



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

ΔΠΜΣ Επιστήμη Δεδομένων & Μηχανική Μάθηση

**Μελέτη μοντέλων μηχανικής μάθησης για την
πρόβλεψη προθεσμιακών συμβολαίων σε
αγροτικά προϊόντα με τηλεσκοπικά δεδομένα.**

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΒΑΣΙΛΕΙΟΣ Γ. ΜΠΡΙΛΛΑΚΗΣ

Επιβλέπων : Κωνσταντίνος Καράντζαλος

Καθηγητής Ε.Μ.Π

Αθήνα, Νοέμβριος 2023



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

ΔΠΜΣ Επιστήμη Δεδομένων & Μηχανική Μάθηση

Μελέτη μοντέλων μηχανικής μάθησης για την πρόβλεψη προθεσμιακών συμβολαίων σε αγροτικά προϊόντα με τηλεσκοπικά δεδομένα.

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΒΑΣΙΛΕΙΟΣ Γ. ΜΠΡΙΛΛΑΚΗΣ

Επιβλέπων : Κωνσταντίνος Καράντζαλος

Καθηγητής Ε.Μ.Π

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 3^η Νοεμβρίου 2023.

.....
Κωνσταντίνος Καράντζαλος
Καθηγητής Ε.Μ.Π.

.....
Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

.....
Αθανάσιος Βουλόδημος
Επ. Καθηγητής Ε.Μ.Π.

Αθήνα, Νοέμβριος 2023

Βασίλειος Γ. Μπριλλάκης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Copyright © Βασίλειος Γ. Μπριλλάκης, 2023.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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 την χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται στον συγγραφέα.

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

Περίληψη

Η γεωργία ακριβείας έχει ξεπεράσει τις παραδοσιακές μεθόδους καλλιέργειας, προσφέροντας μια προοδευτική προσέγγιση που βασίζεται στην τεχνολογία και στις πληροφορίες που βασίζονται σε δεδομένα. Στη σφαίρα ενός απρόβλεπτου κλίματος σε συνδυασμό με τις κλιμακούμενες παγκόσμιες απαιτήσεις σε τρόφιμα, η σημασία της ενίσχυσης της γεωργικής παραγωγικότητας δεν μπορεί να υπερτιμηθεί. Μια ιδιαίτερα υποσχόμενη εφαρμογή στο πλαίσιο της γεωργίας ακριβείας είναι η πρόβλεψη των τιμών των βασικών προϊόντων. Στην παρούσα διπλωματική μελετώνται τεχνικές μηχανικής μάθησης όπου σε συνδυασμό με δεδομένα από γεωργία ακριβείας μπορεί να οδηγήσουν στην ανάπτυξη προηγμένων μοντέλων ικανών να προβλέπουν τις τάσεις στις τιμές αγροτικών προϊόντων. Τέτοια εργαλεία πρόβλεψης μπορούν να αποδειχθούν σημαντικά για τους παραγωγούς, προσφέροντάς τους πρόγνωση των πιθανών διακυμάνσεων των τιμών. Με αυτές τις προγνωστικές γνώσεις, οι παραγωγοί είναι σε θέση να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με το ποιες καλλιέργειες θα φυτέψουν, πότε θα πουλήσουν τα προϊόντα τους και πώς θα χαράξουν στρατηγική για την παρουσία τους στην αγορά. Αυτό όχι μόνο ενισχύει την κερδοφορία τους, αλλά και την ανθεικτικότητα έναντι των διακυμάνσεων της αγοράς. Στην ουσία, η ενσωμάτωση της μηχανικής μάθησης για την πρόβλεψη των τιμών στη γεωργία ακριβείας μπορεί να λειτουργήσει ως καταλύτης για έναν πιο προληπτικό και οικονομικά βιώσιμο γεωργικό τομέα.

Λέξεις κλειδιά: Μηχανική Μάθηση, Βαθιά Μάθηση, Συνελικτικά Νευρωνικά Δίκτυα (ΣΝΔ), Αναδρομικά Νευρωνικά Δίκτυα (ΑΝΔ), Ακριβή Γεωργία., Τηλεσκοπικά Δεδομένα.

Abstract

Precision agriculture has transcended traditional farming methods, offering a forward-thinking approach that hinges on technology and data-driven insights. In the realm of an unpredictable climate coupled with escalating global food requirements, the importance of enhancing agricultural productivity cannot be overstated. One particularly promising application within precision agriculture is commodity price forecasting. As this thesis elucidates, the fusion of machine learning techniques with precision agriculture can lead to the development of advanced models capable of predicting market trends. Such forecasting tools can prove invaluable for farmers, offering them foresight into potential price fluctuations. With these predictive insights, farmers are better positioned to make informed decisions about which crops to plant, when to sell their produce, and how to strategize their market presence. This not only bolsters their profitability but also fosters resilience against market volatilities. In essence, the integration of machine learning for price forecasting in precision agriculture can serve as a catalyst for a more proactive and economically viable farming sector.

Key words: Machine Learning, Deep Learning, Convolutional Neural Networks (CNNs), Recursive Neural Networks (RNNs), Precise Agriculture, Remote Sensing Data.

Ευχαριστίες

Αρχικά, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Κωνσταντίνο Καράντζαλο, τους καθηγητές μου Στέφανο Κόλλια και Γιώργο Στάμου που μου έδωσαν τα ερεθίσματα για να ξεκινήσω να ασχολούμαι με τον συγκεκριμένο τομέα και με τις συμβουλές των οποίων κατάφερα να φοιτήσω στο ΔΠΜΣ.

Επίσης, θα ήθελα να ευχαριστήσω τον φίλο μου Στέλιο Ι., ο οποίος συνέβαλε στην έμπνευση για το θέμα αυτής της διπλωματικής εργασίας.

Ακόμα, να ευχαριστήσω τους φίλους και συμφοιτητές Μελκόν Χ. και Δημήτρη Σ. για τη βοήθεια τους και τη συνεργασία μας καθόλη τη διάρκεια φοίτησης μας στο ΔΠΜΣ.

Θα ήθελα να ευχαριστήσω επίσης τους φίλους μου Αργύρη Μ., Γιάννη Ξ., Γιώργο Λ., Γιώργο Τ, Σπύρο Μ., Αριστοτέλη Π., Αλέξανδρο Λ., Λεωνίδα Κ., Βασιλική Π., Βάλια Χ. για την συνεχή στήριξη τους κατά τη διάρκεια φοίτησης μου και εκπόνησης της διπλωματικής.

Τέλος, θα ήθελα να πω ένα μεγάλο ευχαριστώ στη μητέρα μου και στον πατέρα μου για την υλική και συναισθηματική στήριξη τους όλα τα χρόνια φοίτησης. Χωρίς τη βοήθειά τους, η ολοκλήρωση των σπουδών μου δεν θα ήταν εφικτή.

Περιεχόμενα

Ευχαριστίες	9
Κεφάλαιο 1	15
Εισαγωγή	15
1.3 Σκοπός Εργασίας.....	16
1.4 Σχετικές Εργασίες.....	16
1.5 Δομή Διπλωματικής.....	16
Κεφάλαιο 2.....	19
Θεωρητικό Υπόβαθρο	19
2.1 Εισαγωγή.....	19
2.2 Μηχανική Μάθηση	19
2.2.1 Είδη Μηχανικής Μάθησης	20
2.2.2 Επαγωγική Μάθηση	20
2.2.3 Συνάρτηση Κόστους.....	21
2.3 Βαθιά Μάθηση	27
2.3.1 Νευρωνικά Δίκτυα	28
2.3.2 Συναρτήσεις Ενεργοποίησης	30
2.3.3 Αλγόριθμοι Βελτιστοποίησης.....	31
2.3.4 Οπίσθια Διάδοση	34
2.3.4 Συνελκτικά Νευρωνικά Δίκτυα	38
2.3.5 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα	40
2.3.6 LSTM.....	41
2.3.17 Bidirectional LSTM.....	44
Κεφάλαιο 3.....	47
Δεδομένα	47
3.1 Εισαγωγή.....	47
3.2 Τιμές από προθεσμιακά συμβόλαια	47
3.2.1 Προθεσμιακά Συμβόλαια Καλαμποκιού ZCZ3.....	48
3.3 Τηλεσκοπικά Δεδομένα	49
Κεφάλαιο 4.....	53

Επεξεργασία Δεδομένων	53
4.1 Εισαγωγή.....	53
4.2 Καθαρισμός Δεδομένων.....	53
4.2 Μηχανική Χαρακτηριστικών.....	53
4.3 Επιλογή χαρακτηριστικών.....	56
4.4 Κανονικοποίηση Δεδομένων.....	57
Κεφάλαιο 5.....	59
Αρχιτεκτονικές Μοντέλων	59
5.1 Εισαγωγή.....	59
5.2 Ενσωμάτωση των Δεδομένων	59
5.3 Παρουσίαση Αρχιτεκτονικών	61
5.3.1 Linear Regression.....	61
5.3.2 XGBoost.....	61
5.3.3 Vanilla LSTM	62
5.3.4 Stacked LSTM.....	63
5.3.5 Bidirectional LSTM	64
5.3.6 CNN-LSTM	65
5.3.7 Dual Input Attention LSTM.....	66
Κεφάλαιο 6.....	71
Πειραματική Διαδικασία	71
6.1 Εισαγωγή.....	71
6.2 Πρόβλεψη Τιμής Επόμενης Ημέρας.....	71
6.2.1 Linear Regression.....	71
6.2.2 XGBoost.....	75
6.2.2 Vanilla LSTM	77
6.2.3 Stacked LSTM.....	79
6.2.4 Bidirectional LSTM	81
6.2.5 CNN-LSTM	83
6.2.6 Dual Input Attention LSTM.....	85
6.3 Πρόβλεψη Τιμών Επόμενων Ημερών.....	87
6.3.1 Linear Regression.....	87

6.3.1 XGBoost.....	88
6.3.3 Vanilla LSTM	89
6.3.4 Stacked LSTM.....	90
6.3.5 Bidirectional LSTM	91
6.3.6 Dual Input Attention LSTM.....	92
Κεφάλαιο 7.....	95
Συμπεράσματα και Μελλοντική Εργασία.....	95
Κατάλογος Σχημάτων	97
Κατάλογος Πινάκων	100
Βιβλιογραφία.....	102

Κεφάλαιο 1

Εισαγωγή

1.1 Ακριβή Γεωργία

Η γεωργία ακριβείας αντιπροσωπεύει μια επαναστατική αλλαγή στον γεωργικό τομέα, συνδυάζοντας τις παραδοσιακές μεθόδους καλλιέργειας με τεχνολογίες αιχμής για την επίτευξη υψηλών επιπέδων αποδοτικότητας και παραγωγικότητας. Στον πυρήνα αυτής της προσέγγισης βρίσκονται προηγμένα εργαλεία όπως τα παγκόσμια συστήματα εντοπισμού θέσης (GPS), μη επανδρωμένα αεροσκάφη εξοπλισμένα με κάμερες υψηλής ανάλυσης, δορυφορικές εικόνες και μια σειρά από συσκευές του Διαδικτύου των Πραγμάτων (IoT), συμπεριλαμβανομένων αισθητήρων εδάφους και οργάνων παρακολούθησης του κλίματος. Οι τεχνολογίες αυτές συνεργάζονται για να καταγράψουν μια πληθώρα δεδομένων σε πραγματικό χρόνο, που κυμαίνονται από τα επίπεδα θρεπτικών συστατικών του εδάφους έως την περιεκτικότητα σε υγρασία, ακόμη και τις συνθήκες μικροκλίματος. Αυτός ο πλούτος δεδομένων, στη συνέχεια αναλύεται και ερμηνεύεται, συχνά με τη βοήθεια εξελιγμένων αλγορίθμων, ώστε να παρέχονται στους παραγωγούς αξιοποιήσιμες πληροφορίες. Με αυτές τις γνώσεις, οι γεωργικές πρακτικές, όπως η φύτευση, το πότισμα, η λίπανση και η καταπολέμηση των παρασίτων, μπορούν να βελτιστοποιηθούν σε εξαιρετικά λεπτομερές επίπεδο. Το αποτέλεσμα είναι ένα αγροτικό οικοσύστημα που ελαχιστοποιεί τα απόβλητα και τις περιβαλλοντικές επιπτώσεις, ενώ μεγιστοποιεί την απόδοση και την αποδοτικότητα των πόρων. Εν τέλει, η γεωργία ακριβείας επιδιώκει να φέρει επανάσταση στην παραδοσιακή γεωργία, καθιστώντας την ως μια πιο καθοδηγούμενη από δεδομένα, αποτελεσματική και βιώσιμη προσπάθεια.

1.2 Προσφορά Ακριβής Γεωργίας

Τα πλεονεκτήματα που προσφέρει η γεωργία ακριβείας είναι τόσο άμεσα όσο και εκτεταμένα, επηρεάζοντας όχι μόνο το οικονομικό αποτέλεσμα για τους παραγωγούς, αλλά και συμβάλλοντας στην επίτευξη ευρύτερων στόχων βιωσιμότητας και περιβάλλοντος. Η αποδοτικότητα των πόρων είναι ένα από τα πιο συναρπαστικά οφέλη: με τη χρήση δεδομένων που καθοδηγούν την ακριβή εφαρμογή νερού, λιπασμάτων και φυτοφαρμάκων, οι αγρότες μπορούν να μειώσουν σημαντικά το λειτουργικό τους κόστος. Αυτή η προσέγγιση όχι μόνο ελαχιστοποιεί τη σπατάλη, αλλά μειώνει επίσης το περιβαλλοντικό τίμημα, μειώνοντας την απορροή και την υποβάθμιση του εδάφους. Επιπλέον, το μοντέλο της γεωργίας ακριβείας οδηγούμενο από τα δεδομένα επιτρέπει την αποτελεσματικότερη διαχείριση των καλλιεργειών. Σε μια εποχή όπου η κλιματική αλλαγή δημιουργεί ακανόνιστα καιρικά φαινόμενα και απρόβλεπτες προκλήσεις, η

προσαρμοστικότητα και η λήψη αποφάσεων σε πραγματικό χρόνο που διευκολύνει η γεωργία ακριβείας είναι ανεκτίμητης αξίας. Τέλος, τα δεδομένα που παράγονται μέσω αυτών των τεχνολογιών μπορούν να χρησιμεύσουν ως κρίσιμος πόρος για τη συνεχιζόμενη γεωργική έρευνα, την ανάπτυξη πολιτικής και τις στρατηγικές διαχείρισης της γης, θέτοντας τις βάσεις για ένα πιο βιώσιμο μέλλον στη γεωργία.

1.3 Σκοπός Εργασίας

Η παρούσα διπλωματική εργασία επικεντρώνεται κυρίως στη διερεύνηση των δυνατοτήτων των αλγορίθμων μηχανικής μάθησης για την πρόβλεψη των τιμών αγροτικών προϊόντων. Ενώ η γεωργία ακριβείας έχει ήδη σημειώσει σημαντικά βήματα στη βελτιστοποίηση των γεωργικών πρακτικών, η ικανότητα ακριβούς πρόβλεψης των τιμών της αγοράς μπορεί να δώσει τη δυνατότητα στους παραγωγούς να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με την επιλογή των καλλιεργειών και τις στρατηγικές πώλησης, μεγιστοποιώντας έτσι την κερδοφορία και παρέχοντας ασφάλεια και σταθερότητα. Η διπλωματική αυτή φιλοδοξεί να φέρει περαιτέρω επανάσταση στον γεωργικό τομέα, καθιστώντας τον πιο προσαρμοστικό, κερδοφόρο και βιώσιμο.

1.4 Σχετικές Εργασίες

Στο πλαίσιο της εργασίας μελετήθηκαν μέθοδοι που αφορούν σε βαθιά νευρωνικά δίκτυα και αρχιτεκτονικές LSTM[1], CNN-RNN [2], σε ανίχνευση και χρήση υποσυστημάτων προσοχής [2]–[4].

1.5 Δομή Διπλωματικής

- Στο **Κεφάλαιο 2** παρουσιάζεται η απαραίτητη θεωρία για την εξοικείωση του αναγνώστη με βασικές έννοιες της μηχανικής μάθησης καθώς επίσης αναλύονται μοντέλα βαθιάς μάθησης.
- Στο **Κεφάλαιο 3** περιγράφεται η διαδικασία που ακολούθηθηκε για την εξόρυξη (mining) των δεδομένων και η ανάλυση των δεδομένων.
- Στο **Κεφάλαιο 4** περιγράφεται η διαδικασία της μηχανικής των χαρακτηριστικών (feature engineering) καθώς και η διαδικασία επιλογής των χαρακτηριστικών (feature selection).
- Στο **Κεφάλαιο 5** παρουσιάζονται οι αρχιτεκτονικές των Τεχνητών Νευρωνικών Δικτύων (ΤΝΔ) που χρησιμοποιήθηκαν κατά την διάρκεια των υπολογισμών.
- Στο **Κεφάλαιο 6** αναλύεται αρχικά η πειραματική διαδικασία που ακολούθηθηκε για κάθε μοντέλο και έπειτα παρουσιάζονται τα αποτελέσματα.

- Στο **Κεφάλαιο 7** παρατίθενται τα βασικά συμπεράσματα της εργασίας καθώς προτείνονται ιδέες για μελλοντική εργασία.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

2.1 Εισαγωγή

Αρχικά, πραγματοποιείται η εισαγωγή στην μηχανική μάθηση, δηλαδή την ευρύτερη κατηγορία στην οποία ανήκουν τα Τεχνητά Νευρωνικά Δίκτυα (ΤΝΔ). Αναλυτικότερα, παρουσιάζονται οι κατηγορίες στις οποίες διακρίνονται οι αλγόριθμοι μηχανικής μάθησης, δίνοντας έμφαση στην κατηγορία αλγορίθμων μάθησης, μέσω των οποίων επιλύεται το πρόβλημα που πραγματεύεται η παρούσα διπλωματική εργασία. Έπειτα, γίνεται η μετάβαση στην βαθιά μάθηση που ανήκουν οι αρχιτεκτονικές νευρωνικών δικτύων. Συγκεκριμένα, αναλύεται η δομή και οι μονάδες των νευρωνικών δικτύων, αναφέρονται τεχνικές και οι μέθοδοι που χρησιμοποιούνται σε αυτά και τέλος παρουσιάζονται εξειδικευμένες κατηγορίες νευρωνικών δικτύων.

2.2 Μηχανική Μάθηση

Η μηχανική μάθηση είναι ο κλάδος της επιστήμης των υπολογιστών που ασχολείται με τη μελέτη και τη κατασκευή αλγορίθμων που δίνουν την δυνατότητα στον υπολογιστή να μαθαίνει από την εμπειρία. Το βασικό χαρακτηριστικό της μηχανικής μάθησης είναι ότι αναζητεί πρότυπα και σχέσεις στα δεδομένα με σκοπό να μοντελοποιήσει και να κάνει προβλέψεις πάνω σε αυτά. Σύμφωνα με τον πρώιμο ορισμό που έδωσε ο Arthur Samuel το 1959 [5] η μηχανική μάθηση είναι:

Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί.

Ένας πιο επίσημος ορισμός που δόθηκε από τον Tom M. Mitchell [6] είναι ο ακόλουθος:

Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P , αν η επίδοση του σε εργασίες μια κλάσης T , όπως αποτιμάται από το μέτρο P , βελτιώνεται με την εμπειρία.

Όλοι οι αλγόριθμοι μηχανικής μάθησης περνάνε από δυο στάδια, το στάδιο της εκπαίδευσης (training) και το στάδιο του ελέγχου (testing). Στο πρώτο στάδιο ο αλγόριθμος δέχεται ένα σύνολο δεδομένων (training set) από τα οποία αποικτάει γνώση για την επίτευξη συγκεκριμένου σκοπού. Ακόμα, μπορεί να υπάρχει υπάρχουσα γνώση από προηγούμενη εκπαίδευση του αλγορίθμου, την οποία μπορεί να λάβει υπόψιν του ή να την αγνοήσει κατά την εκπαίδευση. Στην συνέχεια, πραγματοποιείται η πιστοποίηση της

παραγόμενης γνώσης. Συγκεκριμένα, έχοντας ένα άλλο σύνολο δεδομένων (test set) ξένο ως προς το πρώτο, ελέγχεται η παραγόμενη γνώση κάνοντας ανάκληση του συνόλου μέσω του αλγορίθμου και στην συνέχεια γίνεται αξιολόγηση της απόδοσης του στο συγκεκριμένο σύνολο.

2.2.1 Είδη Μηχανικής Μάθησης

Οι αλγόριθμοι μηχανικής μάθησης μπορούν να χωριστούν σε τρεις κατηγορίες ανάλογα με το είδος της εμπειρίας από την οποία μαθαίνουν. Αυτές είναι ανάλογες με τους τρόπους με τους οποίους μαθαίνει ο άνθρωπος: την μη επιβλεπόμενη μάθηση, την επιβλεπόμενη μάθηση και την ενισχυτική μάθηση.

Μη επιβλεπόμενη μάθηση (Unsupervised Learning) είναι η διαδικασία κατά την οποία δοθέντος ενός κλειστού συνόλου δεδομένων με πολλά χαρακτηριστικά εξάγεται χρήσιμη πληροφορία για την δομή του συνόλου. Χρησιμοποιείται κυρίως σε προβλήματα συσταδιοποίησης (clustering).

Επιβλεπόμενη μάθηση (Supervised Learning), είναι η διαδικασία κατά την οποία δοθέντος ενός κλειστού συνόλου δεδομένων με πολλά χαρακτηριστικά, κάθε στοιχείο του οποίου σχετίζεται με μια ετικέτα (label), γίνεται προσέγγιση μιας συνάρτησης που απεικονίζει εισόδους (στοιχεία) σε εξόδους (labels) βάσει της εμπειρίας που έχει αποκτηθεί από το σύνολο δεδομένων. Χρησιμοποιείται κυρίως σε προβλήματα κατηγοριοποίησης (classification), πρόβλεψης (prediction) και διερμηνείας (interpretation).

Ενισχυμένη μάθηση (Reinforcement Learning), είναι η διαδικασία κατά την οποία κατασκευάζεται μια στρατηγική μέσω της αλληλεπίδρασης με ένα ανοικτό σύνολο δεδομένων.

Είναι φανερό ότι για κάθε πρόβλημα υπάρχει ένας κατάλληλος τρόπος μάθησης και για κάθε τρόπο μάθησης υπάρχει τουλάχιστον ένας κατάλληλος αλγόριθμος.

2.2.2 Επαγωγική Μάθηση

Για το πρόβλημα της παρούσας διπλωματικής, αυτό της πρόβλεψης τιμών (forecasting), θα χρησιμοποιήσουμε έναν αλγόριθμο επιβλεπόμενης επαγωγικής μάθησης. Η επαγωγική μάθηση στηρίζεται στην “ υπόθεση επαγωγικής μάθησης ” (inductive learning hypothesis), σύμφωνα με την οποία “ *κάθε υπόθεση h που προσεγγίζει καλά την συνάρτηση στόχο για ένα αρκετά μεγάλο σύνολο παραδειγμάτων, θα προσεγγίζει το ίδιο καλά τη συνάρτηση στόχο και για περιπτώσεις που δεν έχει εξετάσει*”. Στην επαγωγική μάθηση το σύστημα μαθαίνει μια συνάρτηση που ονομάζεται συνάρτηση στόχος (target function) και αποτελεί έκφραση του μοντέλου που περιγράφει τα δεδομένα. Η συνάρτηση στόχος απεικονίζει τις τιμές ενός συνόλου μεταβλητών, που ονομάζονται ανεξάρτητες μεταβλητές

ή μεταβλητές εισόδου ή χαρακτηριστικά, σε μια τιμή που ονομάζεται εξαρτημένη μεταβλητή ή μεταβλητή εξόδου. Το σύνολο των διαφορετικών δυνατών τιμών εισόδου της συνάρτησης, δηλαδή το πεδίο ορισμού της, ονομάζεται σύνολο **περιπτώσεων** ή **στιγμιότυπων** (instances). Κάθε στιγμιότυπο περιγράφεται από ένα σύνολο **χαρακτηριστικών** (features). Το σύνολο των εισόδων που δέχεται ο αλγόριθμος κατά την εκπαίδευση του ονομάζεται **σύνολο εκπαίδευσης** (training set) ή **παραδείγματα** και αποτελεί υποσύνολο του συνόλου στιγμιότυπων. Τα υπόλοιπα στιγμιότυπα αποτελούν το **σύνολο ελέγχου** (test set) και χρησιμοποιούνται στο στάδιο της πιστοποίησης. Κατά την διαδικασία της εκπαίδευσης προσεγγίζεται η συνάρτηση στόχος μέσω της **συνάρτησης πρόβλεψης** h (prediction function), η οποία δέχεται ως εισόδους στιγμιότυπα και επιστρέφει μια έξοδο-πρόβλεψη. Για να βελτιωθεί η συμπεριφορά της συνάρτησης πρόβλεψης κατά την μάθηση χρησιμοποιείται η **συνάρτηση κόστους** (cost ή loss function). Αυτή εντοπίζει την διαφορά της εξόδου της συνάρτησης πρόβλεψης με την μεταβλητή στόχο που είναι η επιθυμητή έξοδος. Σκοπός στο στάδιο εκπαίδευσης είναι η ελαχιστοποίηση της συνάρτησης κόστους. Έπειτα από την ολοκλήρωση της μάθησης του συστήματος, ακολουθεί το στάδιο του ελέγχου. Το σύστημα έχει πλέον αποκτήσει νέα γνώση και καλείται να προβλέψει την έξοδο των στιγμιότυπων του συνόλου ελέγχου.

2.2.3 Συνάρτηση Κόστους

Όπως προαναφέρθηκε, κατά την διαδικασία μάθησης χρησιμοποιείται η συνάρτηση κόστους για να υπολογιστεί η διαφορά της επιθυμητής εξόδου με την έξοδο της συνάρτησης πρόβλεψης. Η διαφορά αυτή ονομάζεται **σφάλμα** (loss) και είναι φανερό ότι όσο μικρότερη είναι η τιμή του τόσο καλύτερες είναι οι προβλέψεις του μοντέλου. Η επιλογή της κατάλληλης συνάρτησης κόστους εξαρτάται από την φύση του προβλήματος που καλείται να επιλύσει ένας αλγόριθμος μηχανικής μάθησης. Για προβλήματα κατηγοριοποίησης μια βασική έννοια που χρησιμοποιείται είναι η **Εντροπία**.

Εντροπία (Entropy) μια τυχαίας μεταβλητής X είναι το μέτρο της μέσης αβεβαιότητας που την χαρακτηρίζει. Για κατανομή πιθανότητας $P(X)$ και τυχαία μεταβλητή X , η εντροπία ορίζεται ως εξής:

$$H(X) = \begin{cases} -\int_x P(x)\log P(x), & \text{για συνεχείς τυχαίες μεταβλητές.} \\ -\sum_x P(x)\log P(x), & \text{για διακριτές τυχαίες μεταβλητές.} \end{cases} \quad (2.1)$$

Η εντροπία παίρνει τιμές από 0 ως 1, όσο πιο κοντά είναι η τιμή της στην μονάδα τόσο μεγαλύτερη είναι η αβεβαιότητα της τυχαίας μεταβλητής. Συνεπώς, είναι επιθυμητό η τιμή της να είναι κοντά στο κάτω φράγμα.

Παράδειγμα 2.1. Θεωρείται ένα πείραμα τύχης στο οποίο γίνεται ρίψη ενός αμερόληπτου νομίσματος. Ορίζοντας ως X την τυχαία μεταβλητή που εκφράζει το αποτέλεσμα της ρίψης η μεταβλητή θα παίρνει τις τιμές 0 και 1, που η μια αντιστοιχεί στο ενδεχόμενο το αποτέλεσμα να είναι κορώνα, ενώ η άλλη στο ενδεχόμενο το αποτέλεσμα να είναι γράμματα. Η μεταβλητή αυτή ακολουθεί την διωνυμική κατανομή Bernoulli. Επειδή το νόμισμα είναι αμερόληπτο, οι πιθανότητες των δυο ενδεχομένων θα είναι ίσες. Η εντροπία της X από την σχέση (2.1) υπολογίζεται ως εξής:

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b p(x_i) \Rightarrow$$

$$H(X) = - \sum_{i=1}^2 P(x_i) \log_2 \frac{1}{2} = 1$$

Μπορεί κάποιος να παρατηρήσει πως το αποτέλεσμα είναι αβέβαιο, το οποίο είναι απολύτως λογικό αν ληφθεί υπόψη πως τα δύο ενδεχόμενα στο εν λόγω παράδειγμα είναι ισοπίθανα και επομένως είναι δύσκολο να προβλεφθεί το αποτέλεσμα της επόμενης ρίψης. Αν το νόμισμα ήταν μη αμερόληπτο, η αβεβαιότητα θα μειωνόταν καθώς το ένα ενδεχόμενο θα ήταν πιθανότερο να συμβεί. Ας υποθέσουμε ότι το νόμισμα έχει λυγίσει και είναι πλέον μη αμερόληπτο. Η πιθανότητα το αποτέλεσμα της ρίψης να είναι κορώνα είναι 0.8 ενώ η πιθανότητα να είναι γράμματα είναι 0.2. Από την σχέση (2.1) προκύπτει:

$$H(X) = -0.8 \log_2 0.8 - 0.2 \log_2 0.2 \Rightarrow$$

$$H(X) = 0.72$$

Cross-Entropy Function: Η συνάρτηση σφάλματος εντροπίας ή αλλιώς συνάρτηση λογαριθμικού σφάλματος (log loss function) μετράει την απόδοση ενός μοντέλου ταξινόμησης, του οποίου η έξοδος είναι μια τιμή πιθανότητας μεταξύ 0 και 1 ή και πολλές τιμές μεταξύ 0 και 1 μια για κάθε κλάση του προβλήματος. Το σφάλμα εντροπίας αυξάνεται όσο η προβλεπόμενη τιμή πιθανότητας αποκλίνει από την τιμή της πραγματικής κλάσης. Η κοινή εντροπία και το λογαριθμικό σφάλμα είναι ελαφρώς διαφορετικές έννοιες, αλλά στη μηχανική μάθηση, όταν υπολογίζονται σφάλματα, οι δύο έννοιες ταυτίζονται. Ο τύπος για περισσότερες από δύο κλάσεις ($M > 2$) είναι ο εξής :

$$L_{ce} = - \sum_{i=1}^M y_i \log(p_i) \quad (2.2)$$

Όπου M είναι ο αριθμός των κλάσεων, y_i η πραγματική πιθανότητα αυτής της κλάσης, προβλεπόμενη πιθανότητα για αυτή την κλάση.

Παράδειγμα 2.2. Θεωρείται ένα πρόβλημα κατηγοριοποίησης με 5 κλάσεις. Για το πρόβλημα αυτό έχει υλοποιηθεί ένα σύστημα που προβλέπει για μια είσοδο σε ποια κλάση ανήκει επιστρέφοντας μια πιθανότητα για κάθε κλάση.

P	Y
0.024	0
0.925	1
0.014	0
0.011	0
0.026	0

Σχήμα 2.1 Προβλεπόμενο διάνυσμα P και πραγματικό διάνυσμα Y

Παρατηρούμε ότι το μοντέλο έχει προβλέψει σωστά την κλάση της εισόδου. Αυτό φαίνεται και στην μικρή τιμή που επιστρέφει η συνάρτηση κόστους. Αντικαθιστώντας στην σχέση (2.2) έχουμε:

$$\begin{aligned}
 L_{ce} &= -0 \log(0.024) - 1 \log(0.925) + \\
 &\quad - 0 \log(0.014) - 0 \log(0.011) + \\
 &\quad - 0 \log(0.026) = 0.0339
 \end{aligned}$$

Binary Cross-Entropy Function: Η συνάρτηση αυτή χρησιμοποιείται σε προβλήματα δυαδικής κατηγοριοποίησης. Υπολογίζεται ως το μέσο σφάλμα εντροπίας μεταξύ όλων των στιγμιότυπων.

$$L_{Bin CE} = -\frac{1}{N} \left[\sum_{i=1}^N y_i \log p_i + (1 - y_i) \log(1 - p_i) \right] \quad (2.3)$$

Σε αυτό το σημείο θα ήταν χρήσιμο να αναφερθεί η συνάρτηση **Μέσου Τετραγωνικού Σφάλματος** η οποία αφορά την κατηγορία του προβλήματος της πρόβλεψης τιμών, που θα μελετηθεί στην εργασία αυτή.

Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error): Η συνάρτηση χρησιμοποιείται σε προβλήματα πρόβλεψης τιμών. Υπολογίζεται ως ο μέσος όρος των τετραγώνων των διαφορών της πρόβλεψης \hat{y} με την μεταβλητή στόχο ή πραγματική τιμή y για όλα τα στιγμιότυπα.

$$MSE = \frac{1}{N} [\sum_{i=1}^N (\hat{y}_i - y_i)^2] \quad (2.4)$$

Σφάλμα Μέσης Τετραγωνικής Ρίζας (Root Mean Square Error): Η συνάρτηση χρησιμοποιείται σε προβλήματα πρόβλεψης τιμών. Υπολογίζεται ως ο μέσος όρος της τετραγωνικής ρίζας των τετραγώνων των διαφορών της πρόβλεψης \hat{Y} με την μεταβλητή στόχο ή πραγματική τιμή y για όλα τα στιγμιότυπα.

$$RMSE = \sqrt{\frac{1}{N} [\sum_{i=1}^N (\hat{y}_i - y_i)^2]} \quad (2.5)$$

Μέσο Απόλυτο Σφάλμα (Mean Absolute Error): Η συνάρτηση χρησιμοποιείται σε προβλήματα πρόβλεψης τιμών. Υπολογίζεται ως ο μέσος όρος της απόλυτης τιμής των διαφορών της πρόβλεψης \hat{Y} με την μεταβλητή στόχο ή πραγματική τιμή y για όλα τα στιγμιότυπα.

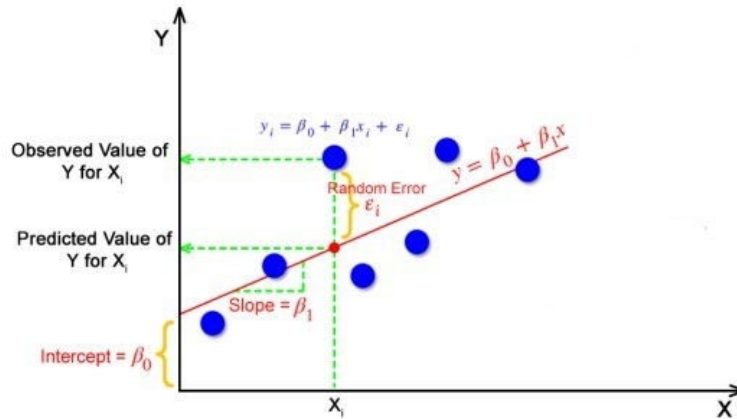
$$MAE = \frac{1}{N} [\sum_{i=1}^N |\hat{y}_i - y_i|] \quad (2.6)$$

Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error): Η συνάρτηση χρησιμοποιείται σε προβλήματα πρόβλεψης τιμών. Υπολογίζεται ως ο μέσος όρος της απόλυτης τιμής των διαφορών της πρόβλεψης \hat{Y} με την μεταβλητή στόχο ή πραγματική τιμή y προς την πραγματική τιμή για όλα τα στιγμιότυπα.

$$MAPE = \frac{1}{N} [\sum_{i=1}^N |\frac{\hat{y}_i - y_i}{y_i}|] \quad (2.7)$$

2.2.4 Γραμμική Παλινδρόμηση

Η γραμμική παλινδρόμηση είναι ένας από τους θεμελιώδεις αλγορίθμους της μηχανικής μάθησης. Αποτελεί την γραμμική προσέγγιση για την μοντελοποίηση της σχέσης μεταξύ της εξαρτημένης μεταβλητής με μία ή περισσότερες ανεξάρτητες μεταβλητές. Στην περίπτωση μιας μόνο ανεξάρτητης μεταβλητής καλείται απλή γραμμική παλινδρόμηση, για περισσότερες από μια ανεξάρτητες μεταβλητές καλείται πολλαπλή γραμμική παλινδρόμηση. Οι σχέσεις που δημιουργούνται για την μοντελοποίηση αυτή καλούνται γραμμικά μοντέλα.



Σχήμα 2.2 Γραμμική παλινδρόμηση

Στην απλή γραμμική παλινδρόμηση (βλέπε Σχήμα 2.2) με μία μόνο ανεξάρτητη μεταβλητή, η σχέση αναπαρίσταται ως εξής:

$$y = \beta_0 + \beta_1 X + \varepsilon \quad (2.8)$$

- y είναι η εξαρτημένη μεταβλητή
- X είναι η ανεξάρτητη μεταβλητή
- β_0 είναι η σταθερά
- β_1 είναι η κλίση
- ε είναι το σφάλμα

Για την πολλαπλή γραμμική παλινδρόμηση, όπου μπορεί να υπάρχουν περισσότερες από μια ανεξάρτητες μεταβλητές η σχέση επεκτείνεται στην ακόλουθη:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (2.9)$$

2.2.5 Πολυωνυμική Παλινδρόμηση

Η πολυωνυμική παλινδρόμηση είναι μια μορφή παλινδρόμησης στην οποία η σχέση μεταξύ της ανεξάρτητης μεταβλητής X και της εξαρτημένης μεταβλητής y μοντελοποιείται ως πολυώνυμο $n^{\text{οστου}}$ βαθμού του X . Η πολυωνυμική παλινδρόμηση προσαρμόζει μια μη γραμμική σχέση μεταξύ της τιμής του X και του αντίστοιχου υπό

συνθήκη μέσου όρου του y , που συμβολίζεται ως $E(y | X)$. Αν και η πολυωνυμική παλινδρόμηση προσαρμόζει ένα μη γραμμικό μοντέλο στα δεδομένα, ως πρόβλημα στατιστικής εκτίμησης, είναι γραμμική, με την έννοια ότι η συνάρτηση παλινδρόμησης $E(y | X)$ είναι γραμμική ως προς τις άγνωστες παραμέτρους που εκτιμώνται από τα δεδομένα. Για το λόγο αυτό, η πολυωνυμική παλινδρόμηση θεωρείται ειδική περίπτωση της πολλαπλής γραμμικής παλινδρόμησης.

Αντί να προσαρμόζει μια ευθεία η πολυωνυμική παλινδρόμηση προσαρμόζει μια πολυωνυμική συνάρτηση στα δεδομένα. Για έναν μόνο προγνωστικό παράγοντα (ένα χαρακτηριστικό feature) x , η πολυωνυμική συνάρτηση δευτέρου βαθμού είναι η παρακάτω:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \varepsilon \quad (2.10)$$

Σε συναρτήσεις υψηλότερου βαθμού προστίθενται και οι αντίστοιχοι όροι

$$\beta_3 x^3, \beta_4 x^4 \dots \beta_n x^n$$

Με την αύξηση του βαθμού της πολυωνυμικής παλινδρόμησης το μοντέλο που προκύπτει θα είναι καλύτερα προσαρμοσμένο στα δεδομένα εκπαίδευσης, προσθέτοντας ρίσκο καθώς το μοντέλο θα προσαρμόζεται και σε τυχόν θόρυβο που περιέχεται στα δεδομένα εκπαίδευσης. Αυτό πιθανότατα να οδηγήσει σε αδυναμία γενίκευσης στο σύνολο ελέγχου αλλά και σε αδυναμία γενίκευσης σε νέα άγνωστα ει των προτέρων δεδομένα.

Στην περίπτωση χρήσης πολλών εκτιμητριών (estimators) το πολυωνυμικό μοντέλο που θα προκύψει από την προσαρμογή στα δεδομένα μπορεί να περιλαμβάνει και σχέσεις μεταξύ των χαρακτηριστικών. Για παράδειγμα, για τους εκτιμητές x_1 και x_2 η συνάρτηση θα είναι:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_1 x_2 + \beta_5 x_2^2 + \varepsilon \quad (2.11)$$

-2.2.6 XGBoost

Το **XGBoost**, που σημαίνει **eXtreme Gradient Boosting**, είναι μια προηγμένη εφαρμογή του αλγορίθμου gradient boosting[7], βελτιστοποιημένη τόσο για την υπολογιστική αποδοτικότητα όσο και για την απόδοση του μοντέλου. Προερχόμενος από ένα ερευνητικό έργο του Tianqi Chen [8], ο XGBoost έχει σχεδιαστεί για να χειρίζεται αραιά δεδομένα (sparse data), να προσφέρει έναν παράλληλο μηχανισμό ενίσχυσης

δέντρων και να παρέχει κανονικοποιημένη ενίσχυση για την αποφυγή της υπερπροσαρμογής. Είναι ιδιαίτερα δημοφιλές σε διαγωνισμούς μηχανικής μάθησης, λόγω της ικανότητάς του να παρέχει ανώτερη ακρίβεια πρόβλεψης με σχετικά γρήγορους χρόνους υπολογισμού. Εκτός από τα προβλήματα ταξινόμησης και παλινδρόμησης, το XGBoost μπορεί επίσης να χειριστεί εργασίες κατάταξης και πρόβλεψης που καθορίζονται από τον χρήστη, καθιστώντας το ευέλικτο για ένα ευρύ φάσμα προβλημάτων της επιστήμης των δεδομένων.

2.2.7 Μετρικές Αξιολόγησης

Για την αξιολόγηση των μοντέλων θα γίνει χρήση ορισμένων δεικτών που ονομάζονται **μετρικές αξιολόγησης** (evaluation metrics). Η επιλογή των κατάλληλων μετρικών εξαρτάται από το πρόβλημα που επιλύεται, στην περίπτωση της πρόβλεψης τιμών οι συνηθέστερες μετρικές είναι το **Σφάλμα Μέσης Τετραγωνικής Ρίζας** (Root Mean Square Error), **Μέσο Τετραγωνικό Σφάλμα** (Mean Squared Error), **Μέσο Απόλυτο Σφάλμα** (Mean Absolute Error) και **Μέσο Απόλυτο Ποσοστιαίο Σφάλμα** (Mean Absolute Percentage Error). Οι σχέσεις για τις μετρικές αξιολόγησης παρουσιάζονται στην υποενότητα 2.2.3. Αυτές οι μετρικές χρησιμοποιούνται και για την αξιολόγηση των τεχνικών βαθιάς μάθησης που θα αναλυθούν στη συνέχεια και χρησιμοποιούνται στην παρούσα διπλωματική.

2.3 Βαθιά Μάθηση

Οι κλασικές τεχνικές μηχανικής μάθησης εμφανίζουν δυσκολία στην επεξεργασία ακατέργαστων δεδομένων. Για δεκαετίες η ανάπτυξη συστημάτων αναγνώρισης προτύπων και μηχανικής μάθησης απαιτούσε αυστηρή μελέτη αλλά και εξειδίκευση στο πεδίο έρευνας για τη σχεδίαση μεθόδων εξαγωγής χαρακτηριστικών που μετατρέπουν ακατέργαστα δεδομένα σε κατάλληλες αναπαραστάσεις, ώστε το σύστημα να είναι ικανό να εντοπίζει ή να κατηγοριοποιεί πρότυπα για τις εισόδους που του δίνονται και να μπορεί να προβλέπει αποτελεσματικά. Η δυσκολία αυτή των παραδοσιακών συστημάτων μηχανικής μάθησης αντιμετωπίστηκε με την χρήση των τεχνικών βαθιάς μάθησης.

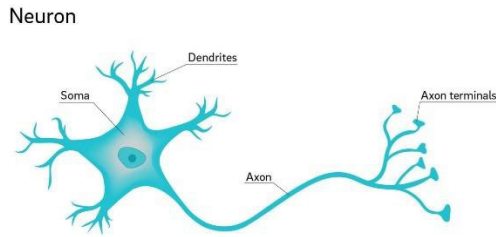
Η βαθιά μάθηση βασίζεται σε μεθόδους αναπαραστασιακής μάθησης (representational learning), οι οποίες επιτρέπουν στο σύστημα να βρῖσκει αυτόματα την αναπαράσταση που χρειάζεται ώστε να είναι εφικτή η κατηγοριοποίηση ή ο εντοπισμός των προτύπων. Συγκεκριμένα, οι μέθοδοι βαθιάς μάθησης χρησιμοποιούν πολλά επίπεδα αναπαράστασης που κατασκευάζονται από μη γραμμικές μονάδες, μετασχηματίζοντας την αναπαράσταση ενός επιπέδου σε αναπαράσταση ενός υψηλότερου και πιο αφαιρετικού

επιπέδου ξεινώνοντας από την ακατέργαστη είσοδο. Με αυτόν τον τρόπο τα συστήματα μπορούν να μάθουν πολύ σύνθετες συναρτήσεις. Στην κατηγοριοποίηση τα υψηλά επίπεδα αναπαράστασης δίνουν έμφαση σε χαρακτηριστικά της εισόδου που είναι σημαντικά για την διάκριση της. Για παράδειγμα, σε μια εικόνα που αναπαρίσταται με ένα πίνακα τιμών pixels τα χαρακτηριστικά που μαθαίνονται στο πρώτο επίπεδο συνήθως σχετίζονται με την παρουσία ή την απουσία ακμών σε συγκεκριμένα σημεία της εικόνας. Στο δεύτερο επίπεδο συνηθίζεται να εντοπίζονται μοτίβα που σχηματίζουν οι ακμές. Στο τρίτο επίπεδο συνδυάζονται τα μοτίβα σε κομμάτια γνωστών μας αντικειμένων και έπειτα σε ανώτερα επίπεδα συνδυάζονται τα αντικείμενα αυτά εντοπίζοντας τα αντικείμενα τα οποία αποτελούν. Το βασικό γνώρισμα της βαθιάς μάθησης είναι ότι αυτά τα επίπεδα δεν έχουν σχεδιαστεί από μηχανικούς, αλλά αναπτύσσονται από δεδομένα μέσω μιας διαδικασίας μάθησης γενικού σκοπού [9].

Αν και σήμερα αποτελεί ίσως τον πιο δραστήριο τομέα της μηχανικής μάθησης, από το 1943 που παρουσιάστηκε το πρώτο μοντέλο νευρωνικού δικτύου από τους McCulloch και Pitts μέχρι το 2006 [10], [11] δεν υπήρχε ενδιαφέρον για τις αρχιτεκτονικές βαθιάς μάθησης. Από τότε μέχρι και σήμερα οι τεχνικές μηχανικής μάθησης έχουν κάνει μεγάλη πρόοδο, βοήθησαν ακόμα και στην επίλυση προβλημάτων τα οποία ήταν άλυτα από την κοινότητα της τεχνητής νοημοσύνης για πολλά χρόνια. Σήμερα βρίσκουν εφαρμογή σε πολλούς επιστημονικούς τομείς μέσω των νευρωνικών δικτύων.

2.3.1 Νευρωνικά Δίκτυα

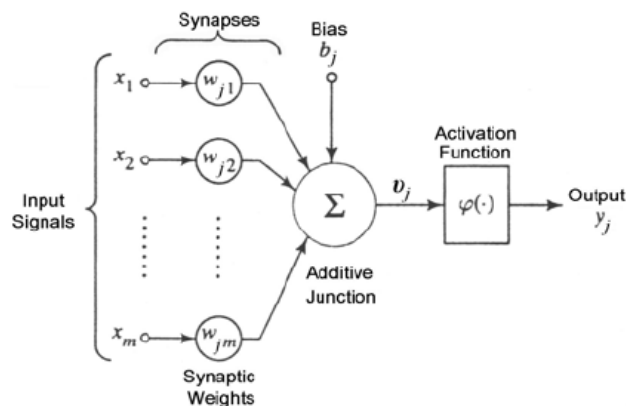
Η βασικότερη αρχιτεκτονική βαθιάς μάθησης είναι αυτή των νευρωνικών δικτύων. Τα νευρωνικά δίκτυα είναι υπολογιστικά μοντέλα εμπνευσμένα από την δομή και την λειτουργία των νευρώνων του εγκεφάλου. Ένας βιολογικός νευρώνας (βλέπε Σχήμα 2.3) αποτελείται από το σώμα που αποτελεί τον πυρήνα του, τους δενδρίτες μέσω των οποίων λαμβάνει σήματα από γειτονικούς νευρώνες (σήματα εισόδου) και τον άξονα που είναι η έξοδος του νευρώνα με τον οποίο συνδέεται με άλλους νευρώνες. Σε κάθε δενδρίτη υπάρχει ένα απειροελάχιστο κενό που ονομάζεται σύναψη. Οι συνάψεις μέσω χημικών διαδικασιών επιβραδύνουν ή επιταχύνουν τη ροή ηλεκτρονίων προς το σώμα του νευρώνα. Η ικανότητα μάθησης και μνήμης του εγκεφάλου στηρίζεται στην ικανότητα των συνάψεων να μεταβάλλουν την αγωγιμότητα τους. Τα ηλεκτρικά σήματα που, εισέρχονται στο σώμα των νευρώνων, συνδυάζονται και αν το αποτέλεσμα ξεπερνά κάποια τιμή κατωφλίου, το σήμα διαδίδεται προς άλλους νευρώνες μέσω του άξονα.



Σχήμα 2.3 Βιολογικός νευρώνας πηγή

Ένας ανθρώπινος εγκέφαλος αποτελείται από περίπου 100 δισεκατομμύρια νευρώνες με κάθε νευρώνα να συνδέεται με περίπου 1000 νευρώνες. Αυτό πραγματοποιείται μέσω του άξονα του νευρώνα στο οποίο καταλήγουν ισάριθμοι δενδρίτες άλλων νευρώνων. Δεδομένου ότι κάθε δενδρίτης έχει μια σύναψη, στον εγκέφαλο υπάρχουν 100 τρισεκατομμύρια συνάψεις οι οποίες επηρεάζουν την λειτουργία του. Αν και ο χρόνος απόκρισης των νευρώνων είναι της τάξης των χιλιοστών του δευτερολέπτου (ms), εντούτοις ο εγκέφαλος είναι σε θέση να λαμβάνει πολύπλοκες αποφάσεις. Αυτό οφείλεται στην υπολογιστική ισχύ του και στο ότι η πληροφορία είναι διαμοιρασμένη σε όλες του τις μονάδες, έτσι μπορεί να επεξεργάζεται παράλληλα και κατανεμημένα. Από τα παραπάνω είναι εμφανής η επιθυμία μοντελοποίησης της δομής και της λειτουργίας του ανθρώπινου εγκεφάλου, προφανώς σε μικρότερη κλίμακα καθώς είναι αδύνατο να μοντελοποιηθεί σε πλήρη κλίμακα. Τα μοντέλα αυτά ονομάζονται τεχνητά νευρωνικά δίκτυα.

Τα τεχνητά νευρωνικά δίκτυα είναι υπολογιστικά μοντέλα τα οποία σε αντιστοιχία με τους βιολογικούς νευρώνες του εγκεφάλου αποτελούνται από τεχνητούς νευρώνες (perceptron's). Όπως φαίνεται και στο Σχήμα 2.3, ένας τεχνητός νευρώνας δέχεται σήματα εισόδου $x_1, x_2, x_3, \dots, x_n$ τα οποία σε αντίθεση με τους ηλεκτρικούς παλμούς είναι συνεχείς μεταβλητές. Σε αντιστοιχία με τις σύναψη στην είσοδο του νευρώνα έχουμε το βάρος w_i το οποίο μεταβάλλει την τιμή της εισόδου.



Σχήμα 2.4 Τεχνητός νευρώνας πηγή

Στο σώμα του νευρωνικού γίνεται η άθροιση των εισόδων επηρεασμένων από τα βάρη, ενώ στην συνέχεια το άθροισμα περνάει από μια συνάρτηση ενεργοποίησης, η οποία διαμορφώνει την τελική τιμή της εξόδου y ανάλογα με την τιμή κατωφλίου που έχει η συνάρτηση ενεργοποίησης. Ένας νευρώνας μπορεί να έχει πολλές εξόδους, όλες όμως θα πρέπει να έχουν την ίδια τιμή. Ορισμένες φορές εκτός από τις εισόδους και τα αντίστοιχα βάρη, ο νευρώνας μπορεί να έχει και κάποιο βάρος w_0 το οποίο ονομάζεται **πόλωση** (bias) ή **παράγοντας προδιάθεσης του νευρώνα**. Η διαφορά της πόλωσης με τα υπόλοιπα βάρη είναι ότι επιδρά σε όλες τις τιμές της εισόδου.

Οι νευρώνες ενός νευρωνικού δικτύου είναι οργανωμένοι σε δομές παρόμοιες αυτών του ανθρώπινου εγκεφάλου. Συγκεκριμένα οργανώνονται σε μια σειρά από στρώματα ή επίπεδα (layers). Το πρώτο επίπεδο του δικτύου ονομάζεται επίπεδο εισόδου (input layer) και χρησιμοποιείται για την εισαγωγή δεδομένων. Τα δομικά στοιχεία αυτού του επιπέδου δεν είναι ουσιαστικά νευρώνες, καθώς δεν υπολογίζουν κάτι, απλώς προωθούν την είσοδο στο επόμενο επίπεδο, συνεπώς δεν έχουν βάρη και συνάρτηση ενεργοποίησης. Στην συνέχεια, ακολουθούν προαιρετικά ένα ή και περισσότερα ενδιάμεσα ή κρυφά επίπεδα (hidden layers), όπου στην περίπτωση που υπάρχουν περισσότερα από ένα κρυφά επίπεδα, το δίκτυο ονομάζεται βαθύ νευρωνικό δίκτυο (deep neural network). Στο τέλος του δικτύου υπάρχει το επίπεδο εξόδου (output layer).

Ιδιαίτερη σημασία σε ένα νευρωνικό δίκτυο έχει η σύνδεση μεταξύ των επιπέδων, ανάλογα με την σύνδεση τα επίπεδα διακρίνονται σε πλήρως και μερικώς συνδεδεμένα. Πλήρως συνδεδεμένα (fully connected) ονομάζονται δύο επίπεδα στα οποία όλοι οι νευρώνες του πρώτου επιπέδου συνδέονται με όλους τους νευρώνες του δεύτερου. Σε κάθε άλλη περίπτωση είναι μερικώς συνδεδεμένα (partially connected). Όταν η έξοδοι των νευρώνων ενός επιπέδου δεν συνδέονται με καμία είσοδο του ίδιου η προηγούμενου επιπέδου, το δίκτυο χαρακτηρίζεται ως πρόσθιας διάδοσης (feedforward network). Ενώ στην αντίθετη περίπτωση όπου υπάρχει σύνδεση με την είσοδο του ίδιου ή προηγούμενου επιπέδου, το δίκτυο χαρακτηρίζεται ως ανατροφοδοτούμενο (recurrent network).

2.3.2 Συναρτήσεις Ενεργοποίησης

Η συνάρτηση ενεργοποίησης είναι μια μαθηματική συνάρτηση η οποία καθορίζει την έξοδο του τεχνητού νευρώνα. Απλή μορφή τέτοιας συνάρτησης είναι η δυαδική συνάρτηση που επιστρέφει 0 ή 1 ανάλογα με την είσοδο, ενεργοποιώντας ή απενεργοποιώντας κατά αυτόν τον τρόπο τον νευρώνα. Επιπλέον, η συνάρτηση ενεργοποίησης χρησιμοποιείται σε περιπτώσεις που είναι επιθυμητή η κανονικοποίηση της εξόδου του νευρώνα, καθώς και σε περιπτώσεις που κρίνεται αναγκαία η εισαγωγή μη γραμμικότητας στο σύστημα. Οι συναρτήσεις ενεργοποίησης χωρίζονται σε γραμμικές και μη γραμμικές. Οι βασικότερες από αυτές είναι οι εξής:

Σιγμοειδής συνάρτηση (sigmoid function): Είναι μια παραγωγίσιμη συνάρτηση με πεδίο ορισμού το σύνολο των πραγματικών αριθμών και σύνολο τιμών το κλειστό σύνολο 0 ως 1. Χρησιμοποιείται συνήθως στο επίπεδο εξόδου νευρωνικών κατηγοριοποίησης δυο κλάσεων. Η συνάρτηση ορίζεται από τον παρακάτω τύπο:

$$S(x) = \frac{e^x}{e^x + 1} \quad (2.12)$$

Συνάρτηση Softmax: Η συνάρτηση αυτή συναντάται συχνά ως συνάρτηση ενεργοποίησης στο επίπεδο εξόδου νευρωνικών δικτύων κατηγοριοποίησης πολλών κλάσεων. Συγκεκριμένα, δέχεται ως είσοδο ένα διάνυσμα από πραγματικές τιμές και επιστρέφει ως έξοδο ένα διάνυσμα με τις κανονικοποιημένες εξόδους για κάθε κλάση, που κάθε μια διαιρείται με το άθροισμα όλων των κανονικοποιημένων εξόδων. Η συνάρτηση ορίζεται από τον παρακάτω τύπο:

$$\text{Softmax}(y_i) = \frac{e^{y_i}}{\sum_{i=1}^N e^{y_i}} \quad (2.13)$$

Συνάρτηση ReLU (Rectified Linear Unit): Αποτελεί τη συνηθέστερη συνάρτηση ενεργοποίησης στα νευρωνικά δίκτυα. Επιστρέφει την τιμή της εισόδου ή 0 στην περίπτωση που η είσοδος είναι μικρότερη από 0. Η συνάρτηση ορίζεται από τον παρακάτω τύπο:

$$\text{ReLU}(x) = \max(x, 0) \quad (2.14)$$

Συνάρτηση υπερβολικής εφαπτομένης (Hyperbolic Tangent Function): Η συνάρτηση αυτή μας επιστρέφει τον λόγο της συνάρτησης υπερβολικού ημιτόνου προς την συνάρτηση υπερβολικού συνημιτόνου. Η συνάρτηση ορίζεται από τον παρακάτω τύπο:

$$\tanh(x) = \frac{e^{2x} - 1}{e^{2x} + 1} \quad (2.15)$$

2.3.3 Αλγόριθμοι Βελτιστοποίησης

Οι αλγόριθμοι μηχανικής μάθησης συνδέονται άμεσα με την βελτιστοποίηση καθώς έχουν στόχο την μεγιστοποίηση ή την ελαχιστοποίηση της συνάρτησης κόστους $L(\theta)$, όπου θ είναι το διάνυσμα των παραμέτρων του νευρωνικού δικτύου. Η βελτιστοποίηση συνάρτησης κόστους επιτυγχάνεται κατά την διάρκεια της εκπαίδευσης

του συστήματος με την ανανέωση των παραμέτρων του. Οι μέθοδοι που καθορίζουν την ανανέωση ονομάζονται αλγόριθμοι βελτιστοποίησης. Είναι φανερό ότι για την αποτελεσματική εκπαίδευση του συστήματος έχει ιδιαίτερη σημασία η επιλογή του κατάλληλου αλγορίθμου. Παρακάτω παρουσιάζονται ορισμένοι από τους συνηθέστερους αλγορίθμους βελτιστοποίησης:

Αλγόριθμος Επικλινούς Καθόδου (Gradient Descent Algorithm). Είναι ένας από τους πιο δημοφιλείς αλγόριθμους βελτιστοποίησης που χρησιμοποιούνται στα νευρωνικά δίκτυα ειδικά σε προβλήματα κατηγοριοποίησης και παλινδρόμησης. Ο αλγόριθμος αυτός ελαχιστοποιεί την συνάρτηση κόστους ακολουθώντας το ελάχιστο της κλίσης (gradient) με την κλίση να υπολογίζεται από τον αλγόριθμο οπίσθιας διάδοσης. Επιπλέον, μέσω της οπίσθιας διάδοσης το σφάλμα διαδίδεται από τα τελευταία επίπεδα του δικτύου προς τα αρχικά, με τα βάρη κάθε επιπέδου να ανανεώνονται με τρόπο τέτοιο, ώστε να ελαχιστοποιηθεί η συνάρτηση κόστους. Αν και είναι εύκολος στην υλοποίηση και στους υπολογισμούς, ο αλγόριθμος αυτός δεν συνιστάται για μεγάλα σύνολα δεδομένων εκπαίδευσης, καθώς τα βάρη ανανεώνονται μετά τον υπολογισμό της κλίσης πάνω σε ολόκληρο το σύνολο δεδομένων. Αυτό έχει ως αποτέλεσμα ο αλγόριθμος να συγκλίνει πολύ αργά και να απαιτεί μεγάλο χώρο στην μνήμη. Τα ανανεωμένα βάρη του νευρωνικού υπολογίζονται ως εξής:

$$\theta_{new} = \theta_{old} - \varepsilon \frac{1}{n} \nabla_{\theta} \sum_{i=1}^n L(f(x_i, \theta_{old}), y_i) \quad (2.16)$$

Όπου $\varepsilon \in R$, είναι μια παράμετρος που ονομάζεται **ρυθμός μάθησης** (learning rate) και καθορίζει το βήμα σύγκλισης σε ένα ελάχιστο, f ή h είναι η συνάρτηση που προσεγγίζει την συνάρτηση στόχος (βλέπε υποενότητα 2.2.2), n είναι το πλήθος των στιγμιότυπων x_i του συνόλου εκπαίδευσης και y_i η μεταβλητή στόχος.

Αλγόριθμος Στοχαστικής Επικλινούς Καθόδου (Stochastic Gradient Descent Algorithm). Ο αλγόριθμος αυτός είναι παραλλαγή του αλγορίθμου επικλινούς καθόδου με την διαφορά ότι τα βάρη ανανεώνονται έπειτα από τον υπολογισμό της κλίσης ενός στιγμιότυπου ή μιας δέσμης στιγμιότυπων του συνόλου δεδομένων. Έτσι συγκλίνει γρηγορότερα σε ένα ελάχιστο και δεν έχει μεγάλες απαιτήσεις σε μνήμη καθώς δεν χρειάζεται να αποθηκεύει τα σφάλματα όλων των στιγμιότυπων μέχρι να γίνει αλλαγή στα βάρη του μοντέλου. Συνεπώς, ο αλγόριθμος αυτός είναι κατάλληλος για μεγάλα σύνολα δεδομένων. Επιλέγοντας μια δέσμη m στιγμιότυπων από το σύνολο εκπαίδευσης και τις αντίστοιχες μεταβλητές στόχους y_i η ανανέωση των βαρών υπολογίζεται από τον παρακάτω τύπο:

$$\theta_{new} = \theta_{old} - \varepsilon \frac{1}{m} \nabla_{\theta} \sum_{i=1}^m L(f(x_i, \theta_{old}), y_i) \quad (2.17)$$

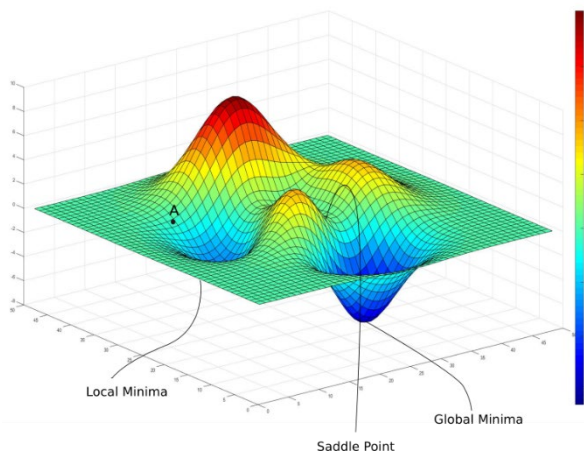
Αλγόριθμοι με ορμή (Momentum) [12] Όταν οι καμπύλες των συναρτήσεων κόστους περιέχουν πολλά ακρότατα (βλέπε Σχήμα 2.5(α)) τότε η εκπαίδευση του μοντέλου αργεί ακόμα και με τον αλγόριθμο στοχαστικής επικλινούς καθόδου. Αυτό συμβαίνει λόγω των έντονων ταλαντώσεων γύρω από τα τοπικά ελάχιστα που συναντάει ο αλγόριθμος (βλέπε Σχήμα 2.5(β)). Το πρόβλημα αυτό αντιμετωπίζεται με την εισαγωγή της μεθόδου της **ορμής u** [13]. Στους αλγορίθμους με ορμή λαμβάνονται υπόψιν στον υπολογισμό των βαρών οι κλίσεις που υπολογίστηκαν σε προηγούμενα βήματα. Κατά αυτό το τρόπο επιταχύνεται η διαδικασία σύγκλισης οδηγώντας τις παραμέτρους προς την σχετική κατεύθυνση και εξομαλύνοντας τις ταλαντώσεις στις άσχετες κατευθύνσεις. Επιλέγοντας μια δέσμη m στιγμιοτύπων από το σύνολο εκπαίδευσης και τις αντίστοιχες μεταβλητές στόχους y_i , η ανανέωση των βαρών ενός μοντέλου προσθέτοντας ορμή στον αλγόριθμο της στοχαστικής επικλινούς καθόδου υπολογίζεται ως εξής:

$$g = \frac{1}{m} \nabla_{\theta} \sum_{i=1}^m L(f(x_i, \theta_{old}), y_i) \quad (2.18)$$

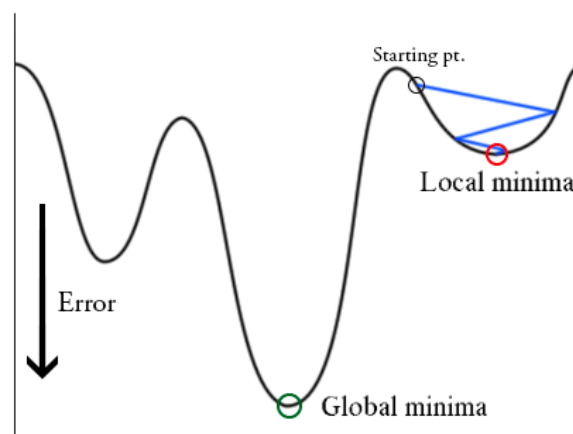
$$u_{new} = \alpha u_{old} - \varepsilon g \quad (2.19)$$

$$\theta_{new} = \theta_{old} + u_{new} \quad (2.20)$$

Όπου $\alpha \in [0,1]$ μια παράμετρος η τιμή της οποία καθορίζει κατά πόσο συνυπολογίζεται η τιμή της ορμής στα βάρη.



Σχήμα 2.5(α) Καμπύλη με πολλά ακρότατα. πηγή : researchgate.com



Σχήμα 2.5(β) Έντονες Ταλαντώσεις γύρω από το τοπικό ελάχιστο, πηγή: re-searchgate.com

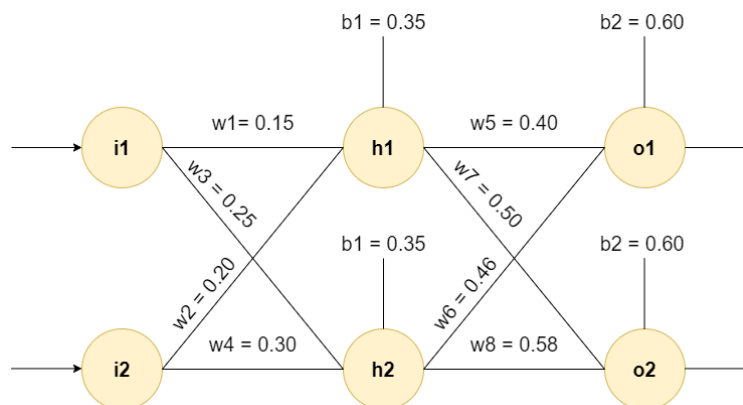
Αλγόριθμος Adam. Ο αλγόριθμος Adam (Adaptive Moment Estimation) [14] είναι ένας αλγόριθμος που περιλαμβάνει την μέθοδο της ορμής. Ο αλγόριθμος χρησιμοποιεί τις

τιμές των ροπών πρώτης (μέσος όρος) και δεύτερης (διακύμανσης) τάξης για την ανανέωση των παραμέτρων. Σε πολλά σύγχρονα συστήματα μηχανικής μάθησης γίνεται χρήση του αλγόριθμου Adam λόγω της ικανότητας του να συγκλίνει πολύ γρήγορα και να αντιμετωπίζει τα προβλήματα που εμφανίζουν οι προηγούμενες τεχνικές.

2.3.4 Οπίσθια Διάδοση

Η **οπίσθια διάδοση** [15] (back propagation) είναι ο αλγόριθμος με τον οποίο εκπαιδεύονται τα νευρωνικά δίκτυα, αρχικά υπολογίζοντας την κλίση της συνάρτησης κόστους και στην συνέχεια εφαρμόζοντας έναν αλγόριθμο βελτιστοποίησης. Για την κατανόηση της μεθόδου θα γίνει περιγραφή της εκπαίδευσης του νευρωνικού δικτύου. Δίνοντας μια είσοδο x σε ένα νευρωνικό δίκτυο η απαραίτητη πληροφορία περνάει από όλους τους νευρώνες του διασχίζοντας όλα τα επίπεδα του και στο τέλος επιστρέφεται μια έξοδος y . Η διαδικασία αυτή ονομάζεται **πρόσθια διάδοση** ή **πρόσθια τροφοδότηση** (forward propagation). Κατά την εκπαίδευση η πρόσθια διάδοση συνεχίζεται μέχρι να υπολογιστεί το σφάλμα της συνάρτησης κόστους $J(\theta)$. Ο αλγόριθμος οπίσθιας διάδοσης επιτρέπει την πληροφορία του σφάλματος να μεταδοθεί προς τα πίσω ξεκινώντας από το τελευταίο επίπεδο, με σκοπό τον υπολογισμό της κλίσης της συνάρτησης κόστους $\nabla_{\theta} J(\theta)$. Είναι σημαντικό εδώ να τονιστεί ότι ο αλγόριθμος οπίσθιας διάδοσης υπολογίζει μόνο την κλίση της συνάρτησης κόστους, η οποία χρησιμοποιείται στην συνέχεια για την ανανέωση των παραμέτρων του μοντέλου από τους αλγόριθμους βελτιστοποίησης. Για τον υπολογισμό της κλίσης γίνεται χρήση του κανόνα της αλυσίδας, ώστε να υπολογιστούν οι εξαρτήσεις όλων των παραγόντων που επιδρούν στην συνάρτηση κόστους.

Παράδειγμα 2.5. Έστω ένα νευρωνικό δίκτυο με ένα επίπεδο εισόδου, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου. Η τοπολογία του δικτύου καθώς και τα βάρη των εισόδων κάθε νευρώνα δίνονται από το παρακάτω σχήμα:



Σχήμα 2.6 Νευρωνικό Δίκτυο Παραδείγματος 2.5

Δίνετε στο δίκτυο το στιγμιότυπο $(i1, i2) = (0.05, 0.10)$ το οποίο έχει επιθυμητή έξοδο $(o1, o2) = (0.01, 0.99)(*)$.

Αρχικά, πραγματοποιείται η διαδικασία της **πρόσθιας διάδοσης** δίνοντας την είσοδο στο δίκτυο. Κάθε νευρώνας, εκτός αυτών του επιπέδου εισόδου, αθροίζει τις εισόδους του και στην συνέχεια χρησιμοποιεί την συνάρτηση ενεργοποίησης για να καθορίσει την έξοδο του, θεωρείται ότι όλοι οι νευρώνες έχουν συνάρτηση ενεργοποίησης την σιγμοειδή. Αναλυτικά:

$$h1_{sum} = w_1 i_1 + w_2 i_2 + b_1 \quad (1)$$

$$h1_{sum} = 0.3775 \quad (2)$$

$$h1_{out} = \frac{e^{h1_{sum}}}{e^{h1_{sum}} + 1} = \frac{1}{e^{-h1_{sum}} + 1} = 0.593269992 \quad (3)$$

$$h2_{sum} = w_3 i_1 + w_4 i_2 + b_1 \quad (4)$$

$$h2_{sum} = 0.3925 \quad (5)$$

$$h2_{out} = 0.596884378 \quad (6)$$

$$o1_{sum} = w_5 h1_{out} + w_6 h2_{out} + b2 = 1.105905967 \quad (7)$$

$$o1_{out} = \frac{e^{o1_{sum}}}{e^{o1_{sum}} + 1} = \frac{1}{e^{-o1_{sum}} + 1} = 0.75136507 \quad (8)$$

$$o2_{sum} = w_7 h1_{out} + w_8 h2_{out} + b2 = 1.224921404 \quad (9)$$

$$o2_{out} = 0.7729284653 \quad (10)$$

Για τον υπολογισμό του σφάλματος, έστω ότι η συνάρτηση κόστους είναι η συνάρτηση **Μέσου Τετραγωνικού Σφάλματος** (Mean Squared Error):

$$L_{total} = \sum_{i=1}^2 \frac{1}{2} (\text{target} - \text{out})^2 \quad (2.21)$$

Τότε από την (2.21):

$$\begin{aligned} L_{total} &= L_{o1} + L_{o2} \Rightarrow \\ L_{total} &= \frac{1}{2} (\text{target} - o1_{out})^2 + \\ &\quad + \frac{1}{2} (\text{target} - o2_{out})^2 \xrightarrow{(*),(8),(10)} \\ L_{total} &= 0.274811083 + 0.023560026 = 0.298371109 \end{aligned} \quad (11)$$

Τώρα με την μέθοδο της οπίσθιας διάδοσης θα υπολογιστεί η κλίση της συνάρτησης κόστους, ώστε στην συνέχεια με τον αλγόριθμο στοχαστικής επικλινούς καθόδου να ανανεωθούν τα βάρη του δικτύου.

Για την παράμετρο w_5 με χρήση του κανόνα της αλυσίδας:

$$\frac{\partial L_{total}}{\partial w_5} = \frac{\partial L_{total}}{\partial o1_{out}} \frac{\partial o1_{out}}{\partial o1_{sum}} \frac{\partial o1_{sum}}{\partial w_5} \quad (12)$$

$$\begin{aligned} \frac{\partial L_{total}}{\partial o1_{out}} &= \frac{\partial(\frac{1}{2}(target - o1_{out})^2 + \frac{1}{2}(target - o2_{out})^2)}{\partial o1_{out}} \\ &= -(target - o1_{out}) = 0.74136507 \end{aligned} \quad (13)$$

$$\begin{aligned} \frac{\partial o1_{out}}{\partial o1_{sum}} &= \frac{\partial(\frac{1}{e^{-o1_{sum}} + 1})}{\partial o1_{sum}} \Rightarrow \\ \frac{\partial o1_{out}}{\partial o1_{sum}} &= o1_{out}(1 - o1_{out}) = 0.18681502 \end{aligned} \quad (14)$$

$$\begin{aligned} \frac{\partial o1_{sum}}{\partial w_5} &= \frac{\partial(w_5 h1_{out} + w_6 h2_{out} + b2)}{\partial w_5} \Rightarrow \\ \frac{\partial o1_{sum}}{\partial w_5} &= h1_{out} = 0.593269992 \end{aligned} \quad (15)$$

Συνεπώς από (12),(13),(14),(15):

$$\frac{\partial L_{total}}{\partial w_5} = \frac{\partial L_{total}}{\partial o1_{out}} \frac{\partial o1_{out}}{\partial o1_{sum}} \frac{\partial o1_{sum}}{\partial w_5} = 0.593269992 \quad (16)$$

Η ανανέωση της παραμέτρου για ρυθμό μάθησης $\varepsilon = 0.5$ με τον αλγόριθμο βελτιστοποίησης της επικλινούς καθόδου υπολογίζεται ως:

$$w_{5_new} = w_{5_old} - 0.5 \frac{\partial L_{total}}{\partial w_{5_old}} = 0.3589648 \quad (17)$$

Ομοίως υπολογίζονται τα βάρη του επιπέδου εξόδου:

$$w_{6_new} = 0,408666186 \quad (18)$$

$$w_{7_new} = 0,511301270 \quad (19)$$

$$w_{8_new} = 0,561370121 \quad (20)$$

Για το κρυφό επίπεδο υπολογίζεται το w_1 :

$$\begin{aligned} \frac{\partial L_{total}}{\partial w_1} &= \frac{\partial L_{total}}{\partial h1_{out}} \frac{\partial h1_{out}}{\partial h1_{sum}} \frac{\partial h1_{sum}}{\partial w_1} \xrightarrow{(25)(26)(27)} \\ \frac{\partial L_{total}}{\partial w_1} &= 0.000438568 \end{aligned} \quad (20)$$

$$\frac{\partial L_{total}}{\partial h1_{out}} = \frac{\partial L_1}{\partial h1_{out}} + \frac{\partial L_2}{\partial h1_{out}} \quad (21)$$

$$\begin{aligned}\frac{\partial L_1}{\partial h_{1_{out}}} &= \frac{\partial L_1}{\partial o_{1_{sum}}} \frac{\partial o_{1_{sum}}}{\partial h_{1_{out}}} = 0.138498562 * 0.4 \Rightarrow \\ \frac{\partial L_1}{\partial h_{1_{out}}} &= 0.055399425\end{aligned}\quad (22)$$

$$\begin{aligned}\frac{\partial L_1}{\partial o_{1_{sum}}} &= \frac{\partial L_1}{\partial o_{1_{out}}} \frac{\partial o_{1_{out}}}{\partial o_{1_{sum}}} \Rightarrow \\ \frac{\partial L_1}{\partial o_{1_{sum}}} &= 0.74136507 * 0.18681502 \Rightarrow \\ \frac{\partial L_1}{\partial o_{1_{sum}}} &= 0.1384981304\end{aligned}\quad (23)$$

$$\begin{aligned}\frac{\partial o_{1_{sum}}}{\partial h_{1_{out}}} &= w_5 = 0.4 \\ \frac{\partial L_2}{\partial h_{1_{out}}} &= \frac{\partial L_2}{\partial o_{2_{out}}} \frac{\partial o_{2_{out}}}{\partial o_{2_{sum}}} \frac{\partial o_{2_{sum}}}{\partial h_{1_{out}}} \Rightarrow \\ \frac{\partial L_2}{\partial h_{1_{out}}} &= -0.019049119\end{aligned}\quad (24)$$

$$\begin{aligned}\frac{\partial L_{total}}{\partial h_{1_{out}}} &= 0.055399425 + (-0.019049119) \Rightarrow \\ \frac{\partial L_{total}}{\partial h_{1_{out}}} &= 0.036350306\end{aligned}\quad (25)$$

$$\begin{aligned}\frac{\partial h_{1_{out}}}{\partial h_{1_{sum}}} &= h_{1_{out}}(1 - h_{1_{out}}) \Rightarrow \\ \frac{\partial h_{1_{out}}}{\partial h_{1_{sum}}} &= 0.593269992(1 - 0.593269992) \Rightarrow \\ \frac{\partial h_{1_{out}}}{\partial h_{1_{sum}}} &= 0.2413007086\end{aligned}\quad (26)$$

$$\frac{\partial h_{1_{sum}}}{\partial w_1} = i_1 = 0.05 \quad (27)$$

Η νέα τιμή της παραμέτρου είναι:

$$w_{1_{new}} = w_{1_{old}} - 0.5 \frac{\partial L_{total}}{\partial w_{1_{old}}} = 0.149780716 \quad (28)$$

Ομοίως υπολογίζονται τα βάρη του κρυφού επιπέδου:

$$w_{2_{new}} = 0.19956143 \quad (29)$$

$$w_{3_{new}} = 0.24975114 \quad (30)$$

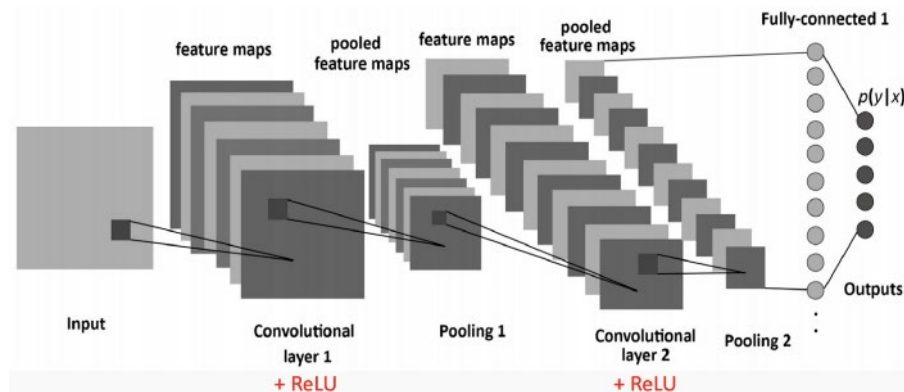
$$w_{4_{new}} = 0.29959229 \quad (31)$$

Ανανεώθηκαν όλα τα βάρη του δικτύου. Το σφάλμα τώρα είναι 0.291027924. Επαναλαμβάνοντας την διαδικασία αυτή 10000 φορές για το παραπάνω παράδειγμα το σφάλμα φτάνει την τιμή 0.0000351085. Σε αυτό το σημείο αν δοθεί η είσοδος (0.05, 0.1) η έξοδος του νευρωνικού θα είναι (0.015912196, 0.984065734), που είναι πολύ κοντά στην επιθυμητή έξοδο (0.15, 0.99).

2.3.4 Συνελικτικά Νευρωνικά Δίκτυα

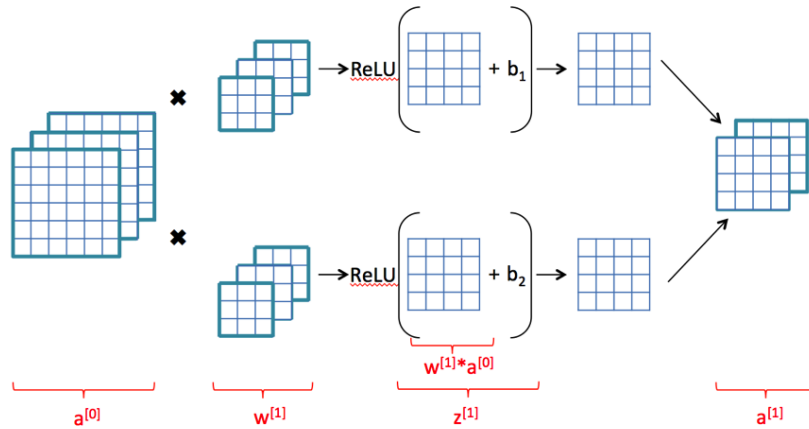
Τα συνελικτικά νευρωνικά δίκτυα [10], [11] είναι μια ειδική κατηγορία πρόσθιου νευρωνικού δικτύου, σχεδιασμένο για την επεξεργασία δεδομένων, τα οποία είναι σε μορφή πίνακα, όπως είναι τα σήματα, οι εικόνες και τα βίντεο. Η λειτουργία των συνελικτικών δικτύων βασίζεται στην ύπαρξη συνδέσεων μεταξύ γειτονικών νευρώνων, στον διαμοιρασμό βαρών, στην χωρική υποδειγματοληψία (pooling) και στη χρήση πολλών επιπέδων.

Αναλυτικότερα, η αρχιτεκτονική (βλέπε Σχήμα 2.7) ενός συνελικτικού δικτύου αποτελείται από στάδια, κάθε ένα από τα οποία περιέχει επίπεδα. Τα πρώτα στάδια περιλαμβάνουν δύο είδη επιπέδων: τα συνελικτικά επίπεδα και τα επίπεδα pooling.



Σχήμα 2.7 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου

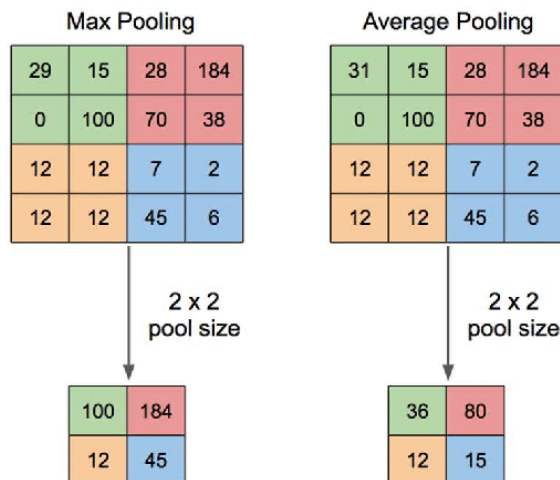
Οι νευρώνες ενός συνελικτικού επιπέδου χωρίζονται σε ομάδες οι οποίες είναι υπεύθυνες για την εξαγωγή ενός χάρτη χαρακτηριστικών (feature map). Συγκεκριμένα, κάθε νευρώνας υπολογίζει την συνέλιξη ενός τμήματος από τους χάρτες χαρακτηριστικών, που εξήγαγε το προηγούμενο επίπεδο, με μια μήτρα φίλτρων (kernel). Το αποτέλεσμα της συνέλιξης περνάει από μια μη γραμμική συνάρτηση ενεργοποίησης, όπως η ReLU, (βλέπε Σχήμα 2.8) πριν γίνει η εισαγωγή στον χάρτη χαρακτηριστικών. Όλοι οι νευρώνες μιας ομάδας μοιράζονται την ίδια μήτρα φίλτρων, συνεπώς διαφορετικοί χάρτες χαρακτηριστικών χρησιμοποιούν διαφορετικές μήτρες.



Σχήμα 2.8 Επίπεδο συνέλιξης

Η αρχιτεκτονική αυτή στηρίζεται στο γεγονός ότι πίνακες δεδομένων όπως οι εικόνες, οι γειτονικές τιμές εικονοστοιχείων (pixels) συσχετίζονται μεταξύ τους σχηματίζοντας μοτίβα που μπορούν να εντοπιστούν εύκολα. Αυτό, σε συνδυασμό με ότι τα μοτίβα αυτά δεν συναντώνται αποκλειστικά σε μια θέση ή μπορεί και να επαναλαμβάνονται σε μια εικόνα καθιστά την χρήση κοινών φίλτρων αποδοτική. Ενώ η χρονική πολυπλοκότητα της πρόσθιας μετάδοσης παραμένει ίδια, μειώνονται σημαντικά οι σημαντικοί οι παράμετροι του μοντέλου.

Το επίπεδο που ακολουθεί το επίπεδο της συνέλιξης είναι το επίπεδο του pooling. Ειεί γίνεται διαλογή των χαρακτηριστικών που παρήγαγε το προηγούμενο επίπεδο διατηρώντας τα σημαντικότερα χαρακτηριστικά και συμπύσσοντας εκείνα που είναι σημασιολογικά παρόμοια. Οι νευρώνες για να κάνουν τη διαλογή (βλέπε σχήμα 2.8) υπολογίζουν το μέγιστο (max) ή τον μέσο (average) γειτονικού τμήματος του χάρτη χαρακτηριστικών. Ακόμα, υπάρχει η δυνατότητα από τους γειτονικούς τους νευρώνες να επιλέξουν τμήμα του χάρτη προσπερνώντας έναν αριθμό γραμμών ή στηλών του χάρτη χαρακτηριστικών. Κατά αυτό τον τρόπο επιτυγχάνεται μείωση στη διάσταση της αναπαράστασης. Για παράδειγμα, για τον εντοπισμό ενός προσώπου σε μια εικόνα δεν χρειάζεται να γνωρίζει κανείς που ακριβώς είναι η θέση των ματιών στην εικόνα παρά μόνο αν υπάρχουν τα μάτια.



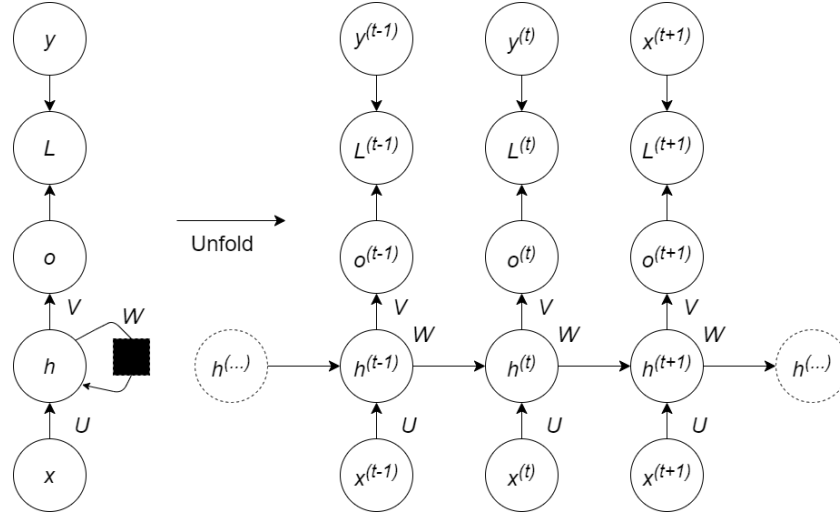
Σχήμα 2.9 Μέγιστη και μέση υποδειγματοληψία

Από όσα προαναφέρθηκαν είναι φανερό ότι γίνεται αποτελεσματικά η εξαγωγή των απαραίτητων χαρακτηριστικών χωρίς να φορτώνεται το σύστημα με παραμέτρους. Η προσθήκη σταδίων αποτελούμενων από επίπεδα συνέλιξης και pooling επιτρέπει την ιεραρχική αναπαράσταση των χαρακτηριστικών, ξεκινώντας από χαρακτηριστικά χαμηλού επιπέδου και καταλήγοντας σε υψηλά χαρακτηριστικά. Τα στάδια αυτά μπορεί να ακολουθούνται από ένα στάδιο με πλήρως συνδεδεμένα επίπεδα, στην περίπτωση που το δίκτυο προορίζεται για κατηγοριοποίηση. Τα συνελικτικά δίκτυα βρίσκουν εφαρμογές σε πολλά επιστημονικά πεδία που αφορούν ήχο, εικόνα και βίντεο. Ένα από αυτά είναι η ιατρική απεικόνιση, η οποία αποτελεί αντικείμενο μελέτης αυτής της εργασίας. Θα αναλυθούν παρακάτω δύο τύποι συνελικτικών νευρωνικών που θα χρησιμοποιηθούν στην συνέχεια.

2.3.5 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα

Τα ανατροφοδοτούμενα νευρωνικά δίκτυα (Recursive Neural Networks) ανήκουν σε μια κατηγορία νευρωνικών δικτύων που εξειδικεύεται στην επεξεργασία ακολουθιακών δεδομένων, όπως για παράδειγμα τα κάρε σε ένα βίντεο ή οι λέξεις σε ένα κείμενο. Προκειμένου να επεξεργαστούν και να αναλύσουν την ακολουθιακή πληροφορία των δεδομένων τα ανατροφοδοτούμενα δίκτυα μιμούνται τον άνθρωπο εισάγοντας την έννοια της μνήμης. Έτσι είναι σε θέση να θυμούνται εισόδους που χρησιμοποιήθηκαν για την παραγωγή προηγούμενων εξόδων. Η έξοδος του δικτύου δεν εξαρτάται μόνο από δεδομένα της εισόδου εκείνη την χρονική στιγμή και τις παραμέτρους του δικτύου, αλλά και από ένα διάνυμα “κρυφής” κατάστασης (hidden state) που αναπαριστά όλες τις παρελθοντικές εισόδους στο δίκτυο. Επομένως, ένα RNN για δύο περιπτώσεις με την ίδια είσοδο μπορεί να αποδώσει διαφορετικές εξόδους όταν τα προηγούμενα δεδομένα εισόδου είναι διαφορετικά σε κάθε περίπτωση.

Για το Σχήμα 2.10 ο υπολογισμός της επόμενης κρυφής κατάστασης εξαρτάται από την είσοδο και την προηγούμενη κρυφή κατάσταση. Παρατηρείται ότι δεν έχει προσδιοριστεί η επιλογή της συνάρτησης ενεργοποίησης για τους νευρώνες της κρυφής κατάστασης, για αυτό θα επιλεγεί αυθαίρετα η συνάρτηση υπερβολικής εφαπτομένης. Επιπλέον, θεωρείται ότι η έξοδος \mathbf{o} είναι διακριτή. Στην συνέχεια, μπορεί να εφαρμοστεί συνάρτηση ενεργοποίησης Softmax. Η πρόσθια διάδοση στο δίκτυο του Σχήματος 2.10 ξεκινά με την αρχικοποίηση της κρυφής μνήμης $\mathbf{h}^{(0)}$. Έπειτα, σε κάθε βήμα από την χρονική στιγμή $t = 1$ ως την $t = \tau$ εφαρμόζονται οι εξής τύποι:



Σχήμα 2.10 Ακολουθητικό Δίκτυο

$$\mathbf{a}^{(t)} = \mathbf{b} + \mathbf{W}\mathbf{h}^{(t-1)} + \mathbf{U}\mathbf{x}^{(t)} \quad (2.21)$$

$$\mathbf{h}^{(t)} = \tanh(\mathbf{a}^{(t)}) \quad (2.22)$$

$$\mathbf{o}^{(t)} = \mathbf{c} + \mathbf{V}\mathbf{h}^{(t)} \quad (2.23)$$

$$\hat{\mathbf{y}} = \text{softmax}(\mathbf{o}^{(t)}) \quad (2.24)$$

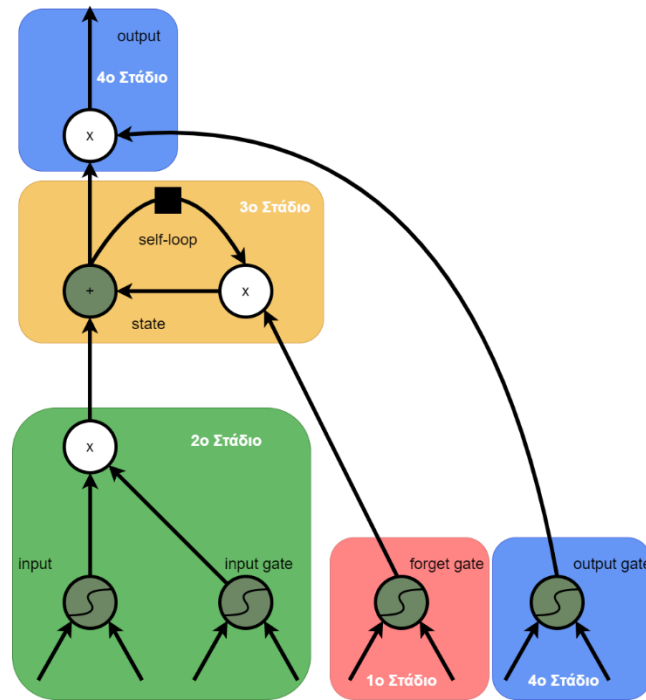
Όπου \mathbf{c} και \mathbf{b} τα βάρη πόλωσης και $\mathbf{U}, \mathbf{V}, \mathbf{W}$ οι πίνακες παραμέτρων για την είσοδο του νευρώνα κρυφής κατάστασης, την έξοδο του νευρώνα κατάστασης και την μετάβαση στον επόμενο νευρώνα κατάστασης αντίστοιχα.

2.3.6 LSTM

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (Long Short-Term Memory) [1] είναι ένα ανατροφοδοτούμενο νευρωνικό δίκτυο το οποίο ανήκει σε μια κατηγορία δικτύων που ονομάζονται ανατροφοδοτούμενα νευρωνικά δίκτυα με πύλες (gated recurrent networks). Τα δίκτυα αυτά σε σχέση με τα απλά RNN επιτρέπουν την διαχείριση πληροφορίας για μεγάλη διάρκεια ενώ αν κρίνεται σκόπιμο, η πληροφορία αυτή μπορεί να διατηρηθεί, να

ανανεωθεί ή να διαγραφεί. Αυτό είναι ιδιαίτερα χρήσιμο σε περιπτώσεις που η ακολουθία που εξετάζεται αποτελείται από υπακολουθίες στις οποίες αναζητείται πληροφορία, οπότε είναι απαραίτητος ένας μηχανισμός που να ανανεώνει την κατάσταση του δικτύου όσο εξετάζει μια υπακολουθία ενώ να την διαγράφει καθώς μεταβαίνει σε μια νέα. Επιπλέον τα LSTM αντιμετωπίζουν το πρόβλημα της εξαφανιζόμενης κλίσης που συναντάται στα απλά RNNs.

Τα LSTM διαφέρουν από τα κλασικά RNNs στην δομή (βλέπε σχήμα 2.11) των μονάδων τους, οι οποίες ονομάζονται κύτταρα (cell). Συγκεκριμένα, εκτός του βρόχου τροφοδότησης προστίθεται εντός της μονάδας ένας εσωτερικός βρόχος. Ο βρόχος αυτός διατηρεί την κατάσταση κυττάρου (cell state) ώστε το δίκτυο να διαχειριστεί πληροφορία για μεγάλα διαστήματα. Ακόμη, τα κύτταρα των LSTM εκτός της εισόδου και της εξόδου, διαθέτουν τρεις πύλες, την πύλη άγνοιας (forget gate), την πύλη εισόδου (input gate) και την πύλη εξόδου (output gate). Η λειτουργία των πυλών και κατά συνέπεια της δομικής μονάδας των LSTM θα γίνει κατανοητή μέσω της διαδικασίας της πρόσθιας διάδοσης, η οποία θα χωριστεί σε τέσσερα στάδια. Αναλυτικά, για δεδομένη χρονική στιγμή t σε ένα κύτταρο i :



Σχήμα 2.11. Δομή κυττάρου LSTM

Στο πρώτο στάδιο η εισόδος $x^{(t)}$ και η προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ του συστήματος εισάγονται στην πύλη άγνοιας, η οποία είναι ένας νευρώνας με βάρη U^f, W^f , βάρη πόλωσης b^f και συνάρτηση ενεργοποίησης την σιγμοειδή. Η πύλη αυτή υπολογίζει μια τιμή $f_i^{(t)}$ μεταξύ του 0 και 1 που καθορίζει την πληροφορία που θα διατηρηθεί ή θα διαγραφεί στην συνέχεια από την νέα κατάσταση κυττάρου $s^{(t)}$.

$$f_i^{(t)} = \sigma(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)}) \quad (2.25)$$

Το δεύτερο στάδιο περιλαμβάνει δύο βήματα. Στο πρώτο βήμα η είσοδος $x^{(t)}$ εισάγεται σε έναν νευρώνα με βάρη U, W , βάρη πόλωσης b και συνάρτηση ενεργοποίησης την σιγμοειδή, προκειμένου να παράγει μια υποψήφια τιμή που μπορεί να προστεθεί μετέπειτα στην τρέχουσα κατάσταση. Στο δεύτερο βήμα η είσοδος $x^{(t)}$ και η προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ του συστήματος εισάγονται στην πύλη εισόδου, η οποία είναι ένας νευρώνας με βάρη U^g, W^g , βάρη πόλωσης b^g και συνάρτηση ενεργοποίησης την σιγμοειδή. Η πύλη αυτή υπολογίζει μια τιμή $g_i^{(t)}$ μεταξύ του 0 και 1 που καθορίζει κατά πόσο θα υπολογιστεί η νέα υποψήφια τιμή στην νέα κατάσταση κυττάρου $s^{(t)}$.

$$i_i^{(t)} = \sigma(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)}) \quad (2.26)$$

$$g_i^{(t)} = \sigma(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)}) \quad (2.27)$$

Στο τρίτο στάδιο θα γίνει ανανέωση της κατάστασης κυττάρου. Συγκεκριμένα, η τιμή $f_i^{(t)}$ πολλαπλασιάζεται με την τρέχουσα κατάσταση κυττάρου $s^{(t-1)}$, έτσι ώστε να διατηρηθεί η πληροφορία που αποφασίστηκε στο πρώτο στάδιο. Στην συνέχεια, προστίθεται το γινόμενο της νέας υποψήφιας τιμής με την τιμή $g_i^{(t)}$. Από το άθροισμα των δυο γινομένων προκύπτει η νέα κατάσταση κυττάρου $s^{(t)}$.

$$s^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} i_i^{(t)} \quad (2.28)$$

Τέλος, στο τέταρτο στάδιο θα υπολογιστεί η έξοδος της πύλης εξόδου $q_i^{(t)}$ και η νέα κρυφή κατάσταση $h_i^{(t)}$. Αρχικά η είσοδος $x^{(t)}$ και η προηγούμενη κρυφή κατάσταση $h^{(t-1)}$ του συστήματος εισάγονται στην πύλη εισόδου, η οποία είναι ένας νευρώνας με βάρη U^o, W^o , βάρη πόλωσης b^o και συνάρτηση ενεργοποίησης την σιγμοειδή. Η πύλη αυτή υπολογίζει μια τιμή $o_i^{(t)}$ μεταξύ του 0 και 1 που καθορίζει κατά πόσο θα υπολογιστεί η προηγούμενη κρυφή κατάσταση τιμή στον υπολογισμό της νέας κρυφής κατάστασης. Έπειτα, η νέα κατάσταση κυττάρου $s^{(t)}$ εισάγεται σε μια σιγμοειδή συνάρτηση ενεργοποίησης και πολλαπλασιάζεται με την έξοδο $q_i^{(t)}$. Το αποτέλεσμα του γινομένου είναι η νέα κρυφή κατάσταση $h_i^{(t)}$.

$$q_i^{(t)} = \sigma(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)}) \quad (2.29)$$

$$h_i^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)} \quad (2.30)$$

2.3.17 Bidirectional LSTM

Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης διπλής κατεύθυνσης [16], [17] είναι μια επέκταση των τυπικών LSTM τα οποία έχουν σημαντικά καλύτερη απόδοση σε προβλήματα κατηγοριοποίησης ακολουθιών. Αυτά τα δίκτυα εκπαιδεύουν συγχρόνως δυο LSTM δίκτυα στην ακολουθία εισόδου, το ένα στην ακολουθία στην αρχική της μορφή και το δεύτερο σε ένα αντίγραφο της ακολουθίας που έχει αντίστροφη φορά. Αυτό το καινούργιο περιεχόμενο εισάγεται στο δίκτυο δημιουργώντας καλύτερα και γρηγορότερα αποτελέσματα.

2.2.17 Γενίκευση

Μια σημαντική πρόκληση που υπάρχει στο τομέα της μηχανικής μάθησης είναι το πόσο καλά αποδίδουν οι αλγόριθμοι σε νέα δεδομένα. Η ικανότητα ενός αλγορίθμου να αποδίδει καλά σε νέα δεδομένα ονομάζεται **γενίκευση** (generalization). Ένα σύστημα μηχανικής μάθησης κατά την εκπαίδευση στοχεύει στην ελαχιστοποίηση του σφάλματος εκπαίδευσης (training error). Πολλές φορές όμως αποτυγχάνει να διατηρήσει ταυτόχρονα χαμηλό και το **σφάλμα γενίκευσης**, δηλαδή το σφάλμα στο σύνολο ελέγχου. Ανάλογα με το σφάλμα εκπαίδευσης και το σφάλμα γενίκευσης ένα μοντέλο μπορεί να χαρακτηριστεί ως **υπερπροσαρμοσμένο** (overfitted) και **υποπροσαρμοσμένο** (underfitted). Το μοντέλο χαρακτηρίζεται ως underfitted όταν έχει υψηλό σφάλμα εκπαίδευσης και υψηλό σφάλμα γενίκευσης. Στην περίπτωση που το μοντέλο διατηρεί χαμηλό σφάλμα εκπαίδευσης και έχει υψηλό σφάλμα γενίκευσης, τότε το μοντέλο έχει μάθει καλά το σύνολο εκπαίδευσης, αλλά δεν μπορεί να το γενίκευση την γνώση του σε νέα δεδομένα. Για την αντιμετώπιση αυτού του προβλήματος χρησιμοποιούνται τεχνικές γενίκευσης.

Εμπλουτισμός Δεδομένων (Data Augmentation). Η τεχνική αυτή αφορά την αύξηση του συνόλου δεδομένων εκπαίδευσης με προσθήκη μετασχηματισμών των υπαρχόντων δεδομένων. Σε εικόνες τέτοιοι μετασχηματισμοί μπορεί να είναι ολισθήσεις. Κατά αυτό τον τρόπο προστίθενται "νέα" δεδομένα που μπορούν να βοηθήσουν στην εκπαίδευση του συστήματος.

Dropout. Κατά την τεχνική του dropout επιλέγονται τυχαία νευρώνες ενός επιπέδου του δικτύου και δεν χρησιμοποιούνται στην εκπαίδευση. Συγκεκριμένα, αυτοί οι κόμβοι δεν συμπεριλαμβάνονται στον υπολογισμό στην πρόσθια καθώς και στην οπίσθια διάδοση.

Αυτό βοηθάει στην αποφυγή της υπερπροσαρμογής των νευρώνων στα χαρακτηριστικά του συνόλου εκπαίδευσης.

Κεφάλαιο 3

Δεδομένα

3.1 Εισαγωγή

Για την παρούσα εργασία δημιουργήθηκε ένα σύνολο δεδομένων το οποίο περιλαμβάνει τιμές χρηματιστηρίου αγροτικών προϊόντων από το χρηματιστήριο της Αμερικής, συγκεκριμένα τιμές προθεσμιακών συμβολαίων καθώς και τηλεσκοπικά δεδομένα από τις κύριες περιοχές παραγωγής αυτών των προϊόντων. Η προσέγγιση αυτή ακολουθεί την λογική της δημοσίευσης **Forecasting Agricultural Commodity Prices Using Dual Input Attention LSTM ([2])**. Αρχικά, έγινε η συλλογή των δεδομένων από την πλατφόρμα του **investing.com** για τα χρηματιστηριακά προϊόντα και χρήση του **API** της **NASA** για συλλογή των τηλεσκοπικών δεδομένων για τις περιοχές ενδιαφέροντος. Έπειτα, έγινε **καθαρισμός των δεδομένων** (data cleaning), **δημιουργία νέων χαρακτηριστικών** (feature engineering) για το σύνολο δεδομένων αξιοποιώντας γνώση από τον χρηματιστηριακό τομέα για την δημιουργία δεικτών που μπορούν να δώσουν χρήσιμη πληροφορία για τα χρηματιστηριακά προϊόντα μελέτης.

3.2 Τιμές από προθεσμιακά συμβόλαια

Ένα συμβόλαιο μελλοντικής εκπλήρωσης είναι μια νομική συμφωνία για την αγορά ή την πώληση ενός συγκεκριμένου περιουσιακού στοιχείου αγαθών ή τίτλων σε προκαθορισμένη τιμή σε συγκεκριμένη χρονική στιγμή στο μέλλον. Τα συμβόλαια μελλοντικής εκπλήρωσης είναι τυποποιημένα ως προς την ποιότητα και την ποσότητα για να διευκολύνουν τις συναλλαγές σε ένα χρηματιστήριο μελλοντικής εκπλήρωσης. Τα συμβόλαια μελλοντικής εκπλήρωσης αγαθών αποτελούν αντικείμενο διαπραγμάτευσης σε χρηματιστήρια εμπορευμάτων, όπως το **Chicago Board of Trade (CBOT)**, το οποίο αποτελεί πλέον μέρος του ομίλου **Chicago Mercantile Exchange (CME)**.

Οι έμποροι μπορούν να αγοράζουν ή να πωλούν συμβόλαια μελλοντικής εκπλήρωσης καλαμποκιού με βάση τις προσδοκίες τους για τις μελλοντικές τιμές του καλαμποκιού. Εάν πιστεύουν ότι οι τιμές του αγαθού θα αυξηθούν, αγοράζουν, ενώ εάν αναμένουν πτώση των τιμών, πωλούν. Οι αγρότες και οι γεωργικοί παραγωγοί χρησιμοποιούν συχνά προθεσμιακά συμβόλαια γεωργικών προϊόντων για να

αντισταθμίσουν την έκθεσή τους στις διακυμάνσεις των τιμών. Για παράδειγμα, ένας αγρότης καλαμποκιού μπορεί να πουλήσει προθεσμιακά συμβόλαια καλαμποκιού για να κλειδώσει μια μελλοντική τιμή πώλησης, εξασφαλίζοντας ότι θα λάβει ένα προκαθορισμένο ποσό για το καλαμπόκι του, ακόμη και αν οι τιμές της αγοράς πέσουν. Ενώ ορισμένοι έμποροι χρησιμοποιούν προθεσμιακά συμβόλαια γεωργικών αγαθών για να τα παραλάβουν πραγματικά, οι περισσότεροι ενδιαφέρονται να κερδοσκοπήσουν επί της τιμής. Ως εκ τούτου, πολλά συμβόλαια μελλοντικής εκπλήρωσης διακανονίζονται σε μετρητά και όχι με φυσική παράδοση. Αυτό σημαίνει ότι στο τέλος της συμβατικής περιόδου, το κέρδος ή η ζημία διακανονίζεται σε μετρητά.

Οι συμβάσεις γεωργικών προϊόντων, όπως και οι συμβάσεις για άλλα εμπορεύματα, τηρούν τυποποιημένες συμβάσεις και κώδικες για να εξασφαλίζουν σαφήνεια και συνέπεια στις συναλλαγές αυτών των προϊόντων. Αυτές οι συμβάσεις περιλαμβάνουν συνήθως μια σειρά αλφαριθμητικών κωδικών που παρέχουν βασικές πληροφορίες σχετικά με τη σύμβαση. Για παράδειγμα, ένας κωδικός υποδεικνύει το συγκεκριμένο εμπόρευμα που αποτελεί αντικείμενο διαπραγμάτευσης, τον μήνα λήξης του συμβολαίου και το έτος κατά το οποίο λήγει. Κάθε εμπόρευμα έχει τον δικό του μοναδικό κωδικό, όπως "ZC" για το καλαμπόκι ή "ZW" για το σιτάρι, ενώ τα γράμματα αντιπροσωπεύουν τον μήνα του συμβολαίου (π.χ. "H" για τον Μάρτιο, "K" για τον Μάιο, "N" για τον Ιούλιο, "U" για τον Σεπτέμβριο, "Z" για τον Δεκέμβριο). Το αριθμητικό ψηφίο υποδηλώνει το έτος λήξης. Αυτοί οι τυποποιημένοι κωδικοί διευκολύνουν τους εμπόρους και τους επενδυτές να εντοπίζουν και να διαπραγματεύονται συγκεκριμένες συμβάσεις στον πολύπλοκο κόσμο των γεωργικών προϊόντων, εξασφαλίζοντας διαφάνεια και αποτελεσματικότητα στις αγορές εμπορευμάτων.

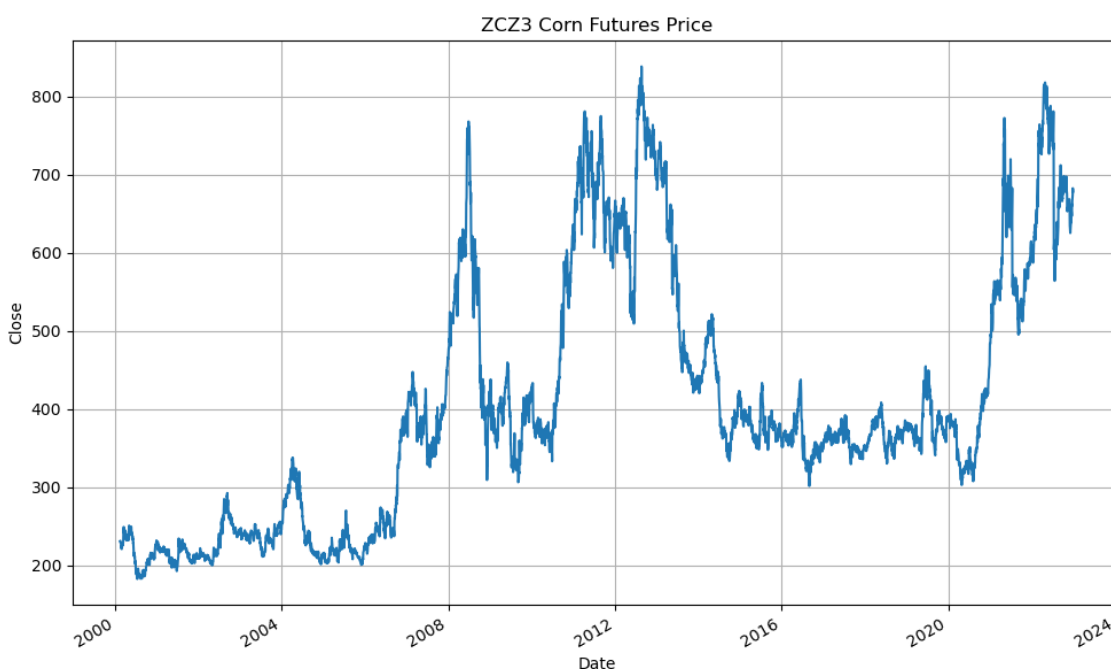
Στην διπλωματική αυτή θα μελετηθούν οι τιμές προθεσμιακών συμβολαίων καλαμποκιού (ZCZ3).

3.2.1 Προθεσμιακά Συμβόλαια Καλαμποκιού ZCZ3

Ο κωδικός προθεσμιακού συμβολαίου ZCZ3 αναφέρεται σε προθεσμιακό συμβόλαιο καλαμποκιού. Συγκεκριμένα ο "ZC" χρησιμεύει ως κωδικός εμπορεύματος, λειτουργώντας ως σαφές αναγνωριστικό για το καλαμπόκι, το οποίο είναι το υποκείμενο περιουσιακό στοιχείο στο συγκεκριμένο συμβόλαιο μελλοντικής εκπλήρωσης. Το πρώτο γράμμα, "Z", υποδηλώνει συνήθως τον μήνα λήξης του συμβολαίου, με κάθε γράμμα να αντιστοιχεί σε συγκεκριμένο μήνα του έτους. Στην προκειμένη περίπτωση, το "Z" υποδηλώνει τον Δεκέμβριο, ενώ άλλα γράμματα χρησιμοποιούνται για να υποδηλώσουν διαφορετικούς μήνες ("H" για τον Μάρτιο, "K" για τον Μάιο, "N" για τον Ιούλιο, "U" για τον Σεπτέμβριο, "Z" για τον Δεκέμβριο). Το αριθμητικό ψηφίο που ακολουθεί τον κωδικό μήνα, στην προκειμένη περίπτωση, το "3", αντιπροσωπεύει το έτος λήξης, το οποίο συχνά αντιστοιχεί στο συγκεκριμένο έτος κατά το οποίο η σύμβαση έχει οριστεί να λήξει.

Τα δεδομένα που χρησιμοποιήθηκαν για το ZCZ3 προέρχονται από το investing.com και αφορούν μια χρονική ακολουθία από τις αρχές του 2000 έως το τέλος του 2022, με ημερήσια συχνότητα. Ειδικότερα για κάθε ημέρα το σύνολο δεδομένων περιλαμβάνει:

- την τιμή ανοίγματος (Open).
- την ελάχιστη τιμή κατά την διάρκεια της ημέρας (Low),
- την μέγιστη τιμή κατά την διάρκεια της ημέρας (High),
- την τιμή κλεισίματος (Price),
- το αριθμό των συναλλαγών που έγιναν κατά την διάρκεια της ημέρας (Volume)



Σχήμα 3.1 Τιμές προθεσμιακών συμβολαίων καλαμποζιού

3.3 Τηλεσκοπικά Δεδομένα

Τα συμβόλαια μελλοντικής εκπλήρωσης για τα αγροτικά προϊόντα επηρεάζονται από διάφορους παράγοντες, όπως οι καιρικές συνθήκες που επηρεάζουν τις αποδόσεις των καλλιεργειών. Στην μελέτη που έγινε επιλέχθηκαν τηλεσκοπικά δεδομένα που επηρεάζουν τις καλλιέργειες των αγροτικών προϊόντων. Τα δεδομένα συλλέχθηκαν με την χρήση του **API** της **NASA**.

Το **API Power and Propulsion Data API** της **NASA**, το οποίο φιλοξενείται στο Ερευνητικό Κέντρο Langley, προσφέρει ένα ευρύ φάσμα τηλεσκοπικών δεδομένων για πολλές τοποθεσίες σε όλο τον κόσμο. Η υπηρεσία αυτή είναι σχεδιασμένη με γνώμονα τον τομέα των ανανεώσιμων πηγών ενέργειας, παρέχει πρόσβαση τόσο σε ιστορικά δεδομένα

όσο και σε δεδομένα σχεδόν πραγματικού χρόνου για διάφορες παραμέτρους, όπως η ηλιακή ακτινοβολία (άμεση και διάχυτη), η θερμοκρασία, η υγρασία και η ταχύτητα του ανέμου. Ο ευέλικτος σχεδιασμός της πλατφόρμας επιτρέπει διαφορετικές χρονικές αναλύσεις που κυμαίνονται από λεπτά έως ετήσιους μέσους όρους. Αυτό που την κάνει να ξεχωρίζει είναι η διεπαφή της με επίκεντρο τον χρήστη, που οι χρήστες μπορούν να καθορίζουν αβίαστα την τοποθεσία, το χρονικό εύρος και τις επιθυμητές παραμέτρους. Αν και το κύριο κοινό του είναι η βιομηχανία ανανεώσιμων πηγών ενέργειας, τα δεδομένα προσελκύουν επίσης μετεωρολόγους, εκπαιδευτικούς και ερευνητές. Η αξιοπιστία των δεδομένων αυτών απορρέει από τις ολοκληρωμένες μεθόδους άντλησης, οι οποίες συνδυάζουν δορυφορικές παρατηρήσεις, επίγειες μετρήσεις και προηγμένα υπολογιστικά μοντέλα. Για όσους είναι έμπειροι στην τεχνολογία και την έρευνα, η NASA διευκολύνει την πρόσβαση σε API για απρόσκοπτη ανάκτηση δεδομένων, καθιστώντας την ενσωμάτωση με άλλες πλατφόρμες και εργαλεία απλή.

Στην παρούσα εργασία τα δεδομένα που αντλήθηκαν μέσω του API αφορούν ημερήσιες μετρήσεις από σημεία της πολιτείας με την μεγαλύτερη παραγωγή για το αγροτικό προϊόν μελέτης. Για το **καλαμπόκι** η περιοχή με την **μεγαλύτερη παραγωγή** είναι η πολιτεία της **Αϊόβα**. Συγκεκριμένα, λήφθηκαν μέσω του API τα ακόλουθα:

- **T2M**: Αναφέρεται στη θερμοκρασία στα 2 μέτρα πάνω από το έδαφος. Δεδομένου ότι αυτό το ύψος είναι κοντά στο μέσο ύψος των ανθρώπων, το T2M χρησιμοποιείται συχνά ως αναπαράσταση της θερμοκρασίας περιβάλλοντος που βιώνουν οι άνθρωποι. Είναι μια κρίσιμη παράμετρος για διάφορες εφαρμογές, από τη γεωργία έως τον αστικό σχεδιασμό.
- **T2MDEW**: Αντιπροσωπεύει τη θερμοκρασία σημείου δρόσου σε ύψος 2 μέτρων πάνω από το έδαφος. Το σημείο δρόσου δείχνει τη θερμοκρασία στην οποία ο αέρας γίνεται κορεσμένος με υγρασία, οδηγώντας στη συμπύκνωση των υδρατμών σε δρόσο. Ένα υψηλότερο σημείο δρόσου υποδηλώνει περισσότερη υγρασία στον αέρα, η οποία μπορεί να αποτελεί μέτρο της υγρασίας και του επιπέδου άνεσης. Είναι σημαντικό στην πρόγνωση του καιρού, ιδίως όταν προβλέπεται ομίχλη ή παγετός.
- **T2M_MAX**: Πρόκειται για τη μέγιστη θερμοκρασία που καταγράφεται σε ύψος 2 μέτρων πάνω από το έδαφος κατά τη διάρκεια μιας συγκεκριμένης περιόδου (συνήθως μιας ημέρας). Η παρακολούθηση της ημερήσιας μέγιστης θερμοκρασίας είναι απαραίτητη για διάφορους τομείς, από την υγεία (προειδοποιήσεις για καύσιμα) έως την ενέργεια (πρόβλεψη αιχμής ζήτησης ενέργειας).
- **T2M_MIN**: Αντιπροσωπεύει την ελάχιστη θερμοκρασία στα 2 μέτρα πάνω από το έδαφος κατά τη διάρκεια μιας συγκεκριμένης περιόδου (συνήθως μιας ημέρας). Είναι ζωτικής σημασίας για εφαρμογές όπως η γεωργία, όπου ο παγετός μπορεί να βλάψει τις καλλιέργειες. Έχει επίσης σημασία για την υγεία, όπου οι εξαιρετικά χαμηλές θερμοκρασίες μπορούν να προκαλέσουν κινδύνους υποθερμίας.

- **RH2M:** Δηλώνει τη σχετική υγρασία σε ύψος 2 μέτρων πάνω από το έδαφος. Η σχετική υγρασία είναι ένα μέτρο της περιεκτικότητας του αέρα σε υγρασία σε σχέση με τη μέγιστη ποσότητα που μπορεί να συγκρατήσει ο αέρας σε αυτή τη θερμοκρασία. Αποτελεί βασική παράμετρο στην πρόγνωση του καιρού, τη γεωργία και την κατανόηση των επιπέδων άνεσης. Η υψηλή υγρασία μπορεί επίσης να επηρεάσει την ανάπτυξη μούχλας στα κτίρια και την εξάπλωση ορισμένων ασθενειών.
- **ALLSKY_SFC_SW_DWN:** Αντιπροσωπεύει την επιφανειακή βραχύκυματική ακτινοβολία προς τα κάτω σε όλο τον ουρανό. Με απλούστερους όρους, είναι η ποσότητα της ηλιακής ακτινοβολίας που φτάνει στην επιφάνεια της Γης, λαμβάνοντας υπόψη παράγοντες όπως τα σύννεφα. Αυτή η παράμετρος είναι θεμελιώδης για τις εφαρμογές ηλιακής ενέργειας. Βοηθά στον προσδιορισμό του ηλιακού δυναμικού μιας περιοχής. Επιπλέον, είναι σημαντική σε μελέτες κλιματολογίας, καθώς επηρεάζει το θερμικό ισοζύγιο και τη θερμοκρασία μιας περιοχής.

Κεφάλαιο 4

Επεξεργασία Δεδομένων

4.1 Εισαγωγή

Στο παρόν κεφάλαιο θα αναλυθούν οι διαδικασίες που ακολουθήθηκαν για την επεξεργασία των δεδομένων ώστε να είναι στην κατάλληλη μορφή για την αποτελεσματική αξιοποίηση τους από τα μοντέλα μηχανικής και βαθιάς μάθησης. Οι διαδικασίες αφορούν τον καθαρισμό των δεδομένων (data cleaning), την δημιουργία νέων χαρακτηριστικών από τα ήδη υπάρχοντα (feature engineering), την επιλογή των χαρακτηριστικών (feature selection), και τέλος την κανονικοποίηση των χαρακτηριστικών.

4.2 Καθαρισμός Δεδομένων

Ο καθαρισμός των δεδομένων αποτελεί απαραίτητο στάδιο της επεξεργασίας των δεδομένων καθώς η ποιότητα και η αξιοπιστία κάθε ανάλυσης ή συμπεράσματος εξαρτώνται από την ακεραιότητα των υποκείμενων δεδομένων. Τα ακατέργαστα δεδομένα, συχνά περιέχουν ανακρίβειες, ασυνέπειες ή κενά, τα οποία μπορούν να οδηγήσουν σε παραπλανητικές ή εσφαλμένες ερμηνείες, εάν δεν διορθωθούν. Μέσω του επιμελούς καθαρισμού των δεδομένων, διασφαλίζεται ότι οι επακόλουθες αναλύσεις βασίζονται σε μια στέρεη και αξιόπιστη βάση, ελαχιστοποιώντας τις πιθανές μεροληψίες ή στρεβλώσεις. Από τα δεδομένα που συλλέχθηκαν αφαιρέθηκαν εγγραφές από τις οποίες απουσίαζε κάποιο χαρακτηριστικό (feature).

4.2 Μηχανική Χαρακτηριστικών

Η μηχανική των χαρακτηριστικών είναι μια βασική πτυχή της ροής εργασίας της μηχανικής μάθησης και της μοντελοποίησης δεδομένων, με επίκεντρο τη δημιουργία, την τροποποίηση και τον μετασχηματισμό των χαρακτηριστικών για τη βελτίωση της απόδοσης του μοντέλου. Περιλαμβάνει την αξιοποίηση της γνώσης του τομέα για τη δημιουργία νέων χαρακτηριστικών από τα υπάρχοντα δεδομένα ή τον μετασχηματισμό τους με τρόπο που τα καθιστά πιο χρήσιμα και κατάλληλα για μοντελοποίηση. Η τέχνη της μηχανικής των χαρακτηριστικών έγκειται στον εντοπισμό μοτίβων, σχέσεων ή

αποχρώσεων στα δεδομένα που μπορεί να μην είναι άμεσα εμφανή, αλλά μπορούν να επηρεάσουν σημαντικά την προβλεπτική ικανότητα ενός μοντέλου. Με τη δημιουργία χαρακτηριστικών που αντιπροσωπεύουν καλύτερα τις υποκείμενες δομές και τις περιπλοκές των δεδομένων, μπορεί κανείς συχνά να επιτύχει βελτιωμένη ακρίβεια και γενίκευση του μοντέλου. Πέρα από τη βελτίωση των επιδόσεων, η αποτελεσματική μηχανική των χαρακτηριστικών μπορεί να οδηγήσει σε μοντέλα που είναι περισσότερο ερμηνεύσιμα και διορατικά, ρίχνοντας φως στη δυναμική του προβλήματος που αντιμετωπίζεται. Στο πεδίο της μηχανικής μάθησης, η μηχανική των χαρακτηριστικών λέγεται συχνά ότι είναι τόσο τέχνη όσο και επιστήμη, υποστηρίζοντας τη λεπτή ισορροπία μεταξύ της αναπαράστασης των δεδομένων και της αποτελεσματικότητας του μοντέλου.

Για το σκοπό αυτό θα χρησιμοποιηθεί γνώση από τον χρηματοοικονομικό τομέα για την δημιουργία και προσθήκη νέων χαρακτηριστικών που μπορούν να δώσουν ωφέλιμη πληροφορία για τα προθεσμιακά συμβόλαια αγροτικών προϊόντων. Οι δείκτες που παρατίθενται χρησιμοποιούνται κυρίως στην τεχνική ανάλυση, μια μεθοδολογία που χρησιμοποιείται για την πρόβλεψη της κατεύθυνσης των τιμών των χρηματοπιστωτικών αγορών με βάση ιστορικά δεδομένα. Οι δείκτες που προστέθηκαν είναι οι παρακάτω:

- Ο **EMA_9** αντιπροσωπεύει τον Εκθετικό Κινητό Μέσο Όρο 9 περιόδων, ο οποίος δίνει μεγαλύτερη βαρύτητα στις πρόσφατες τιμές, καθιστώντας τον πιο ευαίσθητο στις μεταβολές των τιμών σε σχέση με έναν απλό κινητό μέσο όρο (SMA).
- Οι **SMA_5**, **SMA_10**, **SMA_15** και **SMA_30** είναι απλοί κινητοί μέσοι όροι που υπολογίζονται για 5, 10, 15 και 30 περιόδους αντίστοιχα, οι οποίοι χρησιμοποιούνται συνήθως για τον εντοπισμό αναστροφών τάσεων με σύγκριση με τις τρέχουσες τιμές.
- Ο **RSI**, ή Δείκτης Σχετικής Δύναμης, είναι ένας ταλαντωτής ορμής που μετρά την ταχύτητα και την αλλαγή των κινήσεων των τιμών, ο οποίος συνήθως ταλαντώνεται μεταξύ 0 και 100, με τιμές πάνω από 70 που υποδεικνύουν συνθήκες υπεραγοράς και κάτω από 30 που υποδεικνύουν συνθήκες υπερπώλησης.
- Ο **MACD**, ο Moving Average Convergence Divergence, είναι ένας άλλος δείκτης δυναμικής που δείχνει τη σχέση μεταξύ δύο κινητών μέσων όρων της τιμής ενός προθεσμιακού συμβολαίου, συγκεκριμένα του **EMA_12** και του **EMA_26**.
- Η γραμμή **MACD_signal** είναι μια εξομαλυμένη έκδοση της γραμμής **MACD** και όταν η **MACD** διασταυρώνεται πάνω ή κάτω από αυτή τη γραμμή σήματος, μπορεί να σηματοδοτήσει πιθανές ευκαιρίες αγοράς ή πώλησης.
- Τέλος, ο δείκτης **Williams % R**, που συχνά αποκαλείται ως Williams Percent Range, είναι ένας δείκτης δυναμικής που κινείται μεταξύ 0 και -100, σχεδιασμένος για να προσδιορίζει αν ένα περιουσιακό στοιχείο είναι υπεραγορασμένο ή υπερπουλημένο. Γενικά, οι ενδείξεις πάνω από το -20 θεωρούνται υπεραγορασμένες, ενώ οι ενδείξεις κάτω από το -80 θεωρούνται υπερπουλημένες. Συλλογικά, αυτοί οι δείκτες είναι εργαλεία που χρησιμοποιούν οι έμποροι και οι

αναλυτές για να προβλέπουν τις τάσεις της αγοράς και να λαμβάνουν τειμηριωμένες αποφάσεις συναλλαγών.

Τα μοντέλα που μελετήθηκαν στην παρούσα διπλωματική προβλέπουν τις τιμές των προθεσμιακών συμβολαίων για τα αγροτικά προϊόντα μελέτης από 1 έως n μέρες από την μέρα αναφοράς που δίνεται ως είσοδος. Συνεπώς σε κάθε παράδειγμα του συνόλου δεδομένων θα περιέχονται χαρακτηριστικά που δεν συσχετίζονται με την μεταβλητή στόχο, δηλαδή την τιμή του προθεσμιακού συμβολαίου. Για αυτό σε κάθε εγγραφή του συνόλου θα προστεθούν επιπλέον χαρακτηριστικά που αφορούν την προηγούμενη μέρα από την μέρα αναφοράς. Συγκεκριμένα, προστέθηκαν τα εξής:

- **prev_low**: Αναφέρεται στην χαμηλότερη τιμή της προηγούμενης ημέρας.
- **prev_close**: Αναφέρεται στην τιμή κλεισίματος της προηγούμενης ημέρας.
- **prev_open**: Αναφέρεται στην τιμή ανοίγματος της προηγούμενης ημέρας.
- **prev_high**: Αναφέρεται στην υψηλότερη τιμή της προηγούμενης ημέρας.
- **prev_Vol.**: Αναφέρεται στον αριθμό των συναλλαγών της προηγούμενης ημέρας.
- **prev_T2M**: Αναφέρεται στην θερμοκρασία T2M της προηγούμενης ημέρας.
- **prev_T2M_MAX**: Αναφέρεται στην μέγιστη θερμοκρασία T2M της προηγούμενης ημέρας.
- **prev_T2M_MIN**: Αναφέρεται στην ελάχιστη θερμοκρασία T2M της προηγούμενης ημέρας.
- **prev_RH2M**: Αναφέρεται στην υγρασία RH2M της προηγούμενης ημέρας.
- **prev_ALLSKY_SFC_SW_DWN**: Αναφέρεται στην συσώρευση επιφανειακής βραχυκυματικής ακτινοβολίας της προηγούμενης ημέρας.
- Οι **SMA_5**, **SMA_10**, **SMA_15** και **SMA_30** για την θερμοκρασία T2M είναι απλοί κινητοί μέσοι όροι που υπολογίζονται για 5, 10, 15 και 30 περιόδους αντίστοιχα.
- **MAX_days_T2M_MAX**: Αναφέρεται στην μέγιστη θερμοκρασία τις προηγούμενες **days=5,10,15,30** μέρες.
- **MIN_days_T2M_MIN**: Αναφέρεται στην ελάχιστη θερμοκρασία τις προηγούμενες **days=5,10,15,30** μέρες.

Ακόμα δημιουργήθηκαν τρία νέα χαρακτηριστικά που προκύπτουν από το χαρακτηριστικό **Date** (ημέρα) της εγγραφής. Αυτά είναι τα εξής:

- **Day**: Ημέρα Από την ημερομηνία εγγραφής.
- **Month**: Μήνας Από την ημερομηνία της εγγραφής.
- **Year**: Από την ημερομηνία της εγγραφής.

4.3 Επιλογή χαρακτηριστικών

Η επιλογή χαρακτηριστικών είναι ένα κρίσιμο βήμα στη διαδικασία της μηχανικής μάθησης και της μοντελοποίησης δεδομένων, με στόχο τον εντοπισμό και τη διατήρηση των πιο συναφών χαρακτηριστικών από ένα σύνολο δεδομένων. Με την επιλογή των πιο σχετικών χαρακτηριστικών ή μεταβλητών, μπορεί κανείς να βελτιώσει την απόδοση ενός μοντέλου, να μειώσει την υπερπροσαρμογή και να επιταχύνει τη διαδικασία εκπαίδευσης. Η επιλογή χαρακτηριστικών είναι απαραίτητη επειδή δεν συμβάλλουν όλα τα χαρακτηριστικά με νόημα στο αποτέλεσμα της πρόβλεψης, ορισμένα μπορεί να είναι περιττά ή ακόμη και επιζήμια, εισάγοντας θόρυβο αντί για σαφήνεια. Οι αποτελεσματικές τεχνικές επιλογής χαρακτηριστικών μπορούν να βοηθήσουν στην απλούστευση των μοντέλων, καθιστώντας τα ευκολότερα στην ερμηνεία και πιο διαφανή. Επιπλέον, με τη μείωση της διαστατικότητας των δεδομένων, το υπολογιστικό κόστος μπορεί να ελαχιστοποιηθεί, οδηγώντας σε ταχύτερη και αποτελεσματικότερη εκπαίδευση και επικύρωση μοντέλων. Στην ουσία, η επιλογή χαρακτηριστικών χρησιμεύει ως γέφυρα, διασφαλίζοντας ότι τα μοντέλα είναι τόσο υπολογιστικά αποδοτικά όσο και βέλτιστα ακριβή. Από τα δεδομένα που συλλέχθηκαν αρχικά και από τις μεταβλητές που δημιουργήθηκαν στο στάδιο της μηχανικής χαρακτηριστικών αφαιρούνται από κάθε εγγραφή τα εξής:

- **low:** Αναφέρεται στην χαμηλότερη τιμή της ημέρας.
- **close:** Αναφέρεται στην τιμή κλεισίματος της ημέρας.
- **open:** Αναφέρεται στην τιμή ανοίγματος της ημέρας.
- **high:** Αναφέρεται στην υψηλότερη τιμή της ημέρας.
- **Vol.:** Αναφέρεται στον αριθμό των συναλλαγών της ημέρας.
- **T2M:** Αναφέρεται στην θερμοκρασία T2M της ημέρας.
- **T2M_MAX:** Αναφέρεται στην μέγιστη θερμοκρασία T2M της ημέρας.
- **T2M_MIN:** Αναφέρεται στην ελάχιστη θερμοκρασία T2M της ημέρας.
- **RH2M:** Αναφέρεται στην υγρασία RH2M της ημέρας.
- **ALLSKY_SFC_SW_DWN:** Αναφέρεται στην συσσώρευση επιφανειακής βραχυκυματικής ακτινοβολίας της ημέρας.
- **Date:** Η ημερομηνία της εγγραφής.

Τα χαρακτηριστικά αυτά αφαιρούνται καθώς είναι εκ των προτέρων άγνωστα για τις μέρες που γίνεται η πρόβλεψη.

4.4 Κανονικοποίηση Δεδομένων

Η κανονικοποίηση των δεδομένων είναι ένα θεμελιώδες βήμα προεπεξεργασίας στη μηχανική μάθηση και τη βαθιά μάθηση, με στόχο την τυποποίηση της κλίμακας των χαρακτηριστικών σε ένα σύνολο δεδομένων. Σε πολλούς αλγορίθμους μάθησης, δεδομένα με διαφορετικά μεγέθη, μονάδες ή εύρος μπορούν να επηρεάσουν δυσανάλογα τη διαδικασία εκπαίδευσης, οδηγώντας συχνά σε μη βέλτιστα αποτελέσματα. Για παράδειγμα, τα χαρακτηριστικά με μεγαλύτερα μεγέθη μπορεί να κυριαρχούν σε εκείνα με μικρότερες κλίμακες, με αποτέλεσμα το μοντέλο να μεροληπτεί προς αυτά. Η κανονικοποίηση των δεδομένων μετριάζει αυτές τις αποκλίσεις, διασφαλίζοντας ότι κάθε χαρακτηριστικό συμβάλλει εξίσου στη διαδικασία μάθησης. Αυτό είναι ιδιαίτερα σημαντικό στη βαθιά μάθηση, όπου οι μέθοδοι βελτιστοποίησης που βασίζονται στην κλίση, όπως η οπίσθια διάδοση, μπορούν να συγκλίνουν ταχύτερα και πιο αξιόπιστα σε κανονικοποιημένα δεδομένα.

Μια δημοφιλής τεχνική κανονικοποίησης είναι η κανονικοποίηση MinMax. Σε αυτή τη μέθοδο, τα δεδομένα μετασχηματίζονται γραμμικά ώστε να ταιριάζουν σε ένα επιθυμητό εύρος, συνήθως $[0,1]$. Ο τύπος δίνεται από τη σχέση:

$$\text{scaled value} = \frac{\text{value} - \text{min}}{\text{max} - \text{min}} \quad (4.1)$$

Κεφάλαιο 5

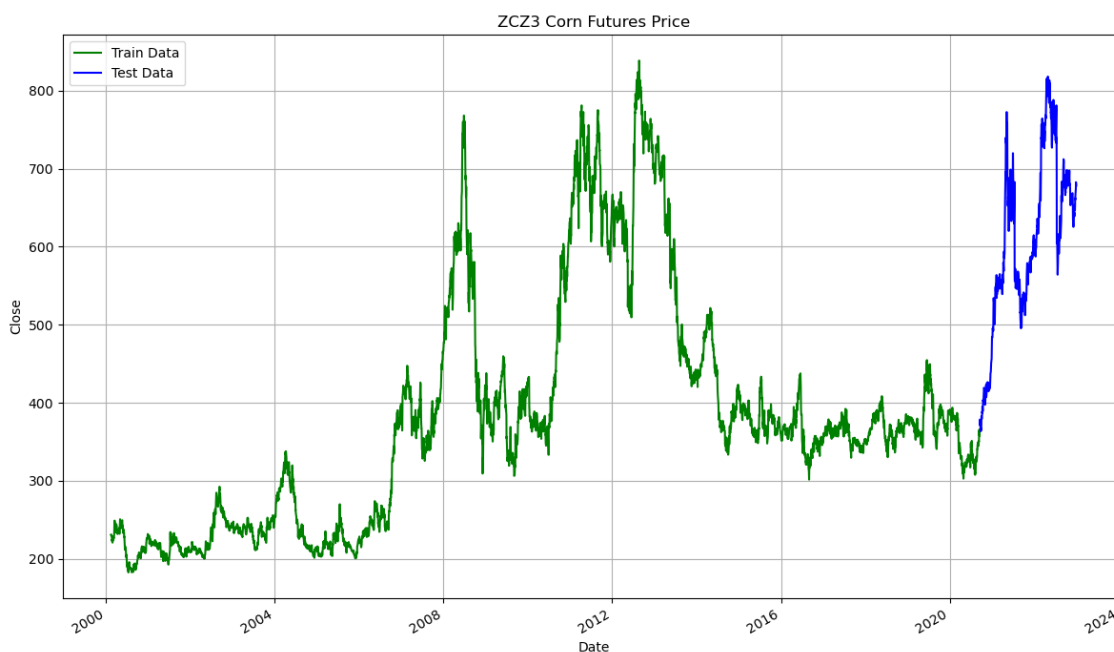
Αρχιτεκτονικές Μοντέλων

5.1 Εισαγωγή

Στο κεφάλαιο αυτό θα αναλυθούν οι αρχιτεκτονικές που μελετήθηκαν και χρησιμοποιήθηκαν στην πειραματική διαδικασία. Οι αρχιτεκτονικές αυτές περιλαμβάνουν την απλούστερη αρχιτεκτονική **LSTM** (vanilla LSTM) και στην συνέχεια επεκτείνονται στις αρχιτεκτονικές **stacked LSTM**, **bidirectional LSTM**, **CNN-LSTM**, την αρχιτεκτονική **Dual Input Attention LSTM** και την αρχιτεκτονική **Transformer**. Πριν αναλυθούν οι αρχιτεκτονικές, θα παρουσιαστεί η διαδικασία ενσωμάτωσης των δεδομένων.

5.2 Ενσωμάτωση των Δεδομένων

Για την ενσωμάτωση των δεδομένων στις αρχιτεκτονικές που μελετήθηκαν τα δεδομένα μετασχηματίστηκαν σε κατάλληλη μορφή ώστε να είναι διαχειρίσιμα από τις αρχιτεκτονικές. Αρχικά, τα δεδομένα χωρίστηκαν στα σύνολα εκπαίδευσης (train set) και ελέγχου (test set). Το πρώτο σύνολο αφορά το 90% των δεδομένων και περιλαμβάνει εγγραφές από τον Φεβρουάριο του 2000 έως τον Αύγουστο του 2020, ενώ το σύνολο ελέγχου αποτελεί το 10% και περιλαμβάνει εγγραφές από τον Σεπτέμβριο του 2020 έως τον Δεκέμβριο του 2022 (βλέπε Σχήμα 5.1).



Σχήμα 5.1 Χωρισμός δεδομένων σε σύνολο εκπαίδευσης και ελέγχου

Οι αρχιτεκτονικές μηχανικής μάθησης που χρησιμοποιήθηκαν δέχονται ως είσοδο μεμονωμένα στιγμιότυπα του συνόλου δεδομένων, οπότε για την ενσωμάτωσή τους στις αρχιτεκτονικές τα σύνολα δεν έγινε κάποιος μετασχηματισμός. Η είσοδος θα είναι:

$$input = X_i,$$

Όπου X_i , είναι το διάνυσμα χαρακτηριστικών της ημέρας i , του συνόλου αναφοράς. Ενώ η έξοδος τους μπορεί να είναι μια τιμή ή ακολουθία τιμών, ανάλογα αν θα χρησιμοποιηθεί για την πρόβλεψη μιας ή πολλών τιμών.

$$output = y_i,$$

όπου y_i είναι η μεταβλητή στόχος της ημέρας i του συνόλου αναφοράς

Οι αρχιτεκτονικές βαθιάς μάθησης που χρησιμοποιήθηκαν δέχονται ως είσοδο ακολουθίες δεδομένων (time sequences), οι ακολουθίες αυτές αποτελούνται από 1 ή περισσότερες χρονικές στιγμές (time steps). Στα δεδομένα μελέτης κάθε χρονική στιγμή είναι μια ημέρα. Τα δεδομένα χωρίστηκαν σε ακολουθίες που περιλαμβάνουν n , χρονικές στιγμές, με κάθε ακολουθία έχει την ακόλουθη μορφή:

$$input_sequence_1 = [X_1, X_2, \dots, X_n]$$

$$input_sequence_2 = [X_2, X_3, \dots, X_{n+1}]$$

Όπου X_i , είναι το διάνυσμα χαρακτηριστικών της ημέρας i , του συνόλου αναφοράς. Οι ακολουθίες εισόδου θα είναι επικαλυπτόμενες, κατά αυτό το τρόπο γίνεται εμπλουτισμός των δεδομένων (data augmentation) αυξάνοντας τον αριθμό δειγμάτων εκπαίδευσης καθώς και δειγμάτων ελέγχου. Αντίστοιχα πρέπει να γίνει προσαρμογή της μεταβλητής στόχου (target variable), ώστε τα μοντέλα να δέχονται κάθε φορά τα σωστά ζεύγη ακολουθιών και μεταβλητών στόχων, για να εξασφαλιστεί η σωστή εκπαίδευση τους. Η μεταβλητή στόχος θα δίνεται αντίστοιχα σε ακολουθίες, ανάλογα με το παράθυρο πρόβλεψης της αρχιτεκτονικής μελέτης. Για πρόβλεψη της επόμενης χρονικής στιγμής θα δίνεται μια μόνο χρονική στιγμή της μεταβλητής στόχος y_i , ενώ για πρόβλεψη k χρονικών στιγμών θα δίνεται μια ακολουθία χρονικών στιγμών της μεταβλητής στόχος. Οι ακολουθίες αυτές θα έχουν την παρακάτω μορφή:

$$output_sequence_1 = [y_{n+1}, y_{n+2}, \dots, y_{n+k}]$$

$$output_sequence_2 = [y_{n+2}, y_{n+3}, \dots, y_{n+k+1}],$$

Οι ακολουθίες της μεταβλητής στόχου ξεκινάνε από την $n + 1$ εγγραφή, καθώς στα μοντέλα πρόβλεψης δεν θα πρέπει να υπάρχει άμεση συσχέτιση της μεταβλητής στόχος με τις ανεξάρτητες μεταβλητές, για αυτό για την πρόβλεψη χρησιμοποιούνται παρελθοντικά δεδομένα.

Τα ζεύγη εισόδου-εξόδου θα είναι:

$$\begin{aligned} \text{input_sequence}_1 &= [X_1, X_2, \dots, X_n] \\ \text{output_sequence}_1 &= [y_{n+1}, y_{n+2}, \dots, y_{n+k}] \end{aligned}$$

5.3 Παρουσίαση Αρχιτεκτονικών

5.3.1 Linear Regression

Στην συγκεκριμένη αρχιτεκτονική προβλέπεται η τιμή του παραγώγου της επόμενης μέρα χρησιμοποιώντας τις μεταβλητές που αναλύθηκαν στο Κεφάλαιο 4 και επηρεάζουν τις τιμές των γεωργικών προϊόντων. Η είσοδος που θα δίνεται κάθε φορά αφορά την ημέρα για την οποία θα γίνει η πρόβλεψη και θα είναι:

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$$

Ενώ η έξοδος αφορά την πρόβλεψη της επόμενης ημέρας:

$$\widehat{y_{t+1}} = f(x_t)$$

Όπου f είναι η γραμμική συνάρτηση.

5.3.2 XGBoost

Ομοίως με την αρχιτεκτονική Linear Regression η αρχιτεκτονική, που χρησιμοποιεί το XGBoost, προβλέπει την τιμή του παραγώγου της επόμενης ημέρας χρησιμοποιώντας τις μεταβλητές που αναλύθηκαν στο Κεφάλαιο 4.

Η είσοδος που θα δίνεται κάθε φορά αφορά την ημέρα για την οποία θα γίνει η πρόβλεψη και θα είναι:

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$$

Ενώ η έξοδος αφορά την πρόβλεψη της επόμενης ημέρας:

$$\widehat{y_{t+1}} = XGB(x_t)$$

Όπου XGB είναι το μοντέλο XGBoost.

5.3.3 Vanilla LSTM

Η πρώτη αρχιτεκτονική αναδρομικού δικτύου που χρησιμοποιήθηκε είναι αυτή του Vanilla LSTM. Στην συγκεκριμένη αρχιτεκτονική προβλέπεται η τιμή του παραγώγου της επόμενης ημέρας ή των επόμενων ημερών χρησιμοποιώντας τις μεταβλητές που αναλύθηκαν στο κεφάλαιο 4. Οι n μεταβλητές αυτές μπορούν να εκφραστούν ως εξής:

$$I = (x^1, x^2, \dots, x^n),$$

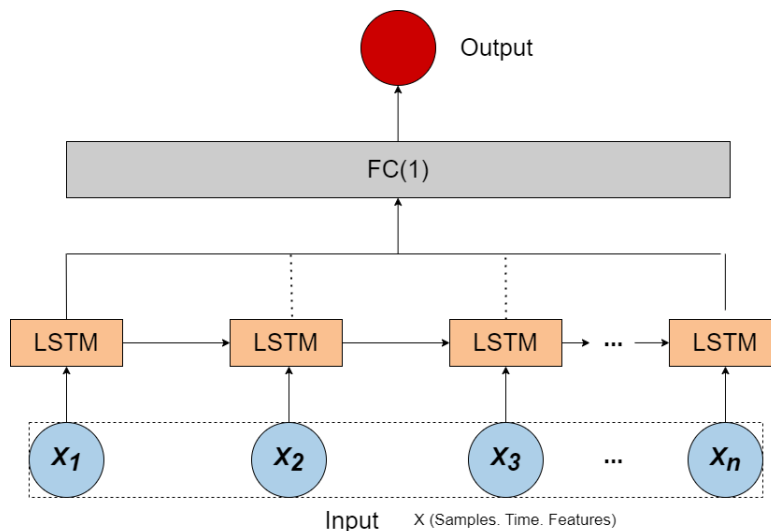
όπου x_n είναι η χρονοσειρά για την $n^{\text{οστη}}$ μεταβλητή που επηρεάζει την τιμή. Δηλαδή, $X^T = (x_1, x_2, \dots, x_T) \in R^{n \times T}$

όπου T είναι το μήκος του χρονικού βήματος (ή το μέγεθος του παραθύρου). Δηλαδή, η τιμή του επόμενου μήνα προβλέπεται χρησιμοποιώντας δεδομένα από το παρελθόν T

ημέρες. Η $k^{\text{οστη}}$ μεταβλητή εισόδου της οποίας το χρονικό βήμα είναι T

εκφράζεται ως $x^k = (x_1^k, x_2^k, \dots, x_T^k)^T \in R^T$, και οι τιμές των n μεταβλητών εισόδου τη χρονική στιγμή t εκφράζονται ως:

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$$



Σχήμα 5.2 Αρχιτεκτονική Vanilla LSTM

Η αρχιτεκτονική Vanilla LSTM (βλέπε Σχήμα 5.2) αποτελείται από το επίπεδο εισόδου από το οποίο εισάγονται τα ακολουθιακά δεδομένα, στην συνέχεια δίνονται σε ένα στρώμα LSTM που περιέχει Z αριθμό μονάδων LSTM και τέλος οι έξοδοι δίνονται σε ένα συνελικτικό επίπεδο. Για την πρόβλεψη, ο αριθμός των νευρώνων στο τελευταίο

πλήρως συνδεδεμένο στρώμα ορίζεται σε 1 για πρόβλεψη μιας τιμής, ενώ σε n για πρόβλεψη n τιμών.

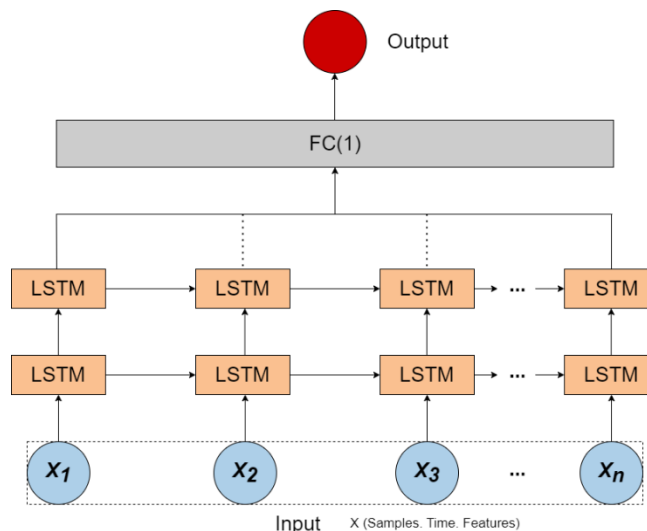
5.3.4 Stacked LSTM

Η αρχιτεκτονική stacked LSTM είναι επέκταση της αρχιτεκτονικής Vanilla-LSTM με την προσθήκη πρόσθετων LSTM επιπέδων. Στην συγκεκριμένη αρχιτεκτονική προβλέπει την τιμή του παραγώγου την επόμενη ημέρα ή επόμενων ημερών ημερών χρησιμοποιώντας τις μεταβλητές που αναλύθηκαν στο κεφάλαιο 4 και επηρεάζουν τις τιμές των γεωργικών προϊόντων. Οι n μεταβλητές που επηρεάζουν την τιμή των γεωργικών προϊόντων μπορούν να εκφραστούν ως εξής:

$$I = (x^1, x^2, \dots, x^n),$$

όπου x_n είναι η χρονοσειρά για την $n^{\text{οστη}}$ μεταβλητή που επηρεάζει την τιμή. Δηλαδή, $X^T = (x_1, x_2, \dots, x_T) \in R^{n \times T}$ όπου T είναι το μήκος του χρονικού βήματος (ή το μέγεθος του παραθύρου). Δηλαδή, η τιμή του επόμενου μήνα προβλέπεται χρησιμοποιώντας δεδομένα από το παρελθόν T ημέρες. Η $k^{\text{οστη}}$ μεταβλητή εισόδου της οποίας το χρονικό βήμα είναι T εκφράζεται ως $x^k = (x_1^k, x_2^k, \dots, x_T^k)^T \in R^T$, και οι τιμές των n μεταβλητών εισόδου τη χρονική στιγμή t εκφράζονται ως:

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$$



Σχήμα 5.3 Αρχιτεκτονική Stacked LSTM

Η αρχιτεκτονική Stacked-LSTM αποτελείται από το επίπεδο εισόδου από το οποίο εισάγονται τα ακολουθιακά δεδομένα, στην συνέχεια ακολουθεί μια σειρά από στρώματα LSTM που περιέχει z αριθμό μονάδων LSTM και στο τέλος οι έξοδοι δίνονται σε ένα συνελικτικό επίπεδο. Για την πρόβλεψη της τελικής ενιαίας πραγματικής τιμής, ο αριθμός των νευρώνων στο τελευταίο πλήρως συνδεδεμένο στρώμα ορίζεται σε 1 για πρόβλεψη μιας τιμής ενώ n για πρόβλεψη n τιμών.

5.3.5 Bidirectional LSTM

Η αρχιτεκτονική Bidirectional LSTM, που συνήθως αναφέρεται ως BiLSTM, επεκτείνει την παραδοσιακή αρχιτεκτονική LSTM εισάγοντας ένα δεύτερο στρώμα LSTM που επεξεργάζεται τις ακολουθίες εισόδου με την αντίστροφη σειρά. Αυτή η διπλή κατευθυντικότητα επιτρέπει στο μοντέλο να συλλαμβάνει πληροφορίες τόσο για το παρελθόν (ακολουθία προς τα εμπρός) όσο και για το μέλλον (ακολουθία προς τα πίσω). Στην συγκεκριμένη αρχιτεκτονική προβλέπεται η τιμή του παραγώγου της επόμενης ημέρα ή επόμενων ημερών χρησιμοποιώντας τις μεταβλητές που αναλύθηκαν στο κεφάλαιο 4 και επηρεάζουν τις τιμές των γεωργικών προϊόντων. Οι n μεταβλητές που επηρεάζουν την τιμή των γεωργικών προϊόντων μπορούν να εκφραστούν ως εξής:

$$I = (x^1, x^2, \dots, x^n),$$

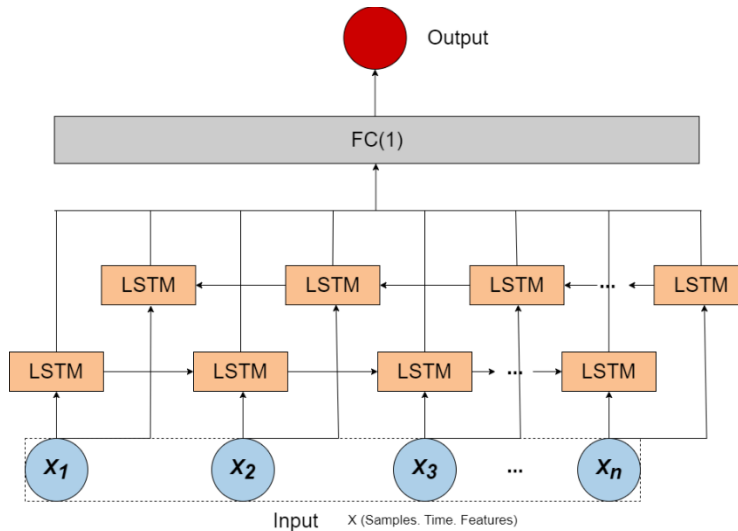
όπου x_n είναι η χρονοσειρά για την $n^{\text{οστη}}$ μεταβλητή που επηρεάζει την τιμή. Δηλαδή, $X^T = (x_1, x_2, \dots, x_T) \in R^{n \times T}$

όπου T είναι το μήκος του χρονικού βήματος (ή το μέγεθος του παραθύρου). Δηλαδή, η τιμή του επόμενου μήνα προβλέπεται χρησιμοποιώντας δεδομένα από το παρελθόν T

ημέρες. Η $k^{\text{οστη}}$ μεταβλητή εισόδου της οποίας το χρονικό βήμα είναι T

εκφράζεται ως $x^k = (x_1^k, x_2^k, \dots, x_T^k)^T \in R^T$, και οι τιμές των n μεταβλητών εισόδου τη χρονική στιγμή t εκφράζονται ως:

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$$



Σχήμα 5.4 Αρχιτεκτονική Bidirectional LSTM

Η αρχιτεκτονική Bidirectional-LSTM αποτελείται από το επίπεδο εισόδου από το οποίο εισάγονται τα ακολουθιακά δεδομένα, στην συνέχεια ακολουθούν δυο στρώματα LSTM στο πρώτο τροφοδοτείται η ακολουθία με την σειρά εισόδου ενώ στο δεύτερο η σειρά τροφοδοτείται με την αντίστροφη σειρά. Οι έξοδοι των δυο στρωμάτων LSTM δίνονται ως είσοδο σε ένα πλήρως συνδεδεμένο επίπεδο. Για την πρόβλεψη της τελικής ενιαίας πραγματικής τιμής, ο αριθμός των νευρώνων στο τελευταίο πλήρως συνδεδεμένο στρώμα ορίζεται σε 1 για πρόβλεψη μιας τιμής ενώ n για πρόβλεψη n τιμών.

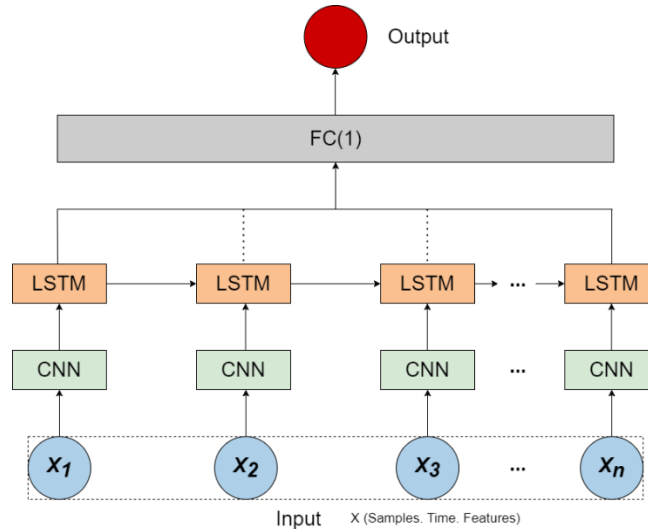
5.3.6 CNN-LSTM

Η αρχιτεκτονική CNN-LSTM είναι μια ενσωμάτωση των συνελικτικών νευρωνικών δικτύων (CNN) και των δικτύων LSTM. Αυτός ο συνδυασμός αξιοποιεί τις δυνατότητες εξαγωγής χωρικών χαρακτηριστικών των CNN με τις δυνατότητες χρονικής μοντελοποίησης των LSTM, καθιστώντας τον ιδιαίτερα ισχυρό για εφαρμογές που περιλαμβάνουν ακολουθίες με μεγάλο πλήθος χωρικών χαρακτηριστικών, όπως η πρόβλεψη χρονοσειρών με πολλές εισόδους. Στην συγκεκριμένη αρχιτεκτονική προβλέπει την τιμή του παράγωγου την επόμενη ημέρα ή επόμενων ημερών χρησιμοποιώντας τις μεταβλητές που αναλύθηκαν στο κεφάλαιο 4 και επηρεάζουν τις τιμές των γεωργικών προϊόντων. Οι n μεταβλητές που επηρεάζουν την τιμή των γεωργικών προϊόντων μπορούν να εκφραστούν ως εξής:

$$I = (x^1, x^2, \dots, x^n),$$

όπου x_n είναι η χρονοσειρά για την $n^{\text{οστη}}$ μεταβλητή που επηρεάζει την τιμή. Δηλαδή, $X^T = (x_1, x_2, \dots, x_T) \in R^{n \times T}$ όπου T είναι το μήκος του χρονικού βήματος (ή το μέγεθος του παραθύρου). Δηλαδή, η τιμή του επόμενου μήνα προβλέπεται χρησιμοποιώντας δεδομένα από το παρελθόν T ημέρες. Η $k^{\text{οστη}}$ μεταβλητή εισόδου της οποίας το χρονικό βήμα είναι T εκφράζεται ως $x^k = (x_1^k, x_2^k, \dots, x_T^k)^T \in R^T$, και οι τιμές των n μεταβλητών εισόδου τη χρονική στιγμή t εκφράζονται ως:

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$$



Σχήμα 5.5 Αρχιτεκτονική CNN-LSTM

Η αρχιτεκτονική CNN-LSTM αποτελείται από το επίπεδο εισόδου από το οποίο εισάγονται τα ακολουθιακά δεδομένα, στην συνέχεια ακολουθεί ένα σειρά από στρώματα LSTM που περιέχει z αριθμό μονάδων LSTM και στην τέλος οι έξοδοι δίνονται σε ένα συνελικτικό επίπεδο. Για την πρόβλεψη της τελικής ενιαίας πραγματικής τιμής, ο αριθμός των νευρώνων στο τελευταίο πλήρως συνδεδεμένο στρώμα ορίζεται σε 1 για πρόβλεψη μιας τιμής ενώ n για πρόβλεψη n τιμών.

5.3.7 Dual Input Attention LSTM

Η αρχιτεκτονική (DIA-LSTM)[2] προβλέπει την τιμή της επόμενης μέρας ή επόμενων ημερών χρησιμοποιώντας τις μεταβλητές που αναλύθηκαν στο κεφάλαιο 4 και επηρεάζουν τις τιμές των γεωργικών προϊόντων. Οι n μεταβλητές που επηρεάζουν την τιμή των γεωργικών προϊόντων μπορούν να εκφραστούν ως εξής:

$$I = (x^1, x^2, \dots, x^n),$$

όπου x_n είναι η χρονοσειρά για την $n^{\text{οστη}}$ μεταβλητή που επηρεάζει την τιμή. Δηλαδή, $X^T = (x_1, x_2, \dots, x_T) \in R^{n \times T}$

όπου T είναι το μήκος του χρονικού βήματος (ή το μέγεθος του παραθύρου). Δηλαδή, η τιμή του επόμενου μήνα προβλέπεται χρησιμοποιώντας δεδομένα από το παρελθόν T

ημέρες. Η $k^{\text{οστη}}$ μεταβλητή εισόδου της οποίας το χρονικό βήμα είναι T

εικφράζεται ως $x^k = (x_1^k, x_2^k, \dots, x_T^k)^T \in R^T$, και οι τιμές των n μεταβλητών εισόδου τη χρονική στιγμή t εκφράζονται ως:

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)^T \in R^n$$

Η αρχιτεκτονική DIA-LSTM χρησιμοποιεί την τιμή των συμβολαίων των γεωργικών προϊόντων του παρελθόντος, $y = (y_1, y_2, \dots, y_T)$, όπου $y_t \in R$ και παρελθοντικές τιμές του n μεταβλητών εισόδου, (x_1, x_2, \dots, x_T) , με $x_t \in R^n$, για την πρόβλεψη της τιμής του επόμενου χρονικού βήματος. Για παράδειγμα, την επόμενη μέρα, θα είναι $y_{T+1} \in \mathbb{R}$. Αυτό εκφράζεται στην εξίσωση (1), όπου \mathcal{M} εκφράζει την προτεινόμενη DIA-LSTM:

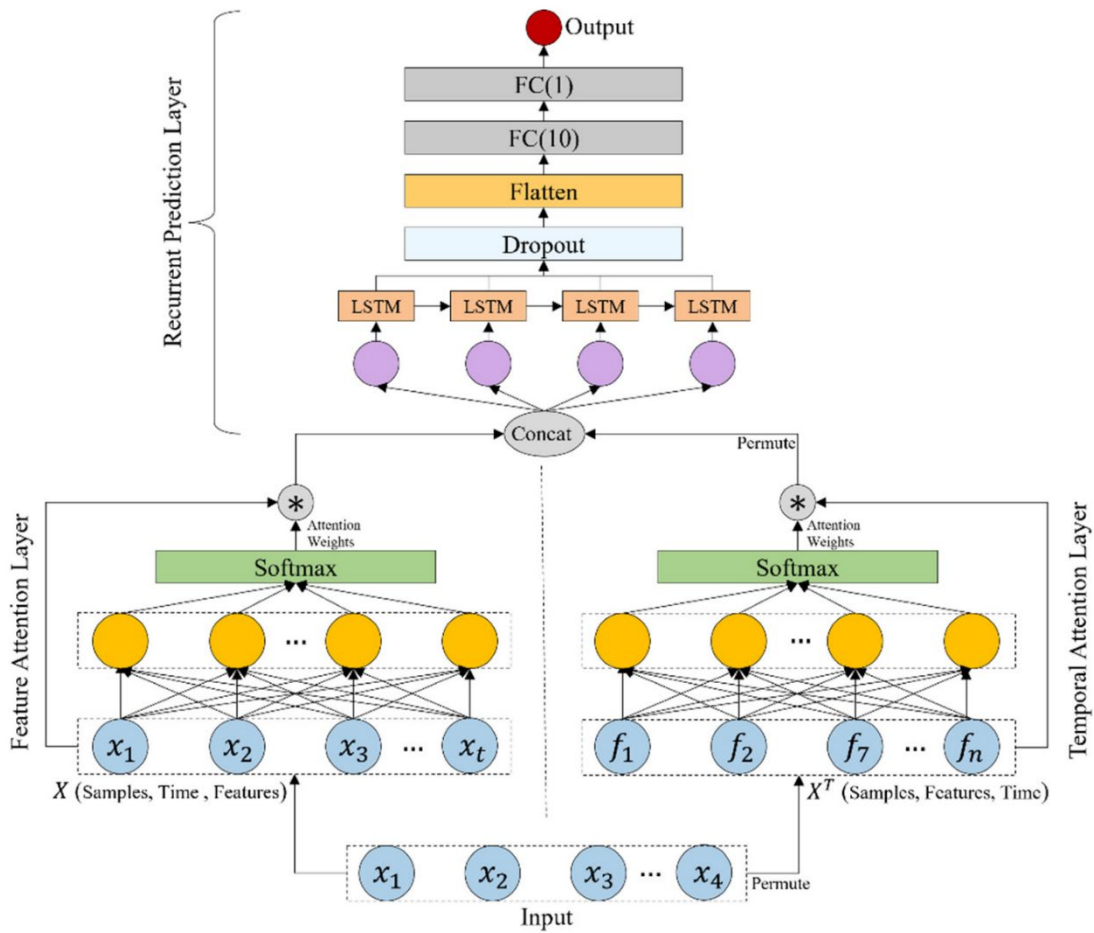
$$\hat{y}_{T+1} = M(x_1, x_2, \dots, x_T, y_1, y_2, \dots, y_T).$$

Το DIA-LSTM αποτελείται από ένα στρώμα προσοχής χαρακτηριστικών (feature attention layer), στρώματα χρονικής προσοχής (temporal attention layers) και ένα επαναλαμβανόμενο στρώμα πρόβλεψης (recurrent prediction layer). Η δομή του παρουσιάζεται στο παρακάτω σχήμα. Το στρώμα προσοχής χαρακτηριστικών μαθαίνει τη συσχέτιση χαρακτηριστικών στα δεδομένα εισόδου $X = (x_1, x_2, \dots, x_t)$, ενώ το στρώμα χρονικής προσοχής μοντελοποιεί τη χρονική σχέση με βάση τα μετατοπισμένα δεδομένα εισόδου $X^T = (f_1, f_2, \dots, f_n)$. Η έξοδος του στρώματος χαρακτηριστικών και χρονικής προσοχής παράγεται κάνοντας έναν στοιχειομετρικό (element wise) πολλαπλασιασμό (συμβολίζεται με $*$) των βαρών προσοχής με τα δεδομένα εισόδου. Το στρώμα επαναλαμβανόμενης πρόβλεψης προβλέπει την τελική τιμή του αποτελέσματος χρησιμοποιώντας τη συνδυασμένη τιμή της εξόδου από τα στρώματα χαρακτηριστικής και χρονικής προσοχής.

Το στρώμα προσοχής χαρακτηριστικών και το στρώμα χρονικής προσοχής υλοποιήθηκαν απλοποιώντας το perceptron ενός στρώματος. Αυτό ήταν εμπνευσμένο από τον μηχανισμό αυτοπροσοχής (self-attention) που μπορεί να κατασκευάσει την προσοχή (attention) χρησιμοποιώντας μόνο τιμές εισόδου. Τα βάρη προσοχής εφαρμόστηκαν σε κάθε μεταβλητή εισόδου στο στρώμα προσοχής χαρακτηριστικών, καθώς και σε κάθε χρονικό βήμα στο στρώμα χρονικής προσοχής.

Το στρώμα επαναλαμβανόμενης πρόβλεψης αποτελείται από ένα στρώμα stateful LSTM ενός στρώματος και δύο πλήρως συνδεδεμένα στρώματα (συμβολίζονται με FC

στο Σχήμα 5.6). Το μοντέλο stateful LSTM σημαίνει ότι η κρυφή κατάσταση h_t που μαθαίνεται στο τρέχον χρονικό βήμα μεταφέρεται στην αρχική κατάσταση κατά την επόμενη μάθηση. Το μοντέλο LSTM λαμβάνει τη συνένωση (concatenation) της εξόδου X_f του επιπέδου προσοχής χαρακτηριστικών και της εξόδου X_t του στρώματος χρονικής προσοχής. Στη συνέχεια, εφαρμόζεται dropout στην έξοδο του LSTM και, μετά την ισοπέδωση (συμβολίζεται με Flatten στο Σχήμα 5.6), εισάγεται στο πλήρως συνδεδεμένο στρώμα (FC). Στον Πίνακα 5.1 παρουσιάζονται οι ρυθμίσεις των υπερπαραμέτρων που χρησιμοποιούνται σε κάθε στρώμα. Για την πρόβλεψη της τελικής ενιαίας πραγματικής τιμής, ο αριθμός των νευρώνων στο τελευταίο πλήρως συνδεδεμένο στρώμα ορίζεται σε 1 για πρόβλεψη μιας τιμής ενώ n για πρόβλεψη n τιμών.



Σχήμα 5.6 Αρχιτεκτονική DLA-LSTM

Layer Name	Parameter Name	Value
LSTM	Unit size	6
	Activation function	Tanh
	Stateful	True
Dropout	Dropout rate	0.2
Fully connected	Number of neurons in the 1st FC layer	10
	Activation functions in the 1st FC layer	None
	Number of neurons in the 2nd FC layer	1
	Activation function in the 2nd FC layer	None

Πίνακας 5.1 Χαρακτηριστικά αρχιτεκτονικής DLA-LSTM

Κεφάλαιο 6

Πειραματική Διαδικασία

6.1 Εισαγωγή

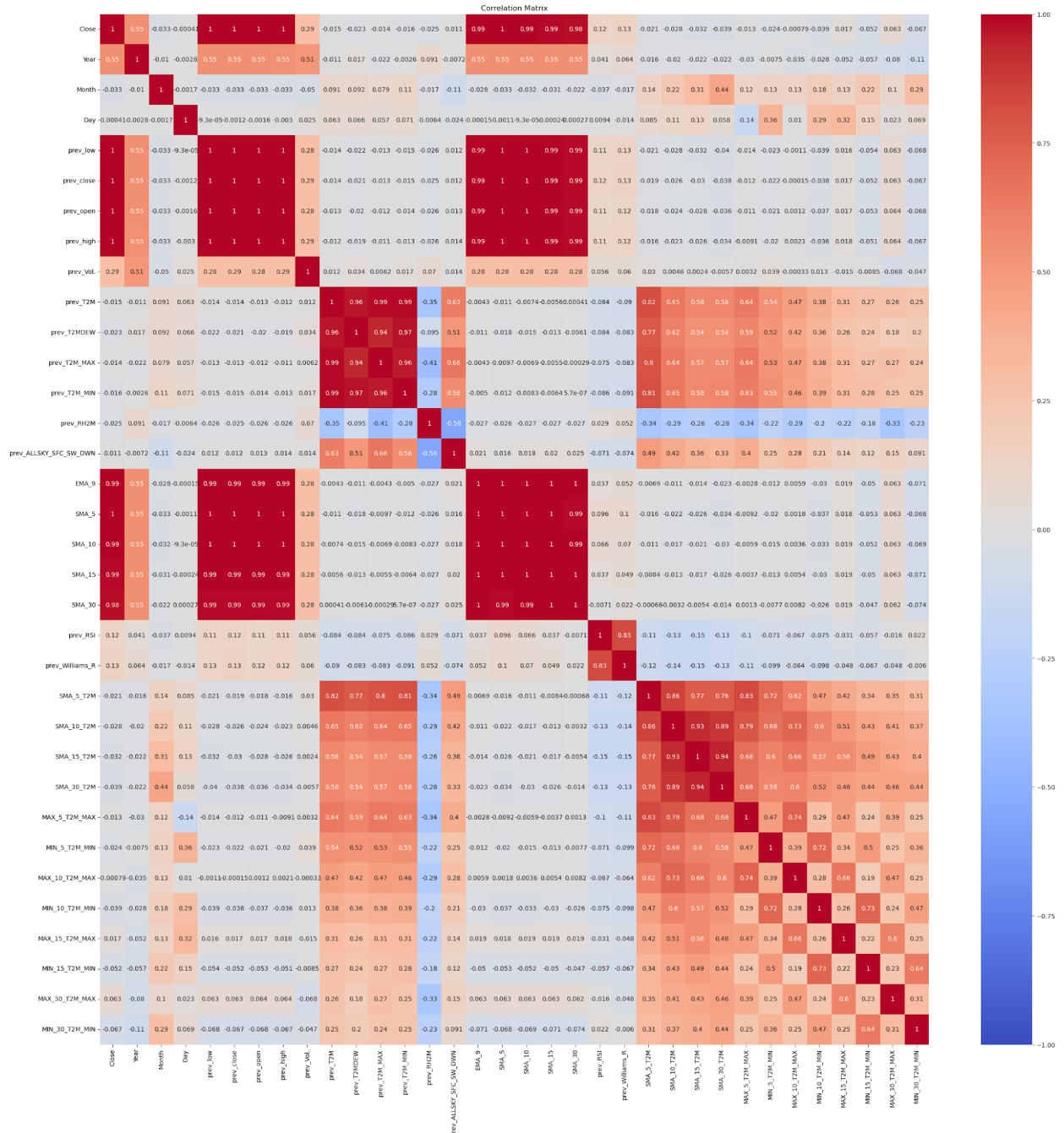
Στο κεφάλαιο αυτό θα αναλυθούν τα πειράματα που διεξήχθησαν για την πρόβλεψη τιμών επόμενης ημέρας καθώς και την πρόβλεψη των τιμών των n επόμενων ημερών. Αρχικά, θα αναλυθούν τα πειράματα με χρήση των αρχιτεκτονικών που αναλύθηκαν στο κεφάλαιο 5 για την πρόβλεψη της επόμενης ημέρας. Έπειτα, θα χρησιμοποιηθούν αρχιτεκτονικές που δίνουν την δυνατότητα πρόβλεψης των n επόμενων ημερών.

6.2 Πρόβλεψη Τιμής Επόμενης Ημέρας

Για την πρόβλεψη της τιμής των προθεσμιακών συμβολαίων της επόμενης ημέρας αρχικά θα μελετηθούν οι δυο αρχιτεκτονικές μηχανικής μάθησης **Linear Regression** και **XGBoost**, ενώ στην συνέχεια θα μελετηθούν οι αρχιτεκτονικές αναδρομικών νευρωνικών δικτύων **Vanilla LSTM**, **Stacked LSTM**, **Bidirectional LSTM**, **CNN-LSTM**, **Dual Input Attention LSTM** (οι αρχιτεκτονικές παρουσιάζονται αναλυτικά στο Κεφάλαιο 5).

6.2.1 Linear Regression

Η αρχιτεκτονική **Linear Regression** χρησιμοποιείται για την εύρεση της βέλτιστης γραμμικής συνάρτησης προσαρμόζοντας την στα δεδομένα εκπαίδευσης με εφαρμογή της μεθόδου της γραμμικής παλινδρόμησης. Αρχικά, θα υπολογιστεί η συσχέτιση των χαρακτηριστικών του συνόλου δεδομένων χρησιμοποιώντας την μέθοδο Pearson. Παρατηρούνται (Σχήμα 6.1) ισχυρές συσχετίσεις της μεταβλητής στόχος ή εξαρτημένης μεταβλητής close με τις μεταβλητές prev_open, prev_close, prev_low, prev_high, EMA_9, SMA_5, SMA_10, SMA_15, SMA_30, ενώ ακολουθεί με μικρότερη συσχέτιση η μεταβλητή Year και η μεταβλητή prev_Vol. Οι υπόλοιπες μεταβλητές έχουν χαμηλή συσχέτιση.



Σχήμα 6.1 Πίνακας συσχέτισης μεταβλητών

Θα διατηρηθούν όλες οι μεταβλητές και θα εφαρμοστεί ο αλγόριθμος της γραμμική παλινδρόμησης προσαρμόζοντας ένα πολλαπλό γραμμικό μοντέλο στα δεδομένα εκπαίδευσης. Στόχος της μεθόδου είναι η ελαχιστοποίηση του αθροίσματος των τετραγωνικών σφαλμάτων.

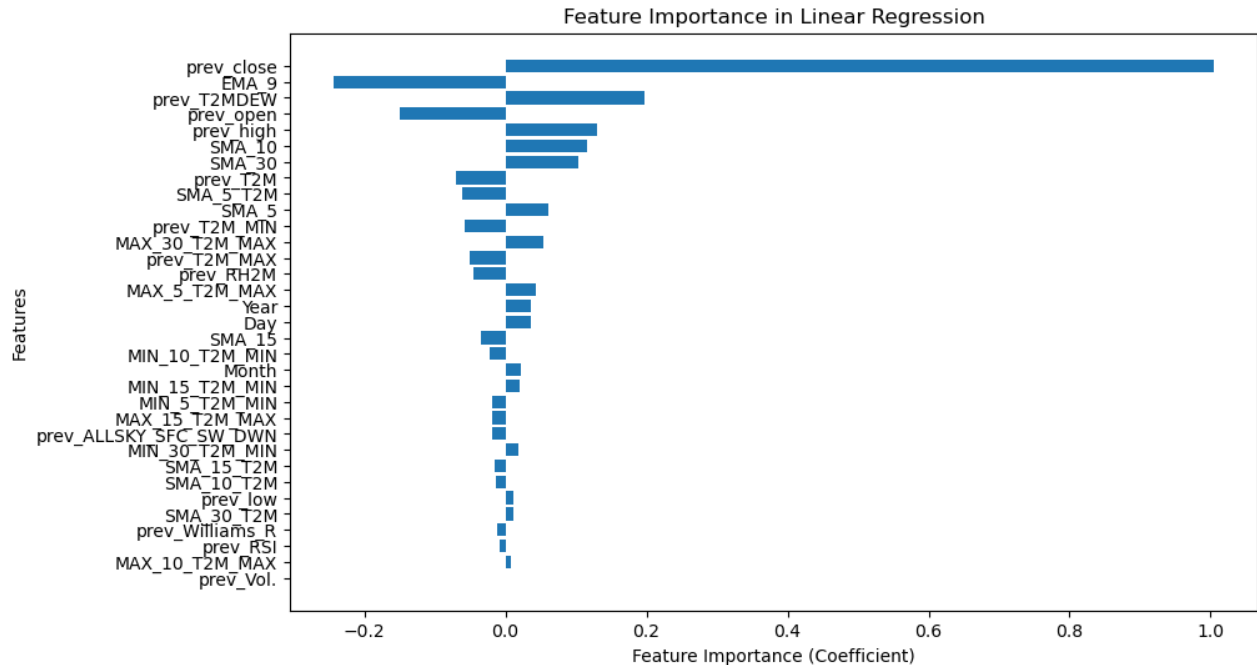
$$J(\beta) = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik}))^2$$

Όπου:

- $J(\beta)$ είναι η συνάρτηση κόστους, Άθροισμα Τετραγωνικών Σφαλμάτων.
- n είναι ο αριθμός των παρατηρήσεων του συνόλου δεδομένων.
- y_i είναι η πραγματική τιμή της εξαρτημένης μεταβλητής για την $i^{\text{οστη}}$ παρατήρηση.
- x_{ij} είναι η τιμή της j ανεξάρτητης μεταβλητής για την $i^{\text{οστη}}$ παρατήρηση.
- β_j είναι ο συντελεστής της j ανεξάρτητης μεταβλητής.

Για την προσαρμογή του μοντέλου στα δεδομένα χρησιμοποιήθηκε η συνάρτηση `LinearRegression()` της βιβλιοθήκης `sklearn` της Python. Το μοντέλο που προκύπτει με την προσαρμογή στα δεδομένα είναι το ακόλουθο:

$$y = -70.1030 + (0.0363 * x_1) + (0.0215 * x_2) + (0.0358 * x_3) + (0.0121 * x_4) + (1.0041 * x_5) - (0.1499 * x_6) + (0.1294 * x_7) - (0.0000 * x_8) - (0.0704 * x_9) + (0.1976 * x_{10}) - (0.0514 * x_{11}) - (0.0579 * x_{12}) - (0.0445 * x_{13}) - (0.0189 * x_{14}) - (0.2430 * x_{15}) + (0.0607 * x_{16}) + (0.1154 * x_{17}) - (0.0353 * x_{18}) + (0.1028 * x_{19}) - (0.0083 * x_{20}) - (0.0111 * x_{21}) - (0.0608 * x_{22}) - (0.0140 * x_{23}) - (0.0149 * x_{24}) + (0.0118 * x_{25}) + (0.0431 * x_{26}) - (0.0191 * x_{27}) + (0.0073 * x_{28}) - (0.0221 * x_{29}) - (0.0191 * x_{30}) + (0.0207 * x_{31}) + (0.0533 * x_{32}) + (0.0186 * x_{33})$$



Σχήμα 6.2 Σημαντικότητα χαρακτηριστικών για γραμμικό μοντέλο

Στο μοντέλο γραμμικής παλινδρόμησης που προκύπτει παρατηρείται ότι η μεταβλητή με την μεγαλύτερη επιρροή (Σχήμα 6.2) είναι η τιμή κλεισίματος της

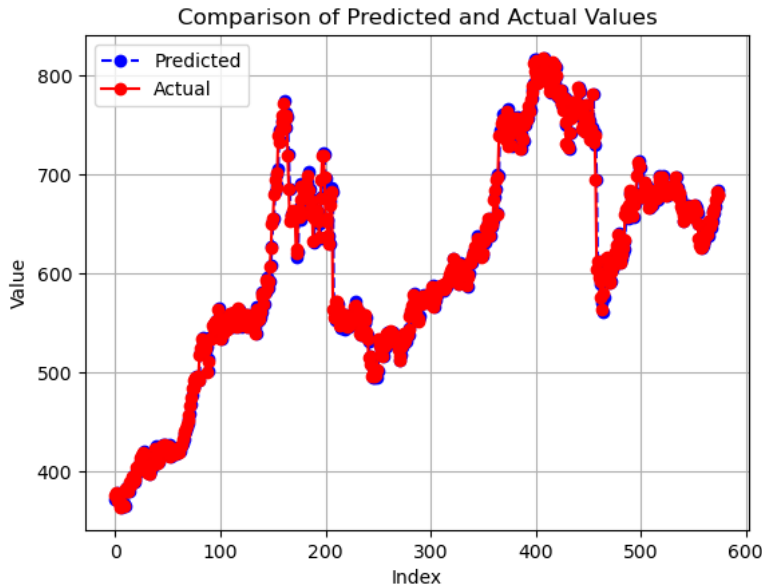
προηγούμενης ημέρας $prev_close$, η οποία είχε υψηλή συσχέτιση με την μεταβλητή στόχος, ενώ ακολουθεί η μεταβλητή EMA_9 , η μεταβλητή $T2MDEW$ η οποία δεν είχε υψηλή συσχέτιση, είναι τρίτη σε σημαντικότητα. Ακολουθούν οι $prev_open$, $prev_high$, SMA_10 , SMA_30 όπου παρουσίαζαν υψηλή συσχέτιση. Στις 10 μεταβλητές με την μεγαλύτερη επιρροή στο μοντέλο είναι και οι μεταβλητές $prev_T2M$ και SMA_5_T2M αν και παρουσίαζαν χαμηλή συσχέτιση, παρατηρείται ότι η μεταβλητή $prev_Vol$, η οποία περιέχει τον αριθμό των συναλλαγών, δεν συμμετέχει στην πρόβλεψη. Είναι φανερό η επιρροή των τηλεσκοπικών δεδομένων στο γραμμικό μοντέλο που προκύπτει, καθώς 2 από αυτές βρίσκονται στις 10 μεταβλητές με την μεγαλύτερη σημαντικότητα.

Για την αξιολόγηση της απόδοσης του μοντέλου θα χρησιμοποιηθεί το σύνολο ελέγχου και οι μετρικές αξιολόγησης RMSE, MSE, MAE, MAPE. Οι τιμές των μετρικών αξιολόγησης παρουσιάζονται στον παρακάτω πίνακα:

Μετρική	Τιμή
MSE	173.8564513016124
MAE	8.667592893647761
RMSE	13.18546363620227
MAPE	0.013962770326628725

Πίνακας 6.1 Μετρικές linear regression

Το συγκεκριμένο μοντέλο παρουσιάζει καλή προσέγγιση του συνόλου ελέγχου το οποίο φαίνεται και από τις τιμές των μετρικών αξιολόγησης. Στο Σχήμα 6.3 παρουσιάζονται οι πραγματικές τιμές των προθεσμιακών συμβολαίων σε σχέση με τις τιμές πρόβλεψης.

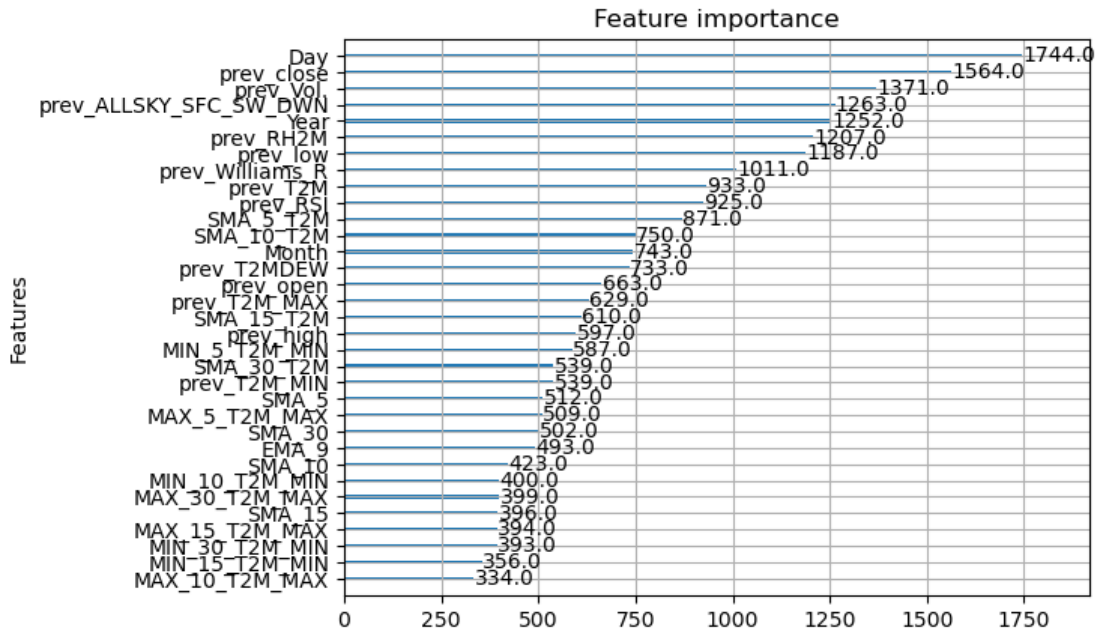


Σχήμα 6.3 Προβλέψεις γραμμικού μοντέλου

6.2.2 XGBoost

Η αρχιτεκτονική **XGBoost** θα χρησιμοποιηθεί για την εύρεση της τιμής του προθεσμιακού συμβολαίου της επόμενης ημέρας. Για την προσαρμογή του μοντέλου XGBoost στα δεδομένα εκπαίδευσης θα χρησιμοποιηθεί η βιβλιοθήκη xgboost της Python. Ομοίως με τη γραμμική παλινδρόμηση στόχος είναι η ελαχιστοποίηση του αθροίσματος τετραγωνικών σφαλμάτων. Για την εύρεση του βέλτιστου μοντέλου θα χρησιμοποιηθεί η τεχνική της αναζήτησης πλέγματος (grid search) αναζητώντας τις βέλτιστες παραμέτρους. Οι παράμετροι στις οποίες θα εφαρμοστεί η αναζήτηση πλέγματος είναι ο αριθμός των εκτιμητών δέντρων που θα δημιουργηθούν (`n_estimators`), ο ρυθμός μάθησης (`learning rate`) το βήμα με το οποίο θα προσεγγίζεται ένα ελάχιστο από την συνάρτηση κόστους, το μέγιστο βάθος (`max_depth`) των δέντρων που θα δημιουργηθούν και η παράμετρος γάμα (`gamma`) αποτελεί μια παράμετρο κανονικοποίησης και καθορίζει έναν όρο ρύθμισης της συνάρτησης κόστους. Στην αναζήτηση πλέγματος για το `n_estimators` έγινε αναζήτηση μεταξύ των τιμών [100, 200, 300, 400], για το `learning rate` μεταξύ των [0.001, 0.005, 0.01, 0.05, 0.1] για το `max_depth` μεταξύ των [1,2,3,4,5,6,7,8,9,10,11,12,13,14,15] και για το `gamma` μεταξύ των [0.0001, 0.001, 0.005, 0.01, 0.02]. Οι παράμετροι που επιλέχθηκαν είναι για `learning_rate` 0.05, για `n_estimators` 200, για `max_depth` 8 και για το `gamma` 0.001. Παρακάτω στο Σχήμα

παρουσιάζεται ο πίνακας (Πίνακας 6.4) με την σημαντικότητα των χαρακτηριστικών για το τελικό μοντέλο:



Σχήμα 6.4 Σημαντικότητα χαρακτηριστικών για XGBoost μοντέλο

Στην σημαντικότητα των μεταβλητών στο Σχήμα 6.4 παρατηρείται ότι πρώτη είναι η ημέρα της πρόβλεψης Day η οποία δεν είχε υψηλή γραμμική συσχέτιση και δεν ήταν στις σημαντικότερες μεταβλητές του γραμμικού μοντέλου, στην συνέχεια ακολουθεί η προηγούμενη τιμή κλεισίματος του προθεσμιακού συμβολαίου το οποίο συμφωνεί με την γραμμική συσχέτιση και με το προηγούμενο μοντέλο. Παρατηρείται και σε αυτή την αρχιτεκτονική ότι υπάρχει σημαντική επιρροή από τα τηλεσκοπικά δεδομένα στο μοντέλο πρόβλεψης, όπως και στην περίπτωση της γραμμικής παλινδρόμησης.

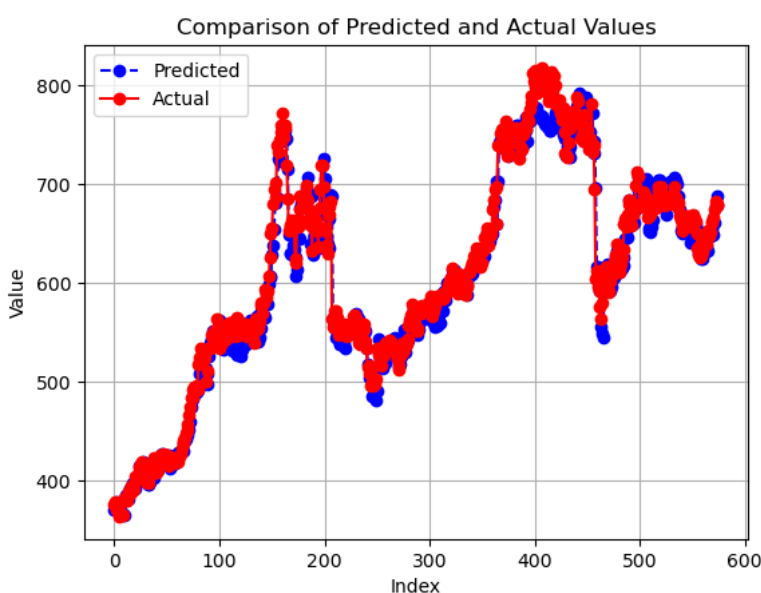
Για την αξιολόγηση της απόδοσης του μοντέλου θα χρησιμοποιηθεί το σύνολο ελέγχου και οι μετρικές αξιολόγησης RMSE, MSE, MAE, MAPE. Οι τιμές των μετρικών αξιολόγησης στο σύνολο ελέγχου παρουσιάζονται στον παρακάτω πίνακα:

Μετρική	Τιμή
MSE	314.0564552326549
MAE	12.433651547639267

RMSE	17.72163805162082
MAPE	0.0196675903769575 3

Πίνακας 6.2 Μετρικές XGBoost

Παρατηρείται ότι οι τιμές των μετριοών για την αρχιτεκτονική XGBoost είναι υψηλότερες σε σχέση με αυτές της Linear Regression, το οποίο μεταφράζεται σε χειρότερη προσαρμογή στο σύνολο ελέγχου και επομένως χειρότερες προβλέψεις. Αυτό φαίνεται και από το Σχήμα 6.5 όπου οι προβλέψεις δεν προσεγγίζουν τόσο καλά τις πραγματικές τιμές σε σχέση με το Σχήμα της γραμμική παλινδρόμηση.

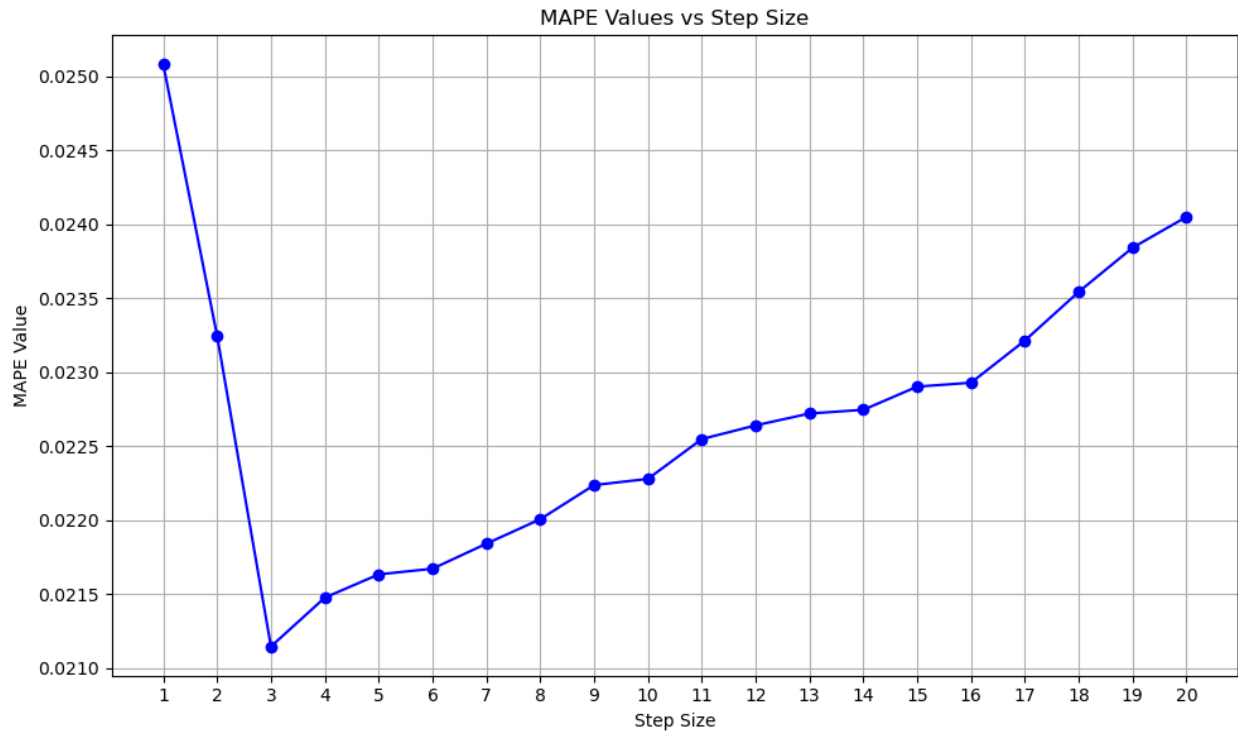


Σχήμα 6.5 Προβλέψεις XGBoost

6.2.2 Vanilla LSTM

Από τις αρχιτεκτονικές νευρωνικών δικτύων η πρώτη που μελετήθηκε είναι η απλούστερη αρχιτεκτονική αναδρομικού νευρωνικού δικτύου **Vanilla LSTM**. Στην εκπαίδευση αυτού του νευρωνικού δικτύου έπειτα από χρήση της μεθόδου αναζήτησης πλέγματος επιλέχθηκε ο αριθμός των νευρώνων του αναδρομικού στρώματος LSTM ίσος με 100 και αριθμός δέσμης παρτίδας (batch size) ίσος με 64. Ως αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο Adam και το μοντέλο εκπαιδεύτηκε για 100 εποχές (epochs). Στην πρόβλεψη τιμών με χρονοσειρές, σημαντικό είναι να προσδιοριστεί η παράμετρος του χρονικού βήματος ή παραθύρου (time step), των παρελθοντικών δειγμάτων που θα λάβει υπόψιν το νευρωνικό δίκτυο για την πρόβλεψη μελλοντικών τιμών. Για το σκοπό αυτό χρησιμοποιήθηκε η μέθοδος αναζήτησης πλέγματος για την εύρεση

της βέλτιστης τιμής του χρονικού παραθύρου που ελαχιστοποιεί τις μετρικές αξιολόγησης. Δόθηκαν οι τιμές $t \in \{1,2,3, \dots, 20\}$. Παρακάτω παρουσιάζεται η απόδοση του μοντέλου βάσει της μετρικής MAE σε σχέση με το μήκος παραθύρου:



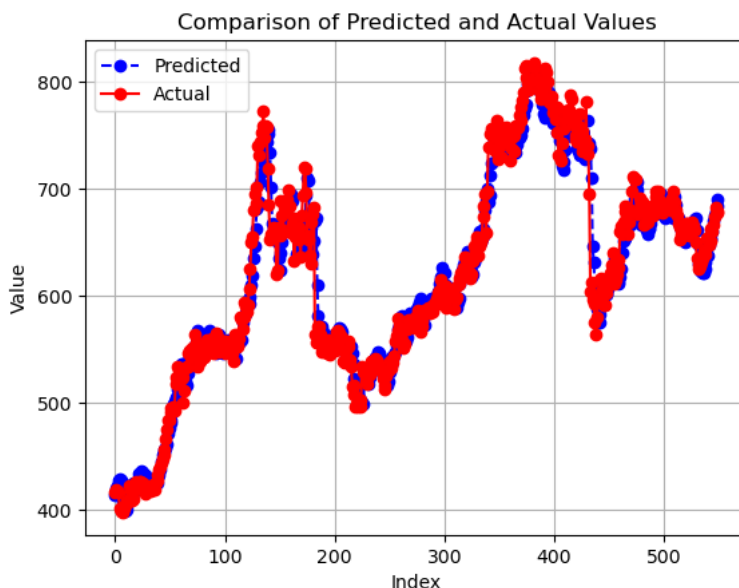
Σχήμα 6.6 MAPE σε σχέση με μήκος παραθύρου για Vanilla LSTM

Από το Σχήμα 6.6 παρατηρείται ότι το ελάχιστο MAPE καταγράφεται στο μοντέλο που χρησιμοποιεί μήκος παραθύρου $t = 3$. Από την καμπύλη του Σχήματος φαίνεται ότι μετά από αυτό το μήκος παραθύρου η απόδοση του μοντέλου μειώνεται. Οι μετρικές αξιολόγησης στο σύνολο ελέγχου παρουσιάζονται στον παρακάτω πίνακα:

Μετρική	Τιμή
MSE	378.5843705619372
MAE	13.005053508552166
RMSE	19.457244680630843
MAPE	0.02114495684069417 3

Πίνακας 6.3 Μετρικές Vanilla LSTM

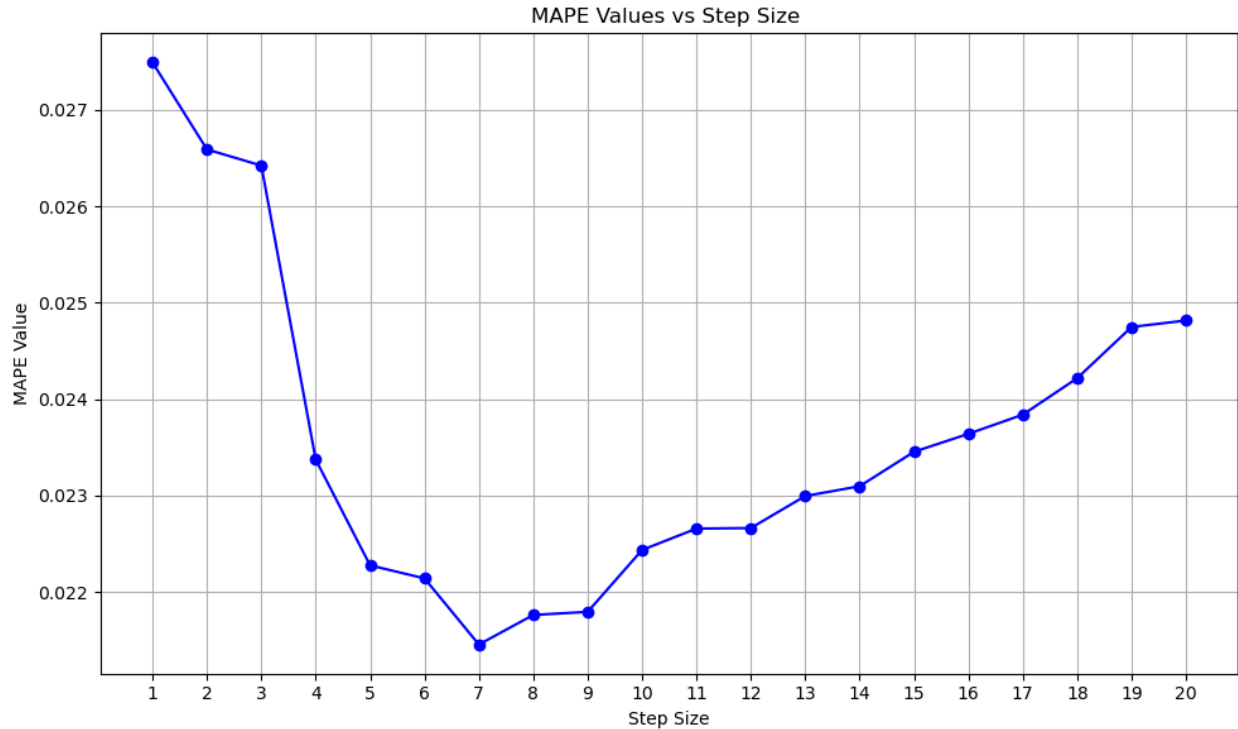
Οι τιμές των μετρικών είναι υψηλότερες σε σχέση με τα προηγούμενα μοντέλα, επομένως η απόδοση του μοντέλου είναι χειρότερη σε σχέση με τα μοντέλα που εξετάστηκαν προηγουμένως. Στο Σχήμα 6.7 παρουσιάζονται οι προβλεπόμενες τιμές σε σχέση με τις πραγματικές.



Σχήμα 6.7 Προβλέψεις Vanilla LSTM

6.2.3 Stacked LSTM

Στην αρχιτεκτονική **stacked LSTM** επεκτείνει την αρχιτεκτονική Vanilla LSTM προσθέτοντας ένα ακόμα αναδρομικό επίπεδο με των ίδιο αριθμό μονάδων LSTM με την αρχιτεκτονική Vanilla LSTM. Ως αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο Adam και το μοντέλο εκπαιδεύτηκε για 100 εποχές (epochs). Για την υπερπαραμέτρο του μήκους παραθύρου χρησιμοποιήθηκε η μέθοδος αναζήτησης πλέγματος για την εύρεση της βέλτιστης τιμής του χρονικού παραθύρου που ελαχιστοποιεί τις μετρικές αξιολόγησης. Δόθηκαν οι τιμές $t \in \{1,2,3, \dots, 20\}$. Παρακάτω παρουσιάζεται η απόδοση του μοντέλου βάσει της μετρικής MAE σε σχέση με το μήκος παραθύρου:



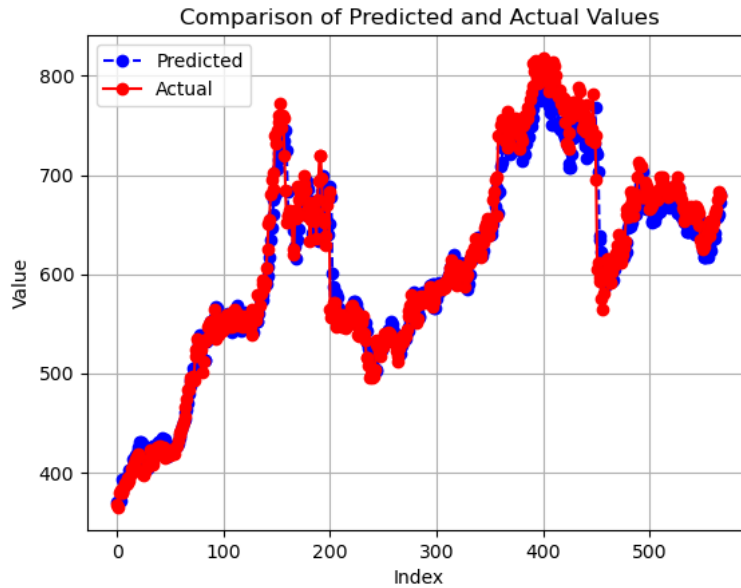
Σχήμα 6.8 MAPE σε σχέση με μήκος παραθύρου για stacked LSTM

Από το Σχήμα 6.8 φαίνεται ότι το ελάχιστο MAPE καταγράφεται από το μοντέλο που χρησιμοποιεί μήκος παραθύρου $t = 7$. Ομοίως με το μοντέλο Vanilla LSTM παρουσιάζεται μείωση της απόδοσης της αρχιτεκτονικής μετά από ένα μήκος παραθύρου. Οι μετρικές αξιολόγησης στο σύνολο ελέγχου παρουσιάζονται στον παρακάτω πίνακα:

Μετρική	Τιμή
MSE	397.06925145935537
MAE	13.230284625490855
RMSE	19.92659658495036
MAPE	0.02145596685234409 8

Πίνακας 6.4 Μετρικές stacked LSTM

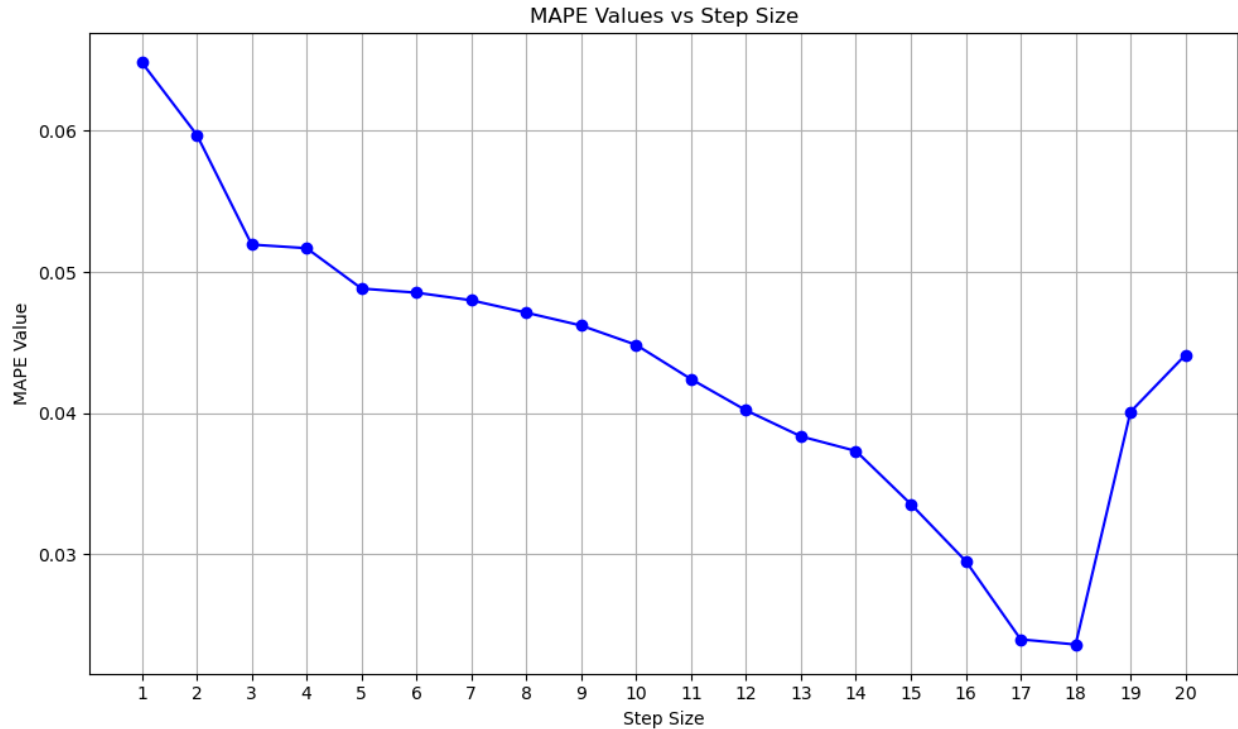
Από τον Πίνακα παρατηρείται ότι η αρχιτεκτονική stacked LSTM παρουσιάζει χαμηλότερη απόδοση σε σχέση με την αρχιτεκτονική Vanilla LSTM. Στο Σχήμα 6.9 παρουσιάζονται οι προβλεπόμενες τιμές σε σχέση με τις πραγματικές.



Σχήμα 6.9 Προβλέψεις *stacked LSTM*

6.2.4 Bidirectional LSTM

Η αρχιτεκτονική **bidirectional LSTM** προσθέτει ένα ακόμα αναδρομικό επίπεδο στην αρχιτεκτονική Vanilla LSTM με 100 νευρώνες, που δέχεται ως είσοδο τις ακολουθίες με την αντίστροφη σειρά. Ως αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο Adam και το μοντέλο εκπαιδεύτηκε για 100 εποχές (epochs). Για την υπερπαραμέτρο του μήκους παραθύρου χρησιμοποιήθηκε η μέθοδος αναζήτησης πλέγματος για την εύρεση της βέλτιστης τιμής του χρονικού παραθύρου που ελαχιστοποιεί τις μετρικές αξιολόγησης. Δόθηκαν οι τιμές $t \in \{1,2,3, \dots, 20\}$. Παρακάτω παρουσιάζεται η απόδοση του μοντέλου βάσει της μετρικής MAE σε σχέση με το μήκος παραθύρου:



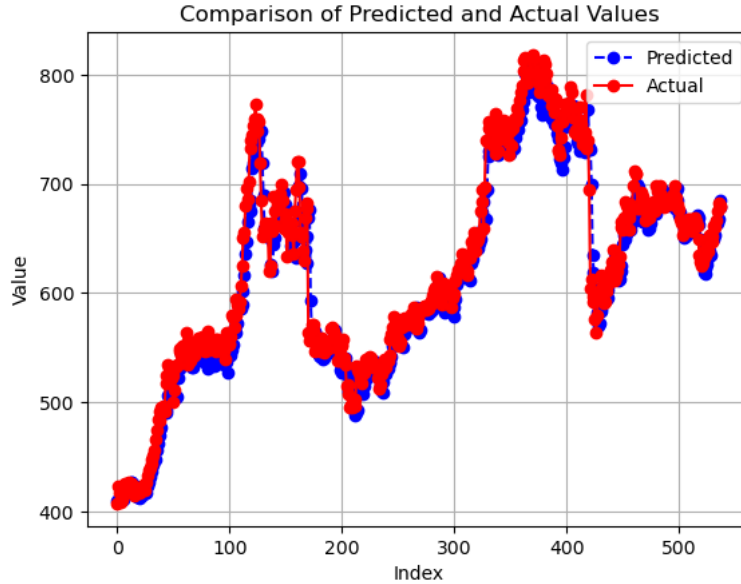
Σχήμα 6.10 MAPE σε σχέση με μήκος παραθύρου για *bidirectional LSTM*

Το ελάχιστο MAPE καταγράφεται στο μοντέλο που χρησιμοποιεί μήκος παραθύρου $t = 18$. Για μήκη παραθύρου $t = 17, 18$ καλύτερη απόδοση της αρχιτεκτονικής ενώ για τιμές μεγαλύτερες η απόδοση μειώνεται. Οι μετρικές αξιολόγησης στο σύνολο ελέγχου παρουσιάζονται στον παρακάτω πίνακα:

Μετρική	Τιμή
MSE	458.7370879871646
MAE	14.684382120768229
RMSE	21.4181485658113
MAPE	0.02360304312061616 7

Πίνακας 6.5 Μετρικές *bidirectional LSTM*

Η αρχιτεκτονική stacked LSTM παρουσιάζει χαμηλότερη απόδοση σε σχέση με τις προηγούμενες αρχιτεκτονικές. Στο Σχήμα 6.11 παρουσιάζονται οι προβλεπόμενες τιμές σε σχέση με τις πραγματικές.



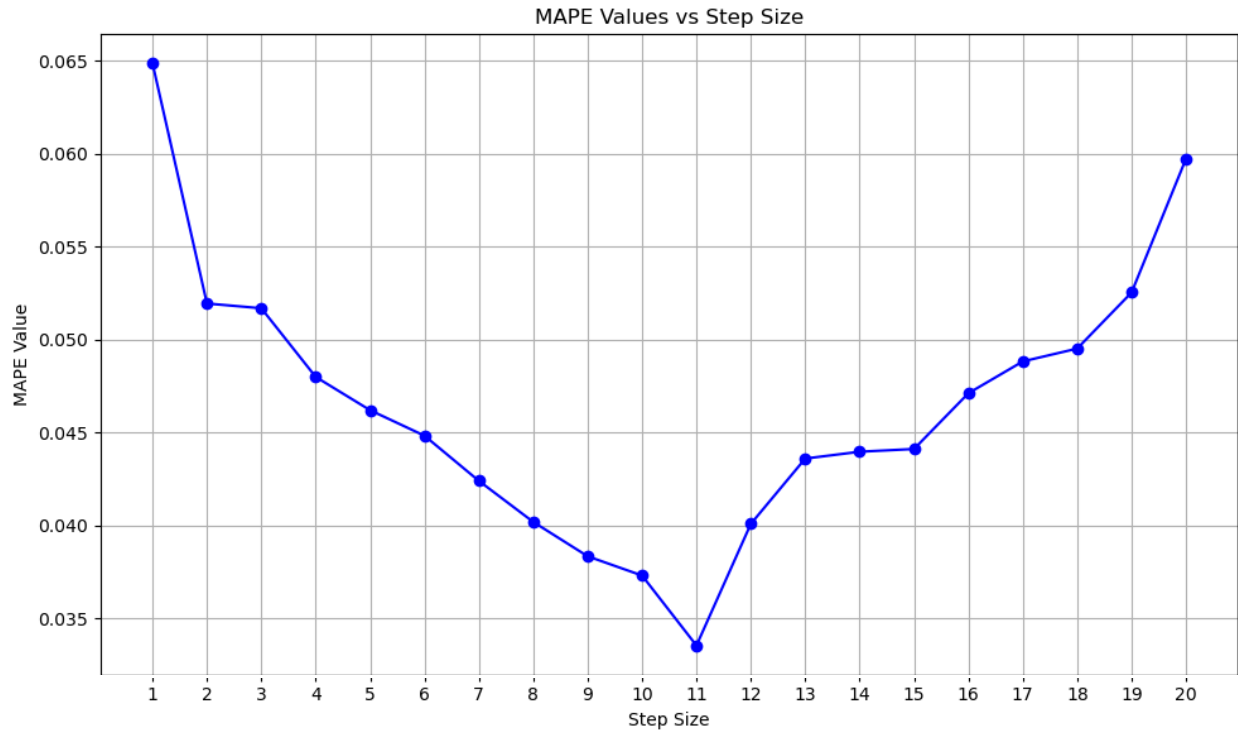
Σχήμα 6.11 Προβλέψεις *bidirectional LSTM*

6.2.5 CNN-LSTM

Στην αρχιτεκτονική **CNN-LSTM** προστίθεται ένα συνελικτικό επίπεδο στην αρχιτεκτονική *Vanilla LSTM*, η οποία χρησιμοποιείται για την εξαγωγή χαρακτηριστικών μειώνοντας την διάσταση των χαρακτηριστικών που εισάγονταν αρχικά στο *Vanilla LSTM*. Ως αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο *Adam* και το μοντέλο εκπαιδεύτηκε για 100 εποχές (epochs). Για την υπερπαραμέτρο του μήκους παραθύρου χρησιμοποιήθηκε η μέθοδος αναζήτησης πλέγματος για την εύρεση της βέλτιστης τιμής του χρονικού παραθύρου που ελαχιστοποιεί τις μετρικές αξιολόγησης. Δόθηκαν οι τιμές

$t \in \{1,2,3, \dots, 20\}$. Στο Σχήμα 6.12 παρουσιάζεται η απόδοση του μοντέλου βάσει της μετρικής MAE σε σχέση με το μήκος παραθύρου.

Το ελάχιστο MAPE καταγράφεται στο μοντέλο που χρησιμοποιεί μήκος παραθύρου $t = 11$. Ομοίως με τις προηγούμενες αρχιτεκτονικές παρατηρείται μείωση

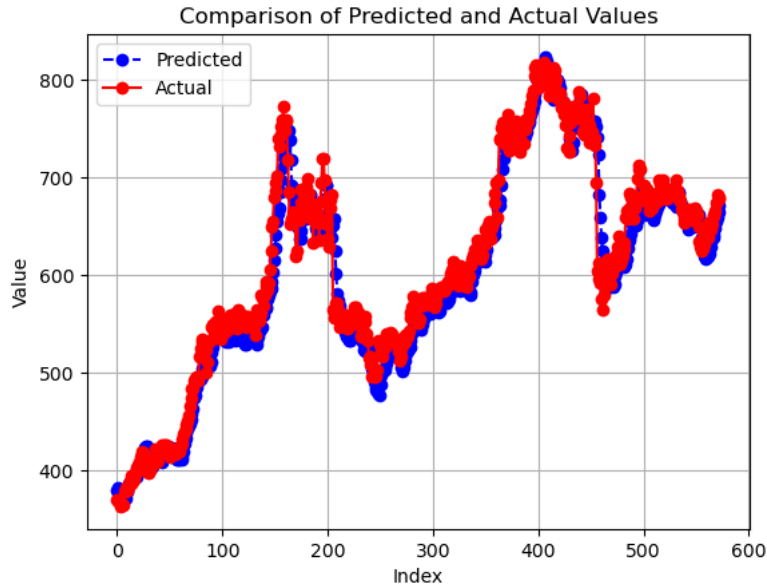


Σχήμα 6.12 MAPE σε σχέση με μήκος παραθύρου για CNN-LSTM

της απόδοσης μετά από το βέλτιστο μήκος παραθύρου. Οι μετρικές αξιολόγησης στο σύνολο ελέγχου παρουσιάζονται στον παρακάτω πίνακα:

Μετρική	Τιμή
MSE	828.798728998074
MAE	14.684382120768229
RMSE	21.4181485658113
MAPE	0.03355269999332785 6

Πίνακας 6.6 Μετρικές CNN-LSTM

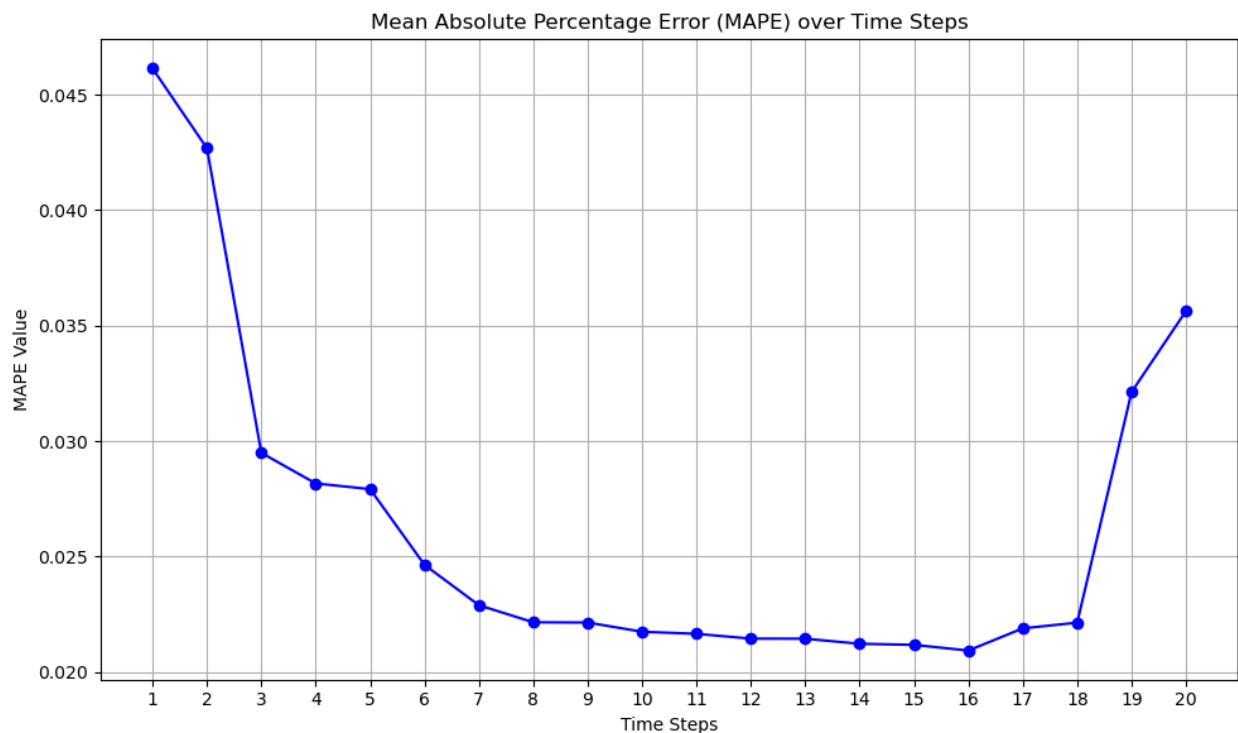


Σχήμα 6.13 Προβλέψεις CNN- LSTM

Η αρχιτεκτονική CNN-LSTM παρουσιάζει χαμηλότερη απόδοση σε σχέση με τις προηγούμενες αρχιτεκτονικές. Στο Σχήμα 6.13 παρουσιάζονται οι προβλεπόμενες τιμές σε σχέση με τις πραγματικές.

6.2.6 Dual Input Attention LSTM

Η αρχιτεκτονική **Dual Input Attention LSTM**, περιέχει 100 νευρώνες στο αναδρομικό επίπεδο και αναλύεται στο κεφάλαιο 5. Ως αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο Adam και το μοντέλο εκπαιδεύτηκε για 100 εποχές (epochs). Για την υπερπαράμετρο του μήκους παραθύρου χρησιμοποιήθηκε η μέθοδος αναζήτησης πλέγματος για την εύρεση της βέλτιστης τιμής του χρονικού παραθύρου που ελαχιστοποιεί τις μετρικές αξιολόγησης. Δόθηκαν οι τιμές $t \in \{1,2,3, \dots, 20\}$. Στο παρακάτω σχήμα (Σχήμα 6.14) παρουσιάζεται η απόδοση του μοντέλου βάσει της μετρικής MAE σε σχέση με το μήκος παραθύρου:



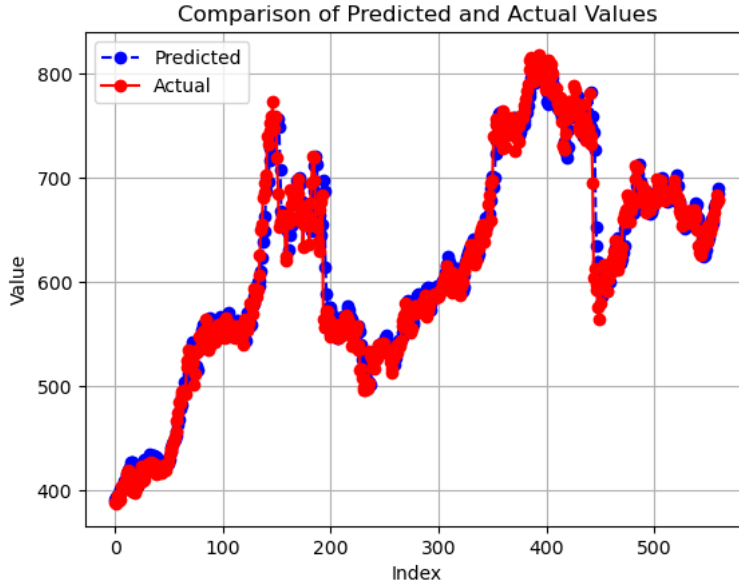
Σχήμα 6.14 MAPE σε σχέση με μήκος παραθύρου για DIA-LSTM

Το ελάχιστο MAPE καταγράφεται στο μοντέλο που χρησιμοποιεί μήκος παραθύρου $t = 16$. Στην συγκεκριμένη αρχιτεκτονική παρατηρείται ένα διάστημα τιμών παραθύρου, οι οποίες έχουν υψηλή απόδοση συγκεκριμένα στο διάστημα 7-18, βέβαια μετά από αυτό το διάστημα το σφάλμα αυξάνεται. Οι μετρικές αξιολόγησης στο σύνολο ελέγχου παρουσιάζονται στον παρακάτω πίνακα:

Μετρική	Τιμή
MSE	379.51176660826223
MAE	12.993800354003904
RMSE	19.481061742324577
MAPE	0.0209339327740532 3

Πίνακας 6.7 Μετρικές DIA-LSTM

Η αρχιτεκτονική DIA-LSTM παρουσιάζει την καλύτερη επίδοση σε σχέση με τις προηγούμενες αρχιτεκτονικές LSTM, παραμένει όμως παραμένει χαμηλότερη από τις αρχιτεκτονικές Linear Regression και XGBoost.



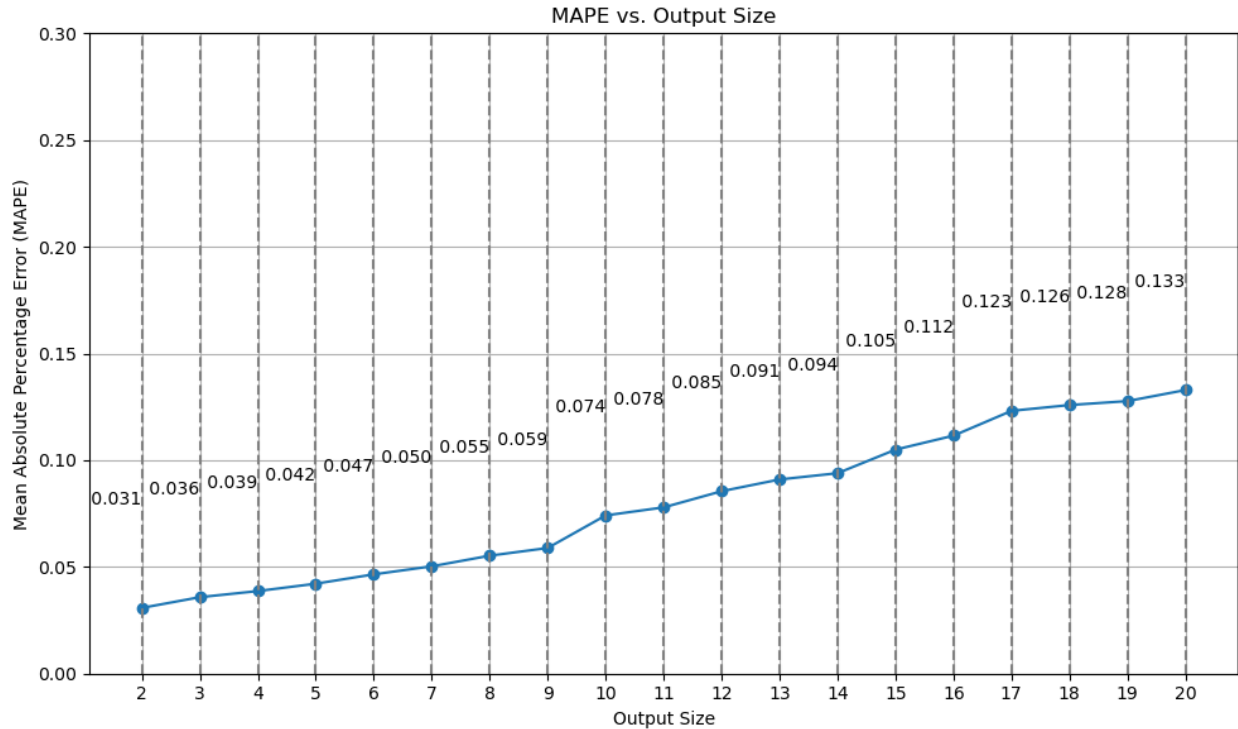
Σχήμα 6.15 Προβλέψεις DLA-LSTM

6.3 Πρόβλεψη Τιμών Επόμενων Ημερών

Στη ενότητα αυτή θα μελετηθεί η ικανότητα των αρχιτεκτονικών μηχανική μάθησης καθώς και η ικανότητα των αρχιτεκτονικών βαθιάς μάθησης να προβλέπουν τις τιμές των επόμενων n ημερών.

6.3.1 Linear Regression

Αρχικά θα εξεταστεί η δυνατότητα πρόβλεψης των τιμών προθεσμιακών συμβολαίων για τις n επόμενες μέρες με την αρχιτεκτονική Linear Regression. Το n θα παίρνει τιμές από 1 έως 20. Για κάθε n θα εκπαιδεύεται ένα μοντέλο γραμμικής παλινδρόμησης. Στο παρακάτω Σχήμα 6.16 παρουσιάζεται η απόδοση των μοντέλων ως προς την μετρική MAPE για $n \in \{1,2,3, \dots, 20\}$.

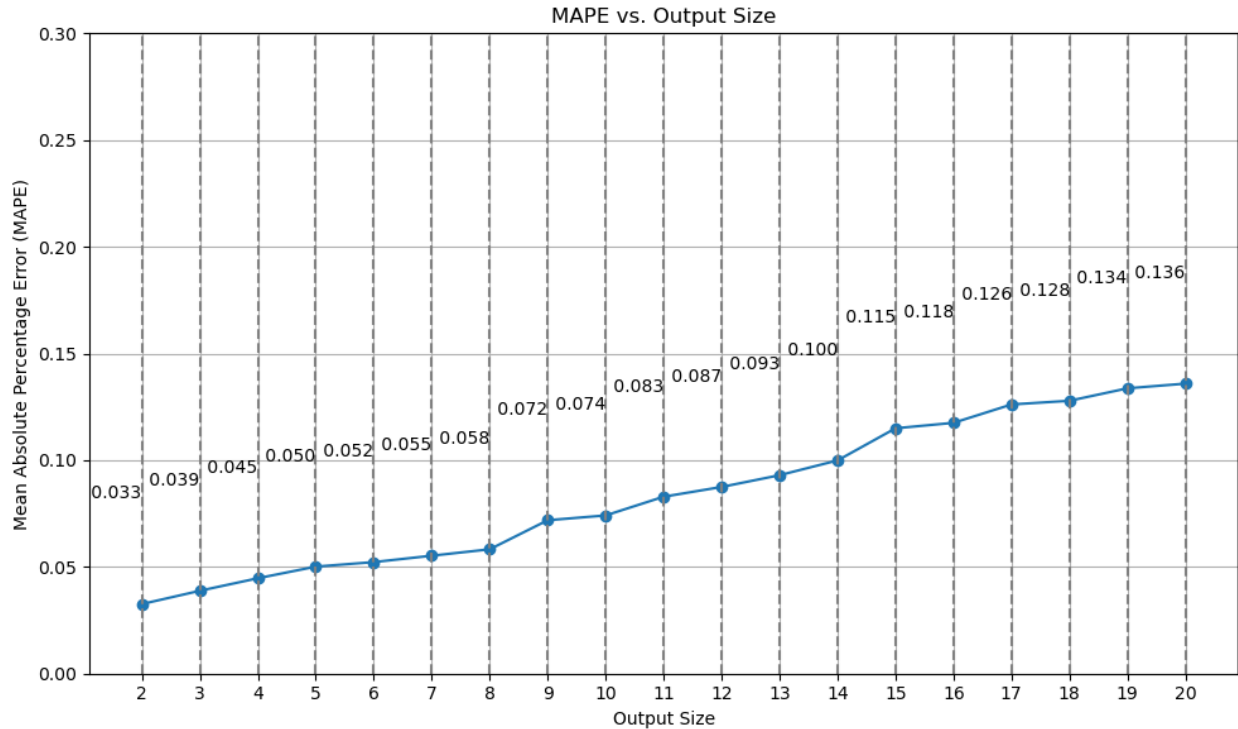


Σχήμα 6.16 MAPE για πρόβλεψη πολλών ημερών με Linear Regression

Όπως φαίνεται από το Σχήμα 6.16, για πρόβλεψη μέχρι και 6 ημέρες παρατηρείται σφάλμα κάτω από 0.05, ενώ από την 14^η ημέρα και έπειτα το σφάλμα είναι μεγαλύτερο του 0.10.

6.3.1 XGBoost

Θα εξεταστεί η δυνατότητα πρόβλεψης των τιμών προθεσμιακών συμβολαίων για τις n επόμενες μέρες με την αρχιτεκτονική XGBoost. Το n θα παίρνει τιμές από 1 έως 20. Για κάθε n θα εκπαιδευτεί ένα μοντέλο XGBoost χρησιμοποιώντας την μέθοδο αναζήτησης πλέγματος. Στο παρακάτω Σχήμα παρουσιάζεται η απόδοση των μοντέλων ως προς την μετρική MAPE για $n \in \{1,2,3, \dots, 20\}$.

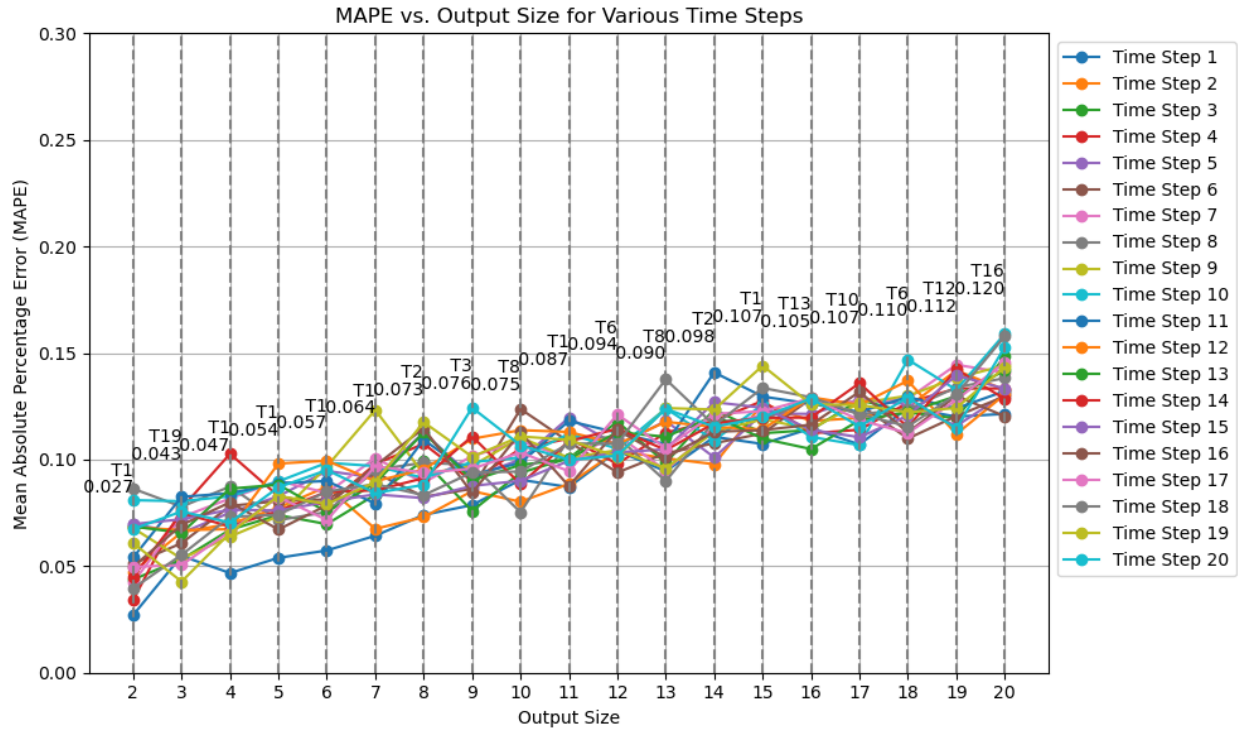


Σχήμα 6.17 MAPE για πρόβλεψη πολλών ημερών με XGBoost

Όπως φαίνεται από το Σχήμα 6.17, για πρόβλεψη μέχρι και 4 ημέρες παρατηρείται σφάλμα κάτω από 0.05, ενώ από την 13^η ημέρα και έπειτα το σφάλμα είναι μεγαλύτερο του 0.10.

6.3.3 Vanilla LSTM

Αρχικά, θα εξεταστεί η δυνατότητα πρόβλεψης των τιμών προθεσμιακών συμβολαίων για τις n επόμενες μέρες με την αρχιτεκτονική Vanilla LSTM. Θα χρησιμοποιηθεί η αρχιτεκτονική που εφαρμόστηκε και για την πρόβλεψη της επόμενης ημέρας που περιλαμβάνει 100 μονάδες LSTM στο αναδρομικό επίπεδο. Ως αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο Adam και τα μοντέλα εκπαιδεύτηκαν για 100 εποχές (epochs). Η πρόβλεψη γίνεται για $n \in \{1,2,3, \dots, 20\}$ μέρες μπροστά και θα αναζητηθεί η βέλτιστη τιμή χρονικού παραθύρου με χρήση αναζήτησης πλέγματος για τιμές $t \in \{1,2,3, \dots, 20\}$.



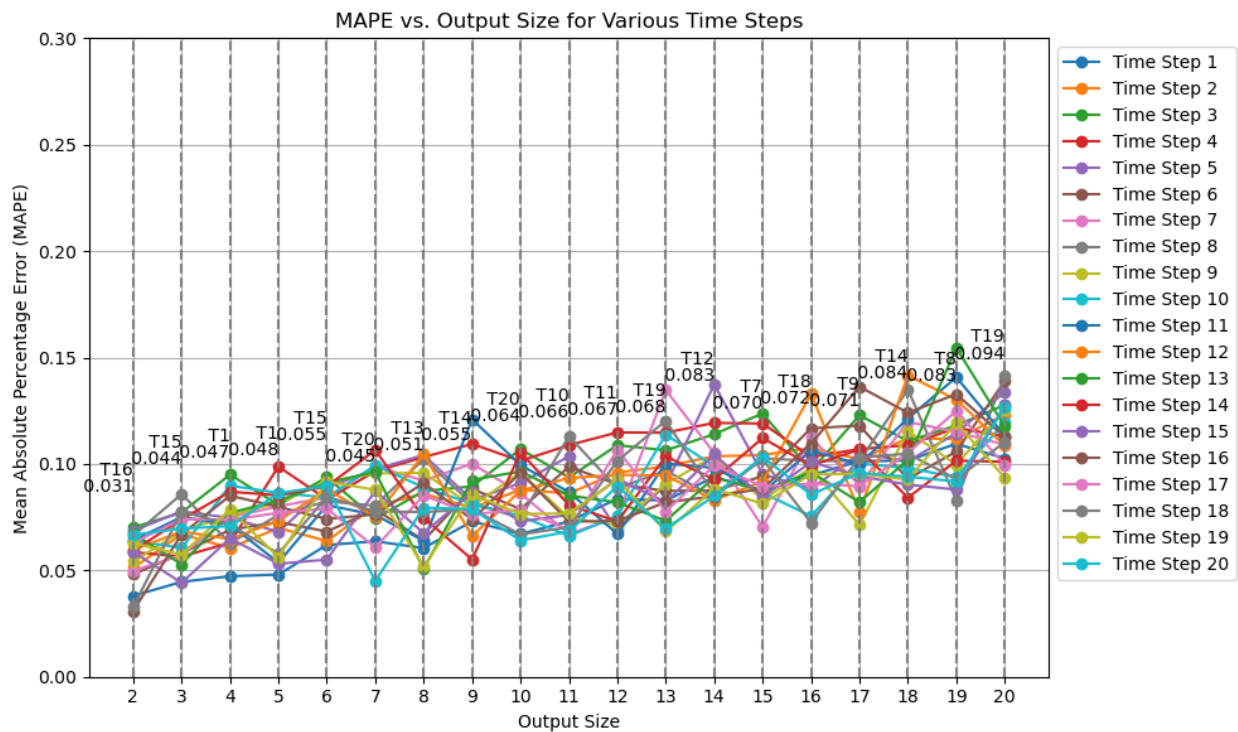
Σχήμα 6.18 MAPE για πρόβλεψη πολλών ημερών με Vanilla LSTM

Στο παραπάνω Σχήμα (Σχήμα 6.18) παρουσιάζεται η απόδοση των μοντέλων βάσει της μετρικής MAPE σε σχέση με το μήκος της ακολουθίας εξόδου, δηλαδή των αριθμό των ημερών στο μέλλον για τις οποίες γίνεται πρόβλεψη. Κάθε καμπύλη αντιστοιχεί σε ένα μοντέλο που χρησιμοποιεί ένα συγκεκριμένο μήκος παραθύρου για την ακολουθία εισόδου. Παρατηρείται ότι για πρόβλεψη 2 ημερών μετά το μοντέλο που χρησιμοποιεί μήκος παραθύρου 1 πετυχαίνει υψηλή επίδοση με $MAPE = 0.027$. Όσο αυξάνεται το παράθυρο πρόβλεψης το σφάλμα αυξάνεται αλλά παραμένει σε ικανοποιητικά πλαίσια μέχρι και την πρόβλεψη της τιμής της 4^{ης} ημέρας όπου το MAPE είναι ίσο με 0.047. Μετά τις 5 ημέρες το MAPE γίνεται μεγαλύτερο από 0.05 και παραμένει μικρότερο από 0.10 μέχρι και την 13^η μέρα.

6.3.4 Stacked LSTM

Θα χρησιμοποιηθεί η αρχιτεκτονική που εφαρμόστηκε και για την πρόβλεψη της επόμενης ημέρας που περιλαμβάνει 100 μονάδες LSTM στο αναδρομικό επίπεδο. Ως

αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο Adam και τα μοντέλα εκπαιδεύτηκαν για 100 εποχές (epochs). Η πρόβλεψη γίνεται για $n \in \{1,2,3, \dots, 20\}$ μέρες μπροστά και θα αναζητηθεί η βέλτιστη τιμή χρονικού παραθύρου με χρήση αναζήτησης πλέγματος για τιμές $t \in \{1,2,3, \dots, 20\}$.



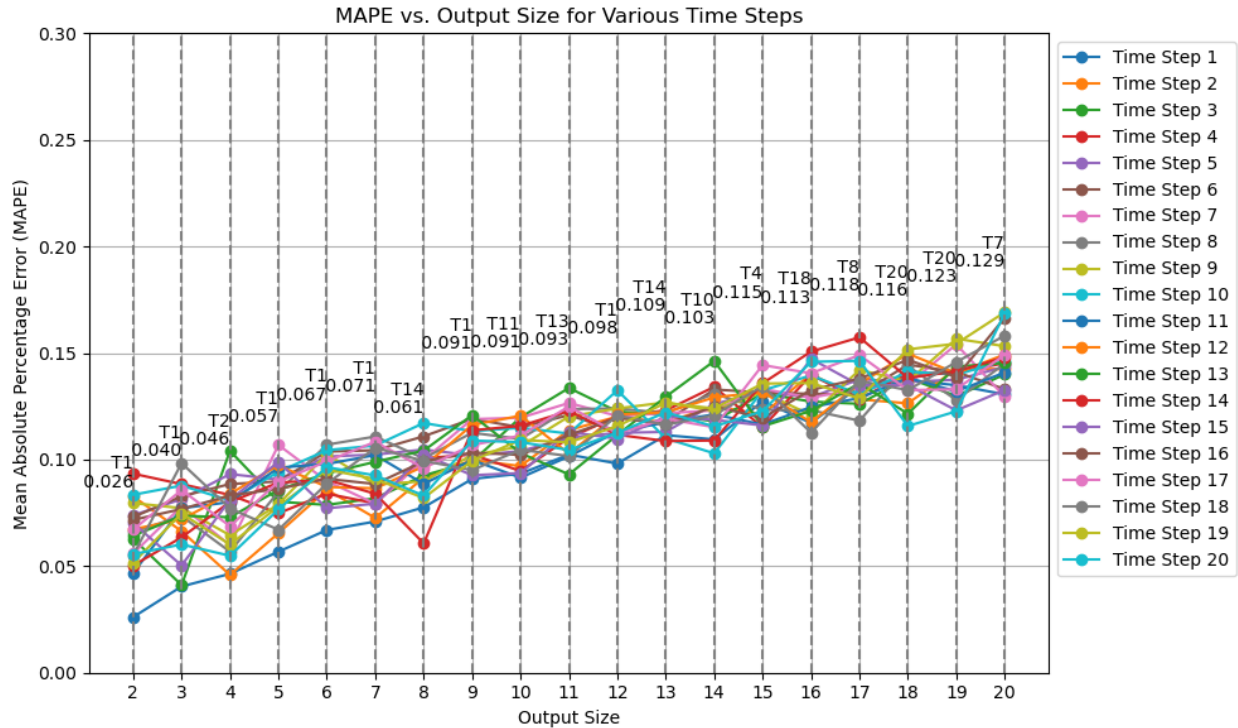
Σχήμα 6.19 MAPE για πρόβλεψη πολλών ημερών με Stacked LSTM

Στο παραπάνω Σχήμα (Σχήμα 6.19) παρατηρείται ότι για πρόβλεψη 2 ημερών μετά το μοντέλο που χρησιμοποιεί μήκος παραθύρου 16 πετυχαίνει επίδοση με MAPE = 0.031. Όσο αυξάνεται το παράθυρο πρόβλεψης, το σφάλμα αυξάνεται αλλά παραμένει σε ικανοποιητικά πλαίσια μέχρι και την πρόβλεψη της τιμής 5^{ης} ημέρας όπου το MAPE είναι ίσο με 0.048. Μετά την 5^η ημέρα το MAPE γίνεται μεγαλύτερο από 0.05 και παραμένει μέχρι και την 20^η μέρα κάτω από 0.10. Συνολικά, οι προβλέψεις αυτής της αρχιτεκτονικής έχουν χαμηλότερο σφάλμα από αυτές τις Vanilla LSTM.

6.3.5 Bidirectional LSTM

Θα χρησιμοποιηθεί η αρχιτεκτονική που εφαρμόστηκε και για την πρόβλεψη της επόμενης ημέρας που περιλαμβάνει 100 μονάδες LSTM στο αναδρομικό επίπεδο. Ως

αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο Adam και τα μοντέλα εκπαιδεύτηκαν για 100 εποχές (epochs). Η πρόβλεψη γίνεται για $n \in \{1,2,3, \dots, 20\}$ μέρες μπροστά και θα αναζητηθεί η βέλτιστη τιμή χρονικού παραθύρου με χρήση αναζήτησης πλέγματος για τιμές $t \in \{1,2,3, \dots, 20\}$.



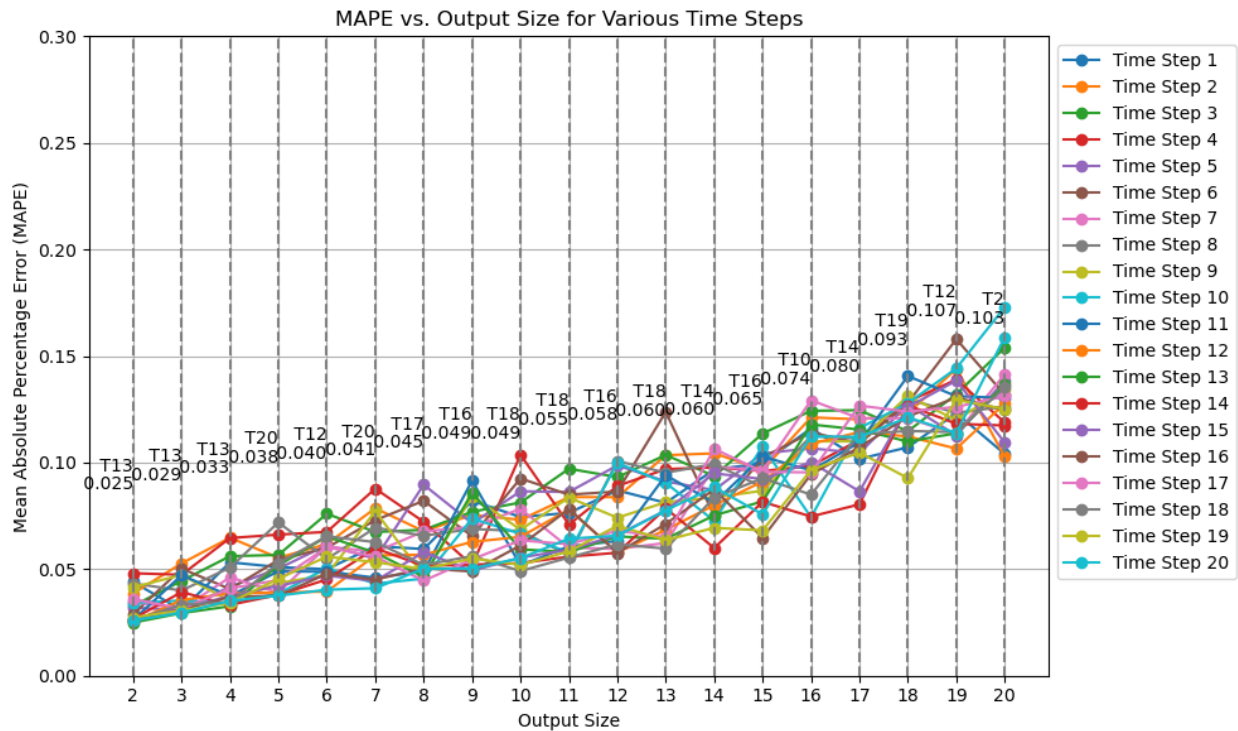
Σχήμα 6.20 MAPE για πρόβλεψη πολλών ημερών με Bidirectional LSTM

Στο παραπάνω Σχήμα (Σχήμα 6.20) παρατηρείται ότι για πρόβλεψη 2 ημερών μετά το μοντέλο που χρησιμοποιεί μήκος παραθύρου 1 πετυχαίνει υψηλή επίδοση με MAPE = 0.026, όπως και το μοντέλο της αρχιτεκτονικής Vanilla LSTM. Όσο αυξάνεται το παράθυρο πρόβλεψης το σφάλμα αυξάνεται αλλά παραμένει σε ικανοποιητικά πλαίσια μέχρι και την πρόβλεψη της τιμής της 4^{ης} ημέρας όπου το MAPE είναι ίσο με 0.046. Μετά την 4^η ημέρα το MAPE γίνεται μεγαλύτερο από 0.05 και παραμένει μέχρι και την 12^η μέρα κάτω από 0.10. Συνολικά, οι προβλέψεις αυτής της αρχιτεκτονικής έχουν υψηλότερο σφάλμα από τις αυτές των προηγούμενων αρχιτεκτονικών.

6.3.6 Dual Input Attention LSTM

Θα χρησιμοποιηθεί η αρχιτεκτονική που εφαρμόστηκε και για την πρόβλεψη της επόμενης ημέρας που περιλαμβάνει 100 μονάδες LSTM στο αναδρομικό επίπεδο. Ως αλγόριθμος βελτιστοποίησης (optimizer) επιλέχθηκε ο Adam και τα μοντέλα

εκπαιδεύτηκαν για 100 εποχές (epochs). Η πρόβλεψη γίνεται για $n \in \{1,2,3, \dots, 20\}$ μέρες μπροστά και θα αναζητηθεί η βέλτιστη τιμή χρονικού παραθύρου με χρήση αναζήτησης πλέγματος για τιμές $t \in \{1,2,3, \dots, 20\}$.



Σχήμα 6.21 MAPE για πρόβλεψη πολλών ημερών με DIA-LSTM

Στο παραπάνω Σχήμα (Σχήμα 6.21) παρατηρείται ότι οι προβλέψεις για την 2^η, 3^η και 4^η ημέρα παρουσιάζουν σφάλμα MAPE ίσο με 0.25, 0.29 και 0.33 για τα μοντέλο που χρησιμοποιεί μήκος παραθύρου ίσο με 13. Από τις προβλέψεις των πρώτων ημερών παρατηρείται υψηλότερη επίδοση σε σχέση με τις προηγούμενες αρχιτεκτονικές που εξετάστηκαν. Ομοίως, όσο αυξάνεται το παράθυρο πρόβλεψης, αυξάνεται και το σφάλμα αλλά παραμένει σε ικανοποιητικά πλαίσια μέχρι και την πρόβλεψη της τιμής της 10^{ης} ημέρας όπου το MAPE είναι ίσο με 0.049. Μετά την 10^η ημέρα το MAPE γίνεται μεγαλύτερο από 0.05 και παραμένει μέχρι και την 18^η μέρα κάτω από 0.10. Συνολικά, οι προβλέψεις αυτής της αρχιτεκτονικής έχουν χαμηλότερο σφάλμα από τις αυτές των προηγούμενων αρχιτεκτονικών.

Κεφάλαιο 7

Συμπεράσματα και Μελλοντική Εργασία

Στην παρούσα διπλωματική εξετάστηκαν μοντέλα μηχανικής μάθησης, καθώς και μοντέλα βαθιάς μάθησης για την πρόβλεψη τιμών προθεσμιακών προϊόντων χρησιμοποιώντας τηλεσκοπικά δεδομένα και δεδομένα χρηματιστηρίου. Τα μοντέλα αξιολογήθηκαν ως προς την δυνατότητα τους να προβλέπουν τις τιμές των προθεσμιακών συμβολαίων αγροτικών προϊόντων σε διάστημα 1 έως n ημέρες. Τα μοντέλα μηχανικής μάθησης και συγκεκριμένα το μοντέλο Linear Regression εμφάνιζε υψηλότερη ακρίβεια στις προβλέψεις επόμενης ημέρας ενώ ακολουθούσαν το μοντέλο DIA-LSTM, Vanilla LSTM, Stacked LSTM, Bidirectional LSTM, CNN-LSTM. Σε προβλέψεις επόμενων ημερών τα μοντέλα βαθιάς μάθησης παρουσίασαν καλύτερη επίδοση από τα μοντέλα μηχανικής μάθησης, ενώ εκεί ξεχώρισε το μοντέλο DIA-LSTM.

Σημαντική κρίνεται η διερεύνηση της αρχιτεκτονικής Transformer για την πρόβλεψη των τιμών προθεσμιακών συμβολαίων σε μελλοντική εργασία. Επιπλέον, μια εξίσου πολλά υποσχόμενη οδός για διερεύνηση θα ήταν η ενσωμάτωση ενός συνόλου δεδομένων που θα περιέχει την ακριβή θέση των καλλιεργειών για το υπό μελέτη προϊόν. Η αξιοποίηση τέτοιων δεδομένων θα επέτρεπε την ακριβέστερη εξαγωγή τηλεσκοπικών δεδομένων, βελτιώνοντας έτσι τις προβλέψεις και προσφέροντας στους ενδιαφερόμενους πιο αξιόπιστες γνώσεις σχετικά με τη δυναμική των αγορών γεωργικών προϊόντων.

Κατάλογος Σχημάτων

Σχήμα 2.1 Προβλεπόμενο διάνυσμα P και πραγματικό διάνυσμα Y	22
Σχήμα 2.2 Γραμμική παλινδρόμηση	24
Σχήμα 2.3 Βιολογικός νευρώνας πηγή	28
Σχήμα 2.4 Τεχνητός νευρώνας πηγή	28
Σχήμα 2.5 Καμπύλη με πολλά ακρότατα. πηγή : researchgate.com	32
Σχήμα 2.6 Νευρωνικό Δίκτυο Παραδείγματος 2.5	33
Σχήμα 2.7 Αρχιτεκτονική Συνελικτικού Νευρωνικού Δικτύου	37
Σχήμα 2.8 Επίπεδο συνέλιξης	38
Σχήμα 2.9 Μέγιστη και μέση υποδειγματοληψία	39
Σχήμα 2.10 Ακολουθιακό Δίκτυο	40
Σχήμα 2.11 Δομή κυττάρου LSTM	41
Σχήμα 3.1 Τιμές προθεσμιακών συμβολαίων καλαμποκιού	Error! Bookmark not defined.
Σχήμα 5.1 Χωρισμός δεδομένων σε σύνολο εκπαίδευσης και ελέγχου	57
Σχήμα 5.2 Αρχιτεκτονική Vanilla LSTM	60
Σχήμα 5.3 Αρχιτεκτονική Stacked LSTM	61
Σχήμα 5.4 Αρχιτεκτονική Bidirectional LSTM	63
Σχήμα 5.5 Αρχιτεκτονική CNN-LSTM	64
Σχήμα 5.6 Αρχιτεκτονική DIA-LSTM	66
Σχήμα 6.1 Πίνακας συσχέτισης μεταβλητών	70
Σχήμα 6.2 Σημαντικότητα χαρακτηριστικών για γραμμικό μοντέλο	71
Σχήμα 6.3 Προβλέψεις γραμμικού μοντέλου	73
Σχήμα 6.4 Σημαντικότητα χαρακτηριστικών μοντέλου XGBoost	Error! Bookmark not defined.
Σχήμα 6.5 Προβλέψεις XGBoost	75
Σχήμα 6.6 MAPE σε σχέση με μήκος παραθύρου για Vanilla LSTM	76
Σχήμα 6.7 Προβλέψεις Vanilla LSTM	Error! Bookmark not defined.
Σχήμα 6.8 MAPE σε σχέση με μήκος παράθυρου για stacked LSTM	78
Σχήμα 6.9 Προβλέψεις stacked LSTM	79
Σχήμα 6.10 MAPE σε σχέση με μήκος παράθυρου για bidirectional LSTM	80
Σχήμα 6.11 Προβλέψεις bidirectional LSTM	81
Σχήμα 6.12 MAPE σε σχέση με μήκος παράθυρου για CNN-LSTM	Error! Bookmark not defined.
Σχήμα 6.13 Προβλέψεις για CNN-LSTM	Error! Bookmark not defined.
Σχήμα 6.14 MAPE σε σχέση με μήκος παράθυρου για DIA-LSTM	83
Σχήμα 6.15 Προβλέψεις DIA-LSTM	84
Σχήμα 6.16 MAPE για πρόβλεψη πολλών ημερών με Linear Regression	85
Σχήμα 6.17 MAPE για πρόβλεψη πολλών ημερών με XGBoost	86

Σχήμα 6.18 MAPE για πρόβλεψη πολλών ημερών με Vanilla LSTM	87
Σχήμα 6.19 MAPE για πρόβλεψη πολλών ημερών με stacked LSTM	Error! Bookmark not defined.
Σχήμα 6.20 MAPE για πρόβλεψη πολλών ημερών με bidirectional LSTM	Error!
Σχήμα 6.21 MAPE για πρόβλεψη πολλών ημερών με DIA-LSTM	90

Κατάλογος Πινάκων

Πίνακας 5.1 Χαρακτηριστικά αρχιτεκτονικής DIA-LSTM	69
Πίνακας 6.1 Μετρικές linear regression	74
Πίνακας 6.2 Μετρικές XGBoost	77
Πίνακας 6.3 Μετρικές Vanilla LSTM	78
Πίνακας 6.4 Μετρικές stacked LSTM	80
Πίνακας 6.5 Μετρικές bidirectional LSTM	82
Πίνακας 6.6 Μετρικές CNN-LSTM	84
Πίνακας 6.7 Μετρικές DIA-LSTM	86

Βιβλιογραφία

- [1] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Comput*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [2] J. He, H. Yang, S. Zhou, J. Chen, and M. Chen, “A Dual-Attention-Mechanism Multi-Channel Convolutional LSTM for Short-Term Wind Speed Prediction,” *Atmosphere* 2023, *Vol. 14, Page 71*, vol. 14, no. 1, p. 71, Dec. 2022, doi: 10.3390/ATMOS14010071.
- [3] X. Zhang, X. Liang, A. Zhiyuli, S. Zhang, R. Xu, and B. Wu, “AT-LSTM: An Attention-based LSTM Model for Financial Time Series Prediction,” *IOP Conf Ser Mater Sci Eng*, vol. 569, no. 5, p. 052037, Jul. 2019, doi: 10.1088/1757-899X/569/5/052037.
- [4] H. Yin, D. Jin, Y. H. Gu, C. J. Park, S. K. Han, and S. J. Yoo, “STL-ATTLSTM: Vegetable Price Forecasting Using STL and Attention Mechanism-Based LSTM,” *Agriculture* 2020, *Vol. 10, Page 612*, vol. 10, no. 12, p. 612, Dec. 2020, doi: 10.3390/AGRICULTURE10120612.
- [5] A. L. Samuel, “Some studies in machine learning using the game of checkers,” *IBM J Res Dev*, vol. 44, no. 1–2, pp. 207–219, 2000, doi: 10.1147/rd.441.0206.
- [6] “Machine Learning textbook.” Accessed: Nov. 01, 2020. [Online]. Available: <http://www.cs.cmu.edu/~tom/mlbook.html>
- [7] J. H. Friedman, Friedman, and J. H., “Stochastic gradient boosting,” *Comput Stat Data Anal*, vol. 38, no. 4, pp. 367–378, 2002, Accessed: Oct. 30, 2023. [Online]. Available: <https://EconPapers.repec.org/RePEc:eee:csdana:v:38:y:2002:i:4:p:367-378>
- [8] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System”, Accessed: Oct. 30, 2023. [Online]. Available: <https://github.com/dmlc/xgboost>
- [9] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.
- [10] Y. Lecun *et al.*, “Handwritten digit recognition with a back-propagation network.” Morgan Kaufmann, 1990. Accessed: Oct. 10, 2020. [Online]. Available: <https://nyuscholars.nyu.edu/en/publications/handwritten-digit-recognition-with-a-back-propagation-network-2>
- [11] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998, doi: 10.1109/5.726791.

- [12] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, vol. 44, no. 1. MIT Press, 2016. doi: 10.1088/1751-8113/44/8/085201.
- [13] B. T. Polyak, “Some methods of speeding up the convergence of iteration methods,” *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964, doi: 10.1016/0041-5553(64)90137-5.
- [14] D. P. Kingma and J. L. Ba, “Adam: A method for stochastic optimization,” in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, International Conference on Learning Representations, ICLR, Dec. 2015. Accessed: Oct. 18, 2020. [Online]. Available: <https://arxiv.org/abs/1412.6980v9>
- [15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, “Learning representations by back-propagating errors,” *Nature 1986 323:6088*, vol. 323, no. 6088, pp. 533–536, 1986, doi: 10.1038/323533a0.
- [16] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997, doi: 10.1109/78.650093.
- [17] L. Y. Pratt, “Discriminability-Based Transfer between Neural Networks,” in *Advances in neural information processing systems*, 1993, pp. 204–211.