



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ
ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Εφαρμογή Μηχανικής Μάθησης για την Πρόβλεψη Επίδρασης Έργων Ενεργειακής Ανακαίνισης Κτιρίων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωνσταντίνος Α. Κεφάλας

Επιβλέπων : Ευάγγελος Μαρινάκης

Επίκουρος Καθηγητής ΕΜΠ

Αθήνα, Φεβρουάριος 2024



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ
ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Εφαρμογή Μηχανικής Μάθησης για την Πρόβλεψη Επίδρασης Έργων Ενεργειακής Ανακαίνισης Κτιρίων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Κωνσταντίνος Α. Κεφάλας

Επιβλέπων : Ευάγγελος Μαρινάκης

Επίκουρος Καθηγητής ΕΜΠ

Ευάγγελος Μαρινάκης

Ιωάννης Ψαρράς

Δημήτριος Ασκούνης

Επίκουρος Καθηγητής Ε.Μ.Π.

Καθηγητής Ε.Μ.Π.

Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2024

Κωνσταντίνος Κ. Κεφάλας

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών, Ε.Μ.Π.

Copyright © Κωνσταντίνος Κ. Κεφάλας, 2024

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

*Στην Αλεξάνδρα, τον Ανδρέα, τον Ηλία, την Χρυσούλα,
την Μαριαλένα, στους παππούδες και στη γιαγιά μου.....*

Περίληψη

Με την αύξηση της παγκόσμιας ενεργειακής κατανάλωσης και την επιδείνωση των περιβαλλοντικών προβλημάτων, η ανάγκη για αποτελεσματικές λύσεις που θα μειώσουν την κατανάλωση ενέργειας στα κτίρια είναι επιτακτική. Οι ενεργειακές ανακαινίσεις παίζουν καθοριστικό ρόλο στην επίτευξη αυτού του στόχου, επιτρέποντας τη βελτίωση της ενεργειακής απόδοσης των κτιρίων και τη μείωση του αντίκτυπού τους στο περιβάλλον. Ωστόσο, παρά τη σημασία τους, ο ρυθμός των ενεργειακών ανακαινίσεων παραμένει συχνά χαμηλός. Συνεπώς, η ανάπτυξη και η εφαρμογή νέων τεχνολογιών και προσεγγίσεων για την ενίσχυση των ενεργειακών ανακαινίσεων αποτελεί κρίσιμη προτεραιότητα προκειμένου να επιτευχθεί η ενεργειακή αποδοτικότητα στα κτίρια και να μειωθεί η συνολική ενεργειακή κατανάλωση.

Η παρούσα διπλωματική εργασία ασχολείται με την εφαρμογή αλγορίθμων μηχανικής μάθησης για την πρόβλεψη της μείωσης της ενεργειακής κατανάλωσης και των εκπομπών διοξειδίου του άνθρακα σε κτίρια ως αποτέλεσμα ενεργειακών ανακαινίσεων. Αρχικά, προσδιορίζεται το πρόβλημα που αφορά την ενεργειακή αποδοτικότητα στα κτίρια και επισημαίνονται τα εμπόδια στην υλοποίηση έργων βελτίωσης της ενεργειακής αποδοτικότητας. Στη συνέχεια, πραγματοποιείται ανάλυση των διαθέσιμων δεδομένων από έργα ανακαίνισης που έχουν υλοποιηθεί στο παρελθόν, ενώ εφαρμόζονται αλγόριθμοι μηχανικής μάθησης (K-Nearest Neighbors, Random Forest, Lightgbm, Support Vector Machine, Τεχνητό Νευρωνικό Δίκτυο) για την εκπαίδευση μοντέλων πρόβλεψης με βάση αυτά τα δεδομένα. Συνεπώς καταγράφονται οι αλγόριθμοι με το μεγαλύτερο βαθμό απόδοσης και αναλύονται μέθοδοι βελτιστοποίησης των τεχνολογιών που εφαρμόστηκαν. Τα εξαγόμενα αποτελέσματα τονίζουν την σημασία της μηχανικής μάθησης στον τομέα της ενεργειακής αποδοτικότητας αλλά και της ενέργειας γενικότερα, καθώς αναδεικνύονται οι προοπτικές μοντέλων βασισμένων στη μηχανική μάθηση να υποστηρίξουν τους εμπλεκόμενους φορείς σε διαδικασίες λήψης αποφάσεων, χρησιμοποιώντας τα υπάρχοντα δεδομένα, πραγματοποιώντας συστηματική ανάλυση με βάση τις διαθέσιμες πληροφορίες και προσφέροντας χρήσιμες προβλέψεις για την επίδραση συγκεκριμένων παρεμβάσεων. Επομένως, μπορούν να συνδράμουν σε περαιτέρω έρευνες εμβαθύνοντας τόσο σε πρακτικό όσο και σε ερευνητικό επίπεδο.

Λέξεις κλειδιά: Μηχανική Μάθηση, Ενεργειακή αποδοτικότητα, Ενεργειακές ανακαινίσεις κτιρίων, Τεχνητή νοημοσύνη, Διαχείριση ενέργειας

Abstract

With global energy consumption increasing and environmental problems worsening, the need for effective solutions to reduce energy consumption in buildings is urgent. Energy renovation plays a key role in achieving this objective, allowing the energy efficiency of buildings to be improved and their negative impact on the environment to be mitigated. However, despite their importance, the rate of energy renovations often remains low, partly due to the financial and technical challenges faced by building owners and managers. Therefore, the development and implementation of new technologies and approaches to enhance energy renovation is a critical priority in order to achieve energy efficiency in buildings and reduce overall energy consumption.

This thesis evaluates the application of machine learning algorithms to predict the reduction of energy consumption and carbon emissions in buildings. To begin with, the problem is defined, referring to energy efficiency in buildings, and the barriers to implementing energy efficiency projects are highlighted. Then, an analysis of the available data from energy efficiency projects that have been implemented in the past is carried out and machine learning algorithms (K-Nearest Neighbors, Random Forest, Lightgbm, Support Vector Machine, Artificial Neural Network) are applied to train prediction models using these data. Subsequently, the algorithms with the highest degree of efficiency are identified and optimization methods of the technologies applied are analyzed. The results obtained emphasize the importance of machine learning in the field of energy efficiency and energy in general, since the potential of machine learning models to support stakeholders in decision making procedures are highlighted, taking advantage of existing data, conducting systematic analysis of available information and providing useful insights and predictions of the impact of specific interventions. Therefore, they can contribute to further research by deepening both practical and research level.

Keywords: Machine learning, Energy efficiency, Energy renovation of buildings, Artificial intelligence, Energy management, Energy efficiency, Energy renovations

Πρόλογος

Η συγκεκριμένη διπλωματική εργασία εκπονήθηκε στο πλαίσιο των ερευνητικών δραστηριοτήτων του Εργαστηρίου Συστημάτων Αποφάσεων και Διοίκησης, το οποίο υπάγεται στη Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Ηλεκτρονικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου (ΕΜΠ), κατά το ακαδημαϊκό έτος 2023-2024 και την περίοδο Οκτωβρίου 2023-Φεβρουαρίου 2024. Η εργασία πραγματοποιήθηκε υπό την επίβλεψη του κ. Βαγγέλη Μαρινάκη, Επίκουρου Καθηγητή ΕΜΠ, στην Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, τον οποίο θα ήθελα να ευχαριστήσω θερμά για την ευκαιρία που μου έδωσε να ασχοληθώ με το συγκεκριμένο θέμα.

Στόχος της συγκεκριμένης διπλωματικής είναι η εφαρμογή μοντέλων τεχνητής νοημοσύνης και μηχανικής μάθησης, με απώτερο σκοπό την εύρεση των βέλτιστων δράσεων εξοικονόμησης ενέργειας σε κτιριακές δομές. η εργασία αποσκοπεί στον εντοπισμό των βέλτιστων δράσεων εξοικονόμησης ενέργειας σε κτίρια, λαμβάνοντας υπόψη τα δεδομένα που προέρχονται από προηγούμενες ενεργειακές ανακαινίσεις. Μέσω της εφαρμογής αυτών των μοντέλων και της εκτέλεσης αναλύσεων, προβλέπεται ότι θα επιτευχθεί η μέγιστη δυνατή εξοικονόμηση ενέργειας σε κάθε κτιριακή δομή, συνεισφέροντας έτσι στη μείωση της ενεργειακής κατανάλωσης και στην ενίσχυση της βιωσιμότητας του κτιριακού τομέα.

Θα ήθελα να ευχαριστήσω ιδιαιτέρως τους ερευνητές του εργαστηρίου Συστημάτων Αποφάσεων και Διοίκησης Κατερίνα Παπαποστόλου, Ιωάννα Ανδρεουλάκη και Ντανιέλα Στοϊάν για την εξαιρετική συνεργασία και επικοινωνία που είχαμε, την επίλυση αποριών και τη σημαντική βοήθεια που μου προσέφεραν σε όλη τη διάρκεια εκπόνησης της παρούσας διπλωματικής εργασίας.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου και το στενό φιλικό μου κύκλο για την αδιάκοπη και πολύπλευρη υποστήριξη που μου προσέφεραν κατά την διάρκεια των σπουδών μου και κατά την εκπόνηση της παρούσας εργασίας.

Κωνσταντίνος Κεφάλας,

Φεβρουάριος 2024

Πίνακας περιεχομένων

Ευρετήριο εικόνων	11
Ευρετήριο διαγραμμάτων	12
Ευρετήριο πινάκων	14
1 Εισαγωγή	18
1.1 Αντικείμενο-Σκοπός.....	18
1.2 Φάσεις υλοποίησης	19
1.3 Οργάνωση εργασίας	22
2 Ενεργειακή αποδοτικότητα στα κτίρια: Ευρωπαϊκοί Στόχοι και Εμπόδια.....	25
3 Μηχανική Μάθηση στην Ενέργεια	33
3.1 Μηχανική Μάθηση: Βασικές Έννοιες	33
3.2 Μηχανική Μάθηση στην Ενέργεια: Επιστημονικά Άρθρα και Εφαρμογές .34	
3.2.1 Μηχανική Μάθηση στα Ενεργειακά Συστήματα	38
3.2.2 Μηχανική Μάθηση στην Ενεργειακή Αποδοτικότητα Μη Οικιστικών Κτιρίων40	
3.2.3 Μηχανική Μάθηση στην Ενεργειακή Αποδοτικότητα Οικιστικών Κτιρίων 41	
3.2.4 Μηχανική Μάθηση στην Ενεργειακή Αποδοτικότητα εκτός Κτιριακού Τομέα 42	
4 Υλοποίηση Μοντέλων Μηχανικής Μάθησης για την Πρόβλεψη της Αποτελεσματικότητας Ενεργειακών Ανακαινίσεων.....	48
4.1 Εργαλεία Συστήματος	48
4.2 Μηχανική Μάθηση & Ανάλυση Δεδομένων.....	50
4.2.1 Αρχικά Δεδομένα	50
4.2.2 Ανάλυση Δεδομένων.....	59
4.3 Εφαρμογή Αλγορίθμων Μηχανικής Μάθησης	68
4.3.1 Αλγόριθμος K-Nearest-Neighbors (KNN)	74
4.3.2 Random Forest	81
4.3.3 Lightgbm	87
4.3.4 Lightgbm (v.2)	94
4.3.5 Support Vector Machine (SVM).....	101
4.3.6 Artificial Neural Network (ANN).....	107
5 Σχολιασμός των Αποτελεσμάτων	119
6 Συμπεράσματα και Προτάσεις για Περαιτέρω Έρευνα	127

6.1	Συμπεράσματα για τους αλγορίθμους	127
6.2	Συμπεράσματα για την έρευνα.....	129
6.3	Προτάσεις για περαιτέρω έρευνα	130
7	Βιβλιογραφία	134
8	Παράρτημα	141
8.1	Παράρτημα 1: Συντομογραφίες.....	141
8.2	Παράρτημα 2: Περιγραφή επιστημονικών άρθρων βιβλιογραφικής ανασκόπησης	142

Ευρετήριο εικόνων

<i>Εικόνα 1: Μεθοδολογία της παρούσας διπλωματικής εργασίας</i>	<i>21</i>
<i>Εικόνα 2: Θεματικοί τομείς του ML στα ενεργειακά συστήματα (Mosavi et al., 2019)[73].....</i>	<i>39</i>
<i>Εικόνα 3: Δεδομένα από το πρώτο dataset</i>	<i>52</i>
<i>Εικόνα 4: Δεδομένα από το πρώτο dataset</i>	<i>52</i>
<i>Εικόνα 5: Δεδομένα από το δεύτερο dataset</i>	<i>54</i>
<i>Εικόνα 6: Δεδομένα από το δεύτερο dataset</i>	<i>54</i>
<i>Εικόνα 7: Δεδομένα από το τρίτο dataset</i>	<i>56</i>
<i>Εικόνα 8: Δεδομένα από το τρίτο dataset</i>	<i>56</i>
<i>Εικόνα 9: Δεδομένα από το τρίτο dataset</i>	<i>57</i>
<i>Εικόνα 10: Δεδομένα από το τρίτο dataset</i>	<i>57</i>
<i>Εικόνα 11: Δεδομένα από το τρίτο dataset</i>	<i>58</i>
<i>Εικόνα 12: Δεδομένα από το τρίτο dataset</i>	<i>58</i>
<i>Εικόνα 13: Αρχιτεκτονική νευρωνικού δικτύου</i>	<i>112</i>
<i>Εικόνα 14: Προβλέψεις vs Πραγματικές τιμές για τον αλγόριθμο Lightgbm(v2).....</i>	<i>123</i>

Ευρετήριο διαγραμμάτων

Διάγραμμα 1: Μηχανική Μάθηση	34
Διάγραμμα 2: Βιβλιογραφική ανασκόπηση	35
Διάγραμμα 3: Κατηγοριοποίηση επιστημονικών άρθρων για τη μηχανική μάθηση στην ενέργεια.....	37
Διάγραμμα 4: Μεθοδολογία ανάλυσης δεδομένων.....	60
Διάγραμμα 5: Κατανομή των διαφόρων ειδών ανακαίνισης στα κτίρια που μελετήθηκαν	62
Διάγραμμα 6: Εμφάνιση πιθανών λέξεων στην κατηγορία "Project Activity"	63
Διάγραμμα 7: Εμφάνιση των συγκεκριμένων λέξεων στην κατηγορία "Project Activity"	64
Διάγραμμα 8: Μοναδικές εμφανίσεις κάθε λέξης στην κατηγορία "Project Activity"	64
Διάγραμμα 9: Εμφάνιση των συγκεκριμένων λέξεων στην κατηγορία "Type of activity: renovation of the building's enclosing structures"	65
Διάγραμμα 10: Μεθοδολογία εφαρμογής των αλγόριθμων μηχανικής μάθησης.....	68
Διάγραμμα 11: Χάρτης συσχέτισης χαρακτηριστικών.....	71
Διάγραμμα 12: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	75
Διάγραμμα 13: Iterations για εύρεση καλύτερου k	76
Διάγραμμα 14: Προβλέψεις vs Πραγματικές τιμές.....	77
Διάγραμμα 15: Κατανομή του "Sum of % CO2 reduction" συγκριτικά με τις προβλέψεις	77
Διάγραμμα 16: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	78
Διάγραμμα 17: Iterations για εύρεση καλύτερου k	79
Διάγραμμα 18: Προβλέψεις vs Πραγματικές τιμές.....	79
Διάγραμμα 19: Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις	80
Διάγραμμα 20: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	82
Διάγραμμα 21: Iterations για εύρεση καλύτερου n	83
Διάγραμμα 22: Προβλέψεις vs Πραγματικές τιμές.....	83
Διάγραμμα 23: Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις	84
Διάγραμμα 24: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	85
Διάγραμμα 25: Iterations για την επιλογή βέλτιστου n	85
Διάγραμμα 26: : Προβλέψεις vs Πραγματικές τιμές.....	86
Διάγραμμα 27: : Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις	86
Διάγραμμα 28: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	89
Διάγραμμα 29: Προβλέψεις vs Πραγματικές τιμές.....	90
Διάγραμμα 30: Κατανομή του "Sum of % CO2 reduction" συγκριτικά με τις προβλέψεις	90
Διάγραμμα 31: Τα 5 χαρακτηριστικά με τη μεγαλύτερη σημασία για το μοντέλο	91
Διάγραμμα 32: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	92

Διάγραμμα 33: Προβλέψεις vs Πραγματικές τιμές.....	93
Διάγραμμα 34: Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις	93
Διάγραμμα 35: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	96
Διάγραμμα 36: Προβλέψεις vs Πραγματικές τιμές.....	97
Διάγραμμα 37: Κατανομή του "Sum of % CO2 reduction" συγκριτικά με τις προβλέψεις	97
Διάγραμμα 38: Τα 5 χαρακτηριστικά με τη μεγαλύτερη σημασία για το μοντέλο	98
Διάγραμμα 39: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	99
Διάγραμμα 40: Προβλέψεις vs Πραγματικές τιμές.....	99
Διάγραμμα 41: Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις	100
Διάγραμμα 42: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	103
Διάγραμμα 43: Προβλέψεις vs Πραγματικές τιμές.....	104
Διάγραμμα 44: Κατανομή του "Sum of % CO2 reduction" συγκριτικά με τις προβλέψεις	104
Διάγραμμα 45: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	105
Διάγραμμα 46: Προβλέψεις vs Πραγματικές τιμές.....	106
Διάγραμμα 47: Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις	106
Διάγραμμα 48: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	113
Διάγραμμα 49: Προβλέψεις vs Πραγματικές τιμές.....	113
Διάγραμμα 50: Κατανομή του "Sum of % CO2 reduction" συγκριτικά με τις προβλέψεις	114
Διάγραμμα 51: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα	115
Διάγραμμα 52: Προβλέψεις vs Πραγματικές τιμές.....	115
Διάγραμμα 53: Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις	116

Ευρετήριο πινάκων

<i>Πίνακας 1: Εμπόδια στην ενεργειακή ανακαίνιση κτιρίων.....</i>	<i>26</i>
<i>Πίνακας 2: Λέξεις κλειδιά για την αναζήτηση επιστημονικών άρθρων</i>	<i>36</i>
<i>Πίνακας 3: Κατηγοριοποίηση επιστημονικών άρθρων.....</i>	<i>36</i>
<i>Πίνακας 4: Κατηγοριοποίηση Λέξεων.....</i>	<i>66</i>
<i>Πίνακας 5: Χαρακτηριστικά που επιλέχθηκαν.....</i>	<i>68</i>
<i>Πίνακας 6: Συγκεντρωτικά αποτελέσματα των αλγορίθμων.....</i>	<i>120</i>
<i>Πίνακας 7: Καλύτεροι και χειρότεροι αλγόριθμοι</i>	<i>122</i>
<i>Πίνακας 8: Πίνακας κατάταξης των αλγορίθμων.....</i>	<i>123</i>
<i>Πίνακας 9: Πίνακας Συντομογραφιών</i>	<i>141</i>
<i>Πίνακας 10: Σύντομη περιγραφή των επιστημονικών άρθρων.....</i>	<i>142</i>

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

ΚΕΦΑΛΑΙΟ 1

Εισαγωγή

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

1 Εισαγωγή

1.1 Αντικείμενο-Σκοπός

Η ενεργειακή αποδοτικότητα στον κτιριακό τομέα αναδεικνύεται ως ένας από τους σημαντικότερους παράγοντες που επηρεάζουν τη βιωσιμότητα και την οικονομική αποδοτικότητα των κτιρίων στην Ευρωπαϊκή Ένωση. Παρά την αναγνώριση της σημασίας της, ο ρυθμός εφαρμογής ενεργειακών ανακαινίσεων παραμένει σε σχετικά χαμηλά επίπεδα. Ως αποτέλεσμα, η ανάγκη για αποτελεσματικότερες και πιο καινοτόμες προσεγγίσεις για τη βελτίωση της ενεργειακής απόδοσης στα κτίρια είναι επίκαιρη και αναπόφευκτη.

Σε αυτό το πλαίσιο, η προσθήκη νέων τεχνολογιών και η εφαρμογή μοντέλων τεχνητής νοημοσύνης και μηχανικής μάθησης προκύπτει ως κρίσιμη στρατηγική. Η ενσωμάτωση αυτών των τεχνολογιών στη διαδικασία λήψης αποφάσεων για ενεργειακές αναβαθμίσεις μπορεί να οδηγήσει σε σημαντική αύξηση του ρυθμού ανακαινίσεων και στην μέγιστη αξιοποίηση των δυνατοτήτων εξοικονόμησης ενέργειας που προσφέρονται από αυτές. Επιπλέον, η υιοθέτηση τεχνολογιών τεχνητής νοημοσύνης και μηχανικής μάθησης στον τομέα της ενεργειακής ανακαίνισης κτιρίων προσφέρει μια προοδευτική προσέγγιση προς τη βελτίωση της αποδοτικότητας της ενέργειας. Αντί για την παραδοσιακή προσέγγιση που βασίζεται σε απλούς μηχανικούς υπολογισμούς και εμπειρία, η χρήση αυτών των προηγμένων τεχνολογιών επιτρέπει την ανάλυση μεγάλων όγκων δεδομένων και την εξαγωγή συμπερασμάτων με βάση πιο σύνθετους αλγόριθμους. Συγχρόνως, η ενσωμάτωση τεχνητής νοημοσύνης στη διαδικασία λήψης αποφάσεων για την ενεργειακή ανακαίνιση των κτιρίων δίνει τη δυνατότητα για προσαρμοστική και δυναμική βελτιστοποίηση των στρατηγικών. Αυτό σημαίνει ότι οι αποφάσεις λαμβάνονται με βάση τα πραγματικά δεδομένα και τις συνθήκες λειτουργίας του κάθε κτιρίου, προσφέροντας έτσι μια πιο αποτελεσματική προσέγγιση.

Η παρούσα διπλωματική εργασία έχει ως στόχο την ανάπτυξη και εφαρμογή μοντέλων τεχνητής νοημοσύνης και μηχανικής μάθησης στον κτιριακό τομέα. Συγκεκριμένα, η εργασία αποσκοπεί στον εντοπισμό των βέλτιστων δράσεων εξοικονόμησης ενέργειας σε κτίρια, λαμβάνοντας υπόψη τα δεδομένα που προέρχονται από προηγούμενες ενεργειακές ανακαινίσεις. Αυτή η προσέγγιση επιτρέπει την ανάπτυξη εξατομικευμένων στρατηγικών ανακαίνισης για κάθε κτιριακή δομή, βασισμένη στα ειδικά χαρακτηριστικά και τις ανάγκες του κάθε έργου. Μέσω της εφαρμογής αυτών των μοντέλων και της εκτέλεσης αναλύσεων, προβλέπεται ότι θα επιτευχθεί η μέγιστη δυνατή εξοικονόμηση ενέργειας σε κάθε κτιριακή δομή, συνεισφέροντας έτσι στη μείωση της ενεργειακής κατανάλωσης και στην ενίσχυση της βιωσιμότητας του κτιριακού τομέα.

1.2 Φάσεις υλοποίησης

Η υλοποίηση μιας διπλωματικής εργασίας που αποσκοπεί στην ανάπτυξη και εφαρμογή μοντέλων τεχνητής νοημοσύνης και μηχανικής μάθησης για τη βελτίωση της ενεργειακής απόδοσης στον κτιριακό τομέα απαιτεί μια συστηματική προσέγγιση και ένα πολυσύνθετο σχέδιο εργασίας. Σε αυτό το πλαίσιο, οι φάσεις υλοποίησης αποτελούν κρίσιμα στάδια που οδηγούν στην επίτευξη των στόχων της έρευνας και την πρακτική εφαρμογή των μοντέλων που αναπτύσσονται. Ως βασικές φάσεις υλοποίησης που προβλέπονται για την παρούσα διπλωματική εργασία είναι οι ακόλουθες:

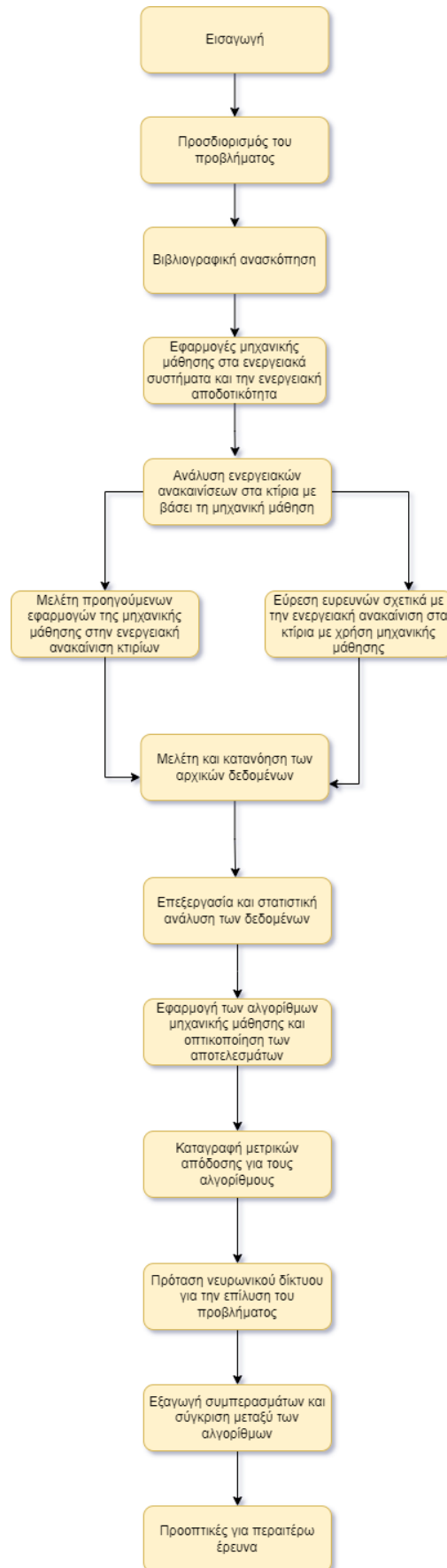
1. **Θεωρητικό υπόβαθρο:** Ως πρώτο βήμα πραγματοποιείται μια εκτενής ανάλυση του θεωρητικού υποβάθρου που αφορά τον κτιριακό τομέα και τις συναφείς τεχνολογίες εξοικονόμησης ενέργειας. Προσδιορίζεται αρχικά το πρόβλημα που αφορά την ενεργειακή αποδοτικότητα στα κτίρια το οποίο περιλαμβάνει πτυχές όπως οι περιορισμένες ανακαινίσεις που υλοποιούνται αλλά και τα εμπόδια στην υλοποίηση του ενεργειακού αποδοτικού σχεδιασμού γενικότερα. Περιγράφονται οι βασικές έννοιες και οι μεθοδολογίες που αφορούν στην ενεργειακή αποδοτικότητα των κτιρίων, παρέχοντας χρήσιμες πληροφορίες για την κατανόηση του πεδίου έρευνας.
2. **Βιβλιογραφική ανασκόπηση:** Σε αυτή τη φάση πραγματοποιείται μια λεπτομερής ανασκόπηση της βιβλιογραφίας σχετικά με την έννοια της μηχανικής μάθησης και τη χρήση της στα ενεργειακά συστήματα, την εφαρμογή της στον τομέα της ενεργειακής αποδοτικότητας αλλά και τις ενεργειακές ανακαινίσεις στα κτίρια και την εφαρμογή τεχνολογιών τεχνητής νοημοσύνης και μηχανικής μάθησης σε αυτό το πλαίσιο. Αναλύονται προηγούμενες μελέτες, ερευνητικά έργα και τεχνολογικές εξελίξεις που σχετίζονται με το θέμα της ενεργειακής απόδοσης στα κτίρια.
3. **Ανάλυση Δεδομένων:** Επόμενο βήμα αποτελεί η ανάλυση των δεδομένων που σχετίζονται με την ενεργειακή αποδοτικότητα των κτιρίων τα οποία προέρχονται από έργα ενεργειακής ανακαίνισης στη Λετονία. Η ανάλυση αυτή περιλαμβάνει την επεξεργασία στοιχείων σχετικά με την κατανάλωση ενέργειας, τα χαρακτηριστικά του κτιρίου και τις ενεργειακές ανακαινίσεις που έχουν πραγματοποιηθεί με στόχο τη δημιουργία ενός πακέτου δεδομένων εισόδου το οποίο θα καθιστά τη λειτουργία των αλγορίθμων μηχανικής μάθησης αποδοτική.
4. **Εφαρμογή Αλγορίθμων:** Έχοντας αναλύσει και μοντελοποιήσει τα δεδομένα ακολουθεί η ανάπτυξη και η εφαρμογή αλγορίθμων τεχνητής νοημοσύνης και μηχανικής μάθησης για την ανάλυση των δεδομένων και την πρόβλεψη βέλτιστων δράσεων εξοικονόμησης ενέργειας. Παράλληλα παρατίθενται και ορισμένες γραφικές παραστάσεις για την καλύτερη κατανόηση της λειτουργίας και των αποτελεσμάτων των αλγορίθμων.
5. **Αξιολόγηση και Σχολιασμός των Αποτελεσμάτων:** Σε αυτή τη φάση πραγματοποιείται η αξιολόγηση της απόδοσης των αναπτυγμένων μοντέλων και αλγορίθμων, καθώς και ο σχολιασμός των παραγόμενων αποτελεσμάτων

ενώ επιλέγεται και η ιδανικότερη στρατηγική για την πρόβλεψη των ενεργειακών ανακαινίσεων στα κτίρια.

6. **Αναγνώριση Προκλήσεων και Αδυναμιών:** Αναλύονται οι πιθανές αδυναμίες της μεθοδολογίας ή των μοντέλων που χρησιμοποιήθηκαν και οι περιορισμοί των αποτελεσμάτων.
7. **Συμπεράσματα:** Στο κεφάλαιο αυτό διατυπώνονται τα συμπεράσματα τα οποία εξάγονται από την παρούσα έρευνα και το πώς συνδράμει με μία καινοτόμα και πολυδιάστατη προσέγγιση στην δημιουργία ενεργειακά αποδοτικών κτιρίων.
8. **Προτάσεις για Περαιτέρω Έρευνα:** Διατυπώνονται προτάσεις για περαιτέρω έρευνα σε πιθανά επιμέρους θέματα που απαιτούν εμβάθυνση ή διερεύνηση. Αυτό μπορεί να περιλαμβάνει την πρόταση νέων μεθόδων ή τεχνικών, την ανάγκη για περισσότερα πειράματα ή αναλύσεις δεδομένων, ή ακόμα και την ανάγκη για περαιτέρω έρευνα σε συγκεκριμένους τομείς του κτιριακού τομέα.

Στη επόμενη σελίδα παρατίθεται ένα σχήμα που αποτυπώνει την αναλυτική μεθοδολογία της παρούσας διπλωματικής εργασίας.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Εικόνα 1: Μεθοδολογία της παρούσας διπλωματικής εργασίας

1.3 Οργάνωση εργασίας

Σε συνέχεια του τρέχοντος εισαγωγικού κεφαλαίου, η δομή της διπλωματικής εργασίας και μία σύνοψη του περιεχομένου κάθε κεφαλαίου δίνονται παρακάτω: **Κεφάλαιο 2: Θεωρητικό υπόβαθρο και Προσδιορισμός του Προβλήματος** Σε αυτό το κεφάλαιο πραγματοποιείται λεπτομερής ανάλυση του θεωρητικού υποβάθρου που αφορά την ενεργειακή αποδοτικότητα στα κτίρια. Περιγράφονται εκτενώς το πρόβλημα που συνδέονται με την αποδοτικότητα ενέργειας στα κτίρια, περιλαμβανομένων των περιορισμών στις ανακαινίσεις και των εμποδίων στην υλοποίηση του ενεργειακού σχεδιασμού και παρέχονται βασικές πληροφορίες για την κατανόηση του πεδίου έρευνας.

Κεφάλαιο 3: Μηχανική Μάθηση στην Ενέργεια Στο κεφάλαιο αυτό αναλύεται μέσω μιας εκτενούς βιβλιογραφικής ανασκόπησης η έννοια της μηχανικής μάθησης, η συμβολή της στα συστήματα ενέργειας αλλά και στον τομέα της ενεργειακής αποδοτικότητας γενικότερα και πιο συγκεκριμένα πώς οι αλγόριθμοι μηχανικής μάθησης έχουν εφαρμοστεί εντός και εκτός κτιριακού τομέα και πιο εξειδικευμένα στον οικιστικό τομέα στο κομμάτι της ενεργειακής αποδοτικότητας.

Κεφάλαιο 4: Υλοποίηση Μοντέλων Στο κεφάλαιο αυτό παρουσιάζονται συνοπτικά οι βιβλιοθήκες που χρησιμοποιήθηκαν για την υλοποίηση της εργασίας και καταγράφονται αναλυτικά οι ενέργειες που πραγματοποιήθηκαν για την επεξεργασία των αρχικών δεδομένων ώστε να αποκτήσουν την απαραίτητη μορφή για να χρησιμοποιηθούν ως δεδομένα εισόδου για τους αλγόριθμους μηχανικής μάθησης. Συγχρόνως παρουσιάζονται πληροφορίες για καθένα αλγόριθμο που εφαρμόστηκε ξεχωριστά, το πώς εφαρμόστηκε ο κάθε αλγόριθμος, τα αποτελέσματα τα οποία επέστρεψε αλλά και κάποιες μετρικές απόδοσης ώστε να μπορούν να αξιολογηθούν στη συνέχεια.

Κεφάλαιο 5: Σχολιασμός των αποτελεσμάτων Στο κεφάλαιο αυτό παρατίθενται τα αποτελέσματα κάθε αλγορίθμου όπως φαίνονται μέσα από τις μετρικές απόδοσης, συγκρίνονται μεταξύ τους και επιλέγεται ο βέλτιστος αλγόριθμος για την επιλογή ενεργειών που στόχο έχουν την ενεργειακή ανακαίνιση κτιρίων. Παράλληλα γίνονται παρατηρήσεις για κάθε αλγόριθμο και προτείνονται βελτιστοποιήσεις.

Κεφάλαιο 6: Συμπεράσματα και προτάσεις για περαιτέρω έρευνα Στο κεφάλαιο αυτό διατυπώνονται τα συμπεράσματα που εξάγονται από την παρούσα διπλωματική εργασία, καθώς και οι προοπτικές και προτάσεις για περαιτέρω έρευνα.

Κεφάλαιο 7: Βιβλιογραφία Στο κεφάλαιο αυτό παρουσιάζονται αναλυτικά οι πηγές από τις οποίες εξήχθησαν πληροφορίες για την υλοποίηση της παρούσας διπλωματικής εργασίας.

ΚΕΦΑΛΑΙΟ 2

Θεωρητικό Υπόβαθρο και Προσδιορισμός του Προβλήματος

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

2 Ενεργειακή αποδοτικότητα στα κτίρια: Ευρωπαϊκοί Στόχοι και Εμπόδια

Το θεωρητικό υπόβαθρο επικεντρώνεται στην ανάλυση του προβλήματος των ενεργειακών ανακαινίσεων στα κτίρια, αναδεικνύοντας τα εμπόδια που υφίστανται σε αυτήν την πορεία και τους στόχους που έχει θέσει η Ευρωπαϊκή Ένωση σε αυτόν τον τομέα. Εστιάζει στους περιορισμούς που συναντώνται στις ανακαινίσεις και στην ανάγκη προσδιορισμού του προβλήματος για την αποτελεσματική αντιμετώπισή του. Μέσω της ανάλυσης αυτής, καθορίζονται οι κύριες προκλήσεις που καλείται να αντιμετωπίσει ο κτιριακός τομέας προκειμένου να επιτύχει την ενεργειακή αποδοτικότητα και τη βιωσιμότητα στο μέλλον. Μέσα από αυτό το πλαίσιο, επιδιώκεται η κατανόηση των βασικών παραμέτρων και των προκλήσεων που διαμορφώνουν την ενεργειακή κατάσταση των κτιρίων, προκειμένου να οριστεί η κατεύθυνση προς την οποία θα κινηθούν οι ενεργειακές ανακαινίσεις και ο σχεδιασμός αποδοτικών ενεργειακών συστημάτων με τη βοήθεια της μηχανικής μάθησης.

Η αναγκαιότητα ενίσχυσης της ενεργειακής αποδοτικότητας έχει αναγνωριστεί από την Ευρωπαϊκή Ένωση (ΕΕ), με αποτέλεσμα να έχει αναχθεί σε προτεραιότητα και να έχει ενσωματωθεί ως αναπόσπαστο κομμάτι της πολιτικής της τα τελευταία χρόνια. Για παράδειγμα, το Green Paper της Ευρωπαϊκής Ένωσης είναι ένα έγγραφο πολιτικής που εκδίδεται από την Ευρωπαϊκή Επιτροπή με σκοπό την προώθηση δημόσιας συζήτησης και συλλογής απόψεων και πληροφοριών σχετικά σε ένα συγκεκριμένο θέμα. Το Green Paper της ΕΕ που αφορά την ενεργειακή αποδοτικότητα προτρέπει σε δράση για την επίτευξη των περιβαλλοντικών στόχων, την ενίσχυση της ασφάλειας του εφοδιασμού και την αύξηση της ανταγωνιστικότητας μέσω της μείωσης της χρήσης ενέργειας. Αρκετές μελέτες εξετάζουν το ρόλο που πρέπει να διαδραματίσουν η ενεργειακή αποδοτικότητα και οι πολιτικές που την προωθούν, σύμφωνα με την Ευρωπαϊκή Επιτροπή [1].

Προκειμένου να βοηθηθούν στην επίτευξη των στόχων εξοικονόμησης ενέργειας, τα κράτη μέλη της ΕΕ καλούνται να εφαρμόσουν συστήματα υποχρεώσεων ενεργειακής απόδοσης (Energy efficiency obligation schemes-EEOs). Σύμφωνα με τους (Fawcett et al., 2019) το 2018 τα ΕΕOs της ΕΕ σε εφαρμογή έφτασαν τα 15, σε αντίθεση με μόλις έξι πριν από την εισαγωγή της οδηγίας για την ενεργειακή αποδοτικότητα του 2012. Ταυτόχρονα, οι ανησυχίες για την αύξηση των δαπανών έχουν δημιουργήσει δυσκολίες για τα καθιερωμένα ΕΕOs στη Δανία και το Ηνωμένο Βασίλειο. Ο κύριος λόγος για τον οποίο υποστηρίχθηκαν τα ΕΕOs σε επίπεδο ΕΕ ήταν τα ισχυρά αποδεικτικά στοιχεία ότι τα καλά σχεδιασμένα ΕΕOs μπορούν να οδηγήσουν σε μεγάλη, μακροπρόθεσμη και οικονομικά αποδοτική εξοικονόμηση ενέργειας. Υπάρχει σημαντικός και διευρυμένος όγκος δεδομένων που υποστηρίζουν τα κοινωνικά και οικονομικά οφέλη των ΕΕOs. Παρόλα αυτά, αρκετά κράτη μέλη έχουν εκφράσει δημόσιες πολιτικές ανησυχίες λόγω της αυξανόμενης φιλοδοξίας των στόχων εξοικονόμησης και του κόστους των συστημάτων για τους πληρωτές των λογαριασμών. Οι στόχοι εξοικονόμησης μειώθηκαν πρόσφατα ως αποτέλεσμα

αυτού. Έτσι, ακόμη και αν τα ΕΕΟs ήταν επιτυχημένα στο παρελθόν και έχουν υιοθετηθεί από τα περισσότερα κράτη μέλη της ΕΕ, η μελλοντική τους λειτουργία παραμένει ένας γρίφο [2].

Ένα παράδειγμα χώρας που επιβεβαιώνει όσα προαναφέρθηκαν είναι η Γερμανία, η οποία έχει θέσει υψηλούς στόχους για τη δραστική βελτίωση της θερμικής ενεργειακής απόδοσης των παλαιών κατοικιών της, τη μείωση των εκπομπών διοξειδίου του άνθρακα και τη μείωση του κόστους θέρμανσης - καθώς τα σπίτια αυτά κατοικούνται σε δυσανάλογα μεγάλο ποσοστό από άτομα με χαμηλό εισόδημα. Ακόμα και όταν η ανακαίνιση γίνεται στο πιο βασικό επίπεδο ενεργειακής αποδοτικότητας, είναι αμφίβολο αν η μείωση του ενεργειακού κόστους ισοδυναμεί με το κόστος της ανακαίνισης. Η διαφορά μεταξύ δαπανών και εξοικονόμησης διευρύνεται περαιτέρω με την ανακαίνιση σε υψηλότερα πρότυπα [3]. Επομένως το κόστος των ανακαίνισεων προβληματίζει αρκετές από τις χώρες οι οποίες επιδιώκουν τη μετάβαση σε μία πιο ενεργειακά βιώσιμη πραγματικότητα.

Τα εμπόδια τα οποία συναντώνται στις ενεργειακές ανακαίνισεις κτιρίων είναι πολλά και ποικίλουν ανάλογα με τους εμπλεκόμενους φορείς. Στη συνέχεια καταγράφονται σε έναν πίνακα τα βασικότερα προβλήματα που συναντώνται στις ανακαίνισεις κτιρίων με στόχο να γίνουν πιο ενεργειακά αποδοτικά σύμφωνα με τις παρακάτω μελέτες [4], [5], [6], [7].

Πίνακας 1: Εμπόδια στην ενεργειακή ανακαίνιση κτιρίων

Κατηγορία	Θεωρητικά Εμπόδια	Σχόλια
Οργάνωση της αγοράς	Κατακερματισμένη αγορά	Κατακερματισμένος κλάδος και προσωρινοί συνασπισμοί, ζητήματα με υπεργολαβίες. Σχεδιαστές, σύμβουλοι και υπεργολάβοι δεν ενδιαφέρονται μακροπρόθεσμα για ένα κτίριο Έλλειψη ολοκλήρωσης του έργου και επικοινωνίας μεταξύ των εμπλεκόμενων φορέων Οι προτροπές για την εξέταση του κόστους όλης της ζωής είναι αδύναμες.
	Μοιρασμένα κίνητρα	Ποικίλα ενδιαφερόμενα μέρη με ποικίλα συμφέροντα . Οι εργολάβοι οικοδομικών υπηρεσιών έχουν κίνητρα για υπερδιαστασιολόγηση του εξοπλισμού
	Έλλειψη χρόνου	Έλλειψη χρόνου, επαναχρησιμοποίηση των προσφορών λόγω έλλειψης χρόνου,

		και δεν υπάρχει χρόνος για καταιγισμό ιδεών ή δημιουργικότητα, και τάση διατήρησης των υφιστάμενων πρακτικών. Λιγότερο χρονοβόρος ο σχεδιασμός ενός ακατέργαστου κτιρίου αντί να δαπανάται χρόνος για νέες λύσεις ομοιόμορφα.
Πληροφορίες	Ατελής πληροφόρηση	Έλλειψη πληροφοριών και γνώσεων σχετικά με τα ενεργειακά αποδοτικά και βιώσιμα υλικά και προϊόντα
Εμπόδια συμπεριφοράς	Έλλειψη κοινών στόχων	Τα ενεργειακά μέτρα πρέπει να συμβαδίζουν με τους στόχους
	Άλλες προτεραιότητες	Ακόμη και αν τα μέτρα μειώνουν την ενέργεια, είναι υψηλό το αρχικό κόστος και μακρά η απόσβεση-προτεραιότητα των ανταγωνιστικών διαγωνισμών με τη χαμηλότερη τιμή.
	Αδράνεια	Συντηρητισμός στην οικοδομική βιομηχανία- έλλειψη πίεσης μετασχηματισμού, αποστροφή στην αλλαγή, εξάρτηση από την πορεία, περιορισμένη σκέψη
	Περιορισμένη ορθολογικότητα	Αντί να λαμβάνονται με βάση, για παράδειγμα, την τέλεια πληροφόρηση και τον πλήρη ορθολογισμό, οι αποφάσεις συχνά λαμβάνονται σε περιορισμένα περιβάλλοντα που οδηγούν σε περιορισμένες, ή περιορισμένες, αποφάσεις, δηλαδή μη βέλτιστες από μια πλήρως ορθολογική άποψη, δηλαδή χρησιμοποιείται ο κανόνας του αντίχειρα.
Τεχνικά		Αποκλίσεις μεταξύ προβλεπόμενης και πραγματικής εξοικονόμησης, π.χ. λόγω κενών μόνωσης
Οικονομικά	Η ενεργειακή αποδοτικότητα αποτιμάται λιγότερο από το κόστος της επένδυσης	Η ενεργειακή αποδοτικότητα του κτιρίου θεωρείται λιγότερο σημαντική από την επενδυτική του αξία

	Υψηλό κόστος επένδυσης και καμία προοπτική LCC	Πριμοδότηση αρχικού κόστους, υψηλό κόστος, μακρύς χρόνος απόσβεσης, έλλειψη προοπτικής κύκλου ζωής και έλλειψη οικονομικών κινήτρων
	Εξωτερικός κίνδυνος	Η επιτυχία της εξοικονόμησης ενέργειας εξαρτάται από την τιμή της ενέργειας

Στο πλαίσιο της συμφωνίας του Παρισιού [8], η Ευρωπαϊκή Ένωση (ΕΕ) πρέπει να καθορίσει με σαφήνεια τους στόχους απαλλαγής από τις εκπομπές του διοξειδίου του άνθρακα μέχρι το 2050 [15]. Ωστόσο, το κατά πόσο οι στόχοι αυτοί είναι εφικτοί είναι ακόμη υπό συζήτηση. Τα σχέδια δράσης για την ενέργεια και το κλίμα που έχει προτείνει η Ευρωπαϊκή Επιτροπή ποικίλουν. Μία έρευνα με τη χρήση μιας ειδικής διαδικασίας μοντελοποίησης στο πλαίσιο της πρωτοβουλίας SET-Nav εντοπίζει τέσσερις διακριτούς δρόμους που οδηγούν σε ένα καθαρό, ασφαλές και αποδοτικό ενεργειακό σύστημα. Ο βαθμός συνεργασίας (δηλαδή, συνεργασία έναντι περιχαράκωσης) και ο βαθμός αποκέντρωσης (δηλαδή, αποκέντρωση έναντι εξάρτησης από τη διαδρομή) είναι οι δύο κύριες αβεβαιότητες που διαμορφώνουν τους δρόμους SET-Nav. Μέχρι το 2050, και οι τέσσερις διαδρομές μειώνουν τις εκπομπές κατά 85-95% [9].

Σύμφωνα με πρόσφατες μελέτες [10], [11] αλλά και ανακοινώσεις η Ευρωπαϊκή Ένωση συνεχώς ανανεώνει τους στόχους της και δίνει μεγάλη έμφαση στην ενεργειακή αποδοτικότητα όπως φαίνεται και από τις πρόσφατες αναθεωρήσεις της οδηγίας για την ενεργειακή απόδοση των κτιρίων (Energy Performance of Buildings Directive) [12] και της οδηγίας για την ενεργειακή αποδοτικότητα (Energy Efficiency Directive)[13].

Στις βασικές διατάξεις της οδηγίας για την ενεργειακή αποδοτικότητα περιλαμβάνεται ο καθορισμός δεσμευτικού στόχου της ΕΕ για τη μείωση της τελικής κατανάλωσης ενέργειας κατά 11,7% έως το 2030, ενώ τα κράτη μέλη υποχρεούνται να καθορίσουν ενδεικτικές εθνικές συνεισφορές. Οι ετήσιοι στόχοι εξοικονόμησης ενέργειας αυξάνονται σταδιακά, με έμφαση στους ευάλωτους πελάτες και την κοινωνική στέγαση. Η οδηγία επιβάλλει στόχο ετήσιας μείωσης της κατανάλωσης ενέργειας κατά 1,9% για τον δημόσιο τομέα και επεκτείνει τις υποχρεώσεις ανακαίνισης σε όλα τα επίπεδα της δημόσιας διοίκησης. Οι επιχειρήσεις υποχρεούνται να εφαρμόζουν συστήματα διαχείρισης ενέργειας ή να διενεργούν ενεργειακούς ελέγχους με βάση τα επίπεδα κατανάλωσης ενέργειας. Επιπλέον, τα κέντρα δεδομένων υποχρεούνται να παρακολουθούν την ενεργειακή απόδοση, με τα δεδομένα να συλλέγονται σε μια βάση δεδομένων σε επίπεδο ΕΕ. Προωθούνται τοπικά σχέδια θέρμανσης και ψύξης, παράλληλα με μέτρα για την αύξηση της ενεργειακής αποδοτικότητας στην παροχή θερμότητας ή ψύξης, συμπεριλαμβανομένης της τηλεθέρμανσης. Συνολικά, η οδηγία δίνει έμφαση στην αρχή “Energy Efficiency First” και αποσκοπεί στην επιτάχυνση της προόδου προς την επίτευξη των στόχων ενεργειακής αποδοτικότητας της ΕΕ.

Όσον αφορά την αναθεωρημένη οδηγία για την ενεργειακή απόδοση των κτιρίων (EPBD), εκείνη αποσκοπεί στην επιτάχυνση των ανακαινίσεων των κτιρίων, στη βελτίωση της ποιότητας του αέρα και στη στήριξη των υποδομών βιώσιμης κινητικότητας σε όλες τις χώρες της ΕΕ. Αναγνωρίζοντας τις εθνικές διαφορές, επιτρέπει στις κυβερνήσεις να προσαρμόζουν τα μέτρα ανακαίνισης στα ειδικά τους πλαίσια, με εξαιρέσεις για ορισμένες κατηγορίες κτιρίων. Επίσης, είναι σημαντικό ότι ενισχύει τη στοχευμένη χρηματοδότηση για επενδύσεις σε κτίρια, την καταπολέμηση της ενεργειακής φτώχειας και τη διασφάλιση των δικαιωμάτων των ενοικιαστών. Ευθυγραμμισμένη με την ευρωπαϊκή Πράσινη Συμφωνία (Green Paper), η οδηγία στοχεύει στη μείωση των εκπομπών στον κτιριακό τομέα κατά 60% έως το 2030 και στην κλιματική ουδετερότητα έως το 2050. Συγχρόνως, συμπληρώνει άλλες πολιτικές της Πράσινης Συμφωνίας και εισάγει μέτρα όπως ελάχιστα πρότυπα ενεργειακής αποδοτικότητας, εθνικές τροχιές για τη μείωση της ενέργειας και πρότυπα για τα νέα κτίρια που πρέπει να είναι μηδενικών εκπομπών. Επιπλέον, προωθεί την αξιοπιστία των πιστοποιητικών ενεργειακής απόδοσης, την ψηφιοποίηση και εισάγει έννοιες όπως η βαθιά ανακαίνιση (deep renovation) και τα διαβατήρια ανακαίνισης κτιρίων (building renovation passports). Η οδηγία επιβάλλει τέλος νέα κτίρια με δυνατότητα ηλιακής ενέργειας, σταδιακή κατάργηση των λεβήτων που λειτουργούν με ορυκτά καύσιμα, υπηρεσίες μιας στάσης (one-stop shops) για ενεργειακές ανακαινίσεις κτιρίων και προωθεί τον εκσυγχρονισμό και την ενσωμάτωση των ενεργειακών συστημάτων.

Οι δημόσιες αρχές και οι διεθνείς οργανισμοί ανησυχούν πολύ για το πώς η κλιματική αλλαγή μπορεί να επηρεάσει τη συμπεριφορά των επιχειρήσεων, των επενδυτών, των καταναλωτών και άλλων ενδιαφερόμενων μερών, προκειμένου να αντιμετωπίσουν αποτελεσματικά το πρόβλημα αυτό. Τα σχέδια ανάκαμψης πρέπει να δρομολογήσουν την απαλλαγή από τις ανθρακούχες εκπομπές και την πράσινη μετάβαση για την αντιμετώπιση της μετα-πανδημικής κλιματικής έκτακτης ανάγκης. Αυτό περιλαμβάνει την πραγματοποίηση στοχευμένων επενδύσεων και την προσφορά μιας πιο προσαρμοστικής δομής για τις πηγές ενέργειας σε διάφορες τοποθεσίες, ώστε να ικανοποιηθεί η ανάγκη για μια συστημική στροφή προς μια πιο βιώσιμη οικονομία που θα ωφελεί τους ανθρώπους και το περιβάλλον. Μια έρευνα που πραγματοποιήθηκε μελετώντας τα έτη 1990-2018 έδειξε ότι οι επαγγελματίες ενδιαφέρονται όλο και περισσότερο για την αλλαγή του μοντέλου κατανάλωσης ενέργειας που προκαλείται από την αύξηση του ποσοστού της πράσινης ενέργειας, με αρκετά μεγάλες διαφορές μεταξύ των 42 ευρωπαϊκών χωρών που εξετάστηκαν. Διαπιστώθηκαν διαφορές πολιτικής με βάση το εμπορικό άνοιγμα, το ΑΕΠ/κάτοικο, τον δείκτη παγκοσμιοποίησης και το ποσοστό των ανανεώσιμων πηγών ενέργειας στη συνολική κατανάλωση ενέργειας [14].

Παρόλα αυτά χρησιμοποιώντας ανανεώσιμες πηγές ενέργειας όπως ανεμογεννήτριες, φωτοβολταϊκά τα οποία δεν εγκαθίστανται απαραίτητα σε εδάφη αλλά και στις στέγες κτιρίων, ενδεχομένως πλήρης εξηλεκτρισμός του τομέα των μεταφορών, η χρήση μηχανικής μάθησης στις ενεργειακές ανακαινίσεις είναι μερικά

από τα μέτρα τα οποία φαντάζουν τον στόχο της απαλλαγής από τις εκπομπές του διοξειδίου του άνθρακα εφικτό για την Ευρωπαϊκή Ένωση. Επιπλέον, η ενεργειακή μετάβαση στην ΕΕ θα έχει σημαντικό θετικό αντίκτυπο στις οικονομικές, περιβαλλοντικές και κοινωνικές πτυχές της βιωσιμότητας, με τη δημιουργία περισσότερων από 1,5 εκατομμυρίου νέων θέσεων εργασίας σε ολόκληρη την ΕΕ τα επόμενα 30 χρόνια [15].

Δεδομένου ότι η τρέχουσα εργασία επικεντρώνεται στο πώς η μηχανική μάθηση μπορεί να χρησιμοποιηθεί για την επίτευξη των προαναφερθέντων στόχων της ενεργειακής μετάβασης, στο επόμενο κεφάλαιο αναλύονται σχετικές προσπάθειες χρήσης της μηχανικής μάθησης σε ενεργειακές εφαρμογές σε ερευνητικό και πρακτικό επίπεδο.

ΚΕΦΑΛΑΙΟ 3

Μηχανική Μάθηση στην Ενέργεια

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

3 Μηχανική Μάθηση στην Ενέργεια

Η μηχανική μάθηση αποτελεί ένα αναπτυσσόμενο πεδίο της επιστήμης που έχει αποκτήσει ολοένα και μεγαλύτερη σημασία στον τομέα της ενεργειακής αποδοτικότητας. Αναδεικνύει νέες δυνατότητες και προσεγγίσεις για την αντιμετώπιση των προκλήσεων που σχετίζονται με την ενεργειακή αποδοτικότητα, επιτρέποντας την ανάπτυξη αυτόνομων συστημάτων και την πρόβλεψη των ενεργειακών αναγκών με ακρίβεια. Η εφαρμογή της μηχανικής μάθησης στα ενεργειακά συστήματα δημιουργεί προοπτικές για τη βελτίωση της απόδοσης και τη μείωση της κατανάλωσης ενέργειας. Ειδικότερα, η ενσωμάτωση της μηχανικής μάθησης στον κτιριακό τομέα, τόσο σε οικιακό επίπεδο όσο και εκτός αυτού, ανοίγει νέες προοπτικές για την ανάπτυξη έξυπνων και βιώσιμων κτιρίων. Στο πλαίσιο αυτό, η παρούσα ενότητα εξετάζει τη σημασία της μηχανικής μάθησης για την ενεργειακή αποδοτικότητα, καθώς και την εφαρμογή της τόσο εντός όσο και εκτός του κτιριακού τομέα, προσφέροντας ένα επίκαιρο και πολύτιμο πλαίσιο για την κατανόηση της συμβολής της στη βελτίωση της ενεργειακής απόδοσης.

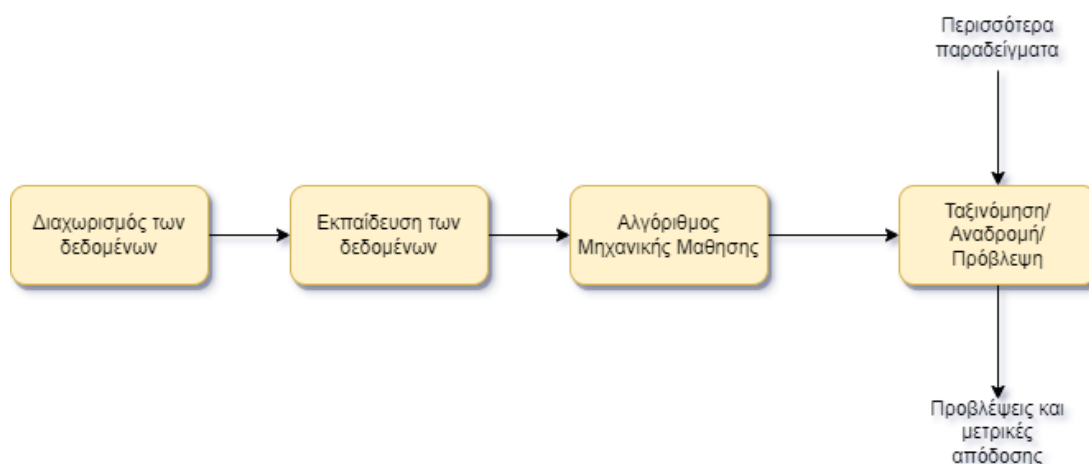
3.1 Μηχανική Μάθηση: Βασικές Έννοιες

Η Μηχανική Μάθηση (Machine Learning-ML) είναι ένα υποπεδίο υπολογιστικών αλγορίθμων που στοχεύει στο να μιμηθεί την ανθρώπινη νοημοσύνη λαμβάνοντας πληροφορίες από το περιβάλλον [16]. Στην ολοκαίνουργια εποχή των «μεγάλων δεδομένων», θεωρείται ως βασικό εργαλείο. Οι τεχνικές μηχανικής μάθησης έχουν χρησιμοποιηθεί αποτελεσματικά σε διάφορους τομείς, όπως η υπολογιστική βιολογία [17], τα οικονομικά [18], η ψυχαγωγία [19], η αναγνώριση προτύπων [20], η όραση υπολογιστών [21], η αεροδιαστημική μηχανική [22] και οι βιοϊατρικές και ιατρικές εφαρμογές [23].

Πολλές διαφορετικές τεχνολογίες χρησιμοποιούνται στη μηχανική μάθηση για την εξαγωγή συμπερασμάτων από τα δεδομένα. Η ανάλυση κύριων συνιστωσών [24], οι μέθοδοι ομαδοποίησης [25], οι μηχανές διανυσμάτων υποστήριξης [26], τα τεχνητά νευρωνικά δίκτυα [27] και πολλά άλλα είναι παραδείγματα τέτοιων τεχνολογιών. Η πρόκληση της άντλησης γνώσεων από τα δεδομένα συχνά δεν είναι απλή, αντίθετα απαιτεί τον προοδευτικό μετασχηματισμό των βασικών δεδομένων σε ολοένα και πληρέστερη πληροφορία μέσω μιας ποικιλίας φάσεων. Με βάση τα διαθέσιμα δεδομένα και τους στόχους που επιθυμούμε να πετύχουμε, κάθε εφαρμογή μηχανικής μάθησης απαιτεί συνήθως μια μοναδική σειρά διαδικασιών.

Τα μοντέλα μηχανικής μάθησης βασίζονται σε ένα συγκεκριμένο μοτίβο. Οι προσεγγίσεις της ML συχνά συνεπάγονται μια διαδικασία μάθησης με στόχο την εκτέλεση μιας εργασίας μέσω της μάθησης από την "εμπειρία" (η μάθηση επιτυγχάνεται μέσα από δεδομένα εκπαίδευσης). Στη μηχανική μάθηση, τα δεδομένα είναι μια σειρά παραδειγμάτων. Συνήθως, μια συλλογή χαρακτηριστικών, που μερικές φορές αναφέρονται ως χαρακτηριστικά ή μεταβλητές, περιγράφει καλύτερα μια μεμονωμένη περίπτωση. Ένα χαρακτηριστικό μπορεί να είναι αριθμητικό (ακέραιος, πραγματικός αριθμός κ.λπ.), ταξινομημένο (A+ ή B-), δυαδικό (δηλαδή 0 ή

1) ή ονομαστικό (απαρίθμηση). Μια μετρική απόδοσης που βελτιώνεται με την πάροδο του χρόνου με την εμπειρία χρησιμοποιείται για να μετρήσει πόσο καλά αποδίδει το μοντέλο μηχανικής μάθησης σε κάθε εργασία. Το εκπαιδευμένο μοντέλο μπορεί να χρησιμοποιηθεί για την κατηγοριοποίηση, την πρόβλεψη ή την ομαδοποίηση όταν ολοκληρωθεί η διαδικασία εκμάθησης. Ανάλογα με τον τύπο μάθησης (με επίβλεψη/χωρίς επίβλεψη), τα μοντέλα μάθησης (ταξινόμηση, παλινδρόμηση, ομαδοποίηση και μείωση διαστάσεων) ή τα μοντέλα μάθησης που χρησιμοποιούνται για την επίτευξη του επιλεγμένου στόχου, οι εργασίες ML συνήθως χωρίζονται σε ορισμένες μεγάλες κατηγορίες [28], [29], [30].



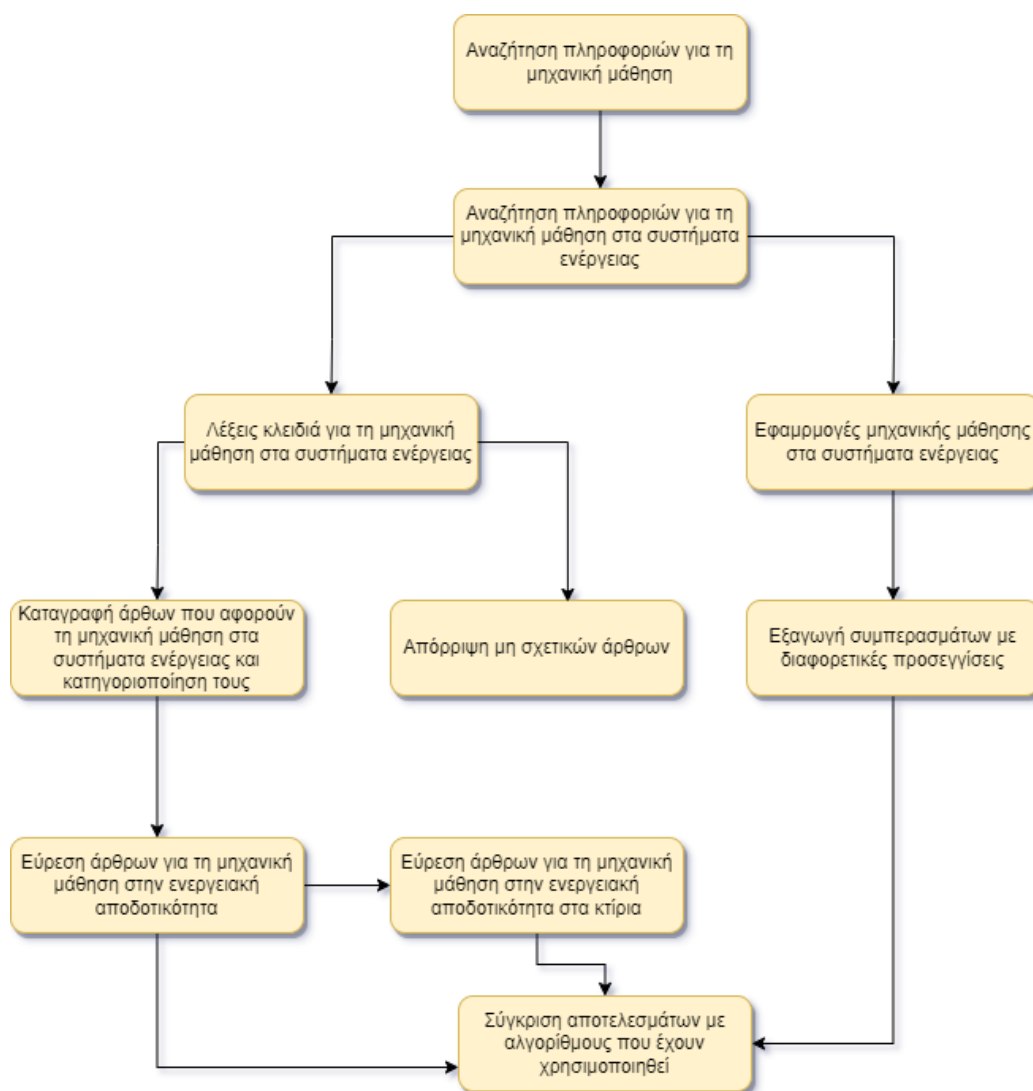
Διάγραμμα 1: Μηχανική Μάθηση

3.2 Μηχανική Μάθηση στην Ενέργεια: Επιστημονικά Άρθρα και Εφαρμογές

Έχοντας παρουσιάσει την εισαγωγή της μηχανικής μάθησης και των εφαρμογών της, οδηγούμαστε στο επιστημονικό ερώτημα της παρούσας διπλωματικής εργασίας, το οποίο δεν είναι άλλο από τη χρήση της μηχανικής μάθησης σε ενεργειακές εφαρμογές, με έμφαση στην ενεργειακή αποδοτικότητα των κτιρίων.

Σε αυτό το κεφάλαιο, σκοπός είναι να προσδιοριστούν οι πιθανές χρήσεις της μηχανικής μάθησης στα ενεργειακά συστήματα, οι οποίες στοχεύουν στο να ενισχυθεί η ενεργειακή αποδοτικότητα και να αυξηθεί η αποδοτικότητα της λειτουργίας διαφόρων συστημάτων. Για τον σκοπό αυτό, πραγματοποιείται βιβλιογραφική ανασκόπηση. Προκειμένου να μην παρουσιαστεί μόνο η θεωρητική προσέγγιση του συγκεκριμένου θέματος, αλλά και η πρακτική εφαρμογή της μηχανικής μάθησης σε εφαρμογές ενεργειακής απόδοσης, μελετήθηκαν σχετικά έργα που υλοποιήθηκαν από διάφορους οργανισμούς, όπως ερευνητικά ιδρύματα, καθώς και εταιρείες που ασχολούνται με την ενέργεια ή την τεχνολογία. Τέλος, για την καλύτερη εποπτεία της μηχανικής μάθησης στον ενεργειακό τομέα, συγκρίνουμε τα αποτελέσματα με έργα που έχουν ήδη υλοποιηθεί τόσο στον κτιριακό τομέα όσο και σε άλλες εφαρμογές ενεργειακής απόδοσης. Ο τρόπος με τον οποίο καταρτίστηκε η βιβλιογραφική ανασκόπηση παρουσιάζεται παρακάτω.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 2: Βιβλιογραφική ανασκόπηση

Παρακάτω παρουσιάζεται η εφαρμογή της μηχανικής μάθησης στον τομέα της ενέργειας, όπως διαπιστώθηκε μέσω της παρατήρησης και της ανάγνωσης διαφόρων επιστημονικών άρθρων. Στο παρελθόν έχουν πραγματοποιηθεί μελέτες σχετικά με τη χρήση και την εφαρμογή της μηχανικής μάθησης στον τομέα της ενέργειας [31] και της ενεργειακής αποδοτικότητας [32], κατά τις οποίες χρησιμοποιήθηκαν διάφοροι αλγόριθμοι. Η ανασκόπηση που ακολουθεί, βασίστηκε στις λέξεις-κλειδιά που δίνονται στον πίνακα 2. Οι όροι παρουσιάζονται στην αγγλική γλώσσα, δεδομένου ότι η ανασκόπηση βασίζεται στη διεθνή βιβλιογραφία.

Πίνακας 2: Λέξεις κλειδιά για την αναζήτηση επιστημονικών άρθρων

<u>Λέξεις Κλειδιά</u>
Machine learning, Artificial intelligence, Machine Learning Algorithms, Data-driven models, Knowledge-driven models, Regression, Optimization, Artificial neural networks, Neural networks, Predicting models, Deep Learning, Data analysis, Data mining
Energy management, Energy systems, Machine Learning in Energy systems, Barriers in energy renovations in buildings
Energy efficiency, Renewable energy systems, Energy benchmarking, Renewable energies
Energy Efficiency in public sector, Energy conservation, EU Energy transition goals (net zero by 2050), Energy efficiency directive, Energy performance of buildings directive, Sustainable development goals
Building energy efficiency, Smart home, Sustainability, Building energy consumption, Building load forecasting, Energy-efficient buildings, Smart grid

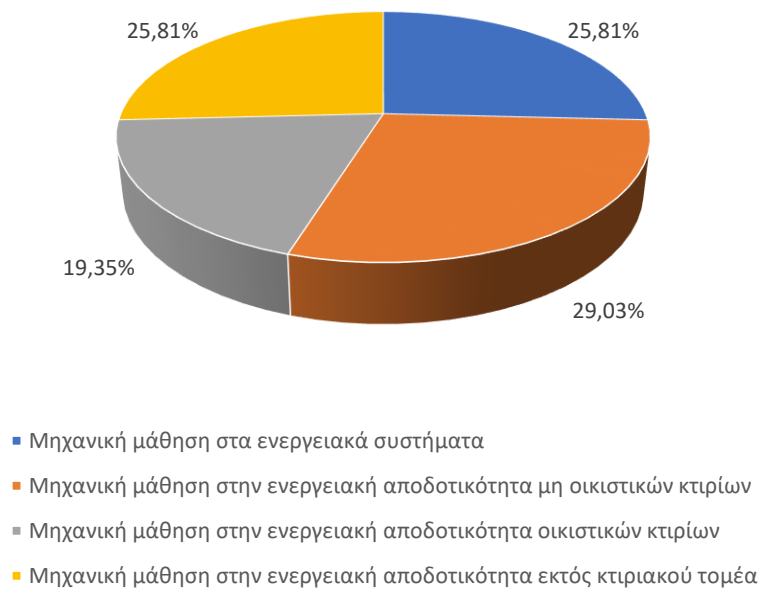
Μετά την έρευνά μας σε διάφορες επιστημονικές εργασίες, συνοψίσαμε τα άρθρα σε πέντε διαφορετικές κατηγορίες: εφαρμογές μηχανικής μάθησης γενικότερα, μηχανική μάθηση στα ενεργειακά συστήματα, μηχανική μάθηση στην ενεργειακή αποδοτικότητα το οποίο χωρίζεται σε δύο υποκατηγορίες, την μηχανική μάθηση στην ενεργειακή αποδοτικότητα κτιρίων, το οποίο με τη σειρά του χωρίζεται σε άλλες δύο υποκατηγορίες, τα μη οικιστικά κτίρια (σχολεία, νοσοκομεία, κυβερνητικά κτίρια κ.λπ.) και τα οικιστικά κτίρια στα οποία ανήκουν και τα έξυπνα σπίτια (smart homes), και την μηχανική μάθηση στην ενεργειακή αποδοτικότητα εκτός κτιριακού τομέα. Σε αυτό το σημείο είναι απαραίτητο να επισημανθεί ότι ορισμένα από τα παρακάτω άρθρα μπορεί να ανήκουν σε περισσότερες από μία κατηγορίες, αλλά έχουν ταξινομηθεί με βάση το κυρίαρχο θέμα τους. Στο παράρτημα 2 δίνονται σύντομες περιγραφές για τα πιο σημαντικά επιστημονικά άρθρα.

Πίνακας 3: Κατηγοριοποίηση επιστημονικών άρθρων

<u>Κατηγορίες</u>	<u>Επιστημονικά άρθρα</u>
Εφαρμογές μηχανικής μάθησης εκτός του τομέα της ενέργειας	[17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28], [30], [32], [59], [60], [61], [62], [63], [64], [65], [67], [68], [69], [70], [71], [72], [76]
Μηχανική μάθηση στα ενεργειακά συστήματα	[31], [35], [36], [37], [38], [33], [73], [76]
Μηχανική μάθηση στην ενεργειακή αποδοτικότητα μη οικιστικών κτιρίων	[36], [39], [41], [42], [40], [43], [44], [74], [77]
Μηχανική μάθηση στην ενεργειακή αποδοτικότητα οικιστικών κτιρίων	[46], [47], [48], [49], [50], [51]
Μηχανική μάθηση στην ενεργειακή αποδοτικότητα εκτός κτιριακού τομέα	[52], [53], [54], [55], [56], [57], [58], [75]

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Η κατηγοριοποίηση των επιστημονικών άρθρων φαίνεται και στο παρακάτω διάγραμμα.



Διάγραμμα 3: Κατηγοριοποίηση επιστημονικών άρθρων για τη μηχανική μάθηση στην ενέργεια

3.2.1 Μηχανική Μάθηση στα Ενεργειακά Συστήματα

Η άνοδος της οικονομικής ανάπτυξης σε συνδυασμό με τη μεταστροφή της ανθρώπινης συμπεριφοράς που αναζητά την άνεση οδήγησε σε αύξηση της ζήτησης ενέργειας. Η βελτιστοποίηση των ενεργειακών συστημάτων είναι ιδανική λύση επειδή οι ενεργειακοί πόροι είναι περιορισμένοι και πρέπει να προστατευθούν για τις μελλοντικές γενιές. Παρόλα αυτά, αυτό δεν είναι καθόλου εύκολο εγχείρημα δεδομένης της πολυπλοκότητας των ολοκληρωμένων ενεργειακών συστημάτων. Εδώ είναι που ο προσδιορισμός του ιδανικού σημείου για την προσφορά και τη ζήτηση μπορεί να βελτιωθεί σημαντικά με τη λήψη αποφάσεων με τη βοήθεια υπολογιστή. Τον 20ό αιώνα, αναπτύχθηκαν οι ιδέες της τεχνητής νοημοσύνης (AI) και της μηχανικής μάθησης (ML) για να μπορέσουν οι υπολογιστές να μιμηθούν τις ανθρώπινες διαδικασίες λήψης αποφάσεων και μάθησης. Έκτοτε, η τεχνητή νοημοσύνη και η εξόρυξη δεδομένων έχουν αποκτήσει όλο και μεγαλύτερη σημασία σε ένα ευρύ φάσμα επιστημονικών τομέων. Φυσικά, ένας τομέας όπου χρησιμοποιείται η τεχνητή νοημοσύνη και η μηχανική μάθηση είναι ο τομέας της ενέργειας.

Τα ενεργειακά συστήματα περιλαμβάνουν δομές όπως μηχανήματα, κατασκευές, οικοσυστήματα όπως η χλωρίδα, ακόμη και συστήματα έξυπνης ενέργειας (όπως τα έξυπνα ηλεκτρικά δίκτυα). Με άλλα λόγια, ως ενεργειακό σύστημα ορίζεται κάθε σύστημα που απαιτεί ενέργεια για να λειτουργήσει, να διατηρήσει μια δεδομένη κατάσταση ή να μετακινήσει ενέργεια μεταξύ σημείων [31]. Η ύπαρξη ενός έξυπνου συστήματος διαχείρισης που μπορεί να προβλέψει μελλοντικά γεγονότα για να λειτουργήσει τα στοιχεία του δικτύου στο μέγιστο δυνατό βαθμό ή να ανταποκριθεί σε απότομες αλλαγές στις εισροές (όπως η αύξηση ή η μείωση της ζήτησης) μπορεί να είναι πολύ χρήσιμη όταν πρόκειται για τη μεταφορά ή την κατανάλωση ενέργειας. Για παράδειγμα, η βραχυπρόθεσμη πρόβλεψη της ποσότητας ηλεκτρικής ενέργειας που απαιτείται να παραχθεί από σταθμούς παραγωγής ενέργειας από ορυκτά καύσιμα και της ποσότητας καυσίμου που χρησιμοποιούν, εκτιμώντας τη ζήτηση για το δίκτυο και τη διαθεσιμότητα ανανεώσιμων πηγών ενέργειας, όπως η ηλιακή ή η αιολική. Αυτές οι βραχυπρόθεσμες προβλέψεις απαιτούν δυναμικά μοντέλα για τη λήψη αποφάσεων. Η τεχνητή νοημοσύνη έχει τη δυνατότητα να αποτελέσει ένα πολύ χρήσιμο εργαλείο για την προσομοίωση της ανθρώπινης λήψης αποφάσεων και τη διαχείριση έξυπνων ενεργειακών συστημάτων χωρίς την παρέμβαση του χειριστή. Με τη μηχανική μάθηση, οι υπολογιστές μπορούν να εκπαιδευτούν ώστε να γίνονται πιο ακριβείς και να λαμβάνουν καλύτερες αποφάσεις σε σημείο που, με αρκετά δεδομένα, να μπορούν εύκολα να ξεπεράσουν τους ανθρώπους [33], [34].

Πολλά από τα άρθρα που εξετάστηκαν πραγματοποιήσαν μια ολοκληρωμένη ανασκόπηση του τρόπου με τον οποίο η τεχνητή νοημοσύνη και η μηχανική μάθηση μας βοήθησαν να κατασκευάσουμε ενεργειακά συστήματα τα τελευταία 20 ή περισσότερα χρόνια [35]. Η πλειονότητά τους χρησιμοποίησε δύο διαφορετικές προσεγγίσεις, την καθοδηγούμενη από δεδομένα [36] και την καθοδηγούμενη από γνώση [37]. Η πρώτη βασίζεται σε πραγματικά δεδομένα που αντλήσαμε από

3.2.2 *Μηχανική Μάθηση στην Ενεργειακή Αποδοτικότητα Μη Οικιστικών Κτιρίων*

Επειδή τα κτίρια συμβάλλουν σημαντικά στην παγκόσμια κατανάλωση ενέργειας και στις εκπομπές αερίων του θερμοκηπίου, πρέπει να είναι ενεργειακά αποδοτικά και βιώσιμα. Προκειμένου να επιλυθούν τα ζητήματα της ενεργειακής απόδοσης των κτιρίων και να αντιμετωπιστούν οι σημερινές ανησυχίες για την ανθρώπινη άνεση, την επέκταση της αστικοποίησης και την επακόλουθη αύξηση της κατανάλωσης ενέργειας, η μοντελοποίηση και η πρόβλεψη της ενεργειακής κατανάλωσης των κτιρίων είναι ζωτικής σημασίας. Ευέλικτες λύσεις για την πρόβλεψη της ενεργειακής κατανάλωσης των κτιρίων παρέχονται από μεθοδολογίες βασισμένες στα δεδομένα, οι οποίες εξαρτώνται από τη μηχανική μάθηση και την ανάλυση δεδομένων στο πλαίσιο ολοκληρωμένων έξυπνων υποδομών. Στην παρούσα μελέτη παρέχεται μια επισκόπηση έργων που δημιουργούν μοντέλα με βάση τα δεδομένα για εφαρμογές μεγάλης κλίμακας [38][36]. Ορισμένες από αυτές τις εφαρμογές μηχανικής μάθησης χρησιμοποιούν διαφορετικές προσεγγίσεις, όπως η πρόβλεψη Random Forest [39], για την πρόβλεψη της ενεργειακής κατανάλωσης και τον εντοπισμό των μοτίβων που τείνουν να εμφανιστούν σε κάθε κτίριο, ενώ άλλες χρησιμοποιούν προσεγγίσεις ML, όπως το τεχνητό νευρωνικό δίκτυο, η μηχανή διανυσμάτων υποστήριξης, οι παλινδρομήσεις με βάση τη Γκαουσιανή και η ομαδοποίηση, οι οποίες έχουν συνήθως εφαρμοστεί στην πρόβλεψη και τη βελτίωση της ενεργειακής απόδοσης των κτιρίων, για να συνεπάγονται ενεργειακές προβλέψεις. [40]. Άλλοι χρησιμοποιούν διαφορετικούς ταξινομητές για να δουν την κατανάλωση μονοκατοικιών σε μικρές κατοικημένες περιοχές [41]. Επιπλέον, ορισμένοι από τους επιστήμονες προσεγγίζουν την ενεργειακή αποδοτικότητα σκεπτόμενοι το φορτίο θέρμανσης που έχει ένα κτίριο, το οποίο καταλήγει στο σύστημα εξαερισμού, στον κλιματισμό, στην επαναληπτική θέρμανση [42]. Κάθε αλγόριθμος βασισμένος σε δεδομένα που έχει χρησιμοποιηθεί για τη βελτίωση της ενεργειακής απόδοσης ενός κτιρίου χρησιμοποιεί μετρήσεις όπως το RMSE, το R-squared και το προσαρμοσμένο R-squared για να καταλήξει σε ένα ασφαλές και πειστικό αποτέλεσμα.

Έχουν διεξαχθεί πολλές μελέτες για να διευκρινιστεί πώς οι αλγόριθμοι ML μπορούν να υποστηρίξουν την ενεργειακή αποδοτικότητα στα κτίρια, και όλες μοιράζονται ένα κοινό δίλημμα. Οι μελέτες αυτές υποστηρίζουν ότι για να κάνουμε μια επιτυχημένη αξιολόγηση δεν πρέπει να υιοθετήσουμε έναν αλγόριθμο ML χωρίς αιτιολογημένο σκεπτικό, αλλά πρέπει να δοκιμάσουμε διαφορετικούς αλγορίθμους με βάση δεδομένα από διαφορετικές πηγές δεδομένων, προκειμένου να επιλέξουμε τον αλγόριθμο με τις καλύτερες επιδόσεις [43]. Επιπλέον οι Tien et al. (2022) προσπάθησαν να καταγράψουν τα σημερινά εργαλεία Μηχανικής Μάθησης που έχουμε στα χέρια μας, προς την κατεύθυνση της επιλογής του καλύτερου μοντέλου μηχανικής μάθησης και βαθιάς μάθησης. Σύμφωνα με τη βιβλιογραφία, οι προηγούμενες έρευνες εξέτασαν μια μεγάλη ποικιλία πεδίων (από ένα στοιχείο HVAC έως αστικές περιοχές) και χρονικών κλιμάκων (από ένα λεπτό έως ένα έτος). Επίσης, η χρήση ενός ποικίλου συνόλου κριτηρίων αξιολόγησης καθιστά το εγχείρημα πιο δύσκολο και χρειάζεται περισσότερες εξελίξεις και πιο λεπτομερείς απαιτήσεις ειδικά για ένα δομημένο περιβάλλον. Ενώ η μηχανική και η βαθιά μάθηση έχουν

εφαρμοστεί με επιτυχία στην έρευνα για την ενεργειακή αποδοτικότητα των κτιρίων, οι περισσότερες μελέτες βρίσκονται ακόμη σε πειραματικό ή δοκιμαστικό στάδιο και υπάρχουν μόνο λίγες μελέτες που εφάρμοσαν στρατηγικές μηχανικής και βαθιάς μάθησης σε πραγματικά κτίρια και διεξήγαγαν αξιολογήσεις μετά τη χρήση τους [44].

Αρκετοί ενδιαφερόμενοι φορείς θεώρησαν ότι η ενσωμάτωση της χρηματοδότησης της ενεργειακής απόδοσης αποτελεί ύψιστη προτεραιότητα κατά τα τελευταία δέκα χρόνια. Μία έρευνα η οποία πραγματοποιήθηκε από το Εργαστήριο Συστημάτων Αποφάσεων και Διοίκησης του Εθνικού Μετσόβιου Πολυτεχνείου στην Ελλάδα προτάσσει ένα υγιές μεθοδολογικό πλαίσιο για την υποστήριξη της διαδικασίας χρηματοδότησης των επενδύσεων ενεργειακής απόδοσης. Αυτό επιτυγχάνεται λαμβάνοντας υπόψη μια σειρά από κρίσιμους παράγοντες που είναι απαραίτητοι για τη μακροπρόθεσμη βιωσιμότητα αυτών των πρωτοβουλιών και την μείωση του επενδυτικού κινδύνου. Χρησιμοποιώντας ένα αξιόπιστο σύνολο δεδομένων ενεργειακά αποδοτικών έργων από πολλές πόλεις της Λετονίας, η προτεινόμενη προσέγγιση εφαρμόστηκε για τη μεγιστοποίηση της πραγματικής απόδοσης της επένδυσης. Από την εφαρμογή της μεθοδολογίας προέκυψε ένα σχέδιο χρηματοδότησης που μειώνει το κόστος των έργων ενεργειακής απόδοσης κατά 15%, ενώ επιτυγχάνεται περίπου η ίδια εξοικονόμηση ενέργειας [45].

3.2.3 Μηχανική Μάθηση στην Ενεργειακή Αποδοτικότητα Οικιστικών Κτιρίων

Η ανάγκη εξισορρόπησης της άνεσης των κατοίκων με την ενεργειακή αποδοτικότητα μας οδήγησε στη λύση των έξυπνων ενεργειακά αποδοτικών κατοικιών. Τα έξυπνα σπίτια που εξασφαλίζουν τη συμβατότητα και τη συνεργασία μεταξύ των διαφόρων συσκευών, αποτρέποντας την ακούσια κατανάλωση ενέργειας, ήταν το κύριο μέρος των περισσότερων άρθρων που περιλάμβαναν την ενεργειακή αποδοτικότητα των κτιρίων [46], [47]. Χρησιμοποίησαν μοντέλα επεξεργασίας δεδομένων προκειμένου να αξιολογήσουν και να προτείνουν στρατηγικές εξοικονόμησης ενέργειας που περιλαμβάνουν τόσο την υποδομή του κτιρίου όσο και την κατανομή της κατανάλωσης ενέργειας στο εσωτερικό του. Μία από αυτές τις μελέτες [47] προτείνει ένα έξυπνο σύστημα ενεργειακά αποδοτικού οικιακού αυτοματισμού που μπορεί να έχει πρόσβαση και να λειτουργεί τον οικιακό εξοπλισμό από οποιοδήποτε σημείο του πλανήτη. Για το σύστημα αυτό, μια μονάδα συνδεσιμότητας στο Διαδίκτυο συνδέεται με την κύρια μονάδα τροφοδοσίας του οικιακού συστήματος και μπορεί να έχει πρόσβαση μέσω του Διαδικτύου. Η στατική διεύθυνση IP είναι απαραίτητη για την ενεργοποίηση της ασύρματης συνδεσιμότητας. Ο οικιακός αυτοματισμός βασίζεται σε μια πολυτροπική εφαρμογή που μπορεί να λειτουργεί μέσω της εντολής αναγνώρισης ομιλίας του χρήστη μέσω του Google Assistant ή μέσω μιας διαδικτυακής εφαρμογής.

Μια άλλη προσέγγιση για έξυπνα ενεργειακά αποδοτικά σπίτια είναι η χρήση διαφόρων ειδών αισθητήρων στα σπίτια μας. Οι Lee et al. (2014), διεξήγαγαν μια έρευνα όπου πρώτα συνέλεξαν και ανέλυσαν τα δεδομένα από διάφορους

αισθητήρες στο σπίτι και στη συνέχεια, αφού παρουσίασαν την αρχιτεκτονική του δικτύου μέσω διαφόρων συσκευών και διαφορετικών τύπων δικτύων, προχώρησαν σε μέτρα εξοικονόμησης ενέργειας, δημιουργώντας παράλληλα ένα φιλικό προς το χρήστη περιβάλλον [48]. Στην πραγματικότητα, λόγω της συμφωνίας του Παρισιού για την κλιματική αλλαγή (2015), είναι ζωτικής σημασίας για πολλά έθνη να εφαρμόσουν έξυπνη τεχνολογία με στόχο τη μείωση της χρήσης ενέργειας. Η Σιγκαπούρη είναι ένα από αυτά. Γι' αυτό το λόγο επιστήμονες από τη Σιγκαπούρη διατύπωσαν ένα έγγραφο [49] όπου επικεντρώθηκαν στις αντιλήψεις των οικογενειών της Σιγκαπούρης για την έξυπνη τεχνολογία και την εφαρμογή της για την εξοικονόμηση ενέργειας. Το άρθρο περιλάμβανε τα εξής θέματα: (1) κατανάλωση ενέργειας στα νοικοκυριά της Σιγκαπούρης, (2) δημόσιες πρωτοβουλίες και πολιτικές εξοικονόμησης ενέργειας, (3) χρήση τεχνολογίας εξοικονόμησης ενέργειας και (4) αντίληψη των νοικοκυριών για την εξοικονόμηση ενέργειας στα έξυπνα σπίτια. Επιπλέον, αναλύθηκαν τρεις μελέτες περιπτώσεων που σχετίζονται με έξυπνα σπίτια και έξυπνες τεχνολογίες, καθώς και μια συζήτηση σχετικά με την ωριμότητα των υφιστάμενων λύσεων.

Η κατανάλωση ενέργειας από οικιακές συσκευές θεωρείται ένα από τα σημαντικότερα θέματα που απαιτούν την προσοχή των ερευνητών. Η εξοικονόμηση ενέργειας μπορεί να επιτευχθεί με επιτυχία με τη σωστή διαχείριση της κατανομής ενέργειας για τις οικιακές συσκευές με βάση τη δραστηριότητα του χρήστη. Η αναγνώριση των ανθρώπινων δραστηριοτήτων και η παροχή ενέργειας για συσκευές που συνδέονται με αυτές τις δραστηριότητες μπορεί να οδηγήσει σε καλή χρήση και εξοικονόμηση ενέργειας. Το σημερινό σύστημα χρησιμοποιεί διάφορους αισθητήρες και υπολογιστές για την παρακολούθηση των ανθρώπινων συμπεριφορών, το οποίο μπορεί να είναι περίπλοκο για τους χρήστες. Γι' αυτό προτείνεται μια απλή προσέγγιση που βασίζεται στο Διαδίκτυο των πραγμάτων (IoT) για την αναγνώριση της ανθρώπινης δραστηριότητας με τη χρήση επεξεργασίας εικόνας [50]. Πρόκειται για μια τεχνική διαχείρισης ενέργειας σε πραγματικό χρόνο που χρησιμοποιεί την επικοινωνία μεταξύ μηχανών.

Ένα πρόβλημα που προκύπτει με το IoT και τα έξυπνα σπίτια σχετίζεται με την προστασία των προσωπικών δεδομένων. Αυτός είναι ο λόγος για τον οποίο πολλοί ερευνητές παράλληλα με τη μελέτη των ήδη υπάρχουσών τεχνολογιών, όπως τα πρωτόκολλα αυτοματισμού και τις έξυπνες συσκευές που βασίζονται στο IoT, εστιάζουν επίσης στο μοντέλο ασφάλειας δεδομένων, την κρυπτογράφηση δεδομένων και τις πρακτικές προσεγγίσεις τους στα έξυπνα σπίτια [51].

3.2.4 Μηχανική Μάθηση στην Ενεργειακή Αποδοτικότητα εκτός Κτιριακού Τομέα

Ο στόχος της καθαρής και ανανεώσιμης ενέργειας μας έδωσε την ευκαιρία να εξερευνήσουμε νέες μεθόδους στην εκτίμηση της ισχύος και στη χρήση των ήδη υπάρχουσών ανανεώσιμων πηγών. Το IoT (Internet of Things) χρησιμοποιείται σε πολλές μελέτες για την παρακολούθηση δεδομένων σε πραγματικό χρόνο και την

οπτικοποίησή τους με σκοπό την πρόληψη της απώλειας ενέργειας. Στο [52] (Demir, 2023) χρησιμοποιούνται δέντρα αποφάσεων, ο αλγόριθμος k-κοντινότερου γείτονα, ο αλγόριθμος μηχανής διανυσμάτων υποστήριξης και η γραμμική παλινδρόμηση για να παρουσιαστεί ένα σύστημα τηλεπαρακολούθησης για ηλιακά φωτοβολταϊκά ενεργειακά συστήματα μαζί με έναν εκτιμητή φωτοβολταϊκής ισχύος βασισμένο στη μηχανική μάθηση. Η παραγωγή ηλιακής ενέργειας είναι το κύριο θέμα ενός ακόμα άρθρου (Acuña Acurio et al., 2023), αυτή τη φορά όμως με τη χρήση ενός τροποποιημένου νευρωνικού δικτύου VGG-16 για την αποφυγή των ψευδών εισχωρήσεων δεδομένων από κυβερνοεπιθέσεις για τη δημιουργία ενός μοντέλου που δεν χρειάζεται ανθρώπινη επίβλεψη [53].

Οι Fiona Burlig, Christopher Knittel, David Rapson, Mar Reguant και Catherine Wolfram αξιολόγησαν τον αντίκτυπο των ανακαινίσεων ενεργειακής απόδοσης σε σχολεία K-12 στην Καλιφόρνια, χρησιμοποιώντας δεδομένα πάνελ υψηλής συχνότητας σχετικά με τη χρήση ενέργειας. Χρησιμοποιώντας μια μέθοδο σταθερών επιδράσεων πάνελ, ανακάλυψαν ότι, ανάλογα με τις προδιαγραφές και την αντιμετώπιση των ακραίων τιμών, οι βελτιώσεις αυτές παράγουν μεταξύ 12% και 86% της προβλεπόμενης εξοικονόμησης. Εκτίμησαν ένα στενότερο εύρος 52%-98%, με κεντρική εκτίμηση 60%, χρησιμοποιώντας μηχανική μάθηση για να καθοδηγήσουν την απόφαση προδιαγραφών τους [54]. Μια άλλη εφαρμογή της ML για τη διαχείριση της ενεργειακής απόδοσης πραγματοποιήθηκε στην Κροατία, όπου χρησιμοποιήθηκαν βαθιά νευρωνικά δίκτυα, δέντρα παλινδρόμησης Rpart και Random Forest με διαδικασίες μείωσης των μεταβλητών για την ανάπτυξη μοντέλων πρόβλεψης της συγκεκριμένης χρήσης ενέργειας από το δημόσιο της Κροατίας. Στόχος τους ήταν να επιλύσουν το ερώτημα πώς να συμπεριλάβουν μια πλατφόρμα μεγάλων δεδομένων και μηχανικής μάθησης σε ένα ευφύες σύστημα για τη διαχείριση της ενεργειακής απόδοσης του δημόσιου τομέα ως σημαντικό συστατικό της ιδέας της έξυπνης πόλης, όσο και να προσδιορίσουν τις απαιτήσεις για την ανάπτυξη μιας τέτοιας πλατφόρμας που θα μπορούσε να χρησιμοποιηθεί από τη δημόσια διοίκηση [55]. Η τεχνική Random Forest παρείχε το πιο ακριβές μοντέλο και πραγματοποιήθηκε σύγκριση των σχετικών προβλεπτικών παραγόντων που ανακτήθηκαν και με τις τρεις μεθόδους.

Είναι ευρέως γνωστό ότι τα ορυκτά καύσιμα, που αποτελούν την κύρια πηγή ενέργειας για τον πλανήτη μας, θα εξαντληθούν κάποια στιγμή. Για το λόγο αυτό, ένα μείζον παγκόσμιο πρόβλημα είναι η μετάβαση από τα ορυκτά καύσιμα στις ανανεώσιμες πηγές ενέργειας, η οποία απαιτεί εξελίξεις στα υλικά, τις τεχνολογίες και τα συστήματα για την αποτελεσματική συλλογή, αποθήκευση, μετατροπή και διαχείριση της ανανεώσιμης ενέργειας. Οι προσεγγίσεις μηχανικής μάθησης (ML) έχουν χρησιμοποιηθεί από ειδικούς σε θέματα ενέργειας για την επιτάχυνση αυτών των εξελίξεων. Υπήρξαν πολλές έρευνες που έδειξαν τα αποτελέσματα της χρήσης της μηχανικής μάθησης (ML) για την ενίσχυση της συλλογής ενέργειας (φωτοβολταϊκά), της αποθήκευσης (μπαταρίες), της μετατροπής (ηλεκτροκατάλυση) και της διαχείρισης (έξυπνα δίκτυα) [56], [38].

Εν μέσω μιας παγκόσμιας συζήτησης για τους ενεργειακούς πόρους, είναι κρίσιμο να δοθούν στους αποφασίζοντες διαφορετικές λύσεις για το πώς να γίνουμε πιο αποδοτικοί ενεργειακά. Μία από αυτές τις λύσεις προτάθηκε από δύο επιστήμονες στην Ελλάδα και περιλάμβανε τη διαχείριση των εγκαταστάσεων επεξεργασίας λυμάτων (ΕΕΛ) με τη χρήση μηχανικής μάθησης [57]. Η παρούσα μελέτη εξέτασε δεδομένα από πολλά δημοτικά έργα ΕΕΛ στην Ελλάδα προκειμένου να αναπτυχθούν μοντέλα ενεργειακά αποδοτικών εκτιμήσεων. Τα δεδομένα αξιολογήθηκαν αρχικά στατιστικά στο πλαίσιο των χαρακτηριστικών του έργου και της ενεργειακής κατανάλωσης και χρησιμοποιήθηκε ανάλυση συσχέτισης για τον εντοπισμό των κατάλληλων προβλεπτικών μεταβλητών του έργου. Η μελέτη υπογραμμίζει τη δυνητική δυνατότητα γενίκευσης αυτών των μοντέλων εντός και εκτός του ελληνικού πλαισίου, παρέχοντας πολύτιμα εργαλεία στους ενδιαφερόμενους φορείς για την ενημέρωση της λήψης αποφάσεων, την αποτελεσματική κατανομή των πόρων και τη βελτίωση των ενεργειακά αποδοτικών σχεδίων, με αποτέλεσμα την εξοικονόμηση κόστους και τα οφέλη της βιωσιμότητας.

Μια άλλη καινοτόμος προσέγγιση του θέματος διατυπώθηκε από τους Fischer et al. (2023) οι οποίοι προσπάθησαν να δημιουργήσουν ένα ενοποιημένο πλαίσιο για την αξιολόγηση της ενεργειακής απόδοσης της μηχανικής μάθησης. Δήλωσαν ότι οι περιβαλλοντικά ευαισθητοποιημένοι καταναλωτές απαιτούν ένα ενιαίο πλαίσιο για την αξιολόγηση, τη σύγκριση και την αναφορά των συμβιβασμών αποδοτικότητας και απόδοσης διαφόρων μεθοδολογιών και μοντέλων. Έτσι, παρουσίασαν καινοτόμους αλγορίθμους συγκέντρωσης αποδοτικότητας, ευρετηρίωσης και διαβάθμισης για εφαρμογές ML σε περισσότερα από 20 μοντέλα SOTA σε μια ποικιλία αρχιτεκτονικών υλικού. Για τον σκοπό αυτό, πρότειναν ένα σύνολο μέτρων που επιτρέπουν μια ολοκληρωμένη εικόνα, λαμβάνοντας υπόψη τον τύπο της εργασίας, το αφηρημένο μοντέλο, το λογισμικό και το υλικό. Ως αποτέλεσμα, τα συστήματα ML παρουσιάζονταν πλέον παρόμοια σε όλες τις ρυθμίσεις εκτέλεσης. Προσέφεραν επίσης μια πρόταση για τη γραφική κοινοποίηση πληροφοριών αποδοτικότητας στο κοινό με κατανοητό τρόπο, εμπνευσμένη από το σύστημα ενεργειακών ετικετών της ΕΕ [58].

Η έρευνα που πραγματοποιήθηκε στο πλαίσιο της παρούσας διπλωματικής είναι καινοτόμα και διαφέρει από τις προσεγγίσεις που περιγράφονται παραπάνω για αρκετούς λόγους. Καταρχάς, αντί να βασίζεται σε δεδομένα μεγάλης συχνότητας δειγμάτων ή σε πραγματικό χρόνο, όπως συνήθως συμβαίνει στις προβλέψεις κατανάλωσης με μηχανική μάθηση, εργάζεται με δεδομένα που είναι εύκολα προσβάσιμα στους ιδιοκτήτες ή/και διαχειριστές κτιρίων καθώς προκύπτουν από τον λογαριασμό κατανάλωσης ενέργειας του κτιρίου. Αυτό επιτρέπει να εφαρμόζονται τα μοντέλα σε ένα ευρύ φάσμα κτιρίων χωρίς την ανάγκη για ειδικά συστήματα μέτρησης ή συλλογής δεδομένων. Συγχρόνως, η προσέγγισή που ακολουθείται είναι πολυδιάστατη γιατί πειραματίζεται με μια πληθώρα αλγορίθμων μηχανικής μάθησης, συγκρίνοντας τα αποτελέσματα τους προκειμένου να εντοπιστεί η πιο αποτελεσματική μέθοδος. Αυτό επιτρέπει να προσαρμόζεται η μοντελοποίηση σε

διάφορες συνθήκες και να επιλέγεται ο καταλληλότερος αλγόριθμος για κάθε περίπτωση.

Επιπλέον, αντί να περιοριστεί απλώς στην πρόβλεψη της κατανάλωσης ενέργειας, συσχετίζονται τα αποτελέσματα με τις συγκεκριμένες παρεμβάσεις που έχουν πραγματοποιηθεί στο κτίριο. Η συγκεκριμένη μεθοδολογία επιτρέπει όχι μόνο να πραγματοποιούνται προβλέψεις, αλλά και να αξιολογείται η επίδραση διαφορετικών δράσεων και να προτεραιοποιούνται οι μελλοντικές παρεμβάσεις με βάση την τελική επιρροή που θα έχουν. Τέλος, λαμβάνεται υπόψη όχι μόνο η κατανάλωση ενέργειας αλλά και οι εκπομπές αερίων θερμοκηπίου δίνοντας τη δυνατότητα για μια ολοκληρωμένη αξιολόγηση για το περιβαλλοντικό αποτύπωμα των παρεμβάσεων που πραγματοποιούνται στα κτίρια.

ΚΕΦΑΛΑΙΟ 4

Υλοποίηση Μοντέλων Μηχανικής Μάθησης για την Πρόβλεψη της Αποτελεσματικότητας Ενεργειακών Ανακαινίσεων

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

4 Υλοποίηση Μοντέλων Μηχανικής Μάθησης για την Πρόβλεψη της Αποτελεσματικότητας Ενεργειακών Ανακαινίσεων

Στο κεφάλαιο αυτό θα παρουσιαστούν αναλυτικά τα εργαλεία που χρησιμοποιήθηκαν, η μέθοδος που ακολουθήθηκε για την ανάλυση των αρχικών δεδομένων, οι αλγόριθμοι μηχανικής μάθησης που εφαρμόστηκαν αλλά και ένα νευρωνικό δίκτυο που δημιουργήθηκε με σκοπό την εύρεσης της βέλτιστης στρατηγικής ενεργειακής ανακαίνισης κτιρίων και τη μείωση τους σφάλματος των προηγούμενων μεθόδων μηχανικής μάθησης.

4.1 Εργαλεία Συστήματος

Για την ανάπτυξη μιας εφαρμογής, χρειάζεται ένα σύνολο εργαλείων που θα βοηθήσουν στην υλοποίηση των λειτουργιών και τη διαχείριση του κώδικα, τα οποία περιλαμβάνουν αναπτυξιακά πλαίσια (frameworks), βιβλιοθήκες και προγράμματα που επιτρέπουν την αποτελεσματική ανάπτυξή της. Στη συνέχεια περιγράφονται αναλυτικά τα αντίστοιχα εργαλεία που χρησιμοποιήθηκαν και τον σκοπό που εξυπηρετούν.

Οι βιβλιοθήκες είναι συλλογές προκαθορισμένου κώδικα που παρέχουν έτοιμες λειτουργίες για συγκεκριμένες ανάγκες. Μπορούν να χρησιμοποιηθούν για να επιταχύνουν την ανάπτυξη και να παρέχουν πρόσθετες δυνατότητες. Στη συνέχεια αναφέρονται μερικές από τις σημαντικότερες βιβλιοθήκες που χρησιμοποιήθηκαν στην ανάπτυξη της εφαρμογής.

scikit-learn (sklearn): Είναι μια δημοφιλής βιβλιοθήκη μηχανικής μάθησης στην Python που παρέχει πολλούς αλγόριθμους μηχανικής μάθησης και εργαλεία για προεπεξεργασία δεδομένων, αξιολόγηση μοντέλων, επιλογή χαρακτηριστικών και άλλα.

matplotlib: Είναι μια βιβλιοθήκη γραφικών για τη γλώσσα προγραμματισμού Python. Χρησιμοποιείται για τη δημιουργία γραφημάτων και απεικονίσεων δεδομένων.

NumPy: Είναι μια βιβλιοθήκη Python που παρέχει υψηλής απόδοσης δομές δεδομένων, όπως πίνακες και πίνακες πολλαπλών διαστάσεων, και εργαλεία για να επεξεργάζεται αυτές τις δομές.

pandas: Είναι μια βιβλιοθήκη Python για διαχείριση και ανάλυση δεδομένων. Παρέχει δομές δεδομένων όπως τα DataFrame που είναι πολύ ευέλικτα για την εργασία με δομές δεδομένων σε μορφή πίνακα.

TensorFlow και Keras: Το TensorFlow είναι μια ανοικτού κώδικα πλατφόρμα μηχανικής νευρωνικών δικτύων που παρέχει εργαλεία για την ανάπτυξη και την εκπαίδευση μοντέλων μηχανικής μάθησης. Το Keras είναι μια υψηλού επιπέδου βιβλιοθήκη που παρέχει απλότητα και ταυτόχρονα ισχυρές δυνατότητες για την ανάπτυξη νευρωνικών δικτύων.

textwrap: Είναι μια βιβλιοθήκη για τη διαχείριση του αναδίπλωσης κειμένου στη γλώσσα προγραμματισμού Python.

pickle: Είναι ένα πρωτόκολλο σειριοποίησης δεδομένων για τη μεταφορά δεδομένων μεταξύ διαφορετικών προγραμματιστικών περιβαλλόντων στη γλώσσα Python.

openpyxl: Είναι μια βιβλιοθήκη Python για την εργασία με αρχεία Excel (.xlsx), που επιτρέπει την ανάγνωση, την εγγραφή και την επεξεργασία αυτών των αρχείων.

re (Regular Expressions): Είναι μια βιβλιοθήκη που παρέχει εργαλεία για την εκτέλεση λειτουργιών αντιστοίχισης προτύπων (pattern matching) σε κείμενο με τη χρήση κανόνων ορισμένων από τα κανονικά εκφραστικά.

datetime: Είναι μια βιβλιοθήκη που παρέχει εργαλεία για την εργασία με ημερομηνίες και ώρες στη γλώσσα προγραμματισμού Python.

collections: Παρέχει πρόσθετες δομές δεδομένων πέρα από τις βασικές δομές της Python, όπως οι λίστες, οι πλειάδες και οι λεξικοί.

LightGBM: Είναι μια βιβλιοθήκη που παρέχει υψηλή απόδοση υλοποιήσεων των αλγορίθμων μηχανικής μάθησης για προβλήματα επιτάχυνσης Gradient Boosting.

4.2 Μηχανική Μάθηση & Ανάλυση Δεδομένων

Κατά τη διάρκεια των προηγούμενων δεκαετιών, οι σημαντικές εξελίξεις στα συστήματα υπολογιστών είχαν ως αποτέλεσμα την ανάπτυξη νέων τεχνολογιών για την εκπαίδευση μέσω διαδικτύου. Η αυξανόμενη άνοδος των βάσεων δεδομένων δημιούργησε τη ζήτηση για την ανάπτυξη συστημάτων που χρησιμοποιούν αποτελεσματικά τις πληροφορίες και τη γνώση. Υπάρχουν διαφορετικοί τρόποι ανάλυσης δεδομένων όπως η γενίκευση [59], ο χαρακτηρισμός [60] και η ταξινόμηση [61].

Για να μπορεί να επιτευχθεί ο στόχος της επιτυχούς πρόβλεψης της μείωσης της κατανάλωσης ενέργειας αλλά και των εκπομπών διοξειδίου του άνθρακα με βάσει τα χαρακτηριστικά του κτιρίου και τις ανακαινίσεις που έχει υποστεί έπρεπε να επεξεργαστούν τα αρχικά ανομοιόμορφα δεδομένα από κτίρια που είχαν δοθεί. Γενικότερα στη μηχανική μάθηση το κομμάτι της ανάλυσης δεδομένων είναι πολύ σημαντικό γιατί η μορφή των δεδομένων μπορεί να αλλάξει το αποτέλεσμα για οποιοδήποτε αλγόριθμο μηχανικής μάθησης χρησιμοποιείται για την επίλυση ενός προβλήματος.

Ξεκινώντας είναι σημαντικό να αναφερθεί ότι τα δεδομένα τα οποία δόθηκαν αφορούσαν κτιριακές μονάδες στη Λετονία οι οποίες συμμετείχαν σε ενεργειακές ανακαινίσεις κτιρίων με σκοπό να γίνουν πιο ενεργειακά αποδοτικά. Τα δεδομένα είναι ανώνυμα και χρησιμοποιούνται αποκλειστικά για ερευνητικούς σκοπούς. Αναλυτικότερα αναφέρονται τα τεχνικά χαρακτηριστικά κάθε κτιρίου, διάφορες καταναλώσεις που αφορούν τη θέρμανση, το ηλεκτρικό ρεύμα αλλά και το ζεστό νερό, οι ενέργειες που πραγματοποιήθηκαν για να ανακαινιστεί ενεργειακά το κάθε κτίριο καθώς επίσης και οικονομικά στοιχεία για κάθε project.

Στη συνέχεια παρουσιάζονται αναλυτικά τα βήματα τα οποία ακολουθήθηκαν μέσα από διαφορετικές προσεγγίσεις αλλά και γραφικές παραστάσεις προκειμένου να είναι πιο κατανοητοί οι αλγόριθμοι αλλά και οι μέθοδοι που υλοποιήθηκαν.

4.2.1 Αρχικά Δεδομένα

Ξεκινώντας με το κομμάτι του data analysis της παρούσας διπλωματικής τα δεδομένα που μελετήθηκαν αφορούσαν κτίρια στη Λετονία. Πρόκειται για δεδομένα τα οποία χρησιμοποιήθηκαν στο πλαίσιο του Ευρωπαϊκού έργου ENERGATE (Homepage | Energate. (n.d.). <https://energate-project.eu>) και τα οποία είναι ανώνυμα, έτσι ώστε να μπορούν να χρησιμοποιηθούν για ερευνητικούς σκοπούς. Ειδικότερα, τα δεδομένα περιείχαν τρία αρχεία Excel καθένα από τα οποία περιείχε και ένα ξεχωριστό σύνολο δεδομένων(dataset). Πιο συγκεκριμένα:

- Το πρώτο dataset(*Data on activities*) περιείχε δεδομένα για την ανακαίνιση κάθε κτιρίου όπως :
 - Την κατηγορία project (Tender),
 - τον αριθμό project στο οποίο ανήκε (Project No.),
 - τον τίτλο του κτιρίου (Project building),
 - το είδος του κτιρίου (Type of building),

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

- το είδος ενεργειακής ανακαίνισης που πραγματοποιήθηκε (Type of activity),
- τις δραστηριότητες που περιλάμβανε αυτή η ανακαίνιση (Project activity),
- την μείωση που επέφερε η ανακαίνιση τόσο στις εκπομπές διοξειδίου του άνθρακα όσο και στην ενεργειακή κατανάλωση (CO₂ reduction (in tons, Energy reduction (in MWh)),
- τον συντελεστή εκπομπής διοξειδίου του άνθρακα (CO₂ emission factor)

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Tender	Project No.	Project building	Project activity
KPFI-1	KPFI-1/1	Cesvaine preschool educational institution (Builders iela 1, Cesvaine, C	Expanded clay wall insulation
KPFI-1	KPFI-1/1	Cesvaine preschool educational institution (Builders iela 1, Cesvaine, C	Silicate brick wall insulation
KPFI-1	KPFI-1/1	Cesvaine preschool educational institution (Builders iela 1, Cesvaine, C	Exterior door replacement of glazed doors
KPFI-1	KPFI-1/1	Cesvaine preschool educational institution (Builders iela 1, Cesvaine, C	Wooden window replacement with double-glazed windows
KPFI-1	KPFI-1/1	Cesvaine preschool educational institution (Builders iela 1, Cesvaine, C	Attic ceiling insulation
KPFI-1	KPFI-1/1	Cesvaine preschool educational institution (Builders iela 1, Cesvaine, C	Basement ceiling insulation
KPFI-1	KPFI-1/1	Cesvaine preschool educational institution (Builders iela 1, Cesvaine, C	Heating pipe insulation
KPFI-1	KPFI-1/1	Cesvaine preschool educational institution (Builders iela 1, Cesvaine, C	Roof change
KPFI-1	KPFI-1/10	Ventspils 2. Primary School	School building attic ceiling cleaning and heat insulation, mineral w
KPFI-1	KPFI-1/10	Ventspils 2. Primary School	Building foundation heat insulation around the perimeter, above a
KPFI-1	KPFI-1/10	Ventspils 2. Primary School	Building brick exterior wall heat insulation, 100mm polystyrene / m
KPFI-1	KPFI-1/10	Ventspils 2. Primary School	Light reconstruction washrooms, installing a motion sensor (with a
KPFI-1	KPFI-1/10	Ventspils 2. Primary School	Lighting reconstruction of basement floor corridor and changing ro

Εικόνα 3: Δεδομένα από το πρώτο dataset

Type of building	Type of activity	CO2 reduction (in tons)	Energy reduction (in MWh)	CO2 emission factor
preschool educational institution	renovation of the building's enclosing structures	23,193	87,854	0,264
preschool educational institution	renovation of the building's enclosing structures	15,712	59,514	0,264
preschool educational institution	renovation of the building's enclosing structures	2,245	8,502	0,264
preschool educational institution	renovation of the building's enclosing structures	5,985	22,672	0,264
preschool educational institution	renovation of the building's enclosing structures	29,927	113,360	0,264
preschool educational institution	renovation of the building's enclosing structures	11,223	42,510	0,264
preschool educational institution	renovation of the building's enclosing structures	7,482	28,340	0,264
preschool educational institution	renovation of the building's enclosing structures	0,000	0,000	0,264
school	renovation of the building's enclosing structures	5,811	22,010	0,264
school	renovation of the building's enclosing structures	2,113	8,002	0,264
school	renovation of the building's enclosing structures	29,567	111,996	0,264
school	energy efficient lighting	0,792	1,994	0,397
school	energy efficient lighting	0,479	1,206	0,397

Εικόνα 4: Δεδομένα από το πρώτο dataset

- Το δεύτερο dataset (*Data on projects*) παρέχει πληροφορίες για τις διάφορες κατηγορίες project(*Tender*) όπως :
 - το όνομα του project group που ανήκει κάθε κτίριο (Name of Tender),
 - τον αριθμό του project (Project No.),
 - το συνολικό κόστος των ανακαίνισεων κάθε κατηγορίας (Total project costs),
 - τη διάρκεια τους (Project duration in months),
 - τις ακριβείς ημερομηνίες έναρξης και λήξης (Start date of the project, End date of the project),
 - το ποσό της χρηματοδότησης (Grant financing),
 - την εκτιμώμενη μείωση στις εκπομπές διοξειδίου του άνθρακα κάθε κατηγορίας (Project planned CO2 reduction (in tons)).

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Tender	Name of Tender	Project No.	Project planned CO2 reduction (in tons)	Start date of the project	End date of the project	Project duration in months	Total project costs	Grant financing	% of Grant
KPFI-1	Increase of energy efficiency in I KPFI-1/1		95,750	2009-11-19	2010-12-21	13,07	367.882,94 €	155.643,93 €	42,31%
KPFI-1	Increase of energy efficiency in I KPFI-1/10		261,425	2009-11-11	2010-12-29	13,60	1.226.770,81 €	1.041.574,24 €	84,90%
KPFI-1	Increase of energy efficiency in I KPFI-1/11		64,500	2009-11-16	2010-12-29	13,43	220.956,24 €	92.388,59 €	41,81%
KPFI-1	Increase of energy efficiency in I KPFI-1/12		44,200	2009-12-01	2010-06-10	6,30	273.418,26 €	203.322,82 €	74,36%
KPFI-1	Increase of energy efficiency in I KPFI-1/13		73,900	2009-12-27	2011-02-02	13,17	396.496,04 €	170.638,71 €	43,04%
KPFI-1	Increase of energy efficiency in I KPFI-1/14		48,120	2009-11-27	2010-08-11	8,47	325.296,71 €	119.693,40 €	36,80%
KPFI-1	Increase of energy efficiency in I KPFI-1/15		470,400	2009-12-04	2011-12-29	24,83	1.850.566,56 €	1.331.837,57 €	71,97%
KPFI-1	Increase of energy efficiency in I KPFI-1/16		423,700	2009-12-04	2012-03-28	27,80	1.638.058,31 €	1.178.442,85 €	71,94%
KPFI-1	Increase of energy efficiency in I KPFI-1/17		107,500	2009-11-27	2011-08-29	21,07	667.100,98 €	416.055,67 €	62,37%
KPFI-1	Increase of energy efficiency in I KPFI-1/18		73,400	2009-11-11	2010-11-24	12,43	264.685,92 €	127.148,86 €	48,04%
KPFI-1	Increase of energy efficiency in I KPFI-1/19		357,450	2009-12-02	2012-02-17	26,50	1.595.940,92 €	1.103.442,62 €	69,14%
KPFI-1	Increase of energy efficiency in I KPFI-1/22		55,200	2009-11-11	2010-12-28	13,57	170.469,26 €	72.992,44 €	42,82%
KPFI-1	Increase of energy efficiency in I KPFI-1/23		92,580	2009-12-27	2010-12-22	11,83	398.404,11 €	222.487,59 €	55,84%

Εικόνα 5: Δεδομένα από το δεύτερο dataset

Tender group	Number of Projects	Sum of Total project costs	Sum of Grant financing	% grant	Sum of Project planned CO2 reduction (in tons)
KPFI-1					
Increase of energy efficiency in municipal buildings (I round)	56	46.119.547,71 €	31.690.120,72 €	68,71%	10799,035
KPFI-10					
Low energy consumption buildings	6	6.667.213,47 €	3.359.914,96 €	50,39%	1426,041
KPFI-15					
Complex solutions for greenhouse gas emission reduction (I round)	6	2.312.276,20 €	714.794,19 €	30,91%	442,835
KPFI-15.1					
Complex solutions for greenhouse gas emission reduction (II round)	34	14.249.825,65 €	6.622.702,23 €	46,48%	3942,831
KPFI-15.2					
Complex solutions for greenhouse gas emission reduction (III round)	101	42.662.976,61 €	18.179.089,28 €	42,61%	9833,309

Εικόνα 6: Δεδομένα από το δεύτερο dataset

- Το τρίτο dataset(*Detailed implementation data*) περιέχει αναλυτικά δεδομένα για κάθε κτίριο όπως :
- η έκταση του (Total heating are, m²),
 - ο αριθμός ορόφων του (Number of floors),
 - το είδος λειτουργίας του κτιρίου (House functions),
 - το πότε κατασκευάστηκε το εκάστοτε κτίριο (Year of building (first year and years of renovations/additions where applicable)),
 - η πόλη στην οποία είναι χτισμένο (Town),
 - η κομητεία στην οποία είναι χτισμένο (County)
 - οι καταναλώσεις που αφορούν τη θέρμανση (Source of heat (Heating), CO2 emission factor (tons CO2/MWh) (Heating), CO2 emission factor (audit) (tons CO2/MWh) (Heating)),
 - το ηλεκτρικό ρεύμα αλλά και το ζεστό νερό σε διάφορες μονάδες μέτρησης τόσο πριν την έναρξη του project αλλά και μετά την ολοκλήρωση του,
 - θερμοκρασίες και καταναλώσεις θέρμανσης, ηλεκτρικού ρεύματος αλλά και ζεστού νερού αναλυτικά για κάθε μήνα ενός έτους.

Στο τελευταίο αυτό dataset εν τέλει περιέχονται και πληροφορίες σε ποσοστιαία κλίμακα για την μείωση που επιτεύχθηκε στην κατανάλωση ενέργειας και στις εκπομπές διοξειδίου του άνθρακα για κάθε κατηγορία project ξεχωριστά (Sum of % energy reduction, Sum of % CO2 reduction).

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Reporting year	Tender	Project No.	Planned CO2 reduction per year	House Nr.	Year of building (first year and years of renovations/additions where applicable)	Number of floors	Total heating area, m2	House functions	Town	County
2011	KPFI-1	KPFI-1/1	95,767	1	1972, 1982	3	2834,00	Preschool educational institution	Cesvaine	Cesvaine
2011	KPFI-1	KPFI-1/10	31,686	1	1962	4	2462,00	Educational institution	Ventspils	Ventspils
2011	KPFI-1	KPFI-1/10	59,076	2	1958	5	4633,00	Educational institution	Ventspils	Ventspils
2011	KPFI-1	KPFI-1/10	36,516	3	1963, 1967	2	1729,00	Preschool educational institution	Ventspils	Ventspils
2011	KPFI-1	KPFI-1/10	50,067	4	1965	2	1756,00	Preschool educational institution	Ventspils	Ventspils
2011	KPFI-1	KPFI-1/10	42,317	5	1973, 1996	2	1781,00	Preschool educational institution	Ventspils	Ventspils
2011	KPFI-1	KPFI-1/10	30,777	6	1968	2	1761,00	Preschool educational institution	Ventspils	Ventspils
2011	KPFI-1	KPFI-1/10	10,986	7	N/A	1 to 2	429,00	Preschool educational institution	Ventspils	Ventspils
2011	KPFI-1	KPFI-1/11	64,501	1	N/A	N/A	2673,70	Educational institution	Vecumnieki	Vecumnieku novads
2011	KPFI-1	KPFI-1/14	48,181	1	1974	3	2198,30	Preschool educational institution	Salaspils	Salaspils
2011	KPFI-1	KPFI-1/18	73,396	1	1966	3	6950,40	Educational institution	Ikšķile	Ikšķiles novads
2011	KPFI-1	KPFI-1/22	29,042	1	1998	4	1203,60	Municipal administrative building	N/A	Engures novads
2011	KPFI-1	KPFI-1/22	26,140	2	1986	2	1060,10	Preschool educational institution	N/A	Engures novads

Εικόνα 7: Δεδομένα από το τρίτο dataset

Energy consumption before project (MWh)	Energy consumption before project (kWh/m2)	CO2 emission before project (tons)	Meteorology station	Source of heat (Heating)	CO2 emission factor (tons CO2/MWh) (Heating)	CO2 emission factor (audit) (tons CO2/MWh) (Heating)	Source of heat (Hot water)	CO2 emission factor (tons CO2/MWh) (Hot water)	CO2 emission factor (tons CO2/MWh) (Electricity)	Total energy consumption, MWh (Heating before project)	CO2 emission, tons (Heating before project)	Total energy consumption, MWh (Heating reporting year)	CO2 emission, tons (Heating reporting year)	Total energy consumption, MWh (Heating difference)	CO2 emission, tons (Heating difference)
578,703	204,200	152,778	Zilāni	Central heating	0,264	0,286	Diesel fuel	0,000	0,000	578,703	152,778	263,316	69,515	-315,387	-83,263
460,394	187,000	121,544	Liepāja	Central heating	0,264	0,123	Diesel fuel	0,000	0,000	460,394	121,544	187,728	49,560	-272,666	-71,984
820,041	177,000	216,491	Liepāja	Central heating	0,264	0,400	Diesel fuel	0,000	0,000	820,041	216,491	336,974	88,961	-483,067	-127,530
435,708	252,000	115,027	Liepāja	Central heating	0,264	0,400	Diesel fuel	0,000	0,000	435,708	115,027	154,063	40,673	-281,645	-74,354
477,632	272,000	126,095	Liepāja	Central heating	0,264	0,400	Diesel fuel	0,000	0,000	477,632	126,095	156,523	41,322	-321,109	-84,773
463,060	260,000	122,248	Liepāja	Central heating	0,264	0,428	Diesel fuel	0,000	0,000	463,060	122,248	155,303	41,000	-307,757	-81,248
419,118	238,000	110,647	Liepāja	Central heating	0,264	0,400	Diesel fuel	0,000	0,000	419,118	110,647	151,978	40,122	-267,140	-70,525
120,120	280,000	31,712	Liepāja	Central heating	0,264	0,392	Diesel fuel	0,000	0,000	120,120	31,712	70,049	18,493	-50,071	-13,219
569,498	213,000	150,347	Rīga	Central heating	0,264	0,264	Diesel fuel	0,000	0,000	569,498	150,347	337,859	89,195	-231,639	-61,152
430,207	195,700	113,575	Rīga	Central heating	0,264	0,264	Diesel fuel	0,000	0,000	430,207	113,575	261,756	69,104	-168,451	-44,471
959,155	138,000	253,217	Rīga	Central heating	0,264	0,264	Diesel fuel	0,000	0,000	959,155	253,217	680,342	179,610	-278,813	-73,607
210,630	175,000	55,606	Mērsrags	Electricity, kWh	0,264	0,265	Diesel fuel	0,000	0,000	210,630	55,606	96,804	25,556	-113,826	-30,050
240,643	227,000	63,530	Mērsrags	Central heating	0,264	0,258	Diesel fuel	0,000	0,000	240,643	63,530	97,191	25,658	-143,452	-37,872

Εικόνα 8: Δεδομένα από το τρίτο dataset

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Total energy consumption, MWh (Hotwater before project)	CO2 emission, tons (Hotwater before project)	Total energy consumption, MWh (Hotwater reporting year)	CO2 emission, tons (Hotwater reporting year)	Total energy consumption, MWh (Hotwater difference)	CO2 emission, tons (Hotwater difference)	Total energy consumption, MWh (Electricity before project)	CO2 emission, tons (Electricity before project)	Total energy consumption, MWh (Electricity reporting year)	CO2 emission, tons (Electricity reporting year)	Total energy consumption, MWh (Electricity difference)	CO2 emission, tons (Electricity difference)	Produced energy, MWh (Renewable energy)	CO2 emission, tons (Renewable energy)	Total energy consumption (MWh)	Total energy consumption (kWh/m2)	Total CO2 emission in reporting year (tons)	CO2 reduction (CO2 in reporting year - CO2 before project) (tons per year)
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	263,316	92,913	69,515	-83,263
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	187,728	76,250	49,560	-71,984
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	336,974	72,733	88,961	-127,530
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	154,063	89,105	40,673	-74,354
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	156,523	89,136	41,322	-84,773
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	155,303	87,200	41,000	-81,248
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	151,978	86,302	40,122	-70,525
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	70,049	163,284	18,493	-13,219
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	337,859	126,364	89,195	-61,152
0,000	0,000	65,590	0,000	65,590	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	327,346	148,909	69,104	-44,471
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	680,342	97,885	179,610	-73,607
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	96,804	80,429	25,556	-30,050
0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	97,191	91,681	25,658	-37,872

Εικόνα 9: Δεδομένα από το τρίτο dataset

Temperat ure-Jan	Temperat ure-Feb	Temperat ure-Ma	Temperat ure-Apr	Temperat ure-Ma	Temperat ure-Jun	Temperat ure-Jul	Temperat ure-Aug	Temperat ure-Sep	Temperat ure-Oct	Temperat ure-Nov	Temperat ure-Dec	Temperat ure-Averag	Heating period, days-Jan	Heating period, days-Feb	Heating period, days-M	Heating period, days-Apr	Heating period, days-Ma	Heating period, days-Ju
23	23	23	23	23	23	23	23	23	23	23	23	23	31	28	31	22	0	0
20,3	20,3	20,3	20,3	20,3	20,3	20,3	20,3	20,3	20,3	20,3	20,3	20,3	31	28	31	30	9	0
20,5	20,5	20,5	20,5	20,5	20,5	20,5	20,5	20,5	20,5	20,5	20,5	20,5	31	28	31	30	9	0
21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	31	28	31	30	10	0
20,6	20,6	20,6	20,6	20,6	20,6	20,6	20,6	20,6	20,6	20,6	20,6	20,6	31	28	31	30	10	0
21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	21,4	31	28	31	30	10	0
21,3	21,3	21,3	21,3	21,3	21,3	21,3	21,3	21,3	21,3	21,3	21,3	21,3	31	28	31	30	10	0
19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	31	28	31	30	10	0
22	22	22	22	22	22	22	22	22	22	22	22	22	31	28	31	20	0	0
18	18	18	18	18	18	18	18	18	18	18	18	18	31	28	31	15	0	0
19	19	19	19	19	19	19	19	19	19	19	19	19	31	28	31	7	0	0
22	22	22	22	22	22	22	22	22	22	22	22	22	31	28	31	30	0	0
23	23	23	23	23	23	23	23	23	23	23	23	23	31	28	31	20	0	0

Εικόνα 10: Δεδομένα από το τρίτο dataset

Tender group	Sum of Energy consumption before project (M	Sum of Total energy consumption (MWh)	Sum of % energy reduction
KPFI-1	520821,466	283301,896	-45,60%
KPFI-10	25963,476	3154,125	-87,85%
KPFI-15	16604,345	11197,15	-32,56%
KPFI-15.1	126144,165	67343,545	-46,61%
KPFI-15.2	299675,713	277970,885	-7,24%
KPFI-15.3	231692,943	119453,62	-48,44%
KPFI-15.4	87034,748	58852,193	-32,38%
KPFI-3	80886,132	38595,351	-52,28%
KPFI-5	128867,688	46411,729	-63,98%
KPFI-6	209530,092	146115,909	-30,26%
KPFI-7	279952,022	133074,355	-52,47%
Grand Total	2007172,79	1185470,758	-40,94%

Εικόνα 11: Δεδομένα από το τρίτο dataset

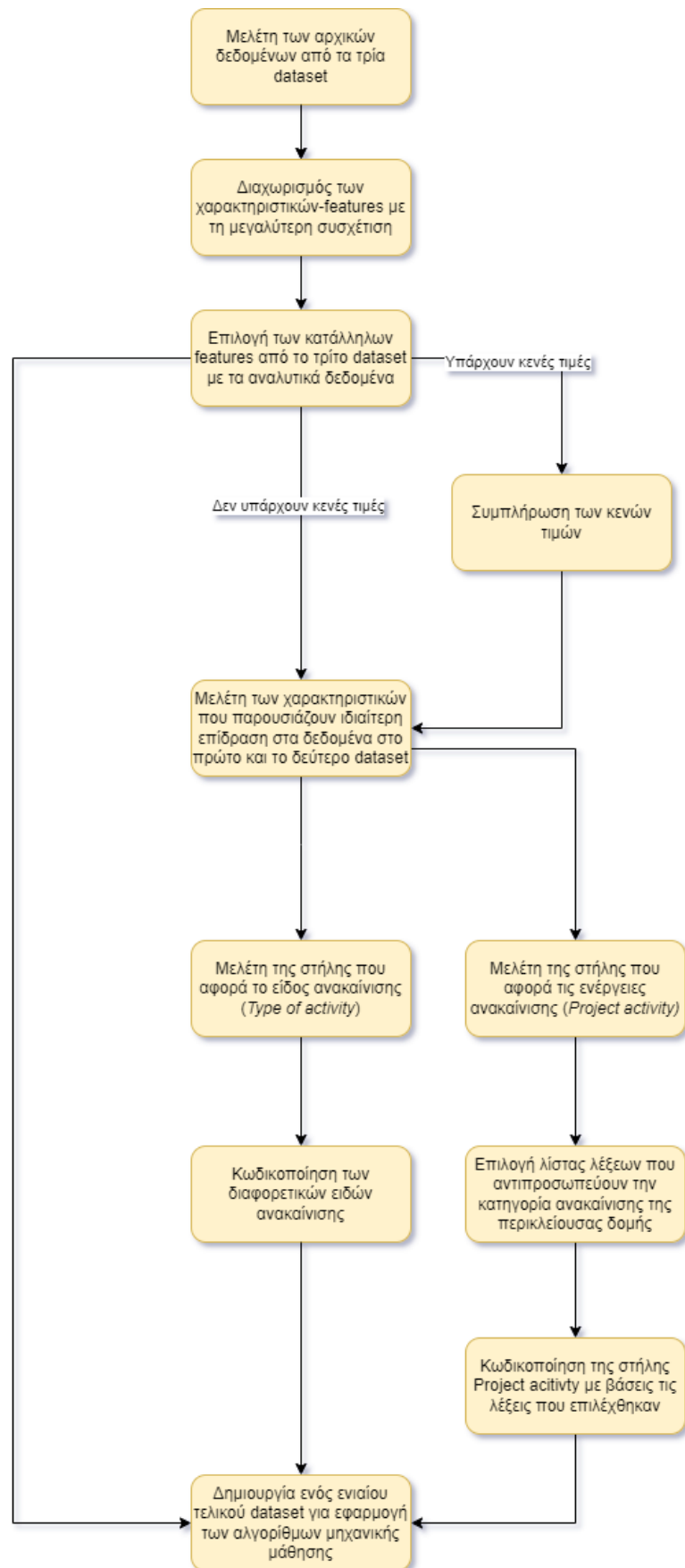
Tender group	Sum of CO2 emission before project (tons)	Sum of Total CO2 emission in reporting year (tons)	Sum of % CO2 reduction
KPFI-1	137496,83	64479,951	-53,10%
KPFI-10	6955,542	737,532	-89,40%
KPFI-15	3018,802	1206,322	-60,04%
KPFI-15.1	37793,604	21139,743	-44,07%
KPFI-15.2	73060,422	31547,636	-56,82%
KPFI-15.3	57962,438	24007,559	-58,58%
KPFI-15.4	21517,247	8737,574	-59,39%
KPFI-3	22648,89	10718,948	-52,67%
KPFI-5	37694,67	14015,78	-62,82%
KPFI-6	60323,319	27436,039	-54,52%
KPFI-7	78146,348	37277,285	-52,30%
Grand Total	536618,112	241304,369	-55,03%

Εικόνα 12: Δεδομένα από το τρίτο dataset

4.2.2 Ανάλυση Δεδομένων

Η μελέτη των αρχικών δεδομένων οδήγησε στο συμπέρασμα ότι ο όγκος πληροφορίας που περιείχαν ήταν πολύ μεγάλος για λειτουργήσει σωστά το μοντέλο και να εξασφαλιστεί η ασφαλής εξαγωγή συμπερασμάτων. Οι πολλές πληροφορίες ελλοχεύουν τον κίνδυνο της υπερπροσαρμογής(overfitting) [62] δηλαδή το μοντέλο να μάθει πολύ καλά τα δεδομένα εκπαίδευσης συμπεριλαμβανομένων και των θορυβωδών παρατηρήσεων και να αποτυγχάνει να εξάγει γενικές πληροφορίες. Στόχος ήταν η δημιουργία ενός καινούριου dataset, το οποίο θα περιείχε χαρακτηριστικά-features και πληροφορίες από τα τρία σύνολα δεδομένων που προαναφέρθηκαν τα οποία θα είχαν τον καθοριστικότερο ρόλο στην εκπαίδευση των μοντέλων και θα οδηγούσαν στο να έχουμε μια επιτυχής πρόβλεψη. Η μεθοδολογία η οποία εφαρμόστηκε φαίνεται στο παρακάτω διάγραμμα.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 4: Μέθοδος ανάλυσης δεδομένων

Για τον λόγο αυτό από το τρίτο dataset(*KPFI_Detailed implementation data*) επιλέχθηκαν τα εξής χαρακτηριστικά:

- η χρονιά που αξιολογήθηκε το κτίριο (Reporting year),
- η κατηγορία project στην οποία ανήκει (Tender), ο αριθμός project (Project No.),
- ο κωδικός κάθε κτιρίου στο αντίστοιχο project (House Nr.),
- η χρονιά κατά την οποία κατασκευάστηκε το κτίριο και οι χρονιές που έγιναν οι πρώτες προσθήκες ή ανακαινίσεις (Year of building (first year and years of renovations/additions where applicable)),
- η ηλικία του κτιρίου η οποία προστέθηκε ύστερα από κάποιον υπολογισμό (Building Age),
- ο αριθμός ορόφων που έχει το κτίριο (Number of floors),
- η επιφάνεια θέρμανσης σε τετραγωνικά μέτρα (Total heating are, m²),
- η κατανάλωση ενέργειας πριν από το project σε MWh και σε kWh/m²(Energy consumption before project (MWh), Energy consumption before project (kWh/m²)),
- οι εκπομπές διοξειδίου του άνθρακα(CO₂) σε τόνους (CO₂ emission before project (tons))
- η συγκεντρωτική μείωση στην κατανάλωση ενέργειας και στις εκπομπές διοξειδίου του άνθρακα για κάθε project τόσο σε αριθμητικό πλαίσιο όσο και σε ποσοστιαία κλίμακα (Sum of Energy consumption before project (MWh), Sum of % energy reduction, Sum of CO₂ emission before project (tons), Sum of % CO₂ reduction).

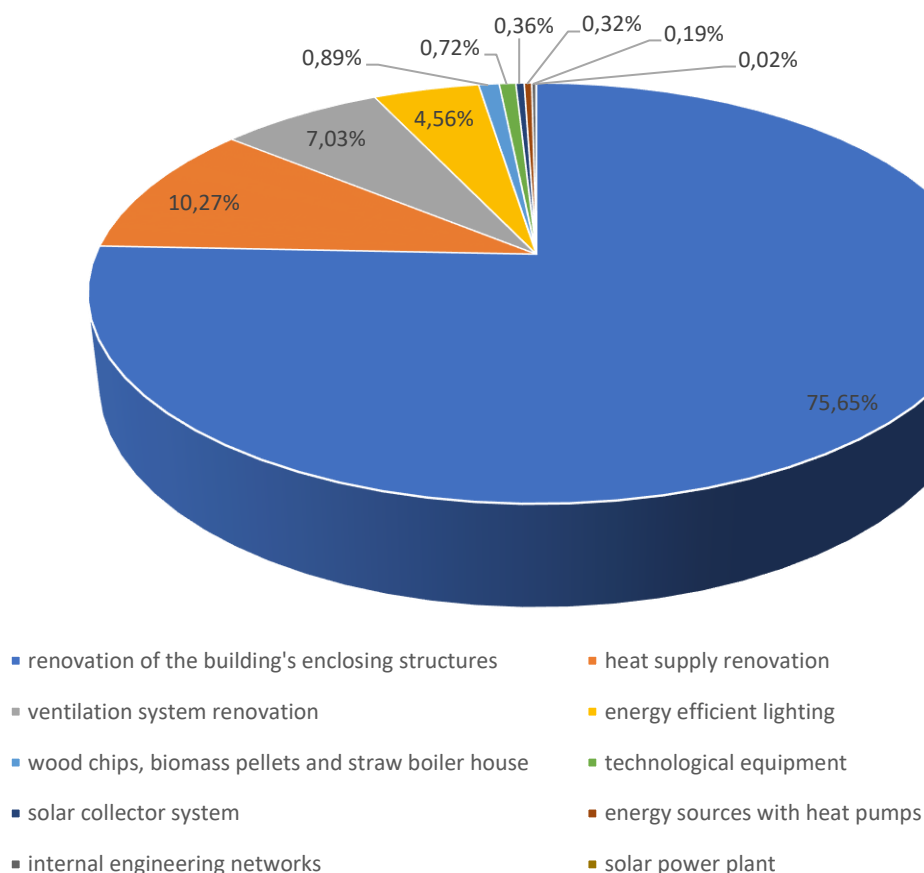
Συγχρόνως οι κενές τιμές συμπληρώθηκαν σε δύο στήλες συμπληρώθηκαν χρησιμοποιώντας την συνάρτηση *random.choice* μέσα από μια λίστα αριθμών που αντιπροσώπευε σωστά το αντίστοιχο feature . Αναλυτικότερα στη στήλη “*Year of building (first year and years of renovations/additions where applicable)*” η λίστα περιείχε αριθμούς από το 1960 έως το 1998 ενώ για τη στήλη “*Number of floors*” η λίστα περιείχε αριθμούς από το 1 έως το 10. Οι τιμές αυτές αντιστοιχούν στις τιμές min και max που παρουσιάζονται στις αντίστοιχες στήλες.

Όσον αφορά το πρώτο σύνολο δεδομένων (*Data on activities*) όπως προαναφέρθηκε περιείχε αναλυτικές πληροφορίες για το κάθε κτίριο και τις ανακαινίσεις που έγιναν σε αυτό. Η επιλογή των κατάλληλων χαρακτηριστικών έγινε ξεκινώντας από το είδος ή τα είδη ανακαίνισης που έλαβαν χώρα σε κάποιο κτίριο (*Type of activity*) . Το συγκεκριμένο feature χρήζει ιδιαίτερης προσοχής καθώς όπως παρατηρήθηκε διαφορετικό *Type of activity* συνεπάγεται και διαφορετική πρόβλεψη για το κάθε κτίριο. Για αυτό επιλέχθηκε η περαιτέρω ανάλυση του. Το *Type of activity* χωρίζεται στις εξής κατηγορίες:

1. renovation of the building's enclosing structures
2. energy efficient lighting
3. ventilation system renovation

4. internal engineering networks
5. wood chips, biomass pellets and straw boiler house
6. solar collector system
7. technological equipment
8. energy sources with heat pumps
9. solar power plant
10. heat supply renovation

Στη συνέχεια φαίνεται η ποσοστιαία κλίμακα κάθε κατηγορίας ανακαίνισης στα κτίρια που εξετάστηκαν.



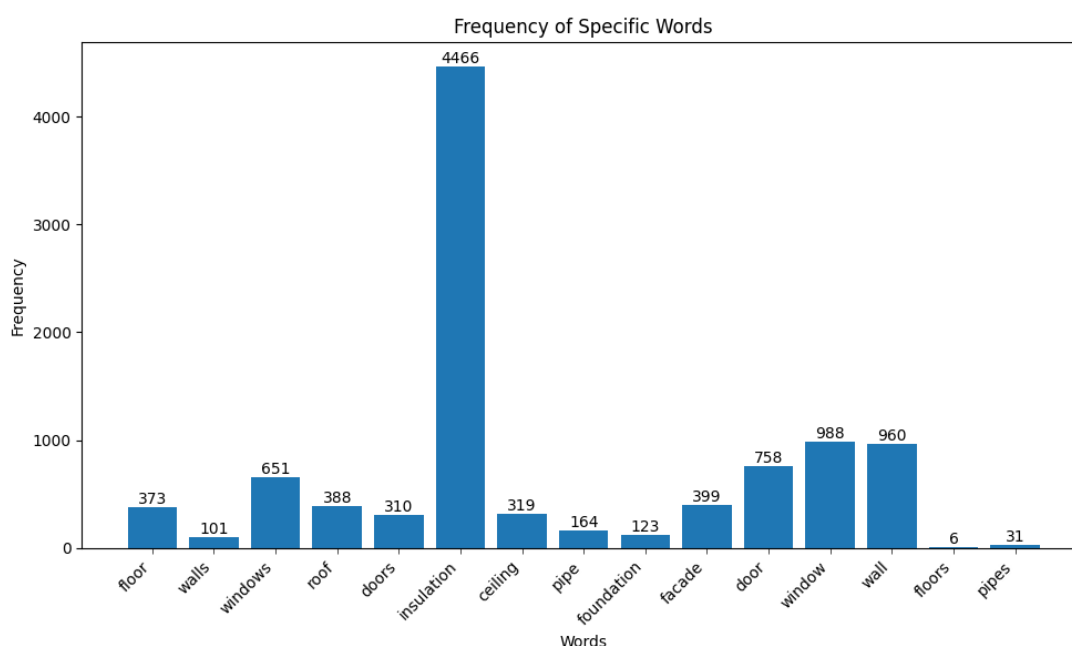
Διάγραμμα 5: Κατανομή των διαφόρων ειδών ανακαίνισης στα κτίρια που μελετήθηκαν

Όπως φαίνεται και από το παραπάνω διάγραμμα το μεγαλύτερο ποσοστό καταλαμβάνουν οι περιπτώσεις ανακαίνισης της περικλείουσας δομής των κτιρίων (renovation of building's enclosing structure), σε ποσοστό 76%. Εύκολα εξάγεται το συμπέρασμα ότι η συγκεκριμένη μέθοδος ανακαίνισης απαιτεί περισσότερη παρατήρηση και μελέτη για τις διάφορες τεχνικές που χρησιμοποιήθηκαν. Για αυτό επιλέχθηκε να αναλυθεί περισσότερο η συγκεκριμένη κατηγορία. Αναλυτικότερα παρατηρήθηκε ότι στις περιπτώσεις των κτιρίων που είχε ανακαινιστεί η περικλείουσα δομή τους η ανακαίνιση αυτή μπορεί να περιελάμβανε ποικίλες υποκατηγορίες δράσεων εξοικονόμησης ενέργειας, όπως είναι η αλλαγή στους τοίχους, τα πατώματα, τις πόρτες τα παράθυρα, τις σωληνώσεις.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Project activity	Type of building	Type of activity
Expanded clay wall insulation	preschool educational institution	renovation of the building's enclosing structures
Silicate brick wall insulation	preschool educational institution	renovation of the building's enclosing structures
Exterior door replacement of glazed doors	preschool educational institution	renovation of the building's enclosing structures
Wooden window replacement with double-glazed windows	preschool educational institution	renovation of the building's enclosing structures
Attic ceiling insulation	preschool educational institution	renovation of the building's enclosing structures
Basement ceiling insulation	preschool educational institution	renovation of the building's enclosing structures
Heating pipe insulation	preschool educational institution	renovation of the building's enclosing structures
Roof change	preschool educational institution	renovation of the building's enclosing structures
School building attic ceiling cleaning and heat insulation, mineral w	school	renovation of the building's enclosing structures
Building foundation heat insulation around the perimeter, above ai	school	renovation of the building's enclosing structures
Building brick exterior wall heat insulation, 100mm polystyrene / m	school	renovation of the building's enclosing structures
Light reconstruction washrooms, installing a motion sensor (with a	school	energy efficient lighting
Lighting reconstruction of basement floor corridor and changing ro	school	energy efficient lighting
Input node reconstruction / replacement door	school	renovation of the building's enclosing structures
Single storey part of the building, roof covering additional heat ins	school	renovation of the building's enclosing structures

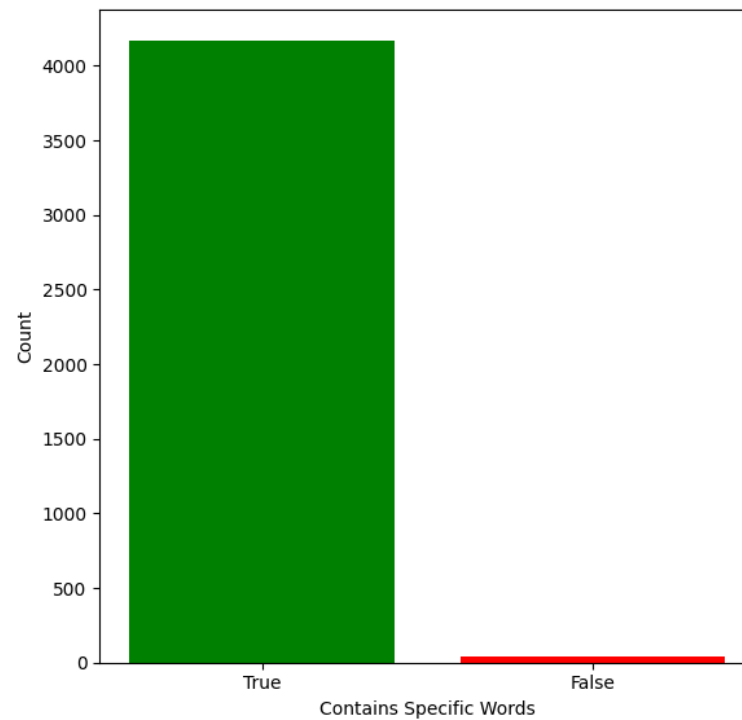
Πιο συγκεκριμένα η στήλη *Project activity* παρουσιάζει τις μεμονωμένες αλλαγές που έγιναν στην δομή του κτιρίου με απώτερο σκοπό την ενεργειακή ανακαίνιση. Όπως είναι ευδιάκριτο τα παραπάνω δεδομένα δεν παρουσιάζουν κάποια ομοιομορφία. Δεν είναι εύκολη η κατηγοριοποίηση τους τόσο γιατί δεν μπορούν να χωριστούν σε υποκατηγορίες όσο και γιατί δεν αναγράφεται ποια χρονιά υλοποιήθηκαν καθώς κάποια κτίρια αξιολογήθηκαν παραπάνω από μία χρονιές. Παρόλα αυτά μετά από μελέτη των δεδομένων αυτών παρατηρήθηκε ότι κάποιες βασικές έννοιες που περιγράφουν αλλαγές στην δομή ενός κτιρίου επαναλαμβάνονται στις διάφορες ενέργειες που πραγματοποιήθηκαν.



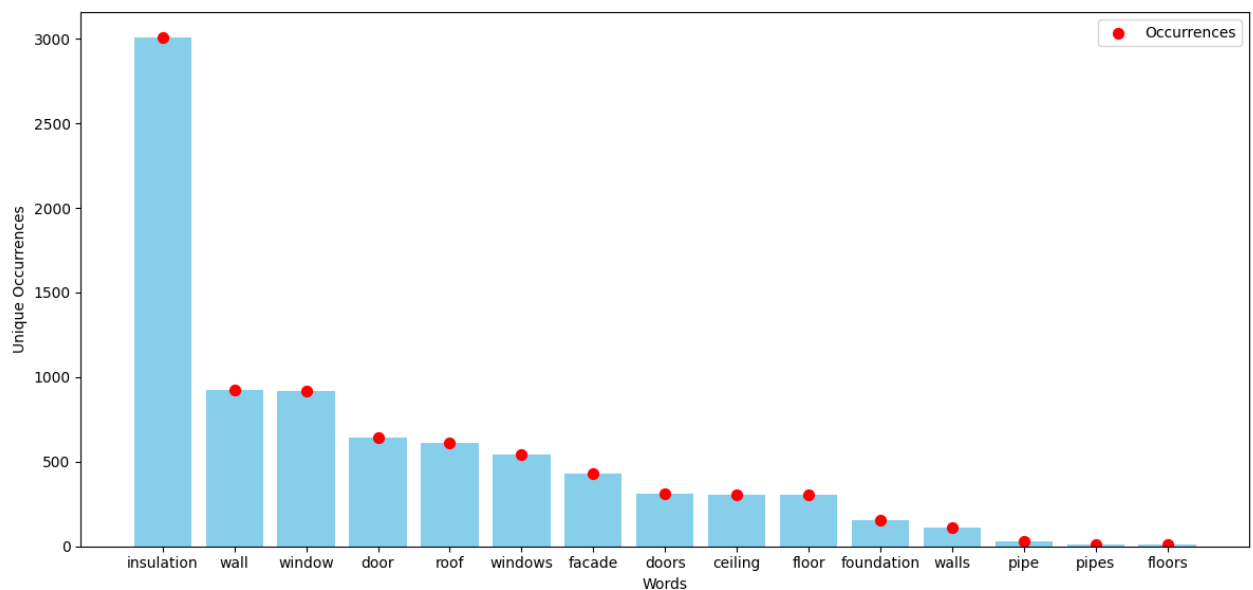
Διάγραμμα 6: Εμφάνιση πιθανών λέξεων στην κατηγορία "Project Activity"

Δημιουργήθηκε έτσι μια λίστα με πιθανές λέξεις οι οποίες χρησιμοποιούνται που περιγράφουν πιο συγκεκριμένα τον τύπο ανακαίνισης της περικλείουσας δομής ενός κτιρίου. Η λίστα ήταν η εξής : floor, walls, windows, roof, doors, insulation, ceiling, pipe, foundation, facade, door, window, wall, floors, pipes. Επιβεβαιώθηκε ότι οι λέξεις αυτές περιέχονταν σε αρκετές περιπτώσεις που ανακαινίσουν όλων των project ανεξάρτητα από το είδος της ανακαίνισης στο οποίο είχαν προβεί όπως φαίνεται από τα παρακάτω διαγράμματα.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

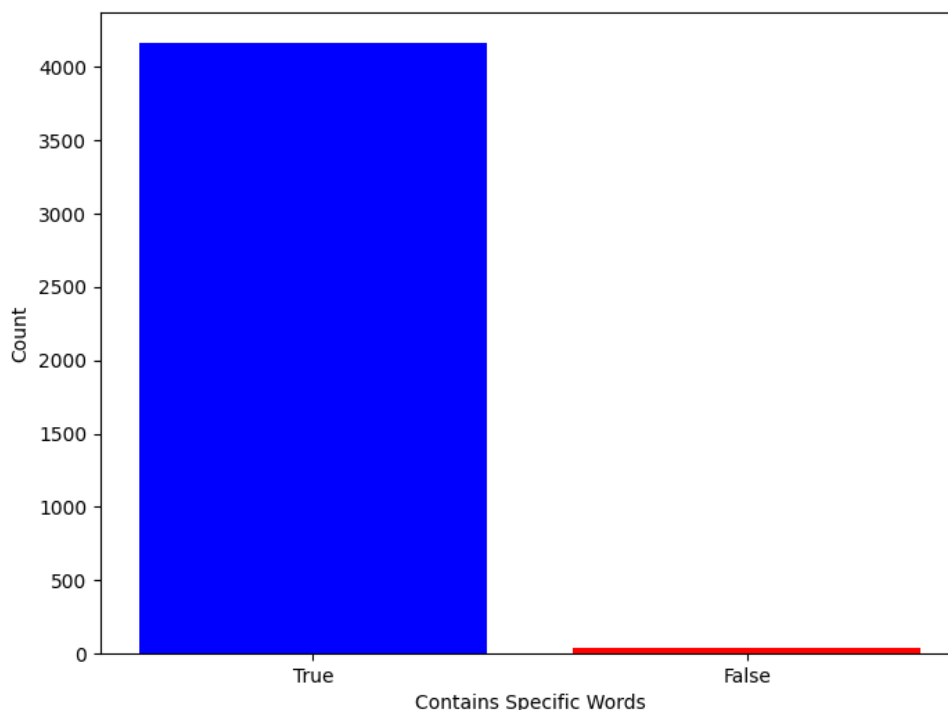


Διάγραμμα 7: Εμφάνιση των συγκεκριμένων λέξεων στην κατηγορία “Project Activity”



Διάγραμμα 8: Μοναδικές εμφανίσεις κάθε λέξης στην κατηγορία “Project Activity”

Με την παρακάτω ανάλυση στην κατηγορία ανακαίνισης της περικλείουσας δομής διαπιστώθηκε ότι είναι ελάχιστες οι περιπτώσεις που δεν εμφανίζονται οι παραπάνω λέξεις .



Διάγραμμα 9: Εμφάνιση των συγκεκριμένων λέξεων στην κατηγορία “Type of activity: renovation of the building’s enclosing structures”

Με βάση τα όσα προαναφέρθηκαν αποφασίστηκε να εφαρμοστεί μια κωδικοποίηση τόσο των κατηγοριών ανακαίνισης όσο και των λέξεων που προαναφέρθηκαν. Ειδικότερα στο συγκεντρωτικό αρχείο με τα απαραίτητα χαρακτηριστικά (features) για τα μοντέλα μας προστέθηκαν αρχικά δέκα στήλες η κάθε μία με το όνομα ενός είδους ανακαίνισης. Η κωδικοποίηση έγινε με δυαδικό τρόπο. Για κάθε κτίριο συγκεντρώθηκαν τα διάφορα είδη ανακαίνισης που έλαβαν χώρα σε αυτό από τις διάφορες γραμμές του αρχείου και συνενώθηκαν σε ένα κελί. Ύστερα με βάσει τις τιμές αυτού του κελιού οι οποίες διαχωρίζονταν με κόμμα τοποθετήθηκε ένα (1) στην αντίστοιχη στήλη ανακαίνισης αν είχε πραγματοποιηθεί για το εκάστοτε κτίριο και μηδέν (0) αν δεν είχε πραγματοποιηθεί. Η υλοποίηση φαίνεται από τον κώδικα παρακάτω.

```
activity_df['Aggregated Type of Activity'] = activity_df['Project
building'].map(activity_map)
# Split the aggregated text based on commas, keep only unique values, and join
them back
activity_df['Aggregated Type of Activity'] = activity_df['Aggregated Type of
Activity'].apply(
lambda x: ', '.join(set(x.split(', '))))
# Create columns for each unique 'Type of Activity'

for activity in unique_activity_values:
    activity_df[activity] = 0
# Split the aggregated text based on commas and use get_dummies to encode values
for index, row in activity_df.iterrows():
```

```
activities = row['Aggregated Type of Activity'].split(',')
for activity in activities:
    if activity in unique_activity_values:
        activity_df.at[index, activity] = 1
```

Με την ίδια ακριβώς δυαδική λογική έγινε και η κωδικοποίηση της λίστας λέξεων . Πιο συγκεκριμένα οι λέξεις χωρίστηκαν σε κατηγορίες όπως φαίνεται από τον παρακάτω πίνακα.

Πίνακας 4: Κατηγοριοποίηση Λέξεων

Κατηγορίες	Λέξεις
Κατηγορία 1	wall
Κατηγορία 2	insulation
Κατηγορία 3	doors
Κατηγορία 4	door
Κατηγορία 5	windows
Κατηγορία 6	window
Κατηγορία 7	ceiling
Κατηγορία 8	roof
Κατηγορία 9	foundation
Κατηγορία 10	floor
Κατηγορία 11	façade
Άλλο	pipe, pipes, floors

Στη συνέχεια πραγματοποιήθηκε η κωδικοποίηση ως εξής. Για κάθε κτίριο συγκεντρώθηκαν οι διάφορες ενέργειες ανακαίνισης που έγιναν σε αυτό από τις διάφορες γραμμές του αρχείου και συνενώθηκαν σε ένα κελί. Ύστερα πραγματοποιήθηκε αναζήτηση σε κάθε κελί για το αν περιέχει μία ή περισσότερες από τις λέξεις αυτές και τοποθετήθηκε ένα (1) στην αντίστοιχη στήλη κατηγορίας αν περιείχε την αντίστοιχη λέξη/λέξεις και μηδέν(0) αν δεν περιείχε κάποια από αυτές. Η υλοποίηση φαίνεται από τον κώδικα παρακάτω.

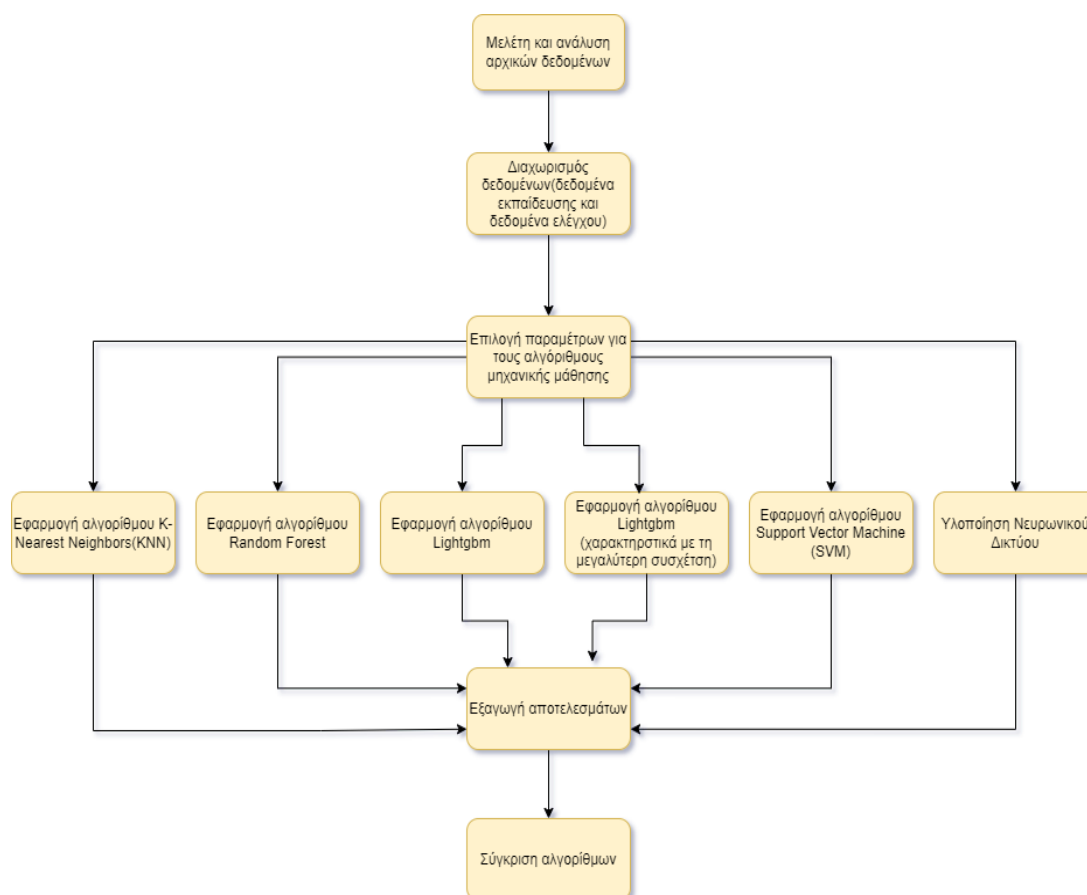
```
# Function to assign categories to words
def categorize_words(word):
    return word_categories.get(word, 'Other Words') # Assign 'Other Words' category if
word not found in word_categories
def create_dummies(text):
    words = set(text.lower().split())
    categories = {key: 0 for key in word_categories.values()} # Initialize categories
dictionary with zeros
    category = categorize_words(word)
```

```
for word in words:
    if category in word_categories.values():
        categories[category] = 1 # Set 1 for the specific category
    return pd.Series(categories)
# Apply the function to create DataFrame with dummies for categories on 'Aggregated
Project Activity' column
result['Aggregated Project Activity'] = result['Aggregated Project Activity'].astype(str)
#Ensure 'Aggregated Project Activity' is string type
dummies_df = result['Aggregated Project Activity'].apply(create_dummies)
# Concatenate the original DataFrame with the dummies DataFrame
result = pd.concat([result, dummies_df], axis=1)
```

Στη συνέχεια χρησιμοποιώντας τον αριθμό του κάθε project σαν κριτήριο (Project No.) πραγματοποιήθηκε η αντιστοίχιση ανάμεσα στα χαρακτηριστικά των κτιρίων που επιλέχθηκαν από το τρίτο αρχείο και στα αποτελέσματα της επεξεργασίας των δεδομένων από το πρώτο αρχείο για να δημιουργηθεί το τελικό αρχείο(Finalized data.xlsx) το οποίο χρησιμοποιήθηκε έπειτα στην υλοποίηση των αλγορίθμων μηχανικής μάθησης. Κάποιοι αριθμοί project επαναλαμβάνονταν καθώς τα αποτελέσματα των ανακαινίσεων σε ορισμένα κτίρια αξιολογήθηκαν παραπάνω από μία χρονιές, ώστε να καταγραφεί η θετική επίδραση της ανακαίνισης σε βάθος χρόνου. Επιπλέον, μία ακόμη πρόκληση που έπρεπε να αντιμετωπιστεί ήταν το γεγονός ότι στο πρώτο αρχείο που αναφέρονται οι ενέργειες ανακαίνισης για κάθε κτιριακή μονάδα δεν αναγράφεται ποιες ενέργειες πραγματοποιήθηκαν μια συγκεκριμένη χρονιά, το οποίο κατέστησε την υλοποίηση πιο δύσκολη. Επομένως, προκειμένου να μπορέσουν να χρησιμοποιηθούν τα δεδομένα θεωρήθηκε ότι οι ενέργειες που αναγράφονται πραγματοποιήθηκαν όσες φορές επανεμφανίζεται ένα έργο.

4.3 Εφαρμογή Αλγορίθμων Μηχανικής Μάθησης

Προτού δημιουργηθεί ένα νευρωνικό δίκτυο από την αρχή για την προσπάθεια εύρεσης της βέλτιστης στρατηγικής ενεργειακής ανακαίνισης κτιρίων εφαρμόστηκαν μοντέλα τεχνητής νοημοσύνης και μηχανικής μάθησης, με απώτερο σκοπό την εύρεση των βέλτιστων δράσεων εξοικονόμησης ενέργειας σε κτιριακές δομές. Τα αποτελέσματα των αλγορίθμων αυτών χρησιμοποιήθηκαν ως benchmarks για το νευρωνικό δίκτυο. Η μεθοδολογία που ακολουθήθηκε κατά την εφαρμογή των αλγορίθμων μηχανικής μάθησης φαίνεται στο παρακάτω διάγραμμα.



Διάγραμμα 10: Μεθοδολογία εφαρμογής των αλγορίθμων μηχανικής μάθησης

Για τους αλγόριθμους αυτούς χρησιμοποιήθηκαν τα features που επιλέχθηκαν από το τρίτο dataset αλλά και η κωδικοποιημένες στήλες τόσο για το *Type of activity* όσο και τις λέξεις που επιλέχθηκαν για τη στήλη *Project activity*. Αναλυτικότερα τα features που χρησιμοποιήθηκαν είναι τα εξής:

Πίνακας 5: Χαρακτηριστικά που επιλέχθηκαν

Feature	Label	Description
1	Reporting Year	Η χρονιά που αξιολογήθηκε το κτίριο
2	Tender	Η κατηγορία project στην οποία ανήκει
3	Project No.	Ο αριθμός project

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

4	House Nr.	Ο κωδικός κάθε κτιρίου στο αντίστοιχο project
5	Year of building (first year and years of renovations/additions where applicable)	Η χρονιά κατά την οποία κατασκευάστηκε το κτίριο και οι χρονιές που έγιναν οι πρώτες προσθήκες ή ανακαινίσεις
6	Building Age	Η ηλικία του κτιρίου η οποία προστέθηκε ύστερα από κάποιον υπολογισμό
7	Number of floors	Ο αριθμός ορόφων που έχει το κτίριο
8	Total heating area, m ²	Η επιφάνεια θέρμανσης σε τετραγωνικά μέτρα
9	Energy consumption before project (MWh), Energy consumption before project (kWh/m ²)	Η κατανάλωση ενέργειας πριν από το project σε MWh και σε kWh/m ²
10	CO ₂ emission before project (tons)	Οι εκπομπές διοξειδίου του άνθρακα(CO ₂) σε τόνους
11	Sum of Energy consumption before project (MWh), Sum of % energy reduction, Sum of CO ₂ emission before project (tons), Sum of % CO ₂ reduction	Η συγκεντρωτική μείωση στην κατανάλωση ενέργειας και στις εκπομπές διοξειδίου του άνθρακα για κάθε project τόσο σε αριθμητικό πλαίσιο όσο και σε ποσοστιαία κλίμακα
12	renovation of the building's enclosing structures	Ένα από τα είδη ανακαίνισης
13	energy efficient lighting	Ένα από τα είδη ανακαίνισης
14	ventilation system renovation	Ένα από τα είδη ανακαίνισης
15	Internal engineering networks	Ένα από τα είδη ανακαίνισης
16	wood chips, biomass pellets and straw boiler house	Ένα από τα είδη ανακαίνισης
17	solar collection system	Ένα από τα είδη ανακαίνισης
18	technological equipment	Ένα από τα είδη ανακαίνισης

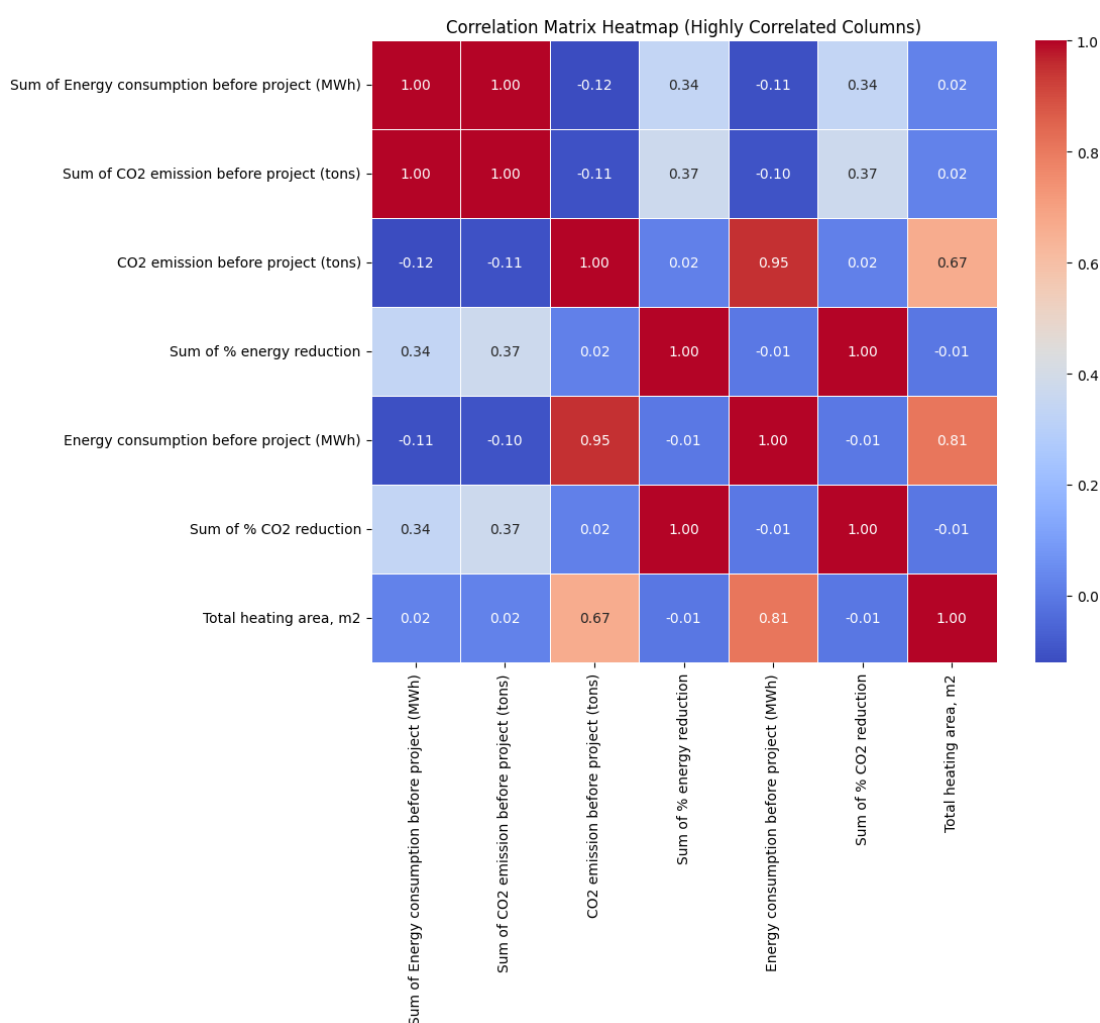
Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

19	energy sources with heat pumps	Ένα από τα είδη ανακαίνισης
20	solar power plant	Ένα από τα είδη ανακαίνισης
21	heat supply renovation	Ένα από τα είδη ανακαίνισης
22	Category 1	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη wall
23	Category 2	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη insulation
24	Category 3	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη doors
25	Category 4	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη door
26	Category 5	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη windows
27	Category 6	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη window
28	Category 7	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη ceiling
29	Category 8	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη roof
30	Category 9	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη foundation
31	Category 10	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη floor

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

32	Category 11	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τη λέξη facade
33	Other	Προτάσεις στην κατηγορία “Project Activity” που περιέχουν τις λέξεις pipe, pipes, floors

Από τα παραπάνω χαρακτηριστικά που επιλέχθηκαν τα χαρακτηριστικά με τη μεγαλύτερη συσχέτιση μεταξύ τους είναι:



Διάγραμμα 11: Χάρτης συσχέτισης χαρακτηριστικών

Να σημειωθεί ότι για το συγκεκριμένο διάγραμμα χρησιμοποιήθηκε κατώφλι 0.7 (threshold = 0.7).

Οι αλγόριθμοι που εφαρμόστηκαν είναι ο αλγόριθμος των κ-κοντινότερων γειτόνων (k-nearest neighbors – KNN) [63] , ο αλγόριθμος Random Forest [64], ο

αλγόριθμος Lightgbm [65] και ο αλγόριθμος Support Vector Machine (SVM) [66]. Για αυτούς τους αλγόριθμους χρησιμοποιήθηκαν τα δεδομένα που διαμορφώθηκαν από την επεξεργασία που αναλύθηκε στο προηγούμενο κεφάλαιο.

Για την μέτρηση της απόδοσης των αλγορίθμων χρησιμοποιήθηκαν κάποιες μετρικές. Σε όλους τους αλγόριθμους που εφαρμόστηκαν υπολογίστηκε το μέσο τετραγωνικό σφάλμα(mean squared error), το μέσο απόλυτο σφάλμα(mean absolute error), το r-τετράγωνο(r-squared) και το μέσο απόλυτο ποσοστιαίο σφάλμα (mean absolute percentage error).

Το μέσο τετραγωνικό σφάλμα(mean squared error) [67] είναι ο μέσος όρος των τετραγώνων των διαφορών μεταξύ των πραγματικών και των προβλεπόμενων τιμών.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Όπου:

- n είναι ο αριθμός των σημείων δεδομένων.
- y_i είναι η πραγματική τιμή.
- $\hat{y}_i$ είναι η προβλεπόμενη τιμή.

Το μέσο απόλυτο σφάλμα(mean absolute error) [67] είναι ο μέσος όρος των απόλυτων διαφορών μεταξύ των πραγματικών και των προβλεπόμενων τιμών.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Όπου:

- n είναι ο αριθμός των σημείων δεδομένων.
- y_i είναι η πραγματική τιμή.
- $\hat{y}_i$ είναι η προβλεπόμενη τιμή.

Το r-τετράγωνο(r-squared) [68]μετρά το ποσοστό της διακύμανσης στην εξαρτημένη μεταβλητή που μπορεί να προβλεφθεί από τις ανεξάρτητες μεταβλητές.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Όπου:

- n είναι ο αριθμός των σημείων δεδομένων.
- y_i είναι η πραγματική τιμή.
- $\hat{y}_i$ είναι η προβλεπόμενη τιμή.
- $\bar{y}$ είναι το μέσον των πραγματικών τιμών

Το εύρος για το R^2 είναι μεταξύ 0 και 1, όπου το 1 υποδεικνύει την τέλεια εφαρμογή.

Το μέσο απόλυτο ποσοστιαίο σφάλμα(mean absolute percentage error) [69] υπολογίζει τον μέσο ποσοστιαίο διαφοροποιητικό όρο μεταξύ προβλεπόμενων και πραγματικών τιμών.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

Όπου:

- n είναι ο αριθμός των σημείων δεδομένων.
- y_i είναι η πραγματική τιμή.
- $\hat{y}_i$ είναι η προβλεπόμενη τιμή.

Το μέσο τετραγωνικό και το μέσο απόλυτο σφάλμα λαμβάνουν τιμές ανάλογα με τη μονάδα μέτρησης της τιμής που προβλέπεται. Στην προκειμένη περίπτωση όλες οι προβλεπόμενες τιμές εκτός από το μέσο τετραγωνικό σφάλμα παρουσιάζουν ποσοστό στη δεκαδική κλίμακα. Το μέσο τετραγωνικό σφάλμα (MSE) είναι όπως προαναφέρθηκε ένα μέτρο της μέσης τετραγωνικής διαφοράς μεταξύ των πραγματικών τιμών και των προβλεπόμενων τιμών σε ένα σύνολο δεδομένων. Δεδομένου ότι τα δεδομένα μας αναπαρίστανται σε δεκαδική μορφή, για παράδειγμα, 0,56 για να δηλώσει το 56%, το MSE θα υπολογιστεί στο τετράγωνο αυτών των δεκαδικών μονάδων. Για παράδειγμα, εάν η προβλεπόμενη τιμή μας είναι 0,56 και η πραγματική τιμή είναι 0,75, η τετραγωνική διαφορά θα είναι $(0,56 - 0,75)^2$. Επομένως οι τιμές των σφαλμάτων που παρατίθενται αφορούν ποσοστά στη δεκαδική κλίμακα.

Οι παραπάνω μετρικές απόδοσης εκφράστηκαν και ως ποσοστά. Όταν εκφράζονται ως ποσοστά, κλιμακώνονται σε σχέση με τη μεγαλύτερη τάξη μεγέθους των πραγματικών τιμών. Για να επιτευχθεί αυτό χρησιμοποιήθηκε το Εύρος Πραγματικών Τιμών (*Target Range*). Το εύρος των πραγματικών τιμών αντιπροσωπεύει τη διακύμανση ή τη διαφορά μεταξύ των μεγαλύτερων και μικρότερων τιμών της στήλης προς πρόβλεψη και υπολογίζεται ως εξής

$$Target\ range = Max(testing\ set) - Min(testing\ set)$$

Επομένως όλες οι ποσοστιαίες εκφράσεις είναι :

$$MAE(\%) = \frac{MAE}{Target\ range}$$

$$MSE(\%) = \frac{MSE}{Target\ range}$$

$$R^2(\%) = \frac{R^2}{Target\ range}$$

4.3.1 Αλγόριθμος K-Nearest-Neighbors (KNN)

Ο πρώτος αλγόριθμος που υλοποιήθηκε ήταν ο αλγόριθμος κ-κοντινότερων γειτόνων. Από προεπιλογή, η συνάρτηση `knn()` χρησιμοποιεί την ευκλείδεια απόσταση, η οποία μπορεί να προσδιοριστεί χρησιμοποιώντας την ακόλουθη εξίσωση:

$$D(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2}$$

όπου p και q είναι τα θέματα που πρέπει να συγκριθούν με η χαρακτηριστικά. Μια άλλη σημαντική έννοια είναι η παράμετρος k , η οποία ελέγχει πόσοι γείτονες επιλέγονται για τον αλγόριθμο KNN. Η βέλτιστη τιμή k επηρεάζει σημαντικά τη διαγνωστική απόδοση του αλγορίθμου KNN. Ένα υψηλό k ελαχιστοποιεί την επιρροή της τυχαίας διακύμανσης του σφάλματος, αλλά αυξάνει την πιθανότητα να αγνοηθούν δευτερεύοντα αλλά σημαντικά μοτίβα. Το κλειδί για την επιλογή μιας κατάλληλης τιμής k είναι η εξεύρεση μιας ισορροπίας μεταξύ υπερπροσαρμογής και υποπροσαρμογής. Ορισμένοι συγγραφείς συνιστούν να ορίζεται το k ίσο με την τετραγωνική ρίζα του αριθμού των παρατηρήσεων στο σύνολο δεδομένων εκπαίδευσης [63].

Όσον αφορά τα δεδομένα, αρχικά πραγματοποιήθηκε κανονικοποίηση σε αυτά, γιατί τα δεδομένα ήταν σε διάφορες μονάδες μέτρησης και ο KNN είναι ευαίσθητος στην κλίμακα γιατί λειτουργεί με αποστάσεις όπως φάνηκε παραπάνω, προκειμένου να μπορούν να χρησιμοποιηθούν σωστά από τον αλγόριθμο KNN και εν συνέχεια χωρίστηκαν σε δεδομένα εκπαίδευσης (training set) και σε δεδομένα ελέγχου (testing set) ενώ επιλέχθηκε σαν μεταβλητή πρόβλεψης η ποσοστιαία μείωση του διοξειδίου του άνθρακα ("Sum of % CO2 reduction") πρώτα και η ποσοστιαία μείωση στην ενεργειακή κατανάλωση έπειτα ("Sum of % energy reduction").

```
# Scale features
scaler = StandardScaler()
scaled_features = scaler.fit_transform(df_finalized)
# Split the data
target_variable = df_finalized['Sum of % CO2 reduction']
X_train, X_test, y_train, y_test = train_test_split(scaled_features, target_variable,
test_size=
0.30)
```

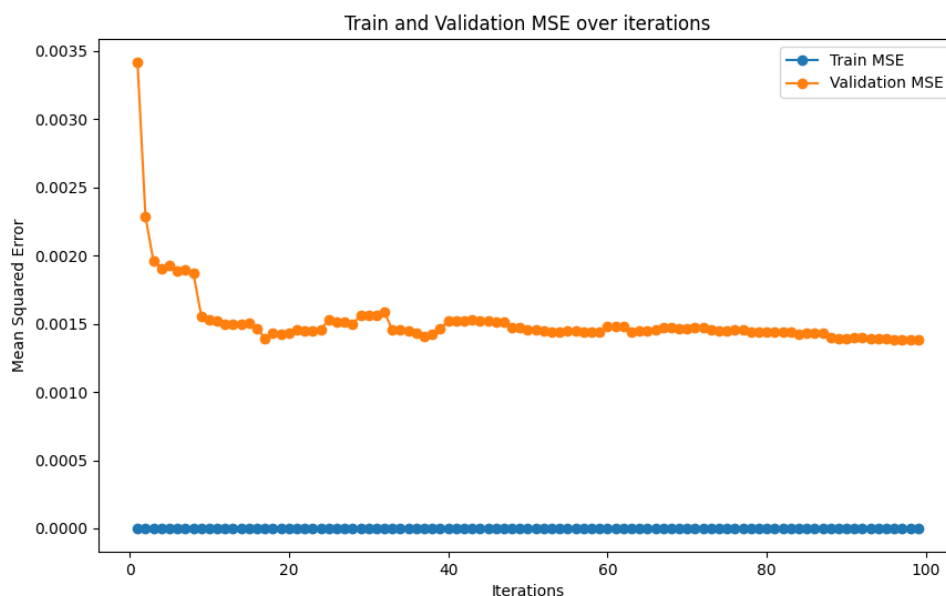
Έπειτα εφαρμόστηκε ο αλγόριθμος KNN με επιλογή $k=1$, το οποίο θα αιτιολογηθεί στη συνέχεια, έγιναν οι προβλέψεις και τυπώθηκαν κάποιες μετρικές για την αξιολόγηση της ορθής λειτουργίας του αλγορίθμου.

```
knn = KNeighborsRegressor(n_neighbors=1)
pred = knn.predict(X_test)
r_squared = knn.score(X_test, y_test)
```

```
print(f"R-squared: {r_squared}")
mse = mean_squared_error(y_test, pred)
print(f"Mean Squared Error: {mse}")
mae = mean_absolute_error(y_test, pred)
print(f"Mean Absolute Error: {mae}")
```

Όσον αφορά τις μετρικές του αλγόριθμου KNN παρουσιάζονται παρακάτω :

Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0013844890756302519	0.3847832383948112	0.01753352007469654	
0.3054%	38.4783%	3.8680%	3.0407%



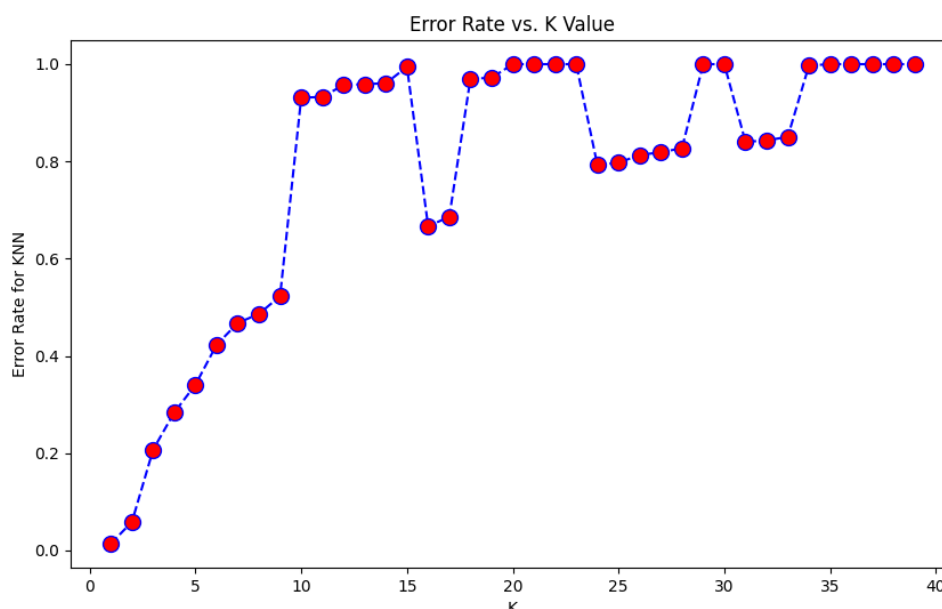
Διάγραμμα 12: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

Για την επιλογή του $k=1$ όπως παρουσιάζεται στο παρακάτω διάγραμμα επιλέχθηκε γιατί δοκιμάστηκε ο αλγόριθμος για k από 1 έως 40 και όπως φαίνεται το $k=1$ έχει το μικρότερο σφάλμα από όλα τα iterations που εκτελέστηκαν. Αξιοσημείωτο είναι ότι ο αλγόριθμος δεν παρουσιάζει υπερπροσαρμογή (overfitting) παρά την επιλογή $k=1$. Αυτό οφείλεται στους εξής λόγους :

- Τοπική Γενίκευση: Με $k=1$, η πρόβλεψη για ένα νέο δεδομένο βασίζεται αποκλειστικά στον πλησιέστερο γείτονα. Αυτή η τοπική προσέγγιση επιτρέπει

στο μοντέλο να προσαρμόζεται καλά στην τοπική δομή των δεδομένων χωρίς να επηρεάζεται υπερβολικά από τον θόρυβο ή τις λεπτομέρειες που θα μπορούσαν να οδηγήσουν σε overfitting.

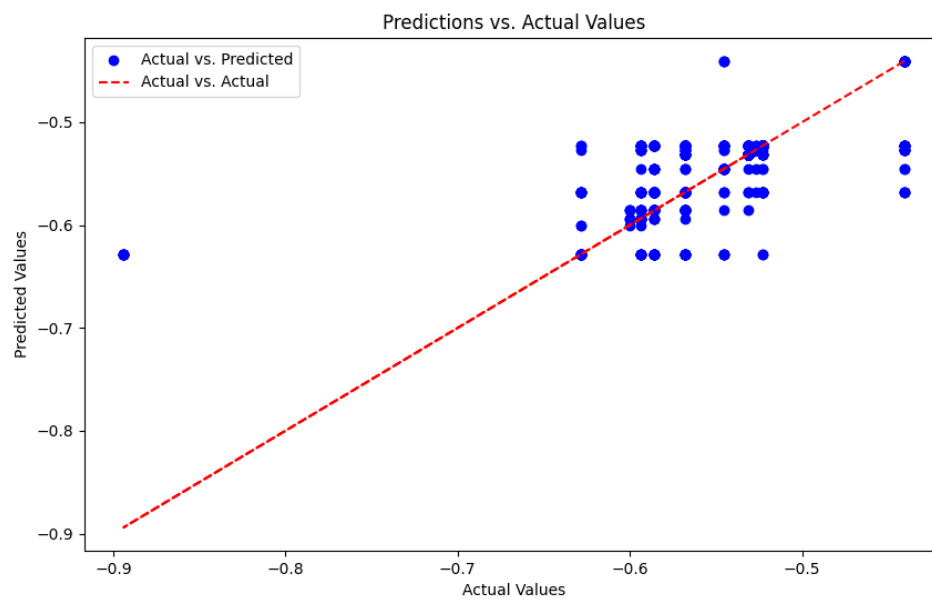
- Smoothing effect: Έχοντας μόνο έναν γείτονα σημαίνει ότι οι προβλέψεις είναι ευαίσθητες σε κάθε μεμονωμένο σημείο δεδομένων, αλλά αυτό μπορεί επίσης να λειτουργήσει ως μορφή smoothing. Αυτό αποτρέπει το μοντέλο από το να αιχμαλωτίσει θόρυβο που ενδέχεται να υπάρχουν στα δεδομένα εκπαίδευσης, συνεισφέροντας σε ένα πιο αξιόπιστο και λιγότερο υπερεκπαιδευμένο μοντέλο.
- Αποφυγή Πολύπλοκων Συνόρων Απόφασης: Το KNN με $k=1$ τείνει να δημιουργεί σύνορα απόφασης που ακολουθούν στενά τα δεδομένα. Αντίθετα από μοντέλα με υψηλότερες τιμές του k που ενδέχεται να δημιουργήσουν πολύπλοκα σύνορα απόφασης, το $k=1$ τείνει να αποφεύγει περιττή πολυπλοκότητα, μειώνοντας τον κίνδυνο overfitting.



Διάγραμμα 13: Iterations για εύρεση καλύτερου k

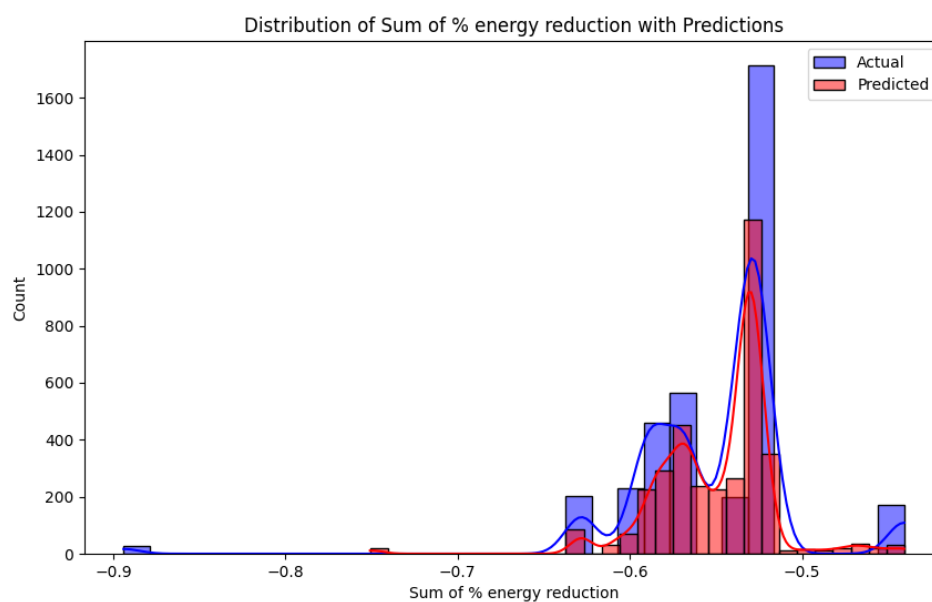
Στη συνέχεια παρουσιάζεται το διάγραμμα το οποίο συγκρίνει τις τιμές που προέβλεψε ο αλγόριθμος συγκριτικά με τις πραγματικές τιμές στην στήλη “Sum of % CO2 reduction”.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 14: Προβλέψεις vs Πραγματικές τιμές

Τέλος παρουσιάζεται η κατανομή των πραγματικών τιμών της μείωσης διοξειδίου του άνθρακα συγκριτικά με τις προβλέψεις.

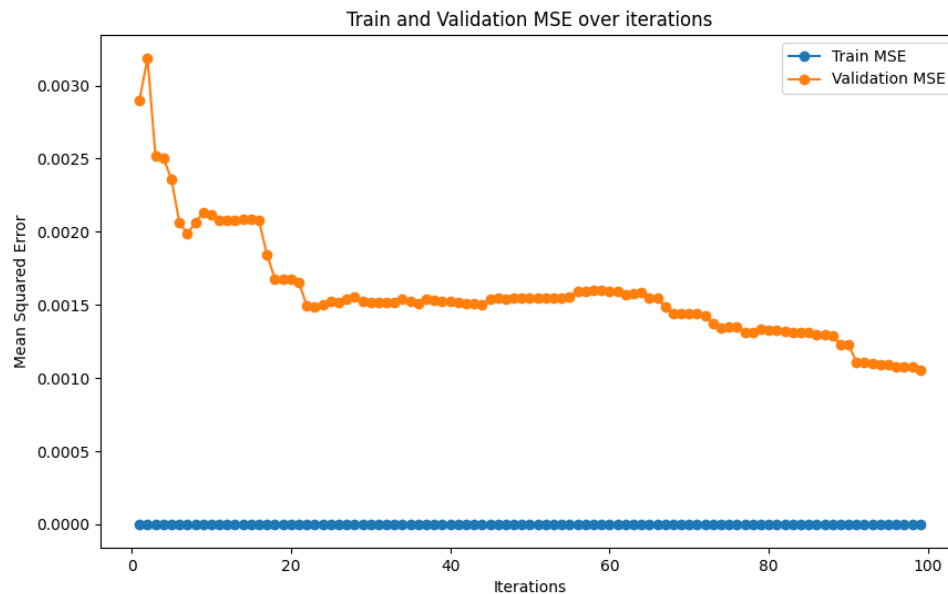


Διάγραμμα 15: Κατανομή του “Sum of % CO2 reduction” συγκριτικά με τις προβλέψεις

Πέρα από τη στήλη “Sum of % CO2 reduction” με τις εκπομπές διοξειδίου του άνθρακα ο αλγόριθμος KNN εκτελέστηκε και για τις τιμές για την στήλη “Sum of % energy reduction”. Οι μετρικές απόδοσης φαίνονται παρακάτω.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

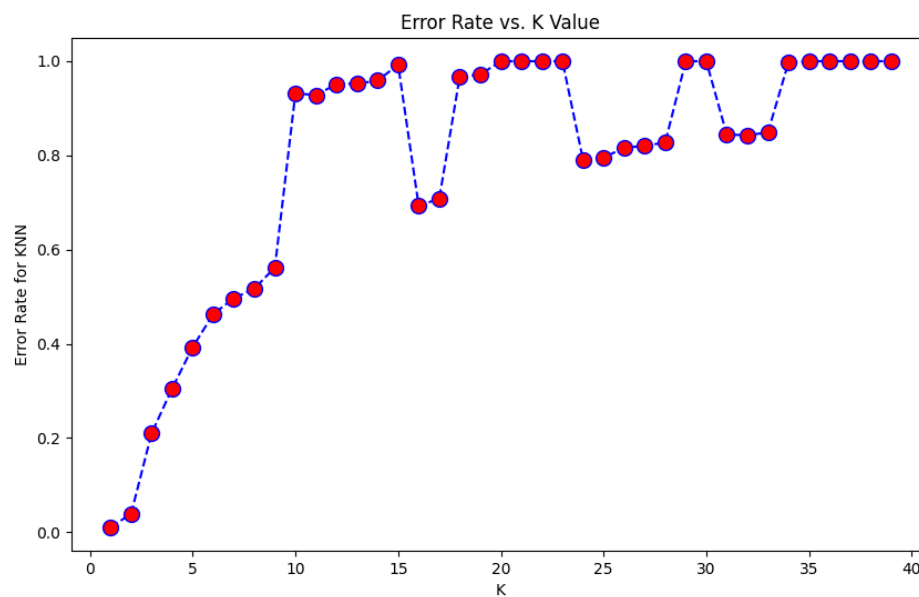
Mean Squared Error	R-Squared	Mean absolute error	Mean Absolute Percentage Error
0.0010553810457516336	0.5798622595140075	0.013434360410830994	
0.2328%	57.9862%	2.9637%	2.2674%



Διάγραμμα 16: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

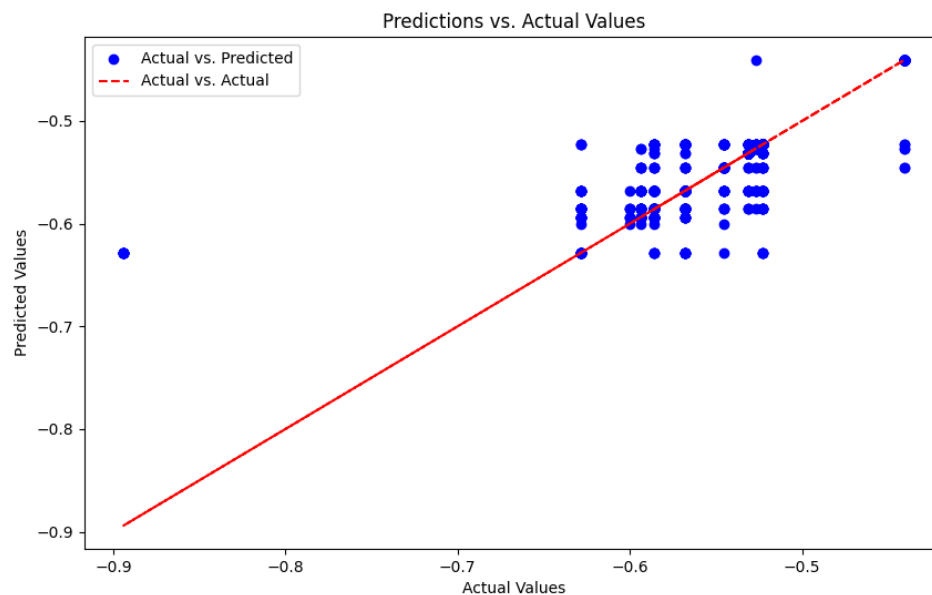
Επιλέχθηκε ξανά $k=1$ για τον ίδιο λόγο που εξηγήθηκε παραπάνω όπως φαίνεται από το επόμενο διάγραμμα. Στην συγκεκριμένη περίπτωση αποφεύχθηκε ξανά η υπερπροσαρμογή (overfitting) για τους ίδιους λόγους που αναφέρθηκαν παραπάνω, δηλαδή λόγω της τοπικής γενίκευσης, του smoothing effect και της αποφυγής πολύπλοκων συνόρων απόφασης.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



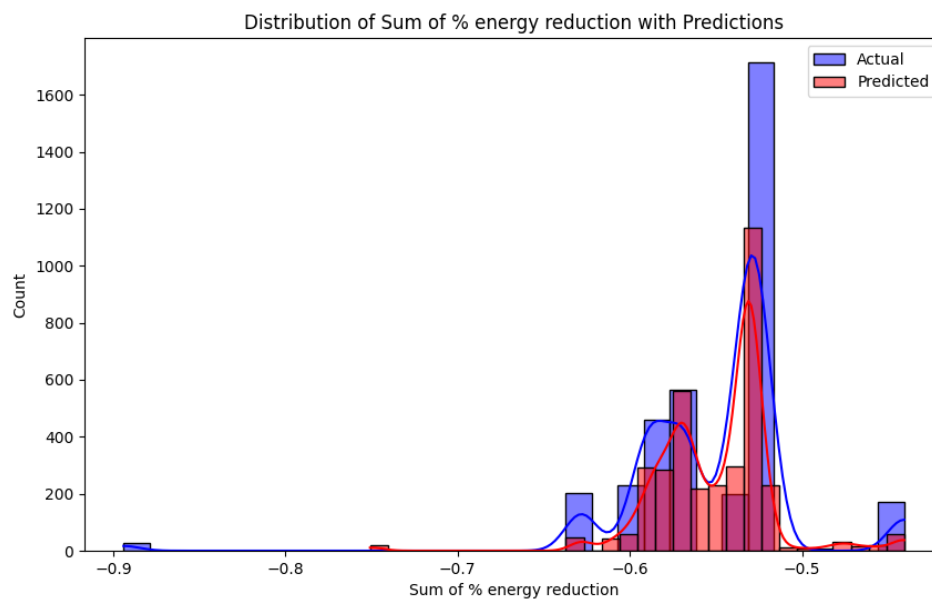
Διάγραμμα 17: Iterations για εύρεση καλύτερου k

Στα επόμενα διαγράμματα παρουσιάζεται η σύγκριση των πραγματικών τιμών με τις προβλέψεις για την στήλη "Sum of % energy reduction" και η κατανομή των πραγματικών τιμών της μείωσης ενεργειακής κατανάλωσης συγκριτικά με τις προβλέψεις.



Διάγραμμα 18: Προβλέψεις vs Πραγματικές τιμές

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 19: Κατανομή του “Sum of % energy reduction” συγκριτικά με τις προβλέψεις

4.3.2 *Random Forest*

Ο αλγόριθμος Random Forest επινοήθηκε από τον Breiman και χρησιμοποιεί την τυχαιότητα για να δημιουργήσει έναν τεράστιο αριθμό δέντρων απόφασης. Η έξοδος αυτών των δέντρων συνδυάζεται σε μια ενιαία έξοδο με ψηφοφορία όταν το πρόβλημα αφορά εργασίες ταξινόμησης και με μέσο όρο όταν αφορά θέματα παλινδρόμησης. Η τυχαιοποίηση επιτυγχάνεται με δύο τρόπους. Πρώτον, το σύνολο δεδομένων δειγματοληπτείται με τη χρήση μιας μεθόδου αντικατάστασης. Η πράξη της συλλογής ενός νέου δείγματος με αυτόν τον τρόπο είναι γνωστή ως "συνάθροιση" ή "bagging". Για παράδειγμα, εάν υπάρχουν έξι εθελοντές σε μια έρευνα, μπορούν να χρησιμοποιηθούν έξι ζαριές για την επιλογή ενός νέου, τυχαίου δείγματος. Η δεύτερη τυχαιοποίηση πραγματοποιείται στους κόμβους απόφασης. Σε κάθε κόμβο επιλέγεται ένας συγκεκριμένος αριθμός προβλεπτών. Για μια συλλογή p προβλεπτών, μια συνήθης ποσότητα είναι η στρογγυλεμένη τετραγωνική ρίζα του p , ωστόσο η παράμετρος αυτή επιλέγεται από τον κάθε αναλυτή. Στη συνέχεια, ο αλγόριθμος αξιολογεί όλα τα πιθανά κατώφλια για όλες τις καθορισμένες μεταβλητές και επιλέγει τον συνδυασμό μεταβλητής-κατωφλίου που παράγει τον βέλτιστο διαχωρισμό [70].

Για την εφαρμογή του αλγορίθμου Random Forest χρησιμοποιήθηκαν τα αρχικά δεδομένα τα οποία αναλύθηκαν παραπάνω. Όσον αφορά τα δεδομένα, αρχικά εφαρμόστηκε ένας Imputer στα δεδομένα μας σε περίπτωση που είχαν κενές τιμές με στόχο να τις αντικαταστήσει χρησιμοποιώντας την τεχνική του μέσου όρου και εν συνεχεία πραγματοποιήθηκε scaling σε αυτά καθώς παρατηρήθηκε ότι βρίσκονται σε διαφορετικές κλίμακες, γεγονός που βοηθάει στην ταχύτητα εκπαίδευσης του μοντέλου στον αλγόριθμο Random Forest, προκειμένου να μπορούν να χρησιμοποιηθούν σωστά από τον αλγόριθμο. Συγχρόνως επιλέχθηκε σαν μεταβλητή πρόβλεψης η ποσοστιαία μείωση του διοξειδίου του άνθρακα ("Sum of % CO2 reduction") πρώτα και η ποσοστιαία μείωση στην ενεργειακή κατανάλωση έπειτα ("Sum of % energy reduction").

```
X = df_finalized.drop('Sum of % CO2 reduction', axis=1)
y = df_finalized['Sum of % CO2 reduction']
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3,
random_state=42)
# Assuming X_train and y_train have NaN values
imputer = SimpleImputer(strategy='mean')
X_train_imputed = imputer.fit_transform(X_train)
X_test_imputed = imputer.transform(X_test) # Apply imputer transformation to the
test data
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train_imputed)
X_test_scaled = scaler.transform(X_test_imputed) # Apply the same scaling to the
test data as done on the training data
```

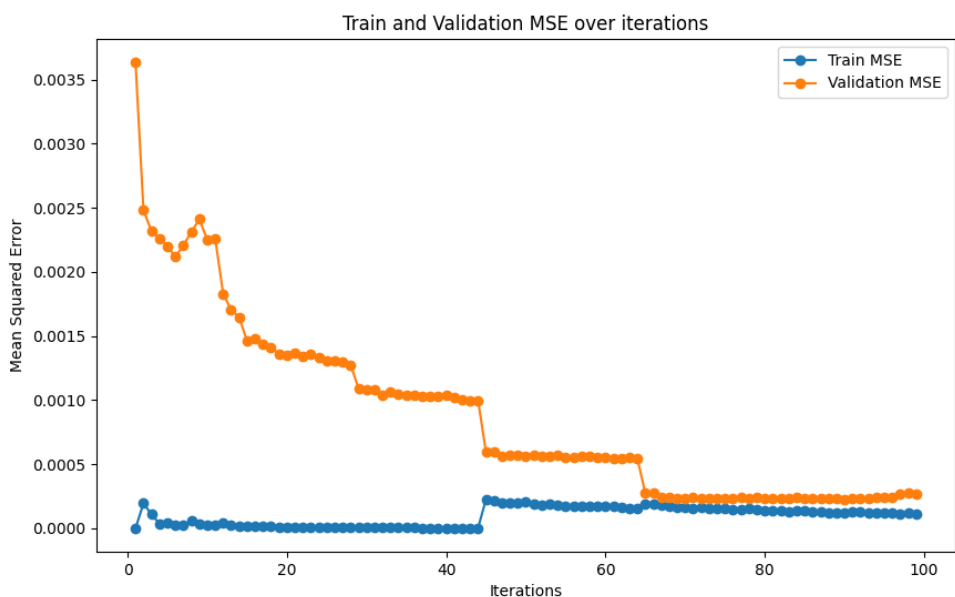
Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Έπειτα εφαρμόστηκε ο αλγόριθμος Random Forest θέτοντας ως αριθμό εκτιμητών ίσο με 300. Να σημειωθεί ότι οι εκτιμητές στον αλγόριθμο Random Forest είναι τα δέντρα απόφασης τα οποία χρησιμοποιηθεί ο αλγόριθμος για να πραγματοποιήσει προβλέψεις.

```
# Train the RandomForestRegressor on the scaled training data
rf_regressor = RandomForestRegressor(n_estimators=300, random_state=42)
rf_regressor.fit(X_train_scaled, y_train)
# Lists to store train and test MSE values
train_mse_list = []
test_mse_list = []
results = []
```

Όσον αφορά τις μετρικές του αλγορίθμου παρουσιάζονται παρακάτω:

Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0002645738523204	0.89529221458703	0.002794452225336	
6745	6	1337	
0.0584%	89.5292%	0.6165%	0.4060%

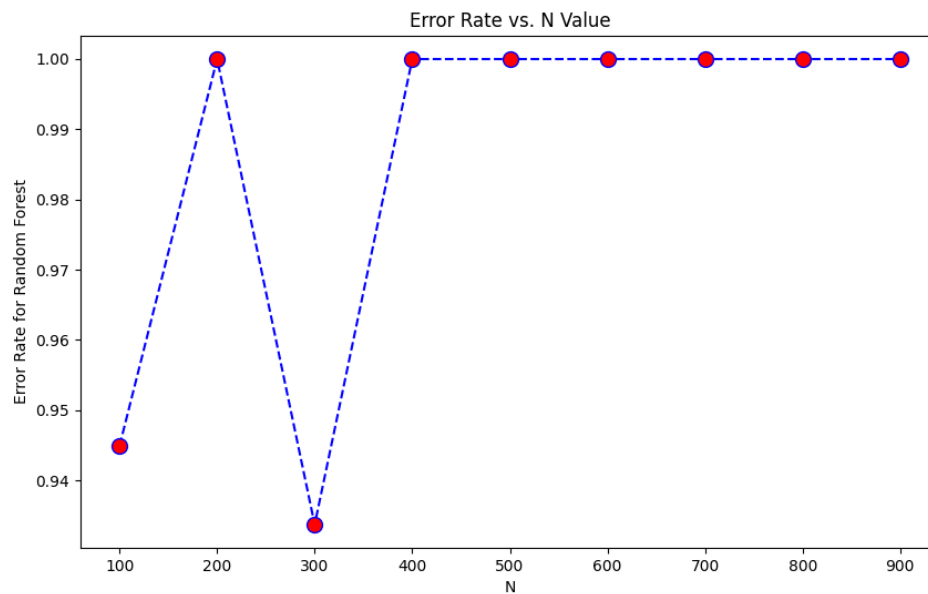


Διάγραμμα 20: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

Για την επιλογή του $n_estimators=300$ όπως παρουσιάζεται στο παρακάτω διάγραμμα επιλέχθηκε γιατί δοκιμάστηκε ο αλγόριθμος για $n_estimators$ από 100

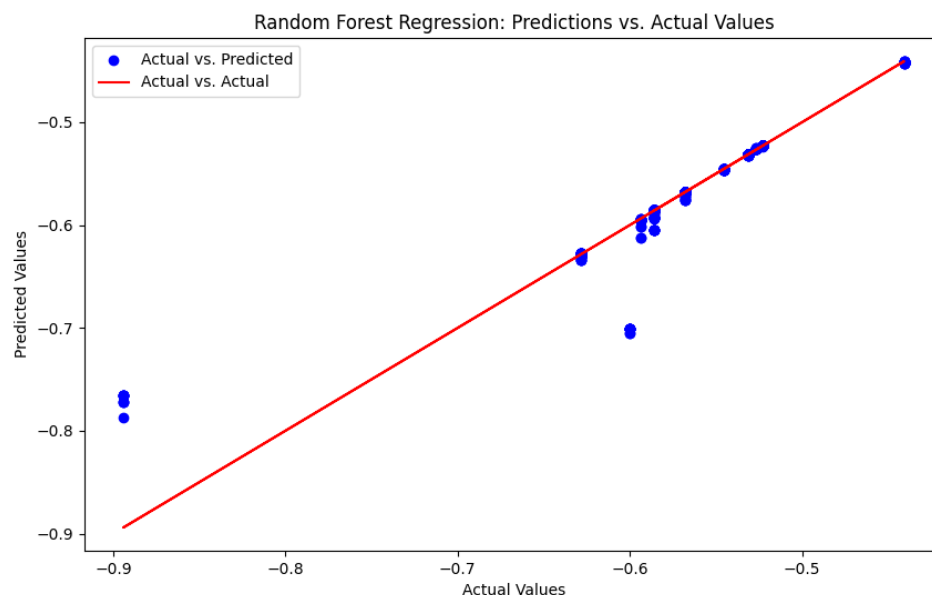
Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

έως 1000 και όπως φαίνεται το $n_estimators=300$ έχει το μικρότερο σφάλμα από όλα τα iterations που εκτελέστηκαν.



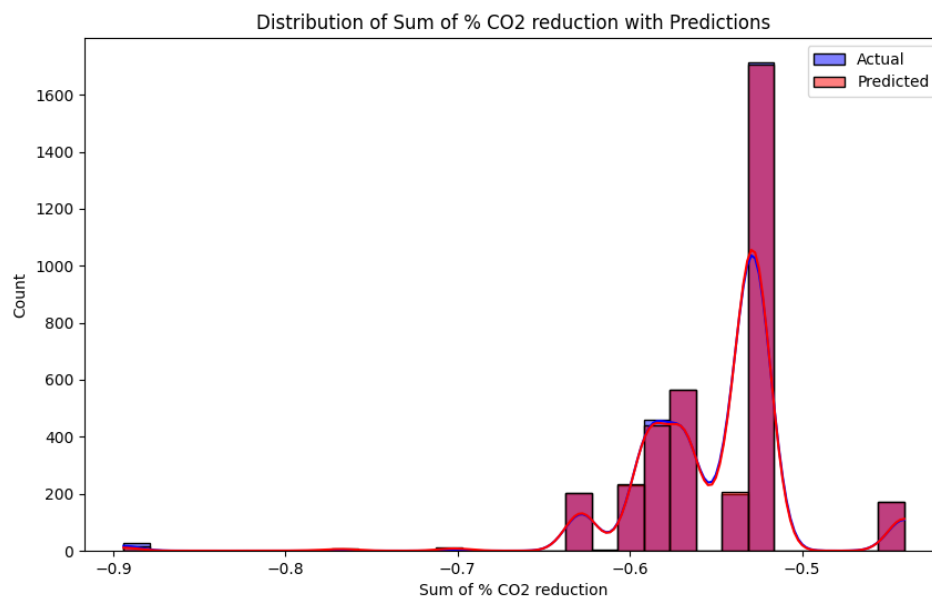
Διάγραμμα 21: Iterations για εύρεση καλύτερου n

Στη συνέχεια παρουσιάζεται το διάγραμμα το οποίο συγκρίνει τις τιμές που πρόβλεψε ο αλγόριθμος συγκριτικά με τις πραγματικές τιμές στην στήλη “Sum of % CO2 reduction”.



Διάγραμμα 22: Προβλέψεις vs Πραγματικές τιμές

Τέλος παρουσιάζεται η κατανομή των πραγματικών τιμών της μείωσης διοξειδίου του άνθρακα συγκριτικά με τις προβλέψεις.

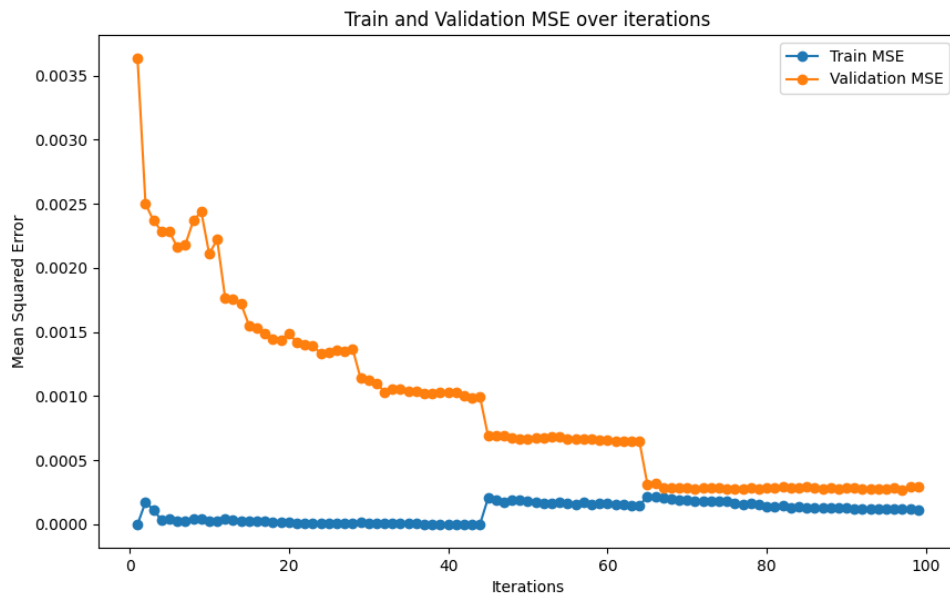


Διάγραμμα 23: Κατανομή του “Sum of % energy reduction” συγκριτικά με τις προβλέψεις

Πέρα από τη στήλη “Sum of % CO2 reduction” με τις εκπομπές διοξειδίου του άνθρακα ο αλγόριθμος Random Forest εκτελέστηκε και για τις τιμές για την στήλη “Sum of % energy reduction”. Οι μετρικές απόδοσης φαίνονται παρακάτω.

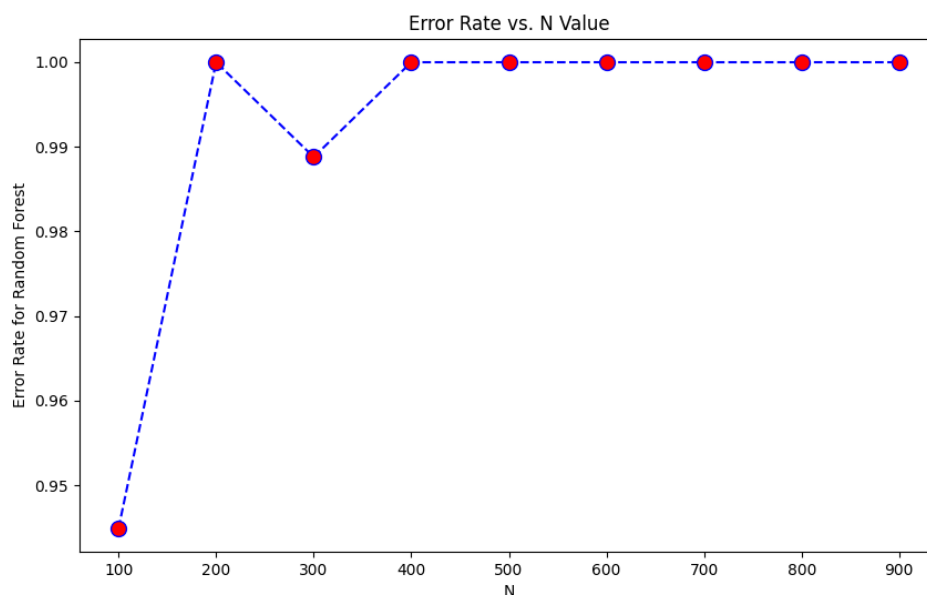
Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0002982072739495782	0.881981446860908	0.002858106442577554	
0.0658%	88.1981%	0.6305%	0.4176%

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 24: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

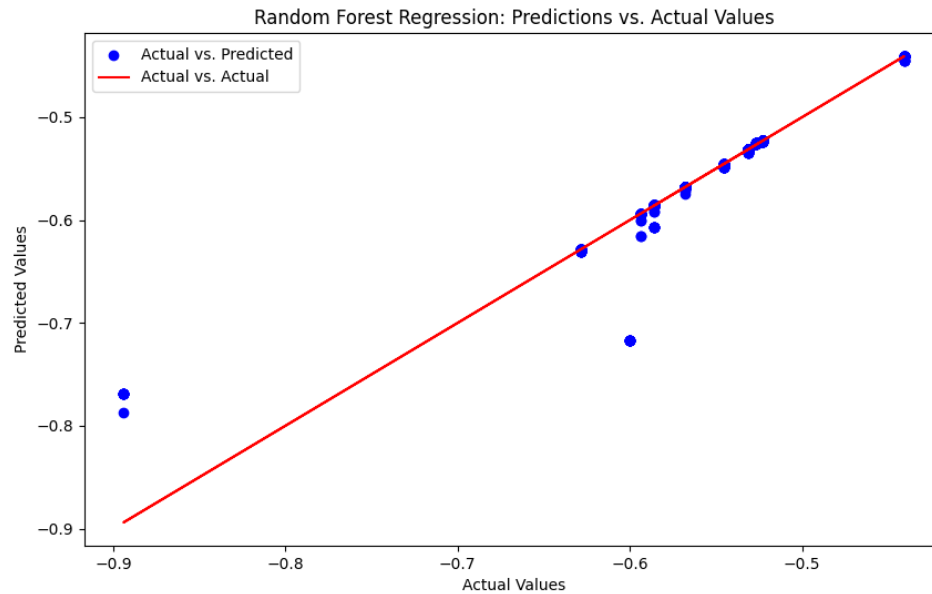
Όπως αναφέρθηκε και προηγουμένως στις προβλέψεις για τη στήλη "Sum of % CO2 reduction" για την επιλογή της τιμής του $n_estimators$ όπως παρουσιάζεται στο παρακάτω διάγραμμα επιλέχθηκε η τιμή 100 για τις προβλέψεις που αφορούν τη στήλη "Sum of % energy reduction" γιατί δοκιμάστηκε ο αλγόριθμος για $n_estimators$ από 100 έως 1000 και όπως φαίνεται το $n_estimators=100$ έχει το μικρότερο σφάλμα από όλα τα iterations που εκτελέστηκαν.



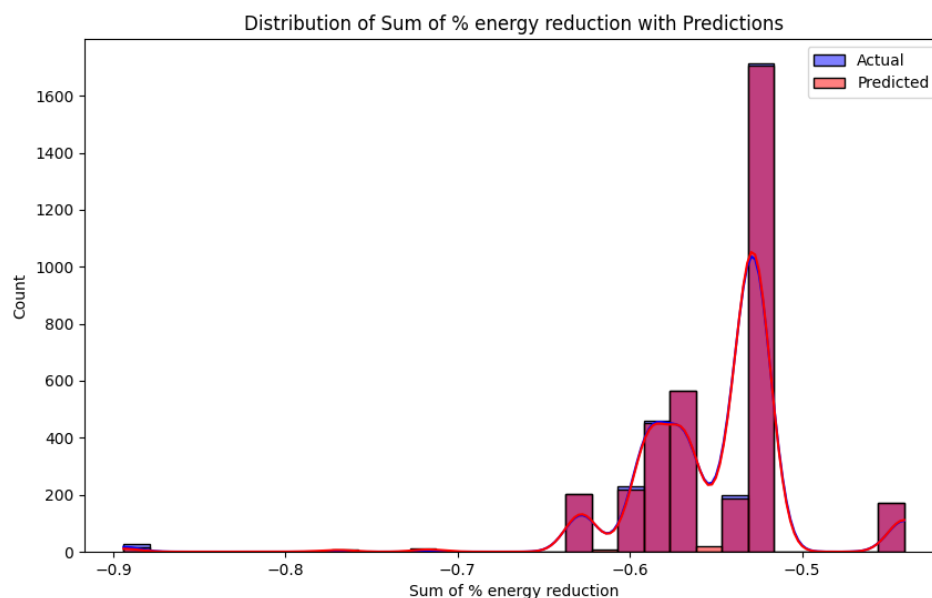
Διάγραμμα 25: Iterations για την επιλογή βέλτιστου n

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Στη συνέχεια παρουσιάζεται ένα διάγραμμα σύγκρισης των πραγματικών τιμών με τις προβλέψεις αλλά και ένα διάγραμμα κατανομή των πραγματικών τιμών της μείωσης της ενεργειακής κατανάλωσης συγκριτικά με τις προβλέψεις.



Διάγραμμα 26: : Προβλέψεις vs Πραγματικές τιμές



Διάγραμμα 27: : Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις

4.3.3 *Lightgbm*

Η μέθοδος Gradient boosting είναι ένας τύπος μεθόδου μηχανικής μάθησης που χρησιμοποιείται για ταξινόμηση και παλινδρόμηση. Ενσωματώνει μοντέλα από πολλές μεθόδους για τη δημιουργία ενός νέου επαναληπτικού μοντέλου. Το Gradient boosting είναι μια από τις πιο συχνά χρησιμοποιούμενες μεθόδους μηχανικής μάθησης λόγω της υψηλής ακρίβειας και αποδοτικότητάς της. Ξεκίνησε με το Adaptive Boosting (AdaBoost), το οποίο εξελίχθηκε σε μια ποικιλία αλγορίθμων και μεθοδολογιών, συμπεριλαμβανομένων των GBM και Model-Based Boosting (MBoost), καθώς και των CatBoost, XGBoost και LightGBM.

Το LightGBM είναι ένα νέο πλαίσιο βαθμωτής ενίσχυσης βασισμένο σε τεχνικές δέντρων απόφασης που αναπτύχθηκε από τη Microsoft. Υποστηρίζει διάφορους αλγορίθμους, συμπεριλαμβανομένων των GBM, GBDT και Gradient Boosted Regression. Χρησιμοποιούνται Gradient Boosted Regression Tree (GBRT) και Multiple Additive Regression Tree (MART) για να παρέχουν επεκτασιμότητα, ακρίβεια και αποτελεσματικότητα. Σε αυτή την τεχνική, το δέντρο απόφασης αναπτύσσεται κατά φύλλο, γεγονός που βελτιστοποιεί την απώλεια που δημιουργεί κλαδιά. Η βασική ιδέα πίσω από το LightGBM είναι να εκπαιδεύσει ένα σύνολο από απλά δέντρα αποφάσεων (decision trees) και να συνδυάσει τα αποτελέσματά τους για να παράγει την τελική πρόβλεψη. Κατά τη διάρκεια της εκπαίδευσης, το LightGBM χρησιμοποιεί μια στρατηγική που επικεντρώνεται στα δείγματα που είναι πιο δύσκολα να εκπαιδευτούν, επικεντρώνοντας στα σημεία του χώρου των χαρακτηριστικών όπου ο αλγόριθμος κάνει λάθη. Ως αποτέλεσμα, αυτή η προσέγγιση είναι ταχύτερη και λιγότερο περίπλοκη από την ανάπτυξη κατά επίπεδο, η οποία αυξάνει το βάθος του δέντρου [71].

Για την εφαρμογή του αλγορίθμου Lightgbm χρησιμοποιήθηκαν τα αρχικά δεδομένα τα οποία χρησιμοποιήθηκαν και στους προηγούμενους αλγορίθμους. Αρχικά, οι στήλες του DataFrame (df) μετονομάζονται χρησιμοποιώντας ένα λειτουργικό της βιβλιοθήκης rename για να αφαιρέσουν τυχόν ειδικούς χαρακτήρες που δεν είναι γράμματα, αριθμοί ή κάτω παύλες. Η μετονομασία γίνεται γιατί ο Lightgbm δεν μπορεί να χειριστεί στήλες που περιέχουν ειδικούς χαρακτήρες ή κενά. Στη συνέχεια, δημιουργείται ένας χάρτης αντιστοίχισης μεταξύ των αρχικών και μετονομασμένων στηλών.

```
df = df.rename(columns = lambda x:re.sub('[^A-Za-z0-9_]+', '', x))  
# Create a mapping between original and renamed columns  
column_mapping = {new_name: original_name for original_name, new_name in  
zip(original_column_names,  
df.columns)}
```

Τα δεδομένα διαχωρίζονται σε σύνολα εκπαίδευσης (X_train, y_train) και ελέγχου (X_test, y_test) χρησιμοποιώντας τη συνάρτηση train_test_split. Ο χειρισμός των τελικών δεδομένων γίνεται με τον SimpleImputer, χρησιμοποιώντας τη μέση τιμή (strategy='mean') για την αντικατάσταση των τιμών που λείπουν στο σύνολο

εκπαίδευσης και στο σύνολο ελέγχου. Τέλος χρησιμοποιείται η GridSearchCV για την αναζήτηση των βέλτιστων υπερπαραμέτρων για το μοντέλο LightGBM, με βάση ένα πλέγμα πιθανών τιμών (param_grid) και το μοντέλο εκπαιδεύεται με τις βέλτιστες υπερπαραμέτρους χρησιμοποιώντας τα δεδομένα εκπαίδευσης ενώ συγχρόνως εφαρμόζεται πρόωρος τερματισμός (early_stopping) για να αποφευχθεί η υπερβολική εκπαίδευση. Αρχικά το μοντέλο εκπαιδεύτηκε για τη στήλη "Sum of % CO2 reduction".

```
# Separate features and target variable
X = df.drop('SumofCO2reduction', axis=1) # Features (all columns except the target column)
y = df['SumofCO2reduction'] # Target variable
# Splitting the dataset into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Assuming X_train and y_train have NaN values
imputer = SimpleImputer(strategy='mean')
X_train_imputed = imputer.fit_transform(X_train)
X_test_imputed = imputer.transform(X_test) # Apply imputer transformation to the test data
# Handle missing values in the target variable
y_train_imputed = imputer.fit_transform(y_train.values.reshape(-1, 1)).ravel()
y_test_imputed = imputer.transform(y_test.values.reshape(-1, 1)).ravel()
# Hyperparameter tuning with GridSearchCV
param_grid = {
    'num_leaves': [20, 30, 40],
    'learning_rate': [0.05, 0.1, 0.2],
    'n_estimators': [50, 100, 200],
    'reg_alpha': [0.1, 0.2, 0.5],
    'reg_lambda': [0.1, 0.2, 0.5],
}
model = lgb.LGBMRegressor()
grid_search = GridSearchCV(model, param_grid, cv=3, scoring='neg_mean_squared_error')
grid_search.fit(X_train_imputed, y_train_imputed)
best_params = grid_search.best_params_
eval_set = [(X_test, y_test)]
# Training the model with best parameters
best_model = lgb.LGBMRegressor(**best_params)
best_model.fit(X_train_imputed, y_train_imputed, eval_set=eval_set, eval_metric='rmse', callbacks=[lgb.early_stopping(stopping_rounds=10)])
# Making predictions on the test set
predicted_y = best_model.predict(X_test_imputed)
```


Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

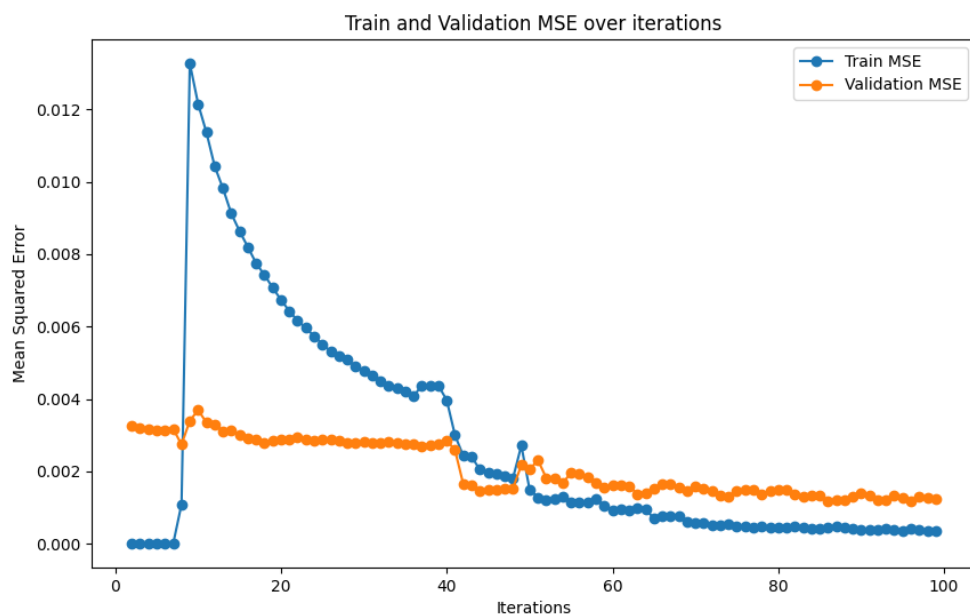
Ο αλγόριθμος επιστρέφει και τις καλύτερες παραμέτρους που επέλεξε για την εκτέλεση του.

Best Hyperparameters: {'learning_rate': 0.2, 'n_estimators': 200, 'num_leaves': 20, 'reg_alpha': 0.1, 'reg_lambda': 0.1}

Οι μετρικές απόδοσης φαίνονται παρακάτω για τον συγκεκριμένο αλγόριθμο φαίνονται παρακάτω.

Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0045	0.9923	0.0010	
0.9901463472167101 %	99.2298272831528 6%	0.227568049654496 62%	0.1605%

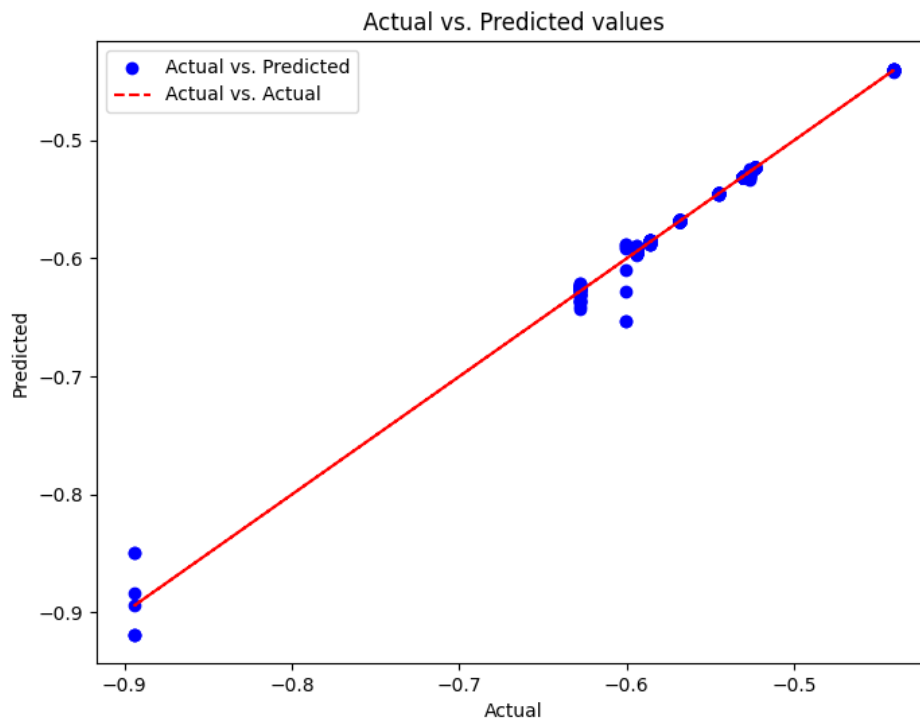
Εν συνεχεία παρουσιάζεται το διάγραμμα που αφορά το σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα.



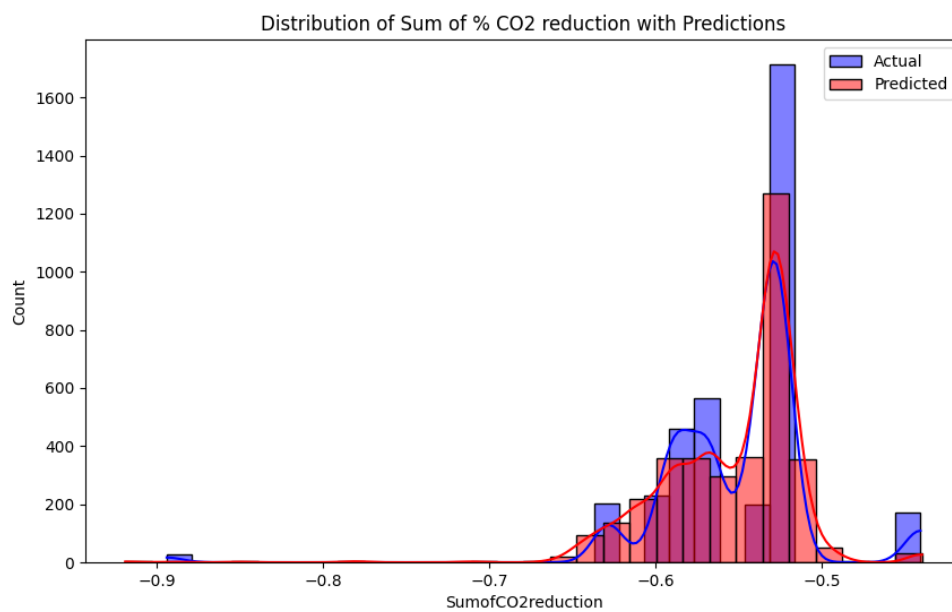
Διάγραμμα 28: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

Έπειτα παρατίθενται ένα διάγραμμα το οποίο συγκρίνει τις τιμές που προβλέφθηκαν με τις πραγματικές τιμές αλλά και ένα διάγραμμα για την κατανομή του “Sum of % CO2 reduction” συγκριτικά με τις προβλέψεις.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

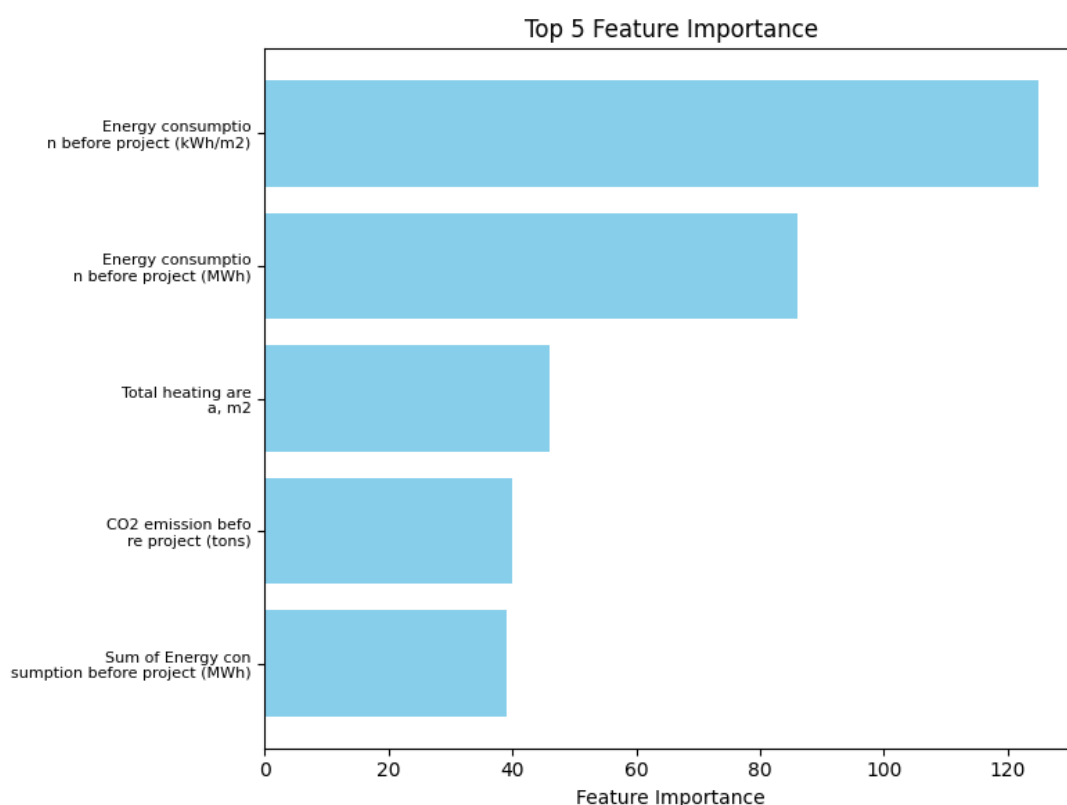


Διάγραμμα 29: Προβλέψεις vs Πραγματικές τιμές



Διάγραμμα 30: Κατανομή του “Sum of % CO2 reduction” συγκριτικά με τις προβλέψεις

Αξίζει να σημειωθεί ότι για τον συγκεκριμένο αλγόριθμο εξετάστηκαν και τα χαρακτηριστικά τα οποία είχαν την μεγαλύτερη επιρροή στο μοντέλο Lightgbm . Το feature importance (σημασία των χαρακτηριστικών) είναι ένα μέτρο που χρησιμοποιείται για να κατανοηθεί ποια χαρακτηριστικά των δεδομένων είναι πιο σημαντικά για το μοντέλο μηχανικής μάθησης κατά τη λήψη των αποφάσεών του. Για το LightGBM, η σημασία των χαρακτηριστικών υπολογίζεται με βάση τον τρόπο με τον οποίο τα χαρακτηριστικά χρησιμοποιούνται κατά τη διαδικασία του δέντρου απόφασης που εκπαιδεύεται το μοντέλο. Τα 5 χαρακτηριστικά με την μεγαλύτερη σημασία για το μοντέλο που εφαρμόστηκε είναι τα παρακάτω:

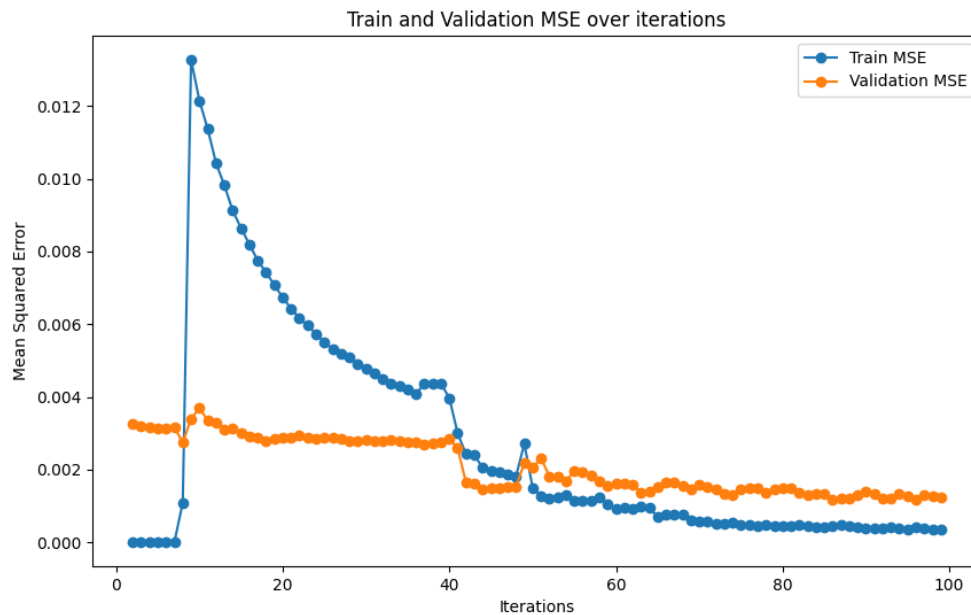


Διάγραμμα 31: Τα 5 χαρακτηριστικά με τη μεγαλύτερη σημασία για το μοντέλο

Όπως και για τους δύο προηγούμενους αλγόριθμους πέρα από τη στήλη “Sum of % CO2 reduction” με τις εκπομπές διοξειδίου του άνθρακα ο αλγόριθμος Lightgbm εκτελέστηκε και για τις τιμές για την στήλη “Sum of % energy reduction”. Οι μετρικές απόδοσης φαίνονται παρακάτω.

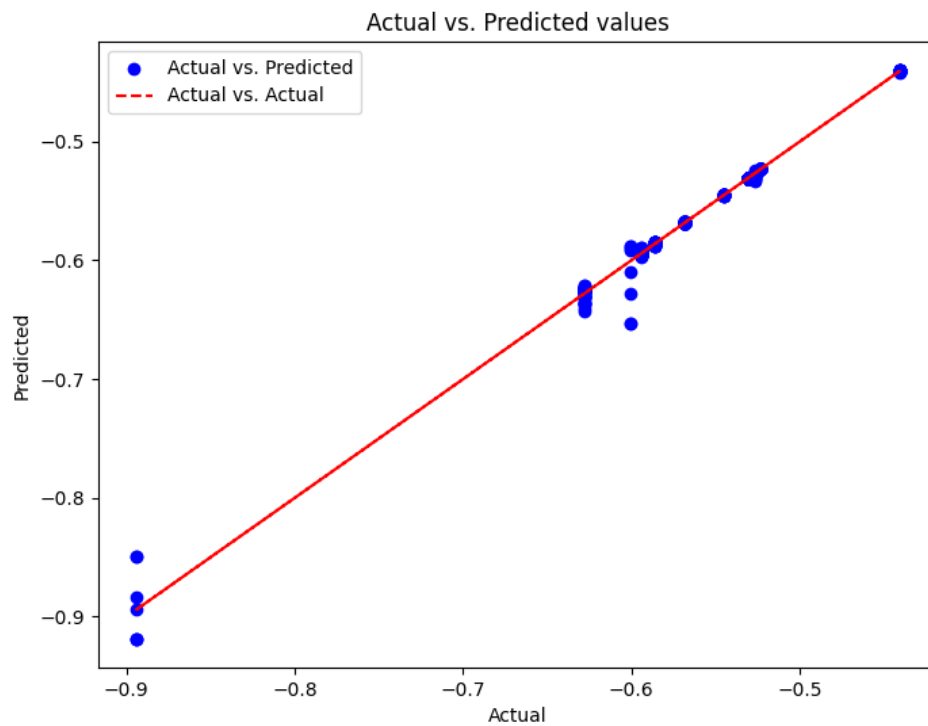
Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0045	0.9923	0.0010	
0.9901463472167101 %	99.2298272831528 6%	0.227568049654496 62%	0.1605%

Όπως παρατηρείται ο αλγόριθμος επέστρεψε παρόμοια αποτελέσματα με τη στήλη “Sum of % CO2 reduction”. Στη συνέχεια παρατίθεται τα διαγράμματα που αφορούν το σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα, τη σύγκριση των πραγματικών τιμών με τις προβλέψεις αλλά και την κατανομή της στήλης “Sum of % energy reduction” συγκριτικά με τις προβλέψεις.

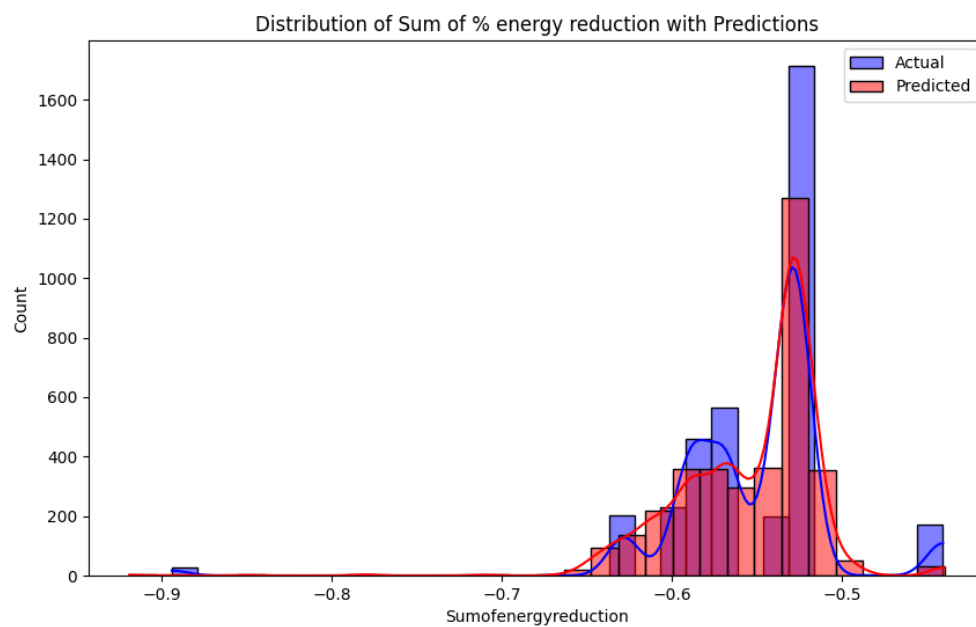


Διάγραμμα 32: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 33: Προβλέψεις vs Πραγματικές τιμές



Διάγραμμα 34: Κατανομή του “Sum of % energy reduction” συγκριτικά με τις προβλέψεις

4.3.4 *Lightgbm (v.2)*

Ο αλγόριθμος Lightgbm πέρα από την εκτέλεση με τα αρχικά δεδομένα επιλέχθηκε να εκτελεστεί και με τις στήλες που ήταν περισσότερο συσχετισμένες μεταξύ τους. Δηλαδή πρώτα μελετήθηκαν τα δεδομένα και επιλέχθηκαν τα χαρακτηριστικά με τη μεγαλύτερη συσχέτιση θέτοντας ως κατώφλι ίσο με 0.7 και εν συνεχεία εκπαιδεύτηκε το μοντέλο. Με τον τρόπο αυτό επιδιώχθηκε η βελτίωση του μοντέλου στις προβλέψεις και κατά συνέχεια και η μείωση των σφαλμάτων.

```
# Calculate the correlation matrix for all columns
correlation_matrix = df.corr()
# Set your threshold for correlation coefficient
threshold = 0.7
high_corr_cols = set()
# Iterate through columns in the correlation matrix to find highly correlated columns
for i in range(len(correlation_matrix.columns)):
    for j in range(i):
        if abs(correlation_matrix.iloc[i, j]) > threshold:
            colname_i = correlation_matrix.columns[i]
            colname_j = correlation_matrix.columns[j]
            high_corr_cols.add(colname_i)
            high_corr_cols.add(colname_j)
# Create a new dataframe with columns having high correlations using original
column names
subset_df = df[list(high_corr_cols)]
# Get the correlation matrix for the subset of column
subset_corr_matrix = subset_df.corr()
```

Εφόσον επιλέχθηκαν τα στοιχεία με τη μεγαλύτερη συσχέτιση μεταξύ τους έπειτα εκπαιδεύτηκε το μοντέλο όπως ακριβώς και στην προηγούμενη έκδοση του Lightgbm. Τα δεδομένα διαχωρίζονται σε σύνολα εκπαίδευσης (X_train_filtered, y_train_filtered) και ελέγχου (X_test_filtered, y_test_filtered) χρησιμοποιώντας τη συνάρτηση train_test_split. Ο χειρισμός των τελικών δεδομένων γίνεται με τον SimpleImputer, χρησιμοποιώντας τη μέση τιμή (strategy='mean') για την αντικατάσταση των τιμών που λείπουν στο σύνολο εκπαίδευσης και στο σύνολο ελέγχου. Τέλος χρησιμοποιείται η GridSearchCV για την αναζήτηση των βέλτιστων υπερπαραμέτρων για το μοντέλο LightGBM, με βάση ένα πλέγμα πιθανών τιμών (param_grid) και το μοντέλο εκπαιδεύεται με τις βέλτιστες υπερπαραμέτρους χρησιμοποιώντας τα δεδομένα εκπαίδευσης ενώ συγχρόνως εφαρμόζεται πρόωρος τερματισμός (early_stopping) για να αποφευχθεί η υπερβολική εκπαίδευση.

```
Get the original column names for highly correlated columns
original_high_corr_cols = [column_mapping[col] for col in high_corr_cols]
# Filter the dataset to include only columns with high correlations
df_filtered = df[list(high_corr_cols)]
```

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

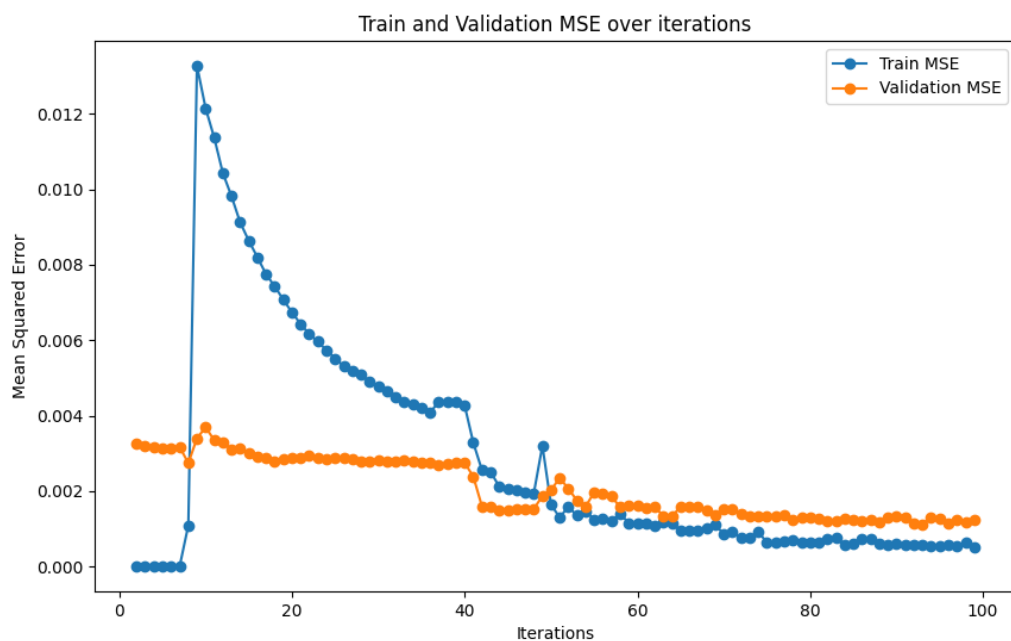
```
# Splitting the dataset into training and testing sets using only highly correlated columns
X_filtered = df_filtered.drop('SumofCO2reduction', axis=1) # Features (all columns except the target column)
y_filtered = df_filtered['SumofCO2reduction'] # Target variable
# Splitting the dataset into training and testing sets
X_train_filtered, X_test_filtered, y_train_filtered, y_test_filtered = train_test_split(X_filtered, y_filtered, test_size=0.2, random_state=42)
# Assuming X_train_filtered and y_train_filtered have NaN values
imputer_filtered = SimpleImputer(strategy='mean')
X_train_imputed_filtered = imputer_filtered.fit_transform(X_train_filtered)
X_test_imputed_filtered = imputer_filtered.transform(X_test_filtered) # Apply imputer transformation to the test data
# Handle missing values in the target variable
y_train_imputed_filtered = imputer_filtered.fit_transform(y_train_filtered.values.reshape(-1, 1)).ravel()
y_test_imputed_filtered = imputer_filtered.transform(y_test_filtered.values.reshape(-1, 1)).ravel()
param_grid = {
    'num_leaves': [20, 30, 40],
    'learning_rate': [0.05, 0.1, 0.2],
    'n_estimators': [50, 100, 200],
    'reg_alpha': [0.1, 0.2, 0.5],
    'reg_lambda': [0.1, 0.2, 0.5],
}
model = lgb.LGBMRegressor()
grid_search = GridSearchCV(model, param_grid, cv=3, scoring='neg_mean_squared_error')
grid_search.fit(X_train_imputed_filtered, y_train_imputed_filtered)
best_params = grid_search.best_params_
eval_set = [(X_test_filtered, y_test_filtered)]
# Training the model with best parameters using the filtered dataset
best_model_filtered = lgb.LGBMRegressor(**best_params)
best_model_filtered.fit(X_train_imputed_filtered, y_train_imputed_filtered, eval_set=eval_set, eval_metric='rmse', callbacks=[lgb.early_stopping(stopping_rounds=10)])
# Making predictions on the filtered test set
predicted_y_filtered = best_model_filtered.predict(X_test_imputed_filtered)
```

Αξίζει να σημειωθεί ότι τα features με τη μεγαλύτερη συσχέτιση που χρησιμοποιήθηκαν εν τέλει είναι: *'EnergyconsumptionbeforeprojectMWh', 'SumofEnergyconsumptionbeforeprojectMWh', 'SumofCO2emissionbeforeprojectttons', 'Totalheatingaream2', Sumofenergyreduction', 'CO2emissionbeforeprojectttons', 'SumofCO2reduction'.*

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

Οι μετρικές απόδοσης για την συγκεκριμένη έκδοση του αλγορίθμου Lightgbm παρουσιάζονται παρακάτω.

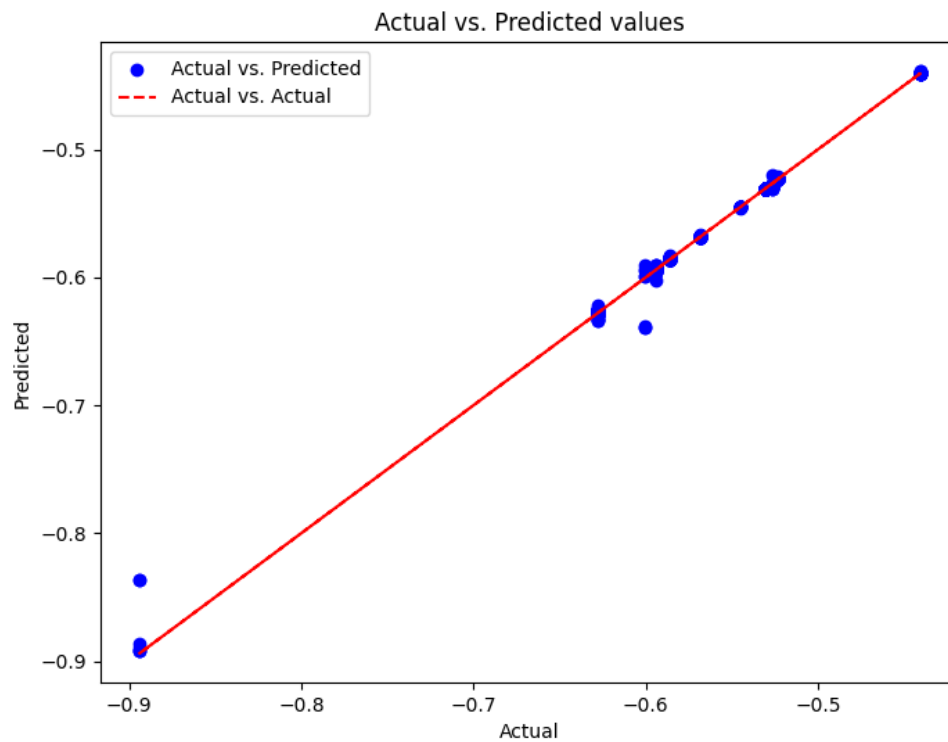
Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0041	0.9935	0.0008	
0.9086%	99.3514%	0.1736%	0.1246%



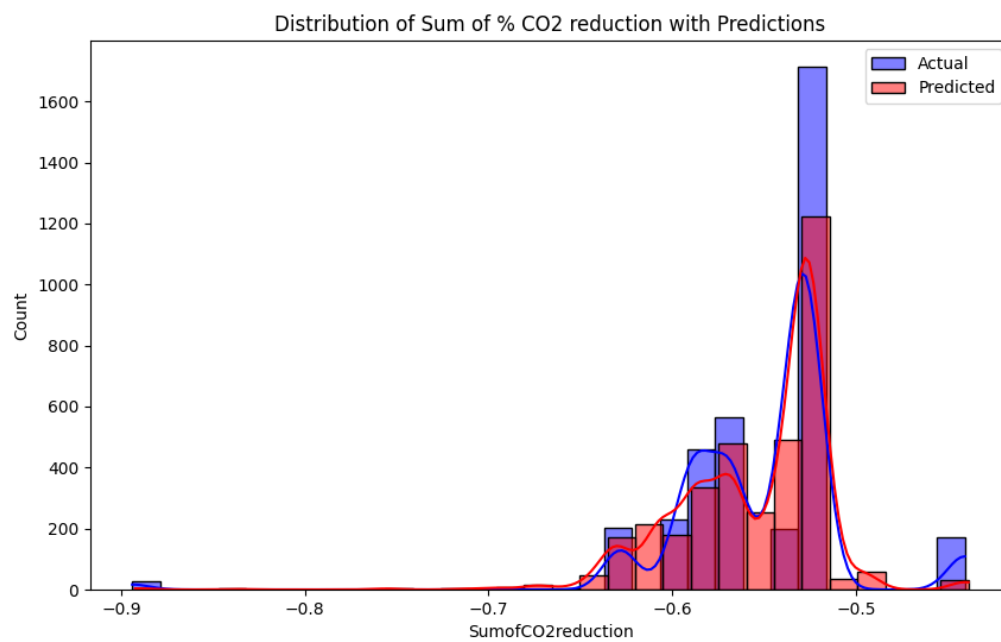
Διάγραμμα 35: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

Στη συνέχεια παρουσιάζονται ένα διάγραμμα το οποίο συγκρίνει τις τιμές που προβλέφθηκαν με τις πραγματικές τιμές αλλά και ένα διάγραμμα για την κατανομή του “Sum of % CO2 reduction” συγκριτικά με τις προβλέψεις.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

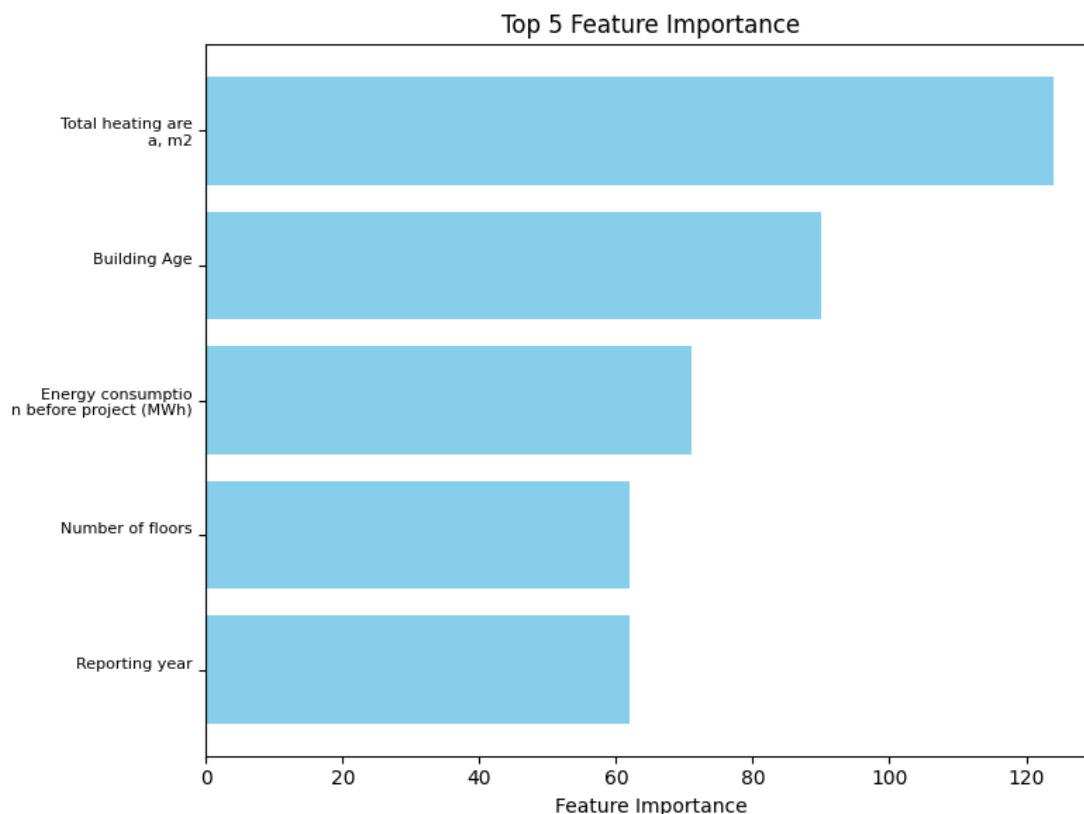


Διάγραμμα 36: Προβλέψεις vs Πραγματικές τιμές



Διάγραμμα 37: Κατανομή του "Sum of % CO2 reduction" συγκριτικά με τις προβλέψεις

Όπως προαναφέρθηκε και στο μοντέλο του Lightgbm κατά την εκπαίδευση του οποίου χρησιμοποιήθηκαν όλα τα αρχικά δεδομένα έτσι και εδώ αξιολογήθηκε το feature importance για τα features που χρησιμοποιήθηκαν. Τα 5 χαρακτηριστικά με τη μεγαλύτερη επιρροή στο μοντέλο είναι :

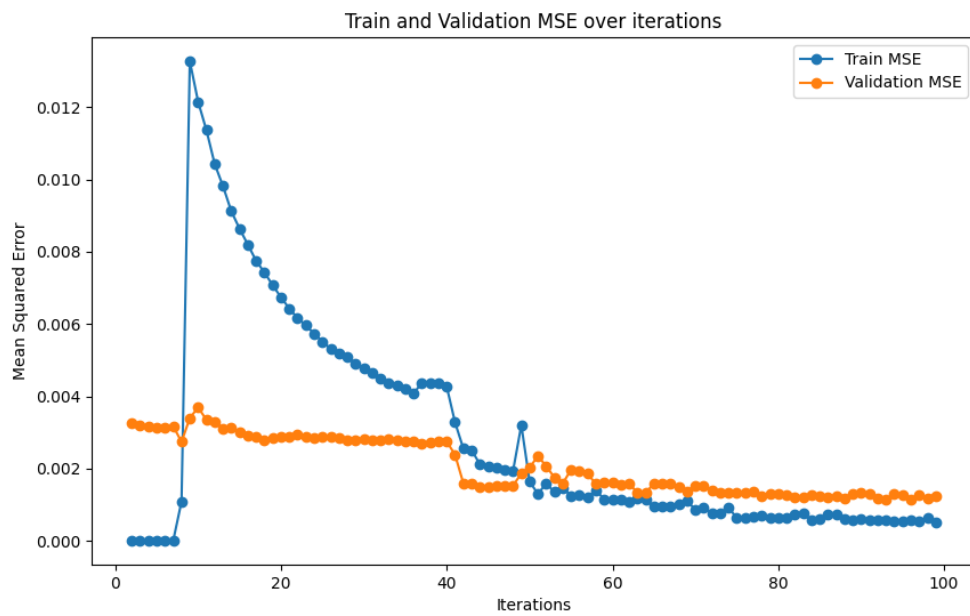


Διάγραμμα 38: Τα 5 χαρακτηριστικά με τη μεγαλύτερη σημασία για το μοντέλο

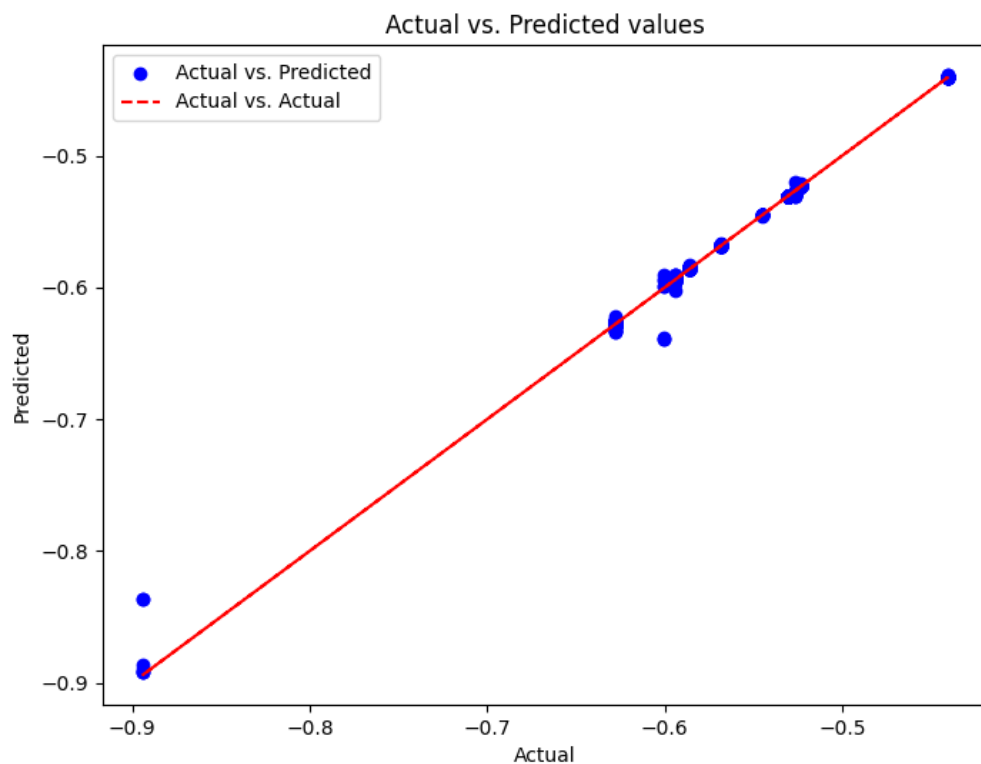
Παρατηρείται ότι στην συγκεκριμένη έκδοση του Lightgbm τα χαρακτηριστικά με τη μεγαλύτερη σημασία για το μοντέλο διαφέρουν από εκείνα που εμφανίστηκαν όταν επιλέχθηκαν όλα τα αρχικά δεδομένα για την εκπαίδευση του. Αυτό οφείλεται στο γεγονός ότι στην προκειμένη περίπτωση χρησιμοποιήθηκαν μόνο features τα οποία είναι περισσότερο συσχετισμένα μεταξύ τους.

Ο αλγόριθμος αυτός εφαρμόστηκε και για τη στήλη “Sum of % energy reduction”. Οι μετρικές απόδοσης που επέστρεψε είναι ίδιες με αυτές για τη στήλη “Sum of % CO2 reduction” ωστόσο παρατίθενται παρακάτω μαζί με το διάγραμμα για το σφάλμα εκπαίδευσης και επικύρωσης, ένα συγκριτικό διάγραμμα των πραγματικών τιμών με τις προβλέψεις και ένα διάγραμμα για την κατανομή του “Sum of % energy reduction” συγκριτικά με τις προβλέψεις του μοντέλου.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

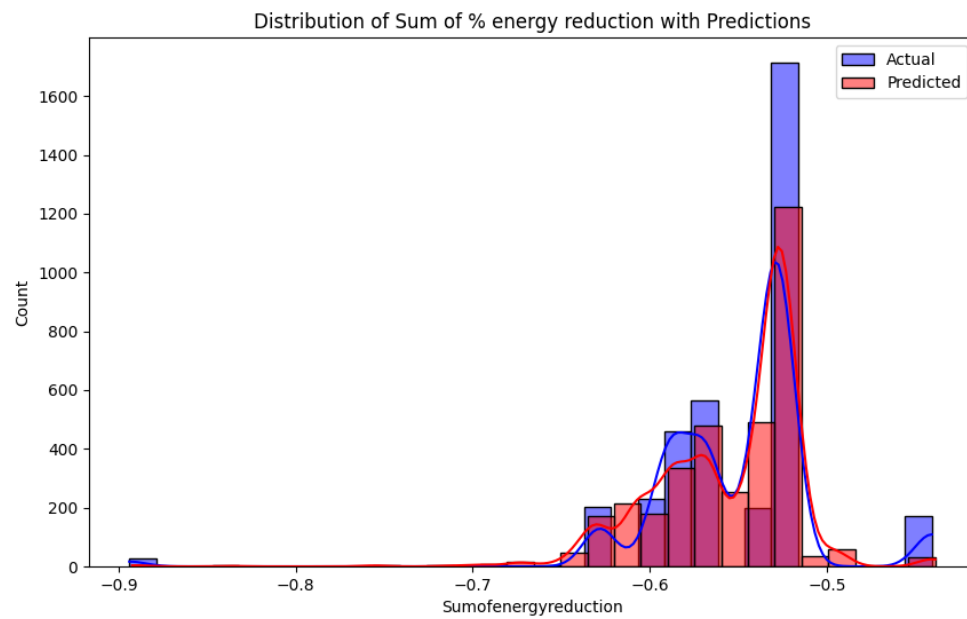


Διάγραμμα 39: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα



Διάγραμμα 40: Προβλέψεις vs Πραγματικές τιμές

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 41: Κατανομή του “Sum of % energy reduction” συγκριτικά με τις προβλέψεις

4.3.5 *Support Vector Machine (SVM)*

Ο αλγόριθμος Support Vector Machine(SVM), επίσης γνωστός ως δίκτυο διανυσμάτων υποστήριξης (support vector network), είναι μια τεχνική μάθησης με επίβλεψη για ταξινόμηση και παλινδρόμηση. Δεδομένης μιας συλλογής εκπαιδευτικών επισημασμένων παραδειγμάτων από δύο κλάσεις, η μέθοδος εκπαίδευσης SVM δημιουργεί ένα όριο απόφασης μεταξύ αυτών των δειγμάτων. Η SVM το κάνει αυτό μεγιστοποιώντας το περιθώριο μεταξύ δύο κατηγοριών δεδομένων. Σε έναν χώρο N διαστάσεων, η SVM δημιουργεί ένα διαχωριστικό υπερεπίπεδο για τη διάκριση μεταξύ δύο κλάσεων [66].

Κατά τη διάρκεια της εκπαίδευσης, το SVM βρίσκει το υπερεπίπεδο που μεγιστοποιεί το περιθώριο, το οποίο είναι η ελάχιστη απόσταση μεταξύ του υπερεπιπέδου και των πλησιέστερων δεδομένων (support vectors). Η διατύπωση του προβλήματος είναι μια μαθηματική βελτιστοποίηση που στοχεύει στην εύρεση αυτού του υπερεπιπέδου. Η μαθηματική διατύπωση του προβλήματος του SVM συνήθως περιλαμβάνει εξισώσεις για την εύρεση του βέλτιστου υπερεπιπέδου και των παραμέτρων που απαιτούνται. Οι εξισώσεις αυτές χρησιμοποιούν μεθόδους βελτιστοποίησης όπως η κλίση και οι μέθοδοι Newton-Raphson για να επιτύχουν την ελαχιστοποίηση μιας συνάρτησης κόστους ή το μέγιστο του περιθωρίου, ανάλογα με το πρόβλημα ταξινόμησης ή παλινδρόμησης.

Για την εφαρμογή του αλγορίθμου SVM χρησιμοποιήθηκαν τα αρχικά δεδομένα χωρίς να υποστούν κάποια επεξεργασία. Ωστόσο αξίζει να σημειωθεί ότι χρησιμοποιήθηκαν οι πρώτες 900 γραμμές δεδομένων καθώς από οποιονδήποτε αριθμό πάνω από αυτό ο αλγόριθμος γινόταν πολύ αργός και καθυστερούσε αρκετά να παράγει τις προβλέψεις. Συγχρόνως επειδή ακριβώς ο αλγόριθμος καθυστερούσε την εκτέλεση του χρησιμοποιήθηκαν τα καλύτερα 5 χαρακτηριστικά.

Επομένως αρχικά, τα δεδομένα διαιρούνται σε χαρακτηριστικά (X) και μεταβλητή-στόχο (y). Τα χαρακτηριστικά είναι οι στήλες του πίνακα *subset_data* χωρίς τη στήλη “Sum of % CO2 reduction”, ενώ η μεταβλητή-στόχος είναι η στήλη “Sum of % CO2 reduction”. Στη συνέχεια, ορίζεται ένα πλέγμα υπερπαραμέτρων για την αναζήτηση των βέλτιστων υπερπαραμέτρων του μοντέλου SVM (Support Vector Machine). Οι υπερπαραμέτροι περιλαμβάνουν το C (παράμετρος κανονικοποίησης), τον πυρήνα (kernel), και το γκάμμα (gamma), που είναι η παράμετρος πυρήνα για τον πυρήνα 'rbf' (RBF kernel) και τα δεδομένα διαιρούνται σε σύνολα εκπαίδευσης και ελέγχου.

```
subset_data = df.head(900) #Let's have a subset of the data
# Setting a target variable
X = subset_data.drop('Sum of % CO2 reduction', axis=1) # Features
y = subset_data['Sum of % CO2 reduction'] # Target variable
#Define a grid of hyperparameters to search (example values)
params = {
```

```
'C': 0.1, # Regularization parameter
'kernel': 'linear', # Kernel type
'gamma': 'auto' # Kernel coefficient for 'rbf' kernel
}
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
```

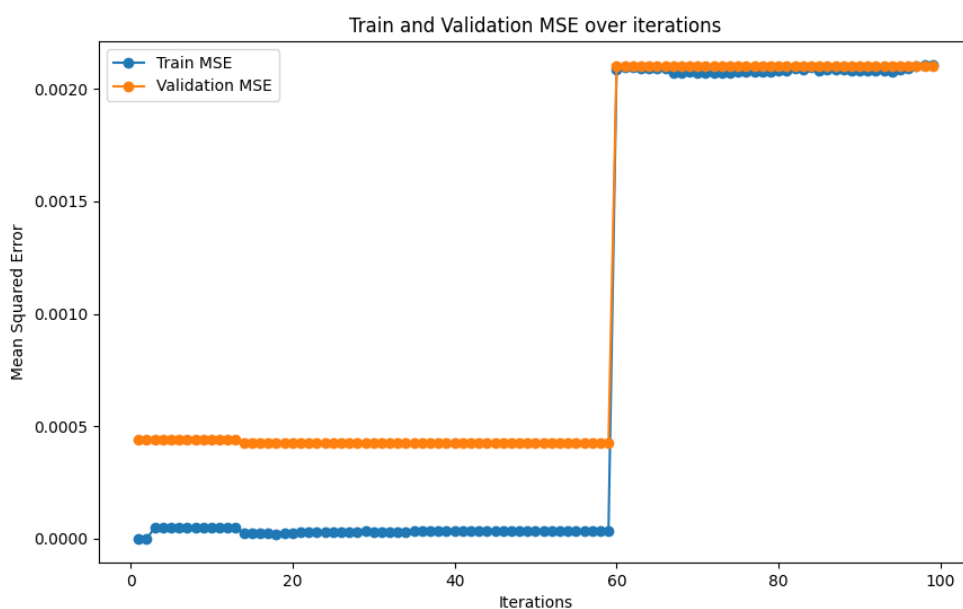
Έπειτα χρησιμοποιείται η μέθοδος *SelectKBest* για να επιλεγούν τα κορυφαία χαρακτηριστικά. Η μέθοδος *SelectKBest* αποτελεί ένα εργαλείο επιλογής χαρακτηριστικών στο πλαίσιο της μηχανικής μάθησης. Χρησιμοποιείται για να επιλέξει ένα υποσύνολο των κορυφαίων χαρακτηριστικών με βάση κάποιο σκορ που υπολογίζεται για κάθε χαρακτηριστικό. Το σκορ μπορεί να είναι οποιοδήποτε μέτρο που μπορεί να χρησιμοποιηθεί για την αξιολόγηση της σημαντικότητας των χαρακτηριστικών. Για προβλήματα παλινδρόμησης όπως αυτό που εξετάζεται συνήθως χρησιμοποιείται η *f_regression*, η οποία εκτελεί μια ανάλυση της διακύμανσης (ANOVA) μεταξύ κάθε χαρακτηριστικού και της μεταβλητής-στόχου. Στη συνέχεια, τα χαρακτηριστικά ταξινομούνται με βάση το σκορ που υπολογίστηκε και επιλέγονται τα κορυφαία *k* χαρακτηριστικά, όπου *k* είναι ο αριθμός των χαρακτηριστικών που επιθυμούμε να επιλέξουμε. Στη συνέχεια εκπαιδεύεται το μοντέλο και γίνονται προβλέψεις στα δεδομένα ελέγχου.

```
# Use SelectKBest to select top k features
k = 5 # Number of top features to select
selector = SelectKBest(score_func=f_regression, k=k) # Using f_regression for
regression task
X_train_selected = selector.fit_transform(X_train, y_train)
X_test_selected = selector.transform(X_test)
# Get the mask of selected features
selected_mask = selector.get_support()
# Retrieve the indices/column names of the selected features
selected_features_indices = [i for i, selected in enumerate(selected_mask) if
selected]
selected_features_names = X.columns[selected_features_indices]
print("Selected Feature Indices:", selected_features_indices)
print("Selected Feature Names:", selected_features_names)
# Create an SVR model
svm_regressor = SVR(**params)
# Fit the SVR model on the selected features
svm_regressor.fit(X_train_selected, y_train)
# Make predictions using the SVR model
y_pred = svm_regressor.predict(X_test_selected)
```

Οι μετρικές απόδοσης του αλγορίθμου SVM παραθέτονται στον επόμενο πίνακα.

Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0021025980555555 526	-3.923946606328 7054	0.045609444444444 41	
1.9987%	0.0000%	43.3550%	8.5394%

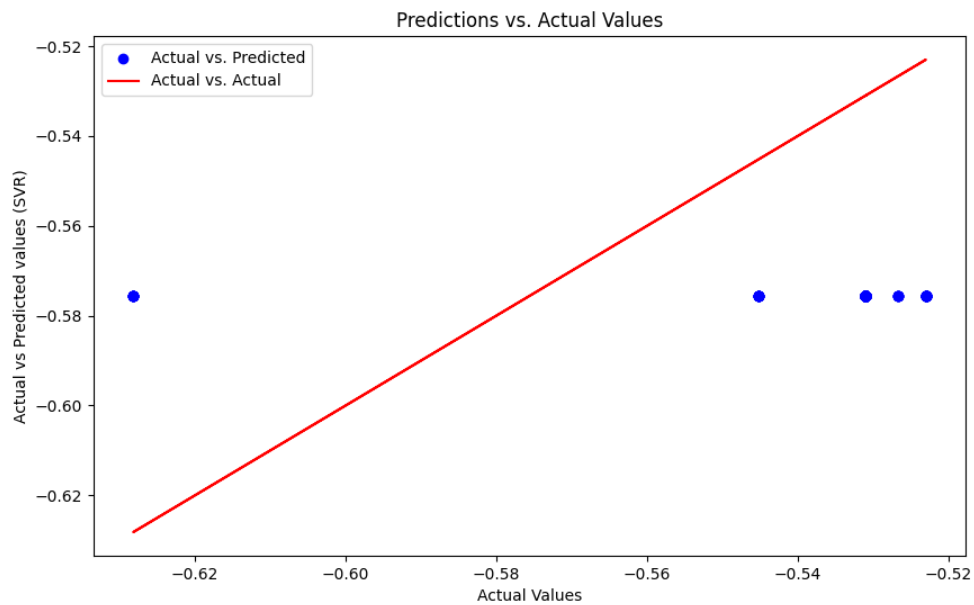
Ο απαραίτητος σχολιασμός για τις μετρικές κάθε αλγορίθμου θα γίνει σε επόμενο κεφάλαιο όπου θα υπάρξει και σύγκριση μεταξύ των αποτελεσμάτων. Στη συνέχεια παρουσιάζονται τα αντίστοιχα διαγράμματα για τον αλγόριθμο SVM όπως παρουσιάστηκαν και για τους προηγούμενους αλγορίθμους ξεκινώντας από ένα διάγραμμα που απεικονίζει το σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα.



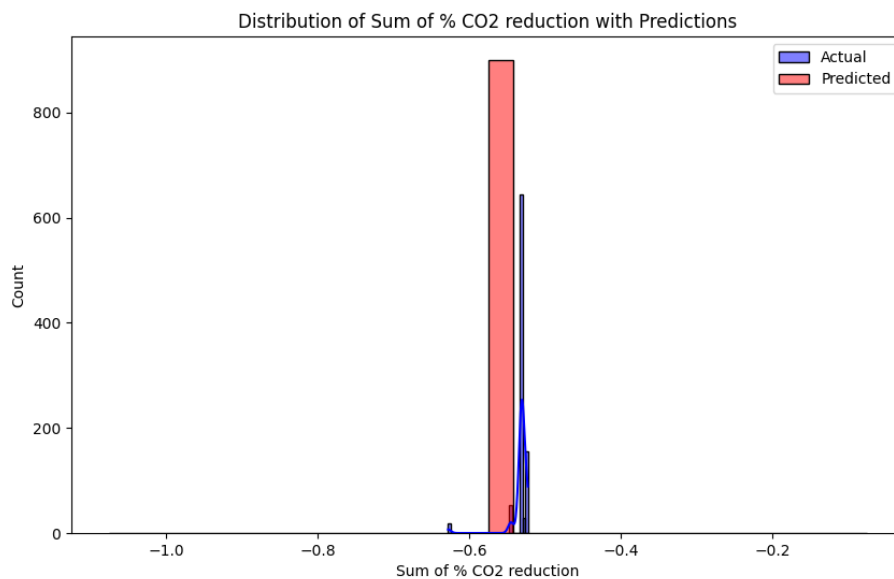
Διάγραμμα 42: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

Έπειτα παρατίθενται ένα διάγραμμα το οποίο συγκρίνει τις τιμές που προβλέφθηκαν με τις πραγματικές τιμές αλλά και ένα διάγραμμα για την κατανομή του “Sum of % CO2 reduction” συγκριτικά με τις προβλέψεις.

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 43: Προβλέψεις vs Πραγματικές τιμές



Διάγραμμα 44: Κατανομή του "Sum of % CO2 reduction" συγκριτικά με τις προβλέψεις

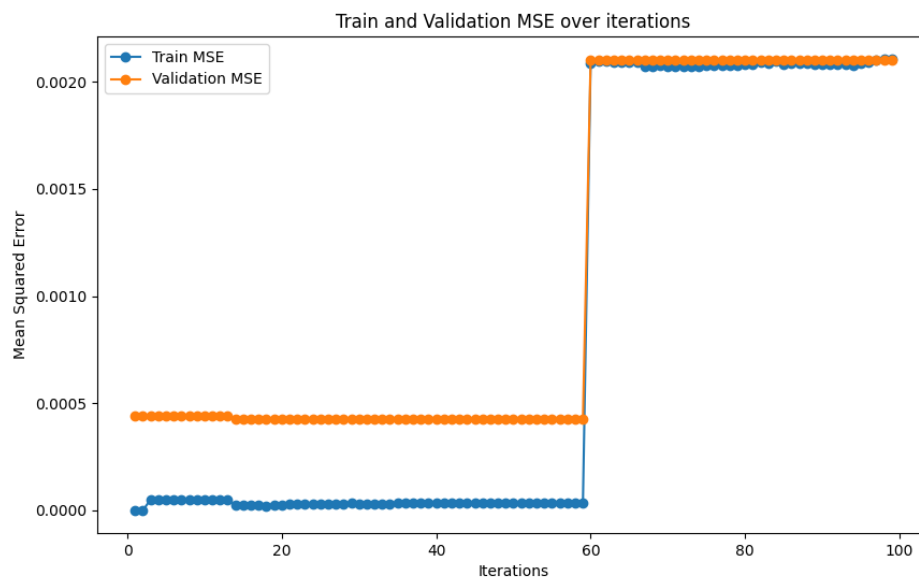
Ο αλγόριθμος SVM αξιολογήθηκε και για την προβλεπτική του ικανότητα στη στήλη "Sum of % energy reduction". Οι μετρικές παρουσιάζονται στην επόμενο πίνακα και στη συνέχεια προβάλλονται τα διαγράμματα που αφορούν το σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα, τη σύγκριση των προβλέψεων με τις

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

πραγματικές τιμές και την κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις.

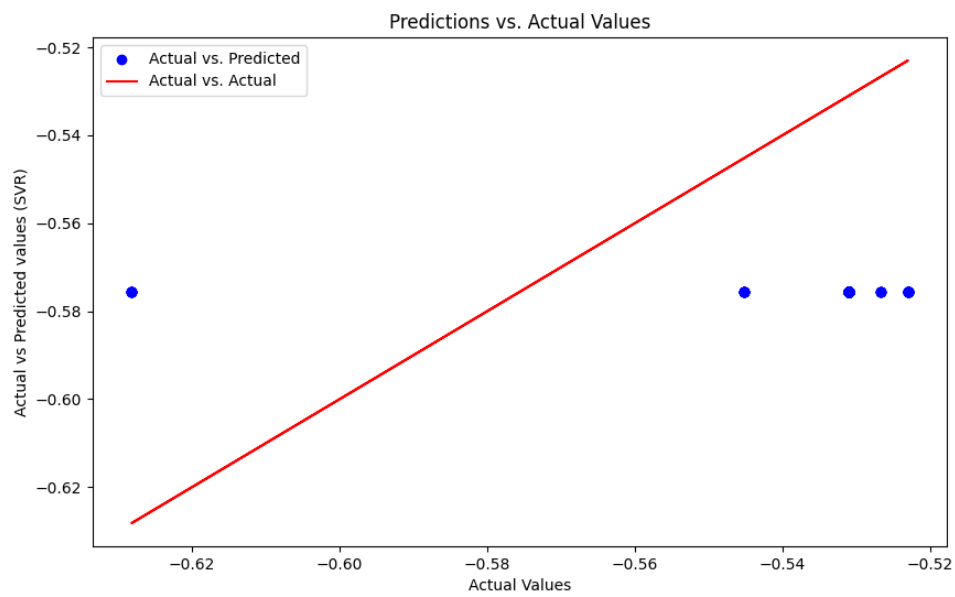
Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0021025980555555 526	-3.923946606328 7054	0.0456094444444444 41	
1.9987%	0.0000%	43.3550%	8.5394%

Και σε αυτή την περίπτωση παρατηρήθηκε το φαινόμενο ότι οι τιμές επαναλαμβάνονται για τις μετρικές απόδοσης.

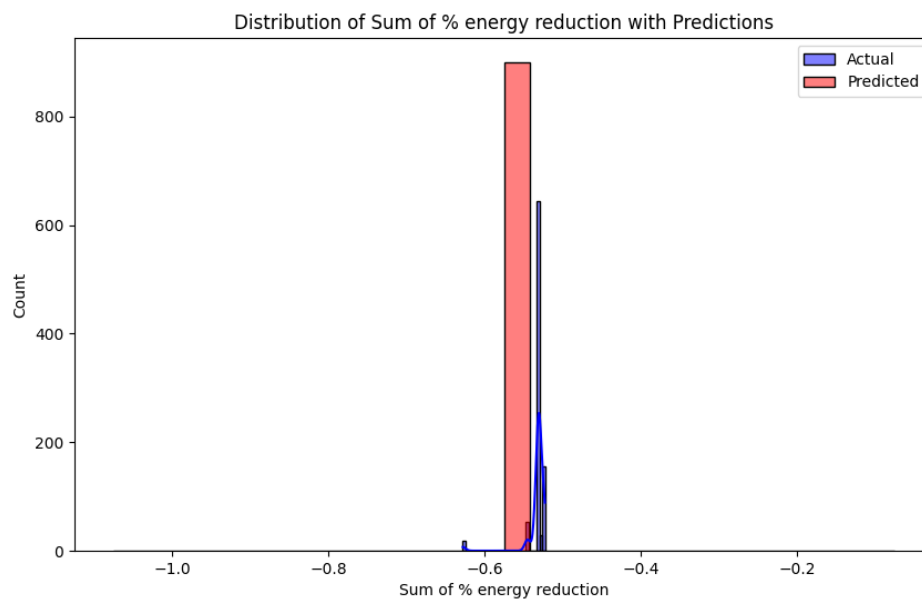


Διάγραμμα 45: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 46: Προβλέψεις vs Πραγματικές τιμές



Διάγραμμα 47: Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις

4.3.6 Artificial Neural Network (ANN)

Το τεχνητό νευρωνικό δίκτυο (Artificial Neural Network-ANN) είναι μια τεχνική μηχανικής μάθησης που έχει ως πρότυπο τα βιολογικά νευρωνικά δίκτυα. Κάθε τεχνητό νευρωνικό δίκτυο(TNN) έχει κόμβους (παρόμοιοι με τα κυτταρικά σώματα) που επικοινωνούν μεταξύ τους μέσω συνδέσεων (παρόμοιες με τους άξονες και τους δενδρίτες). Σε ένα TNN, οι συνδέσεις μεταξύ των κόμβων σταθμίζονται με βάση την ικανότητά τους να παρέχουν ένα επιθυμητό αποτέλεσμα, παρόμοια με τον τρόπο με τον οποίο ενισχύονται οι συνάψεις μεταξύ των νευρώνων σε ένα βιολογικό νευρωνικό δίκτυο (η θεωρία Hebbian υποστηρίζει ότι "νεύρα που πυροδοτούνται μαζί, καλωδιώνονται μαζί") [72].

Τα τεχνητά νευρωνικά δίκτυα αποτελούνται από:

- **Νευρώνες:** Ο νευρώνας είναι η βασική μονάδα επεξεργασίας ενός νευρωνικού δικτύου. Κάθε νευρώνας λαμβάνει εισόδους από άλλους νευρώνες ή από εξωτερικές πηγές, εκτελεί μια αριθμητική λειτουργία σε αυτές τις εισόδους και παράγει μια έξοδο.
- **Βάρη:** Τα βάρη είναι παράμετροι που ρυθμίζουν τη σημασιολογική σημασία των εισόδων σε κάθε νευρώνα. Κατά την εκπαίδευση, τα βάρη προσαρμόζονται έτσι ώστε το δίκτυο να μπορεί να προβλέπει την επιθυμητή έξοδο για δεδομένες εισόδους.
- **Συναρτήσεις ενεργοποίησης:** Οι συναρτήσεις ενεργοποίησης εισάγονται σε κάθε νευρώνα για να προσδώσουν μη γραμμικότητα στο μοντέλο. Αυτό επιτρέπει στο δίκτυο να μάθει πιο πολύπλοκες σχέσεις μεταξύ των δεδομένων εισόδου και εξόδου.
- **Δίκτυα επιπέδων:** Τα νευρωνικά δίκτυα αποτελούνται από πολλά επίπεδα νευρώνων. Κάθε επίπεδο λαμβάνει εισόδους από το προηγούμενο επίπεδο, εκτελεί μια σειρά πράξεων και παράγει έξοδο που περνά στο επόμενο επίπεδο.

Προκειμένου να δοθεί λύση στο πρόβλημα που εξετάζεται δημιουργήθηκε ένα νευρωνικό δίκτυο το οποίο εκπαιδεύτηκε χρησιμοποιώντας τα αρχικά δεδομένα προκειμένου να προβλέψει τόσο την μείωση στην ενεργειακή κατανάλωση των κτιρίων όσο και στις εκπομπές διοξειδίου του άνθρακα. Ο τρόπος δημιουργίας του νευρωνικού δικτύου αλλά και η επεξεργασία που υπέστησαν τα δεδομένα προτού χρησιμοποιηθούν παρουσιάζονται παρακάτω.

Αρχικά όπως φαίνεται στη συνέχεια επιλέχθηκαν συγκεκριμένες στήλες από τα αρχικά δεδομένα, ορίστηκε σαν μεταβλητή πρόβλεψης η στήλη “Sum of % CO2 reduction” και χωρίστηκαν τα δεδομένα σε δεδομένα εκπαίδευσης και δεδομένα ελέγχου. Τα δεδομένα εκπαίδευσης χωρίστηκαν σε δύο υποκατηγορίες, σε δεδομένα εκπαίδευσης(training data) και δεδομένα επικύρωσης(validation data) σε ποσοστό 60% και 20% αντίστοιχα.

```
selected_columns = ['Building Age', 'Total heating area, m2', 'Energy consumption before project (MWh)', 'CO2 emission before project (tons)', 'renovation of the building's enclosing structures', 'energy efficient lighting', 'heat supply renovation', 'ventilation system renovation', 'internal engineering networks', 'wood chips, biomass pellets and straw boiler house', 'solar collector system', 'technological equipment', 'energy sources with heat pumps', 'solar power plant', 'Category 1', 'Category 2', 'Category 3', 'Category 4', 'Category 5', 'Category 6', 'Category 7', 'Category 8', 'Category 9', 'Category 10', 'Category 11', 'Other']
input_cols = selected_columns # Input columns
target_col = 'Sum of % CO2 reduction' # Target column name
X = data[input_cols]
y = data[target_col]
# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# Further split the training set into training and validation sets (60% train, 20% validation)
X_train, X_val, y_train, y_val = train_test_split(X_train, y_train, test_size=0.25, random_state=42)
```

Στη συνέχεια εφαρμόστηκε η τεχνική του embedding στις στήλες που αγοράζαν την κατηγοριοποίηση λέξεων ('Category 1', 'Category 2', 'Category 3', 'Category 4', 'Category 5', 'Category 6', 'Category 7', 'Category 8', 'Category 9', 'Category 10', 'Category 11', 'Other'). Το embedding είναι μια τεχνική που χρησιμοποιείται στη μηχανική μάθηση και στα νευρωνικά δίκτυα για την αναπαράσταση δεδομένων κατηγορικών μεταβλητών σε μια χώρα χαμηλής διάστασης. Κάθε κατηγορική τιμή μετατρέπεται σε έναν διάνυσμα χαμηλής διάστασης (embedding vector), το οποίο μπορεί να μάθει το μοντέλο κατά τη διάρκεια της εκπαίδευσης. Η κύρια ιδέα είναι ότι οι παρόμοιες κατηγορικές τιμές θα έχουν παρόμοια embedding vectors. Πιο συγκεκριμένα στον κώδικα που παρατίθεται παρακάτω η λογική είναι η εξής:

- Κωδικοποίηση κατηγορικών μεταβλητών:
 - Για κάθε κατηγορική μεταβλητή στη λίστα ['Category 1', 'Category 2', ..., 'Category 11', 'Other'], ένα αντικείμενο *LabelEncoder* δημιουργείται για την κωδικοποίηση των κατηγορικών τιμών σε ακέραιους αριθμούς.
 - Οι κωδικοποιημένες τιμές αντικαθιστώνται στο DataFrame X, με τη χρήση της μεθόδου *fit_transform* του αντικειμένου *LabelEncoder*.
 - Οι *LabelEncoder* αποθηκεύονται στο λεξικό *label_encoders* με τα ονόματα των κατηγορικών μεταβλητών ως κλειδιά.
- Δημιουργία ενσωματωμένων επιπέδων:
 - Για κάθε κατηγορική μεταβλητή, δημιουργείται ένα ενσωματωμένο επίπεδο (*embedding layer*) στο οποίο οι κωδικοποιημένες κατηγορικές τιμές εισάγονται.
 - Η διάσταση του ενσωματωμένου επιπέδου υπολογίζεται με βάση τον αριθμό των μοναδικών κατηγοριών στην κάθε μεταβλητή, όπου η

επιλεγμένη μονάδα είναι ο μισός αριθμός των μοναδικών κατηγοριών (αλλά όχι περισσότερο από 50).

- Το ενσωματωμένο επίπεδο στη συνέχεια «ισοπεδώνεται» (flatten) για να μπορεί να ενσωματωθεί σε ένα πλήρες συνδεδεμένο επίπεδο στο τέλος του δικτύου.

```
# Create label encoders for categorical columns
label_encoders = {}
embedded_layers = []
for col in ['Category 1', 'Category 2', 'Category 3', 'Category 4', 'Category 5', 'Category 6', 'Category 7', 'Category 8', 'Category 9', 'Category 10', 'Category 11', 'Other']:
    le = LabelEncoder()
    X[col] = le.fit_transform(X[col])
    label_encoders[col] = le
# Create an input layer for each categorical column
input_layer = Input(shape=(1,))
# Create an embedding layer for each categorical column
embed_dim = min(np.ceil((X[col].nunique()) / 2), 50)
embed_dim = int(embed_dim)
embed_layer = Embedding(input_dim=X[col].nunique(), output_dim=embed_dim,
input_length=1)(input_layer)
embed_layer = Flatten()(embed_layer)
embedded_layers.append(embed_layer)
# Create input layer for numerical columns
numerical_input = Input(shape=(X.shape[1] - len(label_encoders),))
# Concatenate the embedded layers and numerical input
concatenated = Concatenate()(embedded_layers + [numerical_input])
```

Έπειτα ορίζεται το μοντέλο ως ένα Sequential model, το οποίο σημαίνει ότι οι στρώσεις προστίθενται με σειριακό τρόπο. Η επιλογή ενός sequential μοντέλου έγινε γιατί είναι κατάλληλο για απλά σταθερά μονοπάτια από επίπεδα, όπου η πληροφορία κινείται μόνο μπροστά, από την είσοδο στην έξοδο, και δεν υπάρχουν παράλληλα ή εναλλασσόμενα μονοπάτια. Εν συνεχεία προστίθενται έξι κρυφά επίπεδα με 32 κρυφές μονάδες για αρχή σε κάθε ένα, και η συνάρτηση ενεργοποίησης που χρησιμοποιείται είναι η ReLU (Rectified Linear Activation) και ένα τελευταίο επίπεδο εξόδου με μία μονάδα, καθώς πρόκειται για ένα πρόβλημα παλινδρόμησης. Το μοντέλο συντίθεται καλώντας τη μέθοδο compile, όπου καθορίζονται ο βελτιστοποιητής (optimizer) και η συνάρτηση κόστους (loss function) για την εκπαίδευση του μοντέλου. Σε αυτήν την περίπτωση, χρησιμοποιείται ο βελτιστοποιητής Adam και η συνάρτηση κόστους είναι η μέση τετραγωνική απόκλιση (mean squared error), που είναι ιδανικό για ένα πρόβλημα παλινδρόμησης. Τέλος το μοντέλο εκπαιδεύεται με τα δεδομένα εκπαίδευσης (X_train, y_train) για έναν αριθμό εποχών που ορίζονται από τον παράμετρο epochs και πραγματοποιούνται και οι προβλέψεις. Χρησιμοποιείται η μέθοδος επαλήθευσης (validation) με τα δεδομένα επαλήθευσης (X_val, y_val) για τον έλεγχο της απόδοσης του μοντέλου κατά την εκπαίδευση. Οποιαδήποτε υπερβολική εκπαίδευση αποφεύγεται με τη χρήση του

callback EarlyStopping, το οποίο διακόπτει την εκπαίδευση αν η απώλεια στα δεδομένα επαλήθευσης δεν μειώνεται για έναν ορισμένο αριθμό εποχών (που καθορίζεται από την παράμετρο patience)

```
model = Sequential()
# Add 6 hidden layers with 32 hidden units in each neuron
model.add(Dense(32, activation='relu', input_shape=(X.shape[1],)))
model.add(Dense(32, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(32, activation='relu'))
# Add the output layer
model.add(Dense(1)) # Regression task
# Compile the model
model.compile(optimizer='adam', loss='mean_squared_error', metrics = ['mae']) #
Change loss for classification tasks
# Early stopping callback
early_stopping = EarlyStopping(monitor='val_loss', patience=5,
restore_best_weights=True)
# Train the model and collect MSE values for each epoch
epochs = 50
train_loss = []
val_loss = []
# Train the model using validation data for monitoring
history = model.fit(X_train, y_train, epochs=50, batch_size=32, verbose=1,
validation_data=(X_val, y_val), callbacks=[early_stopping])
# Make predictions
predictions = model.predict(X_test)
```

Οι μετρικές απόδοσης του νευρωνικού δικτύου παρατίθενται στον παρακάτω πίνακα.

Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.1957477700393562	-73.836718451443 32	0.268742394975988 1	
43.18%	0.0000%	59.29%	48.6606%

Όπως είναι εμφανές τα νούμερα είναι πολύ υψηλά για να θεωρηθεί ο αλγόριθμος αποδοτικός. Επιλέχθηκε λοιπόν να προστεθεί ένα ακόμα hidden layer στο νευρωνικό δίκτυο και να αυξηθούν και οι κρυφές μονάδες κάθε layer από 32 σε 128.

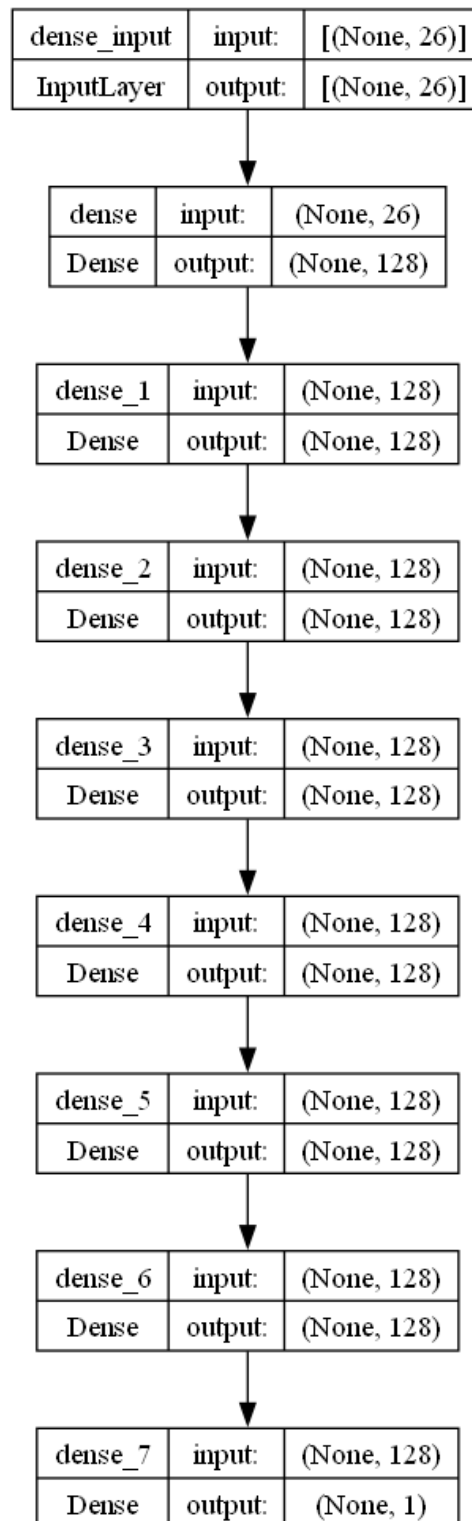
```
# Add 7 hidden layers with 128 hidden units in each neuron
model.add(Dense(128, activation='relu', input_shape=(X.shape[1],)))
model.add(Dense(128, activation='relu'))
model.add(Dense(128, activation='relu'))
model.add(Dense(128, activation='relu'))
```

```
model.add(Dense(128, activation='relu'))  
model.add(Dense(128, activation='relu'))  
model.add(Dense(128, activation='relu'))
```

Οι μετρικές απόδοσης του νέου νευρωνικού δικτύου παρουσιάζονται στη συνέχεια.

Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0760964973053221	-28.092602908500	0.219872344131713	
1	222	5	
16.79%	0.0000%	48.50%	39.7686%

Παρατηρείται ότι το μέσο τετραγωνιαίο σφάλμα μειώθηκε αρκετά, ωστόσο δεν σημειώθηκε τόσο σημαντική μείωση και στο μέσο απόλυτο σφάλμα. Εν συνεχεία δοκιμάστηκαν και άλλες προσεγγίσεις με την προσθήκη layers οι οποίες μείωναν ελάχιστα το μέσο απόλυτο σφάλμα και συνεπώς δεν αποτελούν αποτελεσματικές μεθόδους αντιμετώπισης του προβλήματος. Για αυτό επιλέχθηκε η τελική μορφή του δικτύου να είναι αυτή που προαναφέρθηκε και διακρίνεται στο επόμενο σχήμα.

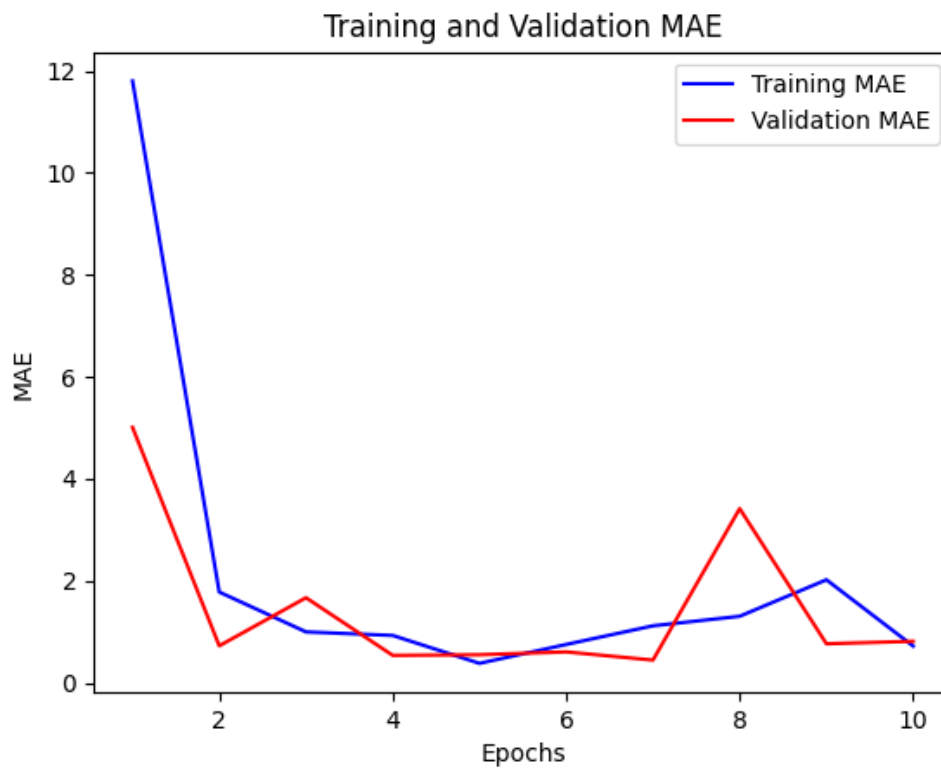


Εικόνα 13: Αρχιτεκτονική νευρωνικού δικτύου

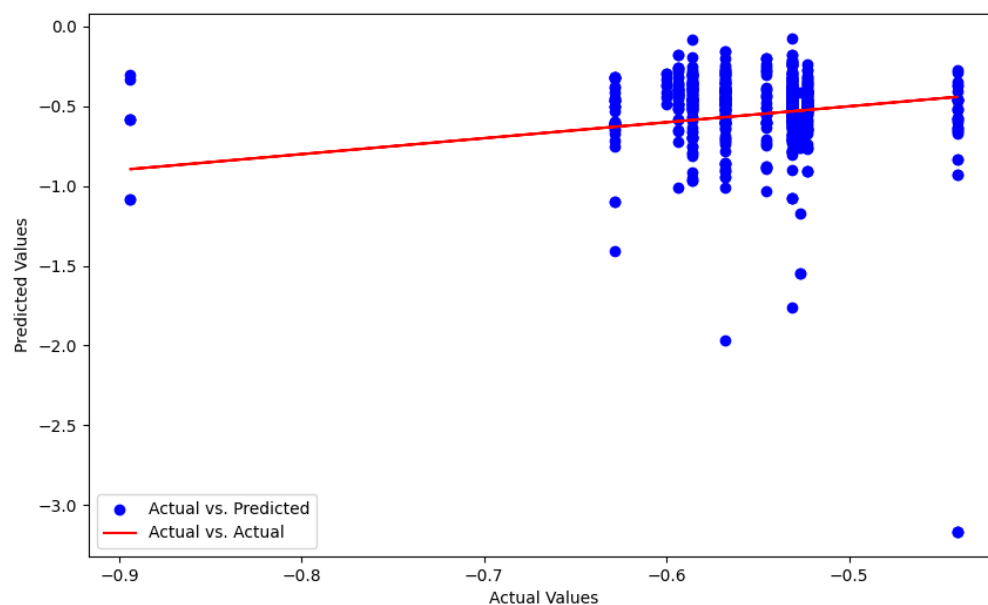
Η αξιολόγηση του νευρωνικού δικτύου έγινε με παρόμοιο τρόπο με τους αλγόριθμους μηχανικής μάθησης που παρατέθηκαν νωρίτερα σε αυτό το κεφάλαιο. Για αυτό παρουσιάζονται τα διαγράμματα που αφορούν το σφάλμα εκπαίδευσης και επικύρωσης, ένα συγκριτικό διάγραμμα των πραγματικών τιμών με τις προβλέψεις

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

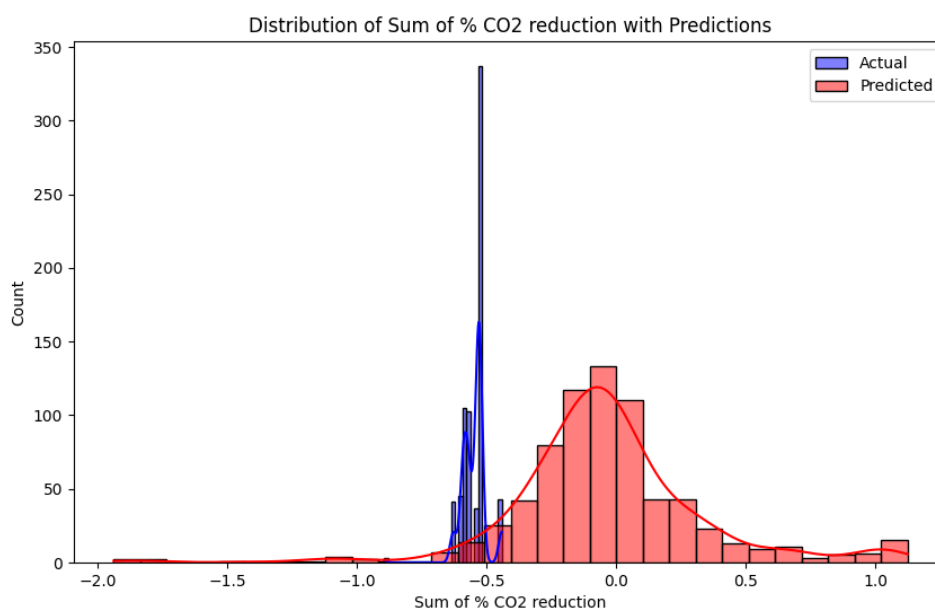
και ένα διάγραμμα για την κατανομή του “Sum of % energy reduction” συγκριτικά με τις προβλέψεις του νευρωνικού δικτύου.



Διάγραμμα 48: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα



Διάγραμμα 49: Προβλέψεις vs Πραγματικές τιμές



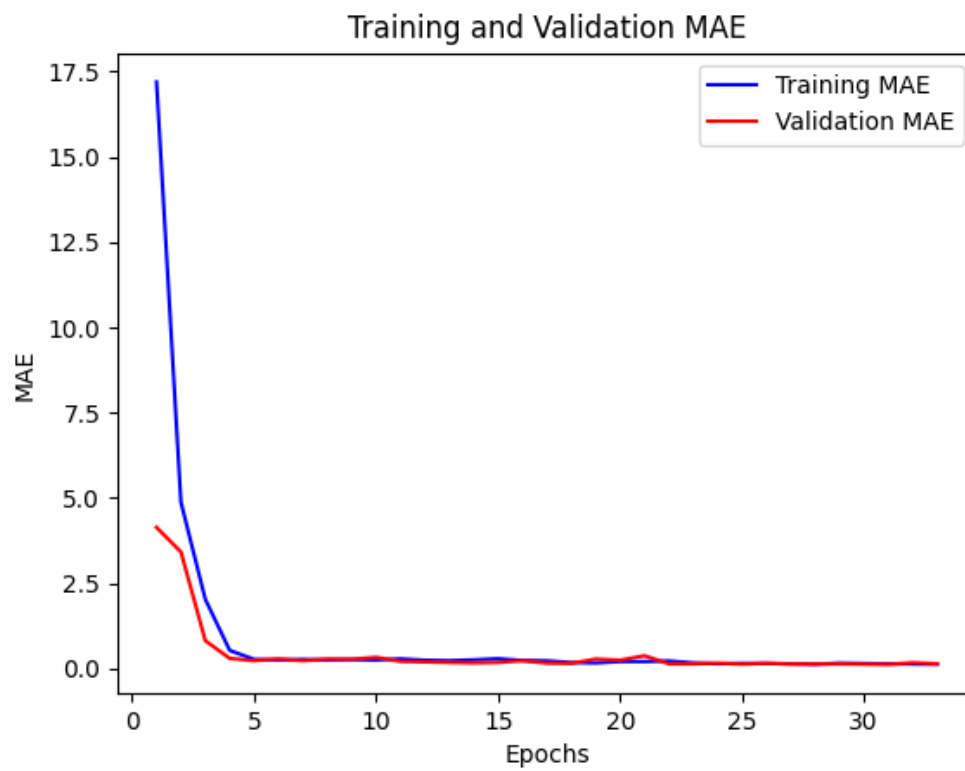
Διάγραμμα 50: Κατανομή του "Sum of % CO2 reduction" συγκριτικά με τις προβλέψεις

Τέλος οι προβλέψεις πραγματοποιήθηκαν και τη στήλη "Sum of % energy reduction", για την οποία παρατίθενται τόσο οι μετρικές απόδοσης του νευρωνικού δικτύου όσο και τα αντίστοιχα διαγράμματα παρακάτω.

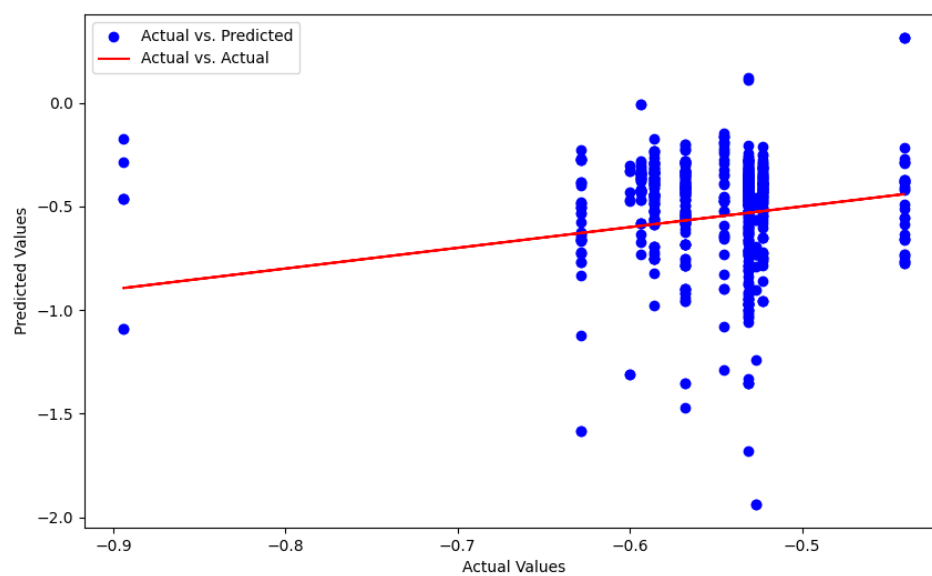
Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
0.0313030936901690	-10.967547873875	0.139869299720448	
5	324	98	
6.91%	0.0000%	30.86%	25.0582%

Στις προβλέψεις της στήλης "Sum of % energy reduction" παρατηρούνται καλύτερες τιμές των μετρικών συγκριτικά με τη στήλη "Sum of % CO2 reduction". Στη συνέχεια παρουσιάζονται τα διαγράμματα για τα σφάλματα του νευρωνικού δικτύου, η σύγκριση των προβλέψεων με τις πραγματικές τιμές αλλά και η κατανομή των προβλέψεων της στήλης "Sum of % energy reduction" συγκριτικά με τις πραγματικές τιμές της στήλης.

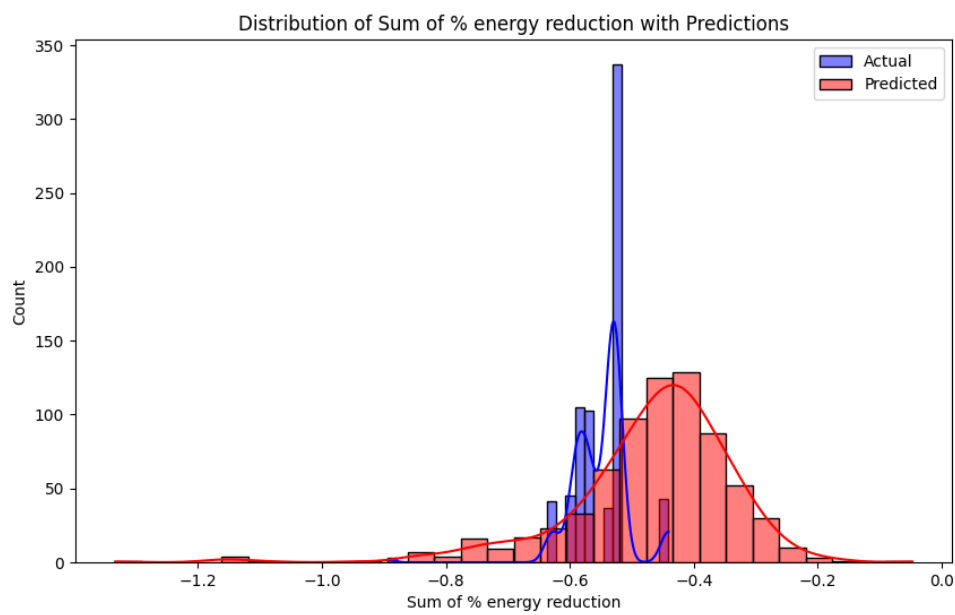
Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων



Διάγραμμα 51: Σφάλμα εκπαίδευσης και επικύρωσης για τα δεδομένα



Διάγραμμα 52: Προβλέψεις vs Πραγματικές τιμές



Διάγραμμα 53: Κατανομή του "Sum of % energy reduction" συγκριτικά με τις προβλέψεις

ΚΕΦΑΛΑΙΟ 5

Σχολιασμός των Αποτελεσμάτων

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

5 Σχολιασμός των Αποτελεσμάτων

Η σύγκριση των μετρικών απόδοσης μεταξύ διαφορετικών αλγορίθμων μηχανικής μάθησης είναι κρίσιμη για την αξιολόγηση της απόδοσής τους σε συγκεκριμένα προβλήματα. Αυτή η σύγκριση μπορεί να παρέχει πολύτιμες ενδείξεις σχετικά με την αποτελεσματικότητα και την καταλληλότητα του κάθε αλγορίθμου για το συγκεκριμένο πρόβλημα. Στο πλαίσιο αυτού του κεφαλαίου, θα αναλυθούν και θα συγκριθούν οι μετρικές απόδοσης όπως το Mean Absolute Error (MAE), το Mean Squared Error (MSE), το Mean Absolute Percentage Error (MAPE) και το R-squared (R^2) μεταξύ των αλγορίθμων μηχανικής μάθησης και του νευρωνικού δικτύου που εφαρμόστηκαν. Αυτή η σύγκριση θα παράσχει μια πλήρη εικόνα της απόδοσης των αλγορίθμων και θα διευκολύνει την επιλογή του κατάλληλου αλγορίθμου για το συγκεκριμένο πρόβλημα ενεργειακής ανακαίνισης κτιρίων. Επιπλέον, η ανάλυση αυτή θα προσφέρει επίσης ενδείξεις σχετικά με την αποτελεσματικότητα των μεθόδων μηχανικής μάθησης στην πρόβλεψη της μείωσης της ενεργειακής κατανάλωσης και των εκπομπών αερίων του θερμοκηπίου σε ανακαινισμένα κτίρια. Στη συνέχεια παρατίθεται συγκεντρωτικά τα αποτελέσματα των μετρικών απόδοσης για κάθε τεχνική και για κάθε στήλη που πραγματοποιήθηκαν οι προβλέψεις.

Για να εξαχθούν ασφαλή συμπεράσματα παρατίθενται οι ιδανικές τιμές για κάθε μετρική απόδοσης. Οι ιδανικές τιμές για τις μετρικές απόδοσης διαφέρουν ανάλογα με το πρόβλημα και τις συνθήκες της κάθε εφαρμογής. Ωστόσο, γενικά ιδανικές τιμές για κάθε μετρική είναι οι εξής:

- **Mean Squared Error (MSE):** Η ιδανική τιμή του MSE είναι 0. Αυτό σημαίνει ότι οι προβλεπόμενες τιμές είναι ακριβώς ίσες με τις πραγματικές τιμές.
- **R-Squared (R^2):** Η ιδανική τιμή του R^2 είναι 1. Αυτό σημαίνει ότι οι προβλεπόμενες τιμές εξηγούν πλήρως τη διακύμανση των πραγματικών τιμών.
- **Mean Absolute Error (MAE):** Η ιδανική τιμή του MAE είναι 0. Όσο κοντά στο 0 είναι η τιμή του MAE, τόσο καλύτερη είναι η ακρίβεια του μοντέλου.
- **Mean Absolute Percentage Error (MAPE):** Δεν υπάρχει απόλυτη ιδανική τιμή για το MAPE, καθώς εξαρτάται από τη φύση των δεδομένων. Ωστόσο, για πρακτικούς σκοπούς, μια χαμηλή τιμή του MAPE (κοντά στο 0) υποδεικνύει ότι το μοντέλο παράγει προβλέψεις που είναι κοντά στις πραγματικές τιμές. Είναι σημαντικό να ληφθεί υπόψη ότι το MAPE μπορεί να παρουσιάζει προβλήματα όταν οι πραγματικές τιμές είναι πολύ κοντά στο μηδέν.

Πίνακας 6: Συγκεντρωτικά αποτελέσματα των αλγορίθμων

Algorithm	Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
KNN "Sum of % CO2 reduction"	0.001384489 0756302519	0.3847832383 948112	0.017533520074 69654	
	0.3054%	38.4783%	3.8680%	3.0407%
KNN "Sum of % energy reduction"	0.001055381 0457516336	0.5798622595 140075	0.013434360410 830994	
	0.2328%	57.9862%	2.9637%	2.2674%
Random Forest "Sum of % CO2 reduction"	0.000264573 8523204674 5	0.8952922145 87036	0.002794452225 3361337	
	0.0584%	89.5292%	0.6165%	0.4060%
Random Forest "Sum of % energy reduction"	0.000298207 2739495782	0.8819814468 60908	0.002858106442 577554	
	0.0658%	88.1981%	0.6305%	0.4176%
Lightgbm "Sum of % CO2 reduction"	0.0045	0.9923	0.0010	
	0.990146347 2167101%	99.229827283 15286%	0.227568049654 49662%	0.1605%
Lightgbm "Sum of % energy reduction"	0.0045	0.9923	0.0010	
	0.990146347 2167101%	99.229827283 15286%	0.227568049654 49662%	0.1605%
Lightgbm(v2) "Sum of % CO2 reduction"	0.0041	0.9935	0.0008	
	0.9086%	99.3514%	0.1736%	0.1246%
Lightgbm(v2) "Sum of % energy reduction"	0.0041	0.9935	0.0008	
	0.9086%	99.3514%	0.1736%	0.1246%

SVM "Sum of % CO2 reduction"	0.002102598 0555555526	-3.9239466 063287054	0.045609444444 44441	
	1.9987%	0.0000%	43.3550%	8.5394%
SVM "Sum of % energy reduction"	0.002102598 0555555526	-3.9239466 063287054	0.045609444444 44441	
	1.9987%	0.0000%	43.3550%	8.5394%
Neural Network "Sum of % CO2 reduction"	0.076096497 30532211	-28.0926029 08500222	0.219872344131 7135	
	16.79%	0.0000%	48.50%	39.7686%
Neural Network "Sum of % energy reduction"	0.031303093 69016905	-10.967547 873875324	0.139869299720 44898	
	6.91%	0.0000%	30.86%	25.0582%

Ξεκινώντας από το μέσο τετραγωνικό σφάλμα παρατηρείται ότι το μικρότερο μέσο τετραγωνικό σφάλμα παρουσιάζει ο αλγόριθμος Random Forest με τιμή ίση με 0.00026457385232046745. Το R-Squared (R^2), η διακύμανση των πραγματικών τιμών πρέπει να είναι κοντά στη μονάδα. Σε γενικές γραμμές, ένα αρνητικό R^2 υποδεικνύει ότι το μοντέλο τείνει να είναι αναξιόπιστο και δεν παρέχει κάποια χρήσιμη πληροφορία για την εξήγηση της διακύμανσης της εξαρτημένης μεταβλητής. Κοντινότερη τιμή στη μονάδα έχει ο αλγόριθμος Lightgbm στον οποία χρησιμοποιήθηκαν ως δεδομένα εισόδου τα χαρακτηριστικά με τη μεγαλύτερη συσχέτιση μεταξύ τους με τιμή 0.9935. Όσον αφορά το μέσο απόλυτο σφάλμα τη μικρότερη τιμή φαίνεται ότι έχει ο αλγόριθμος Lightgbm, η δεύτερη έκδοση του αρχικού, με τιμή ίση με 0.0008. Τέλος για τον ποσοστιαίο μέσο τετραγωνικό σφάλμα παρουσιάζει πάλι ο αλγόριθμος Lightgbm(v2) με τιμή ίση με 0.1246%. Οι καλύτεροι(min) αλλά και οι χειρότεροι(max) αλγόριθμοι φαίνονται στον παρακάτω πίνακα.

Πίνακας 7: Καλύτεροι και χειρότεροι αλγόριθμοι

	Mean Squared Error	R-Squared	Mean Absolute Error	Mean Absolute Percentage Error
Min	Random Forest	Lightgbm(v2)	Lightgbm(v2)	Lightgbm(v2)
Max	Neural Network	Neural Network	Neural Network	Neural Network

Όπως αναφέρθηκε προηγουμένως η λειτουργία του νευρωνικού δικτύου μπορεί να βελτιωθεί προσθέτοντας και άλλα layers αλλά δεν είναι η ιδανικότερη επιλογή. Αυτό συμβαίνει γιατί η προσθήκη πολλών layers μπορεί να οδηγήσει σε υπερβολική προσαρμογή στα δεδομένα εκπαίδευσης, με αποτέλεσμα το μοντέλο να μην γενικεύει επαρκώς σε νέα δεδομένα ενώ συγχρόνως μπορεί να αυξήσει την απόδοση του μοντέλου στα δεδομένα εκπαίδευσης, αλλά μπορεί επίσης να αυξήσει τον υπολογιστικό φόρτο που απαιτείται για την εκπαίδευση και την πρόβλεψη.

Συνεπώς ο αλγόριθμος Lightgbm όπου χρησιμοποιήθηκαν τα features με την μεγαλύτερη συσχέτιση μεταξύ τους αναδεικνύεται ως η καλύτερη επιλογή. Για την εκτέλεση του παρατίθενται και μια εικόνα του αρχείου Excel που παριστάνονται οι πραγματικές τιμές και οι προβλέψεις που έκανε ο αλγόριθμος. Είναι εμφανής η πολύ μικρή απόκλιση που έχουν οι δύο στήλες μεταξύ τους, γεγονός που επιβεβαιώνει το πολύ μικρό μέσο απόλυτο σφάλμα που παρουσιάζει αυτή η εκδοχή του αλγορίθμου Lightgbm.

Actual	Predicted
-0,523	-0,523094846
-0,5858	-0,585657183
-0,531	-0,53100396
-0,6282	-0,621829133
-0,523	-0,523051257
-0,4407	-0,440901235
-0,5452	-0,545193547
-0,531	-0,530992312
-0,523	-0,523100706
-0,6004	-0,594299518
-0,6282	-0,625421384
-0,531	-0,531074027
-0,5682	-0,568686018
-0,6282	-0,629549202
-0,523	-0,523205066
-0,6282	-0,628739235
-0,5682	-0,568185336
-0,5682	-0,568192361
-0,6282	-0,624988605
-0,5682	-0,568185336
-0,5682	-0,568451646
-0,5452	-0,545175073
-0,5939	-0,593752076
-0,531	-0,530973967
-0,5682	-0,568253729
-0,5682	-0,568509701
-0,523	-0,523058022
-0,5858	-0,585555628
-0,531	-0,530975217
-0,523	-0,523187976

Εικόνα 14: Προβλέψεις vs Πραγματικές τιμές για τον αλγόριθμο Lightgbm(v2)

Τέλος παρουσιάζεται η κατάταξη των αλγορίθμων μηχανικής μάθησης που χρησιμοποιήθηκαν με βάσει το μέσο απόλυτο σφάλμα, την απόκλιση της πρόβλεψης που πραγματοποιήσαν δηλαδή από τις πραγματικές τιμές.

Πίνακας 8: Πίνακας κατάταξης των αλγορίθμων

Rank	Algorithm	Feature	Mean Absolute Error
1	Lightgbm(v2)	"Sum of % energy reduction" & "Sum of % CO2 reduction"	0.0008
2	Lightgbm	"Sum of % energy reduction" & "Sum	0.0010

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

		of % CO2 reduction"	
3	Random Forest	"Sum of % CO2 reduction"	0.0027944522253361337
	KNN	"Sum of % energy reduction"	0.013434360410830994
4	SVM	"Sum of % energy reduction" & "Sum of % CO2 reduction"	0.04560944444444441
5	Neural Network	"Sum of % energy reduction"	0.13986929972044898

ΚΕΦΑΛΑΙΟ 6

Συμπεράσματα και Προτάσεις για Περαιτέρω Έρευνα

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

6 Συμπεράσματα και Προτάσεις για Περαιτέρω Έρευνα

6.1 Συμπεράσματα για τους αλγόριθμους

Στην παρούσα εργασία πραγματοποιείται διεξοδική ανάλυση και εφαρμογή της μηχανικής μάθησης στον τομέα της ενέργειας. Αξίζει να σημειωθεί ότι όπως αναφέρθηκε και στη βιβλιογραφική ανασκόπηση στο κεφάλαιο 3, ο τομέας της ενέργειας κυριαρχεί στις περισσότερες έρευνες που χρησιμοποιείται η μηχανική μάθηση και η τεχνητή νοημοσύνη καθώς ο στόχος για ένα βιώσιμο και λιγότερο επιβλαβές μέλλον αποτελεί προτεραιότητα όλων.

Για την υλοποίηση της παρούσας διπλωματικής εργασίας επιλέχθηκαν αλγόριθμοι μηχανικής μάθησης που λειτουργούν καλύτερα σε προβλήματα παλινδρόμησης (regression). Συγκεκριμένα χρησιμοποιήθηκαν οι αλγόριθμοι K-Nearest Neighbors(KNN), Random Forest, Lightgbm, Support Vector Machine(SVM) ενώ παράλληλα αναπτύχθηκε και ένα νευρωνικό δίκτυο από την αρχή με στόχο την πρόβλεψη της μείωσης της ενεργειακής κατανάλωσης και των εκπομπών διοξειδίου του άνθρακα σε κτίρια που πραγματοποίησαν ενέργειες ανακαίνισης με στόχο να γίνουν πιο ενεργειακά αποδοτικά. Οι αλγόριθμοι αυτοί όπως καταγράφονται στο προηγούμενο κεφάλαιο κατατάσσονται από τον καλύτερο προς το χειρότερο με την παρακάτω σειρά.

1. Lightgbm (v.2)
2. Lightgbm
3. Random Forest
4. K-Nearest Neighbors (KNN)
5. Support Vector Machine (SVM)
6. Νευρωνικό Δίκτυο

Ξεκινώντας από τον χειρότερο αποτελεσματικά αλγόριθμο, το νευρωνικό δίκτυο το οποίο δημιουργήθηκε δεν μπόρεσε να προβλέψει επιτυχώς ούτε την μείωση της ενεργειακής κατανάλωσης αλλά ούτε και την μείωση των εκπομπών του διοξειδίου του άνθρακα επιφέροντας αποκλίσεις της τάξεως του 25% και 39% αντίστοιχα. Οι αποκλίσεις αυτές είναι πολύ μεγάλες αν αναλογιστεί κανείς ότι η τιμή πρόβλεψης έχει να κάνει με την ποσοστιαία μείωση στην ενεργειακή κατανάλωση. Συγχρόνως ο συντελεστής προσδιορισμού (R-squared) ήταν αρνητικός, γεγονός που υποδηλώνει ότι το επιλεγμένο μοντέλο έχει χειρότερη απόδοση από ένα μοντέλο που προβλέπει απλώς τον μέσο όρο της εξαρτημένης μεταβλητής. Με άλλα λόγια, οι ανεξάρτητες μεταβλητές στο μοντέλο εξηγούν λιγότερο από τη διακύμανση της εξαρτημένης μεταβλητής από ό,τι αν χρησιμοποιούσαμε απλώς τον μέσο όρο. Ωστόσο με την προσθήκη περισσότερων επιπέδων στο νευρωνικό δίκτυο οι προβλέψεις εμφάνιζαν βελτίωση, ωστόσο δεν είναι ο ιδανικός τρόπος αντιμετώπισης του προβλήματος καθώς μπορεί να οδηγήσει σε υπερβολική προσαρμογή στα δεδομένα εκπαίδευσης και σε αυξημένο υπολογιστικό φόρτο. Ωστόσο οι δυσκολίες που παρουσιάστηκαν κατά την υλοποίηση με παρόμοιο νευρωνικό δίκτυο εμφανίζονται και σε άλλες

έρευνες, υποδεικνύοντας την ύπαρξη μιας κοινής πρόκλησης στον τομέα. Αυτό ενισχύει την ανάγκη για περαιτέρω μελέτη και καινοτομία σε επίπεδο μεθοδολογίας και αλγορίθμων.

Ο αλγόριθμος SVM παρά το γεγονός ότι εμφάνισε αρνητική διακύμανση (R-squared), είχε σαν αποτέλεσμα αποκλίσεις της τάξεως του 8%, εμφανίζοντας μεγαλύτερη βελτίωση και καλύτερη ικανότητα πρόβλεψης από το νευρωνικό δίκτυο. Παρόλα αυτά, ο αλγόριθμος SVM καθυστερούσε πολύ να επιστρέψει αποτελέσματα όταν το μέγεθος των δεδομένων εισόδου ξεπερνούσε τις 1000 εγγραφές. Αυτό οφείλεται συνήθως στο γεγονός ότι ο αλγόριθμος SVM αναζητά το βέλτιστο υπερεπίπεδο που διαχωρίζει τις κλάσεις στο χώρο των χαρακτηριστικών. Αυτό καθιστά τον αλγόριθμο μη λειτουργικό καθώς σε πραγματική ροή δεδομένων θα επιφέρει σοβαρά χρονικά προβλήματα.

Ο επόμενος αλγόριθμος στη σειρά κατάταξης είναι ο αλγόριθμος KNN ο οποίος εμφάνισε αποκλίσεις της τάξης του 2,2% και 3%. Παρότι οι τιμές αυτές που αφορούν την ποσοστιαία έκφραση του μέσο τετραγωνικού σφάλματος είναι πάρα πολύ καλές η τιμή της διακύμανσης (R-squared) της εξαρτημένης μεταβλητής είναι 57%. Η ιδανική τιμή για την διακύμανση είναι 1 ή αντίστοιχα 100%. Όπως είναι προφανές το 57% απέχει αρκετά από την ιδανική τιμή, γεγονός που υποδηλώνει ότι το μοντέλο του KNN δεν εξηγεί καλά την εξαρτημένη μεταβλητή που εξετάζεται και χρήζει βελτίωσης.

Οι αλγόριθμοι Random Forest και Lightgbm που ακολουθούν στη σειρά κατάταξης λειτούργησαν εξαιρετικά όσο να αφορά την ικανότητα πρόβλεψης τόσο για την μείωση της ενεργειακής κατανάλωσης αλλά και των εκπομπών του διοξειδίου του άνθρακα. Αναλυτικότερα οι αποκλίσεις ήταν 0,40% για τον αλγόριθμο Random Forest και 0,16% αντίστοιχα για τον Lightgbm, ενώ η διακύμανση της εξαρτημένης μεταβλητής είχε τιμές 90% και 99% αντίστοιχα. Όμως ο αλγόριθμος Random Forest ήταν αρκετά ευαίσθητος στην υπερπροσαρμογή (overfitting), έχοντας ως αποτέλεσμα να παρουσιάζει τιμές ακόμα και 100% στο R-squared και 0 στις τιμές που αφορούσαν τα σφάλματα, κάτι το οποίο υποδεικνύει ότι ο αλγόριθμος δεν λειτουργεί σωστά. Για αυτό ακριβώς τον λόγο αλλά και γιατί ο αλγόριθμος Lightgbm εμφάνισε καλύτερες μετρικές απόδοσης επιλέχθηκε αυτός ως η καλύτερη μέθοδος για την πρόβλεψη της ενεργειακής κατανάλωσης και των εκπομπών του διοξειδίου του άνθρακα. Επιλέχθηκε ωστόσο να εξεταστεί και μία τελευταία περίπτωση του αλγορίθμου Lightgbm χρησιμοποιώντας ως δεδομένου εισόδου τα χαρακτηριστικά με τη μεγαλύτερη συσχέτιση μεταξύ τους. Η εκδοχή αυτή του αλγορίθμου Lightgbm, είχε σαν αποτέλεσμα οι αποκλίσεις των πραγματικών τιμών με τις προβλέψεις να είναι της τάξεως του 0,1246%, όπως φαίνεται και από την εικόνα στο προηγούμενο κεφάλαιο, ενώ η διακύμανση είχε την μέγιστη τιμή της, 99,35%. Επομένως επιλέχθηκε αυτός ο αλγόριθμος ως ο καλύτερος για την εφαρμογή του για την πρόβλεψη της μείωσης στην ενεργειακή κατανάλωση και τις εκπομπές του διοξειδίου του άνθρακα σε κτίρια τα οποία προχώρησαν σε ενέργειες με στόχο να γίνουν πιο ενεργειακά αποδοτικά. Παράλληλα αποδεικνύεται ότι σημασία δεν έχει μόνο η επιλογή του κατάλληλου αλγορίθμου αλλά και η χρήση των χαρακτηριστικών με τη μεγαλύτερη

συσχέτιση μεταξύ τους, καθώς αυτό το μοντέλο απέδωσε με τα καλύτερα αποτελέσματα.

6.2 Συμπεράσματα για την έρευνα

Η έρευνα που πραγματοποιήθηκε στο πλαίσιο της παρούσας διπλωματικής εργασίας προσφέρει μια πολύ καινοτόμα προσέγγιση στον τομέα της εφαρμογής της μηχανικής μάθησης στην ενέργεια. Η εφαρμογή της τεχνολογίας αυτής ανοίγει νέες προοπτικές και δυνατότητες.

Αρχικά, οι διαχειριστές και ιδιοκτήτες κτιρίων αντιμετωπίζουν διάφορες προκλήσεις στη διαχείριση της ενέργειας και των εκπομπών, όπως η απόδοση των συστημάτων θέρμανσης και ψύξης, η βελτιστοποίηση της χρήσης φωτισμού και η διαχείριση της κατανάλωσης ενέργειας των ηλεκτρικών συσκευών. Πολλές φορές επιλέγουν να προβούν σε ενεργειακές ανακαινίσεις με στόχο να λύσουν κάποιο από τα παραπάνω προβλήματα. Το αποδοτικότερο μοντέλο στο οποίο κατέληξε η έρευνα αυτή, δηλαδή ο αλγόριθμος Lightgbm με τη χρήση των χαρακτηριστικών με τη μεγαλύτερη συσχέτιση μεταξύ τους, μπορεί να αποβεί ένας αποτελεσματικός και αποδοτικός τρόπος καθώς προβλέπει την μείωση τόσο στην ενεργειακή κατανάλωση ενός κτιρίου όσο και στις εκπομπές διοξειδίου του άνθρακα με βάση τεχνικά χαρακτηριστικά του κτιρίου και κάποια δεδομένα όπως η κατανάλωση ενέργειας του κτιρίου. Συγχρόνως, μπορούν να εξετάσουν τις επιπτώσεις που θα έχουν διαφορετικές παρεμβάσεις στην ενεργειακή κατανάλωση, επιλέγοντας στο τέλος τις πιο χρήσιμες και ενεργειακά αποδοτικές. Αυτό μπορεί να επιτρέψει στους διαχειριστές να λαμβάνουν προληπτικά μέτρα για τη μείωση της κατανάλωσης ενέργειας και τον περιορισμό του κόστους λειτουργίας του κτιρίου.

Επίσης, η χρήση του μοντέλου που εξετάστηκε μπορεί να παρέχει χρήσιμες πληροφορίες και εργαλεία σε ειδικούς που ασχολούνται με τις ανακαινίσεις κτιρίων (εταιρίες συμβουλευτικής σε θέματα ενέργειας, παρόχους ενεργειακών υπηρεσιών). Πιο συγκεκριμένα, ανάλογα με το αντικείμενο ενασχόλησης τους μπορούν να αξιολογήσουν την απόδοση διαφόρων τεχνικών ανακαίνισης, όπως η εγκατάσταση μονωτικών υλικών, η αντικατάσταση των κλιματιστικών συστημάτων με πιο αποδοτικά ή η χρήση οικολογικών υλικών κατασκευής. Έτσι μπορούν αρχικά να συγκρίνουν τα αποτελέσματα με τις μεθόδους που χρησιμοποιούσαν ήδη για την πρόβλεψη της επίδρασης των παρεμβάσεων τους αλλά και να προβλέψουν ποιες είναι οι πιο αποτελεσματικές για συγκεκριμένα κτίρια και συνθήκες. Η προσέγγιση αυτή μειώνει τον κίνδυνο ανεπιθύμητων αποτελεσμάτων αυξάνοντας τις πιθανότητες επιτυχούς ανακαίνισης. Επιπλέον, συμβάλλει στην επίτευξη περιβαλλοντικών στόχων και την προαγωγή της αειφορίας στον τομέα των κατασκευών και της ανακαίνισης.

Όσον αφορά τους ειδικούς που ασχολούνται με τη μηχανική μάθηση, η έρευνα αυτή αποτελεί σημαντική αφορμή προκειμένου να δουν τα διαφορετικά αποτελέσματα που προκύπτουν από τους αλγόριθμους που χρησιμοποιήθηκαν και ενδεχομένως να επαναξιολογήσουν τους ήδη υπάρχοντες αλγόριθμους. Παράλληλα

αποτελεί και μια ευκαιρία για περαιτέρω έρευνα για περισσότερη ανάλυση και εξειδίκευση στα μοντέλα μηχανικής μάθησης που χρησιμοποιούνται.

Τέλος, το σημαντικότερο όλων είναι ότι η έρευνα αυτή ανοίγει νέες δυνατότητες για τη βελτίωση της απόδοσης των κτιρίων, τη μείωση του ανθρωπίνου αποτυπώματος στο περιβάλλον και την επίτευξη οικονομικών οφελών καθώς χρησιμοποιεί δεδομένα τα οποία μπορεί ο χρήστης να καταγράψει γνωρίζοντας απλά τεχνικά χαρακτηριστικά του κτιρίου και ορισμένα στοιχεία από το λογαριασμό ενέργειας του κτιρίου. Αυτό επιτρέπει να εφαρμόζονται τα μοντέλα σε ένα ευρύ φάσμα κτιρίων χωρίς την ανάγκη για ειδικά συστήματα μέτρησης ή συλλογής δεδομένων. Μπορεί επομένως πολύ εύκολα να βρει πιο βιώσιμες λύσεις που εξοικονομούν ενέργεια και μειώνουν το κόστος λειτουργίας καθιστώντας το κτίριο το οποίο μελετάται άμεσα πιο ενεργειακά αποδοτικό συμβάλλοντας έτσι στο να δημιουργηθούν πιο έξυπνα, βιώσιμα και αποδοτικά κτίρια για το μέλλον.

6.3 Προτάσεις για περαιτέρω έρευνα

Η παρούσα διπλωματική εργασία προσφέρει μια σημαντική βάση για περαιτέρω έρευνα στον τομέα της ενεργειακής αποδοτικότητας στα κτίρια, ειδικά με έμφαση στην εφαρμογή μοντέλων τεχνητής νοημοσύνης και μηχανικής μάθησης. Μια σημαντική προοπτική για περαιτέρω έρευνα στον τομέα της ενεργειακής αποδοτικότητας σε κτιριακές δομές είναι η ανάλυση των δεδομένων με βάση τις διάφορες κατηγορίες κτιρίων. Η κατηγοριοποίηση των δεδομένων ανά είδος κτιρίου, όπως κατοικημένα, σχολεία, εμπορικά, βιομηχανικά κλπ., μπορεί να παρέχει πολύτιμες πληροφορίες σχετικά με τις ενεργειακές ανάγκες και τις βέλτιστες πρακτικές για κάθε κατηγορία. Μια περαιτέρω ανάλυση θα μπορούσε να εστιάσει στην εξέταση των παραγόντων που επηρεάζουν την ενεργειακή αποδοτικότητα σε κάθε κατηγορία κτιρίου, όπως η γεωγραφική τοποθεσία, το μέγεθος, η ηλιακή έκθεση, η χρήση του χώρου και άλλοι παράγοντες. Μια τέτοια ανάλυση θα διευκόλυνε την ανάπτυξη πιο εξατομικευμένων και αποτελεσματικών στρατηγικών για τη βελτίωση της ενεργειακής απόδοσης σε κάθε είδος κτιρίου.

Επιπλέον, μια ενδιαφέρουσα πτυχή που θα μπορούσε να εξεταστεί είναι η σύγκριση της αποτελεσματικότητας των διαφόρων μοντέλων μηχανικής μάθησης σε διαφορετικές κατηγορίες κτιρίων. Μια συστηματική σύγκριση των αλγορίθμων και τεχνικών μηχανικής μάθησης, όπως τα νευρωνικά δίκτυα, τα δέντρα απόφασης και οι μέθοδοι ομαδοποίησης, με βάση την ακρίβεια και την απόδοσή τους σε διαφορετικά περιβάλλοντα κτιρίων, θα μπορούσε να αποκαλύψει την καταλληλότερη προσέγγιση για κάθε κατηγορία κτιρίου.

Τέλος, μια περαιτέρω έρευνα θα μπορούσε να επικεντρωθεί στην εφαρμογή προηγμένων μεθόδων μηχανικής μάθησης, όπως η ενισχυτική μάθηση ή η μάθηση με επίβλεψη, για την ανάπτυξη εξειδικευμένων συστημάτων πρόβλεψης και ελέγχου της ενεργειακής απόδοσης σε κτίρια. Αυτές οι προηγμένες μέθοδοι μπορούν να επιτύχουν υψηλότερη ακρίβεια και απόδοση, ιδίως όταν οι προβλέψεις πρέπει να

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

λαμβάνουν υπόψη πολύπλοκους παράγοντες και δεδομένα. Η εφαρμογή αυτών των προηγμένων τεχνικών μπορεί να οδηγήσει σε πιο αποτελεσματικές στρατηγικές για τη βελτίωση της ενεργειακής απόδοσης στον κτιριακό τομέα.

ΚΕΦΑΛΑΙΟ 7

Βιβλιογραφία

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

7 Βιβλιογραφία

- [1] Nilsson, M. (2007). Red light for Green Paper: The EU policy on energy efficiency. *Energy Policy*, 35(1), 540–547. <https://doi.org/10.1016/j.enpol.2005.12.023>
- [2] Fawcett, T., Rosenow, J., & Bertoldi, P. (2018). Energy efficiency obligation schemes: their future in the EU. *Energy Efficiency*, 12(1), 57–71. <https://doi.org/10.1007/s12053-018-9657-1>
- [3] Galvin, R. (2024). The economic losses of energy-efficiency renovation of Germany's older dwellings: The size of the problem and the financial challenge it presents. *Energy Policy*, 184, 113905. <https://doi.org/10.1016/j.enpol.2023.113905>
- [4] Palm, J., & Reindl, K. (2017). Understanding barriers to energy-efficiency renovations of multifamily dwellings. *Energy Efficiency*, 11(1), 53–65. <https://doi.org/10.1007/s12053-017-9549-9>
- [5] Bertone, E., Sahin, O., Stewart, R. A., Zou, P., Alam, M., & Blair, E. (2016). State-of-the-art review revealing a roadmap for public building water and energy efficiency retrofit projects. *International Journal of Sustainable Built Environment*, 5(2), 526–548. <https://doi.org/10.1016/j.ijsbe.2016.09.004>
- [6] Cuerda, E., Guerra-Santin, O., Sendra, J. J., & Neila, Fco. J. (2020). Understanding the performance gap in energy retrofitting: Measured input data for adjusting building simulation models. *Energy and Buildings*, 209, 109688. <https://doi.org/10.1016/j.enbuild.2019.109688>
- [7] Sardianou, E. (2008). Barriers to industrial energy efficiency investments in Greece. *Journal of Cleaner Production*, 16(13), 1416–1423. <https://doi.org/10.1016/j.jclepro.2007.08.002>
- [8] https://treaties.un.org/pages/ViewDetails.aspx?src=TREATY&mtdsg_no=XXVII-7-d&chapter=27&clang=en
- [9] Crespo del Granado, P., Resch, G., Holz, F., Welisch, M., Geipel, J., Hartner, M., Forthuber, S., Sensfuss, F., Olmos, L., Bernath, C., Lumbreras, S., Kranzl, L., Müller, A., Heitel, S., Herbst, A., Wilson, C., & Ramose, A. (2020). Energy Transition Pathways to a low-carbon Europe in 2050: the degree of cooperation and the level of decentralization. *Economics of Energy & Environmental Policy*, 9(1). <https://doi.org/10.5547/2160-5890.9.1.pcre>
- [10] von Malmberg, F. (2023). First and last and always: Politics of the 'energy efficiency first' principle in EU energy and climate policy. *Energy Research & Social Science*, 101, 103126. <https://doi.org/10.1016/j.erss.2023.103126>
- [11] Olasolo-Alonso, P., López-Ochoa, L. M., Las-Heras-Casas, J., & López-González, L. M. (2023). Energy Performance of Buildings Directive implementation in Southern European countries: A review. *Energy and Buildings*, 281, 112751. <https://doi.org/10.1016/j.enbuild.2022.112751>
- [12] https://energy.ec.europa.eu/topics/energy-efficiency/energy-efficient-buildings/energy-performance-buildings-directive_en
- [13] https://energy.ec.europa.eu/news/new-energy-efficiency-directive-published-2023-09-20_en
- [14] Apostu, S. A., Panait, M., & Vasile, V. (2022). The energy transition in Europe—a solution for net zero carbon? *Environmental Science and Pollution Research*, 29(47), 71358–71379. <https://doi.org/10.1007/s11356-022-20730-z>
- [15] Agrawal, R., De Tommasi, L., Lyons, P., Zanoni, S., Papagiannis, G. K., Karakosta, C., Papapostolou, A., Durand, A., Martinez, L., Fragidis, G., Corbella, M., Sileni, L., Neusel,

- L., Repetto, M., Mariuzzo, I., Kakardakos, T., & Güemes, E. L. (2023). Challenges and opportunities for improving energy efficiency in SMEs: learnings from seven European projects. *Energy Efficiency*, 16(3). <https://doi.org/10.1007/s12053-023-10090-z>
- [16] El Naqa, I., & Murphy, M. J. (2015). *What is machine learning?* (pp. 3-11). Springer International Publishing.
- [17] Chicco, D. Ten quick tips for machine learning in computational biology. *BioData Mining* 10, 35 (2017). <https://doi.org/10.1186/s13040-017-0155-3>
- [18] John W. Goodell, Satish Kumar, Weng Marc Lim, Debidutta Pattnaik, Artificial intelligence and machine learning in finance: Identifying foundations, themes, and research clusters from bibliometric analysis, <https://doi.org/10.1016/j.jbef.2021.100577>.
- [19] Suresh, S., Sinha, N., & Prusty, S. (2020). Latent Approach in Entertainment Industry Using Machine Learning. *International Research Journal on Advanced Science Hub*, 2(Special Issue ICARD 2020), 304-307.
- [20] Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: Springer.
- [21] Islam, A. R. (2022). Machine learning in computer vision. In *Applications of Machine Learning and Artificial Intelligence in Education* (pp. 48-72). IGI Global.
- [22] Brunton, S. L., Nathan Kutz, J., Manohar, K., Aravkin, A. Y., Morgansen, K., Klemisch, J., ... & McDonald, D. (2021). Data-driven aerospace engineering: reframing the industry with machine learning. *AIAA Journal*, 59(8), 2820-2847.
- [23] Shehab, M., Abualigah, L., Shambour, Q., Abu-Hashem, M. A., Shambour, M. K. Y., Alsalibi, A. I., & Gandomi, A. H. (2022). Machine learning in medical applications: A review of state-of-the-art methods. *Computers in Biology and Medicine*, 145, 105458.
- [24] Kherif, F., & Latypova, A. (2020). Principal component analysis. In *Machine Learning* (pp. 209-225). Academic Press.
- [25] Rokach, L., & Maimon, O. (2005). Clustering methods. *Data mining and knowledge discovery handbook*, 321-352.
- [26] Shetty, S., & Rao, Y. S. (2016, August). SVM based machine learning approach to identify Parkinson's disease using gait analysis. In *2016 International conference on inventive computation technologies (ICICT)* (Vol. 2, pp. 1-5). IEEE.
- [27] Zou, J., Han, Y., & So, S. S. (2009). Overview of artificial neural networks. *Artificial neural networks: methods and applications*, 14-22.
- [28] Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983). AN OVERVIEW OF MACHINE LEARNING. *Machine Learning*, 3-23. <https://doi.org/10.1016/b978-0-08-051054-5.50005-4>
- [29] Liakos KG, Busato P, Moshou D, Pearson S, Bochtis D. Machine Learning in Agriculture: A Review. *Sensors*. 2018; 18(8):2674. <https://doi.org/10.3390/s18082674>
- [30] Alzubi, J., Nayyar, A., & Kumar, A. (2018, November). Machine learning from theory to algorithms: an overview. In *Journal of physics: conference series* (Vol. 1142, p. 012012). IOP Publishing.
- [31] Entezari, A., Aslani, A., Zahedi, R., & Noorollahi, Y. (2023). Artificial intelligence and machine learning in energy systems: A bibliographic perspective. *Energy Strategy Reviews*, 45, 101017.
- [32] Application of machine learning tools for energy efficiency in industry: A review, Diogo A.C. Narciso, F.G. Martins. <https://doi.org/10.1016/j.egy.2020.04.035>

- [33]Forootan MM, Larki I, Zahedi R, Ahmadi A. Machine Learning and Deep Learning in Energy Systems: A Review. *Sustainability*. 2022; 14(8):4832. <https://doi.org/10.3390/su14084832>
- [34]Rätz, M., Javadi, A. P., Baranski, M., Finkbeiner, K., & Müller, D. (2019). Automated data-driven modeling of building energy systems via machine learning algorithms. *Energy and Buildings*, 202, 109384.
- [35]Zhao, Y., Li, T., Zhang, X., & Zhang, C. (2019). Artificial intelligence-based fault detection and diagnosis methods for building energy systems: Advantages, challenges, and the future. *Renewable and Sustainable Energy Reviews*, 109, 85-101. <https://doi.org/10.1016/j.rser.2019.04.021>
- [36]Bourdeau, M., Zhai, X. qiang, Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society*, 48, 101533. <https://doi.org/10.1016/j.scs.2019.101533>
- [37]Jiangyan Liu, Guannan Li, Bin Liu, Kuining Li, Huanxin Chen, Knowledge discovery of data-driven-based fault diagnostics for building energy systems: A case study of the building variable refrigerant flow system, <https://doi.org/10.1016/j.energy.2019.02.161>
- [38]Yao, Z., Lum, Y., Johnston, A. *et al.* Machine learning for a sustainable energy future. *Nat Rev Mater* 8, 202–215 (2023). <https://doi.org/10.1038/s41578-022-00490-5>
- [39]Anh-Duc Pham, Ngoc-Tri Ngo, Thi Thu Ha Truong, Nhat-To Huynh, Ngoc-Son Truong, Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability, *Journal of Cleaner Production*, <https://doi.org/10.1016/j.jclepro.2020.121082>
- [40]Seyedzadeh, S., Rahimian, F., Glesk, I. *et al.* Machine learning for estimation of building energy consumption and performance: a review. *Vis. in Eng.* 6, 5 (2018). <https://doi.org/10.1186/s40327-018-0064-7>
- [41]Benavente-Peces C, Ibadah N. Buildings Energy Efficiency Analysis and Classification Using Various Machine Learning Technique Classifiers. *Energies*. 2020; 13(13):3497. <https://doi.org/10.3390/en13133497>
- [42]Moayed, H.; Bui, D.T.; Dounis, A.; Lyu, Z.; Foong, L.K. Predicting Heating Load in Energy-Efficient Buildings Through Machine Learning Techniques. *Appl. Sci.* **2019**, 9, 4338. <https://doi.org/10.3390/app9204338>
- [43]Egwim, C. N., Egunjobi, O. O., Gomes, A., & Alaka, H. (2021). A Comparative Study on Machine Learning Algorithms for Assessing Energy Efficiency of Buildings. *Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 546–566. https://doi.org/10.1007/978-3-030-93733-1_41
- [44]Tien, P. W., Wei, S., Darkwa, J., Wood, C., & Calautit, J. K. (2022). Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality – A Review. *Energy and AI*, 10, 100198. <https://doi.org/10.1016/j.egyai.2022.100198>
- [45]Sarmas, E., Marinakis, V., & Doukas, H. (2022). A data-driven multicriteria decision making tool for assessing investments in energy efficiency. *Operational Research*, 22(5), 5597–5616. <https://doi.org/10.1007/s12351-022-00727-9>
- [46]De Paola, A., Ortolani, M., Lo Re, G., Anastasi, G., & Das, S. K. (2014). Intelligent Management Systems for Energy Efficiency in Buildings. *ACM Computing Surveys*, 47(1), 1–38. <https://doi.org/10.1145/2611779>

- [47]S. K. Vishwakarma, P. Upadhyaya, B. Kumari and A. K. Mishra, "Smart Energy Efficient Home Automation System Using IoT," 2019 4th International Conference on Internet of Things: Smart Innovation and Usages (IoT-SIU), Ghaziabad, India, 2019, pp. 1-4, doi: 10.1109/IoT-SIU.2019.8777607.
- [48]H. Lee, W. -K. Park and I. -W. Lee, "A Home Energy Management System for Energy-Efficient Smart Homes," 2014 International Conference on Computational Science and Computational Intelligence, Las Vegas, NV, USA, 2014, pp. 142-145, doi: 10.1109/CSCI.2014.109.
- [49]Bhati, A., Hansen, M., & Chan, C. M. (2017). Energy conservation through smart homes in a smart city: A lesson for Singapore households. *Energy Policy*, 104, 230–239. <https://doi.org/10.1016/j.enpol.2017.01.032>
- [50]J. JeyaPadmini and K. R. Kashwan, "Effective power utilization and conservation in smart homes using IoT," 2015 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), Melmaruvathur, India, 2015, pp. 0195-0199, doi: 10.1109/ICCPEIC.2015.7259463.
- [51]Singh, P. P., Khosla, P. K., & Mittal, M. (2019). Energy Conservation in IoT-Based Smart Home and Its Automation. *Studies in Systems, Decision and Control*, 155–177. https://doi.org/10.1007/978-981-13-7399-2_7
- [52]Demir, B.E. A New Low-Cost Internet of Things-Based Monitoring System Design for Stand-Alone Solar Photovoltaic Plant and Power Estimation. *Appl. Sci.* **2023**, *13*, 13072. <https://doi.org/10.3390/app132413072>
- [53]Acuña Acurio, B.A.; Chérrez Barragán, D.E.; López, J.C.; Grijalva, F.; Rodríguez, J.C.; da Silva, L.C.P. Visual State Estimation for False Data Injection Detection of Solar Power Generation. *Eng. Proc.* **2023**, *47*, 5. <https://doi.org/10.3390/engproc2023047005>
- [54]Burlig, F., Knittel, C., Rapson, D., Reguant, M., & Wolfram, C. (2020). Machine Learning from Schools about Energy Efficiency. *Journal of the Association of Environmental and Resource Economists*, 7(6), 1181–1217. <https://doi.org/10.1086/710606>
- [55]Zekić-Sušac, M., Mitrović, S., & Has, A. (2021). Machine learning based system for managing energy efficiency of public sector as an approach towards smart cities. *International Journal of Information Management*, 58, 102074. <https://doi.org/10.1016/j.ijinfomgt.2020.102074>
- [56]Li, H., & Liu, Z. (2018). Performance Prediction and Optimization of Solar Water Heater via a Knowledge-Based Machine Learning Method. *Advances in Computer and Electrical Engineering*, 55–74. <https://doi.org/10.4018/978-1-5225-3935-3.ch002>
- [57]Karadimos, P., & Anthopoulos, L. (2023). Machine Learning-Based Energy Consumption Estimation of Wastewater Treatment Plants in Greece. *Energies*, 16(21), 7408. <https://doi.org/10.3390/en16217408>
- [58]Fischer, R., Jakobs, M., Mücke, S., & Morik, K. (2023). A Unified Framework for Assessing Energy Efficiency of Machine Learning. *Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, 39–54. https://doi.org/10.1007/978-3-031-23618-1_3
- [59]Han, J., Nishio, S., Kawano, H., & Wang, W. (1998). Generalization-based data mining in object-oriented databases using an object cube model. *Data & Knowledge Engineering*, 25(1-2), 55-97.
- [60]Linden, A., & Yarnold, P. R. (2016). Using data mining techniques to characterize participation in observational studies. *Journal of evaluation in clinical practice*, 22(6), 839-847.

- [61] Gorade, S. M., Deo, A., & Purohit, P. (2017). A study of some data mining classification techniques. *International Research Journal of Engineering and Technology*, 4(4), 3112-3115.
- [62] Ying, X. (2019, February). An overview of overfitting and its solutions. In *Journal of physics: Conference series* (Vol. 1168, p. 022022). IOP Publishing.
- [63] Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. *Annals of translational medicine*, 4(11).
- [64] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- [65] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., ... & Liu, T. Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.
- [66] Veisi, H. (2023). Introduction to SVM. In *Learning with Fractional Orthogonal Kernel Classifiers in Support Vector Machines: Theory, Algorithms and Applications* (pp. 3-18). Singapore: Springer Nature Singapore.
- [67] Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE). *Geoscientific model development discussions*, 7(1), 1525-1534.
- [68] Miles, J. (2005). R-squared, adjusted R-squared. *Encyclopedia of statistics in behavioral science*.
- [69] De Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean absolute percentage error for regression models. *Neurocomputing*, 192, 38-48.
- [70] Rigatti, S. J. (2017). Random Forest. *Journal of Insurance Medicine*, 47(1), 31-39. <https://doi.org/10.17849/insm-47-01-31-39.1>
- [71] Al-kasassbeh, M., Abbadi, M. A., & Al-Bustanji, A. M. (2020). LightGBM Algorithm for Malware Detection. *Intelligent Computing*, 391-403. https://doi.org/10.1007/978-3-030-52243-8_28
- [72] Rene Y. Choi, Aaron S. Coyner, Jayashree Kalpathy-Cramer, Michael F. Chiang, J. Peter Campbell; Introduction to Machine Learning, Neural Networks, and Deep Learning. *Trans. Vis. Sci. Tech.* 2020;9(2):14. <https://doi.org/10.1167/tvst.9.2.14>.
- [73] Mosavi A, Salimi M, Faizollahzadeh Ardabili S, Rabczuk T, Shamshirband S, Varkonyi-Koczy AR. State of the Art of Machine Learning Models in Energy Systems, a Systematic Review. *Energies*. 2019; 12(7):1301. <https://doi.org/10.3390/en12071301>
- [74] Fan, C., Xiao, F., Yan, C., Liu, C., Li, Z., & Wang, J. (2019). A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. *Applied Energy*, 235, 1551-1560. <https://doi.org/10.1016/j.apenergy.2018.11.081>
- [75] Stecuła, K., Wolniak, R., & Grebski, W. W. (2023). AI-Driven Urban Energy Solutions—From Individuals to Society: A Review. *Energies*, 16(24), 7988. <https://doi.org/10.3390/en16247988>
- [76] Liao, S.-H., Chu, P.-H., & Hsiao, P.-Y. (2012). Data mining techniques and applications – A decade review from 2000 to 2011. *Expert Systems with Applications*, 39(12), 11303-11311. <https://doi.org/10.1016/j.eswa.2012.02.063>
- [77] Seyedzadeh, S., Pour Rahimian, F., Oliver, S., Rodriguez, S., & Glesk, I. (2020). Machine learning modelling for predicting non-domestic buildings energy performance: A model to support deep energy retrofit decision-making. *Applied Energy*, 279, 115908. <https://doi.org/10.1016/j.apenergy.2020.115908>

ΚΕΦΑΛΑΙΟ 8

Παράρτημα

Εφαρμογή μηχανικής μάθησης για την πρόβλεψη επίδρασης έργων ενεργειακής ανακαίνισης κτιρίων

8 Παράρτημα

8.1 Παράρτημα 1: Συντομογραφίες

Στο παράρτημα που ακολουθεί παρουσιάζεται μια λίστα με τις συντομογραφίες που χρησιμοποιήθηκαν κατά την υλοποίηση της παρούσας διπλωματικής εργασίας αλλά και ένας πίνακας με τα κυριότερα επιστημονικά άρθρα που μελετήθηκαν με μία σύντομη περίληψη για το καθένα από αυτά.

Πίνακας 9: Πίνακας Συντομογραφιών

<u>Συντομογραφία</u>	<u>Πλήρης έκφραση</u>
EE	Ευρωπαϊκή Ένωση
ΕΕΛ	Εγκαταστάσεις Επεξεργασίας Λυμάτων
MM	Μηχανική Μάθηση
EEOs	Energy Efficiency Obligation Schemes
EPBD	Energy Performance of Buildings Directive
EED	Energy Efficiency Directive
ML	Machine Learning
EU	European Union
IoT	Internet of Things
RMSE	Root Mean Squared Error
IP	Internet Protocol
SOTA	State-of-the-Art
MSE	Mean Squared Error
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
GBRT	Gradient Boosted Regression Tree
MART	Multiple Additive Regression Tree
MBoost	Model Based Boosting
KNN	K-Nearest Neighbors
SVM	Support Vector Machine
ANN	Artificial Network
ReLU	Support Rectified Linear Activation

8.2 Παράρτημα 2: Περιγραφή επιστημονικών άρθρων βιβλιογραφικής ανασκόπησης

Πίνακας 10: Σύντομη περιγραφή των κυριότερων επιστημονικών άρθρων

<u>Επιστημονικά Άρθρα</u>	<u>Περίληψη</u>
[32]	Το παρόν άρθρο παρουσιάζει μια βιβλιογραφική ανασκόπηση σχετικά με τις μεθοδολογίες που αναφέρονται στην επιστημονική βιβλιογραφία και διερευνούν τη δυνητική αξία των βιομηχανικών δεδομένων μέσω της αξιοποίησης εργαλείων μηχανικής μάθησης για στόχους που σχετίζονται με την ενεργειακή αποδοτικότητα. Η εργασία αυτή εντοπίζει συγχρόνως και εξετάζει λεπτομερώς τις επιστημονικές συνεισφορές που έχουν δημοσιευτεί μέχρι σήμερα.
[31]	Η συγκεκριμένη εργασία χρησιμοποιεί το λογισμικό VOSviewer για να διερευνήσει και να επανεξετάσει τη χρήση της τεχνητής νοημοσύνης και της μηχανικής μάθησης στον τομέα της ενέργειας και προτείνει υποσχόμενους αλλά παραμελημένους ή ανεξερεύνητους τομείς στους οποίους μπορούν να χρησιμοποιηθούν αυτές οι έννοιες. Τα αποτελέσματα έδειξαν ότι από εμπορικής πλευράς, οι πατέντες που υποβλήθηκαν για την τεχνητή νοημοσύνη και τη μηχανική μάθηση σε τομείς που σχετίζονται με την ενέργεια είχαν κατακόρυφη αύξηση.
[35]	Η παρούσα εργασία έχει ως στόχο να κάνει μια ολοκληρωμένη βιβλιογραφική ανασκόπηση των μεθόδων ανίχνευσης και διάγνωσης σφαλμάτων (FDD) που βασίζονται στην τεχνητή νοημοσύνη για ενεργειακά συστήματα κτιρίων τα τελευταία είκοσι χρόνια από το 1998 έως το 2018, να συνοψίσει τα δυνατά σημεία και τις ελλείψεις των υφιστάμενων μεθόδων που βασίζονται στην τεχνητή νοημοσύνη και να αποκαλύψει τα σημαντικότερα ερευνητικά καθήκοντα στο μέλλον.
[36]	Η παρούσα μελέτη παρέχει μια επισκόπηση των ερευνών που έχουν διεξαχθεί για την κατασκευή μοντέλων μηχανικής μάθησης βασισμένων σε δεδομένα για εφαρμογές σε κλίμακα κτιρίου. Περιγράφονται οι κοινές μεθοδολογίες, με έμφαση στα χαρακτηριστικά των δεδομένων εισόδου και στις μεθόδους προεπεξεργασίας των δεδομένων, στις τυπολογίες κτιρίων που εξετάζονται, στις στοχευόμενες τελικές χρήσεις ενέργειας και στα χρονοδιαγράμματα πρόβλεψης, καθώς και στην αξιολόγηση της ακρίβειας.
[37]	Η παρούσα εργασία προτείνει μια μεθοδολογία η οποία είναι σε θέση να διεξάγει τόσο τη διάγνωση βλαβών όσο και την ανακάλυψη διαγνωστικής ελέγχων για ενεργειακά συστήματα κτιρίων. Μια μελέτη εφαρμόζεται σε ένα πειραματικό σύστημα μεταβλητής ροής ψυκτικού μέσου (VRF).
[56]	Σε αυτό το επιστημονικό άρθρο, οι συγγραφείς παρουσιάζουν πώς χρησιμοποιούν τα μοντέλα μηχανικής μάθησης που εκπαιδεύονται από μια μεγάλη πειραματική βάση δεδομένων για να εκτελέσουν ακριβή πρόβλεψη και βελτιστοποίηση σε ένα σύστημα ηλιακού θερμοσίφωνα (SWH). Συγχρόνως προτείνεται μια νέα στρατηγική βελτιστοποίησης ενεργειακού συστήματος που βασίζεται σε μια διαδικασία διαλογής υψηλής απόδοσης (HTS).

[38]	Σε αυτό το άρθρο επισημαίνονται οι πρόσφατες εξελίξεις στην έρευνα στον τομέα της ενέργειας με βάση την Μηχανική Μάθηση, υπογραμμίζονται οι τρέχουσες και μελλοντικές προκλήσεις και περιγράφεται τι απαιτείται για την καλύτερη δυνατή χρήση των τεχνικών ML. Ειδική μνεία γίνεται με βάσει ένα σύνολο βασικών δεικτών επιδόσεων στους διάφορους μηχανισμούς ML στην ενεργειακή έρευνα.
[73]	Το παρόν επιστημονικό άρθρο παρουσιάζει την κατάσταση της τεχνολογίας των μοντέλων ML που χρησιμοποιούνται στα ενεργειακά συστήματα μαζί με μια νέα ταξινόμηση των μοντέλων και των εφαρμογών. Καταλήγει περαιτέρω στο συμπέρασμα ότι υπάρχει εξαιρετική αύξηση της ακρίβειας, της ευρωστίας και της ικανότητας γενίκευσης των μοντέλων ML στα ενεργειακά συστήματα με τη χρήση υβριδικών μοντέλων ML.
[29]	Αυτό το έγγραφο παρουσιάζει μια ολοκληρωμένη ανασκόπηση της έρευνας που είναι αφιερωμένη στις εφαρμογές της μηχανικής μάθησης στα συστήματα γεωργικής παραγωγής. Οι εργασίες που αναλύθηκαν κατηγοριοποιήθηκαν σε (α) διαχείριση καλλιεργειών, (β) διαχείριση ζωικού κεφαλαίου, (γ) διαχείριση νερού- και (δ) διαχείριση εδάφους.
[39]	Η παρούσα μελέτη προσθέτει (i) πληροφορίες εξετάζοντας τη γενίκευση και την αποτελεσματικότητα των μοντέλων ML στην πρόβλεψη των προτύπων ενεργειακής κατανάλωσης των κτιρίων και (ii) στην κατάσταση της πρακτικής, προτείνοντας ένα αποτελεσματικό εργαλείο για να βοηθήσει τους ιδιοκτήτες κτιρίων και τους διαχειριστές εγκαταστάσεων στην κατανόηση της ενεργειακής απόδοσης των κτιρίων με σκοπό τη βελτίωση της ενεργειακής απόδοσης των κτιρίων.
[41]	Στο παρόν έγγραφο, παρουσιάζεται πώς μπορούν να εφαρμοστούν οι τεχνολογίες και οι τεχνικές των ΤΠΕ και της επιστήμης των δεδομένων για την αξιολόγηση της ενεργειακής απόδοσης των κτιρίων. Συγκεκριμένα, εφαρμόζονται τεχνικές μηχανικής μάθησης για την ταξινόμηση των κτιρίων με βάση την ενεργειακή τους απόδοση. Ειδικότερα, εστιάζει σε κτίρια μίας οικογένειας σε κατοικημένες περιοχές.
[42]	Στην παρούσα μελέτη, χρησιμοποιήθηκαν έξι τεχνικές μηχανικής μάθησης, δηλαδή τον multi-layer perceptron regressor (MLPr), lazy locally weighted learning (LLWL), alternating model tree (AMT), random forest (RF), ElasticNet (ENet), and radial basis function regression (RBFr) για το πρόβλημα του σχεδιασμού ενεργειακά αποδοτικών κτιρίων. Στη συνέχεια, οι προσεγγίσεις αυτές χρησιμοποιήθηκαν για τον προσδιορισμό μιας σχέσης μεταξύ των παραμέτρων εισόδου και εξόδου όσον αφορά την ενεργειακή αποδοτικότητα των κτιρίων ενώ εν τέλει ιεραρχήθηκαν με βάσει κάποιες μετρικές απόδοσης.
[46]	Σκοπός του παρόντος άρθρου επισκόπησης είναι να παράσχει μια δομημένη και ολοκληρωμένη αντιμετώπιση της υπάρχουσας βιβλιογραφίας σχετικά με τα ευφυή συστήματα διαχείρισης ενέργειας σε κτίρια, με ιδιαίτερη έμφαση στις διαθέσιμες αρχιτεκτονικές και μεθοδολογίες που υποστηρίζουν ένα όραμα που υπερβαίνει το καθιερωμένο όραμα του έξυπνου σπιτιού και εισέρχεται σε μία νέα οπτική της νοημοσύνης περιβάλλοντος.

[40]	Η παρούσα μελέτη παρουσιάζει μια διεξοδική εξέταση των τεσσάρων κύριων μεθοδολογιών μηχανικής μάθησης (ML), συμπεριλαμβανομένων των τεχνητών νευρωνικών δικτύων, των μηχανών διανυσμάτων υποστήριξης, των παλινδρομήσεων με βάση τη Γκαουσιανή και της ομαδοποίησης, που έχουν χρησιμοποιηθεί ευρέως για την πρόβλεψη και τη βελτιστοποίηση της ενεργειακής απόδοσης των κτιρίων.
[52]	Στην παρούσα μελέτη, παρουσιάζεται ένα οικονομικά αποδοτικό σύστημα τηλεπαρακολούθησης με βάση το Διαδίκτυο των πραγμάτων (IoT) για ηλιακά φωτοβολταϊκά ενεργειακά συστήματα, καθώς και ένας εκτιμητής φωτοβολταϊκής ισχύος με βάση τη μηχανική μάθηση.
[53]	Η έρευνα αυτή προτείνει την ανίχνευση των FDIAs με τη χρήση οπτικών δεδομένων. Η οπτική εκτίμηση της κατάστασης αποσκοπεί στη βελτίωση της ανθεκτικότητας και της ασφάλειας των συστημάτων ανανεώσιμων πηγών ενέργειας. Η λύση αυτή προσθέτει ένα ακόμη επίπεδο προστασίας έναντι απειλών στον κυβερνοχώρο, διατηρώντας την ακεραιότητα και την αξιοπιστία των δεδομένων παραγωγής ηλιακής ενέργειας και υποστηρίζοντας την αποτελεσματική και ασφαλή λειτουργία των ΣΠΔ.
[47]	Η παρούσα μελέτη προτείνει ένα έξυπνο, ενεργειακά αποδοτικό σύστημα οικιακού αυτοματισμού που μπορεί να έχει πρόσβαση και να χειρίζεται τον οικιακό εξοπλισμό από οποιοδήποτε σημείο του πλανήτη κάνοντας έτσι το σπίτι μας πιο «έξυπνο» και ασφαλές.
[48]	Στην παρούσα έρευνα προτείνεται ένα σύστημα οικιακής ενεργειακής διαχείρισης για ένα ενεργειακά αποδοτικό έξυπνο σπίτι. Πρώτον, συλλέγονται και αξιολογούνται δεδομένα χρήσης ενέργειας για ένα έξυπνο σπίτι και έπειτα αναπτύσσονται και εφαρμόζονται μέθοδοι εξοικονόμησης ενέργειας, ενώ διατηρείται η άνεση των χρηστών.
[43]	Στην παρούσα εργασία παρέχεται μια ιδέα για τη σύγκριση της απόδοσης αλγορίθμων ML για την αξιολόγηση της ενεργειακής απόδοσης κτιρίων. Αρχικά, χρησιμοποιούνται συγκεντρωτικές αξιολογήσεις ενεργειακής απόδοσης κτιρίων από διάφορες πηγές δεδομένων για τη δημιουργία μοντέλων πρόβλεψης με τη χρήση διαφόρων μεθόδων μηχανικής μάθησης. Στη συνέχεια, προσδιορίζεται το μοντέλο με τις καλύτερες επιδόσεις με τη σύγκριση κριτηρίων αξιολόγησης, όπως το RMSE, το R-Squared και το Adjusted R-Squared.
[54]	Σε αυτό το άρθρο αξιολογείται ο αντίκτυπος των ανακαινίσεων ενεργειακής απόδοσης σε σχολεία K-12 στην Καλιφόρνια, χρησιμοποιώντας δεδομένα πάνελ υψηλής συχνότητας σχετικά με την ηλεκτρική ενέργεια.
[55]	Το συγκεκριμένο άρθρο προσπαθεί να λύσει το ερώτημα πώς μπορεί να συμπεριληφθεί μια πλατφόρμα μεγάλων δεδομένων και η μηχανική μάθηση σε ένα ευφύες σύστημα διαχείρισης της ενεργειακής απόδοσης του δημόσιου τομέα ως σημαντικό στοιχείο της ιδέας της έξυπνης πόλης. Για τη δημιουργία μοντέλων πρόβλεψης της συγκεκριμένης ενεργειακής κατανάλωσης των κτηρίων του δημόσιου τομέα της Κροατίας χρησιμοποιήθηκαν βαθιά νευρωνικά δίκτυα, Rpart regression tree και Random Forest με διαδικασίες μείωσης των μεταβλητών.
[57]	Η παρούσα μελέτη εξετάζει δεδομένα από πολλαπλά δημοτικά έργα ΕΕΛ (εγκαταστάσεις επεξεργασίας λυμάτων) στην Ελλάδα προκειμένου να

	αναπτυχθούν μοντέλα εκτίμησης ΕΚ. Η μελέτη υπογραμμίζει τη δυνητική δυνατότητα γενίκευσης αυτών των μοντέλων στην Ελλάδα και πέραν αυτής, παρέχοντας πολύτιμα εργαλεία στους ενδιαφερόμενους φορείς για την ενημέρωση της λήψης αποφάσεων, την αποτελεσματική κατανομή των πόρων και τη βελτίωση των ενεργειακά αποδοτικών σχεδιασμών, με αποτέλεσμα την εξοικονόμηση κόστους και τα οφέλη της βιωσιμότητας.
[58]	Στην παρούσα εργασία παρουσιάζονται καινοτόμοι αλγόριθμοι συγκέντρωσης απόδοσης, ευρετηρίασης και ταξινόμησης για εφαρμογές Μηχανικής Μάθησης. Για τον σκοπό αυτό, προτείνεται ένα σύνολο μέτρων που επιτρέπουν μια ολοκληρωμένη εικόνα, λαμβάνοντας υπόψη τον τύπο της εργασίας, το αφηρημένο μοντέλο, το λογισμικό και το υλικό.
[49]	Το άρθρο αυτό επικεντρώνεται στις αντιλήψεις των οικογενειών της Σιγκαπούρης για την έξυπνη τεχνολογία και την εφαρμογή της για την εξοικονόμηση ενέργειας. Τα τρέχοντα θέματα μελέτης περιλαμβάνουν: (1) η κατανάλωση ενέργειας στα νοικοκυριά της Σιγκαπούρης, (2) οι δημόσιες πρωτοβουλίες και πολιτικές εξοικονόμησης ενέργειας, (3) η χρήση τεχνολογίας εξοικονόμησης ενέργειας και (4) η αντίληψη των νοικοκυριών για την εξοικονόμηση ενέργειας στα έξυπνα σπίτια. Επιπλέον, αναλύονται τρεις μελέτες περιπτώσεων που σχετίζονται με έξυπνα σπίτια και έξυπνες τεχνολογίες, καθώς παρουσιάζεται και μια συζήτηση σχετικά με την ωριμότητα των υφιστάμενων λύσεων.
[51]	Το κεφάλαιο αυτό εξετάζει ολόκληρο το ηλεκτρικό δίκτυο, από την παραγωγή έως τη μεταφορά και τη διανομή στον χρήστη. Εξετάζονται ο σχεδιασμός του ηλεκτρικού δικτύου και τα σημεία παρακολούθησης της ενέργειας, καθώς και οι πηγές των απωλειών ενέργειας και οι δυνατότητες μείωσης των απωλειών ενέργειας καθώς επίσης και τα συστήματα αυτοματοποιημένης διαχείρισης της ενεργειακής ζήτησης και η λειτουργία των κτιρίων στην εξοικονόμηση ενέργειας.
[50]	Η παρούσα έρευνα προτείνει μια απλή προσέγγιση βασισμένη στο Διαδίκτυο των πραγμάτων (IoT) για τον εντοπισμό της ανθρώπινης δραστηριότητας με τη χρήση ανάλυσης εικόνας. Πρόκειται για μια μέθοδο διαχείρισης ενέργειας σε πραγματικό χρόνο που περιλαμβάνει την επικοινωνία μεταξύ μηχανών.
[44]	Η παρούσα μελέτη παρέχει μια κριτική σύνθεση της βιβλιογραφίας της προηγούμενης δεκαετίας σχετικά με τις μεθόδους μηχανικής και βαθιάς μάθησης για το δομημένο περιβάλλον, με έμφαση στις ολιστικές προσεγγίσεις. Εξετάστηκαν διάφορες στρατηγικές βασισμένες στην ΤΝ που χρησιμοποιούνται για την αντιμετώπιση αλληλένδετων προβλημάτων με συστήματα θέρμανσης, εξαερισμού και κλιματισμού (HVAC) και τη βελτίωση της απόδοσης των κτιρίων. Η παρούσα μελέτη εξετάζει επίσης τις υπάρχουσες μεθοδολογίες στον τομέα της ενέργειας που βασίζονται στην ΤΝ, δηλαδή το πλαίσιο, τη μεθοδολογία και τις επιδόσεις.
[74]	Το παρόν άρθρο παρουσιάζει ένα πλήρες πλαίσιο για την επεξήγηση και αξιολόγηση των μοντέλων ενεργειακής απόδοσης κτιρίων που βασίζονται σε δεδομένα, προκειμένου να αυξηθεί η πρακτική χρησιμότητα των σύγχρονων τεχνικών μηχανικής μάθησης στον τομέα των κτιρίων. Η τεχνική βασίζεται σε μια ερμηνεύσιμη αρχιτεκτονική μηχανικής μάθησης.
[77]	Η παρούσα έρευνα προτείνει ένα μοντέλο πρόβλεψης της ενεργειακής απόδοσης μη οικιακών κτιρίων με βάση τη μηχανική μάθηση. Ο στόχος του

	μοντέλου είναι να παρέχει μια γρήγορη μηχανή εκτίμησης της ενεργειακής απόδοσης για να βοηθήσει στην πολύπλευρη αντικειμενική βελτιστοποίηση των σχεδίων ενεργειακής αναβάθμισης μη οικιακών κτιρίων. Η εργασία περιγράφει τη διαδικασία δημιουργίας του μοντέλου, από τη διερεύνηση των απαιτήσεων και την εξαγωγή χαρακτηριστικών έως την εφαρμογή σε μια μελέτη περίπτωσης.
[75]	Η παρούσα μελέτη παρουσιάζει μια ολοκληρωμένη αξιολόγηση των λύσεων που βασίζονται στην τεχνητή νοημοσύνη (TN) στον τομέα της αστικής ενέργειας, με έμφαση στις εφαρμογές και τις επιπτώσεις τους. Στη μελέτη αυτή χρησιμοποιείται μια τεχνική βιβλιογραφικής ανασκόπησης για να εξετάσει την πρόσφατη έρευνα σχετικά με τη συμμετοχή της τεχνητής νοημοσύνης σε λύσεις που σχετίζονται με την ενέργεια από το 2019 έως το 2023.