



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ  
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ  
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ  
ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Ανάπτυξη αρχιτεκτονικών νευρωνικών δικτύων  
πρόβλεψης με χρήση μηχανισμών προσοχής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΚΟΥΛΟΥΡΗ Χ. ΣΠΥΡΙΔΩΝΟΣ

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

Αθήνα, Ιούλιος 2024

---





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων  
Και Συστημάτων Αποφάσεων

## Ανάπτυξη αρχιτεκτονικών νευρωνικών δικτύων πρόβλεψης με χρήση μηχανισμών προσοχής

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

**ΚΟΥΛΟΥΡΗ Χ. ΣΠΥΡΙΔΩΝΟΣ**

**Επιβλέπων:** Ασημακόπουλος Βασίλειος  
Ομότιμος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 16η Ιουλίου 2024.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

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

.....  
Ιωάννης Ψαρράς  
Καθηγητής Ε.Μ.Π.

.....  
Δημήτριος Ασκούνης  
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2024





Εθνικό Μετσόβιο Πολυτεχνείο  
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών  
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων  
Και Συστημάτων Αποφάσεων

(Υπογραφή)

.....  
Χρήστος Σπυρίδων  
Κουλούρης  
Διπλωματούχος  
Ηλεκτρολόγος και  
Μηχανικός Υπολογιστών  
Ε.Μ.Π  
16 Ιουλίου 2024

Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.

Χρήστος Σπυρίδων Κουλούρης  
Διπλωματούχος Ηλεκτρολόγος και Μηχανικός Υπολογιστών Ε.Μ.Π, 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. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.



*στην οικογένειά μου*





## Ευχαριστίες

---

Θα ήθελα να ευχαριστήσω τον Καθηγητή του τομέα Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων κ. Βασίλειο Ασημακόπουλο ο οποίος μου έδωσε την ευκαιρία να εκπονήσω υπό την επίβλεψή του την παρούσα εργασία, δίνοντάς μου την ευκαιρία να ασχοληθώ με έναν πολύ ενδιαφέρον ερευνητικό χώρο. Θα ήθελα επίσης να ευχαριστήσω τον συντονιστή της Μονάδας Προβλέψεων και στρατηγικής Δρ. Ευάγγελο Σπηλιώτη για την υποστήριξή του, καθώς και τον υπεύθυνο της παρούσας εργασίας υπ. Διδάκτορα Αρτέμιο-Ανάργυρο Σεμένογλου, του οποίου η βοήθεια και καθοδήγηση κατά την εκπόνηση της εργασίας μου έδινε ώθηση να εξερευνήσω το αντικείμενο στον μέγιστο δυνατό βαθμό. Θα ήθελα, ακόμα, να ευχαριστήσω τον Καθηγητή κ. Ιωάννη Ψαρρά και τον Καθηγητή κ. Δημήτριο Ασκούνη για τη συμμετοχή τους στην επιτροπή εξέτασης της εργασίας.

Τέλος, δεν θα μπορούσα να μην αναφερθώ σε αυτό το σημείο στον πατέρα μου Παύλο, στη μητέρα μου Μαρία, στην αδερφή μου Αδαμαντία και στη Σοφία, των οποίων η συναισθηματική υποστήριξη υπήρξε καταλυτική καθ'όλη τη διάρκεια των σπουδών μου στο Εθνικό Μετσόβιο Πολυτεχνείο.

Αθήνα, Ιούλιος 2024

*Χρήστος Σπυρίδων Κουλούρης*

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



## Περίληψη

---

Η παρούσα διπλωματική εργασία επιχειρεί την πρόβλεψη του φορτίου ηλεκτρικής ενέργειας ανά την Ευρώπη, χρησιμοποιώντας μηχανισμούς προσοχής. Πιο συγκεκριμένα, αρχικά γίνεται μία εκτενής βιβλιογραφική έρευνα για τον εντοπισμό των μέχρι τώρα ερευνητικών προσπαθειών στην κατεύθυνση αυτή, είτε με χρήση μηχανισμών προσοχής είτε χωρίς. Έπειτα, έχοντας γνώση των δυνατοτήτων αλλά και των αδυναμιών των δημοσιευμένων ερευνών στον τομέα αυτό, εξετάζεται η ικανότητα αρχιτεκτονικών με μηχανισμούς προσοχής στην πρόβλεψη του συγκεκριμένου είδους χρονοσειρών. Γίνεται μία εκτενής έρευνα των νευρωνικών αρχιτεκτονικών που μπορούν να τοποθετηθούν πριν αλλά και μετά τον μηχανισμό προσοχής ενώ μελετάται και ο τρόπος με τον οποίο μπορούν να συνδυαστούν μηχανισμοί προσοχής στην ίδια νευρωνική αρχιτεκτονική, έτσι ώστε να προκύψουν οι πιο ακριβείς προβλέψεις. Παράλληλα, ιδιαίτερη έμφαση δίνεται και στην δυνατότητα παραγωγής επεξηγήσιμων προβλέψεων, κάτι που καθιστά τα παρουσιαζόμενα μοντέλα πιο αξιόπιστα. Επιπλέον, κατά την προσπάθεια αύξησης της ακρίβειας των προβλέψεων, διαπιστώνεται η σημασία της πρόβλεψης της εμφάνισης ειδικών γεγονότων που διαταράσσουν την κανονικότητα των χρονοσειρών, και προτείνονται μεθοδολογίες προς αυτή την κατεύθυνση. Τέλος, η παρούσα έρευνα προτείνει και έναν νέο Transformer, ο οποίος έχει τη δυνατότητα να παρέχει πολύ πιο αξιόπιστες προβλέψεις από αυτές των αντίστοιχων παρόχων ηλεκτρικής ενέργειας, ενώ φαίνεται να μπορεί να προβλέψει σε αξιοσημείωτο βαθμό την εμφάνιση ειδικών γεγονότων, χωρίς την χρήση επιπλέον πληροφοριών που στην βιβλιογραφία προκύπτουν από την χρήση ημερολογιακών ή καιρικών μεταβλητών.

## Λέξεις Κλειδιά

Πρόβλεψη Χρονοσειρών, Ενεργειακό Φορτίο, Νευρωνικά Δίκτυα, Μηχανισμός Προσοχής



# Abstract

---

This thesis attempts to forecast electricity load across Europe using Attention Mechanisms. Specifically, an extensive literature survey is first carried out to identify the research efforts in this direction so far, both with and without Attention Mechanisms. Then, being aware of the strengths and weaknesses of published research in this area, this thesis examines the ability of neural architectures with Attention Mechanisms to predict this type of time series. Extensive research is conducted and multiple experiments are carried out to identify the neural architectures that can be placed before and after an attention mechanism, aiming to maximize the accuracy of the forecasts produced by the models presented. Furthermore, this research also studies how attention mechanisms can be combined within the same neural architecture to obtain the most accurate predictions. At the same time, special emphasis is placed on the explainability of the predictions produced, making the presented models more reliable. Additionally, this research highlights the importance of predicting the occurrence of special events that disrupt the regularity of the time series and proposes methodologies in this direction. Finally, this research proposes a new Transformer that is able to provide more reliable forecasts than those of the corresponding electricity providers and appears to have the ability to predict outliers in the time series to a remarkable extent without the use of additional information, which in the literature often comes from calendar or weather variables.

## Keywords

Timeseries Forecasting, Energy Load, Neural Networks, Attention Mechanism



# Περιεχόμενα

---

Ευχαριστίες	3
Περίληψη	5
Abstract	7
<b>1 Εισαγωγή</b>	<b>19</b>
1.1 Αντικείμενο της διπλωματικής	19
1.2 Οργάνωση του τόμου	19
<b>2 Παραδοσιακές Μέθοδοι Πρόβλεψης Χρονοσειρών</b>	<b>21</b>
2.1 Εισαγωγή στις Προβλέψεις Χρονοσειρών	21
2.2 Χαρακτηριστικά Χρονοσειρών	22
2.3 Συνήθεις Στατιστικές Μέθοδοι	24
2.3.1 Αφελής Μέθοδος	26
2.3.2 Μέθοδοι Μέσου Όρου	27
2.3.3 Μέθοδοι Εκθετικής Εξομάλυνσης	28
2.3.4 Μοντέλα ARIMA	34
2.3.5 Μέθοδος $\Theta$	35
2.3.6 Απλή και Πολλαπλή Παλινδρόμηση	37
2.4 Μέθοδοι Αξιολόγησης Μοντέλων Πρόβλεψης	39
<b>3 Μηχανική Μάθηση και Προβλέψεις Χρονοσειρών</b>	<b>41</b>
3.1 Εισαγωγή στη Μηχανική Μάθηση	41
3.1.1 Εποπτευόμενη Μάθηση	42
3.1.2 Μη Εποπτευόμενη Μάθηση	43
3.1.3 Ενισχυτική Μάθηση	45
3.2 SVM και SVR	46
3.2.1 Θεωρητικό Πλαίσιο	46
3.2.2 SVR και Προβλέψεις Χρονοσειρών	47
3.3 Δέντρα Απόφασης και Τυχαία Δάση	48
3.3.1 Θεωρητικό Πλαίσιο	49
3.3.2 Δέντρα Αποφάσεων, Τυχαία Δάση και Πρόβλεψη Χρονοσειρών	50

<b>4 Βαθιά Μάθηση και Προβλέψεις Χρονοσειρών</b>	<b>53</b>
4.1 Εισαγωγή στα Νευρωνικά Δίκτυα	53
4.2 Τεχνητός Νευρώνας	53
4.3 MLP	56
4.4 Εκπαίδευση Νευρωνικών Δικτύων	57
4.5 Πρόβλεψη Χρονοσειρών με MLPs	59
4.6 RNN	60
4.6.1 Αρχιτεκτονική	60
4.6.2 Εκπαίδευση RNN (BPTT)	61
4.6.3 LSTMs	62
4.6.4 LSTMs στην Πρόβλεψη Χρονοσειρών	64
4.7 CNN	65
4.7.1 Αρχιτεκτονική	65
4.7.2 Εκπαίδευση CNN	68
4.7.3 CNN στην Πρόβλεψη Χρονοσειρών	69
<b>5 Πρόβλεψη Ηλεκτρικού Φορτίου</b>	<b>73</b>
5.1 Η Σημασία της Πρόβλεψης Ηλεκτρικού Φορτίου	73
5.2 Παράγοντες που επηρεάζουν την Ακρίβεια των Προβλέψεων	75
5.3 Ταξινόμηση Προβλέψεων Ζήτησης Ηλεκτρικής Ενέργειας	76
5.4 Μέθοδοι Πρόβλεψης Ζήτησης Ηλεκτρικής Ενέργειας	77
<b>6 Μηχανισμός Προσοχής και Προβλέψεις Χρονοσειρών</b>	<b>83</b>
6.1 Εισαγωγή στο Μηχανισμό Προσοχής	83
6.2 Αρχιτεκτονική Μηχανισμού Προσοχής	84
6.2.1 Προσοχή Μιας Κεφαλής	84
6.2.2 Προσοχή Πολλαπλών Κεφαλών	85
6.3 Αυτοπροσοχή	85
6.4 Αυτοπροσοχή για Πρόβλεψη Χρονοσειρών	87
6.5 Κωδικοποιήσεις Θέσης	88
<b>7 Πειραματική Διαδικασία</b>	<b>91</b>
7.1 Δεδομένα και Προεπεξεργασία	91
7.2 Υπολογιστικοί Πόροι και Πειραματική Διάταξη	94
7.3 Νευρωνικές Αρχιτεκτονικές	94
7.3.1 Αποκλειστικά Μηχανισμοί Προσοχής	95
7.3.2 Αρχιτεκτονικές που προηγούνται του Μηχανισμού Προσοχής	97
7.3.3 Αρχιτεκτονικές που έπονται του Μηχανισμού Προσοχής	102
7.3.4 Ο Transformer	108
<b>8 Αποτελέσματα</b>	<b>111</b>
8.1 Αποκλειστικά Μηχανισμοί Προσοχής	111
8.2 Αρχιτεκτονικές που προηγούνται του Μηχανισμού Προσοχής	117
8.2.1 Αποτελέσματα από τη χρήση MLP πριν τον μηχανισμό προσοχής	117



---

8.2.2	Αποτελέσματα από τη χρήση κωδικοποιήσεων θέσης . . . . .	118
8.2.3	Αποτελέσματα από τη χρήση LSTM πριν τον μηχανισμό προσοχής. . .	121
8.2.4	Αποτελέσματα από τη χρήση CNN πριν τον μηχανισμό προσοχής. . . .	123
8.3	Αρχιτεκτονικές που έπονται του Μηχανισμού Προσοχής . . . . .	125
8.3.1	Χρήση υπολειμματικής σύνδεσης μετά τον Μηχανισμό Προσοχής . . . .	126
8.3.2	Χρήση MLP, LSTM, CNN μετά τον μηχανισμό προσοχής . . . . .	128
8.4	Ο Transformer . . . . .	130
<b>9</b>	<b>Επίλογος</b>	<b>133</b>
9.1	Συμπεράσματα . . . . .	133
9.2	Μελλοντικές Επεκτάσεις . . . . .	135
	<b>Βιβλιογραφία</b>	<b>149</b>
	<b>Συντομογραφίες - Αρχιτεκτονικές - Ακρωνύμια</b>	<b>151</b>



## Κατάλογος Σχημάτων

---

7.1	Το επίπεδο προσοχής, έτσι όπως κατασκευάστηκε στην παρούσα εργασία. . .	95
7.2	Η βασική αρχιτεκτονική προσοχής πάνω στην οποία χτίστηκαν οι υπόλοιπες. .	97
7.3	Η οικογένεια αρχιτεκτονικών MLP Attention Models . . . . .	98
7.4	Η οικογένεια αρχιτεκτονικών MLP Attention Models εμπλουτισμένη με υπο- λειμματικές συνδέσεις . . . . .	99
7.5	Η οικογένεια αρχιτεκτονικών Positional Attention Models . . . . .	99
7.6	Η οικογένεια αρχιτεκτονικών Positional Attention Models εμπλουτισμένη με υπολειμματικές συνδέσεις . . . . .	100
7.7	Η οικογένεια αρχιτεκτονικών LSTM Attention Models . . . . .	101
7.8	Η οικογένεια αρχιτεκτονικών Convolutional Attention Models . . . . .	102
7.9	Είδη υπολειμματικών συνδέσεων που εφαρμόστηκαν Γύρω από τον μηχανισμό προσοχής στα Convolutional Attention Models . . . . .	104
7.10	Είδη υπολειμματικών συνδέσεων που εφαρμόστηκαν Γύρω από τον μηχανισμό προσοχής στα Positional Attention Models . . . . .	104
7.11	Χρήση MLP μετά τον μηχανισμό προσοχής . . . . .	105
7.12	Χρήση LSTM μετά τον μηχανισμό προσοχής . . . . .	106
7.13	Χρήση CNN μετά τον μηχανισμό προσοχής . . . . .	107
7.14	Ο Transformer που κατασκευάστηκε . . . . .	110



## Κατάλογος Εικόνων

---

2.1	Χαρακτηριστικά Χρονοσειρών . . . . .	24
2.2	Κινητός Μέσος Όρος . . . . .	28
2.3	Εκθετική Εξομάλυνση . . . . .	34
2.4	Μέθοδος Theta . . . . .	37
3.1	Το SVM . . . . .	47
3.2	Το Random Forest . . . . .	50
3.3	Πρόβλεψη χρονοσειράς με δέντρο απόφασης . . . . .	51
4.1	Ο Τεχνητός Νευρώνας . . . . .	54
4.2	Το MLP . . . . .	57
4.3	Η γενική δομή των RNN . . . . .	61
4.4	Η δομική μονάδα των LSTM . . . . .	63
4.5	Πιθανές έξοδοι ενός επιπέδου LSTM . . . . .	65
4.6	Τυπική αρχιτεκτονική συνελικτικού νευρωνικού δικτύου . . . . .	66
4.7	Συνελικτικό νευρωνικό δίκτυο για πρόβλεψη χρονοσειρών . . . . .	71
6.1	Η αρχιτεκτονική του Transformer, έτσι όπως παρουσιάστηκε το 2017. . . . .	83
6.2	Ο μηχανισμός προσοχής. . . . .	84
6.3	Ο μηχανισμός προσοχής πολλαπλών κεφαλών. . . . .	85
7.1	Μία τυπική εβδομάδα στην Πολωνία . . . . .	93
7.2	Η επίδραση της ημέρας του Συντάγματος της 3ης Μαΐου στην κατανάλωση ρεύματος στην Πολωνία. . . . .	93
7.3	Οι εορτασμοί της ημέρας Ανεξαρτησίας συνέπεσαν με Σαββατοκύριακο το 2022 στην Πολωνία. . . . .	94
7.4	Η υπολειμματική σύνδεση που χρησιμοποιείται στον αρχικό Transformer . . . . .	103
8.1	Η επίδραση του πλήθους κεφαλών προσοχής στην προγνωστική ικανότητα του μοντέλου. . . . .	112
8.2	Η επίδραση του πλήθους κεφαλών προσοχής στον χρόνο εκπαίδευσης του μοντέλου. . . . .	112
8.3	Η επίδραση των βημάτων εκπαίδευσης στην προγνωστική ικανότητα του μοντέλου. . . . .	113
8.4	Η επίδραση του μεγέθους των πινάκων $Q, K, V$ στην προγνωστική ικανότητα του μοντέλου. . . . .	113

8.5	Η επίδραση του μεγέθους των πινάκων $Q, K, V$ στην ταχύτητα εκπαίδευσης του μοντέλου. . . . .	113
8.6	Οι κεφαλές ενός Basic Attention Model διαμορφωμένες μετά την είσοδο μίας τυπικής εβδομάδας. . . . .	114
8.7	Η εβδομάδα κατά την είσοδο της οποίας στο μοντέλο προκύπτουν οι κεφαλές της εικόνας 8.6 . . . . .	115
8.8	Οι κεφαλές ενός Basic Attention Model διαμορφωμένες μετά την είσοδο μίας εβδομάδας με ειδικό γεγονός. . . . .	116
8.9	Η εβδομάδα κατά την είσοδο της οποίας στο μοντέλο προκύπτουν οι κεφαλές της εικόνας 8.8 . . . . .	116
8.10	Το Basic Attention Model αδυνατεί να προβλέψει ειδικά γεγονότα. . . . .	117
8.11	Η επίδραση των πλήρως συνδεδεμένων επιπέδων στα MLP Attention Models. . . . .	118
8.12	Η επίδραση των κωδικοποιήσεων θέσης. . . . .	119
8.13	Απεικόνιση των κεφαλών προσοχής μετά την χρήση κωδικοποιήσεων θέσης. . . . .	120
8.14	Η είσοδος που χορηγήθηκε στο μοντέλο για να προκύψουν οι κεφαλές της εικόνας 8.13. . . . .	120
8.15	Η επίδραση του πλήθους των μονάδων του LSTM. . . . .	121
8.16	Η αύξηση του χρόνου εκπαίδευσης καθώς αυξάνονται οι μονάδες του LSTM επιπέδου. . . . .	122
8.17	Η επίδραση των πολλαπλών επιπέδων LSTM. . . . .	122
8.18	Η επίδραση του μεγέθους του συνελικτικού πυρήνα. . . . .	123
8.19	Η επίδραση του του πλήθους συνελικτικών φίλτρων. . . . .	124
8.20	Η επίδραση των διαφορετικών ειδών pooling. . . . .	124
8.21	Η επίδραση του πλήθους πλήρως συνδεδεμένων επιπέδων νευρώνων μετά το συνελικτικό επίπεδο. . . . .	125
8.22	Η επίδραση των υπολειμματικών συνδέσεων στην πρόβλεψη ειδικού γεγονότος στην Ιταλία. . . . .	127
8.23	Η επίδραση των υπολειμματικών συνδέσεων στην πρόβλεψη ειδικού γεγονότος στην Πολωνία. . . . .	128
8.24	Η επίδραση των υπολειμματικών συνδέσεων στην πρόβλεψη ειδικού γεγονότος στην Αυστρία. . . . .	128
8.25	Η τελική μορφή του Transformer . . . . .	132

## Κατάλογος Πινάκων

---

8.1	Σύγκριση των Basic Attention Models με τις προβλέψεις των παρόχων . . .	114
8.2	Σύγκριση πλήρως συνδεδεμένων επιπέδων νευρώνων με και χωρίς υπολειμματικές συνδέσεις. . . . .	118
8.3	Υπολειμματικές συνδέσεις μετά την εισαγωγή κωδικοποιήσεων θέσης . . . .	119
8.4	Σύγκριση με τις προβλέψεις των παρόχων . . . . .	119
8.5	Σύγκριση MLP, Positional, LSTM Attention Models με τις προβλέψεις των παρόχων. . . . .	122
8.6	Σύγκριση αποτελεσμάτων μετά την εξέταση όλων των βασικών οικογενειών μοντέλων. . . . .	125
8.7	Σύγκριση υπολειμματικών συνδέσεων στα Convolutional Attention Models .	126
8.8	Σύγκριση υπολειμματικών συνδέσεων στα Positional Attention Models . . .	127
8.9	Συγκεντρωτικά αποτελέσματα μετά και την προσθήκη υπολειμματικών συνδέσεων γύρω από τους μηχανισμούς προσοχής . . . . .	129
8.10	Παρουσίαση αποτελεσμάτων από την προσθήκη επιπλέον επιπέδων επεξεργασίας πάνω από τον μηχανισμό προσοχής . . . . .	129
8.11	Παρουσίαση της προγνωστικής ικανότητας του Transformer . . . . .	130
8.12	Χρόνοι εκπαίδευσης των υπόλοιπων μοντέλων συγκριτικά με τον Transformer	131





# Κεφάλαιο **1**

## Εισαγωγή

---

### 1.1 Αντικείμενο της διπλωματικής

Η ηλεκτρική ενέργεια είναι το όχημα για τη διατήρηση της εξαιρετικά προηγμένης τεχνολογικά παγκόσμιας οικονομίας. Σχεδόν κάθε δραστηριότητα στη σημερινή εποχή εξαρτάται από την ηλεκτρική ενέργεια. Παράλληλα, η πρόβλεψη της ζήτησης ηλεκτρικής ενέργειας αποτελεί κρίσιμο παράγοντα για την αποτελεσματική διαχείρισή της, την αξιοπιστία του δικτύου και τα βιώσιμα ενεργειακά συστήματα. Παρέχοντας πληροφορίες για τα μελλοντικά πρότυπα ζήτησης ηλεκτρικής ενέργειας, η πρόβλεψη φορτίου διευκολύνει την προληπτική λήψη αποφάσεων και τον σχεδιασμό υποδομών σε όλη την αλυσίδα παραγωγής και διανομής της ενέργειας. Είναι εμφανές πως η πρόβλεψη του φορτίου ηλεκτρικής ενέργειας σχετίζεται σε μεγάλο βαθμό με την ανάπτυξη της οικονομίας και την καθημερινή λειτουργία της κοινωνίας. Ως εκ τούτου, η ακρίβεια της πρόβλεψης ηλεκτρικού φορτίου έχει μεγάλη σημασία για τη γενικότερη ευημερία της κοινωνίας.

Για τους παραπάνω λόγους, η παρούσα εργασία ερευνά την κατασκευή νευρωνικών αρχιτεκτονικών που περιέχουν την πολύ ισχυρή δομή του Μηχανισμού Προσοχής με στόχο την παραγωγή όσο το δυνατόν πιο αξιόπιστων προβλέψεων ζήτησης ηλεκτρικής ενέργειας. Παράλληλα, πέραν της παραγωγής προβλέψεων ακριβείας, ένας ακόμη σημαντικός στόχος της παρούσας διπλωματικής εργασίας είναι και η διερεύνηση της επεξηγησιμότητας των προβλέψεων που προαναφέρθηκαν, έτσι ώστε να αυξηθεί και η αξιοπιστία των νευρωνικών αρχιτεκτονικών που θα κατασκευαστούν.

### 1.2 Οργάνωση του τόμου

Η εργασία αυτή είναι οργανωμένη σε εννιά κεφάλαια: Στο Κεφάλαιο 2 δίνεται ένα βασικό θεωρητικό υπόβαθρο σχετικά με τις χρονοσειρές και την πρόβλεψη των μελλοντικών τιμών τους, ενώ παρουσιάζονται και οι πιο δημοφιλείς παραδοσιακές στατιστικές μέθοδοι που χρησιμοποιούνται για την πρόβλεψη χρονοσειρών. Στο Κεφάλαιο 3 περιγράφεται συνοπτικά ο χώρος της Μηχανικής Μάθησης και γίνεται ανάλυση των μεθοδολογιών του συγκεκριμένου χώρου που συμβάλουν στην παραγωγή προβλέψεων μελλοντικών τιμών χρονοσειρών. Στο Κεφάλαιο 4 γίνεται μία εισαγωγή στην Βαθιά Μάθηση, ενώ παράλληλα ερευνώνται και οι νευρωνικές αρχιτεκτονικές που χρησιμοποιούνται συνήθως για την πρόβλεψη χρονοσειρών. Αφού έχουν παρουσιαστεί οι χρονοσειρές και οι βασικές μέθοδοι πρόβλεψης των επόμενων

τιμών τους, στο Κεφάλαιο 5 γίνεται μία εκτενής ανάλυση του προβλήματος της πρόβλεψης ηλεκτρικού φορτίου και παρουσιάζονται οι βασικότερες μεθοδολογίες που έχουν προταθεί από την ερευνητική κοινότητα για τον σκοπό αυτό. Στο Κεφάλαιο 6 παρουσιάζεται ο Μηχανισμός Προσοχής, κεντρικό θέμα της παρούσας διπλωματικής, ενώ εξηγείται και πώς μπορεί να συνδράμει στην πρόβλεψη χρονοσειρών. Στο Κεφάλαιο 7 παρουσιάζονται τα πειράματα που εκπονήθηκαν στην συγκεκριμένη έρευνα, ενώ στο Κεφάλαιο 8 δίνονται τα αποτελέσματα που προέκυψαν από τα πειράματα αυτά. Τέλος, στο Κεφάλαιο 9 παρουσιάζονται τα συμπεράσματα και η συνεισφορά της διπλωματικής εργασίας, ενώ περιγράφονται και μελλοντικές επεκτάσεις της συγκεκριμένης έρευνας.

# Παραδοσιακές Μέθοδοι Πρόβλεψης Χρονοσειρών

---

## 2.1 Εισαγωγή στις Προβλέψεις Χρονοσειρών

Η παραγωγή προβλέψεων αποτελεί ακρογωνιαίο λίθο των διαδικασιών λήψης αποφάσεων, όχι μόνο στον επιχειρηματικό τομέα, αλλά και σε όλο το ευρύ φάσμα της επιστημονικής έρευνας. Η ικανότητα πρόβλεψης μελλοντικών γεγονότων ή τάσεων διαδραματίζει κρίσιμο ρόλο στην καθοδήγηση στρατηγικών επιλογών τόσο για μικρές επιχειρήσεις όσο και για πολυεθνικές εταιρείες. Αυτές οι εκτιμήσεις για το μέλλον συχνά επικεντρώνονται σε βασικές μετρήσεις όπως ο όγκος παραγωγής, οι προσφορές υπηρεσιών και οι επενδυτικές στρατηγικές. Η ακρίβεια που παρέχουν αυτές οι προβλέψεις μπορούν να επηρεάσουν σημαντικά τη μελλοντική πορεία ενός οργανισμού, διαμορφώνοντας την κερδοφορία και την επιτυχία του. Κατά συνέπεια, ο κλάδος των προβλέψεων έχει συγκεντρώσει σημαντική προσοχή και ενδιαφέρον από όλη την επιστημονική κοινότητα και όχι μόνο, σημειώνοντας τις τελευταίες δεκαετίες αξιοσημείωτη πρόοδο στη μεθοδολογία και την αποτελεσματικότητα των τεχνικών προβλέψεων.

Η εξέλιξη των μεθοδολογιών πρόβλεψης και η ανάπτυξη συγκριτικών μετρήσεων για την αξιολόγηση της ακρίβειας των προβλέψεων έχουν αποτελέσει αντικείμενο εντατικής έρευνας και καινοτομίας. Ανάμεσα στις διάφορες έννοιες που είναι κεντρικές για την επιστήμη της πρόβλεψης, οι χρονοσειρές κατέχουν εξέχουσα θέση. Ορισμένες ως μια ακολουθία τιμών που καταγράφονται σε διαδοχικές χρονικές στιγμές, οι χρονοσειρές παρουσιάζουν ένα μοναδικό σύνολο δεδομένων όπου κάθε παρατήρηση ταξινομείται χρονολογικά. Αυτή η χρονική διάταξη πολλές φορές δίνει την εντύπωση πως οι τιμές μέσα σε μια χρονοσειρά εξαρτώνται εγγενώς η μία από την άλλη, υποδηλώνοντας ότι οι μελλοντικές τιμές της σειράς μπορούν να προβλεφθούν από τις προηγούμενες παρατηρήσεις της - μια αρχή γνωστή ως ντετερμινισμός.

Ωστόσο, η ντετερμινιστική οπτική στον τομέα των χρονοσειρών δεν ενσωματώνει πλήρως την πολυπλοκότητα που είναι εγγενής στα δεδομένα χρονοσειρών του πραγματικού κόσμου. Ενώ τα ιστορικά δεδομένα παρέχουν πράγματι τη βάση για την πρόβλεψη μελλοντικών παρατηρήσεων, η επιρροή εξωτερικών μεταβλητών - που κυμαίνονται από απρόβλεπτες οικονομικές αλλαγές, τεχνολογικές ανακαλύψεις έως τυχαία γεγονότα - εισάγει ένα επίπεδο αβεβαιότητας που δεν μπορεί να αγνοηθεί. Αυτοί οι παράγοντες συμβάλλουν στη στοχαστική φύση των χρονοσειρών, όπου η τυχαιότητα παίζει σημαντικό ρόλο στη διαμόρφωση των μελλοντικών παρατηρήσεων. Ως εκ τούτου, η πρόβλεψη χρονοσειρών συχνά προσεγγίζεται ως μια στοχα-

στική διαδικασία, αναγνωρίζοντας το μείγμα ντετερμινισμού και απρόβλεπτου στοιχείου που χαρακτηρίζει τα δεδομένα του πραγματικού κόσμου.

Η αλληλεπίδραση μεταξύ ντετερμινιστικών προτύπων και στοχαστικών επιρροών υπογραμμίζει τις υπάρχουσες προκλήσεις στον τομέα της πρόβλεψης χρονοσειρών. Καθώς οι επιχειρήσεις και οι επιστημονικές προσπάθειες βασίζονται όλο και περισσότερο στην ακρίβεια και την αξιοπιστία αυτών των προβλέψεων, το πεδίο συνεχίζει να εξελίσσεται, καθοδηγούμενο από τον στόχο της ενίσχυσης της ακρίβειας πρόβλεψης. Αυτή η συνεχής επιδίωξη βελτίωσης αντανακλά τον κρίσιμο ρόλο που διαδραματίζει η πρόβλεψη στην αντιμετώπιση των αβεβαιοτήτων του μέλλοντος.

Οι προβλέψεις χρονοσειρών μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες μεθόδων πρόβλεψης, τις ποιοτικές μεθόδους και τις ποσοτικές μεθόδους [1].

Οι ποιοτικές προσεγγίσεις βασίζονται σε υποκειμενική κρίση και όχι σε μαθηματικούς υπολογισμούς. Η ισχύς των ποιοτικών μεθόδων έγκειται στην ευελιξία τους να ενσωματώνουν νέες πληροφορίες όποτε κρίνεται απαραίτητο από το άτομο ή την επιτροπή ατόμων που κάνει την πρόβλεψη, καθώς και στην ικανότητά τους να αποσαφηνίζουν το σχεπτικό πίσω από κάθε πρόβλεψη. Τέτοιου είδους προβλέψεις είναι ιδιαίτερα χρήσιμες σε σενάρια όπου τα ιστορικά δεδομένα είναι περιορισμένα ή όταν τα μοτίβα στα δεδομένα χρονοσειρών φαίνονται ακανόνιστα και απρόβλεπτα. Από την άλλη πλευρά, η υποκειμενική φύση αυτών των μεθόδων, δηλαδή το γεγονός πως εξαρτώνται από τις ανθρώπινες γνώσεις και αντιλήψεις για τη δημιουργία προβλέψεων, τις καθιστά πολλές φορές αναξιόπιστες [2].

Αντίθετα, οι τεχνικές ποσοτικής πρόβλεψης χρησιμοποιούν μαθηματικά και υπολογιστικά μοντέλα για την ανάλυση και την επεξεργασία δεδομένων. Αυτές οι μέθοδοι χαρακτηρίζονται από την αντικειμενικότητά τους, καθώς βασίζονται αποκλειστικά στην ανάλυση αριθμητικών δεδομένων, χωρίς ανθρώπινη προκατάληψη ή συναισθήματα. Για να είναι αποτελεσματικά τα ποσοτικά μοντέλα, τα δεδομένα που χρησιμοποιούνται πρέπει όχι μόνο να είναι επαρκή αλλά και να παρουσιάζουν ένα συνεπές πρότυπο με την πάροδο του χρόνου. Όπως ήδη γίνεται αντιληπτό ένα σημαντικό μειονέκτημα αυτών των μεθόδων είναι η εξάρτησή τους από ιστορικά δεδομένα, καθιστώντας τις λιγότερο κατάλληλες σε χρονοσειρές με λίγα και θορυβώδη δεδομένα. Σε τέτοιες περιπτώσεις συχνά επιλέγονται ποιοτικές προσεγγίσεις πρόβλεψης, καθώς θεωρούνται πιο κατάλληλες.

Αυτή η εργασία θα επικεντρωθεί στην ανάπτυξη προβλέψεων με χρήση ποσοτικών μεθόδων, καθώς τα διαθέσιμα δεδομένα επαρκούν και οι χρονοσειρές που θα χρησιμοποιηθούν ακολουθούν εμφανή μοτίβα.

## 2.2 Χαρακτηριστικά Χρονοσειρών

Όπως έγινε αντιληπτό παραπάνω, οι χρονοσειρές διαδραματίζουν κεντρικό ρόλο σε διάφορους επιστημονικούς κλάδους καθώς και στην επιχειρηματική καθημερινότητα. Αυτή η ευρεία δυνατότητα εφαρμογής εξηγεί και την σημασία της ανάλυσης των χρονοσειρών. Η ανάλυση χρονοσειρών περιλαμβάνει την εξέταση των δεδομένων που συλλέγονται ή καταγράφονται σε διαδοχικές χρονικές στιγμές για να εξαχθούν συμπεράσματα και να προσδιοριστούν τα χαρακτηριστικά των δεδομένων. Οι παραδοσιακές στατιστικές προσεγγίσεις για την ανάλυση χρονοσειρών εστιάζουν στην αποσύνθεση μιας χρονοσειράς. Αυτή η διαδικασία αποσύνθεσης

διαίρει τις χρονοσειρές σε διακριτά στοιχεία, καθένα από τα οποία αντιπροσωπεύει υποκείμενα μοτίβα όπως η τάση, η εποχικότητα, οι κυκλικές διακυμάνσεις και οι ακανόνιστες διακυμάνσεις. Με τον εντοπισμό και την αξιολόγηση αυτών των παραγόντων, οι ερευνητές και οι αναλυτές μπορούν να βελτιώσουν την ακρίβεια των προβλέψεών τους. Επιπλέον, μια τέτοια αποσύνθεση όχι μόνο διευκολύνει την εξαγωγή προβλέψεων από τα δεδομένα, αλλά βοηθά επίσης στη διερεύνηση των ποιοτικών πτυχών της χρονοσειράς.

Παρακάτω, παρουσιάζονται τα κύρια ποιοτικά χαρακτηριστικά των χρονοσειρών των οποίων ο εντοπισμός είναι συχνά απαραίτητος για την παραγωγή ακριβών προβλέψεων. Αυτά τα χαρακτηριστικά εξάγονται κατά τη διαδικασία της αποσύνθεσης για να κατανοήσουμε καλύτερα τη δυναμική και τα εγγενή μοτίβα μέσα στα δεδομένα.

### **Τάση**

Η τάση αναφέρεται στη μακροπρόθεσμη κίνηση του μέσου επιπέδου των δεδομένων με την πάροδο του χρόνου, σηματοδοτώντας εάν η συμπεριφορά της σειράς είναι σε ανοδική ή καθοδική τροχιά [3]. Αν και οι τάση μπορεί να μην χαρακτηρίζεται πάντα από γραμμικότητα, η ανάλυση της χρονοσειράς μπορεί να αποκαλύψει τη γενική κατεύθυνση της κίνησης των τιμών. Για παράδειγμα, σε μια χρονοσειρά που αναλύει τη μέση παγκόσμια θερμοκρασία, θα μπορούσε κανείς να παρατηρήσει μια σταδιακή αύξηση κατά την πάροδο των χρόνων, υποδεικνύοντας μια τάση υπερθέρμανσης του πλανήτη.

### **Εποχικότητα**

Η εποχικότητα χαρακτηρίζεται από περιοδικές διακυμάνσεις σε μια χρονοσειρά μέσα σε μια καθορισμένη περίοδο, συνήθως λιγότερο από ένα έτος [4]. Αυτό το χαρακτηριστικό δεν υπάρχει σε όλες τις χρονικές σειρές. Μπορεί να παρατηρηθεί μόνο σε εκείνες τις χρονοσειρές με συχνότητες παρατήρησης που υποστηρίζουν εποχιακά μοτίβα (ώρα, ημέρα, μήνας). Οι λόγοι πίσω από την εποχικότητα είναι συχνά προφανείς και η συσχέτιση μεταξύ των παρατηρήσεων σε σταθερά διαστήματα καθιστά εύκολη την απομόνωση της εποχικότητας βελτιστοποιώντας με αυτό τον τρόπο την ανάλυση της χρονοσειράς και τις αντίστοιχες προβλέψεις. Ένα κλασικό παράδειγμα αντιπροσωπευτικό της ύπαρξης εποχικότητας σε μία χρονοσειρά είναι τα δεδομένα πωλήσεων ενός καταστήματος λιανικής, τα οποία συνήθως κορυφώνονται κατά την περίοδο των διακοπών κάθε χρόνο λόγω της αυξημένης διάθεσης για κατανάλωση.

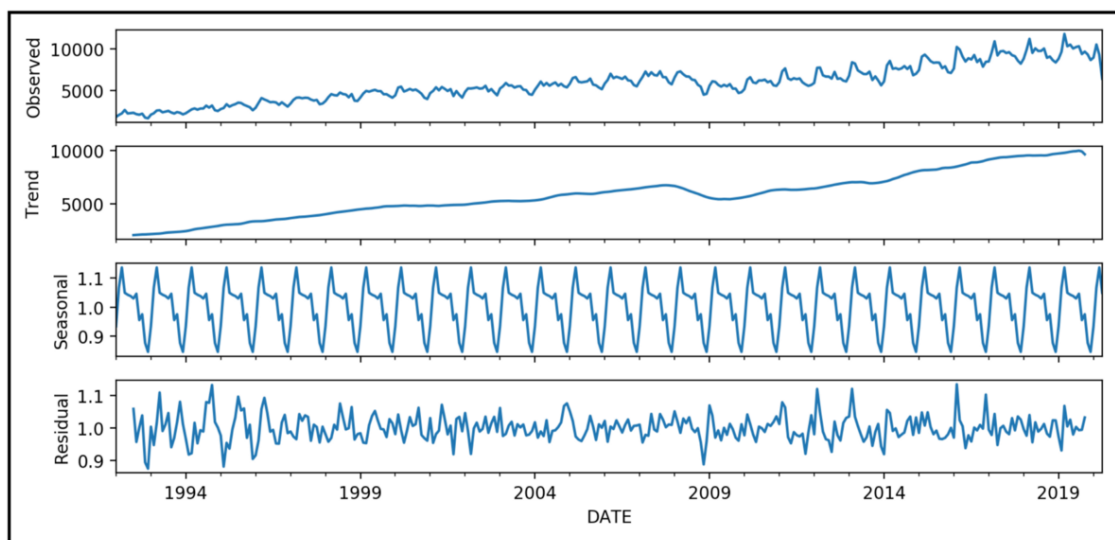
### **Κυκλικότητα**

Η κυκλικότητα γίνεται αντιληπτή από την ύπαρξη διακυμάνσεων στις τιμές των χρονοσειρών λόγω διαφόρων εξωτερικών παραγόντων [5]. Αυτές οι διακυμάνσεις είναι εν γένει πιο ακανόνιστες και απρόβλεπτες. Είναι σημαντικό να διακρίνουμε την κυκλικότητα από την εποχικότητα. Είναι γεγονός πως και οι δύο, όπως αναφέρθηκε και παραπάνω, δημιουργούν μοτίβα στις τιμές της χρονοσειράς. Ωστόσο, κατά τα άλλα, διαφέρουν σημαντικά καθώς ο εντοπισμός τους προκύπτει από την διερεύνηση διαφορετικών συχνοτήτων παρατήρησης. Η κυκλικότητα παρατηρείται σε διαστήματα συνήθως πολύ μεγαλύτερα του χρόνου [6], ενώ η εποχικότητα μπορεί να παρατηρηθεί σε διάστημα μηνών, ημερών αλλά ακόμη και ωρών. Συνεπώς, απαιτείται η ύπαρξη πληθώρας παρατηρήσεων για να μπορεί κανείς με βεβαιότητα να αναγνωρίσει το

χαρακτηριστικό της κυκλικότητας [3]. Ένα παράδειγμα κυκλικότητας θα μπορούσαν να είναι οι οικονομικοί κύκλοι που βιώνει μια χώρα, όπου οι περίοδοι ταχείας οικονομικής ανάπτυξης ακολουθούνται από ύφεση για αρκετά χρόνια.

### Ασυνέχειες

Οι ασυνέχειες προκύπτουν από την ύπαρξη δεδομένων εντός της χρονοσειράς που δεν μπορούν να προβλεφθούν από την ύπαρξη και τον εντοπισμό της εποχικότητας, τάσης ή κυκλικότητας. Πρόκειται για ξαφνικές μετατοπίσεις στη συμπεριφορά των τιμών της χρονοσειράς και συνήθως εκδηλώνονται ως ακραίες τιμές (outliers), μετατοπίσεις επιπέδου (level shifts) ή τυχαίες τιμές (randomness) [7]. Οι ακραίες τιμές, για παράδειγμα, ξεχωρίζουν σημαντικά από τα καθιερωμένα πρότυπα. Ένα παράδειγμα ακραίας τιμής θα μπορούσε να είναι μια ξαφνική αύξηση της ζήτησης ηλεκτρικής ενέργειας μετά από έναν απροσδόκητο καύσωνα. Αντίστοιχα, αλλαγή επιπέδου παρατηρείται όταν υφίσταται μία απότομη άνοδος ή πτώση στις τιμές μιας χρονοσειράς, από μία χρονική στιγμή και μετά [8]. Ένα παράδειγμα που θα μπορούσε να διευκολύνει την κατανόηση της έννοιας της αλλαγής επιπέδου είναι μία ασυνέχεια σε μια χρονοσειρά που αναπαριστά τη χρήση του διαδικτύου, στην οποία εμφανίζεται μια ξαφνική και μόνιμη αύξηση της κατανάλωσης δεδομένων κινητής τηλεφωνίας μετά την εισαγωγή μιας νέας υπηρεσίας. Όσον αφορά την τυχαιότητα, χαρακτηρίζει χρονοσειρές όπου οι παρατηρήσεις είναι ανεξάρτητες μεταξύ τους. Παρά τα διάφορα μοντέλα πρόβλεψης και αλγόριθμους, η πρόβλεψη της κατανομής των τιμών της χρονοσειράς είναι δύσκολη όταν τα δεδομένα προσομοιάζουν τον λευκό θόρυβο ή ακολουθούν αυτό που έχει γίνει γνωστό στη βιβλιογραφία ως τυχαίος περίπατος (random walk) [9].



Εικόνα 2.1: Χαρακτηριστικά Χρονοσειρών

## 2.3 Συνήθειες Στατιστικές Μέθοδοι

Τα παραπάνω χαρακτηριστικά, όπως μπορεί να γίνει αντιληπτό, αποτελούν την βάση της ανάλυσης των χρονοσειρών και συνεπώς και της πρόβλεψης των μελλοντικών τιμών τους.

Ωστόσο, εμβαθύνοντας στον τρόπο με τον οποίο οι στατιστικές μέθοδοι χρησιμοποιούνται για την πρόβλεψη των επόμενων τιμών των χρονοσειρών, ιδιαίτερη σημασία έχει η παρουσίαση των δύο βασικότερων «οικογενειών» στατιστικών μεθόδων πρόβλεψης.

Στο επίκεντρο των μεθόδων πρόβλεψης χρονοσειρών βρίσκεται η εκμετάλλευση της εξάρτησης από δεδομένα προηγούμενων τιμών της χρονοσειράς. Αυτή η προσέγγιση υποστηρίζεται προφανώς από την αμφιλεγόμενη υπόθεση ότι τα μοτίβα που παρατηρήθηκαν στο παρελθόν θα συνεχιστούν και στο μέλλον, επιτρέποντας έτσι να γίνουν προβλέψεις με επέκταση αυτών των προτύπων [10]. Όπως γίνεται εύκολα αντιληπτό, αυτή η λογική είναι άμεσα συνδεδεμένη με την αποσύνθεση των χρονοσειρών και τα εξαγόμενα χαρακτηριστικά τους που περιγράφηκαν προηγουμένως, καθώς η επέκταση της συμπεριφοράς αυτών των χαρακτηριστικών είναι που εν τέλει δίνει τις προβλέψεις. Τυπικά, αυτή η οικογένεια μεθόδων μπορεί να περιγραφεί με τον τύπο  $Y = f(X_1, X_2, \dots, X_n)$ , όπου το  $Y$  αντιπροσωπεύει την προβλεπόμενη μεταβλητή και τα  $X_1, X_2, \dots, X_n$  υποδηλώνουν ιστορικά σημεία δεδομένων που συλλέγονται σε αντίστοιχες χρονικές στιγμές. Η δύναμη αυτής της φιλοσοφίας έγκειται στην απλότητα και την αποτελεσματικότητά της σε καταστάσεις όπου η χρονοσειρά παρουσιάζει σταθερή και συνεπή συμπεριφορά με την πάροδο του χρόνου. Ωστόσο, το κύριο μειονέκτημά της είναι η αδυναμία να ληφθούν υπόψη οι εξωτερικές επιρροές - όπως οι οικονομικές διακυμάνσεις ή γενικότερα οι αλλαγές των περιβαλλοντικών συνθηκών - που μπορεί να επηρεάσουν τα προβλεπόμενα αποτελέσματα. Παρά αυτόν τον περιορισμό, αυτή η προσέγγιση παραμένει δημοφιλής για εφαρμογές όπου η πρόβλεψη των μελλοντικών τιμών μιας χρονοσειράς θεωρείται πιο κρίσιμη από την κατανόηση των λόγων πίσω από τις συγκεκριμένες προβλέψεις. Η χρησιμότητα αυτών των μοντέλων είναι ιδιαίτερα εμφανής σε τομείς όπου οι τάσεις του παρελθόντος είναι ισχυροί προγνωστικοί παράγοντες για τα μελλοντικά αποτελέσματα, όπως στην πρόβλεψη ζήτησης στη λιανική.

Σε αντίθεση με τα μοντέλα που βασίζονται καθαρά σε ιστορικά δεδομένα της ίδιας της χρονοσειράς, τα αιτιοκρατικά μοντέλα βασίζονται στην υπόθεση ότι η προβλεπόμενη μεταβλητή,  $y$ , επηρεάζεται από ένα σύνολο ανεξάρτητων μεταβλητών ( $X_1, X_2, \dots, X_n$ ). Αυτή η προσέγγιση αποκλίνει σημαντικά από την φιλοσοφία της οικογένειας μοντέλων που παρουσιάστηκε προηγουμένως, καθώς δεν υποθέτει μια προκαθορισμένη σχέση μεταξύ των εξαρτημένων και ανεξάρτητων μεταβλητών [11]. Η διαδικασία ξεκινά με τον εντοπισμό μεταβλητών που επηρεάζουν σημαντικά την προβλεπόμενη ποσότητα [12]. Έπειτα ακολουθεί ο προσδιορισμός της φύσης και των ιδιοτήτων της σχέσης μεταξύ των ανεξάρτητων μεταβλητών και της μεταβλητής-στόχου. Ένα ενδεικτικό παράδειγμα θα μπορούσε να είναι η πρόβλεψη ζήτησης ηλεκτρικής ενέργειας σε αστικές περιοχές, όπου η εξαρτημένη μεταβλητή (ζήτηση ενέργειας) επηρεάζεται από ανεξάρτητες μεταβλητές όπως η πληθυσμιακή αύξηση, οι εποχιακές αλλαγές, οι καιρικές συνθήκες και οι δραστηριότητες αστικής ανάπτυξης. Αυτό το μοντέλο απαιτεί πρόσβαση σε ένα ολοκληρωμένο σύνολο δεδομένων που περιλαμβάνει όχι μόνο τις ιστορικές τιμές της προβλεπόμενης μεταβλητής αλλά και λεπτομερείς πληροφορίες για όλες τις σχετικές ανεξάρτητες μεταβλητές. Παρά τη μεγαλύτερη πολυπλοκότητα και τις υψηλότερες υπολογιστικές απαιτήσεις, τα συγκεκριμένα μοντέλα προσφέρουν το πλεονέκτημα της προσαρμοστικότητας στις αλλαγές στο περιβάλλον του συστήματος, καθιστώντας τα ανεκτίμητα για σχεδιασμό και ανάλυση πολιτικής σε δυναμικά περιβάλλοντα.

Η διάκριση μεταξύ των μοντέλων που χρησιμοποιούν παρελθοντικά δεδομένα της ίδιας

της χρονοσειράς και των μοντέλων που χρησιμοποιούν ανεξάρτητες μεταβλητές για τον προσδιορισμό των μελλοντικών τιμών της χρονοσειράς αντιπροσωπεύει μια θεμελιώδη ταξινόμηση, ωστόσο το τοπίο των μεθόδων προβλέψεων είναι πολύ πιο πλούσιο. Τα υβριδικά μοντέλα που συνδυάζουν την προγνωστική δύναμη των ιστορικών δεδομένων με τις επεξηγηματικές ικανότητες των ανεξάρτητων παραγόντων κατέχουν ιδιαίτερα σημαντική θέση στον τομέα των προβλέψεων χρονοσειρών [13]. Αυτά τα μοντέλα συνδυάζουν τα πλεονεκτήματα των μεθόδων που παρουσιάστηκαν προηγουμένως προβλέποντας με ακρίβεια μελλοντικά γεγονότα, παρέχοντας παράλληλα πληροφορίες για τις σχέσεις μεταξύ διαφόρων παραγόντων.

Όπως είναι φυσικό, μετά την ανάλυση των δύο βασικότερων μεθοδολογιών πίσω από τις ποσοτικές στατιστικές μεθόδους, έχει ιδιαίτερη σημασία η παρουσίαση και η εξέταση των συνηθέστερων τεχνικών και από τις δύο οικογένειες μεθοδολογιών.

### 2.3.1 Αφελής Μέθοδος

Η αφελής μέθοδος πρόβλεψης, που συχνά θεωρείται η πιο στοιχειώδης προσέγγιση στο φάσμα των τεχνικών πρόβλεψης χρονοσειρών, χρησιμεύει κυρίως ως σημείο αναφοράς λόγω της απλότητάς της. Αυτή η μέθοδος λειτουργεί με μια απλή αρχή: θέτει ότι ο καλύτερος προγνωστικός παράγοντας της μελλοντικής τιμής σε οποιαδήποτε δεδομένη στιγμή  $t$  είναι η τιμή από την αμέσως προηγούμενη χρονική περίοδο ( $t - 1$ ). Επομένως, η προβλεπόμενη τιμή για το χρόνο  $t$ , που συμβολίζεται ως  $F(t)$ , αντιπροσωπεύεται μαθηματικά ως  $F(t) = Y(t - 1)$ , όπου  $Y(t - 1)$  συμβολίζει την παρατηρούμενη τιμή της χρονοσειράς τη χρονική στιγμή  $t - 1$  [14].

Η κύρια δύναμη της αφελούς μεθόδου έγκειται στις ελάχιστες υπολογιστικές της απαιτήσεις, καθιστώντας την μια αποτελεσματική επιλογή για τη δημιουργία προβλέψεων σε σύντομους ορίζοντες, συνήθως μια περίοδο μπροστά. Λόγω της εγγενούς απλότητάς της, η μέθοδος δεν παρέχει ιδιαίτερα ακριβείς προβλέψεις, ειδικά σε μεγαλύτερους χρονικούς ορίζοντες ή σε δεδομένα που παρουσιάζουν σημαντική μεταβλητότητα ή εποχικότητα. Κατά συνέπεια, η αφελής προσέγγιση χρησιμοποιείται σπάνια ως η μόνη μέθοδος πρόβλεψης σε πρακτικές εφαρμογές.

Επεκτείνοντας το βασικό μοντέλο της συγκεκριμένης μεθόδου, η εποχιακή παραλλαγή της ανταποκρίνεται στην ύπαρξη περιοδικών διακυμάνσεων που είναι εγγενείς σε πολλές χρονοσειρές. Αυτό το μοντέλο προβλέπει μελλοντικές τιμές χρησιμοποιώντας την τελευταία παρατηρούμενη τιμή από την ίδια εποχή ή περίοδο στον προηγούμενο εποχιακό κύκλο. Μαθηματικά, η εποχιακή αφελής πρόβλεψη εκφράζεται ως  $F(t) = Y(t - m(k + 1))$ , όπου το  $m$  αντιπροσωπεύει τη διάρκεια της εποχικής περιόδου και  $k$  είναι ο ακέραιος αριθμός που δηλώνει τον αριθμό των πλήρων εποχών που έχουν περάσει [15]. Αυτή η προσαρμογή της αφελούς μεθόδου είναι ιδιαίτερα κατάλληλη για σύνολα δεδομένων με έντονες εποχιακές διακυμάνσεις, αν και χρησιμοποιείται κυρίως ως σημείο αναφοράς για την αξιολόγηση πιο σύνθετων μεθόδων.

Εν ολίγοις, η αφελής μέθοδος και το εποχιακό αντίστοιχό της προσφέρουν μια βάση για συγκριτική αξιολόγηση στην πρόβλεψη χρονοσειρών. Ενώ η απλότητά τους είναι μια αρετή όσον αφορά την υπολογιστική απόδοση, είναι ακριβώς αυτό το χαρακτηριστικό που περιορίζει την εφαρμογή τους ως αυτόνομα εργαλεία πρόβλεψης σε πολύπλοκα ή εξαιρετικά μεταβαλλόμενα σύνολα δεδομένων.



### 2.3.2 Μέθοδοι Μέσου Όρου

Μεταξύ των διαφόρων τεχνικών που χρησιμοποιούνται για την πρόβλεψη, οι μέθοδοι που χρησιμοποιούν το μέσο όρο προηγούμενων τιμών, η μέθοδος απλού μέσου όρου και η μέθοδος κινούμενου μέσου ξεχωρίζουν για την απλότητα και την αποτελεσματικότητά τους, ειδικά στον χειρισμό δεδομένων που παρουσιάζουν σταθερά μοτίβα με την πάροδο του χρόνου χωρίς σημαντικές εποχιακές διακυμάνσεις.

Η Απλή Μέθοδος Μέσου Όρου χρησιμοποιεί μια απλή προσέγγιση όπου η πρόβλεψη για την επερχόμενη περίοδο ( $t + 1$ ) προσδιορίζεται με τον υπολογισμό του μέσου όρου όλων των διαθέσιμων παρατηρήσεων μέχρι την τρέχουσα στιγμή  $t$ . Αυτή η μέθοδος εκφράζεται μαθηματικά ως εξής:

$$F_{t+1} = \frac{1}{t} \sum_{i=1}^t Y_i$$

όπου  $F_{t+1}$  είναι η πρόβλεψη για την επόμενη χρονική στιγμή και το  $Y_i$  αντιπροσωπεύει τις προηγούμενες χρονικά παρατηρήσεις στη χρονοσειρά [16]. Το κύριο πλεονέκτημα αυτής της μεθόδου έγκειται στην απλότητά της και στην ίση στάθμιση που δίνεται σε όλες τις παρατηρήσεις, καθιστώντας την ιδιαίτερα κατάλληλη για χρονοσειρές που επιδεικνύουν ένα σταθερό μοτίβο με την πάροδο του χρόνου χωρίς έντονες εποχιακές διακυμάνσεις. Ωστόσο, η εξάρτησή από μεγάλο όγκο δεδομένων σημαίνει ότι μπορεί να μην προσαρμοστεί γρήγορα στις πρόσφατες αλλαγές στο πρότυπο δεδομένων, περιορίζοντας την αποτελεσματικότητά του στην πρόβλεψη μη στάσιμων χρονοσειρών.

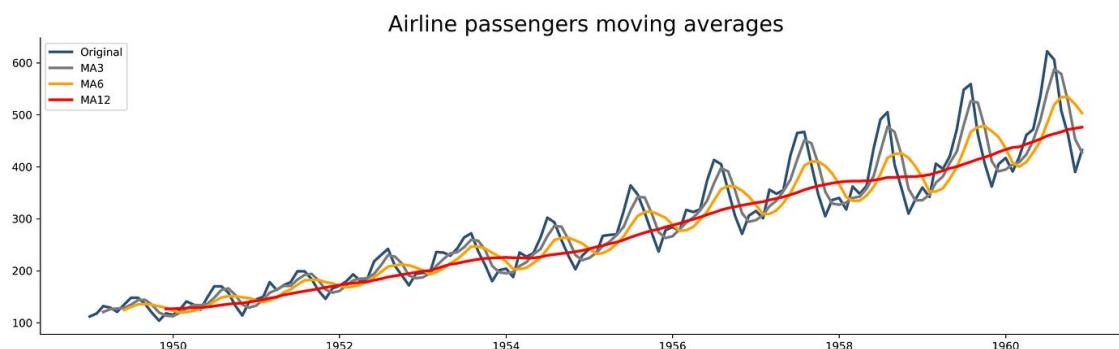
Η μέθοδος κινητού μέσου όρου ενισχύει την απλή μέθοδο μέσου όρου εστιάζοντας σε ένα υποσύνολο των πιο πρόσφατων παρατηρήσεων. Αυτή η μέθοδος υπολογίζει την πρόβλεψη για την επόμενη χρονική στιγμή υπολογίζοντας τον μέσο όρο των  $k$  πιο πρόσφατων παρατηρήσεων. Ο τύπος για την πρόβλεψη του κινητού μέσου όρου δίνεται από το

$$F_{t+1} = \frac{1}{k} \sum_{i=t-k+1}^t Y_i$$

Αυτή η προσέγγιση επιτρέπει μια πιο δυναμική διαδικασία πρόβλεψης, καθώς ενημερώνει συνεχώς το σύνολο δεδομένων που χρησιμοποιείται για τον υπολογισμό των μελλοντικών τιμών, ενσωματώνοντας τα πιο πρόσφατα δεδομένα και αγνοώντας τα παλαιότερα. Η μέθοδος του κινητού μέσου όρου είναι ιδιαίτερα ικανή να εξομαλύνει τις βραχυπρόθεσμες διακυμάνσεις και να επισημαίνει μακροπρόθεσμες τάσεις ή κύκλους εντός της χρονοσειράς ενώ παράλληλα έχει την δυνατότητα να προσαρμόζεται πιο εύκολα σε δυναμικώς μεταβαλλόμενες καταστάσεις. Ωστόσο, όπως και η μέθοδος του απλού μέσου όρου, αδυνατεί να προβλέψει ξαφνικές αλλαγές, καθώς η πρόβλεψη βασίζεται αποκλειστικά στον μέσο όρο των πρόσφατων παρατηρήσεων χωρίς να λαμβάνονται υπόψη άλλοι παράγοντες ή μοτίβα που μπορεί να επηρεάσουν τις μελλοντικές τιμές [16].

Συμπερασματικά, τόσο ο απλός όσο και ο κινητός μέσος όρος χρησιμεύουν ως θεμελιώδη εργαλεία στο οπλοστάσιο των τεχνικών πρόβλεψης χρονοσειρών. Η χρησιμότητά τους είναι ιδιαίτερα εμφανής σε σταθερές χρονοσειρές, όπου μπορούν να παρέχουν αξιόπιστες προβλέψεις ή να χρησιμεύσουν ως σημεία αναφοράς για την αξιολόγηση πιο σύνθετων μοντέλων. Αξίζει

επιπλέον να σημειωθεί ότι ο καθορισμός του βέλτιστου μήκους  $\chi$  για τον υπολογισμό του κινητού μέσου όρου απαιτεί προσεκτική εξέταση για να εξισορροπηθεί η δυναμική απόκριση του μοντέλου με την αποτελεσματικότητα εξομάλυνσης ακραίων τιμών που δεν είναι ενδεικτικές ως προς τις μελλοντικές τιμές της χρονοσειράς. Όπως συμβαίνει με κάθε μέθοδο πρόβλεψης, η επιλογή του συγκεκριμένου μοντέλου θα πρέπει να βασίζεται στα χαρακτηριστικά των δεδομένων και τους στόχους της άσκησης πρόβλεψης.



Εικόνα 2.2: Κινητός Μέσος Όρος

### 2.3.3 Μέθοδοι Εκθετικής Εξομάλυνσης

Οι μέθοδοι εκθετικής εξομάλυνσης αντιπροσωπεύουν μια κομβική προσέγγιση στον κόσμο των χρονοσειρών. Γνωστές για την απλή εφαρμογή τους, τις ελάχιστες υπολογιστικές απαιτήσεις και τον μικρό σχετικά όγκο δεδομένων που απαιτείται για τη δημιουργία αξιόπιστων προβλέψεων, αυτές οι τεχνικές έχουν εδραιωθεί ως απαραίτητα εργαλεία στον τομέα των προβλέψεων. Η ουσία της εκθετικής εξομάλυνσης έγκειται στην ικανότητά της να φιλτράρει τον θόρυβο που είναι εγγενής στα δεδομένα χρονοσειρών, αποσπώντας έτσι τα υποκείμενα μοτίβα για πιο ακριβείς μελλοντικές προβλέψεις. Οι εμπειρικές μελέτες έχουν αποδείξει την ακρίβεια των τεχνικών εκθετικής εξομάλυνσης, ακόμη και όταν αντιπαρατίθενται με πιο εξελιγμένες μεθόδους πρόβλεψης.

Η μεθοδολογία στην οποία βασίζεται η εκθετική εξομάλυνση εκτείνεται πέρα από τις απλές τεχνικές υπολογισμού του μέσου όρου, όπου παραδοσιακά, κάθε παρατήρηση συμβάλλει εξίσου στο προβλεπόμενο αποτέλεσμα. Στο επίκεντρο της εκθετικής εξομάλυνσης βρίσκεται η αρχή ότι οι πρόσφατες παρατηρήσεις έχουν μεγαλύτερη πληροφοριακή αξία για το μέλλον, οδηγώντας σε ένα σχήμα στάθμισης όπου η σημασία των τιμών της χρονοσειράς μειώνεται εκθετικά καθώς κάποιος μετακινείται πίσω στο χρόνο. Αυτό το μεθοδολογικό πλαίσιο είναι ιδιαίτερα επωφελές για βραχυπρόθεσμες έως και μεσοπρόθεσμες προβλέψεις.

Παρακάτω παρουσιάζονται τα βασικότερα μοντέλα εκθετικής εξομάλυνσης.

#### Απλή Εκθετική Εξομάλυνση

Η μέθοδος της Απλής Εκθετικής Εξομάλυνσης (Simple Exponential Smoothing, SES) είναι μια από τις βασικότερες προσεγγίσεις στην πρόβλεψη χρονοσειρών, ιδιαίτερα αποτελεσματική για χρονοσειρές που δεν έχουν έντονες τάσεις ή εποχικότητα. Η SES χαρακτηρίζεται

από την απλότητα και την προσαρμοστικότητα της, καθιστώντας την ένα ευέλικτο εργαλείο στην πρόβλεψη.

Στον πυρήνα της SES βρίσκονται δύο βασικές εξισώσεις που ορίζουν τη λειτουργία του μοντέλου:

- Ο όρος σφάλματος  $e_t$  είναι η απόκλιση μεταξύ της πραγματικής παρατηρούμενης τιμής  $Y_t$  τη στιγμή  $t$  και της προβλεπόμενης τιμής  $F_t$  για την ίδια περίοδο. Αυτή η απόκλιση ορίζεται και μαθηματικά ως:

$$e_t = Y_t - F_t$$

- Η πρόβλεψη  $F_t$  για οποιαδήποτε δεδομένη στιγμή αναθεωρείται με βάση την προηγούμενη πρόβλεψη  $F_{t-1}$  και τον τρέχοντα όρο σφάλματος  $e_t$ , με τον συντελεστή εξομάλυνσης  $\alpha$  να μεσολαβεί σε αυτήν την προσαρμογή. Η διαδικασία αναθεώρησης περιγράφεται ως εξής:

$$F_t = F_{t-1} + \alpha \cdot e_t$$

Αυτή η προσαρμογή διασφαλίζει ότι η πρόβλεψη για την επόμενη περίοδο λαμβάνει υπόψη το σφάλμα της προηγούμενης πρόβλεψης, ενσωματώνοντας έτσι διορθώσεις με βάση τα πιο πρόσφατα παρατηρούμενα δεδομένα [17].

Συνεπώς, το μοντέλο SES καταλήγει σε έναν απλοποιημένο τύπο όπου η πρόβλεψη  $F_t$  είναι ένα σταθμισμένο άθροισμα της τρέχουσας παρατήρησης  $S_t$  και της προηγούμενης πρόβλεψης  $F_{t-1}$ , και ορίζεται ως:

$$F_t = \alpha S_t + (1 - \alpha) F_{t-1}$$

Το αρχικό βήμα για την εφαρμογή του μοντέλου SES περιλαμβάνει τον καθορισμό της αρχικής πρόβλεψης  $F_1$  και του επιπέδου  $S_0$ , καθοριστικής σημασίας για την έναρξη της διαδικασίας πρόβλεψης. Μπορούν να χρησιμοποιηθούν διάφορες στρατηγικές για τον προσδιορισμό αυτών των αρχικών τιμών, όπως η χρήση του μέσου όρου όλων των παρατηρήσεων, του μέσου όρου των αρχικών  $n$  παρατηρήσεων ή η χρήση της τελευταίας παρατήρησης. Η επιλογή του αρχικού επιπέδου είναι κρίσιμη και θα πρέπει να αντανακλά τα χαρακτηριστικά της χρονοσειράς για να διασφαλιστεί η ακρίβεια της πρόβλεψης και η βέλτιστη επιλογή παραμέτρων [18].

Ο προσδιορισμός της βέλτιστης τιμής του  $\alpha$  περιλαμβάνει την εξισορρόπηση δύο βασικών παραγόντων: την ποσότητα του θορύβου στη χρονοσειρά και τη σταθερότητα του μέσου όρου της σειράς. Ένα χαμηλότερο  $\alpha$  προτιμάται για πιο θορυβώδη δεδομένα για να μειώσει τον αντίκτυπο των τυχαίων διακυμάνσεων. Αντίθετα, ένα υψηλότερο  $\alpha$  επιτρέπει στο μοντέλο να ανταποκρίνεται πιο γρήγορα σε σημαντικές αποκλίσεις από τη μέση τιμή της χρονοσειράς.

Συνοψίζοντας, η SES ξεχωρίζει για την αποτελεσματικότητά της σε σενάρια όπου τα δεδομένα χρονοσειρών δεν έχουν σαφείς τάσεις ή εποχιακά μοτίβα. Μέσω της εξάρτησής της από έναν αναδρομικό τύπο που βελτιώνει σταδιακά τις προβλέψεις με κάθε νέα παρατήρηση, η SES έχει εδραιώσει τη θέση της στον χώρο των προβλέψεων καθώς προσφέρει ένα ισορροπημένο και προσαρμοστικό πλαίσιο.

## Γραμμική Εκθετική Εξομάλυνση

Η μέθοδος Holt που είναι γνωστή και ως γραμμική εκθετική εξομάλυνση είναι ένα προηγμένο μοντέλο πρόβλεψης που έχει σχεδιαστεί για χρονοσειρές που παρουσιάζουν μια συνιστώσα τάσης. Αυτή η μέθοδος επεκτείνει την τεχνική απλής εκθετικής εξομάλυνσης (SES) ενσωματώνοντας έναν παράγοντα τάσης, καθιστώντας την ικανή στο χειρισμό τόσο των αλλαγών επιπέδου όσο και των τάσεων με την πάροδο του χρόνου [19].

Στο επίκεντρο της μεθόδου του Holt βρίσκονται δύο κρίσιμες εξισώσεις που καθορίζουν τη λειτουργία της:

- Η εξίσωση επιπέδου ενημερώνει το τρέχον επίπεδο  $S_t$  λαμβάνοντας υπόψη το τρέχον σφάλμα καθώς και την τάση που παρατηρείται:

$$S_t = S_{t-1} + T_{t-1} + \alpha \cdot e_t$$

- Όπως γίνεται αντιληπτό, κεντρική θέση έχει η τάση  $T_t$ , η οποία προσαρμόζεται με βάση το τρέχον σφάλμα και τη σταθερά εξομάλυνσης για την τάση  $\beta$ :

$$T_t = T_{t-1} + \beta \cdot \alpha \cdot e_t$$

Η συνάρτηση πρόβλεψης επιχειρώντας να εισάγει στον υπολογισμό την έννοια της κλίσης, ορίζει την πρόβλεψη  $F_{t+m}$  για  $m$  βήματα μπροστά να είναι το άθροισμα της τιμής του τελευταίου επιπέδου συν  $m$  φορές την τελευταία εκτιμώμενη τιμή τάσης. Αυτή η διατύπωση επιτρέπει γραμμικά αύξουσες ή φθίνουσες προβλέψεις που αντικατοπτρίζουν με μεγαλύτερη ακρίβεια τις υποκείμενες τάσεις των δεδομένων:

$$F_{t+m} = S_t + m \cdot T_t$$

Για να την αρχικοποίηση του μοντέλου του Holt, είναι απαραίτητος ο ορισμός ενός αρχικού επιπέδου  $S_0$  και μιας αρχικής τάσης  $T_0$ . Αυτές οι αρχικές τιμές μπορούν να επηρεάσουν σημαντικά την ακρίβεια του μοντέλου και την επακόλουθη ποιότητα πρόβλεψης. Οι συνήθειες πρακτικές για τη ρύθμιση του  $S_0$  περιλαμβάνουν τη χρήση του μέσου όρου των τελευταίων παρατηρήσεων ή της ίδιας της τελευταίας τιμής, παρόμοια με το SES [20]. Για το  $T_0$ , οι επιλογές περιλαμβάνουν τη διαφορά μεταξύ των δύο τελευταίων παρατηρήσεων ή μια κλίση που προέρχεται από ένα απλό μοντέλο γραμμικής παλινδρόμησης σε ένα σύνολο παρατηρήσεων.

Η επιλογή των σταθερών εξομάλυνσης,  $\alpha$  για το επίπεδο και  $\beta$  για την τάση, είναι κρίσιμη. Αυτές οι παράμετροι, και οι δύο περιορισμένες στο διάστημα  $[0, 1]$ , εξισορροπούν τη βαρύτητα που δίνεται στα πρόσφατα παρατηρούμενα δεδομένα έναντι της ιστορικής τάσης πρόβλεψης [21]. Η επιλογή αυτών των σταθερών, παράλληλα με το αρχικό επίπεδο και την τάση, πρέπει να γίνει με σύνεση με βάση τα χαρακτηριστικά των ιστορικών δεδομένων για να διασφαλιστεί η αποτελεσματικότητα του μοντέλου στην πρόβλεψη. Συνοψίζοντας, η μέθοδος εκθετικής εξομάλυνσης του Holt είναι ιδιαίτερα χρήσιμη για μεσοπρόθεσμες προβλέψεις όπου τα δεδομένα παρουσιάζουν μια αναγνωρίσιμη γραμμική τάση. Ενσωματώνοντας πληροφορία για την τάση της χρονοσειράς, το μοντέλο του Holt προσφέρει μια πιο διαφοροποιημένη προσέγγιση από

το SES, παρέχοντας προβλέψεις που αντιστοιχούν τόσο στο τρέχον επίπεδο όσο και στην κατεύθυνση προς την οποία κινούνται τα δεδομένα. Η προσαρμοστικότητα και η ακρίβεια αυτού του μοντέλου το καθιστούν μια προτιμώμενη επιλογή για πρόβλεψη σε πολλές εφαρμογές, που κυμαίνονται από τη διαχείριση αποθεμάτων έως την ανάλυση χρηματοοικονομικών αγορών.

### Μη Γραμμική Εκθετική Εξομάλυνση

Η μέθοδος μη γραμμικής εκθετικής εξομάλυνσης, ή αλλιώς Damped Exponential Smoothing (Damped), που εισήχθη από τους Gardner και McKenzie το 1985, αντιπροσωπεύει μια σημαντική βελτίωση του μοντέλου γραμμικής τάσης του Holt [21, 22, 23]. Αυτή η ενίσχυση προέκυψε ως απάντηση στην παρατηρούμενη τάση του μοντέλου γραμμικής τάσης να παράγει υπερβολικά αισιόδοξες προβλέψεις, ιδιαίτερα για μεσοπρόθεσμες έως μακροπρόθεσμες προβλέψεις. Η μέθοδος Δαμπεδ εισάγει έναν παράγοντα απόσβεσης, που συμβολίζεται με  $\phi$ , στις εξισώσεις πρόβλεψης, εγκαθιδρύοντας έτσι έναν μηχανισμό για να μετριάσει τη θετική προκατάληψη που χαρακτηρίζει την μέθοδο του Holt. Αυτή η μέθοδος έχει αποτελέσει αντικείμενο εκτενούς μελέτης και εφαρμογής, αποδεικνύοντας τη χρησιμότητά της σε διάφορους τομείς πρόβλεψης.

Το μοντέλο μη γραμμικής εκθετικής εξομάλυνσης προσαρμόζει τις προβλεπόμενες τιμές μέσω ενός συνόλου εξισώσεων που λαμβάνουν υπόψη τα παρατηρούμενα σφάλματα, το επίπεδο και την τάση της σειράς καθώς και την μεταβολή αυτής. Αυτές οι εξισώσεις επιτρέπουν στο μοντέλο να προσαρμόζει δυναμικά τις προβλέψεις του. Η αναλυτική μαθηματική έκφραση του μοντέλου παρουσιάζεται παρακάτω:

1. Διόρθωση σφάλματος:  $e_t = Y_t - F_t$ , όπου το  $e_t$  αντιπροσωπεύει την απόκλιση μεταξύ της παρατηρούμενης τιμής  $Y_t$  και της προβλεπόμενης τιμής  $F_t$ .
2. Προσαρμογή επιπέδου:  $S_t = S_{t-1} + \phi \cdot T_{t-1} + \alpha \cdot e_t$ , αυτή η εξίσωση ενημερώνει το επίπεδο της σειράς.
3. Ενημέρωση τάσης:  $T_t = \phi \cdot T_{t-1} + \alpha \cdot \beta \cdot e_t$ , η οποία αναθεωρεί την τάση με βάση την προηγούμενη τάση, τροποποιημένη από τον παράγοντα απόσβεσης και το τρέχον σφάλμα.
4. Μελλοντική πρόβλεψη:  $F_{t+m} = S_t + (\sum_{i=1}^m \phi^i) \cdot T_t$ , πρόβλεψη μελλοντικών τιμών προσαρμόζοντας το επίπεδο και τάση στον προβλεπόμενο ορίζοντα.

Ο συντελεστής απόσβεσης  $\phi$  παίζει καθοριστικό ρόλο στη λειτουργικότητα του μοντέλου:

- Με  $\phi = 0$ , το μοντέλο απλοποιείται υποθέτοντας πως δεν υπάρχει τάση (παρόμοια με την απλή εκθετική εξομάλυνση).
- Μια τιμή  $\phi$  μεταξύ 0 και 1 υποδηλώνει μειούμενη μη γραμμική εκθετική εξομάλυνση, η οποία είναι ιδιαίτερα ικανή να χειρίζεται μεσοπρόθεσμες προβλέψεις συγκρατώντας την υπερβολική αισιοδοξία.
- Με  $\phi = 1$  προκύπτει το μοντέλο γραμμικής τάσης του Holt, εφαρμόζοντας μια σταθερή τάση με την πάροδο του χρόνου.

- Τιμές  $\phi$  μεγαλύτερες από 1 γενικά αποφεύγονται λόγω της τάσης τους να παράγουν προβλέψεις με υψηλή θετική προκατάληψη, αντίστοιχη μιας μη ρεαλιστικής εκθετικής τάσης [24].

Ο προσδιορισμός των βέλτιστων τιμών για το  $\phi$ , μαζί με το  $\alpha$  (ο συντελεστής εξομάλυνσης του επιπέδου) και το  $\beta$  (ο συντελεστής εξομάλυνσης τάσης), είναι κρίσιμος για την επιτυχία του μοντέλου [25]. Αυτές οι παράμετροι τυπικά καθορίζονται μέσω τεχνικών βελτιστοποίησης που στοχεύουν στην ελαχιστοποίηση των σφαλμάτων της πρόβλεψης, διασφαλίζοντας έτσι την προγνωστική ακρίβεια του μοντέλου. Η εκκίνηση του μοντέλου περιλαμβάνει την εκτίμηση του αρχικού επιπέδου και της τάσης, συχνά μέσω μεθόδων παρόμοιες με εκείνες που χρησιμοποιούνται σε άλλα μοντέλα εκθετικής εξομάλυνσης, όπως η απλή γραμμική παλινδρόμηση.

Οι πρακτικές εφαρμογές της μεθόδου Damped είναι τεράστιες και ποικίλες, καλύπτοντας τη διαχείριση αποθεμάτων, την οικονομική πρόβλεψη και όχι μόνο. Η ανάπτυξή του όχι μόνο σηματοδότησε ένα ορόσημο στην πρόοδο των μεθοδολογιών πρόβλεψης, αλλά υπογράμμισε επίσης τη σημασία της προσαρμοστικότητας και της ακρίβειας στην προγνωστική μοντελοποίηση [26].

Συμπερασματικά, το μοντέλο μη γραμμικής εκθετικής εξομάλυνσης από τους Gardner και McKenzie [24] έχει επηρεάσει βαθιά το πεδίο της πρόβλεψης χρονοσειρών, προσφέροντας μια εξελιγμένη και πιο αξιόπιστη εναλλακτική λύση στο μοντέλο γραμμικής τάσης. Η ικανότητα αυτής της μεθόδου να παράγει συγκρατημένες και πιο ακριβείς μακροπρόθεσμες προβλέψεις υπογραμμίζει τη συνεχή ανάγκη για καινοτομία και βελτίωση στις μεθοδολογίες πρόβλεψης.

## Εποχιακή Εξομάλυνση

Τα μοντέλα που περιγράφηκαν στις προηγούμενες παραγράφους είναι μη εποχιακά μοντέλα και είναι κατάλληλα για χρονοσειρές που εμφανίζουν μόνο πρότυπα τάσεων. Στην περίπτωση που η εποχιακή συνιστώσα είναι έντονη στα ιστορικά δεδομένα, τα εποχιακά μοντέλα επεκτείνονται με την προσθήκη ενός εξομαλυσμένου εποχιακού παράγοντα. Η ακριβής μαθηματική περιγραφή των εποχιακών διακυμάνσεων είναι πρωταρχικής σημασίας για τη δημιουργία αξιόπιστων προβλέψεων. Οι εποχιακές διακυμάνσεις μπορούν να επηρεάσουν σημαντικά τη συνολική τάση και τα μοτίβα σε ένα σύνολο δεδομένων, καθιστώντας τη διάκριση μεταξύ προσθετικής και πολλαπλασιαστικής εποχικότητας κρίσιμη. Ενώ η προσθετική εποχικότητα υποστηρίζει ότι οι εποχιακές διακυμάνσεις παραμένουν σταθερές με την πάροδο του χρόνου, η πολλαπλασιαστική εποχικότητα υποδηλώνει ότι αυτές οι διακυμάνσεις είναι ανάλογες με το επίπεδο της σειράς [27]. Αυτή η λεπτή διαφορά των εποχιακών επιδράσεων είναι κεντρικής σημασίας για την εφαρμογή των μεθόδων Winters, οι οποίες έχουν σχεδιαστεί για να προσαρμόζονται στην περίπλοκη δυναμική των χρονοσειρών που παρουσιάζουν εποχιακά μοτίβα.

Η λογική της προσθετικής εποχικότητας υποστηρίζει ότι οι εποχιακές διακυμάνσεις είναι συνεπείς και δεν ποικίλλουν με το επίπεδο της χρονοσειράς. Αυτό το χαρακτηριστικό είναι εμφανές σε σενάρια όπου εξωτερικοί, επαναλαμβανόμενοι παράγοντες επηρεάζουν τα δεδομένα με ομοίμορφο τρόπο σε διαφορετικές περιόδους, όπως οι σταθερές αλλαγές θερμοκρασίας που επηρεάζουν την κατανάλωση ενέργειας ή οι σταθερές επιπτώσεις των διακοπών στις πωλήσεις. Στην προσθετική εποχικότητα, η εποχιακή συνιστώσα απλώς προστίθεται ή αφαιρείται από τη συνιστώσα τάσης.

Η πολλαπλασιαστική εποχικότητα υποθέτει ότι τα εποχιακά φαινόμενα κλιμακώνονται με την πάροδο της χρονοσειράς. Ένα τέτοιο μοτίβο παρατηρείται συχνά σε οικονομικά, περιβαλλοντικά και λιανικά δεδομένα, όπου ο αντίκτυπος της εποχικότητας εντείνεται ή μειώνεται καθώς αλλάζει το συνολικό επίπεδο δραστηριότητας [28]. Για λόγους συντομίας και απλότητας παρακάτω θα γίνει αναφορά μόνο στην λογική της πολλαπλασιαστικής εποχικότητας.

Η μέθοδος Ωιντερς [29] αναπτύχθηκε για να αντιμετωπίσει τις προκλήσεις της εποχικότητας, καθώς βελτιώνει τις προβλέψεις ενσωματώνοντας έναν εποχιακό παράγοντα προσαρμογής που προσαρμόζεται στο επίπεδο της σειράς. Έτσι, προσαρμόζοντας αυτή τη λογική στις εξισώσεις της απλής εκθετικής εξομάλυνσης, προκύπτουν τα παρακάτω:

1. Συντελεστής εποχικής προσαρμογής:

$$I_t = I_{t-p} + \gamma \cdot \left( \frac{e_t}{S_{t-1}} \right)$$

Εδώ, το  $I_t$  είναι ο εποχιακός δείκτης για την περίοδο  $t$ , που αντικατοπτρίζει την εποχιακή επίδραση σε σχέση με το επίπεδο της σειράς.

2. Εξομαλυμένο επίπεδο:

$$S_t = S_{t-1} + \alpha \cdot \left( \frac{e_t}{I_{t-p}} \right)$$

Το επίπεδο,  $S_t$ , ενημερώνεται με βάση το σφάλμα πρόβλεψης,  $e_t$  και τον εποχιακό δείκτη, προσφέροντας μια τρέχουσα εκτίμηση που λαμβάνει υπόψη τις εποχιακές προσαρμογές.

3. Εξίσωση πρόβλεψης:

$$F_{t+m} = S_t \cdot I_{t-p+m}$$

Οι μελλοντικές τιμές προβλέπονται πολλαπλασιάζοντας το εξομαλυμένο επίπεδο με τον εποχιακό δείκτη, διασφαλίζοντας ότι οι προβλέψεις αντικατοπτρίζουν τόσο το επίπεδο των τιμών όσο και τις εποχιακές διακυμάνσεις.

Τα παραπάνω ωστόσο μπορούν να εφαρμοστούν και σε όλα τα μοντέλα εκθετικής εξομάλυνσης που περιγράφηκαν προηγουμένως.

Η μέθοδος Holt-Winters [30] επεκτείνει τη λογική της μεθόδου Winters συμπεριλαμβάνοντας μια συνιστώσα γραμμικής τάσης. Αυτή η προσθήκη επιτρέπει στο μοντέλο να καταγράφει όχι μόνο τις εποχιακές διακυμάνσεις αλλά και τις τάσεις στα δεδομένα, παρέχοντας ένα πιο ολοκληρωμένο εργαλείο πρόβλεψης. Συνεπώς, προκύπτουν οι παρακάτω εξισώσεις:

$$S_t = S_{t-1} + T_{t-1} + \alpha \cdot \left( \frac{e_t}{I_{t-p}} \right)$$












$$T_t = T_{t-1} + \beta \cdot \left( \frac{e_t}{I_{t-p}} \right)$$

$$I_t = I_{t-p} + \gamma \cdot \left( \frac{e_t}{S_{t-1} + T_{t-1}} \right)$$

$$F_{t+m} = (S_t + m \cdot T_t) \cdot I_{t-p+m}$$

Η εξίσωση πρόβλεψης συνθέτει το εξομαλυνόμενο επίπεδο, την τάση και τον εποχιακό δείκτη για να προσφέρει πιο ακριβείς προβλέψεις.

Συμπερασματικά, ενώ τα προσθετικά μοντέλα βρίσκουν την εφαρμογή τους σε σενάρια με συνεχείς εποχιακές διακυμάνσεις, η πολλαπλασιαστική προσέγγιση παρέχει μια μεγαλύτερη ευελιξία που μπορεί να χρησιμεύσει σε κάποιες περιπτώσεις. Με τις κατάλληλες επιλογές, οι μέθοδοι Winters προσφέρουν βελτιωμένες προβλέψεις σε σχέση με τις απλές πρακτικές εκθετικής εξομάλυνσης, κάτι που τις καθιστά απαραίτητες για την λήψη αποφάσεων σε ένα φάσμα τομέων, από τα χρηματοοικονομικά έως τις περιβαλλοντικές μελέτες.

	Nonseasonal	Additive Seasonal	Multiplicative Seasonal
Constant Level	(SIMPLE) — NN	 NA	 NM
Linear Trend	(HOLT)  LN	 LA	(WINTERS)  LM
Damped Trend (0.95)	 DN	 DA	 DM
Exponential Trend (1.05)	 EN	 EA	 EM

Εικόνα 2.3: Εκθετική Εξομάλυνση

### 2.3.4 Μοντέλα ARIMA

Τα μοντέλα ARIMA αποτελούν μία από τις πιο συχνά χρησιμοποιούμενες παραδοσιακές τεχνικές στον κόσμο των προβλέψεων χρονοσειρών. Αυτά τα μοντέλα που αναπτύχθηκαν από τους Box και Jenkins το 1970 εξαρτώνται πλήρως από τις ιστορικές τιμές της ίδιας της σειράς, καθιστώντας τα ιδιαίτερα χρήσιμα σε σενάρια όπου οι εξωτερικοί παράγοντες είναι άγνωστοι ή δύσκολο να ποσοτικοποιηθούν [31]. Στον πυρήνα του, ένα μοντέλο ARIMA χαρακτηρίζεται από τρία κύρια στοιχεία: τους αυτοπαλινδρομικούς όρους (AR), την διαφόριση (I) για την επίτευξη σταθερότητας και τον κινητό μέσο όρο (MA) [32]. Αυτή η τριάδα συμβολίζεται ως ARIMA(p, d, q), όπου:

- Το  $p$  αντιπροσωπεύει την τάξη των όρων AR,
- Το  $d$  υποδεικνύει τον βαθμό διαφόρισης που απαιτείται για να καταστεί η χρονοσειρά σταθερή,
- Το  $q$  αναφέρεται στην τάξη των όρων MA.

1. **Αυτοπαλινδρόμηση AR:** Το τμήμα AR καταγράφει την επίδραση των προηγούμενων τιμών της χρονοσειράς στην τρέχουσα πρόβλεψη, που διατυπώνεται ως ένας γραμμικός



συνδυασμός προηγούμενων παρατηρήσεων. Το μοντέλο AR τάξης  $p$  (AR( $p$ )) εκφράζεται ως:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \epsilon_t$$

όπου  $c$  είναι μια σταθερά,  $\phi_1, \phi_2, \dots, \phi_p$  είναι οι συντελεστές και  $\epsilon_t$  είναι λευκός θόρυβος.

2. **Διαφόρηση I:** Για να επιτευχθεί σταθερότητα, η σειρά διαφορίζεται  $d$  φορές, κάτι που συμβολίζεται ως  $I(d)$ . Η διαφόρηση περιλαμβάνει την αφαίρεση της τρέχουσας παρατήρησης από την προηγούμενη:

$$y'_t = y_t - y_{t-1}$$

Αυτό το βήμα βοηθά στη σταθεροποίηση του μέσου όρου της σειράς, αφαιρώντας τις τάσεις και τις εποχιακές συνιστώσες.

3. **Κινητός Μέσος (MA):** Το τμήμα MA μοντελοποιεί τον όρο σφάλματος ως γραμμικό συνδυασμό όλων των όρων σφάλματος σε διάφορες χρονικές στιγμές στο παρελθόν. Το μοντέλο MA της τάξης  $q$  (MA( $q$ )) δίνεται από:

$$y_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q}$$

όπου  $\theta_1, \theta_2, \dots, \theta_q$  είναι οι συντελεστές των όρων σφάλματος.

Η ενσωμάτωση αυτών των στοιχείων δίνει το μοντέλο ARIMA( $p, d, q$ ), το οποίο μπορεί να αναπαρασταθεί ως:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Εδώ, το  $y'_t$  υποδηλώνει τη διαφορισμένη σειρά, καταγράφοντας τόσο τη δυναμική της αυτοπαλιδρομικής μεθόδου όσο και του κινούμενου μέσου πάνω στη διαφορισμένη χρονοσειρά.

Γενικά μιλώντας, για να είναι αποτελεσματικό ένα μοντέλο ARIMA, η χρονοσειρά θα πρέπει ιδανικά να παρουσιάζει σταθερότητα, όπου οι στατιστικές της ιδιότητες όπως ο μέσος όρος και η διακύμανση δεν αλλάζουν με την πάροδο του χρόνου. Η διαφόρηση (I) παίζει καθοριστικό ρόλο στην επίτευξη αυτού του στόχου, αφαιρώντας τις τάσεις και τις εποχιακές συνιστώσες. Παράλληλα, η επιλογή των κατάλληλων τιμών για τα  $p$ ,  $d$  και  $q$  είναι κρίσιμη. Μετρικές όπως το Akaike Information Criterion (AIC) μπορούν να αυτοματοποιήσουν αυτή τη διαδικασία, βοηθώντας στον εντοπισμό του μοντέλου που ταιριάζει καλύτερα στα δεδομένα.

Συνοψίζοντας, τα μοντέλα ARIMA προσφέρουν ένα ισχυρό και ευέλικτο εργαλείο για την ανάλυση χρονοσειρών καθώς και για την πρόβλεψη μελλοντικών τιμών τους, ενώ υπερέρχουν στη βραχυπρόθεσμη πρόβλεψη. Αξιοποιώντας ιστορικά δεδομένα και επιλέγοντας προσεκτικά τις παραμέτρους του μοντέλου, η ARIMA επιτρέπει ακριβείς και διορατικές προβλέψεις σε διάφορους τομείς, από την οικονομία έως την περιβαλλοντική επιστήμη [33, 34, 35].

### 2.3.5 Μέθοδος Θ

Μία από τις πλέον προηγμένες παραδοσιακές στατιστικές μεθόδους που χρησιμοποιούνται στον τομέα των προβλέψεων χρονοσειρών είναι η μέθοδος Theta, η οποία προτάθηκε από

τους Ασημακόπουλο και Νικολόπουλο το 2000 [36]. Η μέθοδος αυτή βασίζεται στη μεταβολή των τοπικών καμπυλοτήτων της χρονοσειράς, παράγοντας νέες χρονοσειρές, τις λεγόμενες γραμμές Theta, των οποίων οι προβλεπόμενες τιμές συνδυάζονται εν τέλει για να προκύψει η συνολική τελική τιμή.

Εμβαιθύνοντας στη μεθοδολογία του συγκεκριμένου μοντέλου, το πρώτο βήμα που ακολουθείται είναι ο έλεγχος εποχικότητας. Αρχικά η χρονοσειρά ελέγχεται ως προς την ύπαρξη ευδιάκριτης στατιστικά εποχικής συμπεριφοράς, και αν αποδειχθεί κάτι τέτοιο αποεποχικοποιείται μέσω της μεθόδου πολλαπλασιαστικής αποσύνθεσης που περιγράφηκε παραπάνω [16].

Έπειτα, στην καρδιά της συγκεκριμένης μεθοδολογίας βρίσκεται η αποσύνθεση της χρονοσειράς σε διαφορετικές γραμμές Theta. Αυτό πραγματοποιείται με χρήση των παρακάτω εξισώσεων:

$$Y_{t,\theta} = \theta \cdot Y_t''$$

όπου

$$Y_t'' = Y_t - 2 \cdot Y_{t-1} + Y_{t-2}$$

Η κλασική μέθοδος Theta βασίζεται στην αποσύνθεση σε δύο γραμμές Theta, την ευθεία γραμμικής παλινδρόμησης ( $\theta = 0$ ) και τη γραμμή Theta με παράμετρο  $\theta = 2$ . Για λόγους απλότητας, η μετέπειτα παρουσίαση της μεθόδου θα υποθέτει πως πρόκειται για αυτή την κλασική μορφή της εν λόγω μεθοδολογίας.

Έπειτα, η γραμμή Theta με παράμετρο  $\theta = 0$ , που αναπαριστά την ευθεία γραμμικής παλινδρόμησης, προεκτείνεται με τον συνηθισμένο τρόπο, ενώ η δεύτερη γραμμή προεκτείνεται μέσω της απλής εκθετικής εξομάλυνσης που παρουσιάστηκε σε προηγούμενη ενότητα. Ακολούθως, οι δύο προβλέψεις συντίθενται, όπως γίνεται αντιληπτό από την παρακάτω εξίσωση:

$$Y_t = \frac{1}{2}(Y_{t,\theta=0} + Y_{t,\theta=2})$$

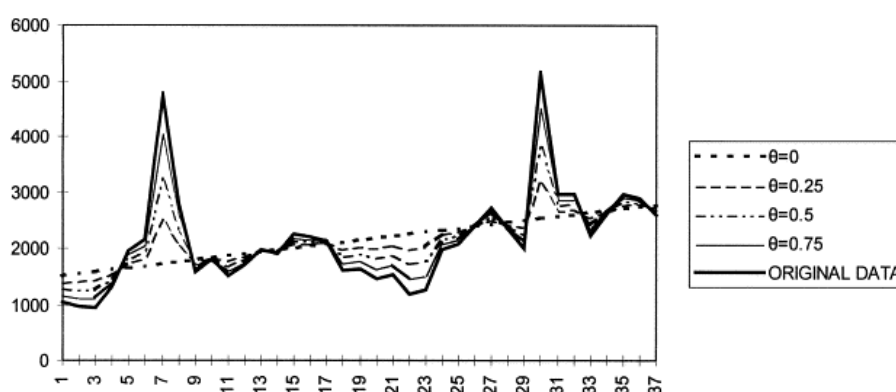
Τέλος, οι τελικές προβλέψεις εποχικοποιούνται, χρησιμοποιώντας τους δείκτες εποχικότητας που υπολογίστηκαν στα πρώτα βήματα της μεθοδολογίας πολλαπλασιαστικά.

Ανάλογα με την τιμή του  $\theta$ , παρατηρούνται διαφορετικές συμπεριφορές των γραμμών Theta:

- Για  $\theta = 0$ , η σειρά ευθυγραμμίζεται με τη γραμμή γραμμικής παλινδρόμησης, δίνοντας έμφαση στις μακροπρόθεσμες τάσεις.
- Για  $\theta = -1$ , η σειρά που προκύπτει είναι συμμετρική με την αρχική γραμμή παλινδρόμησης, εμφανίζοντας μια ανεστραμμένη καμπυλότητα.
- Για  $\theta > 1$ , η καμπυλότητα ενισχύεται, δίνοντας μεγαλύτερη έμφαση στις βραχυπρόθεσμες διακυμάνσεις.

Όσον αφορά την κλασική μέθοδο Theta, η μακροπρόθεσμη τάση εξασφαλίζεται από την προέκταση της γραμμής  $\theta = 0$ . Ταυτόχρονα, η ύπαρξη και της γραμμής  $\theta = 2$  λειτουργεί σαν αντίβαρο στην χρησιμοποίηση μόνο της απλής γραμμικής παλινδρόμησης και εξασφαλίζει την αξιοποίηση και της βραχυπρόθεσμης πληροφορίας. Παράλληλα, αξίζει να σημειωθεί πως η δύναμη της μεθόδου έγκειται στην ικανότητά της να επεκτείνεται πέρα από το κλασικό

μοντέλο δύο γραμμών ενσωματώνοντας πρόσθετες γραμμές Theta για διαφορετικές τιμές του  $\theta$ , παρουσιάζοντας έτσι την ικανότητα να χειρίζεται ένα ευρύτερο φάσμα χρονοσειρών [37].



Εικόνα 2.4: Μέθοδος Theta

### 2.3.6 Απλή και Πολλαπλή Παλινδρόμηση

Σε αντίθεση με τις μεθόδους που προαναφέρθηκαν που εστιάζουν αποκλειστικά στις προηγούμενες παρατηρήσεις της χρονοσειράς για την πρόβλεψη μελλοντικών τιμών, η γραμμική παλινδρόμηση και η επέκτασή της, η πολλαπλή γραμμική παλινδρόμηση, εισάγουν την έννοια της χρήσης ανεξάρτητων μεταβλητών για την πρόβλεψη. Αυτές οι ανεξάρτητες μεταβλητές, μπορεί να είναι οποιοδήποτε παράγοντες πιστεύεται ότι έχουν αντίκτυπο στη μεταβλητή στόχο. Για παράδειγμα, στην πρόβλεψη των τιμών των μετοχών, θα μπορούσε κανείς να εξετάσει δείκτες όπως τα επιτόκια ή τα ποσοστά πληθωρισμού. Η συμπερίληψη αυτών των μεταβλητών προσφέρει ένα σημαντικό πλεονέκτημα: παρέχει έναν τρόπο να ληφθούν υπόψη οι εξωτερικές επιρροές που θα μπορούσαν να επηρεάσουν την προβλεπόμενη μεταβλητή, οδηγώντας σε πιθανώς πιο ακριβείς αλλά και εξηγήσιμες προβλέψεις.

Τα μοντέλα γραμμικής παλινδρόμησης λειτουργούν με την προϋπόθεση ότι υπάρχει μια γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής (η χρονική σειρά ενδιαφέροντος) και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Η απλή γραμμική παλινδρόμηση εξετάζει μια ανεξάρτητη μεταβλητή, προσφέροντας μια απλή ανάλυση του τρόπου με τον οποίο οι αλλαγές σε αυτόν τον προγνωστικό παράγοντα επηρεάζουν τη μεταβλητή στόχο. Ωστόσο, τα περισσότερα φαινόμενα του πραγματικού κόσμου επηρεάζονται από πολλούς παράγοντες ταυτόχρονα. Τα πολλαπλά μοντέλα γραμμικής παλινδρόμησης αντιμετωπίζουν αυτή την πολυπλοκότητα ενσωματώνοντας πολλές ανεξάρτητες μεταβλητές, επιτρέποντας μια ολοκληρωμένη ανάλυση του συλλογικού τους αντίκτυπου στην εξαρτημένη μεταβλητή.

Πιο συγκεκριμένα, η απλή γραμμική παλινδρόμηση είναι μια θεμελιώδης στατιστική μέθοδος που χρησιμοποιείται για τη μοντελοποίηση της σχέσης μεταξύ δύο μεταβλητών: μιας εξαρτημένης μεταβλητής  $y$  και μιας ανεξάρτητης μεταβλητής  $x$ . Ο στόχος είναι να βρεθεί η καλύτερη προσαρμοσμένη γραμμή μέσω των σημείων δεδομένων που ελαχιστοποιεί τη διαφορά (σφάλμα) μεταξύ των παρατηρούμενων τιμών και των τιμών που προβλέπονται από το γραμμικό μοντέλο.

Το μοντέλο περιγράφεται από την εξίσωση:

$$y = \beta_0 + \beta_1 x + \epsilon$$

Το  $\epsilon$  υποδηλώνει τον όρο σφάλματος, λαμβάνοντας υπόψη την απόκλιση των παρατηρούμενων τιμών από τη γραμμή. Οι βέλτιστες τιμές των  $\beta_0$  και  $\beta_1$  καθορίζονται χρησιμοποιώντας το κριτήριο των ελαχίστων τετραγώνων, το οποίο στοχεύει στην ελαχιστοποίηση του αθροίσματος των τετραγωνικών διαφορών μεταξύ των παρατηρούμενων τιμών και των τιμών που προβλέπονται από το μοντέλο:

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

Μέσω αυτής της μεθόδου, οι συντελεστές  $\beta_0$  και  $\beta_1$  μπορούν να υπολογιστούν ως:

$$\beta_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

όπου τα  $\bar{x}$  και  $\bar{y}$  είναι οι μέσοι όροι του δείγματος των  $x$  και  $y$ , αντίστοιχα.

Από την άλλη, η πολλαπλή γραμμική παλινδρόμηση επεκτείνει το μοντέλο απλής γραμμικής παλινδρόμησης για να ενσωματώσει πολλαπλές ανεξάρτητες μεταβλητές. Αυτό επιτρέπει μια πιο λεπτομερή ανάλυση του τρόπου με τον οποίο διάφοροι παράγοντες επηρεάζουν ταυτόχρονα την εξαρτημένη μεταβλητή. Το μοντέλο πολλαπλής γραμμικής παλινδρόμησης αναπαρίσταται ως:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon$$

Εδώ,  $y$  είναι η εξαρτημένη μεταβλητή,  $x_1, x_2, \dots, x_k$  είναι οι ανεξάρτητες μεταβλητές, ενώ  $\beta_1, \beta_2, \dots, \beta_k$  είναι οι συντελεστές των ανεξάρτητων μεταβλητών και  $\epsilon$  είναι ο όρος σφάλματος. Οι συντελεστές και πάλι υπολογίζονται με τη μέθοδο των ελαχίστων τετραγώνων, με στόχο την ελαχιστοποίηση του αθροίσματος των τετραγωνικών σφαλμάτων:

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

Αυτό περιλαμβάνει την επίλυση ενός συνόλου γραμμικών εξισώσεων για την εύρεση των τιμών των  $\beta_0, \beta_1, \dots, \beta_k$  που ελαχιστοποιούν τη διαφορά μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών.

Συνολικά, η γραμμική παλινδρόμηση και η πολλαπλή γραμμική παλινδρόμηση είναι απαραίτητα εργαλεία στην πρόβλεψη χρονοσειρών, προσφέροντας μια συστηματική προσέγγιση στη μοντελοποίηση και την πρόβλεψη δεδομένων. Κατανοώντας και εφαρμόζοντας αυτές τις τεχνικές, οι αναλυτές μπορούν να αποκαλύψουν σημαντικές σχέσεις μεταξύ μεταβλητών και να κάνουν ακριβείς προβλέψεις που βοηθούν στη λήψη αποφάσεων σε διάφορους τομείς.

## 2.4 Μέθοδοι Αξιολόγησης Μοντέλων Πρόβλεψης

Η ακρίβεια της πρόβλεψης είναι ζωτικής σημασίας στην ανάλυση χρονοσειρών, καθορίζοντας την αποτελεσματικότητα των μοντέλων πρόβλεψης σε διάφορες εφαρμογές. Δεδομένης της πληθώρας των διαθέσιμων μεθόδων πρόβλεψης, από την απλή εκθετική εξομάλυνση έως τα πολύπλοκα μοντέλα ARIMA και Theta, η δυνατότητα ποσοτικής αξιολόγησης της απόδοσής τους είναι ζωτικής σημασίας. Αυτό απαιτεί ένα ισχυρό πλαίσιο μετρικών σφάλματος, καθεμία από τις οποίες προσφέρει μοναδικές πληροφορίες σχετικά με τις προγνωστικές δυνατότητες ενός μοντέλου. Παρακάτω θα παρουσιαστούν οι κύριες μετρικές σφάλματος που χρησιμοποιούνται για την αξιολόγηση μοντέλων πρόβλεψης χρονοσειρών.

1. **Μέσο Σφάλμα (ME):** Το Μέσο Σφάλμα (ME) προσφέρει ένα απλό μέτρο της μεροληψίας ενός μοντέλου, υπολογίζοντας τη μέση απόκλιση των προβλέψεων από τις πραγματικές παρατηρήσεις. Μαθηματικά εκφράζεται ως:

$$ME = \frac{1}{n} \sum_{t=1}^n (Y_t - F_t)$$

όπου  $Y_t$  αντιπροσωπεύει την πραγματική τιμή τη στιγμή  $t$ ,  $F_t$  είναι η προβλεπόμενη τιμή και  $n$  είναι ο αριθμός των παρατηρήσεων. Το ME βοηθά στον προσδιορισμό του εάν ένα μοντέλο υπερεκτιμά ή υποτιμά με συνέπεια τις πραγματικές τιμές. Ένα ME κοντά στο μηδέν υποδηλώνει ελάχιστη μεροληψία, ενώ οι θετικές ή αρνητικές τιμές υποδεικνύουν μια τάση προς υπερπρόβλεψη ή υποπρόβλεψη, αντίστοιχα.

2. **Μέσο Απόλυτο Σφάλμα (MAE):** Το μέσο απόλυτο σφάλμα (MAE) ποσοτικοποιεί το μέσο μέγεθος των σφαλμάτων πρόβλεψης, αγνοώντας την κατεύθυνσή τους. Ορίζεται ως:

$$MAE = \frac{1}{n} \sum_{t=1}^n |Y_t - F_t|$$

Το MAE είναι ιδιαίτερα χρήσιμο για την κατανόηση της ακρίβειας του μοντέλου χωρίς να λαμβάνεται υπόψη η κατεύθυνση του σφάλματος, παρέχοντας μια σαφέστερη εικόνα του συνολικού μεγέθους του σφάλματος.

3. **Μέσο Τετράγωνο Σφάλμα (MSE) και Ριζικό Μέσο Τετράγωνο Σφάλμα (RMSE):** Το MSE και το RMSE είναι ζωτικής σημασίας για την αξιολόγηση της διακύμανσης των σφαλμάτων πρόβλεψης, με το RMSE να παρέχει ένα μέτρο σε κατάλληλη κλίμακα λαμβάνοντας την τετραγωνική ρίζα του MSE. Αυτά υπολογίζονται ως εξής:

$$MSE = \frac{1}{n} \sum_{t=1}^n (Y_t - F_t)^2$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{t=1}^n (Y_t - F_t)^2}$$

Αυτές οι μετρήσεις είναι ευαίσθητες σε μεγάλα σφάλματα, καθιστώντας τις κατάλληλες για μοντέλα όπου η ελαχιστοποίηση μεγάλων αποκλίσεων είναι ζωτικής σημασίας. Το

RMSE, που βρίσκεται στην ίδια κλίμακα με τις παρατηρήσεις, είναι ιδιαίτερα βολικό για άμεση σύγκριση με τον μέσο όρο ή την τυπική απόκλιση των παρατηρούμενων τιμών.

4. **Μέσο Απόλυτο Ποσοστό Σφάλματος (MAPE) και Συμμετρικό Μέσο Απόλυτο Ποσοστό Σφάλματος (sMAPE):** Το MAPE εκφράζει τα σφάλματα πρόβλεψης ως ποσοστό των παρατηρούμενων τιμών, παρέχοντας ένα διαισθητικό μέτρο της ακρίβειας του μοντέλου σε σχέση με το μέγεθος των δεδομένων. Ορίζεται ως:

$$MAPE = \frac{100}{n} \sum_{t=1}^n \left| \frac{Y_t - F_t}{Y_t} \right|$$

Ωστόσο, το MAPE μπορεί να είναι παραπλανητικό για τιμές κοντά στο μηδέν. Το sMAPE αντιμετωπίζει αυτόν και άλλους περιορισμούς τροποποιώντας τον τύπο σε:

$$sMAPE = \frac{100}{n} \sum_{t=1}^n \frac{2|Y_t - F_t|}{|Y_t| + |F_t|}$$

Η συμμετρική αντιμετώπιση υπερεκτιμήσεων και υποεκτιμήσεων του sMAPE, μαζί με το οριοθετημένο εύρος του, προσφέρει μια ισορροπημένη άποψη της απόδοσης του μοντέλου.

5. **Προηγμένες Μετρήσεις Σχετικού Σφάλματος: MASE και RMSSE:** Το μέσο απόλυτο σφάλμα κλίμακας (MASE) και το ριζικό μέσο τετράγωνο κλίμακας σφάλματος (RMSSE) κανονικοποιούν το σφάλμα πρόβλεψης συγκρίνοντάς το με ένα απλό μοντέλο αναφοράς. Το MASE ορίζεται ως:

$$MASE = \frac{\sum_{t=1}^n |Y_t - F_t|}{\frac{n}{n-1} \sum_{t=2}^n |Y_t - Y_{t-1}|}$$

Το RMSSE, ομοίως, κλιμακώνει το RMSE έναντι της απόδοσης της αφελούς μεθόδου, παρέχοντας ένα μέτρο σφάλματος που είναι ανεξάρτητο από την κλίμακα δεδομένων και υπογραμμίζει τη σχετική βελτίωση του μοντέλου από μία αφελή πρόβλεψη.

Συμπερασματικά, οι μετρικές σφαλμάτων είναι απαραίτητες για την αξιολόγηση και σύγκριση μοντέλων πρόβλεψης χρονοσειρών. Εφαρμόζοντας σχολαστικά αυτές τις μετρικές, οι αναλυτές μπορούν να μετρήσουν την ακρίβεια, την προκατάληψη και την αξιοπιστία ενός μοντέλου, προσαρμόζοντας τις στρατηγικές πρόβλεψής τους ώστε να ανταποκρίνονται στις συγκεκριμένες απαιτήσεις των εργασιών τους. Καθώς το πεδίο της πρόβλεψης χρονοσειρών συνεχίζει να εξελίσσεται, η συνετή εφαρμογή αυτών των μετρήσεων παραμένει μια κρίσιμη πτυχή της επιλογής και της αξιολόγησης μοντέλων, καθοδηγώντας τους αναλυτές σε πιο ακριβή και αξιόπιστα αποτελέσματα πρόβλεψης.

## Κεφάλαιο **3**

# Μηχανική Μάθηση και Προβλέψεις Χρονοσειρών

---

### 3.1 Εισαγωγή στη Μηχανική Μάθηση

Η Μηχανική Μάθηση (ML) βρίσκεται στην πρώτη γραμμή των εξελίξεων στον τομέα της τεχνητής νοημοσύνης (AI), παρέχοντας ένα πλαίσιο έτσι ώστε τα συστήματα να μαθαίνουν και να λαμβάνουν αποφάσεις με βάση παρελθοντικά δεδομένα. Σε αντίθεση με τις παραδοσιακές υπολογιστικές προσεγγίσεις που απαιτούν ντετερμινιστικό προγραμματισμό για να λάβουν μία απόφαση, οι αλγόριθμοι μηχανικής μάθησης έχουν σχεδιαστεί έτσι ώστε να μαθαίνουν από μοτίβα δεδομένων και να κάνουν προβλέψεις ή αποφάσεις με βάση τα δεδομένα πάνω στα οποία έχουν εκπαιδευτεί. Η μηχανική μάθηση είναι μια συγχώνευση εννοιών από τους τομείς των πιθανοτήτων, της στατιστικής, της άλγεβρας και της θεωρίας βελτιστοποίησης. Στον τομέα της μηχανικής μάθησης διακρίνονται τρεις διαφορετικοί χώροι [38]:

1. **Εποπτευόμενη μάθηση:** Στην εποπτευόμενη μάθηση (Supervised Learning), οι αλγόριθμοι εκπαιδεύονται με βάση ένα σύνολο δεδομένων που περιλαμβάνει τόσο τα δεδομένα εισόδου όσο και τις επιθυμητές εξόδους. Το μοντέλο μαθαίνει να χαρτογραφεί τις εισόδους στις εξόδους, με στόχο να προβλέψει τις εξόδους για νέα, άγνωστα δεδομένα. Τυπικά παραδείγματα περιλαμβάνουν την ταξινόμηση και την παλινδρόμηση.
2. **Μη εποπτευόμενη μάθηση:** Στη μη εποπτευόμενη μάθηση (Unsupervised Learning), οι αλγόριθμοι εκπαιδεύονται αποκλειστικά με βάση δεδομένα εισόδου χωρίς να παρέχονται επιθυμητές εξόδοι. Ο στόχος είναι η εύρεση κρυφών δομών ή μοτίβων στα δεδομένα. Παραδείγματα περιλαμβάνουν την ομαδοποίηση και τη μείωση διαστάσεων.
3. **Ενισχυτική μάθηση:** Στην ενισχυτική μάθηση (Reinforcement Learning), οι αλγόριθμοι μαθαίνουν μέσω αλληλεπιδράσεων με το περιβάλλον τους, λαμβάνοντας ανταμοιβές ή τιμωρίες ως ανταπόκριση στις ενέργειές τους. Ο στόχος είναι η μεγιστοποίηση της συνολικής ανταμοιβής με την πάροδο του χρόνου. Αυτή η μέθοδος χρησιμοποιείται συχνά σε εφαρμογές όπου οι αποφάσεις πρέπει να λαμβάνονται σε αλληλουχίες, όπως στη ρομποτική και στα παίγνια.

### 3.1.1 Εποπτευόμενη Μάθηση

Η εποπτευόμενη μάθηση, ο ακρογωνιαίος λίθος της μηχανικής μάθησης (ML), περιλαμβάνει την εκπαίδευση ενός υπολογιστικού συστήματος για να κάνει προβλέψεις ή να παίρνει αποφάσεις χρησιμοποιώντας δεδομένα στα οποία η μεταβλητή στόχος είναι γνωστή (labelled data). Σύμφωνα με αυτή την προσέγγιση, ο αλγόριθμος μαθαίνει την αντιστοίχιση μεταξύ των μεταβλητών εισόδου (ανεξάρτητες μεταβλητές) και των μεταβλητών εξόδου (εξαρτημένες μεταβλητές) χρησιμοποιώντας παρελθοντικά δεδομένα. Ο απώτερος στόχος της εποπτευόμενης μάθησης είναι η κατασκευή ενός μοντέλου που μπορεί να κάνει ακριβείς προβλέψεις σε νέα, μη παρουσιασμένα προηγουμένως στον αλγόριθμο δεδομένα με βάση τη γνώση που αποκτήθηκε κατά τη φάση της εκπαίδευσής του [39].

Στον πυρήνα της, η εποπτευόμενη μάθηση βασίζεται στη θεωρία της προσέγγισης συναρτήσεων [40]. Οι συγκεκριμένοι αλγόριθμοι μπορούν να αντιμετωπιστούν ως προσπάθεια εύρεσης μιας συνάρτησης που αντιστοιχίζει καλύτερα τις μεταβλητές εισόδου σε μεταβλητές εξόδου. Αυτή η συνάρτηση μαθαίνεται μέσω της έκθεσης σε ένα σύνολο δεδομένων που περιλαμβάνει παραδείγματα ζευγών εισόδου-εξόδου. Ο αλγόριθμος μάθησης επαναλαμβάνεται μέσω αυτού του συνόλου δεδομένων, προσαρμόζοντας τις παραμέτρους του για να ελαχιστοποιήσει την απόκλιση μεταξύ των προβλέψεών του και των πραγματικών αποτελεσμάτων.

Πιο επίσημα, έστω ένα σύνολο δεδομένων  $D$  που αποτελείται από  $n$  δείγματα. Κάθε δείγμα είναι ένα ζεύγος  $(\mathbf{x}_i, y_i)$ , όπου το  $\mathbf{x}_i$  είναι ένα διάνυσμα χαρακτηριστικών εισόδου από το χώρο εισόδου  $\mathcal{X}$  και  $y_i$  είναι η αντίστοιχη έξοδος ή ετικέτα από το χώρο εξόδου  $\mathcal{Y}$ . Ο στόχος της εποπτευόμενης μάθησης είναι να μάθει μια αντιστοίχιση  $f : \mathcal{X} \rightarrow \mathcal{Y}$  που προβλέπει την έξοδο  $y$  για μια νέα είσοδο  $\mathbf{x}$ .

Η διαδικασία εκμάθησης βασίζεται σε μια συνάρτηση απώλειας  $L(y, f(\mathbf{x}))$  για τη μέτρηση της διαφοράς μεταξύ της προβλεπόμενης τιμής  $f(\mathbf{x})$  και της πραγματικής τιμής  $y$ . Η επιλογή της συνάρτησης απώλειας εξαρτάται από τη φύση της επιθυμητής πρόβλεψης [41]. Για προβλήματα παλινδρόμησης, όπου η έξοδος είναι μια συνεχής μεταβλητή, χρησιμοποιείται συνήθως το μέσο τετραγωνικό σφάλμα (MSE):

$$L_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - f(\mathbf{x}_i))^2$$

Σε προβλήματα ταξινόμησης, όπου η έξοδος είναι μια κατηγορική μεταβλητή, η απώλεια διασταυρούμενης εντροπίας (cross-entropy loss) χρησιμοποιείται συχνά [42]:

$$L_{\text{cross-entropy}} = -\frac{1}{n} \sum_{i=1}^n [y_i \log(f(\mathbf{x}_i)) + (1 - y_i) \log(1 - f(\mathbf{x}_i))]$$

Ο αλγόριθμος εκμάθησης επιδιώκει να βρει τη συνάρτηση  $f$  που ελαχιστοποιεί τη συνάρτηση απώλειας σε όλο το σύνολο δεδομένων εκπαίδευσης. Αυτό συνήθως επιτυγχάνεται μέσω τεχνικών βελτιστοποίησης, όπως η gradient descent. Οι παράμετροι του μοντέλου προσαρμόζονται επαναληπτικά προς την κατεύθυνση που μειώνει ταχύτερα την απώλεια.

Μια ουσιαστική πτυχή της εποπτευόμενης μάθησης είναι η αντιστάθμιση μεταξύ της πολυπλοκότητας του μοντέλου και της ικανότητας γενίκευσης από τα δεδομένα της εκπαίδευσης



σε δεδομένα που δεν ανήκουν στο σύνολο εκπαίδευσης. Τα εξαιρετικά πολύπλοκα μοντέλα μπορούν να μάθουν εξαιρετικά τα δεδομένα εκπαίδευσης, οδηγώντας δυνητικά σε υπερπροσαρμογή, όπου το μοντέλο μαθαίνει θόρυβο ή τυχαίες διακυμάνσεις στα δεδομένα εκπαίδευσης αντί για την πραγματικά υποκείμενη κατανομή [43]. Από την άλλη πλευρά, τα υπερβολικά απλοϊκά μοντέλα ενδέχεται να μην αποτυπώνουν επαρκώς τη δομή των δεδομένων, με αποτέλεσμα την υποπροσαρμογή [43].

Συχνά χρησιμοποιούνται τεχνικές κανονικοποίησης για τον έλεγχο της πολυπλοκότητας του μοντέλου. Αυτές οι τεχνικές προσθέτουν έναν όρο ποινής στη συνάρτηση απώλειας, αποθαρρύνοντας το μοντέλο να γίνει πολύ περίπλοκο και επομένως το καθιστούν πιο ικανό να γενικεύει καλά σε νέα πρωτοεμφανιζόμενα στον αλγόριθμο δεδομένα. Για παράδειγμα, η κανονικοποίηση L2 προσθέτει έναν όρο ανάλογο του τετραγώνου του μεγέθους των συντελεστών, τιμωρώντας αποτελεσματικά τους μεγάλους συντελεστές:

$$L_{\text{reg}} = L(y, f(\mathbf{x})) + \lambda \|\mathbf{w}\|^2$$

όπου το  $\lambda$  είναι μια παράμετρος κανονικοποίησης που ελέγχει την ισορροπία μεταξύ της προσαρμογής των δεδομένων εκπαίδευσης και της διατήρησης των παραμέτρων του μοντέλου σε χαμηλό επίπεδο.

Ως προς την επιλογή του αλγορίθμου εποπτευόμενης μάθησης, το θεώρημα «No Free Lunch» [44] υποστηρίζει ότι κανένας αλγόριθμος δεν είναι η καλύτερη επιλογή για κάθε πρόβλημα. Η απόδοση ενός εποπτευόμενου αλγορίθμου μάθησης εξαρτάται από την αντιστοίχιση μεταξύ της δομής του προβλήματος και των ιδιομορφιών του κάθε αλγορίθμου. Η κατανόηση των ιδιοτήτων των διαφορετικών μοντέλων και των χαρακτηριστικών των δεδομένων είναι ζωτικής σημασίας για την επιλογή του κατάλληλου αλγορίθμου.

Συνοψίζοντας, η εποπτευόμενη μάθηση παρέχει ένα ισχυρό πλαίσιο για προγνωστική μοντελοποίηση, με ευρείες εφαρμογές που κυμαίνονται από την αναγνώριση εικόνας και την επεξεργασία φυσικής γλώσσας έως την παραγωγή προβλέψεων ακριβείας σε χρονοσειρές. Η αποτελεσματικότητά του εξαρτάται από τον προσεκτικό σχεδιασμό του αλγορίθμου μάθησης, συμπεριλαμβανομένης της επιλογής μοντέλου, συνάρτησης απώλειας, τεχνικής βελτιστοποίησης και στρατηγικής κανονικοποίησης. Εξισορροπώντας την πολυπλοκότητα του μοντέλου με την ικανότητά του να γενικεύει, η εποπτευόμενη μάθηση στοχεύει στη δημιουργία προγνωστικών μοντέλων που αποδίδουν καλά όχι μόνο στα δεδομένα στα οποία εκπαιδεύτηκαν αλλά και σε νέα δεδομένα που ο αλγόριθμος συναντά πρώτη φορά.

### 3.1.2 Μη Εποπτευόμενη Μάθηση

Η μάθηση χωρίς επίβλεψη εστιάζει στην εξαγωγή συμπερασμάτων από σύνολα δεδομένων χωρίς μία γνωστή μεταβλητή-στόχο. Σε αντίθεση με την εποπτευόμενη μάθηση όπου τα δεδομένα εκπαίδευσης συνοδεύονται από ετικέτες (labels) που υποδεικνύουν το σωστό αποτέλεσμα, οι αλγόριθμοι μάθησης χωρίς επίβλεψη λειτουργούν με δεδομένα που δεν έχουν τέτοια καθοδήγηση. Ο πρωταρχικός στόχος της μάθησης χωρίς επίβλεψη είναι να αποκαλύψει την υποκείμενη δομή ή κατανομή στα δεδομένα έτσι ώστε να προκύψει γνώση για τα ίδια τα δεδομένα.

Τα θεμέλια της μάθησης χωρίς επίβλεψη περιστρέφονται γύρω από την ανακάλυψη προ-

τύπων, συσχετισμών και δομών σε δεδομένα χωρίς τη βοήθεια ετικετών. Αυτό μπορεί να περιλαμβάνει τον εντοπισμό ομάδων παρόμοιων παραδειγμάτων μέσα στα δεδομένα, γνωστά ως ομαδοποίηση, ή την εύρεση μιας συμπαγούς αναπαράστασης που αποτυπώνει τα πιο σημαντικά χαρακτηριστικά των δεδομένων, γνωστή ως μείωση διαστάσεων. Η απουσία ετικετών σημαίνει ότι η αξιολόγηση των αλγορίθμων μάθησης χωρίς επίβλεψη είναι πιο σύνθετη από ό,τι στην εποπτευόμενη μάθηση, συχνά βασιζόμενη σε μετρήσεις που μετρούν την ποιότητα της δομής που ανακαλύπτεται από τον αλγόριθμο.

Πιο επίσημα, στην μάθηση χωρίς επίβλεψη μας δίνεται ένα σύνολο δεδομένων  $D$  με  $n$  δείγματα, όπου κάθε δείγμα είναι ένα διάνυσμα εισόδου  $\mathbf{x}_i$  στο χώρο  $\mathcal{X}$ . Ο στόχος είναι η μάθηση κάποιας συνάρτησης  $f : \mathcal{X} \rightarrow \mathcal{Y}$ , όπου το  $\mathcal{Y}$  αντιπροσωπεύει ένα σύνολο ιδιοτήτων, μοτίβων ή δομών που ανακαλύφθηκε στα δεδομένα [45].

1. **Ομαδοποίηση:** Η ομαδοποίηση στοχεύει να χωρίσει το σύνολο δεδομένων σε ομάδες, έτσι ώστε τα στοιχεία μέσα στο ίδιο σύμπλεγμα να είναι πιο παρόμοια μεταξύ τους παρά με εκείνα σε άλλα συμπλέγματα. Μια κοινή προσέγγιση είναι ο αλγόριθμος  $k$ -means, ο οποίος ελαχιστοποιεί το άθροισμα τετραγώνων εντός του συμπλέγματος (WCSS):

$$\min_{\mathbf{S}} \sum_{i=1}^k \sum_{\mathbf{x} \in S_i} \|\mathbf{x} - \mu_i\|^2$$

Εδώ, το  $k$  είναι ο αριθμός των συστάδων, το  $\mathbf{S} = \{S_1, S_2, \dots, S_k\}$  αντιπροσωπεύει τη διαίρεση των δεδομένων σε  $k$  συμπλέγματα και το  $\mu_i$  είναι ο μέσος όρος των σημείων στο  $S_i$ .

2. **Μείωση Διαστάσεων:** Η μείωση διαστάσεων επιδιώκει την προβολή δεδομένων από έναν χώρο υψηλής διαστατικότητας σε έναν χώρο χαμηλότερης διαστατικότητας, διατηρώντας παράλληλα όσο το δυνατόν περισσότερες σχετικές πληροφορίες. Το Principal Component Analysis (PCA) [46] είναι μια τέτοια δημοφιλής τεχνική που μεγιστοποιεί την διακύμανση των δεδομένων μειώνοντας την διαστατικότητα τους, όπως παρακάτω:

$$\max_{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_k} \sum_{i=1}^k \text{Var}(\mathbf{w}_i^\top \mathbf{X})$$

όπου πρέπει  $\mathbf{w}_i^\top \mathbf{w}_j = 0$  για  $i \neq j$ .

Στην μη-εποπτευόμενη μάθηση, η αξιολόγηση των αποτελεσμάτων μπορεί να είναι δύσκολη, καθώς δεν υπάρχει απλός τρόπος μέτρησης της ακρίβειας του μοντέλου χωρίς δεδομένα με ετικέτα. Η επιλογή του αλγορίθμου και των παραμέτρων εξαρτάται σε μεγάλο βαθμό από τα ειδικά χαρακτηριστικά των δεδομένων και το επιθυμητό αποτέλεσμα της ανάλυσης [47].

Συνοψίζοντας, η μάθηση χωρίς επίβλεψη είναι ένας κρίσιμος τομέας στην μηχανική μάθηση που παρέχει γνώση σχετικά με τη δομή των δεδομένων. Επιτρέπει την ανακάλυψη προτύπων και σχέσεων σε σύνολα δεδομένων χωρίς προϋπάρχουσες ετικέτες, προσφέροντας έναν τρόπο κατανόησης της εγγενούς πολυπλοκότητας των δεδομένων. Από την ομαδοποίηση πελατών με βάση την αγοραστική συμπεριφορά έως τη μείωση των διαστάσεων μεγάλων συνόλων δεδο-

μένων για οπτικοποίηση, οι αλγόριθμοι μάθησης χωρίς επίβλεψη διαδραματίζουν ζωτικό ρόλο στην εξερεύνηση δεδομένων και στην ανακάλυψη γνώσης.

### 3.1.3 Ενισχυτική Μάθηση

Η Ενισχυτική Μάθηση (RL) αντιπροσωπεύει μια ξεχωριστή κατηγορία στο ευρύτερο φάσμα μηχανικής μάθησης, εστιάζοντας στον τρόπο με τον οποίο διάφοροι πράκτορες (agents) πρέπει να πραγματοποιούν ενέργειες και να λαμβάνουν αποφάσεις σε ένα περιβάλλον για να μεγιστοποιήσουν κάποια ανταμοιβή. Σε αντίθεση με την εποπτευόμενη μάθηση, όπου τα μοντέλα μαθαίνουν από ένα σύνολο δεδομένων που παρέχει σωστές απαντήσεις, και την μάθηση χωρίς επίβλεψη, όπου τα μοντέλα συμπεραίνουν μοτίβα χωρίς γνωστές σωστές απαντήσεις, η ενισχυτική μάθηση αφορά τη μάθηση από τις συνέπειες των ενεργειών σε ένα δυναμικό περιβάλλον [48].

Στη βάση της ενισχυτικής μάθησης βρίσκεται η έννοια των πρακτόρων που μαθαίνουν να λαμβάνουν αποφάσεις μέσω δοκιμών, αλληλεπιδρώντας με ένα περιβάλλον στο οποίο πρέπει να εκτελούν ενέργειες και να παρατηρούν τα αποτελέσματα. Αυτή η διαδικασία καθοδηγείται από μια πολιτική - μια στρατηγική που καθορίζει τη δράση που πρέπει να επιλέξει ο πράκτορας σε μια δεδομένη κατάσταση - την οποία ο πράκτορας επιδιώκει να βελτιστοποιήσει με την πάροδο του χρόνου. Η ενισχυτική μάθηση προέρχεται από το μαθηματικό πλαίσιο των Διαδικασιών Αποφάσεων Markov (Markov Decision Processes, MDPs) [49], οι οποίες παρέχουν μια τυποποίηση για διαδοχικά προβλήματα λήψης αποφάσεων.

Μία MDP ορίζεται από ένα σύνολο καταστάσεων  $S$ , ένα σύνολο ενεργειών  $A$ , μια συνάρτηση μετάβασης  $T : S \times A \times S \rightarrow [0, 1]$  που μοντελοποιεί την πιθανότητα μετακίνησης από μια κατάσταση σε άλλη δεδομένης μιας ενέργειας και μια συνάρτηση ανταμοιβής  $R : S \times A \rightarrow \mathbb{R}$  που δίνει μια ανταμοιβή σε κάθε ενέργεια που πραγματοποιείται σε μια κατάσταση. Ο στόχος στην ενισχυτική μάθηση είναι να μαθευτεί μια πολιτική  $\pi : S \rightarrow A$  που μεγιστοποιεί την ανταμοιβή, που συχνά εκφράζεται ως το άθροισμα των ανταμοιβών με την πάροδο του χρόνου:

$$\max_{\pi} \mathbb{E} \left[ \sum_{t=0}^{\infty} \gamma^t R(s_t, \pi(s_t)) \right]$$

όπου  $s_t$  είναι η κατάσταση τη στιγμή  $t$ ,  $\pi(s_t)$  είναι η ενέργεια που πραγματοποιείται σύμφωνα με την πολιτική  $\pi$  και  $\gamma$  είναι ένας συντελεστής που μοντελοποιεί την παρούσα αξία των μελλοντικών ανταμοιβών [50].

Μια θεμελιώδης πρόκληση στην ενισχυτική μάθηση είναι ο συμβιβασμός μεταξύ της εξερεύνησης —δοκιμάζοντας νέες ενέργειες για την ανακάλυψη των επιπτώσεών τους— και της εκμετάλλευσης—της λήψης ενεργειών που είναι γνωστό ότι αποφέρουν υψηλές ανταμοιβές. Η αποτελεσματική μάθηση απαιτεί την εξισορρόπηση αυτών των ανταγωνιστικών στόχων για να ανακαλυφθούν οι πιο ανταποδοτικές ενέργειες και να συγκεντρωθεί όσο το δυνατόν περισσότερη ανταμοιβή.

Επιπλέον, η ενισχυτική μάθηση ενέχει υψηλή υπολογιστική πολυπλοκότητα καθιστώντας δύσκολη την εφαρμογή της σε προβλήματα του πραγματικού κόσμου. Η πολυπλοκότητα του σχεδιασμού ενός περιβάλλοντος, η διαμόρφωση της ανταμοιβής και η διασφάλιση ότι οι πολιτικές που έχουν μαθευτεί είναι ασφαλείς και ανθεκτικές σε αβεβαιότητες του μελλοντικού

περιβάλλοντος αποτελούν σημαντικά εμπόδια.

Το θεώρημα ‘No free lunch’ ισχύει επίσης και για την ενισχυτική μάθηση, υποδεικνύοντας ότι καμία στρατηγική δεν είναι καλύτερη για όλα τα σενάρια. Η επιτυχία μιας προσέγγισης εξαρτάται σε μεγάλο βαθμό από το πλαίσιο του προβλήματος, συμπεριλαμβανομένων των χαρακτηριστικών της κατάστασης και των χώρων δράσης και της δυναμικής του περιβάλλοντος.

Συνολικά, η ενισχυτική μάθηση προσφέρει τη δυνατότητα σε υπολογιστικά συστήματα να μάθουν σύνθετες συμπεριφορές πέρα από στατικά σύνολα δεδομένων, επιτρέποντάς τους να προσαρμοστούν σε εξελισσόμενα περιβάλλοντα και να λαμβάνουν αλληλουχίες αποφάσεων που οδηγούν σε έναν στόχο. Οι εφαρμογές του κυμαίνονται από αυτόνομα οχήματα έως αλγόριθμους διαχείρισης επενδυτικών χαρτοφυλακίων [51].

Συνοψίζοντας, οι σφαίρες της εποπτευόμενης, χωρίς επίβλεψη και της ενισχυτικής μάθησης περιλαμβάνουν το τεράστιο και δυναμικό πεδίο της μηχανικής μάθησης, προσφέροντας ποικίλες μεθοδολογίες για την εξαγωγή προτύπων, τη δημιουργία προβλέψεων και τη διευκόλυνση της λήψης αποφάσεων από μεγάλο όγκο δεδομένων. Στην παρούσα διπλωματική εργασία, κύριο αντικείμενο ενασχόλησης θα είναι η εποπτευόμενη μάθηση.

Ο τομέας της πρόβλεψης χρονοσειρών, με τον οποίο ουσιαστικά ασχολείται η παρούσα εργασία, έχει επωφεληθεί σημαντικά από την ανάπτυξη μεθοδολογιών επιβλεπόμενης μηχανικής μάθησης. Παρακάτω παρουσιάζονται οι κυριότεροι αλγόριθμοι επιβλεπόμενης μάθησης που χρησιμοποιούνται για την πρόβλεψη χρονοσειρών.

## 3.2 SVM και SVR

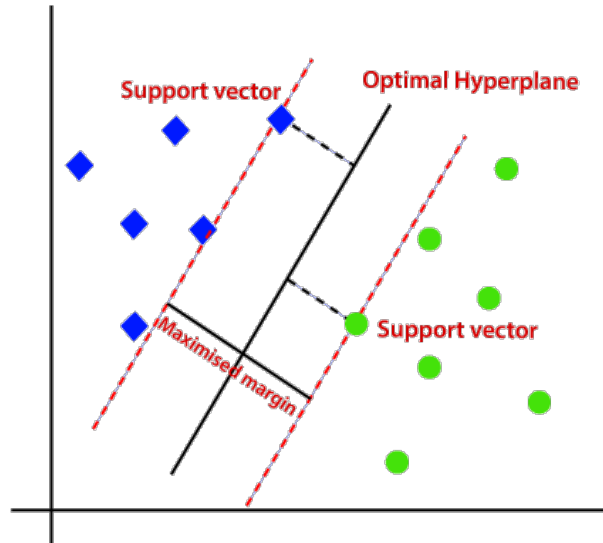
Οι Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines, SVMs) είναι ένα σύνολο εποπτευόμενων μεθόδων μηχανικής μάθησης που χρησιμοποιούνται για ταξινόμηση, παλινδρόμηση και ανίχνευση ακραίων τιμών. Στην πιο κλασική περίπτωση, το SVM βρίσκει το βέλτιστο υπερεπίπεδο που διαχωρίζει διαφορετικές κλάσεις στον χώρο με το μέγιστο δυνατό περιθώριο μεταξύ τους. Τα υποστηρικτικά διανύσματα παλινδρόμησης (SVR) επεκτείνουν τα SVM σε προβλήματα παλινδρόμησης, με στόχο την προσαρμογή του σφάλματος εντός ενός συγκεκριμένου διαστήματος. Για να γίνουν αντιληπτά τα SVR, που μας απασχολούν περισσότερο στην πρόβλεψη χρονοσειρών, ιδιαίτερα σημαντική κρίνεται μία σύντομη εισαγωγή στα SVM.

### 3.2.1 Θεωρητικό Πλαίσιο

Έστω ότι σε ένα πρόβλημα δυαδικής ταξινόμησης, δίνεται ένα σύνολο δεδομένων εκπαίδευσης με  $n$  σημεία της μορφής  $(\mathbf{x}_i, y_i)$ , όπου  $\mathbf{x}_i \in \mathbb{R}^p$  και  $y_i \in \{1, -1\}$ . Σε αυτό το πρόβλημα το SVM αναζητά ένα υπερεπίπεδο που ορίζεται από το διάνυσμα  $\mathbf{w}$  και το  $b$  που διαχωρίζει τις δύο κλάσεις. Το βέλτιστο υπερεπίπεδο μεγιστοποιεί το περιθώριο μεταξύ των δύο κλάσεων, που ορίζεται ως η απόσταση μεταξύ του υπερεπιπέδου και του πλησιέστερου σημείου κάθε κατηγορίας.

Το συγκεκριμένο πρόβλημα μπορεί να μοντελοποιηθεί και σαν μία περίπτωση βελτιστοποίησης [52]:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2$$



Εικόνα 3.1: Το SVM

με περιορισμό  $y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1, \forall i$

Για μη γραμμικά διαχωρίσιμα δεδομένα, το SVM χρησιμοποιεί έναν επιπλέον μηχανισμό (kernel trick) για να αντιστοιχίσει τα χαρακτηριστικά εισόδου σε χώρο υψηλών διαστάσεων όπου είναι δυνατός ένας γραμμικός διαχωρισμός [52].

Το SVR εφαρμόζει τη λογική του SVM σε προβλήματα παλινδρόμησης. Αντί να βρει ένα υπερεπίπεδο για να χωρίσει δύο κλάσεις, το SVR βρίσκει μια συνάρτηση  $f(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + b$  έτσι ώστε τα σφάλματα να μην υπερβαίνουν ένα καθορισμένο όριο  $\epsilon$ , ενώ ταυτόχρονα ελαχιστοποιείται το  $\|\mathbf{w}\|$  για να διασφαλιστεί ότι το μοντέλο είναι όσο το δυνατόν πιο επίπεδο [53].

Η συγκεκριμένη προσέγγιση εισάγει μεταβλητές  $\xi_i, \xi_i^*$  για τη μέτρηση της απόκλισης εκτός της ζώνης- $\epsilon$ :

$$\min_{\mathbf{w}, b, \xi, \xi^*} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*)$$

όπου ισχύουν οι περιορισμοί:

$$\begin{cases} y_i - \mathbf{w}^\top \mathbf{x}_i - b \leq \epsilon + \xi_i \\ \mathbf{w}^\top \mathbf{x}_i + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \forall i \end{cases}$$

### 3.2.2 SVR και Προβλέψεις Χρονοσειρών

Η πρόβλεψη χρονοσειρών στην μορφή με την οποία ασχολείται η παρούσα εργασία, περιλαμβάνει τη χρήση ιστορικών δεδομένων της χρονοσειράς για την πρόβλεψη μελλοντικών τιμών. Το SVR είναι ιδιαίτερα κατάλληλο για αυτήν την εργασία λόγω της ικανότητάς του να μοντελοποιεί πολύπλοκες μη γραμμικές σχέσεις και την ανθεκτικότητά του στην υπερπροσαρμογή, ειδικά σε χώρους υψηλών διαστάσεων [54, 55].

Το πρώτο βήμα για την εφαρμογή του SVR στην πρόβλεψη χρονοσειρών περιλαμβάνει

την προεπεξεργασία των δεδομένων έτσι ώστε να προκύψει μια μορφή κατάλληλη για παλινδρόμηση. Μια κοινή προσέγγιση είναι η μετατροπή των δεδομένων χρονοσειρών σε ένα εποπτευόμενο πρόβλημα μηχανικής μάθησης. Αυτό γίνεται συνήθως με τη δημιουργία διανυσμάτων εισόδου από τις προηγούμενες τιμές της χρονοσειράς και η τιμή που θα προβλεφθεί αντιμετωπίζεται ως μεταβλητή εξόδου (στόχος).

Πιο συγκεκριμένα, δεδομένης μιας χρονοσειράς  $Y = \{y_1, y_2, \dots, y_T\}$ , μπορούμε να δημιουργήσουμε ένα σύνολο δεδομένων για ένα εποπτευόμενο πρόβλημα μάθησης όπου κάθε δείγμα αποτελείται από  $n$  παρατηρήσεις ως μεταβλητές εισόδου  $X = \{x_1, x_2, \dots, x_n\}$  και η μελλοντική τιμή  $y_{t+1}$  ως μεταβλητή εξόδου. Αυτός ο μετασχηματισμός μπορεί να αναπαρασταθεί μαθηματικά ως:

$$X_t = [y_{t-n+1}, \dots, y_t] \rightarrow y_{t+1}$$

όπου  $X_t$  είναι το διάνυσμα εισόδου τη στιγμή  $t$ , που αποτελείται από τις  $n$  πιο πρόσφατες παρατηρήσεις και  $y_{t+1}$  είναι η τιμή στόχος που πρέπει να προβλεφθεί.

Από τη στιγμή που τα δεδομένα είναι στην επιθυμητή μορφή, μπορεί να χρησιμοποιηθεί το μοντέλο SVR που παρουσιάστηκε παραπάνω. Όπως γίνεται εύκολα αντιληπτό, λόγω της μη γραμμικότητας που παρατηρείται στις χρονοσειρές συνήθως, ιδιαίτερης σημασίας είναι η χρήση μη γραμμικής παλινδρόμησης, κάτι που επιτυγχάνεται όπως φάνηκε και παραπάνω με τις συναρτήσεις πυρήνα (Kernel) [56]. Μία δημοφιλής επιλογή πυρήνα στην πρόβλεψη χρονοσειρών είναι η Radial Basis Function (RBF) [57].

Για την πρόβλεψη πολλαπλών μελλοντικών τιμών, δύο βασικές προσεγγίσεις χρησιμοποιούνται με το SVR:

1. **Άμεση Μέθοδος:** Ξεχωριστά μοντέλα SVR εκπαιδεύονται ώστε να προβλέπουν απευθείας κάθε μελλοντικό χρονικό βήμα. Για την πρόβλεψη των βημάτων  $h$  μπροστά, εκπαιδεύονται  $h$  διαφορετικά μοντέλα, καθένα από τα οποία στοχεύει ένα διαφορετικό βήμα στο μέλλον.
2. **Επαναληπτική (Αναδρομική) Μέθοδος:** Ένα μεμονωμένο μοντέλο SVR εκπαιδεύεται να προβλέπει ένα βήμα μπροστά και οι προβλέψεις του χρησιμοποιούνται στο διάνυσμα εισόδου για την πρόβλεψη των επόμενων βημάτων.

Συνολικά, η ικανότητα του SVR να μοντελοποιεί πολύπλοκες μη γραμμικές σχέσεις, μαζί με την ανθεκτικότητά του έναντι της υπερπροσαρμογής, το καθιστά μια πολύτιμη μέθοδο για την πρόβλεψη χρονοσειρών. Ωστόσο, ενώ το SVR παρέχει ένα ισχυρό εργαλείο για την πρόβλεψη χρονοσειρών, πρέπει να αντιμετωπιστούν αρκετές προκλήσεις, όπως η αντιμετώπιση της υπολογιστικής πολυπλοκότητας για μεγάλα σύνολα δεδομένων. Επιπλέον, η επιλογή μεταξύ της άμεσης και της επαναληπτικής μεθόδου για την πρόβλεψη πολλαπλών μελλοντικών τιμών περιλαμβάνει μια αντιστάθμιση μεταξύ της υπολογιστικής αποτελεσματικότητας και της πιθανής συσσώρευσης σφαλμάτων πρόβλεψης στην επαναληπτική προσέγγιση.

### 3.3 Δέντρα Απόφασης και Τυχαία Δάση

Τα Δέντρα Αποφάσεων και τα Τυχαία Δάση είναι δύο μεθοδολογίες μηχανικής μάθησης που αναγνωρίζονται ευρέως για την επεξηγηματικότητα, την ευελιξία και την ισχυρή απόδοση

τους σε διάφορες εργασίες, συμπεριλαμβανομένης της ταξινόμησης, της παλινδρόμησης και κατ' επέκταση της πρόβλεψης χρονοσειρών.

### 3.3.1 Θεωρητικό Πλαίσιο

Από την πλευρά της ταξινόμησης, όπου χρησιμοποιήθηκε αρχικά, ένα δέντρο απόφασης είναι μια δομή που μοιάζει με διάγραμμα ροής όπου κάθε εσωτερικός κόμβος αντιπροσωπεύει ένα "τεστ" για ένα χαρακτηριστικό, κάθε διακλάδωση αντιπροσωπεύει το αποτέλεσμα του τεστ και κάθε φύλλο (τελικός κόμβος) αντιπροσωπεύει μια ετικέτα κλάσης. Τα τυχαία δάση επεκτείνουν τη μεθοδολογία των δέντρων απόφασης δημιουργώντας ένα σύνολο δέντρων στοχεύοντας την αποφυγή της υπερπροσαρμογής και τη βελτίωση της ακρίβειας πρόβλεψης.

Η κατασκευή ενός Δέντρου Αποφάσεων περιλαμβάνει τον αναδρομικό διαχωρισμό των δεδομένων με βάση συγκεκριμένα χαρακτηριστικά για να μεγιστοποιηθεί η ομοφωνία στις ετικέτες καθώς προχωράμε σε θυγατρικούς κόμβους. Η απόφαση για το πώς και πού θα διαχωριστούν τα δεδομένα βασίζεται σε μέτρα όπως το Gini Impurity ή η Εντροπία (Information Gain) [58].

Επεκτείνοντας αυτόν τον τρόπο σκέψης στις περιπτώσεις παλινδρόμησης, οι οποίες μας ενδιαφέρουν και περισσότερο στην πρόβλεψη χρονοσειρών, ένα δέντρο αποφάσεων προβλέπει συνεχή αποτελέσματα με βάση τα χαρακτηριστικά εισόδου μέσω μιας σειράς δυαδικών αποφάσεων, διαχωρίζοντας αποτελεσματικά τον χώρο εισόδου σε περιοχές με κοντινές τιμές απόκρισης.

Ένα δέντρο παλινδρόμησης δημιουργείται μέσω μιας διαδικασίας γνωστής ως αναδρομικός δυαδικός διαχωρισμός (recursive binary splitting). Ο στόχος είναι να βρεθούν οι διαχωριστικές τιμές που ελαχιστοποιούν τη συνολική διακύμανση τιμών-στόχων μέσα στα φύλλα. Δεδομένου ενός συνόλου δεδομένων  $D$  με χαρακτηριστικά  $X$  και μια συνεχή μεταβλητή στόχο  $Y$ , ο αλγόριθμος επιδιώκει να χωρίσει το  $D$  σε  $M$  περιοχές  $R_1, R_2, \dots, R_M$  που ελαχιστοποιούν το συνολικό άθροισμα των τετραγωνικών υπολειμμάτων:

$$\min_{\{R_m\}} \sum_{m=1}^M \sum_{i \in R_m} (y_i - \hat{y}_{R_m})^2$$

Το δέντρο απόφασης μεγαλώνει επιλέγοντας αναδρομικά το χαρακτηριστικό και το σημείο διαχωρισμού που έχουν ως αποτέλεσμα τη μεγαλύτερη μείωση της διακύμανσης της τιμής-στόχου [59].

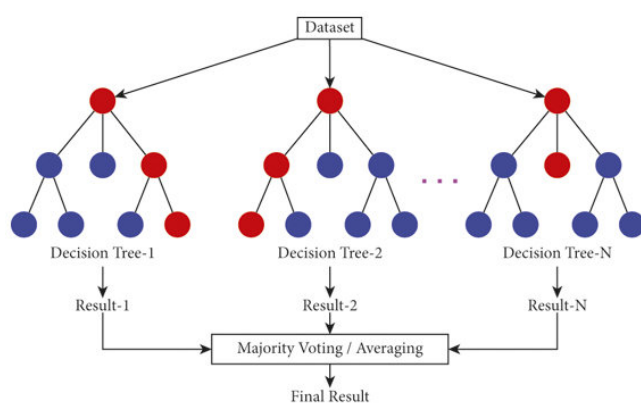
Για να αποφευχθεί η υπερπροσαρμογή, τα δέντρα κλαδεύονται από ένα μέγιστο βάθος και έπειτα. Το κλάδεμα πολυπλοκότητας, γνωστό και ως κλάδεμα ασθενέστερου συνδέσμου (weakest link pruning) [60], εισάγει μια παράμετρο πολυπλοκότητας  $\alpha$  που τιμωρεί τον αριθμό των φύλλων, βρίσκοντας αποτελεσματικά μια ισορροπία μεταξύ του βάθους του δέντρου και της προσαρμογής του στα δεδομένα.

Τα τυχαία δάση (random forests) επεκτείνουν την έννοια του δέντρου απόφασης σε μια μέθοδο ensemble, όπου αναπτύσσονται πολλά δέντρα και υπολογίζεται ο μέσος όρος των προβλέψεών τους, με στόχο τη βελτίωση της προγνωστικής ακρίβειας και την μείωση της πιθανότητας για υπερπροσαρμογή. Η θεμελιώδης ιδέα πίσω από τα Random Forests είναι ότι μια ομάδα «αδύναμων μοντέλων» μπορεί να ομαδοποιηθεί για να σχηματίσει ένα «ισχυρό

μοντέλο». Τα τυχαία δάση σε εργασίες παλινδρόμησης βασίζονται στην έννοια του bagging και της τυχαίας επιλογής χαρακτηριστικών. Δεδομένου ενός συνόλου δεδομένων  $D$ , επιλέγονται  $B$  αρχικά δείγματα. Ένα δέντρο απόφασης αναπτύσσεται από κάθε δείγμα, αλλά σε κάθε κόμβο, επιλέγεται ένα τυχαίο υποσύνολο χαρακτηριστικών  $k$  ως υποψήφια για τον διαχωρισμό από το πλήρες σύνολο των χαρακτηριστικών  $p$  [61]. Η τελική πρόβλεψη του Τυχαίου Δάσους είναι ο μέσος όρος των προβλέψεων των  $B$  δέντρων:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B \hat{y}^{(b)}(x)$$

όπου  $\hat{y}^{(b)}(x)$  είναι η πρόβλεψη του  $b$  δέντρου.



Εικόνα 3.2: Το Random Forest

Η εισαγωγή της τυχαιότητας στην επιλογή χαρακτηριστικών σε κάθε σημείο διαίρεσης μειώνει τη συσχέτιση μεταξύ των δέντρων και αυξάνει τη διακύμανση του συνόλου, οδηγώντας σε πιο ισχυρές συνολικές προβλέψεις [62].

### 3.3.2 Δέντρα Αποφάσεων, Τυχαία Δάση και Πρόβλεψη Χρονοσειρών

Για την προσαρμογή των δέντρων αποφάσεων και των τυχαίων δασών έτσι ώστε να μπορούν να συμβάλουν στην πρόβλεψη χρονοσειρών, πρέπει, όπως και στην περίπτωση των SVR, να γίνει προσεκτική επιλογή χαρακτηριστικών (feature engineering) για να γίνει μάθηση των χρονικών εξαρτήσεων.

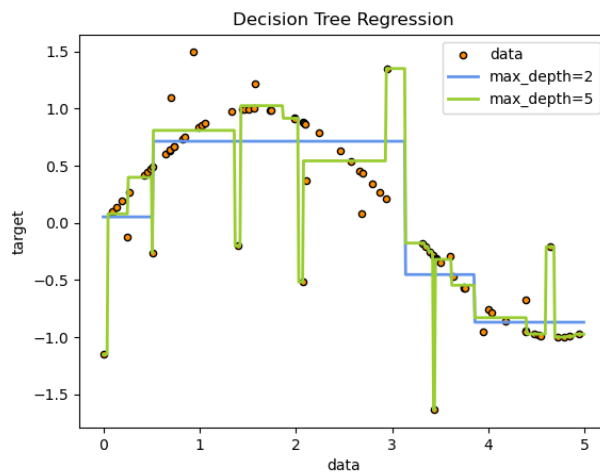
Για άλλη μία φορά, στο είδος προβλέψεων που εξετάζει η παρούσα εργασία, οι παρελθοντικές τιμές της χρονοσειράς χρησιμοποιούνται ως προγνωστικοί παράγοντες. Ωστόσο, θα μπορούσαν να εισαχθούν και επιπλέον διαστάσεις στο διάνυσμα εισόδου, εισάγοντας χαρακτηριστικά όπως κινητούς μέσους όρους, διαφορές ( $Y_t - Y_{t-1}$ ) και άλλους μετασχηματισμούς για την καταγραφή των τάσεων και της εποχικότητας.

Όπως και στα SVR, όταν θέλουμε να προβλέψουμε πάνω από μία μελλοντικές τιμές υπάρχουν δύο βασικές επιλογές. Στην περίπτωση της άμεσης πρόβλεψης, δημιουργούνται ξεχωριστά μοντέλα για κάθε ορίζοντα πρόβλεψης. Στην αναδρομική πρόβλεψη, ένα μοντέλο



εκπαιδεύεται να προβλέπει ένα βήμα μπροστά και η έξοδος του χρησιμοποιείται ως χαρακτηριστικό εισόδου για επόμενες προβλέψεις.

Συνολικά, τα δέντρα απόφασης και τα τυχαία δάση είναι ισχυρά εργαλεία που χρησιμεύουν ιδιαίτερα στην παλινδρόμηση, συμπεριλαμβανομένης και της πρόβλεψης χρονοσειρών. Η ευελιξία τους να μοντελοποιούν πολύπλοκες, μη γραμμικές σχέσεις και αλληλεπιδράσεις μεταξύ των παρελθοντικών τιμών τα καθιστά ιδιαίτερα κατάλληλα για την πρόβλεψη μελλοντικών τιμών σε δεδομένα χρονοσειρών [63]. Ωστόσο, σημαντική είναι η επιλογή των σωστών υπερπαραμέτρων, όπως το βάθος δέντρου, έτσι ώστε να εξασφαλιστούν υψηλού επιπέδου προβλέψεις, καθώς πάντα ελλοχεύουν κίνδυνοι όπως η υπερπροσαρμογή [64].



Εικόνα 3.3: Πρόβλεψη χρονοσειράς με δέντρο απόφασης



## Κεφάλαιο **4**

# Βαθιά Μάθηση και Προβλέψεις Χρονοσειρών

---

### 4.1 Εισαγωγή στα Νευρωνικά Δίκτυα

Ο ανθρώπινος εγκέφαλος, ένα όργανο ιδιαίτερης πολυπλοκότητας και αποτελεσματικότητας, χρησιμεύει ως το βιολογικό πρότυπο για το σχεδιασμό των νευρωνικών δικτύων. Αποτελούμενος από περίπου 86 δισεκατομμύρια νευρώνες, η λειτουργικότητα του εγκεφάλου προκύπτει από ένα περίπλοκο δίκτυο συνδέσεων που διευκολύνει την επικοινωνία μεταξύ αυτών των κυττάρων. Αυτό το δίκτυο αποτελεί το θεμέλιο για όλες τις λειτουργίες του εγκεφάλου, συμπεριλαμβανομένης της μάθησης, της μνήμης και των διαδικασιών λήψης αποφάσεων. Οι νευρώνες επικοινωνούν μέσω συνάψεων, όπου τα χημικά σήματα μετατρέπονται σε ηλεκτρικά, επιτρέποντας τη μετάδοση πληροφοριών στον εγκέφαλο. Αυτή η δυναμική διαδικασία, όπου η είσοδος από πολλούς νευρώνες ενσωματώνεται για να παραχθεί μια έξοδος, αποτέλεσε τη βάση για τους τεχνητούς νευρώνες [65].

Η κατανόηση του τρόπου με τον οποίο ο εγκέφαλος μαθαίνει και προσαρμόζεται έχει κεντρική σημασία για την αναπαραγωγή των διαδικασιών του. Η μάθηση στον εγκέφαλο περιλαμβάνει την ενίσχυση ή την αποδυνάμωση των συναπτικών συνδέσεων, μια έννοια γνωστή ως συναπτική πλαστικότητα. Αυτή η ικανότητα τροποποίησης των συνδέσεων με βάση την εμπειρία αντικατοπτρίζεται στην προσαρμοστικότητα των νευρωνικών δικτύων μέσω των αλγορίθμων μάθησης.

Το άλμα από τις βιολογικές διεργασίες στην τεχνητή αναπαραγωγή ξεκίνησε με την εξερεύνηση των βασικών λειτουργιών του νευρώνα. Ένας τεχνητός νευρώνας μιμείται τις λειτουργικές αρχές του βιολογικού ομολόγου του, λαμβάνοντας εισόδους, επεξεργαζόμενος αυτές μέσω ενός σταθμισμένου αθροίσματος και χρησιμοποιώντας μια συνάρτηση ενεργοποίησης για τον προσδιορισμό της εξόδου. Αυτό το μοντέλο συμπυκνώνει την ουσία της νευρωνικής δραστηριότητας: χρήση των εισόδων και παραγωγή ενός σήματος εξόδου.

### 4.2 Τεχνητός Νευρώνας

Παρακάτω θα εμβαθύνουμε στη δομή και τα χαρακτηριστικά του τεχνητού νευρώνα, που είναι αναπόσπαστο μέρος του σχεδιασμού και της λειτουργίας των νευρωνικών δικτύων. Ο τεχνητός νευρώνας, που συχνά αναφέρεται ως perceptron στην απλούστερη μορφή του, αποτελεί τον ακρογωνιαίο λίθο των τεχνητών νευρωνικών δικτύων, αντικατοπτρίζοντας τα χαρακτηριστικά επεξεργασίας πληροφοριών των βιολογικών νευρώνων.

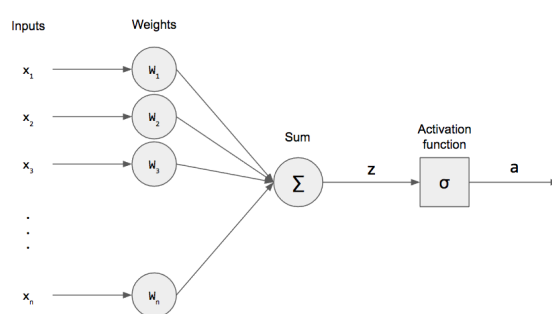
Ένας τεχνητός νευρώνας έχει σχεδιαστεί για να προσομοιώνει την ικανότητα επεξεργασίας σήματος ενός βιολογικού νευρώνα. Λαμβάνει πολλαπλές εισόδους, καθεμία από τις οποίες αντιπροσωπεύει ένα χαρακτηριστικό ή ένα σήμα παρόμοιο με τις ηλεκτρικές ώσεις που συλλαμβάνονται από τους δενδρίτες ενός νευρώνα από τους προηγούμενους νευρώνες. Αυτές οι εισόδοι συμβολίζονται ως  $x_1, x_2, \dots, x_n$ , όπου το  $n$  αντιπροσωπεύει τον αριθμό των εισόδων του νευρώνα.

Κάθε είσοδος  $x_i$  συνδέεται με ένα βάρος  $w_i$ , το οποίο είναι μια αριθμητική τιμή που αντιπροσωπεύει την ισχύ ή τη σημασία της συγκεκριμένης εισόδου στην έξοδο του νευρώνα. Αυτή η έννοια αντικατοπτρίζει τη βιολογική αποτελεσματικότητα των συναπτικών συνδέσεων, όπου η ισχύς μιας σύναψης καθορίζει πόσο μεγάλη επιρροή έχει ένας νευρώνας σε έναν άλλο.

Εκτός από τα βάρη, το μοντέλο ενσωματώνει έναν όρο μεροληψίας  $b$ , ένα κρίσιμο στοιχείο που επιτρέπει την προσαρμογή της συνάρτησης ενεργοποίησης. Η μεροληψία μπορεί να θεωρηθεί ως το επίπεδο κατωφλίου του νευρώνα που πρέπει να ξεπεραστεί για να πυροδοτηθεί ή να ενεργοποιηθεί ο νευρώνας, θυμίζοντας το δυναμικό κατωφλίου στους βιολογικούς νευρώνες.

Το σταθμισμένο άθροισμα των εισόδων συν τη μεροληψία υπολογίζεται από τη συνάρτηση άθροισης, η οποία ενσωματώνει τα σήματα που λαμβάνει ο νευρώνας. Μαθηματικά, η διαδικασία αυτή αναπαρίσταται ως  $\sum_{i=1}^n w_i x_i + b$ , η οποία συνδυάζει τις εισόδους σταθμισμένες με την αντίστοιχη σημασία τους, προσαρμοσμένες από την προκατάληψη  $b$ , για να παράγει μια μοναδική τιμή που χρησιμεύει ως είσοδος για τη συνάρτηση ενεργοποίησης [66].

Η συνάρτηση ενεργοποίησης  $\phi$ , ένα θεμελιώδες στοιχείο του τεχνητού νευρώνα, επεξεργάζεται αυτή την είσοδο για να παράγει την έξοδο του νευρώνα. Με την εισαγωγή μη γραμμικότητας, επιτρέπει στον νευρώνα να μαθαίνει πολύπλοκα μοτίβα και σχέσεις μέσα στα δεδομένα. Η μαθηματική έκφραση της συνάρτησης,  $y = \phi(\sum_{i=1}^n w_i x_i + b)$ , αναδεικνύει το ρόλο της στη μετατροπή του γραμμικού συνδυασμού των εισόδων σε μη γραμμική έξοδο  $y$ , η οποία μπορεί στη συνέχεια να μεταδοθεί σε επόμενους νευρώνες ή να χρησιμεύσει ως η τελική έξοδος του δικτύου.



Εικόνα 4.1: Ο Τεχνητός Νευρώνας

Εξερευνώντας περαιτέρω τις συναρτήσεις ενεργοποίησης, συναντάμε μια ποικιλία επιλογών, η καθεμία με ξεχωριστά χαρακτηριστικά και επίδραση στην έξοδο του νευρώνα [67]:

- **Γραμμική Συνάρτηση Ενεργοποίησης:** Ορίζεται ως  $\phi(z) = z$ , όπου εξάγεται απευθείας η είσοδος χωρίς μετασχηματισμό. Η απλότητά της αποτελεί βασικό πλεονέκτημα, καθώς επεξεργάζεται απ' ευθείας τα δεδομένα εισόδου, αποφεύγοντας επιπλοκές όπως το πρόβλημα της εξαφάνισης ή της έκρηξης των κλίσεων (vanishing or

exploding gradients) λόγω της σταθερής παραγώγου της. Ωστόσο, η αδυναμία της να μοντελοποιήσει πολύπλοκα πρότυπα δεδομένων, λόγω της γραμμικότητάς της, περιορίζει την εφαρμογή της κυρίως στα επίπεδα εξόδου σε μοντέλα παλινδρόμησης, καθιστώντας την αναποτελεσματική για εργασίες που απαιτούν την αποτύπωση μη γραμμικών σχέσεων.

- **Σιγμοειδής Συνάρτηση Ενεργοποίησης:** Δίνεται από τη σχέση  $\phi(z) = \frac{1}{1+e^{-z}}$ , παράγοντας εξόδους στο εύρος  $(0, 1)$ . Αυτή η συνάρτηση είναι ιδιαίτερα πλεονεχτική για μοντέλα όπου οι εξόδοι ερμηνεύονται ως πιθανότητες, χάρη στην ομαλή, συνεχώς παραγωγίσιμη κλίση της που βοηθά τις μεθόδους βελτιστοποίησης με βάση την κλίση. Παρ' όλα αυτά, πάσχει από το πρόβλημα της εξαφανιζόμενης κλίσης για εισόδους μεγάλου μεγέθους [68].
- **ReLU (διορθωμένη γραμμική μονάδα):** Ορίζεται από τη σχέση  $\phi(z) = \max(0, z)$ . Ενεργοποιεί τους νευρώνες μόνο εάν η είσοδος είναι θετική, συμβάλλοντας στην ταχύτερη και αποτελεσματικότερη εκπαίδευση σε αρχιτεκτονικές νευρωνικών δικτύων. Η υπολογιστική απλότητα της ReLU και η λύση που δίνει στο πρόβλημα της εξαφανιζόμενης κλίσης, παρέχοντας γραμμική απόκριση για θετικές εισόδους, την έχει καταστήσει δημοφιλή επιλογή. Ωστόσο, η ReLU είναι επιρρεπής στο πρόβλημα "dying ReLU", όπου οι νευρώνες μπορεί να γίνουν ανενεργοί για όλες τις εισόδους εάν λαμβάνουν μόνο αρνητικά σήματα, οδηγώντας σε απώλεια ικανότητας μάθησης σε τμήματα του δικτύου [69]. Όπως και οι προηγούμενες συναρτήσεις ενεργοποίησης, οι εξόδοι του ReLU δεν είναι μηδενικοκεντρικές, γεγονός που μπορεί να περιπλέξει τη διαδικασία βελτιστοποίησης.
- **Συνάρτηση υπερβολικής εφαπτομένης:** Περιγράφεται από τη σχέση  $\phi(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ , προσφέροντας τιμές εξόδου στο εύρος από  $-1$  έως  $1$ . Η μηδενικοκεντρική της φύση βελτιώνει τη σύγκλιση κατά τη διάρκεια της εκπαίδευσης, καθιστώντας την κατάλληλη για χρήση σε κρυφά στρώματα ενός νευρωνικού δικτύου όπου πρέπει να αποτυπωθούν πολύπλοκα μοτίβα δεδομένων. Παρά τα πλεονεκτήματά της, η tanh μοιράζεται το πρόβλημα της εξαφανιζόμενης κλίσης με τη σιγμοειδή συνάρτηση, όπου εισόδοι μεγάλου μεγέθους μπορούν να οδηγήσουν σε κορεσμένες εξόδους, επιβραδύνοντας τη διαδικασία μάθησης στα δίκτυα [70].
- **Leaky ReLU:** Προσπαθεί να διορθώσει τους περιορισμούς της ReLU ως εξής:  $\phi(z) = \max(\alpha z, z)$ , όπου  $\alpha$  είναι μια μικρή σταθερά. Αυτή η τροποποίηση επιτρέπει μια μικρή κλίση όταν η είσοδος είναι αρνητική, διατηρώντας έτσι τη συμμετοχή ενεργών νευρώνων σε όλο το δίκτυο και ενισχύοντας την ικανότητα μάθησης του μοντέλου [71]. Η εξάρτηση της απόδοσης από την παράμετρο  $\alpha$ , ωστόσο, εισάγει μια πρόκληση, καθώς η βέλτιστη τιμή της ποικίλλει σημαντικά ανά περίπτωση και επηρεάζει σημαντικά την αποτελεσματικότητα του δικτύου.

Γενικά, η επιλογή μιας συνάρτησης ενεργοποίησης είναι αναπόσπαστο μέρος του σχεδιασμού των νευρωνικών δικτύων, επηρεάζοντας την ικανότητά τους να μαθαίνουν και να μοντελοποιούν σύνθετες σχέσεις. Ενώ οι γραμμικές συναρτήσεις προσφέρουν απλότητα και άμεση επεξεργασία για εργασίες παλινδρόμησης, οι συναρτήσεις sigmoid και tanh παρέχουν ομαλές

εξόδους κατάλληλες για ταξινόμηση και λεπτομερή μοντελοποίηση στα κρυφά στρώματα, αν και εμφανίζουν τάση για εξαφανιζόμενες κλίσεις. Η ReLU και η παραλλαγή της, η Leaky ReLU, αντιμετωπίζουν ορισμένες από αυτές τις προκλήσεις.

### 4.3 MLP

Μετά την παρουσίαση του τεχνητού νευρώνα και των συναρτήσεων ενεργοποίησης, θα εμβραθύνουμε στον τρόπο με τον οποίο αυτοί οι νευρώνες συνδέονται για να σχηματίσουν νευρωνικά δίκτυα, με ιδιαίτερη έμφαση στην αρχιτεκτονική του Multilayer Perceptron (MLP).

Το MLP ανήκει στην ευρύτερη κατηγορία των τεχνητών νευρωνικών δικτύων (ANN), τα οποία διακρίνονται από την οργάνωσή τους σε πολλαπλά στρώματα νευρώνων: ένα στρώμα εισόδου, διάφορα κρυφά στρώματα και ένα στρώμα εξόδου. Το χαρακτηριστικό γνώρισμα της αρχιτεκτονικής του MLP είναι η πλήρως συνδεδεμένη φύση του, όπου κάθε νευρώνας σε ένα δεδομένο στρώμα δημιουργεί συνδέσεις με κάθε νευρώνα στο επόμενο στρώμα. Αυτή η ολοκληρωμένη διασύνδεση είναι καθοριστική για την ικανότητα του MLP να συλλαμβάνει και να μοντελοποιεί τις πολύπλοκες, μη γραμμικές σχέσεις που συχνά υπάρχουν μεταξύ των δεδομένων εισόδου και των επιθυμητών εξόδων.

Η αρχιτεκτονική αρχίζει με το στρώμα εισόδου, όπου οι επιμέρους νευρώνες αντιστοιχούν άμεσα στα διάφορα χαρακτηριστικά που υπάρχουν στα δεδομένα εισόδου. Η διάσταση του διανύσματος εισόδου  $X = [x_1, x_2, \dots, x_n]$  υπαγορεύει τον αριθμό των νευρώνων σε αυτό το επίπεδο, δημιουργώντας μια σχέση ένα προς ένα μεταξύ των χαρακτηριστικών εισόδου και των νευρώνων. Το στρώμα εισόδου λειτουργεί ως η πύλη μέσω της οποίας τα δεδομένα εισέρχονται στο δίκτυο, θέτοντας τις βάσεις για την επακόλουθη επεξεργασία.

Μετά το στρώμα εισόδου ακολουθούν ένα ή περισσότερα κρυφά στρώματα, τα οποία αποτελούν τον πυρήνα της υπολογιστικής ισχύος του MLP. Οι νευρώνες σε αυτά τα στρώματα αναλαμβάνουν το κρίσιμο έργο του μετασχηματισμού των σταθμισμένων εισόδων που λαμβάνονται από το προηγούμενο στρώμα. Ο μετασχηματισμός αυτός περιλαμβάνει την άθροιση των σταθμισμένων εισόδων, την ενσωμάτωση ενός όρου προκατάληψης και, στη συνέχεια, την εφαρμογή μιας μη γραμμικής συνάρτησης ενεργοποίησης.

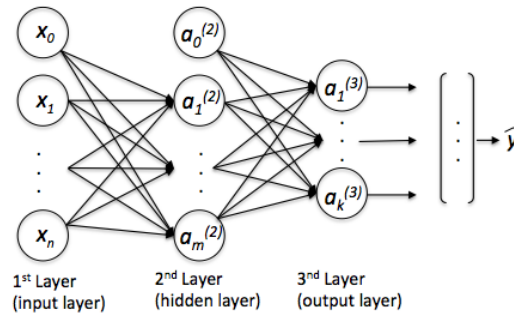
Για κάθε νευρώνα σε ένα επόμενο κρυφό στρώμα  $k$ , η έξοδος  $h_{jk}^{(k)}$  προκύπτει από το

$$h_{jk}^{(k)} = \phi \left( \sum_m w_{jm}^{(k)} h_m^{(k-1)} + b_j^{(k)} \right),$$

απεικονίζοντας τη διαδοχική, πολυεπίπεδη επεξεργασία των δεδομένων μέσω του δικτύου [8].

Το αποκορύφωμα αυτής της προς τα εμπρός διάδοσης των δεδομένων παρατηρείται στο στρώμα εξόδου, το οποίο προσαρμόζεται στις ειδικές απαιτήσεις της εκάστοτε εργασίας. Το αν ο στόχος είναι η παλινδρόμηση, η δυαδική ταξινόμηση ή η ταξινόμηση πολλαπλών κλάσεων υπαγορεύει τόσο τον αριθμό των νευρώνων σε αυτό το επίπεδο όσο και την επιλογή της συνάρτησης ενεργοποίησης. Σε προβλήματα ταξινόμησης, για παράδειγμα, χρησιμοποιείται συχνά η συνάρτηση softmax για τη μετατροπή της εξόδου σε πιθανότητες που αθροίζουν στη μονάδα, διευκολύνοντας την πιθανολογική ερμηνεία της εξόδου [72].

Το ταξίδι των δεδομένων από το στρώμα εισόδου, μέσω των κρυφών στρωμάτων, στο



Εικόνα 4.2: Το MLP

στρώμα εξόδου διέπεται από τις αρχές της προς τα εμπρός διάδοσης. Η ικανότητα του MLP να μαθαίνει και να βελτιώνει το μοντέλο του ενισχύεται περαιτέρω μέσω της οπισθοδιάδοσης (backpropagation) και της βαθμωτής καθόδου (gradient descent), όπου οι παράμετροι του δικτύου (βάρη και προκαταλήψεις) προσαρμόζονται επαναληπτικά ώστε να ελαχιστοποιείται η απόκλιση μεταξύ των πραγματικών και των προβλεπόμενων εξόδων. Αυτή η επαναληπτική βελτιστοποίηση καθοδηγείται από μια σαφώς καθορισμένη συνάρτηση απωλειών.

## 4.4 Εκπαίδευση Νευρωνικών Δικτύων

Στην αρχή, η συνάρτηση απώλειας  $L$  ποσοτικοποιεί την απόδοση του δικτύου μετρώντας τη διαφορά μεταξύ των προβλεπόμενων εξόδων ( $y_{pred}$ ) και των πραγματικών εξόδων ( $y_{true}$ ). Για ένα σύνολο δεδομένων με  $n$  καταχωρήσεις, η συνάρτηση απωλειών μπορεί να αναπαρασταθεί ως εξής:

$$L = \frac{1}{n} \sum_{i=1}^n L_i(y_{pred,i}, y_{true,i})$$

όπου  $L_i$  συμβολίζει την απώλεια που σχετίζεται με το  $i$ -οστό σημείο δεδομένων. Οι συνήθειες επιλογές για το  $L_i$  περιλαμβάνουν το μέσο τετραγωνικό σφάλμα (MSE)  $L_i = \frac{1}{2}(y_{pred,i} - y_{true,i})^2$  για εργασίες παλινδρόμησης, και το Cross-Entropy  $L_i = -[y_{true,i} \log(y_{pred,i}) + (1 - y_{true,i}) \log(1 - y_{pred,i})]$  για εργασίες δυαδικής ταξινόμησης [73].

Η οπισθοδιάδοση διευκολύνει τον υπολογισμό του  $\frac{\partial L}{\partial w}$ , της κλίσης της συνάρτησης απωλειών ως προς κάθε βάρος, εφαρμόζοντας τον κανόνα της αλυσίδας. Για ένα δεδομένο βάρος  $w_{jk}^{(l)}$  που συνδέει τον νευρώνα  $k$  στο στρώμα  $l - 1$  με τον νευρώνα  $j$  στο στρώμα  $l$ , η κλίση είναι μέρος μιας αλυσίδας που εκτείνεται από το στρώμα εξόδου πίσω στο στρώμα εισόδου:

$$\frac{\partial L}{\partial w_{jk}^{(l)}} = \frac{\partial L}{\partial a_j^{(l)}} \cdot \frac{\partial a_j^{(l)}}{\partial z_j^{(l)}} \cdot \frac{\partial z_j^{(l)}}{\partial w_{jk}^{(l)}}$$

όπου  $a_j^{(l)}$  είναι η έξοδος ενεργοποίησης του νευρώνα  $j$  στο στρώμα  $l$ , και  $z_j^{(l)}$  είναι η σταθμισμένη είσοδος στο νευρώνα  $j$ . Αυτή η κλίση μας λέει πώς μια μικρή αλλαγή στο  $w_{jk}^{(l)}$  επηρεάζει τη συνολική απώλεια, επιτρέποντας ακριβείς προσαρμογές του βάρους προς την κατεύθυνση που μειώνει την απώλεια.

Ο υπολογισμός ξεκινά από το επίπεδο εξόδου, όπου το σφάλμα της πρόβλεψης του δικτύου είναι άμεσα μετρήσιμο. Ο όρος σφάλματος  $\delta_j^{(L)}$  για έναν νευρώνα  $j$  στο στρώμα εξόδου  $L$  δίνεται από την παράγωγο της συνάρτησης απώλειας ως προς την έξοδο ενεργοποίησης του νευρώνα:

$$\delta_j^{(L)} = \frac{\partial L}{\partial a_j^{(L)}} \cdot \phi'(z_j^{(L)})$$

όπου  $\phi'(z_j^{(L)})$  είναι η παράγωγος της συνάρτησης ενεργοποίησης ως προς τη σταθμισμένη είσοδο  $z_j^{(L)}$ . Ο όρος  $\frac{\partial L}{\partial a_j^{(L)}}$  αντιπροσωπεύει την κλίση της συνάρτησης απωλειών ως προς την ενεργοποίηση του νευρώνα  $j$  και μπορεί να ποικίλλει ανάλογα με τη συνάρτηση απωλειών που χρησιμοποιείται. Για το MSE, για παράδειγμα, αυτή η κλίση μπορεί να εκφραστεί ως  $(a_j^{(L)} - y_j)$  όπου  $y_j$  είναι η πραγματική τιμή-στόχος για τον νευρώνα  $j$ .

Για τα κρυφά στρώματα, ο όρος σφάλματος  $\delta_j^{(l)}$  για κάθε νευρώνα  $j$  στο στρώμα  $l$  υπολογίζεται με τη διάδοση των σφαλμάτων από το στρώμα πάνω από αυτό ( $l + 1$ ) προς τα πίσω:

$$\delta_j^{(l)} = \left( \sum_k \delta_k^{(l+1)} w_{jk}^{(l+1)} \right) \phi'(z_j^{(l)})$$

Αυτή η εξίσωση δείχνει πώς ο όρος σφάλματος για έναν δεδομένο νευρώνα επηρεάζεται από τους όρους σφάλματος των νευρώνων στο επόμενο επίπεδο ( $\delta_k^{(l+1)}$ ) και τα βάρη που τον συνδέουν με αυτούς τους νευρώνες ( $w_{jk}^{(l+1)}$ ), διαμορφωμένα από την παράγωγο της συνάρτησης ενεργοποίησης.

Με τους όρους σφάλματος να έχουν καθοριστεί, οι κλίσεις της συνάρτησης απωλειών σε σχέση με τα βάρη και τις προκαταλήψεις μπορούν να υπολογιστούν ως εξής:

- Για τα βάρη:

$$\frac{\partial L}{\partial w_{jk}^{(l)}} = a_k^{(l-1)} \delta_j^{(l)}$$

- Για τις προκαταλήψεις:

$$\frac{\partial L}{\partial b_j^{(l)}} = \delta_j^{(l)}$$

Αυτές οι κλίσεις αποκαλύπτουν την κατεύθυνση προς την οποία πρέπει να προσαρμοστεί κάθε παράμετρος για να μειωθεί η συνολική απώλεια, χρησιμεύοντας ως βάση για την ενημέρωση των βαρών και των προκαταλήψεων του δικτύου μέσω της καθόδου κλίσης.

Τελικά, τα βάρη και οι προκαταλήψεις ενημερώνονται προς την κατεύθυνση που ελαχιστοποιεί την απώλεια, χρησιμοποιώντας τις υπολογισμένες κλίσεις και έναν προκαθορισμένο ρυθμό μάθησης  $\eta$ :

- Κανόνας ενημέρωσης για τα βάρη:

$$w_{jk}^{(l)} := w_{jk}^{(l)} - \eta \frac{\partial L}{\partial w_{jk}^{(l)}}$$



- Κανόνας ενημέρωσης για τις προκαταλήψεις:

$$b_j^{(l)} := b_j^{(l)} - \eta \frac{\partial L}{\partial b_j^{(l)}}$$

Σε αυτή την εξίσωση,  $\eta$  είναι ο ρυθμός μάθησης, μια κομβική παράμετρος που ελέγχει το μέγεθος του βήματος κατά τη διαδικασία διόρθωσης των βαρών [74].

Οι έννοιες των εποχών και του μεγέθους των δεσμών (batches) είναι ζωτικής σημασίας στη διάθρωση της διαδικασίας εκπαίδευσης [75]. Μια εποχή αντιπροσωπεύει ένα πλήρες πέρασμα σε ολόκληρο το σύνολο δεδομένων εκπαίδευσης, επιτρέποντας στο δίκτυο πολλαπλές ευκαιρίες να μάθει από τα δεδομένα. Το μέγεθος δέσμης αναφέρεται στον αριθμό των δειγμάτων δεδομένων που εισάγονται στο νευρωνικό δίκτυο πριν από την ενημέρωση του μοντέλου.

## 4.5 Πρόβλεψη Χρονοσειρών με MLPs

Τα Multilayer Perceptrons (MLPs), με τη βαθιά αρχιτεκτονική τους από διασυνδεδεμένα στρώματα νευρώνων, αποτελούν ένα ισχυρό εργαλείο για την πρόβλεψη χρονοσειρών [76]. Η πρόβλεψη αυτή περιλαμβάνει την πρόβλεψη μελλοντικών τιμών σε μια ακολουθία τιμών, όπου η εγγενής πολυπλοκότητα και η μη γραμμικότητα των δεδομένων απαιτούν εξειλιγμένες τεχνικές μοντελοποίησης. Τα MLP υπερέρχουν σε αυτόν τον τομέα αξιοποιώντας την ικανότητά τους να συλλαμβάνουν και να μοντελοποιούν μη γραμμικές σχέσεις εντός των δεδομένων, προσφέροντας ακριβείς προβλέψεις.

Ένα MLP που έχει σχεδιαστεί για την πρόβλεψη χρονοσειρών περιλαμβάνει συνήθως ένα στρώμα εισόδου, ένα ή περισσότερα κρυφά στρώματα και ένα στρώμα εξόδου. Το στρώμα εισόδου λαμβάνει χαρακτηριστικά που προέρχονται από τα δεδομένα της χρονοσειράς, συμπεριλαμβανομένων των προηγούμενων τιμών της ίδιας της σειράς και ενδεχομένως εξωγενών μεταβλητών, εάν είναι διαθέσιμες [76].

Εστιάζοντας στην περίπτωση που θα αποτελέσει αντικείμενο της παρούσας εργασίας, θα εξετάσουμε τα είδη νευρωνικών αρχιτεκτονικών που χρησιμοποιούν μόνο παρελθοντικές τιμές της ίδιας χρονοσειράς (univariate time series). Για μια χρονοσειρά  $Y = \{y_1, y_2, \dots, y_t\}$ , γίνεται χρήση ενός ολισθαίνοντος παραθύρου για τη δημιουργία χαρακτηριστικών εισόδου στο νευρωνικό δίκτυο. Δηλαδή, για να προβλέψει την τιμή τη χρονική στιγμή  $t + 1$ , το MLP μπορεί να χρησιμοποιήσει τις  $p$  προηγούμενες τιμές  $[y_t, y_{t-1}, \dots, y_{t-p+1}]$  ως εισόδους. Η διάσταση αυτών των εισόδων υπαγορεύει τον αριθμό των νευρώνων στο επίπεδο εισόδου, που αντιστοιχεί σε  $p$ , στην προκειμένη εξίσωση.

Ο στόχος του MLP σε αυτό το πλαίσιο είναι η προσέγγιση της συνάρτησης  $f$  που απεικονίζει το διάλυσμα εισόδου  $X_t = [y_t, y_{t-1}, \dots, y_{t-p+1}]$  στην έξοδο  $y_{t+1}$ . Αυτή η αντιστοίχιση μπορεί να αναπαρασταθεί ως εξής:

$$\hat{y}_{t+1} = f(X_t; \Theta)$$

όπου  $\hat{y}_{t+1}$  δηλώνει την προβλεπόμενη τιμή για τη χρονική στιγμή  $t + 1$ , και  $\Theta$  συμβολίζει τις παραμέτρους του MLP, που περιλαμβάνουν τόσο τα βάρη όσο και τις προκαταλήψεις.

Η εκπαίδευση περιλαμβάνει την ελαχιστοποίηση μιας συνάρτησης απωλειών, η οποία στις προβλέψεις χρονοσειρών είναι συχνά το μέσο τετραγωνικό σφάλμα (MSE), που ορίζεται ως εξής:

$$MSE = \frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2$$

Χρησιμοποιώντας οπισθοδιάδοση, οι παράμετροι  $\Theta$  του MLP προσαρμόζονται επαναληπτικά για να μειώσουν το MSE, βελτιώνοντας την ικανότητα πρόβλεψης του δικτύου.

Όταν ο στόχος επεκτείνεται στην πρόβλεψη πολλαπλών βημάτων μπροστά, η δομή του στρώματος εξόδου υφίσταται μία τροποποίηση. Αντί να προβλέπει μια ενιαία μελλοντική τιμή  $y_{t+1}$ , το δίκτυο στοχεύει στην πρόβλεψη μιας ακολουθίας μελλοντικών τιμών  $[y_{t+1}, y_{t+2}, \dots, y_{t+m}]$ , όπου  $m$  είναι ο αριθμός των βημάτων μπροστά. Κατά συνέπεια, το επίπεδο εξόδου πρέπει να έχει  $m$  νευρώνες, καθένας από τους οποίους αντιστοιχεί σε ένα από τα  $m$  μελλοντικά χρονικά βήματα που προβλέπονται. Η έξοδος του δικτύου μπορεί επομένως να αναπαρασταθεί ως εξής:

$$[\hat{y}_{t+1}, \hat{y}_{t+2}, \dots, \hat{y}_{t+m}] = f(X_t; \Theta)$$

Αυτή η αλλαγή απαιτεί ένα πιο πολύπλοκο σχήμα εκπαίδευσης και υπολογισμού απωλειών, καθώς το δίκτυο ελαχιστοποιεί τώρα το σφάλμα σε όλα τα προβλεπόμενα χρονικά βήματα, χρησιμοποιώντας ενδεχομένως μια συνάρτηση απωλειών που αθροίζεται ή υπολογίζεται κατά μέσο όρο στις εξόδους  $m$ .

Τα MLP προσφέρουν ένα στιβαρό πλαίσιο για την πρόβλεψη χρονοσειρών, ικανό να αντιμετωπίσει τόσο τις προβλέψεις ενός βήματος όσο και τις προβλέψεις πολλαπλών βημάτων μέσω της προσαρμοσμένης αρχιτεκτονικής τους. Με την προσαρμογή των χαρακτηριστικών εισόδου και της δομής του επιπέδου εξόδου, τα MLP μπορούν να προσαρμοστούν σε ένα ευρύ φάσμα εργασιών πρόβλεψης [77], από βραχυπρόθεσμες προβλέψεις έως μακροπρόθεσμες προβολές. Το κλειδί της αποτελεσματικότητάς τους έγκειται στην ικανότητά τους να μαθαίνουν τα υποκείμενα μοτίβα στα δεδομένα χρονοσειρών, η οποία διευκολύνεται από την περίπλοκη διαδικασία της οπισθοδιάδοσης. Μέσω προσεκτικού σχεδιασμού, εκπαίδευσης και συντονισμού των παραμέτρων, τα MLP μπορούν έτσι να παρέχουν πολύτιμες προβλέψεις, βοηθώντας στη λήψη αποφάσεων σε διάφορους τομείς.

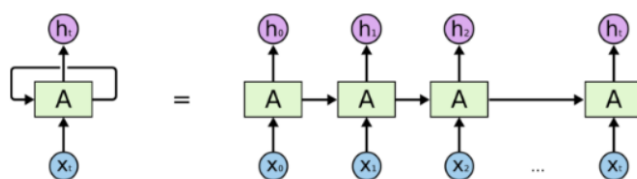
Ωστόσο, ενώ τα MLP προσφέρουν ένα ισχυρό πλαίσιο για ένα ευρύ φάσμα εργασιών, πολλές φορές φαίνεται να μην είναι αρκετά αποδοτικά σε προβλήματα με σειριακά δεδομένα, όπως αυτό της πρόβλεψης χρονοσειρών, γεγονός που υπογραμμίζει τη σημασία της διερεύνησης και άλλων νευρωνικών αρχιτεκτονικών.

## 4.6 RNN

### 4.6.1 Αρχιτεκτονική

Τα αναδρομικά νευρωνικά δίκτυα (RNN) είναι μια κατηγορία τεχνητών νευρωνικών δικτύων που έχουν σχεδιαστεί για να αναγνωρίζουν μοτίβα σε ακολουθιακά δεδομένα, όπως κείμενο, γονιδιώματα ή αριθμητικά δεδομένα χρονοσειρών. Σε αντίθεση με τα παραδοσιακά νευρωνικά δίκτυα (MLPs), τα RNN αξιοποιούν τη διαδοχική φύση της εισόδου τους, επιτρέπο-

ντάς τους να διατηρήσουν μια μορφή μνήμης χρησιμοποιώντας την έξοδο τους ως είσοδο για το επόμενο βήμα της υπό εξέταση αλληλουχίας.



Εικόνα 4.3: Η γενική δομή των RNN

Στον πυρήνα του, ένα RNN επεξεργάζεται τις ακολουθίες διατρέχοντας τα στοιχεία της ακολουθίας και διατηρώντας ένα διάνυσμα κατάστασης που ενσωματώνει πληροφορίες για τα στοιχεία που έχει δει μέχρι τώρα. Οι θεμελιώδεις διαδικασίες σε ένα RNN για ένα δεδομένο χρονικό βήμα  $t$  μπορούν να περιγραφούν ως εξής [78]:

1. **Ενημέρωση κατάστασης:** Η κρυφή κατάσταση  $h_t$  τη στιγμή  $t$  ενημερώνεται με βάση την προηγούμενη κρυφή κατάσταση  $h_{t-1}$  και την είσοδο στο τρέχον βήμα  $x_t$ . Αυτή η ενημέρωση είναι συνήθως μια μη γραμμική συνάρτηση  $f$ :

$$h_t = f(W_{hh}h_{t-1} + W_{xh}x_t + b_h)$$

Εδώ, το  $W_{hh}$  είναι ο πίνακας βαρών που συνδέει την προηγούμενη κρυφή κατάσταση  $h_{t-1}$  με την τρέχουσα κρυφή κατάσταση  $h_t$ ,  $W_{xh}$  είναι πάλι τα βάρη που συνδέουν αυτή τη φορά την είσοδο  $x_t$  με την κρυφή κατάσταση και το  $b_h$  είναι ένας όρος μεροληψίας. Η συνάρτηση  $f$  είναι συχνά μια συνάρτηση tanh ή σιγμοειδής, παρέχοντας στο μοντέλο μη γραμμικές δυνατότητες.

2. **Δημιουργία εξόδου:** Η έξοδος  $y_t$  τη στιγμή  $t$  δημιουργείται με βάση την τρέχουσα κρυφή κατάσταση  $h_t$ . Αυτό περιλαμβάνει πάλι έναν μετασχηματισμό, ακολουθούμενο από μια συνάρτηση ενεργοποίησης:

$$y_t = g(W_{hy}h_t + b_y)$$

Σε αυτήν την εξίσωση, το  $W_{hy}$  είναι ο πίνακας βαρών που συνδέει την κρυφή κατάσταση  $h_t$  στην έξοδο  $y_t$  και ο  $b_y$  είναι όρος μεροληψίας. Η συνάρτηση  $g$  θα μπορούσε να είναι μια συνάρτηση softmax.

Η επαναληπτική ενημέρωση της κρυφής κατάστασης επιτρέπει στα RNN να διατηρούν, θεωρητικά, πληροφορίες από όλες τις προηγούμενες εισόδους στην ακολουθία. Ωστόσο, στην πράξη, τα τυπικά RNN αντιμετωπίζουν συχνά προκλήσεις στη διατήρηση μακροπρόθεσμων εξαρτήσεων λόγω ζητημάτων όπως η εξαφάνιση και η έκρηξη των κλίσεων κατά τη διάρκεια της εκπαίδευσης με οπισθοδιάδοση στο χρόνο (BPTT).

#### 4.6.2 Εκπαίδευση RNN (BPTT)

Η εκπαίδευση των RNN γίνεται μέσω του αλγορίθμου BPTT [79], μια επέκταση του backpropagation που παρουσιάστηκε νωρίτερα για τα MLPs. Δεδομένης μιας ακολουθίας,

οι έξοδοι του δικτύου συγκρίνονται με την επιθυμητή έξοδο και υπολογίζεται μια απώλεια σε κάθε βήμα. Η συνολική απώλεια είναι τότε το άθροισμα αυτών των απωλειών. Οι παράγωγοι αυτής της συνάρτησης απώλειας υπολογίζονται στη συνέχεια σε σχέση με κάθε παράμετρο στο δίκτυο. Λόγω της διαδοχικής φύσης του RNN, αυτές οι παράγωγοι θα περιλαμβάνουν όρους από κάθε χρονικό βήμα, καθιστώντας το υπολογιστικό γράφημα πολύ μεγαλύτερο από αυτό ενός τυπικού MLP.

Πιο συγκεκριμένα, ο BPTT υπολογίζει τις παραγώγους μιας συνάρτησης απώλειας  $L$  που μετρά την απόκλιση μεταξύ των προβλεπόμενων εξόδων  $Y$  και των πραγματικών τιμών στόχου  $Y^*$ . Για απλότητα, μπορούμε να θεωρήσουμε τη συνάρτηση απώλειας ως το άθροισμα των απωλειών σε όλα τα χρονικά βήματα:

$$L = \sum_{t=1}^T L_t(y_t, y_t^*)$$

Για να ενημερωθούν οι παράμετροι του δικτύου ( $W_{hh}$ ,  $W_{xh}$ ,  $W_{hy}$ ,  $b_h$ ,  $b_y$ ), ο BPTT υπολογίζει τη μερική παράγωγο του  $L$  σε σχέση με αυτές τις παραμέτρους.

Για κάθε παράμετρο, η κλίση περιλαμβάνει την άθροιση όλων των χρονικών βημάτων, επειδή η απώλεια σε κάθε χρονικό βήμα δυνητικά επηρεάζει όλα τα επόμενα χρονικά βήματα. Για παράδειγμα, η παράγωγος του  $L$  ως προς το  $W_{hy}$  υπολογίζεται εφαρμόζοντας τον κανόνα της αλυσίδας:

$$\frac{\partial L}{\partial W_{hy}} = \sum_{t=1}^T \frac{\partial L_t}{\partial y_t} \frac{\partial y_t}{\partial W_{hy}}$$

Εφόσον  $y_t = g(W_{hy}h_t + b_y)$ , μπορούμε να επεκτείνουμε περαιτέρω το  $\frac{\partial y_t}{\partial W_{hy}}$  με βάση τη συγκεκριμένη μορφή του  $g$ .

Οι όροι σφάλματος διαδίδονται προς τα πίσω από το επίπεδο εξόδου στο επίπεδο εισόδου. Η παράγωγος σφάλματος ως προς την κρυφή κατάσταση τη στιγμή  $t$ ,  $\frac{\partial L}{\partial h_t}$ , επηρεάζεται από το σφάλμα τη στιγμή  $t$  και το σφάλμα που διαδόθηκε από το χρόνο  $t + 1$  [80]:

$$\frac{\partial L}{\partial h_t} = \left( \frac{\partial L_t}{\partial y_t} \frac{\partial y_t}{\partial h_t} \right) + \left( \frac{\partial L}{\partial h_{t+1}} \frac{\partial h_{t+1}}{\partial h_t} \right)$$

Αυτή η εξίσωση υπογραμμίζει την αναδρομική φύση του BPTT, όπου η κλίση σε κάθε βήμα εξαρτάται από τις διαβαθμίσεις σε μελλοντικά βήματα.

Μετά τον υπολογισμό των παραγώγων, οι παράμετροι ενημερώνονται κατά τα γνωστά με τη χρήση gradient descent ή μίας από τις παραλλαγές του:

$$W := W - \eta \frac{\partial L}{\partial W}$$

### 4.6.3 LSTMs

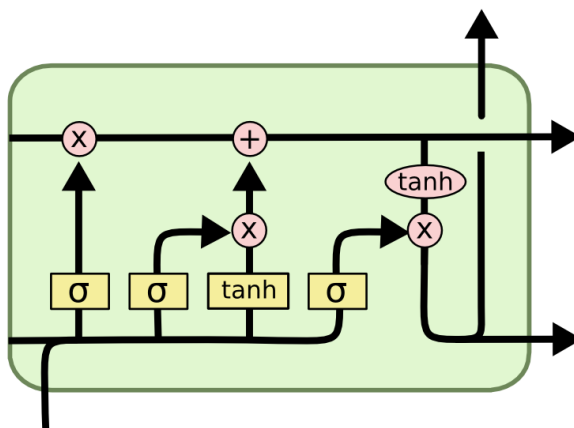
Τα δίκτυα Long Short-Term Memory (LSTM), ένα ειδικό είδος αναδρομικού νευρωνικού δικτύου (RNN), σχεδιάστηκαν για να ξεπεράσουν τους περιορισμούς των παραδοσιακών RNN, ειδικά τη δυσκολία τους στην εκμάθηση μακροπρόθεσμων εξαρτήσεων. Τα LSTM το επιτυγχάνουν αυτό εισάγοντας μια πιο σύνθετη υπολογιστική μονάδα: τη μονάδα LSTM

(LSTM cell), η οποία περιλαμβάνει μηχανισμούς που διευκολύνουν την ύπαρξη μνήμης [81]. Αυτό τους επιτρέπει να θυμούνται πληροφορίες για μεγάλες περιόδους, καθιστώντας τις ιδιαίτερα κατάλληλες για σειριακά δεδομένα μεγάλου μήκους όπως αυτά που εμφανίζονται στην πρόβλεψη χρονοσειρών, στην επεξεργασία φυσικής γλώσσας και άλλα [82].

Η μονάδα LSTM αποτελείται από κάποια βασικά στοιχεία: την παρούσα κατάσταση μονάδας, την πύλη εισόδου, την πύλη απώλειας μνήμης και την πύλη εξόδου. Αυτά τα στοιχεία συνεργάζονται για να ρυθμίσουν τη ροή πληροφοριών μέσα και έξω από την μονάδα, καθώς και για να ενημερώσουν την εσωτερική κατάσταση της μονάδας [83].

Η κατάσταση μονάδας  $C_t$  είναι το κεντρικό στοιχείο της μονάδας LSTM, που λειτουργεί ως μεταφορικός ιμάντας για τη μεταφορά σχετικών πληροφοριών σε όλη την ακολουθία δεδομένων. Υποβάλλεται σε αλλαγές σε κάθε χρονικό βήμα, επιτρέποντας στις πληροφορίες να ρέουν κατά μήκος του αμετάβλητες εάν χρειάζεται.

Οι πύλες είναι ένας τρόπος για να περάσουν προαιρετικά πληροφορίες. Αποτελούνται από ένα στρώμα νευρώνων και μια πράξη πολλαπλασιασμού κατά σημείο. Η σιγμοειδής συνάρτηση ενεργοποίησης που πλαισιώνει το στρώμα νευρώνων εξάγει αριθμούς μεταξύ μηδέν και ενός, περιγράφοντας πόσο από κάθε στοιχείο πρέπει να περάσει. Σε μία μονάδα LSTM, υπάρχουν τρεις τέτοιες πύλες:



Εικόνα 4.4: Η δομική μονάδα των LSTM

1. **Πύλη απώλειας μνήμης ( $f_t$ ):** Αποφασίζει ποιες πληροφορίες θα απορριφθούν από την παρούσα κατάσταση μονάδας. Εξετάζει το  $h_{t-1}$  (προηγούμενη έξοδος) και το  $x_t$  (τρέχουσα είσοδος) και εξάγει έναν αριθμό μεταξύ 0 και 1 για κάθε αριθμό στην κατάσταση μονάδας  $C_{t-1}$ .

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

2. **Πύλη εισόδου ( $i_t$ ) και Υποψήφιες τιμές ( $\tilde{C}_t$ ):** Προσδιορίζει ποιες νέες πληροφορίες θα αποθηκευτούν στην κατάσταση μονάδας. Η πύλη εισόδου αποφασίζει ποιες τιμές θα ενημερωθούν και δημιουργείται ένα διάνυσμα νέων υποψήφιων τιμών που θα

μπορούσαν να προστεθούν στην κατάσταση.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

3. **Πύλη εξόδου ( $o_t$ ):** Καθορίζει ποια θα πρέπει να είναι η επόμενη κρυφή κατάσταση  $h_t$  (η επόμενη έξοδος). Η κρυφή κατάσταση φιλτράρεται επίσης από την πύλη εξόδου, καθορίζοντας την επόμενη έξοδο.

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

Η κατάσταση της μονάδας ενημερώνεται σε δύο βήματα. Πρώτα, η παλιά κατάσταση  $C_{t-1}$  πολλαπλασιάζεται επί  $f_t$ , ξεχνώντας τα πράγματα που αποφασίστηκε να ξεχαστούν νωρίτερα. Στη συνέχεια, προσθέτει  $i_t * \tilde{C}_t$ , που είναι οι νέες υποψήφιας τιμές, οι οποίες κλιμακώνονται ανάλογα με το πόσο αποφασίσαμε να ενημερώσουμε κάθε τιμή κατάστασης.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t$$

Τα LSTM έχουν εφαρμοστεί σε ένα ευρύ φάσμα εργασιών με μεγάλη επιτυχία. Ωστόσο, η απόδοσή τους εξαρτάται σε μεγάλο βαθμό από την επιλογή των υπερπαραμέτρων, την αρχιτεκτονική του δικτύου (όπως ο αριθμός των επιπέδων LSTM και το μέγεθος κάθε επιπέδου) καθώς και τις ιδιαιτερότητες της κάθε εργασίας. Επιπλέον, ενώ τα LSTM αντιμετωπίζουν το ζήτημα των μακροπρόθεσμων εξαρτήσεων καλύτερα από τα παραδοσιακά RNN, μπορούν ακόμα να αντιμετωπίσουν προβλήματα με πολύ μεγάλες ακολουθίες ή με ακολουθίες όπου οι σχετικές πληροφορίες είναι εξαιρετικά διασκορπισμένες [84]. Προηγμένες τεχνικές και παραλλαγές, όπως μηχανισμοί προσοχής, έχουν εισαχθεί για την αντιμετώπιση ορισμένων από αυτές τις προκλήσεις, προσφέροντας ένα διαφορετικό έλεγχο στη ροή πληροφοριών εντός του δικτύου.

#### 4.6.4 LSTMs στην Πρόβλεψη Χρονοσειρών

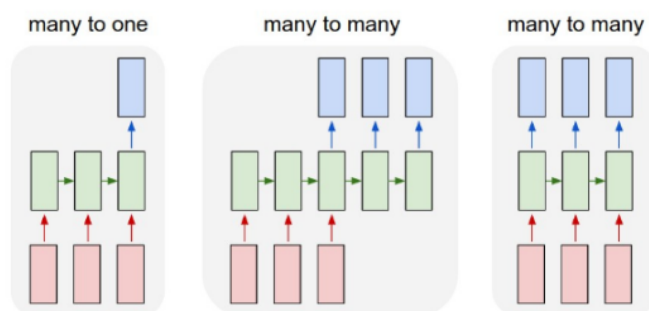
Στην πρόβλεψη χρονοσειρών, υπό το πρίσμα που αντιμετωπίζεται αυτό το πρόβλημα στην παρούσα εργασία, ο στόχος είναι η πρόβλεψη μελλοντικών τιμών με βάση προηγούμενες παρατηρήσεις. Τα LSTM υπερέρχουν σε αυτό το έργο καθώς επεξεργάζονται αλληλουχίες δεδομένων και μαθαίνουν από τη χρονική δυναμική και τις εξαρτήσεις εντός της σειράς [85].

Ένα μοντέλο LSTM για πρόβλεψη χρονοσειρών αποτελείται γενικά από ένα ή περισσότερα επίπεδα LSTM που ακολουθούνται από ένα ή περισσότερα πυκνά στρώματα νευρώνων για την έξοδο των προβλεπόμενων τιμών. Η είσοδος στο μοντέλο είναι μια ακολουθία προηγούμενων παρατηρήσεων και η έξοδος είναι η πρόβλεψη μελλοντικών τιμών.

Το μοντέλο LSTM εκπαιδεύεται σε ιστορικά δεδομένα χρονοσειρών. Το δίκτυο μαθαίνει να ελαχιστοποιεί μια συνάρτηση απώλειας, συνήθως το Μέσο Τετραγωνικό Σφάλμα (MSE) που μετρά τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών μελλοντικών τιμών.

Το Backpropagation Through Time (BPTT) χρησιμοποιείται για τον υπολογισμό των παραγώγων της συνάρτησης απώλειας που είναι απαραίτητες για την ενημέρωση των βαρών του μοντέλου.

Μόλις εκπαιδευτεί, το μοντέλο LSTM μπορεί να χρησιμοποιηθεί για να κάνει προβλέψεις. Για πρόβλεψη ενός βήματος, το μοντέλο προβλέπει την επόμενη τιμή στην ακολουθία. Για την πρόβλεψη πολλαπλών βημάτων, το μοντέλο μπορεί είτε να εκπαιδευτεί ώστε να προβλέπει πολλαπλά βήματα ταυτόχρονα είτε να χρησιμοποιήσει μια επαναληπτική προσέγγιση όπου οι προβλέψεις του μοντέλου ανατροφοδοτούνται ως εισροές για περαιτέρω πρόβλεψη στο μέλλον.



Εικόνα 4.5: Πιθανές έξοδοι ενός επιπέδου LSTM

Ανάμεσα στα προτερήματα των LSTM στην πρόβλεψη χρονοσειρών συγκαταλέγεται και η ευελιξία τους. Τα LSTM μπορούν να χειριστούν χρονοσειρές τόσο μίας μεταβλητής όσο και πολλαπλών μεταβλητών [86]. Παράλληλα, ενώ τα LSTM προσφέρουν σημαντικά πλεονεκτήματα για την πρόβλεψη χρονοσειρών, η εφαρμογή τους συνοδεύεται από προκλήσεις. Η αρχιτεκτονική και οι υπερπαραμέτροι του μοντέλου LSTM, όπως ο αριθμός των μονάδων και των επιπέδων LSTM, ο ρυθμός εκμάθησης και το μήκος ακολουθίας, χρειάζονται προσεκτικό συντονισμό για την επίτευξη βέλτιστης απόδοσης. Η υπερπροσαρμογή είναι μια άλλη ανησυχία, ιδιαίτερα με μικρότερα σύνολα δεδομένων [87].

Επιπλέον, τα LSTM απαιτούν σημαντικούς υπολογιστικούς πόρους για εκπαίδευση, ειδικά με μεγάλα σύνολα δεδομένων. Παρά αυτές τις προκλήσεις, τα οφέλη των LSTM, ιδιαίτερα η ικανότητά τους να μαθαίνουν και να θυμούνται σε μεγάλες ακολουθίες, τα καθιστούν ένα ισχυρό εργαλείο για την πρόβλεψη χρονοσειρών, παρέχοντας προβλέψεις που οδηγούν στη λήψη αποφάσεων σε όλους τους κλάδους.

## 4.7 CNN

### 4.7.1 Αρχιτεκτονική

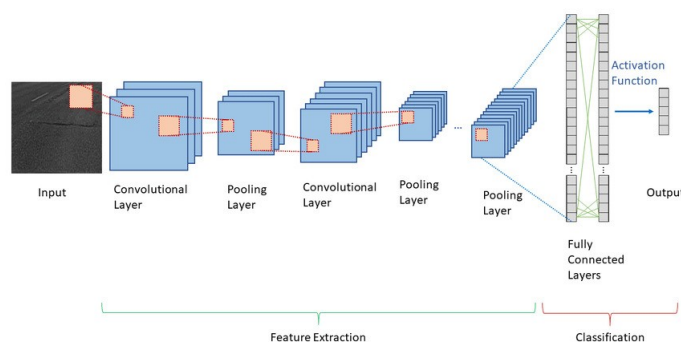
Τα συνελικτικά νευρωνικά δίκτυα (CNN) είναι μια κατηγορία νευρωνικών δικτύων, που χρησιμοποιούνται κυρίως στον τομέα της όρασης υπολογιστών, ωστόσο η εφαρμογή τους επεκτείνεται και σε άλλους τομείς όπως η πρόβλεψη χρονοσειρών [88]. Η θεμελιώδης καινοτομία των CNN έγκειται στην ικανότητά τους να μαθαίνουν μέσω της χρήσης συνελίξεων λαμβάνοντας ως είσοδο ακολουθίες.

Το βασικό συστατικό των CNN είναι το συνελικτικό επίπεδο που εφαρμόζει μια συνελικτι-

κή διαδικασία στην είσοδο, περνώντας το αποτέλεσμα στο επόμενο επίπεδο. Για μια δεδομένη εικόνα εισόδου ή ακολουθία  $X$  και ένα φίλτρο (κερνελ)  $F$  μεγέθους  $K \times K$ , η συνελικτική διαδικασία σε μια θέση  $(i, j)$  ορίζεται ως:

$$(X * F)(i, j) = \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} X(i + m, j + n) \cdot F(m, n)$$

όπου το  $*$  υποδηλώνει τη συνέλιξη. Αυτή η διαδικασία επαναλαμβάνεται για κάθε θέση πάνω από τον πίνακα εισόδου (εικόνα) σύροντας το φίλτρο κατά μήκος της εισόδου, δημιουργώντας έναν χάρτη χαρακτηριστικών.



Εικόνα 4.6: Τυπική αρχιτεκτονική συνελικτικού νευρωνικού δικτύου

Στην πράξη, τα φίλτρα στα CNN μαθαίνονται από τα δεδομένα κατά τη διάρκεια της εκπαιδευτικής διαδικασίας. Ξεκινώντας από την τυχαία αρχικοποίηση, οι τιμές μέσα σε κάθε φίλτρο προσαρμόζονται μέσω backpropagation για να ελαχιστοποιηθεί η συνάρτηση απώλειας του δικτύου. Έτσι, το CNN μαθαίνει αυτόματα τα φίλτρα που είναι πιο χρήσιμα για τη συγκεκριμένη εργασία.

Τα CNN συχνά περιλαμβάνουν επίπεδα συγκέντρωσης (pooling layers) για να μειώσουν τη διάσταση των χαρτών χαρακτηριστικών. Αυτή η μείωση διαστάσεων όχι μόνο μειώνει τον υπολογιστικό φόρτο και τον αριθμό των παραμέτρων στο δίκτυο, καθιστώντας το λιγότερο επιρρεπές σε υπερπροσαρμογή, αλλά βοηθά επίσης να γίνει η αναπαράσταση πιο αφηρημένη διατηρώντας τις πιο βασικές πληροφορίες [89]. Υπάρχουν διάφοροι τύποι ομαδοποίησης, με το Max Pooling και το Average Pooling να είναι οι πιο διαδεδομένοι. Κάθε τύπος στρώματος συγκέντρωσης λειτουργεί με έναν ξεχωριστό τρόπο, όπως θα φανεί παρακάτω [90]:

- **Max Pooling:** Το Max Pooling περιλαμβάνει την ολίσηση ενός παραθύρου (ή φίλτρου) πάνω από τα δεδομένα εισόδου του στρώματος και τη λήψη της μέγιστης τιμής μέσα στο παράθυρο σε κάθε βήμα. Για μια διδιάστατη είσοδο  $X$  με διαστάσεις  $M \times N$ , ένα max pooling layer με μέγεθος παραθύρου  $k \times k$  μπορεί να οριστεί μαθηματικά ως εξής: Για κάθε υποπεριοχή  $S$  από την είσοδο  $X$  που καλύπτεται από το παράθυρο,

$$P(S) = \max(S)$$

όπου το  $P(S)$  αντιπροσωπεύει τη συγκεντρωτική έξοδο για την υποπεριοχή  $S$  και το  $\max(S)$  υπολογίζει τη μέγιστη τιμή εντός του παραθύρου. Το παράθυρο μετακινείται



κατά μήκος της μήτρας εισόδου με ένα συγκεκριμένο βήμα, μειώνοντας τη διάσταση της εξόδου.

- **Average Pooling:** Το Average Pooling, σε αντίθεση με το Max Pooling, υπολογίζει τον μέσο όρο των στοιχείων μέσα στο παράθυρο για κάθε υποπεριοχή της εισόδου. Με δεδομένη την ίδια είσοδο  $X$  και ένα παράθυρο μεγέθους  $k \times k$ , η λογική του Average Pooling ορίζεται ως: Για κάθε υποπεριοχή  $S$  της εισόδου  $X$ ,

$$P(S) = \frac{1}{k^2} \sum_{i \in S} x_i$$

όπου  $P(S)$  είναι η ομαδοποιημένη έξοδος για την υποπεριοχή  $S$ ,  $x_i$  είναι τα στοιχεία εντός του  $S$  και  $k^2$  αντιπροσωπεύει το συνολικό αριθμό στοιχείων στο παράθυρο. Αυτή η λειτουργία μειώνει αποτελεσματικά την είσοδο υπολογίζοντας τον μέσο όρο στο παράθυρο, οδηγώντας σε μια πιο ομαλή αναπαράσταση της εισόδου.

- **Global Pooling:** Το Global Pooling μετατρέπει κάθε χάρτη χαρακτηριστικών σε μία μόνο τιμή. Αυτό επιτυγχάνεται με την εφαρμογή είτε της μέγιστης είτε της μέσης λειτουργίας σε ολόκληρο τον χάρτη χαρακτηριστικών. Μαθηματικά, για έναν χάρτη χαρακτηριστικών  $F$  διαστάσεων  $M \times N$ :

- Global Max Pooling:

$$P(F) = \max(F)$$

- Global Average Pooling:

$$P(F) = \frac{1}{M \times N} \sum_{i=1}^M \sum_{j=1}^N f_{ij}$$

όπου  $P(F)$  είναι η ομαδοποιημένη έξοδος για τον χάρτη χαρακτηριστικών  $F$  και  $f_{ij}$  αντιπροσωπεύει τα μεμονωμένα στοιχεία του  $F$ .

- **Lp Pooling:** Το Lp Pooling είναι μια γενίκευση του Max Pooling και του Average Pooling. Για μέγεθος παραθύρου  $k \times k$ , η λειτουργία του ορίζεται ως: Για κάθε υποπεριοχή  $S$  του  $X$ ,

$$P(S) = \left( \frac{1}{k^2} \sum_{i \in S} |x_i|^p \right)^{\frac{1}{p}}$$

όπου  $p > 0$ . Όταν  $p = 2$ , αντιστοιχεί σε μια νόρμα  $L_2$ , υπολογίζοντας ουσιαστικά το Ευκλείδειο μέτρο μέσα στο παράθυρο. Καθώς το  $p$  πλησιάζει το άπειρο, το Lp Pooling πλησιάζει το Max Pooling.

Κάθε τύπος ομαδοποίησης προσφέρει διαφορετικό τρόπο σύνοψης των πληροφοριών στους χάρτες χαρακτηριστικών εισόδου. Η επιλογή της κατάλληλης ομαδοποίησης μπορεί να επηρεάσει την απόδοση του CNN σε συγκεκριμένες εργασίες, επηρεάζοντας τόσο την ακρίβεια όσο και την υπολογιστική απόδοση του δικτύου [91, 92, 93]. Με την επιλεκτική μείωση

δειγματοληψίας των δεδομένων εισόδου, τα επίπεδα συγκέντρωσης συμβάλλουν στην ικανότητα εξαγωγής χαρακτηριστικών των CNN, επιτρέποντάς τους να επιτύχουν ανθεκτικότητα σε παραλλαγές και μετασχηματισμούς στα δεδομένα εισόδου.

Μετά από πολλά συνελικτικά στρώματα και ομαδοποιήσεις, τα CNN συνήθως καταλήγουν με ένα ή περισσότερα πλήρως συνδεδεμένα στρώματα, παρόμοια με ένα παραδοσιακό Multilayer Perceptron (MLP).

Αυτά τα επίπεδα έχουν σχεδιαστεί για να εκτελούν συλλογισμούς υψηλού επιπέδου με βάση τα χαρακτηριστικά που εξήχθησαν και συγκεντρώθηκαν από τα προηγούμενα επίπεδα.

Σε ένα Συνελικτικό Νευρωνικό Δίκτυο (CNN), οι παράμετροι που μεταβάλλονται κατά την εκπαίδευση (learnable parameters), που συχνά αναφέρονται ως βάρη, βρίσκονται κυρίως σε δύο τύπους επιπέδων: στα συνελικτικά στρώματα και στα πλήρως συνδεδεμένα στρώματα. Αυτές οι παράμετροι προσαρμόζονται κατά τη διάρκεια της διαδικασίας εκπαίδευσης για να ελαχιστοποιηθεί η συνάρτηση απώλειας του δικτύου.

#### 4.7.2 Εκπαίδευση CNN

Τα CNN εκπαιδεύονται χρησιμοποιώντας backpropagation, παρόμοια με τα άλλα νευρωνικά δίκτυα που έχουν ήδη παρουσιαστεί, όπου ο στόχος είναι να ελαχιστοποιηθεί μια συνάρτηση απώλειας. Όπως έχει αναφερθεί προηγουμένως, η συγκεκριμένη διαδικασία περιλαμβάνει δύο κύριες φάσεις: ένα πέρασμα προς τα εμπρός, όπου τα δεδομένα εισόδου περνούν μέσα από το δίκτυο για τη δημιουργία προβλέψεων, και ένα πέρασμα προς τα πίσω, όπου υπολογίζεται η κλίση της συνάρτησης απώλειας σε σχέση με κάθε παράμετρο (βάρος και προκατάληψη) και οι παράμετροι ενημερώνονται ανάλογα. Η μόνη πτυχή που διαφοροποιείται λίγο από τη διαδικασία της οπισθοδιάδοσης στα MLPs που περιγράφηκε προηγουμένως είναι η οπίσθια διάδοση μέσω των συνελικτικών στρωμάτων [94]. Ως εκ τούτου, θα επικεντρωθούμε σε αυτήν την πτυχή.

Η οπισθοδιάδοση μέσω συνελικτικών επιπέδων είναι μια κρίσιμη διαδικασία που επιτρέπει στα Συνελικτικά Νευρωνικά Δίκτυα (CNN) να μάθουν τα βέλτιστα βάρη για τα φίλτρα τους, επιτρέποντάς τους έτσι να εξάγουν σημαντικά χαρακτηριστικά από δεδομένα εισόδου.

Κατά τη διάρκεια του περάσματος προς τα εμπρός, ένα φίλτρο  $F$  με βάρη  $W$  συνελίσσεται πάνω στα δεδομένα εισόδου  $X$  για να παραχθεί ένας χάρτης χαρακτηριστικών εξόδου  $O$ . Για μια δεδομένη θέση  $(i, j)$ , η λειτουργία συνέλιξης μπορεί να αναπαρασταθεί μαθηματικά ως:

$$O(i, j) = (X * F)(i, j) + b = \sum_m \sum_n X(i + m, j + n) \cdot W(m, n) + \beta$$

όπου  $b$  είναι ο όρος μεροληψίας και τα  $m, n$  διατρέχουν τις διαστάσεις του φίλτρου  $F$ .

Η οπίσθια διάδοση σε συνελικτικά επίπεδα περιλαμβάνει τον υπολογισμό των κλίσεων της συνάρτησης απώλειας  $L$  σε σχέση με τα βάρη του φίλτρου  $W$  και τις προκαταλήψεις  $b$ , που συμβολίζονται ως  $\frac{\partial L}{\partial W}$  και  $\frac{\partial L}{\partial b}$ , αντίστοιχα. Αυτές οι παράγωγοι χρησιμοποιούνται στη συνέχεια για την ενημέρωση των παραμέτρων του φίλτρου.

Η κλίση της απώλειας  $L$  σε σχέση με ένα συγκεκριμένο βάρος  $W(m, n)$  στο φίλτρο μπορεί να βρεθεί εφαρμόζοντας τον κανόνα της αλυσίδας:

$$\frac{\partial L}{\partial W(m, n)} = \sum_i \sum_j \frac{\partial L}{\partial O(i, j)} \cdot \frac{\partial O(i, j)}{\partial W(m, n)}$$

Μπορούμε να εκφράσουμε το  $\frac{\partial O(i,j)}{\partial W(m,n)}$  ως:

$$\frac{\partial O(i,j)}{\partial W(m,n)} = X(i+m, j+n)$$

Έτσι, η κλίση της απώλειας ως προς το  $W(m,n)$  γίνεται:

$$\frac{\partial L}{\partial W(m,n)} = \sum_i \sum_j \frac{\partial L}{\partial O(i,j)} \cdot X(i+m, j+n)$$

Η κλίση της συνάρτησης απώλειας ως προς την προκατάληψη  $b$  είναι απλούστερη, καθώς ο όρος μεροληψίας προστίθεται ομοιόμορφα σε ολόκληρο τον χάρτη χαρακτηριστικών:

$$\frac{\partial L}{\partial b} = \sum_i \sum_j \frac{\partial L}{\partial O(i,j)}$$

Δεδομένου ότι η μεροληψία  $b$  επηρεάζει εξίσου κάθε έξοδο  $O(i,j)$ , η κλίση  $\frac{\partial L}{\partial b}$  είναι το άθροισμα των κλίσεων της συνάρτησης απώλειας ως προς το χάρτη χαρακτηριστικών  $O$ , σε όλες τις θέσεις.

Αφού υπολογιστούν οι παράγωγοι, τα βάρη και οι προκαταλήψεις του φίλτρου μπορούν να ενημερωθούν χρησιμοποιώντας την λογική του gradient descent:

$$W_{new}(m,n) = W_{old}(m,n) - \eta \frac{\partial L}{\partial W(m,n)}$$

$$b_{new} = b_{old} - \eta \frac{\partial L}{\partial b}$$

όπου το  $\eta$  αντιπροσωπεύει το ρυθμό μάθησης.

Τα συνελκτικά νευρωνικά δίκτυα (CNN), παραδοσιακά γνωστά για την ικανότητά τους στην επεξεργασία χωρικών δεδομένων όπως εικόνων, έχουν επίσης αποδειχθεί αποτελεσματικά στο χειρισμό άλλων μορφών σειριακών δεδομένων, συμπεριλαμβανομένης της πρόβλεψης μονομεταβλητών χρονοσειρών [95]. Αυτή η εφαρμογή αξιοποιεί την ικανότητα του CNN να εξάγει μοτίβα ή χαρακτηριστικά από δεδομένα, όπου οι χρονικές ακολουθίες σε χρονοσειρές μπορούν να παραλληλιστούν με τις χωρικές ακολουθίες στις εικόνες.

### 4.7.3 CNN στην Πρόβλεψη Χρονοσειρών

Στο πλαίσιο της πρόβλεψης μονομεταβλητών χρονοσειρών, τα δεδομένα εκπαίδευσης είναι μια ακολουθία τιμών  $X = [x_1, x_2, \dots, x_T]$ , που αντιπροσωπεύουν τις τιμές της σειράς σε προηγούμενα χρονικά βήματα. Ο στόχος είναι να προβλέψουμε μελλοντικές τιμές  $Y = [y_{T+1}, y_{T+2}, \dots, y_{T+N}]$  με βάση αυτήν την ιστορική ακολουθία, όπου  $N$  είναι ο ορίζοντας πρόβλεψης.

Το πρώτο βήμα στην εφαρμογή των CNN για τη μονομεταβλητή πρόβλεψη χρονοσειρών περιλαμβάνει τη μετατροπή των δεδομένων χρονοσειρών σε μια μορφή κατάλληλη για συνελκτικές λειτουργίες. Αυτό επιτυγχάνεται συχνά μέσω μιας διαδικασίας όπου οι συνεχείς χρονοσειρές τμηματοποιούνται σε μια σειρά επικαλυπτόμενων παραθύρων, καθένα από τα

οποία χρησιμεύει ως μια είσοδος στο μοντέλο:

$$X_t = [x_{t-n+1}, \dots, x_{t-1}, x_t]$$

Εδώ, το  $n$  αντιπροσωπεύει το μέγεθος του παραθύρου. Κάθε παράθυρο  $X_t$  αντιστοιχίζεται σε μια αντίστοιχη έξοδο  $Y_t$ , τις μελλοντικές τιμές που στοχεύει να προβλέψει το μοντέλο. Για μονομεταβλητές σειρές, αυτή η διαδικασία μετατρέπει τη σειρά σε μια διάταξη  $2\Delta$  όπου η μία διάσταση αντιστοιχεί στο μέγεθος του παραθύρου και η άλλη είναι μια διάσταση ενός καναλιού (παρόμοια με τις εικόνες σε κλίμακα του γκρι).

Μέσα σε ένα CNN, συνελικτικά επίπεδα εφαρμόζουν φίλτρα (πυρήνες) στα παράθυρα εισόδου για εξαγωγή χρονικών χαρακτηριστικών. Ένα φίλτρο  $F$  μεγέθους  $k$  λειτουργεί στην είσοδο με παράθυρο  $X_t$  ως εξής:

$$(F * X_t)(i) = \sum_{m=0}^{k-1} F(m) \cdot X_t(i + m)$$

όπου το  $(F * X_t)(i)$  υποδηλώνει τη λειτουργία συνέλιξης που εφαρμόζεται στη θέση  $i$  του παραθύρου  $X_t$ , δημιουργώντας έναν χάρτη χαρακτηριστικών που επισημαίνει τα χρονικά μοτίβα που ανιχνεύονται από το φίλτρο.

Εφαρμόζονται πολλαπλά φίλτρα για την καταγραφή των διαφόρων πτυχών της χρονικής δομής των δεδομένων, με κάθε φίλτρο να ειδικεύεται ενδεχομένως στην αναγνώριση διαφορετικών τύπων μοτίβων (π.χ. τάσεις, εποχιακά μοτίβα) [96]. Για να μειωθεί η διάσταση του μοντέλου και να ενισχυθεί η εστίαση του μοντέλου σε σημαντικά χαρακτηριστικά, συχνά χρησιμοποιούνται στρώματα ομαδοποίησης τα οποία ακολουθούν τα συνελικτικά επίπεδα. Στο πρόβλημα της πρόβλεψης χρονοσειρών χρησιμοποιείται συνήθως το μαξ ποολινγ [97], όπου διατηρείται η μέγιστη τιμή σε μια υποπεριοχή του χάρτη χαρακτηριστικών:

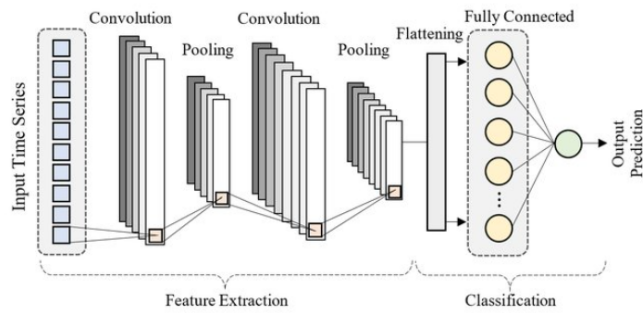
$$P(i) = \max_{m=i}^{i+k-1} (F * X_t)(m)$$

Με αυτόν τον τρόπο το μαξ ποολινγ μειώνει αποτελεσματικά τους χάρτες χαρακτηριστικών, μειώνοντας το μέγεθός τους, διατηρώντας παράλληλα τα πιο σημαντικά χαρακτηριστικά που εντοπίζονται από τα συνελικτικά φίλτρα.

Η έξοδος από το τελικό στρώμα συγκέντρωσης ή συνέλιξης τυπικά επιπεδοποιείται και τροφοδοτείται σε ένα ή περισσότερα πλήρως συνδεδεμένα στρώματα τεχνητών νευρώνων, καταλήγοντας σε ένα στρώμα εξόδου σχεδιασμένο να ταιριάζει με τις απαιτήσεις της εργασίας πρόβλεψης. Το επίπεδο εξόδου αποτελείται συχνά από έναν μόνο νευρώνα για πρόβλεψη ενός βήματος ή πολλούς νευρώνες για πρόβλεψη πολλαπλών βημάτων, ο καθένας από τους οποίους παράγει μια προβλεπόμενη τιμή.

Όπως είναι γνωστό, το CNN εκπαιδεύεται προσαρμόζοντας τα βάρη των φίλτρων και των πλήρως συνδεδεμένων επιπέδων για την ελαχιστοποίηση μιας συνάρτησης απώλειας, η οποία στην περίπτωση της πρόβλεψης χρονοσειρών είναι το μέσο τετραγωνικό σφάλμα (MSE):

$$MSE = \frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2$$



Εικόνα 4.7: Συνελικτικό νευρωνικό δίκτυο για πρόβλεψη χρονοσειρών

όπου  $N$  είναι ο αριθμός των προβλεπόμενων χρονικών βημάτων,  $\hat{y}_t$  είναι η προβλεπόμενη τιμή και  $y_t$  είναι η πραγματική τιμή τη στιγμή  $t$ .

Συνοπτικά, τα CNN μπορούν να χρησιμοποιηθούν αποτελεσματικά για πρόβλεψη μονομεταβλητών χρονοσειρών μετατρέποντας τις σειρές σε μορφή που επιδέχεται συνελικτική επεξεργασία, εφαρμόζοντας συνελικτικά και ομαδοποιητικά επίπεδα για εξαγωγή και μείωση διάστασης χρονικών χαρακτηριστικών και χρησιμοποιώντας πλήρως συνδεδεμένα επίπεδα για τις τελικές προβλέψεις. Αυτή η προσέγγιση επιτρέπει στα CNN να συλλαμβάνουν τόσο τοπικές όσο και άλλες, πιο μακρινές εξαρτήσεις στα δεδομένα, προσφέροντας ακριβείς προβλέψεις ακόμη και με την παρουσία πολύπλοκων χρονικών δυναμικών [98].



## Κεφάλαιο **5**

# Πρόβλεψη Ηλεκτρικού Φορτίου

---

### 5.1 Η Σημασία της Πρόβλεψης Ηλεκτρικού Φορτίου

Η ηλεκτρική ενέργεια είναι το όχημα για τη διατήρηση της εξαιρετικά προηγμένης τεχνολογικά παγκόσμιας οικονομίας [99, 100, 101]. Σχεδόν κάθε δραστηριότητα στη σημερινή εποχή εξαρτάται από την ηλεκτρική ενέργεια. Η οικονομική ανάπτυξη κάθε κράτους σχετίζεται σε μεγάλο βαθμό με την ηλεκτρική υποδομή, το δίκτυο και τη διαθεσιμότητά του. Ως εκ τούτου, η ζήτηση και η χρήση ηλεκτρικής ενέργειας για οικιακούς και εμπορικούς σκοπούς αυξάνεται παγκοσμίως [102]. Πιο συγκεκριμένα, η παγκόσμια ζήτηση ηλεκτρικής ενέργειας προβλέπεται να αυξηθεί κατά 85% το 2040 καθώς το βιοτικό επίπεδο αυξάνεται και οι οικονομίες επεκτείνονται [103].

Ωστόσο, η διαδικασία παραγωγής, μετάδοσης και διανομής ηλεκτρικής ενέργειας παραμένει περίπλοκη και δαπανηρή. Ως εκ τούτου, η αποτελεσματική διαχείριση του δικτύου παίζει ουσιαστικό ρόλο στη μείωση του κόστους παραγωγής ενέργειας και στην αύξηση της παραγωγικής ικανότητας για την κάλυψη της αυξανόμενης ζήτησης ηλεκτρικής ενέργειας [104].

Αντίστοιχα, η αποτελεσματική διαχείριση του δικτύου περιλαμβάνει σωστό σχεδιασμό ζήτησης φορτίου, επαρκές πρόγραμμα συντήρησης για τις γραμμές παραγωγής, μεταφοράς και διανομής και αποτελεσματική κατανομή φορτίου μέσω των γραμμών τροφοδοσίας. Ως εκ τούτου, μια ακριβής πρόβλεψη ηλεκτρικού φορτίου θα βοηθούσε πολύ στη μεγιστοποίηση της αποτελεσματικότητας της διαδικασίας σχεδιασμού στις βιομηχανίες παραγωγής ενέργειας [104, 105].

Η πρόβλεψη του φορτίου ηλεκτρικής ενέργειας αποτελεί βασική πτυχή της σύγχρονης διαχείρισης και διανομής ενέργειας. Περιλαμβάνει τη διαδικασία πρόβλεψης της ζήτησης ηλεκτρικής ενέργειας για συγκεκριμένες περιόδους, βοηθώντας στην κατανομή πόρων, σχεδιασμό υποδομής και διασφάλιση της αποτελεσματικής λειτουργίας των ηλεκτρικών δικτύων. Η σημασία της πρόβλεψης φορτίου ηλεκτρικής ενέργειας είναι πολύπλευρη.

Πιο συγκεκριμένα, η ακριβής πρόβλεψη φορτίου επιτρέπει στους παρόχους ενέργειας να βελτιστοποιούν την κατανομή των πόρων ταιριάζοντας την παροχή ηλεκτρικής ενέργειας με την αναμενόμενη ζήτηση. Με τις προβλέψεις αυτές, οι παραγωγοί ενέργειας μπορούν να αναπτύξουν στρατηγικά πόρους, όπως σταθμούς ηλεκτροπαραγωγής και ανανεώσιμες πηγές ενέργειας, διασφαλίζοντας αποτελεσματική χρήση [106]. Αυτή η βελτιστοποίηση όχι μόνο ενισχύει τη λειτουργική απόδοση, αλλά συμβάλλει επίσης στη μείωση του κόστους και στην περιβαλλοντική βιωσιμότητα, μειώνοντας την ανάγκη για πλεονάζουσα παραγωγή.

Επιπλέον, όπως είναι γνωστό, τα δίκτυα ηλεκτρικής ενέργειας λειτουργούν σε δυναμικά περιβάλλοντα που χαρακτηρίζονται από κυμαινόμενη ζήτηση. Η πρόβλεψη φορτίου διαδραματίζει κρίσιμο ρόλο στη διατήρηση της σταθερότητας και της αξιοπιστίας του δικτύου διευκολύνοντας την προληπτική διαχείριση και τον έλεγχο του [107, 108]. Οι φορείς διαχείρισης του δικτύου χρησιμοποιούν προβλέψεις φορτίου για να προβλέψουν περιόδους ζήτησης αιχμής, διακυμάνσεις τάσης και πιθανές υπερφορτίσεις, επιτρέποντας έγκαιρες προσαρμογές και προληπτικά μέτρα για την αποφυγή διακοπών ρεύματος.

Ωστόσο, οι ακριβείς προβλέψεις στον συγκεκριμένο τομέα δεν χρησιμεύουν μόνο για τους λόγους που αναφέρθηκαν παραπάνω. Η ακριβής πρόβλεψη φορτίου δίνει τη δυνατότητα στους καταναλωτές και τις επιχειρήσεις να εφαρμόζουν στρατηγικές διαχείρισης από την πλευρά της ζήτησης με στόχο τη βελτιστοποίηση της κατανάλωσης ενέργειας, τη μείωση του κόστους και την προώθηση της ενεργειακής απόδοσης. Οπλισμένοι με ακριβείς προβλέψεις φορτίου, οι καταναλωτές μπορούν να προγραμματίσουν διάφορες δραστηριότητες, όπως η χρήση συσκευών και η φόρτιση ηλεκτρικών οχημάτων, σε ώρες εκτός αιχμής, όταν οι τιμές του ηλεκτρικού ρεύματος είναι χαμηλότερες. Επιπλέον, η πρόβλεψη φορτίου διευκολύνει την εφαρμογή προγραμμάτων, όπου οι καταναλωτές προσαρμόζουν οικειοθελώς τη χρήση ηλεκτρικής ενέργειας ανάλογα με τις συνθήκες του δικτύου, μειώνοντας έτσι τις πιέσεις αιχμής της ζήτησης και προάγοντας τη σταθερότητα του δικτύου. Ως εκ τούτου, όπως γίνεται αντιληπτό, οι πρωτοβουλίες διαχείρισης από την πλευρά της ζήτησης όχι μόνο ωφελούν τους μεμονωμένους καταναλωτές αλλά συμβάλλουν επίσης στη συνολική εξοικονόμηση ενέργειας και στην περιβαλλοντική βιωσιμότητα.

Επιπροσθέτως, η ακριβής πρόβλεψη φορτίου υποστηρίζει την ενοποίηση ανανεώσιμων πηγών ενέργειας, όπως η ηλιακή και η αιολική ενέργεια, ευθυγραμμίζοντας τα χρονοδιαγράμματα παραγωγής με τα πρότυπα ζήτησης. Οι διαχειριστές συστημάτων διανομής και μεταφοράς βασίζονται στις προβλέψεις προκειμένου να αντιμετωπίσουν τις στοχαστικές διακυμάνσεις των καταναμημένων ανανεώσιμων πηγών ενέργειας που συνδέονται στο δίκτυο [109]. Αυτό ισχύει τόσο για το συνολικό φορτίο του συστήματος (δηλαδή, σε βάση χώρας) όσο και για το φορτίο που καλύπτεται από τα μικρότερα δίκτυα. Αν και το κύριο σώμα της αντίστοιχης βιβλιογραφίας ασχολείται με την πρόβλεψη του συνολικού φορτίου σε επίπεδο χώρας, περιοχής ή νομού/κοινοότητας, σημαντική προσοχή στρέφεται στο φορτίο των συστημάτων μεταφοράς και διανομής, τα οποία επηρεάζονται περισσότερο από τη στοχαστική φύση μεμονωμένων φορτίων [110].

Όπως επίσης είναι γνωστό, οι αγορές ηλεκτρικής ενέργειας λειτουργούν σε πολύπλοκα ρυθμιστικά πλαίσια που χαρακτηρίζονται από το δυναμικό σύστημα προσφοράς-ζήτησης, μηχανισμούς τιμολόγησης και ανταγωνισμό στην αγορά. Η πρόβλεψη φορτίου χρησιμεύει σημαντικά στις λειτουργίες της αγοράς, επιτρέποντας στους συμμετέχοντες στην αγορά να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με το εμπόριο ενέργειας, τις στρατηγικές τιμολόγησης και τη διαχείριση κινδύνου. Οι παραγωγοί και οι έμποροι ενέργειας βασίζονται σε προβλέψεις φορτίου για να βελτιστοποιήσουν τα χρονοδιαγράμματα παραγωγής, να αντισταθμίσουν την αστάθεια των τιμών και να μεγιστοποιήσουν τις ευκαιρίες εσόδων στις αγορές χονδρικής ηλεκτρικής ενέργειας. Ομοίως, οι έμποροι λιανικής χρησιμοποιούν την πρόβλεψη φορτίου για να αναπτύξουν σχέδια τιμολόγησης, να διαχειριστούν χαρτοφυλάκια πελατών και να ενισχύσουν την ανταγωνιστικότητα της αγοράς. Η ακριβής πρόβλεψη φορτίου προάγει την αποτελεσματικότητα, τη ρευστότητα και τη διαφάνεια της αγοράς, προωθώντας έτσι ένα ζωντανό και



βιώσιμο οικοσύστημα αγοράς ενέργειας.

Τέλος, ως προς τις προβλέψεις με ορίζοντα αρκετών ετών, η πρόβλεψη φορτίου ηλεκτρικής ενέργειας ενημερώνει τον μακροπρόθεσμο ενεργειακό σχεδιασμό και τις πρωτοβουλίες ανάπτυξης πολιτικής που στοχεύουν στη διασφάλιση της ενεργειακής αξιοπιστίας και βιωσιμότητας. Οι κυβερνητικές υπηρεσίες, οι επιχειρήσεις κοινής ωφέλειας και οι υπεύθυνοι χάραξης πολιτικής χρησιμοποιούν προβλέψεις φορτίου για να αξιολογήσουν τις μελλοντικές τάσεις της ζήτησης ενέργειας, να προσδιορίσουν τις ανάγκες υποδομής και να διαμορφώσουν ενεργειακές πολιτικές και κανονισμούς. Ενσωματώνοντας πληροφορίες πρόβλεψης φορτίου στις διαδικασίες ενεργειακού σχεδιασμού, οι ενδιαφερόμενοι μπορούν να δώσουν προτεραιότητα στις επενδύσεις σε υποδομές μεταφοράς και διανομής, να προωθήσουν την ανάπτυξη ανανεώσιμων πηγών ενέργειας και να αντιμετωπίσουν τις αναδυόμενες προκλήσεις, όπως η κλιματική αλλαγή και η ενεργειακή μετάβαση. Επιπλέον, η ακριβής πρόβλεψη φορτίου υποστηρίζει την ενσωμάτωση αναδυόμενων τεχνολογιών, όπως τα συστήματα αποθήκευσης ενέργειας και τα έξυπνα δίκτυα, στο ενεργειακό τοπίο, ενισχύοντας την καινοτομία και την ανθεκτικότητα.

Συμπερασματικά, η πρόβλεψη φορτίου ηλεκτρικής ενέργειας αποτελεί κρίσιμο παράγοντα για την αποτελεσματική διαχείριση ενέργειας, την αξιοπιστία του δικτύου και τα βιώσιμα ενεργειακά συστήματα [107, 108]. Παρέχοντας πληροφορίες για τα μελλοντικά πρότυπα ζήτησης ηλεκτρικής ενέργειας, η πρόβλεψη φορτίου διευκολύνει την προληπτική λήψη αποφάσεων, τη βελτιστοποίηση των πόρων και τον σχεδιασμό υποδομών σε όλη την αλυσίδα αξίας της ενέργειας. Επιπλέον, όπως κανείς μπορεί να υποθέσει από τα προηγούμενα, η πρόβλεψη του φορτίου ηλεκτρικής ενέργειας σχετίζεται σε μεγάλο βαθμό με την ανάπτυξη της οικονομίας, με την εθνική ασφάλεια καθώς και την καθημερινή λειτουργία της κοινωνίας [111]. Ως εκ τούτου, η ακρίβεια της πρόβλεψης ηλεκτρικού φορτίου έχει μεγάλη σημασία [112] για τη γενικότερη ευημερία της κοινωνίας.

Ωστόσο, η ακρίβεια της πρόβλεψης ηλεκτρικού φορτίου δεν μπορεί συχνά να εκπληρώσει το επιθυμητό αποτέλεσμα επειδή επηρεάζεται από διάφορους αβέβαιους και ανεξέλεγκτους παράγοντες όπως η οικονομική ανάπτυξη, οι ανθρώπινες κοινωνικές δραστηριότητες, οι πολιτικές της χώρας και η κλιματική αλλαγή [113].

## 5.2 Παράγοντες που επηρεάζουν την Ακρίβεια των Προβλέψεων

Πιο συγκεκριμένα, εξετάζοντας τους παράγοντες που συνήθως επηρεάζουν την ακρίβεια των προβλέψεων αξίζει κανείς να εμβαθύνει κυρίως στους οικονομικούς παράγοντες και τις καιρικές συνθήκες.

Οι οικονομικοί συντελεστές συμπεριλαμβάνουν παράγοντες όπως η βιομηχανική ανάπτυξη, η αύξηση του πληθυσμού, το Ακαθάριστο Εγχώριο Προϊόν (ΑΕΠ) και το κόστος ηλεκτρικής ενέργειας κ.λπ. Η μακροπρόθεσμη πρόβλεψη φορτίου επηρεάζεται σημαντικά από οικονομικούς παράγοντες, ωστόσο είναι επίσης σημαντική για μεσοπρόθεσμα και βραχυπρόθεσμη πρόβλεψη [114]. Σύμφωνα με τους διαφορετικούς ορίζοντες των προβλέψεων, οι διαφορετικοί οικονομικοί παράγοντες συμβάλουν με διαφορετικό τρόπο π.χ. χρόνος χρήσης για βραχυπρόθεσμες προβλέψεις, αγοραστική δύναμη για μεσοπρόθεσμες προβλέψεις και ΑΕΠ για μακροπρόθεσμες

προβλέψεις κ.λπ. Ως προς το ΑΕΠ, καθώς ως δείκτης υποδεικνύει το μέγεθος της οικονομικής δραστηριότητας, είναι προφανές πως η αύξησή του είναι η κύρια ώθηση για την τόνωση της ζήτησης ηλεκτρικής ενέργειας. Παράλληλα έχει παρατηρηθεί πως η βιομηχανική ανάπτυξη αλλά και ο υψηλός ρυθμός αύξησης του πληθυσμού σε μια συγκεκριμένη περιοχή αυξάνει την κατανάλωση ενέργειας. Το κόστος ηλεκτρικής ενέργειας είναι ένας ακόμη παράγοντας που επηρεάζει το φορτίο. Η ποσότητα της μη απαραίτητης κατανάλωσης ηλεκτρικής ενέργειας αυξάνεται όταν η ηλεκτρική ενέργεια γίνεται φθηνότερη. Επιπλέον, η τιμή του πετρελαίου δραματίζει σημαντικό ρόλο καθώς η διακύμανση της τιμής του πετρελαίου θα μεταβάλλει το κόστος παραγωγής ηλεκτρικής ενέργειας και κατά συνέπεια το κόστος ηλεκτρικής ενέργειας, το οποίο τελικά επηρεάζει την ζήτηση.

Ως προς τις καιρικές συνθήκες που επηρεάζουν την ακρίβεια των προβλέψεων, διάφορες μεταβλητές καιρού θα μπορούσαν να ληφθούν υπόψη για την πρόβλεψη της ζήτησης ηλεκτρικού ρεύματος. Κάποιες από αυτές είναι η θερμοκρασία, η υγρασία, η ταχύτητα του ανέμου και η πυκνότητα νεφών.

Πιο συγκεκριμένα, όπως έχει παρατηρηθεί, η κατανάλωση ηλεκτρικού ρεύματος και η θερμοκρασία συνδέονται σε σημαντικό βαθμό. Υπάρχει μια θετική συσχέτιση μεταξύ της καμπύλης θερμοκρασίας και της ζήτησης ηλεκτρικού ρεύματος ειδικά κατά τη θερινή περίοδο [114]. Αυτό οφείλεται στο γεγονός ότι κατά τη διάρκεια του καλοκαιριού, η αλλαγή της θερμοκρασίας θα επηρεάσει το αίσθημα της απαίτησης επιπέδου άνεσης των ανθρώπων. Κατά τη διάρκεια του καλοκαιριού καθώς αυξάνεται η θερμοκρασία, η αυξημένη χρήση ψυκτικών συσκευών αυξάνει επίσης την κατανάλωση ρεύματος.

Επιπλέον, η υγρασία επηρεάζει τη βραχυπρόθεσμη πρόβλεψη ζήτησης ηλεκτρικής ενέργειας καθώς αυξάνει την αίσθηση της σοβαρότητας της θερμοκρασίας κατά τη διάρκεια του καλοκαιριού και της περιόδου των βροχών. Έτσι, η κατανάλωση ρεύματος αυξάνεται κατά τη διάρκεια της καλοκαιρινής υγρής ημέρας.

Επιπροσθέτως, οι ταχύτητες ανέμου επηρεάζουν την κατανάλωση ηλεκτρικού φορτίου. Κατά τη διάρκεια της ημέρας με αέρα το ανθρώπινο σώμα αισθάνεται τη θερμοκρασία πολύ πιο κάτω και απαιτείται συσκευή θέρμανσης αυξάνοντας έτσι την κατανάλωση ρεύματος. Τέλος, οι επιπτώσεις των έντονων νεφώσεων στη χρήση ηλεκτρικής ενέργειας εξαρτώνται από την ώρα. Κατά τη διάρκεια της ημέρας η νέφωση μπορεί να διαταράξει το ηλιακό φως, με αποτέλεσμα τη μείωση της θερμοκρασίας και ως εκ τούτου τη μείωση της κατανάλωσης ηλεκτρικής ενέργειας.

### 5.3 Ταξινόμηση Προβλέψεων Ζήτησης Ηλεκτρικής Ενέργειας

Όπως έχει γίνει ήδη αντιληπτό, οι διάφορες περιπτώσεις πρόβλεψης του ενεργειακού φορτίου διαφέρουν μεταξύ τους σε ένα βαθμό, και όπως προκύπτει και από την προηγούμενη ανάλυση, η βασική ειδοποιός διαφορά είναι ο ορίζοντας πρόβλεψης.

Πιο συγκεκριμένα, ορισμένοι ερευνητές έχουν χωρίσει την πρόβλεψη φορτίου ως προς τον ορίζοντα της πρόβλεψης σε τρεις κατηγορίες [108, 115, 116, 117]: βραχυπρόθεσμες προβλέψεις, μεσοπρόθεσμες προβλέψεις και μακροπρόθεσμες προβλέψεις. Άλλοι ερευνητές ταξινομούν την πρόβλεψη φορτίου σε τέσσερις ομάδες [107, 118, 119, 120]: μακροπρόθεσμες

προβλέψεις, μεσοπρόθεσμες προβλέψεις, βραχυπρόθεσμες προβλέψεις και πολύ βραχυπρόθεσμες προβλέψεις.

Οι μακροπρόθεσμες προβλέψεις φορτίου (*LTLF*) αναφέρονται συνήθως σε περισσότερο από ένα έτος έως και 20 χρόνια μπροστά. Αυτός ο τύπος πρόβλεψης είναι θεμελιώδης για τον στρατηγικό σχεδιασμό και την ανάπτυξη του συστήματος παροχής και παράδοσης ηλεκτρικής ενέργειας. Μία μεσοπρόθεσμη πρόβλεψη φορτίου (*MTLF*) είναι συνήθως για μια εβδομάδα έως ένα έτος. Χρησιμοποιείται για τον προγραμματισμό συντήρησης και τον προγραμματισμό αγορών καυσίμων καθώς και για την εμπορία ενέργειας και την αξιολόγηση εσόδων για τις επιχειρήσεις παροχής ηλεκτρικού ρεύματος. Παράλληλα, ο όρος βραχυπρόθεσμη πρόβλεψη φορτίου (*STLF*) χρησιμοποιείται για διαστήματα που κυμαίνονται από μία ώρα έως μία εβδομάδα. Αυτού του είδους η πρόβλεψη είναι πολύ σημαντική για τις καθημερινές λειτουργίες μιας εταιρείας παροχής ηλεκτρικής ενέργειας και για τον προγραμματισμό, την παραγωγή και τη μεταφορά ηλεκτρικής ενέργειας. Τέλος η πολύ βραχυπρόθεσμη πρόβλεψη φορτίου (*VSTLF*) κυμαίνεται από λίγα λεπτά έως μία ώρα μπροστά και χρησιμοποιείται για έλεγχο σε πραγματικό χρόνο.

Η παρούσα εργασία θα εστιάσει στις βραχυπρόθεσμες προβλέψεις φορτίου ηλεκτρικής ενέργειας.

Όπως γίνεται αντιληπτό, το συγκεκριμένο είδος προβλέψεων, οι βραχυπρόθεσμες προβλέψεις φορτίου, έχουν απασχολήσει εκτενώς την ακαδημαϊκή κοινότητα, λόγω της ιδιαίτερης χρησιμότητάς τους. Το συγκεκριμένο πρόβλημα μελετάται εδώ και δεκαετίες και για την αντιμετώπισή του έχουν χρησιμοποιηθεί πληθώρα μεθόδων.

## 5.4 Μέθοδοι Πρόβλεψης Ζήτησης Ηλεκτρικής Ενέργειας

Πριν την ανάπτυξη των σύγχρονων υπολογιστικών εργαλείων που κινούνται κυρίως στην περιοχή της μηχανικής μάθησης και της βαθιάς μάθησης κατ' επέκταση, αυτού του είδους οι προβλέψεις πραγματοποιούνταν χρησιμοποιώντας παραδοσιακές (συμβατικές) μαθηματικές τεχνικές. Οι πιο συχνά χρησιμοποιούμενες παραδοσιακές τεχνικές συμπεριλαμβάνουν την παλινδρόμηση, την πολλαπλή παλινδρόμηση και την εκθετική εξομάλυνση. Μετέπειτα, οι στατιστικές μέθοδοι εξελίχθηκαν και προέκυψαν μοντέλα όπως τα μοντέλα ARMA και ARIMA.

Η παλινδρόμηση είναι μια από τις πιο ευρέως χρησιμοποιούμενες στατιστικές τεχνικές, ενώ πολλές φορές παρουσιάζεται και ως μέθοδος μηχανικής μάθησης, όπως παρουσιάστηκε και στα προηγούμενα κεφάλαια, καθώς είναι αρκετά εύκολο να εφαρμοστεί [121, 122, 123, 124]. Οι μέθοδοι παλινδρόμησης χρησιμοποιούνται συνήθως για τη μοντελοποίηση της σχέσης κατανάλωσης φορτίου και ενός άλλου παράγοντα όπως οι καιρικές συνθήκες, οι τύποι ημερών και οι κατηγορίες πελατών. Αυτή η μέθοδος προϋποθέτει ότι το φορτίο μπορεί να χωριστεί σε μια τυπική τάση φορτίου και μια τάση που εξαρτάται γραμμικά από έναν άλλο παράγοντα που επηρεάζει το φορτίο.

Η πολλαπλή παλινδρόμηση είναι η πιο δημοφιλής μέθοδος και χρησιμοποιείται συχνά για την πρόβλεψη του φορτίου, ιδιαίτερα όταν τα διαθέσιμα δεδομένα δίνουν πληροφορίες για μία σειρά από παράγοντες που επηρεάζουν την πρόβλεψη όπως μετεωρολογικές επιδράσεις, τιμές ηλεκτρικής ενέργειας και άλλοι οικονομικοί παράγοντες. Η τεχνική της πολλαπλής παλινδρόμησης χρησιμοποιεί την μέθοδο ελαχίστων τετραγώνων για την εύρεση της σχέσης που

διέπει το συνολικό φορτίο και τις υπόλοιπες ανεξάρτητες μεταβλητές. Ιδιαίτερα γνωστή στον συγκεκριμένο χώρο είναι η δουλειά των Mbamalu και El-Hawary [125] οι οποίοι χρησιμοποίησαν ένα μοντέλο το οποίο δέχεται ως είσοδο ένα διάλυμα μεταβλητών που περιλαμβάνει στοιχεία όπως η θερμοκρασία, η ένταση φωτός, η ταχύτητα ανέμου, η υγρασία αλλά και ο τύπος της ημέρας (εργάσιμη ημέρα, Σαββατοκύριακο). Συνήθως, ο πολυωνυμικός βαθμός μέσω του οποίου η κάθε μεταβλητή επηρεάζει το αποτέλεσμα επιλέγεται από ένα εύρος τιμών που συνήθως δεν ξεπερνά τον 5ο βαθμό. Στις περισσότερες περιπτώσεις, η γραμμική εξάρτηση δίνει τα καλύτερα αποτελέσματα, ενώ υπάρχει πληθώρα ερευνών που διαπιστώνουν την αποτελεσματικότητα αυτής της τεχνικής ως προς την βραχυπρόθεσμη πρόβλεψη φορτίου [126, 127]. Παραλλάσσοντας κάπως την αρχική ιδέα, οι Παπαλεξόπουλος και Hesterberg ανέπτυξαν ένα μοντέλο που παράγει μια αρχική ημερήσια πρόβλεψη και στη συνέχεια χρησιμοποιεί αυτήν την αρχική πρόβλεψη για να παράγει ωριαίες προβλέψεις [128]. Στο επόμενο βήμα, χρησιμοποιούν το μέγιστο της αρχικής ωριαίας πρόβλεψης, το πιο πρόσφατο σφάλμα και τα εκθετικά εξομαλυσμένα σφάλματα ως μεταβλητές σε ένα μοντέλο πολλαπλής παλινδρόμησης. Επιπλέον, μια προσέγγιση της μεθόδου ελαχίστων τετραγώνων χρησιμοποιήθηκε από τους Varadan και Makram για να προσδιορίσουν το φορτίο σε γραμμές και υποσταθμούς ηλεκτρικής ενέργειας [129]. Τέλος, ο Broadwater (1997) εισήγαγε μια νέα μέθοδο που βασίζεται στην παλινδρόμηση, τον Nonlinear Load Research Estimator (NLRE) [130, 131, 132].

Η εκθετική εξομάλυνση χρησιμοποιεί μια προσέγγιση που δίνει μεγαλύτερη έμφαση σε πιο πρόσφατες παρατηρήσεις για την πρόβλεψη μελλοντικών τιμών, καθιστώντας την ιδιαίτερα κατάλληλη για δεδομένα χρονοσειρών με εγγενή μοτίβα όπως η τάση ή η εποχικότητα, κάτι που υπό προϋποθέσεις μπορεί να φανεί χρήσιμο στην βραχυπρόθεσμη πρόβλεψη ηλεκτρικού φορτίου. Ένα σημαντικό πλεονέκτημα της εκθετικής εξομάλυνσης είναι η απλότητα και το χαμηλό υπολογιστικό της κόστος, παρέχοντας μια ισχυρή μέθοδο για βραχυπρόθεσμες προβλέψεις σε διάφορες εφαρμογές, συμπεριλαμβανομένης της πρόβλεψης ηλεκτρικού φορτίου. Για παράδειγμα, η μέθοδος του Winter, η οποία αναλύθηκε σε προηγούμενο κεφάλαιο, έχει χρησιμοποιηθεί ευρέως στον συγκεκριμένο τομέα. Αυτή η μέθοδος έχει αποδειχθεί ότι αποδίδει καλά στην ανάλυση και την πρόβλεψη δεδομένων εποχιακών χρονοσειρών, καθιστώντας την ανεκτίμητη στο πλαίσιο της πρόβλεψης φορτίου όπου η κατανόηση και η πρόβλεψη εποχιακών προτύπων είναι ζωτικής σημασίας [126, 127]. Επιπλέον, έχουν προταθεί και διάφορες υβριδικές μέθοδοι που χρησιμοποιούν και εκθετική εξομάλυνση [133] ενώ η τεχνική αφαίρεσης τάσης με βάση την εξομάλυνση των Infield and Hill [134] έχει αποδειχθεί ότι μπορεί να ανταγωνιστεί τις υπόλοιπες συμβατικές μεθόδους πρόβλεψης φορτίου.

Εξέλιξη των συμβατικών μεθόδων που παρουσιάστηκαν προηγουμένως αποτελεί το μαθηματικό μοντέλο ARMA. Καθώς η μέθοδος έχει αναλυθεί εκτενώς σε προηγούμενο κεφάλαιο, δεν θα γίνει περαιτέρω εμβάθυνση στον τρόπο που λειτουργεί. Κλασικό παράδειγμα χρήσης αυτής της τεχνικής στον υπό μελέτη τομέα είναι η προσέγγιση των Fan και McDonald [129] οι οποίοι χρησιμοποίησαν τον αλγόριθμο WRLS (Weighted Recursive Least Squares) για να ενημερώσουν τις παραμέτρους του μοντέλου ARMA. Επιπλέον, ο Chen [135] χρησιμοποίησε ένα μοντέλο ARMA για την βραχυπρόθεσμη πρόβλεψη φορτίου, στο οποίο τα διαθέσιμα σφάλματα πρόβλεψης χρησιμοποιούνται για την ενημέρωση του μοντέλου. Χρησιμοποιώντας το ελάχιστο μέσο τετραγωνικό σφάλμα για την εξαγωγή συντελεστών εκμάθησης σφάλματος, το προσαρμοσμένο υπολογιστικό σχήμα ξεπέρασε τα συμβατικά μοντέλα ARMA.

Το μοντέλο ARIMA είναι ένα βασικό εργαλείο στη βραχυπρόθεσμη πρόβλεψη φορτίου, καθώς είναι ιδιαίτερα αποτελεσματικό για δυναμικές και μη σταθερές χρονοσειρές. Χρησιμοποιώντας διαφόριση, όπως έχει παρουσιαστεί και στο αντίστοιχο κεφάλαιο, η ARIMA μετατρέπει τα μη στάσιμα δεδομένα σε μια σταθερή σειρά, η οποία είναι ζωτικής σημασίας για τον εντοπισμό υποκείμενων προτύπων όπως οι τάσεις και η εποχικότητα. Επεκτείνοντας λοιπόν την ισχύ των μοντέλων ARMA, τα συγκεκριμένα μοντέλα έχουν χρησιμοποιηθεί από πολλούς ερευνητές για την βραχυπρόθεσμη πρόβλεψη φορτίου [127, 136, 137].

Όπως γίνεται αντιληπτό οι συμβατικές μέθοδοι αποτελούν ένα σημαντικό πρώτο βήμα προς την επίτευξη του στόχου των αποτελεσματικών βραχυπρόθεσμων προβλέψεων ηλεκτρικού φορτίου. Ωστόσο, η πρόοδος που σημειώθηκε τα τελευταία χρόνια στον χώρο της μηχανικής μάθησης επηρέασε αναπόφευκτα και αυτού του είδους τις προβλέψεις.

Μία από τις συχνά χρησιμοποιούμενες μεθόδους μηχανικής μάθησης που χρησιμοποιούνται για την βραχυπρόθεσμη πρόβλεψη ηλεκτρικού φορτίου είναι τα SVMs, καθώς υπάρχει εκτενής έρευνα στον τομέα [138, 139, 140]. Παράλληλα, πολλές έρευνες συγκρίνουν την αποτελεσματικότητά τους στο πεδίο των προβλέψεων με αυτή των νευρωνικών δικτύων [141].

Ένας ακόμη αλγόριθμος μηχανικής μάθησης ο οποίος χρησιμοποιείται αρκετά για βραχυπρόθεσμες προβλέψεις φορτίου είναι το δέντρο απόφασης (decision tree) [142, 143], ενώ εξελίσσοντας τον αλγόριθμο του δέντρου απόφασης, συμβάλει συχνά στις βραχυπρόθεσμες προβλέψεις φορτίου ο αλγόριθμος τυχαίου δάσους (Random Forest). Συνδυάζοντας πολλαπλά δέντρα αποφάσεων για να κάνει προβλέψεις συνήθως αποφεύγει τον κίνδυνο του overfitting [144], κάτι ιδιαίτερα χρήσιμο και στον συγκεκριμένο τομέα [145, 146].

Προχωρώντας σε πιο πρόσφατες εξελίξεις στον χώρο, η εφαρμογή των μοντέλων βαθιάς μάθησης, ιδίως των αναδρομικών νευρωνικών δικτύων (RNN), των δικτύων μακράς βραχυπρόθεσμης μνήμης (LSTM) και των νευρωνικών δικτύων συνελικτικού τύπου (CNN), των οποίων η λειτουργία έχει αναλυθεί και εξηγηθεί εκτενώς σε προηγούμενο κεφάλαιο, έχει προωθήσει σημαντικά τον τομέα της βραχυπρόθεσμης πρόβλεψης φορτίου (STLF) τα τελευταία χρόνια καθώς οι αρχιτεκτονικές βαθιάς μάθησης είναι ικανές να αποτυπώσουν σύνθετες μη γραμμικές σχέσεις μεταξύ του φορτίου και διαφόρων μεταβλητών που το επηρεάζουν [147]. Οι μοναδικές ικανότητες αυτών των μοντέλων στο χειρισμό σειριακών δεδομένων και στη σύλληψη χρονικών εξαρτήσεων τα καθιστούν ιδανικά για την πρόβλεψη της ζήτησης ηλεκτρικής ενέργειας, η οποία είναι εγγενώς ένα πρόβλημα με δεδομένα χρονοσειρών που χαρακτηρίζονται από εποχικότητα και τάσεις.

Τα RNN, με την ικανότητά τους να επεξεργάζονται ακολουθίες δεδομένων, αποτέλεσαν ακρογωνιαίο λίθο στις εφαρμογές STLF. Οι Shi κ.ά. [148] κατέδειξαν την αποτελεσματικότητα τόσο των ρηχών όσο και των βαθιών RNNs στην καταγραφή των δυναμικών μοτίβων που είναι εγγενή στην πρόβλεψη φορτίου. Η διαδοχική φύση των RNNs, η οποία επιτρέπει την επεξεργασία δεδομένων εισόδου σε μια χρονική ακολουθία, τα καθιστά ιδιαίτερα κατάλληλα για εργασίες πρόβλεψης χρονοσειρών, όπου η τρέχουσα έξοδος εξαρτάται από προηγούμενες εισόδους.

Τα LSTM, ένα ειδικό είδος RNN, έχουν σχεδιαστεί για να ξεπερνούν τους περιορισμούς των παραδοσιακών RNN, ιδίως όσον αφορά τη μάθηση των μακροπρόθεσμων εξαρτήσεων. Με την ενσωμάτωση κυττάρων μνήμης, τα LSTM είναι ικανά να θυμούνται πληροφορίες για μεγάλες ακολουθίες, καθιστώντας τα ιδιαίτερα αποτελεσματικά σε εφαρμογές STLF όπου η

κατανόηση της χρονικής δυναμικής είναι ζωτικής σημασίας. Οι Narayan και Hipel [149] εφάρμοσαν με επιτυχία μοντέλα LSTM σε ένα σύνολο δεδομένων φορτίου ηλεκτρικής ενέργειας 10 ετών από το Οντάριο του Καναδά, αναδεικνύοντας την επάρκεια του μοντέλου στην αντιμετώπιση μακροπρόθεσμων ιστορικών δεδομένων. Παρομοίως, οι Kumar κ.ά. [150] διαπίστωσαν ότι η GRU, μια άλλη παραλλαγή του RNN που απλοποιεί την αρχιτεκτονική LSTM, υπερέχει των LSTM στην πρόβλεψη της ζήτησης ηλεκτρικής ενέργειας, υπογραμμίζοντας τη δυνατότητα των μοντέλων GRU να αποτυπώνουν αποτελεσματικά τις χρονικές εξαρτήσεις με μικρότερη υπολογιστική πολυπλοκότητα.

Η ενσωμάτωση των CNN στο STLF άνοιξε νέους δρόμους για τη διερεύνηση χωροχρονικών σχέσεων σε δεδομένα φορτίου ηλεκτρικής ενέργειας. Σε αντίθεση με τα παραδοσιακά RNNs και LSTMs που εστιάζουν κυρίως σε χρονικά χαρακτηριστικά, τα CNNs υπερέχουν στην εξαγωγή ιεραρχικών χωρικών χαρακτηριστικών από τα δεδομένα, γεγονός που μπορεί να είναι ιδιαίτερα επωφελές στο STLF όταν πρόκειται για μεγάλα σύνολα δεδομένων από έξυπνα δίκτυα ή κατανεμημένους αισθητήρες. Οι Cai κ.ά. [151] χρησιμοποίησαν μοντέλα CNN με πύλες (GCNN) για να επιτύχουν αξιοσημείωτη ακρίβεια πρόβλεψης, τονίζοντας την ευρωστία και την ικανότητα γενίκευσης των CNNs στον χειρισμό της περιοδικότητας και των δυναμικών προτύπων που υπάρχουν στα δεδομένα φορτίου ηλεκτρικής ενέργειας. Ομοίως, η εργασία των Dong κ.ά. [152] οι οποίοι συνδύασαν την ομαδοποίηση k-means με CNNs, αναδεικνύει την ικανότητα του μοντέλου να ομαδοποιεί και να προβλέπει αποτελεσματικά δεδομένα φορτίου ηλεκτρικής ενέργειας από διαφορετικές περιοχές, καταδεικνύοντας περαιτέρω την ευελιξία των CNNs στην καταγραφή τόσο των χωρικών όσο και των χρονικών εξαρτήσεων. Επιπλέον, τα CNN έχουν επιδείξει επιτυχία και στην πρόβλεψη φορτίου βασισμένη αποκλειστικά σε καιρικές συνθήκες και ιστορικά δεδομένα χρονοσειράς [153].

Φυσικά πολύ συχνός είναι και ο συνδυασμός διαφόρων μοντέλων για την επίτευξη ακριβέστερων προβλέψεων φορτίου. Τα υβριδικά μοντέλα, όπως συχνά ανφέρονται στη βιβλιογραφία, χρησιμοποιούν μια σύνθεση παραδοσιακών στατιστικών μεθόδων και μεθόδων μηχανικής μάθησης, προσφέροντας μια πολλά υποσχόμενη οδό για την ενίσχυση της ακρίβειας στην βραχυπρόθεση πρόβλεψη φορτίου σε εφαρμογές ενεργειακών συστημάτων. Ανάμεσα στα συνηθισμένα υβριδικά μοντέλα, τα ARIMA-SVM και ES-ANN ξεχωρίζουν για την αποτελεσματικότητά τους. Το ARIMA-SVM ενσωματώνει άψογα την αναγνώριση χρονικών προτύπων του ARIMA με την ικανότητα των SVM να αντλήσουν μη-γραμμικές σχέσεις ζήτησης φορτίου [154]. Ομοίως, το ES-NN συνδυάζει την Εκθετική Εξομάλυνση (ES) και τα Νευρωνικά Δίκτυα (NN) για να εξάγει με επιδεξιότητα και τις εποχιακές μεταβολές και τα μη-γραμμικά πρότυπα ζήτησης φορτίου [141]. Ένα ακόμη αξιοσημείωτο υβριδικό μοντέλο είναι το υβριδικό μοντέλο ARIMA-NN [155]. Η χρήση υβριδικών μοντέλων έχει δείξει σημαντικές βελτιώσεις τόσο στην ακρίβεια όσο και στην δυνατότητα ερμηνείας των βραχυπρόθεσμων αυτών προβλέψεων.

Επιπλέον, η καινοτόμος χρήση υβριδικών μοντέλων που συνδυάζουν τα πλεονεκτήματα των RNNs, των LSTMs και των CNNs οδήγησε σε σημαντικές βελτιώσεις στην απόδοση πρόβλεψης. Για παράδειγμα, ο He [156] ανέπτυξε μια παράλληλη δομή CNN-RNN για την αξιοποίηση των δυνατοτήτων εξαγωγής χαρακτηριστικών των CNNs μαζί με τη δύναμη της διαδοχικής επεξεργασίας δεδομένων των RNNs, με αποτέλεσμα τη βελτίωση της ακρίβειας πρόβλεψης φορτίου. Αυτή η υβριδική προσέγγιση συλλαμβάνει αποτελεσματικά τόσο τα χωρι-

κά χαρακτηριστικά μέσω των CNNs όσο και τη χρονική δυναμική μέσω των RNNs/LSTMs, προσφέροντας ένα ολοκληρωμένο μοντέλο για το STLF.

Ωστόσο, είναι σημαντικό να σημειωθεί ότι τα υβριδικά μοντέλα απαιτούν περισσότερους υπολογιστικούς πόρους σε σύγκριση με τις καθαρά στατιστικές προσεγγίσεις και μπορεί να παρουσιάζουν προκλήσεις στην υλοποίηση σε σύγκριση με τα αυτόνομα μοντέλα μηχανικής μάθησης [157]. Παρ' όλα αυτά, έχουν προταθεί ευρέως για την ενίσχυση της ακρίβειας πρόβλεψης φορτίου εκμεταλλευόμενα τα πλεονεκτήματα διαφορετικών τεχνικών μοντελοποίησης [158].



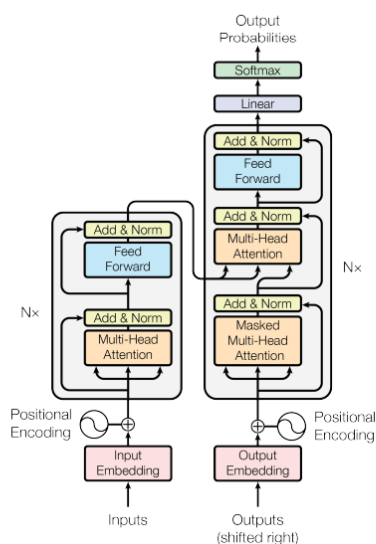


## Κεφάλαιο 6

# Μηχανισμός Προσοχής και Προβλέψεις Χρονοσειρών

### 6.1 Εισαγωγή στο Μηχανισμό Προσοχής

Το μοντέλο Transformer, που παρουσιάστηκε στο δημοσίευση ορόσημο 'Attention is All You Need' των Vaswani et al. το 2017 [159], αντιπροσωπεύει μια σημαντική απόκλιση από τα αναδρομικά ή συνελικτικά νευρωνικά δίκτυα, ιδιαίτερα στην επεξεργασία φυσικής γλώσσας (NLP). Η αρχιτεκτονική του αξιοποιεί μηχανισμούς αυτοπροσοχής, αποφεύγοντας εντελώς την αναδρομή και τις συνελίξεις, για να δημιουργήσει εξαρτήσεις μεταξύ εισόδου και εξόδου. Ο σχεδιασμός του Transformer επιτρέπει σημαντικά μεγαλύτερη παραλληλοποίηση κατά τη διάρκεια της εκπαίδευσης και δίνει τη δυνατότητα βαθιάς κατανόησης των σχέσεων μεταξύ διαφορετικών τμημάτων μιας ακολουθίας, καθιστώντας τον εξαιρετικά αποτελεσματικό για ένα ευρύ φάσμα εργασιών [159].



Εικόνα 6.1: Η αρχιτεκτονική του Transformer, έτσι όπως παρουσιάστηκε το 2017.

Ο μηχανισμός προσοχής, θεμελιώδες στοιχείο του Transformer, έχει γίνει ένα θεμελιώδες συστατικό στην αρχιτεκτονική των νευρωνικών δικτύων, ειδικά στον τομέα της επεξεργασίας φυσικής γλώσσας (NLP) [160]. Επιτρέπει στα μοντέλα να εστιάζουν δυναμικά σε διαφορε-

τικά μέρη της ακολουθίας εισόδου κατά την παραγωγή κάθε μέρους της ακολουθίας εξόδου, μιμούμενος την ανθρώπινη ικανότητα να δίνει προσοχή σε συγκεκριμένες πληροφορίες, ενώ αγνοεί άσχετες λεπτομέρειες.

## 6.2 Αρχιτεκτονική Μηχανισμού Προσοχής

### 6.2.1 Προσοχή Μιας Κεφαλής

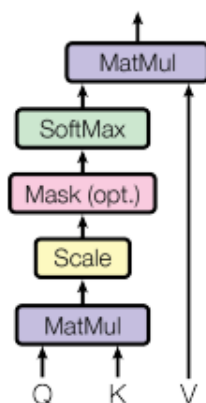
Η βασική ιδέα πίσω από τον μηχανισμό προσοχής είναι ο υπολογισμός ενός σταθμισμένου αθροίσματος ορισμένων διανυσμάτων, με τα weights να έχουν σχεδιαστεί για να λένε στο μοντέλο πού να εστιάσει την προσοχή του. Μαθηματικά, δεδομένου ενός συνόλου διανυσμάτων ερωτήματος (query)  $Q$ , διανυσμάτων κλειδιών (key)  $K$  και διανυσμάτων τιμής (value)  $V$  (τα οποία στην πράξη θα μπορούσαν να είναι τα ίδια διανύσματα στην αυτοπροσοχή), η συνάρτηση προσοχής μπορεί να περιγραφεί ως εξής:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Εδώ, το  $d_k$  αντιπροσωπεύει τη διάσταση των βασικών διανυσμάτων, η οποία χρησιμοποιείται ως παράγοντας κλιμάκωσης για να αποτρέψει το όρισμα της συνάρτησης softmax από το να γίνει πολύ μεγάλο σε μέγεθος. Η συνάρτηση softmax είναι ζωτικής σημασίας στον μηχανισμό προσοχής καθώς μετατρέπει τις βαθμολογίες (που προκύπτουν από  $QK^T$ ) σε πιθανότητες:

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

όπου  $z_i$  είναι ο βαθμός προσοχής για το  $i$ -οστό στοιχείο και ο παρονομαστής είναι το άθροισμα των εκθετικών τιμών όλων των βαθμολογιών. Αυτό διασφαλίζει ότι η έξοδος της συνάρτησης softmax είναι μια κατανομή των πιθανοτήτων στα keys, αντικατοπτρίζοντας την εστίαση του μοντέλου.



Εικόνα 6.2: Ο μηχανισμός προσοχής.

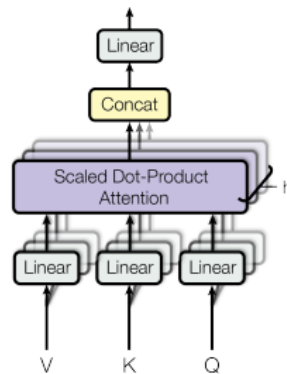
## 6.2.2 Προσοχή Πολλαπλών Κεφαλών

Για να επιτραπεί στο μοντέλο να παρακολουθεί από κοινού πληροφορίες από διαφορετικούς υποχώρους, ο μηχανισμός προσοχής μπορεί να πάρει την μορφή της προσοχής πολλαπλών κεφαλών. Αυτό περιλαμβάνει την παράλληλη εκτέλεση πολλών πράξεων προσοχής, καθεμία με το δικό της σύνολο γραμμικών μετασχηματισμών που έχουν μάθει για τα  $Q$ ,  $K$  και  $V$ , ακολουθούμενες από συνένωση και έναν άλλο γραμμικό μετασχηματισμό:

$$\text{Multihead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

$$\text{όπου } \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Κάθε κεφαλή καταγράφει διαφορετικές πτυχές της εισόδου και οι έξοδοι τους συνδυάζονται για να παράγουν την τελική έξοδο, επιτρέποντας στο μοντέλο να καταγράψει μια πιο πλούσια αναπαράσταση των δεδομένων εισόδου.



Εικόνα 6.3: Ο μηχανισμός προσοχής πολλαπλών κεφαλών.

## 6.3 Αυτοπροσοχή

Η αυτοπροσοχή είναι ένας νευρωνικός μηχανισμός που επιτρέπει σε κάθε θέση σε μια ακολουθία να παρακολουθεί και να ερευνά τη σχέση της με όλες τις θέσεις εντός της ίδιας ακολουθίας. Αυτός ο μηχανισμός αποτελεί βασικό στοιχείο του μοντέλου Transformer, το οποίο έχει επηρεάσει σημαντικά την επεξεργασία φυσικής γλώσσας (NLP) και όχι μόνο [161]. Η αυτοπροσοχή διευκολύνει τη μοντελοποίηση των σχέσεων και των εξαρτήσεων μεταξύ λέξεων ή υπολέξεων σε μια πρόταση, ανεξάρτητα από τη θέση τους, υπολογίζοντας τις βαθμολογίες προσοχής με αυτοαναφορικό τρόπο.

Αντίστοιχα με το μηχανισμό προσοχής που παρουσιάστηκε προηγουμένως, έστω ότι δίνεται μια ακολουθία που αντιπροσωπεύεται από έναν πίνακα  $X \in \mathbb{R}^{n \times d}$ , όπου  $n$  είναι το μήκος της ακολουθίας και  $d$  είναι η διάσταση κάθε διανύσματος της ακολουθίας. Η αυτοπροσοχή παράγει μια νέα ακολουθία  $Z$  του ίδιου μεγέθους.

Η διαδικασία αυτοπροσοχής περιλαμβάνει τρία βασικά βήματα:

1. **Γραμμικές προβολές:** Πρώτον, το  $X$  προβάλλεται γραμμικά σε τρεις διαφορετικούς

χώρους για να ληφθούν τα  $Q$ ,  $K$  και  $V$  χρησιμοποιώντας πίνακες βαρών:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V$$

Εδώ,  $W^Q, W^K, W^V \in \mathbb{R}^{d \times d_k}$

2. **Υπολογισμός βαθμού προσοχής:** Ο βαθμός προσοχής μεταξύ κάθε ζεύγους θέσεων υπολογίζεται λαμβάνοντας το εσωτερικό γινόμενο του ερωτήματός τους και των διανυσμάτων κλειδιών, ακολουθούμενο από έναν παράγοντα κλιμάκωσης  $\frac{1}{\sqrt{d_k}}$ :

$$\text{Self-Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$

3. **Σταθμισμένο άθροισμα τιμών:** Τέλος, για κάθε θέση στην ακολουθία, η έξοδος είναι ένα σταθμισμένο άθροισμα διανυσμάτων, όπου τα βάρη δίνονται από τις βαθμολογίες προσοχής που υπολογίστηκαν στο προηγούμενο βήμα. Αυτό έχει ως αποτέλεσμα την ακολουθία εξόδου  $Z$ , όπου κάθε στοιχείο είναι μια σύνθεση πληροφοριών από όλες τις θέσεις στην ακολουθία εισόδου, σταθμισμένη ανάλογα με τη συνάφειά τους.

Ο Transformer ενισχύει τον μηχανισμό αυτοπροσοχής χρησιμοποιώντας πολλαπλή αυτοπροσοχή, επιτρέποντας στο μοντέλο να παρακολουθεί από κοινού πληροφορίες από διαφορετικούς υποχώρους. Στην αυτοπροσοχή πολλών κεφαλών, οι ανεξάρτητες πράξεις αυτοπροσοχής (κεφαλές)  $h$  υπολογίζονται παράλληλα και οι έξοδοι τους συνδέονται και μετασχηματίζονται γραμμικά, αντίστοιχα με την πολυκέφαλη προσοχή:

$$\text{MultiHead Self-Attention}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

$$\text{όπου } \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Η αυτοπροσοχή επιτρέπει στο μοντέλο να προσαρμόζει δυναμικά την εστίαση σε διαφορετικά μέρη της ακολουθίας εισόδου, καταγράφοντας τοπικές και καθολικές εξαρτήσεις χωρίς να περιορίζεται από το μήκος της ακολουθίας. Αυτή η ικανότητα καθιστά την αυτοπροσοχή εξαιρετικά αποτελεσματική για εργασίες που απαιτούν κατανόηση των συμφραζόμενων σχέσεων μεταξύ λέξεων ή υπολέξεων, όπως η αυτόματη μετάφραση, η σύνοψη κειμένου και η απάντηση ερωτήσεων, αλλά και για πρόβλεψη χρονοσειρών όπως θα δείξει η παρούσα εργασία [162].

Καθώς ο μηχανισμός προσοχής και ο μηχανισμός αυτοπροσοχής έχουν πολλά κοινά μεταξύ τους, είναι ιδιαίτερα χρήσιμο να σταχυολογήσουμε τις βασικές τους διαφορές. Η κύρια διάκριση έγκειται στην εφαρμογή και την παραγωγή των  $Q$ ,  $K$  και  $V$ :

- Στους γενικούς μηχανισμούς προσοχής, τα  $Q$ ,  $K$  και  $V$  ενδέχεται να προέρχονται από διαφορετικές ακολουθίες ή στάδια επεξεργασίας (π.χ.  $Q$  από τον αποκωδικοποιητή,  $K$  και  $V$  από τον κωδικοποιητή). Αυτή η ρύθμιση είναι σημαντική για μοντέλα που πρέπει να γεφυρώνουν πληροφορίες μεταξύ διαφορετικών τμημάτων του μοντέλου, όπως η μετάφραση μεταξύ δύο γλωσσών.
- Στην αυτοπροσοχή, τα  $Q$ ,  $K$  και  $V$  υπολογίζονται από την ίδια ακολουθία στο ίδιο στάδιο επεξεργασίας. Αυτός ο μηχανισμός είναι ζωτικής σημασίας για την κατανόηση

και την επεξεργασία της εσωτερικής δομής της ίδιας της ακολουθίας, χωρίς αναφορά σε άλλη ακολουθία ή πλαίσιο.

## 6.4 Αυτοπροσοχή για Πρόβλεψη Χρονοσειρών

Για να διερευνήσουμε αναλυτικά πώς η αυτοπροσοχή μπορεί να χρησιμοποιηθεί για την πρόβλεψη χρονοσειρών, θα εμβαθύνουμε στην εφαρμογή του μηχανισμού αυτοπροσοχής, ιδιαίτερα στο πλαίσιο μοντέλων όπως ο Transformer. Έστω μια μονομεταβλητή χρονοσειρά  $X = [x_1, x_2, \dots, x_T]$ , όπου κάθε  $x_t$  αντιπροσωπεύει την παρατηρούμενη τιμή τη χρονική στιγμή  $t$ . Ο στόχος της πρόβλεψης είναι να προβλέψει τις μελλοντικές τιμές  $X_{future} = [x_{T+1}, x_{T+2}, \dots, x_{T+N}]$  με βάση την παρατηρούμενη ακολουθία. Για να εφαρμοστεί αυτοπροσοχή σε αυτό το πρόβλημα, πολλοί ερευνητές μετατρέπουν πρώτα τις χρονοσειρές σε μια κατάλληλη μορφή για επεξεργασία.

Το αρχικό βήμα συχνά περιλαμβάνει την ενσωμάτωση των χρονικών βημάτων σε έναν χώρο υψηλότερης διάστασης, ενισχύοντας την αναπαράσταση της χρονοσειράς [163]. Αυτό μπορεί να αναπαρασταθεί μαθηματικά ως:

$$E_t = \text{Embed}(x_t)$$

όπου το  $E_t$  είναι το embedding για το χρονικό βήμα  $t$  και το  $\text{Embed}(\cdot)$  είναι μια συνάρτηση, δυνητικά τόσο απλή όσο μια γραμμική προβολή.

Με τις χρονοσειρές κατάλληλα ενσωματωμένες σε ένα χώρο υψηλότερων διαστάσεων (ή και αυτούσιες ανάλογα την προσέγγιση), η αυτοπροσοχή εφαρμόζεται στην ακολουθία  $E = [E_1, E_2, \dots, E_T]$ . Ακολουθώντας τον μηχανισμό αυτοπροσοχής του Transformer, για κάθε χρονικό βήμα  $t$ , το μοντέλο υπολογίζει τους βαθμούς προσοχής που αντικατοπτρίζουν τη συνάφεια κάθε άλλου σημείου της σειράς με το τρέχον σημείο:

1. Υπολογισμός queries, keys και values για κάθε χρονικό βήμα:

$$Q_t = W^Q E_t, \quad K_t = W^K E_t, \quad V_t = W^V E_t$$

2. Υπολογισμός των βαθμών προσοχής: Χρησιμοποιείται το εσωτερικό γινόμενο των queries και keys, ακολουθούμενο από το softmax layer για να προκύψουν οι πιθανότητες:

$$\text{Attention scores}_{t,\tau} = \text{softmax} \left( \frac{Q_t \cdot K_\tau^T}{\sqrt{d_k}} \right)$$

3. Συγκεντρωτικό Άθροισμα: Παράγεται το τελικό αποτέλεσμα για κάθε χρονικό βήμα:

$$O_t = \sum_{\tau=1}^T \text{Attention scores}_{t,\tau} V_\tau$$

Εδώ, τα  $W^Q$ ,  $W^K$  και  $W^V$  είναι πίνακες βάρους για την προβολή των embeddings (ή των πραγματικών τιμών της χρονοσειράς) σε queries, keys και values, αντίστοιχα.

Η έξοδος του μηχανισμού αυτοπροσοχής,  $O = [O_1, O_2, \dots, O_T]$ , χρησιμεύει ως μια πλούσια αναπαράσταση ολόκληρης της σειράς, ενσωματώνοντας πληροφορίες από όλα τα χρονικά σημεία.

ία σταθμισμένα ανάλογα με τη συνάφειά τους. Αυτή η αναπαράσταση μπορεί στη συνέχεια να τροφοδοτηθεί σε περαιτέρω επίπεδα του μοντέλου - όπως πρόσθετα επίπεδα προσοχής, MLPs ή ακόμα και απευθείας σε ένα απλό επίπεδο νευρώνων για να δημιουργήσει τις προβλεπόμενες μελλοντικές τιμές.

$$X_{future} = \text{Decoder}(O)$$

όπου το  $\text{Decoder}(\cdot)$  αντιπροσωπεύει τα επόμενα επίπεδα επεξεργασίας που μετατρέπουν τις εξόδους προσοχής σε προβλέψεις μελλοντικών σειρών.

Η αυτοπροσοχή επιτρέπει στο μοντέλο να συσχετίζει άμεσα οποιαδήποτε δύο σημεία της χρονοσειράς, ανεξάρτητα από τη χρονική τους απόσταση, καταγράφοντας μακροπρόθεσμες εξαρτήσεις που μπορεί να χάνουν τα μοντέλα που βασίζονται σε ένα πιο βραχυπρόθεσμο πλαίσιο. Επιπλέον, οι βαθμοί προσοχής παρέχουν πληροφορίες για τον τρόπο με τον οποίο το μοντέλο εξάγει τις προβλέψεις του, υποδεικνύοντας ποιες προηγούμενες παρατηρήσεις θεωρούνται πιο σχετικές για την πρόβλεψη μελλοντικών τιμών.

## 6.5 Κωδικοποιήσεις Θέσης

Οι κωδικοποιήσεις θέσης είναι ένα κρίσιμο συστατικό στην αρχιτεκτονική των μοντέλων που χρησιμοποιούν μηχανισμούς αυτοπροσοχής, όπως ο Transformer, ο οποίος εγγενώς στερείται της ικανότητας αντίληψης της θέσης της κάθε τιμής (στοιχείου) [159]. Οι κωδικοποιήσεις θέσης εμποτίζουν αυτά τα μοντέλα με την ικανότητα αναγνώρισης και χρήσης της σειράς των παρατηρήσεων στις ακολουθίες εισόδου, η οποία είναι απαραίτητη για την επεξεργασία δεδομένων χρονοσειρών, κειμένου ή οποιασδήποτε διαδοχικής πληροφορίας.

Οι μηχανισμοί αυτοπροσοχής αντιμετωπίζουν κάθε είσοδο ανεξάρτητα και υπολογίζουν τους βαθμούς προσοχής με βάση το περιεχόμενο των εισόδων χωρίς να λαμβάνουν υπόψη τις θέσεις τους στην ακολουθία. Αν και αυτό επιτρέπει σημαντική παραλληλοποίηση και ευελιξία, σημαίνει επίσης ότι το μοντέλο από μόνο του δεν μπορεί να διακρίνει μεταξύ ακολουθιών που έχουν τα ίδια στοιχεία σε διαφορετική σειρά. Οι κωδικοποιήσεις θέσης επιλύουν αυτό το πρόβλημα προσθέτοντας πληροφορίες σχετικά με τη σχετική ή την απόλυτη θέση των στοιχείων στην ακολουθία στα embeddings προτού υποβληθούν σε επεξεργασία από τα επίπεδα προσοχής.

Κάτι αντίστοιχο προτείνει και το αρχικό μοντέλο Transformer [159]. Αυτές οι κωδικοποιήσεις έχουν την ίδια διάσταση  $d_{\text{model}}$  με τα embeddings. Η κωδικοποίηση θέσης για μια θέση  $pos$  και μια embedding-διάσταση  $i$  ορίζεται ως:

$$PE(pos, 2i) = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE(pos, 2i + 1) = \cos(pos/10000^{2i/d_{\text{model}}})$$

Αυτή η διατύπωση δημιουργεί μια μοναδική κωδικοποίηση για κάθε θέση και επιτρέπει στο μοντέλο να μάθει εύκολα να παρακολουθεί τις σχετικές θέσεις αφού, για οποιαδήποτε σταθερή μετατόπιση  $k$ , η  $PE(pos + k)$  μπορεί να αναπαρασταθεί ως γραμμική συνάρτηση των  $PE(pos)$ .

Η χρήση ημιτονοειδών συναρτήσεων για κωδικοποίηση θέσης υποκινείται από διάφορους παράγοντες:

- Αποσύνθεση συχνότητας: Οι συναρτήσεις ημιτόνου και συνημιτόνου επιτρέπουν σε κάθε διάσταση της κωδικοποίησης θέσης να αντιστοιχεί σε ένα ημιτονοειδές διαφορετικού μήκους κύματος, που κυμαίνεται από  $2\pi$  έως  $10000 \cdot 2\pi$ . Αυτό επιτρέπει στο μοντέλο να μάθει εύκολα να παρακολουθεί τις σχετικές θέσεις, καθώς η διαφορά στην κωδικοποίηση μεταξύ οποιωνδήποτε δύο θέσεων είναι συνεπής [164].
- Γραμμική προβλεψιμότητα: Οποιαδήποτε σταθερή μετατόπιση  $k$ ,  $PE(pos + k)$  μπορεί να προβλεφθεί γραμμικά από το  $PE(pos)$ , κάτι που είναι επωφελές για το μοντέλο που προσπαθεί να μάθει σχετικές θέσεις. Η γραμμική σχέση μεταξύ των θέσεων και των κωδικοποιήσεών τους διευκολύνει την εκμάθηση βαρών προσοχής που αξιοποιούν αποτελεσματικά τις πληροφορίες θέσης.

Ενώ ο αρχικός Transformer χρησιμοποιεί σταθερές, ημιτονοειδείς κωδικοποιήσεις θέσης, μεταγενέστερες εργασίες έχουν εξερευνήσει εναλλακτικές μεθόδους [164]:

- Κωδικοποιήσεις που προκύπτουν από διαδικασίες μάθησης [165]: Αντί να χρησιμοποιείται μια προκαθορισμένη μαθηματική συνάρτηση, οι κωδικοποιήσεις θέσης μπορεί να είναι παράμετροι που μαθαίνει το μοντέλο κατά τη διάρκεια της εκπαίδευσης. Αυτή η προσέγγιση αντιμετωπίζει τις θέσεις ως embeddings και μπορεί ενδεχομένως να προσαρμοστεί καλύτερα σε συγκεκριμένες εργασίες.
- Σχετικές κωδικοποιήσεις θέσης: Ορισμένες παραλλαγές μηχανισμών προσοχής, όπως αυτοί που χρησιμοποιούνται στα μοντέλα Transformer-XL [166] ή T5 [167], ενσωματώνουν σχετικές πληροφορίες θέσης απευθείας στον υπολογισμό αυτοπροσοχής, επιτρέποντας στο μοντέλο να εστιάζει στις σχετικές θέσεις του στοιχείου κατά τον υπολογισμό των βαθμών προσοχής.

Συνοπτικά, η ενσωμάτωση κωδικοποιήσεων θέσης σε μοντέλα για την πρόβλεψη χρονοσειρών αξιοποιεί τη χρονική δομή των δεδομένων, επιτρέποντας σε μοντέλα όπως τα Transformers να κατανοήσουν την αλληλουχία των γεγονότων και τις σχετικές θέσεις τους στο χρόνο. Αυτή η προσέγγιση είναι κρίσιμη επειδή η σειρά των σημείων δεδομένων σε μια χρονοσειρά επηρεάζει άμεσα το αποτέλεσμα της πρόβλεψης.





## Κεφάλαιο **7**

### Πειραματική Διαδικασία

---

#### 7.1 Δεδομένα και Προεπεξεργασία

Όπως έχει αναφερθεί και στην εισαγωγή της εργασίας, η παρούσα μελέτη έχει ως στόχο την ανάπτυξη και σύγκριση διαφόρων νευρωνικών αρχιτεκτονικών που χρησιμοποιούν μηχανισμούς αυτοπροσοχής για την πρόβλεψη του ηλεκτρικού φορτίου. Η ανάλυση των επιδόσεων των διαφόρων αρχιτεκτονικών διεξήχθη με τη χρήση ενός συνόλου δεδομένων που περιλαμβάνει το συνολικό ηλεκτρικό φορτίο του συστήματος, σε ωριαία βάση, από 24 διαφορετικές ευρωπαϊκές χώρες. Τα δεδομένα αυτά προήλθαν από τον αντίστοιχο Διαχειριστή κάθε χώρας και είναι διαθέσιμα μέσω της πλατφόρμας του Ευρωπαϊκού Δικτύου Διαχειριστών Συστημάτων Μεταφοράς Ηλεκτρικής Ενέργειας (ENTSO-E).

Πιο συγκεκριμένα, η παρούσα έρευνα χρησιμοποίησε ιστορικά δεδομένα φορτίου που εκτείνονται από τον Ιανουάριο του 2016 έως τον Δεκέμβριο του 2021. Αυτή η περίοδος χρησιμοποιήθηκε για την εκπαίδευση των διαφόρων νευρωνικών μοντέλων. Για την εξέταση της ακρίβειας πρόβλεψης των μοντέλων χρησιμοποιήθηκαν δεδομένα από το 2022.

Όπως γίνεται αντιληπτό, ένα σύνολο δεδομένων τέτοιας έκτασης χρειάζεται προεργασία για να χρησιμοποιηθεί αποδοτικά κατά τη διάρκεια της υπόλοιπης έρευνας.

Η αρχική προετοιμασία των δεδομένων περιλάμβανε την προσαρμογή της συχνότητας των δεδομένων, όπου ήταν απαραίτητο. Για τις χώρες που παρέχουν δεδομένα για το φορτίο του συστήματός τους ανά 15 λεπτά ή 30 λεπτά της ώρας, οι πληροφορίες αυτές συγκεντρώθηκαν ώστε να αντιστοιχούν σε ωριαίο ρυθμό. Αυτό το βήμα εξασφάλισε την ύπαρξη συνοχής σε όλο το σύνολο των δεδομένων, επιτρέποντας μια τυποποιημένη προσέγγιση για τη μετέπειτα επεξεργασία.

Δεδομένης της εκτεταμένης χρονικής περιόδου (έξι έτη) και του εύρους (24 χώρες) των δεδομένων, ήταν αναμενόμενο να προκύψουν ορισμένα προβλήματα ποιότητας των δεδομένων, όπως η ύπαρξη διαφόρων ελλειπόντων ή αλλοιωμένων τιμών στις υπό εξέταση χρονοσειρές. Ως εκ τούτου, υιοθετήθηκε μια μεθοδική προσέγγιση για τον καθαρισμό των δεδομένων:

Για χρονικές ακολουθίες όπου έλειπαν λιγότερα από έξι διαδοχικά σημεία δεδομένων, χρησιμοποιήθηκαν μέθοδοι παρεμβολής για την εκτίμηση των ελλειπόντων τιμών με βάση τις πιο κοντινές παρατηρήσεις. Εάν το κενό υπερέβαινε αυτό το όριο, τα δεδομένα για τη συγκεκριμένη περίοδο αποκλείονταν από το σύνολο δεδομένων για να διατηρηθεί η ακεραιότητα της ποιότητας των δεδομένων.

Επιπλέον, οι ακραίες τιμές που οφείλονται σε λανθασμένες μετρήσεις του φορτίου μπορούν

να αλλοιώσουν τα αποτελέσματα της προγνωστικής μοντελοποίησης. Για να μετριάσει αυτό, τα σημεία δεδομένων που έπεφταν εκτός των εκατοστημορίων 0,1% και 99,9% των τιμών της κάθε χρονοσειράς αναγνωρίστηκαν ως λανθασμένες ακραίες τιμές και περικλύθηκαν σε αυτά τα όρια. Αυτή η προσέγγιση διαχειρίστηκε αποτελεσματικά τις λανθασμένες τιμές χωρίς να διακυβεύεται η συνολική μεταβλητότητα του συνόλου δεδομένων.

Προχωρώντας στα μετέπειτα στάδια της αρχικής επεξεργασίας των δεδομένων των χρονοσειρών, αξίζει να αναφερθεί πως η συνέπεια στην κλίμακα των τιμών είναι ζωτικής σημασίας κατά την εκπαίδευση νευρωνικών δικτύων, ιδίως σε περιπτώσεις όπου τα δεδομένα προέρχονται από πολλαπλές πηγές και διαφέρουν σημαντικά σε μέγεθος, όπως δηλαδή γίνεται και στην περίπτωση των υπό εξέταση χρονοσειρών, τα οποία προέρχονται από διαφορετικούς παρόχους και δίκτυα από όλη την Ευρώπη. Για να αντιμετωπιστεί αυτό, τα δεδομένα φορτίου κάθε χώρας κανονικοποιήθηκαν: Αρχικά υπολογίστηκαν ο μέσος όρος και η τυπική απόκλιση κάθε σειράς χρησιμοποιώντας δεδομένα έως και τον Δεκέμβριο του 2021, έτσι ώστε να μην μεταφερθούν πληροφορίες από το σύνολο δεδομένων αξιολόγησης στο σύνολο δεδομένων εκπαίδευσης. Στη συνέχεια, κάθε τιμή των χρονοσειρών τυποποιήθηκε αφαιρώντας τη μέση τιμή και διαιρώντας με την τυπική απόκλιση. Αυτός ο μετασχηματισμός αντιστράφηκε στις εξόδους του μοντέλου για να αποδώσουν τις προβλέψεις στην αρχική τους κλίμακα.

Αφού τα δεδομένα μετατράπηκαν σε μορφή ικανή να τροφοδοτηθούν σε ένα νευρωνικό δίκτυο, το τελευταίο βήμα είναι ο χωρισμός σε σύνολο εκπαίδευσης και αξιολόγησης, το οποίο παρουσιάζεται παρακάτω αναλυτικότερα.

Πιο συγκεκριμένα, εφαρμόστηκε η μεθοδολογία ολισθαίνοντος παραθύρου για τη δημιουργία συνόλων δεδομένων τόσο εκπαίδευσης όσο και δοκιμής, εξασφαλίζοντας ότι τα μοντέλα θα αξιολογηθούν υπό συνθήκες που προσομοιάζουν τις πραγματικές επιχειρησιακές συνθήκες. Τα παράθυρα δεδομένων εκπαίδευσης αποτελούνται από 204 διαδοχικές τιμές ωριαίων δεδομένων για το ηλεκτρικό φορτίο. Από αυτές, οι πρώτες 168 ώρες χρησιμοποιήθηκαν για την πρόβλεψη των επόμενων 36 ωρών. Αυτή η επιλογή αντικατοπτρίζει δεδομένα μιας εβδομάδας που χρησιμοποιούνται για την πρόβλεψη περίπου μιάμισης μέρας μπροστά. Τα δεδομένα αξιολόγησης δομήθηκαν παρομοίως, αλλά χρησιμοποιούν ένα μεγαλύτερο βήμα 24 ωρών μεταξύ των παραθύρων για να μιμηθούν τις καθημερινές δραστηριότητες πρόβλεψης.

Εν κατακλείδι, προετοιμάστηκαν συνολικά 101.562 δείγματα εκπαίδευσης, ενώ επιπλέον 8.760 δείγματα προορίστηκαν για την αξιολόγηση των μοντέλων που κατασκευάστηκαν.

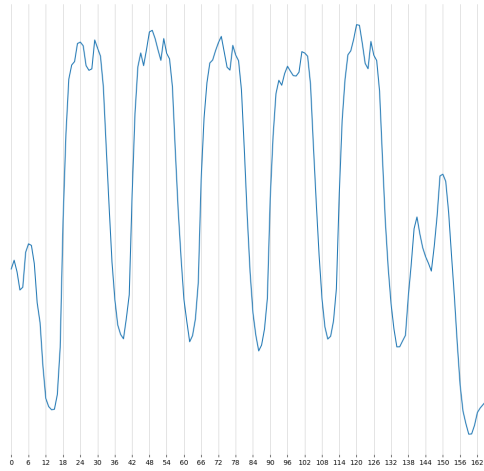
Τα δεδομένα παρουσιάζουν εν γένει έντονη εποχικότητα, όπως φαίνεται στην απεικόνιση μιας τυπικής εβδομάδας στην Πολωνία, που παρουσιάζεται στην εικόνα 7.1.

Στις περισσότερες χώρες, όπως στην Πολωνία, μπορεί κανείς να διακρίνει εύκολα τις πέντε καθημερινές αλλά και την Κυριακή (πρώτη μέρα με μειωμένη κατανάλωση στην εικόνα 7.1) και το Σάββατο (τελευταία μέρα με μειωμένη κατανάλωση στην εικόνα 7.1).

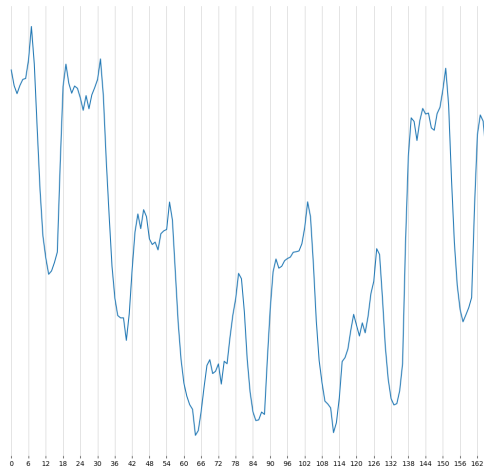
Φυσικά, σε όλες τις χώρες, παρατηρούνται και έντονες διαταράξεις της παραπάνω καθαρής εποχικής συμπεριφοράς λόγω ύπαρξης ειδικών γεγονότων, των οποίων η ύπαρξη αξίζει ιδιαίτερης προσοχής και μελέτης καθώς, όπως θα φανεί στην επόμενη ενότητα, επηρεάζει σημαντικά την ακρίβεια των προβλέψεων.

Στην εικόνα 7.2 παρατηρούμε την επίδραση των εορτασμών για την ημέρα του Συντάγματος της 3ης Μαΐου στην Πολωνία (η συγκεκριμένη εβδομάδα είναι από το έτος 2022).

Ωστόσο, η εμφάνιση διαταραχών στην εποχική συμπεριφορά των δεδομένων δεν είναι μόνο

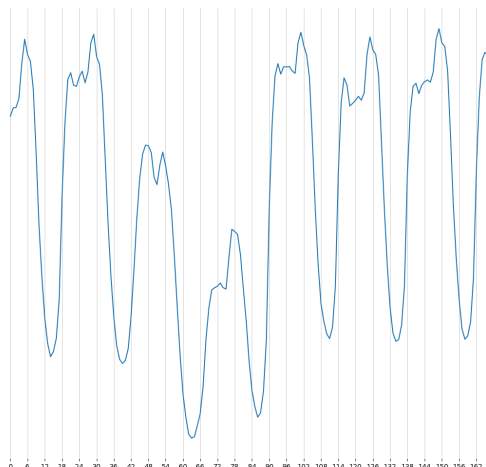


Εικόνα 7.1: Μία τυπική εβδομάδα στην Πολωνία



Εικόνα 7.2: Η επίδραση της ημέρας του Συντάγματος της 3ης Μαΐου στην κατανάλωση ρεύματος στην Πολωνία.

συνάρτηση της αντίστοιχης ημέρας και μήνα του ειδικού γεγονότος. Στην εικόνα 7.3 βλέπουμε την κατανάλωση ενέργειας κατά την εθνική αργία του εορτασμού της ανεξαρτησίας της Πολωνίας (11 Νοεμβρίου) το 2022. Ωστόσο, η συγκεκριμένη αργία το 2022 δεν επηρεάζει το σύνηθες μοτίβο κατανάλωσης στην Πολωνία, για το έτος 2022. Αυτό συμβαίνει καθώς η συγκεκριμένη γιορτή συνέπεσε χρονικά, εκείνο το έτος με Σαββατοκύριακο, οπότε η κατανάλωση ήταν έτσι και αλλιώς μειωμένη.



Εικόνα 7.3: Οι εορτασμοί της ημέρας Ανεξαρτησίας συνέπεσαν με Σαββατοκύριακο το 2022 στην Πολωνία.

## 7.2 Υπολογιστικοί Πόροι και Πειραματική Διάταξη

Όπως γίνεται αντιληπτό, το σύνολο δεδομένων που παρουσιάστηκε παραπάνω προσφέρει μία στέρεα βάση για την εξερεύνηση της προγνωστικής ικανότητας διαφόρων νευρωνικών αρχιτεκτονικών. Για την εξερεύνηση αυτή χρησιμοποιήθηκε μία V-100 GPU με 16GB RAM, καθώς και μία CPU με 46GB RAM.

Επιπλέον, αξίζει να σημειωθεί πως στα πειράματα που θα περιγραφούν στις επόμενες παραγράφους εκπαιδεύονταν πολλαπλά μοντέλα για κάθε αρχιτεκτονική. Πιο συγκεκριμένα, όπως θα φανεί και παρακάτω, καθώς για την αρχικοποίηση των παραμέτρων των αρχιτεκτονικών χρησιμοποιούνται τυχαίες διαδικασίες, μία αρχιτεκτονική που χρησιμοποιεί τις ίδιες υπερπαραμέτρους μπορεί να δώσει δύο εντελώς διαφορετικά αποτελέσματα αν εκπαιδευτεί δύο φορές. Για αυτό το λόγο, για κάθε αρχιτεκτονική εκπαιδεύονται 30 διαφορετικά μοντέλα και έπειτα, κατά την παραγωγή προβλέψεων, προκύπτει ο μέσος όρος των προβλέψεων των μοντέλων. Το συγκεκριμένο πλήθος μοντέλων επιλέχθηκε έπειτα από αρκετά πειράματα, όπως θα παρουσιαστεί αναλυτικά και στην ενότητα των αποτελεσμάτων.

Κατά την εκπαίδευση του κάθε μοντέλου χρησιμοποιούταν η GPU για μεγαλύτερη ταχύτητα εκπαίδευσης, ενώ για την παραγωγή προβλέψεων και την αξιολόγηση αυτών χρησιμοποιήθηκε η CPU λόγω της υπεροχής της σε χώρο στη RAM.

## 7.3 Νευρωνικές Αρχιτεκτονικές

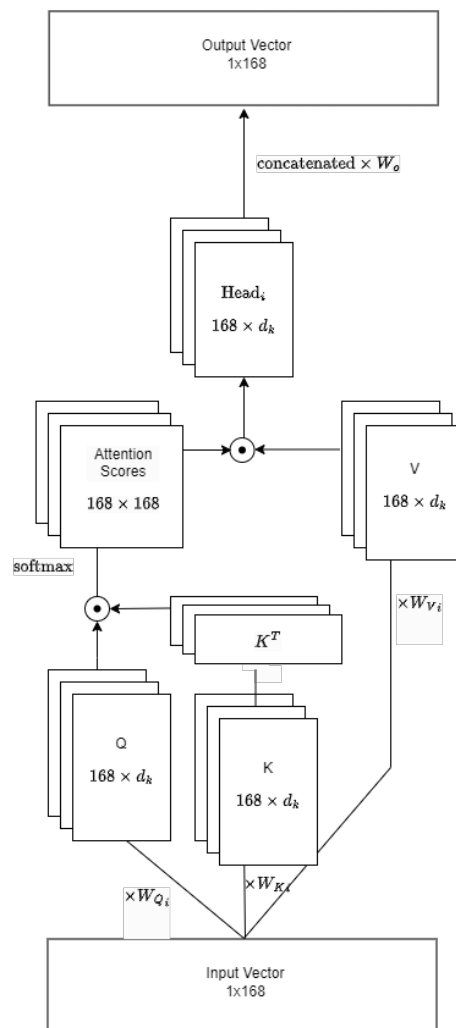
Η συγκεκριμένη εργασία εξέτασε και χρησιμοποίησε για την πρόβλεψη των χρονοσειρών που αναλύθηκαν παραπάνω μία πληθώρα από αρχιτεκτονικές που χρησιμοποιούν μηχανισμούς αυτοπροσοχής. Φυσικά, για να εξαχθούν χρήσιμα συμπεράσματα και να προκύψουν τα βέλτιστα δυνατά αποτελέσματα, η πειραματική διαδικασία χαρακτηρίστηκε από μία συγκεκριμένη οργανωτική δομή: Αρχικά εξετάστηκαν αρχιτεκτονικές οι οποίες χρησιμοποιούν αποκλειστι-

κά μηχανισμό προσοχής. Έπειτα, ερευνήθηκε η χρήση διαφόρων αρχιτεκτονικών πριν τον μηχανισμό προσοχής και πως μπορούν να συνδεθούν κατάλληλα, ενώ στο τέλος έγινε και μία ενδελεχής έρευνα για το τι θα μπορούσε να ακολουθεί μετά τον μηχανισμό αυτοπροσοχής.

### 7.3.1 Αποκλειστικά Μηχανισμοί Προσοχής

Οι πρώτες αρχιτεκτονικές που δοκιμάστηκαν εστίασαν στη λειτουργικότητα του μηχανισμού προσοχής αυτού καθ'αυτού, δηλαδή οι παρελθοντικές τιμές μίας εβδομάδας που χρησιμοποιούνται για την πρόβλεψη της επόμενης μιάμισης μέρας τροφοδοτήθηκαν απευθείας στον μηχανισμό προσοχής. Η έξοδος του μηχανισμού προσοχής συνδέεται έπειτα με ένα πλήρως συνδεδεμένο επίπεδο νευρώνων οι οποίοι απαραίτητως χρειάζονται για τον σχηματισμό του τελικού διανύσματος προβλέψεων.

Ως προς τον μηχανισμό προσοχής που χρησιμοποιήθηκε σε αυτό το σημείο της πειραματικής διαδικασίας αναπτύχθηκε ένας απλός μηχανισμός αυτοπροσοχής (scaled dot-product self-attention), ενώ εξετάστηκαν και μηχανισμοί αυτοπροσοχής με πολλαπλές κεφαλές.



Σχήμα 7.1: Το επίπεδο προσοχής, έτσι όπως κατασκευάστηκε στην παρούσα εργασία.

Πιο συγκεκριμένα, έστω το διάνυσμα εισόδου  $X$ , όπου  $X \in \mathbb{R}^{1 \times 168}$ . Το  $X$  προβάλλεται γραμμικά σε τρεις διαφορετικούς χώρους για να ληφθούν τα  $Q$ ,  $K$  και  $V$  χρησιμοποιώντας

πίνακες βαρών:

$$Q = XW^Q, \quad K = XW^K, \quad V = XW^V$$

Στον συγκεκριμένο μηχανισμό αυτοπροσοχής που αναπτύχθηκε στην παρούσα εργασία ισχύει πως  $W^Q, W^K, W^V \in \mathbb{R}^{1 \times d_k}$ , όπου το  $d_k$  αντιπροσωπεύει το μέγεθος του πίνακα βαρών  $K$ . Συνεπώς προκύπτει ότι  $Q, K, V \in \mathbb{R}^{168 \times d_k}$ . Τα βάρη προσοχής υπολογίζονται μετέπειτα ως εξής:

$$\text{Attention}_{\text{weights}}(Q, K) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right)$$

όπου  $\text{Attention}_{\text{weights}} \in \mathbb{R}^{168 \times 168}$ . Η τελική έξοδος του απλού μηχανισμού αυτοπροσοχής προκύπτει ως εξής:

$$\text{Self-Attention}(Q, K, V) = \text{Attention}_{\text{weights}} V$$

όπου  $\text{Self-Attention}(Q, K, V) \in \mathbb{R}^{168 \times d_v}$ . Το  $d_v$  αναπαριστά το μέγεθος του πίνακα  $V$ , το οποίο στην παρούσα εργασία επιλέγεται ίσο με το  $d_k$ .

Ως προς την επέκταση του απλού μηχανισμού αυτοπροσοχής σε μηχανισμό πολλαπλών κεφαλών, ισχύουν τα παρακάτω:

$$\text{MultiHead Self-Attention}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

$$\text{όπου } \text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

Αξίζει να σημειωθεί πως στην παρούσα εργασία το  $W^O \in \mathbb{R}^{h \cdot d_v \times 1}$ , ενώ  $\text{MultiHead Self-Attention}(Q, K, V) \in \mathbb{R}^{168 \times 1}$ . Συνεπώς, για το υπόλοιπο της παρούσας εργασίας, όταν γίνεται αναφορά σε μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών, θα εννοείται ο παραπάνω μηχανισμός με τις συγκεκριμένες διαστάσεις.

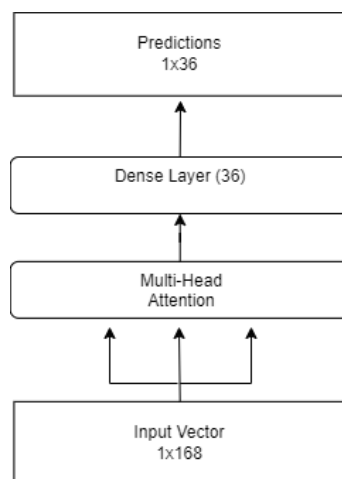
Τέλος, για την εξαγωγή των τελικών προβλέψεων χρησιμοποιείται το απαραίτητο πλήρως συνδεδεμένο επίπεδο νευρώνων, οπότε:

$$F = \text{ReLU}(W^{(\text{layer})} \cdot \text{MultiHead Self-Attention}(Q, K, V)), \quad W^{(\text{layer})} \in \mathbb{R}^{36 \times 168}$$

Οι παραπάνω αρχιτεκτονικές συνοψίζονται και στο διάγραμμα 7.2.

Ανάμεσα στις παραμέτρους που εξετάστηκαν στις παραπάνω αρχιτεκτονικές είναι το μέγεθος των πινάκων query, key και value ( $d_k$ ) αλλά και το πλήθος των κεφαλών του μηχανισμού αυτοπροσοχής. Αυτή η οικογένεια αρχιτεκτονικών θα ονομάζεται από εδώ και πέρα Basic Attention Models.

Τα συγκριτικά αποτελέσματα των αρχιτεκτονικών αυτών θα παρουσιαστούν στην επόμενη ενότητα, ωστόσο, για να αιτιολογηθεί η χρήση του μηχανισμού αυτοπροσοχής πολλαπλών κεφαλών στις επόμενες αρχιτεκτονικές που θα παρουσιαστούν, αξίζει να σημειωθεί σε αυτό το σημείο πως οι πιο ακριβείς προβλέψεις σημειώθηκαν με χρήση πολλαπλών κεφαλών, και μάλιστα 6.



Σχήμα 7.2: Η βασική αρχιτεκτονική προσοχής πάνω στην οποία χτίστηκαν οι υπόλοιπες.

### 7.3.2 Αρχιτεκτονικές που προηγούνται του Μηχανισμού Προσοχής

Η εξέταση της ικανότητας των νευρωνικών δικτύων με μηχανισμούς προσοχής στην πρόβλεψη μελλοντικών τιμών ηλεκτρικού φορτίου συνεχίστηκε με την εξερεύνηση και την υλοποίηση διαφόρων αρχιτεκτονικών που μπορούν να προηγούνται του μηχανισμού προσοχής. Η χρήση τέτοιων αρχιτεκτονικών αναμενόταν και είναι απαραίτητη για την παραγωγή προβλέψεων μεγαλύτερης ακρίβειας. Όπως μπορεί να γίνει αντιληπτό από τις προηγούμενες θεωρητικές ενότητες, οι μηχανισμοί προσοχής έχουν την ικανότητα να αντιληφθούν σε εξαιρετικό βαθμό σε ποιες παρελθοντικές τιμές της χρονοσειράς πρέπει να δοθεί περισσότερη προσοχή από το μοντέλο για την πρόβλεψη των επόμενων τιμών, ωστόσο δεν έχει ιδιαίτερη αντίληψη των σχετικών τους θέσεων. Για τον λόγο αυτό, απαιτείται η χρήση μίας αρχιτεκτονικής η οποία θα εμπλουτίζει την είσοδο του μηχανισμού προσοχής με πληροφορία που αφορά τη θέση των τιμών μέσα στη χρονοσειρά.

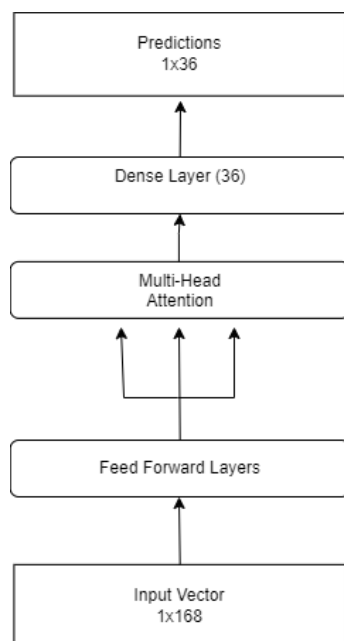
#### MLP πριν τον Μηχανισμό Προσοχής

Για το σκοπό αυτό, αρχικά χρησιμοποιήθηκε ένα MLP. Η κανονικοποιημένη είσοδος τροφοδοτήθηκε στην πρώτη στρώση νευρώνων και η έξοδος του MLP τροφοδοτήθηκε στον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών που περιγράφηκε στην προηγούμενη υποενότητα. Πιο συγκεκριμένα, η αρχική είσοδος  $X$  τροφοδοτείται σε ένα MLP του οποίου κάθε στρώση ( $l$ ) είναι της μορφής:

$$h^{(l)} = \text{ReLU}(W^{(l)}h^{(l-1)})$$

όπου  $h^{(l)} \in \mathbb{R}^{1 \times 168}$ . Τέλος, το αποτέλεσμα του MLP  $h^{(L)} \in \mathbb{R}^{1 \times 168}$  (από την τελευταία στρώση  $L$ ) τροφοδοτείται όπως και προηγουμένως στον μηχανισμό προσοχής πολλαπλών κεφαλών, ενώ η έξοδος του μηχανισμού αυτοπροσοχής πολλαπλών κεφαλών τροφοδοτείται στο απαραίτητο τελικό στρώμα νευρώνων για την εξαγωγή των προβλέψεων, όπως φαίνεται στο διάγραμμα 7.3.

Ανάμεσα στις παραμέτρους των οποίων εξερευνήθηκε η επίδραση στην ακρίβεια των προβλέψεων είναι το πλήθος των στρωμάτων νευρώνων στο Multi-Layer Perceptron. Τα αποτελέσματα των πειραμάτων αυτών θα παρουσιαστούν αναλυτικά στην επόμενη ενότητα, ενώ



Σχήμα 7.3: Η οικογένεια αρχιτεκτονικών MLP Attention Models

αυτή η οικογένεια αρχιτεκτονικών θα ονομάζεται από εδώ και πέρα MLP Attention Models.

Επιπλέον, η παραπάνω αρχιτεκτονική εμπλουτίστηκε και με υπολειμματικές συνδέσεις (residual connections) ανάμεσα στα πλήρως συνδεδεμένα επίπεδα νευρώνων. Μία εκδοχή αυτής της αρχιτεκτονικής παρουσιάζεται στο διάγραμμα 7.4, ενώ η συγκριτική μελέτη της προγνωστικής ικανότητας των MLP Attention Models με ή χωρίς υπολειμματικές συνδέσεις θα παρουσιαστεί στην επόμενη ενότητα.

Ένα παράδειγμα χρήσης υπολειμματικών συνδέσεων μεταξύ των επιπέδων του πλήρως συνδεδεμένου δικτύου παρουσιάζεται στην παρακάτω εικόνα.

### Κωδικοποιήσεις θέσης πριν τον Μηχανισμό Προσοχής

Ένας ακόμα μηχανισμός που δοκιμάστηκε πριν το επίπεδο της αυτοπροσοχής πολλαπλών κεφαλών είναι οι κωδικοποιήσεις θέσης. Όπως παρουσιάστηκε σε προηγούμενη ενότητα, στην αρχική αρχιτεκτονική του Transformer η κωδικοποίηση θέσης για μια θέση  $pos$  και μια embedding-διάσταση  $i$  ορίζεται ως:

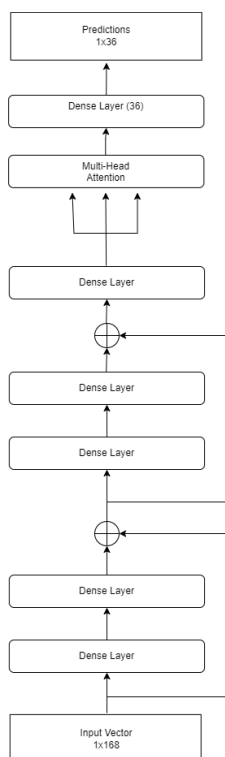
$$PE(pos, 2i) = \sin(pos/10000^{2i/d_{\text{model}}})$$

$$PE(pos, 2i + 1) = \cos(pos/10000^{2i/d_{\text{model}}})$$

Ωστόσο, στην περίπτωση των προβλέψεων χρονοσειρών, τα embeddings είναι μονοδιάστατα, δηλαδή η είσοδος  $X \in \mathbb{R}^{1 \times 168}$ . Συνεπώς, δεν έχει νόημα να εισάγουμε την έννοια των embeddings στον υπολογισμό των κωδικοποιήσεων θέσης. Ως εκ τούτου, στην παρούσα μελέτη, ο παραπάνω τύπος για την εξαγωγή των κωδικοποιήσεων θέσης μετασχηματίστηκε και χρησιμοποιήθηκε ως εξής:

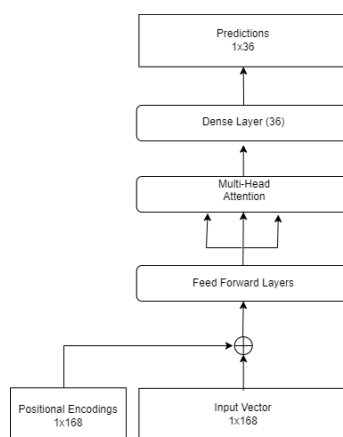
$$PE(pos) = \sin(pos/1000)$$





Σχήμα 7.4: Η οικογένεια αρχιτεκτονικών MLP Attention Models εμπλουτισμένη με υπολειματικές συνδέσεις

Με αυτόν τον τρόπο, εμπλουτίστηκε η είσοδος με πληροφορία θέσης και το εμπλουτισμένο διάνυσμα τροφοδοτήθηκε αρχικά στον μηχανισμό προσοχής ενώ μετέπειτα δοκιμάστηκε και η τροφοδότηση του εμπλουτισμένου διανύσματος σε ένα MLP και έπειτα στον μηχανισμό προσοχής.

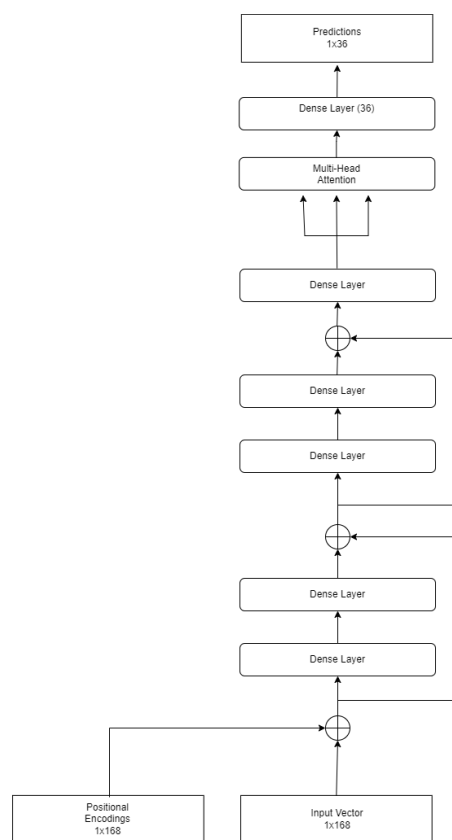


Σχήμα 7.5: Η οικογένεια αρχιτεκτονικών Positional Attention Models

Όπως και παραπάνω εξετάστηκε λεπτομερώς ο βέλτιστος αριθμός στρώσεων νευρώνων και τα αποτελέσματα θα παρουσιαστούν αναλυτικά στην επόμενη ενότητα, ενώ από εδώ και πέρα αυτή η οικογένεια αρχιτεκτονικών θα ονομάζεται Positional Attention Models (σχεδιάγραμμα 7.5).

Παράλληλα, και στην περίπτωση αυτής της αρχιτεκτονικής, όπως και προηγουμένως, δοκι-

μάστηκε και η χρήση υπολειμματικών συνδέσεων μεταξύ των επιπέδων πλήρως συνδεδεμένων νευρώνων. Μία μορφή της παραπάνω αρχιτεκτονικής παρουσιάζεται στο σχεδιάγραμμα 7.6.



Σχήμα 7.6: Η οικογένεια αρχιτεκτονικών *Positional Attention Models* εμπλουτισμένη με υπολειμματικές συνδέσεις

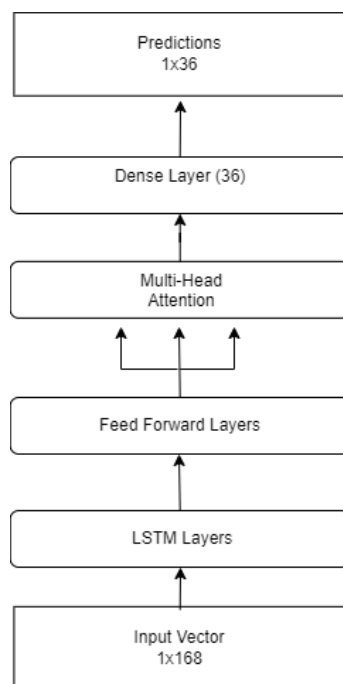
## LSTM πριν τον Μηχανισμό Προσοχής

Ωστόσο, η εξερεύνηση των αρχιτεκτονικών που θα μπορούσαν να προηγηθούν του μηχανισμού προσοχής έτσι ώστε να δώσουν πληροφορία θέσης δεν σταμάτησε στις κωδικοποιήσεις θέσης, καθώς δοκιμάστηκε και η χρήση LSTM.

Η ακριβής αρχιτεκτονική και ο τρόπος λειτουργίας των LSTM γενικότερα, παρουσιάστηκε σε προηγούμενη ενότητα. Εμβαθύνοντας στη χρήση LSTM στην συγκεκριμένη περίπτωση, το διάνυσμα εισόδου  $X \in \mathbb{R}^{1 \times 168}$  τροφοδοτείται στο LSTM το οποίο αποτελείται από  $h$  μονάδες (cells). Το διάνυσμα  $\text{lstm}_{out} \in \mathbb{R}^{1 \times h}$  που προκύπτει στην έξοδο του LSTM είναι η τελευταία κρυφή κατάσταση  $h_t = o_t * \tanh(c_t)$  που προκύπτει, με τρόπο που έχει εξηγηθεί αναλυτικά σε προηγούμενη ενότητα. Τέλος, το διάνυσμα  $\text{lstm}_{out}$  τροφοδοτείται στον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών όπως έχει περιγραφεί και παραπάνω, ενώ πάντα ακολουθεί το απαραίτητο στρώμα πλήρως συνδεδεμένων νευρώνων έτσι ώστε να προκύψουν οι τελικές 36 προβλέψεις. Εδώ αξίζει να σημειωθεί πως σε αυτή την οικογένεια μοντέλων εξετάστηκαν και αρχιτεκτονικές οι οποίες περιλαμβάνουν περισσότερα επίπεδα LSTM πριν τον μηχανισμό προσοχής, με την έξοδο του ενός να τροφοδοτεί την είσοδο του άλλου. Ωστόσο, το τελικό διάνυσμα το οποίο τροφοδοτείται στον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών παραμένει το ίδιο ( $\text{lstm}_{out} \in \mathbb{R}^{1 \times h}$ ) καθώς όλα τα επίπεδα LSTM περιέχουν τον ίδιο αριθμό από

cells.

Φυσικά, ένα από τα χαρακτηριστικά του LSTM που εξετάστηκε ενδελεχώς ως προς την επίδρασή του στην ακρίβεια των προβλέψεων είναι το πλήθος των μονάδων (cells) καθώς και η επίδραση αυτού στον χρόνο εκπαίδευσης της συνολικής αρχιτεκτονικής, ενώ δοκιμάστηκε και η επίδραση των πλήρως συνδεδεμένων επιπέδων νευρώνων μετά το επίπεδο LSTM και πριν τον μηχανισμό προσοχής στην προγνωστική ικανότητα του μοντέλου.



Σχήμα 7.7: Η οικογένεια αρχιτεκτονικών LSTM Attention Models

Η παραπάνω οικογένεια αρχιτεκτονικών θα ονομάζεται πλέον LSTM Attention Models και τα αποτελέσματα που αντανακλούν την προγνωστική της ικανότητα θα παρουσιαστούν αναλυτικά στην επόμενη ενότητα.

## CNN πριν τον Μηχανισμό Προσοχής

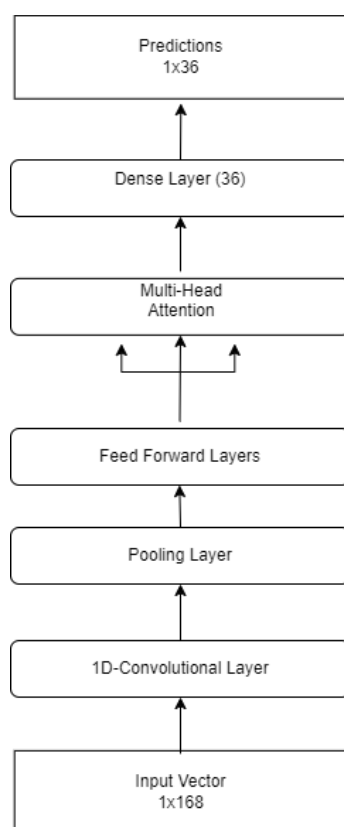
Μία ακόμα αρχιτεκτονική που δοκιμάστηκε πριν τον μηχανισμό προσοχής είναι τα συνελικτικά επίπεδα. Σε αυτή την περίπτωση, το διάνυσμα εισόδου  $X \in \mathbb{R}^{1 \times 168}$  εισέρχεται στο πρώτο συνελικτικό επίπεδο το οποίο χαρακτηρίζεται από πυρήνες μεγέθους  $k$  (kernel size) και από  $f$  στο πλήθος φίλτρα.

Στην έξοδο του συνελικτικού επιπέδου προκύπτει διάνυσμα  $c_{out} \in \mathbb{R}^{1 \times (l_{in}-k+1) \times f}$ , όπου  $l_{in}$  είναι το μέγεθος του διανύσματος που εισέρχεται στο συνελικτικό επίπεδο. Έπειτα από κάθε συνελικτικό επίπεδο ακολουθεί ένα pooling επίπεδο, το οποίο δέχεται ως είσοδο το διάνυσμα  $c_{out}$  και προκύπτει το διάνυσμα  $p_{out} \in \mathbb{R}^{1 \times \lfloor \frac{(l_{in}-k+1)}{p} \rfloor \times f}$ .

Αυτό είτε ανατροφοδοτείται σε ένα νέο ζεύγος συνελικτικού και pooling επιπέδου είτε επιπεδοποιείται και τροφοδοτείται σε ένα MLP πριν εισέλθει στον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών. Αξίζει να σημειωθεί πως στην περίπτωση της ανατροφοδότησης σε επιπλέον συνελικτικά επίπεδα, η τελευταία διάσταση, αυτή που αντιστοιχεί στον αριθμό των

φίλτρων διατηρεί σταθερό το μέγεθος της, καθώς ο αριθμός των φίλτρων που χρησιμοποιούνται είναι ο ίδιος σε όλα τα συνελικτικά επίπεδα.

Κατά την εξέταση της προγνωστικής ικανότητας της παραπάνω αρχιτεκτονικής σε συνδυασμό με τον μηχανισμό προσοχής, δοκιμάστηκαν διάφορες εκδοχές της. Εξετάστηκαν διαφορετικά μεγέθη πυρήνα, μία πληθώρα αριθμού φίλτρων καθώς και διάφορα πλήρως συνδεδεμένα επίπεδα νευρώνων μετά τα συνελικτικά επίπεδα και πριν τον μηχανισμό προσοχής.



Σχήμα 7.8: Η οικογένεια αρχιτεκτονικών *Convolutional Attention Models*

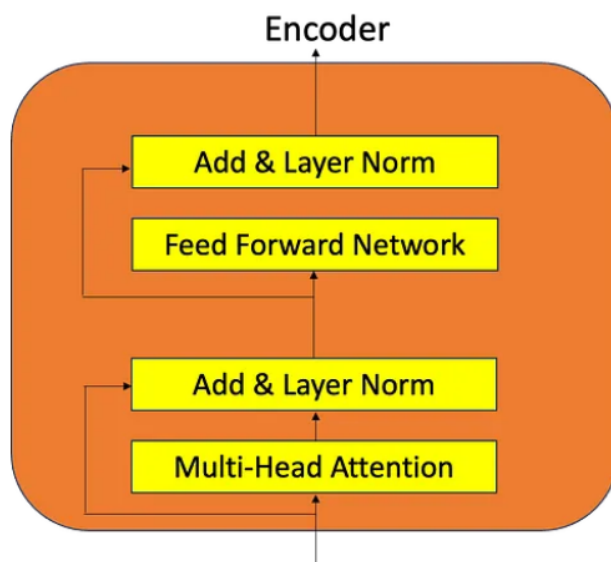
Η παραπάνω οικογένεια αρχιτεκτονικών θα ονομάζεται πλέον *Convolutional Attention Models* και τα αποτελέσματα που αντανακλούν την προγνωστική της ικανότητα θα παρουσιαστούν αναλυτικά στην επόμενη ενότητα.

### 7.3.3 Αρχιτεκτονικές που έπονται του Μηχανισμού Προσοχής

Στην παραπάνω υποενότητα έγινε μία αναλυτική παρουσίαση των μηχανισμών και αρχιτεκτονικών που θεωρήθηκε σκόπιμο να εξεταστούν και να υλοποιηθούν πριν από τον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών. Η παρουσίαση της προγνωστικής ικανότητας της κάθε αρχιτεκτονικής θα γίνει στην επόμενη ενότητα, ωστόσο, για να αιτιολογηθεί η χρήση τους από αυτό το σημείο και έπειτα, αξίζει να σημειωθεί πως το μικρότερο σφάλμα πρόβλεψης προέκυψε από τη χρήση της οικογένειας μοντέλων *Positional Attention Models* και *Convolutional Attention Models*.

### Χρήση υπολειμματικής σύνδεσης μετά το μηχανισμό προσοχής

Στην αρχιτεκτονική του Transformer που παρουσιάστηκε στη δημοσίευση “Attention is all You Need”, ο μηχανισμός προσοχής πλαισιωνόταν από μία υπολειμματική σύνδεση, όπως μπορεί να φανεί στην εικόνα 7.4:



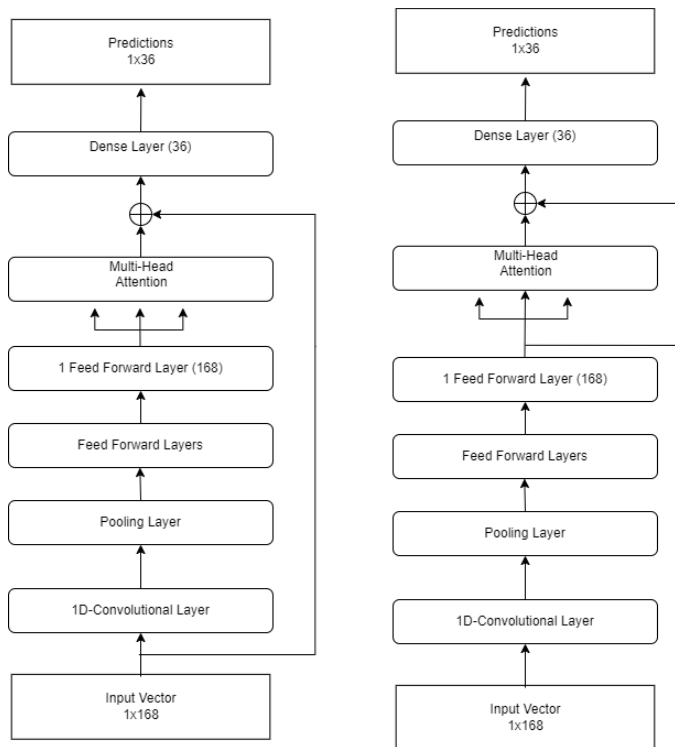
Εικόνα 7.4: Η υπολειμματική σύνδεση που χρησιμοποιείται στον αρχικό Transformer

Πιο συγκεκριμένα, προστίθεται η είσοδος του μηχανισμού αυτοπροσοχής με την έξοδο αυτού και προωθείται σε ένα πλήρως συνδεδεμένο δίκτυο. Χρησιμοποιώντας αυτή την ιδέα πάνω στις καλύτερες από τις προηγούμενες αρχιτεκτονικές προκύπτουν νέα μοντέλα. Αρχικά χρησιμοποιούμε την οικογένεια Convolutional Attention Models. Οι νέες αρχιτεκτονικές που προκύπτουν παρουσιάζονται στο σχεδιάγραμμα 7.9:

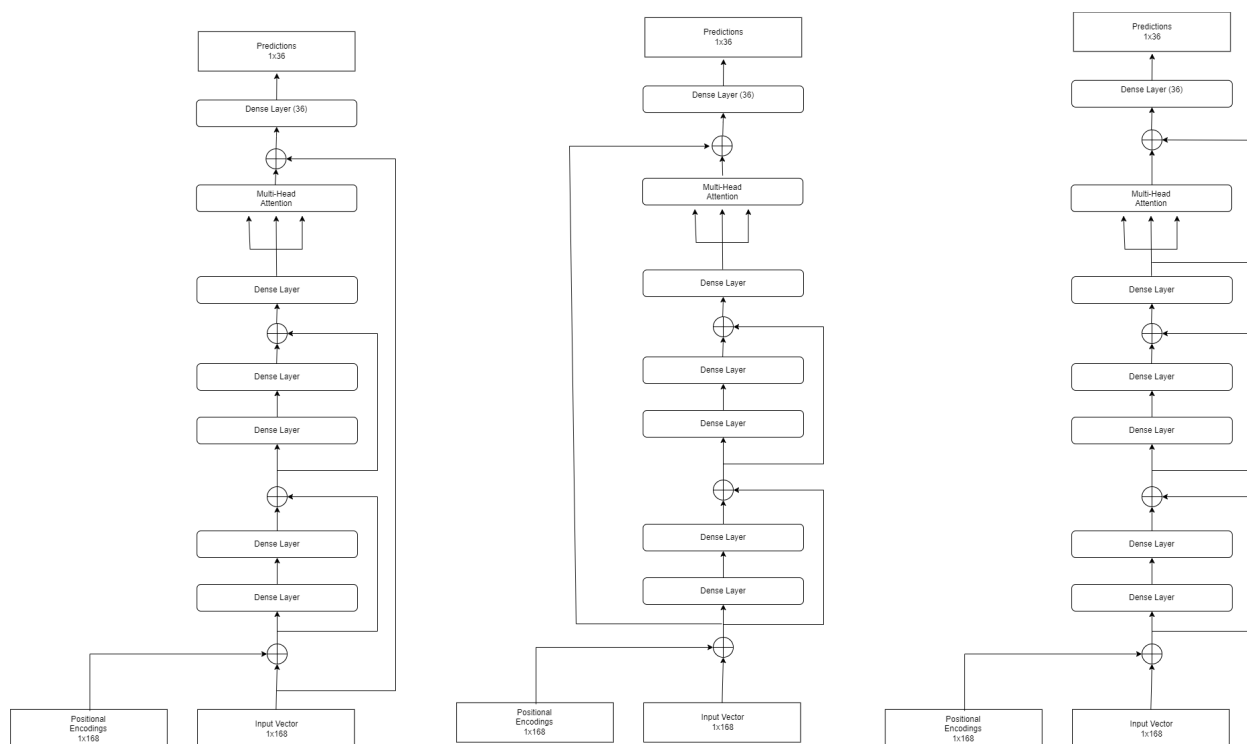
Όπως γίνεται αντιληπτό, υλοποιήθηκαν δύο διαφορετικές προσεγγίσεις. Αρχικά, δοκιμάστηκε η χρήση υπολειμματικής σύνδεσης μεταξύ της εισόδου του δικτύου και της εξόδου του μηχανισμού αυτοπροσοχής (παρουσιάζεται στην αριστερή πλευρά του σχεδιαγράμματος 7.9). Επιπλέον, χρησιμοποιήθηκε και η χρήση υπολειμματικής σύνδεσης μεταξύ της εισόδου του μηχανισμού προσοχής και της εξόδου αυτού, προσθέτοντας τις δύο αναπαραστάσεις και προωθώντας τις στο πλήρως συνδεδεμένο επίπεδο νευρώνων για να προκύψουν οι προβλέψεις. Και στις δύο περιπτώσεις, η άθροιση αυτή ήταν δυνατή καθώς σε όλα αυτά τα σημεία του δικτύου που προαναφέρονται τα διανύσματα που προκύπτουν ανήκουν στον ίδιο διανυσματικό χώρο  $\mathbb{R}^{1 \times 168}$ .

Επιπλέον, χρησιμοποιήθηκε και η οικογένεια Positional Attention Models που παρουσιάστηκε στην προηγούμενη υποενότητα εμπλουτισμένη αυτή τη φορά με υπολειμματική σύνδεση μετά τον μηχανισμό προσοχής. Ωστόσο, δοκιμάστηκαν τρεις διαφορετικές εσωτερικές αναπαραστάσεις του μοντέλου οι οποίες θα μπορούσαν να προστεθούν στην έξοδο του μηχανισμού προσοχής.

Όπως φαίνεται στο σχεδιάγραμμα 7.10, αρχικά χρησιμοποιήθηκε μία υπολειμματική σύνδεση μεταξύ της εισόδου του δικτύου και της εξόδου του μηχανισμού προσοχής. Το άθροισμα



Σχήμα 7.9: Είδη υπολειμματικών συνδέσεων που εφαρμόστηκαν Γύρω από τον μηχανισμό προσοχής στα Convolutional Attention Models



Σχήμα 7.10: Είδη υπολειμματικών συνδέσεων που εφαρμόστηκαν Γύρω από τον μηχανισμό προσοχής στα Positional Attention Models

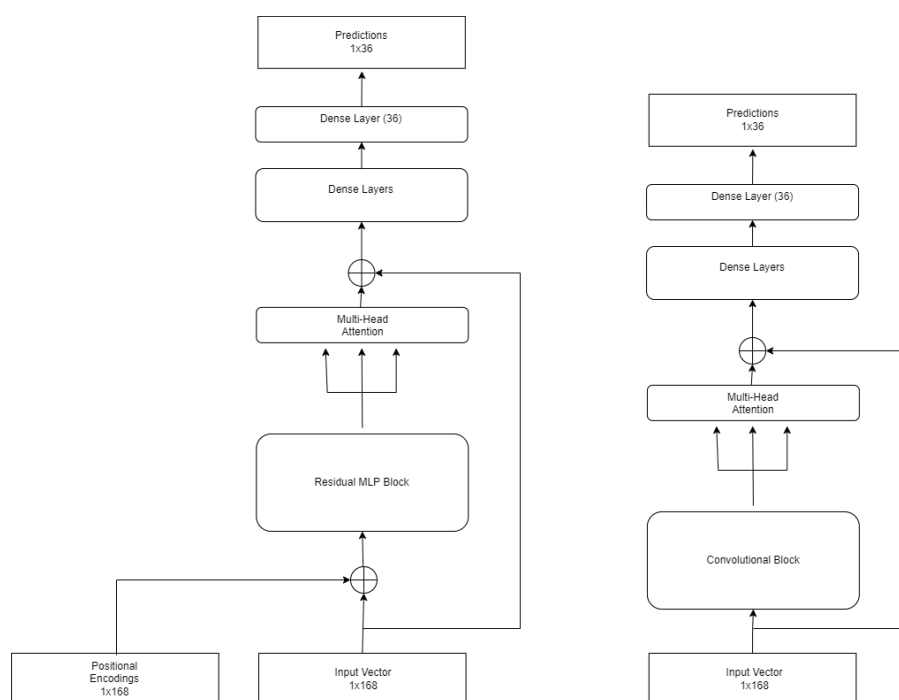
αυτών των δύο διανυσμάτων τροφοδοτείται έπειτα στο τελευταίο επίπεδο πλήρως συνδεδεμένων νευρώνων έτσι ώστε να προκύψουν οι προβλέψεις. Οι δύο υπόλοιπες υπολειμματικές

συνδέσεις που δοκιμάστηκαν ακολουθούν περισσότερο τη λογική των αντίστοιχων συνδέσεων στην αρχική αρχιτεκτονική του Transformer. Δοκιμάστηκε η άθροιση του εμπλουτισμένου με κωδικοποίηση θέσης διανύσματος με το διάνυσμα εξόδου του μηχανισμού προσοχής, ενώ όπως φαίνεται στο τρίτο μοντέλο στο σχεδιάγραμμα 7.10 δοκιμάστηκε και η άθροιση της εισόδου με την έξοδο του μηχανισμού. Το άθροισμα αυτό, σε κάθε περίπτωση τροφοδοτείται στο τελευταίο επίπεδο πλήρως συνδεδεμένων νευρώνων.

Η συγκριτική μελέτη της προγνωστικής ικανότητας των μοντέλων αυτών χωρίς ή με υπολειμματικές συνδέσεις γύρω από τον μηχανισμό προσοχής θα παρουσιαστεί στην επόμενη ενότητα, ωστόσο, αξίζει να αναφερθεί σε αυτό το σημείο, πως η χρήση αυτών των συνδέσεων βελτιώνει σημαντικά την ακρίβεια των προβλέψεων. Για το λόγο αυτό, οι επόμενες αρχιτεκτονικές που θα παρουσιαστούν θα βασίζονται στα παραπάνω μοντέλα.

### Χρήση MLP μετά τον μηχανισμό προσοχής

Χτίζοντας πάνω στις αρχιτεκτονικές που παρουσιάστηκαν παραπάνω, δοκιμάστηκε η χρήση πλήρως συνδεδεμένων επιπέδων νευρώνων στην έξοδο του μηχανισμού προσοχής, η έξοδος των οποίων τροφοδοτεί το τελευταίο επίπεδο 36 νευρώνων για την εξαγωγή προβλέψεων. Η παραπάνω λογική εφαρμόστηκε και στην οικογένεια Positional Attention Models αλλά και στην οικογένεια Convolutional Attention Models, όπως παρουσιάζεται και στο σχεδιάγραμμα 7.11.



Σχήμα 7.11: Χρήση MLP μετά τον μηχανισμό προσοχής

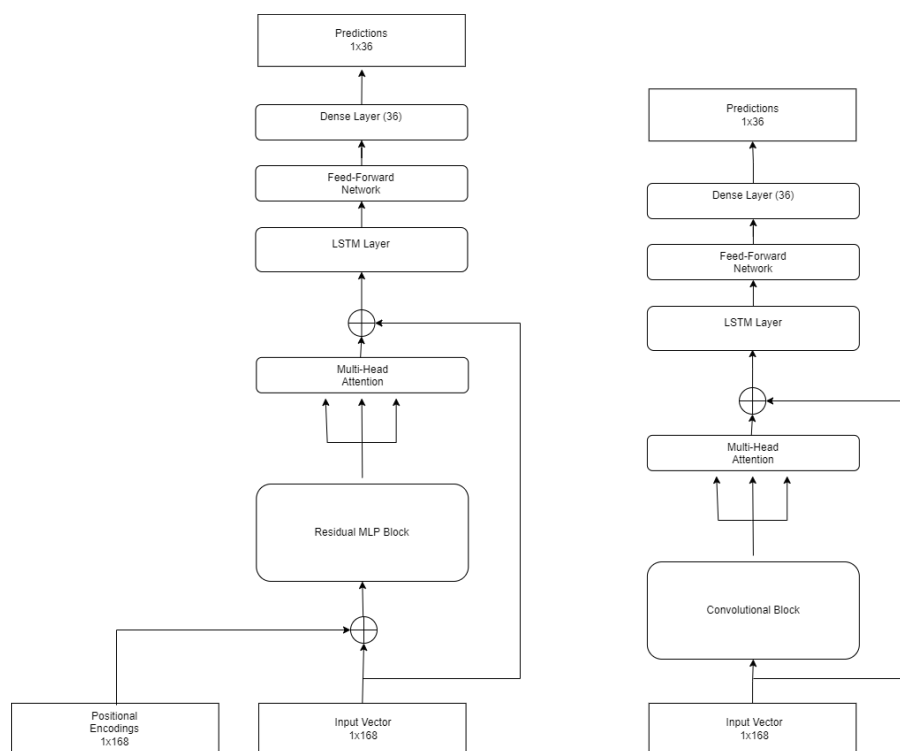
Όπως και παραπάνω, εξετάστηκε λεπτομερώς ο βέλτιστος αριθμός στρώσεων νευρώνων καθώς και ο αριθμός νευρώνων ανά επίπεδο. Παράλληλα, και στην περίπτωση αυτής της αρχιτεκτονικής, όπως και προηγουμένως όταν δοκιμάζονταν τα MLP πριν τον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών, δοκιμάστηκε και η χρήση υπολειμματικών συνδέσεων μεταξύ των

επιπέδων πλήρως συνδεδεμένων νευρώνων. Τα αποτελέσματα θα παρουσιαστούν αναλυτικά στην επόμενη ενότητα.

### Χρήση LSTM μετά τον μηχανισμό προσοχής

Εκτός από τα πλήρως συνδεδεμένα επίπεδα νευρώνων, μετά τον μηχανισμό προσοχής δοκιμάστηκε και η χρήση επιπέδων LSTM. Πιο συγκεκριμένα, το διάνυσμα  $X \in \mathbb{R}^{1 \times 168}$  που εξέρχεται από τον μηχανισμό προσοχής τροφοδοτείται στο LSTM το οποίο αποτελείται από  $h$  μονάδες (cells). Το διάνυσμα  $lstm_{out} \in \mathbb{R}^{1 \times h}$  που προκύπτει στην έξοδο του LSTM είναι η τελευταία κρυφή κατάσταση  $h_t = o_t * \tanh(c_t)$  που προκύπτει, με τρόπο που έχει εξηγηθεί αναλυτικά σε προηγούμενη ενότητα. Τέλος, το διάνυσμα  $lstm_{out}$  τροφοδοτείται στο απαραίτητο στρώμα πλήρως συνδεδεμένων νευρώνων έτσι ώστε να προκύψουν οι τελικές 36 προβλέψεις. Εδώ αξίζει να σημειωθεί πως σε αυτή την οικογένεια μοντέλων εξετάστηκαν και αρχιτεκτονικές οι οποίες περιλαμβάνουν περισσότερα επίπεδα LSTM πριν τον μηχανισμό προσοχής, με την έξοδο του ενός να τροφοδοτεί την είσοδο του άλλου. Ωστόσο, το τελικό διάνυσμα το οποίο τροφοδοτείται στο τελικό πλήρως συνδεδεμένο επίπεδο νευρώνων παραμένει το ίδιο ( $lstm_{out} \in \mathbb{R}^{1 \times h}$ ) καθώς όλα τα επίπεδα LSTM περιέχουν τον ίδιο αριθμό από cells.

Φυσικά, ένα από τα χαρακτηριστικά του LSTM που εξετάστηκε ενδελεχώς ως προς την επίδρασή του στην ακρίβεια των προβλέψεων είναι το πλήθος των μονάδων (cells) καθώς και η επίδραση αυτού στον χρόνο εκπαίδευσης της συνολικής αρχιτεκτονικής, ενώ δοκιμάστηκε και η επίδραση των πλήρως συνδεδεμένων επιπέδων νευρώνων (μετά το επίπεδο LSTM και πριν το τελικό επίπεδο νευρώνων) στην προγνωστική ικανότητα του μοντέλου.



Σχήμα 7.12: Χρήση LSTM μετά τον μηχανισμό προσοχής

Τα αποτελέσματα από τη χρήση των παραπάνω αρχιτεκτονικών θα παρουσιαστούν αναλυ-



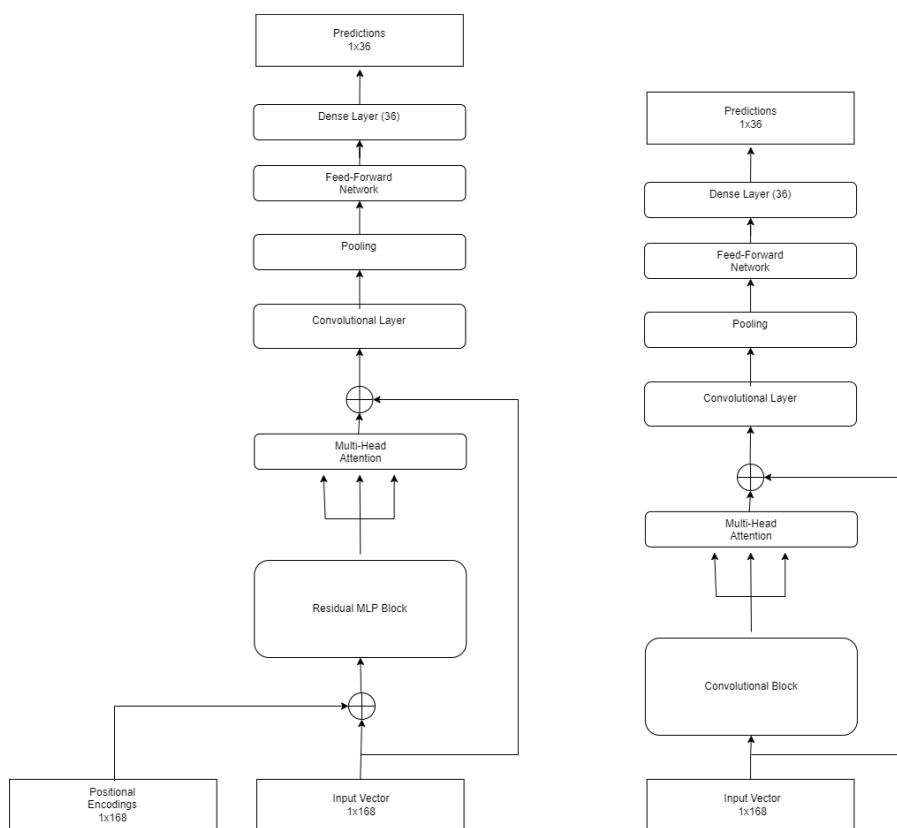
τικά στην επόμενη ενότητα.

### Χρήση CNN μετά τον μηχανισμό προσοχής

Τέλος, μετά τον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών δοκιμάστηκε και η χρήση συνελικτικών επιπέδων. Σε αυτή την περίπτωση, το διάνυσμα  $X \in \mathbb{R}^{1 \times 168}$  που εξέρχεται από τον μηχανισμό προσοχής εισέρχεται στο πρώτο συνελικτικό επίπεδο το οποίο χαρακτηρίζεται από πυρήνες μεγέθους  $k$  (kernel size) και από  $f$  στο πλήθος φίλτρα.

Στην έξοδο του συνελικτικού επιπέδου προκύπτει διάνυσμα  $c_{\text{out}} \in \mathbb{R}^{1 \times (l_{\text{in}} - k + 1) \times f}$ , όπου  $l_{\text{in}}$  είναι το μέγεθος του διανύσματος που εισέρχεται στο συνελικτικό επίπεδο. Έπειτα από κάθε συνελικτικό επίπεδο ακολουθεί ένα pooling επίπεδο, το οποίο δέχεται ως είσοδο το διάνυσμα  $c_{\text{out}}$  και προκύπτει το διάνυσμα  $p_{\text{out}} \in \mathbb{R}^{1 \times \lfloor \frac{(l_{\text{in}} - k + 1)}{p} \rfloor \times f}$ .

Αυτό είτε ανατροφοδοτείται σε ένα νέο ζεύγος συνελικτικού και pooling επιπέδου είτε επιπεδοποιείται και τροφοδοτείται σε ένα MLP πριν εισέλθει στο τελικό επίπεδο 36 πλήρως συνδεδεμένων νευρώνων, έτσι ώστε να προκύψουν οι προβλέψεις. Αξίζει να σημειωθεί πως στην περίπτωση της ανατροφοδότησης σε επιπλέον συνελικτικά επίπεδα, η τελευταία διάσταση, αυτή που αντιστοιχεί στον αριθμό των φίλτρων, διατηρεί σταθερό το μέγεθός της, καθώς ο αριθμός των φίλτρων που χρησιμοποιούνται είναι ο ίδιος σε όλα τα συνελικτικά επίπεδα.



Σχήμα 7.13: Χρήση CNN μετά τον μηχανισμό προσοχής

Κατά την εξέταση της προγνωστικής ικανότητας των παραπάνω αρχιτεκτονικών δοκιμάστηκαν διάφορες εκδοχές τους. Εξετάστηκαν διαφορετικά μεγέθη πυρήνα, μία πληθώρα αριθμού φίλτρων καθώς και διάφορα πλήρως συνδεδεμένα επίπεδα νευρώνων μετά τα συνελι-

κτικά επίπεδα και πριν το τελικό επίπεδο 36 νευρώνων.

Τα αποτελέσματα που ανταναχλούν την προγνωστική ικανότητα των παραπάνω αρχιτεκτονικών θα παρουσιαστούν αναλυτικά στην ενότητα των αποτελεσμάτων. Ωστόσο, για να γίνουν αντιληπτές οι σχεδιαστικές επιλογές στην αρχιτεκτονική που θα παρουσιαστεί παρακάτω, αξίζει να σημειωθεί σε αυτό το σημείο πως η χρήση MLP, LSTM ή συνελικτικών επιπέδων μετά τον μηχανισμό προσοχής δεν φαίνεται να συμβάλλει ιδιαίτερα στην προγνωστική ικανότητα των μοντέλων.

### 7.3.4 Ο Transformer

Η χρήση μηχανισμών προσοχής πολύ συχνά πλαισιώνεται από την αρχιτεκτονική του Transformer, έτσι όπως αυτή παρουσιάστηκε στην δημοσιευμένη έρευνα “Attention is All You Need”. Αυτή η λογική ήταν και ο λόγος που στα πλαίσια της παρούσας διπλωματικής εργασίας επιλέχθηκε η ανάπτυξη μίας αρχιτεκτονικής Transformer, χρησιμοποιώντας ως δομικά στοιχεία τις αρχιτεκτονικές που έχουν παρουσιαστεί μέχρι στιγμής.

Πιο συγκεκριμένα, για την κατασκευή του Encoder χρησιμοποιήθηκε μία αρχιτεκτονική της οικογένειας Convolutional Attention Models ενώ για την κατασκευή του Decoder χρησιμοποιήθηκε μία αρχιτεκτονική της οικογένειας Positional Attention Models. Όπως θα φανεί παρακάτω, ο Decoder χρησιμοποιείται με έναν αρκετά διαφορετικό τρόπο από αυτόν που χρησιμοποιείται στην έρευνα “Attention is All You Need”, ωστόσο χρησιμοποιείται αυτή η ονομασία για λόγους ευκολίας στην σύγκριση των δύο αρχιτεκτονικών.

Αρχικά, όσον αφορά τον Encoder, το διάνυσμα εισόδου  $X \in \mathbb{R}^{1 \times 168}$  εισέρχεται στο συνελικτικό Block. Αυτό ανήκει στην οικογένεια Convolutional Attention Models και περιέχει, όπως έχει προαναφερθεί, ένα μονοδιάστατο συνελικτικό επίπεδο, ένα pooling επίπεδο και τέλος, πλήρως συνδεδεμένα επίπεδα νευρώνων. Ο λόγος που σε αυτό το σημείο επιλέγεται να διαφοροποιηθεί η αρχιτεκτονική από αυτή του κλασικού Transformer θα γίνει αντιληπτός κατά την παρουσίαση των αποτελεσμάτων. Το διάνυσμα που προκύπτει στην έξοδο αυτού του Block έχει προσαρμοστεί από το τελευταίο επίπεδο νευρώνων, έτσι ώστε να έχει αντίστοιχες διαστάσεις με την υπόλοιπη αρχιτεκτονική, δηλαδή  $out_{conv} \in \mathbb{R}^{1 \times 168}$ . Έπειτα, αυτό τροφοδοτείται στον πρώτο μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών και προκύπτει το διάνυσμα  $out_{enc} \in \mathbb{R}^{1 \times 168}$  στο οποίο όμως έχει προστεθεί και η αρχική είσοδος. Το εύρος αυτής της ανάδρασης είναι άλλο ένα σημείο στο οποίο διαφοροποιείται η νέα αυτή αρχιτεκτονική από τον Transformer της έρευνας Attention is All You Need.

Προχωρώντας στον Decoder, παρατηρούμε κάποιες επιπλέον αλλαγές, των οποίων φυσικά η χρησιμότητα θα γίνει αντιληπτή κατά την παρουσίαση των αποτελεσμάτων. Ως είσοδος στον Decoder χρησιμοποιείται και πάλι το ίδιο διάνυσμα εισόδου, αυτή τη φορά εμπλουτισμένο με κωδικοποιήσεις θέσης. Αυτό εισέρχεται σε ένα Residual MLP Block το οποίο αποτελείται από πλήρως συνδεδεμένα επίπεδα νευρώνων τα οποία όμως πλαισιώνονται και από υπολειμματικές συνδέσεις, όπως έχει παρουσιαστεί και παραπάνω. Το διάνυσμα εξόδου αυτού του Block εισέρχεται έπειτα σε έναν μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών ο οποίος όμως χρησιμοποιεί μάσκα.

Ο μηχανισμός προσοχής με μάσκα αποτελεί βασικό συστατικό των Transformer, ιδίως σε εργασίες όπως η μοντελοποίηση γλώσσας. Ο μηχανισμός επιτρέπει στα μοντέλα να λαμ-

βάνουν υπόψη μόνο ορισμένα τμήματα της εισόδου κατά τον υπολογισμό των αποτελεσμάτων προσοχής.

Σε εργασίες όπως η γλωσσική μοντελοποίηση, όπου το μοντέλο παράγει κείμενο μία λέξη κάθε φορά, είναι ζωτικής σημασίας να μην μπορεί το μοντέλο να «κλέψει» εξετάζοντας μελλοντικές λέξεις κατά τη διάρκεια της εκπαίδευσης. Για παράδειγμα, κατά την πρόβλεψη της επόμενης λέξης σε μια πρόταση, το μοντέλο θα πρέπει να λαμβάνει υπόψη μόνο την τρέχουσα και την προηγούμενη λέξη, όχι τις λέξεις που ακολουθούν. Φυσικά, σε αυτό το σημείο εντοπίζεται άλλη μία σημαντική διαφορά με τις κλασικές αρχιτεκτονικές Transformer, καθώς η αρχιτεκτονική που παρουσιάζεται δεν θα παράγει μία ωριαία πρόβλεψη τη φορά και μετά θα τη χρησιμοποιεί σαν είσοδο για τις επόμενες προβλέψεις. Αντίθετα, θα παράγει και τις 36 ωριαίες προβλέψεις μαζί. Συνεπώς, η μάσκα δεν χρησιμοποιείται για να αποτρέψει τον μηχανισμό προσοχής να δει δεδομένα που δεν θα έπρεπε, αλλά εξασφαλίζεται να είναι απαραίτητη σε συγκεκριμένα σημεία για να παράγονται ακριβείς προβλέψεις, όπως θα φανεί και στην παρουσίαση των αποτελεσμάτων. Στην παρούσα εργασία, η μάσκα υλοποιείται ως εξής: Πρόκειται για έναν πίνακα που αντιστοιχεί στις διαστάσεις των αποτελεσμάτων προσοχής. Δηλαδή, ο πίνακας μάσκας  $M$  στην συγκεκριμένη περίπτωση έχει μέγεθος  $168 \times 168$ . Πιο συγκεκριμένα, η μάσκα είναι ένας άνω τριγωνικός πίνακας γεμάτος με  $-\infty$  (ή έναν πολύ μεγάλο αρνητικό αριθμό, στην συγκεκριμένη περίπτωση  $-10^9$ ) πάνω από τη διαγώνιο και μηδενικά αλλού. Αυτή η ρύθμιση διασφαλίζει ότι το μοντέλο δίνει προσοχή μόνο σε σχέσεις με παρελθοντικές τιμές.

$$M = \begin{pmatrix} 0 & -\infty & -\infty & \dots & -\infty \\ 0 & 0 & -\infty & \dots & -\infty \\ 0 & 0 & 0 & \dots & -\infty \\ \vdots & \vdots & \vdots & \ddots & -\infty \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Η μάσκα προστίθεται στις βαθμολογίες προσοχής πριν από την εφαρμογή της συνάρτησης softmax. Οι μεγάλες αρνητικές τιμές εξασφαλίζουν ότι οι αντίστοιχες θέσεις λαμβάνουν μηδενική πιθανότητα μετά την softmax.

Δεδομένου του πίνακα αποτελεσμάτων προσοχής  $A$  (που προκύπτει από το γινόμενο των  $Q$  και  $K$ , κλιμακωμένο κατά  $\sqrt{d_k}$ ) και του πίνακα μάσκας  $M$ :

$$\text{Masked Attention Scores} = A + M$$

Στη συνέχεια, εφαρμόζεται συνάρτηση softmax για να προκύψουν τα βάρη προσοχής

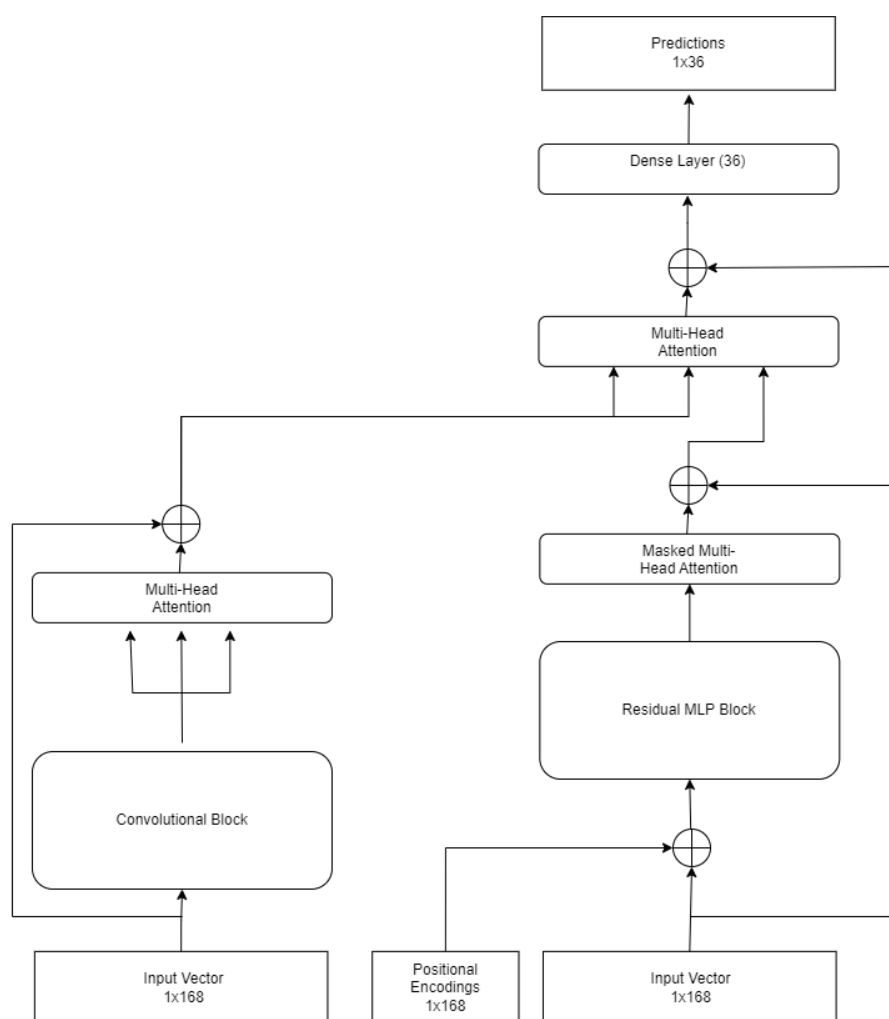
$$\text{Attention Weights} = \text{softmax}(A + M).$$

Η έξοδος της κεφαλής είναι κατά τα γνωστά

$$\text{Output} = \text{Attention Weights} \times V.$$

Όπως γίνεται αντιληπτό, το διάνυσμα εξόδου αυτού του τροποποιημένου μηχανισμού προσοχής δεν αλλάζει μέγεθος, άρα χρησιμοποιείται χωρίς να χρειαστεί να αλλάξει κάτι στην

αρχιτεκτονική. Συνεπώς, το διάνυσμα εξόδου από τον μηχανισμό προσοχής με μάσκα προστίθεται μέσω μιας υπολειμματικής σύνδεσης με την είσοδο και τροφοδοτείται ως διάνυσμα  $V$  σε έναν ακόμα μηχανισμό προσοχής, αυτή τη φορά χωρίς μάσκα. Αυτός ο μηχανισμός χρησιμοποιεί ως  $Q, K$  την έξοδο του Encoder. Τέλος, η έξοδος του δεύτερου μηχανισμού προσοχής του Decoder προστίθεται και πάλι με την είσοδο, κάτι που αποτελεί άλλη μία διαφοροποίηση από τον κλασικό Transformer. Οι προβλέψεις προκύπτουν από ένα τελικό επίπεδο 36 πλήρως συνδεδεμένων νευρώνων.



Σχήμα 7.14: Ο Transformer που κατασκευάστηκε

Όπως θα αναλυθεί και στην ενότητα των αποτελεσμάτων, εξετάστηκαν πολλές εκδόσεις της παραπάνω αρχιτεκτονικής, οι οποίες συμπεριελάμβαναν χρήση μηχανισμών με μάσκα σε διαφορετικά σημεία, διάφορες υπολειμματικές συνδέσεις καθώς και διαφορετικές επιλογές για το κύριο σώμα του Encoder και του Decoder.

## Κεφάλαιο 8

### Αποτελέσματα

---

Όπως έγινε αντιληπτό στην παραπάνω ενότητα, η χρήση μηχανισμών προσοχής στην πρόβλεψη χρονοσειρών, ιδιαίτερα στη πρόβλεψη του ηλεκτρικού φορτίου στην οποία εστιάζει η παρούσα εργασία, μπορεί να πάρει πολλές μορφές, καθώς ο μηχανισμός από μόνος του διαφοροποιείται αρκετά ανάλογα με τις σχεδιαστικές επιλογές που αφορούν κυρίως τις διαστάσεις του και τον αριθμό των κεφαλών προσοχής που μπορεί να χρησιμοποιηθούν. Ωστόσο, η μεγάλη ποικιλία αρχιτεκτονικών που εξετάζονται στην παρούσα εργασία δεν έχει να κάνει μόνο με τη σχεδίαση του μηχανισμού αυτού καθ'αυτού αλλά και με τις νευρωνικές αρχιτεκτονικές που τον πλαισιώνουν, είτε πριν από αυτόν, τροφοδοτώντας τον με ένα πιο εμπλουτισμένο και επεξεργασμένο διάνυσμα, είτε μετά από τον μηχανισμό, τονώνοντας την προγνωστική του ικανότητα. Στην παρούσα ενότητα θα παρουσιαστούν τα αποτελέσματα που προκύπτουν από την λεπτομερή εξέταση της προγνωστικής ικανότητας όλων αυτών των αρχιτεκτονικών.

#### 8.1 Αποκλειστικά Μηχανισμοί Προσοχής

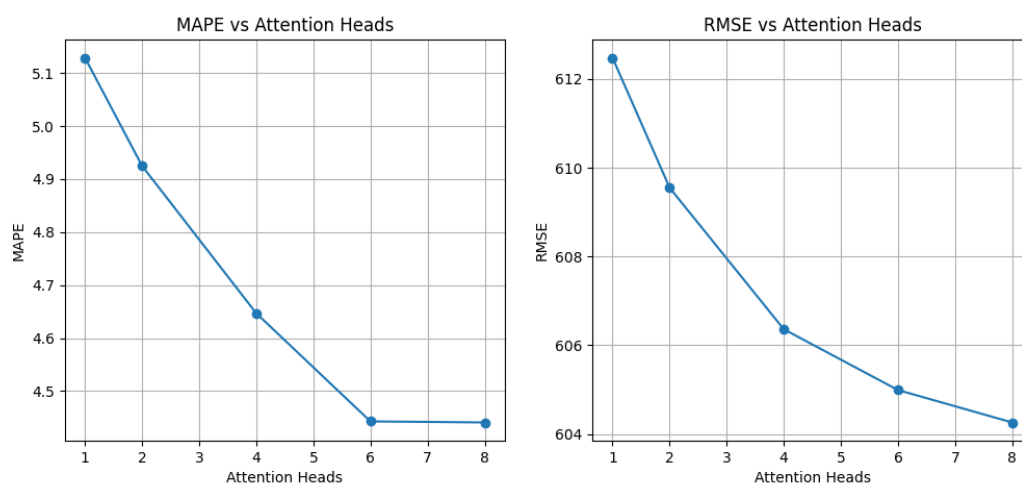
Αρχικά, όπως παρουσιάστηκε αναλυτικά και στην προηγούμενη ενότητα, αναπτύχθηκαν αρχιτεκτονικές κατά τις οποίες το διάνυσμα εισόδου εισέρχεται στον μηχανισμό αυτοπροσοχής χωρίς προηγούμενη προεπεξεργασία και οι προβλέψεις προκύπτουν από την έξοδο του μηχανισμού και ένα πλήρως συνδεδεμένο επίπεδο 36 νευρώνων.

Αρχικά εξετάστηκε η επίδραση του πλήθους των κεφαλών προσοχής στην προγνωστική ικανότητα του μοντέλου.

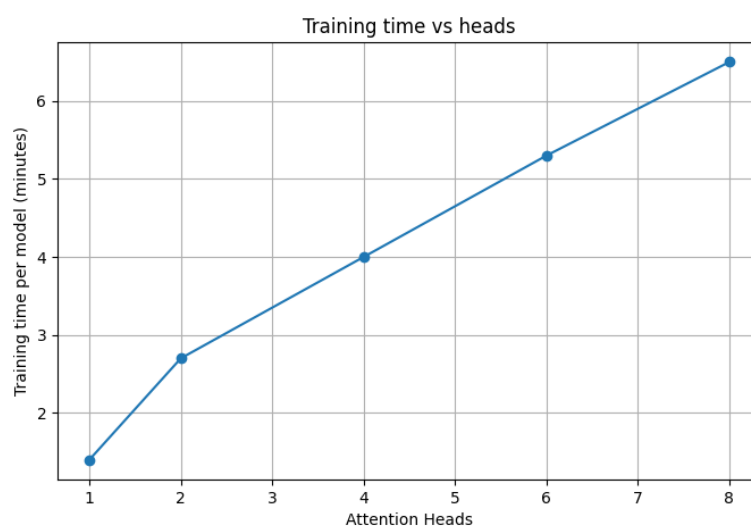
Όπως γίνεται αντιληπτό από τις γραφικές παραστάσεις της εικόνας 8.1, η χρήση παραπάνω κεφαλών φαίνεται να βοηθάει μέχρι ενός σημείου. Κάτι τέτοιο ήταν απολύτως αναμενόμενο καθώς κάθε κεφαλή μπορεί να εστιάσει ανεξάρτητα σε διαφορετικές πτυχές των δεδομένων εισόδου, παρέχοντας μια πιο πλούσια, πιο ποικίλη αναπαράσταση των χαρακτηριστικών. Αυτή η ποικιλομορφία ενισχύει την ικανότητα του μοντέλου να μαθαίνει παράλληλα ποικίλα χαρακτηριστικά δεδομένων, βελτιώνοντας τη γενίκευση σε νέα δεδομένα με την ενσωμάτωση πολλαπλών προοπτικών.

Φυσικά, όπως φαίνεται και από τις γραφικές παραστάσεις της εικόνας 8.1, η αύξηση των κεφαλών άνω των 6 δεν έχει ιδιαίτερος εντυπωσιακά αποτελέσματα, ενώ όπως παρουσιάζεται στο διάγραμμα 8.2 αυξάνεται και σημαντικά ο χρόνος που απαιτείται για την εκπαίδευση του μοντέλου.

Τα παραπάνω παρουσιαζόμενα μοντέλα εκπαιδεύονται για 20000 βήματα, παράμετρος η



Εικόνα 8.1: Η επίδραση του πλήθους κεφαλών προσοχής στην προγνωστική ικανότητα του μοντέλου.



Εικόνα 8.2: Η επίδραση του πλήθους κεφαλών προσοχής στον χρόνο εκπαίδευσης του μοντέλου.

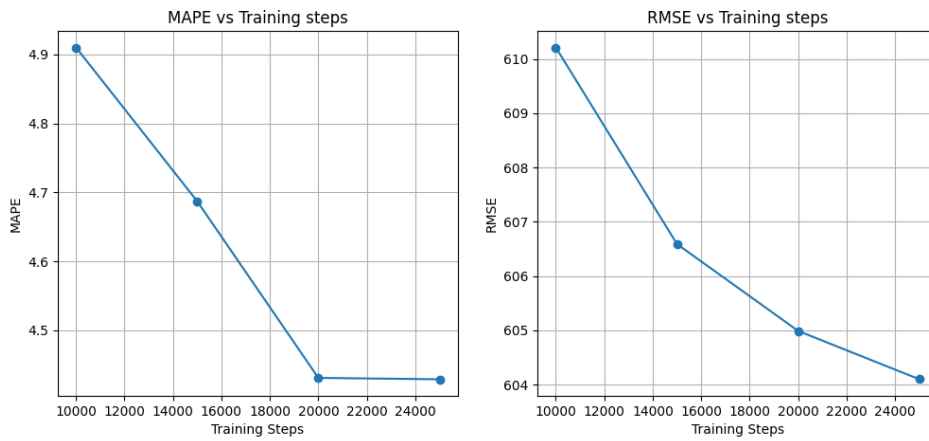
οποία εξετάστηκε διεξοδικά. Όπως φαίνεται στην γραφική παράσταση της εικόνας 8.3, τα συγκεκριμένα βήματα εκπαίδευσης κρίθηκαν αρκετά για την εκπαίδευση των αρχιτεκτονικών, καθώς η περαιτέρω αύξηση των βημάτων δεν προσφέρει ιδιαίτερη ακρίβεια στις προβλέψεις ενώ αυξάνει σημαντικά και τον χρόνο εκπαίδευσης.

Συνεπώς, στη συνέχεια αυτής της εργασίας, οι υπόλοιπες αρχιτεκτονικές των οποίων η προγνωστική ικανότητα θα διερευνηθεί, θα χρησιμοποιούν 6 κεφαλές προσοχής.

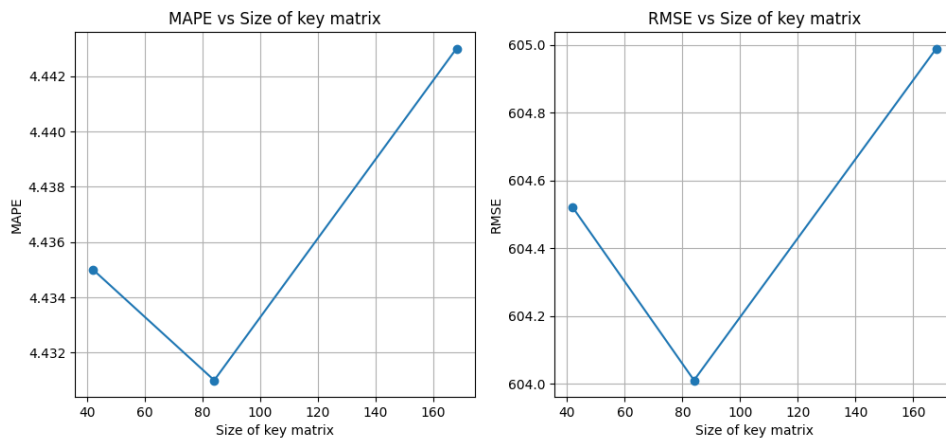
Ένα ακόμη χαρακτηριστικό των μηχανισμών αυτοπροσοχής πολλαπλών κεφαλών που εξετάστηκε είναι η διάσταση των πινάκων  $Q, K$ .

Όπως γίνεται αντιληπτό από τα διαγράμματα της εικόνας 8.4, η βέλτιστη διάσταση είναι 84, χωρίς όμως οι διαφορές να είναι πολύ μεγάλες.

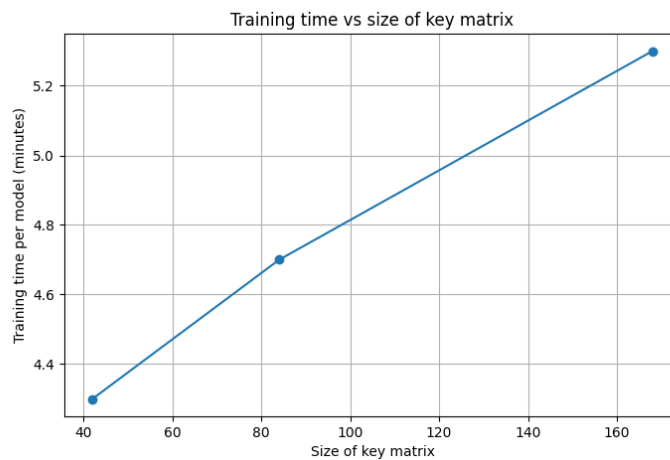
Οστόσο, όπως φαίνεται και στο διάγραμμα 8.5, τυχόν αύξηση της διάστασης σε 168 θα



Εικόνα 8.3: Η επίδραση των βημάτων εκπαίδευσης στην προγνωστική ικανότητα του μοντέλου.



Εικόνα 8.4: Η επίδραση του μεγέθους των πινάκων  $Q$ ,  $K$ ,  $V$  στην προγνωστική ικανότητα του μοντέλου.



Εικόνα 8.5: Η επίδραση του μεγέθους των πινάκων  $Q$ ,  $K$ ,  $V$  στην ταχύτητα εκπαίδευσης του μοντέλου.

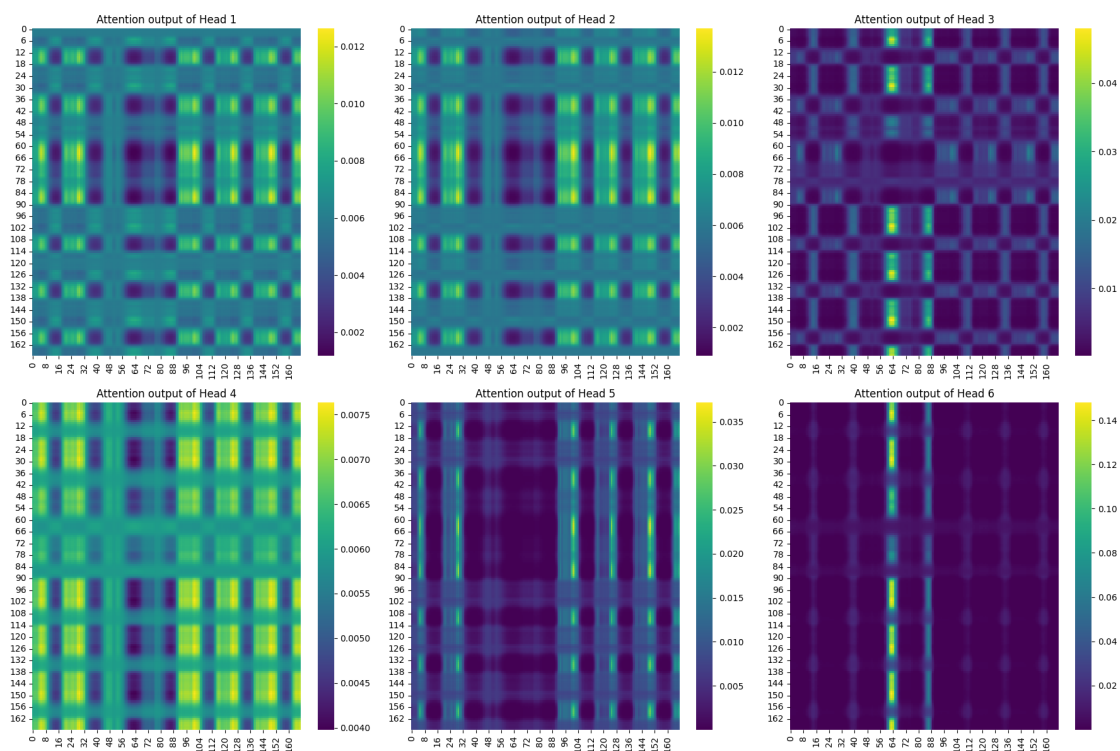
αύξανε πολύ τον χρόνο εκπαίδευσης. Εδώ αξίζει να υπενθυμιστεί ότι για κάθε αρχιτεκτονική εκπαιδεύονται 30 μοντέλα, συνεπώς μία διαφορά της τάξης του ενός λεπτού ανά μοντέλο σημαίνει συνολική αύξηση μισής ώρας στον συνολικό χρόνο εκπαίδευσης της αρχιτεκτονικής.

Συνοψίζοντας τα παραπάνω, ο μηχανισμός αυτοπροσοχής που προκύπτει πως έχει την καλύτερη προγνωστική ικανότητα αλλά και συνδυάζει την χρονική αποδοτικότητα είναι ένας μηχανισμός πολλαπλών κεφαλών και συγκεκριμένα 6 κεφαλών, ενώ η καλύτερη διάσταση για τους πίνακες  $Q, K$  φαίνεται να είναι 84. Φυσικά, τα παραπάνω αποτελούν την βάση του υπό κατασκευή μοντέλου και δεν επαρκούν από μόνα τους για ακριβείς προβλέψεις, όπως μπορεί να φανεί στον πίνακα 8.1, όπου συγκρίνουμε την μέχρι τώρα ακρίβεια προβλέψεων με αυτή των παρόχων ηλεκτρικής ενέργειας.

Πίνακας 8.1: Σύγκριση των *Basic Attention Models* με τις προβλέψεις των παρόχων

Model	MAPE	RMSE
Βέλτιστο Basic Attention Model	4.431	604
Προβλέψεις TSO	2.921	352

Ωστόσο, τα παραπάνω αποτελούν μία καλή αρχή πάνω στην οποία θα βασιστούμε για τα επόμενα. Η σημασία χρήσης μηχανισμού προσοχής και συγκεκριμένα πολλαπλών κεφαλών φαίνεται πολύ καθαρά και στην εικόνα 8.6, όπου έχει εκτυπωθεί σε ποια σημεία δίνεται σημασία από κάθε κεφαλή. Στην πράξη βλέπουμε τα βάρη του πίνακα έτσι όπως διαμορφώνεται μετά την έξοδο της softmax συνάρτησης εντός του μηχανισμού προσοχής.

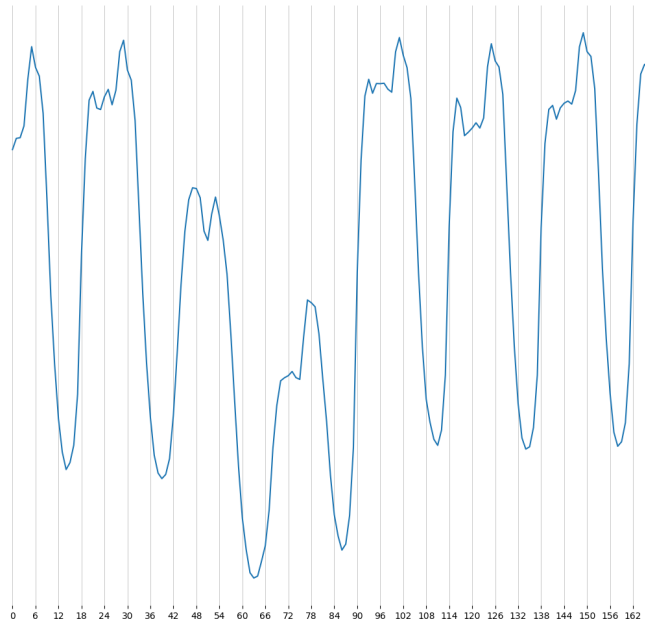


Εικόνα 8.6: Οι κεφαλές ενός *Basic Attention Model* διαμορφωμένες μετά την είσοδο μίας τυπικής εβδομάδας.

Το αποτέλεσμα της εικόνας 8.6 δημιουργήθηκε από την τροφοδότηση στο δίκτυο της



εισόδου που απεικονίζεται στην εικόνα 8.7 (μία τυπική εβδομάδα στην Πολωνία, όπου το Σαββατοκύριακο εντοπίζεται στη μέση του δείγματος):



Εικόνα 8.7: Η εβδομάδα κατά την είσοδο της οποίας στο μοντέλο προκύπτουν οι κεφαλές της εικόνας 8.6

Όπως παρατηρούμε, η έκτη κεφαλή δίνει ιδιαίτερη έμφαση στις τιμές του Σαββάτου και της Κυριακής, ενώ οι υπόλοιπες κεφαλές δίνουν μεγαλύτερη έμφαση, με διαφορετικό τρόπο η κάθε μία, στις καθημερινές. Αξίζει να σημειωθεί, πως με πιο προσεκτική εξέταση των παραπάνω διαγραμμάτων φαίνεται πως ο μηχανισμός προσοχής δίνει, ανεξαρτήτως ημέρας, μεγαλύτερη έμφαση στις μεσημεριανές ώρες, κάτι το αναμενόμενο, αφού αυτές είναι και οι πιο αντιπροσωπευτικές ώρες.

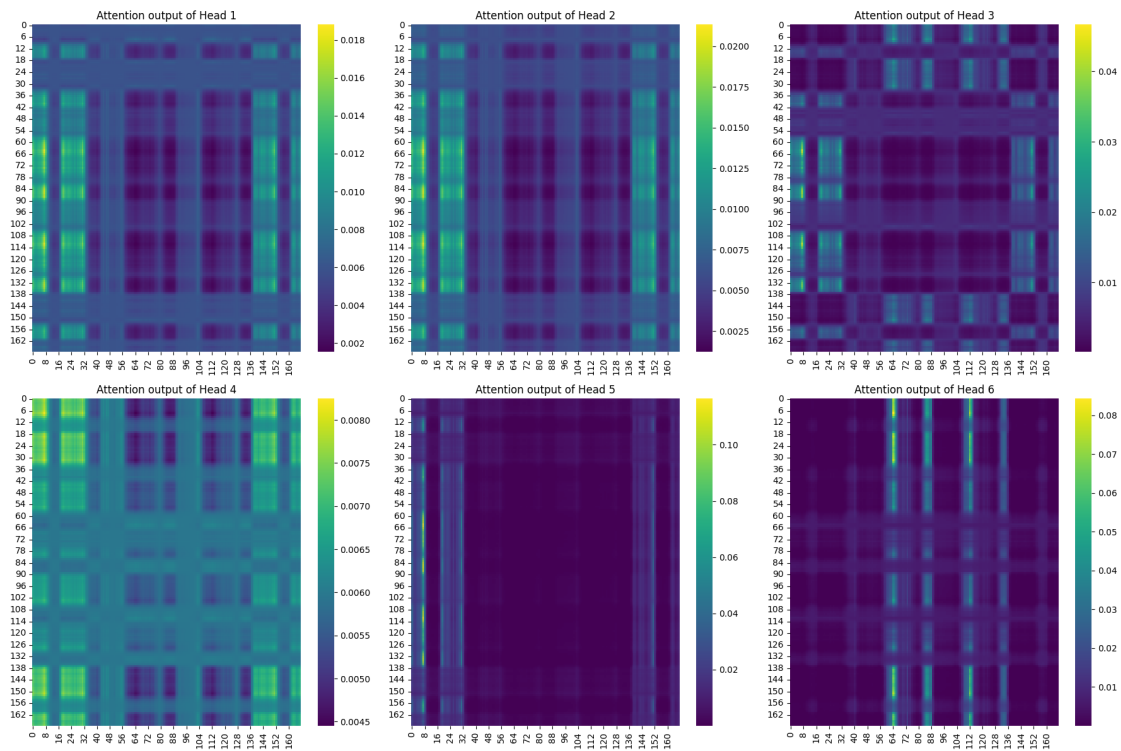
Επιπλέον, στην εικόνα 8.8 φαίνεται η οπτικοποίηση του πίνακα προσοχής όταν ως είσοδος δίνεται μία εβδομάδα που περιέχει ένα ειδικό γεγονός.

Η είσοδος που προκάλεσε αυτό το αποτέλεσμα φαίνεται στην εικόνα 8.9 (εβδομάδα στην οποία φαίνεται η επίδραση της ημέρας πολιτεύματος, η οποία είναι αργία στην Πολωνία).

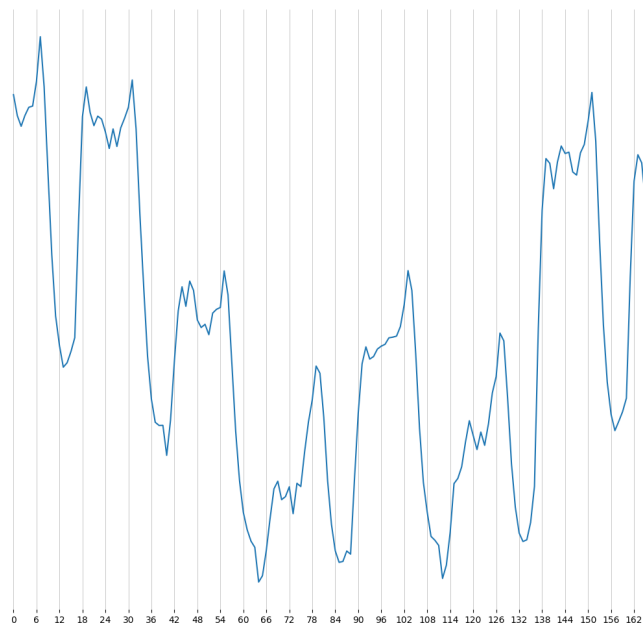
Όπως γίνεται αντιληπτό, ο μηχανισμός προσοχής μπορεί και εντοπίζει αυτό το γεγονός στην είσοδο του.

Ωστόσο, είναι ιδιαίτερα σημαντικό να αναφερθεί πως όταν στην είσοδο του μοντέλου δίνεται η εβδομάδα (168 τιμές) πριν το ειδικό γεγονός και το μοντέλο καλείται να προβλέψει το ειδικό γεγονός, τότε αποτυγχάνει. Με άλλα λόγια, σε αυτή την περίπτωση το μοντέλο δίνει έμφαση στην κλασική εβδομαδιαία εποχική διακύμανση αδυνατώντας να προβλέψει το ειδικό γεγονός. Αυτό γίνεται εύκολα αντιληπτό και στην εικόνα 8.10.

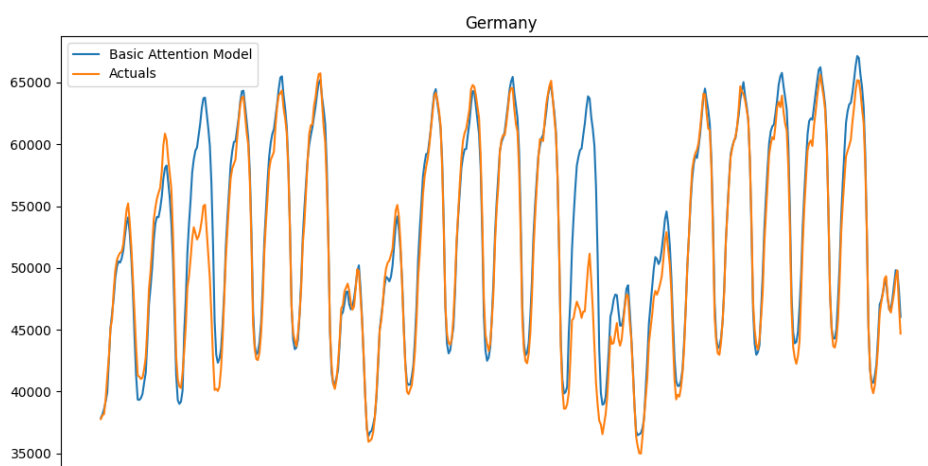
Όπως φαίνεται και στην εικόνα 8.10, ήδη το μοντέλο έχει μία πολύ ισχυρή προγνωστική ικανότητα. Ωστόσο, την Τρίτη της πρώτης εβδομάδας και την Παρασκευή της δεύτερης στα



Εικόνα 8.8: Οι κεφαλές ενός *Basic Attention Model* διαμορφωμένες μετά την είσοδο μίας εβδομάδας με ειδικό γεγονός.



Εικόνα 8.9: Η εβδομάδα κατά την είσοδο της οποίας στο μοντέλο προκύπτουν οι κεφαλές της εικόνας 8.8



Εικόνα 8.10: Το *Basic Attention Model* αδυνατεί να προβλέψει ειδικά γεγονότα.

δεδομένα που απεικονίζονται στην εικόνα και στην εικόνα 8.10, παρατηρούνται ειδικά γεγονότα τα οποία το μοντέλο μας δεν μπορεί ακόμα να προβλέψει. Πάνω σε αυτό το βασικό μοντέλο προσοχής θα χτίσουμε τα υπόλοιπα που ακολουθούν, επιχειρώντας να μειώσουμε στο μέγιστο δυνατό βαθμό τέτοιου είδους σφάλματα.

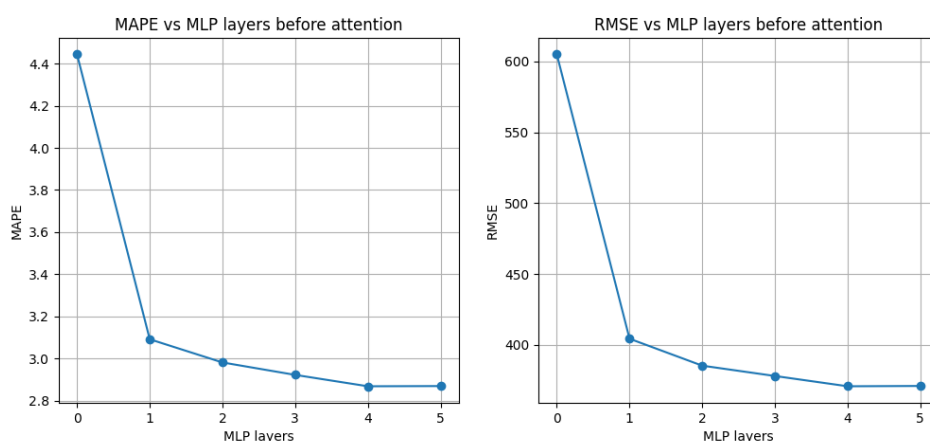
## 8.2 Αρχιτεκτονικές που προηγούνται του Μηχανισμού Προσοχής

Πριν τον μηχανισμό προσοχής, όπως παρουσιάστηκε αναλυτικά στην προηγούμενη ενότητα, χρησιμοποιήθηκαν διάφοροι μηχανισμοί έτσι ώστε να τροποποιήσουν την είσοδο του μοντέλου κατάλληλα για να εξάγει ο μηχανισμός προσοχής την μέγιστη δυνατή πληροφορία. Πιο συγκεκριμένα, χρησιμοποιήθηκαν πλήρως συνδεδεμένα επίπεδα νευρώνων, χωρίς ή με υπολειμματικές συνδέσεις μεταξύ τους, επίπεδα LSTM καθώς και συνελικτικά επίπεδα. Στην παρούσα ενότητα θα παρουσιαστούν αναλυτικά τα αποτελέσματα της έρευνας αυτής.

### 8.2.1 Αποτελέσματα από τη χρήση MLP πριν τον μηχανισμό προσοχής

Στην περίπτωση του πλήρως συνδεδεμένου δικτύου, η παράμετρος η οποία βρέθηκε στο επίκεντρο των πειραματισμών ήταν το πλήθος των πλήρως συνδεδεμένων επιπέδων πριν τον μηχανισμό προσοχής. Τα αποτελέσματα αυτής της έρευνας παρουσιάζονται στην εικόνα 8.11.

Όπως γίνεται αντιληπτό, η προσθήκη τέτοιων επιπέδων βοηθά πολύ στις προγνώσεις, ωστόσο μετά τα 5 επίπεδα η προσθήκη νέων επιπέδων δεν φαίνεται να έχει ιδιαίτερο αντίκτυπο. Ωστόσο, δοκιμάστηκε και η προσθήκη υπολειμματικών συνδέσεων ανάμεσα σε αυτά τα πλήρως συνδεδεμένα επίπεδα, όπως παρουσιάστηκε αναλυτικά στην προηγούμενη ενότητα. Ο πίνακας 8.2 συνοψίζει την αποδοτικότητα αυτών των συνδέσεων μεταξύ των πλήρως συνδεδεμένων επιπέδων.



Εικόνα 8.11: Η επίδραση των πλήρως συνδεδεμένων επιπέδων στα MLP Attention Models.

Πίνακας 8.2: Σύγκριση πλήρως συνδεδεμένων επιπέδων νευρώνων με και χωρίς υπολειμματικές συνδέσεις.

Είδος δικτύου πριν την είσοδο στο μηχανισμό προσοχής	MAPE	RMSE
4 πλήρως συνδεδεμένα επίπεδα	2.869 %	370
5 πλήρως συνδεδεμένα επίπεδα	2.873 %	371
5 πλήρως συνδεδεμένα επίπεδα με υπολειμματικές συνδέσεις	2.854 %	369
7 πλήρως συνδεδεμένα επίπεδα με υπολειμματικές συνδέσεις	2.822 %	365
9 πλήρως συνδεδεμένα επίπεδα με υπολειμματικές συνδέσεις	2.826 %	364

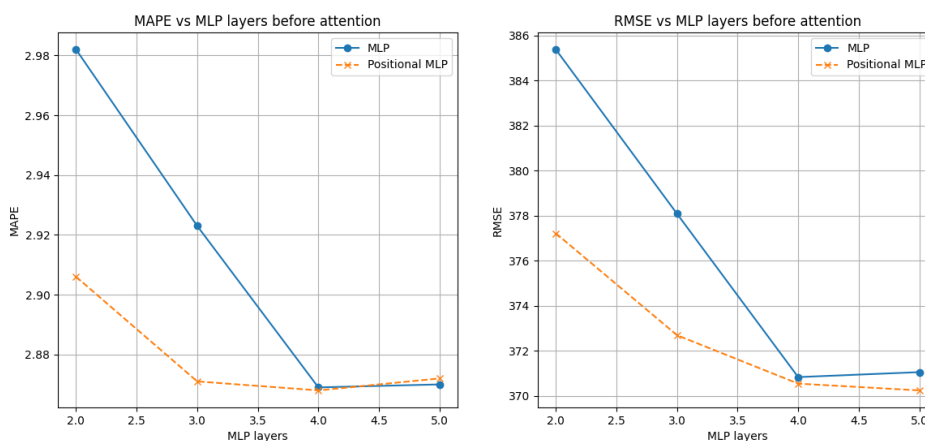
Όπως γίνεται αντιληπτό, καθώς αυξάνονται τα απλά πλήρως συνδεδεμένα επίπεδα πριν την είσοδο στον μηχανισμό προσοχής άνω των τεσσάρων η προγνωστική ικανότητα του μοντέλου μειώνεται. Αυτό συμβαίνει λόγω του προβλήματος της εξαφανιζόμενης κλίσης (vanishing gradient). Το πρόβλημα αυτό αντιμετωπίζεται με τη χρήση υπολειμματικών συνδέσεων, οι οποίες προσθέτουν προηγούμενα βάρη σε επόμενα. Για αυτό το λόγο επιλέγεται και σταθερός αριθμός νευρώνων ανά επίπεδο, έτσι ώστε να μπορεί να γίνει αυτή η άθροιση.

Καταλήγουμε λοιπόν πως το καλύτερο, μέχρι στιγμής, μοντέλο μας είναι ένα δίκτυο με 7 πλήρως συνδεδεμένα στρώματα νευρώνων τα οποία εμπλουτίζονται με υπολειμματικές συνδέσεις και παρέχουν την έξοδό τους στον μηχανισμό αυτοπροσοχής.

## 8.2.2 Αποτελέσματα από τη χρήση κωδικοποιήσεων θέσης

Τα επόμενα πειράματα που έγιναν περιείχαν την χρήση κωδικοποιήσεων θέσης. Όπως παρουσιάστηκε εκτενώς στην προηγούμενη ενότητα, στην οικογένεια αρχιτεκτονικών Positional Attention Models συμπεριλαμβάνονται μοντέλα τα οποία το εμπλουτισμένο με πληροφορία θέσης διάνυσμα εισόδου τροφοδοτούν σε πλήρως συνδεδεμένα δίκτυα. Όπως είναι λογικό εξετάστηκε και πάλι η επίδραση του πλήθους των επιπέδων αυτών.

Όπως φαίνεται στην εικόνα 8.12, όταν τα πλήρως συνδεδεμένα επίπεδα είναι λιγότερα (π.χ. 2) τότε η επίδραση των κωδικοποιήσεων θέσης είναι καταλυτική. Ωστόσο, καθώς τα επίπεδα αυξάνονται φαίνεται να έχουν από μόνα τους την ικανότητα να δημιουργούν διανύσματα με ενσωματωμένη την πληροφορία θέσης.



Εικόνα 8.12: Η επίδραση των κωδικοποιήσεων θέσης.

Επιπλέον των παραπάνω πειραμάτων, όπως παρουσιάστηκε και στην προηγούμενη ενότητα, δοκιμάστηκε και η χρήση υπολειμματικών συνδέσεων ανάμεσα στα επίπεδα που εισάγεται το εμπλουτισμένο με πληροφορία θέσης διάνυσμα.

Πίνακας 8.3: Υπολειμματικές συνδέσεις μετά την εισαγωγή κωδικοποιήσεων θέσης

Πλήρως συνδεδεμένα επίπεδα νευρώνων μετά τον εμπλουτισμό θέσης	MAPE	RMSE
4 επίπεδα	2.868 %	370
5 επίπεδα με υπολειμματικές συνδέσεις	2.854 %	369
7 επίπεδα με υπολειμματικές συνδέσεις	2.818 %	363
9 επίπεδα με υπολειμματικές συνδέσεις	2.819 %	363

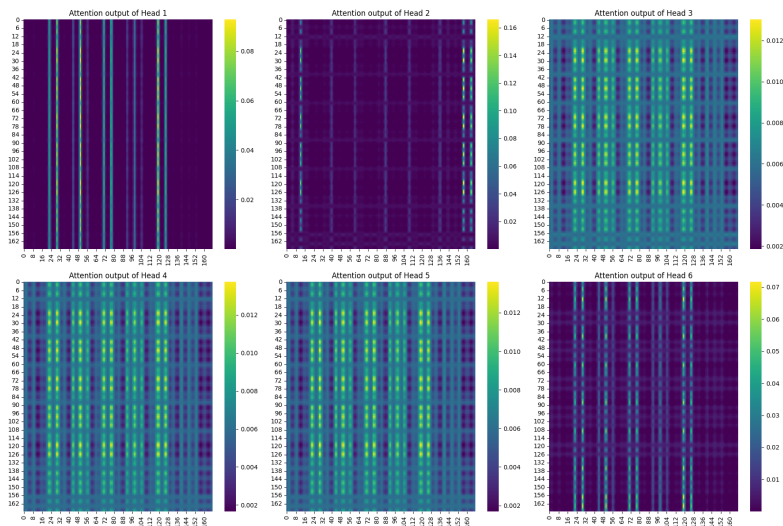
Όπως μπορεί να γίνει αντιληπτό από τον πίνακα 8.3, πλέον, η καλύτερη αρχιτεκτονική που έχει δοκιμαστεί περιέχει τον εμπλουτισμό του διανύσματος εισόδου με κωδικοποιήσεις θέσης και την είσοδό του σε ένα δίκτυο 7 πλήρως συνδεδεμένων επιπέδων νευρώνων με υπολειμματικές συνδέσεις, και την εισαγωγή του αποτελέσματος σε μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών, ενώ το μοντέλο μας παρέχει ήδη καλύτερες προβλέψεις από αυτό των Ευρωπαίων παρόχων, όπως φαίνεται στον πίνακα 8.4:

Πίνακας 8.4: Σύγκριση με τις προβλέψεις των παρόχων

Μοντέλο	MAPE	RMSE
Προβλέψεις παρόχων TSO	2.921 %	352
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια MLP Attention Models	2.822 %	365
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια Positional Attention Models	2.818 %	363

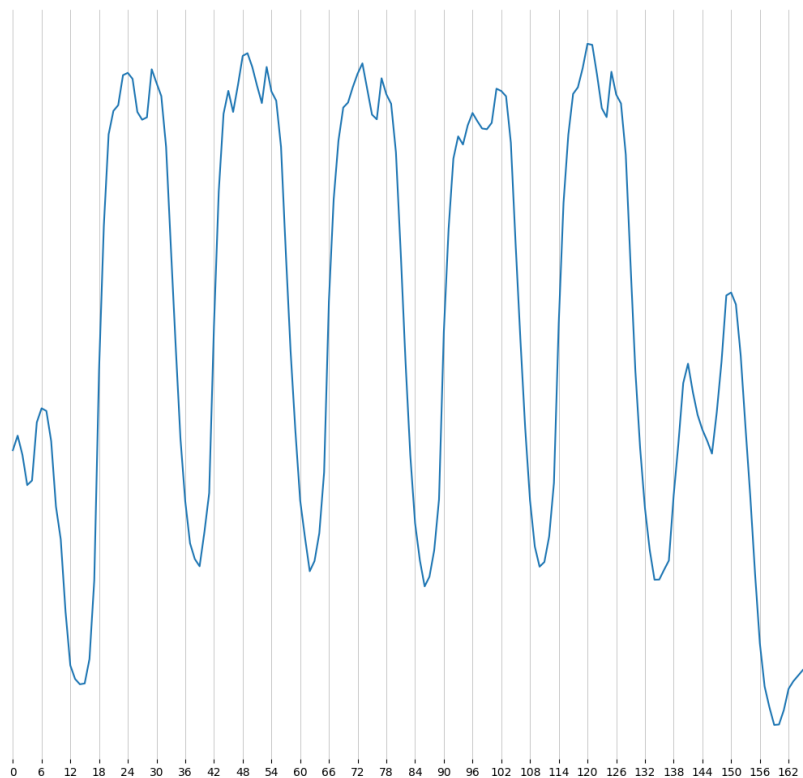
Παράλληλα, εκτυπώνοντας τον πίνακα προσοχής μπορούμε να αντιληφθούμε και σε ένα βαθμό να επεξηγήσουμε την επίδραση των κωδικοποιήσεων θέσης.

Για να προκύψουν οι κεφαλές που φαίνονται στην εικόνα 8.13 εισάγαμε στο μοντέλο μας μία τυπική εβδομάδα στην Πολωνία όπου η Κυριακή παρατηρείται στην αρχή του δείγματος



Εικόνα 8.13: Απεικόνιση των κεφαλών προσοχής μετά την χρήση κωδικοποιήσεων θέσης.

και το Σάββατο στο τέλος αυτού, όπως φαίνεται και στην εικόνα 8.14.



Εικόνα 8.14: Η είσοδος που χορηγήθηκε στο μοντέλο για να προκύψουν οι κεφαλές της εικόνας 8.13.

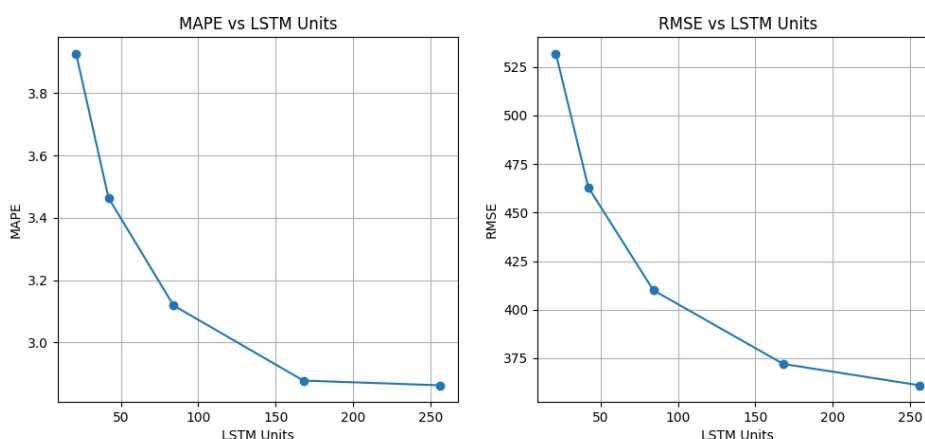
Όπως γίνεται και πάλι αντιληπτό, όπως και κατά την χρήση μηχανισμού προσοχής χωρίς

κωδικοποιήσεις θέσης, υπάρχει κεφαλή που δίνει μεγαλύτερη προσοχή στα σαββατοκύριακα και άλλες που δίνουν μεγαλύτερη προσοχή στις καθημερινές ημέρες. Το σημείο στο οποίο διαφοροποιείται ωστόσο η αρχιτεκτονική που χρησιμοποιεί και κωδικοποιήσεις θέσης είναι η λεπτομέρεια με την οποία επιλέγονται οι συγκεκριμένες ώρες στις οποίες πρέπει να δοθεί μεγαλύτερη έμφαση.

### 8.2.3 Αποτελέσματα από τη χρήση LSTM πριν τον μηχανισμό προσοχής.

Συνεχίζοντας τα πειράματα για την εύρεση του καλύτερου μηχανισμού που μπορεί να φανεί πιο χρήσιμος για την προεπεξεργασία της εισόδου πριν αυτή τροφοδοτηθεί στον μηχανισμό προσοχής, εξετάστηκαν και τα επίπεδα LSTM.

Τα βασικότερα στοιχεία υπό εξέταση της οικογένειας LSTM Attention Models, που παρουσιάστηκε λεπτομερώς στην προηγούμενη ενότητα, είναι το πλήθος των μονάδων στο επίπεδο LSTM καθώς και το πλήθος των επιπέδων LSTM που εισήχθησαν πριν τον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών. Παρακάτω φαίνεται η επίδραση του διαφορετικού πλήθους μονάδων LSTM στην προγνωστική ικανότητα του μοντέλου.



Εικόνα 8.15: Η επίδραση του πλήθους των μονάδων του LSTM.

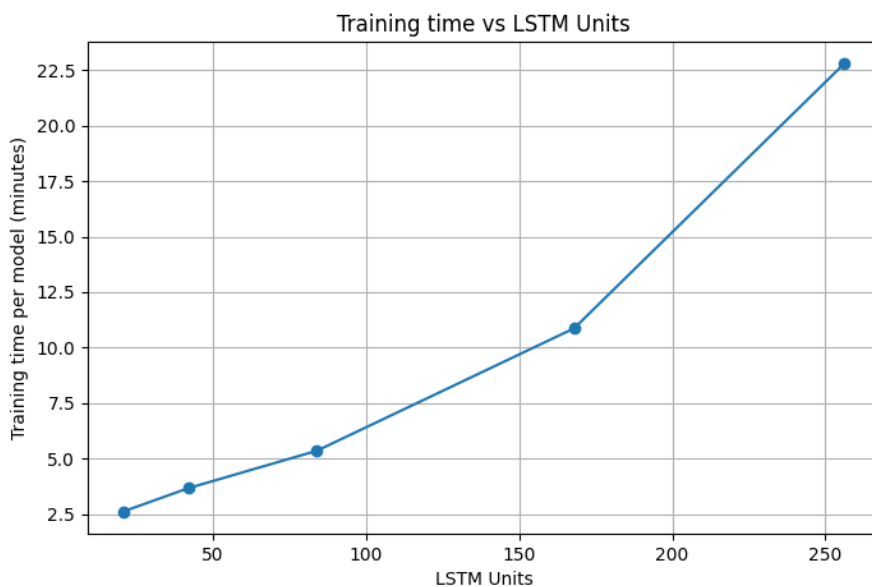
Όπως γίνεται αντιληπτό από την εικόνα 8.15, η ακρίβεια του μοντέλου αυξάνεται καθώς προστίθενται επιπλέον μονάδες, φτάνοντας μέχρι και τις 256.

Ωστόσο, αξίζει να σημειωθεί πως η αύξηση των μονάδων αυξάνει εκθετικά και τον χρόνο εκπαίδευσης ενός μοντέλου, σε βαθμό που δεν δικαιολογεί την προγνωστική ικανότητα της οικογένειας LSTM Attention Models σε σύγκριση με την οικογένεια Positional Attention Models που εξετάστηκε προηγουμένως (εικόνα 8.16).

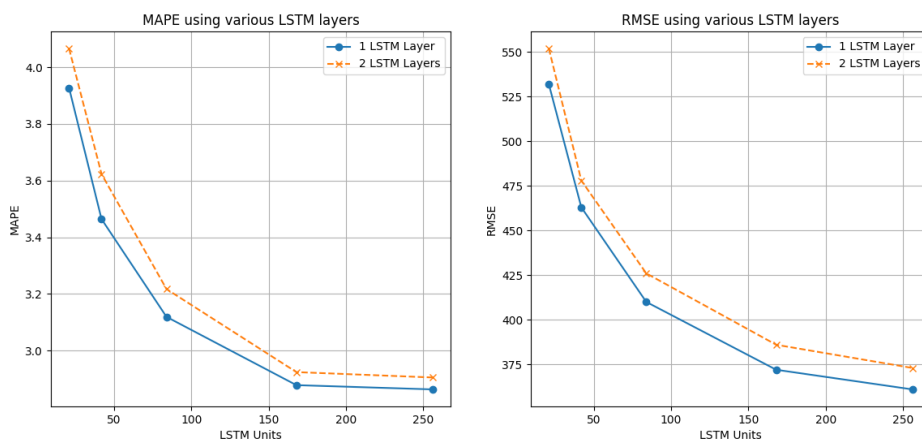
Παράλληλα, δοκιμάστηκε και η εισαγωγή επιπλέον επιπέδων LSTM, όπως φαίνεται στην εικόνα 8.17.

Όπως γίνεται αντιληπτό, η προσθήκη επιπλέον LSTM δεν μειώνει το σφάλμα πρόβλεψης. Επιπλέον αξίζει να σημειωθεί πως δοκιμάστηκε και η χρήση πλήρως συνδεδεμένων επιπέδων νευρώνων μετά το επίπεδο LSTM και πριν τον μηχανισμό προσοχής χωρίς αυτό και πάλι να βελτιώσει τις προβλέψεις που δίνει στην έξοδό του το μοντέλο.

Τα συγκεντρωτικά αποτελέσματα της μέχρι τώρα έρευνας παρουσιάζονται στον πίνακα 8.5:



Εικόνα 8.16: Η αύξηση του χρόνου εκπαίδευσης καθώς αυξάνονται οι μονάδες του LSTM επιπέδου.



Εικόνα 8.17: Η επίδραση των πολλαπλών επιπέδων LSTM.

Πίνακας 8.5: Σύγκριση MLP, Positional, LSTM Attention Models με τις προβλέψεις των παρόχων.

Μοντέλο	MAPE	RMSE
Προβλέψεις παρόχων TSO	2.921 %	352
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια LSTM Attention Models	2.863 %	372
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια MLP Attention Models	2.822 %	365
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια Positional Attention Models	2.818 %	363



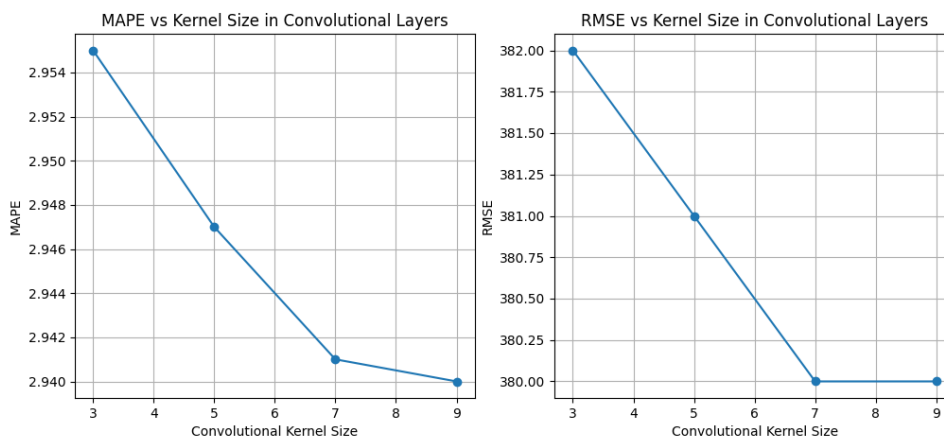
Όπως φαίνεται και στον πίνακα 8.5, τα μοντέλα της οικογένειας LSTM Attention Models εκμεταλλεύονται την δύναμη του μηχανισμού προσοχής και καταφέρνουν να δώσουν πιο ακριβείς προβλέψεις από αυτές των Ευρωπαϊών παρόχων ηλεκτρικής ενέργειας. Ωστόσο, δεν καταφέρνουν να ανταγωνιστούν την προγνωστική ικανότητα της οικογένειας Positional Attention Models.

Εδώ αξίζει να σημειωθεί, πως εξαιτίας των παραπάνω αποτελεσμάτων, δοκιμάστηκε να εισαχθεί στο LSTM επίπεδο διάνυσμα εμπλουτισμένο με κωδικοποιήσεις θέσης, ωστόσο και πάλι το αποτέλεσμα δεν ήταν καλύτερο από αυτό που παρουσιάστηκε παραπάνω.

#### 8.2.4 Αποτελέσματα από τη χρήση CNN πριν τον μηχανισμό προσοχής.

Η τελευταία αρχιτεκτονική που δοκιμάστηκε για την προεπεξεργασία του διανύσματος εισόδου, πριν αυτό εισέλθει στον μηχανισμό προσοχής είναι τα συνελικτικά επίπεδα. Η οικογένεια Convolutional Attention Models παρουσιάστηκε εκτενώς στην προηγούμενη ενότητα, ενώ εκτενείς ήταν και οι πειραματισμοί που έγιναν με τα συγκεκριμένα μοντέλα, των οποίων τα αποτελέσματα παρουσιάζονται παρακάτω. Ανάμεσα στα σημαντικότερα στοιχεία που εξετάστηκαν ως προς την επίδραση στην προγνωστική ικανότητα του μοντέλου ήταν το μέγεθος του πυρήνα συνέλιξης, το πλήθος των συνελικτικών φίλτρων καθώς και το είδος και η έκταση του pooling που εφαρμόστηκε μετά το συνελικτικό επίπεδο.

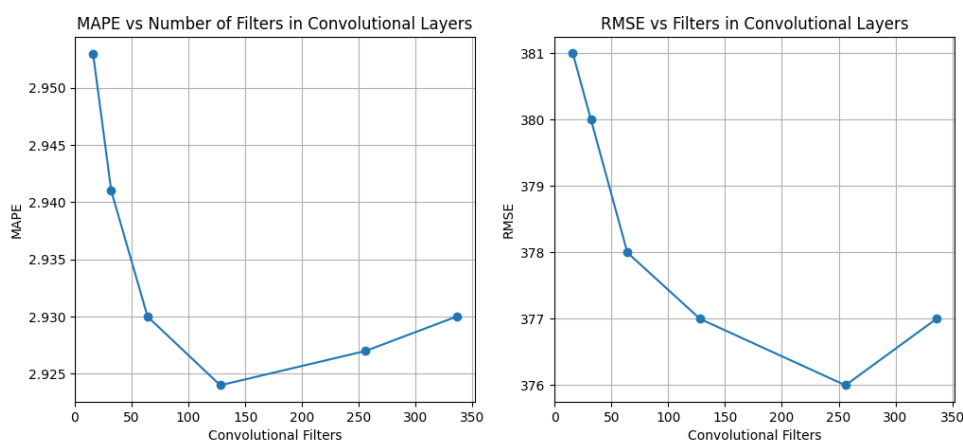
Πιο συγκεκριμένα, στην εικόνα 8.18 φαίνεται η επίδραση του μεγέθους του συνελικτικού πυρήνα στην προγνωστική ικανότητα του μοντέλου.



Εικόνα 8.18: Η επίδραση του μεγέθους του συνελικτικού πυρήνα.

Όπως φαίνεται, η αύξηση του πυρήνα μέχρι ένα σημείο βελτιώνει τα αποτελέσματα, ωστόσο, από το 9 και μετά τα πειράματα που έγιναν έδειξαν μια απότομη πτώση της ακρίβειας των προγνώσεων. Αυτό συμβαίνει καθώς, μέσω της πράξης της συνέλιξης το μοντέλο μαθαίνει υψηλότερου επιπέδου χαρακτηριστικά, εστιάζοντας στο επίπεδο των τιμών σε όλη τη γειτονιά της κάθε τιμής, χωρίς να εστιάζει περισσότερο από αυτό που χρειάζεται στην κάθε τιμή ξεχωριστά. Ωστόσο, η υπερβολική αύξηση αυτού του παραθύρου (άνω των 9 ωρών) φαίνεται να αφαιρεί χρήσιμη λεπτομέρεια από την εισερχόμενη πληροφορία.

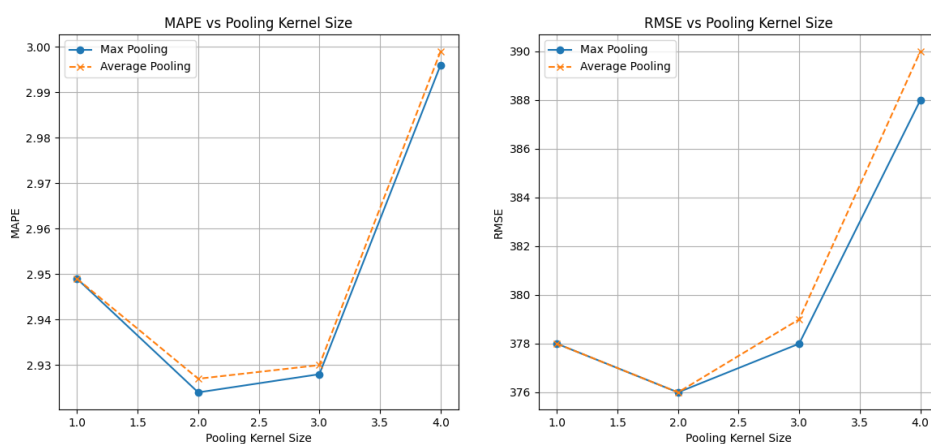
Επιπλέον, στην εικόνα 8.19 παρουσιάζεται και επίδραση του πλήθους των φίλτρων στην ακρίβεια των προβλέψεων.



Εικόνα 8.19: Η επίδραση του του πλήθους συνελικτικών φίλτρων.

Και πάλι φαίνεται πως η αύξηση μέχρι ενός σημείου είναι επωφελής.

Επιπροσθέτως εξετάστηκε και η επίδραση του pooling μετά το συνελικτικό επίπεδο στις προβλέψεις. Στην εικόνα 8.20, το μοναδιαίο μέγεθος πυρήνα αντιστοιχεί σε μη χρήση pooling.



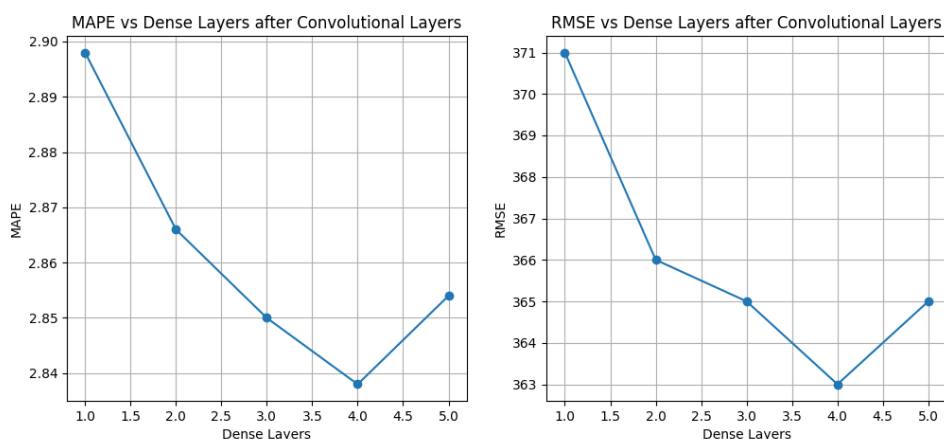
Εικόνα 8.20: Η επίδραση των διαφορετικών ειδών pooling.

Όπως γίνεται αντιληπτό, η χρήση max pooling περιορισμένης έκτασης φαίνεται να είναι επωφελής για την ακρίβεια των προβλέψεων.

Τέλος, εξετάστηκε και η χρήση πλήρως συνδεδεμένων επιπέδων πριν τον μηχανισμό προσοχής και μετά το συνελικτικό επίπεδο (εικόνα 8.21).

Φαίνεται πως η χρήση πλήρως συνδεδεμένων επιπέδων σε αυτό σημείο βοηθάει την παραγωγή πιο ακριβών προβλέψεων. Αξίζει να σημειωθεί πως δοκιμάστηκε και η χρήση πλήρως συνδεδεμένων επιπέδων και πριν το συνελικτικό επίπεδο, ωστόσο δεν είχαν αντίστοιχη συνεισφορά στην προγνωστική ακρίβεια. Τέλος, δοκιμάστηκε η χρήση και επιπλέον συνελικτικών επιπέδων, τα οποία όμως και πάλι δεν βελτίωσαν τις προβλέψεις.

Συγκεντρωτικά, τα αποτελέσματα μέχρι στιγμής παρουσιάζονται στον πίνακα 8.6.



Εικόνα 8.21: Η επίδραση του πλήθους πλήρως συνδεδεμένων επιπέδων νευρώνων μετά το συνελκτικό επίπεδο.

Πίνακας 8.6: Σύγκριση αποτελεσμάτων μετά την εξέταση όλων των βασικών οικογενειών μοντέλων.

	MAPE	RMSE
Προβλέψεις παρόχων TSO	2.921 %	352
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια LSTM Attention Models	2.863 %	372
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια Convolutional Attention Models	2.838 %	363
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια Positional Attention Models	2.818 %	363

### 8.3 Αρχιτεκτονικές που έπονται του Μηχανισμού Προσοχής

Οι αρχιτεκτονικές που συγκαταλέγονται στις οικογένειες Convolutional Attention Models και Positional Attention Models φαίνεται να συνδυάζουν χαμηλό χρόνο εκπαίδευσης (από 5 μέχρι 7 λεπτά ανά μοντέλο) ενώ παράλληλα παράγουν και ακριβείς προβλέψεις. Για το λόγο αυτό, τα πειράματα που θα ακολουθήσουν θα επεκτείνουν αυτές τις δύο αρχιτεκτονικές. Εδώ αξίζει να σημειωθεί πως οι δύο αρχιτεκτονικές μπορεί να παράγουν ακριβείς προβλέψεις, ωστόσο αυτές προκύπτουν με διαφορετικό τρόπο, όπως έχει αναλυθεί και παραπάνω. Εκτυπώνοντας τον πίνακα προσοχής (σε προηγούμενη υποενότητα) διαπιστώσαμε πως οι αρχιτεκτονικές που ανήκουν στην οικογένεια Positional Attention Models εκμεταλλεύονται λεπτομέρειες που ποικίλουν από ώρα σε ώρα στα δεδομένα εισόδου. Απο την άλλη, οι αρχιτεκτονικές που ανήκουν στην οικογένεια Convolutional Attention Models λόγω της χρήσης συνέλιξης μαθαίνουν λαμβάνοντας σταθμισμένη πληροφορία από την στάθμη των τιμών όλου του παραθύρου γειτόνων της κάθε τιμής. Με αυτό τον τρόπο, αποφεύγουν την εκμάθηση λεπτομερειών που δεν θα ήταν χρήσιμες για την δημιουργία ακριβών προβλέψεων. Η σημασία της διαφοράς αυτής θα γίνει αντιληπτή στις επόμενες υποενότητες.

### 8.3.1 Χρήση υπολειμματικής σύνδεσης μετά τον Μηχανισμό Προσοχής

Όπως εξηγήθηκε αναλυτικά στην προηγούμενη ενότητα, δοκιμάστηκαν διάφορα είδη υπολειμματικών συνδέσεων γύρω από τον μηχανισμό αυτοπροσοχής πολλαπλών κεφαλών, χρησιμοποιώντας τα καλύτερα μοντέλα από τις οικογένειες Positional Attention Models και Convolutional Attention Models.

Αρχικά, ξεκινώντας από τα συνελικτικά δίκτυα, δοκιμάστηκε η χρήση υπολειμματικής σύνδεσης μεταξύ της εισόδου του δικτύου και της εξόδου του μηχανισμού αυτοπροσοχής. Επιπλέον, χρησιμοποιήθηκε και η χρήση υπολειμματικής σύνδεσης μεταξύ της εισόδου του μηχανισμού προσοχής και της εξόδου αυτού, προσθέτοντας τις δύο αναπαραστάσεις και προωθώντας τις στο πλήρως συνδεδεμένο επίπεδο νευρώνων για να προκύψουν οι προβλέψεις. Τα αποτελέσματα των δύο διαφορετικών προσεγγίσεων παρουσιάζονται στον πίνακα 8.7.

Πίνακας 8.7: Σύγκριση υπολειμματικών συνδέσεων στα *Convolutional Attention Models*

Είδος υπολειμματικής σύνδεσης	MAPE	RMSE
Χωρίς υπολειμματική σύνδεση	2.838 %	363
Είσοδος μηχανισμού προσοχής - Έξοδος μηχανισμού προσοχής	2.888 %	367
Είσοδος μοντέλου - Έξοδος μηχανισμού προσοχής	2.766 %	354

Αντίστοιχοι πειραματισμοί έγιναν και στην περίπτωση της οικογένειας μοντέλων Positional Attention Models. Όπως αναφέρθηκε και στην προηγούμενη ενότητα, δοκιμάστηκαν τρεις διαφορετικές εσωτερικές αναπαραστάσεις του μοντέλου οι οποίες θα μπορούσαν να προστεθούν στην έξοδο του μηχανισμού προσοχής.

Αρχικά χρησιμοποιήθηκε μία υπολειμματική σύνδεση μεταξύ της εισόδου του δικτύου και της εξόδου του μηχανισμού προσοχής. Το άθροισμα αυτών των δύο διανυσμάτων τροφοδοτείται έπειτα στο τελευταίο επίπεδο πλήρως συνδεδεμένων νευρώνων έτσι ώστε να προκύψουν οι προβλέψεις. Οι δύο υπόλοιπες υπολειμματικές συνδέσεις που δοκιμάστηκαν ακολουθούν περισσότερο την λογική των αντίστοιχων συνδέσεων στην αρχική αρχιτεκτονική του Τρανσφορμερ έτσι όπως παρουσιάστηκε στην έρευνα "Attention Is All You Need". Δοκιμάστηκε η άθροιση του εμπλουτισμένου με κωδικοποίηση θέσης διανύσματος με το διάνυσμα εξόδου του μηχανισμού προσοχής, ενώ δοκιμάστηκε και η άθροιση της εισόδου με την έξοδο του μηχανισμού. Το άθροισμα αυτό, σε κάθε περίπτωση τροφοδοτείται στο τελευταίο επίπεδο πλήρως συνδεδεμένων νευρώνων. Στον πίνακα 8.8 παρουσιάζονται και τα αποτελέσματα από αυτά τα πειράματα.

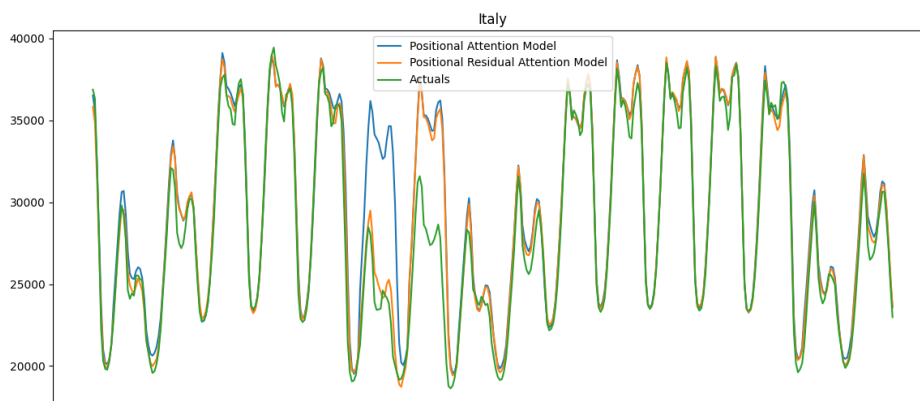
Όπως γίνεται αντιληπτό και στις δύο περιπτώσεις (πίνακες 8.7, 8.8), οι υπολειμματικές συνδέσεις ενισχύουν σε πολύ μεγάλο βαθμό την προγνωστική ικανότητα και των δύο οικογενειών μοντέλων. Πιο συγκεκριμένα, η σύνδεση που φαίνεται να δίνει τα καλύτερα αποτελέσματα είναι αυτή μεταξύ της αρχικής εισόδου του μοντέλου και της εξόδου του μηχανισμού προσοχής. Αυτό είναι και ένα από τα σημεία στα οποία διαφοροποιείται η παρούσα εργασία από την υπάρχουσα δημοσιευμένη έρευνα. Συνήθως, οι υπολειμματικές συνδέσεις που χρησιμοποιούνται συνδέουν την είσοδο με την έξοδο του μηχανισμού προσοχής.

Πίνακας 8.8: Σύγκριση υπολειμματικών συνδέσεων στα *Positional Attention Models*

Είδος υπολειμματικής σύνδεσης	MAPE	RMSE
Χωρίς υπολειμματική σύνδεση	2.818 %	363
Είσοδος μηχανισμού προσοχής - Έξοδος μηχανισμού προσοχής	2.831 %	363
Σημείο προσθήκης κωδικοποιήσεων θέσης - Έξοδος μηχανισμού προσοχής	2.786 %	360
Είσοδος μοντέλου - Έξοδος μηχανισμού προσοχής	2.770 %	359

Αυτή η διαφοροποίηση φαίνεται να δίνει καλύτερα αποτελέσματα καθώς με αυτόν τον τρόπο δίνεται μεγαλύτερη ευελιξία στο μοντέλο να διαχειριστεί τις περιπτώσεις τις οποίες ο μηχανισμός προσοχής δεν μπορεί να αντιμετωπίσει τόσο αποδοτικά. Στην αρχή της προηγούμενης ενότητας παρουσιάστηκε πως ο μηχανισμός προσοχής φαίνεται να έχει την ικανότητα να προβλέπει τιμές με έντονη εποχικότητα. Ωστόσο, λόγω της προσοχής που μαθαίνει να δείχνει στις εποχικές διακυμάνσεις φαίνεται να αδυνατεί να προβλέψει ειδικά γεγονότα που εμφανίζονται σχετικά σπάνια. Προωθώντας τις αρχικές τιμές της χρονοσειράς στο τελευταίο επίπεδο επεξεργασίας της αρχιτεκτονικής δίνουμε την ευελιξία στο τελευταίο πλήρως συνδεδεμένο επίπεδο να διαχειριστεί αυτές τις ιδιαίχουσες περιπτώσεις χωρίς να δίνεται περισσότερη προσοχή από όση χρειάζεται στην εποχικότητα της χρονοσειράς ή να “μπερδεύουμε” το μοντέλο με κωδικοποιήσεις θέσης που θα το προέτρεπαν να ακολουθήσει το εποχικό μοτίβο που έχει μάθει.

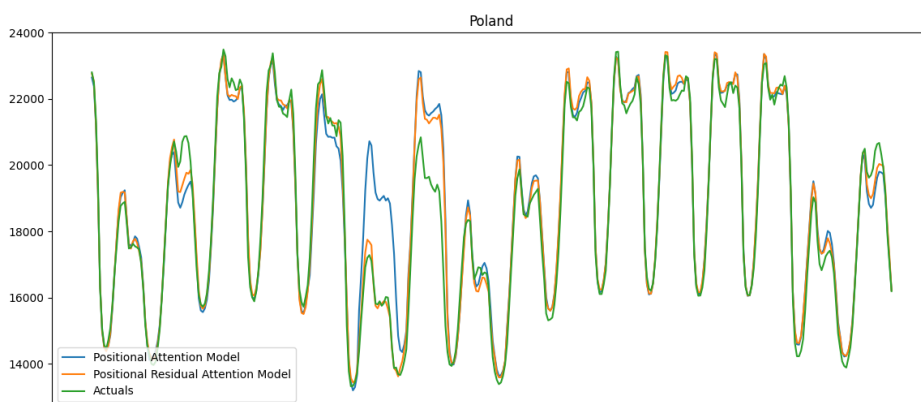
Η ορθότητα του παραπάνω επιχειρήματος φαίνεται στα παραδείγματα που παρουσιάζονται παρακάτω.



Εικόνα 8.22: Η επίδραση των υπολειμματικών συνδέσεων στην πρόβλεψη ειδικού γεγονότος στην Ιταλία.

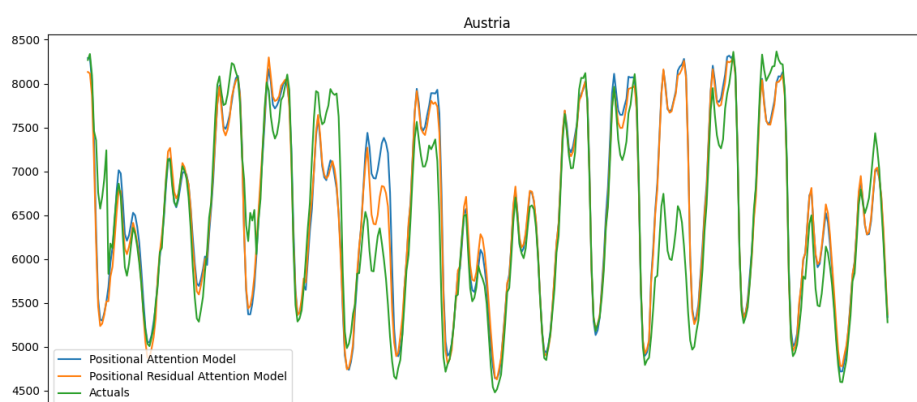
Στο παράδειγμα της εικόνας 8.22 από την Ιταλία, η κατανάλωση ηλεκτρικής ενέργειας την πρώτη εβδομάδα που παρουσιάζεται και συγκεκριμένα την Πέμπτη πέφτει απροσδόκητα. Αυτό το ειδικό γεγονός φαίνεται να μπορεί να προβλεφθεί με τη χρήση της υπολειμματικής σύνδεσης, ενώ χωρίς αυτή, το μοντέλο θεωρεί την συγκεκριμένη μέρα ως μία τυπική Πέμπτη. Φυσικά αυτό δεν είναι πάντα εφικτό όπως γίνεται αντιληπτό την αμέσως επόμενη μέρα (Παρασκευή).

Ένα ακόμη αντίστοιχο παράδειγμα φαίνεται στην εικόνα 8.23 από τη χρονοσειρά της Πολωνίας, το οποίο μάλιστα απεικονίζει τις ίδιες εβδομάδες με το προηγούμενο παράδειγμα.



Εικόνα 8.23: Η επίδραση των υπολειμματικών συνδέσεων στην πρόβλεψη ειδικού γεγονότος στην Πολωνία.

Τέλος, μία ακόμη Πέμπτη η οποία εμφάνισε ιδιόμορφη συμπεριφορά παρουσιάζεται στην εικόνα 8.24 από την Αυστρία.



Εικόνα 8.24: Η επίδραση των υπολειμματικών συνδέσεων στην πρόβλεψη ειδικού γεγονότος στην Αυστρία.

Την πρώτη εβδομάδα που απεικονίζεται, η χρήση της υπολειμματικής σύνδεσης φαίνεται να δίνει την δυνατότητα πρόβλεψης του ειδικού γεγονότος.

Συγκεντρωτικά, τα καλύτερα μοντέλα μέχρι στιγμής παρουσιάζονται στον πίνακα 8.9:

### 8.3.2 Χρήση MLP, LSTM, CNN μετά τον μηχανισμό προσοχής

Όπως αναφέρθηκε και στην προηγούμενη ενότητα, μετά την υπολειμματική σύνδεση ερευνήθηκε ποιές νευρωνικές αρχιτεκτονικές θα μπορούσαν να ακολουθήσουν το μηχανισμό προσοχής έτσι ώστε να χρησιμοποιήσουν την χρήσιμη πληροφορία που δίνει αλλά και την

Πίνακας 8.9: Συγκριτικά αποτελέσματα μετά και την προσθήκη υπολειμματικών συνδέσεων γύρω από τους μηχανισμούς προσοχής

	MAPE	RMSE
Προβλέψεις παρόχων TSO	2.921 %	352
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια Convolutional Attention Models	2.766 %	354
Προβλέψεις με χρήση της βέλτιστης αρχιτεκτονικής από την οικογένεια Positional Attention Models	2.770 %	359

πληροφορία του διανύσματος εισόδου με το οποίο προστίθεται η έξοδος του μηχανισμού προσοχής. Δοκιμάστηκαν πλήρως συνδεδεμένα επίπεδα νευρώνων, επίπεδα LSTM καθώς και συνελικτικά επίπεδα.

Η γενική παρατήρηση που προκύπτει είναι πώς και στις δύο οικογένειες μοντέλων η προσθήκη επιπλέον αρχιτεκτονικών μετά τον μηχανισμό προσοχής δεν βοηθάει ιδιαίτερα στην παραγωγή πιο ακριβών προβλέψεων, ενώ πολλές φορές δυσχεραίνει και την προγνωστική ικανότητα του μοντέλου. Συνεπώς, μετά και την υπολειμματική σύνδεση, το πλήρως συνδεδεμένο επίπεδο 36 νευρώνων που χρησιμοποιείται φαίνεται να είναι αρκετό για την δημιουργία αξιόπιστων προβλέψεων. Για το λόγο αυτό, στην παρούσα υποενοότητα δεν θα γίνει εκτενής περιγραφή όλων των παραμέτρων που εξετάστηκαν, καθώς κάτι τέτοιο δεν προσφέρει κάτι στην παρουσίαση της συγκεκριμένης έρευνας.

Επιγραμματικά, ως προς την ικανότητα των πλήρως συνδεδεμένων επιπέδων νευρώνων να βοηθήσουν στην παραγωγή προβλέψεων μετά τον μηχανισμό προσοχής, διαπιστώθηκε πως όσα περισσότερα επίπεδα προστίθενται, τόσο περισσότερο μειώνεται και η ακρίβεια των προβλέψεων. Επιπλέον, δοκιμάστηκε και η χρήση υπολειμματικών συνδέσεων μεταξύ των επιπέδων αυτών, πρακτική που είχε βοηθήσει στην αντίστοιχη εξερεύνηση πριν τον μηχανισμό προσοχής ωστόσο και πάλι δεν παρατηρήθηκε αύξηση στην ακρίβεια των παραγόμενων προβλέψεων.

Επιπλέον, όπως αναφέρθηκε παραπάνω, εξετάστηκε και η χρήση LSTM καθώς και συνελικτικών επιπέδων. Ωστόσο και οι δύο αρχιτεκτονικές μείωσαν εμφανώς την ακρίβεια του μοντέλου, ενώ παράλληλα αύξησαν και τον χρόνο εκπαίδευσης.

Ενδεικτικά, δίνονται στον πίνακα τα καλύτερα αποτελέσματα που προέκυψαν από αυτή την εξερεύνηση, χρησιμοποιώντας ως βάση την αρχιτεκτονική των Convolutional Attention Models:

Πίνακας 8.10: Παρουσίαση αποτελεσμάτων από την προσθήκη επιπλέον επιπέδων επεξεργασίας πάνω από τον μηχανισμό προσοχής

Βέλτιστη αρχιτεκτονική μετά τον μηχανισμό προσοχής	MAPE	RMSE
Χωρίς επιπλέον επεξεργασία	2.766 %	354
Χρήση ενός επιπλέον επιπέδου νευρώνων	2.774 %	360
Χρήση επιπέδου LSTM	2.835 %	364
Χρήση συνελικτικού επιπέδου ακολουθούμενο από pooling	2.820 %	364

Επιπλέον, αξίζει να σημειωθεί πως δεδομένης της αδυναμίας των παραπάνω αρχιτεκτονικών

να χρησιμοποιήσουν την έξοδο του μηχανισμού προσοχής για να βελτιώσουν τις προβλέψεις των μοντέλων, δοκιμάστηκε και η χρήση του ενός Βλοσκ (είτε Convolutional είτε Positional) μετά το άλλο. Ωστόσο και πάλι τα αποτελέσματα έμειναν στάσιμα.

## 8.4 Ο Transformer

Όπως περιγράφηκε και στην προηγούμενη ενότητα, οι δύο καλύτερες αρχιτεκτονικές από τους προηγούμενους πειραματισμούς επιλέχθηκαν για την κατασκευή μίας αρχιτεκτονικής που θυμίζει την δομή του γνωστού Transformer, ωστόσο διαφέρει σημαντικά σε διάφορα σημεία του, τα οποία αναλύθηκαν εκτενώς. Η επιλογή των οικογενειών Positional Attention Models και Convolutional Attention Models ως δομικά στοιχεία για την κατασκευή του Transformer δεν προέκυψε τυχαία.

Εκτυπώνοντας τον πίνακα προσοχής (σε προηγούμενη υποενότητα) διαπιστώσαμε πως οι αρχιτεκτονικές που ανήκουν στην οικογένεια Positional Attention Models εκμεταλλεύονται λεπτομέρειες που ποικίλουν από ώρα σε ώρα στα δεδομένα εισόδου. Από την άλλη, οι αρχιτεκτονικές που ανήκουν στην οικογένεια Convolutional Attention Models λόγω της χρήσης συνέλιξης μαθαίνουν λαμβάνοντας σταθμισμένη πληροφορία από την στάθμη των τιμών όλου του παραθύρου γειτόνων της κάθε τιμής. Με αυτό τον τρόπο, αποφεύγουν την εκμάθηση λεπτομερειών που δεν θα ήταν χρήσιμες για την δημιουργία ακριβών προβλέψεων.

Ως εκ τούτου, επιλέχθηκε να χρησιμοποιηθεί η οικογένεια Convolutional Attention Models στην θέση του Encoder, έτσι ώστε να μάθει χαρακτηριστικά υψηλότερου επιπέδου και το αποτέλεσμα που θα προκύψει να δοθεί ως διανύσματα Q, K στον Decoder, ο οποίος επιλέχθηκε να δομηθεί από αρχιτεκτονική της οικογένειας Positional Attention Models, έτσι ώστε μετά την εκμάθηση των χαρακτηριστικών υψηλότερου επιπέδου στον Encoder να προκύψει και η εκμάθηση, πιο στοχευμένα πλέον, λεπτομερειών. Η εγκυρότητα των παραπάνω επιχειρημάτων αλλά και η προγνωστική ικανότητα του Transformer που κατασκευάστηκε προκύπτει και από τον πίνακα 8.11 όπου παρουσιάζονται συνοπτικά αποτελέσματα από τα διάφορα πειράματα που έχουν γίνει.

Πίνακας 8.11: Παρουσίαση της προγνωστικής ικανότητας του Transformer

	MAPE	RMSE
Προβλέψεις παρόχων TSO	2.921 %	352
Decoder-only Positional Attention Models	2.770 %	359
Decoder-only Convolutional Attention Models	2.766 %	354
Transformer με χρήση αποκλειστικά Positional Attention Models	2.776 %	357
Transformer με χρήση αποκλειστικά Convolutional Attention Models	2.773 %	358
Transformer με Positional Encoder και Convolutional Decoder	2.787 %	357
<b>Transformer με Convolutional Encoder και Positional Decoder</b>	<b>2.754 %</b>	<b>354</b>

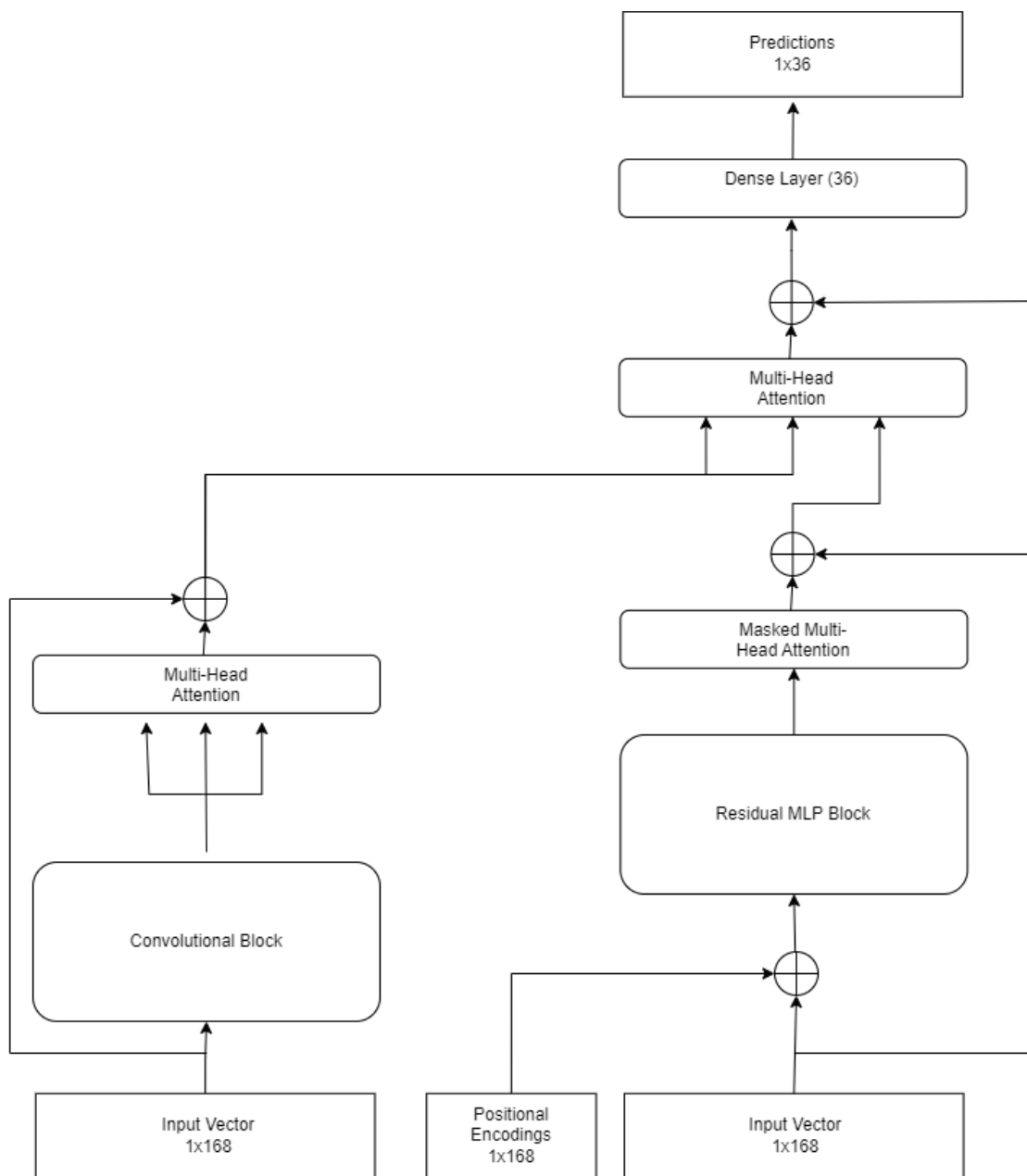
Τέλος, όπως μπορεί να φανεί και στον πίνακα 8.12, λόγω της χρήσης περισσότερων μηχανο-



νισμών προσοχής στην προτεινόμενη αρχιτεκτονική του Transformer, ο χρόνος εκπαίδευσης αυξάνεται αρκετά. Ωστόσο, αυτού του είδους η αύξηση θεωρήθηκε αποδεκτή, δεδομένης της αύξησης της προγνωστικής ικανότητας του μοντέλου. Σε προηγούμενες ενότητες φάνηκε πως όταν η αύξηση του χρόνου εκπαίδευσης δεν συνοδεύεται από αύξηση της προγνωστικής ικανότητας του μοντέλου, τότε η συγκεκριμένη αρχιτεκτονική δεν θεωρήθηκε πετυχημένη. Κλασσικό παράδειγμα αποτελεί και η περίπτωση των LSTM-Attention Models.

Πίνακας 8.12: Χρόνοι εκπαίδευσης των υπόλοιπων μοντέλων συγκριτικά με τον Transformer

Βέλτιστη Αρχιτεκτονική ανά οικογένεια	Χρόνος εκπαίδευσης ανά μοντέλο	MAPE
Decoder-only Basic Attention Models	4.71 λεπτά	4.431 %
Decoder-only Positional Attention Models	5.58 λεπτά	2.770 %
Decoder-only Convolutional Attention Models	6.13 λεπτά	2.766 %
Decoder-only LSTM Attention Models	13.52 λεπτά	2.863 %
<b>Transformer με Convolutional Encoder και Positional Decoder</b>	<b>17.60 λεπτά</b>	<b>2.754 %</b>



Εικόνα 8.25: Η τελική μορφή του Transformer

## Κεφάλαιο 9

### Επίλογος

---

#### 9.1 Συμπεράσματα

Στα προηγούμενα δύο κεφάλαια παρουσιάστηκαν όλα τα πειράματα που έγιναν στην παρούσα έρευνα, ενώ αναλύθηκαν και τα αποτελέσματα αυτών. Παρακάτω, σε συνέχεια της ροηγούμενης ανάλυσης, έχουν συλλεχθεί τα βασικότερα συμπεράσματα που προέκυψαν κατά την εκπόνηση της παρούσας διπλωματικής εργασίας:

- Κατά την κατασκευή νευρωνικών αρχιτεκτονικών με μηχανισμό προσοχής που στόχο έχουν την πρόβλεψη χρονοσειρών, ιδιαίτερα ευεργετική ως προς την ακρίβεια των προβλέψεων φαίνεται να είναι η χρήση πολλαπλών κεφαλών, καθώς η κάθε μία από τις κεφαλές φαίνεται να εξειδικεύεται στην αναγνώριση διαφορετικών χαρακτηριστικών του διανύσματος εισόδου. Χαρακτηριστικό είναι το παράδειγμα που παρουσιάζεται στην εικόνα 8.8, όπου είναι εμφανές πως κάποιες κεφαλές εστιάζουν την προσοχή τους σε καθημερινές μέρες της εβδομάδας εισόδου, ενώ άλλες κεφαλές εστιάζουν σε πληροφορία που μπορεί να εξαχθεί από την κατανάλωση ενέργειας το προηγούμενο Σαββατοκύριακο.
- Παράλληλα, κατά την μελέτη που παρουσιάστηκε στα προηγούμενα κεφάλαια, έγινε σαφές πως οι αρχιτεκτονικές που περιέχουν μηχανισμούς προσοχής εμφανίζουν ιδιαίτερη ικανότητα στην εκμάθηση εποχιακών μοτίβων, ενώ δυσκολεύονται αρκετά στην πρόβλεψη γεγονότων που διαταράσσει αυτή την εποχική συμπεριφορά. Η εικόνα 8.10 επιβεβαιώνει την εγκυρότητα του παραπάνω ισχυρισμού.
- Ως προς τις νευρωνικές αρχιτεκτονικές των οποίων η τοποθέτηση πριν τον μηχανισμό προσοχής μειώνει το σφάλμα πρόβλεψης, η χρήση κωδικοποιήσεων θέσης και πλήρως συνδεδεμένων επιπέδων νευρώνων φαίνεται να επιτελεί αυτή τη λειτουργία. Επίσης πολύ χρήσιμη φαίνεται να είναι και η τοποθέτηση συνελικτικού επιπέδου πριν τον μηχανισμό προσοχής που φυσικά απολυνθείται από τα απαραίτητα πλήρως συνδεδεμένα επίπεδα νευρώνων. Ωστόσο, η χρήση LSTM πριν τον μηχανισμό προσοχής δεν φαίνεται να μπορεί να ανταγωνιστεί τις προηγούμενες δύο αρχιτεκτονικές (πίνακας 8.6).
- Κατά τη χρήση συνελικτικού επιπέδου, αλλά και κατά τη χρήση κωδικοποιήσεων θέσης έγινε σαφές πως η αύξηση των επιπέδων πλήρως συνδεδεμένων νευρώνων που τα ακολουθούν είναι ευεργετική για την αύξηση της ακρίβειας των προβλέψεων. Ωστόσο, η προσθήκη επιπλέον επιπέδων νευρώνων έχει νόημα μέχρι ενός σημείου, καθώς μετά

από ένα συγκεκριμένο πλήθος παρατηρείται το φαινόμενο του vanishing gradient και αυτό μειώνει την απόδοση του μοντέλου. Για την αντιμετώπιση του προβλήματος αυτού μπορούν να χρησιμοποιηθούν υπολειμματικές συνδέσεις μεταξύ των επιπέδων νευρώνων (πίνακας 8.3).

- Επιπλέον, έγινε αρκετά σαφές πως οι κωδικοποιήσεις θέσης έχουν την ικανότητα να δίνουν χρήσιμη πληροφορία στον μηχανισμό προσοχής την οποία δεν μπορεί να αποκτήσει μόνος του, καθώς δεν έχει την ικανότητα να αναγνωρίσει τις σχετικές θέσεις των τιμών στο διάλυσμα εισόδου. Ωστόσο, αν τα επίπεδα πλήρως συνδεδεμένων νευρώνων που ακολουθούν μετά την προσθήκη κωδικοποιήσεων θέσης είναι αρκετά, τότε η συνεισφορά των κωδικοποιήσεων είναι σχεδόν αμελητέα.
- Όπως αναφέρθηκε παραπάνω, η χρήση κωδικοποιήσεων θέσης και πλήρως συνδεδεμένων νευρώνων αλλά και η χρήση συνελικτικών επιπέδων πριν τον μηχανισμό προσοχής αυξάνουν ιδιαίτερα την προγνωστική ικανότητα του μηχανισμού. Ωστόσο, αυτό γίνεται με διαφορετικό τρόπο. Εκτυπώνοντας τον πίνακα προσοχής (εικόνα 8.13) διαπιστώσαμε πως οι αρχιτεκτονικές που ανήκουν στην οικογένεια Positional Attention Models εκμεταλλεύονται λεπτομέρειες που ποικίλουν από ώρα σε ώρα στα δεδομένα εισόδου, δηλαδή ο μηχανισμός σε αυτή την περίπτωση μαθαίνει να δίνει προσοχή σε συγκεκριμένες ζώνες ωρών (κυρίως μεσημβρινές). Αυτό μπορεί να γίνει ιδιαίτερος αντιληπτό κατά την σύγκριση των κεφαλών προσοχής των οικογενειών Positional Attention Models και Basic Attention Models. Απο την άλλη, οι αρχιτεκτονικές που ανήκουν στην οικογένεια Convolutional Attention Models λόγω της χρήσης συνέλιξης αλλά και max pooling μαθαίνουν λαμβάνοντας σταθμισμένη πληροφορία από την στάθμη των τιμών όλου του παραθύρου γειτόνων της κάθε τιμής. Με αυτό τον τρόπο, αποφεύγουν την εκμάθηση ιδιαίτερων λεπτομερειών που μπορεί να μην είναι χρήσιμες για την δημιουργία ακριβών προβλέψεων.
- Η αδυναμία των μηχανισμών προσοχής στην πρόβλεψη γεγονότων που διαταράσσουν την ομαλή εποχική συμπεριφορά των χρονοσειρών μπορεί να αντιμετωπιστεί σε μεγάλο βαθμό με τη χρήση υπολειμματικών συνδέσεων μεταξύ της εισόδου του μοντέλου και της εξόδου του μηχανισμού προσοχής. Αυτό είναι και ένα από τα σημεία στα οποία διαφοροποιείται η παρούσα εργασία από την υπάρχουσα δημοσιευμένη έρευνα. Συνήθως, οι υπολειμματικές συνδέσεις που χρησιμοποιούνται συνδέουν την είσοδο με την έξοδο του μηχανισμού προσοχής. Προωθώντας τις αρχικές τιμές της χρονοσειράς στο τελευταίο επίπεδο επεξεργασίας της αρχιτεκτονικής δίνουμε την ευελιξία στο τελευταίο πλήρως συνδεδεμένο επίπεδο να διαχειριστεί αυτές τις ιδιόζουσες περιπτώσεις διακοπής της εβδομαδιαίας εποχικότητας χωρίς να δίνεται περισσότερη προσοχή από όση χρειάζεται στην εποχικότητα ή να “μπερδεύουμε” το μοντέλο με κωδικοποιήσεις θέσης που θα το προέτρεπαν να ακολουθήσει το εποχικό μοτίβο που έχει μάθει. Η εγκυρότητα του παραπάνω επιχειρήματος επιβεβαιώνεται από τους πίνακες 8.7, 8.8 και από τις εικόνες 8.22, 8.23 και 8.24.
- Επιπλέον, μετά από αρκετά πειράματα, η παρούσα έρευνα κατέληξε πως η χρήση MLP, CNN ή LSTM μετά τον μηχανισμό προσοχής δεν συμβάλει στην περαιτέρω αύξηση

της ακρίβειας των προβλέψεων (πίνακας 8.10). Συνεπώς, μετά και την υπολειμματική σύνδεση, το πλήρως συνδεδεμένο επίπεδο 36 νευρώνων (όσες είναι και οι τιμές του ορίζοντα πρόβλεψης) που χρησιμοποιείται φαίνεται να είναι αρκετό για την δημιουργία αξιόπιστων προβλέψεων.

- Τέλος, η βασικότερη συνεισφορά της παρούσας έρευνας είναι η κατασκευή ενός νέου Transformer. Στην συγκεκριμένη αρχιτεκτονική που προτίνεται επιλέχθηκε να χρησιμοποιηθεί η οικογένεια Convolutional Attention Models στην θέση του Encoder, έτσι ώστε να μάθει χαρακτηριστικά υψηλότερου επιπέδου και το αποτέλεσμα που θα προκύψει να δοθεί ως διανύσματα Q, K στον Decoder, ο οποίος επιλέχθηκε να δομηθεί από αρχιτεκτονική της οικογένειας Positional Attention Models, έτσι ώστε μετά την εκμάθηση των χαρακτηριστικών υψηλότερου επιπέδου στον Encoder να προκύψει και η εκμάθηση, πιο στοχευμένα πλέον, λεπτομερειών. Η εγκυρότητα των παραπάνω επιχειρημάτων αλλά και η προγνωστική ικανότητα του Transformer που κατασκευάστηκε προκύπτει και από τον πίνακα 8.11 όπου παρουσιάζονται συνοπτικά αποτελέσματα από διάφορα πειράματα που έγιναν.

## 9.2 Μελλοντικές Επεκτάσεις

Όπως έχει γίνει σαφές, η παρούσα εργασία παρουσιάζει την ικανότητα των μηχανισμών προσοχής να παράγουν ακριβείς προβλέψεις ενώ παράλληλα προτείνει και έναν νέο Transformer ο οποίος φαίνεται να έχει πολύ ισχυρές προβλεπτικές ικανότητες. Οι μηχανισμοί προσοχής φαίνεται να μαθαίνουν πολύ καλά τα εποχικά δεδομένα, ωστόσο, όπως φαίνεται από το σύνολο δεδομένων που χρησιμοποιήθηκε στην παρούσα έρευνα, δυσκολεύεται να αναγνωρίσει και να προβλέψει ειδικά γεγονότα. Η συγκεκριμένη εργασία προσέφερε διάφορους τρόπους για την αντιμετώπιση αυτού του προβλήματος, ωστόσο υπάρχει εκτενής έρευνα πάνω στο συγκεκριμένο αντικείμενο η οποία εντείνεται τα τελευταία χρόνια.

Μία από τις ιδέες που φαίνεται να είναι πολύ ενδιαφέρουσες και θα άξιζε να διερευνηθεί η εφαρμογή της στην υπό εξέταση αρχιτεκτονική είναι η χρήση παραγωγικών μοντέλων (generative models) για την πρόβλεψη ανωμαλιών.

Σε πολλαπλές μελέτες, συμπεριλαμβανομένων των TranAD [168], MT-RVAE [169], και TransAnomaly [170], οι ερευνητές πρότειναν τον συνδυασμό του Transformer με νευρωνικά παραγωγικά μοντέλα, όπως τα VAEs [171] και τα GANs [172], για καλύτερη απόδοση στην ανίχνευση ανωμαλιών. Πολύ ενδιαφέρουσα θα ήταν και η χρήση Diffusion μοντέλων, καθώς αυτή η ιδέα φαίνεται να μην έχει αναπτυχθεί ιδιαίτερα στην παρούσα βιβλιογραφία.

Επιπλέον, κατά τη μοντελοποίηση χρονοσειρών, η πληροφορία των χρονοσφραγίδων (timestamps) είναι συνήθως ιδιαίτερα χρήσιμη, συμπεριλαμβανομένων των ημερολογιακών χρονοσφραγίδων (π.χ. δευτερόλεπτο, λεπτό, ώρα, εβδομάδα, μήνας και έτος) και των ειδικών χρονοσφραγίδων (π.χ. αργίες και γεγονότα). Αυτές οι χρονοσφραγίδες είναι αρκετά κατατοπιστικές σε πραγματικές εφαρμογές, αλλά δύσκολα αξιοποιούνται σε αρχιτεκτονικές Transformer. Για τον μετριασμό του προβλήματος, ο Informer [173] πρότεινε την κωδικοποίηση των χρονοσφραγίδων ως επιπρόσθετων κωδικοποιήσεων θέσης. Ένα παρόμοιο σύστημα κωδικοποίησης χρονοσφραγίδων χρησιμοποιήθηκε στον Autoformer [174] και FEDformer [175].

Αντίστοιχες ιδέες θα μπορούσαν να χρησιμοποιηθούν και στην περίπτωση της παρούσας έρευνας, έτσι ώστε να διαπιστωθεί η ικανότητα τους να προβλέπουν ειδικά γεγονότα.

Ωστόσο, αξίζει να σημειωθεί πως η παρούσα έρευνα επιχείρησε να παράγει αξιόπιστες προβλέψεις χωρίς την χρήση “εξωτερικών” βοηθημάτων χρησιμοποιώντας μόνο αρχιτεκτονικές που ακολουθούν τη λογική του Τρανσφορμερ και του μηχανισμού προσοχής, και σε ένα πολύ μεγάλο βαθμό το πέτυχε.

## Βιβλιογραφία

---

- [1] S. A. Ebrahim, J. Poshtan, S. M. Jamali και N. A. Ebrahim. *Quantitative and qualitative analysis of time-series classification using deep learning*. *IEEE Access*, 8:90202–90215, 2020.
- [2] C. K. Lee, H. J. Song και J. W. Mjelde. *The forecasting of International Expo tourism using quantitative and qualitative techniques*. *Tourism Management*, 29(6):1084–1098, 2008.
- [3] R. J. Hyndman και G. Athanasopoulos. *Forecasting: Principles and Practice*. OTexts, Australia, 2η έκδοση, 2018.
- [4] G. E. Box, S. C. Hillmer και G. C. Tiao. *Analysis and modeling of seasonal time series*, σελίδες 309–344. NBER, 1978.
- [5] B. B. Mandelbrot. *Statistical methodology for nonperiodic cycles: from the covariance to  $r/s$  analysis*. *Annals of Economic and Social Measurement*, 1:259–290, 1972.
- [6] S. Martin. *Characterization of financial time series*. Research note, University College London, London, 2011. PN/11/01.
- [7] C. Chen και L. M. Liu. *Forecasting time series with outliers*. *Journal of Forecasting*, 12(1):13–35, 1993.
- [8] ESRI. *Understanding outliers in time series analysis*, 2020.
- [9] T. Bex. *How to Detect Random Walk and White Noise in Time Series Forecasting. Find out if the target is worth forecasting*, 2021.
- [10] J. Du Preez και S. F. Witt. *Univariate versus multivariate time series forecasting: an application to international tourism demand*. *International Journal of Forecasting*, 19(3):435–451, 2003.
- [11] K. I. Stergiou, E. D. Christou και G. Petrakis. *Modelling and forecasting monthly fisheries catches: comparison of regression, univariate and multivariate time series methods*. *Fisheries Research*, 29(1):55–95, 1997.
- [12] H. D. Tran, N. Muttill και B. J. C. Perera. *Selection of significant input variables for time series forecasting*. *Environmental Modelling & Software*, 64:156–163, 2015.
- [13] M. A. Castán-Lascorz, P. Jiménez-Herrera, A. Troncoso και G. Asencio-Cortés. *A new hybrid method for predicting univariate and multivariate time series based on pattern forecasting*. *Information Sciences*, 586:611–627, 2022.

- [14] L. Kilian και M. P. Taylor. *Why is it so difficult to beat the random walk forecast of exchange rates?* *Journal of International Economics*, 60(1):85–107, 2003.
- [15] V. Cerqueira, L. Torgo και C. Soares. *Machine learning vs statistical methods for time series forecasting: Size matters.* *arXiv preprint arXiv:1909.13316*, 2019.
- [16] Φ. Ασημακόπουλος και Β. Πετρόπουλος. *Επιχειρησιακές προβλέψεις*. Εκδόσεις Συμμετρία, 2015.
- [17] E. Ostertagova και O. Ostertag. *Forecasting using simple exponential smoothing method.* *Acta Electrotechnica et Informatica*, 12(3):62, 2012.
- [18] T. B. Fomby. *Exponential smoothing models. Manual SAS/ETS Software: Time Series Forecasting System. Version 6.* 2008.
- [19] C. C. Holt. *Forecasting Trends and Seasonal by Exponentially Weighted Averages.* Pittsburgh, USA, 1957. Δημοσιεύθηκε στο *Journal of Forecasting* 2004, 20, 5–13.
- [20] J. D. Bermúdez, A. Corberán-Vallet και E. Vercher. *Forecasting time series with missing data using Holt’s model.* *Journal of Statistical Planning and Inference*, 139(8):2791–2799, 2009.
- [21] B. Billah, M. L. King, R. D. Snyder και A. B. Koehler. *Exponential smoothing model selection for forecasting.* *International Journal of Forecasting*, 22(2):239–247, 2006.
- [22] E. S. Gardner Jr. *Exponential smoothing: The state of the art — Part II.* *International Journal of Forecasting*, 22:637–666, 2006.
- [23] E. S. Gardner Jr. και E. McKenzie. *Forecasting trends in time series.* *Management Science*, 31:1237–1246, 1985.
- [24] E. S. Gardner Jr. και E. McKenzie. *Model identification in exponential smoothing.* *Journal of the Operational Research Society*, 39:863–867, 1988.
- [25] E. McKenzie και E. S. Gardner Jr. *Damped trend exponential smoothing: a modelling viewpoint.* *International Journal of Forecasting*, 26(4):661–665, 2010.
- [26] S. Makridakis και M. Hibon. *The M3-Competition: Results, conclusions and implications.* *International Journal of Forecasting*, 16:451–476, 2000.
- [27] C. Carl Pegels. *Exponential forecasting: some new variations.* *Management Science*, σελίδες 311–315, 1969.
- [28] P. J. Brockwell και R. A. Davis. *Introduction to Time Series and Forecasting.* Springer, 2016.
- [29] P. R. Winters. *Forecasting sales by exponentially weighted moving averages.* *Management Science*, 6(3):324–342, 1960.
- [30] C. Chatfield. *The Holt-winters forecasting procedure.* *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 27(3):264–279, 1978.



- [31] G. E. P. Box και G. M. Jenkins. *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1970.
- [32] R. Nau. *The mathematical structure of arima models*, 2014. Δυκε Υνιερσιτη Ονλινε Αρτιζλε, 1(1), 1-8.
- [33] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti και M. Ciccozzi. *Application of the ARIMA model on the COVID-2019 epidemic dataset*. *Data in Brief*, 29:105340, 2020.
- [34] M. R. Abonazel και A. I. Abd-Elftah. *Forecasting Egyptian GDP using ARIMA models*. *Reports on Economics and Finance*, 5(1):35–47, 2019.
- [35] A. A. Ariyo, A. O. Adewumi και C. K. Ayo. *Stock price prediction using the ARIMA model*. *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*, σελίδες 106–112. IEEE, 2014.
- [36] V. Assimakopoulos και K. Nikolopoulos. *The theta model: a decomposition approach to forecasting*. *International Journal of Forecasting*, 16(4):521–530, 2000.
- [37] K. I. Nikolopoulos και D. D. Thomakos. *Forecasting with the Theta Method: Theory and Applications*. John Wiley & Sons, 2019.
- [38] E. F. Morales και H. J. Escalante. *A brief introduction to supervised, unsupervised, and reinforcement learning*, σελίδες 111–129. Academic Press, 2022.
- [39] P. Cunningham, M. Cord και S. J. Delany. *Supervised learning*, σελίδες 21–49. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [40] Z. Li, D. Meunier, M. Mollenhauer και A. Gretton. *Optimal rates for regularized conditional mean embedding learning*. *Advances in Neural Information Processing Systems*, τόμος 35, σελίδες 4433–4445, 2022.
- [41] R. Caruana και A. Niculescu-Mizil. *An empirical comparison of supervised learning algorithms*. *Proceedings of the 23rd International Conference on Machine Learning*, σελίδες 161–168, 2006.
- [42] R. Saravanan και P. Sujatha. *A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification*. *2018 Second International Conference on Intelligent Computing and Control Systems (I-CICCS)*, σελίδες 945–949. IEEE, 2018.
- [43] T. Hastie, R. Tibshirani και J. Friedman. *Overview of supervised learning*, σελίδες 9–41. 2009.
- [44] D. H. Wolpert. *The supervised learning no-free-lunch theorems*, σελίδες 25–42. 2002.
- [45] Z. Ghahramani. *Unsupervised learning*, σελίδες 72–112. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

- [46] A. Maćkiewicz και W. Ratajczak. *Principal components analysis (PCA)*. *Computers & Geosciences*, 19(3):303–342, 1993.
- [47] T. Hastie, R. Tibshirani και J. Friedman. *Unsupervised learning*, σελίδες 485–585. 2009.
- [48] K. Arulkumaran, M. P. Deisenroth, M. Brundage και A. A. Bharath. *Deep reinforcement learning: A brief survey*. *IEEE Signal Processing Magazine*, 34(6):26–38, 2017.
- [49] M. Van Otterlo και M. Wiering. *Reinforcement learning and markov decision processes*, σελίδες 3–42. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [50] F. Garcia και E. Rachelson. *Markov decision processes*, σελίδες 1–38. 2013.
- [51] Y. Li. *Reinforcement learning applications*, 2019. arXiv preprint arXiv:1908.06973.
- [52] A. Patle και D. S. Chouhan. *SVM kernel functions for classification*. *2013 International Conference on Advances in Technology and Engineering (ICATE)*, σελίδες 1–9. IEEE, 2013.
- [53] M. Awad και R. Khanna. *Support vector regression*, σελίδες 67–80. 2015.
- [54] K. R. Müller, A. J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen και V. Vapnik. *Predicting time series with support vector machines*. *International Conference on Artificial Neural Networks*, σελίδες 999–1004, Berlin, Heidelberg, 1997. Springer Berlin Heidelberg.
- [55] C. J. Lu, T. S. Lee και C. C. Chiu. *Financial time series forecasting using independent component analysis and support vector regression*. *Decision Support Systems*, 47(2):115–125, 2009.
- [56] A. Altan και S. Karasu. *The effect of kernel values in support vector machine to forecasting performance of financial time series*. *The Journal of Cognitive Systems*, 4(1):17–21, 2019.
- [57] Z. Ramedani, M. Omid, A. Keyhani, S. Shamshirband και B. Khoshnevisan. *Potential of radial basis function based support vector regression for global solar radiation prediction*. *Renewable and Sustainable Energy Reviews*, 39:1005–1011, 2014.
- [58] L. E. Raileanu και K. Stoffel. *Theoretical comparison between the gini index and information gain criteria*. *Annals of Mathematics and Artificial Intelligence*, 41:77–93, 2004.
- [59] Priyanka και D. Kumar. *Decision tree classifier: a detailed survey*. *International Journal of Information and Decision Sciences*, 12(3):246–269, 2020.
- [60] J. Mingers. *An empirical comparison of pruning methods for decision tree induction*. *Machine Learning*, 4:227–243, 1989.

- [61] M. H. Roy και D. Larocque. *Robustness of random forests for regression*. *Journal of Nonparametric Statistics*, 24(4):993–1006, 2012.
- [62] V. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo και M. J. Chica-Rivas. *Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines*. *Ore Geology Reviews*, 71:804–818, 2015.
- [63] E. Spiliotis. *Decision trees for time-series forecasting*. *Foresight*, 64:30–44, 2022.
- [64] E. H. A. Rady, H. Fawzy και A. M. A. Fattah. *Time series forecasting using tree based methods*. *Journal of Statistical Applications and Probability*, 10(1):229–244, 2021.
- [65] R. Hecht-Nielsen. *Neurocomputing: picking the human brain*. *IEEE Spectrum*, 25(3):36–41, 1988.
- [66] F. Rosenblatt. *The perceptron: a probabilistic model for information storage and organization in the brain*. *Psychological Review*, 65(6):386, 1958.
- [67] S. Sharma, S. Sharma και A. Athaiya. *Activation functions in neural networks*. *Towards Data Science*, 6(12):310–316, 2017.
- [68] A. A. Waoo και B. K. Soni. *Performance analysis of sigmoid and relu activation functions in deep neural network*, σελίδες 39–52. Springer Singapore, 2021.
- [69] K. Eckle και J. Schmidt-Hieber. *A comparison of deep networks with ReLU activation function and linear spline-type methods*. *Neural Networks*, 110:232–242, 2019.
- [70] B. Karlik και A. V. Olgac. *Performance analysis of various activation functions in generalized MLP architectures of neural networks*. *International Journal of Artificial Intelligence and Expert Systems*, 1(4):111–122, 2011.
- [71] A. K. Dubey και V. Jain. *Comparative study of convolution neural network’s ReLU and leaky-ReLU activation functions*, σελίδες 873–880. Springer Singapore, 2019.
- [72] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu και F. E. Alsaadi. *A survey of deep neural network architectures and their applications*. *Neurocomputing*, 234:11–26, 2017.
- [73] R. Rojas. *The backpropagation algorithm*, σελίδες 149–182. 1996.
- [74] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao και J. Han. *On the variance of the adaptive learning rate and beyond*, 2019. arXiv preprint arXiv:1908.03265.
- [75] K. Fukumizu. *Effect of batch learning in multilayer neural networks*. *Gen*, 1(4):1E–03, 1998.
- [76] J. J. M. Moreno. *Artificial neural networks applied to forecasting time series*. *Psicothema*, 23(2):322–329, 2011.

- [77] N. K. Ahmed, A. F. Atiya, N. E. Gayar και H. El-Shishiny. *An empirical comparison of machine learning models for time series forecasting*. *Econometric Reviews*, 29(5):594–621, 2010.
- [78] H. Salehinejad, S. Sankar, J. Barfett, E. Colak και S. Valaee. *Recent advances in recurrent neural networks*, 2017. αρΧι προπριντ αρΧι:1801.01078.
- [79] P. J. Werbos. *Backpropagation through time: what it does and how to do it*. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.
- [80] J. Guo. *Backpropagation through time*, 2013. Unpublished manuscript, Harbin Institute of Technology, 40, 1-6.
- [81] Y. Yu, X. Si, C. Hu και J. Zhang. *A review of recurrent neural networks: LSTM cells and network architectures*. *Neural Computation*, 31(7):1235–1270, 2019.
- [82] K. Smagulova και A. P. James. *A survey on LSTM memristive neural network architectures and applications*. *The European Physical Journal Special Topics*, 228(10):2313–2324, 2019.
- [83] A. Sherstinsky. *Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network*. *Physica D: Nonlinear Phenomena*, 404:132306, 2020.
- [84] G. Van Houdt, C. Mosquera και G. Nápoles. *A review on the long short-term memory model*. *Artificial Intelligence Review*, 53(8):5929–5955, 2020.
- [85] B. Lindemann, T. Müller, H. Vietz, N. Jazdi και M. Weyrich. *A survey on long short-term memory networks for time series prediction*. *Procedia CIRP*, 99:650–655, 2021.
- [86] K. Benidis, S. S. Rangapuram, V. Flunkert, Y. Wang, D. Maddix, C. Turkmen και T. Januschowski. *Deep learning for time series forecasting: Tutorial and literature survey*. *ACM Computing Surveys*, 55(6):1–36, 2022.
- [87] Z. Chen, M. Ma, T. Li, H. Wang και C. Li. *Long sequence time-series forecasting with deep learning: A survey*. *Information Fusion*, 97:101819, 2023.
- [88] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma και L. Farhan. *Review of deep learning: concepts, CNN architectures, challenges, applications, future directions*. *Journal of Big Data*, 8:1–74, 2021.
- [89] Z. Li, F. Liu, W. Yang, S. Peng και J. Zhou. *A survey of convolutional neural networks: analysis, applications, and prospects*. *IEEE Transactions on Neural Networks and Learning Systems*, 33(12):6999–7019, 2021.
- [90] N. Akhtar και U. Ragavendran. *Interpretation of intelligence in CNN-pooling processes: a methodological survey*. *Neural Computing and Applications*, 32(3):879–898, 2020.

- [91] A. Zafar, M. Aamir, N. Mohd Nawi, A. Arshad, S. Riaz, A. Alruban και S. Almotairi. *A comparison of pooling methods for convolutional neural networks*. *Applied Sciences*, 12(17):8643, 2022.
- [92] G. Habib και S. Qureshi. *Optimization and acceleration of convolutional neural networks: A survey*. *Journal of King Saud University-Computer and Information Sciences*, 34(7):4244–4268, 2022.
- [93] H. Gholamalinezhad και H. Khosravi. *Pooling methods in deep neural networks, a review*, 2020. arXiv preprint arXiv:2009.07485.
- [94] H. Mizutani. *The back propagation method for CNN*. *Proceedings of IEEE International Symposium on Circuits and Systems-ISCAS'94*, τόμος 6, σελίδες 463–466. IEEE, 1994.
- [95] S. Mehtab και J. Sen. *Analysis and forecasting of financial time series using CNN and LSTM-based deep learning models*, σελίδες 405–423. Springer Singapore, 2022.
- [96] I. Koprinska, D. Wu και Z. Wang. *Convolutional neural networks for energy time series forecasting*. *2018 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8. IEEE, 2018.
- [97] A. Dingli και K. S. Fournier. *Financial time series forecasting-a deep learning approach*. *International Journal of Machine Learning and Computing*, 7(5):118–122, 2017.
- [98] X. Jin, X. Yu, X. Wang, Y. Bai, T. Su και J. Kong. *Prediction for Time Series with CNN and LSTM*, σελίδες 631–641. Springer Singapore, 2020.
- [99] S. S. Aung. *Electric power is the main driving force for industrialization*. <http://www.globalnewlightofmyanmar.com/electric-power-is-the-main-driving-force-for-industrialization/>, 2015.
- [100] A. Dedinec, S. Filiposka, A. Dedinec και L. Kocarev. *Deep belief network based electricity load forecasting: an analysis of Macedonian case*. *Energy*, 115:1688–1700, 2016.
- [101] A. Hussain, M. Rahman και J. A. Memon. *Forecasting electricity consumption in Pakistan: the way forward*. *Energy Policy*, 90:73–80, 2016.
- [102] S. Jevgenijs, Joeri deW, A. Kochnakyan και F. Vivien. *Forecasting electricity demand: an aid for practitioners*. <http://www.worldbank.org/energy/livewire>, 2017.
- [103] IEA. *World Energy Outlook 2015*. International Energy Agency, Paris, France, 2015.
- [104] M. U. Zaman, A. Islam και N. Sultana. *Short term load forecasting based on internet of things (IoT)*. *Διδακτορική Διατριβή*, BRAC University, Dhaka, 2018.

- [105] C. H. J. Kumar και M. Veerakumari. *Load forecasting of Andhra Pradesh grid using PSO, DE algorithms. International Journal of Advanced Research in Computer Engineering and Technology*, 1:179–184, 2012.
- [106] Q. Wang, Y. Wang και L. Zhang. *Research on Post-Evaluation Model and Method of Electric Power Demand Forecasting. Chinese Control and Decision Conference (CCDC)*, China, 2008.
- [107] G. Nalcaci, A. Özmen και G. W. Weber. *Long-term Load Forecasting: Models Based on MARS, ANN and LR methods. Central European Journal of Operations Research*, 27(2019):1033–1049, 2018.
- [108] E. Almeshaiei και H. Soltan. *A Methodology for Electric Power Load Forecasting. Alexandria Engineering Journal*, 50:137–144, 2011.
- [109] K. B. Sahay και M. M. Tripathi. *Day ahead hourly load forecast of PJM electricity market and ISO New England market by using artificial neural network. Proceedings of the 2013 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, σελίδες 1–5, Bangalore, India, 2013.
- [110] I. P. Panapakidis. *Clustering based day-ahead and hour-ahead bus load forecasting models. International Journal of Electrical Power & Energy Systems*, 80:171–178, 2016.
- [111] J. Zhang. *Research on Power Load Forecasting Based on the Improved Elman Neural Network. The Italian Association of Chemical Engineering (AIDIC)*, 51:589–594, 2016.
- [112] Y. Lin, H. Luo, D. Wang, H. Guo και K. Zhu. *An Ensemble Model Based on Machine Learning Methods and Data Preprocessing for Short-Term Electric Load Forecasting. Energies*, 10:1186, 2017.
- [113] X. Zhang, J. Wang και K. Zhang. *Short-Term Electric Load Forecasting Based on Singular Spectrum Analysis and Support Vector Machine Optimized by Cuckoo Search Algorithm. Electric Power Systems Research*, 146:270–285, 2017.
- [114] S. Khatoun, Ibraheem, A. K. Sing και Priti. *Effects of various factors on electric load forecasting: An overview. IEEE Power India International Conference (PIICON)*, σελίδες 1–5, 2014.
- [115] M. Y. Khamaira, A. S. Krzma και A. M. Alnass. *Long Term Peak Load Forecasting for the Libyan Network. Conference for Engineering Sciences and Technology (CEST)*, τόμος 1, σελίδες 185–193. AIJR Publisher, 2018.
- [116] Ü. B. Filik, Ö. N. Gerek και M. Kurban. *A Novel Modeling Approach for Hourly Forecasting of Long-Term Electric Energy Demand. Energy Conversion and Management*, 52:199–211, 2011.

- [117] R. Gordillo-Orquera, L. M. Lopez-Ramos, S. Muñoz-Romero, P. Iglesias-Casarrubios, D. Arcos-Avilés, A. G. Marques και J. L. Rojo-Álvarez. *Analyzing and Forecasting Electrical Load Consumption in Healthcare Buildings*. *Energies*, 11:493, 2018.
- [118] L. Friedrich και A. Afshari. *Short-Term Forecasting of the Abu Dhabi Electricity Load Using Multiple Weather Variables*. *7th International Conference on Applied Energy (ICAE)*, 2015.
- [119] N. Abu-Shikhah και F. Elkarmi. *Medium-Term Electric Load Forecasting Using Singular Value Decomposition*. *Energy Conversion and Management*, 36(7):4259–4271, 2011.
- [120] R. Wang, J. Wang και Y. Xu. *A Novel Combined Model Based on Hybrid Optimization Algorithm for Electrical Load Forecasting*. *Applied Soft Computing Journal*, 82:105548, 2019.
- [121] G. Gross και F. D. Galiana. *Short Term Load Forecasting*. *Proceedings of the IEEE*, 75:1558–1573, 1987.
- [122] Q. Ding. *Long-Term Load Forecast Using Decision Tree Method*. *Power Systems Conference and Exposition, PSCE 06, IEEE PES*, τόμος 1, σελίδες 1541–1543, 2006.
- [123] G. T. Heineman, D. A. Nordman και E. C. Plant. *The Relationship Between Summer Weather and Summer Loads - A Regression Analysis*. *IEEE Transaction Power Apparatus System*, ΠΑΣ-85(11):1144–1154, 1966.
- [124] J. H. Park, Y. M. Park και K. Y. Lee. *Composite Modeling for Adaptive Short-Term Load Forecasting*. *IEEE Transaction Power System*, 6(2):450–457, 1991.
- [125] G. A. N. Mbamalu και M. E. El-Hawary. *Load forecasting via suboptimal seasonal autoregressive models and iteratively reweighted least squares estimation*. *IEEE Transactions on Power Systems*, 8(1):343–348, 1993.
- [126] I. Moghram και S. Rahman. *Analysis And Evaluation Of Five Short-Term Load Forecasting Techniques*. *IEEE Transaction on Power System*, 4:1484–1491, 1989.
- [127] E. H. Barakat, M. A. Qayyum, M. N. Hamed και S. A. Al-Rashed. *Short-Term Peak Demand Forecasting in Fast Developing Utility with Inherent Dynamic Load Characteristics*. *IEEE Transactions on Power Systems*, 5:813–824, 1990.
- [128] A. D. Papalexopoulos και T. C. Hesterberg. *A Regression-based Approach to Short-Term Load Forecasting*. *IEEE Transmission on Power System*, 5:1214–1221, 1990.
- [129] S. Varadan και E. B. Makram. *Harmonic Load Identification and Determination of Load Composition Using a Least Squares Method*. *Electric Power Systems Research*, 37:203–208, 1996.
- [130] O. Hyde και P. F. Hodnett. *An Adaptable Automated Procedure for Short-Term Electricity Load Forecasting*. *IEEE Transactions on Power Systems*, 12:84–93, 1997.

- [131] R. P. Broadwater, A. Sargent, A. Yarali, H. E. Shaalan και J. Nazarko. *Estimating Substation Peaks from Research Data. IEEE Transaction on Power Delivery*, 12:451–456, 1997.
- [132] H. K. Alfares και M. Nazeeruddin. *Regression-Based Methodology for Daily Peak Load Forecasting. Proceedings of the 2nd International Conference on Operations and Quantitative Management*, σελίδες 468–471, Ahmedabad, India, 1999.
- [133] A. A. El-Keib, X. Ma και H. Ma. *Advancement of Statistical Based Modeling for Short-Term Load Forecasting. Electric Power Systems Research*, 35:51–58, 1995.
- [134] D. G. Infield και D. C. Hill. *Optimal Smoothing for Trend Removal in Short term Electricity Demand Forecasting. IEEE Transaction on Power System*, 13:1115–1120, 1998.
- [135] J. F. Chen, W. M. Wang και C. M. Huang. *Analysis of an Adaptive Time-Series Autoregressive Moving-Average (ARMA) Model for Short-Term Load Forecasting. Electric Power Systems Research*, 34:187–196, 1995.
- [136] G. N. Shilpa και G. S. Sheshadri. *Short-term load forecasting using ARIMA model for Karnataka state electrical load. International Journal of Engineering Research and Development*, 13(7):75–79, 2017.
- [137] G. Juberias, R. Yunta, J. Garcia Morino και C. Mendivil. *A New ARIMA Model for Hourly Load Forecasting. IEEE Transmission and Distribution Conference Proceedings*, τόμος 1, σελίδες 314–319, 1999.
- [138] G. Li, C. T. Cheng, J. Y. Lin και Y. Zeng. *Short-term load forecasting using support vector machine with SCE-UA algorithm. Third International Conference on Natural Computation (ICNC 2007)*, τόμος 1, σελίδες 290–294. IEEE, 2007.
- [139] M. Chauhan, S. Gupta και M. Sandhu. *Short-term electric load forecasting using support vector machines. ECS Transactions*, 107(1):9731, 2022.
- [140] B. E. Türkay και D. Demren. *Electrical load forecasting using support vector machines: A case study. International Review of Electrical Engineering*, 6(5), 2011.
- [141] S. Akhtar, S. Shahzad, A. Zaheer, H. S. Ullah, H. Kilic, R. Gono και Z. Leonowicz. *Short-term load forecasting models: A review of challenges, progress, and the road ahead. Energies*, 16(10):4060, 2023.
- [142] A. Kumar Srivastava, D. Singh και A. S. Pandey. *Short term load forecasting using regression trees: Random forest, bagging and m5p*, 2020.
- [143] A. R. Íyàndá, O. A. Odéjobí και A. O. Kómoláfé. *Short-term electric load forecasting in uncertain domain: A fuzzy decision tree approach. INFOCOMP Journal of Computer Science*, 10(4):29–39, 2011.



- [144] W. Guo, L. Che, M. Shahidehpour και X. Wan. *Machine-Learning based methods in short-term load forecasting*. *The Electricity Journal*, 34(1):106884, 2021.
- [145] G. Dudek. *A comprehensive study of random forest for short-term load forecasting*. *Energies*, 15(20):7547, 2022.
- [146] G. Dudek. *Short-term load forecasting using random forests*. *Intelligent Systems' 2014: Proceedings of the 7th IEEE International Conference Intelligent Systems I-S'2014, September 24-26, 2014, Warsaw, Poland, Volume 2: Tools, Architectures, Systems, Applications*, σελίδες 821–828. Springer International Publishing, 2015.
- [147] M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati και H. Abu-Rub. *A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting*. *Energy*, 214:118874, 2021.
- [148] H. Shi, M. Xu, Q. Ma, C. Zhang, R. Li και F. Li. *A whole system assessment of novel deep learning approach on short-term load forecasting*. *Energy Procedia*, 142:2791–2796, 2017.
- [149] A. Narayan και K. W. Hipel. *Long short term memory networks for short-term electric load forecasting*. *2017 IEEE International Conference on Systems, Man, and Cybernetics. SMC, IEEE*, σελίδες 2573–2578, 2017.
- [150] S. Kumar, L. Hussain, S. Banarjee και M. Reza. *Energy load forecasting using deep learning approach-LSTM and GRU in spark cluster*. *2018 Fifth International Conference on Emerging Applications of Information Technology. EAIT, IEEE*, σελίδες 1–4, 2018.
- [151] M. Cai, M. Pipattanasomporn και S. Rahman. *Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques*. *Applied Energy*, 236:1078–1088, 2019.
- [152] X. Dong, L. Qian και L. Huang. *Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach*. *2017 IEEE International Conference on Big Data and Smart Computing. BigComp, IEEE*, σελίδες 119–125, 2017.
- [153] K. Yan, W. Li, Z. Ji, M. Qi και Y. Du. *A Hybrid LSTM Neural Network for Energy Consumption Forecasting of Individual Households*. *IEEE Access*, 7:157633–157642, 2019.
- [154] P. Li, K. Zhou, X. Lu και S. Yang. *A hybrid deep learning model for short-term PV power forecasting*. *Applied Energy*, 259:114216, 2019.
- [155] X. Wang, Z. Yao και M. Papaefthymiou. *A real-time electrical load forecasting and unsupervised anomaly detection framework*. *Applied Energy*, 330:120279, 2023.
- [156] W. He. *Load forecasting via deep neural networks*. *Procedia Computer Science*, 122:308–314, 2017.

- [157] J. F. Torres, D. Hadjout, A. Sebaa, F. Martínez-Álvarez και A. Troncoso. *Deep Learning for Time Series Forecasting: A Survey*. *Data*, 9:3–21, 2021.
- [158] T. Ouyang, Y. He, H. Li, Z. Sun και S. Baek. *Modeling and Forecasting Short-Term Power Load With Copula Model and Deep Belief Network*. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 3:127–136, 2019.
- [159] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez και I. Polosukhin. *Attention is all you need*. *Advances in Neural Information Processing Systems*, 30, 2017.
- [160] S. Lei, W. Yi, C. Ying και R. Ruibin. *Review of attention mechanism in natural language processing*. *Data Analysis and Knowledge Discovery*, 4(5):1–14, 2020.
- [161] S. Chaudhari, V. Mithal, G. Polatkan και R. Ramanath. *An attentive survey of attention models*. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(5):1–32, 2021.
- [162] D. Hu. *An introductory survey on attention mechanisms in NLP problems*, σελίδες 432–448. Springer International Publishing, 2020.
- [163] Y. Hu και F. Xiao. *Network self attention for forecasting time series*. *Applied Soft Computing*, 124:109092, 2022.
- [164] J. Zheng, S. Ramasinghe και S. Lucey. *Rethinking positional encoding*, 2021. arXiv preprint arXiv:2107.02561.
- [165] G. Wang, Y. Lu, L. Cui, T. Lv, D. Florencio και C. Zhang. *A simple yet effective learnable positional encoding method for improving document transformer model*. *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2022*, σελίδες 453–463, 2022.
- [166] Z. Dai, Z. Yang, Y. Yang, W. W. Cohen, J. Carbonell, Q. V. Le και R. Salakhutdinov. *Transformer-XL: Language modeling with longer-term dependency*, 2018.
- [167] L. Xue, N. Constant, A. Roberts, M. Kale, R. Al-Rfou, A. Siddhant και C. Raffel. *mT5: A massively multilingual pre-trained text-to-text transformer*, 2020. arXiv preprint arXiv:2010.11934.
- [168] Shreshth Tuli, Giuliano Casale και Nicholas R. Jennings. *TranAD: Deep transformer networks for anomaly detection in multivariate time series data*. *VLDB*, 2022.
- [169] Xixuan Wang, Dechang Pi και Xiangyan Zhang. *Variational transformer-based anomaly detection approach for multivariate time series*. *Measurement*, σελίδα 110791, 2022.
- [170] Hongwei Zhang και Yuanqing Xia. *Unsupervised anomaly detection in multivariate time series through transformer-based variational autoencoder*. *CCDC*, 2021.

- 
- [171] Diederik P. Kingma και Max Welling. *Auto-encoding variational bayes*. *ICLR*, 2014.
- [172] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley και Sherjil Ozair. *Generative adversarial nets*. *NeurIPS*, 2014.
- [173] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong και Wancai Zhang. *Informer: Beyond efficient transformer for long sequence timeseries forecasting*. *AAAI*, 2021.
- [174] Haixu Wu, Jiehui Xu, Jianmin Wang και Mingsheng Long. *Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting*. *NeurIPS*, 2021.
- [175] Tian Zhou, Ziqing Ma, Qingsong Wen, Xue Wang, Liang Sun και Rong Jin. *FEDformer: Frequency enhanced decomposed transformer for long-term series forecasting*. *ICML*, 2022.



## Συντομογραφίες - Αρχιτικόλεξα - Ακρωνύμια

---

AI	Artificial Intelligence
ARIMA	Auto Regressive Integrated Moving Average
CNN	Convolutional Neural Network
CPU	Central Processing Unit
DL	Deep Learning
GPU	Graphic Processing Unit
LSTM	Long Short Term Memory
ML	Machine Learning
MLP	Multi-Layer Perceptron
SES	Simple Exponential Smoothing