



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΙΑΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Εκπαίδευση μοντέλου μηχανικής μάθησης για πρόβλεψη μεταβολής τιμών καυσίμων

ΕΥΑΓΓΕΛΟΣ ΡΟΚΚΑΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

ΚΟΛΟΜΒΑΤΣΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ
ΑΝΑΠΛΗΡΩΤΗΣ ΚΑΘΗΓΗΤΗΣ

Λαμία 2026



UNIVERSITY OF
THESSALY

SCHOOL OF SCIENCE

DEPARTMENT OF INFORMATICS & TELECOMMUNICATIONS

TRAINING A MACHINE LEARNING MODEL TO PREDICT FUEL PRICE CHANGES

EVANGELOS ROKKAS

FINAL THESIS

ADVISOR
KOLOMVASOS KONSTANTINOS
ASSOCIATE PROFESSOR

Lamia 2026

«Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις ⁽¹⁾, που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:

1. Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί **χωρίς να τα περικλείω σε εισαγωγικά** και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάστηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.
2. Δέχομαι ότι η αυτολεξεί **παράθεση χωρίς εισαγωγικά**, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.
3. Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια
4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.

Ημερομηνία:/...../20.....

Ο – Η Δηλ.

(1) «Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»

ΠΕΡΙΛΗΨΗ

Η πρόβλεψη των τιμών καυσίμων αποτελεί ένα ιδιαίτερα σημαντικό αντικείμενο μελέτης, καθώς επηρεάζει άμεσα την οικονομική δραστηριότητα, τον τομέα των μεταφορών και τον ενεργειακό σχεδιασμό. Η πολυπλοκότητα των παραγόντων που διαμορφώνουν τις τιμές των καυσίμων καθιστά αναγκαία την αξιοποίηση σύγχρονων μεθόδων ανάλυσης και πρόβλεψης, ικανών να επεξεργάζονται αποτελεσματικά μεγάλους όγκους ιστορικών δεδομένων.

Σκοπός της παρούσας εργασίας είναι η ανάπτυξη ενός αυτοματοποιημένου συστήματος συλλογής, επεξεργασίας και πρόβλεψης τιμών καυσίμων, αξιοποιώντας τεχνικές μηχανικής μάθησης και ομαδικής μάθησης (Ensemble Learning). Το σύστημα πραγματοποιεί αυτόματη συλλογή ημερήσιων τιμών καυσίμων από διαδικτυακές πηγές και δημιουργεί ένα οργανωμένο ιστορικό σύνολο δεδομένων, το οποίο χρησιμοποιείται για την εκπαίδευση και αξιολόγηση μοντέλων πρόβλεψης.

Για την εκτίμηση των μελλοντικών τιμών εφαρμόζονται πολλαπλοί αλγόριθμοι μηχανικής μάθησης, οι οποίοι εκπαιδεύονται σε χαρακτηριστικά που προκύπτουν από ιστορικές τιμές, χρονικές υστερήσεις (lag features) και στατιστικά κυλιόμενων παραθύρων. Ως μοντέλο αναφοράς χρησιμοποιείται η Ridge Regression, ενώ παράλληλα αξιοποιούνται αλγόριθμοι γραμμικής παλινδρόμησης, μοντέλα βασισμένα σε δέντρα αποφάσεων, καθώς και μη γραμμικές μέθοδοι πρόβλεψης.

Η απόδοση των μοντέλων αξιολογείται μέσω του Σφάλματος Μέσης Απόλυτης Απόκλισης (Mean Absolute Error – MAE). Με βάση τα αποτελέσματα της αξιολόγησης επιλέγονται τα πέντε μοντέλα με την καλύτερη απόδοση, των οποίων οι προβλέψεις συνδυάζονται μέσω τεχνικής Ensemble Learning για την παραγωγή της τελικής πρόβλεψης των επόμενων επτά ημερών.

Τα αποτελέσματα δείχνουν ότι ο συνδυασμός πολλαπλών μοντέλων συμβάλλει στη βελτίωση της ακρίβειας και της σταθερότητας των προβλέψεων, μειώνοντας την εξάρτηση από ένα μόνο μοντέλο. Η προτεινόμενη προσέγγιση αποτελεί μια αξιόπιστη και επεκτάσιμη λύση για την πρόβλεψη τιμών καυσίμων και μπορεί να αξιοποιηθεί ως βάση για περαιτέρω έρευνα και ανάπτυξη συστημάτων πρόβλεψης στον ενεργειακό τομέα.

ABSTRACT

Fuel price forecasting constitutes an important research area, as fuel prices directly affect economic activity, transportation systems, and energy planning. The complexity of the factors influencing fuel price fluctuations necessitates the use of advanced analytical and forecasting techniques capable of effectively processing large volumes of historical data.

The objective of this thesis is the development of an automated system for the collection, processing, and forecasting of fuel prices using Machine Learning and Ensemble Learning techniques. The proposed system performs automated daily collection of fuel price data from online sources and constructs a structured historical dataset that is subsequently utilized for model training and evaluation.

To estimate future fuel prices, multiple machine learning algorithms are employed and trained using features derived from historical observations, lag-based variables, and rolling statistical indicators. Ridge Regression is used as a baseline model, while additional predictive approaches include linear regression techniques, tree-based algorithms, and non-linear forecasting methods.

The performance of the developed models is evaluated using the Mean Absolute Error (MAE) metric. Based on the evaluation results, the five best-performing models are selected, and their forecasts are combined through an Ensemble Learning approach to generate the final seven-day fuel price forecast.

The experimental results indicate that the combination of multiple predictive models improves forecasting accuracy and robustness while reducing dependence on a single algorithm. The proposed approach provides a reliable and extensible solution for fuel price forecasting and may serve as a foundation for future research and the development of advanced predictive systems in the energy sector.

Table of Contents

ΠΕΡΙΛΗΨΗ	I
ABSTRACT	II
ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΗ	3
1.1 Η ΣΗΜΑΣΙΑ ΤΗΣ ΠΡΟΒΛΕΨΗΣ ΤΙΜΩΝ ΚΑΥΣΙΜΩΝ	3
1.2 Η ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΣΤΗΝ ΑΝΑΛΥΣΗ ΚΑΙ ΠΡΟΒΛΕΨΗ ΧΡΟΝΟΣΕΙΡΩΝ	3
1.3 ΣΚΟΠΟΣ ΚΑΙ ΑΝΤΙΚΕΙΜΕΝΟ ΤΗΣ ΕΡΓΑΣΙΑΣ	4
1.4 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	4
ΚΕΦΑΛΑΙΟ 2 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ANALYTICS	6
2.1 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΠΩΣ ΛΕΙΤΟΥΡΓΕΙ;	6
2.2 ΚΑΤΗΓΟΡΙΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	6
2.2.1 SUPERVISED LEARNING	6
2.2.2 UNSUPERVISED LEARNING	7
2.2.3 REINFORCEMENT LEARNING	8
2.3 ΑΛΓΟΡΙΘΜΟΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ ΠΡΟΒΛΕΨΗ ΧΡΟΝΟΣΕΙΡΩΝ	8
2.3.1 ΓΡΑΜΜΙΚΑ ΜΟΝΤΕΛΑ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (LINEAR, RIDGE, LASSO, ELASTIC NET)	9
2.3.2 ΑΛΓΟΡΙΘΜΟΙ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΔΕΝΤΡΑ (DECISION TREE, RANDOM FOREST, EXTRA TREES, GRADIENT BOOSTING, ADABOOST, BAGGING)	9
2.3.3 ΑΛΓΟΡΙΘΜΟΙ ΒΑΣΙΣΜΕΝΟΙ ΣΕ ΠΥΡΗΝΕΣ ΚΑΙ ΑΠΟΣΤΑΣΗ (SVR, KNN)	10
2.4 ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ DATA ANALYTICS	11
2.5 ΕΙΔΗ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	11
2.6 ΟΜΑΔΙΚΗ ΜΑΘΗΣΗ (ENSEMBLE LEARNING)	12
ΚΕΦΑΛΑΙΟ 3 ΠΡΟΚΛΗΣΕΙΣ ΣΤΟΝ ΟΙΚΟΝΟΜΙΚΟ ΤΟΜΕΑ	13
3.1 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΤΩΝ ΟΙΚΟΝΟΜΙΚΩΝ ΔΕΔΟΜΕΝΩΝ	13
3.2 ΠΡΟΗΓΟΥΜΕΝΕΣ ΕΡΓΑΣΙΕΣ (RELATED WORK)	13

<u>3.3 ΔΥΣΚΟΛΙΕΣ ΣΤΗΝ ΠΡΟΒΛΕΨΗ ΟΙΚΟΝΟΜΙΚΩΝ ΔΕΔΟΜΕΝΩΝ</u>	<u>14</u>
<u>3.4 ΕΞΩΤΕΡΙΚΟΙ ΠΑΡΑΓΟΝΤΕΣ ΠΟΥ ΕΠΗΡΕΑΖΟΥΝ ΤΙΣ ΟΙΚΟΝΟΜΙΚΕΣ ΑΓΟΡΕΣ</u>	<u>15</u>
<u>3.5 ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΟΙΚΟΝΟΜΙΚΩΝ ΔΕΔΟΜΕΝΩΝ</u>	<u>15</u>
<u>ΚΕΦΑΛΑΙΟ 4 ΠΡΟΤΕΙΝΟΜΕΝΟ ΣΥΣΤΗΜΑ</u>	<u>17</u>
<u>4.1 ΕΙΣΑΓΩΓΗ.....</u>	<u>17</u>
4.1.1 ΣΤΟΧΟΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ	17
4.1.2 ΓΕΝΙΚΗ ΛΕΙΤΟΥΡΓΙΑ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ	17
<u>4.2 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ.....</u>	<u>17</u>
4.2.1 ΤΑ ΔΕΔΟΜΕΝΑ ΚΑΙ Η ΠΗΓΗ ΤΟΥΣ	17
4.2.2 ΔΙΑΔΙΚΑΣΙΑ ΕΞΑΓΩΓΗΣ ΔΕΔΟΜΕΝΩΝ	18
4.2.3 ΑΠΟΘΗΚΕΥΣΗ ΤΩΝ ΕΞΑΓΟΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ	19
<u>4.3 ΔΗΜΙΟΥΡΓΙΑ ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΤΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ</u>	<u>20</u>
4.3.1 ΟΡΓΑΝΩΣΗ ΚΑΙ ΣΥΜΠΛΗΡΩΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ	20
4.3.2 ΔΗΜΙΟΥΡΓΙΑ ΣΤΑΤΙΣΤΙΚΩΝ ΔΕΔΟΜΕΝΩΝ	21
<u>4.4 ΔΗΜΙΟΥΡΓΙΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΓΙΑ ΤΑ ΜΟΝΤΕΛΑ ΠΡΟΒΛΕΨΗΣ</u>	<u>21</u>
4.4.1 ΧΡΗΣΗ ΠΡΟΗΓΟΥΜΕΝΩΝ ΤΙΜΩΝ (LAGS)	22
4.4.2 ΥΠΟΛΟΓΙΣΜΟΣ ΚΙΝΗΤΟΥ ΜΕΣΟΥ ΟΡΟΥ	22
4.4.3 ΔΗΜΙΟΥΡΓΙΑ ΤΟΥ ΤΕΛΙΚΟΥ ΣΥΝΟΛΟΥ ΔΕΔΟΜΕΝΩΝ	22
<u>4.5 ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΕΦΑΡΜΟΓΗ ΤΩΝ ΜΟΝΤΕΛΩΝ ΠΡΟΒΛΕΨΗΣ</u>	<u>23</u>
<u>4.6 ΣΥΝΔΥΑΣΜΟΣ ΜΟΝΤΕΛΩΝ ΠΡΟΒΛΕΨΗΣ(ENSEMBLE)</u>	<u>24</u>
<u>ΚΕΦΑΛΑΙΟ 5 ΑΠΟΤΕΛΕΣΜΑΤΑ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗ ΜΟΝΤΕΛΩΝ</u>	<u>25</u>
<u>5.1 ΕΙΣΑΓΩΓΗ ΚΑΙ ΔΙΑΔΙΚΑΣΙΑ ΑΞΙΟΛΟΓΗΣΗΣ</u>	<u>25</u>
<u>5.2 ΑΠΟΤΕΛΕΣΜΑΤΑ ΜΟΝΤΕΛΩΝ ΚΑΙ ΠΡΟΒΛΕΨΕΙΣ</u>	<u>25</u>
<u>5.3 ΑΝΑΛΥΣΗ ΕΠΙΛΕΓΜΕΝΩΝ ΜΟΝΤΕΛΩΝ</u>	<u>26</u>
5.3.1 ΠΑΡΑΔΕΙΓΜΑ 1 ^ο	26

5.3.2 ΠΑΡΑΔΕΙΓΜΑ 2 ^ο	27
<u>5.4 ΤΕΛΙΚΗ ΠΡΟΒΛΕΨΗ ΤΙΜΩΝ ΚΑΥΣΙΜΩΝ (NEXT 7 DAYS)</u>	<u>28</u>
5.4.1 ΕΦΑΡΜΟΓΗ ΤΟΥ PENALTY MECHANISM.....	28
5.4.2 ΤΕΛΙΚΗ ΠΡΟΒΛΕΨΗ ΓΙΑ ΤΗΝ ΕΛΛΑΔΑ (BENZINΗ)	29
5.4.3 ΤΕΛΙΚΗ ΠΡΟΒΛΕΨΗ ΓΙΑ ΤΗ ΓΑΛΛΙΑ (ΠΕΤΡΕΛΑΙΟ).....	29
5.4.4 ΤΕΛΙΚΗ ΠΡΟΒΛΕΨΗ ΓΙΑ ΤΗ ΕΛΛΑΔΑ (ΥΓΡΑΕΡΙΟ LPG)	30
5.4.5 ΤΕΛΙΚΗ ΠΡΟΒΛΕΨΗ ΓΙΑ ΤΗ ΓΕΡΜΑΝΙΑ (BENZINΗ)- PENALTY.....	30
5.5 ΣΥΓΚΡΙΣΗ ΠΡΟΒΛΕΠΟΜΕΝΩΝ ΚΑΙ ΠΡΑΓΜΑΤΙΚΩΝ ΤΙΜΩΝ ΚΑΥΣΙΜΩΝ.....	31
<u>ΚΕΦΑΛΑΙΟ 6 ΣΥΜΠΕΡΑΣΜΑΤΑ</u>	<u>33</u>
<u>6.1 ΣΥΜΠΕΡΑΣΜΑΤΑ.....</u>	<u>33</u>
<u>6.2 ΜΕΛΛΟΝΤΙΚΕΣ ΒΕΛΤΙΩΣΕΙΣ.....</u>	<u>33</u>
<u>ΒΙΒΛΙΟΓΡΑΦΙΑ</u>	<u>35</u>

ΚΕΦΑΛΑΙΟ 1 Εισαγωγή

1.1 Η σημασία της πρόβλεψης τιμών καυσίμων

Οι τιμές των καυσίμων αποτελούν έναν από τους σημαντικότερους παράγοντες που επηρεάζουν την οικονομική δραστηριότητα, τις μεταφορές και την καθημερινή λειτουργία της κοινωνίας. Η συνεχής μεταβολή των τιμών της βενζίνης, του πετρελαίου κίνησης και άλλων ενεργειακών προϊόντων επηρεάζει άμεσα το κόστος μετακίνησης, τη λειτουργία των επιχειρήσεων και τη συνολική οικονομική σταθερότητα.

Η εξέλιξη των τιμών καυσίμων καθορίζεται από ένα σύνολο σύνθετων και αλληλεξαρτώμενων παραγόντων, όπως η προσφορά και η ζήτηση στην παγκόσμια αγορά ενέργειας, οι γεωπολιτικές εξελίξεις, οι φορολογικές πολιτικές των κρατών και οι διακυμάνσεις στις διεθνείς τιμές του πετρελαίου. Η πολυπλοκότητα των παραγόντων αυτών καθιστά ιδιαίτερα δύσκολη την ακριβή πρόβλεψη των τιμών με τη χρήση παραδοσιακών στατιστικών μεθόδων.

Τα τελευταία χρόνια, η αξιοποίηση σύγχρονων υπολογιστικών τεχνικών, και ειδικότερα της μηχανικής μάθησης, έχει προσφέρει νέες δυνατότητες στην ανάλυση οικονομικών και ενεργειακών δεδομένων. Τα μοντέλα μηχανικής μάθησης έχουν τη δυνατότητα να επεξεργάζονται μεγάλα σύνολα ιστορικών δεδομένων και να εντοπίζουν κρυφές σχέσεις και πρότυπα, τα οποία μπορούν να αξιοποιηθούν για την πρόβλεψη μελλοντικών τιμών.



1.2 Η μηχανική μάθηση στην ανάλυση και πρόβλεψη χρονοσειρών

Η μηχανική μάθηση αποτελεί έναν από τους βασικότερους κλάδους της τεχνητής νοημοσύνης και έχει αναδειχθεί σε ένα ιδιαίτερα ισχυρό εργαλείο για την ανάλυση δεδομένων και την υποστήριξη διαδικασιών πρόβλεψης. Στον τομέα της ανάλυσης οικονομικών και ενεργειακών δεδομένων, τα δεδομένα συνήθως παρουσιάζονται με τη μορφή χρονοσειρών, δηλαδή διαδοχικών παρατηρήσεων που καταγράφονται σε συγκεκριμένες χρονικές στιγμές.

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να εκπαιδευτούν χρησιμοποιώντας ιστορικά δεδομένα και να δημιουργήσουν μοντέλα πρόβλεψης που είναι ικανά να εκτιμήσουν τη

μελλοντική εξέλιξη μιας μεταβλητής. Μέσω της διαδικασίας εκπαίδευσης, τα μοντέλα αυτά αναγνωρίζουν σχέσεις μεταξύ των δεδομένων και μαθαίνουν να προβλέπουν τις μελλοντικές τιμές με αυξανόμενη ακρίβεια.

Ιδιαίτερο ενδιαφέρον παρουσιάζουν οι τεχνικές ομαδικής μάθησης (ensemble learning), οι οποίες συνδυάζουν τα αποτελέσματα πολλαπλών αλγορίθμων με σκοπό τη βελτίωση της συνολικής απόδοσης του συστήματος πρόβλεψης. Με τον τρόπο αυτό μειώνεται η επίδραση των αδυναμιών ενός μεμονωμένου μοντέλου και αυξάνεται η αξιοπιστία των παραγόμενων προβλέψεων.

1.3 Σκοπός και αντικείμενο της εργασίας

Σκοπός της παρούσας πτυχιακής εργασίας είναι η ανάπτυξη ενός ολοκληρωμένου συστήματος πρόβλεψης των τιμών καυσίμων μέσω της αξιοποίησης τεχνικών μηχανικής μάθησης και ανάλυσης χρονοσειρών. Το προτεινόμενο σύστημα βασίζεται στη συλλογή ιστορικών δεδομένων από διαδικτυακές πηγές, στη δημιουργία και επεξεργασία ενός κατάλληλα διαμορφωμένου συνόλου δεδομένων και στην εκπαίδευση μοντέλων πρόβλεψης με στόχο την εκτίμηση μελλοντικών μεταβολών των τιμών.

Στο πλαίσιο της εργασίας υλοποιείται μια ολοκληρωμένη μεθοδολογία η οποία περιλαμβάνει την εφαρμογή και αξιολόγηση πολλαπλών αλγορίθμων μηχανικής μάθησης. Η συγκριτική ανάλυση των μοντέλων πραγματοποιείται με τη χρήση κατάλληλων μετρικών αξιολόγησης, όπως το Σφάλμα Μέσης Απόλυτης Απόκλισης (Mean Absolute Error – MAE), προκειμένου να διασφαλιστεί η αντικειμενική αποτίμηση της προγνωστικής τους ικανότητας και η επιλογή των πλέον αποδοτικών προσεγγίσεων.

Ιδιαίτερη έμφαση δίνεται στην αξιοποίηση τεχνικών ομαδικής μάθησης (Ensemble Learning), μέσω των οποίων συνδυάζονται οι προβλέψεις των μοντέλων με την καλύτερη απόδοση. Η προσέγγιση αυτή αποσκοπεί στη βελτίωση της ακρίβειας, της σταθερότητας και της αξιοπιστίας των τελικών προβλέψεων, περιορίζοντας παράλληλα την εξάρτηση από ένα μεμονωμένο μοντέλο.

Το προτεινόμενο σύστημα φιλοδοξεί να αποτελέσει ένα αξιόπιστο και επεκτάσιμο εργαλείο για την ανάλυση και πρόβλεψη τιμών καυσίμων, συμβάλλοντας στην καλύτερη κατανόηση της δυναμικής των ενεργειακών δεδομένων και υποστηρίζοντας τη λήψη τεκμηριωμένων αποφάσεων σε οικονομικό και ενεργειακό επίπεδο.

1.4 Δομή της εργασίας

Η παρούσα εργασία οργανώνεται σε έξι κεφάλαια, τα οποία παρουσιάζουν σταδιακά το θεωρητικό υπόβαθρο, τη μεθοδολογία ανάπτυξης, την υλοποίηση του προτεινόμενου συστήματος και την αξιολόγηση των αποτελεσμάτων.

Στο πρώτο κεφάλαιο παρουσιάζεται το αντικείμενο της εργασίας, η σημασία της πρόβλεψης τιμών καυσίμων στον ενεργειακό και οικονομικό τομέα, καθώς και οι στόχοι και η συνεισφορά της προτεινόμενης προσέγγισης.

Στο δεύτερο κεφάλαιο αναλύονται οι βασικές έννοιες της μηχανικής μάθησης και της ανάλυσης δεδομένων, ενώ παρουσιάζονται οι κυριότερες κατηγορίες αλγορίθμων που χρησιμοποιούνται σε προβλήματα πρόβλεψης χρονοσειρών. Επιπλέον, γίνεται αναφορά στις τεχνικές ομαδικής μάθησης (Ensemble Learning), οι οποίες αποτελούν βασικό στοιχείο της προτεινόμενης λύσης.

Στο τρίτο κεφάλαιο παρουσιάζονται σχετικές ερευνητικές εργασίες που αφορούν την πρόβλεψη τιμών καυσίμων και ενεργειακών δεδομένων, καθώς και οι κυριότερες προκλήσεις που εμφανίζονται κατά την ανάλυση οικονομικών χρονοσειρών και την ανάπτυξη μοντέλων πρόβλεψης.

Στο τέταρτο κεφάλαιο περιγράφεται αναλυτικά το προτεινόμενο σύστημα. Ειδικότερα, παρουσιάζονται η διαδικασία συλλογής δεδομένων από διαδικτυακές πηγές, η αποθήκευση και προεπεξεργασία τους, η δημιουργία χαρακτηριστικών (features) με βάση ιστορικές τιμές και στατιστικά μεγέθη, καθώς και η εκπαίδευση των μοντέλων μηχανικής μάθησης.

Στο πέμπτο κεφάλαιο παρουσιάζονται τα αποτελέσματα της πειραματικής διαδικασίας. Αναλύεται η απόδοση των επιμέρους αλγορίθμων με βάση τη μετρική Mean Absolute Error (MAE), η διαδικασία επιλογής των καλύτερων μοντέλων και η παραγωγή των τελικών προβλέψεων μέσω της τεχνικής Ensemble Learning για τις επόμενες επτά ημέρες.

Τέλος, στο έκτο κεφάλαιο συνοψίζονται τα βασικά συμπεράσματα της εργασίας, αξιολογούνται τα πλεονεκτήματα και οι περιορισμοί της προτεινόμενης προσέγγισης και διατυπώνονται προτάσεις για μελλοντική έρευνα και περαιτέρω βελτίωση του συστήματος.

ΚΕΦΑΛΑΙΟ 2 Μηχανική Μάθηση και Analytics

2.1 Μηχανική Μάθηση και Πώς Λειτουργεί;

Η μηχανική μάθηση (Machine Learning) [15] αποτελεί έναν σημαντικό υποκλάδο της τεχνητής νοημοσύνης (Artificial Intelligence – AI), ο οποίος έχει αλλάξει σημαντικά τον τρόπο με τον οποίο οι υπολογιστές επεξεργάζονται και αξιοποιούν τα δεδομένα. Η βασική ιδέα της μηχανικής μάθησης είναι η δυνατότητα των υπολογιστικών συστημάτων να μαθαίνουν από δεδομένα, να αναγνωρίζουν πρότυπα και να εξάγουν χρήσιμες πληροφορίες χωρίς να απαιτείται ρητός προγραμματισμός για κάθε πιθανή περίπτωση.

Μέσω της ανάλυσης μεγάλων συνόλων δεδομένων, τα μοντέλα μηχανικής μάθησης μπορούν να εντοπίζουν σχέσεις μεταξύ των μεταβλητών και να χρησιμοποιούν αυτές τις πληροφορίες για την πρόβλεψη μελλοντικών τιμών ή τη λήψη αποφάσεων. Η ικανότητα αυτή καθιστά τη μηχανική μάθηση ένα ιδιαίτερα ισχυρό εργαλείο σε πολλούς επιστημονικούς και τεχνολογικούς τομείς.

Με την αύξηση του όγκου των δεδομένων στη σύγχρονη ψηφιακή εποχή, η ανάγκη για συστήματα που μπορούν να διαχειρίζονται και να αναλύουν μεγάλα και πολύπλοκα σύνολα δεδομένων έχει γίνει ιδιαίτερα σημαντική. Τα δεδομένα αυτά, τα οποία συχνά αναφέρονται ως Big Data, προέρχονται από διάφορες πηγές, όπως αισθητήρες, διαδικτυακές υπηρεσίες, κινητές συσκευές και κοινωνικά δίκτυα.

Κατά τη διαδικασία εκπαίδευσης ενός μοντέλου μηχανικής μάθησης, οι αλγόριθμοι αναλύουν τα διαθέσιμα δεδομένα με στόχο την αναγνώριση μοτίβων και τη δημιουργία ενός μαθηματικού μοντέλου που μπορεί να χρησιμοποιηθεί για την πρόβλεψη νέων δεδομένων. Με την επαναλαμβανόμενη εκπαίδευση και αξιολόγηση, τα μοντέλα αυτά βελτιώνουν σταδιακά την απόδοσή τους, οδηγώντας σε πιο ακριβή και αξιόπιστα αποτελέσματα.

Η μηχανική μάθηση εφαρμόζεται σήμερα σε πολλούς τομείς, όπως η οικονομία, η υγεία, οι μεταφορές, η ενέργεια και η ανάλυση δεδομένων, προσφέροντας σημαντικές δυνατότητες για την κατανόηση σύνθετων συστημάτων και τη λήψη τεκμηριωμένων αποφάσεων.

2.2 Κατηγορίες Μηχανικής Μάθησης

2.2.1 Supervised Learning

Η Supervised Learning [15] (επιβλεπόμενη μάθηση) είναι ένα υποσύνολο του machine learning, όπου τα μοντέλα εκπαιδεύονται με ετικετοποιημένα σύνολα δεδομένων (labeled datasets). Τα ετικετοποιημένα σύνολα δεδομένων χρησιμοποιούνται για την εκπαίδευση αλγορίθμων που ταξινομούν δεδομένα ή κάνουν προβλέψεις με υψηλή ακρίβεια. Ο όρος "ετικετοποιημένο" υποδηλώνει ότι ορισμένα δεδομένα εισόδου έχουν ήδη τη σωστή έξοδο. Κάθε σημείο δεδομένων, συνεπώς, συσχετίζεται με την αντίστοιχη ετικέτα εξόδου. Η

επιβλεπόμενη μάθηση μπορεί να χρησιμοποιηθεί για διάφορες περιπτώσεις, όπως αξιολόγηση κινδύνου, ανίχνευση απάτης, φιλτράρισμα ανεπιθύμητων μηνυμάτων (spam) και πολλά άλλα.

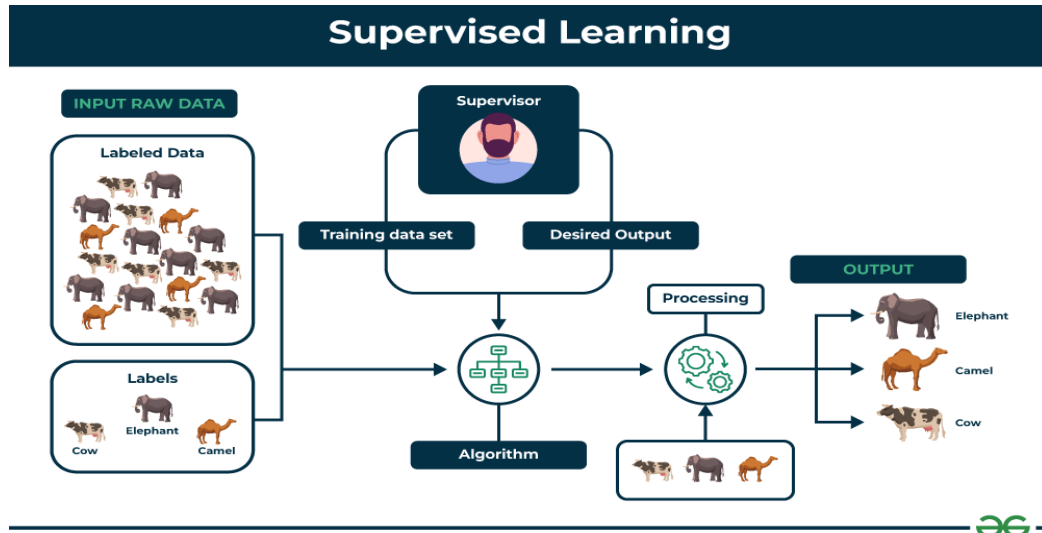


Figure 1 Supervised Learning

2.2.2 Unsupervised Learning

Η Unsupervised Learning [16] (Μη επιβλεπόμενη μάθηση) αναφέρεται σε μια κατηγορία τεχνικής μηχανικής μάθησης όπου αλγόριθμοι τεχνητής νοημοσύνης αναζητούν πρότυπα σε σύνολα δεδομένων χωρίς προκαθορισμένες κατηγορίες ή labels. Τα μοντέλα αυτά δεν απαιτούν επίβλεψη κατά την εκπαίδευση με δεδομένα, καθιστώντας την ιδανική για την εντοπισμό προτύπων, ομάδων και διαφορών σε μη δομημένα σύνολα δεδομένων. Είναι κατάλληλη για διαδικασίες όπως ο διαχωρισμός πελατών, η ανάλυση ανακάλυψης δεδομένων ή η αναγνώριση εικόνων.

Οι αλγόριθμοι Unsupervised Learning μπορούν να οργανώσουν και να ομαδοποιήσουν τα δεδομένα χωρίς εξωτερική καθοδήγηση.

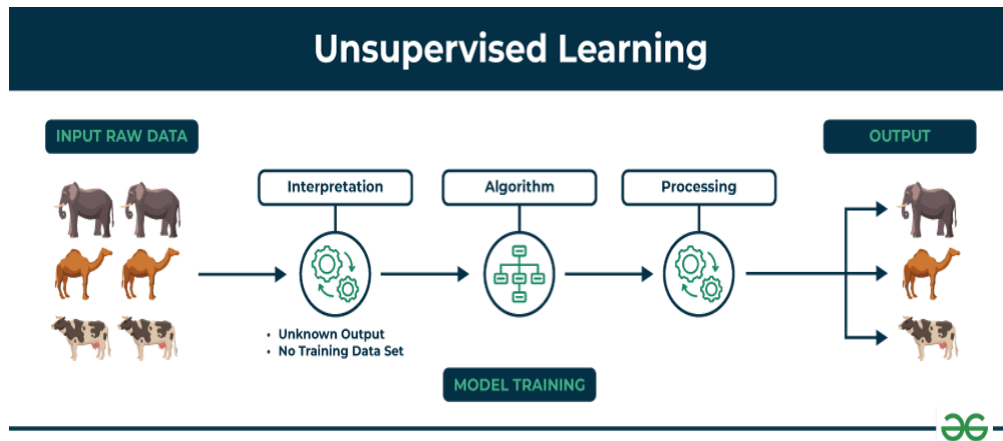


Figure 2 Unsupervised Learning

2.2.3 Reinforcement Learning

Η Reinforcement Learning [17] (Ενισχυτική μάθηση) είναι η επιστημονική μεθοδολογία της λήψης αποφάσεων, που επιδιώκει να μάθει πώς να συμπεριφέρεται βέλτιστα σε ένα περιβάλλον για να αποκτήσει τη μεγαλύτερη δυνατή ανταμοιβή. Αυτή η βέλτιστη συμπεριφορά αποκτάται μέσω της δοκιμής και του λάθους, αντίστοιχα με το πώς τα παιδιά εξερευνούν και μαθαίνουν τις ενέργειες που τους οδηγούν σε επιτυχία. Χωρίς επόπτη, ο μαθητής πρέπει να ανακαλύψει ανεξάρτητα την καλύτερη ακολουθία ενεργειών που θα του φέρουν την μέγιστη ανταμοιβή, κάτι που γίνεται με την παρατήρηση των αντιδράσεων του περιβάλλοντος. Η ισχύς της ενίσχυσης μάθησης έγκειται στην ικανότητά της να μαθαίνει από την εμπειρία χωρίς εξωτερική καθοδήγηση.



REINFORCEMENT LEARNING MODEL

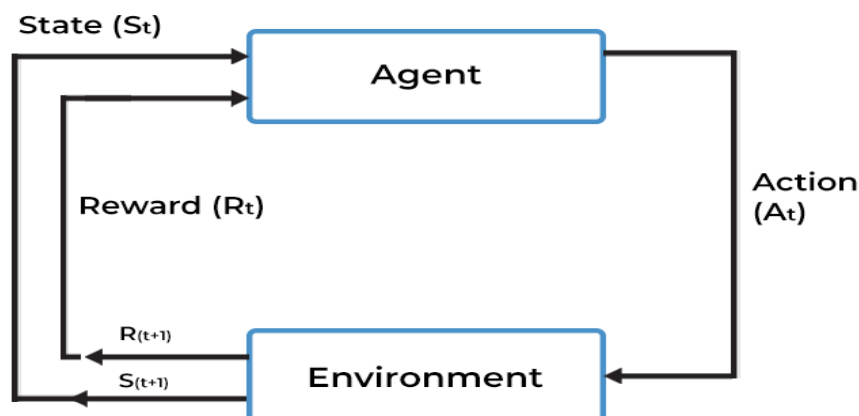


Figure 3 Reinforcement Learning

2.3 Αλγόριθμοι μηχανικής μάθησης για πρόβλεψη χρονοσειρών

Η πρόβλεψη χρονοσειρών αποτελεί ένα σημαντικό πεδίο εφαρμογής της μηχανικής μάθησης, καθώς επιτρέπει την εκτίμηση μελλοντικών τιμών με βάση ιστορικά δεδομένα. Στον οικονομικό τομέα, η ανάλυση χρονοσειρών χρησιμοποιείται ευρέως για την πρόβλεψη τιμών, την κατανόηση των τάσεων της αγοράς και τη λήψη αποφάσεων.

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να αξιοποιήσουν τα ιστορικά δεδομένα για να εντοπίσουν πρότυπα και σχέσεις μεταξύ των μεταβλητών. Μέσω της εκπαίδευσης των μοντέλων είναι δυνατόν να δημιουργηθούν συστήματα που μπορούν να προβλέψουν την πιθανή εξέλιξη μιας μεταβλητής στο μέλλον.

Στην παρούσα εργασία χρησιμοποιούνται διάφορες κατηγορίες αλγορίθμων μηχανικής μάθησης που εφαρμόζονται συχνά σε προβλήματα πρόβλεψης αριθμητικών τιμών. Οι αλγόριθμοι αυτοί μπορούν να ταξινομηθούν σε διαφορετικές κατηγορίες ανάλογα με τη μεθοδολογία που χρησιμοποιούν για την ανάλυση των δεδομένων.

Στις επόμενες ενότητες παρουσιάζονται ορισμένες βασικές κατηγορίες αλγορίθμων μηχανικής μάθησης που χρησιμοποιούνται συχνά σε προβλήματα πρόβλεψης χρονοσειρών.

2.3.1 Γραμμικά μοντέλα παλινδρόμησης (Linear, Ridge, Lasso, Elastic Net)

Στην κατηγορία αυτή ανήκει η Linear Regression,[1],[2] η οποία χρησιμοποιείται για την πρόβλεψη αριθμητικών τιμών μέσω μιας απλής γραμμικής σχέσης μεταξύ των δεδομένων. Η Ridge Regression λειτουργεί με παρόμοιο τρόπο, αλλά εφαρμόζει κανονικοποίηση ώστε να αποφεύγεται η υπερπροσαρμογή. Η Lasso Regression μπορεί επιπλέον να μειώσει την επίδραση λιγότερο σημαντικών χαρακτηριστικών, ενώ η Elastic Net Regression συνδυάζει τα πλεονεκτήματα των Ridge και Lasso, προσφέροντας μεγαλύτερη ευελιξία κατά την εκπαίδευση του μοντέλου.

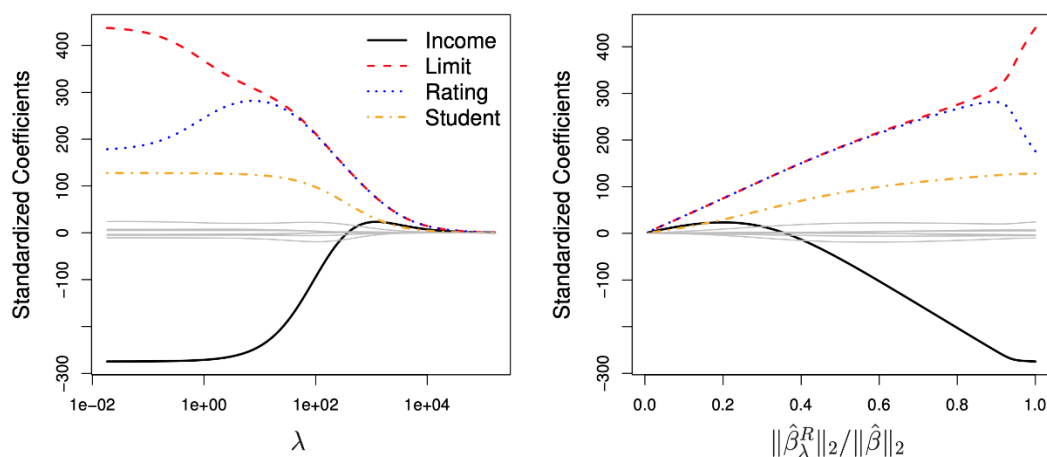


Figure 4 Μεταβολή των συντελεστών με την κανονικοποίηση

2.3.2 Αλγόριθμοι βασισμένοι σε δέντρα (Decision Tree, Random Forest, Extra Trees, Gradient Boosting, AdaBoost, Bagging)

Οι αλγόριθμοι αυτής της κατηγορίας βασίζονται στη δημιουργία δέντρων αποφάσεων για την ανάλυση των δεδομένων. Ο **Decision Tree Regressor** παράγει προβλέψεις μέσω διαδοχικών κανόνων διαχωρισμού των δεδομένων, ενώ ο **Random Forest Regressor** και ο **Extra Trees Regressor** συνδυάζουν πολλά δέντρα για μεγαλύτερη σταθερότητα και ακρίβεια. Οι μέθοδοι **Gradient Boosting**, **AdaBoost** και **Bagging** βασίζονται στη συνεργασία πολλών μοντέλων, με στόχο τη βελτίωση της τελικής πρόβλεψης και τη μείωση των σφαλμάτων.[3],[4],[5]

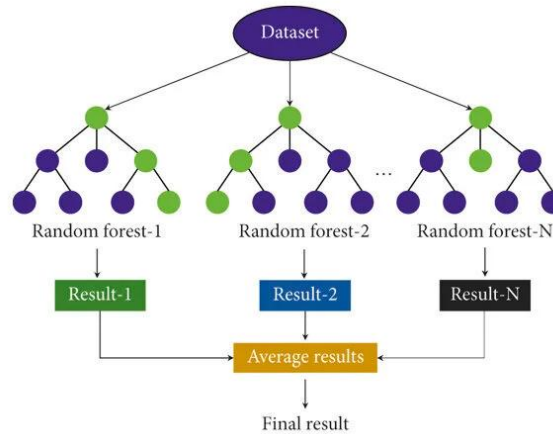


Figure 5 Random Forest

2.3.3 Αλγόριθμοι βασισμένοι σε πυρήνες και απόσταση (SVR, KNN)

Η Support Vector Regression (SVR) [6] χρησιμοποιείται για την πρόβλεψη αριθμητικών τιμών και μπορεί να μοντελοποιήσει πιο σύνθετες σχέσεις μεταξύ των δεδομένων. Ο K-Nearest Neighbors (KNN) [7] βασίζεται στην ιδέα ότι παρόμοιες παρατηρήσεις έχουν συνήθως παρόμοια συμπεριφορά και πραγματοποιεί προβλέψεις χρησιμοποιώντας τις κοντινότερες εγγραφές του συνόλου δεδομένων.

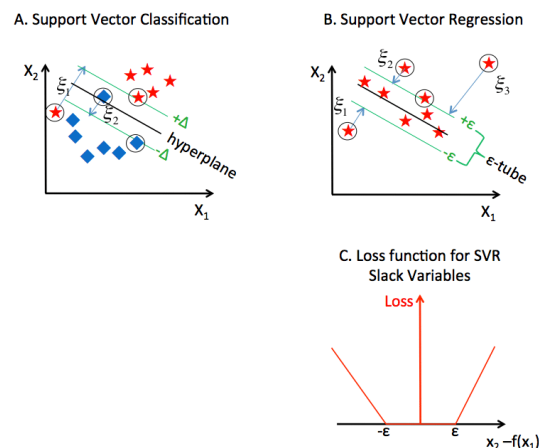


Figure 6 Support Vector Regression (SVR)

2.4 Ανάλυση Δεδομένων και Data Analytics

Η ανάλυση δεδομένων (Data Analytics) [8] αποτελεί ένα από τα σημαντικότερα στάδια στη διαδικασία ανάπτυξης μοντέλων μηχανικής μάθησης. Μέσω της ανάλυσης, τα δεδομένα συλλέγονται, οργανώνονται και προετοιμάζονται κατάλληλα πριν από την εκπαίδευση των μοντέλων.

Η διαδικασία περιλαμβάνει τον καθαρισμό των δεδομένων, την αντιμετώπιση ελλειπών ή λανθασμένων τιμών, καθώς και τη μετατροπή τους σε μορφή κατάλληλη για επεξεργασία. Επιπλέον, εφαρμόζονται τεχνικές όπως η κανονικοποίηση και η τυποποίηση, ώστε να διασφαλιστεί η ομοιομορφία των δεδομένων.

Σημαντικό ρόλο διαδραματίζουν και οι τεχνικές οπτικοποίησης, οι οποίες επιτρέπουν την καλύτερη κατανόηση της δομής των δεδομένων και την αναγνώριση τάσεων και προτύπων. Μέσω γραφημάτων και στατιστικών αναλύσεων, οι ερευνητές μπορούν να εντοπίσουν χρήσιμες πληροφορίες που συμβάλλουν στη βελτίωση της απόδοσης των μοντέλων.

Η ποιότητα των δεδομένων επηρεάζει άμεσα την αποτελεσματικότητα των μοντέλων μηχανικής μάθησης, καθιστώντας την προεπεξεργασία ένα ιδιαίτερα κρίσιμο στάδιο.

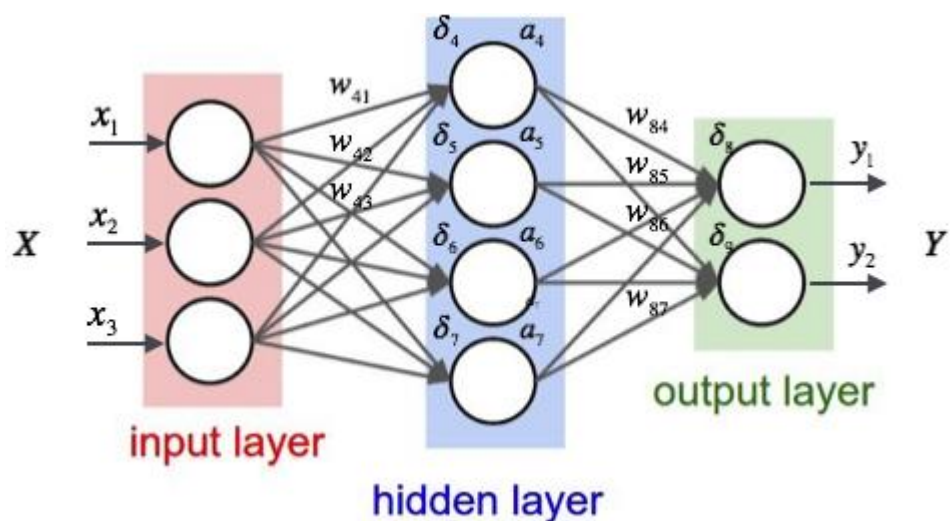


Figure 7 Neural Network Layers

2.5 Είδη αλγορίθμων μηχανικής μάθησης

Η αξιολόγηση των μοντέλων μηχανικής μάθησης αποτελεί ένα κρίσιμο στάδιο στη διαδικασία ανάπτυξης ενός συστήματος πρόβλεψης. Μέσω της αξιολόγησης είναι δυνατή η εκτίμηση της απόδοσης κάθε μοντέλου και η σύγκρισή του με εναλλακτικές προσεγγίσεις.

Για την αξιολόγηση μοντέλων που προβλέπουν αριθμητικές τιμές χρησιμοποιούνται διάφορες μετρικές σφάλματος. Μία από τις πιο διαδεδομένες είναι το σφάλμα μέσης απόλυτης

απόκλισης (Mean Absolute Error – MAE) [1] , το οποίο υπολογίζει τη μέση απόλυτη διαφορά μεταξύ των πραγματικών και των προβλεπόμενων τιμών.

Η χρήση της μετρικής MAE επιτρέπει την αντικειμενική σύγκριση διαφορετικών μοντέλων και βοηθά στον εντοπισμό των πιο αποδοτικών αλγορίθμων για το συγκεκριμένο πρόβλημα.

Στο πλαίσιο της παρούσας εργασίας, η αξιολόγηση των μοντέλων βασίζεται στη σύγκριση των τιμών MAE, με στόχο την επιλογή των μοντέλων που παρουσιάζουν τη μικρότερη απόκλιση και, συνεπώς, τη μεγαλύτερη ακρίβεια πρόβλεψης

2.6 Ομαδική Μάθηση (Ensemble Learning)

Η ομαδική μάθηση (Ensemble Learning) [9] αποτελεί μια σημαντική τεχνική στη μηχανική μάθηση, κατά την οποία συνδυάζονται πολλαπλά μοντέλα πρόβλεψης με σκοπό τη βελτίωση της συνολικής απόδοσης του συστήματος. Η βασική ιδέα της μεθόδου είναι ότι διαφορετικοί αλγόριθμοι μπορεί να έχουν διαφορετικά πλεονεκτήματα και αδυναμίες, επομένως ο συνδυασμός των προβλέψεών τους μπορεί να οδηγήσει σε πιο αξιόπιστα αποτελέσματα.

Σε πολλές περιπτώσεις, ένα μεμονωμένο μοντέλο μπορεί να παρουσιάζει σφάλματα πρόβλεψης λόγω περιορισμών στη δομή του ή λόγω της φύσης των δεδομένων. Με τη χρήση της ομαδικής μάθησης, οι προβλέψεις πολλών διαφορετικών μοντέλων συνδυάζονται, μειώνοντας την πιθανότητα σφάλματος και αυξάνοντας τη σταθερότητα των αποτελεσμάτων.

Υπάρχουν διάφορες τεχνικές ensemble learning, όπως το bagging, το boosting και ο μέσος όρος προβλέψεων πολλαπλών μοντέλων. Στην τελευταία περίπτωση, οι προβλέψεις διαφορετικών αλγορίθμων συνδυάζονται υπολογίζοντας τον μέσο όρο των αποτελεσμάτων τους, δημιουργώντας έτσι μια τελική πρόβλεψη που βασίζεται στη συλλογική απόδοση των μοντέλων.

Η προσέγγιση αυτή χρησιμοποιείται ευρέως σε προβλήματα πρόβλεψης χρονοσειρών και οικονομικών δεδομένων, καθώς μπορεί να βελτιώσει σημαντικά την ακρίβεια των προβλέψεων. Με τον τρόπο αυτό, η ομαδική μάθηση συμβάλλει στη δημιουργία πιο αξιόπιστων συστημάτων πρόβλεψης και στη μείωση της εξάρτησης από ένα μόνο μοντέλο.

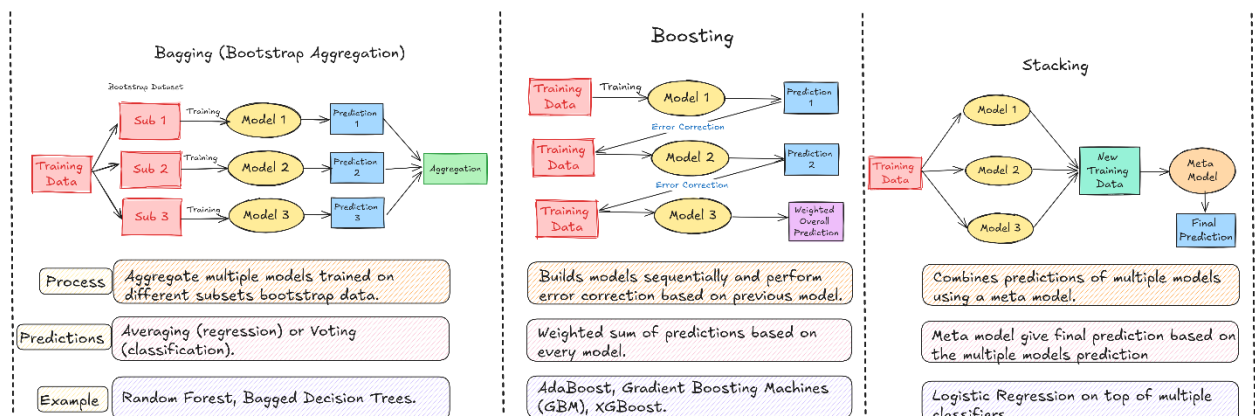


Figure 8 Ensemble Learning

ΚΕΦΑΛΑΙΟ 3 Προκλήσεις στον Οικονομικό Τομέα

3.1 Χαρακτηριστικά των οικονομικών δεδομένων

Ο οικονομικός τομέας χαρακτηρίζεται από μεγάλη πολυπλοκότητα και συνεχή μεταβολή των δεδομένων. Οι τιμές προϊόντων και υπηρεσιών επηρεάζονται από πολλούς παράγοντες, όπως η προσφορά και η ζήτηση, οι διεθνείς αγορές, οι κυβερνητικές πολιτικές και οι μεταβολές στις τιμές των πρώτων υλών. Για τον λόγο αυτό, η ανάλυση οικονομικών δεδομένων αποτελεί μια ιδιαίτερα απαιτητική διαδικασία.

Στον ενεργειακό τομέα, και ειδικότερα στις τιμές των καυσίμων, οι μεταβολές αυτές εμφανίζονται συχνά και σε σύντομο χρονικό διάστημα. Οι τιμές της βενζίνης, του πετρελαίου και άλλων ενεργειακών προϊόντων επηρεάζουν άμεσα την οικονομία, καθώς σχετίζονται με το κόστος μεταφορών, την παραγωγή προϊόντων και τη λειτουργία των επιχειρήσεων.

Τα οικονομικά δεδομένα συχνά εμφανίζουν υψηλή μεταβλητότητα, γεγονός που σημαίνει ότι οι τιμές μπορεί να παρουσιάζουν μεγάλες αυξήσεις ή μειώσεις μέσα σε μικρό χρονικό διάστημα. Επιπλέον, τα δεδομένα αυτά μπορεί να επηρεάζονται από πολλούς παράγοντες ταυτόχρονα, γεγονός που καθιστά δύσκολη την κατανόηση των πραγματικών αιτιών που προκαλούν τις μεταβολές στις τιμές.

3.2 Προηγούμενες Εργασίες (Related Work)

- **Zhang et al. – Fuel Price Forecasting Using Machine Learning:** Στην εργασία των Zhang et al. [11] εξετάζεται η χρήση τεχνικών μηχανικής μάθησης για την πρόβλεψη των τιμών καυσίμων με βάση ιστορικά δεδομένα. Οι ερευνητές χρησιμοποίησαν διαφορετικούς αλγόριθμους παλινδρόμησης προκειμένου να αναλύσουν τις μεταβολές των τιμών μέσα στον χρόνο και να εντοπίσουν πρότυπα που μπορούν να χρησιμοποιηθούν για την πρόβλεψη μελλοντικών τιμών. Στο πλαίσιο της έρευνας αναλύθηκαν δεδομένα που αφορούν τις τιμές καυσίμων σε διαφορετικές χρονικές περιόδους και εφαρμόστηκαν διάφορες μέθοδοι πρόβλεψης για τη σύγκριση της απόδοσής τους. Τα αποτελέσματα της μελέτης έδειξαν ότι οι τεχνικές μηχανικής μάθησης μπορούν να προσφέρουν σημαντικά βελτιωμένες προβλέψεις σε σχέση με παραδοσιακές στατιστικές μεθόδους. Η συγκεκριμένη εργασία αναδεικνύει τη σημασία της χρήσης υπολογιστικών μοντέλων για την κατανόηση της συμπεριφοράς των τιμών καυσίμων και τη δημιουργία συστημάτων πρόβλεψης που μπορούν να χρησιμοποιηθούν σε οικονομικές εφαρμογές.
- **Deb et al. – Forecasting Energy Prices Using Machine Learning :** Η εργασία των Deb et al. [10] επικεντρώνεται στην πρόβλεψη των τιμών ενέργειας μέσω της χρήσης τεχνικών μηχανικής μάθησης. Οι ερευνητές ανέλυσαν δεδομένα από αγορές ενέργειας και χρησιμοποίησαν αλγόριθμους όπως Random Forest και Support Vector Machines για τη δημιουργία μοντέλων πρόβλεψης. Στην έρευνα αυτή πραγματοποιήθηκε σύγκριση μεταξύ

διαφορετικών αλγορίθμων με στόχο τον εντοπισμό των μοντέλων που παρουσιάζουν την καλύτερη απόδοση στην πρόβλεψη τιμών. Τα αποτελέσματα έδειξαν ότι τα μοντέλα μηχανικής μάθησης μπορούν να εντοπίσουν πολύπλοκες σχέσεις μεταξύ των δεδομένων και να παρέχουν πιο αξιόπιστες προβλέψεις σε σχέση με απλούστερες μεθόδους ανάλυσης. Η μελέτη αυτή επιβεβαιώνει ότι η χρήση τεχνικών μηχανικής μάθησης αποτελεί μια αποτελεσματική προσέγγιση για την ανάλυση οικονομικών δεδομένων που παρουσιάζουν έντονες διακυμάνσεις.

- **Lago et al. (2021) – Forecasting Energy Prices:** Η εργασία των Lago et al. [11] αποτελεί μια εκτενή μελέτη σχετικά με τις μεθόδους πρόβλεψης τιμών ενέργειας. Στην έρευνα αυτή εξετάζονται διάφορες τεχνικές ανάλυσης χρονοσειρών και μοντέλα μηχανικής μάθησης που χρησιμοποιούνται για την πρόβλεψη τιμών στις αγορές ενέργειας. Οι ερευνητές αξιολόγησαν την απόδοση διαφορετικών μοντέλων χρησιμοποιώντας πραγματικά δεδομένα αγορών και εξέτασαν τα πλεονεκτήματα και τα μειονεκτήματα κάθε μεθόδου. Τα αποτελέσματα της μελέτης έδειξαν ότι η χρήση συνδυασμού διαφορετικών μοντέλων μπορεί να οδηγήσει σε καλύτερα αποτελέσματα πρόβλεψης, καθώς αξιοποιεί τα πλεονεκτήματα κάθε προσέγγισης. Η εργασία αυτή παρέχει μια ολοκληρωμένη επισκόπηση των σύγχρονων τεχνικών πρόβλεψης τιμών ενέργειας και αποτελεί σημαντική αναφορά για την ανάπτυξη συστημάτων πρόβλεψης στον ενεργειακό τομέα.
- **Breiman (2001) – Random Forest:** Η εργασία του Leo Breiman το 2001 [12] εισήγαγε τον αλγόριθμο Random Forest, ο οποίος αποτελεί έναν από τους πιο γνωστούς και ευρέως χρησιμοποιούμενους αλγορίθμους μηχανικής μάθησης. Ο αλγόριθμος αυτός βασίζεται στη δημιουργία πολλαπλών δέντρων αποφάσεων, τα οποία εκπαιδεύονται σε διαφορετικά υποσύνολα δεδομένων. Η τελική πρόβλεψη προκύπτει από τον συνδυασμό των προβλέψεων όλων των δέντρων, γεγονός που επιτρέπει τη μείωση των σφαλμάτων και τη βελτίωση της ακρίβειας του μοντέλου. Το Random Forest έχει χρησιμοποιηθεί σε πολλές εφαρμογές ανάλυσης δεδομένων, καθώς είναι ιδιαίτερα ανθεκτικό σε θόρυβο και μπορεί να μοντελοποιήσει πολύπλοκες σχέσεις μεταξύ των μεταβλητών. Η συγκεκριμένη μέθοδος χρησιμοποιείται συχνά σε προβλήματα πρόβλεψης οικονομικών δεδομένων και αποτελεί σημαντικό εργαλείο για την ανάλυση μεγάλων συνόλων δεδομένων.
- **Friedman (2001) – Gradient Boosting:** Η εργασία του Jerome Friedman το 2001 [16] παρουσίασε την τεχνική Gradient Boosting, μια μέθοδο που βασίζεται στη δημιουργία διαδοχικών μοντέλων πρόβλεψης με στόχο τη σταδιακή βελτίωση της απόδοσης του συστήματος. Στη μέθοδο αυτή, κάθε νέο μοντέλο εκπαιδεύεται με στόχο τη διόρθωση των σφαλμάτων που έχουν προκύψει από τα προηγούμενα μοντέλα. Με τον τρόπο αυτό δημιουργείται ένα ισχυρό μοντέλο πρόβλεψης που μπορεί να επιτύχει υψηλή ακρίβεια. Το Gradient Boosting χρησιμοποιείται ευρέως σε εφαρμογές ανάλυσης δεδομένων και πρόβλεψης οικονομικών τιμών, καθώς μπορεί να χειριστεί πολύπλοκες σχέσεις μεταξύ των μεταβλητών και να παράγει αξιόπιστα αποτελέσματα.

3.3 Δυσκολίες στην πρόβλεψη οικονομικών δεδομένων

Η πρόβλεψη οικονομικών δεδομένων αποτελεί μια ιδιαίτερα απαιτητική διαδικασία, καθώς τα δεδομένα αυτά χαρακτηρίζονται από υψηλό βαθμό αβεβαιότητας και

μεταβλητότητας. Οι τιμές των καυσίμων επηρεάζονται από πολλούς παράγοντες, οι οποίοι μεταβάλλονται συνεχώς και συχνά με μη προβλέψιμο τρόπο.

Ένα από τα βασικά προβλήματα είναι η μη γραμμικότητα των δεδομένων, καθώς οι σχέσεις μεταξύ των μεταβλητών δεν είναι πάντα απλές ή σταθερές. Επιπλέον, τα οικονομικά δεδομένα συχνά περιέχουν θόρυβο, ελλιπείς τιμές και ακραίες μεταβολές, γεγονός που δυσκολεύει την εκπαίδευση αξιόπιστων μοντέλων.

Ακόμη, οι ξαφνικές αλλαγές στην αγορά, όπως οικονομικές κρίσεις ή γεωπολιτικά γεγονότα, μπορούν να επηρεάσουν σημαντικά τις τιμές, καθιστώντας δύσκολη τη δημιουργία μοντέλων που να γενικεύουν σωστά σε νέα δεδομένα.

Για τους λόγους αυτούς, η πρόβλεψη οικονομικών δεδομένων απαιτεί τη χρήση προηγμένων τεχνικών ανάλυσης και μοντέλων που μπορούν να προσαρμόζονται σε δυναμικά περιβάλλοντα .[13]

3.4 Εξωτερικοί παράγοντες που επηρεάζουν τις οικονομικές αγορές

Εκτός από τους οικονομικούς παράγοντες που επηρεάζουν τις τιμές των προϊόντων, σημαντικό ρόλο παίζουν και εξωτερικά γεγονότα που συμβαίνουν σε διεθνές επίπεδο. Γεωπολιτικές συγκρούσεις, οικονομικές κρίσεις, πανδημίες και αλλαγές στην ενεργειακή πολιτική μπορούν να προκαλέσουν σημαντικές μεταβολές στις αγορές.

Ένα χαρακτηριστικό παράδειγμα αποτελεί ο πόλεμος στην Ουκρανία, ο οποίος επηρέασε σημαντικά τις διεθνείς αγορές ενέργειας. Η σύγκρουση αυτή προκάλεσε περιορισμούς στην προμήθεια φυσικού αερίου και πετρελαίου, οδηγώντας σε αύξηση των τιμών ενέργειας σε πολλές χώρες της Ευρώπης. Παρόμοια γεγονότα μπορούν να δημιουργήσουν αβεβαιότητα στις αγορές και να προκαλέσουν έντονες διακυμάνσεις στις τιμές.

Επιπλέον, οικονομικές κρίσεις ή παγκόσμιες υγειονομικές κρίσεις, όπως η πανδημία COVID-19, μπορούν να επηρεάσουν σημαντικά την οικονομική δραστηριότητα και τη ζήτηση για ενέργεια. Οι αλλαγές αυτές μπορεί να οδηγήσουν σε μεγάλες αυξήσεις ή μειώσεις στις τιμές των καυσίμων.

Οι παράγοντες αυτοί καθιστούν απαραίτητη τη χρήση σύγχρονων μεθόδων ανάλυσης δεδομένων, οι οποίες μπορούν να βοηθήσουν στην κατανόηση των τάσεων της αγοράς και στην εκτίμηση πιθανών μελλοντικών εξελίξεων.

3.5 Σύγχρονες τεχνολογίες και μέθοδοι ανάλυσης οικονομικών δεδομένων

Τα τελευταία χρόνια έχουν αναπτυχθεί πολλές τεχνολογίες και υπολογιστικές μέθοδοι για την ανάλυση οικονομικών δεδομένων. Η αυξημένη διαθεσιμότητα δεδομένων και η ανάγκη για καλύτερη κατανόηση των οικονομικών τάσεων έχουν οδηγήσει στην ανάπτυξη νέων εργαλείων που επιτρέπουν την επεξεργασία και την ανάλυση μεγάλων συνόλων δεδομένων.

Οι μέθοδοι αυτές περιλαμβάνουν τόσο παραδοσιακές στατιστικές τεχνικές όσο και σύγχρονες προσεγγίσεις ανάλυσης δεδομένων. Οι στατιστικές μέθοδοι χρησιμοποιούνται

συχνά για την κατανόηση των βασικών χαρακτηριστικών των δεδομένων, όπως οι τάσεις, οι μεταβολές και οι σχέσεις μεταξύ διαφορετικών οικονομικών μεταβλητών.

Παράλληλα, σημαντικό ρόλο στην ανάλυση οικονομικών δεδομένων παίζουν και τα μοντέλα παλινδρόμησης, τα οποία επιτρέπουν την εκτίμηση της σχέσης μεταξύ διαφορετικών μεταβλητών και την πρόβλεψη μελλοντικών τιμών. Στον τομέα της πρόβλεψης τιμών καυσίμων χρησιμοποιούνται συχνά τέτοια μοντέλα για την ανάλυση ιστορικών δεδομένων και την εκτίμηση της μελλοντικής εξέλιξης των τιμών.

Επιπλέον, χρησιμοποιούνται και αλγόριθμοι που βασίζονται σε δέντρα αποφάσεων, όπως τα Random Forest και Gradient Boosting, οι οποίοι μπορούν να αναγνωρίσουν πιο σύνθετες σχέσεις μεταξύ των δεδομένων. Οι αλγόριθμοι αυτοί έχουν χρησιμοποιηθεί ευρέως σε προβλήματα πρόβλεψης οικονομικών δεδομένων, καθώς μπορούν να προσφέρουν υψηλή ακρίβεια.

Σε πολλές περιπτώσεις εφαρμόζονται επίσης συνδυασμοί διαφορετικών μοντέλων πρόβλεψης, με στόχο τη βελτίωση της ακρίβειας των αποτελεσμάτων. Οι προσεγγίσεις αυτές επιτρέπουν τη δημιουργία πιο αξιόπιστων συστημάτων ανάλυσης και πρόβλεψης οικονομικών δεδομένων. [14]

ΚΕΦΑΛΑΙΟ 4 Προτεινόμενο Σύστημα

4.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζεται αναλυτικά το προτεινόμενο σύστημα που αναπτύχθηκε στο πλαίσιο της παρούσας εργασίας. Το σύστημα έχει ως στόχο τη συλλογή δεδομένων τιμών καυσίμων από διαδικτυακές πηγές, την αποθήκευση και επεξεργασία των δεδομένων αυτών και την ανάπτυξη μοντέλων πρόβλεψης για την εκτίμηση της μελλοντικής εξέλιξης των τιμών.

Η λειτουργία του συστήματος βασίζεται σε μια αυτοματοποιημένη διαδικασία που περιλαμβάνει τη συλλογή δεδομένων, τη δημιουργία ενός ιστορικού συνόλου δεδομένων και την εφαρμογή αλγορίθμων ανάλυσης δεδομένων για την πρόβλεψη μελλοντικών τιμών. Μέσω της διαδικασίας αυτής δημιουργείται ένα σύστημα που επιτρέπει την ανάλυση της εξέλιξης των τιμών καυσίμων σε διαφορετικές χώρες και χρονικές περιόδους.

4.1.1 Στόχος του συστήματος

Ο βασικός στόχος του προτεινόμενου συστήματος είναι η ανάπτυξη ενός εργαλείου που θα μπορεί να συλλέγει και να αναλύει δεδομένα τιμών καυσίμων με σκοπό την πρόβλεψη της μελλοντικής τους εξέλιξης. Η ανάλυση των δεδομένων αυτών μπορεί να συμβάλει στην κατανόηση των τάσεων της αγοράς καυσίμων και στη λήψη οικονομικών αποφάσεων.

Επιπλέον, το σύστημα επιδιώκει να δημιουργήσει ένα οργανωμένο ιστορικό σύνολο δεδομένων το οποίο μπορεί να χρησιμοποιηθεί για την εκπαίδευση υπολογιστικών μοντέλων και την ανάπτυξη μεθόδων πρόβλεψης.

4.1.2 Γενική λειτουργία του συστήματος

Η λειτουργία του συστήματος βασίζεται σε μια σειρά από στάδια που εκτελούνται διαδοχικά. Αρχικά πραγματοποιείται η συλλογή δεδομένων από διαδικτυακές πηγές. Στη συνέχεια τα δεδομένα αποθηκεύονται σε αρχείο ώστε να δημιουργηθεί ένα ιστορικό σύνολο δεδομένων.

Μετά την αποθήκευση των δεδομένων ακολουθεί η επεξεργασία και η προετοιμασία τους για την εφαρμογή μοντέλων ανάλυσης δεδομένων. Στο τελικό στάδιο εφαρμόζονται αλγόριθμοι πρόβλεψης που χρησιμοποιούνται για την εκτίμηση της μελλοντικής εξέλιξης των τιμών καυσίμων.

4.2 Συλλογή δεδομένων

4.2.1 Τα Δεδομένα και η πηγή τους

Οι πληροφορίες για τις τιμές καυσίμων συλλέγονται από διαδικτυακές πηγές που δημοσιεύουν δεδομένα σχετικά με τις τιμές ενέργειας σε διάφορες χώρες. Στην παρούσα εργασία χρησιμοποιείται ως βασική πηγή δεδομένων η ιστοσελίδα **GlobalPetrolPrices** <https://www.globalpetrolprices.com/>, η οποία παρέχει ενημερωμένες πληροφορίες σχετικά με

τις τιμές καυσίμων, ηλεκτρικής ενέργειας και φυσικού αερίου σε περισσότερες από 150 χώρες παγκοσμίως.

Η ιστοσελίδα αυτή συγκεντρώνει δεδομένα από διάφορες πηγές, όπως κυβερνητικούς οργανισμούς, ενεργειακές εταιρείες και επίσημες αναφορές αγοράς, τα οποία ελέγχονται και επαληθεύονται πριν δημοσιευθούν. Οι τιμές που παρέχονται αντιστοιχούν συνήθως στις λιανικές τιμές καυσίμων (retail prices) ανά λίτρο και περιλαμβάνουν φόρους και άλλες χρεώσεις.

Από τις σελίδες αυτές εξάγονται πληροφορίες όπως:

- η τιμή καυσίμου ανά λίτρο
- η ημερομηνία τελευταίας ενημέρωσης των δεδομένων
- η χώρα στην οποία αναφέρεται η τιμή
- ο τύπος καυσίμου

Στο πλαίσιο της παρούσας εργασίας συλλέγονται δεδομένα για τρεις βασικούς τύπους καυσίμων:

- **βενζίνη (Gasoline)**
- **πετρέλαιο κίνησης (Diesel)**
- **υγραέριο (LPG)**

Τα δεδομένα συλλέγονται για διαφορετικές χώρες της Ευρώπης, όπως η **Ελλάδα**, η **Γερμανία** και η **Γαλλία**, ώστε να δημιουργηθεί ένα σύνολο δεδομένων που επιτρέπει τη σύγκριση των τιμών μεταξύ διαφορετικών αγορών.

Τα δεδομένα που συλλέγονται αποτελούν τη βάση για τη δημιουργία του συνόλου δεδομένων που χρησιμοποιείται στο σύστημα για την ανάλυση και την πρόβλεψη των τιμών καυσίμων.

4.2.2 Διαδικασία εξαγωγής δεδομένων

Η διαδικασία εξαγωγής δεδομένων του συστήματος βασίζεται στην τεχνική του web scraping, δηλαδή στην αυτοματοποιημένη ανάκτηση και επεξεργασία πληροφοριών από ιστοσελίδες. Στο πλαίσιο της παρούσας εργασίας, η τεχνική αυτή χρησιμοποιείται για τη συλλογή τιμών καυσίμων από την ιστοσελίδα GlobalPetrolPrices, με στόχο τη δημιουργία ενός οργανωμένου συνόλου δεδομένων για περαιτέρω ανάλυση και πρόβλεψη.

Για την υλοποίηση της διαδικασίας χρησιμοποιείται αρχικά η βιβλιοθήκη requests, η οποία επιτρέπει στο πρόγραμμα να αποστέλλει αιτήματα HTTP προς την ιστοσελίδα και να λαμβάνει το περιεχόμενό της. Μέσω της βιβλιοθήκης αυτής, το σύστημα συνδέεται στη σελίδα που αντιστοιχεί σε κάθε χώρα και σε κάθε τύπο καυσίμου, ανακτώντας τον HTML κώδικα της σελίδας. Η χρήση της requests προσφέρει έναν απλό και αξιόπιστο τρόπο επικοινωνίας με τον απομακρυσμένο εξυπηρετητή και αποτελεί βασικό στάδιο της διαδικασίας συλλογής των δεδομένων.

Στη συνέχεια, για την ανάλυση του περιεχομένου της σελίδας χρησιμοποιείται η βιβλιοθήκη BeautifulSoup. Η βιβλιοθήκη αυτή επιτρέπει την επεξεργασία του HTML κώδικα και τη μετατροπή του σε μια δομή που μπορεί να αναλυθεί ευκολότερα από το πρόγραμμα. Με τον τρόπο αυτό, το σύστημα μπορεί να εντοπίσει τα σημεία της σελίδας στα οποία εμφανίζονται οι πληροφορίες που ενδιαφέρουν, όπως η τιμή του καυσίμου και η ημερομηνία τελευταίας ενημέρωσης.

Μετά την ανάκτηση και ανάλυση του περιεχομένου, το σύστημα εφαρμόζει ελέγχους πάνω στο κείμενο της ιστοσελίδας ώστε να εξάγει τις επιθυμητές πληροφορίες. Πιο συγκεκριμένα, αναζητούνται στοιχεία όπως η τρέχουσα τιμή ανά λίτρο και η τελευταία ημερομηνία ενημέρωσης. Τα δεδομένα αυτά οργανώνονται σε δομημένη μορφή και επιστρέφονται από το πρόγραμμα ώστε να αποθηκευτούν στη συνέχεια στο αρχείο ιστορικού. Η διαδικασία αυτή επαναλαμβάνεται για κάθε χώρα και για κάθε τύπο καυσίμου που περιλαμβάνει το σύστημα.

Ένα σημαντικό χαρακτηριστικό της υλοποίησης είναι ότι η διαδικασία εξαγωγής δεν περιορίζεται σε μία μόνο τιμή, αλλά εφαρμόζεται συστηματικά σε πολλαπλές περιπτώσεις. Στην παρούσα εργασία συλλέγονται δεδομένα για βενζίνη, πετρέλαιο κίνησης και υγραέριο, ενώ η διαδικασία εκτελείται για χώρες όπως η Ελλάδα, η Γερμανία και η Γαλλία. Με αυτόν τον τρόπο δημιουργείται ένα πιο ολοκληρωμένο σύνολο δεδομένων, το οποίο επιτρέπει τη μελέτη και τη σύγκριση των τιμών μεταξύ διαφορετικών χωρών και διαφορετικών καυσίμων.

Η χρήση του web scraping προσφέρει σημαντικά πλεονεκτήματα, καθώς επιτρέπει την αυτοματοποίηση της διαδικασίας συλλογής δεδομένων, μειώνει την ανάγκη χειροκίνητης καταχώρισης και διευκολύνει τη συνεχή ενημέρωση του συνόλου δεδομένων. Παράλληλα, η χρήση των βιβλιοθηκών requests και BeautifulSoup καθιστά την υλοποίηση ευέλικτη και επεκτάσιμη, επιτρέποντας τη μελλοντική προσαρμογή του συστήματος και σε άλλες πηγές δεδομένων, εφόσον αυτό κριθεί απαραίτητο.

```
def fetch_country_fuel(country: str, fuel_path: str) -> dict:

    url = f"https://www.globalpetrolprices.com/{country}/{fuel_path}/"
    r = requests.get(url, headers=HEADERS, timeout=30)
    r.raise_for_status()

    soup = BeautifulSoup(r.text, "html.parser")
    text = soup.get_text("\n", strip=True)

    # "Last update 2025-12-15"
    last_update = None
    m = re.search(r"Last update\s+(\d{4}-\d{2}-\d{2})", text)
    if m:
        last_update = m.group(1)

    # "Current price 1.75 -"
    eur = None
    m = re.search(r"Current price\s+([\d-]+(?:\.\d+)?)\s+-", text)
    if m:
        eur = float(m.group(1))

    if eur is None:
        raise RuntimeError(f"Could not parse EUR/L price from page: {url}")

    return {
        "url": url,
        "last_update": last_update,
        "eur_per_liter": eur,
    }
```

Figure 9 Απόσπασμα κώδικα Python για την εξαγωγή τιμών καυσίμων μέσω της πλατφόρμα Global Petrol Prices

4.2.3 Αποθήκευση των εξαγόμενων δεδομένων

Μετά την εξαγωγή των δεδομένων από τις διαδικτυακές πηγές, το επόμενο στάδιο του συστήματος αφορά την αποθήκευσή τους σε οργανωμένη μορφή. Στο προτεινόμενο σύστημα τα δεδομένα αποθηκεύονται σε αρχείο μορφής **CSV (Comma-Separated Values)** με όνομα **fuel_prices.csv**, το οποίο λειτουργεί ως το βασικό σύνολο δεδομένων της εφαρμογής.

Το αρχείο αυτό περιλαμβάνει πληροφορίες σχετικά με τις τιμές καυσίμων που συλλέγονται από τις διαδικτυακές πηγές. Για κάθε εγγραφή καταγράφονται στοιχεία όπως:

- η ημερομηνία συλλογής των δεδομένων (fetch_date)
- η χώρα (country)
- ο τύπος καυσίμου (fuel)
- η τιμή καυσίμου ανά λίτρο (price_eur_per_liter)
- η ημερομηνία τελευταίας ενημέρωσης της τιμής (last_update)
- η πηγή των δεδομένων (source_url)

Η αποθήκευση των δεδομένων πραγματοποιείται αυτόματα κάθε φορά που το σύστημα συλλέγει νέες τιμές από την ιστοσελίδα. Με τον τρόπο αυτό δημιουργείται ένα **ιστορικό**

Κατά την επεξεργασία των δεδομένων παρατηρείται ότι σε ορισμένες περιπτώσεις υπάρχουν ελλείψεις τιμών για συγκεκριμένες ημερομηνίες. Αυτό συμβαίνει επειδή οι τιμές δεν ενημερώνονται καθημερινά ή ενδέχεται να υπάρχουν καθυστερήσεις στη δημοσίευσή τους.

Για την αντιμετώπιση του προβλήματος αυτού και για τη δημιουργία μιας πλήρους χρονοσειράς, εφαρμόζεται διαδικασία συμπλήρωσης των ελλিপών τιμών. Συγκεκριμένα, σε περιπτώσεις όπου δεν υπάρχουν διαθέσιμα δεδομένα για κάποια ημερομηνία, χρησιμοποιούνται τεχνικές παρεμβολής (interpolation) ή διατήρησης της προηγούμενης τιμής (forward fill).

4.3.2 Δημιουργία στατιστικών δεδομένων

Μετά την οργάνωση και επεξεργασία των δεδομένων, το σύστημα δημιουργεί επιπλέον στατιστικές πληροφορίες που βοηθούν στην καλύτερη κατανόηση της εξέλιξης των τιμών καυσίμων. Η διαδικασία αυτή πραγματοποιείται κατά την επεξεργασία του αρχείου **fuel_prices.csv**, στο οποίο έχουν αποθηκευτεί οι ημερήσιες τιμές καυσίμων.

Κατά τη διαδικασία αυτή, το πρόγραμμα διαβάζει τα δεδομένα από το αρχείο CSV και τα ομαδοποιεί με βάση τον μήνα και το έτος. Στη συνέχεια υπολογίζονται στατιστικές τιμές για κάθε χρονική περίοδο.

Πιο συγκεκριμένα, για κάθε μήνα υπολογίζονται:

- η μέση τιμή καυσίμου
- η μέγιστη τιμή που καταγράφηκε
- η ελάχιστη τιμή του μήνα
- η πρώτη και η τελευταία τιμή που εμφανίζεται μέσα στον μήνα

Η διαδικασία αυτή πραγματοποιείται στο στάδιο της επεξεργασίας των δεδομένων πριν την εφαρμογή των μοντέλων πρόβλεψης. Με τον τρόπο αυτό δημιουργούνται επιπλέον πληροφορίες που βοηθούν στην ανάλυση των τάσεων των τιμών καυσίμων και μπορούν να χρησιμοποιηθούν για την καλύτερη κατανόηση της συμπεριφοράς των δεδομένων.

Οι παραπάνω υπολογισμοί μπορούν να εκφραστούν μαθηματικά ως εξής:

- Μέση Τιμή = $(1 / \text{Πλήθος Ημερήσιων Καταγραφών}) \times \Sigma (\text{Ημερήσια Τιμή Καυσίμου})$
- Μέγιστη τιμή: $p_{\max} = \max(p_i)$
- Ελάχιστη τιμή: $p_{\min} = \min(p_i)$

όπου $p_{\max}$ είναι η μεγαλύτερη τιμή καυσίμου που καταγράφηκε μέσα στον μήνα και $p_{\min}$ η μικρότερη. Τα στατιστικά αυτά μεγέθη συμβάλλουν στην καλύτερη κατανόηση της συμπεριφοράς των δεδομένων και αποτελούν χρήσιμη βάση για τα επόμενα στάδια της ανάλυσης και της πρόβλεψης των τιμών καυσίμων.

4.4 Δημιουργία χαρακτηριστικών για τα μοντέλα πρόβλεψης

Μετά την επεξεργασία και οργάνωση των δεδομένων, το επόμενο στάδιο του συστήματος αφορά τη δημιουργία των χαρακτηριστικών (features) που θα χρησιμοποιηθούν για την εκπαίδευση των μοντέλων πρόβλεψης. Τα χαρακτηριστικά αποτελούν τις μεταβλητές εισόδου των αλγορίθμων μηχανικής μάθησης και παίζουν σημαντικό ρόλο στην ακρίβεια των προβλέψεων.

Στο πλαίσιο της παρούσας εργασίας, τα χαρακτηριστικά δημιουργούνται με βάση τις ιστορικές τιμές καυσίμων που έχουν αποθηκευτεί στο αρχείο **fuel_prices.csv**. Η βασική ιδέα είναι ότι οι προηγούμενες τιμές των καυσίμων μπορούν να χρησιμοποιηθούν για την πρόβλεψη των μελλοντικών τιμών.

4.4.1 Χρήση προηγούμενων τιμών (lags)

Ένα από τα βασικά χαρακτηριστικά που χρησιμοποιούνται είναι οι προηγούμενες τιμές των καυσίμων. Για κάθε χρονική στιγμή δημιουργούνται νέες μεταβλητές που αντιστοιχούν στις τιμές των προηγούμενων ημερών. Για παράδειγμα, για να προβλεφθεί η τιμή μιας συγκεκριμένης ημέρας, χρησιμοποιούνται ως είσοδος οι τιμές των προηγούμενων ημερών.

Η διαδικασία αυτή είναι γνωστή ως δημιουργία lag features και χρησιμοποιείται ευρέως στην ανάλυση χρονοσειρών. Τα χαρακτηριστικά αυτά επιτρέπουν στα μοντέλα να κατανοήσουν την εξέλιξη των τιμών μέσα στον χρόνο και να εντοπίσουν πιθανές σχέσεις μεταξύ των διαδοχικών τιμών.

Η χρήση των προηγούμενων τιμών μπορεί να εκφραστεί μαθηματικά ως:

- $X_t = [p_{t-1}, p_{t-2}, \dots, p_{t-7}]$

όπου X_t είναι το σύνολο των χαρακτηριστικών εισόδου που χρησιμοποιούνται για την πρόβλεψη της τιμής τη χρονική στιγμή t , p_{t-1} είναι η τιμή καυσίμου μία ημέρα πριν από τη χρονική στιγμή t , p_{t-2} η τιμή δύο ημέρες πριν, ενώ p_{t-7} η τιμή επτά ημέρες πριν. Στην παρούσα εργασία χρησιμοποιούνται οι τιμές των επτά προηγούμενων ημερών ως χαρακτηριστικά εισόδου για την εκπαίδευση των μοντέλων μηχανικής μάθησης. Η προσέγγιση αυτή επιτρέπει στα μοντέλα να αξιοποιούν τη χρονική πληροφορία που περιέχεται στα δεδομένα και να παράγουν πιο αξιόπιστες προβλέψεις για τις μελλοντικές τιμές καυσίμων.

4.4.2 Υπολογισμός κινητού μέσου όρου

Εκτός από τις προηγούμενες τιμές, δημιουργούνται και επιπλέον χαρακτηριστικά που βοηθούν τα μοντέλα να κατανοήσουν καλύτερα τη γενική πορεία των τιμών καυσίμων. Ένα από τα σημαντικότερα χαρακτηριστικά είναι ο κινητός μέσος όρος (Moving Average), ο οποίος υπολογίζεται από τον μέσο όρο των τιμών των προηγούμενων ημερών.

Ο κινητός μέσος όρος υπολογίζεται ως:

- $MA_t = (p_{t-1} + p_{t-2} + \dots + p_{t-n}) / n$

όπου MA_t είναι ο κινητός μέσος όρος τη χρονική στιγμή t , $p_{t-1}, p_{t-2}, \dots, p_{t-n}$ είναι οι τιμές καυσίμου των προηγούμενων ημερών και n είναι το πλήθος των ημερών που χρησιμοποιούνται στον υπολογισμό. Με τον τρόπο αυτό προκύπτει μια πιο ομαλή εικόνα της εξέλιξης των τιμών μέσα στον χρόνο.

Η χρήση του κινητού μέσου όρου μειώνει την επίδραση των μικρών καθημερινών διακυμάνσεων και βοηθά στην ανάδειξη της γενικότερης τάσης των δεδομένων. Έτσι, τα μοντέλα μηχανικής μάθησης μπορούν να αξιοποιήσουν πιο αποτελεσματικά τις πληροφορίες που περιέχονται στη χρονοσειρά και να παράγουν πιο αξιόπιστες προβλέψεις.

4.4.3 Δημιουργία του τελικού συνόλου δεδομένων

Αφού δημιουργηθούν τα χαρακτηριστικά, τα δεδομένα οργανώνονται σε ένα τελικό σύνολο δεδομένων το οποίο χρησιμοποιείται για την εκπαίδευση των μοντέλων πρόβλεψης. Στο σύνολο αυτό περιλαμβάνονται τόσο τα χαρακτηριστικά εισόδου όσο και η μεταβλητή στόχος, δηλαδή η τιμή καυσίμου που επιδιώκεται να προβλεφθεί.

Στη συνέχεια το σύνολο δεδομένων χωρίζεται σε δύο βασικά τμήματα: το σύνολο εκπαίδευσης (training set) και το σύνολο ελέγχου (test set). Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση των μοντέλων μηχανικής μάθησης, ενώ το σύνολο ελέγχου χρησιμοποιείται για την αξιολόγηση της απόδοσης των μοντέλων.

Η δημιουργία ενός κατάλληλα οργανωμένου συνόλου δεδομένων αποτελεί βασικό βήμα για την ανάπτυξη αξιόπιστων μοντέλων πρόβλεψης και επηρεάζει σημαντικά την ποιότητα των αποτελεσμάτων που παράγει το σύστημα.

4.5 Εκπαίδευση και εφαρμογή των μοντέλων πρόβλεψης

Μετά τη δημιουργία των χαρακτηριστικών και την προετοιμασία του συνόλου δεδομένων, το επόμενο στάδιο του συστήματος αφορά την εκπαίδευση των μοντέλων μηχανικής μάθησης που χρησιμοποιούνται για την πρόβλεψη των μελλοντικών τιμών καυσίμων. Στο στάδιο αυτό χρησιμοποιούνται τα ιστορικά δεδομένα που έχουν αποθηκευτεί στο αρχείο **fuel_prices.csv**, τα οποία περιλαμβάνουν πληροφορίες σχετικά με την τιμή του καυσίμου, τη χώρα και τη χρονική στιγμή καταγραφής της τιμής.

Η διαδικασία εκπαίδευσης των μοντέλων βασίζεται στη χρήση των χαρακτηριστικών που δημιουργήθηκαν στο προηγούμενο στάδιο. Τα χαρακτηριστικά αυτά περιλαμβάνουν προηγούμενες τιμές καυσίμων καθώς και στατιστικές πληροφορίες που περιγράφουν τη συμπεριφορά των δεδομένων μέσα στον χρόνο.

Για την εκπαίδευση των μοντέλων το σύνολο δεδομένων χωρίζεται σε δύο βασικά μέρη:

- **σύνολο εκπαίδευσης (training set)**
- **σύνολο ελέγχου (test set)**

Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση των αλγορίθμων μηχανικής μάθησης, ενώ το σύνολο ελέγχου χρησιμοποιείται για την αξιολόγηση της απόδοσης των μοντέλων σε δεδομένα που δεν έχουν χρησιμοποιηθεί κατά την εκπαίδευση.

Στο προτεινόμενο σύστημα εφαρμόζεται ένα σύνολο από **δώδεκα (12) διαφορετικά μοντέλα μηχανικής μάθησης**, με στόχο τη σύγκριση της απόδοσής τους στο πρόβλημα της πρόβλεψης τιμών καυσίμων. Τα μοντέλα που δοκιμάζονται είναι τα εξής:

- **Linear Regression**
- **Ridge Regression**
- **Lasso Regression**
- **Elastic Net Regression**
- **Decision Tree Regressor**
- **Random Forest Regressor**
- **Extra Trees Regressor**
- **Gradient Boosting Regressor**
- **AdaBoost Regressor**
- **Bagging Regressor**
- **Support Vector Regression (SVR)**
- **K-Nearest Neighbors (KNN) Regressor**

Όλα τα μοντέλα εκπαιδεύονται χρησιμοποιώντας τα ίδια δεδομένα ώστε να είναι δυνατή η σύγκριση της απόδοσής τους. Μετά την εκπαίδευση, τα μοντέλα χρησιμοποιούνται για την παραγωγή προβλέψεων σχετικά με τις μελλοντικές τιμές καυσίμων.

Από τα παραπάνω μοντέλα, ιδιαίτερη σημασία δίνεται στα **Ridge Regression, Random Forest, Gradient Boosting, Support Vector Regression (SVR)** και **K-Nearest Neighbors (KNN)**, καθώς τα συγκεκριμένα μοντέλα παρουσιάζουν καλύτερη απόδοση στο συγκεκριμένο πρόβλημα πρόβλεψης.

Η απόδοση των μοντέλων αξιολογείται με τη χρήση της μετρικής Mean Absolute Error (MAE), η οποία υπολογίζει τη μέση απόλυτη διαφορά μεταξύ των πραγματικών και των προβλεπόμενων τιμών. Η μετρική αυτή χρησιμοποιείται ευρέως σε προβλήματα παλινδρόμησης, καθώς παρέχει μια εύκολα κατανοητή εκτίμηση του σφάλματος πρόβλεψης. Ο υπολογισμός του MAE δίνεται από τη σχέση:

$$\bullet \quad \text{MAE} = (|y_1 - \hat{y}_1| + |y_2 - \hat{y}_2| + \dots + |y_n - \hat{y}_n|) / N$$

όπου $y_1, y_2, \dots, y_n$ είναι οι πραγματικές τιμές, $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$ οι αντίστοιχες τιμές που προβλέπει το μοντέλο και N το συνολικό πλήθος των παρατηρήσεων. Όσο μικρότερη είναι η τιμή του MAE, τόσο πιο κοντά βρίσκονται οι προβλέψεις στις πραγματικές τιμές.

Η μετρική αυτή επιτρέπει την αντικειμενική σύγκριση των διαφορετικών μοντέλων και βοηθά στην επιλογή των πιο αποδοτικών αλγορίθμων για το συγκεκριμένο πρόβλημα.

4.6 Συνδυασμός μοντέλων πρόβλεψης(Ensemble)

Μετά την εκπαίδευση και την αξιολόγηση των διαφορετικών μοντέλων μηχανικής μάθησης, το επόμενο στάδιο του συστήματος αφορά τον συνδυασμό των αποτελεσμάτων των μοντέλων με σκοπό τη βελτίωση της ακρίβειας των προβλέψεων. Η προσέγγιση αυτή βασίζεται στην ιδέα ότι διαφορετικά μοντέλα μπορεί να αποδίδουν καλύτερα σε διαφορετικές περιπτώσεις και ότι ο συνδυασμός των προβλέψεών τους μπορεί να οδηγήσει σε πιο αξιόπιστα αποτελέσματα.

Στο προτεινόμενο σύστημα χρησιμοποιείται μια μέθοδος **ensemble**, κατά την οποία οι προβλέψεις των καλύτερων μοντέλων συνδυάζονται για την παραγωγή μιας τελικής πρόβλεψης. Η μέθοδος αυτή επιτρέπει την αξιοποίηση των πλεονεκτημάτων πολλών διαφορετικών αλγορίθμων και συμβάλλει στη μείωση των σφαλμάτων πρόβλεψης.

Η διαδικασία λειτουργεί ως εξής: αρχικά αξιολογούνται όλα τα μοντέλα που έχουν εκπαιδευτεί και επιλέγονται εκείνα που παρουσιάζουν την καλύτερη απόδοση με βάση τα αποτελέσματα στο σύνολο ελέγχου. Στη συνέχεια οι προβλέψεις των επιλεγμένων μοντέλων συνδυάζονται ώστε να δημιουργηθεί μια τελική πρόβλεψη για την τιμή του καυσίμου.

Η τελική πρόβλεψη προκύπτει από τον **μέσο όρο των προβλέψεων των επιλεγμένων μοντέλων**, γεγονός που οδηγεί σε ένα πιο σταθερό και αξιόπιστο αποτέλεσμα. Η χρήση της τεχνικής ensemble επιτρέπει στο σύστημα να μειώνει την επίδραση πιθανών σφαλμάτων που μπορεί να εμφανίζονται σε μεμονωμένα μοντέλα.

Με τον τρόπο αυτό το προτεινόμενο σύστημα συνδυάζει τα αποτελέσματα πολλών μοντέλων μηχανικής μάθησης και δημιουργεί μια πιο αξιόπιστη εκτίμηση για την εξέλιξη των τιμών καυσίμων. Στο επόμενο κεφάλαιο παρουσιάζονται τα πειραματικά αποτελέσματα και η αξιολόγηση της απόδοσης των μοντέλων που χρησιμοποιήθηκαν.

ΚΕΦΑΛΑΙΟ 5 Αποτελέσματα και Αξιολόγηση Μοντέλων

5.1 Εισαγωγή και διαδικασία αξιολόγησης

Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα της εφαρμογής των αλγορίθμων μηχανικής μάθησης για την πρόβλεψη των τιμών καυσίμων. Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από διαδικτυακές πηγές και έχουν αποθηκευτεί στο αρχείο `fuel_prices.csv`. Για κάθε χώρα (Ελλάδα, Γερμανία και Γαλλία) και για κάθε τύπο καυσίμου (βενζίνη, diesel και LPG), το σύστημα επεξεργάζεται τα ιστορικά δεδομένα και εκπαιδεύει πολλαπλά μοντέλα μηχανικής μάθησης.

Πριν από την εκπαίδευση των μοντέλων, πραγματοποιείται η διαδικασία συλλογής και αποθήκευσης των πιο πρόσφατων διαθέσιμων τιμών καυσίμων. Οι τιμές αυτές αποτελούν το σημείο εκκίνησης για την ενημέρωση του ιστορικού συνόλου δεδομένων και χρησιμοποιούνται για τη δημιουργία των χαρακτηριστικών που τροφοδοτούν τα μοντέλα πρόβλεψης. Το Σχήμα 5.1 παρουσιάζει τις τελευταίες τιμές καυσίμων που συλλέχθηκαν από το σύστημα για τις χώρες που εξετάζονται στην παρούσα εργασία.

Αρχικά, το σύστημα συλλέγει τις πιο πρόσφατες διαθέσιμες τιμές καυσίμων και ενημερώνει το ιστορικό σύνολο δεδομένων. Η παρακάτω εικόνα παρουσιάζει τις τρέχουσες τιμές καυσίμων για κάθε χώρα, καθώς και το πλήθος των διαθέσιμων δεδομένων εκπαίδευσης (*train points*) που χρησιμοποιούνται για την εκπαίδευση των μοντέλων μηχανικής μάθησης.

```
Fetch date: 2026-06-03
== Greece ==
Gasoline: EUR/L 2.06 | Last update: 2026-06-01
Diesel   : EUR/L 1.74 | Last update: 2026-06-01
LPG      : EUR/L 1.12 | Last update: 2026-06-01

== Germany ==
Gasoline: EUR/L 1.9 | Last update: 2026-06-01
Diesel   : EUR/L 1.86 | Last update: 2026-06-01
LPG      : EUR/L 1.18 | Last update: 2026-06-01

== France ==
Gasoline: EUR/L 2.02 | Last update: 2026-06-01
Diesel   : EUR/L 2.05 | Last update: 2026-06-01
LPG      : EUR/L 1.08 | Last update: 2026-06-01
```

Figure 11 Δεδομένα εισόδου και αριθμός παρατηρήσεων 1

5.2 Αποτελέσματα μοντέλων και προβλέψεις

Στο στάδιο αυτό αξιολογείται η απόδοση των μοντέλων μηχανικής μάθησης με χρήση του δείκτη **MAE (Mean Absolute Error)**, ο οποίος εκφράζει το μέσο απόλυτο σφάλμα μεταξύ των πραγματικών και των προβλεπόμενων τιμών.

Η αξιολόγηση πραγματοποιείται ξεχωριστά για κάθε χώρα και τύπο καυσίμου, ώστε να εντοπιστούν τα μοντέλα με την καλύτερη συμπεριφορά.
Για παράδειγμα:

```
Greece | Gasoline: last=2.060 on 2026-06-03 | train_points=90 | supervised_samples=83
ranked(MAE): huber:0.0223, ridge:0.0405, elastic:0.0417, lasso:0.0440, adaboost:0.1386, knn:0.1386,
kept: ['huber', 'ridge', 'elastic', 'lasso', 'adaboost']
excluded: ['rf', 'extra_trees', 'gboost', 'hgb', 'knn', 'svr_rbf', 'gpr_rbf']
-- per-model next7 --
```

Κατάταξη μοντέλων με βάση το MAE της Ελλάδας για την βενζίνη

Σύμφωνα με τα αποτελέσματα που παρουσιάζονται στην εικόνα, για την περίπτωση της Ελλάδας και της βενζίνης, το μοντέλο **Huber** παρουσίασε τη μικρότερη τιμή σφάλματος (**MAE = 0.0223**) και κατέλαβε την πρώτη θέση στην κατάταξη των μοντέλων. Ακολούθησαν τα μοντέλα **Ridge** (**MAE = 0.0405**), **Elastic Net** (**MAE = 0.0417**) και **Lasso** (**MAE = 0.0440**). Το μοντέλο **AdaBoost** κατέγραψε τιμή **MAE = 0.1386**, παρόμοια με τα μοντέλα **KNN** (**0.1386**) και **Gradient Boosting** (**0.1386**). Τα μοντέλα **Extra Trees** (**0.1394**), **Random Forest** (**0.1399**), **Gaussian Process Regression** (**0.1434**), **HistGradientBoosting** (**0.1471**) και **SVR-RBF** (**0.1751**) παρουσίασαν υψηλότερες τιμές σφάλματος. Με βάση τα αποτελέσματα αυτά, τα μοντέλα **Huber**, **Ridge**, **Elastic Net**, **Lasso** και **AdaBoost** επιλέχθηκαν για τη δημιουργία του τελικού σχήματος Ensemble Learning, καθώς εμφάνισαν τις χαμηλότερες τιμές MAE.

5.3 Ανάλυση επιλεγμένων μοντέλων

5.3.1 Παράδειγμα 1°

Μετά την αξιολόγηση των μοντέλων με βάση το σφάλμα MAE, το σύστημα επιλέγει τα μοντέλα με την καλύτερη απόδοση, τα οποία χρησιμοποιούνται στη συνέχεια για την πρόβλεψη των τιμών καυσίμων.

```
Greece | Gasoline: last=2.060 on 2026-06-03 | train_points=90 | supervised_samples=83
ranked(MAE): huber:0.0223, ridge:0.0405, elastic:0.0417, lasso:0.0440, adaboost:0.1386
kept: ['huber', 'ridge', 'elastic', 'lasso', 'adaboost']
excluded: ['rf', 'extra_trees', 'gboost', 'hgb', 'knn', 'svr_rbf', 'gpr_rbf']
-- per-model next7 --
huber      : [2.129, 2.164, 2.168, 2.177, 2.184, 2.187, 2.181]
ridge      : [2.111, 2.129, 2.119, 2.122, 2.122, 2.113, 2.103]
elastic     : [2.053, 2.048, 2.042, 2.037, 2.033, 2.028, 2.020]
lasso       : [2.051, 2.045, 2.038, 2.033, 2.028, 2.023, 2.016]
adaboost    : [2.105, 2.105, 2.105, 2.105, 2.105, 2.105, 2.105]
```

Figure 12 Επιλεγμένα μοντέλα για την βενζίνη στην Ελλάδα

Στην περίπτωση της Ελλάδας (βενζίνη), τα επιλεγμένα μοντέλα (Top-5) είναι τα εξής:

- **Huber**

- **Ridge**
- **Elastic Net**
- **Lasso**
- **AdaBoost**

Τα μοντέλα αυτά παρουσιάζουν τις μικρότερες τιμές MAE και συνεπώς θεωρούνται τα καταλληλότερα για τη δημιουργία του τελικού σχήματος Ensemble Learning.

Συγκεκριμένα, Το μοντέλο **Huber** παρουσίασε την καλύτερη απόδοση με τιμή **MAE = 0.0223**, ακολουθούμενο από τα μοντέλα **Ridge (0.0405)**, **Elastic Net (0.0417)** και **Lasso (0.0440)**. Το μοντέλο **AdaBoost** κατέγραψε τιμή **MAE = 0.1386** και συμπλήρωσε την πεντάδα των επιλεγμένων μοντέλων.

Οι προβλέψεις των επιλεγμένων μοντέλων παρουσιάζουν μικρές διαφοροποιήσεις μεταξύ τους. Η τελική πρόβλεψη προέκυψε από τον μέσο όρο των επιμέρους προβλέψεων και για τις επόμενες επτά ημέρες διαμορφώθηκε στις τιμές **[2.090, 2.098, 2.094, 2.095, 2.094, 2.091, 2.085] €/L**. Η χρήση του μέσου όρου συμβάλλει στη μείωση της επίδρασης πιθανών ακραίων προβλέψεων και οδηγεί σε πιο σταθερά και αξιόπιστα αποτελέσματα.

5.3.2 Παράδειγμα 2°

Ως δεύτερο παράδειγμα, εξετάζεται η περίπτωση της Γαλλίας (diesel), όπως παρουσιάζεται στην παρακάτω εικόνα, όπου εμφανίζεται η κατάταξη των μοντέλων με βάση το σφάλμα.

```
France | Diesel: last=2.050 on 2026-06-03 | train_points=90 | supervised_samples=83
ranked(MAE): huber:0.0145, ridge:0.0253, elastic:0.0260, lasso:0.0266, gpr_rbf:0.0430, knn:0.0664,
kept: ['huber', 'ridge', 'elastic', 'lasso', 'gpr_rbf']
excluded: ['rf', 'extra_trees', 'gboost', 'adaboost', 'hgb', 'knn', 'svr_rbf']
-- per-model next7 --
huber      : [2.117, 2.155, 2.160, 2.168, 2.175, 2.179, 2.177]
ridge      : [2.094, 2.111, 2.098, 2.105, 2.107, 2.099, 2.088]
elastic    : [2.054, 2.057, 2.051, 2.050, 2.049, 2.047, 2.044]
lasso      : [2.054, 2.056, 2.053, 2.051, 2.050, 2.048, 2.045]
gpr_rbf    : [2.110, 2.138, 2.136, 2.130, 2.127, 2.122, 2.121]
```

Figure 13 Επιλεγμένα μοντέλα για την πετρέλαιο στην Γαλλία

Στην περίπτωση της Γαλλίας (πετρέλαιο), τα επιλεγμένα μοντέλα (Top-5) είναι τα εξής:

- **Huber**
- **Ridge**
- **Elastic Net**
- **Lasso**
- **GPR-RBF**

Τα παραπάνω μοντέλα παρουσιάζουν τις μικρότερες τιμές MAE και συνεπώς θεωρούνται τα καταλληλότερα για την πρόβλεψη των τιμών πετρελαίου.

Τα παραπάνω μοντέλα παρουσίασαν τις μικρότερες τιμές MAE και επιλέχθηκαν για τη δημιουργία του τελικού σχήματος Ensemble Learning. Το μοντέλο **Huber** εμφάνισε την καλύτερη απόδοση με **MAE = 0.0145**, παρουσιάζοντας τη μικρότερη απόκλιση μεταξύ πραγματικών και προβλεπόμενων τιμών. Ακολούθησαν τα μοντέλα **Ridge (MAE = 0.0253)**, **Elastic Net (MAE = 0.0260)** και **Lasso (MAE = 0.0266)**, τα οποία παρουσίασαν παρόμοια επίπεδα ακρίβειας. Το μοντέλο **GPR-RBF** συμπλήρωσε την πεντάδα των επιλεγμένων μοντέλων με **MAE = 0.0430**. Τα αποτελέσματα δείχνουν ότι το μοντέλο Huber προσαρμόζεται

αποτελεσματικότερα στα συγκεκριμένα δεδομένα, ενώ και τα υπόλοιπα επιλεγμένα μοντέλα παρουσιάζουν ικανοποιητική απόδοση.

5.4 Τελική πρόβλεψη τιμών καυσίμων (Next 7 Days)

5.4.1 Εφαρμογή του Penalty Mechanism

Για την ενίσχυση της αξιοπιστίας των τελικών προβλέψεων, το προτεινόμενο σύστημα ενσωματώνει έναν μηχανισμό διόρθωσης (**Penalty Mechanism**), ο οποίος εφαρμόζεται μετά την ολοκλήρωση της διαδικασίας **Ensemble Learning**. Στόχος του μηχανισμού είναι ο περιορισμός ακραίων ή μη ρεαλιστικών τιμών που ενδέχεται να προκύψουν κατά την πρόβλεψη των μελλοντικών τιμών καυσίμων.

Η λειτουργία του μηχανισμού βασίζεται στην ανάλυση των πιο πρόσφατων ιστορικών δεδομένων της χρονοσειράς. Συγκεκριμένα, λαμβάνονται υπόψη οι **14 τελευταίες διαθέσιμες παρατηρήσεις (lookback = 14)**, από τις οποίες υπολογίζονται ο μέσος όρος και η τυπική απόκλιση των τιμών. Με βάση τα στατιστικά αυτά χαρακτηριστικά προσδιορίζεται ένα αποδεκτό εύρος μεταβολής, το οποίο εκφράζει τη φυσιολογική συμπεριφορά της χρονοσειράς κατά το πρόσφατο χρονικό διάστημα.

Το εύρος αυτό υπολογίζεται χρησιμοποιώντας **συντελεστή απόκλισης ίσο με 1.0 (penalty factor = 1.0)**, ο οποίος πολλαπλασιάζεται με την τυπική απόκλιση των πρόσφατων δεδομένων. Παράλληλα, εφαρμόζεται ένα **ελάχιστο ποσοστιαίο εύρος ίσο με 2%** του μέσου όρου των πρόσφατων τιμών (**min_pct_band = 0.02**), ώστε να επιτρέπονται λογικές διακυμάνσεις ακόμη και σε περιόδους χαμηλής μεταβλητότητας. Με τον τρόπο αυτό διαμορφώνεται ένα δυναμικό εύρος αποδεκτών τιμών, το οποίο προσαρμόζεται στα χαρακτηριστικά της εκάστοτε χρονοσειράς.

Για κάθε νέα πρόβλεψη πραγματοποιείται έλεγχος ως προς το αν η τιμή βρίσκεται εντός των επιτρεπτών ορίων. Σε περίπτωση που η προβλεπόμενη τιμή υπερβαίνει το ανώτερο ή το κατώτερο όριο, προσαρμόζεται αυτόματα στην πλησιέστερη αποδεκτή τιμή. Με τον τρόπο αυτό αποτρέπεται η εμφάνιση ακραίων διακυμάνσεων που δεν δικαιολογούνται από τη συμπεριφορά των ιστορικών δεδομένων και μειώνεται η πιθανότητα παραγωγής μη ρεαλιστικών προβλέψεων.

Σημαντικό χαρακτηριστικό της υλοποίησης είναι ότι ο μηχανισμός διόρθωσης **δεν εφαρμόζεται στα επιμέρους μοντέλα μηχανικής μάθησης**. Αρχικά παράγονται οι προβλέψεις όλων των επιλεγμένων μοντέλων, στη συνέχεια υπολογίζεται η **τελική πρόβλεψη μέσω του μέσου όρου** των αποτελεσμάτων τους και μόνο τότε εφαρμόζεται η διαδικασία διόρθωσης. Η προσέγγιση αυτή διατηρεί τη συνεισφορά κάθε αλγορίθμου στο τελικό αποτέλεσμα, ενώ παράλληλα εξασφαλίζει ότι η τελική πρόβλεψη παραμένει συμβατή με τη συμπεριφορά των ιστορικών δεδομένων.

Η χρήση του **Penalty Mechanism** συμβάλλει ουσιαστικά στη βελτίωση της **σταθερότητας**, της **αξιοπιστίας** και της **συνέπειας** των προβλέψεων. Οι τελικές τιμές παραμένουν εντός ρεαλιστικών ορίων, χωρίς να περιορίζεται υπερβολικά η δυνατότητα του συστήματος να ανιχνεύει πραγματικές τάσεις και μεταβολές της αγοράς. Ως αποτέλεσμα, το σύστημα παράγει πιο αξιόπιστες εκτιμήσεις για τις μελλοντικές τιμές καυσίμων, βελτιώνοντας τη συνολική ποιότητα και χρηστικότητα των αποτελεσμάτων.

5.4.2 Τελική πρόβλεψη για την Ελλάδα (βενζίνη)

```
Greece | Gasoline: last=2.060 on 2026-06-03 | train_points=90 | supervised_samples=83
ranked(MAE): huber:0.0223, ridge:0.0405, elastic:0.0417, lasso:0.0440, adaboost:0.1386
kept: ['huber', 'ridge', 'elastic', 'lasso', 'adaboost']
excluded: ['rf', 'extra_trees', 'gboost', 'hgb', 'knn', 'svr_rbf', 'gpr_rbf']
-- per-model next7 --
huber      : [2.129, 2.164, 2.168, 2.177, 2.184, 2.187, 2.181]
ridge      : [2.111, 2.129, 2.119, 2.122, 2.122, 2.113, 2.103]
elastic    : [2.053, 2.048, 2.042, 2.037, 2.033, 2.028, 2.020]
lasso      : [2.051, 2.045, 2.038, 2.033, 2.028, 2.023, 2.016]
adaboost   : [2.105, 2.105, 2.105, 2.105, 2.105, 2.105, 2.105]
FINAL MEAN next7=[2.090, 2.098, 2.094, 2.095, 2.094, 2.091, 2.085] (no penalty)
```

Figure 14 Τελική Πρόβλεψη Τιμής Βενζίνης στην Ελλάδα

Όπως φαίνεται στην παραπάνω εικόνα, για την περίπτωση της Γαλλίας και του πετρελαίου, η τελική πρόβλεψη των επιλεγμένων μοντέλων διαμορφώνεται σε τιμές που κυμαίνονται από **2.086 €/L έως 2.104 €/L** για τις επόμενες επτά ημέρες. Οι προβλεπόμενες τιμές παρουσιάζουν μικρές διακυμάνσεις και παραμένουν σε σχετικά σταθερά επίπεδα, γεγονός που υποδηλώνει ότι τα μοντέλα συγκλίνουν σε παρόμοιες εκτιμήσεις. Η περιορισμένη μεταβλητότητα των προβλέψεων δείχνει ότι δεν αναμένονται σημαντικές μεταβολές στην τιμή του πετρελαίου κατά το εξεταζόμενο χρονικό διάστημα. Επιπλέον, η χρήση της μεθόδου Ensemble Learning συμβάλλει στην εξομάλυνση των επιμέρους προβλέψεων και οδηγεί σε πιο σταθερά και αξιόπιστα αποτελέσματα.

Επιπλέον, όπως φαίνεται στην εικόνα, ο μηχανισμός διόρθωσης (*Penalty Mechanism*) δεν ενεργοποιήθηκε κατά τη συγκεκριμένη πρόβλεψη, καθώς οι παραγόμενες τιμές βρίσκονταν ήδη εντός των αποδεκτών ορίων που έχουν καθοριστεί από τα πρόσφατα ιστορικά δεδομένα. Για τον λόγο αυτό η τελική πρόβλεψη εμφανίζεται με την ένδειξη **"(no penalty)"**, γεγονός που υποδηλώνει ότι οι προβλέψεις των μοντέλων θεωρήθηκαν ρεαλιστικές και δεν απαιτήθηκε καμία επιπλέον προσαρμογή.

5.4.3 Τελική πρόβλεψη για τη Γαλλία (πετρέλαιο)

```
France | Diesel: last=2.050 on 2026-06-03 | train_points=90 | supervised_samples=83
ranked(MAE): huber:0.0145, ridge:0.0253, elastic:0.0260, lasso:0.0266, gpr_rbf:0.0430, knn:0.0664,
kept: ['huber', 'ridge', 'elastic', 'lasso', 'gpr_rbf']
excluded: ['rf', 'extra_trees', 'gboost', 'adaboost', 'hgb', 'knn', 'svr_rbf']
-- per-model next7 --
huber      : [2.117, 2.155, 2.160, 2.168, 2.175, 2.179, 2.177]
ridge      : [2.094, 2.111, 2.098, 2.105, 2.107, 2.099, 2.088]
elastic    : [2.054, 2.057, 2.051, 2.050, 2.049, 2.047, 2.044]
lasso      : [2.054, 2.056, 2.053, 2.051, 2.050, 2.048, 2.045]
gpr_rbf    : [2.110, 2.138, 2.136, 2.130, 2.127, 2.122, 2.121]
FINAL MEAN next7=[2.086, 2.104, 2.099, 2.101, 2.102, 2.099, 2.095] (no penalty)
```

Figure 15 Τελική Πρόβλεψη Τιμής Πετρελαίου στην Γαλλία

Στην παραπάνω εικόνα παρατηρείται ότι οι προβλεπόμενες τιμές πετρελαίου για τη Γαλλία κυμαίνονται μεταξύ **2.086 €/L και 2.104 €/L** κατά τις επόμενες επτά ημέρες. Οι διακυμάνσεις είναι περιορισμένες και οι τιμές παραμένουν σε σχετικά σταθερά επίπεδα, χωρίς να εμφανίζεται κάποια έντονη ανοδική ή καθοδική τάση. Η συμπεριφορά αυτή υποδηλώνει ότι

τα μοντέλα συγκλίνουν σε παρόμοιες προβλέψεις και εκτιμούν ότι η αγορά θα διατηρηθεί σε κατάσταση σχετικής σταθερότητας κατά το εξεταζόμενο χρονικό διάστημα. Η χρήση του Ensemble Learning συμβάλλει στην εξομάλυνση των επιμέρους προβλέψεων και οδηγεί σε πιο αξιόπιστα αποτελέσματα. Επιπλέον, όπως φαίνεται από την ένδειξη **"(no penalty)"**, ο μηχανισμός διόρθωσης δεν ενεργοποιήθηκε στη συγκεκριμένη περίπτωση, καθώς οι τελικές προβλέψεις βρίσκονταν ήδη εντός των αποδεκτών ορίων που έχουν καθοριστεί από τα πρόσφατα ιστορικά δεδομένα. Το γεγονός αυτό υποδηλώνει ότι οι παραγόμενες προβλέψεις θεωρήθηκαν ρεαλιστικές και δεν απαιτήθηκε κάποια επιπλέον προσαρμογή.

5.4.4 Τελική πρόβλεψη για τη Ελλάδα (υγραέριο LPG)

```
Greece | LPG: last=1.120 on 2026-06-03 | train_points=90 | supervised_samples=83
ranked(MAE): lasso:0.0138, elastic:0.0138, huber:0.0147, ridge:0.0156, svr_rbf:0.0188, hgb:0.0226,
kept: ['lasso', 'elastic', 'huber', 'ridge', 'svr_rbf']
excluded: ['rf', 'extra_trees', 'gboost', 'adaboost', 'hgb', 'knn', 'gpr_rbf']
-- per-model next7 --
lasso      : [1.150, 1.169, 1.176, 1.184, 1.191, 1.197, 1.202]
elastic    : [1.140, 1.155, 1.160, 1.174, 1.184, 1.190, 1.197]
huber      : [1.118, 1.118, 1.105, 1.104, 1.108, 1.110, 1.109]
ridge      : [1.140, 1.148, 1.136, 1.154, 1.167, 1.179, 1.186]
svr_rbf    : [1.172, 1.172, 1.172, 1.175, 1.176, 1.177, 1.178]
FINAL MEAN next7=[1.144, 1.152, 1.150, 1.158, 1.165, 1.171, 1.175] (no penalty)
```

Figure 16 Τελική Πρόβλεψη Τιμής Αερίου στην Ελλάδα

Στην παραπάνω εικόνα παρατηρείται ότι οι προβλεπόμενες τιμές LPG για την Ελλάδα κυμαίνονται μεταξύ **1.144 €/L** και **1.175 €/L** κατά τις επόμενες επτά ημέρες. Οι τιμές παρουσιάζουν ήπια ανοδική τάση, καθώς η πρόβλεψη αυξάνεται σταδιακά από την πρώτη έως την τελευταία ημέρα. Η συμπεριφορά αυτή υποδηλώνει ότι τα επιλεγμένα μοντέλα εκτιμούν μια μικρή αλλά σταθερή αύξηση της τιμής του υγραερίου στο άμεσο μέλλον. Η χρήση της μεθόδου Ensemble Learning συμβάλλει στην εξομάλυνση των επιμέρους προβλέψεων και οδηγεί σε πιο σταθερά και αξιόπιστα αποτελέσματα.

5.4.5 Τελική πρόβλεψη για τη Γερμανία (βενζίνη)- Penalty

```
Germany | Gasoline: last=1.900 on 2026-06-03 | train_points=90 | supervised_samples=83
ranked(MAE): elastic:0.0180, huber:0.0181, lasso:0.0182, extra_trees:0.0216, ridge:0.0238,
.0268, hgb:0.0315, rf:0.0348, gboost:0.0350, knn:0.0350, adaboost:0.0350
kept: ['elastic', 'huber', 'lasso', 'extra_trees', 'ridge']
excluded: ['rf', 'gboost', 'adaboost', 'hgb', 'knn', 'svr_rbf', 'gpr_rbf']
-- per-model next7 --
elastic    : [1.939, 1.965, 1.983, 1.995, 2.004, 2.009, 2.013]
huber      : [1.938, 1.966, 1.964, 1.964, 1.964, 1.964, 1.965]
lasso      : [1.940, 1.967, 1.985, 1.997, 2.005, 2.010, 2.014]
extra_trees : [1.943, 1.960, 1.965, 1.966, 1.965, 1.965, 1.965]
ridge      : [1.940, 1.962, 1.970, 1.977, 1.985, 1.992, 2.004]
RAW FINAL MEAN next7=[1.940, 1.964, 1.973, 1.980, 1.984, 1.988, 1.992]
FINAL MEAN next7=[1.955, 1.964, 1.973, 1.980, 1.984, 1.988, 1.992]
```

Figure 17 Τελική Πρόβλεψη Τιμής Βενζίνης στην Γερμανία

Στην παραπάνω εικόνα παρατηρείται ότι οι προβλεπόμενες τιμές βενζίνης για τη Γερμανία κυμαίνονται μεταξύ **1.955 €/L** και **1.992 €/L** κατά τις επόμενες επτά ημέρες. Οι τιμές παρουσιάζουν ήπια ανοδική τάση, καθώς η πρόβλεψη αυξάνεται σταδιακά από την πρώτη έως την τελευταία ημέρα. Η συμπεριφορά αυτή υποδηλώνει ότι τα επιλεγμένα μοντέλα εκτιμούν μια μικρή αλλά σταθερή αύξηση της τιμής της βενζίνης στο άμεσο μέλλον. Η χρήση της μεθόδου Ensemble Learning συμβάλλει στην εξομάλυνση των επιμέρους προβλέψεων και οδηγεί σε πιο σταθερά και αξιόπιστα αποτελέσματα.

Επιπλέον, στην παρούσα περίπτωση παρατηρείται η λειτουργία του μηχανισμού διόρθωσης (*Penalty Mechanism*). Συγκεκριμένα, η αρχική πρόβλεψη του Ensemble Learning (**RAW FINAL MEAN**) για την πρώτη ημέρα ήταν **1.940 €/L**, ενώ μετά την εφαρμογή του μηχανισμού η τελική πρόβλεψη (**FINAL MEAN**) προσαρμόστηκε σε **1.955 €/L**. Οι υπόλοιπες προβλέψεις παρέμειναν αμετάβλητες, καθώς βρίσκονταν ήδη εντός των αποδεκτών ορίων που έχουν καθοριστεί από τα πρόσφατα ιστορικά δεδομένα. Το συγκεκριμένο παράδειγμα αναδεικνύει τη λειτουργία του Penalty Mechanism, το οποίο παρεμβαίνει μόνο όταν εντοπίζονται αποκλίσεις που ενδέχεται να οδηγήσουν σε μη ρεαλιστικές προβλέψεις. Με τον τρόπο αυτό διασφαλίζεται ότι οι τελικές τιμές παραμένουν συμβατές με τη συμπεριφορά της χρονοσειράς, ενισχύοντας τη σταθερότητα και την αξιοπιστία του συστήματος πρόβλεψης.

5.5 Σύγκριση Προβλεπόμενων και Πραγματικών Τιμών Καυσίμων

Για την αξιολόγηση της αποτελεσματικότητας του προτεινόμενου συστήματος πραγματοποιήθηκε σύγκριση των προβλεπόμενων τιμών με τις πραγματικές τιμές καυσίμων που δημοσιεύονται στο Παρατηρητήριο Τιμών Υγρών Καυσίμων του Υπουργείου Ανάπτυξης και Ανταγωνιστικότητας και είναι διαθέσιμες μέσω της διαδικτυακής πλατφόρμας FuelPrices.gr. Η σύγκριση πραγματοποιήθηκε για το χρονικό διάστημα από 26/05/2026 έως 01/06/2026 και περιλαμβάνει τις τρεις βασικές κατηγορίες καυσίμων, δηλαδή τη βενζίνη, το πετρέλαιο κίνησης και το υγραέριο (LPG).

Το Παρατηρητήριο Τιμών Υγρών Καυσίμων αποτελεί τον επίσημο μηχανισμό συλλογής και δημοσίευσης στοιχείων λιανικών τιμών καυσίμων στην Ελλάδα, παρέχοντας καθημερινά επικαιροποιημένα δεδομένα για το σύνολο της ελληνικής επικράτειας. Για τον λόγο αυτό τα συγκεκριμένα στοιχεία χρησιμοποιήθηκαν ως σημείο αναφοράς για την επαλήθευση και αξιολόγηση των προβλέψεων του συστήματος.

Στόχος της διαδικασίας αξιολόγησης είναι η εκτίμηση του βαθμού προσέγγισης των πραγματικών τιμών από το προτεινόμενο μοντέλο πρόβλεψης, καθώς και η διερεύνηση της αξιοπιστίας και της πρακτικής εφαρμοσιμότητάς του σε πραγματικές συνθήκες λειτουργίας.

Σύγκριση προβλέψεων με πραγματικές τιμές (fuelprices.gr)

Τιμές σε ευρώ ανά λίτρο (€/L) – Από 26/5 έως 1/6

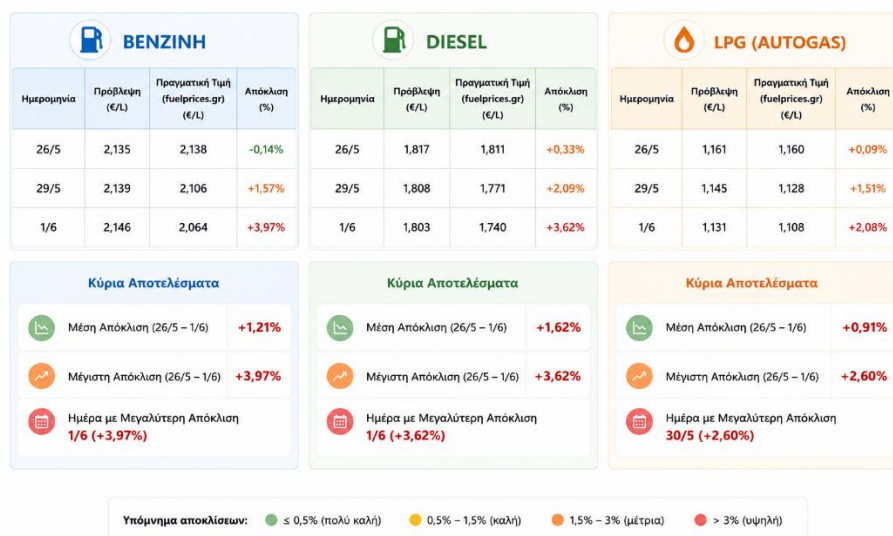


Figure 18 Σύγκριση προβλεπόμενων και πραγματικών τιμών καυσίμων για το χρονικό διάστημα 26/05/2026–01/06/2026.

Όπως φαίνεται στο παραπάνω σχήμα, το προτεινόμενο σύστημα παρουσίασε ικανοποιητική ακρίβεια και για τις τρεις κατηγορίες καυσίμων. Η μικρότερη μέση απόκλιση καταγράφηκε στο LPG, όπου η μέση διαφορά μεταξύ προβλεπόμενων και πραγματικών τιμών ήταν περίπου **0,91%**. Στη βενζίνη η μέση απόκλιση διαμορφώθηκε σε **1,21%**, ενώ στο πετρέλαιο κίνησης σε **1,62%**. Τα αποτελέσματα αυτά δείχνουν ότι το μοντέλο κατάφερε να προσεγγίσει με αρκετά υψηλή ακρίβεια τις πραγματικές τιμές της αγοράς.

Αναφορικά με τις μέγιστες αποκλίσεις, η μεγαλύτερη καταγράφηκε στη βενζίνη την **01/06/2026**, όπου η διαφορά μεταξύ πρόβλεψης και πραγματικής τιμής έφτασε το **3,97%**. Αντίστοιχα, στο πετρέλαιο η μέγιστη απόκλιση ήταν **3,62%**, επίσης την **01/06/2026**, ενώ στο LPG η μεγαλύτερη απόκλιση εμφανίστηκε στις **30/05/2026** και ανήλθε σε **2,60%**. Παρά τις αποκλίσεις αυτές, οι προβλέψεις διατήρησαν σωστά τη γενική κατεύθυνση της αγοράς και ακολούθησαν τη συνολική τάση μεταβολής των τιμών.

Συνολικά, τα αποτελέσματα επιβεβαιώνουν ότι η προτεινόμενη μεθοδολογία μπορεί να παρέχει αξιόπιστες βραχυπρόθεσμες προβλέψεις τιμών καυσίμων. Οι αποκλίσεις παρέμειναν σε χαμηλά επίπεδα, με μέγιστη τιμή μικρότερη του 4%, γεγονός που θεωρείται ιδιαίτερα ικανοποιητικό για δεδομένα που επηρεάζονται από πλήθος εξωτερικών οικονομικών και ενεργειακών παραγόντων. Η χρήση της μεθόδου Ensemble Learning σε συνδυασμό με τον μηχανισμό Penalty Mechanism συνέβαλε σημαντικά στη βελτίωση της σταθερότητας και της αξιοπιστίας των τελικών προβλέψεων.

ΚΕΦΑΛΑΙΟ 6 Συμπεράσματα

6.1 Συμπεράσματα

Στην παρούσα εργασία αναπτύχθηκε και αξιολογήθηκε ένα σύστημα πρόβλεψης τιμών καυσίμων με χρήση τεχνικών μηχανικής μάθησης. Για την εκπαίδευση και αξιολόγηση του συστήματος χρησιμοποιήθηκαν ιστορικά δεδομένα τιμών καυσίμων από διαφορετικές χώρες και τύπους καυσίμων, ενώ εξετάστηκαν πολλαπλοί αλγόριθμοι παλινδρόμησης με στόχο την επιλογή των μοντέλων που παρουσιάζουν τη μεγαλύτερη ακρίβεια.

Τα αποτελέσματα έδειξαν ότι δεν υπάρχει ένα μοντέλο το οποίο να υπερέχει σε όλες τις περιπτώσεις, καθώς η απόδοση κάθε αλγορίθμου εξαρτάται από τα χαρακτηριστικά των δεδομένων. Για τον λόγο αυτό υιοθετήθηκε η προσέγγιση Ensemble Learning, όπου οι προβλέψεις των πέντε καλύτερων μοντέλων συνδυάζονται μέσω υπολογισμού του μέσου όρου. Η μέθοδος αυτή οδήγησε σε πιο σταθερές και αξιόπιστες προβλέψεις σε σχέση με τη χρήση ενός μόνο μοντέλου.

Παράλληλα, ενσωματώθηκε ένας μηχανισμός διόρθωσης προβλέψεων (Penalty Mechanism), ο οποίος εφαρμόζεται στην τελική πρόβλεψη με σκοπό τον περιορισμό ακραίων τιμών και τη διασφάλιση της συμβατότητας των αποτελεσμάτων με τη συμπεριφορά των ιστορικών δεδομένων. Η εφαρμογή του μηχανισμού αυτού συνέβαλε στη βελτίωση της αξιοπιστίας του συστήματος και στην αποφυγή μη ρεαλιστικών προβλέψεων.

Συνολικά, η εργασία απέδειξε ότι οι τεχνικές μηχανικής μάθησης μπορούν να χρησιμοποιηθούν αποτελεσματικά για την πρόβλεψη τιμών καυσίμων, παρέχοντας αξιόπιστες εκτιμήσεις για βραχυπρόθεσμο χρονικό ορίζοντα και συμβάλλοντας στην καλύτερη κατανόηση της εξέλιξης των τιμών.

6.2 Μελλοντικές Βελτιώσεις

Παρόλο που το προτεινόμενο σύστημα παρουσίασε ικανοποιητικά αποτελέσματα, υπάρχουν αρκετές δυνατότητες περαιτέρω βελτίωσης και επέκτασης. Μία σημαντική κατεύθυνση αφορά την ενσωμάτωση πρόσθετων παραγόντων που επηρεάζουν τις τιμές των καυσίμων, όπως οι τιμές του αργού πετρελαίου, οι συναλλαγματικές ισοτιμίες, οικονομικοί δείκτες και γεωπολιτικά γεγονότα.

Επιπλέον, θα μπορούσε να εξεταστεί η χρήση πιο προηγμένων μοντέλων χρονοσειρών και τεχνικών βαθιάς μάθησης, όπως τα δίκτυα LSTM και οι αρχιτεκτονικές Transformer, τα οποία έχουν παρουσιάσει ιδιαίτερα υψηλές επιδόσεις σε προβλήματα πρόβλεψης χρονικών σειρών.

Μία ακόμη πιθανή επέκταση είναι η βελτίωση του μηχανισμού Ensemble Learning μέσω δυναμικής επιλογής μοντέλων και προσαρμογής των βαρών τους ανάλογα με την απόδοσή τους. Παράλληλα, ο μηχανισμός Penalty θα μπορούσε να εξελιχθεί ώστε να προσαρμόζει αυτόματα τις παραμέτρους του σύμφωνα με τη μεταβλητότητα των δεδομένων.

Τέλος, η ανάπτυξη μιας διαδικτυακής εφαρμογής με αυτόματη ενημέρωση δεδομένων και παραγωγή προβλέψεων σε πραγματικό χρόνο θα μπορούσε να αυξήσει σημαντικά την

πρακτική αξιοποίηση του συστήματος και να το καταστήσει χρήσιμο εργαλείο για την παρακολούθηση της αγοράς καυσίμων.

ΒΙΒΛΙΟΓΡΑΦΙΑ

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer.
2. Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*.
3. Mitchell, T. M. (1997). *Machine Learning*. McGraw-Hill.
4. Breiman, L. (2001). Random Forests. *Machine Learning Journal*.
5. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*.
6. Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*.
7. Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*.
8. Provost, F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media.
9. Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms*. Chapman & Hall.
10. Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*.
11. Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.
12. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 29(5), 1189–1232.
13. Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
14. Jordan, M. I., & Mitchell, T. M. (2015). *Machine Learning: Trends, Perspectives, and Prospects*. Science.
15. <https://bigblue.academy/en/supervised-learning>
16. <https://www.techtarget.com/searchenterpriseai/definition/unsupervised-learning>
17. <https://www.synopsys.com/ai/what-is-reinforcement-learning.html>