



ΣΧΟΛΗ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΝΑΛΥΣΗ ΚΑΙ ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΤΗΣ ΨΗΦΙΑΚΗΣ
ΠΡΟΟΔΟΥ ΤΩΝ ΚΡΑΤΩΝ-ΜΕΛΩΝ ΤΗΣ Ε.Ε. ΜΕ ΧΡΗΣΗ ΤΟΥ
ΔΕΙΚΤΗ DESI ΚΑΙ ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

ΚΛΕΟΠΑΤΡΑ ΔΕΤΣΗ

ΑΜ 531578

ΕΠΙΒΛΕΠΟΝΤΕΣ ΚΑΘΗΓΗΤΕΣ

ΙΩΑΝΝΗΣ ΛΑΚΟΥΜΕΝΤΑΣ

ΑΛΕΞΑΝΔΡΟΣ ΤΖΑΛΛΑΣ

2025

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

Περίληψη

Η εργασία διερευνά την ψηφιακή πρόοδο των κρατών-μελών της Ευρωπαϊκής Ένωσης αξιοποιώντας τον Δείκτη Ψηφιακής Οικονομίας και Κοινωνίας (DESI) του 2022, ως εργαλείο αποτίμησης και κατηγοριοποίησης. Στο πλαίσιο της ανάλυσης, εφαρμόζονται προηγμένες τεχνικές μηχανικής μάθησης τόσο επιβλεπόμενης (supervised learning) όσο και μη επιβλεπόμενης (unsupervised learning), με στόχο την ανακάλυψη προτύπων και διαφοροποιήσεων στον ψηφιακό μετασχηματισμό των κρατών της Ε.Ε.

Μεθοδολογικά, αξιοποιήθηκαν αλγόριθμοι ταξινόμησης όπως Decision Tree (J48), Random Forest, Adaboost, Multilayer Perceptron και Support Vector Machine για την κατηγοριοποίηση των χωρών, ενώ για την ομαδοποίησή τους χρησιμοποιήθηκε ο αλγόριθμος K-means. Η ανάλυση υλοποιήθηκε στο λογισμικό Weka, με υποστηρικτική χρήση του R και του Excel για απεικονίσεις και οργάνωση των δεδομένων. Έγινε επίσης επιλογή χαρακτηριστικών (feature selection) με μεθόδους Wrapper-Greedy Stepwise και Wrapper-Genetic Search, με στόχο τη βελτίωση της αποδοτικότητας των μοντέλων και την αποφυγή της υπερπροσαρμογής.

Η εργασία δεν περιορίζεται σε τεχνική επεξεργασία δεδομένων, αλλά επιχειρεί και τη συσχέτιση της ψηφιακής επίδοσης των κρατών με δείκτες οικονομικής και κοινωνικής ανάπτυξης, αντλώντας στοιχεία από πρόσφατες επιστημονικές μελέτες. Μέσω της συγκριτικής αυτής προσέγγισης, αναδεικνύονται τάσεις ψηφιακής σύγκλισης αλλά και υφιστάμενες ανισότητες μεταξύ των κρατών-μελών.

Τα αποτελέσματα καταδεικνύουν ότι η ψηφιακή ωριμότητα των χωρών συσχετίζεται σημαντικά με τη γενικότερη αναπτυξιακή τους πορεία και προτείνουν την ανάγκη χάραξης στοχευμένων πολιτικών, ώστε να επιτευχθεί μια περισσότερο ισόρροπη ψηφιακή Ευρώπη στο πλαίσιο της Ψηφιακής Δεκαετίας 2030.

Λέξεις - Κλειδιά

Δείκτης Ψηφιακής Οικονομίας και Κοινωνίας (DESI), ψηφιακός μετασχηματισμός, Ευρωπαϊκή Ένωση, μηχανική μάθηση, ομαδοποίηση, ταξινόμηση, ψηφιακή στρατηγική, δημόσια διοίκηση, τεχνητή νοημοσύνη, ανθρώπινο κεφάλαιο, συνδεσιμότητα, ψηφιακές δημόσιες υπηρεσίες, ενσωμάτωση ψηφιακής τεχνολογίας, Weka, επιλογή χαρακτηριστικών, επιβλεπόμενη μάθηση, μη επιβλεπόμενη μάθηση, ψηφιακή ετοιμότητα.

Abstract

This thesis explores the digital progress of European Union member states using the Digital Economy and Society Index (DESI) 2022 as a key instrument for assessment and classification. The study applies advanced machine learning techniques—both supervised and unsupervised—to identify patterns and disparities in the digital transformation of EU countries.

Methodologically, the analysis employs classification algorithms such as Decision Tree (J48), Random Forest, Adaboost, Multilayer Perceptron, and Support Vector Machine, while the K-means algorithm is used for clustering. The implementation was carried out in Weka, with R and Excel utilized for visualizations and data organization. Feature selection techniques, including Wrapper-Greedy Stepwise and Wrapper-Genetic Search, were applied to enhance model accuracy and prevent overfitting.

Beyond technical data analysis, the thesis correlates digital performance with economic and social development indicators by drawing on recent academic studies. This comparative approach highlights both convergence trends and persistent inequalities across EU member states.

The findings indicate a strong relationship between digital maturity and broader developmental trajectories, emphasizing the need for targeted policy-making to achieve a more balanced and inclusive digital Europe within the framework of the Digital Decade 2030.

Keywords

Digital Economy and Society Index (DESI), digital transformation, European Union, machine learning, clustering, classification, digital strategy, public administration, artificial intelligence, human capital, connectivity, digital public services, digital technology integration, Weka, feature selection, supervised learning, unsupervised learning, digital readiness.

Πίνακας περιεχομένων

1. Εισαγωγή	8
2. Αναλυτική Επισκόπηση του Δείκτη DESI (Digital Economy and Society Index) (European Commission, 2022)	10
2.1 Ανθρώπινο Κεφάλαιο (Human Capital).....	10
2.1.1 Υποδιαστάσεις & Δείκτες:.....	10
2.2 Συνδεσιμότητα (Connectivity)	11
2.2.1 Υποδιαστάσεις & Δείκτες:.....	11
2.3 Ενσωμάτωση Ψηφιακής Τεχνολογίας (Integration of Digital Technology)	13
2.3.1 Υποδιαστάσεις και Δείκτες.....	13
2.4 Ψηφιακές Δημόσιες Υπηρεσίες (Digital Public Services)	14
2.4.1 Υποδιαστάσεις και Δείκτες:.....	14
2.5 Δομή και χαρακτηριστικά του συνόλου δεδομένων DESI 2022	15
3. Αρχές Μηχανικής Μάθησης	17
3.1 Βασικές Κατηγορίες Μηχανικής Μάθησης	17
3.1.1 Επιβλεπόμενη Μάθηση και Ανάλυση του Δείκτη DESI.....	18
3.1.2 Μη Επιβλεπόμενη Μάθηση και Ομαδοποίηση Χωρών	24
4. Βιβλιογραφική Επισκόπηση για συναφείς εργασίες.....	26
5. Συσχέτιση δεδομένων DESI με άλλες μελέτες με επιβλεπόμενη μάθηση.	33
5.1 Bánhidi & Dobos (2023) – Digital Development	33
5.1.1 Decision Tree (J48).....	33
5.1.2 Random Forest	34
5.1.3 Adaboost	35
5.1.4 Multilayer Perceptron	35
5.1.5 Support Vector Machine	36
5.2 Martin & Carnero (2019) – Sustainable Development	39
5.2.1 Decision Tree (J48).....	39

5.2.2	Random Forest	40
5.2.3	Adaboost	41
5.2.4	Multilayer Perceptron	41
5.2.5	Support Vector Machine	42
5.3	Onuferová et al. (2020) – Economic Development.....	45
5.3.1	Decision Tree (J48).....	45
5.3.2	Random Forest	46
5.3.3	Adaboost	46
5.3.4	Multilayer Perceptron	47
5.3.5	Support Vector Machine	48
5.4	Συμπεράσματα και συγκριτική μελέτη	51
6.	Ανάλυση Ομαδοποίησης (Clustering) με βάση τον Δείκτη DESI 2022.....	54
6.1	Μεθοδολογία.....	54
6.2	Αποτελέσματα για το σύνολο των χαρακτηριστικών	57
6.3	Αποτελέσματα ανά συνιστώσα του DESI - Ανθρώπινο Κεφάλαιο (Human Capital) 58	
6.4	Ανάλυση ανά συνιστώσα του DESI-Συνδεσιμότητα (Connectivity).....	60
6.5	Ανάλυση ανά συνιστώσα του DESI - Ενσωμάτωση Ψηφιακής Τεχνολογίας (Integration of Digital Technology).....	61
6.6	Ανάλυση ανά συνιστώσα του DESI Ψηφιακές Δημόσιες Υπηρεσίες (Digital Public Services).....	62
6.7	Συσχέτιση με προηγούμενες έρευνες	63
6.8	Συμπεράσματα	65
7.	Ανάλυση ταξινόμησης με βάση τον δείκτη DESI	67
7.1	Μεθοδολογία.....	67
7.2	Αποτελέσματα για το σύνολο των χαρακτηριστικών του δείκτη DESI.....	68
7.2.1	Decision Tree (J48).....	68
7.2.2	Random Forest	69

7.2.3	Adaboost	70
7.2.4	Multilayer Perceptron	70
7.2.5	Support Vector Machine	71
7.3	Αποτελέσματα για τη συνιστώσα «Ανθρώπινο Κεφάλαιο»	74
7.3.1	Decision Tree (J48).....	74
7.3.2	Random Forest	74
7.3.3	Adaboost	75
7.3.4	Multilayer Perceptron	76
7.3.5	Support Vector Machine	77
7.4	Αποτελέσματα για τη συνιστώσα «Συνδεσιμότητα»	78
7.4.1	Decision Tree (J48).....	78
7.4.2	Random Forest	79
7.4.3	Adaboost	79
7.4.4	Multilayer Perceptron	80
7.4.5	Support Vector Machine	81
7.5	Αποτελέσματα για τη συνιστώσα «Ενσωμάτωση Ψηφιακής Τεχνολογίας»	82
7.5.1	Decision Tree (J48).....	82
7.5.2	Random Forest	83
7.5.3	Adaboost	84
7.5.4	Multilayer Perceptron	85
7.5.5	Support Vector Machine	86
7.6	Αποτελέσματα για τη συνιστώσα «Ψηφιακές Δημόσιες Υπηρεσίες»	87
7.6.1	Decision Tree (J48).....	87
7.6.2	Random Forest	88
7.6.3	Adaboost	89
7.6.4	Multilayer Perceptron	89
7.6.5	Support Vector Machine	90

7.7 Παρατηρήσεις	91
8. Συμπεράσματα	94
9. Περιορισμοί της μελέτης και προτάσεις για περαιτέρω έρευνα	99
Βιβλιογραφία	101

1. Εισαγωγή

Η Ευρωπαϊκή Ένωση έχει θέσει ως προτεραιότητα να προετοιμάσει τις χώρες κράτη-μέλη για την νέα ψηφιακή εποχή, ενδυναμώνοντας και εκπαιδεύοντας τους πολίτες και τις επιχειρήσεις και παρέχοντας εργαλεία στην Δημόσια Διοίκηση για μια ομαλή μετάβαση στη νέα εποχή της πληροφορίας. Η ψηφιακή πρόοδος των κρατών-μελών αποτελεί βασικό άξονα πολιτικής της Ευρωπαϊκής Ένωσης για την επόμενη δεκαετία, με στόχο να διασφαλιστεί η πλήρης ψηφιακή μετάβαση, να ενισχυθεί η κοινωνική συνοχή και να προαχθεί η βιώσιμη οικονομική ανάπτυξη. Στο πλαίσιο αυτό, η Ευρωπαϊκή Επιτροπή έχει θεσπίσει τον Δείκτη Ψηφιακής Οικονομίας και Κοινωνίας (DESI), προκειμένου να μελετά και να παρακολουθεί την πρόοδο των κρατών-μελών στους τομείς της ψηφιακής ικανότητας του ανθρώπινου δυναμικού, της συνδεσιμότητας, της ενσωμάτωσης ψηφιακών τεχνολογιών και της παροχής ψηφιακών δημόσιων υπηρεσιών. Ο δείκτης DESI λειτουργεί ταυτόχρονα τόσο ως εργαλείο αξιολόγησης, αλλά και ως σημείο αναφοράς για τον σχεδιασμό πολιτικών που θα ενισχύσουν την ψηφιακή σύγκλιση εντός της Ε.Ε.

Η αξιοποίηση των δεδομένων του DESI προσφέρει τη δυνατότητα όχι μόνο αποτύπωσης της υφιστάμενης κατάστασης αλλά και το πεδίο έρευνας για την εξαγωγή προγνωστικών και ταξινομητικών συμπερασμάτων με χρήση διαφόρων μεθόδων στατιστικής ή εργαλείων μηχανικής μάθησης (machine learning). Η μηχανική μάθηση, η οποία ανήκει στον ευρύτερο κλάδο της τεχνητής νοημοσύνης, επιτρέπει αφενός την ανάδειξη προτύπων μέσα από μεγάλα σύνολα δεδομένων και αφετέρου, καθιστά εφικτή τη συγκριτική ανάλυση, την ομαδοποίηση και την πρόβλεψη του ψηφιακού μέλλοντος των κρατών-μελών.

Η παρούσα εργασία επιχειρεί να αναλύσει και να κατηγοριοποιήσει την ψηφιακή πρόοδο των χωρών της Ε.Ε., με βάση τα δεδομένα του δείκτη DESI για το έτος 2022, χρησιμοποιώντας τη μέθοδο της μηχανικής μάθησης, εφαρμόζοντας τόσο επιβλεπόμενες όσο και μη επιβλεπόμενες μεθόδους μηχανικής μάθησης με τη βοήθεια του λογισμικού Weka (Witten et al., 2016). Ως βοηθητικά εργαλεία χρησιμοποιήθηκαν το R (R Core Team, 2024) και το Excel, το πρώτο για επιμέρους απεικονίσεις και το δεύτερο για την οργάνωση και την επιμέρους οπτικοποίηση του υλικού. Παράλληλα, διερευνάται η συσχέτιση της ψηφιακής ανάπτυξης με άλλους δείκτες κοινωνικής, οικονομικής και βιώσιμης ανάπτυξης, αξιοποιώντας σχετικές επιστημονικές μελέτες που κατηγοριοποιούν τις χώρες της Ε.Ε. βάσει των ανωτέρω αναπτυξιακών προφίλ. Η προσέγγιση αυτή θα αναδείξει τους παράγοντες που επηρεάζουν περισσότερο την ψηφιακή εξέλιξη των κρατών καθώς και τις ανισότητες που εξακολουθούν να υπάρχουν εντός της Ε.Ε.

Επιδιώκεται, επομένως, η απάντηση στα παρακάτω ερευνητικά ερωτήματα: Ποια είναι σήμερα η χαρτογράφηση του ψηφιακού τοπίου της Ε.Ε.; Ποιες είναι οι βασικές μεταβλητές που καθορίζουν την ψηφιακή ετοιμότητα των κρατών-μελών της Ε.Ε.; Σε ποιο βαθμό τα δεδομένα του δείκτη DESI μπορούν να προβλέψουν ή να εξηγήσουν τις διαφοροποιήσεις μεταξύ των κρατών-μελών σε επίπεδο συνολικής ανάπτυξης;

Περίληπτικά, η εργασία οργανώνεται ως εξής:

Το **Κεφάλαιο 2** παρουσιάζει αναλυτικά τον δείκτη DESI και τις τέσσερις βασικές συνιστώσες του.

Το **Κεφάλαιο 3** εισάγει τις βασικές αρχές της μηχανικής μάθησης και αναλύει την μεθοδολογική προσέγγιση της παρούσας εργασίας.

Το **Κεφάλαιο 4** παρουσιάζει τη βιβλιογραφική επισκόπηση σχετικών ερευνών και την επιλογή τριών εξ αυτών για συσχέτιση με τα δεδομένα DESI 2022.

Στο **Κεφάλαιο 5** πραγματοποιείται η συσχέτιση των δεδομένων DESI με τα αποτελέσματα των παραπάνω ερευνών με τη χρήση μεθόδων επιβλεπόμενης μάθησης.

Στο **Κεφάλαιο 6** επιχειρείται ομαδοποίηση των κρατών-μελών με τεχνικές μη επιβλεπόμενης μάθησης.

Στο **Κεφάλαιο 7** εφαρμόζονται τεχνικές επιβλεπόμενης μάθησης στα δεδομένα DESI με τη χρήση ετικετών κλάσης που προέκυψαν στο κεφάλαιο 6

Στο **Κεφάλαιο 8** παρουσιάζονται τα γενικά συμπεράσματα της μελέτης και εν συνεχεία στο **Κεφάλαιο 9** η εργασία ολοκληρώνεται με τις προκλήσεις της έρευνας και προτάσεις για περαιτέρω μελέτη.

2. Αναλυτική Επισκόπηση του Δείκτη DESI (Digital Economy and Society Index) (European Commission, 2022)

Ο Δείκτης Ψηφιακής Οικονομίας και Κοινωνίας (DESI) είναι ένας σύνθετος δείκτης που δημιουργήθηκε από την Ευρωπαϊκή Επιτροπή για να παρακολουθεί την ψηφιακή πρόοδο των κρατών-μελών της Ευρωπαϊκής Ένωσης. Ο δείκτης αξιολογεί την ψηφιακή ετοιμότητα και την απόδοση των χωρών σε βασικούς τομείς που σχετίζονται με τις ψηφιακές τεχνολογίες και αποτελείται από τέσσερις βασικές συνιστώσες, οι οποίες περιλαμβάνουν διάφορους επιμέρους δείκτες, όπως περιγράφονται παρακάτω:

2.1 Ανθρώπινο Κεφάλαιο (Human Capital)

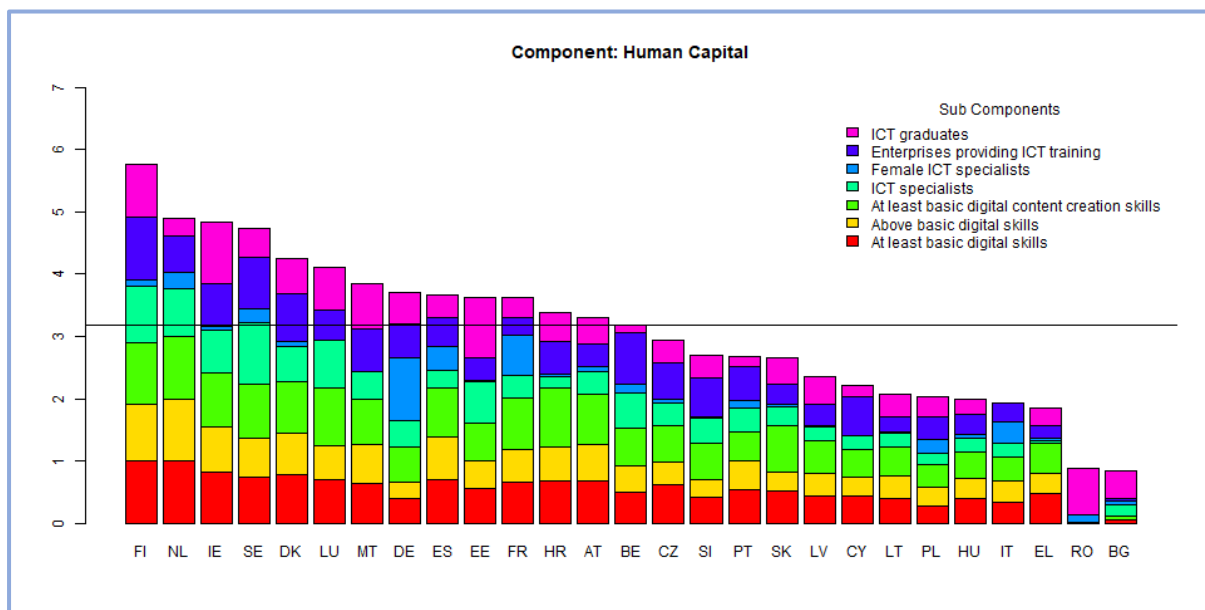
Αξιολογεί τις ψηφιακές δεξιότητες του πληθυσμού και την ανάπτυξη επαγγελματικών δεξιοτήτων στον τομέα των Τεχνολογιών Πληροφορικής και Επικοινωνίας (ΤΠΕ).

2.1.1 Υποδιαστάσεις & Δείκτες:

- Δεξιότητες χρήσης διαδικτύου (Internet User Skills)
 - Τουλάχιστον βασικές ψηφιακές δεξιότητες (At least basic digital skills) (% ατόμων)
 - Ψηφιακές δεξιότητες πάνω από το βασικό επίπεδο (Above basic digital skills) (% ατόμων)
 - Τουλάχιστον βασικές δεξιότητες δημιουργίας ψηφιακού περιεχομένου (At least basic digital content creation skills) (% ατόμων)
- Προχωρημένες δεξιότητες & ανάπτυξη (Advanced Skills & Development)
 - Ειδικοί ΤΠΕ (ICT specialists) (% απασχολούμενων ατόμων ηλικίας 15-74 ετών)
 - Γυναίκες ειδικοί ΤΠΕ (Female ICT specialists) (% των ειδικών ΤΠΕ)
 - Επιχειρήσεις που παρέχουν εκπαίδευση ΤΠΕ στο προσωπικό (Enterprises providing ICT training)(% επιχειρήσεων)
 - Απόφοιτοι ΤΠΕ (ICT graduates) (% των συνολικών αποφοίτων)

Σύμφωνα με τα την έκθεση (European Commission, 2022), το 87% των Ευρωπαίων χρησιμοποιεί το διαδίκτυο, αλλά μόνο το 54% διαθέτει βασικές ψηφιακές δεξιότητες, ταυτόχρονα, υπάρχει έλλειψη ειδικών ΤΠΕ (ICT specialists), ενώ μόνο το 19% αυτών είναι γυναίκες. Οι χώρες με τις υψηλότερες επιδόσεις είναι η Φινλανδία και η Ολλανδία, ενώ η Ρουμανία και η Βουλγαρία βρίσκονται στο χαμηλότερο επίπεδο. Στόχος της ΕΕ μέχρι το 2030

είναι, τουλάχιστον το 80% του πληθυσμού, να διαθέτει βασικές ψηφιακές δεξιότητες και να φτάσει τους 20 εκατομμύρια ειδικούς ΤΠΕ.



Γράφημα 1. Δεξιότητες ανθρώπινου κεφαλαίου ανά κράτος-μέλος της Ε.Ε. Οι τιμές κάθε δείκτη (ICT graduates, ICT specialists κ.λπ.) έχουν τυποποιηθεί στην ίδια κλίμακα (min-max normalization) για να επιτρέπουν ευθέως συγκρίσιμες μετρήσεις μεταξύ των διαφορετικών δεικτών. Οι χώρες είναι διατεταγμένες (από αριστερά προς τα δεξιά) με βάση το συνολικό άθροισμα της απόδοσής τους σε όλα τα χαρακτηριστικά. Η οριζόντια γραμμή αντιπροσωπεύει τη διάμεσο αυτού του αθροίσματος. Συνολικά τις ανώτερες επιδόσεις εμφανίζουν η Φινλανδία, η Ολλανδία και η Ιρλανδία και τις χαμηλότερες η Βουλγαρία, η Ρουμανία και η Ελλάδα.

2.2 Συνδεσιμότητα (Connectivity)

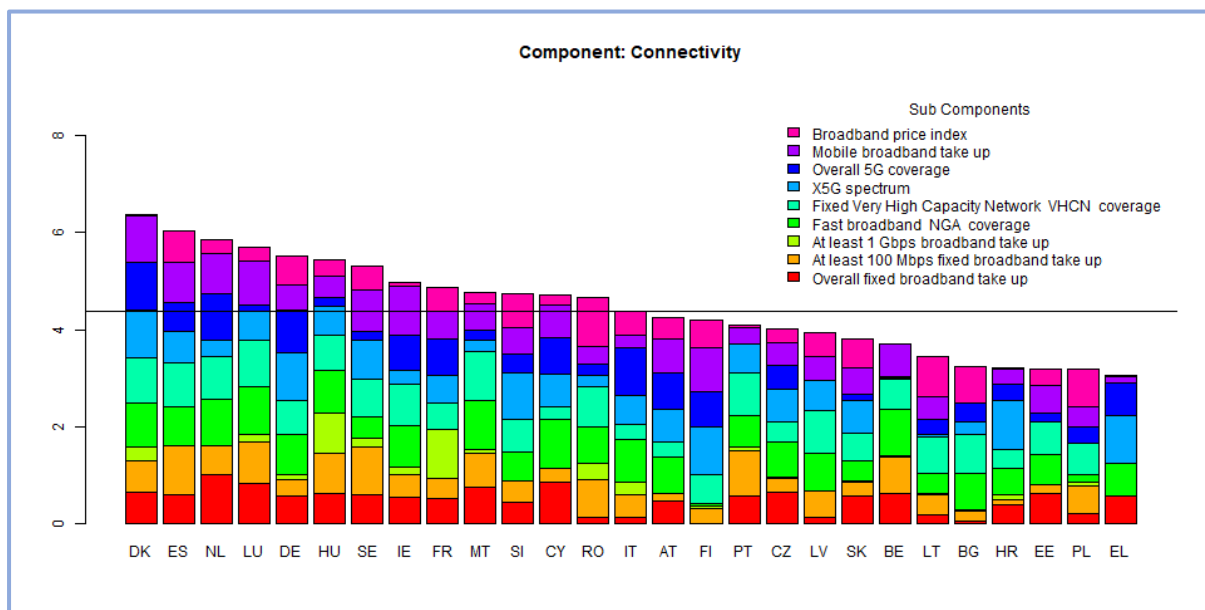
Αξιολογεί την ποιότητα και προσβασιμότητα των δικτύων σταθερής και κινητής ευρυζωνικότητας.

2.2.1 Υποδιαστάσεις & Δείκτες:

- Υιοθέτηση σταθερής ευρυζωνικότητας (Fixed Broadband Take-up)
 - Συνδρομές σταθερής ευρυζωνικής σύνδεσης (At least 100 Mbps fixed broadband take-up) (% νοικοκυριών)
 - Συνδρομές σταθερής ευρυζωνικής σύνδεσης ≥ 100 Mbps (At least 100 Gbps broadband take-up) (% νοικοκυριών)
 - Συνδρομές σταθερής ευρυζωνικής σύνδεσης ≥ 1 Gbps (At least 100 Gbps broadband take-up) (% νοικοκυριών)
- Κάλυψη σταθερής ευρυζωνικότητας (Fixed Broadband Coverage)
 - Κάλυψη ταχείας ευρυζωνικότητας (Fast broadband (NGA) coverage) (% νοικοκυριών)

- Κάλυψη δικτύων πολύ υψηλής ταχύτητας (Fixed very high capacity network (VHCN) coverage) (% νοικοκυριών)
- Κινητή ευρυζωνικότητα (Mobile Broadband)
 - Φάσμα 5G (5G spectrum) (% εκχωρημένου φάσματος για 5G)
 - Γενική κάλυψη 5G (Overall 5G coverage)(% πληθυσμού)
 - Χρήση κινητής ευρυζωνικότητας (Mobile broadband take-up) (συνδρομές ανά 100 άτομα)
- Τιμές ευρυζωνικότητας (Broadband Prices)
 - Δείκτης τιμών ευρυζωνικού δικτύου (Βαθμολογία 0-100)

Σήμερα η κάλυψη σταθερών ευρυζωνικών δικτύων υψηλής ταχύτητας (VHCN) έχει φτάσει το 70% των νοικοκυριών στην ΕΕ. Το 5G καλύπτει το 66% των κατοικημένων περιοχών, αλλά μόνο το 56% του διαθέσιμου φάσματος 5G έχει κατανομηθεί στα κράτη μέλη. Η Δανία, η Ισπανία και η Ολλανδία παρουσιάζουν τις καλύτερες επιδόσεις όσον αφορά στη συνδεσιμότητα, ενώ η Ελλάδα παραμένει μία από τις χώρες με τις χαμηλότερες επιδόσεις (Γράφημα 2.). Στόχος της ΕΕ είναι η πλήρης κάλυψη gigabit και η ανάπτυξη του 5G σε όλες τις αστικές περιοχές μέχρι το 2030.



Γράφημα 2. Συνδεσιμότητα ανά κράτος-μέλος της Ε.Ε. Οι τιμές κάθε δείκτη (Broadband price index, 5G spectrum κ.λπ.) έχουν τυποποιηθεί στην ίδια κλίμακα (min-max normalization) για να επιτρέπουν ευθέως συγκρίσιμες μετρήσεις μεταξύ των διαφορετικών δεικτών. Οι χώρες είναι διατεταγμένες (από αριστερά προς τα δεξιά) με βάση το συνολικό άθροισμα της απόδοσής τους σε όλα τα χαρακτηριστικά. Η οριζόντια γραμμή αντιπροσωπεύει τη διάμεσο αυτού του αθροίσματος. Συνολικά τις ανώτερες επιδόσεις εμφανίζουν η Δανία, η Ισπανία και η Ολλανδία και τις χαμηλότερες η Ελλάδα, η Πολωνία και η Εσθονία.

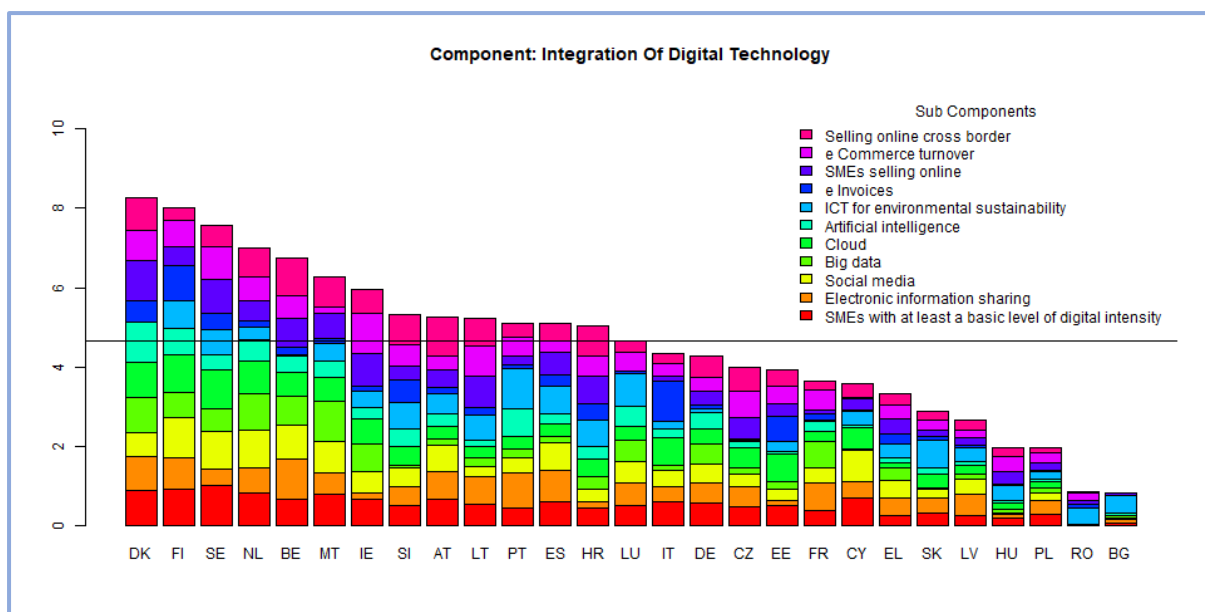
2.3 Ενσωμάτωση Ψηφιακής Τεχνολογίας (Integration of Digital Technology)

Μετρά την υιοθέτηση και χρήση ψηφιακών τεχνολογιών από επιχειρήσεις.

2.3.1 Υποδιαστάσεις και Δείκτες

- Ψηφιακή ένταση (Digital Intensity)
 - ΜΜΕ με τουλάχιστον βασικό επίπεδο ψηφιακής έντασης (SMEs with at least a basic level of digital intensity) (% ΜΜΕ)
- Ψηφιακές τεχνολογίες στις επιχειρήσεις (Digital Technologies for Businesses)
 - Ηλεκτρονική ανταλλαγή πληροφοριών (Electronic information sharing) (% επιχειρήσεων)
 - Χρήση Κοινωνικών Δικτύων (Social media) (% επιχειρήσεων)
 - Ανάλυση μεγάλων δεδομένων (Big Data) (% επιχειρήσεων)
 - Υπηρεσίες Cloud (Cloud)(% επιχειρήσεων)
 - Τεχνητή Νοημοσύνη (Artificial Intelligence) (% επιχειρήσεων)
 - ΤΠΕ για περιβαλλοντική βιωσιμότητα (ICT for environmental sustainability) (% επιχειρήσεων)
 - Ηλεκτρονικά τιμολόγια (e-Invoices) (% επιχειρήσεων)
- Ηλεκτρονικό εμπόριο (E-Commerce)
 - ΜΜΕ που πωλούν online (SMEs selling online) (% ΜΜΕ)
 - Πωλήσεις ηλεκτρονικού εμπορίου (e-commerce turnover) (% συνολικών πωλήσεων)
 - Διασυνοριακές online πωλήσεις (selling online cross-border) (% ΜΜΕ)

Σύμφωνα με στοιχεία της έκθεσης (European Commission, 2022), μόνο το 55% των μικρομεσαίων επιχειρήσεων (ΜΜΕ) διαθέτει βασικό επίπεδο ψηφιακής έντασης, με στόχο της Ευρωπαϊκής Επιτροπής το ποσοστό αυτό να φτάσει το 90% έως το 2030. Επιπλέον, μόνο το 8% των επιχειρήσεων χρησιμοποιεί τεχνητή νοημοσύνη, 34% cloud computing και 14% big data. Η Φινλανδία, η Δανία και η Σουηδία διαθέτουν τον υψηλότερο βαθμό ψηφιακής ενσωμάτωσης στις επιχειρήσεις.



Γράφημα 3. Ενσωμάτωση Ψηφιακών τεχνολογιών ανά κράτος-μέλος της Ε.Ε. Οι τιμές κάθε δείκτη (Selling online cross border, e-commerce turnover κ.λπ.) έχουν τυποποιηθεί στην ίδια κλίμακα (min-max normalization) για να επιτρέπουν ευθέως συγκρίσιμες μετρήσεις μεταξύ των διαφορετικών δεικτών. Οι χώρες είναι διατεταγμένες (από αριστερά προς τα δεξιά) με βάση το συνολικό άθροισμα (sum) της απόδοσής τους σε όλα τα χαρακτηριστικά. Η οριζόντια γραμμή αντιπροσωπεύει τη διάμεσο αυτού του αθροίσματος. Συνολικά τις ανώτερες επιδόσεις εμφανίζουν η Δανία, η Φινλανδία και η Σουηδία και τις χαμηλότερες η Βουλγαρία, η Ρουμανία και η Πολωνία.

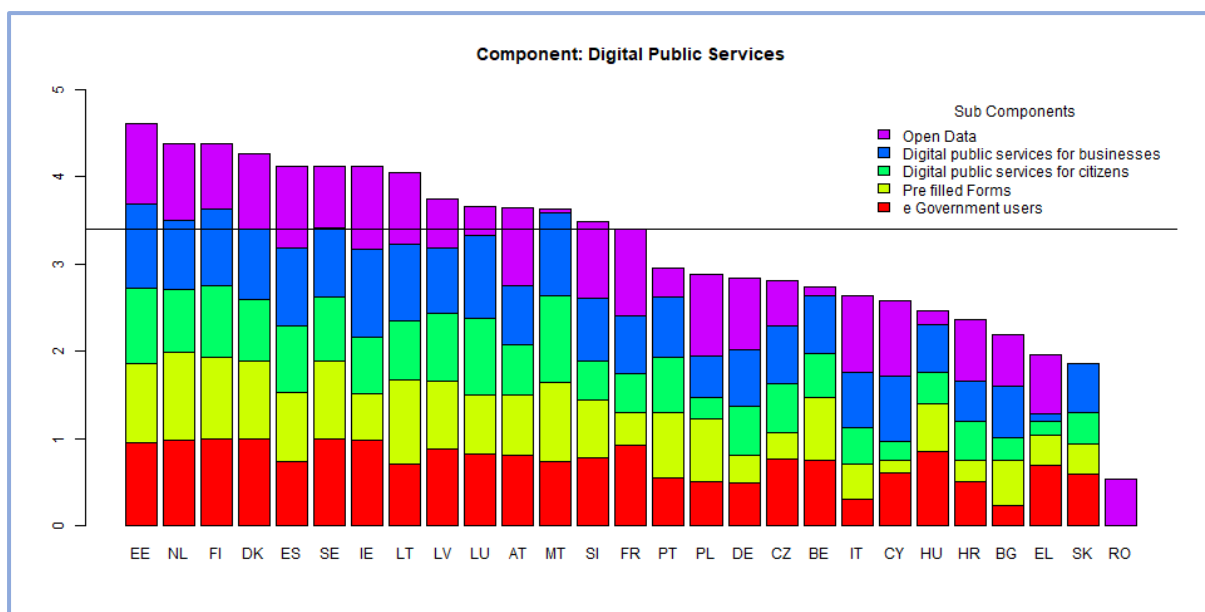
2.4 Ψηφιακές Δημόσιες Υπηρεσίες (Digital Public Services)

Αξιολογεί την παροχή ψηφιακών υπηρεσιών από το δημόσιο.

2.4.1 Υποδιαστάσεις και Δείκτες:

- Ηλεκτρονική διακυβέρνηση (e-Government)
 - Χρήστες ηλεκτρονικής διακυβέρνησης (e-Government users) (% ατόμων)
 - Προσυμπληρωμένες φόρμες (Pre-filled forms) (Βαθμολογία 0-100)
 - Ψηφιακές δημόσιες υπηρεσίες για πολίτες (Digital public services for citizens) (Βαθμολογία 0-100)
 - Ψηφιακές δημόσιες υπηρεσίες για επιχειρήσεις (Digital public services for businesses) (Βαθμολογία 0-100)
 - Ανοιχτά δεδομένα (Open data) (% μέγιστης δυνατής βαθμολογίας)

Σήμερα το ποσοστό των πολιτών που χρησιμοποιεί e-government υπηρεσίες φτάνει το 75%, ενώ για τις επιχειρήσεις το αντίστοιχο ποσοστό φτάνει το 82%. Οι χώρες ηγέτες στην ηλεκτρονική διακυβέρνηση είναι η Εσθονία, η Φινλανδία και η Ολλανδία, ενώ η Ελλάδα, Σλοβακία και η Ρουμανία απέχουν ακόμα από τους στόχους της ΕΕ για πλήρη παροχή ψηφιακών δημόσιων υπηρεσιών έως το 2030.



Γράφημα 4. Ψηφιακές δημόσιες υπηρεσίες ανά κράτος-μέλος της Ε.Ε. Οι τιμές κάθε δείκτη (Open Data, e-Government Users κ.λπ.) έχουν τυποποιηθεί στην ίδια κλίμακα (min-max normalization) για να επιτρέπουν ευθέως συγκρίσιμες μετρήσεις μεταξύ των διαφορετικών δεικτών. Οι χώρες είναι διατεταγμένες (από αριστερά προς τα δεξιά) με βάση το συνολικό άθροισμα (sum) της απόδοσής τους σε όλα τα χαρακτηριστικά. Η οριζόντια γραμμή αντιπροσωπεύει τη διάμεσο αυτού του αθροίσματος. Συνολικά τις ανώτερες επιδόσεις εμφανίζουν η Εσθονία, η Ολλανδία και η Φινλανδία και τις χαμηλότερες η Ελλάδα, η Σλοβακία και η Ρουμανία.

Συνολικά, λαμβάνονται μέτρα για την συνεχή ψηφιακή πρόοδο των κρατών μελών, και οι διαφορές μεταξύ των χωρών με υψηλές και χαμηλές επιδόσεις μειώνονται σταδιακά, ωστόσο εξακολουθούν να υπάρχουν ανισότητες, οι οποίες απαιτούν επιπλέον επενδύσεις αλλά και πολιτικές τομές. Προς αυτή την κατεύθυνση, η Ευρωπαϊκή Ένωση ενισχύει τη στρατηγική της για την Ψηφιακή Δεκαετία 2030, με έμφαση στις υποδομές, τις δεξιότητες και τις ψηφιακές υπηρεσίες.

2.5 Δομή και χαρακτηριστικά του συνόλου δεδομένων DESI 2022

Το σύνολο δεδομένων που χρησιμοποιείται στην παρούσα μελέτη βασίζεται στον Δείκτη Ψηφιακής Οικονομίας και Κοινωνίας (DESI) για το έτος 2022, όπως δημοσιεύεται από την Ευρωπαϊκή Επιτροπή (European Commission, 2022). Το αρχείο δεδομένων είναι διαρθρωμένο σε μορφή πίνακα (tabular form), όπου κάθε γραμμή αντιστοιχεί σε ένα κράτος-μέλος και κάθε στήλη αντιπροσωπεύει έναν δείκτη-χαρακτηριστικό του DESI 2022. Τα σύνολο δεδομένων αποτελείται εξ ολοκλήρου από αριθμητικά στοιχεία, χωρίς ελλιπείς τιμές, γεγονός που τα καθιστά κατάλληλα για επεξεργασία με μεθόδους μηχανικής μάθησης, τόσο επιβλεπόμενης όσο και μη επιβλεπόμενης.

Ωστόσο, παρά τα πλεονεκτήματά του, το σύνολο δεδομένων (dataset) παρουσιάζει έναν σημαντικό περιορισμό που αφορά στις προκλήσεις που προκύπτουν όταν το πλήθος των

χαρακτηριστικών (features) αυξάνεται σημαντικά σε σχέση με τον αριθμό των παρατηρήσεων. Το φαινόμενο αυτό ονομάζεται «κατάρα της διαστατικότητας» (curse of dimensionality) (Bishop, 2006). Καθώς ο DESI περιλαμβάνει πολυάριθμες μεταβλητές με κάθε συνιστώσα να αναλύεται περαιτέρω σε επιμέρους δείκτες, η ανάλυση πραγματοποιείται σε έναν χώρο υψηλής διάστασης. Ως εκ τούτου, η πυκνότητα των δεδομένων μειώνεται δραστικά καθώς αυξάνονται οι διαστάσεις — τα σημεία απομακρύνονται μεταξύ τους, κάνοντας πιο δύσκολη τη διάκριση και την εύρεση συστάδων. Αυτό καθιστά τις εργασίες ομαδοποίησης και ταξινόμησης λιγότερο αποδοτικές αλλά αναγκαίες, καθώς οι παραδοσιακές στατιστικές τεχνικές αποτυγχάνουν όταν δεν υπάρχουν επαρκείς βαθμοί ελευθερίας. (Aggarwal, Hinneburg, & Keim, 2001). Συνέπεια των παραπάνω, αποτελεί και η αύξηση του κινδύνου υπερπροσαρμογής (overfitting) καθώς το μοντέλο ενδέχεται να ταιριάζει υπερβολικά στα δεδομένα εκπαίδευσης χωρίς να επιτυγχάνει να γενικεύει σωστά σε νέα δεδομένα (Bishop, 2006).

Για τον λόγο αυτό, εφαρμόζονται τεχνικές μείωσης διαστατικότητας, όπως η επιλογή χαρακτηριστικών (feature selection) με βάση τη σπουδαιότητα των μεταβλητών (Hastie, Tibshirani, & Friedman, 2009), η οποία θα χρησιμοποιηθεί στο πλαίσιο της εργασίας αυτής, με στόχο τη βελτίωση της ακρίβειας και της δυνατότητας ερμηνείας των αποτελεσμάτων.

3. Αρχές Μηχανικής Μάθησης

Η Μηχανική Μάθηση (Machine Learning) αποτελεί ένα ταχέως αναπτυσσόμενο κλάδο της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI), ο οποίος χρησιμοποιείται ευρέως για την ανάλυση δεδομένων, τη μοντελοποίηση συστημάτων καθώς και την πρόβλεψη τάσεων, με στόχο να προσφέρει πολύτιμες πληροφορίες στη λήψη αποφάσεων αλλά και τη χάραξη πολιτικών. Η παρούσα διπλωματική εργασία, επικεντρώνεται στη χρήση της μηχανικής μάθησης για τη δημιουργία συσχετίσεων μεταξύ της ψηφιακής προόδου των κρατών-μελών (με δεδομένα του δείκτη DESI για το 2022) και της ανάπτυξης των χωρών αυτών τόσο σε κοινωνικό και οικονομικό επίπεδο όσο και στα ευρύτερα πλαίσια της βιώσιμης ανάπτυξης. Η μηχανική μάθηση μπορεί να αποτελέσει ιδιαίτερα χρήσιμο εργαλείο στη μελέτη μεγάλων συνόλων δεδομένων, στην ανίχνευση προτύπων και να συμβάλλει στη λήψη τεκμηριωμένων αποφάσεων (Mitchell, 1997).

Η εφαρμογή τεχνικών μηχανικής μάθησης στην ανάλυση αυτών των δεδομένων του δείκτη DESI, διευκολύνει την κατηγοριοποίηση των κρατών-μελών της Ε.Ε., την ανίχνευση ομάδων χωρών με παρόμοια χαρακτηριστικά και την πρόβλεψη της μελλοντικής τους εξέλιξης. Στην παρούσα μελέτη, χρησιμοποιούνται τόσο επιβλεπόμενες όσο και μη επιβλεπόμενες μέθοδοι προκειμένου να διερευνηθεί η δυναμική της ψηφιακής ανάπτυξης στην Ε.Ε.

3.1 Βασικές Κατηγορίες Μηχανικής Μάθησης

Οι αλγόριθμοι μηχανικής μάθησης ταξινομούνται κατά κύριο λόγο σε επιβλεπόμενους (supervised learning), μη επιβλεπόμενους (unsupervised learning) και ενισχυτικούς (reinforcement learning), ανάλογα με τη φύση των δεδομένων που επεξεργάζονται και τον τρόπο με τον οποίο "εκπαιδεύονται" (Hastie, Tibshirani, & Friedman, 2009).

Στην επιβλεπόμενη μάθηση, το σύστημα εκπαιδεύεται με δεδομένα που έχουν προκαθορισμένες ετικέτες κλάσης (labels) – στόχους, και σκοπός της ανάλυσης είναι η πρόβλεψη του παραπάνω στόχου (Provost & Fawcett, 2013/2019). Αυτό καθιστά τους αλγορίθμους επιβλεπόμενης μάθησης κατάλληλους για προβλήματα ταξινόμησης (όταν ο στόχος είναι κατηγορικός, όπως η πρόβλεψη του επιπέδου ανάπτυξης μιας χώρας βάσει των ιστορικών δεδομένων ψηφιακής προόδου) και παλινδρόμησης όταν ο στόχος είναι αριθμητικός.

Ωστόσο, στη μη επιβλεπόμενη μάθηση, δεν υπάρχουν προκαθορισμένες κατηγορίες ή ετικέτες. Οι αλγόριθμοι, στην περίπτωση αυτή, αναζητούν μοτίβα, κοινά χαρακτηριστικά και συσχετίσεις μεταξύ των δεδομένων, ώστε να ανιχνεύσουν δομές που δεν ήταν προηγουμένως προφανείς και τη δημιουργία ομάδων με όμοια χαρακτηριστικά (Bishop, 2006). Η συσταδοποίηση (clustering), για παράδειγμα, επιτρέπει την ομαδοποίηση των χωρών της Ε.Ε. με βάση την ομοιότητά τους στον δείκτη DESI.

Η ενισχυτική μάθηση (Reinforcement Learning), αφορά την αλληλεπίδραση ενός αλγορίθμου με το περιβάλλον του και τη συνεχή προσαρμογή του με βάση ενός συστήματος ανταμοιβών (Sutton & Barto, 2018) αλλά δεν εφαρμόζεται στην ανάλυσή μας.

3.1.1 Επιβλεπόμενη Μάθηση και Ανάλυση του Δείκτη DESI

Η επιβλεπόμενη μάθηση είναι ιδιαίτερα χρήσιμη για την πρόβλεψη της ψηφιακής εξέλιξης των κρατών-μελών και την ταξινόμησή τους σε προκαθορισμένες κατηγορίες. Στην παρούσα μελέτη χρησιμοποιούνται διάφοροι αλγόριθμοι ταξινόμησης, με σκοπό την επίτευξη υψηλότερης ακρίβειας αλλά και την ανάδειξη των αλγορίθμων που επιτυγχάνουν τα καλύτερα αποτελέσματα. Οι αλγόριθμοι που επιλέχθηκαν και οι οποίοι παρουσιάζουν διαφορετικά χαρακτηριστικά και εφαρμογές είναι οι Decision Tree (J48), Random Forest (RDF), Adaboost (ADA), Multilayer Perceptron (MLP) και Support Vector Machine (SVM). Για την αξιολόγηση της απόδοσής τους, εφαρμόζουμε τη μέθοδο της διασταυρούμενης επικύρωσης (*cross-validation*), ώστε να διασφαλίσουμε την αξιοπιστία των αποτελεσμάτων. Οι μετρικές που χρησιμοποιούμε για την αποτίμηση της απόδοσης των μοντέλων είναι οι εξής (Walker, 2023):

- **Accuracy:** Το συνολικό ποσοστό σωστών ταξινομήσεων, είτε θετικών είτε αρνητικών.
- **Precision:** Μετρά την ακρίβεια των θετικών προβλέψεων. Απαντά στο ερώτημα: «Από όλα τα στοιχεία που το μοντέλο χαρακτήρισε ως θετικά, πόσα ήταν πραγματικά θετικά;».
- **Recall:** Μετρά την ικανότητα του μοντέλου να βρίσκει όλες τις θετικές περιπτώσεις. Απαντά στο ερώτημα: «Από όλα τα πραγματικά θετικά, πόσα αναγνώρισε σωστά το μοντέλο;».
- **F1-score:** Ο αρμονικός μέσος των Precision και Recall του μοντέλου, ο οποίος εξισορροπεί τις δύο μετρικές σε έναν ενιαίο αριθμό.

Περιγραφή Αλγορίθμων

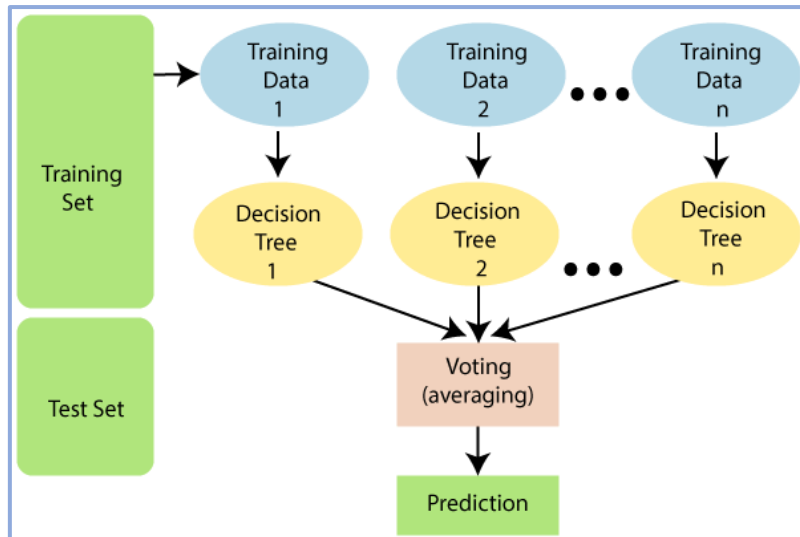
Αλγόριθμος Decision Tree J48

Ο αλγόριθμος Decision Tree (J48), προτάθηκε από τον Quinlan (1993) και αποτελεί υλοποίηση του αλγορίθμου C4.5 Decision Tree. Ο αλγόριθμος αυτός, δημιουργεί δενδροειδείς δομές που επιτρέπουν την εύκολη ερμηνεία των αποφάσεων, γεγονός που τον καθιστά ιδιαίτερα χρήσιμο στην ανάλυση και ερμηνεία των αποτελεσμάτων ταξινόμησης (Quinlan, 1993). Το μοντέλο διαχωρίζει τα δεδομένα σε κόμβους και κλαδιά βάσει των πιο σημαντικών χαρακτηριστικών τους, προκειμένου να τα ταξινομήσει. Στη φάση της εκπαίδευσής του, επιλέγει το χαρακτηριστικό που διαχωρίζει καλύτερα τα δεδομένα (ρίζα) και στη συνέχεια δημιουργούνται κόμβοι και κλαδιά για κάθε δυνατή τιμή του χαρακτηριστικού. Η παραπάνω διαδικασία επαναλαμβάνεται για κάθε κόμβο με επιλογή νέων χαρακτηριστικών μέχρι να δημιουργηθεί ένα ολοκληρωμένο δέντρο. Στη συνέχεια, ακολουθεί το κλάδεμα (pruning) του δέντρου όπου αφαιρούνται τα μη χρήσιμα κλαδιά προς αποφυγή της υπερπροσαρμογής. Η διαδικασία της εκπαίδευσης ακολουθείται από τη διαδικασία της ταξινόμησης, όπου ο αλγόριθμος ξεκινώντας από τη ρίζα του δέντρου και προχωρώντας σε κόμβους και κλαδιά, προσπαθεί να προβλέψει την κατηγορία μιας νέας περίπτωσης. Ο αλγόριθμος Decision Tree (J48) είναι ένας απλός στην λειτουργία ταξινομητής, γρήγορος και εύκολα ερμηνεύσιμος, ταυτόχρονα όμως παρουσιάζει περιορισμούς. Οι σημαντικότερες προκλήσεις για την αποτελεσματικότητα του αλγορίθμου είναι η δημιουργία κενών κλαδιών που δεν συμβάλλουν στην ταξινόμηση, η ύπαρξη κλαδιών που δεν παρέχουν σημαντικές πληροφορίες για την ταξινόμηση και αυξάνουν το πρόβλημα της υπερπροσαρμογής (overfitting) και τέλος η τάση του αλγορίθμου να ταξινομεί τέλεια τα δεδομένα εκπαίδευσης αλλά να μην έχει καλή προβλεπτική ικανότητα σε νέα δεδομένα (υπερπροσαρμογή) (Saravanan & Gayathri, 2018).

Αλγόριθμος Random Forest

Ο αλγόριθμος Random Forest είναι μια μέθοδος που βασίζεται στη δημιουργία δέντρων απόφασης (Decision Trees) όπως και ο J48 με τη διαφορά ότι συνδυάζει πολλά δέντρα απόφασης, με αποτέλεσμα να βελτιώνει την ακρίβεια της πρόβλεψης και γι' αυτό αποτελεί ένα πιο αξιόπιστο αλγόριθμο ταξινόμησης (Breiman, 2001). Κατά τη διάρκεια της εκπαίδευσής του, ο αλγόριθμος δημιουργεί τυχαία υποσύνολα δεδομένων μέσω bagging και έπειτα για κάθε υποσύνολο κατασκευάζει ένα ανεξάρτητο δέντρο απόφασης. Εν τέλει, η προβλεπτική ικανότητα του μοντέλου βασίζεται στην πλειοψηφική ψηφοφορία (Breiman, 2001) (Εικόνα 1). Ο αλγόριθμος Random Forest είναι κατάλληλος για μεγάλα σύνολα δεδομένων καθώς έχει την

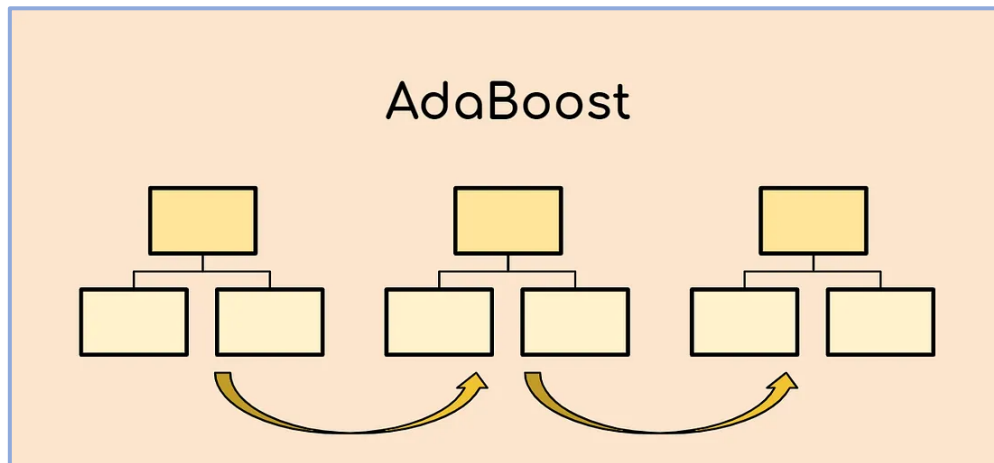
ικανότητα να μειώνει το πρόβλημα της υπερπροσαρμογής (overfitting) (Breiman, 2001) αλλά είναι πιο αργός από ένα απλό δέντρο απόφασης.



Εικόνα 1. Διάγραμμα λειτουργίας αλγόριθμου Random Forest: ο αλγόριθμος Random Forest, κατά τη διάρκεια της εκπαίδευσής του δημιουργεί πολλαπλά δέντρα απόφασης και συνδυάζοντας τις εξόδους τους βελτιώνει την ακρίβεια της πρόβλεψης και περιορίζει την υπερπροσαρμογή. (“Random Forest Algorithm - Javatpoint,” 2021).

Αλγόριθμος Adaboost

Ο αλγόριθμος Adaboost (συντομογραφία του Adaptive Boosting) δημιουργήθηκε από τους Freund και Schapire (1997) και αποτελεί έναν αλγόριθμο που έχει την ικανότητα να βελτιώνει-ενισχύει την απόδοση των αδύναμων ταξινομητών, δημιουργώντας ένα ισχυρότερο μοντέλο. Η φάση εκπαίδευσης ξεκινά από τη δημιουργία ενός απλού ταξινομητή (π.χ. ένα δέντρο απόφασης), στη συνέχεια τα δεδομένα που ταξινομούνται εσφαλμένα λαμβάνουν αυξημένο βάρος, ενώ εκείνα που ταξινομούνται σωστά λαμβάνουν μειωμένο βάρος σε σχέση με τα προηγούμενα. Η διαδικασία επαναλαμβάνεται με τον αλγόριθμο να δίνει μεγαλύτερη έμφαση στα δεδομένα που ταξινομήθηκαν λανθασμένα, ώστε να διορθωθούν από τους επόμενους ταξινομητές (Εικόνα 2). Το τελικό μοντέλο είναι ένας γραμμικός συνδυασμός των επιμέρους ταξινομητών (Hastie, Rosset, Zhu, & Zou, 2009). Ο αλγόριθμος Adaboost παρέχει αυξημένη ακρίβεια, δεν είναι ιδιαίτερα ευαίσθητος στην υπερπροσαρμογή και λειτουργεί ικανοποιητικά χωρίς παραμετροποίηση. Ωστόσο, είναι ευαίσθητος σε θόρυβο και ακραίες τιμές και μπορεί να είναι αργός για μεγάλα σύνολα δεδομένων (Kurama, 2024).

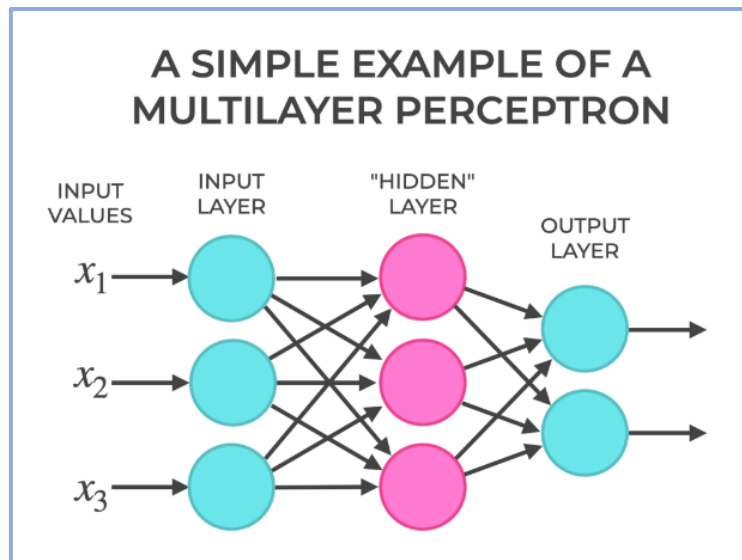


Εικόνα 2. Διάγραμμα λειτουργίας Adaboost: το διάγραμμα παρουσιάζει τη δυνατότητα του αλγορίθμου Adaboost να συνδυάζει πολλαπλούς αδύναμους ταξινομητές για να σχηματίσει έναν ισχυρό ταξινομητή προσαρμόζοντας κατ' επανάληψη τα βάρη των λανθασμένα ταξινομημένων περιπτώσεων (Verma, 2022).

Αλγόριθμος Multilayer Perceptron

Ο Multilayer Perceptron (MLP) ανήκει στην κατηγορία των τεχνητών νευρωνικών δικτύων (Artificial Neural Networks - ANN) και αποτελεί ιδανική επιλογή για την αναγνώριση μη γραμμικών σχέσεων μεταξύ των μεταβλητών, επιτρέποντας την κατασκευή μοντέλων υψηλής προγνωστικής ικανότητας και μεγάλης ακρίβειας (Haykin, 2009). Ο Multilayer Perceptron αποτελείται από τρία κύρια μέρη: α. Την είσοδο, όπου γίνεται η υποδοχή των δεδομένων, β. τα κρυφά επίπεδα (hidden layers), όπου ο αλγόριθμος ανακαλύπτει μοτίβα και σχέσεις ανάμεσα στα δεδομένα χρησιμοποιώντας μαθηματικές συναρτήσεις και γ. την έξοδο, που παρέχει το τελικό αποτέλεσμα (Εικόνα 3). Ο αλγόριθμος εκπαιδεύεται με μία μέθοδο που ονομάζεται backpropagation. Ουσιαστικά, ο αλγόριθμος κάνει μια εικασία, αξιολογεί πόσο έκανε λάθος και έπειτα επιστρέφει για να διορθώσει τις παραμέτρους και να μην επαναλάβει το ίδιο λάθος, με τη διαδικασία αυτή να επαναλαμβάνεται πολλές φορές (Bishop, 2006).

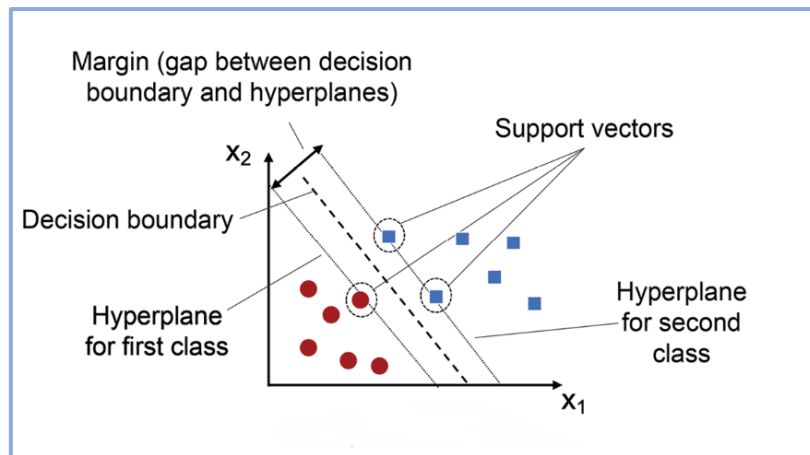
Η βασική μορφή του MLP που χρησιμοποιήθηκε στην μελέτη αυτή και χρησιμοποιείται και σε πολλές εφαρμογές σήμερα χαρακτηρίζεται ως "shallow neural network", καθώς περιλαμβάνει λίγα κρυφά επίπεδα (στην περίπτωση μας 3). Αντιθέτως, τα σύγχρονα βαθιά νευρωνικά δίκτυα (deep neural networks – DNNs), τα οποία αποτελούν τη μετεξέλιξη του MLP, περιλαμβάνουν πολλαπλά κρυφά στρώματα και αποτελούν τη βάση της τεχνολογίας deep learning, η οποία κυριαρχεί σήμερα σε ποικίλα πεδία τεχνητής νοημοσύνης (LeCun, Bengio, & Hinton, 2015).



Εικόνα 3. Διάγραμμα λειτουργίας Multilayer Perceptron: Απεικονίζεται αρχιτεκτονική ενός MLP, δείχνοντας τα επίπεδα εισόδου, κρυφά επίπεδα και το επίπεδο εξόδου, καθώς και τις συνδέσεις μεταξύ των νευρώνων (Ebner, 2023).

Αλγόριθμος Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης, ο οποίος είναι κατάλληλος τόσο για ταξινόμηση όσο και για παλινδρόμηση. Η αρχή της λειτουργίας του είναι απλή, επιδιώκει να βρίσκει κάθε φορά ένα υπερεπίπεδο (hyperlane) που χωρίζει τα δεδομένα διαφορετικών κλάσεων με το μεγαλύτερο δυνατό περιθώριο (Εικόνα 4) (Cortes & Vapnik, 1995). Στην περίπτωση που τα δεδομένα είναι γραμμικά διαχωρίσιμα, ο SVM βρίσκει μια ευθεία (ή υπερεπίπεδο) που τα χωρίζει με το μέγιστο περιθώριο. Ωστόσο, αν τα δεδομένα δεν διαχωρίζονται γραμμικά, ο SVM χρησιμοποιεί συναρτήσεις πυρήνα (kernels, π.χ. RBF, πολυωνυμική) για να επιτύχει το διαχωρισμό. Ο SVM είναι επομένως ιδιαίτερος αποτελεσματικός σε περιπτώσεις όπου τα δεδομένα είναι υψηλής διάστασης και απαιτείται μεγάλη ακρίβεια στις ταξινομήσεις, είναι ανθεκτικός στην υπερπροσαρμογή αλλά μπορεί να είναι αργός σε μεγάλα σύνολα δεδομένων και ευαίσθητος στην επιλογή του πυρήνα (Cortes & Vapnik, 1995, Bishop, 2006)



Εικόνα 4. Διάγραμμα λειτουργίας του αλγορίθμου SVM: Απεικονίζεται ο τρόπος με τον οποίο ο αλγόριθμος SVM βρίσκει το υπερεπίπεδο που διαχωρίζει τα δεδομένα σε διαφορετικές κατηγορίες με το μέγιστο περιθώριο (Jainvidip, 2024)

Μέθοδος 10-fold Cross Validation

Η μέθοδος k-fold cross-validation, του προγράμματος Weka (Witten et al., 2016), είναι μια μέθοδος αξιολόγησης της απόδοσης ενός μοντέλου μηχανικής μάθησης. Η μέθοδος αυτή χρησιμοποιείται για να ελεγχθεί πόσο καλά γενικεύει το μοντέλο σε νέα άγνωστα δεδομένα, χωρίς να απαιτεί ξεχωριστό σύνολο ελέγχου (test set). Η μέθοδος βοηθά στην αποφυγή του προβλήματος της υπερπροσαρμογής, καθώς αξιολογεί την απόδοση του μοντέλου σε πολλαπλά σύνολα δεδομένων επικύρωσης κατά τη διάρκεια της εκπαίδευσης. Για το σκοπό αυτό, χωρίζει το σύνολο δεδομένων σε υποομάδες για την εκπαίδευση και τη δοκιμή.

Ειδικότερα, η μέθοδος 10-fold cross-validation στο Weka (Witten et al., 2016), χωρίζει το σύνολο των δεδομένων σε 10 ίσα μέρη (folds). Στη συνέχεια το μοντέλο εκπαιδεύεται σε 9 από τα 10 μέρη και δοκιμάζεται στο 1 τελευταίο μέρος. Η παραπάνω διαδικασία επαναλαμβάνεται 10 φορές χρησιμοποιώντας διαφορετικά υποσύνολα δοκιμών. Υπολογίζονται με τον τρόπο αυτό 10 διαφορετικά σύνολα μετρικών απόδοσης (π.χ. ακρίβεια, precision, recall) και λαμβάνεται ο μέσος όρος αυτών. Η μέθοδος αυτή είναι ιδιαίτερα αποτελεσματική καθώς χρησιμοποιεί το σύνολο των δεδομένων τόσο για εκπαίδευση όσο και για αξιολόγηση και παρέχει μια αξιόπιστη εκτίμηση της γενικής απόδοσης του μοντέλου (Kohavi, 1995).

Επιλογή Χαρακτηριστικών

Για τη βελτίωση της απόδοσης των αλγορίθμων αλλά και αποφυγή της «κατάρτας της διαστατικότητας», εφαρμόστηκε διαδικασία επιλογής χαρακτηριστικών (feature selection)

χρησιμοποιώντας τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise και έπειτα σε συνδυασμό με τον αλγόριθμο Genetic search. Η μέθοδος Wrapper στο WEKA είναι μια τεχνική επιλογής χαρακτηριστικών που χρησιμοποιεί έναν αλγόριθμο μηχανικής μάθησης για να αξιολογήσει τα υποσύνολα χαρακτηριστικών και να βρει τα πιο σημαντικά. Κατά την εφαρμογή της μεθόδου, δημιουργούνται υποσύνολα των χαρακτηριστικών και για κάθε ένα από τα υποσύνολα αυτά, εκπαιδεύεται ένα μοντέλο μηχανικής μάθησης (π.χ. Decision Tree, Random Forest). Ακολουθεί η αξιολόγηση με τη μέθοδο 10-fold cross-validation και επιλέγεται, εν τέλει, το υποσύνολο των χαρακτηριστικών που δίνει στο μοντέλο τη μεγαλύτερη ακρίβεια και την καλύτερη απόδοση (Kohavi & John, 1997).

Η μέθοδος Wrapper συνδυάστηκε αρχικά με τον αλγόριθμο Greedy Stepwise του οποίου η αρχή λειτουργίας του βασίζεται στο ότι επιλέγει σε κάθε βήμα το χαρακτηριστικό που προσφέρει τη μεγαλύτερη βελτίωση στην απόδοση ενός προκαθορισμένου αξιολογητή, όπως στην περίπτωση της παρούσας εργασία ο Wrapper (Kohavi & John, 1997).

Στη συνέχεια, για ενίσχυση της ακρίβειας αλλά και της δυνατότητας γενίκευσης των μοντέλων, εφαρμόστηκε εκ νέου η μέθοδος Wrapper, αυτή τη φορά σε συνδυασμό με τον αλγόριθμο Genetic Search. Ο γενετικός αλγόριθμος αποτελεί μια μέθοδο εμπνευσμένη από τη βιολογική εξέλιξη και επιδιώκει την επιλογή ενός βέλτιστου υποσυνόλου χαρακτηριστικών μέσω μίας εξελικτικής διαδικασίας (Goldberg D. E, 1989). Ο Genetic Search είναι πιο απαιτητικός υπολογιστικά και πιο αργός αλλά σε αντίθεση με τον Greedy Stepwise μπορεί να οδηγήσει σε καλύτερα αποτελέσματα αφού καταφέρνει να αποφύγει τα τοπικά βέλτιστα που αναφέρθηκαν παραπάνω.

Όλες οι μετρικές απόδοσης που προέκυψαν από τις παραπάνω αναλύσεις, παρουσιάζονται στην συνέχεια σε συγκριτικούς πίνακες, αποδεικνύοντας τη βελτίωση της απόδοσης του ταξινομητή μετά την επιλογή χαρακτηριστικών.

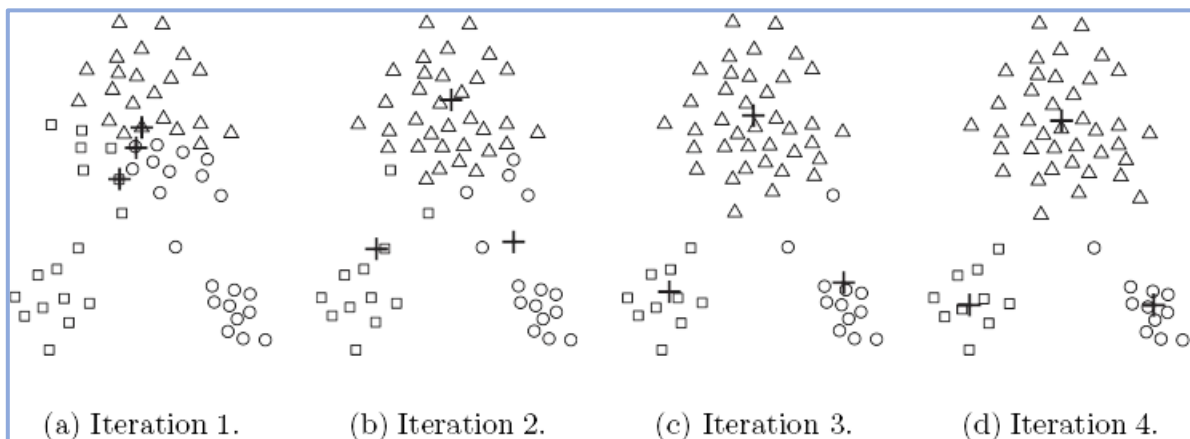
3.1.2 Μη Επιβλεπόμενη Μάθηση και Ομαδοποίηση Χωρών

Η μη επιβλεπόμενη μάθηση επιτρέπει την κατηγοριοποίηση των κρατών-μελών της Ε.Ε. με βάση παρόμοια χαρακτηριστικά ψηφιακής ανάπτυξης. Στην ανάλυση των δεδομένων του DESI, ομαδοποιήσαμε τις χώρες με τη βοήθεια του αλγορίθμου K-means.

Αλγόριθμος K-means

Ο K-means είναι ένας αλγόριθμος μη επιβλεπόμενης μάθησης, ο οποίος χρησιμοποιείται για την ομαδοποίηση (clustering) συνόλου δεδομένων και χρησιμοποιείται συχνά στην εξόρυξη δεδομένων και τη μηχανική μάθηση (McQueen, 1967). Ο αλγόριθμος λειτουργεί ακολουθώντας κάποια βασικά βήματα. Αρχικά επιλέγονται k κεντροειδή (centroids), η επιλογή μπορεί να είναι τυχαία ή να βασίζεται σε κάποια στρατηγική (π.χ. K-means++). Στη συνέχεια κάθε χαρακτηριστικό ανατίθεται στο πλησιέστερο κεντροειδές (βάσει Ευκλείδειας απόστασης, όπου ως τέτοια ορίζεται η απόσταση ανάμεσα σε δύο σημεία $\mathbf{x} = (x_1, x_2, \dots, x_n)$ και $\mathbf{y} = (y_1, y_2, \dots, y_n)$): $d_E(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}$ είτε βάσει απόστασης Manhattan $d_M(\mathbf{x}, \mathbf{y}) = (|x_1 - y_1| + |x_2 - y_2| + \dots + |x_n - y_n|)$ και δημιουργούνται οι πρώτες k ομάδες. Εν συνεχεία υπολογίζεται το νέο κεντροειδές για κάθε ομάδα και η διαδικασία επαναλαμβάνεται ωσότου σταθεροποιηθούν τα κεντροειδή ή μέχρι να ολοκληρωθεί ένας μέγιστος αριθμός επαναλήψεων (Bishop, 2006) (Εικόνα 5). Ουσιαστικά ο K-means προσπαθεί να ελαχιστοποιήσει το Sum of Squared Errors (SSE), δηλαδή το άθροισμα των τετραγωνικών αποστάσεων των σημείων από το κεντροειδές τους (Tan et al., 2019).

Ο αλγόριθμος είναι απλός, εύκολα εφαρμόσιμος και αποδοτικός σε μεγάλα σύνολα δεδομένων, όμως απαιτεί προκαθορισμένο αριθμό ομάδων (k), είναι ευαίσθητος στην αρχική επιλογή των κέντρων και δυσκολεύεται στον χειρισμό μη σφαιρικών ανομοιόμορφων ομάδων καθώς βασίζεται στην απόσταση των χαρακτηριστικών από τα κεντροειδή.



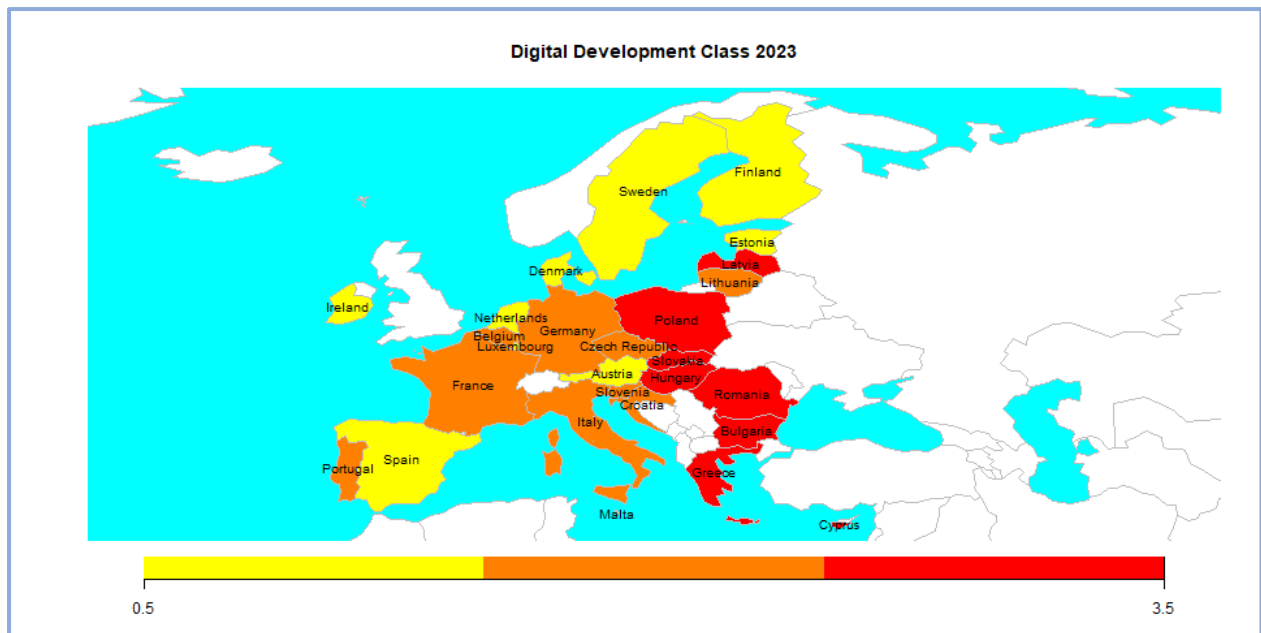
Εικόνα 5. Χρήση του αλγορίθμου K-means για ομαδοποίηση σε 3 clusters (ομάδες) (Tan et al., 2005)

4. Βιβλιογραφική Επισκόπηση για συναφείς εργασίες

Για το σκοπό της παρούσας μελέτης, πραγματοποιήθηκε βιβλιογραφική έρευνα με στόχο την αναγνώριση και κατηγοριοποίηση των χωρών της Ευρωπαϊκής Ένωσης (ΕΕ) βάσει του επιπέδου ανάπτυξής τους. Η αναζήτηση εστίασε σε επιστημονικές εργασίες που αξιολογούν τη θέση των κρατών-μελών στη βάση διαφορετικών αναπτυξιακών δεικτών, όπως η βιώσιμη ανάπτυξη, η κοινωνικοοικονομική εξέλιξη και ο ψηφιακός μετασχηματισμός. Η σύγκριση αυτών των μελετών με τα δεδομένα του DESI 2022 επιτρέπει την κατανόηση των τάσεων και διαφορών μεταξύ των κρατών της ΕΕ.

Ειδικότερα, η μελέτη των Bánhidí & Dobos (2023) εστιάζει στην ψηφιακή ανάπτυξη των κρατών-μελών της ΕΕ, αξιοποιώντας τον Δείκτη Ψηφιακής Οικονομίας και Κοινωνίας (DESI). Οι συγγραφείς εφαρμόζουν τρεις μεθοδολογικές προσεγγίσεις – Poset Analysis, Tiered DEA και Cluster Analysis – με στόχο την ομαδοποίηση των χωρών σύμφωνα με τους τέσσερις βασικούς πυλώνες του DESI: Συνδεσιμότητα, Ανθρώπινο Κεφάλαιο, Ενσωμάτωση Ψηφιακής Τεχνολογίας και Ψηφιακές Δημόσιες Υπηρεσίες. Η μελέτη καταλήγει πως η Δανία, η Φινλανδία, η Σουηδία, η Ολλανδία και η Εσθονία βρίσκονται στις κορυφαίες θέσεις, ενώ χώρες όπως η Βουλγαρία, η Ελλάδα και η Πολωνία εμφανίζουν χαμηλότερες επιδόσεις στον τομέα της ψηφιακής προόδου. Ιδιαίτερη σημασία έχει το γεγονός ότι η Εσθονία, παρά το μικρό της μέγεθος, κατατάσσεται μεταξύ των πιο ανεπτυγμένων ψηφιακά χωρών, υπογραμμίζοντας την επιτυχία των στρατηγικών ψηφιακού μετασχηματισμού που έχει θεσμοθετήσει. Οι ερευνητές αναδεικνύουν τη υψηλή συσχέτιση μεταξύ ψηφιακής και οικονομικής ανάπτυξης και προτείνουν περαιτέρω έρευνα για την κατανόηση των παραγόντων που επηρεάζουν τον ρυθμό της ψηφιακής προόδου στις διάφορες χώρες.

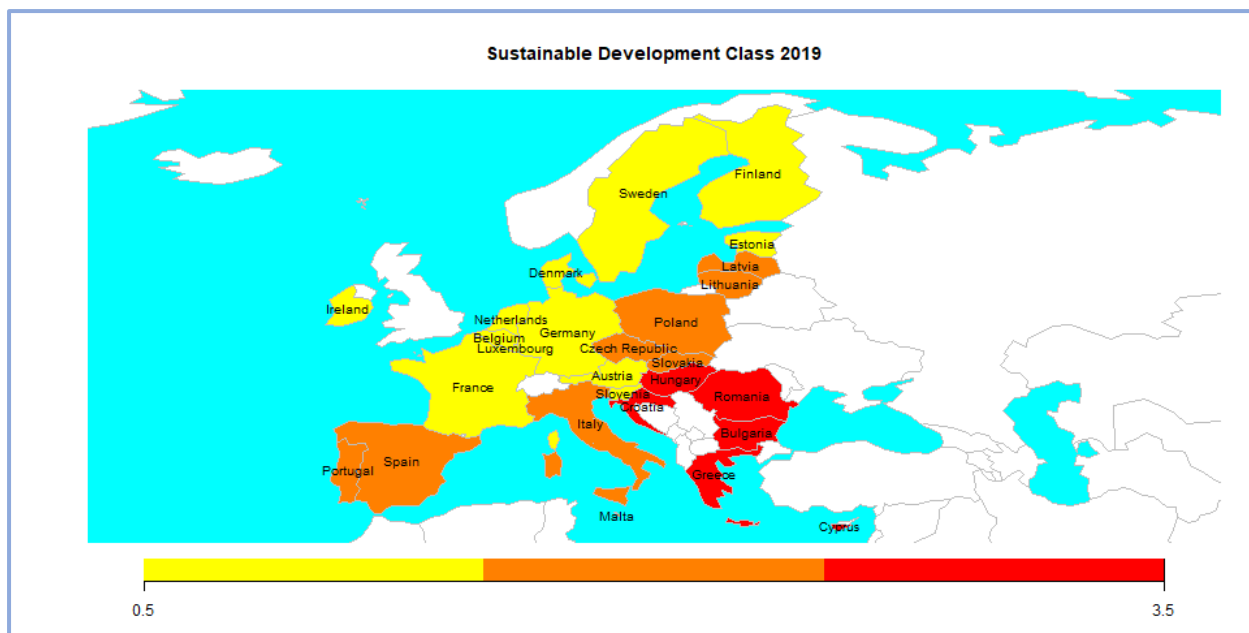
Βάσει της μελέτης αυτής, οι 27 χώρες – μέλη της Ευρωπαϊκής Ένωσης κατηγοριοποιήθηκαν ανάλογα με την ψηφιακή τους πρόοδο στις κατηγορίες High (Κίτρινο χρώμα), Medium (Πορτοκαλί χρώμα), Low (Κόκκινο χρώμα) όπως φαίνεται στον Χάρτη 1.



Χάρτης 1. Κατηγοριοποίηση των χωρών κρατών-μελών της ΕΕ βάσει της ψηφιακής τους προόδου κατά Bánhidí & Dobos (2023).

Παράλληλα, η μελέτη των Martín & Carnero (2019) εξετάζει τη βιώσιμη ανάπτυξη στις χώρες της ΕΕ χρησιμοποιώντας ένα σύνθετο δείκτη βιώσιμης ανάπτυξης βασισμένο σε 79 επιμέρους δείκτες, οι οποίοι ομαδοποιήθηκαν σε τρεις κύριες διαστάσεις: οικονομική, κοινωνική και περιβαλλοντική. Η ανάλυση των δεδομένων πραγματοποιήθηκε με τη μέθοδο της Ανάλυσης Κύριων Συνιστωσών (Principal Component Analysis - PCA), η οποία επέτρεψε τη μείωση της διάστασης των δεδομένων και την κατασκευή του σύνθετου δείκτη. Τα αποτελέσματα υπογραμμίζουν τις σημαντικές διαφοροποιήσεις που υπάρχουν μεταξύ των κρατών-μελών, με τις Σκανδιναβικές χώρες να εμφανίζουν τις υψηλότερες επιδόσεις και τις υπόλοιπες χώρες να κατατάσσονται με φθίνουσα σειρά καθώς απομακρύνονται από το πρότυπο της βιώσιμης ανάπτυξης. Η μελέτη προτείνει την εφαρμογή βέλτιστων πρακτικών και στοχευμένων πολιτικών για την ενίσχυση της βιωσιμότητας σε χώρες που υστερούν στους δείκτες αυτούς.

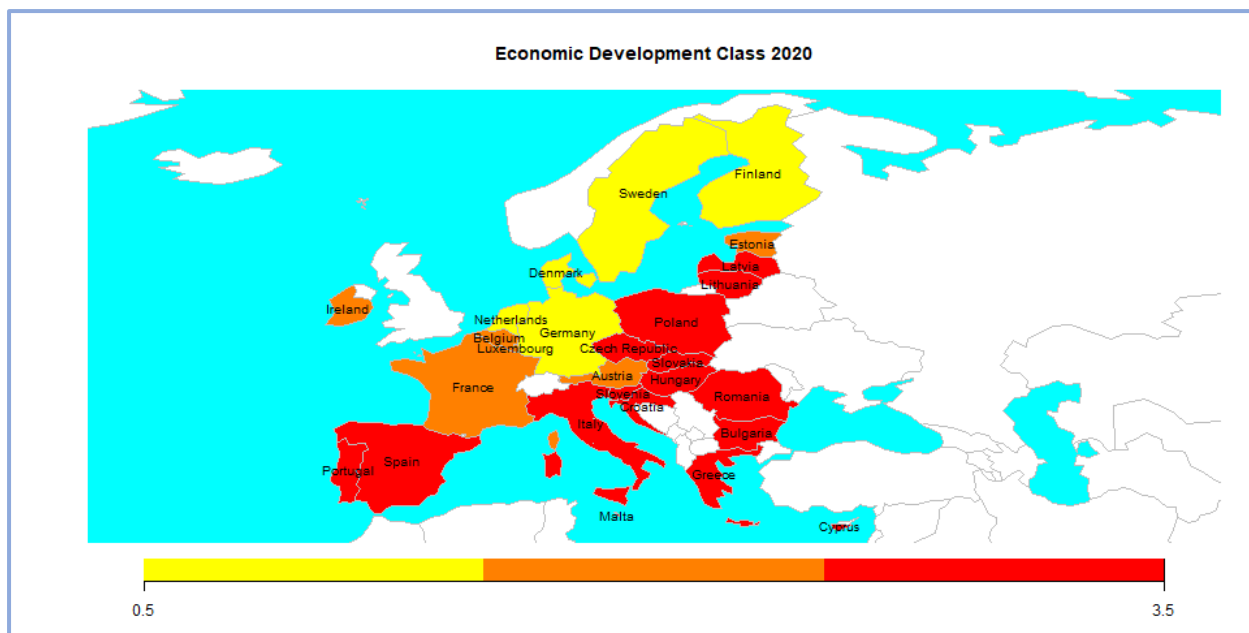
Βάσει της μελέτης αυτής, οι 27 χώρες – μέλη της Ευρωπαϊκής Ένωσης κατηγοριοποιήθηκαν εκ νέου ανάλογα με την ψηφιακή τους πρόοδο στις κατηγορίες High (Κίτρινο χρώμα), Medium (Πορτοκαλί χρώμα), Low (Κόκκινο χρώμα) όπως φαίνεται στο Χάρτη 2.



Χάρτης 2. Κατηγοριοποίηση των χωρών της ΕΕ βάσει δεικτών βιώσιμης ανάπτυξης κατά Martín & Carnero (2019).

Η μελέτη των Onuferoná, Čabinová & Matijoná (2020), κατατάσσει τα κράτη-μέλη της ΕΕ βάσει διεθνών δεικτών πολλαπλών κριτηρίων, χρησιμοποιώντας ανάλυση συστάδων (Cluster Analysis). Ειδικότερα, εξετάζονται οι δείκτες Global Competitiveness Index (GCI) – Ανταγωνιστικότητα, Economic Freedom Index (EFI) – Οικονομική ελευθερία, Global Innovation Index (GII) – Καινοτομία, Corruption Perceptions Index (CPI) – Διαφθορά και Human Development Index (HDI) – Ανθρώπινη ανάπτυξη, προκειμένου να διαμορφωθεί μια σαφής εικόνα της αναπτυξιακής θέσης κάθε χώρας. Τα ευρήματα της έρευνας υπογραμμίζουν τις έντονες ανισότητες μεταξύ των κρατών-μελών, με τις χώρες της δυτικής Ευρώπης να κατατάσσονται υψηλότερα από τις χώρες της ανατολικής Ευρώπης. Η μελέτη προτείνει την εφαρμογή στοχευμένης πολιτικής παρέμβασης για τη μείωση των αναπτυξιακών ανισοτήτων και την ενίσχυση της ανταγωνιστικότητας στις λιγότερο ανεπτυγμένες οικονομίες της ΕΕ.

Όμοια, οι 27 χώρες – μέλη της Ευρωπαϊκής Ένωσης κατηγοριοποιήθηκαν όπως και παραπάνω, ανάλογα με την ψηφιακή τους πρόοδο στις κατηγορίες High (Κίτρινο χρώμα), Medium (Πορτοκαλί χρώμα), Low (Κόκκινο χρώμα) όπως φαίνεται στο Χάρτη 3.



Χάρτης 3. Κατηγοριοποίηση των χωρών της ΕΕ βάσει δεικτών οικονομικής ανάπτυξης και ανταγωνιστικότητας κατά Onuferev et al. (2020).

Παρότι εξετάστηκαν αρκετές ακόμη μελέτες, επιλέχθηκαν οι παραπάνω τρεις λόγω της επικαιρότητάς τους και της διαφοροποίησης στη θεματική τους προσέγγιση. Άλλες μελέτες, όπως αυτή των Grzebyk & Stec (2015), αν και παρέχουν μια χρήσιμη θεώρηση της βιώσιμης ανάπτυξης στην ΕΕ, αναλύουν δεδομένα της περιόδου (2005-2012), καθιστώντας τα ευρήματά τους λιγότερο συναφή με τις σύγχρονες τάσεις. Όμοια, η εργασία των Stec et al. (2014) επικεντρώνεται στην κοινωνικοοικονομική ανάπτυξη των κρατών-μελών μέχρι το 2010, προσφέροντας μια σημαντική ανασκόπηση αλλά όχι μια πρόσφατη εικόνα της κατάστασης. Στην ίδια κατηγορία ανήκει και η εργασία της Bluszcz (2015), που ταξινομεί τις χώρες της ΕΕ βάσει της βιώσιμης ανάπτυξης, αλλά βασίζεται σε δεδομένα πριν το 2015, περιορίζοντας τη χρησιμότητά της ως συγκριτικό εργαλείο για τη σύγχρονη κατάσταση.

Η μελέτη των Laitsou & Xenakis (2023), που εξετάζει τη σχέση μεταξύ ψηφιακής ανάπτυξης και ανθρώπινης ευημερίας, χρησιμοποιώντας τον δείκτη DESI και τον δείκτη Ανθρώπινης Ανάπτυξης (HDI), εστιάζει περισσότερο στην αλληλεπίδραση μεταξύ των δύο αυτών μεταβλητών παρά στην κατηγοριοποίηση των χωρών, γεγονός που την καθιστά λιγότερο χρήσιμη για τους σκοπούς της παρούσας έρευνας. Η μελέτη Masoura & Malefaki (2023) εξετάζει την εξέλιξη του Δείκτη Ψηφιακής Οικονομίας και Κοινωνίας (DESI) στην Ευρωπαϊκή Ένωση (ΕΕ) για την περίοδο 2014-2019, χρησιμοποιώντας δεδομένα από τις εκθέσεις DESI, και καταλήγει στην κατηγοριοποίηση των κρατών μελών σε δύο ομάδες ψηφιακής ανάπτυξης και στην εξαγωγή συμπερασμάτων για τις εφαρμοζόμενες ψηφιακές πολιτικές της ΕΕ. Η

μελέτη δεν χρησιμοποιήθηκε τελικά καθώς κρίνεται συναφής αν και προγενέστερη, με το αντικείμενο της παρούσας εργασίας. Τέλος, η μελέτη των Leogrande et al. (2022), αν και παρέχει μια ενδιαφέρουσα ανάλυση για την ηλεκτρονική διακυβέρνηση στην Ευρώπη μέσω μηχανικής μάθησης, επικεντρώνεται περισσότερο στην τεχνολογική πρόοδο παρά στη συνολική ανάπτυξη των κρατών-μελών.

Έτσι, για τις ανάγκες της παρούσας έρευνας, επιλέχθηκαν οι μελέτες των Bánhidí & Dobos (2023), Martín & Carnero (2019) και Onuferová et al. (2020), καθώς παρέχουν πρόσφατα και πολυδιάστατα δεδομένα που καλύπτουν τόσο τον ψηφιακό όσο και τον οικονομικό και κοινωνικό τομέα ανάπτυξης. Η συγκριτική ανάλυση των ευρημάτων αυτών των ερευνών με τα δεδομένα του DESI 2022 θα συμβάλει στην εξαγωγή συμπερασμάτων σχετικά με τον ρόλο της ψηφιακής προόδου στη συνολική ανάπτυξη των κρατών-μελών.

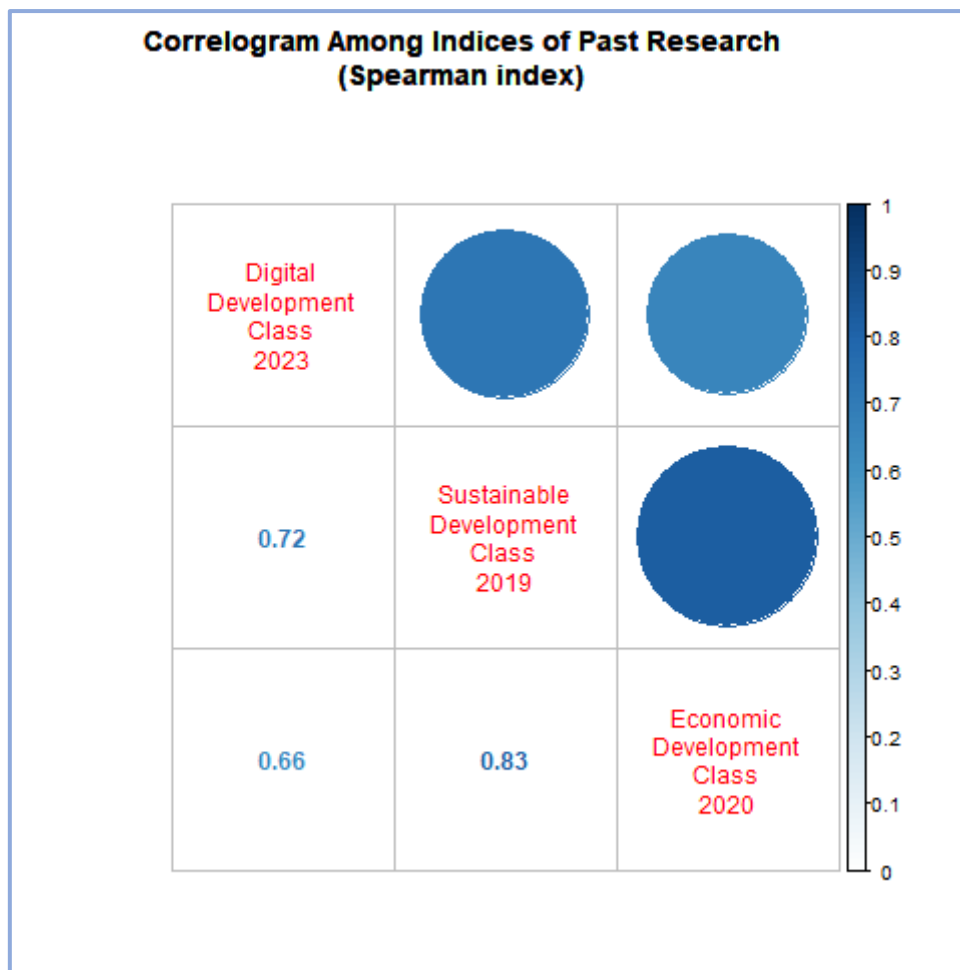
Η κατηγοριοποίηση των χωρών βάσει των παραπάνω δεικτών στις τρεις μελέτες αναλύεται στον Πίνακα 1 και εν συνεχεία ακολουθεί διάγραμμα συσχέτισης των τριών μελετών (Εικόνα 6). Το διάγραμμα συσχέτισης “Spearman correlogram” είναι ένα διάγραμμα που απεικονίζει τον βαθμό συσχέτισης μεταξύ μεταβλητών, χρησιμοποιώντας τον συντελεστή συσχέτισης κατά Spearman. Ο δείκτης Spearman αποτελεί ένα μέτρο συσχέτισης που βασίζεται στις τάξεις των τιμών και όχι στις απόλυτες αριθμητικές τιμές, όπως συμβαίνει με τον συντελεστή Pearson. Είναι κατάλληλος για αναλύσεις κοινωνικών επιστημών, καθώς διερευνά συσχετίσεις μεταξύ μη παραμετρικών μεταβλητών ή σε περιπτώσεις που τα δεδομένα ανήκουν σε διατακτική (ordinal) κλίμακα, δηλαδή σε αναλύσεις που δεν πληρούνται οι προϋποθέσεις της κανονικότητας και της γραμμικότητας (Conover, 1999). Το correlogram απεικονίζει όλους τους πιθανούς συνδυασμούς συσχετίσεων μεταξύ ζευγών μεταβλητών ενός dataset, με έγχρωμη ένδειξη της έντασης και του πρόσημου της συσχέτισης (Spearman C, 1987). Οι αποχρώσεις και τα μεγέθη των κύκλων υποδεικνύουν το επίπεδο συσχέτισης (Spearman Index), όπου τα πιο σκούρα και μεγαλύτερα σύμβολα υποδηλώνουν ισχυρότερες σχέσεις μεταξύ των μεταβλητών.

Όπως αναδεικνύεται από το διάγραμμα συσχέτισης υπάρχει θετική συσχέτιση μεταξύ της Οικονομικής Ανάπτυξης 2020 και της Ψηφιακής Ανάπτυξης 2023, γεγονός το οποίο υποδηλώνει ότι οι χώρες με ισχυρή οικονομία το 2020 έχουν επενδύσει σε σχέδια για την πρόοδο της ψηφιακής ανάπτυξης το 2023. Επιπρόσθετα, υπάρχει ισχυρή θετική συσχέτιση της Βιώσιμης ανάπτυξης 2019 τόσο με την Ψηφιακή Ανάπτυξη 2023 όσο και με την Οικονομική Ανάπτυξη 2020, αναδεικνύοντας την άμεση σύνδεση των δεικτών ανάπτυξης κάτι που

υποδεικνύει ότι οι χώρες που είχαν προωθήσει τη βιώσιμη ανάπτυξη το 2019 πιθανώς διαθέτουν και ισχυρές οικονομικές και ψηφιακές επιδόσεις. Η γενική εικόνα του correlogram υποστηρίζει την υπόθεση ότι η οικονομική, βιώσιμη και ψηφιακή ανάπτυξη είναι αλληλένδετες και ενισχύουν η μία την άλλη.

Country	Digital Development Class 2023	Sustainable Development Class 2019	Economic Development Class 2020
Austria	1	1	2
Belgium	2	1	2
Bulgaria	3	3	3
Cyprus	3	3	3
Czech Republic	2	2	3
Germany	2	1	1
Denmark	1	1	1
Estonia	1	1	2
Greece	3	3	3
Spain	1	2	3
Finland	1	1	1
France	2	1	2
Croatia	2	3	3
Hungary	3	3	3
Ireland	1	1	2
Italy	2	2	3
Lithuania	2	2	3
Luxembourg	1	1	2
Latvia	3	2	3
Malta	1	2	3
Netherlands	1	1	1
Poland	3	2	3
Portugal	2	2	3
Romania	3	3	3
Sweden	1	1	1
Slovenia	2	1	3
Slovakia	3	2	3

Πίνακας 1. Ομαδοποίηση των χωρών βάσει δεικτών ανάπτυξης σύμφωνα με τις μελέτες των Bánhidi & Dobos (2023), Martín & Carnero (2019) και Onuferová et al. (2020) (HIGH=1, MEDIUM=2, LOW=3).



Εικόνα 6. Correlogram που απεικονίζει τις συσχετίσεις μεταξύ δεικτών ψηφιακής, οικονομικής και βιώσιμης ανάπτυξης (Spearman Index). Οι αριθμητικές τιμές αντιπροσωπεύουν τον συντελεστή συσχέτισης μεταξύ των ζευγών μεταβλητών, ενώ οι αποχρώσεις και τα μεγέθη των κύκλων υποδεικνύουν το επίπεδο της συσχέτισης: όσο πιο σκούρο και μεγαλύτερο το σύμβολο, τόσο ισχυρότερη η συσχέτιση.

5. Συσχέτιση δεδομένων DESI με άλλες μελέτες με επιβλεπόμενη μάθηση.

Στην παρούσα ενότητα εφαρμόζονται τεχνικές επιβλεπόμενης μάθησης (supervised learning), με στόχο την πρόβλεψη της ψηφιακής κατηγορίας στην οποία ανήκει κάθε κράτος-μέλος της Ευρωπαϊκής Ένωσης. Για κάθε ένα από τα άρθρα των Bánhidi & Dobos (2023), Martín & Carnero (2019) και Onuferoná et al. (2020), δημιουργήθηκαν αντίστοιχα τρία αρχεία ARFF, με κάθε χώρα της ΕΕ να αποτελεί ένα δείγμα (instance), τα χαρακτηριστικά του DESI 2022 χρησιμοποιήθηκαν ως attributes και η ετικέτα κλάσης καθορίζεται από την ταξινόμηση των χωρών της ΕΕ όπως αναφέρεται στο κάθε άρθρο (Πίνακας 1). Τα αρχεία εκτελέστηκαν με το πρόγραμμα WEKA (Witten et al., 2016) και εφαρμόστηκαν οι αλγόριθμοι επιβλεπόμενης μάθησης: Decision Tree (J48), Random Forest (RDF), Adaboost (ADA), Multilayer Perceptron (MLP) και Support Vector Machine (SVM). Στη συνέχεια ακολούθησε επιλογή χαρακτηριστικών τύπου Wrapper σε συνδυασμό με τους αλγόριθμους Greedy Stepwise και Genetic Search και οι αλγόριθμοι επιβλεπόμενης μάθησης εφαρμόστηκαν εκ νέου.

5.1 Bánhidi & Dobos (2023) – Digital Development

Η μελέτη των Bánhidi & Dobos (2023): “Measurement of digital development with partial orders, Tiered DEA, and cluster analysis for the European Union” εστιάζει στην ψηφιακή ανάπτυξη των κρατών-μελών της ΕΕ

5.1.1 Decision Tree (J48)

Ο αλγόριθμος με το δένδρο απόφασης (J48) εφαρμόστηκε στα δεδομένα χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον Πίνακα 2.

Precision	Recall	F1-score	Class	Accuracy
0.875	0.700	0.778	HIGH	
0.500	0.667	0.572	MEDIUM	
0.571	0.500	0.533	LOW	
			Total	0.6296

Πίνακας 2. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48).

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 3) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 3). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 16. Ο Decision

Tree (J48) εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.727	0.800	0.762	HIGH	
0.700	0.778	0.737	MEDIUM	
1.000	0.750	0.857	LOW	
			Total	0.7778

Πίνακας 3. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) μετά από επιλογή χαρακτηριστικών τύπου Wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.900	0.900	0.900	HIGH	
0.800	0.889	0.842	MEDIUM	
0.857	0.750	0.800	LOW	
			Total	0.8519

Πίνακας 4. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών τύπου Wrapper και τον αλγόριθμο Genetic Search.

5.1.2 Random Forest

Ο αλγόριθμος Random Forest εφαρμόστηκε στα δεδομένα χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.889	0.800	0.842	HIGH	
0.538	0.778	0.636	MEDIUM	
0.800	0.500	0.615	LOW	
			Total	0.7037

Πίνακας 5. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest.

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 6) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 7). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 16. Ο Random Forest εφαρμόστηκε ξανά με τα παραπάνω επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.900	0.947	HIGH	
0.818	1.000	0.900	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.9259

Πίνακας 6. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest μετά από επιλογή χαρακτηριστικών τύπου wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.818	0.900	0.857	HIGH	
0.889	0.889	0.889	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.8889

Πίνακας 7. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών τύπου Wrapper και τον αλγόριθμο Genetic Search.

5.1.3 Adaboost

Ομοίως εφαρμόστηκε ο αλγόριθμος Adaboost, και κατέληξε στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.875	0.700	0.778	HIGH	
0.538	0.778	0.572	MEDIUM	
0.667	0.500	0.533	LOW	
			Total	0.6667

Πίνακας 8. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost.

Η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο GreedyStepwise (Πίνακας 9) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 9) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 16. Ο Adaboost εφαρμόστηκε ξανά με τα επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.818	0.900	0.857	HIGH	
0.889	0.889	0.889	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.8889

Πίνακας 9. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost μετά από επιλογή χαρακτηριστικών τύπου Wrapper (μετρικές όμοιες κατά την εφαρμογή τόσο του αλγορίθμου Greedy Stepwise όσο και του αλγορίθμου Genetic Search).

5.1.4 Multilayer Perceptron

Ο αλγόριθμος Multilayer Perceptron εφαρμόστηκε στα δεδομένα με την παραμετροποίηση “hidden layers=3”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το

υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.750	0.900	0.818	HIGH	
0.571	0.444	0.500	MEDIUM	
0.750	0.750	0.750	LOW	
			Total	0.7407

Πίνακας 10. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron.

Ακολούθησε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 11) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 12). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 16. Ο Multilayer Perceptron εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.900	0.947	HIGH	
0.818	1.000	0.900	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.9259

Πίνακας 11. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron μετά από επιλογή χαρακτηριστικών τύπου wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.909	1.000	0.952	HIGH	
0.889	0.889	0.889	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.9259

Πίνακας 12. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron μετά από επιλογή χαρακτηριστικών τύπου Wrapper και τον αλγόριθμο Genetic Search.

5.1.5 Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) εφαρμόστηκε στα δεδομένα με την παραμετροποίηση “Kernel type=linear”), και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.800	0.800	0.800	HIGH	
0.500	0.556	0.527	MEDIUM	
0.714	0.625	0.667	LOW	
			Total	0.6667

Πίνακας 13. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine.

Τέλος, η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 14) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 15) κατέληξε στα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 16. Ο Support Vector Machine εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.900	0.947	HIGH	
0.800	0.889	0.842	MEDIUM	
0.875	0.875	0.875	LOW	
			Total	0.8889

Πίνακας 14. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine μετά από επιλογή χαρακτηριστικών τύπου wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.900	0.900	0.900	HIGH	
0.889	0.889	0.889	MEDIUM	
0.875	0.875	0.875	LOW	
			Total	0.8889

Πίνακας 15. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine μετά από επιλογή χαρακτηριστικών τύπου wrapper και τον αλγόριθμο Genetic search.

Variable	J48	RDF	ADA	MLP	SVM	J48	RDF	ADA	MLP	SVM
At least basic digital skills										
Above basic digital skills										
At least basic digital content creation skills										
ICT specialists										
Female ICT specialists										
Enterprises providing ICT training										
ICT graduates										
Overall fixed broadband take-up										
At least 100 Mbps fixed broadband take-up										
At least 1 Gbps broadband take-up										
Fast broadband (NGA) coverage										
Fixed Very High Capacity Network (VHCN) coverage										
5G spectrum										
Overall 5G coverage										
Mobile broadband take-up										
Broadband price index										
SMEs with at least a basic level of digital intensity										
Electronic information sharing										
Social media										
Big data										
Cloud										
Artificial intelligence										
ICT for environmental sustainability										
e-Invoices										
SMEs selling online										
e-Commerce turnover										
Selling online cross-border										
e-Government users										
Pre-filled Forms										
Digital public services for citizens										
Digital public services for businesses										
Open data										

Πίνακας 16. Αποτελέσματα Επιλογής Χαρακτηριστικών με Μεθόδους Wrapper για Δεδομένα DESI 2022 με ετικέτες κλάσης από τη μελέτη Bánhidí & Dobos (2023). Τα πράσινα κελιά υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τους αλγόριθμους Greedy Stepwise, ενώ τα γκρι υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Genetic Search. Τα χαρακτηριστικά που

επιλέχθηκαν κατ' επανάληψη ήταν τα **Basic digital skills, Mobile broadband take-up, Digital public services for citizens**.

5.2 Martin & Carnero (2019) – Sustainable Development

Η μελέτη "Evaluation of Sustainable Development in European Union Countries" των Carlos Javier Martín και María Carmen Carnero, δημοσιεύθηκε το 2019 και αξιολογεί τη βιώσιμη ανάπτυξη των κρατών-μελών της Ευρωπαϊκής Ένωσης (ΕΕ) χρησιμοποιώντας έναν σύνθετο δείκτη που περιλαμβάνει τόσο οικονομικές, κοινωνικές, όσο και περιβαλλοντικές διαστάσεις.

5.2.1 Decision Tree (J48)

Ο αλγόριθμος με το δένδρο απόφασης (J48) εφαρμόστηκε στα δεδομένα χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.917	0.917	0.917	HIGH	
0.500	0.778	0.999	MEDIUM	
0.000	0.000	0.000	LOW	
			Total	0.6667

Πίνακας 17. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48).

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 18) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 19). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 32. Ο Decision Tree (J48) εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.818	0.750	0.783	HIGH	
0.636	0.778	0.700	MEDIUM	
0.600	0.500	0.545	LOW	
			Total	0.7037

Πίνακας 18. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) μετά από επιλογή χαρακτηριστικών τύπου wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.917	0.917	0.917	HIGH	
0.667	0.889	0.762	MEDIUM	
1.000	0.500	0.667	LOW	
			Total	0.8148

Πίνακας 19. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) μετά από επιλογή χαρακτηριστικών τύπου wrapper και τον αλγόριθμο Genetic Search.

5.2.2 Random Forest

Ο αλγόριθμος Random Forest εφαρμόστηκε στα δεδομένα χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται παρακάτω:

Precision	Recall	F1-score	Class	Accuracy
0.769	0.833	0.800	HIGH	
0.400	0.444	0.421	MEDIUM	
0.500	0.333	0.400	LOW	
			Total	0.5926

Πίνακας 20. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest.

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 21) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 22). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 32. Ο Random Forest εφαρμόστηκε ξανά με τα παραπάνω επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.818	0.750	0.783	HIGH	
0.778	0.778	0.778	MEDIUM	
0.714	0.833	0.769	LOW	
			Total	0.7778

Πίνακας 21. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest μετά από επιλογή χαρακτηριστικών τύπου wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.846	0.917	0.880	HIGH	
0.600	0.667	0.632	MEDIUM	
0.750	0.500	0.600	LOW	
			Total	0.7407

Πίνακας 22. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest μετά από επιλογή χαρακτηριστικών τύπου wrapper και τον αλγόριθμο Genetic Search.

5.2.3 Adaboost

Εν συνεχεία η εφαρμογή του αλγορίθμου Adaboost, στο σύνολο των δεδομένων οδήγησε στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.857	1.000	0.923	HIGH	
0.429	0.333	0.375	MEDIUM	
0.500	0.333	0.400	LOW	
			Total	0.6296

Πίνακας 23. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost.

Η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 24) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 25) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 32. Ο Adaboost εφαρμόστηκε ξανά με τα επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.667	1.000	0.800	HIGH	
0.667	0.667	0.667	MEDIUM	
?	0.000	?	LOW	
			Total	0.6667

Πίνακας 24. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost μετά από επιλογή χαρακτηριστικών τύπου Wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.786	0.917	0.846	HIGH	
0.700	0.778	0.737	MEDIUM	
1.000	0.500	0.667	LOW	
			Total	0.7778

Πίνακας 25. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost μετά από επιλογή χαρακτηριστικών τύπου Wrapper και τον αλγόριθμο Genetic Search.

5.2.4 Multilayer Perceptron

Ο αλγόριθμος Multilayer Perceptron εφαρμόστηκε στα δεδομένα με την παραμετροποίηση “hidden layers=3”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.636	0.583	0.608	HIGH	
0.333	0.444	0.381	MEDIUM	
0.500	0.333	0.400	LOW	
			Total	0.4815

Πίνακας 26. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron.

Ακολούθησε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 27) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 28). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 32. Ο Multilayer Perceptron εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.667	0.833	0.741	HIGH	
0.667	0.667	0.667	MEDIUM	
1.000	0.500	0.667	LOW	
			Total	0.7037

Πίνακας 27. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.917	0.917	0.917	HIGH	
0.750	0.667	0.706	MEDIUM	
0.571	0.667	0.615	LOW	
			Total	0.7778

Πίνακας 28. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper και τον αλγόριθμο Genetic Search.

5.2.5 Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) εφαρμόστηκε στα δεδομένα με την παραμετροποίηση “Kernel type=linear”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.600	0.500	0.545	HIGH	
0.417	0.556	0.477	MEDIUM	
0.600	0.500	0.545	LOW	
			Total	0.5185

Πίνακας 29. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine.

Τέλος, η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 30) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 31) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 32. Ο Support Vector Machine εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.750	0.857	HIGH	
0.643	1.000	0.783	MEDIUM	
1.000	0.667	0.800	LOW	
			Total	0.8148

Πίνακας 30. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper και τον αλγόριθμό Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.833	0.833	0.833	HIGH	
0.700	0.778	0.737	MEDIUM	
1.000	0.833	0.909	LOW	
			Total	0.8148

Πίνακας 31. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper και τον αλγόριθμό Genetic Search.

Variable	J48	RDF	ADA	MLP	SVM	J48	RDF	ADA	MLP	SVM
At least basic digital skills										
Above basic digital skills										
At least basic digital content creation skills										
ICT specialists										
Female ICT specialists										
Enterprises providing ICT training										
ICT graduates										
Overall fixed broadband take-up										
At least 100 Mbps fixed broadband take-up										
At least 1 Gbps broadband take-up										
Fast broadband (NGA) coverage										
Fixed Very High Capacity Network (VHCN) coverage										
5G spectrum										
Overall 5G coverage										
Mobile broadband take-up										
Broadband price index										
SMEs with at least a basic level of digital intensity										
Electronic information sharing										
Social media										
Big data										
Cloud										
Artificial intelligence										
ICT for environmental sustainability										
e-Invoices										
SMEs selling online										
e-Commerce turnover										
Selling online cross-border										
e-Government users										
Pre-filled Forms										
Digital public services for citizens										
Digital public services for businesses										
Open data										

Πίνακας 32. Αποτελέσματα Επιλογής Χαρακτηριστικών με Μεθόδους Wrapper για Δεδομένα DESI 2022 με ετικέτες κλάσης από τη μελέτη των Martín & Carnero (2019). Τα πράσινα κελιά υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τους αλγόριθμους Greedy Stepwise, ενώ τα γκρι υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Genetic Search. Τα χαρακτηριστικά που επιλέχθηκαν συχνότερα είναι τα ICT specialists, At least 100 Mbps fixed broadband take-up, Electronic information sharing, e-Commerce turnover.

5.3 Onuferová et al. (2020) – Economic Development

Η μελέτη " Categorization of the EU Member States in the Context of Selected Multicriteria International Indices Using Cluster Analysis " των Onuferová, E., Čabinová, V., & Matijoná, M. του 2020, αξιολογεί την οικονομική ανάπτυξη και την ανταγωνιστικότητα των κρατών-μελών της Ευρωπαϊκής Ένωσης (ΕΕ).

5.3.1 Decision Tree (J48)

Ο αλγόριθμος με το δένδρο απόφασης (J48) εφαρμόστηκε στα δεδομένα χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F-1 score	Class	Accuracy
0.750	0.600	0.667	HIGH	
0.286	0.333	0.308	MEDIUM	
0.750	0.750	0.750	LOW	
			Total	0.6296

Πίνακας 33. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48).

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 34) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 35). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 48. Ο Decision Tree (J48) εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F-1 score	Class	Accuracy
0.800	0.800	0.800	HIGH	
0.667	0.667	0.667	MEDIUM	
0.875	0.875	0.875	LOW	
			Total	0.8148

Πίνακας 34. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών με τη μέθοδο wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F-1 score	Class	Accuracy
0.667	0.800	0.727	HIGH	
0.800	0.667	0.727	MEDIUM	
0.813	0.813	0.813	LOW	
			Total	0.7778

Πίνακας 35. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών με τη μέθοδο wrapper και τον αλγόριθμο Genetic Search.

5.3.2 Random Forest

Ο αλγόριθμος Random Forest εφαρμόστηκε στα δεδομένα χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται παρακάτω:

Precision	Recall	F-1 score	Class	Accuracy
1.000	0.800	0.889	HIGH	
0.500	0.333	0.400	MEDIUM	
0.737	0.875	0.800	LOW	
			Total	0.7407

Πίνακας 36. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest.

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 37) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 38). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 48. Ο Random Forest εφαρμόστηκε ξανά με τα παραπάνω επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F-1 score	Class	Accuracy
1.000	0.800	0.889	HIGH	
1.000	0.667	0.800	MEDIUM	
0.842	1.000	0.914	LOW	
			Total	0.8889

Πίνακας 37. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F-1 score	Class	Accuracy
1.000	0.600	0.750	HIGH	
0.800	0.667	0.727	MEDIUM	
0.842	1.000	0.914	LOW	
			Total	0.8518

Πίνακας 38. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

5.3.3 Adaboost

Εν συνεχεία εφαρμόστηκε ο αλγόριθμος Adaboost, στο σύνολο των δεδομένων και οδήγησε στα παρακάτω αποτελέσματα:

Precision	Recall	F-1 score	Class	Accuracy
0.500	0.400	0.444	HIGH	
0.375	0.500	0.429	MEDIUM	
0.800	0.750	0.774	LOW	
			Total	0.6296

Πίνακας 39. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost.

Η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο GreedyStepwise (Πίνακας 40) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 41) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 48. Ο Adaboost εφαρμόστηκε ξανά με τα επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F-1 score	Class	Accuracy
1.000	0.800	0.889	HIGH	
0.800	0.667	0.727	MEDIUM	
0.833	0.938	0.882	LOW	
			Total	0.8519

Πίνακας 40. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F-1 score	Class	Accuracy
0.800	0.800	0.800	HIGH	
0.750	0.500	0.600	MEDIUM	
0.833	0.938	0.882	LOW	
			Total	0.8148

Πίνακας 41. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

5.3.4 Multilayer Perceptron

Ο αλγόριθμος Multilayer Perceptron εφαρμόστηκε στα δεδομένα με την παραμετροποίηση “hidden layers=3”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F-1 score	Class	Accuracy
0.667	0.800	0.727	HIGH	
0.000	0.000	0.000	MEDIUM	
0.722	0.813	0.765	LOW	
			Total	0.6296

Πίνακας 42. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron.

Ακολούθησε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον αλγόριθμο Greedy Stepwise (Πίνακας 43) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 44). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 48. Ο Multilayer Perceptron εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F-1 score	Class	Accuracy
0.800	0.800	0.800	HIGH	
0.800	0.667	0.727	MEDIUM	
0.941	1.000	0.970	LOW	
			Total	0.8889

Πίνακας 43. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F-1 score	Class	Accuracy
1.000	0.800	0.889	HIGH	
0.857	1.000	0.923	MEDIUM	
1.000	1.000	1.000	LOW	
			Total	0.9630

Πίνακας 44. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

5.3.5 Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) εφαρμόστηκε στα δεδομένα με την παραμετροποίηση “Kernel type=linear”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.500	0.400	0.444	HIGH	
0.250	0.167	0.200	MEDIUM	
0.684	0.813	0.743	LOW	
			Total	0.5926

Πίνακας 45. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine.

Τέλος, η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 46) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 47) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 48. Ο Support Vector Machine εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	1.000	1.000	HIGH	
0.714	0.833	0.769	MEDIUM	
0.933	0.875	0.903	LOW	
			Total	0.8889

Πίνακας 46. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.600	0.600	0.600	HIGH	
0.667	0.667	0.667	MEDIUM	
0.875	0.875	0.875	LOW	
			Total	0.7778

Πίνακας 47. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

Variable	J48	RDF	ADA	MLP	SVM	J48	RDF	ADA	MLP	SVM
At least basic digital skills										
Above basic digital skills										
At least basic digital content creation skills										
ICT specialists										
Female ICT specialists										
Enterprises providing ICT training										
ICT graduates										
Overall fixed broadband take-up										
At least 100 Mbps fixed broadband take-up										
At least 1 Gbps broadband take-up										
Fast broadband (NGA) coverage										
Fixed Very High Capacity Network (VHCN) coverage										
5G spectrum										
Overall 5G coverage										
Mobile broadband take-up										
Broadband price index										
SMEs with at least a basic level of digital intensity										
Electronic information sharing										
Social media										
Big data										
Cloud										
Artificial intelligence										
ICT for environmental sustainability										
e-Invoices										
SMEs selling online										
e-Commerce turnover										
Selling online cross-border										
e-Government users										
Pre-filled Forms										
Digital public services for citizens										
Digital public services for businesses										
Open data										

Πίνακας 48. Αποτελέσματα Επιλογής Χαρακτηριστικών με Μεθόδους Wrapper για Δεδομένα DESI 2022 με ετικέτες κλάσης από τη μελέτη Onuferoná et al. (2020). Τα πράσινα κελιά υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τους αλγόριθμους Greedy Stepwise, ενώ τα γκρι υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Genetic Search. Τα χαρακτηριστικά που ανέδειξε η μέθοδος κατ' επανάληψη είναι ICT specialists, SMEs with at least a basic level of digital intensity, Cloud, e-Government users.

5.4 Συμπεράσματα και συγκριτική μελέτη

Διερευνήθηκε παραπάνω η συσχέτιση των δεικτών ψηφιακής ανάπτυξης (DESI 2022) με τρεις διαφορετικές κατηγοριοποιήσεις χωρών της ΕΕ, όπως αυτές προτάθηκαν στις εργασίες των Bánhidi & Dobos (2023), Martín & Carnero (2019) και Onuferová et al. (2020). Εξετάστηκε μάλιστα ανωτέρω, η δυνατότητα πρόβλεψης των προτεινόμενων ταξινομήσεων με βάση τα χαρακτηριστικά του DESI, χρησιμοποιώντας την εφαρμογή Weka και αλγόριθμους επιβλεπόμενης μάθησης. Τα αποτελέσματα φανερώνουν σημαντικές διαφορές στην ικανότητα πρόβλεψης ανάλογα με τον τύπο της μελέτης αλλά και τον επιλεγμένο ταξινομητή.

Στην περίπτωση της μελέτης Bánhidi & Dobos (2023), η οποία εστιάζει στην ψηφιακή πρόοδο, παρατηρήθηκε η υψηλότερη ακρίβεια πρόβλεψης. Ο αλγόριθμος Multilayer Perceptron επέδειξε τη μεγαλύτερη ακρίβεια 74,07% ανάμεσα στους υπό συζήτηση αλγορίθμους, ενώ μετά την επιλογή χαρακτηριστικών η απόδοση του έφθασε το 92,59% τόσο για τον Random Forest όσο και για το Multilayer Perceptron. Οι ψηφιακές δεξιότητες του πληθυσμού (at least and above basic digital skills) για τη συνιστώσα «Ανθρώπινο Κεφάλαιο», η χρήση κινητής ευρυζωνικότητας (mobile broadband take-up) για τη «Συνδεσιμότητα» και οι ψηφιακές δημόσιες υπηρεσίες για πολίτες (digital public services for citizens) για τη συνιστώσα «Ψηφιακές Δημόσιες Υπηρεσίες» ήταν τα χαρακτηριστικά που επηρέασαν περισσότερο την απόδοση των ταξινομητών.

Ωστόσο, στην ανάλυση της βιώσιμης ανάπτυξης (Martín & Carnero, 2019) τα αποτελέσματα ήταν λιγότερο ικανοποιητικά. Πριν από την επιλογή χαρακτηριστικών, ο αλγόριθμος Decision Tree (J48) έδειξε ακρίβεια 66,67%, παρουσίασε όμως σημαντικές αδυναμίες στην ταξινόμηση της κλάσης LOW. Μετά την επιλογή χαρακτηριστικών, το μοντέλο Decision Tree (J48) αλλά και ο SVM επέδειξαν τη μεγαλύτερη ακρίβεια 81,48%, με τα πιο σημαντικά χαρακτηριστικά να είναι οι ειδικοί ΤΠΕ (ICT specialists) όσον αφορά στο «Ανθρώπινο Κεφάλαιο», οι συνδρομές ευρυζωνικής σύνδεσης τουλάχιστον 100 Mbps (At least 100 Mbps fixed broadband take-up) όσον αφορά στη «Συνδεσιμότητα», η ηλεκτρονική ανταλλαγή πληροφοριών (Electronic information sharing) και οι πωλήσεις ηλεκτρονικού εμπορίου (e-commerce turnover) όσον αφορά στην «Ενσωμάτωση ψηφιακών τεχνολογιών από επιχειρήσεις». Η διαφορά στην απόδοση των ταξινομητών σε σύγκριση με την προηγούμενη μελέτη υποδηλώνει ότι η βιωσιμότητα, η οποία είναι μια πολυδιάστατη έννοια που αφορά κοινωνικούς, οικονομικούς και περιβαλλοντικούς δείκτες, δεν μπορεί να προβλεφθεί πλήρως μέσω των ψηφιακών δεικτών μόνο.

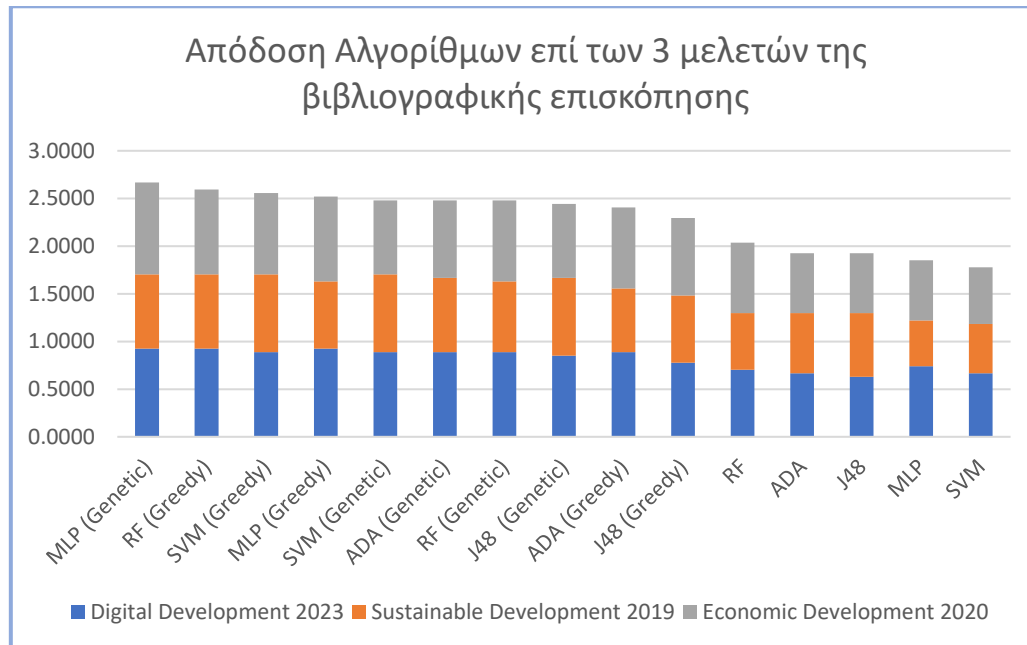
Στην τρίτη μελέτη (Onuferoná et al., 2020), που αφορά κατά κύριο λόγο την οικονομική ανάπτυξη, τα αποτελέσματα ήταν εμφανώς καλύτερα. Χωρίς επιλογή χαρακτηριστικών, ο Random Forest έφθασε σε ακρίβεια 74,07%, η οποία βελτιώθηκε σε 88,89% μετά την επιλογή χαρακτηριστικών με τον αλγόριθμο Greedy. Αντίστοιχα, εξαιρετικά αποτελέσματα έδωσε ο αλγόριθμος Multilayer Perceptron μετά από επιλογή χαρακτηριστικών με τον αλγόριθμο Genetic με ακρίβεια 96,30%. Στην περίπτωση αυτής της μελέτης, οι ειδικοί ΤΠΕ (ICT specialists) όσον αφορά στη συνιστώσα «Ανθρώπινο Κεφάλαιο», οι ΜΜΕ με τουλάχιστον βασικό επίπεδο ψηφιακής έντασης (SMEs with at least a basic level of digital intensity) και η χρήση υπολογιστικού σύννεφου (Cloud) όσον αφορά στη συνιστώσα της «Ενσωμάτωσης Ψηφιακών Υπηρεσιών», και οι χρήστες ηλεκτρονικής διακυβέρνησης (e-Government users) όσον αφορά στις «Ψηφιακές Δημόσιες Υπηρεσίες», ήταν τα κατ' επανάληψη επιλεγμένα χαρακτηριστικά που βελτίωσαν την ακρίβεια της ταξινόμησης. Η σχετικά υψηλή απόδοση υποδεικνύει ότι οι ψηφιακοί δείκτες έχουν σημαντική συσχέτιση με την οικονομική ανάπτυξη, αν και πιθανώς δεν αποτελούν πλήρες προγνωστικό μοντέλο από μόνοι τους.

Συγκρίνοντας τις τρεις μελέτες, διαπιστώνεται ότι η ικανότητα πρόβλεψης των μοντέλων ποικίλλει ανάλογα με τον τύπο της εξαρτημένης μεταβλητής. Η υψηλότερη ακρίβεια παρατηρήθηκε στην περίπτωση της ψηφιακής προόδου, όπου οι δείκτες DESI σχεδιάστηκαν ειδικά για αυτόν τον σκοπό. Αντίθετα, στις περιπτώσεις της βιώσιμης και οικονομικής ανάπτυξης, όπου οι έννοιες είναι ευρύτερες και πολυσύνθετες, η απόδοση των μοντέλων ήταν χαμηλότερη, αν και βελτιώθηκε σημαντικά με την εφαρμογή μεθόδων επιλογής χαρακτηριστικών.

Από πλευράς αλγορίθμων, οι Random Forest, Multilayer Perceptron και SVM επέδειξαν την καλύτερη συνολική απόδοση, ειδικότερα μετά την επιλογή χαρακτηριστικών. Ο αλγόριθμος με δέντρο απόφασης (J48) και ο Adaboost παρείχαν χαμηλότερη ακρίβεια σε σχέση με άλλους αλγορίθμους. Η επιλογή χαρακτηριστικών με τη μέθοδο wrapper απέδειξε την αξία της σε όλες τις περιπτώσεις και στις τρεις μελέτες, με βελτιώσεις στην ακρίβεια που κυμαίνονταν από 15% έως 25% με τον αλγόριθμο Genetic Search να έχει λίγο καλύτερες συνολικές επιδόσεις από τον Greedy Stepwise, αναδεικνύοντας ταυτόχρονα μεγαλύτερη ποικιλία επιλεγμένων χαρακτηριστικών.

Εν κατακλείδι, τα ευρήματα της παραπάνω ανάλυσης καταδεικνύουν πως τα χαρακτηριστικά του DESI μπορούν να χρησιμοποιηθούν για αξιόπιστη ταξινόμηση σε πολλαπλές διαστάσεις ανάπτυξης αλλά ταυτόχρονα αποκαλύπτουν πως η ανάλυση πιο σύνθετων εννοιών, όπως η

βιωσιμότητα, χρειάζεται ενδεχομένως συνδυασμό περισσότερων εργαλείων για ακριβέστερη πρόβλεψη. Επιπρόσθετα, η βελτίωση της ακρίβειας μετά από την επιλογή χαρακτηριστικών, δεν είναι απλά ένα στατιστικό εύρημα, αλλά αποδεικνύει την ανάγκη για στρατηγική επιλογή μεταβλητών σε εφαρμογές πολιτικής ανάλυσης.



Γράφημα 5. Απόδοση Αλγορίθμων DESI επί των 3 μελετών της βιβλιογραφικής επισκόπησης. Η σειρά Digital Development αφορά στη μελέτη Bánhidí & Dobos (2023), η Sustainable Development 2019 στη μελέτη Martín & Carnero (2019) και η σειρά Economic Development 2020 στη μελέτη Onuferová et al. (2020).

6. Ανάλυση Ομαδοποίησης (Clustering) με βάση τον Δείκτη DESI 2022

Στο παρόν κεφάλαιο, εφαρμόζεται η μέθοδος της ομαδοποίησης (clustering) με στόχο την ταξινόμηση των χωρών της ΕΕ σε ομοιογενείς ομάδες ανάλογα με το επίπεδο ψηφιακής τους ανάπτυξης. Για το σκοπό της ανάλυσης χρησιμοποιήθηκαν δεδομένα του DESI για το έτος 2022 και τα αποτελέσματα αφορούν τόσο στον συνολικό δείκτη όσο και σε κάθε συνιστώσα ξεχωριστά (Ανθρώπινο Κεφάλαιο, Συνδεσιμότητα, Ενσωμάτωση Ψηφιακής Τεχνολογίας, Ψηφιακές Δημόσιες Υπηρεσίες).

6.1 Μεθοδολογία

Για το σκοπό της διερεύνησης της ομοιογένειας και παράλληλα της διαφοροποίησης των ευρωπαϊκών χωρών ως προς την ψηφιακή τους ανάπτυξη, εφαρμόστηκε η μέθοδος ομαδοποίησης K-means clustering μέσω του λογισμικού Weka (Witten et al., 2017). Στόχος ήταν η ταξινόμηση των χωρών σε τρεις διακριτές ομάδες βάσει των επιδόσεών τους στον συνολικό δείκτη DESI (Digital Economy and Society Index) καθώς και στις τέσσερις επιμέρους διαστάσεις-συνιστώσες του (συνδεσιμότητα, ανθρώπινο κεφάλαιο, ενσωμάτωση ψηφιακής τεχνολογίας, ψηφιακές δημόσιες υπηρεσίες).

Τα δεδομένα DESI 2022 που χρησιμοποιήθηκαν, τυποποιήθηκαν με τιμές από 0 έως 1 (standardized) χρησιμοποιώντας την κατανομή minmax: $x' = \frac{x - \min(X)}{\max(X) - \min(X)}$. Όπου:

- x : η αρχική τιμή του χαρακτηριστικού
- x' : η κανονικοποιημένη τιμή
- $\min(X)$: η ελάχιστη τιμή του χαρακτηριστικού
- $\max(X)$: η μέγιστη τιμή του χαρακτηριστικού

Η τυποποίηση προέκυψε από την ανάγκη εξάλειψης των προβλημάτων που σχετίζονται με διαφορετικές μονάδες μέτρησης και διαφορετικές κλίμακες και η εξασφάλιση της ισότιμης συνεισφοράς των μεταβλητών στην ανάλυση, χωρίς να επηρεάζονται από μεγάλες διακυμάνσεις στις αρχικές τους τιμές (Mohamad & Usman, 2013).

Ο αριθμός των ομάδων ορίστηκε σε $k=3$, ώστε να διαχωρίζονται με επάρκεια οι χώρες σε επίπεδα υψηλής, μέτριας και χαμηλής ψηφιακής ανάπτυξης (HIGH=1, MEDIUM=2, LOW=3) και να υπάρχει συγκρισιμότητα με τις μελέτες των Bánhidí & Dobos (2023), Martín & Carnero

(2019) και Onuferoná et al. (2020). Ο αριθμός αυτός βοηθά ερμηνευτικά και στη βάση της πολιτικής ρητορικής της Ε.Ε. περί «ψηφιακού χάσματος» και «ψηφιακών πρωτοπόρων», διευκολύνοντας την κατηγοριοποίηση σε ουσιαστικά διαφοροποιημένες ομάδες.

Για τον υπολογισμό των αποστάσεων μεταξύ των παρατηρήσεων χρησιμοποιήθηκε η Manhattan distance (L1 norm). Ο K-means δοκιμάστηκε σε ολόκληρο το dataset, τόσο με τη χρήση Ευκλείδειας απόστασης και όσο και με τη χρήση της απόστασης Manhattan, με τη δεύτερη να δίνει μικρότερο και συνεπώς προτιμητέο intracluster sum of squared errors. Η επιλογή αυτή προτιμήθηκε επίσης, έναντι της πιο κοινής Ευκλείδειας απόστασης, καθώς η απόσταση Manhattan αποδίδει καλύτερα όταν τα δεδομένα είναι πολλών διαστάσεων ή οι μεταβλητές είναι ανεξάρτητες, κάτι που ισχύει συχνά σε κοινωνικοοικονομικά δεδομένα (Kaufman & Rousseeuw, 2005; Aggarwal, Hinneburg, & Keim, 2001). Επιπλέον, η παρουσία ακραίων τιμών, μπορεί να επηρεάσει δυσανάλογα την Ευκλείδεια απόσταση (Duda, Hart, & G.Stork, 2001).

Η τεχνική K-means++, χρησιμοποιήθηκε για την αρχικοποίηση των κέντρων καθώς οδηγεί συνήθως σε πιο σταθερά και ποιοτικά αποτελέσματα και πιο γρήγορη σύγκλιση σε σχέση με την τυχαία αρχικοποίηση. Ο K-means++ είναι μια βελτιωμένη έκδοση του αλγορίθμου K-means, η οποία επιλέγει αυτόματα καλύτερα σημεία εκκίνησης αντί να τα επιλέγει τυχαία. Η βασική ιδέα πίσω από τον K-means++ είναι ότι επιλέγει τα αρχικά κέντρα συστάδων με έξυπνο τρόπο, ώστε να διασφαλίζει ότι είναι διασκορπισμένα, γεγονός που βοηθά τον αλγόριθμο να συγκλίνει ταχύτερα και να δίνει καλύτερα αποτελέσματα ομαδοποίησης (Arthur & Vassilvitskii, 2007).

Για να διασφαλιστεί η αναπαραγωγικότητα των αποτελεσμάτων, ορίστηκε προκαθορισμένη τιμή τυχαίας αρχικοποίησης (seed = 100). Η πρακτική αυτή είναι σημαντική σε μεθόδους όπως το K-means, καθώς επιτρέπει τη σταθερή επανάληψη των ίδιων αποτελεσμάτων σε κάθε εκτέλεση του αλγορίθμου ως εκ τούτου χρησιμοποιήθηκε σταθερό και προκαθορισμένο seed το οποίο είναι απαραίτητο για την αξιοπιστία αλλά και την συνέπεια των αναλύσεων (Bishop, 2006). Η βελτιστοποίηση του μοντέλου αξιολογήθηκε με το Sum of Squared Errors, το οποίο μετρά την εσωτερική συνοχή των ομάδων (Kaufman & Rousseeuw, 2009, p.43), και έπειτα από δοκιμές επιλέχθηκε το seed που οδηγεί σε ελαχιστοποίηση του Sum of Squared Errors. Τέλος, ο αλγόριθμος ρυθμίστηκε με μέγιστο αριθμό επαναλήψεων (maxIterations) 1.000.000 για να επιτευχθεί βέλτιστη σύγκλιση.

Η ανάλυση διεξήχθη σε δύο φάσεις:

1. Συνολικός Δείκτης DESI 2022: Ομαδοποίηση με βάση όλες τις μεταβλητές του δείκτη.
2. Ανάλυση ανά Συνιστώσα: Ομαδοποίηση για κάθε μία από τις τέσσερις συνιστώσες (Human Capital, Connectivity, Integration of Digital Technology, Digital Public Services).

Τα αποτελέσματα της συσταδοποίησης των κρατών-μελών της Ε.Ε. παρουσιάζονται στον Πίνακα (49) και ο οποίος θα αναλυθεί στις ενότητες 6.2-6.7.

Country	DESI 2022 Total	DESI 2022 Human Capital Component	DESI 2022 Connectivity Component	DESI 2022 Integration Of Digital Technology Component	DESI 2022 Digital Public Services Component
Austria	2	2	2	2	1
Belgium	1	2	1	1	2
Bulgaria	3	3	3	3	3
Cyprus	2	3	2	3	3
Czech Republic	2	3	2	2	3
Germany	2	3	2	2	3
Denmark	1	2	1	1	1
Estonia	2	2	3	3	1
Greece	3	3	2	3	3
Spain	2	2	1	2	1
Finland	1	1	2	1	1
France	2	2	2	3	3
Croatia	2	2	2	2	3
Hungary	3	3	1	3	2
Ireland	1	2	1	1	1
Italy	3	3	2	3	3
Lithuania	2	3	3	2	1
Luxembourg	1	2	1	2	2
Latvia	3	3	3	3	2
Malta	1	2	1	1	2
Netherlands	1	1	1	1	1
Poland	3	3	3	3	3
Portugal	2	3	1	2	2
Romania	3	3	3	3	3
Sweden	1	1	1	1	1
Slovenia	2	3	3	2	1
Slovakia	3	3	2	3	3

Πίνακας 49. Ομαδοποίηση κρατών-μελών βάσει του DESI 2022 και των συνιστωσών του (HIGH=1, MEDIUM=2, LOW=3).

6.2 Αποτελέσματα για το σύνολο των χαρακτηριστικών

Η εφαρμογή του αλγορίθμου K-means στο σύνολο των μεταβλητών του DESI 2022 παρήγαγε τρεις ομάδες κρατών-μελών της ΕΕ, οι οποίες διακρίνονται βάσει του συνολικού επιπέδου ψηφιακής ανάπτυξης. Τα αποτελέσματα παρουσιάζονται στον πίνακα των συστάδων (Πίνακας 49) και απεικονίζονται στον αντίστοιχο χάρτη (Χάρτης 4), όπου παρουσιάζεται η γεωγραφική κατανομή των ψηφιακά ανεπτυγμένων, μεσαίου επιπέδου και λιγότερο αναπτυγμένων χωρών.

Οι χώρες με υψηλή ψηφιακή ανάπτυξη ανήκουν κυρίως στη Βόρεια και Κεντρική Ευρώπη, ενώ οι χώρες της Νότιας και Ανατολικής Ευρώπης συγκεντρώνονται κυρίως στις κατηγορίες μεσαίας και χαμηλής ανάπτυξης. Η συγκεκριμένη κατηγοριοποίηση παρουσιάζει συσχετίσεις με προηγούμενες μελέτες που εξετάζουν τη γενικότερη οικονομική και βιώσιμη ανάπτυξη.

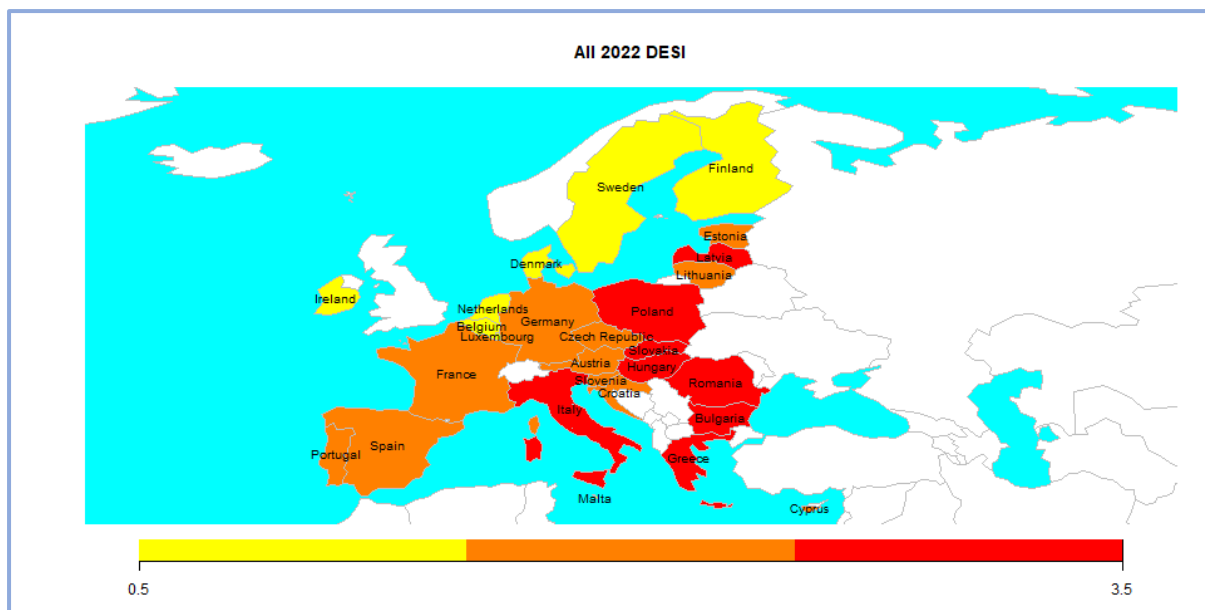
Η εφαρμογή του K-means στον συνολικό δείκτη DESI 2022 ανέδειξε τρεις ομάδες χωρών, οι οποίες αντιστοιχούν σε:

Υψηλό Επίπεδο Ψηφιακής Ανάπτυξης: Χώρες της Βόρειας/Δυτικής Ευρώπης (π.χ. Φινλανδία, Σουηδία, Δανία, Ιρλανδία).

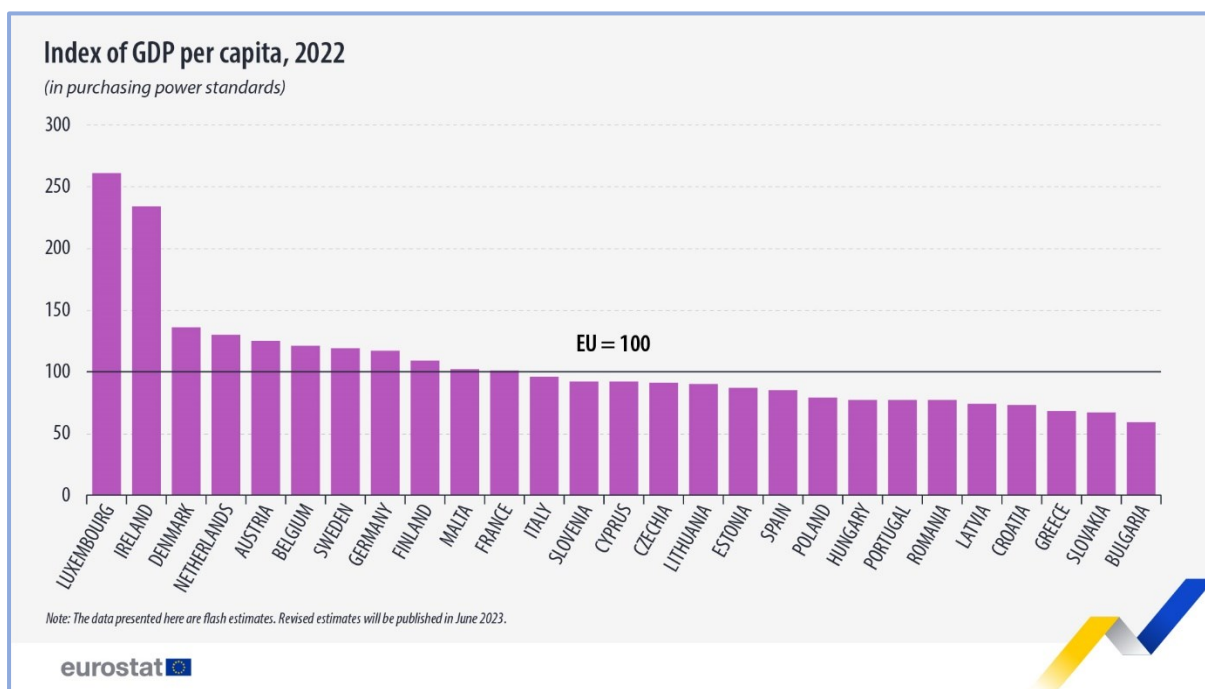
Μέσο Επίπεδο Ψηφιακής Ανάπτυξης: Χώρες της Νότιας/Κεντρικής Ευρώπης (π.χ. Γερμανία, Αυστρία).

Χαμηλό Επίπεδο Ψηφιακής Ανάπτυξης: Χώρες της Ανατολικής Ευρώπης (π.χ. Βουλγαρία, Ρουμανία).

Η γεωγραφική κατανομή των ομάδων (Χάρτης 4) επιβεβαιώνει τις γνωστές διακρίσεις μεταξύ Βορρά και Νότου, Δύσης και Ανατολής, καθώς και τη συσχέτιση μεταξύ ψηφιακής ωριμότητας και οικονομικής ανάπτυξης (Γράφημα 6), ταυτόχρονα όμως αναδεικνύεται και επιτυχία πρακτικών με προσανατολισμό την ψηφιακή πρόοδο, χωρών λιγότερο αναπτυγμένων οικονομικά, π.χ. Ισπανία και Πορτογαλία.



Χάρτης 4. Κατηγοριοποίηση των χωρών της ΕΕ βάσει των συνολικών χαρακτηριστικών του Δείκτη DESI 2022.

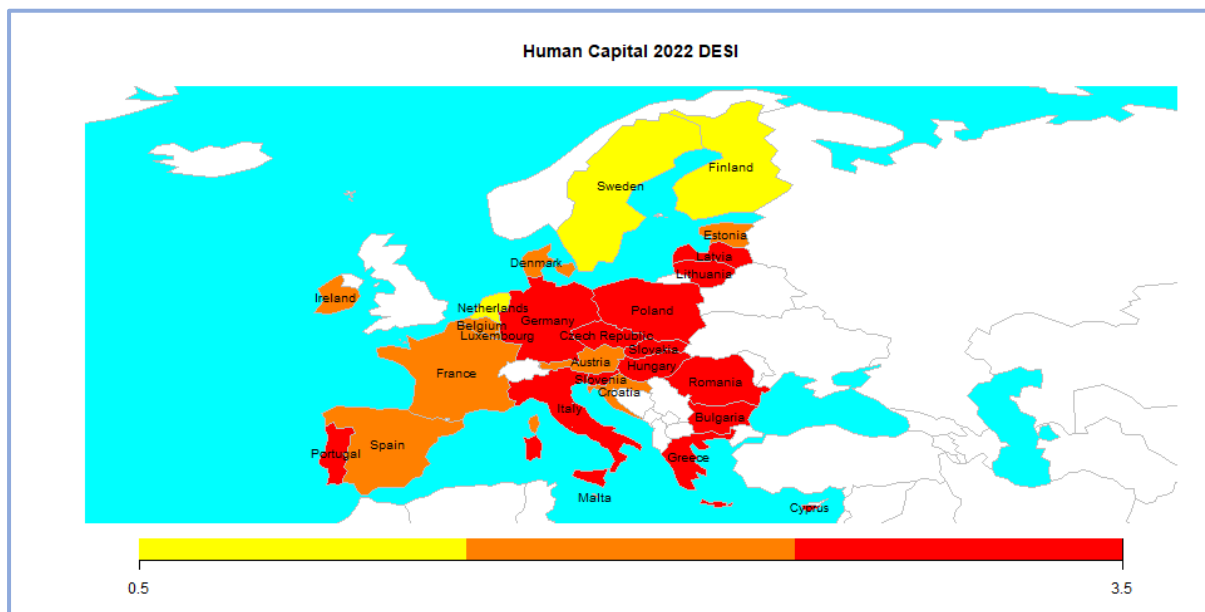


Γράφημα 6. Κατάταξη των κρατών-μελών της Ε.Ε. βάσει του κατά κεφαλήν ΑΕΠ (Eurostat)

6.3 Αποτελέσματα ανά συνιστώσα του DESI - Ανθρώπινο Κεφάλαιο (Human Capital)

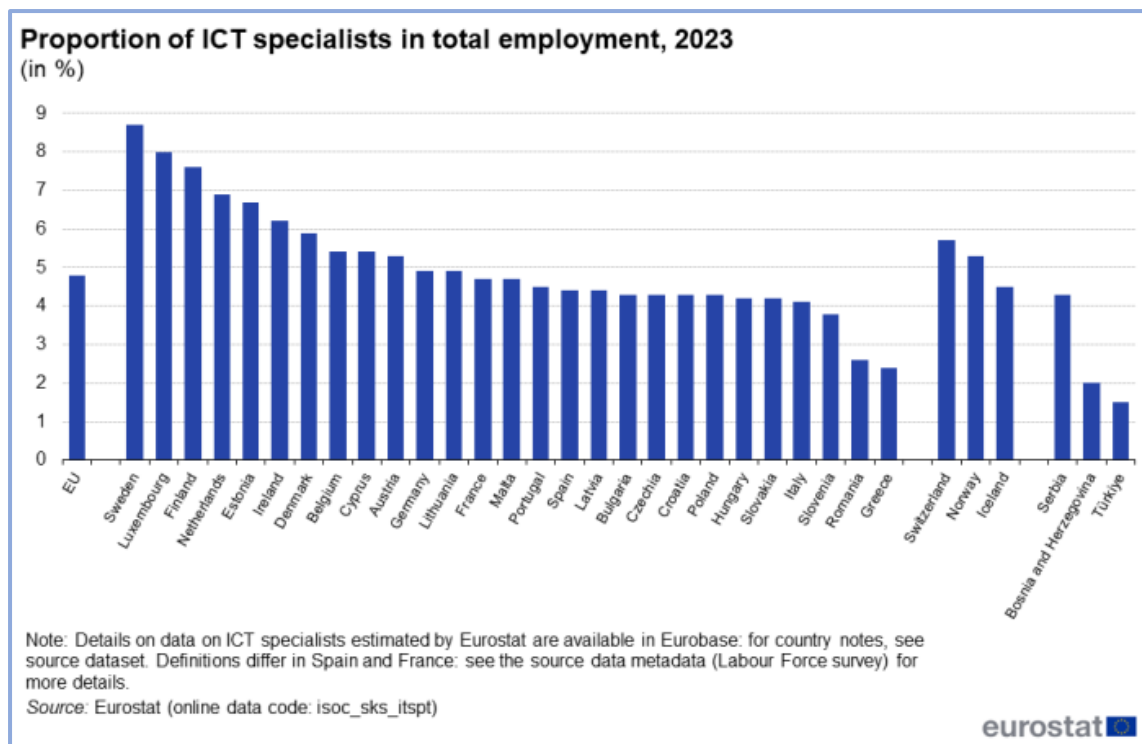
Η συνιστώσα του Ανθρώπινου Κεφαλαίου μετρά την ανάπτυξη ψηφιακών δεξιοτήτων του πληθυσμού, την απασχόληση ειδικών σε τεχνολογίες πληροφορικής και επικοινωνιών (ΤΠΕ) και τη συμμετοχή σε σχετικά προγράμματα εκπαίδευσης.

Η ομαδοποίηση των χωρών βάσει της συνιστώσας του Ανθρώπινου Κεφαλαίου ανέδειξε διαφοροποιήσεις στις ψηφιακές δεξιότητες του πληθυσμού και την παρουσία εξειδικευμένων επαγγελματιών ΤΠΕ. Οι χώρες με τις υψηλότερες επιδόσεις είναι η Φινλανδία, η Σουηδία και η Ολλανδία, ενώ τις χαμηλότερες επιδόσεις παρουσιάζουν οι χώρες σχεδόν όλη η ανατολική Ευρώπη και μέρος της κεντρικής Ευρώπης. Τα αποτελέσματα αποτυπώνονται στον σχετικό χάρτη.



Χάρτης 5. Κατηγοριοποίηση των χωρών της ΕΕ βάσει των χαρακτηριστικών της συνιστώσας «Ανθρώπινο Κεφάλαιο» του Δείκτη DESI 2022

Αξίζει να σημειωθεί πως οι χώρες με την υψηλότερη απόδοση (π.χ. Σκανδιναβικές χώρες) διαθέτουν ισχυρό εκπαιδευτικό σύστημα και υψηλά ποσοστά απασχόλησης στον τομέα των ΤΠΕ (Εικόνα).

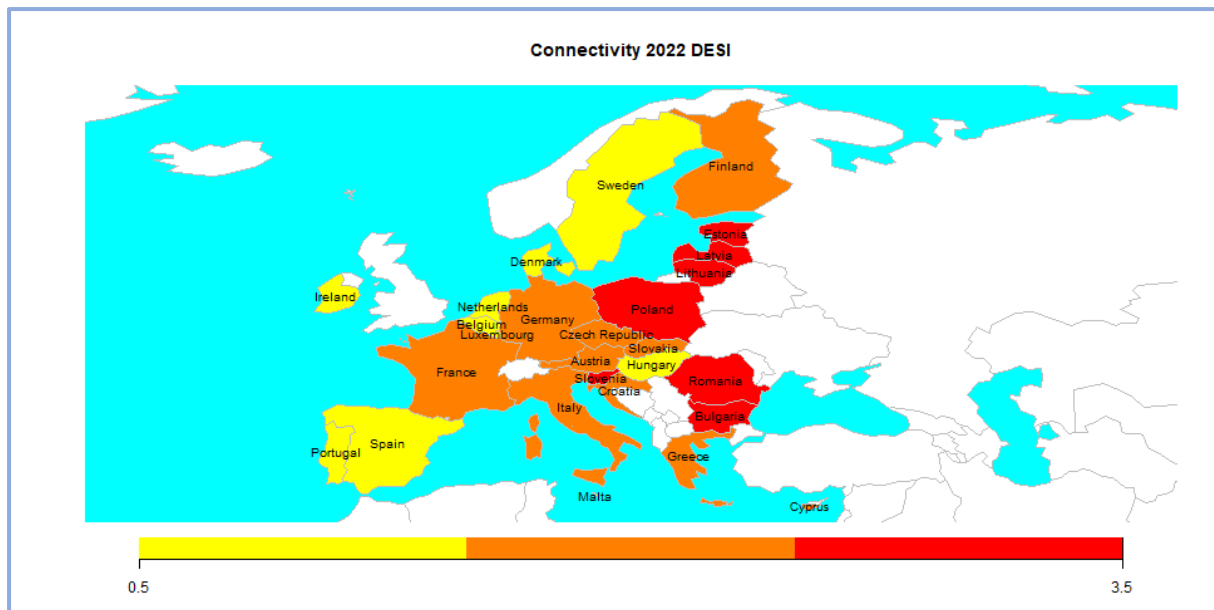


Εικόνα 7. Ποσοστό ειδικών ICT επί του συνόλου των απασχολούμενων (Eurostat, 2023)

6.4 Ανάλυση ανά συνιστώσα του DESI-Συνδεσιμότητα (Connectivity)

Η συνδεσιμότητα εξετάζει τη διαθεσιμότητα και τη χρήση ευρυζωνικών δικτύων υψηλής ταχύτητας, καθώς και την κάλυψη 5G. Οι χώρες με ανεπτυγμένες υποδομές, όπως η Ολλανδία η Δανία, η Σουηδία, η Ιρλανδία, η Ισπανία, η Πορτογαλία και η Ουγγαρία κατατάσσονται στις υψηλότερες κλάσεις, ενώ χώρες της Ανατολικής Ευρώπης εμφανίζουν χαμηλότερες επιδόσεις.

Η ομαδοποίηση ανέδειξε ανισότητες στο πλαίσιο της συνδεσιμότητας και στην πρόσβαση σε γρήγορες και αξιόπιστες συνδέσεις, όπως απεικονίζεται στον αντίστοιχο χάρτη (Χάρτης 6). Αυτές οι ανισότητες σχετίζονται με τη γενικότερη οικονομική ανάπτυξη και τις εθνικές επενδύσεις σε τηλεπικοινωνιακές υποδομές.

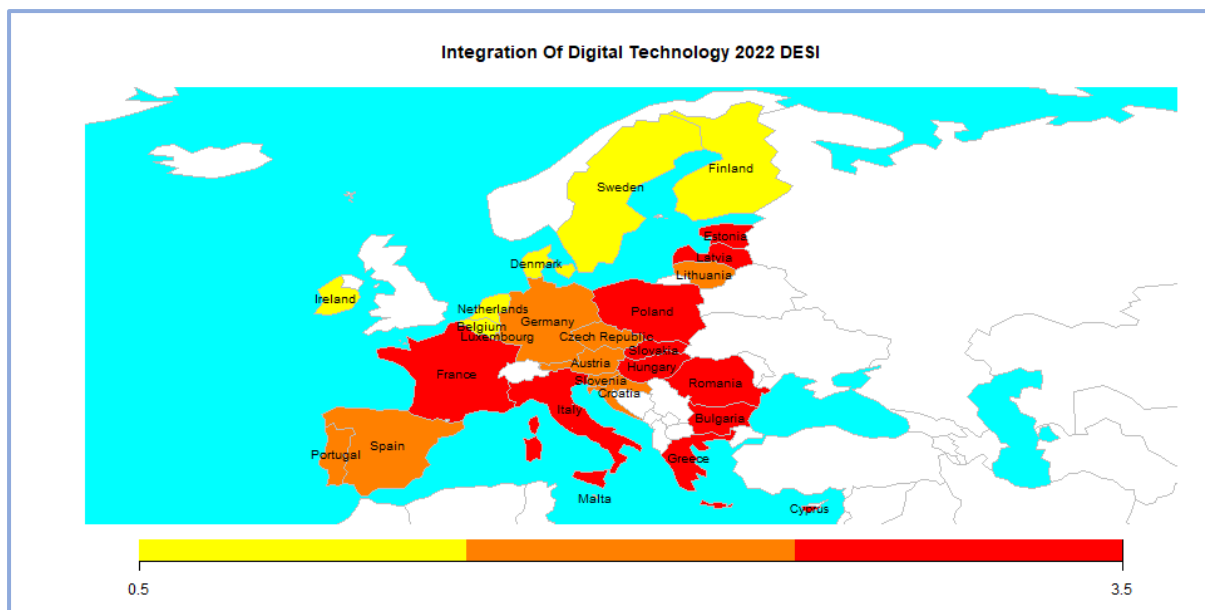


Χάρτης 6. Κατηγοριοποίηση των χωρών της ΕΕ βάσει των χαρακτηριστικών της συνιστώσας «Συνδεσιμότητα» του Δείκτη DESI 2022

6.5 Ανάλυση ανά συνιστώσα του DESI - Ενσωμάτωση Ψηφιακής Τεχνολογίας (Integration of Digital Technology)

Η συσταδοποίηση των χωρών βάσει της χρήσης ψηφιακών τεχνολογιών από επιχειρήσεις αποκάλυψε διακριτές ομάδες. Οι προηγμένες οικονομίες, της βορειοδυτικής Ευρώπης, εμφανίζουν υψηλά ποσοστά ενσωμάτωσης τεχνολογιών (Σκανδιναβία, Ολλανδία, Ιρλανδία), ενώ χώρες με χαμηλότερο ΑΕΠ (χώρες της νοτιοανατολική Ευρώπης αλλά και η Γαλλία) εμφανίζουν μικρότερη ενσωμάτωση των ψηφιακών εργαλείων.

Οι διαφορετικές κλάσεις παρουσιάζονται στον σχετικό χάρτη (Χάρτης 7), όπου αποτυπώνονται οι ανισότητες μεταξύ των επιχειρηματικών τομέων των κρατών-μελών.

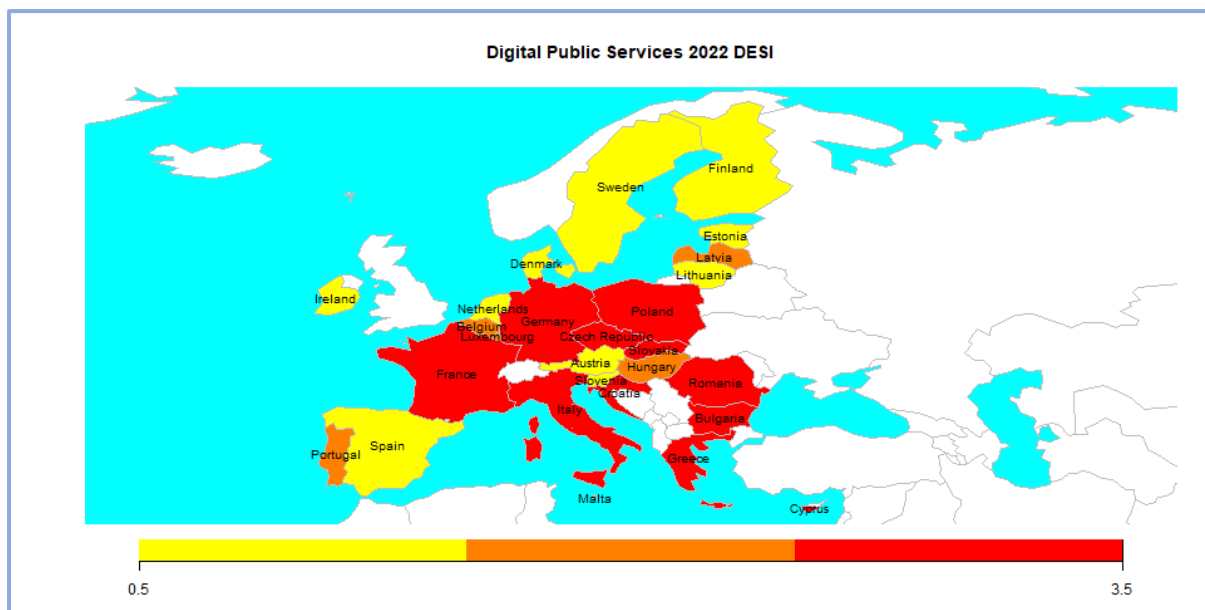


Χάρτης 7. Κατηγοριοποίηση των χωρών της ΕΕ βάσει των χαρακτηριστικών της συνιστώσας «Ενσωμάτωση της Ψηφιακής Τεχνολογίας» του Δείκτη DESI 2022

6.6 Ανάλυση ανά συνιστώσα του DESI Ψηφιακές Δημόσιες Υπηρεσίες (Digital Public Services)

Η ανάλυση αυτής της συνιστώσας, ανέδειξε τρεις διακριτές ομάδες χωρών με βάση την παροχή και χρήση, την προσβασιμότητα και την ποιότητα των ψηφιακών δημόσιων υπηρεσιών. Οι βόρειες και δυτικοευρωπαϊκές χώρες παρουσιάζουν υψηλή απόδοση, ενώ οι χώρες της Ανατολικής και Νότιας Ευρώπης εμφανίζουν χαμηλότερα επίπεδα ψηφιοποίησης των δημοσίων υπηρεσιών.

Η ταξινόμηση (Χάρτης 8) αντισταθμίζει τις διαφορές στη διακυβερνητική ψηφιακή στρατηγική και παρατηρείται ότι χώρες όπως η Εσθονία, η Φινλανδία, η Σουηδία, η Δανία, Ολλανδία κτλ. αποτελούν ηγέτες στον τομέα των ψηφιακών δημόσιων υπηρεσιών, ενώ κρίνεται απαραίτητο να πραγματοποιηθούν τομές στην παραπάνω συνιστώσα, στο μεγαλύτερο μέρος της κεντρικής και ανατολικής Ευρώπης, γεγονός που επιβεβαιώνεται και από αντίστοιχη μελέτη των Ηνωμένων εθνών σε σχέση με την ηλεκτρονική διακυβέρνηση (United Nations, 2022).



Χάρτης 8. Κατηγοριοποίηση των χωρών της ΕΕ βάσει των χαρακτηριστικών της συνιστώσας «Ψηφιακές Δημόσιες Υπηρεσίες» του Δείκτη DESI 2022

6.7 Συσχέτιση με προηγούμενες έρευνες

Η ανάλυση του διαγράμματος συσχέτισης (Εικόνα 8) αποκαλύπτει σημαντικές θετικές συσχετίσεις (0-1) μεταξύ των δεικτών ψηφιακής ανάπτυξης (DESI 2022) και άλλων μετρήσεων οικονομικής, βιώσιμης και ψηφιακής προόδου.

Παρατηρείται ισχυρή θετική συσχέτιση ($\rho=0.83$) μεταξύ του συνολικού DESI 2022 και του δείκτη Digital Development.Class.2023, γεγονός που επιβεβαιώνει τη σταθερότητα των αποτελεσμάτων και αποτελεί μέτρο αξιολόγησης για την παρούσα έρευνα.

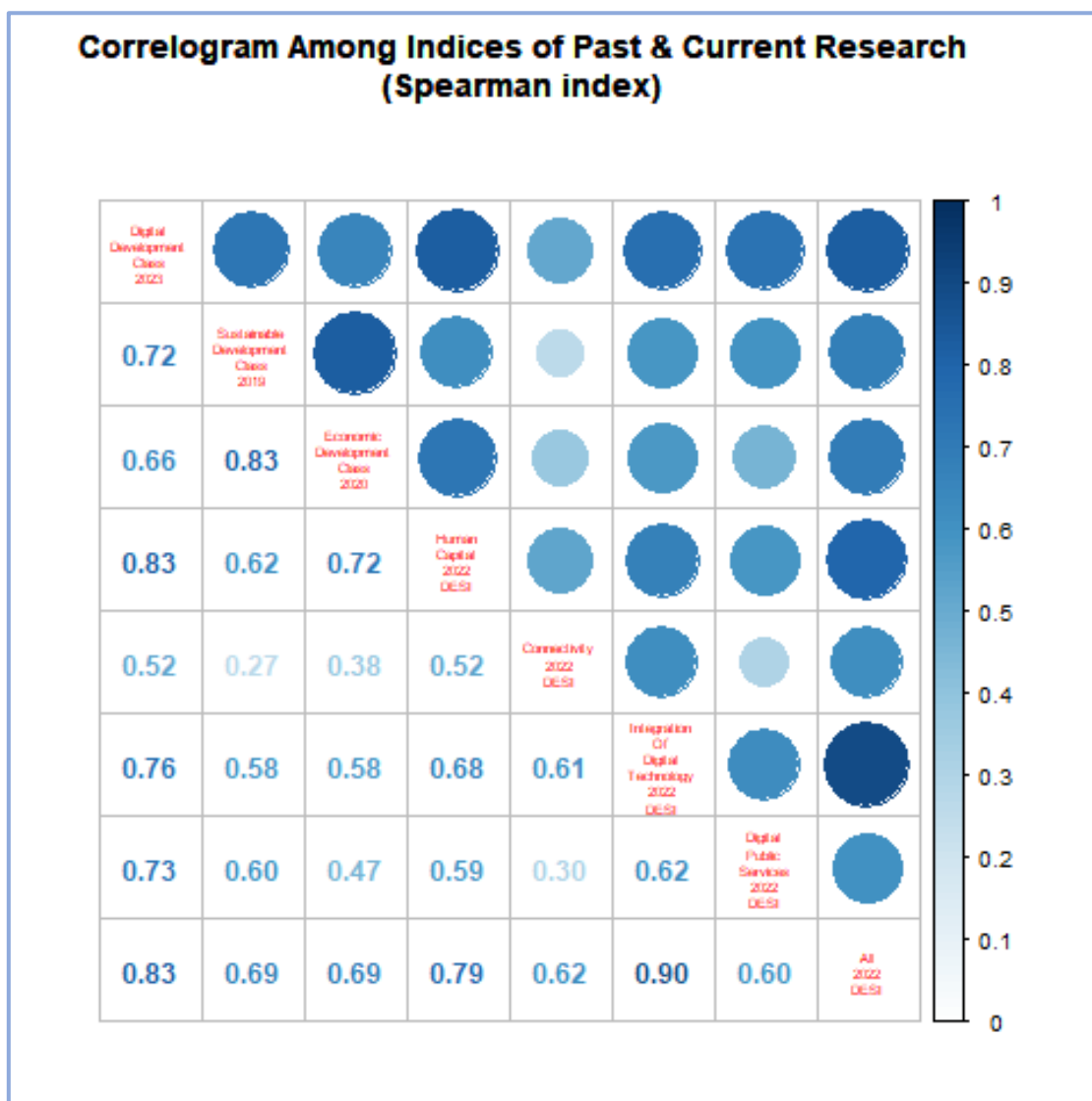
Ο δείκτης οικονομικής ανάπτυξης παρουσιάζει μέτρια προς υψηλή συσχέτιση με τον DESI 2022 ($\rho=0.69$). Αυτό υποδηλώνει ότι οι χώρες με ισχυρή οικονομία π.χ. Δανία, Ολλανδία, Λουξεμβούργο κτλ. τείνουν να έχουν καλύτερες ψηφιακές υποδομές και υψηλότερη ενσωμάτωση τεχνολογιών. Ωστόσο, υπάρχουν εξαιρέσεις (π.χ. Ισπανία, Πορτογαλία), όπου η εξαιρετική τους ψηφιακή πρόοδος δεν συνάδει με το οικονομικό τους επίπεδο, και αυτό οφείλεται πιθανά σε στρατηγικές επενδύσεις ψηφιακής ετοιμότητας αλλά και σε πολιτική πρωτοβουλία για ψηφιακές μεταρρυθμίσεις. Παρατηρείται επίσης ότι ο δείκτης οικονομικής ανάπτυξης έχει μέτρια προς υψηλή συσχέτιση με την ψηφιακή κατάρτιση του πληθυσμού (Human Capital) και την ενσωμάτωση της ψηφιακής τεχνολογίας.

Η συσχέτιση του δείκτη DESI 2022 με τη Βιώσιμη Ανάπτυξη είναι θετική αλλά ασθενέστερη ($\rho=0.69$), γεγονός που συμβαδίζει με τη θεωρία ότι οι χώρες που επενδύουν συνολικότερα σε κοινωνικές, οικονομικές και περιβαλλοντικές πολιτικές τείνουν να έχουν και πιο προηγμένα

ψηφιακά συστήματα. Ωστόσο, η μικρότερη έντασης συσχέτιση, σε σχέση με την οικονομική ανάπτυξη, υποδεικνύει ότι η ψηφιακή μετάβαση μπορεί να προωθηθεί και ανεξάρτητα από περιβαλλοντικούς στόχους.

Οι δείκτες Human Capital και Integration of Digital Technology εμφανίζουν τις υψηλότερες συσχετίσεις με την οικονομική και βιώσιμη ανάπτυξη, καθώς αντικατοπτρίζουν βασικούς παράγοντες παραγωγικότητας (εκπαίδευση, επιχειρήσεις).

Οι Ψηφιακές Δημόσιες Υπηρεσίες σχετίζονται άμεσα με την ψηφιακή πρόοδο αλλά και με τη βιώσιμη ανάπτυξη, καθώς σχετίζονται συνολικότερα με θεσμικούς παράγοντες (π.χ. ποιότητα διακυβέρνησης), γεγονός που εξηγεί τη διαφοροποίηση μεταξύ χωρών με παρόμοιο ΑΕΠ (π.χ. Ισπανία vs. Πολωνία).



Εικόνα 8. Συσχέτιση μεταξύ των δεικτών DESI 2022 και παλαιότερων ταξινομήσεων ψηφιακής, οικονομικής και βιώσιμης ανάπτυξης. Οι αριθμητικές τιμές αντιπροσωπεύουν τον συντελεστή συσχέτισης

μεταξύ των ζευγών μεταβλητών, ενώ οι αποχρώσεις και τα μεγέθη των κύκλων υποδεικνύουν το επίπεδο της συσχέτισης: όσο πιο σκούρο και μεγαλύτερο το σύμβολο, τόσο ισχυρότερη η συσχέτιση.

6.8 Συμπεράσματα

Τα αποτελέσματα της ομαδοποίησης των κρατών-μελών της Ε.Ε. βάσει του δείκτη DESI 2022 επιβεβαιώνουν την ύπαρξη τριών διακριτών ομάδων μεταξύ των κρατών-μελών της ΕΕ, όσον αφορά την ψηφιακή τους ανάπτυξη. Ειδικότερα, παρατηρείται και γεωγραφική κατανομή των παραπάνω ομάδων:

Ομάδα 1 (Υψηλό Επίπεδο): Περιλαμβάνει τις χώρες της Βόρειας/Δυτικής Ευρώπης (Φινλανδία, Σουηδία, Ολλανδία, Δανία, Ιρλανδία, Λουξεμβούργο) με συστηματικές επενδύσεις σε ψηφιακές δεξιότητες, υποδομές και καινοτομία.

Ομάδα 2 (Μέσο Επίπεδο): Περιλαμβάνει τις χώρες της Νότιας/Κεντρικής Ευρώπης (Γερμανία, Γαλλία, Τσεχία, Ισπανία, Πορτογαλία) με ετερογενείς επιδόσεις ανά συνιστώσα, συχνά λόγω θεσμικών ή γεωγραφικών ανισοτήτων.

Ομάδα 3 (Χαμηλό Επίπεδο): Περιλαμβάνει τις χώρες της Ανατολικής και Νότιας Ευρώπης (Βουλγαρία, Ρουμανία, Πολωνία, Ελλάδα) με το χαμηλότερο επίπεδο ψηφιακής προόδου, και οι οποίες έχουν ελλείψεις σε βασικούς τομείς της ψηφιακής ανάπτυξης (π.χ. εκπαίδευση ΤΠΕ, ευρυζωνική κάλυψη).

Συγκρίνοντας αυτά τα αποτελέσματα με παλαιότερες έρευνες, διαπιστώνεται ότι οι χώρες που προηγούνται ψηφιακά έχουν επίσης υψηλές επιδόσεις στη βιώσιμη και οικονομική ανάπτυξη. Αυτό επιβεβαιώνεται από τη συσχέτιση των κλάσεων του DESI 2022 με τις κατηγορίες των Digital Development 2023, Sustainable Development 2019 και Economic Development 2020. Τα διαγράμματα συσχέτισης αναδεικνύουν θετικές σχέσεις μεταξύ αυτών των δεικτών, ενισχύοντας την άποψη ότι η ψηφιακή πρόοδος συνδέεται άμεσα με τη γενικότερη ανάπτυξη των κρατών.

Η Φινλανδία, η Δανία, η Σουηδία και η Ολλανδία βρίσκονται σταθερά στην υψηλότερη κλάση ψηφιακής ωριμότητας, ανεξαρτήτως της συνιστώσας του δείκτη που αναλύεται. Οι χώρες αυτές παρουσιάζουν ισχυρές επιδόσεις τόσο σε τεχνικούς δείκτες (όπως η συνδεσιμότητα) όσο και σε κοινωνικούς (όπως οι ψηφιακές δεξιότητες του πληθυσμού και η χρήση e-government υπηρεσιών). Στον αντίποδα, η Βουλγαρία, η Ρουμανία και σε αρκετές περιπτώσεις και η Ελλάδα κατατάσσονται στις λιγότερο αναπτυγμένες ψηφιακά χώρες, παρουσιάζοντας σημαντικά χαμηλότερες τιμές σε όλους σχεδόν τους δείκτες του DESI.

Ιδιαίτερο ενδιαφέρον παρουσιάζει το γεγονός ότι η ανάλυση ανά συνιστώσα φέρνει στο φως διαφοροποιήσεις που δεν είναι ορατές όταν εξετάζεται το σύνολο των χαρακτηριστικών. Για παράδειγμα, η Ισπανία εμφανίζει μεσαίου επιπέδου ψηφιακής ωριμότητας συνολικά, ωστόσο διαθέτει εξαιρετικά υψηλή επίδοση στη συνιστώσα «Ψηφιακές Δημόσιες Υπηρεσίες», λόγω της μακρόχρονης επένδυσης σε e-government (Bánhidí & Dobos, 2023). Αντίστοιχα, χώρες όπως η Ελλάδα, παρόλο που κατατάσσεται στους ουραγούς της ψηφιακής προόδου, έχει σημειώσει βελτίωση στις υποδομές συνδεσιμότητας. Η Πορτογαλία από την άλλη, παρόλο που έχει επιδείξει μεγάλη πρόοδο στην ψηφιακή μετάβαση, φαίνεται να υστερεί στον τομέα του Ανθρωπίνου Κεφαλαίου.

Τα ευρήματα της μη επιβλεπόμενης μάθησης, μπορούν να χρησιμοποιηθούν για τη σχεδίαση εστιασμένων πολιτικών που θα ενισχύσουν τη ψηφιακή σύγκλιση και την καινοτομία, σύμφωνα με τους στόχους της Ψηφιακής Δεκαετίας 2030 (European Commission, 2021).

7. Ανάλυση ταξινόμησης με βάση τον δείκτη DESI

7.1 Μεθοδολογία

Στην παρούσα ενότητα εφαρμόζονται τεχνικές επιβλεπόμενης μάθησης (supervised learning), με στόχο την πρόβλεψη της ψηφιακής κατηγορίας στην οποία ανήκει κάθε κράτος-μέλος της Ευρωπαϊκής Ένωσης. Οι ετικέτες κλάσης, οι οποίες χρησιμοποιήθηκαν για την εκπαίδευση των αλγορίθμων προέρχονται από τα αποτελέσματα της μη επιβλεπόμενης μάθησης που παρουσιάστηκαν στο Κεφάλαιο 6. Ειδικότερα, οι ομάδες χωρών (clusters) που δημιουργήθηκαν μέσω του αλγορίθμου K-means, με βάση τα χαρακτηριστικά του δείκτη DESI 2022, αποτέλεσαν το στόχο πρόβλεψης για τα ταξινομητικά μοντέλα. Πιο συγκεκριμένα, η συσταδοποίηση του κεφαλαίου 6, αποδίδει κατηγορίες σύμφωνα με τις ομοιότητες μεταξύ των χωρών, από την άλλη, η ταξινόμηση επιχειρεί να εντοπίσει ποια χαρακτηριστικά του δείκτη DESI συμβάλλουν περισσότερο στην πρόβλεψη της κλάσης, δηλαδή στην κατηγοριοποίηση ψηφιακής προόδου.

Η διαδικασία περιλαμβάνει τα εξής στάδια:

Δημιουργία νέων αρχείων δεδομένων τύπου .arff, όπου ως ετικέτα κλάσης έχει τεθεί η κλάση που προέκυψε από τον K-means, είτε στο σύνολο του δείκτη, είτε ανά συνιστώσα.

Εκπαίδευση και αξιολόγηση αλγορίθμων με χρήση της μεθόδου 10-fold cross-validation, που εξασφαλίζει την εγκυρότητα και σταθερότητα των αποτελεσμάτων (Kohavi, 1995).

Εφαρμόζονται ξανά πέντε διαφορετικοί αλγόριθμοι ταξινόμησης: Decision Tree (J48) (υλοποίηση του C4.5 Decision Tree), Random Forest (RDF), Adaboost (ADA), Multilayer Perceptron (MLP), Support Vector Machine (SVM)

Για τη βελτίωση της ακρίβειας και της ερμηνείας των μοντέλων, εφαρμόστηκε επιλογή χαρακτηριστικών (feature selection) με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise και εν συνεχεία με τον αλγόριθμο Genetic Search, με στόχο την ανάδειξη των σημαντικότερων χαρακτηριστικών του δείκτη DESI για κάθε ταξινομητή (Kohavi & John, 1997).

Η απόδοση των ταξινομητών αξιολογήθηκε με τις μετρικές Accuracy, Precision, Recall και F1-Score, οι οποίες παρέχουν πλήρη εικόνα της ικανότητας του μοντέλου να διακρίνει τις κατηγορίες και να αποφεύγει ψευδείς ταξινομήσεις.

Η συγκεκριμένη μεθοδολογία επιτρέπει την επικύρωση των αποτελεσμάτων της συσταδοποίησης, παρέχοντας εξηγήσεις για τη συγκρότηση κάθε συστάδας – cluster. Παράλληλα, εντοπίζει τα χαρακτηριστικά του δείκτη DESI που διαφοροποιούν τις χώρες μεταξύ τους ως προς την ψηφιακή τους πρόοδο. Η συνδυαστική αυτή προσέγγιση αποτελεί απόδειξη της συμπληρωματικότητας της επιβλεπόμενης και μη επιβλεπόμενης μάθησης στην ανάλυση σύνθετων κοινωνικοτεχνικών φαινομένων.

7.2 Αποτελέσματα για το σύνολο των χαρακτηριστικών του δείκτη DESI

7.2.1 Decision Tree (J48)

Ο αλγόριθμος Decision Tree (J48) εφαρμόστηκε στο σύνολο των δεδομένων του δείκτη DESI 2022, χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.727	0.727	0.727	HIGH	
0.727	1.000	0.842	MEDIUM	
1.000	0.625	0.769	LOW	
			Total	0.7778

Πίνακας 50. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48).

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 51) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 52). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 65. Ο Decision Tree (J48) εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.900	0.857	0.857	HIGH	
0.800	0.889	0.889	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.8889

Πίνακας 51. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.900	0.857	0.857	HIGH	
0.800	0.889	0.889	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.8889

Πίνακας 52. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

7.2.2 Random Forest

Ο αλγόριθμος Random Forest εφαρμόστηκε στο σύνολο των δεδομένων του δείκτη DESI 2022, χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.600	0.818	0.692	HIGH	
1.000	0.625	0.769	MEDIUM	
0.714	0.625	0.667	LOW	
			Total	0.7037

Πίνακας 53. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest.

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο GreedyStepwise (Πίνακας 54) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 55). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 65. Ο Random Forest εφαρμόστηκε ξανά με τα παραπάνω επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.833	0.909	0.870	HIGH	
0.875	0.875	0.875	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.8889

Πίνακας 54. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.909	0.909	0.909	HIGH	
1.000	0.875	0.933	MEDIUM	
0.889	1.000	0.941	LOW	
			Total	0.9259

Πίνακας 55. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

7.2.3 Adaboost

Εν συνεχεία εφαρμόστηκε ο αλγόριθμος Adaboost, στο σύνολο των δεδομένων DESI 2022, χωρίς περαιτέρω παραμετροποίηση, και οδήγησε στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.583	0.636	0.609	HIGH	
0.700	0.875	0.778	MEDIUM	
0.800	0.500	0.615	LOW	
			Total	0.6667

Πίνακας 56. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost.

Η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 57) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 58) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 65. Ο Adaboost εφαρμόστηκε ξανά με τα επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.833	0.909	0.870	HIGH	
0.778	0.875	0.824	MEDIUM	
1.000	0.750	0.857	LOW	
			Total	0.8519

Πίνακας 57. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.818	0.818	0.818	HIGH	
0.778	0.875	0.824	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.8519

Πίνακας 58. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

7.2.4 Multilayer Perceptron

Ο αλγόριθμος Multilayer Perceptron εφαρμόστηκε στο σύνολο των δεδομένων DESI 2022 με την παραμετροποίηση “hidden layers=3”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.800	0.727	0.762	HIGH	
0.875	0.875	0.875	MEDIUM	
0.778	0.875	0.824	LOW	
			Total	0.8148

Πίνακας 59. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron.

Ακολούθησε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 60) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 61). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 65. Ο Multilayer Perceptron εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.786	1.000	0.880	HIGH	
1.000	0.750	0.857	MEDIUM	
1.000	0.875	0.933	LOW	
			Total	0.8889

Πίνακας 60. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
1.000	0.909	0.952	HIGH	
0.889	1.000	0.941	MEDIUM	
1.000	1.000	1.000	LOW	
			Total	0.9630

Πίνακας 61. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

7.2.5 Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) εφαρμόστηκε στα δεδομένα με την παραμετροποίηση “Kernel type=linear”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.818	0.818	0.818	HIGH	
0.875	0.875	0.875	MEDIUM	
0.875	0.875	0.875	LOW	
			Total	0.8519

Πίνακας 62. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine.

Τέλος, η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο GreedyStepwise (Πίνακας 63) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 64) κατέληξε στα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 65. Ο Support Vector Machine εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.909	0.952	HIGH	
1.000	1.000	1.000	MEDIUM	
0.889	1.000	0.941	LOW	
			Total	0.9630

Πίνακας 63. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
1.000	1.000	1.000	HIGH	
1.000	1.000	1.000	MEDIUM	
1.000	1.000	1.000	LOW	
			Total	1.0000

Πίνακας 64. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

Variable	J48	RDF	ADA	MLP	SVM	J48	RDF	ADA	MLP	SVM
At least basic digital skills										
Above basic digital skills										
At least basic digital content creation skills										
ICT specialists										
Female ICT specialists										
Enterprises providing ICT training										
ICT graduates										
Overall fixed broadband take-up										
At least 100 Mbps fixed broadband take-up										
At least 1 Gbps broadband take-up										
Fast broadband (NGA) coverage										
Fixed Very High Capacity Network (VHCN) coverage										
5G spectrum										
Overall 5G coverage										
Mobile broadband take-up										
Broadband price index										
SMEs with at least a basic level of digital intensity										
Electronic information sharing										
Social media										
Big data										
Cloud										
Artificial intelligence										
ICT for environmental sustainability										
e-Invoices										
SMEs selling online										
e-Commerce turnover										
Selling online cross-border										
e-Government users										
Pre-filled Forms										
Digital public services for citizens										
Digital public services for businesses										
Open data										

Πίνακας 65. Αποτελέσματα Επιλογής Χαρακτηριστικών με Μεθόδους Wrapper για Δεδομένα DESI 2022. Τα πράσινα κελιά υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Greedy Stepwise και τα γκρι κελιά τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Genetic Search. Τα

χαρακτηριστικά που επιλέχθηκαν κατ' επανάληψη ήταν τα 5G spectrum, SMEs with at least a basic level of digital intensity, Big data & Selling online cross-border.

7.3 Αποτελέσματα για τη συνιστώσα «Ανθρώπινο Κεφάλαιο»

7.3.1 Decision Tree (J48)

Ο αλγόριθμος με το δένδρο απόφασης (J48) εφαρμόστηκε στα δεδομένα της συνιστώσας «Ανθρώπινο κεφάλαιο», χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.909	0.714	0.800	HIGH	
0.333	0.333	0.333	MEDIUM	
0.538	0.538	0.609	LOW	
			Total	0.6667

Πίνακας 66. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48).

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 67) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 68). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 77. Ο Decision Tree (J48) εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.857	0.923	HIGH	
0.500	0.333	0.400	MEDIUM	
0.692	0.900	0.783	LOW	
			Total	0.8148

Πίνακας 67. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (ο αλγόριθμος Greedy Stepwise και ο αλγόριθμος Genetic Search έδωσαν όμοια αποτελέσματα).

7.3.2 Random Forest

Ο αλγόριθμος Random Forest εφαρμόστηκε στα δεδομένα της συνιστώσας «Ανθρώπινο Κεφάλαιο», χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.929	0.929	0.929	HIGH	
1.000	0.333	0.500	MEDIUM	
0.750	0.900	0.818	LOW	
			Total	0.8519

Πίνακας 68. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest.

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 69) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 69). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 77. Ο Random Forest εφαρμόστηκε ξανά με τα παραπάνω επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.929	0.963	HIGH	
0.667	0.667	0.667	MEDIUM	
0.818	0.900	0.857	LOW	
			Total	0.8889

Πίνακας 69. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (ο αλγόριθμος Greedy Stepwise και ο αλγόριθμος Genetic Search έδωσαν όμοια αποτελέσματα).

7.3.3 Adaboost

Ακολούθως, εφαρμόστηκε ο αλγόριθμος Adaboost, στα δεδομένα της συνιστώσας «Ανθρώπινο Κεφάλαιο», χωρίς περαιτέρω παραμετροποίηση, και οδήγησε στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.923	0.857	0.889	HIGH	
0.500	0.333	0.400	MEDIUM	
0.667	0.800	0.727	LOW	
			Total	0.7778

Πίνακας 70. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost.

Η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 71) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 71) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 77. Ο Adaboost εφαρμόστηκε ξανά με τα επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.929	0.929	0.929	HIGH	
0.500	0.333	0.400	MEDIUM	
0.727	0.800	0.762	LOW	
			Total	0.8148

Πίνακας 71. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών (ο αλγόριθμος Greedy Stepwise και ο αλγόριθμος Genetic Search έδωσαν όμοια αποτελέσματα).

7.3.4 Multilayer Perceptron

Ο αλγόριθμος Multilayer Perceptron εφαρμόστηκε στα δεδομένα της συνιστώσας του δείκτη DESI «Ανθρώπινο Κεφάλαιο» με την παραμετροποίηση “hidden layers=3”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.857	0.857	0.857	HIGH	
1.000	1.000	1.000	MEDIUM	
0.800	0.800	0.800	LOW	
			Total	0.8519

Πίνακας 72. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron.

Ακολούθησε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 73) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 73). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 77. Ο Multilayer Perceptron εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.929	0.929	0.929	HIGH	
1.000	0.667	0.800	MEDIUM	
0.818	0.900	0.857	LOW	
			Total	0.8889

Πίνακας 73. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
1.000	1.000	1.000	HIGH	
1.000	0.667	0.800	MEDIUM	
0.909	1.000	0.952	LOW	
			Total	0.9630

Πίνακας 74. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper και τον αλγόριθμο Genetic Search.

7.3.5 Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) εφαρμόστηκε στα δεδομένα της συνιστώσας «Ανθρώπινο κεφάλαιο» με την παραμετροποίηση “Kernel type=linear”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.867	0.929	0.897	HIGH	
?	0.000	?	MEDIUM	
0.667	0.800	0.727	LOW	
			Total	0.7778

Πίνακας 75. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine.

Τέλος, η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 76) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 76) κατέληξε στα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 77. Ο Support Vector Machine εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.933	1.000	0.966	HIGH	
?	0.000	?	MEDIUM	
0.750	0.900	0.818	LOW	
			Total	0.8519

Πίνακας 76. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών (όμοια για τους δύο αλγόριθμους Greedy Stepwise και Genetic search).

Variable	J48	RDF	ADA	MLP	SVM	J48	RDF	ADA	MLP	SVM
At least basic digital skills										
Above basic digital skills										
At least basic digital content creation skills										
ICT specialists										
Female ICT specialists										
Enterprises providing ICT training										
ICT graduates										

Πίνακας 77. Αποτελέσματα επιλογής χαρακτηριστικών για τη συνιστώσα «Ανθρώπινο Κεφάλαιο» του δείκτη DESI 2022 με μεθόδους Wrapper. Τα πράσινα κελιά υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Greedy Stepwise και τα γκρι κελιά τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Genetic Search. Τα χαρακτηριστικά που επιλέχθηκαν κατ' επανάληψη ήταν above basic digital skills και at least basic digital skills.

7.4 Αποτελέσματα για τη συνιστώσα «Συνδεσιμότητα»

7.4.1 Decision Tree (J48)

Ο αλγόριθμος με το δένδρο απόφασης Decision Tree (J48) εφαρμόστηκε στα δεδομένα της συνιστώσας «Συνδεσιμότητα», χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.941	1.000	0.970	HIGH	
1.000	0.600	0.750	MEDIUM	
0.857	1.000	0.923	LOW	
			Total	0.9259

Πίνακας 78. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48).

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 79) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 79). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 88. Ο Decision Tree (J48) εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.941	1.000	0.970	HIGH	
1.000	0.600	0.750	MEDIUM	
0.857	1.000	0.923	LOW	
			Total	0.9259

Πίνακας 79. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών (όμοια για τους δύο αλγόριθμους Greedy Stepwise και Genetic search).

7.4.2 Random Forest

Ο αλγόριθμος Random Forest εφαρμόστηκε στα δεδομένα της συνιστώσας «Συνδεσιμότητα», χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.842	1.000	0.914	HIGH	
0.500	0.200	0.286	MEDIUM	
0.833	0.833	0.833	LOW	
			Total	0.8148

Πίνακας 80. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest.

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 81) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 81). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακας 88. Ο Random Forest εφαρμόστηκε ξανά με τα παραπάνω επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.889	1.000	0.941	HIGH	
0.750	0.600	0.667	MEDIUM	
1.000	0.833	0.909	LOW	
			Total	0.8889

Πίνακας 81. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών (όμοια για τους δύο αλγόριθμους Greedy Stepwise και Genetic search).

7.4.3 Adaboost

Εν συνεχεία εφαρμόστηκε ο αλγόριθμος Adaboost, στο σύνολο των δεδομένων της συνιστώσας «Συνδεσιμότητα» και οδήγησε στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.875	0.875	0.875	HIGH	
0.500	0.600	0.545	MEDIUM	
1.000	0.833	0.909	LOW	
			Total	0.8148

Πίνακας 82. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost.

Η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 83) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 83) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 88. Ο Adaboost εφαρμόστηκε ξανά με τα επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.889	0.889	1.000	HIGH	
0.750	0.750	0.600	MEDIUM	
1.000	1.000	0.833	LOW	
			Total	0.8889

Πίνακας 83. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (οι μετρικές ήταν οι παραπάνω και για τους δύο αλγόριθμους Greedy Stepwise και Genetic Search).

7.4.4 Multilayer Perceptron

Ο αλγόριθμος Multilayer Perceptron εφαρμόστηκε στα δεδομένα της συνιστώσας του δείκτη DESI «Συνδεσιμότητα» με την παραμετροποίηση “hidden layers=3”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
1.000	0.938	0.968	HIGH	
1.000	1.000	1.000	MEDIUM	
0.857	1.000	0.923	LOW	
			Total	0.9630

Πίνακας 84. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron.

Ακολούθησε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 85) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 85). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 88. Ο Multilayer Perceptron εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	1.000	1.000	HIGH	
1.000	1.000	1.000	MEDIUM	
1.000	1.000	1.000	LOW	
			Total	1.0000

Πίνακας 85. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με επιλογή χαρακτηριστικών (όμοια για τους δύο αλγόριθμους Greedy Stepwise και Genetic search).

7.4.5 Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) εφαρμόστηκε στα δεδομένα της συνιστώσας «Συνδεσιμότητα» με την παραμετροποίηση “Kernel type=linear”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.842	1.000	0.914	HIGH	
1.000	0.400	0.571	MEDIUM	
1.000	1.000	1.000	LOW	
			Total	0.8889

Πίνακας 86. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine.

Τέλος, η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο GreedyStepwise (Πίνακας 87) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 87) κατέληξε στα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 88. Ο Support Vector Machine εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.842	1.000	1.000	HIGH	
1.000	0.857	0.923	MEDIUM	
1.000	1.000	0.952	LOW	
			Total	0.9630

Πίνακας 87. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (όμοια για τους δύο αλγόριθμους Greedy Stepwise και Genetic search).

Variable	J48	RDF	ADA	MLP	SVM	J48	RDF	ADA	MLP	SVM
Overall fixed broadband take-up										
At least 100 Mbps fixed broadband take-up										
At least 1 Gbps broadband take-up										
Fast broadband (NGA) coverage										
Fixed Very High Capacity Network (VHCN) coverage										
5G spectrum										
Overall 5G coverage										
Mobile broadband take-up										
Broadband price index										

Πίνακας 88. Αποτελέσματα επιλογής χαρακτηριστικών με μεθόδους Wrapper για τη συνιστώσα «Συνδεσιμότητα» του δείκτη DESI 2022. Τα πράσινα κελιά υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Greedy Stepwise και τα γκρι κελιά τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Genetic Search. Τα χαρακτηριστικά που επιλέχθηκαν κατ' επανάληψη ήταν τα overall fixed broadband take-up, at least 100 Mbps fixed broadband take-up, 5G spectrum και fixed very high capacity network coverage.

7.5 Αποτελέσματα για τη συνιστώσα «Ενσωμάτωση Ψηφιακής Τεχνολογίας»

7.5.1 Decision Tree (J48)

Ο αλγόριθμος με το δένδρο απόφασης Decision Tree (J48) εφαρμόστηκε στα δεδομένα της συνιστώσας «Ενσωμάτωση Ψηφιακής Τεχνολογίας», χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.333	0.222	0.267	HIGH	
0.875	1.000	0.933	MEDIUM	
0.462	0.545	0.500	LOW	
			Total	0.5556

Πίνακας 89. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48).

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 90) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 90). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό πίνακα Πίνακας 102. Ο Decision Tree (J48) εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.857	0.667	0.750	HIGH	
0.778	1.000	0.875	MEDIUM	
0.818	0.818	0.818	LOW	
			Total	0.8148

Πίνακας 90. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper και τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.778	0.778	0.778	HIGH	
1.000	1.000	1.000	MEDIUM	
0.818	0.818	0.818	LOW	
			Total	0.8519

Πίνακας 91. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper και τον αλγόριθμο Genetic Search.

7.5.2 Random Forest

Ο αλγόριθμος Random Forest εφαρμόστηκε στα δεδομένα της Ενσωμάτωσης της Ψηφιακής Τεχνολογίας, χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.818	1.000	0.900	HIGH	
1.000	1.000	1.000	MEDIUM	
1.000	0.818	0.900	LOW	
			Total	0.9259

Πίνακας 92. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest.

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 93) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 94). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 102. Ο Random Forest εφαρμόστηκε ξανά με τα παραπάνω επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.818	1.000	0.900	HIGH	
1.000	1.000	1.000	MEDIUM	
1.000	0.818	0.900	LOW	
			Total	0.8889

Πίνακας 93. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
1.000	0.889	0.941	HIGH	
0.875	1.000	0.933	MEDIUM	
0.909	0.909	0.909	LOW	
			Total	0.9259

Πίνακας 94. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

7.5.3 Adaboost

Εν συνεχεία εφαρμόστηκε ο αλγόριθμος Adaboost, στο σύνολο των δεδομένων της συνιστώσας «Ενσωμάτωση Ψηφιακής Τεχνολογίας» και οδήγησε στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
0.667	0.889	0.762	HIGH	
0.857	0.857	0.857	MEDIUM	
0.875	0.636	0.737	LOW	
			Total	0.7778

Πίνακας 95. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost.

Η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 96) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 97) κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 102. Ο Adaboost εφαρμόστηκε ξανά με τα επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.818	1.000	0.900	HIGH	
0.857	0.857	0.857	MEDIUM	
0.889	0.727	0.800	LOW	
			Total	0.8519

Πίνακας 96. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Greedy Stepwise.

Precision	Recall	F1-score	Class	Accuracy
0.900	1.000	0.947	HIGH	
1.000	1.000	1.000	MEDIUM	
1.000	0.909	0.952	LOW	
			Total	0.9630

Πίνακας 97. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper σε συνδυασμό με τον αλγόριθμο Genetic Search.

7.5.4 Multilayer Perceptron

Ο αλγόριθμος Multilayer Perceptron εφαρμόστηκε στα δεδομένα της Ενσωμάτωσης Ψηφιακής Τεχνολογίας με την παραμετροποίηση “hidden layers=3”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.889	0.889	0.889	HIGH	
1.000	1.000	1.000	MEDIUM	
0.909	0.909	0.909	LOW	
			Total	0.9259

Πίνακας 98. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron.

Ακολούθησε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 99) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 99). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 102. Ο Multilayer Perceptron εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	1.000	1.000	HIGH	
1.000	1.000	1.000	MEDIUM	
1.000	1.000	1.000	LOW	
			Total	1.0000

Πίνακας 99. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (όμοια για τους δύο αλγόριθμους Greedy Stepwise και Genetic search) .

7.5.5 Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) εφαρμόστηκε στα δεδομένα της συνιστώσας «Ενσωμάτωση Ψηφιακής Τεχνολογίας» με την παραμετροποίηση “Kernel type=linear”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.875	0.778	0.824	HIGH	
1.000	1.000	1.000	MEDIUM	
0.833	0.909	0.870	LOW	
			Total	0.8889

Πίνακας 100. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine.

Τέλος, η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 101) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 101) κατέληξε στα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 102. Ο Support Vector Machine εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.889	0.941	HIGH	
1.000	1.000	1.000	MEDIUM	
0.917	1.000	0.957	LOW	
			Total	0.9630

Πίνακας 101. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών.

Variable	J48	RDF	ADA	MLP	SVM	J48	RDF	ADA	MLP	SVM
SMEs with at least a basic level of digital intensity										
Electronic information sharing										
Social media										
Big data										
Cloud										
Artificial intelligence										
ICT for environmental sustainability										
e-Invoices										
SMEs selling online										
e-Commerce turnover										
Selling online cross-border										

Πίνακας 102. Αποτελέσματα Επιλογής Χαρακτηριστικών με Μεθόδους Wrapper για τη συνιστώσα «Ενσωμάτωση Ψηφιακής Τεχνολογίας» του δείκτη DESI 2022. Τα πράσινα κελιά υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τους αλγόριθμους Greedy Stepwise, ενώ τα γκρι υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Genetic Search. Σημαντικότερα χαρακτηριστικά αναδείχθηκαν τα Big data, Cloud, Artificial intelligence, SMEs selling online, e-Commerce turnover, Selling online cross-border.

7.6 Αποτελέσματα για τη συνιστώσα «Ψηφιακές Δημόσιες Υπηρεσίες»

7.6.1 Decision Tree (J48)

Ο αλγόριθμος με το δένδρο απόφασης Decision Tree (J48) εφαρμόστηκε στα δεδομένα που αφορούν τις Ψηφιακές Δημόσιες Υπηρεσίες χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
1.000	0.727	0.842	HIGH	
0.714	1.000	0.833	MEDIUM	
1.000	0.833	0.909	LOW	
			Total	0.8519

Πίνακας 103. Αξιολόγηση απόδοσης του αλγόριθμου Decision Tree (J48).

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 104) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 104). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 113.

Ο Decision Tree (J48) εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.818	0.900	HIGH	
0.769	1.000	0.870	MEDIUM	
1.000	0.833	0.909	LOW	
			Total	0.8889

Πίνακας 104. Αξιολόγηση απόδοσης του αλγορίθμου Decision Tree (J48) με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (όμοια για τους δύο αλγόριθμους Greedy Stepwise και Genetic search).

7.6.2 Random Forest

Ο αλγόριθμος Random Forest εφαρμόστηκε στα δεδομένα των Ψηφιακών Δημόσιων Υπηρεσιών, χωρίς περαιτέρω παραμετροποίηση, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
1.000	0.909	0.952	HIGH	
0.846	1.000	0.909	MEDIUM	
1.000	0.833	0.909	LOW	
			Total	0.9259

Πίνακας 105. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest.

Εν συνεχεία, πραγματοποιήθηκε η επιλογή χαρακτηριστικών τύπου Wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 106) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 106). Τα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 113. Ο Random Forest εφαρμόστηκε ξανά με τα παραπάνω επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
0.909	0.909	0.909	HIGH	
0.833	1.000	0.909	MEDIUM	
1.000	0.667	0.800	LOW	
			Total	0.8889

Πίνακας 106. Αξιολόγηση απόδοσης του αλγορίθμου Random Forest με επιλογή χαρακτηριστικών (όμοια και για τις δύο ομάδες χαρακτηριστικών από Greedy Stepwise και Genetic Search).

7.6.3 Adaboost

Εν συνεχεία εφαρμόστηκε ο αλγόριθμος Adaboost, στα δεδομένα των Ψηφιακών Δημόσιων Υπηρεσιών και οδήγησε στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.636	0.778	HIGH	
0.769	1.000	0.870	MEDIUM	
0.095	0.833	0.769	LOW	
			Total	0.8148

Πίνακας 107. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost.

Η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο GreedyStepwise (Πίνακας 108) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 108), κατέληξε στα επιλεγμένα χαρακτηριστικά που παρουσιάζονται στον περιληπτικό Πίνακα 113. Ο Adaboost εφαρμόστηκε ξανά με τα επιλεγμένα χαρακτηριστικά, και καταλήξαμε στις παρακάτω μετρικές:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.727	0.842	HIGH	
0.769	1.000	0.870	MEDIUM	
0.833	0.833	0.909	LOW	
			Total	0.8519

Πίνακας 108. Αξιολόγηση απόδοσης του αλγορίθμου Adaboost με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (όμοια για τα χαρακτηριστικά που προέκυψαν από Greedy Stepwise και Genetic search).

7.6.4 Multilayer Perceptron

Ο αλγόριθμος Multilayer Perceptron εφαρμόστηκε στα δεδομένα των Ψηφιακών Δημόσιων Υπηρεσιών με την παραμετροποίηση “hidden layers=3”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγορίθμου.

Precision	Recall	F1-score	Class	Accuracy
0.900	0.818	0.857	HIGH	
0.769	1.000	0.870	MEDIUM	
1.000	0.667	0.800	LOW	
			Total	0.8519

Πίνακας 109. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron.

Ακολούθησε η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 110) και έπειτα με τον αλγόριθμο Genetic Search (Πίνακας 110). Τα

επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 113. Ο Multilayer Perceptron εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.818	0.900	HIGH	
0.769	1.000	0.870	MEDIUM	
1.000	0.833	0.909	LOW	
			Total	0.8889

Πίνακας 110. Αξιολόγηση απόδοσης του αλγορίθμου Multilayer Perceptron με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (όμοια για τα χαρακτηριστικά που προέκυψαν από Greedy Stepwise και Genetic search).

7.6.5 Support Vector Machine

Ο αλγόριθμος Support Vector Machine (SVM) εφαρμόστηκε στα δεδομένα με την παραμετροποίηση “Kernel type=linear”, και αξιολογήθηκε με τη χρήση 10-fold cross-validation σε ολόκληρο το υπάρχον σύνολο δεδομένων. Τα αποτελέσματα αποτυπώνονται στον παρακάτω σχετικό πίνακα με τις επιδόσεις του αλγόριθμου.

Precision	Recall	F1-score	Class	Accuracy
0.900	0.818	0.857	HIGH	
0.833	1.000	0.909	MEDIUM	
0.800	0.667	0.727	LOW	
			Total	0.8519

Πίνακας 111. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine.

Τέλος, η επιλογή χαρακτηριστικών τύπου wrapper, πρώτα με τον Αλγόριθμο Greedy Stepwise (Πίνακας 112) και έπειτα με τον αλγόριθμο Genetic search (Πίνακας 112) κατέληξε στα επιλεγμένα χαρακτηριστικά παρουσιάζονται στον περιληπτικό Πίνακα 113. Ο Support Vector Machine εφαρμόστηκε ξανά στα επιλεγμένα χαρακτηριστικά, καταλήγοντας στα παρακάτω αποτελέσματα:

Precision	Recall	F1-score	Class	Accuracy
1.000	0.818	0.900	HIGH	
0.833	1.000	0.909	MEDIUM	
0.833	0.833	0.833	LOW	
			Total	0.8889

Πίνακας 112. Αξιολόγηση απόδοσης του αλγορίθμου Support Vector Machine με επιλογή χαρακτηριστικών με τη μέθοδο Wrapper (όμοια για τα χαρακτηριστικά που προέκυψαν από Greedy Stepwise και Genetic search).

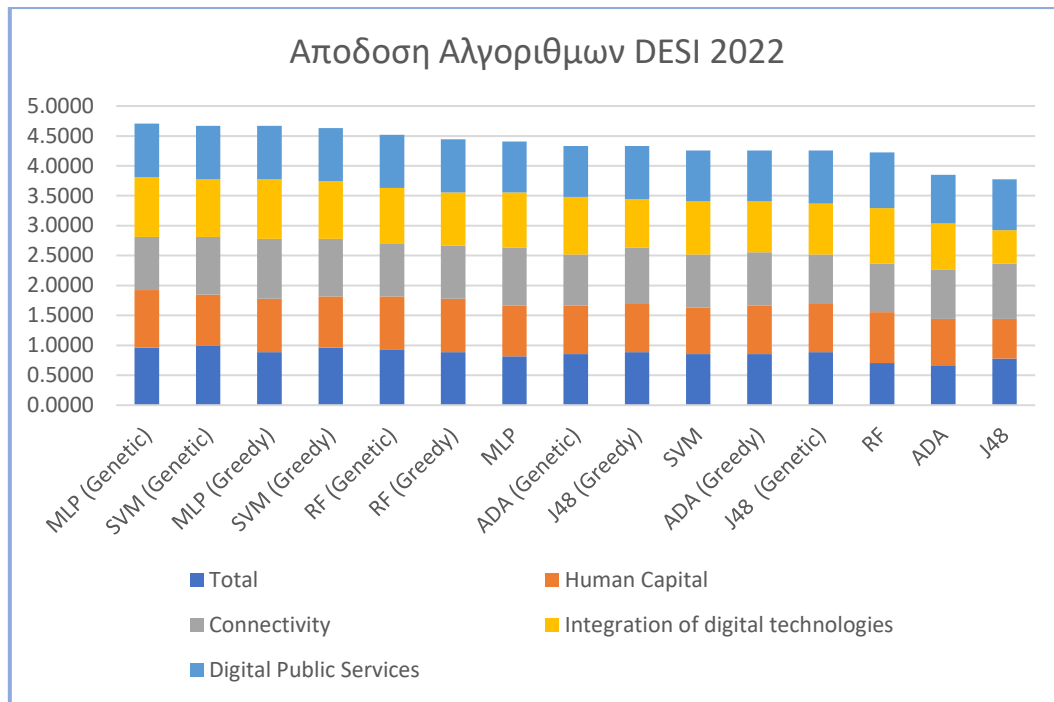
Variable	J48	RDF	ADA	MLP	SVM	J48	RDF	ADA	MLP	SVM
e-Government users										
Pre-filled Forms										
Digital public services for citizens										
Digital public services for businesses										
Open data										

Πίνακας 113. Αποτελέσματα Επιλογής Χαρακτηριστικών με Μεθόδους Wrapper για τη συνιστώσα «Ψηφιακές Δημόσιες Υπηρεσίες» του δείκτη DESI 2022. Τα πράσινα κελιά υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τους αλγόριθμους Greedy Stepwise, ενώ τα γκρι υποδεικνύουν τα χαρακτηριστικά που επιλέχθηκαν από τον αλγόριθμο Genetic Search. Τα χαρακτηριστικά που επέλεξαν οι αλγόριθμοι είναι τα e-Government users, Pre-filled Forms, Digital public services for businesses, Open data.

7.7 Παρατηρήσεις

Η εφαρμογή τεχνικών επιβλεπόμενης μάθησης στο σύνολο των δεδομένων του δείκτη DESI για το έτος 2022 ανέδειξε ουσιαστικές παρατηρήσεις οι οποίες σχετίζονται με την προγνωστική ικανότητα των χαρακτηριστικών του δείκτη να κατηγοριοποιεί τα κράτη-μέλη της Ε.Ε. βάσει της ψηφιακής τους προόδου. Η ομαδοποίηση που προέκυψε από τη ανάλυση, με τεχνικές μη επιβλεπόμενης μάθησης, έδωσε ετικέτες κλάσης και επέτρεψε την αντίστροφη διερεύνηση των παραγόντων που διαφοροποιούν τις ομάδες χωρών, μέσω της επιβλεπόμενης μάθησης αναδεικνύοντας τα πιο σημαντικά χαρακτηριστικά.

Αρχικά, οι αλγόριθμοι που χρησιμοποιήθηκαν παρουσίασαν ικανοποιητικές έως εξαιρετικές επιδόσεις, ιδίως όταν είχε προηγηθεί επιλογή χαρακτηριστικών. Ειδικά οι Random Forest, Adaboost και Multilayer Perceptron, σημείωσαν υψηλή ακρίβεια (Accuracy), υπερβαίνοντας το 90% σε ορισμένα σενάρια, γεγονός που καταδεικνύει και επιβεβαιώνει τη σαφή διαφοροποίηση των ομάδων που προέκυψαν από την ανάλυση clustering.



Γράφημα 7. Συνολική απόδοση αλγορίθμων μηχανικής μάθησης στην ταξινόμηση των κρατών-μελών της Ε.Ε. βάσει του δείκτη DESI 2022. Οι αλγόριθμοι εφαρμόστηκαν με διαφορετικές τεχνικές επιλογής χαρακτηριστικών (Genetic Search και Greedy Stepwise), και η απόδοσή τους εκφράζεται με άθροισμα επιμέρους συνιστωσών: Human Capital, Connectivity, Integration of Digital Technologies και Digital Public Services. Η κατηγορία “Total” αναφέρεται στην ακρίβεια για το σύνολο των χαρακτηριστικών του δείκτη DESI 2022.

Η επιλογή χαρακτηριστικών ανέδειξε επαναλαμβανόμενες μεταβλητές με υψηλή σημασία. Ειδικότερα στο σύνολο των δεδομένων του δείκτη DESI 2022, σημαντικότερα χαρακτηριστικά αναδείχτηκαν η ανάλυση μεγάλων δεδομένων (Big data), το φάσμα 5G (5G spectrum, οι MME με τουλάχιστον βασικό επίπεδο ψηφιακής έντασης (SMES with at least a basic level of digital intensity) και διασυνοριακές online πωλήσεις (Selling online cross-border). Η ανάλυση της συνιστώσας «Ανθρώπινο Κεφάλαιο», ανέδειξε ως σημαντικότερα χαρακτηριστικά το βασικό και άνω του βασικού επίπεδο ψηφιακών δεξιοτήτων (at least basic digital skills, above basic digital skills). Η ανάλυση της συνιστώσας «Συνδεσιμότητα», εστίασε στα χαρακτηριστικά: συνδρομές νοικοκυριών με σύνδεση >100 Mbps (100 Mbps fixed broadband take-up), κάλυψη 5G (5G spectrum), χρήση σταθερής ευρυζωνικότητας (fixed broadband take-up) και κάλυψη δικτύων πολύ υψηλής ταχύτητας (fixed very high capacity network coverage). Επιπρόσθετα, η συνιστώσα «Ενσωμάτωση Ψηφιακής Τεχνολογίας» ανέδειξε τα χαρακτηριστικά: υπηρεσίες Cloud, τεχνητή νοημοσύνη (artificial intelligence), ανάλυση μεγάλων δεδομένων (Big data), MME που πωλούν online (SMEs selling online), πωλήσεις ηλεκτρονικού εμπορίου (e-Commerce turnover) και διασυνοριακές online πωλήσεις (Selling online cross-border). Τέλος, η ανάλυση της συνιστώσας «Ψηφιακές Δημόσιες

Υπηρεσίες» έδειξε πως τα περισσότερα χαρακτηριστικά αυτής της συνιστώσας είναι σημαντικά για την ταξινόμηση, και ιδιαίτερα τα ανοιχτά δεδομένα (Open data) και οι ψηφιακές δημόσιες υπηρεσίες επιχειρήσεις (Digital public services for businesses), οι χρήστες ψηφιακών δημοσίων υπηρεσιών (e-Government users) και οι προσυμπληρωμένες φόρμες (Pre-filled Forms).

Η παραπάνω αποτίμηση των δεδομένων αποδεικνύει ότι η ποιότητα και η ένταση της ψηφιακής δραστηριότητας, τόσο στον δημόσιο όσο και στον ιδιωτικό τομέα, συνιστούν καθοριστικούς παράγοντες διαφοροποίησης μεταξύ των χωρών.

Επιπλέον, η δυνατότητα των αλγορίθμων να προβλέψουν με ακρίβεια τις ομάδες που προέκυψαν από την συσταδοποίηση του Κεφαλαίου 6 προσδίδει αξιοπιστία και εγκυρότητα στα αποτελέσματα της μη επιβλεπόμενης μάθησης, εφόσον αυτές μπορούν να αναπαραχθούν με επιβλεπόμενες μεθόδους. Η συνέργεια των δύο προσεγγίσεων ενισχύει τη μεθοδολογική αρτιότητα της μελέτης και επιτρέπει την ερμηνεία των αποτελεσμάτων σε μία σταθερή βάση. Συνεπώς, οι προσεγγίσεις αυτές, μπορούν να λειτουργήσουν ως εργαλεία υποστήριξης πολιτικής, εντοπίζοντας τους τομείς όπου απαιτούνται παρεμβάσεις για την ψηφιακή σύγκλιση.

8. Συμπεράσματα

Η παρούσα εργασία είχε ως βασικό στόχο τη μελέτη και την κατηγοριοποίηση της ψηφιακής προόδου των κρατών-μελών της Ευρωπαϊκής Ένωσης, με χρήση των δεδομένων του Δείκτη Ψηφιακής Οικονομίας και Κοινωνίας (DESI) για το έτος 2022 όπως αυτά δημοσιεύθηκαν από την Ευρωπαϊκή Επιτροπή. Μέσα από την αξιοποίηση τεχνικών επιβλεπόμενης και μη επιβλεπόμενης μηχανικής μάθησης στο περιβάλλον του λογισμικού Weka, επιχειρήθηκε όχι μόνο η χαρτογράφηση του ψηφιακού τοπίου της Ε.Ε. αλλά και η σύγκριση των ευρημάτων με αποτελέσματα προηγούμενων ερευνών. Η ανάλυση αυτή ανέδειξε κρίσιμα συμπεράσματα που αφορούν τόσο τη σημερινή κατάσταση όσο και τις προοπτικές της ψηφιακής ανάπτυξης στην Ευρώπη και τη συσχέτιση της ψηφιακής προόδου με άλλους αναπτυξιακούς δείκτες.

Οι τεχνικές συσταδοποίησης (clustering), με κύρια μέθοδο τον αλγόριθμο K-means, κατέδειξαν την ύπαρξη διακριτών ομάδων κρατών με κοινά χαρακτηριστικά όσον αφορά στην ψηφιακή πρόοδο, ενισχύοντας τη διαπίστωση ότι οι ανισότητες εντός της Ε.Ε. παραμένουν, αν και μειούμενες.

- **Ομάδα 1 (Υψηλό Επίπεδο):** Η ομάδα περιλαμβάνει Χώρες της Βόρειας και Δυτικής Ευρώπης (Φινλανδία, Σουηδία, Δανία, Ολλανδία, Ιρλανδία, Λουξεμβούργο) που διακρίνονται για τις υψηλές επενδύσεις τους σε όλες τις συνιστώσες του δείκτη DESI.
- **Ομάδα 2 (Μέσο Επίπεδο):** Η ομάδα περιλαμβάνει κατά κύριο λόγο χώρες της Κεντρικής και Νότιας Ευρώπης (Γερμανία, Γαλλία, Τσεχία, Ισπανία, Πορτογαλία) με επιδόσεις που ποικίλουν στις διάφορες συνιστώσες. Παρά τις θετικές προοπτικές εξέλιξης, αντιμετωπίζουν ανισότητες, είτε λόγω θεσμικών ελλείψεων (π.χ. στην εκπαίδευση ΤΠΕ), είτε λόγω γεωγραφικών προκλήσεων (π.χ. μεγάλες αγροτικές ή απομακρυσμένες περιοχές με περιορισμένη ευρυζωνική κάλυψη).
- **Ομάδα 3 (Χαμηλό Επίπεδο):** Η ομάδα που επιδεικνύει τις χαμηλότερες επιδόσεις περιλαμβάνει χώρες της Ανατολικής και Νοτιοανατολικής Ευρώπης (Βουλγαρία, Ρουμανία, Ελλάδα, Πολωνία) που υστερούν σε όλους τους βασικούς τομείς, όπως οι ψηφιακές δεξιότητες του πληθυσμού, η υιοθέτηση των νέων τεχνολογιών από επιχειρήσεις, η ποιότητα των ψηφιακών δημόσιων υπηρεσιών και οι συνδεσιμότητες.

Το αποτέλεσμα αυτό ενίσχυσε την υπόθεση ότι τόσο η γεωγραφική θέση όσο και το οικονομικό επίπεδο, εξακολουθούν να διαδραματίζουν καθοριστικό ρόλο στην ψηφιακή σύγκλιση της Ευρώπης.

Ο συνδυασμός τεχνικών συσταδοποίησης με τη χρήση ταξινομητών επιβλεπόμενης μάθησης (J48, Random Forest, Adaboost, Multilayer Perceptron, Support Vector Machine), επιβεβαίωσε την εγκυρότητα των ομαδοποιήσεων και αυτή η σύγκλιση των αποτελεσμάτων ενισχύει την αξιοπιστία της μεθόδου. Η ταξινόμηση ανέδειξε τους αλγόριθμους Random Forest, Multilayer Perceptron και Support Vector Machine ως τους πιο ακριβείς, ιδιαίτερα μετά τη διαδικασία επιλογής χαρακτηριστικών με τη μέθοδο Wrapper, σε συνδυασμό με τους αλγόριθμους (Greedy Stepwise ή Genetic Search). Η μέθοδος αυτή, βελτίωσε αισθητά την αποδοτικότητα των μοντέλων και ανέδειξε κρίσιμα χαρακτηριστικά για κάθε συνιστώσα του DESI. Μέσω της επιλογής χαρακτηριστικών και της ανάλυσης των μοντέλων μηχανικής μάθησης, αναδείχθηκαν οι εξής καθοριστικοί παράγοντες: Για τη συνιστώσα «Ανθρώπινο Κεφάλαιο» σημαντικές είναι οι βασικές και προχωρημένες ψηφιακές δεξιότητες αλλά και ο αριθμός ειδικών ΤΠΕ. Όσον αφορά στη συνιστώσα «Συνδεσιμότητα» οι μεταβλητές Ευρυζωνική κάλυψη (100 Mbps και 5G) και χρήση σταθερής και κινητής ευρυζωνικής σύνδεσης ξεχώρισαν ανάμεσα στα εν λόγω χαρακτηριστικά. Η χρήση cloud, τεχνητής νοημοσύνης και big data καθώς και το διασυννοριακό ηλεκτρονικό εμπόριο είναι οι μεταβλητές που φέρονται ως οι πιο κρίσιμες για τη συνιστώσα «Ενσωμάτωση Ψηφιακών Υπηρεσιών» και τέλος η ποιότητα e-government (π.χ. προπληρωμένες φόρμες, ανοιχτά δεδομένα) και η χρήση από πολίτες και επιχειρήσεις ήταν οι μεταβλητές της συνιστώσας «Ψηφιακές Δημόσιες Υπηρεσίες» που βελτίωσαν την ακρίβεια των αλγορίθμων. Αυτοί οι παράγοντες μπορούν να χρησιμεύσουν ως στρατηγικοί πυλώνες για εθνικές και ευρωπαϊκές πολιτικές ψηφιακής μετάβασης.

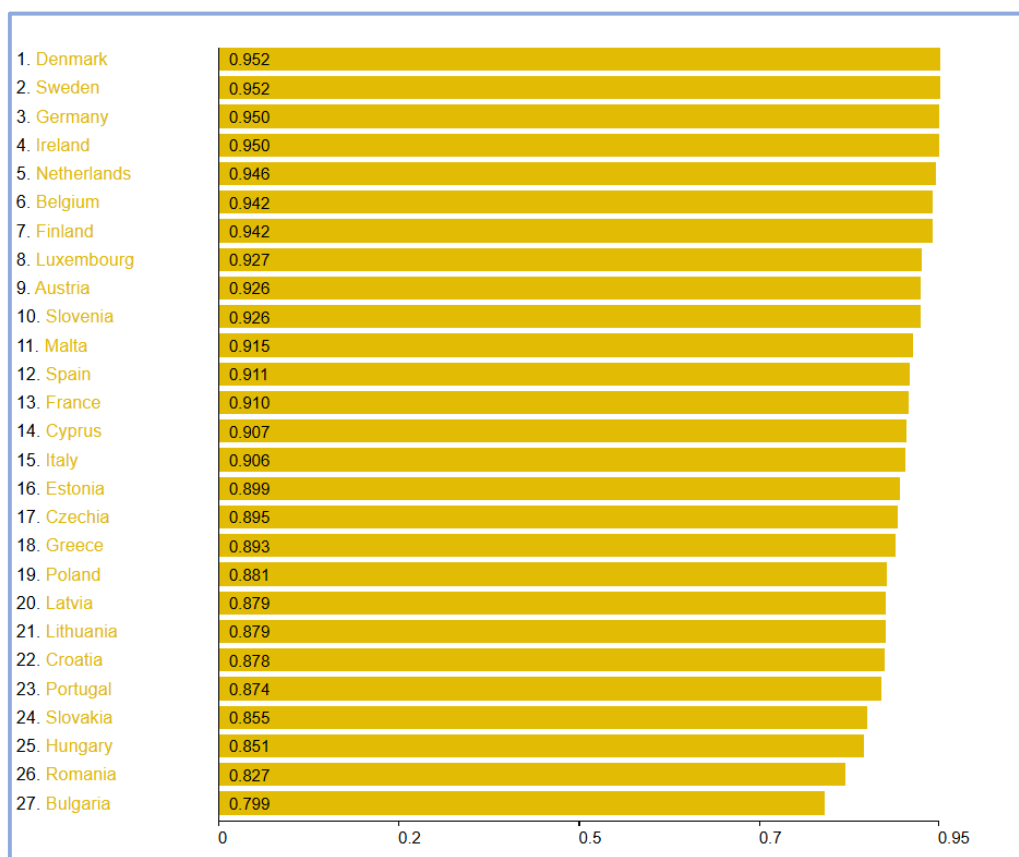
Επιπλέον, η ανάλυση των επιμέρους συνιστωσών του δείκτη DESI ανέδειξε ενδιαφέρουσες διαφοροποιήσεις. Για παράδειγμα, χώρες όπως η Ισπανία, που συνολικά κατατάσσονται στο μεσαίο επίπεδο, σημειώνουν πολύ υψηλές επιδόσεις σε επιμέρους τομείς, όπως οι ψηφιακές δημόσιες υπηρεσίες. Αντίστοιχα, η Ελλάδα, ενώ καταγράφει βελτίωση σε τομείς όπως η συνδεσιμότητα, εξακολουθεί να υπολείπεται στον τομέα του ανθρώπινου κεφαλαίου.

Η μελέτη διερεύνησε επίσης, τη σχέση του DESI με τρεις διαφορετικές ετικέτες κλάσης των χωρών, που αφορούσαν την ψηφιακή πρόοδο και προέκυψαν από τη μελέτη Bánhidí & Dobos (2023), τη βιώσιμη ανάπτυξη (Martín & Carnero, 2019) και την οικονομική ανάπτυξη και

ανταγωνιστικότητα (Onuferoná et al., 2020). Τα αποτελέσματα ήταν διαφορετικά ανάλογα με τον τομέα:

- **Ψηφιακή Πρόοδος:** Οι δείκτης DESI είχε υψηλή προγνωστική ικανότητα (ακρίβεια έως 92,59%).
- **Βιώσιμη Ανάπτυξη:** Η συσχέτιση της ψηφιακής προόδου με την βιώσιμη ανάπτυξη ήταν πολύ καλή (ακρίβεια ~81,48%), όμως η πρόβλεψη της βιωσιμότητας απαιτεί ευρύτερους δείκτες (κοινωνικούς, περιβαλλοντικούς) που δεν καλύπτονται πλήρως από το DESI.
- **Οικονομική Ανάπτυξη και Ανταγωνιστικότητα:** Αναδεικνύεται η σημαντική σύνδεση ψηφιακής προόδου και οικονομικής ανάπτυξης και ανταγωνιστικότητας καθώς παρατηρήθηκε ισχυρή συσχέτιση των δεικτών (ακρίβεια έως 96,30%).

Αυτά τα ευρήματα υποστηρίζουν ότι η ψηφιακή μετάβαση είναι ένας κρίσιμος πυλώνας της οικονομικής ανάπτυξης, αλλά η σύνδεσή της με τη βιωσιμότητα απαιτεί ευρύτερη ανάλυση. Η πολυπλοκότητα κοινωνικών, οικονομικών και περιβαλλοντικών παραμέτρων απαιτεί τη συνδυαστική χρήση περισσότερων δεδομένων και δεικτών, προκειμένου να επιτευχθεί μία πλήρης εικόνα. Σε κάθε περίπτωση, η θετική συσχέτιση με τους παραπάνω δείκτες καταδεικνύει πως η ψηφιακή μετάβαση αποτελεί αναπόσπαστο μέρος της σύγχρονης ανάπτυξης. (Αντίστοιχα συμπεράσματα μπορούμε να εξάγουμε αν παρατηρήσουμε στο Γράφημα 8 την κατάταξη των χωρών βάσει του δείκτη Ανθρώπινης Ανάπτυξης).



Γράφημα 8. Κατάταξη των χωρών κρατών μελών της Ε.Ε. βάσει του Δείκτη Ανθρώπινης Ανάπτυξης (“Human Development in the European Union,” 2023)

Συνοψίζοντας, η εργασία αυτή επιβεβαιώνει την ύπαρξη ανισοτήτων στον ψηφιακό χάρτη της Ε.Ε. και την ανάγκη ευρύτερων πολιτικών για την επίτευξη ουσιαστικής σύγκλισης. Ειδικότερα, κρίνεται πως υπάρχει ανάγκη δημιουργίας εκπαιδευτικών προγραμμάτων για τον πληθυσμό, ενίσχυση των υποδομών συνδεσιμότητας με επέκταση της ευρυζωνικής κάλυψης σε αγροτικές και απομακρυσμένες περιοχές και ανάπτυξη των δικτύων 5G, υποστήριξη και παροχή κινήτρων για την ψηφιακή μετάβαση των μικρομεσαίων επιχειρήσεων, απλοποίηση και ανάπτυξη των ψηφιακών υπηρεσιών του δημοσίου και προώθηση της διαφάνειας (ανοιχτά δεδομένα). Επιπρόσθετα, η χρήση μεθόδων μηχανικής μάθησης προσφέρει αξιόπιστα εργαλεία για την κατηγοριοποίηση και πρόβλεψη της ψηφιακής εξέλιξης των κρατών.

Οι ανισότητες που καταγράφηκαν αντανakλούν δομικά προβλήματα (οικονομικά, θεσμικά), αλλά και ευκαιρίες για στρατηγικές συνεργασίες εντός της ΕΕ, ιδίως μέσω κοινοτικών προγραμμάτων (π.χ. Digital Europe).

Η ερευνητική αυτή προσπάθεια ανέδειξε τη χρησιμότητα του δείκτη DESI ως εργαλείο ανάλυσης και χάραξης πολιτικής, ανοίγοντας ταυτόχρονα τον δρόμο για μελλοντικές

εφαρμογές και συσχετίσεις σε ακόμη περισσότερα πεδία της κοινωνικής και οικονομικής ζωής. Επιπλέον, η χρήση ετικετών κλάσης από προηγούμενες μελέτες και η προσπάθεια πρόβλεψής τους μέσω των δεδομένων του DESI 2022 αποτέλεσε μια καινοτόμο προσέγγιση, ενισχύοντας τη συζήτηση γύρω από την αξιοπιστία και τη δυνατότητα χρήσης του δείκτη ως εργαλείου ανάλυσης.

9. Περιορισμοί της μελέτης και προτάσεις για περαιτέρω έρευνα

Παρότι η ανάλυση παρείχε χρήσιμα συμπεράσματα, είναι απαραίτητο να επισημανθούν ορισμένοι περιορισμοί με στόχο να διερευνηθεί το πεδίο για περαιτέρω έρευνα.

Αρχικά, η χρήση δεδομένων ενός μόνο έτους (2022) δεν επιτρέπει την ανάλυση των δυναμικών εξελίξεων σε βάθος χρόνου π.χ. τη διαχρονικότητα των ευρημάτων ή την ανάλυση τάσεων στη διάρκεια του χρόνου. Επίσης είναι σημαντικό να αναφερθεί ότι η ανάλυση έχει πραγματοποιηθεί σε επίπεδο κράτους μέλους, χωρίς να εξετάζονται οι περιφερειακές ανισότητες, οι οποίες είναι ενδεχομένως σημαντικές. Επιπλέον, το σύνολο δεδομένων αποτελείται από πολλά χαρακτηριστικά για κάθε χώρα και ως εκ τούτου προκύπτει το πρόβλημα της «κατάρας της διαστατικότητας» και ο κίνδυνος υπερπροσαρμογής των μοντέλων. Τέλος, ο προκαθορισμένος αριθμός ομάδων «clusters» για λόγους συγκρισιμότητας με τις επιλεγμένες μελέτες, εισάγει ένα στοιχείο υποκειμενικότητας το οποίο μπορεί να επηρεάσει τη σύσταση αλλά και τη σταθερότητα των ομάδων.

Παρά τους περιορισμούς αυτούς, η μελέτη ανέδειξε τη μεγάλη δυναμική που παρουσιάζει ο συνδυασμός δεδομένων ψηφιακής ανάπτυξης και τεχνικών μηχανικής μάθησης στην ανάλυση διαφορών αλλά και ομοιοτήτων στην ανάπτυξη των χωρών και στη διερεύνηση των παραγόντων που αποτελούν καταλύτες της ψηφιακής ανάπτυξης.

Η συστηματική καταγραφή των παραπάνω συμπερασμάτων οδηγεί στην ανάγκη διατύπωσης προτάσεων για μελλοντική έρευνα, οι οποίες θα μπορούσαν να διευρύνουν περαιτέρω το πεδίο ανάλυσης και να καλύψουν υφιστάμενα κενά.

Ειδικότερα προτείνεται η επέκταση της ανάλυσης και η χρήση δεδομένων περισσότερων ετών του δείκτη DESI, ώστε να αναδειχθεί η ψηφιακή πορεία των κρατών-μελών σε βάθος χρόνου. Επιπλέον, κρίνεται χρήσιμη η ενσωμάτωση κοινωνικών και οικονομικών δεικτών, για βαθύτερη κατανόηση των αιτίων διαφοροποίησης.

Επιπρόσθετα, θα ήταν χρήσιμη η επιμέρους ανάλυση ανά γεωγραφικές περιφέρειες (π.χ. Βόρεια Ευρώπη, Ανατολική Ευρώπη) αλλά και τοπικά εντός των χωρών (πόλεις – περιφέρειες) για την κατανόηση τοπικών ιδιαιτεροτήτων. Είναι απαραίτητη η διερεύνηση των αιτιών πίσω από τις διαφορές ανά περιοχή, με έμφαση σε παράγοντες όπως η οικονομική κατάσταση, η

πολιτική σταθερότητα και οι ιδιαιτερότητες του εκπαιδευτικού συστήματος αλλά και οι δυνατότητες εγκατάστασης σύγχρονων υποδομών συνδεσιμότητας.

Όσον αφορά στο τεχνικό (αλγοριθμικό) μέρος της μελέτης, προτείνεται η εφαρμογή πιο μοντέρνων αλγόριθμων [π.χ. Deep Learning ή XGBoost (Chen & Guestrin, 2016; LeCun, Bengio, & Hinton, 2015), ως μετεξέλιξη της AdaBoost], ή χρήση τεχνικών όπως η ανάλυση συστάδων με δυναμικό αριθμό ομάδων (k clusters), για πιο λεπτομερή αποτελέσματα. Το ίδιο ισχύει και για την επιλογή σχεδόν βέλτιστου υποσυνόλου χαρακτηριστικών γνωρισμάτων, με πιο state-of-the-art μεθόδους. Η τάση σήμερα στη μηχανική μάθηση λέγεται επεξηγήσιμη (explainable) μηχανική μάθηση, και με εφαρμογή μεθόδων όπως η SHAP (Lundberg & Lee, 2017), διασφαλίζει την επεξηγησιμότητα και τη διαφάνεια της διαδικασίας της μάθησης. Ως συνέπεια, με τη SHAP (για παράδειγμα) μπορεί κανείς όχι μόνο να επιλέξει τα καλύτερα χαρακτηριστικά γνωρίσματα που βελτιώνουν την ακρίβεια κατά τη μάθηση, αλλά και να απεικονίσει το «γιατί» και το «πώς» αυτά τα γνωρίσματα επηρεάζουν το τελικό αποτέλεσμα της κατηγοριοποίησης. Το Deep Learning και η XGBoost είναι σύγχρονες μέθοδοι που η κάθε μία τους μπορεί να συνδυαστεί άριστα με τη SHAP για μια σχεδόν βέλτιστη ακρίβεια και επιλογή χαρακτηριστικών γνωρισμάτων, αλλά και για επεξηγησιμότητα της διαδικασίας. Στην παρούσα φάση, καμία από τις παραπάνω μεθόδους δεν έχει υλοποιηθεί στο περιβάλλον του WEKA.

Τέλος προτείνεται η σύγκριση των αποτελεσμάτων της ΕΕ με άλλες οικονομίες (π.χ., ΗΠΑ, Ασία) για να εντοπιστούν οι αποδοτικότερες πρακτικές και να αντιμετωπιστούν οι επί μέρους προκλήσεις.

Η συνέχεια της έρευνας με τον τρόπο αυτό, θα μπορούσε να συμβάλει ουσιαστικά στη διαμόρφωση στοχευμένων πολιτικών ψηφιακής ανάπτυξης εντός της Ευρωπαϊκής Ένωσης και στην ανάπτυξη εργαλείων πολιτικής, όπως πλαίσια αξιολόγησης για εθνικές στρατηγικές ψηφιακής μετάβασης.

Βιβλιογραφία

- Aggarwal, C. C., Hinneburg, A., & Keim, D. A. (2001). On the Surprising Behavior of Distance Metrics in High Dimensional Space. In Van & V. Vianu (Eds.), *Database Theory — ICDT 2001* (pp. 420–434). Berlin, Heidelberg: Springer Berlin Heidelberg.
- Annoni, P., Dominicus, de, & Khabirpour, N. (2019). Location matters: A spatial econometric analysis of regional resilience in the European Union. *Growth and Change*, 50. <https://doi.org/10.1111/grow.12311>
- Arthur, D., & Vassilvitskii, S. (2007). K-Means++: The Advantages of Careful Seeding. In *Proc. of the Annu. ACM SIAM Symp. on Discrete Algorithms* (Vol. 8, pp. 1027–1035). <https://doi.org/10.1145/1283383.1283494>
- Bánhidí, Z., & Dobos, I. (2023). Measurement of digital development with partial orders, Tiered DEA, and cluster analysis for the European Union. *International Review of Applied Sciences and Engineering*, 14(3), 392–401. <https://doi.org/10.1556/1848.2023.00612>
- Bishop, C. (2006). Pattern Recognition and Machine Learning. In *Journal of Electronic Imaging* (Vol. 16, p. 758). Springer. <https://doi.org/10.1117/1.2819119>
- Bluszczyk, A. (2015). Classification of the European Union member states according to the relative level of sustainable development. *Quality & Quantity*, 50(6), 2591–2605. <https://doi.org/10.1007/s11135-015-0278-x>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297. <https://doi.org/10.1007/BF00994018>
- Chen, T., & Guestrin, C. (2016). XGBoost: a Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, 1(1), 785–794. <https://doi.org/10.1145/2939672.2939785>
- Duda, R., Hart, P., & Stork, D. (2001). Pattern Classification. In *Pattern Classification*. Wiley.
- Ebner, J. (2023, July 31). Perceptrons, Explained - Sharp Sight. Retrieved from www.sharpsightlabs.com website: <https://www.sharpsightlabs.com/blog/perceptrons-explained/>
- European Commission. (2022). DESI | Shaping Europe's digital future. Retrieved from [digital-strategy.ec.europa.eu](https://digital-strategy.ec.europa.eu/en/policies/desi) website: <https://digital-strategy.ec.europa.eu/en/policies/desi>

- European Commission. (2022). Digital Economy and Society Index (DESI) 2022: Methodological Note. Publications Office of the European Union. Retrieved from <https://ec.europa.eu/digital-strategy/desi>.
- Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Goldberg D. E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley Professional.
- Grzebyk, M., & Stec, M. (2015). Sustainable Development in EU Countries: Concept and Rating of Levels of Development. *Sustainable Development*, 23. <https://doi.org/10.1002/sd.1577>
- Guyon, I., & Elisseeff, A. (2003). An Introduction of Variable and Feature Selection. *J. Machine Learning Research Special Issue on Variable and Feature Selection*, 3, 1157–1182. <https://doi.org/10.1162/153244303322753616>
- Hastie, T., Rosset, S., Zhu, J., & Zou, H. (2009). Multi-class AdaBoost. *Statistics and Its Interface*, 2(3), 349–360. <https://doi.org/10.4310/sii.2009.v2.n3.a8>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. In *Springer Series in Statistics*. New York, NY: Springer New York. <https://doi.org/10.1007/978-0-387-84858-7>
- Haykin, S. (2009). *Neural Networks and Learning Machines*. Upper Saddle River, N.J.: Pearson Education.
- Human development in the European Union. (2023). Retrieved from TheGlobalEconomy.com website: https://www.theglobaleconomy.com/rankings/human_development/European-union/
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651–666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Jainvidip. (2024, June 28). Understanding Support Vector Machines (SVMs) - Jainvidip - Medium. Retrieved March 8, 2025, from Medium website: <https://medium.com/%40jainvidip/understanding-support-vector-machines-svms-1f7c78bad934>
- Kaufman, L., & Rousseeuw, P. (1990). Finding Groups in Data: An Introduction to Cluster Analysis. In *Wiley, New York. ISBN 0471878766*. <https://doi.org/10.2307/2532178>

- Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, 2, 1137–1143.
- Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1-2), 273–324. [https://doi.org/10.1016/s0004-3702\(97\)00043-x](https://doi.org/10.1016/s0004-3702(97)00043-x)
- Kurama, V. (2024). A Guide to AdaBoost: Boosting To Save The Day | DigitalOcean. Retrieved from Digitalocean.com website: <https://www.digitalocean.com/community/tutorials/adaboost-optimizer>
- Laitsou, E., & Xenakis, A. (2023). *The impact of digital development on human well-being and vice versa*. International Telecommunications Society (ITS). Retrieved from <https://hdl.handle.net/10419/277995>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Leogrande, A., Magaletti, N., Cosoli, G., & Massaro, A. (2022). *Munich Personal RePEc Archive e-Government in Europe. A Machine Learning Approach e-Government in Europe. An Augmented Data Approach*. Retrieved from https://mpra.ub.uni-muenchen.de/112242/1/MPRA_paper_112242.pdf
- Lundberg, S., & Lee, S. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30. NeurIPS. <https://doi.org/10.48550/arXiv.1705.07874>
- McQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume I: Statistics*, 281–297.
- Martín, C. J., & Carnero, M. C. (2019). Evaluation of Sustainable Development in European Union Countries. *Applied Sciences*, 9(22). <https://doi.org/10.3390/app9224880>
- Masoura, M., & Malefaki, S. (2023). Evolution of the Digital Economy and Society Index in the European Union: A Socioeconomic Perspective. *TalTech Journal of European Studies*, 13(2), 177–203. <https://doi.org/10.2478/bjes20230020>
- Mitchell, T. M. (1997). *Machine Learning*. McGrawHill. Retrieved from <https://books.google.gr/books?id=EoYBngEACAAJ>
- Mohamad, I., & Usman, D. (2013). Standardization and Its Effects on K-Means Clustering Algorithm. *Research Journal of Applied Sciences, Engineering and Technology*, 6, 3299–3303. <https://doi.org/10.19026/rjaset.6.3638>

- OECD. (2019). Indicators. Retrieved April 5, 2025, from OECD website:
<https://www.oecd.org/en/data/indicators.html?orderBy=mostRelevant&page=4&facetTags=oecd-policy-areas%3Apa6%2Coecd-policy-areas%3Apa7>
- Onuferová, E., Čabinová, V., & Matijová, M. (2020). Categorization of the EU Member States in the Context of Selected Multicriteria International Indices Using Cluster Analysis. *Review of Economic Perspectives*, 20(3), 379–401.
<https://doi.org/10.2478/revecp20200018>
- Provost, F., & Fawcett, T. (2019a). *Η επιστήμη των δεδομένων για επιχειρήσεις* (p. 526). Κλειδάριθμος. (Original work published 2013)
- Provost, F., & Fawcett, T. (2019b). *Η επιστήμη των δεδομένων για επιχειρήσεις* (p. 526). Κλειδάριθμος. (Original work published 2013)
- Quinlan, J. R. (1993). *C4.5 programs for machine learning*. San Mateo, Calif. M. Kaufmann. Retrieved from <https://doi.acm.org/10.1145/152181>
- R Core Team. (2024). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.rproject.org/>
- Random Forest Algorithm - Javatpoint. (2021). Retrieved from www.javatpoint.com website:
<https://www.tpointtech.com/machine-learning-random-forest-algorithm>
- Saravanan, N., & Gayathri, V. (2018). Performance and Classification Evaluation of J48 Algorithm and Kendall's Based J48 Algorithm (KNJ48). *International Journal of Computer Trends and Technology*, 59(2), 73–80.
<https://doi.org/10.14445/22312803/ijctt-v59p112>
- Spearman C. (1987). The proof and measurement of association between two things. By C. Spearman, 1904. *PubMed*, 100(3-4), 441–471.
- Stec, M. A., Filip, P. B., Grzebyk, M. J., & Pierscieniak, A. (2014). Socio-economic Development in the EU Member States – Concept and Classification. *Engineering Economics*, 25(5). <https://doi.org/10.5755/j01.ee.25.5.6413>
- Sutton, R. S., & Barto, A. (2018). *Reinforcement learning: an introduction*. Cambridge, Ma; London: The MIT Press.
- Tan, PangNing, Steinbach, M., Michael, A. O., Kumar, V., & Vipin. (2005). *Introduction to Data Mining*. Pearson.
- United Nations. (2022). UN E-Government Survey 2022. Retrieved from Un.org website:
<https://publicadministration.un.org/egovkb/en-us/Reports/UN-E-Government-Survey-2022>

- Verma, N. (2022). Adaboost Algorithm explained in less than 5 minutes. Retrieved from www.medium.com
- Walker, S. (2023, July 4). F-Score: What are Accuracy, Precision, Recall, and F1-Score? — Klu. Retrieved from klu.ai website: <https://klu.ai/glossary/accuracy-precision-recall-f1>
- Witten, I. H., & AI, E. (2017). *Data mining: practical machine learning tools and techniques*. Amsterdam: Morgan Kaufmann.

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.