πρακτικά αμελητέα. Το αποτέλεσμα αυτό είναι συνεπές με την ανάλυση ιστορικού, η οποία χαρακτήρισε το διαθέσιμο ιστορικό ως μέτριας σταθερότητας και πρότεινε χαμηλό συντελεστή προσαρμογής.



Σχήμα 21 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των τριών εκδοχών παρουσιάζεται στον Πίνακα 22, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE, καθώς βασικός στόχος της αξιολόγησης είναι η ακρίβεια της ωριαίας καμπύλης πρόβλεψης.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	LGBM	84,99	95,12	30,53	7,375	6,733	0,642
with-history	XGB	69,52	86,03	25,02	8,058	6,733	1,325
with-history adaptive	XGB	70,08	86,56	25,23	8,054	6,733	1,321

Πίνακας 22 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home316

Για την κατοικία home316, η χρήση ιστορικών δεδομένων οδηγεί σε βελτίωση των βασικών ωριαίων μετρικών, αλλά όχι σε βελτίωση της συνολικής ημερήσιας εκτίμησης. Το RMSE μειώνεται από 95,12 Wh στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων, σε 86,03 Wh στο απλό σενάριο με χρήση ιστορικών δεδομένων. Ωστόσο, το ημερήσιο σφάλμα αυξάνεται από 0,642 kWh σε 1,325 kWh, γεγονός που δείχνει ότι το σενάριο με χρήση ιστορικών δεδομένων βελτιώνει το σχήμα της ωριαίας καμπύλης, αλλά υπερεκτιμά περισσότερο το συνολικό ημερήσιο επίπεδο κατανάλωσης.

Το αποτέλεσμα αυτό είναι ερμηνεύσιμο με βάση την ανάλυση ιστορικού. Το πρόσφατο ιστορικό της κατοικίας παρουσίαζε μέση ημερήσια κατανάλωση 12,389 kWh και διάμεση τιμή 13,639 kWh, δηλαδή τιμές σημαντικά υψηλότερες από την πραγματική κατανάλωση της ημέρας αξιολόγησης, η οποία ήταν 6,733 kWh. Επιπλέον, η πραγματική ημέρα στόχος βρισκόταν ελαφρώς κάτω ακόμη και από την ελάχιστη ημερήσια τιμή του πρόσφατου ιστορικού. Επομένως, η υπερεκτίμηση της κατανάλωσης στο σενάριο με χρήση ιστορικών δεδομένων δεν οφείλεται απαραίτητα σε λανθασμένη λειτουργία του μοντέλου, αλλά στο γεγονός ότι το διαθέσιμο ιστορικό υποδήλωνε αυξημένο επίπεδο κατανάλωσης σε σχέση με την ημέρα στόχο.

Η προσαρμοστική διόρθωση δεν προσφέρει ουσιαστική πρόσθετη βελτίωση στη συγκεκριμένη κατοικία. Η μεταβολή του RMSE είναι ελαφρώς αρνητική, ενώ το ημερήσιο σφάλμα παραμένει σχεδόν αμετάβλητο. Το αποτέλεσμα αυτό είναι αναμενόμενο, καθώς η ανάλυση ιστορικού χαρακτήρισε το διαθέσιμο ιστορικό ως μέτριας σταθερότητας και πρότεινε χαμηλό συντελεστή $\max_alpha$. Συνεπώς, το σύστημα απέφυγε να μεταβάλει έντονα την αρχική πρόβλεψη βάσει ενός ιστορικού που δεν ήταν αρκετά σταθερό.

Επιπλέον, πραγματοποιήθηκε δοκιμαστικός έλεγχος στην ανάλυση ιστορικού, με μεγαλύτερο παράθυρο ιστορικού 14 ημερών, ώστε να εξεταστεί αν η χρήση περισσότερων προηγούμενων ημερών θα οδηγούσε σε πιο σταθερή εικόνα της κατανάλωσης. Το αποτέλεσμα έδειξε ότι το ιστορικό παρέμεινε μέτριας σταθερότητας, με προτεινόμενο εύρος $\max_alpha$ 0,10–0,15. Παράλληλα, η μέση ημερήσια κατανάλωση του 14ήμερου ιστορικού ήταν 13,119 kWh και η διάμεση τιμή 13,598 kWh, δηλαδή σημαντικά υψηλότερες από την πραγματική κατανάλωση της ημέρας αξιολόγησης. Το εύρημα αυτό ενισχύει την ερμηνεία ότι η υπερεκτίμηση της κατανάλωσης στο σενάριο με χρήση ιστορικών δεδομένων, δεν οφείλεται στο μικρό μήκος του ιστορικού παραθύρου, αλλά στο γεγονός ότι η ημέρα στόχος είχε ασυνήθιστα χαμηλότερη κατανάλωση σε σχέση με το πρόσφατο ιστορικό της κατοικίας.

Η κατοικία home316 αποτελεί περίπτωση όπου η χρήση ιστορικών δεδομένων βοηθά ως προς την ωριαία ακρίβεια, αλλά δεν αρκεί για τη διόρθωση της υπερεκτίμησης της ημερήσιας κατανάλωσης. Σε τέτοιες περιπτώσεις, η ποιότητα, η σταθερότητα και κυρίως η αντιπροσωπευτικότητα του πρόσφατου ιστορικού ως προς την ημέρα στόχο αποτελούν κρίσιμους παράγοντες για την αποτελεσματικότητα του σεναρίου με χρήση ιστορικών δεδομένων και της προσαρμοστικής διόρθωσης.

9.3 Εποχικές μελέτες περίπτωσης

Μετά την αξιολόγηση των επιλεγμένων κατοικιών στο ανεξάρτητο σύνολο ελέγχου του Ιουνίου 2018, πραγματοποιήθηκαν πρόσθετες εποχικές μελέτες περίπτωσης. Σκοπός της ενότητας αυτής είναι να εξεταστεί η συμπεριφορά του συστήματος σε διαφορετικές χρονικές περιόδους του έτους και να διερευνηθεί αν τα συμπεράσματα της κύριας αξιολόγησης διατηρούνται και σε ημέρες με διαφορετικά εποχικά χαρακτηριστικά.

Για τον σκοπό αυτό επιλέχθηκαν ενδεικτικές ημέρες από διαφορετικές εποχές, ώστε να καλυφθούν περιπτώσεις χειμερινής, εαρινής, θερινής και φθινοπωρινής περιόδου. Οι μελέτες περίπτωσης πραγματοποιήθηκαν για τις ίδιες επιλεγμένες κατοικίες που εξετάστηκαν στην προηγούμενη ενότητα, με στόχο να υπάρχει συνέχεια στην αξιολόγηση και να είναι δυνατή η σύγκριση της συμπεριφοράς κάθε κατοικίας σε διαφορετικές περιόδους.

Στις εποχικές μελέτες περίπτωσης εξετάζονται δύο βασικές εκδοχές πρόβλεψης: το σενάριο χωρίς χρήση ιστορικών δεδομένων και το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση. Το απλό σενάριο με χρήση ιστορικών δεδομένων χωρίς προσαρμοστική διόρθωση δεν παρουσιάζεται αναλυτικά στην ενότητα αυτή, ώστε η αξιολόγηση να επικεντρωθεί στη σύγκριση μεταξύ της βασικής πρόβλεψης του συστήματος και της τελικής μορφής του σεναρίου με αξιοποίηση ιστορικού.

Όπως και στην προηγούμενη ενότητα, για κάθε περίπτωση εξετάζονται οι μετρικές MAE, RMSE, MAPE και ημερήσιο σφάλμα. Παράλληλα, λαμβάνεται υπόψη η ανάλυση του πρόσφατου ιστορικού κατανάλωσης, ώστε να ερμηνευθεί αν η χρήση ιστορικών δεδομένων ήταν χρήσιμη, περιορισμένη ή πιθανώς μη αντιπροσωπευτική για την ημέρα στόχο.

Οι εποχικές μελέτες περίπτωσης χρησιμοποιούνται ως συμπληρωματικά λειτουργικά παραδείγματα αξιολόγησης και όχι ως κύρια απόδειξη γενικευτικής ικανότητας των μοντέλων. Η κύρια αξιολόγηση της γενίκευσης βασίζεται στις κατοικίες και στις ημέρες του ανεξάρτητου συνόλου ελέγχου του Ιουνίου 2018, ενώ οι εποχικές περιπτώσεις αξιοποιούνται κυρίως για τη διερεύνηση της συμπεριφοράς του συστήματος υπό διαφορετικές εποχικές συνθήκες και διαφορετικά επίπεδα σταθερότητας ιστορικού.

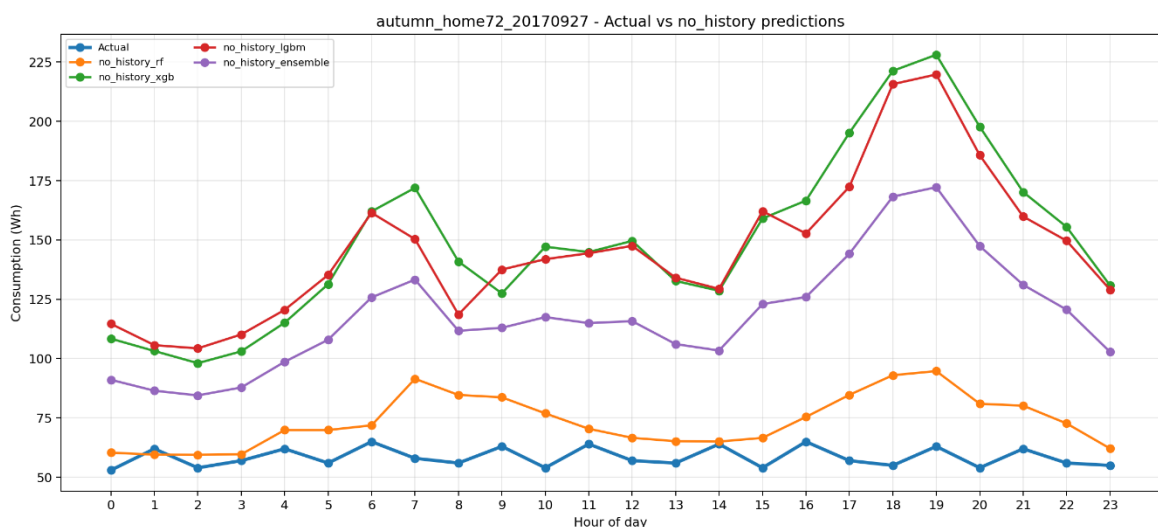
9.3.1 Φθινοπωρινή αξιολόγηση για την κατοικία home72

Η πρώτη εποχική μελέτη περίπτωσης αφορά την κατοικία home72 και ημερομηνία αξιολόγησης την 27η Σεπτεμβρίου 2017, η οποία αντιστοιχεί σε φθινοπωρινή περίοδο. Η κατοικία αντιστοιχεί σε διαμέρισμα με δύο ενοίκους, συνολικό εμβαδόν 80 m² και 9 καταγεγραμμένες ηλεκτρικές συσκευές. Ως προς την κατηγορία αστικότητας, η κατοικία ανήκει στην κατηγορία 1, η οποία αντιστοιχεί σε μεγάλη αστική περιοχή. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 1,402 kWh. Στον Πίνακα 23 παρουσιάζονται τα αποτελέσματα του σεναρίου χωρίς χρήση ιστορικών δεδομένων.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	15,34	18,69	26,64	1,765	1,402	0,363
XGBoost	91,07	97,47	156,84	3,588	1,402	2,186
LightGBM	87,48	92,39	150,66	3,502	1,402	2,100
Ensemble	59,62	63,83	102,77	2,833	1,402	1,431

Πίνακας 23 Απόδοση μοντέλων για την κατοικία home72 στη φθινοπωρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το Random Forest εμφανίζει την καλύτερη απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του Random Forest είναι 18,69 Wh, ενώ το ημερήσιο σφάλμα είναι 0,363 kWh. Τα υπόλοιπα μοντέλα υπερεκτιμούν σημαντικά τη συνολική ημερήσια κατανάλωση, γεγονός που οδηγεί σε αρκετά υψηλότερες τιμές σφάλματος, τόσο σε ωριαίο όσο και σε ημερήσιο επίπεδο.



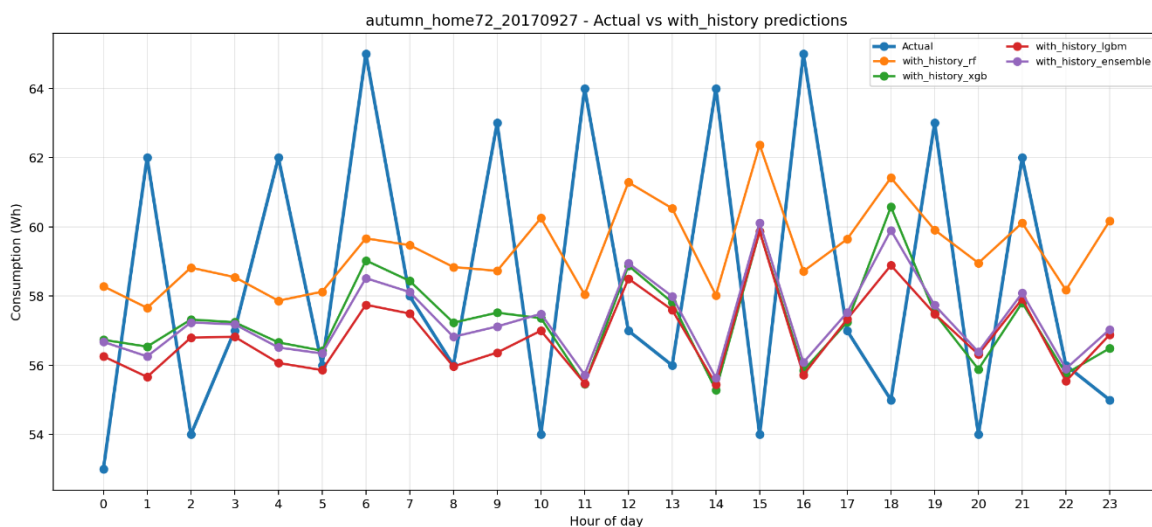
Σχήμα 22 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στη φθινοπωρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 24.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	4,34	4,69	7,48	1,424	1,402	0,022
XGBoost	3,75	4,66	6,25	1,375	1,402	0,027
LightGBM	3,74	4,75	6,19	1,365	1,402	0,037
Ensemble	3,76	4,64	6,26	1,375	1,402	0,027

Πίνακας 24 Απόδοση μοντέλων για την κατοικία home72 στη φθινοπωρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, παρατηρείται πολύ μεγάλη βελτίωση σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Το Ensemble παρουσιάζει το χαμηλότερο RMSE, ίσο με 4,64 Wh, ενώ το LightGBM εμφανίζει οριακά χαμηλότερο MAE και MAPE. Παράλληλα, όλα τα μοντέλα με χρήση ιστορικών δεδομένων εκτιμούν με μεγάλη ακρίβεια τη συνολική ημερήσια κατανάλωση, καθώς τα ημερήσια σφάλματα κυμαίνονται περίπου από 0,022 έως 0,037 kWh.



Σχήμα 23 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home72 στη φθινοπωρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η ανάλυση σταθερότητας έδειξε εξαιρετικά υψηλή σταθερότητα ιστορικού. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 1,413 kWh, η διάμεση τιμή ήταν 1,418 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν ιδιαίτερα περιορισμένο, από 1,387 έως 1,439 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 1,402 kWh, βρισκόταν εντός του ιστορικού εύρους και πολύ κοντά τόσο στη μέση όσο και στη διάμεση τιμή. Επιπλέον, οι δείκτες μεταβλητότητας της ημερήσιας κατανάλωσης, του ωριαίου προφίλ και των αιχμών ήταν πολύ χαμηλοί, γεγονός που δείχνει ότι το πρόσφατο ιστορικό παρουσίαζε εξαιρετικά επαναλαμβανόμενη συμπεριφορά. Με βάση την ανάλυση αυτή, το σύστημα κατέταξε το ιστορικό στην κατηγορία Excellent και πρότεινε υψηλή τιμή για τον συντελεστή max_alpha, συγκεκριμένα 0,60.

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 25, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	RF	15,34	18,69	26,64	1,765	1,402	0,363
with-history adaptive	Ensemble	3,76	4,64	6,26	1,375	1,402	0,027

Πίνακας 25 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home72 στη φθινοπωρινή περίπτωση

Για τη φθινοπωρινή περίπτωση της home72, η χρήση ιστορικών δεδομένων οδηγεί σε πολύ σημαντική βελτίωση της πρόβλεψης. Το RMSE μειώνεται από 18,69 Wh στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων σε 4,64 Wh στο καλύτερο μοντέλο του σεναρίου με χρήση ιστορικών δεδομένων, δηλαδή παρουσιάζει μείωση περίπου 75%. Παράλληλα, το ημερήσιο σφάλμα μειώνεται από 0,363 kWh σε 0,027 kWh, γεγονός που δείχνει ότι το ιστορικό συνέβαλε τόσο στη βελτίωση της ωριαίας καμπύλης όσο και στην ακριβέστερη εκτίμηση της συνολικής ημερήσιας κατανάλωσης.

Η περίπτωση αυτή είναι ιδιαίτερα ενδεικτική της χρησιμότητας του ιστορικού όταν αυτό είναι πλήρες, εξαιρετικά σταθερό και αντιπροσωπευτικό της ημέρας στόχου. Σε αντίθεση με την κύρια αξιολόγηση της κατοικίας home72 στο test set του Ιουνίου 2018, όπου το ιστορικό είχε μόνο μέτρια σταθερότητα και το σενάριο χωρίς χρήση ιστορικών δεδομένων παρέμεινε οριακά καλύτερο, στη φθινοπωρινή περίπτωση το πρόσφατο ιστορικό ήταν πολύ πιο επαναλαμβανόμενο και κοντά στο πραγματικό ημερήσιο επίπεδο της ημέρας πρόβλεψης. Για τον λόγο αυτό, το σενάριο με χρήση ιστορικών δεδομένων αποδίδει σαφώς καλύτερα.

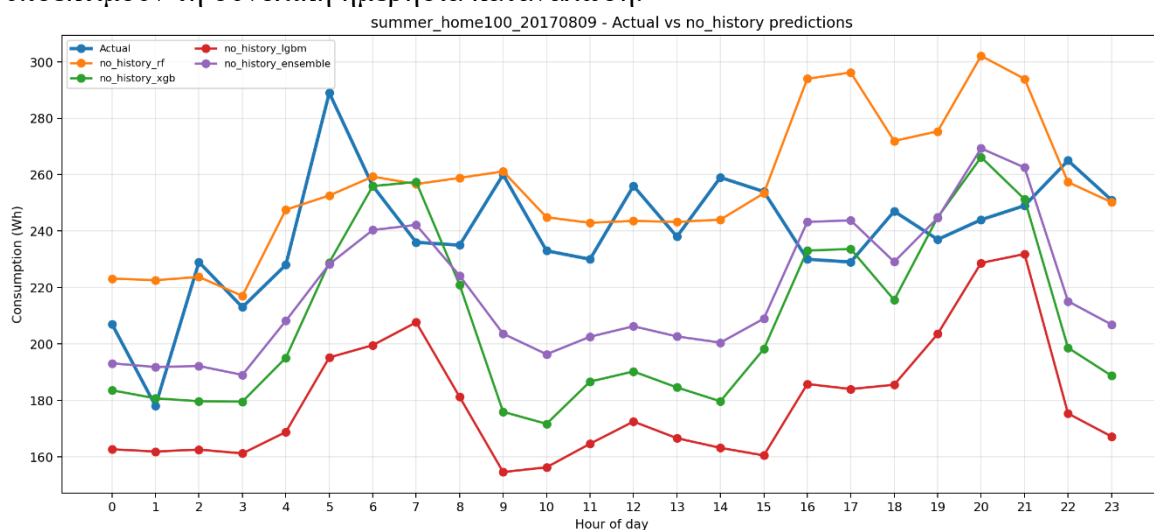
9.3.2 Θερινή αξιολόγηση για την κατοικία home100

Η δεύτερη εποχική μελέτη περίπτωσης αφορά την κατοικία home100 και ημερομηνία αξιολόγησης την 9η Αυγούστου 2017, η οποία αντιστοιχεί σε θερινή περίοδο. Η κατοικία αντιστοιχεί σε διαμέρισμα με έναν ένοικο, συνολικό εμβαδόν 72,5 m² και 11 καταγεγραμμένες ηλεκτρικές συσκευές. Ως προς την κατηγορία αστικότητας, η κατοικία ανήκει στην κατηγορία 3+, η οποία αντιστοιχεί σε πιο απομακρυσμένη μικρή πόλη ή αγροτική περιοχή. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 5,753 kWh. Στον Πίνακα 26 παρουσιάζονται τα αποτελέσματα του σεναρίου χωρίς χρήση ιστορικών δεδομένων.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	22,45	30,20	9,63	6,136	5,753	0,383
XGBoost	36,71	45,03	14,92	5,000	5,753	0,753
LightGBM	60,54	65,94	24,84	4,300	5,753	1,453
Ensemble	29,06	33,73	11,83	5,244	5,753	0,509

Πίνακας 26 Απόδοση μοντέλων για την κατοικία home100 στη θερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το Random Forest παρουσιάζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του Random Forest είναι 30,20 Wh και το MAPE 9,63%. Η συνολική ημερήσια κατανάλωση που προβλέπει το Random Forest είναι 6,136 kWh, έναντι πραγματικής τιμής 5,753 kWh, με ημερήσιο σφάλμα 0,383 kWh. Τα υπόλοιπα μοντέλα εμφανίζουν μεγαλύτερες αποκλίσεις, με το XGBoost, το LightGBM και το Ensemble να υποεκτιμούν τη συνολική ημερήσια κατανάλωση.



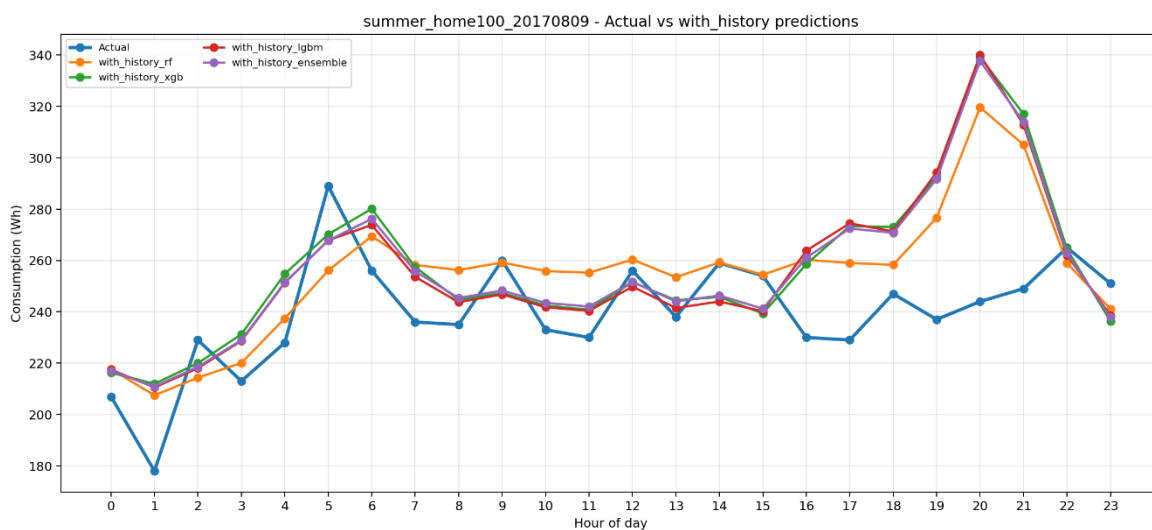
Σχήμα 24 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στη θερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 27.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	20,38	26,97	8,64	6,113	5,753	0,360
XGBoost	23,99	32,34	10,17	6,152	5,753	0,399
LightGBM	23,54	32,00	9,96	6,127	5,753	0,374
Ensemble	23,39	31,48	9,91	6,137	5,753	0,384

Πίνακας 27 Απόδοση μοντέλων για την κατοικία home100 στη θερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, το Random Forest παραμένει το καλύτερο μοντέλο ως προς τις βασικές μετρικές σφάλματος. Το RMSE μειώνεται από 30,20 Wh στο σενάριο χωρίς ιστορικά δεδομένα σε 26,97 Wh, ενώ το MAPE μειώνεται από 9,63% σε 8,64%. Παράλληλα, το ημερήσιο σφάλμα του Random Forest μειώνεται ελαφρά από 0,383 kWh σε 0,360 kWh. Η βελτίωση είναι υπαρκτή, αλλά πιο περιορισμένη σε σχέση με τη φθινοπωρινή περίπτωση της home72.



Σχήμα 25 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home100 στη θερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η ανάλυση σταθερότητας έδειξε υψηλή σταθερότητα ιστορικού. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 6,338 kWh, η διάμεση τιμή ήταν 6,289 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 6,086–6,643 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 5,753 kWh, ήταν χαμηλότερη από το πρόσφατο ιστορικό εύρος. Το στοιχείο αυτό δείχνει ότι, παρότι το ιστορικό ήταν σταθερό, το πρόσφατο ημερήσιο επίπεδο κατανάλωσης ήταν υψηλότερο από αυτό της ημέρας στόχου. Επιπλέον, οι δείκτες μεταβλητότητας της ημερήσιας κατανάλωσης και του ωριαίου προφίλ ήταν χαμηλοί, ενώ η σταθερότητα των αιχμών ήταν πιο περιορισμένη. Με βάση την ανάλυση αυτή, το σύστημα κατέταξε το ιστορικό στην κατηγορία High με προτεινόμενη τιμή max_alpha 0,25.

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 28, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	RF	22,45	30,20	9,63	6,136	5,753	0,383
with-history adaptive	RF	20,38	26,97	8,64	6,113	5,753	0,360

Πίνακας 28 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home100 στη θερινή περίπτωση

Για τη θερινή περίπτωση της κατοικίας home100, η χρήση ιστορικών δεδομένων και προσαρμοστικής διόρθωσης οδηγεί σε ήπια αλλά συνεπή βελτίωση της πρόβλεψης. Το RMSE μειώνεται από 30,20 Wh σε 26,97 Wh, δηλαδή παρουσιάζει μείωση περίπου 11%. Παράλληλα, το ημερήσιο σφάλμα μειώνεται από 0,383 kWh σε 0,360 kWh. Επομένως, το πρόσφατο ιστορικό συνέβαλε στη βελτίωση της ωριαίας ακρίβειας, αλλά η επίδρασή του στο συνολικό ημερήσιο επίπεδο ήταν περιορισμένη.

Το αποτέλεσμα αυτό είναι ερμηνεύσιμο με βάση την ανάλυση ιστορικού. Το ιστορικό της κατοικίας home100 ήταν πλήρες και σταθερό, ωστόσο η μέση και η διάμεση ημερήσια κατανάλωση του πρόσφατου ιστορικού ήταν υψηλότερες από την πραγματική κατανάλωση της ημέρας στόχου. Συνεπώς, το σύστημα είχε διαθέσιμη αξιόπιστη ιστορική πληροφορία ως προς τη σταθερότητα, αλλά το πρόσφατο επίπεδο κατανάλωσης δεν ταυτιζόταν πλήρως με την ημέρα πρόβλεψης. Για τον λόγο αυτό, η χρήση ιστορικών δεδομένων βελτίωσε την πρόβλεψη, χωρίς όμως να οδηγήσει σε τόσο έντονη βελτίωση όσο στην περίπτωση της home72, όπου το ιστορικό ήταν ταυτόχρονα εξαιρετικά σταθερό και πολύ κοντά στο πραγματικό ημερήσιο επίπεδο της ημέρας στόχου.

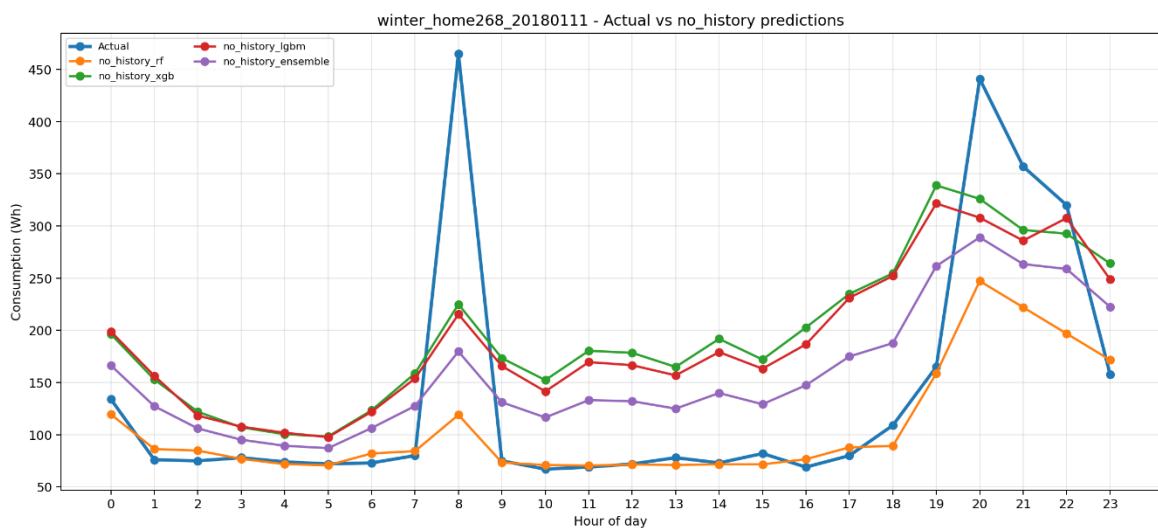
9.3.3 Χειμερινή αξιολόγηση για την κατοικία home268

Η τρίτη εποχική μελέτη περίπτωσης αφορά την κατοικία home268 και ημερομηνία αξιολόγησης την 11η Ιανουαρίου 2018, η οποία αντιστοιχεί σε χειμερινή περίοδο. Η κατοικία αντιστοιχεί σε μονοκατοικία, με έναν ένοικο, συνολικό εμβαδόν 89,5 m² και 13 καταγεγραμμένες ηλεκτρικές συσκευές. Ως προς την κατηγορία αστικότητας, η κατοικία ανήκει στην κατηγορία 3+, η οποία αντιστοιχεί σε πιο απομακρυσμένη μικρή πόλη ή αγροτική περιοχή. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 3,342 kWh. Στον Πίνακα 29 παρουσιάζονται τα αποτελέσματα του σεναρίου χωρίς χρήση ιστορικών δεδομένων.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	38,82	89,45	14,23	2,546	3,342	0,796
XGBoost	93,84	106,73	93,18	4,708	3,342	1,366
LightGBM	89,45	103,03	87,02	4,557	3,342	1,215
Ensemble	68,32	87,28	57,58	3,798	3,342	0,456

Πίνακας 29 Απόδοση μοντέλων για την κατοικία home268 στη χειμερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το Ensemble παρουσιάζει την καλύτερη απόδοση ως προς τη μετρική RMSE, με τιμή 87,28 Wh και ταυτόχρονα εμφανίζει το μικρότερο ημερήσιο σφάλμα, ίσο με 0,456 kWh. Το Random Forest έχει χαμηλότερο MAE και MAPE, όμως το Ensemble αποδίδει καλύτερα ως προς την ωριαία καμπύλη με βάση το RMSE και προσεγγίζει καλύτερα τη συνολική ημερήσια κατανάλωση. Τα XGBoost και LightGBM υπερεκτιμούν περισσότερο το ημερήσιο επίπεδο κατανάλωσης.



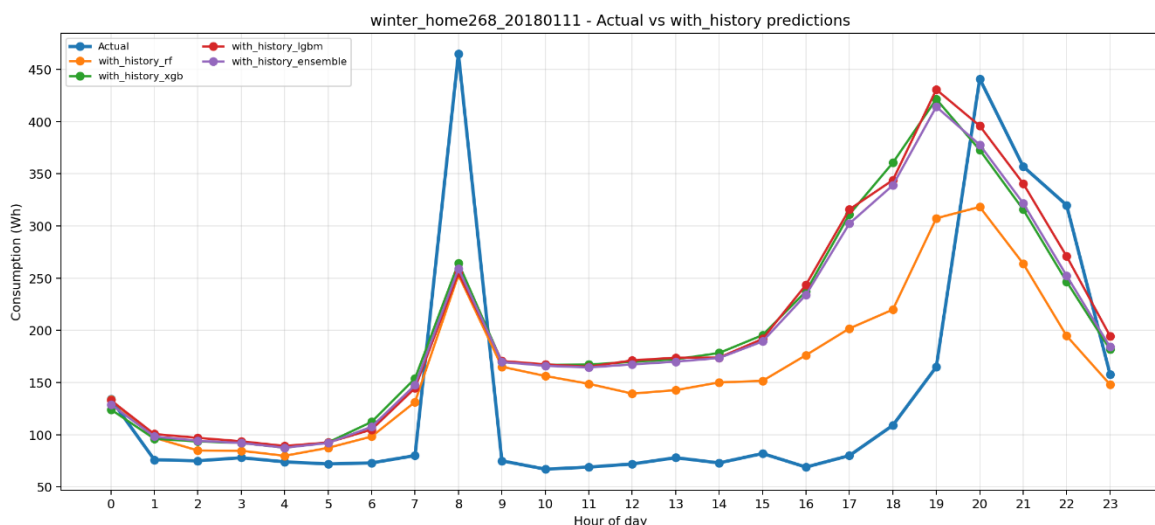
Σχήμα 26 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στη χειμερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 30.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	71,63	88,83	64,78	3,935	3,342	0,593
XGBoost	92,84	119,55	93,91	4,784	3,342	1,442
LightGBM	90,07	118,84	92,86	4,861	3,342	1,519
Ensemble	89,39	115,30	90,48	4,734	3,342	1,392

Πίνακας 30 Απόδοση μοντέλων για την κατοικία home268 στη χειμερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, το Random Forest παρουσιάζει την καλύτερη απόδοση μεταξύ των μοντέλων, με RMSE ίσο με 88,83 Wh και ημερήσιο σφάλμα 0,593 kWh. Ωστόσο, η απόδοση αυτή δεν είναι καλύτερη από το καλύτερο αποτέλεσμα του σεναρίου χωρίς χρήση ιστορικών δεδομένων. Συγκεκριμένα, το RMSE παραμένει ελαφρώς υψηλότερο από το 87,28 Wh του μοντέλου Ensemble χωρίς χρήση ιστορικών δεδομένων, ενώ και το ημερήσιο σφάλμα είναι μεγαλύτερο.



Σχήμα 27 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home268 στη χειμερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η ανάλυση σταθερότητας έδειξε χαμηλή σταθερότητα ιστορικού. Το διαθέσιμο ιστορικό ήταν πλήρες, ωστόσο, η μέση ημερήσια κατανάλωση του ιστορικού ήταν 3,797 kWh, η διάμεση τιμή ήταν 3,711 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν ιδιαίτερα μεγάλο, από 1,456 έως 8,881 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 3,342 kWh, βρισκόταν εντός του ιστορικού εύρους, αλλά το εύρος αυτό ήταν τόσο μεγάλο ώστε δεν παρείχε σταθερή εικόνα της τυπικής ημερήσιας συμπεριφοράς της κατοικίας. Επιπλέον, οι δείκτες μεταβλητότητας της ημερήσιας κατανάλωσης, του ωριαίου προφίλ και των αιχμών ήταν υψηλοί, με αποτέλεσμα οι αντίστοιχοι δείκτες σταθερότητας να είναι μηδενικοί. Με βάση την ανάλυση αυτή, το σύστημα κατέταξε το ιστορικό στην κατηγορία Low και πρότεινε πολύ χαμηλή τιμή max_alpha, συγκεκριμένα 0,05.

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 31, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	Ensemble	68,32	87,28	57,58	3,798	3,342	0,456
with-history adaptive	RF	71,63	88,83	64,78	3,935	3,342	0,593

Πίνακας 31 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home268 στη χειμερινή περίπτωση

Για τη χειμερινή περίπτωση της home268, το σενάριο χωρίς χρήση ιστορικών δεδομένων εμφανίζει οριακά καλύτερη απόδοση από το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση. Το RMSE στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων είναι 87,28 Wh, ενώ το αντίστοιχο RMSE στο καλύτερο μοντέλο του σεναρίου με χρήση ιστορικών δεδομένων είναι 88,83 Wh. Παράλληλα, το ημερήσιο σφάλμα αυξάνεται από 0,456 kWh σε 0,593 kWh όταν χρησιμοποιείται ιστορικό κατανάλωσης.

Το αποτέλεσμα αυτό είναι συνεπές με την ανάλυση ιστορικού. Παρότι το ιστορικό ήταν πλήρες, παρουσίαζε πολύ μεγάλη μεταβλητότητα ως προς τη συνολική ημερήσια κατανάλωση, το ωριαίο προφίλ και τις αιχμές. Το μεγάλο εύρος των ημερήσιων τιμών, από 1,456 έως 8,881 kWh, δείχνει ότι οι προηγούμενες ημέρες δεν ακολουθούσαν σταθερό και επαναλαμβανόμενο πρότυπο. Για τον λόγο αυτό, το ιστορικό δεν αποτέλεσε αξιόπιστη βάση για ουσιαστική προσαρμογή της πρόβλεψης.

Η περίπτωση αυτή παρουσιάζει ιδιαίτερο ενδιαφέρον, καθώς διαφοροποιείται από την κύρια αξιολόγηση της ίδιας κατοικίας στο test set του Ιουνίου 2018. Εκεί, το πρόσφατο ιστορικό της home268 ήταν εξαιρετικά σταθερό και αντιπροσωπευτικό της ημέρας στόχου, με αποτέλεσμα το σενάριο με χρήση ιστορικών δεδομένων να βελτιώσει σημαντικά την πρόβλεψη. Αντίθετα, στη χειμερινή περίπτωση το ιστορικό ήταν ασταθές και μη επαναλαμβανόμενο, με αποτέλεσμα το σενάριο χωρίς ιστορικά δεδομένα να παραμείνει ελαφρώς καλύτερο.

Συνεπώς, η χειμερινή περίπτωση της home268 επιβεβαιώνει ότι η χρήση ιστορικών δεδομένων δεν είναι από μόνη της επαρκής για τη βελτίωση της πρόβλεψης. Η αποτελεσματικότητά της εξαρτάται από την ποιότητα, τη σταθερότητα και την αντιπροσωπευτικότητα του πρόσφατου ιστορικού. Όταν το ιστορικό παρουσιάζει χαμηλή σταθερότητα και μη επαναλαμβανόμενες αιχμές, η προσαρμοστική διόρθωση πρέπει να παραμένει περιορισμένη, όπως πράγματι συνέβη μέσω της χαμηλής προτεινόμενης τιμής max_alpha.

9.3.4 Εαρινή αξιολόγηση για την κατοικία home316

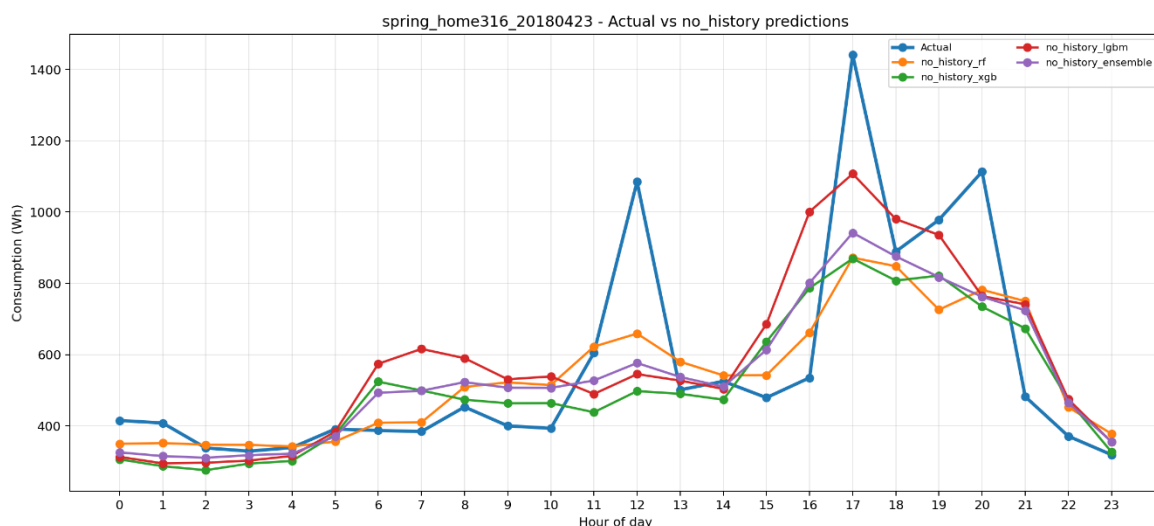
Η τέταρτη εποχική μελέτη περίπτωσης αφορά την κατοικία home316 και ημερομηνία αξιολόγησης την 23η Απριλίου 2018, η οποία αντιστοιχεί σε εαρινή περίοδο. Η κατοικία αντιστοιχεί σε μονοκατοικία, με τέσσερις ενοίκους, συνολικό εμβαδόν 143,5 m² και 13

καταγεγραμμένες ηλεκτρικές συσκευές. Ως προς την κατηγορία αστικότητας, η κατοικία ανήκει στην κατηγορία 3+, η οποία αντιστοιχεί σε πιο απομακρυσμένη μικρή πόλη ή αγροτική περιοχή. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 13,560 kWh. Στον Πίνακα 32 παρουσιάζονται τα αποτελέσματα του σεναρίου χωρίς χρήση ιστορικών δεδομένων.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	119,34	186,23	17,80	12,871	13,560	0,689
XGBoost	145,86	212,82	22,72	12,290	13,560	1,270
LightGBM	155,62	210,15	27,50	13,863	13,560	0,303
Ensemble	133,50	193,02	21,38	12,994	13,560	0,566

Πίνακας 32 Απόδοση μοντέλων για την κατοικία home316 στην εαρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το Random Forest παρουσιάζει την καλύτερη απόδοση ως προς την ακρίβεια της ωριαίας καμπύλης, επιτυγχάνοντας το χαμηλότερο RMSE, ίσο με 186,23 Wh. Παράλληλα, έχει και το χαμηλότερο MAE και MAPE μεταξύ των μοντέλων. Ωστόσο, ως προς τη συνολική ημερήσια κατανάλωση, το LightGBM εμφανίζει το μικρότερο ημερήσιο σφάλμα, καθώς προβλέπει 13,863 kWh έναντι πραγματικής τιμής 13,560 kWh, με απόκλιση 0,303 kWh. Το αποτέλεσμα αυτό δείχνει ότι το καλύτερο μοντέλο ως προς την ωριαία καμπύλη δεν ταυτίζεται απαραίτητα με το καλύτερο μοντέλο ως προς τη συνολική ημερήσια εκτίμηση.



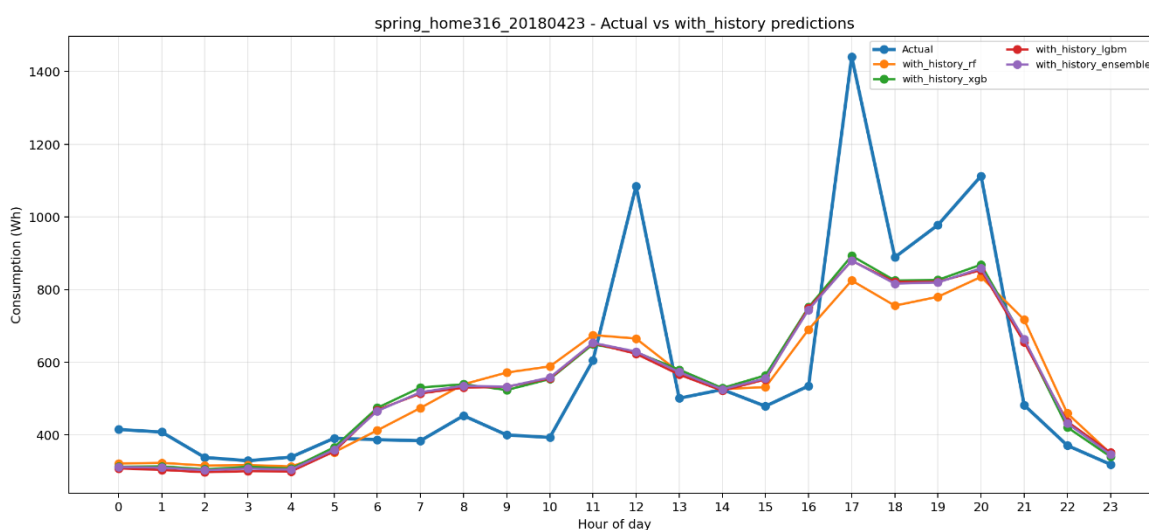
Σχήμα 28 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στην εαρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 33.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	133,13	192,91	20,94	12,913	13,560	0,647
XGBoost	127,13	181,82	20,58	13,077	13,560	0,483
LightGBM	130,11	184,76	21,20	12,949	13,560	0,611
Ensemble	128,95	183,91	20,87	13,003	13,560	0,557

Πίνακας 33 Απόδοση μοντέλων για την κατοικία home316 στην εαρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, το XGBoost παρουσιάζει την καλύτερη απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE μειώνεται σε 181,82 Wh και το ημερήσιο σφάλμα είναι 0,483 kWh. Το Ensemble και το LightGBM εμφανίζουν επίσης κοντινή απόδοση ως προς το RMSE, όμως το XGBoost διατηρεί την καλύτερη συνολική ισορροπία μεταξύ ωριαίων μετρικών και ημερήσιας εκτίμησης.



Σχήμα 29 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home316 στην εαρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η ανάλυση σταθερότητας έδειξε υψηλή σταθερότητα ιστορικού. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 14,229 kWh, η διάμεση τιμή ήταν 13,720 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 10,824–18,668 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 13,560 kWh, βρισκόταν εντός του ιστορικού εύρους και αρκετά κοντά στη διάμεση τιμή του πρόσφατου ιστορικού. Παρότι το εύρος των ημερήσιων τιμών ήταν σχετικά μεγάλο, οι δείκτες σταθερότητας έδειξαν αποδεκτή επαναληψιμότητα, ιδιαίτερα ως προς τις αιχμές, ενώ η ημερήσια και η ωριαία σταθερότητα ήταν μέτρια προς υψηλή. Με βάση την ανάλυση αυτή, το σύστημα κατέταξε το ιστορικό στην κατηγορία High με προτεινόμενη τιμή max_alpha 0,25.

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 34, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	RF	119,34	186,23	17,80	12,871	13,560	0,689
with-history adaptive	XGB	127,13	181,82	20,58	13,077	13,560	0,483

Πίνακας 34 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home316 στην εαρινή περίπτωση

Για την εαρινή περίπτωση της home316, η χρήση ιστορικών δεδομένων και προσαρμοστικής διόρθωσης οδηγεί σε μικρή βελτίωση της ωριαίας ακρίβειας. Το RMSE μειώνεται από 186,23 Wh στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων, σε 181,82 Wh στο καλύτερο μοντέλο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση. Η μείωση αυτή είναι περιορισμένη, περίπου 2,4%, ωστόσο δείχνει ότι το ιστορικό συνέβαλε σε ελαφρά βελτίωση της μορφής της ωριαίας καμπύλης. Παράλληλα, το ημερήσιο σφάλμα του καλύτερου μοντέλου με βάση το RMSE μειώνεται από 0,689 kWh σε 0,483 kWh. Ωστόσο, πρέπει να σημειωθεί ότι στο σενάριο χωρίς χρήση ιστορικών δεδομένων το LightGBM εμφάνισε ακόμη μικρότερο ημερήσιο σφάλμα, ίσο με 0,303 kWh, χωρίς όμως να έχει την καλύτερη ωριαία ακρίβεια. Επομένως, η αξιολόγηση της συγκεκριμένης περίπτωσης δείχνει ξανά ότι η επιλογή του καλύτερου μοντέλου εξαρτάται από το αν δίνεται προτεραιότητα στην ωριαία καμπύλη ή στη συνολική ημερήσια κατανάλωση.

Το αποτέλεσμα είναι συνεπές με την ανάλυση ιστορικού. Το πρόσφατο ιστορικό της home316 ήταν πλήρες και χαρακτηρίστηκε ως υψηλής σταθερότητας, ενώ η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν κοντά στη διάμεση ημερήσια κατανάλωση του ιστορικού. Για τον λόγο αυτό, το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση είχε θετική επίδραση στην πρόβλεψη. Η βελτίωση, όμως, παρέμεινε περιορισμένη, πιθανώς επειδή το ιστορικό παρουσίαζε ακόμη σημαντικό εύρος ημερήσιων τιμών και επειδή η ίδια η κατοικία εμφανίζει πιο σύνθετη και μεταβλητή ενεργειακή συμπεριφορά.

Η εαρινή περίπτωση της home316 διαφοροποιείται από την περίπτωση της ίδιας κατοικίας στο σύνολο ελέγχου του Ιουνίου. Τον Ιούνιο, το πρόσφατο ιστορικό υποδήλωνε πολύ υψηλότερο ημερήσιο επίπεδο κατανάλωσης από την ημέρα στόχο, με αποτέλεσμα το σενάριο με ιστορικά δεδομένα να βελτιώσει τις ωριαίες μετρικές αλλά να αυξήσει το ημερήσιο σφάλμα. Αντίθετα, στην εαρινή περίπτωση, η ημέρα στόχος βρίσκεται πιο κοντά στο πρόσφατο ιστορικό επίπεδο, με αποτέλεσμα η χρήση ιστορικών δεδομένων να οδηγεί σε μικρή αλλά πιο ισορροπημένη βελτίωση.

9.4 Συνολική αποτίμηση κύριας και εποχικής αξιολόγησης

Μετά την αναλυτική παρουσίαση τόσο της κύριας αξιολόγησης του Ιουνίου 2018 όσο και των εποχικών μελετών περίπτωσης, πραγματοποιείται συνολική αποτίμηση των

αποτελεσμάτων. Στόχος της ενότητας αυτής είναι να συγκεντρωθούν τα βασικά ευρήματα και να εξεταστεί συνολικά η συμπεριφορά του συστήματος στα σενάρια χωρίς χρήση ιστορικών δεδομένων και με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση. Στον Πίνακα 35 παρουσιάζονται συγκεντρωτικά τα καλύτερα αποτελέσματα της κύριας αξιολόγησης για τις τέσσερις επιλεγμένες κατοικίες του συνόλου ελέγχου του Ιουνίου 2018. Η επιλογή του καλύτερου σεναρίου έγινε με κύριο κριτήριο τη μετρική RMSE, καθώς βασικός στόχος της αξιολόγησης είναι η ακρίβεια της ωριαίας καμπύλης πρόβλεψης.

Κατοικία	Ημερομηνία	Κατηγορία ιστορικού	Καλύτερο σενάριο	Καλύτερο μοντέλο	RMSE (Wh)	Ημερήσιο σφάλμα (kWh)
home72	15/06/2018	Medium	no-history	RF	28,48	0,028
home100	28/06/2018	High	with-history adaptive	LightGBM	15,67	0,074
home268	29/06/2018	Excellent	with-history adaptive	LightGBM	8,89	0,086
home316	30/06/2018	Medium	with-history	XGBoost	86,03	1,325

Πίνακας 35 Συγκεντρωτική σύγκριση καλύτερων σεναρίων στην κύρια αξιολόγηση

Από την κύρια αξιολόγηση προκύπτει ότι η χρήση ιστορικών δεδομένων δεν έχει ομοιόμορφη επίδραση σε όλες τις κατοικίες. Στις κατοικίες home100 και home268, όπου το πρόσφατο ιστορικό ήταν σταθερό ή εξαιρετικά σταθερό και παράλληλα αντιπροσωπευτικό της ημέρας στόχου, τα σενάρια με χρήση ιστορικών δεδομένων οδήγησαν σε σαφή βελτίωση της πρόβλεψης. Ειδικά στην κατοικία home268, το RMSE μειώθηκε από 59,02 Wh στο καλύτερο μοντέλο του σεναρίου χωρίς χρήση ιστορικών δεδομένων σε 8,89 Wh μετά την εφαρμογή της προσαρμοστικής διόρθωσης, γεγονός που δείχνει την έντονη θετική επίδραση ενός αξιόπιστου ιστορικού προφίλ.

Αντίθετα, στην κατοικία home72, το ιστορικό χαρακτηρίστηκε μόνο μέτριας σταθερότητας. Παρότι η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν εντός του πρόσφατου ιστορικού εύρους, η χρήση ιστορικών δεδομένων δεν προσέφερε ουσιαστική βελτίωση. Το σενάριο χωρίς ιστορικά δεδομένα παρέμεινε οριακά καλύτερο, γεγονός που δείχνει ότι, όταν το ιστορικό δεν είναι αρκετά επαναλαμβανόμενο, τα στατικά, χρονικά και περιβαλλοντικά χαρακτηριστικά μπορούν σε ορισμένες περιπτώσεις να αποδώσουν εξίσου καλά ή και ελαφρώς καλύτερα.

Η κατοικία home316 αποτελεί διαφορετική περίπτωση. Εκεί, το σενάριο με χρήση ιστορικών δεδομένων βελτίωσε το RMSE σε σχέση με το σενάριο χωρίς χρήση ιστορικού, όμως αύξησε το ημερήσιο σφάλμα. Η ερμηνεία αυτού του αποτελέσματος συνδέεται με το γεγονός ότι το πρόσφατο ιστορικό της κατοικίας υποδήλωνε πολύ υψηλότερο ημερήσιο επίπεδο κατανάλωσης από την ημέρα στόχο. Συνεπώς, η υπερεκτίμηση της συνολικής ημερήσιας κατανάλωσης στο σενάριο με χρήση ιστορικού είναι ερμηνεύσιμη και δεν οφείλεται απαραίτητα σε λανθασμένη λειτουργία του μοντέλου.

Στον Πίνακα 36 παρουσιάζονται τα αντίστοιχα συγκεντρωτικά αποτελέσματα των εποχικών μελετών περίπτωσης. Και εδώ η επιλογή του καλύτερου σεναρίου έγινε με κύριο κριτήριο τη μετρική RMSE.

Εποχική περίπτωση	Κατηγορία ιστορικού	Καλύτερο σενάριο	Καλύτερο μοντέλο	RMSE (Wh)	Ημερήσιο σφάλμα (kWh)
home72 – Φθινόπωρο	Excellent	with-history adaptive	Ensemble	4,64	0,027
home100 – Καλοκαίρι	High	with-history adaptive	RF	26,97	0,360
home268 – Χειμώνας	Low	no-history	Ensemble	87,28	0,456
home316 – Άνοιξη	High	with-history adaptive	XGBoost	181,82	0,483

Πίνακας 36 Συγκεντρωτική σύγκριση καλύτερων σεναρίων στις εποχικές μελέτες περίπτωσης

Από τη συγκεντρωτική σύγκριση των εποχικών μελετών προκύπτει ότι το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρουσίασε καλύτερη απόδοση σε τρεις από τις τέσσερις περιπτώσεις, με βάση τη μετρική RMSE. Συγκεκριμένα, στις περιπτώσεις home72, home100 και home316, η αξιοποίηση του πρόσφατου ιστορικού οδήγησε σε βελτίωση της ωριαίας ακρίβειας σε σχέση με το σενάριο χωρίς χρήση ιστορικών δεδομένων. Αντίθετα, στη χειμερινή περίπτωση της κατοικίας home268, όπου το ιστορικό χαρακτηρίστηκε χαμηλής σταθερότητας, το σενάριο χωρίς χρήση ιστορικών δεδομένων παρέμεινε οριακά καλύτερο.

Η μεγαλύτερη βελτίωση παρατηρήθηκε στη φθινοπωρινή περίπτωση της κατοικίας home72. Εκεί, το ιστορικό ήταν εξαιρετικά σταθερό και η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν πολύ κοντά στη μέση και στη διάμεση ημερήσια κατανάλωση του πρόσφατου ιστορικού. Ως αποτέλεσμα, το RMSE μειώθηκε από 18,69 Wh σε 4,64 Wh, ενώ το ημερήσιο σφάλμα μειώθηκε από 0,363 kWh σε 0,027 kWh. Η περίπτωση αυτή δείχνει με σαφήνεια ότι όταν το πρόσφατο ιστορικό είναι πλήρες, σταθερό και αντιπροσωπευτικό, η χρήση ιστορικών δεδομένων μπορεί να βελτιώσει σημαντικά τόσο την ωριαία πρόβλεψη όσο και τη συνολική ημερήσια εκτίμηση.

Στη θερινή περίπτωση της home100, η χρήση ιστορικών δεδομένων οδήγησε επίσης σε βελτίωση, αλλά η επίδραση ήταν πιο περιορισμένη. Το ιστορικό χαρακτηρίστηκε υψηλής σταθερότητας, όμως η πραγματική ημερήσια κατανάλωση της ημέρας στόχου ήταν χαμηλότερη από το πρόσφατο ιστορικό εύρος. Για τον λόγο αυτό, το σενάριο με χρήση ιστορικών δεδομένων βελτίωσε το RMSE και το ημερήσιο σφάλμα, χωρίς όμως να οδηγήσει σε τόσο έντονη βελτίωση όσο στην περίπτωση της home72.

Η χειμερινή περίπτωση της κατοικίας home268 αποτελεί το βασικό παράδειγμα όπου η χρήση ιστορικών δεδομένων δεν προσέφερε βελτίωση. Παρότι το ιστορικό ήταν πλήρες, παρουσίαζε πολύ μεγάλη μεταβλητότητα ως προς τη συνολική ημερήσια κατανάλωση, το ωριαίο προφίλ και τις αιχμές. Το εύρος των ημερήσιων τιμών ήταν ιδιαίτερα μεγάλο, γεγονός που έδειχνε ότι οι προηγούμενες ημέρες δεν ακολουθούσαν σταθερό και επαναλαμβανόμενο πρότυπο. Σε αυτή την περίπτωση, το σύστημα κατέταξε το ιστορικό στην κατηγορία Low και πρότεινε πολύ χαμηλή τιμή max_alpha. Το αποτέλεσμα επιβεβαιώνει ότι η προσαρμοστική διόρθωση πρέπει να παραμένει περιορισμένη όταν το διαθέσιμο ιστορικό δεν είναι αξιόπιστο.

Στην εαρινή περίπτωση της κατοικίας home316, το ιστορικό χαρακτηρίστηκε υψηλής σταθερότητας και η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν

κοντά στη διάμεση ημερήσια κατανάλωση του ιστορικού. Η χρήση ιστορικών δεδομένων και προσαρμοστικής διόρθωσης οδήγησε σε μικρή βελτίωση του RMSE και σε μείωση του ημερήσιου σφάλματος του καλύτερου μοντέλου με βάση την ωριαία ακρίβεια. Ωστόσο, η βελτίωση παρέμεινε περιορισμένη, γεγονός που δείχνει ότι ακόμη και όταν το ιστορικό είναι σχετικά σταθερό, η πολυπλοκότητα της ενεργειακής συμπεριφοράς μιας κατοικίας μπορεί να περιορίσει το μέγεθος της βελτίωσης.

Συγκρίνοντας την κύρια αξιολόγηση με τις εποχικές μελέτες περίπτωσης, προκύπτει ότι η απόδοση του σεναρίου με χρήση ιστορικών δεδομένων δεν εξαρτάται μόνο από την ίδια την κατοικία, αλλά και από την ποιότητα του πρόσφατου ιστορικού που προηγείται της ημέρας πρόβλεψης. Ειδικότερα, κρίσιμοι παράγοντες είναι η σταθερότητα του ιστορικού, η επαναληψιμότητα του ωριαίου προφίλ και το κατά πόσο το ημερήσιο επίπεδο κατανάλωσης του ιστορικού είναι αντιπροσωπευτικό της ημέρας στόχου. Για παράδειγμα, στην κύρια αξιολόγηση του Ιουνίου η κατοικία home72 είχε ιστορικό μέτριας σταθερότητας και το σενάριο χωρίς ιστορικά δεδομένα παρέμεινε οριακά καλύτερο. Αντίθετα, στη φθινοπωρινή εποχική περίπτωση της ίδιας κατοικίας, το ιστορικό ήταν εξαιρετικά σταθερό και αντιπροσωπευτικό, με αποτέλεσμα το σενάριο με χρήση ιστορικών δεδομένων να υπερέχει καθαρά.

Αντίστοιχα, η κατοικία home268 παρουσίασε πολύ καλή συμπεριφορά στο σενάριο με χρήση ιστορικών δεδομένων κατά την κύρια αξιολόγηση του Ιουνίου, όπου το ιστορικό ήταν εξαιρετικά σταθερό και κοντά στην ημέρα στόχο. Στη χειμερινή εποχική περίπτωση, όμως, το ιστορικό της ίδιας κατοικίας ήταν χαμηλής σταθερότητας και μη επαναλαμβανόμενο, με αποτέλεσμα το σενάριο χωρίς ιστορικά δεδομένα να αποδώσει καλύτερα. Το εύρημα αυτό δείχνει ότι η αποτελεσματικότητα της πρόβλεψης με χρήση ιστορικών δεδομένων μπορεί να διαφοροποιείται ακόμη και για την ίδια κατοικία, ανάλογα με την περίοδο εξέτασης και τη σταθερότητα της κατανάλωσης των προηγούμενων ημερών. Αντίστοιχα συμπεράσματα προκύπτουν και για τις κατοικίες home100 και home316. Στην κατοικία home100, τόσο στην κύρια όσο και στη θερινή αξιολόγηση, το ιστορικό ήταν υψηλής σταθερότητας και το σενάριο με ιστορικά δεδομένα παρουσίασε καλύτερη απόδοση. Ωστόσο, στη θερινή περίπτωση η βελτίωση ήταν πιο περιορισμένη, επειδή η πραγματική ημερήσια κατανάλωση της ημέρας στόχου ήταν χαμηλότερη από το πρόσφατο ιστορικό εύρος. Στην κατοικία home316, η χρήση ιστορικών δεδομένων βελτίωσε την ωριαία ακρίβεια τόσο στην κύρια όσο και στην εαρινή αξιολόγηση, όμως η ερμηνεία των αποτελεσμάτων διαφοροποιείται. Τον Ιούνιο, το πρόσφατο ιστορικό υποδήλωνε πολύ υψηλότερη κατανάλωση από την ημέρα στόχο, οδηγώντας σε μεγαλύτερο ημερήσιο σφάλμα. Αντίθετα, στην εαρινή περίπτωση, η ημέρα στόχος βρισκόταν πιο κοντά στο πρόσφατο ιστορικό επίπεδο, με αποτέλεσμα η βελτίωση να είναι πιο ισορροπημένη.

Συνολικά, τα αποτελέσματα δείχνουν ότι η αποτελεσματικότητα του σεναρίου με χρήση ιστορικών δεδομένων εξαρτάται από δύο βασικούς παράγοντες. Ο πρώτος είναι η σταθερότητα του πρόσφατου ιστορικού, δηλαδή το κατά πόσο η ημερήσια κατανάλωση, το ωριαίο προφίλ και οι αιχμές παρουσιάζουν επαναλαμβανόμενη συμπεριφορά. Ο δεύτερος είναι η αντιπροσωπευτικότητα του ιστορικού ως προς την ημέρα στόχο, δηλαδή το κατά πόσο το πρόσφατο ημερήσιο επίπεδο κατανάλωσης βρίσκεται κοντά στην πραγματική ή αναμενόμενη κατανάλωση της ημέρας πρόβλεψης. Η σταθερότητα από μόνη της δεν είναι πάντα επαρκής· είναι επίσης σημαντικό το ημερήσιο επίπεδο κατανάλωσης του ιστορικού να είναι κοντά στο αναμενόμενο επίπεδο της ημέρας στόχου.

Η προσαρμοστική διόρθωση βάσει ιστορικού κατανάλωσης λειτούργησε κυρίως ως ελεγχόμενος μηχανισμός βελτίωσης της ωριαίας καμπύλης. Η επίδρασή της ήταν μεγαλύτερη όταν το ιστορικό ήταν υψηλής ή εξαιρετικής σταθερότητας και ταυτόχρονα κοντά στο πραγματικό ημερήσιο επίπεδο της ημέρας στόχου, όπως στις περιπτώσεις

home268 τον Ιούνιο και home72 το φθινόπωρο. Αντίθετα, παρέμεινε περιορισμένη ή δεν προσέφερε βελτίωση όταν το ιστορικό ήταν μέτριας ή χαμηλής σταθερότητας ή όταν δεν ήταν πλήρως αντιπροσωπευτικό του ημερήσιου επιπέδου της ημέρας πρόβλεψης. Το αποτέλεσμα αυτό είναι σύμφωνο με τη λογική του μηχανισμού $\max_alpha$, ο οποίος περιορίζει την επίδραση του ιστορικού όταν αυτό δεν θεωρείται αρκετά αξιόπιστο.

Σε πραγματικές συνθήκες χρήσης, περιπτώσεις όπου η ημέρα στόχος αναμένεται να διαφοροποιηθεί σημαντικά από το πρόσφατο ιστορικό μπορούν να αντιμετωπιστούν συμπληρωματικά μέσω της λειτουργίας συμπεριφορικής προσαρμογής (behavioral adjustment). Αν ο χρήστης γνωρίζει εκ των προτέρων ότι θα υπάρξει αλλαγή στη συμπεριφορά της κατοικίας, όπως απουσία από το σπίτι ή αυξημένη ή μειωμένη χρήση ηλεκτρικών συσκευών, μπορεί να προσαρμόσει την πρόβλεψη σε συγκεκριμένες ώρες ή χρονικές ζώνες. Με αυτόν τον τρόπο, το σύστημα μπορεί να συνδυάσει τα δεδομένα του ιστορικού με τη γνώση του χρήστη για την αναμενόμενη συμπεριφορά της ημέρας πρόβλεψης.

Επομένως, το βασικό συμπέρασμα της συνολικής αξιολόγησης είναι ότι η χρήση πρόσφατου ιστορικού κατανάλωσης μπορεί να βελτιώσει σημαντικά την πρόβλεψη, αλλά μόνο όταν εφαρμόζεται ελεγχόμενα και όταν το ιστορικό είναι τόσο σταθερό όσο και αντιπροσωπευτικό της ημέρας στόχου. Σε αντίθετη περίπτωση, η χρήση ιστορικών δεδομένων μπορεί να έχει περιορισμένο όφελος ή να βελτιώνει την ωριαία μορφή της καμπύλης χωρίς να βελτιώνει απαραίτητα τη συνολική ημερήσια εκτίμηση. Η ανάλυση σταθερότητας του ιστορικού και ο περιορισμός μέσω του $\max_alpha$ λειτουργούν ως σημαντικοί μηχανισμοί ασφαλείας, καθώς επιτρέπουν μεγαλύτερη αξιοποίηση του ιστορικού όταν αυτό είναι αξιόπιστο και περιορίζουν την επίδρασή του όταν παρουσιάζει αστάθεια ή μη αντιπροσωπευτική συμπεριφορά.

9.5 Συνολικά συμπεράσματα κεφαλαίου

Στο παρόν κεφάλαιο αξιολογήθηκε η πρακτική λειτουργία του συστήματος πρόβλεψης σε επίπεδο μεμονωμένης κατοικίας, τόσο στο ανεξάρτητο σύνολο ελέγχου όσο και σε πρόσθετες εποχικές μελέτες περίπτωσης. Η ανάλυση έδειξε ότι η απόδοση του συστήματος δεν εξαρτάται αποκλειστικά από το επιλεγμένο μοντέλο, αλλά και από τη διαθεσιμότητα, τη σταθερότητα και την αντιπροσωπευτικότητα του πρόσφατου ιστορικού κατανάλωσης.

Το βασικό συμπέρασμα που προκύπτει είναι ότι η χρήση ιστορικών δεδομένων μπορεί να βελτιώσει σημαντικά την πρόβλεψη, όταν το πρόσφατο ιστορικό είναι πλήρες, σταθερό και αντιπροσωπευτικό της ημέρας στόχου. Αντίθετα, όταν το ιστορικό παρουσιάζει χαμηλή σταθερότητα, έντονη μεταβλητότητα ή ημερήσιο επίπεδο κατανάλωσης που διαφέρει σημαντικά από την ημέρα πρόβλεψης, η επίδρασή του μπορεί να είναι περιορισμένη ή ακόμη και να οδηγήσει σε μεγαλύτερο ημερήσιο σφάλμα. Για τον λόγο αυτό, η αξιοποίηση ιστορικών δεδομένων πρέπει να γίνεται ελεγχόμενα και να συνοδεύεται από ανάλυση της ποιότητας του διαθέσιμου ιστορικού.

Η ανάλυση σταθερότητας του ιστορικού αποδείχθηκε ιδιαίτερα χρήσιμη για την ερμηνεία των αποτελεσμάτων. Οι δείκτες πληρότητας, ημερήσιας σταθερότητας, σταθερότητας ωριαίου προφίλ και σταθερότητας αιχμών επέτρεψαν την αξιολόγηση του κατά πόσο το πρόσφατο ιστορικό μπορούσε να χρησιμοποιηθεί ως αξιόπιστη βάση για την προσαρμογή της πρόβλεψης. Επιπλέον, η εμφάνιση της μέσης, της διάμεσης και του εύρους ημερήσιας κατανάλωσης βοήθησε στην καλύτερη κατανόηση του αν η ημέρα στόχος βρισκόταν κοντά ή μακριά από το πρόσφατο επίπεδο κατανάλωσης της κατοικίας.

Η προσαρμοστική διόρθωση βάσει ιστορικού κατανάλωσης λειτούργησε ως ελεγχόμενος μηχανισμός αξιοποίησης του ιστορικού. Ο περιορισμός μέσω του συντελεστή $\max_alpha$ επέτρεψε μεγαλύτερη επίδραση του ιστορικού όταν αυτό ήταν υψηλής ή εξαιρετικής σταθερότητας, ενώ περιορίσε την επίδρασή του όταν το ιστορικό ήταν μέτριας ή χαμηλής σταθερότητας. Με τον τρόπο αυτό, το σύστημα απέφυγε την υπερβολική προσαρμογή σε ασταθή ή μη επαναλαμβανόμενα ιστορικά προφίλ.

Σε επίπεδο επιλογής μοντέλου, τα αποτελέσματα δεν επιτρέπουν τη διατύπωση απόλυτου κανόνα που να ισχύει για όλες τις κατοικίες και όλες τις περιόδους. Ωστόσο, μπορούν να εξαχθούν ορισμένες πρακτικές κατευθύνσεις. Στο σενάριο χωρίς διαθέσιμα ιστορικά δεδομένα ή όταν το ιστορικό είναι χαμηλής σταθερότητας, το Random Forest μπορεί να θεωρηθεί ως μια ασφαλής και σταθερή επιλογή αναφοράς, καθώς παρουσίασε συχνά την καλύτερη ή ιδιαίτερα ανταγωνιστική απόδοση.

Αντίστοιχα, στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, το LightGBM μπορεί να θεωρηθεί ως κατάλληλη προεπιλεγμένη επιλογή όταν το πρόσφατο ιστορικό είναι υψηλής ή εξαιρετικής σταθερότητας και αντιπροσωπευτικό της ημέρας στόχου. Στις περιπτώσεις αυτές, το μοντέλο αξιοποιεί αποτελεσματικά τα χαρακτηριστικά ιστορικής κατανάλωσης και μπορεί να οδηγήσει σε σημαντική βελτίωση της ωριαίας πρόβλεψης και του ημερήσιου σφάλματος. Παρ' όλα αυτά, η επιλογή αυτή δεν πρέπει να ερμηνευθεί ως απόλυτος κανόνας, καθώς σε ορισμένες περιπτώσεις το XGBoost ή το Ensemble παρουσίασαν καλύτερη συμπεριφορά.

Για τον λόγο αυτό, σε πρακτική εφαρμογή είναι σημαντικό το σύστημα να εξακολουθεί να εμφανίζει τις προβλέψεις περισσότερων μοντέλων. Η σύγκριση μεταξύ μοντέλων δίνει στον χρήστη καλύτερη εικόνα της αβεβαιότητας της πρόβλεψης και επιτρέπει πιο ενημερωμένη επιλογή. Ως πρακτική προεπιλογή, όμως, μπορούν να χρησιμοποιούνται το Random Forest για το σενάριο χωρίς ιστορικά δεδομένα και το LightGBM για το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, υπό την προϋπόθεση ότι το ιστορικό είναι σταθερό και αντιπροσωπευτικό.

Ιδιαίτερη σημασία έχει και η λειτουργία συμπεριφορικής προσαρμογής. Σε περιπτώσεις όπου ο χρήστης γνωρίζει εκ των προτέρων ότι η συμπεριφορά της κατοικίας θα διαφοροποιηθεί από το πρόσφατο ιστορικό, για παράδειγμα λόγω απουσίας από την κατοικία ή αλλαγής στη χρήση ηλεκτρικών συσκευών, μπορεί να προσαρμόσει την πρόβλεψη σε συγκεκριμένες ώρες ή χρονικές ζώνες. Η δυνατότητα αυτή είναι σημαντική, επειδή υπάρχουν πληροφορίες για τη μελλοντική συμπεριφορά των ενοίκων που δεν είναι διαθέσιμες στα δεδομένα και δεν μπορούν να προβλεφθούν αποκλειστικά από τα μοντέλα. Έτσι, το σύστημα συνδυάζει τη μηχανική μάθηση με την εμπειρική γνώση του χρήστη.

Τα αποτελέσματα της αξιολόγησης πρέπει επίσης να ερμηνευθούν λαμβάνοντας υπόψη τη δυσκολία του προβλήματος της πρόβλεψης σε επίπεδο μεμονωμένης κατοικίας. Σε αντίθεση με την πρόβλεψη σε επίπεδο συνόλου κατοικιών, περιοχής ή υποσταθμού, όπου η συμπεριφορά πολλών καταναλωτών εξομαλύνεται μέσω της συνάθροισης, η κατανάλωση μιας μεμονωμένης κατοικίας επηρεάζεται έντονα από τυχαίους και μη πλήρως παρατηρήσιμους παράγοντες. Τέτοιοι παράγοντες είναι η παρουσία των ενοίκων, οι καθημερινές συνήθειες, η χρήση συγκεκριμένων συσκευών και οι απότομες αιχμές φορτίου. Η διεθνής βιβλιογραφία επιβεβαιώνει ότι η πρόβλεψη σε επίπεδο οικιακού καταναλωτή αποτελεί απαιτητικό πρόβλημα. Οι Sousa και Bernardo (2022), εξετάζοντας δεδομένα έξυπνων μετρητών από οικιακούς καταναλωτές, επισημαίνουν ότι η σύγκριση μεταξύ διαφορετικών εργασιών πρέπει να γίνεται με προσοχή, λόγω διαφορών στα σύνολα δεδομένων, στη χρονική ανάλυση, στον ορίζοντα πρόβλεψης και στις χρησιμοποιούμενες μετρικές. Αντίστοιχα, οι Han και Wang (2023), σε εργασία που αφορά την προσαρμοστική πρόβλεψη φορτίου σε επίπεδο μεμονωμένης κατοικίας, τονίζουν ότι η πρόβλεψη σε επίπεδο

μεμονωμένης κατοικίας χαρακτηρίζεται από έντονη μεταβλητότητα και αβεβαιότητα. Τα ευρήματα αυτά είναι συμβατά με τα αποτελέσματα της παρούσας εργασίας, όπου παρατηρήθηκαν περιπτώσεις σημαντικής βελτίωσης αλλά και περιπτώσεις περιορισμένου οφέλους από τη χρήση ιστορικών δεδομένων.

Στο πλαίσιο αυτό, τα αποτελέσματα της παρούσας εργασίας θεωρούνται ρεαλιστικά για πρόβλεψη σε επίπεδο μεμονωμένης κατοικίας. Η αξιολόγηση έδειξε ότι δεν είναι πάντα εφικτή η ταυτόχρονη βελτίωση όλων των μετρικών. Σε ορισμένες περιπτώσεις βελτιώθηκε η ωριαία ακρίβεια, όπως αποτυπώνεται στο RMSE, αλλά δεν βελτιώθηκε αντίστοιχα η συνολική ημερήσια εκτίμηση. Σε άλλες περιπτώσεις το ημερήσιο σφάλμα ήταν χαμηλό, αλλά η ωριαία καμπύλη παρουσίαζε μεγαλύτερες αποκλίσεις. Το γεγονός αυτό αναδεικνύει ότι υπάρχει συχνά συμβιβασμός μεταξύ της ακρίβειας της ωριαίας καμπύλης, της σωστής αποτύπωσης των αιχμών και της ακριβούς εκτίμησης της συνολικής ημερήσιας κατανάλωσης.

Ιδιαίτερη προσοχή απαιτεί η ερμηνεία της μετρικής MAPE. Παρότι η μετρική αυτή είναι χρήσιμη, επειδή εκφράζει το σφάλμα σε ποσοστιαία μορφή, μπορεί να λάβει υψηλές τιμές σε επίπεδο μεμονωμένης κατοικίας, ειδικά όταν οι πραγματικές τιμές κατανάλωσης είναι χαμηλές ή όταν εμφανίζονται απότομες αιχμές. Για τον λόγο αυτό, η MAPE δεν πρέπει να εξετάζεται απομονωμένα, αλλά σε συνδυασμό με τις απόλυτες μετρικές σφάλματος, όπως MAE και RMSE, καθώς και με το ημερήσιο σφάλμα σε kWh.

Στην παρούσα αξιολόγηση, οι τιμές MAPE των καλύτερων μοντέλων στο σενάριο χωρίς χρήση ιστορικών δεδομένων παρουσίασαν σημαντική διαφοροποίηση ανάλογα με την κατοικία και την εξεταζόμενη ημέρα. Η καλύτερη επίδοση στο σενάριο χωρίς ιστορικά δεδομένα εμφανίστηκε στη θερινή περίπτωση της κατοικίας home100, όπου το Random Forest πέτυχε MAPE 9,63%. Αντίστοιχα, στην κύρια αξιολόγηση του Ιουνίου, οι κατοικίες home72 και home100 παρουσίασαν MAPE 22,20% και 18,53% αντίστοιχα στα καλύτερα no-history μοντέλα. Επιπλέον, η κατοικία home316 στην κύρια αξιολόγηση παρουσίασε MAPE 30,53%, δηλαδή τιμή οριακά υψηλότερη από το όριο του 30%. Οι τιμές αυτές δείχνουν ότι, σε αρκετές περιπτώσεις, ακόμη και η πρόβλεψη χωρίς ιστορικά δεδομένα μπορεί να δώσει ικανοποιητική ή οριακά αποδεκτή σχετική ακρίβεια.

Υψηλότερες τιμές MAPE παρατηρήθηκαν κυρίως σε δυσκολότερες περιπτώσεις. Ενδεικτικά, στην κύρια αξιολόγηση της κατοικίας home268 το καλύτερο no-history μοντέλο παρουσίασε MAPE 50,89%, ενώ στη χειμερινή εποχική περίπτωση της ίδιας κατοικίας το αντίστοιχο MAPE ήταν 57,58%. Οι τιμές αυτές δείχνουν ότι το σενάριο χωρίς ιστορικά δεδομένα δυσκολεύεται περισσότερο όταν τα στατικά, χρονικά και περιβαλλοντικά χαρακτηριστικά δεν επαρκούν για την ακριβή αποτύπωση του ημερήσιου επιπέδου κατανάλωσης ή όταν η συμπεριφορά της κατοικίας παρουσιάζει έντονη μεταβλητότητα.

Η εικόνα αυτή είναι συμβατή με τη διεθνή βιβλιογραφία για την πρόβλεψη οικιακού φορτίου σε επίπεδο μεμονωμένου καταναλωτή. Οι Sousa και Bernardo (2022) επισημαίνουν ότι η άμεση σύγκριση μετρικών μεταξύ διαφορετικών εργασιών πρέπει να γίνεται με προσοχή, καθώς διαφοροποιούνται τα σύνολα δεδομένων, η χρονική ανάλυση, ο ορίζοντας πρόβλεψης και οι μονάδες μέτρησης. Παράλληλα, αναφέρουν ότι σε σχετικές εργασίες πρόβλεψης οικιακού φορτίου έχουν εμφανιστεί τιμές MAPE περίπου από 36,75% έως 55,8%, ανάλογα με το μοντέλο και το σύνολο δεδομένων. Με βάση αυτό το πλαίσιο, οι χαμηλότερες τιμές MAPE της παρούσας αξιολόγησης είναι ιδιαίτερα ικανοποιητικές, ενώ ακόμη και οι υψηλότερες τιμές βρίσκονται σε παρόμοια τάξη μεγέθους με δύσκολες περιπτώσεις που αναφέρονται στη βιβλιογραφία.

Συνεπώς, ένα MAPE άνω του 30% μπορεί να θεωρηθεί ένδειξη περιορισμένης ακρίβειας, ιδίως όταν εξετάζεται μεμονωμένα. Ωστόσο, στην πρόβλεψη ηλεκτρικής κατανάλωσης σε

επίπεδο κατοικίας, η ερμηνεία του πρέπει να γίνεται με βάση το συνολικό πλαίσιο της περίπτωσης. Η ύπαρξη ορισμένων υψηλών τιμών MAPE δεν συνεπάγεται συνολική αποτυχία του συστήματος, αλλά αναδεικνύει τις δυσκολίες της πρόβλεψης σε ατομικό οικιακό επίπεδο και την ανάγκη συνδυαστικής αξιολόγησης με MAE, RMSE, ημερήσιο σφάλμα και ανάλυση σταθερότητας ιστορικού.

Επιπλέον, η ύπαρξη σημαντικής αβεβαιότητας σε επίπεδο κατοικίας δείχνει ότι μελλοντική επέκταση του συστήματος θα μπορούσε να περιλαμβάνει πιθανοτική πρόβλεψη. Αντί για μία μόνο σημειακή πρόβλεψη, το σύστημα θα μπορούσε να εμφανίζει ένα εύρος αναμενόμενης κατανάλωσης, όπως χαμηλό, πιθανό και υψηλό σενάριο. Η κατεύθυνση αυτή είναι ιδιαίτερα χρήσιμη σε προβλήματα οικιακής κατανάλωσης, όπου η αβεβαιότητα της ανθρώπινης συμπεριφοράς είναι υψηλή και δεν μπορεί να αποτυπωθεί πλήρως από τα διαθέσιμα χαρακτηριστικά (vom Scheidt et al., 2021).

Συνολικά, το Κεφάλαιο 9 έδειξε ότι το προτεινόμενο σύστημα μπορεί να παράγει χρήσιμες και ερμηνεύσιμες προβλέψεις σε επίπεδο κατοικίας. Η πρόβλεψη χωρίς ιστορικά δεδομένα παρέχει μια πρακτική λύση όταν δεν υπάρχουν διαθέσιμες προηγούμενες μετρήσεις ή όταν το ιστορικό είναι χαμηλής ποιότητας. Αντίστοιχα, το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση μπορεί να βελτιώσει σημαντικά την πρόβλεψη όταν το ιστορικό είναι σταθερό και αντιπροσωπευτικό. Η ενσωμάτωση της ανάλυσης ιστορικού, του $\max_alpha$ και της δυνατότητας συμπεριφορικής προσαρμογής ενισχύει την πρακτική αξία του συστήματος, καθώς επιτρέπει στον χρήστη να κατανοήσει, να ελέγξει και, όπου χρειάζεται, να προσαρμόσει την τελική πρόβλεψη.

10. Το σύνολο δεδομένων PLEGMA

Το Κεφάλαιο 10 εισάγει το σύνολο δεδομένων PLEGMA, το οποίο χρησιμοποιείται στο δεύτερο μέρος της εργασίας για την εφαρμογή της προτεινόμενης μεθοδολογίας σε δεδομένα ελληνικών κατοικιών. Αρχικά, παρουσιάζεται ο λόγος ένταξης του συγκεκριμένου συνόλου δεδομένων στην εργασία και η σημασία του για τη διερεύνηση της εφαρμοσιμότητας της προσέγγισης πέρα από το περιβάλλον του IDEAL. Στη συνέχεια, περιγράφονται τα βασικά χαρακτηριστικά του PLEGMA, όπως η δομή των ηλεκτρικών και περιβαλλοντικών μετρήσεων, η χρονική ανάλυση των δεδομένων, η περίοδος καταγραφής και τα διαθέσιμα κοινωνικοδημογραφικά και κτηριακά χαρακτηριστικά. Τέλος, αποσαφηνίζεται ο ρόλος του PLEGMA στην παρούσα εργασία, καθώς δεν χρησιμοποιείται ως απλή επέκταση του IDEAL, αλλά ως ξεχωριστό πειραματικό και εφαρμοστικό πλαίσιο, στο οποίο εκπαιδεύονται εκ νέου οι ίδιοι βασικοί αλγόριθμοι μηχανικής μάθησης και υλοποιείται αντίστοιχο σύστημα πρόβλεψης προσαρμοσμένο στα διαθέσιμα δεδομένα ελληνικών κατοικιών.

10.1 Εισαγωγή

Πέρα από τη μελέτη και ανάπτυξη μοντέλων πρόβλεψης με βάση το σύνολο δεδομένων IDEAL, κρίθηκε σκόπιμη η αξιολόγηση της προτεινόμενης μεθοδολογίας σε δεδομένα που προέρχονται από πραγματικές ελληνικές κατοικίες και επιτρέπουν μια πρώτη διερεύνηση της εφαρμογής της μεθοδολογίας σε ελληνικό ενεργειακό πλαίσιο. Για τον σκοπό αυτό επιλέχθηκε το σύνολο δεδομένων PLEGMA, το οποίο περιλαμβάνει μετρήσεις ηλεκτρικής κατανάλωσης και περιβαλλοντικών παραμέτρων από πραγματικές κατοικίες στην Ελλάδα. Η ενσωμάτωση του PLEGMA στην εργασία δεν αποσκοπεί στην αντικατάσταση του IDEAL ως βασικού συνόλου ανάπτυξης της μεθοδολογίας, αλλά στη διερεύνηση του κατά πόσο η ίδια λογική πρόβλεψης μπορεί να εφαρμοστεί σε ένα διαφορετικό γεωγραφικό, κλιματικό και κοινωνικοοικονομικό πλαίσιο. Με τον τρόπο αυτό, εξετάζεται η δυνατότητα προσαρμογής της προτεινόμενης προσέγγισης σε δεδομένα ελληνικών κατοικιών, τα οποία παρουσιάζουν διαφορετικά πρότυπα κατανάλωσης, διαφορετική επίδραση των περιβαλλοντικών μεταβλητών και διαφορετική χρονική κάλυψη σε σχέση με το IDEAL. Στο πλαίσιο αυτό, το PLEGMA δεν αντιμετωπίζεται ως απλό σύνολο ελέγχου των μοντέλων που εκπαιδεύτηκαν στο IDEAL. Αντίθετα, αξιοποιείται ως δεύτερο ανεξάρτητο πειραματικό περιβάλλον, στο οποίο εφαρμόζονται εκ νέου τα βασικά στάδια της εργασίας: προεπεξεργασία και συγχρονισμός δεδομένων, δημιουργία ωριαίου συνόλου δεδομένων, διερευνητική ανάλυση, εκπαίδευση μοντέλων μηχανικής μάθησης και υλοποίηση συστήματος πρόβλεψης μέσω API και γραφικής διεπαφής χρήστη.

10.2 Περιγραφή του συνόλου δεδομένων PLEGMA

Το σύνολο δεδομένων PLEGMA αποτελεί ένα ολοκληρωμένο σύνολο δεδομένων ενεργειακής κατανάλωσης που συλλέχθηκε στο πλαίσιο ερευνητικής δραστηριότητας με στόχο την ανάλυση της κατανάλωσης ηλεκτρικής ενέργειας σε ελληνικά νοικοκυριά. Περιλαμβάνει δεδομένα από 13 κατοικίες στην Ελλάδα, οι οποίες παρουσιάζουν

διαφοροποιήσεις ως προς τα χαρακτηριστικά του νοικοκυριού, το κτήριο, την περίοδο καταγραφής και το προφίλ χρήσης ενέργειας (Athanasoulas et al., 2024). Για κάθε κατοικία, τα δεδομένα οργανώνονται σε τρεις βασικές κατηγορίες:

- **Ηλεκτρικά δεδομένα (Electric_data)**

Τα ηλεκτρικά δεδομένα περιλαμβάνουν μετρήσεις ηλεκτρικής ισχύος υψηλής χρονικής ανάλυσης, με χρονικό βήμα 10 δευτερολέπτων. Η βασική μεταβλητή που αξιοποιείται στην παρούσα εργασία είναι η συνολική ισχύς κατανάλωσης (P_{agg}), από την οποία υπολογίζεται η καταναλισκόμενη ηλεκτρική ενέργεια. Η ύπαρξη τόσο πυκνών μετρήσεων επιτρέπει λεπτομερή αποτύπωση της συμπεριφοράς της κατανάλωσης, αλλά απαιτεί κατάλληλη επεξεργασία και χρονική συνάθροιση πριν από τη χρήση των δεδομένων σε μοντέλα πρόβλεψης επόμενης ημέρας.

- **Περιβαλλοντικά δεδομένα (Environmental_data)**

Τα περιβαλλοντικά δεδομένα περιλαμβάνουν μετρήσεις εσωτερικής και εξωτερικής θερμοκρασίας, καθώς και εσωτερικής και εξωτερικής σχετικής υγρασίας. Οι μετρήσεις αυτές καταγράφονται σε διαφορετική χρονική ανάλυση από τα ηλεκτρικά δεδομένα, γεγονός που καθιστά αναγκαία τη διαδικασία συγχρονισμού και χρονικής ευθυγράμμισης. Οι περιβαλλοντικές μεταβλητές αποτελούν σημαντικούς εξωγενείς παράγοντες, καθώς σχετίζονται με τις ανάγκες θέρμανσης και ψύξης των κατοικιών και επομένως επηρεάζουν άμεσα ή έμμεσα την ηλεκτρική κατανάλωση.

- **Κοινωνικοδημογραφικά και κτηριακά χαρακτηριστικά**

Για κάθε κατοικία διατίθενται επίσης στατικά χαρακτηριστικά που αφορούν το κτήριο και το νοικοκυριό. Τα χαρακτηριστικά αυτά περιλαμβάνουν, μεταξύ άλλων, τον τύπο κατοικίας, τον αριθμό δωματίων, τον αριθμό ενοίκων, την περίοδο κατασκευής, το σύστημα θέρμανσης, το σύστημα παραγωγής ζεστού νερού, την ύπαρξη βασικών ηλεκτρικών συσκευών και άλλες πληροφορίες που μπορούν να επηρεάσουν το επίπεδο και το πρότυπο κατανάλωσης. Τα στοιχεία αυτά αξιοποιούνται ως στατικές εισοδοί στα μοντέλα πρόβλεψης, αντίστοιχα με τη χρήση μεταδεδωμένων κατοικίας στο IDEAL.

Η περίοδος καταγραφής των δεδομένων εκτείνεται γενικά από τον Ιούλιο του 2022 έως τον Σεπτέμβριο του 2023, αν και η διάρκεια και η πληρότητα των διαθέσιμων μετρήσεων διαφοροποιούνται ανά κατοικία. Η ετερογένεια αυτή αποτελεί σημαντικό χαρακτηριστικό του PLEGMA και επηρεάζει τόσο τη διαδικασία προεπεξεργασίας όσο και τη διαδικασία εκπαίδευσης και αξιολόγησης των μοντέλων. Για τον λόγο αυτό, στα επόμενα κεφάλαια πραγματοποιείται αναλυτικός έλεγχος ποιότητας, αποκλεισμός κατοικιών με εκτεταμένα προβλήματα και δημιουργία ενιαίου ωριαίου συνόλου δεδομένων κατάλληλου για μοντελοποίηση.

10.3 Ρόλος του PLEGMA dataset στην παρούσα εργασία

Ο ρόλος του PLEGMA στην παρούσα εργασία είναι συμπληρωματικός αλλά ουσιαστικός. Το IDEAL χρησιμοποιείται ως το βασικό σύνολο δεδομένων για την ανάπτυξη της μεθοδολογίας, τη διερεύνηση διαφορετικών στρατηγικών μοντελοποίησης και την αξιολόγηση της εφαρμογής σε μεγαλύτερο αριθμό κατοικιών. Το PLEGMA, από την άλλη πλευρά, αξιοποιείται ως δεύτερο πειραματικό πλαίσιο, στο οποίο εξετάζεται κατά πόσο η ίδια γενική προσέγγιση μπορεί να προσαρμοστεί σε ένα μικρότερο αλλά διαφορετικό σύνολο δεδομένων, το οποίο αφορά ελληνικές κατοικίες.

Λόγω των διαφορών στη δομή, στην κλίμακα και στη χρονική κάλυψη των δύο συνόλων δεδομένων, δεν εφαρμόστηκε μεταφορά μάθησης (transfer learning) από το IDEAL στο PLEGMA. Αντίθετα, επιλέχθηκε η εκ νέου εκπαίδευση των ίδιων βασικών αλγορίθμων μηχανικής μάθησης, δηλαδή των Random Forest, XGBoost και LightGBM, απευθείας πάνω στο τελικό επεξεργασμένο σύνολο δεδομένων PLEGMA. Με τον τρόπο αυτό, τα μοντέλα προσαρμόζονται στα χαρακτηριστικά των ελληνικών κατοικιών και στις ιδιαιτερότητες των διαθέσιμων μετρήσεων, χωρίς να εισάγονται πιθανές ασυμβατότητες από τα δεδομένα του IDEAL.

Η μεθοδολογία που εφαρμόζεται στο PLEGMA διατηρεί την ίδια βασική λογική με εκείνη του IDEAL. Εξετάζονται δύο βασικά σενάρια πρόβλεψης: το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης και το σενάριο με χρήση πρόσφατου ιστορικού κατανάλωσης. Στο πρώτο σενάριο, τα μοντέλα βασίζονται σε χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά. Στο δεύτερο, ενσωματώνονται επιπλέον ιστορικά χαρακτηριστικά, όπως υστερήσεις κατανάλωσης και κυλιόμενοι μέσοι όροι, ώστε να αξιοποιείται η πρόσφατη δυναμική της κατανάλωσης κάθε κατοικίας.

Παράλληλα, το PLEGMA χρησιμοποιείται και για την ανάπτυξη ξεχωριστής εφαρμογής πρόβλεψης, αντίστοιχης με εκείνη που δημιουργήθηκε για το IDEAL. Η εφαρμογή περιλαμβάνει εκπαιδευμένα μοντέλα, μηχανισμό επιλογής μοντέλου, API και γραφική διεπαφή χρήστη. Για το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, το σύστημα μπορεί να αξιοποιεί αρχείο ιστορικού κατανάλωσης μίας κατοικίας, επιτρέποντας πιο ρεαλιστική προσομοίωση εφαρμογής σε νέα ή επιλεγμένη κατοικία.

Με τον τρόπο αυτό, το PLEGMA συμβάλλει στη διεύρυνση της εργασίας πέρα από το αρχικό περιβάλλον του IDEAL και επιτρέπει την εξέταση της πρακτικής εφαρμογής της προτεινόμενης προσέγγισης σε ένα ελληνικό ενεργειακό πλαίσιο. Παρά το μικρότερο μέγεθος και τους περιορισμούς του συνόλου δεδομένων, η αξιοποίησή του προσφέρει χρήσιμη εικόνα για τη συμπεριφορά των μοντέλων σε διαφορετικές συνθήκες και για τις προκλήσεις που εμφανίζονται κατά την πρόβλεψη ηλεκτρικής κατανάλωσης σε επίπεδο μεμονωμένης κατοικίας.

11. Προεπεξεργασία και συγχρονισμός δεδομένων PLEGMA

Στο κεφάλαιο αυτό παρουσιάζεται η διαδικασία προεπεξεργασίας, καθαρισμού, συγχρονισμού και ενοποίησης των δεδομένων του συνόλου PLEGMA. Σε αντίθεση με το σύνολο δεδομένων IDEAL, τα αρχικά δεδομένα του PLEGMA χαρακτηρίζονται από διαφορετικές χρονικές αναλύσεις μεταξύ ηλεκτρικών και περιβαλλοντικών μετρήσεων, καθώς οι ηλεκτρικές μετρήσεις καταγράφονται ανά 10 δευτερόλεπτα, ενώ οι περιβαλλοντικές μεταβλητές είναι διαθέσιμες ανά 15 λεπτά. Για τον λόγο αυτό, απαιτήθηκε ειδική διαδικασία χρονικής ευθυγράμμισης και μετατροπής των δεδομένων σε κοινή ωριαία μορφή.

Αρχικά, περιγράφεται ο έλεγχος ποιότητας των διαθέσιμων χρονοσειρών, η διαχείριση των ελλিপών τιμών και ο αποκλεισμός κατοικιών με εκτεταμένα προβλήματα πληρότητας. Στη συνέχεια, αναλύεται η διαδικασία υπολογισμού της καταναλισκόμενης ενέργειας από τις μετρήσεις ισχύος, η ενοποίηση των ηλεκτρικών και περιβαλλοντικών δεδομένων και η μετατροπή τους σε ωριαία ανάλυση. Τέλος, παρουσιάζεται η δημιουργία του τελικού συνόλου χαρακτηριστικών, το οποίο περιλαμβάνει χρονικές, περιβαλλοντικές και στατικές μεταβλητές και αποτελεί τη βάση για τη διερευνητική ανάλυση και τη μοντελοποίηση στα επόμενα κεφάλαια.

11.1 Προετοιμασία και καθαρισμός δεδομένων

Η προεπεξεργασία των δεδομένων του PLEGMA αποτέλεσε κρίσιμο στάδιο της παρούσας εργασίας, καθώς το αρχικό σύνολο δεδομένων χαρακτηρίζεται από υψηλή χρονική ανάλυση, ετερογένεια μεταξύ των κατοικιών και παρουσία ελλিপών ή μη συνεχών καταγραφών. Στόχος της διαδικασίας ήταν η δημιουργία ενός καθαρού, χρονικά συνεπούς και ενοποιημένου συνόλου δεδομένων, κατάλληλου για την εκπαίδευση και αξιολόγηση μοντέλων πρόβλεψης ηλεκτρικής κατανάλωσης.

Η συνολική προσέγγιση σχεδιάστηκε με βάση τη μεθοδολογική λογική που εφαρμόστηκε και στο IDEAL, αλλά προσαρμόστηκε στις ιδιαιτερότητες του PLEGMA. Σε αντίθεση με το IDEAL, όπου τα δεδομένα είχαν ήδη διαμορφωθεί σε πιο άμεσα αξιοποιήσιμη ωριαία μορφή, το PLEGMA περιλαμβάνει ηλεκτρικές μετρήσεις πολύ υψηλής χρονικής ανάλυσης και περιβαλλοντικές μετρήσεις διαφορετικού χρονικού βήματος. Για τον λόγο αυτό, πριν από τη μοντελοποίηση απαιτήθηκε αναλυτικός έλεγχος ποιότητας, διαχείριση ελλিপών τιμών και χρονική ενοποίηση των διαθέσιμων αρχείων.

Αρχικά, πραγματοποιήθηκε έλεγχος της ποιότητας των δεδομένων ανά κατοικία, με έμφαση στον εντοπισμό ελλিপών τιμών, μεγάλων χρονικών κενών και μη ρεαλιστικών καταγραφών. Η ανάλυση αυτή έδειξε ότι ορισμένες κατοικίες παρουσίαζαν εκτεταμένα προβλήματα πληρότητας, τα οποία θα μπορούσαν να επηρεάσουν αρνητικά την αξιοπιστία της εκπαίδευσης και της αξιολόγησης των μοντέλων. Για τον λόγο αυτό, οι κατοικίες House_06, House_08 και House_12 αποκλείστηκαν από την περαιτέρω ανάλυση, καθώς εμφάνιζαν μεγάλο ποσοστό ελλিপών δεδομένων και δεν παρείχαν επαρκώς αξιόπιστη χρονική κάλυψη.

Για τις υπόλοιπες κατοικίες, επιλέχθηκε η συμπλήρωση των ελλিপών τιμών αντί της πλήρους διαγραφής των αντίστοιχων χρονικών διαστημάτων. Η επιλογή αυτή κρίθηκε απαραίτητη, καθώς η πλήρης αφαίρεση των διαστημάτων με ελλείψεις θα οδηγούσε σε σημαντική μείωση του διαθέσιμου όγκου δεδομένων και θα δημιουργούσε ασυνέχειες στις

χρονοσειρές. Επιπλέον, η διατήρηση συνεχών χρονοσειρών είναι ιδιαίτερα σημαντική στο σενάριο με ιστορικά δεδομένα, όπου τα μοντέλα αξιοποιούν χαρακτηριστικά χρονικής καθυστέρησης και κυλιόμενους μέσους όρους.

Η διαχείριση των ελλিপών τιμών πραγματοποιήθηκε με διαφορετικές στρατηγικές, ανάλογα με τη φύση των δεδομένων. Στα περιβαλλοντικά δεδομένα, όπως η εσωτερική και εξωτερική θερμοκρασία και υγρασία, εφαρμόστηκε γραμμική παρεμβολή, καθώς οι μεταβλητές αυτές μεταβάλλονται συνήθως ομαλά στον χρόνο. Με τον τρόπο αυτό διατηρείται η φυσική συνέχεια των περιβαλλοντικών χρονοσειρών και αποφεύγονται απότομες μη ρεαλιστικές μεταβολές.

Αντίθετα, στα ηλεκτρικά δεδομένα, όπου η κατανάλωση μπορεί να παρουσιάζει απότομες μεταβολές και αιχμές, η απλή γραμμική παρεμβολή δεν θεωρήθηκε πάντα κατάλληλη. Για τον λόγο αυτό, οι ελλείπουσες τιμές καλύφθηκαν με βάση προηγούμενες διαθέσιμες τιμές από αντίστοιχες χρονικές στιγμές της ίδιας ημέρας της εβδομάδας, όπου αυτό ήταν εφικτό. Η στρατηγική αυτή επιλέχθηκε ώστε να διατηρηθεί η εβδομαδιαία περιοδικότητα της κατανάλωσης και να αποφευχθεί η υπερβολική εξομάλυνση των προφίλ φορτίου.

Ιδιαίτερη περίπτωση αποτέλεσε η κατοικία House 13, στην οποία παρατηρήθηκε πλήρης απουσία δεδομένων για συγκεκριμένους μήνες. Για τη διατήρηση της χρονικής συνέχειας της χρονοσειράς, δημιουργήθηκαν συμπληρωματικές εγγραφές για τα ελλείποντα χρονικά διαστήματα, αξιοποιώντας διαθέσιμες τιμές από γειτονικές χρονικές περιόδους και διατηρώντας, κατά το δυνατόν, τη διάκριση μεταξύ εργάσιμων ημερών και Σαββατοκύριακου. Οι εγγραφές αυτές δεν αντιμετωπίζονται ως πραγματικές μετρήσεις, αλλά ως ανακατασκευασμένες τιμές που χρησιμοποιήθηκαν αποκλειστικά για την αποφυγή πλήρους διακοπής της χρονοσειράς. Για τον λόγο αυτό, η συγκεκριμένη επιλογή λαμβάνεται υπόψη κατά την ερμηνεία των αποτελεσμάτων.

Το τελικό αποτέλεσμα της διαδικασίας καθαρισμού ήταν η δημιουργία ενός συνεκτικού συνόλου δεδομένων ανά κατοικία, με βελτιωμένη πληρότητα και χρονική συνέχεια. Το σύνολο αυτό αποτέλεσε τη βάση για τα επόμενα στάδια της προεπεξεργασίας, δηλαδή τον συγχρονισμό των ηλεκτρικών και περιβαλλοντικών μετρήσεων, τη μετατροπή της ισχύος σε ενέργεια, τη χρονική συνάθροιση και τη δημιουργία του τελικού ωριαίου συνόλου δεδομένων που χρησιμοποιήθηκε για τη μοντελοποίηση.

11.2 Συγχρονισμός και χρονική ευθυγράμμιση δεδομένων

Η ενοποίηση και χρονική ευθυγράμμιση των δεδομένων του PLEGMA αποτέλεσε βασικό στάδιο της προεπεξεργασίας, καθώς τα διαθέσιμα δεδομένα προέρχονται από διαφορετικές πηγές και χαρακτηρίζονται από διαφορετική χρονική ανάλυση. Συγκεκριμένα, οι ηλεκτρικές μετρήσεις καταγράφονται ανά 10 δευτερόλεπτα, ενώ τα περιβαλλοντικά δεδομένα καταγράφονται ανά 15 λεπτά. Επομένως, πριν από τη δημιουργία του τελικού συνόλου δεδομένων, ήταν απαραίτητη η μετατροπή των δύο κατηγοριών μετρήσεων σε κοινό χρονικό βήμα.

Στο πλαίσιο αυτό, επιλέχθηκε αρχικά η χρονική ανάλυση των 30 λεπτών. Η επιλογή αυτή επέτρεψε τη σημαντική μείωση του όγκου των αρχικών ηλεκτρικών μετρήσεων, χωρίς να χαθεί η βασική πληροφορία για τη χρονική εξέλιξη της κατανάλωσης. Παράλληλα, το χρονικό βήμα των 30 λεπτών διευκόλυνε την ευθυγράμμιση με τα περιβαλλοντικά δεδομένα, τα οποία ήταν διαθέσιμα σε ανάλυση 15 λεπτών.

11.2.1 Υπολογισμός κατανάλωσης ενέργειας

Η βασική μεταβλητή των ηλεκτρικών δεδομένων είναι η συνολική ισχύς κατανάλωσης (P_{agg}), η οποία εκφράζεται σε Watt και καταγράφεται ανά χρονικά διαστήματα των 10 δευτερολέπτων. Για τον υπολογισμό της καταναλισκόμενης ενέργειας σε κάθε χρονικό παράθυρο εφαρμόστηκε αριθμητική ολοκλήρωση της ισχύος.

Συγκεκριμένα, για κάθε χρονικό διάστημα 30 λεπτών, η ενέργεια υπολογίστηκε ως:

$$Energy(Wh) = \frac{\sum Pi \cdot \Delta t}{3600} \quad (15)$$

όπου:

- $\sum Pi$: το άθροισμα των τιμών ισχύος (σε Watt) εντός του χρονικού διαστήματος
- Δt : η χρονική διάρκεια κάθε μέτρησης (10 δευτερόλεπτα)

Η διαίρεση με το 3600 χρησιμοποιείται για τη μετατροπή της ενέργειας από Joules σε Watt-hours (Wh).

Λαμβάνοντας υπόψη ότι $\Delta t = 10$ δευτερόλεπτα, ο παραπάνω τύπος απλοποιείται ως εξής:

$$Energy(Wh) = \frac{\sum Pi \cdot 10}{3600} = \frac{\sum Pi}{360} \quad (16)$$

Με τον τρόπο αυτό, η συνολική κατανάλωση ενέργειας για κάθε 30λεπτο προκύπτει άμεσα από το άθροισμα των μετρήσεων ισχύος.

Η χρονοσήμανση κάθε εγγραφής αντιστοιχίστηκε στο τέλος του χρονικού διαστήματος (π.χ. το διάστημα 00:00–00:29:59 αντιστοιχεί σε timestamp 00:30), ώστε να διατηρείται συνέπεια με τη λογική της συσσωρευτικής κατανάλωσης και να διασφαλίζεται η σωστή χρονική ευθυγράμμιση με τις υπόλοιπες μεταβλητές.

11.2.2 Ενοποίηση περιβαλλοντικών δεδομένων

Τα περιβαλλοντικά δεδομένα περιλαμβάνουν μετρήσεις εσωτερικής και εξωτερικής θερμοκρασίας, καθώς και εσωτερικής και εξωτερικής σχετικής υγρασίας. Οι μετρήσεις αυτές είναι διαθέσιμες σε χρονική ανάλυση 15 λεπτών. Για την προσαρμογή τους στο κοινό χρονικό βήμα των 30 λεπτών, οι τιμές των περιβαλλοντικών μεταβλητών υπολογίστηκαν ως μέσοι όροι των αντίστοιχων 15λεπτων μετρήσεων που ανήκαν σε κάθε 30λεπτο διάστημα.

11.2.3 Συγχώνευση δεδομένων

Μετά την ομογενοποίηση της χρονικής ανάλυσης, τα ηλεκτρικά και τα περιβαλλοντικά δεδομένα συγχωνεύθηκαν με βάση τη χρονοσήμανση. Το αποτέλεσμα της διαδικασίας ήταν η δημιουργία ενός ενοποιημένου συνόλου δεδομένων ανά κατοικία, το οποίο περιλαμβάνει τη συνολική κατανάλωση ενέργειας ανά χρονικό διάστημα, τις περιβαλλοντικές μεταβλητές και πλήρη χρονική αντιστοίχιση μεταξύ των διαφορετικών πηγών δεδομένων.

Τα ενοποιημένα αρχεία ανά κατοικία αποτέλεσαν τη βάση για την επόμενη φάση επεξεργασίας, δηλαδή τη μετατροπή των δεδομένων σε ωριαία ανάλυση και την προσθήκη των χαρακτηριστικών που χρησιμοποιήθηκαν στα μοντέλα πρόβλεψης.

11.2.4 Μετατροπή σε ωριαία ανάλυση

Δεδομένου ότι η μεθοδολογία του IDEAL, καθώς και η τελική λογική του API και της γραφικής διεπαφής χρήστη, βασίζονται σε ωριαίες προβλέψεις επόμενης ημέρας, κρίθηκε απαραίτητη η μετατροπή των δεδομένων του PLEGMA σε αντίστοιχη ωριαία μορφή. Η επιλογή αυτή διασφαλίζει συνέπεια μεταξύ των δύο πειραματικών πλαισίων και επιτρέπει την εφαρμογή παρόμοιας λογικής μοντελοποίησης και αξιολόγησης.

Για τον σκοπό αυτό, η κατανάλωση ενέργειας αθροίστηκε εντός κάθε ώρας, ώστε να παραχθούν ωριαίες τιμές κατανάλωσης. Αντίστοιχα, οι περιβαλλοντικές μεταβλητές υπολογίστηκαν ως ωριαίοι μέσοι όροι των διαθέσιμων τιμών. Η μεταβλητή ενέργειας μετονομάστηκε σε `consumption_Wh`, ώστε να υπάρχει συνέπεια με την ονοματολογία που χρησιμοποιήθηκε και στο IDEAL.

Η διαδικασία αυτή οδήγησε στη δημιουργία ενός ενιαίου ωριαίου συνόλου δεδομένων, το οποίο αποτελεί την τελική είσοδο για τη διερευνητική ανάλυση και τη μοντελοποίηση. Η σωστή χρονική αντιστοίχιση της κατανάλωσης με τις περιβαλλοντικές μεταβλητές αποτελεί βασική προϋπόθεση για την εκπαίδευση αξιόπιστων μοντέλων πρόβλεψης, καθώς επιτρέπει στα μοντέλα να αξιοποιούν με συνέπεια τόσο τις χρονικές όσο και τις εξωγενείς περιβαλλοντικές πληροφορίες.

11.3 Μετασχηματισμός χαρακτηριστικών και δημιουργία τελικού συνόλου δεδομένων

Μετά την ολοκλήρωση του καθαρισμού, του συγχρονισμού και της χρονικής ευθυγράμμισης των δεδομένων, ακολούθησε το στάδιο του μετασχηματισμού των χαρακτηριστικών (feature engineering) και της δημιουργίας του τελικού συνόλου δεδομένων που χρησιμοποιήθηκε για τη διερευνητική ανάλυση και τη μοντελοποίηση.

Η διαδικασία αυτή σχεδιάστηκε με στόχο τη διατήρηση της συνέπειας με τη μεθοδολογία που εφαρμόστηκε στο σύνολο δεδομένων IDEAL, αλλά και την προσαρμογή στις ιδιαιτερότητες του συνόλου δεδομένων PLEGMA. Συγκεκριμένα, δημιουργήθηκε ένα ενιαίο ωριαίο σύνολο δεδομένων, στο οποίο κάθε γραμμή αντιστοιχεί σε μία κατοικία και μία ωριαία χρονική στιγμή. Το σύνολο αυτό περιλαμβάνει τη μεταβλητή στόχο, δηλαδή την ωριαία ηλεκτρική κατανάλωση, καθώς και χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά.

Η δημιουργία των βασικών χαρακτηριστικών πραγματοποιήθηκε με τρόπο που να υποστηρίζει τα δύο σενάρια πρόβλεψης που εξετάζονται στην εργασία: το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης και το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης. Στο πρώτο σενάριο αξιοποιούνται μόνο χαρακτηριστικά που είναι διαθέσιμα χωρίς προηγούμενες μετρήσεις κατανάλωσης, όπως χρονικές, περιβαλλοντικές και στατικές μεταβλητές. Στο δεύτερο σενάριο, κατά τη φάση διαμόρφωσης των δεδομένων εκπαίδευσης των προβλεπτικών μοντέλων, προστίθενται επιπλέον ιστορικά χαρακτηριστικά, τα οποία αποτυπώνουν την πρόσφατη δυναμική της κατανάλωσης.

11.3.1 Χρονικά χαρακτηριστικά

Στο ενοποιημένο ωριαίο σύνολο δεδομένων προστέθηκαν βασικά χρονικά χαρακτηριστικά, τα οποία αποτυπώνουν την περιοδικότητα της κατανάλωσης. Συγκεκριμένα, δημιουργήθηκαν οι ακόλουθες μεταβλητές:

- ώρα της ημέρας (hour),
- ημέρα της εβδομάδας (dayofweek),
- ένδειξη Σαββατοκύριακου (is_weekend),
- μήνας (month),
- εποχή (season),
- ένδειξη επίσημης αργίας (is_holiday).

Τα χαρακτηριστικά αυτά επιτρέπουν στα μοντέλα να αναγνωρίζουν επαναλαμβανόμενα πρότυπα κατανάλωσης, όπως διαφοροποιήσεις μεταξύ πρωινών και βραδινών ωρών, καθημερινών και Σαββατοκύριακων, καθώς και εποχιακές διαφορές. Η χρήση της ένδειξης αργίας είναι επίσης σημαντική, καθώς οι ημέρες αυτές μπορεί να παρουσιάζουν διαφορετική συμπεριφορά κατανάλωσης σε σχέση με τις τυπικές εργάσιμες ημέρες.

11.3.2 Περιβαλλοντικά χαρακτηριστικά

Στο τελικό σύνολο δεδομένων ενσωματώθηκαν οι βασικές περιβαλλοντικές μεταβλητές:

- εξωτερική θερμοκρασία (external_temperature),
- εσωτερική θερμοκρασία (internal_temperature),
- εξωτερική υγρασία (external_humidity),
- εσωτερική υγρασία (internal_humidity).

Οι μεταβλητές αυτές αποτελούν κρίσιμους εξωγενείς παράγοντες, καθώς επηρεάζουν άμεσα τις ανάγκες θέρμανσης και ψύξης των κατοικιών και, κατ' επέκταση, την κατανάλωση ηλεκτρικής ενέργειας.

11.3.3 Στατικά χαρακτηριστικά κατοικίας και νοικοκυριού

Παράλληλα, στο τελικό σύνολο δεδομένων ενσωματώθηκαν στατικά χαρακτηριστικά που προκύπτουν από τα κοινωνικοδημογραφικά και κτηριακά δεδομένα του συνόλου δεδομένων PLEGMA. Τα χαρακτηριστικά αυτά περιγράφουν τη γενική δομή της κατοικίας και του νοικοκυριού και χρησιμοποιούνται ώστε τα μοντέλα να διαφοροποιούν τις προβλέψεις τους ανάλογα με τα βασικά χαρακτηριστικά κάθε περίπτωσης. Τα χαρακτηριστικά αυτά περιλαμβάνουν πληροφορίες όπως:

- τύπος κατοικίας (building_type),
- αριθμός δωματίων (num_rooms),
- αριθμός ενοίκων (residents),
- περίοδος κατασκευής (build_era),
- σύστημα θέρμανσης (heating_type),
- σύστημα παραγωγής ζεστού νερού (water_heater_type),
- ύπαρξη βασικών ηλεκτρικών συσκευών.

Η ενσωμάτωση αυτών των μεταβλητών είναι ιδιαίτερα σημαντική στο σενάριο χωρίς χρήση ιστορικών δεδομένων, όπου το μοντέλο δεν έχει πρόσβαση σε προηγούμενες τιμές κατανάλωσης. Σε αυτή την περίπτωση, τα στατικά χαρακτηριστικά βοηθούν το μοντέλο να

εκτιμήσει το γενικό επίπεδο κατανάλωσης μιας κατοικίας με βάση τη δομή και τα χαρακτηριστικά του νοικοκυριού.

11.3.4 Διαχείριση ανωμαλιών και μη ρεαλιστικών τιμών

Κατά τη δημιουργία του τελικού συνόλου δεδομένων πραγματοποιήθηκε επιπλέον έλεγχος για ανωμαλίες στις χρονοσειρές κατανάλωσης, όπως εκτεταμένα διαστήματα μηδενικών τιμών. Σε περιπτώσεις όπου τα διαστήματα αυτά δεν ήταν ρεαλιστικά (π.χ. συνεχόμενη μηδενική κατανάλωση για μεγάλο χρονικό διάστημα), αντιμετωπίστηκαν ως ελλιπή δεδομένα και αντικαταστάθηκαν μέσω παρεμβολής.

Η επιλογή αυτή κρίθηκε απαραίτητη, καθώς τέτοιου είδους φαινόμενα δεν αντανακλούν πραγματική συμπεριφορά κατανάλωσης, αλλά πιθανές αστοχίες καταγραφής ή εξωτερικά γεγονότα, τα οποία δεν μπορούν να προβλεφθούν από τα μοντέλα.

11.3.5 Δημιουργία τελικού συνόλου δεδομένων

Το τελικό αποτέλεσμα της διαδικασίας ήταν η δημιουργία ενός ενιαίου ωριαίου συνόλου δεδομένων, στο οποίο κάθε γραμμή αντιστοιχεί σε μία κατοικία και μία συγκεκριμένη ώρα. Το σύνολο αυτό περιλαμβάνει τη μεταβλητή στόχο (*consumption_Wh*), χρονικά χαρακτηριστικά, περιβαλλοντικά χαρακτηριστικά και στατικά χαρακτηριστικά κατοικίας και νοικοκυριού. Η διαμόρφωσή του σε ωριαία ανάλυση διασφαλίζει συνέπεια με το αντίστοιχο σύνολο δεδομένων του IDEAL και διευκολύνει την ανάπτυξη κοινής λογικής πρόβλεψης, αξιολόγησης, API και γραφικής διεπαφής χρήστη.

12. Διερευνητική ανάλυση δεδομένων (EDA) στο PLEGMA

Στο κεφάλαιο αυτό παρουσιάζεται η διερευνητική ανάλυση δεδομένων του τελικού ωριαίου συνόλου δεδομένων PLEGMA, με στόχο την κατανόηση των βασικών προτύπων ηλεκτρικής κατανάλωσης και τη σύνδεσή τους με τα χαρακτηριστικά που χρησιμοποιούνται στη μοντελοποίηση. Μετά την ολοκλήρωση της προεπεξεργασίας, του συγχρονισμού και της δημιουργίας του τελικού συνόλου δεδομένων, η διερευνητική ανάλυση επιτρέπει την αποτύπωση της χρονικής συμπεριφοράς της κατανάλωσης και την ανάδειξη των παραγόντων που επηρεάζουν τη ζήτηση ηλεκτρικής ενέργειας στις κατοικίες του PLEGMA.

Αρχικά, εξετάζονται τα ημερήσια και ωριαία προφίλ κατανάλωσης, ώστε να αναγνωριστούν επαναλαμβανόμενα μοτίβα μέσα στην ημέρα. Στη συνέχεια, αναλύονται οι διαφοροποιήσεις μεταξύ καθημερινών και Σαββατοκύριακου, καθώς οι διαφορετικές συνήθειες των ενοίκων μπορούν να οδηγήσουν σε μεταβολές τόσο στο επίπεδο όσο και στη χρονική κατανομή της κατανάλωσης. Επιπλέον, διερευνάται η σχέση της ηλεκτρικής κατανάλωσης με την εξωτερική θερμοκρασία, η οποία αποτελεί σημαντικό περιβαλλοντικό παράγοντα, ιδιαίτερα σε ελληνικές κατοικίες όπου η χρήση συστημάτων ψύξης μπορεί να επηρεάζει έντονα το φορτίο κατά τις θερμές περιόδους.

Τέλος, εξετάζεται η κατανομή της κατανάλωσης και τα βασικά στατιστικά χαρακτηριστικά της, με στόχο την κατανόηση της μεταβλητότητας του συνόλου δεδομένων και των δυσκολιών που προκύπτουν κατά την πρόβλεψη. Η ανάλυση αυτή λειτουργεί ως ενδιάμεσο στάδιο μεταξύ της προεπεξεργασίας και της μοντελοποίησης, καθώς τεκμηριώνει τη σημασία των χρονικών, περιβαλλοντικών και στατικών χαρακτηριστικών που χρησιμοποιούνται στα επόμενα κεφάλαια.

12.1 Προσέγγιση διερευνητικής ανάλυσης

Η διερευνητική ανάλυση πραγματοποιήθηκε πάνω στο τελικό ωριαίο σύνολο δεδομένων του PLEGMA, το οποίο προέκυψε μετά τη διαδικασία καθαρισμού, συγχρονισμού και χρονικής ενοποίησης των ηλεκτρικών και περιβαλλοντικών μετρήσεων. Στόχος της ανάλυσης ήταν η κατανόηση της γενικής συμπεριφοράς της ηλεκτρικής κατανάλωσης πριν από τη φάση της μοντελοποίησης, καθώς και η τεκμηρίωση της επιλογής των βασικών χαρακτηριστικών εισόδου.

Δεδομένου ότι οι κατοικίες του PLEGMA δεν διαθέτουν όλες την ίδια χρονική διάρκεια καταγραφής, οι συγκεντρωτικές απεικονίσεις ερμηνεύονται με προσοχή. Κατοικίες με μεγαλύτερη χρονική κάλυψη μπορούν να επηρεάσουν περισσότερο τους συνολικούς μέσους όρους, όταν η ανάλυση βασίζεται στο σύνολο των διαθέσιμων ωριαίων εγγραφών. Για τον λόγο αυτό, τα ευρήματα της διερευνητικής ανάλυσης χρησιμοποιούνται κυρίως για την κατανόηση γενικών τάσεων του συνόλου δεδομένων και όχι για την εξαγωγή συμπερασμάτων σε επίπεδο μεμονωμένης κατοικίας.

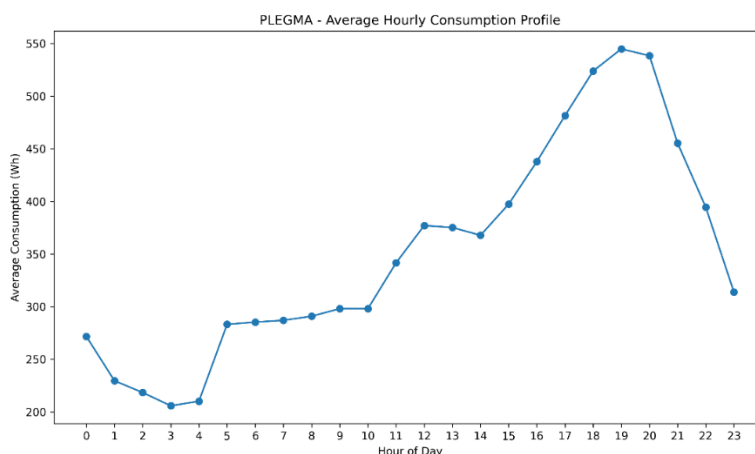
Η ανάλυση βασίστηκε κυρίως σε συγκεντρωτικές απεικονίσεις και στατιστικές περιγραφές της κατανάλωσης. Εξετάστηκαν η μέση ωριαία κατανάλωση, οι διαφορές μεταξύ καθημερινών και Σαββατοκύριακου, η σχέση της κατανάλωσης με την εξωτερική θερμοκρασία και η συνολική κατανομή των τιμών κατανάλωσης. Με τον τρόπο αυτό αξιολογήθηκε κατά πόσο το σύνολο δεδομένων παρουσιάζει χρονικά και περιβαλλοντικά πρότυπα που μπορούν να αξιοποιηθούν από τα μοντέλα πρόβλεψης.

Η προσέγγιση που ακολουθήθηκε είναι αντίστοιχη με εκείνη που εφαρμόστηκε στο IDEAL, ώστε να διατηρείται η μεθοδολογική συνέπεια μεταξύ των δύο συνόλων δεδομένων. Ωστόσο, η ανάλυση προσαρμόστηκε στις ιδιαιτερότητες του PLEGMA, όπως το μικρότερο πλήθος κατοικιών, η διαφοροποιημένη χρονική κάλυψη ανά κατοικία και η διαθεσιμότητα διαφορετικών χαρακτηριστικών εισόδου σε σχέση με το IDEAL.

Τα ευρήματα αξιοποιούνται στα επόμενα κεφάλαια για την ερμηνεία της συμπεριφοράς των μοντέλων και των δυσκολιών που εμφανίζονται κατά την πρόβλεψη της οικιακής κατανάλωσης.

12.2 Ημερήσια και ωριαία προφίλ κατανάλωσης

Η ανάλυση της μέσης ωριαίας κατανάλωσης στο σύνολο δεδομένων PLEGMA ανέδειξε σαφή ημερήσια περιοδικότητα. Όπως παρουσιάζεται στο Σχήμα 30, η κατανάλωση παραμένει σε χαμηλότερα επίπεδα κατά τις νυχτερινές ώρες, όταν η δραστηριότητα των ενοίκων είναι περιορισμένη. Στη συνέχεια, παρατηρείται σταδιακή αύξηση κατά τις πρωινές ώρες, η οποία μπορεί να συνδεθεί με την έναρξη της ημερήσιας δραστηριότητας των νοικοκυριών.



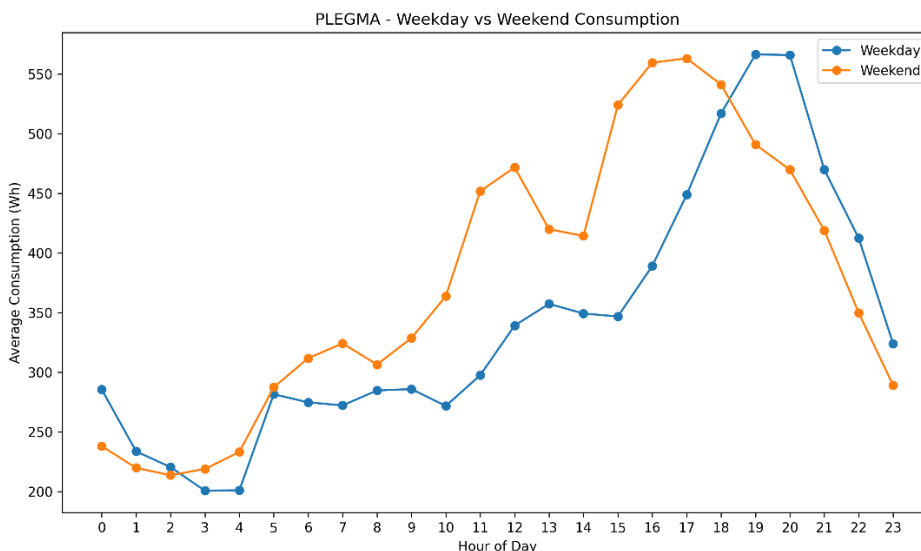
Σχήμα 30 Μέση ωριαία κατανάλωση ηλεκτρικής ενέργειας στο σύνολο δεδομένων PLEGMA

Κατά τις μεσημεριανές ώρες η κατανάλωση εμφανίζει σχετική σταθεροποίηση ή μικρή υποχώρηση, ενώ αυξάνεται εκ νέου από το απόγευμα και κυρίως κατά τις βραδινές ώρες. Η βραδινή περίοδος παρουσιάζει την υψηλότερη αιχμή της ημέρας, γεγονός που συνδέεται με την επιστροφή των ενοίκων στην κατοικία και τη συγκέντρωση δραστηριοτήτων που απαιτούν χρήση ηλεκτρικών συσκευών, όπως μαγείρεμα, φωτισμός, ψυχαγωγία και λειτουργία συστημάτων θέρμανσης ή ψύξης.

Το παρατηρούμενο πρότυπο είναι συμβατό με τη γενική συμπεριφορά οικιακής κατανάλωσης, όπου η ζήτηση ηλεκτρικής ενέργειας επηρεάζεται έντονα από το ημερήσιο πρόγραμμα των χρηστών. Η ύπαρξη σαφούς ημερήσιας περιοδικότητας επιβεβαιώνει τη σημασία των χρονικών χαρακτηριστικών και ιδιαίτερα της ώρας της ημέρας, στη διαδικασία πρόβλεψης. Τα μοντέλα μηχανικής μάθησης μπορούν να αξιοποιήσουν τα χαρακτηριστικά αυτά ώστε να αναγνωρίσουν επαναλαμβανόμενα μοτίβα και να προσαρμόσουν την πρόβλεψη ανάλογα με τη χρονική στιγμή της ημέρας.

12.3 Σύγκριση καθημερινών και Σαββατοκύριακου

Η διαφοροποίηση της κατανάλωσης μεταξύ καθημερινών και Σαββατοκύριακου παρουσιάζεται στο Σχήμα 31, όπου απεικονίζεται η μέση ωριαία κατανάλωση για τις δύο κατηγορίες ημερών.



Σχήμα 31 Μέση ωριαία κατανάλωση για καθημερινές και Σαββατοκύριακα στο σύνολο δεδομένων PLEGMA

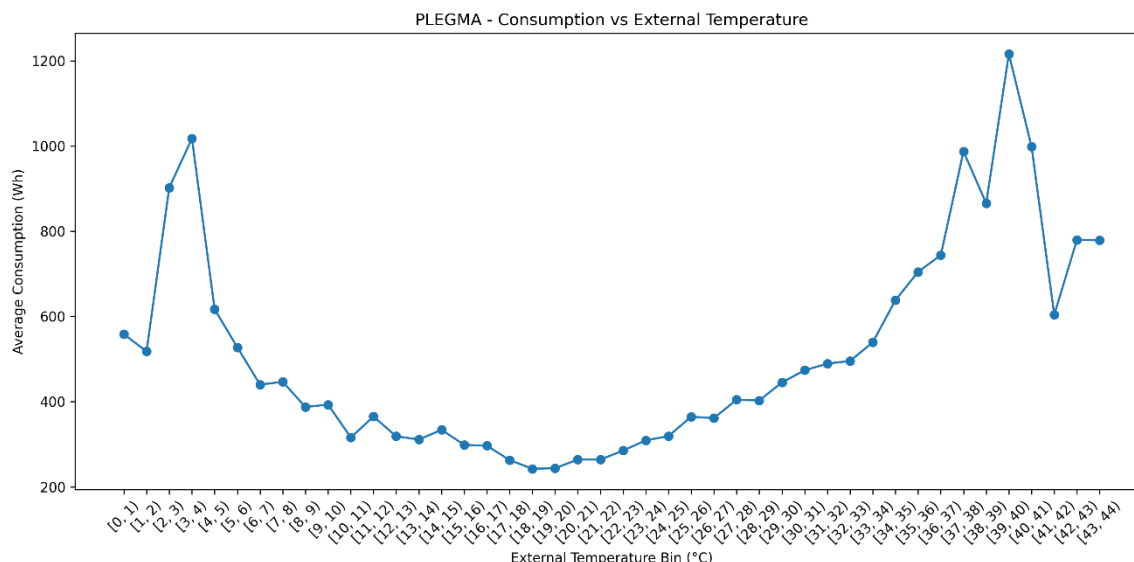
Παρατηρείται ότι κατά τις καθημερινές ημέρες η κατανάλωση εμφανίζει εντονότερη αύξηση κατά τις απογευματινές και βραδινές ώρες, με κορύφωση περίπου μεταξύ 18:00 και 21:00. Το πρότυπο αυτό μπορεί να συνδεθεί με την επιστροφή των ενοίκων στην κατοικία μετά την εργασία ή τις καθημερινές δραστηριότητες, καθώς και με τη συγκέντρωση ενεργειακών χρήσεων μέσα σε περιορισμένο χρονικό διάστημα.

Αντίθετα, κατά τα Σαββατοκύριακα η κατανάλωση εμφανίζεται σχετικά αυξημένη και κατά τις μεσημεριανές ώρες, περίπου μεταξύ 11:00 και 15:00. Η συμπεριφορά αυτή υποδηλώνει μεγαλύτερη παραμονή των ενοίκων στην κατοικία και πιο ομοιόμορφη κατανομή των δραστηριοτήτων κατά τη διάρκεια της ημέρας. Παράλληλα, η βραδινή αιχμή παραμένει εμφανής, αλλά παρουσιάζει μικρότερη συγκέντρωση σε σχέση με τις καθημερινές.

Η διαφοροποίηση αυτή δείχνει ότι η ηλεκτρική κατανάλωση επηρεάζεται σημαντικά από το ημερήσιο πρόγραμμα και τις συνήθειες των χρηστών. Για τον λόγο αυτό, η διάκριση μεταξύ καθημερινών και Σαββατοκύριακου αποτελεί χρήσιμο χρονικό χαρακτηριστικό εισόδου στα μοντέλα πρόβλεψης, καθώς επιτρέπει την καλύτερη αποτύπωση διαφορετικών προτύπων χρήσης της κατοικίας.

12.4 Επίδραση εξωτερικής θερμοκρασίας στην κατανάλωση

Η επίδραση της εξωτερικής θερμοκρασίας στην κατανάλωση ηλεκτρικής ενέργειας παρουσιάζεται στο Σχήμα 32, όπου απεικονίζεται η μέση κατανάλωση για διαφορετικά εύρη τιμών εξωτερικής θερμοκρασίας.



Σχήμα 32 Μέση κατανάλωση ηλεκτρικής ενέργειας σε συνάρτηση με την εξωτερική θερμοκρασία στο σύνολο δεδομένων PLEGMA

Από την ανάλυση προκύπτει ότι η κατανάλωση παρουσιάζει εμφανή σχέση με την εξωτερική θερμοκρασία. Συγκεκριμένα, παρατηρούνται αυξημένες τιμές κατανάλωσης τόσο σε χαμηλές όσο και σε υψηλές θερμοκρασίες, ενώ στα ενδιάμεσα επίπεδα θερμοκρασίας η μέση κατανάλωση εμφανίζεται χαμηλότερη. Η μορφή αυτή υποδηλώνει μη γραμμική σχέση μεταξύ κατανάλωσης και θερμοκρασίας, η οποία προσεγγίζει καμπύλη τύπου U.

Στις χαμηλές θερμοκρασίες, η αυξημένη κατανάλωση μπορεί να σχετίζεται με τη χρήση συστημάτων θέρμανσης ή βοηθητικών ηλεκτρικών συσκευών. Αντίστοιχα, στις υψηλές θερμοκρασίες, η αύξηση της κατανάλωσης μπορεί να συνδέεται με τη χρήση κλιματιστικών μονάδων και γενικότερα συστημάτων ψύξης, τα οποία αποτελούν σημαντικό παράγοντα ηλεκτρικού φορτίου σε θερμότερα κλιματικά περιβάλλοντα. Η σχέση αυτή δεν πρέπει να ερμηνευθεί ως αποκλειστικά αιτιακή, καθώς η ηλεκτρική κατανάλωση επηρεάζεται ταυτόχρονα από χρονικούς, εποχικούς και συμπεριφορικούς παράγοντες. Παρ' όλα αυτά, η αύξηση της μέσης κατανάλωσης σε χαμηλές και υψηλές θερμοκρασίες υποδηλώνει ότι η εξωτερική θερμοκρασία αποτελεί σημαντικό περιβαλλοντικό χαρακτηριστικό εισόδου για τα μοντέλα πρόβλεψης.

Η συμπεριφορά αυτή παρουσιάζει ενδιαφέρον σε σχέση με το IDEAL, καθώς στο PLEGMA η σχέση της κατανάλωσης με την εξωτερική θερμοκρασία εμφανίζεται εντονότερη και στις υψηλές θερμοκρασίες. Το εύρημα αυτό είναι συμβατό με το ελληνικό κλιματικό πλαίσιο, όπου οι ανάγκες ψύξης κατά τους θερμούς μήνες μπορούν να επηρεάσουν σημαντικά την ηλεκτρική κατανάλωση. Αντίθετα, στο IDEAL η επίδραση των

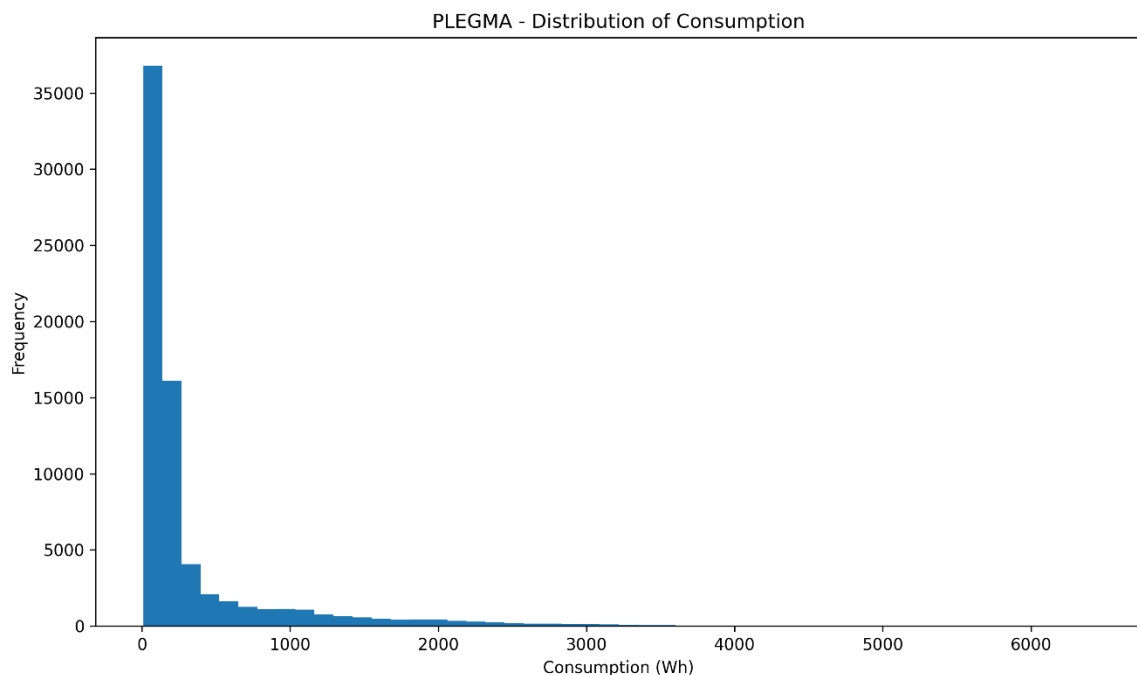
χαμηλών θερμοκρασιών είναι πιο χαρακτηριστική, καθώς το σύνολο δεδομένων προέρχεται από διαφορετικό γεωγραφικό και κλιματικό περιβάλλον.

Η παρατηρούμενη μη γραμμική σχέση κατανάλωσης και εξωτερικής θερμοκρασίας υποδηλώνει ότι τα μοντέλα πρόβλεψης πρέπει να είναι ικανά να αποτυπώνουν σύνθετες αλληλεπιδράσεις μεταξύ περιβαλλοντικών μεταβλητών και ηλεκτρικού φορτίου. Το εύρημα αυτό ενισχύει τη χρησιμότητα αλγορίθμων βασισμένων σε δέντρα αποφάσεων, όπως Random Forest, XGBoost και LightGBM, οι οποίοι μπορούν να χειριστούν μη γραμμικές σχέσεις χωρίς να απαιτείται εκ των προτέρων γραμμική υπόθεση.

Η ανάλυση επιβεβαιώνει τη σημασία της εξωτερικής θερμοκρασίας ως βασικού περιβαλλοντικού χαρακτηριστικού εισόδου στα μοντέλα πρόβλεψης. Παράλληλα, αναδεικνύει την ανάγκη προσαρμογής της μεθοδολογίας στο εκάστοτε γεωγραφικό και ενεργειακό πλαίσιο, καθώς οι περιβαλλοντικοί παράγοντες δεν επηρεάζουν με τον ίδιο τρόπο την κατανάλωση σε όλα τα σύνολα δεδομένων.

12.5 Κατανομή κατανάλωσης και βασικά στατιστικά χαρακτηριστικά

Η κατανομή της κατανάλωσης ηλεκτρικής ενέργειας παρουσιάζεται στο Σχήμα 33.



Σχήμα 33 Κατανομή της κατανάλωσης ηλεκτρικής ενέργειας στο σύνολο δεδομένων PLEGMA

Από την κατανομή παρατηρείται έντονη ασυμμετρία προς τα δεξιά, με μεγάλο αριθμό χαμηλών τιμών κατανάλωσης και σταδιακή μείωση της συχνότητας όσο αυξάνεται το επίπεδο κατανάλωσης. Η μορφή αυτή είναι χαρακτηριστική σε δεδομένα οικιακής ηλεκτρικής κατανάλωσης, όπου οι περισσότερες ώρες αντιστοιχούν σε χαμηλό ή μέτριο φορτίο, ενώ οι υψηλές τιμές εμφανίζονται πιο σπάνια και συνδέονται με περιόδους έντονης δραστηριότητας ή λειτουργίας ενεργοβόρων συσκευών.

Η συγκέντρωση πολλών τιμών σε χαμηλά επίπεδα μπορεί να συνδεθεί με την παρουσία βασικού φορτίου, το οποίο αντιστοιχεί σε περιόδους περιορισμένης δραστηριότητας, όπως οι νυχτερινές ώρες ή ώρες απουσίας των ενοίκων από την κατοικία. Παράλληλα, πολύ χαμηλές ή μηδενικές τιμές μπορεί σε ορισμένες περιπτώσεις να σχετίζονται με ιδιαιτερότητες της καταγραφής ή της προεπεξεργασίας, όπως τεχνικά κενά, μηδενικές μετρήσεις ή την επίδραση της χρονικής συνάθροισης.

Αντίθετα, οι υψηλές τιμές κατανάλωσης εμφανίζονται με σημαντικά μικρότερη συχνότητα και αντιστοιχούν σε περιόδους αιχμής. Οι αιχμές αυτές μπορούν να προκύπτουν από τη λειτουργία ενεργοβόρων συσκευών, συστημάτων θέρμανσης ή ψύξης, ή από τη συγκέντρωση πολλών ενεργειακών δραστηριοτήτων μέσα σε μικρό χρονικό διάστημα.

Η μορφή της κατανομής δείχνει ότι το σύνολο δεδομένων χαρακτηρίζεται από ανισορροπία μεταξύ συχνών χαμηλών τιμών και σπάνιων υψηλών τιμών κατανάλωσης. Το γεγονός αυτό έχει άμεση επίδραση στη συμπεριφορά των μοντέλων πρόβλεψης, καθώς τα μοντέλα τείνουν να προσαρμόζονται καλύτερα στις συχνότερες περιοχές της κατανομής, ενώ η ακριβής πρόβλεψη των σπάνιων αιχμών παραμένει δυσκολότερη.

Η παρατήρηση αυτή είναι σημαντική και για την αξιολόγηση των μοντέλων. Οι πολύ χαμηλές πραγματικές τιμές μπορούν να οδηγήσουν σε υψηλά ποσοστιαία σφάλματα, ακόμη και όταν η απόλυτη απόκλιση είναι μικρή. Για τον λόγο αυτό, η αξιολόγηση δεν πρέπει να βασίζεται αποκλειστικά σε μία μετρική, αλλά να συνδυάζει δείκτες όπως MAE, RMSE και MAPE, ώστε να αποτυπώνονται τόσο τα απόλυτα σφάλματα όσο και οι σχετικές αποκλίσεις. Συνολικά, η κατανομή της κατανάλωσης αναδεικνύει τη δυσκολία του προβλήματος πρόβλεψης στο PLEGMA και ενισχύει την ανάγκη χρήσης μοντέλων που μπορούν να αποτυπώσουν μη γραμμικές σχέσεις και σπάνιες αιχμές φορτίου.

13. Μοντελοποίηση στο σύνολο δεδομένων PLEGMA

Στο κεφάλαιο αυτό παρουσιάζεται η διαδικασία μοντελοποίησης που εφαρμόστηκε στο σύνολο δεδομένων PLEGMA, με στόχο την αξιολόγηση της προτεινόμενης μεθοδολογίας σε δεδομένα ελληνικών κατοικιών. Μετά την ολοκλήρωση της προεπεξεργασίας και της διερευνητικής ανάλυσης, το τελικό ωριαίο σύνολο δεδομένων χρησιμοποιήθηκε για την εκπαίδευση και σύγκριση μοντέλων μηχανικής μάθησης, ακολουθώντας τη γενική λογική που εφαρμόστηκε και στο σύνολο δεδομένων IDEAL.

Η μοντελοποίηση στο PLEGMA δεν βασίστηκε σε άμεση μεταφορά των μοντέλων που είχαν εκπαιδευτεί στο IDEAL, αλλά σε εκ νέου εκπαίδευση των ίδιων βασικών αλγορίθμων πάνω στα δεδομένα του PLEGMA. Η επιλογή αυτή κρίθηκε απαραίτητη λόγω των διαφορών μεταξύ των δύο συνόλων δεδομένων ως προς το γεωγραφικό και κλιματικό πλαίσιο, τη χρονική κάλυψη, το πλήθος κατοικιών και τη δομή των διαθέσιμων χαρακτηριστικών.

Στο πλαίσιο του κεφαλαίου παρουσιάζονται η μεθοδολογική προσαρμογή στο PLEGMA, τα σενάρια πρόβλεψης που εξετάστηκαν, η δημιουργία των κατάλληλων χαρακτηριστικών για κάθε σενάριο, τα μοντέλα μηχανικής μάθησης που χρησιμοποιήθηκαν, καθώς και τα βασικά μοντέλα αναφοράς. Επιπλέον, περιγράφεται η διαδικασία εκπαίδευσης και αποθήκευσης των μοντέλων, ώστε να καταστεί δυνατή η αξιοποίησή τους τόσο στην πειραματική αξιολόγηση όσο και στην τελική εφαρμογή πρόβλεψης.

13.1 Μεθοδολογική προσέγγιση και προσαρμογή στο PLEGMA

Όπως και στην περίπτωση του συνόλου δεδομένων IDEAL, το προβλεπτικό πρόβλημα ορίζεται ως η εκτίμηση της ωριαίας ηλεκτρικής κατανάλωσης για την επόμενη ημέρα σε επίπεδο κατοικίας. Για κάθε κατοικία και για κάθε ώρα του εξεταζόμενου ημερήσιου ορίζοντα, το μοντέλο παράγει μία πρόβλεψη της κατανάλωσης σε Wh, αξιοποιώντας τα διαθέσιμα χρονικά, περιβαλλοντικά, στατικά και, όπου είναι διαθέσιμα, ιστορικά χαρακτηριστικά.

Η διαδικασία μοντελοποίησης στο σύνολο δεδομένων PLEGMA βασίστηκε στην ίδια γενική μεθοδολογική προσέγγιση που εφαρμόστηκε στο σύνολο δεδομένων IDEAL, με στόχο τη διατήρηση συνέπειας μεταξύ των δύο πειραματικών πλαισίων και τη δυνατότητα άμεσης σύγκρισης των αποτελεσμάτων. Ωστόσο, η εφαρμογή της μεθοδολογίας προσαρμόστηκε στα ιδιαίτερα χαρακτηριστικά του PLEGMA, τα οποία διαφοροποιούνται σημαντικά από εκείνα του IDEAL.

Αρχικά, το σύνολο δεδομένων PLEGMA περιλαμβάνει σημαντικά μικρότερο αριθμό κατοικιών και μικρότερη συνολική ποσότητα διαθέσιμων δεδομένων. Επιπλέον, η χρονική διάρκεια καταγραφής δεν είναι ίδια για όλες τις κατοικίες, γεγονός που οδηγεί σε διαφορετικό βαθμό χρονικής κάλυψης μεταξύ των επιμέρους χρονοσειρών. Παράλληλα, το σύνολο δεδομένων προέρχεται από ελληνικές κατοικίες και αποτυπώνει διαφορετικές κλιματικές και καταναλωτικές συνθήκες σε σχέση με το IDEAL, το οποίο βασίζεται σε κατοικίες του Ηνωμένου Βασιλείου.

Για τους λόγους αυτούς δεν εφαρμόστηκε μεταφορά μάθησης (transfer learning) από τα μοντέλα που εκπαιδεύτηκαν στο IDEAL. Αντίθετα, οι ίδιοι αλγόριθμοι μηχανικής μάθησης εκπαιδεύτηκαν εκ νέου χρησιμοποιώντας αποκλειστικά τα δεδομένα του PLEGMA. Η επιλογή αυτή επιτρέπει την ανεξάρτητη αξιολόγηση της προτεινόμενης μεθοδολογίας σε

διαφορετικό ενεργειακό και γεωγραφικό περιβάλλον, χωρίς να εισάγονται πρόσθετες παραδοχές ή πιθανές μεροληψίες που θα μπορούσαν να προκύψουν από τη χρήση μοντέλων εκπαιδευμένων σε διαφορετικό σύνολο δεδομένων.

Παρά τις διαφοροποιήσεις αυτές, διατηρήθηκαν βασικά στοιχεία της μεθοδολογίας που εφαρμόστηκε στο IDEAL. Συγκεκριμένα, χρησιμοποιήθηκαν παρόμοιες κατηγορίες χαρακτηριστικών, όπως χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά κατοικίας και νοικοκυριού, ενώ υιοθετήθηκαν τα ίδια βασικά σενάρια πρόβλεψης. Επιπλέον, εφαρμόστηκε αντίστοιχη λογική αξιολόγησης και σύγκρισης των μοντέλων, ώστε τα αποτελέσματα των δύο συνόλων δεδομένων να είναι άμεσα συγκρίσιμα.

Η προσέγγιση αυτή επιτρέπει την ανεξάρτητη αξιολόγηση της μεθοδολογίας στο PLEGMA και τη διερεύνηση του βαθμού στον οποίο οι τεχνικές που εφαρμόστηκαν στο IDEAL μπορούν να προσαρμοστούν σε ένα μικρότερο σύνολο δεδομένων ελληνικών κατοικιών.

13.2 Σενάρια πρόβλεψης

Στο σύνολο δεδομένων PLEGMA εξετάστηκαν δύο βασικά σενάρια πρόβλεψης, όπως και στο IDEAL. Η διάκριση αυτή επιτρέπει την αξιολόγηση της απόδοσης των μοντέλων υπό διαφορετικά επίπεδα διαθέσιμης πληροφορίας και βοηθά στην κατανόηση της σημασίας που έχει το ιστορικό κατανάλωσης για την πρόβλεψη της ηλεκτρικής ζήτησης επόμενης ημέρας.

Το πρώτο σενάριο είναι το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης. Σε αυτό το σενάριο, τα μοντέλα δεν έχουν πρόσβαση σε προηγούμενες μετρήσεις κατανάλωσης της κατοικίας. Η πρόβλεψη βασίζεται αποκλειστικά σε χαρακτηριστικά που μπορούν να είναι διαθέσιμα χωρίς την ύπαρξη ιστορικού κατανάλωσης, όπως χρονικά χαρακτηριστικά, περιβαλλοντικές μεταβλητές και στατικά χαρακτηριστικά κατοικίας και νοικοκυριού. Το σενάριο αυτό προσομοιώνει την περίπτωση όπου το σύστημα καλείται να εκτιμήσει την κατανάλωση για μία κατοικία χωρίς διαθέσιμο πρόσφατο ενεργειακό ιστορικό.

Το δεύτερο σενάριο είναι το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης. Σε αυτή την περίπτωση, εκτός από τα χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά, τα μοντέλα αξιοποιούν και πληροφορία από προηγούμενες τιμές κατανάλωσης. Η πληροφορία αυτή ενσωματώνεται μέσω ιστορικών χαρακτηριστικών, όπως χρονικές υστερήσεις και κυλιόμενοι στατιστικοί δείκτες, τα οποία δημιουργούνται κατά τη φάση διαμόρφωσης των δεδομένων εκπαίδευσης. Με τον τρόπο αυτό, το μοντέλο μπορεί να λάβει υπόψη την πρόσφατη δυναμική της κατανάλωσης και να προσαρμόσει την πρόβλεψη στο τρέχον επίπεδο ενεργειακής χρήσης της κατοικίας.

Η σύγκριση των δύο σεναρίων είναι ιδιαίτερα σημαντική, καθώς επιτρέπει να εκτιμηθεί η προστιθέμενη αξία των ιστορικών δεδομένων στην πρόβλεψη. Αναμένεται ότι το σενάριο με χρήση ιστορικών δεδομένων θα παρουσιάζει καλύτερη απόδοση, καθώς η πρόσφατη κατανάλωση αποτελεί ισχυρή ένδειξη της μελλοντικής συμπεριφοράς, ιδίως σε οικιακές χρονοσειρές όπου εμφανίζονται επαναλαμβανόμενα πρότυπα και βραχυχρόνια αυτοσυσχέτιση. Ωστόσο, το σενάριο χωρίς χρήση ιστορικών δεδομένων παραμένει σημαντικό, καθώς αντιπροσωπεύει μία πιο απαιτητική και πρακτικά χρήσιμη περίπτωση εφαρμογής, όπου δεν υπάρχει ακόμη διαθέσιμο ιστορικό για τη συγκεκριμένη κατοικία.

13.3 Δημιουργία χαρακτηριστικών για τα σενάρια πρόβλεψης

Η δημιουργία των χαρακτηριστικών για τη μοντελοποίηση στο PLEGMA βασίστηκε στο τελικό ωριαίο σύνολο δεδομένων που προέκυψε από τη διαδικασία προεπεξεργασίας και συγχρονισμού. Το σύνολο αυτό περιλαμβάνει τη μεταβλητή στόχο `consumption_Wh`, η οποία εκφράζει την ωριαία ηλεκτρική κατανάλωση, καθώς και χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά κατοικίας και νοικοκυριού.

Όπως και στο IDEAL, δεν εφαρμόστηκε ξεχωριστό στάδιο ομαδοποίησης των κατοικιών. Οι διαφοροποιήσεις μεταξύ των κατοικιών ενσωματώθηκαν μέσω των διαθέσιμων στατικών χαρακτηριστικών, όπως ο τύπος κατοικίας, ο αριθμός δωματίων, ο αριθμός ενοίκων και οι διαθέσιμες πληροφορίες για το νοικοκυριό και τις συσκευές. Με τον τρόπο αυτό, τα μοντέλα μπορούν να μάθουν διαφορετικά πρότυπα κατανάλωσης χωρίς να απαιτείται προκαθορισμένη κατηγοριοποίηση των κατοικιών.

Για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης αξιοποιήθηκαν μόνο χαρακτηριστικά που δεν απαιτούν προηγούμενες μετρήσεις κατανάλωσης. Συγκεκριμένα, χρησιμοποιήθηκαν χρονικά χαρακτηριστικά, όπως η ώρα της ημέρας, η ημέρα της εβδομάδας, ο μήνας, η εποχή, η ένδειξη Σαββατοκύριακου και η ένδειξη αργίας. Παράλληλα, χρησιμοποιήθηκαν περιβαλλοντικές μεταβλητές, όπως η εξωτερική και εσωτερική θερμοκρασία και υγρασία, καθώς και στατικά χαρακτηριστικά που περιγράφουν την κατοικία και το νοικοκυριό, όπως ο τύπος κατοικίας, ο αριθμός δωματίων, ο αριθμός ενοίκων, το σύστημα θέρμανσης, το σύστημα παραγωγής ζεστού νερού και η ύπαρξη βασικών ηλεκτρικών συσκευών.

Για το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, το ίδιο βασικό σύνολο χαρακτηριστικών εμπλουτίστηκε με χαρακτηριστικά που υπολογίζονται από προηγούμενες τιμές κατανάλωσης. Τα χαρακτηριστικά αυτά περιλαμβάνουν χρονικές υστερήσεις και κυλιόμενους στατιστικούς δείκτες, οι οποίοι αποτυπώνουν τη βραχυχρόνια συμπεριφορά της κατανάλωσης κάθε κατοικίας. Η δημιουργία τους πραγματοποιήθηκε χωριστά για κάθε κατοικία και μόνο με βάση παρελθοντικές τιμές, ώστε να αποφεύγεται η χρήση μελλοντικής πληροφορίας.

Η διάκριση των χαρακτηριστικών ανά σενάριο επιτρέπει τη σαφή σύγκριση της πρόβλεψης με και χωρίς αξιοποίηση πρόσφατης κατανάλωσης.

13.4 Μοντέλα μηχανικής μάθησης

Για την πρόβλεψη της ηλεκτρικής κατανάλωσης στο σύνολο δεδομένων PLEGMA χρησιμοποιήθηκαν οι ίδιοι βασικοί αλγόριθμοι μηχανικής μάθησης που εφαρμόστηκαν και στο σύνολο δεδομένων IDEAL. Συγκεκριμένα, εκπαιδεύτηκαν μοντέλα Random Forest, XGBoost και LightGBM, τα οποία ανήκουν στην κατηγορία των μοντέλων που βασίζονται σε δέντρα αποφάσεων και είναι κατάλληλα για προβλήματα πρόβλεψης με δομημένα δεδομένα σε μορφή πίνακα.

Η επιλογή των ίδιων αλγορίθμων με το IDEAL έγινε για λόγους μεθοδολογικής συνέπειας και συγκρισιμότητας. Με αυτόν τον τρόπο, η αξιολόγηση στο PLEGMA επικεντρώνεται κυρίως στην επίδραση των διαφορών των δεδομένων, όπως το πλήθος κατοικιών, η χρονική κάλυψη και τα διαθέσιμα χαρακτηριστικά.

Το Random Forest χρησιμοποιήθηκε ως ισχυρό και αξιόπιστο μοντέλο συνόλου, το οποίο συνδυάζει πολλαπλά δέντρα αποφάσεων και μπορεί να αποτυπώσει μη γραμμικές σχέσεις

μεταξύ των χαρακτηριστικών εισόδου και της κατανάλωσης. Το XGBoost και το LightGBM χρησιμοποιήθηκαν ως πιο εξελιγμένες μέθοδοι ενισχυτικής μάθησης δέντρων, οι οποίες εκπαιδεύονται διαδοχικά με στόχο τη σταδιακή μείωση του σφάλματος πρόβλεψης.

Η χρήση αυτών των μοντέλων κρίνεται κατάλληλη για το σύνολο δεδομένων PLEGMA, καθώς η διερευνητική ανάλυση έδειξε ότι η ηλεκτρική κατανάλωση επηρεάζεται από μη γραμμικές σχέσεις, όπως η σχέση της κατανάλωσης με την εξωτερική θερμοκρασία, καθώς και από χρονικά πρότυπα που διαφοροποιούνται ανά ώρα και τύπο ημέρας. Τα μοντέλα που βασίζονται σε δέντρα αποφάσεων μπορούν να χειριστούν τέτοιες σχέσεις χωρίς να απαιτείται εκ των προτέρων γραμμική υπόθεση για τη μορφή των δεδομένων.

Για κάθε μοντέλο εξετάστηκαν τα δύο βασικά σενάρια πρόβλεψης που έχουν αναλυθεί. Με τον τρόπο αυτό, κάθε αλγόριθμος αξιολογείται τόσο σε συνθήκες περιορισμένης πληροφορίας όσο και σε συνθήκες όπου είναι διαθέσιμο πρόσφατο ιστορικό κατανάλωσης. Όπως και στο IDEAL, η εκπαίδευση των μοντέλων πραγματοποιήθηκε με χρήση λογαριθμικού μετασχηματισμού της μεταβλητής στόχου, ώστε να περιοριστεί η επίδραση πολύ υψηλών τιμών κατανάλωσης και να βελτιωθεί η σταθερότητα της εκπαίδευσης. Μετά την παραγωγή των προβλέψεων, οι τιμές μετατράπηκαν ξανά στην αρχική κλίμακα Wh.

13.5 Βασικά μοντέλα αναφοράς

Πριν από την αξιολόγηση των μοντέλων μηχανικής μάθησης, χρησιμοποιήθηκαν βασικά μοντέλα αναφοράς, ώστε να υπάρχει ένα απλό και σαφές σημείο σύγκρισης. Τα μοντέλα αυτά δεν έχουν ως στόχο να επιτύχουν την καλύτερη δυνατή πρόβλεψη, αλλά να ορίσουν ένα αρχικό επίπεδο απόδοσης, με βάση το οποίο μπορεί να εκτιμηθεί αν οι πιο σύνθετοι αλγόριθμοι προσφέρουν ουσιαστική βελτίωση.

Η χρήση μοντέλων αναφοράς είναι ιδιαίτερα σημαντική στο PLEGMA, καθώς το σύνολο δεδομένων είναι μικρότερο από το IDEAL και παρουσιάζει διαφοροποιήσεις ως προς τη χρονική κάλυψη και τη συμπεριφορά των κατοικιών. Επομένως, η σύγκριση των μοντέλων μηχανικής μάθησης με απλούστερες προσεγγίσεις επιτρέπει να αξιολογηθεί αν τα πιο σύνθετα μοντέλα αξιοποιούν πράγματι αποτελεσματικά τα διαθέσιμα χαρακτηριστικά.

Στο πλαίσιο της εργασίας εξετάστηκαν μοντέλα αναφοράς που βασίζονται σε απλές χρονικές σχέσεις της κατανάλωσης. Ένα από αυτά είναι το μοντέλο εμμονής, στο οποίο η πρόβλεψη βασίζεται στην αμέσως προηγούμενη διαθέσιμη τιμή κατανάλωσης. Η προσέγγιση αυτή είναι απλή, αλλά συχνά αποδεικνύεται ισχυρή σε χρονοσειρές ηλεκτρικής κατανάλωσης, καθώς η ζήτηση παρουσιάζει βραχυχρόνια συνέχεια και αυτοσυσχέτιση.

Επιπλέον, εξετάστηκαν μοντέλα που βασίζονται στην κατανάλωση της ίδιας ώρας της προηγούμενης ημέρας, καθώς και μοντέλα που χρησιμοποιούν μέσες τιμές κατανάλωσης ανά ώρα. Οι προσεγγίσεις αυτές αξιοποιούν την ημερήσια περιοδικότητα της οικιακής κατανάλωσης και παρέχουν απλές προβλέψεις χωρίς σύνθετη διαδικασία εκπαίδευσης. Σε ορισμένες περιπτώσεις, η μέση ωριαία κατανάλωση υπολογίστηκε συνολικά για το σύνολο των δεδομένων, ενώ σε άλλες περιπτώσεις διαφοροποιήθηκε ανά κατοικία ή ανά τύπο ημέρας, ώστε να ληφθούν υπόψη βασικές διαφορές στη συμπεριφορά κατανάλωσης.

Τα βασικά μοντέλα αναφοράς χρησιμοποιήθηκαν αποκλειστικά ως σημείο σύγκρισης για την αξιολόγηση των μοντέλων Random Forest, XGBoost και LightGBM. Η αναλυτική παρουσίαση της απόδοσής τους, με βάση δείκτες όπως MAE, RMSE και MAPE,

πραγματοποιείται στο επόμενο κεφάλαιο, όπου εξετάζονται συνολικά τα αποτελέσματα της μοντελοποίησης στο PLEGMA.

13.6 Διαδικασία εκπαίδευσης και χρονικός διαχωρισμός δεδομένων

Για την αξιολόγηση των μοντέλων στο σύνολο δεδομένων PLEGMA εφαρμόστηκε χρονικός διαχωρισμός των δεδομένων σε σύνολα εκπαίδευσης, επικύρωσης και ελέγχου. Αρχικά, όλες οι παρατηρήσεις ταξινομήθηκαν χρονολογικά και στη συνέχεια διαχωρίστηκαν σε διαδοχικές χρονικές περιόδους, ώστε κάθε μοντέλο να εκπαιδεύεται σε παλαιότερα δεδομένα και να αξιολογείται σε μεταγενέστερες χρονικές περιόδους.

Η χρήση χρονικού, αντί τυχαίου, διαχωρισμού είναι απαραίτητη σε προβλήματα πρόβλεψης χρονοσειρών, καθώς αποτρέπει τη διαρροή πληροφορίας από το μέλλον προς το παρελθόν και προσομοιώνει πιο ρεαλιστικά τη λειτουργία ενός πραγματικού συστήματος πρόβλεψης. Στην πράξη, ένα τέτοιο σύστημα καλείται να παράγει προβλέψεις για χρονικές στιγμές που δεν έχουν ακόμη παρατηρηθεί, χρησιμοποιώντας μόνο πληροφορίες που είναι διαθέσιμες μέχρι τη στιγμή της πρόβλεψης.

Για τα τελικά μοντέλα LightGBM και XGBoost χρησιμοποιήθηκε ο ακόλουθος χρονικός διαχωρισμός:

- Σύνολο εκπαίδευσης για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης: από 11 Ιουλίου 2022 14:00 έως 2 Αυγούστου 2023 01:00.
- Σύνολο εκπαίδευσης για το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης: από 18 Ιουλίου 2022 14:00 έως 2 Αυγούστου 2023 01:00.
- Σύνολο επικύρωσης για τον έλεγχο υπερπροσαρμογής των μοντέλων: από 2 Αυγούστου 2023 02:00 έως 17 Αυγούστου 2023 01:00.
- Σύνολο επικύρωσης για βαθμονόμηση των προβλέψεων και διαγνωστικό έλεγχο: από 17 Αυγούστου 2023 02:00 έως 1 Σεπτεμβρίου 2023 01:00.
- Σύνολο ελέγχου: από 1 Σεπτεμβρίου 2023 02:00 έως 1 Οκτωβρίου 2023 01:00.

Στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, το σύνολο εκπαίδευσης ξεκινά μεταγενέστερα, επειδή απαιτείται προηγούμενο ιστορικό για τη δημιουργία των χαρακτηριστικών υστέρησης και των κυλιόμενων στατιστικών. Για τα μοντέλα LightGBM και XGBoost χρησιμοποιήθηκε ιστορικό έως 168 ώρες, ενώ στο Random Forest χρησιμοποιήθηκαν και μεγαλύτερες υστέρησης. Για τον λόγο αυτό, στο Random Forest με χρήση ιστορικών δεδομένων το σύνολο εκπαίδευσης ξεκινά στις 25 Ιουλίου 2022 14:00.

Η διάκριση των δύο συνόλων επικύρωσης επιτρέπει αφενός τον έλεγχο υπερπροσαρμογής κατά την εκπαίδευση των μοντέλων διαδοχικής ενίσχυσης δέντρων αποφάσεων και αφετέρου τη χρήση ξεχωριστού χρονικού διαστήματος για βαθμονόμηση ή διαγνωστική αξιολόγηση των προβλέψεων, χωρίς να χρησιμοποιείται το τελικό σύνολο ελέγχου.

Για τα μοντέλα Random Forest δεν εφαρμόζεται διαδικασία πρόωρης διακοπής της εκπαίδευσης, καθώς ο αλγόριθμος δεν υποστηρίζει αντίστοιχο μηχανισμό. Ωστόσο, διατηρήθηκε η ίδια γενική χρονική λογική διαχωρισμού, ώστε τα αποτελέσματα να είναι συγκρίσιμα με εκείνα των μοντέλων LightGBM και XGBoost.

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, τα τελικά μοντέλα εκπαιδεύτηκαν εκ νέου χρησιμοποιώντας όλα τα διαθέσιμα δεδομένα πριν από το σύνολο ελέγχου, δηλαδή έως την 1η Σεπτεμβρίου 2023 01:00. Αντίστοιχα, το τελικό σύνολο

ελέγχου παρέμεινε κοινό για όλα τα μοντέλα και για τα δύο σενάρια πρόβλεψης. Με τον τρόπο αυτό εξασφαλίζεται η άμεση συγκρισιμότητα των αποτελεσμάτων μεταξύ διαφορετικών αλγορίθμων και διαφορετικών σεναρίων, καθώς όλα τα μοντέλα αξιολογούνται στην ίδια χρονική περίοδο και υπό τις ίδιες συνθήκες.

13.7 Αποθήκευση μοντέλων και σύνδεση με την εφαρμογή πρόβλεψης

Μετά την ολοκλήρωση της εκπαίδευσης, τα τελικά μοντέλα αποθηκεύτηκαν σε κατάλληλη μορφή, ώστε να μπορούν να χρησιμοποιηθούν από την εφαρμογή πρόβλεψης. Για κάθε αλγόριθμο και για κάθε σενάριο πρόβλεψης αποθηκεύτηκαν τα απαραίτητα αρχεία του μοντέλου, η διαδικασία προεπεξεργασίας των χαρακτηριστικών, καθώς και οι αντίστοιχες ρυθμίσεις που απαιτούνται για τη σωστή παραγωγή προβλέψεων.

Η αποθήκευση των μοντέλων πραγματοποιήθηκε με τέτοιο τρόπο ώστε η διαδικασία πρόβλεψης να είναι συνεπής με τη διαδικασία εκπαίδευσης. Αυτό σημαίνει ότι τα ίδια χαρακτηριστικά, η ίδια λογική κωδικοποίησης των κατηγορικών μεταβλητών και η ίδια διαχείριση των αριθμητικών μεταβλητών εφαρμόζονται τόσο κατά την εκπαίδευση όσο και κατά τη χρήση του μοντέλου στην εφαρμογή. Με τον τρόπο αυτό αποφεύγονται ασυνέπειες μεταξύ του σταδίου εκπαίδευσης και του σταδίου παραγωγής προβλέψεων.

Ανάλογα με το επιλεγμένο σενάριο πρόβλεψης, η εφαρμογή χρησιμοποιεί το αντίστοιχο αποθηκευμένο μοντέλο και τα κατάλληλα χαρακτηριστικά εισόδου. Στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης αξιοποιούνται χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά, ενώ στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης προστίθενται χαρακτηριστικά υστέρησης και κυλιόμενοι στατιστικοί δείκτες που υπολογίζονται από τις πρόσφατες μετρήσεις κατανάλωσης.

Ιδιαίτερη σημασία δόθηκε στο γεγονός ότι τα τελικά μοντέλα δεν βασίζονται στο αναγνωριστικό της κατοικίας ως χαρακτηριστικό εισόδου. Το αναγνωριστικό χρησιμοποιείται μόνο εσωτερικά, όπου απαιτείται, για την οργάνωση των χρονοσειρών και τη δημιουργία ιστορικών χαρακτηριστικών. Η επιλογή αυτή καθιστά τη μεθοδολογία πιο κατάλληλη για χρήση σε πραγματική εφαρμογή, καθώς η πρόβλεψη δεν εξαρτάται από την ύπαρξη μιας συγκεκριμένης κατοικίας στο σύνολο εκπαίδευσης.

Η σύνδεση των αποθηκευμένων μοντέλων με την εφαρμογή πρόβλεψης επιτρέπει την πρακτική αξιοποίηση της μεθοδολογίας που αναπτύχθηκε στο πλαίσιο της εργασίας. Ο χρήστης μπορεί να επιλέξει το κατάλληλο σενάριο πρόβλεψης, να εισαγάγει τα απαιτούμενα χαρακτηριστικά και, όπου είναι διαθέσιμο, να χρησιμοποιήσει πρόσφατο ιστορικό κατανάλωσης. Στη συνέχεια, η εφαρμογή χρησιμοποιεί το αντίστοιχο αποθηκευμένο μοντέλο και επιστρέφει την προβλεπόμενη ηλεκτρική κατανάλωση για το χρονικό διάστημα που εξετάζεται.

Με αυτόν τον τρόπο, η μοντελοποίηση στο PLEGMA δεν περιορίζεται στην πειραματική αξιολόγηση των αλγορίθμων, αλλά συνδέεται άμεσα με την τελική εφαρμογή πρόβλεψης.

14. Αξιολόγηση βασικών μοντέλων αναφοράς και μοντέλων μηχανικής μάθησης

Σε αυτό το κεφάλαιο παρουσιάζεται η αξιολόγηση των βασικών μοντέλων αναφοράς και των μοντέλων μηχανικής μάθησης στο σύνολο δεδομένων PLEGMA. Αρχικά, περιγράφεται το πειραματικό πλαίσιο που ακολουθήθηκε, με έμφαση στον χρονικό διαχωρισμό των δεδομένων, στην προεπεξεργασία της μεταβλητής στόχου και στη χρήση κοινών μετρικών αξιολόγησης, ώστε τα αποτελέσματα να είναι συγκρίσιμα με εκείνα του IDEAL. Στη συνέχεια, παρουσιάζονται τα αποτελέσματα των μοντέλων μηχανικής μάθησης στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης και συγκρίνονται με το ισχυρότερο βασικό μοντέλο αναφοράς. Ακολουθεί η αξιολόγηση του σεναρίου με χρήση ιστορικών δεδομένων κατανάλωσης, προκειμένου να διερευνηθεί η συμβολή της πρόσφατης κατανάλωσης στη βελτίωση της ακρίβειας των προβλέψεων. Επιπλέον, εξετάζεται η σημασία των χαρακτηριστικών, ώστε να ερμηνευθεί ποιοι παράγοντες επηρεάζουν περισσότερο την απόδοση των μοντέλων στα δύο σενάρια. Τέλος, συνοψίζονται τα βασικά συμπεράσματα της αξιολόγησης, αναδεικνύοντας τόσο τη χρησιμότητα των ιστορικών και πρόσθετων χαρακτηριστικών όσο και τους περιορισμούς που προκύπτουν από τη μεταβλητότητα και το μικρότερο μέγεθος του συνόλου δεδομένων PLEGMA.

14.1 Ρυθμίσεις αξιολόγησης και πειραματικό πλαίσιο

Η αξιολόγηση των μοντέλων στο σύνολο δεδομένων PLEGMA πραγματοποιήθηκε με την ίδια βασική λογική που εφαρμόστηκε και στο σύνολο δεδομένων IDEAL, ώστε να διατηρείται συνέπεια στη μεθοδολογία εκπαίδευσης, αξιολόγησης και σύγκρισης των σεναρίων πρόβλεψης. Όλα τα μοντέλα αξιολογήθηκαν στο ίδιο ανεξάρτητο σύνολο ελέγχου, το οποίο αντιστοιχεί στο τελευταίο διαθέσιμο χρονικό διάστημα του συνόλου δεδομένων PLEGMA, από 1 Σεπτεμβρίου 2023 02:00 έως 1 Οκτωβρίου 2023 01:00.

Η χρήση κοινού συνόλου ελέγχου για όλα τα μοντέλα και για τα δύο σενάρια πρόβλεψης επιτρέπει την άμεση σύγκριση των αποτελεσμάτων. Με τον τρόπο αυτό, οι διαφορές στην απόδοση μπορούν να αποδοθούν κυρίως στη φύση των χαρακτηριστικών που χρησιμοποιούνται και στις ιδιότητες των αλγορίθμων και όχι σε διαφορετικές χρονικές περιόδους αξιολόγησης.

Η αξιολόγηση πραγματοποιήθηκε ξεχωριστά για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης και για το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης. Η σύγκριση των δύο σεναρίων επιτρέπει να εκτιμηθεί η συμβολή του ιστορικού κατανάλωσης στη βελτίωση της ακρίβειας των προβλέψεων.

Για την αποτίμηση της απόδοσης χρησιμοποιήθηκαν οι ίδιες βασικές μετρικές με το IDEAL: το Μέσο Απόλυτο Σφάλμα (MAE), η Ρίζα του Μέσου Τετραγωνικού Σφάλματος (RMSE) και το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (MAPE). Το MAE αποτυπώνει το μέσο μέγεθος του σφάλματος σε Wh, το RMSE δίνει μεγαλύτερη βαρύτητα στα μεγάλα σφάλματα και στις αποκλίσεις κατά τις ώρες αιχμής, ενώ το MAPE εκφράζει το σφάλμα σε ποσοστιαία μορφή.

Επιπλέον, εξετάστηκε η απόκλιση στη συνολική ημερήσια κατανάλωση, μέσω του μέσου απόλυτου ημερήσιου σφάλματος σε kWh. Η μετρική αυτή είναι χρήσιμη επειδή αξιολογεί

κατά πόσο το μοντέλο μπορεί να εκτιμήσει σωστά όχι μόνο την ωριαία κατανομή της κατανάλωσης, αλλά και τη συνολική ενεργειακή ζήτηση της ημέρας.

Ιδιαίτερη προσοχή απαιτείται στην ερμηνεία του MAPE, καθώς το σύνολο δεδομένων PLEGMA περιλαμβάνει αρκετές χαμηλές τιμές κατανάλωσης. Σε τέτοιες περιπτώσεις, ακόμη και μικρές απόλυτες αποκλίσεις μπορούν να οδηγήσουν σε υψηλά ποσοστιαία σφάλματα. Για τον λόγο αυτό, η αξιολόγηση δεν βασίζεται σε μία μόνο μετρική, αλλά στη συνδυαστική εξέταση των MAE, RMSE, MAPE και ημερήσιου σφάλματος κατανάλωσης. Στις ενότητες που ακολουθούν, οι τιμές που παρουσιάζονται στους πίνακες αποτελεσμάτων αποτελούν μέσους όρους των αντίστοιχων μετρικών για όλες τις κατοικίες και όλες τις διαθέσιμες χρονικές εγγραφές του συνόλου ελέγχου, αποτυπώνοντας έτσι τη συνολική μέση απόδοση κάθε μοντέλου στο ανεξάρτητο σύνολο αξιολόγησης.

14.2 Αποτελέσματα βασικών μοντέλων αναφοράς

Πριν από την αξιολόγηση των μοντέλων μηχανικής μάθησης, εξετάστηκε η απόδοση βασικών μοντέλων αναφοράς, με στόχο να οριστεί ένα απλό και σαφές σημείο σύγκρισης. Τα μοντέλα αυτά βασίζονται σε απλές στατιστικές ή χρονικές σχέσεις της κατανάλωσης και δεν αξιοποιούν σύνθετη διαδικασία εκπαίδευσης. Η χρήση τους είναι σημαντική, καθώς επιτρέπει να αξιολογηθεί αν τα μοντέλα μηχανικής μάθησης προσφέρουν ουσιαστική βελτίωση σε σχέση με απλούστερες προσεγγίσεις.

Στο πλαίσιο της αξιολόγησης στο σύνολο δεδομένων PLEGMA εξετάστηκαν πέντε βασικά μοντέλα αναφοράς. Τα μοντέλα αυτά περιλαμβάνουν προσεγγίσεις που βασίζονται στην ίδια ώρα της προηγούμενης ημέρας, σε μέσες τιμές ανά ώρα, σε μέσες τιμές ανά ώρα με διάκριση τύπου ημέρας, σε μέση τιμή ανά ώρα ανά κατοικία, καθώς και στο μοντέλο εμμονής, όπου η πρόβλεψη βασίζεται στην αμέσως προηγούμενη διαθέσιμη τιμή κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)
Baseline 05 – Persistence (lag_1)	249,99	528,53	62,74
Baseline 04 – Μέση τιμή ανά ώρα (ανά κατοικία)	327,68	540,97	117,56
Baseline 02 – Μέση τιμή ανά ώρα (συνολικά)	343,23	611,73	98,12
Baseline 03 – Μέση τιμή ανά ώρα + τύπος ημέρας	343,61	612,21	98,41
Baseline 01 – Ίδια ώρα προηγούμενης ημέρας (lag_{24h})	364,82	693,20	115,40

Πίνακας 37 Απόδοση βασικών μοντέλων αναφοράς στο σύνολο δεδομένων PLEGMA

Όπως φαίνεται στον Πίνακα 37, το μοντέλο εμμονής Persistence (lag_1) παρουσιάζει την καλύτερη απόδοση μεταξύ των βασικών μοντέλων αναφοράς, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Το αποτέλεσμα αυτό επιβεβαιώνει ότι η ηλεκτρική

κατανάλωση εμφανίζει ισχυρή βραχυχρόνια αυτοσυσχέτιση, καθώς η αμέσως προηγούμενη τιμή αποτελεί ιδιαίτερα ισχυρή ένδειξη για τη μελλοντική κατανάλωση.

Τα υπόλοιπα βασικά μοντέλα, τα οποία βασίζονται σε μέσες τιμές ή σε απλοποιημένες χρονικές σχέσεις, παρουσιάζουν υψηλότερα σφάλματα. Ιδιαίτερα τα μοντέλα που βασίζονται σε μέσες ωριαίες τιμές δυσκολεύονται να αποτυπώσουν τη βραχυχρόνια μεταβλητότητα της κατανάλωσης, καθώς εξομαλύνουν τις διακυμάνσεις και δεν λαμβάνουν υπόψη τις άμεσες μεταβολές στη συμπεριφορά της κατοικίας.

Με βάση τα παραπάνω, το μοντέλο Persistence (lag_1) επιλέγεται ως το βασικό σημείο αναφοράς για την αξιολόγηση των μοντέλων μηχανικής μάθησης που παρουσιάζονται στη συνέχεια. Τα μοντέλα Random Forest, XGBoost και LightGBM θα πρέπει να επιτύχουν χαμηλότερα σφάλματα από αυτό το baseline, ώστε να τεκμηριωθεί η προστιθέμενη αξία της χρήσης χρονικών, περιβαλλοντικών, στατικών και ιστορικών χαρακτηριστικών.

14.3 Απόδοση μοντέλων μηχανικής μάθησης στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης

Στην ενότητα αυτή παρουσιάζεται η απόδοση των μοντέλων μηχανικής μάθησης στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης. Το σενάριο αυτό αποτελεί την πιο απαιτητική περίπτωση πρόβλεψης, καθώς τα μοντέλα δεν έχουν πρόσβαση σε πρόσφατες μετρήσεις κατανάλωσης της κατοικίας και βασίζονται αποκλειστικά στα χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά.

Ο Πίνακας 38 παρουσιάζει την απόδοση των τριών μοντέλων μηχανικής μάθησης στο ανεξάρτητο σύνολο ελέγχου για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	Ημερήσιο σφάλμα (kWh)
Random Forest	190,29	429,11	55,28	2,70
XGBoost	187,67	428,01	56,10	2,66
LightGBM	187,64	427,47	55,12	2,68

Πίνακας 38 Απόδοση μοντέλων μηχανικής μάθησης στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης

Τα αποτελέσματα δείχνουν ότι και τα τρία μοντέλα μηχανικής μάθησης υπερτερούν του ισχυρότερου βασικού μοντέλου αναφοράς, δηλαδή του Persistence (lag_1), το οποίο είχε MAE 249,99 Wh και RMSE 528,53 Wh. Η μείωση του MAE σε τιμές κοντά στα 188–190 Wh δείχνει ότι τα μοντέλα μπορούν να αξιοποιήσουν αποτελεσματικά την πληροφορία που περιέχεται στα χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά, ακόμη και χωρίς πρόσβαση σε προηγούμενες μετρήσεις κατανάλωσης.

Μεταξύ των τριών μοντέλων, το LightGBM παρουσιάζει οριακά την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Το XGBoost εμφανίζει σχεδόν ισοδύναμη συμπεριφορά, με πολύ μικρές διαφορές σε σχέση με το LightGBM. Το Random Forest παρουσιάζει ελαφρώς υψηλότερα σφάλματα, ωστόσο παραμένει ιδιαίτερα ανταγωνιστικό και σαφώς καλύτερο από το βασικό μοντέλο αναφοράς.

Η μικρή διαφορά μεταξύ XGBoost και LightGBM δείχνει ότι, στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, τα μοντέλα διαδοχικής ενίσχυσης δέντρων αποφάσεων αξιοποιούν με παρόμοιο τρόπο τα διαθέσιμα χαρακτηριστικά. Παράλληλα, η σχετικά κοντινή απόδοση του Random Forest υποδηλώνει ότι η επιλογή του αλγορίθμου δεν είναι ο μοναδικός καθοριστικός παράγοντας της απόδοσης. Η ποιότητα και η πληροφοριακή αξία των χαρακτηριστικών εισόδου παραμένουν εξίσου σημαντικές. Τα αποτελέσματα του σεναρίου χωρίς χρήση ιστορικών δεδομένων κατανάλωσης δείχνουν ότι τα μοντέλα μηχανικής μάθησης προσφέρουν σαφή βελτίωση έναντι των απλών στατιστικών προσεγγίσεων. Ωστόσο, η απουσία πρόσφατης πληροφορίας κατανάλωσης περιορίζει την ακρίβεια των προβλέψεων, γεγονός που καθιστά αναγκαία την εξέταση του σεναρίου με χρήση ιστορικών δεδομένων κατανάλωσης στην επόμενη ενότητα.

14.4 Απόδοση μοντέλων μηχανικής μάθησης στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης

Στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, τα μοντέλα αξιοποιούν πρόσθετη πληροφορία από προηγούμενες μετρήσεις της ίδιας κατοικίας, μέσω χαρακτηριστικών υστέρησης και κυλιόμενων στατιστικών δεικτών.

Ο Πίνακας 39 παρουσιάζει την απόδοση των τριών μοντέλων μηχανικής μάθησης στο ανεξάρτητο σύνολο ελέγχου για το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	Ημερήσιο σφάλμα (kWh)
Random Forest	166,96	395,86	48,75	1,78
XGBoost	165,00	395,12	46,03	1,81
LightGBM	162,80	395,13	43,99	1,81

Πίνακας 39 Απόδοση μοντέλων μηχανικής μάθησης στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης

Τα αποτελέσματα δείχνουν ότι η χρήση ιστορικών δεδομένων κατανάλωσης οδηγεί σε βελτίωση της απόδοσης των μοντέλων σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Το MAE μειώνεται σε όλες τις περιπτώσεις, γεγονός που δείχνει ότι η πρόσφατη κατανάλωση παρέχει σημαντική πληροφορία για την πρόβλεψη της επόμενης ημέρας. Η βελτίωση είναι ιδιαίτερα εμφανής στο LightGBM, το οποίο επιτυγχάνει το χαμηλότερο MAE και το χαμηλότερο MAPE μεταξύ των τριών μοντέλων.

Το XGBoost παρουσιάζει επίσης πολύ ανταγωνιστική απόδοση, με RMSE σχεδόν ταυτόσημο με εκείνο του LightGBM. Η μικρή διαφορά μεταξύ των δύο μοντέλων δείχνει ότι οι μέθοδοι διαδοχικής ενίσχυσης δέντρων αποφάσεων αξιοποιούν με παρόμοιο τρόπο τα ιστορικά χαρακτηριστικά. Το Random Forest εμφανίζει ελαφρώς υψηλότερα ωριαία σφάλματα, αλλά παραμένει ανταγωνιστικό και παρουσιάζει το χαμηλότερο μέσο ημερήσιο σφάλμα κατανάλωσης.

Η μείωση του MAPE στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης δείχνει ότι τα ιστορικά χαρακτηριστικά βοηθούν τα μοντέλα να προσαρμόζονται καλύτερα το επίπεδο της πρόβλεψης στη συμπεριφορά κάθε κατοικίας. Παρ' όλα αυτά, οι τιμές του MAPE

παραμένουν σχετικά υψηλές, γεγονός που συνδέεται με την ύπαρξη αρκετών χαμηλών τιμών κατανάλωσης στο σύνολο δεδομένων PLEGMA.

Συνολικά, τα αποτελέσματα της ενότητας δείχνουν ότι η πρόσφατη κατανάλωση αποτελεί σημαντικό παράγοντα βελτίωσης της προβλεπτικής ακρίβειας. Η χρήση χαρακτηριστικών υστέρησης και κυλιόμενων στατιστικών δεικτών επιτρέπει στα μοντέλα να αποτυπώνουν καλύτερα τη βραχυχρόνια συμπεριφορά της κατοικίας, οδηγώντας σε χαμηλότερα σφάλματα σε σχέση με το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

14.5 Ερμηνεία της σημασίας των χαρακτηριστικών

Για την καλύτερη ερμηνεία των αποτελεσμάτων πραγματοποιήθηκε ανάλυση της σημασίας των χαρακτηριστικών με χρήση της μεθόδου permutation importance. Η μέθοδος αυτή επιτρέπει την εκτίμηση της συμβολής κάθε μεταβλητής στην απόδοση του μοντέλου, εξετάζοντας πόσο επηρεάζεται το σφάλμα πρόβλεψης όταν οι τιμές ενός χαρακτηριστικού αναδιατάσσονται τυχαία.

Τα αποτελέσματα της ανάλυσης δείχνουν ότι η σχετική σημασία των χαρακτηριστικών διαφοροποιείται ανάλογα με τη διαθεσιμότητα ιστορικών δεδομένων κατανάλωσης. Στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, τα μοντέλα βασίζονται κυρίως σε χρονικές και περιβαλλοντικές μεταβλητές, όπως η ώρα της ημέρας, η εξωτερική θερμοκρασία, η εσωτερική θερμοκρασία και οι μεταβλητές υγρασίας. Το εύρημα αυτό είναι αναμενόμενο, καθώς όταν δεν υπάρχει διαθέσιμη πληροφορία προηγούμενης κατανάλωσης, τα μοντέλα αξιοποιούν περισσότερο τις εξωγενείς συνθήκες και τα επαναλαμβανόμενα ημερήσια πρότυπα.

Αντίθετα, στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, τα σημαντικότερα χαρακτηριστικά σχετίζονται κυρίως με προηγούμενες τιμές κατανάλωσης και κυλιόμενους στατιστικούς δείκτες. Χαρακτηριστικά όπως η κατανάλωση της προηγούμενης ώρας, της προηγούμενης ημέρας ή προηγούμενων ημερών παρέχουν άμεση πληροφορία για το πρόσφατο επίπεδο ενεργειακής χρήσης της κατοικίας. Η αυξημένη σημασία αυτών των μεταβλητών εξηγεί σε μεγάλο βαθμό τη βελτίωση της απόδοσης που παρατηρείται στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης.

Η ανάλυση της σημασίας των χαρακτηριστικών επιβεβαιώνει ότι η πρόσφατη κατανάλωση αποτελεί έναν από τους ισχυρότερους προγνωστικούς παράγοντες στη βραχυπρόθεσμη πρόβλεψη ηλεκτρικής κατανάλωσης. Παράλληλα, δείχνει ότι τα χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά εξακολουθούν να παρέχουν χρήσιμη πληροφορία, ιδιαίτερα στις περιπτώσεις όπου δεν υπάρχει διαθέσιμο ιστορικό κατανάλωσης.

Ιδιαίτερη προσοχή απαιτείται στην ερμηνεία των στατικών χαρακτηριστικών που εμφανίζονται ως σημαντικά. Η υψηλή σημασία ενός χαρακτηριστικού σε μια ανάλυση permutation importance δείχνει ότι το χαρακτηριστικό αυτό συμβάλλει στην προβλεπτική απόδοση του μοντέλου, αλλά δεν συνεπάγεται απαραίτητα άμεση αιτιώδη σχέση με την κατανάλωση ηλεκτρικής ενέργειας (Fisher et al., 2019· Molnar, 2022). Η ερμηνεία γίνεται ακόμη πιο σύνθετη όταν τα χαρακτηριστικά είναι συσχετισμένα μεταξύ τους, καθώς ένα στατικό χαρακτηριστικό μπορεί να λειτουργεί ως έμμεσος δείκτης άλλων παραγόντων που δεν αποτυπώνονται πλήρως στο μοντέλο (Strobl et al., 2008). Για τον λόγο αυτό, χαρακτηριστικά όπως ο αριθμός ενοίκων, ο εξοπλισμός της κατοικίας ή η διάρκεια διαμονής πρέπει να ερμηνεύονται κυρίως ως ενδείξεις συσχέτισης με το γενικό επίπεδο ενεργειακής χρήσης και όχι ως αυτόνομες αιτιώδεις μεταβλητές.

Η ανάλυση της σημασίας των χαρακτηριστικών υποστηρίζει τα συμπεράσματα της συγκριτικής αξιολόγησης. Όταν δεν χρησιμοποιείται ιστορικό κατανάλωσης, η πρόβλεψη εξαρτάται περισσότερο από εξωγενείς και χρονικούς παράγοντες. Όταν όμως είναι διαθέσιμη πρόσφατη κατανάλωση, τα μοντέλα αξιοποιούν κυρίως τη δυναμική της ίδιας της χρονοσειράς, γεγονός που οδηγεί σε ακριβέστερες προβλέψεις.

14.6 Ερμηνεία αποτελεσμάτων

Τα αποτελέσματα της αξιολόγησης των μοντέλων μηχανικής μάθησης στο σύνολο δεδομένων PLEGMΑ δείχνουν ότι τα μοντέλα αυτά υπερτερούν των βασικών μοντέλων αναφοράς. Η βελτίωση αυτή επιβεβαιώνει ότι η πρόβλεψη της οικιακής ηλεκτρικής κατανάλωσης δεν μπορεί να βασιστεί μόνο σε απλές χρονικές σχέσεις, αλλά ωφελείται από τον συνδυασμό χρονικών, περιβαλλοντικών και στατικών χαρακτηριστικών.

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, τα μοντέλα Random Forest, XGBoost και LightGBM κατάφεραν να μειώσουν αισθητά τα σφάλματα σε σχέση με το ισχυρότερο βασικό μοντέλο αναφοράς. Το αποτέλεσμα αυτό δείχνει ότι ακόμη και χωρίς πρόσφατες μετρήσεις κατανάλωσης, τα διαθέσιμα χαρακτηριστικά της κατοικίας, του χρόνου και των περιβαλλοντικών συνθηκών παρέχουν χρήσιμη πληροφορία για την πρόβλεψη.

Η χρήση ιστορικών δεδομένων κατανάλωσης οδήγησε σε περαιτέρω βελτίωση της ακρίβειας των μοντέλων. Η μείωση των σφαλμάτων δείχνει ότι η πρόσφατη κατανάλωση της ίδιας κατοικίας αποτελεί ιδιαίτερα σημαντικό προγνωστικό παράγοντα. Τα χαρακτηριστικά υστέρησης και οι κυλιόμενοι στατιστικοί δείκτες επιτρέπουν στα μοντέλα να προσαρμόζουν καλύτερα την πρόβλεψη στο πρόσφατο επίπεδο ενεργειακής χρήσης και στη βραχυχρόνια δυναμική της κατανάλωσης.

Οι διαφορές μεταξύ των τριών αλγορίθμων παραμένουν σχετικά περιορισμένες. Τα μοντέλα XGBoost και LightGBM εμφανίζουν οριακά καλύτερη απόδοση στις περισσότερες ωριαίες μετρικές, ενώ το Random Forest παραμένει ανταγωνιστικό και παρουσιάζει ικανοποιητική συμπεριφορά, ιδιαίτερα λαμβάνοντας υπόψη τη μικρότερη πολυπλοκότητά του. Το εύρημα αυτό δείχνει ότι η επιλογή των χαρακτηριστικών και η διαθεσιμότητα πρόσφατου ιστορικού επηρεάζουν περισσότερο την τελική ακρίβεια από ό,τι η επιλογή μεταξύ των συγκεκριμένων δένδρων αλγορίθμων.

Η ανάλυση της σημασίας των χαρακτηριστικών ενισχύει την παραπάνω ερμηνεία, καθώς δείχνει ότι η πρόσφατη κατανάλωση αποκτά κυρίαρχο ρόλο όταν είναι διαθέσιμη, ενώ στα σενάρια χωρίς ιστορικά δεδομένα τα μοντέλα βασίζονται περισσότερο σε χρονικούς, περιβαλλοντικούς και στατικούς παράγοντες.

Παρά τα θετικά αποτελέσματα, η αξιολόγηση στο PLEGMΑ ανέδειξε και ορισμένους περιορισμούς. Το μικρότερο πλήθος κατοικιών, η διαφοροποιημένη χρονική κάλυψη και η έντονη μεταβλητότητα της οικιακής κατανάλωσης καθιστούν το πρόβλημα πιο απαιτητικό. Συμπερασματικά, η εφαρμογή της μεθοδολογίας στο PLEGMΑ επιβεβαιώνει ότι τα μοντέλα μηχανικής μάθησης μπορούν να αξιοποιήσουν αποτελεσματικά τα διαθέσιμα χαρακτηριστικά για την πρόβλεψη της ηλεκτρικής κατανάλωσης σε ελληνικές κατοικίες. Η χρήση ιστορικών δεδομένων βελτιώνει περαιτέρω την ακρίβεια, ενώ τα αποτελέσματα δείχνουν ότι η μεθοδολογία μπορεί να προσαρμοστεί σε διαφορετικό σύνολο δεδομένων από το IDEAL. Παράλληλα, οι περιορισμοί του PLEGMΑ αναδεικνύουν την ανάγκη

προσεκτικής ερμηνείας των αποτελεσμάτων και ενισχύουν τη σημασία της αξιολόγησης σε επίπεδο πραγματικών κατοικιών και ημερήσιων προφίλ κατανάλωσης.

15. Προσαρμογή της εφαρμογής πρόβλεψης στο σύνολο δεδομένων PLEGMA

Μετά την εκπαίδευση και αξιολόγηση των μοντέλων μηχανικής μάθησης στο σύνολο δεδομένων PLEGMA, ακολούθησε η ενσωμάτωσή τους στην εφαρμογή πρόβλεψης που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας. Στόχος του κεφαλαίου αυτού είναι να παρουσιαστεί ο τρόπος με τον οποίο η ήδη διαμορφωμένη αρχιτεκτονική πρόβλεψης προσαρμόστηκε στις ιδιαιτερότητες του PLEGMA, ώστε να μπορεί να υποστηρίξει την παραγωγή προβλέψεων ηλεκτρικής κατανάλωσης για ελληνικές κατοικίες.

Η εφαρμογή διατηρεί τη βασική λογική που παρουσιάστηκε στην περίπτωση του IDEAL, δηλαδή τον διαχωρισμό μεταξύ των αποθηκευμένων μοντέλων, του επιπέδου υπηρεσιών και της γραφικής διεπαφής χρήστη. Ωστόσο, απαιτήθηκαν προσαρμογές στα χαρακτηριστικά εισόδου, στη διαχείριση των διαθέσιμων μοντέλων και στον τρόπο αξιοποίησης των ιστορικών δεδομένων κατανάλωσης, ώστε το σύστημα να είναι συμβατό με τη δομή και τις μεταβλητές του PLEGMA.

Στο κεφάλαιο αυτό παρουσιάζονται συνοπτικά οι βασικές αλλαγές που πραγματοποιήθηκαν για την προσαρμογή της εφαρμογής στο νέο σύνολο δεδομένων. Έμφαση δίνεται στη διαχείριση των μοντέλων πρόβλεψης, στην υποστήριξη των σεναρίων λειτουργίας, στη χρήση ιστορικών δεδομένων κατανάλωσης και στη λειτουργία της γραφικής διεπαφής χρήστη. Με τον τρόπο αυτό, το κεφάλαιο λειτουργεί ως σύνδεση ανάμεσα στην πειραματική αξιολόγηση των μοντέλων και στην επόμενη φάση της εργασίας, όπου οι παραγόμενες προβλέψεις θα αξιολογηθούν μέσα από συγκεκριμένα παραδείγματα χρήσης.

15.1 Σκοπός της εφαρμογής και σχέση με την υλοποίηση του IDEAL

Η εφαρμογή πρόβλεψης που χρησιμοποιήθηκε για το PLEGMA βασίστηκε στη γενική αρχιτεκτονική που είχε ήδη αναπτυχθεί για το σύνολο δεδομένων IDEAL. Η αρχική υλοποίηση περιλάμβανε την αποθήκευση των εκπαιδευμένων μοντέλων, την εκτέλεση των προβλέψεων μέσω API και την αλληλεπίδραση με τον χρήστη μέσω γραφικής διεπαφής. Στο πλαίσιο του PLEGMA, η λογική αυτή διατηρήθηκε, αλλά προσαρμόστηκε ώστε να ανταποκρίνεται στη διαφορετική δομή των δεδομένων και στα χαρακτηριστικά των ελληνικών κατοικιών.

Βασικός σκοπός της εφαρμογής είναι η παραγωγή πρόβλεψης ηλεκτρικής κατανάλωσης επόμενης ημέρας, αξιοποιώντας τα μοντέλα που εκπαιδεύτηκαν και αξιολογήθηκαν στα προηγούμενα κεφάλαια. Η εφαρμογή επιτρέπει στον χρήστη να επιλέγει το σενάριο πρόβλεψης, να εισάγει τα απαραίτητα χαρακτηριστικά της κατοικίας και, όπου είναι διαθέσιμο, να αξιοποιεί ιστορικά δεδομένα κατανάλωσης.

Η προσαρμογή στο PLEGMA ήταν απαραίτητη, καθώς το συγκεκριμένο σύνολο δεδομένων περιλαμβάνει διαφορετικά περιβαλλοντικά και στατικά χαρακτηριστικά σε σχέση με το IDEAL. Επιπλέον, τα μοντέλα που εκπαιδεύτηκαν για το PLEGMA δεν βασίζονται στην ίδια ακριβώς δομή εισόδων με τα μοντέλα του IDEAL, γεγονός που απαιτεί αντίστοιχη προσαρμογή στη διαδικασία προετοιμασίας των δεδομένων εισόδου και στη διαχείριση των διαθέσιμων μοντέλων.

Η ενσωμάτωση των μοντέλων PLEGMA στην εφαρμογή επιτρέπει την αξιοποίηση της ίδιας βασικής υποδομής πρόβλεψης σε ένα διαφορετικό σύνολο δεδομένων. Με αυτόν τον

τρόπο εξετάζεται όχι μόνο η απόδοση των μοντέλων σε επίπεδο πειραματικής αξιολόγησης, αλλά και η δυνατότητα προσαρμογής της συνολικής μεθοδολογίας σε ένα νέο περιβάλλον εφαρμογής.

15.2 Προσαρμογή της αρχιτεκτονικής API/UI στο PLEGMA

Η προσαρμογή της εφαρμογής στο σύνολο δεδομένων PLEGMA βασίστηκε στη διατήρηση της ίδιας γενικής αρχιτεκτονικής που χρησιμοποιήθηκε και στην περίπτωση του IDEAL. Η αρχιτεκτονική αυτή περιλαμβάνει τρία βασικά επίπεδα: το επίπεδο των αποθηκευμένων μοντέλων, το επίπεδο υπηρεσιών μέσω API και τη γραφική διεπαφή χρήστη. Η διατήρηση αυτής της δομής επέτρεψε την ενσωμάτωση των νέων μοντέλων χωρίς να απαιτηθεί πλήρης ανασχεδιασμός του συστήματος.

Οι βασικές αλλαγές αφορούσαν κυρίως την προσαρμογή των εισόδων της εφαρμογής στα χαρακτηριστικά του συνόλου δεδομένων PLEGMA. Σε αντίθεση με το IDEAL, το PLEGMA περιλαμβάνει διαφορετικό σύνολο περιβαλλοντικών και στατικών μεταβλητών, όπως εσωτερική και εξωτερική θερμοκρασία, εσωτερική και εξωτερική υγρασία, καθώς και κοινωνικοδημογραφικά και κτηριακά χαρακτηριστικά που προέρχονται από τα διαθέσιμα μεταδεδομένα των κατοικιών. Για τον λόγο αυτό, η διαδικασία δημιουργίας των χαρακτηριστικών εισόδου προσαρμόστηκε ώστε να αντιστοιχεί στη δομή των μοντέλων που εκπαιδεύτηκαν στο PLEGMA.

Παράλληλα, προσαρμόστηκε η διαχείριση των διαθέσιμων μοντέλων στο σύστημα. Τα τελικά μοντέλα Random Forest, XGBoost και LightGBM αποθηκεύτηκαν σε μορφή κατάλληλη για χρήση από το API, μαζί με τις απαραίτητες πληροφορίες προεπεξεργασίας και τα χαρακτηριστικά που απαιτούνται σε κάθε σενάριο πρόβλεψης. Με αυτόν τον τρόπο διασφαλίζεται ότι κατά την εκτέλεση μιας πρόβλεψης εφαρμόζεται η ίδια λογική μετασχηματισμού που χρησιμοποιήθηκε κατά την εκπαίδευση.

Το επίπεδο υπηρεσιών αναλαμβάνει την παραλαβή των δεδομένων εισόδου, την εφαρμογή των απαραίτητων ελέγχων και τη δρομολόγηση του αιτήματος προς το κατάλληλο μοντέλο. Ανάλογα με το επιλεγμένο σενάριο, το API δημιουργεί είτε μόνο τα βασικά χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά είτε επιπλέον χαρακτηριστικά που προκύπτουν από το ιστορικό κατανάλωσης. Η διαδικασία αυτή επιτρέπει στην εφαρμογή να υποστηρίζει με ενιαίο τρόπο τόσο το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης όσο και το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης.

Η γραφική διεπαφή χρήστη προσαρμόστηκε αντίστοιχα, ώστε να επιτρέπει την εισαγωγή των απαραίτητων χαρακτηριστικών για το PLEGMA και την επιλογή του επιθυμητού σεναρίου πρόβλεψης. Μέσω της διεπαφής, ο χρήστης μπορεί να επιλέξει μοντέλο, να εισαγάγει ή να αξιοποιήσει διαθέσιμα δεδομένα ιστορικού και να λάβει την προβλεπόμενη κατανάλωση σε ωριαία και ημερήσια μορφή.

Η προσαρμογή της αρχιτεκτονικής στο PLEGMA δείχνει ότι η εφαρμογή μπορεί να λειτουργήσει ως ευέλικτη υποδομή πρόβλεψης, η οποία δεν περιορίζεται σε ένα μόνο σύνολο δεδομένων. Η επαναχρησιμοποίηση της βασικής αρχιτεκτονικής και η αντικατάσταση των μοντέλων και των χαρακτηριστικών εισόδου επιτρέπουν τη μεταφορά της ίδιας μεθοδολογικής λογικής σε διαφορετικά δεδομένα και διαφορετικά περιβάλλοντα εφαρμογής.

15.3 Διαχείριση μοντέλων και σεναρίων πρόβλεψης

Η εφαρμογή πρόβλεψης για το PLEGMA υποστηρίζει τα τελικά μοντέλα μηχανικής μάθησης που εκπαιδεύτηκαν και αξιολογήθηκαν στα προηγούμενα κεφάλαια. Συγκεκριμένα, ενσωματώθηκαν μοντέλα Random Forest, XGBoost και LightGBM, ώστε η εφαρμογή να μπορεί να παράγει προβλέψεις με διαφορετικές μεθόδους και να επιτρέπει τη σύγκριση των αποτελεσμάτων τους.

Η διαχείριση των μοντέλων πραγματοποιείται μέσω μηχανισμού καταχώρισης μοντέλων, ο οποίος περιλαμβάνει πληροφορίες για τα διαθέσιμα μοντέλα, τα υποστηριζόμενα σενάρια πρόβλεψης και τα χαρακτηριστικά εισόδου που απαιτούνται σε κάθε περίπτωση. Με αυτόν τον τρόπο, η εφαρμογή δεν εξαρτάται άμεσα από συγκεκριμένες διαδρομές αρχείων ή από χειροκίνητη επιλογή αντικειμένων μοντέλου, αλλά μπορεί να εντοπίζει και να χρησιμοποιεί το κατάλληλο μοντέλο ανάλογα με την επιλογή του χρήστη.

Η εφαρμογή υποστηρίζει δύο βασικά σενάρια λειτουργίας. Στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, η πρόβλεψη βασίζεται στα διαθέσιμα χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά της κατοικίας. Το σενάριο αυτό είναι κατάλληλο για περιπτώσεις όπου δεν υπάρχει διαθέσιμο ενεργειακό ιστορικό ή όταν ο χρήστης επιθυμεί μια αρχική εκτίμηση της κατανάλωσης με βάση τα γενικά χαρακτηριστικά της κατοικίας και τις περιβαλλοντικές συνθήκες.

Στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, η εφαρμογή αξιοποιεί επιπλέον πρόσφατες μετρήσεις κατανάλωσης της ίδιας κατοικίας. Τα δεδομένα αυτά χρησιμοποιούνται για τη δημιουργία χαρακτηριστικών υστέρησης και κυλιόμενων στατιστικών δεικτών, τα οποία είναι απαραίτητα για τα αντίστοιχα μοντέλα. Με τον τρόπο αυτό, η πρόβλεψη μπορεί να προσαρμοστεί καλύτερα στο πρόσφατο επίπεδο και στη βραχυχρόνια δυναμική της κατανάλωσης.

Η επιλογή του σεναρίου πρόβλεψης πραγματοποιείται από τον χρήστη μέσω της γραφικής διεπαφής. Ανάλογα με την επιλογή αυτή, το σύστημα ενεργοποιεί την κατάλληλη διαδικασία προετοιμασίας των δεδομένων και δρομολογεί το αίτημα προς το αντίστοιχο μοντέλο. Η ενιαία αυτή διαχείριση επιτρέπει στην εφαρμογή να υποστηρίζει διαφορετικές περιπτώσεις χρήσης χωρίς να απαιτείται αλλαγή στον βασικό τρόπο λειτουργίας της.

Επιπλέον, η ύπαρξη πολλαπλών μοντέλων επιτρέπει τη σύγκριση των προβλέψεων που παράγονται από διαφορετικούς αλγόριθμους. Η δυνατότητα αυτή είναι χρήσιμη τόσο για ερευνητικούς σκοπούς όσο και για την καλύτερη κατανόηση της συμπεριφοράς των μοντέλων σε συγκεκριμένα σενάρια πρόβλεψης. Παράλληλα, η εφαρμογή μπορεί να υποστηρίξει συνδυαστικές προβλέψεις, στις οποίες οι προβλέψεις επιμέρους μοντέλων συνδυάζονται ώστε να παραχθεί ένα ενιαίο αποτέλεσμα.

Για τη συνδυαστική πρόβλεψη χρησιμοποιήθηκαν προεπιλεγμένα βάρη 40% για το LightGBM, 35% για το XGBoost και 25% για το Random Forest. Η επιλογή αυτή βασίστηκε στη συνολική απόδοση των μοντέλων στα δύο σενάρια πρόβλεψης. Το LightGBM εμφάνισε την καλύτερη συνολική συμπεριφορά στους περισσότερους δείκτες, ενώ το XGBoost παρουσίασε πολύ κοντινή απόδοση. Το Random Forest διατηρήθηκε με μικρότερο αλλά μη μηδενικό βάρος, καθώς σε ορισμένες επιμέρους κατοικίες και ημερομηνίες εμφάνισε ανταγωνιστικά αποτελέσματα. Με αυτόν τον τρόπο, η συνδυαστική πρόβλεψη δίνει μεγαλύτερη βαρύτητα στα ισχυρότερα μοντέλα, χωρίς να αποκλείει πλήρως τη συνεισφορά του Random Forest.

Η διαχείριση των μοντέλων και των σεναρίων πρόβλεψης αποτελεί βασικό στοιχείο της προσαρμογής της εφαρμογής στο PLEGMA. Μέσω αυτής της οργάνωσης, τα εκπαιδευμένα μοντέλα μπορούν να αξιοποιηθούν με συνεπή τρόπο από το API και τη γραφική διεπαφή,

διατηρώντας σαφή διάκριση ανάμεσα στις προβλέψεις χωρίς ιστορικό κατανάλωσης και στις προβλέψεις που αξιοποιούν πρόσφατα δεδομένα της κατοικίας.

15.4 Διαχείριση και ανάλυση ιστορικών δεδομένων κατανάλωσης

Η διαχείριση των ιστορικών δεδομένων κατανάλωσης αποτελεί βασικό στοιχείο της εφαρμογής, καθώς επιτρέπει την εκτέλεση προβλέψεων στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης. Στην τελική υλοποίηση, κάθε αρχείο ιστορικού αντιστοιχεί σε μία μόνο κατοικία και περιλαμβάνει δύο βασικές στήλες: τη χρονική σήμανση της μέτρησης και την αντίστοιχη κατανάλωση ηλεκτρικής ενέργειας. Με αυτόν τον τρόπο, η εφαρμογή μπορεί να χρησιμοποιήσει το ιστορικό ως ανεξάρτητη είσοδο για την παραγωγή χαρακτηριστικών υστέρησης και κυλιόμενων στατιστικών δεικτών.

Για την εκτέλεση πρόβλεψης με ιστορικά δεδομένα απαιτείται επαρκές πρόσφατο ιστορικό κατανάλωσης. Η απαίτηση αυτή είναι απαραίτητη, επειδή τα μοντέλα που λειτουργούν σε αυτό το σενάριο χρειάζονται προηγούμενες τιμές κατανάλωσης για τη δημιουργία των αντίστοιχων χαρακτηριστικών εισόδου. Εάν το διαθέσιμο ιστορικό δεν επαρκεί, η εφαρμογή δεν μπορεί να δημιουργήσει αξιόπιστα τα απαιτούμενα χαρακτηριστικά και ενημερώνει τον χρήστη για τον περιορισμό αυτό.

Η εισαγωγή των ιστορικών δεδομένων πραγματοποιείται μέσω αρχείου CSV, το οποίο μπορεί να επιλεγεί από τον χρήστη μέσα από τη γραφική διεπαφή. Μετά την επιλογή του αρχείου, η εφαρμογή εκτελεί βασικούς ελέγχους συμβατότητας ως προς τη δομή των δεδομένων και τη διαθεσιμότητα των απαιτούμενων χρονικών εγγραφών. Στη συνέχεια, το ιστορικό χρησιμοποιείται από το API για τη δημιουργία των κατάλληλων χαρακτηριστικών εισόδου και την παραγωγή της πρόβλεψης.

Πριν από την παραγωγή πρόβλεψης με χρήση ιστορικών δεδομένων, η εφαρμογή παρέχει επίσης δυνατότητα ανάλυσης του διαθέσιμου ιστορικού κατανάλωσης. Η λειτουργία αυτή εξετάζει βασικά χαρακτηριστικά του ιστορικού, όπως το χρονικό εύρος των διαθέσιμων μετρήσεων, την επάρκεια των ωριαίων εγγραφών, τη μέση και διάμεση ημερήσια κατανάλωση, καθώς και τη μεταβλητότητα του πρόσφατου προφίλ κατανάλωσης. Με τον τρόπο αυτό, ο χρήστης μπορεί να αξιολογήσει αν το ιστορικό που έχει επιλεγεί είναι επαρκές και αντιπροσωπευτικό για την παραγωγή πρόβλεψης.

Η ανάλυση αυτή δεν αντικαθιστά τη διαδικασία πρόβλεψης, αλλά λειτουργεί ως διαγνωστικό εργαλείο πριν από την εκτέλεσή της. Είναι ιδιαίτερα χρήσιμη στις περιπτώσεις όπου το πρόσφατο ιστορικό παρουσιάζει έντονη μεταβλητότητα ή μη τυπική συμπεριφορά, καθώς βοηθά στην ερμηνεία της τελικής πρόβλεψης και στην κατανόηση πιθανών αποκλίσεων από τη συνήθη ημερήσια κατανάλωση της κατοικίας.

Με βάση τα αποτελέσματα της ανάλυσης ιστορικού, η εφαρμογή μπορεί επίσης να προτείνει κατάλληλη τιμή για τον μέγιστο συντελεστή προσαρμογής ($\max_alpha$). Ο συντελεστής αυτός χρησιμοποιείται στη λειτουργία προσαρμοστικής διόρθωσης με βάση το πρόσφατο ιστορικό (Adaptive History Correction) και καθορίζει το μέγιστο επιτρεπτό βάρος με το οποίο το πρόσφατο ιστορικό κατανάλωσης μπορεί να επηρεάσει την τελική πρόβλεψη. Με αυτόν τον τρόπο, η επίδραση του ιστορικού παραμένει ελεγχόμενη: όταν το ιστορικό είναι πιο σταθερό και αντιπροσωπευτικό, επιτρέπεται μεγαλύτερη προσαρμογή της πρόβλεψης, ενώ σε περιπτώσεις ασταθούς ή περιορισμένου ιστορικού η διόρθωση παραμένει πιο συντηρητική.

Η λειτουργία Adaptive History Correction εφαρμόζεται ως προαιρετικό στάδιο μεταγενέστερης προσαρμογής της πρόβλεψης και όχι ως διαδικασία επανεκπαίδευσης του

μοντέλου. Επομένως, η τελική πρόβλεψη εξακολουθεί να βασίζεται στο εκπαιδευμένο μοντέλο μηχανικής μάθησης, ενώ η πρόσθετη διόρθωση λειτουργεί βοηθητικά, αξιοποιώντας τη σταθερότητα και το πρόσφατο προφίλ κατανάλωσης της κατοικίας.

Η χρήση ιστορικών δεδομένων επιτρέπει την καλύτερη προσαρμογή της πρόβλεψης στη συμπεριφορά της συγκεκριμένης κατοικίας. Το σύστημα δεν βασίζεται μόνο στα γενικά χαρακτηριστικά του νοικοκυριού και στις περιβαλλοντικές συνθήκες, αλλά λαμβάνει υπόψη και το πρόσφατο επίπεδο κατανάλωσης, γεγονός που συμβάλλει στη βελτίωση της ακρίβειας των προβλέψεων.

15.5 Γραφική διεπαφή χρήστη και βασική ροή πρόβλεψης

Η γραφική διεπαφή χρήστη προσαρμόστηκε ώστε να υποστηρίζει την εκτέλεση προβλέψεων ηλεκτρικής κατανάλωσης με τα μοντέλα που εκπαιδεύτηκαν στο σύνολο δεδομένων PLEGMA. Στόχος της διεπαφής είναι να επιτρέπει στον χρήστη να διαμορφώνει εύκολα ένα σενάριο πρόβλεψης, να επιλέγει το επιθυμητό μοντέλο και να λαμβάνει την προβλεπόμενη ωριαία κατανάλωση της επόμενης ημέρας χωρίς να απαιτείται άμεση αλληλεπίδραση με τον υποκείμενο κώδικα.

Η βασική ροή χρήσης ξεκινά με την εισαγωγή των απαιτούμενων στοιχείων της κατοικίας, του νοικοκυριού και της ημέρας πρόβλεψης. Ο χρήστης επιλέγει την πόλη και την ημερομηνία πρόβλεψης και στη συνέχεια συμπληρώνει τα διαθέσιμα χαρακτηριστικά της κατοικίας και του νοικοκυριού. Τα χαρακτηριστικά αυτά περιλαμβάνουν, μεταξύ άλλων, τον αριθμό δωματίων, τον αριθμό ενοίκων, τη σύνθεση του νοικοκυριού, τον τύπο κατοικίας, την περίοδο κατασκευής, το είδος θέρμανσης, τον τύπο θερμοσίφωνα, την ύπαρξη βασικών ηλεκτρικών συσκευών και λοιπά διαθέσιμα στατικά χαρακτηριστικά.

Παράλληλα, η διεπαφή επιτρέπει την εισαγωγή περιβαλλοντικών δεδομένων που είναι συμβατά με το PLEGMA. Ο χρήστης μπορεί να αξιοποιήσει διαθέσιμες τιμές εσωτερικής και εξωτερικής θερμοκρασίας, καθώς και εσωτερικής και εξωτερικής υγρασίας. Όταν δεν υπάρχουν πλήρεις ωριαίες τιμές εξωτερικής θερμοκρασίας, η εφαρμογή μπορεί να αξιοποιήσει ελάχιστη και μέγιστη ημερήσια θερμοκρασία ή εναλλακτικούς μηχανισμούς ανάκτησης/εκτίμησης μετεωρολογικών δεδομένων.

Η διεπαφή υποστηρίζει τόσο το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης όσο και το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης. Στην πρώτη περίπτωση, η πρόβλεψη βασίζεται στα χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά που εισάγει ο χρήστης. Στη δεύτερη περίπτωση, δίνεται επιπλέον η δυνατότητα επιλογής αρχείου ιστορικού κατανάλωσης σε μορφή CSV, το οποίο χρησιμοποιείται από την εφαρμογή για τη δημιουργία χαρακτηριστικών υστέρησης και κυλιόμενων στατιστικών δεικτών.

Η συνολική μορφή της γραφικής διεπαφής χρήστη παρουσιάζεται στις Εικόνες 9 και 10, όπου φαίνονται οι βασικές επιλογές εισαγωγής δεδομένων, επιλογής μοντέλου, χρήσης ιστορικού κατανάλωσης και παραγωγής προβλέψεων.

PLEGMA Load Forecasting (Demo UI)

Συμπληρώστε τα πεδία και πάτα **Predict** για ένα μοντέλο, **Compare** για RF / XGB / LGBM ή **Combined Prediction** για weighted ensemble.

☒ API: OK | Ready models: 6 | No history: lgbm_no_history_default | With history: lgbm_with_history_default

Αν τα ενεργοποιήσετε, το API θα χρησιμοποιήσει model="with_history".

☒ Πρόβλεψη με χρήση ιστορικών δεδομένων

Μοντέλο

auto — Auto default

Optimization

balanced

City

athens

Target date (YYYY-MM-DD)

2026-06-23

Environmental inputs (24h)

internal_temperature_24h (default visible, editable)

21.0, 20.8, 20.7, 20.6, 20.5, 20.8, 21.2, 21.5, 21.7, 21.8, 21.9, 22.0, 22.0, 21.9, 21.8, 21.9, 22.0, 22.0, 21.8, 21.5, 21.3, 21.1

external_temperature_24h (optional, 24 values)

Λάστο κενό για city-based lookup / t_min-t_max fallback

internal_humidity_24h (default visible, editable)

50.0, 50.8, 49.5, 49.5, 49.0, 49.0, 50.5, 50.5, 51.0, 51.0, 51.5, 52.0, 52.0, 51.5, 51.0, 50.5, 50.5, 50.0, 49.5, 49.5, 49.5, 50.0

external_humidity_24h (optional, 24 values)

Λάστο κενό για Open-Meteo / default humidity profile από API

Αν δεν βρεθεί εξωτερική θερμοκρασία από Open-Meteo, συμπληρώστε t_min και t_max.

Household inputs

building_type

apartment

build_era

1970_1990

num_rooms

3

residents

2

has_dryer

0

has_ac

1

has_washing_machine

1

has_dishwasher

1

has_microwave

1

heating_type

air_conditioner

water_heater_type

electric_boiler

years_in_house

3_to_4_years

Εικόνα 9 Γραφική διεπαφή χρήστη της εφαρμογής πρόβλεψης για το PLEGMA

With history inputs

To history CSV πρέπει να είναι single-home αρχείο με στήλες `timestamp` και `consumption_wh`. Δεν χρειάζεται και δεν ζητείται `home_id`.

Select history CSV

Selected history CSV path

Θα συμπληρωθεί αυτόματα μετά την επιλογή αρχείου ή γράφει path χειροκίνητα

No history CSV selected.

☒ Use default history_store.csv if no CSV/manual history is provided

min_history_hours

168

history_consumption_wh manual vector (optional, comma-separated)

n.g. 100, 120, 95, ...

☒ Apply adaptive history correction

history_correction_days

7

manual_max_alpha

0.2

History stability recommendation

Behavioral adjustment (optional)

Weighted ensemble settings

☐ Save prediction CSV

Predict

Compare RF / XGB / LGBM

Combined Prediction

Summary

Status

Prediction chart

Prediction / Display table

Weather table

Comparison table

Predictions

1

2

3

Εικόνα 10 Γραφική διεπαφή χρήστη της εφαρμογής πρόβλεψης για το PLEGMA

Πριν από την εκτέλεση πρόβλεψης με ιστορικά δεδομένα, ο χρήστης μπορεί να πραγματοποιήσει ανάλυση του επιλεγμένου ιστορικού. Μέσω της λειτουργίας αυτής εμφανίζονται πληροφορίες σχετικά με τη χρονική κάλυψη, την επάρκεια των διαθέσιμων μετρήσεων, τη μέση και διάμεση ημερήσια κατανάλωση και τη γενικότερη σταθερότητα του πρόσφατου προφίλ κατανάλωσης. Η πληροφορία αυτή βοηθά τον χρήστη να αξιολογήσει αν το διαθέσιμο ιστορικό είναι κατάλληλο για χρήση στην πρόβλεψη.

Μέσω της γραφικής διεπαφής, ο χρήστης μπορεί επίσης να ενεργοποιήσει προαιρετικά τη λειτουργία Adaptive History Correction. Η λειτουργία αυτή εφαρμόζεται μετά την αρχική πρόβλεψη του μοντέλου και αξιοποιεί το πρόσφατο ιστορικό κατανάλωσης, με βάση τον προτεινόμενο ή επιλεγμένο συντελεστή $\max_alpha$. Με τον τρόπο αυτό παρέχεται δυνατότητα ελεγχόμενης προσαρμογής της τελικής πρόβλεψης στο πρόσφατο προφίλ της κατοικίας.

Η εφαρμογή παρέχει επίσης δυνατότητα σύγκρισης των διαθέσιμων μοντέλων και δημιουργίας συνδυαστικών προβλέψεων. Ο χρήστης μπορεί να συγκρίνει τις προβλέψεις των μοντέλων Random Forest, XGBoost και LightGBM για το ίδιο σενάριο εισόδου ή να δημιουργήσει συνδυαστική πρόβλεψη με σταθμισμένο μέσο όρο των επιμέρους μοντέλων. Τα βάρη των μοντέλων μπορούν να παραμείνουν στις προεπιλεγμένες τιμές ή να τροποποιηθούν από τον χρήστη.

Επιπλέον, η διεπαφή περιλαμβάνει λειτουργία προσαρμογής της αναμενόμενης συμπεριφοράς κατανάλωσης. Μέσω της λειτουργίας αυτής ο χρήστης μπορεί να δηλώσει ώρες κατά τις οποίες αναμένεται αυξημένη ή μειωμένη κατανάλωση, εφαρμόζοντας έναν συντελεστή προσαρμογής στην τελική πρόβλεψη. Η δυνατότητα αυτή επιτρέπει την ενσωμάτωση πληροφορίας που μπορεί να γνωρίζει ο χρήστης για την ημέρα πρόβλεψης, αλλά δεν αποτυπώνεται απαραίτητα στα δεδομένα εισόδου του μοντέλου.

Μετά την ολοκλήρωση της πρόβλεψης, τα αποτελέσματα παρουσιάζονται σε μορφή πίνακα και γραφήματος. Ο χρήστης μπορεί να εξετάσει την προβλεπόμενη κατανάλωση για κάθε ώρα της ημέρας, τη συνολική ημερήσια κατανάλωση και πρόσθετες πληροφορίες σχετικά με το μοντέλο, το σενάριο πρόβλεψης και τις πηγές των δεδομένων εισόδου. Παράλληλα, παρέχεται δυνατότητα αποθήκευσης των αποτελεσμάτων σε αρχείο CSV για περαιτέρω ανάλυση ή τεκμηρίωση.

Η γραφική διεπαφή χρήστη λειτουργεί ως το τελικό επίπεδο αλληλεπίδρασης μεταξύ του χρήστη και των εκπαιδευμένων μοντέλων. Μέσω αυτής, η μεθοδολογία πρόβλεψης που αναπτύχθηκε στα προηγούμενα κεφάλαια μετατρέπεται σε πρακτικό εργαλείο παραγωγής και ερμηνείας προβλέψεων ηλεκτρικής κατανάλωσης.

15.6 Διαχείριση μετεωρολογικών και περιβαλλοντικών δεδομένων

Για τη λειτουργία της εφαρμογής πρόβλεψης απαιτείται η παροχή περιβαλλοντικών δεδομένων που αντιστοιχούν στην ημέρα πρόβλεψης. Σε αντίθεση με την περίπτωση του IDEAL, όπου η εξωτερική θερμοκρασία αποτέλεσε τη βασική περιβαλλοντική μεταβλητή, στο PLEGMA η εφαρμογή προσαρμόστηκε ώστε να υποστηρίζει πλουσιότερο σύνολο περιβαλλοντικών εισόδων. Συγκεκριμένα, λαμβάνονται υπόψη η εξωτερική και εσωτερική θερμοκρασία, καθώς και η εξωτερική και εσωτερική υγρασία.

Η διαχείριση των περιβαλλοντικών εισόδων ακολουθεί ευέλικτη λογική, ώστε η εφαρμογή να μπορεί να λειτουργεί ακόμη και όταν δεν είναι διαθέσιμα όλα τα απαιτούμενα δεδομένα σε πλήρη ωριαία μορφή. Για την εξωτερική θερμοκρασία, ο χρήστης μπορεί να εισαγάγει απευθείας πλήρες 24ωρο προφίλ τιμών. Όταν δεν παρέχεται τέτοιο προφίλ, η εφαρμογή μπορεί να αξιοποιήσει εναλλακτικές πηγές ή εκτιμήσεις, όπως την αυτόματη ανάκτηση μετεωρολογικών δεδομένων από την υπηρεσία Open-Meteo, ή τη δημιουργία προσεγγιστικού ημερήσιου προφίλ με βάση την ελάχιστη και τη μέγιστη θερμοκρασία της ημέρας.

Η δυνατότητα χρήσης ελάχιστης και μέγιστης ημερήσιας θερμοκρασίας λειτουργεί ως εφεδρικός μηχανισμός. Στην περίπτωση αυτή, το σύστημα δημιουργεί ένα συνθετικό 24ωρο

προφίλ εξωτερικής θερμοκρασίας, το οποίο προσεγγίζει την ημερήσια διακύμανση της θερμοκρασίας. Η προσέγγιση αυτή δεν αντικαθιστά πλήρως τις πραγματικές ωριαίες μετεωρολογικές τιμές, αλλά επιτρέπει τη λειτουργία της εφαρμογής όταν δεν υπάρχουν διαθέσιμα πλήρη δεδομένα πρόγνωσης.

Αντίστοιχη λογική εφαρμόζεται και για τις υπόλοιπες περιβαλλοντικές μεταβλητές. Η εφαρμογή επιτρέπει την εισαγωγή πλήρων 24ωρων προφίλ για την εξωτερική υγρασία, την εσωτερική θερμοκρασία και την εσωτερική υγρασία. Όταν ο χρήστης δεν παρέχει πλήρη εσωτερικά περιβαλλοντικά δεδομένα, μπορούν να χρησιμοποιηθούν προκαθορισμένα ημερήσια προφίλ, ώστε να διατηρείται η συμβατότητα των εισόδων με τα χαρακτηριστικά που χρησιμοποιήθηκαν κατά την εκπαίδευση των μοντέλων.

Η υποστήριξη εσωτερικών περιβαλλοντικών μεταβλητών είναι ιδιαίτερα σημαντική για το PLEGMA, καθώς το τελικό σύνολο δεδομένων περιλαμβάνει πληροφορία τόσο για τις εξωτερικές όσο και για τις εσωτερικές συνθήκες της κατοικίας. Με αυτόν τον τρόπο, η εφαρμογή δεν περιορίζεται μόνο στις καιρικές συνθήκες της περιοχής, αλλά μπορεί να αξιοποιήσει και πληροφορία που σχετίζεται με το εσωτερικό περιβάλλον, εφόσον αυτή είναι διαθέσιμη.

Η διαχείριση των περιβαλλοντικών δεδομένων σχεδιάστηκε ώστε να διατηρείται συνέπεια ανάμεσα στη διαδικασία εκπαίδευσης των μοντέλων και στη διαδικασία παραγωγής προβλέψεων. Τα μοντέλα που εκπαιδεύτηκαν στο PLEGMA αξιοποίησαν περιβαλλοντικά χαρακτηριστικά θερμοκρασίας και υγρασίας· επομένως, κατά τη χρήση της εφαρμογής, οι αντίστοιχες εισοδοί πρέπει να παρέχονται ή να συμπληρώνονται με τρόπο συμβατό με τη δομή των δεδομένων εκπαίδευσης.

Συνολικά, η εφαρμογή συνδυάζει δυνατότητα χειροκίνητης εισαγωγής περιβαλλοντικών δεδομένων, αυτόματη ανάκτηση ή εφεδρική εκτίμηση των εξωτερικών συνθηκών και χρήση προκαθορισμένων προφίλ όπου απαιτείται. Η προσέγγιση αυτή ενισχύει την ευελιξία και την ανθεκτικότητα του συστήματος, επιτρέποντας την παραγωγή προβλέψεων ακόμη και σε περιπτώσεις όπου ο χρήστης δεν διαθέτει πλήρες σύνολο ωριαίων περιβαλλοντικών μετρήσεων.

15.7 Περιορισμοί και λειτουργικές παραδοχές της εφαρμογής

Παρά τη λειτουργικότητα που παρέχει η εφαρμογή πρόβλεψης, η χρήση της βασίζεται σε ορισμένες λειτουργικές παραδοχές και περιορισμούς, οι οποίοι πρέπει να λαμβάνονται υπόψη κατά την ερμηνεία των αποτελεσμάτων. Οι περιορισμοί αυτοί συνδέονται τόσο με τη φύση των διαθέσιμων δεδομένων όσο και με τον τρόπο με τον οποίο τα εκπαιδευμένα μοντέλα αξιοποιούνται κατά την παραγωγή νέων προβλέψεων.

Αρχικά, η ποιότητα της πρόβλεψης εξαρτάται άμεσα από την ποιότητα και τη συμβατότητα των δεδομένων εισόδου. Στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, η εφαρμογή βασίζεται στα χαρακτηριστικά που εισάγει ο χρήστης, όπως τα στοιχεία της κατοικίας, του νοικοκυριού και των περιβαλλοντικών συνθηκών. Επομένως, ελλιπή ή μη ρεαλιστικά δεδομένα εισόδου μπορούν να επηρεάσουν την αξιοπιστία της παραγόμενης πρόβλεψης.

Στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, η πρόβλεψη εξαρτάται επιπλέον από την επάρκεια και την αντιπροσωπευτικότητα του διαθέσιμου ιστορικού. Η ύπαρξη πρόσφατων μετρήσεων δεν συνεπάγεται απαραίτητα ότι το ιστορικό είναι κατάλληλο για την ημέρα πρόβλεψης. Για παράδειγμα, ένα ιστορικό που περιλαμβάνει

ασυνήθιστες ημέρες, έντονες αιχμές ή μεταβολές στη συμπεριφορά των ενοίκων μπορεί να οδηγήσει σε προβλέψεις που αποκλίνουν από την πραγματική κατανάλωση της ημέρας στόχου.

Η λειτουργία Adaptive History Correction εφαρμόζεται ως προαιρετικό στάδιο μεταγενέστερης προσαρμογής της πρόβλεψης και όχι ως νέο μοντέλο ή διαδικασία επανεκπαίδευσης. Η χρήση του μέγιστου συντελεστή προσαρμογής ($\max_alpha$) επιτρέπει την ελεγχόμενη επίδραση του πρόσφατου ιστορικού, ωστόσο η αποτελεσματικότητα της διόρθωσης εξαρτάται από τη σταθερότητα και την αντιπροσωπευτικότητα του ιστορικού κατανάλωσης. Για τον λόγο αυτό, η λειτουργία πρέπει να ερμηνεύεται ως βοηθητικός μηχανισμός προσαρμογής και όχι ως εγγύηση βελτίωσης σε κάθε περίπτωση.

Επιπλέον, η εφαρμογή αξιοποιεί τα μοντέλα που εκπαιδεύτηκαν στο διαθέσιμο σύνολο δεδομένων PLEGMA. Συνεπώς, οι προβλέψεις είναι πιο αξιόπιστες όταν τα δεδομένα εισόδου βρίσκονται κοντά στο εύρος και στα χαρακτηριστικά των δεδομένων εκπαίδευσης. Περιπτώσεις κατοικιών με πολύ διαφορετική συμπεριφορά κατανάλωσης, διαφορετικό εξοπλισμό ή ιδιαίτερες συνθήκες χρήσης μπορεί να οδηγήσουν σε μεγαλύτερη αβεβαιότητα.

Ένας ακόμη περιορισμός αφορά τα μετεωρολογικά και περιβαλλοντικά δεδομένα. Όταν δεν είναι διαθέσιμες πλήρεις ωριαίες τιμές θερμοκρασίας ή υγρασίας, η εφαρμογή μπορεί να χρησιμοποιήσει εναλλακτικές προσεγγίσεις εκτίμησης ή προκαθορισμένα προφίλ. Οι προσεγγίσεις αυτές διασφαλίζουν τη λειτουργικότητα του συστήματος, αλλά δεν αντικαθιστούν πλήρως την ακρίβεια πραγματικών ή αξιόπιστων μετεωρολογικών και περιβαλλοντικών δεδομένων.

Τέλος, οι πρόσθετες λειτουργίες της διεπαφής, όπως η συνδυαστική πρόβλεψη και η προσαρμογή της αναμενόμενης συμπεριφοράς κατανάλωσης, παρέχουν ευελιξία στον χρήστη, αλλά απαιτούν προσεκτική χρήση. Οι λειτουργίες αυτές μπορούν να βοηθήσουν όταν ο χρήστης διαθέτει πρόσθετη γνώση για τη συγκεκριμένη ημέρα πρόβλεψης, ωστόσο η υπερβολική ή μη τεκμηριωμένη προσαρμογή μπορεί να οδηγήσει σε λιγότερο αξιόπιστα αποτελέσματα.

Εν κατακλείδι, η εφαρμογή αποτελεί ένα πρακτικό εργαλείο αξιοποίησης των εκπαιδευμένων μοντέλων πρόβλεψης, αλλά τα αποτελέσματά της πρέπει να ερμηνεύονται σε συνάρτηση με την ποιότητα των εισόδων, τη διαθεσιμότητα ιστορικών δεδομένων και τους περιορισμούς του συνόλου δεδομένων PLEGMA. Οι παραπάνω παραδοχές λαμβάνονται υπόψη στο επόμενο κεφάλαιο, όπου οι προβλέψεις της εφαρμογής αξιολογούνται μέσα από συγκεκριμένα παραδείγματα κατοικιών και ημερήσιων προφίλ κατανάλωσης.

16. Αξιολόγηση προβλέψεων σε μεμονωμένες κατοικίες (PLEGMA)

Στο παρόν κεφάλαιο παρουσιάζεται η αξιολόγηση του συστήματος πρόβλεψης ηλεκτρικής κατανάλωσης που προσαρμόστηκε στο σύνολο δεδομένων PLEGMA. Όπως και στην περίπτωση του IDEAL, στόχος της αξιολόγησης είναι να εξεταστεί η πρακτική συμπεριφορά της εφαρμογής σε ρεαλιστικά σενάρια χρήσης, όπως αυτά υποστηρίζονται από τη γραφική διεπαφή χρήστη και το API.

Η αξιολόγηση πραγματοποιείται αρχικά σε τέσσερις επιλεγμένες κατοικίες του συνόλου ελέγχου του PLEGMA. Οι κατοικίες αυτές επιλέχθηκαν ώστε να καλύπτουν διαφορετικές περιπτώσεις πρόβλεψης: κατοικίες στις οποίες το σύστημα αποδίδει ικανοποιητικά, κατοικίες με χαμηλότερη ή μεσαία κατανάλωση, καθώς και μία δυσκολότερη περίπτωση που αναδεικνύει τους περιορισμούς της προσέγγισης. Στη συνέχεια, η αξιολόγηση επεκτείνεται και σε εποχικές μελέτες περίπτωσης, ώστε να εξεταστεί η συμπεριφορά του συστήματος σε διαφορετικές χρονικές περιόδους.

Τα αποτελέσματα που παρουσιάζονται στο κεφάλαιο αυτό επιτρέπουν την καλύτερη κατανόηση της πρακτικής λειτουργίας του συστήματος και αναδεικνύουν τον βαθμό στον οποίο η χρήση πρόσφατου ιστορικού μπορεί να βελτιώσει την πρόβλεψη. Ιδιαίτερη έμφαση δίνεται στη σχέση μεταξύ της σταθερότητας του ιστορικού, της αντιπροσωπευτικότητας της ημέρας πρόβλεψης και της τελικής ακρίβειας των μοντέλων.

16.1 Μεθοδολογία αξιολόγησης

Για κάθε κατοικία αξιολογήθηκαν δύο βασικά σενάρια πρόβλεψης. Το πρώτο είναι το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, στο οποίο τα μοντέλα αξιοποιούν χρονικά χαρακτηριστικά, περιβαλλοντικά χαρακτηριστικά και στατικά χαρακτηριστικά της κατοικίας και του νοικοκυριού. Το δεύτερο είναι το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, στο οποίο χρησιμοποιούνται επιπλέον πρόσφατες μετρήσεις κατανάλωσης της ίδιας κατοικίας. Στην παρούσα αξιολόγηση, το σενάριο αυτό εξετάζεται στην τελική μορφή της εφαρμογής, δηλαδή με ενεργοποιημένη την προσαρμοστική διόρθωση βάσει ιστορικού κατανάλωσης.

Για την υποστήριξη του σεναρίου με χρήση ιστορικών δεδομένων χρησιμοποιήθηκε ιστορικό παράθυρο 168 ωρών πριν από την ημέρα πρόβλεψης. Το ιστορικό αυτό περιλαμβάνει πραγματικές μετρήσεις κατανάλωσης που προηγούνται χρονικά της ημέρας στόχου και δεν περιλαμβάνει δεδομένα από την ίδια την ημέρα πρόβλεψης. Με αυτόν τον τρόπο αποφεύγεται η διαρροή πληροφορίας από την ημέρα αξιολόγησης προς τη διαδικασία πρόβλεψης.

Πριν από την εκτέλεση των προβλέψεων με χρήση ιστορικού πραγματοποιήθηκε ανάλυση της σταθερότητας του διαθέσιμου ιστορικού. Η ανάλυση αυτή περιλαμβάνει τον έλεγχο πληρότητας των ωριαίων εγγραφών, τη μέση και διάμεση ημερήσια κατανάλωση, το εύρος των ημερήσιων τιμών, καθώς και δείκτες σταθερότητας της ημερήσιας κατανάλωσης, του ωριαίου προφίλ και των αιχμών. Με βάση τα αποτελέσματα αυτά, το σύστημα προτείνει κατάλληλη τιμή για τον συντελεστή $\max_alpha$, ο οποίος περιορίζει τον βαθμό επίδρασης της προσαρμοστικής διόρθωσης στην τελική πρόβλεψη.

Οι προβλέψεις πραγματοποιήθηκαν με τα μοντέλα Random Forest, XGBoost και LightGBM, καθώς και με συνδυαστική πρόβλεψη. Για τη συνδυαστική πρόβλεψη

χρησιμοποιήθηκαν προεπιλεγμένα βάρη 25% για το Random Forest, 35% για το XGBoost και 40% για το LightGBM. Η επιλογή αυτή βασίστηκε στη συνολική απόδοση των μοντέλων, δίνοντας μεγαλύτερη βαρύτητα στα μοντέλα διαδοχικής ενίσχυσης δέντρων αποφάσεων, χωρίς όμως να αποκλείεται η συνεισφορά του Random Forest.

Η αξιολόγηση πραγματοποιήθηκε με τις μετρικές MAE, RMSE, MAPE και ημερήσιο απόλυτο σφάλμα. Οι μετρικές MAE και RMSE χρησιμοποιούνται για την αξιολόγηση της ωριαίας ακρίβειας της πρόβλεψης, ενώ το ημερήσιο απόλυτο σφάλμα δείχνει την απόκλιση μεταξύ της πραγματικής και της προβλεπόμενης συνολικής ημερήσιας κατανάλωσης. Η MAPE χρησιμοποιείται συμπληρωματικά, καθώς σε κατοικίες ή ημέρες με χαμηλή κατανάλωση μπορεί να λάβει υψηλές τιμές ακόμη και όταν οι απόλυτες αποκλίσεις είναι σχετικά μικρές.

Οι κώδικες σε γλώσσα Python που χρησιμοποιήθηκαν για την αξιολόγηση των αποτελεσμάτων βρίσκονται στον φάκελο PLEGMA/Evaluation του δημόσιου αποθετηρίου GitHub, στον ακόλουθο σύνδεσμο: <https://github.com/stdGeorgios/Forecasting-the-electricity-consumption-of-buildings->.

16.2 Αξιολόγηση επιλεγμένων κατοικιών του συνόλου ελέγχου

Η κύρια αξιολόγηση του συστήματος πραγματοποιήθηκε σε τέσσερις επιλεγμένες κατοικίες του συνόλου ελέγχου του PLEGMA. Οι ημερομηνίες αξιολόγησης ανήκουν στο test set των μοντέλων και αντιστοιχούν σε πραγματικές ημέρες για τις οποίες ήταν διαθέσιμες τόσο οι πραγματικές τιμές κατανάλωσης όσο και τα απαιτούμενα δεδομένα εισόδου για την εκτέλεση των προβλέψεων μέσω της εφαρμογής.

Οι κατοικίες που επιλέχθηκαν είναι οι home01, home04, home05 και home07. Η επιλογή έγινε με στόχο να καλυφθούν διαφορετικές περιπτώσεις αξιολόγησης.

Στις υποενότητες που ακολουθούν παρουσιάζονται αναλυτικά τα αποτελέσματα για κάθε κατοικία.

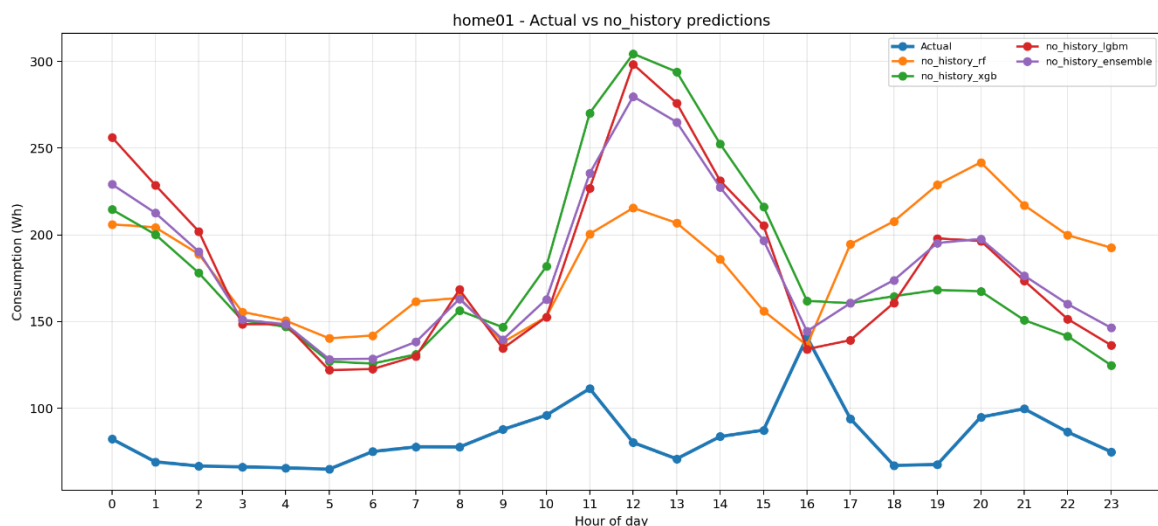
16.2.1 Αποτελέσματα για την κατοικία home01

Η πρώτη κατοικία που εξετάστηκε είναι η home01, με ημέρα αξιολόγησης την Τετάρτη 20 Σεπτεμβρίου 2023. Η κατοικία είναι διαμέρισμα, κατασκευασμένο την περίοδο 1950–1970, με έναν ένοικο και τρία δωμάτια. Διαθέτει πλυντήριο ρούχων, ενώ δεν διαθέτει πλυντήριο πιάτων, στεγνωτήριο ή φούρνο μικροκυμάτων. Η θέρμανση βασίζεται σε θερμαντικά σώματα λαδιού και η παραγωγή ζεστού νερού γίνεται μέσω ηλεκτρικού θερμοσίφωνα. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 1,983 kWh. Στον Πίνακα 40 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	100,52	106,63	129,52	4,386	1,983	2,403
XGBoost	98,00	111,11	124,82	4,335	1,983	2,352
LightGBM	98,80	111,95	127,37	4,341	1,983	2,357
Ensemble	98,62	108,40	126,79	4,350	1,983	2,367

Πίνακας 40 Απόδοση μοντέλων για την κατοικία home01 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, όλα τα μοντέλα υπερεκτιμούν σημαντικά τη συνολική ημερήσια κατανάλωση. Η πραγματική ημερήσια κατανάλωση ήταν 1,983 kWh, ενώ οι προβλέψεις κυμαίνονται περίπου από 4,335 έως 4,386 kWh. Ως αποτέλεσμα, το ημερήσιο σφάλμα υπερβαίνει τα 2,35 kWh σε όλα τα μοντέλα. Το Random Forest παρουσιάζει το χαμηλότερο RMSE, ίσο με 106,63 Wh, ενώ το XGBoost εμφανίζει το χαμηλότερο MAE και MAPE. Παρ' όλα αυτά, οι διαφορές μεταξύ των μοντέλων είναι μικρές και η συνολική εικόνα δείχνει ότι, χωρίς πρόσφατο ιστορικό, το σύστημα δυσκολεύεται να εκτιμήσει σωστά το χαμηλό ημερήσιο επίπεδο κατανάλωσης της συγκεκριμένης κατοικίας.



Σχήμα 34 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home01 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

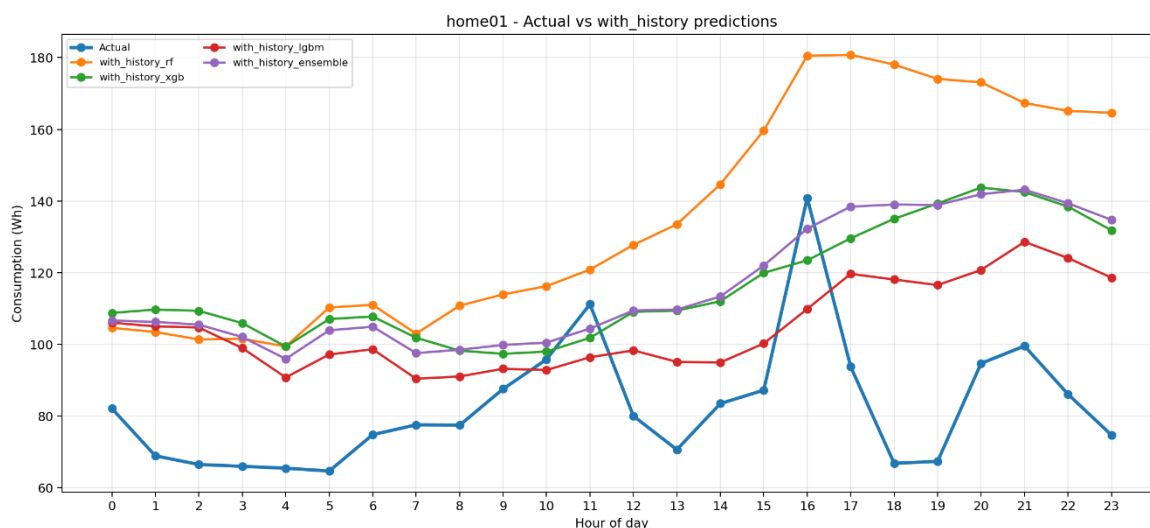
Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 41.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	52,62	59,48	66,87	3,246	1,983	1,263
XGBoost	35,42	39,25	46,38	2,780	1,983	0,797
LightGBM	26,02	28,95	33,84	2,510	1,983	0,527
Ensemble	34,83	39,12	45,61	2,789	1,983	0,805

Πίνακας 41 Απόδοση μοντέλων για την κατοικία home01 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρατηρείται σαφής βελτίωση σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Το LightGBM παρουσιάζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE μειώνεται σε 28,95 Wh, ενώ το ημερήσιο σφάλμα περιορίζεται σε 0,527 kWh. Το XGBoost και το Ensemble εμφανίζουν επίσης σημαντική βελτίωση σε σχέση με το σενάριο χωρίς ιστορικό, ενώ το Random Forest παραμένει πιο αδύναμο από τα υπόλοιπα μοντέλα στο συγκεκριμένο σενάριο.

Η ανάλυση σταθερότητας του ιστορικού έδειξε ότι το διαθέσιμο ιστορικό ήταν πλήρες, ωστόσο, η κατανάλωση των προηγούμενων ημερών παρουσίαζε μεγάλη μεταβλητότητα. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 5,813 kWh, η διάμεση τιμή ήταν 4,684 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 2,131–10,828 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 1,983 kWh, ήταν χαμηλότερη ακόμη και από την ελάχιστη τιμή του πρόσφατου ιστορικού. Επιπλέον, οι δείκτες ημερήσιας σταθερότητας και σταθερότητας ωριαίου προφίλ ήταν μηδενικοί, ενώ η σταθερότητα των αιχμών ήταν περιορισμένη. Για τον λόγο αυτό, το ιστορικό κατατάχθηκε στην κατηγορία Low και το σύστημα πρότεινε χαμηλή τιμή max_alpha, ίση με 0,05.



Σχήμα 35 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home01 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 42, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	RF	100,52	106,63	129,52	4,386	1,983	2,403
with-history adaptive	LGBM	26,02	28,95	33,84	2,510	1,983	0,527

Πίνακας 42 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home01

Για την κατοικία home01, το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση εμφανίζει σαφώς καλύτερη απόδοση από το σενάριο χωρίς ιστορικά δεδομένα. Το RMSE μειώνεται από 106,63 Wh στο καλύτερο μοντέλο χωρίς ιστορικό σε 28,95 Wh στο καλύτερο μοντέλο με χρήση ιστορικού, ενώ το ημερήσιο σφάλμα μειώνεται από 2,403 kWh σε 0,527 kWh. Η μείωση αυτή δείχνει ότι η αξιοποίηση του πρόσφατου ιστορικού βοήθησε σημαντικά στη διόρθωση της υπερεκτίμησης που παρατηρήθηκε στο σενάριο χωρίς ιστορικά δεδομένα.

Ωστόσο, η πρόβλεψη εξακολουθεί να υπερεκτιμά τη συνολική ημερήσια κατανάλωση. Το αποτέλεσμα αυτό είναι ερμηνεύσιμο με βάση την ανάλυση ιστορικού, καθώς το πρόσφατο ιστορικό της κατοικίας παρουσίαζε σημαντικά υψηλότερο επίπεδο κατανάλωσης από την ημέρα στόχο και χαρακτηρίστηκε χαμηλής σταθερότητας. Επομένως, παρότι το ιστορικό βελτίωσε την πρόβλεψη, δεν ήταν πλήρως αντιπροσωπευτικό της ημέρας αξιολόγησης. Η χαμηλή τιμή max_alpha λειτούργησε περιοριστικά, αποτρέποντας υπερβολική προσαρμογή σε ένα ασταθές ιστορικό, αλλά η επίδραση του ιστορικού παρέμεινε αρκετή ώστε να βελτιώσει αισθητά τόσο τις ωριαίες μετρικές όσο και το ημερήσιο σφάλμα.

16.2.2 Αποτελέσματα για την κατοικία home04

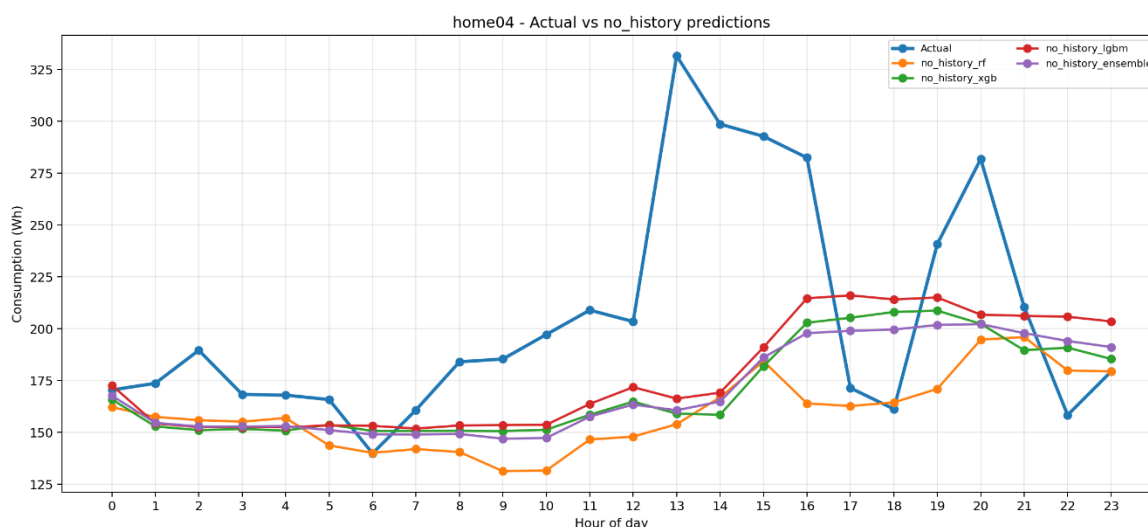
Η δεύτερη κατοικία που εξετάστηκε είναι η home04, με ημέρα αξιολόγησης την Τρίτη 5 Σεπτεμβρίου 2023. Η κατοικία είναι διαμέρισμα, κατασκευασμένο την περίοδο 1990–2010, με δύο ενοίκους και τέσσερα δωμάτια. Διαθέτει πλυντήριο ρούχων και φούρνο μικροκυμάτων, ενώ δεν διαθέτει πλυντήριο πιάτων ή στεγνωτήριο. Η θέρμανση βασίζεται σε φορητά ηλεκτρικά θερμαντικά σώματα και η παραγωγή ζεστού νερού γίνεται μέσω ηλεκτρικού θερμοσίφωνα. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 4,925 kWh.

Στον Πίνακα 43 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	47,75	66,67	19,95	3,830	4,925	1,096
XGBoost	45,36	61,78	19,53	4,097	4,925	0,828
LightGBM	43,60	58,65	19,16	4,249	4,925	0,677
Ensemble	45,00	61,02	19,34	4,091	4,925	0,835

Πίνακας 43 Απόδοση μοντέλων για την κατοικία home04 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το LightGBM παρουσιάζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του LightGBM είναι 58,65 Wh, ενώ το ημερήσιο σφάλμα είναι 0,677 kWh. Όλα τα μοντέλα υποεκτιμούν τη συνολική ημερήσια κατανάλωση, καθώς οι προβλέψεις κυμαίνονται από 3,830 έως 4,249 kWh, έναντι πραγματικής τιμής 4,925 kWh.



Σχήμα 36 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home04 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

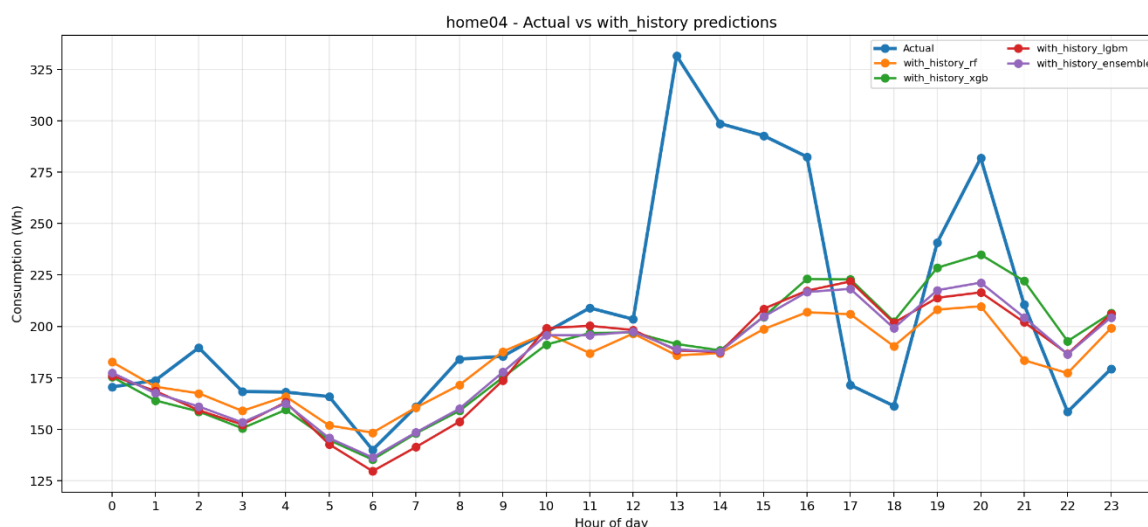
Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 44.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	32,40	49,93	13,24	4,398	4,925	0,527
XGBoost	33,10	47,87	14,42	4,472	4,925	0,453
LightGBM	34,33	49,37	14,86	4,408	4,925	0,518
Ensemble	32,79	48,62	13,96	4,428	4,925	0,498

Πίνακας 44 Απόδοση μοντέλων για την κατοικία home04 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρατηρείται βελτίωση σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Το XGBoost παρουσιάζει το χαμηλότερο RMSE, ίσο με 47,87 Wh και ταυτόχρονα το μικρότερο ημερήσιο σφάλμα, ίσο με 0,453 kWh. Το Random Forest εμφανίζει το χαμηλότερο MAE και MAPE, ενώ το Ensemble βρίσκεται πολύ κοντά στα καλύτερα επιμέρους μοντέλα. Τα αποτελέσματα δείχνουν ότι η χρήση ιστορικών δεδομένων βελτίωσε την ωριαία ακρίβεια και περιόρισε την υποεκτίμηση της ημερήσιας κατανάλωσης.

Η ανάλυση σταθερότητας του ιστορικού έδειξε ότι το διαθέσιμο ιστορικό ήταν πλήρες, ωστόσο, η κατανάλωση των προηγούμενων ημερών παρουσίαζε έντονη μεταβλητότητα. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 8,786 kWh, η διάμεση τιμή ήταν 7,552 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 3,887–17,081 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 4,925 kWh, βρισκόταν εντός του ιστορικού εύρους, αλλά ήταν σημαντικά χαμηλότερη από τη μέση και τη διάμεση τιμή του πρόσφατου ιστορικού. Επιπλέον, οι δείκτες ημερήσιας σταθερότητας και σταθερότητας ωριαίου προφίλ ήταν μηδενικοί, ενώ η σταθερότητα των αιχμών ήταν περιορισμένη. Για τον λόγο αυτό, το ιστορικό κατατάχθηκε στην κατηγορία Low και το σύστημα πρότεινε χαμηλή τιμή max_alpha, ίση με 0,05.



Σχήμα 37 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home04 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 45, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	LGBM	43,60	58,65	19,16	4,249	4,925	0,677
with-history adaptive	XGB	33,10	47,87	14,42	4,472	4,925	0,453

Πίνακας 45 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home04

Για την κατοικία home04, το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρουσιάζει καλύτερη απόδοση από το σενάριο χωρίς ιστορικά δεδομένα. Το RMSE μειώνεται από 58,65 Wh στο καλύτερο μοντέλο χωρίς ιστορικό σε 47,87 Wh στο καλύτερο μοντέλο με χρήση ιστορικού. Παράλληλα, το ημερήσιο σφάλμα μειώνεται από 0,677 kWh σε 0,453 kWh. Η βελτίωση αυτή δείχνει ότι το πρόσφατο ιστορικό συνέβαλε θετικά στην πρόβλεψη, παρά το γεγονός ότι χαρακτηρίστηκε χαμηλής σταθερότητας. Ωστόσο, η επίδραση του ιστορικού παρέμεινε ελεγχόμενη, καθώς το σύστημα πρότεινε χαμηλή τιμή $\max_alpha$. Το αποτέλεσμα αυτό είναι λογικό, επειδή το πρόσφατο ιστορικό παρουσίαζε μεγάλη μεταβλητότητα και υψηλότερο μέσο ημερήσιο επίπεδο κατανάλωσης από την ημέρα στόχο. Επομένως, η προσαρμοστική διόρθωση βοήθησε στη βελτίωση της πρόβλεψης, χωρίς όμως να μεταβάλλει υπερβολικά την αρχική καμπύλη με βάση ένα ιστορικό που δεν ήταν πλήρως σταθερό ή αντιπροσωπευτικό.

16.2.3 Αποτελέσματα για την κατοικία home05

Η τρίτη κατοικία που εξετάστηκε είναι η home05, με ημέρα αξιολόγησης τη Δευτέρα 4 Σεπτεμβρίου 2023. Η κατοικία είναι μονοκατοικία, κατασκευασμένη την περίοδο 1970–1990, με τρεις ενοίκους και τέσσερα δωμάτια. Διαθέτει πλυντήριο ρούχων και πλυντήριο πιάτων, ενώ δεν διαθέτει στεγνωτήριο ή φούρνο μικροκυμάτων. Η θέρμανση βασίζεται σε

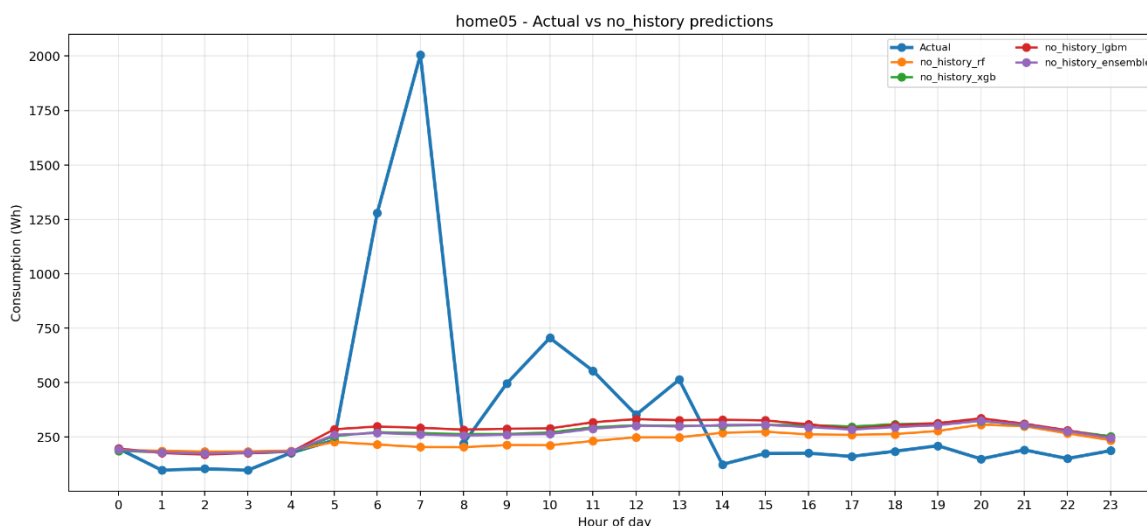
θερμαντικά σώματα λαδιού και η παραγωγή ζεστού νερού γίνεται μέσω λέβητα φυσικού αερίου. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 8,737 kWh. Η κατοικία αυτή αποτελεί τη δυσκολότερη περίπτωση της κύριας αξιολόγησης, καθώς παρουσιάζει υψηλότερη ημερήσια κατανάλωση και διαφορετικό τύπο κατοικίας σε σχέση με τις προηγούμενες περιπτώσεις.

Στον Πίνακα 46 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	234,86	457,55	55,89	5,644	8,737	3,093
XGBoost	230,06	437,35	59,68	6,372	8,737	2,364
LightGBM	226,64	429,50	60,54	6,651	8,737	2,086
Ensemble	229,36	438,86	58,84	6,302	8,737	2,435

Πίνακας 46 Απόδοση μοντέλων για την κατοικία home05 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το LightGBM παρουσιάζει την καλύτερη συνολική απόδοση ως προς το MAE, το RMSE και το ημερήσιο σφάλμα. Συγκεκριμένα, το RMSE του LightGBM είναι 429,50 Wh, ενώ το ημερήσιο σφάλμα είναι 2,086 kWh. Ωστόσο, όλα τα μοντέλα υποεκτιμούν σημαντικά τη συνολική ημερήσια κατανάλωση. Η πραγματική κατανάλωση της ημέρας ήταν 8,737 kWh, ενώ οι προβλέψεις κυμαίνονται από 5,644 έως 6,651 kWh. Το αποτέλεσμα αυτό δείχνει ότι, χωρίς πρόσφατο ιστορικό, το σύστημα δυσκολεύεται να εκτιμήσει το υψηλότερο ημερήσιο επίπεδο κατανάλωσης της συγκεκριμένης μονοκατοικίας.



Σχήμα 38 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home05 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

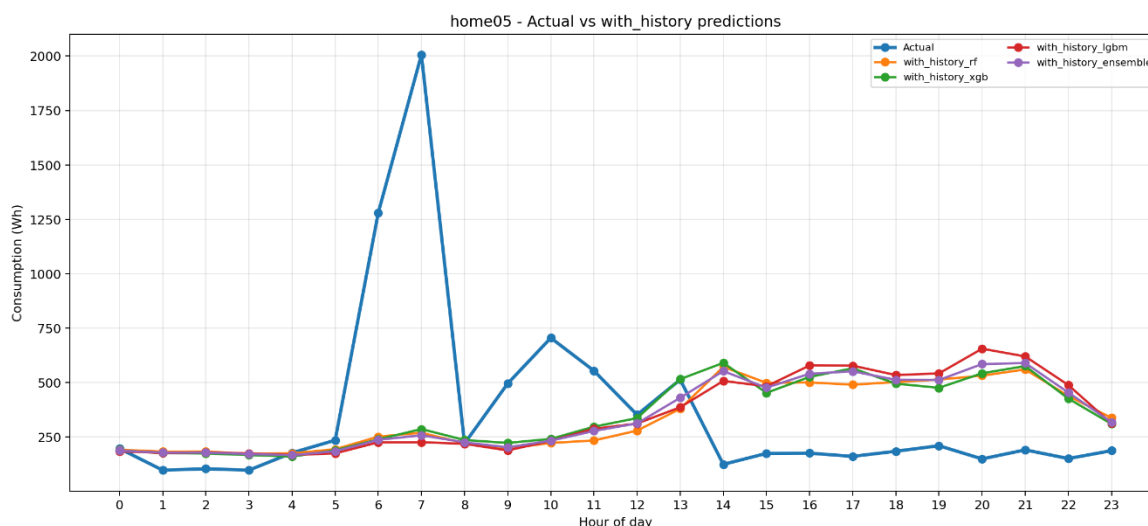
Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 47.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	316,77	485,00	110,06	8,094	8,737	0,643
XGBoost	305,52	481,31	108,25	8,384	8,737	0,353
LightGBM	331,58	503,84	118,20	8,427	8,737	0,310
Ensemble	318,63	490,65	112,64	8,329	8,737	0,408

Πίνακας 47 Απόδοση μοντέλων για την κατοικία home05 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρατηρείται διαφορετική εικόνα σε σχέση με τις προηγούμενες κατοικίες. Το XGBoost παρουσιάζει το χαμηλότερο MAE και RMSE, με RMSE ίσο με 481,31 Wh, ενώ το LightGBM εμφανίζει το μικρότερο ημερήσιο σφάλμα, ίσο με 0,310 kWh. Παρότι η χρήση ιστορικών δεδομένων βελτιώνει σημαντικά την εκτίμηση της συνολικής ημερήσιας κατανάλωσης, δεν βελτιώνει την ωριαία ακρίβεια. Αντίθετα, οι τιμές MAE, RMSE και MAPE είναι υψηλότερες σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα.

Η ανάλυση σταθερότητας του ιστορικού έδειξε ότι το διαθέσιμο ιστορικό ήταν πλήρες, καθώς περιλάμβανε και τις 168 αναμενόμενες ωριαίες εγγραφές χωρίς ελλείψεις. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 12,837 kWh, η διάμεση τιμή ήταν 12,534 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 11,098–15,114 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 8,737 kWh, ήταν χαμηλότερη από το πρόσφατο ιστορικό εύρος. Παρότι η ημερήσια κατανάλωση του ιστορικού παρουσίαζε σχετικά μικρή μεταβλητότητα, το ωριαίο προφίλ δεν ήταν σταθερό, καθώς ο αντίστοιχος δείκτης σταθερότητας ήταν μηδενικός. Για τον λόγο αυτό, το ιστορικό κατατάχθηκε στην κατηγορία Medium και το σύστημα πρότεινε χαμηλή τιμή max_alpha, ίση με 0,10.



Σχήμα 39 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home05 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 48, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	LGBM	226,64	429,50	60,54	6,651	8,737	2,086
with-history adaptive	XGB	305,52	481,31	108,25	8,384	8,737	0,353

Πίνακας 48 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home05

Για την κατοικία home05, η χρήση ιστορικών δεδομένων και προσαρμοστικής διόρθωσης δεν οδηγεί σε βελτίωση της ωριαίας ακρίβειας. Το RMSE αυξάνεται από 429,50 Wh στο καλύτερο μοντέλο χωρίς ιστορικό σε 481,31 Wh στο καλύτερο μοντέλο με χρήση ιστορικού. Αντίθετα, το ημερήσιο σφάλμα μειώνεται σημαντικά, από 2,086 kWh σε 0,353 kWh, γεγονός που δείχνει ότι το πρόσφατο ιστορικό βοήθησε στην καλύτερη εκτίμηση του συνολικού ημερήσιου επιπέδου κατανάλωσης, αλλά όχι στην ακριβή αναπαραγωγή της ωριαίας καμπύλης.

Το αποτέλεσμα αυτό είναι ιδιαίτερα χρήσιμο για την αξιολόγηση των περιορισμών του συστήματος. Το πρόσφατο ιστορικό της home05 είχε υψηλότερο ημερήσιο επίπεδο κατανάλωσης από την ημέρα στόχο, ενώ το ωριαίο προφίλ του δεν παρουσίαζε σταθερή επαναληψιμότητα. Έτσι, η χρήση ιστορικών δεδομένων βοήθησε να διορθωθεί η έντονη υποεκτίμηση της ημερήσιας κατανάλωσης που εμφανίστηκε στο σενάριο χωρίς ιστορικό, αλλά δεν ήταν αρκετή για να βελτιώσει τις ωριαίες μετρικές. Η κατοικία home05 αναδεικνύει ότι η αξιολόγηση πρέπει να λαμβάνει υπόψη ταυτόχρονα τόσο την ωριαία ακρίβεια όσο και το ημερήσιο σφάλμα, καθώς η βελτίωση της συνολικής ημερήσιας κατανάλωσης δεν συνεπάγεται απαραίτητα και βελτίωση της μορφής της ωριαίας καμπύλης.

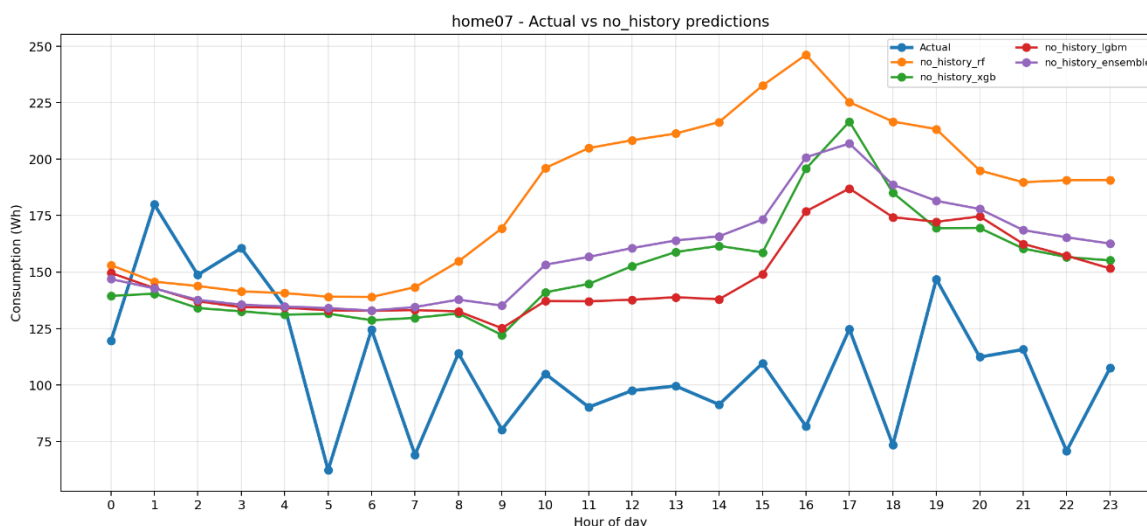
16.2.4 Αποτελέσματα για την κατοικία home07

Η τέταρτη κατοικία που εξετάστηκε είναι η home07, με ημέρα αξιολόγησης την Παρασκευή 22 Σεπτεμβρίου 2023. Η κατοικία είναι διαμέρισμα, κατασκευασμένο την περίοδο 1970–1990, με έναν ένοικο και τέσσερα δωμάτια. Διαθέτει πλυντήριο ρούχων, πλυντήριο πιάτων και φούρνο μικροκυμάτων, ενώ δεν διαθέτει στεγνωτήριο. Η θέρμανση βασίζεται σε αντλία θερμότητας και η παραγωγή ζεστού νερού γίνεται μέσω ηλεκτρικού θερμοσίφωνα. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 2,620 kWh. Στον Πίνακα 49 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	79,30	90,47	85,41	4,407	2,620	1,787
XGBoost	49,93	58,08	54,41	3,647	2,620	1,027
LightGBM	44,97	51,62	49,37	3,549	2,620	0,929
Ensemble	55,18	62,87	60,06	3,798	2,620	1,178

Πίνακας 49 Απόδοση μοντέλων για την κατοικία home07 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το LightGBM παρουσιάζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του LightGBM είναι 51,62 Wh, ενώ το ημερήσιο σφάλμα είναι 0,929 kWh. Παρότι το LightGBM είναι το καλύτερο μοντέλο στο συγκεκριμένο σενάριο, όλα τα μοντέλα υπερεκτιμούν τη συνολική ημερήσια κατανάλωση. Οι προβλέψεις κυμαίνονται από 3,549 έως 4,407 kWh, ενώ η πραγματική ημερήσια κατανάλωση είναι 2,620 kWh.



Σχήμα 40 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home07 στο σενάριο χωρίς χρήση ιστορικών δεδομένων

Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 50.

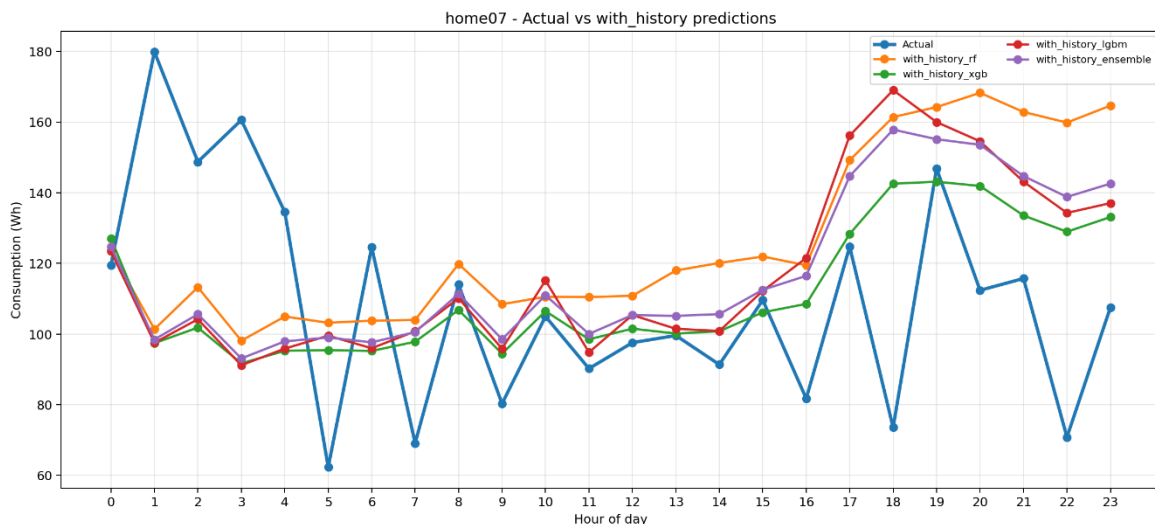
Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	35,67	43,25	36,41	3,022	2,620	0,402
XGBoost	25,80	34,95	25,27	2,676	2,620	0,056
LightGBM	30,62	39,87	30,45	2,820	2,620	0,200
Ensemble	29,86	38,42	29,85	2,820	2,620	0,200

Πίνακας 50 Απόδοση μοντέλων για την κατοικία home07 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρατηρείται σαφής βελτίωση σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Το XGBoost παρουσιάζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Επιπλέον, εμφανίζει και το μικρότερο ημερήσιο σφάλμα, καθώς προβλέπει συνολική ημερήσια κατανάλωση 2,676 kWh έναντι πραγματικής τιμής 2,620 kWh, με απόκλιση μόλις 0,056 kWh. Το αποτέλεσμα αυτό δείχνει ότι η χρήση πρόσφατου ιστορικού βοήθησε σημαντικά τόσο στην ωριαία ακρίβεια όσο και στην εκτίμηση της συνολικής ημερήσιας κατανάλωσης.

Η ανάλυση σταθερότητας του ιστορικού έδειξε ότι το διαθέσιμο ιστορικό ήταν πλήρες. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 2,920 kWh, η διάμεση τιμή ήταν 2,795 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 2,465–4,136 kWh. Η πραγματική

κατανάλωση της ημέρας αξιολόγησης, ίση με 2,620 kWh, βρισκόταν εντός του ιστορικού εύρους και κοντά στη διάμεση τιμή του πρόσφατου ιστορικού. Παρότι η ημερήσια σταθερότητα και η σταθερότητα των αιχμών δεν ήταν πολύ υψηλές, το ωριαίο προφίλ παρουσίασε καλύτερη σταθερότητα. Για τον λόγο αυτό, το ιστορικό κατατάχθηκε στην κατηγορία Medium και το σύστημα πρότεινε χαμηλή τιμή max_alpha, ίση με 0,10.



Σχήμα 41 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home07 στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 51, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	LGBM	44,97	51,62	49,37	3,549	2,620	0,929
with-history adaptive	XGB	25,80	34,95	25,27	2,676	2,620	0,056

Πίνακας 51 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home07

Για την κατοικία home07, το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρουσιάζει σαφή βελτίωση σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Το RMSE μειώνεται από 51,62 Wh στο καλύτερο μοντέλο χωρίς ιστορικό σε 34,95 Wh στο καλύτερο μοντέλο με χρήση ιστορικού. Παράλληλα, το ημερήσιο σφάλμα μειώνεται από 0,929 kWh σε 0,056 kWh, γεγονός που δείχνει ότι η χρήση ιστορικού βοήθησε ιδιαίτερα στην εκτίμηση του συνολικού ημερήσιου επιπέδου κατανάλωσης.

Το αποτέλεσμα αυτό είναι ερμηνεύσιμο με βάση την ανάλυση ιστορικού. Σε αντίθεση με τις κατοικίες home01 και home04, όπου η πραγματική ημερήσια κατανάλωση της ημέρας στόχου ήταν αρκετά χαμηλότερη από τη μέση και τη διάμεση κατανάλωση του πρόσφατου ιστορικού, καθώς και με την κατοικία home05, όπου η ημέρα στόχος ήταν χαμηλότερη ακόμη και από το πρόσφατο ιστορικό εύρος, στην κατοικία home07, η ημέρα αξιολόγησης

βρισκόταν εντός του πρόσφατου ιστορικού εύρους και σχετικά κοντά στη διάμεση ημερήσια τιμή. Επομένως, το πρόσφατο ιστορικό ήταν πιο αντιπροσωπευτικό της ημέρας στόχου, γεγονός που επέτρεψε στο σενάριο με χρήση ιστορικών δεδομένων να βελτιώσει ουσιαστικά την πρόβλεψη. Η χαμηλή τιμή $\max_alpha$ διατήρησε την προσαρμοστική διόρθωση ελεγχόμενη, χωρίς να οδηγήσει σε υπερβολική προσαρμογή στο ιστορικό.

16.3 Εποχικές μελέτες περίπτωσης

Μετά την κύρια αξιολόγηση των επιλεγμένων κατοικιών του συνόλου ελέγχου, πραγματοποιήθηκαν πρόσθετες εποχικές μελέτες περίπτωσης. Σκοπός της ενότητας αυτής είναι να εξεταστεί η συμπεριφορά του συστήματος πρόβλεψης σε διαφορετικές χρονικές περιόδους του έτους και να διερευνηθεί αν τα συμπεράσματα της προηγούμενης αξιολόγησης διατηρούνται και σε ημέρες με διαφορετικά εποχικά χαρακτηριστικά.

Για τον σκοπό αυτό επιλέχθηκαν ενδεικτικές ημέρες από διαφορετικές εποχές για τις ίδιες κατοικίες που εξετάστηκαν στην κύρια αξιολόγηση. Συγκεκριμένα, εξετάζονται η home01 σε θερινή περίπτωση, η home04 σε φθινοπωρινή περίπτωση, η home05 σε χειμερινή περίπτωση και η home07 σε εαρινή περίπτωση. Με αυτόν τον τρόπο διατηρείται η συνέχεια με την προηγούμενη ενότητα, ενώ παράλληλα αξιολογείται η συμπεριφορά του συστήματος σε διαφορετικές περιόδους και επίπεδα κατανάλωσης.

Οι εποχικές μελέτες περίπτωσης χρησιμοποιούνται ως συμπληρωματικά λειτουργικά παραδείγματα αξιολόγησης και όχι ως κύρια απόδειξη γενικευτικής ικανότητας των μοντέλων. Η κύρια αξιολόγηση της γενίκευσης βασίζεται στις κατοικίες και στις ημέρες του συνόλου ελέγχου, ενώ οι εποχικές περιπτώσεις αξιοποιούνται κυρίως για τη διερεύνηση της συμπεριφοράς του συστήματος υπό διαφορετικές εποχικές συνθήκες και διαφορετικά επίπεδα σταθερότητας ιστορικού.

16.3.1 Θερινή αξιολόγηση για την κατοικία home01

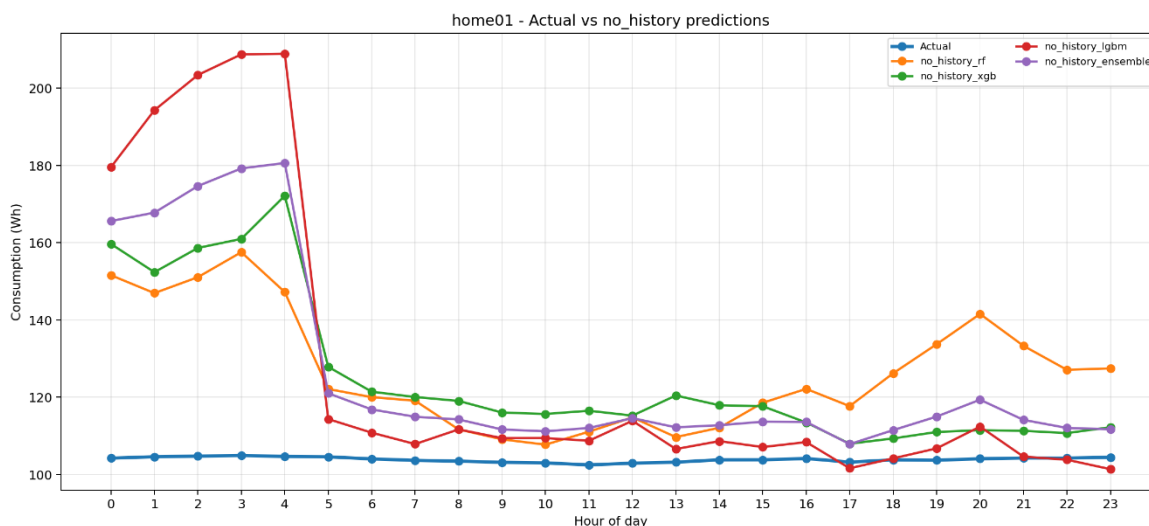
Η πρώτη εποχική μελέτη περίπτωσης αφορά την κατοικία home01, με ημέρα αξιολόγησης το Σάββατο 13 Αυγούστου 2022. Η ημερομηνία αυτή αντιστοιχεί σε θερινή περίοδο και επιλέχθηκε ώστε να εξεταστεί η συμπεριφορά του συστήματος σε διαφορετική χρονική περίοδο από την κύρια αξιολόγηση της ίδιας κατοικίας. Η κατοικία είναι διαμέρισμα, κατασκευασμένο την περίοδο 1950–1970, με έναν ένοικο και τρία δωμάτια. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 2,494 kWh.

Στον Πίνακα 52 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	22,76	26,97	21,83	3,040	2,494	0,546
XGBoost	21,08	28,18	20,22	3,000	2,494	0,506
LightGBM	23,47	43,63	22,46	3,047	2,494	0,553
Ensemble	22,29	32,89	21,36	3,029	2,494	0,535

Πίνακας 52 Απόδοση μοντέλων για την κατοικία home01 στη θερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, τα μοντέλα παρουσιάζουν σχετικά περιορισμένα ωριαία σφάλματα, αλλά υπερεκτιμούν τη συνολική ημερήσια κατανάλωση. Η πραγματική ημερήσια κατανάλωση ήταν 2,494 kWh, ενώ οι προβλέψεις κυμαίνονται περίπου από 3,000 έως 3,047 kWh. Το Random Forest παρουσιάζει το χαμηλότερο RMSE, ίσο με 26,97 Wh, ενώ το XGBoost εμφανίζει το χαμηλότερο MAE και MAPE. Παρ' όλα αυτά, όλα τα μοντέλα εμφανίζουν ημερήσιο σφάλμα περίπου 0,51–0,55 kWh, γεγονός που δείχνει ότι χωρίς ιστορικό το σύστημα τείνει να εκτιμά υψηλότερο ημερήσιο επίπεδο κατανάλωσης.



Σχήμα 42 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home01 στη θερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

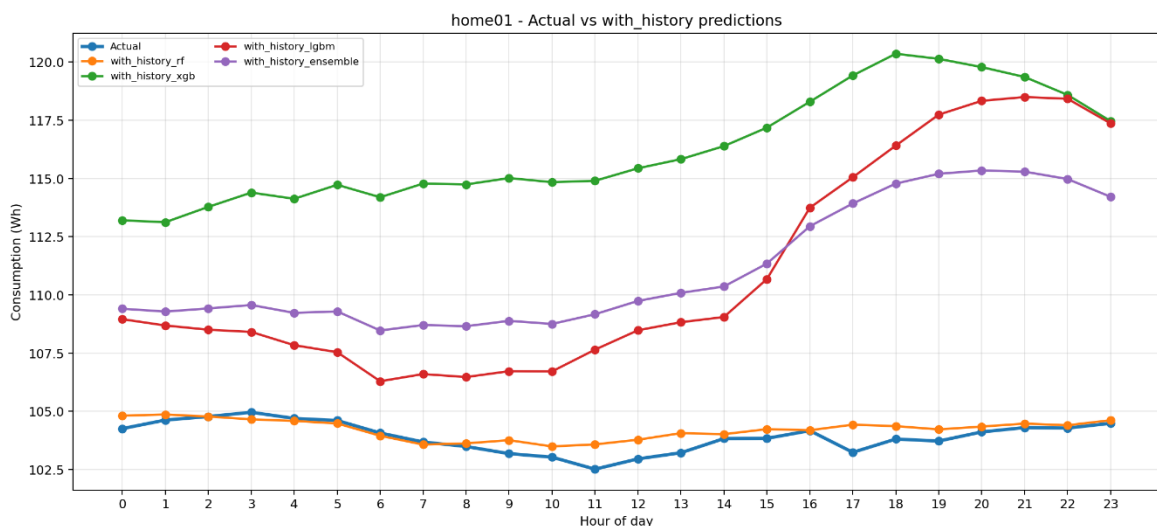
Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 53.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	0,37	0,49	0,36	2,501	2,494	0,007
XGBoost	12,34	12,58	11,88	2,790	2,494	0,296
LightGBM	7,05	8,29	6,78	2,663	2,494	0,169
Ensemble	7,22	7,65	6,95	2,667	2,494	0,173

Πίνακας 53 Απόδοση μοντέλων για την κατοικία home01 στη θερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρατηρείται πολύ μεγάλη βελτίωση σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Το Random Forest παρουσιάζει την καλύτερη συνολική απόδοση, με MAE 0,37 Wh, RMSE 0,49 Wh και MAPE 0,36%. Παράλληλα, εκτιμά σχεδόν με ακρίβεια τη συνολική ημερήσια κατανάλωση, καθώς η προβλεπόμενη τιμή είναι 2,501 kWh έναντι πραγματικής τιμής 2,494 kWh, με ημερήσιο σφάλμα μόλις 0,007 kWh. Τα υπόλοιπα μοντέλα με χρήση ιστορικού παρουσιάζουν επίσης καλύτερη απόδοση σε σχέση με το αντίστοιχο σενάριο χωρίς ιστορικά δεδομένα, αν και δεν φτάνουν την ακρίβεια του Random Forest.

Η ανάλυση σταθερότητας έδειξε ότι το διαθέσιμο ιστορικό ήταν πλήρες, καθώς περιλάμβανε και τις 168 αναμενόμενες ωριαίες εγγραφές χωρίς ελλείψεις. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 2,493 kWh, η διάμεση τιμή ήταν 2,491 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν ιδιαίτερα περιορισμένο, από 2,483 έως 2,505 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 2,494 kWh, βρισκόταν εντός του ιστορικού εύρους και ήταν σχεδόν ταυτόσημη με τη μέση και τη διάμεση τιμή του πρόσφατου ιστορικού. Επιπλέον, οι δείκτες σταθερότητας της ημερήσιας κατανάλωσης, του ωριαίου προφίλ και των αιχμών ήταν εξαιρετικά υψηλοί. Για τον λόγο αυτό, το ιστορικό κατατάχθηκε στην κατηγορία Excellent και το σύστημα πρότεινε υψηλή τιμή max_alpha, ίση με 0,60.



Σχήμα 43 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home01 στη θερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 54, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	RF	22,76	26,97	21,83	3,040	2,494	0,546
with-history adaptive	RF	0,37	0,49	0,36	2,501	2,494	0,007

Πίνακας 54 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home01 στη θερινή περίπτωση

Για τη θερινή περίπτωση της home01, η χρήση ιστορικών δεδομένων και προσαρμοστικής διόρθωσης οδηγεί σε εξαιρετικά μεγάλη βελτίωση της πρόβλεψης. Το RMSE μειώνεται από 26,97 Wh στο καλύτερο μοντέλο χωρίς ιστορικό σε 0,49 Wh στο καλύτερο μοντέλο με χρήση ιστορικού. Παράλληλα, το ημερήσιο σφάλμα μειώνεται από 0,546 kWh σε 0,007 kWh, γεγονός που δείχνει ότι το πρόσφατο ιστορικό συνέβαλε τόσο στη βελτίωση της ωριαίας καμπύλης όσο και στην ακριβέστερη εκτίμηση της συνολικής ημερήσιας κατανάλωσης.

Το ιδιαίτερα χαμηλό σφάλμα της συγκεκριμένης περίπτωσης πρέπει να ερμηνευθεί σε συνάρτηση με την εξαιρετικά υψηλή επαναληψιμότητα του πρόσφατου ιστορικού. Δεν αποτελεί γενικό αναμενόμενο επίπεδο απόδοσης για όλες τις κατοικίες, αλλά δείχνει ότι, όταν το ιστορικό είναι πλήρες, σταθερό και σχεδόν ταυτόσημο με τη συμπεριφορά της ημέρας στόχου, η χρήση ιστορικών δεδομένων μπορεί να οδηγήσει σε πολύ ακριβείς προβλέψεις.

16.3.2 Φθινοπωρινή αξιολόγηση για την κατοικία home04

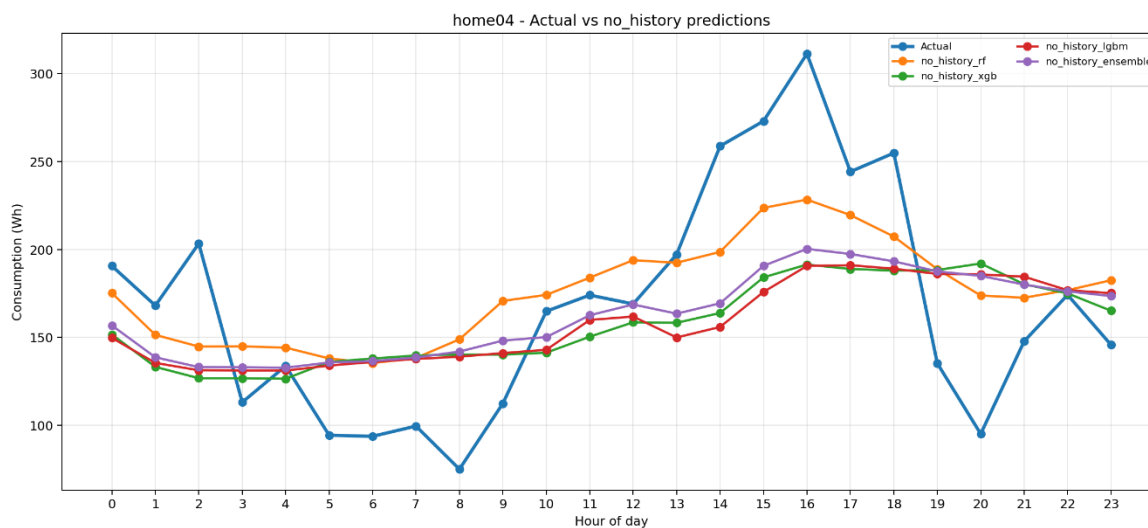
Η δεύτερη εποχική μελέτη περίπτωσης αφορά την κατοικία home04, με ημέρα αξιολόγησης την Κυριακή 23 Οκτωβρίου 2022. Η ημερομηνία αυτή αντιστοιχεί σε φθινοπωρινή περίοδο και επιλέχθηκε ώστε να εξεταστεί η συμπεριφορά του συστήματος σε διαφορετική χρονική περίοδο από την κύρια αξιολόγηση της ίδιας κατοικίας. Η κατοικία είναι διαμέρισμα, κατασκευασμένο την περίοδο 1990–2010, με δύο ενοίκους και τέσσερα δωμάτια. Διαθέτει πλυντήριο ρούχων και φούρνο μικροκυμάτων, ενώ δεν διαθέτει πλυντήριο πιάτων ή στεγνωτήριο. Η θέρμανση βασίζεται σε φορητά ηλεκτρικά θερμαντικά σώματα και η παραγωγή ζεστού νερού γίνεται μέσω ηλεκτρικού θερμοσίφωνα. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 4,029 kWh.

Στον Πίνακα 55 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	37,47	44,13	27,24	4,208	4,029	0,179
XGBoost	46,49	55,67	30,30	3,784	4,029	0,245
LightGBM	46,65	56,19	30,06	3,792	4,029	0,237
Ensemble	43,16	52,19	28,73	3,893	4,029	0,136

Πίνακας 55 Απόδοση μοντέλων για την κατοικία home04 στη φθινοπωρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το Random Forest παρουσιάζει την καλύτερη ωριαία απόδοση, επιτυγχάνοντας το χαμηλότερο MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του Random Forest είναι 44,13 Wh. Ωστόσο, ως προς το ημερήσιο σφάλμα, η συνδυαστική πρόβλεψη εμφανίζει την καλύτερη επίδοση, με απόκλιση 0,136 kWh από την πραγματική ημερήσια κατανάλωση. Συνολικά, τα μοντέλα εμφανίζουν σχετικά καλή απόδοση, καθώς οι προβλέψεις βρίσκονται κοντά στην πραγματική τιμή των 4,029 kWh.



Σχήμα 44 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home04 στη φθινοπωρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

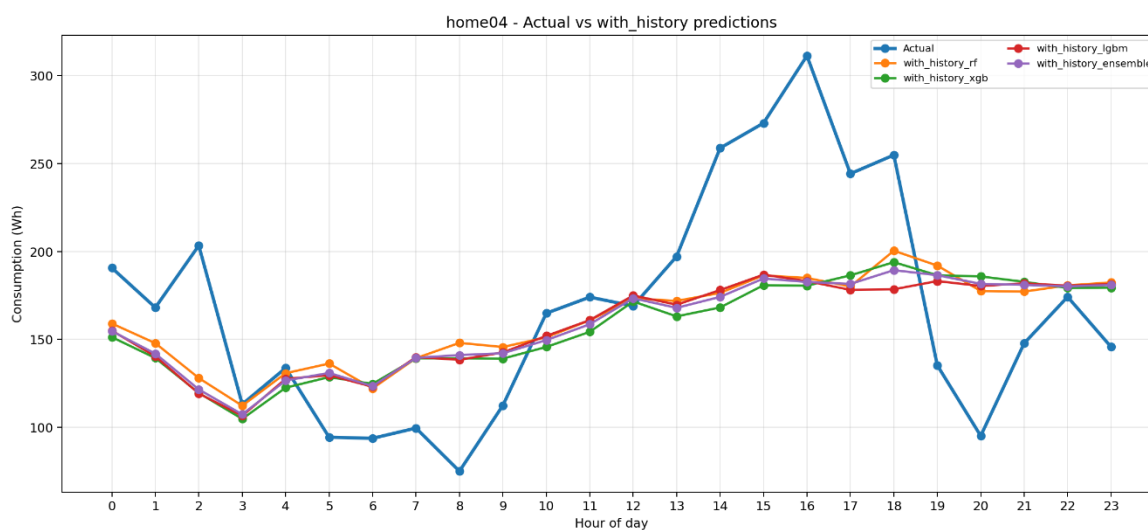
Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 56.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	42,90	53,35	27,95	3,864	4,029	0,165
XGBoost	45,45	55,60	29,01	3,766	4,029	0,263
LightGBM	44,39	54,63	28,24	3,791	4,029	0,238
Ensemble	44,39	54,55	28,44	3,801	4,029	0,228

Πίνακας 56 Απόδοση μοντέλων για την κατοικία home04 στη φθινοπωρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, το Random Forest παρουσιάζει την καλύτερη συνολική απόδοση, με το χαμηλότερο MAE, RMSE και MAPE. Το RMSE του Random Forest είναι 53,35 Wh, ενώ το ημερήσιο σφάλμα είναι 0,165 kWh. Παρότι οι αποκλίσεις παραμένουν σχετικά περιορισμένες, η χρήση ιστορικών δεδομένων δεν οδηγεί σε βελτίωση σε σχέση με το καλύτερο αποτέλεσμα του σεναρίου χωρίς ιστορικά δεδομένα. Αντίθετα, οι ωριαίες μετρικές είναι ελαφρώς χειρότερες, ενώ και το ημερήσιο σφάλμα παραμένει κοντά αλλά όχι χαμηλότερο από το καλύτερο no-history αποτέλεσμα.

Η ανάλυση σταθερότητας έδειξε ότι το διαθέσιμο ιστορικό ήταν πλήρες. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 4,294 kWh, η διάμεση τιμή ήταν 4,423 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 3,284–5,014 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 4,029 kWh, βρισκόταν εντός του ιστορικού εύρους και σχετικά κοντά στο μέσο επίπεδο του πρόσφατου ιστορικού. Οι δείκτες σταθερότητας της ημερήσιας κατανάλωσης, του ωριαίου προφίλ και των αιχμών ήταν μέτριοι προς υψηλοί. Για τον λόγο αυτό, το ιστορικό κατατάχθηκε στην κατηγορία High και το σύστημα πρότεινε τιμή max_alpha ίση με 0,25.



Σχήμα 45 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home04 στη φθινοπωρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 57, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	RF	37,47	44,13	27,24	4,208	4,029	0,179
with-history adaptive	RF	42,90	53,35	27,95	3,864	4,029	0,165

Πίνακας 57 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home04 στη φθινοπωρινή περίπτωση

Για τη φθινοπωρινή περίπτωση της home04, η χρήση ιστορικών δεδομένων και προσαρμοστικής διόρθωσης δεν οδηγεί σε βελτίωση της ωριαίας ακρίβειας. Το RMSE αυξάνεται από 44,13 Wh στο καλύτερο μοντέλο χωρίς ιστορικό σε 53,35 Wh στο καλύτερο μοντέλο με χρήση ιστορικού. Ως προς το ημερήσιο σφάλμα, το with-history Random Forest εμφανίζει οριακά μικρότερη απόκλιση από το αντίστοιχο no-history Random Forest, καθώς το σφάλμα μειώνεται από 0,179 kWh σε 0,165 kWh. Ωστόσο, η καλύτερη ημερήσια εκτίμηση της συγκεκριμένης περίπτωσης παραμένει η συνδυαστική πρόβλεψη του σεναρίου χωρίς ιστορικά δεδομένα, με ημερήσιο σφάλμα 0,136 kWh.

Το αποτέλεσμα αυτό δεν αναιρεί τη χρησιμότητα του ιστορικού, αλλά δείχνει ότι η θετική επίδρασή του δεν είναι δεδομένη σε κάθε περίπτωση και σε κάθε μετρική. Στην home04, το πρόσφατο ιστορικό ήταν πλήρες και κατατάχθηκε στην κατηγορία High, ωστόσο δεν ήταν απόλυτα αντιπροσωπευτικό της ωριαίας μορφής της ημέρας στόχου. Παράλληλα, το σενάριο χωρίς χρήση ιστορικών δεδομένων και ειδικότερα το Random Forest, είχε ήδη αποδώσει αρκετά καλά την καμπύλη της συγκεκριμένης ημέρας.

Η χρήση ιστορικού οδήγησε σε οριακά μικρότερο ημερήσιο σφάλμα, γεγονός που δείχνει ότι το πρόσφατο ιστορικό βοήθησε σε επίπεδο συνολικής ημερήσιας κατανάλωσης. Ωστόσο, οι ωριαίες μετρικές δεν βελτιώθηκαν, καθώς η μορφή της καμπύλης που προέκυψε μετά τη διόρθωση δεν ταίριαζε καλύτερα με την πραγματική ωριαία κατανομή της ημέρας

στόχου. Συνεπώς, η περίπτωση αυτή δείχνει ότι η σταθερότητα του ιστορικού αποτελεί σημαντική αλλά όχι επαρκή συνθήκη για τη βελτίωση της πρόβλεψης. Εξίσου σημαντικό είναι το πρόσφατο ιστορικό να είναι αντιπροσωπευτικό όχι μόνο ως προς το συνολικό ημερήσιο επίπεδο, αλλά και ως προς τη χρονική κατανομή της κατανάλωσης μέσα στην ημέρα.

16.3.3 Χειμερινή αξιολόγηση για την κατοικία home05

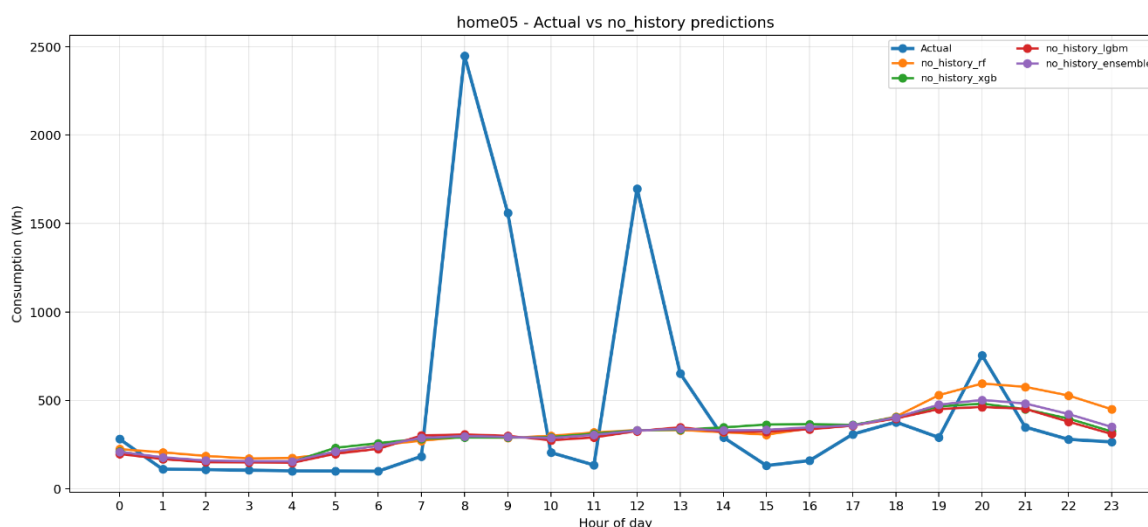
Η τρίτη εποχική μελέτη περίπτωσης αφορά την κατοικία home05, με ημέρα αξιολόγησης την Τετάρτη 15 Φεβρουαρίου 2023. Η ημερομηνία αυτή αντιστοιχεί σε χειμερινή περίοδο και επιλέχθηκε ώστε να εξεταστεί η συμπεριφορά του συστήματος σε περίοδο κατά την οποία η κατανάλωση μπορεί να επηρεάζεται περισσότερο από τις ανάγκες θέρμανσης και τη γενικότερη εποχική συμπεριφορά του νοικοκυριού. Η κατοικία είναι μονοκατοικία, κατασκευασμένη την περίοδο 1970–1990, με τρεις ενοίκους και τέσσερα δωμάτια. Διαθέτει πλυντήριο ρούχων και πλυντήριο πιάτων, ενώ δεν διαθέτει στεγνωτήριο ή φούρνο μικροκυμάτων. Η θέρμανση βασίζεται σε θερμαντικά σώματα λαδιού και η παραγωγή ζεστού νερού γίνεται μέσω λέβητα φυσικού αερίου. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 11 kWh.

Στον Πίνακα 58 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	317,56	600,02	70,69	7,960	11	3,040
XGBoost	305,89	598,10	65,41	7,428	11	3,572
LightGBM	295,27	592,98	58,45	7,177	11	3,823
Ensemble	304,56	595,96	63,94	7,461	11	3,540

Πίνακας 58 Απόδοση μοντέλων για την κατοικία home05 στη χειμερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το LightGBM παρουσιάζει την καλύτερη ωριαία απόδοση, επιτυγχάνοντας τα χαμηλότερα MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του LightGBM είναι 592,98 Wh. Ωστόσο, όλα τα μοντέλα υποεκτιμούν σημαντικά τη συνολική ημερήσια κατανάλωση. Η πραγματική κατανάλωση της ημέρας ήταν 11,000 kWh, ενώ οι προβλέψεις κυμαίνονται από 7,177 έως 7,960 kWh. Ως προς το ημερήσιο σφάλμα, το Random Forest εμφανίζει την καλύτερη επίδοση, με απόκλιση 3,040 kWh, αν και η τιμή αυτή παραμένει υψηλή.



Σχήμα 46 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home05 στη χειμερινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

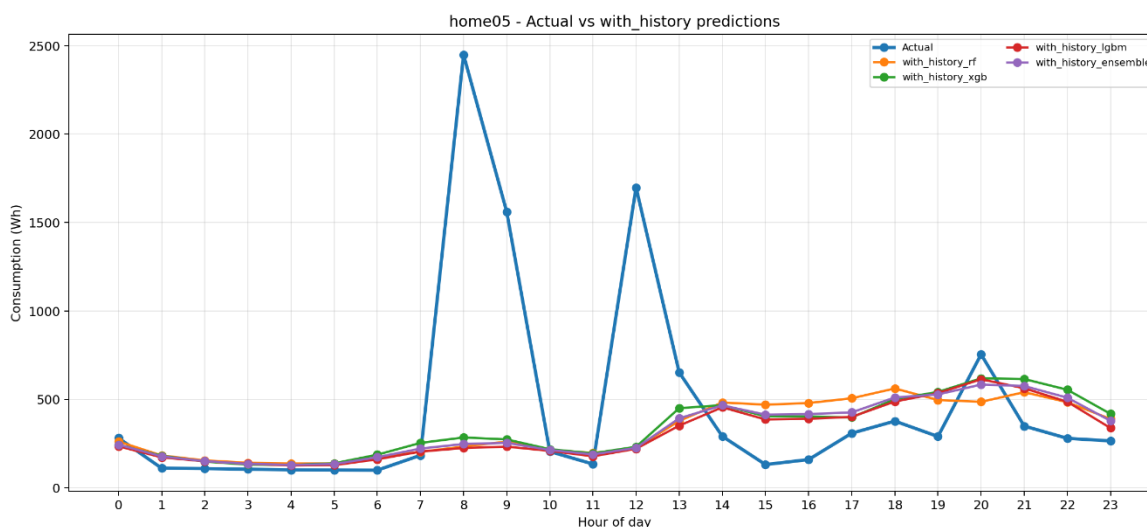
Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 59.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	329,14	625,71	65,64	7,801	11	3,199
XGBoost	315,75	612,24	61,97	7,989	11	3,012
LightGBM	309,26	623,73	55,20	7,388	11	3,613
Ensemble	316,50	619,68	60,18	7,701	11	3,299

Πίνακας 59 Απόδοση μοντέλων για την κατοικία home05 στη χειμερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση, το XGBoost παρουσιάζει το χαμηλότερο RMSE, ίσο με 612,24 Wh, ενώ το LightGBM εμφανίζει το χαμηλότερο MAE και MAPE. Παρ' όλα αυτά, η χρήση ιστορικών δεδομένων δεν βελτιώνει την ωριαία ακρίβεια σε σχέση με το καλύτερο αποτέλεσμα του σεναρίου χωρίς ιστορικά δεδομένα. Οι προβλέψεις εξακολουθούν να υποεκτιμούν τη συνολική ημερήσια κατανάλωση, καθώς κυμαίνονται από 7,388 έως 7,989 kWh, έναντι πραγματικής τιμής 11,000 kWh. Το μικρότερο ημερήσιο σφάλμα στο σενάριο με ιστορικό εμφανίζεται στο XGBoost και είναι 3,012 kWh, δηλαδή μόνο οριακά χαμηλότερο από το καλύτερο ημερήσιο σφάλμα του σεναρίου χωρίς ιστορικά δεδομένα.

Η ανάλυση σταθερότητας έδειξε ότι το διαθέσιμο ιστορικό ήταν πλήρες. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 11,448 kWh, η διάμεση τιμή ήταν 11,408 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 8,850–13,217 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 11 kWh, βρισκόταν εντός του ιστορικού εύρους και σχετικά κοντά στη μέση και τη διάμεση τιμή του πρόσφατου ιστορικού. Ωστόσο, η σταθερότητα του ωριαίου προφίλ ήταν πολύ χαμηλή, καθώς ο αντίστοιχος δείκτης ήταν μόλις 0,019. Αυτό δείχνει ότι, παρότι το ιστορικό ήταν αρκετά αντιπροσωπευτικό ως προς το συνολικό ημερήσιο επίπεδο, δεν ήταν εξίσου αντιπροσωπευτικό ως προς την κατανομή της κατανάλωσης μέσα στην ημέρα. Για τον λόγο αυτό, το ιστορικό κατατάχθηκε στην κατηγορία Medium και το σύστημα πρότεινε χαμηλή τιμή max_alpha, ίση με 0,10.



Σχήμα 47 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home05 στη χειμερινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 60, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	LGBM	295,27	592,98	58,45	7,177	11	3,823
with-history adaptive	XGB	315,75	612,24	61,97	7,989	11	3,012

Πίνακας 60 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home05 στη χειμερινή περίπτωση

Για τη χειμερινή περίπτωση της home05, το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση δεν οδηγεί σε βελτίωση της ωριαίας ακρίβειας. Το RMSE αυξάνεται από 592,98 Wh στο καλύτερο μοντέλο χωρίς ιστορικό σε 612,24 Wh στο καλύτερο μοντέλο με χρήση ιστορικού. Παράλληλα, το ημερήσιο σφάλμα μειώνεται από 3,823 kWh σε 3,012 kWh όταν η σύγκριση γίνεται με βάση τα καλύτερα μοντέλα ως προς το RMSE. Ωστόσο, η βελτίωση στο ημερήσιο σφάλμα είναι περιορισμένη, καθώς το καλύτερο ημερήσιο σφάλμα του σεναρίου χωρίς ιστορικά δεδομένα είναι ήδη 3,040 kWh με το Random Forest.

Η περίπτωση αυτή δείχνει ότι η εγγύτητα του ιστορικού ως προς το συνολικό ημερήσιο επίπεδο κατανάλωσης δεν αρκεί από μόνη της για τη βελτίωση της πρόβλεψης. Παρότι η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν κοντά στη μέση και τη διάμεση τιμή του πρόσφατου ιστορικού, το ωριαίο προφίλ του ιστορικού ήταν ασταθές. Επομένως, η προσαρμοστική διόρθωση δεν μπόρεσε να βελτιώσει τη μορφή της ωριαίας καμπύλης. Έτσι επιβεβαιώνεται ότι η αξιοποίηση ιστορικών δεδομένων είναι πιο αποτελεσματική όταν το ιστορικό είναι αντιπροσωπευτικό όχι μόνο ως προς τη συνολική ημερήσια κατανάλωση, αλλά και ως προς την ωριαία κατανομή και τις αιχμές της ημέρας στόχου.

16.3.4 Εαρινή αξιολόγηση για την κατοικία home07

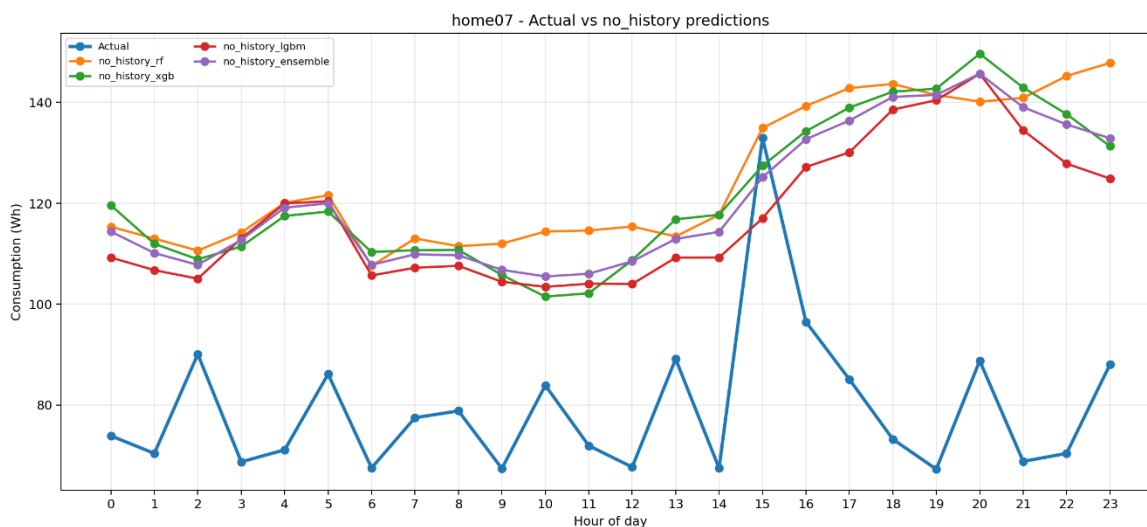
Η τέταρτη εποχική μελέτη περίπτωσης αφορά την κατοικία home07, με ημέρα αξιολόγησης την Τρίτη 2 Μαΐου 2023. Η ημερομηνία αυτή αντιστοιχεί σε εαρινή περίοδο και επιλέχθηκε ώστε να εξεταστεί η συμπεριφορά του συστήματος σε διαφορετική χρονική περίοδο από την κύρια αξιολόγηση της ίδιας κατοικίας. Η κατοικία είναι διαμέρισμα, κατασκευασμένο την περίοδο 1970–1990, με έναν ένοικο και τέσσερα δωμάτια. Διαθέτει πλυντήριο ρούχων, πλυντήριο πιάτων και φούρνο μικροκυμάτων, ενώ δεν διαθέτει στεγνωτήριο. Η θέρμανση βασίζεται σε αντλία θερμότητας και η παραγωγή ζεστού νερού γίνεται μέσω ηλεκτρικού θερμοσίφωνα. Η πραγματική συνολική ημερήσια κατανάλωση της συγκεκριμένης ημέρας ήταν 1,902 kWh.

Στον Πίνακα 61 παρουσιάζονται τα αποτελέσματα για το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	45,32	48,46	60,55	2,990	1,902	1,088
XGBoost	42,81	46,18	57,22	2,919	1,902	1,016
LightGBM	39,35	42,21	52,46	2,815	1,902	0,912
Ensemble	42,01	45,04	56,12	2,895	1,902	0,993

Πίνακας 61 Απόδοση μοντέλων για την κατοικία home07 στην εαρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων, το LightGBM παρουσιάζει την καλύτερη συνολική απόδοση, επιτυγχάνοντας τα χαμηλότερα MAE, RMSE και MAPE. Συγκεκριμένα, το RMSE του LightGBM είναι 42,21 Wh, ενώ το ημερήσιο σφάλμα είναι 0,912 kWh. Παρότι το LightGBM υπερέχει έναντι των υπόλοιπων μοντέλων, όλα τα μοντέλα υπερεκτιμούν τη συνολική ημερήσια κατανάλωση. Η πραγματική κατανάλωση της ημέρας ήταν 1,902 kWh, ενώ οι προβλέψεις κυμαίνονται από 2,815 έως 2,990 kWh.



Σχήμα 48 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home07 στην εαρινή περίπτωση χωρίς χρήση ιστορικών δεδομένων

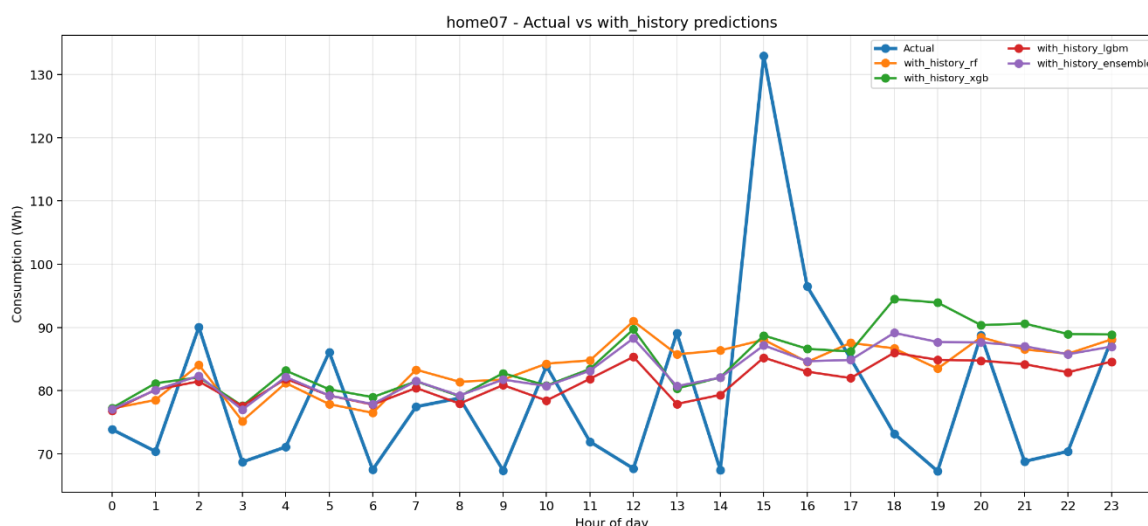
Στη συνέχεια εξετάστηκε το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και προσαρμοστική διόρθωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 62.

Μοντέλο	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
RF	10,64	14,28	13,57	2,008	1,902	0,106
XGBoost	11,92	15,54	15,37	2,029	1,902	0,127
LightGBM	10,90	14,13	13,58	1,955	1,902	0,052
Ensemble	11,00	14,52	13,98	1,994	1,902	0,092

Πίνακας 62 Απόδοση μοντέλων για την κατοικία home07 στην εαρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Στο σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρατηρείται σημαντική βελτίωση σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Το LightGBM παρουσιάζει το χαμηλότερο RMSE, ίσο με 14,13 Wh, καθώς και το μικρότερο ημερήσιο σφάλμα, ίσο με 0,052 kWh. Το Random Forest εμφανίζει οριακά χαμηλότερο MAE, ενώ οι διαφορές μεταξύ των μοντέλων με χρήση ιστορικού είναι σχετικά μικρές. Συνολικά, όλα τα μοντέλα του σεναρίου με ιστορικό προσεγγίζουν πολύ καλύτερα τόσο την ωριαία καμπύλη όσο και τη συνολική ημερήσια κατανάλωση.

Η ανάλυση σταθερότητας έδειξε ότι το διαθέσιμο ιστορικό ήταν πλήρες. Η μέση ημερήσια κατανάλωση του ιστορικού ήταν 2,042 kWh, η διάμεση τιμή ήταν 1,906 kWh, ενώ το εύρος των ημερήσιων τιμών ήταν 1,836–2,595 kWh. Η πραγματική κατανάλωση της ημέρας αξιολόγησης, ίση με 1,902 kWh, βρισκόταν εντός του ιστορικού εύρους και ήταν σχεδόν ταυτόσημη με τη διάμεση τιμή του πρόσφατου ιστορικού. Επιπλέον, η σταθερότητα του ωριαίου προφίλ ήταν υψηλή, ενώ η ημερήσια σταθερότητα ήταν επίσης ικανοποιητική. Για τον λόγο αυτό, το ιστορικό κατατάχθηκε στην κατηγορία High και το σύστημα πρότεινε τιμή max_alpha ίση με 0,25.



Σχήμα 49 Σύγκριση πραγματικής και προβλεπόμενης κατανάλωσης για την κατοικία home07 στην εαρινή περίπτωση με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση

Η συνολική σύγκριση των δύο σεναρίων παρουσιάζεται στον Πίνακα 63, λαμβάνοντας υπόψη το καλύτερο μοντέλο κάθε σεναρίου με κύριο κριτήριο τη μετρική RMSE.

Σενάριο	Best Model	MAE (Wh)	RMSE (Wh)	MAPE (%)	pred_daily (kWh)	real_daily (kWh)	Ημερήσιο σφάλμα (kWh)
no-history	LGBM	39,35	42,21	52,46	2,815	1,902	0,912
with-history adaptive	LGBM	10,90	14,13	13,58	1,955	1,902	0,052

Πίνακας 63 Συγκριτική αξιολόγηση καλύτερων μοντέλων για την κατοικία home07 στην εαρινή περίπτωση

Για την εαρινή περίπτωση της home07, η χρήση ιστορικών δεδομένων και προσαρμοστικής διόρθωσης οδηγεί σε σαφή βελτίωση της πρόβλεψης. Το RMSE μειώνεται από 42,21 Wh στο καλύτερο μοντέλο χωρίς ιστορικό σε 14,13 Wh στο καλύτερο μοντέλο με χρήση ιστορικού. Παράλληλα, το ημερήσιο σφάλμα μειώνεται από 0,912 kWh σε 0,052 kWh, γεγονός που δείχνει ότι το πρόσφατο ιστορικό συνέβαλε τόσο στη βελτίωση της ωριαίας ακρίβειας όσο και στην ακριβέστερη εκτίμηση του ημερήσιου επιπέδου κατανάλωσης. Και αυτή η περίπτωση επιβεβαιώνει ότι το ιστορικό μπορεί να είναι ιδιαίτερα χρήσιμο όταν είναι πλήρες, σχετικά σταθερό και αντιπροσωπευτικό της ημέρας στόχου. Σε αντίθεση με τη χειμερινή περίπτωση της home05, όπου το ημερήσιο επίπεδο του ιστορικού ήταν κοντά στην ημέρα στόχο αλλά το ωριαίο προφίλ ήταν ασταθές, στην home07 το πρόσφατο ιστορικό ήταν αντιπροσωπευτικό τόσο ως προς το συνολικό επίπεδο κατανάλωσης όσο και ως προς την ωριαία κατανομή. Για τον λόγο αυτό, η προσαρμοστική διόρθωση βελτίωσε ουσιαστικά την πρόβλεψη χωρίς να οδηγήσει σε υπερβολική προσαρμογή.

16.4 Συγκεντρωτική αποτίμηση κύριας και εποχικής αξιολόγησης

Μετά την αναλυτική παρουσίαση της κύριας αξιολόγησης και των εποχικών μελετών περίπτωσης, πραγματοποιείται συνολική αποτίμηση των αποτελεσμάτων για το σύνολο δεδομένων PLEGMA. Στόχος της ενότητας αυτής είναι να συγκεντρωθούν τα βασικά ευρήματα και να εξεταστεί συνολικά η συμπεριφορά του συστήματος στα δύο βασικά σενάρια πρόβλεψης.

Στον Πίνακα 64 παρουσιάζονται συγκεντρωτικά τα καλύτερα αποτελέσματα της κύριας αξιολόγησης για τις τέσσερις επιλεγμένες κατοικίες του συνόλου ελέγχου. Η επιλογή του καλύτερου σεναρίου έγινε με κύριο κριτήριο τη μετρική RMSE, καθώς βασικός στόχος της αξιολόγησης είναι η ακρίβεια της ωριαίας καμπύλης πρόβλεψης.

Κατοικία	Ημερομηνία	Κατηγορία ιστορικού	Καλύτερο σενάριο	Καλύτερο μοντέλο	RMSE (Wh)	Ημερήσιο σφάλμα (kWh)
home01	20/09/2023	Low	with-history adaptive	LightGBM	28,95	0,527
home04	05/09/2023	Low	with-history adaptive	XGBoost	47,87	0,453
home05	04/09/2023	Medium	no-history	LightGBM	429,50	2,086
home07	22/09/2023	Medium	with-history adaptive	XGBoost	34,95	0,056

Πίνακας 64 Συγκεντρωτική σύγκριση καλύτερων σεναρίων στην κύρια αξιολόγηση του PLEGMA

Από την κύρια αξιολόγηση προκύπτει ότι το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρουσίασε καλύτερη ωριαία απόδοση σε τρεις από τις τέσσερις κατοικίες. Συγκεκριμένα, στις κατοικίες home01, home04 και home07, η χρήση ιστορικού οδήγησε σε μείωση του RMSE σε σχέση με το σενάριο χωρίς ιστορικά δεδομένα. Η βελτίωση ήταν ιδιαίτερα εμφανής στην home07, όπου η ημέρα στόχος βρισκόταν εντός του ιστορικού εύρους και κοντά στη διάμεση ημερήσια τιμή του πρόσφατου ιστορικού. Στην περίπτωση αυτή, το ημερήσιο σφάλμα μειώθηκε σε μόλις 0,056 kWh.

Ωστόσο, η κύρια αξιολόγηση ανέδειξε και τους περιορισμούς της προσέγγισης. Στην κατοικία home05, το σενάριο χωρίς χρήση ιστορικών δεδομένων παρέμεινε καλύτερο ως προς την ωριαία ακρίβεια, καθώς το LightGBM πέτυχε χαμηλότερο RMSE σε σχέση με τα μοντέλα του σεναρίου με ιστορικό. Παρ' όλα αυτά, η χρήση ιστορικού βελτίωσε σημαντικά την εκτίμηση της συνολικής ημερήσιας κατανάλωσης, μειώνοντας το ημερήσιο σφάλμα. Το αποτέλεσμα αυτό δείχνει ότι η βελτίωση του ημερήσιου επιπέδου κατανάλωσης δεν συνεπάγεται απαραίτητα και βελτίωση της μορφής της ωριαίας καμπύλης.

Στον Πίνακα 65 παρουσιάζονται τα αντίστοιχα συγκεντρωτικά αποτελέσματα των εποχικών μελετών περίπτωσης. Και εδώ η επιλογή του καλύτερου σεναρίου έγινε με κύριο κριτήριο τη μετρική RMSE.

Εποχική περίπτωση	Κατηγορία ιστορικού	Καλύτερο σενάριο	Καλύτερο μοντέλο	RMSE (Wh)	Ημερήσιο σφάλμα (kWh)
home01 – Καλοκαίρι	Excellent	with-history adaptive	Random Forest	0,49	0,007
home04 – Φθινόπωρο	High	no-history	Random Forest	44,13	0,179
home05 – Χειμώνας	Medium	no-history	LightGBM	592,98	3,823
home07 – Άνοιξη	High	with-history adaptive	LightGBM	14,13	0,052

Πίνακας 65 Συγκεντρωτική σύγκριση καλύτερων σεναρίων στις εποχικές μελέτες περίπτωσης του PLEGMA

Από τις εποχικές μελέτες περίπτωσης προκύπτει πιο σύνθετη εικόνα. Το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση παρουσίασε την καλύτερη ωριαία απόδοση σε δύο από τις τέσσερις περιπτώσεις, δηλαδή στη θερινή περίπτωση της κατοικίας home01 και στην εαρινή περίπτωση της κατοικίας home07. Αντίθετα, στη φθινοπωρινή περίπτωση της κατοικίας home04 και στη χειμερινή περίπτωση της κατοικίας home05, το σενάριο χωρίς χρήση ιστορικών δεδομένων παρέμεινε καλύτερο ως προς τη μετρική RMSE. Η θερινή περίπτωση της κατοικίας home01 αποτελεί το πιο χαρακτηριστικό παράδειγμα επιτυχούς αξιοποίησης ιστορικού. Το πρόσφατο ιστορικό ήταν πλήρες, εξαιρετικά σταθερό και σχεδόν ταυτόσημο με την ημέρα στόχο ως προς το ημερήσιο επίπεδο κατανάλωσης. Για τον λόγο αυτό, η προσαρμοστική διόρθωση οδήγησε σε πολύ μεγάλη μείωση τόσο του RMSE όσο και του ημερήσιου σφάλματος. Αντίστοιχα, στην εαρινή περίπτωση της κατοικίας home07, το ιστορικό ήταν πλήρες, αρκετά σταθερό και αντιπροσωπευτικό της ημέρας στόχου, με αποτέλεσμα το σενάριο με χρήση ιστορικών δεδομένων να βελτιώσει ουσιαστικά την πρόβλεψη.

Αντίθετα, η φθινοπωρινή περίπτωση της κατοικίας home04 δείχνει ότι ακόμη και ένα ιστορικό που κατατάσσεται σε υψηλή κατηγορία δεν εγγυάται απαραίτητα καλύτερη ωριαία πρόβλεψη. Στην περίπτωση αυτή, το σενάριο χωρίς ιστορικά δεδομένα και ειδικότερα το Random Forest, είχε ήδη αποτυπώσει αρκετά καλά την ωριαία καμπύλη της ημέρας στόχου. Η χρήση ιστορικού οδήγησε σε οριακά μικρότερο ημερήσιο σφάλμα για το αντίστοιχο μοντέλο Random Forest, αλλά δεν βελτίωσε το RMSE. Το αποτέλεσμα αυτό δείχνει ότι η σταθερότητα του ιστορικού αποτελεί σημαντική αλλά όχι επαρκή συνθήκη για τη βελτίωση της πρόβλεψης. Εξίσου σημαντικό είναι το ιστορικό να είναι αντιπροσωπευτικό της πραγματικής ωριαίας κατανομής της ημέρας στόχου.

Η χειμερινή περίπτωση της κατοικίας home05 αποτελεί τη δυσκολότερη εποχική περίπτωση. Παρότι η πραγματική ημερήσια κατανάλωση της ημέρας στόχου βρισκόταν κοντά στη μέση και στη διάμεση τιμή του πρόσφατου ιστορικού, το ωριαίο προφίλ του ιστορικού ήταν ιδιαίτερα ασταθές. Ως αποτέλεσμα, η χρήση ιστορικών δεδομένων δεν βελτίωσε την ωριαία ακρίβεια, ενώ η βελτίωση του ημερήσιου σφάλματος ήταν περιορισμένη. Η home05 επιβεβαιώνει ότι η αντιπροσωπευτικότητα του ημερήσιου επιπέδου δεν αρκεί από μόνη της, αν το ωριαίο σχήμα της κατανάλωσης δεν παρουσιάζει επαναληψιμότητα.

Συγκρίνοντας την κύρια αξιολόγηση με τις εποχικές μελέτες περίπτωσης, προκύπτει ότι η αποτελεσματικότητα του ιστορικού διαφοροποιείται όχι μόνο ανά κατοικία, αλλά και ανά χρονική περίοδο. Η ίδια κατοικία εμφανίζει διαφορετική συμπεριφορά σε διαφορετικές εποχές, ανάλογα με τη σταθερότητα των προηγούμενων ημερών και το κατά πόσο η ημέρα στόχος ακολουθεί το πρόσφατο ιστορικό πρότυπο. Για παράδειγμα, η κατοικία home01 στην κύρια αξιολόγηση είχε ιστορικό χαμηλής σταθερότητας και η χρήση ιστορικού βελτίωσε μεν την πρόβλεψη, αλλά η τελική πρόβλεψη παρέμεινε υπερεκτιμημένη. Αντίθετα, στη θερινή εποχική περίπτωση της ίδιας κατοικίας, το ιστορικό ήταν εξαιρετικά σταθερό και οδήγησε σε σχεδόν τέλεια πρόβλεψη.

Αντίστοιχα, η κατοικία home07 παρουσίασε θετική συμπεριφορά τόσο στην κύρια όσο και στην εαρινή αξιολόγηση, επειδή και στις δύο περιπτώσεις η ημέρα στόχος βρισκόταν εντός του ιστορικού εύρους και κοντά στο πρόσφατο επίπεδο κατανάλωσης. Αντίθετα, η κατοικία home05 παρέμεινε δύσκολη περίπτωση και στις δύο αξιολογήσεις, γεγονός που συνδέεται με το υψηλότερο επίπεδο κατανάλωσης, τον τύπο κατοικίας και την αστάθεια της ωριαίας συμπεριφοράς.

Συνολικά, τα αποτελέσματα του PLEGMA επιβεβαιώνουν ότι η χρήση πρόσφατου ιστορικού κατανάλωσης μπορεί να βελτιώσει την πρόβλεψη, αλλά η επίδρασή της δεν είναι

ομοιόμορφη. Η βελτίωση είναι πιο έντονη όταν το ιστορικό είναι πλήρες, σταθερό και αντιπροσωπευτικό τόσο ως προς το συνολικό ημερήσιο επίπεδο όσο και ως προς την ωριαία κατανομή της κατανάλωσης. Όταν κάποιο από αυτά τα στοιχεία απουσιάζει, η προσαρμοστική διόρθωση μπορεί να έχει περιορισμένο όφελος ή να βελτιώνει μόνο ορισμένες μετρικές, όπως το ημερήσιο σφάλμα, χωρίς να βελτιώνει απαραίτητα την ωριαία ακρίβεια.

Η εικόνα αυτή πρέπει να ερμηνευθεί λαμβάνοντας υπόψη και το μικρό μέγεθος του συνόλου δεδομένων PLEGMA. Σε αντίθεση με μεγαλύτερα σύνολα δεδομένων, το PLEGMA περιλαμβάνει περιορισμένο αριθμό κατοικιών και μικρότερη ποικιλία προτύπων κατανάλωσης. Ως εκ τούτου, τα μοντέλα έχουν λιγότερα παραδείγματα για να μάθουν γενικεύσιμα μοτίβα συμπεριφοράς, γεγονός που καθιστά την απόδοσή τους πιο ευαίσθητη στις ιδιαιτερότητες κάθε κατοικίας και κάθε ημέρας αξιολόγησης. Ο περιορισμός αυτός εξηγεί σε σημαντικό βαθμό γιατί τα αποτελέσματα παρουσιάζουν μεγαλύτερη μεταβλητότητα και γιατί η χρήση ιστορικού δεν οδηγεί πάντα σε σταθερή βελτίωση όλων των μετρικών.

Επομένως, το βασικό συμπέρασμα της συγκεντρωτικής αποτίμησης είναι ότι το σύστημα πρόβλεψης μπορεί να εφαρμοστεί και στο PLEGMA, παρέχοντας χρήσιμες και ερμηνεύσιμες προβλέψεις σε επίπεδο κατοικίας. Ωστόσο, λόγω του μικρότερου μεγέθους του dataset και της μεγαλύτερης ευαισθησίας των αποτελεσμάτων ανά κατοικία, η αξιολόγηση πρέπει να ερμηνεύεται προσεκτικά. Η ανάλυση του πρόσφατου ιστορικού, ο περιορισμός μέσω του $\max_alpha$ και η ταυτόχρονη εξέταση πολλαπλών μετρικών αποτελούν απαραίτητα στοιχεία για την ορθή αποτίμηση της τελικής πρόβλεψης.

16.5 Συμπεράσματα κεφαλαίου

Στο παρόν κεφάλαιο αξιολογήθηκε η πρακτική λειτουργία του συστήματος πρόβλεψης στο σύνολο δεδομένων PLEGMA, σε επίπεδο μεμονωμένης κατοικίας. Η αξιολόγηση πραγματοποιήθηκε τόσο σε επιλεγμένες κατοικίες του συνόλου ελέγχου όσο και σε πρόσθετες εποχικές μελέτες περίπτωσης, με στόχο να εξεταστεί η συμπεριφορά της εφαρμογής σε διαφορετικές χρονικές περιόδους και σε διαφορετικά πρότυπα κατανάλωσης. Τα αποτελέσματα επιβεβαίωσαν ότι το σύστημα μπορεί να εφαρμοστεί και στο PLEGMA, παρέχοντας ερμηνεύσιμες προβλέψεις και χρήσιμη σύγκριση μεταξύ διαφορετικών μοντέλων και σεναρίων. Το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης προσφέρει μια πρακτική λύση σε περιπτώσεις όπου δεν υπάρχει διαθέσιμο ενεργειακό ιστορικό ή όπου το διαθέσιμο ιστορικό δεν είναι αρκετά αξιόπιστο. Αντίστοιχα, το σενάριο με χρήση ιστορικών δεδομένων και προσαρμοστική διόρθωση μπορεί να βελτιώσει σημαντικά την πρόβλεψη, όταν το πρόσφατο ιστορικό είναι ποιοτικό και συμβατό με τη συμπεριφορά της ημέρας στόχου.

Παράλληλα, η αξιολόγηση ανέδειξε ότι η πρόβλεψη ηλεκτρικής κατανάλωσης σε επίπεδο μεμονωμένης κατοικίας παραμένει σύνθετο πρόβλημα. Η κατανάλωση επηρεάζεται από παράγοντες που δεν είναι πάντα διαθέσιμοι στα δεδομένα, όπως η παρουσία των ενοίκων, η χρήση συγκεκριμένων συσκευών και οι έκτακτες αλλαγές στη συμπεριφορά του νοικοκυριού. Για τον λόγο αυτό, η απόδοση των μοντέλων μπορεί να διαφοροποιείται σημαντικά ανά κατοικία και ανά ημέρα αξιολόγησης.

Ιδιαίτερη σημασία έχει και το μέγεθος του συνόλου δεδομένων. Το PLEGMA είναι σαφώς μικρότερο από το IDEAL, τόσο ως προς τον αριθμό των κατοικιών όσο και ως προς την ποικιλία των διαθέσιμων προτύπων κατανάλωσης. Επομένως, τα μοντέλα διαθέτουν

λιγότερα παραδείγματα εκπαίδευσης για να μάθουν γενικεύσιμες σχέσεις και είναι αναμενόμενο η απόδοσή τους να παρουσιάζει μεγαλύτερη μεταβλητότητα. Ο περιορισμός αυτός δεν ακυρώνει τα αποτελέσματα, αλλά πρέπει να λαμβάνεται υπόψη κατά την ερμηνεία τους.

Η αξιολόγηση στο PLEGMA λειτουργεί ως εφαρμογή και επικύρωση της προτεινόμενης μεθοδολογίας σε ελληνικά δεδομένα. Τα αποτελέσματα δείχνουν ότι η μεθοδολογία είναι λειτουργική και μπορεί να υποστηρίξει προβλέψεις επόμενης ημέρας μέσω API και γραφικής διεπαφής. Ωστόσο, λόγω του περιορισμένου μεγέθους του dataset και της έντονης διαφοροποίησης μεταξύ κατοικιών, τα αποτελέσματα πρέπει να ερμηνεύονται ως ενδεικτική αξιολόγηση της πρακτικής λειτουργίας του συστήματος και όχι ως πλήρης γενίκευση σε όλες τις πιθανές κατοικίες ή συνθήκες κατανάλωσης.

17. Συμπεράσματα και μελλοντικές επεκτάσεις

Στο παρόν κεφάλαιο παρουσιάζονται τα συνολικά συμπεράσματα της διπλωματικής εργασίας και πραγματοποιείται η τελική αποτίμηση της μεθοδολογίας που αναπτύχθηκε για την πρόβλεψη της ηλεκτρικής κατανάλωσης κατοικιών. Η ανάλυση που προηγήθηκε βασίστηκε στην εφαρμογή μοντέλων μηχανικής μάθησης στα σύνολα δεδομένων IDEAL και PLEGMA, καθώς και στην αξιολόγηση της πρακτικής λειτουργίας ενός συστήματος πρόβλεψης μέσω API και γραφικής διεπαφής χρήστη.

Στις ενότητες που ακολουθούν συνοψίζονται αρχικά τα βασικά στάδια και ευρήματα της εργασίας. Στη συνέχεια πραγματοποιείται συγκριτική αποτίμηση των αποτελεσμάτων μεταξύ IDEAL και PLEGMA, διατυπώνονται τα κύρια συμπεράσματα, αναφέρονται οι βασικοί περιορισμοί της μελέτης και παρουσιάζονται πιθανές κατευθύνσεις για μελλοντική έρευνα και επέκταση του συστήματος.

17.1 Συνοπτική ανακεφαλαίωση της εργασίας

Η παρούσα διπλωματική εργασία είχε ως αντικείμενο την πρόβλεψη της ηλεκτρικής κατανάλωσης κατοικιών με χρήση τεχνικών μηχανικής μάθησης. Η προσέγγιση που ακολουθήθηκε δεν περιορίστηκε μόνο στην εκπαίδευση και σύγκριση μοντέλων, αλλά επεκτάθηκε στην ανάπτυξη μιας ολοκληρωμένης μεθοδολογίας, η οποία περιλαμβάνει την επεξεργασία πραγματικών δεδομένων, τη δημιουργία κατάλληλων χαρακτηριστικών, την αξιολόγηση διαφορετικών σεναρίων πρόβλεψης και την υλοποίηση λειτουργικού συστήματος πρόβλεψης μέσω API και γραφικής διεπαφής χρήστη.

Η μεθοδολογία εφαρμόστηκε αρχικά στο σύνολο δεδομένων IDEAL, το οποίο αποτέλεσε την κύρια βάση ανάπτυξης των μοντέλων και της εφαρμογής. Το IDEAL προσέφερε μεγάλο αριθμό κατοικιών, εκτεταμένο χρονικό εύρος μετρήσεων και πλούσια ποικιλία προτύπων κατανάλωσης. Για τον λόγο αυτό, χρησιμοποιήθηκε ως κύριο πεδίο εκπαίδευσης, αξιολόγησης και βελτίωσης της μεθοδολογίας. Στη συνέχεια, η ίδια γενική λογική προσαρμόστηκε στο σύνολο δεδομένων PLEGMA, με στόχο να εξεταστεί η δυνατότητα εφαρμογής του συστήματος σε ελληνικά δεδομένα και σε διαφορετικό ενεργειακό και κλιματικό περιβάλλον.

Στο πλαίσιο της εργασίας εξετάστηκαν δύο βασικά σενάρια πρόβλεψης. Το πρώτο ήταν το σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, στο οποίο τα μοντέλα βασίζονται σε χρονικά, περιβαλλοντικά και στατικά χαρακτηριστικά της κατοικίας και του νοικοκυριού. Το σενάριο αυτό είναι ιδιαίτερα σημαντικό σε περιπτώσεις όπου δεν υπάρχει διαθέσιμο προηγούμενο ιστορικό κατανάλωσης για μια κατοικία ή όπου το διαθέσιμο ιστορικό δεν είναι επαρκές. Το δεύτερο ήταν το σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης, στο οποίο αξιοποιούνται πρόσφατες μετρήσεις της ίδιας κατοικίας, ώστε η πρόβλεψη να προσαρμόζεται καλύτερα στη συγκεκριμένη ενεργειακή συμπεριφορά.

Για την εκπαίδευση των μοντέλων χρησιμοποιήθηκαν τρεις αλγόριθμοι μηχανικής μάθησης βασισμένοι σε δέντρα αποφάσεων: Random Forest, XGBoost και LightGBM. Τα μοντέλα αυτά επιλέχθηκαν επειδή είναι κατάλληλα για δομημένα δεδομένα, μπορούν να διαχειριστούν μη γραμμικές σχέσεις και επιτρέπουν βασική ερμηνεία μέσω της ανάλυσης σημαντικότητας χαρακτηριστικών. Παράλληλα, εξετάστηκε και η συνδυαστική πρόβλεψη μέσω ensemble, ώστε να αξιοποιηθεί η συμπληρωματική συμπεριφορά των επιμέρους μοντέλων.

Η αξιολόγηση πραγματοποιήθηκε σε δύο επίπεδα. Αρχικά, εξετάστηκε η συνολική απόδοση των μοντέλων στα διαθέσιμα σύνολα εκπαίδευσης και ελέγχου, με χρήση μετρικών όπως MAE, RMSE και MAPE. Στη συνέχεια, η αξιολόγηση επεκτάθηκε σε επίπεδο πρακτικής εφαρμογής, όπου το σύστημα παρήγαγε προβλέψεις επόμενης ημέρας για συγκεκριμένες κατοικίες και συγκεκριμένες ημερομηνίες. Με αυτόν τον τρόπο αξιολογήθηκε όχι μόνο η αριθμητική ακρίβεια των μοντέλων, αλλά και η ικανότητα του συστήματος να αποτυπώνει την ωριαία καμπύλη και τη συνολική ημερήσια κατανάλωση. Ιδιαίτερη συμβολή της εργασίας αποτελεί η ενσωμάτωση μηχανισμού προσαρμοστικής διόρθωσης βάσει πρόσφατου ιστορικού κατανάλωσης. Ο μηχανισμός αυτός επιτρέπει στο σύστημα να αξιοποιεί το ιστορικό μιας κατοικίας όταν αυτό είναι επαρκές και αξιόπιστο, αλλά ταυτόχρονα να περιορίζει την επίδρασή του όταν το ιστορικό παρουσιάζει αστάθεια ή δεν είναι αντιπροσωπευτικό της ημέρας στόχου. Η προσέγγιση αυτή ενισχύει την πρακτική αξία του συστήματος, καθώς δεν αντιμετωπίζει το ιστορικό ως δεδομένα χρήσιμη πληροφορία σε κάθε περίπτωση, αλλά το αξιολογεί πριν το ενσωματώσει στην τελική πρόβλεψη.

Η ανάπτυξη API και γραφικής διεπαφής χρήστη μετέτρεψε τη μεθοδολογία από απλή πειραματική διαδικασία σε λειτουργικό πρωτότυπο σύστημα πρόβλεψης. Μέσω της εφαρμογής, ο χρήστης μπορεί να εισάγει χαρακτηριστικά κατοικίας, να επιλέγει σενάριο πρόβλεψης, να αξιοποιεί διαθέσιμο ιστορικό κατανάλωσης και να εξετάζει την προβλεπόμενη ωριαία καμπύλη. Με αυτόν τον τρόπο, η εργασία συνδέει τη θεωρητική ανάπτυξη μοντέλων με ένα πιο πρακτικό περιβάλλον χρήσης.

Συνολικά, η εργασία ανέδειξε ότι τα μοντέλα μηχανικής μάθησης μπορούν να χρησιμοποιηθούν αποτελεσματικά για την πρόβλεψη ηλεκτρικής κατανάλωσης κατοικιών, αλλά η απόδοσή τους εξαρτάται σημαντικά από το διαθέσιμο επίπεδο πληροφορίας, την ποιότητα των δεδομένων, τη σταθερότητα της κατανάλωσης και το μέγεθος του συνόλου δεδομένων. Η σύγκριση των σεναρίων χωρίς και με χρήση ιστορικών δεδομένων έδειξε ότι το ιστορικό μπορεί να προσφέρει σημαντική βελτίωση, όταν είναι πλήρες, σταθερό και αντιπροσωπευτικό. Παράλληλα, όμως, φάνηκε ότι η πρόβλεψη σε επίπεδο μεμονωμένης κατοικίας παραμένει απαιτητικό πρόβλημα, καθώς επηρεάζεται από συμπεριφορικούς και τυχαίους παράγοντες που δεν αποτυπώνονται πάντα στα διαθέσιμα χαρακτηριστικά.

17.2 Συγκριτική αποτίμηση IDEAL και PLEGMA

Η εφαρμογή της μεθοδολογίας στα δύο σύνολα δεδομένων, IDEAL και PLEGMA, επέτρεψε την αξιολόγηση της ίδιας βασικής προσέγγισης σε δύο διαφορετικά πλαίσια. Και στις δύο περιπτώσεις ακολουθήθηκε κοινή μεθοδολογική λογική, η οποία περιλάμβανε προεπεξεργασία των δεδομένων, δημιουργία χρονικών, περιβαλλοντικών, στατικών και ιστορικών χαρακτηριστικών, εκπαίδευση των μοντέλων Random Forest, XGBoost και LightGBM, καθώς και αξιολόγηση σε σενάρια χωρίς και με χρήση ιστορικών δεδομένων κατανάλωσης. Ωστόσο, τα δύο σύνολα δεδομένων διαφέρουν σημαντικά ως προς το μέγεθος, τη δομή και την ποικιλία των διαθέσιμων δεδομένων, γεγονός που επηρέασε την τελική απόδοση των μοντέλων.

Το IDEAL αποτέλεσε το κύριο σύνολο ανάπτυξης και αξιολόγησης της μεθοδολογίας. Περιλαμβάνει μεγαλύτερο αριθμό κατοικιών, μεγαλύτερο χρονικό εύρος μετρήσεων και μεγαλύτερη ποικιλία προτύπων κατανάλωσης. Το γεγονός αυτό επέτρεψε στα μοντέλα να εκπαιδευτούν σε περισσότερες περιπτώσεις και να μάθουν πιο γενικεύσιμες σχέσεις ανάμεσα στα χαρακτηριστικά εισόδου και στην ηλεκτρική κατανάλωση. Ως αποτέλεσμα, η

συμπεριφορά των μοντέλων στο IDEAL ήταν γενικά πιο σταθερή, ενώ η χρήση ιστορικών δεδομένων και προσαρμοστικής διόρθωσης οδήγησε συχνότερα σε σαφή βελτίωση της πρόβλεψης, όταν το πρόσφατο ιστορικό ήταν σταθερό και αντιπροσωπευτικό.

Αντίθετα, το PLEGMA αποτέλεσε ένα μικρότερο και πιο περιορισμένο σύνολο δεδομένων. Περιλαμβάνει μικρότερο αριθμό κατοικιών και μικρότερη ποικιλία προτύπων κατανάλωσης, γεγονός που περιορίζει τη δυνατότητα των μοντέλων να μάθουν γενικεύσιμα μοτίβα συμπεριφοράς. Για τον λόγο αυτό, τα αποτελέσματα στο PLEGMA παρουσίασαν μεγαλύτερη ευαισθησία ανά κατοικία και ανά ημέρα αξιολόγησης. Η απόδοση των μοντέλων εξαρτήθηκε περισσότερο από τα ιδιαίτερα χαρακτηριστικά κάθε κατοικίας, τη σταθερότητα του διαθέσιμου ιστορικού και το κατά πόσο η ημέρα στόχος ακολουθούσε το πρόσφατο πρότυπο κατανάλωσης.

Στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης, τα δύο σύνολα δεδομένων ανέδειξαν διαφορετικές πτυχές του προβλήματος. Στο IDEAL, το μεγαλύτερο πλήθος κατοικιών και η μεγαλύτερη ποικιλία στατικών και συμπεριφορικών χαρακτηριστικών επέτρεψαν στα μοντέλα να αξιοποιήσουν καλύτερα πληροφορίες όπως ο τύπος κατοικίας, η επιφάνεια, ο αριθμός ενοίκων και τα χρονικά χαρακτηριστικά. Στο PLEGMA, το αντίστοιχο σενάριο παρέμεινε χρήσιμο, ιδιαίτερα όταν δεν υπήρχε αξιόπιστο ιστορικό, όμως η περιορισμένη ποσότητα δεδομένων μείωσε τη δυνατότητα των μοντέλων να εκτιμήσουν με συνέπεια το επίπεδο κατανάλωσης κάθε κατοικίας.

Στο σενάριο με χρήση ιστορικών δεδομένων κατανάλωσης και τα δύο σύνολα δεδομένων επιβεβαίωσαν τη σημασία του πρόσφατου ιστορικού. Η χρήση ιστορικών χαρακτηριστικών και προσαρμοστικής διόρθωσης μπορεί να βελτιώσει ουσιαστικά την πρόβλεψη, όταν το ιστορικό είναι πλήρες, σταθερό και αντιπροσωπευτικό της ημέρας στόχου. Ωστόσο, η επίδραση αυτή ήταν πιο σταθερή στο IDEAL και πιο μεταβλητή στο PLEGMA. Στο IDEAL, η ύπαρξη περισσότερων κατοικιών και μεγαλύτερου όγκου δεδομένων επέτρεψε στα μοντέλα να αξιοποιήσουν αποτελεσματικότερα τα ιστορικά μοτίβα. Στο PLEGMA, η χρήση ιστορικού οδήγησε σε σημαντικές βελτιώσεις σε ορισμένες περιπτώσεις, αλλά σε άλλες είχε περιορισμένο όφελος ή βελτίωνε μόνο το ημερήσιο σφάλμα χωρίς να βελτιώνει απαραίτητα την ωριαία ακρίβεια.

Ένα κοινό εύρημα και στα δύο σύνολα δεδομένων είναι ότι η ποιότητα του ιστορικού δεν πρέπει να αξιολογείται μόνο με βάση την πληρότητα των διαθέσιμων μετρήσεων. Εξίσου σημαντική είναι η σταθερότητα της ημερήσιας κατανάλωσης, η επαναληψιμότητα του ωριαίου προφίλ και η αντιπροσωπευτικότητα του πρόσφατου ιστορικού ως προς την ημέρα πρόβλεψης. Η ανάλυση των μεμονωμένων κατοικιών έδειξε ότι ακόμη και για την ίδια κατοικία, η αποτελεσματικότητα του ιστορικού μπορεί να διαφοροποιείται ανάλογα με την εποχή και τη συμπεριφορά των προηγούμενων ημερών. Επομένως, η χρήση ιστορικών δεδομένων δεν πρέπει να θεωρείται αυτόματα πλεονέκτημα, αλλά πρέπει να συνοδεύεται από έλεγχο ποιότητας και από περιορισμό της επίδρασής της όταν το ιστορικό δεν είναι αξιόπιστο.

Η σύγκριση των δύο συνόλων δεδομένων δείχνει επίσης ότι η άμεση αριθμητική σύγκριση των μετρικών πρέπει να γίνεται με προσοχή. Τα δύο σύνολα δεδομένων διαφέρουν ως προς τον αριθμό κατοικιών, το χρονικό εύρος, τις διαθέσιμες μεταβλητές, τις κλιματικές συνθήκες και τη γενικότερη ενεργειακή συμπεριφορά των νοικοκυριών. Επομένως, η σύγκριση IDEAL και PLEGMA δεν έχει ως στόχο να δείξει ότι κάποιο από αυτά οδηγεί απόλυτα σε καλύτερα ή χειρότερα αποτελέσματα, αλλά να αναδείξει πώς η ίδια μεθοδολογία συμπεριφέρεται σε διαφορετικά περιβάλλοντα δεδομένων.

Το IDEAL ανέδειξε τη δυνατότητα της μεθοδολογίας να εκπαιδεύει πιο σταθερά και γενικεύσιμα μοντέλα, λόγω του μεγαλύτερου όγκου και της μεγαλύτερης ποικιλίας δεδομένων. Το PLEGMA, από την άλλη πλευρά, ανέδειξε την πρακτική δυνατότητα

προσαρμογής της μεθοδολογίας σε ελληνικά δεδομένα, αλλά και τους περιορισμούς που προκύπτουν όταν το σύνολο δεδομένων είναι μικρότερο και λιγότερο αντιπροσωπευτικό. Η συνδυαστική χρήση των δύο συνόλων δεδομένων ενισχύει την αξία της εργασίας, καθώς επιτρέπει τόσο την ανάπτυξη και αξιολόγηση της μεθοδολογίας σε ένα πλούσιο σύνολο δεδομένων όσο και την εφαρμογή της σε ένα μικρότερο ελληνικό σύνολο, όπου οι περιορισμοί είναι πιο εμφανείς.

Η βασική διαπίστωση που προκύπτει από τη σύγκριση IDEAL και PLEGMA είναι ότι η επιτυχία ενός συστήματος πρόβλεψης ηλεκτρικής κατανάλωσης δεν εξαρτάται αποκλειστικά από την επιλογή του αλγορίθμου. Εξαρτάται σε μεγάλο βαθμό από την ποσότητα και ποιότητα των διαθέσιμων δεδομένων, τη σταθερότητα της συμπεριφοράς κάθε κατοικίας και τη δυνατότητα αξιοποίησης αξιόπιστου πρόσφατου ιστορικού. Για τον λόγο αυτό, η προτεινόμενη μεθοδολογία πρέπει να ερμηνεύεται ως ένα ευέλικτο πλαίσιο πρόβλεψης, το οποίο μπορεί να προσαρμόζεται σε διαφορετικά σύνολα δεδομένων, αλλά απαιτεί προσεκτική αξιολόγηση των ιδιαίτερων χαρακτηριστικών κάθε περίπτωσης.

17.3 Κύρια συμπεράσματα και σύνδεση με τη βιβλιογραφία

Τα ευρήματα της παρούσας εργασίας ευθυγραμμίζονται σε μεγάλο βαθμό με τη διεθνή βιβλιογραφία σχετικά με την πρόβλεψη ηλεκτρικής κατανάλωσης σε οικιακά περιβάλλοντα. Η κατανάλωση ηλεκτρικής ενέργειας σε επίπεδο κατοικίας επηρεάζεται από συνδυασμό χρονικών, περιβαλλοντικών, κτηριακών και συμπεριφορικών παραγόντων. Το συμπέρασμα αυτό επιβεβαιώθηκε και στην παρούσα εργασία, καθώς χαρακτηριστικά όπως η ώρα της ημέρας, η ημέρα της εβδομάδας, οι περιβαλλοντικές συνθήκες, ο τύπος κατοικίας και τα στατικά χαρακτηριστικά του νοικοκυριού συνέβαλαν στη διαμόρφωση των προβλέψεων, ιδιαίτερα στο σενάριο χωρίς χρήση ιστορικών δεδομένων κατανάλωσης. Το εύρημα αυτό είναι σύμφωνο με προηγούμενες ανασκοπήσεις στον τομέα της ενεργειακής πρόβλεψης κτηρίων, οι οποίες τονίζουν τη σημασία των μετεωρολογικών, κτηριακών και συμπεριφορικών χαρακτηριστικών στην πρόβλεψη φορτίου (Ahmad et al., 2018; Hong & Fan, 2016).

Παράλληλα, η εργασία επιβεβαίωσε τον καθοριστικό ρόλο της χρονικής εξάρτησης της κατανάλωσης. Στο σενάριο με χρήση ιστορικών δεδομένων, οι προηγούμενες τιμές κατανάλωσης και οι κυλίσμενοι στατιστικοί δείκτες αποτέλεσαν βασική πηγή πληροφορίας για τα μοντέλα. Η παρατήρηση αυτή είναι συμβατή με τη βιβλιογραφία της βραχυπρόθεσμης πρόβλεψης φορτίου, όπου η πρόσφατη συμπεριφορά της κατανάλωσης θεωρείται ένας από τους σημαντικότερους προγνωστικούς παράγοντες. Ωστόσο, η παρούσα εργασία έδειξε ότι η χρήση ιστορικού δεν είναι πάντοτε εξίσου αποτελεσματική. Η συμβολή του εξαρτάται από το αν το πρόσφατο ιστορικό είναι πλήρες, σταθερό και αντιπροσωπευτικό της ημέρας στόχου. Επομένως, η ύπαρξη ιστορικών δεδομένων δεν αρκεί από μόνη της· απαιτείται και αξιολόγηση της ποιότητας και της καταλληλότητάς τους.

Ένα ακόμη βασικό συμπέρασμα είναι ότι η πρόβλεψη σε επίπεδο μεμονωμένης κατοικίας παραμένει ιδιαίτερα απαιτητικό πρόβλημα. Σε αντίθεση με την πρόβλεψη συνολικού φορτίου, όπου η συμπεριφορά πολλών καταναλωτών εξομαλύνεται, η κατανάλωση μιας κατοικίας επηρεάζεται έντονα από μη πλήρως παρατηρήσιμους παράγοντες, όπως η παρουσία των ενοίκων, η χρήση συγκεκριμένων ηλεκτρικών συσκευών και οι έκτακτες αλλαγές στη ρουτίνα του νοικοκυριού. Το εύρημα αυτό συμφωνεί με τη βιβλιογραφία, σύμφωνα με την οποία η πρόβλεψη οικιακού φορτίου σε επίπεδο μεμονωμένου καταναλωτή

χαρακτηρίζεται από υψηλή μεταβλητότητα και αβεβαιότητα (Sousa & Bernardo, 2022; Han & Wang, 2023).

Η σύγκριση των αποτελεσμάτων της παρούσας εργασίας με αντίστοιχες μελέτες πρέπει να γίνεται με προσοχή, καθώς οι εργασίες διαφέρουν ως προς το σύνολο δεδομένων, τη χρονική ανάλυση, τον ορίζοντα πρόβλεψης, το επίπεδο συνάθροισης και τις χρησιμοποιούμενες μετρικές. Για τον λόγο αυτό, η άμεση αριθμητική σύγκριση των τιμών MAE, RMSE ή MAPE δεν είναι πάντοτε δίκαιη ή μεθοδολογικά ισοδύναμη. Παρ' όλα αυτά, τα αποτελέσματα της παρούσας εργασίας βρίσκονται σε λογική συμφωνία με παρόμοιες εργασίες πρόβλεψης οικιακής κατανάλωσης, στις οποίες αναφέρεται ότι οι τιμές σφάλματος μπορεί να παρουσιάζουν σημαντική διακύμανση ανάλογα με το προφίλ της κατοικίας, το μέγεθος του συνόλου δεδομένων και τον βαθμό μεταβλητότητας της κατανάλωσης (Bourdeau et al., 2019; Sousa & Bernardo, 2022; Han & Wang, 2023). Οι περιπτώσεις με υψηλότερα σφάλματα στην παρούσα αξιολόγηση δεν πρέπει επομένως να ερμηνεύονται ως αποτυχία της μεθοδολογίας, αλλά ως ένδειξη της εγγενούς δυσκολίας του προβλήματος.

Η επιλογή των δενδροειδών μοντέλων αποδείχθηκε κατάλληλη για το αντικείμενο της εργασίας. Τα μοντέλα Random Forest, XGBoost και LightGBM μπόρεσαν να αξιοποιήσουν ετερογενή χαρακτηριστικά, να διαχειριστούν μη γραμμικές σχέσεις και να προσφέρουν ικανοποιητική ερμηνευσιμότητα μέσω της ανάλυσης σημαντικότητας χαρακτηριστικών. Η χρήση τους είναι συνεπής με προηγούμενες μελέτες, στις οποίες τα μοντέλα μηχανικής μάθησης και ιδιαίτερα οι δενδροειδείς μέθοδοι εμφανίζονται αποτελεσματικά σε προβλήματα πρόβλεψης ενεργειακής κατανάλωσης με δομημένα δεδομένα (Amasyali & El-Gohary, 2018; Ahmad & Chen, 2020). Παρότι η σύγχρονη βιβλιογραφία αναδεικνύει τη δυναμική των μοντέλων βαθιάς μάθησης σε προβλήματα χρονοσειρών, τα μοντέλα αυτά συνήθως απαιτούν μεγαλύτερο όγκο δεδομένων και αυξημένους υπολογιστικούς πόρους. Για τον λόγο αυτό, τα δενδροειδή μοντέλα παραμένουν ιδιαίτερα χρήσιμη και πρακτική επιλογή σε εφαρμογές με δομημένα δεδομένα, περιορισμένο αριθμό παρατηρήσεων ή ανάγκη βασικής ερμηνευσιμότητας.

Η εργασία ανέδειξε επίσης τη σημασία του μεγέθους και της δομής του διαθέσιμου συνόλου δεδομένων. Το IDEAL, ως μεγαλύτερο και πλουσιότερο σύνολο δεδομένων, επέτρεψε πιο σταθερή εκπαίδευση και καλύτερη γενίκευση των μοντέλων. Αντίθετα, το PLEGMA, λόγω του μικρότερου αριθμού κατοικιών και της μικρότερης ποικιλίας προτύπων κατανάλωσης, οδήγησε σε μεγαλύτερη ευαισθησία των αποτελεσμάτων ανά κατοικία και ανά ημέρα αξιολόγησης. Η παρατήρηση αυτή είναι σύμφωνη με τη βιβλιογραφία, όπου επισημαίνεται ότι τα μικρά και ετερογενή σύνολα δεδομένων μπορούν να περιορίσουν τη σταθερότητα και τη γενικευσιμότητα των μοντέλων πρόβλεψης (Bourdeau et al., 2019).

Τα κύρια συμπεράσματα της εργασίας επιβεβαιώνουν ότι η μηχανική μάθηση μπορεί να αποτελέσει χρήσιμο εργαλείο για την πρόβλεψη της οικιακής ηλεκτρικής κατανάλωσης, υπό την προϋπόθεση ότι τα αποτελέσματα ερμηνεύονται με βάση το διαθέσιμο επίπεδο πληροφορίας και την ποιότητα των δεδομένων. Η προτεινόμενη μεθοδολογία έδειξε ότι η χρήση ιστορικών δεδομένων μπορεί να βελτιώσει σημαντικά την πρόβλεψη, αλλά μόνο όταν συνοδεύεται από έλεγχο σταθερότητας και αντιπροσωπευτικότητας. Παράλληλα, η ανάπτυξη λειτουργικού συστήματος μέσω API και γραφικής διεπαφής έδειξε ότι τα μοντέλα πρόβλεψης μπορούν να ενσωματωθούν σε πρακτικό περιβάλλον χρήσης, προσφέροντας μια βάση για μελλοντικές εφαρμογές σε έξυπνα κτήρια και συστήματα ενεργειακής διαχείρισης.

17.4 Σύνδεση με τα ερευνητικά ερωτήματα

Τα αποτελέσματα της εργασίας απαντούν στα βασικά ερευνητικά ερωτήματα που τέθηκαν στην Ενότητα 1.3. Αρχικά, επιβεβαιώθηκε ότι μοντέλα μηχανικής μάθησης, όπως Random Forest, XGBoost και LightGBM, μπορούν να χρησιμοποιηθούν αποτελεσματικά για την πρόβλεψη της ωριαίας ηλεκτρικής κατανάλωσης κατοικιών, αξιοποιώντας χρονικά, περιβαλλοντικά, στατικά και ιστορικά χαρακτηριστικά. Επιπλέον, η σύγκριση των δύο σεναρίων πρόβλεψης έδειξε ότι η χρήση πρόσφατου ιστορικού κατανάλωσης μπορεί να βελτιώσει σημαντικά την ακρίβεια των προβλέψεων, υπό την προϋπόθεση ότι το ιστορικό είναι πλήρες, σταθερό και αντιπροσωπευτικό της ημέρας στόχου.

Παράλληλα, η εφαρμογή της μεθοδολογίας στα σύνολα δεδομένων IDEAL και PLEGMA έδειξε ότι η ίδια βασική λογική μπορεί να προσαρμοστεί σε διαφορετικά δεδομένα και διαφορετικά γεωγραφικά περιβάλλοντα, αν και η απόδοση επηρεάζεται σημαντικά από το μέγεθος και την ποιότητα του διαθέσιμου συνόλου δεδομένων. Τέλος, η ανάπτυξη API και γραφικής διεπαφής χρήστη έδειξε ότι τα εκπαιδευμένα μοντέλα μπορούν να ενσωματωθούν σε λειτουργικό πρωτότυπο σύστημα πρόβλεψης, το οποίο υποστηρίζει διαφορετικά σενάρια χρήσης και παρέχει στον χρήστη τη δυνατότητα αξιοποίησης ιστορικών δεδομένων και ερμηνείας των προβλέψεων.

17.5 Περιορισμοί της μελέτης

Παρότι η παρούσα εργασία ανέπτυξε και αξιολόγησε μια ολοκληρωμένη μεθοδολογία για την πρόβλεψη της ηλεκτρικής κατανάλωσης κατοικιών, υπάρχουν ορισμένοι περιορισμοί που πρέπει να ληφθούν υπόψη κατά την ερμηνεία των αποτελεσμάτων. Οι περιορισμοί αυτοί σχετίζονται κυρίως με τη φύση των διαθέσιμων δεδομένων, τη δυσκολία της πρόβλεψης σε επίπεδο μεμονωμένης κατοικίας και τις τεχνικές επιλογές που έγιναν στο πλαίσιο της υλοποίησης.

Ένας βασικός περιορισμός αφορά το μέγεθος και τη δομή των συνόλων δεδομένων. Παρότι το IDEAL αποτελεί σχετικά πλούσιο σύνολο δεδομένων, το PLEGMA είναι σαφώς μικρότερο, τόσο ως προς τον αριθμό των κατοικιών όσο και ως προς την ποικιλία των διαθέσιμων προτύπων κατανάλωσης. Το γεγονός αυτό περιορίζει τη δυνατότητα των μοντέλων να μάθουν γενικεύσιμες σχέσεις και καθιστά τα αποτελέσματα περισσότερο ευαίσθητα στις ιδιαιτερότητες κάθε κατοικίας. Επομένως, τα αποτελέσματα του PLEGMA πρέπει να ερμηνεύονται ως ενδεικτική αξιολόγηση της μεθοδολογίας σε ελληνικά δεδομένα και όχι ως πλήρης γενίκευση σε όλες τις κατοικίες ή σε όλες τις συνθήκες κατανάλωσης.

Επίσης, τα αποτελέσματα της εργασίας πρέπει να ερμηνεύονται λαμβάνοντας υπόψη τον γεωγραφικό και κλιματικό περιορισμό των διαθέσιμων δεδομένων. Το IDEAL αφορά κατοικίες στο Ηνωμένο Βασίλειο, ενώ το PLEGMA αφορά ελληνικές κατοικίες. Συνεπώς, τα συμπεράσματα δεν μπορούν να γενικευθούν αυτόματα σε διαφορετικές χώρες, κλιματικές ζώνες ή ενεργειακές συνθήκες χωρίς περαιτέρω αξιολόγηση σε πρόσθετα σύνολα δεδομένων.

Ένας ακόμη περιορισμός σχετίζεται με τη μεταβλητότητα της οικιακής κατανάλωσης. Η πρόβλεψη σε επίπεδο μεμονωμένης κατοικίας επηρεάζεται από παράγοντες που δεν είναι πάντα διαθέσιμοι στα δεδομένα, όπως η παρουσία των ενοίκων, οι καθημερινές συνήθειες, η χρήση συγκεκριμένων συσκευών και οι έκτακτες αλλαγές στη συμπεριφορά του νοικοκυριού. Οι παράγοντες αυτοί μπορούν να οδηγήσουν σε απότομες αιχμές ή σε ημέρες

που αποκλίνουν σημαντικά από το πρόσφατο ιστορικό, γεγονός που δυσκολεύει την ακριβή πρόβλεψη ακόμη και όταν τα μοντέλα έχουν εκπαιδευτεί σωστά.

Ένας επιπλέον περιορισμός αφορά την αναδρομική παραγωγή προβλέψεων στο σενάριο με χρήση ιστορικών δεδομένων. Στο πλαίσιο της πρόβλεψης ολόκληρης της επόμενης ημέρας, οι πραγματικές τιμές κατανάλωσης των ενδιάμεσων ωρών της ημέρας στόχου δεν είναι διαθέσιμες κατά τη στιγμή της πρόβλεψης. Για τον λόγο αυτό, οι προβλεπόμενες τιμές των προηγούμενων ωρών χρησιμοποιούνται προσωρινά για τον υπολογισμό των ιστορικών και κυλιόμενων χαρακτηριστικών των επόμενων ωρών. Η προσέγγιση αυτή είναι συνεπής με το πλαίσιο της πρόβλεψης επόμενης ημέρας, ωστόσο ενδέχεται να οδηγήσει σε σταδιακή συσσώρευση σφάλματος κατά τη διάρκεια του 24ώρου. Η επίδραση αυτή θα μπορούσε να διερευνηθεί μελλοντικά μέσω ανάλυσης του σφάλματος ανά ώρα πρόβλεψης.

Επιπλέον, η αξιοποίηση ιστορικών δεδομένων κατανάλωσης εξαρτάται από την ποιότητα και την αντιπροσωπευτικότητα του διαθέσιμου ιστορικού. Η εργασία έδειξε ότι το πρόσφατο ιστορικό μπορεί να βελτιώσει σημαντικά την πρόβλεψη όταν είναι πλήρες, σταθερό και κοντά στη συμπεριφορά της ημέρας στόχου. Ωστόσο, όταν το ιστορικό είναι ασταθές ή δεν αντιπροσωπεύει την ημέρα πρόβλεψης, η επίδρασή του μπορεί να είναι περιορισμένη ή να βελτιώνει μόνο ορισμένες μετρικές. Συνεπώς, η χρήση ιστορικού δεν μπορεί να θεωρηθεί αυτόματα ωφέλιμη σε κάθε περίπτωση.

Περιορισμοί υπάρχουν επίσης ως προς τις διαθέσιμες μεταβλητές εισόδου. Παρότι χρησιμοποιήθηκαν χρονικά, περιβαλλοντικά, στατικά και ιστορικά χαρακτηριστικά, δεν ήταν διαθέσιμες λεπτομερείς πληροφορίες για την πραγματική παρουσία των ενοίκων, τη χρήση επιμέρους συσκευών σε πραγματικό χρόνο ή μελλοντικές αλλαγές στη συμπεριφορά του νοικοκυριού. Επίσης, τα περιβαλλοντικά δεδομένα και οι μετεωρολογικές εισοδοί ενδέχεται να μην αποτυπώνουν πάντα με πλήρη ακρίβεια τις πραγματικές συνθήκες που επηρεάζουν την κατανάλωση.

Ένας ακόμη περιορισμός αφορά το είδος των μοντέλων που χρησιμοποιήθηκαν. Η εργασία επικεντρώθηκε σε δένδροειδή μοντέλα μηχανικής μάθησης, τα οποία είναι κατάλληλα για δομημένα δεδομένα και παρέχουν ικανοποιητική απόδοση και βασική ερμηνευσιμότητα. Ωστόσο, δεν εξετάστηκαν συστηματικά μοντέλα βαθιάς μάθησης ή πιθανοτικές προσεγγίσεις πρόβλεψης, οι οποίες θα μπορούσαν να αποτυπώσουν διαφορετικές πτυχές της χρονικής εξάρτησης και της αβεβαιότητας. Η επιλογή αυτή ήταν συνειδητή, καθώς τα δένδροειδή μοντέλα είναι πιο κατάλληλα για τα διαθέσιμα δεδομένα και τους υπολογιστικούς περιορισμούς της εργασίας.

Ένας επιπλέον περιορισμός αφορά το επίπεδο ερμηνευσιμότητας των μοντέλων. Στην παρούσα εργασία η ερμηνεία των αποτελεσμάτων βασίστηκε κυρίως στην ανάλυση σημαντικότητας χαρακτηριστικών, η οποία παρέχει χρήσιμη εικόνα για τη σχετική συμβολή των μεταβλητών στη διαδικασία πρόβλεψης. Ωστόσο, δεν αξιοποιήθηκαν πιο προηγμένες τεχνικές ερμηνευσιμότητας, όπως οι τιμές SHAP (Shapley Additive Explanations), οι οποίες μπορούν να προσφέρουν λεπτομερέστερη ερμηνεία τόσο σε συνολικό επίπεδο μοντέλου όσο και σε επίπεδο μεμονωμένης πρόβλεψης (Moon et al., 2022).

Τέλος, η αξιολόγηση βασίστηκε κυρίως σε σημειακές προβλέψεις. Παρότι οι μετρικές MAE, RMSE, MAPE και ημερήσιο απόλυτο σφάλμα παρέχουν χρήσιμη εικόνα για την απόδοση των μοντέλων, δεν αποτυπώνουν πλήρως την αβεβαιότητα της πρόβλεψης. Σε ένα πραγματικό περιβάλλον χρήσης, θα μπορούσε να είναι χρήσιμη η παροχή εύρους πιθανών τιμών ή σεναρίων κατανάλωσης, ώστε ο χρήστης να κατανοεί καλύτερα το επίπεδο αβεβαιότητας που συνοδεύει την πρόβλεψη.

Οι παραπάνω περιορισμοί δεν αναιρούν τη χρησιμότητα της μεθοδολογίας, αλλά οριοθετούν το πλαίσιο μέσα στο οποίο πρέπει να ερμηνευθούν τα αποτελέσματα. Η εργασία δείχνει ότι η μηχανική μάθηση μπορεί να υποστηρίξει πρακτικά συστήματα πρόβλεψης

οικιακής κατανάλωσης, αλλά η ακρίβεια και η αξιοπιστία των προβλέψεων εξαρτώνται σε μεγάλο βαθμό από την ποιότητα των δεδομένων, τη σταθερότητα της συμπεριφοράς της κατοικίας και τη διαθεσιμότητα αξιόπιστου ιστορικού.

17.6 Προτάσεις για μελλοντικές επεκτάσεις

Με βάση τα αποτελέσματα και τους περιορισμούς της παρούσας εργασίας, μπορούν να διατυπωθούν αρκετές κατευθύνσεις για μελλοντική έρευνα και περαιτέρω ανάπτυξη του συστήματος πρόβλεψης. Οι κατευθύνσεις αυτές αφορούν τόσο τη βελτίωση της μεθοδολογίας πρόβλεψης όσο και την επέκταση της πρακτικής λειτουργίας της εφαρμογής. Μια πρώτη κατεύθυνση αφορά την αξιοποίηση μεγαλύτερων και πιο αντιπροσωπευτικών συνόλων δεδομένων. Η παρούσα εργασία έδειξε ότι το μέγεθος και η ποικιλία των δεδομένων επηρεάζουν σημαντικά την απόδοση των μοντέλων. Επομένως, η χρήση περισσότερων κατοικιών, μεγαλύτερου χρονικού εύρους μετρήσεων και πλουσιότερων πληροφοριών για το νοικοκυριό και τη χρήση συσκευών θα μπορούσε να βελτιώσει την ικανότητα γενίκευσης των μοντέλων. Η κατεύθυνση αυτή είναι ιδιαίτερα σημαντική για το PLEGMA, όπου ο περιορισμένος αριθμός κατοικιών αποτέλεσε βασικό παράγοντα μεταβλητότητας των αποτελεσμάτων.

Μια δεύτερη κατεύθυνση αφορά την περαιτέρω διερεύνηση μοντέλων βαθιάς μάθησης για χρονοσειρές, όπως LSTM, GRU, Temporal Convolutional Networks και αρχιτεκτονικές βασισμένες σε μηχανισμούς προσοχής. Τα μοντέλα αυτά μπορούν να αποτυπώσουν σύνθετες χρονικές εξαρτήσεις, ιδίως όταν υπάρχει μεγάλος όγκος δεδομένων και επαρκές ιστορικό κατανάλωσης (Qiu, 2024; Biswal et al., 2024). Στην παρούσα εργασία προτιμήθηκαν δένδροειδή μοντέλα, καθώς ήταν πιο κατάλληλα για τα διαθέσιμα δεδομένα, τους υπολογιστικούς περιορισμούς και την ανάγκη βασικής ερμηνευσιμότητας. Ωστόσο, σε μελλοντική εργασία, η συστηματική σύγκριση των δένδροειδών μοντέλων με μοντέλα βαθιάς μάθησης θα μπορούσε να προσφέρει πληρέστερη εικόνα για τα πλεονεκτήματα και τους περιορισμούς κάθε προσέγγισης.

Μια τρίτη κατεύθυνση αφορά τη διερεύνηση τρόπων αξιοποίησης γνώσης από μεγαλύτερα σύνολα δεδομένων σε μικρότερα και πιο περιορισμένα περιβάλλοντα. Στην παρούσα εργασία τα μοντέλα εκπαιδεύτηκαν ανεξάρτητα στα σύνολα δεδομένων IDEAL και PLEGMA, καθώς η άμεση μεταφορά μάθησης δεν αποτελεί τυπική ή απλή διαδικασία για τα δένδροειδή μοντέλα που χρησιμοποιήθηκαν. Ωστόσο, σε μελλοντική έρευνα θα μπορούσαν να εξεταστούν εναλλακτικές προσεγγίσεις, όπως η αξιοποίηση κοινών χαρακτηριστικών, η προεκπαίδευση μοντέλων βαθιάς μάθησης σε μεγαλύτερα σύνολα δεδομένων ή η χρήση γνώσης από ένα πλουσιότερο σύνολο δεδομένων, όπως το IDEAL, για τον σχεδιασμό καλύτερων χαρακτηριστικών και στρατηγικών προσαρμογής σε μικρότερα σύνολα δεδομένων, όπως το PLEGMA. Μια τέτοια κατεύθυνση θα μπορούσε να είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου υπάρχουν περιορισμένα δεδομένα για νέες κατοικίες ή νέα κτηριακά περιβάλλοντα (Spencer et al., 2025).

Ιδιαίτερο ενδιαφέρον παρουσιάζει επίσης η ενίσχυση της ερμηνευσιμότητας των μοντέλων. Στην παρούσα εργασία η ερμηνεία βασίστηκε κυρίως στη σημαντικότητα χαρακτηριστικών. Μελλοντικά, θα μπορούσαν να ενσωματωθούν πιο προηγμένες τεχνικές, όπως οι τιμές SHAP, ώστε να παρέχεται λεπτομερέστερη κατανόηση της επίδρασης κάθε χαρακτηριστικού τόσο σε συνολικό επίπεδο όσο και σε επίπεδο μεμονωμένης πρόβλεψης (Moon et al., 2022). Η δυνατότητα αυτή θα ήταν ιδιαίτερα χρήσιμη σε πρακτικό περιβάλλον

εφαρμογής, καθώς θα βοηθούσε τον χρήστη να κατανοεί γιατί το σύστημα παράγει μια συγκεκριμένη πρόβλεψη.

Μια ακόμη σημαντική επέκταση αφορά την πιθανοτική πρόβλεψη. Η παρούσα εργασία βασίστηκε σε σημειακές προβλέψεις, δηλαδή σε μία εκτιμώμενη τιμή κατανάλωσης για κάθε ώρα. Ωστόσο, η κατανάλωση μιας μεμονωμένης κατοικίας χαρακτηρίζεται από σημαντική αβεβαιότητα. Για τον λόγο αυτό, μελλοντικά το σύστημα θα μπορούσε να παρέχει εύρος πιθανών τιμών ή εναλλακτικά σενάρια κατανάλωσης, όπως χαμηλό, πιθανό και υψηλό σενάριο. Η προσέγγιση αυτή θα επέτρεπε πιο ρεαλιστική αποτύπωση της αβεβαιότητας και θα βοηθούσε τον χρήστη να αξιολογεί καλύτερα τον βαθμό εμπιστοσύνης της πρόβλεψης.

Σε επίπεδο εφαρμογής, μελλοντική εργασία θα μπορούσε να επικεντρωθεί στη δυναμική προσαρμογή των μοντέλων με την πάροδο του χρόνου. Η περιοδική επανεκπαίδευση ή η διαδικτυακή μάθηση με νέα δεδομένα από έξυπνους μετρητές θα μπορούσε να επιτρέψει στο σύστημα να προσαρμόζεται σε αλλαγές στη συμπεριφορά των ενοίκων, στις εποχικές συνθήκες ή στη χρήση ηλεκτρικών συσκευών. Επιπλέον, σε περιπτώσεις όπου υπάρχει επαρκές ιστορικό κατανάλωσης για μία συγκεκριμένη κατοικία, θα μπορούσε να εξεταστεί και η εξατομικευμένη εκπαίδευση μοντέλων ανά κατοικία, ώστε η πρόβλεψη να προσαρμόζεται καλύτερα στα ιδιαίτερα πρότυπα κατανάλωσης του συγκεκριμένου νοικοκυριού. Η ενσωμάτωση των μοντέλων σε πιο ολοκληρωμένα περιβάλλοντα εφαρμογής μέσω υπηρεσιών και διεπαφών χρήστη αποτελεί σημαντική κατεύθυνση για την πρακτική αξιοποίηση τέτοιων συστημάτων πρόβλεψης (Olu-Ajayi et al., 2023).

Επιπλέον, η γραφική διεπαφή χρήστη μπορεί να επεκταθεί με πιο εξελιγμένα εργαλεία οπτικοποίησης και παρακολούθησης. Ενδεικτικά, θα μπορούσαν να προστεθούν dashboards για την παρακολούθηση της ημερήσιας, εβδομαδιαίας και μηνιαίας κατανάλωσης, ειδοποιήσεις για αναμενόμενες αιχμές φορτίου, καθώς και εργαλεία εντοπισμού ασυνήθιστης συμπεριφοράς. Παράλληλα, η λειτουργία προσαρμογής συμπεριφοράς (behavioral adjustment) θα μπορούσε να επεκταθεί, ώστε ο χρήστης να εισάγει πιο εύκολα πληροφορίες για αναμενόμενες αλλαγές στη συμπεριφορά της κατοικίας, όπως απουσία από το σπίτι ή αυξημένη χρήση συγκεκριμένων συσκευών.

Τέλος, μια σημαντική μελλοντική κατεύθυνση αφορά τη σύνδεση των προβλέψεων με εφαρμογές ενεργειακής βελτιστοποίησης. Οι προβλέψεις επόμενης ημέρας θα μπορούσαν να αξιοποιηθούν για τον προγραμματισμό λειτουργίας συσκευών, τη διαχείριση φορτίων, τη χρήση φωτοβολταϊκών και συστημάτων αποθήκευσης ή τη μετατόπιση κατανάλωσης σε οικονομικότερες χρονικές ζώνες. Με αυτόν τον τρόπο, το σύστημα πρόβλεψης θα μπορούσε να αποτελέσει μέρος μιας ευρύτερης λύσης ενεργειακής διαχείρισης σε έξυπνα κτήρια και οικιακά ενεργειακά συστήματα (Lucas Segarra et al., 2020).

Εν κατακλείδι, οι μελλοντικές επεκτάσεις μπορούν να κινηθούν σε δύο βασικούς άξονες: αφενός στη βελτίωση των μοντέλων και της ερμηνευσιμότητας της πρόβλεψης και αφετέρου στην ενίσχυση της πρακτικής λειτουργίας της εφαρμογής. Η παρούσα εργασία παρέχει μια λειτουργική βάση πάνω στην οποία μπορούν να αναπτυχθούν πιο εξελιγμένα, προσαρμοστικά και χρήσιμα συστήματα πρόβλεψης ηλεκτρικής κατανάλωσης.

Βιβλιογραφία

Ahmad, T., & Chen, H. (2020). A review on machine learning forecasting growth trends and their real-time applications in different energy systems. *Sustainable Cities and Society*, 54, 102010. <https://doi.org/10.1016/j.scs.2019.102010>

Ahmad, T., Mourshed, M., & Rezgui, Y. (2018). Trees vs neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy and Buildings*, 147, 77–89. <https://doi.org/10.1016/j.enbuild.2017.04.038>

Al-Shargabi, A. A., Almhafdy, A., Ibrahim, D. M., Alghieth, M., & Chiclana, F. (2022). Buildings' energy consumption prediction models based on buildings' characteristics: Research trends, taxonomy, and performance measures. *Journal of Building Engineering*, 54, 104577. <https://doi.org/10.1016/j.jobbe.2022.104577>

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 1192–1205. <https://doi.org/10.1016/j.rser.2017.04.095>

Athanasoulas, S., Guasselli, F. A., Doulamis, N. D., Doulamis, A. D., Kokos, I., Ipiotis, N., & Kiprakis, A. (2024). The Plegma dataset: Domestic appliance-level and aggregate electricity demand with metadata from Greece. *Scientific Data*, 11, 376. <https://doi.org/10.1038/s41597-024-03208-0>

Biswal, B., Deb, S., Datta, S., Ustun, T. S., & Cali, U. (2024). Review on smart grid load forecasting for smart energy management using machine learning and deep learning techniques. *Energy Reports*, 11, 3654–3670. <https://doi.org/10.1016/j.egy.2024.09.056>

Bourdeau, M., Zhai, X. Q., Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, 48, 101533. <https://doi.org/10.1016/j.scs.2019.101533>

Challu, C., Olivares, K. G., Oreshkin, B. N., Ramirez, F. G., Canseco, M. M., Dubrawski, A., & Dholakia, K. (2023). *N-HITS: Neural hierarchical interpolation for time series forecasting*. Proceedings of the AAAI Conference on Artificial Intelligence, 37(6), 6989–6997. <https://doi.org/10.48550/arXiv.2201.12886>

Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>

Clean Energy Wire. (2025). *Interest in smart meter installation rising sharply in Germany – survey*. <https://www.cleanenergywire.org/news/interest-smart-meter-installation-rising-sharply-germany-survey>

Fisher, A., Rudin, C., & Dominici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20(177), 1–81. <https://doi.org/10.48550/arXiv.1801.01489>

Forschungsgesellschaft für Energiewirtschaft. (2025). *How is the German smart meter rollout progressing?* <https://www.ffe.de/en/publications/how-is-the-german-smart-meter-rollout-progressing/>

Han, F., & Wang, X. (2023). Adaptive individual residential load forecasting based on deep learning and dynamic mirror descent. *Frontiers in Energy Research*, 10, Article 986146. <https://doi.org/10.3389/fenrg.2022.986146>

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3), 914–938. <https://doi.org/10.1016/j.ijforecast.2015.11.011>

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. <https://otexts.com/fpp2>

IDEAL Household Energy Project. (2021). *IDEAL household energy dataset* [Data set]. University of Edinburgh. <https://datashare.ed.ac.uk/handle/10283/3647>

Kasprzyk, Ł., Pękała, P., Oreshkin, B. N., & Dudek, G. (2025). Enhanced N-BEATS for mid-term electricity demand forecasting. *Applied Soft Computing*, 182, 113575. <https://doi.org/10.1016/j.asoc.2025.113575>

Kim, M. K., Kim, Y.-S., Fu, N., Liu, J., Wang, J., Lee, S., & Srebric, J. (2025). Advanced techniques for electricity consumption prediction in buildings using comparative correlation analysis, data normalization, and Long Short-Term Memory (LSTM) networks: A case study of a U.S. commercial building. *Energy Reports*, 14, 56–65. <https://doi.org/10.1016/j.egyr.2025.05.074>

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid*, 10(1), 841–851. <https://doi.org/10.1109/TSG.2017.2753802>

Li, R., Tu, Q., Feng, H., & Zou, Z. (2025). Demand response based battery energy storage systems design and operation optimization. *Energy and Buildings*, 320, 115738. <https://doi.org/10.1016/j.enbuild.2025.115738>

Lim, B., Arik, S. O., Loeff, N., & Pfister, T. (2021). *Temporal Fusion Transformers for interpretable multi-horizon time series forecasting*. International Journal of Forecasting, 37(4), 1748–1764. <https://doi.org/10.1016/j.ijforecast.2021.03.012>

Lucas Segarra, E., Ramos Ruiz, G., & Fernández Bandera, C. (2020). Probabilistic load forecasting for building energy models. *Sensors*, 20(22), 6525. <https://doi.org/10.3390/s20226525>

Molnar, C. (2022). Interpretable machine learning: A guide for making black box models explainable (2nd ed.). Self-published. <https://christophm.github.io/interpretable-ml-book/>

Moon, J., Rho, S., & Baik, S. W. (2022). Toward explainable electrical load forecasting of buildings: A comparative study of tree-based ensemble methods with Shapley values. *Sustainable Energy Technologies and Assessments*, 54, 102888. <https://doi.org/10.1016/j.seta.2022.102888>

Olu-Ajayi, R., Alaka, H., Owolabi, H., Akanbi, L., & Ganiyu, S. (2023). Data-driven tools for building energy consumption prediction: A review. *Energies*, 16(6), 2574. <https://doi.org/10.3390/en16062574>

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering*, 45, 103406. <https://doi.org/10.1016/j.jobbe.2021.103406>

Pullinger, M., Kilgour, J., Skatova, A., Frontuto, V., Krüger, J., Shi, Y., Dunlop, F., & Staddon, J. (2021). The IDEAL household energy dataset, electricity, gas, contextual sensor data and survey data for 255 UK homes. *Scientific Data*, 8, 146. <https://doi.org/10.1038/s41597-021-00921-y>

Qiu, L. (2024). Deep learning approaches for building energy consumption prediction. *Frontiers in Environmental Research*, 2(3), 11–17. <https://doi.org/10.61784/fer3012>

Runge, J., & Zmeureanu, R. (2019). Forecasting energy use in buildings using artificial neural networks: A review. *Energies*, 12(17), 3254. <https://doi.org/10.3390/en12173254>

Sousa, J. C., & Bernardo, H. (2022). Benchmarking of load forecasting methods using residential smart meter data. *Applied Sciences*, 12(19), 9844. <https://doi.org/10.3390/app12199844>

Spencer, R., Ranathunga, S., Boulic, M., van Heerden, A. (H.), & Susnjak, T. (2025). Transfer learning on transformers for building energy consumption forecasting: A comparative study. *Energy and Buildings*, 336, 115632. <https://doi.org/10.1016/j.enbuild.2025.115632>

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9, Article 307. <https://doi.org/10.1186/1471-2105-9-307>

UN Environment Programme. (2023). *Global status report for buildings and construction 2023*. <https://www.unep.org/resources/report/global-status-report-buildings-and-construction>

vom Scheidt, F., Dong, X., Bartos, A., Staudt, P., & Weinhardt, C. (2021). Probabilistic forecasting of household loads: Effects of distributed energy technologies on forecast quality. In *Proceedings of the Twelfth ACM International Conference on Future Energy Systems (e-Energy '21)* (pp. 231–238). Association for Computing Machinery. <https://doi.org/10.1145/3447555.3464861>

Wang, J., & Li, X. (2025). Optimization of machine learning load forecasting models based on building operation characteristics: A case study for office buildings. *Energy and Buildings*, 346, 116233. <https://doi.org/10.1016/j.enbuild.2025.116233>

Yao, J. (2021). *Electricity consumption and temperature: Evidence from satellite data* (IMF Working Paper No. WP/21/22). International Monetary Fund. <https://www.imf.org/-/media/files/publications/wp/2021/english/wpica2021022-print-pdf.pdf>

Zhang, J., Wei, Y.-M., Li, D., Tan, Z., & Zhou, J. (2018). Short term electricity load forecasting using a hybrid model. *Energy*, 158, 774–781. <https://doi.org/10.1016/j.energy.2018.06.012>

Παράρτημα Α: Βιβλιοθήκες και Αρχεία με κώδικα σε γλώσσα Python

Όλα τα αρχεία υλοποίησης που έχουν αναπτυχθεί σε γλώσσα Python, συμπεριλαμβανομένων των αρχείων Jupyter Notebook (.ipynb) και των αρχείων Python (.py), είναι διαθέσιμα στο αποθετήριο GitHub του έργου στον σύνδεσμο:

[GitHub - stdGeorgios/Forecasting-the-electricity-consumption-of-buildings-: Thesis](https://github.com/stdGeorgios/Forecasting-the-electricity-consumption-of-buildings-: Thesis).

Στο ίδιο αποθετήριο διατίθεται, ως ξεχωριστό συνοδευτικό αρχείο, η αναλυτική τεχνική περιγραφή του κώδικα για τα βασικά εκτελέσιμα αρχεία της υλοποίησης. Στο παρόν παράρτημα παρέχεται συνοπτική περιγραφή της δομής του αποθετηρίου και των βασικών βιβλιοθηκών Python που χρησιμοποιήθηκαν.

Δομή αποθετηρίου GitHub

Το αποθετήριο οργανώνεται σε δύο βασικούς φακέλους, οι οποίοι αντιστοιχούν στα δύο σύνολα δεδομένων που χρησιμοποιήθηκαν στην εργασία: το IDEAL και το PLEGMA. Ο φάκελος IDEAL περιλαμβάνει τα αρχεία που αφορούν την προεπεξεργασία, τη διερευνητική ανάλυση, την εκπαίδευση των μοντέλων, την αξιολόγηση και την υλοποίηση της εφαρμογής πρόβλεψης για το σύνολο δεδομένων IDEAL. Αντίστοιχα, ο φάκελος PLEGMA περιλαμβάνει τα αρχεία που χρησιμοποιήθηκαν για την προσαρμογή της ίδιας μεθοδολογικής λογικής στο ελληνικό σύνολο δεδομένων PLEGMA.

Σε κάθε φάκελο, τα αρχεία είναι οργανωμένα ανάλογα με τον ρόλο τους στη συνολική ροή της υλοποίησης. Περιλαμβάνονται υποφάκελοι για την προεπεξεργασία και δημιουργία χαρακτηριστικών, τη διερευνητική ανάλυση δεδομένων, τα βασικά μοντέλα αναφοράς, τα μοντέλα μηχανικής μάθησης, την αξιολόγηση αποτελεσμάτων και την ανάπτυξη της προγραμματιστικής και γραφικής διεπαφής χρήστη. Με αυτόν τον τρόπο, η δομή του αποθετηρίου ακολουθεί τη μεθοδολογική πορεία της εργασίας, από την προετοιμασία των δεδομένων έως την παραγωγή προβλέψεων μέσω της εφαρμογής.

Τα πλήρη σύνολα δεδομένων δεν περιλαμβάνονται στο αποθετήριο λόγω μεγέθους και περιορισμών διάθεσης. Στο αποθετήριο παρέχονται ο κώδικας, τα συνοδευτικά αρχεία υλοποίησης και η τεκμηρίωση που απαιτούνται για την κατανόηση και αναπαραγωγή της διαδικασίας, υπό την προϋπόθεση ότι τα αντίστοιχα δεδομένα τοποθετηθούν στις κατάλληλες διαδρομές.

Βιβλιοθήκες Python

- | | |
|---------------|---|
| pandas: | Διαχείριση, καθαρισμός και μετασχηματισμός δεδομένων σε μορφή DataFrame. |
| numpy: | Αριθμητικοί υπολογισμοί και διαχείριση πινάκων. |
| scikit-learn: | Προεπεξεργασία δεδομένων, εκπαίδευση baseline και tree-based μοντέλων, καθώς και αξιολόγηση μετρικών. |

xgboost:	Υλοποίηση και εκπαίδευση του μοντέλου XGBoost.
lightgbm:	Υλοποίηση και εκπαίδευση του μοντέλου LightGBM.
joblib:	Αποθήκευση και φόρτωση εκπαιδευμένων μοντέλων και αντικειμένων προεπεξεργασίας.
matplotlib:	Δημιουργία γραφημάτων και στατικών οπτικοποιήσεων.
seaborn:	Στατιστικές οπτικοποιήσεις και αισθητική παρουσίαση δεδομένων.
plotly:	Διαδραστικές οπτικοποιήσεις, κυρίως στο περιβάλλον της διεπαφής χρήστη.
fastapi:	Ανάπτυξη του backend API για την παραγωγή προβλέψεων.
pydantic:	Ορισμός και έλεγχος της δομής των αιτημάτων και αποκρίσεων του API.
uvicorn:	Εκτέλεση του FastAPI server.
gradio:	Δημιουργία διαδραστικών web interfaces για μοντέλα ML, APIs ή άλλες Python συναρτήσεις.
datetime:	Χειρισμός ημερομηνιών, ωρών και χρονικών διαφορών.
requests:	Επικοινωνία με εξωτερικές υπηρεσίες και ανταλλαγή δεδομένων μέσω HTTP.
json:	Ανάγνωση και εγγραφή δομημένων αρχείων ρυθμίσεων και μεταδεδομένων.
pathlib / os:	Διαχείριση αρχείων, φακέλων και διαδρομών συστήματος.

Υπεύθυνη Δήλωση Συγγραφέα:

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν.1599/1986, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλει κάθε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχει έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης.