ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ ΜΕΤΑΛΛΕΙΩΝ ΜΕΤΑΛΛΟΥΡΓΩΝ

ΤΟΜΕΑΣ ΜΕΤΑΛΛΕΥΤΙΚΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΠΡΟΒΛΕΨΗ ΚΙΝΔΥΝΟΥ ΕΝΕΡΓΕΙΑΚΗΣ ΦΤΩΧΕΙΑΣ ΜΕ ΧΡΗΣΗ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Τσουκνίδας Αθανάσιος

Επιβλέπων διπλωματικής εργασίας:

Καθηγητής Δημήτρης Καλιαμπάκος

ΑΘΗΝΑ, ΙΟΥΝΙΟΣ 2022

Περιεχόμενα

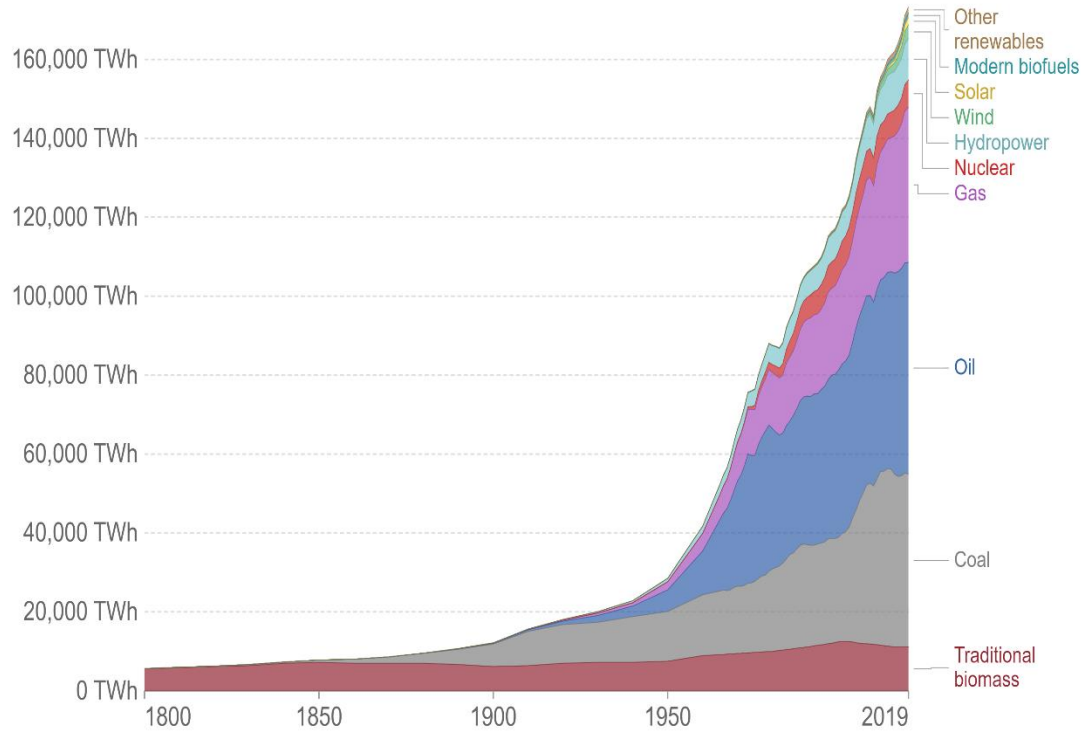
Εισαγωγή 5
Κεφάλαιο 1. Ενεργειακή φτώχεια 9
1.1. Ενεργειακή Κρίση 9
1.2. Ενεργειακή κρίση στην Ελλάδα 12
1.3. Ενεργειακή φτώχεια 14
1.4. Επιπτώσεις ενεργειακής φτώχειας 16
1.4.1.Επιπτώσεις στην ανθρώπινη υγεία 16
1.4.2 Επιπτώσεις στην ποιότητα ζωής 18
1.4.3 Επιπτώσεις στο περιβάλλον 19
1.4.4. Επιπτώσεις στην οικονομία 19
1.5.Η εξέλιξη της έρευνας στον ορισμό και τον υπολογισμό της ενεργειακής φτώχειας 20
Κεφάλαιο 2. Ορεινές περιοχές 29
2.1. Η σημασία των ορεινών περιοχών 29
2.1.1.Ορεινές περιοχές και διεθνής προσέγγιση 29
2.1.2.Συνεισφορά ορεινών οικοσυστημάτων 31
2.2. Το ενεργειακό πρόβλημα των ορεινών περιοχών 33
Κεφάλαιο 3. Αλγόριθμοι μηχανικής μάθησης 39
3.1 Εισαγωγή στους Αλγόριθμους μηχανικής μάθησης 39
3.1.1. Ορισμός αλγορίθμων μηχανικής μάθησης 39
3.1.2. Τύποι μοντέλων Μηχανικής Μάθησης 40
3.1.3.Διαδεδομένοι αλγόριθμοι μηχανικής μάθησης 41
3.1.4. Εφαρμογές Αλγορίθμων Μηχανικής μάθησης 48
3.2. Αλγόριθμοι μηχανικής μάθησης στο ζήτημα της ενεργειακής φτώχειας 51
3.3.WEKA 52
3.3.1.Εισαγωγικά 52
3.3.2.Προεπεξεργασία 55
3.3.3.Ταξινομητές- Αλγόριθμοί μηχανικής μάθησης 57
Κεφάλαιο 4. Μεθοδολογία 64
Κεφάλαιο 5. Αποτελέσματα 68
5.1. Αποτελέσματα ανά αλγόριθμο 68
5.1.1 Naïve Bayes 68
5.1.2. IBk (Instance-based learning with parameter k) 73

5.1.3.Decision Table Classifier 78
5.1.4. J48..... 83
5.1.5.AdaBoost.M1- Decision Table 90
5.1.6. Cost-Sensitive Classifier- J48 96
5.2. Συγκριτική αξιολόγηση αποτελεσμάτων..... 101
Κεφάλαιο 6. Συμπεράσματα 105
Βιβλιογραφία 110

Εισαγωγή

Η αφθονία και η διαθεσιμότητα ενέργειας έχει μεταμορφώσει την πορεία της ανθρωπότητας τους τελευταίους αιώνες, ειδικότερα από τον 19^{ο} αιώνα μέχρι σήμερα. Με το πέρασμα των ετών, ο άνθρωπος έχει «εκμεταλλευτεί» πληθώρα πηγών ενέργειας, όπως τα ορυκτά καύσιμα (άνθρακας, πετρέλαιο, φυσικό αέριο), τις Ανανεώσιμες Πηγές Ενέργειας (ηλιακή, αιολική, γεωθερμία, κλπ.), την πυρηνική και την υδροηλεκτρική ενέργεια. Παράλληλα με την εξέλιξη στις μεθόδους παραγωγής ενέργειας, υπήρξαν και σημαντικές μεταβολές στις ποσότητες ενέργειας που παράγονται και καταναλώνονται παγκοσμίως.

Το ενεργειακό σύστημα έχει μεταβληθεί δραματικά μετά τις αλλαγές που επέφερε η Βιομηχανική Επανάσταση, αυξάνοντας σημαντικά τις δυνατότητες παραγωγής αλλά και της κατανάλωσης ενέργειας σε παγκόσμιο επίπεδο. Στο παρακάτω διάγραμμα απεικονίζεται η μεταβολή της παγκόσμιας πρωτογενούς κατανάλωσης ηλεκτρικής ενέργεια από το 1840 έως και σήμερα (Εικ. 1). Η πρωτογενής κατανάλωση ενέργειας υπολογίζει την συνολική ενεργειακή ζήτηση μίας χώρας. Περιλαμβάνει την κατανάλωση του ίδιου του ενεργειακού τομέα, τις απώλειες από μετατροπές ενέργειας (π.χ. από πετρέλαιο ή φυσικό αέριο σε ηλεκτρική ενέργεια) και την τελική κατανάλωση από τους τελικούς χρήστες.



Εικόνα 1: Παγκόσμια κατανάλωση ηλεκτρικής ενέργειας ανά πηγή παραγωγής ηλεκτρικής ενέργειας . (Πηγή: Our World in Data, 2021)

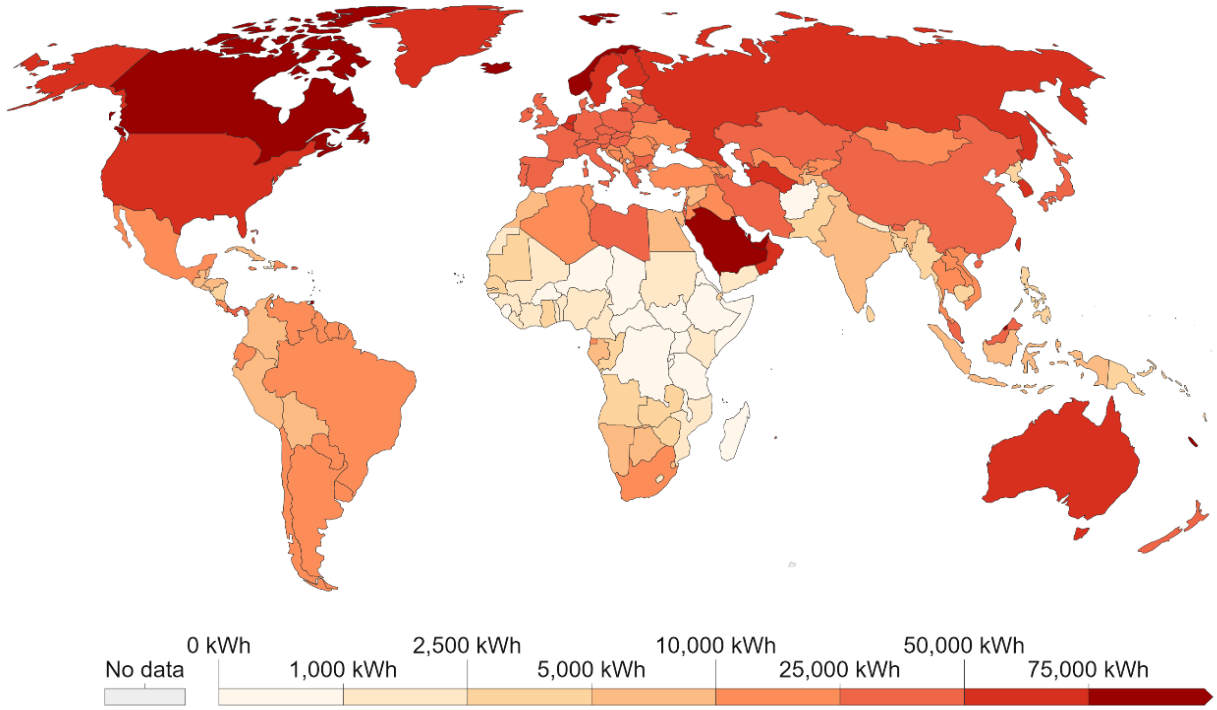
Είναι απαραίτητο να αναφερθεί ότι το παραπάνω διάγραμμα (Εικ.1) αναπαριστά την κύρια ενεργειακή κατανάλωση, η οποία υπολογίζεται με τη μέθοδο της αντικατάστασης (Substitution Method). Η συγκεκριμένη μέθοδος -σε σύγκριση με την ''άμεση μέθοδο''– επιχειρεί να διορθώσει τις ενεργειακές απώλειες (ενέργεια που χάνεται ως θερμότητα κατά την καύση) στις διεργασίες μετατροπής ενέργειας στα ορυκτά καύσιμα και στη βιομάζα. Καλύπτει αυτές τις απώλειες, μετατρέποντας την ενεργειακή παραγωγή από πυρηνική ενέργεια και Ανανεώσιμες Πηγές Ενέργειας στην ισοδύναμη πρωτογενή παραγωγή, αν αυτή η ενέργεια παραγόταν από ορυκτά καύσιμα.

Όπως φαίνεται, λοιπόν, στο παραπάνω διάγραμμα (Εικ. 1), με τη σταδιακή προσθήκη νέων πηγών ενέργειας, η παγκόσμια κατανάλωση εκτοξεύτηκε, ειδικότερα μετά το 1950, όταν η κυριαρχία του πετρελαίου και του φυσικού αερίου αύξησαν την παγκόσμια κατανάλωση κατά 320% έως το 2000 (από 28.516 σε 122.073 TWh) ενώ, χαρακτηριστικά, το 2019 η παγκόσμια κατανάλωση εμφανίζεται αυξημένη κατά 500% σε σχέση με το 1950, χάριν και της προσθήκης των Ανανεώσιμων Πηγών Ενέργειας. Γίνεται λοιπόν προφανές ότι με τη συνεχή ανάπτυξη νέων τεχνολογιών στον τομέα της παραγωγής ενέργειας, αλλά και τη δυνατότητα παραγωγής τεράστιων ποσοτήτων ενέργειας, η «δίψα» του παγκόσμιου πληθυσμού και της βιομηχανίας για ενέργεια αυξάνεται με φρενήρεις ρυθμούς. Ενδεικτικά, για το 2020, η μέση ημερήσια κατανάλωση ανά άτομο έφτασε τις 58 kWh (BP, 2022) και η ενεργειακή ζήτηση αναμένεται να αυξηθεί κατά 48% έως το 2040 (IEA, 2016).

Ωστόσο, παρά τη ραγδαία αύξηση στα ενεργειακά μεγέθη, οι υφιστάμενες ανισότητες είναι τεράστιες και το ενεργειακό χάσμα εντείνεται διαρκώς. Πιο συγκεκριμένα, ενώ στην πλειονότητα των ανεπτυγμένων χωρών παρατηρείται εκτεταμένη σπατάλη ενέργειας, ο αναπτυσσόμενος κόσμος αντιμετωπίζει σοβαρά προβλήματα ακόμη και σε θεμελιώδεις υπηρεσίες ενέργειας, όπως η πρωταρχική ανάγκη για ασφαλή και σταθερή πρόσβαση σε ενέργεια.

Στον παρακάτω χάρτη (Εικ.2) παρουσιάζεται η ετήσια κατά κεφαλήν κατανάλωση ανά την υφήλιο, η οποία περιλαμβάνει ηλεκτρική ενέργεια, μεταφορές, θέρμανση και ενεργειακές δαπάνες για μαγείρεμα.

a



Εικόνα 2: Μέση ετήσια κατά κεφαλήν ενεργειακή κατανάλωση (kWh). (Πηγή: Our World in Data, 2021)

Από το παραπάνω διάγραμμα (Εικ.2) παρατηρούνται εύκολα οι ανισότητες που επικρατούν σε ενεργειακό επίπεδο μεταξύ ανεπτυγμένων και αναπτυσσόμενων χωρών. Χαρακτηριστική σύγκριση μπορεί να γίνει μεταξύ των χωρών που αντιπροσωπεύουν τα δύο άκρα του φαινομένου. Ενδεικτικά, για το 2020, το κράτος του Κατάρ εμφάνισε κατά κεφαλήν ενεργειακή κατανάλωση 165.044 kWh, ενώ για το ίδιο έτος, στην Ινδία, η κατά κεφαλήν κατανάλωση ανήλθε μόλις στις 6.438 kWh. Σε γενικές γραμμές, η λίστα των χωρών με τη μεγαλύτερη ενεργειακή κατανάλωση περιλαμβάνει την Ιρλανδία, τη Νορβηγία, τον Καναδά, τις Η.Π.Α. και τις εύπορες χώρες της Μέσης Ανατολής όπως, Κατάρ, Ομάν και Σαουδική Αραβία. Ο μέσος άνθρωπος αυτών των χωρών καταναλώνει έως και 100 φορές περισσότερη ενέργεια από τον μέσο άνθρωπο των φτωχότερων χωρών όπως το Μπαγκλαντές και η Ινδία.

Ταυτόχρονα, οι αδύναμες οικονομίες αυτών των χωρών έχουν καταδικάσει μεγάλα ποσοστά των πληθυσμών τους να ζουν χωρίς πρόσβαση σε ηλεκτρική ενέργεια και κατάλληλες εγκαταστάσεις μαγειρέματος. Η αδυναμία πρόσβασης ενός πληθυσμού σε ηλεκτρική ενέργεια αποτελεί έναν αντιπροσωπευτικό δείκτη για την κοινωνική και οικονομική κατάσταση ενός κράτους. Η έννοια ‘’πρόσβαση σε ηλεκτρική ενέργεια’’, παρόλο που δεν υπάρχει ως κοινά συμφωνημένος ορισμός, στις περισσότερες περιπτώσεις περιλαμβάνει έννοιες όπως η παροχή ηλεκτρικής ενέργειας, οι ασφαλείς εγκαταστάσεις μαγειρέματος και η ύπαρξη ενός ελάχιστου ορίου κατανάλωσης ηλεκτρικής ενέργειας.

Ενώ σε παγκόσμιο επίπεδο η πρόσβαση σε ηλεκτρική ενέργεια αυξάνεται σταθερά τις τελευταίες δεκαετίες, με 1,26 δισεκατομμύρια ανθρώπους να αποκτούν πρόσβαση από το 2005 έως το 2016, ένα τεράστιο μέρος του πληθυσμού εξακολουθεί να παραμένει χωρίς πρόσβαση σε

ηλεκτρική ενέργεια. Πιο συγκεκριμένα, για το 2019, περισσότερα από 760 εκατομμύρια άνθρωποι (10% του παγκόσμιου πληθυσμού) δεν έχουν πρόσβαση σε ηλεκτρικό ρεύμα ενώ, παράλληλα, περισσότεροι από 2,5 δισεκατομμύρια άνθρωποι δεν έχουν πρόσβαση σε κατάλληλες εγκαταστάσεις μαγειρέματος. Ο μεγαλύτερος όγκος πληθυσμού που αντιμετωπίζει αυτά τα προβλήματα συγκεντρώνεται στην Υποσαχάρια Αφρική και στις αναπτυσσόμενες περιοχές της Ασίας.

Το φαινόμενο αυτό αποτελεί μία πλευρά του ευρύτατου κοινωνικού και οικονομικού προβλήματος της ενεργειακής φτώχειας (energy poverty). Η ενεργειακή φτώχεια αποτελεί ένα ζήτημα το οποίο επηρεάζει έντονα τόσο τις κοινωνίες του αναπτυσσόμενου κόσμου, όσο και τις πιο εύπορες του ανεπτυγμένου κόσμου. Όσον αφορά τον τρόπο έκφρασή της στις ανεπτυγμένες κοινωνίες, η ενεργειακή φτώχεια σχετίζεται κυρίως με τα υψηλά κόστη που διαμορφώνονται σε σχέση με τα διαθέσιμα εισοδήματα των νοικοκυριών, ενώ στις αναπτυσσόμενες κοινωνίες το πρόβλημα συνδέεται κυρίως με την αδυναμία πρόσβασης σε ηλεκτρική ενέργεια. Ωστόσο, ενώ η ενεργειακή φτώχεια αποτελεί ένα κρίσιμο ζήτημα για τις αναπτυσσόμενες κοινωνίες, σημαντική όξυνση του φαινομένου παρατηρείται και στις ανεπτυγμένες κοινωνίες τα τελευταία χρόνια, και ειδικότερα στην Ευρώπη, με την Ελλάδα να πλήττεται ιδιαίτερα έντονα. Εν μέσω οικονομικής κρίσης και αδυναμίας του κράτους να απορροφήσει τις υψηλές τιμές στη βιομηχανία της ενέργειας, οι καταναλωτές έρχονται αντιμέτωποι με επιβαρυμένους λογαριασμούς ηλεκτρικούς ρεύματος, τους οποίους είτε αδυνατούν είτε καθυστερούν να εξοφλήσουν. Αυτή η δυσχέρεια έχει ως αποτέλεσμα οι καταναλωτές να περικόπτουν έξοδα από άλλες βασικές ανάγκες προκειμένου να εξασφαλίζουν βασικές παροχές όπως η θέρμανση της κατοικίας τους, να πληρώνουν σημαντικό κομμάτι του εισοδήματός τους σε δαπάνες ηλεκτρικής ενέργειας ή, σε κάποιες περιπτώσεις, να στρέφονται σε φθηνότερα και λιγότερο αποδοτικά συστήματα θέρμανσης (π.χ. τζάκια, σόμπες), τα οποία αποτελούν πηγές εξωτερικής και εσωτερικής ατμοσφαιρικής ρύπανσης (Παπαδά, 2017).

Συνολικά, ανεξαρτήτως της μορφής με την οποία εμφανίζεται η ενεργειακή φτώχεια, είτε πλήττοντας φτωχότερα κράτη, είτε ανεπτυγμένα, οι οικονομικές, κοινωνικές και περιβαλλοντικές επιπτώσεις του προβλήματος, καθώς και οι συνέπειες στην ανθρώπινη υγεία και ποιότητα ζωής έχουν πλέον αναγνωριστεί διεθνώς ως μείζον ζήτημα. Η έρευνα, η ανάλυση και η αντιμετώπιση του φαινομένου αποτελεί αντικείμενο υψηλού ενδιαφέροντος ανά τον κόσμο, με αποτέλεσμα οι μελέτες επί του θέματος να επεκτείνονται συνεχώς (Παπαδά, 2017).

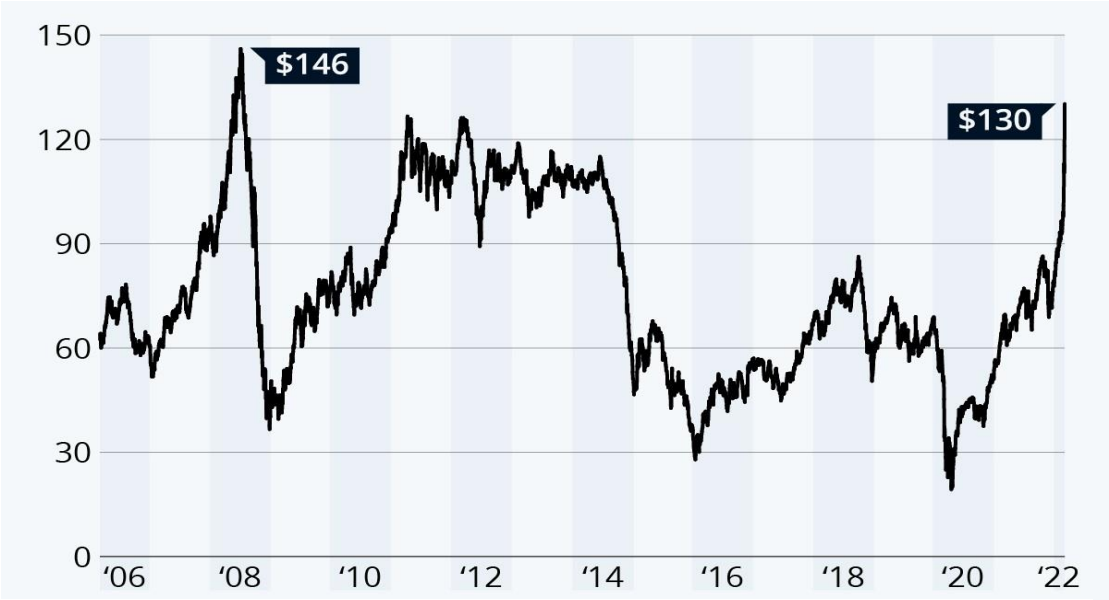
Οι μέχρι τώρα μελέτες που αφορούν την ενεργειακή φτώχεια εστιάζουν στην ανάλυση του φαινομένου στις αστικές περιοχές, ενώ οι ορεινές περιοχές έχουν παραμεληθεί, παρόλο που αποτελούν περιοχές ιδιαίτερα ευάλωτες στην ενεργειακή φτώχεια λόγων των εγγενών χαρακτηριστικών τους. Πιο συγκεκριμένα, το ψυχρό κλίμα, η απομόνωση, η ασθενής οικονομική/παραγωγική δραστηριότητα και η παλαιότητα των κατοικιών αποτελούν γνωρίσματα τα οποία ευνοούν την όξυνση του φαινομένου και καθιστούν τον υπολογισμό της έκτασης της ενεργειακής φτώχειας ουσιώδη (Katsoulakos et al., 2014; Papada & Kaliampakos, 2017). Σύμφωνα με πρόσφατη έρευνα, τα νοικοκυριά των ορεινών περιοχών στον ελλαδικό χώρο είναι πιο ευάλωτα στην ενεργειακή φτώχεια από τα νοικοκυριά στο επίπεδο της χώρας κατά 89,7% (Papada & Kaliampakos, 2019).

Στην παρούσα εργασία θα πραγματοποιηθεί επισκόπηση των ιδιαίτερων χαρακτηριστικών του προβλήματος την ενεργειακής φτώχειας στις ορεινές περιοχές, ξεκινώντας από την παρουσίαση των βασικών γνωρισμάτων του φαινομένου και την εξέλιξη της έρευνας πάνω στο αντικείμενο και, εν συνεχεία, εξετάζοντας εκτενέστερα τη σημασία του φαινομένου για τις ορεινές περιοχές. Επιπλέον, η εργασία θα επικεντρωθεί στη χρήση των αλγορίθμων μηχανικής μάθησης και την υιοθέτηση τους για την εξαγωγή χρήσιμων συμπερασμάτων αναφορικά με τους δείκτες πρόβλεψης του κινδύνου ενεργειακής φτώχειας. Τέλος, θα παρουσιαστούν τα πειραματικά αποτελέσματα που προέκυψαν από τη σύγκριση των βασικών δεικτών της ενεργειακής φτώχειας, όπως υπολογίστηκαν με τη χρήση αλγορίθμων μηχανικής μάθησης, και θα πραγματοποιηθεί μια σύνοψη της σημασίας των αποτελεσμάτων.

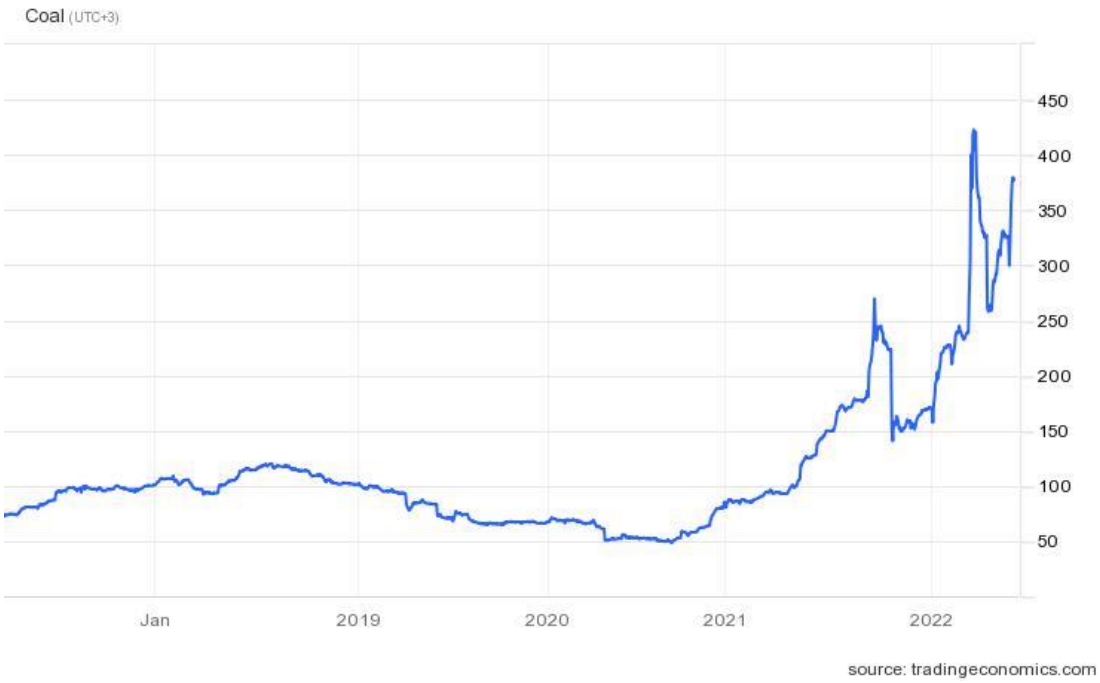
Κεφάλαιο 1. Ενεργειακή φτώχεια

1.1. Ενεργειακή Κρίση

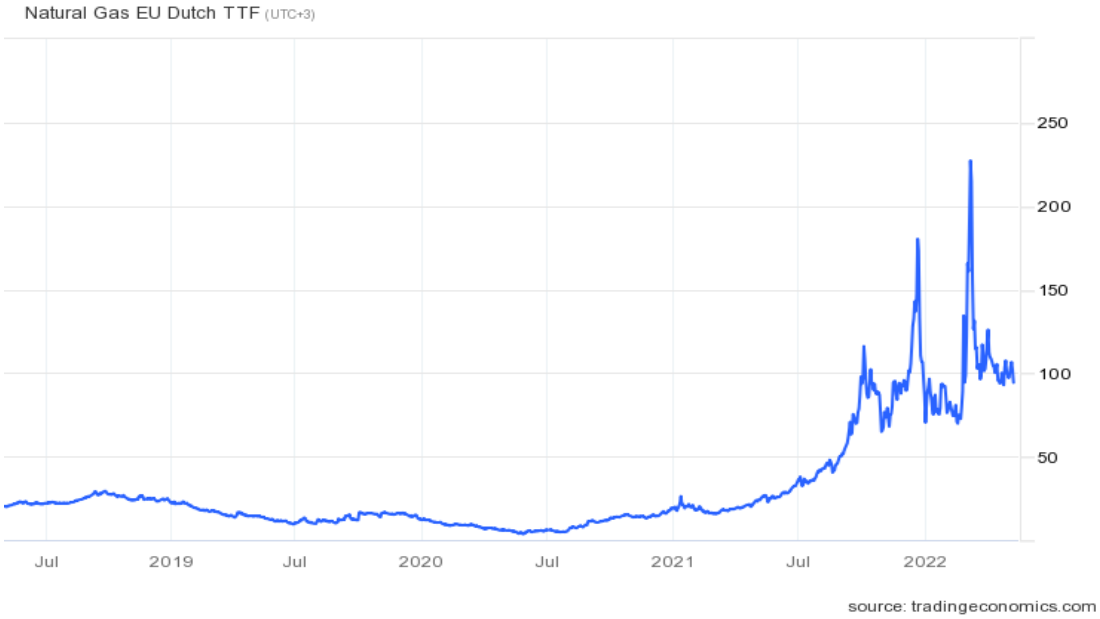
Ο πλανήτης βρίσκεται εν μέσω μια ταχείας παγκόσμιας ενεργειακής κρίσης. Η υφιστάμενη κρίση θα μπορούσε να εξελιχθεί ως η χειρότερη ενεργειακή κρίση, ξεπερνώντας την πετρελαϊκή κρίση του 1970. Οι τιμές του φυσικού αερίου στην Ασία και την Ευρώπη έχουν φτάσει σε επίπεδα-ρεκόρ (Εικ. 4). Οι τιμές του πετρελαίου έχουν εκτοξευθεί, ως απόρροια των ελλείψεων στον εφοδιασμό φυσικού αερίου (Εικ. 3) φτάνοντας σε υψηλό ρεκόρ τριετίας. Αυτές οι ελλείψεις έχουν συμπαρασύρει σε νέα υψηλά επίπεδα και την τιμή του άνθρακα (Εικ.5)



Εικόνα 3 : Ημερήσια τιμή ανά βαρέλι για το αργό πετρέλαιο Brent σε USD (Πηγή: Statista, 2022)



Εικόνα 4 : Τιμή Ολλανδικών futures φυσικού αερίου ανά MWh σε EUR για την Ευρωπαϊκή Ένωση.(Dutch TTF). (Πηγή: Trading Economics, 2022)



Εικόνα 5: Τιμή futures (Newcastle) άνθρακα USD/Τn. (Πηγή: Trading Economics, 2022)

Η ενεργειακή κρίση οφείλεται σε έναν σπάνιο συνδυασμό γεγονότων που δημιούργησαν τις ιδανικές συνθήκες για την αποσταθεροποίηση της παγκόσμιας αγοράς ενέργειας. Οι κύριοι παράγοντες που οδήγησαν σε αυτή την κρίση ήταν δύο. Πρώτον, η απελευθέρωση της συσσωρευμένης ζήτησης που συνόδευσε την επαναλειτουργία της παγκόσμιας οικονομίας. Η πανδημία Covid-19 έθεσε σε παύση την παγκόσμια οικονομική δραστηριότητα, με αποτέλεσμα την κατάρρευση της ζήτησης ενέργειας το 2020, με τις τιμές του πετρελαίου να καταγράφουν ιστορικό χαμηλό. Αυτή η τεχνητή μείωση της ζήτησης διορθώθηκε με την απελευθέρωση και ανάκαμψη της οικονομίας, με τις κεντρικές τράπεζες να εκδίδουν πακέτα δημοσιονομικής τόνωσης σε όλο τον κόσμο. Αντιπροσωπευτικό παράδειγμα αποτελεί η αυξημένη ασιατική ζήτηση για υγροποιημένο φυσικό αέριο (LNG) το 2021 εξαιτίας της ισχυρής κινέζικης ζήτησης, η οποία δημιούργησε ισχυρό ανταγωνισμό με την Ευρώπη για την προμήθεια υγροποιημένου φυσικού αέριου, με αποτέλεσμα την αύξηση των τιμών του φυσικού αερίου. Δεύτερον, οι τελευταίοι χειμώνες, με αποκορύφωμα το χειμώνα του 2021, ήταν ασυνήθιστα κρύοι, προκαλώντας εκ νέου αύξηση της ζήτησης για ενέργεια και για υψηλότερες ποσότητες καυσίμων. Συμπληρωματικά, μετά τον ιδιαίτερα ψυχρό χειμώνα του 2021, το βόρειο ημισφαίριο βίωσε έναν ''δεύτερο χειμώνα'', εξαιτίας θυελλωδών ανέμων από την Αρκτική και του φαινομένου Λα Νινιά, που διήρκησε μέχρι το τέλος της άνοιξης, οδηγώντας τα επίπεδα του αποθηκευμένου αέρια σε νέα χαμηλά επίπεδα για την εποχή του έτους.

Η πλευρά της προσφοράς επηρεάστηκε από τρεις βασικά παράγοντες. Πρώτον, η κατάρρευση της ζήτησης ενέργειας εν μέσω πανδημίας οδήγησε πολλές εταιρίες να μειώσουν την παραγωγή τους κατά τη διάρκεια του 2020, προκαλώντας προβλήματα σε θέματα αποθήκευσης, συντήρησης και εφοδιαστικών ελλείψεων. Για παράδειγμα, η Νορβηγία παρείχε περιορισμένες ποσότητες φυσικού αερίου στην Ευρώπη εξαιτίας προβλημάτων συντήρησης. Δεύτερον, οι ΑΠΕ, όπως τα αιολικά πάρκα και οι υδροηλεκτρικές μονάδες, υπολειτουργούσαν. Εν μέσω αυτής της κρίσης αναδείχθηκε η βασική αδυναμία των ΑΠΕ, ήτοι η αδυναμία να παράγουν μια συνεχή και σταθερή ποσότητα ενέργειας, αρκετή ώστε να μπορεί να καλύψει επαρκώς και εξ' ολοκλήρου τις ενεργειακές ανάγκες μιας ολόκληρης χώρας. Ειδικά κατά το πρόσφατο καλοκαίρι του 2021 αναδείχθηκε αυτή η αδυναμία των ΑΠΕ και, ειδικότερα των αιολικών πάρκων, καθώς χαρακτηρίσθηκε ως το ''καλοκαίρι χωρίς ανέμους''.

Τέλος, οι βασικοί προμηθευτές ενέργειας, κυρίως η Ρωσία, δεν κάλυψαν τις ελλείψεις παροχής φυσικού αερίου τόσο στην ευρωπαϊκή, όσο και στην ασιατική αγορά. Αυτή η κίνηση μπορεί να αποδοθεί στην αυξημένη εγχώρια ζήτηση και στην πιθανότητα να χρησιμοποιεί τη θέση της ως βασικός προμηθευτής για να επιταχύνει την έγκριση του αγωγού Nord Stream 2.

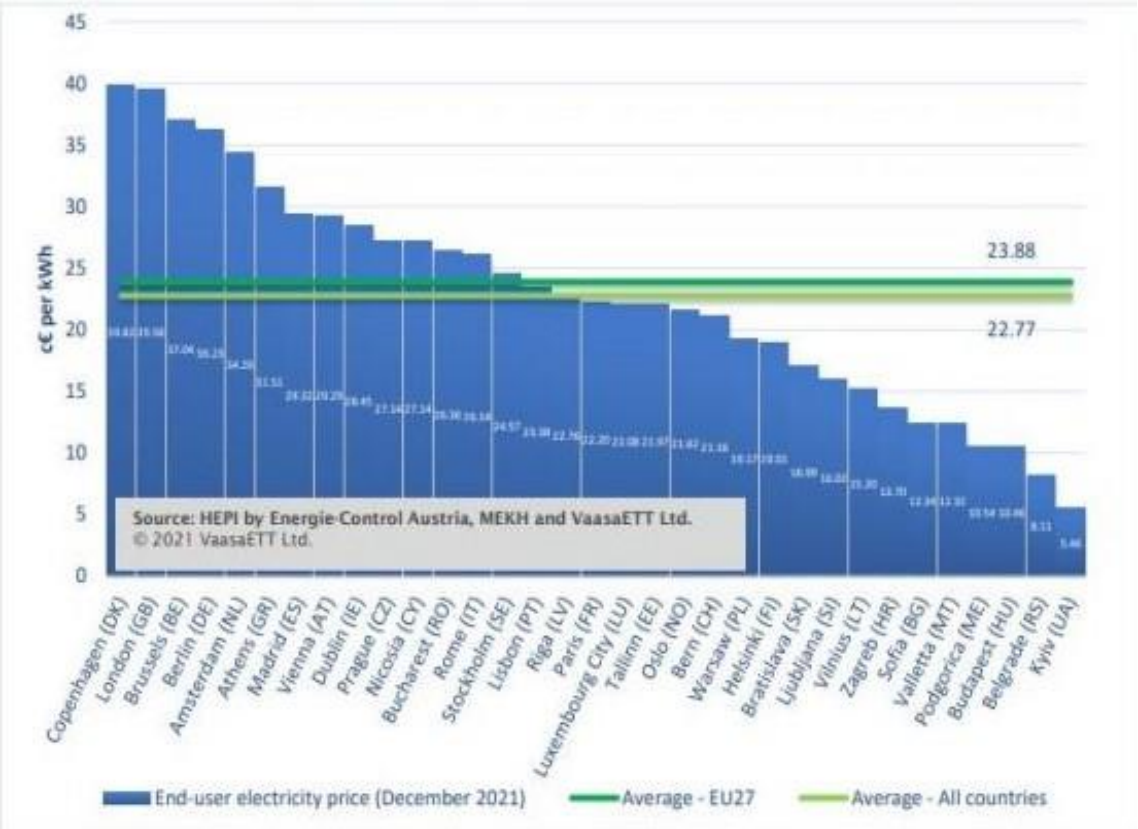
Φυσικά, με την εισβολή της Ρωσίας στην Ουκρανία, η κατάσταση στην αγορά ενέργειας αποσταθεροποιήθηκε ακόμη περισσότερο και συνεχίζει να επιδεινώνεται καθημερινά. Η εισβολή έφερε ακόμη μεγαλύτερη αστάθεια στην γεωπολιτική ισορροπία και τις αγορές ενέργειας, οδηγώντας την τιμή του πετρελαίου και του φυσικού αερίου σε υψηλά ρεκόρ δεκαετίας. Η Ευρώπη βρέθηκε εκτεθειμένη καθώς οι ενεργειακές τις ανάγκες καλύπτονται κατά ένα σημαντικό μέρος από τη Ρωσία. Η Ευρωπαϊκή Ένωση (ΕΕ) εισάγει από τη Ρωσία περίπου το 40% του φυσικού αερίου, το 25% του πετρελαίου και το 50% του άνθρακα εκ των συνολικών της εισαγωγών. Η ΕΕ, εν μέσω κυρώσεων προς τη Ρωσία, αποφάσισε να μειώσει τις εισαγωγές της από τη Ρωσία και σχεδιάζει να τις μειώσει συνολικά κατά 66% μέχρι το τέλος του 2022. Εντούτοις, η κίνηση αυτή απαιτεί την εξασφάλιση των ενεργειακών αναγκών της από άλλους προμηθευτές,

γεγονός το οποίο δεν αναμένεται να μειώσει τις τιμές των καυσίμων, τουλάχιστον σε σύντομο χρονικό διάστημα. Ως εκ τούτου, οι τιμές τόσο του φυσικού αερίου όσο και της ηλεκτρικής ενέργειας αναμένεται να παραμείνουν σε υψηλά και μη βιώσιμα προς τους καταναλωτές επίπεδα για το προσεχές χρονικό διάστημα.

1.2. Ενεργειακή κρίση στην Ελλάδα

Η Ελλάδα καθ' όλη τη διάρκεια της ενεργειακής κρίσης, έχει «χτυπηθεί» σκληρά από τις εξελίξεις στις ενεργειακές αγορές και την ανοδική τάση των τιμών, με χαρακτηριστικό παράδειγμα την πρώτη θέση που κατέχει για αρκετούς μήνες με την ακριβότερη τιμή ηλεκτρικής ενέργειας στη χονδρεμπορική αγορά της Ευρώπης.

Στο παρακάτω διάγραμμα (Εικ. 6) παρουσιάζονται οι τιμές ηλεκτρικής ενέργειας για τους καταναλωτές της ΕΕ για τον Δεκέμβριο 2021, με την Ελλάδα να κατέχει την 6η θέση υψηλότερης τιμής.



Εικόνα 6: Τελική τιμή ηλεκτρικής ενέργειας καταναλωτών (της ΕΕ για τον Δεκέμβριο 2021 με φόρους) σε c€ /kWh). (Πηγή: Household Energy Price Index, 2022)

Γίνεται κατανοητό ότι, ακόμη και πριν τις ακραίες τιμές που επέφερε η αποσταθεροποίηση λόγω της εισβολής στην Ουκρανία, οι Έλληνες καταναλωτές προμηθεύονταν ηλεκτρικό ρεύμα το οποίο συγκαταλεγόταν στα ακριβότερα της Ευρώπης. Με την έναρξη του πολέμου στην Ουκρανία, οι Έλληνες καταναλωτές ήρθαν αντιμέτωποι με πρωτόγνωρους λογαριασμούς ρεύματος, δίνοντάς τους την πρώτη θέση ανάμεσα σε όλες τις χώρες της ΕΕ για αρκετούς μήνες εντός του 2022. Αντιπροσωπευτικό παράδειγμα της απότομης αύξησης των τιμών του ηλεκτρικού ρεύματος αποτελεί η διαφορά που παρουσιάζεται μεταξύ των Μεσοσταθμικών Τιμών Αγοράς Ηλεκτρικής ενέργειας στο Διασυνδεδεμένο Σύστημα για του μήνες Φεβρουάριο και Μάρτιο. Για τον Φεβρουάριο του 2022 -πριν την έναρξη της εισβολής στη Ουκρανία- η Μεσοσταθμική Τιμή Αγοράς Ηλεκτρικής Ενέργειας ανήλθε σε 229,87 €/MWh, ενώ για το Μάρτιο η τιμή εκτοξεύθηκε στα 297,87 €/MWh (ΑΔΜΗΕ, 2022). Έκτοτε, η Μεσοσταθμική Τιμή του ηλεκτρικού ρεύματος παρέμεινε σε υψηλά επίπεδα με το μέσο όρο για τους πέντε μήνες του 2022 να ανέρχεται 258,25 €/MWh (ΑΔΜΗΕ, 2022). Συγκρίνοντας τον μέσο όρο των πέντε μηνών για το 2022, με το μέσο όρο των τιμών για τους ίδιους μήνες του 2021 (Ιανουάριος-Φεβρουάριος-Μάρτιος,-Απρίλιος-Μάιος) όπου κατά μέσο όρο η μεσοσταθμική του ηλεκτρικού ρεύματος κινούνταν στα 69,19 €/MWh (ΑΔΜΗΕ, 2022), παρατηρείται αύξηση των τιμών ηλεκτρικής ενέργειας κατά 273%. Αυτός ο υπερδιπλασιασμός των τιμών αγοράς ενέργειας μεταφέρθηκε απευθείας στο καταναλωτές, καθώς με βάση αυτές τις τιμές διαμορφώνεται και οι τιμές ενέργειας της λιανικής αγοράς των καταναλωτών.

Συγκεκριμένα, οι αυξήσεις της χονδρεμπορικής αγοράς ηλεκτρικής ενέργειας μεταβιβάζονται στους καταναλωτές και στους λογαριασμούς αυτών, μέσω της «ρήτρας αναπροσαρμογής». Ως ενδεικτικό παράδειγμα για το μήνα Μάρτιο του 2022, οι αυξήσεις των τιμών ηλεκτρικής ενέργειας μεταφράστηκαν στους καταναλωτές ως εξής (ΔΕΗ, 2022):

Η βασική τιμή ενός κιλοβατώρας (kWh) για το βασικό πάροχο ηλεκτρικής ενέργειας (ΔΕΗ) στο βασικό οικιακό τιμολόγιο Γ1 ήταν στα 0,111 €/kWh για κατανάλωση μέχρι 2.000 kWh και 0,119 €/kWh για κατανάλωση μεγαλύτερη των 2.000 kWh

Εν συνεχεία, προστίθεται η Ρήτρα Αναπροσαρμογής Χρέωσης Προμήθειας η οποία υπολογίζεται για τον μήνα t βάσει του τύπου: Y= α*x+β, όπου

- α: Συντελεστής Προσαύξησης, ίσος με 1,15
- x: Αριθμητικός μέσος όρος Τιμής Εκκαθάρισης Αγοράς (Market Cleaning Price – MCP) της Προ-Ημερήσιας Αγοράς (Day Ahead Market-DAM) του προηγούμενου μήνα (t-1), που δημοσιεύεται στην ιστοσελίδα του Ελληνικού Χρηματιστηρίου Ενέργειας
- β: Συντελεστής Προσαύξησης, ίσος με 0,0115 €/kWh
- L_u: Άνω όριο αναφοράς, ίσο με 0,050
- L_d: Κάτω όριο αναφοράς, ίσο με 0,040

και ισούται με:

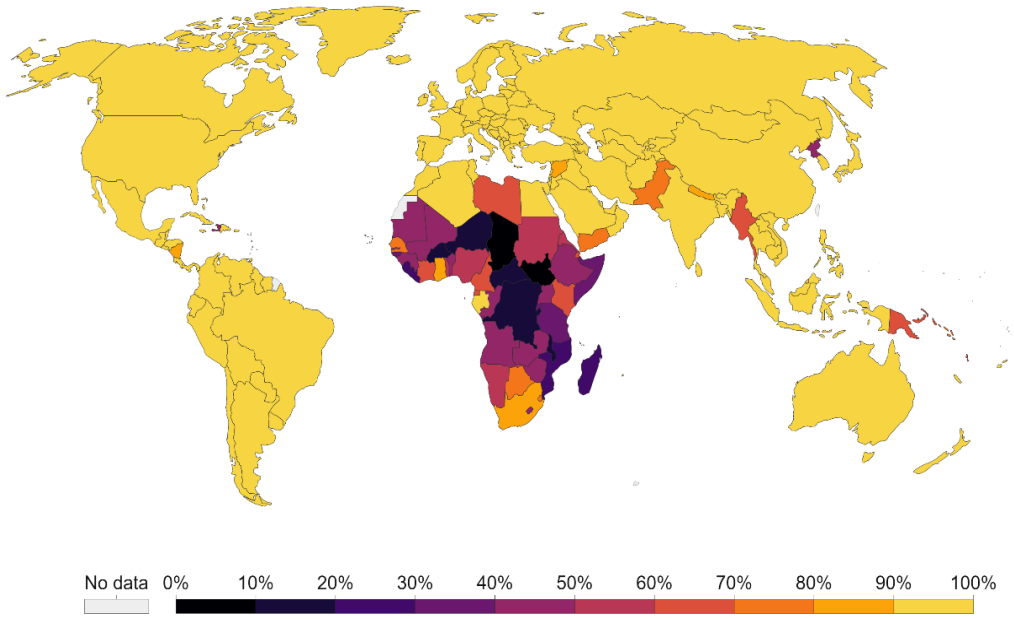
- Χρέωση ίση με Y-L_u, όταν το Y είναι μεγαλύτερο από το άνω όριο L_u
- Πίστωση ίση με Y-L_d όταν το Y είναι μικρότερο από το κάτω όριο L_d
- Μηδενική χρέωση όταν το Y βρίσκεται εντός των ορίων L_d και L_u

Για ένα μέσο νοικοκυριό με μέση μηνιαία κατανάλωση 300 kWh και την τιμή εκκαθάρισης αγοράς για τον Μάρτιο του 2022 να φθάνει τα 0,27 €/kWh, η ρήτρα αναπροσαρμογής υπολογίζεται στα 0,28 €/kWh μέσω της παραπάνω μεθόδου υπολογισμού. Αθροίζοντας τη βασική τιμή της κιλοβατώρας με την ρήτρα αναπροσαρμογής το κόστος ανέρχεται σε 0,39 €/kWh. Συμπληρώνοντας και τις ρυθμιζόμενες χρώσεις Σύστημα Μεταφοράς, Δίκτυο Διανομής, ΕΤΜΕΑΡ, ΥΚΩ) με 0,09 €/kWh, το συνολικό κόστος φτάνει τα 0,48 €/kWh. Συνολικά, λοιπόν, για κατανάλωση 300 kWh, ο καταναλωτής έρχεται αντιμέτωπος με μηνιαίο λογαριασμό ίσο με περίπου 145 €.

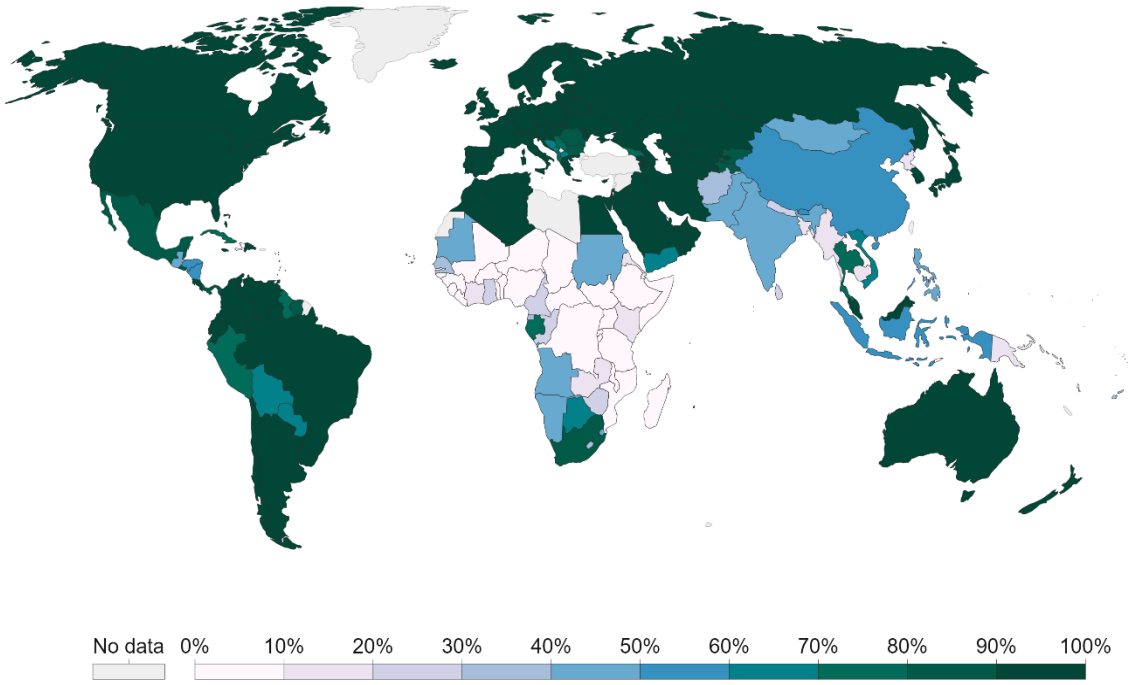
Γίνεται εύκολα αντιληπτό λοιπόν πως, όταν σε μία χώρα όπως η Ελλάδα, η οποία βρίσκεται ακόμη σε οικονομική κρίση και εξακολουθεί να εφαρμόζει πολιτικές λιτότητας, εμφανίζονται τέτοιες ενεργειακές δαπάνες για πολίτες με περιορισμένα εισοδήματα, οι περιπτώσεις καταναλωτών οι οποίοι θεωρούνται ενεργειακά φτωχοί, αναμένεται να πολλαπλασιαστούν.

1.3. Ενεργειακή φτώχεια

Όπως προαναφέρθηκε, το ενεργειακό πρόβλημα στις φτωχές και αναπτυσσόμενες χώρες σχετίζεται κυρίως, με την προσβασιμότητα στην ενέργεια και την έλλειψη πρόσβασης σε καθαρές εγκαταστάσεις μαγειρέματος. Στους παρακάτω χάρτες παρουσιάζεται η έλλειψη πρόσβασης σε ηλεκτρικό ρεύμα (Εικ. 7) και καθαρές μεθόδους μαγειρέματος (Εικ. 8) σε ποσοστό επί τοις εκατό του πληθυσμού κάθε χώρας.



Εικόνα 7: Ποσοστό πληθυσμού ανά χώρα με πρόσβαση σε ηλεκτρικό ρεύμα για το 2019. (Πηγή:Our World in Data, 2019)



Εικόνα 8: Ποσοστό πληθυσμού με πρόσβαση σε καθαρά καύσιμα μαγειρέματος για το 2019. (Πηγή: Our World in Data, 2019)

Οι διεθνείς στατιστικές μελέτες υιοθετούν μια πολύ χαμηλή βάση για το τι εστί πρόσβαση σε ηλεκτρικό ρεύμα: έχει καθοριστεί ως η πρόσβαση σε μία πηγή ηλεκτρικής ενέργειας η οποία μπορεί να προσφέρει βασικό φωτισμό και φόρτιση ενός κινητού ή ραδιόφωνού για 4 ώρες ανά ημέρα.

Η πρόσβαση σε καθαρά καύσιμα ή καθαρές τεχνολογίες όπως ηλεκτρικές εστίες μαγειρέματος μειώνουν σημαντικά την έκθεση σε ρύπους στην εσωτερική ατμόσφαιρα ενός νοικοκυριού, αιτία πολλών θανάτων σε νοικοκυριά με χαμηλό εισόδημα.

Όμως, το πρόβλημα της ενεργειακής φτώχειας διαφοροποιείται σημαντικά στις πιο εύπορες χώρες του ανεπτυγμένου κόσμου, στις οποίες το πρόβλημα της πρόσβασης σε ηλεκτρικό ρεύμα, πλέον, δεν υφίστανται. Η μορφή λοιπόν που λαμβάνει η ενεργειακή φτώχεια, αφορά το κόστος του ηλεκτρικού ρεύματος και των αναγκαίων υπηρεσιών από τις οποίες συνοδεύεται, από την οποία έλαβε και το όνομα του το υπό μελέτη φαινόμενο.

Η ενεργειακή φτώχεια στις ανεπτυγμένες χώρες είναι συνάρτηση των εξής τριών παραγώγων (Legendre & Ricci, 2015; Palmer et al., 2008; IEA, 2011) :

- Το χαμηλό εισόδημα ενός νοικοκυριού
- Το υψηλό κόστος ενέργειας
- Τη χαμηλή ενεργειακή απόδοση του σπιτιού.

Ο συνδυασμός, της οικονομικής κρίσης που μαστίζει την Ευρώπη-και κυρίως την Ελλάδα- με τις σκληρές πολιτικές λιτότητας, έχουν αναδείξει την ενεργειακή φτώχεια ως ένα από τα πιο κρίσιμα οικονομικά, κοινωνικά και περιβαλλοντικά προβλήματα της εποχής, με προεκτάσεις σε πολλούς

τομείς. Παρά τα όποια μέτρα λαμβάνονται από κυβερνήσεις και κεντρικές τράπεζες για βραχυπρόθεσμες μειώσεις στις τιμές του ρεύματος, οι ολοένα αυξανόμενες τιμές στα ορυκτά καύσιμα δείχνουν ότι το πρόβλημα αναμένεται να οξυνθεί στο εγγύς μέλλον (IEA, 2021).

Σύμφωνα με έρευνα της ΕΕ και της Κομισιόν για το 2019, 30 εκατομμύρια Ευρωπαίοι θεωρούνταν ενεργειακά φτωχοί (Jacques Delores Institute, 2021), καθώς δεν είχαν την οικονομική δυνατότητα να κρατήσουν τα νοικοκυριά τους σε άνετη θερμοκρασία διαβίωσης, ενώ για το 2020 αυτό το ποσό αυξήθηκε σε 35,8 εκατομμύρια Ευρωπαίους (8% του πληθυσμού)(European Commission, 2021). Ωστόσο, σύμφωνα με παλαιότερες ακόμη μελέτες (πχ. EPEE 2009$_a$; BPIE 2014), εκτιμήθηκε ότι η ενεργειακή φτώχεια μπορεί να πλήττει από 50 έως 125 εκατομμύρια ανθρώπους στην Ευρωπαϊκή Ένωση και σύμφωνα με την τωρινή κατάσταση στην αγορά ενέργειας αυτή η εκτίμηση αποδεικνύεται ακριβής, καθώς τον Ιούλιο του 2021 οι ενεργειακά φτωχοί Ευρωπαίοι έφτασαν τα 50 εκατομμύρια (ODYSSEE-MURE, 2021).

Βαθύτερη ανάλυση του ορισμού της ενεργειακής φτώχειας θα πραγματοποιηθεί στο κεφάλαιο 1.3.

1.4. Επιπτώσεις ενεργειακής φτώχειας

Η ενεργειακή φτώχεια μπορεί να επηρεάσει πολλούς τομείς της ανθρώπινης ζωής και να προκαλέσει σοβαρή ζημιά στο μοντέλο λειτουργίας του σύγχρονου κόσμου, θέτοντας σε κίνδυνο την ανθρώπινη ζωή, την τοπική και παγκόσμια οικονομία αλλά και το περιβάλλον. Οι κύριες επιπτώσεις, λοιπόν, αφορούν την ανθρώπινη υγεία, την μόλυνση του περιβάλλοντος και οικονομικές/κοινωνικές ζημιές (Παπαδά,2017).

1.4.1.Επιπτώσεις στην ανθρώπινη υγεία

Η ενεργειακή φτώχεια επιφέρει σημαντικά προβλήματα στην δημόσια υγεία. Αυτός ο κίνδυνος προκύπτει κυρίως με την ρύπανση του αέρα εσωτερικών χώρων ή αλλιώς οικιακή ατμοσφαιρική ρύπανση. Η συγκεκριμένη πτυχή των επιπτώσεων συνδέεται άμεσα με την αυξημένη θνησιμότητα το χειμώνα και τα κυκλοφορικά, αναπνευστικά ακόμη και ψυχικά νοσήματα (Healy 2004; BPIE 2014; Liddell & Morris 2010; Carrere et.al, 2021).

Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (WHO,2021), περίπου 2,6 δισεκατομμύρια άνθρωποι παγκοσμίως μαγειρεύουν χρησιμοποιώντας ρυπογόνες μεθόδους όπως, ανοιχτές εστίες φωτιάς ή απλούς φούρνους τροφοδοτούμενους με κηροζίνη, βιομάζα και κάρβουνο. Οι περισσότεροι εξ' αυτών των ανθρώπων είναι φτωχοί και ζουν σε χώρες χαμηλού ή μέσου εισοδήματος. Αυτές οι μέθοδοι μαγειρέματος είναι μη αποδοτικές και χρησιμοποιούν καύσιμα και τεχνολογίες, που παράγουν υψηλά επίπεδα οικιακής ατμοσφαιρικής ρύπανσης με μεγάλο εύρος επικίνδυνων ρύπων στη σύσταση τους, συμπεριλαμβανομένων και των μικρών σωματιδίων αιθάλης που διεισδύουν βαθιά στους πνεύμονες. Σε ανεπαρκώς αεριζόμενες κατοικίες, ο εσωτερικός καπνός μπορεί να είναι έως και 100 φορές υψηλότερος από τα αποδεκτά

επίπεδα. Η έκθεση είναι ιδιαίτερα υψηλή ανάμεσα σε γυναίκες και μικρής ηλικίας παιδιά, που ξοδεύουν τον περισσότερο χρόνο τους καθημερινά κοντά σε εστίες θέρμανσης ή μαγειρέματος (WHO, 2021).

Κάθε χρόνο περισσότεροι από 3,8 εκατομμύρια άνθρωποι πεθαίνουν πρόωρα από ασθένειες που οφείλονται σε οικιακή ατμοσφαιρική ρύπανση από μη αποτελεσματικές μεθόδους μαγειρέματος. Μεταξύ αυτών των θανάτων(WHO, 2021):

- 27% οφείλεται σε πνευμονία
- 18% σε εγκεφαλικό επεισόδιο
- 27% σε ισχαιμικές καρδιακές νόσους
- 20% σε χρόνιες αποφρακτικές πνευμονοπάθειες
- 8% σε καρκίνο των πνευμόνων

- **Πνευμονία**

Έκθεση σε οικιακή ατμοσφαιρική ρύπανση σχεδόν διπλασιάζει το ρίσκο για παιδική πνευμονία και είναι υπεύθυνη για % όλων των θανάτων από πνευμονία σε παιδιά κάτω των 5 ετών. Η οικιακή ατμοσφαιρική ρύπανση αποτελεί, επίσης, αιτία οξειών λοιμώξεων του κάτω αναπνευστικού για τους ενήλικες και συνεισφέρει στο 28% όλων των θανάτων ενηλίκων από πνευμονία (WHO, 2021).

- **Χρόνιες αποφρακτικές πνευμονοπάθειες**

Το 25% των θανάτων από χρόνιες αποφρακτικές πνευμονοπάθειες ενηλικών σε χώρες χαμηλού και μέσου εισοδήματος οφείλονται στην οικιακή ατμοσφαιρική ρύπανση. Γυναίκες που εκτίθενται σε υψηλά επίπεδα εσωτερικού καπνού έχουν τις διπλάσιες πιθανότητες να παρουσιάσουν χρόνια αποφρακτική πνευμονοπάθεια, σε σχέση με τις γυναίκες που χρησιμοποιούν καθαρά καύσιμα και τεχνολογίες. Μεταξύ αντρών (οι οποίοι ήδη έχουν υψηλότερο ρίσκο να εμφανίσουν χρόνια αποφρακτική πνευμονοπάθεια λόγω υψηλοτέρων ποσοστό καπνιστών), η έκθεση σε οικιακή ατμοσφαιρική ρύπανση, διπλασιάζει το ρίσκο (WHO, 2021).

- **Εγκεφαλικό επεισόδιο**

12% των θανάτων από εγκεφαλικό επεισόδιο μπορούν να αποδοθούν σε καθημερινή έκθεση σε οικιακή ατμοσφαιρική ρύπανση, η οποία προέρχεται από μαγείρεμα με στερεά καύσιμα και κηροζίνη (WHO, 2021).

- **Ισχαιμικές καρδιακές νόσοι**

Περίπου 11% όλων των θανάτων από ισχαιμικές καρδιακές νόσους με πάνω από 1 εκατομμύριο πρόωρους θανάτους ετησίως, μπορεί να αποδοθεί σε έκθεση σε οικιακή ατμοσφαιρική ρύπανση (WHO, 2021).

- **Καρκίνος του πνεύμονα**

Περίπου 17% των θανάτων από καρκίνο των πνευμόνων για ενήλικες αποδίδονται σε έκθεση σε καρκινογόνα προερχόμενα από οικιακή ατμοσφαιρική ρύπανση, τα οποία προκαλούνται από μαγείρεμα με κηροζίνη ή με στερεά καύσιμα ξύλα και κάρβουνα. Το ρίσκο για τις γυναίκες είναι υψηλότερο, εξαιτίας του ρόλου τους στην προετοιμασία του φαγητού ενός, 2021).

<u>Άλλες επιπτώσεις στην ανθρώπινη υγεία</u>

Σε γενικότερη βάση, μικρά αιωρούμενα σωματίδια και άλλοι ρύποι ερεθίζουν τους αεραγωγούς και τους πνεύμονες , ενώ ακόμη αποδυναμώνουν το ανοσοποιητικό σύστημα και μειώνουν την ικανότητα του αίματος να μεταφέρει οξυγόνο. Υπάρχουν ακόμη στοιχεία που συνδέουν την οικιακή ατμοσφαιρική ρύπανση με το χαμηλό βάρος γέννησης, περιστατικά φυματίωσης, καταρράκτη, καρκίνο του ρινοφάρυγγα και του φάρυγγα (WHO, 2021).

1.4.2 Επιπτώσεις στην ποιότητα ζωής

Στις αναπτυσσόμενες και φτωχές χώρες , η συλλογή της απαιτούμενης για θέρμανση και μαγείρεμα βιομάζα πραγματοποιείται κυρίως από γυναίκες και παιδιά. Ο χρόνος που απαιτείται για τη διαδικασία συλλογής των καυσίμων περιορίζει τις γυναίκες από τη συμμετοχή τους σε άλλες βιοποριστικές δραστηριότητες, ενώ εμποδίζει παιδιά από το να συμμετέχουν ακόμη και σε στοιχειώδη εκπαίδευση, καθώς ακόμη κι παιδιά πολύ μικρής ηλικίας συμμετέχουν στη συλλογή καύσιμης βιομάζας. Αξίζει, επίσης να σημειωθεί, ότι λόγω της πολύ χαμηλής απόδοσης αυτών τον καυσίμων , είναι απαραίτητη η συλλογή μεγάλων ποσοτήτων για την εξασφάλιση επαρκούς θέρμανσης και συνθηκών μαγειρέματος. Συμπληρωματικά, σε λιγότερο ασφαλή περιβάλλοντά όπως της Υποσαχάριας Αφρικής, οι γυναίκες και τα παιδιά εκτίθενται σε σοβαρό κίνδυνο τραυματισμού ή και θανάτου, καθώς σε αυτές τις χώρες η διαδικασία συλλογής βιομάζας αποτελεί ένα ανταγωνιστικό περιβάλλον μεταξύ των συλλογέων, αφού ουσιαστικά αποτελεί έναν αγώνα επιβίωσης για αυτούς, με αποτέλεσμα η βία να είναι σχεδόν καθημερινό φαινόμενο (WHO, 2021).

Επίσης, έχει αποδειχθεί ότι η ενεργειακή φτώχεια μπορεί να επηρεάσει την ανάπτυξη των παιδιών. Οι ανεπαρκείς συνθήκες στέγασης, πέραν της επίδρασης που μπορεί να έχουν στη διατροφή των παιδιών, μπορούν να επηρεάσουν την συναισθηματική υγεία και το μορφωτικό/πνευματικό επίπεδο. Ενώ, με βελτιωμένες συνθήκες στέγασης , πέραν των βασικών ωφελειών, βελτιώνονται τις επιδόσεις των παιδιών στο σχολείο (Sule et al..2022).

Στις ανεπτυγμένες χώρες, τα νοικοκυριά που είναι αναγκασμένα να ξοδεύουν μεγάλό μέρος του εισοδήματος τους σε δαπάνες ενέργειας είναι αναγκασμένα να περιορίσουν άλλες βασικές ανάγκες. Μία εκ των αναγκών που τείνουν να περιορίζουν τα περισσότερα νοικοκυριά είναι η σωστή διατροφή, φαινόμενο γνωστό ως ''heat or eat''. Χαρακτηριστικό παράδειγμα του φαινομένου αποτελεί η απότομη αύξηση των ανθρώπων στη Γαλλία οι οποίοι καταφεύγουν σε τράπεζες τροφίμων όπως η '' Banques Alimentaires'', κατά 30% από την έναρξη της πανδημίας. Παρόμοια εικόνα επικρατεί σε όλη την Ευρώπη και ειδικότερα στη Βουλγαρία, όπου περισσότερο από το 25% του πληθυσμού δεν έχει την οικονομική δυνατότητα να διατηρεί το σπίτι αρκετά ζεστό χωρίς να περικόπτει χρήματα από άλλες πτυχές της ζωής τους.

1.4.3 Επιπτώσεις στο περιβάλλον

Το μεθάνιο και το διοξείδιο του άνθρακα που εκλύονται από την μη αποδοτική καύση στερεών καυσίμων σε σόμπες ή ανοιχτές εστίες είναι ρύποι που δρουν ως αέρια του θερμοκηπίου και επιδρούν έντονα στην κλιματική αλλαγή (WHO, 2021) . Ωστόσο, εξίσου σημαντική είναι και η επίδραση της ενεργειακής φτώχειας στην αλλαγή χρήσεων γης. Όπως ειπώθηκε και παραπάνω, η συλλογή βιομάζας προς καύση είναι η κυριότερη πηγή ενέργειας των νοικοκυριών στις φτωχότερες χώρες του κόσμου. Η υπερεκμετάλλευση αυτών των πόρων έχει αυξήσει δραστικά την αποψίλωση, την ερημοποίηση και την υποβάθμιση πολλών εδαφών, ωστόσο όπως είναι φυσικό, δεν αποτελεί την κυριότερη πηγή κινδύνου για τα δάση. Υπάρχει, λοιπόν, μια αξιόλογη απειλή λόγω της υπερεκμετάλλευσης , καθώς μειώνεται η χωρητικότητα των δασών για δέσμευση διοξειδίου του άνθρακα της ατμόσφαιρας, η οποία δρα υπέρ της κλιματικής αλλαγής(Gonzalez-Eguino, 2015).

1.4.4. Επιπτώσεις στην οικονομία

Η ενεργειακή φτώχεια επηρεάζει όλους τους παραγωγικούς τομείς και περιορίζει τις δυνατότητες ανάπτυξης. Για παράδειγμα στη γεωργία, έναν ουσιώδη τομέα των περισσότερων οικονομιών, η εισροή ενέργειας στις φτωχές χώρες είναι πολύ χαμηλή και προέρχεται κυρίως από την εργασία ζώων και ανθρώπου. Αντίθετα, στις πλούσιές χώρες υπάρχουν υψηλά επίπεδα άμεσων ενεργειακών εισροών (μηχανήματα και καύσιμα), καθώς και έμμεσων εισροών (χημικά και λιπάσματα). Στις ΗΠΑ, για παράδειγμα, τα λιπάσματα με βάση το άζωτο αντιπροσωπεύουν το 45% του συνόλου των ενεργειακών εισροών κατά τη διαδικασία παραγωγής καλαμποκιού, ενώ η φυσική εργασία, αντιπροσωπεύει μόλις το 3% (Παπαδά, 2017)

Από την άλλη πλευρά, ακόμη και μικρές βελτιώσεις στην πρόσβαση και την κατανάλωση ενέργειας μπορούν να έχουν σημαντικό αντίκτυπο. Στην εκπαίδευση, για παράδειγμα, οι στατιστικές δείχνουν ότι οι πληθυσμοί με υψηλότερα επίπεδα πρόσβασης στην ηλεκτρική ενέργεια και καλύτερο φωτισμό δρόμων έχουν υψηλότερα επίπεδα αλφαβητισμού, χαμηλότερα

ποσοστά εγκατάλειψης και αφιερώνουν περισσότερο χρόνο στο διάβασμα και τη μελέτη (Khandler et al.,2014).

Επιπροσθέτως, αν και η ιδέα ίσως μοιάζει ακόμη μακρινή για τις φτωχές χώρες, η πρόσβαση στις τεχνολογίες της πληροφορίας και της επικοινωνίας θα μπορούσε να ενθαρρύνει τη δημιουργία νέων επιχειρήσεων, την δωρεάν πρόσβαση σε ποιοτικά μαθήματα κατάρτισης στο διαδίκτυο και να ενθαρρύνει την σταδιακή βελτίωση της κοινωνίας (Gonzalez-Eguino, 2015).

1.5.Η εξέλιξη της έρευνας στον ορισμό και τον υπολογισμό της ενεργειακής φτώχειας

Σε ευρωπαϊκό επίπεδο, σε σχετικές μελέτες και σε επίσημες ανακοινώσεις Ευρωπαϊκών οργανισμών (όπως η Ευρωπαϊκή Επιτροπή), επικρατεί η χρήση δύο διαφορετικών ορολογιών για την περιγραφή του φαινομένου: «fuel poverty» και «energy poverty». Ο όρος «fuel poverty» εισήχθη στο Ηνωμένο Βασίλειο και αποτελούσε μια ''υποκατηγορία '' του φαινομένου, η οποία αφορούσε τη διασφάλιση επαρκούς θέρμανσης στο σπίτι, μέσω επαρκούς χρήσης καυσίμων. Στην πορεία όμως ο όρος αμφισβητήθηκε λόγω της σημασίας που έδινε στο κομμάτι της θέρμανσης, ενώ υποβάθμιζε σημαντικά το ρόλο των υπόλοιπων οικιακών ενεργειακών χρήσεων, τον οποία καλύπτει πλήρως ο όρος «energy poverty». Η σημερινή έννοια της ενεργειακής φτώχειας συμπεριλαμβάνει όλες τις δυνατές ενεργειακές χρήσεις ενός νοικοκυριού. Παρόλα αυτά, οι δύο έννοιες συνεχίζουν να χρησιμοποιούνται εκτεταμένα, αναφερόμενες πλέον, στο ίδιο περιεχόμενο, με τον όρο «energy poverty» να αποτελεί την προτιμώμενη ορολογία των επιστημονικών μελετών για την ενεργειακή φτώχεια (Παπαδά,2017).

Το σύνολο των ορισμών και των τρόπων υπολογισμού της ενεργειακής φτώχειας, από την αρχή μελέτης του φαινομένου έως και σήμερα μπορούν να ταξινομηθούν σε τέσσερις ευρείες κατηγορίες:

- 1η κατηγορία: Θεωρητικές αναφορές για τον ορισμό και τον υπολογισμό της ενεργειακής φτώχειας
- 2η κατηγορία: Αναφορές με χρήση ποσοτικών δεικτών για τον υπολογισμό της ενεργειακής φτώχειας
- 3η κατηγορία: Αναφορές που εισάγουν τη χρήση υποκειμενικών δεικτών για την εκτίμηση της ενεργειακής φτώχειας
- 4η κατηγορία: Αναφορές που μελετούν την ενεργειακή φτώχεια από μια ευρύτερη πλευρά τρωτότητας

Οι παραπάνω κατηγορίες δεν είναι ανεξάρτητες μεταξύ τους, πχ. οι θεωρητικοί ορισμοί (1η κατηγορία) οδηγούν σε μεθόδους ποσοτικοποίησης του φαινομένου (2η Κατηγορία). Η κατηγοριοποίηση που πραγματοποιείται, χρησιμεύει σε αποδοτικότερη παρακολούθηση της έρευνας. Η ανάλυση που ακολουθεί, πραγματοποιείται σύμφωνα με τη χρονολογική σειρά εξέλιξης της έρευνας του αντικειμένου(Παπαδά,2017).

- **1η κατηγορία**

Οι πρώτες έρευνες που ασχολήθηκαν με το ζήτημα της ενεργειακής φτώχειας ξεκινούν τη δεκαετία του ’80, με τους πρώτους ορισμούς να βασίζονται στην λογική ‘’της επαρκούς θέρμανσης ενός νοικοκυριού και στην ‘’επαρκή ζέστη’’ (Lewis, 1982; Bradshaw & Hutton 1983; Boardman 1986), με ποσοτικοποίηση των νοικοκυριών που δεν επιτυγχάνουν τις ελάχιστες προδιαγραφές θέρμανσης (Παπαδά, 2017).

Το 1991, η Έκθεση της Αγγλικής Έρευνας Συνθηκών Στέγασης (English House Condition Survey – EHCS) , υιοθετήθηκε το όριο του 10% του εισοδήματος ως ‘’πάτωμα’’ του ενεργειακού κόστους (DOE 1996), με μία σημαντική διόρθωση. Αντί του πραγματικού ενεργειακού κόστους , εισήχθη η έννοια της ‘’απαιτούμενης δαπάνης’’ για την κάλυψη των ενεργειακών αναγκών. Η βασική διαφορά που προσέφερε αυτή η διόρθωση αφορά την εισαγωγή της ικανοποιητικής θέρμανσης στον ορισμό. Έτσι, ένα νοικοκυριό δε θεωρούταν πλέον ενεργειακά φτωχό όταν ξοδεύει πάνω από το 10% του εισοδήματος του για καύσιμα, άλλα όταν χρειάζεται να ξοδέψει πάνω από 10 % του εισοδήματος του, προκειμένου να επιτύχει ικανοποιητικό επίπεδο θέρμανσης για την υγεία και τη θερμική άνεση, συμπεριλαμβανομένων των απαιτήσεων για κατάλληλο φωτισμό, μαγείρεμα και οικιακές συσκευές (Moore, 2012;EPEE 2009_b) (Παπαδά, 2017).

Το 1996, η αντίστοιχη Έκθεση της Αγγλικής Έρευνας ενίσχυσε την παραπάνω διόρθωση, εισάγοντας την έννοια της ‘’απαιτούμενης ικανοποιητικής θέρμανσης’’ (DETR 2000). Οι εκτιμήσεις της ενεργειακής φτώχειας βάσει της σχετικής έρευνας (EHCS) του 1996 αποτέλεσαν τη βάση στην οποία στηρίχθηκε η Στρατηγική για την Ενεργειακή Φτώχεια του Ηνωμένου Βασιλείου το 2001 (UK Fuel Poverty Strategy), με την οποία εισήχθη ουσιαστικά ο πρώτος επίσημος ορισμός για την ενεργειακή φτώχεια (Παπαδά, 2017). Ο ορισμός αυτός υιοθετήθηκε και από τη Σκωτία το 2001, συμπεριλαμβάνοντας σε αυτό, , όλες τις ενεργειακές χρήσεις ενός νοικοκυριού και είναι ο εξής:

Ένα νοικοκυριό βρίσκεται σε κατάσταση ενεργειακής φτώχειας εάν, προκειμένου να διατηρήσει ένα ικανοποιητικό επίπεδο θέρμανσης, θα απαιτούνταν να ξοδέψει πάνω από το 10% του εισοδήματός του (συμπεριλαμβανομένου του στεγαστικού επιδόματος της ενίσχυσης εισοδήματός για την αποπληρωμή στεγαστικού δανείου) για όλες τις οικιακές χρήσεις καυσίμων/ενέργειας. Όταν για αυτές απαιτείται πάνω από 20% του εισοδήματος, η κατάσταση αποκαλείται ‘’κατάσταση ακραίας ενεργειακής φτώχειας’’.

Το ικανοποιητικό επίπεδο θέρμανση ορίζεται ως εξής: Για τα ευάλωτα νοικοκυριά (ηλικιωμένους και ασθενείς ομάδες) 23°C στο σαλόνι και 18°C στα άλλα δωμάτια. Για τα υπόλοιπα νοικοκυριά, 21 ^{o}C στο σαλόνι και 18 ^{o}C στα άλλα δωμάτια.

Ο αντικειμενικός αυτός ορισμός της ενεργειακής φτώχειας επισημοποιήθηκε το 1998, όταν η τότε κυβέρνηση υιοθέτησε τον όρο «ενεργειακή φτώχεια» (fuel poverty) και αποδέχτηκε το όριο του 10% ως όριο απαιτούμενων δαπανών ενέργειας πάνω από το οποίο ένα νοικοκυριό χαρακτηρίζεται ως ενεργειακά φτωχό, και χαρακτήρισε το 10% ως επίσημο δείκτη υπολογισμού της ενεργειακής φτώχειας (EPEE, 2009_b).

Με τη πάροδο του χρόνου, και πέραν του επίσημου ορισμού από το Ηνωμένο Βασίλειο, νέοι ορισμοί για την ενεργειακή φτώχεια έχουν προταθεί από την επιστημονική κοινότητα,

ενισχύοντας ή αναιρώντας προηγούμενες προσεγγίσεις. Σύμφωνα με τον Healy (2004), η ενεργειακή φτώχεια προκαλείται από ένα συνδυασμό παραμέτρων όπως το χαμηλό εισόδημα και η ενεργειακή ανεπάρκεια του σπιτιού, ενώ παράλληλα διαφέρει από τη γενικότερη φτώχεια, η οποία μπορεί να αντιμετωπιστεί πιο εύκολα, με ενίσχυση του εισοδήματος. Μία ακόμη κοινωνική διάσταση του φαινομένου εισήχθη από τον Buzar (2007), ο οποίος ορίζει την ενεργειακή φτώχεια ως κατάσταση κατά την οποία ένα νοικοκυριό δεν είναι σε θέση να έχει πρόσβαση σε υπηρεσίες ενέργειας στο σπίτι, σε ένα κοινωνικά και υλικά επαρκές επίπεδο (Παπαδά, 2017).

Σύμφωνα με τον Tennakoon (2008), ο ορισμός της ενεργειακής φτώχειας συνδέεται άμεσα με τον τρόπο που αυτή υπολογίζεται, καθώς υπάρχει πληθώρα τρόπων μέτρησης του φαινομένου, όπως για παράδειγμα με βάση το ενεργειακό κόστος ενός νοικοκυριού, τα ποσά ενέργειας που καταναλώνει ένα νοικοκυριό, η πρόσβαση σε υπηρεσίες ενέργειας κ.ά.. Οι διαφορές μεταξύ των τρόπων μέτρησης έχουν προκύψει εξαιτίας της έλλειψης διεθνούς συμφωνίας για έναν αποδεκτό μέτρο αναφορικά με την απαιτούμενη ενέργεια για την κάλυψη των βασικών ανθρώπινων αναγκών (Pachauri et al., 2004). Ο οργανισμός Practical Action (2009) ορίζει την ενεργειακή φτώχεια ως την έλλειψη επαρκούς σύγχρονης μορφής ενέργειας για τις βασικές ανάγκες του μαγειρέματος, της θέρμανσης, του φωτισμού, καθώς και την έλλειψη των βασικών υπηρεσιών ενέργειας για τα σχολεία, τα κέντρα υγείας και την παραγωγή εισοδήματος (Παπαδά,2017).

Νεότερος ορισμός από την Ευρωπαϊκή Οικονομική και Κοινωνική Επιτροπή (ΕΟΚΕ 2013) ορίζει την ενεργειακή φτώχεια ως (Γνωμοδότηση TEN/420):

Τη δυσκολία ή αδυναμία ενός νοικοκυριού να εξασφαλίσει επαρκή θέρμανση στην κατοικία του και να έχει πρόσβαση σε άλλες βασικές υπηρεσίες ενέργειας σε λογική τιμή.

Αξίζει, επίσης, να επισημανθεί ότι η ενεργειακή φτώχεια πλήττει την πλειονότητα των ατόμων που βρίσκονται σε κατάσταση γενικής φτώχειας, καθώς οι δυσμενείς συνθήκες επιδρούν μεταξύ τους και επιδεινώνουν, τις συνθήκες διαβίωσης αυτών των ανθρώπων.

- **2η Κατηγορία**

Το Ηνωμένο Βασίλειο αποτέλεσε την πρώτη χώρα με επίσημο, αντικειμενικό ορισμό του φαινομένου. Ο αντικειμενικός δείκτης υπολογισμού της ενεργειακής φτώχειας βασίστηκε στον ήδη θεσμοθετημένο ορισμό της Στρατηγικής για την Ενεργειακή Φτώχεια του 2001 και αποτυπώθηκε με μαθηματικό τύπο το 2009, στο πλαίσιο εκθέσεων ενεργειακής φτώχειας για Ηνωμένο Βασίλειο ως εξής (DECC, 2015):

$$\text{Fuel Poverty Ratio} = \frac{\text{Modelled fuel costs (i.e., modelled consumption x price)}}{\text{Income}} > 10\%$$

Στη μαθηματική έκφραση του δείκτη, το εισόδημα συμπεριλαμβάνει όλες τις πηγές εισοδήματος και επιδομάτων, απαλλαγμένα από φόρους και ασφαλιστικές εισφορές (DECC, 2013). Η ενεργειακή κατανάλωση περιλαμβάνει τη θέρμανση, το ζεστό νερό χρήσης, το μαγείρεμα, το φωτισμό και τις ηλεκτρικές συσκευές (DECC, 2009). Το ''ικανοποιητικό επίπεδο θέρμανσης'' ορίζεται με διατήρηση θερμοκρασίας 21 ^{o}C στο σαλόνι και στους 18 ^{o}C στα άλλα δωμάτια (DECC, 2015).

Ο Hills (2011) πρότεινε ως ενεργειακά φτωχά τα νοικοκυριά εκείνα των οποίων το ισοδύναμο εισόδημα, χωρίς το κόστος στέγασης και το οικιακό ενεργειακό κόστος, βρίσκονται υπό του 60% του ισοδύναμου εθνικού διαμέσου εισοδήματος, απαλλαγμένο από το κόστος στέγασης και το ενεργειακό οικιακό κόστος. Η μαθηματική προσέγγιση που σχηματίστηκε για αυτήν την έκφραση είναι ο εξής (DECC, 2015):

Fuel Poverty = Equivalized (Income – Housing costs – Domestic fuel costs) < 60% Equivalized (National Median Income – Housing costs – Domestic Fuel Costs)

Ο παραπάνω ορισμός ωστόσο, κατατάσσει σχεδόν όλα τα νοικοκυριά με χαμηλό εισόδημα ως ενεργειακά φτωχά, ανεξάρτητα από τις ενεργειακές τους δαπάνες. Αυτή η προσέγγιση αντιτίθεται κάθετα σε όποια άλλη προσέγγιση έχει γίνει για την ενεργειακή φτώχεια, και φαίνεται να προσαρμόζεται καλύτερα στην γενική έννοια της φτώχειας (Παπαδά,2017).

Η Αγγλία υιοθέτησε το 2013, ένα νέο δείκτη για την μέτρηση της ενεργειακής φτώχειας αντικαθιστώντας το δείκτη 10%. Ο νέος δείκτης ''Low Income/ High Costs- LIHC'' εισήχθη από τον Hills το 2012. Ο δείκτης ορίζει ως ενεργειακά φτωχό ένα νοικοκυριό, αν έχει απαιτούμενα ενεργειακά κόστη πάνω από το εθνικό διάμεσο επίπεδο και, αν αφού ξοδέψει το ποσό αυτό, απομείνει με ένα εισόδημα κάτω από το όριο της φτώχειας (Παπαδά, 2017).

Ο δείκτης 10% παραμένει ο πιο ευρέως χρησιμοποιούμενος δείκτης και έχει καθιερωθεί ως γενικός δείκτης υπολογισμού της ενεργειακής φτώχειας σε μία χώρα ή περιοχή (Atsalis, 2016). Ο δείκτης αυτός, όμως, παρουσιάζει δυσκολία στον υπολογισμό της απαιτούμενης ενεργειακής κατανάλωσης, όπως απαιτεί ο επίσημος ορισμός και έχει αναγκάσει σχεδόν όλη την επιστημονική κοινότητα να χρησιμοποιεί κατά τους υπολογισμούς την πραγματοποιθείσα (actual) κατανάλωση αντί της απαιτούμενης (required), καθώς η συλλογή στοιχείων για την πρώτη περίπτωση είναι εφικτή, μέσω των εθνικών στατιστικών ερευνών ή μέσω πρωτογενούς έρευνας (Παπαδά, 2017).

Η μοναδική χώρα η οποία διαθέτει ειδικό πρόγραμμα μοντελοποίησης της απαιτούμενης ενεργειακής κατανάλωσης ανά νοικοκυριό, προσαρμοσμένο στα δεδομένα της χώρας είναι η Αγγλία. Η συγκεκριμένη μοντελοποίηση είναι αδύνατο να υιοθετηθεί στα δεδομένα των υπόλοιπων χωρών, με αποτέλεσμα, ο δείκτης 10% πραγματοποιηθείσας κατανάλωσης να είναι ευρέως καθιερωμένος ως συμβατικός αντικειμενικός δείκτης ενεργειακής φτώχειας σε ευρωπαϊκό επίπεδο (Roberts et al. 2015, EC, 2010).

Αναφέρεται συχνά στη διεθνή επιστημονική κοινότητα ότι, ο συμβατικός δείκτης 10% υποεκτιμά την πραγματική διάσταση του προβλήματος, καθώς αποτυγχάνει να εντοπίσει τα νοικοκυριά εκείνα που, λόγω οικονομικής αδυναμίας, μειώνουν την ενεργειακή κατανάλωση στο σπίτι (Dubois, 2012; Fahmy et al., 2011; Legendre & 2015; Thomson, 2013), και επομένως, θεωρείται ένας ''φτωχός'' δείκτης της ενεργειακής φτώχειας (Moore, 2012; Liddell et al., 2012). Έχει επίσης αναφερθεί ότι, ως δείκτης είναι ''ευαίσθητος'' στη μεταβολή της τιμής των καυσίμων σε σχέση με την επίδραση των άλλων δύο παραγόντων (ενεργειακή κατανάλωση, εισόδημα) (DECC 2013), με αποτέλεσμα οι μεταβολές της τιμής να είναι αυτές που ''κατευθύνουν'' το δείκτη. Η πρώτη αδυναμία διορθώνεται με την εισαγωγή της απαιτούμενης ενεργειακής κατανάλωσης. Ωστόσο και η χρήση της απαιτούμενης κατανάλωσης έχει δεχθεί κριτική, με τους (Hills 2011, Moore 2012) να επισημαίνουν πως με τη χρήση της απαιτούμενης κατανάλωσης, μεγάλος αριθμός νοικοκυριών υψηλού εισοδήματος θα καταγράφονται ως ενεργειακά φτωχά νοικοκυριά, εφόσον δεν έχει προσδιοριστεί ένα ανώτατο όριο απαιτούμενης κατανάλωσης για τις περιπτώσεις άσκοπης χρήσης ενέργειας (Παπαδά, 2017).

Ο δείκτης «LIHC» του Hills επιδίωξε να διορθώσει την αδυναμία που προκύπτει από τη χρήση της απαιτούμενης κατανάλωσης, συνδυάζοντας τη διάσταση της φτώχειας με το υψηλό ενεργειακό κόστος. Ωστόσο, η εφαρμογή του ορισμού έχει χαρακτηριστεί ως πολύπλοκη, και σε μεγάλο βαθμό μη διαφανής (Moore, 2012), ενώ παράλληλα, παραμένει αμετάβλητη από τις διακυμάνσεις στην τιμή των καυσίμων (Moore 2012; Roberts et al., 2015).

Πιο συγκεκριμένα, έχει αναφερθεί ότι ο δείκτης LHIC υπολογίζει τα ενεργειακά φτωχά νοικοκυριά στο χρόνο, χωρίς να επηρεάζεται σε έντονο βαθμό από το εισόδημα ή την τιμή των καυσίμων σε σχέση με το δείκτη 10% -είτε τον συμβατικό, είτε τον επίσημο- (DECC, 2013). Συμπερασματικά, μια αύξηση στο εισόδημα ή μια μείωση στην τιμή των καυσίμων είναι πιθανό να μειώσει την έκταση της ενεργειακής φτώχειας σύμφωνα με τον δείκτη 10%, αλλά σύμφωνα με τον δείκτη LIHC, οι ίδιες μεταβολές δεν θα επηρεάσουν τον δείκτη. Συνεπώς κατά τον δείκτη «LIHC», η έκταση της ενεργειακής φτώχειας θα παραμείνει η ίδια, εκτός αν οι μεταβολές στα υπό μελέτη νοικοκυριά είναι μεγαλύτερες από τις μέσες εθνικές μεταβολές (Παπαδά, 2017).

Αναφορικά με το όριο 10% η επιστημονική κοινότητα εκφράζει δυσπιστία ως προς την εγκυρότητα χρήσης του συγκεκριμένου ορίου (Lawson et al., 2015), καθώς υποστηρίζεται ότι η μπορεί να είναι προβληματική σε περιπτώσεις μελετών μεγάλης κλίμακας (ΕΟΚΕ, 2013), και το όριο θα πρέπει να διαφοροποιείται σε εθνικό επίπεδο, ανάλογα με τις ενεργειακές ανάγκες και την οικονομική κατάσταση κάθε χώρας (Παπαδά, 2017).

Στα πλαίσια της διεθνούς έρευνας , έχουν παρουσιαστεί και άλλοι δείκτες μέτρησης της ενεργειακής φτώχειας στην διεθνή βιβλιογραφία και είναι οι εξής:

- Δείκτης αυξημένης θνησιμότητας στη διάρκεια του χειμώνα (Excess Winter Death Index) ή δείκτης εποχικής μεταβλητότητας της θνησιμότητας (Index of Seasonal Mortality Variability) (Healy, 2004; BPIE, 2014; Πάνας, 2012).

- Κτιριακός Δείκτης Ενεργειακής Φτώχειας (Building Fuel Poverty Index), σχετίζεται με την ενεργειακή απόδοση των κτιρίων, με σκοπό την απόκτηση δυνατότητας προσδιορισμού μη ενεργειακά αποδοτικών κτιρίων που απαιτούν άμεση επέμβαση (Fabbri, 2015).

- Ενεργειακή Τρωτότητα (παραλλαγή ορισμού Hills, 2011). Σύμφωνα με αυτό τον ορισμό, τα νοικοκυριά είναι ενεργειακά ευάλωτα όταν δεν είναι εκ των προτέρων φτωχά (το ισοδύναμο εισόδημα, απαλλαγμένο από το κόστος στέγασης, δεν είναι μικρότερο του 60 % του ισοδύναμου εθνικού διάμεσου εισοδήματος), αλλά θεωρούνται φτωχά όταν συμπεριληφθεί στις δαπάνες τους, το ενεργειακό κόστος. Όταν δηλαδή, ο παράγοντας του ενεργειακού κόστους είναι ο κρίσιμος παράγοντας που υποβαθμίζει τα νοικοκυριά σε κατάσταση φτώχειας (Παπαδά, 2017).

- Δείκτης ενεργειακής φτώχειας βάσει Απαιτούμενου κόστους (Energy Poverty Index based on Required cost – EpI_{RC}). Ο δείκτης αυτός στηρίζεται στον υπολογισμό των απαιτούμενων αναγκών ενεργειακής κατανάλωσης μέσω εφαρμογής του Στοχαστικού Μοντέλου Ενεργειακής Φτώχειας (Stochastic Model of Energy Poverty-SMEP) (Παπαδά, 2017). Το συγκεκριμένο μοντέλο αναπτύχθηκε με σκοπό την αντιμετώπιση της αδυναμίας που εμφανίζει ο συμβατικός δείκτης 10% βάσει των πραγματοποιηθέντων αναγκών , κατά την οποία υποεκτιμούνται οι πραγματικές ενεργειακές ανάγκες των νοικοκυριών. Η μαθηματική εξίσωση εστιάζει στον υπολογισμό των απαιτούμενων καταναλώσεων για θέρμανση και ψύξη μέσω της μεθόδου των βαθμοημερών και εκτιμά τα πραγματικά επίπεδα της ενεργειακής φτώχειας ποσοτικοποιώντας την επίδραση των διαφόρων παραμέτρων που την επηρεάζουν (Παπαδά, 2017; Papada & Kaliampakos, 2018). Ο συγκεκριμένος δείκτης αποτελεί έναν από τους τρεις δείκτες που θα υιοθετηθούν για τη μοντελοποίηση των δεδομένων στην συγκεκριμένη διπλωματική εργασία, ώστε να εκτιμηθεί η ακρίβεια που εμφανίζει σε σχέση με τους άλλους χρησιμοποιούμενους δείκτες.

EPI_{RC} = (Απαιτούμενο ενεργειακό κόστος/ Εισόδημα) >10%

- Δείκτης DCEN (Degree of Coverage of Energy Needs). Ο συγκεκριμένος δείκτης εισήχθη για πρώτη φορά από την Δρ. Λευκοθέα Παπαδά στη διδακτορική της διατριβή (Παπαδά, 2017). Ο συγκεκριμένος δείκτης μελετάει, σε ποσοτικό επίπεδο, την υποτιμημένη αλλά ιδιαίτερα σημαντική μορφή της ενεργειακής φτώχειας, αυτή της συμπίεσης των ενεργειακών αναγκών, όπως και το φαινόμενο της ενεργειακής σπατάλης, κατηγοριοποιώντας τα νοικοκυριά σε τρεις ομάδες (Συμπίεση αναγκών - Ικανοποίηση αναγκών - Σπατάλη), διερευνώντας σχετιζόμενες παραμέτρους. Πιο συγκεκριμένα, ο δείκτης ορίζεται ως ο λόγος του καταβληθέντος ενεργειακού κόστους (actual energy expenditure) προς το απαιτούμενο ενεργειακό κόστος (required energy expenditure) (Παπαδά, 2017; Papada & Kaliampakos, 2020).

DCEN= Actual Energy Expenditure/ Required Energy Expenditure

Ανάλογα με τη τιμή του δείκτη DCEN, η κάλυψη των ενεργειακών αναγκών των νοικοκυριών κατηγοριοποιείται βάσει της ακόλουθης διαβάθμισης:

- DCEN ≤ 0,8: Συμπίεση αναγκών
- 0,8 ≤DCEN≤ 1,0: Ικανοποίηση αναγκών
- DCEN ≥ 1,0: Σπατάλη

Το όριο του 80% βασίστηκε στη μελέτη του Tourkolias (2016), ο οποίος εισήγαγε ως Δείκτη ενεργειακής φτώχειας το λόγο «Πραγματική / Θεωρητικά απαιτούμενη ενεργειακή κατανάλωση <80%» (Actual energy consumption/Theoretically required energy consumption<80%), θεωρώντας ότι κάτω από το όριο αυτό δεν είναι δυνατόν να ικανοποιηθούν επαρκώς οι ενεργειακές ανάγκες. Με στόχο τον εντοπισμό της καθαρής συμπίεσης των αναγκών, «επετράπη» στο δείκτη DCEN να θεωρεί ικανοποίηση των αναγκών μια μείωση του απαιτούμενου κόστους έως 20%.

Ο δείκτης DCEN αποτελεί έναν από τους τρεις δείκτες που θα υιοθετηθούν για τη μοντελοποίηση των δεδομένων στην συγκεκριμένη διπλωματική εργασία, ώστε να μελετηθεί δευτερευόντως, και η πραγματική ικανότητα ταξινόμησης των νοικοκυριών ως ενεργειακά ή μη ενεργειακά φτωχά σύμφωνα με τον ορισμό του συγκεκριμένου δείκτη.

- **3η κατηγορία**

Πέραν των αντικειμενικών τρόπων μέτρησης της ενεργειακούς φτώχειας, έχει σχηματιστεί και μια υποκειμενική προσέγγιση του φαινομένου, η οποία βασίζεται στη προσωπική εκτίμηση των νοικοκυριών για τις συνθήκες θερμικής άνεσης ή χρησιμοποιεί έμμεσα δεδομένα (πχ. Καθυστερημένοι λογαριασμοί). Οι κυριότεροι και ευρέως χρησιμοποιούμενοι υποκειμενικοί δείκτες σε ευρωπαϊκό επίπεδο είναι οι κάτωθι (Παπαδά, 2017):

- Αδυναμία να διατηρηθεί το σπίτι επαρκώς ζεστό (Inability to keep home adequately warm)
- Καθυστερημένοι πάγιοι λογαριασμοί (Arrears on utility bills)
- Διαρροή στη στέγη, υγρασία σε τοίχους, πατώματα, θεμέλια ή σάπια κουφώματα (Dwelling with a leaking roof ,damp walls, floors or foundations, or rotten window frames or floor)

Παρ' ότι οι συγκεκριμένοι υποκειμενικοί δείκτες, δεν έχουν αποδειχθεί να έχουν άμεση σχέση με την ενεργειακή φτώχεια, έχει πραγματοποιηθεί η χρήση τους σε έρευνες (πχ. Healy & Clinch, 2002; Thomson& Snell, 2013), με βασικό τους πλεονέκτημα, την ικανότητα να εντοπίζουν τα νοικοκυριά που συμπιέζουν τις ενεργειακές τους ανάγκες λόγω οικονομικής αδυναμίας (Dubois, 2012), κάτι το οποίο οι αντικειμενικοί δείκτες δεν έχουν την δυνατότητα να πράξουν.

Σύμφωνα με τον Fahmy (2011), λόγω της πολυπλοκότητας της επίσημης μεθοδολογίας μέτρησης, είναι δύσκολο να καθοριστεί κατά πόσο επικαλύπτονται οι δύο μέθοδοι και κατά πόσο ο συνδυασμός τους είναι ακριβής (αντικειμενική-υποκειμενική μέτρηση) (Price et al. 2007;DECC 2009). Παρ' όλα αυτά, σύμφωνα με τον Fahmy, η συμπληρωματική χρήση των υποκειμενικών

δεικτών μπορεί να είναι χρήσιμη για την επαλήθευση των αποτελεσμάτων από τον αντικειμενικό δείκτη, αλλά και για τη σύγκριση ακρίβειας μεταξύ αντικειμενικών δεικτών (Παπαδά,2017).

Επίσης, οι υποκειμενικοί δείκτες, είναι πιθανό να έχουν τη δυνατότητα να αποτυπώνουν πτυχές της ενεργειακής φτώχειας που αντικειμενικοί δείκτες αδυνατούν (Price et al., 2012; Roberts et al., 2015, Phimister et al., 2015).

Αξιοσημείωτη έρευνα πάνω στην χρήση των υποκειμενικών δεικτών έχει πραγματοποιηθεί τα τελευταία χρόνια, με πρωτοπόρο στο συγκεκριμένο τομέα την δημοσίευση των Nussbaumer et al. Με τίτλο «Measuring Energy Poverty: Focusing on what matters», όπου εισάγεται για πρώτη φορά τις πολύκριτηριακός συνθετικός δείκτης για τον υπολογισμό της ενεργειακής φτώχειας σε χώρες της Αφρικής. Ο δείκτης «MEPI- Multidimensional Energy Poverty Index» εστιάζει στην έλλειψη σύγχρονων υπηρεσιών ρεύματος. Ο «MEPI», καταγράφει το σύνολο των ενεργειακών στερήσεων που μπορεί να επηρεάσει ένα άτομο. Αποτελείται από πέντε διαστάσεις που αντιπροσωπεύουν βασικές ενεργειακές υπηρεσίες με τη χρήση έξι δεικτών. Ένα άτομο χαρακτηρίζεται ως ενεργειακά φτωχό εάν ο συνδυασμός των στερήσεων που αντιμετωπίζει υπερβαίνει ένα προκαθορισμένο όριο. Το MEPI είναι προϊόν ενός ποσοστού ατόμων (ποσοστό των ατόμων που χαρακτηρίζονται ως ενεργειακά φτωχό) και της μέσης έντασης ενεργειακής φτώχειας αυτών των ατόμων. Οι διαστάσεις και οι δείκτες που λαμβάνει υπόψιν ο συγκεκριμένος δείκτης είναι οι εξής (Πίνακας 1) (Nussbaumer, 2012):

Πίνακας 1: Διαστάσεις και δείκτες του ''MEPI'' . (Πηγή: Nussbaumer, 2012)

Διάσταση	Δείκτης (βαρύτητα)	Παράμετρος
Μαγείρεμα	Σύγχρονες μέθοδοι μαγειρέματος	Τύπος καυσίμου μαγειρέματος
	Εσωτερική οικιακή αέρια ρύπανση	Μαγείρεμα σε φούρνους ή ανοιχτές εστίες φωτιάς (χωρίς κάλυμμα ή καμινάδα, αν δεν χρησιμοποιείται ως καύσιμο το ηλεκτρικό ρεύμα, LPG, φυσικό αέριο ή βιοαέριο
Φωτισμός	Πρόσβαση σε ενέργεια	Υπάρχει πρόσβαση σε ηλεκτρική ενέργεια
Υπηρεσίες οικιακών συσκευών	Ιδιοκτησία οικιακών συσκευών	Ύπαρξη ψυγείου
Ψυχαγωγία/ Εκπαίδευση	Ιδιοκτησία οικιακών συσκευών ψυχαγωγίας/εκπαίδευσης	Ύπαρξη ραδιοφώνου ή τηλεόρασης
Επικοινωνίες	Μέσα τηλεπικοινωνιών	Επίγεια τηλεφωνική γραμμή ή κινητό τηλέφωνο

Η μεθοδολογία του «MEPI» παρέχει μια σειρά πλεονεκτημάτων. Ειδικότερα, επικεντρώνεται στις ενεργειακές υπηρεσίες και βασίζεται σε δεδομένα που σχετίζονται με τις ενεργειακές στερήσεις, σε αντίθεση με την έμμεση άντληση πληροφοριών μέσω μεταβλητών που είναι θεωρητικά συσχετιζόμενες (π.χ. κατανάλωση ενέργειας ή ηλεκτρικής ενέργειας). Επιπλέον, καταγράφει την

έκταση (αριθμός των ενεργειακά φτωχών ατόμων) όσο και την ένταση (πόσο φτωχοί ενεργειακά είναι) (Nussbaumer, 2012).

- **4η κατηγορία**

Σε αυτή τη κατηγορία εντάσσονται οι μελέτες του Bouzarovski (2014) και των Middlemiss & Gillard (2015), όπου σύμφωνα με αυτές , η ενεργειακή φτώχεια θα πρέπει να προσεγγιστεί ως ζήτημα με κύριο άξονα την έννοια της ''ενεργειακής τρωτότητας'' (energy vulnerability) υπό τη λογική πως έτσι θα προσεγγιστούν ορθότερα οι παράγοντες που επηρεάζουν το ενεργειακό πρόβλημα των νοικοκυριών. Οι Bouzakovski & Petrova (2015) και Thomson (2017), μελέτησαν το πρόβλημα εντός του πλαισίου της τρωτότητας, σε σχέση με νέες μεθόδους που αναπτύχθηκαν, λαμβάνοντας υπόψιν την χωρική και χρονική μεταβλητότητα του φαινομένου, όπου οι συνθήκες μεταβάλλονται (Παπαδά,2017).

Με βάση τα παραπάνω, η ενεργειακή φτώχεια, ως ένα πολυδιάστατο φαινόμενο μπορεί να προσεγγιστεί:

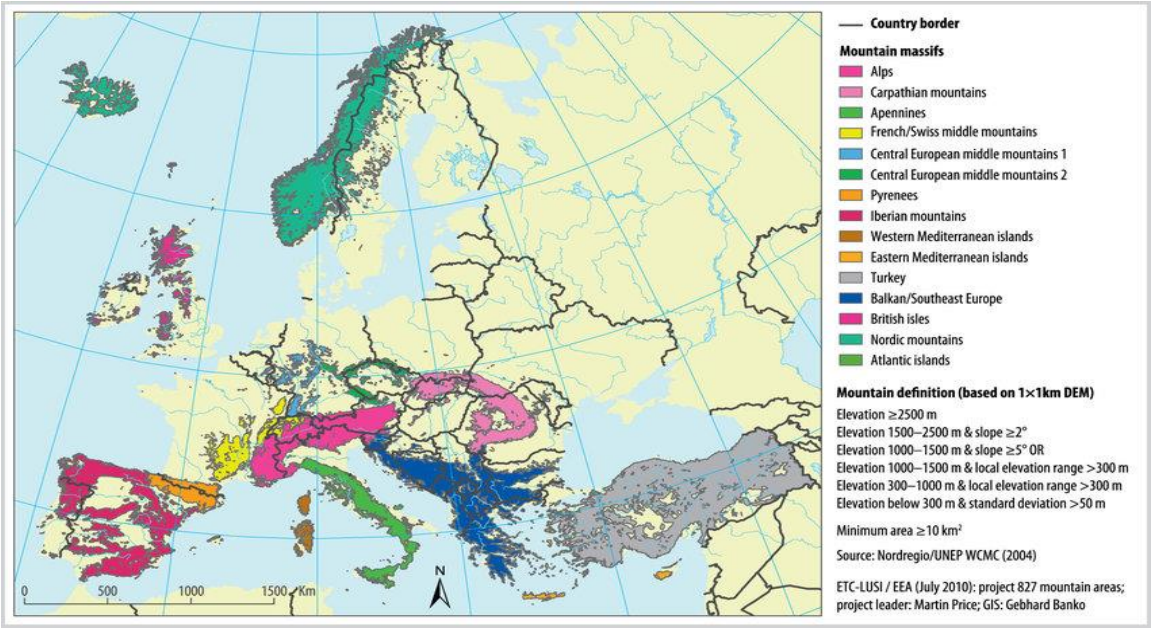
- Με βάση τις δαπάνες, σε συνδυασμό με κάποιο αντικειμενικό δείκτη (π.χ. συμβατικό 10%, επίσημο βάσει απαιτούμενης κατανάλωσης 10 %, LIHC), με στοιχεία στατιστικών δεδομένων ή πρωτογενών ερευνών.
- Με χρήση πολλαπλών υποκειμενικών δεικτών, με στοιχεία εθνικών ή ευρωπαϊκών στατιστικών δεδομένων και πρωτογενών ερευνών (Bouzarovski 2011)
- Με συνδυασμό αντικειμενικών και υποκειμενικών δεικτών (πχ. Lawson et al., 2015)
- Με διαφορετικές, συνολικές προσεγγίσεις, προσαρμοσμένες στο πλαίσιο της ενεργειακής τρωτότητας(π.χ. Bouzarovski, 2014; Thomson et al., 2017)

Κεφάλαιο 2. Ορεινές περιοχές

2.1. Η σημασία των ορεινών περιοχών

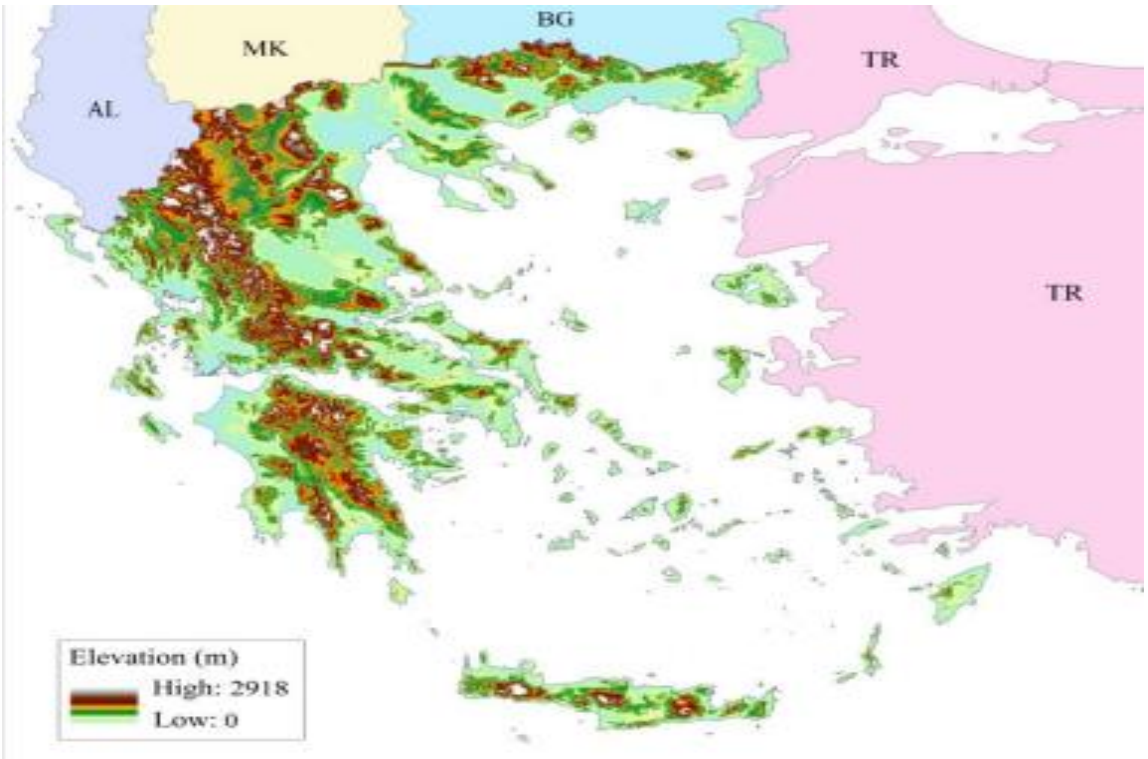
2.1.1.Ορεινές περιοχές και διεθνής προσέγγιση

Τα βουνά αποτελούν γεωγραφικά στοιχεία καθοριστικής σημασίας για την ανθρωπότητα ανά τους αιώνες. Το 24% της επιφάνειας της γης θεωρείται ορεινή, με την Ευρώπη να καλύπτεται κατά 35% από βουνά και την ΕΕ κατά 30% (Εικ. 9), με 112 εκατομμύρια Ευρωπαίους και 64 εκατομμύρια κατοίκους της ΕΕ αντίστοιχα να διαμένουν σε αυτές τις περιοχές. Ωστόσο, έχουν πραγματοποιηθεί εκτεταμένες συζητήσεις σε επίπεδο ΕΕ, οι οποίες κινούνται γύρω από το θέμα της απομόνωσης/περιθωριοποίησης των ορεινών περιοχών από ευκαιρίες προσαρμοσμένης ανάπτυξης και συμμετοχής τους στην αντιμετώπιση της κλιματικής αλλαγής.



Εικόνα 9: Ορεινές περιοχές της Ευρώπης.(Πηγή: Nordregio/WCMC 2004)

Η Ελλάδα αποτελεί, και αυτή, μια κατ' εξοχήν ορεινή χώρα, με την ορεινή της έκταση να καλύπτει περισσότερο από το 78% της συνολικής της έκτασης της χώρας (European Final Report, 2004). Η οροσειρά της Πίνδου διασχίζει το μεγαλύτερο μέρος της ηπειρωτικής χώρας από τα ελληνοαλβανικά σύνορα έως την Πελοπόννησο, ενώ παράλληλα και η υπόλοιπη χωρά, συμπεριλαμβανομένης της Κρήτης, διατρέχεται από όρη (Εικ. 10). Σύμφωνα με την απογραφή του 2001, όπου υπήρξε στατιστική μελέτη των ορεινών και ημιορεινών περιοχών, υπολογίστηκε ότι ο πληθυσμός αυτών των περιοχών ανερχόταν σε περισσότερα από 3 εκατομμύρια, με τους κατοίκους των καθαρά ορεινών περιοχών να ανέρχονται σε 883.500 (ΕΣΥΕ, 2001).



Εικόνα 10: Χάρτης Ορεινών περιοχών, Ελλάδα, (Πηγή: Παπαδά,2017)

Η θετική επίδραση και η αναγκαιότητα των βουνών αναγνωρίζεται διεθνώς από το 1992, στο πλαίσιο της Διάσκεψης των Ηνωμένων Εθνών για το Περιβάλλον και την Ανάπτυξη στο Rio de Janeiro. Το σχέδιο Δράσης (Agenda 21) περιλάμβανε ένα ειδικό κεφάλαιο με τίτλο ‘’ Η διαχείριση των εύθραυστων οικοσυστημάτων: Αειφόρος Ορεινή Ανάπτυξη’’ (United nations General Assembly, 2013).

Στη διάσκεψη του ΟΗΕ το 2012, πραγματοποιήθηκαν εκτεταμένες συζητήσεις για την υιοθέτηση επιπλέον δράσεων και πολιτικών, για τη μείωση της φτώχειας στις ορεινές περιοχές, καθώς η αύξηση κατά 30%, της επισιτιστικής ανασφάλειας κατά το διάστημα 2000-2012 δημιούργησε ανησυχία (FAO, 2015). Έκτοτε, η Αειφόρος Ορεινή Ανάπτυξη κατέχει εξέχουσα θέση στην παγκόσμια ατζέντα, αλλά, παρόλη την παγκόσμια συνεισφορά υπάρχει εκτεταμένη απουσία πληροφοριών σχετικά με την κατάσταση των ορεινών περιοχών και των κατοίκων τους, σύμφωνα με τον Οργανισμό Τροφίμων και Γεωργίας του ΟΗΕ (FAO, 2015).

2.1.2.Συνεισφορά ορεινών οικοσυστημάτων

Τα ορεινά οικοσυστήματα συνεισφέρουν σημαντικά στην επιβίωση της βιοποικιλότητας και στην οικονομική ανάπτυξη (Εικ. 11). Οι κύριες υπηρεσίες που προσφέρουν τα ορεινά οικοσυστήματα κατά τους Kornr et.al (2005), Menash et al. (2017) , Wondimagegn (2020) και Grid Arendal (2022) είναι οι εξής:

- Υπηρεσίες τροφοδοσίας: πρώτες ύλες, ορυκτοί και ενεργειακοί πόροι, προϊόντα παραγωγής των οικοσυστημάτων (ξυλεία, φρέσκο πόσιμο νερό, μη ξύλινα δασικά προϊόντα, φαρμακευτικά προϊόντα ιαματικά φυτά, κλπ.)
- Ρυθμιστικές και υποστηρικτικές υπηρεσίες: γονιμότητα και ρύθμιση του εδάφους, ρύθμιση του κλίματος, βιοποικιλότητα και υποστήριξη ευάλωτων οικοσυστημάτων, έδαφος ως δεξαμενή αποθήκευσης άνθρακα και νερού, ορεινή βλάστηση ως προστασία από φυσικούς κινδύνους (χιονοστιβάδες, κατολισθήσεις), περιορισμός διάβρωσης εδαφών, σύστημα προειδοποίησης φυσικών κινδύνων
- Πολιτισμικές υπηρεσίες: πολιτιστική ποικιλομορφία, αναψυχή και τουρισμός, ιεροί τόποι στην παράδοση και την θρησκεία των περισσότερων λαών ανά την υφήλιο, αισθητικές και εκπαιδευτικές αξίες

Τα ορεινά γεωγραφικά χαρακτηριστικά σε συνδυασμό με τις κλιματικές συνθήκες των βουνών και τον τύπο διαμόρφωσης των ορεινών κοινωνιών, δημιουργούν ευνοϊκές συνθήκες για την ύπαρξη πλούσιου ενεργειακού δυναμικού (Funnell & Parish 2001, Price 2002). Ορισμένα βασικά χαρακτηριστικά των ορεινών περιοχών τα καθιστούν ως εξαιρετικές πηγές για παραγωγή ενέργειας μέσω ανανεώσιμων πηγών ενέργειας.

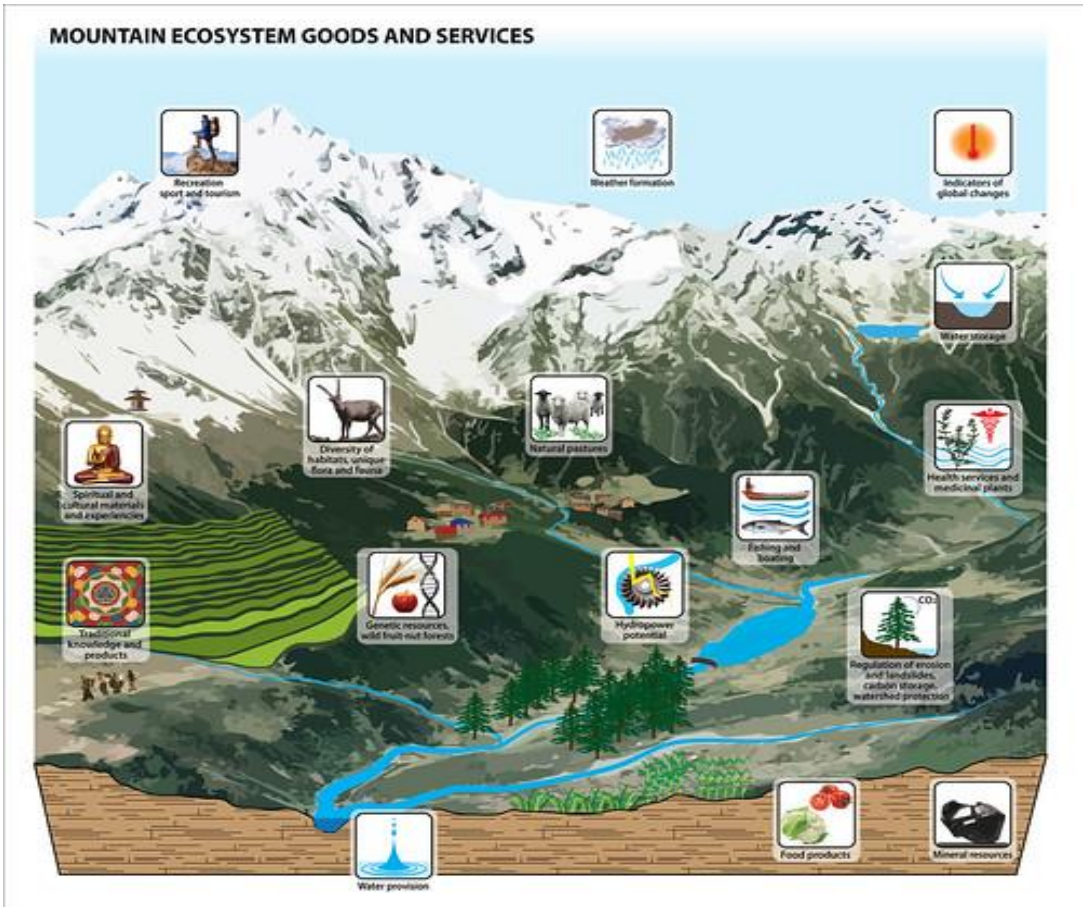
Αιολική ενέργεια: Οι αυξημένες ταχύτητες των ανέμων σε υψηλά υψόμετρα και φαινόμενα όπως οι άνεμοι των κοιλάδων συνήθως ενισχύουν τη ενεργειακή δυναμικότητα των ορεινών περιοχών. Για παράδειγμα, στην Ελλάδα, όσον αφορά την αναμενόμενη αιολική δυναμικότητα, τα βουνά και τα νησιά, θεωρούνται οι πιο ιδανικές περιοχές για την εγκατάσταση ανεμογεννητριών. Σε χωροταξικό σχέδιο για τις ανανεώσιμες πηγές ενέργειας Ελλάδα, υπάρχουν πολλές περιπτώσεις περιοχών που έχουν κατηγοριοποιηθεί ως περιοχές ‘’αιολικής προτεραιότητας’’, και στην πλειονότητα τους αυτές οι περιοχές είναι ημί-ορεινές ή ορεινές (Katsoulakos & Kaliampakos, 2013).

Ηλιακή ενέργεια: Στις ορεινές περιοχές, η αξιοποίηση της ηλιακής ακτινοβολίας για την παραγωγή ενέργειας μπορεί να είναι ιδιαίτερα αποτελεσματική, λόγω της αύξησης της ηλιακής ακτινοβολίας σε σχέση με το υψόμετρο. Εκτός από την αύξηση της ηλιακής ακτινοβολίας σε σχέση με το υψόμετρο, η υψηλότερη απόδοση των φωτοβολταϊκών συστημάτων στις ορεινές περιοχές αποδίδεται στις χαμηλότερες θερμοκρασίες των ορεινών περιοχών. Πιο συγκεκριμένα, ένας από τους παράγοντες που επηρεάζουν τη συνολική απόδοση ενός φωτοβολταϊκού πλαισίου σχετίζεται με τη διαφορά μεταξύ της θερμοκρασίας, κάτω από την οποία το λειτουργεί το πάνελ, με την θερμοκρασία μέγιστης απόδοσης να βρίσκεται περίπου στους 25°C. Από την άλλη πλευρά, υπάρχει και η θεώρηση, ότι τα φωτοβολταϊκά συστήματα σε ορεινές περιοχές παρουσιάζουν χαμηλή απόδοση, βάσει αποτελεσμάτων, γι’ αυτό και η χρήση τους σε ορεινές περιοχές παραμένει περιορισμένη σε μικρές οικιακές εγκαταστάτης. Η περιορισμένη χρήση

δικαιολογείται και από την έλλειψη εκτενούς ανοικτού επίπεδου χώρου στις ορεινές περιοχές, με τις απότομες κλίσεις να επικρατούν στο ανάγλυφο των ορεινών περιοχών και να καθιστούν αδύνατη την εγκατάσταση τέτοιων εγκαταστάσεων (Katsoulakos & Kaliampakos, 2013).

Υδροηλεκτρική ενέργεια: Ο ορεινός χώρος περικλείει όλες τις απαραίτητες συνθήκες για την ανάπτυξη υδροηλεκτρικών σταθμών ενέργειας, ειδικά μέσω μικρής κλίμακας υδροηλεκτρικών έργων. Οι διαθέσιμοι υδατικοί όγκοι σε συνδυασμό με τα μεγάλα ύψη πτώσης των βουνών, καθιστά τις ορεινές περιοχές αποδοτικές τοποθεσίες για την εγκατάσταση οικονομικά βιώσιμων υδροηλεκτρικών σταθμών (Katsoulakos & Kaliampakos, 2013).

Βιομάζα: Η ορεινή βιομάζα συνιστά σημαντική ανανεώσιμη πηγή ενέργειας για την παραγωγή θερμικής ενέργειας, καθώς περιλαμβάνει προϊόντα και παραπροϊόντα από την υλοτομική εκμετάλλευση, την κτηνοτροφία και την γεωργία, αν και οι διαθέσιμες ποσότητες αυτών εξαρτώνται από τη δυναμικότητα των ανθρώπινων δραστηριοτήτων (Katsoulakos & Kaliampakos, 2013).



Εικόνα 11: Αγαθά και Υπηρεσίες των ορεινών οικοσυστημάτων.(Πηγή : GRID ARENDAL 2018)

Εν κατακλείδι, αποδεικνύεται ότι οι ορεινές περιοχές, χαρακτηρίζονται από πλούσια αιολικό, ηλιακό, υδροηλεκτρικό δυναμικό και δασική βιομάζα, κυρίως λόγω των ιδιαίτερων χαρακτηριστικών του ορεινού κλίματος. Χαρακτηριστικό παράδειγμα αποτελούν ορεινές περιοχές του ελλαδικού χώρου. Σύμφωνα με τους Katsoulakos & Kaliampakos (2013), μια ορεινή περιοχή είναι πιθανότερο να έχει περισσότερο εκμεταλλεύσιμο ενεργειακό δυναμικό από μια αντίστοιχη πεδινή. Αυτό απεικονίζεται στην την περίπτωση του Μετσόβου και των Ιωαννίνων, στην επαρχία Ηπείρου. Ο ορεινός δήμος του Μετσόβου εκτιμάται ότι διαθέτει 72% υψηλότερο συνολικό δυναμικό ανανεώσιμης ενέργειας από τον γειτονικό, πεδινό δήμο Ιωαννίνων. Ως εκ τούτου, προκύπτει μια ένδειξη εγγενούς σχέσης μεταξύ των βουνών και του δυναμικού ανανεώσιμης ενέργειας (Παπαδά, 2017).

2.2. Το ενεργειακό πρόβλημα των ορεινών περιοχών

Παρά τις παραπάνω ευνοϊκές συνθήκες των ορεινών περιοχών και του πλούσιου ενεργειακού δυναμικού που τα χαρακτηρίζει, οι ορεινοί πληθυσμοί πάσχουν από σοβαρά προβλήματα ενεργειακής ανεπάρκειας και ενεργειακής φτώχειας. Αυτά τα προβλήματα προκύπτουν από τα βασικά δυσμενή χαρακτηριστικά των ορεινών περιοχών, όπως οι ψυχρές κλιματικές συνθήκες, η γεωγραφική απομόνωση, το παλαιό κτιριακό δυναμικό και η ασθενής παραγωγική βάση, τα οποία καθιστούν τα ορεινά νοικοκυριά ευάλωτα στην ενεργειακή φτώχεια (Katsoulakos et al., 2014). Από την άλλη πλευρά, βρίσκονται τα νοικοκυριά των αστικών κέντρων, που είναι ευάλωτα σε ‘’προσωρινή ενεργειακή φτώχεια’’ (Papada et al., 2016), η οποία οξύνεται ή υποχωρεί ανάλογα με τις τρέχουσες οικονομικές και κοινωνικές συνθήκες.

Τα βασικά χαρακτηριστικά που επιβαρύνουν τα ορεινά νοικοκυριά είναι:

- **Υψόμετρο**

Το υψόμετρο αποτελεί το βασικότερο χαρακτηριστικό γνώρισμα του ορεινού χώρου. Το 48% των παγκόσμιων εδαφών βρίσκεται σε υψόμετρα μεγαλύτερα των 500 m, το 11% σε υψόμετρα μεγαλύτερα των 2.000 m και το 2% σε υψόμετρα άνω των 4.000 m (Καλιαμπάκος et al.2009, Κατσουλάκος 2013). Για την περίπτωση της Ελλάδας, το 20% των οικισμών βρίσκεται σε υψόμετρα άνω των 600 m, οι οποίοι φιλοξενούν το 6% του πληθυσμού. Σε γενική βάση, οι δυσμενείς συνθήκες που χαρακτηρίζουν την ορεινή διαβίωση, ειδικά στα υψηλά υψόμετρα, προκαλούν εκθετική μείωση του πληθυσμού σε σχέση με την αύξηση του υψομέτρου (Κατσουλάκος 2013). Σε παγκόσμιο επίπεδο, ο ορεινός πληθυσμός αντιπροσωπεύει, συνολικά, το ένα δέκατο του παγκόσμιου πληθυσμού (FAO, 2015)

Παράλληλα, το υψόμετρο αποτελεί καθοριστική παράμετρο στη διαμόρφωση του κλίματος, με το ορεινό κλίμα να χαρακτηρίζεται από δριμείς χειμώνες και σύντομα καλοκαίρια. Πιο συγκεκριμένα, η θερμοκρασία μειώνεται με την αύξηση του υψομέτρου (μείωση 2°C-3 ^{o}C κατά μέσο όρο ανά 300 m), γεγονός που οδηγεί σε σημαντική αύξηση των θερμικών αναγκών των ορεινών περιοχών συγκριτικά με τις πεδινές. Ενδεικτικά στην Ελλάδα, η θερμική ενεργειακή ζήτηση στα 1.000 m είναι κατά 170% υψηλότερη σε σχέση με το επίπεδο της θάλασσας (Κατσουλάκος, 2013).

- **Απομόνωση**

Η γεωγραφική απομόνωση συνδέεται με τη δυσκολία πρόσβασης σε συγκεκριμένες ανάγκες και αποτελεί ένα σημαντικό χαρακτηριστικό των ορεινών περιοχών, που παρεμποδίζει την ομαλή, σταθερή ενεργειακή τροφοδοσία τους. Μια γεωγραφικά απομονωμένη περιοχή έχει μειωμένη προσβασιμότητα σε υπηρεσίες κοινής ωφέλειας και στις μεγάλες αγορές ενέργειας. Αυτές οι βασικές ελλείψεις επηρεάζουν την ενεργειακή κάλυψη των νοικοκυριών, καθώς επικρατεί μειωμένη πρόσβαση σε βασικές ενεργειακές υπηρεσίες (πχ. Δίκτυο ηλεκτρισμού) (Rodrigue et al. 2013) ενώ, και το κόστος ενέργειας αυξάνεται σημαντικά λόγω των μεγάλων αποστάσεων από τις μεγάλες αγορές (MacDonald, 2019; Marvin, 1994). Οι ορεινοί οικισμοί βρίσκονται σε μεγάλη απόσταση από τα αστικά και βιομηχανικά κέντρα, όπου είναι συνήθως εγκατεστημένα τα διυλιστήρια και οι μονάδες αποθήκευσης άλλων καυσίμων. Κατά συνέπεια, η μεταφορά καυσίμων αυξάνει σημαντικά το κόστος τους στον τελικό προορισμό, όπου σε αυτή τη περίπτωση είναι οι ορεινές περιοχές, με την κακή κατάσταση του οδικού δικτύου να δυσχεραίνει περισσότερο τη κατάσταση (Κατσουλάκος, 2013).

Ταυτόχρονα, σημαντικό ζήτημα αποτελεί και η πρόσβαση των απομονωμένων οικισμών στο δίκτυο ηλεκτρισμού, καθώς αποτελεί και αυτό ένα σημαντικό κόστος, το οποίο μεταφέρεται στους ευάλωτους κατοίκους των περιοχών αυτών. Αυτό συμβαίνει γιατί, στην πλειονότητα των περιπτώσεων, τα κεντρικά δίκτυα μεταφοράς ηλεκτρικής ενέργειας τροφοδοτούν τις μεγάλες κεντρικές ενεργειακές μονάδες και εν συνεχεία επεκτείνονται στους οικισμούς δίνοντας προτεραιότητα σε αυτούς που απαιτούν τα μικρότερα κόστη επένδυσης για την επέκταση του δικτύου. Έτσι, οι απομονωμένες και ορεινές περιοχές είναι αυτές που συνήθως ηλεκτροδοτούνται τελευταίες (UNDP, 2000). Η μεθοδολογία αυτή, υιοθετήθηκε και στην Ελλάδα, με χαρακτηριστικό παράδειγμα της ορεινής περιοχής των Άγραφων στην Ήπειρο, όπου ηλεκτροδοτήθηκε μόλις στα μέσα της δεκαετίας του 1980 (Κατσουλάκος, 2013). Στο κόστος επένδυσης για ηλεκτροδότηση, προστίθενται και άλλοι παράγοντες που καθιστούν την ηλεκτροδότηση των ορεινών περιοχών ακόμη πιο δύσκολη, και βασικότερα , ασύμφορη. Η τύπου δενδροειδής ανάπτυξη του ηλεκτρικού δικτύου στις ορεινές περιοχές συνοδεύονται από υψηλές ενεργειακές απώλειες στα δίκτυα μεταφοράς και διανομής, μη αποδοτική αξιοποίηση της διαθέσιμης ισχύος λόγω χαμηλής ζήτησης και η αιχμιακή μορφή ζήτησης ενέργειας στους μικρούς οικισμούς. Πράγματι, συχνές είναι οι περιπτώσεις δικτύων ηλεκτροδότησης ορεινών περιοχών παλαιάς, ανεπαρκούς κατασκευής, με αποτέλεσμα τις συχνές διακοπές ηλεκτροδότησης, τις πτώσεις τάσης στις περιόδους συγκέντρωσης μεγάλου τμήματος του πληθυσμού και τα υψηλά κόστη για την αποκατάσταση των ζημιών (Κατσουλάκος, 2013).

Τα παραπάνω προβλήματα έχουν ως αποτέλεσμα εκτενείς περιοχές του πλανήτη να έχουν περιορισμένη ή μηδενική πρόσβαση σε ενεργειακές υπηρεσίες. Στην Ελλάδα, δεν σημειώνεται πλέον το πρόβλημα έλλειψης πρόσβασης σε υπηρεσίες ενέργειας με εξαίρεση το δίκτυο μεταφοράς και διανομής φυσικού αερίου, το οποίο είναι εξαιρετικά περιορισμένο στις ορεινές περιοχές. Ωστόσο, η μειωμένη προσβασιμότητα συγκεκριμένων περιοχών, κυρίως των ορεινών και νησιωτικών, δημιουργεί συχνά προβλήματα που αφορούν την επαρκή ενεργειακή κάλυψη του πληθυσμού και τις υψηλές τιμές των καυσίμων.

- **Κλίση εδάφους**

Η κλίση του εδάφους είναι τυπικό χαρακτηριστικό των περισσότερων ορεινών περιοχών και αποτελεί σημαντικό εμπόδιο στην αξιοποίηση του άφθονου ενεργειακού δυναμικού, καθώς μειώνει σημαντικά το διαθέσιμο χώρο προς εγκατάσταση ενεργειακών μονάδων, όπως οι φωτοβολταϊκές μονάδες ηλεκτροπαραγωγής. Συμπληρωματικά, η έλλειψη διαθέσιμου χώρου, σε συνδυασμό με τα κατολισθητικά φαινόμενα που συνήθως ευνοούνται από τις μεγάλες κλίσεις, δημιουργούν προβλήματα και στη χωροθέτηση μικρότερων ενεργειακών μονάδων, όπως είναι οι εγκαταστάσεις βιομάζας και οι ανεμογεννήτριες (Papada & Kaliampakos, 2015; Katsoulakos & Kaliampakos, 2016).

- **Ασθενής παραγωγική βάση- Φτώχεια**

Τα εισοδήματα των ορεινών πληθυσμών είναι, κατά κανόνα, χαμηλότερα από αυτά των πεδινών και αστικών περιοχών. Οι παραγωγικές/οικονομικές δραστηριότητες, και άρα, οι θέσεις απασχόλησης είναι συγκεντρωμένες στις πυκνοκατοικημένες πεδινές περιοχές. Πιο συγκεκριμένα, η γεωργική παραγωγή αποτελεί την κύρια παραγωγική δραστηριότητα στα ορεινά τμήματα του πλανήτη. Ωστόσο, η έλλειψη μεγάλων καλλιεργήσιμων εκτάσεων, το ψυχρό κλίμα και συνεχείς συγκρούσεις χρήσεων στις ορεινές περιοχές δρουν εναντίον της γεωργικής ανάπτυξης στις ορεινές περιοχές. Μία άλλη βασική δραστηριότητα των ορεινών περιοχών είναι η κτηνοτροφία, η οποία παλαιότερα αποτελούσε μια ανεπτυγμένη παραγωγική βάση , αλλά πλέον έχει παρακμάσει σημαντικά. Αιτία αυτής της παρακμής, αποτελούν οι μεγάλες, οργανωμένες κτηνοτροφικές μονάδες, οι οποίες έχουν αντικαταστήσει σε μεγάλο βαθμό την ημί-νομαδική εκτροφή ζώων. Ο δευτερογενής τομέας, ο οποίος αφορά κυρίως μεταποιητικές μονάδες είναι εξαιρετικά περιορισμένος λόγω της απομόνωσης και του μικρού πληθυσμού. Όσον αφορά, τον τριτογενή τομέα, οι εμπορικές δραστηριότητες και παροχή υπηρεσιών είναι σχεδόν ανύπαρκτες, με τον τουρισμό να αποτελεί τη μοναδική αξιόλογη δραστηριότητα, αν και ο ορεινός τουρισμός βρίσκεται σε πρώιμο στάδιο ανάπτυξης (Papada, 2017).

Συνεπώς, γίνεται κατανοητό ότι η παραγωγική δυναμικότητα των περιοχών είναι αδύναμη και οι δυνατότητες απασχόλησης περιορισμένες, με την ανεργία στις ορεινές περιοχές να είναι αυξημένη λόγω της ελάχιστων οικονομικών δραστηριοτήτων (Κατσουλάκος, 2013). Χαρακτηριστικό παράδειγμα για την Ελλάδα αποτελεί η πιο ορεινή της περιφέρεια, η Ήπειρος, η οποία αποτελεί και την φτωχότερη περιφέρεια της χώρας, εμφανίζοντας τα χαμηλότερα εισοδήματα (Παπαδά, 2017) . Στον παρακάτω πίνακα (Πίνακας 2) παρουσιάζονται οι διαφορές στα εισοδήματα μεταξύ ορεινών νομών σε σύγκριση με το Νομό Αττικής, σύμφωνα με τα τελευταία διαθέσιμα στοιχειά για τα δηλωμένα εισοδήματα κατοίκων ανά δήμο για το 2011.

Πίνακας 2: Ποσοστιαία μεταβολή των εισοδημάτων ορεινών Νομών σε σχέση με το νομό Αττικής. Οικονομικό έτος 2011.(Πηγή: ΓΓΠΣ 2011)

Νομός	Μέσο δηλωθέν οικογενειακό εισόδημα (€)	Σύγκριση με νομό Αττικής (%)
Αχαΐας	16.929	-21%
Ιωαννίνων	16.756	-22%
Αρκαδίας	16.502	-23%
Βοιωτίας	15.594	-27%
Άρτας	15.059	-30%
Ευρυτανίας	13.884	-35%
Γρεβενών	13.639	-36%
Λακωνίας	12.659	-41%
Αττική	**21.447**	

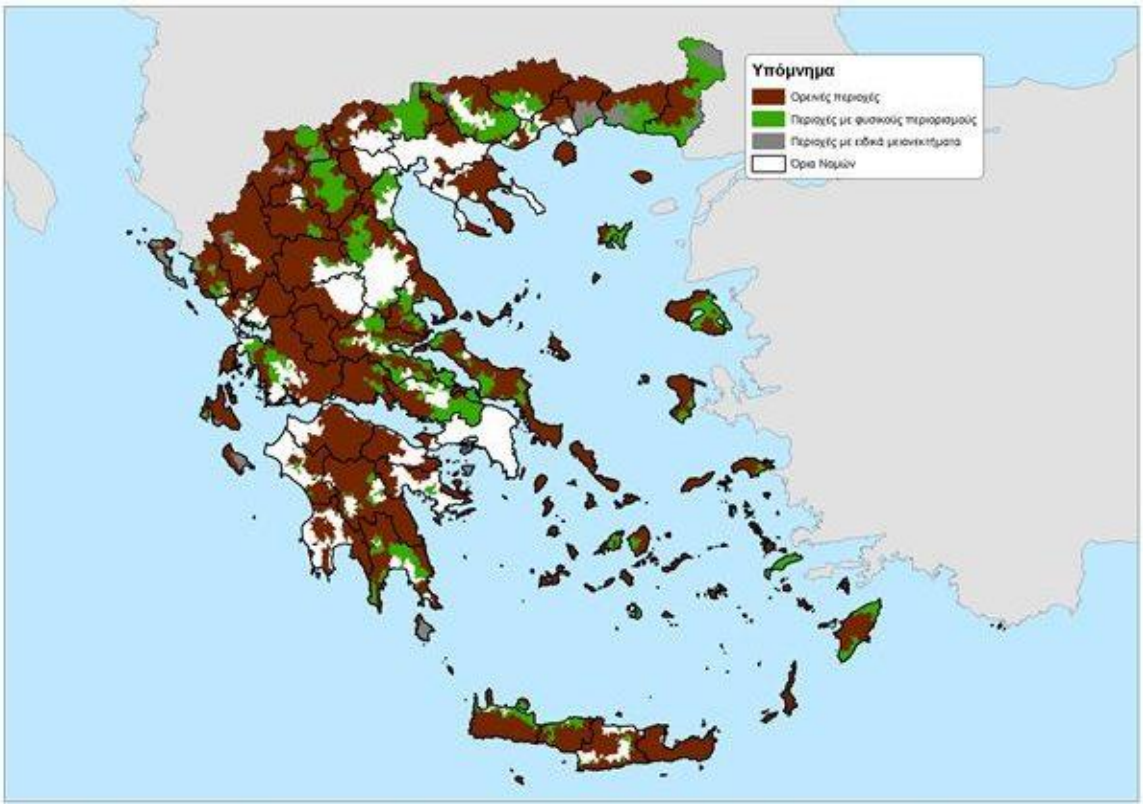
- **Παλαιό κτιριακό δυναμικό**

Μία ακόμη συνθήκη που δυσχεραίνει τις ενεργειακές δυσκολίες των ορεινών περιοχών είναι το παλαιό κτιριακό δυναμικό, το οποίο επικρατεί στα περισσότερα ορεινά κτίρια και ισοδυναμεί με χαμηλή ενεργειακή απόδοση κτιρίων και εκτεταμένες ενεργειακές απώλειες. Οι ορεινές περιοχές εμφανίζουν υψηλά ποσοστά κτιρίων χτισμένων πριν το 1980, τα οποία στις περισσότερες περιπτώσεις δεν έχουν θερμομονωτική προστασία, καθώς ο πρώτος Κανονισμός για τη θερμομονωτική προστασία των κτιρίων στη χώρα εισήχθη το 1979 (Παπαδά, 2017) . Στον κάτωθι πίνακα (Πίνακας 3) φαίνονται τα ποσοστά κτιρίων ορεινών περιοχών, κατασκευασμένα πριν το 1960 και το 1980, σύμφωνα με την απογραφή του 2011.

Πίνακας 3: Ποσοστά κτιρίων κατασκευασμένα πριν το 1960 και πριν το 1980 σε κύριους ορεινούς της Ελλάδας, σε σχέση με το σύνολο της χώρας (Πηγή: Katsoulakos et al.2014, ΕΛΣΤΑΤ 2011)

Ορεινοί οικισμοί (Υψόμετρο)	Ποσοστά κτιρίων χτισμένων:	
	Πριν το 1960	Πριν το 1980
Αράχοβα (1075 m)	18%	41%
Δημητσάνα (950 m)	36%	46%
Καλάβρυτα(750 m)	28%	70%
Καρπενήσι(980 m)	25%	57%
Λιβάδι Ολύμπου(1180 m)	35%	76%
Μέτσοβο (1240 m)	27%	55%
Σύνολο χώρας	**17%**	**38%**

Σύμφωνα με τις παραπάνω παραμέτρους, οι οποίες επιβαρύνουν δριμύτατα τα ορεινά νοικοκυριά, γίνεται κατανοητό ότι οι ορεινές περιοχές είναι ευάλωτες στην ενεργειακή φτώχεια. Τα τελευταία χρόνια, εν μέσω συνεχών αυξήσεων στις τιμές της ενέργειας, τα χαμηλόμισθα τμήματα του πληθυσμού καταφεύγουν σε μεθόδους μείωσης της οικιακής ενεργειακής κατανάλωσης κάτω από τα αποδεκτά πρότυπα διαβίωσης και σε χρήση μη ασφαλών καυσίμων (Γιακουμίδης & Ιατρίδης, 2009), με την οικονομική κρίση κατά την πενταετία (2009-2014) να οδηγεί σε όξυνση του φαινομένου στον ελλαδικό χώρο. Σε αυτό το διάστημα, οι τιμές του πετρελαίου αυξήθηκαν δραματικά (από 42 $/βαρέλι τον Ιανουάριο του 2009 σε 105 $/βαρέλι τον Ιούλιο του 2014, για το αργό πετρέλαιο (Trading Economics,2022), με αποτέλεσμα τη μείωση της ζήτησης και τη ''στροφή'' των καταναλωτών – ειδικά αυτών που κατοικούν σε ορεινές περιοχές- σε εναλλακτικές πηγές καυσίμων όπως η βιομάζα (καυσόξυλα, πέλλετς, μπρικέτες κλπ.) (Katsoulakos et al.,2015). Παράλληλα, πολλαπλασιάστηκαν τα περιστατικά παράνομης υλοτομίας και η χρήση φθηνής, ανεπεξέργαστης βιομάζας, η οποία οδήγησε εμφάνιση σημαντικής ατμοσφαιρικής ρύπανσης (αιθαλομίχλης) και προβλημάτων υγείας (Katsoulakos et al. 2015). Συγκρίνοντας την Εικ. 10 με την Εικ. 12, παρατηρείται εύκολα η ταύτιση της οριοθέτησης των ορεινών περιοχών της Ελλάδας με την οριοθέτηση των μειονεκτικών περιοχών της χώρας για το 2018 σύμφωνα με το Υπουργείου Αγροτικής Ανάπτυξης και Τροφίμων και την Ευρωπαϊκή Επιτροπή. Η εξ' ολοκλήρου κάλυψη των ορεινών περιοχών και η καθιέρωση τους ως μειονεκτικές περιοχές, αναδεικνύει περαιτέρω την δυσχερή οικονομική και παραγωγική κατάσταση που επικρατεί στον ορεινό πληθυσμό της χώρας.



Εικόνα 12: Χάρτης οριοθέτησης μειονεκτικών περιοχών , 2018. (Πηγή: Υπουργείο Αγροτικής Ανάπτυξης και Τροφίμων 2018)

Όσον αφορά τις ορεινές περιοχές του αναπτυσσόμενου κόσμου, όπως αναφέρθηκε και στο Κεφάλαιο 1.3., η ενεργειακή φτώχεια εμφανίζεται με τη μορφή αδυναμίας πρόσβασης σε ενεργειακές υπηρεσίες και δεν αφορά τόσο το κόστος των υπηρεσιών αυτών, όπως στις ανεπτυγμένες χώρες. Η Υποσαχάρια Αφρική, οι περιοχές των Ιμαλάιων (Ινδία, Νεπάλ) και των Άνδεων (Περού, Βολιβία) αποτελούν τα πιο χαρακτηριστικά παραδείγματα επίδρασης της ενεργειακής φτώχειας με σημαντικά ποσοστά του πληθυσμού να μην έχουν πρόσβαση σε βασικές υπηρεσίες ενέργειας, με αποτέλεσμα την επιβράδυνση της ανάπτυξης της οικονομίας (Παπαδά, 2017).

Από τη δεκαετία του ’80 έχει εκφραστεί ότι, μία κατά κεφαλήν αύξηση ηλεκτρικής ισχύος της τάξεως των 100 W σε περιοχές του αναπτυσσόμενου κόσμου (15% της μέσης κατά κεφαλήν ισχύος του ανεπτυγμένου κόσμου) μπορεί να οδηγήσει σε ριζική βελτίωση του βιοτικού επιπέδου των πληθυσμών αυτών (Goldemberg et al., 1985). Παρ’ όλα αυτά, 37 χρόνια μετά την παραπάνω εκτίμηση, υπολογίζεται ότι 1,2 δισεκατομμύρια άνθρωποι παγκοσμίως δεν έχουν πρόσβαση στην ηλεκτρική ενέργεια και πάνω από 2,7 δισεκατομμύρια άνθρωποι δεν έχουν πρόσβαση σε καθαρές εγκαταστάσεις μαγειρέματος, το μεγαλύτερο μέρος των οποίων ζουν σε αγροτικές και ορεινές περιοχές των παραπάνω περιοχών (WEO 2022).

Κεφάλαιο 3. Αλγόριθμοι μηχανικής μάθησης

3.1 Εισαγωγή στους Αλγόριθμους μηχανικής μάθησης

3.1.1. Ορισμός αλγορίθμων μηχανικής μάθησης

Η μηχανική μάθηση είναι ένας κλάδος της τεχνητής νοημοσύνης (AI) και της επιστήμης των υπολογιστών που επικεντρώνεται στη χρήση δεδομένων και αλγορίθμων για τη μίμηση της μεθόδου μάθησης του ανθρώπινου εγκεφάλου, βελτιώνοντας σταδιακά την ακρίβειά του.

Η βασική έννοια της μηχανικής μάθησης στην επιστήμη των δεδομένων περιλαμβάνει τη χρήση μεθόδων στατιστικής μάθησης και βελτιστοποίησης που επιτρέπουν στους υπολογιστές να αναλύουν σύνολα δεδομένων και να εντοπίζουν μοτίβα. Οι τεχνικές μηχανικής μάθησης αξιοποιούν την «εξόρυξη» δεδομένων για τον εντοπισμό ιστορικών τάσεων και την ενημέρωση μελλοντικών μοντέλων.

Στη μηχανική μάθηση, κύριο μέλημα αποτελεί η εκπαίδευση αλγορίθμων μάθησης, όπως η γραμμική παλινδρόμηση (Linear regression), Δέντρα αποφάσεων (Decision Trees), διαχωριστής K-means, αλγόριθμος Random Forest, αλγόριθμος K-nearest neighbors (KNN) και ο SVM (Support vector machine).

Αυτοί οι αλγόριθμοι φιλτράρουν σύνολα δεδομένων, μέσω των οποίων ‘’μαθαίνουν’’, προσαρμόζονται σε νέες καταστάσεις και αναζητούν συνεχώς νέα μοτίβα προσέγγισης των δεδομένων. Τα δεδομένα αυτά αποτελούν το βασικό υπόστρωμα για τη καλύτερη δυνατή λειτουργία των αλγορίθμων.

Ο τυπικός αλγόριθμος μηχανικής μάθησης αποτελείται από τρία περίπου στοιχεία:

1. **Διαδικασία λήψης αποφάσεων:** Μια ‘’συνταγή’’ υπολογισμών ή άλλων βημάτων που λαμβάνει τα δεδομένα και "μαντεύει" τι είδους μοτίβο ψάχνει να βρει ο αλγόριθμος.

2. **Συνάρτηση σφάλματος:** Συγκρίνοντάς την με γνωστά παραδείγματα (όταν αυτά είναι διαθέσιμα). Η διαδικασία λήψης αποφάσεων ήταν σωστή; Αν όχι, πώς ποσοτικοποιείται το "πόσο κακή" ήταν η αστοχία;
3. **Διαδικασία ενημέρωσης ή βελτιστοποίησης:** Μια μέθοδος κατά την οποία ο αλγόριθμος εξετάζει την αστοχία και στη συνέχεια ενημερώνει τον τρόπο με τον οποίο η διαδικασία λήψης αποφάσεων καταλήγει στην τελική απόφαση, ώστε την επόμενη φορά η αστοχία να μην είναι τόσο μεγάλη.

Δεδομένου ότι ένας αλγόριθμος μηχανικής μάθησης ενημερώνεται αυτόνομα, η αναλυτική ακρίβεια βελτιώνεται με κάθε πραγματοποιούμενη εκτέλεση, καθώς ο αλγόριθμος μαθαίνει μόνος του από τα δεδομένα που αναλύει. Αυτός ο επαναληπτικός χαρακτήρας της μάθησης είναι μοναδικός και πολύτιμος, επειδή συμβαίνει χωρίς ανθρώπινη παρέμβαση δίνοντας τη δυνατότητα στον αλγόριθμο να αποκαλύπτει κρυφές γνώσεις χωρίς να είναι ειδικά προγραμματισμένος για να το κάνει.

3.1.2. Τύποι μοντέλων Μηχανικής Μάθησης

Υπάρχουν πολλοί τύποι μοντέλων μηχανικής μάθησης που ορίζονται από την παρουσία ή την απουσία ανθρώπινης επιρροής στα ακατέργαστα δεδομένα, αν προσφέρεται ‘’ανταμοιβή’’, αν δίνεται συγκεκριμένη ανατροφοδότηση ή αν χρησιμοποιούνται ετικέτες.

- **<u>Εποπτευόμενη μάθηση:</u>** Το χρησιμοποιούμενο σύνολο δεδομένων έχει επισημανθεί και ταξινομηθεί εκ τω προτέρων από τον χρήστη , ώστε ο αλγόριθμο να αξιολογήσει την απόδοσή του. Τα σύνολα δεδομένων που χρησιμοποιούνται για εκπαίδευση των αλγορίθμων μηχανικής μάθησης χαρακτηρίζονται με ετικέτες, δηλαδή τα δεδομένα εισάγονται στον αλγόριθμο μαζί ένα φύλλο απαντήσεων, το οποίο ενημερώνει τον υπολογιστή για την σωστή ή επιθυμητή απάντηση. Για παράδειγμα, ένας υπολογιστής που σαρώνει τα εισερχόμενα μηνύματα για ανεπιθύμητα μηνύματα μπορεί να ανατρέξει σε ένα σύνολο δεδομένων με ετικέτες για να καταλάβει ποια μηνύματα είναι ανεπιθύμητα και ποια νόμιμα. Αυτό, ουσιαστικά ονομάζεται μάθηση ‘’υπό επίβλεψη’’(Prieto et al.2016).

- **<u>Μη εποπτευόμενη μάθηση:</u>** Το ακατέργαστο σύνολο δεδομένων που χρησιμοποιείται είναι μη ταξινομημένο από το χρήστη και ένας αλγόριθμος εντοπίζει μοτίβα και σχέσεις στα δεδομένα χωρίς τη βοήθεια του χρήστη. Σε αυτή την κατηγορία, οι μέθοδοι μάθησης είναι οι εξής (Prieto et al.2016):

    (1) **Ανακάλυψη συστάδων** (Clusters discovering) με ομαδοποίηση δεδομένων σε ομάδες).
    (2) **Εξαγωγή λανθανόντων παραγόντων** (Extraction of latent factors). Αυτή η μέθοδος αποσκοπεί στη μείωση της διαστατικότητας των λανθανόντων παραγόντων, προβάλλοντας τα δεδομένα σε ένα υπό-χώρο μικρότερης διάστασης, εξάγοντας τους παράγοντες που αλλοιώνουν τις κύριες πληροφορίες των δεδομένων.
    (3) **Ανακάλυψη δομών γραφημάτων** (Discovering graph structures) για τον προσδιορισμό του βαθμού συσχέτισης μεταξύ διαφορετικών μεταβλητών.
    (4) **Ολοκλήρωση μήτρας** (Matrix completion) και **σχηματισμός δέσμης**(beam forming) όπου, καθίσταται δυνατή η εκτίμηση τιμών για δεδομένα που λείπουν ή την ανακατασκευή ενός σήματος, δηλαδή την εύρεση ενός προηγουμένως άγνωστου σήματος από αλλοιωμένες ή ατελείς παρατηρήσεις.

- **<u>Ημί-εποπτευόμενη μάθηση:</u>** Το σύνολο δεδομένων περιέχει δομημένα και μη δομημένα δεδομένα, τα οποία καθοδηγούν τον αλγόριθμο στην πορεία του προς την εξαγωγή ανεξάρτητων συμπερασμάτων. Ο συνδυασμός των δύο τύπων δεδομένων σε ένα σύνολο δεδομένων εκπαίδευσης επιτρέπει στους αλγορίθμους μηχανικής μάθησης να μάθουν να επισημαίνουν τα μη επισημασμένα δεδομένα.
- **<u>Ενισχυτική μάθηση:</u>** Το σύνολο δεδομένων χρησιμοποιεί ένα σύστημα "ανταμοιβών/τιμωριών", προσφέροντας ‘’κριτική’’ στον αλγόριθμο για να μάθει από τις δικές του εμπειρίες μέσω δοκιμής και λάθους.

Τέλος, υπάρχει η έννοια της βαθιάς μάθησης (Deep Learning), η οποία είναι ένας νεότερος τομέας της μηχανικής μάθησης που μαθαίνει αυτόματα από σύνολα δεδομένων χωρίς την εισαγωγή ανθρώπινων κανόνων ή γνώσεων. Αυτό απαιτεί τεράστιες ποσότητες ακατέργαστων δεδομένων για επεξεργασία - και όσο περισσότερα δεδομένα λαμβάνονται, τόσο περισσότερο βελτιώνεται το μοντέλο πρόβλεψης.

3.1.3.Διαδεδομένοι αλγόριθμοι μηχανικής μάθησης

Σκοπός της μηχανικής μάθησης είναι η χρήση αλγορίθμων μηχανικής μάθησης για την ανάλυση δεδομένων. Αξιοποιώντας τη μηχανική μάθηση, ο χρήστης μπορεί να βελτιώσει την αποτελεσματικότητα μιας εργασίας που περιλαμβάνει μεγάλες ποσότητες δεδομένων χωρίς την ανάγκη χειροκίνητης ανθρώπινης εισόδου. Σε όλο τον κόσμο, οι ισχυροί αλγόριθμοι μηχανικής μάθησης μπορούν να χρησιμοποιηθούν για τη βελτίωση της παραγωγικότητας των επιστημόνων που εργάζονται στην επιστήμη των δεδομένων, στην επιστήμη των υπολογιστών και σε πολλούς άλλους τομείς.

- **<u>Γραμμική παλινδρόμηση (Linear Regression)</u>**

Η γραμμική παλινδρόμηση είναι ένας αλγόριθμος που χρησιμοποιείται για την ανάλυση της σχέσης μεταξύ ανεξάρτητων μεταβλητών εισόδου και τουλάχιστον μιας μεταβλητής στόχου. Αυτό το είδος παλινδρόμησης χρησιμοποιείται για την πρόβλεψη συνεχών αποτελεσμάτων - μεταβλητών που μπορούν να λάβουν οποιοδήποτε αριθμητικό αποτέλεσμα. Για παράδειγμα, με δεδομένα για τη γειτονιά και ένα ακίνητο, μπορεί ένα μοντέλο να προβλέψει την αξία πώλησης αυτού του ακινήτου (Berkeley, School of Information, 2020).

Οι γραμμικές σχέσεις εμφανίζονται όταν η παρατηρούμενη σχέση δεδομένων τείνει να ακολουθεί συνολικά μια ευθεία γραμμή και ως εκ τούτου, αυτό το μοντέλο μπορεί να χρησιμοποιηθεί για να παρατηρηθεί εάν ένα σημείο δεδομένων αυξάνεται, μειώνεται ή παραμένει το ίδιο σε σχέση με κάποια ανεξάρτητη μεταβλητή, όπως ο χρόνος που παρήλθε ή η θέση του σημείου δεδομένων (Berkeley,School of Information,2020).

Τα μοντέλα μηχανικής μάθησης μπορούν να χρησιμοποιηθούν για την ανάλυση δεδομένων προκειμένου να παρατηρηθούν και να χαρτογραφηθούν οι γραμμικές παλινδρομήσεις. Οι ανεξάρτητες μεταβλητές και οι μεταβλητές-στόχοι μπορούν να εισαχθούν σε ένα μοντέλο μηχανικής μάθησης γραμμικής παλινδρόμησης, όπου το μοντέλο θα χαρτογραφήσει τους συντελεστές της γραμμής που προσαρμόζεται καλύτερα στα δεδομένα. Με άλλα λόγια, τα μοντέλα γραμμικής παλινδρόμησης προσπαθούν να χαρτογραφήσουν μια ευθεία γραμμή ή μια γραμμική σχέση μέσω του συνόλου δεδομένων (Berkeley,School of Information,2020).

Τα μοντέλα γραμμικής παλινδρόμησης είναι συνήθως απλά και παρέχουν μία εύκολα κατανοητή συνάρτηση για την πραγματοποίηση προβλέψεων. Οι γραμμική παλινδρόμηση χρησιμοποιείται σε πληθώρα επιστημονικών πεδίων, από τη βιολογία και τα οικονομικά έως τις κοινωνικές/ ανθρωπιστικές επιστήμες, καθώς έχει κατοχυρωθεί ως ένας αποδεδειγμένα αξιόπιστος τρόπος για την πραγματοποίηση προβλέψεων (IBM Cloud Education, 2020).

- **<u>Λογιστική Παλινδρόμηση (Logistic Regression)</u>**

Η λογιστική παλινδρόμηση είναι ένας αλγόριθμος μάθησης με επίβλεψη που χρησιμοποιείται για προβλήματα ταξινόμησης. Αντί για συνεχή έξοδο όπως στη γραμμική παλινδρόμηση, ένα λογιστικό μοντέλο προβλέπει την πιθανότητα εμφάνισης ενός δυαδικού γεγονότος. Για παράδειγμα, μελετώντας ένα μήνυμα ηλεκτρονικού ταχυδρομείου, μπορεί ένα μοντέλο να προβλέψει αν το περιεχόμενό του είναι Spam ή όχι;

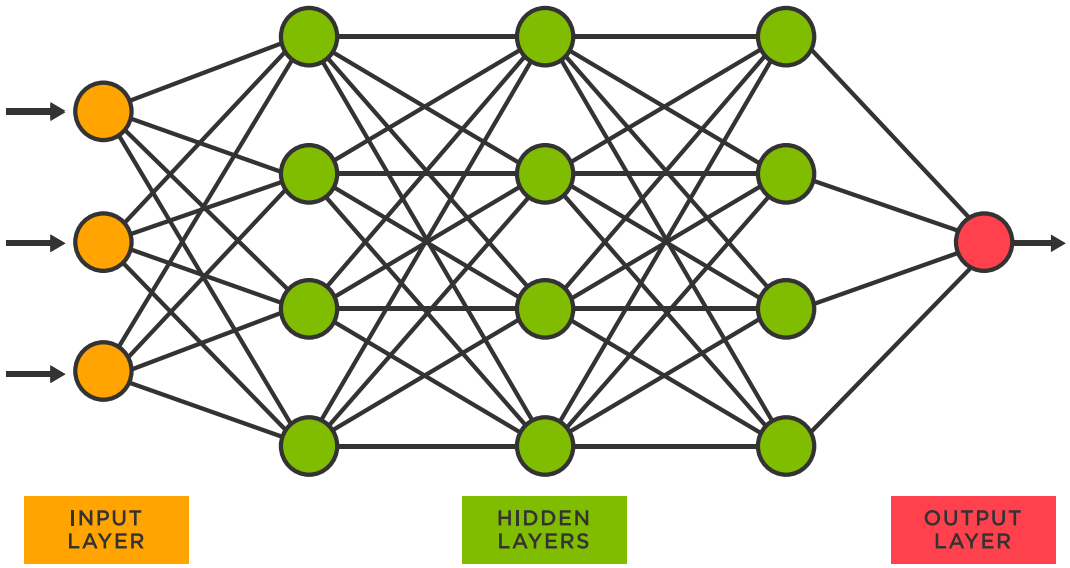
Οι αλγόριθμοι μηχανικής μάθησης μπορούν να χρησιμοποιούν μοντέλα λογιστικής παλινδρόμησης για τον προσδιορισμό αποτελεσμάτων ταξινόμησης σε κατηγορίες. Όταν δίνεται ένα σύνολο δεδομένων, το μοντέλο λογιστικής παλινδρόμησης μπορεί να ελέγξει τυχόν ‘’βάρη’’ και ‘’προκαταλήψεις’’ στα δεδομένα και στη συνέχεια να χρησιμοποιήσει τις εισηγμένες μεταβλητές κατηγοριοποίησης για να κατανοήσει πώς να κατηγοριοποιήσει σωστά το εν λόγω σύνολο δεδομένων (Berkeley,School of Information,2020).

- **<u>Νευρωνικά δίκτυα (Neural Networks)</u>**

Η ανάπτυξη τεχνητών νευρωνικών δικτύων (Artificial Neural Networks) ξεκίνησε ως μια προσπάθεια κατανόησης της λειτουργίας του ανθρώπινου εγκεφάλου και μίμησης της αξιολογητικής του ικανότητας και αποτελούν ένα υποσύνολο του κλάδου της μηχανικής μάθησης (machine learning) και επίκεντρο της υποκατηγορίας των αλγορίθμων βαθιάς μάθησης (deep learning algorithms). Με άλλα λόγια, τη ανάπτυξη των Νευρωνικών Δικτύων με τέτοιο τρόπο, ώστε να μπορούν να αποφασίζουν και να δρουν υπό συνθήκες αβεβαιότητας ή να επιλύουν προβλήματα έχοντας περιορισμένη εμπειρία πάνω σε αυτά. Τα Τεχνητά Νευρωνικά Δίκτυα είναι μαθηματικά μοντέλα τα οποία αποτελούνται από διασυνδεδεμένους κόμβους επεξεργασίας (Νευρώνες) με προκαθορισμένη τοπολογία (επίπεδα) (Benardos, 2008).

Τα νευρωνικά δίκτυα παρουσιάζουν μεγάλη ομοιότητα με τον βιολογικό εγκέφαλο και για αυτό το λόγο, μεγάλο μέρος της ορολογίας που χρησιμοποιείται σε αυτά, πηγάζει από τη νευροεπιστήμη. Το βασικό τους χαρακτηριστικό είναι η ικανότητα τους να πραγματοποιούν μαζική παράλληλη επεξεργασία των δεδομένων που λαμβάνουν (input data), σε αντίθεση με τα συνήθη μαθηματικά μοντέλα, τα οποία βασίζονται σε σειριακή επεξεργασία μαθηματικών και λογικών πράξεων. Ακόμη, ένα σημαντικό πλεονέκτημα των Νευρωνικών δικτύων αποτελεί η ευελιξία τους κατά την επεξεργασία δεδομένων, όπου δεν απαιτούν κάποια ντετερμινιστική μαθηματική σχέση μεταξύ των εξεταζόμενων συνιστωσών, αλλά λειτουργούν ως εξής: Με την εισαγωγή δεδομένων, υπό μορφή αιτίου-αποτελέσματος, το νευρωνικό δίκτυο τακτοποιεί τις υπάρχουσες συσχετίσεις μεταξύ των δεδομένων, και εν συνεχεία, ''μαθαίνει'' και ''μιμείται'' τη συμπεριφορά αυτών των συσχετίσεων, ρυθμίζοντας τους συντελεστής βαρύτητας αυτών των σχέσεων ή αλλιώς, ρυθμίζοντας την ισχύ των δεσμών μεταξύ των νευρώνων. Γι' αυτό το λόγο, τα Νευρωνικά δίκτυα δεν μπορούν να προγραμματιστούν, αλλά ''εκπαιδεύονται'' και μαθαίνουν'' μέσω απόκτησης εμπειρίας από καταστάσεις. Ως αποτέλεσμα, αφού ολοκληρωθεί η εκπαίδευση ενός νευρωνικού δικτύου, με ένα ήδη υπαρκτό σύνολο δεδομένων, μπορούν να υπολογιστούν εκτιμήσεις για ένα διαφορετικό σύνολο εισαγόμενων δεδομένων. Το νευρωνικό δίκτυο, αφού εκπαιδευτεί, έχει την ικανότητα να δώσει σχετικά ακριβείς εκτιμήσεις ακόμη και υπό συνθήκες αβεβαιότητας ή ημιτελών συνόλων δεδομένων. Το κύριο μειονέκτημα των Τεχνητών Νευρωνικών Δικτύων είναι η αδυναμία τους να προσαρμόσουν και να ρυθμίσουν τα συντελεστές βαρύτητας των παραμέτρων με ακρίβεια, ή και σε πολλές περιπτώσεις να αδυνατούν εξαρχής να δώσουν μια εκτίμηση για τους συντελεστές αυτούς, όταν η δομή του δικτύου είναι εκτενής και πολύπλοκη.

Σε γενική βάση, η τυπική τοπολογία ενός Τεχνητού Νευρωνικού δικτύου συνίσταται από ένα σύστημα επιπέδων αποτελούμενο, από το επίπεδο (layer) εισόδου δεδομένων (input), ένα ή περισσότερα ''κρυφά'' υπό-επίπεδα και το επίπεδο εξόδου (output), τα οποία περιέχουν ένα συγκεκριμένων αριθμό νευρώνων (Εικ. 13). Από εκεί και πέρα, κάθε νευρώνας συνδέεται με γειτονικούς του με διαφορετικούς συντελεστές συνδεσιμότητας , που αντιπροσωπεύουν τη βαρύτητα αυτών των συνδέσεων, ενώ ταυτόχρονα περιέχει και ένα κατώτατο όριο. Όταν μια τιμή των δεδομένων εξόδου ενός νευρώνα υπερβαίνει το κατώτατο όριο, τότε ο νευρώνας ενεργοποιείται, στέλνοντας δεδομένα στο επόμενο επίπεδο του δικτύου (IBM Cloud Learn Hub,2020). Επίσης, κάθε νευρώνας των κρυφών υπό-επιπέδων (hidden layers) είναι διασυνδεμένος με όλους τους νευρώνες των επιπέδων εισόδου και εξόδου(Benardos 2008).



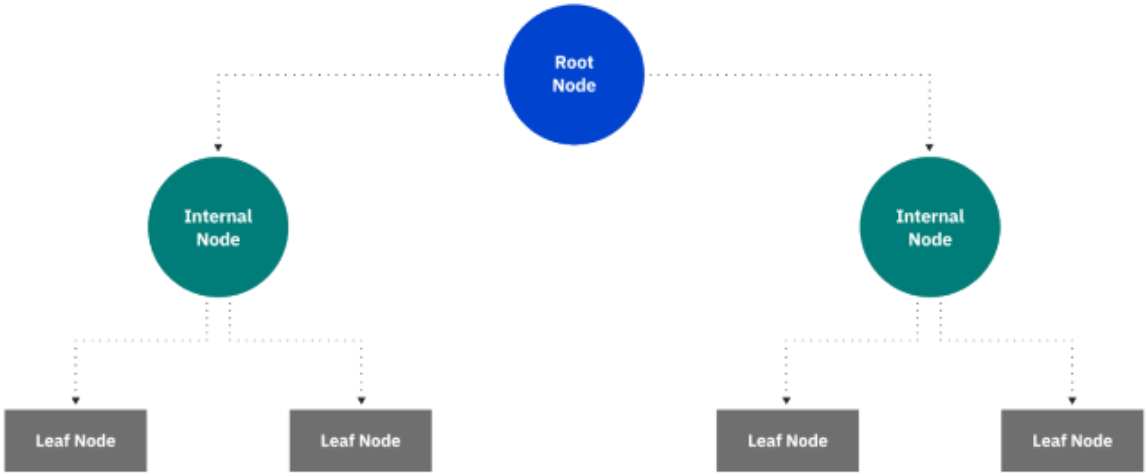
Εικόνα 13: Τυπική δομή ενός Τεχνητού Νευρωνικού Δικτύου (Πηγή: TIBCO, Neural Networks 2022)

Τα Νευρωνικά δίκτυα εξαρτώνται από τα δεδομένα εκπαίδευσης που θα τους δοθούν για να ‘’μάθουν’’ και να βελτιώσουν την ακρίβεια τους με το πέρασμα του χρόνου. Από τη στιγμή, όμως, πού αυτοί οι αλγόριθμοι μάθησης ρυθμιστούν σωστά για τη μέγιστη δυνατή ακρίβεια, αποτελούν ισχυρότατα εργαλεία, τα οποία επιτρέπουν την ταχύτατη ταξινόμηση (classifying) και ομαδοποίηση (clustering) μεγάλου όγκου δεδομένων. Χαρακτηριστικό παράδειγμα, μέσω της χρήσης Νευρωνικών Δικτύων, εργασίες αναγνώρισης ομιλίας ή εικόνας μπορούν να διαρκέσουν λίγα λεπτά έναντι ωρών σε σύγκριση με τη συμβατική αναγνώρισης από εμπειρογνώμονες (IBM Cloud Education, 2020).

Οι αλγόριθμοι μάθησης νευρωνικών δικτύων χρησιμοποιούνται για την αναγνώριση μοτίβων στην ανθρώπινη ομιλία, στη μεταγλώττιση γλωσσών και στην πραγματοποίηση οικονομικών προβλέψεων μέσω των χιλιάδων ή και εκατομμυρίων διασυνδεδεμένων νευρώνων επεξεργασίας.

- **<u>Δέντρα αποφάσεων (Decision Trees)</u>**

Το δέντρο αποφάσεων είναι ένας μη παραμετρικός αλγόριθμος μάθησης με επίβλεψη, ο οποίος χρησιμοποιείται τόσο για εργασίες ταξινόμησης όσο και για εργασίες παλινδρόμησης. Έχει μια ιεραρχική, δενδροειδή δομή, η οποία αποτελείται από έναν κόμβο ρίζας(root node) , κλάδους (branches), εσωτερικούς κόμβους και κόμβους φύλλων (leaf nodes) (IBM, 2020).



Εικόνα 14: Διάγραμμα μορφής δέντρου αποφάσεων (Decision tree), (Πηγή: IBM ,2020)

Όπως φαίνεται και στο παραπάνω διάγραμμα (Εικ. 14), ένα δέντρο αποφάσεων ξεκινά με έναν κόμβο ρίζας, ο οποίος δεν έχει εισερχόμενα κλαδιά. Οι εξερχόμενοι κλάδοι από τον κόμβο ρίζας τροφοδοτούν στη συνέχεια τους εσωτερικούς κόμβους, γνωστούς και ως κόμβους απόφασης. Με βάση τα διαθέσιμα χαρακτηριστικά, και οι δύο τύποι κόμβων διεξάγουν αξιολογήσεις για να σχηματίσουν ομοιογενή υποσύνολα, τα οποία συμβολίζονται με κόμβους φύλλων ή τερματικούς κόμβους. Οι κόμβοι φύλλων αντιπροσωπεύουν όλα τα πιθανά αποτελέσματα εντός του συνόλου δεδομένων (IBM Cloud Education, 2020).

Η εκμάθηση δέντρων αποφάσεων χρησιμοποιεί μια στρατηγική "διαίρει και βασίλευε" με τη διεξαγωγή μιας ''άπληστης αναζήτησης'' (Greedy Search) για τον εντοπισμό των βέλτιστων σημείων διαχωρισμού μέσα σε ένα δέντρο. Αυτή η διαδικασία διάσπασης επαναλαμβάνεται στη συνέχεια με αναδρομικό τρόπο από πάνω προς τα κάτω, έως ότου όλα ή η πλειονότητα των εγγραφών ταξινομηθούν σε συγκεκριμένες ετικέτες κλάσης. Το κατά πόσον όλα τα σημεία δεδομένων ταξινομούνται ως ομοιογενή σύνολα εξαρτάται σε μεγάλο βαθμό από την πολυπλοκότητα του δέντρου απόφασης. Τα μικρότερα δέντρα είναι πιο εύκολα σε θέση να επιτύχουν αμιγείς κόμβους φύλλων, δηλαδή σημεία δεδομένων σε μία μόνο κλάση. Ωστόσο, καθώς ένα δέντρο μεγαλώνει σε μέγεθος γίνεται όλο και πιο δύσκολο να διατηρηθεί αυτή η καθαρότητα και αυτό συνήθως έχει ως αποτέλεσμα να εμπίπτουν πολύ λίγα δεδομένα σε ένα δεδομένο υπό-δέντρο. Όταν συμβαίνει αυτό, περιγράφεται ως κατακερματισμός δεδομένων

(data fragmentation) και μπορεί συχνά να οδηγήσει σε υπερ-προσαρμογή (Overfitting). « Με άλλα λόγια, τα δέντρα αποφάσεων θα πρέπει να προσθέτουν πολυπλοκότητα μόνο αν είναι απαραίτητο, καθώς η απλούστερη εξήγηση είναι συχνά η καλύτερη. Για να μειωθεί η πολυπλοκότητα και να αποφευχθεί η υπερβολική προσαρμογή, χρησιμοποιείται συνήθως το κλάδεμα, όπου πρόκειται για μια διαδικασία, η οποία αφαιρεί κλάδους που χωρίζονται σε χαρακτηριστικά με χαμηλή σημασία (IBM Cloud Education, 2020).

- **<u>‘‘Τυχαία δάση’’ (Random Forest)</u>**

Το τυχαίο δάσος είναι της ευρέως χρησιμοποιούμενος αλγόριθμος μηχανικής μάθησης ο οποίος συνδυάζει τα δεδομένα εξόδου πολλαπλών δέντρων απόφασης για να καταλήξει σε ένα ενιαίο αποτέλεσμα. Η ευκολία χρήσης και η ευελιξία του έχουν τροφοδοτήσει την υιοθέτησή του, καθώς χειρίζεται τόσο προβλήματα ταξινόμησης όσο και παλινδρόμησης.

Ο αλγόριθμος Random Forest αποτελεί επέκταση της μεθόδου «bagging», καθώς χρησιμοποιεί τόσο το «bagging» όσο και την τυχαιότητα των χαρακτηριστικών για τη δημιουργία ενός ασυσχέτιστου δάσους δέντρων απόφασης. Η τυχαιότητα των χαρακτηριστικών, επίσης γνωστή ως «feature bagging» (μέθοδος του τυχαίου υπό-χώρου) δημιουργεί ένα τυχαίο υποσύνολο χαρακτηριστικών, το οποίο εξασφαλίζει χαμηλή συσχέτιση μεταξύ των δέντρων απόφασης. Αυτή είναι μια βασική διαφορά μεταξύ των δέντρων απόφασης και των τυχαίων δασών. Ενώ τα δέντρα αποφάσεων εξετάζουν όλες τις πιθανές διαχωριστικές ομάδες χαρακτηριστικών, τα τυχαία δάση επιλέγουν μόνο ένα υποσύνολο αυτών των χαρακτηριστικών.

Οι αλγόριθμοι τυχαίου δάσους έχουν τρεις κύριες παραμέτρους, οι οποίες πρέπει να οριστούν πριν από την εκπαίδευση. Αυτές περιλαμβάνουν το μέγεθος των κόμβων, τον αριθμό των δέντρων και τον αριθμό των δειγματοληπτικών χαρακτηριστικών. Από εκεί και πέρα, ο ταξινομητής τυχαίου δάσους μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων παλινδρόμησης ή ταξινόμησης.

Ο αλγόριθμος «Random Forest» αποτελείται από μια συλλογή δέντρων απόφασης και κάθε δέντρο στο σύνολο αποτελείται από ένα δείγμα δεδομένων που αντλείται από ένα σύνολο εκπαίδευσης με αντικατάσταση, το οποίο ονομάζεται δείγμα «bootstrap». Από αυτό το δείγμα εκπαίδευσης, το ένα τρίτο αυτού του δείγματος τίθεται στην άκρη ως δεδομένα δοκιμής, γνωστό ως δείγμα «out-of-bag (oob)». Ένα άλλο παράδειγμα τυχαιότητας εισάγεται στη συνέχεια μέσω του «feature bagging», προσθέτοντας μεγαλύτερη ποικιλομορφία στο σύνολο δεδομένων και μειώνοντας τη συσχέτιση μεταξύ των δέντρων απόφασης. Ανάλογα με τον τύπο του προβλήματος, ο προσδιορισμός της πρόβλεψης θα διαφέρει. Τέλος, το δείγμα «oob» χρησιμοποιείται στη συνέχεια για διασταυρούμενη επικύρωση (cross validation), οριστικοποιώντας την εν λόγω πρόβλεψη.

Συμπληρωματικά στην κύρια μεθοδολογία των αλγορίθμων μηχανικής μάθησης, χρησιμοποιούνται και τεχνικές προ-επεξεργασίας των δεδομένων προτού εισαχθούν στον αλγόριθμο μηχανικής μάθησης, με σκοπό να ελαχιστοποιηθούν οι ανακρίβειες, που μπορεί να προκύψουν από τη μορφή των δεδομένων όπως:

- **Περιορισμένη ποσότητα δεδομένων για εκπαίδευση**: Χρησιμοποιούνται τεχνικές διασταυρούμενης επικύρωσης «cross-validation», όπου τα διαθέσιμα δεδομένα διαιρούνται σε δύο ομάδες, μία για μάθηση και μία για επικύρωση της συμπεριφοράς του δικτύου. Προκειμένου να αποκτηθεί καλύτερη γνώση του αλγορίθμου, του μέγεθος και του αριθμού των στοιχείων μπορεί να τροποποιηθεί για την εκπαίδευση και για την αξιολόγηση του δικτύου σε διάφορες καταστάσεις.

- **Ανισοβαρή δεδομένα**: Πρόβλημα το οποίο εμφανίζεται κατά τη εκπαίδευση του δικτύου, συνήθως σε μοντέλα ταξινόμησης όπου (classification) εμφανίζονται περισσότερα δεδομένα σε μία κλάση των δεδομένων από ότι σε μία άλλη. Υπάρχουν πολλές τεχνικές επίλυσης αυτού του φαινομένου, οι οποίες επικεντρώνονται στο επίπεδο των δεδομένων (μέθοδοι δειγματοληψίας), είτε στο επίπεδο του ταξινομητή (εσωτερικές τροποποιήσεις). Οι μέθοδοι δειγματοληψίας στην ανισόρροπη μάθηση, προσπαθούν να τροποποιήσουν το ανισοβαρές σύνολο δεδομένων με κάποιους μηχανισμούς προκειμένου να παρέχουν μια ισορροπημένη κατανομή λαμβάνοντας υπόψιν τα αντιπροσωπευτικά ποσοστά των παραδειγμάτων της κλάσης στην κατανομή δεδομένων.
- **Ελλιπή δεδομένα**: Σε μερικές περιπτώσεις, η συλλογή δεδομένων για την επίλυση μίας συγκεκριμένης εργασίας είναι δυνατή, αλλά χαρακτηρίζεται ως ελλιπής εξαιτίας απόντων δεδομένων ή ελλιπούς γνώσης πάνω στα χαρακτηριστικά των δεδομένων. Η επίλυση αυτού του προβλήματος εστιάζει στην προσέγγιση των τιμών που λείπουν και στην ανακάλυψη μίας σχέσης που να συνδέει τα γνωστά και άγνωστα δεδομένα. Τέτοια προσέγγιση πραγματοποιεί η τεχνική Kernel Learning, η οποία βασίζεται, αυτή στη χρήση αλγορίθμων μηχανικής μάθησης για την αντιμετώπιση των ελλιπών δεδομένων.

- **Υψηλή διαστασιολόγηση δεδομένων**: Τα δεδομένα σε εφαρμογές του πραγματικού κόσμου είναι σε πολλές περιπτώσεις υπερβολικά άφθονα για την επίλυση ενός συγκεκριμένου προβλήματος. Λύση αυτού, αποτελεί η χρήση της μεθόδου «εξαγωγή λανθανόντων παραγόντων» που παρουσιάστηκε παραπάνω, όπου ανακαλύπτονται αυτοί οι παράγοντες (λανθάνοντες) σε μεγάλης έκτασης δεδομένα, μειώνοντας τις ‘’διαστάσεις’’ τους, προβάλλοντάς τα δεδομένα σε έναν υπό-χώρο μικρότερης διάστασης, με αποτέλεσμα να εξάγεται η πραγματική ‘’ουσία’’ των δεδομένων.

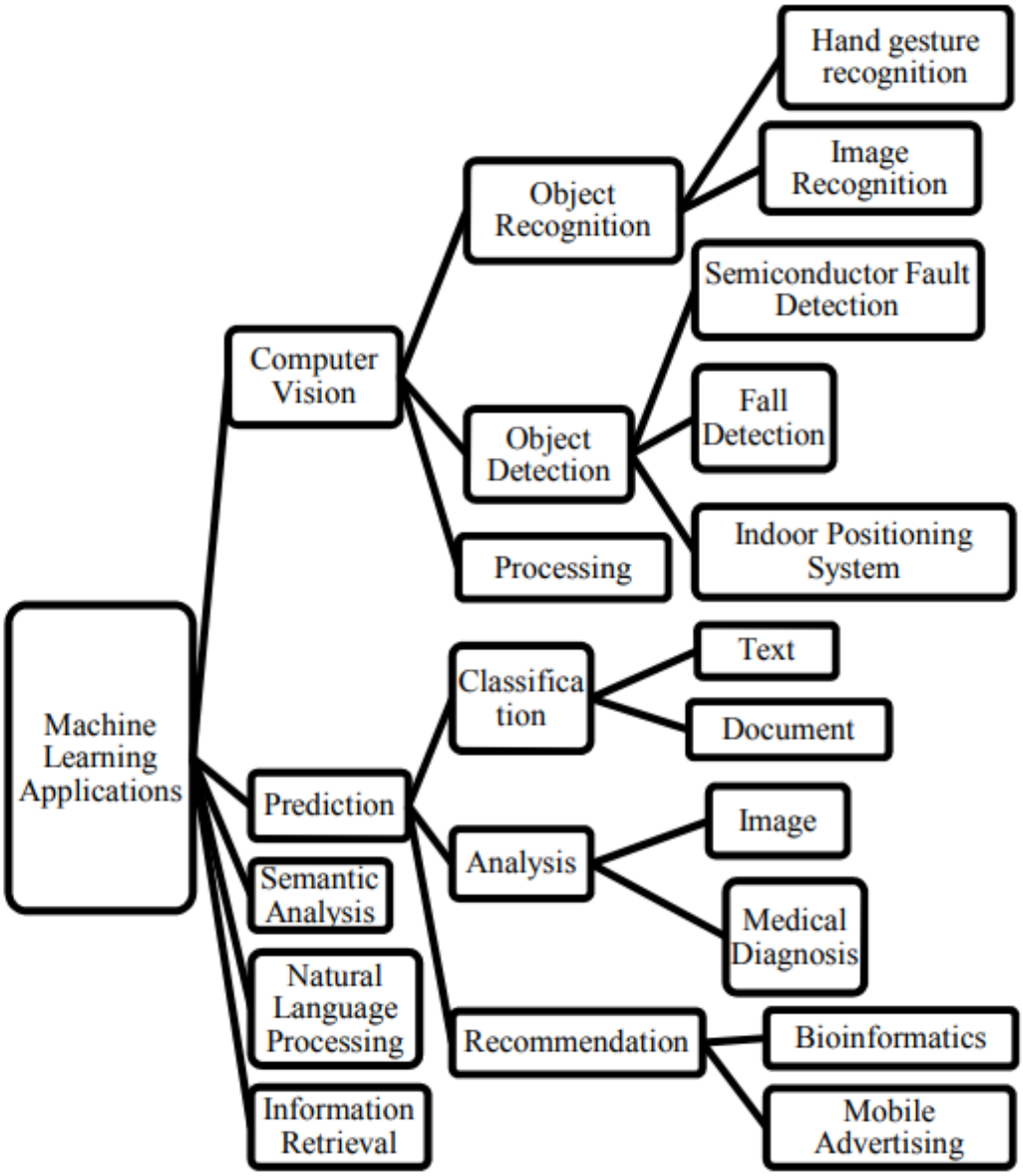
3.1.4. Εφαρμογές Αλγορίθμων Μηχανικής μάθησης

Οι αλγόριθμοι μηχανικής μάθησης αποτελούν ένα απαραίτητο εργαλείο για πολλούς ερευνητές, εταιρίες και οργανισμούς για δύο λόγους:

- **Κλίμακα δεδομένων**: Οι εταιρείες αλλά και πολλές ερευνητικές μελέτες έρχονται αντιμέτωπες με τεράστιους όγκους και τύπους δεδομένων που πρέπει να επεξεργαστούν. Η επεξεργαστική ισχύς των αλγορίθμων μηχανικής μάθησης είναι πιο αποτελεσματική και άμεσα διαθέσιμη. Τα μοντέλα μπορούν να προγραμματιστούν ώστε να επεξεργάζονται αυτόνομα τα δεδομένα, να καθορίζουν συμπεράσματα και να εντοπίζουν μοτίβα υψηλής σημασίας για την σύγχρονη ερευνητική και εταιρική κοινότητα.
- **Απροσδόκητα ευρήματα**: Δεδομένου ότι της αλγόριθμος μηχανικής μάθησης ενημερώνεται αυτόνομα, η αναλυτική ακρίβεια βελτιώνεται με κάθε εκτέλεση, καθώς μαθαίνει μόνος του από τα σύνολα δεδομένων που αναλύει. Αυτή η επαναληπτική φύση της μηχανικής μάθησης είναι μοναδική και πολύτιμη καθώς, χωρίς ανθρώπινη παρέμβαση, οι αλγόριθμοι μηχανικής μάθησης μπορούν να αποκαλύψουν κρυφές γνώσεις χωρίς να έχουν προγραμματιστεί ειδικά για να το κάνουν.

Με την ανακάλυψη των παραπάνω δυνατότητων από την επιστημονική κοινότητα, οι αλγόριθμοι μηχανικής μάθησης ξεκίνησαν να δοκιμάζονται για την επίλυση προβλημάτων του πραγματικού κόσμου από τις αρχές της δεκαετίας του ’90 σε συνδυασμό με άλλες τεχνικές επεξεργασίας δεδομένων, όπως στατιστική θεωρία μάθησης (statistical learning theory) και τη θεωρίας της πληροφορίας (Information theory). Πλέον, τα συστήματα επεξεργασίας που βασίζονται σε αλγόριθμους μηχανικής μάθησης θεωρούνται τυποποιημένα και ιδιαίτερα χρήσιμα εργαλεία για την επίλυση προβλημάτων με μεγάλο εύρος. Χαρακτηριστικά, μερικά από τα πεδία εφαρμογής των αλγορίθμων μηχανικής μάθησης είναι η διαγνωστική ιατρική, η βιολογία, η διαφημιστική βιομηχανία, η αυτόνομη οδήγηση κ.α..

Στο παρακάτω διάγραμμα (Εικ. 15) παρουσιάζονται μερικές εφαρμογές των αλγορίθμων μηχανικής μάθησης, κατηγοριοποιημένες σύμφωνα με τους Shinde & Shah (2018), ανά τομείς και υπό-τομείς εφαρμογών.



Εικόνα 15: Εφαρμογές αλγορίθμων μηχανικής μάθησης. (Πηγή: Shinde & Shah 2018)

Εμβαθύνοντας στο παραπάνω διάγραμμα (Εικ. 15), οι Shinde & Shah (2018) αναφέρουν τους αλγορίθμους που χρησιμοποιήθηκαν για την ανάπτυξη των μοντέλων κάθε εφαρμογής. Πιο συγκεκριμένα:

1. **Αναγνώριση αντικειμένων (Object recognition):** Χρησιμοποιούνται οι αλγόριθμοι είναι οι Naive Bayes, k-Nearest Neighbor, Artificial Neural Networks. Οι εφαρμογές αναγνώρισης αντικειμένων περιλαμβάνουν την αναγνώριση εικόνας και αναγνώριση χειραψιών χεριών.
2. **Ανίχνευση αντικειμένων (Object detection):** Αλγόριθμοι Decision trees, Bagging, Naïve Bayes, k-Nearest Neighbor. Οι εφαρμογές ανίχνευσης αντικειμένων περιλαμβάνουν, ανίχνευση σφάλματος ημι-αγωγών, ανίχνευσης πτώσης αντικειμένων, Ανίχνευση εσωτερικής τοποθεσίας αντικειμένων.
3. **Ταξινόμηση κειμένου (Text Classification):** Πραγματοποιείται με τη χρήση των αλγορίθμων Naïve Bayes, Linear Classifier, Artificial Neural Networks.
4. **Ανίχνευση καρκίνου του μαστού (Breast cancer detection and diagnosis):** Χρήση Bayesian Network και Random Forest.

Επιπλέον παραδείγματα αποτελούν :

- η χρήση αλγορίθμων γραμμικής παλινδρόμησης και λογιστικής παλινδρόμησης από το εκπαιδευτικό σύστημα των Η.Π.Α. για την πρόβλεψη των μαθησιακών δεξιοτήτων και ψυχολογικών παραμέτρων των Αμερικάνων μαθητών.
- Η χρήση νευρωνικών δικτύων στο χρηματιστήριο, όπου νευρωνικά δίκτυα εκπαιδεύονται στην πρόβλεψη των τάσεων επενδυτικών οχημάτων (μετοχές, ομόλογα, κ.α.) σε πραγματικό χρόνο.

Οι αλγόριθμοι μηχανικής μάθησης εξαπλώνονται ταχύτατα και βρίσκουν πλέον χρήση σχεδόν σε κάθε κύριο επιστημονικό και επιχειρηματικό πεδίο, με εφαρμογές στα οικονομικά, την γεωργία, τις μεταφορές και το μάρκετινγκ. Αυτή η εκτεταμένη υιοθέτηση σε μεγάλο εύρος βιομηχανιών, αποτελεί πειστήριο της υψηλής αξίας των αλγορίθμων μηχανικής μάθησης.

Οπλισμένοι με πληροφορίες από τεράστια σύνολα δεδομένων, τα οποία συχνά αναλύονται σε πραγματικό χρόνο, οι οργανισμοί μπορούν να λειτουργούν πιο αποτελεσματικά και να αποκτούν ανταγωνιστικό πλεονέκτημα. Η συνεχής ψηφιοποίηση των περισσότερων τομέων της κοινωνίας και της βιομηχανίας συνεπάγεται την παραγωγή αυξανόμενων όγκων δεδομένων. Η ικανότητα απόκτησης γνώσεων από αυτά ογκώδη σύνολα δεδομένων είναι ένα από τα κλειδιά για την αντιμετώπιση ενός τεράστιου φάσματος ζητημάτων από τον αποτελεσματικότερο εντοπισμό και τη θεραπεία ασθενειών έως την καταπολέμηση εγκληματιών στον κυβερνοχώρο και την βελτιστοποίηση λειτουργίας μεγάλων εταιριών.

3.2. Αλγόριθμοι μηχανικής μάθησης στο ζήτημα της ενεργειακής φτώχειας

Βάσει του κεφαλαίου 3.1.4., γίνεται ξεκάθαρο ότι η χρήση των αλγορίθμων μηχανικής μάθησης έχει αυξηθεί δραματικά, ειδικά τα τελευταία είκοσι χρόνια, όπου η αλματώδης εξέλιξη στο χώρο της πληροφορικής και της ανάπτυξης συστημάτων τεχνητής νοημοσύνης, έχει φέρει μια επανάσταση στις συμβατικές μεθόδους ανάλυσης δεδομένων και φαίνεται να δίνει νέες δυνατότητες σε ένα τεράστιο εύρος επιστημονικών πεδίων. Φυσικά σε αυτά τα πεδία, περιλαμβάνεται και ο ενεργειακός τομέας, όπου οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούνται για έρευνες και αναλύσεις που αφορούν πληθώρα θεμάτων όπως τη μοντελοποίηση λειτουργίας ενεργειακών συστημάτων (Kalogirou, 2013; Booth et al., 1998; Wong et al., 2014) και τη βραχυπρόθεσμη πρόβλεψη ηλεκτρικών φορτίων για περιοχές (Hippert et al., 2001). Ωστόσο, αυτά τα άρθρα αφορούν κυρίως θέματα σχεδιασμού ενεργειακών συστημάτων ή του ηλεκτρικού δικτύου, ενώ για το φαινόμενο της ενεργειακής φτώχειας υπάρχει πολύ περιορισμένη εφαρμογή των αλγορίθμων μηχανικής μάθησης, πέραν κάποιων πολύ πρόσφατων εξαιρέσεων. Ειδικότερα, οι Rajic et al.(2020), παρουσίασαν ένα μοντέλο πρόβλεψης της ενεργειακής ζήτησης της Σερβίας όσον αφορά τον ηλεκτρισμό και τη θέρμανση, βασιζόμενοι σε δεκαπέντε παραμέτρους (πληθυσμός, ανεργία, μέσος μισθός, βιομηχανία κ.α.), στοχεύοντας σε μια μακρό-οικονομική ανάλυση του ζητήματος. Για αυτή την ανάλυση χρησιμοποίησαν ένα νευρωνικό δίκτυο επονομαζόμενο «Endocrine NARX neural network (NARX» το οποίο λειτουργούσε μέσω του MATLAB.

Επίσης, οι Longa et al. (2021), χρησιμοποίησαν ένα ταξινομητή (classifier) μηχανικής μάθησης ονόματι «XGBoost/ gradient boosting decision tree», έναν αλγόριθμο δέντρου αποφάσεων, προκειμένου να προβλέψουν τον κίνδυνο ενεργειακής φτώχειας των νοικοκυριών στην Ολλανδία, βάσει ενός εύρους κοινωνικοοικονομικών παραμέτρων (οικονομική αξία σπιτιού, ιδιοκτησιακό καθεστώς και ηλικία σπιτιού, μέγεθος νοικοκυριού, μέση πυκνότητα πληθυσμού). Στην ανάλυση τους αναφέρουν ότι η μηχανική μάθηση θα μπορούσε να χρησιμοποιηθεί ως ένα αποτελεσματικό μέσο για την παρακολούθηση της ενεργειακής φτώχειας και να βοηθήσει ουσιαστικά στο σχεδιασμό και την εφαρμογή κατάλληλων μέτρων πολιτικής.

Ακόμη πιο πρόσφατο και πιο σχετικό είναι το άρθρο των Khizar Abbas et al. (2022), όπου γίνεται μέτρηση της ακραίας ενεργειακής φτώχειας στις αναπτυσσόμενες χώρες της Αφρικής και της Ασίας. Πιο συγκεκριμένα, ανέπτυξαν και εκπαίδευσαν ένα τεχνητό νευρωνικό δίκτυο (Artificial Neural Network) εξοπλισμένο με ταξινομητή «Multilayer Perceptron approach» για ομαδοποιήσουν (classifier) και να προβλέψουν τις κοινωνικοοικονομικές παραμέτρους που επηρεάζουν τις υπό μελέτη περιοχές και χρησιμοποίησαν τον δείκτη MEPI (Multidimensional Energy Poverty Index) (Nussbaumer et al.2012) για τον υπολογισμό της ενεργειακής φτώχειας.

Τέλος, αναφέρονται οι μελέτες των Παπαδά (2021) και Papada & Kaliampakos (2022), στις οποίες βασίζεται και η παρούσα διπλωματική εργασία. Ειδικότερα, στις παραπάνω μελέτες χρησιμοποιείται το πρόγραμμα μηχανικής μάθησης «Waikato Environment for Knowledge Analysis» (WEKA), με σκοπό την πρόβλεψη αντικειμενικών δεικτών ενεργειακής φτώχειας βάσει επιλεγμένων υποκειμενικών δεικτών: Αδυναμία να διατηρηθεί το σπίτι ζεστό (Inability to heat home), καθυστερήσεις στους λογαριασμούς ενέργειας (Arrears in energy bills), προβλήματα υγρασίας/μούχλας (Damp problems), προβλήματα υγείας λόγω κακών συνθηκών θέρμανσης

(Health problems) και ο περιορισμός άλλων βασικών αναγκών προς ικανοποίησης των αναγκών θέρμανσης (Restriction of other needs). Ως αντικειμενικοί δείκτες υπολογισμού της ενεργειακής φτώχειας επιλέχθηκαν ο δείκτης 10% με βάση τις πραγματοποιηθείσες δαπάνες, ο δείκτης 10% με βάση το απαιτούμενο κόστος και ο δείκτης συμπίεσης των ενεργειακών αναγκών (Degree of Coverage of Energy Needs – DCEN). Όσον αφορά στη μέθοδο μηχανικής μάθησης που χρησιμοποιήθηκε, επιλέχθηκε ο ταξινομητής (classifier) «Multilayer Perceptron», δηλαδή ένα νευρωνικό δίκτυο που χρησιμοποιεί οπίσθια διάδοση (back propagation) για την εκπαίδευση του δικτύου.

3.3.WEKA

3.3.1.Εισαγωγικά

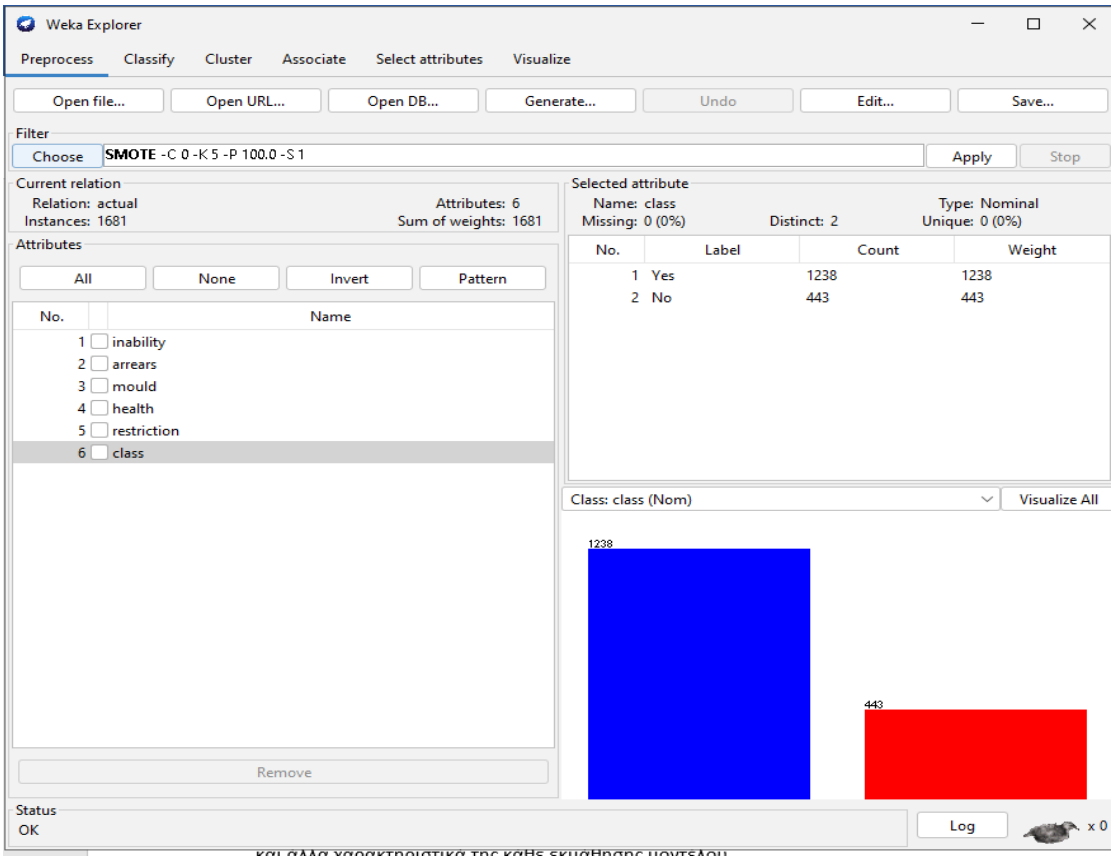
Το WEKA αποτελεί έναν χώρο εργασίας στους αλγορίθμους μηχανικής μάθησης που περιέχει μια συλλογή αλγορίθμων μηχανικής μάθησης και εργαλείων προ-επεξεργασίας δεδομένων. Έχει σχεδιαστεί με τέτοιο τρόπο ώστε να είναι δυνατό, με εύκολο και γρήγορο τρόπο, να δοκιμάζονται όλες οι διαθέσιμες μέθοδοι, χωρίς να είναι απαραίτητη η χρήση οποιασδήποτε μορφής προγραμματισμού (π.χ. γραφή κώδικα). Το πρόγραμμα αναπτύχθηκε στο Πανεπιστήμιο του Waikato στη Νέα Ζηλανδία υπό την ονομασία « Waikato Environment for Knowledge Analysis» (WEKA) και είναι γραμμένο σε γλώσσα προγραμματισμού Java. Παρέχει εύκολη υλοποίηση αλγορίθμων μηχανικής μάθησης για το εισαχθέν σύνολο δεδομένων, περιλαμβάνει εργαλεία για διάκριση (discretization) και δειγματοληψία (sampling) και προ επεξεργασία δεδομένων, ενώ δίνει τη δυνατότητα για ανάλυση της ταξινόμησης (classifying) που πραγματοποιήθηκε, μελετώντας την απόδοση και άλλα χαρακτηριστικά της κάθε εκμάθησης μοντέλου.

Πιο συγκεκριμένα υπάρχουν τρεις δυνατές χρήσεις του WEKA:

1. Εφαρμογή μεθόδων εκμάθησης σε σύνολο δεδομένων και ανάλυση των αποτελεσμάτων
2. Χρήση μεθόδων εκμάθησης για τη δημιουργία προβλέψεων σε νέες αντίστοιχες περιπτώσεις
3. Χρήση διαφορετικών μεθόδων εκμάθησης προς σύγκριση των επιδόσεων τους και επιλογή καταλληλότερου για πρόβλεψη

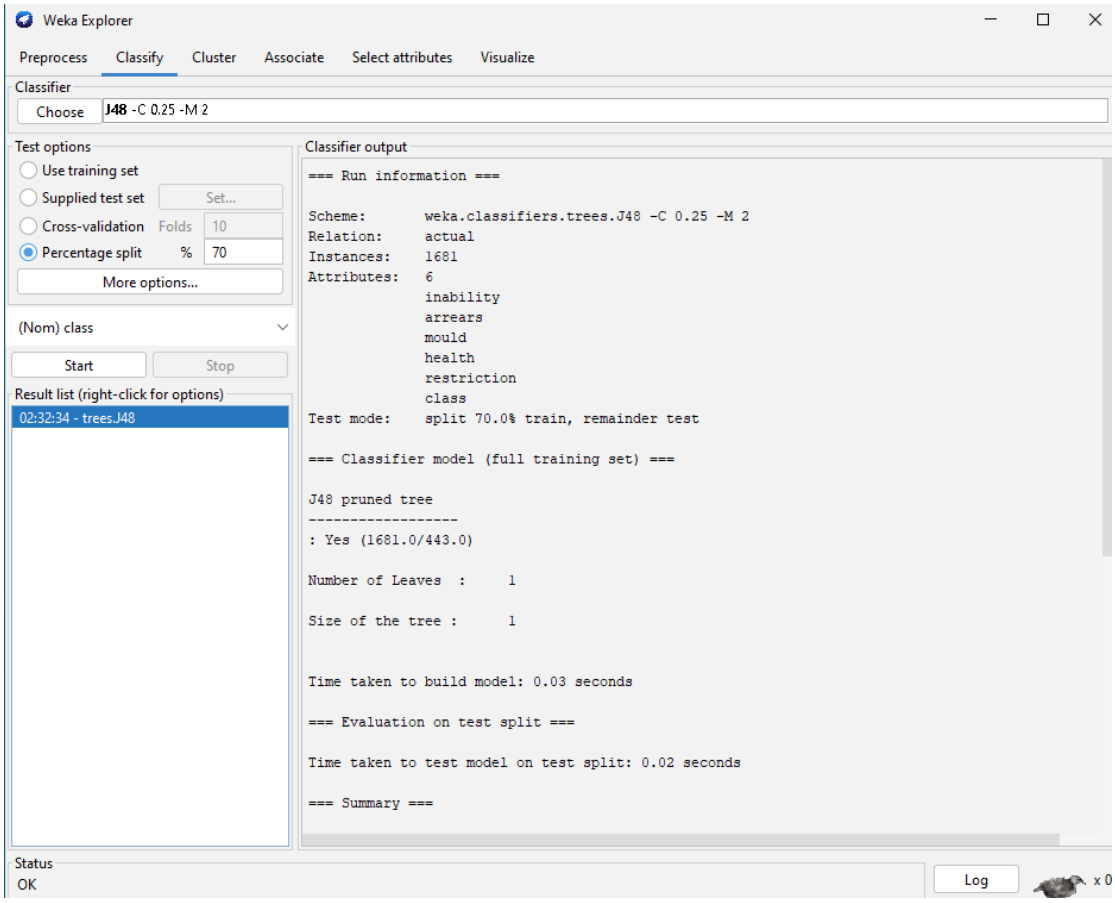
Το WEKA περιλαμβάνει πέντε διαφορετικά γραφικά περιβάλλοντα εργασίας για τον χρήστη: Explorer, Knowledge Flow, Experimenter, Workbench, command-line interfaces. Από αυτά, το Explorer, αποτελεί την πιο διαδεδομένη γραφική επιφάνεια, καθώς είναι εύχρηστη και μέσω αυτής είναι δυνατόν να εξαχθούν ακριβή αποτελέσματα σε μικρό χρονικό διάστημα, αν βέβαια, ο χρήστης έχει επίγνωση των ρυθμίσεων και των παραμέτρων που θέτει. Στη παρούσα εργασία, το Explorer αποτέλεσε την επιφάνεια εργασίας που χρησιμοποιήθηκε για την επεξεργασία των δεδομένων, όπως και στο άρθρο στο οποίο βασίζεται και η διπλωματική εργασία (Papada & Kaliampakos, 2022).

Στο επόμενο υπό-κεφάλαιο (3.3.2) θα περιγράφουν οι διαδικασίες προ-επεξεργασίας και οι αλγόριθμοί που χρησιμοποιήθηκαν για την επεξεργασία των δεδομένων στη συγκεκριμένη εργασία. Το WEKA περιλαμβάνει μεγάλη γκάμα διαθέσιμων μέσων προ-επεξεργασίας και αλγορίθμων μηχανικής μάθησης, ωστόσο για την παρούσα εργασία χρησιμοποιήθηκαν έξι διαφορετικοί αλγόριθμοι και ένα συγκεκριμένο μέσο προ επεξεργασίας, το οποίο κρίθηκε απαραίτητο για την διόρθωση των αρχικών δεδομένων, πριν την εισαγωγή τους στα μοντέλα αλγορίθμων μηχανικής μάθησης



Εικόνα 16:Κεντρική γραφική επιφάνεια προ-επεξεργασίας του Explorer. (Πηγή: Ίδια αρχεία)

Όπως φαίνεται στην παραπάνω εικόνα (Εικόνα 16), στην κεντρική γραφική επιφάνεια προ επεξεργασίας του Explorer, απεικονίζονται τα στοιχεία του συνόλου δεδομένων . Στην αριστερή στήλη βρίσκονται τα χαρακτηριστικά (attributes), όπου στη συγκεκριμένη περίπτωση περιλαμβάνουν πέντε υποκειμενικούς δείκτες (θα αναλυθούν στο επόμενο κεφάλαιο) και η κλάση(class) που καθορίζεται από τον αντικειμενικό δείκτη 10%. Στην δεξιά στήλη φαίνεται η μέτρηση των περιπτώσεων (instances) για κάθε χαρακτηριστικό (attribute), ενώ στο γράφημα από κάτω παρουσιάζονται γραφικά αυτές οι περιπτώσεις. Επιπλέον, στην συγκεκριμένη επιφάνεια βρίσκεται η περιοχή επιλογής φίλτρων προ επεξεργασίας. Αφού ολοκληρωθούν οι διαδικασίες φόρτωσης των δεδομένων, ελέγχου και εφαρμογής φίλτρων, η διαδικασία συνεχίζεται στην επιλογή ταξινομητή επιλέγοντας την γραφική επιφάνεια «classify».



Εικόνα 17: Γραφική επιφάνεια ''Classify'' του Explorer. (Πηγή: Ιδία αρχεία)

Στην γραφική επιφάνεια «classify» (Εικ. 17), γίνεται η επιλογή ταξινομητή (classifier) στην αντίστοιχη μπάρα όπως φαίνεται στην παραπάνω εικόνα και η διαχείριση διαθέσιμων επιλογών αναφορικά με το σύνολο των δεδομένων και την μέθοδο επιλογής των δεδομένων για την εκπαίδευση του επιλεγμένου αλγορίθμου.

Αφού ρυθμιστούν όλες οι παράμετροι έτσι ώστε να επεξεργασθεί και να εκπαιδευτεί το δίκτυο σωστά, δίνεται η εντολή στο πρόγραμμα, να ''τρέξει'' τον επιλεγμένο ταξινομητή. Τα δεδομένα εξόδου (classifier output) που εξέρχονται από τον ταξινομητή εμφανίζονται αναλυτικά στη δεξιά στήλη της εικόνας, όπου περιλαμβάνεται η μεθοδολογία που ακολούθησε ο επιλεγμένος ταξινομητής και δείκτες που αντιπροσωπεύουν την ακρίβεια των υπολογισμών για κάθε κλάση.

3.3.2.Προεπεξεργασία

Προ επεξεργασία και φίλτρα

Αρχικά, για την προ-επεξεργασία των δεδομένων χρησιμοποιήθηκαν φίλτρα για την διόρθωση χαρακτηριστικών των συλλεγμένων δεδομένων, τα οποία θα αλλοίωναν την ακρίβεια των αποτελεσμάτων που θα παρουσίαζε ο χρησιμοποιούμενος αλγόριθμος. Το WEKA περιλαμβάνει διαφορετικές κατηγορίες φίλτρων, για προ επεξεργασία ξεχωριστών μερών των δεδομένων εισόδου (input data), μεταμορφώνοντας ή αφαιρώντας τα με συγκεκριμένο τρόπο, έτσι ώστε να πάρουν την επιθυμητή μορφή, ρυθμίζοντας παραμέτρους οι οποίες επηρεάζουν τη δραστικότητα του χρησιμοποιούμενου φίλτρου.

Το πρόγραμμα περιλαμβάνει δύο τύπους φίλτρων, τα εποπτευόμενα (supervised) και μη εποπτευόμενα (unsupervised) φίλτρα, όπου τα εποπτευόμενα φίλτρα λαμβάνουν υπόψιν την τιμή κάθε κλάσης για εξαγωγή καλών διαστημάτων διάκρισης, ενώ τα μη εποπτευόμενα δεν ακολουθούν αυτή την μέθοδο. Ωστόσο, τα εποπτευόμενα φίλτρα απαιτούν ιδιαίτερη προσοχή στη χρήση τους, καθώς εάν χρησιμοποιηθούν, εφαρμόζονται και στο σύνολο εκπαίδευσης του δικτύου αλλά και στο δοκιμαστικό μέρος των δεδομένων, με αποτέλεσμα να επηρεάζονται τα αποτελέσματα αλλοιώνοντας την ακρίβεια τους.

Σε δεύτερο στάδιο, τα φίλτρα χωρίζονται σε κατηγορίες, ανάλογα σε ποιο μέρος των δεδομένων εφαρμόζονται. Έτσι, τα φίλτρα χωρίζονται σε: εποπτευόμενα ή μη εποπτευόμενα για τα χαρακτηριστικά (attributes) ή και για τις περιπτώσεις (instances).

Στη συγκεκριμένη εργασία χρησιμοποιήθηκε το εποπτευόμενο φίλτρο περιπτώσεων (Supervised Instance Filter) ονόματι SMOTE (Synthetic Minority Oversampling Technique). Το συγκεκριμένο φίλτρο χρησιμοποιείται σε περιπτώσεις ανισοβαρούς κατανομής των δεδομένων στις κλάσεις, με σκοπό τη διόρθωση της ανισορροπίας μεταξύ των κλάσεων. Αυτό επιτυγχάνεται λαμβάνοντας δείγματα από τα δεδομένα και φτιάχνοντας συνθετικές περιπτώσεις (Synthetic classes) για να μεταβάλλει την κατανομή των κλάσεων. Οι συνθετικές περιπτώσεις δημιουργούνται χρησιμοποιώντας την προσέγγιση k-nearest neighbor (k-κοντινότερων γειτόνων), όπου ρυθμίζεται η συχνότητα μεταξύ των κλάσεων, ενώ παράλληλα πραγματοποιείται υπο-δειγματοληψία της κλάσης πλειονότητας και υπέρ-δειγματοληψία της κλάσης μειονότητας. Ο χρήστης έχει τη δυνατότητα να καθορίσει το ποσοστό υπέρ-δειγματοληψίας και τον αριθμό των γειτόνων που θα χρησιμοποιηθούν για τη δημιουργία συνθετικών περιπτώσεων, όπως και έγινε όντως κατά την επεξεργασία των δεδομένων αυτής της εργασίας, όπου για την επίτευξη της καλύτερης δυνατής κατανομής δεδομένων έγινε ρύθμιση και των δύο παραμέτρων.

Μέθοδος επιλογής δεδομένων προς εκπαίδευση και δοκιμή

Στο WEKA υπάρχει η δυνατότητα χρήσης των δεδομένων με διαφορετικούς τρόπους για την εκπαίδευσης των αλγορίθμων μηχανικής μάθησης, όπου κάθε τρόπος είναι πιο κατάλληλος για χρήση σε διαφορετικές εφαρμογές, ανάλογα με τη μορφή των δεδομένων και τον χρησιμοποιούμενο ταξινομητή. Οι διαθέσιμες μέθοδοι χρήσης των δεδομένων για εκπαίδευση και δοκιμή είναι οι εξής:

- **Χρήση παρεχόμενου συνόλου δοκιμής (Supplied test set)**, όπου υπάρχει διαθέσιμο ξεχωριστό αρχείο με σύνολο δεδομένων αποκλειστικά για τη δοκιμή του μοντέλου.

- **Ποσοστιαίος διαχωρισμός του συνόλου των δεδομένων (Percentage (%) split)**, όπου επιλέγεται ένα μέρος των δεδομένων για εκπαίδευση και κρατείται το υπόλοιπο για δοκιμή.

- **N-fold cross validation**, όπου τα δεδομένα χωρίζονται τυχαία σε N ίσα μέρη. Από αυτά τα N μέρη, χρησιμοποιούνται τα N-1 μέρη χρησιμοποιούνται για την εκπαίδευση του αλγορίθμου, και το εναπομένον μέρος για δοκιμή. Η πιο ευρέως χρησιμοποιούμενη τιμή του N είναι 10 (10-fold cross validation), με τη συγκεκριμένη μέθοδο να είναι ιδιαίτερα δημοφιλής στις εφαρμογές «data mining».

- **Χρήση του συνόλου των δεδομένων, τόσο για την εκπαίδευση όσο και για τη δοκιμή του αλγορίθμου (Use training set)** . Δεν αποτελεί ιδιαίτερα αποδεκτή μέθοδο, καθώς ο αλγόριθμός απλώς ‘’απομνημονεύει’’ τα δεδομένα δοκιμής, χωρίς στη πραγματικότητα να '' μαθαίνει'' από το σύνολο των δεδομένων, ώστε να μπορεί να επεξεργαστεί άγνωστα δεδομένα.

Ο βασικός σκοπός της χρήσης αλγορίθμων μηχανικής μάθησης είναι, να δοκιμάζονται οι αλγόριθμοι σε δεδομένα διαφορετικά από αυτά που εκπαιδεύτηκαν, και όχι απλά να απομνημονεύουν τα στοιχεία των δεδομένων. Γι’ αυτό το λόγο, από τις παραπάνω μεθόδους, οι τρεις πρώτες είναι γενικά αποδεκτές στους περισσότερους τύπους δεδομένων και εφαρμογών, ενώ η χρήση του συνόλου των δεδομένων, τόσο για την εκπαίδευση όσο και για τη δοκιμή του αλγορίθμου (Use training set), υιοθετείται σε πολύ περιορισμένο επίπεδο. Για την παρούσα εργασία και την επεξεργασία των δεδομένων , επιλέχθηκε η μέθοδος percentage (%) split , με διαχωρισμό 70%/30%, δηλαδή, χρήση 70% των δεδομένων για εκπαίδευση των αλγορίθμων, και 30% για δοκιμή. Αυτή η μέθοδος επιλέχθηκε αρχικά λόγω της ευρείας αποδοχής της ως αξιόπιστης για εφαρμογές σε αλγορίθμους μηχανικής μάθησης, και κατά δεύτερον, για να είναι δυνατό να επιτευχθεί μια ταύτιση των συνθηκών υπό τις οποίες εκπαιδεύονται οι αλγόριθμοι, με τις συνθήκες που χρησιμοποιήθηκαν και στα άρθρα της Δρ. Παπαδά «Artificial intelligence in the field of energy poverty. First evidence for Greece,2021», ώστε εν τέλει, να είναι εφικτή η σύγκριση των αποτελεσμάτων και η εξαγωγή συμπερασμάτων σε κοινό πλαίσιο, όπως πραγματοποιείται στο κεφάλαιο 4 παρακάτω.

3.3.3.Ταξινομητές- Αλγόριθμοί μηχανικής μάθησης

Οι ταξινομητές στο WEKA χωρίζονται σε:

- «Bayesian» ταξινομητές (Bayesian classifiers)
- Δέντρα αποφάσεων (Decision Trees)
- Κανόνες (Rules)
- Meta-learning algorithms
- Συναρτήσεις (Functions)
- ''Τεμπέληδες'' ταξινομητές (Lazy classifiers)
- Κατηγορία διαφόρων (Miscellaneous)

Για την επιλογή ταξινομητών για την εργασία αυτή, έγινε μια συνολική δοκιμή έξι ταξινομητών, διαλέγοντας έναν ταξινομητή από κάθε κατηγορία, αποσκοπώντας στην συλλογή δεδομένων από την κάθε δοκιμή και την εξαγωγή συμπερασμάτων, αναφορικά με τον πιο αποτελεσματικό και κατάλληλο αλγόριθμο για την μορφή των υπαρχόντων δεδομένων. Επίσης, η δοκιμή ευρείας γκάμας αλγορίθμων μπορεί να δώσει κάποιες ενδείξεις, αναφορικά με την μέγιστη δυνατή ακρίβεια που μπορούν να δώσουν οι αλγόριθμοι μηχανικής μάθησης στο συγκεκριμένο θέμα μελέτης, και ίσως δώσουν κάποια σημαντικά στοιχεία που να δείχνουν το μέγιστο σημείο χρήσης ολόκληρης της μεθοδολογίας αλγορίθμων μηχανικής μάθησης για το θέμα της ενεργειακής φτώχειας, ενώ παράλληλα μπορούν να κρίνουν μερικώς, την ίδια την μέθοδο υπολογισμού της ενεργειακής φτώχειας. Συνεπώς, στο τελικό στάδιο της ανάλυσης των δεδομένων θα υπάρχουν στοιχεία τα οποία θα υποδεικνύουν αν υπάρχει χρησιμότητα και πραγματική ακρίβεια στις μεθόδους υπολογισμού της ενεργειακής φτώχειας.

Σε αυτό το σημείο, είναι απαραίτητο να αναφερθούν οι τύποι αλγορίθμων μηχανικής μάθησης που χρησιμοποιήθηκαν και να περιγράφει ο τρόπος λειτουργίας του κάθε αλγορίθμου.

1. Bayesian Classifiers

Naive Bayes: Το Θεώρημα Bayes με ανεξάρτητες υποθέσεις μεταξύ των προβλεπτών είναι η βασική έννοια του Naive Bayes ταξινομητή. Η απλούστερη προσέγγιση του δικτύου Bayes στο WEKA, είναι ο Naive Bayes στον οποίο όλα τα χαρακτηριστικά (Attributes) ενός συνόλου δεδομένων είναι ανεξάρτητα από την τιμή της μεταβλητής κλάσης του. Έτσι, ο Naïve Bayes είναι ένα δίκτυο Bayes όπου η κλάση(class) δεν έχει ''γονέα'', δηλαδή δεν λαμβάνει τις ιδιότητες της από κάποιο άλλο σύνολο δεδομένων και κάθε χαρακτηριστικό (attribute) έχει την κλάση(class) ως μοναδικό γονέα. Ο Naive Bayes δημιουργεί το μοντέλο ταξινόμησης χωρίς περίπλοκες επαναληπτικές εκτιμήσεις παραμέτρων. Λόγω αυτού του χαρακτηριστικού, οι ταξινομητές τύπου Bayes είναι χρήσιμοι και για πολύ μεγάλα σύνολα δεδομένων. Παρά την απλότητά τους, ο Naive Bayes παρέχει αξιόπιστα αποτελέσματα για πολύπλοκα προβλήματα του πραγματικού κόσμου (Yugal, 2012).

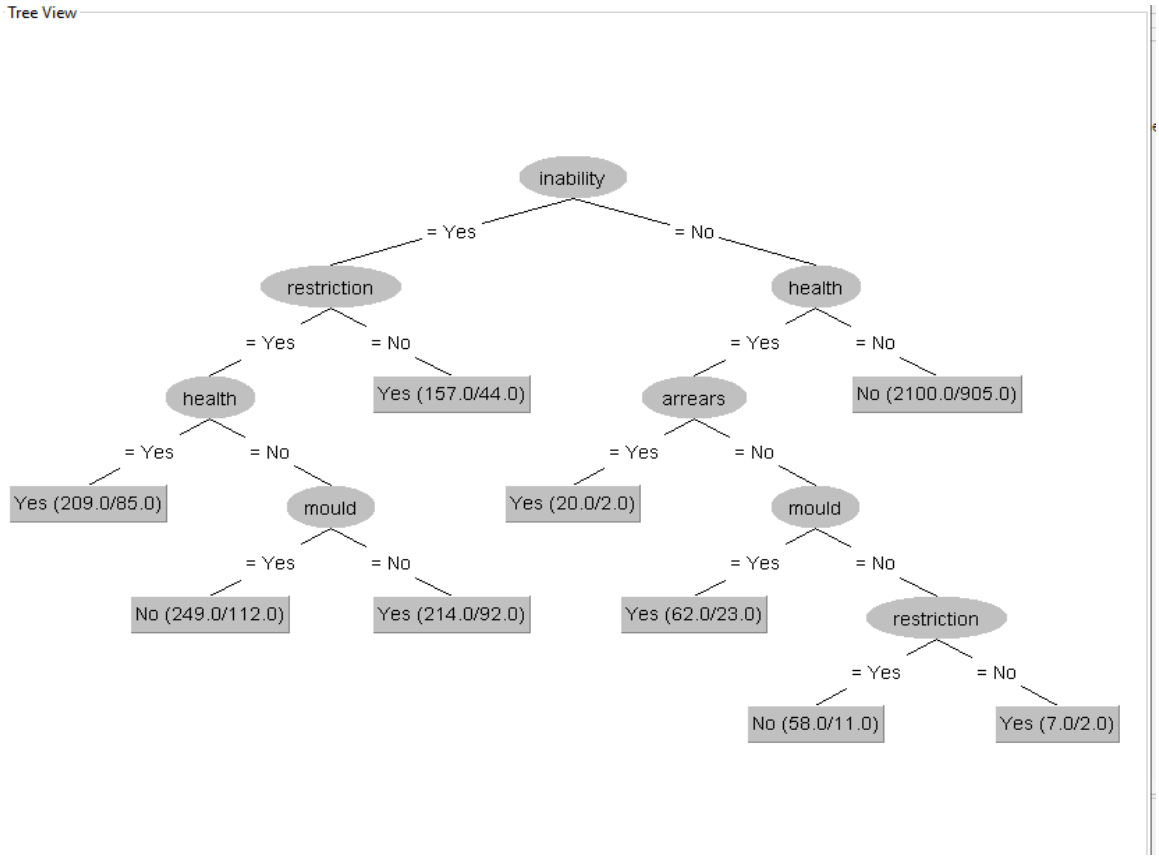
Ο Naive Bayes αποτελεί έναν μη παραμετρικό αλγόριθμο, ο οποίος μέσω χρήσης κανονικής κατανομής μοντελοποιεί τα αριθμητικά χαρακτηριστικά των δεδομένων, ενώ επίσης μπορεί να χρησιμοποιήσει τη μέθοδο «KDE» (kernel density estimation) για να βελτιώσει την απόδοσή του, αν η υπόθεση για κανονική κατανομή είναι λανθασμένη. Η «KDA» αποτελεί έναν μη παραμετρικό τρόπο εκτίμησης της συνάρτησης πυκνότητας πιθανότητας μιας τυχαίας μεταβλητής. Τέλος, ο συγκεκριμένος ταξινομήτης μπορεί να χειριστεί τόσο αριθμητικές, όσο και ονομαστικές τιμές.

2. Δέντρα αποφάσεων (Decision Trees)

J48: Ο αλγόριθμος ταξινόμησης J48 βασίζεται στον αλγόριθμο C4.5, επέκταση του ID3 του Ross Quinlan, ο οποίος αποτελεί ένα στατιστικό ταξινομητή μορφής δέντρου αποφάσεων. Ο J48 είναι εξαιρετικά χρήσιμος για την εξέταση δεδομένων κατά κατηγορία και συνεχόμενα. Στο WEKA ο J48 περιλαμβάνει πολλά πρόσθετα χαρακτηριστικά που μπορεί να ρυθμίσει ο χρήστης όπως, η εκτίμηση των ελλιπών τιμών, το ‘’κλάδεμα’’ (pruning) των δέντρων αποφάσεων και τα εύρη τιμών των χαρακτηριστικών (attribute values). Συγκεκριμένα, το ‘’κλάδεμα’’ αποτελεί μια ουσιώδη διαδικασία , η οποία αντιμετωπίζει το φαινόμενο του ‘’overfitting’’, όπου ο αλγόριθμος εκπαιδεύεται υπερβολικά καλά και μαθαίνει τις λεπτομέρειες και τα πυκνά δεδομένα υπερβολικά καλά, με αποτέλεσμα να μπορεί να αναλύσει μόνο τα δεδομένα εκπαίδευσης και να έχει πολύ χαμηλή απόδοση σε άγνωστα σύνολα δεδομένων. Το ‘’κλάδεμα’’ είναι μια διαδικασία με την οποία επιλέγεται το μεγαλύτερο δέντρο -το πιο γενικεύσιμο- και κλαδεύονται όλα τα κλαδιά κάτω από αυτό το επίπεδο. Αυτό βελτιώνει σημαντικά την απόδοση σε νέα δεδομένα.

Ο αλγόριθμος J48 κατασκευάζει δέντρα αποφάσεων με βάση ένα σύνολο δεδομένων εκπαίδευσης με τον ίδιο τρόπο που το κάνει ο αλγόριθμος ID3, χρησιμοποιώντας την έννοια της εντροπίας πληροφορίας. Τα δεδομένα εκπαίδευσης είναι ένα σύνολο S={s1, s2, ...} ήδη ταξινομημένων δειγμάτων. Κάθε δείγμα si αποτελείται από ένα p-διάστατο διάνυσμα (x1, i, x2, i, ..., xp, i) όπου το xj αντιπροσωπεύει τις τιμές των χαρακτηριστικών ή τα χαρακτηριστικά του αντίστοιχου δείγματος, καθώς και την κλάση στην οποία ανήκει το δείγμα. Για να επιτευχθεί η υψηλότερη ακρίβεια ταξινόμησης, το καλύτερο χαρακτηριστικό για διαχωρισμό είναι το χαρακτηριστικό με την περισσότερη πληροφορία.

Τα δέντρα αποφάσεων χρησιμοποιούνται για να περιγράψουν τις διαδικασίες λήψης αποφάσεων. Είναι ένας ταξινομητής που λειτουργεί σαν ένα διάγραμμα ροής όπως η κατασκευή δέντρων για την απεικόνιση μοντέλων συσχέτισης. Τα δέντρα αποφάσεων χρησιμοποιούνται για την κατηγοριοποίηση των περιπτώσεων με την ταξινόμησή τους κατά μήκος του δέντρου από την προέλευση σε έναν μικρό κόμβο φύλλου. Κάθε κόμβος προσδιορίζει μια εξέταση της περίπτωσης και κάθε διαίρεση αντιστοιχεί σε ένα από τα πιθανά οφέλη για αυτό το χαρακτηριστικό. Διαιρεί ένα σύνολο δεδομένων σε όλο και μικρότερα υποσύνολα και αναπτύσσεται σταδιακά. Το τελικό αποτέλεσμα είναι ένα δέντρο μαζί με κόμβους απόφασης και κόμβους φύλλων. Κάθε κόμβος απόφασης έχει δύο ή περισσότερες διαιρέσεις και ο κόμβος φύλλου ενσωματώνει μια συσχέτιση ή απόφαση. Σημαντικό χαρακτηριστικό του αλγορίθμου αποτελεί και η δυνατότητα γραφικής απεικόνισης του δέντρου αποφάσεων (Εικόνα 18), όπου απεικονίζεται ο τρόπος με τον οποίο κατηγοριοποιεί ο αλγόριθμος τα δεδομένα:



Εικόνα 18 Γραφική απεικόνιση των δεδομένων με τη χρήση ταξινομήτη J48.

Σε κάθε κόμβο του δέντρου, ο αλγόριθμος C4.5 επιλέγει το χαρακτηριστικό των δεδομένων που διαχωρίζει αποτελεσματικότερα το σύνολο των δειγμάτων σε υποσύνολα, εμπλουτισμένα στη μία ή την άλλη κατηγορία. Το κριτήριο διαχωρισμού είναι το κανονικοποιημένο κέρδος πληροφορίας, το οποίο υπολογίζεται από τη διαφορά της εντροπίας. Για τη λήψη της απόφασης επιλέγεται το χαρακτηριστικό με το μεγαλύτερο κανονικοποιημένο κέρδος πληροφορίας. Στη συνέχεια, ο αλγόριθμος C4.5 ανατρέχει στις διαιρεμένες υπό-λίστες και δημιουργεί ένα δέντρο απόφασης. Οι εσωτερικοί κόμβοι αυτού του δέντρου απόφασης υποδηλώνουν τα διάφορα χαρακτηριστικά, και οι διακλαδώσεις μεταξύ των κόμβων ενημερώνουν για τις πιθανές τιμές που μπορούν να έχουν. Το προβλεπόμενο χαρακτηριστικό είναι η εξαρτημένη μεταβλητή και τα άλλα χαρακτηριστικά στο δέντρο είναι οι ανεξάρτητες μεταβλητές του συνόλου δεδομένων.

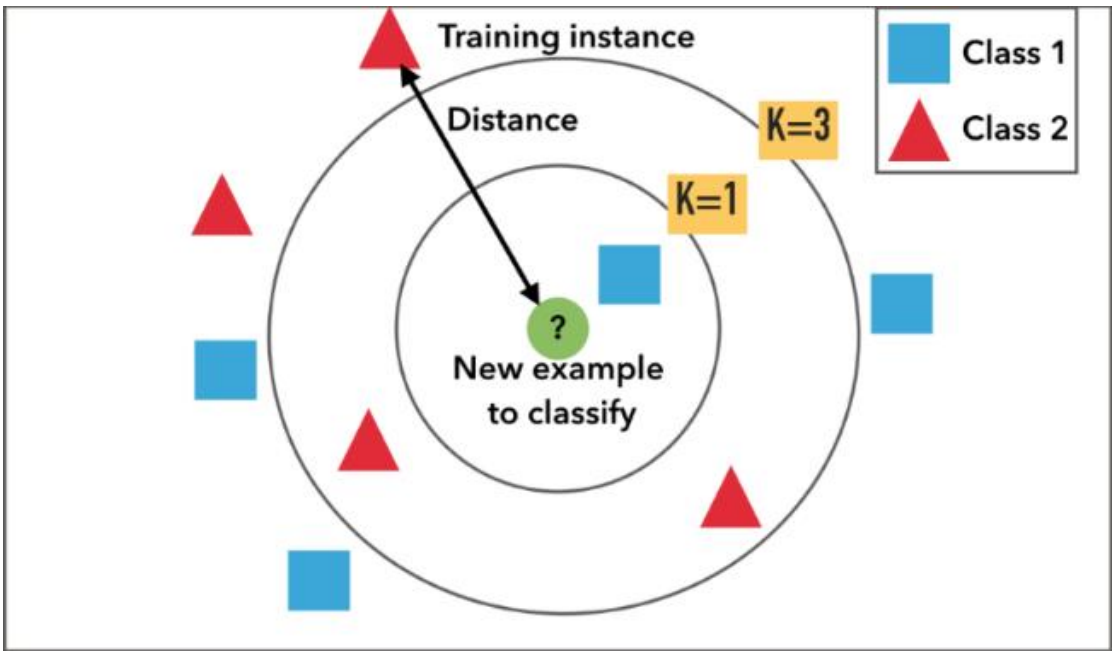
Κατά τη δημιουργία ενός δέντρου, το J48 αγνοεί τις ελλείπουσες τιμές με την τιμή αυτού του στοιχείου να μπορεί να προβλεφθεί από τις τιμές των χαρακτηριστικών για τις άλλες εγγραφές. Η βασική ιδέα είναι η διαίρεση των δεδομένων σε περιοχές με βάση τις τιμές των χαρακτηριστικών που βρέθηκαν στο δείγμα εκπαίδευσης.

3. Lazy Classifiers

IBk (Instance-based learning with parameter k): Ο αλγόριθμος IBk δεν κατασκευάζει ένα μοντέλο, αλλά παράγει μια πρόβλεψη για μια περίπτωση δοκιμής. Ουσιαστικά, αποτελεί τον αλγόριθμο kNN (k nearest neighbor). Κατά την ταξινόμηση kNN, μια τιμή ταξινομείται σύμφωνα με την πλειοψηφία πλήθους των γειτόνων του, με την τιμή να κατατάσσεται στην κλάση που είναι πιο κοινή μεταξύ των πλησιέστερων γειτόνων με το k να είναι θετικός ακέραιος αριθμός. Για παράδειγμα αν k=1, τότε η τιμή ταξινομείται στην κλάση του μοναδικού πλησιέστερου γείτονα.

Ο αλγόριθμος IBk χρησιμοποιεί ένα μέτρο απόστασης για να εντοπίσει k "κοντινές" περιπτώσεις στα δεδομένα εκπαίδευσης για κάθε περίπτωση δοκιμής και χρησιμοποιεί αυτές τις επιλεγμένες περιπτώσεις για να κάνει μια πρόβλεψη. Μια ποικιλία διαφορετικών αλγορίθμων αναζήτησης μπορούν να χρησιμοποιηθούν για την επιτάχυνση του έργου της εύρεσης των πλησιέστερων γειτόνων. Η γραμμική αναζήτηση είναι η προεπιλεγμένη, αλλά περιλαμβάνουν kD-trees, ball trees και τα λεγόμενα "cover trees" (Beygelzimer et al., 2006).

Γραφική αναπαράσταση της μεθόδου ταξινόμησης kNN (IBk στο WEKA) σύμφωνα με τους k-κοντινότερους γείτονες παρουσιάζεται στο παρακάτω σχήμα (Εικόνα 19):



Εικόνα 19 : Γραφική αναπαράσταση ταξινόμησης k-nearest neighbor. (Πηγή: Towards Data science 2018)

4. Rules

Decision Table: Ο αλγόριθμός αυτός δημιουργεί ένα ταξινομητή με τη μορφή πίνακα αποφάσεων και αποτελεί έναν πολύ απλό ταξινομητή, όσον αφορά τη μεθοδολογία ταξινόμησης. Αξιολογεί υποσύνολα χαρακτηριστικών χρησιμοποιώντας την αναζήτηση με την πρώτη σε σειρά καλύτερη επιλογή και μπορεί να χρησιμοποιήσει διασταυρούμενη επικύρωση για την αξιολόγηση. Συμπληρωματική επιλογή επιτρέπει στον αλγόριθμο χρησιμοποιεί τη μέθοδο k-πλησιέστερου γείτονα για τον προσδιορισμό της κλάσης για κάθε περίπτωση που δεν καλύπτεται από μια εγγραφή του πίνακα αποφάσεων, εναντίον της συνολικής πλειοψηφίας του πίνακα, με βάση το ίδιο σύνολο χαρακτηριστικών. Ουσιαστικά, μέσω του πίνακα αποφάσεων δημιουργείται ένα σύστημα όπου περιλαμβάνονται όλα τα πιθανά σενάρια συνθηκών όπου μπορεί να γίνει η ταξινόμηση, απ' όπου ο αλγόριθμός επιλέγει αυτή με τη μεγαλύτερη ακρίβεια και πιθανότητα να συμβεί.

5. Meta-learning algorithms

Οι αλγόριθμοι meta-learning, μετατρέπουν τους αλγορίθμους ταξινόμησης σε πιο ισχυρά εργαλεία ταξινόμησης, βελτιώνοντας σημαντικά τις ικανότητες μάθησής τους. Αυτή η βελτίωση επιτυγχάνεται με διαφορετικές μεθόδους, όπως ο συνδυασμός διαφορετικών δεδομένων εξόδων από ξεχωριστούς ταξινομητές, αλλά και άλλες προσεγγίσεις. Το WEKA περιλαμβάνει πληθώρα αλγορίθμων meta-learning, οι οποίοι ακολουθούν διαφορετικές μορφές βελτίωσης των αλγορίθμων, με τις κυριότερες να είναι οι διαδικασίες : Boosting, Bagging and randomization, Combining classifiers, Cost-Sensitive learning. Από τους διαθέσιμους αλγορίθμους βελτίωσης, επιλέχθηκε ένας αλγόριθμος τύπου Boosting και ένας τύπου Cost Sensitive για εφαρμογή στα πλαίσια της ανάλυσης της διπλωματικής εργασίας, των οποίων η λειτουργία εξηγείται παρακάτω:

- **Boosting (Ενίσχυση):** Η ενίσχυση είναι μια γενική μέθοδος βελτίωσης που δημιουργεί έναν ισχυρό ταξινομητή από έναν αριθμό αδύναμων ταξινομητών. Αυτό γίνεται με τη δημιουργία ενός μοντέλου από τα δεδομένα εκπαίδευσης και στη συνέχεια με τη δημιουργία ενός δεύτερου μοντέλου που προσπαθεί να διορθώσει τα σφάλματα του πρώτου μοντέλου. Τα μοντέλα προστίθενται έως ότου γίνει βέλτιστη πρόβλεψη του συνόλου εκπαίδευσης ή έως ότου προστεθεί ένας μέγιστος αριθμός μοντέλων (Machine Learning Mastery, Brownlee 2020). Κλασσική περίπτωση αλγορίθμου boosting αποτελεί ο AdaBoost.M1 (ή AdaBoost).
  Ο AdaBoost.M1 μπορεί να χρησιμοποιηθεί για την ενίσχυση της απόδοσης οποιουδήποτε αλγορίθμου μηχανικής μάθησης, αλλά λειτουργεί πιο αποδοτικά σε αλγόριθμους με χαμηλή ικανότητα εκπαίδευσης. Έτσι, ο AdaBoost παραδίδει τη μέγιστη απόδοση όταν χρησιμοποιείται για να ενισχύσει δέντρων αποφάσεων (Decision Trees) (Machine Learning Mastery, Brownlee 2020). Σύμφωνα με τα παραπάνω, ο AdaBoost αποτέλεσε έναν εκ των αλγορίθμων που χρησιμοποιήθηκαν για την επεξεργασία των δεδομένων σε αυτή την εργασία, όπου χρησιμοποιήθηκε για να βελτιώσει τον αλγόριθμο μορφής πίνακα αποφάσεων «Decision Table». Στο επόμενο κεφάλαιο, αναλύονται τα αποτελέσματα εξόδου κάθε αλγορίθμου, όπου θα σχολιασθεί και η βελτίωση που

επετεύχθη από τον AdaBoost στον Decision Table, συγκρίνοντας τα αποτελέσματα του Decision Table στην περίπτωση που χρησιμοποιήθηκε χωρίς τη χρήση αλγορίθμου ενίσχυσης.

Cost-Sensitive learning: Αυτή η μέθοδος είναι από μόνη της, ένα υπό-πεδίο της μηχανικής μάθησης που λαμβάνει υπόψη το κόστος των σφαλμάτων πρόβλεψης κατά την εκπαίδευση ενός μοντέλου μηχανικής μάθησης. Η μέθοδος Cost-Sensitive learning αποτελεί ένα πεδίο που επικεντρώνεται στην εκπαίδευση και χρήση μοντέλων σε δεδομένα που επιβάλλουν ''άνισες ποινές'' ή ''κόστη'' όταν κάνουν προβλέψεις. Παραδοσιακά, οι αλγόριθμοι μηχανικής μάθησης εκπαιδεύονται σε ένα σύνολο δεδομένων και επιδιώκουν την ελαχιστοποίηση του σφάλματος, όπου μια σειρά από συναρτήσεις μπορούν να χρησιμοποιηθούν για τον υπολογισμό του σφάλματος ενός μοντέλου σε δεδομένα εκπαίδευσης. Ο όρος ''ελαχιστοποίηση του σφάλματος'' , αναφέρεται στο συμβατικό και λογικό στόχο κατά τη διάρκεια εκπαίδευσης ενός αλγορίθμου μηχανικής μάθησης να ελαχιστοποιηθούν τα σφάλματα του μοντέλου όταν ταξινομεί τα δεδομένα εκπαίδευσης. Επίσης, ο όρος ''cost'' χρησιμοποιείται για να περιγράψει την έννοια της ποινής για κάθε λανθασμένη πρόβλεψη του μοντέλου.
Το WEKA περιλαμβάνει τον ταξινομητή «CostSensitiveClassifier » στη κατηγορία meta-learning, όπου δίνεται η δυνατότητα να διορθώνονται οι λανθασμένες προβλέψεις που έκανε το μοντέλο, σταθμίζοντας εκ νέου τις περιπτώσεις εκπαίδευσης σύμφωνα με το συνολικό κόστος που ανατίθεται σε κάθε κλάση ή προβλέποντας την κλάση με το ελάχιστο αναμενόμενο κόστος λανθασμένης ταξινόμησης αντί για την πιο πιθανή. Αυτή η διαδικασία ανάθεσης κόστους, πραγματοποιείται παρατηρώντας το confusion matrix (βλ. Κεφ.3.3.4.) και αναθέτοντας κόστη για κάθε ένα από τα κελιά του μέσω του λεγόμενης μήτρας κόστους (cost matrix). Τα κόστη εισάγονται υπό τη μορφή συντελεστών σε κάθε κελί του «confusion matrix», ώστε να μπορεί να μεταβληθεί η ένταση του κόστους που επιβάλλεται σε κάθε κελί προβλέψεων, ανάλογα με το ποσοστό λανθασμένων προβλέψεων και να πραγματοποιηθεί η μέγιστη δυνατή βελτιστοποίηση του αλγορίθμου. Στην συγκεκριμένη εργασία ο CostSensitiveClassifier χρησιμοποιήθηκε για ενισχύσει την απόδοση του αλγορίθμου J48 (Decision Tree), από την οποία διαδικασία κατέστη δυνατή και η σύγκριση μεταξύ δύο αλγορίθμων meta learning αναφορικά με το ποιος εκ των δύο ενισχύει περισσότερο τον ενισχυόμενο αλγόριθμο.

6. Νευρωνικά Δίκτυα (Neural Networks)

Multilayer Perceptron: Το Multi-Layer Perceptron (MLP) είναι ένας τύπος τεχνητού νευρωνικού δικτύου που αποτελείται από μια μη-γραμμική συνάρτηση ενεργοποίησης στο κρυφό στρώμα (hidden layer). Το δίκτυο MLP παρέχει μη γραμμική απεικόνιση μεταξύ των διανυσμάτων εισόδου και εξόδου. Όπως αναφέρθηκε και στο Κεφάλαιο 3.1.3, η κύρια δομή του νευρωνικού δικτύου αποτελείται από αρχιτεκτονική τριών στρωμάτων, δηλαδή, ένα στρώμα εισόδου που ορίζει την τιμή εισόδου, ένα ή περισσότερα κρυφά στρώματα ορίζουν τη μαθηματική συνάρτηση και ένα στρώμα εξόδου ορίζει το τελικό αποτέλεσμα. Κάθε στρώμα αποτελείται από μεγάλο αριθμό νευρώνων που είναι συνδέονται μεταξύ τους μέσω βαρών. Κάθε νευρώνας έχει μαθηματική συνάρτηση (γνωστή και ως ενεργοποίηση) που δέχεται δεδομένα εισόδου από το

προηγούμενο επίπεδο και παράγει έξοδο για το επόμενο επίπεδο. Έτσι, στα νευρωνικά δίκτυα η πρόβλεψη ορίζεται από τη συνάρτηση ενεργοποίησης (Yugal, 2012).

Ο συγκεκριμένος αλγόριθμος χρησιμοποιήθηκε στη μελέτη των Papada & Kaliampakos (2022) και τα αποτελέσματα που παρήγαγε θα ενταχθούν στη συγκεκριμένη εργασία, στο πλαίσιο της σύγκρισης αλγορίθμων που χρησιμοποιήθηκαν από το WEKA, συμπεριλαμβάνοντας περιπτώσεις αλγορίθμων από κάθε κατηγορία που διαθέτει το πρόγραμμα.

Όπως αναφέρθηκε και στο κεφάλαιο 3.1, η γραφική επιφάνεια του Explorer παρέχει στο χρήστη πληροφορίες σχετικά με τη μεθοδολογία που ακολούθησε ο αλγόριθμος κατά την εκπαίδευση του, καθώς και αναλυτικές λεπτομέρειες σχετικά με την ακρίβεια του μοντέλου που δημιούργησε ο αλγόριθμός ταξινόμησης για κάθε κλάση. Πιο συγκεκριμένα, στην επιφάνεια «classifier output» παρουσιάζονται στατιστικά αποτελέσματα σχετικά με τις ταξινομημένες περιπτώσεις με τα ποσοστά έγκυρων και λανθασμένων προβλέψεων, ειδικούς στατιστικούς δείκτες οι οποίοι παρουσιάζουν αναλυτικότερα την ακρίβεια των περιπτώσεων και το «confusion matrix» (Εικόνα 20). Οι εν λόγω στατιστικοί δείκτες αναλύονται με λεπτομέρεια στο Κεφάλαιο 4.

```
=== Summary ===

Correctly Classified Instances         546               59.1549 %
Incorrectly Classified Instances       377               40.8451 %
Kappa statistic                          0.1503
Mean absolute error                      0.4793
Root mean squared error                  0.4875
Relative absolute error                 96.2037 %
Root relative squared error             97.7004 %
Total Number of Instances              923

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0,275    0,130    0,650      0,275    0,387      0,182    0,593     0,550     Yes
                 0,870    0,725    0,577      0,870    0,694      0,182    0,593     0,592     No
Weighted Avg.    0,592    0,446    0,611      0,592    0,550      0,182    0,593     0,573

=== Confusion Matrix ===

   a   b   <-- classified as
 119 313 |   a = Yes
  64 427 |   b = No
```

Εικόνα 20: Αναλυτική ακρίβεια Ταξινομητή.

Κεφάλαιο 4. Μεθοδολογία

Η βάση δεδομένων που εισήχθη στο WEKA στο πλαίσιο της παρούσας διπλωματικής εργασίας αποτελεί την ίδια βάση δεδομένων που χρησιμοποιήθηκε και στη μελέτη των Papada & Kaliampakos (2022). Η βάση δεδομένων δημιουργήθηκε από τον συνδυασμό πέντε διακριτών πρωτογενών ερευνών για την ενεργειακή φτώχεια στην Ελλάδα. Η πρώτη αφορούσε το σύνολο της χώρας (Ελλάδα) (Papada & Kaliampakos, 2016), μία με αναφορά στις ελληνικές ορεινές περιοχές (Papada & Kaliampakos, 2017), δύο με αναφορά την ορεινή πόλη του Μετσόβου και μία τελευταία αναφερόμενη στον ορεινό οικισμό των Άγραφων, με τις τρεις τελευταίες να είναι πιο αντιπροσωπευτικές σχετικά με την περίπτωση αυτής της εργασίας. Το μεγαλύτερο ποσοστό του δείγματος επιλέχθηκε να αποτελείται από νοικοκυριά που ζουν σε ορεινές περιοχές της χώρας, με σκοπό την ανάδειξη της οξύτητας και της έκτασης του προβλήματος της ενεργειακής φτώχειας, ειδικά στις ορεινές περιοχές.

Οι παραπάνω έρευνες πραγματοποιήθηκαν σε επίπεδο νοικοκυριού κατά τα έτη 2016-2020, με τυχαία μέθοδο δειγματοληψίας και διάστημα εμπιστοσύνης 95%. Η βάση δεδομένων περιλαμβάνει στο σύνολο της 1754 σειρές δεδομένων, δηλαδή 400 νοικοκυριά που ζουν σε ολόκληρη την Ελλάδα, 400 νοικοκυριά που ζουν σε ορεινές περιοχές της χώρας, 643 νοικοκυριά που ζουν στο Μέτσοβο και 311 που ζουν στα Άγραφα. Από το ευρύ φάσμα δεδομένων, το οποίο περιλαμβάνει πληροφορίες σχετικές με τη στέγαση, τις συνθήκες διαβίωσης, τα συστήματα θέρμανσης, το ενεργειακό κόστος, τα εισοδήματα κ.λπ. επιλέχθηκαν συγκεκριμένοι αντικειμενικοί και υποκειμενικοί δείκτες, καθώς και συγκεκριμένες κοινωνικοοικονομικές παράμετροι για την πραγματοποίηση της συγκεκριμένης ανάλυσης.

Η ανάλυση περιλαμβάνει την πρόβλεψη αντικειμενικών δεικτών με βάση υποκειμενικούς δείκτες, μέσω της εκπαίδευσης διαφόρων αλγορίθμων μηχανικής μάθησης. Όσον αφορά την μεταβλητή εξόδου, χρησιμοποιήθηκαν τρεις δείκτες αντικειμενικού υπολογισμού της ενεργειακής φτώχειας. Πιο συγκεκριμένα, δημιουργήθηκαν ξεχωριστά μοντέλα για κάθε μεταβλητή εξόδου, όπου πραγματοποιήθηκε πρόβλεψη με βάση τους υποκειμενικούς δείκτες για κάθε αντικειμενικό δείκτη. Οι αντικειμενικοί δείκτες είναι οι εξής:

1. 10% Actual : Δείκτης 10 % βάσει των πραγματοποιηθέντων ενεργειακών δαπανών
2. 10% Required: Δείκτης 10 % βάσει του Απαιτούμενου ενεργειακού κόστους
3. DCEN (Degree of Coverage of Energy Needs): Βαθμός Κάλυψης Ενεργειακών Αναγκών

Οι πέντε υποκειμενικοί δείκτες που επιλέχθηκαν ως μεταβλητές εισόδου ήταν:

1. Η αδυναμία να διατηρηθεί το σπίτι ζεστό (Inability to heat home)
2. Οι καθυστερήσεις στους λογαριασμούς ενέργειας (Arrears in energy bills)
3. Τα προβλήματα υγρασίας/μούχλας (Damp problems)
4. Τα προβλήματα υγείας λόγω ανεπαρκών συνθηκών θέρμανσης (Health Problems
5. Ο περιορισμός άλλων βασικών αναγκών προς κάλυψη των αναγκών θέρμανσης (Restriction of other needs)

Οι υποκειμενικοί δείκτες λήφθηκαν απευθείας από τις πρωτογενείς έρευνες που αναφέρθηκαν παραπάνω, ενώ οι επιλεχθέντες αντικειμενικοί δείκτες υπολογίστηκαν με τη χρήση δεδομένων

ενεργειακού κόστους και διαθέσιμου εισοδήματος των νοικοκυριών. Ειδικότερα, ο 1ος αντικειμενικός δείκτης (10% actual) υπολογίστηκε ως ο λόγος ενεργειακού κόστους προς το εισόδημα νοικοκυριών, ο 2ος αντικειμενικός δείκτης (10% required) βάσει της μεθοδολογίας που αναπτύχθηκε στη μελέτη των Papada & Kaliampakos (2018) και ο 3ος αντικειμενικός δείκτης (DCEN) για τη συμπίεση των αναγκών βάσει της μεθοδολογίας που αναπτύχθηκε στη μελέτη των Papada & Kaliampakos (2020).

Μετά τον καθορισμό της βάσης δεδομένων και, όπως προαναφέρθηκε και στο κεφάλαιο 3.3., πραγματοποιήθηκε προ-επεξεργασία των δεδομένων προτού δοθεί εντολή στους χρησιμοποιούμενους αλγορίθμους να ταξινομήσουν τα δεδομένα, ώστε να αποφευχθεί μείωση της ακρίβειας ταξινόμησης λόγω της μορφής των δεδομένων. Τα δεδομένα φορτώθηκαν στο WEKA και έγιναν οι δοκιμές εκπαίδευσης των αλγορίθμων. Οι αλγόριθμοι μηχανικής μάθησης (Ταξινομητές / Classifiers) που χρησιμοποιήθηκαν στην παρούσα εργασία είναι αυτοί που περιεγραφήκαν στο Κεφάλαιο 3.3.3., ήτοι:

1) Naïve Bayes
2) IBk
3) Decision Table
4) J48
5) AdaBoost.M1 (Meta-learning Algorithm) για ενίσχυση του Decision Table
6) Cost Sensitive Classifier (Meta-learning Algorithm) για ενίσχυση του J48

Για την εκτίμηση της ακρίβειας των μοντέλων λήφθηκε υπόψιν ένα σύνολο δεικτών:

- **Percentage of Correctly/Incorrectly Classified Instances:** Αποτελεί το κυριότερο στατιστικό αποτέλεσμα και την πρώτη τιμή που λαμβάνεται υπόψιν για την αξιολόγηση. Το WEKA παρουσιάζει τον αριθμό των σωστών προβλέψεων και το ποσοστό αυτών επί το συνολικών προβλέψεων που πραγματοποιήθηκαν. Στο κεφάλαιο 4, το συγκεκριμένο ποσοστό θα αναφέρεται απλώς ως « Accuracy» (Ακρίβεια) του μοντέλου.

Accuracy = True Instances / (True Instances + False Instances)

- **Detailed accuracy by class:** Σημαντικά επιμέρους ειδικά στατιστικά αποτελέσματα δίνονται ως επέκταση της αξιολόγησης ακρίβειας του ταξινομητή. Αυτά που θα ενταχθούν στην μελέτη της εργασίας είναι τα εξής:

  1. **<u>Precision:</u>** Αναφέρεται στην αναλογία σωστών θετικών προβλέψεων προς τις συνολικές θετικές (σωστές και λανθασμένες) προβλέψεις της κάθε κλάσης.

Precision = True Positive Instances/ (True Positive Instances + False Positive Instances)

2. **<u>Recall:</u>** Αναφέρεται στην αναλογία σωστών θετικών προβλέψεων προς το άθροισμα των σωστών θετικών και λανθασμένων αρνητικών προβλέψεων της κάθε κλάσης.

Recall = True Positive Instances / (True Positive Instances + False Negative Instances)

3. **<u>F-Measure:</u>** Εκφράζει τον αρμονικό μέσο όρο του ''Precision'' και ''Recall'' δίνοντας και στα δύο τον ίδιο συντελεστή βαρύτητας κατά τον συνδυασμό τους. Επιτρέπει την αξιολόγηση ενός μοντέλου λαμβάνοντας υπόψιν και τους δύο παράγοντες με τη χρήση μιας ενιαίας βαθμολόγησης, η οποία είναι χρήσιμη κατά την περιγραφή απόδοσης του μοντέλου, σε σύγκριση με άλλα μοντέλα, και κατά την εκτίμηση του κόστους ψευδών προβλέψεων (ψευδών θετικών και ψευδών αρνητικών).

F-Measure = (2* Precision*Recall) / (Precision + Recall)

4. **<u>ROC (Receiver Operating Characteristic) Area</u>**: Αποτελεί ένα μέτρο ακρίβειας, το οποίο υποδεικνύει την ικανότητας ακριβούς πρόβλεψης του μοντέλου σε τυχαία δεδομένα, και αποτελεί την περιοχή κάτω από την καμπύλη ROC (ROC curve). Τόσο η περιοχή όσο και η καμπύλη ROC αποτελούν γραφική μέθοδο απεικόνισης της ευαισθησίας των ορίων για κάθε μοντέλο, ενώ επίσης δίνει κάποιες ενδείξεις για το όφελος χρήσης του συγκεκριμένου τρόπου μοντελοποίησης (στη συγκεκριμένη περίπτωση την απόδοση ταξινόμησης) για τα μελετόμενα δεδομένα.
   Οι τιμές των συγκεκριμένων δεικτών εμφανίζονται σε παρόμοια επίπεδα με την τιμή της ακρίβειας «Accuracy». Συνεπώς, ένα μοντέλο με αυξημένη σχετικά ακρίβεια, χαρακτηρίζεται και από σχετικά αυξημένες τιμές για τους ειδικούς στατιστικούς δείκτες, και αντίστοιχα μοντέλο με μειωμένη ακρίβεια χαρακτηρίζεται από μειωμένες τιμές των ειδικών στατιστικών δεικτών. Αυτό είναι λογικό καθώς, οι συγκεκριμένοι δείκτες αποτελούν επιμέρους χαρακτηριστικά αξιολόγησης της ακρίβειας ενός μοντέλου

- Confusion Matrix: Αποτελεί πίνακα στον οποίο παρουσιάζονται τα ποσοστά των σωστών και λανθασμένων προβλέψεων για κάθε κλάση, με τα διαγώνια στοιχεία κάθε σειράς να αντιπροσωπεύουν τις σωστές προβλέψεις (σωστές θετικές και σωστές αρνητικές) για κάθε κλάση της μεταβλητής εξόδου.
  Μοντέλα με χαμηλή ακρίβεια χαρακτηρίζονται και από έντονη ανισορροπία μεταξύ των σωστών προβλέψεων ανά κλάση, η οποία σηματοδοτεί την μέτρια απόδοση και χαμηλή

εγκυρότητα της δοκιμής. Αυτό το αποτέλεσμα μεταφράζεται ως λανθασμένη ταξινόμηση των νοικοκυριών ως ενεργειακά ή με ενεργειακά φτωχά.

Για την εκπαίδευση και τον έλεγχο των μοντέλων, επιλέχθηκε η μέθοδος «Percentage Split», και πιο συγκεκριμένα, καθορίστηκε ότι το 70% των δεδομένων θα χρησιμοποιηθεί για εκπαίδευση και το 30% για δοκιμή (test) του μοντέλου. Η επιλογή αυτή έγινε διατηρώντας την ίδια μέθοδο που ακολουθήθηκε και στη μελέτη των Papada & Kaliampakos (2022), ώστε να είναι δυνατή και η σύγκριση των αποτελεσμάτων σε τελική ανάλυση.

Συνολικά, πραγματοποιήθηκαν 144 δοκιμές με συνδυασμό δεικτών ως μεταβλητές εισόδου, δοκιμάζοντας διαφορετικούς συνδυασμούς υποκειμενικών δεικτών, ώστε να διερευνηθεί η επίδραση του κάθε υποκειμενικού δείκτη, αλλά και της συνδυασμένης χρήσης τους στους αντικειμενικούς δείκτες. Χάριν ευκολίας, χρησιμοποιούνται οι εξής συντομογραφίες για κάθε υποκειμενικό δείκτη στα αποτελέσματα της μελέτης:

- IW = Inability to keep the house warm
- AB = Arrears in Energy bills
- DL = Damp problems
- HP = Health Problems
- RN = Restriction of other Needs

Αναφορικά με τη βελτιστοποίηση των μοντέλων που «έτρεξαν», αλλά και των αποτελεσμάτων, εφαρμόστηκε το φίλτρο SMOTE, με σκοπό την εξομάλυνση των δεδομένων στις μεταβλητές εξόδου: 10% actual, 10% required, DCEN, αλλά και γενικά σε όποια περίπτωση κλάσεων τα δεδομένα ήταν ανισοβαρή. Επιπλέον, έγιναν κάποιες ρυθμίσεις στις παραμέτρους εφαρμογής του SMOTE και την προσέγγιση k-nearest neighbor που χρησιμοποιεί για τη δημιουργία συνθετικών περιπτώσεων, προς εξομάλυνση των δεδομένων. Πιο συγκεκριμένα, για τη συντριπτική πλειονότητα των μοντέλων που δημιουργήθηκαν, επιλέχθηκε k=4 για SMOTE, δηλαδή δόθηκε εντολή στο φίλτρο να δημιουργεί συνθετικές περιπτώσεις βάσει των τεσσάρων κοντινότερων γειτονικών περιπτώσεων.

Κεφάλαιο 5. Αποτελέσματα

Στο παρόν Κεφάλαιο παρουσιάζονται τα αποτελέσματα που κατέγραψε το WEKA από την εκπαίδευση των επιλεχθέντων αλγορίθμων μηχανικής μάθησης.

5.1. Αποτελέσματα ανά αλγόριθμο

5.1.1 Naïve Bayes

Στους παρακάτω πίνακες (Πίν.1, 2, 3) παρουσιάζονται τα αποτελέσματα (output) που έδωσε ο ταξινομητής Naive Bayes για τους αντικειμενικούς δείκτες που επιλέχθηκαν, ήτοι 10% actual (με βάση τις πραγματοποιθείσες δαπάνες), 10% required (με βάση το απαιτούμενο κόστος) και δείκτης συμπίεσης DCEN. Όπως φαίνεται, δημιουργήθηκαν μοντέλα για διαφορετικούς συνδυασμούς υποκειμενικών δεικτών, τα οποία έδωσαν μια μέτρια προς σχετικά καλή ακρίβεια για τους περισσότερους στατιστικούς δείκτες.

Πίνακας 1: Πρόβλεψη Naive Bayes ταξινομητή με δείκτη 10% actual

Prediction of indicator 10% actual								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,546	0,695	0,612	0,566	Yes	**56,38%**	**69%**	31%
	0,593	0,435	0,502	0,566	No		56%	**44%**
	0,570	0,564	0,556	0,566	Average			
IW-AB-DL-RN	0,546	0,695	0,612	0,568	Yes	**56,38%**	**69%**	31%
	0,593	0,435	0,502	0,568	No		56%	**44%**
	0,570	0,564	0,556	0,568	Average			
IW-AB-RN	0,552	0,684	0,611	0,563	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,563	No		54%	**46%**
	0,575	0,569	0,564	0,563	Average			
IW-DL-RN	0,552	0,684	0,611	0,581	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,581	No		54%	**46%**
	0,575	0,569	0,564	0,581	Average			
IW-HP-RN	0,552	0,684	0,611	0,577	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,577	No		54%	**46%**
	0,575	0,569	0,564	0,577	Average			
IW-RN	0,552	0,684	0,611	0,573	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,573	No		54%	**46%**

	0,575	0,569	0,564	0,573	Average			
AB-RN	0,532	0,678	0,596	0,575	Yes	**56,14%**	**67%**	33%
	0,608	0,455	0,52	0,575	No		54%	**46%**
	0,571	0,561	0,557	0,575	Average			
RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571	Average			

Πίνακας 2 Πρόβλεψη Naive Bayes ταξινομητή με δείκτη 10% required

Prediction of indicator 10% required								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,588	0,505	0,543	0,591	Yes	**57,48%**	**51%**	49%
	0,565	0,645	0,603	0,591	No		35%	**65%**
	0,576	0,575	0,573	0,591				
IW-AB-DL-RN	0,583	0,498	0,537	0,592	Yes	**57,02%**	**49%**	51%
	0,561	0,643	0,599	0,592	No		36%	**64%**
	0,572	0,57	0,568	0,592				
IW-AB-RN	0,607	0,395	0,479	0,587	Yes	**56,91%**	**37%**	63%
	0,551	0,744	0,633	0,587	No		26%	**74%**
	0,579	0,569	0,569	0,587				
IW-DL-RN	0,579	0,415	0,483	0,553	Yes	**55,10%**	**41%**	59%
	0,535	0,691	0,603	0,553	No		31%	**69%**
	0,557	0,551	0,542	0,553				
IW-HP-RN	0,588	0,505	0,543	0,591	Yes	**57,48%**	**50%**	50%
	0,565	0,645	0,603	0,591	No		35%	**65%**
	0,576	0,575	0,573	0,591				
IW-RN	0,607	0,331	0,428	0,581	Yes	**55,77%**	**33%**	67%
	0,539	0,785	0,639	0,581	No		22%	**78%**
	0,573	0,558	0,534	0,581				
AB-RN	0,588	0,505	0,543	0,591	Yes	**57,48%**	**50%**	50%
	0,565	0,645	0,603	0,591	No		35%	**65%**
	0,576	0,575	0,573	0,591				
RN	0,534	0,674	0,596	0,543	Yes	**54,28%**	**67%**	33%
	0,557	0,412	0,474	0,543	No		59%	**41%**
	0,546	0,543	0,535	0,543				

Πίνακας 3 Πρόβλεψη Naive Bayes ταξινομητή με δείκτη DCEN

Prediction of indicator DCEN								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,613	0,439	0,512	0,626	Yes	**56,77%**	**44%**	56%
	0,541	0,705	0,612	0,626	No		48%	**52%**
	0,578	0,568	0,56	0,626				
IW-AB-DL-RN	0,597	0,687	0,639	0,623	Yes	**59,92%**	**69%**	31%
	0,603	0,506	0,550	0,623	No		49%	**51%**
	0,600	0,599	0,506	0,623				
IW-AB-RN	0,603	0,665	0,632	0,607	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,607	No		47%	**53%**
	0,601	0,601	0,599	0,607				
IW-DL-RN	0,603	0,665	0,632	0,625	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,625	No		47%	**53%**
	0,601	0,601	0,599	0,625				
IW-HP-RN	0,604	0,709	0,652	0,620	Yes	**61,04%**	**71%**	29%
	0,620	0,506	0,557	0,620	No		49%	**51%**
	0,612	0,610	0,606	0,620				
IW-RN	0,603	0,665	0,632	0,612	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,612	No		47%	**53%**
	0,601	0,601	0,599	0,612				
AB-RN	0,620	0,428	0,506	0,571	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,571	No		28%	**72%**
	0,582	0,57	0,561	0,571				
RN	0,620	0,428	0,506	0,574	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,574	No		28%	**72%**
	0,582	0,570	0,561	0,574				

Ακολουθεί η ανάλυση των παραπάνω στοιχείων ανά πίνακα:

- **Δείκτης 10% actual (Πίνακας 1)**

Εύρος ακρίβειας: 56,14% έως 56,92%.

Δοκιμή χαμηλότερη τιμής ακρίβειας: ‘’IW-RN’’ με 56,14%

Δοκιμή υψηλότερης τιμής ακρίβειας: ‘’IW-AB-RN’’ , ‘’IW-DL-RN’’ , ‘’IW-HP-RN’’, ‘’RN’’ με 56,92%

Σύμφωνα με τα παραπάνω αποτελέσματα, η απουσία του δείκτη ‘’IW-Inability to heat home’’ από κάποια δοκιμή φαίνεται να μειώνει την απόδοση του ταξινομητή (δοκιμή AB-RN) , ενώ

παράλληλα η παρουσία του δείκτη ‘’RN-Restriction of other needs’’ φαίνεται να είναι καθοριστική για την επίτευξη ικανοποιητικών αποδόσεων άνω του 50% σε όλες τις δοκιμές.

Επίσης , σε όλες τις περιπτώσεις οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area παρουσιάζουν ικανοποιητικές τιμές για τον σταθμισμένο μέσο όρο, καθώς υπερβαίνουν την τιμή 0,5 (από 0,556 έως 0,581), όπου σε μελέτη τέτοιων κοινωνικών φαινομένων μια ακρίβεια της τάξης του 60% θεωρείται ικανοποιητική.

Όσον αφορά τα confusion matrix, παρατηρείται ότι ο ταξινομήτης λειτουργεί σε πολύ καλό επίπεδο, προβλέποντας σωστά το 67% με 69% των προβλέψεων για τις περιπτώσεις των ενεργειακά φτωχών νοικοκυριών, ενώ το ποσοστό πρόβλεψης για τα μη ενεργειακά φτωχά νοικοκυριά πέφτει στο 44%. Από αυτά τα αποτελέσματα, συμπεραίνεται ότι το 67% με 69% των περιπτώσεων που εμφανίζει οποιονδήποτε συνδυασμό εκ των υποκειμενικών δεικτών στα νοικοκυριά τους είναι ενεργειακά φτωχό. Αυτές οι ενδείξεις δείχνουν την πιθανή έκταση του προβλήματος της ενεργειακής φτώχειας και των περαιτέρω επιπτώσεων που μπορεί να επιφέρει.

- **Δείκτης 10% required (Πίνακας 2)**

Εύρος ακρίβειας: 54,28% έως 57,48%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’IW-DL-RN ‘’ με 54,28%

Δοκιμή υψηλότερης τιμής ακρίβειας: ‘’IW-AB-DL-HP-RN’’, ‘’IW-HP-RN’’ , ‘’IW-AB-RN’’ με 57,48%

Όσον αφορά το μοντέλο για τον δείκτη 10% required (με βάση το απαιτούμενο κόστος), τα επίπεδα ακρίβειας των δοκιμών είναι σχετικά ικανοποιητικά (άνω του 54%), με τις δοκιμές όπου περιλαμβάνονται οι δείκτες ‘’HP- Health problems’’, ‘’AB- Arrears in energy bills’’, να δίνουν τα καλύτερα αποτελέσματα με ακρίβεια άνω του 56,90% , ενώ ο υποκειμενικός δείκτης που φαίνεται να μειώνει την ακρίβεια των δοκιμών είναι ο ‘’RN-Restriction of other needs’’, καθώς η δοκιμή όπου χρησιμοποιείται ως ο μοναδικός υποκειμενικός δείκτης εμφανίζει τη χαμηλότερη ακρίβεια. Στις υπόλοιπες δοκιμές η μειωμένη ακρίβεια του συγκεκριμένου δείκτη καλύπτεται από την καλύτερη απόδοση των υπόλοιπων δεικτών.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area παρουσιάζουν ικανοποιητικές τιμές για τον σταθμισμένο μέσο όρο, καθώς υπερβαίνουν την τιμή 0,5 για όλες τις περιπτώσεις, κυμαινόμενοι στο σύνολό τους στο εύρος 0,542 έως 0,591. Όπως προαναφέρθηκε, ακρίβεια κοντά στο 60% θεωρείται ικανοποιητική σε τέτοιου είδους μελέτες κοινωνικών φαινομένων.

Τα confusion matrix για το μοντέλο του δείκτη 10% required φαίνεται να είναι ελαφρώς χειρότερα από αυτά του δείκτη 10% actual, καθώς στις περισσότερες περιπτώσεις παρατηρείται αρκετά έντονη ανισορροπία μεταξύ του ποσοστού σωστών προβλέψεων για κάθε κλάση. Οι δοκιμές ‘’IW-AB-DL-HP-RN’’, ‘’AB-RN’’ αποτελούν τις δύο περιπτώσεις όπου τα ποσοστά σωστών προβλέψεων είναι αρκετά ικανοποιητικά, καθώς για κάθε κλάση το ποσοστό σωστών προβλέψεων είναι μεγαλύτερο ή ίσο του 50%, υποδηλώνοντας σχετική εμπιστοσύνη σχετικά με τα αποτελέσματα της δοκιμής. Αντίθετα, στις υπόλοιπες περιπτώσεις επικρατεί ανισορροπία

μεταξύ των περιπτώσεων, ειδικά στις δοκιμές ‘’IW-AB-RN’’, ‘’IW-RN’’, όπου τα ποσοστά σωστών προβλέψεων για κάθε κλάση έχουν σημαντικές διαφορές. Για παράδειγμα, στην περίπτωση ‘’IW-RN’’, το ποσοστό σωστών θετικών (ενεργειακά φτωχών) προβλέψεων περιορίζεται στο 33%, ενώ το ποσοστό των σωστών αρνητικών προβλέψεων (μη ενεργειακά φτωχών) ανέρχεται στο υψηλό 78%. Μεταφράζοντας τα αποτελέσματα των confusion matrix ως προς τις συνέπειες που μπορούν να επιφέρουν σε πραγματικό επίπεδο, πολλές περιπτώσεις ενεργειακά φτωχών νοικοκυριών θα θεωρηθούν μη ενεργειακά φτωχά και πιθανώς θα στερηθούν οποιαδήποτε μορφή οικονομικής υποστήριξης.

- **Δείκτης DCEN (Πίνακας 3)**

Εύρος ακρίβειας : 56,77% έως 61,04%

Δοκιμή χαμηλότερης τιμής ακρίβειας: IW-AB-DL-HP-RN ΜΕ 56,77%

Δοκιμή υψηλότερης τιμής ακρίβειας: IW-HP-RN με 61,04%

Στη μοντελοποίηση των δεδομένων με τη χρήση του δείκτη DCEN (Degree of Coverage of Energy Needs) εμφανίζονται τα καλύτερα αποτελέσματα για τους περισσότερους στατιστικούς δείκτες συγκριτικά με τα μοντέλα των άλλων δεικτών. Η ακρίβεια των μοντέλων κυμαίνεται άνω του 56,70%, ενώ είναι το μοναδικό μοντέλο του Naive Bayes το οποίο επιτυγχάνει ακρίβεια άνω του 60% σε ορισμένες περιπτώσεις. Συγκεκριμένα, οι δοκιμές τριών υποκειμενικών δεικτών ‘’IW-AB-RN’’, ‘’IW-DL-RN’’, ‘’ IW-HP-RN’’ έδωσαν ακρίβεια άνω του 60%. Επίσης, στο συγκεκριμένο μοντέλο δε φαίνεται να υπάρχει κάποιος εκ των υποκειμενικών δεικτών ο οποίος να δρα αρνητικά στην ακρίβεια των δοκιμών, καθώς δεν παρατηρείται κάποια σημαντική μείωση στην ακρίβεια κάποιας δοκιμής. Θα μπορούσε, πιθανώς, να θεωρηθεί ότι ο συγκεκριμένος αλγόριθμος, υπό τη χρήση της συγκεκριμένης μορφής δεδομένων λειτουργεί στο βέλτιστο δυνατό επίπεδο με τη χρήση τριών υποκειμενικών δεικτών, ενώ με την πρόσθεση ή αφαίρεση δεικτών, ο Naive Bayes ελαφρώς υπολειτουργεί.

Υψηλές τιμές σε σχέση με τα υπόλοιπα μελετόμενα μοντέλα των αντικειμενικών δεικτών παρατηρούνται και για τους στατιστικούς δείκτες Precision, Recall, F-Measure και ROC Area, όπου οι τιμές όλων των δεικτών για τον σταθμισμένο μέσο όρο κυμαίνονται από 0,506 έως 0,626. Παρά το μεγάλο εύρος τιμών των ειδικών στατιστικών δεικτών που παρατηρείται για τον δείκτη DCEN, αποτελεί την πρώτη περίπτωση στη μελέτη των αντικειμενικών δεικτών όπου παρατηρούνται τιμές άνω του 60% (>0,6) και στους ειδικούς στατιστικούς δείκτες.

Η βελτιωμένη απόδοση του δείκτη DCEN φαίνεται και στα confusion matrix των δοκιμών, στα οποία εμφανίστηκαν, συνολικά, τα καλύτερα ποσοστά σωστών προβλέψεων σε σύγκριση με τα μοντέλα των υπόλοιπων αντικειμενικών δεικτών. Συγκεκριμένα, στις δοκιμές υποκειμενικών δεικτών ‘’IW-AB-DL-RN’’, ‘’ IW-AB-RN’’ ,’’ IW-DL-RN’’, ‘’ IW-HP-RN’’, ‘’ IW-RN’’ σημειώθηκαν ποσοστά σωστών προβλέψεων άνω του 50% και για τις δύο κλάσεις, ενώ παράλληλα τα ποσοστά σωστών θετικών προβλέψεων για αυτές τις δοκιμές κυμάνθηκαν σε υψηλές τιμές από 67%-71%, οι οποίες είναι αρκετά ικανοποιητικές και σαφώς ανώτερες από αυτές των προηγούμενων μοντέλων.

5.1.2. IBk (Instance-based learning with parameter k)

Στους παρακάτω πίνακες (Πίν. 4,5,6) παρατηρούνται τα αποτελέσματα (output) που έδωσε ο ταξινομητής IBk για τους τρεις μελετώμενους αντικειμενικός δείκτες, για τις ίδιους συνδυασμούς υποκειμενικών δεικτών που μελετήθηκαν και με τον ταξινομητή Naive Bayes. Τα στατιστικά αποτελέσματα του IBk θεωρούνται σχετικά ικανοποιητικά όπως και στην προηγούμενη περίπτωση, με κάποιες διακυμάνσεις συγκριτικά με τον Naive Bayes. Πιο αναλυτικά:

Πίνακας 4: Πρόβλεψη IBk ταξινομητή με δείκτη 10% actual

Prediction of indicator 10% actual								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,570	0,558	0,564	0,585	Yes	**57,33%**	**56%**	44%
	0,576	0,589	0,582	0,585	No		41%	**59%**
	0,573	0,573	0,573	0,585				
IW-AB-DL-RN	0,549	0,626	0,585	0,578	Yes	**56,11%**	**63%**	37%
	0,576	0,497	0,534	0,578	No		50%	**50%**
	0,563	0,561	0,559	0,578				
IW-AB-RN	0,552	0,684	0,611	0,572	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,572	No		54%	**46%**
	0,575	0,569	0,564	0,572				
IW-DL-RN	0,544	0,613	0,576	0,577	Yes	**55,43%**	**61%**	39%
	0,567	0,497	0,530	0,577	No		50%	**50%**
	0,556	0,554	0,553	0,577				
IW-HP-RN	0,561	0,657	0,605	0,573	Yes	**57,61%**	**66%**	34%
	0,597	0,497	0,543	0,573	No		50%	**50%**
	0,579	0,576	0,573	0,573				
IW-RN	0,552	0,684	0,611	0,573	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,573	No		54%	**46%**
	0,575	0,569	0,564	0,573				
AB-RN	0,552	0,684	0,611	0,568	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,568	No		54%	**46%**
	0,575	0,569	0,564	0,568				
RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				

Πίνακας 5: Πρόβλεψη IBk ταξινομητή με δείκτη 10% required

Prediction of indicator 10% required								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,547	0,642	0,590	0,599	Yes	**55,42%**	**64%**	36%
	0,565	0,467	0,511	0,599	No		53%	**47%**
	0,556	0,554	0,511	0,599				
IW-AB-DL-RN	0,599	0,402	0,481	0,593	Yes	**56,57%**	**40%**	60%
	0,549	0,730	0,627	0,593	No		27%	**73%**
	0,574	0,566	0,554	0,593				
IW-AB-RN	0,537	0,685	0,602	0,578	Yes	**54,62%**	**68%**	32%
	0,563	0,407	0,473	0,578	No		59%	**41%**
	0,55	0,546	0,537	0,578				
IW-DL-RN	0,582	0,452	0,509	0,588	Yes	**56,34%**	**45%**	55%
	0,551	0,675	0,607	0,588	No		32%	**68%**
	0,567	0,563	0,558	0,588				
IW-HP-RN	0,601	0,372	0,46	0,582	Yes	**56,22%**	**37%**	63%
	0,545	0,753	0,632	0,582	No		25%	**75%**
	0,573	0,562	0,546	0,582				
IW-RN	0,542	0,749	0,629	0,582	Yes	**55,77%**	**75%**	25%
	0,593	0,366	0,453	0,582	No		63%	**37%**
	0,567	0,558	0,541	0,582				
AB-RN	0,533	0,71	0,609	0,553	Yes	**54,40%**	**71%**	29%
	0,565	0,378	0,453	0,553	No		62%	**38%**
	0,549	0,544	0,531	0,553				
RN	0,534	0,674	0,596	0,543	Yes	**54,28%**	**67%**	33%
	0,557	0,412	0,474	0,543	No		59%	**41%**
	0,546	0,543	0,535	0,543				

Πίνακας 6: Πρόβλεψη IBk ταξινομητή με δείκτη DCEN

Prediction of indicator DCEN								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,617	0,406	0,490	0,615	Yes	**56,40%**	**41%**	59%
	0,537	0,732	0,619	0,615	No		27%	**73%**
	0,578	0,564	0,553	0,615				
IW-AB-DL-RN	0,609	0,622	0,616	0,625	Yes	**59,92%**	**62%**	38%

	0,588	0,575	0,581	0,625	No		43%	**57%**
	0,599	0,599	0,599	0,625				
IW-AB-RN	0,584	0,651	0,616	0,614	Yes	**58,07%**	**65%**	35%
	0,576	0,506	0,539	0,614	No		49%	**51%**
	0,580	0,581	0,578	0,614				
IW-DL-RN	0,603	0,665	0,632	0,624	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,624	No		47%	**53%**
	0,601	0,601	0,599	0,624				
IW-HP-RN	0,604	0,709	0,652	0,627	Yes	**61,04%**	**71%**	29%
	0,620	0,506	0,557	0,627	No		49%	**51%**
	0,612	0,610	0,606	0,627				
IW-RN	0,603	0,665	0,632	0,612	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,612	No		47%	**53%**
	0,601	0,601	0,599	0,612				
AB-RN	0,620	0,428	0,506	0,571	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,571	No		28%	**72%**
	0,582	0,570	0,561	0,571				
RN	0,620	0,428	0,506	0,574	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,574	No		28%	**72%**
	0,582	0,570	0,561	0,574				

Ακολουθεί η ανάλυση των παραπάνω στοιχείων ανά πίνακα:

- **Δείκτης 10% actual (Πίνακας 4)**

Εύρος ακρίβειας: 55,43% έως 57,61%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’IW-DL-RN’’ με 55,43%

Δοκιμή υψηλότερης τιμής ακρίβειας: ‘’ IW-HP-RN’’ με 57,61%

Στον πίνακα 4 παρατηρούνται τα αποτελέσματα του IBk για μοντέλο του δείκτη 10% actual, και μπορούν να χαρακτηριστούν ως ελαφρώς καλύτερα σε σχέση με την ίδια μοντελοποίηση του Naïve Bayes όσον αφορά την ακρίβεια των δοκιμών. Ο IBk εμφάνισε πιο μεγάλο εύρος τιμών ακρίβειας μεταξύ των δοκιμών, ωστόσο, μόνο η δοκιμή ‘’IW-DL-RN’’ είχε μια σχετικά χαμηλότερη ακρίβεια (55,43%), ενώ οι υπόλοιπες δοκιμές σημείωσαν ακρίβεια μεγαλύτερη του 56,9% , καθιστώντας έτσι τα αποτελέσματα της ακρίβειας ως ικανοποιητικά. Πιθανώς, ο δείκτης ‘’DL-Damp problems’’, να αποτελεί την αιτία χαμηλότερης ακρίβειας της δοκιμής ‘’IW-DL-RN’’, καθώς οι υπόλοιπες δοκιμές σημειώνουν μεγαλύτερη ακρίβεια κατά, τουλάχιστον, 0,60 ποσοστιαίες μονάδες και δεν περιλαμβάνουν τον δείκτη ‘’DL’’. Εξαίρεση σε αυτό το συλλογισμό αποτελεί η

δοκιμή του συνόλου, των 5 υποκειμενικών δεικτών ‘’ IW-AB-DL-HP-RN’’,όπου η πιο αποδοτική λειτουργία των υπόλοιπων δεικτών υπερκαλύπτει και εξισορροπεί την ‘’υπολειτουργία’’ του δείκτη ‘’DL’’ επιτυγχάνοντας την καλή απόδοση της τάξης του 57,33%. Σε γενική βάση όλες οι δοκιμές επιτυγχάνουν σχετικά καλή ακρίβεια άνω του 55%

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area, παραμένουν σε ικανοποιητικό επίπεδο για την φύση των δεικτών αυτών, άνω του 0,5 (από 0,553 έως 0,585) για τον σταθμισμένο μέσο όρο.

Τα confusion matrix των προβλέψεων εμφανίζονται ελαφρώς βελτιωμένα σε σχέση με τις προβλέψεις του NaïveBayes, όσον αφορά το ποσοστό των σωστών προβλέψεων για μη ενεργειακά φτωχές περιπτώσεις όπου τα ποσοστά κυμαίνονται από 46% έως 59%, ενώ παράλληλα και τα ποσοστά των σωστών προβλέψεων για ενεργειακά φτωχές περιπτώσεις παραμένουν άνω του 50% (από 56% έως 68%). Συμπερασματικά, ο IBk φαίνεται να λειτούργησε ελαφρώς καλύτερα από τον Naive Bayes, σε όλες τις μελετώμενες παραμέτρους περί ακρίβειας και σωστών προβλέψεων, από το οποίο μπορεί να εξαχθεί το συμπέρασμα ότι λειτουργεί καλύτερα σε αυτές τις μορφές φαινομένων. Άλλωστε, και ο δείκτης ROC Area, οποίος δείχνει την ακρίβεια του ταξινόμητη σε τυχαία δεδομένα είναι υψηλότερος στην περίπτωση του IBk.

- **Δείκτης 10% required (Πίνακας 5)**

Εύρος ακρίβειας : 54,28% ΄έως 56,57%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’RN’’ με 54,28%

Δοκιμή υψηλότερης τιμής ακρίβειας : ‘’IW-AB-DL-RN’’ ΜΕ 56,57%

Η ακρίβεια των δοκιμών του IBk για τον δείκτη 10% required είναι σχετικά ικανοποιητική, με όλες τις δοκιμές να σημειώνουν ακρίβεια άνω του 54%, αλλά είναι ελαφρώς μειωμένη σε σχέση με την ακρίβεια των δοκιμών του δείκτη 10% required με τη χρήση του Naive Bayes. Ο δείκτης ‘’RN-Restriction of other needs’’ φαίνεται να είναι ο δείκτης που δυσχεραίνει περισσότερο τη διαδικασία ταξινόμησης του αλγορίθμου λόγω τη χαμηλότερης ακρίβειας που εμφανίζει στη δοκιμή που χρησιμοποιείται ως ο μοναδικός υποκειμενικός δέκτης της δοκιμής, ενώ καθώς προστίθενται υποκειμενικοί δείκτες για τις επόμενες δοκιμές, η ακρίβεια τους αυξάνεται ως ενός σημείου. Τέλος αξίζει να σημειωθεί ότι, οι δοκιμές όπου γίνεται χρήση τριών υποκειμενικών δεικτών για τη μοντελοποίηση, παρουσιάζουν τη μεγαλύτερη ακρίβεια.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area, παραμένουν σε ικανοποιητικό επίπεδο δεδομένου του τύπου του μελετώμενου φαινομένου, άνω του 0,5 (από 0,511 έως 0,599) για τον σταθμισμένο μέσο όρο, σημειώνοντας ελαφρώς μεγαλύτερο εύρος τιμών.

Όσον αφορά τα confusion matrix των δοκιμών, παρατηρείται ανισορροπία μεταξύ των σωστών και λανθασμένων προβλέψεων ανά κλάση, με συγκεκριμένες περιπτώσεις να σημειώνουν μέτρια έως και απογοητευτικά αποτελέσματα. Ειδικότερα, ενώ στις περισσότερες περιπτώσεις

οι προβλέψεις θετικών περιπτώσεων (ενεργειακά φτωχών) επιτυγχάνουν ικανοποιητικά αποτελέσματα σωστών προβλέψεων, την ίδια στιγμή, οι αρνητικές σωστές (μη ενεργειακά φτωχές) προβλέψεις περιορίζονται σε πού χαμηλά ποσοστά. Φυσικά σε κάποιες δοκιμές, συμβαίνει και το ακριβώς αντίθετο, με τις σωστές αρνητικές προβλέψεις να επιτυγχάνουν υψηλά ποσοστά, ενώ οι σωστές θετικές να κυμαίνονται σε χαμηλά επίπεδα. Αυτή η έντονη ανισορροπία μεταξύ των σωστών προβλέψεων μειώνει σημαντικά την αποτελεσματικότητα του μοντέλου και καθιστά δύσκολη τη χρήση του για εξαγωγή συμπερασμάτων για την σύνδεση μεταξύ αντικειμενικών και υποκειμενικών δεικτών όσον αφορά την ενεργειακή φτώχεια. Τα αποτελέσματα, λοιπόν του συγκεκριμένου μοντέλου δίνουν κάποια στοιχεία για την μελέτη της ενεργειακής φτώχειας σύμφωνα με τον δείκτη 10% required (με βάση το απαιτούμενο κόστος), αλλά αναφορικά με τον χρησιμοποιούμενο αλγόριθμο, ο IBk φαίνεται να μην λειτουργεί στο επιθυμητό επίπεδο.

- **Δείκτης DCEN (Πίνακας 6)**

Εύρος ακρίβειας : 56,77% έως 61,04%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’ IW-AB-DL-HP-RN’’ με 56,77%

Δοκιμή υψηλότερης τιμής ακρίβειας : ‘’ IW-HP-RN’’ με 61,03%

Οι δοκιμές που πραγματοποιούνται με βάση τον δείκτη DCEN, φαίνεται να αποδίδουν ικανοποιητικά και για τον IBk όπου επιτυγχάνεται ακρίβεια άνω του 60% σε ορισμένες δοκιμές. Ο αλγόριθμος προβλέπει ικανοποιητικά όλες τις δοκιμές, με την χαμηλότερη ακρίβεια στην δοκιμή ‘’IW-AB-DL-HP-RN’’ (56,77%), πιθανώς να οφείλεται στον ‘’θόρυβο’’, δηλαδή τον μεγαλύτερο όγκο δεδομένων που λαμβάνεται υπόψιν στη δοκιμή. Για ακόμη μία φορά, οι δοκιμές με τη μεγαλύτερη ακρίβεια είναι αυτές όπου χρησιμοποιούνται μόνο τρεις υποκειμενικοί δείκτες για τη δοκιμή (ακρίβεια > 60%).

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area παρουσιάζουν ικανοποιητικές τιμές για τον σταθμισμένο μέσο όρο άνω του 0,55 (55%) και κυμαίνονται μεταξύ 0,560 και 0,625 για όλους του στατιστικούς δείκτες. Επιπλέον, η καλή απόδοση του αλγορίθμου φαίνεται και από τιμές του ROC Area, οι οποίες κυμαίνονται από 0,571 έως και 0,625 για το σταθμισμένο μέσο όρο, υποδεικνύοντας την ικανοποιητική ακρίβεια του αλγορίθμου και σε τυχαία δεδομένα.

Τέλος όσον αφορά τα confusion matrix, σε γενικό επίπεδο, επιτυγχάνονται καλά ποσοστά σωστών προβλέψεων, και οι περιπτώσεις ανισορροπίας είναι περιορισμένες σε ήπια επίπεδα. Άλλο ένα χαρακτηριστικό παράδειγμα της πιθανής ισχυρής σχέσης των υποκειμενικών και αντικειμενικών δεικτών και τη σημασίας συνδυασμού και των δύο για τον υπολογισμό της ενεργειακής φτώχεια αποτελεί η δοκιμή ‘’ IW-HP-RN’’,όπου επιτυγχάνεται ποσοστό 71% για σωστές θετικές (ενεργειακά φτωχές) περιπτώσεις, ενώ παράλληλα διατηρείται και ικανοποιητικό ποσοστό σωστών αρνητικών (μη ενεργειακά φτωχών) περιπτώσεων. Από αυτή τη δοκιμή, εξάγεται με μια σχετική ασφάλεια όσον αφορά την εγκυρότητα των δεδομένων το εξής

συμπέρασμα. Το 71% των ορεινών νοικοκυριών που δήλωσαν ότι πάσχουν από προβλήματα υγείας, είναι κατά πάσα πιθανότητα και ενεργειακά φτωχά.

5.1.3.Decision Table Classifier

Στους παρακάτω πίνακες (Πίν. 7,8,9) σημειώνεται η στατιστική απόδοση που σημείωσε ο ταξινομητής Decision Table για τα μοντέλα των τριών αντικειμενικών δεικτών. Τα αποτελέσματα θεωρούνται και σε αυτή τη περίπτωση θεωρητικά ικανοποιητικά και παρακάτω αναλύονται περαιτέρω.

Πίνακας 7: Πρόβλεψη Decision Table ταξινομητή με δείκτη 10% actual

Prediction of indicator 10% actual								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,553	0,555	0,554	0,579	Yes	**55,84%**	**55%**	45%
	0,563	0,562	0,563	0,579	No		44%	**56%**
	0,558	0,558	0,558	0,579				
IW-AB-DL-RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				
IW-AB-RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				
IW-DL-RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				
IW-HP-RN	0,561	0,657	0,605	0,573	Yes	**57,61%**	**66%**	34%
	0,597	0,497	0,543	0,573	No		50%	**50%**
	0,579	0,576	0,573	0,573				
IW-RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				
AB-RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				
RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				

Πίνακας 8: Πρόβλεψη Decision Table ταξινομητή με δείκτη 10% required

Prediction of indicator 10% required								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,546	0,639	0,589	0,593	Yes	**55,31%**	**74%**	36%
	0,564	0,467	0,511	0,593	No		53%	**47%**
	0,555	0,553	0,55	0,593				
IW-AB-DL-RN	0,582	0,452	0,509	0,588	Yes	**56,34%**	**45%**	55%
	0,551	0,675	0,607	0,588	No		32%	**68%**
	0,567	0,563	0,558	0,588				
IW-AB-RN	0,542	0,749	0,629	0,582	Yes	**55,77%**	**75%**	25%
	0,593	0,366	0,453	0,582	No		63%	**33%**
	0,567	0,558	0,541	0,582				
IW-DL-RN	0,582	0,452	0,509	0,588	Yes	**56,34%**	**45%**	55%
	0,551	0,675	0,607	0,588	No		32%	**68%**
	0,567	0,563	0,558	0,588				
IW-HP-RN	0,601	0,372	0,460	0,582	Yes	**56,22%**	**37%**	63%
	0,545	0,753	0,632	0,582	No		25%	**75%**
	0,573	0,562	0,546	0,582				
IW-RN	0,542	0,749	0,629	0,582	Yes	**55,77%**	**75%**	25%
	0,593	0,366	0,453	0,582	No		63%	**37%**
	0,567	0,558	0,541	0,582				
AB-RN	0,533	0,71	0,609	0,553	Yes	**54,40%**	**71%**	29%
	0,565	0,378	0,453	0,553	No		62%	**38%**
	0,549	0,544	0,531	0,553				
RN	0,534	0,674	0,596	0,543	Yes	**54,28%**	**67%**	33%
	0,557	0,412	0,474	0,543	No		59%	**41%**
	0,546	0,543	0,535	0,543				

Πίνακας 9 Πρόβλεψη Decision Table ταξινομητή με δείκτη DCEN

Prediction of indicator DCEN								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,604	0,709	0,652	0,628	Yes	**61,03%**	**71%**	29%
	0,620	0,506	0,557	0,628	No		49%	**51%**
	0,612	0,61	0,606	0,628				
IW-AB-DL-RN	0,584	0,651	0,616	0,614	Yes	**58,07%**	**65%**	35%
	0,576	0,506	0,539	0,614	No		49%	**51%**
	0,580	0,581	0,578	0,614				
IW-AB-RN	0,584	0,651	0,616	0,614	Yes	**58,07%**	**65%**	35%
	0,576	0,506	0,539	0,614	No		49%	**51%**
	0,58	0,581	0,578	0,614				
IW-DL-RN	0,603	0,665	0,632	0,612	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,612	No		47%	**53%**
	0,601	0,601	0,599	0,612				
IW-HP-RN	0,604	0,709	0,652	0,628	Yes	**61,03%**	**71%**	29%
	0,620	0,506	0,557	0,628	No		49%	**51%**
	0,612	0,610	0,606	0,628				
IW-RN	0,603	0,665	0,632	0,612	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,612	No		47%	**53%**
	0,601	0,601	0,599	0,612				
AB-RN	0,620	0,428	0,506	0,574	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,574	No		29%	**72%**
	0,582	0,570	0,561	0,574				
RN	0,620	0,428	0,506	0,574	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,574	No		28%	**72%**
	0,582	0,570	0,561	0,574				

Ακολουθεί η ανάλυση των παραπάνω στοιχείων ανά πίνακα:

- **Δείκτης 10 % actual (Πίνακας 7)**

Εύρος ακρίβειας: 55,84% έως 56,92%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’IW-AB-DL-HP-RN’’ με 55,84%

Δοκιμή υψηλότερης τιμής ακρίβειας: ‘’IW-HP-RN’’ με 57,61 %

Στον πίνακα 7 παρατηρούνται ίδια αποτελέσματα με τον IBk (Πίνακας 2), με τον Decision Table να επιτυγχάνει ακρίβεια στο εύρος 55,84% έως 57,61%, το οποίο θεωρείται σχετικά καλό. Η μεγαλύτερη ακρίβεια συναντάται στη δοκιμή συνδυασμού των τριών υποκειμενικών δεικτών ‘’IW-HP-RN’’ με 57,61%, ενώ η χαμηλότερη στην περίπτωση συνδυασμού και των πέντε υποκειμενικών δεικτών ‘’IW-AB-DL-HP-RN’’ με 55,84%. Σε αυτή τη περίπτωση ταξινομητή, τη μεγαλύτερη θετική επίδραση φαίνεται να έχει ο δείκτης ‘’HP-Health Problems’’, όπου στην περίπτωση με τη μεγαλύτερη ακρίβεια ‘’IW-HP-RN’’, είναι πιθανό να αύξησε ελαφρώς την ακρίβεια, ενώ οι υποκειμενικοί δείκτες ‘’IW- Inability to heat home’’ και ‘’RN- Restrictions of other needs’’ φαίνεται να δίνουν μια σταθερή ικανοποιητική ακρίβεια σε κάθε περίπτωση.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area παραμένουν σε ικανοποιητικά επίπεδα άνω του 0,5 (0,564 έως 0,579) για το σταθμισμένο μέσο όρο. Αξιοσημείωτο χαρακτηριστικό είναι το μικρότερο εύρος αυτών των στατιστικών δεικτών σε σχέση με τους προηγούμενους ταξινομητές (Εύρος τιμών: 0,013 μονάδες). Αυτό το χαρακτηριστικό δείχνει τη σχετική ομοιομορφία όλων των δοκιμών όσον αφορά τη συνολική ακρίβεια και ανάκληση του αλγορίθμου και την παρόμοια συμπεριφορά του αλγορίθμου απέναντι σε όλες τις δοκιμές. Αυτή η σχετική ομοιομορφία επικρατεί ως ένα επίπεδο σε όλους τους αλγορίθμους που χρησιμοποιήθηκαν μέχρι τώρα για τον δείκτη 10% actual και δείχνει ότι οι τρεις πρώτοι αλγόριθμοι, δίνουν σχετικά όμοια αποτελέσματα.

Τα confusion matrix κάθε δοκιμής παραμένουν σε ικανοποιητικό επίπεδο. Πιο συγκεκριμένα, ο συνδυασμός πέντε υποκειμενικών δεικτών ‘’IW-AB-DL-HP-RN’’ και ο συνδυασμός τριών υποκειμενικών δεικτών ‘’IW-HP-RN’’ έδωσε τα πιο ισορροπημένα αποτελέσματα, σχετικά με τις σωστές προβλέψεις, όπου επετεύχθη ποσοστό σωστών προβλέψεων άνω του 50% και για τις δύο κλάσεις, με τη δοκιμή ‘’IW-HP-RN’’ να επιτυγχάνει 66% σωστών προβλέψεων ενεργειακά φτωχών δειγμάτων. Σύμφωνα με τον ταξινομητή Decision Table, το 66% των περιπτώσεων που υποφέρουν από τους τρεις υποκειμενικούς δείκτες είναι ενεργειακά φτωχές. Συνεπώς, σε αυτό το σημείο, και οι τρεις αλγόριθμοι έχουν δείξει ότι την πιθανή ισχυρή σύνδεση που υπάρχει μεταξύ υποκειμενικών δεικτών και ενεργειακής φτώχειας

- **Δείκτης 10% required (Πίνακας 8)**

Εύρος ακρίβειας : 54,28% έως 56,34%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’RN’’ με 54,28%

Δοκιμή υψηλότερης τιμής ακρίβειας: ‘’IW-DL-RN’’ 56,34%

Όπως και στα προηγούμενα μοντέλα η ακρίβεια των δοκιμών παραμένει σε ικανοποιητικά επίπεδα, αλλά όπως και στα μοντέλα των προηγούμενων αλγορίθμων η ακρίβεια των δοκιμών

για τον δείκτη 10% required είναι ελαφρώς μειωμένη σε σχέση με τα μοντέλα του αντικειμενικού δείκτη 10% actual και αρκετά πιο μειωμένη σε σχέση με τα μοντέλα του δείκτη DCEN, για ορισμένες δοκιμές. Σύμφωνα, λοιπόν με τις δοκιμές , η μοντελοποιήση με σκοπό την πρόβλεψη του αντικειμενικού δείκτη 10% required με τη χρήση των πέντε υποκειμενικών δεικτών είναι ελαφρώς πιο ''άκαρπη'' αναφορικά με την ακρίβεια των προβλέψεων του αλγορίθμου. Ο ''RN-Restriction of other needs'' φαίνεται να παραμένει ο δείκτης ο οποίος, στα μοντέλα με βάση τον 10% required σημειώνει την μικρότερη τιμή ακρίβειας, κάτι το οποίο πιθανώς να οφείλεται στην δυσκολία των αλγορίθμων να προβλέψουν με υψηλότερα ποσοστά ακρίβειας όταν χρησιμοποιούν και τις τιμές του δείκτη ''RN''.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area έχουν ικανοποιητικές τιμές για τον σταθμισμένο μέσο όρο του συνόλου τους, άνω του 0,5, με εύρος από 0,531 έως 0,593. Συμπληρωματικό στοιχείο για την αδυναμία των αλγορίθμων να λειτουργήσουν αποδοτικά όταν στα δεδομένα εισόδου περιλαμβάνεται και ο δείκτης ''RN'' αποτελούν οι τιμές του ROC Area για την δοκιμή με μοναδικό υποκειμενικό δείκτη τον ''RN'', όπου τιμή ισούται με 0,543, ενώ με τον συνδυασμό και άλλων υποκειμενικών δεικτών η τιμή αυξάνεται σταδιακά σε 0,553 και φτάνει το 0,593 με χρήση και των πέντε δεικτών.

Τέλος στα confusion matrix της κάθε δοκιμής, παρατηρείται και πάλι ανισορροπία μεταξύ των ποσοστών σωστών προβλέψεων για κάθε κλάση, με τις δοκιμές ''IW-AB-RN'' , ''IW-HP-RN'', ''IW-RN'', ''AB-RN'' να εμφανίζουν έντονη ανισορροπία, με αποτέλεσμα η μία κλάση να προβλέπεται σωστά σε μεγάλό ποσοστό, ενώ η άλλη να έχει σε συντριπτική πλειονότητα λανθασμένες προβλέψεις. Το γεγονός αυτό, σε πραγματικό επίπεδο, μεταφράζεται ως λανθασμένη ταξινόμηση ενεργειακά φτωχών νοικοκυριών ως μη ενεργειακά φτωχά, προκαλώντας έντονες κοινωνικές αδικίες, λόγω λανθασμένης ταξινόμησης του αλγορίθμου μηχανικής μάθησης. Σε αυτό το σημείο γίνεται ξεκάθαρό και το αντίκτυπο που μπορεί να έχει μια λανθασμένη ταξινόμηση των ενεργειακά φτωχών - και μη- νοικοκυριών.

- **Δείκτης DCEN (Πίνακας 9)**

Εύρος ακρίβειας: 56,95 έως 61,03%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ''AB-RN'',''RN '' με 56,95%

Δοκιμή υψηλότερης τιμής ακρίβειας: ''IW-AB-DL-HP-RN'', ''IW-HP-RN''

Σε ένα ακόμη αλγόριθμο φαίνεται η υψηλή ακρίβεια των υπολογισμών που προκύπτουν από την μοντελοποίηση με βάση τον δείκτη DCEN, όπου η ακρίβεια ξεπερνάει σε όλες τις περιπτώσεις το 56%. Σύμφωνα λοιπόν με τα αποτελέσματα, οι αλγόριθμοι του WEKA μοντελοποιούν και προβλέπουν σαφώς ακριβέστερα, όταν λαμβάνουν ως αντικειμενικό δείκτη τον DCEN.Επίσης, οι δείκτες ''AB-Arrears in Bills'',''RN-Restriction of other needs'' , φαίνεται να λειτουργούν ελαφρώς χειρότερα από τους υπόλοιπους, καθώς στις δοκιμές που περιλαμβάνονται μόνοι τους ή ως ζευγάρι, ακρίβεια της δοκιμής μειώνεται έως και 4%.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area έχουν ικανοποιητικές τιμές για τον σταθμισμένο μέσο όρο του συνόλου τους, άνω του 0,5 , με το εύρος να κυμαίνεται από 0,561 έως 0,628. Η μέτρια απόδοση των δεικτών ''AB'',''RN'' εξηγείται και από την τιμή του ROC Area για τις περιπτώσεις ''AB-RN'', ''RN'', όπου η τιμή φτάνει το 0,574 , σαφώς μικρότερη από αυτή των άλλων περιπτώσεων (> 0,612), αναδεικνύοντας την μέτρια απόδοση τους σε γενικό επίπεδο και σε τυχαία δεδομένα.

Τα confusion matrix λαμβάνουν αρκετά καλές τιμές για όλες τις περιπτώσεις πέραν των ''AB-RN'', ''RN'', όπου εντοπίζεται σαφής ανισορροπία μεταξύ των σωστών προβλέψεων για κάθε κλάση, με 72% σωστές προβλέψεις για τις αρνητικές περιπτώσεις (μη ενεργειακά φτωχά νοικοκυριά) και 43% σωστές προβλέψεις για θετικές περιπτώσεις (ενεργειακά φτωχά νοικοκυριά). Αντίθετα στις υπόλοιπες περιπτώσεις, και στις δύο κλάσεις το ποσοστό των σωστών προβλέψεων είναι άνω του 50%,με μία εκ των δύο κλάσεων να σημειώνει και ποσοστά σωστών προβλέψεων άνω του 65%, υποδηλώνοντας την σωστή πρόβλεψη της πλειοψηφίας των περιπτώσεων.

5.1.4. J48

Στους επόμενους πίνακες (Πίν. 10,11,12) παρουσιάζονται τα αποτελέσματα μοντελοποίησης των δεδομένων με χρήση του ταξινομητή J48 (Decision tree algorithm) για τους τρεις αντικειμενικούς δείκτες. Τα τελικά στατιστικά αποτελέσματα είναι σχετικά ικανοποιητικά με συγκεκριμένες περιπτώσεις να λειτουργούν λιγότερο αποδοτικά συγκριτικά με τις δοκιμές των ίδιων περιπτώσεων στους προηγούμενους ταξινομητές.

Πίνακας 10 : Πρόβλεψη J48 ταξινομητή με δείκτη 10% actual

Prediction of indicator 10% actual								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,565	0,581	0,573	0,564	Yes	**55,06%**	**58%**	42%
	0,534	0,518	0,526	0,564	No		48%	**52%**
	0,550	0,551	0,55	0,564				
IW-AB-DL-RN	0,561	0,671	0,611	0,553	Yes	**55,73%**	**67%**	33%
	0,551	0,435	0,486	0,553	No		57%	**43%**
	0,556	0,557	0,551	0,553				
IW-AB-RN	0,561	0,671	0,611	0,553	Yes	**55,73%**	**67%**	33%
	0,551	0,435	0,486	0,553	No		57%	**43%**
	0,556	0,557	0,551	0,553				
IW-DL-RN	0,561	0,671	0,611	0,553	Yes	**55,73%**	**67%**	33%
	0,551	0,435	0,486	0,553	No		57%	**43%**
	0,556	0,557	0,551	0,553				

IW-HP-RN	0,553	0,527	0,539	0,548	Yes	**53,33%**	**53%**	47%
	0,515	0,540	0,527	0,548	No		46%	**54%**
	0,534	0,533	0,533	0,548				
IW-RN	0,561	0,671	0,611	0,553	Yes	**55,73%**	**67%**	33%
	0,551	0,435	0,486	0,553	No		57%	**43%**
	0,556	0,557	0,551	0,553				
AB-RN	0,561	0,671	0,611	0,553	Yes	**55,73%**	**67%**	33%
	0,551	0,435	0,486	0,553	No		57%	**43%**
	0,556	0,557	0,551	0,553				
RN	0,561	0,671	0,611	0,553	Yes	**55,73%**	**67%**	33%
	0,551	0,435	0,486	0,553	No		57%	**43%**
	0,556	0,557	0,551	0,553				

Πίνακας 11: Πρόβλεψη J48 ταξινομητή με δείκτη 10% required

Prediction of indicator 10% required								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,584	0,407	0,479	0,573	Yes	**55,59%**	**41%**	59%
	0,541	0,707	0,613	0,573	No		29%	**71%**
	0,562	0,556	0,546	0,573				
IW-AB-DL-RN	0,582	0,452	0,509	0,584	Yes	**56,34%**	**45%**	55%
	0,551	0,675	0,607	0,584	No		32%	**68%**
	0,567	0,563	0,558	0,584				
IW-AB-RN	0,607	0,331	0,428	0,558	Yes	**55,77%**	**33%**	67%
	0,539	0,785	0,639	0,558	No		22%	**78%**
	0,573	0,558	0,534	0,558				
IW-DL-RN	0,545	0,678	0,604	0,588	Yes	**55,54%**	**68%**	32%
	0,573	0,432	0,493	0,588	No		57%	**43%**
	0,559	0,555	0,549	0,588				
IW-HP-RN	0,607	0,331	0,428	0,558	Yes	**55,74%**	**33%**	67%
	0,539	0,785	0,639	0,558	No		22%	**78%**
	0,573	0,558	0,534	0,558				
IW-RN	0,607	0,331	0,428	0,558	Yes	**55,77%**	**33%**	67%
	0,539	0,785	0,639	0,558	No		22%	**78%**
	0,573	0,558	0,534	0,558				
AB-RN	0,545	0,678	0,604	0,588	Yes	**55,54%**	**68%**	32%

	0,573	0,432	0,493	0,588	No		57%	**43%**
	0,559	0,555	0,549	0,588				
RN	0,534	0,674	0,596	0,543	Yes	**54,28%**	**67%**	33%
	0,557	0,412	0,474	0,543	No		59%	**41%**
	0,546	0,543	0,535	0,543				

Πίνακας 12: Πρόβλεψη J48 ταξινομητή με δείκτη DCEN

Prediction of indicator 10% DCEN								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,604	0,709	0,652	0,621	Yes	**61,03%**	**71%**	29%
	0,620	0,506	0,557	0,621	No		49%	**51%**
	0,612	0,61	0,606	0,621				
IW-AB-DL-RN	0,603	0,665	0,632	0,606	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,606	No		47%	**53%**
	0,601	0,601	0,599	0,606				
IW-AB-RN	0,603	0,665	0,632	0,606	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,606	No		47%	**53%**
	0,601	0,601	0,599	0,606				
IW-DL-RN	0,603	0,665	0,632	0,606	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,606	No		47%	**53%**
	0,601	0,601	0,599	0,606				
IW-HP-RN	0,604	0,709	0,652	0,621	Yes	**61,04%**	**71%**	29%
	0,620	0,506	0,557	0,621	No		49%	**51%**
	0,612	0,610	0,606	0,621				
IW-RN	0,603	0,665	0,632	0,606	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,606	No		47%	**53%**
	0,601	0,601	0,599	0,606				
AB-RN	0,620	0,428	0,506	0,574	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,574	No		28%	**72%**
	0,582	0,570	0,561	0,574				
RN	0,620	0,428	0,506	0,574	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,574	No		28%	**72%**
	0,582	0,570	0,561	0,574				

Ακολουθεί η ανάλυση των παραπάνω στοιχείων ανά πίνακα:

- **Δείκτης 10% actual (Πίνακας 10)**

Εύρος ακρίβειας: 53,73% έως 55,73%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’ IW-HP-RN’’ με 53,33%

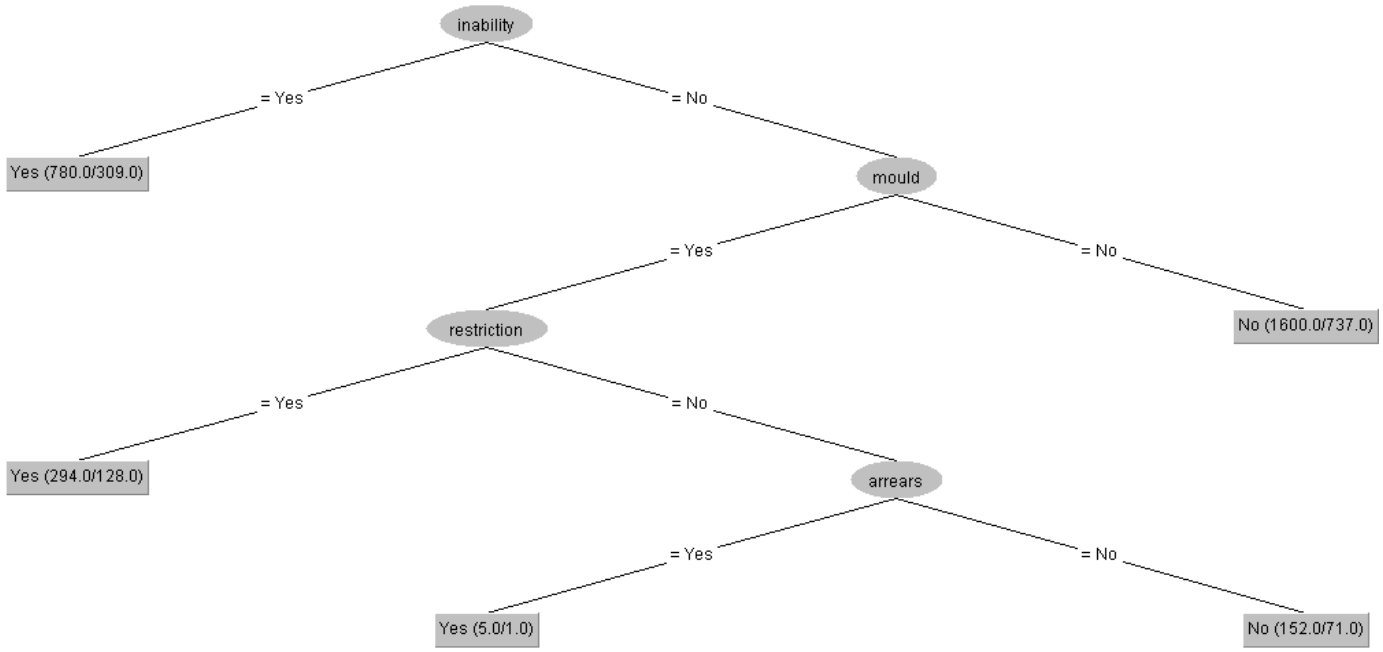
Δοκιμή υψηλότερης τιμής ακρίβειας: ‘’IW-AB-DL-RN’’, ‘’IW-AB-RN’’, ‘’IW-DL-RN’’, ‘’IW-RN’’, ’’RN’’

Στον πίνακα 10 δίνονται τα αποτελέσματα των προβλέψεων του ταξινομήτη J48 για τον δείκτη 10% actual. Όπως φαίνεται, η ακρίβεια του ταξινομήτη είναι ελαφρώς χαμηλότερη σε σχέση με τους προηγουμένους ταξινομητές, καθώς κυμαίνεται από 55,06% έως 55,73%. Η περίπτωση με τη χαμηλότερη ακρίβεια είναι αυτή με τη χρήση και των πέντε υποκειμενικών δεικτών ‘’IW-AB-DL-HP-RN’’, με 55,06% , ενώ παρατηρούνται πέντε δοκιμές με την υψηλότερη τιμή ακρίβειας για το συγκεκριμένο αλγόριθμο , με τις περιπτώσεις ‘’ IW-AB,DL,RN’’, ‘’IW-AB-RN’’, ‘’IW-DL-RN’’, ‘’IW-RN’’, ‘’AB-RN’’, ‘’RN’’ να επιτυγχάνουν ακρίβεια 55,73%. Επίσης, παρατηρείται ότι το εύρος της ακρίβειας είναι περιορισμένο σε 0,67 ποσοστιαίες μονάδες, υποδεικνύοντας την σχετικά αμετάβλητη λειτουργία αλγορίθμου ανεξάρτητα από τον αριθμό υποκειμενικών δεικτών που λαμβάνονται ως δεδομένα εισόδου. Συμπληρωματική σημείωση θα μπορούσε να προστεθεί για τον υποκειμενικό δείκτη ‘’HP-Health Problems’’, όπου στις δοκιμές όπου λαμβάνεται ως ένας εκ των δεδομένων εισόδου, η ακρίβεια είναι ελαφρώς μειωμένη, συμπεραίνοντας ότι ο αλγόριθμος εκπαιδεύεται λιγότερο αποτελεσματικά με την παρουσία των δεδομένων του δείκτη αυτού. Τελικά, όσον αφορά την ακρίβεια ο J48 αποτελεί τον λιγότερό λειτουργικό δείκτη από τους έξι μελετώμενους.

Όσον αφορά, τους στατιστικούς δείκτες Precision, Recall, F-Measure και ROC Area, και σε αυτόν τον αλγόριθμό παραμένουν οριακά σε ικανοποιητικό επίπεδο άνω του 0,5 με εύρος από 0,551 έως 0,557 στο σταθμισμένο μέσο όρο των δύο κλάσεων. Μπορούν να θεωρηθούν ως λίγο πιο αδύναμοι σε σχέση με τα αποτελέσματα που έχουν δώσει στους προηγούμενες αλγορίθμους, αλλά εντός δεκτών επιπέδων.

Τα confusion matrix είναι και αυτά σε σχετικά αποδεκτά επίπεδα για τη φύση του μελετώμενου φαινομένου , εκτός συγκεκριμένων περιπτώσεων. Πιο συγκεκριμένα, οι δοκιμές ‘’ IW-AB-DL-RN’’, ‘’IW-AB-RN’’, ’’IW-DL-RN’’, ‘’IW-RN’’, ’’AB-RN’’, ‘’RN’’ παρουσιάζουν μια ανισορροπία στις σωστές προβλέψεις μεταξύ των δύο κλάσεων, όπου οι σωστές θετικές προβλέψεις φτάνουν το 67% , ενώ οι αρνητικές σωστές προβλέψεις το 43%. Αυτή η ανισορροπία μεταφράζεται ως, ακριβής πρόβλεψη αρκετών ενεργειακά φτωχών περιπτώσεων, αλλά και λάθος ταξινόμηση ορισμένων ενεργειακά φτωχών περιπτώσεων ως μη-ενεργειακά φτωχές.

Στο παρακάτω διάγραμμα (Διάγ. 1) δίνεται το δέντρο αποφάσεων που σχημάτισε ο J48, βάσει του οποίου ταξινομήθηκαν τα δεδομένα για την περίπτωση υψηλότερης ακρίβειας



Διάγραμμα 1: Δέντρο αποφάσεων J48 για την περίπτωση μοντελοποιήσης με βάση τον δείκτη 10% actual για τη δοκιμή ''IW-AB-DL--RN''

- **Δείκτης 10% required (Πίνακας 11):**

Εύρος ακρίβειας: 54,28% έως 56,34%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ''RN'' με 54,28%

Δοκιμή υψηλότερης τιμής ακρίβειας: ''IW-AB-DL-RN'' με 56,34%

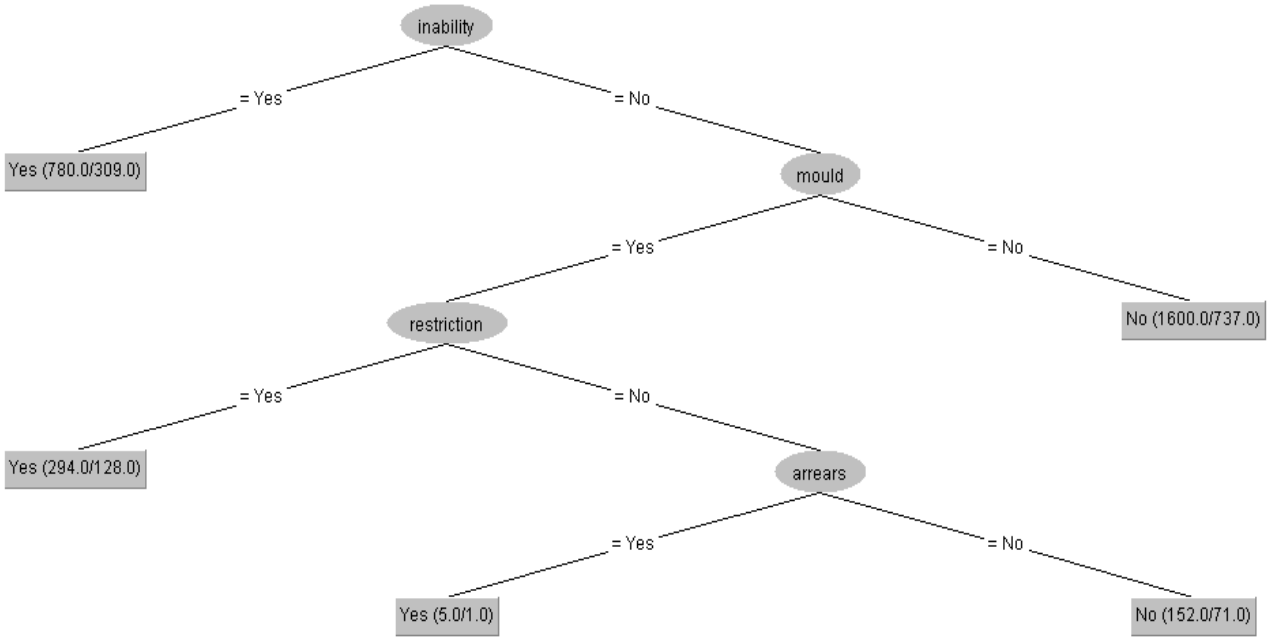
Το εύρος ακρίβειας για την μοντελοποίηση με τον δείκτη 10% required κινείται σε σχετικά ικανοποιητικά επίπεδα, καθώς σύμφωνα και με τα αποτελέσματα των προηγούμενων ταξινομητών, τα μοντέλα για τον δείκτη αυτόν σημειώνουν ελαφρώς χειρότερα αποτελέσματα ακρίβειας σε σχέση με τα μοντέλα των άλλων δύο αντικειμενικών δεικτών. Αξιοσημείωτη είναι η ακρίβεια που σημειώνει η δοκιμή ''IW-AB-DL-RN'' με 56,34%, όπου φαίνεται να υπερτερεί σε σχέση με τις άλλες δοκιμές που κινούνται σε ελαφρώς χαμηλότερα επίπεδα, με τον J48 να δημιουργεί καλύτερες προβλέψεις χρησιμοποιώντας τέσσερις υποκειμενικούς δείκτες. Από την άλλη πλευρά, ο δείκτης ''RN'' φαίνεται να υπολειτουργεί και να περιορίζει την ακρίβεια των δοκιμών και σε αυτόν τον αλγόριθμο, υποδεικνύοντας ότι ο συγκεκριμένος είναι ακατάλληλος

για χρήση στις συγκεκριμένες μοντελοποιήσεις, περιπλέκοντας την επεξεργασία του αλγόριθμου και υποβαθμίζοντας τα αποτελέσματα.

Όσον αφορά, τους στατιστικούς δείκτες Precision, Recall, F-Measure και ROC Area, οι τιμές του κυμαίνονται σε ικανοποιητικά επίπεδο για τον σταθμισμένο μέσο όρο άνω του 0,5 (0,534 έως 0,573), χωρίς ωστόσο, κάποιον δείκτη να ξεπερνάει τη τιμή 0,6, υποδεικνύοντας ότι οι στατιστικοί δείκτες βρίσκονται σε απλώς ικανοποιητικά επίπεδα, χωρίς κάποια δοκιμή να ξεχωρίζει. Επίσης οι τιμές ROC δείχνουν, ότι ο ταξινομητής λειτουργεί ικανοποιητικά και σε τυχαία δεδομένα.

Τα confusion matrix κυμαίνονται σε μέτρια επίπεδα, με την ανισορροπία να βρίσκεται παρούσα σε κάθε περίπτωση, είτε έντονή ή πιο ανεπαίσθητη. Πιο συγκεκριμένα, σε καμία περίπτωση δεν παρατηρείται ποσοστό άνω του 50% για σωστές προβλέψεις και για τις δύο κλάσεις, στατιστικό το οποίο θα έδειχνε μια έγκυρη δοκιμή των δεδομένων. Αντίθετα, η πιο ισορροπημένη περίπτωση αποτελεί η δοκιμή ‘’ IW-AB-DL-RN’’ όπου οι σωστές θετικές προβλέψεις είναι στο 45% και οι σωστές αρνητικές προβλέψεις σε ποσοστό 68%, όπου η ανισορροπία είναι σχετικά περιορισμένη , αλλά όχι απούσα. Στις υπόλοιπες περιπτώσεις επικρατεί εντονότερη ανισορροπία μεταξύ των σωστών προβλέψεων.

Στο παρακάτω διάγραμμα (Διάγ. 2) δίνεται το δέντρο αποφάσεων που σχημάτισε ο J48, βάσει του οποίου ταξινομήθηκαν τα δεδομένα για τη δοκιμή με το μεγαλύτερο ποσοστό ακρίβειας.



Διάγραμμα 2: Δέντρο αποφάσεων J48 για την περίπτωση μοντελοποιήσης με βάση τον δείκτη 10% required για τη δοκιμή ‘’IW-AB-DL--RN’’

- **Δείκτης DCEN (Πίνακας 12)**

Εύρος ακρίβειας : 56,95% έως 61,04%

Δοκιμή χαμηλότερης τιμής ακρίβειας : ‘’ AB-RN’’ , ‘’RN’’ με 56,95%

Δοκιμή υψηλότερης τιμής ακρίβειας : ‘’IW-HP-RN’’ με 61,04%

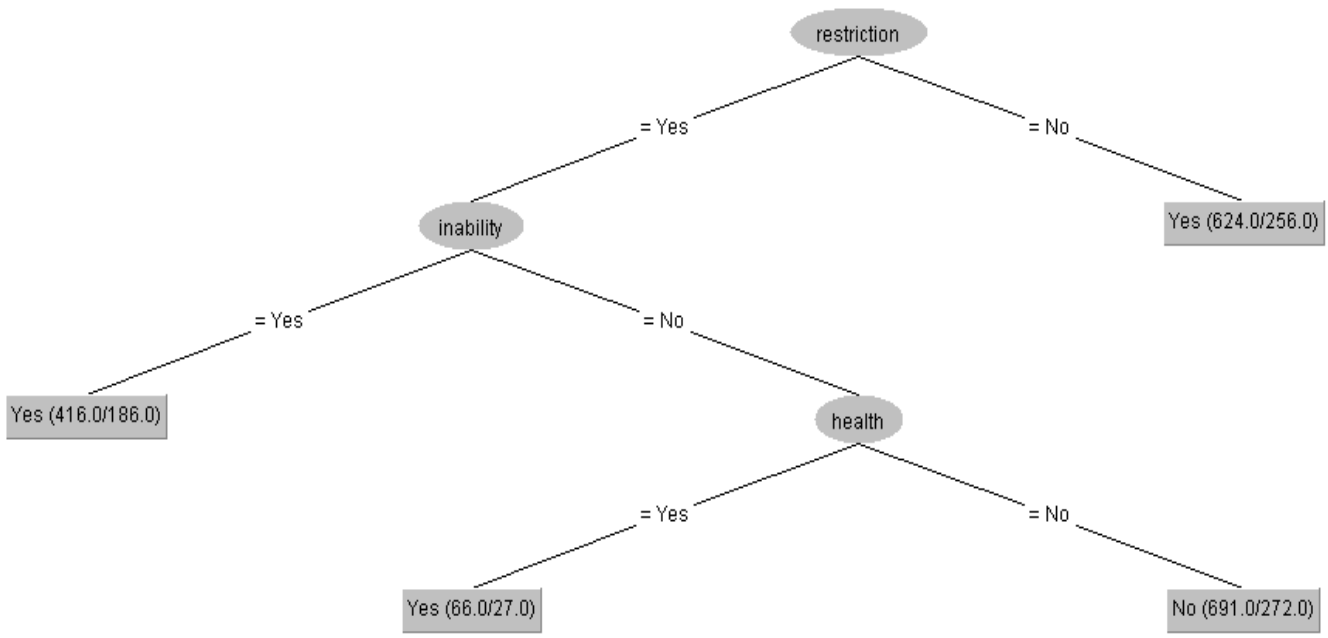
Άλλη μία περίπτωση όπου αναδεικνύεται η αυξημένη ακρίβεια μοντελοποίησης των δεδομένων με βάση τον δείκτη DCEN συγκριτικά με τους υπόλοιπους αντικειμενικούς δείκτες, με την ακρίβεια για όλες τις δοκιμές να ξεπερνάει το 56,90 %, αναδεικνύοντας τον δείκτη DCEN ως τον ακριβέστερο, στη μέθοδο πρόβλεψης αντικειμενικών δεικτών με τη χρήση υποκειμενικών δεικτών, μέσω αλγορίθμων μηχανικής μάθησης. Επιπλέον, για άλλη μια φορά, ο δείκτης ‘’RN’’ εμφανίζει ελαφρώς χειρότερη ακρίβεια σε σχέση με τους υπόλοιπους αντικειμενικούς δείκτες, ενισχύοντας την υπόθεση ότι πιθανόν να είναι ασύμβατος με την εκπαίδευση των αλγορίθμων μηχανικής μάθησης του WEKA. Ωστόσο, ακόμη και με αυτή την υπόθεση οι δοκιμές επιτυγχάνουν σχετικά καλή ακρίβεια.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area βρίσκονται σε αρκετά ικανοποιητικά επίπεδα άνω του 0,6 (0,599 έως 0,621) για τους περισσότερους δείκτες σε κάθε δοκιμή, εκτός των ελαφρώς πιο αδύναμων δοκιμών ‘’AB-RN’’ ΚΑΙ ‘’RN’’ που κυμαίνονται από 0,561 έως 0,584. Όπως αναφέρθηκε και προηγουμένως, επίτευξη τιμών κοντά στο 0,6 ή και υψηλοτέρων είναι αρκετά ικανοποιητικά, σε περιπτώσεις μελέτης κοινωνικών φαινομένων όπως είναι και η ενεργειακή φτώχεια.

Τέλος τα confusion matrix δείχνουν αρκετά ικανοποιητικά και ισορροπημένα αποτελέσματα σωστών προβλέψεων, με συγκεκριμένες περιπτώσεις να επιτυγχάνουν και αρκετά υψηλότερες τιμές σωστών προβλέψεων. Εξαίρεση στα συνολικά ικανοποιητικά ποσοστά, αποτελούν οι δοκιμές ‘’ AB-RN’’, ‘’RN’’ όπου παρατηρείται ανισορροπία και χαμηλή εγκυρότητα προβλέψεων. Πιο συγκεκριμένα, η δοκιμή ‘’RN’’ παρουσιάζει ποσοστό 72% σωστών θετικών (ενεργειακά φτωχών προβλέψεων), το οποίο αποτελεί υψηλό ποσοστό, ενώ ταυτόχρονα το ποσοστό σωστών αρνητικών (μη ενεργειακά φτωχών) περιπτώσεων περιορίζεται στο χαμηλό 43%, το οποίο μπορεί να μην είναι το χαμηλότερο ποσοστό που έχει παρατηρηθεί στην έως τώρα μελέτη, αλλά δίνει ένα σαφές παράδειγμα ανισορροπίας. Από τη άλλη πλευρά, η πιο έγκυρη πρόβλεψη σημειώνει ποσοστό 71% σωστών θετικών προβλέψεων, ενώ ταυτόχρονα οι αρνητικές σωστές προβλέψεις διατηρούν ικανοποιητικό ποσοστό 51%.

Στο παρακάτω διάγραμμα (Διάγ. 3) δίνεται το δέντρο αποφάσεων που σχημάτισε ο J48, βάσει του οποίου ταξινομήθηκαν τα δεδομένα, για την δοκιμή ‘’IW-HP-RN’’ με το υψηλότερο ποσοστό ακρίβειας

.



Διάγραμμα 3: Δέντρο αποφάσεων J48 για την περίπτωση μοντελοποίησης με βάση τον δείκτη DCEN για τη δοκιμή ''IW-HP-RN''

5.1.5.AdaBoost.M1- Decision Table

Στους παρακάτω πίνακες (Πίν. 13,14,15) καταγράφονται τα στατιστικά αποτελέσματα των δοκιμών μοντελοποίησης των δεδομένων με βάση τους τρεις αντικειμενικούς δείκτες με τη χρήση του meta-learning ταξινομητή AdaBoost.M1, ο οποίος επιλέχθηκε για να ενισχύσει (boosting) τη ικανότητα ακριβούς μοντελοποίησης του ήδη χρησιμοποιημένου αλγορίθμου Decision Table. Τα αποτελέσματα που παρουσίασε για τους τρεις αντικειμενικούς δείκτες θεωρούνται σχετικά ικανοποιητικά, ενώ παρακάτω μελετάται και η πραγματική ενίσχυση που προσέφερε στον ταξινομητή Decision Table.

Πίνακας 12: Πρόβλεψη AdaBoost.M1- Decision Table ταξινομητή με δείκτη 10% actual

Prediction of indicator 10% actual								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,559	0,569	0,564	0,589	Yes	**56.52%**	**57%**	43%
	0,571	0,562	0,566	0,589	No		44%	**56%**
	0,565	0,565	0,565	0,589				

IW-AB-DL-RN	0,553	0,626	0,588	0,573	Yes	**56,52%**	**61%**	39%
	0,580	0,505	0,540	0,573	No		49%	**51%**
	0,567	0,565	0,564	0,573				
IW-AB-RN	0,552	0,684	0,611	0,570	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,570	No		54%	**46%**
	0,575	0,569	0,564	0,570				
IW-DL-RN	0,55	0,588	0,568	0,577	Yes	**55,84%**	**59%**	41%
	0,568	0,53	0,548	0,577	No		47%	**53%**
	0,559	0,558	0,558	0,577				
IW-HP-RN	0,561	0,657	0,605	0,573	Yes	**56,60%**	**66%**	34%
	0,597	0,497	0,543	0,573	No		50%	**50%**
	0,579	0,576	0,573	0,573				
IW-RN	0,552	0,684	0,611	0,573	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,573	No		54%	**46%**
	0,575	0,569	0,564	0,573				
AB-RN	0,552	0,684	0,611	0,568	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,568	No		54%	**46%**
	0,575	0,569	0,564	0,568				
RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				

Πίνακας 13: Πρόβλεψη AdaBoost.M1- Decision Table ταξινομητή με δείκτη 10% required

Prediction of indicator 10% required								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,612	0,324	0,424	0,592	Yes	**55,88%**	**32%**	68%
	0,540	0,794	0,643	0,592	No		21%	**79%**
	0,576	0,559	0,533	0,592				
IW-AB-DL-RN	0,584	0,454	0,511	0,585	Yes	**56,45%**	**45%**	55%
	0,552	0,675	0,608	0,585	No		32%	**68%**
	0,568	0,565	0,559	0,585				
IW-AB-RN	0,537	0,685	0,602	0,577	Yes	**54,62%**	**68%**	32%
	0,563	0,407	0,473	0,577	No		59%	**41%**
	0,550	0,546	0,537	0,577				
IW-DL-RN	0,582	0,452	0,509	0,588	Yes	**56,34%**	**45%**	55%
	0,551	0,675	0,607	0,588	No		32%	**68%**
	0,567	0,563	0,558	0,588				
IW-HP-RN	0,601	0,372	0,460	0,582	Yes	**56,22%**	**37%**	63%
	0,545	0,753	0,632	0,582	No		25%	**75%**
	0,573	0,562	0,546	0,582				

IW-RN	0,542	0,749	0,629	0,582	Yes	**55,77%**	**75%**	25%
	0,593	0,366	0,453	0,582	No		63%	**37%**
	0,567	0,558	0,541	0,582				
AB-RN	0,533	0,710	0,609	0,553	Yes	**54,40%**	**71%**	29%
	0,565	0,378	0,453	0,553	No		62%	**38%**
	0,549	0,544	0,531	0,553				
RN	0,534	0,674	0,596	0,543	Yes	**54,28%**	**67%**	33%
	0,557	0,412	0,474	0,543	No		59%	**41%**
	0,546	0,543	0,535	0,543				

Πίνακας 14: Πρόβλεψη AdaBoost.M1- Decision Table ταξινομητή με δείκτη DCEN

Prediction of indicator DCEN								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,585	0,691	0,634	0,619	Yes	**58,81%**	**69%**	31%
	0,592	0,479	0,53	0,619	No		52%	**48%**
	0,589	0,588	0,583	0,619				
IW-AB-DL-RN	0,584	0,651	0,616	0,621	Yes	**58,07%**	**65%**	35%
	0,576	0,506	0,539	0,621	No		49%	**51%**
	0,580	0,581	0,578	0,621				
IW-AB-RN	0,584	0,651	0,616	0,614	Yes	**58,07%**	**65%**	35%
	0,576	0,506	0,539	0,614	No		49%	**51%**
	0,580	0,581	0,578	0,614				
IW-DL-RN	0,603	0,665	0,632	0,624	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,624	No		47%	**53%**
	0,601	0,601	0,599	0,624				
IW-HP-RN	0,604	0,709	0,652	0,627	Yes	**61,04%**	**71%**	29%
	0,620	0,506	0,557	0,627	No		49%	**51%**
	0,612	0,610	0,606	0,627				
IW-RN	0,603	0,665	0,632	0,612	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,612	No		47%	**53%**
	0,601	0,601	0,599	0,612				
AB-RN	0,566	0,543	0,554	0,571	Yes	**54,91%**	**54%**	46%
	0,533	0,556	0,544	0,571	No		44%	**56%**
	0,55	0,549	0,549	0,571				
RN	0,620	0,428	0,506	0,574	Yes	**56,95%**	**43%**	57%
	0,542	0,720	0,618	0,574	No		28%	**72%**
	0,582	0,570	0,561	0,574				

Ακολουθεί η ανάλυση των παραπάνω στοιχείων ανά πίνακα:

- **Δείκτης 10% actual (Πίνακας 13)**

Εύρος ακρίβειας: 55,84% έως 56,92%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ''IW-DL-RN'' με 55,84%

Δοκιμή υψηλότερης τιμής ακρίβειας: ''IW-AB-RN'', ''IW-RN'', ''AB-RN'', ''RN'' με 56,92%

Στον Πίνακα 13, καταγράφονται τα αποτελέσματα της πρώτης περίπτωσης χρήσης αλγορίθμου meta-learning για την ενίσχυση της ικανότητας ταξινόμησης άλλου αλγορίθμου. Στη συγκεκριμένη περίπτωση χρησιμοποιήθηκε ο AdaBoost.M1 (Boosting algorithm), με σκοπό την ενίσχυση της λειτουργίας του αλγορίθμου Decision Table για την μοντελοποίηση βάσει του δείκτη 10% actual. Συγκρίνοντας τα αποτελέσματα του Πίνακα 13 με αυτά του Πίνακα 7 , παρατηρείται, πράγματι, μια περιορισμένη βελτίωση των επιδόσεων του Decision Table σε μερικούς από τους αναλυόμενους δείκτες, αλλά και μείωση της αποτελεσματικότητας σε κάποιούς άλλους.

Αρχικά παρατηρείται αύξηση της ακρίβειας σε μία μόνο περίπτωση, αυτή της δοκιμής με πέντε υποκειμενικούς δείκτες ''IW-AB-DL-HP-RN'' από 55,84% σε 56,52%. Μείωση της ακρίβειας παρατηρείται στις τρεις δοκιμές: '' IW-AB-DL-RN'' από 56,92% σε 56,52%, ''IW-DL-RN'' από 56,92% σε 55,84% και ''IW-HP-RN'' από 56,61% σε 56,60%. Επίσης, για τις δοκιμές ''IW-AB-RN'', ''IW-RN'', ''AB-RN'', ''RN'' το ποσοστό ακρίβειας των δοκιμών παρέμεινε σταθερό στο 56,92%.Το εύρος των τιμών ακρίβειας για τον AdaBoost.M1 κυμαίνεται από 55,84% έως 56,92%, ελαφρώς μειωμένο σε σχέση με το εύρος 55,84%-57,61% του Decision Table στο μοντέλο ταξινόμησης χωρίς τη χρήση αλγορίθμου boosting.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area κυμάνθηκαν σε παρόμοια ικανοποιητικά επίπεδα με τα προηγούμενα μοντέλα άνω του 0,5 για τον σταθμισμένο μέσο όρο, αλλά ελαφρώς μειωμένα για τις κατώτατες τιμές σε σχέση με τα αποτελέσματα του Decision Table. Πιο συγκεκριμένα οι δείκτες του AdaBoost.M1 κυμαίνονται από 0,558 έως 0,589, ενώ του Decision Table από 0,564 έως 0,579. Συνεπώς, η χρήση του αλγόριθμου ενίσχυσης διεύρυνε το εύρος τιμών των στατιστικών δεικτών, μεταβάλλοντας την απόδοση κάποιων περιπτώσεων ή κρατώντας την σταθερή για κάποιες άλλες.

Τέλος, τα confusion matrix των δοκιμών βρίσκονται σε σχετικά ικανοποιητικά επίπεδα. Πιο συγκεκριμένα, εκτός από τις περιπτώσεις ''IW-AB,DL-HP-RN'', ''IW-AB-DL-RN'', ''IW-DL-RN'', ''IW-HP-RN'', όπου οι σωστές προβλέψεις και για τις δύο κλάσεις είναι άνω του 50%, με τις θετικές σωστές προβλέψεις για τις δοκιμές ''IW-AB-DL-RN'', ''IW-HP-RN'' να υπερβαίνουν και το 60% ακρίβειας, οι υπόλοιπες προβλέψεις έχουν μια μικρή ανισορροπία σωστών προβλέψεων μεταξύ των κλάσεων. Συγκεκριμένα, οι σωστές θετικές προβλέψεις επιτυγχάνουν ακρίβεια 68%, ενώ οι αρνητικές σωστές προβλέψεις περιορίζονται στο 46%.

Συμπερασματικά, η δοκιμή του πρώτου ‘’Boosting meta-learning’’ αλγορίθμου για την ενίσχυση της απόδοσης χρησιμοποιούμενο αλγορίθμου έδειξε ότι αυτού του είδους οι αλγόριθμοι δεν έχουν την δυνατότητα να ενισχύσουν κάθε αλγόριθμο ή κάθε μορφή συνόλου δεδομένων. Η χρήση τους μπορεί να θεωρηθεί σημαντική σε πολλές περιπτώσεις, αλλά πρέπει να δοθεί η απαραίτητη προσοχή στα αποτελέσματα, ώστε να ελεγχθεί αν πράγματι δίνουν βελτιωμένα αποτελέσματα. Στη συγκεκριμένη περίπτωση , ο AdaBoost.M1 βελτίωσε συγκεκριμένα μέρη των αποτελεσμάτων, ενώ μείωσε την απόδυση κάποιων άλλων. Τελικά, η χρήση του θα μπορούσε να θεωρηθεί περιττή για την βελτίωση του αλγορίθμου Decision Table αναφορικά με συγκεκριμένους δείκτες αλλά και χρήσιμος για την ενίσχυση κάποιων άλλων.

- **Δείκτης 10% required (Πίνακας 14)**

Εύρος ακρίβειας: 54,28% έως 56,45%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’RN’’ με 54,28%

Δοκιμή υψηλότερης τιμής ακρίβειας: ‘’IW-AB-DL-RN’’ με 56,45%

Συγκρίνοντας τα αποτελέσματα του Πίνακα 14 με αυτά του Πίνακα 8 παρατηρούνται περιπτώσεις ενίσχυσης, μείωσης αλλά και μη μεταβολής της ακρίβειας μεταξύ των δοκιμών. Το εύρος ακρίβειας εμφανίζει μια λιγοστή επέκταση προς ανώτερες τιμές σε σχέση με την απλή χρήση του Decision Table του οποίου το εύρος κυμαίνεται από 54,28% έως 56,34%. Όσον αφορά της μεταβολές των ποσοστών ακριβείας των δοκιμών, ενίσχυση παρατηρείται στις δοκιμές ‘’IW-AB-DL-HP-RN’’ από 55,31% σε 55,88% και ‘’IW-AB-DL-RN’’ από 56,34% σε 56,45%, ενώ μείωση παρατηρείται στη περίπτωση ‘’IW-AB-DL’’ από 56,34% σε 54,62% . Για τις υπόλοιπες δοκιμές η ακρίβεια παραμένει σταθερή, χωρίς καμία βελτίωση, με τις τιμές να βρίσκονται σε σχετικά ικανοποιητικό επίπεδο άνω του 54%. Ο ενισχυμένος Decision Table,όσον αφορά την ακρίβεια, δεν έδωσε κάποια σαφή βελτίωση και παραμένει ταξινομητής ο οποίος έδωσε εύρος ακρίβειας λιγότερο ικανοποιητικό σε σχέση με τους άλλους ταξινομητές που αναλύθηκαν. Επίσης, όπως και στα αποτελέσματα του Decision Table αλλά και των περισσότερων μοντέλων, ο δείκτης ‘’RN’’ παρουσιάζει την χαμηλότερη ακρίβεια στη δοκιμή που χρησιμοποιείται ως ο μοναδικός υποκειμενικός δείκτης, ενώ η ακρίβεια αυξάνεται σταδιακά όσο προστίθενται και οι υπόλοιποι δείκτες υποδεικνύοντας τη χαμηλή ακρίβεια μοντελοποίησης του και την κατάσταση στην οποία οι υπόλοιποι αντικειμενικοί δείκτες λειτουργούν πιο αποδοτικά, ισορροπώντας την χαμηλή απόδοση του ‘’RN’’.

Αναφορικά με τους στατιστικούς δείκτες Precision, Recall, F-Measure και ROC Area βρίσκονται σε ικανοποιητικό επίπεδο άνω του 0,5 για τον σταθμισμένο μέσο όρο με εύρος από 0,533 έως 0,592, το οποίο ουσιαστικά είναι ίσο με το εύρος του Decision Table στις δοκιμές όπου μοντελοποιεί χωρίς ενίσχυση (0,531 έως 0,593). Συνεπώς, δεν παρατηρείται κάποια βελτίωση με τη χρήση του αλγόριθμου ενίσχυσης AdaBoost.M1..

Τέλος, τα confusion matrix δεν εμφανίζουν καμία βελτίωση με τη χρήση του AdaBoost.M1, καθώς ο αλγόριθμός ενίσχυσης δεν διόρθωσε τη ανισορροπία σωστών προβλέψεων που επικρατεί σε όλες τις περιπτώσεις. Αντίθετα, ο αλγόριθμος ενίσχυσης επιδείνωσε τις περιπτώσεις έντονης ανισορροπίας σε ορισμένες δοκιμές όπως η ''IW-AB-DL-HP-RN'', όπου το ποσοστό σωστών θετικών προβλέψεων βρίσκεται στο 32% και των σωστών αρνητικών προβλέψεων στο 79% , ενώ στην περίπτωση απλής χρήσης του Decision Table, το ποσοστό σωστών προβλέψεων ανέρχεται σε 74% και των σωστών αρνητικών σε 47%. Παρατηρείται λοιπόν, επέκταση της ήδη έντονης ανισορροπίας.

Συνοψίζοντας, η χρήση του AdaBoost.M1 και στην περίπτωση πρόβλεψης βάσει του αντικειμενικού δείκτη 10% required, δεν προσέφερε κάποια ουσιαστική βελτίωση σε οποιοδήποτε μελετώμενο στατιστικό δείκτη, πέραν λιγοστών εξαιρέσεων, ενώ ακόμη, σε ορισμένες περιπτώσεις προκάλεσε μείωση της ακρίβειας και της ικανότητας πρόβλεψης του μοντέλου όπως π.χ. το παράδειγμα που αναφέρθηκε για τα confusion matrix.

- **Δείκτης DCEN (Πίνακας 15)**

Εύρος ακρίβειας : 56,95% έως 61,03%

Δοκιμή χαμηλότερης τιμής ακρίβειας: 54,91%

Δοκιμή υψηλότερης τιμής ακρίβειας : 61,04%

Η υψηλή ακρίβεια του μοντέλου με βάση του αντικειμενικού δείκτη DCEN διατηρείται με τη χρήση του αλγορίθμου ενίσχυσης AdaBoost.M1 με τα στατιστικά αποτελέσματα να παραμένουν ικανοποιητικά, ωστόσο, δεν παρατηρείται κάποια αισθητή βελτίωση μέσω της ενίσχυσης του αλγορίθμου. Αντίθετα, παρατηρούνται συγκεκριμένες περιπτώσεις με μείωση τιμών των στατιστικών δεικτών. Ειδικότερα, αναφορικά με τα ποσοστά ακρίβειας των δοκιμών δεν παρατηρείται κάποια αύξηση στην ακρίβεια οποιασδήποτε δοκιμής, ενώ παρατηρείται μείωση της ακρίβειας στις περιπτώσεις, ''IW-AB-DL-HP-RN'' από 61,03% σε 58,07%, ''AB-RN'' από 56,95% σε 54,91%, με την ακρίβεια των υπόλοιπων δοκιμών να παραμένει σταθερή. Γενικά λοιπόν, τα ποσοστά ακρίβειας παραμένουν σταθερά και σχετικά ικανοποιητικά με ορισμένες περιπτώσεις να ξεπερνούν την 60% ακρίβεια προβλέψεων, με τον αλγόριθμο ενίσχυσης να προκαλεί μειώσεις της τάξης του 2%-3%, χωρίς όμως σοβαρή ''ζημιά'' στο μοντέλο του δείκτη DCEN.

Οι ειδικοί στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area παραμένουν σε ικανοποιητικό επίπεδο για το σταθμισμένο μέσο όρο των κλάσεων άνω του 0,55 , με εύρος 0,570 έως 0,627, χωρίς ιδιαίτερες βελτιώσεις σε σχέση με τιμές των ειδικών στατιστικών δεικτών για την απλή χρήση του Decision Table χωρίς ενίσχυση, όπου το εύρος κυμαίνεται από 0,561 έως 0,628.

Τα confusion matrix παραμένουν στα ίδια ικανοποιητικά και ισορροπημένα επίπεδα, χωρίς κάποια βελτίωση στα ποσοστά σωστών προβλέψεων, χάρη στη χρήση του ταξινομητή ενίσχυσης, με εξαίρεση την δοκιμή ''AB-RN'' όπου, παρόλη τη μείωση ακρίβειας της δοκιμής, τα ποσοστά σωστών προβλέψεων ισορρόπησαν με την τιμή και των δύο να κυμαίνεται άνω του 50

5.1.6. Cost-Sensitive Classifier- J48

Στους παρακάτω πίνακες (Πίν. 16,17,18) καταγράφονται τα στατιστικά αποτελέσματα μοντελοποίησης των δεδομένων βάσει των τριών αντικειμενικών δεικτών με τη χρήση του meta-learning αλγόριθμου Cost Sensitive Classifier, ο οποίος χρησιμοποιήθηκε για να βελτιώσει την απόδοση του J48 με τη μέθοδο άσκησης ποινής στον αλγόριθμο για κάθε λάθος πρόβλεψη.

Πίνακας 16: Πρόβλεψη Cost Sensitive Classifier – J48 ταξινομητή με δείκτη 10% actual

Prediction of indicator 10% actual								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,662	0,556	0,604	0,592	Yes	**56,87%**	**56%**	44%
	0,477	0,587	0,526	0,592	No		41%	**59%**
	0,586	0,569	0,572	0,592				
IW - AB - DL - RN	0,512	0,626	0,563	0,575	Yes	**56,58%**	**63%**	37%
	0,631	0,517	0,568	0,575	No		48%	**52%**
	0,578	0,566	0,566	0,575				
IW - AB - RN	0,451	0,493	0,471	0,561	Yes	**55,42%**	**49%**	51%
	0,636	0,595	0,615	0,561	No		40%	**60%**
	0,561	0,554	0,557	0,561				
IW - DL - RN	0,552	0,684	0,611	0,571	Yes	**56,92%**	**68%**	32%
	0,596	0,457	0,518	0,571	No		54%	**46%**
	0,575	0,569	0,564	0,571				
IW - HP - RN	0,561	0,657	0,605	0,576	Yes	**57,60%**	**66%**	34%
	0,597	0,497	0,543	0,576	No		50%	**50%**
	0,579	0,576	0,573	0,576				
IW - RN	0,618	0,671	0,643	0,547	Yes	**56,72%**	**67%**	33%
	0,481	0,424	0,450	0,547	No		58%	**42%**
	0,560	0,567	0,562	0,547				
AB - RN	0,618	0,671	0,643	0,547	Yes	**56,72%**	**67%**	33%
	0,481	0,424	0,450	0,547	No		58%	**42%**
	0,560	0,567	0,562	0,547				
RN	0,613	0,69	0,649	0,557	Yes	**57,53%**	**69%**	31%
	0,509	0,424	0,463	0,557	No		58%	**42%**
	0,568	0,575	0,569	0,557				

Πίνακας 17: : Πρόβλεψη Cost Sensitive Classifier – J48 ταξινομητή με δείκτη 10% required

Prediction of indicator 10% required								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,639	0,477	0,546	0,575	Yes	**55,79%**	**48%**	52%
	0,500	0,660	0,569	0,575	No		34%	**66%**
	0,578	0,558	0,556	0,575				
IW - AB - DL - RN	0,639	0,472	0,543	0,574	Yes	**55,66%**	**47%**	53%
	0,499	0,664	0,570	0,574	No		34%	**66%**
	0,577	0,557	0,555	0,574				
IW - AB - RN	0,559	0,384	0,455	0,541	Yes	**53,78%**	**38%**	62%
	0,527	0,693	0,599	0,541	No		31%	**69%**
	0,543	0,538	0,527	0,541				
IW - DL - RN	0,554	0,507	0,529	0,589	Yes	**58,24%**	**51%**	49%
	0,604	0,648	0,625	0,589	No		35%	**65%**
	0,580	0,582	0,581	0,589				
IW - HP - RN	0,558	0,369	0,444	0,538	Yes	**53,55%**	**37%**	63%
	0,525	0,705	0,601	0,538	No		30%	**70%**
	0,541	0,536	0,522	0,538				
IW - RN	0,526	0,709	0,604	0,568	Yes	**54,82%**	**71%**	29%
	0,590	0,396	0,474	0,568	No		60%	**40%**
	0,559	0,548	0,537	0,568				
AB - RN	0,516	0,643	0,573	0,536	Yes	**53,33%**	**64%**	36%
	0,560	0,430	0,486	0,536	No		57%	**43%**
	0,539	0,533	0,528	0,536				
RN	0,516	0,643	0,573	0,536	Yes	**53,33%**	**64%**	36%
	0,560	0,430	0,486	0,536	No		57%	**43%**
	0,539	0,533	0,528	0,536				

Πίνακας 18: Πρόβλεψη Cost Sensitive Classifier – J48 ταξινομητή με δείκτη DCEN

Prediction of indicator DCEN								
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Class**	**Accuracy**	**Confusion Matrix**	
IW-AB-DL-HP-RN	0,508	0,586	0,544	0,588	Yes	**55,41%**	**59%**	41%
	0,604	0,528	0,563	0,588	No		47%	**53%**

	0,561	0,554	0,555	0,588				
IW - AB - DL - RN	0,603	0,528	0,563	0,581	Yes	**55,14%**	**53%**	47%
	0,504	0,579	0,539	0,581	No		42%	**58%**
	0,558	0,551	0,552	0,581				
IW - AB - RN	0,616	0,601	0,608	0,620	Yes	**60,11%**	**60%**	40%
	0,586	0,602	0,594	0,620	No		40%	**60%**
	0,602	0,601	0,601	0,620				
IW - DL - RN	0,603	0,665	0,632	0,606	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,606	No		47%	**53%**
	0,601	0,601	0,599	0,606				
IW - HP - RN	0,604	0,709	0,652	0,615	Yes	**61,04%**	**71%**	29%
	0,620	0,506	0,557	0,615	No		49%	**51%**
	0,612	0,610	0,606	0,615				
IW - RN	0,603	0,665	0,632	0,606	Yes	**60,11%**	**67%**	33%
	0,599	0,533	0,564	0,606	No		47%	**53%**
	0,601	0,601	0,599	0,606				
AB - RN	0,566	0,543	0,554	0,569	Yes	**54,91%**	**54%**	46%
	0,533	0,556	0,544	0,569	No		44%	**56%**
	0,550	0,549	0,549	0,569				
RN	0,620	0,428	0,506	0,574	Yes	**56,96%**	**43%**	57%
	0,542	0,720	0,618	0,574	No		28%	**72%**
	0,582	0,570	0,561	0,574				

Ακολουθεί η ανάλυση των παραπάνω στοιχείων ανά πίνακα:

- **Δείκτης 10% actual (Πίνακας 16)**

Εύρος ακρίβειας: 55,42% έως 57,60%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ‘’IW-AB-RN’’ με 55,42%

Δοκιμή υψηλότερης τιμής ακρίβειας: ‘’IW-HP-RN’’ με 57,60%

Στη δεύτερη δοκιμή αλγορίθμου meta-learning για τη βελτίωση ήδη χρησιμοποιούμενου αλγορίθμου, τα αποτελέσματα του Cost-Sensitive classifier παρουσιάζουν μια σχετική ενίσχυση του αλγορίθμου J48 (Decision tree). Συγκεκριμένα, το εύρος ακρίβειας των δοκιμών κυμαίνεται από 55,42% έως 57,60% , αυξημένο σε σχέση με αυτό του J48 στην περίπτωση όπου μοντελοποιεί

χωρίς τη χρήση αλγορίθμου ενίσχυσης, όπου κυμαίνεται από 55,06% έως 55,73%. Η υψηλότερη τιμή ακρίβειας συναντάται στην δοκιμή των υποκειμενικών δεικτών ''IW-HP-RN'' με 57,60%, η οποία κατά τη χρήση του J48 χωρίς ενίσχυση φτάνει το 53,33%, επιτυγχάνοντας έτσι αύξηση της ακρίβειας κατά 3,97 ποσοστιαίες μονάδες. Γενικά σε όλες τις περιπτώσεις παρατηρείται αύξηση της ακρίβειας κατά τουλάχιστον 1,5 ποσοστιαίες μονάδες, με εξαίρεση την δοκιμή ''IW-AB-RN'', όπου παρατηρείται ελαφριά μείωση της ακρίβειας από 55,73% σε 55,42%. Συνοψίζοντας σχετικά με τα αποτελέσματα της βελτίωσης ακρίβειας του μοντέλου, ο Cost Sensitive Classifier, πράγματι, βελτίωσε στις περισσότερες δοκιμές την ακρίβεια, και φαίνεται να λειτουργεί καλύτερα από τον AdaBoost.M1 σε αυτό το μέρος των στατιστικών αποτελεσμάτων.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area βρίσκονται σε ικανοποιητικά επίπεδα για τον σταθμισμένο μέσο όρο άνω του 0,55 (55%) με το εύρος τους να κυμαίνεται από 0,554 έως 0,592, ελαφρώς πιο διευρυμένο και βελτιωμένο σε σχέση με το εύρος κατά την απλή χρήση του J48, όπου κυμαίνεται από 0,551 έως 0,557.

Βελτίωση σημειώθηκε και στα confusion matrix μερικών δοκιμών, με τα ποσοστά σωστών προβλέψεων να ισορροπούνται σε μεγαλύτερο βαθμό σε σχέση με τα ποσοστά του J48 στην περίπτωση όπου χρησιμοποιείται χωρίς αλγόριθμο meta-learning. Πιο αναλυτικά, οι δοκιμές ''IW-DL-RN'', ''IW-RN'', ''AB-RN'', ''RN'' διατήρησαν τα ποσοστά σωστών προβλέψεων σε μία σχετική ανισορροπία, με τον Cost sensitive, να μειώνει σε μικρότερο βαθμό αυτή την ανισορροπία, με τα ποσοστό των σωστών θετικών προβλέψεων των τεσσάρων δοκιμών να φτάνει το 63%, και των σωστών αρνητικών προβλέψεων το 43%. Όπως φαίνεται, η ανισορροπία εξακολουθεί να είναι παρούσα , ωστόσο ο αλγόριθμος μείωσε ελαφρά την διαφορά μεταξύ των σωστών προβλέψεων ανά κλάση (βλ. Πίνακες 10,16). Για τις υπόλοιπες δοκιμές, ο αλγόριθμος αποδεικνύεται καθοριστικός καθώς αυξάνει τα ποσοστά σωστών προβλέψεων των κλάσεων τα οποία είναι κάτω του 50% και εξασφαλίζει την ισορροπία μεταξύ των κλάσεων, καθιστώντας έτσι, τις συγκεκριμένες δοκιμές ως σχετικά έγκυρες προς εξαγωγή συμπερασμάτων. Εξαίρεση στις βελτιώσεις δοκιμών αποτελεί η δοκιμή ''IW-AB-DL-RN'', όπου η χρήση του Cost Sensitive φέρνει σε ανισορροπία τα ποσοστά σωστών προβλέψεων της δοκιμής με 67% σωστές θετικές και 43% σωστές αρνητικές, ενώ χωρίς τη χρήση του, η δοκιμή βρίσκεται σε ισορροπία με 63% σωστές θετικές και 52% σωστές αρνητικές προβλέψεις.

- **Δείκτης 10% required (Πίνακας 17)**

Εύρος ακρίβειας: 53,33% έως 58,24%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ''AB-RN'', ''RN'' με 53,33%

Δοκιμή υψηλότερης τιμής ακρίβειας: ''IW-DL-RN'' με 58,24%

Κατά τη μοντελοποίηση των δεδομένων με βάση τον δείκτη 10% required με χρήση του Cost Sensitive Classifier, τα αποτελέσματα δείχνουν ότι ο αλγόριθμος μείωσε την ακρίβεια των περισσότερων δοκιμών και ανέδειξε την μέτρια απόδοση των αλγορίθμων του WEKA στα

μοντέλα που βασίζονται στον δείκτη 10% required. Με τη χρήση του Cost Sensitive ταξινομητή , η ακρίβεια των δοκιμών παρέμεινε σε σχετικά ικανοποιητικά επίπεδα άνω του 50%, αλλά σημειώθηκε η χαμηλότερη τιμή ακρίβειας δοκιμής από όλα τα μελετόμενα μοντέλα, με τις δοκιμές ''AB-RN'' και ''RN'' να δίνουν ακρίβεια 53,33%. Παρατηρώντας και τα προηγούμενα μοντέλα, οι δοκιμές όπου εισάγονται ένα ή δύο υποκειμενικοί δείκτες ως δεδομένα εισόδου, γενικά εμφανίζουν τα χαμηλότερα ποσοστά ακρίβειας σε σχέση με τις δοκιμές περισσότερων υποκειμενικών δεικτών στα δεδομένα εισόδου, ωστόσο στο συγκεκριμένο μοντέλο καταγράφονται νέα χαμηλά ποσοστά. Επίσης, καταγράφεται και το χαμηλότερο ποσοστό ακρίβειας για την δοκιμή ''IW-AB-RN'' με 53,78%.

Οι στατιστικοί δείκτες Precision, Recall, F-Measure και ROC Area, βρίσκονται σε σχετικά αποδεκτά επίπεδα για το σταθμισμένο μέσο όρο των κλάσεων άνω του 0,5, με εύρος από 0,522 έως 0,589 . Συγκριτικά με τα αποτελέσματα των στατιστικών δεικτών για το μοντέλο του J48, των οποίων το εύρος κυμαίνεται 0,534 έως 0,573, συμπεραίνεται ότι ο Cost Sensitive δεν προσφέρει κάποια σαφή βελτίωση στους δείκτες αυτούς.

Τέλος, τα confusion matrix παρουσιάζουν μέτρια αποτελέσματα με έντονη ανισορροπία μεταξύ των σωστών προβλέψεων κάθε κλάσης για την πλειονότητα των δοκιμών. Η μοναδική δοκιμή με ισορροπημένα ποσοστά είναι η ''IW-DL-RN'' με 51% σωστές θετικές και 65% σωστές αρνητικές προβλέψεις, όπου είναι και η δοκιμή με τη μεγαλύτερη ακρίβεια μεταξύ των υπολοίπων. Όσον αφορά τις υπόλοιπες δοκιμές, ο αλγόριθμός Cost Sensitive προκάλεσε εκτεταμένη ανισορροπία, με την πιο ομαλή ''IW-AB-DL-RN'' εκ των δοκιμών, να σημειώνει, 48% σωστές θετικές και 66% σωστές αρνητικές προβλέψεις, και την πιο ανισόρροπη ''IW-HP-RN'' , 37% σωστές θετικές και 70% σωστές αρνητικές.

- **Δείκτης DCEN (Πίνακας 18)**

Εύρος ακρίβειας : 54,91% έως 61,04%

Δοκιμή χαμηλότερης τιμής ακρίβειας: ''AB-RN'' με 54,91%

Δοκιμή υψηλότερης τιμής ακρίβειας: ''IW-HP-RN'' με 61,04%

Η μοντελοποίηση των δεδομένων με βάση τον αντικειμενικό δείκτη DCEN παραδίδει και πάλι ικανοποιητικά αποτελέσματα με συγκεκριμένες δοκιμές να δίνουν υψηλότερες τιμές σε σχέση με τις ίδιες δοκιμές υπό την μοντελοποίηση άλλων αλγορίθμων. Ωστόσο, ο Cost Sensitive αδυνατεί να προσφέρει αξιόλογη βελτίωση σε οποιοδήποτε στατιστικό αποτέλεσμα, αντίθετα, μειώνει την έως τώρα υψηλή απόδοση του μοντέλου DCEN. Ειδικότερα, το εύρος ακρίβειας κυμαίνεται από 54,91% έως 61,04%, ενώ για το μοντέλο χρήσης του J48 χωρίς ενίσχυση του εύρος κυμαίνεται από 56,95% έως 61,04%, υποδεικνύοντας την μείωση του ποσοστού ακρίβειας ορισμένων δοκιμών, με χαρακτηριστικό παράδειγμα τη δοκιμή ''IW-AB-DL-HP-RN'' όπου καταγράφηκε μείωση ακρίβειας κατά 5,62 ποσοστιαίες μονάδες, από 61,03% σε 55,41%. Επίσης αξίζει να σημειωθεί η μείωση που επήλθε στην ακρίβεια της δοκιμής ''AB-RN'', η οποία επέφερε

την μικρότερή τιμή ακρίβειας που σημειώνεται από τις δοκιμές με βάση τον δείκτη DCEN για όλους του αλγόριθμους που χρησιμοποιήθηκαν. Η μείωση αυτή ανήλθε σε 2,04 ποσοστιαίες μονάδες, από 56,95% σε 54,91%.

Όσον αφορά τους στατιστικούς δείκτες Precision, Recall, F-Measure και ROC Area βρίσκονται σε ικανοποιητικά επίπεδα για τον σταθμισμένο μέσο όρο άνω του 0,5, με το εύρος να κυμαίνεται από 0,549 έως 0,620. Οι δοκιμές ''IW-DL-RN'', ''IW-AB-RN'', ''IW-HP-RN'' και ''IW-RN'' παρουσιάζουν τιμές κοντά ή και υψηλότερες του 0,6 (60%) υποδηλώνοντας την σχετικά καλή ακρίβεια τους. Ωστόσο, αν και τα αποτελέσματα των δεικτών είναι ικανοποιητικά, είναι ελαφρώς μειωμένα σε σχέση με τα αποτελέσματα των στατιστικών δεικτών για τα μοντέλα χρήσης του J48 χωρίς ενίσχυση, όπου το εύρος των τιμών για τον σταθμισμένο μέσο όρο των κλάσεων κυμαίνεται από 0,599 έως 0,621. Συνεπώς, είναι ξεκάθαρο ότι ο Cost Sensitive δεν προσέφερε γενικευμένη βελτίωση των μοντέλων, παρά μόνο σε συγκεκριμένες περιπτώσεις.

Τέλος, τα confusion matrix των δοκιμών παρουσιάζουν σχετικά ικανοποιητικά αποτελέσματα, με την πλειονότητα των δοκιμών να χαρακτηρίζονται από αξιόλογη ισορροπία σωστών προβλέψεων μεταξύ των κλάσεων. Μοναδική εξαίρεση αποτελεί η δοκιμή ''RN'' όπου παρατηρείται ανισορροπία, με το ποσοστό σωστών θετικών προβλέψεων να φτάνει το 43% και των σωστών αρνητικών το 72%. Αξιοσημείωτη είναι και η δοκιμή ''IW-AB-RN'', η οποία αποτελεί τη μοναδική εκ των δοκιμών που πραγματοποιήθηκαν με όλους τους αλγόριθμους με ποσοστό σωστών προβλέψεων 60% και για τις δύο κλάσεις (θετικές και αρνητικές). Αν μεταφραστούν αυτά τα ποσοστά σε πραγματικό επίπεδο, γίνεται κατανοητό ότι η μοντελοποίηση με τους τρεις υποκειμενικούς δείκτες ''IW- Inability to heat home'', ''AB- Arrears in bills'', ''RN-Restriction of other needs'', ταξινομεί σωστά το 60% των ενεργειακά φτωχών και μη ενεργειακά φτωχών νοικοκυριών. Όπως έχει προαναφερθεί, επίτευξη αυτού του ποσοστού σε μελέτες κοινωνικών φαινομένων είναι αρκετά ικανοποιητικές και εξαιρετικά χρήσιμες για αναγνώριση και ανάλυση των φαινομένων, όπως η ενεργειακή φτώχεια. Συμπληρωματικά, ας αναφερθεί ότι συγκριτικά με τα confusion matrix που παρουσιάζει ο J48 χωρίς ενίσχυση, τα αποτελέσματα είναι σχετικά όμοια, με τον Cost-Sensitive στις περισσότερες δοκιμές να εξομαλύνει περαιτέρω την ισορροπία των σωστών προβλέψεων σε μικρό βαθμό.

5.2. Συγκριτική αξιολόγηση αποτελεσμάτων

Προκειμένου να πραγματοποιηθεί μία συνολική αξιολόγηση της απόδοσης των εξεταζόμενων αλγορίθμων, λαμβάνονται υπ' όψιν όλοι οι στατιστικοί δείκτες που εξήχθησαν. Στον Πίνακα 19 παρουσιάζονται η ακρίβεια και οι μέσοι όροι των ειδικών στατιστικών δεικτών για κάθε αλγόριθμο, ανά εξεταζόμενο αντικειμενικό δείκτη.

Πίνακας 19: Μέσοι όροι στατιστικών αποτελεσμάτων αλγορίθμων για κάθε αντικειμενικό δείκτη

Μέσος όρος στατιστικών δεικτών-Naïve Bayes					
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Accuracy**
10% actual	0,573	0,567	0,561	0,572	56,69%
10% required	0,569	0,565	0,558	0,578	56,44%
DCEN	0,595	0,590	0,574	0,607	59,00%
Total Averages	0,579	0,574	0,564	0,586	57,37%
Μέσος όρος στατιστικών δεικτών-IBk					
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Accuracy**
10% actual	0,571	0,568	0,564	0,575	56,77%
10% required	0,560	0,555	0,539	0,577	55,45%
DCEN	0,592	0,587	0,582	0,608	58,69%
Total Averages	0,575	0,570	0,562	0,587	56,97%
Μέσος όρος στατιστικών δεικτών- Decision Table					
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Accuracy**
10% actual	0,573	0,569	0,564	0,572	56,87%
10% required	0,561	0,556	0,545	0,576	55,6%
DCEN	0,594	0,591	0,586	0,607	59,04%
Total Averages	0,576	0,572	0,565	0,585	57,16%
Μέσος όρος στατιστικών δεικτών-J48					
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Accuracy**
10% actual	0,553	0,553	0,549	0,554	55,35%
10% required	0,564	0,556	0,542	0,569	55,57%
DCEN	0,599	0,596	0,591	0,602	59,55%
Total Averages	0,572	0,568	0,561	0,575	56,82%
Μέσος όρος στατιστικών δεικτών-AdaBoost.M1					
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Accuracy**
10% actual	0,571	0,568	0,565	0,574	56,7%
10% required	0,562	0,555	0,543	0,575	55,50%
DCEN	0,587	0,585	0,582	0,608	58,51%
Total Averages	0,573	0,569	0,563	0,586	56,89%
Μέσος όρος στατιστικών δεικτών-Cost Sensitive					
	Precision	**Recall**	**F-Measure**	**ROC Area**	**Accuracy**
10% actual	0,571	0,568	0,566	0,566	56,80%
10% required	0,557	0,548	0,542	0,557	54,81%
DCEN	0,577	0,573	0,567	0,584	57,97%
Total Averages	0,568	0,563	0,558	0,569	56,53%

Αρχικά, όσον αφορά στην ακρίβεια (Accuracy) των αλγορίθμων κατά τις μοντελοποιήσεις του δείκτη 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες), ο αλγόριθμος Cost Sensitive παρουσίασε την υψηλότερη ακρίβεια προβλέψεων με 56,80%, ενώ ο αλγόριθμος J48 την χαμηλότερη με 55,35%. Για τη μοντελοποίηση του δείκτη 10% required (με βάση το απαιτούμενο κόστος), το υψηλότερο ποσοστό ακρίβειας επετεύχθη από τον αλγόριθμο Naïve Bayes με 56,44%, ενώ το χαμηλότερο από τον Cost-Sensitive με 54,81%. Τέλος, της μοντελοποιήσεις βάσει του δείκτη συμπίεσης ενεργειακών αναγκών DCEN, η υψηλότερη ακρίβεια επετεύχθη από τον αλγόριθμο J48 με ποσοστό 59,55% και η χαμηλότερη από τον Cost Sensitive με 57,97%.

Όπως αναφέρθηκε και στο Κεφάλαιο 3.3.4., οι επιδόσεις των στατιστικών δεικτών Precision, Recall, F-Measure συνάδουν με τις επιδόσεις των ποσοστών ακριβείας κάθε αλγορίθμου, συνεπώς η κατάταξη των αλγορίθμων με τις υψηλότερες και χαμηλότερες τιμές επί των ειδικών στατιστικών δεικτών ταυτίζεται με αυτή που πραγματοποιήθηκε στην προηγούμενη παράγραφο για την ακρίβεια των αλγορίθμων.

Ειδική αναφορά γίνεται στις τιμές που παρουσίασε ο ειδικός στατιστικός δείκτης ROC Area στα μοντέλα των αλγορίθμων, καθώς βάσει αυτού γίνεται δυνατή η αξιολόγηση της ικανότητας του αλγορίθμου να ταξινομεί σωστά τα δεδομένα ανά κλάση και να μην πραγματοποιεί τυχαία ταξινόμηση. Πιο συγκεκριμένα, όσον αφορά τα μοντέλα για τον δείκτη 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες), υψηλότερη τιμή σημείωσε ο αλγόριθμος IBk με ROC = 0,572 και χαμηλότερη ο J48 με ROC = 0,554. Για τον δείκτη 10% required (με βάση το απαιτούμενο κόστος), η υψηλότερη τιμή επετεύχθη από τον Naïve Bayes με ROC = 0,578 και η χαμηλότερη από τον Cost Sensitive με ROC = 0,557. Τέλος, στα μοντέλα του δείκτη DCEN, υψηλότερη τιμή επετεύχθη από τους αλγορίθμους IBk και AdaBoost.M1 με ROC = 0,608, ενώ η χαμηλότερη από τον Cost Sensitive με ROC = 0,584.

Μέσω του Πιν.19 γίνεται κατανοητό ότι κανένας από τους δοκιμασμένους αλγόριθμους δεν λειτούργησε το ίδιο ικανοποιητικά για όλα τα μοντέλα. Χαρακτηριστικό παράδειγμα αποτελεί ο αλγόριθμος Cost Sensitive, ο οποίος αποτέλεσε τον αλγόριθμο με τη μεγαλύτερη ακρίβεια για τα μοντέλα του δείκτη 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες), αλλά και τον αλγόριθμο με τη χαμηλότερη ακρίβεια για τα μοντέλα του δείκτη DCEN.

Συνοψίζοντας για την αποτελεσματικότητα μοντελοποίησης των αλγορίθμων, όλοι απέδωσαν σχετικά ικανοποιητικά αποτελέσματα μετά από κατάλληλη ρύθμισή τους πριν τη μοντελοποίηση, πέραν συγκεκριμένων δοκιμών που υπολειτουργούσαν ελαφρώς. Οι χρησιμοποιούμενοι απλοί αλγόριθμοι (Naïve Bayes, IBk, Decision Table, J48), παρουσίασαν παρόμοια στατιστικά αποτελέσματα με το μοντέλο που αναλύθηκε στη σχετική μελέτη των Papada & Kaliampakos (2022) με χρήση νευρωνικών δικτύων για τον δείκτη 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες), απ' όπου συμπεραίνεται ότι το όριο εύρους ακριβείας των μοντέλων του συγκεκριμένου δείκτη κυμαίνεται μεταξύ του 55% και του 57%, όταν χρησιμοποιούνται οι συγκεκριμένοι αλγόριθμοι μηχανικής μάθησης. Είναι λοιπόν πιθανό, και το μέγιστο όριο ακρίβειας που μπορούν να επιτύχουν οι αλγόριθμοι μηχανικής μάθησης κατά την επεξεργασία του συγκεκριμένου συνόλου δεδομένων για την ενεργειακή φτώχεια (εξέταση συνδυασμού αντικειμενικών και υποκειμενικών δεικτών) να κυμαίνεται σχετικά κοντά σε αυτό το εύρος, ανεξάρτητα από τον χρησιμοποιούμενο αλγόριθμο. Τα αποτελέσματα πρόβλεψης του δείκτη 10% required (με βάση το απαιτούμενο κόστος) παρουσιάζονται για πρώτη φορά στην

παρούσα εργασία και χαρακτηρίζονται ως οριακά ικανοποιητικά, καθώς είναι ελαφρώς κατώτερα από τα αποτελέσματα μοντέλων των υπολοίπων αντικειμενικών δεικτών. Επίσης, πραγματοποιήθηκε προσέγγιση της απόδοσης του αντικειμενικού δείκτη DCEN σε τέτοιου είδους μεθόδους ταξινόμησης. Ο δείκτης DCEN (Degree of Coverage of Energy Needs) παρουσίασε τα πιο ενθαρρυντικά αποτελέσματα, καθώς ο μέσος όρος ακρίβειας για όλους τους αλγόριθμους κυμάνθηκε κοντά στο 60%, ξεχωρίζοντας τον ως τον καταλληλότερο αντικειμενικό δείκτη μεταξύ των μελετώμενων για την διεξαγωγή της συγκεκριμένης μοντελοποίησης στη βάση υποκειμενικών δεικτών. Η μοντελοποίηση με βάση τον συγκεκριμένο δείκτη στην παρούσα εργασία εμφάνισε βελτιωμένα στατιστικά αποτελέσματα σε σχέση με όμοια μοντελοποίηση που πραγματοποιήθηκε από τους Papada & Kaliampakos (2022), στην οποία το ποσοστό ακρίβειας των μοντέλων πρόβλεψης βρέθηκε κατά μέσο όρο στο 55% με χρήση νευρωνικών δικτύων, ενώ στην παρούσα εργασία ο μέσος όρος ακρίβειας μοντέλων του DCEN σταθμίστηκε στο 58,8%.

Συμπληρωματική αναφορά γίνεται και στα αποτελέσματα χρήσης αλγόριθμων meta-learning (AdaBoost.M1, Cost Sensitive Classifier). Δεν παρατηρήθηκε αισθητή βελτίωση σε κάποιο μοντέλο αναφορικά με την συνολική του απόδοση, εκτός μεμονωμένων περιπτώσεων δοκιμών. Αντιθέτως, προκάλεσαν μείωση της ακρίβειας σε αρκετές δοκιμές. Γι’ αυτό το λόγο, είναι απαραίτητο να σημειωθεί ότι τέτοιου είδους αλγόριθμοι δεν εγγυώνται την ενίσχυση του χρησιμοποιούμενου αλγορίθμου, είτε λόγω της μορφής των δεδομένων, είτε λόγω της βελτιστοποίησης που ήδη έχει επέλθει στον αλγόριθμο. Ωστόσο, υπάρχουν πολλοί διαθέσιμοι meta learning αλγόριθμοι οι οποίοι πιθανώς μπορούν να βελτιώσουν την ικανότητα ταξινόμησης τον αλγορίθμων σε μεγαλύτερο βαθμό.

Επιπλέον, σημαντικές ενδείξεις δόθηκαν σχετικά με την εγκυρότητα των χρησιμοποιούμενων υποκειμενικών δεικτών καθώς, μέσω των 144 πραγματοποιημένων δοκιμών, ορισμένοι δείκτες αναδείχθηκαν ως πιο έγκυροι και συμβατοί σε σχέση με άλλους. Πιο συγκεκριμένα, ο δείκτης ‘’HP-Health Problems’’ υπήρχε σε όλες τις δοκιμές που παρουσίασαν υψηλότερη ακρίβεια. Αποτέλεσε, δηλαδή, έναν αρκετά έγκυρο δείκτη για την ταξινόμηση νοικοκυριών ως ενεργειακά φτωχά. Συγκεκριμένα, όσα σπίτια αντιμετώπιζαν προβλήματα υγείας, είχαν και υψηλότερη πιθανότητα να ταξινομηθούν έγκυρα ως ενεργειακά φτωχά. Είναι λοιπόν απαραίτητο να αναδειχθεί η σημασία του συγκεκριμένου δείκτη με σκοπό τη χρήση του και σε επόμενες μοντελοποιήσεις. Επιπλέον, η δοκιμή διαφορετικών συνδυασμών υποκειμενικών δεικτών ανέδειξε τη χρήση συνδυασμού τριών υποκειμενικών δεικτών ως την πιο αποδοτική συγκριτικά με δοκιμές όπου χρησιμοποιήθηκαν περισσότεροι ή λιγότεροι υποκειμενικοί δείκτες. Τέλος, ο υποκειμενικός δείκτης ‘’RN-Restriction of other needs’’, αναδείχθηκε ως ο λιγότερο συμβατός υποκειμενικός δείκτης, καθώς αποτέλεσε τον κύριο λόγο μείωσης της ακρίβειας των δοκιμών. Συμπληρωματικά, οι δοκιμές που περιλάμβαναν μόνο το συγκεκριμένο δείκτη ως δεδομένα εισόδου (input data) παρουσίασαν, στην πλειονότητά τους, χαμηλότερη ακρίβεια σε σχέση με τις υπόλοιπες δοκιμές.

Κεφάλαιο 6. Συμπεράσματα

Η ενεργειακή φτώχεια εξακολουθεί να αποτελεί ένα από τα πλέον ουσιώδη οικονομικά και κοινωνικά ζητήματα της σύγχρονης εποχής, το οποίο επηρεάζει τόσο τον αναπτυσσόμενο, όσο και τον ανεπτυγμένο κόσμο με διαφορετικά εμφανιζόμενα χαρακτηριστικά ανά περιοχή. Ενώ λοιπόν, η παγκόσμια παραγωγή και κατανάλωση ενέργειας σημειώνει νέα υψηλά ρεκόρ κάθε έτος, η ενεργειακή φτώχεια συνεχίζει να μαστίζει τις φτωχότερες χώρες με εκτεταμένη έλλειψη πρόσβασης σε υπηρεσίες ενέργειας, καθώς και έλλειψη ασφαλούς και σταθερής παροχή ενέργειας, εγκαταλείποντας εκατοντάδες εκατομμύρια ανθρώπων χωρίς βασικές παροχές άνετης διαβίωσης. Ταυτόχρονα, οι ανεπτυγμένες κοινωνίες έχουν έρθει αντιμέτωπες με τις συνέπειες του επικίνδυνου συνδυασμού υψηλών τιμών ενέργειας και χαμηλότερων εισοδημάτων ανά νοικοκυριό, που εντείνεται επικίνδυνα τα τελευταία χρόνια.

Η υπάρχουσα οικονομική κρίση σε συνδυασμό με τις αλλεπάλληλες γεωπολιτικές συγκρούσεις, έχουν προκαλέσει απότομη άνοδο στις τιμές της ενέργειας σε όλες τις μορφές της, επιφέροντας σημαντικές δυσκολίες στις συνθήκες διαβίωσης των οικονομικά αδύναμων νοικοκυριών. Η Ελλάδα είναι μία από τις χώρες με τις ακριβότερες τιμές ενέργειας σε ευρωπαϊκό επίπεδο, ενώ ταυτόχρονα είναι εξαιρετικά ευάλωτη εξαιτίας της οικονομικής κρίσης που εξακολουθεί να επικρατεί, αλλά και των πολιτικών λιτότητας που εφαρμόζονται για πάνω από μία δεκαετία. Οι συνθήκες αυτές εκθέτουν σε κίνδυνο πολλές πληθυσμιακές ομάδες, καθιστώντας ολοένα και περισσότερα νοικοκυριά «ενεργειακά φτωχά», καθώς αυτά αδυνατούν να ανταπεξέλθουν στις υψηλές τιμές ενέργειας. Μία εκ των ιδιαίτερα ευάλωτων ομάδων είναι οι κάτοικοι των ορεινών περιοχών, λόγω συγκεκριμένων εγγενών αλλά και κοινωνικοοικονομικών χαρακτηριστικών των περιοχών αυτών, όπως το υψόμετρο, η γεωγραφική απομόνωση, η ασθενής παραγωγική βάση κλπ..

Στην παρούσα εργασία επιχειρήθηκε η πρόβλεψη της ενεργειακής φτώχειας με τη χρήση αλγορίθμων μηχανικής μάθησης. Ειδικότερα, πραγματοποιήθηκε η πρόβλεψη τριών αντικειμενικών δεικτών ενεργειακής φτώχειας βάσει πέντε υποκειμενικών δεικτών ενεργειακής φτώχειας. Η συγκεκριμένη μελέτη επιχειρεί να ενισχύσει περαιτέρω τα σχετικά ευρήματα πρόσφατης έρευνας (Papada & Kaliampakos, 2022), αναφορικά με τη σχέση μεταξύ των αντικειμενικών και υποκειμενικών δεικτών της ενεργειακής φτώχειας με τη χρήση των εργαλείων μηχανικής μάθησης, ενώ ταυτόχρονα συνεισφέρει στην εξαγωγή συμπερασμάτων για την ικανότητα των χρησιμοποιούμενων αλγορίθμων μηχανικής μάθησης να δημιουργήσουν έγκυρα μοντέλα ενεργειακής φτώχειας με αποδεκτή ακρίβεια υπολογισμών.

Οι αλγόριθμοι μηχανικής μάθησης αποτελούν εργαλεία με μεγάλο εύρος εφαρμογών σε ένα πλήθος επιστημονικών πεδίων, καθώς βασικά τους χαρακτηριστικά αποτελούν η εκπαίδευση, η επεξεργασία μεγάλου όγκου δεδομένων και η δυνατότητά τους να παραθέτουν στο χειριστή τους σημαντικά στατιστικά στοιχεία και μέτρα ακρίβειας. Για την εκπαίδευση και τη χρήση πολλαπλών αλγορίθμων μηχανικής μάθησης επιλέχθηκε το πρόγραμμα WEKA, μέσω του οποίου πραγματοποιήθηκε η μοντελοποίηση των δεδομένων και εξήχθησαν σημαντικοί στατιστικοί δείκτες για την αξιολόγηση των μοντέλων.

Παρακάτω, παρουσιάζονται τα κύρια αποτελέσματα που εξήχθησαν από τη χρήση των αλγορίθμων μηχανικής μάθησης και τις 144 δοκιμές που πραγματοποιήθηκαν για διάφορους συνδυασμούς υποκειμενικών και αντικειμενικών δεικτών:

1. **Naive Bayes:** Σε γενική βάση, ο συγκεκριμένες αλγόριθμος λειτούργησε ικανοποιητικά, με κάποιες περιπτώσεις να σημειώνουν αρκετά καλύτερα αποτελέσματα σε σχέση με τους υπόλοιπους αλγόριθμους διατηρώντας σταθερότητα μεταξύ των δοκιμών που πραγματοποίησε χωρίς έντονες διακυμάνσεις. Το ποσοστό ακρίβειας κυμάνθηκε από 55% έως 61%, διατηρώντας σχετικά ικανοποιητικά ποσοστά σωστών προβλέψεων ανά κλάση. Οι δοκιμές που πραγματοποιήθηκαν για να προβλέψουν την ενεργειακή φτώχεια σύμφωνα με τον δείκτη DCEN εμφάνισαν την μεγαλύτερη ακρίβεια, με τις δοκιμές τις δείκτη 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες) και τις δοκιμές του δείκτη 10% required (με βάση το απαιτούμενο κόστος) να ακολουθούν με ελαφρώς χειρότερα αποτελέσματα. Το πιο αξιόπιστο αποτέλεσμα του αλγορίθμου αυτού αποτελεί η δοκιμή που περιλάμβανε τους υποκειμενικούς δείκτες ''IW-Inability to heat home'', ''HP-Health problems'', '' RN- Restrictions of other needs'' με βάση τον δείκτη DCEN, όπου επετεύχθη ακρίβεια 61,04%, η οποία μεταφράζεται σε πραγματικό επίπεδο ως εξής: Τα νοικοκυριά τα οποία «πάσχουν» ταυτόχρονα από τους τρεις παραπάνω υποκειμενικούς δείκτες, είναι και ενεργειακά φτωχά βάσει του δείκτη συμπίεσης DCEN με ακρίβεια πρόβλεψης 61,04%.

2. **IBk:** Ο αλγόριθμος δημιούργησε μοντέλα με σχετικά ικανοποιητική ακρίβεια με εύρος ακρίβειας από 54,28% έως 61,04%, ελαφρά μειωμένη σε σχέση με τον Naive Bayes, αλλά διατηρώντας μια σχετική σταθερότητα μεταξύ των μοντέλων ανά αντικειμενικό δείκτη. Οι δοκιμές του δείκτη DCEN εμφάνισαν και σε αυτή τη περίπτωση τα υψηλότερα ποσοστά ακρίβειας καθώς κυμάνθηκαν από 56,95% έως 61,04%. Η σειρά μεγαλύτερης ακρίβειας μεταξύ των μοντέλων κάθε αντικειμενικού δείκτη κατατάσσει τον δείκτη DCEN ως τον πιο ακριβή, με τον δείκτη 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες) να ακολουθεί και τον 10% required (με βάση το απαιτούμενο κόστος) να εμφανίζει κατά μέσο όρο την χαμηλότερη ακρίβεια προβλέψεων. Και σε αυτήν την περίπτωση η δοκιμή ''IW-HP-RN'' σημείωσε την υψηλότερη ακρίβεια μεταξύ των δοκιμών με το 61,04%.

3. **Decision Table:** Ο αλγόριθμος αυτός παρουσιάζει εξίσου ικανοποιητικά αποτελέσματα με τον IBk, όπου το εύρος ακρίβειας κυμαίνεται από 54,28% έως 61,04%. Τα εύρη των μοντέλων για κάθε αντικειμενικό δείκτη είναι σχετικά ικανοποιητικά και σταθερά σε συγκεκριμένες τιμές. Οι τιμές για τα μοντέλα του κάθε αντικειμενικού δείκτη παραμένουν σε παρόμοια επίπεδα με τον IBk και τον Naïve Bayes όσον αφορά τους μέσους όρους ακρίβειας, με ανεπαίσθητες διακυμάνσεις. Και σε αυτόν τον αλγόριθμο η δοκιμή ''IW-HP-RN'' του DCEN σημειώνει την υψηλότερη ακρίβεια μεταξύ των δοκιμών με 61,04%, ενισχύοντας την το συλλογισμό κατά τον οποίο, τα νοικοκυριά τα οποία αντιμετωπίζουν, ταυτόχρονα, τα προβλήματα που αντιπροσωπεύουν οι τρεις υποκειμενικοί δείκτες είναι ενεργειακά φτωχά βάσει του δείκτη συμπίεσης DCEN, με ακρίβεια πρόβλεψης 61,04%.

4. **J48:** Χαμηλότερα ποσοστά ακρίβειας παρουσιάζονται για τις δοκιμές των δεικτών 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες) και 10% required (με βάση το απαιτούμενο κόστος) όπου το εύρος ακρίβειας των δύο μοντέλων κυμαίνεται από 54,28% έως 56,24%, με τις περισσότερες δοκιμές να μην ξεπερνούν το 56%. Τα μοντέλα παραμένουν σχετικά ικανοποιητικά, αλλά κατατάσσονται ως τα χαμηλότερης ακρίβειας μοντέλα σε σχέση με τους υπόλοιπους χρησιμοποιούμενους αλγορίθμους, πέραν των αλγορίθμων ενίσχυσης που σχολιάζονται παρακάτω. Το μοντέλο του DCEN σημειώνει και πάλι υψηλή ακρίβεια με το μέσο όρο των δοκιμών να βρίσκεται άνω του 60%, με τις δοκιμές ''IW-AB-DL-HP-RN'', ''IW-HP-RN'' να επιτυγχάνουν ακρίβεια 61,03% και 61,04% αντίστοιχα. Ανάγοντας και πάλι τις δοκιμές υψηλότερης ακρίβειας σε συμπεράσματα που αφορούν την πραγματικότητα, αποδεικνύεται η υψηλή ακρίβεια πρόβλεψης ενεργειακά φτωχών νοικοκυριών που ''πάσχουν'', ταυτόχρονα, από τα φαινόμενα τα οποία αντιπροσωπεύουν οι δείκτες ''IW'',''HP'' και ''RN'', πάντα με βάση τον αντικειμενικό δείκτη DCEN.

5. **AdaBoost.M1-Decision Table:** Ο συγκεκριμένος αλγόριθμος ενίσχυσης μορφής «boosting», δεν προσφέρει κάποια ουσιαστική ενίσχυση στα στατιστικά αποτελέσματα του Decision Table, καθώς τα εύρη ακρίβειας των δοκιμών παραμένουν όμοια με το μοντέλο χρήσης του Decision Table χωρίς χρήση αλγορίθμου ενίσχυσης. Αντίθετα, σε συγκεκριμένες δοκιμές η επίδραση του αλγορίθμου στα αποτελέσματα είναι αρνητική, καθώς μειώνει την ακρίβεια τους. Το μοντέλο του DCEN, ξεχωρίζει και πάλι για τις υψηλότερες επιδόσεις που καταγράφει σε σχέση με τα μοντέλα των υπόλοιπων αντικειμενικών δεικτών, με τη δοκιμή ''IW-HP-RN'' να δίνει ακρίβεια 61,04%, επιβεβαιώνοντας την ικανοποιητική απόδοσή του με τη χρήση όλων των αλγορίθμων και ενισχύοντας περαιτέρω το συμπέρασμα που εξήχθη από τη χρήση των αλγορίθμων για την δοκιμή υψηλής ακρίβειας ''IW-HP-RN''.

6. **Cost Sensitive Classifier-J48:** Σε αυτή τη μορφή αλγορίθμου, η «μήτρα κόστους» (cost matrix) που διαθέτει ο συγκεκριμένος αλγόριθμος ενίσχυσης προσφέρει βελτιώσεις στα στατιστικά αποτελέσματα συγκεκριμένων δοκιμών, ενώ επιδρά αρνητικά σε κάποιες άλλες. Πιο συγκεκριμένα, το μοντέλο του δείκτη 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες) επιδέχεται κάποιες βελτιώσεις καθώς, ο μέσος όρος ακρίβειας των δοκιμών αυξάνεται από 55,64% σε 56,79% (αύξηση 1,16 ποσοστιαίων μονάδων). Αντίθετα, στο μοντέλο του δείκτη 10% required (με βάση το απαιτούμενο κόστος), ο μέσος όρος ακρίβειας των δοκιμών μειώνεται με τη χρήση του αλγορίθμου ενίσχυσης από 55,57% σε 54,81% (μείωση 0,75 ποσοστιαίων μονάδων). Όσον αφορά το πιο αποδοτικό μοντέλο του δείκτη DCEN, καταγράφεται και σε αυτή τη περίπτωση μείωση του μέσου όρου ακριβείας, με ορισμένες δοκιμές να καταγράφουν μείωση ακρίβειας άνω των 5 ποσοστιαίων μονάδων, με χαρακτηριστικό παράδειγμα τη δοκιμή ''IW-AB-DL-HP-RN''όπου η ακρίβεια μειώνεται από 61,03% σε 55,41%. Συνεπώς, και

αυτός ο αλγόριθμος ενίσχυσης κρίνεται γενικά ακατάλληλος για τα ανεπτυγμένα μοντέλα, με εξαίρεση συγκεκριμένες δοκιμές ανά μοντέλο

Στο πλαίσιο αυτό και, στη βάση μιας συνολικής συγκριτικής αξιολόγησης και, λαμβάνοντας υπ' όψιν τόσο τα ποσοστά ακρίβειας όσο και το σύνολο των ειδικών στατιστικών δεικτών (Precision, Recall, F-Measure, ROC Area), ο δείκτης 10% actual (με βάση τις πραγματοποιηθείσες δαπάνες) προσεγγίστηκε καλύτερα από τον αλγόριθμο Decision Table, ο δείκτης 10% required (με βάση το απαιτούμενο κόστος) προσεγγίστηκε καλύτερα από τον αλγόριθμο Naïve Bayes , ενώ ο δείκτης DCEN πέτυχε τα καλύτερα αποτελέσματα με τον αλγόριθμο J48.

Τέλος, αν διερευνηθεί η συνολική μέση απόδοση κάθε αλγορίθμου στην πρόβλεψη και των τριών δεικτών ενεργειακής φτώχειας που εξετάστηκαν (10% actual, 10% required, DCEN), ο αλγόριθμός Naïve Bayes λειτούργησε συνολικά καλύτερα από τους υπόλοιπους αλγόριθμους, ενώ ο Cost Sensitive Classifier παρουσίασε την χαμηλότερη συνολική ακρίβεια μεταξύ των μοντέλων κάθε δείκτη. Αυτή η σύγκριση δίνει μια γενικότερη εικόνα της απόδοσης των αλγορίθμων στην πρόβλεψη του κινδύνου ενεργειακής φτώχειας. Εντούτοις, γίνεται φανερό ότι η επιλογή του καταλληλότερου αλγόριθμου απαιτεί πολλές δοκιμές, ξεχωριστή ανάλυση και προσαρμογή στα εκάστοτε δεδομένα.

Η χρήση των αλγορίθμων μηχανικής μάθησης στη συγκεκριμένη εργασία επέκτεινε σημαντικά την υφιστάμενη γνώση στον τομέα της μελέτης του φαινομένου της ενεργειακής φτώχειας με τη χρήση τεχνητής νοημοσύνης. Ειδικότερα, η παρούσα έρευνα «πρόσθεσε γνώση» στο κομμάτι της πρόβλεψης του κινδύνου ενεργειακής φτώχειας, όπως αυτή ορίζεται μέσω αντικειμενικών δεικτών, στη βάση διαφόρων υποκειμενικών δεικτών του φαινομένου. Με άλλα λόγια, εκτιμήθηκε κατά πόσο υποκειμενικά χαρακτηριστικά και ανθρώπινες συμπεριφορές (πχ. αίσθηση επαρκούς άνεσης στο σπίτι, έγκαιρη αποπληρωμή λογαριασμών ενέργειας κλπ.) μπορούν να προβλέψουν με ακρίβεια την ενεργειακή φτώχεια βάσει αντικειμενικών δεδομένων, πχ. δεδομένα κόστους και εισοδήματος.

Οι δυνατότητες των σύγχρονων αλγορίθμων μηχανικής μάθησης, πλέον, ξεπερνούν τις συμβατικές μεθόδους μοντελοποίησης μεγάλων συνόλων δεδομένων, γεγονός το οποίο ενθαρρύνει τις απόπειρες χρήσης τους σε όλα τα επιστημονικά πεδία, συμπεριλαμβανομένης και της ενεργειακής φτώχειας, με τα πρώτα ευρήματα των σχετικών μελετών να θέτουν τα θεμέλια για την ανάπτυξη ακριβέστερων ερευνών στον συγκεκριμένο τομέα. Η μελλοντική έρευνα στον τομέα της ενεργειακής φτώχειας με τη χρήση αλγορίθμων μηχανικής μάθησης αναμένεται να επεκτείνει περαιτέρω τα σχετικά ευρήματα, εισάγοντας νέες παραμέτρους και δείκτες στα σχηματιζόμενα μοντέλα. Μελλοντικά, η τεχνητή νοημοσύνη αναμένεται να διαδραματίσει καθοριστικό ρόλο στην κατανόηση των χαρακτηριστικών της ενεργειακής φτώχειας και κατά συνέπεια, στην υιοθέτηση αποτελεσματικών μέτρων και στρατηγικών για την αντιμετώπισή της.

Βιβλιογραφία

Atsalis A., Mirasgedis S., Tourkolias C. & Diakoulaki D. (2016). Fuel poverty in Greece: Quantitative analysis and implications for policy. Energy and Buildings, 131, pp. 87–98.

Benardos, A.: Artificial Intelligence in Underground Development: A Study of TBM Performance. Underground Spaces. WIT Transactions on the Built Environment 102, 21-32 (2008).

Berkeley School of Information, Machine Learning, 2020, https://ischoolonline.berkeley.edu/blog/what-is-machine-learning/

Bouzarovski (2014) Bouzarovski S. (2014). Energy poverty in the European Union: landscapes of vulnerability. WIREs Energy and Environment, 3 (3), pp. 276-289.

BP Statistical Review of World Energy, 2022, 71st Edition, https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2022-full-report.pdf

BPIE (Buildings Performance Institute Europe) (2014). Alleviating fuel poverty in the EU. Investing in home renovation, a sustainable and inclusive solution. Brussels: BPIE.

Bradshaw & Hutton 1983, Bradshaw J. & Hutton S. (1983). Social policy options and fuel poverty. Journal of Economic Psychology, 3, pp. 249–266.

Boardman B. (1986). Seasonal mortality and cold homes. Unhealthy Housing: A Diagnosis conference. Proceedings of the University of Warwick's, 14-16 December 1986.

Boardman B. (1991). Fuel Poverty: From Cold Homes to Affordable Warmth. London: Belhaven Pres

Buzar 2007: Buzar S. (2007). Energy Poverty in Eastern Europe: Hidden Geographies of Deprivation. Aldershot, UK: Ashgate.

Carrere, Peralta et.al, 2021.

DECC (Dept for Energy and Climate Change) (2015). Annual Fuel Poverty Statistics Report 2015. London: DECC.

DECC (Dept. for Energy and Climate Change) (2013) Fuel Poverty Report - Updated August 2013. London: DECC.

DECC (Dept. for Energy and Climate Change) (2009). Annual Report on Fuel Poverty Statistics 2009. London: DECC.

DETR (Department of the Environment, Transport and the Regions) (2000). English House Condition Survey 1996: Energy Report. pp. 119–144. London: DETR

Dubois 2012, Dubois U. (2012). From targeting to implementation: the role of identification of fuel poor households. Energy Policy, 49, pp. 107–115

DOE (Department of the Environment) (1996). English House Condition Survey 1991: Energy Report. pp. 129–137. London: DOE.

Eibe Frank, Mark A. Hall, and Ian H. Witten, The WEKA Workbench, Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques" Morgan Kaufmann, Fourth Edition, 2016

EC 2010: EC (2010). An Energy policy for Customers. Commission Staff Working Paper. European Commission, 11.11.2010. Brussels.

EPEE (European Fuel Poverty and Energy Efficiency) (2009a). Tackling fuel poverty in Europe. Recommendations Guide for Policy Makers. Boulogne, France: Dixit Productions.

Fabbri, 2015 Fabbri K. (2015). Building and fuel poverty, an index to measure fuel poverty: An Italian case study. Energy, 89, pp. 244-258.

Fahmy E. (2011). The definition and measurement of fuel poverty. A Briefing Paper to inform Consumer Focus' submission to the Hills fuel poverty review. University of Bristol, UK.

Fahmy E., Gordon D. & Patsios D. (2011). Predicting fuel poverty at a small-area level in England. Energy Policy, 39, pp. 4370–4377.

Funnell D. & Parish R. (2001). Mountain Environments and Communities. London: Routledge

FAO (Food and Agriculture Organization of the United Nations) (2015). Mapping the vulnerability of mountain peoples to food insecurity. Rome: FAO.

González-Eguino M. (2015). Energy poverty: An overview. Renewable and Sustainable Energy Reviews, 47, pp. 377–385.

Healy J. D. (2004). Housing, Fuel Poverty and Health: A Pan-European Analysis. Aldershot, UK: Ashgate.

Healy & Clinch 2002 Healy J. D. & Clinch P. (2002). Fuel poverty in Europe: A cross-country analysis using a new composite measure. Environmental Studies Research Series Working Paper. University College Dublin

Hills J. (2012). Getting the Measure of Fuel Poverty. Final Report of the Fuel Poverty Review. London School of Economics and Political Science. Report 72. London: Centre for Analysis of Social Exclusion

Hills J. (2011). Fuel Poverty: The Problem and Its Measurement, Interim Report of the Fuel Poverty Review. London School of Economics and Political Science. London: Centre for Analysis of Social Exclusion.

Household Energy Price Index for Europe,2022, https://www.energypriceindex.com/price-data

H. Wang, C. Ma and L. Zhou, "A Brief Review of Machine Learning and Its Application," *2009 International Conference on Information Engineering and Computer Science*, 2009, pp. 1-4, doi: 10.1109/ICIECS.2009.5362936

IBM Cloud Education, Machine Learning ,2020, https://www.ibm.com/cloud/learn/machine-learning

IEA (2021), World Energy Outlook 2021, IEA, Paris https://www.iea.org/reports/world-energy-outlook-2021

IEA (2021), What is behind soaring energy prices and what happens next? IEA, Paris https://www.iea.org/commentaries/what-is-behind-soaring-energy-prices-and-what-happens-next

IEA (2007), Energy Economics: A Place for Energy Poverty in the Agenda? IEA, Paris https://www.iea.org/reports/energy-economics-a-place-for-energy-poverty-in-the-agenda

Jacques Delors Institute, Policy Paper 259, Europe needs a political strategy to end energy poverty, (2021)

J. Brownlee, How to Use Classification Machine Learning Algorithms in Weka, Weka Machine Learning, 2016 (Last update 2019)

Carrere J., Peralta A., Oliveras L., López M., Olmo M.,Benach J., Novoa A., Energy poverty, its intensity and health in vulnerable populations in a Southern European city, Gaceta Sanitaria, Volume 35, Issue 5, 2021, Pages 438-444,

Khandker R. S., Samad H. A., Ali R. & Barnes D. F. (2014). Who benefits most from rural electrification? Evidence in India. Energy Journal, 35 (2), pp. 75-96.

Katsoulakos N. (2011). Combating Energy Poverty in Mountainous Areas Through Energy-saving Interventions. Mountain Research and Development, 31 (4), pp. 284-292.

Korner C., Ohsawa M. (coordinating lead authors) et al. (2005). Mountain Systems. Chapter 24. In Ecosystems and human well-being: current state and trends. Volume 1. Washington, DC: Island Press

Lawson et al. ,2015 Lawson R., Williams J. & Wooliscroft B. (2015). Contrasting approaches to fuel poverty in New Zealand. Energy Policy, 81, pp. 38–42.

Legendre B. & Ricci O. (2015). Measuring fuel poverty in France: Which households are the most fuel vulnerable?. Energy Economics, 49, pp. 620–628

Lewis P. (1982). Fuel poverty can be stopped. Bradford: National Right to Fuel Campaign.

Libbrecht, M., Noble, W. Machine learning applications in genetics and genomics. Nat Rev Genet 16, 321–332 (2015)

Liddell C., Morris C., Mckenzie S. J. P. & Rae G. (2012). Measuring and monitoring fuel poverty in the UK: National and regional perspectives. Energy Policy, 49, pp. 27–32.

Liddell C. & Morris C. (2010). Fuel poverty and human health: a review of recent evidence. Energy Policy, 38, pp. 2987–2997.

Marvin 1994 Marvin S. J. (1994). Accessibility to utility networks: Local policy issues. Local Government Studies, 20 (3), pp. 437-457

Menash et al. (2017) Mensah S., Veldtman R., Assogbadjo A. E., Ham C., Kakaï R. G. & Seifert T. (2017). Ecosystem service importance and use vary with socio-environmental factors: A study from household-surveys in local communities of South Africa. Ecosystem Services, 23, pp. 1–8.

Moore R. (2012). Definitions of fuel poverty: implications for policy. Energy Policy, 49, pp. 19–26

Nordregio/ WCMC (2004) NORDREGIO (2004). Mountain Areas in Europe: Analysis of mountain areas in EU member states, acceding and other European countries. Nordic Center for Spatial Development. European Commission contract No 2002, Final report.

M. Chlechowitz ,M. Reuter, ODYSSEE-MURE, Policy Brief, Energy Poverty in the EU, 2021

Hannah Ritchie, Max Roser and Pablo Rosado (2020) - "Energy". Published online at OurWorldInData.org. Retrieved from: 'https://ourworldindata.org/energy' [Online Resource]

Papada L. & Kaliampakos D. (2016). Measuring energy poverty in Greece. Energy Policy, 94, pp. 157-165.

Papada, L., Kaliampakos, D. Energy poverty in Greek mountainous areas: a comparative study. J. Mt. Sci. 14, 1229–1240 (2017). https://doi.org/10.1007/s11629-016-4095-z

Papada L. & Kaliampakos D. (2015). Quantifying energy demand in mountainous areas. Lecture Notes in Artificial Intelligence 9518, Springer, pp. 31–43. DOI: 10.1007/978-3-319-27430-0_3

Papada, Lefkothea & Kaliampakos, Dimitris, 2018. "A Stochastic Model for energy poverty analysis," Energy Policy, Elsevier, vol. 116(C), pages 153-164.

Papada L. & Kaliampakos D, 2020. Being forced to skimp on energy needs: A new look at energy poverty in Greece, Energy Research & Social Science, Elsevier,vol. 64

P. Nussbaumer, M. Bazilian, Vijay M., Measuring energy poverty: Focusing on what matters, Renewable and Sustainable Energy Reviews, Volume 16, Issue 1, 2012, Pages 231-243,

Pachauri S., Mueller A., Kemmler A. & Spreng D. (2004). On Measuring Energy Poverty in Indian Households. World Development, 32 (12), pp. 2083–2104.

Phimister E., Vera-Toscano E. & Roberts D. (2015). The dynamics of energy poverty: evidence from Spain. Econ. Energy Environ. Policy, 4, pp. 153–166.Practical Action (2009

P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697857.

Practical Action (2009). Energy poverty: the hidden energy crisis. Warwickshire, UK.

Price et al. 2012, Price C. W., Brazier K. & Wang W. (2012). Objective and subjective measures of fuel poverty. Energy Policy, 49, pp. 33–39.

Price et al. 2007 Price C. W., Brazier K., Pham K., Mathieu L. & Wang W. (2007). Identifying fuel poverty using objective and subjective measures. Centre for Competition Policy Working Paper 07-11. Norwich, UK: University of East Anglia.

Price 2002 Price M. F. (2002). Mountain Geology, Natural History and Ecosystems. Stillwater, USA: Voyageur Press

Roberts et al. 2015, Roberts D., Vera-Toscano E. & Phimister E. (2015). Fuel poverty in the UK: Is there a difference between rural and urban areas?. Energy Policy, 87, pp. 216–223.

Scottish Executive (2002). The Scottish Fuel Poverty Statement. http://www.scotland.gov.uk/Resource/Doc/46951/0031675.pdf

Statista.com, Weekly Crude Oil Prices ,2022, https://www.statista.com/statistics/326017/weekly-crude-oil-prices/

Taxheaven,2018,https://www.taxheaven.gr/news/42899/o-neos-xarths-twn-meionektikwn-perioxwn-ths-xwras

Tennakoon D. (2008). Energy poverty: Estimating the level of energy poverty in Sri Lanka. Report Submitted to Practical Action, South Asia.

Thomson H., Bouzarovski S. & Snell C. (2017). Rethinking the measurement of energy poverty in Europe: A critical analysis of indicators and data. Indoor and Built Environment, 0(0), pp. 1–23

Thomson& Snell 2013 Thomson H. & Snell C. (2013). Quantifying the prevalence of fuel poverty across the European Union. Energy Policy, 52, pp. 563–572

Tourkolias C. (2016). Observatory of energy poverty in Greece. 6th CA EED plenary meeting. Center for Renewable Sources and Savings (CRES). Hague

Towards data science, Medium, 2018, https://towardsdatascience.com/

Trading Economics, https://tradingeconomics.com/commodities

UNDP (United Nations Development Programme) (2000). World Energy Assessment: Energy and the Challenge of Sustainability. United Nations Department of Economic and Social Affairs. Washington, D.C.: UNDP

United Nations General Assembly (2013). Sustainable Mountain Development. Report of the Secretary- General. Sixty-eighth session, Item 19

WEO (World Energy Outlook) (2016). http://www.worldenergyoutlook.org/resources/energydevelopment/energyaccessdatabase

WHO (World Health Organization) (2021) Household air pollution and health,2021 , https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health

Wondimagegn M., Teshome S., Gudina L., Ecosystem services research in mountainous regions: A systematic literature review on current knowledge and research gaps, Science of The Total Environment, Volume 702,2020

Ανεξάρτητος Διαχειριστής Μεταφοράς Ηλεκτρικής Ενέργειας (ΑΔΜΗΕ), 2022, 2022-Μεσοσταθμική Τιμή Αγοράς Ενέργειας στο Διασυνδεδεμένο Σύστημα, https://www.admie.gr/sites/default/files/attached-files/type-file/2022/06/202205-Mesostathmiki_Timi_Agoras_Hlektrikhs_energeias_sto_Diasyndedemeno_Systima_v1.pdf

Ανεξάρτητος Διαχειριστής Μεταφοράς Ηλεκτρικής Ενέργειας (ΑΔΜΗΕ), 2022, 2021-Μεσοσταθμική Τιμή Αγοράς Ενέργειας στο Διασυνδεδεμένο Σύστημα, https://www.admie.gr/sites/default/files/attached-files/type-file/2022/02/202112-Mesostathmiki_Timi_Agoras_Hlektrikhs_energeias_sto_Diasyndedemeno_Systima_v3.pdf

Γενική Γραμματεία Πληροφοριακών συστημάτων (ΓΓΠΣ) ,2011, Στατιστικό Δελτίο Φορολογικών Δεδομένων, https://www.gsis.gr/sites/default/files/2019-06/statdeltio2011.pdf

Δημόσια Επιχείρηση Ηλεκτρισμού (ΔΕΗ), 2021,Τιμολόγια Ρυθμιστικού Ελέγχου, Οικιακό Τιμολόγιο Γ1, https://www.dei.gr/media/ywthxibn/timok-xt-aug2021-g1_nd-1.pdf

ΕΛΣΤΑΤ (Ελληνική Στατιστική Αρχή) 2011, Δελτίο τύπου, Απογραφή κτιρίων 2011. Πειραιάς : ΕΛΣΤΑΤ, 2015

ΕΛΣΤΑΤ (Ελληνική Στατιστική Αρχή) 2011, Δελτίο τύπου, Απογραφή κτιρίων 2011,Χαρακτηριστκά και Ανέσεις. Πειραιάς : ΕΛΣΤΑΤ, 2015

ΕΟΚΕ (Ευρωπαϊκή Οικονομική και Κοινωνική Επιτροπή) (Εισηγητές: Coulon & H. Bataller). (2013). Γνωμοδότηση TEN 516: Για μια συντονισμένη ευρωπαϊκή δράση για την πρόληψη και την καταπολέμηση της ενεργειακής φτώχειας. Βρυξέλλες.

Ινστιτούτο Εναλλακτικών Πολιτικών (Ε.Ν.Α.) , Παρατηρητήριο Βιώσιμης Ανάπτυξης,2021, https://www.enainstitute.org/publication

Κατσουλάκος Ν. (2013). Βέλτιστη χρήση ανανεώσιμων πηγών ενέργειας στις ορεινές περιοχές. Η περίπτωση του Μετσόβου. Διδακτορική Διατριβή. Επιβλέπων: Καθ. Δ. Καλιαμπάκος. Σχολή Μηχανικών Μεταλλείων-Μεταλλουργών, Εθνικό Μετσόβιο Πολυτεχνείο, Αθήνα.

Πανάς (2012). Έρευνα για την ενεργειακή φτώχεια στην Ελλάδα. Τμήμα Στατιστικής, Οικονομικό Πανεπιστήμιο Αθηνών, Αθήνα

Παπαδά Λ. (2021). Η τεχνητή νοημοσύνη στον τομέα της ενεργειακής φτώχειας. Πρόδρομα ευρήματα για την Ελλάδα. *Τα βουνά. Σύγχρονες προσεγγίσεις για τις ορεινές περιοχές*. Τόμος 1. ΜΕ.Κ.ΔΕ., ΔΠΜΣ «Περιβάλλον και Ανάπτυξη», 110-121. Εκδόσεις ΕΜΠ. ISBN: 978-618-85675-0-4.