



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



Αξιολόγηση Μεθόδων Πρόβλεψης της Τιμής Ηλεκτρικής Ενέργειας στην Αγορά της Ουγγαρίας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΗΤΣΗ ΕΛΕΝΗ

Επιβλέπων : Βασίλειος Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



Αξιολόγηση Μεθόδων Πρόβλεψης της Τιμής Ηλεκτρικής Ενέργειας στην Αγορά της Ουγγαρίας

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΗΤΣΗ ΕΛΕΝΗ

Επιβλέπων : Βασίλειος Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την x^η Ιουνίου 2020.

(Υπογραφή)

.....
Βασίλειος Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....
Χρυσόστομος Δούκας
Αναπλ. Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2020

(Υπογραφή)

.....

ΕΛΕΝΗ ΜΗΤΣΗ

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

© 2020 – All rights reserved. Με επιφύλαξη παντός δικαιώματος.

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 τους συγγραφείς.

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

Περίληψη

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι η εφαρμογή και η αξιολόγηση διαφόρων μεθόδων για βραχυπρόθεσμη πρόβλεψη της τιμής της ηλεκτρικής ενέργειας στην αγορά της Ουγγαρίας αξιοποιώντας δεδομένα και από διασυνοριακές χώρες ή χώρες με αλληλεξάρτηση. Σε ανταγωνιστικές αγορές ηλεκτρικής ενέργειας, οι συμμετέχοντες όπως είναι οι ηλεκτρικές επιχειρήσεις κοινής ωφέλειας, οι παραγωγοί ηλεκτρικής ενέργειας και οι έμποροι εκτίθενται σε αυξημένους κινδύνους που προκαλούνται από την αστάθεια των τιμών ηλεκτρικής ενέργειας. Επομένως, στην παρούσα διπλωματική εργασία αναπτύσσονται διαφορετικά μοντέλα πρόβλεψης και μέσω συγκριτικής αξιολόγησης προκύπτει ένα τελικό μοντέλο για παραγωγή αξιόπιστων προβλέψεων για την αγορά της επόμενης ημέρας. Η ανάπτυξη των μοντέλων πραγματοποιείται σε πραγματικά ιστορικά δεδομένα για την αγορά της Ουγγαρίας και των γύρω χωρών, ενώ η αξιολόγησή τους γίνεται σε πραγματικές συνθήκες αγοράς. Στο πλαίσιο της παρούσας εργασίας εξετάστηκαν τέσσερα μοντέλα, το πρώτο είναι το μοντέλο της Πολλαπλής Γραμμικής Παλινδρόμησης (MLR), το δεύτερο και το τρίτο μοντέλο ανήκουν στην κατηγορία των Δέντρων Απόφασης (DT) ενώ το τελευταίο ανήκει στην κατηγορία των Νευρωνικών Δικτύων και υλοποιήθηκε μέσω του αλγορίθμου Πολυεπίπεδων Αισθητήρων (MLP). Συμπληρωματικά, σε κάθε αλγόριθμο εφαρμόστηκε η μέθοδος της Σταδιακής Παλινδρόμησης (Stepwise regression) ώστε να επιλεγθούν οι ανεξάρτητες μεταβλητές εισόδου για τα τελικά μοντέλα πρόβλεψης.

Αρχικά, στην παρούσα εργασία πραγματοποιήθηκε μια εκτενής ανάλυση και παρουσίαση των αγορών ηλεκτρικής ενέργειας, δίνοντας ιδιαίτερη έμφαση στην αγορά της Ουγγαρίας που μελετάται. Στην συνέχεια, παρουσιάστηκε η ανάγκη για παραγωγή αξιόπιστων προβλέψεων για κάθε συμμετέχοντα στην αγορά της ηλεκτρικής ενέργειας ενώ έγινε μια κατηγοριοποίηση και ανάλυση αυτών βάσει του ορίζοντα πρόβλεψης. Έπειτα, ακολούθησε η παρουσίαση της επίδρασης των γειτονικών χωρών στην διαμόρφωση της τιμής της ηλεκτρικής ενέργειας για την υπό μελέτη χώρα όπως έχει προκύψει βάσει βιβλιογραφίας καθώς και οι παράγοντες που την διαμορφώνουν. Στο τρίτο κεφάλαιο, έγινε μια ανάλυση των μεθόδων και των μοντέλων πρόβλεψης που χρησιμοποιούνται, ενώ παρουσιάστηκαν εκτενέστερα οι αλγόριθμοι που εφαρμόστηκαν στην παρούσα εργασία. Στην συνέχεια, αναλύθηκαν οι στατιστικοί δείκτες σφάλματος βάσει των οποίων αξιολογείται η ακρίβεια ενός μοντέλου πρόβλεψης και αναπτύχθηκαν οι μέθοδοι της αλυσιδωτής επαλήθευσης και της σταδιακής παλινδρόμησης που εφαρμόστηκαν στην παρούσα εργασία.

Τέλος, στο τέταρτο κεφάλαιο αναλύεται η πειραματική διαδικασία της παρούσας εργασίας. Αρχικά, παρουσιάστηκαν και αναλύθηκαν τα δεδομένα που χρησιμοποιήθηκαν στο παρόν πρόβλημα. Στην συνέχεια, για κάθε αλγόριθμο προσδιορίστηκαν οι υπερπαραμέτροι, όπου αυτό χρειαζόταν, και παρουσιάστηκαν τα αποτελέσματα από την μέθοδο της προς τα πίσω επιλογής μεταβλητών εισόδου. Για την αξιολόγηση των μοντέλων και των μεθόδων χρησιμοποιήθηκε η μέθοδος της αλυσιδωτής επαλήθευσης.

Λέξεις Κλειδιά: Τιμή Ηλεκτρικής Ενέργειας, Ουγγρική Αγορά Ηλεκτρικής Ενέργειας, Μέθοδοι Πρόβλεψης, Πολυεπίπεδοι Αισθητήρες, Δέντρα Απόφασης, Πολλαπλή Γραμμική Παλινδρόμηση

Abstract

The scope of this thesis is to apply and evaluate various methods for short-term forecasting of the electricity price in the Hungarian market also using data from cross-border countries or interdependent countries. In competitive power markets, participants such as, electric utilities, power producers, and traders are exposed to increased risks caused by electricity price volatility. Therefore, in the present thesis different models for forecasting are developed and through comparative evaluation a final reliable model turns out for forecasting the day ahead market results. The development of the models is carried out in real historical data for the market of Hungary and the surrounding countries, while their evaluation is done in real market conditions. In particular, within the frame of this study four models were used, the first being the Multiple Linear Regression Model (MLR), the second and third model belongs to the Decision Tree (DT) class while the latter belongs to the Neural Network class and was implemented through Multi-Layer Perceptron (MLP). Additionally, the Stepwise regression method was applied to each algorithm to select the independent input variables for the final forecasting models.

Initially, in this thesis, an extensive analysis and presentation of the electricity markets took place, with particular emphasis on the Hungarian market being studied. Then, the need for reliable forecasts for each participant in power market was presented while a categorization and analysis of price forecasts was made based on the forecast horizon. This was followed by a presentation of the effect of neighboring countries on the price of electricity for the country under study as shown in the literature and the factors that affect the price. In the third chapter, an analysis of the methods and forecasting models used was made. The algorithms applied in this thesis were presented in more detail. Then, the statistical error indicators were analyzed based on which the accuracy of a prediction model is evaluated while the methods of cross validation and the stepwise selection of input variables applied in the present thesis were developed.

Finally, the fourth chapter analyzes the case study of the present thesis. Initially, the data used in this problem were presented and analyzed. Then for each algorithm, the hyperparameters, where needed, were identified, and the results from the method of backward selection of input variables were presented. The method of cross validation was used to evaluate the models and methods.

Keywords: Electricity Price, Hungarian Electricity Market, Forecasting Methods, Multi-Layer Perceptron, Decision Trees, Multiple Linear Regression

Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε κατά το ακαδημαϊκό έτος 2019-2020 στα πλαίσια της ολοκλήρωσης του Προγράμματος Μεταπτυχιακών Σπουδών «Τεχνο - Οικονομικά Συστήματα» στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα Καθηγητή κ. Ασημακόπουλο Βασίλειο για την ανάθεση αυτής της Διπλωματικής και την ευκαιρία που μου παρείχε να εκπονήσω την εργασία αυτή στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Προβλέψεων και Στρατηγικής. Επιπλέον, θα ήθελα να τον ευχαριστήσω για την συνεργασία που είχαμε κατά την εκπόνησή της και την ευκαιρίας που μου έδωσε να ασχοληθώ με τον τόσο ενδιαφέροντα τομέα των προβλέψεων. Επίσης, ευχαριστώ θερμά τον Καθηγητή κ. Ι. Ψαρρά και τον Αναπληρωτή Καθηγητή κ. Χ. Δούκα για την τιμή που μου έκαναν να συμμετάσχουν στην επιτροπή εξέτασης της εργασίας.

Τέλος, θα ήθελα ιδιαιτέρως να ευχαριστήσω τον ερευνητικό συνεργάτη της Μονάδας Προβλέψεων και Στρατηγικής, Ευάγγελο Σπηλιώτη για την πολύτιμη γνώση και υποστήριξη που πρόθυμα μου προσέφερε σε κάθε στάδιο υλοποίησης της παρούσας εργασίας καθώς και για την άψογη συνεργασία και καθοδήγησή του.

Πίνακας περιεχομένων

1	Εισαγωγή	1
1.1	Αγορά ηλεκτρικής ενέργειας	1
1.2	Δομή της Αγοράς της Ηλεκτρικής Ενέργειας.....	3
1.3	Η Αγορά Ηλεκτρικής Ενέργειας της Ουγγαρίας.....	6
1.4	Οργάνωση κειμένου	14
2	Πρόβλεψη τιμής ηλεκτρικής ενέργειας	17
2.1	Επίδρασης των γειτονικών χωρών στην τιμή της ηλεκτρικής ενέργειας	23
2.2	Παράγοντες που επιδρούν στην τιμή της ηλεκτρικής ενέργειας.....	25
3	Τεχνικές Προβλέψεων	29
3.1	Ποιοτικά Χαρακτηριστικά χρονοσειρών	30
3.2	Κατηγορίες μεθόδων προβλέψεων	32
3.2.1	<i>Βασικές κατηγορίες μεθόδων προβλέψεων</i>	32
3.2.2	<i>Μοντέλα πρόβλεψης</i>	32
3.2.3	<i>Στατιστική πρόβλεψη</i>	34
3.3	Ανάλυση μεθόδων εφαρμογής.....	36
3.3.1	<i>Πολλαπλή γραμμική παλινδρόμηση (MLR)</i>	36
3.3.2	<i>Μοντέλο πολυεπίπεδων αισθητήρων (MLP)</i>	37
3.3.3	<i>Δέντρα αποφάσεων</i>	42
3.4	Στατιστική Ανάλυση και Ακρίβεια Προβλέψεων	46
3.4.1	<i>Σφάλματα Προβλέψεων</i>	46
3.4.2	<i>Cross validation</i>	48
3.5	Σταδιακή παλινδρόμηση (Stepwise regression).....	50
4	Πειραματική Διαδικασία	53
4.1	Παρουσίαση δεδομένων	54
4.2	Πολλαπλή γραμμική παλινδρόμηση (MLR).....	62

4.3	Gradient Boosting Machine	65
4.4	Random Forest	68
4.5	Νευρωνικό δίκτυο (MLP).....	71
4.6	Επιλογή τελικού μοντέλου πρόβλεψης.....	74
5	Επίλογος.....	77
5.1	Σύνοψη και συμπεράσματα.....	77
5.2	Μελλοντικές επεκτάσεις.....	81
6	Βιβλιογραφία	84

Πίνακας εικόνων

Εικόνα 1:Επισκόπηση των διαφόρων τύπων αγορών ηλεκτρικής ενέργειας βάσει του χρόνου.....	6
Εικόνα 2: Μοντέλο αγοράς ηλεκτρικής ενέργειας της Ουγγαρίας	8
Εικόνα 3:Συνολική Κατανάλωση Ηλεκτρικής Ενέργειας[MWh] και ποσοστό Εισαγωγών-Εξαγωγών [%].....	9
Εικόνα 4: Μηνιαία πραγματικά δεδομένα ενέργειας για το διάστημα 1998-2020.....	10
Εικόνα 5: Μείγμα κατανάλωσης ηλεκτρικής ενέργειας Ουγγαρίας	10
Εικόνα 6:Μείγμα παραγωγής ηλεκτρικής ενέργειας Ουγγαρίας	11
Εικόνα 7: Διάγραμμα ζήτησης - προσφοράς πριν και μετά την ζεύξη αγορών.....	13
Εικόνα 8: Κατηγοριοποίηση προσεγγίσεων για πρόβλεψη τιμής ηλεκτρικής ενέργειας	19
Εικόνα 9: Κατανομή επιστημονικών άρθρων σύμφωνα με τη μέθοδο που χρησιμοποιείται για την πρόβλεψη τιμής ηλεκτρικής ενέργειας ^[8]	22
Εικόνα 10: Νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης (MLP).....	38
Εικόνα 11: Λογαριθμική συνάρτηση ενεργοποίησης	39
Εικόνα 12: Γραμμική συνάρτηση ενεργοποίησης	40
Εικόνα 13: Σχέση υπολογισμού βαρών μέσω backpropagation	40
Εικόνα 14:Παράδειγμα επαλήθευσης για n=10 ομάδες	49
Εικόνα 15: Παράδειγμα αλυσιδωτής επαλήθευσης	50
Εικόνα 16: Χρονοσειρά τιμής ηλεκτρικής ενέργειας για την αγορά της Ουγγαρίας...57	
Εικόνα 17: Αναπαράσταση χρονοσειράς τιμής με συχνότητα ημερήσια.....	58
Εικόνα 18: Αναπαράσταση χρονοσειράς τιμής με συχνότητα εβδομαδιαία	58
Εικόνα 19:Διάγραμμα συσχέτισης μεταβλητών	60

Ευρετήριο πινάκων

Πίνακας 1: Αποτελέσματα backward stepwise regression αλγορίθμου MLR	64
Πίνακας 2: Συνδυασμοί υπερπαραμέτρων αλγορίθμου gbm.....	66
Πίνακας 3: Αποτελέσματα συνδυασμός υπερπαραμέτρων του αλγορίθμου gbm.....	67
Πίνακας 4:Αποτελέσματα backward stepwise regression αλγορίθμου GBM	68
Πίνακας 5: Συνδυασμοί υπερπαραμέτρων αλγορίθμου randomforest.....	69
Πίνακας 6:Αποτελέσματα συνδυασμός υπερπαραμέτρων του αλγορίθμου randomforest	70
Πίνακας 7:Αποτελέσματα backward stepwise regression αλγορίθμου Randomforest	71
Πίνακας 8:Συνδυασμοί υπερπαραμέτρων αλγορίθμου MLP	72
Πίνακας 9:Αποτελέσματα συνδυασμός υπερπαραμέτρων του αλγορίθμου MLP.....	73
Πίνακας 10:Αποτελέσματα backward stepwise regression αλγορίθμου MLP	74
Πίνακας 11:Συγκεντρωτικά αποτελέσματα αλγορίθμων	75

1

Εισαγωγή

1.1 Αγορά ηλεκτρικής ενέργειας

Τις τελευταίες δύο δεκαετίες σε πολλές χώρες έχει απελευθερωθεί η αγορά της ηλεκτρικής ενέργειας. Η επιτυχία της απελευθέρωσης σε άλλες βιομηχανίες και η πεποίθηση ότι μπορεί να αναπαραχθεί και στον τομέα της ενέργειας, ως απαίτηση για τον διαχωρισμό των κάθετα ολοκληρωμένων δομών μονοπωλίου, ήταν οι κινητήριες δυνάμεις οι οποίες οδήγησαν στην απελευθέρωση του τομέα της ηλεκτρικής ενέργειας. Η τεχνολογική πρόοδος και οι αλλαγές στις τεχνολογίες παραγωγής και στα δίκτυα ηλεκτρικής ενέργειας επέτρεψαν το διαχωρισμό αυτού που στο παρελθόν είχε υπάρξει ένα φυσικό μονοπώλιο.

Οι μακροπρόθεσμοι στόχοι πίσω από την απελευθέρωση της ηλεκτρικής ενέργειας ήταν να προσελκύσει επενδύσεις, να αυξήσει την αποτελεσματικότητα και να ενθαρρύνει περαιτέρω την τεχνική καινοτομία. Οι μεγάλες επιχειρήσεις ηλεκτρικής ενέργειας που ήταν παλαιότερα μονοπώλιο έχουν πλέον απορρυθμιστεί και διαχωριστεί και σχήματα της αγοράς έχουν ξεκινήσει να παίζουν τον μεγάλο ρόλο στην διαμόρφωση και την εξέλιξη της αγοράς.

Ως αποτέλεσμα της απελευθέρωσης της αγοράς της ηλεκτρικής ενέργειας, η δυναμική του εμπορίου της ηλεκτρικής ενέργειας έχει αναμορφωθεί πλήρως. Ειδικότερα, η ηλεκτρική ενέργεια έχει γίνει εμπόρευμα που εμφανίζει ένα σύνολο χαρακτηριστικών που είναι ασυνήθιστα σε άλλες αγορές. Βασικά από αυτά τα χαρακτηριστικά είναι η σταθερή ισορροπία μεταξύ παραγωγής και κατανάλωσης που πρέπει να επιτευχθεί, φορτίο και παραγωγή που επηρεάζονται από εξωτερικές καιρικές συνθήκες καθώς και η εξάρτηση της κατανάλωσης από την ώρα της ημέρας, την ημέρα της εβδομάδας και την στιγμή στο έτος. Οι χρονοσειρές των τιμών της ηλεκτρικής ενέργειας έχουν σύνθετα χαρακτηριστικά, όπως είναι η μη στασιμότητα, η μη γραμμικότητα και υψηλή μεταβλητότητα, που καθιστούν δύσκολη την πρόβλεψή τους. Τα χαρακτηριστικά αυτά οφείλονται στο ότι στην παρούσα φάση η ηλεκτρική ενέργεια είναι μη αποθηκεύσιμη ενώ σε κάθε χρονική στιγμή σε σύντομο χρονικό διάστημα υπάρχει απαίτηση για συνεχή διατήρηση της ισορροπίας μεταξύ προσφοράς και ζήτησης.

Η πρωτόγνωρη συμπεριφορά που παρουσιάζει η ηλεκτρική ενέργεια έχει ως αποτέλεσμα οι αγορές να έχουν γίνει κεντρικό σημείο έρευνας στον ενεργειακό τομέα και η ακριβής πρόβλεψη των τιμών της ηλεκτρικής ενέργειας έχει αναδειχθεί ως μία από τις μεγαλύτερες προκλήσεις που αντιμετωπίζουν τα μέλη της αγοράς.

Το σύνηθες κίνητρο πίσω από αυτές τις προσπάθειες είναι κυρίως οικονομικό. Όσο αυξάνεται η ακρίβεια των προβλέψεων, οι αρνητικές οικονομικές επιπτώσεις που προκαλεί η αβεβαιότητα των τιμών μετριαζονται και οι παίχτες της αγοράς έχουν οικονομικό κέρδος. Επιπλέον, ένας άλλος παράγοντας που θα πρέπει να ληφθεί υπόψη είναι ότι οι αγορές ηλεκτρικής ενέργειας έχουν δημιουργηθεί για να διατηρείται το δίκτυο σταθερό. Ιδιαίτερα, καθώς οι τιμές γίνονται πιο ασταθείς η ισορροπία του δικτύου τίθεται σε κίνδυνο και ενδέχεται να χρειαστεί να χρησιμοποιηθούν στρατηγικές εφεδρείες ενώ αυξάνεται ο κίνδυνος για γενική διακοπή.

Επομένως, με την ακριβή πρόβλεψη των τιμών της ηλεκτρικής ενέργειας, όχι μόνο μπορούν να αποκομίσουν οι παίχτες οικονομικά κέρδη, αλλά και να βελτιωθεί η σταθερότητα του συστήματος.

1.2 Δομή της Αγοράς της Ηλεκτρικής Ενέργειας

Η δομή των αγορών ηλεκτρικής ενέργειας διαφέρει από χώρα σε χώρα. Τρία είναι τα βασικά μοντέλα που συναντώνται στις αγορές της ηλεκτρικής ενέργειας.

- i. Μοντέλο Κοινοπραξίας Ισχύος: το μοντέλο αυτό ορίζεται ως μία συγκεντρωτική αγορά στην οποία γίνεται εκκαθάριση για παραγωγούς και προμηθευτές ή καταναλωτές. Όλες οι εταιρείες παραγωγής υποβάλλουν ζεύγη προσφορών τιμής και ποσότητας για την ενέργεια που προμηθεύουν το Δίκτυο, τα οποία σχηματίζουν μια καμπύλη προσφοράς. Οι προσφερόμενες τιμές προκύπτουν είτε με βάση προκαθορισμένα μεταβλητά κόστη δημιουργώντας κοινοπραξίες βασισμένες στα κόστη είτε οι γεννήτριες μπορούν ελεύθερα να προσφέρουν σε οποιαδήποτε τιμή επιθυμούν μεταξύ ορισμένων γενικών ελάχιστων και μέγιστων τιμών δημιουργώντας κοινοπραξίες βασισμένες στην τιμή. Ο ανταγωνισμός των παραγωγών έγκειται στο δικαίωμά τους να παρέχουν ενέργεια στο σύστημα, όχι όμως σε συγκεκριμένους προμηθευτές ή καταναλωτές.

Από την πλευρά της ζήτησης, ο Διαχειριστής της Αγοράς προβλέπει καθημερινά το φορτίο ζήτησης και με βάση αυτό γίνεται η κατανομή των μονάδων παραγωγής. Αυτό ονομάζεται Μονομερής Κοινοπραξία Ισχύος (one-sided pool), ενώ είναι δυνατή ύπαρξη και Διμερούς Κοινοπραξία Ισχύος (two-sided pools) στην οποία ο Διαχειριστής κατά την κατανομή λαμβάνει υπ' όψη την αντίστοιχη καμπύλη ζήτησης που έχει σχηματιστεί με βάση τις προσφορές των προμηθευτών και των καταναλωτών. Οι προμηθευτές και οι καταναλωτές έχουν, σε αναλογία με τα προηγούμενα, στόχο την εξασφάλιση ηλεκτρικής ενέργειας.

Οι προσφορές που υποβάλλονται από την μεριά των προμηθευτών/καταναλωτών και αυτές από την μεριά των παραγωγών είναι συνήθως σταθερές, το οποίο σημαίνει ότι εάν ταιριάζουν στην διαδικασία εκκαθάρισης της αγοράς οδηγούν σε υποχρέωση να ληφθούν και αντίστοιχα να παραδοθούν οι αντίστοιχοι όγκοι αλλά και να διευθετηθούν οικονομικά. Ο

Διαχειριστής Συστήματος μέσα σε αυτό το μοντέλο θα κάνει την οικονομική κατανομή, από την οποία θα προκύψει μία συγκεκριμένη τιμή για την ηλεκτρική ενέργεια, δίνοντας έτσι στους συμμετέχοντες ένα σήμα που επηρεάζει τις αποφάσεις τους για κατανάλωση και για επενδύσεις. Στο μοντέλο αυτό οι Παραγωγοί που συμμετέχουν στη συναλλαγή πληρώνονται στην τιμή αυτή.

Η κοινοπραξία μπορεί να λειτουργήσει για την αγορά της επόμενης ημέρας ή μια αγορά κοντά σε πραγματικό χρόνο. Μπορεί επίσης να υπάρχει συνδυασμός πολλών αγορών, επόμενης ημέρας, ενδοημερήσιας και λεπτών μπροστά.

Παράλληλα με το μοντέλο της κοινοπραξίας, οι συμμετέχοντες στην αγορά έχουν την δυνατότητα επίσης να συνάψουν διμερείς συμβάσεις για ηλεκτρική ενέργεια. Αυτές οι συμβάσεις δεν πρέπει να είναι απαραίτητα οικονομικές, αλλά μπορεί επίσης να είναι διμερείς συμβάσεις για φυσική παράδοση ενέργειας. Σε αυτή την περίπτωση, μόνο οι διαφορές διευθετούνται μέσω της κοινοπραξίας, αλλά η παροχή ενέργειας από τους σταθμούς παραγωγής δεν επηρεάζεται από τις φυσικές διμερείς συμβάσεις. Τα καθαρά χρηματοοικονομικά διμερή συμβόλαια προσφέρουν παρόμοια πλεονεκτήματα.

- ii. Μοντέλο Διμερών Συμβάσεων: σε αυτό το μοντέλο Αγοράς οι πωλητές και οι αγοραστές έχουν την δυνατότητα να συνάπτουν ελεύθερα Διμερείς συμβάσεις για παροχή ηλεκτρικής ενέργειας. Οι πωλητές θα είναι συνήθως παραγωγοί και οι αγοραστές θα είναι εταιρείες διανομής και επιλέξιμοι καταναλωτές. Ωστόσο, οι παραγωγοί θα μπορούσαν επίσης να γίνουν αγοραστές (π.χ. σε περίπτωση έλλειψης παραγωγής). Ομοίως, οι καταναλωτές μπορούν να γίνουν πωλητές. Με άλλα λόγια, κάθε συμμετέχων στην αγορά γίνεται έμπορος. Οι μεσίτες μπορούν να λειτουργήσουν ως ενδιάμεσοι μεταξύ αγοραστών και πωλητών.

Αυτοί οι τύποι συναλλαγών αναφέρονται ως εξωχρηματιστηριακές. Στην πραγματικότητα, θα υπάρχουν πάντα διαφορές μεταξύ των όγκων που συναλλάσσονται και των πραγματικών μετρημένων όγκων. Αυτό συνεπάγεται ότι ο διαχειριστής συστήματος θα πρέπει να καθορίσει αυτές τις διαφορές (ή ανισορροπίες) και θα πρέπει να τις διευθετήσει. Σε πιο προχωρημένες αγορές, ο διαχειριστής του συστήματος διαχειρίζεται μια αγορά εξισορρόπησης (ή ρυθμίζει την αγορά ενέργειας) προκειμένου να καθορίζει μια τιμή βασισμένη στην αγορά για τον διακανονισμό αυτών των ανισορροπιών.

Παράλληλα με τις διμερείς συμβάσεις, θα μπορούσε να δημιουργηθεί μια εθελοντική ανταλλαγή ενέργειας ή να αναπτυχθεί στο μέλλον με πρωτοβουλία από τους συμμετέχοντες στην αγορά. Μια ανταλλαγή ενέργειας θα μπορούσε να προσφέρει στην αγορά της επόμενης ημέρας και στην ενδοημερήσια για τους συμμετέχοντες τα ακόλουθα οφέλη:

- Περισσότερη διαφάνεια τιμών
- Αγορά χωρίς κίνδυνο αντισυμβαλλόμενου
- Ανώνυμες συναλλαγές
- Εργαλείο βελτιστοποίησης του χαρτοφυλακίου συναλλαγών.

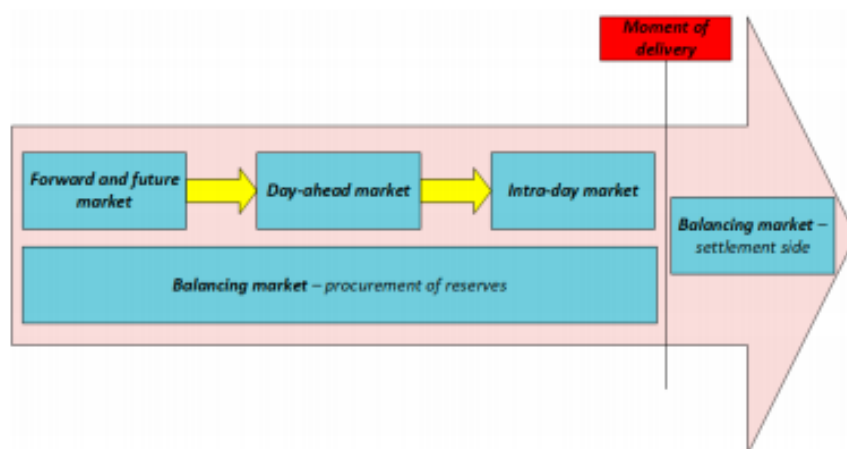
Μια τέτοια ανταλλαγή ενέργειας θα μπορούσε να χειριστεί ο διαχειριστής συστήματος ως κανονικός έμπορος αν και η ανταλλαγή ενέργειας θα ήταν ένας ειδικός τύπος εμπόρου, καθώς το σύνολο των συμβάσεων πώλησης και αγοράς θα πρέπει να είναι πάντα ισορροπημένο.

- iii. Υβριδικό Μοντέλο: πρόκειται για ένα μοντέλο που συνδυάζει τα χαρακτηριστικά των δύο προηγούμενων. Ο προμηθευτής έχει την δυνατότητα να επιλέξει είτε να αγοράσει την ενέργεια στην τιμή της αγοράς είτε να διαπραγματευτεί κατευθείαν με κάποιον παραγωγό μέσω διμερών συμβάσεων.

1.3 Η Αγορά Ηλεκτρικής Ενέργειας της Ουγγαρίας

Η απελευθέρωση της συγγκικής αγοράς ηλεκτρικής ενέργειας και φυσικού αερίου ολοκληρώθηκε το 2008. Συγκεκριμένα στην αγορά ηλεκτρικής ενέργειας της Ουγγαρίας, η οποία θα μας απασχολήσει στην παρούσα διπλωματική, έχει επικρατήσει και εφαρμόζεται το πρώτο μοντέλο της αγοράς. Πρόκειται για το πιο κοινό μοντέλο αγοράς, όπου οι παραγωγοί και οι προμηθευτές/καταναλωτές υποβάλλουν τις προσφορές τους οι οποίες αποτελούνται από ένα σύνολο ποσοτήτων με τιμές. Χρησιμοποιώντας έναν αλγόριθμο εκκαθάρισης αγοράς, ο διαχειριστής της αγοράς τότε καθορίζει τις τιμές για την επόμενη μέρα.

Στην αγορά της ηλεκτρικής ενέργειας στην χώρα της Ουγγαρίας, διαχειριστής του συστήματος μεταφοράς ηλεκτρικής ενέργειας είναι ο MAVIR. Για το σύστημα διανομής ηλεκτρικής ενέργειας υπάρχουν έξι διαχειριστές οι οποίοι είναι οι ακόλουθοι: E.ON South-donau, ÉMÁSZ Distribution System Operator, E.ON North-donau, DÉMÁSZ Distribution System Operator, ELMÜ DSO, E.ON Tisza-region. Το χρηματιστήριο, στο οποίο λαμβάνουν χώρα οι ανταλλαγές ενέργειας για την αγορά της επόμενης ημέρας, την ενδοημερήσια αγορά, οικονομικά συμβόλαια αλλά και βραχυπρόθεσμα συμβόλαια φυσικής παράδοσης (όπως φαίνονται και στο ακόλουθο σχήμα), είναι το Hungarian Power Exchange (HUPX). Τέλος, η ρυθμιστική αρχή ενέργειας είναι η *HUNGARIAN ENERGY AND PUBLIC UTILITY REGULATORY AUTHORITY*.



Εικόνα 1:Επισκόπηση των διαφόρων τύπων αγορών ηλεκτρικής ενέργειας βάσει του χρόνου

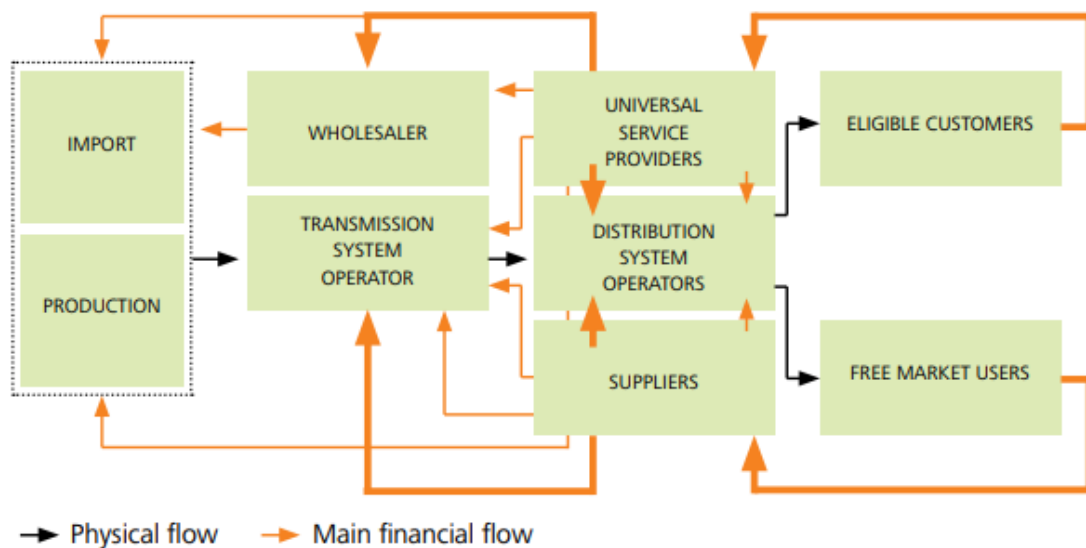
Οι Day-Ahead Markets (DAM) είναι η βάση της διαπραγμάτευσης ηλεκτρικής ενέργειας, όπου η ηλεκτρική ενέργεια εμπορεύεται μια μέρα πριν από την παράδοση. Η μεγάλη σημασία τους προκύπτει από την απαίτηση ισορροπίας για ολόκληρης της ημέρας ζώνη υποβολής προσφορών, δηλαδή, σε μια ζώνη προσφοράς, η προγραμματισμένη παραγωγή ηλεκτρικής ενέργειας πρέπει να είναι ίση με την προβλεπόμενη ζήτηση συν τη διαφορά μεταξύ εισαγωγών και εξαγωγών προς τις άλλες ζώνες υποβολής προσφορών.

Στις αγορές εντός της ημέρας (IM) η ηλεκτρική ενέργεια διατίθεται κατά την ημέρα παράδοσης. Αυτοί οι τύποι αγορών συμπληρώνουν τις DAM και εξασφαλίζουν την ισορροπία μεταξύ της προσφοράς και της ζήτησης που προκαλείται από απροσδόκητα συμβάντα που ενδέχεται να συμβούν μεταξύ του κλεισίματος του DAM και του χρόνου παράδοσης, π.χ. λόγω απροσδόκητης διακοπής ισχύος της εγκατάστασης ή ανακριβής πρόγνωση ανέμου. Χρησιμοποιώντας την πλατφόρμα της ενδοημερήσιας αγοράς, ο συμμετέχων στην αγορά μπορεί να ελαχιστοποιήσει το χάσμα μεταξύ των συμβολαίων της επόμενης ημέρας και του πραγματικού όγκου ηλεκτρικής ενέργειας που παράγεται / καταναλώνεται.

Ο ρόλος του Διαχειριστή Συστήματος (ΔΣΜ) είναι να διατηρήσει ένα υπόλοιπο σε πραγματικό χρόνο ενεργοποιώντας τα αποθεματικά. Εξισορρόπηση Αγορών (BM) χρησιμοποιούνται από τον ΔΣΜ για την προμήθεια αποθεματικών. Η εξισορρόπηση των αγορών μπορεί να χωριστεί σε αποθεματικά προμήθειες και χρηματοοικονομικός διακανονισμός ανισορροπιών. Η προμήθεια αποθεματικών είναι κράτηση ανταλλακτικών χωρητικότητα εκ των προτέρων. Αυτή η χωρητικότητα μπορεί, εάν χρειαστεί, να χρησιμοποιηθεί για εξισορρόπηση ζήτησης και παραγωγής. Ο χρηματοοικονομικός διακανονισμός των ανισορροπιών μπορεί να οριστεί ως κυρώσεις για τους συμμετέχοντες στην αγορά που προκαλούν ανισορροπία και οικονομικά βραβεία για τους συμμετέχοντες στην αγορά που χρησιμοποιούνται για τον περιορισμό της ανισορροπίας (π.χ., πάροχοι αποθεματικών).

Σύμφωνα με τους ισχύοντες κανονισμούς του συγγραφικού νόμου για την ηλεκτρική ενέργεια, υπάρχουν έξι τύποι οντοτήτων που μπορούν να γίνουν μέλη στο HUPX και είναι οι ακόλουθοι. Αρχικά, η πρώτη κατηγορία περιλαμβάνει του Διαπραγματευτές

ηλεκτρικής ενέργειας (Electricity Traders), Προμηθευτές καθολικής υπηρεσίας, Καταναλωτές, Παραγωγοί (με έγκυρη άδεια λειτουργίας μονάδας παραγωγής ενέργειας ή μικρής μονάδας ηλεκτροπαραγωγής), Διαχειριστής Συστήματος Μεταφοράς και τέλος ο Διαχειριστής Συστήματος Διανομής. Η συγκέντρωση της παραγωγής και της χονδρικής αγοράς είναι σχετικά υψηλή και οι κατεστημένοι φορείς όπως οι MVM, E.ON, RWE και NKM (εθνικές εταιρείες κοινής ωφέλειας) κυριαρχούν στην αγορά. Υπάρχουν περισσότεροι από 150 πιστοποιημένοι προμηθευτές στην αγορά, εκ των οποίων οι 30-40 είναι επί του παρόντος ενεργοί. Υπάρχουν περίπου 170 εξουσιοδοτημένοι έμποροι ηλεκτρικής ενέργειας στην ουγγρική αγορά ηλεκτρικής ενέργειας. Η πλειονότητα αυτών των εμπόρων είναι ξένοι ιδιωτικοί φορείς που ασκούν μόνο χονδρική δραστηριότητα εμπορίας ηλεκτρικής ενέργειας. Ωστόσο, ο Όμιλος MVM είναι επίσης σημαντικός παράγοντας στον τομέα της προμήθειας. Στον τομέα της παραγωγής, οι κύριοι παράγοντες της αγοράς είναι οι MVM Group (κρατικοί), MET Group, Opus Global, Alpiq και EP Energy. Στο παρακάτω διάγραμμα φαίνεται το μοντέλο της ηλεκτρικής αγοράς της Ουγγαρίας.



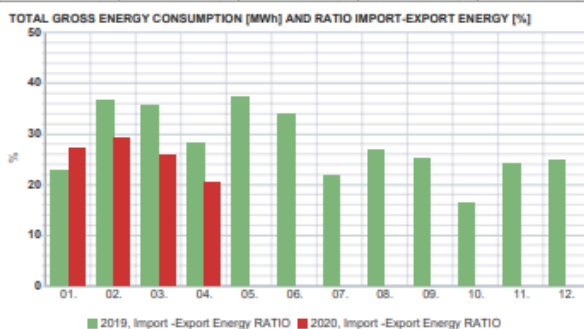
Εικόνα 2: Μοντέλο αγοράς ηλεκτρικής ενέργειας της Ουγγαρίας

Αναλύοντας το μείγμα της κατανάλωσης ηλεκτρικής ενέργειας για την χώρα της Ουγγαρίας, μεγάλο μέρος της ενέργειας εισάγεται από διασυνοριακές αλλά και μη διασυνοριακές χώρες και αναμένεται να ισχύει και για τα επόμενα χρόνια. Στον παρακάτω πίνακα παρουσιάζεται για το 2019 και 2020 η αναλογία εισαγωγών –

εξαγωγών με την κατανάλωση της χώρας. Το ποσοστό της κατανάλωσης που καλύπτουν οι ανταλλαγές ενέργειας είναι πάνω από 20% ενώ τον Μάιο του 2019 έφτασε σχεδόν το 40%.

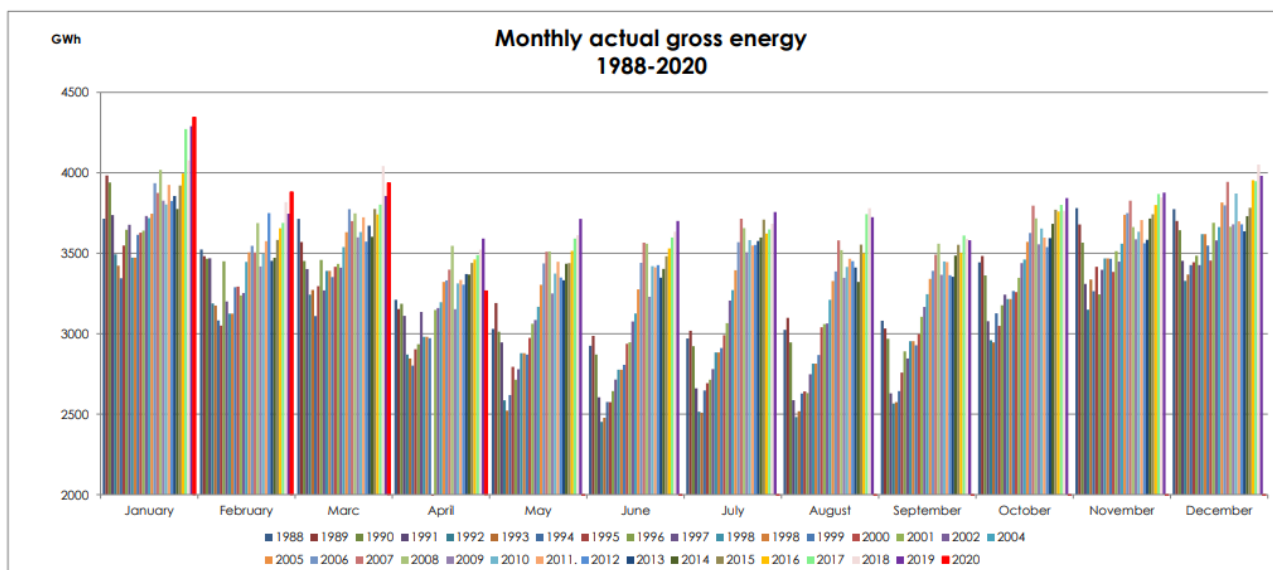
TOTAL GROSS ENERGY CONSUMPTION [MWh] AND RATIO IMPORT-EXPORT ENERGY [%]

Month	2019				2020			
	Import - Export Energy	Domestic Gross Energy	Total Gross Energy Consumption	Import -Export Energy RATIO	Import - Export Energy	Domestic Gross Energy	Total Gross Energy Consumption	Import -Export Energy RATIO
01.	971 857,13	3 318 131,70	4 289 988,83	22,85	1 176 284,20	3 168 952,85	4 345 237,04	27,07
02.	1 367 252,59	2 378 405,10	3 745 657,69	36,50	1 132 737,32	2 748 297,84	3 881 035,16	29,19
03.	1 368 603,00	2 487 333,91	3 855 936,92	35,49	1 017 351,61	2 919 832,22	3 937 183,83	25,84
04.	1 003 311,84	2 588 174,66	3 591 486,49	27,94	658 503,34	2 606 409,99	3 264 913,33	20,17
05.	1 385 540,36	2 328 262,83	3 713 803,19	37,31				
06.	1 249 606,53	2 451 049,00	3 700 655,52	33,77				
07.	807 644,03	2 948 920,57	3 756 564,61	21,50				
08.	1 001 313,77	2 722 911,14	3 724 224,90	26,89				
09.	896 476,32	2 684 731,16	3 581 207,49	25,03				
10.	623 845,02	3 218 611,18	3 842 456,20	16,24				
11.	932 649,13	2 944 538,42	3 877 187,54	24,05				
12.	977 250,86	3 004 085,49	3 981 336,34	24,55				
Yearly sum	12 585 350,57	33 075 155,15	45 660 505,71	27,56	3 984 876,47	11 443 492,89	15 428 369,36	25,83



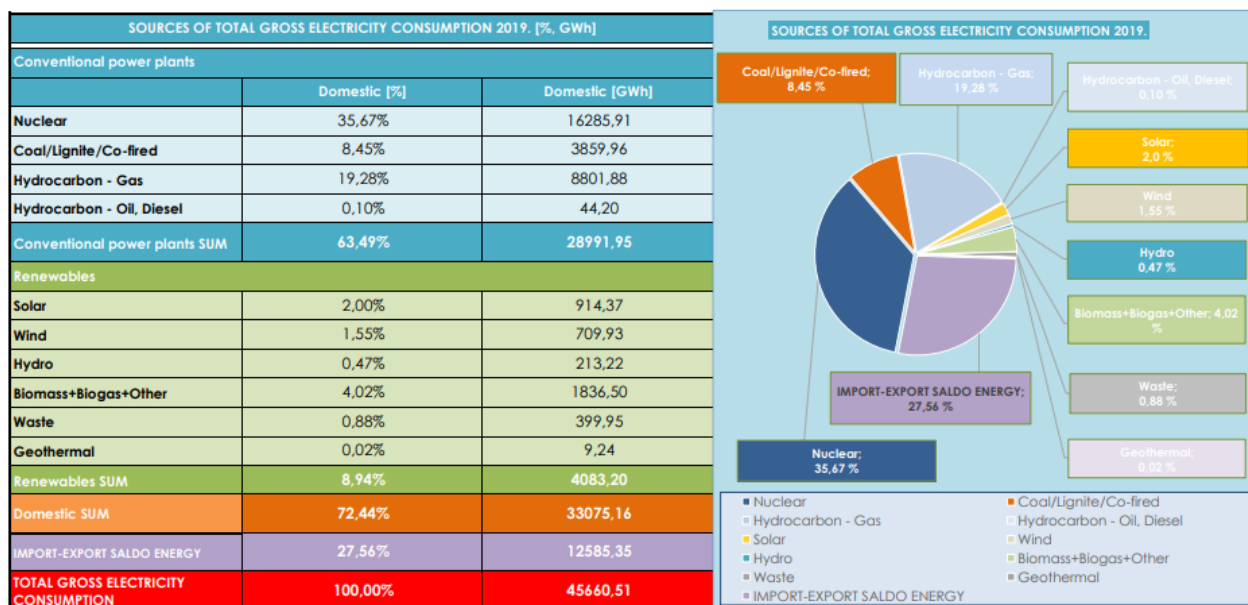
Εικόνα 3:Συνολική Κατανάλωση Ηλεκτρικής Ενέργειας[MWh] και ποσοστό Εισαγωγών-Εξαγωγών [%]

Σύμφωνα με σχετικές προβλέψεις, η συνολική κατανάλωση ενέργειας τα επόμενα χρόνια εκτιμάται ότι θα συνεχίσει να αυξάνεται. Στο παρακάτω, διάγραμμα φαίνεται η εξέλιξη της κατανάλωσης ηλεκτρικής ενέργειας από το 1998. Αν εξαιρεθεί ο Απρίλιος του 2020 που λόγω της πανδημίας η ζήτηση φορτίου μειώθηκε, φαίνεται πως τα τελευταία χρόνια και ιδιαίτερα το 2020 υπάρχει τάση για αυξανόμενη ζήτηση ενέργειας.



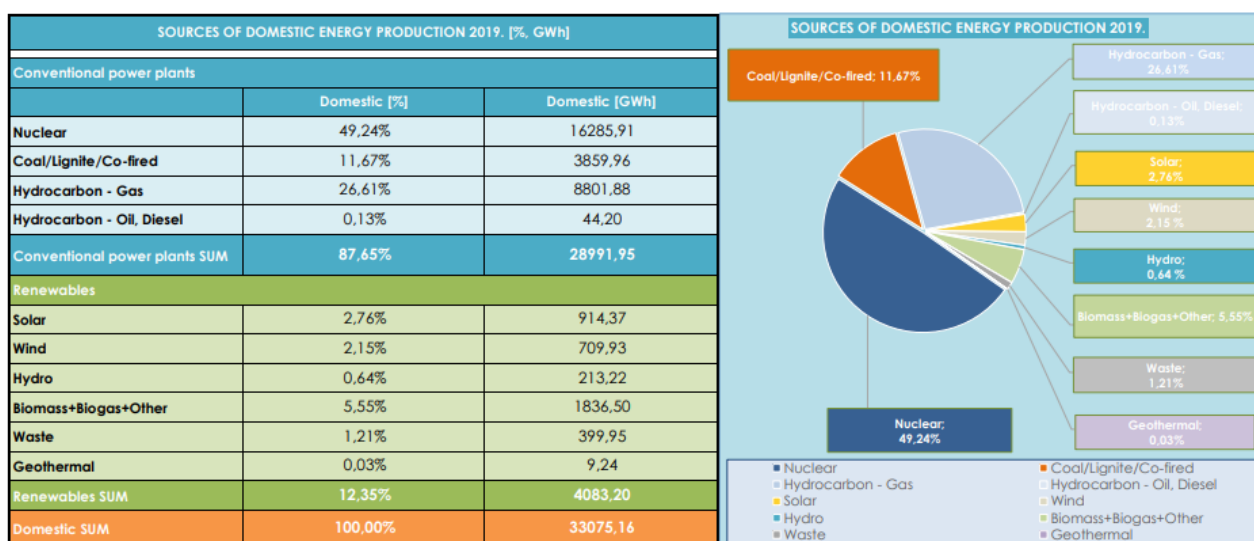
Εικόνα 4: Μηνιαία πραγματικά δεδομένα ενέργειας για το διάστημα 1998-2020

Στον παρακάτω πίνακα και διάγραμμα, φαίνεται το μείγμα ενέργειας της κατανάλωσης για την χώρα για το έτος 2019 όπως αυτά δημοσιεύτηκαν από τον MAVIR. Το μεγαλύτερο ποσοστό στην κατανάλωση ενέργειας κατέχουν τα πυρηνικά με ποσοστό 35,67%, έπειτα ακολουθεί το αέριο με ποσοστό 19,28% ενώ παρατηρείται συγχρόνως ότι οι Ανανεώσιμες πηγές ενέργειας δεν κατέχουν μεγάλο ποσοστό σε αντίθεση με το ισοζύγιο των Εισαγωγών – Εξαγωγών που κατέχουν ένα ποσοστό 27,56%.



Εικόνα 5: Μείγμα κατανάλωσης ηλεκτρικής ενέργειας Ουγγαρίας

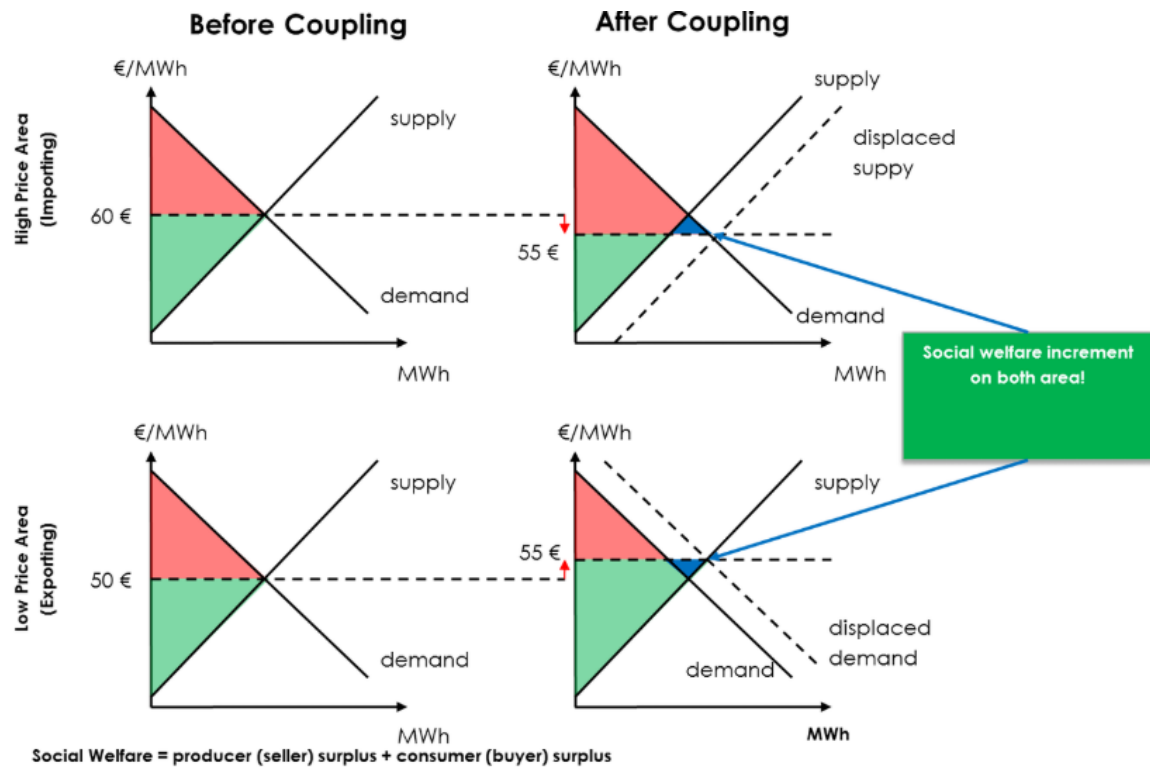
Αντίστοιχα με το μείγμα της κατανάλωσης, στην Ουγγαρία, η παραγωγή ηλεκτρικής ενέργειας βασίζεται κυρίως στην πυρηνική ενέργεια από τους Paks Πυρηνικούς σταθμούς παραγωγής ενέργειας που αντιπροσωπεύει σχεδόν το 50%. Επιπρόσθετα, συμπληρώνεται από παραγωγή από ορυκτά καύσιμα, φυσικό αέριο και άνθρακα. Η παραγωγή από ανανεώσιμες πηγές ενέργειας συνεχίζει να κατέχει μικρό ποσοστό στο μείγμα ενέργειας. Παρόλα αυτά, ο στόχος που έχει τεθεί είναι μέχρι το 2020 το 21% της παραγωγής να αποτελείται από ανανεώσιμες πηγές ενέργειας, εγκαθιστώντας κατά κύριο λόγο μαζικά φωτοβολταϊκή ενέργεια. Επιπλέον, προκειμένου να μειωθούν οι εκπομπές διοξειδίου του άνθρακα, έχουν ως πλάνο την αντικατάσταση το άνθρακα με αέριο διατηρώντας συγχρόνως την πυρηνική ενέργεια. Στην Ουγγαρία, η παραγωγή ηλεκτρικής ενέργειας από ανανεώσιμες πηγές υποστηρίζεται από ένα σύστημα τιμολόγησης τροφοδοσίας (FiT) για εγκαταστάσεις με εγκατεστημένη ισχύ 50-500 kW. Το σύστημα FiT αντικαταστάθηκε από το λεγόμενο σύστημα METÁR από την 1η Ιανουαρίου 2017, το οποίο περιλαμβάνει ένα νέο σύστημα τροφοδοσίας-τιμολόγησης (METÁR-FiT) έως εγκατεστημένη χωρητικότητα 0,5 MW, ένα σύστημα «πράσινο premium χωρίς διαγωνισμό» για εγκατεστημένο χωρητικότητα μεταξύ 0,5-1 MW και σύστημα «πράσινου ασφαλιστρού που παρέχεται μέσω διαγωνισμού» για εγκατεστημένη ισχύ άνω του 1 MW. Στο παλιό σύστημα FiT, η παραγόμενη ηλεκτρική ενέργεια πωλείται στον ΔΣΜ (TSO) σε σταθερή τιμή, ενώ στο νέο σύστημα METÁR η ηλεκτρική ενέργεια πωλείται απευθείας σε εμπόρους ή στο χρηματιστήριο με διόρθωση τιμών. Στο παρακάτω διάγραμμα φαίνεται επίσης το μείγμα της παραγωγής από τον MAVIR.



Εικόνα 6:Μείγμα παραγωγής ηλεκτρικής ενέργειας Ουγγαρίας

Αφού έγινε μια παρουσίαση της γενικής δομής της αγοράς της ηλεκτρικής ενέργειας για την Ουγγαρία, είναι ιδιαίτερα χρήσιμο να γίνει μια εκτενέστερη ανάλυση του μοντέλου της αγοράς της επόμενης ημέρας μέσω του οποίου θα φανεί και η αξία της πρόβλεψης των τιμών των αγορών. Η χονδρική αγορά ηλεκτρικής ενέργειας της Ουγγαρίας, που διαχειρίζεται από το HUPX αποτελεί μέρος του συστήματος Ζεύξης Αγοράς (Market Coupling MC) 4M μαζί με τις αγορές της Τσεχικής Δημοκρατίας, της Σλοβακίας και της Ρουμανίας, η οποία ήταν η τελευταία που ενσωματώθηκε στο σύστημα αυτό το 2014. Το 4M MC αντιπροσωπεύει τη σύνδεση μεταξύ τεσσάρων αγορών ως ενιαία αγορά και χρησιμοποιεί τον ίδιο αλγόριθμο EUPHEMIA όπως οι συζευγμένες αγορές της Δυτικής Ευρώπης για την ταυτόχρονη αντιστοίχιση των τεσσάρων αγορών με ρητή κατανομή της χωρητικότητας διασύνδεσης. Το γεγονός ότι υπάρχει ταυτόχρονη αντιστοίχιση σε αυτές τις τέσσερις αγορές μαζί με τη διαθέσιμη ικανότητα διασύνδεσης προκάλεσε τις τιμές τους να εξελιχθούν αρκετά παράλληλα. Και όχι μόνο η σύνδεση των τεσσάρων αγορών επηρέασε τις τιμές, αλλά η παρουσία μιας σημαντικής αγοράς όπως η γερμανική στα σύνορα της 4M MC είχε ακόμη μεγαλύτερη επίδραση στις τιμές τους. Η χονδρική αγορά ηλεκτρικής ενέργειας της Ουγγαρίας είναι η μία από τις τέσσερις αγορές σε συνδυασμό με 4M MC που είχαν ιστορικά υψηλότερες τιμές. Παρά τις ιστορικές διαφορές τιμών, μπορεί πλέον να φανεί η σαφής σχέση μεταξύ αυτών των τεσσάρων αγορών.

Ο πρωταρχικός στόχος της σύζευξης της αγοράς είναι να εξασφαλίσει μια βαθύτερη ολοκλήρωση των περιφερειακών αγορών ενέργειας. Η διαδικασία στοχεύει στη μεγιστοποίηση της ροής ενέργειας από την περιοχή χαμηλών τιμών στην περιοχή υψηλών τιμών λαμβάνοντας υπόψη την διαθέσιμη διασυνοριακή χωρητικότητα. Τα επίπεδα τιμών μεταξύ των επιμέρους περιοχών της αγοράς συγκλίνουν. Με αυτήν τη διαδικασία, όχι μόνο η ασφάλεια του εφοδιασμού μπορεί να αυξηθεί, αλλά η μεταβλητότητα των τιμών στις σχετικές αγορές ηλεκτρικής ενέργειας μπορεί να μετριαστεί και η πώληση και η αγορά ισχύος να γίνουν πιο αποτελεσματικές. Στο παρακάτω διαγράμμα φαίνεται η ισορροπία που επέρχεται σε δύο αγορές μετά την ζεύξη των αγορών.



Εικόνα 7: Διάγραμμα ζήτησης - προσφοράς πριν και μετά την ζεύξη αγορών

Η εισαγωγή της ζεύξης αγοράς δεν άλλαξε τη μέθοδο κατανομής δυναμικότητας. Για τις διασυνοριακές συνδέσεις, ο εκάστοτε Διαχειριστής Συστήματος υπολογίζει την διαθέσιμη χωρητικότητα μεταξύ των χωρών αλλά αντί να δημοπρατεί ρητά τη διασυνοριακή ικανότητα μεταφοράς μεταξύ των συμμετεχόντων στην αγορά, η σύζευξη της αγοράς καθιστά τις δυνατότητες έμμεσα διαθέσιμες στις ανταλλαγές ισχύος των διαφόρων χρηματιστηρίων. Οι συμμετέχοντες υποβάλουν ζεύγη τιμών και προσφορές παραγωγής ή ζήτησης που αποτελούν τη βάση για τον σχηματισμό της τιμής εκκαθάρισης της αγοράς.

Η πρόβλεψη τιμών για την επόμενη μέρα είναι ένα πολύτιμο εργαλείο για όλους τους συμμετέχοντες της αγοράς καθώς καθορίζει τη λήψη αποφάσεων για την λειτουργία ενός σταθμού παραγωγής ηλεκτρικής ενέργειας και την βελτιστοποίηση του χαρτοφυλακίου στην αγορά. Μέσω μιας ακριβούς πρόβλεψης, ο παραγωγός έχει την δυνατότητα να προσφέρει την ενέργεια σε όσο το δυνατόν υψηλότερη τιμή ώστε να αυξήσει τα κέρδη του, ενώ ο έμπορος ή προμηθευτής έχει την δυνατότητα να αγοράσει ενέργεια σε χαμηλή τιμή.

1.4 Οργάνωση κειμένου

Στο παρόν κεφάλαιο της εργασίας έγινε μια αρχική επισκόπηση της αγοράς της ηλεκτρικής ενέργειας και των χαρακτηριστικών. Στην συνέχεια, αναλύθηκε εκτενέστερα η αγορά της Ουγγαρίας και πώς αυτή λειτουργεί στην συγκεκριμένη χρονική στιγμή.

Στο δεύτερο κεφάλαιο της παρούσας εργασίας πραγματοποιείται μια βιβλιογραφική επισκόπηση σχετικά με την πρόβλεψη της τιμής της ηλεκτρικής ενέργειας. Γίνεται, αρχικά κατηγοριοποίηση της πρόβλεψης βάσει του ορίζοντα καθώς και αναφορά στις μεθόδους που χρησιμοποιούνται βάσει της βιβλιογραφίας και αντίστοιχων εργασιών που έχουν πραγματοποιηθεί. Εν συνεχεία, παρουσιάζεται η επίδραση των γειτονικών χωρών στην πρόβλεψη της τιμής της ενέργειας για την υπό μελέτη αγορά όπως αυτή έχει αναλυθεί σε αντίστοιχες έρευνες. Στο τέλος του κεφαλαίου αναφέρονται οι παράγοντες που επιδρούν στην διαμόρφωση της τιμής της ηλεκτρικής ενέργειας, επομένως οι παράγοντες που πρέπει να ληφθούν υπόψη στην πρόβλεψη της τιμής.

Στο κεφάλαιο που ακολουθεί έπειτα από την βιβλιογραφική επισκόπηση σχετικά με την πρόβλεψη της τιμής των αγορών, γίνεται εκτενής αναφορά στις τεχνικές προβλέψεων, στις κατηγορίες μεθόδων πρόβλεψης και στα χαρακτηριστικά των χρονοσειρών που χρησιμοποιούνται σε αυτές. Παρουσιάζονται αρχικά τα βασικά μοντέλα που χρησιμοποιούνται στην στατιστική πρόβλεψη ενώ στην συνέχεια γίνεται αναφορά στην μηχανική μάθηση όπως αυτή χρησιμοποιείται στον τομέα των προβλέψεων. Ιδιαίτερα, παρουσιάζονται οι αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν στην συγκεκριμένη έρευνα. Το τέλος του κεφαλαίου ασχολείται με τα στατιστικά σφάλματα των μεθόδων, τον τρόπο με τον οποίο μπορούμε να τα εκμεταλλευτούμε για την αξιολόγηση των μεθόδων πρόβλεψης και τα πλεονεκτήματα που το καθένα μας προσφέρει.

Το τέταρτο κεφάλαιο αποτελεί case-study της εργασίας, παρουσιάζει δηλαδή την πρακτική εφαρμογή των αλγορίθμων και των μεθόδων που επιλέχθηκαν στην συγκεκριμένη εργασία. Αρχικά, γίνεται μια ανάλυση και παρουσίαση των δεδομένων που μελετήθηκαν και αναλύθηκαν και στην συνέχεια παρουσιάζονται τα

αποτελέσματα κάθε σεναρίου που υλοποιήθηκε. Στόχος είναι η ανάδειξη της πιο αξιόπιστης μεθόδου πρόβλεψης παρατηρώντας της αποκλίσεις των προβλέψεων από τα πραγματικά δεδομένα και συνεπώς η αξιολόγηση υπό ρεαλιστικές συνθήκες.

Στο τελευταίο κεφάλαιο της εργασίας, συνοψίζονται τα αποτελέσματα και εξάγονται τα κύρια συμπεράσματα βάσει των αποτελεσμάτων από τα πειράματα που υλοποιήθηκαν. Στην συνέχεια προτείνονται θέματα για μελέτη και ανάλυση στο μέλλον.

2

Πρόβλεψη τιμής ηλεκτρικής ενέργειας

Σε οποιασδήποτε εταιρεία ηλεκτρικής ενέργειας για τη λήψη αποφάσεων και την ανάπτυξη στρατηγικής, οι προβλέψεις της τιμής της ηλεκτρικής ενέργειας έχουν καταστεί απαραίτητης συμβολής.

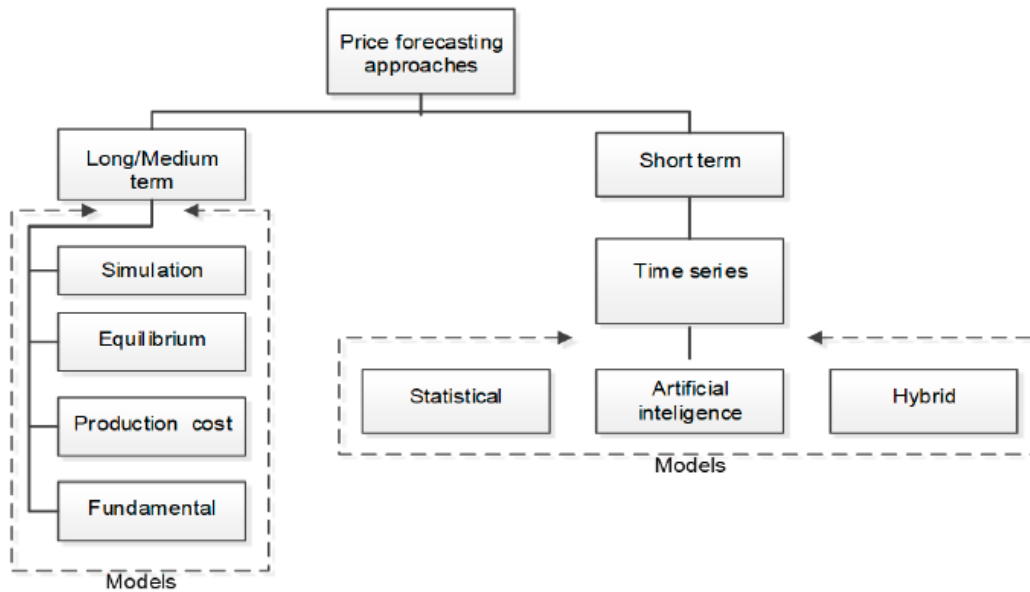
Η έρευνα σχετικά με την πρόβλεψη τιμών ηλεκτρικής ενέργειας μπορεί να ταξινομηθεί με βάση την μεθοδολογία που εφαρμόζεται καθώς και βάσει της διάρκειας του ορίζοντα πρόβλεψης. Αναφορικά με τον ορίζοντα πρόβλεψης μπορεί να γίνει διαχωρισμός στις ακόλουθες τρεις κατηγορίες: α) βραχυπρόθεσμες προβλέψεις, β) μεσοπρόθεσμες προσβλέψεις και γ) μακροπρόθεσμες προβλέψεις.

Στις μακροπρόθεσμες προβλέψεις ο ορίζοντας συνήθως ορίζεται σε χρόνια και κύριος στόχος τους είναι ο προγραμματισμός και η ανάλυση της κερδοφορίας επενδύσεων. Τέτοιες επενδύσεις θα μπορούσαν να είναι η σύναψη κάποιου ετήσιου οικονομικού ή φυσικού συμβολαίου ηλεκτρικής ενέργειας για μια εταιρεία που εμπορεύεται ηλεκτρική ενέργεια ή λήψη αποφάσεων για μελλοντικές επενδύσεις σταθμών παραγωγής ενέργειας για μια εταιρεία που συμμετέχει στην αγορά ως παραγωγός. Αποτελεί ένα πολύ βασικό εργαλείο για τις εταιρείες ηλεκτρικής ενέργειας

προκειμένου να οργανώσουν την στρατηγική τους σχετικά με την παραγωγή της ηλεκτρικής ενέργειας, την τιμολογιακή στρατηγική τους αν πρόκειται για προμηθευτή ηλεκτρικής ενέργειας αλλά και την στρατηγική των μακροπρόθεσμων συμβάσεων και κινήσεων αν πρόκειται για εταιρεία εμπορίας ηλεκτρικής ενέργειας.

Στις μεσοπρόθεσμες προβλέψεις ο χρονικός ορίζοντας συνήθως μετριέται σε μήνες και επικεντρώνεται στους υπολογισμούς του ισολογισμού, στην διαχείριση κινδύνων και στην τιμολόγηση παραγώγων ενέργειας. Για παράδειγμα, μια εταιρεία που εμπορεύεται ηλεκτρική ενέργεια μέσω μια μεσοπρόθεσμης πρόβλεψης για την τιμή της αγοράς τους επόμενους μήνες θα μπορούσε να συνάψει ένα οικονομικό συμβόλαιο ηλεκτρικής ενέργειας για κάποιον από τους επόμενους μήνες ή για κάποιο από τα επόμενα τρίμηνα. Επιπλέον, μέσω των μεσοπρόθεσμων προβλέψεων είναι σε θέση να ορίσει την στρατηγική της σχετικά με τις μεσοπρόθεσμες δημοπρασίες δικαιώματος φυσικής παράδοσης ηλεκτρικής ενέργειας. Μέσω των δημοπρασιών αυτών δίνεται το δικαίωμα σε μία εταιρεία να εισάγει και εξάγει ενέργεια μεταξύ χωρών που υπάρχει η διασύνδεση. Παραδείγματος χάριν, μια εταιρεία εμπορίας θα ήθελε να λάβει δικαίωμα για εισαγωγή ηλεκτρικής ενέργειας για τον επόμενο μήνα από την Αυστρία στην Ουγγαρία αν γνώριζε ότι η τιμή της Αυστρίας θα είναι κάποιες μονάδες πιο κάτω από την αντίστοιχη της Ουγγαρίας. Με αυτόν τον τρόπο θα αγόραζε φθηνή ενέργεια και θα την πουλούσε σε μια πιο ακριβή αγορά. Κατά αντίστοιχο τρόπο λειτουργούν και τα οικονομικά συμβόλαια χωρίς όμως να υπάρχει φυσική παράδοση ηλεκτρικής ενέργειας.

Τα μοντέλα που εφαρμόζονται για μακροπρόθεσμη και μεσοπρόθεσμη πρόβλεψη μπορούν να ταξινομηθούν όπως φαίνονται στο ακόλουθο σχήμα.



Εικόνα 8: Κατηγοριοποίηση προσεγγίσεων για πρόβλεψη τιμής ηλεκτρικής ενέργειας

Τα μοντέλα προσομοίωσης είναι δυνατό εργαλείο για αυτού του τύπου προβλέψεις διότι δημιουργούν διάφορα μελλοντικά σενάρια και υπολογίζουν την πιθανότητα των σεναρίων της τιμής βασισμένα στον αντίστοιχο κίνδυνο. Λαμβάνοντας υπόψη την ισορροπία που πρέπει να διατηρείται σε μια αγορά τα μοντέλα ισορροπίας δημιουργούν σενάρια τιμών βασισμένα στην προϋπόθεση αυτή. Από την άλλη πλευρά, τα μοντέλα κόστους παραγωγής προσομοιώνουν την λειτουργία των μονάδων παραγωγής για να ικανοποιήσουν τη ζήτηση του συστήματος στο ελάχιστο κόστος. Τέλος, οι θεμελιώδεις μέθοδοι καθορίζουν τις μεταβολές των τιμών με μοντελοποίηση του αντίκτυπου σημαντικών οικονομικών και φυσικών παραγόντων στην τιμή της ηλεκτρικής ενέργειας.

Η τρίτη κατηγορία προβλέψεων που αναφέρθηκε αφορά τις βραχυπρόθεσμες προβλέψεις τιμής ηλεκτρικής ενέργειας. Στην κατηγορία αυτή ο ορίζοντας μπορεί να είναι λεπτά μπροστά, ή κάποιες μέρες. Είναι ιδιαίτερα σημαντικές για τις μονάδες παραγωγής από την μεριά των παραγωγών αλλά και για τις εταιρείες εμπορίας ηλεκτρικής ενέργειας καθώς και οι δύο σε καθημερινές δημοπρασίες υποβάλλουν προσφορές για ηλεκτρική ενέργεια ως ζεύγος τιμής με ποσότητα. Οι προσφορές γίνονται δεκτές κατά αύξουσα ή φθίνουσα σειρά έως ότου ικανοποιηθεί η συνολική ζήτηση και προσφορά. Σε μια μονάδα παραγωγής αν οι προβλέψεις της τιμής είναι ακριβείς, δίνεται η δυνατότητα να προσαρμόσει το πρόγραμμα παραγωγής και

επομένως να μεγιστοποιήσει τα κέρδη. Από την άλλη για έναν έμπορο ηλεκτρικής ενέργειας, όσο πιο ακριβείς είναι οι προβλέψεις των τιμών μπορεί να υποβάλει προσφορές που θα μεγιστοποιήσουν αντίστοιχα τα κέρδη του. Η ημερήσια αγορά ενέργειας αποτελείται συνήθως από 24 ωριαίες δημοπρασίες ή 96 15λεπτες που πραγματοποιούνται την προηγούμενη μέρα από την ημέρα παράδοσης της ενέργειας. Κατ' αντιστοιχία με τις μεσοπρόθεσμες προβλέψεις, οι έμποροι ηλεκτρικής ενέργειας χρειάζονται ιδιαίτερα τις βραχυπρόθεσμες προβλέψεις προκειμένου να οργανώσουν την καθημερινή τους στρατηγική και να είναι σε θέση να συμμετάσχουν στις καθημερινές δημοπρασίες κατά τον βέλτιστο δυνατό τρόπο. Μια ανακριβής πρόβλεψη, δηλαδή μια υπερεκτίμηση ή μια υποεκτίμηση της υπό μελέτη αγοράς μπορεί να οδηγήσει σε μια λάθος κίνηση εισαγωγής ή εξαγωγής ενέργειας που θα επιφέρει τελικά ζημιά για την εταιρεία. Επομένως, οι βραχυπρόθεσμες προβλέψεις σε ωριαίο επίπεδο για μία μέρα ή κάποιες μέρες μπροστά είναι ιδιαίτερα σημαντικές για τους συμμετέχοντες στην αγορά της ηλεκτρικής ενέργειας. Σε αυτή την κατηγορία προβλέψεων όπως φαίνεται και στο σχήμα παραπάνω χρησιμοποιούνται κυρίως στατιστικά μοντέλα πρόβλεψης που περιλαμβάνουν χρονοσειρές καθώς και μοντέλα μηχανικής μάθησης που προσφέρουν την δυνατότητα να εισαχθούν διάφορες μεταβλητές εισόδου που επηρεάζουν το υπό μελέτη μέγεθος. Επιπλέον, μπορεί να υλοποιηθεί και να εφαρμοστεί ένα υβριδικό μοντέλο που θα συνδυάζει στατιστική πρόβλεψη με κάποιον αλγόριθμο μηχανικής μάθησης.

Τυπικές στατιστικές μέθοδοι που εφαρμόζονται και θα αναλυθούν και εκτενέστερα στο επόμενο κεφάλαιο είναι τα μοντέλα ARIMA, γραμμική ή εκθετική τάση, εκθετική εξομάλυνση, Holt, μοντέλο Holt-Winter καθώς και συνδυασμός αυτών^{[5],[6],[8]}. Αυτοπαλινδρομικά μοντέλα (ARIMA) και μη στάσιμα μοντέλα χρονοσειρών όπως η δυναμική παλινδρόμηση (DR) χρησιμοποιούνται συνήθως για την στατιστική πρόβλεψη χρονοσειρών. Επιπλέον, έχουν προταθεί σε μελέτες εκτεταμένα μοντέλα ARIMA και συνδυασμός τους με άλλες στατιστικές μεθόδους πρόβλεψης για την εκτίμηση των τιμών στις αγορές της ηλεκτρικής ενέργειας της Ευρώπης, της τιμής αγοράς του Ευρωπαϊκού Χρηματιστηρίου Ενέργειας, της αγοράς της Ισπανίας, του Ηνωμένου Βασιλείου και της Καλιφόρνιας^[8]. Τα στατιστικά μοντέλα πρόβλεψης εφαρμόζονται αφού αφαιρεθεί ο εποχιακός χαρακτήρας που παρουσιάζει η τιμή της ηλεκτρικής ενέργειας. Ο εποχιακός χαρακτήρας μπορεί να

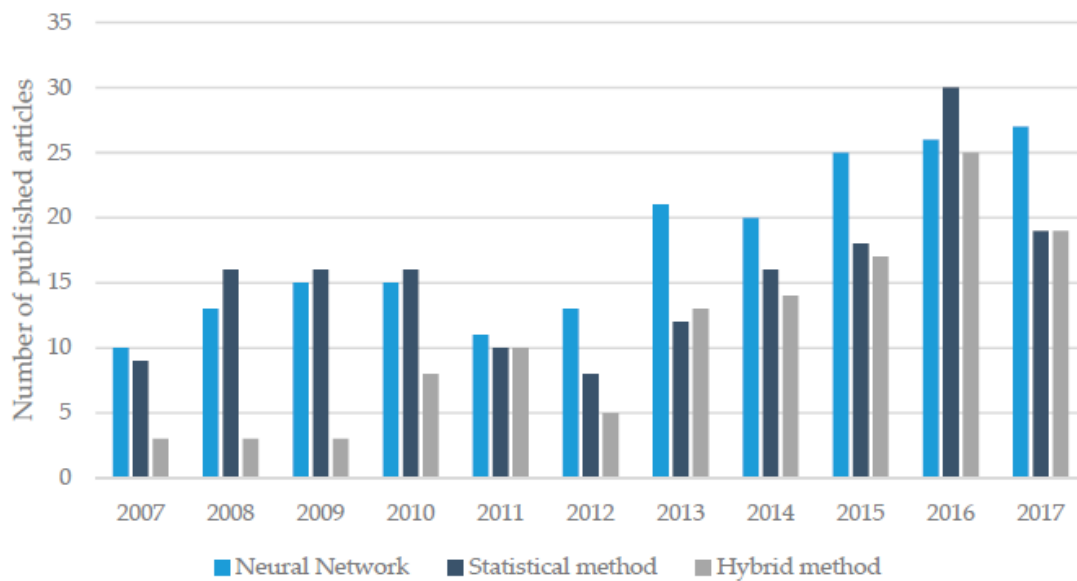
είναι σε ωριαίο επίπεδο, σε ημερήσιο ή και σε εβδομαδιαίο ανάλογα τον χρονικό ορίζοντα της πρόβλεψης.

Τα στατιστικά μοντέλα συνήθως παράγουν επιτυχείς προβλέψεις σε δεδομένα με χαμηλή συχνότητα π.χ. εβδομαδιαία, ωστόσο οι ωριαίες προβλέψεις που εμφανίζουν μη γραμμική συμπεριφορά ως προς τις παραμέτρους εισόδου είναι πιο εύκολο να προβλεφθούν με τις μεθόδους μηχανικής μάθησης. Τέτοιες μέθοδοι συνήθως είναι τα νευρωνικά δίκτυα τα οποία υλοποιούνται μέσω διάφορων αλγορίθμων π.χ Neural networks (NN) ή πολυεπίπεδοι αισθητήρες Multilayer Perceptrons (MLP), αλγόριθμοι που εφαρμόζουν μεθόδους με δέντρα αποφάσεων όπως είναι τα απλά δέντρα αποφάσεων (Decision trees), δάση από δέντρα αποφάσεων (Random forest και Gradient Boosting Machine) , γραμμική παλινδρόμηση απλή ή πολλαπλών εισόδων, υποστηριζόμενοι παλινδρομικοί δείκτες (Support Vector Regressors), ακτινικής βάσης δίκτυα (radial basis function networks), ασαφή δίκτυα (fuzzy networks) καθώς και νευρωνικά βαθιάς αρχιτεκτονικής (Deep Neural Networks)^{[5],[6],[8]}. Ιδιαίτερα, έχουν προταθεί μοντέλα με βάση τα νευρωνικά δίκτυα (NN) και τα ασαφή νευρωνικά δίκτυα (FNN) για την πρόβλεψη τιμών στην ηλεκτρική ενέργεια. Τα μοντέλα αυτά έχουν δοκιμαστεί σε δεδομένα της Ισπανικής αγοράς και της αγοράς της Πενσυλβανίας New Jersey Maryland. Οι υποστηριζόμενοι παλινδρομικοί δείκτες (Support Vector Regressors) και τα τεχνητά νευρωνικά δίκτυα (ANN) έχουν δοκιμαστεί επίσης σε δεδομένα της αγοράς της Πενσυλβανίας^[8].

Οι υβριδικές μέθοδοι συνδυάζουν τόσο γραμμικές όσο και μη γραμμικές στρατηγικές μοντελοποίησης. Για τη βελτίωση των προβλέψεων, οι υβριδικές μέθοδοι χρησιμοποιούν τα συγκεκριμένα χαρακτηριστικά κάθε μοντέλου για να ανιχνεύουν διαφορετικά μοτίβα στα δεδομένα και ταυτόχρονα να αποδίδουν ένα πιο ακριβές και αξιόπιστο αποτέλεσμα. Τα νευρωνικά δίκτυα έχουν συνδυαστεί με τα αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου σε δεδομένα της αγοράς της Ισπανίας ενώ ανάπτυξη και σύγκριση των μοντέλων αυτών έχει πραγματοποιηθεί και στην αυστραλιανή εθνική αγορά ηλεκτρικής ενέργειας^[8].

Στο παρακάτω διάγραμμα φαίνεται μια ποσοτική απεικόνιση των μεθόδων που χρησιμοποιήθηκαν την δεκαετία 2007-2017 για την πρόβλεψη τιμή ηλεκτρικής

ενέργειας στα διάφορα είδη των αγορών. Από το παρακάτω διάγραμμα φαίνεται ότι αυξάνονται σημαντικά οι μελέτες στον τομέα της πρόβλεψης της τιμής ηλεκτρικής ενέργειας ιδιαίτερα μετά την έναρξη της απελευθέρωσης των αγορών. Επιπλέον, είναι ιδιαίτερα εμφανής η υπεροχή των νευρωνικών δικτύων έναντι των στατιστικών και των υβριδικών το 2017 η οποία αποτελεί τάση που αναπτύσσεται με το πέρασμα των χρόνων.



Εικόνα 9: Κατανομή επιστημονικών άρθρων σύμφωνα με τη μέθοδο που χρησιμοποιείται για την πρόβλεψη τιμής ηλεκτρικής ενέργειας^[8]

Πιο συγκεκριμένα, σε πρόσφατο άρθρο οι Jesus Lagoa,b, Fjo De Ridder, Bart De Schutter^[6], προτείνουν τέσσερα διαφορετικά μοντέλα βαθιάς μηχανικήςμάθησης για την πρόβλεψη της ηλεκτρικής ενέργειας τιμές και δείχνουν πώς οδηγούν σε βελτιώσεις στην προγνωστική ακρίβεια. Για την ακρίβεια δείχνουν ότι: (i) οι μέθοδοι μηχανικής μάθησης αποδίδουν, γενικά, καλύτερη ακρίβεια από τα στατιστικά μοντέλα. (ii) οι όροι κινητού μέσου όρου δεν βελτιώνουν την προγνωστική ακρίβεια. (iii) τα υβριδικά μοντέλα δεν ξεπερνούν τα μοντέλα τα απλούστερα αντίστοιχα.

Από την άλλη πλευρά, οι Marin Cerjan , Ana Petrićić and Marko Delimar^[8], εισαγάγουν μια βραχυπρόθεσμη μέθοδο πρόβλεψης τιμών ηλεκτρικής ενέργειας που

αντιμετωπίζει τα προβλήματα της μεταβλητότητας των τιμών. Προτείνουν ένα υβριδικό μοντέλο που αποτελείται από δύο φάσεις. Στην φάση της ανάλυσης, οι θεμελιώδεις παράμετροι που επηρεάζουν την τιμή του ηλεκτρικού ρεύματος προσδιορίζονται ανάλογα με την αγορά που μελετάται. Οι παράμετροι αυτοί χρησιμοποιούνται στο υβριδικό μοντέλο. Το υβριδικό μοντέλο συνδυάζει μια στατιστική προσέγγιση για ανάλυση μεγάλων συνόλων δεδομένων και μέθοδο παρόμοιας ημέρας με εργαλεία νευρικού δικτύου.

Τέλος, οι Aleksandra Dedinec και Aleksandar Dedinec^[10], προβλέπουν 24 ώρες μπροστά της τιμής ηλεκτρικής ενέργειας χρησιμοποιώντας μοντέλα τεχνητής νοημοσύνης. Συγκεκριμένα εξετάζονται νευρωνικά δίκτυα, τα οποία ως είσοδο χρησιμοποιούν όλες τις πληροφορίες που σχετίζονται με την αντίστοιχη πρόβλεψη τιμής ανταλλαγής ισχύος. Αντίστοιχα, οι Jesus Lagoa., Fjo De Ridder, Peter Vrancx, Bart De Schutter^[5], αναλύουν την επίδραση των γειτονικών χωρών στην υπό μελέτη χώρα μέσω των αλγορίθμων μηχανικής μάθησης προτείνοντας ένα βαθύ νευρωνικό δίκτυο.

Στις διάφορες έρευνες που έχουν γίνει, από τα αποτελέσματα που συγκρίνουν την ακρίβεια των παραπάνω μοντέλων προκύπτουν ασαφή αποτελέσματα. Σε γενικές γραμμές, η αποτελεσματικότητα κάθε μοντέλου εξαρτάται από την αγορά και την χρονική περίοδο που εξετάζεται.

2.1 Επίδρασης των γειτονικών χωρών στην τιμή της ηλεκτρικής ενέργειας

Τις τελευταίες δεκαετίες, η Ευρωπαϊκή Ένωση έχει εγκρίνει αρκετούς νόμους που προσπαθούν να επιτύχουν μια ενιαία και ολοκληρωμένη ευρωπαϊκή αγορά ηλεκτρικής ενέργειας. Την δεδομένη χρονική στιγμή, η δημιουργία μιας ενιαίας αγοράς απέχει πολύ από την υπάρχουσα κατάσταση, όμως υπάρχουν στοιχεία που υποδηλώνουν ότι το επίπεδο ολοκλήρωσης μεταξύ διάφορων περιφερειακών αγορών αυξάνεται με την πάροδο του χρόνου. Επιπλέον, αξιοποιώντας στοιχεία των γειτονικών χωρών και των χωρών που γίνονται εισαγωγές και εξαγωγές ενέργειας σε

μεγάλο βαθμό δίνεται η δυνατότητα να προσομοιωθεί η συμπεριφορά των εμπορών που θα έχει επίδραση στην τιμή της αγοράς. Για παράδειγμα, αναμένεται να αυξηθεί η ζήτηση στην περίπτωση εξαγωγής από την υπό εξέταση χώρα σε άλλη χώρα ή αναμένεται να καλυφθεί μέρος του φορτίου οπότε και να μειωθεί η ζήτηση στην περίπτωση των εισαγωγών από άλλες χώρες.

Σε αντίστοιχες μελέτες που έχουν ήδη λάβει χώρα φαίνεται για παράδειγμα ότι στην περίπτωση του Βελγίου και της Γαλλίας, οι τιμές της αγοράς μοιράζονται ισχυρή κοινή δυναμική^[5]. Παρά το γεγονός ότι σύμφωνα με την βιβλιογραφία έχει γίνει αξιολόγηση σχετικά με το επίπεδο ολοκλήρωσης των ευρωπαϊκών χωρών και κάποιοι έχουν προτείνει στατιστικά μοντέλα για την αξιολόγηση της επίδρασης μεταξύ των χωρών της Ευρωπαϊκής ένωσης, οι έρευνες σχετικά με την βελτίωση της ακρίβειας λαμβάνοντας υπόψη στοιχεία των γειτονικών χωρών είναι αρκετά σπάνιες.

Συγκεκριμένα, οι αναλύσεις που έχουν ήδη λάβει χώρα αναλύουν την επίδραση της τιμής της αγοράς κάποιας γειτονικής χώρας ή κάποιας χώρας με την οποία λαμβάνουν χώρα μεγάλες ανταλλαγές ενέργειας σε καθημερινό επίπεδο, στην πρόβλεψη της τιμής της υπό μελέτης αγοράς. Το παραπάνω για να εφαρμοστεί σε ρεαλιστικές συνθήκες συνεπάγεται ότι η τιμή της αγοράς της γειτονικής χώρας είναι γνωστή και διαθέσιμη πριν από την τιμή της υπό μελέτης αγοράς. Επιπλέον, για κάποιο έμπορο ενέργειας που ενδιαφέρεται να αποκτήσει φυσικό δικαίωμα μεταφοράς ενέργειας στις καθημερινές δημοπρασίες ενέργειας που λαμβάνουν χώρα το πρωί όπως έχει ήδη αναλυθεί ένα τέτοιο μοντέλο δεν θα ήταν χρήσιμο. Οι δημοπρασίες αυτές προφανώς κλείνουν πριν γίνει γνωστή η τιμή καμίας αγοράς επομένως θα πρέπει να αποφασίσει την στρατηγική του βάσει κάποιου μοντέλου που θα έχει κάποια άλλη μεταβλητή εισόδου και όχι την τιμή των γειτονικών χωρών. Θα μπορούσε κάποιος να χρησιμοποιήσει πρόβλεψη για την τιμή της γειτονικής χώρας αν θέλει να την εισάγει την τιμή στο μοντέλο με αυτό τον τρόπο. Αυτό όμως θα είχε ως αποτέλεσμα το σφάλμα του τελικού μοντέλου για την υπό μελέτη χώρα να αυξάνεται ακόμα και να πολλαπλασιάζεται. Η πρόβλεψη της γειτονικής χώρας από μόνη της στηρίζεται σε κάποιες προβλέψεις επομένως ένα σφάλμα στο μοντέλο αυτό θα οδηγήσει σε μεγαλύτερο σφάλμα στο τελικό μοντέλο που αφορά την υπό εξέταση χώρα.

Επομένως, στην παρούσα εργασία δεδομένου ότι θεωρείται σημαντική η επίδραση των γειτονικών χωρών και των χωρών με έντονη ανταλλαγή ηλεκτρικής ενέργειας προτείνεται η χρήση των ίδιων δεδομένων που θα αναλυθούν στην συνέχεια τόσο για την υπό μελέτη χώρα όσο και για τις υπόλοιπες. Αυτό συνεπάγεται ένα μοντέλο που μπορεί να εφαρμοστεί και να αποδώσει σε ρεαλιστικές συνθήκες.

2.2 Παράγοντες που επιδρούν στην τιμή της ηλεκτρικής ενέργειας

Κατά την υλοποίηση ενός μοντέλου πρόβλεψης σημαντικό στάδιο είναι αυτό της επιλογής των μεταβλητών που επιδρούν στο υπό μελέτη μέγεθος.

Αρχικά, σημαντική επίδραση στην διαμόρφωση της τιμής της επόμενης μέρας έχει το φορτίο που θα απαιτηθεί από το σύστημα την επόμενη μέρα. Το φορτίο είναι ίσως ο κυριότερος παράγοντας διότι όταν ο διαχειριστής καλείται στο κλείσιμο της αγοράς να λύσει τον αλγόριθμο στόχος είναι να καλυφθεί και να έρθει σε ισορροπία η ζήτηση δηλαδή το φορτίο και οι εξαγωγές από την υπό μελέτη χώρα με την διαθέσιμη παραγωγή οποιασδήποτε μορφής και την αντίστοιχες εισαγωγές που θα γίνουν. Από την μεριά της ζήτησης ο μεγαλύτερος όγκος αφορά το φορτίο της χώρας, ενώ οι εισαγωγές και οι εξαγωγές μπορούν να προσομοιωθούν έμμεσα από τα αντίστοιχα στοιχεία των γύρω χωρών.

Όταν αναφερόμαστε στο φορτίο της επόμενης ημέρας και στην παραγωγή της εκάστοτε χώρας, θα πρέπει να λάβουμε υπόψη της επίδραση από την παραγωγή των ανανεώσιμων πηγών ενέργειας (ΑΠΕ). Τα τελευταία χρόνια με την αυξανόμενη διείσδυση των ανανεώσιμων πηγών ηλεκτρικής ενέργειας η τιμή της ηλεκτρικής ενέργειας εμφανίζει ακόμα πιο ασταθή χαρακτήρα και έντονες διακυμάνσεις. Συγκεκριμένα, ενώ θεωρείται δεδομένη η συμβολή των ΑΠΕ στην δημιουργία ενός πιο βιώσιμο κόσμου, υπάρχει συγχρόνως έντονος προβληματισμός σχετικά με την επίδραση που έχουν στις τιμές της ηλεκτρικής ενέργειας και στην σταθερότητα του δικτύου. Πιο συγκεκριμένα, η διείσδυση των ΑΠΕ αυξάνεται και αναμένεται τα επόμενα χρόνια να αυξηθεί ακόμα περισσότερο σε κάποιες χώρες βάσει των ευρωπαϊκών στόχων που έχουν τεθεί. Αυτό έχει ως αποτέλεσμα ότι η παραγωγή

ηλεκτρικής ενέργειας εξαρτάται ακόμα περισσότερο από τις καιρικές συνθήκες και με την σειρά του την αστάθεια στις τιμές της ηλεκτρικής ενέργειας. Έχουν γίνει διάφορες μελέτες που έχουν οδηγήσει στο συμπέρασμα ότι η διείσδυση της αιολικής ή ηλιακής ενέργειας στο υπό μελέτη δίκτυο ηλεκτρικής ενέργειας οδηγεί σε έντονη αστάθεια των τιμών^{[2],[6],[15]}. Επομένως βάσει των παραπάνω διαφαίνεται η σημασία της επίδρασης του παράγοντα των ανανεώσιμων πηγών ενέργειας στην πρόβλεψη της τιμής της ηλεκτρικής ενέργειας.

Από όλα τα παραπάνω, φαίνεται ότι και ο παράγοντας των καιρικών συνθηκών έχει επίδραση στην τιμή της ηλεκτρικής ενέργειας. Η επίδραση του είναι εμφανής τόσο στο φορτίο της χώρας καθώς σε συνθήκες έντονου ψύχους ή ζέστης αυξάνονται οι απαιτήσεις για ηλεκτρική ενέργεια που πιθανόν να οδηγούσε σε αύξηση της τιμής. Από την άλλη σημαντική είναι η επίδραση όπως αναφέρθηκε και στην παραγωγή των ανανεώσιμων πηγών ενέργειας. Σε συνθήκες έντονης ηλιοφάνειας και ανέμου αναμένεται αύξηση στην παραγωγή των ΑΠΕ η οποία αν προσφερθεί στο δίκτυο ηλεκτρικής ενέργειας θα οδηγήσει με την σειρά της σε μείωση της τιμής. Για τις ανάγκες της παρούσας έρευνας, θεωρήθηκε ότι ο παράγοντας των καιρικών συνθηκών έχει ήδη ληφθεί υπόψη στις προβλέψεις του φορτίου ηλεκτρικής ενέργειας αλλά και στην πρόβλεψη της παραγωγής των ανανεώσιμων πηγών ενέργειας από τον αντίστοιχο διαχειριστή συστήματος.

Σύμφωνα με έρευνες που έχουν γίνει, σημαντική φαίνεται επιπλέον και η επίδραση της τιμής του φυσικού αερίου στην τιμή της αγοράς της ηλεκτρικής ενέργειας^{[4],[15]}. Αν και η αγορά του φυσικού αερίου και αυτή της ηλεκτρικής ενέργειας είναι σαφώς διαχωρισμένες και αφορούν δύο διαφορετικά προϊόντα, σε πολλές χώρες το φυσικό αέριο κατέχει υψηλό μερίδιο στο μείγμα καυσίμου για την παραγωγή ηλεκτρικής ενέργειας και έχει ως αποτέλεσμα την δημιουργία ενός δεσμού μεταξύ των δύο αγορών. Κατά αντιστοιχία, με το φυσικό αέριο σημαντικό παράγοντα αποτελούν οι τιμές όλων των καυσίμων και πετρελαιοειδών που χρησιμοποιούνται από τις μονάδες παραγωγής της εκάστοτε χώρας. Στην τιμή των καυσίμων, θα πρέπει να ληφθεί υπόψη και η τιμολόγηση του άνθρακα. Τα τελευταία χρόνια έχει εφαρμοστεί η πρακτική της τιμής για την ρύπανση από τον άνθρακα ώστε να μειωθούν οι εκπομπές και οδηγήσει σε καθαρότερες μορφές ενέργειας. Σύμφωνα με το Σύστημα Εμπορίας

Εκπομπών της Ευρωπαϊκής Ένωσης μόνο ορισμένες βαριές βιομηχανικές επιχειρήσεις οφείλουν να πληρώσουν την τιμή του άνθρακα. Σε αυτή την λίστα ανήκουν και οι γεννήτριες παραγωγής με αποτέλεσμα να αυξάνεται το κόστος παραγωγής τους το οποίο με την σειρά του θα έχει επίδραση στην τιμή της ηλεκτρικής ενέργειας καθώς θα προσφέρει σε υψηλότερη τιμή βάσει του κόστους.

Αντίστοιχα με τον παραπάνω παράγοντα ιδιαίτερης σημασίας είναι και η επίδραση των υδάτινων αποθεμάτων τόσο για την υπό μελέτη χώρα όσο και για τις γειτονικές. Τα υδροηλεκτρικά κατέχουν επίσης ένα σημαντικό τμήμα του μείγματος καυσίμου για την παραγωγή της ηλεκτρικής ενέργειας. Κάποιες χώρες ιδιαίτερα έχουν υποχρέωση στο καθημερινό μείγμα να έχουν έναν συγκεκριμένο ποσοστό υδάτων που θα καλύψουν στην συνέχεια άλλες ανάγκες της χώρας. Επιπλέον, χώρες όπως η Αυστρία με έντονες βροχοπτώσεις και την παρουσία του ποταμού έχει ιδιαίτερα αυξημένο ποσοστό υδροηλεκτρικών.

Επιπλέον, θα πρέπει να ληφθεί υπόψη και η επίδραση του κόστους του δικτύου ηλεκτρικής ενέργειας καθώς κάθε δίκτυο έχει κόστος κατασκευής, συντήρησης και λειτουργίας. Συμπληρωματικά με το παραπάνω, υπάρχει και το κόστος του δικτύου μεταφοράς και διανομής της ηλεκτρικής ενέργειας καθώς και αυτά με την σειρά τους έχουν κόστος κατασκευής, συντήρησης και λειτουργίας.

Τέλος, σημαντική είναι η επίδραση του εποχιακού χαρακτήρα καθώς και των αργιών της χώρας. Ο εποχιακός χαρακτήρας όπως έχει αναφερθεί είναι εμφανής σε ωριαίο επίπεδο, σε εβδομαδιαίο καθώς και σε μηνιαίο. Παρατηρείται σημαντική διαφοροποίηση τις πρωινές ώρες από τις μεσημεριανές και τις νυχτερινές ενώ ταυτόχρονα παρατηρείται διαφορετική συμπεριφορά τις εργάσιμες ημέρες από τα Σαββατοκύριακα καθώς και μεταξύ διαφορετικών μηνών του έτους λόγω διαφορετικών απαιτήσεων για φορτίο. Από την άλλη πλευρά, και οι αργίες σε μία χώρα οδηγούν σε σημαντική πτώση του φορτίου καθώς μειώνονται οι απαιτήσεις τις ημέρες αυτές. Επιπλέον, όπως γίνεται μελέτη σχετικά με την επίδραση των αργιών και λαμβάνονται υπόψη στην ανάπτυξη πολλών μοντέλων, θα μπορούσε αντίστοιχα να ελεγχθεί και η επίδραση ειδικών γεγονότων αν είναι γνωστή η ύπαρξη τους.

3

Τεχνικές Προβλέψεων

Η παραγωγή προβλέψεων όπως έχει παρουσιαστεί στα προηγούμενα κεφάλαια είναι ένα χρήσιμο εργαλείο στα χέρια μιας μικρής ή μεγάλης εταιρείας ή ακόμα και ενός ανθρώπου μεμονωμένου που θα κληθεί να πάρει μια απλή απόφαση στην καθημερινότητά του. Ο μηχανισμός πρόβλεψης μπορεί να είναι η εμπειρία, η διαίσθηση ή ένα σύνολο μεθόδων, τεχνικών και μοντέλων όπως αυτές αναπτύχθηκαν εν συντομία στο προηγούμενο κεφάλαιο.

Η συστηματική ενασχόληση του ανθρώπου με τις προβλέψεις πηγάζει κυρίως με το μεγαλύτερο εχθρό αυτών, που ήταν και θα είναι η αβεβαιότητα. Με τον όρο αβεβαιότητα, δεν αναφερόμαστε μόνο σε αστοχίες των προβλέψεων λόγω κάποιας μικρής διαφοροποίησης της ροής των γεγονότων από αυτό που αναμέναμε, αλλά και σε απρόβλεπτα γεγονότα που διαφοροποιούν τελείως την αλήθεια του μέλλοντος απ' τις εκτιμήσεις του παρόντος. Σε αυτή την κατηγοριοποίηση της αστοχίας έχει αναφερθεί και ο Μακρυδάκης, ονομάζοντας «αβεβαιότητα του μετρό» την πρώτη κατηγορία αβεβαιότητας και «αβεβαιότητα της καρύδας» τη δεύτερη. Η «αβεβαιότητα του μετρό» περιλαμβάνει, δηλαδή, μικροαστοχίες στις προβλέψεις λόγω κάποιου έκτακτου γεγονότος, όπως η καθυστέρηση του μετρό δύο λεπτά γιατί άργησαν να κλείσουν οι πόρτες στην προηγούμενη στάση λόγω πολυκοσμίας ή λόγω

μιας μικρής τεχνικής βλάβης. Από την άλλη, η «αβεβαιότητα της καρύδας» περιλαμβάνει κάτι το αναπάντεχο και απρόβλεπτο που μπορεί να έχει σημαντικές επιδράσεις και συνέπειες όπως είναι το να περπατάς σε μία εξωτική παραλία και να σου πέσει στο κεφάλι μία καρύδα. Προφανώς, οι επιστημονικοί μέθοδοι πρόβλεψης μπορούν να βελτιώσουν προβλήματα αστοχίας μόνο σαν αυτά του μετρώ, και αυτό όμως είναι από μόνο του πολύ σημαντικό.

Αβεβαιότητες, λοιπόν, σαν αυτές του μετρώ και της καρύδας, που κάνουν τα μοντέλα προβλέψεων συχνά να αποκλίνουν σημαντικά ή και να αστοχούν πλήρως, έχουν κατά καιρούς οδηγήσει τον τομέα των προβλέψεων να δέχεται δυσμενείς κριτικές και να αντιμετωπίζει μεγάλη δυσαρέσκεια σχετικά με την ανικανότητα των μεθόδων. Απ' την άλλη όμως, όσο περισσότερο αλλάζει απρόβλεπτα και πολύπλοκα το περιβάλλον, τόσο πιο αδύναμος είναι ο καθένας να προβλέψει απλοϊκά μόνος του και τόσο πιο αναγκαία φαντάζει η ανάγκη παραγωγής προβλέψεων μέσω συστηματικών μεθόδων. Στο πλαίσιο αυτό αναπτύχθηκε και η παρούσα εργασία ώστε να προτείνει μια νέα προσέγγιση σχετικά με την πρόβλεψη της τιμής της ηλεκτρικής ενέργειας.

3.1 Ποιοτικά Χαρακτηριστικά χρονοσειρών

Οι χρονοσειρές αποτελούν ένα σύνολο διαδοχικών παρατηρήσεων της τιμής κάποιου μεγέθους. Όλες οι χρονοσειρές εμπεριέχουν κάποια βασικά χαρακτηριστικά από τα όποια αποτελούνται. Οι παραδοσιακές μέθοδοι ανάλυσης των χρονοσειρών συνήθως ασχολούνται με την αποσύνθεση της διακύμανσης της χρονοσειράς σε τέσσερα συστατικά της στοιχεία: την τάση, την κυκλικότητα, την εποχικότητα και την τυχαιότητα. Μια τέτοια ανάλυση της χρονοσειράς είναι σημαντική τόσο για την εφαρμογή των στατιστικών μεθόδων πρόβλεψης όσο και για την κατανόηση της χρονοσειράς.

Η τάση ορίζεται ως μία μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών της χρονοσειράς. Το πρόβλημα στον παραπάνω ορισμό είναι πότε θεωρείται μια μεταβολή μακροπρόθεσμη. Η απάντηση στο ερώτημα αυτό διαφέρει ανάλογα με την φύση των δεδομένων επομένως χρειάζεται ένας ικανοποιητικός αριθμός

παρατηρήσεων ώστε να μπορεί με σχετική ακρίβεια να εξαχθεί το συμπέρασμα αν υπάρχει τάση στα δεδομένα. Η τάση μπορεί να είναι ανοδική, σταθερή ή πτωτική και να εκτιμηθεί ανάλογα με την μορφή της σε ευθεία ή εκθετική.

Η κυκλικότητα αντιπροσωπεύει μια «κυματοειδή» μεταβολή που εμφανίζεται εξαιτίας εξωγενών συνθηκών και κατά περιόδους. Πρόκειται για περιόδους που το μήκος τους δεν είναι σταθερό αλλά είναι μεγαλύτερες του έτους. Το συγκεκριμένο ποιοτικό χαρακτηριστικό εμφανίζεται κατά κανόνα στα περισσότερα οικονομικά μεγέθη, όπως το Ακαθάριστο Εθνικό Προϊόν, οι δείκτες βιομηχανικής παραγωγής και οι τιμές των μετοχών. Αυτό συμβαίνει εξαιτίας των γενικότερων οικονομικών συνθηκών δηλαδή των υφέσεων και ανόδων που λαμβάνουν χώρα στο πέρασμα του χρόνου.

Ως εποχικότητα ορίζεται μια περιοδική διακύμανση με σταθερό και μικρότερου ή ίσο του έτος μήκος. Συνήθως, η διακύμανση αυτή είναι κατανοητή και εύκολα αναγνωρίσιμη για τον αναλυτή επομένως είναι σχετικά εύκολο να αναλυθεί και να μετρηθεί η επίδραση στα δεδομένα ώστε στην συνέχεια να απομονωθεί. Η συγκεκριμένη διαδικασία γίνεται μέσω των δεικτών εποχικότητας, οι οποίοι αφού βρεθούν διαιρούν τα πραγματικά δεδομένα ώστε να προκύψει τελικά η αποεποχικοποιημένη χρονοσειρά.

Τέλος, η τυχαιότητα αντιπροσωπεύει τις διακυμάνσεις που προκύπτουν από τα πραγματικά δεδομένα όταν όλα τα παραπάνω χαρακτηριστικά έχουν απομονωθεί. Πρόκειται για διακυμάνσεις που αντιπροσωπεύουν μια εντελώς τυχαία μεταβλητή επομένως θα μπορούσε να οριστεί ως μια στοχαστική διαδικασία και να αντιμετωπιστεί ανάλογα. Εναλλακτικά βεβαία, θα μπορούσε να αφορά και κάποια ασυνέχεια από κάποιο ειδικό γεγονός ή απότομη αλλαγή στο επίπεδο της χρονοσειράς. Το ειδικό γεγονός έχει παροδικό χαρακτήρα ενώ η αλλαγή στο επίπεδο μόνιμο. Ως ειδικό γεγονός ορίζεται μια ασυνήθιστη παρατήρηση που οφείλεται σε κάποιο εξαιρετικά απρόβλεπτο γεγονός. Για παράδειγμα, θα μπορούσε να είναι μια πανδημία όπως αυτή που λαμβάνει χώρα αυτή την περίοδο παγκοσμίως και έχει επηρεάσει, πολλούς τομείς, οικονομικούς δείκτες, μετοχές, πωλήσεις κλπ. Από την άλλη πλευρά, η απότομη αλλαγή στο επίπεδο είναι αποτέλεσμα γεγονότων που

επιφέρουν μόνιμες αλλαγές. Για παράδειγμα, η είσοδος ενός ανταγωνιστή στην αγορά ενός προϊόντος αναμένουμε ότι θα επιφέρει σε πρώτο βαθμό απότομη μείωση των πωλήσεων διότι μερίδιο της αγοράς θα στραφεί στην νέα εταιρεία και αν ο ανταγωνιστής παραμείνει στην αγορά αναμένουμε το επίπεδο πωλήσεων για την εταιρεία μετά από ένα διάστημα να σταθεροποιηθεί σε χαμηλότερο επίπεδο.

3.2 Κατηγορίες μεθόδων προβλέψεων

3.2.1 Βασικές κατηγορίες μεθόδων προβλέψεων

Ανάλογα με τον σκοπό χρήσης των παραγόμενων προβλέψεων αλλά και τα διαθέσιμα μέσα, η διαδικασία παραγωγής προβλέψεων γίνεται με διαφορετικό τρόπο. Οι προβλέψεις χωρίζονται σε τρεις βασικές κατηγορίες: α) στατιστική πρόβλεψη, β) κριτική πρόβλεψη και γ) πρόβλεψη στόχου ή προϋπολογισμού. Στην πρώτη κατηγορία υπάρχει η δυνατότητα συστηματοποίησης και να παράγονται προβλέψεις αυτόματα μέσα από εξειδικευμένα λογισμικά που θα εφαρμόζουν υλοποιημένες στατιστικές μεθόδους πρόβλεψης ή μεθόδους μηχανικές μάθησης. Επιπλέον, μπορεί να παράγονται προβλέψεις με την συμβολή ειδικών και εμπειρογνομόνων το οποίο ονομάζεται *κριτική πρόβλεψη* και αφορά τη δεύτερη κατηγορία. Στις περισσότερες περιπτώσεις, το ένα είδος προβλέψεων ακολουθείται από το άλλο καθώς σχεδόν πάντα είναι απαραίτητα η αξιολόγηση των αποτελεσμάτων από τον ανθρώπινο παράγοντα. Επίσης, μπορεί η πρόβλεψη να αφορά τον επιθυμητό στόχο, που επιθυμεί να φτάσει κάποιος επομένως πρόκειται για *πρόβλεψη στόχου ή προϋπολογισμού*. Η τελική πρόβλεψη όπως αναφέρθηκε και παραπάνω πάνω στην οποία θα βασιστεί η στρατηγική της επιχείρησης θα προκύψει από γραμμικό συνδυασμό ίσων ή άνισων βαρών από την στατιστική πρόβλεψη, την κριτική και την πρόβλεψη στόχου.

3.2.2 Μοντέλα πρόβλεψης

Ως μοντέλο πρόβλεψης ορίζεται η διαδικασία που ακολουθείται προκειμένου να παραχθούν προβλέψεις. Κάθε μοντέλο αντιστοιχεί σε μια συγκεκριμένη τεχνική επομένως υπάρχει και μεγάλη ποικιλία. Οι ποσοτικές μέθοδοι πρόβλεψης

διακρίνονται σε δύο κατηγορίες α) το μοντέλο χρονοσειρών και β) το αιτιοκρατικό ή επεξηγηματικό μοντέλο.

Το μοντέλο χρονοσειρών είναι το πιο διαδεδομένο είδος και μπορεί να εφαρμοστεί μόνο με την προϋπόθεση ότι αφενός υπάρχουν διαθέσιμα ιστορικά στοιχεία για το υπό εξέταση μέγεθος σε προηγούμενες και σταθερές χρονικές περιόδους και αφετέρου η μεταβολή του μεγέθους που εξετάζεται ακολουθεί κάποιο επαναλαμβανόμενο και σταθερό μοτίβο. Οι προβλέψεις παράγονται με την αναγνώριση του εν λόγω μοτίβου και την προέκτασή του στο μέλλον. Οι μέθοδοι που περιγράφονται με το συγκεκριμένο μοντέλο αποτελούν μια ιδιαίτερη κατηγορία των ποσοτικών μεθόδων και ονομάζονται μέθοδοι των χρονοσειρών. Σε αυτές συγκαταλέγονται η αποσύνθεση, η εξομάλυνση και οι αυτοπαλινδρομικές μέθοδοι κινητού μέσου όρου. Το μοντέλο των χρονοσειρών είναι εφικτό είναι να εφαρμοστεί θεωρώντας ότι το μοτίβο μπορεί να αναγνωρισθεί μονοσήμαντα με βάση τα ιστορικά δεδομένα και θεωρείται ότι θα διατηρηθεί σταθερό ώστε να οδηγήσει σε προβλέψεις ικανοποιητικές θεωρώντας ότι η εξέλιξη του υπό μελέτη μεγέθους θα ακολουθήσει τις τρέχουσες συνθήκες. Κάτι τέτοιο δεν ισχύει στην περίπτωση της τιμής της ηλεκτρικής ενέργειας. Όπως αναφέρθηκε και σε προηγούμενο κεφάλαιο είναι έντονη η εξάρτηση από διάφορους παράγοντες που μεταβάλλονται ποικιλομόρφως στο πέρασμα του χρόνου.

Το αιτιοκρατικό μοντέλο βασίζεται στην κύρια υπόθεση ότι το υπό μελέτη μέγεθος (εξαρτημένη μεταβλητή) έχει σταθερή σχέση μεταξύ κάποιων παραμέτρων που το επηρεάζουν (ανεξάρτητες μεταβλητές). Στο μοντέλο αυτό προκύπτει η σχέση που συνδέει τις ανεξάρτητες μεταβλητές με την εξαρτημένη βάσει των ιστορικών στοιχείων και με αντικατάσταση των μελλοντικών τιμών για τις ανεξάρτητες μεταβλητές παράγονται οι προβλέψεις για το υπό μελέτη μέγεθος. Για να εφαρμοστεί το εν λόγω μοντέλο είναι απαραίτητη η ύπαρξη ιστορικών δεδομένων όχι μόνο για το υπό εξέταση μέγεθος αλλά και για τις ανεξάρτητες μεταβλητές. Επιπλέον, πολλές φορές για λόγους αξιοπιστίας θεωρείται απαραίτητη η ύπαρξη ιστορικών δεδομένων μεγάλου μήκους. Στην παρούσα εργασία, θα εφαρμοστούν αλγόριθμοι μηχανικής μάθησης που ανήκουν στην αυτή την κατηγορία προβλέψεων. Βάσει των εξισώσεων που επιλύει ο κάθε αλγόριθμος προσπαθεί να βρει την σχέση που συνδέει την

εξαρτημένη μεταβλητή με τις ανεξάρτητες οπότε στην συνέχεια να προκύψουν οι αντίστοιχες προβλέψεις.

3.2.3 Στατιστική πρόβλεψη

Οι στατιστικές προβλέψεις αναφέρονται στην εφαρμογή στατιστικών μοντέλων χρονοσειρών ή αιτιοκρατικών μοντέλων επί μιας σειράς δεδομένων με σκοπό την αυτοματοποιημένη και συστηματική παραγωγή προβλέψεων. Παρά τα πολλά και σημαντικά πλεονεκτήματα των συγκεκριμένων προβλέψεων υπάρχουν ορισμένα μειονεκτήματα. Αρχικά, η εφαρμογή του προϋποθέτει ότι το μοτίβο που ακολουθούν τα τρέχοντα δεδομένα θα συνεχιστεί και στο μέλλον. Επιπλέον, κατά την ανάλυση και παραγωγή προβλέψεων δεν λαμβάνουν υπόψη ειδικά γεγονότα που μπορεί να λάβουν χώρα στο άμεσο μέλλον που επιθυμούμε να προβλέψουμε. Τέλος, κάποιες από τις στατικές μεθόδους απαιτούν μεγάλο όγκο ιστορικών δεδομένων η συλλογή των οποίων δεν είναι πάντα εύκολη διαδικασία. Παρακάτω αναφέρονται επιγραμματικά οι βασικές στατικές μέθοδοι πρόβλεψης.

- I. Naïve (Απλοϊκή μέθοδος): αποτελεί την πιο απλή μέθοδο πρόβλεψης καθώς δίνει ως πρόβλεψη για την επόμενη χρονική στιγμή την πραγματική τιμή της προηγούμενης χρονικής στιγμής. Η εν λόγω μέθοδος πρόβλεψης πολλές φορές χρησιμοποιείται ως σημείο αναφοράς για την σύγκριση απόδοσης μεταξύ άλλων μεθόδων.
- II. Μέθοδοι Εκθετικής εξομάλυνσης: πρόκειται για μεθόδους που έγιναν ιδιαίτερα δημοφιλείς λόγω της ευκολίας, της ελάχιστης απαίτησης σε υπολογιστικό χρόνο και σε ιστορικά δεδομένα προκειμένου να παραχθούν προβλέψεις. Είναι κατάλληλες κυρίως για βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις. Η απόδοσή τους είναι καλύτερη σε δεδομένα που παρουσιάζουν στασιμότητα ή μικρό ρυθμό ανάπτυξης. Από τις πιο γνωστές μεθόδους της εν λόγω κατηγορίας είναι η απλή εκθετική εξομάλυνση σταθερού επιπέδου, η εκθετική εξομάλυνση γραμμικής τάσης και η εκθετική εξομάλυνση μη γραμμικής ή φθίνουσας τάσης.
- III. Μοντέλα παλινδρόμησης: μέσω των μοντέλων αυτών, εξετάζεται η σχέση μεταξύ μιας εξαρτημένης μεταβλητής με συγκεκριμένες ανεξάρτητες μεταβλητές. Τα μοντέλα αυτά διακρίνονται σε απλά και πολλαπλά, ανάλογα με το πλήθος των

ανεξάρτητων μεταβλητών. Ο πρώτος αλγόριθμος που εφαρμόστηκε στην παρούσα εργασία είναι το μοντέλο της πολλαπλής παλινδρόμηση επομένως θα αναλυθεί εκτενέστερα στην συνέχεια.

- IV. Μέθοδος Theta: βάση της συγκριμένης μεθόδου είναι η μεταβολή των τοπικών καμπυλών μιας χρονοσειράς. Η αρχική χρονοσειρά επομένως αναπτύσσεται σε δύο ή περισσότερες γραμμές Theta εκ των οποίων κάθε μία προεκτείνεται ξεχωριστά και προκύπτουν οι προβλέψεις τους που στο τέλος συνδυάζονται.
- V. Μοντέλα ARIMA: τα ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητών μέσω των οποίων αποτελούν στοχαστικά μοντέλα μέσω των οποίων στόχος είναι η περιγραφή της διαχρονικής εξέλιξης κάποιου φυσικού μεγέθους. Η πλειοψηφία των φυσικών μεγεθών εξαρτάται από μη ντετερμινιστικούς παράγοντες επομένως η ανάλυση τους από ντετερμινιστικά μοντέλα συνήθως είναι δύσκολη και ανακριβείς. Τα εν λόγω στοχαστικά μοντέλα έχουν την δυνατότητα να περιγράψουν πιο αποδοτικά το μέγεθος που μελετάται και την εξάρτηση από τυχαίους στοχαστικούς παράγοντες με αποτέλεσμα να παράγονται πιο ακριβείς προβλέψεις.
- VI. Νευρωνικά δίκτυα: ένα τεχνητό νευρωνικό δίκτυο μπορεί να αποδειχθεί χρήσιμο για μη γραμμικές διαδικασίες, όπου η συσχέτιση δεν είναι εξαρχής γνωστή και επομένως είναι δύσκολο να επιτευχθεί βέλτιστη προσαρμογή. Η βασική ιδέα κατά την υλοποίηση του νευρωνικού είναι το φιλτράρισμα των εισόδων, δηλαδή των ανεξάρτητων μεταβλητών, μέσω ενός ή περισσότερων κρυφών επιπέδων, τα οποία αποτελούνται από κρυφούς κόμβους, πριν παραχθεί η ζητούμενη έξοδος. Η υλοποίηση νευρωνικού δικτύου μέσω του αλγορίθμου πολυεπίπεδων αισθητήρων (MLP) είναι μία από τις εφαρμογές της παρούσας εργασίας οπότε ακολουθεί εκτενής ανάλυση στην συνέχεια.

Τέλος, αξίζει να αναφερθεί και η ύπαρξη μεθόδων συνδυασμού, μέσω των οποίων συνδυάζονται οι προβλέψεις δύο ή περισσότερων μεθόδων με ίσα ή άνισα βάρη για την παραγωγή των τελικών προβλέψεων. Σε αρκετές περιπτώσεις ένας τέτοιος συνδυασμός έχει βελτιώσει σημαντικά την αξιοπιστία της τελικής πρόβλεψης, ενώ μπορεί να συνδράμει σημαντικά στην βελτίωση της προκατάληψης κάποιας μεθόδου.

3.3 Ανάλυση μεθόδων εφαρμογής

3.3.1 Πολλαπλή γραμμική παλινδρόμηση (MLR)

Η πρώτη μέθοδος που εφαρμόστηκε στην εφαρμογή της παρούσας εργασίας είναι το μοντέλο πολλαπλής γραμμικής παλινδρόμησης. Στόχος της παλινδρόμησης είναι η εύρεση συσχετίσεων μεταξύ μιας εξαρτημένης μεταβλητής με μία ή και περισσότερες ανεξάρτητες μεταβλητές. Η πρόβλεψη της εξαρτημένης μεταβλητής είναι μια μαθηματική συνάρτηση των ανεξάρτητων μεταβλητών η οποία καλείται εξίσωση της παλινδρόμησης. Στην πολλαπλή παλινδρόμηση υπάρχει μια εξαρτημένη μεταβλητή της οποίας η τιμή πρέπει να προβλεφθεί βάσει των τιμών δύο ή περισσότερων εξαρτημένων μεταβλητών. Η γενική μορφή της εξίσωσης της γραμμικής παλινδρόμησης είναι η ακόλουθη:

$$Y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k + e$$

Η μεταβλητή Y εκφράζει την εξαρτημένη μεταβλητή, ενώ οι μεταβλητές x_1, x_2, \dots, x_k εκφράζουν τις ανεξάρτητες μεταβλητές. Οι συντελεστές b_0, b_1 έως b_k είναι σταθερές παράμετροι. Τέλος, το e δηλώνει τον τυχαίο παράγοντα, ο οποίος θεωρείται κανονικά κατανομημένος γύρω από το μηδέν. Στην περίπτωση της παρούσας εφαρμογής ως εξαρτημένη τιμή ορίζεται η τιμή της ηλεκτρικής ενέργειας στην αγορά της Ουγγαρίας ενώ ως ανεξάρτητες μεταβλητές ορίζονται όλες οι μεταβλητές εισόδου που επιλέχθηκαν για εξέταση και θα αναφερθούν εκτενέστερα στην συνέχεια όπως είναι το φορτίο, η παραγωγή από τις ΑΠΕ κλπ.

Η εξίσωση της πολλαπλής παλινδρόμησης είναι γραμμική ως προς τους συντελεστές. Ο εκθέτης για κάθε συντελεστή ισούται με την μονάδα επομένως συνεπάγεται η γραμμικότητα, ενώ οι τιμές των εν λόγω συντελεστών μπορεί να προκύψει με την εφαρμογή της μεθόδου των ελαχίστων τετραγώνων. Όταν υπάρχει μια μόνο μεταβλητή X , τότε η μορφή της συνάρτησης που προκύπτει είναι μια ευθεία γραμμή. Όταν οι μεταβλητές είναι δύο, το σχήμα της συνάρτησης γίνεται επίπεδο και καθώς αυξάνονται η παράσταση γίνεται σε υπερεπίπεδο επομένως η μορφή δεν είναι εύκολο να περιγραφεί.

Η μέθοδος η συγκεκριμένη έχει ως στόχο της εφαρμογή της μεθόδου των ελαχίστων τετραγώνων στο σύνολο των δεδομένων ώστε να υπολογιστούν οι άγνωστες παράμετροι b_0, b_1 έως b_k και η διακύμανση του τυχαίου παράγοντα. Στην ουσία η μορφή του στατιστικού μοντέλου παλινδρόμησης είναι:

$$Y = b_0 + b_1 \cdot x_{1,i} + b_2 \cdot x_{2,i} + \dots + b_k \cdot x_{k,i} + e_i$$

Ο υπολογισμός των άγνωστων παραμέτρων προκύπτει από την ελαχιστοποίηση της ακόλουθης σχέσης:

$$\sum_{i=1}^n (y_i - \widehat{y}_i)^2 = \sum_{i=1}^n (y_i - \widehat{b}_0 - \widehat{b}_1 x_{i1} - \dots - \widehat{b}_k x_{ik})^2$$

Επομένως, αρκεί να υπολογιστούν οι μερικές παράγωγοι αυτής για κάθε έναν από τους συντελεστές, να τεθούν οι υπολογισμένες παράγωγοι ίσες με το μηδέν και να λυθεί το γραμμικό σύστημα εξισώσεων.

3.3.2 Μοντέλο πολυεπίπεδων αισθητήρων (MLP)

Το μοντέλο αυτό όπως αναφέρθηκε και στην προηγούμενη ενότητα αποτελεί ειδική εφαρμογή των μοντέλων τεχνητών νευρωνικών δικτύων. Όπως ήδη αναφέρθηκε βασικό πλεονέκτημα της μεθόδου των νευρωνικών δικτύων είναι ότι μπορεί να αναγνωρίσει μη γραμμικές συσχετίσεις ανάμεσα στις εισόδους. Πρόκειται για έναν αλγόριθμο που «μαθαίνει» από τα ιστορικά διαθέσιμα δεδομένα, επομένως δεν μπορούν να δώσουν απάντηση στο πώς προέκυψε το αποτέλεσμα.

Συγκεκριμένα το μοντέλο MLP είναι ένας αλγόριθμος με επίβλεψη δηλαδή αξιοποιεί δεδομένα και υπάρχουσες λύσεις ενός προβλήματος προκειμένου να αποκαλύψει τη σχέση που συνδέει τις μεταβλητές εισόδου ανεξάρτητες μεταβλητές και εξόδου εξαρτημένες μεταβλητές. Συγκεκριμένα είναι ένα δίκτυο εμπρόσθιας τροφοδότησης το οποίο σημαίνει τα δεδομένα εισόδου τροφοδοτούνται στο δίκτυο από τους νευρώνες εισόδου και μεταφέρονται στα επόμενα στρώματα του δικτύου. Στο τελευταίο στρώμα του δικτύου παράγεται η έξοδός του. Κατά την διαδικασία της

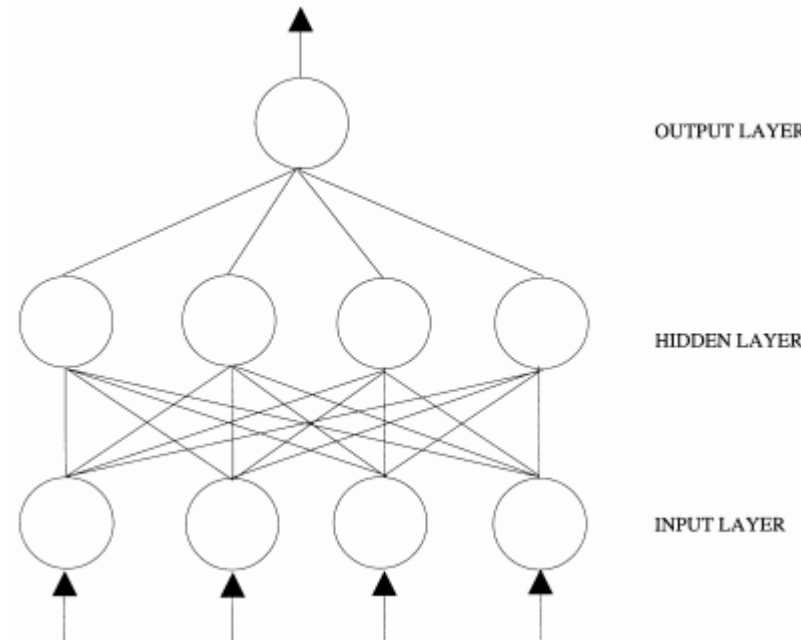
μάθησης τα βάρη που συνδέουν τις νευρώνες προσαρμόζονται έτσι ώστε να ελαχιστοποιείται το σφάλμα μεταξύ εξόδου του δικτύου και την πραγματική έξοδο που επιθυμούμε να προσομοιάσουμε. Το βασικό στοιχείο του συγκεκριμένου αλγορίθμου (perceptron) είναι μια υπολογιστική μονάδα η οποία λαμβάνει τον γραμμικό συνδυασμό των εισόδων της και μέσω της συνάρτησης ενεργοποίησης παράγει την έξοδό της.

Η έξοδος επομένως δίνεται από την ακόλουθη σχέση:

$$o = f\left(\sum_{i=1}^n w_i x_i - b\right)$$

Όπου o η έξοδος, x_i οι εισοδοι, w_i τα βάρη των νευρώνων, b ο συντελεστής πόλωσης και f η συνάρτηση ενεργοποίησης.

Ένα νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης μπορεί να έχει την ακόλουθη μορφή:

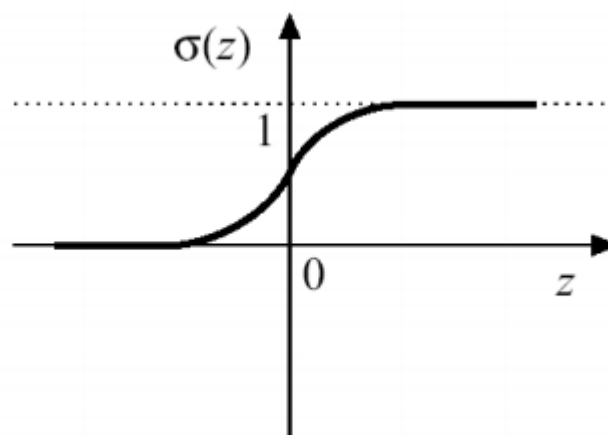


Εικόνα 10: Νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης (MLP)

Κατά την εφαρμογή ενός αλγορίθμου νευρωνικού δικτύου είναι πολύ σημαντικό να οριστεί σωστά η αρχιτεκτονική του δικτύου. Με τον όρο αρχιτεκτονική εννοούμε τον αριθμό των στρωμάτων του δικτύου, τον αριθμό των νευρώνων σε κάθε στρώμα, και τις γραμμές που θα συνδέουν τους νευρώνες. Επιπλέον, σημαντική η επιλογή της συνάρτησης ενεργοποίησης τόσο για τους νευρώνες των ενδιάμεσων στρωμάτων αλλά και αυτών της εξόδου, ο αλγόριθμος εκμάθησης και οι παράμετροι που μπορεί ο ίδιος να έχει.

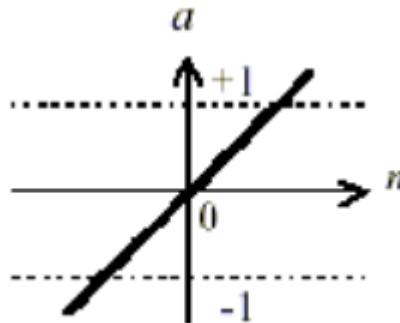
Η μορφή της συνάρτησης ενεργοποίησης κατέχει σημαντικό ρόλο στην αρχιτεκτονική του δικτύου και στην ακρίβεια των προβλέψεων που θα παραχθούν. Η συνάρτηση ενεργοποίησης μπορεί να είναι της ακόλουθης μορφής: γραμμική, βηματική, λογαριθμική και υπερβολική εφαιπτομένη. Παρακάτω, φαίνεται το σχήμα των δύο βασικών συναρτήσεων που εξετάστηκαν στην συγκεκριμένη εφαρμογή.

Η πρώτη συνάρτηση που φαίνεται στο ακόλουθο σχήμα είναι η λογαριθμική. Η μορφή της συγκεκριμένης συνάρτησης μοιάζει με την υπερβολική εφαιπτομένη με την βασική διαφορά ότι στην λογαριθμική οι τιμές είναι από 0 έως 1 ενώ στην υπερβολική από -1 έως 1.



Εικόνα 11: Λογαριθμική συνάρτηση ενεργοποίησης

Η επόμενη συνάρτηση που φαίνεται στο ακόλουθο σχήμα είναι η γραμμική. Η συνάρτηση αυτή αποτυπώνει καλύτερα χρονοσειρές με τάση.



Εικόνα 12: Γραμμική συνάρτηση ενεργοποίησης

Στην Εικόνα 10 παραπάνω, παρουσιάστηκε ένα τυπικό νευρωνικό δίκτυο και φαίνονται τα στρώματα νευρώνων από τα οποία αποτελείται. Κάθε νευρώνας συνδέεται με όλους του επόμενου στρώματος χωρίς να υπάρχουν αναδράσεις καθώς πρόκειται για δίκτυο εμπρόσθιας τροφοδότησης. Η διαδικασία εκμάθησης ενός δικτύου εμπρόσθιας τροφοδότησης είναι ο ακόλουθος. Εισάγεται στο δίκτυο ένα διάνυσμα εισόδων. Στην συνέχεια η είσοδος διαδίδεται προς τα εμπρός στο δίκτυο έως ότου η ενεργοποίηση να φτάσει στο επίπεδο εξόδου. Η διαδικασία αυτή ονομάζεται προς τα εμπρός διάδοση. Η έξοδος από το τελικό στρώμα συγκρίνεται με την είσοδο που έγινε η μάθηση και το επιθυμητό αποτέλεσμα. Το λάθος, δηλαδή η διαφορά ανάμεσα στην έξοδο και στην είσοδο εκμάθησης χρησιμοποιούνται μαζί για να υπολογιστεί η διαφορά που πρέπει να εφαρμοστεί στο βάρος της διασύνδεσης. Για τον υπολογισμό των διαφορών των εσωτερικών μονάδων που δεν είναι διαθέσιμη η μονάδα εισόδου χρησιμοποιείται η μαθηματική σχέση που φαίνεται παρακάτω.

$$\Delta w_{ij} = \eta \delta_j o_i$$

$$\delta_j = \begin{cases} f'_j(\text{net}_j)(t_j - o_j) & \text{if unit } j \text{ is a output-unit} \\ f'_j(\text{net}_j) \sum_k \delta_k w_{jk} & \text{if unit } j \text{ is a hidden-unit} \end{cases}$$

Εικόνα 13: Σχέση υπολογισμού βαρών μέσω backpropagation

Στην παραπάνω σχέση, η είναι ο ρυθμός μάθησης, δ_j είναι η διαφορά μεταξύ πραγματικής εξόδου και νευρώνα εισόδου, t_j δεδομένο εισόδου από νευρώνα j , o_j έξοδος του προηγούμενου νευρώνα i . Με τον παραπάνω τρόπο, τα σφάλματα διαδίδονται προς τα πίσω, οπότε η φάση αυτή ονομάζεται μετάδοση προς τα πίσω. Στην εκμάθηση του νευρωνικού, οι αλλαγές στα βάρη μπορεί να εφαρμόζονται στο δίκτυο μετά από κάθε εκμάθηση, μετά δηλαδή από κάθε πέρασμα προς τα εμπρός και προς τα πίσω. Επίσης, μπορεί να αθροίζονται οι αλλαγές από όλα τα μοτίβα εκμάθησης και το άθροισμα αυτό να εφαρμόζεται μετά από έναν πλήρη κύκλο που ονομάζεται εποχή (epoch).

Σχετικά με τον αλγόριθμο μάθησης του νευρωνικού δικτύου, υπάρχουν διάφοροι με τον δημοφιλέστερο και πιο συχνά χρησιμοποιούμενο αυτό της προς τα πίσω διάδοσης σφάλματος (backpropagation algorithm) που αναλύθηκε προηγουμένως. Ο τυπικός αλγόριθμος είναι μια μέθοδος πτωτικής κλίσης (gradient steepest descent method). Για τον αλγόριθμο πτωτικής κλίσης, εφαρμόζεται ένα βήμα που ονομάζεται ρυθμός μάθησης με τον οποίο τα υπολογιζόμενα σφάλματα αλλάζουν τα βάρη του δικτύου. Η επιλογή της συγκεκριμένης παραμέτρου είναι πολύ σημαντική. Από την μία, η επιλογή ενός μεγάλο ρυθμού μάθησης μπορεί να οδηγήσει σε ταχύτητα σύγκλισης επομένως και υπολογιστική ταχύτητα, από την άλλη πλευρά μπορεί να προσπεράσει την πραγματική θέση του ελαχίστου επομένως να οδηγήσει σε ανακριβή αποτελέσματα. Υπάρχουν αρκετές βελτιώσεις μέσω επιπλέον παραμέτρων που έχουν γίνει ώστε να βελτιωθεί η απόδοση του backpropagation αλγορίθμου όμως στην συγκεκριμένη εφαρμογή επιλέχθηκε η κλασική προσέγγιση.

Ένα επίσης σημαντικό στάδιο πριν την εφαρμογή του αλγορίθμου νευρωνικού δικτύου είναι αυτό της κανονικοποίησης των δεδομένων. Συγκεκριμένα, οι μη γραμμικές συναρτήσεις ενεργοποίησης όπως είναι η λογαριθμική που αναφέρθηκε προηγουμένως που συνήθως χρησιμοποιείται δίνουν μια τιμή εξόδου στα όρια (0,1). Δεδομένου ότι η είσοδος μπορεί να πάρει οποιαδήποτε τιμή και θα είναι συνήθως μεγάλη αυτό θα οδηγήσει σε αδυναμία υπολογισμού καθώς οι προσαρμογές που θα γίνονται στα βάρη δεν θα φέρουν κανένα ουσιαστικό αποτέλεσμα λόγω σύγκρισης ανόμοιων δεδομένων. Προκειμένου να αποφευχθεί το παραπάνω, σημαντική είναι η προεπεξεργασία των δεδομένων. Βασική διαδικασία είναι ο μετασχηματισμός και η

κανονικοποίηση των δεδομένων. Ο μετασχηματισμός αφορά την επεξεργασία των αρχικών δεδομένων για την δημιουργία ενός άλλου τύπου εισόδου στο δίκτυο, ενώ η κανονικοποίηση είναι ένας μετασχηματισμός βάσει των αρχικών δεδομένων ώστε να προσαρμοστούν σε μία κλίμακα. Σε προβλήματα πρόβλεψης χρονοσειρών, συνήθως όλα τα δεδομένα που είναι διαθέσιμα για εκμάθηση μετασχηματίζονται σε μία συγκεκριμένη κλίμακα. Συνήθως, η διαδικασία που χρησιμοποιείται είναι αυτή της γραμμικής κανονικοποίησης σε κλίμακα (0,1). Προκειμένου να εφαρμοστεί η διαδικασία αυτή χρειάζεται να είναι γνωστή η μέγιστη και η ελάχιστη τιμή της χρονοσειράς εισόδου και αντίστοιχα να οριστεί το εύρος των τιμών εισόδου max-min. Τέλος, προκειμένου να αξιολογηθούν τα αποτελέσματα που θα προκύψουν από τον συγκεκριμένο αλγόριθμο χρειάζεται να γίνει μετατροπή των δεδομένων εξόδου βάσει της μέγιστης και της ελάχιστης τιμής του δείγματος.

Ένας άλλος παράγοντας που αναφέρθηκε ότι επηρεάζει την αρχιτεκτονική του δικτύου είναι αυτός του αριθμού των νευρώνων εισόδου. Προκειμένου να αποφευχθεί υπερπροσαρμογή ή υποπροσαρμογή στα δεδομένα το πλήθος των εισόδων θα πρέπει να είναι προσεκτικά επιλεγμένο βάσει των δεδομένων εισόδου. Επιπλέον, σημαντική είναι και η επιλογή του αριθμού των κρυμμένων στρωμάτων που κατ' αντιστοιχία με το παραπάνω δεν πρέπει να είναι πολύ μεγάλος ο αριθμός για να μην επιβαρύνεται το δίκτυο ως προς την υπολογιστική απόδοση αλλά και την αξιοπιστία των αποτελεσμάτων.

Επομένως, βάσει όλης της παραπάνω ανάλυσης μπορεί να «χτιστεί» μια αρχιτεκτονική νευρωνικού δικτύου βάσει της φύσης των δεδομένων και του μεγέθους που μελετάται που θα οδηγήσει σε ιδιαίτερα ικανοποιητικά και αξιόπιστα αποτελέσματα.

3.3.3 Δέντρα αποφάσεων

Οι αλγόριθμοι που υλοποιούν δέντρα απόφασης αποτελούν μοντέλα προβλέψεων που χρησιμοποιούνται είτε σε προβλήματα παλινδρόμησης όπως το πρόβλημα που μελετάται στην παρούσα εργασία είτε σε προβλήματα ταξινόμησης. Αυτό συνεπάγεται ότι μπορούν να διαχειριστούν είτε αριθμητικά είτε κατηγορηματικά δεδομένα. Το

δέντρο απόφασης δημιουργεί το μοντέλο πρόβλεψης με την μορφή δομής δέντρου, το οποίο σημαίνει ότι διασπάται το σύνολο των δεδομένων σε μικρότερα και συνεχώς μικρότερα υποσύνολα ενώ ταυτόχρονα αναπτύσσεται σταδιακά ένα σχετικό δέντρο αποφάσεων. Όπως υποδηλώνει το όνομά του, μπορεί να θεωρηθεί ότι η ανάλυση των δεδομένων βασίζεται σε αποφάσεις που λαμβάνονται έπειτα από μια σειρά ερωτήσεων. Επομένως, πρόκειται για πρόβλεψη ενός στόχου εκμάθησης κανόνων απόφασης από χαρακτηριστικά.

Ένα δέντρο αποφάσεων κατασκευάζεται με αναδρομικό σχεδιασμό. Ξεκινώντας από τον ριζικό κόμβο (γνωστό ως πρώτο γονέα), κάθε κόμβος μπορεί να χωριστεί σε δύο ή περισσότερους θυγατρικούς κόμβους, οι οποίοι μπορούν στην συνέχεια να χωριστούν περαιτέρω επομένως να γίνουν αυτοί γονικοί για τους θυγατρικούς τους κ.ο.κ. Κατά την παραπάνω διαδικασία, τα δεδομένα χωρίζονται με στόχο την βελτιστοποίηση της αντικειμενικής συνάρτησης που επιλέγει ο κάθε αλγόριθμος. Η διαδικασία διαχωρισμού επαναλαμβάνεται έως ότου τα φύλλα να έχουν την ίδια τιμή της μεταβλητής στόχου ή όταν ο διαχωρισμός δεν προσθέτει πλέον αξία στις προβλέψεις. Στην πράξη, η παραπάνω διαδικασία μπορεί να οδηγήσει σε βαθύ δέντρο με πολλούς κόμβους, το οποίο μπορεί να έχει ως αποτέλεσμα υπερπροσαρμογή στα δεδομένα. Για να αποφευχθεί αυτό συνήθως εφαρμόζεται ένα κλάδεμα στο δέντρο θέτοντας ένα όριο για το μέγιστο βάθος του δέντρου.

Το τελικό αποτέλεσμα είναι ένα δέντρο με κόμβους αποφάσεων και κόμβους φύλλων. Ένας κόμβος απόφασης έχει δύο ή περισσότερους κλάδους καθένας από τους οποίους αντιπροσωπεύει τιμές για το χαρακτηριστικό που δοκιμάστηκε. Ο κορυφαίος κόμβος απόφασης σε ένα δέντρο που αντιστοιχεί στον καλύτερο προγνωστικό παράγοντα ονομάζεται κόμβος ρίζα (root node).

Αυτή η διαδικασία της top-down επαγωγής δέντρων απόφασης είναι ένα παράδειγμα ενός άπληστου αλγορίθμου, και είναι η πιο κοινή στρατηγική για την εκμάθηση δέντρων απόφασης από τα δεδομένα. Τα δεδομένα συνήθως έχουν την ακόλουθη μορφή:

$$(x, Y) = (x_1, x_2, \dots, x_k, Y)$$

Η εξαρτημένη μεταβλητή Y είναι η μεταβλητή στόχος που προσπαθούμε να προβλέψουμε. Το διάνυσμα x από τις μεταβλητές εισόδου που χρησιμοποιούνται για την διεργασία και την εξαγωγή αποτελεσμάτων.

Όπως περιγράφεται παραπάνω, τα δέντρα απόφασης είναι μια σειρά διαδοχικών βημάτων που έχουν σχεδιαστεί για να απαντήσουν σε ένα ερώτημα και να παρέχουν πιθανότητα, κόστος ή άλλο χαρακτηριστικό ανάλογα με το πρόβλημα που μελετάται. Είναι σχετικά απλό στην κατανόηση και παρέχει μια σαφή οπτικοποίηση για την διαδικασία της λήψης αποφάσεων. Ωστόσο, η παραπάνω απλότητα μπορεί να οδηγήσει σε κάποια μειονεκτήματα. Αρχικά, όπως αναφέρθηκε και παραπάνω αν δεν γίνει το κατάλληλο κλάδεμα και περιορισμός στους κόμβους των δέντρων μπορεί να οδηγήσει σε υπερπροσαρμογή των δεδομένων καθώς και σε σφάλμα προκατάληψης λόγω περιορισμών στην αντικειμενική συνάρτηση και σφάλμα διακύμανσης καθώς μικρές αλλαγές στα δεδομένα εκμάθησης μπορεί να οδηγήσουν σε μεγάλες αλλαγές στο τελικό αποτέλεσμα. Για τους παραπάνω λόγους, συχνά επιλέγονται αλγόριθμοι που δημιουργούν και επιλύουν παραπάνω από ένα δέντρα, όπως είναι αυτοί που αναλύονται παρακάτω.

3.3.3.1 Random Forest

Ο αλγόριθμος του τυχαίου ταξινομητή (random forest) ανήκει στην κατηγορία των αλγορίθμων δέντρων απόφασης. Πρόκειται για έναν αλγόριθμο που επιλύει ένα σύνολο δέντρων και οδηγεί σε ένα μόνο συγκεντρωτικό αποτέλεσμα. Συνήθως, μειώνουν την διακύμανση που παρατηρείται στα μεμονωμένα δέντρα απόφασης, διότι χρησιμοποιεί διαφορετικά δείγματα για εκπαίδευση, καθορίζει υποσύνολο τυχαίων χαρακτηριστικών και κατασκευάζει συνδυασμό ρηχών δέντρων. Πιο αναλυτικά, κατά την επίλυση ο συγκεκριμένος αλγόριθμος επιλέγει τυχαίες μεταβλητές εισόδου που θα χρησιμοποιήσει πριν από κάθε διαχωρισμό κόμβου. Από τις μεταβλητές αυτές επιλέγει αυτή που θα βελτιστοποιήσει την αντικειμενική συνάρτηση και θα γίνει ο διαχωρισμός. Έπειτα από τον διαχωρισμό, παράγονται δύο κόμβοι που περιέχουν κάποια πληροφορία ανάλογα με τον τύπο των δεδομένων. Η διαδικασία αυτή επαναλαμβάνεται αναδρομικά για κάθε ένα από τα δέντρα που θα επιλυθούν συνολικά. Στο τέλος, η έξοδος είναι ένα μόνο συγκεντρωτικό αποτέλεσμα και

προκύπτει ως ο μέσος όρος των δέντρων που επιλύθηκαν. Ο μεγαλύτερος αριθμός δέντρων οδηγεί ίσως σε ένα πιο αξιόπιστο μοντέλο όμως μπορεί να κάνει την διαδικασία επίλυσης αρκετά πιο αργή. Επίσης, όπως αναφέρθηκε και παραπάνω, ένα μόνο του δέντρο απόφασης είναι εύκολο να μελετηθεί και να κατανοήσει κάποιος πώς προέκυψε η λύση. Στην περίπτωση του τυχαίου δάσος είναι πιο περίπλοκο να ερμηνευτεί.

3.3.3.2 *Gradient Boosting Machine*

Όπως ο παραπάνω αλγόριθμος έτσι ο συγκεκριμένος αποτελεί ένα σύνολο δέντρων αποφάσεων. Οι δύο βασικές διαφορές με το τυχαίο δάσος είναι στον τρόπο που χτίζονται τα δέντρα και στο πώς προκύπτει το τελικό αποτέλεσμα. Στον συγκεκριμένο αλγόριθμο χτίζεται ένα μόνο δέντρο και όχι πολλά ανεξάρτητα όπως στην προηγούμενη κατηγορία. Όπως και άλλοι μέθοδοι ενίσχυσης, έτσι και η ενίσχυση κλίσης συνδυάζει τους αδύναμους μαθητές σε έναν ισχυρό με επαναληπτικό τρόπο. Πρόκειται για ένα προς τα εμπρός προσθετικό μοντέλο που προσπαθεί να διορθώσει συνεχώς το λάθος του προηγούμενου δέντρου δηλαδή να βελτιώσει τις αδυναμίες των μη ισχυρών μαθητών. Επομένως, το αποτέλεσμα από τον αλγόριθμο αυτό προκύπτει κατά την διάρκεια επίλυσης και δεν απαιτείται κάποιος συνδυασμός στο τέλος.

Κατά το χτίσιμο του συγκεκριμένου αλγορίθμου, υπάρχουν σημαντικές παράμετροι που πρέπει να επιλεγθούν. Αρχικά πρέπει να επιλεγθεί ο αριθμός των δέντρων το οποίο πρακτικά δηλώνει τον αριθμό των επαναλήψεων που θα γίνουν. Η επόμενη βασική παράμετρος είναι η συνάρτηση κόστους βάσει της οποίας θα γίνει η αξιολόγηση σε κάθε κόμβο. Οι διαθέσιμες επιλογές είναι "gaussian" (squared error), "laplace" (absolute loss), "bernoulli" (λογιστική παλινδρόμηση για 0-1 αποτελέσματα), "huberized" (άρθρωση απώλεια για 0-1 αποτελέσματα, κλάσεις), "adaboost" (η εκθετική απώλεια για 0-1 αποτελέσματα), "poisson" (μετρήσεις αποτελεσμάτων), "coxph" (παρατηρήσεις σωστής λογοκρισίας), "κβαντική" ή "ζεύγη" (μέτρηση κατάταξης χρησιμοποιώντας τον αλγόριθμο LambdaMart). Η επόμενη μεταβλητή που πρέπει να επιλεγθεί είναι το μέγιστο βάθος κάθε δέντρου, δηλαδή το μέγιστο επιτρεπόμενο επίπεδο μεταβλητών αλληλεπίδρασης. Επιπλέον, θα πρέπει να επιλεγθεί η συρρίκνωση που εφαρμόζεται σε κάθε δέντρο κατά την επέκταση που

είναι γνωστή ως ρυθμός εκμάθησης. Μικρός αριθμός εκμάθησης μπορεί να βελτιώσει σημαντικά τα αποτελέσματα όμως μπορεί συγχρόνως να αυξήσει πολύ τον χρόνο υπολογισμού.

Με προσεκτική επιλογή των υπερπαραμέτρων που έχει ο συγκεκριμένος αλγόριθμος μπορεί να οδηγήσει σε καλύτερα αποτελέσματα από το τυχαίο δάσος.

3.4 Στατιστική Ανάλυση και Ακρίβεια Προβλέψεων

3.4.1 Σφάλματα Προβλέψεων

Βασικό κριτήριο για την αξιολόγηση και την επιλογή ενός μοντέλου πρόβλεψης είναι η ακρίβεια των προβλέψεων που παράγονται από αυτό. Προκειμένου να υπολογίσουμε την ακρίβεια του μοντέλου χρειάζεται να υπολογίσουμε κάποιους στατιστικούς δείκτες σφάλματος. Γενικά, το σφάλμα ορίζεται ως η διαφορά μεταξύ της πραγματικής τιμής Y και της πρόβλεψης F για μια περίοδο, όπως φαίνεται στον ακόλουθο τύπο:

$$e_i = Y_i - F_i$$

Οι στατιστικοί δείκτες σφάλματος μπορεί να υπολογισθούν είτε με τις τιμές που υπάρχουν διαθέσιμες και έχουν προβλεφθεί και από το μοντέλο (*in-sample error*) είτε εκ των υστέρων αφού γίνουν γνωστές οι πραγματικές τιμές οπότε είναι εφικτό να υπολογιστεί τόσο το σφάλμα του μοντέλου όσο και το πραγματικό σφάλμα (*out-of-sample error*). Σε γενικές γραμμές, η ακρίβεια ενός μοντέλου πρόβλεψης αξιολογείται βάσει του σφάλματος που προκύπτει επί των πραγματικών τιμών γιατί μας ενδιαφέρει η απόδοση στο άγνωστο περιβάλλον. Για τον λόγο αυτό, κατά την ανάπτυξη ενός μοντέλου προβλέψεων συνήθως επιλέγεται ένα τμήμα των διαθέσιμων δεδομένων για εκμάθηση του μοντέλου (*training set*) και ένα τμήμα για την αξιολόγηση του μοντέλου σε πραγματικές συνθήκες (*outsample* ή *test set*). Οι βασικοί στατιστικοί δείκτες σφάλματος είναι οι ακόλουθοι:

1. Μέσο σφάλμα (*mean error*). Υπολογίζεται ως ο μέσος όρος των απλών σφαλμάτων. Αν ο δείκτης παίρνει θετικές τιμές, τότε το μοντέλο δίνει απαισιόδοξες προβλέψεις οι οποίες είναι κατά μέσο όρο μικρότερες των πραγματικών τιμών. Το αντίθετο ισχύει στην περίπτωση που ο δείκτης είναι αρνητικός. Ο υπολογισμός γίνεται από τον ακόλουθο τύπο:

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)$$

2. Μέσο απόλυτο σφάλμα (*Mean Absolute Error*). Δηλώνει ένα μέσο μέτρο αστοχίας της πρόβλεψης χωρίς να δίνει κατεύθυνση για αυτή. Όσο μεγαλύτερη η τιμή του δείκτη τόσο μικρότερη η ακρίβεια του μοντέλου.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F_i|$$

3. Μέσο τετραγωνικό σφάλμα (*Mean Squared Error*). Πρόκειται για τον μέσο των απλών σφαλμάτων αφού πρώτα υψωθούν στο τετράγωνο. Είναι ένας δείκτης ακρίβειας που δίνει μεγαλύτερο βάρος στα μεγάλα σφάλματα και μικρότερο στα μικρά αφού υψώνονται στο τετράγωνο.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2$$

4. Ρίζα μέσου τετραγωνικού σφάλματος (*Root Mean Squared Error*). Υπολογίζεται ως η τετραγωνική ρίζα του προηγούμενου σφάλματος επομένως είναι εκφρασμένο σε πραγματικές μονάδες του υπό μελέτη μεγέθους.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2}$$

5. Μέσο απόλυτο ποσοστιαίο σφάλμα (*Mean Absolute Percentage Error*). Είναι ένας χρήσιμος δείκτης για την αξιολόγηση ενός μοντέλου πρόβλεψης που έχει εφαρμοστεί σε διαφορετικά δεδομένα ή στην περίπτωση που το υπό μελέτη μέγεθος παίρνει υψηλές τιμές. Λαμβάνει τιμές μεγαλύτερες ή ίσες του 0, ενώ όσο μικρότερο είναι τόσο πιο ακριβής είναι η μέθοδος.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{Y_i} \right| 100(\%)$$

6. Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (*Symmetric Mean Absolute Percentage Error*). Αποτελεί παραλλαγή του προηγούμενου σφάλματος. Με την διαφοροποίηση που γίνεται στον παρονομαστή του κλάσματος. Δηλαδή γίνεται διαίρεση με το ημιάθροισμα πραγματικής και προβλεπόμενης τιμής και ο εν λόγω δείκτης μπορεί να πάρει τιμές 0 έως 200%.

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{\left(\frac{Y_i + F_i}{2}\right)} \right| 100 = \frac{1}{n} \sum_{i=1}^n \left| \frac{2(Y_i - F_i)}{Y_i + F_i} \right| 100(\%)$$

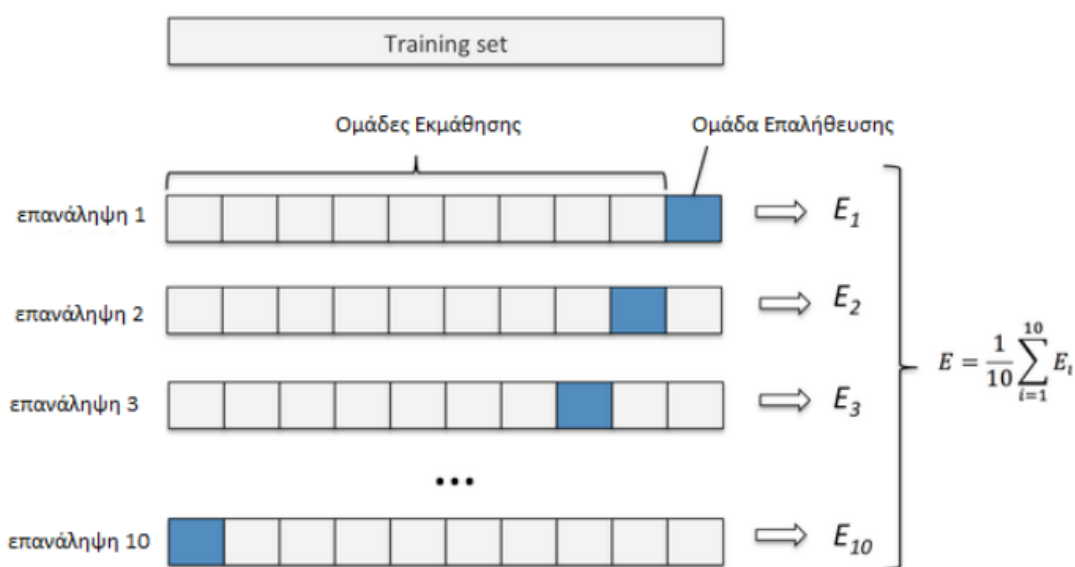
7. Μέσο απόλυτο κανονικοποιημένο σφάλμα (*Mean Absolute Scaled Error*). Πρόκειται για ένα σχετικό μέτρο που στον παρονομαστή χρησιμοποιεί το σφάλμα της Naïve ως σημείο αναφοράς ώστε να μπορεί να γίνει σύγκριση αποδοτικότητας μεταξύ διαφορετικών μεθόδων πρόβλεψης.

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |Y_i - F_i|}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|}$$

3.4.2 Cross validation

Όπως αναφέρθηκε παραπάνω, για την αξιολόγηση ενός μοντέλου πρόβλεψης χρησιμοποιείται το σφάλμα επί των πραγματικών δεδομένων (*out-of-sample*). Ο λόγος είναι ότι το μοντέλο μπορεί να αποδίδει ικανοποιητικά στα διαθέσιμα δεδομένα και πολλές φορές να οδηγεί και σε υπερπροσαρμογή όμως να αστοχεί στις προβλέψεις για μέλλον που μας ενδιαφέρουν.

Προκειμένου να αντιμετωπιστεί το παραπάνω πρόβλημα συχνά εφαρμόζεται η διαδικασία της επαλήθευσης (validation). Για να εφαρμοστεί καλύτερα η διαδικασία διατηρείται ένας ικανοποιητικός αριθμός διαστημάτων από τα διαθέσιμα δεδομένα για την επαλήθευση. Επομένως, τα δεδομένα που θα χρησιμοποιηθούν για την εκπαίδευση του μοντέλου χωρίζονται σε n ομάδες, από τις οποίες κάθε φορά χρησιμοποιούνται οι $n-1$ για την εκπαίδευση του μοντέλου και 1 για την αξιολόγηση του. Η αξιολόγηση σε κάθε επιμέρους ομάδα γίνεται βάσει του στατιστικού σφάλματος που έχει επιλεγεί, ενώ το τελικό σφάλμα προκύπτει ως ο μέσος όρος των επιμέρους. Η διαδικασία απεικονίζεται σχηματικά και παρακάτω.



Εικόνα 14: Παράδειγμα επαλήθευσης για $n=10$ ομάδες

Η παραπάνω διαδικασία εφαρμόζεται κατά κύριο λόγο σε προβλήματα κατηγοριοποίησης. Στην περίπτωση προβλημάτων χρονοσειράς, η χρήση της δεν συνιστάται διότι τα δεδομένα της ομάδας i που αφαιρείται μπορεί να επηρεάζει την ομάδα $i+1$. Για τον λόγο αυτό, στα προβλήματα παλινδρόμησης χρησιμοποιείται η αλυσιδωτή επαλήθευση για τον διαχωρισμό και την αξιολόγηση των επιμέρους ομάδων, ενώ στο τέλος υπολογίζεται ο μέσος όρος. Η τροποποίηση της μεθόδου φαίνεται στο ακόλουθο σχήμα:



Εικόνα 15: Παράδειγμα αλυσιδωτής επαλήθευσης

Η παραπάνω διαδικασία εφαρμόστηκε και στην παρούσα εργασία για την αξιολόγηση των μεθόδων που υλοποιήθηκαν. Σε κάθε επιμέρους ομάδα, υπολογίστηκε ο δείκτης *sMAPE* ενώ το τελικό σφάλμα προέκυψε ως ο μέσος όρος των επιμέρους.

3.5 Σταδιακή παλινδρόμηση (Stepwise regression)

Στα μοντέλα πρόβλεψης που αναλύθηκαν παραπάνω αλλά και στην ανάλυση των προηγούμενων κεφαλαίων γίνεται αναφορά για πρόβλεψη του υπό μελέτη μεγέθους συναρτήσει κάποιων ανεξάρτητων μεταβλητών. Το ερώτημα που τίθεται είναι ποιες από τις ανεξάρτητες μεταβλητές που έχει επιλέξει ο αναλυτής συνεισφέρουν πράγματι στο μοντέλο. Το ερώτημα αυτό έρχεται να επιλύσει η μέθοδος της σταδιακής παλινδρόμησης. Η μέθοδος αυτή αποσκοπεί στην επιλογή των ανεξάρτητων μεταβλητών που θα χρησιμοποιηθούν τελικά από το μοντέλο παλινδρόμησης. Σε κάθε βήμα της μεθόδου, βάσει κάποιου προκαθορισμένου κριτηρίου μια μεταβλητή προστίθεται επομένως πρόκειται για προς τα εμπρός σταδιακή παλινδρόμηση (*forward stepwise regression*) ή αφαιρείται επομένως είναι η προς τα πίσω σταδιακή παλινδρόμηση (*backward stepwise regression*).

Πιο αναλυτικά, κατά την επιλογή προς τα μπροστά το αρχικό μοντέλο εφαρμόζεται χωρίς μεταβλητές εισόδου. Στην συνέχεια, στην πρώτη επανάληψη δοκιμάζεται η προσθήκη κάθε μίας εκ των μεταβλητών και αξιολογείται το αποτέλεσμα βάσει κάποιου κριτηρίου, συνήθως βάσει κάποιου σφάλματος. Αν η προσθήκη κάποιας μεταβλητής βελτιώσει το σφάλμα σε σχέση με αυτό του αρχικού μοντέλου, τότε προστίθενται στο μοντέλο και στην δεύτερη επανάληψη γίνεται εκ νέου έλεγχος για την προσθήκη κάθε μιας από τις υπόλοιπες. Η διαδικασία συνεχίζει κατά αυτόν τον τρόπο και τερματίζεται όταν η προσθήκη μιας νέας μεταβλητής χειροτερεύει το σφάλμα αντί να βελτιώνεται σε σχέση με την προηγούμενη επανάληψη.

Από την άλλη πλευρά, κατά την προς τα πίσω σταδιακή παλινδρόμηση, το αρχικό μοντέλο περιέχει όλες τις μεταβλητές εισόδου. Στην πρώτη επανάληψη αφαιρείται κάθε φορά μία από τις μεταβλητές εισόδου και συγκρίνεται το σφάλμα βάσει του αρχικού. Αν η αφαίρεση κάποιας μεταβλητής βελτίωσε το σφάλμα τότε αφαιρείται από το μοντέλο και συνεχίζεται η διαδικασία με τις υπόλοιπες μεταβλητές. Αντίστοιχα, η διαδικασία τερματίζεται όταν η αφαίρεση μιας μεταβλητής χειροτερεύει το σφάλμα αντί να το βελτιώνει.

Συμπληρωματικά με τις δύο παραπάνω μεθόδους, υπάρχει και η αμφίδρομη διαδικασία κατά την οποία σε κάθε βήμα γίνεται έλεγχος για μεταβλητές που θα συμπεριληφθούν ή θα αφαιρεθούν.

Κατά την εφαρμογή της σταδιακής παλινδρόμησης, αν η αξιολόγηση του μοντέλου γίνεται βάσει κάποιου in-sample σφάλματος ή δείκτη είναι πολύ πιθανόν να οδηγήσει σε υπερπροσαρμογή στα δεδομένα εκμάθησης. Προκειμένου, να αποφευχθεί αυτό εφαρμόζεται ένας συνδυασμός της σταδιακής παλινδρόμησης με την διαδικασία που εφαρμόστηκε στην προηγούμενη υποενότητα. Δηλαδή, κατά την εξέταση για την αφαίρεση ή την προσθήκη της μεταβλητής που είναι προς έλεγχο εφαρμόζεται αλυσιδωτή αξιολόγηση ώστε να αξιολογηθεί η απόδοση του μοντέλου σε ένα εύρος πραγματικών δεδομένων. Επομένως, η αξιολόγηση σχετικά με την επίδραση της αφαίρεσης ή της προσθήκης γίνεται βάσει του τελικού σφάλματος της αλυσιδωτής αξιολόγησης.

4

Πειραματική Διαδικασία

Αντικείμενο της παρούσας εργασίας, όπως έχει αναφερθεί και σε προηγούμενα κεφάλαια είναι η ανάπτυξη μεθόδων πρόβλεψης της τιμής της ηλεκτρικής ενέργειας της επόμενης ημέρας για την αγορά της Ουγγαρίας καθώς και συγκριτική μελέτη των αποτελεσμάτων με σκοπό να γίνει η τελική επιλογή μοντέλου και διαδικασίας. Επομένως, στόχος είναι η παραγωγή ωριαίων προβλέψεων για την επόμενη ημέρα, δηλαδή είκοσι τέσσερις προβλέψεις.

Η επεξεργασία των δεδομένων και η ανάπτυξη των μοντέλων έγινε σε γλώσσα προγραμματισμού R μέσω του περιβάλλοντος Rstudio. Συγκεκριμένα, χρησιμοποιήθηκαν ενδεικτικά οι ακόλουθες βιβλιοθήκες: `library(lubridate)` για επεξεργασία τιμών σε μορφή ημερομηνίας, `library(DT)` για διαχείριση πινάκων, `library(forecast)` για την μέθοδο πρόβλεψης της παλινδρόμησης, `library(plyr)` για συνάθροιση δεδομένων, `library(ggplot2)` για οπτικοποίηση δεδομένων, `library(doBy)` για άθροισμα των δεδομένων ανά μέσο όρο, `library(corrplot)` για διάγραμμα συσχέτισης, `library(gbm)` για την υλοποίηση του αλγορίθμου Gradient Boosting Machine, `library(randomForest)` για την υλοποίηση του αλγορίθμου Random Forest,

library(xlsx) για εξαγωγή των δεδομένων σε αρχείο excel, library(GGally) για διάγραμμα συσχέτισης, library(RSNNS) για την υλοποίηση του αλγορίθμου MLP.

4.1 Παρουσίαση δεδομένων

Στην ανάπτυξη των μοντέλων πρόβλεψης της παρούσας εργασίας έχουν χρησιμοποιηθεί ιστορικά δεδομένα τόσο για την αγορά της υπό μελέτη χώρας όσο και για γειτονικές αλλά και χώρες που υπάρχει έντονη ανταλλαγή ενέργειας.

Όλα τα δεδομένα που χρησιμοποιήθηκαν για την ανάπτυξη της μελέτης αφορούν σχεδόν 1 έτος, και συγκεκριμένα είναι για το διάστημα 9/3/2019-27/3/2020.

Πιο αναλυτικά, τα ιστορικά δεδομένα της τιμής της ηλεκτρικής ενέργειας για την αγορά της Ουγγαρίας αποτελεί βασικό άξονα για την παραγωγή των προβλέψεων που μας ενδιαφέρουν. Πρόκειται για ωριαία δεδομένα συναρτήσει του χρόνου και μονάδα μέτρησης €/MWh. Τα δεδομένα είναι διαθέσιμα είτε στο χρηματιστήριο της αγοράς HUPX είτε στην πλατφόρμα ENTSO-e Transparency στην οποία αναρτώνται τα πιστοποιημένα δεδομένα για τις Ευρωπαϊκές αγορές ηλεκτρικής ενέργειας σχετικά με το φορτίο, την παραγωγή, την ανταλλαγή ηλεκτρικής ενέργειας, τις τιμές των αγορών και των δημοπρασιών κλπ. Η τιμή της αγοράς ηλεκτρικής ενέργειας της Ουγγαρίας αποτελεί την εξαρτημένη μεταβλητή σε όλα τα μοντέλα που αναπτύχθηκαν στην συνέχεια.

Ως ανεξάρτητες μεταβλητές, επιλέχθηκαν δεδομένα φορτίου αλλά και παραγωγής από ΑΠΕ, φωτοβολταϊκά και αιολικά που κατέχουν το μεγαλύτερο μερίδιο τόσο για την Ουγγαρία όσο και για γειτονικές χώρες. Το ηλεκτρικό σύστημα ηλεκτρικής ενέργειας της Ουγγαρίας συνδέεται με διασυνοριακές γραμμές υψηλής τάσης με την Κροατία, τη Σερβία, τη Ρουμανία, την Ουκρανία, την Σλοβακία και την Αυστρία. Από τις χώρες αυτές, για την παρούσα εργασία χρησιμοποιήθηκαν δεδομένα για την Αυστρία, την Σερβία και την Κροατία. Επιπλέον, στην ανάλυση χρησιμοποιήθηκαν δεδομένα και για την χώρα της Γερμανίας καθώς καθημερινά σε βραχυπρόθεσμο και μακροπρόθεσμο επίπεδο λαμβάνουν χώρα συνεχείς ανταλλαγές ηλεκτρικής ενέργειας

που έχουν σημαντική επίδραση στην διαμόρφωση της τιμής της ηλεκτρικής ενέργειας καθώς επίσης και για την Σλοβενία, διότι παρόλο που δεν υπάρχει άμεση διασύνδεση γίνονται εισαγωγές μέσω της Αυστρίας και της Κροατίας. Συγκεκριμένα, τα δεδομένα που χρησιμοποιήθηκαν είναι τα ακόλουθα:

1. Πρόβλεψη φορτίου για την επόμενη μέρα για την Ουγγαρία
(Total_Load_Forecast_Hungary)
2. Πρόβλεψη φορτίου για την επόμενη μέρα για την Αυστρία
(Total_Load_Forecast_Austria)
3. Πρόβλεψη φορτίου για την επόμενη μέρα για την Γερμανία
(Total_Load_Forecast_Germany)
4. Πρόβλεψη φορτίου για την επόμενη μέρα για την Σερβία
(Total_Load_Forecast_Serbia)
5. Πρόβλεψη φορτίου για την επόμενη μέρα για την Κροατία
(Total_Load_Forecast_Croatia)
6. Πρόβλεψη φορτίου για την επόμενη μέρα για την Σλοβενία
(Total_Load_Forecast_Slovenia)
7. Πρόβλεψη παραγωγής φωτοβολταϊκών για την επόμενη μέρα για την Ουγγαρία (Generation Forecast for Solar_Hungary)
8. Πρόβλεψη παραγωγής φωτοβολταϊκών για την επόμενη μέρα για την Αυστρία
(Generation Forecast for Solar_Austria)
9. Πρόβλεψη παραγωγής φωτοβολταϊκών για την επόμενη μέρα για την Γερμανία (Generation Forecast for Solar_Germany)
10. Πρόβλεψη παραγωγής χερσαίων αιολικών για την επόμενη μέρα για την Ουγγαρία (Generation Forecast for Wind_Hungary)
11. Πρόβλεψη παραγωγής χερσαίων αιολικών για την επόμενη μέρα για την Αυστρία (Generation Forecast for Wind_Austria)
12. Πρόβλεψη παραγωγής χερσαίων αιολικών για την επόμενη μέρα για την Γερμανία. (Generation Forecast for Wind Onshore_Germany)
13. Πρόβλεψη παραγωγής υπεράκτιων αιολικών για την επόμενη μέρα για την Γερμανία. (Generation Forecast for Wind Offshore_Germany)

Για το δίκτυο της Σερβίας, της Σλοβενίας και της Κροατίας δεν υπάρχουν διαθέσιμα δεδομένα για την παραγωγή των ΑΠΕ καθώς σύμφωνα με την πλατφόρμα ENTSO-e τα δεδομένα παρέχονται για όλες τις ζώνες υποβολής προσφορών μόνο σε κράτη μέλη με τροφοδοσία άνω του 1% της παραγωγής αιολικής ή ηλιακής ενέργειας ετησίως ή για ζώνες υποβολής προσφορών με περισσότερο από 5% τροφοδοσία παραγωγής αιολικής ή ηλιακής ενέργειας ετησίως.

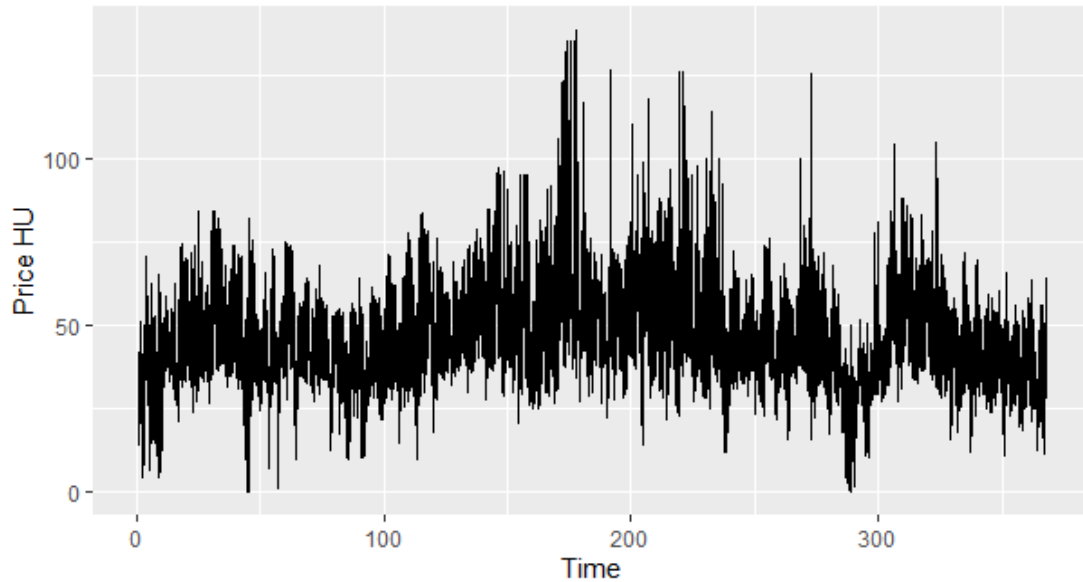
Επιπλέον, των παραπάνω δεδομένων ως ανεξάρτητη μεταβλητή χρησιμοποιήθηκαν και οι αργίες στην χώρα της Ουγγαρίας (Holidays for Hungary), καθώς θεωρείται ότι τόσο τις αργίες όσο και τα Σαββατοκύριακα το δίκτυο αλλά και αγορά έχει μια ιδιαίτερη συμπεριφορά. Για τον λόγο αυτό, εκτός από τις αργίες και για τα Σαββατοκύριακα χρησιμοποιήθηκε βοηθητική μεταβλητή που αναφέρεται στην συνέχεια.

Τα δεδομένα που αναφέρθηκαν παραπάνω έχουν ληφθεί σε ωριαία βάση, ενώ συγκεκριμένα τα δεδομένα για το φορτίο και την παραγωγή των ΑΠΕ της Ουγγαρίας, της Αυστρίας και της Γερμανία είναι διαθέσιμα ανά 15λεπτα οπότε υπολογίστηκε ο μέσος όρος ανά ώρα ώστε να είναι όλα στην ίδια βάση.

Προκειμένου να είναι εύκολη η απόκτηση των παραπάνω δεδομένων αναπτύχθηκε κώδικας σε γλώσσα Python για το κατέβασμα των δεδομένων από την πλατφόρμα ENTSO-e, την ανάγνωση τους, το άθροισμα κατά μέσο όρο στις περιπτώσεις που αυτό χρειαζόταν και την αποθήκευσή τους σε excel αρχείο ώστε στην συνέχεια να χρησιμοποιηθούν στα μοντέλα που αναπτύχθηκαν.

Αναφορικά, με την χρονοσειρά της τιμής της ηλεκτρικής ενέργειας, δεν υπήρχε καμία κενή τιμή οπότε δεν χρειάστηκε να γίνει κάποια διαδικασία σχετικά με συμπλήρωση κενών τιμών. Όσον αφορά, τα υπόλοιπα δεδομένα, υπήρχαν κάποιες κενές τιμές στο φορτίο της Ουγγαρίας και της Γερμανίας παρόλα αυτά επειδή πρόκειται για ανεξάρτητες μεταβλητές επιλέχθηκε να αφαιρεθούν οι γραμμές αυτές από τα ιστορικά δεδομένα και να μην γίνει κάποια διαδικασία για την

συμπλήρωσή τους. Στην συνέχεια φαίνεται η χρονοσειρά της τιμής της ηλεκτρικής ενέργειας.

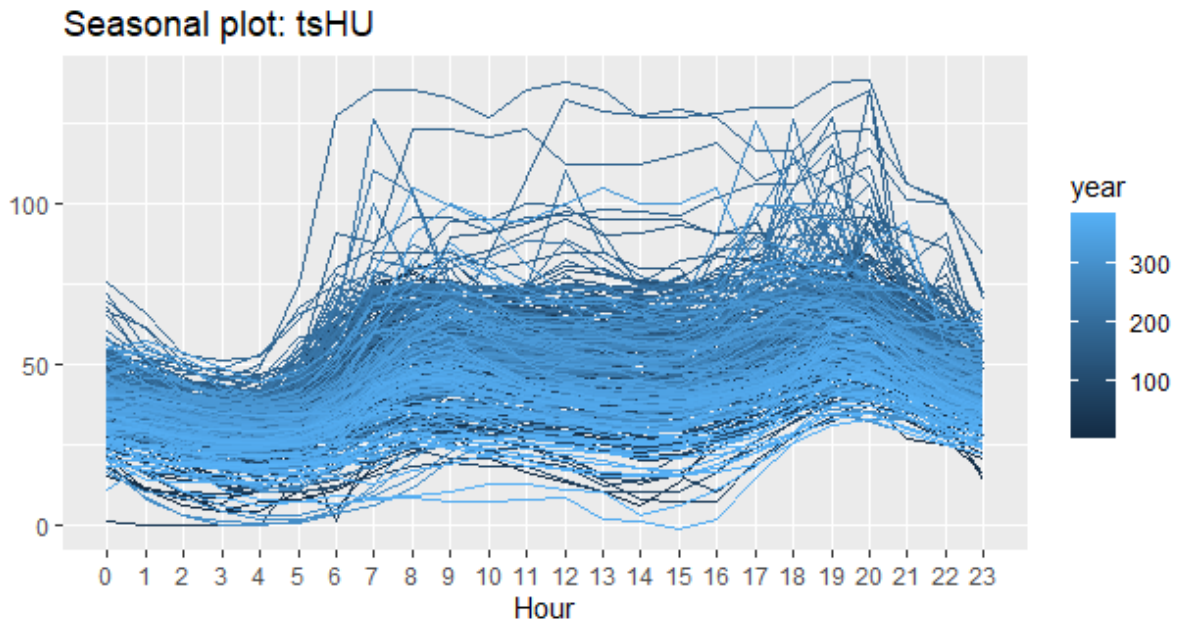


Εικόνα 16: Χρονοσειρά τιμής ηλεκτρικής ενέργειας για την αγορά της Ουγγαρίας

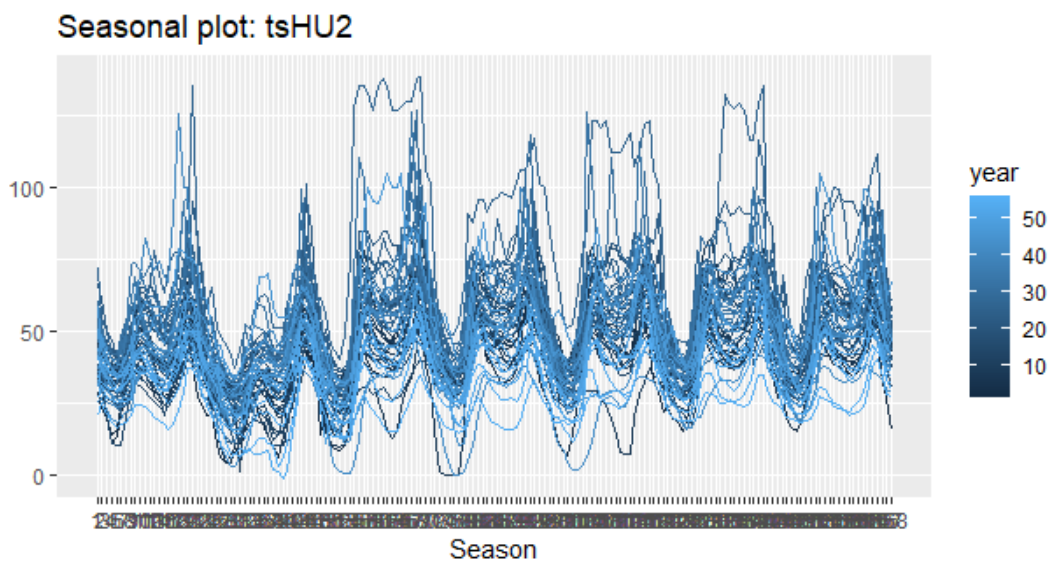
Όπως φαίνεται και από το παραπάνω διάγραμμα, η τιμή της ηλεκτρικής ενέργειας παρουσιάζει αρκετή μεταβλητότητα, κάποιες ασυνήθιστες τιμές καθώς και πολλαπλή εποχικότητα. Η χρονοσειρά δεν παρουσιάζει τάση αλλά σύγκλιση προς τη μέση τιμή. Η εποχικότητα εντοπίζεται όπως έχει αναφερθεί σε προηγούμενο κεφάλαιο τόσο σε επίπεδο ημερήσιο όσο και σε επίπεδο εβδομαδιαίο και αυτό γίνεται εμφανές στα ακόλουθα δύο διαγράμματα. Στο πρώτο διάγραμμα έχει γίνει γραφική αναπαράσταση της υπό μελέτης χρονοσειράς σε ημερήσια συχνότητα ενώ στο δεύτερο σε εβδομαδιαία συχνότητα. Από το πρώτο διάγραμμα, παρατηρούμε ότι τις πρώτες ώρες της ημέρας η τιμή της ενέργειας διατηρείται σε χαμηλά επίπεδα το οποίο εξηγείται από την μειωμένη ζήτηση ηλεκτρικής ενέργειας τις ώρες αυτές. Τις ώρες αιχμής της ημέρας παρατηρείται αύξηση στην τιμή και είναι και οι ώρες που έχουν εμφανίσει κάποιες ασυνήθιστες τιμές, ενώ τις τελευταίες ώρες της ημέρας η τιμή πέφτει εκ νέου.

Αντίστοιχα, στο δεύτερο διάγραμμα φαίνεται το ημερήσιο μοτίβο που περιεγράφηκε προηγουμένως να ακολουθείται κάθε μέρα της εβδομάδας. Επιπλέον, παρατηρείται

ότι τα δεδομένα ξεκινούν από 9/3/2019 που είναι ημέρα Σάββατο επομένως, οι δύο πρώτες επιμέρους κυματομορφές που δημιουργήθηκαν αντιστοιχούν στην ημέρα Σάββατο και Κυριακή αντίστοιχα και φαίνεται μείωση στην τιμή σε σχέση με τις καθημερινές και ιδιαίτερα για την ημέρα Κυριακή (2^η κυματομορφή).



Εικόνα 17: Αναπαράσταση χρονοσειράς τιμής με συχνότητα ημερήσια



Εικόνα 18: Αναπαράσταση χρονοσειράς τιμής με συχνότητα εβδομαδιαία

Ξεκινώντας την ανάλυση των δεδομένων, προστέθηκαν κάποιες βοηθητικές ανεξάρτητες μεταβλητές που θεωρείται ότι βοηθούν στην εκπαίδευση του μοντέλου. Συγκεκριμένα είναι οι ακόλουθες:

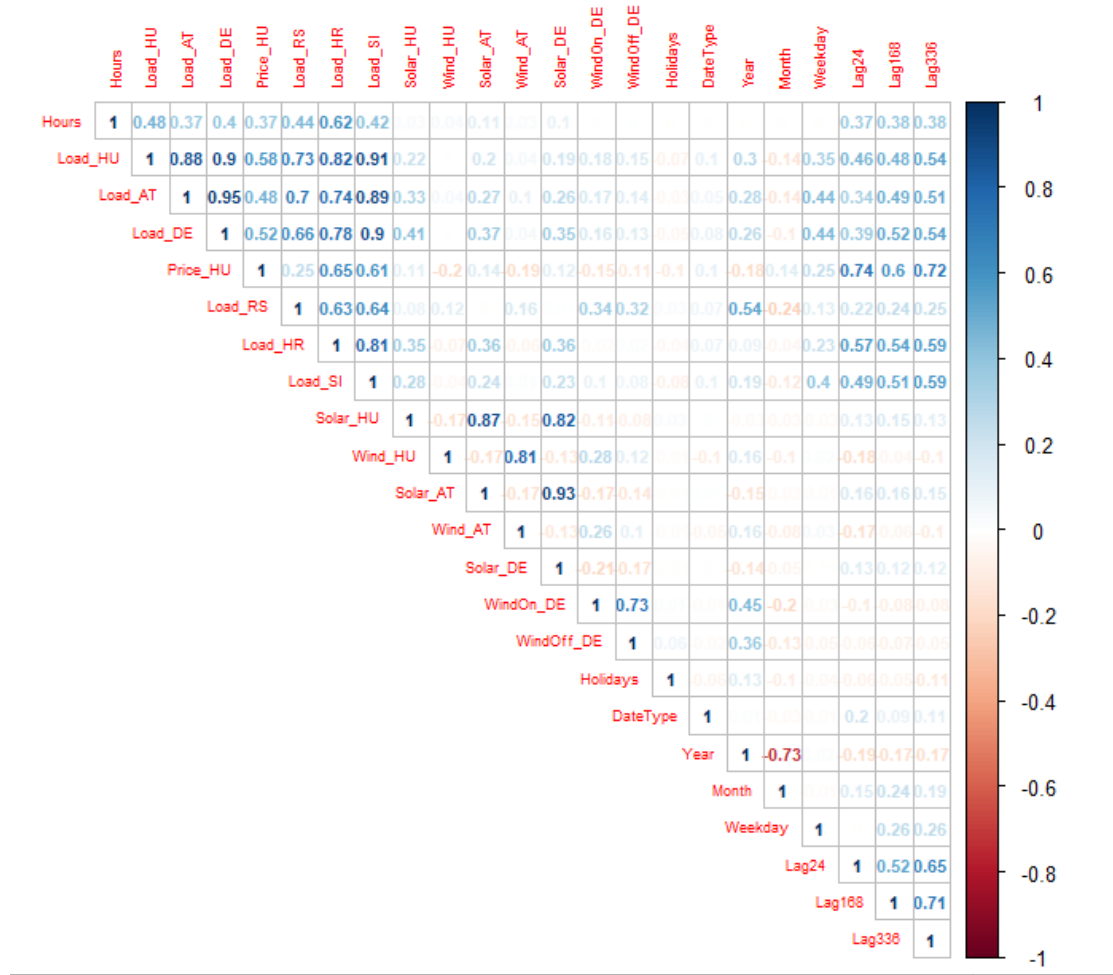
1. Ωρες (Hours)
2. Τύπος ημέρας (Date_type), δηλαδή ποια ημέρας της εβδομάδας είναι η αντίστοιχη ημερομηνία, με τιμές από 1-7. Η τιμή 1 αντιστοιχεί στην Κυριακή ενώ η τιμή 7 στο Σάββατο.
3. Έτος (Year)
4. Μήνας (Month)
5. Καθημερινή (Weekday), δηλαδή αν είναι εργάσιμη ημέρα ή Σαββατοκύριακο. Ο διαχωρισμός αυτός έγινε με την βοήθεια της πρώτης βοηθητικής μεταβλητής. Συγκεκριμένα, στις ημέρες 2-6 δόθηκε η τιμή 1, ενώ στις ημέρες 1 και 7 η τιμή 0.
6. Lag24, η τιμή της προηγούμενης μέρας την ίδια ώρα
7. Lag168, η τιμή της προηγούμενης εβδομάδας την ίδια ώρα
8. Lag336, η τιμή των 2 προηγούμενων εβδομάδων την ίδια ώρα

Επομένως, βάσει όλων των παραπάνω προκύπτει ένα σύνολο δεδομένων (data set) με εξαρτημένη μεταβλητή της τιμή της ηλεκτρικής ενέργειας για την αγορά της Ουγγαρίας και είκοσι δύο (22) ανεξάρτητες μεταβλητές.

Παρακάτω, παρουσιάζεται διάγραμμα συσχέτισης των μεταβλητών για το σύνολο των δεδομένων. Κάθε ένα κελί παρουσιάζει τον συντελεστή συσχέτισης μεταξύ των δύο μεταβλητών. Οι πίνακες συσχέτισης προσφέρουν την δυνατότητα παρουσίασης των συντελεστών συσχέτισης μεταξύ ενός σχετικά μεγάλου αριθμού συνεχών μεταβλητών. Το διάγραμμα έχει δημιουργηθεί μέσω της συνάρτησης *corrplot* από το πακέτο “*corrplot*”.

Αξίζει να σημειωθεί ότι το διάγραμμα προέκυψε για το σύνολο των δεδομένων αφού δημιουργήθηκαν οι βοηθητικές μεταβλητές και αφού αφαιρέθηκαν οι κενές γραμμές.

Όπως αναφέρθηκε και παραπάνω σε κάποιες ανεξάρτητες μεταβλητές όπως είναι το φορτίο της Ουγγαρίας και της Γερμανίας για κάποιες ώρες ή για κάποιες ημέρες υπήρχαν κενές τιμές. Επομένως, επιλέχθηκε η αφαίρεση των γραμμών για τις οποίες έστω και μια ανεξάρτητη μεταβλητή είχε κενή τιμή. Επιπλέον, κατά την προσθήκη των βοηθητικών μεταβλητών Lag24, Lag168, και Lag336 δημιουργήθηκαν κενές τιμές για τις πρώτες γραμμές του δείγματος. Για παράδειγμα για τις δύο πρώτες εβδομάδες του δείγματος δεν υπήρχαν τιμές των δύο προηγούμενων εβδομάδων επομένως η μεταβλητή Lag336 για τις δύο πρώτες εβδομάδες έχει κενές τιμές. Και στην περίπτωση αυτή, επιλέχθηκε οι γραμμές αυτές να αφαιρεθούν από το σύνολο των δεδομένων ώστε να χρησιμοποιηθεί ένα δείγμα που δεν θα έχει καμία κενή τιμή σε κάποια από τις ανεξάρτητες μεταβλητές.



Εικόνα 19: Διάγραμμα συσχέτισης μεταβλητών

Από το παραπάνω διάγραμμα, φαίνεται βάσει του βαθμού συσχέτισης ότι η τιμή της ηλεκτρικής ενέργειας στην τιμή της Ουγγαρίας επηρεάζεται από άλλους παράγοντες περισσότερο επομένως η τιμή είναι του συντελεστή συσχέτισης είναι μεγάλη ενώ από άλλους παράγοντες λιγότερο. Για παράδειγμα, παρατηρούμε ότι υπάρχει μεγάλη συσχέτιση με τα Lag24, Lag168, Lag336, και τα φορτία των χωρών ενώ η συσχέτιση είναι πολύ μικρότερη με το έτος και τον μήνα. Επίσης, παρατηρούμε ότι με κάποιες μεταβλητές ο βαθμός συσχέτισης είναι αρνητικός. Αυτό συνεπάγεται ότι η αύξηση της μιας μεταβλητής οδηγεί σε μείωση στην άλλη. Από τα παραπάνω φαίνεται η ανάγκη της εφαρμογής της σταδιακής παλινδρόμησης όπως αναφέρθηκε στο προηγούμενο κεφάλαιο ώστε να εξεταστεί ποιες μεταβλητές είναι πράγματι χρήσιμες για την παραγωγή των προβλέψεων.

Όπως έχει ήδη αναφερθεί, στόχος του συγκεκριμένου προβλήματος είναι η πρόβλεψη της τιμής ενέργειας ανά ώρα για την επόμενη μέρα για την αγορά της Ουγγαρίας. Για την υλοποίηση του προβλήματος, έγινε διαχωρισμός του data set μετά την αφαίρεση των κενών γραμμών όπως περιεγράφηκε παραπάνω. Το πρώτο τμήμα που αντιστοιχεί στο 85% του δείγματος χρησιμοποιήθηκε για την εκμάθηση των αλγορίθμων δηλαδή για training set ενώ το υπόλοιπο 15% για test set ώστε να γίνει αξιολόγηση των αποτελεσμάτων σε πραγματικά δεδομένα. Προκειμένου να γίνει αυτό εφαρμόστηκε η μέθοδος της αλυσιδωτής επαλήθευσης (cross validation). Το data set μετά την αφαίρεση των κενών γραμμών είχε 8710 γραμμές, επομένως οι 7390 (έως και 31/1/2020) χρησιμοποιήθηκαν ως training set και οι υπόλοιπες 1320 (από 1/2/2020) ως test set). Κατά την υλοποίηση του cross validation, στην πρώτη επανάληψη χρησιμοποιήθηκαν οι 7390 γραμμές για εκμάθηση και έγινε πρόβλεψη για τις επόμενες 24 τιμές. Επομένως η αξιολόγηση έγινε βάσει των πρώτων 24 τιμών από το test set. Στην επόμενη επανάληψη των training set αυξήθηκε κατά 24 γραμμές ενώ συγχρόνως το test set μειώθηκε κατά 24. Αντίστοιχα, έγινε πρόβλεψη για τις επόμενες 24 τιμές και η αξιολόγηση έγινε βάσει των πρώτων 24 τιμών του νέου test set. Αυτή η διαδικασία επαναλαμβάνεται έως ότου το training set να γίνει $7390+1320=8710$. Για την αξιολόγηση επιλέχθηκε ο στατιστικός δείκτης σφάλματος sMAPE ενώ το τελικό σφάλμα προέκυψε από τον μέσο όρο των σφαλμάτων των επαναλήψεων του cross validation. Επομένως, η τελική αξιολόγηση γίνεται σε μακροπρόθεσμο επίπεδο βάσει των προβλέψεων για τις 55 ημέρες ($1320/24=55$) σε ρεαλιστικές συνθήκες,

Παρακάτω εφαρμόστηκαν τέσσερις αλγόριθμοι. Πρώτα εφαρμόστηκε η πολλαπλή γραμμική παλινδρόμηση (MLR), στην συνέχεια οι αλγόριθμοι δέντρων απόφασης Gradient Boosting Machine και Random Forest, ενώ στο τέλος υλοποιήθηκε το νευρωνικό μέσω του αλγορίθμου MLP. Σε κάθε αλγόριθμο εφαρμόστηκε η τεχνική της προς τα πίσω σταδιακής παλινδρόμησης (backward stepwise regression) συνδυαστικά με cross validation ώστε να προκύψει το βέλτιστο μοντέλο πρόβλεψης. Αξίζει να σημειωθεί ότι για την περίπτωση των αλγορίθμων Gradient Boosting Machine, MLP και Random Forest πριν την εφαρμογή της διαδικασίας backward stepwise regression, στο αρχικό μοντέλο που περιείχε όλες τις ανεξάρτητες μεταβλητές εισόδου έγινε μια έρευνα για την επιλογή των βέλτιστων υπερπαραμέτρων του αλγορίθμου. Παρακάτω παρουσιάζονται τα αποτελέσματα των αλγορίθμων και των επιμέρους διαδικασιών που εφαρμόστηκαν.

4.2 Πολλαπλή γραμμική παλινδρόμηση (MLR)

Ο πρώτος αλγόριθμος που εφαρμόστηκε και εξετάστηκε είναι αυτός της πολλαπλής γραμμικής παλινδρόμησης. Ο αλγόριθμος υλοποιήθηκε μέσω της συνάρτησης lm από το πακέτο “forecast” της γλώσσας R. Το πρώτο μοντέλο που εφαρμόστηκε είχε όλες τις ανεξάρτητες μεταβλητές (22 στο σύνολο) για την πρόβλεψη της εξαρτημένης.

All predictors:

Hours+Total_Load_Forecast_Hungary.mean+Total_Load_Forecast_Austria.mean+Total_Load_Forecast_Germany.mean+Total_Load_Forecast_Croatia+Total_Load_Forecast_Serbia+Total_Load_Forecast_Slovenia+Solar_HU+Wind_HU+Solar_AT+Wind_AT+Solar_DE+WindOn_DE+WindOff_DE+Holidays+Weekday+DateType+Year+Month+Lag24+Lag168+Lag336

Από το παραπάνω μοντέλο και την εφαρμογή της διαδικασίας cross validation όπως περιεγράφηκε παραπάνω προέκυψε ένα συνολικό σφάλμα $sMAPE = 17.5178\%$.

Στην συνέχεια εφαρμόστηκε η διαδικασία backward stepwise regression ώστε να εξεταστεί η επίδραση της αφαίρεσης μίας μεταβλητής κάθε φορά από το μοντέλο. Τα αποτελέσματα της διαδικασίας φαίνονται παρακάτω. Στην πρώτη στήλη του

παρακάτω πίνακα αναφέρεται η μεταβλητή που αφαιρείται κάθε φορά από το μοντέλο. Ενώ στις επόμενες στήλες, φαίνεται το συνολικό σφάλμα sMAPE για κάθε μοντέλο έπειτα από την αφαίρεση της εκάστοτε μεταβλητής. Στον πίνακα έχει γίνει σήμανση κάθε φορά των τιμών που είναι μικρότερες από το μοντέλο της προηγούμενης επανάληψης ενώ στην τελευταία γραμμή φαίνεται το ελάχιστο σφάλμα για κάθε επανάληψη. Η μεταβλητή που αντιστοιχεί στο ελάχιστο σφάλμα αφαιρείται από το μοντέλο και στην επόμενη επανάληψη γίνεται η ίδια διαδικασία, έλεγχος για τις υπόλοιπες μεταβλητές. Αξίζει να σημειωθεί ότι η μεταβλητή ώρες (Hours) θεωρήθηκε ιδιαίτερα σημαντική στο μοντέλο πρόβλεψης για την παραγωγή των ωριαίων προβλέψεων που επιθυμούμε οπότε δεν εξετάστηκε στην παραπάνω διαδικασία.

παραμέτρος που αφορείται	SMAPE										SMAPE													
	MLR 1st	MLR 2nd	MLR 3rd	MLR 4th	MLR 5th	MLR 6th	MLR 7th	MLR 8th	MLR 9th	MLR 10th	MLR 11th	MLR 12th	MLR 1st	MLR 2nd	MLR 3rd	MLR 4th	MLR 5th	MLR 6th	MLR 7th	MLR 8th	MLR 9th	MLR 10th	MLR 11th	MLR 12th
Date Type	17,581%	17,060%	16,413%	16,311%	16,223%	16,047%	15,891%	15,759%	15,352%	15,314%														
Lag336	17,683%	17,236%	16,628%	16,477%	16,287%	16,039%	15,901%	15,785%	15,329%															
Lag168	18,513%	18,044%	17,461%	17,432%	17,142%	16,968%	16,826%	16,704%	16,318%	16,712%	16,669%	16,755%												
Lag24	20,201%	19,327%	18,442%	18,255%	18,291%	18,502%	18,402%	18,211%	17,473%	17,606%	17,592%	17,572%												
Month	18,167%	17,677%	17,044%	16,908%	16,772%	16,648%	16,483%	16,337%	15,800%	15,702%	15,797%	15,797%												
Year	19,025%	19,332%	18,582%	18,519%	18,763%	18,566%	18,425%	18,443%	18,554%	18,520%	15,774%	18,946%												
Weekday	17,518%	17,055%	16,359%	16,201%																				
Holidays	17,344%	16,970%	16,348%																					
WindOff_DE	17,435%	17,067%	16,411%	16,285%	16,096%	15,957%	15,827%	15,402%																
WindOn_DE	17,440%	16,995%	16,350%	16,219%	16,103%	15,945%	15,789%																	
Solar_DE	17,571%	17,186%	16,516%	16,395%	16,269%	16,124%	15,926%	15,805%	15,406%	15,330%	15,371%	15,369%												
Wind_AT	17,598%	17,184%	16,440%	16,307%	16,152%	16,007%	15,881%	15,807%	15,458%	15,384%	15,334%	15,333%												
Solar_AT	17,162%	16,857%	16,356%	16,249%	16,084%	15,910%																		
Wind_HU	17,565%	17,170%	16,524%	16,400%	16,246%	16,102%	15,943%	15,794%	15,394%	15,319%	15,353%	15,352%												
Solar_HU	17,566%	17,188%	16,542%	16,424%	16,375%	16,201%	16,115%	16,048%	15,694%	15,597%	15,557%	15,550%												
Total_Load_Forecast_Slovenia	17,510%	17,124%	16,400%	16,270%	16,128%	16,035%	15,907%	15,786%	15,399%	15,331%	15,314%													
Total_Load_Forecast_Serbia	17,129%																							
Total_Load_Forecast_Croatia	17,707%	17,557%	17,009%	16,817%	16,586%	16,460%	16,221%	16,051%	15,534%	15,527%	15,546%	15,620%												
Total_Load_Forecast_Germany	17,157%	16,657%	16,432%	16,306%	16,047%																			
Total_Load_Forecast_Austria	17,263%	16,472%																						
Total_Load_Forecast_Hungary	17,532%	17,060%	16,455%	16,393%	16,272%	15,845%	15,791%	15,733%	15,372%	15,299%	15,363%	15,427%												
min	17,129%	16,472%	16,348%	16,201%	16,047%	15,845%	15,789%	15,402%	15,329%	15,299%	15,314%	15,333%												

Πίνακας 1: Αποτελέσματα backward stepwise regression αλγορίθμου MLR

Στις επαναλήψεις 6 και 10 έχει μικρότερο sMAPE η αφαίρεση της μεταβλητής Total_Load_Forecast_Hungary, παρόλα αυτά, η διαφορά με την αμέσως επόμενη μεταβλητή που αφαιρείται είναι πολύ μικρή. Καθώς το φορτίο της χώρας που μελετάται θεωρείται σημαντική μεταβλητή επιλέχθηκε η αφαίρεση της αμέσως επόμενης.

Από την παραπάνω διαδικασία, το σφάλμα του τελικού μοντέλου είναι ίσο με 15,3137%. Επομένως, επιτεύχθηκε μία μείωση στο σφάλμα κατά 2,2 μονάδες το οποίο αντιστοιχεί σε ποσοστό 12,582%. Το ποσοστό υπολογίστηκε από την ακόλουθη σχέση:

$$\text{Μεταβολή} = \frac{\text{αρχικό} - \text{τελικό}}{\text{αρχικό}}$$

Το τελικό μοντέλο του αλγορίθμου έχει τις ακόλουθες μεταβλητές εισόδου:

Predictors:

Hours+Total_Load_Forecast_Hungary.mean+Total_Load_Forecast_Croatia+Solar_HU+Wind_HU+Wind_AT+Solar_DE+ Year+Month+Lag24+Lag168

4.3 Gradient Boosting Machine

Ο δεύτερος αλγόριθμος που υλοποιήθηκε στην παρούσα εργασία είναι ο Gradient Boosting Machine. Όπως αναφέρθηκε και παραπάνω, πριν την εφαρμογή της διαδικασίας backward stepwise regression εφαρμόστηκε μια έρευνα σχετικά με τις τιμές των υπερπαραμέτρων στο αρχικό μοντέλο που περιέχει όλες τις ανεξάρτητες μεταβλητές εισόδου. Ο αλγόριθμος υλοποιήθηκε μέσω της συνάρτησης *gbm* από το πακέτο “*gbm*” της γλώσσας R.

Συγκεκριμένα, η έρευνα που έγινε στις υπερπαραμέτρους αφορούσε την κατανομή που εφαρμόζεται κατά την αξιολόγηση όπως αναλύθηκε στην ενότητα 3.3.3.1, των αριθμό των δέντρων, το μέγιστο βάθος κάθε δέντρου, δηλαδή το μέγιστο επιτρεπόμενο επίπεδο μεταβλητών αλληλεπίδρασης και τέλος την συρρίκνωση που

εφαρμόζεται σε κάθε δέντρο κατά την επέκταση που είναι γνωστή ως ρυθμός εκμάθησης. Σχετικά με την συνάρτηση κατανομής, αρχικά επιλέχθηκε να ελεγχθεί και να αξιολογηθεί η κατανομή “laplace” και “gaussian”. Η πρώτη κατανομή δημιούργησε έντονο υπολογιστικό πρόβλημα σε σημείο που ήταν ανέφικτη η εκμάθηση του μοντέλου λόγω αναγκαστικού κλεισίματος του προγράμματος. Επομένως, αφαιρέθηκε στην συνέχεια από την έρευνα και την αξιολόγηση. Οι τιμές που δόθηκαν τελικά στις υπόλοιπες υπερπαραμέτρους φαίνονται στον ακόλουθο πίνακα:

distribution	gaussian
n.trees	500
	1000
	2000
shrinkage	0.1
	0.01
	0.001
interaction.depth	2
	3
	4

Πίνακας 2: Συνδυασμοί υπερπαραμέτρων αλγορίθμου gbm

Τα αποτελέσματα των συνδυασμών που φαίνονται στον παραπάνω πίνακα ομαδοποιημένα βάσει του αριθμού των δέντρων που επιλέχθηκε κάθε φορά παρουσιάζονται συγκεντρωτικά παρακάτω. Αξίζει να σημειωθεί ότι από την παραπάνω ανάλυση παρατηρήθηκε ότι όσο αυξάνεται ο αριθμός των δέντρων και ο ρυθμός εκμάθησης αυξάνεται σημαντικά το υπολογιστικό κόστος. Για παράδειγμα, για n.trees=500, ο βέλτιστος συνδυασμός για την διαδικασία cross validation χρειάστηκε 3 λεπτά για την εκμάθηση, για n.trees=1000 χρειάστηκε 6 λεπτά, ενώ για n.trees=2000 χρειάστηκε 15λεπτά. Επιπλέον, όταν εφαρμόστηκαν εκ νέου οι βέλτιστοι συνδυασμοί για κάθε τιμή της μεταβλητής n.trees προέκυψαν λίγο διαφοροποιημένα αποτελέσματα για την ακρίβεια για n.trees=500 15.96796%, n.trees=1000 15.4653% ενώ για n.trees=2000 15.37302%. Αυτό δικαιολογείται από το γεγονός ότι τα δέντρα αποφάσεων περιέχουν τον παράγοντα της τύχης στην επίλυση τους. Επίσης, στην περίπτωση του n.trees=500 επειδή κατά την δεύτερη

δοκιμή το σφάλμα προέκυψε μεγαλύτερο από τον δεύτερο καλύτερο συνδυασμό δοκιμάστηκε εκ νέου ο δεύτερος καλύτερος συνδυασμός για τον οποίο προέκυψε σχεδόν σταθερό σφάλμα ίσο με 15.807%.

	n.trees=500			n.trees=1000			n.trees=2000		
	sMAPE	shrink	inter	sMAPE	shrink	inter	sMAPE	shrink	inter
1	16,252%	0.1	2	15,895%	0.1	2	15,459%	0.1	2
2	15,802%	0.1	3	15,306%	0.1	3	15,301%	0.1	3
3	15,610%	0.1	4	15,374%	0.1	4	14,970%	0.1	4
4	22,123%	0.01	2	19,542%	0.01	2	17,741%	0.01	2
5	21,507%	0.01	3	18,524%	0.01	3	16,811%	0.01	3
6	20,689%	0.01	4	17,934%	0.01	4	16,530%	0.01	4
7	33,581%	0.001	2	30,419%	0.001	2	26,579%	0.001	2
8	33,133%	0.001	3	29,770%	0.001	3	25,810%	0.001	3
9	32,976%	0.001	4	29,343%	0.001	4	25,170%	0.001	4
min	15,610%			15,306%			14,970%		

Πίνακας 3: Αποτελέσματα συνδυασμός υπερπαραμέτρων του αλγορίθμου gbm

Από τα παραπάνω αποτελέσματα συμπεραίνουμε ότι το σφάλμα μειώνεται όσο αυξάνεται ο αριθμός των δέντρων όμως η διαφορά μεταξύ τους δεν είναι σημαντική έναντι του υπολογιστικού κόστους που είναι αρκετά σημαντικό. Επιπλέον, εφόσον δοκιμάζοντας εκ νέου τους καλύτερους συνδυασμούς προέκυψε σφάλμα διαφορετικό από το αρχικό και κάποιες φορές μεγαλύτερο ως καλύτερος συνδυασμός επιλέχθηκε ο ακόλουθος:

```
Hyperparameters values
distribution gaussian
n.trees 500
shrinkage 0.1
interaction.depth 3
```

Αφού έγινε η επιλογή των υπερπαραμέτρων του αλγορίθμου για το αρχικό μοντέλο που περιέχει όλες τις μεταβλητές εισόδου, προχωρήσαμε στην εφαρμογή της διαδικασίας backward stepwise regression και τα αποτελέσματα φαίνονται στον ακόλουθο πίνακα:

sMAPE			
παράμετρος που αφαιρείται	GBM 1st	GBM 2nd	GBM 3rd
DateType	16,010%	15,970%	15,979%
Lag336	15,788%	15,678%	15,608%
Lag168	15,891%	15,759%	15,582%
Lag24	16,479%	16,471%	16,227%
Month	17,099%	17,184%	16,815%
Year	17,007%	17,032%	16,744%
Weekday	15,870%	15,794%	15,366%
Holidays	16,037%	15,151%	
WindOff_DE	15,660%	15,532%	15,378%
WindOn_DE	15,967%	15,501%	15,573%
Solar_DE	16,014%	16,054%	15,839%
Wind_AT	15,472%		
Solar_AT	15,684%	15,599%	15,392%
Wind_HU	15,910%	15,860%	15,797%
Solar_HU	16,377%	15,848%	15,845%
Total_Load_Forecast_Slovenia	16,223%	15,811%	15,639%
Total_Load_Forecast_Serbia	15,951%	15,891%	15,256%
Total_Load_Forecast_Croatia	16,054%	16,013%	16,029%
Total_Load_Forecast_Germany	15,803%	15,567%	15,558%
Total_Load_Forecast_Austria	15,771%	15,690%	15,649%
Total_Load_Forecast_Hungary	15,668%	15,309%	15,459%
min	15,472%	15,151%	15,256%

Πίνακας 4: Αποτελέσματα backward stepwise regression αλγορίθμου GBM

Από τα παραπάνω αποτελέσματα φαίνεται ότι η αφαίρεση μόνο 2 μεταβλητών οδήγησε σε μια βελτίωση του τελικού σφάλματος κατά 0,65645 μονάδες που αντιστοιχεί σε ποσοστό 4,153%. Το παραπάνω αποτέλεσμα επαληθεύει την θεωρία ότι ο αλγόριθμος gbm επιλέγει από μόνους του τις μεταβλητές εισόδου που είναι σημαντικές και βελτιώνουν την ακρίβεια και το αποτέλεσμα της πρόβλεψης.

4.4 Random Forest

Ο επόμενος αλγόριθμος που υλοποιήθηκε στην παρούσα εργασία είναι ο Random Forest. Όπως αναφέρθηκε και παραπάνω, πριν την εφαρμογή της διαδικασίας backward stepwise regression εφαρμόστηκε μια έρευνα σχετικά με τις τιμές των

υπεραπαραμέτρων στο αρχικό μοντέλο που περιέχει όλες τις ανεξάρτητες μεταβλητές εισόδου. Ο αλγόριθμος υλοποιήθηκε μέσω της συνάρτησης *randomForest* από το πακέτο “ *randomForest* ” της γλώσσας R.

Συγκεκριμένα, η έρευνα σχετικά με τις υπερπαραμέτρους για τον αλγόριθμο αυτό αφορούσε αρχικά τον αριθμό των δέντρων καθώς και τον αριθμό των τυχαίων μεταβλητών εισόδου που θα χρησιμοποιήσει πριν από κάθε διαχωρισμό κόμβου. Οι συνδυασμοί των υπερπαραμέτρων που δοκιμάστηκαν φαίνονται στον ακόλουθο πίνακα:

	500
n.tress	1000
	7
	8
	10
	20
	21
mtry	22

Πίνακας 5: Συνδυασμοί υπερπαραμέτρων αλγορίθμου *randomforest*

Σύμφωνα με την βιβλιογραφία και βάσει του ορισμού της συνάρτησης *randomforest* της R η βέλτιστη τιμή για την μεταβλητή *mtry* για προβλήματα regression είναι $p/3$ όπου p ο αριθμός των μεταβλητών εισόδου. Επομένως, για τον παρόν πρόβλημα εφόσον οι μεταβλητές εισόδου είναι 22 προκύπτει βέλτιστη τιμή ίση με 7,333. Επιπλέον, υπάρχει ο ισχυρισμός ότι αν οι μεταβλητές δεν έχουν συσχέτιση μεταξύ τους ίσως θέτοντας το *mtry* ίσο με τον αριθμό των μεταβλητών εισόδου να προκύπτουν καλύτερα αποτελέσματα. Για τους παραπάνω λόγους, επιλέχθηκαν οι συγκεκριμένοι συνδυασμοί για έλεγχο και τα αποτελέσματά τους φαίνονται παρακάτω.

Από την παραπάνω έρευνα, σχετικά με το υπολογιστικό κόστος προκύπτει το συμπέρασμα ότι όσο αυξάνεται ο αριθμός των δέντρων και το *mtry* αντίστοιχα αυξάνεται σημαντικά ο χρόνος επίλυσης και κάποιες φορές διπλασιάζεται κιόλας.

Αξίζει να σημειωθεί ότι ο συγκεκριμένος αλγόριθμος δεν είναι υπολογιστικά αποδοτικός καθώς έχει πολυπλοκότητα $n \cdot \text{trees} \cdot \text{mtry} \cdot \log(\# \text{ objects})$.

	sMAPE	trees	mtry
1	17,257%	500	7
2	17,231%	500	8
3	17,184%	500	10
4	17,266%	1000	7
5	17,204%	1000	8
6	17,169%	1000	10
7	17,194%	500	20
8	17,211%	500	21
9	17,225%	500	22
10	17,166%	1000	20
11	17,173%	1000	21
12	17,210%	1000	22
Min	17,166%		

Πίνακας 6: Αποτελέσματα συνδυασμός υπερπαραμέτρων του αλγορίθμου randomforest

Επομένως για τους παραπάνω λόγους ως βέλτιστος συνδυασμός επιλέχθηκε ο ακόλουθος με σφάλμα 17.184%.

$n.trees = 500$ και $mtry=10$

Αφού έγινε η επιλογή των υπερπαραμέτρων του αλγορίθμου για το αρχικό μοντέλο που περιέχει όλες τις μεταβλητές εισόδου, προχωρήσαμε στην εφαρμογή της διαδικασίας backward stepwise regression και τα αποτελέσματα φαίνονται στον ακόλουθο πίνακα.

sMAPE					
παράμετρος που αφαιρείται	RF 1st	RF 2nd	RF 3rd	RF 4th	RF 5th
DateType	17,035%	16,792%			
Lag336	17,278%	17,079%	16,921%	16,811%	16,739%
Lag168	17,652%	17,434%	17,104%	17,168%	17,061%
Lag24	18,689%	18,406%	18,242%	18,815%	18,646%
Month	17,042%	16,854%	16,589%		
Year	17,446%	17,287%	17,065%	16,903%	16,739%
Weekday	17,296%	17,031%	16,832%	16,682%	16,643%
Holidays	17,140%	16,824%	16,648%	16,494%	
WindOff_DE	16,875%				
WindOn_DE	17,436%	17,366%	17,141%	16,860%	16,852%
Solar_DE	17,199%	16,969%	16,803%	16,678%	16,498%
Wind_AT	17,205%	17,051%	16,724%	16,628%	16,507%
Solar_AT	17,106%	16,978%	16,736%	16,602%	16,494%
Wind_HU	17,067%	16,896%	16,684%	16,530%	16,495%
Solar_HU	17,252%	17,062%	16,830%	16,722%	16,622%
Total_Load_Forecast_Slovenia	17,309%	17,127%	16,880%	17,065%	16,962%
Total_Load_Forecast_Serbia	17,345%	17,215%	16,946%	16,910%	16,877%
Total_Load_Forecast_Croatia	17,365%	17,193%	16,957%	16,832%	16,712%
Total_Load_Forecast_Germany	17,036%	16,869%	16,724%	16,563%	16,496%
Total_Load_Forecast_Austria	17,122%	16,901%	16,730%	16,564%	16,496%
Total_Load_Forecast_Hungary	17,069%	16,865%	16,656%	16,577%	16,540%
min	16,875%	16,792%	16,589%	16,494%	16,494%

Πίνακας 7: Αποτελέσματα backward stepwise regression αλγορίθμου Randomforest

Από τα παραπάνω αποτελέσματα φαίνεται ότι η αφαίρεση μόνο 4 μεταβλητών μείωσε το σφάλμα διότι στην 5η δοκιμή το σφάλμα αυξήθηκε εκ νέου. Συνολικά προέκυψε μία μείωση 0.69 μονάδες από το μοντέλο που περιέχει όλες τις μεταβλητές εισόδου και το τελικό στο οποίο έχουν αφαιρεθεί 4 μεταβλητές. Σε σύγκριση με τους 2 προηγούμενους αλγορίθμους που υλοποιήθηκαν το σφάλμα 16.494% είναι σημαντικά υψηλότερο, επομένως δεν επιλέγεται για την πρόβλεψη που μελετάται.

4.5 Νευρωνικό δίκτυο (MLP)

Το τελευταίο μοντέλο που υλοποιήθηκε είναι το νευρωνικό δίκτυο μέσω του αλγορίθμου Multi-Layer Perceptron (MLP). Όπως αναφέρθηκε και παραπάνω, πριν την εφαρμογή της διαδικασίας backward stepwise regression εφαρμόστηκε μια

έρευνα σχετικά με τις τιμές των υπερπαραμέτρων στο αρχικό μοντέλο που περιέχει όλες τις ανεξάρτητες μεταβλητές εισόδου. Ο αλγόριθμος υλοποιήθηκε μέσω της συνάρτησης mlr από το πακέτο “RSNNS” της γλώσσας R. Όπως αναφέρθηκε και στο κεφάλαιο 3, τα νευρωνικά δίκτυα πριν την εφαρμογή τους στα διαθέσιμα δεδομένα για εκμάθηση απαιτούν την κανονικοποίησή τους. Για τον λόγο αυτό αναπτύχθηκε βοηθητική συνάρτηση η οποία κανονικοποιεί τα δεδομένα σε κλίμακα (0,1) βάσει της μέγιστης και της ελάχιστης τιμής της κάθε μεταβλητής.

Η έρευνα σχετικά με τις υπερπαραμέτρους για τον αλγόριθμο αυτό αφορούσε αρχικά τον αριθμό των στρωμάτων, τον αριθμό των κόμβων σε κάθε στρώμα, τον ρυθμό εκμάθησης για την συνάρτηση εκμάθησης καθώς και την συνάρτηση ενεργοποίησης των κρυμμένων στρωμάτων. Ως συνάρτηση εκμάθησης χρησιμοποιήθηκε η κλασική "Std_Backpropagation" όπως αναφέρθηκε και αναλύθηκε σε προηγούμενο κεφάλαιο. Ενώ για την συνάρτηση ενεργοποίησης δοκιμάστηκε η γραμμική *Act_IdentityPlusBias* και η λογαριθμική (sigmoid) *Act_Logistic*. Οι συνδυασμοί των υπερπαραμέτρων που δοκιμάστηκαν φαίνονται στον ακόλουθο πίνακα:

Hyperparameters	values
layers	1
	3
size	10
	20
	40
learning rate	0.1
	0.01
	0.001
activation function	linear
	sigmoid

Πίνακας 8: Συνδυασμοί υπερπαραμέτρων αλγορίθμου MLP

Τα αποτελέσματα των συνδυασμών που φαίνονται στον παραπάνω πίνακα ομαδοποιημένα βάσει αρχικά της συνάρτησης ενεργοποίησης και έπειτα βάσει του αριθμού των στρωμάτων που επιλέχθηκε κάθε φορά παρουσιάζονται συγκεντρωτικά παρακάτω. Βάσει αποτελεσμάτων, ως καλύτερος συνδυασμός επιλέχθηκε: 1 layer,

size 20, learning rate 0.1 και activation function sigmoid με συνολικό σφάλμα ίσο με 12.9534%.

Linear				Sigmoid			
layers	size	learning rate	sMAPE	layers	size	learning rate	sMAPE
1	10	0.1	13,427%	1	10	0.1	13,038%
		0.01	14,521%			0.01	15,260%
		0.001	18,487%			0.001	33,725%
	20	0.1	13,422%		20	0.1	12,953%
		0.01	14,421%			0.01	14,682%
		0.001	17,630%			0.001	31,222%
	40	0.1	13,398%		40	0.1	13,150%
		0.01	14,061%			0.01	14,490%
		0.001	16,410%			0.001	27,213%
3	(10,10,10)	0.1	13,705%	3	(10,10,10)	0.1	13,368%
		0.01	14,069%			0.01	29,703%
		0.001	17,202%			0.001	38,250%
	(20,20,20)	0.1	13,780%		(20,20,20)	0.1	13,251%
		0.01	13,712%			0.01	20,437%
		0.001	15,765%			0.001	37,093%
	(40,40,40)	0.1	13,959%		(40,40,40)	0.1	13,393%
		0.01	13,636%			0.01	14,838%
		0.001	15,085%			0.001	34,448%
min			13,398%	min			12,953%

Πίνακας 9: Αποτελέσματα συνδυασμός υπερπαραμέτρων του αλγορίθμου MLP

Αφού έγινε η επιλογή των υπερπαραμέτρων του αλγορίθμου για το αρχικό μοντέλο που περιέχει όλες τις μεταβλητές εισόδου, προχωρήσαμε στην εφαρμογή της διαδικασίας backward stepwise regression και τα αποτελέσματα φαίνονται στον ακόλουθο πίνακα.

sMAPE		
παράμετρος που αφαιρείται	NN 1st	NN 2nd
DateType	13,273%	13,237%
Lag336	13,008%	12,806%
Lag168	13,164%	13,188%
Lag24	13,061%	13,068%
Month	13,542%	13,141%
Year	13,013%	13,046%
Weekday	13,388%	13,364%
Holidays	13,131%	13,030%
WindOff_DE	13,198%	13,303%
WindOn_DE	12,881%	12,904%
Solar_DE	13,269%	13,030%
Wind_AT	13,113%	13,129%
Solar_AT	12,974%	12,879%
Wind_HU	12,885%	12,850%
Solar_HU	13,554%	13,249%
Total_Load_Forecast_Slovenia	13,029%	13,265%
Total_Load_Forecast_Serbia	12,716%	
Total_Load_Forecast_Croatia	13,352%	13,289%
Total_Load_Forecast_Germany	12,894%	12,992%
Total_Load_Forecast_Austria	13,243%	12,758%
Total_Load_Forecast_Hungary	12,922%	12,815%
min	12,716%	12,758%

Πίνακας 10: Αποτελέσματα backward stepwise regression αλγορίθμου MLP

Από τα παραπάνω αποτελέσματα φαίνεται ότι η αφαίρεση μόνο 1 μεταβλητής οδήγησε σε μια βελτίωση του τελικού σφάλματος κατά 0,2371 μονάδες που αντιστοιχεί σε ποσοστό 1,83%. Το τελικό μοντέλο του συγκεκριμένου αλγορίθμου παρουσιάζεται σημαντικά καλύτερο τελικό σφάλμα σε σχέση με τους δύο πρώτους αλγορίθμους αφού ο τρίτος αφαιρέθηκε από την διαδικασία της επιλογής.

4.6 Επιλογή τελικού μοντέλου πρόβλεψης

Αφού υλοποιήθηκαν οι τέσσερις αλγόριθμοι, γραμμική παλινδρόμηση (MLR), στην συνέχεια οι αλγόριθμοι δέντρων απόφασης Gradient Boosting Machine και Random Forest, ενώ τέλος το νευρωνικό μέσω του αλγορίθμου MLP βάσει των διαδικασιών

που αναλύθηκαν παραπάνω, στο σημείο αυτό κρίνεται απαραίτητη η τελική επιλογή του μοντέλου πρόβλεψης για το πρόβλημα που μελετάται.

Στον ακόλουθο πίνακα, παρουσιάζονται συνολικά τα αποτελέσματα που προέκυψαν έπειτα από την υλοποιήσεις των αλγορίθμων και των επιμέρους διαδικασιών.

	sMAPE		
	Αρχικό μοντέλο	Τελικό μοντέλο	Κέρδος
MLR	17,518%	15,314%	12,582%
GBM	15,807%	15,151%	4,150%
RF	17,184%	16,494%	4,015%
MLP	12,953%	12,716%	1,833%

Πίνακας 11: Συγκεντρωτικά αποτελέσματα αλγορίθμων

Όπως αναφέρθηκε και φαίνεται και παραπάνω στον πίνακα, ο αλγόριθμος Random Forest παρουσιάζει σημαντικά μεγαλύτερο συνολικό σφάλμα σε σχέση με τους άλλους τρεις αλγορίθμους επομένως αφαιρείται από την διαδικασία της αξιολόγησης. Όσον αφορά τους δύο πρώτους αλγορίθμους τα σφάλματα των τελικών μοντέλων έχουν πολύ κοντινή τιμή επομένως αν και ο τελευταίος αλγόριθμος είχε αντίστοιχα κοντινή τιμή θα εξεταζόταν ένας συνδυασμός των βέλτιστων μοντέλων. Παρόλα αυτά ο αλγόριθμος του νευρωνικού δικτύου έδωσε σημαντικά χαμηλότερο σφάλμα σε σύγκριση με τους άλλους δύο οπότε η εξέταση ενός συνδυασμού μοντέλων θα οδηγήσει σε μεγαλύτερο τελικό σφάλμα.

Συνολικά, από τις παραπάνω υλοποιήσεις βάσει μικρότερου σφάλματος sMAPE επιλέγεται το νευρωνικό δίκτυο το οποίο έχει τελικό σφάλμα ίσο με 12.716%.

5

Επίλογος

Στο παρόν Κεφάλαιο θα ακολουθήσει μια σύνοψη της παρούσας εργασίας και των συμπερασμάτων από τα αποτελέσματα που προέκυψαν. Στην συνέχεια, αναφέρονται κάποιες μελλοντικές επεκτάσεις που θα μπορούσαν να γίνουν στην παρούσα εργασία.

5.1 Σύνοψη και συμπεράσματα

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτελεί η αξιολόγηση μοντέλων πρόβλεψης για την τιμή της ηλεκτρικής ενέργειας στην αγορά της Ουγγαρίας αξιοποιώντας δεδομένα και από διασυννοριακές χώρες ή χώρες με αλληλεξάρτηση.

Στα αρχικό κεφάλαιο της παρούσας εργασίας έγινε μια εκτενής ανάλυση και παρουσίαση των αγορών της ηλεκτρικής ενέργειας. Συγκεκριμένα, παρουσιάστηκε η μορφή που μπορεί να έχει η αγορά της ηλεκτρικής ενέργειας ενώ δόθηκε ιδιαίτερη έμφαση στην αγορά ηλεκτρικής ενέργειας της Ουγγαρίας και την ανάγκη για παραγωγή προβλέψεων σχετικά με την τιμή της ηλεκτρικής ενέργειας. Στην συνέχεια, στο δεύτερο κεφάλαιο έγινε μια εκτενής ανάλυση σχετικά με την σημαντικότητα της πρόβλεψης της τιμής της ηλεκτρικής ενέργειας για κάθε συμμετέχοντα σε αυτή.

Αρχικά, έγινε μια κατηγοριοποίηση των προβλέψεων βάσει του ορίζοντα πρόβλεψης και έπειτα αναφέρθηκαν οι μέθοδοι που χρησιμοποιούνται σε κάθε κατηγορία. Οι όσο το δυνατόν πιο ακριβείς προβλέψεις της τιμής της ηλεκτρικής ενέργειας αποτελούν εργαλείο υψηλής αξίας για τις εταιρείες ηλεκτρικής ενέργειας. Εάν πρόκειται για εταιρείες που δραστηριοποιούνται στην εμπορία της ηλεκτρικής ενέργειας συμμετέχοντας σε αγοραπωλησίες, ενδιαφέρονται για ανάπτυξη στρατηγικών που θα οδηγήσουν σε αύξηση των κερδών. Από την άλλη οι προμηθευτές ηλεκτρικής ενέργειας, προβλέποντας την τιμή της τιμής ηλεκτρικής ενέργειας βάσει της οποίας θα την αγοράσουν στην χονδρεμπορική, ενδιαφέρονται για ανάπτυξη τιμολογιακών στρατηγικών τους τελικούς πελάτες που θα οδηγήσουν από την μία σε αύξηση των κερδών τους και συγχρόνως σε αύξηση του μεριδίου της αγοράς. Επιπλέον, μια ακριβής πρόβλεψη συνεισφέρει στον κατάλληλο προγραμματισμό των παραγωγικών μονάδων για μια εταιρεία παραγωγής ηλεκτρικής ενέργειας που θα οδηγήσει σε μείωση του κόστους παραγωγής και ταυτόχρονη αύξηση των κερδών από την προσφορά της ενέργειας στο δίκτυο. Στην συνέχεια του κεφαλαίου, παρουσιάστηκε η επίδραση των γειτονικών χωρών στην διαμόρφωση της τιμής της ηλεκτρικής ενέργειας για την υπό μελέτη χώρα όπως έχει προκύψει βάσει βιβλιογραφίας. Ενώ στο τελευταίο τμήμα του κεφαλαίου, αναπτύχθηκαν οι παράγοντες που επιδρούν στην διαμόρφωση της τιμής της ηλεκτρικής ενέργειας όπως είναι το προβλεπόμενο φορτίο που θα απαιτηθεί από το δίκτυο, η παραγωγή των ανανεώσιμων πηγών ενέργειας, το κόστος των καυσίμων κ.α.. Στο τρίτο κεφάλαιο, έγινε μια ανάλυση των μεθόδων και των μοντέλων πρόβλεψης που χρησιμοποιούνται. Παρουσιάστηκαν εκτενέστερα οι αλγόριθμοι που εφαρμόστηκαν στην παρούσα εργασία. Στην συνέχεια, αναλύθηκαν οι στατιστικοί δείκτες σφάλματος βάσει των οποίων αξιολογείται η ακρίβεια ενός μοντέλου πρόβλεψης ενώ αναπτύχθηκαν οι μέθοδοι της αλυσιδωτής επαλήθευσης και της σταδιακής επιλογής μεταβλητών εισόδου που εφαρμόστηκαν στην παρούσα εργασία. Τέλος, στο τέταρτο κεφάλαιο αναλύεται η πειραματική διαδικασία της παρούσας εργασίας. Αρχικά, παρουσιάστηκαν και αναλύθηκαν τα δεδομένα που χρησιμοποιήθηκαν στο παρόν πρόβλημα για την παραγωγή των προβλέψεων. Στην συνέχεια για κάθε αλγόριθμο, παρουσιάστηκαν τα αποτελέσματα των αρχικών μοντέλων έπειτα από έρευνα σχετικά με τις υπεραπαραμέτρους τους, όπου αυτό χρειαζόταν, ενώ στην συνέχεια παρουσιάστηκαν τα αποτελέσματα από την μέθοδο της προς τα πίσω επιλογής μεταβλητών εισόδου. Για την αξιολόγηση των μοντέλων και των μεθόδων χρησιμοποιήθηκε η μέθοδος της αλυσιδωτής επαλήθευσης.

Από την παραπάνω ανάλυση προέκυψαν τα ακόλουθα βασικά συμπεράσματα:

- Η χρονοσειρά της τιμής της ηλεκτρικής ενέργειας για την αγορά της Ουγγαρίας, όπως συμβαίνει συνήθως και σε όλες τις αγορές, παρουσιάζει έντονη μεταβλητότητα, πολλαπλή εποχικότητα σε ημερήσια συχνότητα και εβδομαδιαία καθώς και κάποια τυχαιότητα.
- Από την εφαρμογή του αλγορίθμου της πολλαπλής παλινδρόμησης σε συνδυασμό με την σταδιακή παλινδρόμηση (backward stepwise regression) προέκυψε το συμπέρασμα ότι επιτεύχθηκε μία μείωση στο σφάλμα κατά 2,2 μονάδες το οποίο αντιστοιχεί σε ιδιαίτερα σημαντικό ποσοστό ίσο με 12,582%.
- Σε αντίθεση με το παραπάνω, από την εφαρμογή του αλγορίθμου Gradient Boosting Machine σε συνδυασμό με την backward stepwise regression το τελικού σφάλματος μειώθηκε κατά 0,65645 μονάδες που αντιστοιχεί σε ποσοστό 4,153% με την αφαίρεση μόνο δύο μεταβλητών. Βάσει του παραπάνω, προκύπτει το συμπέρασμα ότι ο συγκεκριμένος αλγόριθμος έχει μεγαλύτερη ικανότητα εκμάθησης ώστε να διαχειρίζεται πολλές μεταβλητές εισόδου επιλεκτικά και με μη γραμμικό τρόπο. Επομένως η διαδικασία της σταδιακής επιλογής μεταβλητών δεν οδηγεί σε ιδιαίτερη βελτίωση του τελικού σφάλματος. Επιπλέον, υπολογιστικά πρόκειται για έναν αλγόριθμο λιγότερο αποδοτικό σε σχέση με αυτόν της γραμμικής παλινδρόμησης ενώ το τελικό τους σφάλμα είναι σχεδόν ίδιο.
- Αναφορικά με τον αλγόριθμο Random Forest και την διαδικασία της σταδιακής επιλογής προκύπτει το συμπέρασμα ότι είναι ένα μη αποδοτικός αλγόριθμος υπολογιστικά αφού η εκμάθηση του απαιτεί πολύ χρόνο όσο μεγαλύτερο είναι το data set, ο αριθμός των δέντρων και ο αριθμός των μεταβλητών που επιλέγονται πριν τον διαχωρισμό των κόμβων. Το τελικό σφάλμα της διαδικασίας είναι σημαντικά μεγαλύτερο σε σχέση με τους προηγούμενους δύο αλγορίθμους το οποίο μπορεί να οφείλεται σε

υπερπροσαρμογή των δεδομένων λόγω των τυχαίων δέντρων που επιλύονται. Επιπλέον, η διαδικασία της σταδιακής επιλογής των μεταβλητών εισόδου οδήγησε σε μία μείωση κατά 0.69 μονάδες στο τελικό σφάλμα. Επομένως ούτε στην περίπτωση αυτή η διαδικασία της σταδιακής επιλογής μεταβλητών επέφερε ιδιαίτερη βελτίωση του τελικού σφάλματος.

- Από τον τελευταίο αλγόριθμο που εφαρμόστηκε, ο οποίος είναι το νευρωνικό δίκτυο, προκύπτει αρχικά το συμπέρασμα ότι το σφάλμα του αρχικού μοντέλου είναι μικρότερο για ένα στρώμα, αριθμό νευρώνων κοντινό με τον αριθμό των μεταβλητών εισόδου και συνάρτηση ενεργοποίησης λογαριθμική. Η διαδικασία της σταδιακής παλινδρόμησης δεν αποδείχθηκε ιδιαίτερης σημασίας και για τον αλγόριθμο αυτό. Η αφαίρεση μόνο μίας μεταβλητής οδήγησε σε μια βελτίωση του τελικού σφάλματος κατά 0,2371 μονάδες που αντιστοιχεί σε ποσοστό 1,83%.
- Συνολικά, από όλες τις υλοποιήσεις που πραγματοποιήθηκαν ως βέλτιστο μοντέλο επιλέχθηκε αυτό του νευρωνικού δικτύου (MLP) βάσει μικρότερου σφάλματος sMAPE με τελικό σφάλμα ίσο με 12.716%.
- Ως προς τις ανεξάρτητες μεταβλητές που επιλέχθηκαν για το παρόν πρόβλημα, μέσω της σταδιακής παλινδρόμησης στην περίπτωση του μοντέλου της πολλαπλής παλινδρόμησης προέκυψε ότι σημαντικές ανεξάρτητες μεταβλητές για την παραγωγή των προβλέψεων είναι το φορτίο και η παραγωγή των ανανεώσιμων πηγών ενέργειας για την χώρα που μελετάται, το φορτίο της Κροατίας, η παραγωγή των αιολικών της Αυστρίας και η παραγωγή των φωτοβολταϊκών της Γερμανίας. Επιπλέον, από τις βοηθητικές μεταβλητές στο τελικό μοντέλο, σημαντικές είναι η τιμή της προηγούμενης ημέρας και της προηγούμενης εβδομάδας. Από τις μεταβλητές αυτές αποδεικνύεται ότι υπάρχουν στοιχεία των γειτονικών χωρών που επηρεάζουν την διαμόρφωση της τιμής της χώρας που μελετάται ενώ συγχρόνως είναι μεγάλη η επίδραση από την εξέλιξη και την πορεία της τιμής σε πρόσφατο χρόνο.

- Από την εφαρμογή της διαδικασίας της σταδιακής παλινδρόμησης στους διάφορους αλγόριθμους που υλοποιήθηκαν προκύπτει το συμπέρασμα ότι οι αλγόριθμοι μηχανικής μάθησης έχουν μεγάλη ικανότητα εκμάθησης με αποτέλεσμα το τελικό μοντέλο μηχανικής μάθησης για κάθε αλγόριθμο να περιέχει σχεδόν όλες τις ανεξάρτητες μεταβλητές. Στο σημείο αυτό, επιβεβαιώνεται το γεγονός ότι οι αλγόριθμοι μηχανικής μάθησης έναντι των κλασικών προσεγγίσεων παράγουν πιο ακριβείς και αξιόπιστες προβλέψεις για τις τιμές των αγορών της ηλεκτρικής ενέργειας οι οποίες χαρακτηρίζονται από μη γραμμικότητα.
- Από τις υλοποιήσεις των αλγορίθμων μηχανικής μάθησης προκύπτει το συμπέρασμα ότι πιο σημαντικό στάδιο είναι αυτό της επιλογής των υπερπαραμέτρων του μοντέλου από αυτό της σταδιακής παλινδρόμησης. Μέσω των διάφορων συνδυασμών υπερπαραμέτρων που οδηγούν σε μοντέλα διαφορετικής αρχιτεκτονικής παρατηρήθηκε μεγάλη μεταβολή στο τελικό σφάλμα του μοντέλου. Επομένως η αρχιτεκτονική του μοντέλου είναι αυτή που διαμορφώνει την τελική απόδοση του αλγορίθμου μηχανικής μάθησης.

5.2 Μελλοντικές επεκτάσεις

Στο σημείο αυτό, ολοκληρώνοντας την παρούσα εργασία δίνεται η ευκαιρία να προταθούν κάποιες μελλοντικές υλοποιήσεις και προσεγγίσεις που θα μπορούσαν να βελτιώσουν ακόμα περισσότερο την παραγωγή προβλέψεων για την τιμή της ηλεκτρικής ενέργειας.

Μια μελλοντική επέκταση θα μπορούσε να αφορά την εξέταση νέων μεταβλητών εισόδου βάσει των παραγόντων που επηρεάζουν και διαμορφώνουν την τιμή της ηλεκτρικής ενέργειας όπως αναλύθηκαν στο Κεφάλαιο 2. Τέτοιες μεταβλητές θα μπορούσε να είναι η τιμή των καυσίμων όπως είναι το φυσικό αέριο και τα πετρελαιοειδή, καιρικές συνθήκες κ.α. Επίσης, θα μπορούσαν να εξεταστούν δεδομένα και από τις υπόλοιπες διασυνοριακές χώρες της Ουγγαρίας.

Επιπλέον, μελλοντική επέκταση θα αποτελούσε η εφαρμογή και άλλων μοντέλων και μεθόδων πρόβλεψης χρονοσειρών όπως αυτά που αναπτύχθηκαν στο Κεφάλαιο 2 και στο Κεφάλαιο 3 και να γίνει μια σύγκριση με τα μοντέλα που εξετάστηκαν στην παρούσα εργασία.

Μια άλλη ανάλυση που θα μπορούσε να εξεταστεί αφορά την σύγκριση και αξιολόγηση των μοντέλων που εφαρμόστηκαν διαχωρίζοντας την ακρίβεια των προβλέψεων για εργάσιμες ημέρες και Σαββατοκύριακα. Αντίστοιχα, θα μπορούσε να εξεταστεί η ακρίβεια και η αξιοπιστία των προβλέψεων για τις αργίες και μη αργίες αλλά ακόμα και για τις μέρες που δεν αποτελούν ειδικό γεγονός και αυτές που αποτελούν, όπως την περίοδο που διένυσαν και διανύουν όλες οι χώρες λόγω της πανδημίας.

Η αξία της πρόβλεψης της τιμής της ηλεκτρικής ενέργειας και η επίδραση των γειτονικών χωρών σε αυτή διαφαίνεται από όλη την ανάλυση που έχει προηγηθεί. Επομένως πρόκειται για ένα αντικείμενο που θα απασχολεί συνεχώς τους συμμετέχοντες στην αγορά της ηλεκτρικής ενέργειας αλλά και την επιστημονική κοινότητα με αποτέλεσμα να προκύπτουν συνεχώς επεκτάσεις για την παρούσα διπλωματική εργασία.

6

Βιβλιογραφία

- [1] Πετρόπουλος Φ. & Ασημακόπουλος Β. (2013), *Επιχειρησιακές Προβλέψεις*, Αθήνα: Εκδόσεις:Συμμετρία.
- [2] Saber Talari, Miadreza Shafie-khah, Gerardo J. Osório, Fei Wang, Alireza Heidari and João P. S. Catalão: Price Forecasting of Electricity Markets in the Presence of a High Penetration of Wind Power Generators- Sustainability 2017, 9, 2065, available at <https://www.mdpi.com/2071-1050/9/11/2065/pdf>
- [3] SNNS, Stuttgart Neural Network Simulator, User Manual, Version 4.2, available at: <http://www.ra.cs.unituebingen.de/downloads/SNNS/SNNSv4.2.Manual.pdf>
- [4] Oguzhan Goktug Poyrazoglu, Gokturk Poyrazoglu: Impact of Natural Gas Price on Electricity Price Forecasting in Turkish Day-Ahead Market, 2019 1st Global Power, Energy and Communication Conference (IEEE GPECOM2019), June 12-15, 2019, Cappadocia, Turkey
- [5] Jesus Lago, Fjo De Ridder, Peter Vrancx, Bart De Schutter: Forecasting day-ahead

electricity prices in Europe: The importance of considering market integration. *Applied Energy* 211 (2018) 890–903

[6] Jesus Lago, Fjo De Ridder, Bart De Schutter: Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy* 221 (2018) 386–405.

[7] DG INTERNAL POLICIES OF THE UNION, Policy Department: Economic and Scientific Policy: Price-setting in the Electricity Markets within the EU Single Market. Briefing Note (IP/A/ITRE/SC/2005-174)

[8] Marin Cerjan, Ana Petrićić and Marko Delimar: HIRA Model for Short-Term Electricity Price Forecasting, *Energies* 2019, 12, 568

[9] Guoqiang Zhang, B. Eddy Patuwo, Michael Y. Hu: Forecasting with artificial neural networks: The state of the art, *International Journal of Forecasting* 14 (1998) 35–62

[10] Aleksandra Dedinec, Aleksandar Dedinec: Electricity price forecasting of the South East European power exchanges, 7th International Conference on Information Society and Technology ICIST 2017

[11] Max Kuhn: Building Predictive Models in R Using the caret Package, *Journal of Statistical Software*, November 2008, Volume 28, Issue 5.

[12] Ashkan Yousefi, Omid Ameri Sianaki, Dean Sharafi: Long-Term Electricity Price Forecast Using Machine Learning Techniques, 2019 IEEE PES Innovative Smart Grid Technologies Asia

[13] R Documentation and manuals | R Documentation, <https://www.rdocumentation.org/>

[14] To pool or not to pool?, article available at:

<https://www.powerengineeringint.com/world-regions/asia/to-pool-or-not-to-pool/>

[15] What Factors Influence Business Energy Prices?, article available at:

<https://www.utilityhelpline.co.uk/news/uk-factors-influencing-business-energy-prices>

[16] Decision Tree vs Random Forest vs Gradient Boosting Machines: Explained Simply, article available at:

<https://www.datasciencecentral.com/profiles/blogs/decision-tree-vs-random-forest-vs-boosted-trees-explained>

[17] Stepwise regression – Wikipedia,

available at: https://en.wikipedia.org/wiki/Stepwise_regression

[18] Multilayer perceptron – Wikipedia,

Available at: https://en.wikipedia.org/wiki/Multilayer_perceptron

[19] Decision tree learning – Wikipedia,

Available at: https://en.wikipedia.org/wiki/Decision_tree_learning

[20] Introduction to the Hungarian energy market, available at:

http://www.investhipa.hu/images/hipa_kiadvany_intro_greenenergy_web_201808.pdf

[21] <http://mavir.hu/web/mavir-en>

[22] <https://hupx.hu/en/>

[23] General overview for investors in Hungary's energy market, available at:

<https://hipa.hu/images/HIP/Green%20industry%20overview.pdf>

[24] Viktor Miklos Kiss, Zsolt Hetesi, Tibor Kiss : Issues and solutions relating to Hungary's electricity system, Energy, Volume 116, Part 1, 1 December 2016, Pages 329-340