



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

**ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

**Μοντελοποίηση ηλεκτρικών φορτίων αργιών σε εφαρμογές  
πρόβλεψης ηλεκτρικού φορτίου**

Μεταπτυχιακή Διπλωματική Εργασία

Παππάς Κωνσταντίνος

Επιβλέπουσα: Δασκαλοπούλου Ασπασία

Σεπτέμβριος 2022



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ**

**ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ**

**ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ**

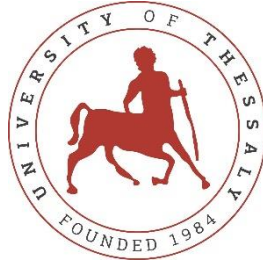
**Μοντελοποίηση ηλεκτρικών φορτίων αργιών σε εφαρμογές  
πρόβλεψης ηλεκτρικού φορτίου**

**Μεταπτυχιακή Διπλωματική Εργασία**

**Παππάς Κωνσταντίνος**

**Επιβλέπουσα: Δασκαλοπούλου Ασπασία**

**Σεπτέμβριος 2022**



**UNIVERSITY OF THESSALY**

**SCHOOL OF ENGINEERING**

**DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING**

**Modeling of electric load of holidays in electrical load forecasting  
applications**

MSc Thesis

Pappas Konstantinos

Supervisor: Daskalopulu Aspassia

September 2022

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπων

**Δασκαλοπούλου Ασπασία**

Αναπληρώτρια Καθηγήτρια, Τμήμα Ηλεκτρολόγων Μηχανικών και  
Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος

**Μπαργιώτας Δημήτριος**

Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών  
Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος

**Ευμορφόπουλος Νέστωρ**

Αναπληρωτής Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και  
Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

**ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ  
ΔΙΚΑΙΩΜΑΤΩΝ**

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα μεταπτυχιακή διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας, αποτελούν αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλουν οποιασδήποτε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχουν έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Δηλώνω επίσης ότι τα αποτελέσματα της εργασίας δεν έχουν χρησιμοποιηθεί για την απόκτηση άλλου πτυχίου. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο/Η Δηλών/ούσα

Κωνσταντίνος Παππάς

## **DISCLAIMER ON ACADEMIC ETHICS AND INTELLECTUAL PROPERTY RIGHTS**

Being fully aware of the implications of copyright laws, I expressly state that this MSc thesis, as well as the electronic files and source codes developed or modified in the course of this thesis, are solely the product of my personal work and do not infringe any rights of intellectual property, personality and personal data of third parties, do not contain work / contributions of third parties for which the permission of the authors / beneficiaries is required and are not a product of partial or complete plagiarism, while the sources used are limited to the bibliographic references only and meet the rules of scientific citing. The points where I have used ideas, text, files and / or sources of other authors are clearly mentioned in the text with the appropriate citation and the relevant complete reference is included in the bibliographic references section. I also declare that the results of the work have not been used to obtain another degree. I fully, individually and personally undertake all legal and administrative consequences that may arise in the event that it is proven, in the course of time, that this thesis or part of it does not belong to me because it is a product of plagiarism.

The Declarant

Konstantinos Pappas

## Ευχαριστίες

Η ολοκλήρωση της διπλωματικής εργασίας αποτελεί και το τελευταίο βήμα για την απόκτηση του μεταπτυχιακού διπλώματος στον τομέα των Ευφύων Δικτύων Ηλεκτρικής Ενέργειας. Αισθάνομαι την ανάγκη να ευχαριστήσω τους ανθρώπους που με βοήθησαν να φτάσω μέχρι εδώ.

Αρχικά, οφείλω να ευχαριστήσω την κυρία Δασκαλοπούλου, για την βοήθεια και τις συμβουλές που πρόσφερε κατά την διάρκεια εκπόνησης της παρούσας εργασίας.

Τον εκλιπόντα Καθηγητή Ιωάννη Παναπακίδη, αρχικό επιβλέποντα και εισηγητή της διπλωματικής μου, για την αμέριστη βοήθεια και την εμπιστοσύνη που μου έδειξε κατά την ανάθεση του θέματος και την υποστήριξη του.

Δεν μπορώ να παραλείψω τις ευχαριστίες μου προς την επιτροπή αποτελούμενη από τον κύριο Μπαργιώτα Δημήτριο και τον κύριο Ευμορφόπουλο Νέστωρα τόσο για την συμμετοχή τους στην τριμελή επιτροπή όσο και για τον χρόνο που διέθεσαν για την αξιολόγηση της διπλωματικής.

Στην οικογένειά μου και στους φίλους μου για την υποστήριξη τους κατά την διάρκεια των σπουδών μου.

Στην Μαρία και την Ιουλία, για τις συμβουλές τους κατά τη συγγραφή του κειμένου.

Στον Κωνσταντίνο Δασκαλόπουλο για τις συμβουλές του στο κομμάτι της συγγραφής του κώδικα που χρησιμοποιήθηκε.

Τέλος, οφείλω να ευχαριστήσω το πανεπιστήμιο Θεσσαλίας, για την δυνατότητα που μου έδωσε να συμμετάσχω στο μεταπτυχιακό πρόγραμμα και για την γνώση που κέρδισα από τη συμμετοχή μου αυτό.

## Μοντελοποίηση ηλεκτρικών φορτίων αργιών σε εφαρμογές πρόβλεψης ηλεκτρικού φορτίου

Κωνσταντίνος Παππάς

### Περίληψη

Η πρόβλεψη φορτίου της επόμενης ημέρας αποτελεί σημαντικό παράγοντα για την εύρυθμη λειτουργία του συστήματος ηλεκτρικής ενέργειας. Έχει δημιουργηθεί ένας μεγάλος αριθμός μοντέλων για την ακριβή πρόβλεψη της ζήτησης με τον βαθμό σφάλματος να βρίσκεται σε επίπεδα της τάξης του 1-2%. Το σφάλμα αυτό είναι αρκετά μεγαλύτερο κατά τις ειδικές ημέρες, ή ημέρες αργιών όπου η ζήτηση διαφοροποιείται λόγω της αλλαγής στη συμπεριφορά των ανθρώπων. Για την μείωση του σφάλματος και κατά επέκταση την αύξηση της ακρίβειας των μοντέλων έχει προταθεί ένας αριθμός λύσεων εφαρμόζοντας διάφορους αλγορίθμους και μοντέλα πρόβλεψης. Στην παρούσα εργασία πραγματοποιείται μια βιβλιογραφική ανασκόπηση των διαφορετικών μεθόδων, αναλύεται το πρόβλημα των ειδικών ημερών προσαρμοσμένο στα ελληνικά δεδομένα και τέλος αναπτύσσεται ένα μοντέλο για την πρόβλεψη του φορτίου της επόμενης ημέρας.

### Λέξεις-κλειδιά:

Πρόβλεψη φορτίου αργιών, ειδικές ημέρες, support vector machines, multi-layer perceptron, XGBoost, βελτιστοποίηση χαρακτηριστικών, αξιολόγηση χαρακτηριστικών.



**MODELING OF ELECTRIC LOAD OF HOLIDAYS IN ELECTRICAL LOAD  
FORECASTING APPLICATIONS**

Konstantinos Pappas

**Abstract**

Next day load forecasting is an important factor for the smooth operation of the power system. Many models have been developed for accurate demand forecasting with the error rate being in the range of 1-2%. This error is considerably higher on special days, or holiday days when demand varies due to changes in people's behavior. To reduce the error and thus increase the accuracy of the models, several solutions have been proposed by applying various algorithms and forecasting models. In this paper a literature review of the different methods is carried out, the problem of special days adapted to the Greek data is analyzed and finally a model for the prediction of the next day's load is developed.

**Keywords:** Load prediction on special days, support vector machines, special days, multi-layer perceptron, XGBoost, Boruta-SHAP feature selection, Bayesian optimization.

Στον Ιωάννη Παναπακίδη.  
*Nanos gigantum humeris insidentes.*

# Πίνακας Περιεχομένων

<i>Ευχαριστίες</i> .....	<i>vii</i>
<i>Περίληψη</i> .....	<i>viii</i>
<i>Abstract</i> .....	<i>ix</i>
<i>Πίνακας Περιεχομένων</i> .....	<i>xi</i>
<i>Λίστα σχημάτων</i> .....	<i>xiii</i>
<i>Λίστα πινάκων</i> .....	<i>xiv</i>
<b>Κεφάλαιο 1 Εισαγωγή</b> .....	<b>1</b>
<b>1.1 Αντικείμενο της διπλωματικής</b> .....	<b>1</b>
1.1.1 Συνεισφορά .....	2
<b>1.2 Οργάνωση του τόμου</b> .....	<b>2</b>
<b>Κεφάλαιο 2 Ευφυή Δίκτυα Ηλεκτρικής Ενέργειας</b> .....	<b>4</b>
<b>2.1 Εισαγωγή</b> .....	<b>4</b>
<b>2.2 Ευφυή δίκτυα</b> .....	<b>5</b>
<b>Κεφάλαιο 3 Πρόβλεψη φορτίου</b> .....	<b>11</b>
<b>3.1 Εισαγωγή</b> .....	<b>11</b>
<b>3.2 Κατηγοριοποίηση βάση χρονικού εύρους</b> .....	<b>11</b>
<b>3.3 Παράγοντες που επηρεάζουν το φορτίο συστήματος</b> .....	<b>12</b>
<b>3.4 Μοντέλα πρόβλεψης φορτίου ηλεκτρικής ενέργειας</b> .....	<b>12</b>
<b>3.5 Μοντέλα Χρονοσειρών</b> .....	<b>15</b>
<b>3.6 Παλινδρόμηση</b> .....	<b>15</b>
<b>3.7 Support Vector machines</b> .....	<b>16</b>
<b>3.8 Μοντέλα Bottom up</b> .....	<b>16</b>
<b>Κεφάλαιο 4 Ανάλυση φορτίου και καθορισμός του προβλήματος των ειδικών ημερών</b> .....	<b>17</b>
<b>4.1 Φορτίο στις κανονικές ημέρες</b> .....	<b>17</b>
<b>4.2 Φορτίο στις ειδικές ημέρες</b> .....	<b>18</b>

4.3 Ειδικές ημέρες στην Ελλάδα.....	19
4.4 Ανάλυση φορτίου ειδικών ημερών. ....	21
4.5 Το πρόβλημα των ειδικών ημερών.....	27
<b>Κεφάλαιο 5 Δεδομένα. ....</b>	<b>33</b>
5.1 Χρονοσειρές.....	33
5.2 Χαρακτηριστικά μεγέθη των χρονοσειρών. ....	33
5.3 Πηγές δεδομένων.....	35
<b>Κεφάλαιο 6 Τεχνητά νευρωνικά δίκτυα, επιλογή χαρακτηριστικών, βελτιστοποίηση.....</b>	<b>37</b>
6.1 Τεχνητά νευρωνικά δίκτυα.....	37
6.2 Perceptron.....	38
6.3 Multilayer Perceptron.....	39
6.4 Extreme gradient boosting (XGBoost).....	39
6.5 Support Vector Machines.....	41
6.6 Boruta-SHAP feature selection.....	42
6.7 Bayesian Optimization.....	43
<b>Κεφάλαιο 7 Υλοποίηση.....</b>	<b>45</b>
7.1 Πρώτη Προσέγγιση.....	45
7.2 Συγκεντρωτική ανάλυση αποτελεσμάτων πρώτης προσέγγισης.....	56
7.3 Ανάλυση χαρακτηριστικών με την χρήση του Boruta-SHAP.....	58
7.4 Προτεινόμενη μεθοδολογία.....	61
7.4.1 Μέρος πρώτο. Εύρεση και επιλογή όμοιων ημερών.....	61
7.4.2 Μέρος Δεύτερο. Βελτιστοποίηση.....	64
7.5 Σχολιασμός της δεύτερης προσέγγισης, και πρόταση μεθοδολογίας.....	68
<b>Κεφάλαιο 8 Συμπεράσματα και προτάσεις για περαιτέρω έρευνα.....</b>	<b>70</b>
8.1 Συμπεράσματα.....	70
8.2 Περαιτέρω έρευνα.....	72
<b>Βιβλιογραφία.....</b>	<b>74</b>

## Λίστα σχημάτων

EU Smart Grid Framework, CEN-CENELEC-ETS 2012.....	7
The European Union's Smart Grid vision -grid operation, Becks et all 2010 .....	10
Classified forecasting models distribution, Kuster et all, 2017 .....	14
Είσοδοι ανάλογα με το μοντέλο πρόβλεψης, Kuster et all, 2017.....	14
Φορτίο Αυγούστου 2019.....	17
Σύγκριση Φορτίου Πρωτοχρονιάς.....	21
Σύγκριση Φορτίου Θεοφανίων .....	22
Σύγκριση Φορτίου Καθαράς Δευτέρας .....	23
Σύγκριση Φορτίου Επετείου της Επανάστασης.....	23
Σύγκριση Φορτίου Πάσχα - Εργατικής Πρωτομαγιάς.....	24
Σύγκριση Φορτίου Αγίου Πνεύματος.....	25
Σύγκριση Φορτίου Κοιμήσεως της Θεοτόκου .....	25
Σύγκριση Φορτίου Επετείου του Όχι .....	26
Σύγκριση Φορτίου Χριστουγέννων και δεύτερης ημέρας Χριστουγέννων .....	26
Φορτίο από 1/1 έως 6/1 2012.....	34
Αντιστοιχία τεχνητών και φυσικών νευρώνων, Wikimedia 2022 .....	37
Perceptron (Sharing Data and Models in Software Engineering, 2015) .....	38
Multilayer Perceptron ANN Hassan et all, 2015.....	39
Regression, tube and functions .....	42
Αποτελέσματα Boruta-SHAP .....	59
Οπτικοποίηση αποτελεσμάτων.....	59
Διάγραμμα ροής μεθοδολογίας.....	69

## Λίστα πινάκων

Πίνακας 2.1 Τομείς και ρόλοι, Gopstein et al, 2021 .....	8
Πίνακας 4.1 Υποχρεωτικές Αργίες.....	19
Πίνακας 4.2 Αργίες κατ' έθιμο.....	19
Πίνακας 4.3 Κατηγοριοποίηση βάση ημέρας .....	20
Πίνακας 7.1 Ειδικές Ημέρες.....	49
Πίνακας 7.2 Αποτελέσματα πρώτης υποπερίπτωσης.....	50
Πίνακας 7.3 Αποτελέσματα δεύτερης υποπερίπτωσης.....	52
Πίνακας 7.4 Αποτελέσματα τρίτης υποπερίπτωσης.....	53
Πίνακας 7.5 Αποτελέσματα τέταρτης υποπερίπτωσης .....	55
Πίνακας 7.6 Συγκεντρωτικά αποτελέσματα Standard Scaler .....	57
Πίνακας 7.7 Αποτελέσματα Boruta-SHAP.....	60
Πίνακας 7.8 Ημέρες που επιλέχθηκαν ως οι πιο όμοιες .....	62
Πίνακας 7.9 Αποτελέσματα ομοίων ημερών .....	63
Πίνακας 7.10 Παράμετροι αναζήτησης MLP.....	64
Πίνακας 7.11 Αποτελέσματα MLP μετά τη βελτιστοποίηση .....	65
Πίνακας 7.12 Παράμετροι βελτιστοποίησης XGBoost.....	65
Πίνακας 7.13 Αποτελέσματα μετά τη βελτιστοποίηση XGBoost.....	66
Πίνακας 7.14 Παράμετροι αναζήτησης SVM .....	67
Πίνακας 7.15 Αποτελέσματα μετά τη βελτιστοποίηση SVM .....	67

## **Κεφάλαιο 1 Εισαγωγή**

Η ηλεκτρική ενέργεια και η εκτεταμένη χρήση της τον 20 αιώνα αποτέλεσε βασικό πυλώνα για την βελτίωση της ζωής εκατομμυρίων ανθρώπων, την ανάπτυξη της βιομηχανίας και την διαμόρφωση της ζωής μας έτσι όπως την γνωρίζουμε. Η αδιάκοπη και ποιοτική παροχή ηλεκτρισμού είναι αναγκαία τόσο σε βιομηχανικό/εμπορικό όσο και σε οικιακό επίπεδο. Το σύστημα που αναλαμβάνει το έργο αυτό ονομάζεται Σύστημα Ηλεκτρικής Ενέργειας (ΣΗΕ). Το ΣΗΕ είναι υπεύθυνο για την παραγωγή, την μεταφορά, και την διανομή της ηλεκτρικής ενέργειας όπου αυτή είναι αναγκαία.

Για τη σωστή λειτουργία του συστήματος απαραίτητη καθίσταται η εκ των προτέρων εκτίμηση της ζήτησης της ηλεκτρικής ενέργειας, καθώς δεν είναι δυνατή η αποθήκευσή της με οικονομικά βιώσιμο τρόπο. Επιπλέον, η διατήρηση της ισορροπίας μεταξύ παραγωγής και κατανάλωσης είναι απαραίτητη για την διατήρηση των ποιοτικών χαρακτηριστικών της παρεχόμενης ενέργειας.

Ακόμη, η απελευθέρωση της αγοράς ενέργειας δημιούργησε ένα ακόμη χώρο όπου η εκτίμηση της ζήτησης αποτελεί καθοριστικό παράγοντα για τους συμμετέχοντες στην αγορά. Αυτό οφείλεται στην αναγκαιότητα της εκτίμησης της ζήτησης για τον καθορισμό των τιμών στο χρηματιστήριο ενέργειας.

Η σημασία λοιπόν στην εκτίμηση της ζήτησης της επόμενης ημέρας είναι μεγάλη τόσο για την ικανοποίηση των λειτουργικών αναγκών του ΣΗΕ όσο και για τη λήψη ζωτικών αποφάσεων των συμμετεχόντων στην αγορά ηλεκτρικής ενέργειας.

### **1.1 Αντικείμενο της διπλωματικής**

Αναλύοντας τα φορτία των ημερών παρατηρείται διαφοροποίηση αναλόγως της χρονικής περιόδου (ημέρα, εποχή κλπ.). Ειδικότερα, οι ημέρες αργιών εμφανίζουν διαφορετική συμπεριφορά εν συγκρίσει με τις υπόλοιπες καθημερινές, πράγμα που είναι υπεύθυνο για την δημιουργία αποκλίσεων στα μοντέλα πρόβλεψης. Γνωρίζοντας

την σημασία της πρόβλεψης με το μικρότερο δυνατό σφάλμα, πραγματοποιείται μια εκτενής βιβλιογραφική ανασκόπηση των μεθόδων αντιμετώπισης του προβλήματος της πρόβλεψης φορτίου των ειδικών ημερών, στην συνέχεια ο ορισμός του προβλήματος βάση των τοπικών δεδομένων της Ελλάδας και τέλος η πρόταση μιας νέας μεθοδολογίας για την πρόβλεψη φορτίου.

### 1.1.1 Συνεισφορά

Δεδομένου του προβλήματος του φορτίου των ειδικών ημερών αφού μελετήθηκε η βιβλιογραφία έγινε μια προσπάθεια για την ανάπτυξη μιας μεθόδου με στόχο την πρόβλεψη του φορτίου της επόμενης ημέρας όσο αφορά τις ειδικές ημέρες. Συνοπτικά:

1. Υλοποιήθηκαν τρία μοντέλα πρόβλεψης με την χρήση Multi-Layer Perceptron Support Vector Machine και XGBoost.
2. Αξιολογήθηκαν οι επιδόσεις τους σε τέσσερα σενάρια με διαφορετικές εισόδους.
3. Αξιολογήθηκαν οι εισοδοί ως προς την σχετικότητά τους με την χρήση του Boruta-SHAP.
4. Υλοποιήθηκε μια μέθοδος πρόβλεψης βάση των εμπειριών που αποκομίστηκαν από την πρώτη προσπάθεια.
5. Εφαρμόστηκε η βελτίωση στις παραμέτρους των τριών αλγορίθμων πρόβλεψης με τη χρήση Bayesian optimization.
6. Αξιολογήθηκε το τελικό μοντέλο.

## 1.2 Οργάνωση του τόμου

Στο 2<sup>ο</sup> κεφάλαιο γίνεται μια εισαγωγή στα ευφυή δίκτυα ηλεκτρικής ενέργειας, τα οποία αποτελούν την εξέλιξη του παρόντος συστήματος μεταφοράς και διανομής, και αναλύονται οι τεχνολογικές καινοτομίες που αναμένονται με την υλοποίησή τους. Στη συνέχεια γίνεται αναφορά στα χαρακτηριστικά και την δομή τους και τέλος δίνεται ένας ορισμός τους.



Το 3<sup>ο</sup> κεφάλαιο αφορά την πρόβλεψη φορτίου. Αρχικά διαχωρίζονται οι τύποι της πρόβλεψης ανάλογα με το χρονικό ορίζοντα και αναλύονται οι διαφορές τους. Στην συνέχεια αναπτύσσονται οι παράγοντες που επηρεάζουν την κατανάλωση ηλεκτρικής ενέργειας, με στόχο την καλύτερη δημιουργία των μοντέλων πρόβλεψης και αναφέρονται οι διάφορες μεθοδολογίες και πραγματοποιείται συγκριτική ανάλυση της χρήσης τους.

Το 4<sup>ο</sup> κεφάλαιο αφορά τις ειδικές ημέρες. Κατηγοριοποιούμε τις αργίες στην Ελλάδα και αναφέρουμε την νομοθεσία πάνω στην οποία στηρίζονται. Στη συνέχεια, ορίζουμε το πρόβλημα των ειδικών ημερών λόγω της διαφοράς στα μοτίβα κατανάλωσης σε σύγκριση με τις κανονικές ημέρες και παρατίθενται τα αποτελέσματα της βιβλιογραφικής έρευνας πάνω στην αντιμετώπιση του προβλήματος.

Το 5<sup>ο</sup> κεφάλαιο αφορά τα δεδομένα. Αναφέρονται τα βασικά χαρακτηριστικά των χρονοσειρών, και στην συνέχεια παρουσιάζονται οι πηγές των δεδομένων τα οποία χρησιμοποιήθηκαν στην παρούσα εργασία. Τέλος, συζητάμε για το πρόβλημα που αντιμετωπίσαμε κατά την επιλογή των δεδομένων.

Στο 6<sup>ο</sup> κεφάλαιο συζητάμε για τα νευρωνικά δίκτυα, το μοντέλο Support Vector Machine και τα μοντέλα δένδρων, τα οποία χρησιμοποιήθηκαν για την εκτίμηση του φορτίου της επόμενης ημέρας. Έπειτα, αναφερόμαστε στην λειτουργία του αλγορίθμου επιλογής χαρακτηριστικών Boruta-SHAP, και τέλος παρουσιάζεται η λειτουργία του αλγορίθμου βελτιστοποίησης Bayesian optimization που χρησιμοποιήθηκε για την εύρεση των παραμέτρων των δύο μοντέλων.

Στο 7<sup>ο</sup> κεφάλαιο γίνεται υλοποίηση ενός αρχικού μοντέλου πρόβλεψης και με τους δύο αλγορίθμους σε τέσσερα διαφορετικά σενάρια εισόδων, παρουσιάζονται τα αποτελέσματα και στη συνέχεια προτείνεται μια μεθοδολογία για την πρόβλεψη φορτίου της επόμενης ημέρας, παρουσιάζονται τα αποτελέσματά της και συγκρίνονται με το καλύτερο μοντέλο της αρχικής υλοποίησης.

Το 8<sup>ο</sup> κεφάλαιο αποτελεί τα συμπεράσματα και τις προτάσεις για μελλοντική έρευνα.

## Κεφάλαιο 2 Ευφυή Δίκτυα Ηλεκτρικής Ενέργειας.

### 2.1 Εισαγωγή

Η χρήση των υδρογονανθράκων αποτέλεσε το έναυσμα της βιομηχανικής επανάστασης με την χρήση του γαιάνθρακα για την παραγωγή έργου στις πρώτες ατμομηχανές. Το κυριότερο ίσως πλεονέκτημά τους δεν είναι άλλο από την μεγάλη ενεργειακή τους πυκνότητα που κυμαίνεται στα επίπεδα των 26 – 33MJ/kg [Fischer and Elert, 2003] για το γαιάνθρακα και τα 46,4 MJ/Kg [IOR Energy, 2003] για την βενζίνη. Η επέκταση της χρήσης τους στους τομείς της παραγωγής ηλεκτρικής ενέργειας, θέρμανσης και μεταφοράς είχε ως αποτέλεσμα την εκθετική αύξηση της κατανάλωσης τους [Fossil Fuels, EESI, 2022].

Το μεγαλύτερο μειονέκτημα της χρήσης τους είναι η εκπομπή ενός ευρέως φάσματος αερίων παραγόμενα κατά την καύση τους τα οποία συνδέονται με την αύξηση της θνησιμότητας και θεωρούνται υπεύθυνα για 1 στους 5 θανάτους παγκοσμίως [Vohra et al., 2021].

Επιπλέον, μέρος των αερίων που εκπέμπονται συντελούν στην αύξηση της συγκέντρωσης του διοξειδίου του άνθρακα στην ατμόσφαιρα με αποτέλεσμα την άνοδο της θερμοκρασίας σε παγκόσμιο επίπεδο [Climate Change, 2020] και κατ' επέκταση στην κλιματική αλλαγή. Η αντίδραση της διεθνούς κοινότητας συνίσταται στην υπογραφή ενός αριθμού πρωτοκόλλων με στόχο τον περιορισμό της χρήσης τους [1. Bell and Cullen, 2015, 2. UNFCCC, 1997].

Η εύρεση εναλλακτικών μορφών ενέργειας είχε ως αποτέλεσμα την ανάπτυξη των ανανεώσιμων μορφών ενέργειας (ΑΠΕ) οι οποίες μετατρέπουν την ηλιακή, αιολική, κυματική κ.α. ενέργεια σε ηλεκτρική. Οι ΑΠΕ προωθήθηκαν έντονα τόσο από την Ε.Ε. [Directive 2009/28/EC, EU, Rev. 2021] όσο και από τις ΗΠΑ [US Energy Administration, 2022] με αποτέλεσμα την αύξηση της συμμετοχής τους στο ενεργειακό μείγμα των χωρών [Share of energy consumption from renewable sources in Europe, 2022].

Η αναγκαιότητα για την εξέλιξη του δικτύου μεταφοράς και διανομής ηλεκτρικής ενέργειας προέκυψε από την είσοδο των ΑΠΕ και την προσπάθεια για μείωση των εκπομπών αερίων του θερμοκηπίου με στόχο την δημιουργία μιας υποδομής ικανής να ανταπεξέλθει στις ανάγκες που προκύπτουν λόγω των παραπάνω, με την ενσωμάτωση νέων τεχνολογιών στην υπάρχουσα υποδομή.

## 2.2 Ευφυή δίκτυα

Τα δίκτυα ηλεκτρικής ενέργειας βελτιώνονται συνεχώς με στόχο την παροχή σταθερής και ποιοτικής ενέργειας (τάση, συχνότητα) στον τελικό χρήστη. Η είσοδος των ΑΠΕ και οι νέες ανάγκες που προέκυψαν από τις ενεργειακές πολιτικές της ΕΕ και των ΗΠΑ με στόχο την μείωση των περιβαλλοντολογικών επιπτώσεων της μετατροπής της χημικής ενέργειας των ορυκτών καυσίμων σε ηλεκτρική στους σταθμούς παραγωγής οδήγησε στην εξέλιξη του παρόντος δικτύου ηλεκτρικής ενέργειας.

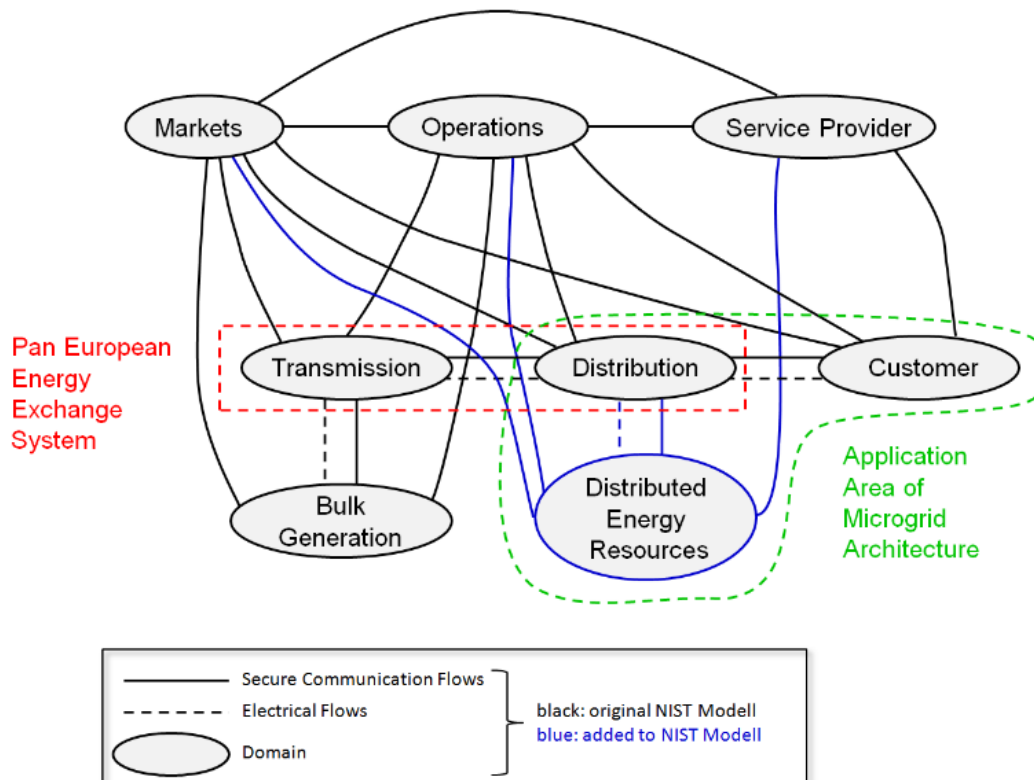
Προοίμιο της αλλαγής αυτής και ταυτόχρονα πρώτος επίσημος ορισμός του ευφυούς δικτύου αποτελεί η νομοθετική ρύθμιση EISA-2007 [ENERGY INDEPENDENCE AND SECURITY ACT OF 2007] όσο αφορά τις ΗΠΑ. Η Ε.Ε με τη σειρά της από το 2005 δημιούργησε την Smart Grid Technology Platform [Iqtiyanillham, Hasanuzzaman and Hosenuzzaman, 2017] με στόχο την δημιουργία ενός πλαισίου για το ευρωπαϊκό ευφυές δίκτυο. Επιπλέον όσο αφορά την Ε.Ε., η περιβαλλοντική πολιτική της με στόχο να είναι carbon neutral με χρονικό όριο το 2050 [The European Green Deal, 2019] έχει θετική επίδραση στην προώθηση των ευφυών δικτύων.

Τόσο η ΕΕ όσο και οι ΗΠΑ έχουν αναπτύξει τον δικό τους ορισμό για τα ευφυή δίκτυα. [1. Office of Electricity, 2008, 2. European Smart Grids Technology Platform, 2006]. Εξετάζοντας τους δύο ορισμούς συμπεραίνουμε πως ταυτίζονται στις βασικές τους αρχές, με την ΕΕ να προωθεί περαιτέρω την διεσπαρμένη παραγωγή. Συνοπτικά η μορφή που προτείνει η Ε.Ε. καθώς και οι διαφορές της από το μοντέλο των ΗΠΑ παρουσιάζεται στην Εικόνα 2.1. Η ανάπτυξη των ευφυών δικτύων αναμένεται να βελτιώσει τα χαρακτηριστικά του δικτύου. Οι τομείς στους οποίους προσβλέπουμε βελτιώσεις παρουσιάζονται ακολούθως.

Η βελτίωση της αξιοπιστίας αποτελεί ένα από τα χαρακτηριστικά του ευφυούς δικτύου με τη χρήση τεχνολογιών που επιτρέπουν την επίγνωση της κατάστασης του δικτύου σε κάθε στιγμή, την αυξημένη ικανότητα εντοπισμού σφαλμάτων καθώς και τη δυνατότητα για αυτόματη επιδιόρθωση. Η ενσωμάτωση των τεχνολογιών αυτών οδηγεί στην αύξηση της αντοχής τους σε επιθέσεις, φυσικές καταστροφές και μείωση των ευπαθειών του [Ardito, Procaccianti, Menga and Morisio, 2013].

Η ενσωμάτωση της διεσπαρμένης παραγωγής κάνει το δίκτυο πιο ευέλικτο. Η ροή ενέργειας από τον καταναλωτή προς το δίκτυο καθίσταται δυνατή (bidirectional energy flow), πέρα από την εισαγωγή μικρών παραγωγικών μονάδων σε διάφορα μέρη του δικτύου. Η μετακίνηση της παραγωγής ενέργειας από το παρόν μοντέλο που θέλει μεγάλους σταθμούς παραγωγής σε μικρούς διεσπαρμένους σταθμούς θα μπορούσε να οδηγήσει στην αύξηση της αξιοπιστίας. Σε αντίθεση με το κλασσικό δίκτυο στο οποίο η αμφίδρομη ροή ενέργειας θα προκαλέσει προβλήματα, το ευφυές δίκτυο θα είναι σε θέση να διαχειριστεί αυτού του είδους την εξέλιξη [Tomoiagă et al., 2013].

Η αύξηση της αποδοτικότητας του δικτύου είναι κάτι που προσδοκάτε με την υιοθέτηση των ευφύων δικτύων. Δυνατότητες όπως η διαχείριση της ζήτησης με την χρήση Voltage/VAR optimization [Dejun Tu, 2014] καθώς και η εφαρμογή load balancing [Jiang et al., 2014] αναμένεται να βελτιώσουν κατά πολύ απόδοση του δικτύου οδηγώντας σε μείωση του κόστους και βελτίωση της ποιότητας της παρεχόμενης ενέργειας. Ένα παράδειγμα load management είναι ο περιορισμός χρήσης κατά την περίοδο της αυξημένης ζήτησης, και το αντίστροφο κάτι που καθίσταται δυνατό με την εφαρμογή μεταβλητής τιμολόγησης στην τιμή της ενέργειας.



Εικόνα 2.1 EU Smart Grid Framework, CEN-CENELEC-ETS 2012

Για την πραγματοποίηση των προαναφερθέντων βελτιώσεων κρίνεται αναγκαία η ενσωμάτωση νέων τεχνολογιών στον δίκτυο, όπως οι έξυπνοι μετρητές οι οποίοι αποτελούν μετεξέλιξη των παραδοσιακών μετρητών ενέργειας με την προσθήκη της δυνατότητας της αμφίδρομης επικοινωνίας. Η δυνατότητα της παροχής πληροφορίας στον διαχειριστή αναμένεται να επιτρέψει σε σχεδόν πραγματικό χρόνο την πληροφόρηση σχετικά με το φορτίο του δικτύου, διευκολύνοντας την διαχείρισή του.

Η συγκέντρωση της και διαχείριση της μεταδιδόμενης πληροφορίας απαιτεί ένα κεντρικό σύστημα ελέγχου το οποίο είναι το Supervisory Control and Data Acquisition (SCADA). Το SCADA είναι μια αρχιτεκτονική που συνδυάζει ηλεκτρονικούς υπολογιστές, αισθητήρες, ενεργοποιητές (actuators) και ένα γραφικό περιβάλλον για τον έλεγχο βιομηχανικών εγκαταστάσεων [Stouffer et al., 2015] και χρησιμοποιείται ήδη από τους διαχειριστές για την εποπτεία της λειτουργίας του δικτύου.

Επιπλέον, ένας αριθμός αισθητήρων είναι αναγκαίος για την παραγωγή των απαραίτητων δεδομένων όπως οι Phasor measurement Units(PMU) [Phasor measurement unit - Wikipedia, 2022] καθώς και αισθητήρες ροής ενέργειας.

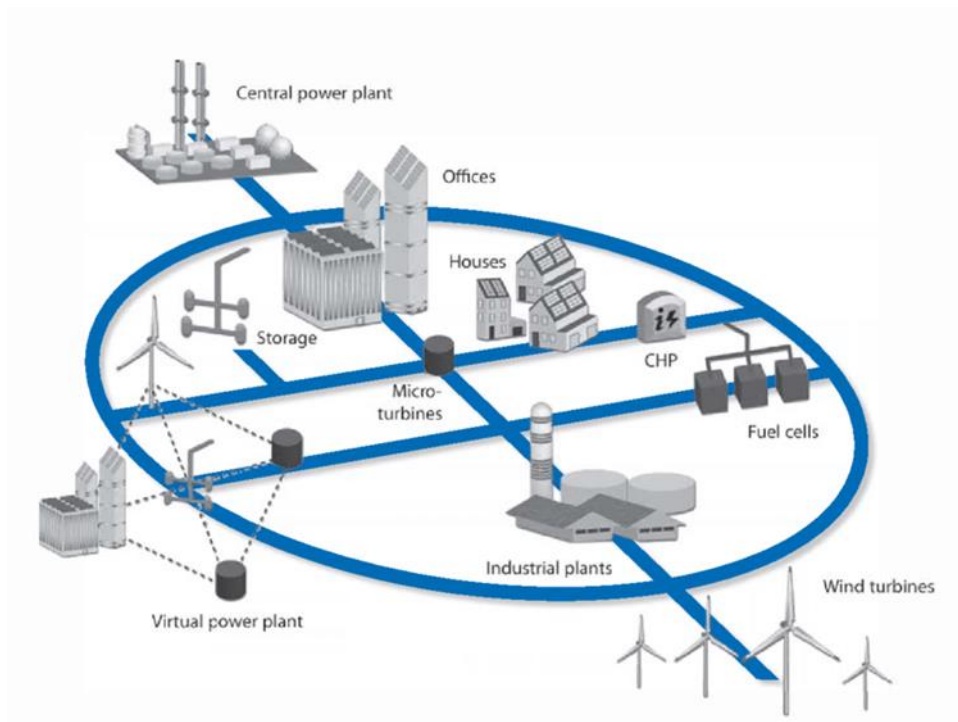
Σε αντίθεση με την κάθετη διαστρωμάτωση του κλασικού ΔΗΕ η οποία αποτελείται από τους παραγωγούς και τους καταναλωτές, η ανάπτυξη των ευφυών δικτύων αναμένεται να εισάγει νέους συμμετέχοντες. Το National Institute of Standards and Technology (NIST) των ΗΠΑ ορίζει τους τομείς του δικτύου και κατηγοριοποιεί τους συμμετέχοντες στις κατηγορίες που παρουσιάζονται στον Πίνακα 2.1, ορίζοντας με σαφήνεια τον ρόλο τους. [Gopstein et al., 2021].

Εξετάζοντας τον Πίνακα 2.1 είναι φανερό η διαφοροποίηση της διαστρωμάτωσης και των συμμετεχόντων εν συγκρίσει με το παρόν ΔΗΕ. Η είσοδος νέων συμμετεχόντων στην αγορά ενέργειας αναμένεται να αυξήσει τον ανταγωνισμό και κατ' επέκταση να μειώσει τις τιμές της ηλεκτρικής ενέργειας. Επιπλέον παρατηρείται η μετατροπή του καταναλωτή σε παραγωγό, δίνοντάς του διττό ρόλο.

Πίνακας 2.1 Τομείς και ρόλοι [Gopstein et all, 2021]

	Τομέας	Ρόλος
1	Πελάτης	Τελικός χρήστης του ηλεκτρισμού. Μπορεί επίσης να παράγει, αποθηκεύει, και να διαχειρίζεται την ενέργεια. Παραδοσιακά, χωρίζονται σε τρεις κατηγορίες : Οικιακοί, εμπορικοί, και βιομηχανικοί πελάτες.
2	Αγορές	Οι οργανωτές και

		συμμετέχοντες στις αγορές ηλεκτρισμού, καθώς και οι διάφοροι οικονομικοί μηχανισμοί που χρησιμοποιούνται για να καθορίσουν τις ενέργειες και να βελτιστοποιήσουν τα αποτελέσματα.
3	Πάροχοι Υπηρεσιών	Οι οργανισμοί που παρέχουν υπηρεσίες στους πελάτες.
4	Διαχειριστές (Operations)	Οι διαχειριστές της μεταφοράς του ηλεκτρικού ρεύματος.
5	Παραγωγή, συμπεριλαμβανομένων της διεσπαρμένης παραγωγής	Οι παραγωγοί του ηλεκτρισμού. Μπορούν επίσης να αποθηκεύουν ενέργεια.
6	Μεταφορά	Οι διαχειριστές της μεταφοράς υψηλής τάσης σε μεγάλες αποστάσεις. Μπορούν επίσης να είναι και παραγωγοί.
7	Διανομή	Οι υπεύθυνοι για την διανομή από και προς τους τελικούς καταναλωτές. Μπορούν επίσης να αποθηκεύουν και να παράγουν



Εικόνα 2.2 The European Union's Smart Grid vision -grid operation [Becks et all 2010]

Βάση των παραπάνω θα μπορούσε να δοθεί ο ακόλουθος ορισμός για το τι είναι το ευφυές δίκτυο ηλεκτρικής ενέργειας, ο οποίος είναι: Το ευφυές δίκτυο ηλεκτρικής ενέργειας αποτελεί την εξέλιξη του παρόντος δικτύου με τη μετατροπή της κάθετης διαστρωμάτωσης σε μια πιο ευέλικτη, όπου οι συμμετέχοντες λαμβάνουν παραπάνω από ένα ρόλο, με την ενσωμάτωση τεχνολογίας για την επικοινωνία όλων των μερών που το αποτελούν με στόχο της αύξηση της αξιοπιστίας, της απόδοσης, της αντοχής σε φυσικές και ψηφιακές επιθέσεις, καθώς και τον καλύτερο έλεγχο, και τέλος της προώθηση της χρήσης ΑΠΕ και διεσπαρμένης παραγωγής. Το όραμα της ΕΕ για το ευφυές δίκτυο παρουσιάζεται στην Εικόνα 2.2.



## Κεφάλαιο 3 Πρόβλεψη φορτίου

### 3.1 Εισαγωγή

Η εκ των προτέρων γνώση του φορτίου του δικτύου αποτελεί κρίσιμο παράγοντα για τη σωστή και εύρυθμη λειτουργία του. Η ακριβής πρόβλεψη είναι απαραίτητη στους διαχειριστές δικτύου για την λήψη αποφάσεων που αφορούν την λειτουργία του, την ένταξη μονάδων, τη διαχείριση φορτίου και την ανάπτυξη της υποδομής. Επιπλέον με την απελευθέρωση της αγοράς ενέργειας και την είσοδο των λιανοπωλητών (retailers) η πρόβλεψη φορτίου παίζει σημαντικό ρόλο στην λήψη αποφάσεων σχετικά με την διαμόρφωση των τιμών [Feinberg and Genethliou, n.d.].

Επιπλέον η αδυναμία αποθήκευσης της παραγόμενης ηλεκτρικής ενέργειας σε επαρκείς ποσότητες με χρήση οικονομικά βιώσιμων συστημάτων καθιστά απαραίτητη την διατήρηση της ισορροπίας μεταξύ της παραγόμενης και καταναλισκόμενης ενέργειας [Jahan, Snasel and Misak, 2020]. Κατά συνέπεια παρουσιάζεται η ανάγκη της συνεχούς παρακολούθησης του φορτίου με στόχο η παραγωγή ενέργειας να είναι όση και η ζητούμενη σε κάθε χρονική στιγμή [Fiot and Dinuzzo, 2018] με στόχο την διατήρηση των ποιοτικών χαρακτηριστικών της παρεχόμενης ενέργειας .

### 3.2 Κατηγοριοποίηση βάση χρονικού εύρους

Η πρόβλεψη φορτίου κατηγοριοποιείται ανάλογα με τον χρόνο πρόβλεψης σε:

A. Μακροπρόθεσμη πρόβλεψη φορτίου. Ο χρονικός ορίζοντας της είναι από 1 έως 10 έτη. Κυρίως χρησιμοποιείται στο μακροχρόνιο σχεδιασμό του συστήματος, στον προγραμματισμό της συντήρησης των μονάδων παραγωγής και του συστήματος διανομής ενέργειας. Αποτελεί βασικό εργαλείο των εταιριών όσο αφορά τον σχεδιασμό και κατασκευή νέων μονάδων παραγωγής και επέκτασης του δικτύου.

B. Μεσοπρόθεσμη πρόβλεψη φορτίου, με ορίζοντα από 1 εβδομάδα έως ένα έτος. Η χρήση της έγκειται στον προγραμματισμό συντήρησης των μονάδων παραγωγής και του δικτύου.

Γ. Βραχυπρόθεσμη πρόβλεψη φορτίου με ορίζοντα πρόβλεψης από μία ημέρα έως μια εβδομάδα. Η χρήση της καθίσταται ζωτική για την ανάλυση της ένταξης μονάδων παραγωγής και κατά συνέπεια τον οικονομικό προγραμματισμό και τον αυτόματο έλεγχο της παραγωγής ενέργειας.

Δ. Πολύ βραχυπρόθεσμη πρόβλεψη, με ορίζοντα χιλιοστών του δευτερολέπτου με 30 λεπτά και χρήση στα συστήματα αυτομάτου ελέγχου των γεννητριών, την δυναμική ανάλυση του συστήματος και στη ανάλυση ευστάθειας του συστήματος.

### **3.3 Παράγοντες που επηρεάζουν το φορτίο συστήματος**

Το φορτίο συστήματος επηρεάζεται από έναν αριθμό παραγόντων οι οποίοι επιδρούν με διαφορετικό τρόπο σε αυτό. Λαμβάνοντας υπόψη τον χρονικό ορίζοντα της πρόβλεψης, η βραχυπρόθεσμη πρόβλεψη επηρεάζεται από γεγονότα τα οποία δεν είναι εποχικά, όπως για παράδειγμα η μεταβολή των καιρικών συνθηκών. Στη μεσοπρόθεσμη πρόβλεψη οι σημαντικοί παράγοντες χαρακτηρίζονται από την εποχικότητα της εμφάνισής τους, για παράδειγμα οι εποχές του έτους. Στη μακροπρόθεσμη, τέλος η κυκλικότητα των παραγόντων είναι μείζονος σημασίας, για παράδειγμα οι οικονομικοί δείκτες, το ΑΕΠ, ο πληθυσμός [Xue and Geng, 2012].

Η κύρια συνιστώσα που επηρεάζει την βραχυπρόθεσμη πρόβλεψη είναι οι καιρικές συνθήκες. Η θερμοκρασία, η υγρασία, η ένταση της ηλιακής ακτινοβολίας, η ταχύτητα του αέρα, η βροχόπτωση επηρεάζουν το φορτίο συστήματος. Η θερμοκρασία αποτελεί την κύρια συνιστώσα, και συνδέεται άμεσα με το φορτίο, για παράδειγμα τον χειμώνα η μείωση της θερμοκρασίας έχει ως αποτέλεσμα την αύξηση της κατανάλωσης ενέργειας για θέρμανση, και τους ζεστούς μήνες του έτους να αυξάνεται η κατανάλωση λόγω της χρήσης κλιματιστικών μονάδων [Khatoon, Ibraheem, Singh and Priti, 2014].

### **3.4 Μοντέλα πρόβλεψης φορτίου ηλεκτρικής ενέργειας.**

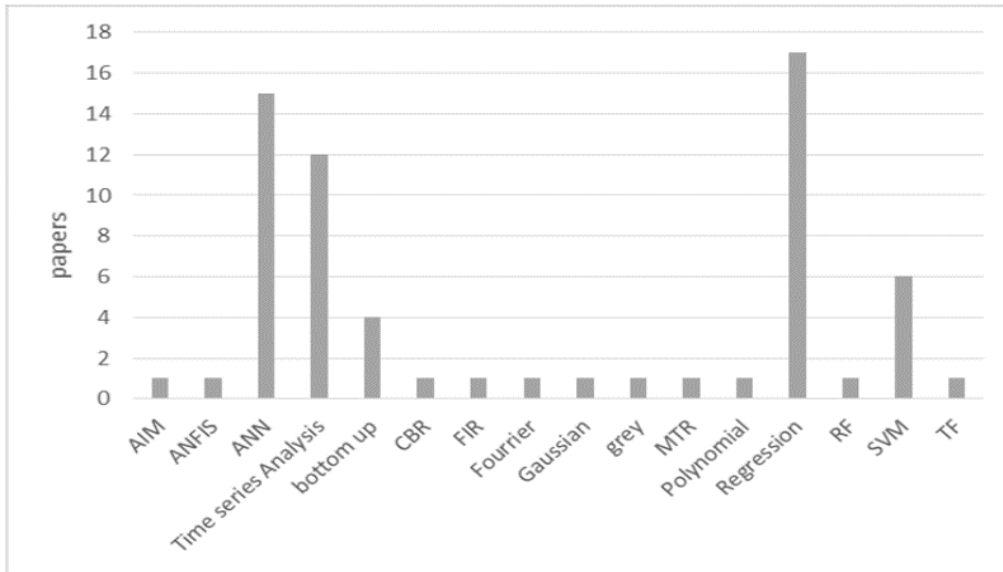
Η αναγκαιότητα της ακριβούς εκτίμησης του φορτίου ενός συστήματος ηλεκτρικής ενέργειας για μια καθορισμένη χρονική περίοδο, για τους λόγους που αναλύθηκαν στο

κεφάλαιο 3.1, οδήγησε στην ανάπτυξη μιας πληθώρας μεθόδων για την πρόβλεψή του. Οι μέθοδοι που εφαρμόστηκαν διαφοροποιούνται ανάλογα με το χρονικό εύρος πρόβλεψης. Η συγκριτική ανάλυση που έγινε από τον Kuster et all [Kuster, Rezgui and Mourshed, 2017] ανέλυσε έναν μεγάλο αριθμό ακαδημαϊκών μελετών, με στόχο τον προσδιορισμό της καταλληλότητας του εκάστοτε μοντέλου ανάλογα με την περίπτωση. Σημαντικός αριθμός παραγόντων λήφθηκε υπόψιν κατά την διαδικασία της αξιολόγησης, όπως οι είσοδοι, οι έξοδοι και το χρονικό πλαίσιο. Οι χρησιμοποιούμενες μέθοδοι κατηγοριοποιήθηκαν ως εξής.

- Μοντέλα χρονοσειρών, η στατιστικά μοντέλα. Η κατηγορία αυτή περιλαμβάνει έναν αριθμό μεθόδων όπως η Autoregression, Moving Average, ARMA, SARMA, SARIMAX.
- Μοντέλα γραμμικής παλινδρόμησης (linear regression models)
- Μοντέλα τεχνητών νευρωνικών δικτύων
- Μοντέλα Support Vector Machines
- Μοντέλα Bottom up

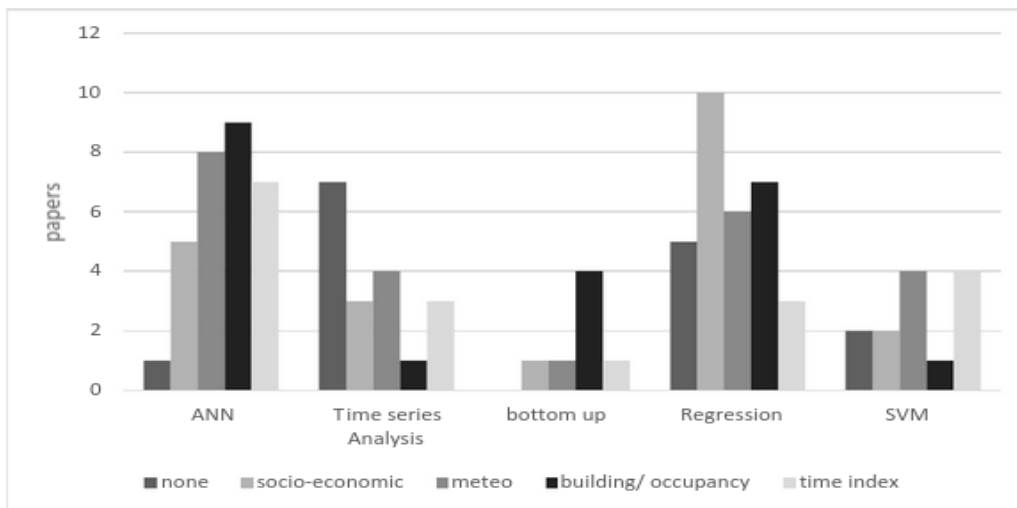
Βάση του γραφήματος της Εικόνας 3.1 είναι εμφανής η προτίμηση των ερευνητών στην παλινδρόμηση, τα τεχνητά νευρωνικά δίκτυα, την ανάλυση χρονοσειρών και τα μοντέλα Support Vector Machine.

Τα αποτελέσματα της ανάλυσης φανερώνουν την αποτελεσματικότητα της παλινδρόμησης όταν εφαρμόζεται σε σενάρια μακροπρόθεσμης πρόβλεψης, των τεχνητών νευρωνικών δικτύων, των Support vector machines και της ανάλυσης χρονοσειρών στην βραχυπρόθεσμη πρόβλεψη φορτίου.



Εικόνα 3.1 Classified forecasting models distribution, [Kuster et all, 2017]

Όσο αφορά τις εισόδους, η αποκλειστική χρήση ιστορικών δεδομένων του ηλεκτρικού φορτίου εφαρμόζεται κατά κόρον στα μοντέλα ανάλυσης χρονοσειρών και στις εφαρμογές παλινδρόμησης. Στα TNN είθισται να χρησιμοποιούνται ένας αριθμός εξωγενών μεταβλητών, όπως τα μετεωρολογικά και κλιματικά στοιχεία. Το διάγραμμα της Εικόνας 3.2 κατηγοριοποιεί τις εισόδους ανάλογα με το μοντέλο πρόβλεψης.



Εικόνα 3.2 Είσοδοι ανάλογα με το μοντέλο πρόβλεψης, [Kuster et all, 2017]

### 3.5 Μοντέλα Χρονοσειρών

Οι χρονοσειρές αποτελούνται από παρατηρήσεις της τιμής μιας μεταβλητής σε ένα δεδομένο χρονικό διάστημα. Η συχνότητα παρατήρησης είναι συνήθως σταθερή. Παραδείγματα χρονοσειρών είναι η κατανάλωση ενέργειας ανά μονάδα χρόνου, οι πωλήσεις ενός προϊόντος, ο αριθμός αυτοκινήτων που διασχίζει τα διόδια ενός αυτοκινητοδρόμου. Οι μεταβλητές που εξετάζονται κατά την διαδικασία της ανάλυσης αφορούν την αυτοσυσχέτιση, την εποχικότητα και την στατικότητα. Η αυτοσυσχέτιση εξετάζει την συσχέτιση των τιμών της μεταβλητής ως συνάρτηση της μεταξύ τους χρονικής υστέρησης. Η εποχικότητα αφορά τις περιοδικές διακυμάνσεις, όπως η αυξημένη κίνηση κατά τις πρώτες πρωινές ώρες όπου οι εργαζόμενοι μεταβαίνουν στον τόπο εργασίας τους και τις απογευματινές λόγω της επιστροφής στην οικία τους. Τέλος, η στατικότητα αναφέρεται στην σταθερότητα των στατιστικών ιδιοτήτων της χρονοσειράς στη διάρκεια του χρόνου, συγκεκριμένα στον σταθερό μέσο και στη διακύμανση.

### 3.6 Παλινδρόμηση.

Η παλινδρόμηση είναι μια τεχνική που εξερευνά την συσχέτιση μεταξύ μιας μεταβλητής η οποία λαμβάνει συνεχείς τιμές και ονομάζεται εξαρτημένη, και ενός αριθμού μεταβλητών που ονομάζονται ανεξάρτητες και μπορεί να λαμβάνουν οποιοδήποτε τύπου τιμές. Το μοντέλο της μαθηματικά εκφράζεται από την σχέση 3.1 όπου  $\beta$  η άγνωστη παράμετρος,  $X$  οι ανεξάρτητη μεταβλητή,  $Y$  η εξαρτημένη μεταβλητή και  $\varepsilon$  το σφάλμα. [Montgomery and Peck, 1992].

$$Y_i = f(X_i, \beta) + \varepsilon_i \quad (3.1)$$

Στην περίπτωση της απλής γραμμικής παλινδρόμησης η σχέση μεταξύ των μεταβλητών είναι γραμμική και εκφράζεται μαθηματικά από την σχέση 3.2, όπου  $y$  η εξαρτημένη μεταβλητή,  $\beta_0$  και  $\beta_1$  οι παράμετροι, και  $x$  η ανεξάρτητη μεταβλητή.

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \text{ με } i = 1, \dots, n \quad (3.2)$$

Στη πολλαπλή γραμμική παλινδρόμηση υπάρχουν περισσότερες από 1 ανεξάρτητες μεταβλητές με την μαθηματική έκφραση να παίρνει την μορφή της σχέσης 3.3.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i \text{ με } i = 1, \dots, n \quad (3.3)$$

Η μη γραμμική παλινδρόμηση εξετάζει μοντέλα των οποίων οι μεταβλητές δεν έχουν γραμμική σχέση μεταξύ τους. Ως μη γραμμικό, ορίζεται το μοντέλο κατά το οποίο ο ρυθμός μεταβολής μεταξύ εξαρτημένης και ανεξάρτητης μεταβλητής δεν είναι σταθερός. Η απλούστερη μορφή ενός μη γραμμικού μοντέλου είναι τα πολυωνυμικά των οποίων η γενική μαθηματική έκφραση παρουσιάζεται στην σχέση 3.4. [Nonlinear Regression Analysis and Its Applications, 1988].

$$E(y) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_p x^p \quad \text{με } p > 3 \quad (3.4)$$

### 3.7 Support Vector machines

Οι support vector machines (SVM) είναι ένα μοντέλο μηχανικής εκμάθησης δημιουργημένο από τον V. Vapnik με χρήση στη γραμμική ταξινόμηση. Μέχρι την έλευση της βαθιάς μηχανικής εκμάθησης η χρήση τους ήταν ευρεία στους τομείς της ταξινόμησης και παλινδρόμησης. Βασίζονται στην θεωρία της στατιστικής μάθησης. [Cristianini and Ricci, 2008]

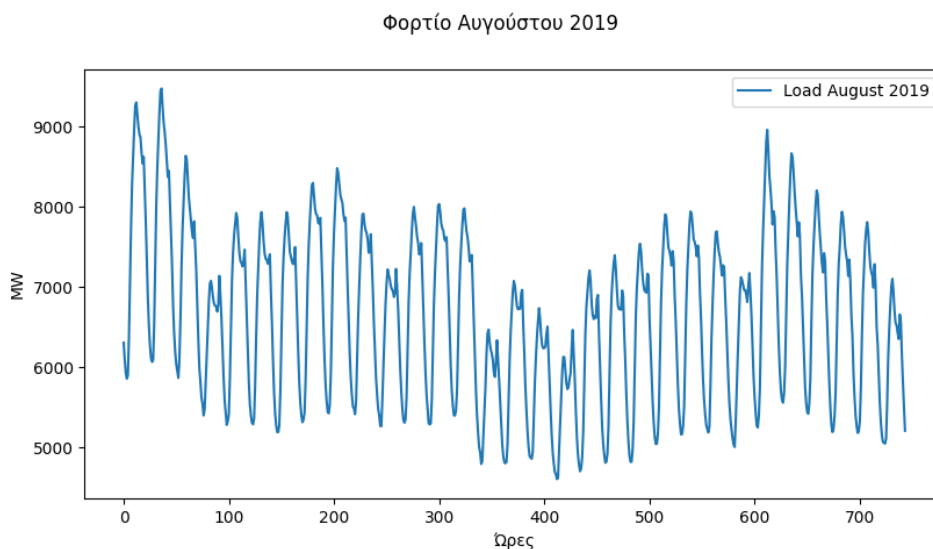
### 3.8 Μοντέλα Bottom up

Το μοντέλο bottom up ακολουθεί τη λογική της σύνθεσης πολύπλοκων μοντέλων από επιμέρους απλά μέρη. Αναπτύχθηκε από τον Carasso [Carasso et al, 1994]. Στον τομέα της πρόβλεψης φορτίου το μοντέλο εξετάζει την κατανάλωση ενός βασικού συστήματος όπως μια οικία, μοντελοποιώντας την χρήση των ηλεκτρικών συσκευών και κατ' επέκταση την ζήτηση φορτίου. Η προσέγγιση εξετάζει την πιθανότητα μια συσκευή να είναι ανοιχτή ή κλειστή, λαμβάνοντας υπόψιν κοινωνικά και οικονομικά κριτήρια, καθώς και παράγοντες που έχουν σχέση με την ψυχολογία του ατόμου και του τρόπου ζωής. Με την επέκταση της σε έναν μεγάλο αριθμό βασικών μονάδων, και δεδομένης της σωστής και ακριβούς πρόβλεψης της βασικής, το μοντέλο μπορεί να γενικευτεί ώστε να προβλέψει το σύνολο της κατανάλωσης ενός τομέα.

## Κεφάλαιο 4 Ανάλυση φορτίου και καθορισμός του προβλήματος των ειδικών ημερών.

### 4.1 Φορτίο στις κανονικές ημέρες

Η κατανάλωση ηλεκτρικής ενέργειας στις κανονικές ημέρες συνδέεται άμεσα με την ανθρώπινη δραστηριότητα. Στο γράφημα της Εικόνας 4.1 απεικονίζεται το φορτίο από την 1<sup>η</sup> έως την 31<sup>η</sup> Αυγούστου το οποίο αποτελεί την απεικόνιση του καλοκαιρινού φορτίου του συστήματος.



Εικόνα 4.1 Φορτίο Αυγούστου 2019

Αρχικά, παρατηρείται η κατανάλωση κατά την διάρκεια της ημέρας. Στις πρωινές ώρες υπάρχει μικρή κατανάλωση, η οποία αυξάνεται στην συνέχεια συνέπεια της εντατικοποίησης της ανθρώπινης δραστηριότητας τόσο σε οικιακό όσο και επαγγελματικό επίπεδο. Παρατηρείται μια τοπική κορύφωση στις πρώτες μεσημβρινές ώρες, και στην συνέχεια μια μείωση κατά τις ώρες της μεσημεριανής ανάπαυσης. Τέλος η κατανάλωση κορυφώνεται τις πρώτες βραδινές ώρες.

Είναι φανερή η διαφοροποίηση κατά την διάρκεια της εβδομάδος και της Κυριακής. Οι Κυριακές ως ημέρες αργίας κατά τις οποίες η δραστηριότητα μειώνεται παρουσιάζουν

αντίστοιχα μειωμένη κατανάλωση ενέργειας. Σε αντίθεση οι εργάσιμες ημέρες παρουσιάζουν την μεγαλύτερη κατανάλωση.

Επιπλέον μπορούμε να παρατηρήσουμε την διαφοροποίηση των καμπυλών φορτίου κατά την διάρκεια των εποχών. Τόσο στην καλοκαιρινή περίοδο, όσο και την χειμερινή οι μεταβολές της θερμοκρασίας συντελούν στην αυξομείωση της απαίτησης για ηλεκτρική ενέργεια όπως αναλύθηκε στο κεφάλαιο 3.3. Συνοψίζοντας, παρατηρούμε τέσσερις διαφορετικούς κύκλους. Τον ημερήσιο, τον εβδομαδιαίο τον εποχικό κύκλο και τον ετήσιο κύκλο.

## **4.2 Φορτίο στις ειδικές ημέρες.**

Η κατανάλωση ηλεκτρικής ενέργειας είναι άρρηκτα συνδεδεμένη με την ανθρώπινη δραστηριότητα. Οι κοινωνικές συμπεριφορές των ανθρώπων αποτελούν βασική μεταβλητή στα ενεργειακά μοντέλα [Pfenninger, Hawkes and Keirstead, 2014], αναγνωρίζοντας επανεμφανιζόμενα μοτίβα κατά την διάρκεια του χρόνου επηρεαζόμενα από καιρικές συνθήκες, πολιτισμικά έθιμα και ενεργειακές πολιτικές, μεταξύ άλλων. Η εμπορική, βιομηχανική και οικιακή δραστηριότητα αποτελούν τους τρεις βασικούς τομείς κατανάλωσης, η καθεμιά με διαφορετική συμβολή στο φορτίο του συστήματος.

Η ανθρώπινη δραστηριότητα μεταβάλλεται κατά την διάρκεια των ειδικών ημερών, στις οποίες λόγω κοινωνικών και πολιτισμικών συνθηκών οι δύο κυριότεροι τομείς κατανάλωσης παύουν τη λειτουργία τους ή λειτουργούν σε χαμηλότερα επίπεδα παραγωγής, και κατά συνέπεια κατανάλωσης. Η κατανάλωση ανά τομέα διαφέρει. Στην περίπτωση της Ελλάδος, η ηλεκτρική ενέργεια καταναλώνεται από τον εμπορικό τομέα σε ποσοστό 34,6%, τον οικιακό τομέα σε ποσοστό 33,6% και τον βιομηχανικό τομέα με ποσοστό της τάξης του 23,8% [Βέττας et al., 2021]. Μπορούμε λοιπόν να σχηματίσουμε την υπόθεση πως η επίδραση των ειδικών ημερών λόγω της παύσης λειτουργίας των εμπορικών και βιομηχανικών καταναλωτών αναμένεται μείωση της κατανάλωσης ηλεκτρικής ενέργειας εν συγκρίσει με τις κανονικές ημέρες.



### 4.3 Ειδικές ημέρες στην Ελλάδα.

Οι αργίες ή ειδικές ημέρες στην Ελλάδα χωρίζονται σε δυο μεγάλες κατηγορίες. Τις υποχρεωτικές αργίες που καθορίζονται από τον νόμο και παρουσιάζονται στον Πίνακα 4.1.

Πίνακας 4.1 Υποχρεωτικές Αργίες

Αργία	Ημερομηνία	Νομοθεσία
Πρωτοχρονιά	1 <sup>η</sup> Ιανουαρίου	Άρθ. 60 Ν.4808/2021
Θεοφάνια	6 <sup>η</sup> Ιανουαρίου	Άρθ. 60 Ν.4808/2021
Η 25 <sup>η</sup> Μαρτίου	25 <sup>η</sup> Μαρτίου	Άρθ. 60 Ν.4808/2021
Δευτέρα του Πάσχα	Κινητή Αργία	Άρθ. 60 Ν.4808/2021
Εργατική Πρωτομαγιά	1 <sup>η</sup> Μαΐου	Άρθ. 14 Ν. 4468/2017
Κοίμηση της Θεοτόκου	15 <sup>η</sup> Αυγούστου	Β.Δ. 748/1966
Εθνική Εορτή της 28 <sup>ης</sup> Οκτωβρίου	28 <sup>η</sup> Οκτωβρίου	Άρθ.60 Ν.4808/2021
Εορτή της Γεννήσεως του Χριστού	25 <sup>η</sup> Δεκεμβρίου	Β.Δ. 748/1966
Δεύτερη ημέρα των χριστουγέννων	26 <sup>η</sup> Δεκεμβρίου	Άρθ42 Ν. 4554/2018

Και οι αργίες κατ' έθιμο οι οποίες παρουσιάζονται στον Πίνακα 4.2.

Πίνακας 4.2 Αργίες κατ' έθιμο

Αργία	Ημερομηνία
Καθαρά Δευτέρα	Κινητή αργία, 48 ημέρες πριν το Πάσχα
Μεγάλη Παρασκευή	Κινητή αργία, δύο ημέρες πριν το Πάσχα
Κυριακή του Πάσχα	Αργία στα πλαίσια της Κυριακής με την ίδια νομοθεσία (Β.Δ 748/1966)

Αγίου Πνεύματος	Κινητή αργία, Δευτέρα μετά την Πεντηκοστή
-----------------	---

Κατά τις αργίες κατ' έθιμο η λειτουργία των επιχειρήσεων εξαρτάται από την βούληση του εργοδότη. Για τον δημόσιο τομέα, επιπλέον αργία είναι το Μεγάλο Σάββατο. Επιπλέον των παραπάνω υπάρχουν τοπικές αργίες που έχουν να κάνουν με τοπικά έθιμα ή πολιούχους των εκάστοτε πόλεων. Τέλος, η Κυριακή θεωρείται αργία βάση του Β.Δ. 748/1966 με τις εξαιρέσεις του άρθρου 2 του ίδιου νόμου.

Όπως μπορούμε να διαπιστώσουμε οι αργίες στην Ελλάδα είναι ένα μείγμα θρησκευτικών εθιμικών και εθνικών εορτών. Οι θρησκευτικές αργίες είναι ένας συνδυασμός σταθερών και κινητών εορτών. Οι εθιμικές εορτές (Πρωτοχρονιά, Εργατική Πρωτομαγιά) είναι σταθερές εορτές. Τέλος, οι δύο εθνικές εορτές είναι σταθερές.

Εκτός της σταθερής ή κινητής κατάστασης, οι αργίες εξετάζονται στο κατά πόσο προκύπτουν σε σταθερή ημέρα ή όχι. Από αυτή την σκοπιά τις διαχωρίζουμε σε δύο κατηγορίες βάση του αν λαμβάνουν χώρα σε σταθερή ημέρα της εβδομάδος ή όχι. Τα αποτελέσματα συνοψίζονται στον Πίνακα 4.3.

Πίνακας 4.3 Κατηγοριοποίηση βάση ημέρας

Αργία	Ημέρα
Πρωτοχρονιά	Διαφορετική ημέρα
Θεοφάνια	Διαφορετική ημέρα
Καθαρά Δευτέρα	Πάντα Δευτέρα
25 <sup>η</sup> Μαρτίου	Διαφορετική ημέρα
Μεγάλη Παρασκευή	Πάντα Παρασκευή
Κυριακή του Πάσχα	Πάντα Κυριακή
Δευτέρα του Πάσχα	Πάντα Δευτέρα
Εργατική Πρωτομαγιά	Διαφορετική Ημέρα
Αγίου Πνεύματος	Πάντα Δευτέρα
Κοίμηση της Θεοτόκου	Διαφορετική ημέρα

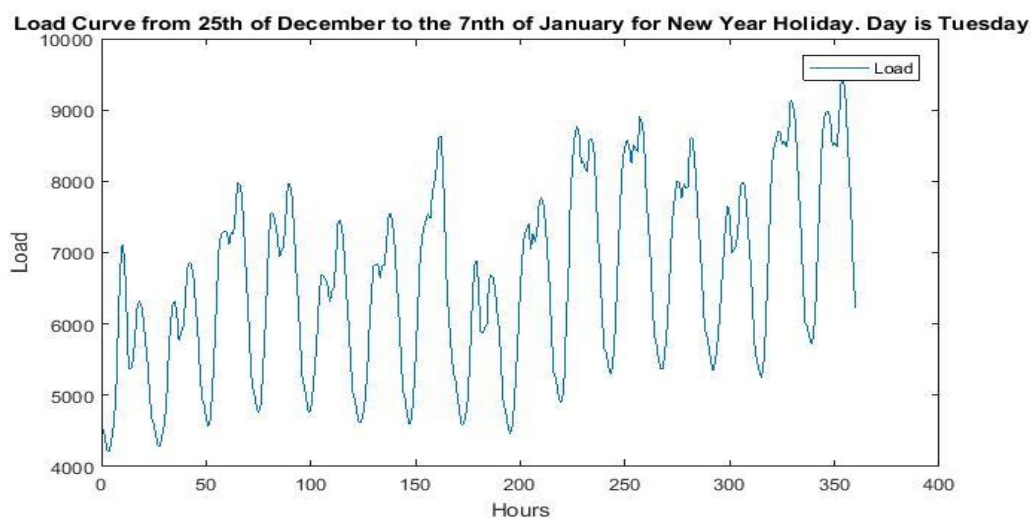
28 <sup>η</sup> Οκτωβρίου	Διαφορετική ημέρα
Χριστούγεννα	Διαφορετική ημέρα
2 <sup>η</sup> ημέρα των Χριστουγέννων	Διαφορετική ημέρα

Από τον Πίνακα 4.3 μπορούμε να συμπεράνουμε πως η πλειονότητα των αργιών λαμβάνει χώρα σε διαφορετικές ημέρες της εβδομάδος κάτι που συνήθως λαμβάνεται υπόψιν κατά την διαδικασία της δημιουργίας του μοντέλου πρόβλεψης.

#### 4.4 Ανάλυση φορτίου ειδικών ημερών.

##### 1. Πρωτοχρονιά.

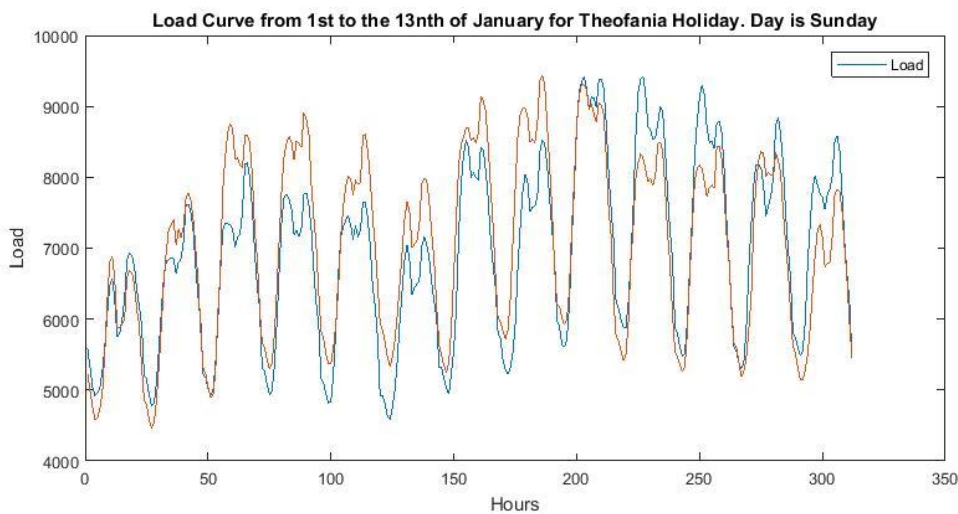
Η Πρωτοχρονιά αποτελεί σταθερή αργία με μεταβλητή ημέρα. Το προφίλ φορτίου διαφέρει, όπως φαίνεται και στο γράφημα της Εικόνας 4.2 στο οποίο παρουσιάζεται η καμπύλη κατανάλωσης ηλεκτρικής ενέργειας από την 25<sup>η</sup> Δεκεμβρίου 2018 έως και την 7<sup>η</sup> Ιανουαρίου του 2019 για λόγους σύγκρισης. Μπορούμε εύκολα να διαπιστώσουμε την μεταβολή της καμπύλης σε σύγκριση τόσο με την προηγούμενη ημέρα όσο και με τις αντίστοιχες  $D_{+7}$  και  $D_{-7}$ . Ειδικότερα η  $D_{+7}$  παρουσιάζει την μεγαλύτερη διαφοροποίηση και αποτελεί εμφανή ένδειξη για την επίδραση των ειδικών ημερών στη ζήτηση φορτίου.



Εικόνα 4.2 Σύγκριση Φορτίου Πρωτοχρονιάς

## 2. Θεοφάνια.

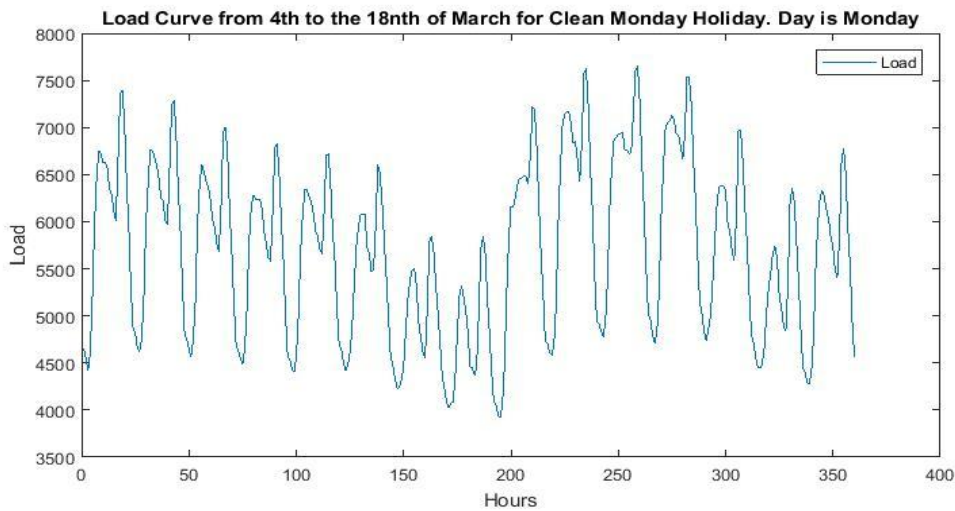
Τα Θεοφάνια είναι μια ειδική ημέρα που εμπίπτει στην κατηγορία των σταθερών εορτών με μεταβαλλόμενη ημέρα. Η αργία το 2019 παρουσιάζεται την Κυριακή, με το φορτίο της Εικόνας 4.3. Για λόγους σύγκρισης, παραθέτουμε και τα δεδομένα του 2017 με κόκκινο χρώμα, όπου η ημέρα κατά την οποία προκύπτει η αργία είναι η Παρασκευή. Παρατηρούμε την μεγάλη διαφοροποίηση του έτους 2017 κατά το οποίο η αργία προκύπτει εργάσιμη ημέρα, και λιγότερο το 2019 όπου προκύπτει Κυριακή.



Εικόνα 4.3 Σύγκριση Φορτίου Θεοφανίων

## 3. Καθαρά Δευτέρα.

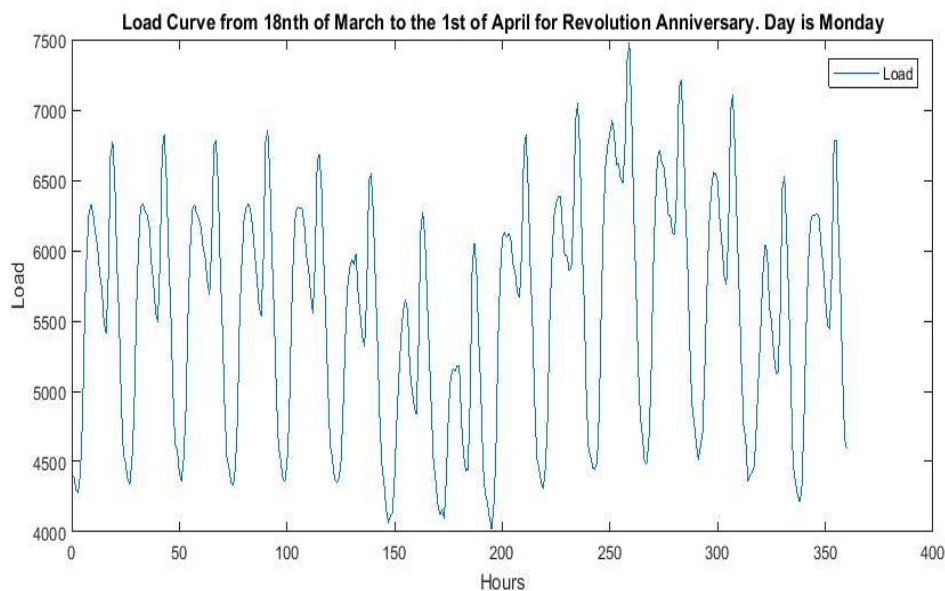
Η ειδική ημέρα ανήκει στις κινητές εορτές με σταθερή ημέρα. Παρουσιάζεται η καμπύλη φορτίου για το έτος 2019 στο γράφημα της Εικόνας 4.4. Παρατηρείται η μεγάλη μείωση της κατανάλωσης η οποία είναι μεγαλύτερη τόσο από την προηγούμενη που είναι Κυριακή όσο και από τις αντίστοιχες  $D_{+7}$  και  $D_{-7}$ .



Εικόνα 4.4 Σύγκριση Φορτίου Καθαράς Δευτέρας

#### 4. 25<sup>η</sup> Μαρτίου.

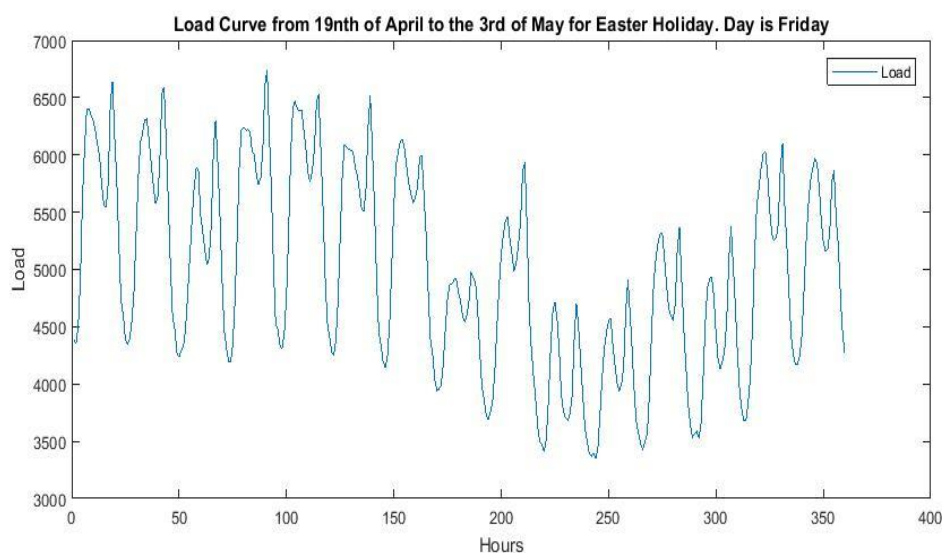
Αποτελεί εθνική εορτή η οποία είναι σταθερή, με μεταβλητή ημέρα. Το 2019 προκύπτει ημέρα Δευτέρα. Παρόμοια παρατηρούμε κατά την εξέταση της Εικόνας 4.5 την μείωση του καταναλισκόμενου φορτίου, τόσο εν συγκρίσει με την προηγούμενη ημέρα όσο και με τις αντίστοιχες  $D_{+7}$  και  $D_{-7}$ .



Εικόνα 4.5 Σύγκριση Φορτίου Επετείου της Επανάστασης

## 5. Πάσχα 2019.

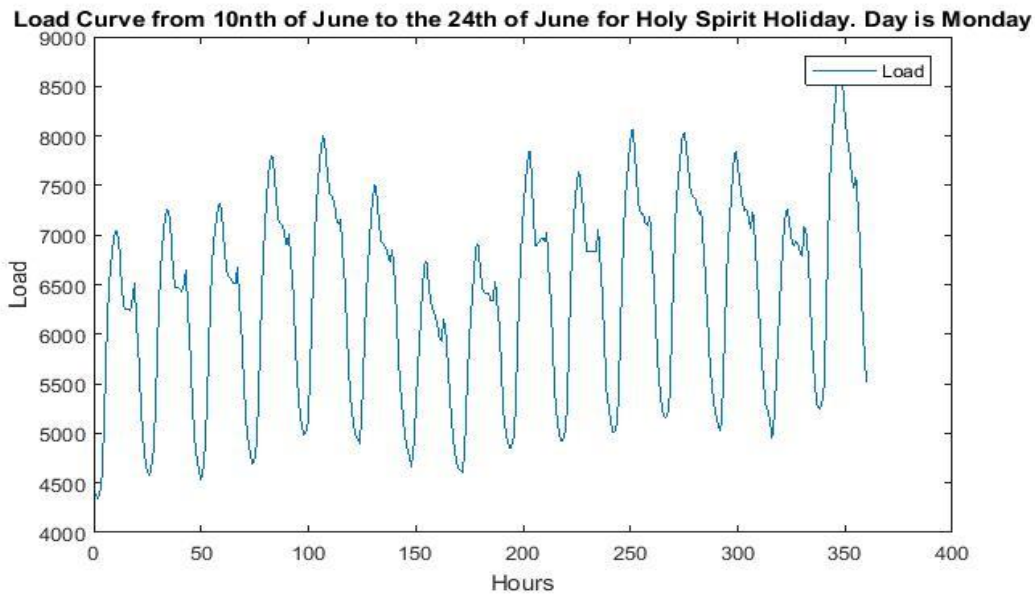
Η εορτή του Πάσχα είναι η πρώτη από τις δυο ιδιαίτερες ημερολογιακά ειδικές ημέρες καθότι συνδυάζει 3 αργίες σε πολύ σύντομο χρονικό διάστημα. Ιδιαίτερα για το έτος 2019, οι αργίες γίνονται 4 καθώς σε πολύ κοντινή ημερομηνία είναι η αργία της εργατικής Πρωτομαγιάς. Αναλύοντας το γράφημα της Εικόνας 4.6 παρατηρείται η μείωση κατανάλωσης κατά την μεγάλη Παρασκευή από την 169<sup>η</sup> ώρα του γραφήματος, στη συνέχεια την αύξηση της ζήτησης κατά το μεγάλο Σάββατο που είναι εργάσιμη, και την μείωση κατά τις ημέρες της Κυριακής και της Δευτέρας, ημέρες που είναι αργίες. Τέλος, φαίνεται και η επίδραση της εργατικής πρωτομαγιάς, όπου το φορτίο μειώνεται.



Εικόνα 4.6 Σύγκριση Φορτίου Πάσχα - Εργατικής Πρωτομαγιάς

## 6. Αργία του Αγίου Πνεύματος.

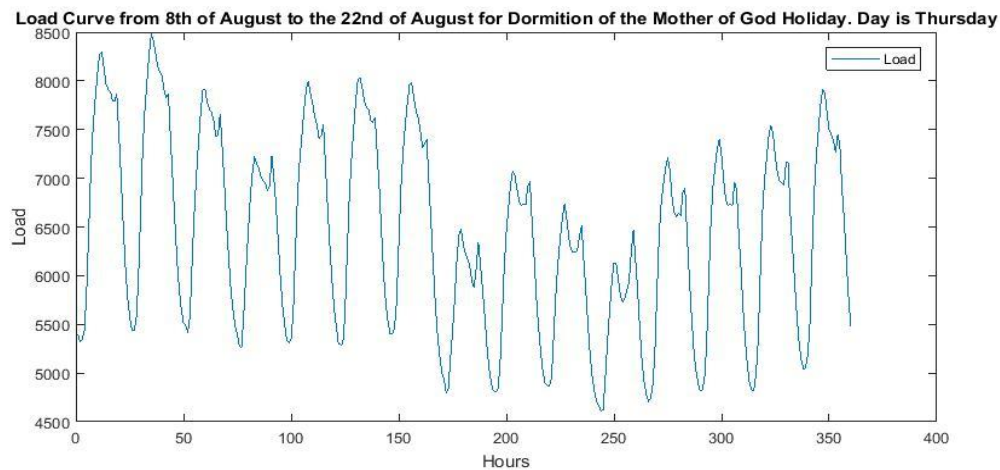
Η συγκεκριμένη αργία είναι στην κατηγορία των κινητών αργιών που προκύπτουν σε σταθερή ημέρα. Προκύπτει πάντα Δευτέρα, μετά την Πεντηκοστή. Η καμπύλη φορτίου παρουσιάζεται στην Εικόνα 4.7 Παρατηρείται η μείωση της ζήτησης συνέπεια της παύσης των δυο εκ των τριών τομέων κατανάλωσης ενέργειας. Επιπλέον, παρατηρούμε την ομοιότητα με την προηγούμενη ημέρα, από την οποία εμφανίζει ελαφρώς αυξημένη κατανάλωση και την διαφοροποίηση των D+7 και D-7.



Εικόνα 4.7 Σύγκριση Φορτίου Αγίου Πνεύματος

#### 7. Κοίμηση της Θεοτόκου.

Η αργία της Κοιμήσεως της Θεοτόκου λαμβάνει χώρα σε σταθερή ημερομηνία με μεταβαλλόμενη ημέρα. Για το 2019 η αργία λαμβάνει χώρα ημέρα Πέμπτη. Εξετάζοντας την Εικόνα 4.8 παρατηρούμε την εμφανή μείωση του φορτίου.

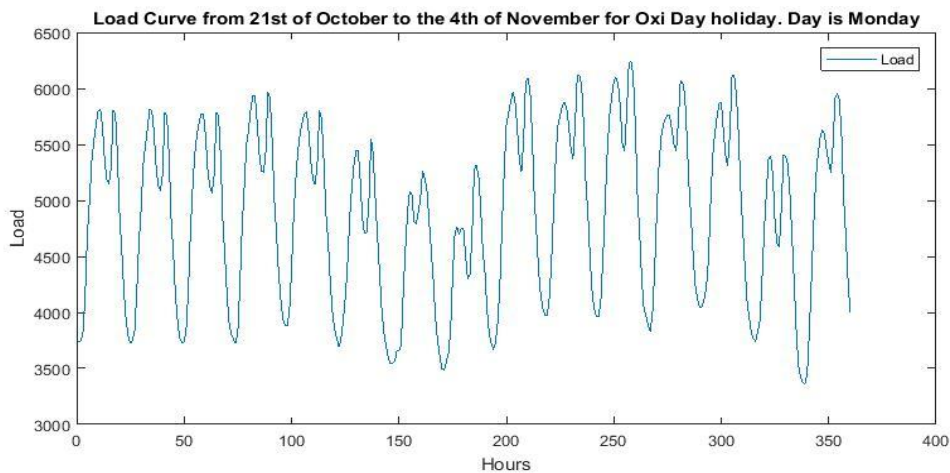


Εικόνα 4.8 Σύγκριση Φορτίου Κοιμήσεως της Θεοτόκου

#### 8. Επέτειος του Όχι, 28<sup>η</sup> Οκτωβρίου.

Ανήκει στην κατηγορία των σταθερών εορτών με μεταβλητή ημέρα. Στο προς εξέταση έτος προκύπτει ημέρα Δευτέρα. Η εξέταση της Εικόνας 4.9 που απεικονίζει το φορτίο

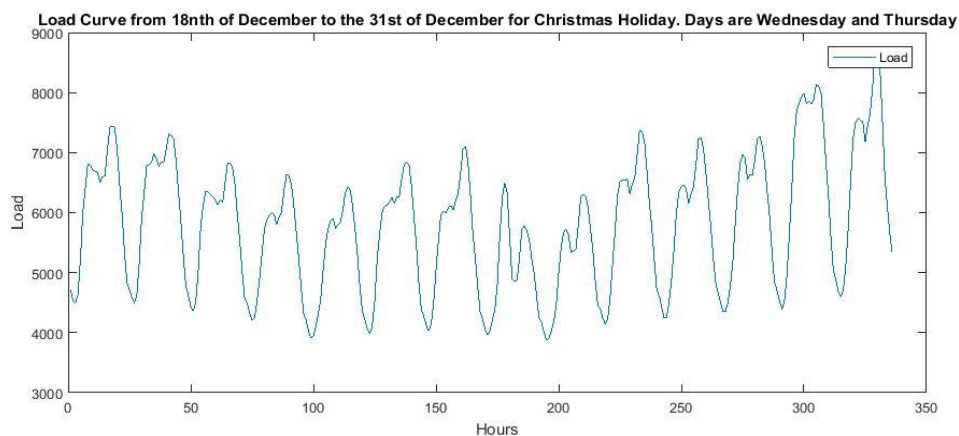
οδηγεί σε παρόμοια συμπεράσματα σε σχέση με τις προηγούμενες αργίες, δηλαδή την μείωση της κατανάλωσης ηλεκτρικής ενέργειας.



Εικόνα 4.9 Σύγκριση Φορτίου Επετείου του Όχι

#### 9. Διήμερο των Χριστουγέννων.

Ο δεύτερος ιδιαίτερος συνδυασμός αργιών προκύπτει κατά την αργία των Χριστουγέννων, όπου εξετάζονται δυο συνεχόμενες αργίες, η ημέρα των Χριστουγέννων και η δεύτερη ημέρα. Πρόκειται για σταθερές αργίες που προκύπτουν σε διαφορετική ημέρα. Η εξέταση του γραφήματος της Εικόνας 4.10 οδηγεί στο συμπέρασμα της μειωμένης κατανάλωσης και τις δυο ημέρες, που είναι Τετάρτη και Πέμπτη. Εμφανής είναι και η διαφορά με τις αντίστοιχες  $D_{+7}$  και  $D_{-7}$ .



Εικόνα 4.10 Σύγκριση Φορτίου Χριστουγέννων και δεύτερης ημέρας Χριστουγέννων



## 4.5 Το πρόβλημα των ειδικών ημερών

Η ανάλυση του φορτίου που πραγματοποιήθηκε στο κεφάλαιο 4.4 οδηγεί στην επιβεβαίωση της υπόθεσης που έγινε στο κεφάλαιο 4.1. Η κατανάλωση κατά τις ειδικές ημέρες φαίνεται να είναι πάντα μικρότερη εν συγκρίσει με τις κανονικές ημέρες, κάτι που είναι φανερό από την εξέταση των διαγραμμάτων των Εικόνων 4.1 έως 4.9. Η εκτίμηση του φορτίου της επόμενης ημέρας για τις ειδικές ημέρες αποτελεί ένα δύσκολο πρόβλημα λόγω της διαφοροποίησης από τις κανονικές ημέρες και κατ' επέκταση των μειωμένων μοτίβων εκπαίδευσης.

Η αντιμετώπιση αυτού του προβλήματος που αναφέρθηκε στην προηγούμενη παράγραφο επιχειρήθηκε με την εφαρμογή μιας σειράς διαφορετικών μεθόδων.

Ο Zhu et all [Zhu, Geng and Wang, 2021] προτείνουν μια μέθοδο αντιστοίχισης μοτίβων και την χρήση του αλγορίθμου XGboost. Η καμπύλη φορτίου στις ειδικές ημέρες πέρα από την σύνθεση του φορτίου επηρεάζεται από τα τοπικά έθιμα και τον τρόπο ζωής, τα οποία αποτελούν και τους μη ποσοτικοποιήσιμους παράγοντες. Η ιδέα των συγγραφέων έγκειται στην ομοιότητα των μελλοντικών καμπυλών φορτίου με την ιστορική τάση υπό από παρόμοιες συνθήκες, με συνέπεια από τις καμπύλες δύο συναπών ημερών να μπορεί να εξαχθεί ένα μοτίβο. Το στάδιο πρόβλεψης χωρίζεται σε δύο μέρη. Στο πρώτο γίνεται η πρόβλεψη της καμπύλης φορτίου, με την επιλογή της ημέρας πρόβλεψης και των ιστορικών ημερών. Έπειτα μετριέται η ομοιότητα της καμπύλης φορτίου μεταξύ των ιστορικών ημερών και της προς πρόβλεψη ημέρας. Στην συνέχεια οι ημέρες ταξινομούνται βάση της ομοιότητάς. Επιλέγονται οι  $M_{max}$  ιστορικές ημέρες ως σύνολο δεδομένων  $H$  και υπολογίζεται το βέλτιστο  $M$ . Τέλος εξάγεται η πρόβλεψη της καμπύλης φορτίου. Στο δεύτερο μέρος, γίνεται η πρόβλεψη των ακραίων τιμών ζήτησης για την κάθε ημέρα. Πρώτα πραγματοποιείται μια κανονικοποίηση των δεδομένων, και έπειτα επιλέγονται τα κοινά χαρακτηριστικά και δημιουργείται το σύνολο χαρακτηριστικών της κάθε αργίας. Στη συνέχεια, γίνεται ταξινόμηση των ειδικών ημερών και επιλέγονται τα αντίστοιχα ιστορικά στοιχεία. Δημιουργείται το μοντέλο XGBOOST και εξάγονται τα αποτελέσματα πρόβλεψης των ακραίων τιμών ζήτησης. Τέλος, τα δύο αυτά αποτελέσματα πρόβλεψης συνδυάζονται και γίνονται οι κατάλληλες διορθώσεις, ώστε

να πάρουμε την τελική πρόβλεψη φορτίου. Το μοντέλο δίνει μέσο σφάλμα της τάξης του 2,98%.

Ο Zeng et al [ZENG, SHENG, and JIN, 2018] προτείνει την χρήση των Support Vector Machines (SVM) ως λύση στα προβλήματα του local optimum και των προβλημάτων ταχύτητας που ταλανίζουν τα TNN καθώς οι SVM μπορούν να αποφύγουν με ευκολία το πρόβλημα του overfitting. Στην εργασία τους χρησιμοποιείται ένα τροποποιημένο SVM που καλείται Support Vector Regression (SVR). Τα ιστορικά δεδομένα των ημερών ταξινομούνται σε τρία σύνολα. Το πρώτο, αφορά τις ημέρες των αργιών, το δεύτερο τις ημέρες πριν από τις αργίες και το τρίτο τις ημέρες μετά τις αργίες. Η ταξινόμηση αυτή είχε ως αποτέλεσμα την βελτίωση των αποτελεσμάτων της πρόβλεψης. Επίσης, χρησιμοποιήθηκε η τεχνική της πρόσθεσης ψευδών δεδομένων αργιών στο μοντέλο για την περαιτέρω βελτίωση της ακρίβειας. Παρόλα αυτά, το πρόβλημα των αραιών δεδομένων παραμένει, και για την λύση του μπορεί να χρησιμοποιηθεί η τεχνική του transfer learning. Πιο συγκεκριμένα, προτείνεται ένα πλαίσιο εκμάθησης βασισμένο σε σταθμισμένη μεταφορά γνώσης (weighted knowledge transfer). Καθώς στη συγκεκριμένη δημοσίευση χρησιμοποιήθηκαν δεδομένα διαφορετικών πόλεων, η μεταβολή των βαρών αντικατοπτρίζουν τον βαθμό της διαφοροποίησης των δεδομένων που χρησιμοποιούνται για την εκπαίδευση του δικτύου καθώς προέρχονται από διαφορετικές πόλεις από αυτή για την οποία γίνεται η πρόβλεψη. Επιπλέον, η χρήση ενός βελτιστοποιημένου αλγορίθμου TrAdaBoost φαίνεται να επιλύει το πρόβλημα της αρνητικής μεταφοράς που προκύπτει σε διαφορετικές πόλεις. Τα κριτήρια Rauta χρησιμοποιούνται για την ανάλυση των δεδομένων και η τεχνική της γραμμικής παρεμβολής χρησιμοποιήθηκε για την διόρθωση τυχών σφαλμάτων των δεδομένων. Το μέσο σφάλμα που αναφέρεται ανέρχεται στο 2,88%.

Ο Arora et al [Arora and Taylor, 2013] προτείνουν μια λύση για το πρόβλημα της πρόβλεψης φορτίου των ειδικών ημερών που βασίζεται σε έναν αριθμό κανόνων, σε συνδυασμό με την χρήση Triple seasonal HWT Exponential Smoothing το οποίο έχει προταθεί για την μοντελοποίηση της εποχικότητας ανάμεσα σε ημέρες, εβδομάδες και έτη των δεδομένων φορτίου. Αυτό το μοντέλο είναι κατάλληλο για χρήση μόνο σε ειδικές ημέρες.

Οι κανόνες που δοκιμάστηκαν είναι:

- Rule 1. Ο κανόνας χρησιμοποιεί τα ιστορικά στοιχεία φορτίου της ίδιας ημέρας, για την πρόβλεψη του φορτίου μιας δεδομένης ημέρας
- Rule 2. Κάθε ημέρα της εβδομάδας έχει ένα συγκεκριμένο αντίκτυπο στο φορτίο της ειδικής ημέρας
- Rule 3. Δεδομένου πως τα σαββατοκύριακα έχουν χαμηλότερο φορτίο σε σχέση με τις υπόλοιπες ημέρες της εβδομάδας, τα σαββατοκύριακα αντιμετωπίζονται σαν να έχουν ένα συγκεκριμένο αντίκτυπο στο φορτίο των ειδικών ημερών.
- Rule 4. Διαφορετικές ημέρες της εβδομάδας που έχουν διακριτό ενδοημερήσιο κύκλο θεωρείται ότι έχουν διακριτές επιπτώσεις στο ανώμαλο φορτίο.
- Διαφορετικές μέρες της εβδομάδας που έχουν έναν διακριτό κύκλο αντιμετωπίζονται σαν να έχουν ξεχωριστό αντίκτυπο στο φορτίο των ειδικών ημερών.

Το μοντέλο που χρησιμοποιήθηκε είναι το SARMA, και ο τρίτος κανόνας φαίνεται να έχει την καλύτερη απόδοση. Το μοντέλο των κανόνων του Arora et al εφαρμόστηκε στα δεδομένα της Γαλλίας [Arora and Taylor, 2018]. Η εποχικότητα κατά τις ημέρες, εβδομάδες και έτη εκφράζεται με την χρήση μιας τριπλής εποχικής προσαρμογής του μοντέλου SARMA (Seasonal autoregressive moving average model). Διαπιστώθηκε πως το μοντέλο χρειάζεται προσαρμογή για τα δεδομένα της Γαλλίας, η οποία πραγματοποιείται με την χρήση δεικτών. Το αποτέλεσμα είναι το μοντέλο RB-SARMA. Στη συνέχεια, γίνεται σύγκριση με έναν αριθμό μοντέλων ως προς την ακρίβεια πρόβλεψης με την χρήση ενός triple rule based Holt-Winters-Taylor exponential smoothing model και TNN. Η εφαρμογή της μεθοδολογίας στην περίπτωση της Γαλλίας δίνει μέσο σφάλμα της τάξης του 4,07%.

Οι Barman and Dev Choudhury [Barman and Dev Choudhury, 2020] δημιούργησαν ένα μοντέλο που βασίζεται στην χρήση ενός SVM και ενός Grey Wolf optimizer (GWO) με αποτέλεσμα ένα υβριδικό σύστημα πρόβλεψης το οποίο συγκρίνεται με έναν αριθμό διαφορετικών μεθόδων. Το μοντέλο αποτελείται από τρία διακριτά μέρη.

- Data pre-processing Unit
- Support Vector Machine

- Grey Wolf optimizer.

Επιπλέον, μελετώνται η εξέταση και αναγνώριση της καταναλωτικής συμπεριφοράς στις ειδικές ημέρες, καθώς και μια νέα μέθοδος ενσωμάτωσης των ευρημάτων της καταναλωτικής συμπεριφοράς και των κλιματικών δεδομένων στο μοντέλο πρόβλεψης. Ο GWO χρησιμοποιείται για την επιλογή των σωστών παραμέτρων που αφορούν το μοντέλο SVM για την πρόβλεψη φορτίου στις ειδικές ημέρες. Το μέσο σφάλμα στις τρεις περιπτώσεις εφαρμογής του μοντέλου είναι της τάξης του 1,73%.

Οι Siah Yap and Jen Yap [Siah Yap and Jen Yap, 2012] προτείνουν τη χρήση μιας Online Sequential Extreme Learning Machine (OSELM) το οποίο είναι ένα νευρωνικό δίκτυο πολλαπλών στρώσεων (multi-layer network) που έχει την δυνατότητα αυξημένης εκμάθησης χωρίς την προσθήκη νέων νευρώνων, πράγμα που το καθιστά κατάλληλο για την διαχείριση μη στατικών προβλημάτων. Επιλέγεται ένα σύστημα OSELM based multi agent system (MAS) with a weighted average (WA) strategy αντί του Bayesian Formalism για την επίλυση του προβλήματος της ταξινόμησης. Η εφαρμογή του μοντέλου γίνεται στην περίπτωση των αργιών της Μαλαισίας με σφάλμα της τάξης του 3,77%.

Ο Song et all [Song, Baek, Hong and Jang, 2005] ανέπτυξαν ένα μοντέλο ασαφούς γραμμικής παλινδρόμησης βασιζόμενο στην προσέγγιση του Tanaka όπου τόσο τα δεδομένα εισόδου όσο και τα δεδομένα εξόδου είναι ασαφείς αριθμοί. Ο ρόλος της ασαφούς παλινδρόμησης έγκειται στην βελτίωση της αποτελεσματικότητας του μοντέλου πρόβλεψης κατά τις ειδικές ημέρες. Αναλύονται τα φορτία των ημερών όσο αφορά τα μοτίβα του φορτίου, και επιλέγονται οι ημέρες με βάση την ομοιότητα των καμπυλών, με την μορφή κανόνα. Το μοντέλο εφαρμόζεται στα δεδομένα των αργιών της Κορέας, με το μέσο μέγιστο σφάλμα να είναι της τάξης του 3,57%.

Ο Mahdavi et all [Mahdavi, Menhaj and Barghinia, 2006] χρησιμοποίησε ένα Bayesian Neural Network για την πρόβλεψη των ωριαίων φορτίων των αργιών σε συνδυασμό με την χρήση μιας υβριδικής μεθόδου μάθησης Monte Carlo με στόχο την χρήση μιας απλούστερης αρχιτεκτονικής. Η μέθοδος δοκιμάστηκε στα δεδομένα φορτίου του Ιράν

των ετών 1996 έως 2004 με το μέσο σφάλμα στις σταθερές αργίες να ανέρχεται στο 1,93% και στις κινητές αργίες στο 2,59%.

Είναι φανερό η πλειάδα των μεθόδων που χρησιμοποιούνται από τους ερευνητές στην προσπάθεια τους για την βελτίωση της ακρίβειας της πρόβλεψης φορτίου των ειδικών ημερών. Η χρήση αξιόπιστων δεδομένων κρίνεται απαραίτητη για την σωστή εκτίμηση και επιβεβαίωση της λειτουργίας των μοντέλων. Στο κεφάλαιο 5 αναλύονται οι πηγές και τα δεδομένα τα οποία χρησιμοποιήθηκαν στην παρούσα εργασία.



## **Κεφάλαιο 5 Δεδομένα.**

### **5.1 Χρονοσειρές**

Η χρήση αξιόπιστων δεδομένων αποτελεί επιτακτική ανάγκη για την σωστή λειτουργία οποιουδήποτε μοντέλου πρόβλεψης. Στην παρούσα εργασία χρησιμοποιούνται δεδομένα θερμοκρασίας και ιστορικά δεδομένα φορτίου τα οποία έχουν τη μορφή χρονοσειρών. Οι χρονοσειρές αποτελούνται από διαδοχικές παρατηρήσεις της τιμής ενός φυσικού μεγέθους ως προς τον χρόνο συνήθως με σταθερό ρυθμό δειγματοληψίας. Παραδείγματα είναι η ένταση της ηλιακής ακτινοβολίας, η ταχύτητα του ανέμου, ο αριθμός των αυτοκινήτων που διασχίζουν ένα σταθμό διοδίων. Η ανάλυση των χρονοσειρών μας επιτρέπει να κάνουμε μια εκτίμηση του συστήματος που τις παράγει και εν συνεχεία μια πρόβλεψη των μελλοντικών τιμών της προς εξέτασης μεταβλητής. Στόχος του κάθε μοντέλου είναι η ακριβέστερη δυνατή πρόβλεψη της μελλοντικής τιμής της εξεταζόμενης μεταβλητής με την αξιοποίηση των ιστορικών πληροφοριών που παρέχονται.

### **5.2 Χαρακτηριστικά μεγέθη των χρονοσειρών.**

Η οπτικοποίηση των τιμών της εκάστοτε χρονοσειράς αποτελεί ένα εξαιρετικό εργαλείο για την εξέταση και την κατανόηση της εξέλιξης της τιμής και κατ' επέκταση της επιλογής της κατάλληλης μεθόδου πρόβλεψης ώστε να επιτευχθεί το βέλτιστο αποτέλεσμα. Ακόμη, καθίσταται εύκολος ο εντοπισμός τυχών ακραίων η ελλειπών τιμών. Στην Εικόνα 5.1 παρουσιάζεται η εξέλιξη του φορτίου κατά τις 6 πρώτες ημέρες του 2012. Είναι εύκολο να παρατηρήσουμε τα ποιοτικά χαρακτηριστικά της χρονοσειράς καθώς και την μεταβολή κατά τις ημέρες των αργιών. Τα βασικά χαρακτηριστικά των χρονοσειρών είναι [Shumway and Stoffer, 2011]:

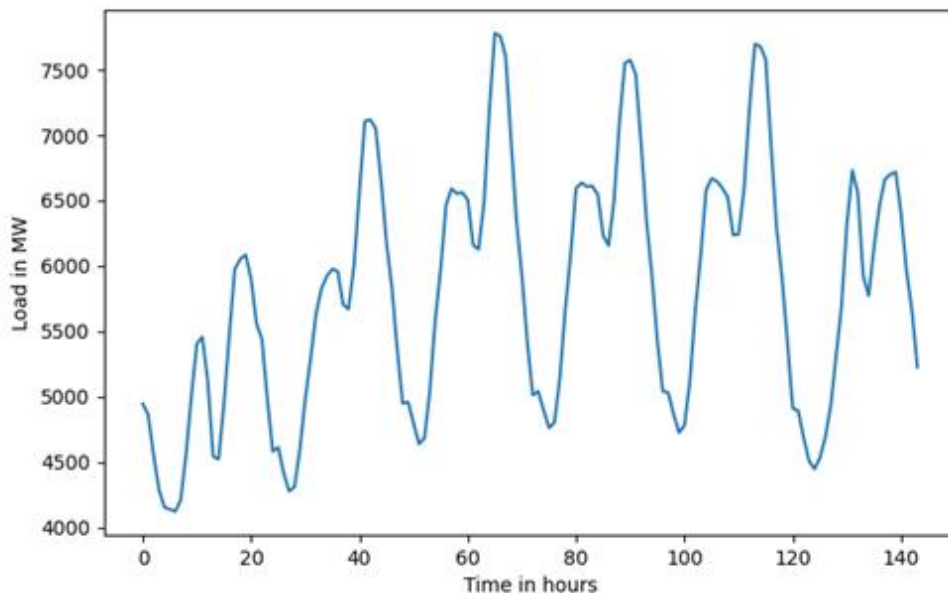
#### **A. Τάση (Trend)**

Η τάση εκφράζει την γενική τάση, είτε αυξητική είτε μειωτική, των προς εξέταση δεδομένων στον μακροπρόθεσμο ορίζοντα. Μπορεί να είναι είτε γραμμική, κάτι που

υπολογίζεται με γραμμική παρεμβολή είτε μη γραμμική όπου εισάγονται λογαριθμικά ή εκθετικά μοντέλα.

#### Β. Κυκλικότητα (Cycle)

Η κυκλικότητα εκφράζει την αυξομείωση των δεδομένων που δεν έχει σταθερή συχνότητα. Οφείλεται κυρίως σε εξωγενείς παράγοντες, και συχνά συγχέεται με την εποχικότητα. Η βασική τους διαφορά είναι η σταθερότητα των διακυμάνσεων αυτών. Σε περίπτωση που η συχνότητά τους δεν είναι σταθερή, πρόκειται για κυκλικότητα, ενώ στην αντίθετη περίπτωση πρόκειται για εποχικότητα.



Εικόνα 5.1 Φορτίο από 1/1 έως 6/1 2012

#### Γ. Εποχικότητα(Seasonality)

Η εποχικότητα εκφράζει την μεταβολή των τιμών της χρονοσειράς επηρεαζόμενη από εποχικούς παράγοντες, όπως η εποχή, η ημέρα της εβδομάδος. Η συχνότητά της είναι σταθερή, κάτι που την διαφοροποιεί επαρκώς από την κυκλικότητα.

#### Δ. Τυχασιότητα (Randomness)

Εκφράζει τις τυχαίες διακυμάνσεις στην εξεταζόμενη χρονοσειρά οι οποίες απομένουν όταν απομονωθούν τα υπόλοιπα χαρακτηριστικά της χρονοσειράς. Προκύπτει είτε από ένα τυχαίο και απρόβλεπτο γεγονός, είτε από κάποιο σφάλμα μέτρησης.



#### Ε. Ασυνέχειες(Outliers)

Πρόκειται για μεταβολές των τιμών της χρονοσειράς που δεν μπορούν να εξηγηθούν. Ο χαρακτήρας τους είναι προσωρινός, και παρότι η επίδρασή τους είναι περιορισμένη, το γεγονός ότι δεν συνάδουν με την υπόλοιπη σειρά μπορεί να οδηγήσει σε αύξηση του σφάλματος κατά την διαδικασία της πρόβλεψης.

#### Στ. Στασιμότητα (Stationarity)

Η στασιμότητα ορίζεται ως η σταθερότητα των στατιστικών ιδιοτήτων της χρονοσειράς. Οι στάσιμες χρονοσειρές, παρότι πρακτικά είναι πολύ σπάνιες, είναι σχετικά πιο εύκολες στην πρόβλεψή τους, καθότι οι στατιστικές τους ιδιότητες είναι σταθερές στο μέλλον.

#### Ζ. Γραμμικότητα.

Ως γραμμικές χαρακτηρίζονται οι χρονοσειρές οι μεταβλητές των οποίων αλληλοεπιδρούν γραμμικά.

### 5.3 Πηγές δεδομένων

#### Α. Ιστορικές τιμές φορτίου συστήματος

Η έρευνα σχετικά με τα δεδομένα φορτίου τα οποία χρησιμοποιήθηκαν στην παρούσα εργασία κατέδειξαν τρεις διαφορετικές πηγές δεδομένων, τον ελληνικό διαχειριστή δικτύου (ΑΔΜΗΕ) [Δεδομένα ΑΔΜΗΕ, 2022], το ευρωπαϊκό δίκτυο διαχειριστών συστημάτων μεταφοράς (ENTSO-E) και τέλος την ιστοσελίδα open-power-system-data.org [Open Power System Data – A platform for open data of the European power system, 2022]. Η αρχική πηγή των δεδομένων, παραμένει ο ΑΔΜΗΕ, ο οποίος με τη σειρά του αποστέλλει δεδομένα στον ENTSO-E. Τέλος, η ιστοσελίδα open-power-system-data.org αναφέρει ως πηγή των χρονοσειρών που παρέχει τον ENTSO-E [Total Load - Day Ahead / Actual, 2022]. Το πρόβλημα που προέκυψε σχετικά με την χρήση δεδομένων του ΑΔΜΗΕ αφορά στην ύπαρξη πολλών κενών στις χρονοσειρές που παρέχει από την ιστοσελίδα του. Συγκεκριμένα, πέρα από την έλλειψη μιας η δύο συνεχόμενων παρατηρήσεων οι οποίες μπορούν να συμπληρωθούν προσεγγιστικά, διαπιστώθηκε η

απουσία καταγραφών ολόκληρων ημερών, κάτι που καθιστά την χρήση των παρεχόμενων δεδομένων αδύνατη.

Τα δεδομένα που παρείχε η ιστοσελίδα [open-power-system-data.org](http://open-power-system-data.org) δεν είχαν κενά μεταξύ των παρατηρήσεων, όχι σε βαθμό που δεν μπορούσαν να συμπληρωθούν, αλλά παρατηρήθηκε ασυμφωνία μεταξύ των παρεχόμενων δεδομένων και αυτών που αναφέρονται ως πηγή, δηλαδή του ENTSO-E. Όσο αφορά τα δεδομένα του ENTSO-E, η σύγκρισή τους με τα αντίστοιχα του ΑΔΜΗΕ που θεωρούνται ως πηγή τους καταδεικνύει διαφοροποιήσεις στις τιμές, οι οποίες θεωρητικά δεν θα έπρεπε να υπάρχουν. Συνοψίζοντας, βρέθηκαν τρεις διαφορετικές πηγές δεδομένων τα δεδομένα των οποίων δεν συμφωνούσαν μεταξύ τους. Κρίθηκε σωστό να επιλεγεί η χρήση των δεδομένων του ENTSO-E καθώς είναι τα μόνα που παρείχαν σχεδόν πλήρεις χρονοσειρές. Αποτελεί πεποίθηση του γράφοντα πως η αναγκαιότητα της δημιουργίας μιας βάσης δεδομένων για το ελληνικό σύστημα ενέργειας αποτελεί επιτακτική ανάγκη, καθώς η ύπαρξή της θα αποτελούσε σημαντική διευκόλυνση για μελλοντική έρευνα στον τομέα της πρόβλεψης φορτίου, δεδομένης της εύκολης πρόσβασης σε επικαιροποιημένα και ακριβή στοιχεία.

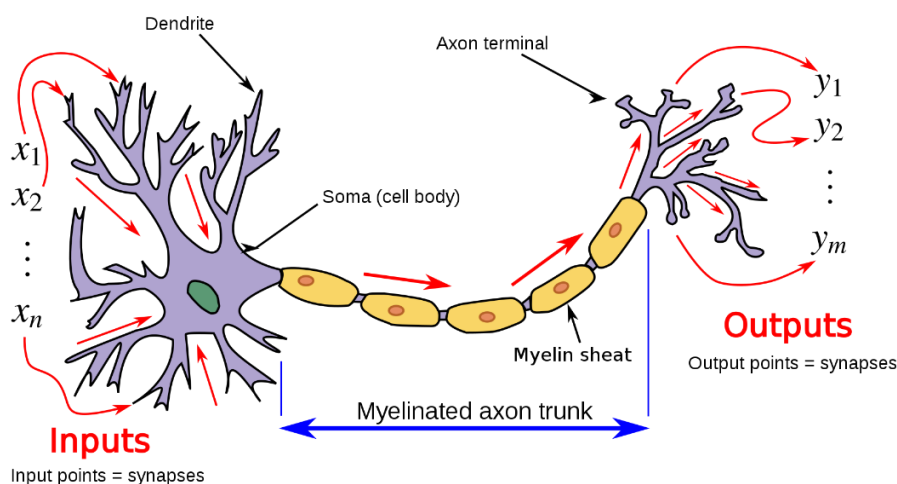
#### B. Ιστορικές τιμές θερμοκρασίας.

Όσο αφορά τα δεδομένα θερμοκρασίας χρησιμοποιήθηκε η χρονοσειρά που παρέχεται από την ιστοσελίδα [open-power-system-data.org](http://open-power-system-data.org). Η αρχική πηγή των δεδομένων αναφέρεται ως το πρόγραμμα MERRA-2 της NASA, και όσο αφορά την Ε.Ε. η ανάλυση πραγματοποιείται από την ιστοσελίδα [renewables.ninja](http://renewables.ninja) [Renewables.ninja, 2022].

## Κεφάλαιο 6 Τεχνητά νευρωνικά δίκτυα, επιλογή χαρακτηριστικών, βελτιστοποίηση.

### 6.1 Τεχνητά νευρωνικά δίκτυα

Τα τεχνητά νευρωνικά δίκτυα (ΤΝΔ) είναι ένα υπολογιστικό σύστημα δημιουργημένο κατ' εικόνα των βιολογικών νευρωνικών δικτύων, μιμούμενο την λειτουργία τους και χρησιμοποιούμενο στη αναγνώριση μοτίβων και τη λύση προβλημάτων στους τομείς της τεχνητής νοημοσύνης, της μηχανικής εκμάθησης και της βαθιάς εκμάθησης [Education, 2022]. Δομικό χαρακτηριστικό των ΤΝΔ είναι ο τεχνητός νευρώνας ο οποίος είναι μια μαθηματική συνάρτηση, η οποία δέχεται weighted δεδομένα εισόδου, τα οποία προσθέτει και με την χρήση μιας συνάρτησης ενεργοποίησης παράγει μια έξοδο. Οι συναρτήσεις είναι συνήθως μη γραμμικές. Η Εικόνα 6.1 απεικονίζει την αντιστοιχία των λειτουργιών μεταξύ των βιολογικών και των τεχνητών νευρώνων.



Εικόνα 6.1 Αντιστοιχία τεχνητών και φυσικών νευρώνων, [Wikimedia 2022]

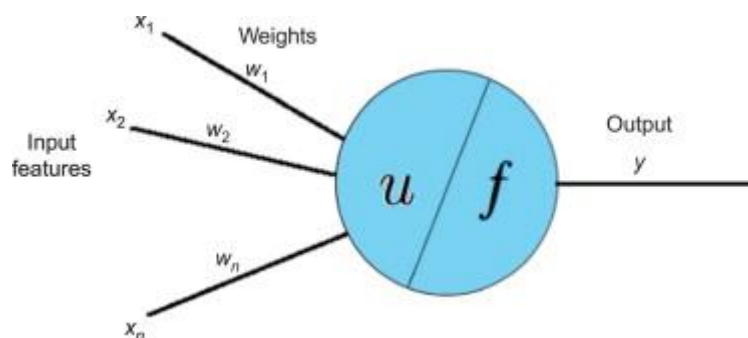
Ένα νευρωνικό δίκτυο είναι ένα σύνολο νευρώνων που επικοινωνούν μεταξύ τους. Οι συνδέσεις, ή κόμβοι, μεταξύ των νευρώνων χαρακτηρίζονται από συγκεκριμένους συντελεστές βαρύτητας, ή βάρη, τα οποία διαμορφώνονται κατά την διαδικασία της εκπαίδευσης. Έχει αναπτυχθεί μεγάλος αριθμός ΤΝΔ ανάλογα με τον τομέα εφαρμογής.

## 6.2 Perceptron

Ο Perceptron ή αντίληπτρο στην Ελληνική βιβλιογραφία, είναι ένας γραμμικός ταξινομητής, δημιουργημένος από τον F. Rosenblatt. [Rosenblatt, 1958]. Στην ουσία πρόκειται για μια συνάρτηση που απεικονίζει την τιμή εισόδου  $x$  σε μια τιμή εξόδου  $f(x)$  ή οποία λαμβάνει δυαδική τιμή. Μαθηματικά απεικονίζεται από την σχέση 6.1 όπου  $w$  είναι η τιμή του βάρους (weight) και  $b$  ένας σταθερός όρος μη εξαρτώμενος από τις τιμές εισόδου.

$$f(x) = \begin{cases} 1 & \text{if } w * x + b > 0 \\ 0 & \text{if } w * x + b \leq 0 \end{cases} \quad (6.1)$$

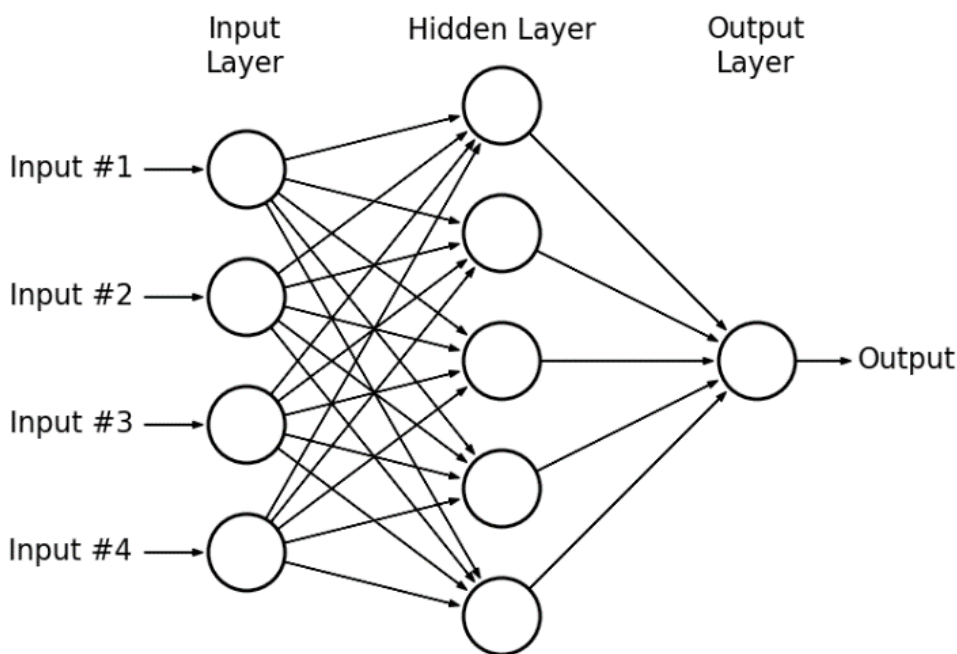
Ο perceptron αποτελείται από τέσσερα μέρη. Τις τιμές εισόδου ή τον input layer που αποτελείται από τα δεδομένα εισόδου. Τα βάρη και το bias, τα πρώτα δημιουργούνται κατά την διαδικασία της εκπαίδευσης του, και το bias που όπως προαναφέρθηκε αποτελεί σταθερό όρο μη σχετιζόμενο με τα δεδομένα εισόδου. Το σύνολο των βαρών και των bias αποτελεί το weighted sum που περνά από την συνάρτηση ενεργοποίησης. Η λειτουργία του χρησιμοποιεί τα τέσσερα μέρη που προαναφέρθηκαν. Αρχικά, όλα τα δεδομένα εισόδου  $x_1, x_2, \dots, x_n$  πολλαπλασιάζονται με τα βάρη και το bias, το αποτέλεσμα των οποίων έστω πως είναι  $k$ . Στη συνέχεια τα  $k$  προστίθενται δημιουργώντας το weighted sum. Τέλος, χρησιμοποιούμε την συνάρτηση ενεργοποίησης για την παραγωγή της εξόδου του συστήματος. Η μορφή του perceptron απεικονίζεται στην Εικόνα 6.2.



Εικόνα 6.2 Perceptron [Sharing Data and Models in Software Engineering, 2015]

### 6.3 Multilayer Perceptron

Ο multilayer perceptron (MLP) είναι ένα feed forward τεχνητό νευρωνικό δίκτυο. Αποτελείται από τρία στρώματα, το στρώμα εισόδου, το κρυφό στρώμα και το στρώμα εξόδου. Το στρώμα εισόδου αποτελεί την είσοδο των προς επεξεργασία δεδομένων. Το κρυφό στρώμα αποτελείται από ένα ή περισσότερα στρώματα στα οποία αποτελούν το υπολογιστικό κομμάτι του δικτύου. Το στρώμα εξόδου δέχεται τις εξόδους του μοντέλου. Η ροή των δεδομένων λαμβάνει χώρα από το στρώμα εισόδου προς το στρώμα εξόδου [Abirami and Chitra, 2020]. Η μορφή του απεικονίζεται στην Εικόνα 6.3.



Εικόνα 6.3 Multilayer Perceptron ANN [Hassan et all, 2015]

Κατά την εκπαίδευση χρησιμοποιείται ο αλγόριθμος back propagation η χρήση του οποίου συνίσταται στον υπολογισμό της ελάχιστης τιμής της συνάρτησης σφάλματος μέσα στο χώρο των βαρών χρησιμοποιώντας τον αλγόριθμο gradient descent ο οποίος είναι ένας αλγόριθμος για τον υπολογισμό των τοπικών ελαχίστων μιας διαφορικής συνάρτησης.

### 6.4 Extreme gradient boosting (XGBoost)

Το XGBoost αποτελεί μια βιβλιοθήκη ανοιχτού λογισμικού δημιουργημένη από τον Chen and Guestrin [Chen and Guestrin, 2016] για την υλοποίηση μοντέλων tree gradient

boosting σε εφαρμογές μηχανικής εκμάθησης. Υλοποιεί χαρακτηριστικά όπως ο έξυπνος τρόπος εισαγωγής ποινής στα δένδρα, η αυτόματη επιλογή χαρακτηριστικών, οι περισσότερες δυνατότητες για τυχαιοποίηση κ.α. τα οποία έχουν ως αποτέλεσμα την αύξηση της επιλογής της για την επίλυση προβλημάτων [Chen and Guestrin, 2016]. Ο αλγόριθμος βρίσκει εφαρμογή τόσο σε προβλήματα κατηγοριοποίησης όσο και σε προβλήματα παλινδρόμησης.

Η λειτουργία του XGBoost δεδομένου του συνόλου εκπαίδευσης  $(x_i, y_i)_{i=1}^n$ , μιας συνάρτησης κόστους  $L(y, F(x))$ , του ρυθμού εκμάθησης  $\alpha$  και του αριθμού μοντέλων  $M$  πραγματοποιείται ως εξής. Το μοντέλο αρχικοποιείται βάση μιας σταθεράς της σχέσης 6.2.

$$f_0(x) = \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^n L(y_i, \theta) \quad (6.2)$$

Στη συνέχεια γίνεται ο υπολογισμός των gradients και των Hessians βάση των σχέσεων 6.3 και 6.4 για  $i = 1$  to  $N$ .

$$g_m(x_i) = \left[ \frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)} \quad (6.3)$$

$$h_m(x_i) = \left[ \frac{\partial^2 L(y_i, f(x_i))}{\partial f(x_i)^2} \right]_{f(x)=f_{m-1}(x)} \quad (6.4)$$

Έπειτα κατασκευάζεται το ασθενές μοντέλο  $f_m(x)$  της σχέσης 6.5, εκπαιδεύεται με το σύνολο δεδομένων εκπαίδευσης, υπολογίζοντας το πρόβλημα βελτιστοποίησης της σχέσης 6.6 και ανανεώνεται βάση της σχέσης 6.7

$$\hat{f}_m(x) = \alpha \hat{\varphi}_m(x) \quad (6.5)$$

$$\varphi_m = \underset{\varphi \in \Phi}{\operatorname{argmin}} \sum_{i=1}^n \frac{1}{2} \widehat{h}_m(x_i) \left[ -\frac{\hat{g}_m(x_i)}{\hat{h}_m(x_i)} - \varphi(x_i) \right]^2 \quad (6.6)$$

$$\hat{f}_m(x) = f_{m-1}(x) + \hat{\varphi}_m(x) \quad (6.7)$$

Τέλος, επιστρέφεται το τελικό μοντέλο  $\hat{f}_M(x) = \sum_{m=0}^M \hat{f}_m(x)$

## 6.5 Support Vector Machines

Οι Support Vector machines (SVM) είναι μοντέλα επιτηρούμενης εκμάθησης δημιουργημένα από τον Vapnik [Cortes and Vapnik, 1995]. Βασίζονται στην θεωρία των πλαισίων στατιστικής μάθησης, μια ιδέα που αναπτύχθηκε από τους Vapnik και Chervonenkis, πράγμα που καθιστά τα SVM μια από τις πιο ισχυρές μεθόδους πρόβλεψης και βρίσκουν εφαρμογή τόσο σε προβλήματα κατηγοριοποίησης όσο και σε προβλήματα παλινδρόμησης.

Η λειτουργία τους συνίσταται στην κατασκευή ενός υπερεπιπέδου (hyperplane) στον πολυδιάστατο χώρο για τον διαχωρισμό των διαφορετικών κλάσεων κατά την λειτουργία της κατηγοριοποίησης, το οποίο χρησιμοποιείται για τον διαχωρισμό των δεδομένων με το μέγιστο δυνατό περιθώριο.

Κατά την παλινδρόμηση χρησιμοποιούνται οι ίδιες αρχές με την ταξινόμηση, δηλαδή η μείωση του σφάλματος με την χρήση του υπερεπιπέδου το οποίο μεγιστοποιεί το περιθώριο, με την παραδοχή πως ένα ποσοστό σφάλματος είναι αποδεκτό. Εκφράζουμε την γραμμική παλινδρόμηση κατά την SVR με την σχέση 6.8 όπου  $x$  είναι ένα σετ ανεξάρτητων μεταβλητών,  $w$  είναι το διάνυσμα του weight,  $b$  είναι το offset και  $e$  το αποδεκτό σφάλμα.

$$f(x) = w^T x + b + e \quad (6.8)$$

Κατά την χρήση του SVR η προσέγγιση της συνάρτησης  $f$  πραγματοποιείται με την ορισμό ενός  $e$ -insensitive tube όσο το δυνατόν πιο επίπεδου, με την ελαχιστοποίηση του  $w$ . Αυτό εκφράζεται στην σχέση 6.9

$$\min_w \frac{1}{2} \|w\|^2 \quad (6.9)$$

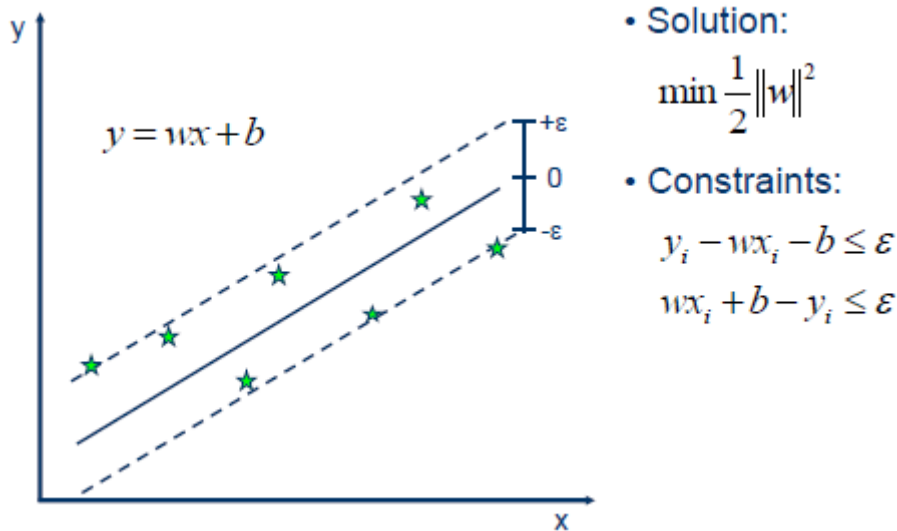
Υποκείμενο στους περιορισμούς των σχέσεων 6.10 και 6.11

$$y_1 - w^T x_i - b \leq e \quad (6.10)$$

$$w^T x_i + b - y_i \leq e \quad (6.11)$$

Συμπεραίνεται πως η ουσία του SVR είναι η πραγματοποίηση της γραμμικής παλινδρόμησης με την χρήση μιας συνάρτησης κόστους που τιμωρεί τις προβλέψεις που απέχουν περισσότερο από  $\epsilon$  από το επιθυμητό αποτέλεσμα, κάτι που φαίνεται καλύτερα στην Εικόνα 6.4. Στην σχέση 6.12 παρουσιάζεται η γραμμική συνάρτηση κόστους.

$$L(y, f(x)) = \begin{cases} 0 & \text{if } |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon & \text{otherwise} \end{cases} \quad (6.12)$$



Εικόνα 6.4 regression, tube and functions

## 6.6 Boruta-SHAP feature selection

Η ανάγκη της επιλογής των πιο κατάλληλων χαρακτηριστικών ενός μοντέλου μηχανικής εκμάθησης οδήγησε στην ανάπτυξη αλγορίθμων οι οποίοι είναι σε θέση να φέρουν εις πέρας την διαδικασία αυτή. Η επιλογή του αλγορίθμου Boruta-SHAP [Pfeifer et al., 2022] πραγματοποιήθηκε λόγω της απόδοσης του. Ο αλγόριθμος χωρίζεται σε δύο μέρη.

Το πρώτο μέρος, ο αλγόριθμος Boruta [Kursa, Jankowski and Rudnicki, 2010] αρχικά αξιολογεί την βαρύτητα των προς εξέταση χαρακτηριστικών, όχι με την σύγκριση μεταξύ τους όπως γίνεται στους υπόλοιπους αλγορίθμους αλλά με την σύγκριση με τυχαιοποιημένες εκδόσεις που αποκαλούνται shadow features. Η βασική ιδέα πίσω από την δημιουργία του είναι η επιλογή χαρακτηριστικών που αποδίδουν καλύτερα από τις



τυχειότητα των shadow features, οπότε συγκρίνεται η βαρύτητα των χαρακτηριστικών με την υψηλότερη τιμή των shadow features. Για την αποφυγή της τυχειότητας στην επιλογή των πιο κατάλληλων χαρακτηριστικών πραγματοποιείται ένας αριθμός επαναλήψεων, και η επιλογή τους πραγματοποιείται με τη χρήση binomial distribution.

Το SHAP (SHapley Additive exPlanations) [An introduction to explainable AI with Shapley values — SHAP latest documentation, 2022] είναι μια προσέγγιση η οποία χρησιμοποιεί τη θεωρία των παιγνίων για την ερμηνεία των εξόδων οποιουδήποτε μοντέλου μηχανικής εκμάθησης αποτελεί το δεύτερο κομμάτι του αλγορίθμου. Η χρήση της έγκειται στην αποφυγή της αδυναμίας του Boruta στην επιλογή των χαρακτηριστικών που βασίζεται στον υπολογισμό της βαρύτητας των χαρακτηριστικών, καθώς υπάρχει πιθανότητα να είναι μεροληπτική ή όχι αρκετά καλή, βάση των δεδομένων.

## 6.7 Bayesian Optimization

Τα μοντέλα της μηχανικής εκμάθησης περιέχουν ένα μεγάλο αριθμό μεταβλητών που επηρεάζουν την συμπεριφορά και την απόδοσή τους. Καθώς η εύρεση των βέλτιστων τιμών των μεταβλητών αυτών έχει ως αποτέλεσμα την περεταίρω βελτίωση της απόδοσης του μοντέλου, κρίθηκε θεμιτή η υλοποίηση της στο μοντέλο που πραγματεύεται η παρούσα εργασία. Ο αριθμός των μεταβλητών είναι αρκετά μεγάλος οπότε η λύση της βελτιστοποίησης του με το χέρι αποδεικνύεται χρονοβόρα. Για τους λόγους αυτούς χρησιμοποιήθηκε ο αλγόριθμος βελτιστοποίησης BayesSearchCV [skopt.BayesSearchCV — scikit-optimize 0.8.1 documentation, 2022] για την βελτιστοποίηση των παραμέτρων του μοντέλου.

Η λειτουργία του αλγορίθμου συνοψίζεται στη δημιουργία ενός μοντέλου πιθανοτήτων της αντικειμενικής συνάρτησης και στη χρήση του για την επιλογή των παραμέτρων για την αξιολόγηση στην πραγματική αντικειμενική συνάρτηση [Frazier, 2018]. Χρησιμοποιείται το θεώρημα του Bayes το οποίο εκφράζεται από την σχέση 6.13 όπου A και B είναι γεγονότα,  $P(A)$  και  $P(B)$  οι πιθανότητες του A και B που είναι ανεξάρτητα

μεταξύ τους,  $P(A|B)$  η υπό συνθήκη πιθανότητα του A να συμβεί δεδομένου πως το B είναι αληθές και  $P(B|A)$  η πιθανότητα του B να συμβεί εφόσον το A είναι αληθές.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (6.13)$$

Επιπλέον του μοντέλου πιθανοτήτων (surrogate model) χρησιμοποιείται για την προσέγγιση της αντικειμενικής συνάρτησης μια συνάρτηση απόκτησης (acquisition function) η οποία κατευθύνει την επιλογή των δειγμάτων στις περιοχές με τη μεγαλύτερη πιθανότητα για την βελτίωση. Όσο αφορά το surrogate model, μια δημοφιλής επιλογή είναι οι Gaussian Processes(GP).

Η επιλογή των σημείων δειγματοληψίας γίνεται με την χρήση των συναρτήσεων απόκτησης οι οποίες αντισταθμίζουν την εκμετάλλευση και εξερεύνηση, όπου ως εκμετάλλευση ορίζεται η λήψη δειγμάτων από ένα σημείο που προτείνεται από το surrogate model και εξερεύνηση η δειγματοληψία από σημεία όπου η αβεβαιότητα πρόβλεψης είναι υψηλή. Έστω μια συνάρτηση  $f(x)$ . Το σημείο δειγματοληψίας ορίζεται από την σχέση 6.14 όπου  $u$  η συνάρτηση απόκτησης και  $D_{1:t-1} = (x_1, y_1) \dots (x_{t-1}, y_{t-1})$  τα δείγματα που έχουν επιλεγεί από την συνάρτηση  $f(x)$  ως τώρα.

$$x_t = \operatorname{argmax}_x u(x|D_{1:t-1}) \quad (6.14)$$

Ο αλγόριθμος βελτιστοποίησης λειτουργεί ακολουθώντας τα βήματα. Για  $t = 1, 2, \dots, n$ ,

- Εύρεση του επομένου σημείου δειγματοληψίας  $x_t$  βελτιστοποιώντας την συνάρτηση απόκτησης 6.13 πάνω στο GP.
- Απόκτηση ενός πιθανού δείγματος με θόρυβο  $y_t = f(x) + e_t$  από την αντικειμενική συνάρτηση  $f$
- Πρόσθεση του δείγματος στα προηγούμενα δείγματα  $D_{1:t} = D_{1:t-1}, (x_t, y_t)$  και ενημέρωση της GP.

## Κεφάλαιο 7 Υλοποίηση

### 7.1 Πρώτη Προσέγγιση

Η προσέγγιση που ακολουθήθηκε στην παρούσα εργασία είχε ως αποτέλεσμα την επιλογή των τεχνητών νευρωνικών δικτύων (ΤΝΔ) για την πρόβλεψη του φορτίου της επόμενης ημέρας. Αυτό πηγάζει από την ικανότητα τους στην εύρεση και εκμετάλλευση μοτίβων ακόμη και στην δική μας περίπτωση που τα δεδομένα είναι πολύ λίγα. Επιπλέον θέλαμε να εξετάσουμε το κατά πόσο είναι δυνατή η βελτίωση των αποτελεσμάτων του εκάστοτε μοντέλου ΤΝΔ με την αναζήτηση των βέλτιστων παραμέτρων του. Αυτό πραγματοποιείται με την υλοποίηση του αλγορίθμου αναζήτησης Bayes Search CV. Ακόμη, αναλύθηκαν τα χαρακτηριστικά εισόδου των μοντέλων στα διάφορα σενάρια που επιλέχθηκαν με τον αλγόριθμο Boruta Shar. Τέλος, η επιλογή των όμοιων ημερών έγινε την χρήση της Discreet Fréchet Distance.

Η υλοποίηση όλων των παραπάνω πραγματοποιήθηκε με την χρήση της γλώσσας προγραμματισμού python. Η επιλογή της έγκειται στην ευκολία της, την απόδοση της και την ύπαρξη μιας μεγάλης κοινότητας με αρκετά παραδείγματα και εξαιρετική τεκμηρίωση για το κάθε κομμάτι λογισμικού που χρησιμοποιήθηκε.

Το IDE που επιλέχθηκε ήταν το Pycharm, δημιουργημένο από την JetBrains [Company - JetBrains, 2022]. Διανέμεται σε δύο εκδόσεις, την community που χρησιμοποιήθηκε στην παρούσα εργασία και είναι διαθέσιμο χωρίς χρέωση, και την professional η οποία έχει περισσότερα χαρακτηριστικά αλλά η χρήση της απαιτεί αντίτιμο. Η χρήση του Pycharm έγινε λόγω της ευκολίας που προσφέρει το περιβάλλον του στην ανάπτυξη λογισμικού, την άμεση διασύνδεση του με το GitHub όπου αποθηκεύτηκε ο κώδικας που χρησιμοποιήθηκε και την δυνατότητα της άμεσης εγκατάστασης του απαραίτητου λογισμικού για την υλοποίηση των εκάστοτε χαρακτηριστικών με την μεθόδων των πακέτων, όπως ακριβώς γίνεται στις διανομές Linux.

Συνοπτικά, δημιουργήθηκαν δύο διαφορετικές υλοποιήσεις κατά τις δοκιμές που πραγματοποιήθηκαν. Και στις δύο υλοποιήσεις η κάθε αργία αντιμετωπίστηκε σαν ξεχωριστή περίπτωση, και για κάθε μια εκπαιδεύτηκε και παρήγαγε προβλέψεις ένα διαφορετικό ΤΝΔ.

Όσο αφορά τις δύο υλοποιήσεις, η πρώτη εξετάζει τις επιπτώσεις των διαφορετικών εισόδων στα αποτελέσματα του μοντέλου, με την δημιουργία τεσσάρων διαφορετικών σεναρίων εισόδου. Δημιουργήθηκε ένα πρόγραμμα με τέτοιο τρόπο ώστε να είναι εύκολη η αλλαγή των χαρακτηριστικών εισόδου χωρίς να είναι αναγκαία η χρονοβόρα τροποποίηση του κώδικα. Η δομή του προγράμματος στηρίζεται στην λογική των modules και αποτελείται από τα εξής στοιχεία:

#### A. dataInput

Εισάγονται τα δεδομένα θερμοκρασίας και υγρασίας για τα έτη 2008 έως 2019, και μετατρέπονται σε κατάλληλη μορφή ώστε να είναι εύκολα διαχειρίσιμα από την ρυθση. Επιπλέον, τα δεδομένα χωρίζονται ανά ημέρα και χρησιμοποιείται ως header η ημερομηνία της εκάστοτε ημέρας. Δηλαδή, στην περίπτωση που ζητηθούν τα δεδομένα θερμοκρασίας της 1/1/2009 τυπώνονται με την χρήση της εντολής:

```
print(load2009['2009-01-01'])
```

#### B. TransformAndCreationFunctions

Εδώ γίνεται η μετατροπή των δεδομένων στην απαραίτητη μορφή για την δημιουργία των σετ δεδομένων, και ορίζονται οι μορφές τους για την κάθε περίπτωση.

#### Γ. PredictionEngine

Σε αυτό το κομμάτι πραγματοποιείται η δημιουργία των σετ εισόδων, η κανονικοποίηση των δεδομένων, και τέλος η υλοποίηση των μοντέλων πρόβλεψης. Στη συνέχεια υπολογίζεται το σφάλμα και δημιουργείται το διάγραμμα που συγκρίνει την πρόβλεψη της επόμενης ημέρας του κάθε μοντέλου με την πραγματική τιμή.

#### Δ. Prediction.

Σε αυτό το κομμάτι υλοποιείται η κάθε αργία ξεχωριστά, και δίνεται η δυνατότητα να επιλεγθεί η κάθε περίπτωση για την κάθε αργία.

Αφού επιλεγθεί η υποπερίπτωση που είναι επιθυμητή, όλη η προσομοίωση εκτελείται από το αρχείο main. Η έξοδος που δίνεται είναι το σφάλμα της πρόβλεψης για την κάθε αργία και με τα τρία μοντέλα που επιλέχθηκαν καθώς και το διάγραμμα τους πρόβλεψής τους, εφόσον το επιθυμούμε.

Η δεύτερη υλοποίηση είναι αρκετά πιο απλή από την πρώτη, καθώς δεν υπάρχει η ανάγκη για την δοκιμή πολλαπλών σεναρίων εισόδου. Η πρόβλεψη της κάθε ειδικής ημέρας πραγματοποιείται από ένα ξεχωριστό αρχείο. Η δομή του προγράμματος είναι ως εξής :

#### A. Functions

##### 1. dataInput.

Όπως και στην προηγούμενη υλοποίηση πραγματοποιείται η είσοδος των δεδομένων θερμοκρασίας και φορτίου καθώς και ο διαχωρισμός τους ανά ημέρα.

##### 2. Discrete\_Frechet\_distance.

Εδώ υπολογίζεται η διακριτή απόσταση Fréchet για την εύρεση των όμοιων ημερών.

##### 3. mlpOptimization.

Στο μέρος του προγράμματος αυτό υλοποιείται η βελτιστοποίηση των χαρακτηριστικών του MLP.

##### 4. SVMoptimization.

Στο μέρος του προγράμματος αυτό υλοποιείται η βελτιστοποίηση των χαρακτηριστικών του SVM.

##### 5. XGBoostOptimization

Στο μέρος του προγράμματος αυτό υλοποιείται η βελτιστοποίηση των χαρακτηριστικών του XGBoost.

## B. Argies

Εδώ γίνεται η υλοποίηση της πρόβλεψης των αργιών, για κάθε ημέρα ξεχωριστά. Πραγματοποιείται η πρόβλεψη φορτίου με την χρήση των τριών ΤΝΔ, υπολογίζεται το σφάλμα της κάθε πρόβλεψης και τέλος δημιουργείται το διάγραμμα για κάθε πρόβλεψη και της πραγματικής τιμής.

Τέλος, για την αξιολόγηση των χαρακτηριστικών χρησιμοποιείται ο αλγόριθμος Boruta-Shap. Δημιουργήθηκε μια ξεχωριστή υλοποίηση για την ανάλυση των χαρακτηριστικών με την ίδια μεθοδολογία στη δόμηση του προγράμματος.

Η πρώτη υλοποίηση αφορούσε την δημιουργία ενός μοντέλου που βασίζεται στην χρήση ενός MLP, ενός XGBoost και ενός SVM ως μοντέλα πρόβλεψης. Χρησιμοποιήθηκε η βιβλιοθήκη scikit τόσο για το μοντέλο όσο και για την κανονικοποίηση των δεδομένων. Παρέχονται δύο επιλογές όσο αφορά την κανονικοποίηση, η χρήση της συνάρτησης MinMax Scaler και του Standard Scaler.

Η λειτουργία του MinMax Scaler συνοψίζεται στην μετατροπή του κάθε χαρακτηριστικού του σετ δεδομένων στον χώρο μεταξύ 0 και 1. Σε αντίθεση, ο Standard Scaler μετατρέπει τα χαρακτηριστικά του σετ δεδομένων αφαιρώντας τον μέσο και διαιρώντας με την τυπική απόκλιση.

Οι ειδικές ημέρες για τις οποίες υλοποιήθηκε το μοντέλο παρουσιάζονται στον Πίνακα 7.1. Για λόγους σύγκρισης οι παράμετροι των μοντέλων σε πρώτη φάση παραμένουν οι ίδιες για όλα τα μοντέλα. Όσο αφορά τα δεδομένα, χρησιμοποιήθηκαν τα έτη 2019, 2018, 2017 ως test sets και τα έτη 2016, 2015, 2014 ως training sets. Εξαίρεση αποτελεί η ειδική ημέρα της κοίμησης της Θεοτόκου του 2016 κατά την οποία χρησιμοποιήθηκαν δεδομένα του 2015, 2014, 2013 για την εκπαίδευση, καθώς του 2016 ήταν ελλιπή σε βαθμό που δεν ήταν δυνατή η συμπλήρωσή τους.

Πίνακας 7.1 Ειδικές Ημέρες

A/A	Ειδική Ημέρα
1	Πρωτοχρονιά
2	Θεοφάνια
3	Καθαρά Δευτέρα
4	Επέτειος της Επανάστασης
5	Μεγάλη Παρασκευή
6	Κυριακή του Πάσχα
7	Δευτέρα του Πάσχα
8	Εργατική Πρωτομαγιά
9	Αγίου Πνεύματος
10	Κοίμηση της Θεοτόκου
11	Επέτειος του Όχι
12	Χριστούγεννα
13	Δεύτερη Ημέρα των Χριστουγέννων

Για την σύγκριση των αποτελεσμάτων των μοντέλων χρησιμοποιήθηκε ο δείκτης MAPE. Ο MAPE (Mean Absolute Percentage Error) εκφράζει το μέσο απόλυτο ποσοστιαίο σφάλμα και υπολογίζεται με την χρήση της βιβλιοθήκης scikit metrics mean absolute percentage error βάσει της σχέσης 7.1 [3.3. Metrics and scoring: quantifying the quality of predictions, 2022].

$$MAPE(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} \frac{|y_i - \hat{y}_i|}{\max(\epsilon, |y_i|)} \quad (7.1)$$

Όπου  $\epsilon$  είναι ένας πολύ μικρός θετικός αριθμός για την αποφυγή διαίρεσης με το 0. Για τις εισόδους του μοντέλου δημιουργήθηκαν τέσσερις διαφορετικές υποπεριπτώσεις οι οποίες αναλύονται στη συνέχεια.

Στην πρώτη υποπερίπτωση οι είσοδοι είναι:

- Τιμές φορτίου προηγούμενης ημέρας (notation : D-1)
- Τιμές φορτίου ίδιας ημέρας την προηγούμενη εβδομάδα (notation : D- 7)

- Τιμές φορτίου ίδιας ημέρας το προηγούμενο έτος (notation : D – 365)

Για τις κινητές αργίες, όσο αφορά την τιμή D-365 χρησιμοποιήθηκε η αντίστοιχη ημέρα αργίας του προηγούμενου έτους. Εκτελέστηκε η προσομοίωση με το MLP καθώς και τον SVM να λαμβάνουν ως παραμέτρους τις default τους και το XGBoost στο οποίο ορίστηκε η gbtree ώστε να χρησιμοποιηθεί μοντέλο δένδρου και όχι γραμμικό. Πραγματοποιήθηκαν δύο προσομοιώσεις, με χρήση του MinMax Scaler και του Standard Scaler με σκοπό την εξερεύνηση της επίδρασης τους στα αποτελέσματα των προβλέψεων. Τα αποτελέσματα παρουσιάζονται συγκεντρωτικά στον Πίνακα 7.2.

Πίνακας 7. 2 Αποτελέσματα πρώτης υποπερίπτωσης

	Standard Scaler			MinMax Scaler		
	MLP	XGBoost	SVM	MLP	XGBoost	SVM
<b>Αργία</b>						
<b>Θεοφάνια</b>	1,71	1,52	1,67	2,4	2,64	2,499
<b>Καθαρά Δευτέρα</b>	3,09	2,56	3,41	3,64	2,78	3,57
<b>Επέτειος της Επανάστασης</b>	4,35	5,2	4,1	7,56	5,87	6,02
<b>Μεγάλη Παρασκευή</b>	1,92	1,86	2,00	1,35	1,82	2,28
<b>Κυριακή του Πάσχα</b>	2,62	2,66	2,32	3,24	2,8	3,026
<b>Δευτέρα του Πάσχα</b>	1,42	2,2	2,45	2,099	3,24	4,22
<b>Εργατική Πρωτομαγιά</b>	1,72	2,38	2,25	4,57	3,41	4,07
<b>Αγίου Πνεύματος</b>	3,29	4,96	3,58	8,91	7,60	6,18
<b>Κοίμηση της Θεοτόκου</b>	3,93	3,14	2,92	2,11	2,99	2,84
<b>Επέτειος του Όχι</b>	3,62	3,55	3,65	3,88	3,01	4,606
<b>Χριστούγεννα</b>	2,98	2,41	2,09	6,78	3,45	5,229



<b>2<sup>η</sup> Ημέρα των Χριστουγέννων</b>	1,66	1,82	1,71	4,11	3,47	3,462
<b>Πρωτοχρονιά</b>	2,07	2,69	2,25	2,82	3,18	2,49
<b>Μέσο σφάλμα</b>	<b>2,64</b>	<b>2,84</b>	<b>2,64</b>	<b>4,10</b>	<b>3,56</b>	<b>3,88</b>

Παρατηρείται η επίδραση της χρήσης των δύο μεθόδων scaling στις τιμές εισόδου του μοντέλου. Και τα τρία μοντέλα συμπεριφέρθηκαν καλύτερα με την χρήση του Standard Scaler δίνοντας σημαντικά μικρότερα σφάλματα. Όσο αφορά την σύγκριση μεταξύ των μοντέλων, το MLP και το SVM δίνουν παρόμοια σφάλματα με την διαφορά τους να βρίσκεται στο τρίτο δεκαδικό ψηφίο. Ο XGBoost δίνει μεγαλύτερο μέσο σφάλμα με την διαφορά να είναι της τάξης του 7,5%.

Στη δεύτερη υποπερίπτωση εισάγονται δείκτες για την ώρα της ημέρας, καθώς και το μέγιστο και ελάχιστο φορτίο της προηγούμενης ημέρας. Οι εισοδοί του μοντέλου διαμορφώνονται ακολούθως:

- Τιμές φορτίου προηγούμενης ημέρας (notation : D-1)
- Τιμές φορτίου ίδιας ημέρας την προηγούμενη εβδομάδα (notation : D- 7)
- Τιμές φορτίου ίδιας ημέρας το προηγούμενο έτος (notation : D – 365)
- Μέγιστο φορτίο προηγούμενης ημέρας (notation : MaxLoad)
- Ελάχιστο φορτίο προηγούμενης ημέρας (notation : MinLoad)
- Δείκτης ώρας, συνημίτονο του 1 - 24

Όπως και στην πρώτη υποπερίπτωση πραγματοποιήθηκαν οι προσομοιώσεις τόσο με τη χρήση του MinMax Scaler, όσο και του Standard Scaler. Οι παράμετροι των τριών μοντέλων παρέμειναν οι ίδιες με την πρώτη υποπερίπτωση. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 7.3.

Παρατηρείται το ίδιο μοτίβο κατά την σύγκριση με τα αποτελέσματα του Πίνακα 7.2, δηλαδή το μοντέλο συμπεριφέρεται καλύτερα με την χρήση του Standard Scaler. Η είσοδος των νέων χαρακτηριστικών δεν φαίνεται να βελτιώνει σε μεγάλο βαθμό τις επιδόσεις του μοντέλου, εν αντιθέσει, σε κάποιες ημέρες τις επιδεινώνει. Τέλος, οι

καλύτερες συνολικά προβλέψεις δόθηκαν με τη χρήση του XGBoost σε συνδυασμό με τον Standard Scaler, αν και η διαφορά είναι μικρή.

Πίνακας 7.3 Αποτελέσματα δεύτερης υποπερίπτωσης

	Standard Scaler			MinMax Scaler		
	MLP	XGBoost	SVM	MLP	XGBoost	SVM
<b>Αργία</b>						
<b>Θεοφάνια</b>	1,67	1,47	1,53	1,68	2,68	2,87
<b>Καθαρά Δευτέρα</b>	3,29	1,44	3,36	3,97	2,78	4,39
<b>Επέτειος της Επανάστασης</b>	4,53	4,75	3,81	4,62	5	5,14
<b>Μεγάλη Παρασκευή</b>	2,04	1,96	2,07	1,39	1,72	2,03
<b>Κυριακή του Πάσχα</b>	2,87	2,61	2,53	5,89	2,78	3,2
<b>Δευτέρα του Πάσχα</b>	1,41	2,07	2,4	3,29	3,06	4,25
<b>Εργατική Πρωτομαγιά</b>	2,3	2,39	2,69	4,62	3,31	4,92
<b>Αγίου Πνεύματος</b>	4,66	5,16	4,8	6,8	6,72	6,19
<b>Κοίμηση της Θεοτόκου</b>	3	3,32	2,75	3,03	2,85	2,22
<b>Επέτειος του Όχι</b>	3,57	3,5	3,65	3,94	2,92	4,61
<b>Χριστούγεννα</b>	3,12	2,39	2,88	8,77	3,55	5,65
<b>2<sup>η</sup> Ημέρα των Χριστουγέννων</b>	2,24	1,78	2,17	2,77	3,45	3,86
<b>Πρωτοχρονιά</b>	2,21	2,82	2,20	2,56	3,14	2,40
<b>Μέσο σφάλμα</b>	<b>2,838</b>	<b>2,831</b>	<b>2,833</b>	<b>4,10</b>	<b>3,38</b>	<b>3,98</b>

Κατά την τρίτη υποπερίπτωση, εισήχθησαν τα δεδομένα θερμοκρασίας στο μοντέλο για την διερεύνηση της επίδρασής της στα αποτελέσματα. Οι είσοδοι διαμορφώνονται ως εξής:

- Τιμές φορτίου προηγούμενης ημέρας (notation : D-1)
- Τιμές φορτίου ίδιας ημέρας την προηγούμενη εβδομάδα (notation : D- 7)
- Τιμές φορτίου ίδιας ημέρας το προηγούμενο έτος (notation : D – 365)
- Τιμές θερμοκρασίας προηγούμενης ημέρας (notation : temper D-1)
- Τιμές θερμοκρασίας ίδιας ημέρας την προηγούμενη εβδομάδα (notation: temperD- 7)
- Τιμές θερμοκρασίας ίδιας ημέρας το προηγούμενο έτος (notation: temperD – 365)
- Τιμές θερμοκρασίας ημέρας προς πρόβλεψη (notation: temperTarget)
- Μέγιστο φορτίο προηγούμενης ημέρας (notation: MaxLoad)
- Ελάχιστο φορτίο προηγούμενης ημέρας (notation: MinLoad)
- Δείκτης ώρας, συνημίτονο του 1 – 24

Πραγματοποιήθηκε η προσομοίωση με τις ίδιες ρυθμίσεις παραμέτρων και χρήση των δύο επιλογών για το scaling όπως και στις δύο προηγούμενες υποπεριπτώσεις. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 7.4

Όπως και στην δεύτερη υποπερίπτωση παρατηρείται μια αύξηση του μέσου σφάλματος σε όλα τα αποτελέσματα των μοντέλων πρόβλεψης. Παρόμοια τα μεγαλύτερα σφάλματα παρατηρούνται με την χρήση του Min Max Scaler όπως και σε προηγούμενες υποπεριπτώσεις. Συνολικά οι προβλέψεις με το μικρότερο σφάλμα δίνονται από το XGBoost μοντέλο.

Πίνακας 7.4 Αποτελέσματα τρίτης υποπερίπτωσης

	Standard Scaler			MinMax Scaler		
Αργία	MLP	XGBoost	SMV	MLP	XGBoost	SVM
Θεοφάνια	3,63	1,49	3,00	3,75	2,67	3,90
Καθαρά Δευτέρα	2,91	2,55	2,98	4,48	2,74	4,74

Επέτειος της Επανάστασης	6,14	6,68	5,48	5,83	7,33	7,63
Μεγάλη Παρασκευή	2,10	2,37	2,64	4,24	2,34	2,83
Κυριακή του Πάσχα	2,69	2,6	2,37	3,28	2,79	3,6
Δευτέρα του Πάσχα	2,03	1,99	1,83	2,42	2,92	2,97
Εργατική Πρωτομαγιά	2,03	2,35	2,70	4,18	3,33	4,65
Αγίου Πνεύματος	4,14	4,56	4,22	6,98	7,95	6,64
Κοίμηση της Θεοτόκου	2,27	2,97	2,91	1,97	3,02	3,34
Επέτειος του Όχι	3,10	3,4	3,09	3,94	3,25	4,56
Χριστούγεννα	4,55	2,49	4,07	7,53	3,75	6,60
2 <sup>η</sup> Ημέρα των Χριστουγέννων	1,80	1,79	2,95	2,71	3,44	4,37
Πρωτοχρονιά	1,67	2,87	2,84	2,16	3,32	3,70
<b>Μέσο σφάλμα</b>	<b>3,13</b>	<b>2,93</b>	<b>3,11</b>	<b>4,11</b>	<b>3,76</b>	<b>4,66</b>

Στην τελευταία υποπερίπτωση, προστέθηκαν επιπλέον δείκτες της ημέρας της εβδομάδας. Οι εισοδοι διαμορφώθηκαν ως εξής:

- Τιμές φορτίου προηγούμενης ημέρας (notation : D-1)
- Τιμές φορτίου ίδιας ημέρας την προηγούμενη εβδομάδα (notation : D- 7)
- Τιμές φορτίου ίδιας ημέρας το προηγούμενο έτος (notation : D – 365)
- Τιμές θερμοκρασίας προηγούμενης ημέρας (notation :temper D-1)
- Τιμές θερμοκρασίας ίδιας ημέρας την προηγούμενη εβδομάδα (notation : temperD- 7)
- Τιμές θερμοκρασίας ίδιας ημέρας το προηγούμενο έτος (notation : temperD – 365)
- Τιμές θερμοκρασίας ημέρας προς πρόβλεψη (notation : temperTarget)
- Μέγιστο φορτίο προηγούμενης ημέρας (notation : MaxLoad)
- Ελάχιστο φορτίο προηγούμενης ημέρας (notation : MinLoad)

- Δείκτης ώρας, συνημίτονο του 1 – 24
- Δείκτης ημέρας φορτίου προηγούμενης ημέρας (notation : weekDayLag1)
- Δείκτης ημέρας φορτίου προηγούμενης εβδομάδας (notation : weekDayLag2)
- Δείκτης ημέρας φορτίου προηγούμενου έτους (notation : weekDayLag3)

Κάνοντας το μοντέλο πιο πολύπλοκο θεωρήθηκε πως θα έχει θετική επίδραση στις προβλέψεις καθώς θα αποτύπωνε σε καλύτερο βαθμό τις μεταβλητές και την διαφορετικότητα των ημερών. Οι παράμετροι παραμένουν οι ίδιες όπως και στις προηγούμενες υποπεριπτώσεις . Τα αποτελέσματα παρουσιάζονται στον Πίνακα 7.5.

Πίνακας 7.5 Αποτελέσματα τέταρτης υποπερίπτωσης

	Standard Scaler			MinMax Scaler		
	MLP	XGBoost	SVM	MLP	XGBoost	SVM
<b>Αργία</b>						
<b>Θεοφάνια</b>	2,401	1,493	3,00	3,0962	2,673	3,60
<b>Καθαρά Δευτέρα</b>	2,643	2,557	2,98	3,696	2,748	5,67
<b>Επέτειος της Επανάστασης</b>	7,17	6,687	5,48	10,782	7,338	7,35
<b>Μεγάλη Παρασκευή</b>	2,337	2,37	2,64	4,612	2,349	2,86
<b>Κυριακή του Πάσχα</b>	2,252	2,608	2,37	2,365	2,794	3,48
<b>Δευτέρα του Πάσχα</b>	1,77	1,99	1,83	2,58	2,921	2,84
<b>Εργατική Πρωτομαγιά</b>	2,621	2,35	2,70	4,665	3,335	4,61
<b>Αγίου Πνεύματος</b>	4,34	4,561	4,22	12,311	7,951	6,63
<b>Κοίμηση της Θεοτόκου</b>	2,572	2,977	2,91	3,576	3,023	3,16
<b>Επέτειος του Όχι</b>	3,1196	3,406	3,09	4,05	3,25	4,49
<b>Χριστούγεννα</b>	3,287	2,49	4,07	12,486	3,751	6,33
<b>2<sup>η</sup> Ημέρα των Χριστουγέννων</b>	1,739	1,797	2,95	1,432	3,446	4,29

<b>Πρωτοχρονιά</b>	2,06	2,87	2,84	3,014	3,326	3,48
<b>Μέσο σφάλμα</b>	<b>2,94</b>	<b>2,93</b>	<b>3,11</b>	<b>5,28</b>	<b>3,76</b>	<b>4,52</b>

Αναλύοντας τα αποτελέσματα της 4<sup>ης</sup> υποπερίπτωσης, παρατηρούμε πως παρότι υπάρχει μια μικρή βελτίωση εν συγκρίσει με την 3<sup>η</sup> όσο αφορά τον MLP τα δυο άλλα μοντέλα δίνουν ίδια αποτελέσματα με την 3<sup>η</sup> περίπτωση, μη επηρεαζόμενα από τις νέες μεταβλητές που εισήχθησαν στο μοντέλο. Η σύγκριση μεταξύ των δύο τρόπων για την πραγματοποίηση του scaling αποδεικνύει για ακόμη μια φορά την υπεροχή του Standard Scaler, με σημαντικά μικρότερες τιμές σφάλματος για όλες τις ημέρες.

## 7.2 Συγκεντρωτική ανάλυση αποτελεσμάτων πρώτης προσέγγισης

Η πρώτη προσέγγιση που πραγματοποιήθηκε παρήγαγε σαφή στοιχεία σχετικά με την συμπεριφορά των εισόδων, της μεθόδου scaling καθώς και των τριών διαφορετικών μοντέλων που χρησιμοποιήθηκαν στην παρούσα εργασία. Αρχικά, όσο αφορά τις εισόδους, η πρώτη προσέγγιση που ταυτόχρονα είναι και η πιο απλή έδωσε τα καλύτερα αποτελέσματα. Παρότι στη βιβλιογραφία αναφέρεται πως η είσοδος των δεδομένων θερμοκρασίας βελτιώνει τα αποτελέσματα του μοντέλου, στην δική μας περίπτωση κάτι τέτοιο δεν παρατηρήθηκε, αντίθετα, όσο πιο πολλά δεδομένα εισήχθησαν τόσο μεγαλύτερο ήταν το σφάλμα.

Η σύγκριση μεταξύ των δύο μορφών scaling απέδειξε την υπεροχή του standard scaler έναντι του min max scaler. Και στις τέσσερις υποπεριπτώσεις, η χρήση του standard scaler ήταν αυτή με τα μικρότερα μέσα σφάλματα, χωρίς καμιά εξαίρεση. Το μεγαλύτερο σφάλμα που παρατηρήθηκε προέκυψε κατά την εφαρμογή του MinMax Scaler στην 4<sup>η</sup> υποπερίπτωση και ήταν της τάξης του 12,5 %. Εν αντιθέσει ο standard scaler έδωσε σαφώς καλύτερα αποτελέσματα σε όλες τις περιπτώσεις.

Όσο αφορά τα μοντέλα που χρησιμοποιήθηκαν, τα αποτελέσματα δείχνουν πως στις περισσότερες περιπτώσεις η χρήση είτε του MLP είτε του SVM δίνει καλύτερα αποτελέσματα από το XGBoost. Επιπλέον, τα δύο πρώτα δίνουν παρόμοιες επιδόσεις.

Όσο αφορά το scaling πιο ωφελημένα είναι τα MPL και SVM, αλλά και τα τρία μοντέλα έδωσαν τα καλύτερα αποτελέσματά τους κατά την χρήση του Standard Scaler. Για πιο εύκολη σύγκριση, παρουσιάζονται τα συγκεντρωτικά στοιχεία των τεσσάρων υποπεριπτώσεων στον Πίνακα 7.6.

Κλείνοντας, η πρώτη προσέγγιση που υλοποιήθηκε έδωσε σαφή στοιχεία για τον σχεδιασμό του επόμενου βήματος. Η χρήση του Standard Scaler είναι η επιλογή για το μοντέλο που θα δημιουργηθεί καθώς δίνει τα καλύτερα αποτελέσματα. Επιπλέον, θα δοκιμαστούν και τα τρία μοντέλα, καθώς τα αποτελέσματα που δίνουν είναι αρκετά κοντά. Όσο αφορά τις εισόδους, η χρήση του φορτίου ως μόνη είσοδο για το μοντέλο είναι αυτή που θα επιλεγεί βάση των αποτελεσμάτων. Η θερμοκρασία, θα αξιοποιηθεί με διαφορετικό τρόπο, ως κύριο στοιχείο για την επιλογή της κατάλληλης ημέρας του προηγούμενου έτους, ώστε τα φορτία να είναι όσο το δυνατόν πιο όμοια. Στη συνέχεια, θα γίνει η αξιολόγηση των εισόδων με την χρήση του Boruta SHAP, για την επιβεβαίωση της επιλογής των εισόδων.

Πίνακας 7. 6 Συγκεντρωτικά αποτελέσματα Standard Scaler

Υποπερίπτωση	1			2			3			4		
Αργία	MPL	XGB	SVM	MPL	XGB	SVM	MPL	XGB	SVM	MPL	XGB	SVM
Θεοφάνια	1,718	1,523	1,677	1,588	1,47	1,53	3,63	1,493	3	2,401	1,493	3
Καθαρά Δευτέρα	3,09	2,56	3,41	3,021	2,58	3,36	2,91	2,55	2,988	2,643	2,557	2,988
Επέτειος της Επανάστασης	4,35	5,2	4,1	5,15	4,75	3,81	6,14	6,687	5,48	7,17	6,687	5,487
Μεγάλη Παρασκευή	1,92	1,86	2	2,19	1,96	2,07	2,098	2,37	2,6404	2,337	2,37	2,64
Κυριακή του Πάσχα	2,62	2,65	2,32	2,45	2,61	2,53	2,689	2,608	2,377	2,252	2,608	2,377
Δευτέρα του Πάσχα	1,42	2,2	2,45	1,63	2,077	2,4	2,038	1,99	1,835	1,77	1,99	1,835
Εργατική Πρωτομαγιά	1,721	2,38	2,25	2,23	2,39	2,69	2,687	2,353	2,701	2,621	2,35	2,701
Αγίου Πνεύματος	3,29	4,96	3,58	4,73	5,16	4,8	4,14	4,561	4,222	4,34	4,561	4,222
Κοίμηση της	3,93	3,14	2,92	3,19	3,32	2,75	2,269	2,977	2,913	2,572	2,977	2,913

<b>Θεοτόκου</b>														
<b>Επέτειος του Όχι</b>	3,62	3,55	3,65	3,79	3,5	3,65	3,099	3,406	3,089	3,1196	3,406	3,089		
<b>Χριστούγεννα</b>	2,98	2,41	2,095	3	2,39	2,88	4,557	2,49	4,077	3,287	2,49	4,077		
<b>2<sup>η</sup> Ημέρα των Χριστουγέννων</b>	1,668	1,82	1,71	2,2	1,779	2,168	1,803	1,797	2,295	1,739	1,797	2,295		
<b>Πρωτοχρονιά</b>	2,077	2,695	2,253	1,73	2,82	2,197	2,677	2,875	2,846	2,06	2,87	2,846		
<b>Μέσο σφάλμα</b>	<b>2,64</b>	<b>2,84</b>	<b>2,64</b>	<b>2,83</b>	<b>2,83</b>	<b>2,83</b>	<b>3,13</b>	<b>2,93</b>	<b>3,11</b>	<b>2,94</b>	<b>2,93</b>	<b>3,11</b>		

### 7.3 Ανάλυση χαρακτηριστικών με την χρήση του Boruta-SHAP

Η χρήση του Boruta SHAP στην ανάλυση των εισόδων του μοντέλου κρίθηκε απαραίτητη για την επιβεβαίωση της επιλογής του φορτίου ως μοναδική είσοδο για το προς ανάπτυξη μοντέλο. Η λειτουργία του αναπτύχθηκε στο κεφάλαιο 6.3, και θεωρείται από τις πιο αποτελεσματικές μεθόδους για την επιλογή των εισόδων ενός μοντέλου. Παρόλο που η κύρια χρήση του είναι ακριβώς αυτή, εδώ θα χρησιμοποιηθεί για την αξιολόγηση όλων των εισόδων. Χρησιμοποιήθηκε η βιβλιοθήκη Boruta [Keaney, E., 2022.]

Για την υλοποίηση του δημιουργήθηκε μια ξεχωριστή εφαρμογή στα πλαίσια της εφαρμογής που γράφτηκε για την πρώτη προσέγγιση. Χρησιμοποιήθηκε το Pycharm ως IDE και η δομή του προγράμματος είναι ως εξής.

Data

Εδώ γίνεται η είσοδος των δεδομένων φορτίου και θερμοκρασίας στο Pycharm.

Functions -> dataInput

Στην μονάδα αυτή διαβάζονται τα δεδομένα από το αρχείο excel και μετατρέπονται σε 24ωρη μορφή με την χρήση ως header της ημερομηνίας.

Holidays

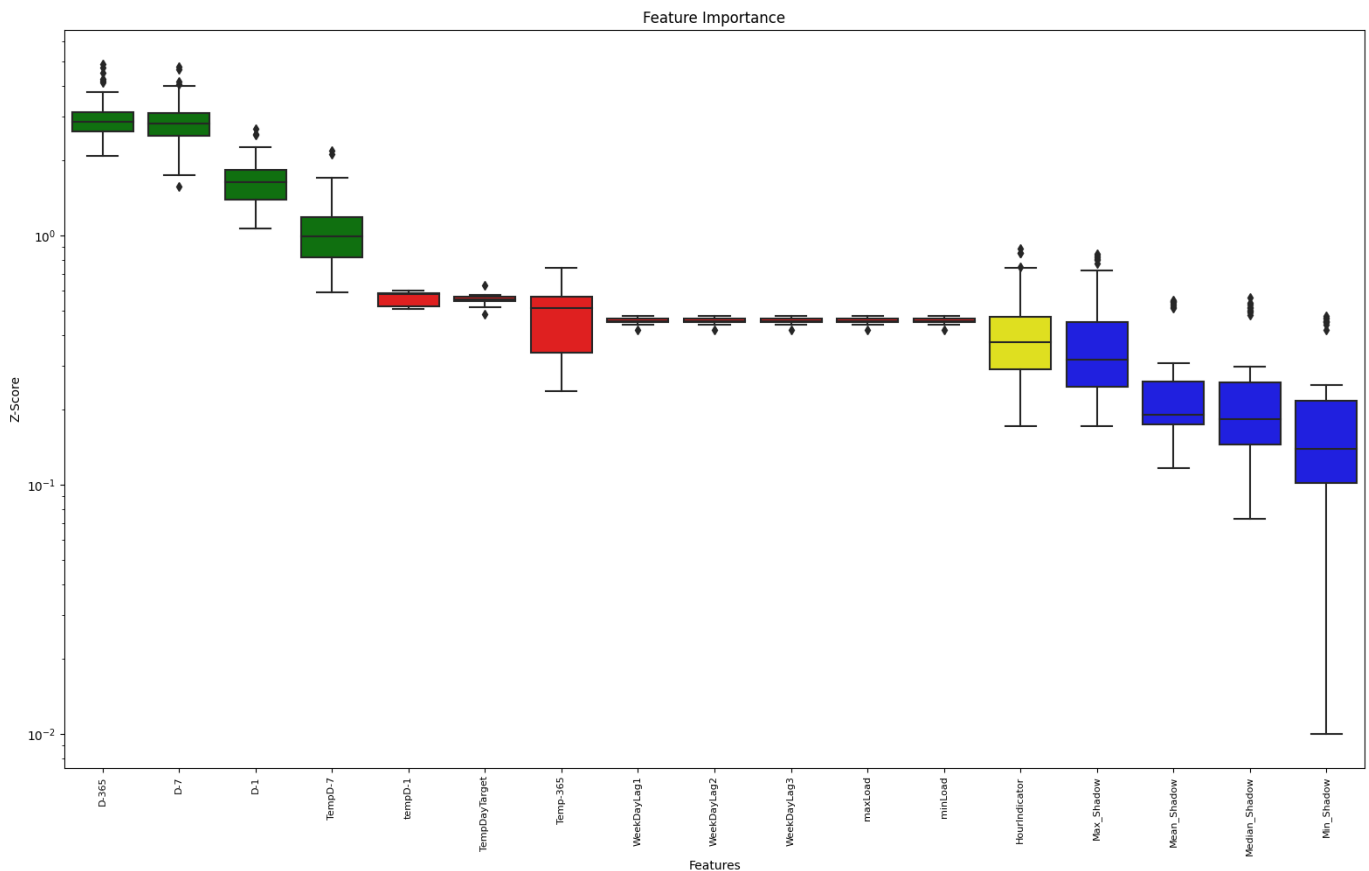
Εδώ δημιουργούμε ξεχωριστά αρχεία για κάθε αργία, όπου στο καθένα επιλέγουμε τα δεδομένα εισόδου, τα μετασχηματίζουμε στην κατάλληλη μορφή και τέλος τρέχουμε τον αλγόριθμο Boruta SHAP. Τα αποτελέσματα έχουν την μορφή της Εικόνας 7.1



```
IPython could not be loaded!  
100%|██████████| 50/50 [00:04<00:00, 11.59it/s]  
4 attributes confirmed important: ['D-365', 'D-1', 'D-7', 'TempD-7']  
8 attributes confirmed unimportant: ['WeekDayLag3', 'maxLoad', 'WeekDayLag2', 'TempDayTarget', 'minLoad', 'tempD-1', 'WeekDayLag1', 'Temp-365']  
1 tentative attributes remains: ['HourIndicator']
```

Εικόνα 7.1 Αποτελέσματα Boruta-SHAP

Όπως παρατηρείται η έξοδος του αλγορίθμου χωρίζεται σε τρεις λίστες. Τα χαρακτηριστικά που είναι σημαντικά, τα χαρακτηριστικά που δεν είναι σημαντικά και



Εικόνα 7.2 Οπτικοποίηση αποτελεσμάτων

τέλος τα χαρακτηριστικά που είναι απροσδιόριστη η επίδρασή τους. Εφόσον το επιθυμούμε είναι δυνατή η οπτικοποίηση των αποτελεσμάτων με τον τρόπο που αυτά παρουσιάζονται στην Εικόνα 7.2

Οι δύο επεξηγηματικές εικόνες αποτελούν την ανάλυση για την ημέρα των Θεοφανίων. Παρατηρούμε πως οι εισοδοί με την μεγαλύτερη επίδραση στο μοντέλο είναι οι εισοδοί

του φορτίου του προηγούμενου έτους, της προηγούμενης εβδομάδας της προηγούμενης ημέρας και της θερμοκρασίας της προηγούμενης εβδομάδας. Την υψηλότερη επίδραση ασκούν οι εισόδοι των φορτίων.

Στη συνέχεια εκτελούμε την αξιολόγηση σε όλες τις αργίες με τα αποτελέσματα να παρουσιάζονται συγκεντρωτικά στον Πίνακα 7.7. Κωδικοποιούμε τα αποτελέσματα ως εξής. Με I χαρακτηρίζουμε τα σημαντικά, με U τα μη σημαντικά και με T τα απροσδιόριστα.

Είναι εμφανής η τάση των πιο σημαντικών παραμέτρων εισόδου να είναι στο φορτίο, με τα υπόλοιπα στοιχεία εισόδου να μην είναι σημαντικά η να είναι απροσδιόριστα. Βάση αυτού, μπορούμε να επιβεβαιώσουμε πως η επιλογή της χρήσης του φορτίου ως μοναδικής εισόδου στο προς ανάπτυξη μοντέλο φαίνεται να είναι σωστή.

Πίνακας 7.7 Αποτελέσματα Boruta-SHAP

Είσοδος	D-1	D-7	D-365	temper D-1	temper D-7	temper D-365	temper Target	WeekdayLag1	WeekdayLag2	WeekdayLag3	Min Load	Max Load	Hour Indicator
<b>Αργία</b>													
<b>Θεοφάνια</b>	I	I	I	U	I	U	U	U	U	U	U	U	T
<b>Καθαρά Δευτέρα</b>	I	I	I	U	U	U	T	U	U	U	U	U	U
<b>Επέτειος της Επανάστασης</b>	U	I	I	U	I	I	U	U	U	U	U	U	U
<b>Μεγάλη Παρασκευή</b>	I	U	I	U	I	U	U	U	U	U	U	U	U
<b>Κυριακή του Πάσχα</b>	I	I	I	U	U	U	U	U	U	U	U	U	U
<b>Δευτέρα του Πάσχα</b>	I	I	I	U	U	U	U	U	U	U	U	U	U
<b>Εργατική Πρωτομαγιά</b>	I	I	I	U	U	U	U	U	U	U	U	U	U

Αγίου Πνεύματος	I	I	I	U	U	I	U	U	U	U	U	U	U
Κοίμηση της Θεοτόκου	I	I	U	U	U	U	U	U	U	U	U	U	U
Επέτειος του Όχι	I	I	I	U	T	U	U	U	U	U	U	U	T
Χριστούγεννα	I	I	I	U	U	U	U	U	U	U	U	U	U
2 <sup>η</sup> Ημέρα των Χριστουγέννων	I	I	I	U	U	T	T	U	U	U	U	U	U
Πρωτοχρονιά	I	I	I	U	U	U	U	U	U	U	U	U	U

## 7.4 Προτεινόμενη μεθοδολογία

Η προτεινόμενη μεθοδολογία στηρίζεται σε δύο μέρη, την εύρεση και χρήση ως ημερών εκπαίδευσης όμοιων θερμοκρασιακά ημερών, και σε δεύτερο χρόνο την προσπάθεια βελτιστοποίησης των παραμέτρων των ΤΝΔ.

7.4.1 Μέρος πρώτο. Εύρεση και επιλογή όμοιων ημερών.

Όπως είδαμε στο κεφάλαιο 4 η θερμοκρασία και η κατανάλωση ενέργειας αποτελούν δυο αλληλοσχετιζόμενες μεταβλητές. Η αδυναμία της αρχικής προσέγγισης να εκμεταλλευτεί την συσχέτιση αυτή είχε ως αποτέλεσμα την χρήση της θερμοκρασίας ως οδηγό για την εύρεση και χρήση των ομοίων ημερών με την χρήση της απόστασης Fréchet οριζόμενη από την σχέση 7.2 όπου  $d$  είναι η distance function του  $S$ , και  $A, B$  δύο καμπύλες στον χώρο  $S$ .

$$F(A, B) = \inf_{a, \beta} \max_{t \in [0,1]} \{d(A(a(t)), B(\beta)(t))\} \quad (7.2)$$

Συγκεκριμένα, χρησιμοποιείται η διακριτή απόσταση Fréchet, η οποία αποτελεί μια προσέγγιση της απόστασης Fréchet για πολυγωνικές καμπύλες [Eiter & Mannila, 1994]. Για την υλοποίησή της στην python γίνεται χρήση του πακέτου similaritymeasures. [similaritymeasures, 2022]. Μετά την επιλογή της πιο κοντινής θερμοκρασιακά ημέρας,

τρέχουμε και τα τρία ΤΝΔ με τις παραμέτρους τους στις default τιμές ώστε να υπάρχει άμεση σύγκριση με τα καλύτερα αποτελέσματα της πρώτης προσέγγισης.

Αρχικά επιλέγονται οι ημέρες. Χωρίζονται τα δεδομένα σε δύο ομάδες. Από το 2009 έως το 2016 έχουμε τον χώρο αναζήτησης των δεδομένων εκπαίδευσης, και από το 2017 έως το 2019 έχουμε τα δεδομένα δοκιμής. Οι ημέρες που επιλέχθηκαν παρουσιάζονται στον Πίνακα 7.8.

Πίνακας 7.8 Ημέρες που επιλέχθηκαν ως οι πιο όμοιες

<b>Αργία</b>	<b>Επιλογή ημέρας (YYYY-MM-DD)</b>
<b>Θεοφάνια</b>	2015-01-06
<b>Καθαρά Δευτέρα</b>	2009-03-02
<b>Επέτειος της Επανάστασης</b>	2011-03-25
<b>Μεγάλη Παρασκευή</b>	2016-04-29
<b>Κυριακή του Πάσχα</b>	2016-05-01
<b>Δευτέρα του Πάσχα</b>	2015-04-13
<b>Εργατική Πρωτομαγιά</b>	2009-05-01
<b>Αγίου Πνεύματος</b>	2009-06-08
<b>Κοίμηση της Θεοτόκου</b>	2011-08-15
<b>Επέτειος του Όχι</b>	2009-10-28
<b>Χριστούγεννα</b>	2013-12-25
<b>2<sup>η</sup> Ημέρα των Χριστουγέννων</b>	2016-12-26
<b>Πρωτοχρονιά</b>	2014-01-01

Στη συνέχεια τρέξαμε το προς κατασκευή μοντέλο με τις νέες εισόδους, για να αξιολογηθεί η επίδραση των νέων εισόδων. Τα δεδομένα παρουσιάζονται στον Πίνακα 7.9.

Πίνακας 7.9 Αποτελέσματα ομοίων ημερών

Αργία	MLP	XGB	SVM
Θεοφάνια	1,22	1,11	1,12
Καθαρά	2,75	2,77	2,95
<b>Δευτέρα</b>			
Επέτειος της	4,63	4,47	4,92
Επανάστασης			
Μεγάλη	2,1	1,9	1,98
Παρασκευή			
Κυριακή του	2,59	2,7	2,33
Πάσχα			
Δευτέρα του	2,47	4,1	2,83
Πάσχα			
Εργατική	1,85	2,6	2,02
Πρωτομαγιά			
Αγίου	1,29	1,91	1,12
Πνεύματος			
Κοίμηση της	1,91	3,06	1,71
Θεοτόκου			
Επέτειος του	3,23	2,75	2,92
Όχι			
Χριστούγεννα	3,19	3,13	3,33
2 <sup>η</sup> Ημέρα των	2,01	1,79	1,7
Χριστουγέννων			
Πρωτοχρονιά	2,83	2,66	2,26
Μέσος όρος	<b>2,46</b>	<b>2,68</b>	<b>2,39</b>

Συγκρίνοντας τα αποτελέσματα με τα καλύτερα που δόθηκαν από την πρώτη υποπερίπτωση της πρώτης προσέγγισης παρατηρούμε την βελτίωση τους και με τα τρία μοντέλα. Πιο συγκεκριμένα το σφάλμα του MLP μειώθηκε κατά 6,78%, του XGB κατά

5,4% και του SVM κατά 9,37% που σημαίνει πως ευνοήθηκε περισσότερο από την επιλογή των ημερών.

Στη συνέχεια, θα πραγματοποιηθεί η βελτιστοποίηση των παραμέτρων των ΤΝΔ σε μια προσπάθεια να μειωθεί περαιτέρω το σφάλμα.

#### 7.4.2 Μέρος Δεύτερο. Βελτιστοποίηση

Για την εύρεση των βέλτιστων παραμέτρων θα χρησιμοποιηθεί η μέθοδος Bayesian optimization με τη χρήση της βιβλιοθήκης BayesSearchCV [Scikit-optimize.github.io. 2022 skopt.BayesSearchCV], η λειτουργία της οποίας αναπτύχθηκε στο 6<sup>ο</sup> κεφάλαιο. Θα αναλυθεί η βελτιστοποίηση των παραμέτρων για κάθε μοντέλο ξεχωριστά.

#### MLP

Οι παράμετροι που καθορίζουν την λειτουργία του MLP στην υλοποίηση που χρησιμοποιείται στην παρούσα εργασία είναι πάνω από 20 [sklearn.neural\_network.MLPRegressor, 2022]. Η εύρεση των βέλτιστων τιμών θα ήταν χρονοβόρα και θα απαιτούσε εμπειρία, στην περίπτωση μη χρήσης ενός αλγορίθμου βελτιστοποίησης.

Η αδυναμία της υλοποίησης του bayesSearchSV στην αναζήτηση arrays είχε ως αποτέλεσμα η αναζήτηση των βέλτιστων τιμών να γίνει σε δύο μέρη. Αρχικά, εφαρμόστηκε μια λύση για την υπέρβαση της αδυναμίας αυτής, και χρησιμοποιήθηκε για την εύρεση του αριθμού των νευρώνων και του αριθμού των κρυφών στρωμάτων. Στη συνέχεια αναζητήθηκαν οι παράμετροι του Πίνακα 7.10

Πίνακας 7.10 Παράμετροι αναζήτησης MLP

Παράμετρος	Εύρος τιμών
Random state	0-1
Activation	relu, tanh, logistic, identity
Solver	Lbfgs, sgd, adam
Early stopping	True, False

Τέλος εισήχθησαν τα νέα δεδομένα στο μοντέλο. Τα αποτελέσματα παρουσιάζονται στον Πίνακα 7.11. Είναι εμφανής η βελτίωση σε όλο το εύρος των αποτελεσμάτων. Το σφάλμα μειώνεται, με το μέσο σφάλμα να αγγίζει το 2,04%, βελτίωση της τάξης του 18,8%.

Πίνακας 7.11 Αποτελέσματα MLP μετά τη βελτιστοποίηση

Αργία	Σφάλμα
Θεοφάνια	1,16
Καθαρά Δευτέρα	2,53
Επέτειος της Επανάστασης	4,19
Μεγάλη Παρασκευή	1,48
Κυριακή του Πάσχα	2,47
Δευτέρα του Πάσχα	1,63
Εργατική Πρωτομαγιά	1,85
Αγίου Πνεύματος	0,78
Κοίμηση της Θεοτόκου	1,74
Επέτειος του Όχι	2,36
Χριστούγεννα	2,91
2 <sup>η</sup> Ημέρα των Χριστουγέννων	1,47
Πρωτοχρονιά	2,06
Μέσος Όρος	<b>2,04</b>

#### *XGBoost*

Στην συνέχεια πραγματοποιήθηκε η βελτιστοποίηση των παραμέτρων όσο αφορά τον XGBoost. Οι παράμετροι που αναζητήθηκαν παρουσιάζονται στον Πίνακα 7.12.

Πίνακας 7.12 Παράμετροι βελτιστοποίησης XGBoost

Παράμετρος	Εύρος τιμών
------------	-------------

<b>Learning rate</b>	0.01 - 2.0
<b>Max depth</b>	1 - 9
<b>Reg lambda</b>	0.1 - 50
<b>Reg alpha</b>	0 - 1
<b>n_estimators</b>	20, 1000
<b>Base score</b>	0.1, 5

Μετά την ολοκλήρωση της βελτιστοποίησης, εκτελέστηκε η προσομοίωση με τα νέα δεδομένα, τα αποτελέσματα της οποίας παρουσιάζονται στον Πίνακα 7.13. Η βελτιστοποίηση δεν είχε θετικό αποτέλεσμα στην περίπτωση του XGBoost, παρότι έγινε προσπάθεια να αναζητηθούν μεγάλο μέρος των παραμέτρων που καθορίζουν τη λειτουργία του μοντέλου. Η διαφορά στο μέσο όρο είναι ελάχιστη, στα όρια του σφάλματος οπότε δεν λαμβάνεται υπόψιν. Ουσιαστικά, τα αποτελέσματα είναι παρόμοια με του Πίνακα 7.9, με μικρές διαφορές.

Πίνακας 7.13 Αποτελέσματα μετά τη βελτιστοποίηση XGBoost

<b>Αργία</b>	<b>Σφάλμα</b>
<b>Θεοφάνια</b>	1,55
<b>Καθαρά Δευτέρα</b>	2,62
<b>Επέτειος της Επανάστασης</b>	3,65
<b>Μεγάλη Παρασκευή</b>	1,97
<b>Κυριακή του Πάσχα</b>	1,88
<b>Δευτέρα του Πάσχα</b>	3,28
<b>Εργατική Πρωτομαγιά</b>	2,24
<b>Αγίου Πνεύματος</b>	2,24
<b>Κοίμηση της Θεοτόκου</b>	2,43
<b>Επέτειος του Όχι</b>	4,82
<b>Χριστούγεννα</b>	3,57
<b>2<sup>η</sup> Ημέρα των Χριστουγέννων</b>	1,8
<b>Πρωτοχρονιά</b>	2,66
<b>Μέσος Όρος</b>	<b>2,67</b>



## SVM

Τέλος, πραγματοποιήθηκε η αναζήτηση των βέλτιστων παραμέτρων για το SVM. Οι παράμετροι και το εύρος τιμών παρουσιάζεται στον Πίνακα 7.14.

Πίνακας 7.14 Παράμετροι αναζήτησης SVM

Παράμετρος	Εύρος τιμών
Kernel	Linear, poly, rbf
C	1e <sup>-6</sup> - 100
Degree	1-5

Εισαγάγαμε τα δεδομένα των νέων παραμέτρων στο μοντέλο και το τρέξαμε. Τα αποτελέσματα φαίνονται στον Πίνακα 7.15

Πίνακας 7.15 Αποτελέσματα μετά τη βελτιστοποίηση SVM

Αργία	Σφάλμα
Θεοφάνια	1,26
Καθαρά Δευτέρα	3,21
Επέτειος της Επανάστασης	4,49
Μεγάλη Παρασκευή	1,75
Κυριακή του Πάσχα	2,55
Δευτέρα του Πάσχα	2,44
Εργατική Πρωτομαγιά	1,95
Αγίου Πνεύματος	1,3
Κοίμηση της Θεοτόκου	1,53
Επέτειος του Όχι	2,31
Χριστούγεννα	2,95
2 <sup>η</sup> Ημέρα των Χριστουγέννων	1,32

Πρωτοχρονιά	2,48
Μέσος Όρος	2,27

Σε αντίθεση με τον XGBoost, ο SVM φαίνεται να ωφελείται με την βελτιστοποίηση των παραμέτρων. Το μέσο σφάλμα πέφτει στο 2,27%, με την διαφορά να είναι της τάξης του 5,5% εν συγκρίσει με τα αποτελέσματα του Πίνακα 7.9.

## 7.5 Σχολιασμός της δεύτερης προσέγγισης, και πρόταση μεθοδολογίας.

Τα αποτελέσματα που πήραμε κατά την δεύτερη προσέγγιση καταδεικνύουν την καταλληλότητα του MLP για την χρήση του στην προτεινόμενη μεθοδολογία, καθώς ανταποκρίθηκε καλύτερα στις απαιτήσεις που τέθηκαν. Αρχικά, οι επιδόσεις του βελτιώθηκαν κατά την επιλογή των ομοίων ημερών, αλλά αυτό που τον κάνει τον κατάλληλο για την χρήση του είναι η πολύ μεγάλη βελτίωση του μετά την βελτιστοποίηση των παραμέτρων του μοντέλου, πράγμα που οδήγησε στην κατακόρυφη πτώση του σφάλματος, το οποίο κινείται στα όρια του 2%, το οποίο είναι χαμηλό συγκρινόμενο με την βιβλιογραφία.

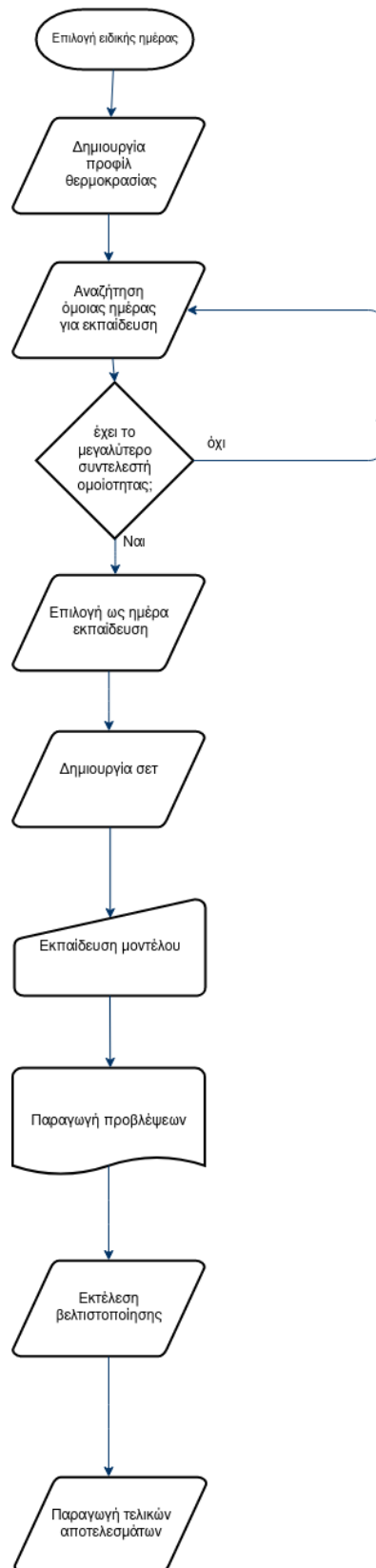
Τα δύο άλλα μοντέλα δεν κατάφεραν να ωφεληθούν από την τεχνική που επιλέχθηκε. Αυτό δεν σημαίνει πως οι επιδόσεις τους είναι χαμηλές, καθώς και η βιβλιογραφία κινείται μεταξύ του 2-4%.

Συνοψίζοντας, η προτεινόμενη μεθοδολογία απαρτίζεται από τα ακόλουθα βήματα.

- Αρχικά, βρίσκουμε την πιο όμοια θερμοκρασιακά ημέρα, και την χρησιμοποιούμε στα δεδομένα εκπαίδευσης με τη χρήση της διακριτής απόστασης Fréchet.
- Στη συνέχεια, εκτελούμε την βελτιστοποίηση των παραμέτρων για τον MLP.
- Τέλος, παράγεται η πρόβλεψη της επόμενης ημέρας.

Η μεθοδολογία παρουσιάζεται στο διάγραμμα ροής της Εικόνας 7.3

Ο κώδικας που αναπτύχθηκε και στην παρούσα εργασία είναι διαθέσιμος στο GitHub στη διεύθυνση <https://github.com/kosPap1/MscThesisLoadPredictionOnHolidays>



Εικόνα 7.3 Διάγραμμα ροής μεθοδολογίας

## Κεφάλαιο 8 Συμπεράσματα και προτάσεις για περαιτέρω έρευνα

### 8.1 Συμπεράσματα

Αρχικά, τα αποτελέσματα που εξήχθησαν στην παρούσα εργασία οδηγούν στο συμπέρασμα πως το φορτίο παραμένει ίσως ο πιο σημαντικός παράγοντας κατά την διαδικασία επιλογής των εισόδων ενός μοντέλου για την πρόβλεψη του φορτίου της επόμενης ημέρας στην ειδική περίπτωση των αργιών. Στο 7.1 δοκιμάστηκαν 4 σετ διαφορετικών παραμέτρων εισόδου, και από αυτά το μικρότερο σφάλμα, και κατά συνέπεια οι πιο ακριβείς μετρήσεις δόθηκαν από το πρώτο σετ, που αποτελείται αποκλειστικά από τα δεδομένα φορτίου. Επιπλέον, κατά την διαδικασία της αξιολόγησης των εισόδων με την χρήση του Boruta-SHAP στην πλειονότητα των περιπτώσεων το φορτίο ήταν η πιο σημαντική είσοδος.

Η κανονικοποίηση των δεδομένων παίζει σημαντικό ρόλο στα αποτελέσματα και των τριών μοντέλων που υλοποιήθηκαν. Στο κεφάλαιο 7.1 δοκιμάστηκαν δυο τρόποι κανονικοποίησης, η χρήση του Standard Scaler και του Min Max Scaler. Σε όλες σχεδόν τις περιπτώσεις ο Standard Scaler έδωσε με διαφορά τα καλύτερα αποτελέσματα, βελτιώνοντας σημαντικά τις επιδόσεις των μοντέλων σε κάθε αργία. Είναι εύκολο λοιπόν το συμπέρασμα πως στη δική μας περίπτωση, η χρήση του Standard Scaler αποτελεί μονόδρομο.

Επιλέγοντας όμοιες θερμοκρασιακά ημέρες είχε ως αποτέλεσμα μείωση των σφαλμάτων οδηγώντας μας στην υιοθέτηση της στο προτεινόμενο μοντέλο. Η υπόθεση που έγινε, πως δηλαδή οι όμοιες θερμοκρασιακά αργίες θα έχουν όμοιο φορτίο, επαληθεύτηκε κατά την εφαρμογή της από τα δεδομένα που πήραμε. Παρότι ο ρόλος των ομοίων ημερών δεν είναι τόσο σημαντικός όσο της βελτιστοποίησης, η μείωση του μέσου σφάλματος δεν είναι αμελητέα. Παρόλα αυτά, το μειονέκτημα είναι η παροχή ακριβών προβλέψεων για τον καιρό της προς εξέτασης ημέρας, ώστε να βρεθεί η αντίστοιχη στο παρελθόν. Σε περίπτωση που η πρόβλεψη θερμοκρασίας είναι λανθασμένη, όλο το

μοντέλο θα δώσει λανθασμένα αποτελέσματα, λόγω της επιλογής διαφορετικής θερμοκρασιακά ημέρας.

Η αναζήτηση των βέλτιστων παραμέτρων λειτουργίας των μοντέλων παίζει εξίσου σημαντικό ρόλο. Η χρήση της τεχνικής Bayesian optimization είναι υπεύθυνη για την μείωση του σφάλματος σε πολύ μεγάλο βαθμό, βελτιώνοντας τις επιδόσεις στο μοντέλο που επιλέχθηκε κατά 15%, ποσοστό που δεν είναι καθόλου αμελητέο. Αυτό όμως δεν έρχεται χωρίς κόστος. Ο χρόνος αναζήτησης παραμένει υψηλός, της τάξης των ~400 δευτερολέπτων κατά μέσο όρο για την αναζήτηση των παραμέτρων του XGBoost, της τάξης των ~300sec για τον MLP και ~250sec για τον SVM με τους χρόνους αυτούς να σημειώνονται σε υπολογιστή με τα χαρακτηριστικά : i5-7500, 16GB RAM, NVme SSD, OS Fedora 35. Επιπλέον, ο αριθμός των interactions για την επιλογή των χαρακτηριστικών ανήλθε στις 100 για τον XGBoost και στις 128 για τον MLP. Σαφώς μεγαλύτερος αριθμός interactions θα έδινε καλύτερα αποτελέσματα, καθώς οι πιθανότητες του αλγορίθμου να βρει καλύτερη λύση θα αυξανόταν. Ενδεικτικά μπορούμε να αναφέρουμε για 512 interactions ο χρόνος εκτέλεσης είναι της τάξης των 14323 sec για τις παραμέτρους του XGBoost και της τάξης των 12817 για τον MLP, χρόνοι απαγορευτικοί.

Κλείνοντας, το προτεινόμενο μοντέλο δείχνει να λειτουργεί αρκετά καλά, δίδοντας χαμηλά ποσοστά σφάλματος, πολλές φορές κάτω από 2%. Η χρήση του MLP σε συνδυασμό με την εύρεση ομοίων θερμοκρασιακά ημερών και την χρήση βελτιστοποίησης είχε ως αποτέλεσμα την συνολική μείωση του μέσου σφάλματος κατά 29,4% εν συγκρίσει με την αρχική προσέγγιση της πρώτης υποπερίπτωσης με αποτέλεσμα την μείωση του μέσου σφάλματος στο 2,04%, ποσοστό που είναι αρκετά καλό εν συγκρίσει με την βιβλιογραφία. Υπάρχουν βέβαια ημέρες, όπως η επέτειος της επανάστασης, όπου το μοντέλο δίνει τα μεγαλύτερα σφάλματά του, της τάξης του 4,19%, αλλά αντίθετα 6 από τις 13 αργίες είναι κάτω του 2% και η αργία του Αγίου Πνεύματος πέφτει κάτω από 1%, ποσοστό εξαιρετικό.

## 8.2 Περεταίρω έρευνα

Η βελτιστοποίηση των παραμέτρων των μοντέλων αποδεικνύεται μια καλή στρατηγική για την μείωση των σφαλμάτων στις προβλέψεις φορτίου της επόμενης ημέρας. Η μέθοδος που χρησιμοποιήθηκε στην παρούσα εργασία δεν είναι η μοναδική. Υπάρχει μια πλειάδα αλγορίθμων, όπως ο grid Search, ο Random Search, ο Gradiend based optimization algorithm και άλλοι. Δεδομένων των αποτελεσμάτων της παρούσας εργασίας θα ήταν καλή επιλογή για μετέπειτα έρευνα η δοκιμή των παραπάνω αλγορίθμων.

Τα τρία μοντέλα που χρησιμοποιήθηκαν δεν είναι τα μοναδικά. Υπάρχει μεγάλος αριθμός ΤΝΔ και θα θέλαμε να δούμε τις επιδόσεις τους κάτω από τις παραμέτρους του προταθέντος μοντέλου. Ο μόνος περιορισμός είναι η δυνατότητα για βελτίωση των παραμέτρων λειτουργίας τους.



## Βιβλιογραφία

2006. *European SmartGrids Technology Platform*. Luxembourg: Publications Office.

2020. Climate Change.

Admie.gr. 2022. *Δεδομένα / ΑΔΜΗΕ*. [online] Available at:

<<https://www.admie.gr/agora/statistika-agoras/dedomena>> [Accessed 29 May 2022].

An introduction to explainable AI with Shapley values — SHAP latest documentation, 2022

Ardito, L., Procaccianti, G., Menga, G. and Morisio, M., 2013. Smart Grid Technologies in Europe: An Overview. *Energies*, 6(1), pp.251-281.

Arora, S. and Taylor, J., 2013. Short-Term Forecasting of Anomalous Load Using Rule-Based Triple Seasonal Methods. *IEEE Transactions on Power Systems*, 28(3), pp.3235-3242.

Arora, S. and Taylor, J., 2018. Rule-based autoregressive moving average models for forecasting load on special days: A case study for France. *European Journal of Operational Research*, 266(1), pp.259-268.

Barman, M. and Dev Choudhury, N., 2020. A similarity-based hybrid GWO-SVM method of power system load forecasting for regional special event days in anomalous load situations in Assam, India. *Sustainable Cities and Society*, 61, p.102311.

Becks, & Bitterer, & Brandstetter, & Buchholz, & Englert, & Glaunsinger, & Hagen, & Kahmann, & Horenkamp, & Kerber, & Kiessling, Andreas & Krause, & Malina, & Meierhofer, & Nickel, & Rehtanz, Christian & Sporer, & Stein, & Steusloff, & Sengbusch. (2010). The German Standardization Roadmap E-Energy/Smart Grid.

Bell, E. and Cullen, J., 2015. *Adoption of the Paris Agreement, 21st Conference of the Parties, Paris: United Nations*. [Framework] United Nations / Framework Convention on Climate Change.

Capasso, A., Grattieri, W., Lamedica, R. and Prudenzi, A., 1994. A bottom-up approach to residential load modeling. *IEEE Transactions on Power Systems*, 9(2), pp.957-964.

Chen, T. and Guestrin, C., 2016. XGBoost. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.



Cristianini, N. and Ricci, E., 2008. Support Vector Machines. *Encyclopedia of Algorithms*, pp.928-932.

Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20(3), pp.273-297.

Dejun Tu, 2014. Automatic control system of voltage and reactive power in the dispatching side and its application. *2014 IEEE Workshop on Electronics, Computer and Applications*.

Education, I., 2022. What are Neural Networks? [online] *Ibm.com*. Available at: <<https://www.ibm.com/cloud/learn/neural-networks>> [Accessed 10 June 2022].

Eesi.org. 2022. *Fossil Fuels | EESI*. [online] Available at: <<https://www.eesi.org/topics/fossil-fuels/description>> [Accessed 23 May 2022].

Eiter, Thomas & Mannila, Heikki. (1994). Computing Discrete Frechet Distance.

En.wikipedia.org. 2022. *Phasor measurement unit - Wikipedia*. [online] Available at: <[https://en.wikipedia.org/wiki/Phasor\\_measurement\\_unit](https://en.wikipedia.org/wiki/Phasor_measurement_unit)> [Accessed 23 May 2022].

*ENERGY INDEPENDENCE AND SECURITY ACT OF 2007*. XIII.

EUROPEAN COMMISSION, 2021. *Directive 2009/28/EC, Promotion of energy from renewable sources*. Brussels: European Union.

Feinberg, E. and Genethliou, D., n.d. Load Forecasting. *Power Electronics and Power Systems*, pp.269-285.

Fiot, J. and Dinuzzo, F., 2018. Electricity Demand Forecasting by Multi-Task Learning. *IEEE Transactions on Smart Grid*, 9(2), pp.544-551.

Fischer, J. and Elert, G., 2003. *The Physics Factbook*.

Gopstein, A., Nguyen, C., O'Fallon, C., Hastings, N. and Wollman, D., 2021. NIST framework and roadmap for smart grid interoperability standards, release 4.0.

<https://www.entsoe.eu/>. 2022. *Total Load - Day Ahead / Actual*. [online] Available at: <<https://www.entsoe.eu/>> [Accessed 29 May 2022].

Iqtiyaniillham, N., Hasanuzzaman, M. and Hosenuzzaman, M., 2017. European smart grid prospects, policies, and challenges. *Renewable and Sustainable Energy Reviews*, 67, pp.776-790.

Jahan, I., Snasel, V. and Misak, S., 2020. Intelligent Systems for Power Load Forecasting: A Study Review. *Energies*, 13(22), p.6105.

Jiang, I., Nam, G., Chang, H., Nassif, S. and Hayes, J., 2014. Smart grid load balancing techniques via simultaneous switch/tie-line/wire configurations. *2014 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*

JetBrains. 2022. *PyCharm: the Python IDE for Professional Developers by JetBrains*. [online] Available at: <<https://www.jetbrains.com/pycharm/>> [Accessed 27 September 2022].

Keany, E., 2022. BorutaShap : A wrapper feature selection method which combines the Boruta feature selection algorithm with Shapley values.. [online] Zenodo. Available at: <<https://zenodo.org/record/4247618#.Yq2VXjVBxB8>> [Accessed 18 June 2022].

Khatoon, S., Ibraheem, Singh, A., and Priti, 2014. Effects of various factors on electric load forecasting: An overview. *2014 6th IEEE Power India International Conference (PIICON)*,.

Kursa, M., Jankowski, A. and Rudnicki, W., 2010. Boruta – A System for Feature Selection. *Fundamenta Informaticae*, 101(4), pp.271-285.

Kuster, C., Rezgui, Y. and Mourshed, M., 2017. Electrical load forecasting models: A critical systematic review. *Sustainable Cities and Society*, 35, pp.257-270.

Mahdavi, N., Menhaj, M. and Barghinia, S., 2006. Short-Term Load Forecasting for Special Days Using Bayesian Neural Networks. *2006 IEEE PES Power Systems Conference and Exposition*.

McCulloch, W. and Pitts, W., 1943. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), pp.115-133.

Montgomery, D. and Peck, E., 1992. *Introduction to linear regression analysis*. New York: Wiley.

Office of Electricity, 2008. *The Smart Grid: An Introduction*. Litos Strategic Communication.

Open-power-system-data.org. 2022. *Open Power System Data – A platform for open data of the European power system*. [online] Available at: <<https://open-power-system-data.org/>> [Accessed 29 May 2022].

Peter I. Frazier. A Tutorial on Bayesian Optimization

Pfeifer, B., Baniecki, H., Saranti, A., Biecek, P. and Holzinger, A., 2022. Network Module Detection from Multi-Modal Node Features with a Greedy Decision Forest for Actionable Explainable AI.

Pfenninger, S., Hawkes, A. and Keirstead, J., 2014. Energy systems modeling for twenty-first century energy challenges. *Renewable and Sustainable Energy Reviews*, 33, pp.74-86.

PyPI. 2022. similaritymeasures. [online] Available at: <<https://pypi.org/project/similaritymeasures/>> [Accessed 21 June 2022].

Renewables.ninja. 2022. *Renewables.ninja*. [online] Available at: <<https://www.renewables.ninja/>> [Accessed 29 May 2022].

Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6), pp.386-408.

scikit-learn. 2022. 3.3. Metrics and scoring: quantifying the quality of predictions. [online] Available at: <[https://scikit-learn.org/stable/modules/model\\_evaluation.html#mean-absolute-percentage-error](https://scikit-learn.org/stable/modules/model_evaluation.html#mean-absolute-percentage-error)> [Accessed 17 June 2022].

scikit-learn. 2022. *sklearn.neural\_network.MLPRegressor*. [online] Available at: <[https://scikit-learn.org/stable/modules/generated/sklearn.neural\\_network.MLPRegressor.html](https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPRegressor.html)> [Accessed 18 August 2022].

Scikit-optimize.github.io. 2022. skopt.BayesSearchCV — scikit-optimize 0.8.1 documentation. [online] Available at: <<https://scikit-optimize.github.io/stable/modules/generated/skopt.BayesSearchCV.html>> [Accessed 15 June 2022].

Shumway, R. and Stoffer, D., 2011. *Time Series Analysis and Its Applications*. New York, NY: Springer Science+Business Media, LLC.

Siah Yap, K. and Jen Yap, H., 2012. Daily maximum load forecasting of consecutive national holidays using OSELM-based multi-agents system with weighted average strategy. *Neurocomputing*, 81, pp.108-112.

Song, K., Baek, Y., Hong, D. and Jang, G., 2005. Short-Term Load Forecasting for the Holidays Using Fuzzy Linear Regression Method. *IEEE Transactions on Power Systems*, 20(1), pp.96-101.

Stouffer, K., Pillitteri, V., Lightman, S., Abrams, M. and Hahn, A., 2015. Guide to Industrial Control Systems (ICS) Security.

*The European Green Deal.*

Tomoiağă, B., Chindriș, M., Sumper, A., Sudria-Andreu, A. and Villafafila-Robles, R., 2013. Pareto Optimal Reconfiguration of Power Distribution Systems Using a Genetic Algorithm Based on NSGA-II. *Energies*, 6(3), pp.1439-1455.

UNFCCC, 1997. *UNFCCC (1997) Kyoto Protocol to the United Nations Framework Convention on Climate Change adopted at COP3 in Kyoto, Japan, on 11 December 1997.* %20http://unfccc.int/resource/docs/cop3/07a01.pdf, Kyoto, Japan.

Vohra, K., Vodonos, A., Schwartz, J., Marais, E., Sulprizio, M. and Mickley, L., 2021. Global mortality from outdoor fine particle pollution generated by fossil fuel combustion: Results from GEOS-Chem. *Environmental Research*, 195, p.110754.

Web.archive.org. 2022. *List of common conversion factors (Engineering conversion factors) - IOR Energy PtyLtd.* [online] Available at: <<https://web.archive.org/web/20100825042309/http://www.ior.com.au/ecflist.html>> [Accessed 23 May 2022].

Wiley Series in Probability and Statistics, 1988. *Nonlinear Regression Analysis and Its Applications.*

www.eea.europa.eu. 2022. *Share of energy consumption from renewable sources in Europe.* [online] Available at: <<https://www.eea.europa.eu/ims/share-of-energy-consumption-from>> [Accessed 23 May 2022].

Xue, B. and Geng, J., 2012. Dynamic transverse correction method of middle and long-term energy forecasting based on statistic of forecasting errors. *2012 10th International Power & Energy Conference (IPEC),.*

ZENG, P., SHENG, C. and JIN, M., 2018. A learning framework based on weighted knowledge transfer for holiday load forecasting. *Journal of Modern Power Systems and Clean Energy*, 7(2), pp.329-339.

Zhu, K., Geng, J. and Wang, K., 2021. A hybrid prediction model based on pattern sequence-based matching method and extreme gradient boosting for holiday load forecasting. *Electric Power Systems Research*, 190, p.106841.

Βέττας, Ν., Danchev, S., Μανιάτης, Γ., Παρατσιώκας, Ν. and Βαλάσκας, Κ., 2021. Ο Τομέας Ενέργειας στην Ελλάδα: Τάσεις, Προοπτικές και Προκλήσεις. Ιδρύμα Οικονομικών και Βιομηχανικών Ερευνών.