



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Πρόβλεψη με χρήση γενικευμένων νευρωνικών δικτύων αξιοποιώντας μεγάλο όγκο δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΜΠΑΛΕΤΣΑΣ Γ. ΝΙΚΟΛΑΟΣ

Επιβλέπων: Βασίλειος Ασημακόπουλος
Ομότιμος Καθηγητής Ε.Μ.Π

Υπεύθυνος: Ευάγγελος Σπηλιώτης
Διδάκτωρ Ε.Μ.Π

Αθήνα, Ιούνιος 2024



Πρόβλεψη με χρήση γενικευμένων νευρωνικών δικτύων αξιοποιώντας μεγάλο όγκο δεδομένων

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΜΠΑΛΕΤΣΑΣ Γ. ΝΙΚΟΛΑΟΣ

Επιβλέπων: Βασίλειος Ασημακόπουλος
Ομότιμος Καθηγητής Ε.Μ.Π

Υπεύθυνος: Ευάγγελος Σπηλιώτης
Διδάκτωρ Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22α Ιουλίου 2024.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

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

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

.....
Ευάγγελος Μαρινάκης
Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούνιος 2024



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Copyright © - All rights reserved. Με την επιφύλαξη παντός δικαιώματος.
Νικόλαος Μπλέτσας, 2024.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....
Νικόλαος Μπλέτσας

29 Μαρτίου 2023

Περίληψη

Η παρούσα διπλωματική εργασία έχει ως στόχο να σχεδιάσει και να εκπαιδεύσει ένα νευρωνικό δίκτυο με χρήση μεγάλου όγκου δεδομένων το οποίο θα είναι σε θέση να προβλέπει χρονοσειρές οι οποίες αναφέρονται σε δεδομένα κατανάλωσης ενέργειας, καθιστώντας το κατάλληλο για γενικευμένη χρήση.

Τα τελευταία χρόνια, με τη ραγδαία βελτίωση των συστημάτων μηχανικής μάθησης και την ανάγκη για πρόβλεψη χρονοσειρών σε ενεργειακά δεδομένα, έχουν γίνει σημαντικές έρευνες χρησιμοποιώντας διάφορα είδη νευρωνικών δικτύων για βέλτιστη πρόβλεψη. Η δημιουργία των μοντέλων βασίστηκε στην καθολική εκπαίδευση και την αξιοποίηση μεγάλου όγκου δεδομένων. Αυτή η μέθοδος επιτρέπει την αξιοποίηση διαφορετικών συνόλων δεδομένων και προηγμένων τεχνικών μοντελοποίησης για τη δημιουργία προβλέψεων υψηλής ακρίβειας. Η αρχιτεκτονική που χρησιμοποιήθηκε είναι το πολυεπίπεδο τεχνητό νευρωνικό δίκτυο, το οποίο, παρά την απλότητά του, παρουσιάζει ικανοποιητικές προβλέψεις σε ενεργειακά δεδομένα.

Η παρούσα εργασία ξεκινά με μια θεωρητική ανάλυση των στατιστικών μεθόδων πρόβλεψης, εξετάζοντας επίσης τις διαφορετικές αρχιτεκτονικές νευρωνικών δικτύων που μπορούν να χρησιμοποιηθούν για την πρόβλεψη χρονοσειρών καθώς και μετρικών που θα εφαρμοστούν κατά τη διάρκεια της έρευνας. Στο πλαίσιο του πειραματικού σταδίου, προηγείται μια σημαντική φάση συλλογής και προεπεξεργασίας των δεδομένων, η οποία είναι ουσιώδης για τη διασφάλιση της κατάλληλης δομής της εισόδου στα νευρωνικά δίκτυα αλλά και της ποιότητας των δεδομένων. Έπειτα, προχωρούμε σε μια εμπειριστατωμένη ανάλυση όλων των δημιουργηθέντων μοντέλων και της αρχιτεκτονικής των νευρωνικών δικτύων τα οποία χρησιμοποιήθηκαν, με στόχο την επίτευξη προβλέψεων υψηλής ακρίβειας, ακολουθούμενη από μια λεπτομερή εξέταση των προκύπτοντων αποτελεσμάτων τόσο σε συνολικό βαθμό όσο και ανάλογα τον τύπο κτιρίου, το είδος ενέργειας και τον τύπο χρήσης ενέργειας.

Παρατηρήθηκε ότι η καθολική εκπαίδευση είναι σημαντική για την αξιοποίηση του μεγάλου συνόλου δεδομένων, ενώ η εστίαση των μοντέλων σε υποσύνολο δεδομένων βελτίωσε τα συνολικά αποτελέσματα. Τέλος, προτείνονται διάφορες μελλοντικές επεκτάσεις για τη βελτίωση της μελέτης και των αποτελεσμάτων.

Λέξεις Κλειδιά

Τεχνικές Πρόβλεψων, Πρόβλεψη Χρονοσειρών, Τεχνητά Νευρωνικά Δίκτυα, Καθολική Εκπαίδευση, Πολυστρωματικό Perceptron

Abstract

The present thesis aims to design and train a neural network using a large dataset that will be able to predict time series referring to energy consumption data, making it suitable for general use.

In recent years with the rapid improvement of machine learning systems as well as the need for time series prediction in energy data, significant research has been done using various kinds of neural networks in order to optimally predict the target energy timeseries. Due to the need to predict such time series and the ever-increasing amount of global data available for analysis, the creation of the models used was based on global training and the exploitation of a large amount of data. Global training of neural networks is an important training method as by using this method it is possible to take advantage of different data sets and advanced modeling techniques aiming to generate accurate predictions. The neural network architecture used during the thesis was that of the multilayer perceptron (MLP) which, despite its simplicity, has demonstrated satisfactory predictions in energy data.

This study begins with a theoretical analysis of statistical forecasting methods, also examining the different neural network architectures that can be used for time series prediction, as well as the metrics that will be applied during the research. In the experimental phase, a significant data collection and preprocessing stage precedes, which is essential for ensuring the proper structure of the input to the neural networks and the quality of the data. Then, we proceed with an in-depth analysis of all the created models and the architecture of the neural networks used, aiming to achieve high-accuracy predictions, followed by a detailed examination of the resulting outcomes both overall and according to building type, energy type, and energy usage type.

It was observed that global training is essential for the correct utilization of the large dataset, while models that focused on subsets of data improved the overall results. Finally, various future extensions are proposed to enhance the study and its outcomes.

Keywords

Forecasting Techniques, Time-series Forecasting, Artificial Neural Networks, Global Training, Multilayer Perceptron

στους γονείς μου

Ευχαριστίες

Θα ήθελα καταρχήν να ευχαριστήσω τον καθηγητή κ. Βασίλειο Ασημακόπουλο για την επίβλεψη αυτής της διπλωματικής εργασίας και για την ευκαιρία που μου έδωσε να εντυπώσω στον κλάδο των τεχνικών προβλέψεων και συγκεκριμένα στις προβλέψεις χρονοσειρών. Επίσης θα ήθελα να ευχαριστήσω τους καθηγητές κ. Χρυσόστομο Δούκα και κ. Ευάγγελο Μαρινάκη για την συμμετοχή τους στην επιτροπή εξέτασης της διπλωματικής εργασίας. Ιδιαίτερες ευχαριστίες θα ήθελα να δώσω στον Δρ. Ευάγγελο Σπηλιώτη για την καθοδήγησή του και την εξαιρετική συνεργασία που είχαμε. Τέλος θα ήθελα να ευχαριστήσω τους γονείς μου για την καθοδήγηση και την ηθική συμπαράσταση που μου προσέφεραν όλα αυτά τα χρόνια.

Αθήνα, Ιούνιος 2024

Νικόλαος Μπλέτσας

Περιεχόμενα

| | |
|--|-----------|
| Περίληψη | 1 |
| Abstract | 3 |
| Ευχαριστίες | 7 |
| 1 Εισαγωγή | 19 |
| 1.1 Αντικείμενο της διπλωματικής | 19 |
| 1.2 Βιβλιογραφική ανασκόπηση | 20 |
| 1.3 Οργάνωση της εργασίας | 23 |
| 2 Μέθοδοι πρόβλεψης χρονοσειρών | 25 |
| 2.1 Εισαγωγή στις χρονοσειρές | 25 |
| 2.2 Ποιοτικά χαρακτηριστικά χρονοσειρών | 25 |
| 2.2.1 Στασιμότητα | 26 |
| 2.2.2 Τάση | 26 |
| 2.2.3 Εποχικότητα | 27 |
| 2.2.4 Κυκλικότητα | 27 |
| 2.2.5 Ασυνέχειες | 28 |
| 2.2.6 Τυχαιότητα | 29 |
| 2.3 Μέθοδοι πρόβλεψης χρονοσειρών | 30 |
| 2.3.1 Αφελής μέθοδος | 30 |
| 2.3.2 Μέθοδος Εκθετικής Εξομάλυνσης | 31 |
| 2.3.3 Μοντέλα ARIMA | 31 |
| 2.3.4 Μοντέλα Γραμμικής και Εκθετικής Τάσης | 32 |
| 2.3.5 Μοντέλα Πολλαπλής Παλινδρόμησης | 32 |
| 2.3.6 Μέθοδος Theta | 33 |
| 2.3.7 Νευρωνικά Δίκτυα | 33 |
| 2.3.8 Συνδυασμός Μεθόδων | 33 |
| 2.4 Μέτρα ακρίβειας προβλέψεων | 34 |
| 2.4.1 Μέσο σφάλμα | 34 |
| 2.4.2 Μέσο απόλυτο σφάλμα | 34 |
| 2.4.3 Μέσο τετραγωνικό σφάλμα | 34 |
| 2.4.4 Ρίζα μέσου τετραγωνικού σφάλματος | 35 |
| 2.4.5 Μέσο Απόλυτο Κανονικοποιημένο Σφάλμα | 35 |
| 2.4.6 Κανονικοποιημένο Μέσο Τετραγωνικό Σφάλμα | 35 |

| | | |
|----------|---|-----------|
| 2.4.7 | Ρίζα Μέσου Τετραγωνικού Κανονικοποιημένου Σφάλματος | 36 |
| 3 | Χρήση νευρωνικών δικτύων για προβλέψεις | 37 |
| 3.1 | Νευρωνικά δίκτυα | 37 |
| 3.2 | Συνάρτηση ενεργοποίησης | 39 |
| 3.2.1 | Linear | 39 |
| 3.2.2 | ReLU | 39 |
| 3.2.3 | Sigmoid | 40 |
| 3.2.4 | tanh | 40 |
| 3.3 | Κατηγορίες νευρωνικών δικτύων | 40 |
| 3.3.1 | Multilayer Perceptrons (MLPs) | 41 |
| 3.3.2 | Επαναλαμβανόμενο νευρωνικό δίκτυο | 42 |
| 3.3.3 | Συνελκτικά δίκτυα | 43 |
| 3.4 | Αρχιτεκτονική νευρωνικών δικτύων | 43 |
| 3.5 | Εκπαίδευση νευρωνικών δικτύων | 44 |
| 3.6 | Είσοδος και έξοδος νευρωνικών δικτύων για πρόβλεψη χρονοσειρών | 45 |
| 3.6.1 | Επίπεδο εισόδου | 45 |
| 3.6.2 | Επίπεδο εξόδου | 45 |
| 3.7 | Εκπαίδευση συνόλου δεδομένων με σκοπό την χρήση τους από νευρωνικά δίκτυα | 45 |
| 3.8 | Καθολική και τοπική εκπαίδευση | 47 |
| 3.8.1 | Τοπική εκπαίδευση | 47 |
| 3.8.2 | Καθολική εκπαίδευση | 47 |
| 3.9 | Εισαγωγή εξωγενών μεταβλητών στα νευρωνικά δίκτυα | 48 |
| 4 | Συλλογή και επεξεργασία δεδομένων | 51 |
| 4.1 | Συλλογή δεδομένων | 51 |
| 4.2 | Παρουσίαση δεδομένων | 52 |
| 4.2.1 | Περιγραφή μεταβλητών συνόλου δεδομένων | 52 |
| 4.2.2 | Οπτικοποίηση δεδομένων | 54 |
| 4.3 | Προεπεξεργασία δεδομένων | 58 |
| 4.3.1 | Αρχική δομή δεδομένων | 58 |
| 4.3.2 | Επεξεργασία δεδομένων με σκοπό την εξαγωγή χαρακτηριστικών | 59 |
| 4.3.3 | Επεξεργασία δεδομένων με σκοπό την βελτίωση της ποιότητας τους | 60 |
| 4.3.4 | Επεξεργασία λόγω όγκου των δεδομένων | 61 |
| 4.3.5 | Τελική δομή δεδομένων | 62 |
| 5 | Εφαρμογή μοντέλων πρόβλεψης και αποτελέσματα | 63 |
| 5.1 | Διαχωρισμός δεδομένων | 63 |
| 5.2 | Κανονικοποίηση δεδομένων | 64 |
| 5.3 | Μοντέλο Βάσης | 65 |
| 5.4 | Καθολικό Μοντέλο | 67 |
| 5.5 | Καθολικά Μοντέλα με εξειδίκευση σε υποσύνολο των δεδομένων | 70 |
| 5.5.1 | Περιγραφή μοντέλων | 70 |
| 5.5.2 | Περιγραφή και αποτελέσματα μοντέλων ανά τύπο κτιρίου | 70 |

| | |
|---|-----------|
| 5.5.3 Περιγραφή και αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας . . . | 71 |
| 5.5.4 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας | 73 |
| 5.5.5 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου | 74 |
| 5.5.6 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας | 75 |
| 5.6 Μοντέλα με επίπεδο ενσωμάτωσης | 77 |
| 5.6.1 Περιγραφή μοντέλων με επίπεδο ενσωμάτωσης | 77 |
| 5.6.2 Περιγραφή και αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων | 78 |
| 5.6.3 Περιγραφή και αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου | 80 |
| 5.6.4 Περιγραφή και αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας | 81 |
| 5.6.5 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας | 82 |
| 5.6.6 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου | 84 |
| 5.7 Σύγκριση και Ανάλυση Αποτελεσμάτων | 85 |
| 5.7.1 Μετρική Αξιολόγησης Αποτελεσμάτων | 85 |
| 5.7.2 Σύγκριση αποτελεσμάτων | 86 |
| 6 Συμπεράσματα και προεκτάσεις | 91 |
| 6.1 Συμπεράσματα | 91 |
| 6.2 Μελλοντικές Επεκτάσεις | 93 |
| Βιβλιογραφία | 99 |

Κατάλογος Σχημάτων

| | | |
|------|--|----|
| 2.1 | Παράδειγμα στάσιμης χρονοσειράς [1] | 26 |
| 2.2 | Παράδειγμα χρονοσειράς με τάση [2] | 26 |
| 2.3 | Παράδειγμα χρονοσειράς με εποχικότητα [3] | 27 |
| 2.4 | Παράδειγμα χρονοσειράς με κυκλικότητα [4] | 28 |
| 2.5 | Παράδειγμα χρονοσειράς με outlier [5] | 28 |
| 2.6 | Παράδειγμα χρονοσειράς με αλλαγή επιπέδου [6] | 29 |
| 2.7 | Παράδειγμα χρονοσειράς με τυχαιότητα [7] | 30 |
| 3.1 | Τεχνητό νευρωνικό δίκτυο [8] | 38 |
| 3.2 | Τεχνητός νευρώνας [9] | 39 |
| 3.3 | Συναρτήσεις ενεργοποίησης [10] | 40 |
| 3.4 | Αρχιτεκτονική MLP [11] | 42 |
| 3.5 | Αρχιτεκτονική RNN [12] | 43 |
| 3.6 | Αρχιτεκτονική CNN [13] | 43 |
| 3.7 | Αρχιτεκτονική MLP με προσθήκη embedding layer [14] | 49 |
| 4.1 | Νοσοκομείο στην Αλάσκα, κατανάλωση ηλεκτρισμού για ψύξη | 55 |
| 4.2 | Νοσοκομείο στην Αλαμπάμα, κατανάλωση ηλεκτρισμού για ψύξη | 55 |
| 4.3 | Μικρό γραφείο στην Μινεσότα, συνολική κατανάλωση ηλεκτρισμού | 55 |
| 4.4 | Νοσοκομείο στην Μινεσότα, συνολική κατανάλωση αερίου | 55 |
| 4.5 | Μεγάλο ξενοδοχείο στην Φλόριντα, κατανάλωση αερίου για θέρμανση | 56 |
| 4.6 | Μεγάλο ξενοδοχείο στην Φλόριντα, κατανάλωση ηλεκτρισμού για θέρμανση | 56 |
| 4.7 | Εστιατόριο στην Νέα Υόρκη, συνολική κατανάλωση αερίου | 56 |
| 4.8 | Εστιατόριο στην Νέα Υόρκη, κατανάλωση αερίου για θέρμανση | 56 |
| 4.9 | Εβδομάδα καλοκαιριού για συνολική κατανάλωση αερίου | 57 |
| 4.10 | Εβδομάδα χειμώνα για συνολική κατανάλωση αερίου | 57 |
| 4.11 | Εβδομάδα καλοκαιριού για κατανάλωση ηλεκτρισμού για ψύξη | 57 |
| 4.12 | Εβδομάδα χειμώνα για κατανάλωση ηλεκτρισμού για ψύξη | 57 |
| 4.13 | Εβδομάδα καλοκαιριού για κατανάλωση αερίου για θέρμανση | 58 |
| 4.14 | Εβδομάδα χειμώνα για κατανάλωση αερίου για θέρμανση | 58 |
| 4.15 | Αρχική δομή δεδομένων | 58 |
| 4.16 | Αρχική μορφή αρχείων CSV | 59 |
| 4.17 | Τελική δομή δεδομένων | 62 |
| 4.18 | Τελική μορφή αρχείων csv | 62 |
| 5.1 | Χάρτης δεδομένων train και test | 64 |

| | | |
|-----|---|----|
| 5.2 | Εστιατόριο στην Νέα Υόρκη, κατανάλωση ηλεκτρισμού για θέρμανση | 66 |
| 5.3 | Εστιατόριο στην Νέα Υόρκη, κατανάλωση φυσικού αερίου για θέρμανση | 66 |
| 5.4 | Αρχιτεκτονική καθολικού μοντέλου | 68 |
| 5.5 | Αρχιτεκτονική μοντέλων με επίπεδο ενσωμάτωσης | 78 |
| 5.6 | Σύγκριση αποτελεσμάτων με βάση τον τύπο κτιρίου | 87 |
| 5.7 | Σύγκριση αποτελεσμάτων με βάση το είδος ενέργειας | 88 |
| 5.8 | Σύγκριση αποτελεσμάτων με βάση τον τύπο χρήσης ενέργειας | 89 |

Κατάλογος Πινάκων

| | | |
|------|--|----|
| 4.1 | Χαρακτηριστικά και περιγραφή του συνόλου δεδομένων | 52 |
| 5.1 | Συνολικά αποτελέσματα Naive | 66 |
| 5.2 | Αποτελέσματα Naive ανά τύπο κτιρίου | 66 |
| 5.3 | Αποτελέσματα Naive ανά τύπο χρήσης ενέργειας | 66 |
| 5.4 | Αποτελέσματα Naive ανά είδος ενέργειας | 67 |
| 5.5 | Συνολικά αποτελέσματα καθολικού μοντέλου | 69 |
| 5.6 | Αποτελέσματα καθολικού μοντέλου ανά τύπο κτιρίου | 69 |
| 5.7 | Αποτελέσματα καθολικού μοντέλου ανά τύπο χρήσης ενέργειας | 69 |
| 5.8 | Αποτελέσματα καθολικού μοντέλου ανά είδος ενέργειας | 69 |
| 5.9 | Συνολικά αποτελέσματα μοντέλων ανά τύπο κτιρίου | 71 |
| 5.10 | Αποτελέσματα μοντέλων ανά τύπο κτιρίου ως προς τον τύπο κτιρίου | 71 |
| 5.11 | Αποτελέσματα μοντέλων ανά τύπο κτιρίου ως προς τον τύπο χρήσης ενέργειας | 71 |
| 5.12 | Αποτελέσματα μοντέλων ανά τύπο κτιρίου ως προς το είδος ενέργειας | 71 |
| 5.13 | Συνολικά αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας | 72 |
| 5.14 | Αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας ως προς τον τύπο κτιρίου | 72 |
| 5.15 | Αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας ως προς τον τύπο χρήσης ενέργειας | 72 |
| 5.16 | Αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας ως προς το είδος ενέργειας | 73 |
| 5.17 | Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας | 73 |
| 5.18 | Αποτελέσματα μοντέλων ανά είδος ενέργειας ως προς τον τύπο κτιρίου | 73 |
| 5.19 | Αποτελέσματα μοντέλων ανά είδος ενέργειας ως προς τον τύπο χρήσης ενέργειας | 74 |
| 5.20 | Αποτελέσματα μοντέλων ανά είδος ενέργειας ως προς το είδος ενέργειας | 74 |
| 5.21 | Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου | 74 |
| 5.22 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου ως προς τον τύπο κτιρίου | 75 |
| 5.23 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου ως προς τον τύπο χρήσης ενέργειας | 75 |
| 5.24 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου ως προς το είδος ενέργειας | 75 |
| 5.25 | Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας | 76 |
| 5.26 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και και τύπο χρήσης ενέργειας ως προς τον τύπο κτιρίου | 76 |

| | | |
|------|---|----|
| 5.27 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας ως προς τον τύπο χρήσης ενέργειας | 76 |
| 5.28 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και και τύπο χρήσης ενέργειας ως προς το είδος ενέργειας | 77 |
| 5.29 | Συνολικά αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων | 79 |
| 5.30 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων ως προς τον τύπο κτιρίου | 79 |
| 5.31 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων ως προς τον τύπο χρήσης ενέργειας | 79 |
| 5.32 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων ως προς το είδος ενέργειας | 79 |
| 5.33 | Συνολικά αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου | 80 |
| 5.34 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς τον τύπο κτιρίου | 80 |
| 5.35 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς τον τύπο χρήσης ενέργειας | 81 |
| 5.36 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς το είδος ενέργειας | 81 |
| 5.37 | Συνολικά αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας | 81 |
| 5.38 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς τον τύπο κτιρίου | 82 |
| 5.39 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς τον τύπο χρήσης ενέργειας | 82 |
| 5.40 | Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς το είδος ενέργειας | 82 |
| 5.41 | Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας | 83 |
| 5.42 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς τον τύπο κτιρίου | 83 |
| 5.43 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς τον τύπο χρήσης ενέργειας | 83 |
| 5.44 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς το είδος ενέργειας | 84 |
| 5.45 | Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου | 84 |
| 5.46 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς τον τύπο κτιρίου | 84 |
| 5.47 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς τον τύπο χρήσης ενέργειας | 85 |

| | | |
|------|--|----|
| 5.48 | Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς το είδος ενέργειας | 85 |
| 5.49 | Σύγκριση συνολικών αποτελεσμάτων και αριθμού μοντέλων | 86 |

Εισαγωγή

1.1 Αντικείμενο της διπλωματικής

Η πρόβλεψη ενεργειακής κατανάλωσης αποτελεί έναν κρίσιμο τομέα, καθώς η ακριβής εκτίμηση της μελλοντικής ζήτησης ενέργειας είναι απαραίτητη για τη διασφάλιση της αποτελεσματικής διαχείρισης πόρων, τη μείωση του κόστους και την ελαχιστοποίηση των περιβαλλοντικών επιπτώσεων. Ειδικότερα στην εποχή μας, λόγω της κλιματικής αλλαγής και άλλων περιβαλλοντολογικών προβλημάτων τα οποία παρατηρούνται, η πρόβλεψη ενεργειακών δεδομένων αποκτά ιδιαίτερη σημασία. Επιπλέον, στις μέρες μας, ολοένα και περισσότερα ενεργειακά δεδομένα είναι διαθέσιμα προς χρήση, οπότε η δημιουργία μοντέλων που θα είναι σε θέση να χρησιμοποιήσουν και να εκμεταλλευτούν όλη αυτή την πληροφορία είναι σημαντική. Τέλος, λόγω της ραγδαίας ανάπτυξης που παρατηρείται στα συστήματα μηχανικής μάθησης και κατά συνέπεια στα νευρωνικά δίκτυα, η δημιουργία και η χρήση μοντέλων που αξιοποιούν αυτές τις αρχιτεκτονικές μοντέλων είναι εξίσου σημαντική, καθώς έτσι θα μπορέσει να επιτευχθεί η βελτίωση των προβλέψεων και η απομάκρυνση από την χρήση παραδοσιακών στατιστικών μεθόδων πρόβλεψης που χρησιμοποιούνται ευρέως έως σήμερα.

Ο σκοπός αυτής της διπλωματικής εργασίας είναι η ανάπτυξη μοντέλων πρόβλεψης μέσω καθολικής εκπαίδευσης, τα οποία θα μπορούν να παράγουν ακριβείς προβλέψεις ενεργειακής κατανάλωσης. Η προσπάθεια αυτή περιλαμβάνει την εκπαίδευση νευρωνικών δικτύων σε ένα μεγάλο και ποικιλόμορφο σύνολο δεδομένων, με στόχο τη βελτίωση της γενίκευσης των μοντέλων και της ικανότητάς τους να παρέχουν αξιόπιστες προβλέψεις σε δεδομένα που αφορούν την κατανάλωση ενέργειας σε διάφορους τύπους κτιρίου. Ακόμη, σκοπός της διπλωματικής είναι η ανάδειξη της σημαντικότητας της καθολικής εκπαίδευσης των μοντέλων καθώς και την προσθήκη πληροφορίας σε αυτά μέσω την εστίαση τους σε ένα συγκεκριμένο υποσύνολο των δεδομένων. Η καθολική εκπαίδευση επιτρέπει στα μοντέλα να εκπαιδευτούν σε ευρύ φάσμα δεδομένων, εξασφαλίζοντας έτσι μια γενική κατανόηση των μοτίβων και των τάσεων στην ενεργειακή κατανάλωση. Παράλληλα, η εξειδίκευση σε συγκεκριμένα υποσύνολα δεδομένων βελτιώνει την ακρίβεια των προβλέψεων, προσαρμόζοντας τα μοντέλα στις ιδιαιτερότητες και τις μοναδικές συνθήκες κάθε επιμέρους σεναρίου. Αυτή η συνδυαστική προσέγγιση αυξάνει την αξιοπιστία και την αποδοτικότητα των προβλέψεων, ενισχύοντας τη λήψη τεκμηριωμένων αποφάσεων στον ενεργειακό τομέα.

Η παρούσα διπλωματική εργασία εστιάζει στην πρόβλεψη χρονοσειρών σε ενεργειακά δεδομένα χρησιμοποιώντας μεγάλο όγκο δεδομένων. Η προεπεξεργασία των δεδομένων είναι

ένα βασικό κομμάτι της διπλωματικής εργασίας, καθώς διασφαλίζει ότι τα δεδομένα είναι κατάλληλα διαμορφωμένα για να αποτελέσουν την είσοδο στα νευρωνικά δίκτυα. Η διαδικασία αυτή περιλαμβάνει διάφορα στάδια, όπως ο καθαρισμός των δεδομένων για την απομάκρυνση σφαλμάτων και ελλείψεων, η κανονικοποίηση ή τυποποίηση των δεδομένων ώστε να έχουν συγκρίσιμες κλίμακες, και η μετατροπή των δεδομένων σε μορφή που μπορεί να χρησιμοποιηθεί αποτελεσματικά από τα μοντέλα μηχανικής μάθησης. Έπειτα η διπλωματική εστιάζει στην δημιουργία μοντέλων με σκοπό την κατάλληλη πρόβλεψη των ενεργειακών χρονοσειρών. Η δημιουργία αυτών των προβλέψεων γίνεται αξιοποιώντας μοντέλα νευρωνικών δικτύων τα οποία κάνουν χρήση της καθολικής εκπαίδευσης. Η καθολική εκπαίδευση είναι μια διαδεδομένη μέθοδος εκπαίδευσης των νευρωνικών μοντέλων μέσω της οποίας μπορεί να γίνει η εκμετάλλευση μεγάλου όγκου δεδομένων και η δημιουργία ακριβών προβλέψεων.

1.2 Βιβλιογραφική ανασκόπηση

Η κοινωνία μας εξαρτάται σημαντικά από την ενέργεια, η οποία αποτελεί τη βάση κάθε μορφής οικονομικής δραστηριότητας. Επομένως, οι διακυμάνσεις στην προσφορά μπορεί να έχουν σημαντικές και ευρείες οικονομικές επιπτώσεις. Αυτό δίνει κίνητρο στις κυβερνήσεις να εξασφαλίσουν ότι ασφαλείς και αξιόπιστες πηγές ενέργειας είναι ευρέως προσβάσιμες. Τα συστήματα που χρησιμοποιούν ανανεώσιμες πηγές ενέργειας είναι ιδανικά τοποθετημένα για να μειώσουν τον κίνδυνο διακοπής της προσφοράς ενέργειας και την τρέχουσα εξάρτηση από εισαγόμενα καύσιμα από πολλές χώρες. Η παραγωγή θερμότητας, ενέργειας και καυσίμων μεταφορών μπορεί να πραγματοποιηθεί με τη βοήθεια ευρέως διασπαρμένων ανανεώσιμων πηγών ενέργειας. Έτσι για τους παραπάνω λόγους, μπορεί να φανεί η σημαντικότητα και η ανάγκη πρόβλεψης χρονοσειρών για δεδομένα που αφορούν την ενεργειακή κατανάλωση.

Λόγω αυτής της σημαντικότητας, έχουν πραγματοποιηθεί διάφορες έρευνες σχετικά με την πρόβλεψη της ενεργειακής κατανάλωσης. Οι μελέτες αυτές εξετάζουν νέες τεχνολογίες και μεθόδους, όπως τα μεγάλα δεδομένα, το Διαδίκτυο των Πραγμάτων (IoT) και τα μοντέλα μηχανικής μάθησης, προκειμένου να βελτιωθεί η ακρίβεια και η αποδοτικότητα των προβλέψεων. Η συνεχιζόμενη έρευνα σε αυτό το πεδίο είναι κρίσιμη για την αντιμετώπιση των προκλήσεων που σχετίζονται με την ενεργειακή κατανάλωση και τη διαχείριση των ενεργειακών πόρων. Ακόμη πολύ σημαντική έρευνα γίνεται λόγω της βελτίωσης των μοντέλων μηχανικής μάθησης, έρευνες που εστιάζουν στην εκπαίδευση των μοντέλων, καθολική εκπαίδευση ή τοπική ή ακόμη και ένας συνδυασμός των δύο αυτών μεθόδων, στην κατηγορία νευρωνικών δικτύων που θα κάνουν χρήση καθώς και στην αρχιτεκτονική του εκάστοτε νευρωνικού μοντέλου. Μερικές σχετικές έρευνες και τα αποτελέσματά τους παρουσιάζονται παρακάτω.

Αρχικά όσον αφορά την εκπαίδευση των μοντέλων με σκοπό την πρόβλεψη χρονοσειρών έχουν υπάρξει αρκετές έρευνες. Οι Verdes et al. [15] στην έρευνα τους σύγκριναν την χρήση καθολικής και τοπικής εκπαίδευσης σε πρόβλεψη απρόβλεπτων χρονοσειρών. Με τον όρο απρόβλεπτες χρονοσειρές εννοούνται χρονοσειρές των οποίων η συμπεριφορά χαρακτηρίζεται από ευαίσθητη εξάρτηση από τις αρχικές συνθήκες, όπου μικρές διαφορές στις αρχικές τιμές μπορούν να οδηγήσουν σε δραματικά διαφορετικές εξελίξεις. Στην έρευνα αυτή κατέληξαν πως με περιορισμένα δεδομένα τα αποτελέσματα των μοντέλων με καθολική εκπαίδευση είναι

αρκετά καλύτερα, ενώ με την αύξηση του μεγέθους του συνόλου δεδομένων τα αποτελέσματα γίνονται συγκρίσιμα. Ακόμη μια έρευνα που αναλύει τις διαφορές της τοπικής εναντίον της καθολικής εκπαίδευσης είναι η έρευνα των Montero-Manso et al. [16]. Συγκεκριμένα στην έρευνα τους κατέληξαν στα εξής βασικά συμπεράσματα. Οι καθολικές και τοπικές μέθοδοι μπορούν να παράγουν τις ίδιες προβλέψεις χωρίς καμία υπόθεση σχετικά με την ομοιότητα των σειρών στο σύνολο. Η πολυπλοκότητα των τοπικών μεθόδων αυξάνεται με το μέγεθος του συνόλου, ενώ παραμένει σταθερή για τις καθολικές μεθόδους. Αυτό το αποτέλεσμα υποστηρίζει τα πρόσφατα ευρήματα και παρέχει αρχές για το σχεδιασμό νέων αλγορίθμων. Και τέλος σε μια εκτενή εμπειρική μελέτη που έκαναν, δείχνουν ότι σκόπιμα απλοϊκοί αλγόριθμοι που πηγάζουν από αυτές τις αρχές εμφανίζουν εξαιρετική ακρίβεια. Συγκεκριμένα, τα καθολικά γραμμικά μοντέλα παρέχουν ανταγωνιστική ακρίβεια με πολύ λιγότερες παραμέτρους από τις πιο απλές τοπικές μεθόδους.

Οι Sousa et al. [17] στην έρευνα τους προσπάθησαν να κάνουν μια ανασκόπηση των καθολικών και τοπικών προσεγγίσεων πρόβλεψης χρονοσειρών, συμπεριλαμβανομένων των μοντέλων βαθιάς μάθησης (DL) που συνδυάζουν καθολικά και τοπικά σήματα, και των μεθόδων δυναμικής ενισχυτικής μάθησης (RL) που βελτιστοποιούν αυτό το συμβιβασμό. Τα κύρια ευρήματα τους δείχνουν ότι τα καθολικά μοντέλα επιτυγχάνουν ισχυρή εκφραστικότητα καταγράφοντας πολύπλοκα μοτίβα ενώ παράλληλα επιτρέπουν πιθανολογικά αποτελέσματα, και οι μέθοδοι ενισχυτικής μάθησης μειριάζουν τους κινδύνους κατά την πρόβλεψη κάθε βήματος για κάθε σειρά. Ακόμη οι Patra et al. [18] στην έρευνα τους προσπάθησαν να συγκρίνουν τα καθολικά και τοπικά μοντέλα με χρήση μοντέλων LSTM πάνω σε δεδομένα τα οποία αφορούν περιφερειακά υπόγεια ύδατα. Στην μελέτη τους προέκυψαν στο συμπέρασμα ότι το τοπικό μοντέλο LSTM δημιουργεί πρόβλεψης υψηλής ακρίβειας σε συνδυασμό με υψηλή υπολογιστική αποδοτικότητα. Σύμφωνα με την πρόσφατη έρευνα των Montero-Manso και Hyndman [16] οποιαδήποτε τοπική μέθοδος που εφαρμόζεται σε ένα σύνολο δεδομένων πολλών χρονοσειρών μπορεί να προσεγγιστεί από ένα καθολικό μοντέλο με επαρκή πολυπλοκότητα, ανεξάρτητα από τη συνάφεια της υποκείμενης χρονοσειράς. Έτσι οι Hewamalage et al. [19] προσπάθησαν να εξερευνήσουν τα οφέλη της καθολικής εναντίον της τοπικής εκπαίδευσης στην έρευνα τους. Κατέληξαν στο γεγονός ότι όσον αφορά τη διαθεσιμότητα των δεδομένων όλα τα καθολικά μοντέλα βελτιώνονται σταδιακά καθώς αυξάνονται οι διαστάσεις και ο αριθμός των χρονοσειρών στο σύνολο δεδομένων. Για τα τοπικά μοντέλα, αυτή η βελτίωση παρατηρείται μόνο με την μεταβολή του μήκους της κάθε μεμονωμένης χρονοσειράς. Η πολυπλοκότητα των τοπικών μοντέλων αυξάνεται ανάλογα με τον αριθμό των χρονοσειρών στο σύνολο δεδομένων, ενδεχομένως ακόμα και υψηλότερα από τη σταθερή πολυπλοκότητα ενός καθολικού μοντέλου που βασίζεται στο ίδιο σύνολο δεδομένων. Γι' αυτό, ακόμα κι αν η προσαρμογή ενός μόνο τοπικού μοντέλου σε μία μόνο χρονοσειρά μπορεί να απαιτεί πολύ λίγο χρόνο, η προσαρμογή πολλών από αυτά σε ένα ολόκληρο σύνολο δεδομένων χρονοσειρών απαιτεί σημαντικό χρονικό διάστημα. Για αυτούς τους λόγους έγινε η επιλογή της χρήσης καθολικής εκπαίδευσης καθόλη την διάρκεια του πειράματος, καθώς από διάφορες άλλες έρευνες είναι εύκολο κανείς να καταλάβει τα πολλαπλά οφέλη αυτού του τρόπου εκπαίδευσης.

Πέρα από τις διάφορες έρευνες που έχουν γίνει στο κομμάτι της εκπαίδευσης των μοντέλων και στις διαφορές μεταξύ καθολικής και τοπικής μεθόδου για πρόβλεψη χρονοσειρών,

έχουν ακόμη δημιουργηθεί πολλές έρευνες που επικεντρώνονται στην πρόβλεψη χρονοσειρών σε ενεργειακά δεδομένα μέσω διαφόρων τεχνικών μηχανικής μάθησης. Οι Wang et al. [20] στην έρευνα τους έκαναν μια ανασκόπηση της χρήσης βαθιάς μηχανικής μάθησης σε δεδομένα ανανεώσιμων πηγών ενέργειας παραθέτοντας τα πολλαπλά οφέλη που συναντώνται στην χρήση τέτοιων συστημάτων. Ακόμη οι Eddaoudi et al. [21] στην δική τους έρευνα έκαναν μια ανασκόπηση της χρήσης συστημάτων μηχανικής μάθησης για την αποτελεσματική πρόβλεψη χρονοσειρών που αφορούν την κατανάλωση ενέργειας. Και στις δύο αυτές έρευνες αναδείχθηκε η σημαντικότητα της χρήσης τέτοιων συστημάτων αντί κάποιων πιο παραδοσιακών στατιστικών μεθόδων καθώς έτσι μπορεί να γίνει εκμετάλλευση ενός μεγάλου όγκου δεδομένων, να γίνει η κατάλληλη επιλογή μοντέλου και η προσαρμογή του ανάλογα με το κάθε πρόβλημα που αντιμετωπίζει, με αποτέλεσμα οι προβλέψεις και τα συνολικά αποτελέσματα να έχουν μεγάλη ακρίβεια.

Οι Eddaoudi et al. [21] στην έρευνα τους σχετικά με την χρήση των μοντέλων μηχανικής μάθησης για την πρόβλεψη της κατανάλωσης ενέργειας ανέδειξαν την σημαντικότητα της χρήσης συστημάτων μηχανικής μάθησης. Συγκεκριμένα κατέληξαν στο συμπέρασμα ότι η βιομηχανία της κτιριακής ενέργειας είναι κατάλληλη για σύγχρονες τεχνολογίες όπως τα Big Data (BD) και το Internet of Things (IoT), τα οποία διαχειρίζονται τεράστια όγκο δεδομένων από αισθητήρες και μετρητές ενέργειας. Τα τελευταία δέκα χρόνια παρατηρείται αύξηση στην εφαρμογή μοντέλων μηχανικής μάθησης. Υπήρξε σημαντική αύξηση στη χρήση υβριδικών και συνόλου μεθόδων εκτός από τις παραδοσιακές τεχνικές μηχανικής μάθησης όπως τα Τεχνητά Νευρωνικά Δίκτυα (ANNs), τα Πολυεπίπεδα Τεχνητά Νευρωνικά Δίκτυα (MLPs) και οι Μηχανές Διανουσμάτων Υποστήριξης (SVMs).

Μια πολύ διαδεδομένη κατηγορία μοντέλων μηχανικής μάθησης είναι το πολυεπίπεδο τεχνητό νευρωνικό δίκτυο (MLP). Είναι μια μέθοδος που έχει επίσης απασχολήσει πολύ το ερευνητικό κοινό και υπάρχουν διάφορες έρευνες για τα οφέλη και την χρήση του όσον αφορά την πρόβλεψη χρονοσειρών. Οι Huang et al. [22] βασίστηκαν στην αρχιτεκτονική του MLP και δημιούργησαν έναν μετασχηματιστή με χρονική εξωτερική προσοχή το οποίο ονόμασαν MEAformer, ο οποίος υιοθετεί μια πλήρως συνδεδεμένη αρχιτεκτονική MLP που αποτελείται από δύο τύπους επιπέδων, επίπεδα κωδικοποιητή (encoder) και επίπεδα MLP. Δημιουργώντας αυτό το μοντέλο μπόρεσαν να επιτύχουν υψηλή ακρίβεια στην πρόβλεψη χρονοσειρών. Παρουσίασε αποδοτική γραμμική πολυπλοκότητα, ειδικά όταν επεξεργάζεται μακροπρόθεσμες εισόδους ακολουθίας. Επιπλέον, λόγω των επιπέδων MLP ενισχύθηκε η αποδοτικότητα στον χρόνο και τη μνήμη. Ακόμη οι Cordeiro-Costas et al. [23] στην έρευνα προσπάθησαν να δημιουργήσουν ένα προγνωστικό μοντέλο διαχείρισης ενέργειας κτιρίων. Έκαναν χρήση ενός υβριδικού μοντέλου το οποίο συνδυάζει τα LSTM και MLP μοντέλα. Η διαχείριση έχει πραγματοποιηθεί σε διαφορετικές συνθήκες χρήσης του κτιρίου, όπως αργίες και εργάσιμες και δύο διαφορετικές εποχές, η άνοιξη και ο χειμώνας, έχουν επίσης ληφθεί υπόψη για να καταδειχθεί η δυνατότητα του μοντέλου σε αντίξοες συνθήκες. Τα αποτελέσματα που προέκυψαν ήταν πολύ θετικά μιας και με την χρήση αυτού του προγνωστικού μοντέλου το ενεργειακό ισοζύγιο του κτιρίου βελτιώνεται, αυξάνοντας τον συντελεστή αυτάρκειας σε όλες τις περιπτώσεις όπου υπάρχει παραγωγή ηλιακής ενέργειας. Επιπλέον, το οικονομικό όφελος είναι παρόν σε όλες τις περιπτώσεις, ανεξαρτήτως εάν υπάρχει υπέρβαση παραγωγής.

Οι Jiang et al. [24] από την πλευρά τους, στην έρευνα τους χρησιμοποίησαν πραγ-

ματικά σύνολα δεδομένων ενεργειακής κατανάλωσης γραφείων για την εκπαίδευση και την δοκιμή του μοντέλου τους. Συνδύασαν τις αρχιτεκτονικές κωδικοποιητή (encoder) και αποκωδικοποιητή (decoder) για τη δημιουργία βαθιών νευρωνικών δικτύων. Ακόμη συνδύασαν την αποσύνθεση TDC και τα βαθιά νευρωνικά δίκτυα για την εκπαίδευση ενός γενικευμένου μοντέλου. Τα αποτελέσματα τους είχαν μεγάλη ακρίβεια, το οποίο υποδηλώνει ότι η αποσύνθεση TDC και η εκπαίδευση μέσω πολλαπλών πηγών ενισχύουν την ικανότητα πρόβλεψης του μοντέλου.

Οι Zheng et al. [25] στην μελέτη τους πρότειναν το DECPR, μια ερμηνεύσιμη μέθοδο αποσύνθεσης, που εξηγεί τις αποσυνθεμένες υποακολουθίες. Ακόμη πρότειναν την χρήση του TFT, μια προσέγγιση βαθιάς μάθησης με ερμηνευσιμότητα, η οποία προβλέπει την ενεργειακή κατανάλωση κάθε ώρας για την επόμενη ημέρα. Η ερμηνευσιμότητα του μοντέλου TFT χρησιμοποιείται για να αναλυθούν οι επιδράσεις των εξωγενών μεταβλητών στην ενεργειακή κατανάλωση. Με την δημιουργία αυτού του μοντέλου μπόρεσαν να έχουν βελτιωμένα αποτελέσματα όσον αφορά την πρόβλεψη της ενεργειακής κατανάλωσης διαφόρων τύπων κτιρίου, τα οποία αποτελούν ασταθή και κυμαινόμενα δεδομένα. Οι Fang et al. [26] από την πλευρά τους πρότειναν ένα πλαίσιο βαθιάς μάθησης για την πρόβλεψη της οικιακής κατανάλωσης ενέργειας, το οποίο κάνει χρήση των επαναλαμβανόμενων νευρωνικών δικτύων (RNNs) με την συνένωση πολλαπλών χαρακτηριστικών. Έτσι με το προτεινόμενο πλαίσιο βαθιάς μάθησης κατάφεραν να συνδυάσουν μεθόδους επεξεργασίας σήματος με αναδρομικά νευρωνικά δίκτυα (RNNs) για να επιτύχουμε προσαρμοστική αποσύνθεση και εξαγωγή πολλαπλών χαρακτηριστικών με αποτελέσματα να έχουν υψηλή ακρίβεια στην πρόβλεψη της ενεργειακής κατανάλωσης.

Ακόμη τα τελευταία χρόνια, η συνεργατική μάθηση (Federated Learning) έχει τραβήξει σημαντικό ενδιαφέρον μεταξύ των ερευνητών λόγω της ικανότητάς της να ενσωματώνει διάφορες πηγές δεδομένων και να διευκολύνει τη συνεργατική κατασκευή μοντέλων μηχανικής μάθησης. Για αυτό τον λόγο οι Abdulla et al. [27] προσπάθησαν να προβλέψουν την ενεργειακή κατανάλωση κτιρίων, χρησιμοποιώντας προσαρμοστική συνεργατική μάθηση. Τα μοντέλα που πρότειναν κάνουν χρήση δύο ξεχωριστών τύπων LSTM και σε συνδυασμό με την προσαρμοστική συνεργατική μάθηση κατάφεραν να μειώσουν τα σφάλμα τους κατά την πρόβλεψη κατά 8%.

Από την ανάλυση της βιβλιογραφίας, έχει φανεί ότι η καθολική εκπαίδευση είναι η πλέον χρησιμοποιούμενη τεχνική εκπαίδευσης των μοντέλων και για αυτόν τον λόγο επιλέχθηκε για την παρούσα εργασία. Επιπλέον, διάφορες έρευνες έχουν δείξει ότι η πρόβλεψη χρονοσειρών δεδομένων ενεργειακής κατανάλωσης με τη χρήση νευρωνικών δικτύων έχει ξεπεράσει τις στατιστικές μεθόδους. Συγκεκριμένα, η αρχιτεκτονική του πολυεπίπεδου τεχνητού νευρωνικού δικτύου έχει αποδειχθεί ότι συνδυάζεται με μεγάλη επιτυχία με την καθολική εκπαίδευση, δημιουργώντας προβλέψεις υψηλής ακρίβειας. Για αυτόν τον λόγο κατά την διάρκεια της εργασίας επιλέχθηκε η χρήση της συγκεκριμένης αρχιτεκτονικής.

1.3 Οργάνωση της εργασίας

Η εργασία αυτή είναι οργανωμένη σε έξι κεφάλαια. Στο 2ο κεφάλαιο δίνεται το θεωρητικό υπόβαθρο σχετικά με τις χρονοσειρές καθώς και τα ποιοτικά χαρακτηριστικά τους. Έπειτα,

παρουσιάζονται εν συντομία οι στατιστικοί μέθοδοι πρόβλεψης χρονοσειρών, παρέχοντας μια επισκόπηση των βασικών τεχνικών. Τέλος, αναλύονται τα μέτρα ακρίβειας των προβλέψεων χρονοσειρών, όπως η μετρική μέσου απόλυτου σφάλματος, που χρησιμοποιούνται για την αξιολόγηση της απόδοσης των προβλεπτικών μοντέλων και έγινε χρήση τους κατά την πειραματική διαδικασία.

Στο κεφάλαιο 3 γίνεται η θεωρητική ανάλυση των νευρωνικών δικτύων για πρόβλεψη χρονοσειρών. Αρχικά, παρουσιάζονται οι διάφορες κατηγορίες και αρχιτεκτονικές των νευρωνικών δικτύων, αναλύοντας τα χαρακτηριστικά και τις ιδιαιτερότητες της κάθε προσέγγισης. Ακόμη, εξετάζεται η διαδικασία εκπαίδευσης των νευρωνικών δικτύων, με έμφαση στην ανάλυση της καθολικής και της τοπικής εκπαίδευσης, και τα οφέλη που έχουν καθεμία από αυτές και τέλος ο τρόπος εισαγωγής εξωγενών μεταβλητών στα διάφορα νευρωνικά μοντέλα.

Στο 4ο κεφάλαιο ξεκινά η πειραματική διαδικασία. Σε αυτό το κεφάλαιο, αναλύσαμε τη συλλογή και την προεπεξεργασία των δεδομένων. Συγκεκριμένα, εξετάσαμε από ποια πηγή συλλέχθηκαν τα δεδομένα καθώς και την αρχική μορφή που αυτά είχαν με σκοπό την καλύτερη κατανόηση τους. Στη συνέχεια, παρουσιάζονται οι διάφορες τεχνικές προεπεξεργασίας που χρησιμοποιήθηκαν με σκοπό τη βελτίωση της ποιότητας των δεδομένων και την κατάλληλη μετατροπή τους ώστε να αποτελέσουν την είσοδο των νευρωνικών δικτύων.

Στο 5ο κεφάλαιο, αρχικά παρουσιάστηκαν και αναλύθηκαν όλα τα διαφορετικά μοντέλα που δημιουργήθηκαν με σκοπό την επίτευξη καλύτερων προβλέψεων. Κάθε μοντέλο εξετάστηκε λεπτομερώς, αναδεικνύοντας τις ιδιαίτερες προσεγγίσεις και τις τεχνικές που χρησιμοποιήθηκαν. Έπειτα, έγινε μια σύγκριση των αποτελεσμάτων που προέκυψαν από τα μοντέλα, τόσο συνολικά όσο και σε σχέση με μερικά από τα χαρακτηριστικά του συνόλου δεδομένων. Αυτή η συγκριτική ανάλυση είχε ως στόχο να αξιολογήσει την αποδοτικότητα των μοντέλων σε διάφορες συνθήκες και να εντοπίσει τα δυνατά και αδύνατα σημεία τους, συμβάλλοντας στην επιλογή των κατάλληλων μοντέλων για την πρόβλεψη ενεργειακών δεδομένων.

Στο 6ο και τελευταίο κεφάλαιο παρουσιάζονται τα συμπεράσματα που προέκυψαν από τη μελέτη. Αναδεικνύονται οι βασικές εκδηλώσεις και αναλύονται τα κύρια αποτελέσματα, ενώ γίνεται σύντομη αξιολόγηση της απόδοσης των μοντέλων και παρουσιάζονται οι επιτευχθείσες επιδόσεις σε σχέση με τους στόχους της μελέτης. Επίσης, προτείνονται μερικές μελλοντικές επεκτάσεις, προκειμένου να βελτιωθεί η μελέτη και να ενισχυθούν οι προβλέψεις ενεργειακών δεδομένων.

Κεφάλαιο 2

Μέθοδοι πρόβλεψης χρονοσειρών

Στο κεφάλαιο αυτό, θα εξετάσουμε εκτενώς την έννοια της χρονοσειράς και τη σημασία της στην ανάλυση δεδομένων. Θα εστιάσουμε στις βασικές κατηγορίες μεθόδων πρόβλεψης, εξετάζοντας τις διάφορες προσεγγίσεις που χρησιμοποιούνται για την πρόβλεψη των τιμών σε μελλοντικά χρονικά σημεία. Επιπλέον, θα διερευνήσουμε τα μέτρα ακρίβειας που χρησιμοποιούνται για να αξιολογήσουμε την απόδοση των μοντέλων πρόβλεψης, προσφέροντας ένα πλήρες πλαίσιο για την κατανόηση και την εφαρμογή των μεθόδων πρόβλεψης στην πράξη.

2.1 Εισαγωγή στις χρονοσειρές

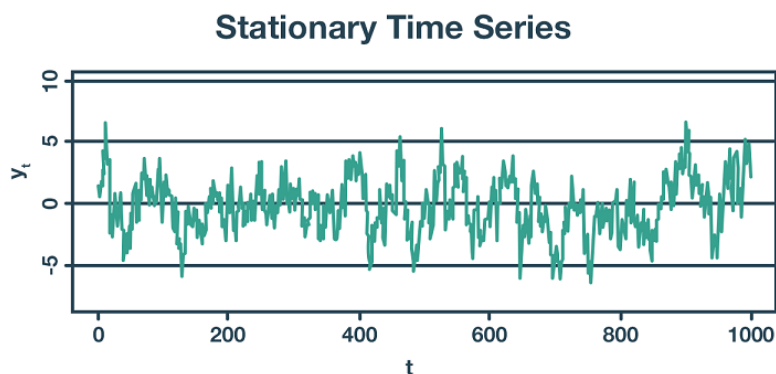
Χρονοσειρά (timeseries) είναι μία ακολουθία παρατηρήσεων, διατεταγμένων σε διαδοχικά ισαπέχουσες χρονικές στιγμές. Οι παρατηρήσεις αυτές αποτελούν συνήθως τιμές (π.χ. η τιμή κατανάλωσης ενέργειας) που καταγράφονται ανά σταθερό χρονικό βήμα το οποίο μπορεί να είναι ωριαίο, ημερήσιο, εβδομαδιαίο μηνιαίο, τριμηνιαίο, ακόμη και ετήσιο. Συναντώνται σε διάφορες επιστήμες, όπως η οικονομία, η μετεωρολογία, η ενέργεια, η ιατρική, και χρησιμοποιούνται ευρέως για την περιγραφή μεγεθών των οποίων η ιστορική πληροφορία μπορεί να έχει κάποια χρήση. Η ανάλυση χρονοσειρών έχει ως στόχο την εξαγωγή πληροφοριών, όπως η τάση και η εποχικότητα που αποτελούν σημαντικές πληροφορίες για την κατανόηση των δεδομένων. Στον τομέα της ενέργειας, στον οποίο εστιάζει η διπλωματική εργασία, οι χρονοσειρές χρησιμοποιούνται για την ανάλυση της κατανάλωσης ενέργειας, την πρόβλεψη της ζήτησης, την αξιολόγηση της απόδοσης των ενεργειακών συστημάτων, καθώς και την ανίχνευση ανωμαλιών στην παραγωγή και διανομή ενέργειας. Ακόμη, η μελέτη και ανάλυση των χρονοσειρών μπορεί να αποκαλύψει τάσεις και πρότυπα χρήσης, επιτρέποντας τη λήψη μέτρων για τη βελτίωση της ενεργειακής απόδοσης. [28]

2.2 Ποιοτικά χαρακτηριστικά χρονοσειρών

Τα ποιοτικά χαρακτηριστικά των χρονοσειρών περιγράφουν τα χαρακτηριστικά και τα πρότυπα που παρουσιάζουν οι χρονοσειρές χωρίς να απαιτείται κάποια στατιστική μοντελοποίηση. Τα ποιοτικά χαρακτηριστικά παίζουν σημαντικό ρόλο στην κατανόηση της χρονοσειράς και την προετοιμασία των δεδομένων για ανάλυση και πρόβλεψη. Παρακάτω θα γίνει μία σύντομη περιγραφή των κύριων ποιοτικών χαρακτηριστικών μιας χρονοσειράς.

2.2.1 Στασιμότητα

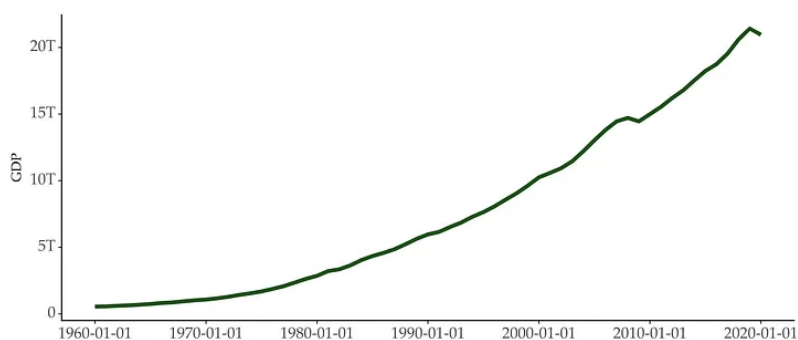
Στασιμότητα σε μια χρονοσειρά παρατηρούμε όταν οι τιμές παραμένουν σχετικά σταθερές στη διάρκεια του χρόνου και δεν παρουσιάζουν συστηματική τάση ή περιοδικές διακυμάνσεις. Αυτό σημαίνει ότι η μέση τιμή, η διακύμανση και άλλα στατιστικά χαρακτηριστικά της χρονοσειράς παραμένουν σταθερά χωρίς να αλλάζουν με την πάροδο του χρόνου. Παρακάτω παρουσιάζεται μια περίπτωση στάσιμης χρονοσειράς. [29]



Σχήμα 2.1: Παράδειγμα στάσιμης χρονοσειράς [1]

2.2.2 Τάση

Η τάση σε μια χρονοσειρά αναφέρεται στη μακροπρόθεσμη αύξηση ή μείωση των τιμών της χρονοσειράς κατά την πάροδο του χρόνου. Η τάση μπορεί να είναι ανοδική, όταν οι τιμές αυξάνονται με την πάροδο του χρόνου, ή καθοδική, όταν οι τιμές μειώνονται. Οι τάσεις μπορεί να είναι γραμμικές ή μη γραμμικές. Μια γραμμική τάση σημαίνει ότι η αύξηση ή η μείωση των τιμών είναι σχεδόν σταθερή κατά τη διάρκεια του χρόνου, ενώ μια μη γραμμική τάση δείχνει μια μεταβολή στην ταχύτητα της αύξησης ή μείωσης των τιμών. Η ανάλυση της τάσης είναι σημαντική για την πρόβλεψη των μελλοντικών τιμών της χρονοσειράς, καθώς μας επιτρέπει να κατανοήσουμε την κατεύθυνση και τη δυναμική των δεδομένων. [30] Παρακάτω φαίνεται μια χρονοσειρά η οποία έχει ανοδική τάση. Η συγκεκριμένη χρονοσειρά απεικονίζει το Ακαθάριστο Εγχώριο Προϊόν (Gross Domestic Product) των Ηνωμένων Πολιτειών της Αμερικής από τον 1η Ιανουαρίου του 1960 έως και την 1η Ιανουαρίου του 2020.

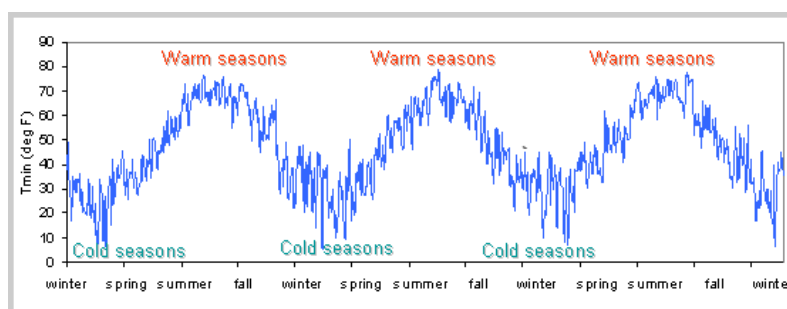


Σχήμα 2.2: Παράδειγμα χρονοσειράς με τάση [2]

2.2.3 Εποχικότητα

Η εποχικότητα σε μια χρονοσειρά αναφέρεται στα κυκλικά μοτίβα ή περιοδικές διακυμάνσεις που εμφανίζονται σε συγκεκριμένες περιόδους του χρόνου και επαναλαμβάνονται σε ετήσια βάση. Αυτές οι περιοδικές διακυμάνσεις μπορεί να οφείλονται σε εποχικούς παράγοντες, όπως οι κλιματικές συνθήκες, τα εορταστικά γεγονότα, οι εποχικές προσφορές και ζήτηση, και άλλοι παράγοντες που επηρεάζουν την καθημερινή ή εποχιακή συμπεριφορά των δεδομένων. Για παράδειγμα, σε τιμές κατανάλωσης ενέργειας για θέρμανση είναι φυσικό το χειμώνα να έχουμε αρκετά υψηλότερες τιμές κατανάλωσης ενέργειας συγκριτικά με το καλοκαίρι, όπου οι αναμενόμενες τιμές είναι αρκετά χαμηλές. Η ανίχνευση και η αντιμετώπιση της εποχικότητας είναι σημαντικές στην ανάλυση χρονοσειρών, καθώς αυτή μπορεί να επηρεάσει σημαντικά τις προβλέψεις και την αξιολόγηση της απόδοσης μοντέλων. Παρακάτω παρουσιάζεται μια χρονοσειρά με έντονη εποχικότητα σε δεδομένα θερμοκρασίας. Συγκεκριμένα απεικονίζεται η εποχιακή μεταβλητότητα της ημερήσιας ελάχιστης θερμοκρασίας σε Φαρενάιτ, για τρία χρόνια (1971-1973) στο Central Park της Νέας Υόρκης [31]

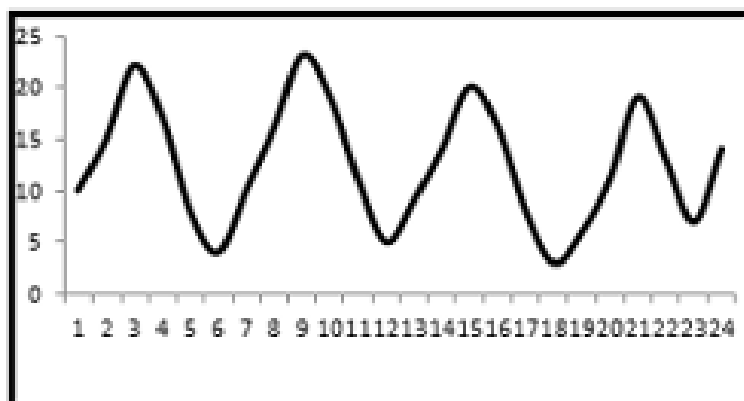
Σεασοναλ ἄριαβιλιτψ οφ Δαιλψ Μινιμυμ Τεμπερατυρε φορ τηρεε ψεαρς (1971-1973) ατ ἔντραλ Παρκ, Νεω Ψορκ



Σχήμα 2.3: Παράδειγμα χρονοσειράς με εποχικότητα [3]

2.2.4 Κυκλικότητα

Η κυκλικότητα σε μια χρονοσειρά αναφέρεται σε μεγαλύτερες κυκλικές διακυμάνσεις που υπερβαίνουν την εποχιακή επίδραση και διαρκούν για περισσότερο χρονικό διάστημα. Ενώ η εποχικότητα επικεντρώνεται σε συγκεκριμένες περιόδους και συνδέεται συνήθως με επαναλαμβανόμενα γεγονότα όπως οι εποχικές προσφορές ή οι κλιματικές συνθήκες, η κυκλικότητα αφορά σε μεγαλύτερους κύκλους που μπορεί να οφείλονται σε οικονομικούς κύκλους, πολιτικές αλλαγές, τάσεις ανάπτυξης ή ύφεσης, και άλλους μακροοικονομικούς παράγοντες. Για παράδειγμα, στον οικονομικό τομέα, η κυκλικότητα μπορεί να εμφανιστεί με κύκλους οικονομικής ευημερίας και ύφεσης που διαρκούν αρκετά χρόνια και επηρεάζουν τις επιχειρήσεις, τις επενδύσεις και την ανεργία. Η κατανόηση αυτών των κυκλικών κινήσεων μπορεί να βοηθήσει στη λήψη αποφάσεων για τη διαχείριση του ρίσκου και την προετοιμασία για μελλοντικές οικονομικές εξελίξεις. Παρακάτω φαίνεται μια χρονοσειρά που έχει δημιουργηθεί ώστε να αποτελέσει ένα παράδειγμα κυκλικότητας. [32]



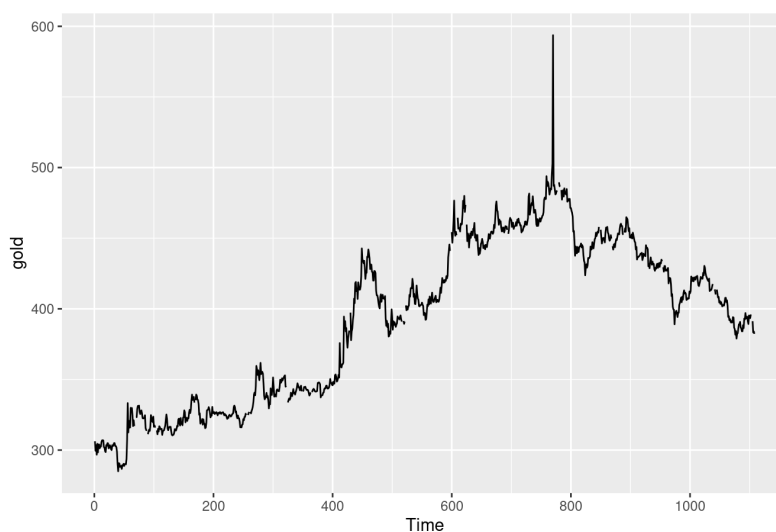
Σχήμα 2.4: Παράδειγμα χρονοσειράς με κυκλικότητα [4]

2.2.5 Ασυνέχειες

Οι ασυνέχειες σε μια χρονοσειρά αναφέρονται σε διακυμάνσεις ή αλλαγές που είναι εκτός του συνήθους προφίλ των δεδομένων και μπορεί να επηρεάσουν την ανάλυση και την πρόβλεψη. Οι κύριες μορφές ασυνέχειας είναι οι εξής.

Ασυνήθιστες τιμές

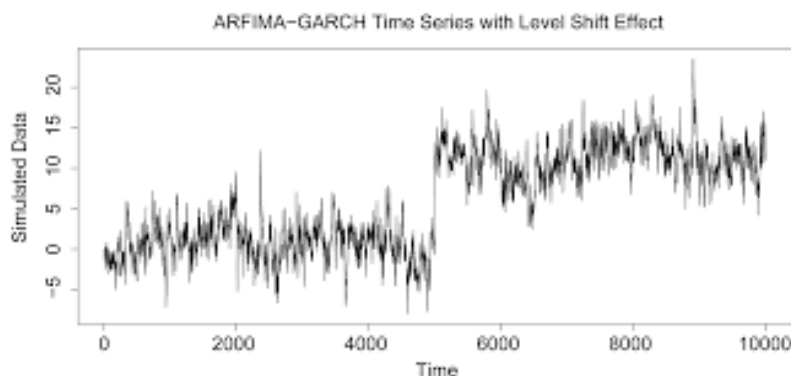
Οι ασυνήθιστες τιμές (outliers) αναφέρονται σε παρατηρήσεις που είναι αισθητά διαφορετικές από τις υπόλοιπες παρατηρήσεις στη χρονοσειρά. Αυτές οι τιμές μπορεί να οφείλονται σε τυχαίες διακυμάνσεις, λάθη μέτρησης ή ακόμα και σε εξαιρετικά σπάνια γεγονότα. Η ανίχνευση και η διαχείριση των ασυνήθιστων τιμών είναι σημαντικές για την αξιοπιστία της ανάλυσης και των προβλέψεων. Παρακάτω φαίνεται μια χρονοσειρά με ασυνήθιστες τιμές. Συγκεκριμένα η παρακάτω χρονοσειρά απεικονίζει την ημερήσια τιμή του χρυσού σε δολάρια των ΗΠΑ, από την 1η Ιανουαρίου 1985 έως τις 31 Μαρτίου 1989. [33]



Σχήμα 2.5: Παράδειγμα χρονοσειράς με outlier [5]

Αλλαγές επιπέδου

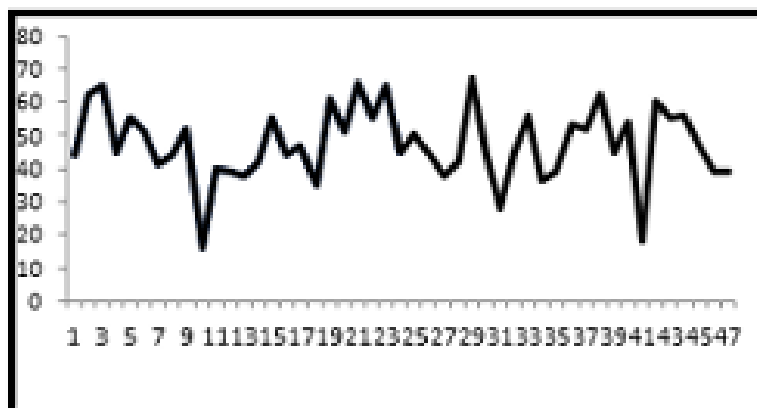
Οι αλλαγές επιπέδου (level shifts) αφορούν τις αιφνίδιες αλλαγές στο γενικό επίπεδο των τιμών της χρονοσειράς. Αυτές οι αλλαγές μπορεί να οφείλονται σε διαρθρωτικές αλλαγές στα δεδομένα, όπως αλλαγές στις πωλήσεις λόγω νέων προϊόντων ή αλλαγές σε μία νομοθεσία που επηρεάζει τον τρόπο λειτουργίας μιας επιχείρησης. Η ανίχνευση αυτών των αλλαγών είναι σημαντική για την ανάλυση της εξέλιξης της χρονοσειράς και την διαχείριση ενδεχομένων αλλαγών στο περιβάλλον. Παρακάτω φαίνεται μια χρονοσειρά με αλλαγές επιπέδου. Παρακάτω απεικονίζεται μια χρονοσειρά που έχει δημιουργηθεί με την χρήση μοντέλου ARFIMA-GARCH και περιέχει προσομοιωμένα δεδομένα στην πάροδο του χρόνου σε δευτερόλεπτα [34]



Σχήμα 2.6: Παράδειγμα χρονοσειράς με αλλαγή επιπέδου [6]

2.2.6 Τυχειότητα

Η τυχειότητα σε μια χρονοσειρά αναφέρεται στην τυχαία ή μη συστηματική διακύμανση των τιμών της χρονοσειράς, η οποία δεν είναι λόγω εποχικότητας, τάσης ή κυκλικότητας. Αυτές οι τυχαίες διακυμάνσεις είναι αποτέλεσμα πολλών διαφόρων παραγόντων που είναι δύσκολο ή αδύνατο να προβλεφθούν, όπως οι τυχαίες εκδηλώσεις, τα λάθη μέτρησης ή ακόμα και η ανακατάταξη των συμβάντων στο χρόνο. Παρακάτω παρουσιάζεται μια χρονοσειρά με τυχειότητα. Η συγκεκριμένη χρονοσειρά έχει σχεδιαστεί κατάλληλα με σκοπό την απεικόνιση της τυχειότητας σε μια χρονοσειρά. [35]



Σχήμα 2.7: Παράδειγμα χρονοσειράς με τυχαιότητα [7]

2.3 Μέθοδοι πρόβλεψης χρονοσειρών

Όπως αναφέρθηκε και προηγουμένως η πρόβλεψη μελλοντικών τιμών είναι ένα βασικό χαρακτηριστικό στην ανάλυση χρονοσειρών και μάλιστα αποτελεί την μεγαλύτερη πρόκληση μιας και η εκτίμηση του τρόπου με τον οποίο θα συνεχιστεί η χρονοσειρά εμπεριέχει διάφορες δυσκολίες. Ζητούμενο είναι να σχεδιαστεί μια διαδικασία που θα εξασφαλίζει ότι θα παράγονται όσο το δυνατόν πιο ακριβείς προβλέψεις, αξιοποιώντας στο έπακρο τη διαθέσιμη πληροφορία.

Υπάρχουν διάφορα είδη προβλέψεων μερικά από τα οποία παρουσιάζονται παρακάτω. Τα είδη αυτά των οποίων θα γίνει μια σύντομη περιγραφή είναι μέθοδοι στατιστικής πρόβλεψης. Η στατιστική πρόβλεψη είναι μια προσέγγιση ανάλυσης δεδομένων που χρησιμοποιεί στατιστικά μοντέλα για την εκτίμηση μελλοντικών τιμών βάσει ιστορικών δεδομένων. Τα πλεονεκτήματά της περιλαμβάνουν την άμεση εφαρμογή, την αρκετά μεγάλη ακρίβεια, την δυνατότητα χρήσης χωρίς εξειδικευμένες γνώσεις καθώς και τις χαμηλές απαιτήσεις σε υπολογιστικούς πόρους. Ωστόσο, έχει και μειονεκτήματα, όπως την προϋπόθεση ότι το ιστορικό πρότυπο της χρονοσειράς θα συνεχιστεί στο μέλλον, κάτι που δεν ισχύει πάντα, καθώς και την αδυναμία λήψης υπόψη ειδικών γεγονότων που μπορεί να επηρεάσουν τα δεδομένα. Επιπλέον, πολλές στατιστικές μέθοδοι απαιτούν μεγάλο αριθμό ιστορικών παρατηρήσεων για την παραγωγή αξιόπιστων προβλέψεων, γεγονός που μπορεί να είναι πρόβλημα σε περιπτώσεις με περιορισμένα δεδομένα. Πέρα από τις στατιστικές μεθόδους, υπάρχουν και μέθοδοι πρόβλεψης βασισμένες στη μηχανική μάθηση, οι οποίες χρησιμοποιούν προηγμένα αλγοριθμικά μοντέλα για την ανάλυση και πρόβλεψη μελλοντικών τάσεων, αξιοποιώντας την ικανότητα μάθησης από δεδομένα και αναγνώρισης πολύπλοκων μοτίβων. Η ανάλυση αυτού του είδους μοντέλων θα γίνει στην συνέχεια.

2.3.1 Αφελής μέθοδος

Η αφελής μέθοδος (Naive) είναι η πιο απλή μέθοδος στατιστικής πρόβλεψης. Δεν παράγει ακριβείς προβλέψεις στις περισσότερες περιπτώσεις αλλά πολλές φορές χρησιμοποιείται ως μέτρο σύγκρισης για άλλες μεθόδους. Η πρόβλεψη θεωρείται πως είναι ίση με την τελευταία

παρατήρηση της διαθέσιμης χρονοσειράς και περιγράφεται με τον παρακάτω τύπο.

$$F(t + 1) = Y(t)$$

Πέρα από την απλή αφελή μέθοδο, υπάρχει και η εποχική αφελής μέθοδος (seasonal Naive) η οποία δεν παίρνει μόνο την τελευταία παρατήρηση για πρόβλεψη, αλλά ένα σύνολο από παρατηρήσεις σε ένα χρονικό διάστημα. Για παράδειγμα μια εποχική αφελής μέθοδος μπορεί να χρησιμοποιεί τις παρατηρήσεις της τελευταίας εβδομάδας για να προβλέψει την επόμενη εβδομάδα. Με αυτόν τον τρόπο μπορεί να λάβει υπόψιν την εποχικότητα των χρονοσειρών που επαναλαμβάνονται εβδομαδιαίως. Η εποχική αφελής μέθοδος είναι αυτή που χρησιμοποιήθηκε και σαν μέτρο σύγκρισης για τα υπόλοιπα μοντέλα κατά την διάρκεια του πειράματος. [36]

2.3.2 Μέθοδος Εκθετικής Εξομάλυνσης

Η μέθοδος εκθετικής εξομάλυνσης είναι δημοφιλής τεχνική πρόβλεψης χρονοσειρών που βασίζεται σε απλούς αλγορίθμους για την εξομάλυνση των δεδομένων και τη δημιουργία προβλέψεων. Η χρήση αυτής της μεθόδου εξομάλυνσης είναι κατάλληλη για βραχυπρόθεσμες προβλέψεις ενός μεγάλου όγκου χρονοσειρών. Αποδίδουν καλύτερα σε δεδομένα που παρουσιάζουν στασιμότητα ή μικρό ρυθμό ανάπτυξης ή μείωσης ως προς το χρόνο. [37] Οι κυριότερες μέθοδοι εξομάλυνσης είναι οι εξής

- Simple Exponential Smoothing (SES)
- Holt
- Damped

Το SES δίνει μεγαλύτερη βαρύτητα στις πιο πρόσφατες παρατηρήσεις και μειώνει εκθετικά το βάρος των παλαιότερων παρατηρήσεων. Η βασική ιδέα είναι να υπολογιστεί μια σταθμισμένη μέση τιμή, όπου οι πιο πρόσφατες τιμές έχουν μεγαλύτερη σημασία. Από την πλευρά της η μέθοδος Holt επεκτείνει το Simple Exponential Smoothing εισάγοντας έναν μηχανισμό για την πρόβλεψη χρονοσειρών που παρουσιάζουν γραμμικές τάσεις. Αυτό γίνεται προσθέτοντας δύο εξισώσεις εξομάλυνσης, μία για το επίπεδο (ή την τιμή) της χρονοσειράς και μία για την τάση της. Τέλος το μοντέλο Damped είναι μια τροποποίηση του μοντέλου Holt που περιορίζει την επίδραση της τάσης στις μακροπρόθεσμες προβλέψεις. Αυτό γίνεται με την εισαγωγή ενός συντελεστή απόσβεσης που μειώνει προοδευτικά την τάση όσο προχωρούν οι προβλέψεις στο μέλλον.

2.3.3 Μοντέλα ARIMA

Τα μοντέλα ARIMA (Autoregressive Integrated Moving Average) είναι στοχαστικά μαθηματικά μοντέλα που χρησιμοποιούνται για την ανάλυση και πρόβλεψη χρονοσειρών. Αυτά τα μοντέλα συνδυάζουν τρεις βασικές συνιστώσες: την αυτοπαλινδρόμηση (autoregression, AR), την ολοκλήρωση (integration, I) και τον κινητό μέσο όρο (moving average, MA). Η σύνθεση αυτών των συνιστωσών επιτρέπει στα μοντέλα ARIMA να αντιμετωπίζουν τις τάσεις, τις εποχικές μεταβολές και τις τυχαίες διακυμάνσεις σε μια χρονοσειρά.

Η αυτοπαλινδρόμηση περιλαμβάνει τη χρήση των προηγούμενων τιμών της χρονοσειράς για την πρόβλεψη των μελλοντικών τιμών, με ένα μοντέλο AR τάξης p (AR(p)) να υποθέτει ότι η τρέχουσα τιμή είναι γραμμικός συνδυασμός των p προηγούμενων τιμών. Η ολοκλήρωση χρησιμοποιείται για να καταστήσει τη χρονοσειρά στάσιμη, εξαλείφοντας τις τάσεις και σταθεροποιώντας τη μέση τιμή μέσω της διαφορικής διαδικασίας, όπου ένα μοντέλο με d ολοκληρώσεις (I(d)) εφαρμόζει τη διαδικασία d φορές. Ο κινητός μέσος όρος χρησιμοποιεί τα προηγούμενα σφάλματα πρόβλεψης για την πρόβλεψη μελλοντικών τιμών, με ένα μοντέλο MA τάξης q (MA(q)) να υποθέτει ότι η τρέχουσα τιμή είναι γραμμικός συνδυασμός των q προηγούμενων σφαλμάτων πρόβλεψης. [38]

2.3.4 Μοντέλα Γραμμικής και Εκθετικής Τάσης

Οι μέθοδοι γραμμικής και εκθετικής τάσης συσχετίζουν το προβλεπόμενο μέγεθος με το χρόνο και είναι κατάλληλες για την παραγωγή μακροπρόθεσμων προβλέψεων. Αυτοί οι μέθοδοι αναφέρονται σε μια απλή παλινδρόμηση με τον χρόνο να είναι η μοναδική μεταβλητή τους. Η βασική μορφή των μοντέλων αυτών έχει την εξής μορφή

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Όπου Y είναι η εξαρτημένη μεταβλητή, και X είναι η μεταβλητή που αναφέρεται στον χρόνο. Το β_0 είναι η σταθερά ή το intercept και β_1 είναι οι συντελεστές παλινδρόμησης. Τέλος, το ε είναι ο όρος του σφάλματος.

2.3.5 Μοντέλα Πολλαπλής Παλινδρόμησης

Η πολλαπλή παλινδρόμηση είναι μια στατιστική τεχνική που χρησιμοποιείται για την εξαγωγή της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής και πολλών ανεξάρτητων μεταβλητών. Αυτή η τεχνική εφαρμόζεται συνήθως σε γραμμικές σχέσεις, όπου προσπαθούμε να προβλέψουμε την τιμή της εξαρτημένης μεταβλητής βάσει πολλών διαφορετικών παραμέτρων. Η βασική μορφή του μοντέλου πολλαπλής γραμμικής παλινδρόμησης είναι η εξής εξίσωση

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

Όπου Y είναι η εξαρτημένη μεταβλητή, X_i είναι οι ανεξάρτητες μεταβλητές, και το β_0 είναι η σταθερά ή το intercept και β_i είναι οι συντελεστές παλινδρόμησης που αντιπροσωπεύουν τη συμβολή κάθε ανεξάρτητης μεταβλητής στην πρόβλεψη της εξαρτημένης μεταβλητής, και ε είναι ο όρος του σφάλματος.

Τα μοντέλα πολλαπλής παλινδρόμησης είναι χρήσιμα όταν θέλουμε να κατανοήσουμε και να ποσοτικοποιήσουμε την επίδραση πολλών παραγόντων σε μια εξαρτημένη μεταβλητή. Χρησιμοποιούνται ευρέως σε διάφορους τομείς, όπως τα οικονομικά, η ιατρική, οι κοινωνικές επιστήμες και η μηχανική.

2.3.6 Μέθοδος Theta

Η μέθοδος πρόβλεψης Theta βασίζεται στην τροποποίηση των τοπικών καμπυλοτήτων της χρονοσειράς. Η αρχική χρονοσειρά αποσυντίθεται σε δύο ή περισσότερες γραμμές Theta. Κάθε μία από αυτές προεκτείνεται ξεχωριστά και οι προβλέψεις τους συνδυάζονται. Η μέθοδος Theta ακόμη είναι μια τεχνική πρόβλεψης χρονοσειρών που συνδυάζει την εκθετική εξομάλυνση με την εποχική αφελή μέθοδο. Αυτή η μέθοδος θεωρείται απλή αλλά αποτελεσματική, καθώς μπορεί να χρησιμοποιηθεί για προβλέψεις λαμβάνοντας υπόψη την εποχικότητα και τις τάσεις των χρονοσειρών. Η εκθετική εξομάλυνση αφορά την ανανέωση των εκτιμήσεων της τάσης και της εποχικότητας βάσει των προηγούμενων παρατηρήσεων, ενώ η εποχική αφελής μέθοδος προβλέπει τις μελλοντικές τιμές ως μέσο των τιμών της περιόδου του προηγούμενου κύκλου εποχικότητας. Η χρήση αυτών των δύο μεθόδων επιτρέπει στη μέθοδο Theta να προβλέπει με επιτυχία τις μελλοντικές τιμές της χρονοσειράς, λαμβάνοντας υπόψη την εξέλιξη των τάσεων και των εποχικών παραμέτρων. [39]

2.3.7 Νευρωνικά Δίκτυα

Τα μοντέλα πρόβλεψης με νευρωνικά δίκτυα είναι ένας από τους πιο προηγμένους τύπους μοντέλων πρόβλεψης που χρησιμοποιούνται στη μηχανική μάθηση και την ανάλυση χρονοσειρών. Μερικά σημαντικά στοιχεία σχετικά με τα μοντέλα πρόβλεψης με νευρωνικά δίκτυα είναι:

Ανάγκη Αρκετών Δεδομένων: Τα νευρωνικά δίκτυα χρειάζονται αρκετά δεδομένα για να εκπαιδευτούν αποτελεσματικά. Αυτό σημαίνει ότι, για να λειτουργήσουν επαρκώς και να παράγουν ακριβείς προβλέψεις, απαιτούν μεγάλο όγκο δεδομένων.

Ακρίβεια Προβλέψεων: Τα νευρωνικά δίκτυα συνήθως παράγουν ακριβείς προβλέψεις, ειδικά σε περιπτώσεις όπου η χρονοσειρά παρουσιάζει μη γραμμικά μοτίβα και πολύπλοκες συσχετίσεις. Αυτό τα καθιστά κατάλληλα για εφαρμογές με υψηλή πολυπλοκότητα.

Ανακάλυψη Μοτίβων: Τα νευρωνικά δίκτυα δεν υποθέτουν κάποιο συγκεκριμένο μοτίβο στα δεδομένα, αλλά προσπαθούν να το ανακαλύψουν μέσω της εκπαίδευσης. Αυτό τα καθιστά ευέλικτα και ικανά να προβλέπουν σε ποικίλες συνθήκες.

Τα νευρωνικά δίκτυα μπορούν να είναι ιδιαίτερα ισχυρά εργαλεία για την πρόβλεψη, αλλά πρέπει να λαμβάνουμε υπόψη την πολυπλοκότητά τους και την ανάγκη για κατάλληλη εκπαίδευση και ρύθμιση προκειμένου να παράγουν αξιόπιστα αποτελέσματα. [40]

Κατά την διάρκεια της μελέτης, επικεντρωθήκαμε σε μεθόδους πρόβλεψης μέσω νευρωνικών δικτύων. Συγκεκριμένα έγινε ανάλυση και σύγκριση των προβλέψεων διάφορων νευρωνικών δικτύων. Σε παρακάτω κεφάλαιο γίνεται λεπτομερής επεξήγηση της χρήσης νευρωνικών δικτύων για προβλέψεις.

2.3.8 Συνδυασμός Μεθόδων

Ο συνδυασμός μεθόδων πρόβλεψης αναφέρεται στην εφαρμογή δύο ή περισσότερων στατιστικών μεθόδων πρόβλεψης με διαφορετικά βάρη. Η επιλογή των μεθόδων που θα συμμε-

τέχουν καθορίζεται από τα χαρακτηριστικά της κάθε μεθόδου και της συγκεκριμένης χρονοσειράς. Αν και μια μέθοδος μπορεί να έχει μικρότερα σφάλματα (όπως το MAE ή το MSE) από μια άλλη, αυτό δεν σημαίνει ότι ο συνδυασμός τους θα έχει ως σφάλμα το μέσο όρο των σφαλμάτων τους. Επομένως, ο συνδυασμός μεθόδων μπορεί να οδηγήσει σε καλύτερες προβλέψεις ανάλογα με την κατάλληλη επιλογή μεθόδων και βαρών, λαμβάνοντας υπόψη τα ιδιαίτερα χαρακτηριστικά των δεδομένων και των μεθόδων πρόβλεψης.

2.4 Μέτρα ακρίβειας προβλέψεων

Τα μέτρα ακρίβειας προβλέψεων είναι μετρικές που χρησιμοποιούνται για να αξιολογήσουν την απόδοση μιας μεθόδου πρόβλεψης σε σχέση με τα πραγματικά δεδομένα. Συγκεκριμένα είναι μετρικές που βρίσκουν ένα είδος σφάλματος. Σφάλμα είναι η διαφορά που έχει η τιμή πρόβλεψης σε σχέση με την πραγματική τιμή.

Παρακάτω παρουσιάζονται οι κυριότερες μετρικές που χρησιμοποιούνται για την αξιολόγηση.

2.4.1 Μέσο σφάλμα

Το μέσο σφάλμα (Mean Error) μετρά την προκατάληψη ή υπολογίζει το μέσο όρο των αποκλίσεων των προβλέψεων από τις πραγματικές τιμές. Αυτό το μέτρο μας δίνει μια γενική ιδέα της κατεύθυνσης των σφαλμάτων, αλλά δεν λαμβάνει υπόψη τη σημασία των αποκλίσεων. Το ME μετρά προκατάληψη [41]

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

2.4.2 Μέσο απόλυτο σφάλμα

Το MAE (Mean Absolute Error) είναι το μέσο του απόλυτου της διαφοράς μεταξύ προβλεπόμενων και πραγματικών τιμών. Είναι μια μετρική η οποία αγνοεί την κατεύθυνση των σφαλμάτων και εστιάζει απλώς στο μέγεθος των αποκλίσεων. Σε αντίθεση με την μετρική του μέσου σφάλματος, το MAE μετρά την ακρίβεια των προβλέψεων και όχι την προκατάληψη. [42]

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

2.4.3 Μέσο τετραγωνικό σφάλμα

Το MSE (Mean Squared Error) είναι το μέσο του τετραγώνου της διαφοράς μεταξύ προβλεπόμενων και πραγματικών τιμών. Το τετράγωνο δίνει μεγαλύτερο βάρος σε μεγάλες αποκλίσεις, κάτι που το καθιστά ευαίσθητο σε ακραίες τιμές. [43]

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

2.4.4 Ρίζα μέσου τετραγωνικού σφάλματος

Το RMSE (Root Mean Squared Error) είναι η τετραγωνική ρίζα του MSE και αντιστοιχεί στην τυπική απόκλιση των αποκλίσεων. Χρησιμοποιείται ευρέως επειδή είναι ευαίσθητο σε μεγάλες αποκλίσεις, αλλά ταυτόχρονα είναι στην ίδια κλίμακα με τις αρχικές μετρικές. [44]

$$RMSE = \sqrt{MSE}$$

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

Πέρα από τους παραπάνω δείκτες σφάλματος, υπάρχουν και τα κανονικοποιημένα σφάλματα. Τα κανονικοποιημένα σφάλματα είναι μετρικές που αναφέρονται στην ακρίβεια μιας μεθόδου πρόβλεψης σε σχέση με την κλίμακα των δεδομένων που αξιολογούνται. Η κανονικοποίηση είναι σημαντική όταν οι τιμές των δεδομένων έχουν μεγάλο εύρος και διαφορετικές μονάδες μέτρησης, καθώς μπορεί να διευκολύνει τη σύγκριση των μετρήσεων ακρίβειας μεταξύ διαφορετικών μοντέλων ή μεθόδων πρόβλεψης. Κάποια από τα πιο συνηθισμένα κανονικοποιημένα μέτρα ακρίβειας είναι τα εξής.

2.4.5 Μέσο Απόλυτο Κανονικοποιημένο Σφάλμα

Το scaled MAE (Mean Absolute Scaled Error - MASE) είναι ο λόγος του MAE προς τη διαφορά των πραγματικών τιμών. Αυτό το μέτρο κανονικοποιεί το MAE με το εύρος των πραγματικών τιμών και μπορεί να χρησιμοποιηθεί για να συγκρίνει την ακρίβεια μεταξύ διαφορετικών δεδομένων. [45]

$$MASE = \frac{\frac{1}{h} \sum_{t=n+1}^{n+h} (|y_t - \hat{y}_t|)}{\frac{1}{n-1} \sum_{t=2}^n (|y_t - y_{t-1}|)}$$

όπου n είναι το δείγμα εκπαίδευσης και h η περίοδος πρόβλεψης

2.4.6 Κανονικοποιημένο Μέσο Τετραγωνικό Σφάλμα

Το scaled MSE είναι ο λόγος του MSE προς τη διαφορά των πραγματικών τιμών. Αντίστοιχα με το scaled MAE, αυτό το μέτρο κανονικοποιεί το MSE με το εύρος των πραγματικών τιμών και βοηθά στην αξιολόγηση της ακρίβειας πρόβλεψης σε διαφορετικά εύρη δεδομένων. Για παράδειγμα στον παρακάτω τύπο γίνεται κανονικοποίηση του MSE με την διαίρεση του με την μέση τιμή της χρονοσειράς

$$scaledMSE = \frac{MSE}{Y_{mean}}$$

2.4.7 Ρίζα Μέσου Τετραγωνικού Κανονικοποιημένου Σφάλματος

Το scaled RMSE (Root Mean Squared Scaled Error - RMSSE) είναι η κανονικοποιημένη εκδοχή του RMSE. Είναι ο λόγος του RMSE προς τη διαφορά των πραγματικών τιμών και χρησιμοποιείται επίσης για τη σύγκριση της ακρίβειας πρόβλεψης σε διαφορετικές κλίμακες δεδομένων. [46]

$$RMSSE = \sqrt{\frac{MSE}{\frac{1}{n-1} \sum_{t=2}^n (y_t - y_{t-1})^2}}$$

όπου n είναι το δείγμα εκπαίδευσης

Κατά την διάρκεια της μελέτης επικεντρωθήκαμε στο μέτρο ακρίβειας scaled MAE για να μπορέσουμε να συγκρίνουμε τα αποτελέσματα των διάφορων μοντέλων πρόβλεψης. Χρησιμοποιήσαμε δηλαδή την μετρική MAE διαιρώντας της όμως με την μέση τιμή των πραγματικών τιμών για να οδηγηθούμε σε ένα κατάλληλο μέτρο ακρίβειας για την αξιολόγηση των μοντέλων.

Κεφάλαιο 3

Χρήση νευρωνικών δικτύων για προβλέψεις

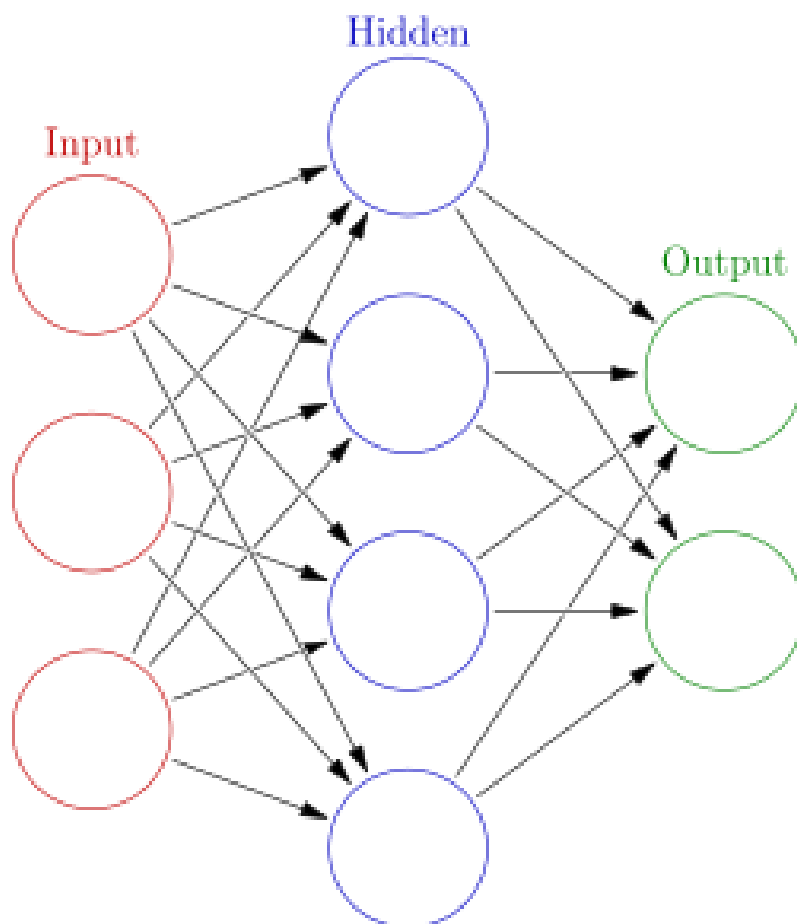
3.1 Νευρωνικά δίκτυα

Τα νευρωνικά δίκτυα (Neural Networks) έχουν εξελιχθεί σε μία από τις πιο διαδεδομένες μεθόδους μηχανικής μάθησης. Ένα σημαντικό χαρακτηριστικό τους είναι η ικανότητά τους να προσδιορίζουν τις παραμέτρους μίας γενικευμένης μη γραμμικής εξίσωσης που μοντελοποιεί τη σχέση μεταξύ των μεταβλητών εισόδου και εξόδου.

Με την αύξηση του πλήθους διαθέσιμων δεδομένων, τη μεγαλύτερη διαθέσιμη υπολογιστική ισχύ, και τη βελτίωση των αλγορίθμων, η χρήση των νευρωνικών δικτύων αυξήθηκε εκθετικά τα τελευταία χρόνια.

Η ευελιξία των νευρωνικών δικτύων τους καθιστά κατάλληλα για εφαρμογές επιβλεπόμενης μάθησης, όπως η ταξινόμηση και η παλινδρόμηση. Μπορούν επίσης να χρησιμοποιηθούν σε εφαρμογές μη επιβλεπόμενης και ημι-επιβλεπόμενης μάθησης. Η μη επιβλεπόμενη μάθηση επιτρέπει στα δίκτυα να ανακαλύπτουν μοτίβα και δομές στα δεδομένα χωρίς ετικέτες εξόδου, ενώ η ημι-επιβλεπόμενη μάθηση συνδυάζει τις δυνατότητες και των δύο προσεγγίσεων, επιτρέποντας την αξιοποίηση ετικετών για καλύτερη εκπαίδευση ενώ ταυτόχρονα εκμεταλλεύεται και την ανεπτυγμένη δυνατότητα ανακάλυψης μοτίβων. Αυτές οι διαφορετικές προσεγγίσεις καθιστούν τα νευρωνικά δίκτυα ιδιαίτερα ευέλικτα και ισχυρά σε ποικίλες εφαρμογές μηχανικής μάθησης. [8]

Με αυτήν την βελτίωση των νευρωνικών δικτύων έτσι άρχισε να αυξάνεται και η χρήση τους για την πρόβλεψη και ανάλυση χρονοσειρών και τα τελευταία χρόνια έχει αναδειχθεί ως ένα από τα πιο αποτελεσματικά εργαλεία στον τομέα της πρόβλεψης δεδομένων, παράγοντας πολλές φορές πιο ακριβής πρόβλεψης συγκριτικά με άλλα στατιστικά μοντέλα [47]. Τα νευρωνικά δίκτυα έχουν τη δυνατότητα να αντιληφθούν πολύπλοκες σχέσεις και προτεραιότητες στα δεδομένα χρονοσειρών, επιτρέποντας την ακριβή πρόβλεψη μελλοντικών τιμών ή προτύπων συμπεριφοράς. Το όφελος της χρήσης νευρωνικών δικτύων για την πρόβλεψη και ανάλυση χρονοσειρών προέρχεται από την ικανότητά τους να αντιμετωπίζουν μεγάλο όγκο δεδομένων με πολύπλοκες δομές και αβεβαιότητες.



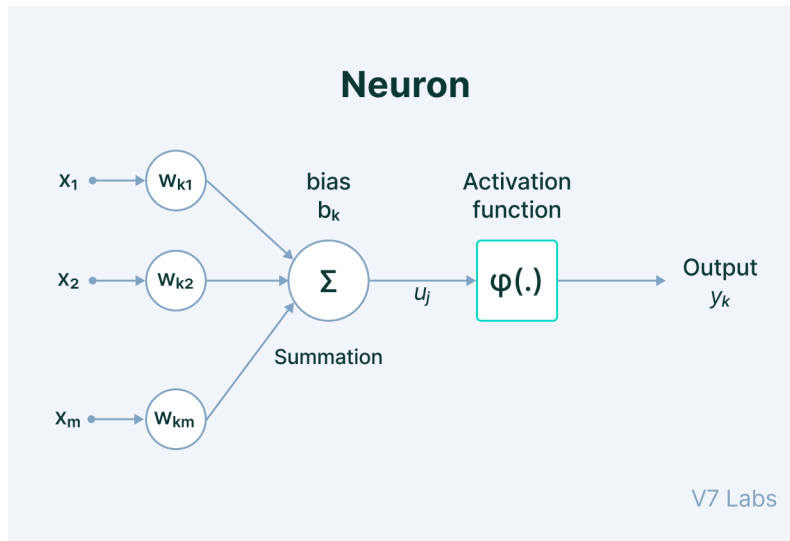
Σχήμα 3.1: Τεχνητό νευρωνικό δίκτυο [8]

Νευρώνες

Ένας τεχνητός νευρώνας ή νευρικός κόμβος είναι ένα μαθηματικό μοντέλο. Στις περισσότερες περιπτώσεις, υπολογίζει το σταθμισμένο μέσο όρο της εισόδου του και στη συνέχεια εφαρμόζει μια προκατάληψη σε αυτό. Έπειτα, περνά αυτόν τον προκύπτον όρο μέσω μιας συνάρτησης ενεργοποίησης. Αυτή η συνάρτηση ενεργοποίησης είναι μια μη γραμμική συνάρτηση όπως η σιγμοειδής συνάρτηση που δέχεται μια γραμμική είσοδο και δίνει μια μη γραμμική έξοδο. [48]

Στην πιο απλή περίπτωση νευρώνων (perceptron) η έξοδος κάθε νευρώνα δίνεται από την παρακάτω εξίσωση:

$$y = f\left(\sum_{i=1}^n (x_i * w_i) + b\right)$$



Σχήμα 3.2: Τεχνητός νευρώνας [9]

3.2 Συνάρτηση ενεργοποίησης

Ένα από τα πιο σημαντικά χαρακτηριστικά των νευρώνων είναι η συνάρτηση ενεργοποίησης (activation function), η οποία δίνει στο νευρωνικό δίκτυο τη μη γραμμική του συμπεριφορά και τη δυνατότητα να μοντελοποιεί πολύπλοκες σχέσεις και παραμέτρους. Η ενεργοποίηση μπορεί να γίνει θεωρητικά μέσω οποιασδήποτε συνάρτησης, όμως παρακάτω παρουσιάζονται οι κυριότερες από αυτές τις συναρτήσης ενεργοποίησης.

3.2.1 Linear

Η γραμμική συνάρτηση ενεργοποίησης είναι μια απλή μέθοδος, όπου η έξοδος εξαρτάται γραμμικά από την είσοδο. Αυτό σημαίνει ότι η αύξηση ή μείωση της εισόδου θα οδηγήσει σε αντίστοιχη αύξηση ή μείωση της εξόδου, με σταθερό ρυθμό. Είναι μια απλή, γρήγορη συνάρτηση, αλλά περιορισμένη σε πολύπλοκα μοντέλα λόγω της γραμμικότητάς της.

Είναι σημαντικό να παρατηρήσουμε ότι αν η γραμμική συνάρτηση επιλεγεί σαν συνάρτηση ενεργοποίησης για όλους τους νευρώνες του δικτύου, τότε το νευρωνικό δίκτυο χάνει την μη γραμμικότητά του και η συμπεριφορά του προσεγγίζει αυτή ενός μοντέλου γραμμικής πολλαπλής παλινδρόμησης.

3.2.2 ReLu

Η συνάρτηση ενεργοποίησης ReLU (Rectified Linear Unit) είναι μια μη γραμμική συνάρτηση που είναι μηδενική για αρνητικές τιμές και γραμμική (είναι ίση με την είσοδο) για μη θετικές τιμές. Αυτό το χαρακτηριστικό της καθιστά αποτελεσματική σε πολλά προβλήματα μηχανικής μάθησης, καθώς προσφέρει ταχύτητα εκτέλεσης και επιτρέπει την εκμάθηση πιο περίπλοκων συναρτήσεων. [10] Ορίζεται από την ακόλουθη συνάρτηση:

$$f(x) = \max(0, x)$$

3.2.3 Sigmoid

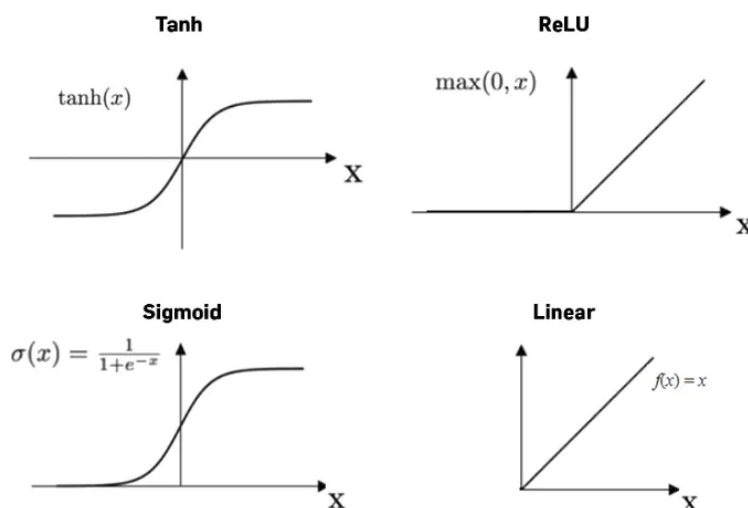
Η σιγμοειδής συνάρτηση ενεργοποίησης είναι μια καμπύλη σχήματος s που αντιστοιχίζει οποιαδήποτε είσοδο σε μια τιμή μεταξύ 0 και 1 καθιστώντας την κατάλληλη για προβλήματα δυαδικής ταξινόμησης. Ωστόσο, μπορεί να αντιμετωπίσει το πρόβλημα της "εξαφάνισης των κλίσεων" (vanishing gradients) σε βαθιά νευρωνικά δίκτυα, όπου οι παράγωγοι μπορεί να γίνουν πολύ μικροί για μεγάλες απόλυτες τιμές. [10] Ορίζεται από την ακόλουθη συνάρτηση: Ορίζεται από την ακόλουθη συνάρτηση:

$$f(x) = 1/(1 + e^{-x})$$

3.2.4 tanh

Η συνάρτηση ενεργοποίησης tanh είναι μια καμπύλη σχήματος s που αντιστοιχίζει οποιαδήποτε είσοδο σε μια τιμή μεταξύ -1 και 1. Χρησιμοποιείται συχνά σε προβλήματα όπου τα δεδομένα είναι κεντραρισμένα γύρω από το μηδέν, ενώ μπορεί να αντιμετωπίσει καλύτερα το πρόβλημα της εξαφάνισης των κλίσεων σε σύγκριση με τη sigmoid. [10] Ορίζεται από την ακόλουθη συνάρτηση:

$$f(x) = 2/(1 + e^{-2x}) - 1$$



Σχήμα 3.3: Συναρτήσεις ενεργοποίησης [10]

3.3 Κατηγορίες νευρωνικών δικτύων

Στον κόσμο των νευρωνικών δικτύων, υπάρχουν διαφορετικές κατηγορίες δικτύων που χρησιμοποιούνται για διαφορετικά είδη προβλημάτων. Από τα συμπαγή πολυεπίπεδα δίκτυα με πολλαπλά στρώματα (MLP) μέχρι τα σύγχρονα συνελκτικά νευρωνικά δίκτυα (CNN) και

τα αναδρομικά νευρωνικά δίκτυα (RNN), κάθε κατηγορία διαθέτει μοναδικά χαρακτηριστικά που την καθιστούν κατάλληλη για συγκεκριμένα προβλήματα.

Στη συνέχεια, θα εξετάσουμε πιο αναλυτικά τις πιο σημαντικές κατηγορίες νευρωνικών δικτύων και τα χαρακτηριστικά τους, προσδίδοντας έτσι μια ολοκληρωμένη εικόνα της ποικιλίας των διαθέσιμων αρχιτεκτονικών που μπορούν να χρησιμοποιηθούν στον τομέα της μηχανικής μάθησης και της τεχνητής νοημοσύνης.

3.3.1 Multilayer Perceptrons (MLPs)

Η πιο απλή κατηγορία νευρωνικών δικτύων αποτελείται αποκλειστικά από perceptrons, τα οποία οργανούνται σε στρώματα (layers). Τα νευρωνικά δίκτυα που έχουν περισσότερα από ένα τέτοια επίπεδα ονομάζονται Multilayer Perceptrons (MLPs). Τα MLPs είναι ένα είδος νευρωνικών δικτύων που αποτελούνται από πολλά επίπεδα νευρώνων που συνδέονται μεταξύ τους. Ανήκουν στην κατηγορία των feedforward νευρωνικών δικτύων, που σημαίνει ότι η πληροφορία κινείται μόνο προς τη μία κατεύθυνση, από τα επίπεδα εισόδου προς τα επίπεδα εξόδου. Τα MLP έχουν χρησιμοποιηθεί ευρέως σε διάφορους τομείς, όπως η αναγνώριση εικόνας, η επεξεργασία φυσικής γλώσσας και φυσικά η πρόβλεψη χρονοσειρών. Η ευελιξία τους στην αρχιτεκτονική και η ικανότητά τους να προσεγγίζουν οποιαδήποτε λειτουργία υπό ορισμένες συνθήκες τα καθιστούν θεμελιώδες δομικό στοιχείο στη βαθιά μάθηση και την έρευνα νευρωνικών δικτύων.

Ένα MLP αποτελείται τουλάχιστον από τρία επίπεδα: το επίπεδο εισόδου, τα κρυφά επίπεδα (hidden layers) και το επίπεδο εξόδου.

Επίπεδο εισόδου

Το επίπεδο εισόδου αποτελείται από κόμβους ή νευρώνες που λαμβάνουν τα αρχικά δεδομένα εισόδου. Κάθε νευρώνας αντιπροσωπεύει ένα χαρακτηριστικό ή μια διάσταση των δεδομένων εισόδου. Ο αριθμός των νευρώνων στο επίπεδο εισόδου καθορίζεται από τη διάσταση των δεδομένων εισόδου. [11]

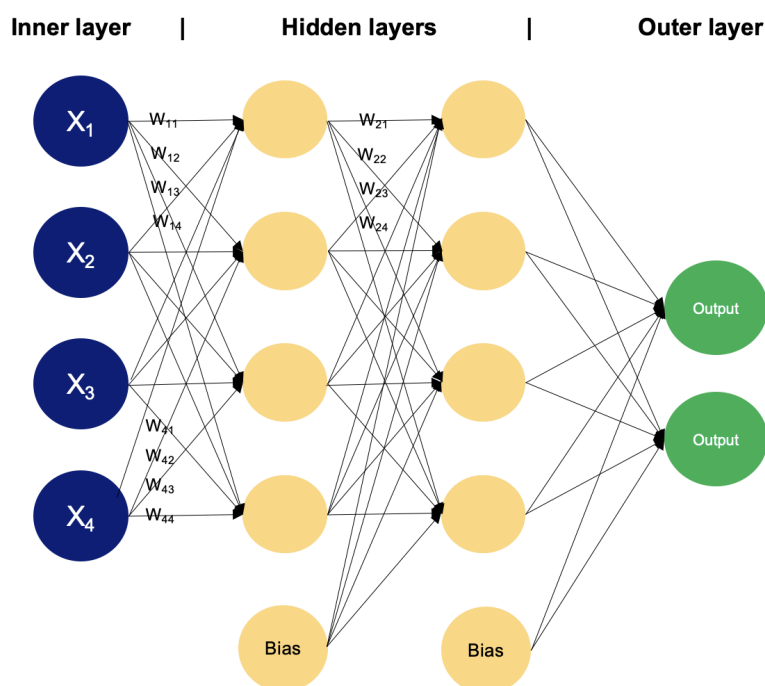
Κρυφά επίπεδα

Μεταξύ των επιπέδων εισόδου και εξόδου, μπορεί να υπάρχουν ένα ή περισσότερα στρώματα νευρώνων. Κάθε νευρώνας σε ένα κρυφό στρώμα λαμβάνει εισόδους από όλους τους νευρώνες του προηγούμενου στρώματος (είτε το στρώμα εισόδου είτε άλλο κρυφό στρώμα) και παράγει μια έξοδο που περνά στο επόμενο επίπεδο. Ο αριθμός των κρυφών επιπέδων και ο αριθμός των νευρώνων σε κάθε κρυφό στρώμα είναι υπερπαραμέτροι που πρέπει να προσδιοριστούν κατά τη φάση σχεδιασμού του μοντέλου. [11]

Επίπεδο εξόδου

Αυτό το επίπεδο αποτελείται από νευρώνες που παράγουν την τελική έξοδο του δικτύου. Ο αριθμός των νευρώνων στο επίπεδο εξόδου εξαρτάται από τη φύση της εργασίας. Στην πρόβλεψη χρονοσειρών μπορούμε να έχουμε, έναν νευρώνα που να προβλέπει μόνο την επόμενη παρατήρηση ή και αρκετά περισσότερους, όπως για παράδειγμα 168 διαφορετικούς

νευρώνες, έναν για κάθε ώρα της επόμενης εβδομάδας. Με αυτό τον τρόπο έχουμε πρόβλεψη για μια ολόκληρη εβδομάδα. [11]



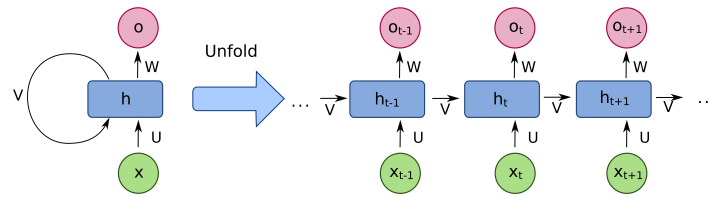
Σχήμα 3.4: Αρχιτεκτονική MLP [11]

Ένα MLP εκπαιδεύεται προσαρμόζοντας τα βάρη (weights) μεταξύ των perceptrons. Κάθε σύνδεση των νευρώνων έχει ένα σχετικό βάρος, το οποίο καθορίζει την αντοχή της σύνδεσης. Αυτά τα βάρη μαθαίνονται κατά τη διάρκεια της εκπαίδευσης.

Αυτή η κατηγορία νευρωνικών μοντέλων είναι και αυτή που χρησιμοποιήθηκε κατά την διάρκεια της μελέτης. Χρησιμοποιήσαμε διάφορα είδη MLP μοντέλων ώστε να έχουμε τα καλύτερα δυνατά αποτελέσματα

3.3.2 Επαναλαμβανόμενο νευρωνικό δίκτυο

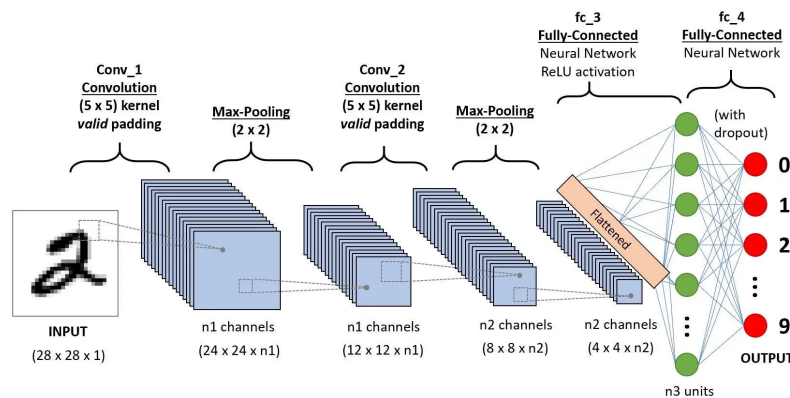
Μια πιο σύνθετη κατηγορία νευρώνων σε σχέση με τα απλά perceptrons είναι οι νευρώνες με δυνατότητα «μνήμης» ή στα αγγλικά Recurrent Neural Networks (RNNs). Η λειτουργία της μνήμης υλοποιείται με χρήση ενός μηχανισμού ανατροφοδότησης. Τα RNN αποτελούν μια ισχυρή τεχνική στον χώρο της μηχανικής μάθησης, σχεδιασμένη ειδικά για την αντιμετώπιση δεδομένων όπου ο χρόνος έχει σημαντική σημασία. Με την ικανότητά τους να διατηρούν μια εσωτερική κατάσταση ή μνήμη, είναι ιδανικά για εφαρμογές πρόβλεψης χρονοσειρών, όπως προβλέψεις τιμών κατανάλωσης ενέργειας, ζήτησης προϊόντων ακόμη και για χρηματιστηριακούς δείκτες. Επιπλέον, η ευελιξία των RNNs επιτρέπει την εφαρμογή διαφορετικών τύπων μνήμης στους νευρώνες, προσφέροντας ποικίλες επιλογές και προσαρμογή σε διαφορετικά προβλήματα. Με αυτά τα χαρακτηριστικά, τα ενPNNs αποτελούν ένα από τα πιο ισχυρά εργαλεία για την ανάλυση και πρόβλεψη δεδομένων που διαμορφώνονται με την πάροδο του χρόνου. [49]



Σχήμα 3.5: Αρχιτεκτονική RNN [12]

3.3.3 Συνελκτικά δίκτυα

Τα συνελκτικά δίκτυα (convolutional networks) προσφέρουν μια διαφορετική προσέγγιση στη δομή και λειτουργία των layers σε σχέση με τα παραδοσιακά νευρωνικά δίκτυα. Αντί να αποτελούνται από μεμονωμένους νευρώνες, τα επίπεδα των συνελκτικών δικτύων αποτελούνται από ένα σύνολο "φίλτρων" που εφαρμόζονται στην είσοδο για να προκύψουν νέα χαρακτηριστικά. Αυτή η δομή είναι ιδιαίτερα αποτελεσματική στην επεξεργασία και αναγνώριση εικόνων, όπως αποδεικνύουν τα επιτυχημένα μοντέλα όπως το VGG και το ResNet. Παρόλα αυτά, η εφαρμογή τους σε εφαρμογές πρόβλεψης χρονοσειρών παραμένει περιορισμένη, κάτι που αντιπροσωπεύει μια πρόκληση για περαιτέρω εξερεύνηση και ανάπτυξη στον τομέα αυτό. [50]



Σχήμα 3.6: Αρχιτεκτονική CNN [13]

3.4 Αρχιτεκτονική νευρωνικών δικτύων

Όπως είδαμε και παραπάνω, με τις διάφορες κατηγορίες νευρωνικών δικτύων, τα επίπεδα καθώς και τις συναρτήσεις ενεργοποίησης, η αρχιτεκτονική ενός δικτύου είναι ένας κρίσιμο βήμα που επηρεάζει σημαντικά την απόδοση του μοντέλου. Η βέλτιστη αρχιτεκτονική εξαρτάται από τη φύση του προβλήματος που επιδιώκουμε να λύσουμε καθώς και από τη διαθεσιμότητα και την ποιότητα των δεδομένων μας.

Η αρχιτεκτονική του δικτύου περιλαμβάνει πολλά στοιχεία που πρέπει να ληφθούν υπόψη, όπως το πλήθος των κρυφών επιπέδων, το πλήθος των νευρώνων σε αυτά τα επίπεδα, τον τρόπο διασύνδεσης των επιπέδων μεταξύ τους και τα χαρακτηριστικά που σχετίζονται με

τον σχεδιασμό των νευρώνων. Η επιλογή αυτών των παραμέτρων επηρεάζει την ικανότητα του μοντέλου να ανακαλύψει πολύπλοκες σχέσεις στα δεδομένα και να γενικεύσει σε νέες περιπτώσεις εκτός του συνόλου εκπαίδευσης. Κατάλληλες επιλογές σε αυτά τα στοιχεία μπορούν να οδηγήσουν σε βελτιωμένη επίδοση και αποτελεσματικότερη εκμάθηση από το μοντέλο νευρωνικού δικτύου. [51]

3.5 Εκπαίδευση νευρωνικών δικτύων

Η εκπαίδευση των νευρωνικών δικτύων είναι ακόμη μια πολύ σημαντική διαδικασία η οποία μπορεί να επηρεάσει σημαντικά τα αποτελέσματα.

Κατά τη διαδικασία εκπαίδευσης ενός νευρωνικού δικτύου, βελτιστοποιούνται οι τιμές των βαρών και ο όρος bias κάθε νευρώνα προκειμένου το δίκτυο να επιτύχει καλύτερη ακρίβεια και απόδοση στα δεδομένα εκπαίδευσης και ελέγχου. Η επιλογή της κατάλληλης συνάρτησης για τον υπολογισμό του σφάλματος είναι επίσης σημαντική και εξαρτάται από το είδος του προβλήματος που αντιμετωπίζουμε.

Κατά την εκπαίδευση, αντιμετωπίζονται διάφορα ζητήματα, όπως μεγάλοι χρόνοι εκτέλεσης, κίνδυνος τοπικών ελαχίστων, απότομες μεταβολές που δυσκολεύουν τη σύγκλιση και άλλα. Ο συντελεστής εκμάθησης (learning rate) παίζει κρίσιμο ρόλο στην εκπαίδευση, καθώς επηρεάζει την ταχύτητα με την οποία τα βάρη ενημερώνονται κατά την εκπαίδευση του δικτύου.

$$w(k) = w(k - 1) + b(d - y)x$$

Τα δεδομένα της εκπαίδευσης οργανώνονται σε πακέτα (batches) για κάθε ένα από τα οποία υπολογίζεται ένα σφάλμα εκπαίδευσης. Η πρόβλεψη ενός πακέτου δεδομένων, ο υπολογισμός του σφάλματος του και η ανανέωση των βαρών του δικτύου με βάση αυτό, ολοκληρώνουν ένα βήμα εκπαίδευσης. Όταν όλα τα διαθέσιμα βατσης περάσουν από την διαδικασία της εκπαίδευσης, και το δίκτυο έχει «δει» όλα τα διαθέσιμα δεδομένα μια φορά, ολοκληρώνεται μια εποχή εκπαίδευσης. Έτσι, πέρα από τα βάρη, πολύ σημαντικό ρόλο παίζει και το μέγεθος των πακέτων των δεδομένων που τροφοδοτούνται τα μοντέλα. Ένα μικρό μέγεθος βατση μπορεί να οδηγήσει σε αργή εκπαίδευση και υψηλό ρίσκο υπερ-προσαρμογής (overfitting), όπου το μοντέλο μαθαίνει τα δεδομένα εκπαίδευσης υπερβολικά καλά αλλά δυσκολεύεται να γενικεύσει σε νέα δεδομένα. Από την άλλη πλευρά, ένα πολύ μεγάλο μέγεθος βατση μπορεί να οδηγήσει σε μειωμένη ακρίβεια κατά την εκπαίδευση. Κατά την επιλογή του μεγέθους των batches, κρίσιμα κριτήρια είναι το πλήθος των δεδομένων εκπαίδευσης, η φύση των δεδομένων, για παράδειγμα αν είναι χρονοσειρές, εικόνες ή κείμενα, η πολυπλοκότητα του δικτύου που εκπαιδεύεται και οι διαθέσιμοι υπολογιστικοί πόροι. Συνήθως, γίνεται πειραματισμός με διαφορετικά μεγέθη batches για να επιλεγεί αυτό που προσφέρει τη βέλτιστη απόδοση στο συγκεκριμένο πρόβλημα που αντιμετωπίζεται. [52]

3.6 Είσοδος και έξοδος νευρωνικών δικτύων για πρόβλεψη χρονοσειρών

Ακόμη ένα πολύ σημαντικό κομμάτι στην κατασκευή των νευρωνικών δικτύων με στόχο την πρόβλεψη χρονοσειρών είναι η σωστή κατασκευή των επιπέδων εισόδου και εξόδου.

3.6.1 Επίπεδο εισόδου

Για την πρόβλεψη χρονοσειρών, τα νευρωνικά δίκτυα λειτουργούν με το να λαμβάνουν υπόψη παρελθοντικές τιμές της χρονοσειράς (lags). Αυτό συνιστά το νευρωνικό δίκτυο ένα μη γραμμικό auto-regression μοντέλο, όπου κάθε μελλοντική παρατήρηση προβλέπεται βάσει των παρελθοντικών παρατηρήσεων. Η επιλογή του αριθμού των παρελθοντικών τιμών που λαμβάνονται υπόψη, γνωστή ως look-back window, είναι μια κρίσιμη παράμετρος του δικτύου καθώς επηρεάζει την ικανότητα του μοντέλου να αντλήσει πληροφορίες από την παρελθοντική συμπεριφορά της χρονοσειράς. Επιπλέον, η είσοδος του νευρωνικού δικτύου μπορεί να περιλαμβάνει και άλλες μεταβλητές, γνωστές ως features.

3.6.2 Επίπεδο εξόδου

Στο επίπεδο εξόδου μπορούμε να έχουμε διαφορετικό αριθμό νευρώνων. Ο οριζόντιος πρόβλεψης, σε προβλήματα πρόβλεψης χρονοσειρών ορίζεται από το χρονικό διάστημα που θέλουμε να προβλέψουμε μπροστά στο μέλλον. Η επιλογή του κατάλληλου οριζόντιου πρόβλεψης εξαρτάται από το είδος του προβλήματος που αντιμετωπίζουμε. Στην πρακτική, χρησιμοποιούμε διαφορετικές στρατηγικές για την παραγωγή προβλέψεων, ανάλογα με το επιθυμητό οριζόντιο πρόβλεψης και τη φύση του προβλήματος. Έτσι ανάλογα με τον οριζόντιο πρόβλεψης που θα επιλέξουμε έχουμε και διαφορετικό αριθμό νευρώνων στο επίπεδο εξόδου. Υπάρχουν τρεις βασικές στρατηγικές σχετικά με την επιλογή οριζόντιου πρόβλεψης. Η πρώτη στρατηγική είναι η "Iterative", όπου κάθε μοντέλο νευρωνικού δικτύου εκπαιδεύεται για να παράγει προβλέψεις μόνο για ένα βήμα μπροστά. Η έξοδος αυτού του μοντέλου γίνεται μέρος της εισόδου για την παραγωγή της επόμενης πρόβλεψης, επαναλαμβάνοντας τη διαδικασία για τον επιθυμητό οριζόντιο πρόβλεψης. Η δεύτερη στρατηγική είναι η "Direct", όπου κάθε μοντέλο νευρωνικού δικτύου εκπαιδεύεται για να παράγει όλες τις προβλέψεις ταυτόχρονα για τον ορισμένο οριζόντιο πρόβλεψης. Η διαδικασία πρόβλεψης συμβαίνει σε ένα βήμα για όλες τις προβλέψεις. Τέλος, η τρίτη στρατηγική είναι η "Multi-NN", όπου κάθε μοντέλο νευρωνικού δικτύου εκπαιδεύεται να παράγει μία πρόβλεψη, αντιστοιχούμενη σε ένα σημείο του οριζόντιου πρόβλεψης. Αυτή η στρατηγική απαιτεί την εκπαίδευση πολλών διαφορετικών νευρωνικών δικτύων για να καλυφθεί ο ζητούμενος οριζόντιος πρόβλεψης.

3.7 Εκπαίδευση συνόλου δεδομένων με σκοπό την χρήση τους από νευρωνικά δίκτυα

Πέρα από τις διάφορες παραμέτρους που πρέπει να ρυθμιστούν με σκοπό την κατάλληλη εκπαίδευση των δεδομένων, πολύ σημαντική είναι και η σωστή μετατροπή του συνόλου

δεδομένων με σκοπό την αποτελεσματική φόρτωση του training set στο νευρωνικό δίκτυο. Για την εκπαίδευση νευρωνικών δικτύων, είναι σημαντική η κανονικοποίηση των δεδομένων, είτε μέσω κάποιο είδος τυποποίησης (standardization) των δεδομένων είτε με την κλιμάκωσή τους (scaling). Αυτό βοηθά στη σταθεροποίηση της εκπαίδευσης και τη μείωση του χρόνου εκπαίδευσης, καθώς βελτιώνει τη σύγκλιση του μοντέλου. Οι κλιμακωτές (scalers) είναι μέθοδοι κανονικοποίησης που χρησιμοποιούνται στην προεπεξεργασία δεδομένων πριν από την εκπαίδευση ενός μοντέλου μηχανικής μάθησης. Μερικοί βασικοί scalars είναι οι εξής

MinMaxScaler

Ο MinMaxScaler κλιμακώνει τα δεδομένα σε ένα διάστημα [0, 1] ή [-1, 1], ανάλογα με την επιλογή, είναι ευαίσθητο σε ακραίες τιμές και μπορεί να επηρεάσει αρνητικά τα δεδομένα σε περιπτώσεις που υπάρχουν ακραίες τιμές. [53] Παρακάτω παρουσιάζεται ο τύπος του.

$$X_{\text{κανονικοποιημένο}} = \frac{X - X_{\text{ελάχιστο}}}{X_{\text{μέγιστο}} - X_{\text{ελάχιστο}}} * (\text{μέγιστο} - \text{ελάχιστο}) + \text{ελάχιστο}$$

Όπου το X είναι η αρχική τιμή των δεδομένων, $X_{\text{ελάχιστο}}$ και $X_{\text{μέγιστο}}$ είναι αντίστοιχα η ελάχιστη και η μέγιστη τιμή των δεδομένων και τέλος το 'μέγιστο' και το 'ελάχιστο' είναι για την ρύθμιση του επιθυμητού εύρους κλιμάκωσης (η προεπιλογή είναι 1 και 0 αντίστοιχα). [54]

Standard Scaler

Ο Standard Scaler είναι ακόμη ένας γνωστός τρόπος κανονικοποίησης δεδομένων, με τον οποίο μετατρέπονται τα δεδομένα ώστε να έχουν μέση τιμή μηδέν και διακύμανση μονάδας. Είναι ανθεκτική σε outliers και συνήθως προτιμάται σε περιπτώσεις που τα δεδομένα δεν ακολουθούν κανονική κατανομή. [53] Ο τύπος αυτής της τεχνικής κανονικοποίησης είναι ο εξής.

$$z = \frac{x - \mu}{\sigma}$$

όπου z είναι η κανονικοποιημένη τιμή, x είναι η αρχική τιμή, μ είναι η μέση τιμή του συνόλου δεδομένων και σ είναι η τυπική απόκλιση του συνόλου δεδομένων. [55]

Robust Scaler

Αυτή η μέθοδος βασίζεται στη μετατροπή των δεδομένων σε μια κατανομή που είναι ανθεκτική σε ακραίες τιμές. Χρησιμοποιείται όταν τα δεδομένα περιέχουν πολλά outliers και ο Standard Scaler δεν λειτουργεί καλά. [53] Ο τύπος για τον Robust Scaler είναι ο εξής.

$$X_{\text{κανονικοποιημένο}} = \frac{X_i - Q_1(x)}{Q_3(x) - Q_1(x)}$$

Όπου $Q_1(x)$ είναι το πρώτο τεταρτημόριο, $Q_3(x)$ είναι το τρίτο τεταρτημόριο και $Q_3(x) - Q_1(x)$ είναι το διατεταγμένο εύρος μεταξύ του πρώτου και τρίτου τεταρτημορίου (interquartile range). [56]

MaxAbs Scaler

Κλιμακώνει τα δεδομένα στο διάστημα $[-1, 1]$ αναγνωρίζοντας τη μέγιστη απόλυτη τιμή των χαρακτηριστικών. Χρησιμοποιείται κυρίως όταν τα δεδομένα είναι κεντραρισμένα στο μηδέν. [53] Παρακάτω παρουσιάζεται ο τύπος της συγκεκριμένης τεχνικής κανονικοποίησης.

$$X_{\text{κανονικοποιημένο}} = \frac{X}{|X_{\text{μέγιστο}}|}$$

Όπου $X_{\text{μέγιστο}}$ είναι η μέγιστη τιμή του συνόλου δεδομένων. [57]

Κατά την διάρκεια της μελέτης, έγινε η χρήση του MinMaxScaler για την σωστή κανονικοποίηση των δεδομένων.

3.8 Καθολική και τοπική εκπαίδευση

Ακόμη μια σημαντική απόφαση κατά την δημιουργία των νευρωνικών δικτύων είναι αν τα μοντέλα θα κάνουν χρήση της καθολικής ή της τοπικής εκπαίδευσης. Δυο αρκετά διαφορετικών τρόπων εκπαίδευσης τα οποία μπορεί να οδηγήσουν σε αρκετά διαφορετικά αποτελέσματα.

3.8.1 Τοπική εκπαίδευση

Η τοπική εκπαίδευση (Localized training) αναφέρεται σε μια προσέγγιση εκπαίδευσης νευρωνικών δικτύων για προβλέψεις χρονοσειρών. Σε αυτήν τη μέθοδο, το μοντέλο εκπαιδεύεται με βάση τα δεδομένα μιας και μόνο συγκεκριμένης χρονοσειράς. Τα δείγματα εκπαίδευσης δημιουργούνται χρησιμοποιώντας την τεχνική των κυλιόμενων παραθύρων.

Αυτή η προσέγγιση θεωρείται παραδοσιακή, αλλά εξακολουθεί να χρησιμοποιείται με επιτυχία σε ορισμένες εφαρμογές πρόβλεψης. Έχει διάφορα πλεονεκτήματα, όπως η γρήγορη εκπαίδευση όταν απαιτείται πρόβλεψη μικρού αριθμού χρονοσειρών, γεγονός που οδηγεί και σε μικρό υπολογιστικό κόστος. Επίσης ακόμη ένα πλεονέκτημα της τοπικής εκπαίδευσης είναι η ικανότητα του νευρωνικού δικτύου να λαμβάνει υπόψιν τις ιδιαιτερότητες και τα μοτίβα των δεδομένων που προβλέπει. Ακόμη ένα πλεονέκτημα αυτού του τρόπου εκπαίδευσης είναι πως δεν υπάρχει ανάγκη για συλλογή και επεξεργασία ενός μεγάλου συνόλου χρονοσειρών. Ωστόσο, υπάρχουν και μειονεκτήματα όπως η αποτελεσματικότητα της εκπαίδευσης όταν οι χρονοσειρές δεν έχουν αρκετές παρατηρήσεις και η δυσκολία επέκτασης της προσέγγισης σε περισσότερες χρονοσειρές. [58]

3.8.2 Καθολική εκπαίδευση

Από την άλλη στην καθολική εκπαίδευση (Global training) το μοντέλο εκπαιδεύεται βάσει όλων των διαθέσιμων δεδομένων. Πρόκειται για μια πιο μοντέρνα προσέγγιση στην ανάπτυξη νευρωνικών δικτύων για προβλέψεις χρονοσειρών. Και αυτή η μέθοδος έχει αρκετά πλεονεκτήματα όπως το γεγονός ότι τα μοντέλα μπορούν να εκμεταλλευτούν γνώση από υπόλοιπα δεδομένα και να βρουν γενικά μοτίβα μέσα σε αυτά. Επίσης ένα νευρωνικό δίκτυο εκπαιδευμένο με αυτόν τον τρόπο μπορεί να παρέχει άμεσα προβλέψεις για μεγάλο αριθμό χρονοσειρών. Συγκεκριμένα οι Mohin et al.[59] στην έρευνα που έκαναν για την

πρόβλεψη χρονοσειρών για την ηλιακή ακτινοβολία, τόνισαν την σημαντικότητα της χρήσης ενός μεγάλου global dataset καθώς έτσι μπορεί το μοντέλο να ανακαλύψει κρυφά μοτίβα, επιτρέποντάς του να κάνει πιο ακριβείς προβλέψεις. Αυτά τα μοντέλα μπορούν να χειριστούν αποτελεσματικά χρονικές εξαρτήσεις, εποχικότητα και μη γραμμικές τάσεις, οδηγώντας σε βελτιωμένη ακρίβεια στις μελλοντικές προβλέψεις.

Ωστόσο και σε αυτή την περίπτωση έχουμε μερικά μειονεκτήματα. Χρειάζονται σημαντικοί υπολογιστικοί πόροι για την εκπαίδευση, καθώς το μέγεθος του training set είναι μεγάλο.

[58]

Επίσης, τα μοντέλα δίνουν μικρότερη σημασία στις ιδιαιτερότητες κάποιων υποσυνόλων των δεδομένων. Αυτό έχει επισημανθεί και από τους Liam Vaughan et al.[60], οι οποίοι χρησιμοποίησαν global dataset στο πείραμα τους για την πρόβλεψη χρονοσειρών για την εξάπλωση του Covid-19 μέσω λυμάτων. Κατέληξαν πως όταν το μοντέλο τους εκπαιδευόταν σε όλο το σύνολο των δεδομένων δεν δημιουργούσε ακριβή αποτελέσματα λόγω της αδυναμίας να προσεγγίσει τα μοτίβα που δημιουργούσαν οι τοπικοί παράγοντες.

Τέλος οι Verdes et al.[15] στην έρευνα του σχετικά με τα οφέλη καθώς και τα μειονεκτήματα της καθολικής και της τοπικής εκπαίδευσης για την πρόβλεψη χρονοσειρών κατέληξαν πως για μικρό σύνολο δεδομένων η χρήση της καθολικής εκπαίδευσης ήταν προτιμότερη ενώ για μεσαίου μεγέθους συνόλου δεδομένων τα αποτελέσματα ήταν συγκρίσιμα. Τόνισαν επίσης πως το υπολογιστικό κόστος της τοπικής εκπαίδευσης είναι αρκετά μικρότερο από αυτό της καθολικής.

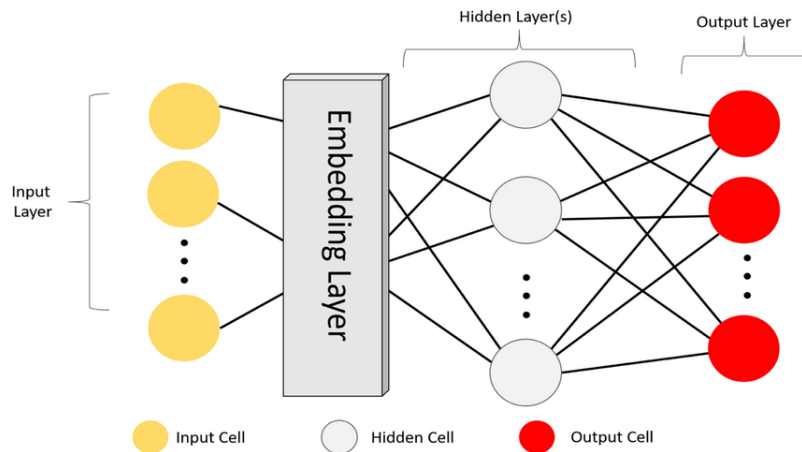
Λαμβάνοντας υπόψιν τις παραπάνω πληροφορίες, κατά την διάρκεια του πειράματος, δημιουργήθηκαν νευρωνικά δίκτυα που έκαναν χρήση της καθολικής εκπαίδευσης, τόσο στο συνολικό αριθμό των δεδομένων όσο και σε εξειδίκευση σε ένα υποσύνολο των δεδομένων, με σκοπό την δημιουργία ενός βέλτιστου μοντέλου βάση του συνόλου δεδομένων που χρησιμοποιήθηκε.

3.9 Εισαγωγή εξωγενών μεταβλητών στα νευρωνικά δίκτυα

Η εισαγωγή εξωγενών μεταβλητών σε ένα νευρωνικό δίκτυο μπορεί να γίνει με διάφορους τρόπους, ανάλογα με τη φύση των μεταβλητών. Ένας από τους τρόπους με τους οποίους μπορεί να επιτευχθεί κάτι τέτοιο είναι η τοπική εκπαίδευση των νευρωνικών δικτύων. Όπως είδαμε και παραπάνω μέσω της τοπικής εκπαίδευσης το μοντέλο μπορεί να εκπαιδευτεί σε ένα υποσύνολο των δεδομένων. Για παράδειγμα όταν έχουμε ένα σετ από χρονοσειρές που αφορούν την κατανάλωση ενέργειας ανά τον κόσμο και γίνει η επιλογή ενός υποσυνόλου αυτών των δεδομένων, όπως η κατανάλωση ενέργειας για μία ήπειρο, προσφέρουμε στο μοντέλο μας παραπάνω γνώση σχετικά με την γεωγραφική περιοχή στην οποία αναφέρονται τα δεδομένα. Ακόμη μια πολύ σημαντική μέθοδος με την οποία μπορεί να γίνει προσθήκη εξωγενών μεταβλητών είναι μέσω embedding layers. Παρακάτω παρουσιάζεται ο τρόπος με τον οποίο γίνεται η προσθήκη πληροφορίας μέσω επιπέδων ενσωμάτωσης.

Εισαγωγή εξωγενών μεταβλητών μέσω επιπέδου ενσωμάτωσης

Ένα επίπεδο ενσωμάτωσης (embedding layer) είναι ένας τύπος κρυφού στρώματος σε ένα νευρωνικό δίκτυο. Η είσοδος μεταβλητών μέσω embeddings αναφέρεται στον τρόπο με τον οποίο μετατρέπουμε κατηγορικές ή διακριτές μεταβλητές σε συνεχείς αναπαραστάσεις που μπορούν να εκπαιδευτούν πιο αποτελεσματικά σε ένα νευρωνικό δίκτυο. Με μια φράση, αυτό το επίπεδο αντιστοιχίζει πληροφορίες εισόδου από έναν χώρο υψηλής διάστασης σε έναν χώρο χαμηλότερης διάστασης, επιτρέποντας στο δίκτυο να μάθει περισσότερα για τη σχέση μεταξύ των εισόδων και να επεξεργάζεται τα δεδομένα πιο αποτελεσματικά. Έτσι για παράδειγμα όταν έχουμε ένα σύνολο δεδομένων το οποίο αναφέρεται σε κατανάλωση ενέργειας για διαφορετικά είδη κτιρίου, μπορούμε μέσω embedding layers να προσφέρουμε στο μοντέλο πληροφορία σχετικά με τον τύπο κτιρίου. [61] Συνολικά, τα embeddings βοηθούν το νευρωνικό δίκτυο να κατανοήσει και να εκμεταλλευτεί καλύτερα τη σημασιολογική πληροφορία των κατηγοριών ή των διακριτών μεταβλητών που εισάγουμε, βελτιώνοντας έτσι την απόδοση και τη γενίκευση του μοντέλου.



Σχήμα 3.7: Αρχιτεκτονική MLP με προσθήκη embedding layer [14]

Κεφάλαιο 4

Συλλογή και επεξεργασία δεδομένων

Σε αυτό το κεφάλαιο περιγράφεται η μορφή των δεδομένων που χρησιμοποιήθηκε για την υλοποίηση της έρευνάς μας. Αρχικά, γίνεται λεπτομερής περιγραφή των χαρακτηριστικών των δεδομένων μας και παρουσιάζεται ο τρόπος συλλογής τους. Τέλος, θα αναφερθούμε στην προεπεξεργασία που έγινε στα δεδομένα με σκοπό την εξαγωγή χαρακτηριστικών, την βελτίωση της ποιότητας των δεδομένων μέσω εύρεσης κενών τιμών και άλλων ελλείψεων στα δεδομένα και την κατάλληλη μετατροπή των δεδομένων με στόχο την προετοιμασία τους για φόρτωση σε μοντέλα μηχανικής μάθησης.

4.1 Συλλογή δεδομένων

Στην ενότητα αυτή παρουσιάζεται η πηγή από την οποία αποκτήσαμε τα δεδομένα. Ειδικότερα, χρησιμοποιήσαμε το Dataset "Commercial and Residential Hourly Load Data Set - USA" [62], το οποίο εντοπίσαμε μέσω της ιστοσελίδας της IEEE, IEEE PES SUBCOMMITTEE ON BIG DATA & ANALYTICS FOR POWER SYSTEM. Επιπλέον, το σύνολο δεδομένων μας διατίθεται μέσω της ιστοσελίδας του Open Energy Information (OpenEI). Το OpenEI είναι μια διαδικτυακή πλατφόρμα που αναπτύχθηκε από το Εθνικό Εργαστήριο Ανανεώσιμης Ενέργειας (National Renewable Energy Laboratory [NREL]) στις Ηνωμένες Πολιτείες. Λειτουργεί ως μια περιεκτική πηγή δεδομένων, εργαλείων και πόρων που σχετίζονται με την ανανεώσιμη ενέργεια, την ενεργειακή απόδοση και τη βιώσιμη ανάπτυξη. Το συγκεκριμένο σύνολο δεδομένων αποτελείται από αρχεία csv δομημένα σε φακέλους. Η λήψη των αρχείων αυτών έγινε μέσω κάποιων συμπιεσμένων φακέλων και έπειτα έγινε η εξαγωγή τους. Τόσο οι ονομασίες των φακέλων όσο και οι ονομασίες των αρχείων προσφέρουν πληροφορία σχετικά με την κατηγοριοποίηση των δεδομένων. Αυτή η κατηγοριοποίηση όμως των δεδομένων, χρησιμοποιώντας μόνο τα ονόματα των φακέλων και των αρχείων δεν είναι τόσο ξεκάθαρη οπότε αποφασίσαμε να οργανώσουμε τα αρχεία μας με έναν αρκετά διαφορετικό τρόπο σε μορφή φακέλων ώστε η κατηγοριοποίησή τους να γίνει ξεκάθαρη και ακριβής όπως θα εξηγηθεί στο κεφάλαιο 1.3.5 Τελική δομή δεδομένων. Αυτή η οργάνωση των δεδομένων πέρα από την ξεκάθαρη κατηγοριοποίηση που προσέφερε, βοήθησε και στην εύρεση ελλείψεων που μπορεί να υπήρχαν στα δεδομένα μας. Σε αυτή την μορφή όλοι οι φάκελοι θα έπρεπε να έχουν δεδομένα και στην περίπτωση που κάποιος εξ' αυτών δεν είχε μπορούσαμε να συμπεράνουμε πως υπάρχουν ελλείψεις, γεγονός που δεν θα μπορούσε να επιτευχθεί τόσο εύκολα χρησιμοποιώντας την αρχική οργάνωση των δεδομένων.

4.2 Παρουσίαση δεδομένων

Στην ενότητα αυτή παρέχονται επιπλέον λεπτομέρειες σχετικά με τα δεδομένα που χρησιμοποιήθηκαν στη μελέτη μας. Συγκεκριμένα τα δεδομένα περιλαμβάνουν ωριαία μέτρηση κατανάλωσης ενέργειας σε όλες τις πολιτείες των Ηνωμένων Πολιτειών της Αμερικής. Αυτά τα δεδομένα καλύπτουν τόσο την εμπορική όσο και την οικιακή κατανάλωση ενέργειας. Κατά τη διάρκεια του πειράματός μας, επικεντρωθήκαμε ειδικότερα στην ανάλυση των δεδομένων που αφορούν την ενεργειακή κατανάλωση σε εμπορικά κτίρια.

4.2.1 Περιγραφή μεταβλητών συνόλου δεδομένων

Τα δεδομένα, το οποία αναφέρονται σε κτήρια εμπορικής χρήσης, όπως αναφέρθηκε προηγουμένως, περιλαμβάνουν ωριαίες μετρήσεις για τη διάρκεια ενός έτους, συγκεκριμένα το διάστημα μεταξύ του 2014 και του 2015. Αυτά τα δεδομένα καλύπτουν διάφορα είδη κτιρίων, δύο πηγές ενέργειας (φυσικό αέριο και ηλεκτρισμό) και διάφορες χρήσεις ενέργειας.

Παρακάτω παρουσιάζεται ένας πίνακας που περιέχει όλα τα διαφορετικά χαρακτηριστικά που καταγράφηκαν μέσω των δεδομένων μας:

Πίνακας 4.1: Χαρακτηριστικά και περιγραφή του συνόλου δεδομένων

| Μεταβλητές Συνόλου Δεδομένων |
|----------------------------------|
| Ημερομηνία και Ώρα των δεδομένων |
| Πολιτεία των Ηνωμένων Πολιτειών |
| Είδος Κτιρίου |
| Τύπος Ενέργειας |
| Τύπος Χρήσης Ενέργειας |
| Κατανάλωση Ενέργειας (KWh) |

Ημερομηνία και ώρα των δεδομένων

Αυτό το χαρακτηριστικό περιλαμβάνει πληροφορίες για την ημερομηνία και την ώρα κατά την οποία πραγματοποιήθηκαν οι μετρήσεις.

Πολιτεία των Ηνωμένων Πολιτειών

Αυτό το χαρακτηριστικό παρέχει πληροφορίες σχετικά με την πολιτεία των Ηνωμένων Πολιτειών στην οποία αναφέρονται τα δεδομένα. Συγκεκριμένα παρέχονται δεδομένα για τις 50 πολιτείες των Ηνωμένων Πολιτειών της Αμερικής

Είδος κτιρίων

Μέσω αυτού του χαρακτηριστικού είχαμε πληροφορία σχετικά με το είδος του κτιρίου. Συγκεκριμένα έχουμε 16 διαφορετικά είδη κτιρίων . Τα οποία είναι τα εξής:

- Νοσοκομείο (Hospital)
- Εστιατόριο (Restaurant)

- Εστιατόρια ταχείας εξυπηρέτησης (Quick Service Restaurant)
- Μικρό γραφείο (Small Office)
- Μεσαίο γραφείο (Medium Office)
- Μεγάλο γραφείο (Large Office)
- Μικρό ξενοδοχείο (Small Hotel)
- Μεγάλο ξενοδοχείο (Large Hotel)
- Διαμέρισμα με 5 έως 10 ορόφους (Midrise Apartment)
- Κτίριο εξωτερικής φροντίδας νοσοκομείου (Out Patient)
- Δημοτικό σχολείο (Primary School)
- Σχολείο δευτεροβάθμιας εκπαίδευσης (Secondary School)
- Αυτόνομο κατάστημα λιανικής (Stand-aloneRetail)
- Σειρά καταστημάτων γύρω από χώρο στάθμευσης (Strip Mall)
- Σουπερμάρκετ (SuperMarket)
- Αποθήκη (Warehouse)

Από αυτά τα 16 διαφορετικά είδη κτιρίων αποφασίσαμε να ασχοληθούμε με 8 από αυτά. Η επιλογή αυτή έγινε μετά από ανάλυση των δεδομένων. Συγκεκριμένα παρατηρήσαμε πως στα περισσότερα είδη κτιρίων που αποκλείσαμε υπήρχαν ελλείψεις στα δεδομένα. Ιδιαίτερα σε μερικά από αυτά τα είδη κτιρίων παρατηρήσαμε ότι δεν υπήρχαν όλες οι διαφορετικές χρήσεις ενέργειας. Έτσι επιλέξαμε να κρατήσουμε ένα υποσύνολο από τα δεδομένα και να ασχοληθούμε με τα εξής 8 είδη κτιρίων:

- Νοσοκομείο (Hospital)
- Εστιατόριο (Restaurant)
- Εστιατόριο ταχείας εξυπηρέτησης (Quick Service Restaurant)
- Μικρό γραφείο (Small Office)
- Μεσαίο γραφείο (Medium Office)
- Μεγάλο γραφείο (Large Office)
- Μικρό ξενοδοχείο (Small Hotel)
- Μεγάλο ξενοδοχείο (Large Hotel)

Τύπος Ενέργειας

Αυτό το χαρακτηριστικό μας προσφέρει την πληροφορία σχετικά με το είδος της ενέργειας και αν οι τιμές μας αναφέρονται σε ηλεκτρισμό ή σε φυσικό αέριο

Τύπος Χρήσης Ενέργειας

Μέσω αυτού του χαρακτηριστικού είχαμε πληροφορία σχετικά με τον τύπο χρήσης της ενέργειας. Ιδιαίτερα είχαμε 7 διαφορετικά είδη χρήσης ενέργειας. Τα διάφορα είδη χρήσης ενέργειας είναι τα εξής:

- Συνολική κατανάλωση ενέργειας (Facility)
- Κατανάλωση ενέργειας σε συστήματα ψύξης (Cooling)
- Κατανάλωση ενέργειας σε συστήματα θέρμανσης (Heating)
- Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό (Interior Equipment)
- Κατανάλωση ενέργειας σε ανεμιστήρες (Fans)
- Κατανάλωση ενέργειας σε θερμοσίφωνα (Water Heater)
- Κατανάλωση ενέργειας σε εσωτερικό φωτισμό (Interior Lights)

Από τα παραπάνω 7 διαφορετικά είδη χρήσης ενέργειας επικεντρωθήκαμε στα 4 από αυτά. Εδώ η επιλογή αυτή έγινε καθώς παρατηρήσαμε όπως και στην περίπτωση των κτιρίων ότι σε πολλές περιπτώσεις δεν είχαμε δεδομένα για τις εναπομείναντες κατηγορίες. Έτσι καταλήξαμε με τα εξής είδη χρήσης ενέργειας:

- Συνολική κατανάλωση ενέργειας (Facility)
- Κατανάλωση ενέργειας σε συστήματα ψύξης (Cooling)
- Κατανάλωση ενέργειας σε συστήματα θέρμανσης (Heating)
- Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό (Interior Equipment)

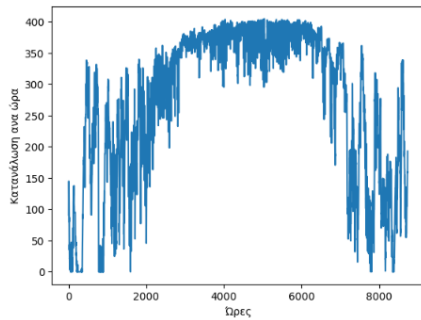
Τιμές των δεδομένων (KWh)

Πληροφορία σχετικά με τις τιμές των δεδομένων οι οποίες τιμές μετρούνται σε KWh. Ακόμη είναι σημαντικό να πούμε πως οι τιμές αυτές είναι τύπου float64.

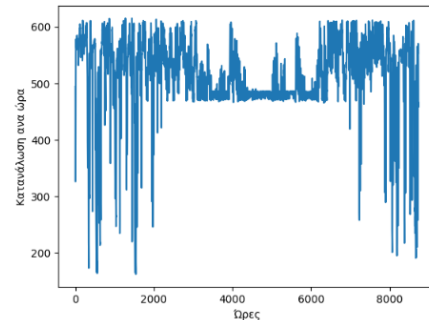
4.2.2 Οπτικοποίηση δεδομένων

Παρακάτω θα παρουσιάσουμε δείγματα από το σύνολο δεδομένων. Συγκεκριμένα θα οπτικοποιήσουμε διάφορες χρονοσειρές οι οποίες αντιστοιχούν σε διαφορετικά κτήρια και χρήσεις ενέργειας. Σκοπός μας είναι να κατανοήσουμε σε βάθος την μορφή των δεδομένων μας.

Διαφοροποίηση στην Πολιτεία της Αμερικής



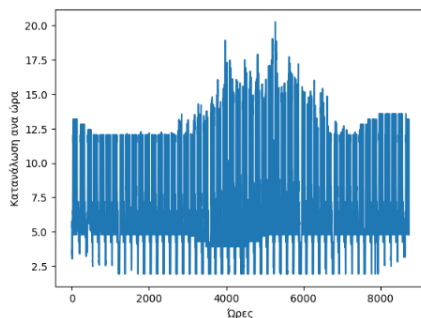
Σχήμα 4.1: Νοσοκομείο στην Αλάσκα, κατανάλωση ηλεκτρισμού για ψύξη



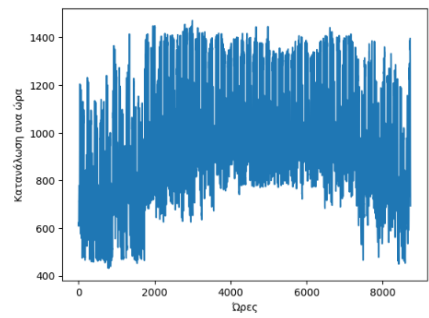
Σχήμα 4.2: Νοσοκομείο στην Αλαμπάμα, κατανάλωση ηλεκτρισμού για ψύξη

Όπως μπορούμε να παρατηρήσουμε από τα δείγματα, η πολιτεία στην οποία αναφέρονται τα δεδομένα μπορεί να διαδραματίσει σημαντικό ρόλο. Αυτό μπορεί να οφείλεται σε διάφορους παράγοντες. Αρχικά οι διαφορές στο κλίμα μεταξύ των πολιτειών μπορούν να έχουν σημαντική επίδραση στην κατανάλωση ενέργειας. Για παράδειγμα, πολιτείες με πιο θερμό κλίμα μπορεί να χρειάζονται περισσότερη ενέργεια για τον κλιματισμό των κτιρίων τους κατά τους θερινούς μήνες. Ακόμη ένας σημαντικός παράγοντας είναι η πληθυσμιακή σύνθεση της εκάστοτε πολιτείας που μπορούν να διαμορφώσουν διαφορετικές ανάγκες και συνήθειες όσον αφορά την ενεργειακή χρήση. Τέλος η οικονομική δραστηριότητα της κάθε πολιτείας της Αμερικής μπορεί να επιφέρει αλλαγές στην κατανάλωση ενέργειας.

Διαφοροποίηση στο είδος του κτιρίου



Σχήμα 4.3: Μικρό γραφείο στην Μινεσότα, συνολική κατανάλωση ηλεκτρισμού

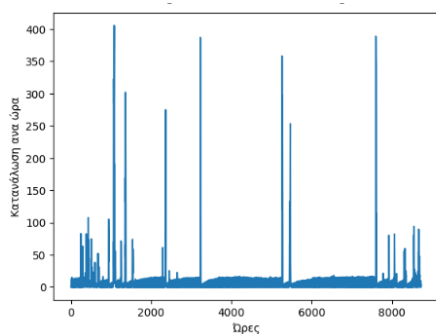


Σχήμα 4.4: Νοσοκομείο στην Μινεσότα, συνολική κατανάλωση αερίου

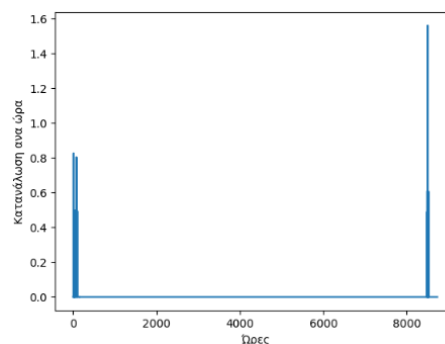
Εδώ μπορούμε να παρατηρήσουμε αρκετά μεγάλες διαφορές στην μορφή των χρονοσειρών μας. Αυτό μπορεί να αποδοθεί στο γεγονός πως τα διαφορετικά είδη κτιρίων χρησιμοποιούνται και για διαφορετικούς σκοπούς, έτσι οι διαφορετικές δραστηριότητες εντός του κτιρίου επηρεάζουν τον τρόπο που χρησιμοποιείται η ενέργεια. Τα κτίρια με διαφορετική χρήση μπορεί να έχουν διαφορετικά ωράρια λειτουργίας. Για παράδειγμα, ένα γραφείο μπορεί να είναι ανοιχτό μόνο κατά τις εργάσιμες ώρες, ενώ ένα ξενοδοχείο λειτουργεί συνεχώς. Ακόμη οι διαφορετικές ομάδες ανθρώπων που χρησιμοποιούν διαφορετικά κτίρια μπορεί να

έχουν διαφορετικές ανάγκες όσον αφορά την άνεση και τις συνθήκες διαβίωσης. Μπορούμε εύκολα λοιπόν να συμπεράνουμε ότι η διαφοροποίηση στο είδος του κτιρίου μπορεί να φέρει μεγάλες αλλαγές στα πρότυπα κατανάλωσης της ενέργειας.

Διαφοροποίηση στο είδος της ενέργειας



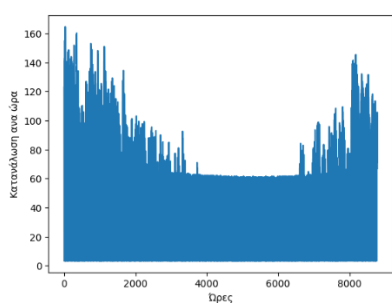
Σχήμα 4.5: Μεγάλο ξενοδοχείο στην Φλόριδα, κατανάλωση αερίου για θέρμανση



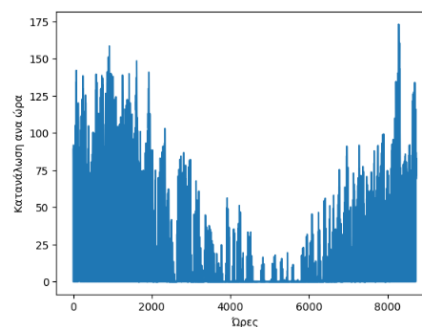
Σχήμα 4.6: Μεγάλο ξενοδοχείο στην Φλόριδα, κατανάλωση ηλεκτρισμού για θέρμανση

Παρατηρώντας τα παραπάνω δείγματα των δεδομένων μας μπορούμε να συμπεράνουμε πως και η μορφή της ενέργειας που καταναλώνεται μπορεί να σχηματίσει σημαντικές διαφορές στα δεδομένα μας. Τέτοιες διαφορές μπορεί να ευθύνονται στις ανάγκες των καταναλωτών ή λόγω των διαφορών που μπορεί να υπάρχουν στην τιμή και την διαθεσιμότητα του κάθε είδους ενέργειας ανά περιόδους.

Διαφοροποίηση στην χρήση ενέργειας



Σχήμα 4.7: Εστιατόριο στην Νέα Υόρκη, συνολική κατανάλωση αερίου



Σχήμα 4.8: Εστιατόριο στην Νέα Υόρκη, κατανάλωση αερίου για θέρμανση

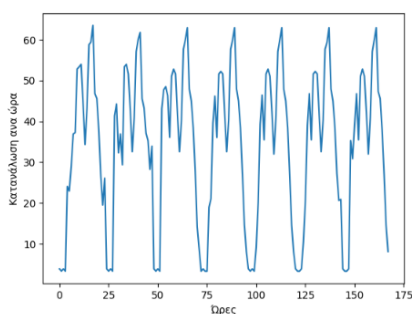
Η διαφοροποίηση στην χρήση της ενέργειας είναι μια σημαντική αλλαγή που αλλάζει σημαντικά τις χρονοσειρές τις οποίες προσπαθούμε να μελετήσουμε και να προβλέψουμε. Εύκολα μπορούμε να παρατηρήσουμε από τις δύο παραπάνω χρονοσειρές πως υπάρχουν διαφορές που σχετίζονται με την εποχικότητα της κάθε χρήσης. Για παράδειγμα η ενέργεια που καταναλώνεται για θέρμανση ή για ψύξη όπως είναι λογικό επηρεάζεται από τις κλιματικές αλλαγές κατά την διάρκεια του έτους. Ακόμη οι ανάγκες της κάθε χρήσης ενέργειας

είναι διαφορετικές και συνεπώς μπορούμε να παρατηρήσουμε αλλαγές στην μέγιστη κατανάλωση κάθε χρονοσειράς. Μια χρονοσειρά που αναφέρεται στην συνολική κατανάλωση θα παρουσιάζει αρκετά μεγαλύτερες τιμές στο σύνολο της σε σχέση με μία που αναφέρεται σε κατανάλωση ενέργειας στον εσωτερικό εξοπλισμό

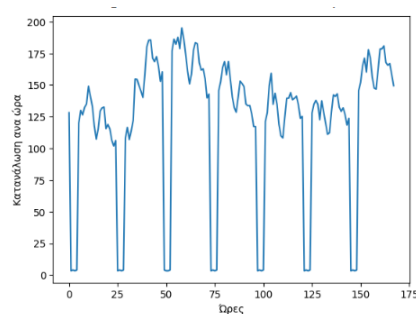
Συνοψίζοντας, μπορούμε εύκολα να κατανοήσουμε ότι η διαφορετική φύση των κτιρίων και οι διαφορετικές χρήσεις ενέργειας επηρεάζουν σημαντικά τα μοτίβα και τις τάσεις των δεδομένων. Κάθε είδος κτιρίου έχει τις δικές του ιδιαιτερότητες και απαιτήσεις σε ενέργεια. Επίσης, οι διαφορετικές χρήσεις ενέργειας, όπως η θέρμανση, η ψύξη, ο εσωτερικός εξοπλισμός και η συνολική κατανάλωση ενέργειας, δημιουργούν διαφορετικά πρότυπα κατανάλωσης. Ακόμη διαφορές στην πολιτεία στην οποία αναφέρονται τα δεδομένα ή στον τύπο ενέργειας που χρησιμοποιείται μπορούν να φέρουν διαφορές στα δεδομένα μας. Αυτή η ποικιλία στις χρονοσειρές τονίζει τη σημασία της προσέγγισης και της ανάλυσης των δεδομένων λαμβάνοντας υπόψη τη διαφοροποίηση αυτών των παραγόντων.

Οπτικοποίηση δεδομένων ανά εβδομάδα

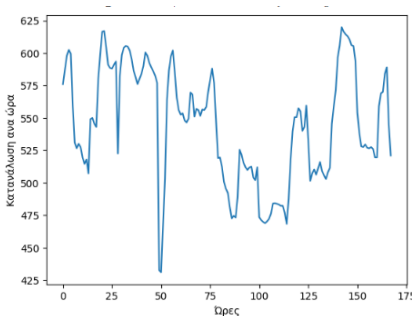
Για να μπορέσουμε να κατανοήσουμε ακόμη περισσότερο την πολυπλοκότητα των δεδομένων μας παρακάτω παρέχονται μερικές εβδομάδες από το σύνολο των δεδομένων μας.



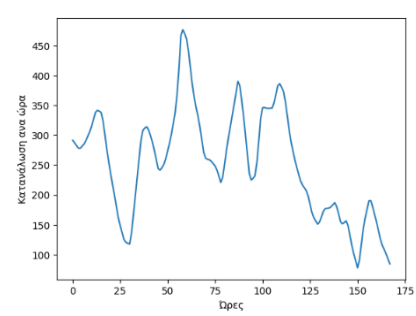
Σχήμα 4.9: Εβδομάδα καλοκαιριού για συνολική κατανάλωση αερίου



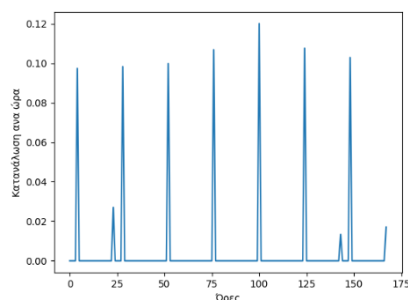
Σχήμα 4.10: Εβδομάδα χειμώνα για συνολική κατανάλωση αερίου



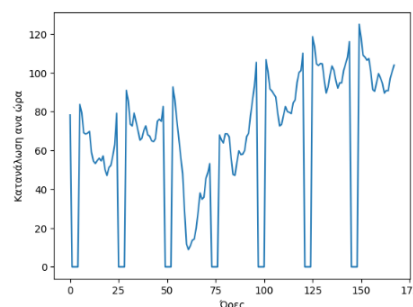
Σχήμα 4.11: Εβδομάδα καλοκαιριού για κατανάλωση ηλεκτρισμού για ψύξη



Σχήμα 4.12: Εβδομάδα χειμώνα για κατανάλωση ηλεκτρισμού για ψύξη



Σχήμα 4.13: Εβδομάδα καλοκαιριού για κατανάλωση αερίου για θέρμανση



Σχήμα 4.14: Εβδομάδα χειμώνα για κατανάλωση αερίου για θέρμανση

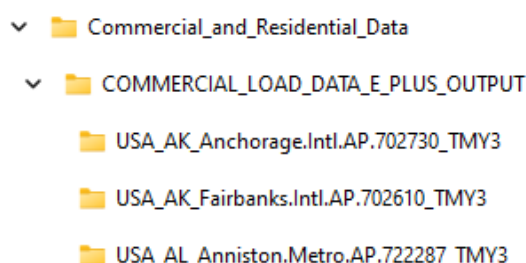
Παρατηρώντας τα παραπάνω δείγματα μπορούμε με ευκολία να καταλάβουμε την τάση και την εποχιακότητα που υπάρχει στα δεδομένα μας τόσο σε βαθμό μήνα, εβδομάδας ακόμη και ημέρας. Για παράδειγμα στις περιπτώσεις της θέρμανσης και ψύξης βλέπουμε πως ανάλογα τον μήνα που αναφέρονται τα δεδομένα οι τιμές αυτών των δεδομένων αλλάζουν σημαντικά. Ακόμη, στην περίπτωση της συνολικής κατανάλωσης μπορούμε εύκολα να δούμε την εποχιακότητα που υπάρχει ανά ημέρα, όπως επίσης παρατηρώντας τις δύο διαφορετικές εβδομάδες μπορούμε να παρατηρήσουμε ότι υπάρχει εποχιακότητα και ανά εβδομάδα.

4.3 Προεπεξεργασία δεδομένων

Στην ενότητα αυτή παρέχονται λεπτομερείς πληροφορίες σχετικά με τις διάφορες τροποποιήσεις που πραγματοποιήθηκαν στα δεδομένα προκειμένου να εξαχθούν τα απαραίτητα χαρακτηριστικά καθώς και λοιπές τροποποιήσεις με σκοπό την βελτίωση της ποιότητας των δεδομένων μας.

4.3.1 Αρχική δομή δεδομένων

Τα δεδομένα μας βρίσκονταν σε μορφή αρχείων CSV και οργανωμένα σε διάφορους φακέλους. Στην παρακάτω εικόνα μπορούμε να παρατηρήσουμε ακριβώς την δομή των δεδομένων.



Σχήμα 4.15: Αρχική δομή δεδομένων

Συνολικά τα δεδομένα μας έχουν 936 διαφορετικούς φακέλους που αναφέρονται σε διαφορετικές πολιτείες των Ηνωμένων Πολιτειών της Αμερικής. Φυσικά για μεγαλύτερες πληθυσμιακά πολιτείες είχαμε περισσότερα δεδομένα συγκριτικά με μικρότερες πολιτείες. Ένα

γεγονός που δυσχεραίνει την δυνατότητα μας να προβλέψουμε όλες τις πολιτείες με μεγάλη ακρίβεια, μιας και ο αριθμός των δεδομένων που έχουμε για να εκπαιδεύσουμε το μοντέλο είναι σημαντικός παράγοντας για την απόδοσή του.

Μέσα σε αυτούς τους φακέλους υπάρχουν τα διάφορα αρχεία CSV. Συγκεκριμένα σε κάθε φάκελο έχουμε 16 διαφορετικά αρχεία, ένα για κάθε διαφορετικό είδος κτιρίου. Τα αρχεία αυτά έχουν χαρακτηριστικό όνομα με σκοπό των διαχωρισμό τους. Για παράδειγμα μερικά ονόματα αρχείων είναι ως εξής:

- **"RefBldgFullServiceRestaurantNew2004_v1.3_7.1_8A_USA_AK_FAIRBANKS"**
- **"RefBldgHospitalNew2004_v1.3_7.1_8A_USA_AK_FAIRBANKS"**
- **"RefBldgLargeHotelNew2004_v1.3_7.1_8A_USA_AK_FAIRBANKS"**
- **"RefBldgLargeOfficeNew2004_v1.3_7.1_8A_USA_AK_FAIRBANKS"**

Παρακάτω παρουσιάζεται ένα δείγμα της αρχικής μορφής που υπήρχε στα αρχεία CSV.

| Date/Time | Electricity:Facility [kW](Hourly) | Fans:Electricity [kW](Hourly) | Cooling:Electricity [kW](Hourly) | Heating:Electricity [kW](Hourly) | InteriorLights:Electricity [kW](Hourly) | InteriorEquipment:Electricity [kW](Hourly) |
|---------------------|--------------------------------------|----------------------------------|-------------------------------------|-------------------------------------|--|---|
| 0 01/01 01:00:00 | 22.035977 | 3.586221 | 0.0 | 0.0 | 4.589925 | 8.1892 |
| 1 01/01 02:00:00 | 14.649757 | 0.000000 | 0.0 | 0.0 | 1.529975 | 7.4902 |
| 2 01/01 03:00:00 | 14.669567 | 0.000000 | 0.0 | 0.0 | 1.529975 | 7.4902 |
| 3 01/01 04:00:00 | 14.677808 | 0.000000 | 0.0 | 0.0 | 1.529975 | 7.4902 |
| 4 01/01 05:00:00 | 14.824794 | 0.000000 | 0.0 | 0.0 | 1.529975 | 7.4902 |

Σχήμα 4.16: Αρχική μορφή αρχείων CSV

Παρατηρώντας το παραπάνω δείγμα μπορούμε αρχικά να δούμε πως υπάρχουν στήλες για κάθε είδος χρήσης ενέργειας και τον τύπο της ενέργειας στον οποίο αναφέρεται η εκάστοτε χρήση, καθώς και μία στήλη που αναφέρεται στην ώρα και ημερομηνία των δεδομένων.

4.3.2 Επεξεργασία δεδομένων με σκοπό την εξαγωγή χαρακτηριστικών

Εξαγωγή είδους κτιρίου

Όπως παρατηρήσαμε και στην παραπάνω περιγραφή των αρχικών μας δεδομένων, είδαμε πως το είδος του κτιρίου δεν υπήρχε μέσα στα τελικά αρχεία, παρά μόνο στα ονόματα των αρχείων. Για παράδειγμα, το αρχείο με όνομα **"RefBldgHospitalNew2004_v1.3_7.1_8A_USA_AK_FAIRBANKS"** αναφέρεται σε Νοσοκομείο(Reference Building Hospital). Θεωρήσαμε αναγκαίο να υπάρχει το είδος του κτιρίου εντός των αρχείων καθώς με αυτήν την προσθήκη βεβαιωθήκαμε ότι όλες οι πληροφορίες που απαιτούνται για την περιγραφή του κάθε αρχείου είναι διαθέσιμες εντός του ίδιου του αρχείου, κάνοντας έτσι τη διαχείριση των δεδομένων μας πιο ευέλικτη και κατανοητή για μετέπειτα αναλύσεις και μοντέλα πρόβλεψης. Για να επιτύχουμε αυτό, δημιουργήσαμε έναν αλγόριθμο που εξάγει το πρώτο μέρος του ονόματος κάθε αρχείου και το χρησιμοποιεί για να αναθέσει το κατάλληλο είδος κτιρίου σε μια νέα στήλη.

Εξαγωγή πολιτείας της Αμερικής

Όμοια με το είδος του κτιρίου, η πληροφορία σχετικά με την πολιτεία που αναφέρονται τα δεδομένα υπήρχε μόνο μέσω συντομογραφίας στο όνομα των φακέλων. Όπως μπορούμε να δούμε στο εξής όνομα φακέλου "USA_AK_Anchorage.Int.AP.702730_TMY3" παρατηρούμε ότι υπάρχει η αναφορά στην πολιτεία μέσω της συντομογραφίας **USA_AK**. Έτσι διασχίζοντας όλους τους φακέλους και ταυτίζοντας την συντομογραφία της πολιτείας με το ολόκληρο όνομα αυτής προσθέσαμε εκ νέου μια νέα στήλη σε κάθε αρχείο που υπήρχε εντός του φακέλου, η οποία συμπεριλάμβανε το όνομα της πολιτείας στην οποία αναφέρονται τα δεδομένα.

Εξαγωγή είδους ενέργειας και χρήσης ενέργειας

Πέρα από την προφανή ανάγκη μας να προσθέσουμε όλα τα δυνατά χαρακτηριστικά εντός των αρχείων αποφασίσαμε πως κάθε αρχείο θα έπρεπε να αναφέρεται σε ένα μόνο είδος ενέργειας και μία μόνο χρήση ενέργειας. Έτσι, διαχωρίσαμε κάθε αρχείο σε 10 διαφορετικά, όσοι και οι διαφορετικοί συνδυασμοί είδους και χρήσης ενέργειας. Για να πετύχουμε αυτή την μετατροπή και να συνεχίσουμε να έχουμε ολόκληρη την πληροφορία εντός των αρχείων διαχωρίσαμε κάθε αρχείο παίρνοντας την στήλη για την ημερομηνία και ώρα, την στήλη για την πολιτεία και το είδος κτιρίου, καθώς και μία στήλη που να περιέχει τις τιμές κατανάλωσης ενέργειας. Φυσικά η πληροφορία σχετικά με το είδος ή και την χρήση ενέργειας έπρεπε να υπάρχει εντός του νέου αρχείου. Για αυτό τον λόγο χρησιμοποιώντας το όνομα της στήλης προσθέσαμε δύο νέες στήλες στο αρχείο που μας έδιναν την απαραίτητη πληροφορία. Για παράδειγμα όταν δημιουργήσαμε ένα νέο αρχείο χρησιμοποιώντας τις τιμές τις εξής στήλης "Electricity:Facility [kW](Hourly)" προσθέσαμε στην στήλη που αναφέρεται στο είδος ενέργειας την τιμή "Ηλεκτρισμό" (Electricity) και στην στήλη για την χρήση της ενέργειας την τιμή "Συνολική κατανάλωση ενέργειας" (Facility).

4.3.3 Επεξεργασία δεδομένων με σκοπό την βελτίωση της ποιότητας τους

Μετατροπή δεδομένων ανά εβδομάδα

Όπως αναφέραμε προηγουμένως τα δεδομένα μας μετά την εξαγωγή των απαραίτητων χαρακτηριστικών αποτελούνταν από μερικές στήλες που παρείχαν λεπτομέρειες σχετικά με αυτά, καθώς και μία στήλη με αριθμητικές τιμές που αναφερόταν στην ανά ώρα κατανάλωση ενέργειας. Με σκοπό την βελτίωση των δεδομένων μας αποφασίσαμε να αλλάξουμε αυτή την μορφή. Συγκεκριμένα μετατρέψαμε τα δεδομένα μας ώστε κάθε γραμμή του αρχείου μας να περιέχει τις αριθμητικές τιμές κατανάλωσης ενέργειας για μία εβδομάδα. Έτσι σε κάθε γραμμή προσθέσαμε 168 νέες στήλες, μια για κάθε ώρα εντός μιας εβδομάδας. Με τον τρόπο αυτό καταφέραμε να παρέχουμε μια καλύτερη οργάνωση των δεδομένων μας καθώς τα εβδομαδιαία δεδομένα με ωριαίες στήλες μπορούν να είναι πιο εύκολα στην ερμηνεία από τις μεμονωμένες ωριαίες γραμμές. Μετατρέποντας έτσι τα δεδομένα παρέχουμε μια έτοιμη είσοδο για τα νευρωνικά δίκτυα σε εβδομάδες, γεγονός που μας βοηθάει να έχουμε μεγαλύτερη ακρίβεια και λάβουμε υπόψιν μας την εποχικότητα σε επίπεδο ημέρας και ώρας. Επίσης, τα μοτίβα και οι τάσεις κατά τις ώρες μέσα σε μια εβδομάδα μπορεί να είναι πιο εμφανείς όταν εξετάζονται με αυτόν τον τρόπο. Έτσι κατά συνέπεια, καταφέραμε να

προσφέρουμε στα μετέπειτα μοντέλα μας μια μορφή δεδομένων, με την οποία θα είναι πιο εύκολο να εκπαιδευτούν και να μάθουν αποτελεσματικά από τα δεδομένα.

Διαγραφή τελευταίας εβδομάδας

Εκτελώντας την παραπάνω επεξεργασία στα δεδομένα μας καταλήξαμε να έχουμε σε κάθε αρχείο 53 γραμμές, μία για κάθε εβδομάδα του έτους. Όπως είναι γνωστό όμως κάθε ημερολογιακό έτος έχει 365 μέρες και κατά συνέπεια μια μόνο μέρα παραπάνω από τις 52 εβδομάδες. Αρχίσαμε την αρίθμηση των δεδομένων από την πρώτη ώρα τους έτους, ανεξάρτητα από το ποια μέρα ήταν και έτσι είχαμε διαθέσιμα δεδομένα για τις 52 πρώτες εβδομάδες. Αυτό δεν συνέβη για την τελευταία εβδομάδα καθώς περιείχε πληροφορία μόνο για μία ημέρα. Αυτό δημιουργούσε πρόβλημα στα δεδομένα μας, μιας και η τελευταία εβδομάδα αποτελούνταν σχεδόν εξολοκλήρου από μηδενικά. Για να αποτρέψουμε να δημιουργήσουμε ψευδές πληροφορίες για τα δεδομένα μας αποφασίσαμε να διαγράψουμε την τελευταία εβδομάδα, ώστε να διατηρήσουμε μια ολοκληρωμένη και ακριβή εικόνα για τις υπόλοιπες 52 εβδομάδες του έτους.

Επεξεργασία κενών τιμών

Κατά την ανάλυση των δεδομένων μας, παρατηρήσαμε ότι κάποιες τιμές ήταν κενές (NaN). Συγκεκριμένα παρατηρήσαμε πως είχαμε στα δεδομένα ένα μικρό ποσοστό κενών τιμών της τάξης του 1,67%. Αυτές οι τιμές συνήθως υπήρχαν μόνο στις τελευταίες ώρες κάποιας εβδομάδας και συνήθως ήταν είτε στην αρχή ή στο τέλος ενός χρόνου. Αντί να απορρίψουμε αυτές τις τιμές και να χάσουμε έτσι πληροφορία, αποφασίσαμε να χρησιμοποιήσουμε μια τεχνική που ονομάζεται γραμμική παρεμβολή (linear interpolation) για να συμπληρώσουμε αυτές τις τιμές. Αυτή η τεχνική επιτρέπει την εκτίμηση των NaN τιμών, λαμβάνοντας υπόψη την τάση ή τη γραμμική συμπεριφορά των γύρω δεδομένων. Με αυτόν τον τρόπο, μπορέσαμε να αξιοποιήσουμε πλήρως τα διαθέσιμα δεδομένα μας και να αποφύγουμε την απώλεια πολύτιμης πληροφορίας.

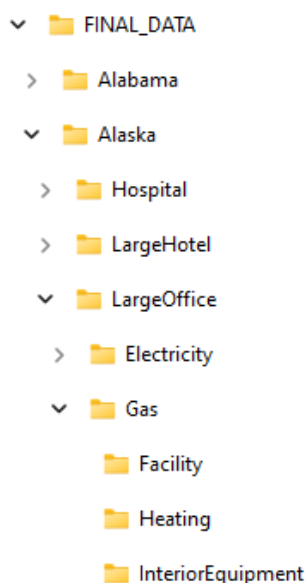
4.3.4 Επεξεργασία λόγω όγκου των δεδομένων

Όπως αναφέραμε και στην περιγραφή του συνόλου δεδομένων, επικεντρωθήκαμε σε ένα υποσύνολο των δεδομένων βάσει τύπου κτιρίου καθώς και είδους ενέργειας. Πέρα από τις διαφορές και ελλείψεις που παρατηρήσαμε σε αυτά τα δεδομένα, αυτό έγινε με σκοπό την μείωση του μεγάλου όγκου των δεδομένων, που θα προκαλούσε αδυναμία φόρτωσης αυτών στα επιμέρους μοντέλα μας. Πέρα από αυτή την μείωση των δεδομένων δημιουργήσαμε έναν αλγόριθμο με τον οποίο καταφέραμε να συμπίεσουμε τα δεδομένα. Συγκεκριμένα γνωρίζουμε ότι οι αριθμητικές τιμές των δεδομένων ήταν τύπου float64, ο οποίος τύπος δεδομένων χρησιμοποιεί και την περισσότερη μνήμη. Έτσι, διαβάσαμε όλα τα δεδομένα μας και ελέγξαμε το εύρος των τιμών σε κάθε γραμμή. Αν αυτό το εύρος μπορούσε να αναπαρασταθεί με κάποιο άλλο τύπο δεδομένων το οποίο να καταλαμβάνει λιγότερη μνήμη(float32, float16), αλλάζαμε τον αρχικό τύπο σε αυτόν. Έτσι καταφέραμε να πετύχουμε μια αρκετή καλή συμπίεση στα δεδομένα μας χωρίς να χάσουμε πληροφορία. Οι παραπάνω ενέργειες συνέβησαν λόγω έλλειψης πόρων σε προσωπικό υπολογιστή. Με την σωστή υποδομή είναι δυνατόν να

μελετήσουμε όλα τα δεδομένα που παρέχονται και να έχουμε μια ακόμη πιο συνολική εικόνα στα αποτελέσματα μας λόγω του μεγαλύτερου όγκου δεδομένων, ως αποτέλεσμα της χρήσης των διαφόρων τύπων κτιρίου καθώς και ειδών χρήσης ενέργειας που δεν χρησιμοποιήσαμε σε αυτό το πείραμα.

4.3.5 Τελική δομή δεδομένων

Μετά από όλες τις παραπάνω μετατροπές στο σύνολο δεδομένων μας καταλήξαμε να έχουμε μια πιο ολοκληρωμένη δομή αυτών. Χρησιμοποιήσαμε μια νέα δομή για να αποθηκεύσουμε τα δεδομένα μας ώστε να είναι ευανάγνωστα και να είναι ξεκάθαρο που αναφέρεται κάθε αρχείο. Συγκεκριμένα καταλήξαμε στην παρακάτω δομή.



Σχήμα 4.17: Τελική δομή δεδομένων

Όπως μπορούμε να παρατηρήσουμε από την παραπάνω εικόνα, είχαμε μια ξεκάθαρη δομή δεδομένων. Αρχικά είχαμε φακέλους για όλες τις διαφορετικές πολιτείες, έπειτα για τα είδη κτιρίων, τον τύπο της ενέργειας και τέλος για το είδος χρήσης της ενέργειας. Εντός των τελευταίων φακέλων υπήρχαν αρχεία csv για τα οποία μπορούσες να λάβεις πληροφορία ακολουθώντας την διαδρομή των φακέλων. Φυσικά οι πληροφορίες αυτές υπάρχουν και εντός των αρχείων. Συγκεκριμένα η τελική μορφή των αρχείων csv είναι η εξής:

| | Weeks | State | BuildingType | Gas/Electricity | Type | Hour 1 | Hour 2 | Hour 3 | Hour 4 | Hour 5 | ... |
|---|--------|---------|--------------|-----------------|---------|--------|--------|--------|--------|--------|-----|
| 0 | Week 0 | Alabama | Hospital | Electricity | Cooling | 494.2 | 351.0 | 326.0 | 326.5 | 346.0 | ... |
| 1 | Week 1 | Alabama | Hospital | Electricity | Cooling | 584.5 | 584.0 | 583.0 | 583.0 | 583.0 | ... |
| 2 | Week 2 | Alabama | Hospital | Electricity | Cooling | 211.2 | 201.6 | 191.1 | 185.4 | 178.6 | ... |
| 3 | Week 3 | Alabama | Hospital | Electricity | Cooling | 564.5 | 543.5 | 483.8 | 428.8 | 392.2 | ... |
| 4 | Week 4 | Alabama | Hospital | Electricity | Cooling | 559.0 | 562.5 | 566.5 | 569.5 | 572.0 | ... |

Σχήμα 4.18: Τελική μορφή αρχείων csv

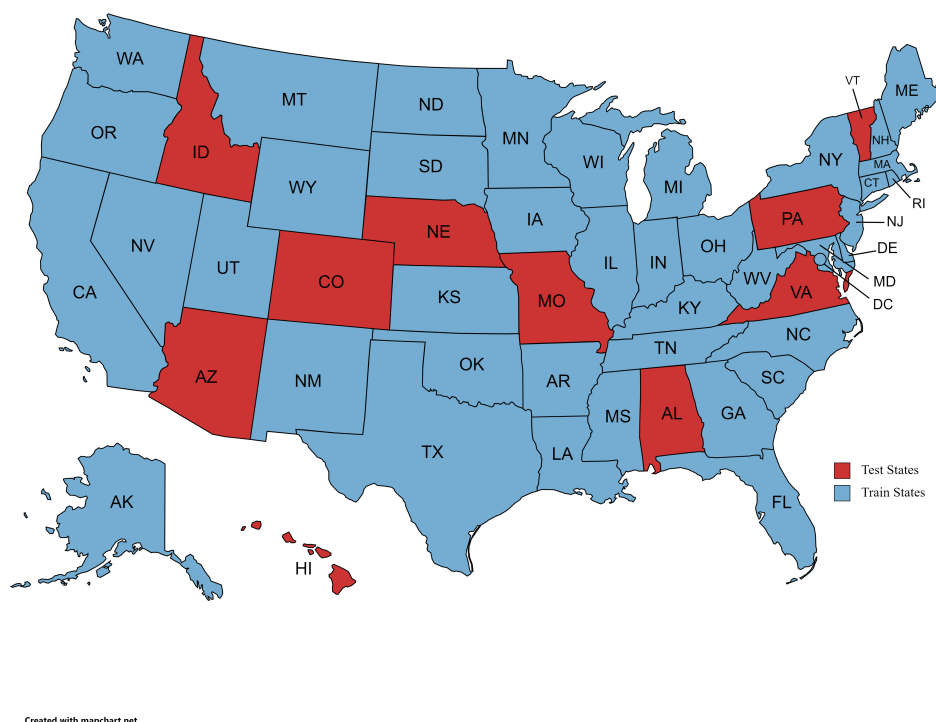
Κεφάλαιο 5

Εφαρμογή μοντέλων πρόβλεψης και αποτελέσματα

Στο κεφάλαιο αυτό παρουσιάζονται αναλυτικά οι τρεις βασικές τεχνολογίες που έχουν σχέση με την εργασία αυτή, δηλαδή τα συστήματα ομότιμων κόμβων, το πλαίσιο RDF και οι γλώσσες ερωτήσεων για RDF.

5.1 Διαχωρισμός δεδομένων

Για να μπορέσουμε να δώσουμε στα μοντέλα που δημιουργήσαμε επαρκής πληροφορία τόσο για την εκπαίδευση των μοντέλων όσο και για τον έλεγχο τους αποφασίσαμε να χωρίσουμε τα δεδομένα μας με ένα ποσοστό 80%-20% όπου το 80% αναφέρεται στο σύνολο των δεδομένων για την εκπαίδευση των μοντέλων ενώ το 20% στο κομμάτι του ελέγχου. Για να πετύχουμε αυτόν τον διαχωρισμό αποφασίσαμε να χωρίσουμε τα δεδομένα βάση πολιτείας των Ηνωμένων Πολιτειών της Αμερικής. Έτσι από τις συνολικά 50 πολιτείες χρησιμοποιήσαμε τις 40 για την εκπαίδευση και τις υπόλοιπες 10 για τον έλεγχο των μοντέλων. Φυσικά η επιλογή των πολιτειών δεν θα μπορούσε να είναι τυχαία. Γνωρίζουμε πως η γεωγραφική τοποθεσία μιας πολιτείας, ο πληθυσμός της καθώς και η γενική της οικονομική ανάπτυξη μπορούν να επηρεάσουν σημαντικά τα δεδομένα μας. Έτσι λαμβάνοντας υπόψιν τις παραπάνω πληροφορίες αποφασίσαμε να κάνουμε τον εξής διαχωρισμό:



Σχήμα 5.1: Χάρτης δεδομένων train και test

Όπως βλέπουμε και από την παραπάνω φωτογραφία οι πολιτείες που επιλέξαμε για το σύνολο των δεδομένων με σκοπό το έλεγχο των μοντέλων μας είναι η Αλαμπάμα, η Αριζόνα, το Κολοράντο, η Χαβάη, το Αϊντάχο, το Μισούρι, η Νεμπράσκα, η Πενσυλβάνια, το Βερμόντ και η Βιρτζίνια. Ενώ όλες οι υπόλοιπες πολιτείες διαμορφώνουν το σύνολο των δεδομένων με σκοπό την εκπαίδευση των μοντέλων

5.2 Κανονικοποίηση δεδομένων

Στο κεφάλαιο για την προεπεξεργασία των δεδομένων, πραγματοποιήσαμε πολλές αλλαγές με στόχο τη βελτίωση και την κατάλληλη προσαρμογή των δεδομένων για τη φόρτωσή τους στα μοντέλα που θα χρησιμοποιήσουμε. Ακόμη μια σημαντική αλλαγή που έπρεπε να γίνει είναι η κανονικοποίηση των δεδομένων. Η κανονικοποίηση είναι σημαντική μιας και μπορεί να βοηθήσει την έρευνα σε διάφορα σημεία. Καταρχάς, η κανονικοποίηση μπορεί να βοηθήσει στην αποφυγή της υπερεκπαίδευσης (overfitting). Αν τα δεδομένα δεν είναι κανονικοποιημένα, το μοντέλο μπορεί να μάθει πολύπλοκες σχέσεις μεταξύ των δεδομένων εκπαίδευσης που δεν είναι γενικεύσιμες, οδηγώντας σε ανακριβή αποτελέσματα. Επιπλέον, μέσω της κανονικοποίησης μπορούμε να επιτύχουμε μια πιο αποτελεσματική εκπαίδευση των δεδομένων, αφού διατηρούνται οι σχέσεις μεταξύ τους σε μια κοινή κλίμακα. Έτσι θεωρήσαμε απαραίτητη την κανονικοποίηση των δεδομένων.

Για την κανονικοποίηση, χρησιμοποιήσαμε τη μέθοδο κανονικοποίησης ελάχιστης-μέγιστης τιμής. Με αυτήν την τεχνική, βρίσκουμε τις μέγιστες και ελάχιστες τιμές του συνόλου δεδομένων μας και στη συνέχεια κλιμακώνουμε τα δεδομένα με βάση την παρακάτω σχέση:

$$X_{\text{κανονικοποιημένο}} = \frac{X - X_{\text{ελάχιστο}}}{X_{\text{μέγιστο}} - X_{\text{ελάχιστο}}}$$

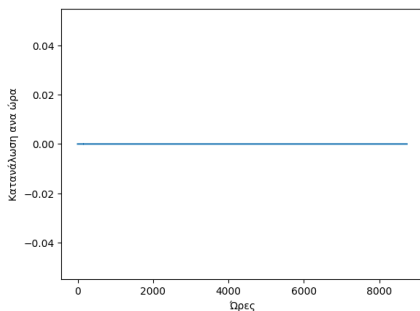
Όπου X είναι η αρχική τιμή, $X_{\text{ελάχιστο}}$ είναι η ελάχιστη τιμή του χαρακτηριστικού και $X_{\text{μέγιστο}}$ είναι η μέγιστη τιμή του χαρακτηριστικού. Αυτή η διαδικασία κλιμάκωσης εξασφαλίζει ότι όλα τα χαρακτηριστικά αντιστοιχούν σε ένα εύρος μεταξύ 0 και 1, με την ελάχιστη τιμή να γίνεται 0 και τη μέγιστη τιμή να γίνεται 1. Διατηρεί τις σχέσεις μεταξύ των δεδομένων αλλά τα μετασχηματίζει σε μία κοινή κλίμακα.

Αποφασίσαμε να κάνουμε αυτή την κανονικοποίηση για κάθε εβδομάδα του έτους για να μπορέσουμε να πετύχουμε πιο ακριβή αποτελέσματα. Οπότε βρίσκοντας το μέγιστο και το ελάχιστο των δεδομένων κάθε εβδομάδας και έπειτα χρησιμοποιώντας την παραπάνω σχέση κανονικοποιήσαμε τα δεδομένα τελειώνοντας έτσι την προετοιμασία των δεδομένων για την φόρτωση τους στα μοντέλα.

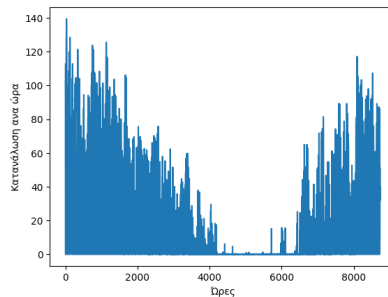
5.3 Μοντέλο Βάσης

Αρχικά για να μπορούμε να έχουμε ένα σημείο αναφοράς ως προς τα αποτελέσματα των μοντέλων, έπρεπε να δημιουργήσουμε κάποιες αρχικές προβλέψεις. Για να το πετύχουμε αυτό χρησιμοποιήσαμε την Naive μέθοδο και ειδικότερα μια μορφή εποχιακής Naive μεθόδου (sNaive). Χρησιμοποιώντας αυτή την τεχνική, πήραμε τις τιμές μιας ολόκληρης εβδομάδας οι οποίες τιμές ήταν και η πρόβλεψη μας για την επόμενη εβδομάδα.

Παρατηρήσαμε πως οι προβλέψεις της sNaive μεθόδου είχαν καλά αποτελέσματα μιας και υπήρχε μια υψηλή συσχέτιση των δεδομένων μιας εβδομάδας με αυτά της προηγούμενης. Πολλές φορές τα δεδομένα για ένα μεγάλο χρονικό διάστημα είναι συνεχώς όμοια. Αυτό συμβαίνει λόγω εποχικότητας ή και χρήσης του είδους ενέργειας. Για παράδειγμα είναι εύκολο να καταλάβει κανείς πως κατά την χειμερινή περίοδο οι τιμές που αφορούν θέρμανση ήταν συνεχώς μηδέν ή πολύ χαμηλές, με αποτέλεσμα το μοντέλο sNaive να έχει ακριβή αποτελέσματα. Ακόμη μια περίπτωση μεγάλης ακρίβειας της sNaive μεθόδου είναι όταν σε κάποιο είδος κτιρίου χρησιμοποιείται ένα από τα δύο είδη ενέργειας για μία συγκεκριμένη χρήση. Για παράδειγμα ένα εστιατόριο μπορεί να χρησιμοποιεί μόνο ηλεκτρισμό για την θέρμανση του κτιρίου, οπότε τα δεδομένα για την θέρμανση με αέριο να είναι συνεχώς μηδέν. Χαρακτηριστικά παρουσιάζουμε δύο χρονοσειρές για ένα εστιατόριο στην Νέα Υόρκη που αφορούν την θέρμανση με την χρήση ηλεκτρισμού και φυσικού αερίου αντίστοιχα



Σχήμα 5.2: Εστιατόριο στην Νέα Υόρκη, κατανάλωση ηλεκτρισμού για θέρμανση



Σχήμα 5.3: Εστιατόριο στην Νέα Υόρκη, κατανάλωση φυσικού αερίου για θέρμανση

Χρησιμοποιώντας την Naive μέθοδο καταλήξαμε στα εξής αποτελέσματα.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 46,07% | 1,408 |

Πίνακας 5.1: Συνοβηκά αποτελέσματα Naive

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 30,65% | 0,526 |
| Εστιατόριο | 32,80% | 0,713 |
| Εστιατόριο ταχείας εξυπηρέτησης | 36,02% | 0,922 |
| Μικρό γραφείο | 60,54% | 2,136 |
| Μεσαίο γραφείο | 85,92% | 1,716 |
| Μεγάλο γραφείο | 52,92% | 1,356 |
| Μικρό ξενοδοχείο | 46,39% | 1,357 |
| Μεγάλο ξενοδοχείο | 47,91% | 2,323 |

Πίνακας 5.2: Αποτελέσματα Naive ανά τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 27,87% | 0,465 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 53,20% | 1,026 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 88,78% | 3,812 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 1,57% | 0,135 |

Πίνακας 5.3: Αποτελέσματα Naive ανά τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 47,85% | 1,644 |
| Φυσικό Αέριο | 45,07% | 1,223 |

Πίνακας 5.4: Αποτελέσματα Naive ανά είδος ενέργειας

5.4 Καθολικό Μοντέλο

Το πρώτο μοντέλο που υλοποιήσαμε είναι ένα καθολικό μοντέλο MLP αρχιτεκτονικής. Το MLP (Multilayer Perceptron) είναι μια από τις πιο βασικές και δημοφιλείς αρχιτεκτονικές νευρωνικών δικτύων στον χώρο της μηχανικής μάθησης και της επεξεργασίας σημάτων. Ανήκει στην κατηγορία των feedforward νευρωνικών δικτύων, που σημαίνει ότι οι πληροφορίες διακινούνται μόνο προς τα εμπρός, από την είσοδο προς την έξοδο. Η βασική δομή του MLP αποτελείται από τρία επίπεδα. Αρχικά από το επίπεδο εισόδου, το οποίο αποτελείται από την είσοδο του δικτύου και δέχεται τις τιμές των χαρακτηριστικών (features) των δεδομένων. Έπειτα από τα κρυφά επίπεδα, τα οποία αποτελούνται από νευρώνες που εκτελούν μη γραμμικές μετασχηματιστικές λειτουργίες στα δεδομένα εισόδου και τέλος το επίπεδο εξόδου που περιέχει την έξοδο του δικτύου και παράγει τις τελικές προβλέψεις.

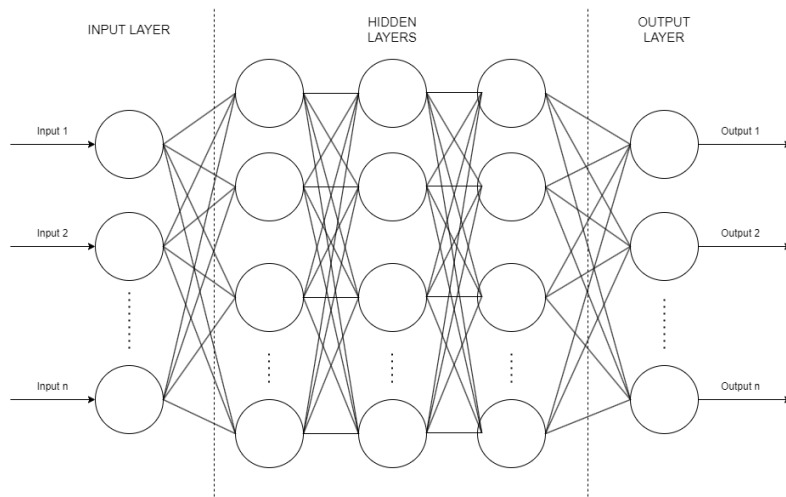
Επιλέξαμε να χρησιμοποιήσουμε ένα μοντέλο MLP με τρία κρυφά επίπεδα. Κάθε κρυφό επίπεδο χρησιμοποιεί τη συνάρτηση ενεργοποίησης ReLU (Rectified Linear Unit), ενώ το επίπεδο εξόδου χρησιμοποιεί γραμμική ενεργοποίηση (activation linear).

Η συνάρτηση ενεργοποίησης ReLU είναι μια μη γραμμική συνάρτηση που εφαρμόζεται στην έξοδο κάθε νευρώνα στα κρυφά επίπεδα. Αυτή η συνάρτηση είναι γνωστή για την ταχύτητα υπολογισμού της και την αποτελεσματικότητά της σε πολλές εφαρμογές.

Το επίπεδο εξόδου χρησιμοποιεί γραμμική ενεργοποίηση καθώς η έξοδος του μοντέλου είναι μια συνεχής μεταβλητή.

Για την εκπαίδευση του μοντέλου μας χρησιμοποιήσαμε τον βελτιστοποιητή Adam (Adam optimizer). Ο βελτιστοποιητής Adam είναι ένας δημοφιλής αλγόριθμος βελτιστοποίησης που συνδυάζει την τεχνική του Momentum με την τεχνική του RMSProp. Αυτό τον κάνει αποτελεσματικό για την εκπαίδευση μοντέλων νευρωνικών δικτύων, καθώς βοηθά στην ταχεία σύγκλιση.

Παρακάτω παρουσιάζεται σχηματικά η αρχιτεκτονική του μοντέλου.



Σχήμα 5.4: Αρχιτεκτονική καθολικού μοντέλου

Η αρχιτεκτονική αυτή του μοντέλου αποτελεί και την βάση μας για όλα τα υπόλοιπα μοντέλα που δημιουργήσαμε. Χρησιμοποιώντας την ίδια αρχιτεκτονική αλλά αλλάζοντας τα δεδομένα εισόδου, τον αριθμό των εκάστοτε μοντέλων καθώς και στην συνέχεια προσθέτοντας κάποια embedding layers καταφέραμε να δημιουργήσουμε τα υπόλοιπα μοντέλα με σκοπό την βελτίωση των αποτελεσμάτων.

Όπως αναφέρεται, το πρώτο αυτό μοντέλο είναι και ένα καθολικό (global) μοντέλο. Τα καθολικά (global) μοντέλα αναφέρονται σε μοντέλα που λαμβάνουν υπόψη τη συνολική δομή και τις γενικές τάσεις των δεδομένων χωρίς να εστιάζουν σε συγκεκριμένες περιοχές ή τμήματα τους. Αυτά τα μοντέλα είναι σχεδιασμένα για να εκπαιδεύονται σε μεγάλα και πλήρη σύνολα δεδομένων και να μάθουν γενικές σχέσεις και μοτίβα. Έτσι αρχίσαμε με αυτό το καθολικό μοντέλο. Δίνοντας στο μοντέλο ένα μεγάλο σύνολο δεδομένων χωρίς κάποια παραπάνω πληροφορία σχετικά με το είδος κτιρίου, την πολιτεία που αναφέρονται τα δεδομένα, το είδος ενέργειας ή και τον τύπο χρήσης ενέργειας θέλαμε να παρατηρήσουμε αν μπορούσε το μοντέλο να αντιληφθεί τη γενική δομή των δεδομένων και να εξάγει προβλέψεις με ακριβή αποτελέσματα. Παρατηρήσαμε πως οι προβλέψεις του καθολικού μοντέλου είχαν ικανοποιητικά αποτελέσματα χωρίς όμως να υπάρχει σημαντική βελτίωση σε σχέση με το μοντέλο βάσης μας, sNaive το οποίο αποτελεί και σημείο αναφοράς. Παρότι το καθολικό αυτό μοντέλο δεν λαμβάνει κάποια πληροφορία σχετικά με κάποιο από τα χαρακτηριστικά των δεδομένων, παρατηρούμε ότι έχει σχετικά ακριβή αποτελέσματα. Σε αυτή την περίπτωση αυτό συμβαίνει καθώς το καθολικό μοντέλο έχει πολλά περισσότερα δεδομένα από τα οποία μπορεί να εκμεταλλευτεί. Έτσι χρησιμοποιώντας ένα μεγάλο σύνολο δεδομένων για να εκπαιδευτεί καταφέρει να επιστρέφει προβλέψεις με καλά αποτελέσματα. Χρησιμοποιώντας το καθολικό μοντέλο καταλήξαμε στα εξής αποτελέσματα.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|-------------------|--------------------|
| 45,20% | 1,078 |

Πίνακας 5.5: Συνολικά αποτελέσματα καθολικού μοντέλου

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|-------------------|--------------------|
| Νοσοκομείο | 24,64% | 0,415 |
| Εστιατόριο | 30,29% | 0,608 |
| Εστιατόριο ταχείας εξυπηρέτησης | 33,54% | 0,626 |
| Μικρό γραφείο | 73,36% | 1,571 |
| Μεσαίο γραφείο | 70,88% | 1,520 |
| Μεγάλο γραφείο | 49,59% | 1,048 |
| Μικρό ξενοδοχείο | 55,83% | 1,178 |
| Μεγάλο ξενοδοχείο | 59,35% | 1,751 |

Πίνακας 5.6: Αποτελέσματα καθολικού μοντέλου ανά τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|-------------------|--------------------|
| Συνολική κατανάλωση ενέργειας | 22,29% | 0,325 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 45,76% | 0,755 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 113,18% | 2,922 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 3,02% | 0,108 |

Πίνακας 5.7: Αποτελέσματα καθολικού μοντέλου ανά τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|------------------------|-------------------|--------------------|
| Ηλεκτρισμός | 49,02% | 1,231 |
| Φυσικό Αέριο | 49,71% | 0,968 |

Πίνακας 5.8: Αποτελέσματα καθολικού μοντέλου ανά είδος ενέργειας

5.5 Καθολικά Μοντέλα με εξειδίκευση σε υποσύνολο των δεδομένων

5.5.1 Περιγραφή μοντέλων

Όπως αναφέραμε προηγουμένως, το πρώτο μοντέλο που δημιουργήσαμε είναι ένα γενικό καθολικό μοντέλο. Ωστόσο, για να δώσουμε στα μοντέλα περισσότερη πληροφορία σχετικά με τα διάφορα χαρακτηριστικά των δεδομένων, στραφήκαμε στη δημιουργία εξειδικευμένων μοντέλων για συγκεκριμένα υποσύνολα των δεδομένων.

Τα μοντέλα αυτά εστιάζουν σε συγκεκριμένες περιοχές ή τμήματα των δεδομένων, αντί να λαμβάνουν υπόψη τη συνολική δομή ή τις γενικές τάσεις. Με αυτόν τον τρόπο, δίνεται η δυνατότητα να γίνει εστίαση σε συγκεκριμένα τμήματα των δεδομένων και να προσαρμοστούν σε τοπικά μοτίβα που παρατηρούνται. Ουσιαστικά με αυτό τον τρόπο γίνεται μια διάσπαση του πρώτου καθολικού μοντέλου σε επιμέρους μικρότερα καθολικά μοντέλα που εστιάζουν σε ένα τμήμα των δεδομένων, όπως για παράδειγμα μοντέλα ανά τύπο κτιρίου. Έτσι ακόμη χρησιμοποιείται όλα το σύνολο των δεδομένων και δημιουργούνται προβλέψεις για όλες τις χρονοσειρές στα επιμέρους μοντέλα. Παρότι αυτή η διάσπαση σε μεγαλύτερο αριθμό μοντέλων μπορεί να οδηγήσει σε απώλεια μερικών γενικών μοτίβων που είναι ευκολότερο να βρεθούν με την χρήση του καθολικού μοντέλου. Επίσης η εκπαίδευση τους θα έχει λιγότερα δεδομένα για να προσαρμοστεί. Ωστόσο, αυτά τα μοντέλα είναι σχεδιασμένα για να προσαρμόζονται σε τοπικά μοτίβα ή παραμέτρους που μπορεί να διαφέρουν ανάλογα με την περιοχή ή τον τύπο δεδομένων που εξετάζουν. Τα τοπικά μοτίβα είναι συχνά πιο δύσκολο να ανακαλυφθούν με τη χρήση του γενικού μοντέλου, γι' αυτό και τα εξειδικευμένα μοντέλα μπορεί να παράγουν καλύτερα αποτελέσματα.

5.5.2 Περιγραφή και αποτελέσματα μοντέλων ανά τύπο κτιρίου

Τα πρώτα μοντέλα που δημιουργήσαμε ήταν ανά τύπο κτιρίου. Συγκεκριμένα χρησιμοποιώντας την MLP αρχιτεκτονική που είδαμε και στο καθολικό μοντέλο δημιουργήσαμε 8 νέα μοντέλα, ένα για κάθε τύπο κτιρίου. Με αυτό τον τρόπο προσπαθήσαμε να δώσουμε παραπάνω πληροφορία στα δεδομένα μας και να παρατηρήσουμε αν η εστίαση στα τοπικά μοτίβα που παρατηρούνται στα διάφορα είδη κτιρίου μας παρέχει πιο ακριβή αποτελέσματα.

Παρατηρήσαμε πως οι προβλέψεις των μοντέλων ανά τύπο κτιρίου είχαν ικανοποιητικά αποτελέσματα, όμως παρατηρήσαμε πως ήταν χειρότερα από το καθολικό μοντέλο και το sNaive μοντέλο ως προς την scaled MAE μετρική. Παρά την παραπάνω πληροφορία με την οποία τροφοδοτήσαμε το μοντέλο μας δεν είδαμε κάποια σημαντική βελτίωση στα συνολικά αποτελέσματα. Αυτό συμβαίνει μιας τα νέα αυτά μοντέλα έχουν αρκετά λιγότερα δεδομένα για να χρησιμοποιήσουν κατά την εκπαίδευση τους. Έτσι παρότι κάθε μοντέλο μαθαίνει ειδικά για κάθε τύπο κτιρίου δεν παρατηρούμε κάποια σημαντική βελτίωση Χρησιμοποιώντας αυτά τα διαφορετικά μοντέλα καταλήξαμε στα εξής αποτελέσματα.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|-------------------|--------------------|
| 46,76% | 1,089 |

Πίνακας 5.9: Συνολικά αποτελέσματα μοντέλων ανά τύπο κτιρίου

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|-------------------|--------------------|
| Νοσοκομείο | 24,67% | 0,410 |
| Εστιατόριο | 28,76% | 0,557 |
| Εστιατόριο ταχείας εξυπηρέτησης | 32,62% | 0,709 |
| Μικρό γραφείο | 70,44% | 1,642 |
| Μεσαίο γραφείο | 73,69% | 1,654 |
| Μεγάλο γραφείο | 47,48% | 1,022 |
| Μικρό ξενοδοχείο | 44,81% | 1,003 |
| Μεγάλο ξενοδοχείο | 55,90% | 1,738 |

Πίνακας 5.10: Αποτελέσματα μοντέλων ανά τύπο κτιρίου ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|-------------------|--------------------|
| Συνολική κατανάλωση ενέργειας | 21,66% | 0,322 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 44,71% | 0,750 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 103,90% | 2,887 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 4,10% | 0,113 |

Πίνακας 5.11: Αποτελέσματα μοντέλων ανά τύπο κτιρίου ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|------------------------|-------------------|--------------------|
| Ηλεκτρισμός | 47,59% | 1,231 |
| Φυσικό Αέριο | 46,01% | 0,960 |

Πίνακας 5.12: Αποτελέσματα μοντέλων ανά τύπο κτιρίου ως προς το είδος ενέργειας

5.5.3 Περιγραφή και αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας

Τα επόμενα μοντέλα που δημιουργήσαμε ήταν τα μοντέλα των οποίων τα δεδομένα εισόδου είναι ανάλογα με τον τύπο χρήσης ενέργειας. Έτσι δημιουργήσαμε 4 διαφορετικά μοντέλα ένα για κάθε διαφορετικό τύπο χρήσης ενέργειας για να μπορέσουμε να βρούμε τα τοπικά μοτίβα που δημιουργούνται λόγω του συγκεκριμένου χαρακτηριστικού των δεδομένων.

Όμοια με τα μοντέλα με εξειδίκευση ανά τύπο κτιρίου παρατηρήσαμε πως παρόλο την εισαγωγή της πληροφορίας στα μοντέλα μας, η βελτίωση των προβλέψεων δεν ήταν σημαντική σε σχέση με το σημείο αναφοράς του sNaive. Είδαμε βελτιώσεις σε διάφορες περιοχές των δεδομένων μας, όμως τα συνολικά αποτελέσματα δεν έδειξαν κάποια σημαντική βελτίωση. Εδώ συγκριτικά με τα υπόλοιπα αρχικά μοντέλα παρατηρήσαμε μια βελτίωση στα αποτελέσματα. Αυτό μας οδήγησε στην σκέψη πως παρατηρούνται αρκετά μοτίβα σε σχέση με τον τύπο χρήσης ενέργειας. Ένας λόγος που μας οδήγησε στην δημιουργία συνδυαστικών καθολικών μοντέλων. Χρησιμοποιώντας αυτά τα διαφορετικά μοντέλα καταλήξαμε στα εξής αποτελέσματα.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 46,11% | 1,118 |

Πίνακας 5.13: Συνολικά αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 25,13% | 0,410 |
| Εστιατόριο | 29,51% | 0,580 |
| Εστιατόριο ταχείας εξυπηρέτησης | 33,52% | 0,739 |
| Μικρό γραφείο | 63,62% | 1,691 |
| Μεσαίο γραφείο | 71,17% | 1,663 |
| Μεγάλο γραφείο | 49,26% | 1,064 |
| Μικρό ξενοδοχείο | 43,05% | 1,029 |
| Μεγάλο ξενοδοχείο | 57,02% | 1,794 |

Πίνακας 5.14: Αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 23,07% | 0,333 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 44,85% | 0,754 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 99,95% | 2,978 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 2,90% | 0,107 |

Πίνακας 5.15: Αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 48,25% | 1,288 |
| Φυσικό Αέριο | 44,16% | 0,964 |

Πίνακας 5.16: Αποτελέσματα μοντέλων ανά τύπο χρήσης ενέργειας ως προς το είδος ενέργειας

5.5.4 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας

Μια επόμενη δοκιμή που κάναμε είναι η δημιουργία μοντέλων με βάση το είδος ενέργειας. Έτσι στην περίπτωση αυτή δημιουργήσαμε δύο νέα μοντέλα, ένα με δεδομένα εισόδου και κατά συνέπεια εστίαση στον ηλεκτρισμό και ένα στο φυσικό αέριο.

Παρατηρήσαμε πως ο διαχωρισμός αυτός των δεδομένων στα δύο μοντέλα παρότι δημιούργησε ικανοποιητικά αποτελέσματα δεν προσέφερε κάποια σημαντική βελτίωση στα συνολικά αποτελέσματα. Εδώ συγκριτικά με τα υπόλοιπα μοντέλα, έχουμε περισσότερη πληροφορία για την εκπαίδευση των μοντέλων μιας και έχουμε μόνο δυο νέα μοντέλα. Παρόλα αυτά τα αποτελέσματα δεν είχαν κάποιο ιδιαίτερη βελτίωση. Χρησιμοποιώντας αυτά τα δύο μοντέλα καταλήξαμε στα εξής αποτελέσματα.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 48,16% | 1,091 |

Πίνακας 5.17: Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 24,99% | 0,407 |
| Εστιατόριο | 30,95% | 0,590 |
| Εστιατόριο ταχείας εξυπηρέτησης | 40,26% | 0,856 |
| Μικρό γραφείο | 75,54% | 1,621 |
| Μεσαίο γραφείο | 78,95% | 1,544 |
| Μεγάλο γραφείο | 48,38% | 0,959 |
| Μικρό ξενοδοχείο | 45,93% | 1,032 |
| Μεγάλο ξενοδοχείο | 53,50% | 1,743 |

Πίνακας 5.18: Αποτελέσματα μοντέλων ανά είδος ενέργειας ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 23,07% | 0,335 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 45,97% | 0,758 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 105,74% | 2,873 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 5,06% | 0,117 |

Πίνακας 5.19: Αποτελέσματα μοντέλων ανά είδος ενέργειας ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 47,25% | 1,254 |
| Φυσικό Αέριο | 48,99% | 0,943 |

Πίνακας 5.20: Αποτελέσματα μοντέλων ανά είδος ενέργειας ως προς το είδος ενέργειας

5.5.5 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου

Έχοντας παρατηρήσει από τις προηγούμενες δοκιμές στην δημιουργία μοντέλων ότι μπορεί να παρατηρηθεί μια βελτίωση των προβλέψεων σε διάφορα τμήματα των δεδομένων μας λόγω της προσθήκης της πληροφορίας στα μοντέλα και στην εστίαση σε ένα τμήμα των δεδομένων, αποφασίσαμε να δημιουργήσουμε εκ νέου κάποια μοντέλα τα οποία θα εστίαζαν σε ακόμη μικρότερο τμήμα των δεδομένων, ανακαλύπτοντας έτσι τα μοτίβα που αυτά τα τμήματα έχουν.

Για αυτόν τον λόγο λοιπόν δημιουργήσαμε ένα μοντέλο για κάθε είδος ενέργειας σε συνδυασμό με τον τύπο κτιρίου. Με αυτό τον τρόπο ουσιαστικά δημιουργήσαμε 16 διαφορετικά μοντέλα, ένα για κάθε διαφορετικό συνδυασμό είδους ενέργειας και τύπου κτιρίου.

Με αυτόν τον νέο τρόπο προσέγγισης παρατηρήσαμε πως είχαμε ικανοποιητικά αποτελέσματα. Σε σύγκριση όμως με το σημείο αναφοράς, το μοντέλο βάσης sNaive δεν μπορούσαμε να παρατηρήσουμε κάποια σημαντική βελτίωση στο σύνολο των αποτελεσμάτων. Αυτό μας οδήγησε στο συμπέρασμα ότι τα μοτίβα που παρατηρούνται ανάλογα με το είδος ενέργειας και τον τύπο κτιρίου δεν είναι ξεκάθαρα και άρα δεν είναι εύκολο να έχουμε ακριβείς προβλέψεις. Χρησιμοποιώντας αυτά τα διαφορετικά μοντέλα καταλήξαμε στα εξής αποτελέσματα.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 46,36% | 1,114 |

Πίνακας 5.21: Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 24,59% | 0,396 |
| Εστιατόριο | 29,69% | 0,666 |
| Εστιατόριο ταχείας εξυπηρέτησης | 32,88% | 0,743 |
| Μικρό γραφείο | 73,36% | 1,941 |
| Μεσαίο γραφείο | 69,47% | 1,514 |
| Μεγάλο γραφείο | 44,21% | 0,920 |
| Μικρό ξενοδοχείο | 47,35% | 1,047 |
| Μεγάλο ξενοδοχείο | 53,35% | 1,727 |

Πίνακας 5.22: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 21,25% | 0,318 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 43,97% | 0,747 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 104,34% | 3,000 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 3,41% | 0,109 |

Πίνακας 5.23: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 46,96% | 1,299 |
| Φυσικό Αέριο | 45,83% | 0,945 |

Πίνακας 5.24: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου ως προς το είδος ενέργειας

5.5.6 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας

Πέρα από τον συνδυασμό του τύπου κτιρίου με το είδος ενέργειας αποφασίσαμε να φτιάξουμε μοντέλα που εστιάζουν σε δεδομένα ανά είδος ενέργειας και τύπο χρήσης ενέργειας, ειδικότερα αφού είχαμε παρατηρήσει κάποια ικανοποιητικά αποτελέσματα στα μοντέλα με βάση τον τύπο χρήσης ενέργειας. Έτσι δημιουργήσαμε 6 νέα διαφορετικά μοντέλα για όλους τους πιθανούς συνδυασμούς. Με αυτό τον τρόπο μπορέσαμε να εστιάσουμε σε μικρότερα και πιο ειδικά τμήματα των δεδομένων.

Χρησιμοποιώντας αυτά τα μοντέλα μπορέσαμε να παρατηρήσουμε τα καλύτερα αποτελέσματα. Είχαμε μια σημαντική βελτίωση των αποτελεσμάτων συγκρίνοντας τόσο με τα απο-

τελέσματα του μοντέλου βάσης όσο και με τα υπόλοιπα μοντέλα που δημιουργήσαμε. Αυτό συνέβη λόγω των διαφορετικών μοτίβων και τάσεων που παρατηρούνται σε αυτό το υποσύνολο δεδομένων. Όπως αναφέραμε και στην ΝΑΙVE μέθοδο με το παράδειγμα της θέρμανσης για διαφορετικό είδος ενέργειας μπορούμε να παρατηρήσουμε ότι απομονώνοντας τα δεδομένα μας βάση αυτών των δύο χαρακτηριστικών μπορούμε να έχουμε μεγαλύτερη ακρίβεια στα αποτελέσματα. Έτσι χρησιμοποιώντας αυτά τα διαφορετικά μοντέλα καταλήξαμε στα εξής αποτελέσματα.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 43,78% | 1,056 |

Πίνακας 5.25: Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 24,45% | 0,413 |
| Εστιατόριο | 29,79% | 0,607 |
| Εστιατόριο ταχείας εξυπηρέτησης | 31,60% | 0,620 |
| Μικρό γραφείο | 62,58% | 1,461 |
| Μεσαίο γραφείο | 66,94% | 1,502 |
| Μεγάλο γραφείο | 47,56% | 1,042 |
| Μικρό ξενοδοχείο | 43,12% | 1,093 |
| Μεγάλο ξενοδοχείο | 48,47% | 1,688 |

Πίνακας 5.26: Αποτελέσματα μοντέλων ανά είδος ενέργειας και και τύπο χρήσης ενέργειας ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 22,25% | 0,325 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 44,86% | 0,753 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 93,43% | 2,790 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 2,91% | 0,108 |

Πίνακας 5.27: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 43,96% | 1,208 |
| Φυσικό Αέριο | 43,62% | 0,919 |

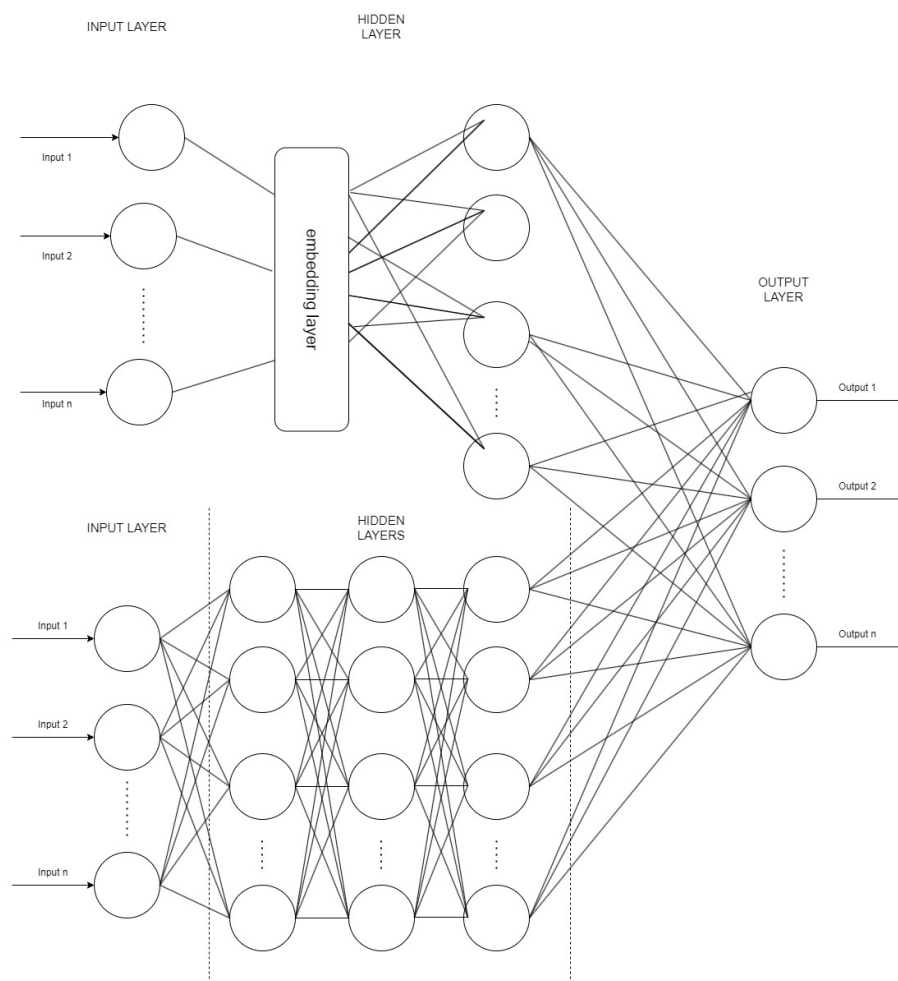
Πίνακας 5.28: Αποτελέσματα μοντέλων ανά είδος ενέργειας και και τύπο χρήσης ενέργειας ως προς το είδος ενέργειας

5.6 Μοντέλα με επίπεδο ενσωμάτωσης

5.6.1 Περιγραφή μοντέλων με επίπεδο ενσωμάτωσης

Δημιουργώντας τα παραπάνω μοντέλα μπορέσαμε να δώσουμε στα μοντέλα μας παραπάνω πληροφορία, μιας και κάθε νέο μοντέλο εστιάζει σε ένα τμήμα των δεδομένων. Πέραν όμως αυτής της προσθήκης πληροφορίας μπορούμε να δώσουμε στα μοντέλα μας περισσότερη πληροφορία και μέσω της χρήσης των επιπέδων ενσωμάτωσης (embedding layers).

Τα επίπεδα ενσωμάτωσης (embedding layers) αποτελούν σημαντικό στοιχείο στα νευρωνικά δίκτυα. Η λειτουργία του επιπέδου ενσωμάτωσης είναι να αντιστοιχίσει κάθε κατηγορία σε ένα διανυσματικό χώρο χαμηλότερης διάστασης. Έτσι εμείς κωδικοποιώντας τα δεδομένα μπορούμε με αυτό τον τρόπο να τα φορτώσουμε στα μοντέλα, παρέχοντας έτσι πληροφορία και βοηθώντας το μοντέλο να αναγνωρίσει τυχόν μοτίβα που υπάρχουν στα διάφορα τμήματα του συνόλου δεδομένων. Ακόμη η προσθήκη του επιπέδου ενσωμάτωσης έγινε πριν το επίπεδο εξόδου με σκοπό να μπορέσουμε να κατηγοριοποιήσουμε τα δεδομένα με μεγαλύτερη ακρίβεια και, χωρίς την παρέμβαση του επιπέδου ενσωμάτωσης στα υπόλοιπα επίπεδα. Παρακάτω μπορούμε να δούμε την αρχιτεκτονική αυτών των μοντέλων.



Σχήμα 5.5: Αρχιτεκτονική μοντέλων με επίπεδο ενσωμάτωσης

5.6.2 Περιγραφή και αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων

Η πρώτη υλοποίηση μοντέλου με την προσθήκη επιπέδου ενσωμάτωσης ήταν η χρησιμοποίηση του αρχικού καθολικού μοντέλου με προσθήκη πληροφορίας για όλους τους διαφορετικούς τύπους κτιρίου, τα είδη ενέργειας και τους τύπους χρήσης ενέργειας. Με αυτό τον τρόπο μπορέσαμε να κατηγοριοποιήσουμε τα δεδομένα μας βάσει όλων των χαρακτηριστικών του συνόλου δεδομένων.

Με αυτές τις προσθήκες στο μοντέλο παρατηρούμε πως τα αποτελέσματα είναι ελαφρώς χειρότερα. Κατηγοριοποιώντας έτσι τα δεδομένα σε συνδυασμό με την χρήση του καθολικού μοντέλου δεν παρατηρούμε κάποια βελτίωση. Παρατηρείται πως η κωδικοποίηση όλων των χαρακτηριστικών και η είσοδος τους στο μοντέλο μέσω επιπέδου ενσωμάτωσης, μπερδεύει στην κατηγοριοποίηση το μοντέλο μας οδηγώντας σε ανακριβή αποτελέσματα συγκριτικά με τα υπόλοιπα μοντέλα. Τα αποτελέσματα του παραπάνω μοντέλου είναι τα εξής.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 52,05% | 1,122 |

Πίνακας 5.29: Συνολικά αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 26,34% | 0,421 |
| Εστιατόριο | 31,21% | 0,568 |
| Εστιατόριο ταχείας εξυπηρέτησης | 36,94% | 0,725 |
| Μικρό γραφείο | 84,25% | 1,760 |
| Μεσαίο γραφείο | 72,38% | 1,527 |
| Μεγάλο γραφείο | 50,49% | 1,040 |
| Μικρό ξενοδοχείο | 66,86% | 1,207 |
| Μεγάλο ξενοδοχείο | 51,77% | 1,735 |

Πίνακας 5.30: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 23,61% | 0,340 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 47,47% | 0,763 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 117,84% | 2,969 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 4,97% | 0,117 |

Πίνακας 5.31: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 48,90% | 1,247 |
| Φυσικό Αέριο | 54,91% | 1,008 |

Πίνακας 5.32: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για όλα τα χαρακτηριστικά του συνόλου δεδομένων ως προς το είδος ενέργειας

5.6.3 Περιγραφή και αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου

Παρατηρώντας τα αποτελέσματα του προηγούμενου μοντέλου, διαπιστώθηκε ότι η προσθήκη όλων των κωδικοποιημένων χαρακτηριστικών του συνόλου δεδομένων στο επίπεδο ενσωμάτωσης δεν έδειχνε κάποια βελτίωση στα αποτελέσματα. Για αυτό τον λόγο δημιουργήσαμε ένα μοντέλο στο οποίο μέσω του επιπέδου ενσωμάτωσης δώσαμε πληροφορία για ένα μόνο χαρακτηριστικό του συνόλου δεδομένων. Σε αυτή την περίπτωση δώσαμε σαν είσοδο στο επίπεδο ενσωμάτωσης τα διαφορετικά είδη κτιρίου.

Παρατηρώντας και εδώ τα αποτελέσματα δεν μπορούμε να παρατηρήσουμε κάποια βελτίωση παρά την κατηγοριοποίηση των δεδομένων μας συγκριτικά με το σημείο αναφοράς της μεθόδου sNaive. Μάλιστα εδώ παρατηρούμε τα χειρότερα αποτελέσματα συγκριτικά με τα υπόλοιπα μοντέλα που χρησιμοποιήσαμε. Αυτό είναι αναμενόμενο μιας και με την χρήση του τύπου κτιρίου για παραπάνω πληροφορία δεν είχαμε παρατηρήσει κάποια σημαντική βελτίωση ούτε στα μοντέλα με εξειδίκευση στον τύπο κτιρίου. Σε συνδυασμό με την κωδικοποίηση των δεδομένων και την προσθήκη τους μέσω επιπέδου ενσωμάτωσης που έχουμε παρατηρήσει πως δεν έχει αρκετά ακριβείς προβλέψεις, είναι αναμενόμενη η πτώση των συνολικών αποτελεσμάτων. Τα αποτελέσματα με την χρήση αυτού του μοντέλου είναι τα εξής.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 53,00% | 1,105 |

Πίνακας 5.33: Συνολικά αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 27,13% | 0,421 |
| Εστιατόριο | 31,28% | 0,567 |
| Εστιατόριο ταχείας εξυπηρέτησης | 35,01% | 0,720 |
| Μικρό γραφείο | 79,14% | 1,690 |
| Μεσαίο γραφείο | 73,49% | 1,528 |
| Μεγάλο γραφείο | 53,35% | 1,050 |
| Μικρό ξενοδοχείο | 52,91% | 1,046 |
| Μεγάλο ξενοδοχείο | 74,04% | 1,827 |

Πίνακας 5.34: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 23,49% | 0,341 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 46,33% | 0,763 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 122,43% | 2,905 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 4,65% | 0,116 |

Πίνακας 5.35: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 57,29% | 1,279 |
| Φυσικό Αέριο | 49,09% | 0,946 |

Πίνακας 5.36: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς το είδος ενέργειας

5.6.4 Περιγραφή και αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας

Μια επόμενη δοκιμή που κάναμε είναι η δημιουργία καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας. Έτσι στην περίπτωση αυτή δημιουργήσαμε ένα νέο μοντέλο προσθέτοντας ένα επίπεδο ενσωμάτωσης που κατηγοριοποιεί τα δεδομένα με βάση τον τύπο χρήσης ενέργειας.

Παρατηρήσαμε και σε αυτή την περίπτωση της χρήσης επιπέδου ενσωμάτωσης σε συνδυασμό με την χρήση καθολικού μοντέλου πως τα αποτελέσματα είναι ελαφρώς χειρότερα από το σημείο αναφοράς. Σε αυτή την περίπτωση επίσης παρατηρήσαμε πως συγκριτικά με τα υπόλοιπα μοντέλα που κάνουν χρήση του επιπέδου ενσωμάτωσης πως είχαμε τα καλύτερα αποτελέσματα. Αυτό επιβεβαιώνει ακόμη μια φορά την ύπαρξη των μοτίβων στα δεδομένα ανάλογα με τον τύπο χρήσης ενέργειας. Χρησιμοποιώντας αυτό το μοντέλο καταλήξαμε στα εξής αποτελέσματα.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 47,11% | 1,084 |

Πίνακας 5.37: Συνολικά αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|-------------------|--------------------|
| Νοσοκομείο | 25,87% | 0,416 |
| Εστιατόριο | 30,77% | 0,565 |
| Εστιατόριο ταχείας εξυπηρέτησης | 35,68% | 0,722 |
| Μικρό γραφείο | 65,91% | 1,635 |
| Μεσαίο γραφείο | 70,30% | 1,528 |
| Μεγάλο γραφείο | 50,52% | 1,045 |
| Μικρό ξενοδοχείο | 43,86% | 1,012 |
| Μεγάλο ξενοδοχείο | 57,72% | 1,762 |

Πίνακας 5.38: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|-------------------|--------------------|
| Συνολική κατανάλωση ενέργειας | 23,73% | 0,341 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 47,33% | 0,765 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 100,07% | 2,828 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 4,52% | 0,115 |

Πίνακας 5.39: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|------------------------|-------------------|--------------------|
| Ηλεκτρισμός | 49,47% | 1,251 |
| Φυσικό Αέριο | 44,95% | 0,931 |

Πίνακας 5.40: Αποτελέσματα καθολικού μοντέλου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς το είδος ενέργειας

5.6.5 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας

Έχοντας παρατηρήσει από τις προηγούμενες δοκιμές με την χρήση επιπέδου ενσωμάτωσης σε καθολικά μοντέλα πως οι τελικές προβλέψεις δεν επέφεραν κάποια βελτίωση στα αποτελέσματα αποφασίσαμε να προχωρήσουμε στον συνδυασμό καθολικών μοντέλων με εξειδίκευση σε κάποιο υποσύνολο των δεδομένων, με την χρήση επιπέδου ενσωμάτωσης. Έτσι πήραμε τα δύο καλύτερα μοντέλα με εξειδίκευση σε υποσύνολο των δεδομένων και προσθέσαμε σε αυτά περαιτέρω πληροφορία μέσω επιπέδων ενσωμάτωσης. Σε αυτήν την περίπτωση χρησιμοποιήσαμε τα μοντέλα που εστιάζουν στο είδος ενέργειας και στον τύπο κτιρίου και προσθέσαμε πληροφορία μέσω επιπέδου ενσωμάτωσης για τον τύπο χρήσης ενέργειας.

Με αυτόν τον νέο τρόπο προσέγγισης παρατηρήσαμε πως είχαμε ικανοποιητικά αποτελέσματα σε σύγκριση με τα υπόλοιπα μοντέλα που περιείχαν επίπεδα ενσωμάτωσης. Παρόλα αυτά σε σύγκριση με το σημείο αναφοράς, το μοντέλο βάσης sNaive δεν μπορούσαμε να παρατηρήσουμε κάποια σημαντική βελτίωση στο σύνολο των αποτελεσμάτων. Εδώ περιμέναμε μία βελτίωση στα αποτελέσματα σε σχέση με τα προηγούμενα μοντέλα με την χρήση επιπέδου ενσωμάτωσης η οποία και έγινε. Παρόλα αυτά δεν είχαμε σημαντική βελτίωση ούτως ώστε να έχουμε ένα καλύτερο μοντέλο πρόβλεψης για τα δεδομένα. Τα αποτελέσματα με την χρήση αυτού του μοντέλου είναι τα εξής.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 48,83% | 1,113 |

Πίνακας 5.41: Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 24,72% | 0,398 |
| Εστιατόριο | 30,57% | 0,667 |
| Εστιατόριο ταχείας εξυπηρέτησης | 35,09% | 0,747 |
| Μικρό γραφείο | 82,54% | 1,945 |
| Μεσαίο γραφείο | 66,43% | 1,485 |
| Μεγάλο γραφείο | 45,61% | 0,921 |
| Μικρό ξενοδοχείο | 49,59% | 1,038 |
| Μεγάλο ξενοδοχείο | 60,42% | 1,745 |

Πίνακας 5.42: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 21,11% | 0,316 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 44,83% | 0,746 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 112,62% | 2,999 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 4,08% | 0,111 |

Πίνακας 5.43: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 50,67% | 1,306 |
| Φυσικό Αέριο | 47,17% | 0,937 |

Πίνακας 5.44: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο κτιρίου με επίπεδο ενσωμάτωσης για τον τύπο χρήσης ενέργειας ως προς το είδος ενέργειας

5.6.6 Περιγραφή και αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου

Σε αυτή την περίπτωση συνδυάσαμε το καλύτερο μοντέλο που είχαμε βάση αποτελεσμάτων με ένα επίπεδο ενσωμάτωσης μέσω του οποίου δώσαμε πληροφορία στα μοντέλα μας σχετικά με το είδος κτιρίου.

Χρησιμοποιώντας αυτά τα μοντέλα δεν μπορέσαμε να παρατηρήσουμε κάποια βελτίωση στα αποτελέσματα. Συγκεκριμένα προσθέτοντας το επίπεδο ενσωμάτωσης για το είδος κτιρίου είχαμε αρκετά χειρότερα αποτελέσματα σε σχέση με το συγκεκριμένο αρχικό μοντέλο. Αυτό μας οδηγεί στο συμπέρασμα πως η είσοδος πληροφορίας μέσω επιπέδου ενσωμάτωσης μπερδεύει το μοντέλο περισσότερο από ότι το βοηθάει. Η κωδικοποίηση και κατηγοριοποίηση αυτή που γίνεται στα δεδομένα όπως φαίνεται δεν οδηγεί σε καλύτερα αποτελέσματα. Τα αποτελέσματα αυτής της μεθόδου είναι τα εξής.

Συνολικά αποτελέσματα

| scaled MAE | scaled RMSE |
|------------|-------------|
| 48,29% | 1,078 |

Πίνακας 5.45: Συνολικά αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου

Αποτελέσματα ανά τύπο κτιρίου

| Τύπος Κτιρίου | scaled MAE | scaled RMSE |
|---------------------------------|------------|-------------|
| Νοσοκομείο | 24,61% | 0,411 |
| Εστιατόριο | 30,24% | 0,605 |
| Εστιατόριο ταχείας εξυπηρέτησης | 32,35% | 0,622 |
| Μικρό γραφείο | 71,43% | 1,538 |
| Μεσαίο γραφείο | 68,25% | 1,512 |
| Μεγάλο γραφείο | 48,91% | 1,042 |
| Μικρό ξενοδοχείο | 54,65% | 1,159 |
| Μεγάλο ξενοδοχείο | 58,25% | 1,712 |

Πίνακας 5.46: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς τον τύπο κτιρίου

Αποτελέσματα ανά τύπο χρήσης ενέργειας

| Τύπος Χρήσης Ενέργειας | scaled MAE | scaled RMSE |
|---|------------|-------------|
| Συνολική κατανάλωση ενέργειας | 22,01% | 0,323 |
| Κατανάλωση ενέργειας σε συστήματα ψύξης | 45,34% | 0,751 |
| Κατανάλωση ενέργειας σε συστήματα θέρμανσης | 110,12% | 2,877 |
| Κατανάλωση ενέργειας σε εσωτερικό εξοπλισμό | 2,55% | 0,107 |

Πίνακας 5.47: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς τον τύπο χρήσης ενέργειας

Αποτελέσματα ανά είδος ενέργειας

| Είδος Ενέργειας | scaled MAE | scaled RMSE |
|-----------------|------------|-------------|
| Ηλεκτρισμός | 49,10% | 1,219 |
| Φυσικό Αέριο | 47,57% | 0,951 |

Πίνακας 5.48: Αποτελέσματα μοντέλων ανά είδος ενέργειας και τύπο χρήσης ενέργειας με επίπεδο ενσωμάτωσης για τον τύπο κτιρίου ως προς το είδος ενέργειας

5.7 Σύγκριση και Ανάλυση Αποτελεσμάτων**5.7.1 Μετρική Αξιολόγησης Αποτελεσμάτων**

Όπως είδαμε στην παρουσίαση των αποτελεσμάτων των παραπάνω μοντέλων χρησιμοποιήσαμε δύο μετρικές για την σύγκριση των αποτελεσμάτων. Αυτές οι μετρικές είναι η scaled MAE μετρική και η scaled RMSE. Η scaled RMSE λαμβάνει την τετραγωνική ρίζα του μέσου των τετραγωνικών σφαλμάτων, ενώ η scaled MAE λαμβάνει το μέσο των απόλυτων τιμών των σφαλμάτων. Αυτές οι δύο μετρικές μπορεί να έχουν διαφορετικές αποκλίσεις, ιδίως όταν τα δεδομένα έχουν διαφορετική κατανομή. Όπως παρατηρήσαμε η μετρική scaled RMSE βγάζει αρκετά χειρότερα αποτελέσματα. Αυτό συμβαίνει λόγω των δεδομένων που χρησιμοποιούμε. Μερικές συνεχείς εναλλαγές που μπορούμε να παρατηρήσουμε στα δεδομένα οδηγούν την scaled RMSE μετρική να έχει χειρότερα αποτελέσματα. Για παράδειγμα η χρήση ενέργειας σε ένα γραφείο μειώνεται απότομα στο τέλος της εργασιακής ημέρας και αντίστροφα ανεβαίνει απότομα στην αρχή της εργασιακής ημέρας. Αυτές οι απότομες εναλλαγές (outliers) οδηγούν έχουν έντονη επίδραση στην μετρική scaled RMSE. Για αυτό τον λόγο κατά την συνολική σύγκριση επικεντρωθήκαμε στην μετρική scaled MAE.

5.7.2 Σύγκριση αποτελεσμάτων

Σύγκριση συνολικών αποτελεσμάτων

| Μοντέλα | scaled MAE | # Models |
|--|---------------|----------|
| NAIVE MODEL | 46,07% | 1 |
| GLOBAL MODEL | 45,21% | 1 |
| BUILDING TYPE MODEL | 46,77% | 8 |
| ENERGY USAGE MODEL | 46,11% | 4 |
| ENERGY TYPE MODEL | 48,16% | 2 |
| BUILDING TYPE & ENERGY TYPE | 46,37% | 16 |
| ENERGY USAGE & ENERGY TYPE | 43,79% | 6 |
| EMBEDDING MODEL(all info provided) | 52,05% | 1 |
| EMBEDDING MODEL(building type) | 53,01% | 1 |
| EMBEDDING MODEL(energy usage) | 47,11% | 1 |
| EMBEDDING MODEL(building type & energy type) | 48,84% | 16 |
| EMBEDDING MODEL(energy usage & energy type) | 48,30% | 6 |
| ENERGY USAGE & ENERGY TYPE WITH NAÏVE HEATING PREDICTIONS | 42,49% | 5 |

Πίνακας 5.49: Σύγκριση συνολικών αποτελεσμάτων και αριθμού μοντέλων

Ανάλυση συνολικών αποτελεσμάτων

Από τον παραπάνω πίνακα, παρατηρούμε ότι τα τρία μοντέλα με τις καλύτερες αποδόσεις είναι το μοντέλο βάσης (sNaive), το καθολικό μοντέλο (GLOBAL) και τέλος το μοντέλο με την καλύτερη απόδοση είναι το μοντέλο με την εξειδίκευση σε δεδομένα βάσει του τύπου κτιρίου και του είδους της ενέργειας.

Το απλοϊκό μοντέλο βάσης έχει αρκετά καλά αποτελέσματα λόγω της ιδιομορφίας των δεδομένων. Παρατηρούνται μοτίβα και πολλαπλές συνεχείς τιμές που καθιστούν το μοντέλο βάσης αξιόπιστο.

Το δεύτερο καλύτερο μοντέλο είναι το καθολικό μοντέλο, το οποίο επωφελείται από τον μεγάλο όγκο δεδομένων που χρησιμοποιεί κατά την εκπαίδευση.

Το βέλτιστο μοντέλο είναι αυτό που τροφοδοτεί τα μοντέλα του με δεδομένα με βάση τον τύπο χρήσης ενέργειας αλλά και το είδος της ενέργειας. Αυτό είναι ένα αναμενόμενο αποτέλεσμα μιας και παρόλο τον μικρό όγκο δεδομένων που χρησιμοποιεί κάθε μοντέλο για την εκπαίδευση του, συγκριτικά με το καθολικό μοντέλο, εδώ τα μοντέλα μπορούν πιο εύκολα να αναγνωρίσουν μοτίβα μέσα σε αυτά τα δεδομένα που το καθολικό μοντέλο, λόγω της υλοποίησης του καθώς και του μεγάλου αριθμού δεδομένων στο οποίο εκπαιδεύεται μπορεί να μην έχει την δυνατότητα να ανιχνεύσει. Έτσι περιμέναμε μια γενικότερη βελτίωση των αποτελεσμάτων με την χρήση μοντέλων με εξειδίκευση σε υποσύνολο των δεδομένων, γεγονός το οποίο δεν έγινε όπως βλέπουμε από τα αποτελέσματα για τις υπόλοιπες περιπτώσεις των μοντέλων. Αυτό συνέβη λόγω της έλλειψης χαρακτηριστικών μοτίβων και τάσεων σε δεδομένα που αφορούν ένα συγκεκριμένο τύπο κτιρίου ή και για ένα είδος ενέργειας.

Αυτό είναι εύκολο να το παρατηρήσει κανείς μιας και σε όλες τις διαφορετικές περιπτώσεις μοντέλων, τα καλύτερα αποτελέσματα έρχονται από τα μοντέλα που εκπαιδεύονται σε δεδομένα που αφορούν τον τύπο χρήσης ενέργειας.

Τέλος, παρατηρώντας τα αποτελέσματα του Naive μοντέλου σε σχέση με τον τύπο χρήσης ενέργειας δημιουργήσαμε ένα νέο συνδυαστικό μοντέλο. Αυτό συνδυάζει τις προβλέψεις

του μοντέλου με την εξειδίκευση των μοντέλων σε δεδομένα με βάση τον τύπο χρήσης ενέργειας καθώς και το είδος της ενέργειας και τις προβλέψεις του Naive μοντέλου μόνο για δεδομένα που αναφέρονται στην θέρμανση στην οποία κατηγορία, όπως θα παρατηρήσουμε και παρακάτω η αφελής μέθοδος έχει αρκετά καλύτερα αποτελέσματα συγκριτικά με τα υπόλοιπα μοντέλα. Έτσι με τον συνδυασμό των προβλέψεων του βέλτιστου μοντέλου μαζί με κάποιες προβλέψεις του Naive μοντέλου καταφέραμε να βελτιώσουμε ακόμη παραπάνω τα αποτελέσματά.

Ανάλυση αποτελεσμάτων ανά τύπο κτιρίου

| Scaled MAE | NAIVE MODEL | GLOBAL MODEL | BUILDING TYPE MODEL | ENERGY USAGE MODEL | ENERGY TYPE MODEL | BUILDING TYPE & ENERGY TYPE MODEL | ENERGY USAGE & ENERGY TYPE MODEL | EMBEDDING MODEL(all) | EMBEDDING MODEL (building type) | EMBEDDING MODEL (energy usage) | EMBEDDING MODEL (building type & energy type) | EMBEDDING MODEL (energy usage & energy type) |
|--------------------------|-------------|--------------|---------------------|--------------------|-------------------|-----------------------------------|----------------------------------|----------------------|---------------------------------|--------------------------------|---|--|
| HOSPITAL | 30,65% | 24,64% | 24,67% | 25,13% | 24,99% | 24,59% | 24,46% | 26,34% | 27,14% | 25,87% | 24,73% | 24,61% |
| RESTAURANT | 32,80% | 30,30% | 28,76% | 29,51% | 30,95% | 29,69% | 29,80% | 31,21% | 31,29% | 30,78% | 30,57% | 30,24% |
| QUICK SERVICE RESTAURANT | 36,02% | 33,54% | 32,62% | 33,52% | 40,27% | 32,88% | 31,61% | 36,95% | 35,01% | 35,69% | 35,10% | 32,36% |
| SMALL HOTEL | 46,39% | 55,84% | 44,82% | 43,06% | 45,93% | 47,36% | 43,12% | 66,86% | 52,92% | 43,86% | 49,60% | 54,65% |
| LARGE HOTEL | 47,91% | 59,36% | 55,91% | 57,03% | 53,50% | 53,35% | 48,48% | 51,78% | 74,04% | 57,72% | 60,42% | 58,26% |
| SMALL OFFICE | 60,54% | 73,37% | 70,45% | 63,63% | 75,54% | 73,37% | 62,58% | 84,26% | 79,14% | 65,91% | 82,55% | 71,44% |
| MEDIUM OFFICE | 68,67% | 70,88% | 73,70% | 71,18% | 78,95% | 69,48% | 66,94% | 72,38% | 73,49% | 70,31% | 66,44% | 68,25% |
| LARGE OFFICE | 52,92% | 49,60% | 47,48% | 49,26% | 48,38% | 44,22% | 47,56% | 50,50% | 53,36% | 50,52% | 45,62% | 48,92% |

Σχήμα 5.6: Σύγκριση αποτελεσμάτων με βάση τον τύπο κτιρίου

Παρατηρώντας τον παραπάνω heatmap μπορούμε να βγάλουμε αρκετά συμπεράσματα. Αρχικά παρατηρούμε πως τα μοντέλα τα οποία είχαν και συνολικά καλά αποτελέσματα έχουν και υψηλά αποτελέσματα σε όλους τους διαφορετικούς τύπους κτιρίου, πέραν από το αφελές μοντέλο βάσης. Στο sNaive μοντέλο παρατηρείται ότι έχει πολύ υψηλές τιμές σε συγκεκριμένα είδη κτιρίου και αρκετά άσχημα αποτελέσματα συγκριτικά με τα υπόλοιπα μοντέλα σε μερικά άλλα. Αυτό είναι αναμενόμενο λόγω της διαφοράς των δεδομένων βάσει του τύπου κτιρίου. Για μερικούς τύπους κτιρίου, όπως τα ξενοδοχεία, τα οποία έχουν συνεχώς επαναλαμβανόμενες τιμές λόγω συγκεκριμένων ωραρίων λειτουργίας καθώς και πιθανόν συνεχόμενες μηδενικές τιμές το αφελές μοντέλο έχει αρκετά καλά αποτελέσματα. Από την άλλη σε τύπους κτιρίων όπως εστιατόρια και νοσοκομεία που κάνουν χρήση ενέργειας καθόλη την μέρα παρατηρείται ότι δεν μπορεί να έχει ικανοποιητικά αποτελέσματα.

Παρατηρείται ότι τα μοντέλα με εξειδίκευση σε ένα υποσύνολο των δεδομένων βάση μόνο μιας παραμέτρου δεν καταφέρνουν να προβλέψουν με αποτελεσματικότητα τις διάφορες χρονοσειρές. Με την περαιτέρω εξειδίκευση των δεδομένων σε μικρότερο τμήμα των δεδομένων,

είτε βάσει τύπου κτιρίου και είδους ενέργειας, είτε βάσει τύπου χρήσης ενέργειας και είδους ενέργειας, παρατηρείται μια συνολική βελτίωση των αποτελεσμάτων. Αυτό γίνεται λόγω της ιδιομορφίας των δεδομένων και των τοπικών μοτίβων που υπάρχουν σε αυτά. Το βέλτιστο μοντέλο, παρόλο το γεγονός ότι δεν έχει κάποια εξειδίκευση στον τύπο κτιρίου, επιτυγχάνει να προβλέψει τις χρονοσειρές με μεγαλύτερη ακρίβεια σε όλα τα είδη κτιρίου συγκριτικά με τα υπόλοιπα μοντέλα. Αυτή η παρατήρηση οδηγεί στο συμπέρασμα ότι οι ιδιομορφίες που παρουσιάζουν τα δεδομένα ανά τύπο χρήσης ενέργειας και ανά είδος ενέργειας διαδραματίζουν σημαντικότερο ρόλο από τις αλλαγές που παρατηρούνται λόγω του διαφορετικού τύπου κτιρίου. Χαρακτηριστικό παράδειγμα είναι πως για ένα συγκεκριμένο τύπο κτιρίου οι ιδιομορφίες των δεδομένων εξαιτίας των υπόλοιπων παραμέτρων μπορεί να οδηγήσουν τα μοντέλα με εξειδίκευση στο υποσύνολο των δεδομένων ανά τύπο κτιρίου σε ανακριβή αποτελέσματα και αδυναμία πρόβλεψης των χρονοσειρών. Τέλος αξίζει να παρατηρήσουμε ότι τα μοντέλα που έκαναν χρήση του επιπέδου ενσωμάτωσης δεν κατάφεραν να προβλέψουν με ακρίβεια τις χρονοσειρές. Και σε αυτά τα είδη μοντέλων όμως, παρατηρείται η βελτίωση των αποτελεσμάτων όταν το μοντέλα αυτά εστιάζουν σε μικρότερο υποσύνολο των δεδομένων.

Ανάλυση αποτελεσμάτων ανά είδος ενέργειας

| Scaled MAE | NAIVE MODEL | GLOBAL MODEL | BUILDING TYPE MODEL | ENERGY USAGE MODEL | ENERGY TYPE MODEL | BUILDING TYPE & ENERGY TYPE MODEL | ENERGY USAGE & ENERGY TYPE MODEL | EMBEDDING MODEL(all) | EMBEDDING MODEL (building type) | EMBEDDING MODEL (energy usage) | EMBEDDING MODEL (building type & energy type) | EMBEDDING MODEL (energy usage & energy type) |
|-------------|-------------|--------------|---------------------|--------------------|-------------------|-----------------------------------|----------------------------------|----------------------|---------------------------------|--------------------------------|--|--|
| ELECTRICITY | 47,86% | 49,02% | 47,59% | 48,25% | 47,25% | 46,96% | 43,96% | 48,90% | 57,30% | 49,48% | 50,68% | 49,11% |
| GAS | 45,08% | 49,72% | 46,02% | 44,16% | 48,99% | 45,83% | 43,63% | 54,92% | 49,10% | 44,96% | 47,17% | 47,57% |

Σχήμα 5.7: Σύγκριση αποτελεσμάτων με βάση το είδος ενέργειας

Στα αποτελέσματα με βάση το είδος της ενέργειας δεν παρατηρείται κάποια ακραία τιμή στα αποτελέσματα σε καμία από τις δύο περιπτώσεις. Και για τα δύο είδη τα αποτελέσματα είναι συγκρίσιμα και βελτιώνονται ανάλογα και με την συνολική βελτίωση του μοντέλου. Το γεγονός αυτό είναι αναμενόμενο μιας και ο όγκος των δεδομένων είναι αρκετά μεγάλος και στις δύο κατηγορίες ενέργειας και έτσι δεν τονίζονται τα δεδομένα όπου παρατηρούνται συνεχείς μηδενικές τιμές ή στα δεδομένα τα οποία εμπεριέχουν διάφορες ιδιομορφίες. Αυτό μπορεί να επαληθευθεί και από τα αποτελέσματα του αφελούς μοντέλου για το οποίο έχουν παρατηρηθεί οι μεγαλύτερες αυξομειώσεις ανάλογα με το τμήμα των δεδομένων στο οποίο αναφέρεται. Σε αυτήν την περίπτωση τα αποτελέσματα του είναι συγκρίσιμα και στις δύο περιπτώσεις.

Όμοια με τις παρατηρήσεις σχετικά με τον τύπο κτιρίου, κι σε αυτή την περίπτωση παρατηρείται βελτίωση των αποτελεσμάτων όταν τα μοντέλα εστιάζουν σε μικρότερο υποσύνολο των δεδομένων. Ακόμη, και σε αυτή την περίπτωση παρατηρείται ότι η συνολική βελτίωση των αποτελεσμάτων οδηγεί στην βελτίωση των προβλέψεων και για τα δύο είδη ενέργειας. Όπως μπορεί να παρατηρηθεί και από το βέλτιστο μοντέλο, το οποίο έχει καλύτερα αποτελέσματα και για τις δύο κατηγορίες ενέργειας συγκριτικά με τα υπόλοιπα μοντέλα. Αντίστοιχα για τα

μοντέλα που έχουν συνολικά χειρότερα αποτελέσματα παρατηρείται η αδυναμία πρόβλεψης των χρονοσειρών και για τα δύο είδη ενέργειας.

Τέλος και σε αυτή την περίπτωση τα μοντέλα που κάνουν χρήση του επιπέδου ενσωμάτωσης έχουν τα χειρότερα αποτελέσματα συγκριτικά με τα υπόλοιπα καθολικά μοντέλα και βελτιώνονται στην περίπτωση της προσθήκης επιπέδου ενσωμάτωσης στα καθολικά μοντέλα που έχουν εξειδίκευση σε ένα μικρότερο υποσύνολο των δεδομένων είτε αυτό είναι ανά τύπο κτιρίου και είδος ενέργειας είτε ανά τύπο χρήσης ενέργειας και είδος ενέργειας.

Ανάλυση αποτελεσμάτων ανά τύπο χρήσης ενέργειας

| Scaled MAE | NAIVE MODEL | GLOBAL MODEL | BUILDING TYPE MODEL | ENERGY USAGE MODEL | ENERGY TYPE MODEL | BUILDING TYPE & ENERGY TYPE MODEL | ENERGY USAGE & ENERGY TYPE MODEL | EMBEDDING MODEL(all) | EMBEDDING MODEL (building type) | EMBEDDING MODEL (energy usage) | EMBEDDING MODEL (building type & energy type) | EMBEDDING MODEL (energy usage & energy type) |
|--------------------|-------------|--------------|---------------------|--------------------|-------------------|-----------------------------------|----------------------------------|----------------------|---------------------------------|--------------------------------|---|--|
| FACILITY | 27,87% | 22,30% | 21,66% | 23,07% | 23,07% | 21,25% | 22,25% | 23,62% | 23,50% | 23,74% | 21,11% | 22,02% |
| COOLING | 53,21% | 45,76% | 44,72% | 44,86% | 45,98% | 43,97% | 44,87% | 47,47% | 46,34% | 47,33% | 44,83% | 45,35% |
| HEATING | 88,78% | 113,18% | 103,91% | 99,96% | 105,75% | 104,35% | 93,43% | 117,84% | 122,44% | 100,08% | 112,62% | 110,13% |
| INTERIOR EQUIPMENT | 1,57% | 3,02% | 4,11% | 2,90% | 5,07% | 3,41% | 2,91% | 4,98% | 4,66% | 4,52% | 4,09% | 2,56% |

Σχήμα 5.8: Σύγκριση αποτελεσμάτων με βάση τον τύπο χρήσης ενέργειας

Παρατηρώντας αυτόν τον πίνακα είναι πλέον ξεκάθαρη η ύπαρξη μοτίβων ανά τύπο χρήσης ενέργειας. Αντίθετα με τα αποτελέσματα ανά είδος ενέργειας, σε αυτήν την περίπτωση παρατηρείται σημαντική διακύμανση των αποτελεσμάτων σε σχέση με τον τύπο χρήσης ενέργειας. Αυτό μπορεί εύκολα να επαληθευθεί από τα αποτελέσματα του αφελούς μοντέλου. Βλέπουμε πως στην περίπτωση του αφελούς μοντέλου, έχουμε τα χειρότερα αποτελέσματα για την συνολική κατανάλωση και για δεδομένα που αναφέρονται σε ψύξη, και αντιθέτως έχουμε τα καλύτερα αποτελέσματα συγκριτικά με όλα τα μοντέλα για τα δεδομένα που αφορούν θέρμανση καθώς και κατανάλωση ενέργειας για εσωτερικό εξοπλισμό.

Ακόμη παρατηρείται πως ανάλογα τον τύπο χρήσης ενέργειας τα αποτελέσματα έχουν σημαντικές διαφορές. Συγκεκριμένα τα αποτελέσματα για τον εσωτερικό εξοπλισμό κυμαίνονται σε ένα ποσοστό από 1,5% έως και 5% ενώ στην περίπτωση της θέρμανσης τα αποτελέσματα κυμαίνονται από 88% έως και 122%. Οι πολύ ακριβείς προβλέψεις όσον αφορά τον εσωτερικό εξοπλισμό είναι εύκολο να ερμηνευθούν μιας και σε αυτή την περίπτωση οι τιμές των παρατηρήσεων είναι αρκετά μικρές και έτσι οι προβλέψεις είναι πάντα σχετικά κοντά. Για αυτόν ακριβώς το λόγο το αφελές μοντέλο έχει τα καλύτερα αποτελέσματα, μιας και οι διακυμάνσεις των παρατηρήσεων ανά εβδομάδα είναι χαμηλές.

Από την άλλη τα αποτελέσματα για την θέρμανση είναι αρκετά χειρότερα. Το γεγονός αυτό προκύπτει από τις μεγάλες εναλλαγές των τιμών καθώς και τις ιδιομορφίες που συναντώνται στα δεδομένα αυτά. Όπως έχει αναφερθεί, στα δεδομένα σχετικά με την θέρμανση παρατηρούνται συνεχόμενες μηδενικές τιμές για την θερινή περίοδο ή για κάποιο είδος ενέργειας το οποίο δεν χρησιμοποιείται με σκοπό την θέρμανση κάποιου τύπου κτιρίου. Ακόμη

ανάλογα με τον τύπο κτιρίου και το ωράριο για το οποίο το κτίριο αυτό είναι εν λειτουργία οι τιμές μπορεί να έχουν σημαντικές εναλλαγές κατά την διάρκεια μιας ημέρας. Αυτό επιβεβαιώνεται και από τα αποτελέσματα του αφελούς μοντέλου το οποίο και σε αυτή την περίπτωση παράγει τις καλύτερες προβλέψεις συγκριτικά με τα υπόλοιπα μοντέλα.

Εν αντιθέσει στην συνολική κατανάλωση και στην κατανάλωση ενέργειας για ψύξη, όπου οι τιμές δεν έχουν τόσες πολλές συνεχόμενες τιμές και μπορούν να παρατηρηθούν μεγαλύτερες διακυμάνσεις στις τιμές των δεδομένων βλέπουμε πως το μοντέλο βάσης έχει τα χειρότερα αποτελέσματα.

Και σε αυτή την περίπτωση για τα υπόλοιπα μοντέλα πέραν του μοντέλου βάσης, μπορεί να παρατηρηθεί ότι η βελτίωση των συνολικών αποτελεσμάτων οδηγεί σε καλύτερα αποτελέσματα για όλους τους τύπους χρήσης ενέργειας. Τα καθολικά μοντέλα με εξειδίκευση σε μικρότερο υποσύνολο των δεδομένων δημιουργούν κι εδώ τις καλύτερες προβλέψεις ενώ τα μοντέλα με επίπεδο ενσωμάτωσης αδυνατούν να προβλέψουν με επιτυχία τις χρονοσειρές.

Παρόλη την σημαντική βελτίωση που έχει το βέλτιστο μοντέλο σε όλους τους τύπους χρήσης ενέργειας, μπορεί να παρατηρηθεί ότι στην περίπτωση των δεδομένων για θέρμανση το αφελές μοντέλο έχει αρκετά καλύτερα αποτελέσματα. Για αυτόν ακριβώς τον λόγο έγινε και ο συνδυασμός του βέλτιστου μοντέλου χρησιμοποιώντας τις προβλέψεις του για τρεις εκ των τεσσάρων κατηγοριών, με τις προβλέψεις του Naive μοντέλου για την κατηγορία της θέρμανσης με σκοπό την δημιουργία ενός μοντέλου το οποίο θα είναι ικανό να προβλέπει τα δεδομένα για όλους τους διαφορετικούς τύπους χρήσης ενέργειας.

Κεφάλαιο 6

Συμπεράσματα και προεκτάσεις

6.1 Συμπεράσματα

Στην εργασία αυτή πραγματοποιήθηκε μια ολοκληρωμένη μελέτη για την πρόβλεψη τιμών κατανάλωσης ενέργειας με τη χρήση γενικευμένων νευρωνικών δικτύων αξιοποιώντας μεγάλο όγκο δεδομένων. Δημιουργήσαμε διάφορα μοντέλα πρόβλεψης ικανά να προσδιορίζουν με ακρίβεια τις μελλοντικές καταναλώσεις.

Ακόμη, μέσω της διπλωματικής εργασίας μελετήσαμε τον τρόπο με τον οποίο είναι δυνατόν να γίνει η κατάλληλη προεπεξεργασία των δεδομένων ενός συνόλου δεδομένων με σκοπό την μετατροπή τους και την προετοιμασία τους για την είσοδο τους στα νευρωνικά δίκτυα. Με τη χρήση μίας μορφής *feature engineering* πάνω στο σύνολο δεδομένων, επιτεύχθηκε η εξαγωγή όσο το δυνατόν περισσότερης πληροφορίας από τα δεδομένα ώστε να μπορεί να χρησιμοποιηθεί από τα επιμέρους μοντέλα.

Τέλος, εξετάστηκε ο τρόπος με τον οποίο οι αλλαγές στο μέγεθος των δεδομένων εισόδου και το είδος της πληροφορίας που παρέχεται στα μοντέλα επηρεάζουν τα αποτελέσματα των μοντέλων πρόβλεψης. Πιο συγκεκριμένα, έγινε διαχωρισμός των δεδομένων με βάση το είδος της ενέργειας, τον τύπο χρήσης της, καθώς και τον τύπο κτιρίου. Χρησιμοποιώντας μόνο ένα από τα τρία χαρακτηριστικά ή συνδυασμό αυτών, διερευνήσαμε την επίδραση της προσθήκης των πληροφοριών μέσω καθολικών μοντέλων νευρωνικών δικτύων τα οποία έχουν εξειδίκευση σε ένα υποσύνολο των δεδομένων. Ακόμη έγινε προσθήκη πληροφορίας στα μοντέλα μέσω εφαρμογής επιπέδου ενσωμάτωσης (*embeddings layer*). Με αυτόν τον τρόπο, αναδείξαμε τις βέλτιστες πρακτικές για την επίτευξη ακριβέστερων και πιο αξιόπιστων προβλέψεων.

Σε πρώτο μέρος επεξηγήθηκε αναλυτικά το θεωρητικό υπόβαθρο στο οποίο βασίστηκε η μελέτη που έγινε στην διπλωματική εργασία. Έγινε αναφορά στα διάφορα μοντέλα, στους αλγόριθμους και τους δείκτες σφάλματος που χρησιμοποιήθηκαν. Στο δεύτερο μέρος πραγματοποιήθηκε μια εκτενής περιγραφή του πειραματικού σταδίου της εργασίας. Αρχικά, έγινε αναφορά στον τρόπο με τον οποίο συλλέχθηκαν τα δεδομένα, καθώς και στο στάδιο της προεπεξεργασίας του μεγάλου όγκου τους με σκοπό την κατάλληλη τροποποίηση τους. Τέλος, έγινε αναφορά στην δημιουργία των μοντέλων μηχανικής μάθησης που χρησιμοποιήθηκαν για την παραγωγή των προβλέψεων, καθώς και η ανάλυση των αποτελεσμάτων των επιμέρους μοντέλων.

Από αυτή την μελέτη εξάγονται κάποια πολύ ενδιαφέροντα συμπεράσματα. Αρχικά χρησιμοποιήθηκε η εποχική αφελής μέθοδος (*Seasonal Naive*) ως σημείο αναφοράς προς τα

υπόλοιπα μοντέλα. Συγκεκριμένα χρησιμοποιήθηκαν τα δεδομένα της προηγούμενης εβδομάδας ως η πρόβλεψη για την επόμενη εβδομάδα. Το μοντέλο αυτό παρά την απλότητα του παρουσίασε αρκετά αποτελέσματα τα οποία ήταν συγκρίσιμα με τα υπόλοιπα μοντέλα που δημιουργήθηκαν. Συγκεκριμένα, συγκριτικά με τα υπόλοιπα μοντέλα είχε την τέταρτη καλύτερη επίδοση. Αυτό έγινε λόγω της ιδιομορφίας των δεδομένων που αφορούν την κατανάλωση ενέργειας. Χαρακτηριστικά, για τύπους χρήσης ενέργειας όπως η θέρμανση και η ψύξη είναι φανερό πως υπάρχει μια έντονη εποχικότητα, που εκφράζεται με πολλές μηδενικές τιμές στο σύνολο δεδομένων, γεγονός που βοήθησε την αφελή μέθοδο στο να έχει πιο ακριβή αποτελέσματα.

Ακόμη τα δεδομένα μας αναφέρονται σε δύο είδη ενέργειας, ηλεκτρισμό και φυσικό αέριο. Αυτό επίσης δημιούργησε μια ιδιομορφία στα δεδομένα, καθώς για ένα συγκεκριμένο τύπο κτιρίου πολλές φορές γινόταν χρήση ενός από τα δύο είδη ενέργειας με αποτέλεσμα την ύπαρξη πολλών μηδενικών τιμών στην άλλη περίπτωση. Τέλος, σε τύπους χρήσης ενέργειας όπως η κατανάλωση ενέργειας για τον εσωτερικό εξοπλισμό ενός κτιρίου, επειδή οι τιμές εκεί ήταν αρκετά μικρές το μοντέλο είχε ακριβείς προβλέψεις. Συγκεκριμένα για δεδομένα που αναφέρονται στον εσωτερικό εξοπλισμό το αφελές μοντέλο είχε μια τιμή του 1,57% στην μετρική scaled MAE. Αυτές ήταν μερικές περιπτώσεις που βοήθησαν το αφελές μοντέλο να έχει καλύτερες προβλέψεις από κάποια από τα υπόλοιπα μοντέλα που μελετήθηκαν. Συγκεκριμένα το εποχικό αφελές μοντέλο παρουσίασε μια τιμή της τάξης του 46,07% για την μετρική scaled MAE, η οποία χρησιμοποιήθηκε καθόλη την διάρκεια της μελέτης για την σύγκριση των μοντέλων.

Το τρίτο καλύτερο μοντέλο ήταν το καθολικό (global). Το μοντέλο αυτό ήταν η πρώτη προσπάθεια για την δημιουργία ενός μοντέλου. Το καθολικό μοντέλο χωρίς κάποια πρότερη πληροφορία εκπαιδεύεται σε όλο το σύνολο δεδομένων. Τα ικανοποιητικά αποτελέσματα αυτού του μοντέλου συγκριτικά με τα υπόλοιπα μοντέλα ήταν αναμενόμενα μιας και έχει πολλά περισσότερα δεδομένα τα οποία χρησιμοποιεί στο στάδιο της εκπαίδευσης και έτσι μπορεί να εντοπίσει μερικά γενικά μοτίβα τα οποία δεν θα ήταν εύκολο να εντοπιστούν από άλλα τοπικά μοντέλα. Το καθολικό μοντέλο παρουσίασε 45,2% στην μετρική scaled MAE, δηλαδή ένα ποσοστό του 2% αυξημένη από αυτή του αφελούς μοντέλου. Παρότι παρατηρείται πως στην περίπτωση των μοντέλων με εξειδίκευση σε ένα υποσύνολο των δεδομένων τα αποτελέσματα βελτιώνονται στις περισσότερες περιπτώσεις, το καθολικό μοντέλο κατάφερε να έχει επαρκή αποτελέσματα συγκριτικά με τα υπόλοιπα μοντέλα λόγω της ικανότητας του να προβλέπει και τα δεδομένα με μεγαλύτερες ιδιομορφίες, όπως δεδομένα που αναφέρονται σε θέρμανση.

Το δεύτερο καλύτερο μοντέλο πρόβλεψης βάσει αποτελεσμάτων είναι αυτό που προέκυψε από την δημιουργία νευρωνικών δικτύων ανά είδος ενέργειας και τύπο χρήσης ενέργειας. Με αυτό τον συνδυασμό καταλήξαμε σε 6 διαφορετικά τοπικά μοντέλα που είχαν πληροφορία τόσο για το είδος ενέργειας όσο και για τον τύπο χρήσης ενέργειας. Τα υψηλής ακρίβειας αποτελέσματα αυτού του μοντέλου ήταν επίσης αναμενόμενα μιας και με αυτόν τον τρόπο τα μοντέλα παρότι υστερούσαν σε γνώση για τα γενικά μοτίβα που υπήρχαν στο σύνολο δεδομένων, μπορούσαν να εντοπίσουν με μεγάλη ακρίβεια μοτίβα και ιδιομορφίες στα δεδομένα, βάσει των δύο αυτών χαρακτηριστικών. Αξίζει επίσης να σημειωθεί ότι ήταν αναμενόμενα τα καλύτερα αποτελέσματα σε αυτό τον τύπο μοντέλου, μιας και όπως αναφέρθηκε προη-

γουμένως, κατά την μελέτη και την επεξεργασία των δεδομένων είχαν παρατηρηθεί τέτοιες ιδιομορφίες ειδικά σε σχέση με αυτά τα δύο χαρακτηριστικά. Το μοντέλο αυτό παρουσίασε μια τιμή του 43,787% στην μετρική scaled MAE και έτσι ήταν κατά 5% βελτιωμένο σε σχέση με το μοντέλο αναφοράς, το αφελές μοντέλο. Ακόμη παρατηρώντας τα αποτελέσματα που είχαν τα μοντέλα σε σχέση με τα χαρακτηριστικά των δεδομένων, παρατηρεί κανείς πως τα βέλτιστα μοντέλα ήταν συνολικά καλύτερα, και όχι μόνο στα δεδομένα για τα οποία είχαν και περισσότερη πληροφορία.

Χαρακτηριστικά, το μοντέλο είχε πιο ακριβή αποτελέσματα ανά συγκεκριμένο τύπο κτιρίου συγκριτικά με άλλα μοντέλα που είχαν πληροφορία για την αναγνώριση του τύπου κτιρίου. Στην μοναδική περίπτωση που αυτό δεν συνέβη ήταν στην περίπτωση του Seasonal Naive μοντέλου. Συγκριτικά με τον τύπο χρήσης ενέργειας παρατηρείται πως αυτό το μοντέλο είχε τα καλύτερα αποτελέσματα για δεδομένα που αφορούσαν την κατανάλωση ενέργειας για θέρμανση καθώς και για εσωτερικό εξοπλισμό, και τα χειρότερα αποτελέσματα για δεδομένα που αναφέρονται στην συνολική κατανάλωση ενέργειας ενός κτιρίου και στην κατανάλωση ενέργειας για ψύξη. Φυσικά και αυτό ήταν αναμενόμενο λόγω της ιδιομορφίας των δεδομένων των οποίων εξετάστηκαν. Ακόμη σημαντικό είναι να παρατηρηθεί ότι το καθολικό μοντέλο με εξειδίκευση σε τύπο κτιρίου και είδος ενέργειας παρουσίασε βελτιωμένες προβλέψεις σε σχέση με τα υπόλοιπα μοντέλα. Αυτό μας οδηγεί στο συμπέρασμα ότι η εξειδίκευση των μοντέλων σε μικρότερο τμήμα των δεδομένων τα βοήθησε να δημιουργήσουν καλύτερες προβλέψεις.

Τέλος, το βέλτιστο αποτέλεσμα προέκυψε από τον συνδυασμό δύο μοντέλων. Του βέλτιστου καθολικού μοντέλου με εξειδίκευση ανά είδος ενέργειας και ανά τύπο χρήσης ενέργειας και του αφελές μοντέλου. Συγκεκριμένα χρησιμοποιήθηκαν οι προβλέψεις του αφελές μοντέλου για τα δεδομένα που αναφέρονται σε θέρμανση και οι υπόλοιπες προβλέψεις έγιναν με την χρήση του καθολικού μοντέλου. Τα αποτελέσματα σε αυτή την περίπτωση ήταν 42,49% στην μετρική scaled MAE γεγονός που καθιστά αυτήν την συνδυαστική μέθοδο την βέλτιστη.

Ακόμη, αξίζει να παρατηρηθεί πως στα μοντέλα στα οποία χρησιμοποιήθηκε επίπεδο ενσωμάτωσης (embedding layer) τα αποτελέσματα συγκριτικά με τα υπόλοιπα μοντέλα ήταν χειρότερα. Αυτό μπορεί να έγινε λόγω πιθανής ανεπάρκειας δεδομένων, μιας και χρειάζεται αυξημένος αριθμός δεδομένων για την εκπαίδευση αυτών των μοντέλων, ή και στην αντίθετη περίπτωση λόγω πιθανού overfitting στα δεδομένα. Η προσθήκη αυτών των επιπέδων ενσωμάτωσης αυξάνει την πολυπλοκότητα του μοντέλου, κάτι που μπορεί να οδηγήσει σε μεγαλύτερο χρόνο εκπαίδευσης και σε δυσκολίες στην επιστροφή στοιχείων που να είναι σημαντικά για την πρόβλεψη.

6.2 Μελλοντικές Επεκτάσεις

Στην μελέτη μας δόθηκε έμφαση κυρίως στην ανάδειξη της σχέσης που υπάρχει στα αποτελέσματα των μοντέλων σε σχέση με τον τρόπο εκπαίδευσης τους και την πληροφορία που παρέχονται σε αυτά. Παρόλα αυτά, τα αποτελέσματα μπορούν να βελτιωθούν ώστε να δοθεί μία πληρέστερη εικόνα σχετικά με την πρόβλεψη χρονοσειρών με την χρήση νευρωνικών μοντέλων.

Στην συγκεκριμένη περίπτωση ένα βήμα που θα οδηγούσε σε μια πιο ολοκληρωμένη εικόνα είναι η χρήση όλου του συνόλου δεδομένων. Όπως αναφέρθηκε λόγω περιορισμένων πόρων χρησιμοποιήθηκε για την μελέτη ένα υποσύνολο των δεδομένων. Με περισσότερους πόρους θα ήταν δυνατή η χρήση ολόκληρου του συνόλου δεδομένων, παρέχοντας έτσι πληροφορία για ακόμη περισσότερους τύπους χρήσης ενέργειας καθώς και τύπους κτιρίου. Αυτό θα οδηγούσε σε ακόμη μεγαλύτερη πολυπλοκότητα στα δεδομένα, σε περισσότερα τοπικά μοτίβα, αλλά και σε πιο λεπτομερείς αναλύσεις. Ακόμη, θα μπορούσε να γίνει χρήση κι άλλων πηγών δεδομένων πέρα από αυτή που χρησιμοποιήθηκε σε αυτό το πείραμα. Έτσι, είναι δυνατόν να έχουμε δεδομένα σχετικά με την κατανάλωση ενέργειας σε ακόμη περισσότερους τύπους κτιρίου, περισσότερα είδη ενέργειας καθώς και τύπους χρήσης ενέργειας. Ακόμη με την συλλογή δεδομένων από άλλες πηγές η ανάλυση θα μπορούσε να επεκταθεί και γεωγραφικά, συλλέγοντας δεδομένα για άλλες χώρες πέραν των Ηνωμένων Πολιτειών της Αμερικής, παράγοντας έτσι μια πιο ολοκληρωμένη παγκόσμια μελέτη.

Επιπλέον, η μελέτη μας θα επωφελούνταν από την κατασκευή πιο πολύπλοκων νευρωνικών μοντέλων. Όπως είναι λογικό η αύξηση του συνόλου δεδομένων θα επιφέρει και μεγαλύτερη ανάγκη σε πόρους. Ωστόσο όμως με τα κατάλληλα μέσα είναι δυνατόν να δημιουργηθούν μοντέλα ικανά να εκπαιδευτούν με βάση αυτά τα δεδομένα, δημιουργώντας έτσι πιο πολύπλοκα νευρωνικά δίκτυα. Με περισσότερους πόρους, είναι δυνατόν να δημιουργηθούν μοντέλα με περισσότερα επίπεδα και νευρώνες, τα οποία θα είναι σε θέση να εκπαιδευτούν αποτελεσματικότερα και να επιτύχουν καλύτερα αποτελέσματα. Βέβαια, η επιλογή αυτή θα πρέπει να είναι προσεκτική, λαμβάνοντας υπόψη το συνολικό κόστος σε πόρους σε σύγκριση με την απόδοση των μοντέλων, καθώς και πιο απλά μοντέλα ενδέχεται να παρέχουν ακριβή αποτελέσματα, κάτι που παρατηρήσαμε και με τα επαρκή αποτελέσματα της εποχικής αφελούς μεθόδου. Μιας και στις μέρες μας παρατηρείται η ανοδική βελτίωση των βαθιών νευρωνικών δικτύων και τα δεδομένα που προσφέρονται είναι όλο και περισσότερα, η δημιουργία τέτοιων μοντέλων θα μπορούσε να βελτιώσει τα συνολικά αποτελέσματα. Για παράδειγμα ένα MLP μοντέλο αλλά με αρκετά περισσότερα κρυφά επίπεδα θα μπορούσε να εκπαιδευτεί καλύτερα στον μεγάλο όγκο των δεδομένων. Ακόμη θα μπορούσε να γίνει χρήση ενός LSTM μοντέλου, το οποίο είναι ένα νευρωνικό βαθιάς μάθησης και ανήκει στην κατηγορία των επαναλαμβανόμενων νευρωνικών δικτύων (RNN). Σε αυτά τα μοντέλα η ροή των δεδομένων δεν είναι μόνο από την είσοδο στην έξοδο όπως στην περίπτωση της αρχιτεκτονικής του νευρωνικού δικτύου που χρησιμοποιήθηκε κατά την διάρκεια της μελέτης, αλλά είναι αμφίδρομη, γεγονός που παρότι δημιουργεί μεγαλύτερη δυσκολία κατά την διάρκεια της εκπαίδευσης μπορεί συνολικά να επιφέρει καλύτερα αποτελέσματα.

Βιβλιογραφία

- [1] *Stationary timeseries*. <https://blog.quantinsti.com/stationarity/>.
- [2] *Trend Timeseries*. <https://towardsdatascience.com/understanding-time-series-trend-addfd9d7764e/>.
- [3] *Seasonality Timeseries*. https://training.weather.gov/pds/climate/pcu2/statistics/Stats/part1/CTS_SeaVar.htm.
- [4] *Cyclical Timeseries*. <https://www.analyticsvidhya.com/blog/2023/02/various-techniques-to-detect-and-isolate-time-series-components-using-python/>.
- [5] *Timeseries with outliers*. <https://robjhyndman.com/hyndsight/tsoutliers/>.
- [6] *Timeseries with level shift*. https://www.scirp.org/pdf/ojs_2020042814313761.pdf.
- [7] *Irregular Timeseries*. <https://www.analyticsvidhya.com/blog/2023/02/various-techniques-to-detect-and-isolate-time-series-components-using-python/>.
- [8] *Neural networks*. [https://en.wikipedia.org/wiki/Neural_network_\(machine_learning\)](https://en.wikipedia.org/wiki/Neural_network_(machine_learning)).
- [9] *schema of artificial neuron*. <https://www.v7labs.com/blog/neural-network-architectures-guide>.
- [10] *Activation functions of neural networks*. <https://itsudit.medium.com/activation-functions-in-deep-learning-understanding-the-role-of-activation-functions-in-neural-423694f7f54e>.
- [11] *Explanation of MLP models*. <https://www.datacamp.com/tutorial/multilayer-perceptrons-in-machine-learning>.
- [12] *Picture of RNN architecture*. <https://ailephant.com/glossary/recurrent-neural-network/>.
- [13] *A Guide to Convolutional Neural Networks*. <https://saturncloud.io/blog/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way/>.
- [14] *Figure of Embedding Layers in MLP*. https://www.researchgate.net/figure/Architecture-of-MLP-with-embedding-layer_fig2_364024533.
- [15] Pablo Verdes, Pablo Granitto, Hugo Navone και Η. Ceccatto. *Forecasting chaotic time series: Global vs. local methods*. 2000.

- [16] Pablo Montero-Manso και Rob J. Hyndman. *Principles and algorithms for forecasting groups of time series: Locality and globality*. *International Journal of Forecasting*, 37(4):1632–1653, 2021.
- [17] João Sousa και Roberto Henriques. *Intersecting reinforcement learning and deep factor methods for optimizing locality and globality in forecasting: A review*. *Engineering Applications of Artificial Intelligence*, 133:108082, 2024.
- [18] Sumriti Ranjan Patra, Hone Jay Chu και Tatas. *Regional groundwater sequential forecasting using global and local LSTM models*. *Journal of Hydrology: Regional Studies*, 47:101442, 2023.
- [19] Hansika Hewamalage, Christoph Bergmeir και Kasun Bandara. *Global models for time series forecasting: A Simulation study*. *Pattern Recognition*, 124:108441, 2022.
- [20] Huaizhi Wang, Zhenxing Lei, Xian Zhang, Bin Zhou και Jianchun Peng. *A review of deep learning for renewable energy forecasting*. *Energy Conversion and Management*, 198:111799, 2019.
- [21] Zahra Eddaoudi, Zineb Aarab, Khadija Boudmen, Asmae Elghazi και Moulay Driss Rahmani. *A Brief Review of Energy Consumption Forecasting Using Machine Learning Models*. *Procedia Computer Science*, 236:33–40, 2024. International Symposium on Green Technologies and Applications (ISGTA2023).
- [22] Siyuan Huang, Yepeng Liu, Haoyi Cui, Fan Zhang, Jinjiang Li, Xiaofeng Zhang, Mingli Zhang και Caiming Zhang. *MEAformer: An all-MLP transformer with temporal external attention for long-term time series forecasting*. *Information Sciences*, 669:120605, 2024.
- [23] Moisés Cordeiro-Costas, Hugo Labandeira-Pérez, Daniel Villanueva, Raquel Pérez-Orozco και Pablo Eguía-Oller. *NSGA-II based short-term building energy management using optimal LSTM-MLP forecasts*. *International Journal of Electrical Power and Energy Systems*, 159:110070, 2024.
- [24] Ben Jiang, Yu Li, Yacine Rezgui, Chengyu Zhang, Peng Wang και Tianyi Zhao. *Multi-source domain generalization deep neural network model for predicting energy consumption in multiple office buildings*. *Energy*, 299:131467, 2024.
- [25] Peijun Zheng, Heng Zhou, Jiang Liu και Yosuke Nakanishi. *Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture*. *Applied Energy*, 349:121607, 2023.
- [26] Lei Fang και Bin He. *A deep learning framework using multi-feature fusion recurrent neural networks for energy consumption forecasting*. *Applied Energy*, 348:121563, 2023.
- [27] Nawaf Abdulla, Mehmet Demirci και Suat Ozdemir. *Smart meter-based energy consumption forecasting for smart cities using adaptive federated learning*. *Sustainable Energy, Grids and Networks*, 38:101342, 2024.

- [28] *The Ultimate Guide to Time-Series Analysis*. <https://www.timescale.com/blog/what-is-time-series-analysis-with-examples-and-applications/>.
- [29] *Stationarity and differencing*. <https://otexts.com/fpp2/stationarity.html>.
- [30] *What is a trend in time series?* <https://www.geeksforgeeks.org/what-is-a-trend-in-time-series/>.
- [31] *Seasonality in Time Series Analysis*. <https://towardsdatascience.com/seasonality-of-time-series-5b45b4809acd>.
- [32] *Time Series Analysis: Understanding Seasonality and Cyclicity*. <https://ogre51.medium.com/in-time-series-forecasting-what-do-you-think-is-the-difference-between-seasonality-and-cyclicity-f4e8d9523d24>.
- [33] *Understanding outliers in time series analysis*. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/space-time-pattern-mining/understanding-outliers-in-time-series-analysis.htm>.
- [34] *Outliers, level shifts, and variance changes in time series*. <https://onlinelibrary.wiley.com/doi/abs/10.1002/for.3980070102>.
- [35] *Irregular time series*. <https://www.ibm.com/docs/pl/informix-servers/12.10?topic=concepts-irregular-time-series>.
- [36] *Time Series - Naive Methods*. https://www.tutorialspoint.com/time_series/time_series_naive_methods.htm.
- [37] *Exponential Smoothing*. <https://otexts.com/fppgr/expsmooth.html>.
- [38] *Autoregressive Integrated Moving Average (ARIMA) Prediction Model*. <https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arma.asp>.
- [39] *Theta Model for Time Series Forecasting*. <https://towardsdatascience.com/theta-model-for-time-series-forecasting-642ad1d00358>.
- [40] *TIME SERIES FORECASTING USING NEURAL NETWORKS*. <https://arxiv.org/pdf/1401.1333>.
- [41] *Mean Error: Definition*. <https://www.statisticshowto.com/mean-error/>.
- [42] *Mean Absolute Error - an overview*. <https://www.sciencedirect.com/topics/engineering/mean-absolute-error>.
- [43] *Mean Squared Error (MSE)*. <https://statisticsbyjim.com/regression/mean-squared-error-mse/>.
- [44] *Root Mean Square Error (RMSE)*. <https://statisticsbyjim.com/regression/root-mean-square-error-rmse/>.

- [45] *Mean Absolute Scaled Error (MASE) in Forecasting*. <https://medium.com/@ashishdce/mean-absolute-scaled-error-mase-in-forecasting-8f3aecc21968>.
- [46] *MASE, RMSSE Metrics*. <https://guangyuwu.wordpress.com/2022/09/16/mase-rmsse-metrics/>.
- [47] *Statistical vs neural network prediction for timeseries*. <https://blog.reachsumit.com/posts/2022/12/stats-vs-ml-for-ts/>.
- [48] *Artificial Neuron*. https://en.wikipedia.org/wiki/Artificial_neuron.
- [49] *Explanation of RNN models*. <https://www.ibm.com/topics/recurrent-neural-networks>.
- [50] *Explanation of CNN models*. <https://www.ibm.com/topics/convolutional-neural-networks>.
- [51] *The Essential Guide to Neural Network Architectures*. <https://www.v7labs.com/blog/neural-network-architectures-guide>.
- [52] *How does Batch Size impact your model learning*. <https://medium.com/geekculture/how-does-batch-size-impact-your-model-learning-2dd34d9fb1fa>.
- [53] *Compare the effect of different scalers on data with outliers*. https://scikit-learn.org/stable/auto_examples/preprocessing/plot_all_scaling.html.
- [54] *MinMaxScaler*. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html>.
- [55] *Using StandardScaler() Function to Standardize Python Data*. <https://www.digitalocean.com/community/tutorials/standardscaler-function-in-python>.
- [56] *StandardScaler, MinMaxScaler and RobustScaler techniques - ML*. <https://www.geeksforgeeks.org/standardscaler-minmaxscaler-and-robustscaler-techniques-ml/>.
- [57] *MaxAbsScaler*. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MaxAbsScaler.html#sklearn.preprocessing.MaxAbsScaler>.
- [58] *Global versus local training strategies*. <https://www.crossknowledge.com/blog/combining-global-and-local-training-strategies-for-the-best-of-both-worlds/>.
- [59] Mahin Mohammadi, Saman Jamshidi, Alireza Rezvanian, Mehdi Gheisari και Ajay Kumar. *Advanced fusion of MTM-LSTM and MLP models for time series forecasting: An application for forecasting the solar radiation*. *Measurement: Sensors*, 33:101179, 2024.
- [60] Liam Vaughan, Muyang Zhang, Haoran Gu, Joan B. Rose, Colleen C. Naughton, Gertjan Medema, Vajra Allan, Anne Roiko, Linda Blackall και Arash Zamyadi. *An exploration of challenges associated with machine learning for time series forecasting of COVID-19 community spread using wastewater-based epidemiological data*. *Science of The Total Environment*, 858:159748, 2023.

- [61] *What Are Embedding Layers in Neural Networks.* <https://www.baeldung.com/cs/neural-nets-embedding-layers>.
- [62] *Commercial and Residential Hourly Load Data Set.* <https://bigdata.seas.gwu.edu/data-set-11-commercial-and-residential-hourly-load-data-set/>.