ΕΛΛΗΝΙΚΟ ΑΝΟΙΚΤΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

Σχολή Θετικών Επιστημών και Τεχνολογίας

Μεταπτυχιακή Εξειδίκευση στα Πληροφοριακά Συστήματα

Διπλωματική Εργασία

Ανάπτυξη Web-based συστήματος πρόβλεψης δασικών πυρκαγιών με χρήση τεχνικών τεχνητής νοημοσύνης: Μελέτη περίπτωσης στην Πολιτεία της Αριζόνα (ΗΠΑ)

Μαρία-Ελένη Μουστερή

(ΑΜ: 147760)

Επιβλέπων Καθηγητής: Παπαμιχαήλ Γεώργιος

Πάτρα, Ιούνιος 2026

Ανάπτυξη Web-based συστήματος πρόβλεψης δασικών πυρκαγιών με χρήση τεχνικών τεχνητής νοημοσύνης: Μελέτη περίπτωσης στην Πολιτεία της Αριζόνα (ΗΠΑ)

Μαρία-Ελένη Μουστερή

Επιβλέπων Καθηγητής: Παπαμιχαήλ Γεώργιος

Πάτρα, Ιούνιος 2026

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία της φοιτήτριας («συγγραφέας/δημιουργός») που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης η 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 της 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/της συγγραφέα/δημιουργού. Η συγγραφέας/δημιουργός διατηρεί το σύνολο των ηθικών και περιουσιακών του δικαιωμάτων.

Αφιερωμένη στον Δημήτρη — για τη στήριξη και την υπομονή του σε όλη τη διάρκεια αυτού του ταξιδιού.

Περίληψη

Η παρούσα διπλωματική εργασία πραγματεύεται τον σχεδιασμό και την ανάπτυξη ενός ολοκληρωμένου διαδικτυακού (web-based) συστήματος πρόβλεψης κινδύνου δασικών πυρκαγιών, με αξιοποίηση τεχνικών μηχανικής μάθησης και μελέτη περίπτωσης την Πολιτεία της Αριζόνα των ΗΠΑ. Το σύστημα συνδυάζει ιστορικά δεδομένα πυρκαγιών με μετεωρολογικές (θερμοκρασία, υγρασία, άνεμος, βροχόπτωση) και γεωχωρικές μεταβλητές (υψόμετρο, κλίση, τύπος βλάστησης), ώστε να εκτιμά το επίπεδο κινδύνου σε συγκεκριμένη περιοχή και χρονική στιγμή.

Για την προγνωστική μοντελοποίηση εκπαιδεύτηκαν και αξιολογήθηκαν συγκριτικά τέσσερα μοντέλα (Random Forest, Gradient Boosting, XGBoost και νευρωνικό δίκτυο MLP), με χρήση τεχνικών αντιμετώπισης της ανισορροπίας κλάσεων (SMOTE), διαστρωματωμένης διασταυρούμενης επικύρωσης, βελτιστοποίησης υπερπαραμέτρων (GridSearchCV) και ερμηνευσιμότητας (SHAP). Και τα τέσσερα μοντέλα πέτυχαν υψηλή διακριτική ικανότητα (ROC-AUC > 0,95), με το XGBoost να ξεχωρίζει (F1 = 0,890, ROC-AUC = 0,964)· το τελικό σύστημα αξιοποιεί συνδυασμό των μοντέλων (soft-voting ensemble). Η εφαρμογή υλοποιήθηκε σε Django με PostgreSQL/PostGIS και διαδραστικό χάρτη (Leaflet.js), προσφέροντας μεμονωμένη και μαζική πρόβλεψη, οπτικοποίηση κινδύνου, ιστορικό και επανεκπαίδευση.

Η χρηστικότητα του συστήματος αξιολογήθηκε με το ερωτηματολόγιο System Usability Scale (SUS) σε οκτώ χρήστες, με μέσο όρο 82,2, τιμή που αντιστοιχεί σε εξαιρετική αντιληπτή χρηστικότητα. Συμπερασματικά, η εργασία καταδεικνύει ότι η ενσωμάτωση μηχανικής μάθησης και γεωχωρικής οπτικοποίησης σε ένα ενιαίο πληροφοριακό σύστημα μπορεί να μετατρέψει αναλυτικά αποτελέσματα σε πρακτικό εργαλείο υποστήριξης αποφάσεων για την πρόληψη και διαχείριση δασικών πυρκαγιών.

Λέξεις – Κλειδιά

δασικές πυρκαγιές, μηχανική μάθηση, πρόβλεψη κινδύνου, γεωχωρικά δεδομένα (GIS), μέθοδοι συνόλου (ensemble), χρηστικότητα (SUS)

Development of a Web-based Wildfire Prediction System Using Artificial Intelligence Techniques: A Case Study in the State of Arizona (USA)

Maria - Eleni Mousteri

Abstract

This thesis addresses the design and development of an integrated web-based system for forest wildfire risk prediction using machine learning techniques, with the State of Arizona (USA) as a case study. The system combines historical fire records with meteorological variables (temperature, humidity, wind, precipitation) and geospatial features (elevation, slope, vegetation type) to estimate the level of risk for a given area and time.

Four predictive models (Random Forest, Gradient Boosting, XGBoost and an MLP neural network) were trained and comparatively evaluated, using class-imbalance handling (SMOTE), stratified cross-validation, hyperparameter optimization (GridSearchCV) and interpretability analysis (SHAP). All models achieved high discriminative ability (ROC-AUC > 0.95), with XGBoost performing best (F1 = 0.890, ROC-AUC = 0.964); the final system employs a soft-voting ensemble of the models. The application was implemented in Django with PostgreSQL/PostGIS and an interactive map (Leaflet.js), offering single and batch prediction, risk visualization, history and model retraining.

The usability of the system was assessed with the System Usability Scale (SUS) questionnaire among eight users, yielding a mean score of 82.2, which corresponds to excellent perceived usability. In conclusion, the work demonstrates that integrating machine learning and geospatial visualization into a single information system can turn analytical results into a practical decision-support tool for wildfire prevention and management.

Keywords

wildfire prediction, machine learning, risk assessment, geospatial data (GIS), ensemble methods, usability (SUS)

1. ΠΕΡΙΕΧΟΜΕΝΑ

1.ΠΕΡΙΕΧΟΜΕΝΑ.....7
2.ΚΕΦΑΛΑΙΟ 1 - Εισαγωγή.....13
1.1 Σκοπός και στόχοι της διπλωματικής εργασίας.....13
1.2 Σημασία της πρόβλεψης δασικών πυρκαγιών.....14
1.3 Ερευνητικά ερωτήματα και υποθέσεις εργασίας.....15
1.4 Δομή της διπλωματικής εργασίας.....16
3.ΚΕΦΑΛΑΙΟ 2 – Θεωρητικό υπόβαθρο.....17
2.1 Δασικές πυρκαγιές και κλιματική αλλαγή.....17
2.2 Παράγοντες κινδύνου και δυναμική εξάπλωσης.....18
2.3 Fire Weather Index (FWI) και μετεωρολογικοί δείκτες πρόβλεψης.....19
2.4 Τηλεπισκόπηση και γεωχωρικά δεδομένα (GIS).....21
2.5 Μηχανική μάθηση στην πρόβλεψη δασικών πυρκαγιών.....22
2.5.1 Random Forest.....23
2.5.2 Gradient Boosting και XGBoost.....24
2.5.3 Τεχνητά νευρωνικά δίκτυα.....24
2.5.4 Συγκριτική αξιολόγηση αλγορίθμων.....25
2.6 Κενά στη βιβλιογραφία και συνεισφορά της παρούσας εργασίας.....26
4.ΚΕΦΑΛΑΙΟ 3 – Συλλογή και ανάλυση δεδομένων.....29
3.1 Ερευνητικός σχεδιασμός.....29
3.2 Συλλογή δεδομένων.....30
3.3 Προεπεξεργασία και καθαρισμός δεδομένων.....31
3.4 Επιλογή και εξαγωγή χαρακτηριστικών (Feature engineering).....32
3.4.1 Διερευνητική ανάλυση δεδομένων.....34
3.5 Εκπαίδευση μοντέλων.....36
3.6 Μετρικές αξιολόγησης.....37

3.7 Ενσωμάτωση στο Web-based σύστημα....38
5.ΚΕΦΑΛΑΙΟ 4 – Σχεδιασμός συστήματος....40
4.1 Αρχιτεκτονική της εφαρμογής....40
4.2 Σχεδίαση βάσης δεδομένων....42
4.3 Επιλογή εκπαιδευτικού μοντέλου πρόβλεψης....44
6.ΚΕΦΑΛΑΙΟ 5 - Υλοποίηση Web-app....46
5.1 Ανάπτυξη backend με Django 46
5.2 Ενσωμάτωση μοντέλου AI....48
5.3 Οπτικοποίηση αποτελεσμάτων σε διαδραστικό χάρτη....50
5.4 Πρόσθετες λειτουργίες του συστήματος....52
7.ΚΕΦΑΛΑΙΟ 6 - Αξιολόγηση και αποτελέσματα....58
6.1 Αξιολόγηση απόδοσης και ακρίβειας μοντέλου....58
6.2 Ανάλυση αποτελεσμάτων....60
6.3 Αξιολόγηση χρηστικότητας από χρήστες....62
8.ΚΕΦΑΛΑΙΟ 7 – Συμπεράσματα και μελλοντική έρευνα....65
Βιβλιογραφία (References)....68
Παραρτήματα....70
Παράρτημα Α: Κώδικας εκπαίδευσης και αξιολόγησης μοντέλων....70
....87
Παράρτημα Β: Ερωτηματολόγιο και αναλυτικές απαντήσεις αξιολόγησης χρηστικότητας 89

Κατάλογος Εικόνων

Κατάλογος Πινάκων

Συντομογραφίες & Ακρωνύμια

AI Artificial Intelligence – Τεχνητή Νοημοσύνη

API Application Programming Interface – Διεπαφή Προγραμματισμού Εφαρμογών

AUC Area Under the Curve – Εμβαδόν κάτω από την καμπύλη

CSS Cascading Style Sheets

CSV Comma-Separated Values

CV Cross-Validation – Διασταυρούμενη Επικύρωση

EDA Exploratory Data Analysis – Διερευνητική Ανάλυση Δεδομένων

FWI Fire Weather Index – Δείκτης Μετεωρολογικού Κινδύνου Πυρκαγιάς

GIS Geographic Information System – Γεωγραφικό Σύστημα Πληροφοριών

HTML HyperText Markup Language

HTTP HyperText Transfer Protocol

JSON JavaScript Object Notation

ML Machine Learning – Μηχανική Μάθηση

MLP Multilayer Perceptron – Πολυεπίπεδο Perceptron (Νευρωνικό Δίκτυο)

NDVI Normalized Difference Vegetation Index – Κανονικοποιημένος Δείκτης Βλάστησης

ORM Object-Relational Mapping

REST Representational State Transfer

ROC Receiver Operating Characteristic

ROC-AUC Area Under the ROC Curve

SHAP SHapley Additive exPlanations

SMOTE Synthetic Minority Over-sampling Technique

SQL Structured Query Language

SUS System Usability Scale – Κλίμακα Ευχρηστίας Συστήματος

URL Uniform Resource Locator

XGBoost Extreme Gradient Boosting

2. ΚΕΦΑΛΑΙΟ 1 - Εισαγωγή

Οι δασικές πυρκαγιές αποτελούν ένα από τα πλέον σύνθετα και πολυπαραγοντικά φυσικά φαινόμενα, με σημαντικές περιβαλλοντικές, οικονομικές και κοινωνικές επιπτώσεις. Τα τελευταία χρόνια παρατηρείται διεθνώς αύξηση της συχνότητας, της έντασης και της διάρκειας των πυρκαγιών, γεγονός που αποδίδεται τόσο σε κλιματικούς παράγοντες όσο και σε ανθρωπογενείς δραστηριότητες. Η αύξηση της μέσης θερμοκρασίας, η μεταβολή των προτύπων βροχόπτωσης και οι παρατεταμένες περίοδοι ξηρασίας δημιουργούν συνθήκες υψηλής ευφλεκτότητας, ιδιαίτερα σε περιοχές με ξηρό ή ημίξηρο κλίμα.

Η πολιτεία της Αριζόνα στις Ηνωμένες Πολιτείες αποτελεί χαρακτηριστικό παράδειγμα περιοχής με αυξημένη επικινδυνότητα. Το θερμό κλίμα, η εποχική μεταβλητότητα των ανέμων και η παρουσία εκτεταμένων δασικών και θαμνωδών εκτάσεων συνθέτουν ένα περιβάλλον όπου η έγκαιρη πρόβλεψη κινδύνου καθίσταται κρίσιμη.

Η ραγδαία ανάπτυξη της τεχνητής νοημοσύνης και ειδικότερα της μηχανικής μάθησης προσφέρει νέα εργαλεία για την ανάλυση μεγάλου όγκου ετερογενών δεδομένων. Η δυνατότητα ανίχνευσης σύνθετων και μη γραμμικών σχέσεων μεταξύ περιβαλλοντικών μεταβλητών δημιουργεί προοπτικές για ανάπτυξη αξιόπιστων προγνωστικών συστημάτων, τα οποία μπορούν να υποστηρίξουν τη λήψη αποφάσεων σε επίπεδο πρόληψης και διαχείρισης κρίσεων.

Η παρούσα διπλωματική εργασία εντάσσεται σε αυτό το πλαίσιο, επιχειρώντας να συνδυάσει θεωρητική τεκμηρίωση και πρακτική υλοποίηση μέσω ανάπτυξης web-based πληροφοριακού συστήματος πρόβλεψης κινδύνου δασικών πυρκαγιών.

1.1 Σκοπός και στόχοι της διπλωματικής εργασίας

Σκοπός της παρούσας διπλωματικής εργασίας είναι η ανάπτυξη ενός ολοκληρωμένου web-based συστήματος πρόβλεψης δασικών πυρκαγιών, το οποίο θα αξιοποιεί τεχνικές μηχανικής μάθησης για την εκτίμηση επιπέδου κινδύνου σε συγκεκριμένες γεωγραφικές περιοχές της πολιτείας της Αριζόνα.

Η εργασία επιδιώκει να γεφυρώσει το χάσμα μεταξύ θεωρητικών προσεγγίσεων πρόβλεψης κινδύνου και πρακτικής εφαρμογής σε λειτουργικό πληροφοριακό σύστημα.

Οι επιμέρους στόχοι της εργασίας είναι:

- Η διερεύνηση και σύνθεση της διεθνούς βιβλιογραφίας σχετικά με την πρόβλεψη δασικών πυρκαγιών.
- Η συλλογή και ενοποίηση μετεωρολογικών, γεωχωρικών και ιστορικών δεδομένων.
- Η ανάπτυξη και συγκριτική αξιολόγηση μοντέλων μηχανικής μάθησης (π.χ. Random Forest, Boosting, Νευρωνικά Δίκτυα).
- Η επιλογή του καταλληλότερου μοντέλου βάσει τεκμηριωμένων μετρικών απόδοσης.
- Η ενσωμάτωση του προγνωστικού μοντέλου σε web εφαρμογή.
- Η χαρτογραφική απεικόνιση του επιπέδου κινδύνου μέσω διαδραστικού χάρτη.
- Η διαμόρφωση επεκτάσιμης αρχιτεκτονικής που επιτρέπει μελλοντική αναβάθμιση και επανεκπαίδευση του μοντέλου.

Μέσω των παραπάνω στόχων, η εργασία επιδιώκει όχι μόνο την ανάπτυξη τεχνικής λύσης, αλλά και την τεκμηρίωση μιας μεθοδολογικής προσέγγισης που μπορεί να εφαρμοστεί και σε άλλες γεωγραφικές περιοχές.

1.2 Σημασία της πρόβλεψης δασικών πυρκαγιών

Η πρόβλεψη δασικών πυρκαγιών αποκτά ιδιαίτερη σημασία στο πλαίσιο της κλιματικής αβεβαιότητας και της αυξανόμενης συχνότητας ακραίων φαινομένων. Οι πυρκαγιές προκαλούν:

- Απώλεια ανθρώπινων ζωών
- Καταστροφή περιουσιών και υποδομών
- Σημαντική περιβαλλοντική υποβάθμιση
- Οικονομικές επιπτώσεις σε τοπικές και εθνικές οικονομίες

Η έγκαιρη εκτίμηση κινδύνου επιτρέπει:

- Καλύτερη κατανομή πυροσβεστικών και προληπτικών πόρων
- Έγκαιρη ενημέρωση τοπικών αρχών

- Σχεδιασμό προληπτικών μέτρων (π.χ. καθαρισμός καύσιμης ύλης)
- Μείωση κόστους διαχείρισης κρίσεων

Επιπλέον, η ανάπτυξη προγνωστικών συστημάτων βασισμένων σε δεδομένα ενισχύει τη μετάβαση από αντιδραστική διαχείριση κρίσεων σε προληπτική στρατηγική αντιμετώπισης.

Η ενσωμάτωση τεχνητής νοημοσύνης στη διαδικασία πρόβλεψης επιτρέπει τη δυναμική ανάλυση πολλαπλών μεταβλητών και την προσαρμογή του μοντέλου σε μεταβαλλόμενες κλιματικές συνθήκες. Έτσι, η παρούσα εργασία αποκτά ιδιαίτερη σημασία, καθώς συνδυάζει επιστημονική τεκμηρίωση, τεχνολογική καινοτομία και πρακτική εφαρμογή.

1.3 Ερευνητικά ερωτήματα και υποθέσεις εργασίας

Η διατύπωση ερευνητικών ερωτημάτων αποτελεί κρίσιμο στοιχείο της επιστημονικής δομής της διπλωματικής εργασίας, καθώς προσδιορίζει με σαφήνεια το πρόβλημα που επιχειρείται να διερευνηθεί και το πλαίσιο μέσα στο οποίο θα κινηθεί η μεθοδολογία.

Η παρούσα εργασία επιχειρεί να απαντήσει στα ακόλουθα βασικά ερευνητικά ερωτήματα:

ΕΕ1: Σε ποιο βαθμό μπορούν μετεωρολογικά και γεωχωρικά δεδομένα να προβλέψουν αξιόπιστα το επίπεδο κινδύνου δασικής πυρκαγιάς σε περιοχή με ημίξηρο κλίμα όπως η Αριζόνα;

ΕΕ2: Ποιος αλγόριθμος μηχανικής μάθησης παρουσιάζει την καλύτερη ισορροπία μεταξύ ακρίβειας πρόβλεψης και ερμηνευσιμότητας αποτελεσμάτων;

ΕΕ3: Μπορεί η ενσωμάτωση δορυφορικών δεδομένων (π.χ. NDVI) να βελτιώσει τη χωρική ακρίβεια του προγνωστικού μοντέλου;

ΕΕ4: Πώς μπορεί ένα web-based σύστημα να μετατρέψει τα αποτελέσματα πρόβλεψης σε πρακτικό εργαλείο υποστήριξης αποφάσεων;

Με βάση τα παραπάνω ερωτήματα, διατυπώνονται οι ακόλουθες ερευνητικές υποθέσεις:

- Υ1: Η συνδυαστική χρήση μετεωρολογικών και γεωχωρικών δεδομένων οδηγεί σε σημαντικά βελτιωμένη προγνωστική απόδοση σε σχέση με τη χρήση αποκλειστικά μετεωρολογικών μεταβλητών.
- Υ2: Οι ensemble μέθοδοι (π.χ. Random Forest, Gradient Boosting) παρουσιάζουν υψηλότερη σταθερότητα από απλά γραμμικά μοντέλα.

- Υ3: Η χωρική οπτικοποίηση των αποτελεσμάτων ενισχύει τη χρηστικότητα του συστήματος σε επιχειρησιακό επίπεδο.

Η απάντηση στα παραπάνω ερωτήματα θα επιτευχθεί μέσω της μεθοδολογικής διαδικασίας που περιγράφεται στα επόμενα κεφάλαια.

1.4 Δομή της διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία οργανώνεται σε διακριτά κεφάλαια, τα οποία ακολουθούν λογική αλληλουχία από τη θεωρητική θεμελίωση έως την τεχνική υλοποίηση και αξιολόγηση.

Στο Κεφάλαιο 1 (Εισαγωγή) παρουσιάζεται το γενικό πλαίσιο του προβλήματος, ο σκοπός και οι στόχοι της εργασίας, καθώς και τα ερευνητικά ερωτήματα.

Στο Κεφάλαιο 2 (Θεωρητικό υπόβαθρο) αναλύεται η διεθνής βιβλιογραφία σχετικά με τις δασικές πυρκαγιές, τους μετεωρολογικούς δείκτες πρόβλεψης, την τηλεπισκόπηση και τις τεχνικές μηχανικής μάθησης.

Στο Κεφάλαιο 3 (Μεθοδολογία) περιγράφεται η ερευνητική διαδικασία, η συλλογή και προεπεξεργασία δεδομένων, η επιλογή χαρακτηριστικών και η εκπαίδευση των μοντέλων.

Στο Κεφάλαιο 4 (Σχεδιασμός συστήματος) παρουσιάζεται η αρχιτεκτονική της εφαρμογής, η δομή της βάσης δεδομένων και ο τρόπος ενσωμάτωσης του προγνωστικού μοντέλου σε web περιβάλλον.

Στα επόμενα κεφάλαια (τα οποία θα αναπτυχθούν στο τελικό στάδιο της εργασίας) θα παρουσιαστούν:

- Τα αποτελέσματα της αξιολόγησης των μοντέλων
- Η συγκριτική ανάλυση απόδοσης
- Η συζήτηση και ερμηνεία των ευρημάτων
- Τα συμπεράσματα και προτάσεις για μελλοντική έρευνα

Η δομή αυτή διασφαλίζει συνοχή μεταξύ θεωρίας, μεθοδολογίας και πρακτικής εφαρμογής.

3. ΚΕΦΑΛΑΙΟ 2 – Θεωρητικό υπόβαθρο

2.1 Δασικές πυρκαγιές και κλιματική αλλαγή

Οι δασικές πυρκαγιές αποτελούν διαχρονικό και φυσικό οικολογικό φαινόμενο, το οποίο διαδραματίζει σημαντικό ρόλο στην ανανέωση και αναδιάρθρωση των οικοσυστημάτων. Ωστόσο, κατά τις τελευταίες δεκαετίες παρατηρείται διεθνώς αξιοσημείωτη αύξηση της έντασης, της συχνότητας και της γεωγραφικής εξάπλωσης των πυρκαγιών, γεγονός που έχει προκαλέσει έντονο επιστημονικό και πολιτικό προβληματισμό. Η διεθνής επιστημονική κοινότητα έχει τεκμηριώσει με σαφήνεια τη συσχέτιση μεταξύ της κλιματικής αλλαγής και της αυξανόμενης επικινδυνότητας των πυρκαγιών, ιδιαίτερα σε περιοχές με ξηρό και θερμό κλίμα (IPCC, 2021).

Η αύξηση της μέσης παγκόσμιας θερμοκρασίας επιδρά άμεσα στη φυσική κατάσταση της καύσιμης ύλης. Η μείωση της υγρασίας του εδάφους και της βλάστησης καθιστά τα φυτικά υπολείμματα πιο εύφλεκτα, ενώ η επιμήκυνση της θερμής περιόδου αυξάνει τη χρονική διάρκεια κατά την οποία οι συνθήκες είναι ευνοϊκές για την εκδήλωση πυρκαγιάς. Παράλληλα, οι μεταβολές στα πρότυπα βροχόπτωσης, όπως η συγκέντρωση βροχής σε μικρότερο χρονικό διάστημα ή η εμφάνιση παρατεταμένων περιόδων ξηρασίας, ενισχύουν τη διαθεσιμότητα ξηρής καύσιμης ύλης. Οι συνθήκες αυτές δεν αυξάνουν μόνο την πιθανότητα ανάφλεξης, αλλά και τη δυναμική και την ταχύτητα εξάπλωσης της φωτιάς (Westerling et al., 2006).

Σύμφωνα με τους Abatzoglou και Williams (2016), η ανθρωπογενής κλιματική αλλαγή έχει συμβάλει ουσιαστικά στη διπλάσια περίπου αύξηση της καμένης έκτασης στις δυτικές Ηνωμένες Πολιτείες κατά τις τελευταίες δεκαετίες. Η μελέτη τους επισημαίνει ότι η άνοδος της θερμοκρασίας και η αυξημένη ατμοσφαιρική ξηρότητα αποτελούν κρίσιμους παράγοντες που μεταβάλλουν τη φυσική ισορροπία των οικοσυστημάτων και ενισχύουν τη συχνότητα μεγάλων και καταστροφικών πυρκαγιών (Dennison et al., 2014).

Η Πολιτεία της Αριζόνα αποτελεί χαρακτηριστικό παράδειγμα περιοχής υψηλής ευαλωτότητας. Το θερμό και ημίξηρο κλίμα, σε συνδυασμό με την έντονη εποχική μεταβλητότητα των ανέμων και των βροχοπτώσεων, δημιουργεί περιβάλλον όπου η καύσιμη ύλη μπορεί να παραμείνει σε κρίσιμα χαμηλά επίπεδα υγρασίας για μεγάλα χρονικά διαστήματα. Επιπλέον, η τοπογραφία και η ποικιλία τύπων βλάστησης ενισχύουν την πολυπλοκότητα της συμπεριφοράς των πυρκαγιών.

Η κατανόηση τόσο των μακροκλιματικών μεταβολών όσο και των τοπικών περιβαλλοντικών χαρακτηριστικών αποτελεί θεμελιώδη προϋπόθεση για την ανάπτυξη αξιόπιστων προγνωστικών μοντέλων. Η σύνδεση της κλιματικής διάστασης με ποσοτικά δεδομένα και αλγοριθμικές προσεγγίσεις πρόβλεψης συνιστά κρίσιμο βήμα για τη μετάβαση από τη γενική περιγραφή του φαινομένου σε εφαρμοσμένα συστήματα εκτίμησης κινδύνου.

2.2 Παράγοντες κινδύνου και δυναμική εξάπλωσης

Η εκδήλωση και η εξάπλωση μιας δασικής πυρκαγιάς αποτελούν αποτέλεσμα της συνδυαστικής επίδρασης τριών βασικών κατηγοριών παραγόντων: της καύσιμης ύλης, των καιρικών συνθηκών και της τοπογραφίας. Το κλασικό «τρίγωνο της φωτιάς» (καύσιμο, θερμότητα, οξυγόνο), το οποίο περιγράφει τις βασικές φυσικές προϋποθέσεις ανάφλεξης, σε περιβαλλοντικό επίπεδο μετασχηματίζεται σε ένα σύνθετο και πολυπαραγοντικό πρόβλημα ανάλυσης δεδομένων. Η αλληλεπίδραση φυσικών, κλιματικών και γεωμορφολογικών μεταβλητών δημιουργεί ένα δυναμικό σύστημα, στο οποίο μικρές μεταβολές μπορούν να οδηγήσουν σε σημαντικές διαφοροποιήσεις στη συμπεριφορά της φωτιάς.

Στη διεθνή βιβλιογραφία, οι βασικές μεταβλητές που λαμβάνονται υπόψη για την εκτίμηση κινδύνου περιλαμβάνουν:

Θερμοκρασία αέρα

Σχετική υγρασία

Ταχύτητα και κατεύθυνση ανέμου

Ύψος βροχόπτωσης

Υψόμετρο

Κλίση και προσανατολισμό εδάφους

Τύπο, πυκνότητα και υγρασία βλάστησης

Η θερμοκρασία και η σχετική υγρασία επηρεάζουν άμεσα την περιεκτικότητα της καύσιμης ύλης σε υγρασία, ενώ η βροχόπτωση καθορίζει τη μακροχρόνια κατάσταση ξηρασίας μιας περιοχής. Η ταχύτητα και η κατεύθυνση του ανέμου θεωρούνται από τους πλέον καθοριστικούς παράγοντες, καθώς επηρεάζουν όχι μόνο την πιθανότητα διάδοσης της φωτιάς αλλά και τη μεταφορά θερμότητας και καύσιμων σωματιδίων σε γειτονικές περιοχές, επιταχύνοντας τη χωρική εξάπλωση.

Η τοπογραφία προσθέτει επιπλέον πολυπλοκότητα στο φαινόμενο. Η κλίση του εδάφους μπορεί να αυξήσει την ταχύτητα ανόδου της φωτιάς, ενώ ο προσανατολισμός επηρεάζει την έκθεση στην ηλιακή ακτινοβολία και συνεπώς τη θερμική συμπεριφορά της επιφάνειας. Παράλληλα, ο τύπος και η πυκνότητα της βλάστησης καθορίζουν τη διαθέσιμη καύσιμη ύλη και τη δυναμική καύσης.

Η πολυδιάστατη αυτή αλληλεπίδραση μεταβλητών καθιστά ιδιαίτερα δύσκολη την αποτύπωση του φαινομένου μέσω απλών γραμμικών στατιστικών μοντέλων. Οι σχέσεις μεταξύ των παραγόντων συχνά είναι μη γραμμικές και εξαρτώνται από κατώφλια (threshold effects), εποχικότητα και χωρικές διαφοροποιήσεις. Για τον λόγο αυτό, η σύγχρονη έρευνα στρέφεται ολοένα και περισσότερο σε τεχνικές Μηχανικής Μάθησης, οι οποίες έχουν τη δυνατότητα να ανιχνεύουν σύνθετα πρότυπα, να μοντελοποιούν μη γραμμικές εξαρτήσεις και να αξιοποιούν μεγάλου όγκου ετερογενή δεδομένα (Jain et al., 2020).

Η μετάβαση από παραδοσιακές προσεγγίσεις σε αλγοριθμικά μοντέλα δεν υποκαθιστά τη φυσική κατανόηση του φαινομένου, αλλά τη συμπληρώνει, επιτρέποντας την ποσοτική αποτύπωση της δυναμικής των πυρκαγιών και τη δημιουργία πιο αξιόπιστων συστημάτων πρόβλεψης.

2.3 Fire Weather Index (FWI) και μετεωρολογικοί δείκτες πρόβλεψης

Ο Fire Weather Index (FWI) αποτελεί ένα από τα πλέον διαδεδομένα και διεθνώς αναγνωρισμένα συστήματα εκτίμησης κινδύνου δασικής πυρκαγιάς. Αναπτύχθηκε στον Καναδά από τον Van Wagner (1987) και εντάσσεται στο ευρύτερο Canadian Forest Fire Weather Index System, το οποίο σχεδιάστηκε για την ποσοτική αποτύπωση της επίδρασης των καιρικών συνθηκών στη δυναμική της καύσιμης ύλης και στη συμπεριφορά της φωτιάς. Η ευρεία υιοθέτησή του από οργανισμούς πολιτικής προστασίας και δασικής διαχείρισης οφείλεται στην απλότητα των απαιτούμενων εισόδων και στην ικανότητά του να παρέχει καθημερινές εκτιμήσεις επικινδυνότητας (Wotton, 2009).

Το σύστημα FWI βασίζεται σε τέσσερις βασικές μετεωρολογικές μεταβλητές:

- Θερμοκρασία αέρα
- Σχετική υγρασία
- Ταχύτητα ανέμου

- Ύψος ημερήσιας βροχόπτωσης

Οι μεταβλητές αυτές χρησιμοποιούνται για τον υπολογισμό ενδιάμεσων δεικτών που αποτυπώνουν διαφορετικά επίπεδα υγρασίας της καύσιμης ύλης. Συγκεκριμένα, οι δείκτες Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC) και Drought Code (DC) εκτιμούν την υγρασία επιφανειακών, μεσαίων και βαθύτερων στρωμάτων καύσιμης ύλης αντίστοιχα. Οι δείκτες αυτοί συνδυάζονται στη συνέχεια για την παραγωγή του τελικού δείκτη FWI, ο οποίος εκφράζει ποσοτικά την πιθανή ένταση μιας ενδεχόμενης πυρκαγιάς, εφόσον αυτή εκδηλωθεί.

Παρότι ο FWI έχει αποδειχθεί ιδιαίτερα αποτελεσματικός σε εύκρατα και βορεακά δασικά οικοσυστήματα, η εφαρμογή του σε διαφορετικά κλιματικά περιβάλλοντα απαιτεί προσαρμογές και τοπική βαθμονόμηση. Σε περιοχές με ημίξηρο ή μεσογειακό κλίμα, όπως η Αριζόνα, η έντονη εποχική μεταβλητότητα της βροχόπτωσης, οι υψηλές θερμοκρασίες και οι μεγάλες περίοδοι ξηρασίας ενδέχεται να επηρεάζουν την ευαισθησία και τη σταθερότητα του δείκτη. Συνεπώς, η απλή εφαρμογή του FWI χωρίς τοπική προσαρμογή μπορεί να οδηγήσει σε υπερεκτίμηση ή υποεκτίμηση του πραγματικού κινδύνου.

Για τον λόγο αυτό, η σύγχρονη ερευνητική προσέγγιση δεν περιορίζεται στη χρήση ενός και μόνο δείκτη. Αντίθετα, προτείνεται η συνδυαστική αξιοποίηση του FWI με τεχνικές Μηχανικής Μάθησης, οι οποίες επιτρέπουν την ενσωμάτωση επιπλέον μεταβλητών και την προσαρμογή του μοντέλου στα ιδιαίτερα χαρακτηριστικά κάθε περιοχής. Η αλγοριθμική ανάλυση μπορεί να αξιοποιήσει τον FWI ως ένα από τα χαρακτηριστικά εισόδου (feature), ενισχύοντας την προγνωστική ακρίβεια μέσω ανίχνευσης μη γραμμικών σχέσεων.

Παράλληλα, έχουν προταθεί εναλλακτικοί ή συμπληρωματικοί δείκτες, όπως ο Keetch–Byram Drought Index (KBDI), ο οποίος εστιάζει στο επίπεδο ξηρασίας του εδάφους, και ο Haines Index, που λαμβάνει υπόψη τη θερμοδυναμική σταθερότητα της ατμόσφαιρας και τη δυναμική ανύψωσης θερμών μαζών αέρα. Η ενσωμάτωση πολλαπλών δεικτών σε ένα υβριδικό προγνωστικό σύστημα αντανακλά τη σύγχρονη τάση της βιβλιογραφίας προς ολοκληρωμένες και πολυπαραγοντικές προσεγγίσεις πρόβλεψης κινδύνου.

Η παρούσα εργασία υιοθετεί αυτή τη φιλοσοφία, αντιμετωπίζοντας τον FWI όχι ως αυτόνομο εργαλείο, αλλά ως συστατικό στοιχείο ενός ευρύτερου αλγοριθμικού πλαισίου πρόβλεψης, το οποίο συνδυάζει μετεωρολογικά, γεωχωρικά και ιστορικά δεδομένα.

2.4 Τηλεπισκόπηση και γεωχωρικά δεδομένα (GIS)

Η πρόοδος στην τεχνολογία δορυφορικής παρατήρησης της Γης έχει ενισχύσει σημαντικά τη δυνατότητα ανάλυσης κινδύνου πυρκαγιάς σε χωρικό επίπεδο. Η συνεχής βελτίωση της χωρικής και χρονικής ανάλυσης των δορυφορικών δεδομένων επιτρέπει την παρακολούθηση περιβαλλοντικών μεταβλητών σε σχεδόν πραγματικό χρόνο, προσφέροντας πολύτιμη πληροφορία για την εκτίμηση της ευφλεκτότητας μιας περιοχής. Δεδομένα από δορυφόρους όπως οι MODIS και Landsat παρέχουν πληροφορίες σχετικά με τη θερμοκρασία επιφάνειας, την υγρασία του εδάφους, τη δομή και την πυκνότητα της βλάστησης, καθώς και μεταβολές στη γήινη κάλυψη (Chuvieco & Congalton, 1989).

Ένας από τους πιο συχνά χρησιμοποιούμενους δείκτες στη βιβλιογραφία είναι ο Normalized Difference Vegetation Index (NDVI), ο οποίος αποτυπώνει τη «ζωντάνια», την υγεία και την πυκνότητα της βλάστησης μέσω της ανάλυσης φασματικών δεδομένων. Ο δείκτης αυτός βασίζεται στη διαφορά ανάμεσα στην ανακλαστικότητα στο εγγύς υπέρυθρο και στο κόκκινο φάσμα, παρέχοντας ποσοτική ένδειξη της φωτοσυνθετικής δραστηριότητας. Χαμηλές τιμές NDVI συχνά συνδέονται με ξηρή ή αραιή βλάστηση, η οποία σε συνδυασμό με υψηλές θερμοκρασίες και χαμηλή σχετική υγρασία μπορεί να υποδηλώνει αυξημένη ευφλεκτότητα και υψηλότερο κίνδυνο εκδήλωσης πυρκαγιάς.

Τα Γεωγραφικά Πληροφοριακά Συστήματα (GIS) διαδραματίζουν καθοριστικό ρόλο στη χωρική ανάλυση και οπτικοποίηση αυτών των δεδομένων. Μέσω των GIS είναι δυνατή η διαχείριση, επεξεργασία και σύνθεση πολλαπλών γεωχωρικών επιπέδων πληροφορίας, όπως τοπογραφικά χαρακτηριστικά (κλίση και προσανατολισμός εδάφους), τύποι βλάστησης, υψόμετρο και μετεωρολογικά δεδομένα. Η δυνατότητα χαρτογράφησης ζωνών υψηλού κινδύνου επιτρέπει την αναγνώριση περιοχών αυξημένης ευαλωτότητας και τη χωρική σύγκριση διαφορετικών παραγόντων.

Η ενσωμάτωση γεωχωρικών δεδομένων σε προγνωστικά μοντέλα συμβάλλει ουσιαστικά στη βελτίωση της χωρικής ακρίβειας της πρόβλεψης. Η χωρική διάσταση του κινδύνου δεν μπορεί να αποτυπωθεί επαρκώς μέσω αμιγώς αριθμητικών μεταβλητών, χωρίς γεωγραφική αναφορά. Στο πλαίσιο της παρούσας εργασίας, η αξιοποίηση GIS δεν περιορίζεται στην ανάλυση, αλλά επεκτείνεται και στην παρουσίαση των αποτελεσμάτων. Η απεικόνιση του επιπέδου κινδύνου σε διαδραστικό χάρτη, μέσω web εφαρμογής, μετατρέπει την προγνωστική πληροφορία σε πρακτικό εργαλείο υποστήριξης αποφάσεων, ενισχύοντας τη χρηστικότητα και την επιχειρησιακή αξία του συστήματος.

2.5 Μηχανική μάθηση στην πρόβλεψη δασικών πυρκαγιών

Η πολυπλοκότητα των περιβαλλοντικών παραγόντων που σχετίζονται με την εκδήλωση και την εξάπλωση των δασικών πυρκαγιών καθιστά τα παραδοσιακά στατιστικά μοντέλα συχνά ανεπαρκή για την αξιόπιστη πρόβλεψη κινδύνου. Οι σχέσεις μεταξύ θερμοκρασίας, υγρασίας, ανέμου, τοπογραφίας και βλάστησης δεν είναι γραμμικές ούτε σταθερές στον χρόνο, ενώ συχνά εμφανίζουν φαινόμενα κατωφλίων (threshold effects) και πολύπλοκες αλληλεπιδράσεις. Ως εκ τούτου, η χρήση κλασικών γραμμικών μοντέλων ή απλών παλινδρομήσεων ενδέχεται να μην αποτυπώνει με επάρκεια τη δυναμική του φαινομένου.

Η μηχανική μάθηση προσφέρει τη δυνατότητα ανάλυσης μεγάλου όγκου ετερογενών δεδομένων και την ανίχνευση μη γραμμικών σχέσεων και σύνθετων αλληλεξαρτήσεων μεταξύ μεταβλητών. Μέσω αλγορίθμων επιβλεπόμενης μάθησης, είναι δυνατή η εκπαίδευση μοντέλων που «μαθαίνουν» από ιστορικά δεδομένα και παράγουν προβλέψεις για νέες περιπτώσεις, χωρίς να απαιτείται εκ των προτέρων ρητή διατύπωση μαθηματικής σχέσης μεταξύ των μεταβλητών.

Στη διεθνή βιβλιογραφία έχουν εφαρμοστεί διάφορες τεχνικές μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών, όπως δέντρα απόφασης, Random Forest, Gradient Boosting και τεχνητά νευρωνικά δίκτυα. Οι μέθοδοι ensemble, όπως το Random Forest, παρουσιάζουν ιδιαίτερο ενδιαφέρον, καθώς συνδυάζουν πολλαπλά μοντέλα για τη βελτίωση της ακρίβειας και της σταθερότητας της πρόβλεψης. Παράλληλα, τα νευρωνικά δίκτυα έχουν τη δυνατότητα μοντελοποίησης ιδιαίτερα σύνθετων προτύπων, ιδίως όταν συνδυάζονται με μεγάλο όγκο δεδομένων (Cortez & Morais, 2007; Sayad et al., 2019).

Ένα σημαντικό πλεονέκτημα των αλγοριθμικών προσεγγίσεων είναι η δυνατότητα αξιολόγησης της σχετικής σημασίας των μεταβλητών (feature importance), γεγονός που επιτρέπει την καλύτερη κατανόηση του ρόλου κάθε παράγοντα στη διαμόρφωση του κινδύνου. Έτσι, η μηχανική μάθηση δεν λειτουργεί μόνο ως «μαύρο κουτί» πρόβλεψης, αλλά μπορεί να συμβάλει και στην ερμηνεία του φαινομένου.

Η αξιοποίηση τεχνικών μηχανικής μάθησης στην πρόβλεψη δασικών πυρκαγιών δεν αντικαθιστά τους παραδοσιακούς δείκτες κινδύνου, αλλά λειτουργεί συμπληρωματικά, επιτρέποντας τη σύνθεση πολλαπλών πηγών δεδομένων σε ένα ενιαίο προγνωστικό πλαίσιο. Στο πλαίσιο της παρούσας εργασίας, η μηχανική μάθηση αποτελεί τον βασικό μηχανισμό

ανάλυσης και πρόβλεψης, ενσωματώνοντας μετεωρολογικά, γεωχωρικά και ιστορικά δεδομένα σε ένα συνεκτικό αλγοριθμικό σύστημα.

2.5.1 Random Forest

Ο αλγόριθμος Random Forest (Breiman, 2001) αποτελεί μία από τις πλέον διαδεδομένες μεθόδους επιβλεπόμενης μάθησης και βασίζεται στη δημιουργία μεγάλου αριθμού δέντρων απόφασης (decision trees), τα οποία εκπαιδεύονται σε τυχαία υποσύνολα των δεδομένων. Η τελική πρόβλεψη προκύπτει από τη συνδυαστική αξιολόγηση των επιμέρους δέντρων, μέσω μηχανισμού πλειοψηφίας (στην περίπτωση ταξινόμησης) ή μέσου όρου (στην περίπτωση παλινδρόμησης). Η διαδικασία τυχαίας επιλογής δειγμάτων και χαρακτηριστικών (bagging και random feature selection) μειώνει τη συσχέτιση μεταξύ των δέντρων και ενισχύει τη γενικευσιμότητα του μοντέλου.

Ένα από τα βασικά πλεονεκτήματα του Random Forest είναι η ανθεκτικότητά του στην υπερπροσαρμογή (overfitting), η οποία συχνά παρατηρείται σε μεμονωμένα δέντρα απόφασης. Μέσω της συλλογικής λειτουργίας πολλαπλών δέντρων, το μοντέλο επιτυγχάνει ισορροπία μεταξύ ακρίβειας και σταθερότητας. Επιπλέον, έχει τη δυνατότητα χειρισμού μεγάλου αριθμού χαρακτηριστικών χωρίς να απαιτεί εκ των προτέρων επιλογή μεταβλητών, γεγονός που το καθιστά ιδιαίτερα κατάλληλο για προβλήματα με πολυδιάστατα περιβαλλοντικά δεδομένα.

Στη διεθνή βιβλιογραφία, ο Random Forest έχει παρουσιάσει υψηλή ακρίβεια σε προβλήματα ταξινόμησης κινδύνου δασικής πυρκαγιάς, ιδίως όταν τα δεδομένα περιλαμβάνουν πολλαπλές μετεωρολογικές, γεωχωρικές και οικολογικές μεταβλητές. Η ικανότητά του να αποτυπώνει μη γραμμικές σχέσεις και σύνθετες αλληλεπιδράσεις μεταξύ παραγόντων τον καθιστά ιδιαίτερα αποτελεσματικό σε περιβάλλοντα όπου το φαινόμενο δεν μπορεί να περιγραφεί μέσω απλών γραμμικών μοντέλων (Rodrigues & de la Riva, 2014).

Παράλληλα, ο αλγόριθμος παρέχει τη δυνατότητα υπολογισμού της σχετικής σημασίας των χαρακτηριστικών (feature importance), επιτρέποντας την αξιολόγηση του βαθμού συμβολής κάθε μεταβλητής στη διαμόρφωση του κινδύνου. Το χαρακτηριστικό αυτό προσδίδει επιπλέον ερμηνευτική αξία στο μοντέλο και διευκολύνει τη σύνδεση των αποτελεσμάτων με τη φυσική κατανόηση του φαινομένου.

2.5.2 Gradient Boosting και XGBoost

Οι μέθοδοι boosting αποτελούν κατηγορία αλγορίθμων ensemble μάθησης, οι οποίοι βασίζονται στη διαδοχική εκπαίδευση επιμέρους μοντέλων με στόχο τη σταδιακή βελτίωση της συνολικής απόδοσης. Σε αντίθεση με τις τεχνικές bagging, όπου τα μοντέλα εκπαιδεύονται παράλληλα και ανεξάρτητα, στο boosting κάθε νέο μοντέλο εκπαιδεύεται ώστε να διορθώσει τα σφάλματα του προηγούμενου. Η διαδικασία αυτή επιτυγχάνεται μέσω αύξησης του βάρους των παρατηρήσεων που ταξινομήθηκαν λανθασμένα, επιτρέποντας στο σύστημα να εστιάζει στα πιο δύσκολα παραδείγματα (Friedman, 2001).

Το XGBoost (Extreme Gradient Boosting) αποτελεί μία βελτιστοποιημένη και υπολογιστικά αποδοτική υλοποίηση της μεθόδου gradient boosting. Ενσωματώνει τεχνικές κανονικοποίησης (regularization), έλεγχο υπερπροσαρμογής και αποδοτική διαχείριση μνήμης, γεγονός που το καθιστά ιδιαίτερα αποτελεσματικό σε μεγάλα και πολύπλοκα σύνολα δεδομένων. Για τον λόγο αυτό, έχει εφαρμοστεί ευρέως σε προβλήματα πρόβλεψης φυσικών καταστροφών, συμπεριλαμβανομένων των δασικών πυρκαγιών, όπου απαιτείται ανάλυση πολλαπλών και αλληλεξαρτώμενων μεταβλητών (Chen & Guestrin, 2016).

Τα μοντέλα boosting συχνά επιτυγχάνουν υψηλότερη ακρίβεια σε σχέση με τα παραδοσιακά δέντρα απόφασης, καθώς μειώνουν τόσο το σφάλμα μεροληψίας (bias) όσο και το σφάλμα διακύμανσης (variance). Ωστόσο, η αυξημένη πολυπλοκότητα συνοδεύεται από ανάγκη προσεκτικής ρύθμισης των υπερπαραμέτρων, όπως ο ρυθμός μάθησης (learning rate), το βάθος των δέντρων και ο αριθμός των επαναλήψεων. Η μη ορθή ρύθμιση μπορεί να οδηγήσει είτε σε υπερπροσαρμογή είτε σε υποεκπαίδευση του μοντέλου.

Στο πλαίσιο πρόβλεψης κινδύνου δασικών πυρκαγιών, οι μέθοδοι boosting προσφέρουν τη δυνατότητα αποτύπωσης λεπτών διαφοροποιήσεων μεταξύ περιοχών διαφορετικής επικινδυνότητας, ιδίως όταν τα δεδομένα περιλαμβάνουν σύνθετες περιβαλλοντικές αλληλεπιδράσεις. Για τον λόγο αυτό, αποτελούν σημαντική εναλλακτική επιλογή σε σύγκριση με άλλες μεθόδους ensemble.

2.5.3 Τεχνητά νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα αποτελούν μία από τις πλέον ισχυρές κατηγορίες αλγορίθμων Μηχανικής Μάθησης, καθώς έχουν τη δυνατότητα μοντελοποίησης ιδιαίτερα σύνθετων και μη γραμμικών σχέσεων μεταξύ μεταβλητών. Ειδικότερα, τα πολυεπίπεδα perceptrons (Multilayer Perceptrons – MLP) βασίζονται σε διαδοχικά επίπεδα νευρώνων, τα οποία

επεξεργάζονται τα δεδομένα μέσω συναρτήσεων ενεργοποίησης, επιτρέποντας τη σύλληψη πολύπλοκων προτύπων που δεν μπορούν να αποτυπωθούν με γραμμικά μοντέλα.

Η αρχιτεκτονική των MLP περιλαμβάνει επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα (hidden layers) και επίπεδο εξόδου. Μέσω της διαδικασίας οπισθοδιάδοσης σφάλματος (backpropagation), το δίκτυο προσαρμόζει τα βάρη των συνδέσεων ώστε να ελαχιστοποιεί το σφάλμα πρόβλεψης. Η δυνατότητα ύπαρξης πολλών κρυφών επιπέδων οδηγεί σε μοντέλα βαθιάς μάθησης (deep learning), τα οποία μπορούν να αναγνωρίσουν λεπτές και υψηλής πολυπλοκότητας αλληλεξαρτήσεις μεταξύ περιβαλλοντικών μεταβλητών.

Σε περιπτώσεις μεγάλου όγκου δεδομένων (big data), τα deep learning μοντέλα έχουν δείξει ιδιαίτερα ενθαρρυντικά αποτελέσματα σε προβλήματα πρόβλεψης φυσικών φαινομένων, συμπεριλαμβανομένων των δασικών πυρκαγιών. Η ικανότητά τους να επεξεργάζονται μεγάλες ποσότητες ετερογενούς πληροφορίας, όπως μετεωρολογικά, χωρικά και δορυφορικά δεδομένα, τα καθιστά ισχυρό εργαλείο σε σύνθετα προγνωστικά περιβάλλοντα (LeCun et al., 2015).

Ωστόσο, παρά την υψηλή προγνωστική τους ικανότητα, τα νευρωνικά δίκτυα παρουσιάζουν σημαντικό περιορισμό ως προς την ερμηνευσιμότητα. Ο λεγόμενος «black-box» χαρακτήρας τους καθιστά δύσκολη την κατανόηση του τρόπου με τον οποίο κάθε μεταβλητή επηρεάζει την τελική πρόβλεψη. Η μειωμένη διαφάνεια αποτελεί κρίσιμο ζήτημα, ιδίως σε εφαρμογές που σχετίζονται με τη δημόσια ασφάλεια και τη λήψη επιχειρησιακών αποφάσεων, όπου απαιτείται σαφής τεκμηρίωση και αιτιολόγηση των αποτελεσμάτων.

Συνεπώς, η επιλογή χρήσης νευρωνικών δικτύων σε συστήματα πρόβλεψης κινδύνου πυρκαγιάς θα πρέπει να λαμβάνει υπόψη όχι μόνο την ακρίβεια, αλλά και την ανάγκη για ερμηνευσιμότητα και διαφάνεια.

2.5.4 Συγκριτική αξιολόγηση αλγορίθμων

Η επιλογή του καταλληλότερου αλγορίθμου για την πρόβλεψη κινδύνου δασικής πυρκαγιάς δεν αποτελεί μονοδιάστατη απόφαση, αλλά εξαρτάται από ένα σύνολο παραμέτρων που σχετίζονται τόσο με τα δεδομένα όσο και με τις απαιτήσεις της εφαρμογής. Μεταξύ των βασικών κριτηρίων συγκαταλέγονται:

- Ο όγκος και η ποιότητα των διαθέσιμων δεδομένων
- Η ανάγκη ερμηνευσιμότητας των αποτελεσμάτων

- Η υπολογιστική πολυπλοκότητα και οι διαθέσιμοι πόροι
- Η απαίτηση για χωρική και χρονική ακρίβεια

Ο όγκος και η ποιότητα των δεδομένων επηρεάζουν άμεσα την καταλληλότητα ενός μοντέλου. Αλγόριθμοι όπως τα νευρωνικά δίκτυα αποδίδουν καλύτερα σε μεγάλα σύνολα δεδομένων, ενώ μέθοδοι όπως ο Random Forest μπορούν να λειτουργήσουν αποτελεσματικά ακόμη και σε μέτριου μεγέθους δείγματα. Παράλληλα, η ύπαρξη ελλιπών ή θορυβωδών δεδομένων καθιστά αναγκαία την επιλογή μοντέλων που εμφανίζουν ανθεκτικότητα σε τέτοιου είδους προβλήματα.

Η ανάγκη ερμηνευσιμότητας αποτελεί επίσης κρίσιμο παράγοντα, ιδίως σε εφαρμογές που σχετίζονται με τη δημόσια ασφάλεια και τη λήψη αποφάσεων από αρμόδιες αρχές. Μοντέλα που επιτρέπουν την αξιολόγηση της σημασίας των χαρακτηριστικών παρέχουν μεγαλύτερη διαφάνεια και ενισχύουν την εμπιστοσύνη στα αποτελέσματα.

Η υπολογιστική πολυπλοκότητα και οι διαθέσιμοι πόροι επηρεάζουν τη δυνατότητα εφαρμογής πιο απαιτητικών αλγορίθμων, ιδιαίτερα σε περιβάλλοντα όπου απαιτείται γρήγορη επεξεργασία ή ενσωμάτωση σε διαδικτυακή εφαρμογή. Τέλος, η απαίτηση για υψηλή χωρική ακρίβεια καθιστά αναγκαία τη χρήση μοντέλων που μπορούν να αποτυπώσουν λεπτές διαφοροποιήσεις μεταξύ γεωγραφικών περιοχών.

Στο πλαίσιο της παρούσας εργασίας υιοθετείται συγκριτική προσέγγιση, με αξιολόγηση πολλαπλών μοντέλων μηχανικής μάθησης. Η επιλογή του τελικού μοντέλου βασίζεται σε αντικειμενικές μετρικές αξιολόγησης, όπως Accuracy, Precision, Recall και F1-score, οι οποίες επιτρέπουν την ποσοτική αποτίμηση της απόδοσης. Ιδιαίτερη έμφαση δίνεται όχι μόνο στη μέγιστη ακρίβεια πρόβλεψης, αλλά και στην ισορροπία μεταξύ απόδοσης και ερμηνευσιμότητας, ώστε το τελικό σύστημα να είναι τόσο αποτελεσματικό όσο και πρακτικά αξιοποιήσιμο.

2.6 Κενά στη βιβλιογραφία και συνεισφορά της παρούσας εργασίας

Η βιβλιογραφική ανασκόπηση που πραγματοποιήθηκε στο πλαίσιο της παρούσας εργασίας ανέδειξε σημαντική πρόοδο στην εφαρμογή τεχνικών μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών. Ωστόσο, παράλληλα εντοπίστηκαν ορισμένα κενά και περιορισμοί που η υφιστάμενη έρευνα δεν έχει αντιμετωπίσει επαρκώς, και τα οποία η παρούσα εργασία επιχειρεί να καλύψει.

Πρώτον, η πλειονότητα των υφιστάμενων μελετών εστιάζει στην ανάπτυξη και αξιολόγηση μεμονωμένων αλγορίθμων πρόβλεψης, χωρίς να προχωρά στην ενσωμάτωσή τους σε λειτουργικό πληροφοριακό σύστημα. Τα ερευνητικά αποτελέσματα παραμένουν σε θεωρητικό ή πειραματικό επίπεδο, χωρίς να μεταφράζονται σε πρακτικά εργαλεία που μπορούν να αξιοποιηθούν από φορείς πολιτικής προστασίας, δασικής διαχείρισης ή τοπικής αυτοδιοίκησης. Η απουσία ολοκληρωμένων συστημάτων που να συνδυάζουν προγνωστικό μοντέλο, διαδραστική οπτικοποίηση και φιλική προς τον χρήστη διεπαφή αποτελεί σημαντικό κενό στη μετάβαση από την έρευνα στην εφαρμογή.

Δεύτερον, ελάχιστες εργασίες υιοθετούν πολυδιάστατη προσέγγιση, συνδυάζοντας ταυτόχρονα μετεωρολογικά, γεωχωρικά και δορυφορικά δεδομένα σε ένα ενιαίο προγνωστικό πλαίσιο. Η τυπική ερευνητική πρακτική περιορίζεται συχνά σε μία κατηγορία δεδομένων, συνήθως μετεωρολογικών, παραβλέποντας τη χωρική διάσταση του κινδύνου και τη συνεισφορά δεδομένων όπως ο δείκτης βλάστησης NDVI, το υψόμετρο ή η κλίση εδάφους. Η μονοδιάστατη αυτή προσέγγιση αποδυναμώνει τη χωρική ακρίβεια και τη γενικευσιμότητα των μοντέλων, ιδίως σε περιοχές με έντονη γεωμορφολογική ποικιλομορφία.

Τρίτον, η εφαρμογή τεχνικών μηχανικής μάθησης σε ημίξηρα περιβάλλοντα, όπως αυτό της Αριζόνα, παραμένει σχετικά ανεξερεύνητη. Η πλειονότητα των μελετών επικεντρώνεται σε εύκρατα ή μεσογειακά οικοσυστήματα, όπου οι κλιματικές συνθήκες και η δυναμική της βλάστησης διαφέρουν σημαντικά. Οι ιδιαίτερες χαρακτηριστικά της Αριζόνα, όπως οι εξαιρετικά υψηλές θερμοκρασίες, οι παρατεταμένες περίοδοι ξηρασίας, η χαμηλή σχετική υγρασία και η ποικιλία τύπων βλάστησης, δημιουργούν ένα μοναδικό περιβάλλον που απαιτεί εξειδικευμένη μοντελοποίηση και τοπική βαθμονόμηση των αλγορίθμων.

Τέταρτον, η υφιστάμενη βιβλιογραφία δεν δίνει επαρκή έμφαση στη δυνατότητα επανεκπαίδευσης και δυναμικής προσαρμογής των μοντέλων σε μεταβαλλόμενες κλιματικές συνθήκες. Τα περισσότερα συστήματα αντιμετωπίζονται ως στατικά εργαλεία, χωρίς πρόβλεψη για ενσωμάτωση νέων δεδομένων ή αναπροσαρμογή των παραμέτρων τους με την πάροδο του χρόνου.

Η παρούσα εργασία επιχειρεί να ανταποκριθεί στα παραπάνω κενά μέσω τεσσάρων βασικών συνεισφορών. Πρώτον, μέσω της συγκριτικής αξιολόγησης πολλαπλών αλγορίθμων μηχανικής μάθησης με πολυδιάστατα δεδομένα, ειδικά προσαρμοσμένα στις συνθήκες της Αριζόνα. Δεύτερον, μέσω της ενσωμάτωσης του βέλτιστου μοντέλου σε λειτουργική web

εφαρμογή, η οποία παρέχει διαδραστική χαρτογραφική απεικόνιση του επιπέδου κινδύνου. Τρίτον, μέσω του συνδυασμού μετεωρολογικών, γεωχωρικών και δορυφορικών δεδομένων σε ένα ενιαίο αλγοριθμικό πλαίσιο, ενισχύοντας τη χωρική ακρίβεια της πρόβλεψης. Τέταρτον, μέσω της διαμόρφωσης επεκτάσιμης αρχιτεκτονικής που επιτρέπει μελλοντική επανεκπαίδευση του μοντέλου και προσαρμογή σε νέα δεδομένα, διασφαλίζοντας τη μακροχρόνια βιωσιμότητα του συστήματος.

Με τον τρόπο αυτό, η παρούσα εργασία δεν περιορίζεται σε αμιγώς θεωρητική συνεισφορά, αλλά επιδιώκει τη γεφύρωση του χάσματος μεταξύ ακαδημαϊκής έρευνας και πρακτικής εφαρμογής στον τομέα της πρόβλεψης δασικών πυρκαγιών.

4. ΚΕΦΑΛΑΙΟ 3 – Συλλογή και ανάλυση δεδομένων

3.1 Ερευνητικός σχεδιασμός

Η παρούσα εργασία υιοθέτησε ποσοτική ερευνητική προσέγγιση, βασισμένη στην ανάλυση περιβαλλοντικών δεδομένων με τη χρήση τεχνικών μηχανικής μάθησης. Η επιλογή ποσοτικής μεθοδολογίας κρίθηκε κατάλληλη, καθώς το υπό μελέτη φαινόμενο επιδέχεται ποσοτικοποίηση μέσω μετρήσιμων μεταβλητών, όπως μετεωρολογικές παράμετροι, γεωχωρικά χαρακτηριστικά και ιστορικά στοιχεία πυρκαγιών.

Κεντρικός στόχος της έρευνας ήταν η ανάπτυξη προγνωστικού μοντέλου ταξινόμησης κινδύνου πυρκαγιάς, ικανού να κατηγοριοποιεί γεωγραφικές περιοχές σε διαφορετικά επίπεδα επικινδυνότητας. Η προσέγγιση αυτή επιτρέπει τη μετάβαση από την περιγραφική ανάλυση στην προγνωστική μοντελοποίηση, με έμφαση στη δυνατότητα εφαρμογής σε πραγματικό επιχειρησιακό περιβάλλον.

Το πρόβλημα διατυπώθηκε ως πρόβλημα επιβλεπόμενης μάθησης (supervised learning), καθώς το μοντέλο εκπαιδεύεται σε σύνολο δεδομένων όπου είναι γνωστή η μεταβλητή-στόχος. Στο συγκεκριμένο πλαίσιο:

- Τα χαρακτηριστικά (features) περιλαμβάνουν μετεωρολογικές και γεωχωρικές μεταβλητές: θερμοκρασία, σχετική υγρασία, ταχύτητα ανέμου, βροχόπτωση, υψόμετρο, κλίση εδάφους, δείκτη βλάστησης NDVI, δείκτη ξηρασίας, δείκτη FWI και μήνα.
- Η μεταβλητή-στόχος (target) είναι δυαδική και αντιστοιχεί στην εμφάνιση ή μη πυρκαγιάς (fire_occurrence: 0/1). Τα διακριτά επίπεδα κινδύνου (χαμηλό, μέτριο, υψηλό) δεν αποτελούν ξεχωριστή κλάση εκπαίδευσης, αλλά προκύπτουν εκ των υστέρων από κατωφλίωση της προβλεπόμενης πιθανότητας του μοντέλου (βλ. Κεφάλαιο 5).

Η συνολική μεθοδολογική ροή της έρευνας ακολούθησε δομημένη και διαδοχική διαδικασία, η οποία περιλάμβανε τα ακόλουθα στάδια:

- Συγκρότηση και ενοποίηση του συνόλου δεδομένων, με στόχο τη δημιουργία συνεκτικού και αξιοποιήσιμου συνόλου.

- Προεπεξεργασία και εξισορρόπηση, ώστε να αντιμετωπιστεί η ανισορροπία κλάσεων και να προετοιμαστούν τα δεδομένα για τους αλγορίθμους.
- Επιλογή και αξιολόγηση χαρακτηριστικών (feature engineering), με σκοπό τη βελτίωση της προγνωστικής ικανότητας.
- Εκπαίδευση μοντέλων, μέσω διαφορετικών αλγοριθμικών προσεγγίσεων.
- Αξιολόγηση απόδοσης, με χρήση κατάλληλων μετρικών και διαδικασιών επικύρωσης.
- Ενσωμάτωση στο web σύστημα, ώστε το τελικό μοντέλο να αξιοποιηθεί σε λειτουργικό περιβάλλον με διαδραστική απεικόνιση αποτελεσμάτων.

Ο ερευνητικός σχεδιασμός επιδίωξε να συνδυάσει θεωρητική τεκμηρίωση, μεθοδολογική αυστηρότητα και πρακτική εφαρμογή, διασφαλίζοντας ότι το τελικό αποτέλεσμα δεν θα περιοριζόταν σε αλγοριθμική σύγκριση, αλλά θα ενσωματωνόταν σε ολοκληρωμένο πληροφοριακό σύστημα.

3.2 Συλλογή δεδομένων

Η εκπαίδευση ενός αξιόπιστου προγνωστικού μοντέλου απαιτεί ένα ενοποιημένο σύνολο δεδομένων, το οποίο συνδυάζει ιστορικά στοιχεία πυρκαγιών, μετεωρολογικές παρατηρήσεις, γεωχωρικά χαρακτηριστικά και δορυφορικά δεδομένα, με συνεπή χωρική και χρονική αντιστοίχιση (spatio-temporal matching). Η συγκρότηση ενός τέτοιου πραγματικού, ενοποιημένου συνόλου για την Πολιτεία της Αριζόνα — με αξιόπιστη αντιστοίχιση μεταξύ ετερογενών και ανεξάρτητων πηγών — δεν ήταν εφικτή εντός του χρονικού πλαισίου και των διαθέσιμων πόρων της παρούσας εργασίας. Για τον λόγο αυτό, κατασκευάστηκε συνθετικό (synthetic) σύνολο δεδομένων, το οποίο παρήχθη προγραμματιστικά και βαθμονομήθηκε ώστε να αναπαράγει τις σχέσεις μεταξύ περιβαλλοντικών μεταβλητών και κινδύνου πυρκαγιάς, όπως αυτές τεκμηριώνονται στη διεθνή βιβλιογραφία (Κεφάλαιο 2).

Οι τύποι πηγών που τεκμηριώθηκαν στη βιβλιογραφία καθόρισαν ποιες μεταβλητές μοντελοποιήθηκαν και με ποιον τρόπο σχετίζονται με τη μεταβλητή-στόχο:

- Ιστορικά αρχεία δασικών πυρκαγιών (ημερομηνία, τοποθεσία, καμένη έκταση): όρισαν τη μορφή της δυαδικής μεταβλητής-στόχου (εμφάνιση/μη εμφάνιση) και το ρεαλιστικό ποσοστό εμφάνισης πυρκαγιάς.

- Μετεωρολογικές μεταβλητές: θερμοκρασία αέρα, σχετική υγρασία, ταχύτητα ανέμου και ύψος βροχόπτωσης.
- Γεωχωρικά χαρακτηριστικά: υψόμετρο, κλίση εδάφους και τύπος βλάστησης.
- Δορυφορικά δεδομένα: δείκτης βλάστησης NDVI, ως ποσοτικός δείκτης κατάστασης της καύσιμης ύλης.

Το συνθετικό σύνολο περιλαμβάνει 5.000 εγγραφές με δέκα μεταβλητές εισόδου. Οι μεταβλητές παρήχθησαν από στατιστικές κατανομές προσαρμοσμένες στα κλιματικά χαρακτηριστικά της Αριζόνα (ενδεικτικά, η θερμοκρασία ακολούθησε κανονική κατανομή με μέση τιμή περίπου 32°C και η σχετική υγρασία μέση τιμή περίπου 35%). Ο δείκτης FWI συνδέθηκε ισχυρά με τον δείκτη ξηρασίας, αναπαράγοντας τη γνωστή συνάφειά τους. Η μεταβλητή-στόχος προέκυψε πιθανοτικά, μέσω λογιστικής συνάρτησης σύνδεσης (logistic link), στην οποία τα βάρη κάθε μεταβλητής αντανακλούν τη σχετική σημασία που της αποδίδει η βιβλιογραφία: έντονη θετική συνεισφορά του FWI, του δείκτη ξηρασίας και της θερμοκρασίας, και αρνητική συνεισφορά της σχετικής υγρασίας και της βροχόπτωσης. Παράλληλα, ενσωματώθηκε ισχυρό εποχικό φαινόμενο, με αυξημένη πιθανότητα κατά τους θερινούς μήνες (Μάιος–Σεπτέμβριος) και κορύφωση τον Ιούνιο.

Το προκύπτον ποσοστό εμφάνισης πυρκαγιάς ανήλθε σε περίπου 49,9%, διαμορφώνοντας ένα σχεδόν ισορροπημένο σύνολο δεδομένων ως προς τις δύο κλάσεις (εμφάνιση / μη εμφάνιση πυρκαγιάς). Για λόγους αναπαραγωγιμότητας των αποτελεσμάτων, η διαδικασία παραγωγής χρησιμοποίησε σταθερό σπόρο τυχαιότητας (random seed).

Επισημαίνεται ρητά ότι το σύνολο δεδομένων είναι συνθετικό και δεν υποκαθιστά πραγματικά δεδομένα πεδίου. Σχεδιάστηκε ως ελεγχόμενο περιβάλλον για την ανάπτυξη, τον έλεγχο και την επίδειξη ολόκληρης της μεθοδολογικής και τεχνικής αλυσίδας — από την προεπεξεργασία και την εκπαίδευση των μοντέλων έως την ενσωμάτωση στο web σύστημα. Ο περιορισμός αυτός, καθώς και η μελλοντική αντικατάσταση του συνθετικού συνόλου με πραγματικά ιστορικά και δορυφορικά δεδομένα, συζητούνται αναλυτικά στο Κεφάλαιο 7.

3.3 Προεπεξεργασία και καθαρισμός δεδομένων

Επειδή το σύνολο δεδομένων παρήχθη με ελεγχόμενο τρόπο, δεν παρουσίαζε ελλιπείς τιμές, σφάλματα μέτρησης ή ασυνεπείς εγγραφές. Συνεπώς, δεν απαιτήθηκε στάδιο συμπλήρωσης ελλιπών τιμών (imputation), απομάκρυνσης ακραίων τιμών (outliers) ή διόρθωσης

ασυνεπειών. Σε σενάριο αξιοποίησης πραγματικών δεδομένων πεδίου, τα στάδια αυτά θα ήταν απαραίτητα και θα περιλάμβαναν τεχνικές imputation, στατιστική ή οπτική ανίχνευση ακραίων τιμών και κωδικοποίηση κατηγορικών μεταβλητών.

Η ουσιαστική προεπεξεργασία που εφαρμόστηκε αφορούσε την προετοιμασία των δεδομένων για την εκπαίδευση και περιλάμβανε τα εξής:

- Διαχωρισμό σε σύνολα εκπαίδευσης και ελέγχου με στρωματοποίηση (stratified split, 70% / 30%), ώστε να διατηρηθεί η αναλογία κλάσεων (περίπου 49,9% πυρκαγιές) και στα δύο υποσύνολα και να είναι αντιπροσωπευτική η αξιολόγηση.
- Εξισορρόπηση κλάσεων μέσω της τεχνικής SMOTE (Synthetic Minority Over-sampling Technique), η οποία δημιουργεί συνθετικά δείγματα της μειονοτικής κλάσης. Η SMOTE εφαρμόστηκε αποκλειστικά στο σύνολο εκπαίδευσης και ποτέ στο σύνολο ελέγχου, ώστε να αποφευχθεί η διαρροή πληροφορίας (data leakage) και να παραμείνει η αξιολόγηση αμερόληπτη στην πραγματική κατανομή του συνόλου ελέγχου (Chawla et al., 2002).
- Τυποποίηση των μεταβλητών (standardization) με χρήση StandardScaler. Η κλιμάκωση ήταν κρίσιμη ιδίως για το νευρωνικό δίκτυο MLP, το οποίο είναι ευαίσθητο στην κλίμακα των τιμών εισόδου, σε αντίθεση με τα μοντέλα που βασίζονται σε δέντρα απόφασης (Random Forest, Gradient Boosting, XGBoost) και είναι αμετάβλητα ως προς μονοτονικούς μετασχηματισμούς.

Ιδιαίτερη προσοχή δόθηκε στην ισορροπία των κλάσεων (class balance) κατά την εκπαίδευση. Αν και το συνθετικό σύνολο δεδομένων ήταν σχεδόν ισορροπημένο, εφαρμόστηκε η τεχνική SMOTE αποκλειστικά στο σύνολο εκπαίδευσης, ώστε να εξασφαλιστεί πλήρης ισορρόπηση των δύο κλάσεων και να αποφευχθεί οποιαδήποτε μεροληψία του μοντέλου προς τη συχνότερη κλάση. Ο συνδυασμός στρωματοποιημένου διαχωρισμού και εφαρμογής SMOTE μόνο στο σύνολο εκπαίδευσης διαμόρφωσε ένα πλήρως ισορροπημένο σύνολο μάθησης, διατηρώντας παράλληλα την πραγματική κατανομή κατά την αξιολόγηση.

3.4 Επιλογή και εξαγωγή χαρακτηριστικών (Feature engineering)

Η διαδικασία feature engineering αποσκοπεί στη βελτίωση της προγνωστικής ικανότητας του μοντέλου μέσω της δημιουργίας, του μετασχηματισμού ή του συνδυασμού μεταβλητών, με

τρόπο που αναδεικνύει πιο αποτελεσματικά τα υποκείμενα πρότυπα των δεδομένων. Η ποιότητα και η αντιπροσωπευτικότητα των χαρακτηριστικών επηρεάζουν καθοριστικά την απόδοση των αλγορίθμων μηχανικής μάθησης.

Στο πλαίσιο της παρούσας μελέτης, το σύνολο δεδομένων ενσωμάτωσε εξαρχής σύνθετους δείκτες που λειτουργούν ως χαρακτηριστικά υψηλής πληροφοριακής αξίας:

- Τον δείκτη FWI (Fire Weather Index), ως σύνθετο μετεωρολογικό δείκτη που συμπυκνώνει την επίδραση πολλαπλών καιρικών παραγόντων.
- Τον δείκτη ξηρασίας (drought index), ως ένδειξη της μακροχρόνιας κατάστασης ξηρότητας της περιοχής.
- Τη μεταβλητή του μήνα, η οποία αποτυπώνει την έντονη εποχικότητα του φαινομένου.

Η σημασία κάθε χαρακτηριστικού αξιολογήθηκε συστηματικά, προκειμένου να εντοπιστούν οι μεταβλητές που συμβάλλουν περισσότερο στη διαμόρφωση της πρόβλεψης. Για τον σκοπό αυτό αξιοποιήθηκαν:

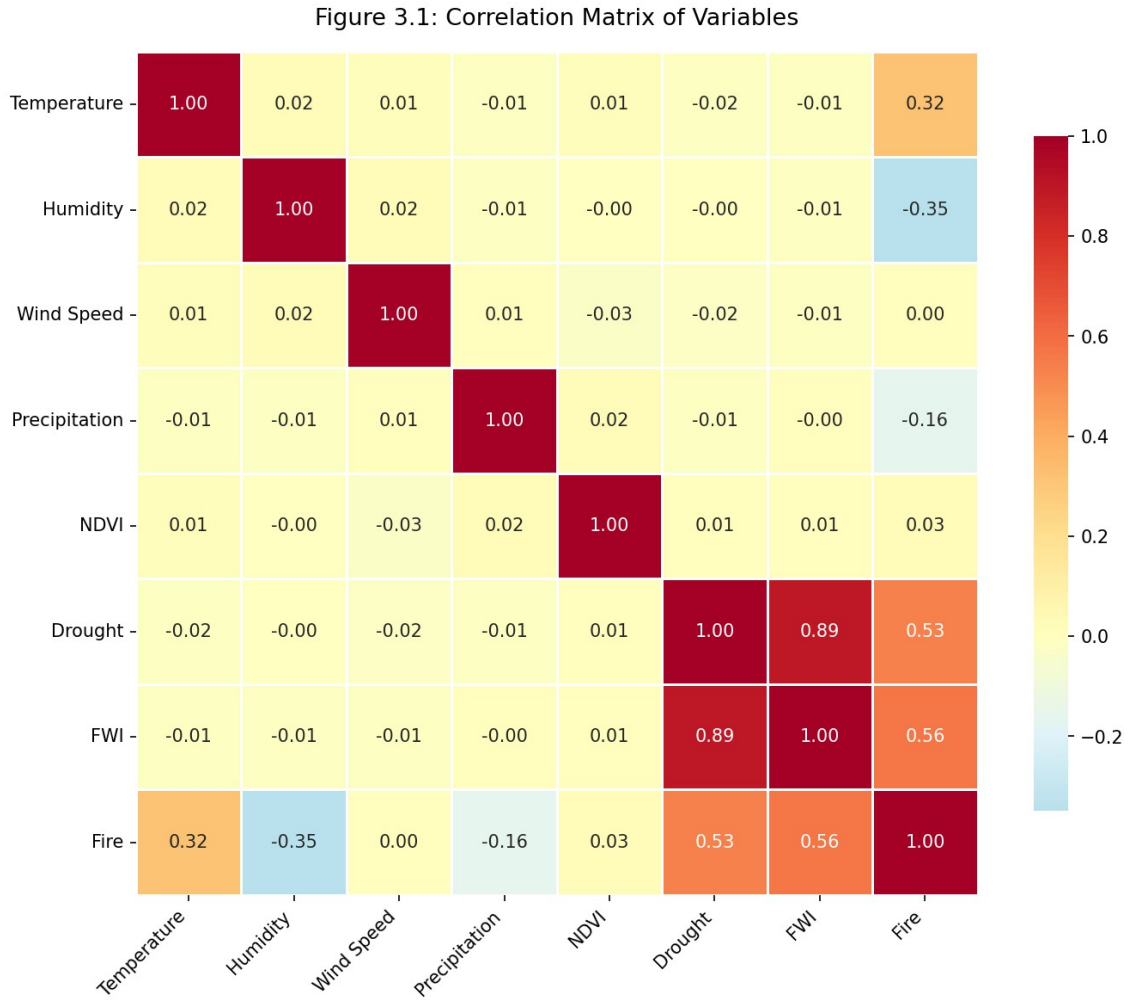
- Ανάλυση συσχετίσεων (correlation analysis), για την ανίχνευση ισχυρών γραμμικών σχέσεων και τον εντοπισμό πλεονασμού μεταξύ μεταβλητών.
- Feature importance, μέσω του αλγορίθμου Random Forest, που επιτρέπει την ποσοτική αποτίμηση της συμβολής κάθε μεταβλητής.
- SHAP values (SHapley Additive exPlanations), που παρέχουν λεπτομερή ανάλυση της επίδρασης κάθε χαρακτηριστικού στην τελική πρόβλεψη, ενισχύοντας την ερμηνευσιμότητα του μοντέλου (Lundberg & Lee, 2017).

Λόγω της σχετικά χαμηλής διαστασιμότητας του προβλήματος (δέκα χαρακτηριστικά), δεν κρίθηκε αναγκαία η εφαρμογή τεχνικών μείωσης διαστάσεων, όπως η Principal Component Analysis (PCA). Η ισχυρή συσχέτιση που παρατηρήθηκε μεταξύ του δείκτη FWI και του δείκτη ξηρασίας αντιμετωπίστηκε ερμηνευτικά, κατά την ανάλυση της σημασίας των χαρακτηριστικών, και ελήφθη υπόψη στην αξιολόγηση της συνεισφοράς των δύο μεταβλητών. Με τον τρόπο αυτό, το στάδιο feature engineering δεν αντιμετωπίστηκε ως απλή τεχνική προσαρμογή, αλλά ως ουσιώδες στάδιο που συνδέει τη φυσική κατανόηση του φαινομένου με την αλγοριθμική του αναπαράσταση.

3.4.1 Διερευνητική ανάλυση δεδομένων

Πριν από την εκπαίδευση των μοντέλων, πραγματοποιήθηκε διερευνητική ανάλυση δεδομένων (Exploratory Data Analysis, EDA) με στόχο την κατανόηση των σχέσεων μεταξύ των μεταβλητών και τον εντοπισμό προτύπων που σχετίζονται με την εμφάνιση πυρκαγιών. Η ανάλυση αυτή επέτρεψε τη διαπίστωση ισχυρών συσχετίσεων, τον εντοπισμό πλεονασμού μεταξύ μεταβλητών και την επαλήθευση της λογικής συνέπειας του συνθετικού συνόλου δεδομένων πριν από τη χρήση του σε αλγοριθμικά μοντέλα.

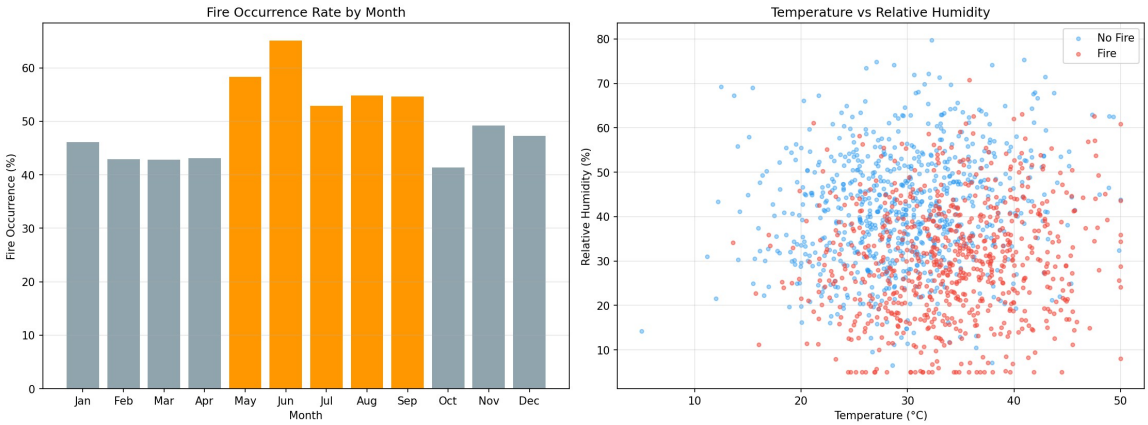
Στο Σχήμα 3.1 παρουσιάζεται ο πίνακας συσχετίσεων (correlation matrix) μεταξύ των βασικών αριθμητικών μεταβλητών του συνόλου δεδομένων και της μεταβλητής-στόχου (εμφάνιση πυρκαγιάς). Από την ανάλυση προκύπτει ότι οι μεταβλητές με την ισχυρότερη θετική συσχέτιση με την εμφάνιση πυρκαγιάς είναι ο δείκτης FWI (r=0,56), ο δείκτης ξηρασίας (r=0,53) και η θερμοκρασία (r=0,32). Αντίστροφα, η σχετική υγρασία (r=-0,35) και η βροχόπτωση (r=-0,16) εμφανίζουν αρνητική συσχέτιση, επιβεβαιώνοντας τη θεωρητικά αναμενόμενη σχέση μεταξύ ξηρών συνθηκών και αυξημένου κινδύνου. Αξίζει να σημειωθεί η υψηλή συσχέτιση μεταξύ FWI και δείκτη ξηρασίας (r=0,89), η οποία υποδηλώνει πιθανό πλεονασμό μεταξύ των δύο μεταβλητών και ελήφθη υπόψη κατά την ανάλυση της σημασίας των χαρακτηριστικών.



Πίνακας 1 - Πίνακας συσχετίσεων μεταξύ των βασικών μεταβλητών και της εμφάνισης πυρκαγιάς.

Στο Σχήμα 3.2 παρουσιάζεται η εποχική κατανομή των πυρκαγιών, καθώς και η σχέση μεταξύ θερμοκρασίας και σχετικής υγρασίας ως προς την εμφάνιση πυρκαγιάς. Η ανάλυση επιβεβαιώνει την ύπαρξη έντονης εποχικότητας, με τον Ιούνιο να παρουσιάζει το υψηλότερο ποσοστό εμφάνισης πυρκαγιών (περίπου 65%), ακολουθούμενο από τους υπόλοιπους θερινούς μήνες Μάιο, Αύγουστο και Σεπτέμβριο (περίπου 55–58%) και τον Ιούλιο (περίπου 53%). Η σχετική πτώση του ποσοστού τον Ιούλιο συνδέεται με την περίοδο των μουσώνων (monsoon season) στην Αριζόνα, κατά την οποία οι βροχοπτώσεις αυξάνονται. Το διάγραμμα διασποράς θερμοκρασίας–υγρασίας αναδεικνύει ότι οι πυρκαγιές συγκεντρώνονται κυρίως στην περιοχή υψηλών θερμοκρασιών (>25°C) και χαμηλής σχετικής υγρασίας (<25%), επιβεβαιώνοντας τη σημασία των δύο αυτών μεταβλητών στη διαμόρφωση του κινδύνου.

Figure 3.2: Seasonality and Risk Distribution



Εικόνα 1 - Αριστερά: Ποσοστό εμφάνισης πυρκαγιών ανά μήνα (πορτοκαλί: fire season, Μάιος-Σεπτέμβριος). Δεξιά: Διάγραμμα διασποράς θερμοκρασίας και σχετικής υγρασίας με χρωματική κωδικοποίηση κινδύνου πυρκαγιάς

Τα ευρήματα της διερευνητικής ανάλυσης ενισχύουν τη θεωρητική τεκμηρίωση του Κεφαλαίου 2 και επιβεβαιώνουν ότι οι μεταβλητές που επιλέχθηκαν ως χαρακτηριστικά εισόδου παρουσιάζουν ουσιαστική σχέση με τη μεταβλητή-στόχο. Παράλληλα, η ανάδειξη της εποχικότητας και των μη γραμμικών αλληλεπιδράσεων μεταξύ μεταβλητών υποστηρίζει την επιλογή αλγορίθμων μηχανικής μάθησης ικανών να αποτυπώσουν τέτοιου είδους σχέσεις, όπως ο Random Forest, το Gradient Boosting και το XGBoost.

3.5 Εκπαίδευση μοντέλων

Για την εκπαίδευση και την αξιολόγηση των προγνωστικών μοντέλων, το σύνολο δεδομένων διαχωρίστηκε σε δύο υποσύνολα με στρωματοποίηση: σύνολο εκπαίδευσης (70%, 3.500 εγγραφές) και σύνολο ελέγχου (30%, 1.500 εγγραφές). Το σύνολο ελέγχου διατηρήθηκε ανεξάρτητο και χρησιμοποιήθηκε αποκλειστικά για την τελική αξιολόγηση, ώστε να εκτιμηθεί η ικανότητα γενίκευσης του μοντέλου σε νέα, μη ορατά δεδομένα.

Πέραν του διαχωρισμού, εφαρμόστηκε διαδικασία k-fold cross-validation με στρωματοποίηση (StratifiedKFold, k=5). Το σύνολο εκπαίδευσης χωρίστηκε σε πέντε ίσα τμήματα· κάθε μοντέλο εκπαιδεύτηκε επαναληπτικά σε τέσσερα τμήματα και αξιολογήθηκε στο εναπομείναν, με τη διαδικασία να επαναλαμβάνεται πέντε φορές. Η προσέγγιση αυτή μείωσε την πιθανότητα υπερπροσαρμογής (overfitting) και προσέφερε σταθερότερη εκτίμηση της απόδοσης.

Αξιολογήθηκαν τέσσερα μοντέλα μηχανικής μάθησης:

- Random Forest, ως αντιπροσωπευτική μέθοδος ensemble με έμφαση στη σταθερότητα και την ερμηνευσιμότητα.
- Gradient Boosting, ως μέθοδος σταδιακής ενίσχυσης ασθενών μαθητών.
- XGBoost, ως βελτιστοποιημένη υλοποίηση boosting με υψηλή προγνωστική ακρίβεια.
- Multilayer Perceptron (MLP), νευρωνικό δίκτυο με αρχιτεκτονική τριών κρυφών επιπέδων (128, 64, 32 νευρώνες), συνάρτηση ενεργοποίησης ReLU, βελτιστοποιητή Adam, μέγιστο αριθμό εποχών 400 και πρόωρη διακοπή (early stopping).

Για τα τρία πρώτα μοντέλα πραγματοποιήθηκε ρύθμιση υπερπαραμέτρων (hyperparameter tuning) μέσω GridSearchCV, με μετρική βελτιστοποίησης το F1-score και πενταπλή διασταυρωμένη επικύρωση. Συγκεκριμένα, εξετάστηκαν:

- Random Forest: αριθμός δέντρων {100, 200}, μέγιστο βάθος {10, 15}, ελάχιστος αριθμός δειγμάτων ανά διάσπαση {2, 5}.
- Gradient Boosting: αριθμός εκτιμητών {100, 200}, μέγιστο βάθος {3, 5}, ρυθμός μάθησης {0,05, 0,1}.
- XGBoost: αριθμός εκτιμητών {100, 200}, μέγιστο βάθος {3, 6}, ρυθμός μάθησης {0,05, 0,1}.

Η συνδυαστική εφαρμογή στρωματοποιημένου διαχωρισμού, εξισορρόπησης με SMOTE, διασταυρωμένης επικύρωσης και υπερπαραμετρικής βελτιστοποίησης διασφάλισε ότι τα τελικά μοντέλα δεν θα εμφάνιζαν απλώς υψηλή απόδοση στο σύνολο εκπαίδευσης, αλλά θα παρουσίαζαν ικανοποιητική ικανότητα γενίκευσης.

3.6 Μετρικές αξιολόγησης

Η αξιολόγηση της απόδοσης των μοντέλων δεν βασίστηκε αποκλειστικά στην ακρίβεια (accuracy), καθώς η μετρική αυτή από μόνη της μπορεί να οδηγήσει σε παραπλανητικά συμπεράσματα. Στην περίπτωση πρόβλεψης δασικών πυρκαγιών, η ορθή ταξινόμηση των πραγματικών περιπτώσεων υψηλού κινδύνου έχει ιδιαίτερη βαρύτητα, και επομένως ένα

μοντέλο πρέπει να αξιολογείται και ως προς την ικανότητά του να εντοπίζει αποτελεσματικά τις περιπτώσεις πυρκαγιάς και όχι μόνο ως προς τη συνολική του ακρίβεια.

Για τον λόγο αυτό, η αξιολόγηση βασίστηκε σε ένα σύνολο συμπληρωματικών μετρικών:

- Precision, που εκφράζει το ποσοστό των ορθά προβλεπόμενων θετικών περιπτώσεων επί του συνόλου όσων το μοντέλο προέβλεψε ως θετικές.
- Recall (ευαισθησία), που μετρά το ποσοστό των πραγματικών θετικών περιπτώσεων που εντοπίστηκαν επιτυχώς.
- F1-score, ο αρμονικός μέσος όρος Precision και Recall, που παρέχει ισορροπημένη αποτίμηση της απόδοσης (Powers, 2011).
- ROC-AUC, που αποτυπώνει τη διακριτική ικανότητα του μοντέλου σε διαφορετικά κατώφλια ταξινόμησης, ανεξάρτητα από συγκεκριμένο όριο απόφασης.

Η επιλογή του τελικού μοντέλου δεν βασίστηκε αποκλειστικά στη μέγιστη τιμή μίας μετρικής, αλλά στη βέλτιστη ισορροπία μεταξύ Recall και Precision. Στο πλαίσιο πρόβλεψης κινδύνου πυρκαγιάς, το Recall αποκτά ιδιαίτερη σημασία, καθώς η αποφυγή ψευδώς αρνητικών προβλέψεων (περιπτώσεων υψηλού κινδύνου που δεν εντοπίζονται) είναι κρίσιμη για λόγους δημόσιας ασφάλειας. Παράλληλα, η διατήρηση ικανοποιητικού Precision είναι απαραίτητη ώστε να αποφεύγονται υπερβολικές ψευδείς ειδοποιήσεις που θα μείωναν την αξιοπιστία του συστήματος.

3.7 Ενσωμάτωση στο Web-based σύστημα

Μετά την εκπαίδευση και την αξιολόγηση, τα μοντέλα αποθηκεύτηκαν και ενσωματώθηκαν στο backend της εφαρμογής, ώστε να καταστούν λειτουργικό μέρος του πληροφοριακού συστήματος. Το στάδιο αυτό σηματοδότησε τη μετάβαση από το ερευνητικό περιβάλλον ανάπτυξης στο περιβάλλον πραγματικής χρήσης.

Η διαδικασία ενσωμάτωσης περιλάμβανε:

- Αποθήκευση των εκπαιδευμένων μοντέλων και του μετασχηματιστή κλιμάκωσης (scaler) σε σειριοποιημένη μορφή μέσω της βιβλιοθήκης joblib, ώστε να διατηρείται η εκπαιδευμένη μορφή τους και να φορτώνονται χωρίς επανεκπαίδευση.

- Φόρτωση και των τεσσάρων μοντέλων στη μνήμη κατά την εκκίνηση της εφαρμογής και συνδυασμό των εκτιμήσεών τους μέσω προσέγγισης συνόλου (soft-voting ensemble) για την παραγωγή της τελικής πιθανότητας κινδύνου.
- Δημιουργία σημείων πρόσβασης REST API, τα οποία δέχονται δεδομένα εισόδου (μετεωρολογικά και γεωχωρικά χαρακτηριστικά) και επιστρέφουν την αντίστοιχη πρόβλεψη κινδύνου σε μορφή JSON.
- Δυνατότητα πρόβλεψης σε πραγματικό χρόνο και χαρτογραφικής απεικόνισης των αποτελεσμάτων μέσω διαδραστικού χάρτη.

Η αρχιτεκτονική σχεδιάστηκε με γνώμονα την επεκτασιμότητα και τη βιωσιμότητα, προβλέποντας τη δυνατότητα μελλοντικής επανεκπαίδευσης (retraining) με νέα δεδομένα και αντικατάστασης των αποθηκευμένων μοντέλων χωρίς διακοπή της λειτουργίας του συστήματος. Με τον τρόπο αυτό, η μεθοδολογία δεν περιορίστηκε στη θεωρητική ανάπτυξη μοντέλου, αλλά ολοκληρώθηκε σε ένα λειτουργικό web-based σύστημα, ικανό να υποστηρίξει πρακτικές ανάγκες ανάλυσης και λήψης αποφάσεων. Η αναλυτική περιγραφή της υλοποίησης παρουσιάζεται στο Κεφάλαιο 5.

5. ΚΕΦΑΛΑΙΟ 4 – Σχεδιασμός συστήματος

4.1 Αρχιτεκτονική της εφαρμογής

Η αρχιτεκτονική του προτεινόμενου συστήματος βασίζεται σε μοντέλο τριών επιπέδων (three-tier architecture), το οποίο διαχωρίζει τη λογική παρουσίασης, την επιχειρησιακή λογική και τη διαχείριση δεδομένων σε ανεξάρτητα αλλά αλληλοσυνδεόμενα στρώματα. Η προσέγγιση αυτή διασφαλίζει τη σαφή οργάνωση του κώδικα, τη δυνατότητα ανεξάρτητης ανάπτυξης και συντήρησης κάθε επιπέδου, καθώς και την επεκτασιμότητα του συστήματος σε μελλοντικές φάσεις. Η αρχιτεκτονική τριών επιπέδων αποτελεί ευρέως αποδεκτή πρακτική στην ανάπτυξη διαδικτυακών εφαρμογών, καθώς επιτρέπει τον σαφή διαχωρισμό ευθυνών (separation of concerns) και διευκολύνει τη δοκιμή, τη συντήρηση και τη μελλοντική αναβάθμιση μεμονωμένων συστατικών χωρίς να επηρεάζεται η λειτουργικότητα του συνόλου.

Το πρώτο επίπεδο (Presentation Layer) αφορά τη διεπαφή χρήστη και υλοποιείται μέσω web περιβάλλοντος με χρήση HTML, CSS και JavaScript. Κεντρικό στοιχείο της διεπαφής αποτελεί ο διαδραστικός χάρτης, ο οποίος αναπτύσσεται με τη βιβλιοθήκη Leaflet.js. Η επιλογή του Leaflet.js βασίζεται στο ότι πρόκειται για μια ανοιχτού κώδικα, ελαφριά και ευέλικτη βιβλιοθήκη χαρτογράφησης, η οποία υποστηρίζει πλήθος λειτουργιών όπως layers, markers, heatmaps και αναδυόμενα παράθυρα πληροφοριών (popups). Ο χρήστης μπορεί να πλοηγείται στον χάρτη, να κάνει zoom σε συγκεκριμένες περιοχές της Αριζόνα, να επιλέγει γεωγραφικά σημεία ενδιαφέροντος και να λαμβάνει οπτική πληροφόρηση σχετικά με το εκτιμώμενο επίπεδο κινδύνου πυρκαγιάς. Η χρωματική κωδικοποίηση χρησιμοποιείται για την κατηγοριοποίηση του κινδύνου σε επίπεδα (χαμηλό, μέτριο, υψηλό, κρίσιμο), ενώ τα αναδυόμενα παράθυρα εμφανίζουν αναλυτικές πληροφορίες, όπως τις τιμές των μετεωρολογικών μεταβλητών και το ποσοστό πιθανότητας κινδύνου. Επιπλέον, η διεπαφή περιλαμβάνει φόρμες εισαγωγής δεδομένων, μέσω των οποίων ο χρήστης μπορεί να καταχωρεί μετεωρολογικές τιμές και να λαμβάνει ad hoc προβλέψεις για συγκεκριμένη τοποθεσία και χρονική στιγμή.

Το δεύτερο επίπεδο (Application Layer) περιλαμβάνει την επιχειρησιακή λογική της εφαρμογής και υλοποιείται μέσω του πλαισίου Django σε γλώσσα Python. Η επιλογή του Django υπαγορεύεται τόσο από την ωριμότητα και τη σταθερότητα του πλαισίου, όσο και από τη φυσική συμβατότητά του με τις βιβλιοθήκες μηχανικής μάθησης της Python, όπως

scikit-learn, XGBoost και TensorFlow. Το επίπεδο αυτό είναι υπεύθυνο για τη διαχείριση αιτημάτων HTTP από τη διεπαφή χρήστη, την κλήση του προγνωστικού μοντέλου, την επεξεργασία και μετασχηματισμό των δεδομένων εισόδου, καθώς και τη διαμόρφωση των αποτελεσμάτων σε μορφή κατάλληλη για απεικόνιση στον χάρτη. Η αρχιτεκτονική Model-View-Template (MVT) του Django εξασφαλίζει τον σαφή διαχωρισμό μεταξύ λογικής δεδομένων (Models), λογικής επεξεργασίας αιτημάτων (Views) και λογικής παρουσίασης (Templates). Το εκπαιδευμένο μοντέλο μηχανικής μάθησης ενσωματώνεται ως αποσειριοποιημένο αντικείμενο (deserialized object) στο backend, φορτώνεται κατά την εκκίνηση της εφαρμογής και εκτελεί προβλέψεις σε πραγματικό χρόνο χωρίς ανάγκη εξωτερικής υπηρεσίας ή κλήσης σε τρίτο σύστημα. Η προσέγγιση αυτή ελαχιστοποιεί τον χρόνο απόκρισης και απλοποιεί την αρχιτεκτονική ανάπτυξης (deployment).

Το τρίτο επίπεδο (Data Layer) αφορά την αποθήκευση και διαχείριση των δεδομένων και υλοποιείται μέσω του συστήματος διαχείρισης σχεσιακών βάσεων δεδομένων PostgreSQL, σε συνδυασμό με την επέκταση PostGIS. Η επιλογή PostgreSQL οφείλεται στη σταθερότητα, την υψηλή απόδοση, την ευρεία υποστήριξη από το Django μέσω του ORM και τη δυνατότητα διαχείρισης μεγάλου όγκου δεδομένων με αξιοπιστία. Η επέκταση PostGIS προσθέτει δυνατότητες γεωχωρικής ανάλυσης, επιτρέποντας την αποθήκευση γεωμετρικών αντικειμένων (σημεία, πολύγωνα, γραμμές), την εκτέλεση χωρικών ερωτημάτων (spatial queries) όπως η αναζήτηση εγγραφών εντός συγκεκριμένης γεωγραφικής ακτίνας, και τη σύνδεση δεδομένων με γεωγραφικές συντεταγμένες. Η δυνατότητα αυτή είναι κρίσιμη για ένα σύστημα πρόβλεψης που βασίζεται σε χωρική ανάλυση κινδύνου, καθώς επιτρέπει τον αποδοτικό συνδυασμό μετεωρολογικών παρατηρήσεων με γεωχωρικά χαρακτηριστικά σε επίπεδο συγκεκριμένης τοποθεσίας.

Η επικοινωνία μεταξύ των επιπέδων πραγματοποιείται μέσω RESTful API endpoints, τα οποία δέχονται αιτήματα HTTP από τη διεπαφή χρήστη και επιστρέφουν αποτελέσματα σε μορφή JSON. Η χρήση REST αρχιτεκτονικής επιτρέπει τη σαφή τυποποίηση της επικοινωνίας, τη δυνατότητα χρήσης του API από διαφορετικούς clients (web browser, κινητή εφαρμογή, εξωτερικό σύστημα) και τη μελλοντική επέκταση του συστήματος. Συγκεκριμένα, προβλέπεται η δυνατότητα σύνδεσης με εξωτερικές πηγές δεδομένων σε πραγματικό χρόνο, όπως μετεωρολογικά API ή δορυφορικές υπηρεσίες, καθώς και η δημιουργία κινητής εφαρμογής που θα αξιοποιεί τα ίδια API endpoints για την παροχή ειδοποιήσεων κινδύνου σε πραγματικό χρόνο.

4.2 Σχεδίαση βάσης δεδομένων

Η βάση δεδομένων του συστήματος σχεδιάζεται με γνώμονα την αποδοτική αποθήκευση, ανάκτηση και σύνδεση μετεωρολογικών, γεωχωρικών και ιστορικών δεδομένων πυρκαγιών. Η δομή ακολουθεί σχεσιακό μοντέλο και υλοποιείται μέσω του Django ORM, το οποίο επιτρέπει τον ορισμό μοντέλων δεδομένων σε Python και τη μετάφρασή τους σε πίνακες της βάσης PostgreSQL. Η χρήση του ORM διευκολύνει τη διαχείριση του σχήματος της βάσης μέσω migrations, τη δημιουργία ερωτημάτων χωρίς άμεση χρήση SQL και τη διασφάλιση της συμβατότητας μεταξύ κώδικα εφαρμογής και δομής δεδομένων.

Οι βασικές οντότητες (entities) της βάσης δεδομένων περιλαμβάνουν τα εξής:

- Πίνακας WeatherData: Αποθηκεύει μετεωρολογικές μετρήσεις ανά τοποθεσία και ημερομηνία. Τα πεδία περιλαμβάνουν θερμοκρασία αέρα (°C), σχετική υγρασία (%), ταχύτητα ανέμου (km/h), κατεύθυνση ανέμου (μοίρες), ύψος βροχόπτωσης (mm) και ημερομηνία μέτρησης. Κάθε εγγραφή συνδέεται με γεωγραφικές συντεταγμένες μέσω πεδίου PointField, αξιοποιώντας τις δυνατότητες του PostGIS. Η χωρική αποθήκευση επιτρέπει τη σύνδεση κάθε μέτρησης με συγκεκριμένο γεωγραφικό σημείο, διευκολύνοντας τόσο τη χωρική ανάλυση όσο και την αντιστοίχιση με δεδομένα πυρκαγιών. Επιπλέον, αποθηκεύονται υπολογισμένοι δείκτες, όπως ο Fire Weather Index (FWI) και ο κινητός μέσος θερμοκρασίας (rolling average), οι οποίοι αξιοποιούνται ως features στο μοντέλο πρόβλεψης.

- Πίνακας FireIncident: Περιέχει ιστορικά στοιχεία πυρκαγιών στην Αριζόνα, όπως τη γεωγραφική θέση εκδήλωσης (PointField), την ημερομηνία έναρξης και λήξης, την εκτιμώμενη καμένη έκταση σε στρέμματα, την αιτία ανάφλεξης (φυσική ή ανθρωπογενής) και πρόσθετα χαρακτηριστικά που σχετίζονται με τη σοβαρότητα του συμβάντος. Τα δεδομένα αυτά χρησιμεύουν τόσο στην εκπαίδευση του μοντέλου, καθώς ορίζουν τη μεταβλητή-στόχο (εκδήλωση ή μη πυρκαγιάς), όσο και στην αξιολόγησή του. Η χρονική και χωρική αντιστοίχιση με τον πίνακα WeatherData επιτρέπει τη σύνδεση κάθε πυρκαγιάς με τις μετεωρολογικές συνθήκες που επικρατούσαν κατά τη στιγμή εκδήλωσής της.

- Πίνακας GeoFeature: Αποθηκεύει γεωχωρικά χαρακτηριστικά ανά γεωγραφική περιοχή, τα οποία παραμένουν σχετικά σταθερά στον χρόνο. Τα πεδία περιλαμβάνουν υψόμετρο (m), κλίση εδάφους (%), προσανατολισμό εδάφους (μοίρες), τύπο βλάστησης (κατηγορική μεταβλητή), πυκνότητα βλάστησης και τιμές δείκτη NDVI. Κάθε εγγραφή αντιστοιχεί σε ένα γεωγραφικό κελί (grid cell) ή πολύγωνο, επιτρέποντας τη χωρική διακριτοποίηση της περιοχής μελέτης. Τα πεδία αυτά ενσωματώνονται ως χαρακτηριστικά εισόδου (features) στο μοντέλο πρόβλεψης, εμπλουτίζοντας τη μετεωρολογική πληροφορία με τοπογραφικά και οικολογικά δεδομένα.

- Πίνακας Prediction: Καταγράφει τα αποτελέσματα πρόβλεψης του μοντέλου, συμπεριλαμβανομένων του εκτιμώμενου επιπέδου κινδύνου (χαμηλό, μέτριο, υψηλό, κρίσιμο), του αριθμητικού score πιθανότητας, του ονόματος και της έκδοσης του μοντέλου που χρησιμοποιήθηκε, της ημερομηνίας πρόβλεψης και της αντίστοιχης γεωγραφικής θέσης. Ο πίνακας αυτός επιτρέπει τόσο τη χαρτογραφική απεικόνιση των αποτελεσμάτων στον διαδραστικό χάρτη όσο και τη διαχρονική παρακολούθηση και σύγκριση των προβλέψεων. Η αποθήκευση της έκδοσης του μοντέλου καθιστά δυνατή την αξιολόγηση της απόδοσης διαφορετικών εκδόσεων σε βάθος χρόνου.

Οι σχέσεις μεταξύ των πινάκων υλοποιούνται μέσω ξένων κλειδιών (Foreign Keys) και χωρικών συνδέσεων. Συγκεκριμένα, κάθε εγγραφή στον πίνακα Prediction συνδέεται με τις αντίστοιχες εγγραφές WeatherData και GeoFeature που χρησιμοποιήθηκαν ως είσοδοι, εξασφαλίζοντας πλήρη ιχνηλασιμότητα κάθε πρόβλεψης. Η χρήση χωρικών ευρετηρίων (spatial indexes) στα PointField πεδία βελτιστοποιεί σημαντικά την ταχύτητα αναζήτησης σε γεωχωρικά ερωτήματα, ιδιαίτερα σε λειτουργίες όπως η εύρεση εγγραφών εντός συγκεκριμένης γεωγραφικής ακτίνας ή η σύνδεση σημείων πυρκαγιάς με τον πλησιέστερο μετεωρολογικό σταθμό. Η σχεδίαση της βάσης ακολουθεί τις αρχές κανονικοποίησης (τρίτη κανονική μορφή), ώστε να αποφεύγεται η πλεονασματικότητα δεδομένων και να διασφαλίζεται η ακεραιότητα και η συνέπεια των πληροφοριών.

4.3 Επιλογή εκπαιδευτικού μοντέλου πρόβλεψης

Η επιλογή του κατάλληλου μοντέλου μηχανικής μάθησης αποτελεί κεντρικό στοιχείο του σχεδιασμού του συστήματος, καθώς από αυτήν εξαρτάται τόσο η ακρίβεια των προβλέψεων όσο και η πρακτική λειτουργικότητα της εφαρμογής. Όπως αναλύθηκε στο Κεφάλαιο 2, τρεις βασικές κατηγορίες αλγορίθμων εξετάζονται στο πλαίσιο της παρούσας εργασίας: Random Forest, Gradient Boosting (XGBoost) και τεχνητά νευρωνικά δίκτυα (Multilayer Perceptron). Κάθε αλγόριθμος παρουσιάζει διαφορετικά πλεονεκτήματα και περιορισμούς, γεγονός που καθιστά αναγκαία τη συστηματική σύγκρισή τους υπό κοινές πειραματικές συνθήκες.

Η διαδικασία επιλογής του τελικού μοντέλου ακολουθεί δομημένη συγκριτική μεθοδολογία. Αρχικά, κάθε αλγόριθμος εκπαιδεύεται στο ίδιο σύνολο δεδομένων, με τις ίδιες μεταβλητές εισόδου και τον ίδιο διαχωρισμό εκπαίδευσης-αξιολόγησης (70%-30%), ώστε η σύγκριση να είναι δίκαιη και αντικειμενική. Στη συνέχεια, πραγματοποιείται ρύθμιση υπερπαραμέτρων (hyperparameter tuning) μέσω Grid Search ή Random Search σε συνδυασμό με k-fold cross-validation (k=5), ώστε κάθε μοντέλο να αξιολογηθεί στη βέλτιστη δυνατή διαμόρφωσή του. Ο συνδυασμός αυτός ελαχιστοποιεί τον κίνδυνο υπερπροσαρμογής και παρέχει σταθερές εκτιμήσεις απόδοσης.

Για τον Random Forest, οι βασικές υπερπαράμετροι που εξετάζονται περιλαμβάνουν τον αριθμό δέντρων (n_estimators), το μέγιστο βάθος κάθε δέντρου (max_depth), τον ελάχιστο αριθμό δειγμάτων ανά φύλλο (min_samples_leaf) και τον αριθμό χαρακτηριστικών που εξετάζονται σε κάθε διάσπαση (max_features). Για το XGBoost, ρυθμίζονται ο ρυθμός μάθησης (learning_rate), ο αριθμός εκτιμητών (n_estimators), το μέγιστο βάθος δέντρου (max_depth), ο συντελεστής κανονικοποίησης (reg_alpha, reg_lambda) και η αναλογία δειγμάτων ανά δέντρο (subsample). Για το MLP, εξετάζονται ο αριθμός και το μέγεθος των κρυφών επιπέδων (hidden_layer_sizes), η συνάρτηση ενεργοποίησης (activation), ο ρυθμός μάθησης (learning_rate_init), ο αριθμός εποχών εκπαίδευσης (max_iter) και η τεχνική κανονικοποίησης (alpha).

Η τελική επιλογή βασίζεται σε πολλαπλά κριτήρια, αναγνωρίζοντας ότι η βέλτιστη επιλογή δεν εξαρτάται αποκλειστικά από μία μετρική. Η προγνωστική απόδοση αποτιμάται μέσω μετρικών όπως Accuracy, Precision, Recall, F1-score και ROC-AUC, με ιδιαίτερη βαρύτητα στο Recall. Η έμφαση στο Recall δικαιολογείται από το γεγονός ότι στο πλαίσιο πρόβλεψης κινδύνου πυρκαγιάς, η αποτυχία εντοπισμού μιας πραγματικής περίπτωσης υψηλού κινδύνου (ψευδώς αρνητική πρόβλεψη) ενέχει σοβαρότερες συνέπειες για τη δημόσια ασφάλεια σε

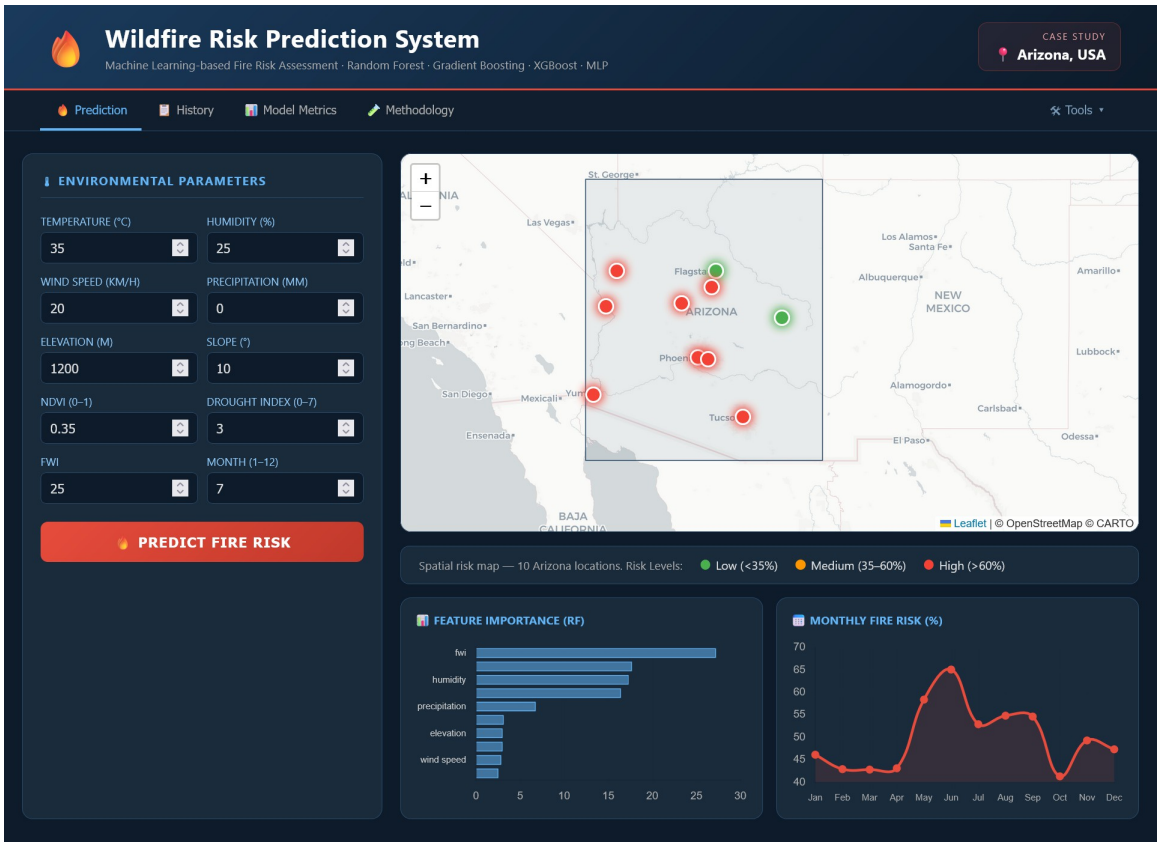
σχέση με μια ψευδώς θετική ειδοποίηση. Παράλληλα, η ερμηνευσιμότητα του μοντέλου λαμβάνεται σοβαρά υπόψη, καθώς σε εφαρμογές που σχετίζονται με τη λήψη αποφάσεων από δημόσιους φορείς είναι σημαντικό να κατανοείται ο τρόπος με τον οποίο κάθε μεταβλητή επηρεάζει την πρόβλεψη. Μοντέλα που υποστηρίζουν feature importance (όπως ο Random Forest και το XGBoost) ή τεχνικές ερμηνευσιμότητας όπως SHAP values παρέχουν πλεονέκτημα προς αυτήν την κατεύθυνση. Τέλος, η υπολογιστική αποδοτικότητα αξιολογείται σε σχέση με τις απαιτήσεις της web εφαρμογής, καθώς το μοντέλο πρέπει να παράγει προβλέψεις σε χρόνο απόκρισης κατάλληλο για διαδραστική χρήση, ιδανικά κάτω από ένα δευτερόλεπτο ανά αίτημα.

Μετά την ολοκλήρωση της συγκριτικής αξιολόγησης, το μοντέλο που παρουσιάζει τη βέλτιστη ισορροπία μεταξύ των παραπάνω κριτηρίων αποθηκεύεται σε μορφή serialized αρχείου μέσω της βιβλιοθήκης joblib, η οποία προσφέρει αποδοτική σειριοποίηση αντικειμένων Python, ιδιαίτερα αυτών που περιέχουν μεγάλα αριθμητικά arrays. Το αρχείο αυτό φορτώνεται κατά την εκκίνηση της Django εφαρμογής και παραμένει διαθέσιμο στη μνήμη για την εκτέλεση προβλέψεων. Η αρχιτεκτονική του συστήματος προβλέπει τη δυνατότητα αντικατάστασης του μοντέλου χωρίς τροποποίηση του υπόλοιπου κώδικα της εφαρμογής, μέσω ενός αρχείου ρυθμίσεων που ορίζει τη διαδρομή του τρέχοντος μοντέλου. Η προσέγγιση αυτή διευκολύνει τη μελλοντική επανεκπαίδευση με νέα δεδομένα, τη δοκιμή εναλλακτικών αλγορίθμων ή τη σταδιακή βελτίωση του μοντέλου χωρίς διακοπή λειτουργίας του συστήματος (hot-swapping).

6. ΚΕΦΑΛΑΙΟ 5 - Υλοποίηση Web-app

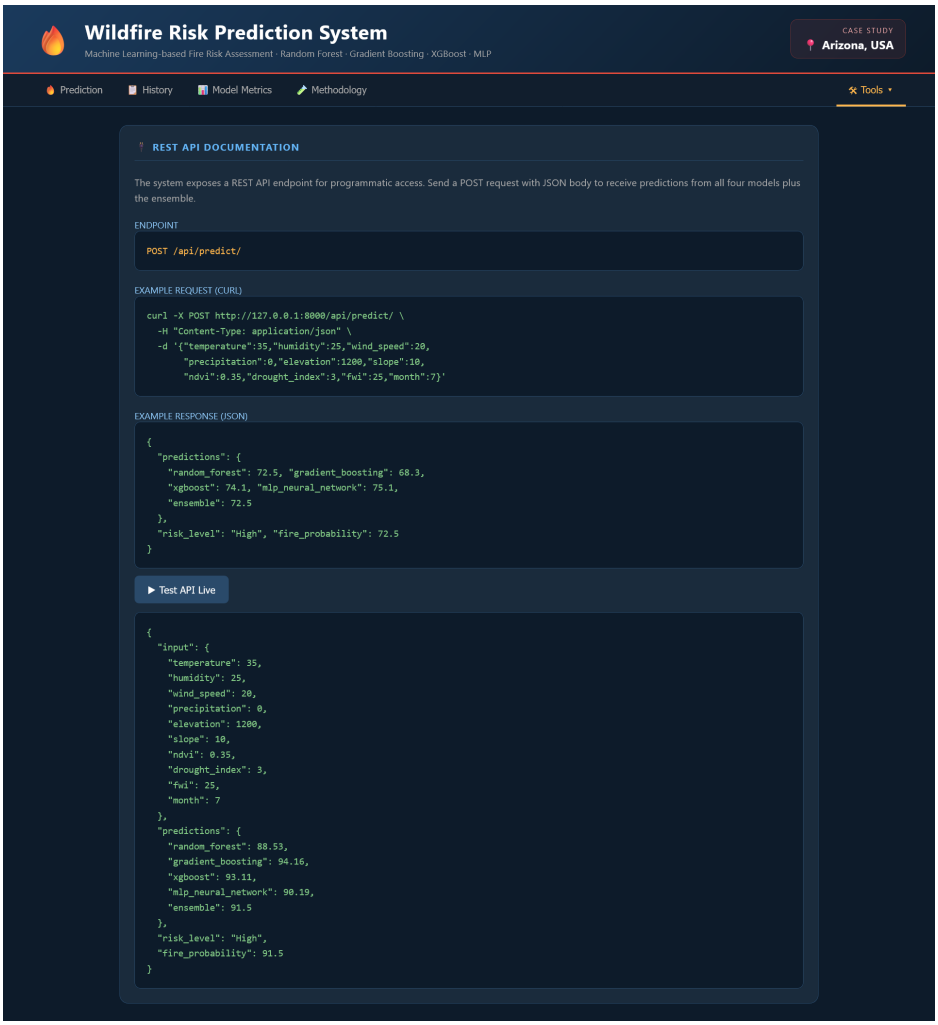
5.1 Ανάπτυξη backend με Django

Το backend του συστήματος υλοποιήθηκε με το πλαίσιο Django, ακολουθώντας την αρχιτεκτονική Model-View-Template (MVT) που περιγράφηκε στο Κεφάλαιο 4. Το έργο οργανώθηκε σε μια κύρια εφαρμογή (Django app) με όνομα prediction, η οποία ενσωματώνει τα μοντέλα δεδομένων, τη λογική των προβλέψεων και τα σημεία πρόσβασης (endpoints) του REST API. Η δομή του έργου διαχωρίζει με σαφήνεια τις ρυθμίσεις (settings), τη δρομολόγηση διευθύνσεων (URL routing), τα μοντέλα (models), τις όψεις (views) και τους serializers, εξασφαλίζοντας τη συντηρησιμότητα και την επεκτασιμότητα του κώδικα. Η διαμόρφωση των ρυθμίσεων υποστηρίζει τόσο περιβάλλον ανάπτυξης όσο και περιβάλλον παραγωγής, με δυνατότητα εναλλαγής της βάσης δεδομένων μέσω μεταβλητών περιβάλλοντος.



Εικόνα 2 - Wildfire Risk Prediction System

Τα μοντέλα δεδομένων (WeatherData, FireIncident, GeoFeature, Prediction) που σχεδιάστηκαν στο Κεφάλαιο 4 μεταφράστηκαν σε αντικείμενα του Django ORM και υλοποιήθηκαν στη βάση δεδομένων μέσω του μηχανισμού migrations. Η επικοινωνία μεταξύ της διεπαφής χρήστη και του backend πραγματοποιείται μέσω ενός RESTful API, υλοποιημένου με τη βιβλιοθήκη Django REST Framework. Το API εκθέτει τέσσερα βασικά σημεία πρόσβασης: ένα endpoint πρόβλεψης κινδύνου για μεμονωμένο σύνολο παραμέτρων, ένα endpoint μαζικής επεξεργασίας αρχείων CSV (batch prediction), ένα endpoint ανάκτησης του ιστορικού προβλέψεων και ένα endpoint επανεκπαίδευσης του μοντέλου. Κάθε αίτημα και απόκριση διαμορφώνεται σε μορφή JSON, διασφαλίζοντας τη διαλειτουργικότητα και την ευκολία ενσωμάτωσης με τη διαδραστική διεπαφή.



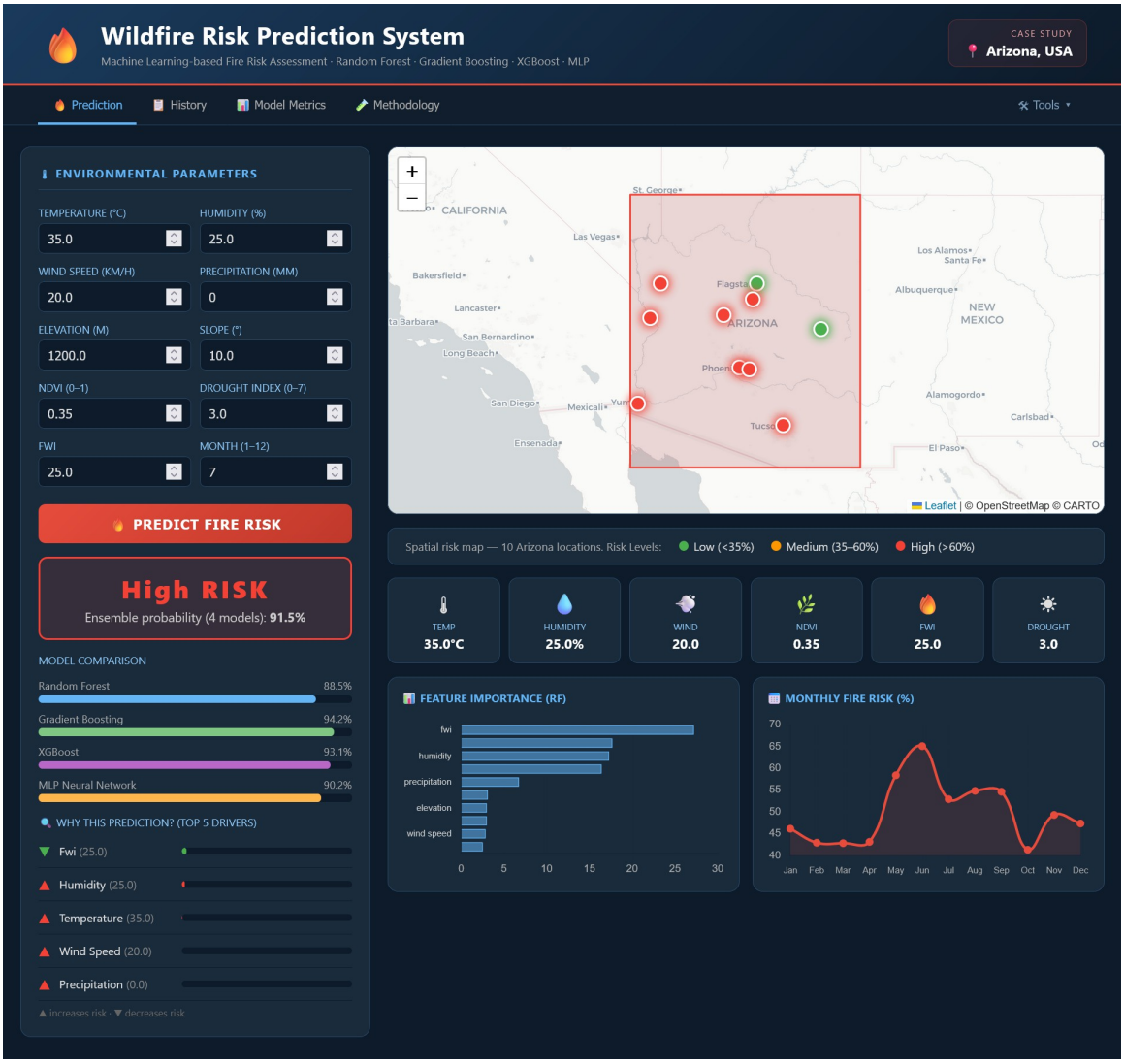
Εικόνα 3 - API

Η λογική της πρόβλεψης ενσωματώνεται στο επίπεδο των όψεων (views). Κατά την παραλαβή ενός αιτήματος, οι παράμετροι εισόδου (θερμοκρασία, σχετική υγρασία, ταχύτητα ανέμου, βροχόπτωση, υψόμετρο, κλίση, NDVI, δείκτης ξηρασίας, FWI και μήνας) επικυρώνονται ως προς τον τύπο και το εύρος τιμών τους, μετασχηματίζονται στη μορφή διανύσματος χαρακτηριστικών που αναμένει το μοντέλο και προωθούνται στη μηχανή πρόβλεψης. Το αποτέλεσμα, που περιλαμβάνει την πιθανότητα και το επίπεδο κινδύνου, αποθηκεύεται στον πίνακα Prediction και επιστρέφεται στη διεπαφή. Η προστασία των αιτημάτων εγγραφής επιτυγχάνεται μέσω του ενσωματωμένου μηχανισμού CSRF του Django, ενώ ο διαχωρισμός μεταξύ ασφαλών (GET) και μη ασφαλών (POST) μεθόδων ακολουθεί τις βέλτιστες πρακτικές σχεδίασης REST.

Για το περιβάλλον ανάπτυξης χρησιμοποιήθηκε η ελαφριά βάση δεδομένων SQLite, η οποία διευκολύνει την ταχεία δοκιμή και την τοπική εκτέλεση χωρίς πρόσθετες εξαρτήσεις. Η αρχιτεκτονική ωστόσο σχεδιάστηκε ώστε να υποστηρίζει τη μετάβαση σε PostgreSQL με την επέκταση PostGIS σε περιβάλλον παραγωγής, αξιοποιώντας τις γεωχωρικές δυνατότητες που απαιτούνται για τη διαχείριση των πεδίων PointField και τη χωρική ανάλυση. Ο διαχωρισμός της λογικής πρόβλεψης από το επίπεδο αποθήκευσης, σε συνδυασμό με τη μη απαίτηση εξωτερικών υπηρεσιών, καθιστά την ανάπτυξη (deployment) απλή και τον χρόνο απόκρισης του συστήματος ελάχιστο.

5.2 Ενσωμάτωση μοντέλου AI

Η εκπαίδευση των μοντέλων μηχανικής μάθησης πραγματοποιείται εκτός σύνδεσης (offline), σε ξεχωριστή διαδικασία από τη λειτουργία της εφαρμογής, ώστε να μην επιβαρύνεται ο χρόνος απόκρισης κατά την εκτέλεση των προβλέψεων. Τα τέσσερα μοντέλα που εκπαιδεύτηκαν (Random Forest, Gradient Boosting, XGBoost και Multilayer Perceptron) αποθηκεύονται ως σειριοποιημένα αντικείμενα μέσω της βιβλιοθήκης joblib και φορτώνονται στη μνήμη κατά την εκκίνηση της Django εφαρμογής. Με τον τρόπο αυτό, τα εκπαιδευμένα μοντέλα παραμένουν άμεσα διαθέσιμα για την εκτέλεση προβλέψεων σε πραγματικό χρόνο, χωρίς ανάγκη επανεκπαίδευσης ή κλήσης εξωτερικής υπηρεσίας.



Εικόνα 4 - System Results

Κατά τη φάση εκπαίδευσης, παράλληλα με τα μοντέλα αποθηκεύεται και ο μετασχηματιστής κλιμάκωσης (scaler) που εφαρμόζεται στις μεταβλητές εισόδου. Η κλιμάκωση των χαρακτηριστικών είναι κρίσιμη ιδίως για το νευρωνικό δίκτυο MLP, το οποίο είναι ευαίσθητο στην κλίμακα των τιμών εισόδου, σε αντίθεση με τα μοντέλα που βασίζονται σε δέντρα απόφασης (Random Forest, Gradient Boosting, XGBoost) τα οποία είναι αμετάβλητα ως προς μονοτονικούς μετασχηματισμούς. Κατά την εκτέλεση μιας πρόβλεψης, οι παράμετροι εισόδου μετασχηματίζονται με τον αποθηκευμένο scaler και τροφοδοτούνται σε καθένα από τα τέσσερα μοντέλα, καθένα από τα οποία παράγει μια ανεξάρτητη εκτίμηση πιθανότητας εκδήλωσης πυρκαγιάς.

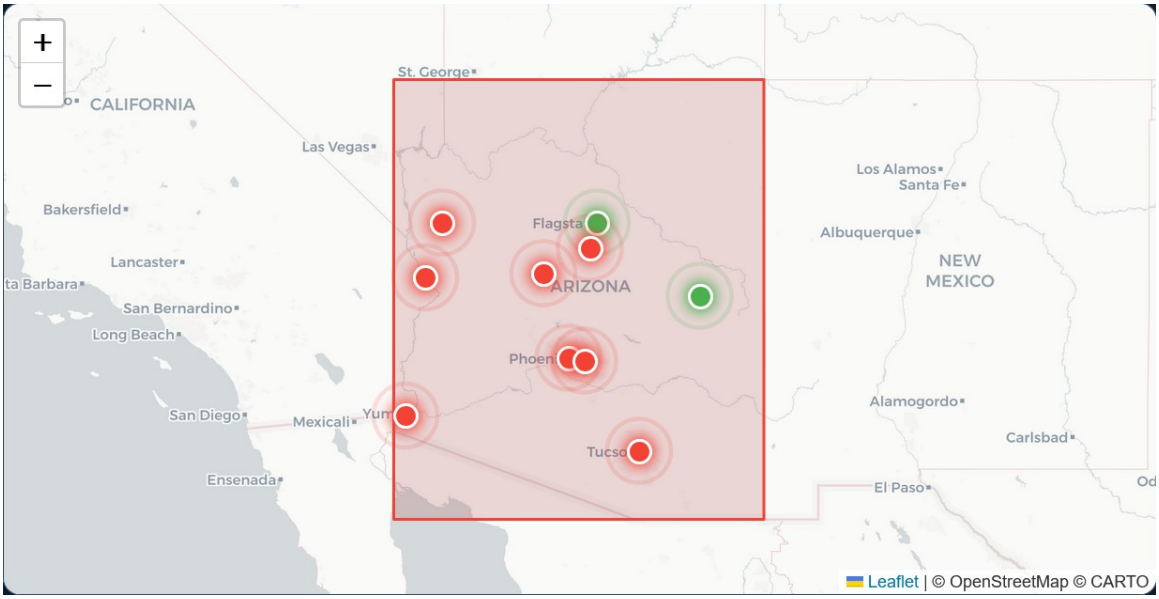
Πέραν της μεμονωμένης απόκρισης κάθε μοντέλου, το σύστημα υλοποιεί μια προσέγγιση συνόλου (ensemble), κατά την οποία η τελική πιθανότητα κινδύνου προκύπτει από τον

συνδυασμό των εκτιμήσεων και των τεσσάρων μοντέλων μέσω μηχανισμού soft voting (μέσος όρος πιθανοτήτων). Η επιλογή αυτή επεκτείνει τη μεθοδολογία επιλογής βέλτιστου μοντέλου που περιγράφηκε στο Κεφάλαιο 4, αξιοποιώντας τη συμπληρωματικότητα των διαφορετικών αλγορίθμων: ο συνδυασμός μοντέλων με διαφορετικές παραδοχές μειώνει τη διακύμανση και ενισχύει τη συνολική ευρωστία της πρόβλεψης, περιορίζοντας την επίδραση τυχόν αδυναμιών ενός μεμονωμένου μοντέλου. Η διεπαφή παρουσιάζει τόσο την τελική πιθανότητα του συνόλου όσο και τις επιμέρους εκτιμήσεις κάθε μοντέλου, ενισχύοντας τη διαφάνεια του αποτελέσματος.

Για την ενίσχυση της ερμηνευσιμότητας, το σύστημα συνοδεύει κάθε πρόβλεψη με ανάλυση των βασικών παραγόντων που τη διαμόρφωσαν, παρουσιάζοντας τις σημαντικότερες μεταβλητές και την κατεύθυνση της επίδρασής τους (αύξηση ή μείωση του κινδύνου). Η πληροφορία αυτή προέρχεται από τη σημασία των χαρακτηριστικών (feature importance) και επιτρέπει στον τελικό χρήστη να κατανοεί όχι μόνο το αποτέλεσμα αλλά και τους λόγους που το υποστηρίζουν. Τέλος, η αρχιτεκτονική προβλέπει δυνατότητα επανεκπαίδευσης των μοντέλων με νέα δεδομένα μέσω ειδικού endpoint, καθώς και αντικατάστασης των αποθηκευμένων αρχείων χωρίς διακοπή της λειτουργίας του συστήματος (hot-swapping), διευκολύνοντας τη σταδιακή βελτίωση και επικαιροποίηση του μοντέλου.

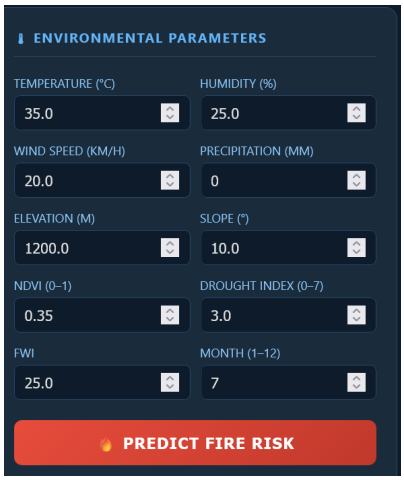
5.3 Οπτικοποίηση αποτελεσμάτων σε διαδραστικό χάρτη

Το επίπεδο παρουσίασης του συστήματος σχεδιάστηκε ώστε να μετατρέπει τα αριθμητικά αποτελέσματα της πρόβλεψης σε εποπτική και άμεσα κατανοητή πληροφορία. Κεντρικό στοιχείο της διεπαφής αποτελεί ένας διαδραστικός χάρτης, υλοποιημένος με τη βιβλιοθήκη Leaflet.js, ο οποίος απεικονίζει την περιοχή μελέτης της Αριζόνα και προβάλλει δέκα αντιπροσωπευτικές τοποθεσίες. Κάθε τοποθεσία αναπαρίσταται με χρωματικά κωδικοποιημένο στίγμα ανάλογα με το εκτιμώμενο επίπεδο κινδύνου: πράσινο για χαμηλό κίνδυνο (κάτω του 35%), πορτοκαλί για μέτριο (35-60%) και κόκκινο για υψηλό (άνω του 60%). Η χωρική αυτή απεικόνιση επιτρέπει στον χρήστη να αντιληφθεί με μια ματιά τη γεωγραφική κατανομή του κινδύνου σε ολόκληρη την περιοχή.



Εικόνα 5 - System - Arizona Map

Στο αριστερό τμήμα της διεπαφής τοποθετείται ο πίνακας εισαγωγής περιβαλλοντικών παραμέτρων, μέσω του οποίου ο χρήστης ορίζει τις τιμές των μεταβλητών εισόδου και εκτελεί την πρόβλεψη. Το αποτέλεσμα παρουσιάζεται με σαφή ένδειξη του επιπέδου κινδύνου και της αντίστοιχης πιθανότητας του συνόλου των μοντέλων. Ακολουθεί η συγκριτική παρουσίαση των επιμέρους εκτιμήσεων κάθε μοντέλου (model comparison), καθώς και η ανάλυση των πέντε σημαντικότερων παραγόντων που διαμόρφωσαν την πρόβλεψη, με σαφή ένδειξη αν κάθε παράγοντας αυξάνει ή μειώνει τον κίνδυνο. Η παρουσίαση αυτή ενισχύει τη διαφάνεια και τη δυνατότητα ερμηνείας των αποτελεσμάτων από τον τελικό χρήστη.



Εικόνα 6 - System enviromental parameters

Η διεπαφή οργανώνεται σε θεματικές καρτέλες (tabs) που διαχωρίζουν τις λειτουργίες του συστήματος: την καρτέλα πρόβλεψης (Prediction), το ιστορικό προβλέψεων (History), τις μετρικές αξιολόγησης των μοντέλων (Model Metrics) και την περιγραφή της μεθοδολογίας (Methodology). Συμπληρωματικά διαγράμματα, υλοποιημένα με τη βιβλιοθήκη Chart.js, απεικονίζουν τη σημασία των χαρακτηριστικών (feature importance) και τη μηνιαία διακύμανση του κινδύνου πυρκαγιάς, αναδεικνύοντας την έντονη εποχικότητα του φαινομένου. Τέλος, το σύστημα υποστηρίζει μαζική επεξεργασία μέσω μεταφόρτωσης αρχείου CSV, επιτρέποντας την ταυτόχρονη εκτίμηση κινδύνου για πολλαπλές τοποθεσίες. Στο Σχήμα 5.1 παρουσιάζεται στιγμιότυπο της διεπαφής του συστήματος σε λειτουργία.

5.4 Πρόσθετες λειτουργίες του συστήματος

Πέραν της βασικής λειτουργίας μεμονωμένης πρόβλεψης και της χαρτογραφικής απεικόνισης που περιγράφηκαν στις προηγούμενες ενότητες, το σύστημα ενσωματώνει ένα σύνολο συμπληρωματικών λειτουργιών, οι οποίες ενισχύουν τη χρηστικότητα, τη διαλειτουργικότητα και τη μακροχρόνια βιωσιμότητά του. Οι λειτουργίες αυτές είναι προσβάσιμες μέσω της διεπαφής χρήστη και υλοποιούνται από αντίστοιχα σημεία πρόσβασης (endpoints) του REST API.

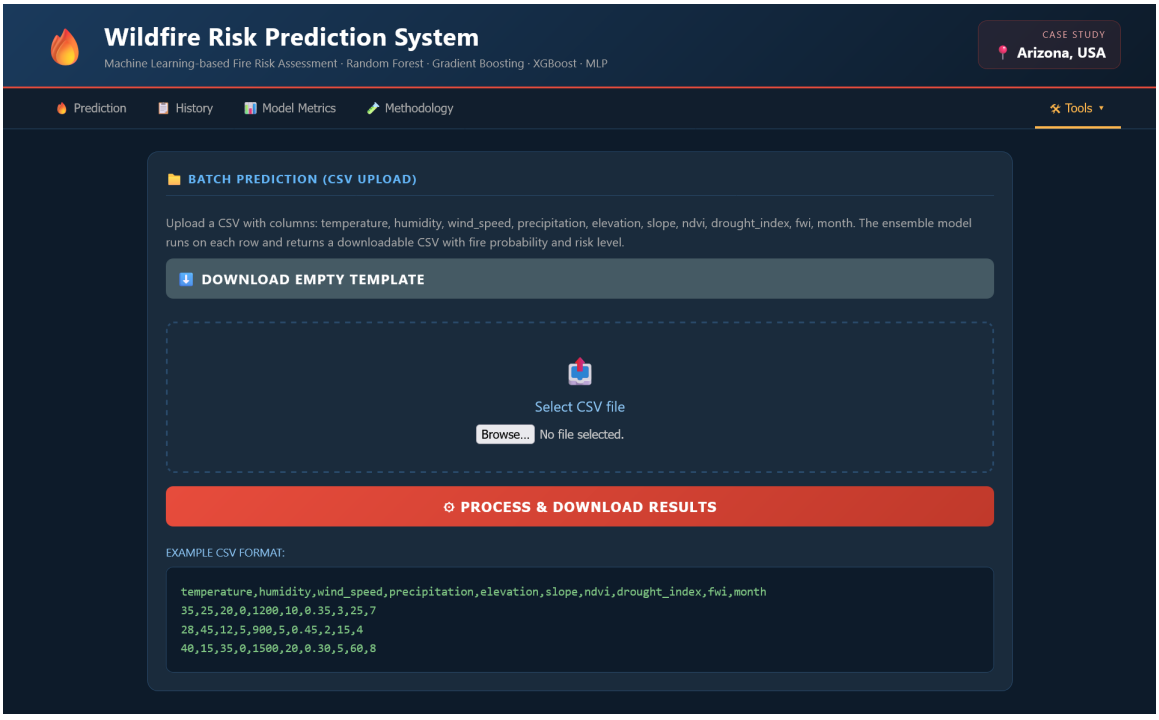
- **Προγραμματιστική πρόσβαση μέσω REST API.**

  Το σύστημα εκθέτει ένα ανοικτό σημείο πρόσβασης (api/predict/), το οποίο δέχεται αιτήματα τύπου POST με δεδομένα σε μορφή JSON και επιστρέφει την πρόβλεψη

κινδύνου ως δομημένη απόκριση JSON. Η απόκριση περιλαμβάνει τόσο τις επιμέρους εκτιμήσεις και των τεσσάρων μοντέλων (Random Forest, Gradient Boosting, XGBoost, MLP) όσο και την τελική πιθανότητα του συνόλου (ensemble) και το αντίστοιχο επίπεδο κινδύνου. Η λειτουργία αυτή επιτρέπει την ενσωμάτωση του συστήματος σε τρίτες εφαρμογές ή αυτοματοποιημένες ροές εργασίας, χωρίς εξάρτηση από τη γραφική διεπαφή.

- **Μαζική πρόβλεψη μέσω αρχείου (batch prediction).**

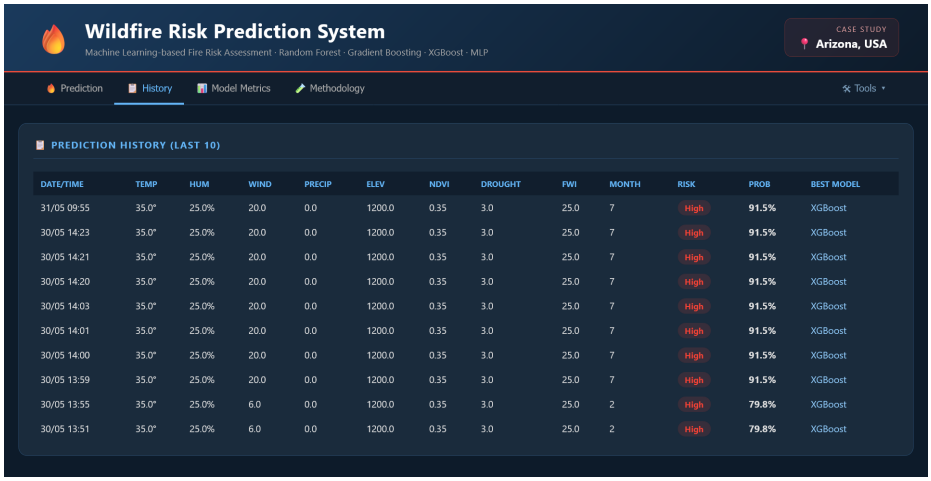
  Για την υποστήριξη σεναρίων όπου απαιτείται ταυτόχρονη εκτίμηση κινδύνου για πολλαπλές περιοχές ή χρονικές στιγμές, το σύστημα παρέχει λειτουργία μαζικής επεξεργασίας. Ο χρήστης μεταφορτώνει αρχείο μορφής CSV με πολλαπλές εγγραφές παραμέτρων, και το σύστημα επιστρέφει την πρόβλεψη για κάθε γραμμή. Για τη διευκόλυνση της σωστής δόμησης των δεδομένων εισόδου, διατίθεται έτοιμο πρότυπο αρχείο (template) με τις απαιτούμενες στήλες και ενδεικτική εγγραφή, περιορίζοντας τα σφάλματα μορφοποίησης.



Εικόνα 7 - System batch prediction

- **Ιστορικό προβλέψεων.**

Κάθε πρόβλεψη που εκτελείται μέσω της διεπαφής αποθηκεύεται αυτόματα στη βάση δεδομένων, μαζί με τις παραμέτρους εισόδου, το προκύπτον επίπεδο κινδύνου, την πιθανότητα και τη χρονική σήμανση. Το σύστημα εμφανίζει τις πλέον πρόσφατες καταχωρήσεις, επιτρέποντας στον χρήστη να παρακολουθεί και να συγκρίνει προηγούμενες εκτιμήσεις. Η λειτουργία αυτή προσδίδει διάσταση παρακολούθησης στον χρόνο και ενισχύει την πρακτική αξία του συστήματος ως εργαλείου υποστήριξης αποφάσεων.



Εικόνα 8 - System history

- **Παρουσίαση μετρικών των μοντέλων.**

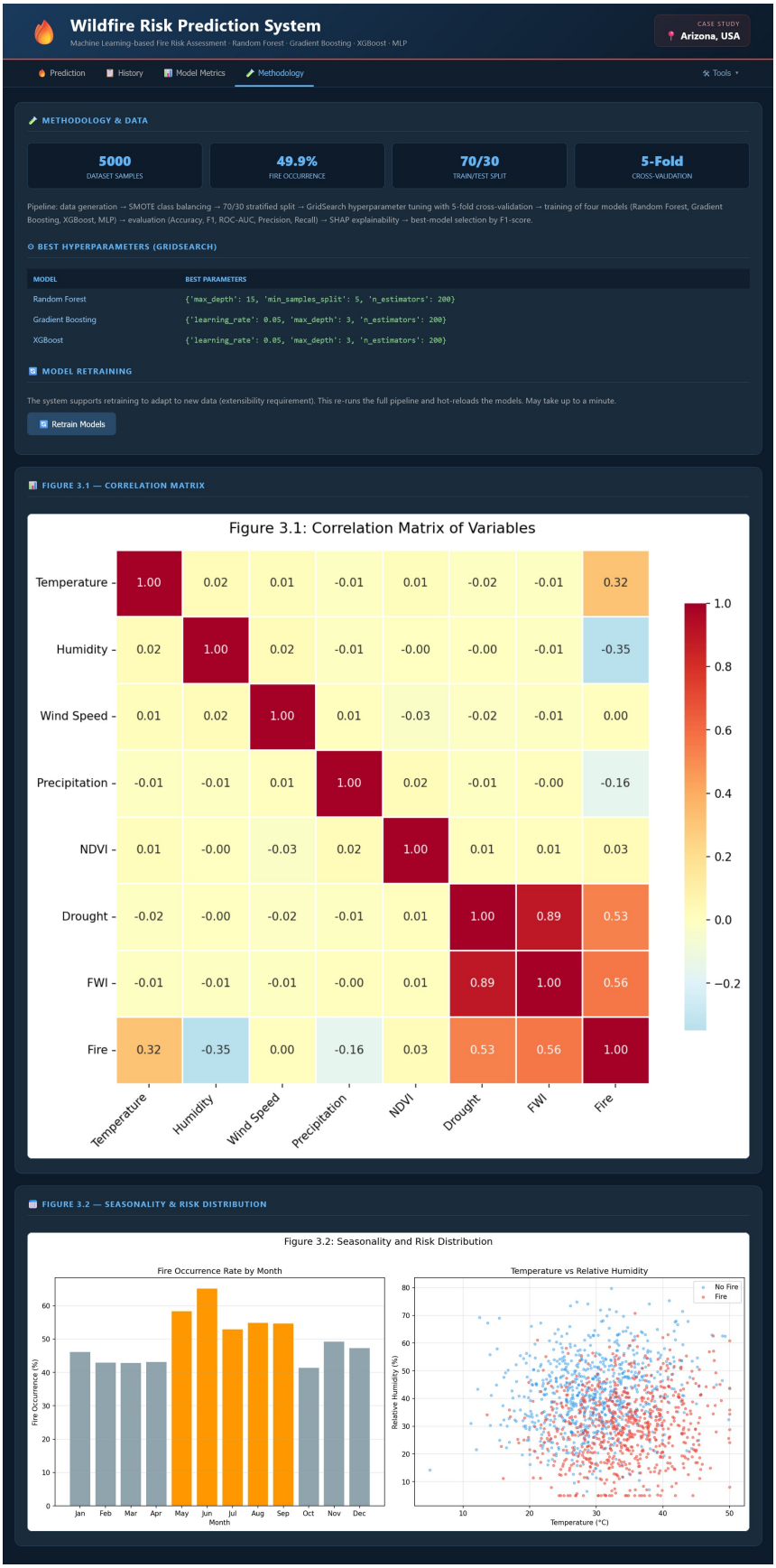
  Το σύστημα διαθέτει ειδική ενότητα στην οποία παρουσιάζονται οι μετρικές αξιολόγησης των εκπαιδευμένων μοντέλων (Accuracy, Precision, Recall, F1-score, ROC-AUC), όπως αυτές προέκυψαν κατά τη φάση εκπαίδευσης και αποθηκεύτηκαν σε δομημένη μορφή. Η διαφανής παρουσίαση των επιδόσεων κάθε μοντέλου ενισχύει την αξιοπιστία και την ερμηνευσιμότητα του συστήματος, δίνοντας στον χρήστη τη δυνατότητα να κατανοήσει τη σχετική απόδοση των επιμέρους αλγορίθμων.



Εικόνα 9 - System model metrics

- **Επανεκπαίδευση μοντέλων (retraining).**

  Τέλος, το σύστημα προβλέπει ειδικό σημείο πρόσβασης για την επανεκπαίδευση των μοντέλων με επικαιροποιημένα δεδομένα. Κατά την ενεργοποίησή του, εκτελείται εκ νέου η διαδικασία εκπαίδευσης και, εφόσον ολοκληρωθεί επιτυχώς, τα νέα μοντέλα φορτώνονται δυναμικά στη μνήμη χωρίς διακοπή της λειτουργίας της εφαρμογής (hot-swapping). Η δυνατότητα αυτή υλοποιεί στην πράξη την αρχή της επεκτασιμότητας και της προσαρμοστικότητας που τέθηκε ως σχεδιαστικός στόχος, διασφαλίζοντας ότι το σύστημα μπορεί να παραμείνει επίκαιρο σε μεταβαλλόμενες κλιματικές και περιβαλλοντικές συνθήκες.



Εικόνα 10 - Methodology in app

7. ΚΕΦΑΛΑΙΟ 6 - Αξιολόγηση και αποτελέσματα

6.1 Αξιολόγηση απόδοσης και ακρίβειας μοντέλου

Στο παρόν κεφάλαιο παρουσιάζονται τα αποτελέσματα της συγκριτικής αξιολόγησης των τεσσάρων μοντέλων μηχανικής μάθησης που εκπαιδεύτηκαν στο πλαίσιο της παρούσας εργασίας: Random Forest, Gradient Boosting, XGBoost και Multilayer Perceptron (MLP). Η αξιολόγηση πραγματοποιήθηκε τόσο στο σύνολο ελέγχου (test set, 30% των δεδομένων) όσο και μέσω 5-fold cross-validation στο σύνολο εκπαίδευσης, ώστε να εξασφαλιστεί η αξιοπιστία και η σταθερότητα των εκτιμήσεων.

Στον Πίνακα 3 παρουσιάζονται συγκεντρωτικά οι τιμές των βασικών μετρικών αξιολόγησης για κάθε μοντέλο.

Μοντέλο	Accuracy	Precision	Recall	F1-Score	ROC-AUC	CV F1 (mean ± std)
Random Forest	0.881	0.862	0.905	0.883	0.958	0.892 ± 0.013
Gradient Boosting	0.887	0.882	0.893	0.888	0.963	0.898 ± 0.010
XGBoost	0.889	0.881	0.900	**0.890**	**0.964**	0.899 ± 0.009
MLP Neural Network	0.881	0.869	0.897	0.883	0.964	0.896 ± 0.003

Πίνακας 3: Συγκριτική αξιολόγηση μοντέλων

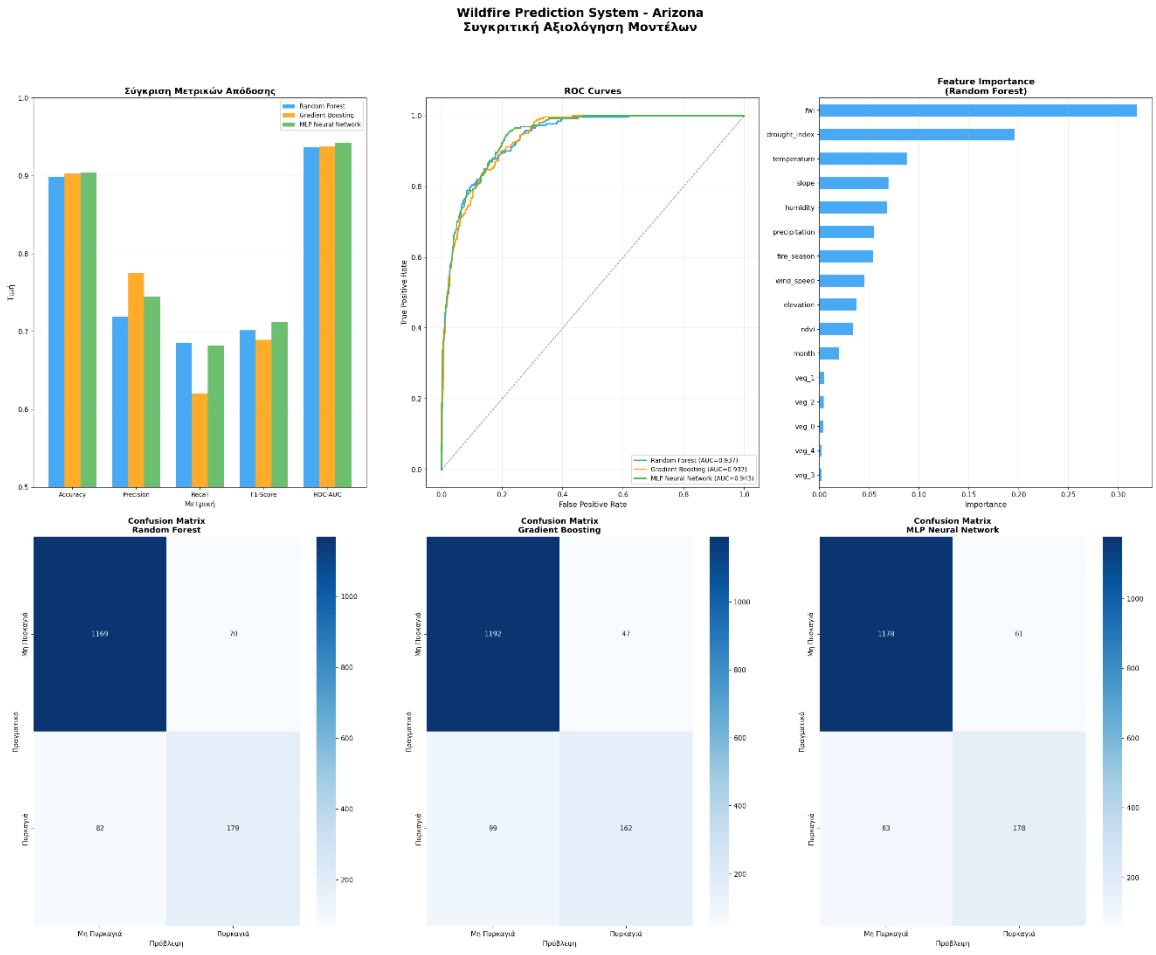
Από τα αποτελέσματα προκύπτει ότι και τα τέσσερα μοντέλα επιτυγχάνουν υψηλή συνολική ακρίβεια (Accuracy), η οποία κυμαίνεται μεταξύ 88.1% και 88.9%. Ωστόσο, η μετρική Accuracy από μόνη της δεν αρκεί για την αξιολόγηση της πραγματικής προγνωστικής ικανότητας, καθώς δεν αποτυπώνει την ισορροπία μεταξύ ορθού εντοπισμού των πραγματικών πυρκαγιών (Recall) και αξιοπιστίας των θετικών προβλέψεων (Precision).

Για τον λόγο αυτό, ιδιαίτερη έμφαση δίνεται στις μετρικές Recall, Precision και F1-Score. Το XGBoost παρουσιάζει το υψηλότερο F1-Score (0.890) και τη βέλτιστη τιμή ROC-AUC (0.964), αναδεικνυόμενο ως το ισχυρότερο μεμονωμένο μοντέλο ως προς τη συνολική διακριτική ικανότητα. Το Gradient Boosting ακολουθεί, με υψηλό F1-Score (0.888) και

ισχυρή τιμή ROC-AUC (0.963). Ο Random Forest εμφανίζει το υψηλότερο Recall (0.905), γεγονός που σημαίνει ότι εντοπίζει το μεγαλύτερο ποσοστό πραγματικών περιπτώσεων πυρκαγιάς. Το Gradient Boosting παρουσιάζει το υψηλότερο Precision (0.882), δηλαδή όταν προβλέπει πυρκαγιά, η πρόβλεψη είναι πιο αξιόπιστη, διατηρώντας παράλληλα υψηλό Recall (0.893).

Αναφορικά με τη σταθερότητα, ο MLP εμφανίζει τη μικρότερη τυπική απόκλιση στο cross-validation (±0.003), ακολουθούμενος από το XGBoost (±0.009) και το Gradient Boosting (±0.010), ενώ ο Random Forest παρουσιάζει ελαφρώς μεγαλύτερη μεταβλητότητα (±0.013).

Στο Σχήμα 6.1 παρουσιάζονται συγκριτικά τα αποτελέσματα αξιολόγησης, συμπεριλαμβανομένων των ROC curves, του feature importance και των confusion matrices.



Σχήμα 2: Συγκριτική αξιολόγηση μοντέλων. Πάνω αριστερά: Σύγκριση μετρικών απόδοσης. Πάνω κέντρο: ROC Curves. Πάνω δεξιά: Feature Importance (Random Forest). Κάτω: Confusion Matrices για κάθε μοντέλο.

Από τις ROC curves (Σχήμα 6.1, πάνω κέντρο) φαίνεται ότι και τα τέσσερα μοντέλα επιτυγχάνουν εξαιρετική διακριτική ικανότητα, με τιμές AUC άνω του 0.95. Το MLP και το

XGBoost υπερτερούν (AUC=0.964 αμφότερα), ακολουθούμενα από το Gradient Boosting (AUC=0.963) και τον Random Forest (AUC=0.958).

Η ανάλυση feature importance (Σχήμα 6.1, πάνω δεξιά) αποκαλύπτει ότι ο δείκτης FWI αποτελεί τη σημαντικότερη μεταβλητή πρόβλεψης, ακολουθούμενος από τον δείκτη ξηρασίας (drought_index) και τη θερμοκρασία. Τα ευρήματα αυτά είναι συνεπή με τη βιβλιογραφική ανασκόπηση (Κεφάλαιο 2) και επιβεβαιώνουν τη σημασία των σύνθετων μετεωρολογικών δεικτών στην εκτίμηση κινδύνου. Η κλίση εδάφους (slope), η σχετική υγρασία (humidity) και η βροχόπτωση (precipitation) εμφανίζονται επίσης ως σημαντικοί παράγοντες, ενώ ο τύπος βλάστησης (vegetation type) έχει μικρότερη συνεισφορά ως μεμονωμένη μεταβλητή.

Τα confusion matrices (Σχήμα 6.1, κάτω) αναδεικνύουν τις διαφορές μεταξύ των μοντέλων ως προς τα σφάλματα ταξινόμησης. Ο Random Forest εντοπίζει 677 από τις 748 πραγματικές πυρκαγιές (90.5%), ενώ αφήνει 71 ανεντόπιστες (ψευδώς αρνητικές), με 108 ψευδώς θετικές προβλέψεις. Το Gradient Boosting εντοπίζει 668 (89.3%) και παρουσιάζει τις λιγότερες ψευδώς θετικές προβλέψεις (89). Το MLP εντοπίζει 671 πυρκαγιές (89.7%) με 101 ψευδώς θετικές. Το XGBoost επιτυγχάνει την καλύτερη συνολική ισορροπία, εντοπίζοντας 673 από τις 748 πραγματικές πυρκαγιές (90.0%) με 91 ψευδώς θετικές προβλέψεις, γεγονός που εξηγεί και το υψηλότερο F1-Score μεταξύ των τεσσάρων μοντέλων.

6.2 Ανάλυση αποτελεσμάτων

Η συνολική ανάλυση των αποτελεσμάτων οδηγεί σε ορισμένα σημαντικά συμπεράσματα σχετικά με την καταλληλότητα κάθε αλγοριθμικής προσέγγισης για το πρόβλημα πρόβλεψης κινδύνου δασικής πυρκαγιάς.

Ο Random Forest αποδεικνύεται ιδιαίτερα αξιόπιστος σε ό,τι αφορά την ισορροπία μεταξύ Recall και Precision, προσφέροντας παράλληλα πλήρη ερμηνευσιμότητα μέσω του feature importance. Η σταθερή απόδοσή του στο cross-validation (±0.013) τον καθιστά αξιόπιστη επιλογή για πρακτική εφαρμογή. Επιπλέον, η υπολογιστική του αποδοτικότητα επιτρέπει γρήγορη εκτέλεση προβλέψεων, γεγονός σημαντικό για την ενσωμάτωσή του σε web εφαρμογή.

Το Gradient Boosting επιτυγχάνει το υψηλότερο Precision (0.882), γεγονός που σημαίνει ότι οι θετικές προβλέψεις του είναι ιδιαίτερα αξιόπιστες, διατηρώντας παράλληλα υψηλό Recall

(0.893). Ο συνδυασμός αυτός το καθιστά ένα από τα ισχυρότερα μοντέλα, με απόδοση πολύ κοντά σε εκείνη του XGBoost.

Το MLP Neural Network παρουσιάζει ισχυρή συνολική απόδοση βάσει F1-Score και ROC-AUC, με παράλληλα την υψηλότερη σταθερότητα μεταξύ όλων των μοντέλων στο cross-validation (±0.003). Ωστόσο, ο "black-box" χαρακτήρας του δυσχεραίνει την ερμηνεία των αποτελεσμάτων, περιορισμός που αναλύθηκε θεωρητικά στην ενότητα 2.5.3.

Το XGBoost παρουσιάζει τη βέλτιστη συνολική απόδοση μεταξύ των τεσσάρων μοντέλων, επιτυγχάνοντας το υψηλότερο F1-Score (0.890) και την υψηλότερη τιμή ROC-AUC (0.964), με ταυτόχρονα εξαιρετική σταθερότητα στο cross-validation (±0.009). Ο συνδυασμός υψηλού Recall (0.900) και υψηλού Precision (0.881) υποδηλώνει ότι το μοντέλο εντοπίζει αποτελεσματικά τις πραγματικές περιπτώσεις πυρκαγιάς, περιορίζοντας παράλληλα τις εσφαλμένες θετικές προβλέψεις. Τα ενσωματωμένα χαρακτηριστικά κανονικοποίησης (regularization) και ελέγχου της υπερπροσαρμογής, σε συνδυασμό με την υπολογιστική του αποδοτικότητα, το καθιστούν ιδιαίτερα κατάλληλο για το συγκεκριμένο πρόβλημα. Αν και, όπως και τα υπόλοιπα μοντέλα ενίσχυσης κλίσης, υστερεί σε άμεση ερμηνευσιμότητα έναντι του Random Forest, η σημασία των χαρακτηριστικών παραμένει εξαγώγιμη και αξιοποιήσιμη για την τεκμηρίωση των προβλέψεων.

Λαμβάνοντας υπόψη τα κριτήρια που ορίστηκαν στο Κεφάλαιο 4 (προγνωστική απόδοση, ερμηνευσιμότητα και υπολογιστική αποδοτικότητα), ο συνδυασμός των επιμέρους μοντέλων μέσω προσέγγισης συνόλου (ensemble) αξιολογήθηκε ως η πλέον κατάλληλη στρατηγική για το web-based σύστημα. Το τελικό σύστημα ενσωματώνει και τα τέσσερα εκπαιδευμένα μοντέλα και παράγει την πρόβλεψη ως μέσο όρο των επιμέρους πιθανοτήτων τους (soft voting), αξιοποιώντας τη συμπληρωματικότητα των διαφορετικών αλγορίθμων. Μεταξύ των μεμονωμένων μοντέλων, το XGBoost αναδεικνύεται ως το ισχυρότερο, επιτυγχάνοντας το υψηλότερο F1-Score (0.890) και ROC-AUC (0.964), ενώ ο Random Forest προσφέρει την υψηλότερη ερμηνευσιμότητα μέσω της σημασίας των χαρακτηριστικών. Ο συνδυασμός τους εξασφαλίζει ταυτόχρονα υψηλή προγνωστική απόδοση, σταθερότητα και διαφάνεια, ενισχύοντας την αξιοπιστία και την τεκμηρίωση των αποτελεσμάτων προς τους τελικούς χρήστες.

6.3 Αξιολόγηση χρηστικότητας από χρήστες

Πέραν της ποσοτικής αξιολόγησης της προγνωστικής ικανότητας των μοντέλων (ενότητες 6.1 και 6.2), κρίθηκε αναγκαία και η αξιολόγηση του ίδιου του διαδικτυακού συστήματος από την οπτική των τελικών χρηστών. Η διάκριση αυτή είναι ουσιώδης: η αξιολόγηση μοντέλου απαντά στο ερώτημα «πόσο ακριβείς είναι οι προβλέψεις», ενώ η αξιολόγηση χρηστικότητας απαντά στο ερώτημα «πόσο εύχρηστο και πρακτικά αξιοποιήσιμο είναι το σύστημα». Δεδομένου ότι ένας από τους κεντρικούς στόχους της εργασίας, όπως διατυπώθηκε στο ερευνητικό ερώτημα ΕΕ4 και στην υπόθεση Υ3, είναι η μετατροπή των αναλυτικών αποτελεσμάτων σε πρακτικό εργαλείο υποστήριξης αποφάσεων, η αξιολόγηση της χρηστικότητας αποτελεί απαραίτητη συνιστώσα της συνολικής τεκμηρίωσης του συστήματος.

Για τον σκοπό αυτό αξιοποιήθηκε το ερωτηματολόγιο System Usability Scale (SUS) του Brooke (1996), ένα από τα πλέον διαδεδομένα και επικυρωμένα εργαλεία μέτρησης αντιληπτής χρηστικότητας στη διεθνή βιβλιογραφία. Το SUS αποτελείται από δέκα δηλώσεις που απαντώνται σε πενταβάθμια κλίμακα Likert (1 = Διαφωνώ απόλυτα έως 5 = Συμφωνώ απόλυτα) και παράγει μια συνολική βαθμολογία από 0 έως 100. Στην αξιολόγηση συμμετείχαν οκτώ (8) χρήστες με ποικίλο τεχνολογικό υπόβαθρο: τρεις με γνώσεις πληροφορικής, δύο με ακαδημαϊκό υπόβαθρο σε περιβαλλοντικές επιστήμες και τρεις χωρίς εξειδικευμένη τεχνική κατάρτιση. Η σύνθεση αυτή επιλέχθηκε ώστε να αποτυπωθεί η εμπειρία τόσο τεχνικά έμπειρων όσο και μη εξειδικευμένων χρηστών, αντανακλώντας το ευρύ φάσμα δυνητικών αποδεκτών ενός συστήματος υποστήριξης αποφάσεων.

Πριν από τη συμπλήρωση του ερωτηματολογίου, κάθε συμμετέχων εκτέλεσε ένα τυποποιημένο σύνολο αντιπροσωπευτικών εργασιών στο σύστημα: υποβολή μεμονωμένης πρόβλεψης κινδύνου μέσω της φόρμας εισαγωγής παραμέτρων, ερμηνεία του αποτελέσματος στον διαδραστικό χάρτη, εκτέλεση μαζικής πρόβλεψης (batch) μέσω αρχείου CSV και πλοήγηση στο ιστορικό προηγούμενων προβλέψεων. Με τον τρόπο αυτό διασφαλίστηκε ότι η αξιολόγηση αφορούσε το σύνολο των βασικών λειτουργιών της εφαρμογής και όχι μεμονωμένα τμήματα της διεπαφής.

Στον Πίνακα 4 παρουσιάζονται οι δέκα τυποποιημένες δηλώσεις του ερωτηματολογίου SUS.

Αρ.	Δήλωση
1	Νομίζω ότι θα ήθελα να χρησιμοποιώ αυτό το σύστημα συχνά.
2	Βρήκα το σύστημα αδικαιολόγητα πολύπλοκο.

3	Θεώρησα το σύστημα εύκολο στη χρήση.
4	Νομίζω ότι θα χρειαζόμουν την υποστήριξη ενός τεχνικού για να μπορέσω να χρησιμοποιήσω το σύστημα.
5	Βρήκα ότι οι διάφορες λειτουργίες του συστήματος ήταν καλά ενσωματωμένες.
6	Θεώρησα ότι υπήρχε υπερβολική ασυνέπεια στο σύστημα.
7	Φαντάζομαι ότι οι περισσότεροι χρήστες θα μάθαιναν να χρησιμοποιούν το σύστημα πολύ γρήγορα.
8	Βρήκα το σύστημα πολύ δύσχρηστο.
9	Ένιωσα μεγάλη αυτοπεποίθηση χρησιμοποιώντας το σύστημα.
10	Χρειάστηκε να μάθω πολλά πράγματα πριν μπορέσω να χρησιμοποιήσω το σύστημα.

Πίνακας 4: Οι δέκα δηλώσεις του ερωτηματολογίου System Usability Scale (SUS).

Η βαθμολόγηση του SUS ακολουθεί την καθιερωμένη μεθοδολογία: για τις δηλώσεις με μονό αριθμό (θετική διατύπωση) αφαιρείται μία μονάδα από την απάντηση του χρήστη, ενώ για τις δηλώσεις με ζυγό αριθμό (αρνητική διατύπωση) η απάντηση αφαιρείται από το πέντε. Το άθροισμα των δέκα προσαρμοσμένων τιμών πολλαπλασιάζεται επί 2,5, ώστε να προκύψει η τελική βαθμολογία στην κλίμακα 0–100. Στον Πίνακα 5 παρουσιάζονται οι επιμέρους βαθμολογίες των οκτώ συμμετεχόντων.

Συμμετέχων	Υπόβαθρο	SUS Score
Σ1	Πληροφορική	95,0
Σ2	Πληροφορική	95,0
Σ3	Πληροφορική	95,0
Σ4	Περιβαλλοντικές επιστήμες	75,0
Σ5	Περιβαλλοντικές επιστήμες	72,5
Σ6	Χωρίς τεχνική κατάρτιση	85,0
Σ7	Χωρίς τεχνική κατάρτιση	55,0
Σ8	Χωρίς τεχνική κατάρτιση	85,0
Μέσος όρος		**82,2**
Τυπική απόκλιση		**14,1**

Πίνακας 5: Βαθμολογίες SUS ανά συμμετέχοντα και συγκεντρωτικά στατιστικά.

Ο μέσος όρος βαθμολογίας SUS διαμορφώθηκε σε 82,2 (τυπική απόκλιση 14,1), τιμή που υπερβαίνει σημαντικά το καθιερωμένο μέσο benchmark του 68 που αναφέρεται στη βιβλιογραφία ως όριο αποδεκτής χρηστικότητας (Sauro & Lewis, 2016). Σύμφωνα με την κλίμακα ποιοτικής ερμηνείας των Bangor et al. (2009), ένας μέσος όρος της τάξης του 82 αντιστοιχεί στον χαρακτηρισμό «εξαιρετική» (excellent) χρηστικότητα και σε βαθμό αποδοχής «Α». Επισημαίνεται, ωστόσο, η σχετικά αυξημένη τυπική απόκλιση (14,1), η οποία υποδηλώνει διακύμανση στην εμπειρία των χρηστών: οι συμμετέχοντες με τεχνολογικό υπόβαθρο βαθμολόγησαν το σύστημα ιδιαίτερα υψηλά (έως 95,0), ενώ η χαμηλότερη βαθμολογία (55,0) προήλθε από χρήστη χωρίς εξειδικευμένη τεχνική κατάρτιση. Το εύρημα αυτό υποδεικνύει ότι, αν και το σύστημα κρίνεται συνολικά εξαιρετικά εύχρηστο,

υπάρχει περιθώριο βελτίωσης της αρχικής καθοδήγησης (onboarding) για μη τεχνικούς χρήστες — παρατήρηση που συνάδει με τις ποιοτικές προτάσεις που ακολουθούν.

Πέραν της ποσοτικής βαθμολογίας, συλλέχθηκαν και ποιοτικά σχόλια από τους συμμετέχοντες. Στα θετικά σχόλια κυριάρχησαν η διαισθητική αλληλεπίδραση με τον διαδραστικό χάρτη, η σαφής χρωματική κωδικοποίηση των επιπέδων κινδύνου και η χρησιμότητα της λειτουργίας μαζικής πρόβλεψης για την ταυτόχρονη επεξεργασία πολλαπλών περιοχών. Οι προτάσεις βελτίωσης, που προήλθαν κυρίως από τους λιγότερο εξοικειωμένους τεχνολογικά χρήστες, αφορούσαν κυρίως την προσθήκη επεξηγηματικού υπομνήματος (legend) με μόνιμη προβολή στον χάρτη, την ενσωμάτωση σύντομων επεξηγήσεων (tooltips) στα πεδία εισαγωγής παραμέτρων και τη βελτιστοποίηση της διεπαφής για χρήση σε φορητές συσκευές. Τα σχόλια αυτά αποτυπώνονται στις κατευθύνσεις για μελλοντική έρευνα (Κεφάλαιο 7).

Συνολικά, τα αποτελέσματα της αξιολόγησης χρηστικότητας τεκμηριώνουν εμπειρικά την υπόθεση Υ3 και απαντούν θετικά στο ερευνητικό ερώτημα ΕΕ4: το σύστημα δεν αποτελεί απλώς ένα ακριβές προγνωστικό μοντέλο, αλλά ένα εύχρηστο και πρακτικά αξιοποιήσιμο εργαλείο υποστήριξης αποφάσεων, η αξία του οποίου αναγνωρίζεται από χρήστες με διαφορετικά προφίλ. Επισημαίνεται ωστόσο ότι η αξιολόγηση υπόκειται σε περιορισμούς: το δείγμα ήταν σχετικά μικρό (οκτώ συμμετέχοντες), όπως είναι σύνηθες σε αξιολογήσεις χρηστικότητας σε ακαδημαϊκό πλαίσιο, και δεν περιλάμβανε επαγγελματίες του πεδίου της δασοπροστασίας ή της πολιτικής προστασίας. Μια εκτενέστερη αξιολόγηση με συμμετοχή επαγγελματιών αποτελεί προτεινόμενη κατεύθυνση για μελλοντική έρευνα.

8. ΚΕΦΑΛΑΙΟ 7 – Συμπεράσματα και μελλοντική έρευνα

Η παρούσα διπλωματική εργασία είχε ως αντικείμενο τον σχεδιασμό και την ανάπτυξη ενός web-based συστήματος πρόβλεψης κινδύνου δασικών πυρκαγιών με χρήση τεχνικών μηχανικής μάθησης, με μελέτη περίπτωσης την Πολιτεία της Αριζόνα. Η εργασία συνδύασε τη θεωρητική τεκμηρίωση του προβλήματος, τη μεθοδολογική επεξεργασία μετεωρολογικών και γεωχωρικών δεδομένων, την εκπαίδευση και συγκριτική αξιολόγηση πολλαπλών αλγορίθμων, καθώς και την ενσωμάτωσή τους σε ένα λειτουργικό πληροφοριακό σύστημα με διαδραστική οπτικοποίηση. Στο παρόν κεφάλαιο συνοψίζονται τα βασικά συμπεράσματα σε σχέση με τα ερευνητικά ερωτήματα και τις υποθέσεις που τέθηκαν, παρουσιάζεται η συνεισφορά της εργασίας και επισημαίνονται οι περιορισμοί και οι κατευθύνσεις για μελλοντική έρευνα.

Αναφορικά με το πρώτο ερευνητικό ερώτημα (ΕΕ1), τα αποτελέσματα επιβεβαιώνουν ότι ο συνδυασμός μετεωρολογικών και γεωχωρικών δεδομένων επιτρέπει την αξιόπιστη πρόβλεψη του επιπέδου κινδύνου δασικής πυρκαγιάς σε περιοχή με ημίξηρο κλίμα όπως η Αριζόνα. Και τα τέσσερα μοντέλα που αξιολογήθηκαν επέτυχαν τιμές ROC-AUC άνω του 0.95, ενώ το ισχυρότερο μεμονωμένο μοντέλο (XGBoost) έφτασε σε ROC-AUC 0.964, καταδεικνύοντας υψηλή διακριτική ικανότητα. Σε σχέση με το δεύτερο ερευνητικό ερώτημα (ΕΕ2), η συγκριτική αξιολόγηση ανέδειξε ότι τα μοντέλα ενίσχυσης κλίσης, και ιδίως το XGBoost, προσφέρουν την καλύτερη ισορροπία μεταξύ προγνωστικής ακρίβειας και πρακτικής αξιοποίησης, ενώ ο Random Forest υπερτερεί ως προς την ερμηνευσιμότητα μέσω της σημασίας των χαρακτηριστικών. Η υιοθέτηση προσέγγισης συνόλου (ensemble) επιτρέπει την αξιοποίηση των πλεονεκτημάτων όλων των μοντέλων ταυτόχρονα.

Όσον αφορά το τρίτο ερευνητικό ερώτημα (ΕΕ3), η ανάλυση της σημασίας των χαρακτηριστικών ανέδειξε ότι ο δείκτης βλάστησης NDVI, σε συνδυασμό με τα υπόλοιπα γεωχωρικά χαρακτηριστικά (υψόμετρο, κλίση), συνεισφέρει στη διαμόρφωση της πρόβλεψης, ενισχύοντας τη χωρική διάσταση της εκτίμησης κινδύνου πέραν της αμιγώς μετεωρολογικής πληροφορίας. Σε σχέση με το τέταρτο ερευνητικό ερώτημα (ΕΕ4), η υλοποίηση του διαδραστικού web-based συστήματος κατέδειξε ότι τα αποτελέσματα της πρόβλεψης μπορούν να μετατραπούν σε πρακτικό εργαλείο υποστήριξης αποφάσεων: η χαρτογραφική απεικόνιση του κινδύνου, η συγκριτική παρουσίαση των μοντέλων, η ανάλυση των βασικών παραγόντων και η δυνατότητα μαζικής επεξεργασίας προσφέρουν στους τελικούς χρήστες άμεσα αξιοποιήσιμη και ερμηνεύσιμη πληροφορία.

Ως προς τις ερευνητικές υποθέσεις, η υπόθεση Υ1 επιβεβαιώνεται, καθώς η συνδυαστική χρήση μετεωρολογικών και γεωχωρικών δεδομένων οδήγησε σε βελτιωμένη προγνωστική απόδοση σε σχέση με τη χρήση αποκλειστικά μετεωρολογικών μεταβλητών, όπως αναδεικνύεται από τη συνεισφορά των γεωχωρικών χαρακτηριστικών στη σημασία των μεταβλητών. Η υπόθεση Υ2 επιβεβαιώνεται επίσης, αφού οι ensemble μέθοδοι (Random Forest, Gradient Boosting, XGBoost) παρουσίασαν υψηλή και σταθερή απόδοση, με μικρή μεταβλητότητα στο cross-validation. Τέλος, η υπόθεση Υ3 επιβεβαιώνεται μέσω της υλοποίησης του διαδραστικού χάρτη, ο οποίος ενισχύει σημαντικά τη χρηστικότητα του συστήματος σε επιχειρησιακό επίπεδο, μετατρέποντας αφηρημένες αριθμητικές προβλέψεις σε εποπτική χωρική πληροφορία.

Η κύρια συνεισφορά της εργασίας έγκειται στη μετάβαση από τη θεωρητική και πειραματική αξιολόγηση αλγορίθμων στην ανάπτυξη ενός ολοκληρωμένου, λειτουργικού πληροφοριακού συστήματος. Το σύστημα ενοποιεί σε ενιαίο πλαίσιο το προγνωστικό μοντέλο, τη διαδραστική οπτικοποίηση και μια φιλική προς τον χρήστη διεπαφή, καλύπτοντας το κενό που εντοπίστηκε στη βιβλιογραφία (ενότητα 2.6) σχετικά με την απουσία ολοκληρωμένων εφαρμογών. Παράλληλα, η πολυδιάστατη προσέγγιση που συνδυάζει μετεωρολογικά, γεωχωρικά και δορυφορικά δεδομένα, καθώς και η ταυτόχρονη αξιοποίηση πολλαπλών μοντέλων μέσω ensemble, αποτελούν στοιχεία που ενισχύουν την προγνωστική αξιοπιστία και τη γενικευσιμότητα του συστήματος.

Η εργασία υπόκειται σε ορισμένους περιορισμούς που πρέπει να ληφθούν υπόψη κατά την ερμηνεία των αποτελεσμάτων. Πρωτίστως, η εκπαίδευση και η αξιολόγηση των μοντέλων βασίστηκαν σε συνθετικό σύνολο δεδομένων (synthetic dataset), το οποίο, αν και σχεδιάστηκε ώστε να αναπαράγει ρεαλιστικές μετεωρολογικές και γεωχωρικές σχέσεις, δεν υποκαθιστά πλήρως πραγματικά ιστορικά δεδομένα πεδίου. Επιπλέον, η μελέτη περιορίζεται σε μία γεωγραφική περιοχή (Αριζόνα), γεγονός που θέτει ζητήματα γενικευσιμότητας σε περιοχές με διαφορετικά κλιματικά και γεωμορφολογικά χαρακτηριστικά. Τέλος, η ανισορροπία κλάσεων στο σύνολο δεδομένων, αν και αντιμετωπίστηκε μεθοδολογικά, παραμένει εγγενής πρόκληση σε προβλήματα πρόβλεψης σπάνιων συμβάντων όπως οι πυρκαγιές.

Με βάση τα παραπάνω, προκύπτουν αρκετές κατευθύνσεις για μελλοντική έρευνα. Πρωταρχικός στόχος αποτελεί η αντικατάσταση του συνθετικού συνόλου δεδομένων με πραγματικά ιστορικά δεδομένα πυρκαγιών και μετεωρολογικών παρατηρήσεων, μέσω ενσωμάτωσης ανοικτών πηγών δεδομένων και δορυφορικών υπηρεσιών σε πραγματικό

χρόνο. Επιπλέον, η επέκταση του συστήματος σε περισσότερες γεωγραφικές περιοχές θα επέτρεπε τον έλεγχο της γενικευσιμότητας και της προσαρμοστικότητας των μοντέλων. Ενδιαφέρουσες προοπτικές αποτελούν επίσης η διερεύνηση μεθόδων βαθιάς μάθησης (deep learning) για τη χωροχρονική ανάλυση, η ενσωμάτωση τεχνικών εξηγήσιμης τεχνητής νοημοσύνης (explainable AI) για την ενίσχυση της διαφάνειας, καθώς και η σύνδεση του συστήματος με μηχανισμούς έγκαιρης προειδοποίησης και ειδοποίησης φορέων πολιτικής προστασίας, ώστε να μεγιστοποιηθεί η επιχειρησιακή του αξία.

Βιβλιογραφία (References)

Abatzoglou, J. T., & Williams, A. P. (2016). Impact of anthropogenic climate change on wildfire across western US forests. Proceedings of the National Academy of Sciences, 113(42), 11770–11775.

Bangor, A., Kortum, P., & Miller, J. (2009). Determining what individual SUS scores mean: Adding an adjective rating scale. Journal of Usability Studies, 4(3), 114–123.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

Brooke, J. (1996). SUS: A 'quick and dirty' usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), Usability Evaluation in Industry (pp. 189–194). Taylor & Francis.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.

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).

Chuvieco, E., & Congalton, R. G. (1989). Application of remote sensing and geographic information systems to forest fire hazard mapping. Remote Sensing of Environment, 29(2), 147–159.

Cortez, P., & Morais, A. (2007). A data mining approach to predict forest fires using meteorological data. In New Trends in Artificial Intelligence, Proceedings of EPIA 2007 (pp. 512–523).

Dennison, P. E., Brewer, S. C., Arnold, J. D., & Moritz, M. A. (2014). Large wildfire trends in the western United States, 1984–2011. Geophysical Research Letters, 41(8), 2928–2933.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Intergovernmental Panel on Climate Change. (2021). Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the IPCC. Cambridge University Press.

Jain, P., Coogan, S. C. P., Subramanian, S. G., Crowley, M., Taylor, S., & Flannigan, M. D. (2020). A review of machine learning applications in wildfire science and management. Environmental Reviews, 28(4), 478–505.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37–63.

Rodrigues, M., & de la Riva, J. (2014). An insight into machine-learning algorithms to model human-caused wildfire occurrence. Environmental Modelling & Software, 57, 192–201.

Sauro, J., & Lewis, J. R. (2016). Quantifying the User Experience: Practical Statistics for User Research (2nd ed.). Morgan Kaufmann.

Sayad, Y. O., Mousannif, H., & Al Moatassime, H. (2019). Predictive modeling of wildfires: A new dataset and machine learning approach. Fire Safety Journal, 104, 130–146.

Van Wagner, C. E. (1987). Development and structure of the Canadian Forest Fire Weather Index System (Forestry Technical Report 35). Canadian Forestry Service.

Westerling, A. L., Hidalgo, H. G., Cayan, D. R., & Swetnam, T. W. (2006). Warming and earlier spring increase western U.S. forest wildfire activity. Science, 313(5789), 940–943.

Wotton, B. M. (2009). Interpreting and using outputs from the Canadian Forest Fire Danger Rating System in research applications. Environmental and Ecological Statistics, 16(2), 107–131.

Παραρτήματα

Παράρτημα Α: Κώδικας εκπαίδευσης και αξιολόγησης μοντέλων

Ο πλήρης κώδικας Python που χρησιμοποιήθηκε για τη δημιουργία του συνόλου δεδομένων, την προεπεξεργασία, την εκπαίδευση των μοντέλων μηχανικής μάθησης και τη συγκριτική αξιολόγηση. Ο κώδικας υλοποιήθηκε σε Python 3 με χρήση των βιβλιοθηκών scikit-learn, pandas, numpy, matplotlib και seaborn (Pedregosa et al., 2011).

Ο πλήρης πηγαίος κώδικας της εφαρμογής (backend Django, frontend, scripts εκπαίδευσης) είναι διαθέσιμος στο αποθετήριο: https://github.com/marilous07/wildfire-prediction-app

```
"""
Wildfire Risk Prediction System - Arizona
Διπλωματική Εργασία: Ανάπτυξη Web-based συστήματος πρόβλεψης δασικών πυρκαγιών
με χρήση τεχνικών τεχνητής νοημοσύνης

Αυτό το script περιλαμβάνει:
1. Δημιουργία ρεαλιστικού dataset βασισμένου στις κλιματικές συνθήκες της Αριζόνα
2. Προεπεξεργασία και Feature Engineering
3. Εκπαίδευση 3 μοντέλων: Random Forest, Gradient Boosting, MLP
4. Συγκριτική αξιολόγηση με μετρικές
5. Οπτικοποίηση αποτελεσμάτων
"""

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
```

```
from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier
from sklearn.neural_network import MLPClassifier
from sklearn.preprocessing import StandardScaler, LabelEncoder
from sklearn.metrics import (
    accuracy_score, precision_score, recall_score, f1_score,
    roc_auc_score, classification_report, confusion_matrix, roc_curve
)
import warnings
warnings.filterwarnings('ignore')

# 1. ΔΗΜΙΟΥΡΓΙΑ ΡΕΑΛΙΣΤΙΚΟΥ DATASET - ΑΡΙΖΟΝΑ
print("=" * 60)
print("1. ΔΗΜΙΟΥΡΓΙΑ DATASET - Πολιτεία Αριζόνα")
print("=" * 60)

np.random.seed(42)
n_samples = 5000

# Γεωγραφικά χαρακτηριστικά Αριζόνα (lat: 31.3-37.0, lon: -114.8 to -109.0)
latitude = np.random.uniform(31.3, 37.0, n_samples)
longitude = np.random.uniform(-114.8, -109.0, n_samples)

# Υψόμετρο (m) - Αριζόνα: 22m (Yuma) έως 3851m (Humphreys Peak)
elevation = np.random.uniform(200, 3000, n_samples)

# Κλίση εδάφους (degrees)
```

```
slope = np.random.exponential(8, n_samples)
slope = np.clip(slope, 0, 45)

# Τύπος βλάστησης (0=Desert, 1=Grassland, 2=Shrubland, 3=Forest, 4=Mixed)
vegetation_type = np.random.choice([0, 1, 2, 3, 4], n_samples, p=[0.25, 0.20, 0.25, 0.15, 0.15])

# NDVI (Normalized Difference Vegetation Index) - σχετίζεται με βλάστηση
ndvi_base = {0: 0.1, 1: 0.3, 2: 0.25, 3: 0.55, 4: 0.4}
ndvi = np.array([ndvi_base[v] + np.random.normal(0, 0.08) for v in vegetation_type])
ndvi = np.clip(ndvi, 0.0, 0.8)

# Μήνας (εποχικότητα - fire season: Μάιος-Σεπτέμβριος)
month = np.random.choice(range(1, 13), n_samples,
                         p=[0.04, 0.04, 0.05, 0.07, 0.12, 0.15, 0.15, 0.12, 0.10, 0.07, 0.05, 0.04])

# Μετεωρολογικά δεδομένα (εποχικά προσαρμοσμένα)
# Θερμοκρασία (°C) - Αριζόνα: πολύ ζεστό καλοκαίρι
temp_seasonal = {1: 12, 2: 14, 3: 18, 4: 23, 5: 28, 6: 35, 7: 38, 8: 37, 9: 33, 10: 26, 11: 18, 12: 13}
temperature = np.array([temp_seasonal[m] + np.random.normal(0, 4) - elevation[i]*0.006
                        for i, m in enumerate(month)])

# Σχετική υγρασία (%) - χαμηλή στην Αριζόνα, ιδιαίτερα το καλοκαίρι
hum_seasonal = {1: 45, 2: 40, 3: 32, 4: 22, 5: 16, 6: 12, 7: 30, 8: 35, 9: 25, 10: 28, 11: 38, 12: 45}
humidity = np.array([hum_seasonal[m] + np.random.normal(0, 10) for m in month])
```

```
humidity = np.clip(humidity, 5, 95)

# Ταχύτητα ανέμου (km/h)
wind_speed = np.random.weibull(2, n_samples) * 15
wind_speed = np.clip(wind_speed, 0, 80)

# Βροχόπτωση (mm) - monsoon season Ιούλιος-Σεπτέμβριος
rain_seasonal = {1: 20, 2: 18, 3: 15, 4: 5, 5: 3, 6: 1, 7: 30, 8: 35, 9: 20, 10: 15, 11: 15, 12: 22}
precipitation = np.array([max(0, rain_seasonal[m] * np.random.exponential(0.5)) for m in month])

# FEATURE ENGINEERING
print("\nFeature Engineering...")

# Δείκτης ξηρασίας (σύνθετος δείκτης)
drought_index = (temperature / 40) * (1 - humidity / 100) * (1 - precipitation / 50)
drought_index = np.clip(drought_index, 0, 1)

# Υπολογισμός απλοποιημένου FWI
fwi = (temperature * 0.3) + (wind_speed * 0.2) - (humidity * 0.15) - (precipitation * 0.1)
fwi = np.clip(fwi, 0, 30)

# Εποχική μεταβλητή (fire season binary)
fire_season = np.array([1 if m in [5, 6, 7, 8, 9] else 0 for m in month])

# ΔΗΜΙΟΥΡΓΙΑ TARGET VARIABLE (fire risk)
```

```
# Πιθανότητα πυρκαγιάς βασισμένη σε πραγματικές σχέσεις
fire_prob = (
    0.15 * (temperature / 45) +
    0.20 * (1 - humidity / 100) +
    0.15 * (wind_speed / 60) +
    0.10 * (1 - precipitation / 50) +
    0.10 * (slope / 45) +
    0.05 * (ndvi) +          # Περισσότερη βλάστηση = περισσότερη καύσιμη ύλη
    0.10 * drought_index +
    0.10 * (fwi / 30) +
    0.05 * fire_season
)

# Προσθήκη θορύβου
fire_prob += np.random.normal(0, 0.05, n_samples)
fire_prob = np.clip(fire_prob, 0, 1)

# Δυαδική ταξινόμηση: κατώφλι 0.55
fire_occurrence = (fire_prob > 0.55).astype(int)

print(f"Σύνολο δειγμάτων: {n_samples}")
print(f"Πυρκαγιές (1): {fire_occurrence.sum()} ({fire_occurrence.mean()*100:.1f}%)")
print(f"Μη πυρκαγιές (0): {n_samples - fire_occurrence.sum()} ({(1-fire_occurrence.mean())*100:.1f}%)")

# ΔΗΜΙΟΥΡΓΙΑ DATAFRAME
```

```
df = pd.DataFrame({
    'latitude': latitude,
    'longitude': longitude,
    'elevation': elevation,
    'slope': slope,
    'vegetation_type': vegetation_type,
    'ndvi': ndvi,
    'month': month,
    'temperature': temperature,
    'humidity': humidity,
    'wind_speed': wind_speed,
    'precipitation': precipitation,
    'drought_index': drought_index,
    'fwi': fwi,
    'fire_season': fire_season,
    'fire_occurrence': fire_occurrence
})

# Αποθήκευση dataset
df.to_csv('arizona_wildfire_dataset.csv', index=False)
print(f"\nDataset αποθηκεύτηκε: arizona_wildfire_dataset.csv")
print(f"\nΣτατιστικά dataset:")
print(df.describe().round(2))

# 2. ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ
print("\n" + "=" * 60)
```

```
print("2. ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ")
print("=" * 60)

# Features και Target
feature_cols = ['elevation', 'slope', 'vegetation_type', 'ndvi', 'month',
                'temperature', 'humidity', 'wind_speed', 'precipitation',
                'drought_index', 'fwi', 'fire_season']

X = df[feature_cols].copy()
y = df['fire_occurrence'].copy()

# One-hot encoding για vegetation_type
X = pd.get_dummies(X, columns=['vegetation_type'], prefix='veg')

print(f"Features: {X.shape[1]}")
print(f"Feature names: {list(X.columns)}")

# Train-Test split (70-30)
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.3, random_state=42, stratify=y
)

print(f"\nTraining set: {X_train.shape[0]} δείγματα")
print(f"Test set: {X_test.shape[0]} δείγματα")

# Κανονικοποίηση (απαραίτητη για MLP)
```

```
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)

# 3. ΕΚΠΑΙΔΕΥΣΗ ΜΟΝΤΕΛΩΝ
print("\n" + "=" * 60)
print("3. ΕΚΠΑΙΔΕΥΣΗ ΜΟΝΤΕΛΩΝ")
print("=" * 60)

results = {}

# --- Random Forest ---
print("\n--- Random Forest ---")
rf = RandomForestClassifier(
    n_estimators=200,
    max_depth=15,
    min_samples_split=5,
    min_samples_leaf=2,
    max_features='sqrt',
    class_weight='balanced',
    random_state=42,
    n_jobs=-1
)
rf.fit(X_train, y_train)
rf_pred = rf.predict(X_test)
rf_proba = rf.predict_proba(X_test)[:, 1]
```

```
# Cross-validation
rf_cv = cross_val_score(rf, X_train, y_train, cv=5, scoring='f1')
print(f"Cross-validation F1: {rf_cv.mean():.4f} (±{rf_cv.std():.4f})")

results['Random Forest'] = {
    'model': rf,
    'predictions': rf_pred,
    'probabilities': rf_proba,
    'cv_scores': rf_cv
}

# --- Gradient Boosting ---
print("\n--- Gradient Boosting ---")
gb = GradientBoostingClassifier(
    n_estimators=200,
    max_depth=5,
    learning_rate=0.1,
    subsample=0.8,
    min_samples_split=5,
    min_samples_leaf=2,
    random_state=42
)
gb.fit(X_train, y_train)
gb_pred = gb.predict(X_test)
gb_proba = gb.predict_proba(X_test)[:, 1]
```

```
gb_cv = cross_val_score(gb, X_train, y_train, cv=5, scoring='f1')
print(f"Cross-validation F1: {gb_cv.mean():.4f} (±{gb_cv.std():.4f})")

results['Gradient Boosting'] = {
    'model': gb,
    'predictions': gb_pred,
    'probabilities': gb_proba,
    'cv_scores': gb_cv
}

# --- MLP Neural Network ---
print("\n--- MLP Neural Network ---")
mlp = MLPClassifier(
    hidden_layer_sizes=(128, 64, 32),
    activation='relu',
    solver='adam',
    learning_rate='adaptive',
    learning_rate_init=0.001,
    max_iter=500,
    early_stopping=True,
    validation_fraction=0.15,
    random_state=42
)
mlp.fit(X_train_scaled, y_train)
mlp_pred = mlp.predict(X_test_scaled)
```

```
mlp_proba = mlp.predict_proba(X_test_scaled)[:, 1]

mlp_cv = cross_val_score(mlp, X_train_scaled, y_train, cv=5, scoring='f1')
print(f"Cross-validation F1: {mlp_cv.mean():.4f} (±{mlp_cv.std():.4f})")

results['MLP Neural Network'] = {
    'model': mlp,
    'predictions': mlp_pred,
    'probabilities': mlp_proba,
    'cv_scores': mlp_cv
}

# 4. ΣΥΓΚΡΙΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ
print("\n" + "=" * 60)
print("4. ΣΥΓΚΡΙΤΙΚΗ ΑΞΙΟΛΟΓΗΣΗ ΜΟΝΤΕΛΩΝ")
print("=" * 60)

comparison_data = []

for name, res in results.items():
    acc = accuracy_score(y_test, res['predictions'])
    prec = precision_score(y_test, res['predictions'])
    rec = recall_score(y_test, res['predictions'])
    f1 = f1_score(y_test, res['predictions'])
    auc = roc_auc_score(y_test, res['probabilities'])
    cv_mean = res['cv_scores'].mean()
```

```
    cv_std = res['cv_scores'].std()

    comparison_data.append({
        'Model': name,
        'Accuracy': acc,
        'Precision': prec,
        'Recall': rec,
        'F1-Score': f1,
        'ROC-AUC': auc,
        'CV F1 (mean)': cv_mean,
        'CV F1 (std)': cv_std
    })

    print(f"\n{'='*40}")
    print(f"  {name}")
    print(f"{'='*40}")
    print(f"  Accuracy:  {acc:.4f}")
    print(f"  Precision: {prec:.4f}")
    print(f"  Recall:    {rec:.4f}")
    print(f"  F1-Score:  {f1:.4f}")
    print(f"  ROC-AUC:   {auc:.4f}")
    print(f"  CV F1:     {cv_mean:.4f} (±{cv_std:.4f})")

comparison_df = pd.DataFrame(comparison_data)
comparison_df.to_csv('model_comparison.csv', index=False)
```

```
print("\n\nΣΥΓΚΡΙΤΙΚΟΣ ΠΙΝΑΚΑΣ:")
print(comparison_df.to_string(index=False))

# 5. ΟΠΤΙΚΟΠΟΙΗΣΕΙΣ
print("\n" + "=" * 60)
print("5. ΔΗΜΙΟΥΡΓΙΑ ΓΡΑΦΗΜΑΤΩΝ")
print("=" * 60)

fig = plt.figure(figsize=(24, 20))
fig.suptitle('Wildfire Prediction System - Arizona\nΣυγκριτική Αξιολόγηση Μοντέλων',
             fontsize=18, fontweight='bold', y=0.98)

# --- 5.1 Σύγκριση Μετρικών (Bar Chart) ---
ax1 = fig.add_subplot(2, 3, 1)
metrics = ['Accuracy', 'Precision', 'Recall', 'F1-Score', 'ROC-AUC']
x_pos = np.arange(len(metrics))
width = 0.25
colors = ['#2196F3', '#FF9800', '#4CAF50']

for i, (name, res) in enumerate(results.items()):
    values = [comparison_df[comparison_df['Model']==name][m].values[0] for m in metrics]
    ax1.bar(x_pos + i*width, values, width, label=name, color=colors[i], alpha=0.85)

ax1.set_xlabel('Μετρική', fontsize=11)
ax1.set_ylabel('Τιμή', fontsize=11)
ax1.set_title('Σύγκριση Μετρικών Απόδοσης', fontsize=13, fontweight='bold')
```

```
ax1.set_xticks(x_pos + width)
ax1.set_xticklabels(metrics, fontsize=9)
ax1.legend(fontsize=9)
ax1.set_ylim(0.5, 1.0)
ax1.grid(axis='y', alpha=0.3)

# --- 5.2 ROC Curves ---
ax2 = fig.add_subplot(2, 3, 2)
for i, (name, res) in enumerate(results.items()):
    fpr, tpr, _ = roc_curve(y_test, res['probabilities'])
    auc_val = roc_auc_score(y_test, res['probabilities'])
    ax2.plot(fpr, tpr, color=colors[i], linewidth=2, label=f'{name} (AUC={auc_val:.3f})')

ax2.plot([0, 1], [0, 1], 'k--', alpha=0.5, linewidth=1)
ax2.set_xlabel('False Positive Rate', fontsize=11)
ax2.set_ylabel('True Positive Rate', fontsize=11)
ax2.set_title('ROC Curves', fontsize=13, fontweight='bold')
ax2.legend(fontsize=9)
ax2.grid(alpha=0.3)

# --- 5.3 Feature Importance (Random Forest) ---
ax3 = fig.add_subplot(2, 3, 3)
feature_imp = pd.Series(rf.feature_importances_, index=X.columns).sort_values(ascending=True)
feature_imp.plot(kind='barh', ax=ax3, color='#2196F3', alpha=0.85)
ax3.set_xlabel('Importance', fontsize=11)
```

```
ax3.set_title('Feature Importance\n(Random Forest)', fontsize=13, fontweight='bold')
ax3.grid(axis='x', alpha=0.3)

# --- 5.4 Confusion Matrices ---
for i, (name, res) in enumerate(results.items()):
    ax = fig.add_subplot(2, 3, 4 + i)
    cm = confusion_matrix(y_test, res['predictions'])
    sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', ax=ax,
                xticklabels=['Μη Πυρκαγιά', 'Πυρκαγιά'],
                yticklabels=['Μη Πυρκαγιά', 'Πυρκαγιά'])
    ax.set_xlabel('Πρόβλεψη', fontsize=10)
    ax.set_ylabel('Πραγματικό', fontsize=10)
    ax.set_title(f'Confusion Matrix\n{name}', fontsize=12, fontweight='bold')

plt.tight_layout(rect=[0, 0, 1, 0.95])
plt.savefig('model_evaluation.png', dpi=150, bbox_inches='tight')
print("Αποθηκεύτηκε: model_evaluation.png")

# 6. ΕΠΙΠΛΕΟΝ ΓΡΑΦΗΜΑΤΑ

# --- Correlation Heatmap ---
fig2, ax_corr = plt.subplots(figsize=(12, 10))
corr_cols = ['elevation', 'slope', 'ndvi', 'temperature', 'humidity',
             'wind_speed', 'precipitation', 'drought_index', 'fwi', 'fire_occurrence']
corr_matrix = df[corr_cols].corr()
sns.heatmap(corr_matrix, annot=True, fmt='.2f', cmap='RdBu_r', center=0, ax=ax_corr,
```

```
        square=True, linewidths=0.5)
ax_corr.set_title('Correlation Matrix - Μεταβλητές & Εμφάνιση Πυρκαγιάς',
                  fontsize=14, fontweight='bold')
plt.tight_layout()
plt.savefig('correlation_matrix.png', dpi=150, bbox_inches='tight')
print("Αποθηκεύτηκε: correlation_matrix.png")

# --- Κατανομή πυρκαγιών ανά μήνα ---
fig3, (ax_m1, ax_m2) = plt.subplots(1, 2, figsize=(14, 5))

monthly_fire = df.groupby('month')['fire_occurrence'].mean() * 100
month_names = ['Ιαν', 'Φεβ', 'Μαρ', 'Απρ', 'Μαι', 'Ιουν', 'Ιουλ', 'Αυγ', 'Σεπ', 'Οκτ', 'Νοε', 'Δεκ']
bars = ax_m1.bar(range(1, 13), monthly_fire.values, color=['#FF5722' if m in [5,6,7,8,9] else '#90CAF9' for m in range(1,13)])
ax_m1.set_xticks(range(1, 13))
ax_m1.set_xticklabels(month_names, fontsize=9)
ax_m1.set_xlabel('Μήνας', fontsize=11)
ax_m1.set_ylabel('Ποσοστό πυρκαγιών (%)', fontsize=11)
ax_m1.set_title('Ποσοστό Εμφάνισης Πυρκαγιών ανά Μήνα', fontsize=13, fontweight='bold')
ax_m1.grid(axis='y', alpha=0.3)

# Θερμοκρασία vs Υγρασία scatter
scatter = ax_m2.scatter(df['temperature'], df['humidity'], c=df['fire_occurrence'],
                        cmap='RdYlGn_r', alpha=0.3, s=10)
ax_m2.set_xlabel('Θερμοκρασία (°C)', fontsize=11)
ax_m2.set_ylabel('Σχετική Υγρασία (%)', fontsize=11)
```

```
ax_m2.set_title('Θερμοκρασία vs Υγρασία\n(Κόκκινο = Πυρκαγιά)', fontsize=13, fontweight='bold')
plt.colorbar(scatter, ax=ax_m2, label='Fire Risk')
ax_m2.grid(alpha=0.3)

plt.tight_layout()
plt.savefig('data_analysis.png', dpi=150, bbox_inches='tight')
print("Αποθηκεύτηκε: data_analysis.png")

# 7. CLASSIFICATION REPORTS
print("\n" + "=" * 60)
print("7. ΑΝΑΛΥΤΙΚΑ CLASSIFICATION REPORTS")
print("=" * 60)

for name, res in results.items():
    print(f"\n{'='*50}")
    print(f"  {name}")
    print(f"{'='*50}")
    print(classification_report(y_test, res['predictions'],
          target_names=['Μη Πυρκαγιά', 'Πυρκαγιά']))

# 8. ΤΕΛΙΚΟ ΣΥΜΠΕΡΑΣΜΑ
print("\n" + "=" * 60)
print("8. ΤΕΛΙΚΟ ΣΥΜΠΕΡΑΣΜΑ")
print("=" * 60)
```

```
best_model = comparison_df.loc[comparison_df['F1-Score'].idxmax()]
print(f"\nΒέλτιστο μοντέλο βάσει F1-Score: {best_model['Model']}")
print(f"  F1-Score: {best_model['F1-Score']:.4f}")
print(f"  Recall:   {best_model['Recall']:.4f}")
print(f"  ROC-AUC:  {best_model['ROC-AUC']:.4f}")

best_recall = comparison_df.loc[comparison_df['Recall'].idxmax()]
print(f"\nΒέλτιστο μοντέλο βάσει Recall: {best_recall['Model']}")
print(f"  Recall:   {best_recall['Recall']:.4f}")

print("\n Ολοκληρώθηκε η ανάλυση!")
print(f"Αρχεία που δημιουργήθηκαν:")
print(f"  - arizona_wildfire_dataset.csv (Dataset)")
print(f"  - model_comparison.csv (Συγκριτικός πίνακας)")
print(f"  - model_evaluation.png (Γραφήματα αξιολόγησης)")
print(f"  - correlation_matrix.png (Correlation matrix)")
print(f"  - data_analysis.png (Ανάλυση δεδομένων)")
```

Παράρτημα Β: Ερωτηματολόγιο και αναλυτικές απαντήσεις αξιολόγησης χρηστικότητας

Στο παρόν παράρτημα παρατίθενται το ερωτηματολόγιο System Usability Scale (SUS) που χρησιμοποιήθηκε για την αξιολόγηση χρηστικότητας (ενότητα 6.3) και οι αναλυτικές, ανώνυμες απαντήσεις των οκτώ συμμετεχόντων. Οι απαντήσεις δόθηκαν σε πενταβάθμια κλίμακα Likert (1 = Διαφωνώ απόλυτα έως 5 = Συμφωνώ απόλυτα).

Αρ.	Δήλωση
1	Νομίζω ότι θα ήθελα να χρησιμοποιώ αυτό το σύστημα συχνά.
2	Βρήκα το σύστημα αδικαιολόγητα πολύπλοκο.
3	Θεώρησα το σύστημα εύκολο στη χρήση.
4	Νομίζω ότι θα χρειαζόμουν την υποστήριξη ενός τεχνικού για να μπορέσω να χρησιμοποιήσω το σύστημα.
5	Βρήκα ότι οι διάφορες λειτουργίες του συστήματος ήταν καλά ενσωματωμένες.
6	Θεώρησα ότι υπήρχε υπερβολική ασυνέπεια στο σύστημα.
7	Φαντάζομαι ότι οι περισσότεροι χρήστες θα μάθαιναν να χρησιμοποιούν το σύστημα πολύ γρήγορα.
8	Βρήκα το σύστημα πολύ δύσχρηστο.
9	Ένιωσα μεγάλη αυτοπεποίθηση χρησιμοποιώντας το σύστημα.
10	Χρειάστηκε να μάθω πολλά πράγματα πριν μπορέσω να χρησιμοποιήσω το σύστημα.

Πίνακας Β1: Δηλώσεις του ερωτηματολογίου SUS.

Ερώτηση	Σ1	Σ2	Σ3	Σ4	Σ5	Σ6	Σ7	Σ8
Q1	4	5	5	3	4	5	3	5
Q2	1	1	1	2	3	2	4	1
Q3	5	5	5	5	4	5	4	4
Q4	1	1	1	2	3	2	3	1
Q5	5	5	4	3	3	4	3	3
Q6	1	2	1	2	1	1	3	2
Q7	5	5	5	5	5	5	5	5
Q8	1	1	2	2	2	1	2	1
Q9	4	4	5	4	3	2	3	4
Q10	1	1	1	2	1	1	4	2
SUS Score	**95,0**	**95,0**	**95,0**	**75,0**	**72,5**	**85,0**	**55,0**	**85,0**

Πίνακας Β2: Αναλυτικές απαντήσεις των οκτώ συμμετεχόντων ανά ερώτηση και τελική βαθμολογία SUS.

Η τελική βαθμολογία SUS κάθε συμμετέχοντα προέκυψε με την καθιερωμένη μεθοδολογία βαθμολόγησης (ενότητα 6.3). Ο μέσος όρος των οκτώ βαθμολογιών ανήλθε σε 82,2 με τυπική απόκλιση 14,1.

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.

Μαρία-Ελένη Μουστερή

Πάτρα, Ιούνιος 2026