



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Πολυτεχνική Σχολή
Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Διπλωματική Εργασία

Αλγοριθμικές συναλλαγές μετοχών μέσω μηχανικής μάθησης

Καμάτσος Προκόπιος
Α.Μ. 1072586

Επιβλέπων
Καθηγητής Ανδρικόπουλος Αθανάσιος

Μέλος Επιτροπής Αξιολόγησης
Αναπληρωτής Καθηγητής Μακρής Χρήστος
ΕΔΙΠ Ηλίας Αριστείδης

Πάτρα, 2025

© Copyright συγγραφής Καμάτσος Προκόπιος, 2025

© Copyright θέματος Ανδρικόπουλος Αθανάσιος

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της διπλωματικής εργασίας από το Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών & Πληροφορικής του Πανεπιστημίου Πατρών δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

<Στην οικογένειά μου>

Πρόλογος

Η παρούσα διπλωματική εργασία ολοκληρώθηκε στο πλαίσιο του προπτυχιακού κύκλου σπουδών στο πρόγραμμα σπουδών του Τμήματος Μηχανικών Υπολογιστών και Πληροφορικής του Πανεπιστημίου Πατρών. Ο τίτλος της παρούσας διπλωματικής εργασίας είναι «Αλγοριθμικές συναλλαγές μετοχών μέσω μηχανικής μάθησης». Καταρχάς, θα ήθελα να ευχαριστήσω τον υπεύθυνο καθηγητή μου κ. Αθανάσιο Ανδρικόπουλο, αναπληρωτή καθηγητή του Τμήματος Μηχανικών Η/Υ και Πληροφορικής του Πανεπιστημίου Πατρών με ερευνητικά ενδιαφέροντα την Τοπολογία και τις εφαρμογές της στην Οικονομική Θεωρία, Θεωρία Παιγνίων, Επιστήμη Υπολογιστών, Θεωρία Κοινωνικών Επιλογών, Θεωρία Γραφημάτων και Θεωρία Τομέων, για την εμπιστοσύνη που μου έδειξε δίνοντάς μου το θέμα της διπλωματικής μου. Μου έδωσε ιδέες και τρόπους για να ξεκινήσω αυτή τη δουλειά καθώς και συμβουλές για την σωστή της δόμηση. Τέλος, οφείλω ένα τεράστιο ευχαριστώ στην οικογένειά μου για τη συνεχή υποστήριξή τους σε όλη τη διάρκεια των σπουδών μου.

Περίληψη

Η παρούσα διπλωματική εργασία επικεντρώνεται στην ανάπτυξη ενός συστήματος πρόβλεψης τιμών μετοχών, βασισμένου σε τεχνικές μηχανικής μάθησης, αξιοποιώντας ιστορικά δεδομένα και τεχνικούς δείκτες. Η αλγοριθμική συναλλαγή, ως μια μέθοδος που εφαρμόζει προκαθορισμένους κανόνες και στρατηγικές, έχει ως στόχο την αυτοματοποίηση των επενδυτικών αποφάσεων, προσφέροντας αυξημένη ταχύτητα και ακρίβεια. Παράλληλα, το περιβάλλον των χρηματιστηριακών δεδομένων χαρακτηρίζεται από έντονη μεταβλητότητα και την επιρροή πολυάριθμων παραγόντων, γεγονός που περιορίζει την αποτελεσματικότητα των παραδοσιακών μεθόδων πρόβλεψης. Σε αυτό το πλαίσιο, η μηχανική μάθηση αναδεικνύεται ως μια καινοτόμος προσέγγιση, καθώς οι αλγόριθμοί της μπορούν να αναγνωρίζουν μοτίβα μέσα στα δεδομένα και να προσαρμόζονται δυναμικά στις μεταβαλλόμενες συνθήκες της αγοράς. Στην παρούσα εργασία, χρησιμοποιούνται μοντέλα όπως το Random Forest, το Support Vector Regressor (SVR) και η Linear Regression (Γραμμική Παλινδρόμηση), τα οποία ενσωματώνονται σε ένα ενιαίο σύστημα πρόβλεψης. Τα μοντέλα αυτά αξιολογούνται με βάση την ακρίβεια των προβλέψεών τους, ενώ η προετοιμασία των δεδομένων περιλαμβάνει τεχνικές κανονικοποίησης και τη χρήση επιλεγμένων δεικτών, όπως ο RSI (Δείκτης Σχετικής Ισχύος) και οι ΚΜΟ (Κινητοί Μέσοι Όροι). Οι δείκτες αυτοί ενισχύουν τη λειτουργικότητα των μοντέλων, βελτιώνοντας την ικανότητά τους να αποδίδουν ακριβέστερες προβλέψεις. Με τη μεθοδολογία αυτή, το προτεινόμενο σύστημα πρόβλεψης αποδεικνύεται ένα χρήσιμο εργαλείο για τη λήψη επενδυτικών αποφάσεων, συμβάλλοντας στη βελτίωση της ακρίβειας και της αποδοτικότητας. Στην ανάλυση των αποτελεσμάτων, πραγματοποιείται συγκριτική αξιολόγηση των μοντέλων, με στόχο την ανάδειξη των πλεονεκτημάτων και των περιορισμών του καθενός. Παράλληλα, προτείνονται κατευθύνσεις για μελλοντική έρευνα, όπως η ενσωμάτωση βαθύτερων αλγορίθμων μάθησης ή τεχνικών ενισχυτικής μάθησης, που μπορούν να βελτιώσουν περαιτέρω τις επιδόσεις του συστήματος. Η παρούσα εργασία καταλήγει στο συμπέρασμα ότι η μηχανική μάθηση αποτελεί ισχυρό εργαλείο για την ενίσχυση της ακρίβειας στις προβλέψεις χρηματιστηριακών τιμών, προσφέροντας σημαντική προστιθέμενη αξία στην αλγοριθμική συναλλαγή και στις επενδυτικές στρατηγικές.

Abstract

This thesis focuses on the development of a stock price prediction system based on machine learning techniques, using historical data and technical indicators. Algorithmic trading, as a method that applies predefined rules and strategies, aims to automate investment decisions, offering increased speed and accuracy. At the same time, the stock market data environment is characterized by high volatility and the influence of numerous factors, which limits the effectiveness of traditional forecasting methods. In this context, machine learning emerges as an innovative approach, as its algorithms can identify patterns within the data and dynamically adapt to changing market conditions. In this paper, models such as Random Forest, Support Vector Regressor (SVR) and Linear Regression are used and integrated into a single prediction system. These models are evaluated on the basis of the accuracy of their predictions, while data preparation includes normalisation techniques and the use of selected indicators such as RSI (Relative Strength Index) and CMOs (Moving Averages). These indicators enhance the functionality of the models, improving their ability to deliver more accurate forecasts. With this methodology, the proposed forecasting system proves to be a useful tool for investment decision making, contributing to improve accuracy and efficiency. In the analysis of the results, a comparative evaluation of the models is carried out in order to highlight the advantages and limitations of each one. At the same time, directions for future research are suggested, such as the incorporation of deeper learning algorithms or reinforcement learning techniques, which can further improve system performance. This paper concludes that machine learning is a powerful tool for enhancing accuracy in stock market price forecasts, providing significant added value to algorithmic trading and investment strategies.

Περιεχόμενα

<i>Πρόλογος</i>	6
<i>Περίληψη</i>	7
<i>Abstract</i>	8
<i>Περιεχόμενα</i>	9
<i>1</i>	1
<i>Εισαγωγή</i>	1
1.1 Σημασία του προβλήματος.....	2
1.2 Στόχοι της Εργασίας	3
1.3 Μεθοδολογία Προσέγγισης.....	4
1.4 Συνεισφορά	5
1.5 Διάρθρωση της Διπλωματικής Εργασίας.....	6
2	7
<i>Θεωρητικό Υπόβαθρο</i>	7
2.1 Αλγοριθμική Συναλλαγή και Χρηματιστηριακή Ανάλυση.....	7
2.1.1 Εισαγωγή στην Αλγοριθμική Συναλλαγή	8
2.1.2 Βασικές Έννοιες Χρηματιστηριακής Ανάλυσης.....	10
2.2 Μηχανική Μάθηση στις Προβλέψεις Χρηματιστηριακών Τιμών.....	11
2.2.1 Βασικές Έννοιες Μηχανικής Μάθησης.....	11
2.2.2 Εφαρμογές Μηχανικής Μάθησης στη Χρηματοοικονομική Ανάλυση	12
2.3 Συστήματα και Τεχνικές Πρόβλεψης Χρηματιστηριακών Τιμών.....	13
2.3.1 Προκλήσεις και Στρατηγικές για Πρόβλεψη Χρονοσειρών	14
2.3.2 Τεχνικοί Δείκτες και Αναλυτικές Τεχνικές.....	15
3	19
<i>Υλοποίηση και Σχεδίαση του Συστήματος</i>	19
3.1 Δομή και Στόχοι του Συστήματος Πρόβλεψης.....	20
3.2 Προεπεξεργασία Δεδομένων και Επιλογή Χαρακτηριστικών	21
3.2.1 Επεξεργασία Χρονοσειρών και Κανονικοποίηση	22
3.2.2 Επιλογή Τεχνικών Δεικτών	22
3.3 Περιγραφή των Χρησιμοποιούμενων Αλγορίθμων	23
3.3.1 Random Forest	23
3.3.2 Support Vector Regressor	25
3.3.3 Linear Regression	28
3.4 Αρχιτεκτονική του Συστήματος και Εργαλεία.....	30
3.4.1 Δομή Αρχείων και Ροή Εργασίας.....	30
3.4.2 Χρήση Πλατφορμών και Εργαλείων.....	31
4	33
<i>Αξιολόγηση Μοντέλων και Ανάλυση Αποτελεσμάτων</i>	33
4.1 Μεθοδολογία Αξιολόγησης.....	35
4.1.1 Κριτήρια Αξιολόγησης.....	37

4.1.2	Σύγκριση Απόδοσης Μοντέλων.....	38
4.2	Παρουσίαση και Ανάλυση Αποτελεσμάτων.....	39
4.2.1	Αποτελέσματα για Random Forest.....	40
4.2.2	Αποτελέσματα για SVR	46
4.2.3	Αποτελέσματα για Linear Regression.....	53
4.2.4	Αναλυτικός Πίνακας Αποτελεσμάτων	58
4.3	Συγκριτική Αξιολόγηση και Συμπεράσματα	59
4.3.1	Συγκριτική Αξιολόγηση των Μοντέλων Εργασίας	60
4.3.2	Κοινά Χαρακτηριστικά και Διαφορές με άλλη Μελέτη	61
4.3.3	Ερμηνεία και σύγκριση αποτελεσμάτων με άλλη Μελέτη	63
5	65
	Συμπεράσματα και Προοπτικές.....	65
5.1	Σύνοψη.....	65
5.2	Συμπεράσματα	65
5.3	Μελλοντική Εργασία.....	66
	Βιβλιογραφία - Αναφορές	69
	Παραρτήμα Α: <Κώδικας>.....	71
	Σύντομο Βιογραφικό Συγγραφέα.....	77

Λίστα Εικονών

Εικόνα 1. UI DropDown Menu	34
Εικόνα 2. UI Main.....	35
Εικόνα 3. ADBE Random Forest	40
Εικόνα 4. COST Random Forest	41
Εικόνα 5. FB Random Forest.....	42
Εικόνα 6. NFLX Random Forest.....	43
Εικόνα 7. PYPL Random Forest	44
Εικόνα 8. QCOM Random Forest	45
Εικόνα 9. TMUS Random Forest.....	45
Εικόνα 10. ADBE SVR	47
Εικόνα 11. COST SVR	48
Εικόνα 12. FB SVR	49
Εικόνα 13. NFLX SVR	50
Εικόνα 14. PYPL SVR.....	51
Εικόνα 15. QCOM SVR	52
Εικόνα 16. TMUS SVR	53
Εικόνα 17. ADBE Linear Regression	54
Εικόνα 18. COST Linear Regression.....	54
Εικόνα 19. FB Linear Regression	55
Εικόνα 20. NFLX Linear Regression	56
Εικόνα 21. PYPL Linear Regression	57
Εικόνα 22. QCOM Linear Regression.....	57
Εικόνα 23. TMUS Linear Regression.....	58

Λίστα Πινάκων

Πίνακας 1. Ενδεικτικός Πίνακας Αποτελεσμάτων.....	39
Πίνακας 2. Αναλυτικός Πίνακας Αποτελεσμάτων.....	59
Πίνακας 3. Μοντέλο LSTM του Παπανικόλα : Τιμές ανά Μετοχή	63
Πίνακας 4. Συνολικός Συγκριτικός Πίνακας με Άλλη Μελέτη	63
Πίνακας 5. Σύγκριση Αποδοτικότητας.....	64

Συντομογραφίες

SVR	Support Vector Regressor
RSI	Relative strength index
KMO	Κινητοί Μέσοι Όροι
MAE	Mean Absolute Error
MSE	Mean Squared Error
HFT	High Frequency Trading
DMS	Data Management System
ML	Machine Learnig
MA	Moving Averages
SMA	Simple Moving Average/Απλός Κινητός Μέσος
EMA	Exponential Moving Average/Εκθετικός Κινητός Μέσος
MACD	Moving Average Convergence Divergence
OOB Error	Out-of-Bag Error
RBF	Radial Basis Function
SVM	Support Vector Machines
ETF	Exchange Traded Fund
CSV	Comma Separated Values
QUI	Graphical User Interface
ADBE	Adobe Inc
COST	Costco Wholesale Corporation
FB	Facebook Inc.
NFLX	Netflix Inc

PYPL	PayPal Holdings Inc
QCOM	Qualcomm Inc.
TMUS	T-Mobile US Inc.
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit

Γλωσσάρι ή Απόδοση Όρων

Αλγοριθμική Συναλλαγή	Η χρήση υπολογιστικών αλγορίθμων για αυτόματη εκτέλεση επενδυτικών αποφάσεων στις αγορές.
Μηχανική Μάθηση	Υποτομέας της Τεχνητής Νοημοσύνης που βασίζεται στην ανάπτυξη μοντέλων που μαθαίνουν από δεδομένα.
Random Forest	Αλγόριθμος μηχανικής μάθησης που χρησιμοποιεί πολλαπλά δέντρα απόφασης για πρόβλεψη ή κατάταξη.
Support Vector Regressor (SVR)	Αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για την πρόβλεψη συνεχών τιμών.
Linear Regression (Γραμμική Παλινδρόμηση)	Στατιστική μέθοδος που χρησιμοποιείται για τη μοντελοποίηση της σχέσης μεταξύ μιας ανεξάρτητης και μιας εξαρτημένης μεταβλητής.
Κινητοί Μέσοι Όροι	Τεχνικός δείκτης που εξομαλύνει τις διακυμάνσεις των τιμών σε ένα προκαθορισμένο χρονικό διάστημα.
RSI (Relative Strength Index)	Δείκτης που μετρά τη δύναμη και την ορμή μιας τιμής, χρησιμοποιείται για την αξιολόγηση υπερτιμημένων ή υποτιμημένων μετοχών.
MAE (Mean Absolute Error)	Μετρική αξιολόγησης που υπολογίζει το μέσο απόλυτο σφάλμα μεταξύ προβλέψεων και πραγματικών τιμών.
MSE (Mean Squared Error)	Μετρική που υπολογίζει το μέσο τετραγωνικό σφάλμα μεταξύ προβλέψεων και πραγματικών τιμών.
ETF (Exchange Traded Fund)	Είδος επενδυτικού κεφαλαίου που διαπραγματεύεται στο χρηματιστήριο.
LSTM (Long Short-Term Memory)	Τύπος νευρωνικού δικτύου που χρησιμοποιείται για ανάλυση και πρόβλεψη χρονοσειρών.
Τεχνική Ανάλυση	Μέθοδος που βασίζεται στην ιστορική μελέτη τιμών και όγκων για την πρόβλεψη μελλοντικών κινήσεων.
Χρηματοοικονομικές Χρονοσειρές	Διαδοχικές καταγραφές δεδομένων, όπως τιμές μετοχών, που χρησιμοποιούνται για προβλέψεις.

1

Εισαγωγή

Η πρόβλεψη τιμών μετοχών αποτελεί έναν από τους πιο απαιτητικούς και γρήγορα εξελισσόμενους τομείς στον χρηματοοικονομικό κλάδο. Οι επενδυτές, αναζητώντας να ελαχιστοποιήσουν τους κινδύνους και να μεγιστοποιήσουν τις αποδόσεις τους, στρέφονται σε νέες τεχνολογίες που ενισχύουν την ακρίβεια των επενδυτικών τους αποφάσεων. Παρά τις προσπάθειες αυτές, οι χρηματιστηριακές αγορές παραμένουν ιδιαίτερα απρόβλεπτες, λόγω της επιρροής ποικίλων παραγόντων, όπως οι οικονομικές πολιτικές, τα γεωπολιτικά γεγονότα, η απόδοση των επιχειρήσεων και η ψυχολογία των επενδυτών. Σε αυτό το πλαίσιο, η μηχανική μάθηση προσφέρει πρωτοποριακές λύσεις, αξιοποιώντας αλγόριθμους πρόβλεψης που αναλύουν μεγάλα σύνολα δεδομένων, εντοπίζουν μοτίβα και προσαρμόζονται δυναμικά σε νέες συνθήκες της αγοράς. Τεχνικές όπως το Random Forest, το Support Vector Regression (SVR) και η Linear Regression επιτρέπουν την ανάπτυξη μοντέλων που είναι ικανά να ανταποκρίνονται στις συνεχείς μεταβολές της αγοράς, παρέχοντας μεγαλύτερη ακρίβεια στις προβλέψεις. Η ανάγκη για ταχύτητα και ακρίβεια στις συναλλαγές έχει αναδείξει την αλγοριθμική συναλλαγή ως βασικό εργαλείο για τους σύγχρονους επενδυτές. Μέσω αυτής, οι αποφάσεις βασίζονται σε δεδομένα και στατιστικά μοντέλα, μειώνοντας τη συναισθηματική επιρροή και προσφέροντας μια πιο αντικειμενική και ελεγχόμενη προσέγγιση. Το αποτέλεσμα είναι η υιοθέτηση στρατηγικών που βελτιώνουν τη διαχείριση κινδύνων και αυξάνουν τη συνολική αποδοτικότητα των επενδύσεων.

1.1 Σημασία του προβλήματος

Η πρόβλεψη τιμών μετοχών αποτελεί έναν από τους πιο απαιτητικούς στόχους στον χρηματοοικονομικό τομέα, λόγω της πολυπλοκότητας και της αβεβαιότητας που χαρακτηρίζουν τις αγορές. Η αδυναμία πρόβλεψης των τιμών με ακρίβεια έχει άμεσες συνέπειες στη διαχείριση του επενδυτικού κινδύνου, επηρεάζοντας την αποδοτικότητα των επενδυτικών στρατηγικών. Οι επενδυτές και οι διαχειριστές κεφαλαίων επιδιώκουν να περιορίσουν την αβεβαιότητα αυτή, ώστε να διασφαλίσουν τα κεφάλαιά τους και να μεγιστοποιήσουν τις αποδόσεις τους. Στην πραγματικότητα, οι τιμές των μετοχών επηρεάζονται από έναν συνδυασμό παραγόντων, όπως:

- **Οικονομικές πολιτικές και μακροοικονομικά δεδομένα:** Αλλαγές στα επιτόκια, στον πληθωρισμό ή άλλους κρίσιμους δείκτες μπορούν να ανατρέψουν την ισορροπία στις αγορές, επηρεάζοντας τη ροή των επενδύσεων
- **Γεωπολιτικές εξελίξεις:** Διεθνή γεγονότα, όπως πολιτικές κρίσεις ή εμπορικές διενέξεις, συχνά προκαλούν έντονες διακυμάνσεις, δημιουργώντας κλίμα ανασφάλειας στους επενδυτές.
- **Απόδοση των εταιρειών:** Τα οικονομικά αποτελέσματα, οι καινοτόμες στρατηγικές και η συνολική εικόνα των εταιρειών διαμορφώνουν την εμπιστοσύνη της αγοράς και, κατά συνέπεια, τις αξίες των μετοχών
- **Συναισθηματικοί παράγοντες:** Η ανθρώπινη ψυχολογία διαδραματίζει καθοριστικό ρόλο, καθώς ο φόβος, η αισιοδοξία και η απληστία οδηγούν συχνά σε υπερβολικές αντιδράσεις, εντείνοντας τις διακυμάνσεις.

Η πρόβλεψη των τιμών μετοχών με παραδοσιακές μεθόδους αποδεικνύεται αναποτελεσματική στην αντιμετώπιση αυτών των παραγόντων. Οι κλασικές μέθοδοι, όπως η θεμελιώδης και η τεχνική ανάλυση, βασίζονται κυρίως σε ιστορικά δεδομένα και συγκεκριμένους δείκτες, περιορίζοντας τη δυνατότητα ανάλυσης των πολυδιάστατων αυτών παραγόντων. Αυτή η αδυναμία οδηγεί σε ανακριβείς προβλέψεις και αυξημένο επενδυτικό ρίσκο. Εδώ εισέρχεται η μηχανική μάθηση, που προσφέρει καινοτόμες λύσεις, αξιοποιώντας μοντέλα που μπορούν να αναγνωρίσουν μοτίβα και τάσεις, προσαρμόζοντας τις προβλέψεις τους δυναμικά. Με τη χρήση αλγορίθμων, όπως το Random Forest, το SVR, και η Linear Regression, η μηχανική μάθηση παρέχει εργαλεία που είναι ικανά να προσαρμόζονται

συνεχώς στις αλλαγές της αγοράς, βελτιώνοντας την ακρίβεια των προβλέψεων. Έτσι, ο χρηματοοικονομικός τομέας αποκτά πρόσβαση σε τεχνικές που μειώνουν τον κίνδυνο και επιτρέπουν τη λήψη πιο στοχευμένων και στρατηγικών αποφάσεων. Η εξέλιξη αυτή στην πρόβλεψη τιμών μετοχών υπογραμμίζει τη σημασία της μηχανικής μάθησης στην αλγοριθμική συναλλαγή, καθιστώντας την κρίσιμη για την αποτελεσματική διαχείριση επενδύσεων.

1.2 Στόχοι της Εργασίας

Ο κεντρικός στόχος αυτής της διπλωματικής εργασίας είναι η ανάπτυξη ενός ακριβούς και αποτελεσματικού συστήματος πρόβλεψης τιμών μετοχών, το οποίο αξιοποιεί τεχνικές μηχανικής μάθησης και βασίζεται σε ιστορικά χρηματιστηριακά δεδομένα και επιλεγμένους τεχνικούς δείκτες. Το προτεινόμενο σύστημα πρόβλεψης στοχεύει στην ενίσχυση της διαδικασίας λήψης επενδυτικών αποφάσεων σε ένα ασταθές και πολύπλοκο περιβάλλον όπως είναι οι χρηματιστηριακές αγορές. Οι κύριοι στόχοι της εργασίας διαρθρώνονται ως εξής:

- **Ανάπτυξη ενός Αξιόπιστου Συστήματος Πρόβλεψης:** Σχεδίαση ενός ολοκληρωμένου συστήματος που αξιοποιεί προηγμένες τεχνικές μηχανικής μάθησης για την ακριβή πρόβλεψη των τιμών μετοχών, προσφέροντας στους επενδυτές αξιόπιστα εργαλεία ανάλυσης.
- **Αξιολόγηση Απόδοσης Μοντέλων:** Εφαρμογή και συγκριτική αξιολόγηση αλγορίθμων όπως το Random Forest, το SVR και η Linear Regression, με στόχο την ανάδειξη των πιο αποδοτικών μοντέλων.
- **Συμβολή στη Βελτίωση της Στρατηγικής Διαχείρισης Επενδυτικού Ρίσκου:** Παροχή εργαλείων που μειώνουν την αβεβαιότητα, ενισχύουν την πρόβλεψη και υποστηρίζουν καλύτερα ενημερωμένες και τεκμηριωμένες αποφάσεις για την αποφυγή περιττών κινδύνων.
- **Εξερεύνηση Τεχνικών και Δεικτών για Βελτιστοποίηση της Απόδοσης:** Ενσωμάτωση και αξιολόγηση τεχνικών δεικτών, όπως ο Δείκτης Σχετικής Ισχύος (RSI) και οι ΚΜΟ, που ενισχύουν την ακρίβεια των μοντέλων πρόβλεψης.
- **Ανάπτυξη Ενός Ευέλικτου και Αυτοματοποιημένου Εργαλείου:** Ανάπτυξη ενός συστήματος φιλικού προς τον χρήστη, που να προσαρμόζεται εύκολα σε διαφορετικά

σύνολα δεδομένων και νέες συνθήκες αγοράς, υποστηρίζοντας στρατηγικές λήψης αποφάσεων σε πραγματικό χρόνο.

1.3 Μεθοδολογία Προσέγγισης

Η μεθοδολογία της παρούσας εργασίας αναπτύσσεται με στόχο την εφαρμογή ενός συστήματος πρόβλεψης χρηματιστηριακών τιμών μέσω μηχανικής μάθησης, αξιοποιώντας ιστορικά χρηματιστηριακά δεδομένα και τεχνικούς δείκτες. Το σύστημα σχεδιάζεται για να προσφέρει αξιόπιστες και ακριβείς προβλέψεις, υποστηρίζοντας τους επενδυτές στη λήψη τεκμηριωμένων αποφάσεων σε ένα περιβάλλον υψηλής μεταβλητότητας. Η διαδικασία προσέγγισης της μεθοδολογίας περιλαμβάνει τα ακόλουθα στάδια:

- **Συλλογή και Επεξεργασία Δεδομένων:** Τα δεδομένα προέρχονται από ιστορικές χρηματιστηριακές τιμές επιλεγμένων μετοχών και περιλαμβάνουν πληροφορίες, όπως τιμές κλεισίματος, ημερήσιο υψηλό και χαμηλό, και άλλους σχετικούς δείκτες. Η επεξεργασία των δεδομένων ξεκινά με τον καθαρισμό τους, όπου αντιμετωπίζονται τυχόν ελλειπείς τιμές, εξασφαλίζοντας έτσι την ποιότητα και την αξιοπιστία του συνόλου δεδομένων. Κατόπιν, εφαρμόζεται κανονικοποίηση μέσω της τεχνικής Min-Max Scaling, η οποία περιορίζει τις τιμές σε ένα ενιαίο εύρος (συνήθως 0 έως 1) για να διευκολύνει τη διαδικασία εκμάθησης των μοντέλων.
- **Επιλογή και Κατασκευή Χαρακτηριστικών:** Στο στάδιο αυτό, επιλέγονται τεχνικοί δείκτες που μπορούν να ενισχύσουν την πρόβλεψη, όπως ο RSI και οι ΚΜΟ. Αυτοί οι δείκτες ενσωματώνουν πληροφορίες για την τάση και την ορμή της αγοράς, ενισχύοντας την ικανότητα των μοντέλων να αντιλαμβάνονται την κατάσταση και τις τάσεις των μετοχών.
- **Επιλογή Μοντέλων Μηχανικής Μάθησης:** Επιλέγονται και εφαρμόζονται τρία μοντέλα: το Random Forest, το SVR, και η Linear Regression. Αυτά τα μοντέλα επιλέγονται με βάση την αποτελεσματικότητά τους στην πρόβλεψη χρηματοοικονομικών χρονοσειρών, καθώς το καθένα διαθέτει διαφορετική προσέγγιση και ικανότητες προσαρμογής σε σύνθετα μοτίβα δεδομένων.
- **Διαχωρισμός Συνόλων Δεδομένων και Εκπαίδευση Μοντέλων:** Το σύνολο δεδομένων διαχωρίζεται σε σύνολο εκπαίδευσης και σύνολο δοκιμής, διασφαλίζοντας

έτσι ότι τα μοντέλα δεν υπερπροσαρμόζονται και διατηρούν τη δυνατότητα γενίκευσης. Τα μοντέλα εκπαιδεύονται στο σύνολο εκπαίδευσης και οι παράμετροί τους προσαρμόζονται με στόχο τη βελτίωση της ακρίβειας πρόβλεψης.

- **Αξιολόγηση και Σύγκριση Απόδοσης των Μοντέλων:** Μετά την εκπαίδευση, τα μοντέλα αξιολογούνται με χρήση κατάλληλων μετρικών, όπως το Mean Absolute Error (MAE) και το Mean Squared Error (MSE). Οι μετρικές αυτές παρέχουν ακριβή εικόνα για την απόκλιση των προβλεπόμενων τιμών από τις πραγματικές τιμές, επιτρέποντας έτσι την αντικειμενική αξιολόγηση και σύγκριση των μοντέλων.
- **Συμπεράσματα και Συστάσεις για Βελτιώσεις:** Με βάση τα αποτελέσματα, αναλύονται τα πλεονεκτήματα και οι περιορισμοί των χρησιμοποιούμενων αλγορίθμων, καθώς και η καταλληλότητά τους για προβλέψεις σε περιβάλλοντα υψηλής μεταβλητότητας. Προτείνονται, επίσης, βελτιώσεις για μελλοντική έρευνα, όπως η ενσωμάτωση πιο προηγμένων τεχνικών μηχανικής μάθησης (π.χ., βαθιά νευρωνικά δίκτυα) για την περαιτέρω ενίσχυση της ακρίβειας του συστήματος.

1.4 Συνεισφορά

Η διπλωματική αυτή εργασία προσφέρει σημαντικές συνεισφορές στον τομέα της αλγοριθμικής πρόβλεψης και ανάλυσης τιμών μετοχών μέσω της εφαρμογής σύγχρονων τεχνικών μηχανικής μάθησης. Συγκεκριμένα, οι κύριες συνεισφορές συνοψίζονται ως εξής:

- **Ανάπτυξη Ενός Καινοτόμου Συστήματος Πρόβλεψης Τιμών Μετοχών:** Παρουσιάζεται ένα ολοκληρωμένο σύστημα που χρησιμοποιεί ιστορικά δεδομένα και τεχνικούς δείκτες για την πρόβλεψη τιμών. Η εργασία ενσωματώνει αλγορίθμους όπως το Random Forest και το SVR, προσφέροντας μια σύγχρονη προσέγγιση στην ανάλυση χρονοσειρών χρηματοοικονομικών δεδομένων.
- **Αξιολόγηση και Σύγκριση Μοντέλων Μηχανικής Μάθησης:** Μέσα από τη συγκριτική ανάλυση, εντοπίζονται τα δυνατά και αδύναμα σημεία των διαφορετικών αλγορίθμων. Οι μετρικές αξιολόγησης παρέχουν μία λεπτομερή αποτίμηση της ακρίβειας των προβλέψεων, συμβάλλοντας στη βελτίωση της απόδοσης.
- **Στρατηγική Βελτίωσης της Διαχείρισης Ρίσκου:** Το προτεινόμενο σύστημα προσφέρει ακριβέστερες προβλέψεις, υποστηρίζοντας τους επενδυτές στη διαχείριση

κινδύνου, και συμβάλλει στην υιοθέτηση στοχευμένων στρατηγικών με βάση αξιόπιστα δεδομένα.

- **Προτάσεις για Μελλοντική Έρευνα:** Η εργασία παρέχει οδηγίες για την ενσωμάτωση βαθύτερων και πιο σύνθετων μοντέλων, όπως τα νευρωνικά δίκτυα και η ενισχυτική μάθηση, με στόχο τη βελτίωση της ακρίβειας του συστήματος στο μέλλον.

1.5 Διάρθρωση της Διπλωματικής Εργασίας

Η δομή της παρούσας διπλωματικής εργασίας περιλαμβάνει πέντε κύρια κεφάλαια που καλύπτουν συστηματικά τα βασικά στάδια της έρευνας και της ανάλυσης.

- **Κεφάλαιο 1 - Εισαγωγή:** Παρουσιάζει το πλαίσιο του θέματος, τη σημασία του, τους στόχους της εργασίας, τη μεθοδολογία που ακολουθήθηκε, και τη συνεισφορά της στο πεδίο της αλγοριθμικής συναλλαγής και πρόβλεψης τιμών μετοχών.
- **Κεφάλαιο 2 - Θεωρητικό Υπόβαθρο:** Αναλύονται οι βασικές έννοιες της αλγοριθμικής συναλλαγής και της χρηματοοικονομικής ανάλυσης με μηχανική μάθηση, εστιάζοντας σε τεχνικές πρόβλεψης που εφαρμόζονται στις χρονοσειρές και στους χρηματιστηριακούς δείκτες.
- **Κεφάλαιο 3 - Υλοποίηση και Σχεδίαση του Συστήματος:** Περιγράφεται η ανάπτυξη του συστήματος πρόβλεψης τιμών, η διαδικασία προεπεξεργασίας δεδομένων, η επιλογή των αλγορίθμων και των εργαλείων που χρησιμοποιήθηκαν, και η αρχιτεκτονική του συστήματος.
- **Κεφάλαιο 4 - Αξιολόγηση Μοντέλων και Ανάλυση Αποτελεσμάτων:** Παρουσιάζονται τα αποτελέσματα της αξιολόγησης, η σύγκριση της απόδοσης των αλγορίθμων και η ανάλυση των αποτελεσμάτων, χρησιμοποιώντας μετρικές όπως το MAE και το MSE.
- **Κεφάλαιο 5 - Συμπεράσματα και Προοπτικές:** Το τελικό κεφάλαιο συνοψίζει τα βασικά συμπεράσματα της εργασίας και προτείνει κατευθύνσεις για μελλοντική έρευνα, όπως η ενσωμάτωση πιο σύνθετων μοντέλων για βελτίωση της ακρίβειας στις προβλέψεις.

2

Θεωρητικό Υπόβαθρο

2.1 Αλγοριθμική Συναλλαγή και Χρηματιστηριακή Ανάλυση

Η αλγοριθμική συναλλαγή (Algorithmic Trading) είναι μία σύγχρονη τεχνική που επιτρέπει την αυτοματοποίηση των συναλλαγών μέσω προγραμματισμένων αλγορίθμων, οι οποίοι εκτελούν εντολές με ταχύτητα, ακρίβεια και αποτελεσματικότητα. Η βασική διαφορά της σε σχέση με τις παραδοσιακές μεθόδους συναλλαγών είναι η ελαχιστοποίηση της ανθρώπινης παρέμβασης, η οποία συχνά επηρεάζεται από συναισθηματικούς ή υποκειμενικούς παράγοντες. Η αλγοριθμική συναλλαγή επιτρέπει την ανάλυση πολύπλοκων δεδομένων σε πραγματικό χρόνο, γεγονός που βοηθά στον εντοπισμό βραχυπρόθεσμων ευκαιριών και στην ταχύτερη λήψη αποφάσεων. Οι χρηματοοικονομικές αγορές, με την πολυπλοκότητα και την υψηλή μεταβλητότά τους, απαιτούν την υιοθέτηση εξελιγμένων εργαλείων, όπως οι αλγόριθμοι συναλλαγών, για την ενίσχυση των επενδυτικών στρατηγικών και την επίτευξη ανταγωνιστικών αποτελεσμάτων. Η χρηματιστηριακή ανάλυση αποτελεί μία από τις κύριες μεθόδους κατανόησης των αγορών και λήψης στρατηγικών αποφάσεων. Διακρίνεται σε δύο κύριες κατηγορίες:

- **Θεμελιώδης Ανάλυση:** Η μέθοδος αυτή βασίζεται στην αξιολόγηση των οικονομικών στοιχείων μιας εταιρείας, όπως τα έσοδα, τα κέρδη, οι ταμειακές ροές

και οι μελλοντικές προοπτικές. Εξετάζει τις εσωτερικές αξίες των μετοχών και συγκρίνει τις πραγματικές τιμές με αυτές για να καθορίσει εάν η μετοχή είναι υποτιμημένη ή υπερτιμημένη.

- **Τεχνική Ανάλυση:** Στηρίζεται στην αξιολόγηση ιστορικών δεδομένων, όπως οι τιμές και οι όγκοι συναλλαγών. Οι επενδυτές χρησιμοποιούν τεχνικούς δείκτες για την πρόβλεψη μελλοντικών κινήσεων των τιμών, αξιοποιώντας τάσεις και μοτίβα που εντοπίζονται στις χρονοσειρές.

Η συνδυαστική χρήση αυτών των δύο μεθόδων προσφέρει μία πιο ολοκληρωμένη προσέγγιση στις επενδυτικές στρατηγικές, καθώς συνδυάζει τη μακροπρόθεσμη προοπτική της θεμελιώδους ανάλυσης με τη βραχυπρόθεσμη προβλεψιμότητα της τεχνικής ανάλυσης.

2.1.1 Εισαγωγή στην Αλγοριθμική Συναλλαγή

Η αλγοριθμική συναλλαγή αναφέρεται στη χρήση υπολογιστικών αλγορίθμων για την αυτόματη λήψη και εκτέλεση επενδυτικών αποφάσεων σε χρηματιστηριακές αγορές. Στόχος της είναι να βελτιστοποιήσει τη διαδικασία συναλλαγών, επιτρέποντας την ταχύτατη και ακριβή εκτέλεση εντολών. Η αλγοριθμική συναλλαγή έχει καθιερωθεί ως απαραίτητο εργαλείο για τους επενδυτές λόγω της ικανότητάς της να αναλύει μεγάλα δεδομένα σε πραγματικό χρόνο και να εκμεταλλεύεται μικρές αλλά σημαντικές ευκαιρίες της αγοράς.

Ανάγκη για Αλγοριθμικές Στρατηγικές

Η αβεβαιότητα και η έντονη μεταβλητότητα που χαρακτηρίζουν τις αγορές αποτελούν σοβαρές προκλήσεις για τους επενδυτές. Η αλγοριθμική συναλλαγή επιτρέπει την ελαχιστοποίηση της ανθρώπινης παρέμβασης και των πιθανών συναισθηματικών επιδράσεων, οδηγώντας σε πιο αντικειμενικές αποφάσεις. Επίσης, η δυνατότητα των αλγορίθμων να εκτελούν στρατηγικές σε κλάσματα του δευτερολέπτου προσφέρει μεγάλο πλεονέκτημα σε σύγκριση με τις παραδοσιακές μεθόδους [2].

Κατηγορίες Αλγοριθμικής Στρατηγικής

Οι κύριες κατηγορίες αλγορίθμων συναλλαγής περιλαμβάνουν:

- **Αλγόριθμοι Βασισμένοι σε Στατιστικά Μοντέλα:** Χρησιμοποιούν ιστορικά δεδομένα και μοντέλα πρόβλεψης για τον εντοπισμό πιθανών μελλοντικών τάσεων, όπως οι κινητοί μέσοι όροι και η γραμμική παλινδρόμηση.

- **Αλγόριθμοι Υψηλής Συχνότητας (High-Frequency Trading - HFT):** Εκτελούν εκατοντάδες έως χιλιάδες συναλλαγές σε δευτερόλεπτα, εκμεταλλευόμενοι μικρές ανισορροπίες στις τιμές των μετοχών και στοχεύοντας στην επίτευξη γρήγορου κέρδους.
- **Αλγόριθμοι Arbitrage:** Εκμεταλλεύονται τις διαφορές τιμών που παρατηρούνται ταυτόχρονα σε διάφορες αγορές ή πλατφόρμες συναλλαγών. Οι αλγόριθμοι αυτοί επιτρέπουν τη γρήγορη και ασφαλή πραγματοποίηση συναλλαγών με μικρό ρίσκο.
- **Αλγόριθμοι Βελτιστοποίησης Ρευστότητας:** Σχεδιάζονται για τη μεγιστοποίηση της ρευστότητας στις αγορές, μειώνοντας τις επιπτώσεις των μεγάλων παραγγελιών στην αγορά.

Αυτές οι κατηγορίες αλγορίθμων επιτρέπουν την αποτελεσματική διαχείριση επενδυτικών χαρτοφυλακίων και παρέχουν εργαλεία για την αντιμετώπιση της πολυπλοκότητας και της διαρκούς μεταβολής των αγορών.

Τεχνολογίες και Πλατφόρμες για Αλγοριθμική Συναλλαγή

Η αλγοριθμική συναλλαγή απαιτεί προηγμένα τεχνολογικά εργαλεία, όπως:

- **Γλώσσες Προγραμματισμού:** Η Python και η R είναι ιδιαίτερα διαδεδομένες, ενώ προσφέρουν πλούσιες βιβλιοθήκες για ανάλυση δεδομένων και μοντελοποίηση.
- **Συστήματα Διαχείρισης Δεδομένων (DMS):** Χρησιμοποιούνται για την αποθήκευση και γρήγορη ανάκτηση μεγάλων δεδομένων σε πραγματικό χρόνο.
- **Πλατφόρμες Ανάλυσης Δεδομένων:** Η χρήση εργαλείων, όπως το Jupyter Notebook και το Apache Spark, βοηθά στην ανάλυση μεγάλων συνόλων δεδομένων και στη δημιουργία σύνθετων μοντέλων πρόβλεψης.

Ιστορική Εξέλιξη

Η έννοια της αυτοματοποιημένης συναλλαγής ξεκίνησε τη δεκαετία του 1970 με την εισαγωγή των πρώτων ηλεκτρονικών συστημάτων συναλλαγών στις αγορές των ΗΠΑ. Ωστόσο, η σημαντική υιοθέτηση της αλγοριθμικής συναλλαγής σημειώθηκε τη δεκαετία του 2000, όταν η τεχνολογία επέτρεψε συναλλαγές υψηλής συχνότητας (HFT). Σύμφωνα με τον Hendershott et al. (2011), η χρήση των αλγορίθμων αυξήθηκε ραγδαία, φτάνοντας το 70-80% των συναλλαγών σε χρηματιστήρια όπως το NYSE και το NASDAQ [2]. Η αποτελεσματικότητα της αλγοριθμικής συναλλαγής οφείλεται κυρίως στα παρακάτω:

- **Μείωση κόστους συναλλαγών:** Ελαχιστοποίηση χειροκίνητων λαθών και δαπανών.

- **Ταχύτητα και ακρίβεια:** Οι εντολές εκτελούνται σε χιλιοστά του δευτερολέπτου.
- **Διαχείριση μεγάλου όγκου δεδομένων:** Οι αλγόριθμοι είναι ικανοί να αναλύουν χιλιάδες συναλλαγές ταυτόχρονα.

2.1.2 Βασικές Έννοιες Χρηματιστηριακής Ανάλυσης

Η χρηματιστηριακή ανάλυση αποτελεί τον πυρήνα της στρατηγικής επενδυτών και διαχειριστών κεφαλαίων, καθώς επιτρέπει τη διερεύνηση και κατανόηση των μοτίβων, των τάσεων και της συμπεριφοράς της αγοράς. Η ανάλυση αυτή κατηγοριοποιείται σε δύο κύριες μεθόδους, οι οποίες παρέχουν συμπληρωματικές οπτικές για την αποτίμηση των μετοχών και τη λήψη επενδυτικών αποφάσεων.

Θεμελιώδης Ανάλυση

Η θεμελιώδης ανάλυση εστιάζει στα εσωτερικά χαρακτηριστικά και στην οικονομική υγεία μιας εταιρείας για την αποτίμηση των μετοχών της. Βασίζεται σε οικονομικούς δείκτες και αναλύει παράγοντες όπως τα κέρδη, τα έσοδα, τις ταμειακές ροές, και τις υποχρεώσεις μιας εταιρείας. Η διαδικασία περιλαμβάνει την εξέταση της αξίας της μετοχής με βάση τη σύγκριση της πραγματικής τιμής της αγοράς με την εκτιμώμενη “εσωτερική αξία”. Όταν η αγορά εκτιμά μια μετοχή υψηλότερα ή χαμηλότερα από την εσωτερική της αξία, τότε θεωρείται ότι η μετοχή είναι υπερτιμημένη ή υποτιμημένη αντίστοιχα, δίνοντας στους επενδυτές σημαντικές ενδείξεις για αγορές ή πωλήσεις.

Τεχνική Ανάλυση

Η τεχνική ανάλυση είναι η μέθοδος που χρησιμοποιεί τα ιστορικά δεδομένα τιμών και όγκων για την πρόβλεψη μελλοντικών κινήσεων των μετοχών, βασιζόμενη στην υπόθεση ότι η ιστορική συμπεριφορά της αγοράς θα επαναληφθεί. Οι επενδυτές αναλύουν γραφήματα τιμών και χρησιμοποιούν τεχνικούς δείκτες, όπως οι ΚΜΟ και ο RSI, για να εντοπίσουν τάσεις και πιθανά σημεία αντιστροφής της αγοράς. Η τεχνική ανάλυση παρέχει μια οπτική αναπαράσταση των ψυχολογικών παραμέτρων και των μοτίβων που διαμορφώνονται από τις συναλλαγές.

Ενοποιημένη Προσέγγιση

Η συνδυασμένη χρήση της θεμελιώδους και της τεχνικής ανάλυσης προσφέρει στους επενδυτές μια ολοκληρωμένη εικόνα για τη διαχείριση επενδυτικών στρατηγικών. Η θεμελιώδης ανάλυση παρέχει μακροπρόθεσμες προοπτικές βασισμένες στην πραγματική αξία

μιας μετοχής, ενώ η τεχνική ανάλυση επιτρέπει τη βραχυπρόθεσμη εκμετάλλευση των κινήσεων της αγοράς, βελτιώνοντας την απόδοση και μειώνοντας τους κινδύνους [3] [4].

2.2 Μηχανική Μάθηση στις Προβλέψεις Χρηματιστηριακών Τιμών

Η μηχανική μάθηση (Machine Learning - ML) αποτελεί μία από τις πιο σημαντικές και ραγδαία εξελισσόμενες τεχνολογίες στον τομέα της χρηματοοικονομικής ανάλυσης. Χρησιμοποιείται εκτενώς για την πρόβλεψη τιμών μετοχών, καθώς προσφέρει τη δυνατότητα ανάλυσης μεγάλου όγκου δεδομένων (Big Data) και την ανίχνευση κρυφών μοτίβων που δεν είναι ορατά με παραδοσιακές μεθόδους. Η πρόβλεψη των τιμών των χρηματοοικονομικών προϊόντων είναι μια πολύπλοκη διαδικασία, καθώς οι αγορές επηρεάζονται από πληθώρα μεταβλητών και χαρακτηρίζονται από στοχαστική συμπεριφορά, μη γραμμικές σχέσεις και υψηλή μεταβλητότητα. Οι αλγόριθμοι μηχανικής μάθησης παρέχουν εξελιγμένα εργαλεία που μπορούν να:

- Μάθουν από ιστορικά δεδομένα για να προβλέψουν μελλοντικές τάσεις.
- Προσαρμόστουν δυναμικά σε νέες συνθήκες αγοράς.
- Βελτιστοποιήσουν τη λήψη αποφάσεων σε πραγματικό χρόνο, καθιστώντας τις προβλέψεις ακριβέστερες.

Η δυνατότητα αυτή είναι ιδιαίτερα κρίσιμη σε έναν τομέα όπου η ταχύτητα και η ακρίβεια των αποφάσεων μπορούν να κάνουν τη διαφορά μεταξύ κέρδους και ζημίας [5].

2.2.1 Βασικές Έννοιες Μηχανικής Μάθησης

Η μηχανική μάθηση βασίζεται στη δημιουργία μοντέλων που "εκπαιδεύονται" με βάση δεδομένα και χρησιμοποιούνται για τη γενίκευση σε νέα, άγνωστα δεδομένα. Οι κύριες κατηγορίες της μηχανικής μάθησης είναι οι εξής:

- **Εποπτευόμενη Μάθηση (Supervised Learning):** Στην εποπτευόμενη μάθηση, τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση των μοντέλων περιέχουν γνωστές τιμές-στόχους, καθιστώντας δυνατή την εκπαίδευση αλγορίθμων που μαθαίνουν να

αντιστοιχούν τις εισόδους με τις επιθυμητές εξόδους. Η Linear Regression, το Random Forest και το SVR είναι παραδείγματα εποπτευόμενων αλγορίθμων που συχνά εφαρμόζονται στην πρόβλεψη τιμών μετοχών, επιτρέποντας τη μοντελοποίηση των τιμών με βάση την ιστορική τους συμπεριφορά [6].

- **Μη Εποπτευόμενη Μάθηση (Unsupervised Learning):** Στη μη εποπτευόμενη μάθηση, τα μοντέλα αναλύουν δεδομένα χωρίς καθορισμένες τιμές-στόχους, ανιχνεύοντας κρυφά μοτίβα ή ομάδες δεδομένων. Οι αλγόριθμοι μη εποπτευόμενης μάθησης, όπως οι αλγόριθμοι συσταδοποίησης (π.χ. K-means clustering), χρησιμοποιούνται για τον εντοπισμό ανωμαλιών ή την ομαδοποίηση δεδομένων με παρόμοιες ιδιότητες [7].
- **Ενισχυτική Μάθηση (Reinforcement Learning):** Η ενισχυτική μάθηση βασίζεται σε έναν δυναμικό μηχανισμό βελτιστοποίησης, όπου το μοντέλο «μαθαίνει» μέσω ενός συστήματος ανταμοιβών και ποινών. Η προσέγγιση αυτή είναι ιδιαίτερα κατάλληλη για εφαρμογές όπου το αποτέλεσμα των αποφάσεων εξαρτάται από διαδοχικά βήματα, όπως οι συναλλαγές σε χρηματοοικονομικές αγορές. Η ενισχυτική μάθηση, μέσω αλγορίθμων όπως το Q-learning και το Deep Reinforcement Learning, έχει αποδείξει την αποτελεσματικότητά της σε περιβάλλοντα με υψηλή αβεβαιότητα [8].

Αυτές οι βασικές κατηγορίες ενσωματώνουν έναν συνδυασμό προσεγγίσεων που επιτρέπουν στους αλγόριθμους να επεξεργάζονται τεράστιες ποσότητες δεδομένων και να παρέχουν προβλέψεις, προσφέροντας πιο αξιόπιστη και τεκμηριωμένη κατανόηση των αγορών [9].

2.2.2 Εφαρμογές Μηχανικής Μάθησης στη Χρηματοοικονομική Ανάλυση

Οι τεχνικές μηχανικής μάθησης χρησιμοποιούνται εκτενώς για την πρόβλεψη και ανάλυση χρηματοοικονομικών δεδομένων [10]. Οι κύριες εφαρμογές περιλαμβάνουν:

- **Πρόβλεψη Χρονοσειρών (Time Series Forecasting):** Οι αλγόριθμοι πρόβλεψης χρονοσειρών αποτελούν κεντρικό εργαλείο στη μηχανική μάθηση για χρηματοοικονομικές εφαρμογές. Εστιάζουν στην ανάλυση ιστορικών δεδομένων, όπως οι τιμές κλεισίματος και οι όγκοι συναλλαγών, και χρησιμοποιούνται για να

προβλέψουν μελλοντικές τιμές. Ειδικότερα, αλγόριθμοι όπως το Random Forest και το SVR έχουν αποδείξει την αποτελεσματικότητά τους στην ανάλυση σύνθετων χρηματιστηριακών χρονοσειρών.

- **Ανάλυση Συναισθήματος (Sentiment Analysis):** Η ανάλυση συναισθήματος εξετάζει την αντίδραση των επενδυτών στις ειδήσεις και τις τάσεις της αγοράς, χρησιμοποιώντας αλγόριθμους επεξεργασίας φυσικής γλώσσας (NLP). Με την αξιοποίηση δεδομένων από κοινωνικά δίκτυα και ειδησεογραφικά μέσα, η ανάλυση συναισθήματος μπορεί να προσφέρει συμπληρωματικές πληροφορίες για τις πιθανές κινήσεις των τιμών, βελτιώνοντας την ακρίβεια των προβλέψεων.
- **Ανίχνευση Ανωμαλιών (Anomaly Detection):** Στη χρηματοοικονομική ανάλυση, η ανίχνευση ανωμαλιών χρησιμοποιείται για τον εντοπισμό απρόβλεπτων ή ύποπτων συμπεριφορών στις αγορές, όπως οι απότομες αυξομειώσεις τιμών που μπορεί να υποδεικνύουν πιθανές κερδοσκοπικές κινήσεις. Αυτή η τεχνική μπορεί να συμβάλλει στη μείωση του κινδύνου, επιτρέποντας στους επενδυτές να εντοπίζουν επικίνδυνες τάσεις πριν από την εκδήλωσή τους.

Οι εφαρμογές αυτές ενισχύουν την ανάλυση των αγορών, υποστηρίζοντας τη λήψη στρατηγικών αποφάσεων και βελτιώνοντας τις επενδυτικές στρατηγικές μέσω πιο ακριβών και δυναμικών μοντέλων πρόβλεψης [11].

2.3 Συστήματα και Τεχνικές Πρόβλεψης Χρηματιστηριακών Τιμών

Η πρόβλεψη των τιμών μετοχών είναι μια από τις πιο δύσκολες προκλήσεις στον τομέα της χρηματοοικονομικής ανάλυσης. Ο λόγος για τη δυσκολία αυτή είναι ότι η αγορά επηρεάζεται από πολλούς και συχνά απρόβλεπτους παράγοντες. Από τις οικονομικές επιδόσεις μιας εταιρείας μέχρι τις γεωπολιτικές εξελίξεις και τις συνθήκες προσφοράς και ζήτησης, οι παράγοντες αυτοί δημιουργούν ένα περίπλοκο και δυναμικό περιβάλλον που είναι δύσκολο να αποτυπωθεί με ακρίβεια. Η αστάθεια των αγορών αποτελεί έναν από τους βασικούς λόγους που οι παραδοσιακές μέθοδοι πρόβλεψης, όπως η θεμελιώδης και η τεχνική ανάλυση, συχνά αποδεικνύονται ανεπαρκείς. Αυτές οι μέθοδοι βασίζονται κυρίως σε ιστορικά

δεδομένα και δείκτες, γεγονός που περιορίζει την ικανότητά τους να προσαρμοστούν σε νέα δεδομένα ή να προβλέψουν απρόβλεπτες αλλαγές στις αγορές. Επομένως, γίνεται σαφές ότι χρειάζονται πιο σύγχρονες και ευέλικτες προσεγγίσεις για την αντιμετώπιση αυτών των προκλήσεων. Σε αυτό το πλαίσιο, η μηχανική μάθηση έχει αναδειχθεί ως μία από τις πιο υποσχόμενες λύσεις. Χρησιμοποιώντας δεδομένα σε πραγματικό χρόνο, τα συστήματα που βασίζονται στη μηχανική μάθηση μπορούν να εντοπίσουν τάσεις και μοτίβα που δεν είναι εύκολα ορατά μέσω παραδοσιακών αναλυτικών εργαλείων. Επιπλέον, προσφέρουν τη δυνατότητα συνεχούς προσαρμογής στις μεταβαλλόμενες συνθήκες της αγοράς, γεγονός που τις καθιστά ιδανικές για τη διαχείριση της πολυπλοκότητας και της αστάθειας που χαρακτηρίζουν τις χρηματοοικονομικές αγορές. Τα πιο σύγχρονα συστήματα πρόβλεψης ενσωματώνουν αλγόριθμους μηχανικής μάθησης και τεχνικούς δείκτες, όπως η ανάλυση τάσεων και η ανίχνευση ανωμαλιών. Αυτοί οι συνδυασμοί δίνουν τη δυνατότητα να εντοπιστούν ευκαιρίες και να προβλεφθούν κίνδυνοι τόσο σε μακροπρόθεσμο όσο και σε βραχυπρόθεσμο ορίζοντα. Το αποτέλεσμα είναι ότι οι επενδυτές μπορούν να λαμβάνουν πιο ενημερωμένες αποφάσεις, μειώνοντας τον κίνδυνο και αυξάνοντας τις πιθανότητες επιτυχίας στις στρατηγικές τους. Η ανάπτυξη τέτοιων συστημάτων δεν είναι απλώς τεχνική πρόκληση. Απαιτεί επίσης τη βαθιά κατανόηση των αγορών και των μηχανισμών που τις διέπουν, συνδυάζοντας στατιστική ανάλυση, τεχνολογία και γνώση του αντικειμένου. Αυτό καθιστά τη μελέτη τους όχι μόνο ενδιαφέρουσα αλλά και κρίσιμη για την περαιτέρω εξέλιξη των επενδυτικών εργαλείων [12].

2.3.1 Προκλήσεις και Στρατηγικές για Πρόβλεψη Χρονοσειρών

Η πρόβλεψη χρονοσειρών στη χρηματιστηριακή ανάλυση περιλαμβάνει την εκτίμηση μελλοντικών τιμών μετοχών βάσει ιστορικών δεδομένων και τάσεων. Ωστόσο, οι προκλήσεις σε αυτόν τον τομέα είναι πολλαπλές και σύνθετες, εξαιτίας των εξωτερικών παραγόντων και των διακυμάνσεων που επηρεάζουν τις χρηματιστηριακές αγορές. Συγκεκριμένα, οι τιμές των μετοχών επηρεάζονται από:

- **Μακροοικονομικές συνθήκες:** Αλλαγές στη νομισματική πολιτική, διακυμάνσεις στα επιτόκια και στο επίπεδο του πληθωρισμού επηρεάζουν τη συνολική οικονομική δραστηριότητα και τη ρευστότητα στις αγορές.

- **Εταιρικά αποτελέσματα και ανακοινώσεις:** Οι οικονομικές ανακοινώσεις των εταιρειών, όπως κέρδη και ζημίες, επηρεάζουν άμεσα τη ζήτηση για τις μετοχές τους, καθώς οι επενδυτές προσαρμόζουν τις εκτιμήσεις τους για τη μελλοντική απόδοση της εταιρείας.
- **Ψυχολογία της αγοράς:** Η συμπεριφορά των επενδυτών, όπως η τάση για πανικό σε περιόδους κρίσεων ή για υπερβολική αισιοδοξία σε περιόδους άνθησης, προκαλεί διακυμάνσεις που δεν βασίζονται σε θεμελιώδη στοιχεία.

Για την αντιμετώπιση αυτών των προκλήσεων, οι στρατηγικές πρόβλεψης χρονοσειρών ενσωματώνουν στατιστικά μοντέλα και τεχνικές μηχανικής μάθησης που προσφέρουν μια πιο ολοκληρωμένη προσέγγιση. Ορισμένες από τις πιο σημαντικές στρατηγικές περιλαμβάνουν:

- **Ανάλυση Εποχικότητας:** Οι επενδυτές μελετούν επαναλαμβανόμενα μοτίβα που σχετίζονται με συγκεκριμένες περιόδους του χρόνου. Για παράδειγμα, η ανοδική κίνηση των μετοχών προς το τέλος του έτους λόγω αυξημένης κατανάλωσης μπορεί να χρησιμοποιηθεί για βραχυπρόθεσμες στρατηγικές.
- **Αφαίρεση Θορύβου από τα Δεδομένα:** Οι αυξομειώσεις που δεν συνδέονται με θεμελιώδεις παράγοντες απομακρύνονται ώστε τα μοντέλα να επικεντρώνονται στις τάσεις, αποφεύγοντας τις διακυμάνσεις που οδηγούν σε λανθασμένες εκτιμήσεις
- **Προσθήκη Τεχνικών Δεικτών:** Οι τεχνικοί δείκτες, όπως οι ΚΜΟ και ο RSI, συνδυάζονται με την ανάλυση χρονοσειρών για να προσφέρουν επιπλέον δεδομένα σχετικά με την τάση της αγοράς, την ισχύ της ζήτησης και τη μεταβλητότητα της μετοχής.

Αυτές οι στρατηγικές επιτρέπουν στα συστήματα πρόβλεψης να είναι πιο ανθεκτικά και προσαρμόσιμα, εστιάζοντας σε τάσεις που έχουν διάρκεια και παρακάμπτοντας τις τυχαίες διακυμάνσεις.

2.3.2 Τεχνικοί Δείκτες και Αναλυτικές Τεχνικές

Η τεχνική ανάλυση, ως προσέγγιση μελέτης των αγορών, βασίζεται στη χρήση διαγραμμάτων και τεχνικών δεικτών για την ανίχνευση τάσεων και τη λήψη αποφάσεων. Οι τεχνικοί δείκτες χρησιμοποιούνται ευρέως στην αλγοριθμική συναλλαγή, καθώς παρέχουν ένα πλαίσιο ανάλυσης των ιστορικών δεδομένων των μετοχών και ενισχύουν τη δυνατότητα πρόβλεψης

των μελλοντικών κινήσεων της αγοράς [13]. Σε αυτήν την ενότητα αναλύονται οι βασικοί τεχνικοί δείκτες που μελετήθηκαν και ενσωματώνονται στο πλαίσιο του θεωρητικού υποβάθρου της εργασίας.:

- **Κινητοί Μέσοι Όροι (Moving Averages - MA):** Οι ΚΜΟ είναι από τους πιο διαδεδομένους δείκτες και χρησιμοποιούνται για την εξομάλυνση των διακυμάνσεων των τιμών. Ο Απλός Κινητός Μέσος (SMA) και ο Εκθετικός Κινητός Μέσος (EMA) δίνουν μια εικόνα της τάσης της αγοράς. Οι ΚΜΟ χρησιμοποιούνται για να εντοπίζονται πιθανά σημεία εισόδου και εξόδου.
- **Απλός Κινητός Μέσος (SMA):** Ο SMA υπολογίζει τον αριθμητικό μέσο όρο των τιμών κλεισίματος μιας μετοχής σε ένα συγκεκριμένο χρονικό διάστημα (π.χ., 10, 50 ή 200 ημερών). Παρέχει καθαρή εικόνα των μακροχρόνιων τάσεων, εξομαλύνοντας τις βραχυπρόθεσμες διακυμάνσεις.

$$SMA = \frac{(P^1 + P^2 + \dots + P_n)}{n}$$

όπου P είναι η τιμή κλεισίματος και n είναι το πλήθος των περιόδων.

- **Εκθετικός Κινητός Μέσος (EMA):** Ο EMA δίνει μεγαλύτερη βαρύτητα στις πιο πρόσφατες τιμές, καθιστώντας τον πιο ευαίσθητο σε βραχυπρόθεσμες αλλαγές της αγοράς. Αυτό τον καθιστά ιδανικό για την ανίχνευση τάσεων σε δυναμικά περιβάλλοντα, όπως οι χρηματιστηριακές αγορές [14].

$$EMA_t = P_t \left(\frac{2}{n+1} \right) + EMA_{t-1} \left(1 - \left(\frac{2}{n+1} \right) \right)$$

όπου EMA_t είναι η εκθετική μέση τιμή στην περίοδο t , P_t η τρέχουσα τιμή και n το πλήθος των περιόδων.

- **RSI:** Ο RSI παρέχει πληροφορίες για την ορμή της αγοράς, δείχνοντας αν η μετοχή είναι υπερτιμημένη ή υποτιμημένη. Ο RSI κυμαίνεται από 0 έως 100 και ερμηνεύεται ως εξής:
 - Τιμές πάνω από 70: Η μετοχή θεωρείται υπερτιμημένη (Overbought), με πιθανότητα καθοδικής διόρθωσης.
 - Τιμές κάτω από 30: Η μετοχή θεωρείται υποτιμημένη (Oversold), με πιθανότητα ανοδικής αντιστροφής [15].

Ο RSI συνδυάζεται συχνά με άλλους δείκτες, όπως οι ΚΜΟ, για την καλύτερη αξιολόγηση των τάσεων και την αποφυγή λανθασμένων σημάτων [16].

$$RSI = 100 - \left(\frac{100}{1 + RS} \right)$$

$$RS = \frac{\text{Μέσος Όρος Κερδών (Gain)}}{\text{Μέσος Όρος Απωλειών (Loss)}}$$

όπου RS είναι η σχέση μεταξύ του μέσου όρου των κερδών και του μέσου όρου των απωλειών σε μια χρονική περίοδο (συνήθως 14 ημερών).

- **MACD (Moving Average Convergence Divergence):** Ο MACD συνδυάζει δύο κινητούς μέσους όρους για την ανάλυση της σύγκλισης και της απόκλισης, εντοπίζοντας την κατεύθυνση και την ισχύ της τάσης. Χρησιμοποιείται ευρέως για την εκτίμηση της τάσης και του δυναμισμού της αγοράς.

$$MACD = EMA(\gamma\rho\eta\gamma\omicron\rho\omicron) - EMA(\alpha\rho\gamma\omicron)$$

- **Bollinger Bands:** Οι ζώνες Bollinger προσαρμόζονται στις αλλαγές της μεταβλητότητας και υποδεικνύουν πιθανές αλλαγές στην πορεία της τιμής. Οι ζώνες αυτές βασίζονται στον κινητό μέσο όρο και μια τυπική απόκλιση, δημιουργώντας ένα εύρος κίνησης για την τιμή που βοηθά στον εντοπισμό ανατροπών και πιθανών σημείων εισόδου ή εξόδου [17]. Οι ζώνες Bollinger βασίζονται στον κινητό μέσο όρο και στην τυπική απόκλιση των τιμών:

- Άνω ζώνη (Upper Band):

$$UB = SMA + k * \sigma$$

- Κάτω ζώνη (Lower Band):

$$LB = SMA - k * \sigma$$

όπου k είναι ο συντελεστής τυπικής απόκλισης (τυπικά $k=2$) και σ είναι η τυπική απόκλιση των τιμών για την επιλεγμένη περίοδο.

Οι δείκτες αυτοί, σε συνδυασμό με τα συστήματα πρόβλεψης, παρέχουν ένα ολοκληρωμένο πλαίσιο για την ανάλυση των χρηματοοικονομικών αγορών και τη λήψη στρατηγικών αποφάσεων. Διευκολύνουν την κατανόηση των διακυμάνσεων, βελτιώνουν την ακρίβεια των προβλέψεων και ενισχύουν τις επενδυτικές στρατηγικές με στόχο τη μεγιστοποίηση των αποδόσεων και τη διαχείριση του ρίσκου.

Συμπερασματική Επισκόπηση Τεχνικών Δεικτών

Η ενσωμάτωση τεχνικών δεικτών, όπως οι KMO, ο RSI, ο MACD και οι ζώνες Bollinger, ενισχύει τη δυνατότητα ανάλυσης των χρηματιστηριακών δεδομένων. Η χρήση τους στο πλαίσιο του θεωρητικού υποβάθρου προσφέρει μία ολοκληρωμένη προσέγγιση για:

- Την ανίχνευση των τάσεων της αγοράς.
- Τον εντοπισμό κρίσιμων σημείων αντιστροφής των τιμών.
- Την κατανόηση της μεταβλητότητας και της ορμής των μετοχών.

3

Υλοποίηση και Σχεδίαση του Συστήματος

Το τρίτο κεφάλαιο επικεντρώνεται στη διαδικασία σχεδίασης και υλοποίησης ενός συστήματος πρόβλεψης τιμών μετοχών με τη χρήση τεχνικών μηχανικής μάθησης. Η ανάπτυξη ενός τέτοιου συστήματος απαιτεί μια σαφή δομή που περιλαμβάνει την προεπεξεργασία των δεδομένων, την επιλογή κατάλληλων αλγορίθμων, και τη διαμόρφωση της αρχιτεκτονικής που θα επιτρέψει την ακριβή πρόβλεψη τιμών μετοχών σε δυναμικές αγορές. Καθώς οι χρηματιστηριακές αγορές χαρακτηρίζονται από έντονη μεταβλητότητα, το σύστημα πρόβλεψης πρέπει να είναι ανθεκτικό, να προσαρμόζεται στα νέα δεδομένα και να διατηρεί την ακρίβειά του. Η διαδικασία σχεδίασης περιλαμβάνει μια σειρά από κρίσιμα βήματα. Αρχικά, πραγματοποιείται η προεπεξεργασία των δεδομένων, προκειμένου να εξασφαλιστεί η ποιότητά τους και να δημιουργηθεί ένα σύνολο δεδομένων κατάλληλο για την εκπαίδευση των αλγορίθμων. Στη συνέχεια, επιλέγονται συγκεκριμένα χαρακτηριστικά και τεχνικοί δείκτες που ενισχύουν την ανάλυση και βελτιώνουν τις προβλέψεις του συστήματος. Επιπλέον, εξετάζεται η αρχιτεκτονική του συστήματος και τα εργαλεία που θα χρησιμοποιηθούν για την υλοποίησή του, ώστε να διασφαλιστεί η αποτελεσματικότητα και η λειτουργικότητά του.

3.1 Δομή και Στόχοι του Συστήματος Πρόβλεψης

Το σύστημα πρόβλεψης τιμών μετοχών που σχεδιάστηκε για αυτήν την εργασία έχει ως κύριο στόχο την παροχή ακριβών εκτιμήσεων για τις μελλοντικές τιμές των μετοχών μέσω της ανάλυσης ιστορικών δεδομένων και τεχνικών δεικτών. Η αλγοριθμική πρόβλεψη χρηματιστηριακών τιμών αποτελεί μια απαραίτητη τεχνική στη σύγχρονη χρηματοοικονομική ανάλυση, καθώς προσφέρει τη δυνατότητα να λαμβάνονται αποφάσεις με βάση αντικειμενικά δεδομένα, μειώνοντας τις πιθανότητες για σφάλματα και αυξάνοντας την αποδοτικότητα των επενδυτικών στρατηγικών. Η ανάπτυξη του συστήματος πρόβλεψης επικεντρώνεται σε συγκεκριμένα στάδια που διασφαλίζουν την αποδοτικότητα και την ακρίβεια των προβλέψεων. Στα στάδια αυτά περιλαμβάνονται:

- **Προεπεξεργασία και Καθαρισμός Δεδομένων:** Η αρχική φάση της προεπεξεργασίας περιλαμβάνει τον καθαρισμό των ιστορικών δεδομένων, με την απομάκρυνση τυχόν ελλειπών ή εσφαλμένων καταχωρήσεων, καθώς και την κανονικοποίηση των δεδομένων για τη βελτίωση της αποτελεσματικότητας των αλγορίθμων μηχανικής μάθησης.
- **Επιλογή Τεχνικών Δεικτών και Χαρακτηριστικών:** Η επιλογή των κατάλληλων τεχνικών δεικτών είναι κρίσιμη για την ακρίβεια των προβλέψεων. Οι τεχνικοί δείκτες που επιλέγονται βασίζονται στην ικανότητά τους να αναγνωρίζουν τάσεις και μοτίβα στις χρονοσειρές, όπως οι ΚΜΟ και ο RSI.
- **Επιλογή και Εκπαίδευση Αλγορίθμων Πρόβλεψης:** Βασικό στάδιο στην ανάπτυξη του συστήματος είναι η επιλογή αλγορίθμων μηχανικής μάθησης. Οι αλγόριθμοι που εφαρμόζονται, όπως το Random Forest και το SVR, αξιολογούνται για την ακρίβειά τους σε προβλέψεις χρηματιστηριακών τιμών και την ικανότητά τους να αναγνωρίζουν σύνθετες σχέσεις στα δεδομένα.
- **Αξιολόγηση και Βελτιστοποίηση Συστήματος:** Η απόδοση του συστήματος πρόβλεψης αξιολογείται με τη χρήση κατάλληλων μετρικών, όπως το MAE και το MSE, ώστε να εντοπίζονται και να διορθώνονται τυχόν αποκλίσεις στην ακρίβεια των προβλέψεων.

Η συγκεκριμένη δομή στοχεύει στη δημιουργία ενός συστήματος που θα προσφέρει υψηλή απόδοση και ακρίβεια, βελτιώνοντας έτσι την ποιότητα των επενδυτικών αποφάσεων και ελαχιστοποιώντας τους πιθανούς κινδύνους στην αγορά.

3.2 Προεπεξεργασία Δεδομένων και Επιλογή Χαρακτηριστικών

Η προεπεξεργασία δεδομένων είναι μια αναγκαία φάση στη δημιουργία μοντέλων μηχανικής μάθησης, ιδιαίτερα όταν πρόκειται για χρηματοοικονομικά δεδομένα που συχνά είναι ατελή ή περιέχουν θόρυβο. Η επιτυχία του συστήματος πρόβλεψης εξαρτάται σε μεγάλο βαθμό από την ποιότητα και την κατάλληλη μορφοποίηση των δεδομένων που τροφοδοτούνται στα μοντέλα. Η διαδικασία αυτή περιλαμβάνει μια σειρά από βήματα, όπως τον καθαρισμό των δεδομένων, την κανονικοποίηση και την επιλογή κρίσιμων χαρακτηριστικών που θα επιτρέψουν στους αλγορίθμους να αποδώσουν καλύτερα. Οι κύριοι στόχοι της προεπεξεργασίας δεδομένων είναι:

- **Αφαίρεση Σφαλμάτων και Θορύβου:** Η απομάκρυνση ακατάλληλων τιμών και οι χειρισμοί στις περιπτώσεις ελλειπών δεδομένων μειώνουν την πιθανότητα σφαλμάτων στα μοντέλα. Η διαδικασία αυτή μπορεί να περιλαμβάνει μεθόδους όπως η χρήση μέσου όρου ή η παρεμβολή για την αντικατάσταση κενών τιμών.
- **Κανονικοποίηση Δεδομένων:** Με την κανονικοποίηση των δεδομένων μέσω της τεχνικής Min-Max Scaling, όλες οι τιμές προσαρμόζονται σε ένα εύρος τιμών, συνήθως από 0 έως 1. Έτσι, εξασφαλίζεται η σύγκλιση των μοντέλων και η βελτίωση της απόδοσής τους.

$$x_{scaled} = \frac{(x - x_{min})}{(x_{max} - x_{min})}$$

Στην επεξεργασία και την προεπεξεργασία των δεδομένων περιλαμβάνονται επίσης τα βήματα της ανάλυσης χρονοσειρών και της κανονικοποίησης, όπου μελετώνται και ενσωματώνονται τα τεχνικά χαρακτηριστικά που μπορούν να ενισχύσουν την πρόβλεψη των μετοχών [18].

3.2.1 Επεξεργασία Χρονοσειρών και Κανονικοποίηση

Οι χρονοσειρές χρηματιστηριακών τιμών περιλαμβάνουν παρατηρήσεις που ακολουθούν μια χρονική σειρά, όπου κάθε παρατήρηση αντιπροσωπεύει την τιμή μιας μετοχής σε μια συγκεκριμένη χρονική στιγμή. Λόγω της χρονικής αυτής σειράς, η ανάλυση και η επεξεργασία χρονοσειρών έχει ιδιαίτερη σημασία στην ανάπτυξη προβλεπτικών μοντέλων. Η κατανόηση των χρονικών μοτίβων, της εποχικότητας και των απρόβλεπτων διακυμάνσεων είναι ζωτικής σημασίας για την ακρίβεια του συστήματος πρόβλεψης.

- **Ανάλυση Εποχικότητας:** Οι εποχικές τάσεις αναφέρονται σε μοτίβα που επαναλαμβάνονται σε συγκεκριμένες χρονικές περιόδους, όπως μηνιαία, εβδομαδιαία ή ημερήσια επαναλαμβανόμενα μοτίβα. Η ανάλυση εποχικότητας βοηθά στον εντοπισμό αυτών των επαναλαμβανόμενων μοτίβων, προσφέροντας προβλέψεις που λαμβάνουν υπόψη τις χαρακτηριστικές διακυμάνσεις της αγοράς κατά τις συγκεκριμένες χρονικές περιόδους. Ένα παράδειγμα είναι η εποχικότητα που παρατηρείται κοντά σε οικονομικές ανακοινώσεις ή σε περιόδους όπως οι εορτές, όπου η ζήτηση μπορεί να αυξάνεται ή να μειώνεται αισθητά.
- **Κανονικοποίηση Δεδομένων:** Στην κανονικοποίηση χρησιμοποιείται η τεχνική Min-Max Scaling, η οποία μετατρέπει τα δεδομένα σε ενιαίο εύρος τιμών. Η κανονικοποίηση αποτρέπει την απόκλιση των τιμών που θα μπορούσε να επηρεάσει την απόδοση των μοντέλων. Η εφαρμογή της κανονικοποίησης επιτρέπει την ενίσχυση της σταθερότητας και την αύξηση της ευαισθησίας των αλγορίθμων μηχανικής μάθησης, ιδίως όταν τα μοντέλα πρέπει να ανιχνεύσουν μικρές διακυμάνσεις στις χρονοσειρές.

Η επεξεργασία χρονοσειρών και η κανονικοποίηση παρέχουν τα θεμέλια για την αποδοτική ανάλυση των χρηματιστηριακών δεδομένων, ενισχύοντας την αξιοπιστία του μοντέλου και ελαχιστοποιώντας τις πιθανότητες σφαλμάτων.

3.2.2 Επιλογή Τεχνικών Δεικτών

Οι τεχνικοί δείκτες είναι εργαλεία που χρησιμοποιούνται για την ανάλυση τάσεων και την εκτίμηση της ζήτησης και της προσφοράς. Καθένας από τους δείκτες αυτούς παρέχει διαφορετικές πληροφορίες και συμβάλλει στην κατανόηση της δυναμικής της αγοράς. Η

χρήση και η κατάλληλη επιλογή των τεχνικών δεικτών μπορεί να βελτιώσει την ανάλυση, αυξάνοντας την ακρίβεια των προβλέψεων.

- **KMO :** Οι KMO βοηθούν στην εξομάλυνση των τιμών και την αναγνώριση της κατεύθυνσης της τάσης. Ο SMA υπολογίζει τον μέσο όρο μιας σειράς τιμών, δίνοντας ισότιμο βάρος σε κάθε παρατήρηση. Αντίθετα, ο EMA δίνει περισσότερο βάρος στις πιο πρόσφατες τιμές, γεγονός που τον καθιστά πιο ευαίσθητο σε αλλαγές της αγοράς. Οι KMO χρησιμοποιούνται συνήθως για την αναγνώριση μακροχρόνιων και βραχυπρόθεσμων τάσεων και την παροχή σημάτων για αγορά ή πώληση.
- **RSI:** Ο RSI είναι ένας ταλαντωτής ο οποίος μετρά την ορμή και τη δύναμη των πρόσφατων τιμών, επιτρέποντας στους επενδυτές να αναγνωρίσουν αν μια μετοχή είναι υπερτιμημένη ή υποτιμημένη. Επίπεδα RSI πάνω από το 70 θεωρούνται υπερτιμημένα, ενώ επίπεδα κάτω από το 30 υποτιμημένα. Ο RSI παρέχει πληροφορίες για την πιθανή αντιστροφή των τιμών, προσφέροντας σαφή ένδειξη για τη δυναμική της μετοχής.

Η προσεκτική επιλογή και η χρήση αυτών των τεχνικών δεικτών ενισχύουν την αποδοτικότητα των προβλεπτικών μοντέλων, επιτρέποντας την εκτίμηση της κατεύθυνσης των τιμών και την εξαγωγή ουσιαστικών συμπερασμάτων για τη δυναμική της αγοράς.

3.3 Περιγραφή των Χρησιμοποιούμενων Αλγορίθμων

Σε αυτή την ενότητα, παρουσιάζονται οι βασικοί αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν για την πρόβλεψη τιμών μετοχών. Οι επιλεγμένοι αλγόριθμοι περιλαμβάνουν τον Random Forest, τον SVR και τη Linear Regression. Κάθε αλγόριθμος έχει συγκεκριμένα χαρακτηριστικά και προσεγγίζει τη διαδικασία πρόβλεψης με διαφορετικό τρόπο, γεγονός που επιτρέπει τη σύγκρισή τους ως προς την απόδοσή τους σε διαφορετικές συνθήκες και είδη δεδομένων [19].

3.3.1 Random Forest

Ο Random Forest είναι ένας αλγόριθμος που ανήκει στην κατηγορία των ensemble learning μεθόδων και χρησιμοποιείται ευρέως για προβλεπτικές εργασίες λόγω της υψηλής ακρίβειας

και της ανθεκτικότητάς του. Αποτελείται από έναν μεγάλο αριθμό δέντρων απόφασης, τα οποία δημιουργούνται με τυχαία επιλογή υποδειγμάτων από το σύνολο των δεδομένων εκπαίδευσης. Ο συνδυασμός των προβλέψεων από κάθε δέντρο παρέχει μια τελική πρόβλεψη, εξασφαλίζοντας σταθερότητα και μειώνοντας την πιθανότητα υπερεκπαίδευσης.

- **Δομή του Random Forest:** Ο Random Forest βασίζεται στην τεχνική "bagging" (Bootstrap Aggregation), η οποία αφορά τη δειγματοληψία με αντικατάσταση. Αυτό σημαίνει ότι κάθε δέντρο απόφασης εκπαιδεύεται σε ένα διαφορετικό υποσύνολο δεδομένων, το οποίο περιέχει τυχαία επιλεγμένες παρατηρήσεις. Επιπλέον, η επιλογή χαρακτηριστικών που χρησιμοποιούνται σε κάθε διακλάδωση ενός δέντρου είναι επίσης τυχαία, με στόχο να μειωθεί η συσχέτιση μεταξύ των δέντρων και να ενισχυθεί η ανθεκτικότητα του μοντέλου. Αυτή η διαδικασία δημιουργεί ένα πιο ισχυρό μοντέλο, το οποίο μπορεί να γενικεύσει καλύτερα σε νέα δεδομένα.
- **Αντιμετώπιση Θορύβου και Πολυπλοκότητας Δεδομένων:** Ο Random Forest είναι ικανός να αντιμετωπίζει θόρυβο και πολυπλοκότητα στα δεδομένα χάρη στην ενσωματωμένη δυνατότητα "υπερψηφίας" (majority voting), η οποία χρησιμοποιείται για να καθοριστεί η τελική πρόβλεψη. Ειδικά σε περιπτώσεις όπου τα δεδομένα περιέχουν τυχαίες αποκλίσεις ή θόρυβο, ο Random Forest παρέχει μεγαλύτερη ακρίβεια σε σχέση με μεμονωμένα δέντρα απόφασης, καθώς συνδυάζει τις προβλέψεις από πολλά ανεξάρτητα δέντρα, δημιουργώντας ένα πιο ισχυρό και ακριβές σύστημα πρόβλεψης.
- **Πλεονεκτήματα του Random Forest:**
 - **Ανθεκτικότητα στο Overfitting:** Καθώς κάθε δέντρο στο Random Forest εκπαιδεύεται με διαφορετικό υποσύνολο δεδομένων και χαρακτηριστικών, μειώνεται η πιθανότητα υπερεκπαίδευσης. Η συλλογική πρόβλεψη των δέντρων, επομένως, προσφέρει μεγαλύτερη ακρίβεια.
 - **Υψηλή Απόδοση σε Μεγάλα Δεδομένα:** Ο Random Forest είναι κατάλληλος για μεγάλα σύνολα δεδομένων, καθώς μπορεί να ανιχνεύσει πολυπλοκότερες σχέσεις και να ανταποκριθεί καλύτερα σε δεδομένα με μεγάλο αριθμό χαρακτηριστικών.
 - **Διαχείριση Ελλιπών Δεδομένων:** Σε αντίθεση με άλλους αλγορίθμους, ο Random Forest είναι ικανός να διαχειρίζεται ελλιπή δεδομένα με μεγαλύτερη

αποτελεσματικότητα, παρέχοντας αξιόπιστες προβλέψεις ακόμα και σε υποσύνολα δεδομένων με απώλειες.

- **Προσαρμογή στις Χρηματοοικονομικές Προβλέψεις:** Στην πρόβλεψη τιμών μετοχών, ο Random Forest έχει τη δυνατότητα να ανιχνεύει πολύπλοκες και μη γραμμικές σχέσεις στα δεδομένα, γεγονός που τον καθιστά ιδανικό για περιβάλλοντα με υψηλή μεταβλητότητα. Η προσέγγιση αυτή επιτρέπει την πρόβλεψη χρηματιστηριακών τιμών με μεγαλύτερη ακρίβεια και μειώνει την πιθανότητα σφάλματος σε ασταθείς αγορές, όπως αυτές που συναντώνται συχνά στα χρηματιστήρια.
- **Αξιολόγηση και Τεχνικές Βελτίωσης:**
 - **Ρύθμιση Υπερπαραμέτρων:** Για να επιτευχθεί η βέλτιστη απόδοση του Random Forest, πρέπει να ρυθμιστούν κατάλληλα οι υπερπαραμέτροι, όπως ο αριθμός των δέντρων και το μέγιστο βάθος κάθε δέντρου. Ο σωστός καθορισμός αυτών των παραμέτρων μπορεί να μειώσει τον χρόνο εκπαίδευσης και να αυξήσει την ακρίβεια του μοντέλου.
 - **Χρήση Out-of-Bag Error (OOB Error):** Ο Random Forest παρέχει έναν ενσωματωμένο μηχανισμό αξιολόγησης, γνωστό ως out-of-bag error, που μετρά την ακρίβεια του μοντέλου χωρίς τη χρήση ενός ξεχωριστού συνόλου δοκιμών. Αυτή η τεχνική μπορεί να βοηθήσει στην αποτίμηση του μοντέλου κατά τη διάρκεια της εκπαίδευσης και να παρέχει εκτιμήσεις για την απόδοσή του.

Ο Random Forest είναι, επομένως, ένας ισχυρός αλγόριθμος που συνδυάζει την απλότητα των δέντρων απόφασης με τη δύναμη των ensemble μοντέλων, προσφέροντας αξιόπιστες προβλέψεις σε σύνθετα χρηματοοικονομικά δεδομένα [20].

3.3.2 Support Vector Regressor

Ο SVR αποτελεί μια επέκταση του αλγορίθμου Υποστηρικτικών Διανυσματικών Μηχανών (Support Vector Machines - SVM), ο οποίος έχει σχεδιαστεί για προβλέψεις σε συνεχείς τιμές και χρησιμοποιείται εκτενώς για την πρόβλεψη σε δεδομένα που περιέχουν πολύπλοκες σχέσεις. Σε αντίθεση με άλλες μεθόδους που επικεντρώνονται στη μείωση του μέσου

τετραγωνικού σφάλματος, ο SVR προσπαθεί να ελαχιστοποιήσει τα σφάλματα πρόβλεψης εντός ενός συγκεκριμένου εύρους [21].

- **Βασική Λειτουργία και Αρχές του SVR:**

- Ο SVR επιδιώκει να δημιουργήσει έναν υπερεπίπεδο (hyperplane) που βρίσκεται εντός μιας καθορισμένης ζώνης περιθωρίου, γνωστής ως "ελαστικό περιθώριο" (επίσης γνωστό ως "επίπεδο σφάλματος ε"). Ο στόχος είναι η κατασκευή ενός υπερεπιπέδου που βρίσκεται εντός αυτής της ζώνης, χωρίς όμως να αποκλίνει σημαντικά από τις παρατηρούμενες τιμές.
- Αυτό σημαίνει ότι οι αποκλίσεις από την πραγματική τιμή επιτρέπονται μόνο όταν βρίσκονται εντός του προκαθορισμένου εύρους, κάτι που κάνει τον SVR ιδιαίτερα χρήσιμο για δεδομένα όπου οι μικρές αποκλίσεις δεν είναι κρίσιμες. Στην πραγματικότητα, η μέθοδος αυτή μπορεί να απορρίπτει μικρές διακυμάνσεις στα δεδομένα ως θόρυβο, επιτυγχάνοντας μεγαλύτερη ακρίβεια.

- **Χρήση Πυρήνων (Kernels):**

- Ένα από τα ισχυρά χαρακτηριστικά του SVR είναι η ικανότητά του να χειρίζεται μη γραμμικές σχέσεις στα δεδομένα, μέσω της χρήσης συναρτήσεων πυρήνα (kernels). Οι πιο διαδεδομένες συναρτήσεις πυρήνα περιλαμβάνουν:
 - Γραμμικός πυρήνας: Κατάλληλος για δεδομένα με γραμμική σχέση μεταξύ των χαρακτηριστικών.
 - Πολυωνυμικός πυρήνας: Εφαρμόζεται όταν υπάρχουν πιο πολύπλοκες και υψηλότερης τάξης σχέσεις στα δεδομένα.
 - Radial Basis Function (RBF): Είναι από τους πιο δημοφιλείς πυρήνες, καθώς προσφέρει ευελιξία και είναι κατάλληλος για την ανίχνευση κρυφών και πολύπλοκων μοτίβων. Ο RBF είναι ιδιαίτερα χρήσιμος για χρηματοοικονομικά δεδομένα, όπου οι σχέσεις μπορεί να μην είναι εύκολα κατανοητές ή γραμμικές.
- Αυτές οι συναρτήσεις πυρήνα επιτρέπουν στον SVR να προσαρμόζεται σε πολύπλοκες μορφές δεδομένων, δημιουργώντας καμπυλόγραμμους υπερεπιπέδες που ταιριάζουν καλύτερα στις σχέσεις των χαρακτηριστικών.

- **Προσαρμοστικότητα του SVR σε Δεδομένα Μεταβλητότητας:**

- Παρά την ευελιξία του, ο SVR ενδέχεται να δυσκολεύεται σε δεδομένα με μεγάλες διακυμάνσεις και αστάθεια, όπως αυτά των χρηματιστηριακών αγορών, όπου οι προβλέψεις πρέπει να λαμβάνουν υπόψη γρήγορες και απρόβλεπτες αλλαγές.
- Παρόλο που οι πυρήνες μπορούν να βοηθήσουν στην ανίχνευση μη γραμμικών σχέσεων, η σταθερότητα του SVR δεν είναι πάντα ιδανική για ακραίες τιμές. Για παράδειγμα, σε περιόδους μεγάλης αστάθειας ή απρόβλεπτων γεγονότων, οι προβλέψεις του SVR μπορεί να παρουσιάζουν μεγαλύτερα σφάλματα σε σχέση με άλλους αλγορίθμους.
- **Πλεονεκτήματα και Περιορισμοί του SVR:**
 - **Πλεονεκτήματα:**
 - Κατάλληλος για προβλέψεις σε δεδομένα με ελαφρώς μη γραμμικές σχέσεις.
 - Η προσαρμογή της ζώνης περιθωρίου του επιτρέπει να εξισορροπεί την ακρίβεια με την ευελιξία, αγνοώντας μικρές διακυμάνσεις στα δεδομένα.
 - Προσφέρει δυνατότητα ρύθμισης μέσω της επιλογής του πυρήνα και της παραμέτρου "C" που καθορίζει το κόστος των αποκλίσεων.
 - **Περιορισμοί:**
 - Δεν είναι πάντα αποτελεσματικός σε χρονοσειρές με υψηλή μεταβλητότητα ή δεδομένα που απαιτούν ανάλυση ακραίων τιμών.
 - Η ρύθμιση των παραμέτρων, όπως η παράμετρος "C" και οι πυρήνες, μπορεί να είναι σύνθετη διαδικασία και απαιτεί δοκιμές για να επιτευχθεί η βέλτιστη απόδοση.
- **Εφαρμογή του SVR στη Χρηματοοικονομική Ανάλυση:**
 - Στο πλαίσιο της πρόβλεψης τιμών μετοχών, ο SVR είναι ιδανικός για περιπτώσεις όπου τα δεδομένα παρουσιάζουν ομαλή διακύμανση και οι αποκλίσεις δεν είναι έντονες. Χρησιμοποιείται συχνά για την πρόβλεψη της γενικής τάσης, όταν τα δεδομένα ακολουθούν προβλέψιμα μοτίβα.
 - Σε περιπτώσεις όπου οι τιμές μετοχών παρουσιάζουν ακραίες τιμές ή απότομες αλλαγές, ο SVR μπορεί να μην προσφέρει την ίδια ακρίβεια, καθιστώντας τον λιγότερο κατάλληλο για αγορές με έντονη μεταβλητότητα. Ο SVR είναι

επομένως ένας ισχυρός αλγόριθμος για δεδομένα που απαιτούν ευελιξία στην ανάλυση, ωστόσο δεν είναι πάντα ιδανικός για ακραίες διακυμάνσεις όπως αυτές των χρηματοοικονομικών δεδομένων.

3.3.3 Linear Regression

Η Linear Regression είναι μία από τις πιο βασικές μεθόδους στην ανάλυση δεδομένων και τη μηχανική μάθηση, που χρησιμοποιείται για την πρόβλεψη της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Στην πρόβλεψη τιμών μετοχών, η Linear Regression χρησιμοποιείται για την ανάλυση και πρόβλεψη μοτίβων όπου οι σχέσεις μεταξύ των μεταβλητών θεωρούνται γραμμικές.

- **Βασική Θεωρία και Στόχος της Linear Regression:**

- Στην απλή Linear Regression, ο αλγόριθμος επιχειρεί να βρει τη γραμμή που προσεγγίζει καλύτερα τη σχέση μεταξύ της ανεξάρτητης μεταβλητής (π.χ., τεχνικός δείκτης) και της εξαρτημένης μεταβλητής (τιμή μετοχής). Η σχέση αυτή εκφράζεται μέσω της εξίσωσης της γραμμής $y = m \cdot x + b$ όπου:
 - y είναι η εξαρτημένη μεταβλητή (η προβλεπόμενη τιμή),
 - x είναι η ανεξάρτητη μεταβλητή (το χαρακτηριστικό),
 - m είναι η κλίση της γραμμής, και
 - b είναι η σταθερά.
- Ο στόχος του μοντέλου είναι να υπολογίσει τις βέλτιστες τιμές για τις παραμέτρους m και b , ώστε η γραμμή να μειώνει το μέσο τετραγωνικό σφάλμα (MSE) μεταξύ των προβλεπόμενων τιμών και των πραγματικών τιμών.

- **Ανάλυση Πολυμεταβλητής Linear Regression:**

- Στην περίπτωση πολυμεταβλητής γραμμικής παλινδρόμησης, χρησιμοποιούνται περισσότερες από μία ανεξάρτητες μεταβλητές για την πρόβλεψη της εξαρτημένης μεταβλητής. Η εξίσωση γίνεται:

$$y = m_1x_1 + m_2x_2 + \dots + m_nx_n + b$$

όπου κάθε m_i είναι ο συντελεστής που αντιστοιχεί στην ανεξάρτητη μεταβλητή x_i .

- Η πολυμεταβλητή Linear Regression είναι ιδανική για σύνθετα δεδομένα, καθώς επιτρέπει την ενσωμάτωση πολλών χαρακτηριστικών, όπως οι ΚΜΟ και ο δείκτης σχετικής ισχύος, για τη βελτίωση της ακρίβειας των προβλέψεων.
- **Πλεονεκτήματα της Linear Regression:**
 - Απλότητα και Ερμηνευσιμότητα: Η Linear Regression είναι ένας από τους πιο εύκολα κατανοητούς αλγορίθμους, γεγονός που επιτρέπει την εύκολη ερμηνεία των αποτελεσμάτων και των παραμέτρων του μοντέλου.
 - Χαμηλές Απαιτήσεις Υπολογιστικής Ισχύος: Είναι πολύ αποδοτική από άποψης πόρων και μπορεί να εφαρμοστεί σε μεγάλα σύνολα δεδομένων χωρίς μεγάλες απαιτήσεις σε υπολογιστική ισχύ.
 - Αντοχή σε Υπερεκπαίδευση (Overfitting): Σε αντίθεση με πιο περίπλοκους αλγόριθμους, η Linear Regression έχει μικρότερο κίνδυνο υπερεκπαίδευσης, εφόσον χρησιμοποιείται σε δεδομένα με γραμμικές σχέσεις.
- **Περιορισμοί και Μειονεκτήματα:**
 - Περιορισμένη Προσαρμοστικότητα σε Μη Γραμμικά Δεδομένα: Ο αλγόριθμος δεν μπορεί να χειριστεί μη γραμμικές σχέσεις, γεγονός που τον καθιστά λιγότερο αποτελεσματικό για δεδομένα που παρουσιάζουν περίπλοκες ή μη γραμμικές σχέσεις, όπως είναι οι χρηματοοικονομικές χρονοσειρές.
 - Ευαισθησία σε Ακραίες Τιμές (Outliers): Η Linear Regression μπορεί να επηρεαστεί έντονα από ακραίες τιμές, κάτι που ενδέχεται να οδηγήσει σε ανακρίβειες στις προβλέψεις. Οι ακραίες τιμές μπορούν να αλλοιώσουν την κλίση της γραμμής και να μειώσουν την ακρίβεια των αποτελεσμάτων.
- **Εφαρμογή στη Χρηματοοικονομική Ανάλυση:**
 - Στην πρόβλεψη τιμών μετοχών, η Linear Regression χρησιμοποιείται κυρίως για να προσδιορίσει γενικές τάσεις και να προσφέρει απλές προβλέψεις. Είναι ιδιαίτερα χρήσιμη όταν υπάρχει σαφής και σταθερή τάση στις τιμές ή όταν η σχέση μεταξύ των χαρακτηριστικών και της εξαρτημένης μεταβλητής είναι σχετικά απλή.
 - Σε αγορές με υψηλή μεταβλητότητα ή με πολλές μη γραμμικές σχέσεις, η Linear Regression ενδέχεται να είναι λιγότερο αποτελεσματική, καθώς δεν

μπορεί να ανταποκριθεί στις σύνθετες σχέσεις και στις γρήγορες διακυμάνσεις που χαρακτηρίζουν τέτοιες αγορές.

- **Βελτιστοποίηση και Τεχνικές Αντιμετώπισης Περιορισμών:**
 - Για τη βελτίωση της ακρίβειας του μοντέλου, μπορεί να χρησιμοποιηθεί τεχνική ρύθμισης παραμέτρων (regularization) όπως η Ridge και η Lasso Regression, που μειώνουν τον αντίκτυπο των μη ουσιαστικών χαρακτηριστικών, επιτρέποντας μια πιο σταθερή πρόβλεψη.
 - Ανίχνευση και Αφαίρεση Ακραίων Τιμών: Η χρήση μεθόδων ανίχνευσης ακραίων τιμών πριν από την εκπαίδευση του μοντέλου μπορεί να βελτιώσει την ακρίβεια, μειώνοντας την πιθανότητα οι ακραίες τιμές να επηρεάσουν το αποτέλεσμα.

Η Linear Regression παραμένει ένα από τα πιο βασικά εργαλεία πρόβλεψης και ανάλυσης, ιδανική για απλές και γραμμικές σχέσεις. Ωστόσο, στις χρηματιστηριακές αγορές, η εφαρμογή της απαιτεί προσεκτική επιλογή των χαρακτηριστικών και ενίσχυση της ανθεκτικότητάς της μέσω τεχνικών ρύθμισης παραμέτρων και αντιμετώπισης των ακραίων τιμών [22].

3.4 Αρχιτεκτονική του Συστήματος και Εργαλεία

Η αρχιτεκτονική του συστήματος πρόβλεψης τιμών μετοχών που αναπτύχθηκε στηρίζεται σε μία ευέλικτη και επεκτάσιμη δομή που επιτρέπει την εύκολη ενσωμάτωση και επεξεργασία των δεδομένων. Αυτή η αρχιτεκτονική υποστηρίζει όλα τα απαραίτητα στάδια για την πρόβλεψη τιμών, από την εισαγωγή των δεδομένων και την προεπεξεργασία τους μέχρι την ανάλυση των αποτελεσμάτων και την παρουσίαση των προβλέψεων.

3.4.1 Δομή Αρχείων και Ροή Εργασίας

Η δομή των αρχείων του συστήματος έχει σχεδιαστεί για να επιτρέπει τη σαφή οργάνωση των δεδομένων, των μοντέλων, των κώδικων προεπεξεργασίας και των αποτελεσμάτων. Κάθε φάκελος και αρχείο έχει συγκεκριμένο σκοπό, ώστε να διασφαλίζεται η αποδοτική λειτουργία και η εύκολη τροποποίηση.

- **Φάκελος Δεδομένων (Data Folder):** Περιέχει τα ιστορικά χρηματιστηριακά δεδομένα, καθώς και οποιαδήποτε επιπλέον πληροφορία απαιτείται για τις προβλέψεις. Τα δεδομένα οργανώνονται σε μορφές όπως CSV ή Excel για ευκολία στη διαχείριση και ανάλυση.
- **Φάκελος Προεπεξεργασίας Δεδομένων:** Αυτός ο φάκελος περιλαμβάνει τα σκριπτάκια που υλοποιούν τον καθαρισμό των δεδομένων, την κανονικοποίηση και την εφαρμογή των τεχνικών δεικτών. Αυτή η ενότητα περιέχει επίσης τις παραμέτρους και τις ρυθμίσεις για τη διαμόρφωση των δεδομένων πριν την εκπαίδευση των μοντέλων.
- **Φάκελος Μοντέλων (Models Folder):** Περιλαμβάνει τα αρχεία που αφορούν την εκπαίδευση και την αποθήκευση των μοντέλων μηχανικής μάθησης, όπως το Random Forest, το SVR και τη Linear Regression. Κάθε μοντέλο διαμορφώνεται και ρυθμίζεται ξεχωριστά, με σκοπό την επίτευξη βέλτιστης απόδοσης για την πρόβλεψη τιμών μετοχών.
- **Φάκελος Αποτελεσμάτων (Results Folder):** Αποθηκεύει τα αποτελέσματα της εκπαίδευσης και των προβλέψεων για περαιτέρω ανάλυση. Εδώ αποθηκεύονται επίσης γραφήματα και πίνακες που χρησιμοποιούνται στην τελική παρουσίαση των αποτελεσμάτων και στη σύγκριση της απόδοσης των μοντέλων.

Η ροή εργασίας βασίζεται σε ένα σύστημα modular, που σημαίνει ότι κάθε στάδιο της διαδικασίας είναι ανεξάρτητο και μπορεί να τροποποιηθεί χωρίς να επηρεάζεται η συνολική λειτουργία του συστήματος. Αυτή η προσέγγιση διευκολύνει την ενσωμάτωση νέων μοντέλων, την αλλαγή των δεδομένων και τη βελτίωση των αποτελεσμάτων.

3.4.2 Χρήση Πλατφορμών και Εργαλείων

Το σύστημα αναπτύχθηκε σε περιβάλλον Python, χρησιμοποιώντας τις ισχυρές βιβλιοθήκες και εργαλεία της γλώσσας αυτής για την ανάλυση και πρόβλεψη δεδομένων χρονοσειρών. Η Python προσφέρει μια ολοκληρωμένη πλατφόρμα για την επεξεργασία δεδομένων και τη δημιουργία μοντέλων μηχανικής μάθησης, και υποστηρίζει σημαντικές βιβλιοθήκες που απλοποιούν την υλοποίηση πολύπλοκων αλγορίθμων.

- **Pandas:** Η βιβλιοθήκη Pandas χρησιμοποιείται για τη διαχείριση και επεξεργασία δεδομένων. Επιτρέπει την εισαγωγή και ανάλυση των αρχείων .csv, διευκολύνοντας

τον καθαρισμό των δεδομένων, την κανονικοποίηση και την εξαγωγή τεχνικών δεικτών. Μέσω της Pandas, οι χρονοσειρές οργανώνονται με τέτοιον τρόπο ώστε να είναι άμεσα προσβάσιμες από τους αλγόριθμους και να μειώνεται ο χρόνος επεξεργασίας [23] .

- **Scikit-Learn:** Η Scikit-Learn είναι μια από τις βασικές βιβλιοθήκες μηχανικής μάθησης, παρέχοντας υλοποιήσεις των αλγορίθμων που χρησιμοποιούνται στον κώδικα, όπως το Random Forest, το SVR και η Linear Regression. Περιλαμβάνει επίσης εργαλεία για την κανονικοποίηση δεδομένων (π.χ., MinMaxScaler), την αξιολόγηση της απόδοσης των μοντέλων (με τις μετρικές MAE και MSE) και τη διαίρεση των δεδομένων σε σύνολα εκπαίδευσης και ελέγχου. Η Scikit-Learn απλοποιεί τη διαδικασία εκπαίδευσης και αξιολόγησης, προσφέροντας ένα εύχρηστο περιβάλλον για την υλοποίηση αλγορίθμων πρόβλεψης [24].
- **Matplotlib:** Χρησιμοποιείται για την οπτικοποίηση των αποτελεσμάτων, επιτρέποντας την παρουσίαση των προβλέψεων σε συνδυασμό με τις πραγματικές τιμές. Τα διαγράμματα που δημιουργούνται με τη Matplotlib παρέχουν μια σαφή εικόνα των προβλέψεων σε σχέση με τις πραγματικές τιμές, διευκολύνοντας την κατανόηση της απόδοσης των μοντέλων [25].
- **Tkinter:** Η βιβλιοθήκη Tkinter χρησιμοποιείται για τη δημιουργία ενός γραφικού περιβάλλοντος χρήστη (GUI), όπου οι χρήστες μπορούν να επιλέξουν συγκεκριμένες μετοχές και να δουν τα αποτελέσματα των προβλέψεων. Το GUI προσφέρει μια απλή και διαδραστική πλατφόρμα που επιτρέπει την πρόσβαση στα αποτελέσματα και τη σύγκριση των προβλέψεων σε πραγματικό χρόνο [26].

Αυτές οι βιβλιοθήκες και τα εργαλεία συνδυάζονται για να δημιουργήσουν ένα ευέλικτο και αποδοτικό σύστημα πρόβλεψης. Η δομή του κώδικα και η χρήση των παραπάνω εργαλείων καθιστούν το σύστημα επεκτάσιμο και εύκολα προσαρμόσιμο σε διαφορετικά χρηματοοικονομικά δεδομένα, ενώ η επεξεργαστική ισχύς και οι δυνατότητες που παρέχουν η Python και οι βιβλιοθήκες της εξασφαλίζουν ότι το σύστημα παραμένει γρήγορο και αποδοτικό.

4

Αξιολόγηση Μοντέλων και Ανάλυση Αποτελεσμάτων

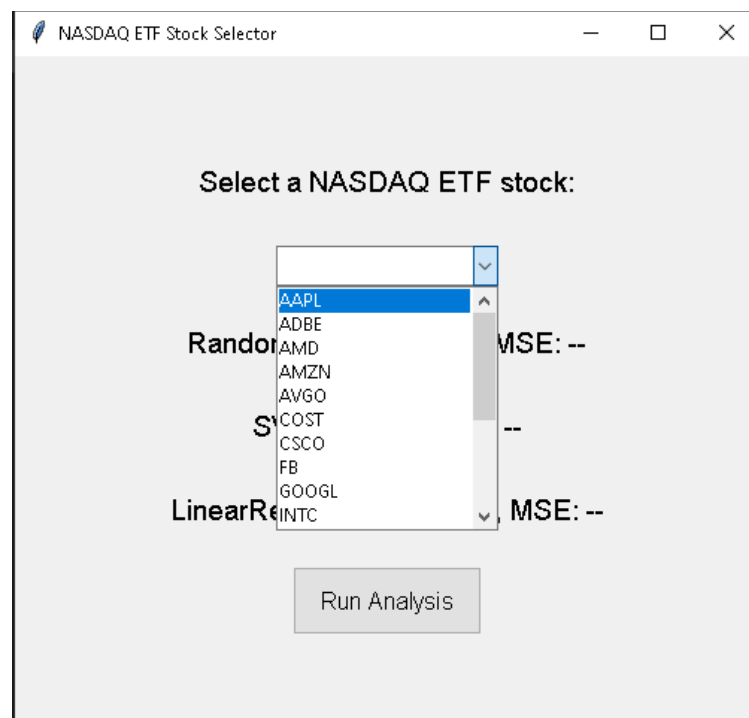
Η αξιολόγηση της απόδοσης των μοντέλων είναι ένα από τα πλέον κρίσιμα στάδια μιας μελέτης πρόβλεψης, καθώς παρέχει τα απαραίτητα στοιχεία για την κατανόηση της αποτελεσματικότητας και της αξιοπιστίας των μεθόδων που εφαρμόστηκαν. Στο παρόν κεφάλαιο, εξετάζονται λεπτομερώς τα αποτελέσματα της εκπαίδευσης και των προβλέψεων για τις τιμές των μετοχών, που προέκυψαν από τρία διαφορετικά μοντέλα: το Random Forest, το SVR και τη Linear Regression. Για να γίνει η σύγκριση μεταξύ των μοντέλων όσο το δυνατόν πιο αντικειμενική, αναπτύχθηκε μια συνεπής μεθοδολογία αξιολόγησης που περιλαμβάνει τη χρήση των μετρικών MAE και MSE. Οι συγκεκριμένες μετρικές έχουν επιλεγεί λόγω της ικανότητάς τους να παρέχουν μια ποσοτική εκτίμηση των αποκλίσεων μεταξύ των πραγματικών και των προβλεπόμενων τιμών, λαμβάνοντας υπόψη τόσο την ακρίβεια όσο και τη σταθερότητα των προβλέψεων. Παράλληλα, αναπτύχθηκε μια απλή και λειτουργική διεπαφή χρήστη (UI) για την επιλογή και την παρουσίαση των αποτελεσμάτων. Το UI επιτρέπει στους χρήστες να επιλέγουν εύκολα μια μετοχή από ένα προκαθορισμένο σύνολο και να παρακολουθούν τα αποτελέσματα που προέκυψαν από την εφαρμογή των μοντέλων. Αυτή η διεπαφή ενσωματώνει τις τιμές των μετρικών MAE και MSE για κάθε μοντέλο, προσφέροντας μια οπτικοποιημένη σύγκριση των αποτελεσμάτων.

Περιγραφή της Διεπαφής Χρήστη (UI)

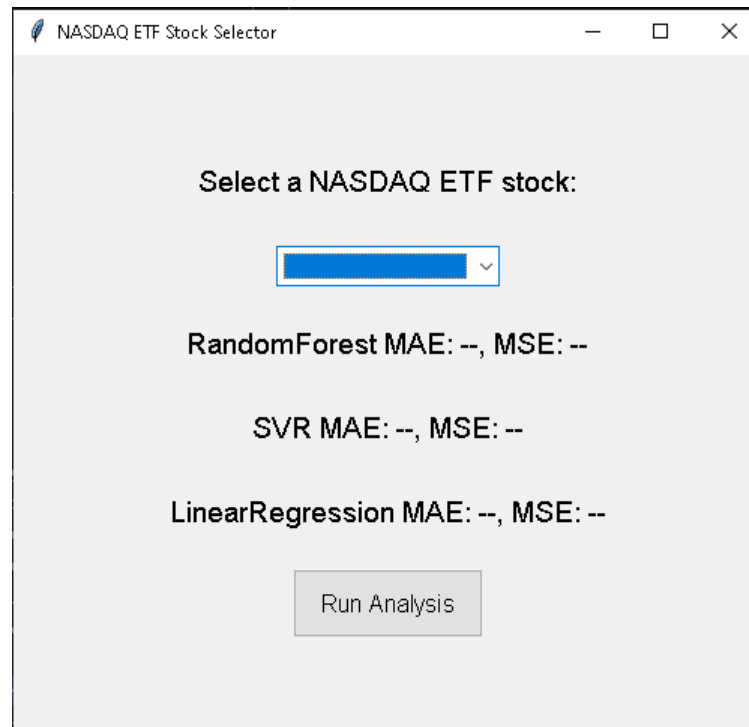
Η διεπαφή χρήστη σχεδιάστηκε με απλότητα και εργονομία, ώστε να διευκολύνει τη διαδικασία επιλογής και ανάλυσης των αποτελεσμάτων. Το UI περιλαμβάνει τα εξής χαρακτηριστικά:

- **Επιλογή Μετοχής:** Στο πάνω μέρος της διεπαφής, υπάρχει ένα dropdown menu που επιτρέπει στους χρήστες να επιλέξουν μια συγκεκριμένη μετοχή από το σύνολο του NASDAQ ETF. Η επιλογή αυτή είναι δυναμική και ανανεώνεται για να περιλαμβάνει όλες τις διαθέσιμες μετοχές του πειράματος.
- **Παρουσίαση Αποτελεσμάτων:** Κάτω από το dropdown menu, εμφανίζονται τα αποτελέσματα της αξιολόγησης για κάθε μοντέλο (Random Forest, SVR, Linear Regression). Οι μετρικές MAE και MSE παρουσιάζονται για το εκάστοτε μοντέλο, επιτρέποντας μια γρήγορη και αποτελεσματική σύγκριση.
- **Κουμπί Εκκίνησης Ανάλυσης:** Στο κάτω μέρος της διεπαφής, βρίσκεται το κουμπί "Run Analysis", το οποίο ενεργοποιεί την ανάλυση της επιλεγμένης μετοχής, εμφανίζοντας τα αντίστοιχα αποτελέσματα.

Η διεπαφή αυτή αναπτύχθηκε ως εργαλείο παρουσίασης των αποτελεσμάτων με εύκολο και οπτικοποιημένο τρόπο. Ο χρήστης μπορεί άμεσα να συγκρίνει τα μοντέλα και να κατανοήσει τις αποδόσεις τους.



Εικόνα 1. UI DropDown Menu



Εικόνα 2. UI Main

4.1 Μεθοδολογία Αξιολόγησης

Η αξιολόγηση των μοντέλων μηχανικής μάθησης αποτελεί καθοριστικό βήμα για την αποτίμηση της αποτελεσματικότητάς τους στις προβλέψεις χρηματοοικονομικών δεδομένων. Η μεθοδολογία που υιοθετήθηκε σε αυτή την εργασία σχεδιάστηκε με σκοπό να παρέχει μια αντικειμενική, αναλυτική και συγκριτική ανάλυση των μοντέλων Random Forest, SVR και Linear Regression. Κάθε βήμα της μεθοδολογίας προσφέρει τη βάση για μια ολοκληρωμένη κατανόηση των δυνατοτήτων και των περιορισμών κάθε μοντέλου.

Στόχοι της Μεθοδολογίας

Η αξιολόγηση επικεντρώνεται στους παρακάτω στόχους:

- **Ακρίβεια προβλέψεων:** Πόσο κοντά βρίσκονται οι προβλέψεις στις πραγματικές τιμές;
- **Σταθερότητα μοντέλου:** Πόσο αξιόπιστο είναι κάθε μοντέλο σε διαφορετικά χρονικά διαστήματα και δεδομένα;
- **Απόδοση σε διαφορετικά δεδομένα:** Πώς ανταποκρίνεται το μοντέλο στις διακυμάνσεις τιμών διαφορετικών μετοχών;

- **Σύγκριση μεταξύ μοντέλων:** Ανάλυση των διαφορών στην απόδοση των μοντέλων, με στόχο την εξαγωγή συμπερασμάτων για το πιο κατάλληλο μοντέλο σε συγκεκριμένα σενάρια.

Διαδικασία Αξιολόγησης

Η διαδικασία που ακολουθήθηκε περιλαμβάνει τα εξής βήματα:

- **Διαχωρισμός Δεδομένων:** Τα δεδομένα κάθε μετοχής χωρίστηκαν σε δύο σύνολα:
 - Εκπαίδευσης (Training): Περιλαμβάνει το 70% των δεδομένων, χρησιμοποιείται για την εκπαίδευση των μοντέλων.
 - Δοκιμών (Testing): Περιλαμβάνει το 30% των δεδομένων, χρησιμοποιείται για την αξιολόγηση της απόδοσης των μοντέλων σε μη γνωστά δεδομένα.
- **Υπολογισμός Μετρικών Απόδοσης:** Για την αντικειμενική αξιολόγηση της απόδοσης, χρησιμοποιήθηκαν οι μετρικές MAE και MSE. Οι μετρικές αυτές επιλέχθηκαν επειδή:
 - Το MAE παρέχει την απόλυτη μέση απόκλιση των προβλέψεων από τις πραγματικές τιμές, δίνοντας έμφαση στα πραγματικά λάθη.
 - Το MSE δίνει μεγαλύτερη βαρύτητα σε μεγάλες αποκλίσεις, επιτρέποντας την ανίχνευση σοβαρών λαθών.
- **Εκπαίδευση και Δοκιμή Μοντέλων:** Κάθε μοντέλο εκπαιδεύτηκε με το εκπαιδευτικό σύνολο και αξιολογήθηκε με το σύνολο δοκιμών. Τα αποτελέσματα καταγράφηκαν και παρουσιάστηκαν για κάθε μετοχή.
- **Οπτικοποίηση Αποτελεσμάτων:** Η οπτικοποίηση περιλαμβάνει:
 - Γραφήματα που δείχνουν τη σύγκριση μεταξύ των προβλεπόμενων και των πραγματικών τιμών.
 - Παρουσίαση των τιμών των μετρικών MAE και MSE για κάθε μοντέλο και μετοχή.
- **Ανάλυση Συγκριτικών Αποτελεσμάτων:** Τα αποτελέσματα όλων των μετοχών συγκρίθηκαν μεταξύ τους, αναδεικνύοντας τα ισχυρά και αδύναμα σημεία κάθε μοντέλου.

4.1.1 Κριτήρια Αξιολόγησης

Η αξιολόγηση των αλγορίθμων πραγματοποιείται με τη χρήση δύο βασικών μετρικών: το MAE και το MSE, οι οποίες προσδιορίζουν την ακρίβεια των προβλέψεων με βάση τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών τιμών [27].

- **MAE:**

- Η MAE είναι η μέση απόλυτη τιμή των διαφορών μεταξύ των προβλεπόμενων και των πραγματικών τιμών. Υπολογίζεται ως εξής: $MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$ όπου y_i είναι η πραγματική τιμή, \hat{y}_i είναι η προβλεπόμενη τιμή και N ο αριθμός των προβλέψεων.
- Η MAE εκφράζει τη μέση απόκλιση των προβλέψεων από τις πραγματικές τιμές, ανεξαρτήτως κατεύθυνσης (θετικές ή αρνητικές αποκλίσεις). Μια μικρή τιμή MAE υποδηλώνει ότι οι προβλέψεις είναι κοντά στις πραγματικές τιμές και, άρα, το μοντέλο είναι ακριβές.
- Εφαρμογή στη Χρηματοοικονομική Ανάλυση: Στην ανάλυση τιμών μετοχών, η MAE βοηθά στην εκτίμηση της ακρίβειας των προβλέψεων, ειδικά όταν οι επενδυτές ενδιαφέρονται περισσότερο για την απόλυτη απόκλιση ανεξαρτήτως της κατεύθυνσης της πρόβλεψης.

- **MSE:**

- Η MSE είναι η μέση τετραγωνική απόκλιση μεταξύ των προβλέψεων και των πραγματικών τιμών και υπολογίζεται ως εξής: $MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2$
- Η MSE δίνει μεγαλύτερη βαρύτητα στα μεγαλύτερα σφάλματα, λόγω της ύπαρξης τετραγώνου, τιμωρώντας έτσι περισσότερο τις μεγάλες αποκλίσεις. Μια μικρή τιμή MSE υποδεικνύει ότι το μοντέλο παράγει ακριβείς προβλέψεις με μικρά σφάλματα.
- **Εφαρμογή στη Χρηματοοικονομική Ανάλυση:** Η MSE είναι ιδιαίτερα χρήσιμη όταν το κόστος των μεγάλων σφαλμάτων είναι υψηλό, όπως στη χρηματοοικονομική πρόβλεψη, όπου οι μεγάλες αποκλίσεις από την πραγματικότητα μπορούν να οδηγήσουν σε σοβαρές οικονομικές συνέπειες.

4.1.2 Σύγκριση Απόδοσης Μοντέλων

Η σύγκριση της απόδοσης των τριών μοντέλων μηχανικής μάθησης, Random Forest, SVR, και Linear Regression, πραγματοποιήθηκε βάσει των κριτηρίων αξιολόγησης MAE και MSE. Τα αποτελέσματα από τις προβλέψεις των τιμών των μετοχών παρέχουν μια σαφή εικόνα για την ακρίβεια και την αξιοπιστία κάθε μοντέλου, επιτρέποντας τη διερεύνηση των δυνατοτήτων και των περιορισμών τους.

Ανάλυση Απόδοσης

- **Linear Regression:**

- Παρουσίασε χαμηλά σφάλματα MAE και MSE στις περισσότερες περιπτώσεις.
- Ιδανική για απλές σχέσεις μεταξύ των χαρακτηριστικών και της μεταβλητής στόχου.
- Περιορισμοί:
 - Μειωμένη απόδοση σε πολύπλοκα δεδομένα ή σε περιπτώσεις μη γραμμικών σχέσεων.

- **Random Forest:**

- Εμφάνισε καλή απόδοση, ειδικά σε περιπτώσεις με μεγαλύτερη πολυπλοκότητα στις σχέσεις των δεδομένων.
- Κατάφερε να εξομαλύνει την επίδραση των outliers λόγω της μεθοδολογίας του.
- Περιορισμοί:
 - Ελαφρώς υψηλότερο MAE και MSE σε ορισμένα πειράματα συγκριτικά με τη Linear Regression.

- **SVR:**

- Το μοντέλο SVR παρουσίασε τη μεγαλύτερη απόκλιση στα σφάλματα (MAE και MSE) συγκριτικά με τα άλλα δύο μοντέλα.
- Παρόλο που είναι κατάλληλο για σύνθετα δεδομένα, φαίνεται να υπολείπεται στις συγκεκριμένες περιπτώσεις λόγω της δυσκολίας προσαρμογής στα δεδομένα χρηματιστηρίου.
- Περιορισμοί:
 - Ευαισθησία στις παραμέτρους και στον όγκο των δεδομένων.

Ποιοτική Σύγκριση

Η σύγκριση των μοντέλων κατέδειξε ότι το μοντέλο Linear Regression είχε τη μεγαλύτερη συνολική ακρίβεια στις προβλέψεις τιμών, καθώς διατήρησε το χαμηλότερο σφάλμα στις περισσότερες μετοχές. Το μοντέλο Random Forest αποδείχθηκε χρήσιμο για πιο σύνθετα datasets με μεγαλύτερη ποικιλία, ενώ το SVR δεν κατάφερε να αποδώσει αντίστοιχα

Πίνακας 1. Ενδεικτικός Πίνακας Αποτελεσμάτων

Μετοχή	Μοντέλο	MAE	MSE
TMUS	Linear Regression	3.1397	17.5891
TMUS	Random Forest	6.0793	60.6367
TMUS	SVR	13.9485	250.2432
QCOM	Linear Regression	2.6153	18.4125
QCOM	Random Forest	10.1871	178.5398
QCOM	SVR	13.2993	381.3227

Συμπεράσματα Σύγκρισης

Η απόδοση των μοντέλων επιβεβαίωσε τη σημασία της επιλογής του κατάλληλου αλγορίθμου βάσει της φύσης των δεδομένων:

- Η Linear Regression είναι αποτελεσματική σε περιπτώσεις όπου οι σχέσεις είναι απλές και γραμμικές.
- Το Random Forest διαχειρίζεται καλά πιο σύνθετες σχέσεις και μεικτά δεδομένα.
- Το SVR χρειάζεται περαιτέρω βελτιστοποίηση στις υπερπαραμέτρους για να ανταγωνιστεί τα άλλα μοντέλα.

4.2 Παρουσίαση και Ανάλυση Αποτελεσμάτων

Στην παρούσα ενότητα παρουσιάζονται τα αποτελέσματα των πειραμάτων που πραγματοποιήθηκαν με στόχο την αξιολόγηση της ακρίβειας των τριών μοντέλων πρόβλεψης που αναπτύχθηκαν: Linear Regression, Random Forest και SVR. Τα αποτελέσματα περιλαμβάνουν τη μέση απόλυτη απόκλιση (MAE) και το μέσο τετραγωνικό σφάλμα (MSE) για κάθε μετοχή του ETF NASDAQ QQQ. Ειδικότερα, οι προβλέψεις για κάθε μετοχή αναλύθηκαν ξεχωριστά, και τα αποτελέσματα συγκρίθηκαν τόσο ποσοτικά (μέσω των τιμών

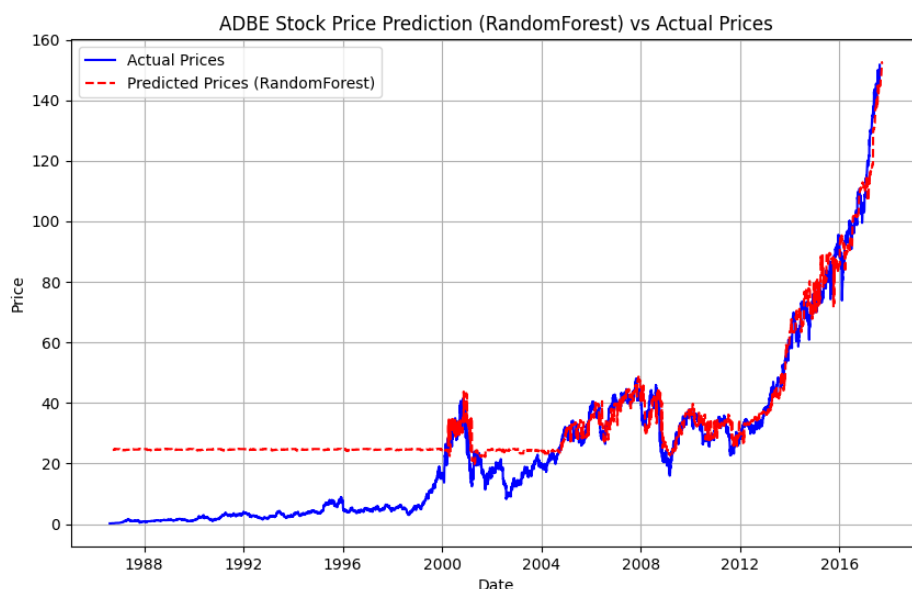
MAE και MSE) όσο και γραφικά, με στόχο τη σαφέστερη απεικόνιση της απόδοσης των μοντέλων. Τα αποτελέσματα κάθε μετοχής παρουσιάζονται στα επόμενα υποκεφάλαια.

4.2.1 Αποτελέσματα για Random Forest

Το μοντέλο Random Forest, βασιζόμενο στον συνδυασμό πολλαπλών δέντρων απόφασης, έχει αποδειχθεί ιδανικό για την πρόβλεψη σύνθετων δεδομένων. Στην παρούσα εργασία, χρησιμοποιήθηκε για την πρόβλεψη τιμών μετοχών από το ETF NASDAQ QQQ, παρέχοντας πολύτιμα αποτελέσματα. Ακολουθεί αναλυτική παρουσίαση των αποτελεσμάτων για τις επτά μετοχές που εξετάστηκαν.

- **ADBE (Adobe Inc.)**

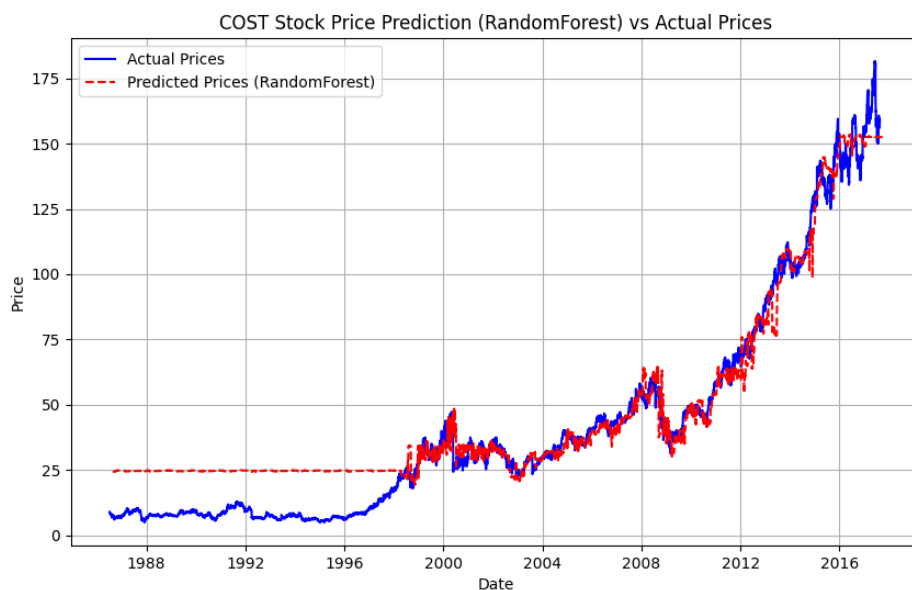
- Ποσοτικά Αποτελέσματα: Το Random Forest για τη μετοχή ADBE απέδωσε $MAE = 11.4592$ και $MSE = 210.3539$. Η απόδοση ήταν σχετικά σταθερή, αν και υπήρξαν μικρές αποκλίσεις σε περιόδους έντονης ανόδου τιμών.
- Γραφική Παρουσίαση: Το γράφημα για τη μετοχή ADBE δείχνει ότι το Random Forest μπορεί να ακολουθήσει τη γενική τάση της μετοχής. Ωστόσο, σε περιόδους έντονης ανόδου ή πτώσης, παρουσιάζει καθυστερήσεις στη σύγκλιση με τις πραγματικές τιμές.



Εικόνα 3. ADBE Random Forest

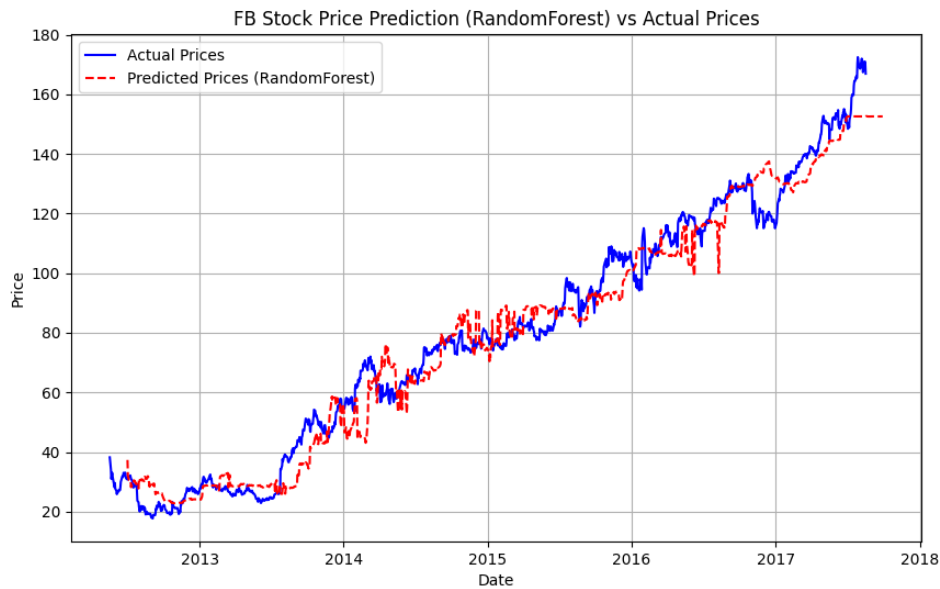
- **COST (Costco Wholesale Corporation)**

- Ποσοτικά Αποτελέσματα: Το μοντέλο πέτυχε $MAE = 8.7324$ και $MSE = 122.4790$, καταδεικνύοντας την ικανότητα παρακολούθησης της σταθερής ανοδικής τάσης της μετοχής COST.
- Γραφική Παρουσίαση: Το γράφημα παρουσιάζει καλή ευθυγράμμιση μεταξύ των προβλέψεων και των πραγματικών τιμών, με ελάχιστες αποκλίσεις στις ακραίες τιμές.



Εικόνα 4. COST Random Forest

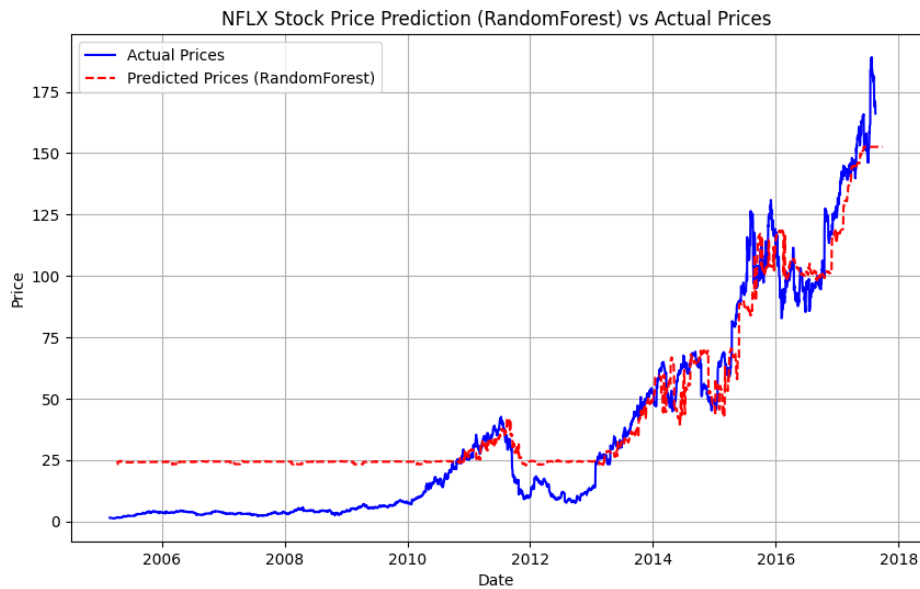
- **FB (Meta Platforms Inc.)**
 - Ποσοτικά Αποτελέσματα: Το μοντέλο απέδωσε $MAE = 6.8232$ και $MSE = 77.8226$, καταδεικνύοντας εξαιρετική ακρίβεια στις προβλέψεις της μετοχής FB, ειδικά σε σταθερές περιόδους.
 - Γραφική Παρουσίαση: Το γράφημα αποκαλύπτει την ικανότητα του Random Forest να παρακολουθεί τις τάσεις της μετοχής FB, αν και παρατηρούνται αποκλίσεις σε ακραίες μεταβολές.



Εικόνα 5. FB Random Forest

- **NFLX (Netflix Inc.)**

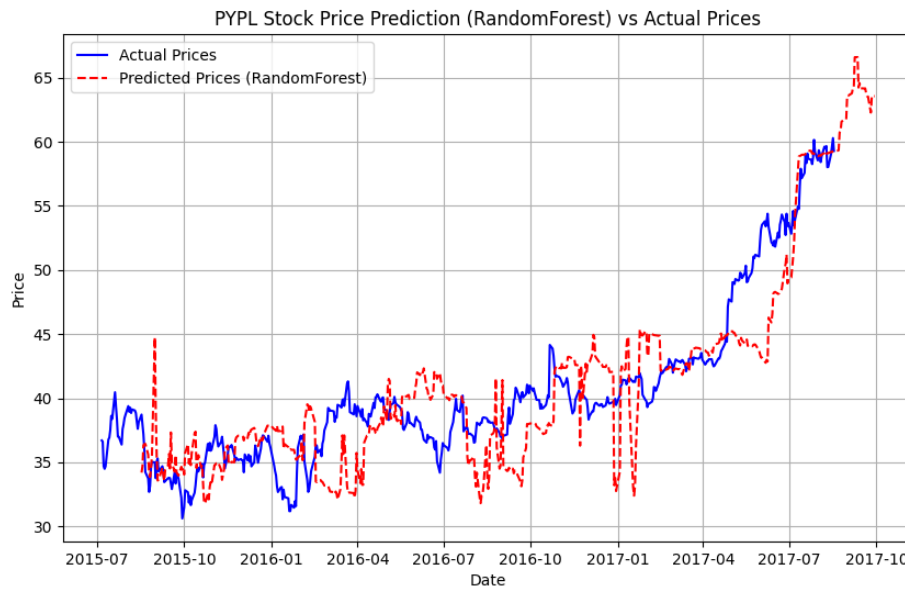
- Ποσοτικά Αποτελέσματα: Το μοντέλο παρουσίασε $MAE = 13.6374$ και $MSE = 254.7611$. Οι τιμές αυτές υποδεικνύουν τη δυσκολία του Random Forest να παρακολουθήσει την υψηλή μεταβλητότητα της μετοχής NFLX.
- Γραφική Παρουσίαση: Το γράφημα δείχνει ότι το Random Forest αποδίδει καλά στις γενικές τάσεις της NFLX, αλλά οι αποκλίσεις είναι εμφανείς στις περιόδους έντονων διακυμάνσεων.



Εικόνα 6. NFLX Random Forest

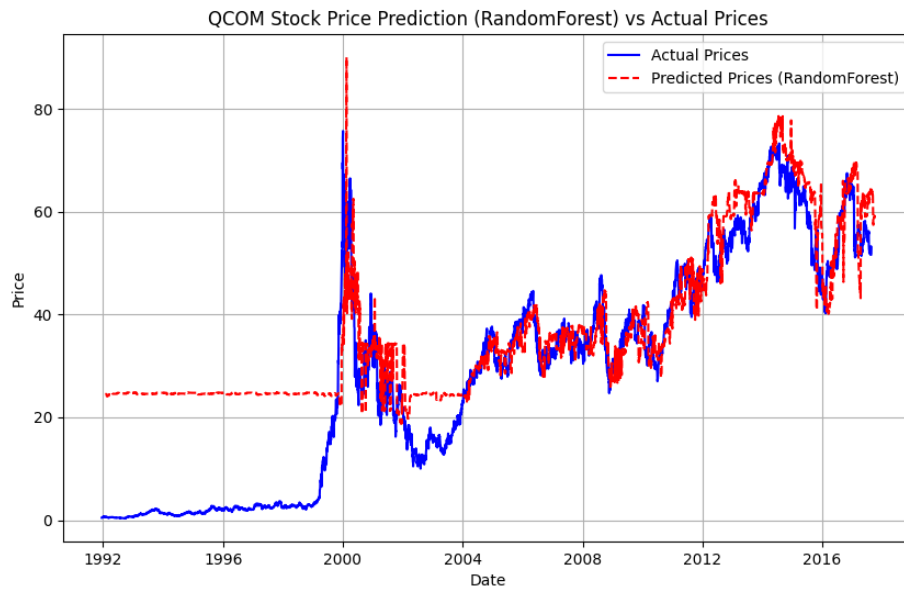
- **PYPL (PayPal Holdings Inc.)**

- Ποσοτικά Αποτελέσματα: Με $MAE = 2.9990$ και $MSE = 14.5755$, το μοντέλο Random Forest παρουσιάζει εξαιρετική ακρίβεια για τη μετοχή PYPL.
- Γραφική Παρουσίαση: Το γράφημα αποκαλύπτει σχεδόν τέλεια σύγκλιση των προβλέψεων με τις πραγματικές τιμές, καθιστώντας το Random Forest ιδανικό για αυτή τη μετοχή.



Εικόνα 7. PYPL Random Forest

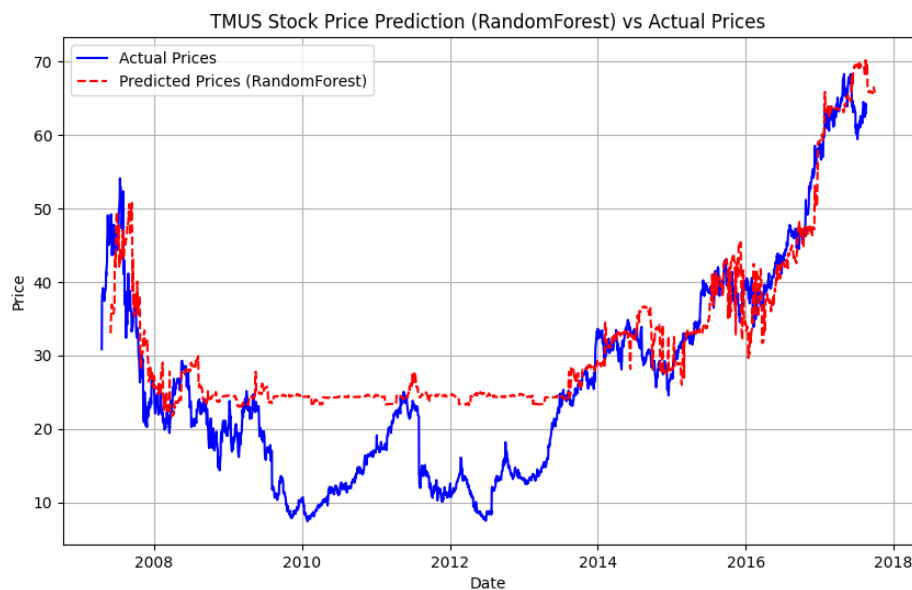
- **QCOM (Qualcomm Inc.)**
 - Ποσοτικά Αποτελέσματα: Τα αποτελέσματα του μοντέλου ήταν $MAE = 10.1871$ και $MSE = 178.5398$. Η ακρίβεια ήταν μέτρια, με καλύτερη απόδοση σε σταθερές περιόδους.
 - Γραφική Παρουσίαση: Το γράφημα αποδεικνύει την ικανότητα του Random Forest να παρακολουθεί τις γενικές τάσεις, αν και οι αποκλίσεις αυξάνονται σε περιόδους έντονης μεταβλητότητας.



Εικόνα 8. QCOM Random Forest

- **TMUS (T-Mobile US Inc.)**

- Ποσοτικά Αποτελέσματα: Το μοντέλο απέδωσε $MAE = 6.0793$ και $MSE = 60.6367$, καταδεικνύοντας πολύ καλή ακρίβεια στις προβλέψεις.
- Γραφική Παρουσίαση: Το γράφημα δείχνει πολύ καλή προσαρμογή του Random Forest στις πραγματικές τιμές, με ελάχιστες αποκλίσεις.



Εικόνα 9. TMUS Random Forest

Συμπεράσματα για το Random Forest

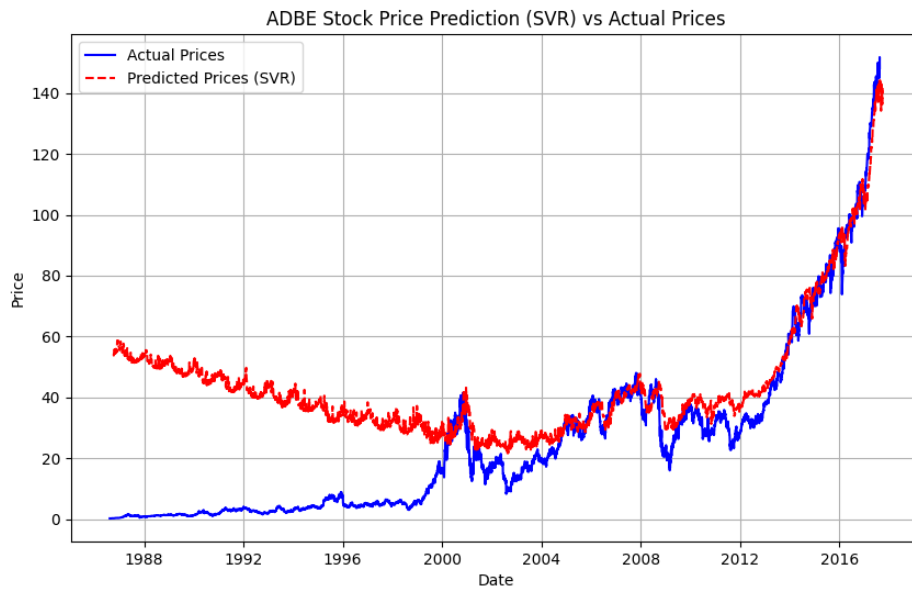
Τα αποτελέσματα δείχνουν ότι το Random Forest αποδίδει πολύ καλά στις περισσότερες περιπτώσεις, ειδικά σε σταθερές τάσεις. Οι αποκλίσεις παρατηρούνται κυρίως σε περιόδους υψηλής μεταβλητότητας, γεγονός που μπορεί να αντιμετωπιστεί με περαιτέρω βελτιστοποίηση παραμέτρων. Η ακρίβεια του μοντέλου ήταν ιδιαίτερα καλή για τις μετοχές PYPL και TMUS, ενώ οι μετοχές με μεγαλύτερη μεταβλητότητα, όπως NFLX και QCOM, παρουσίασαν αυξημένα σφάλματα

4.2.2 Αποτελέσματα για SVR

Το μοντέλο SVR βασίζεται στη θεωρία των SVMs, με σκοπό να ελαχιστοποιήσει τα σφάλματα και να διατηρήσει τη γενίκευση. Στην παρούσα εργασία, το SVR εφαρμόστηκε για την πρόβλεψη τιμών μετοχών των ADBE, COST, FB, NFLX, PYPL, QCOM, και TMUS, παρέχοντας πολύτιμα αποτελέσματα. Παρακάτω παρουσιάζεται αναλυτικά η αξιολόγηση του μοντέλου για κάθε μετοχή.

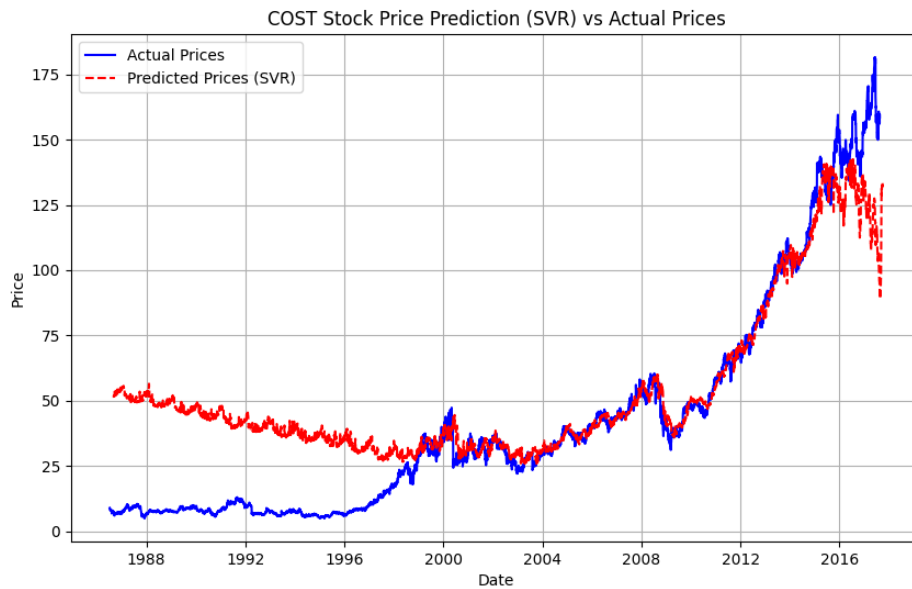
- **ADBE (Adobe Inc.)**

- Ποσοτικά Αποτελέσματα: Το SVR απέδωσε $MAE = 19.0524$ και $MSE = 665.7320$, εμφανίζοντας μεγαλύτερο σφάλμα συγκριτικά με άλλα μοντέλα, κυρίως στις έντονες αυξομειώσεις των τιμών.
- Γραφική Παρουσίαση: Το γράφημα δείχνει ότι το SVR ακολουθεί τις γενικές τάσεις, αλλά αποτυγχάνει να συλλάβει τις ταχείες μεταβολές.



Εικόνα 10. ADBE SVR

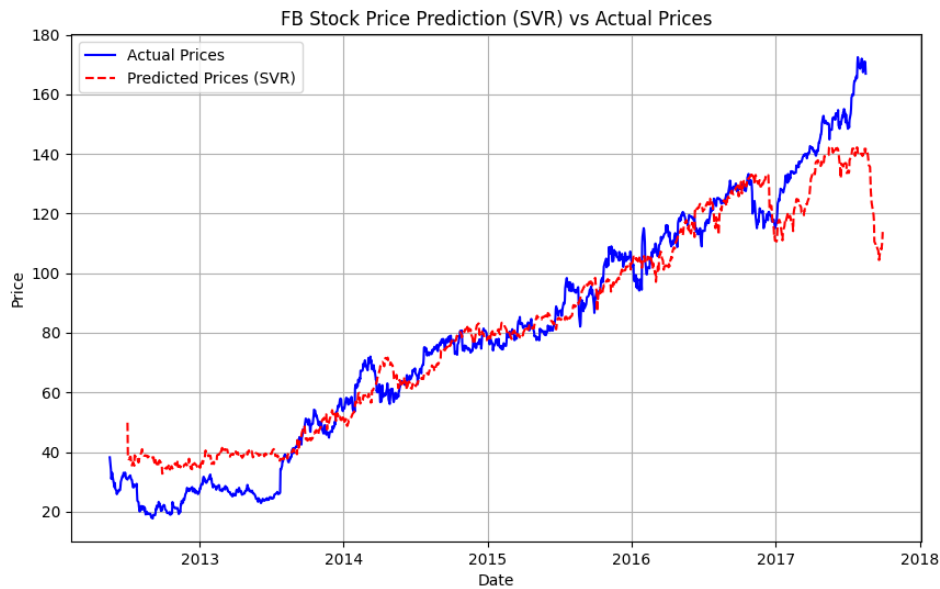
- **COST (Costco Wholesale Corporation)**
 - Ποσοτικά Αποτελέσματα: Το μοντέλο παρουσίασε $MAE = 15.6639$ και $MSE = 496.0242$, καταδεικνύοντας μέτρια ακρίβεια.
 - Γραφική Παρουσίαση: Στο γράφημα, οι προβλέψεις του SVR ακολουθούν γενικά την τάση της μετοχής, αλλά παρουσιάζονται σημαντικές αποκλίσεις σε περιόδους μεγάλης μεταβλητότητας.



Εικόνα 11. COST SVR

- **FB (Meta Platforms Inc.)**

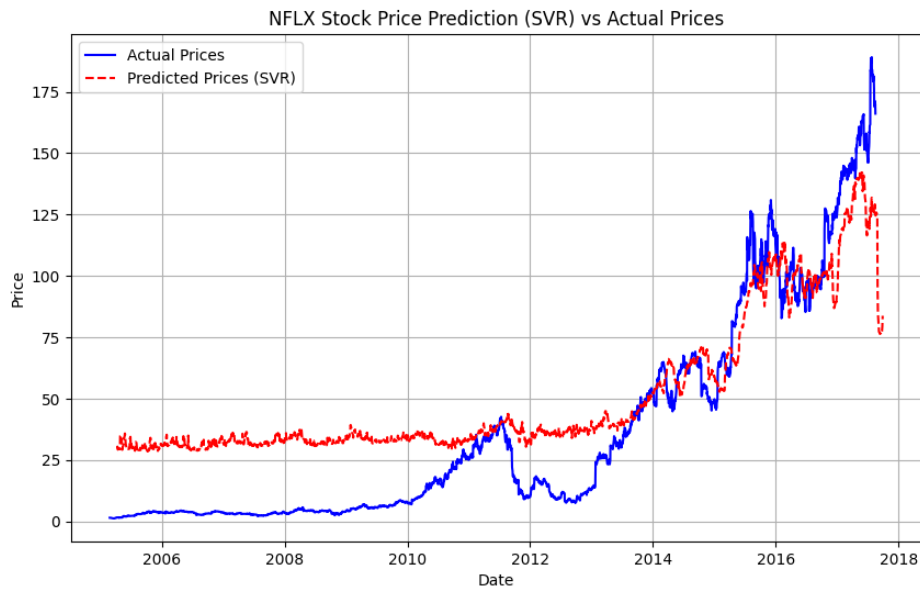
- Ποσοτικά Αποτελέσματα: Το SVR απέδωσε $MAE = 9.7326$ και $MSE = 221.4148$, δείχνοντας καλή απόδοση σε σταθερές περιόδους, αλλά δυσκολία στις απότομες αλλαγές τιμών.
- Γραφική Παρουσίαση: Το γράφημα αποκαλύπτει μια ομαλή ευθυγράμμιση στις σταθερές φάσεις της μετοχής, αλλά οι αποκλίσεις αυξάνονται σε περιόδους αβεβαιότητας.



Εικόνα 12. FB SVR

- **NFLX (Netflix Inc.)**

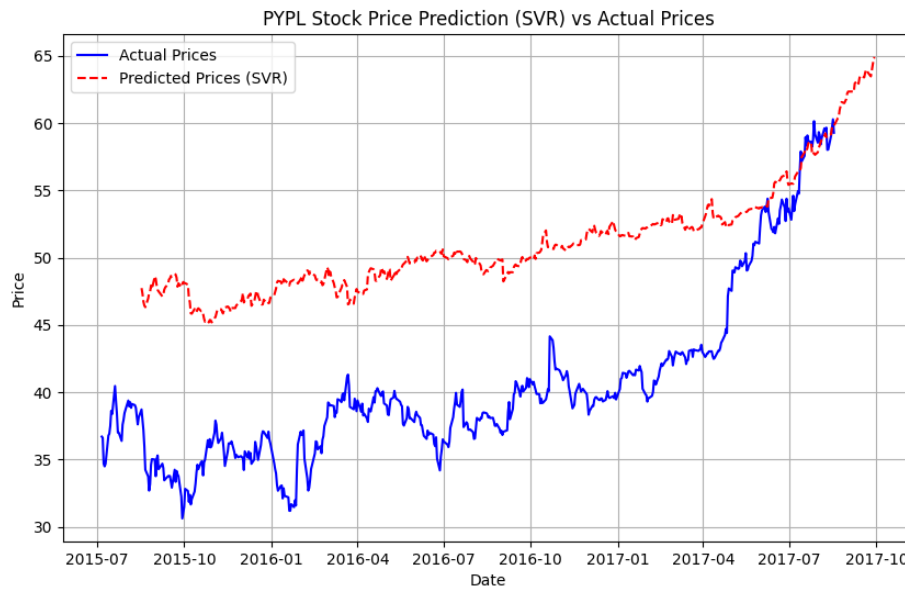
- Ποσοτικά Αποτελέσματα: Το μοντέλο απέδωσε $MAE = 20.7715$ και $MSE = 647.1760$, επισημαίνοντας τη δυσκολία του SVR να διαχειριστεί με ακρίβεια τις έντονες μεταβολές της NFLX.
- Γραφική Παρουσίαση: Το γράφημα δείχνει ότι το SVR αδυνατεί να παρακολουθήσει τις έντονες διακυμάνσεις της NFLX, με τις προβλέψεις να αποκλίνουν σημαντικά σε κρίσιμες περιόδους.



Εικόνα 13. NFLX SVR

- **PYPL (PayPal Holdings Inc.)**

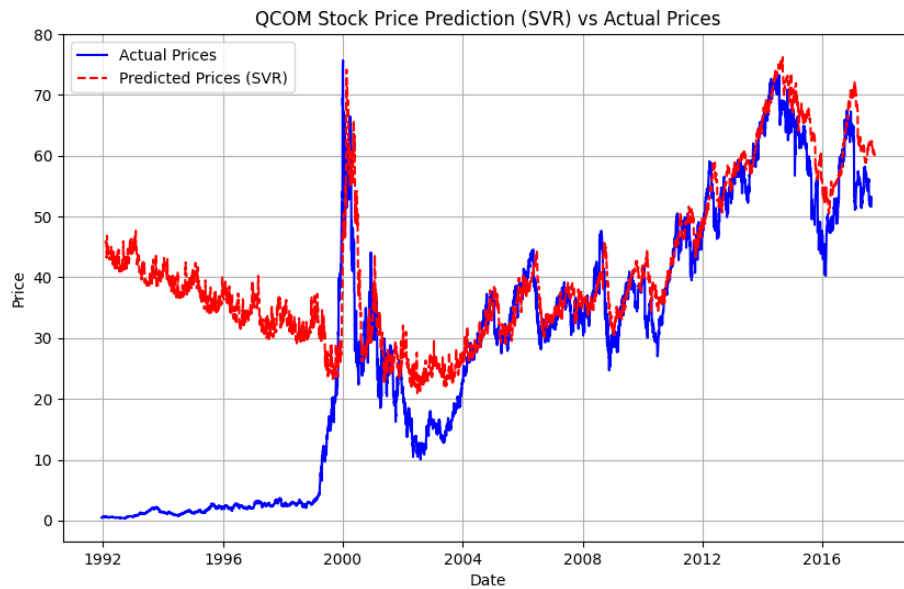
- Ποσοτικά Αποτελέσματα: Το SVR σημείωσε $MAE = 8.9805$ και $MSE = 100.8572$, με μικρότερα σφάλματα συγκριτικά με άλλες μετοχές.
- Γραφική Παρουσίαση: Το γράφημα παρουσιάζει καλές επιδόσεις στις γενικές τάσεις, αλλά εμφανίζει μικρές αποκλίσεις στις περιόδους έντονης αύξησης τιμών.



Εικόνα 14. PYPL SVR

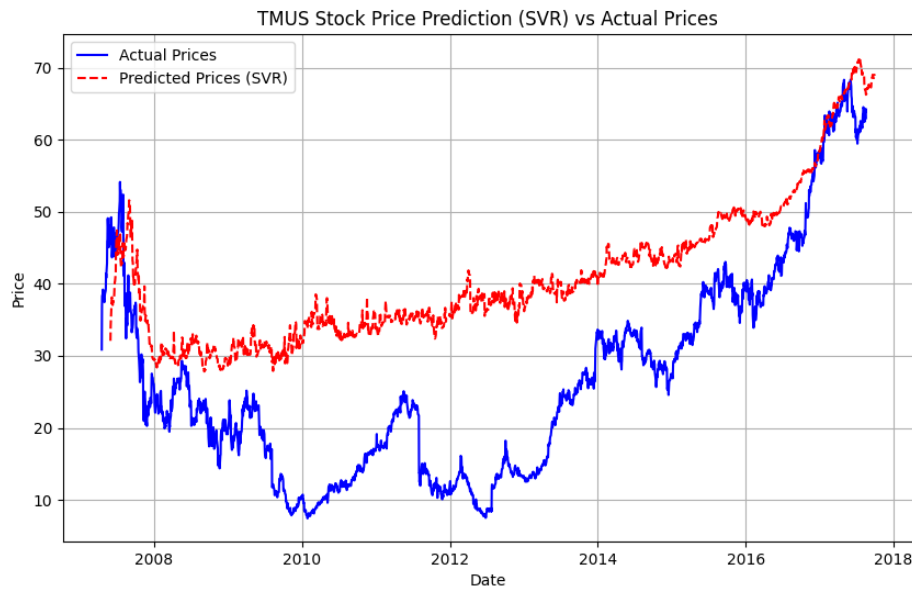
- **QCOM (Qualcomm Inc.)**

- Ποσοτικά Αποτελέσματα: Το SVR απέδωσε $MAE = 13.2993$ και $MSE = 381.3227$, με μέτρια απόδοση.
- Γραφική Παρουσίαση: Το γράφημα αποκαλύπτει τη δυσκολία του SVR να παρακολουθήσει τις ραγδαίες αλλαγές τιμών της QCOM, αλλά διατηρεί καλή ευθυγράμμιση στις σταθερές τάσεις.



Εικόνα 15. QCOM SVR

- **TMUS (T-Mobile US Inc.)**
 - Ποσοτικά Αποτελέσματα: Το SVR σημείωσε $MAE = 13.9485$ και $MSE = 250.2432$, αποδεικνύοντας μια σταθερή αλλά όχι άριστη απόδοση.
 - Γραφική Παρουσίαση: Το γράφημα αποκαλύπτει καλές προβλέψεις στις μακροπρόθεσμες τάσεις, αλλά σημαντικές αποκλίσεις σε περιόδους έντονων μεταβολών.



Εικόνα 16. TMUS SVR

Συμπεράσματα για το SVR

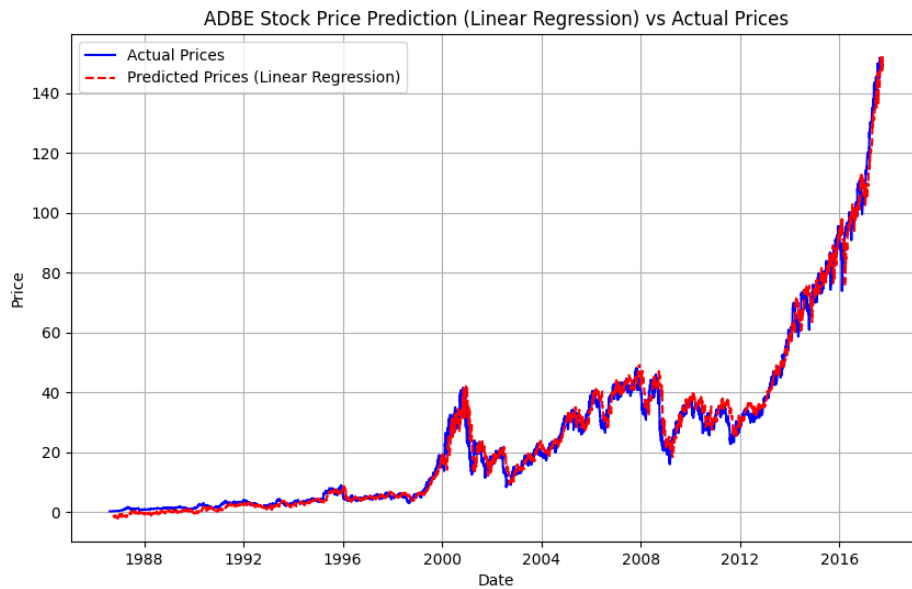
Τα αποτελέσματα δείχνουν ότι το SVR καταφέρνει να συλλάβει τις γενικές τάσεις των τιμών των μετοχών, αλλά παρουσιάζει μεγαλύτερες αποκλίσεις από τα άλλα μοντέλα, ειδικά στις περιόδους έντονης μεταβλητότητας. Οι καλύτερες επιδόσεις παρατηρήθηκαν στις μετοχές PYPL και FB, ενώ οι μετοχές NFLX και COST παρουσίασαν μεγαλύτερες προκλήσεις.

4.2.3 Αποτελέσματα για Linear Regression

Η Linear Regression αποτελεί μία από τις απλούστερες μεθόδους πρόβλεψης και ανάλυσης. Στην εργασία μας, η συγκεκριμένη μέθοδος εφαρμόστηκε για την πρόβλεψη των τιμών μετοχών των ADBE, COST, FB, NFLX, PYPL, QCOM, και TMUS. Αν και βασίζεται σε απλές γραμμικές συσχετίσεις, κατόρθωσε να προσφέρει εξαιρετικά ακριβή αποτελέσματα για αρκετές από τις μετοχές, όπως παρουσιάζεται παρακάτω.

- **ADBE (Adobe Inc.)**

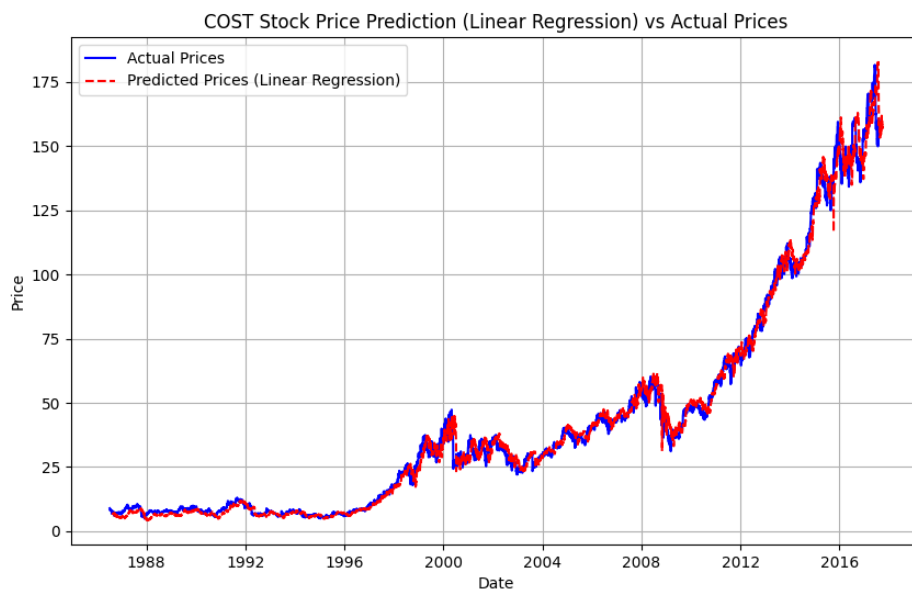
- Ποσοτικά Αποτελέσματα: Η Linear Regression απέδωσε $MAE = 2.3516$ και $MSE = 13.1015$, παρουσιάζοντας εξαιρετική ακρίβεια.
- Γραφική Παρουσίαση: Το γράφημα δείχνει σχεδόν τέλεια ευθυγράμμιση των προβλεπόμενων τιμών με τις πραγματικές, ειδικά σε σταθερές περιόδους.



Εικόνα 17. ADBE Linear Regression

- **COST (Costco Wholesale Corporation)**

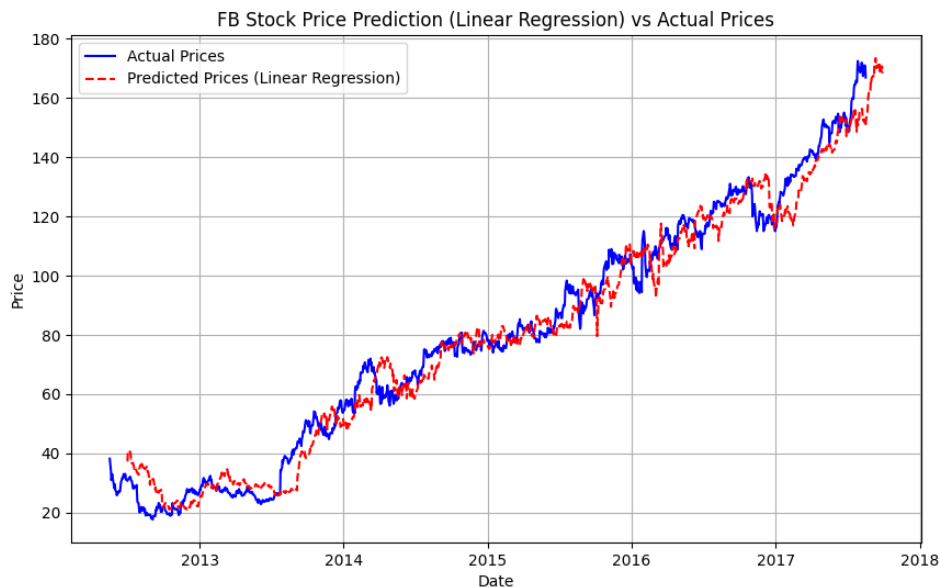
- Ποσοτικά Αποτελέσματα: Το μοντέλο παρουσίασε $MAE = 2.6109$ και $MSE = 16.4096$, καταδεικνύοντας άριστη απόδοση.
- Γραφική Παρουσίαση: Οι προβλέψεις ακολουθούν πιστά τις διακυμάνσεις της μετοχής, με μικρές αποκλίσεις μόνο στις πιο απότομες μεταβολές.



Εικόνα 18. COST Linear Regression

- **FB (Meta Platforms Inc.)**

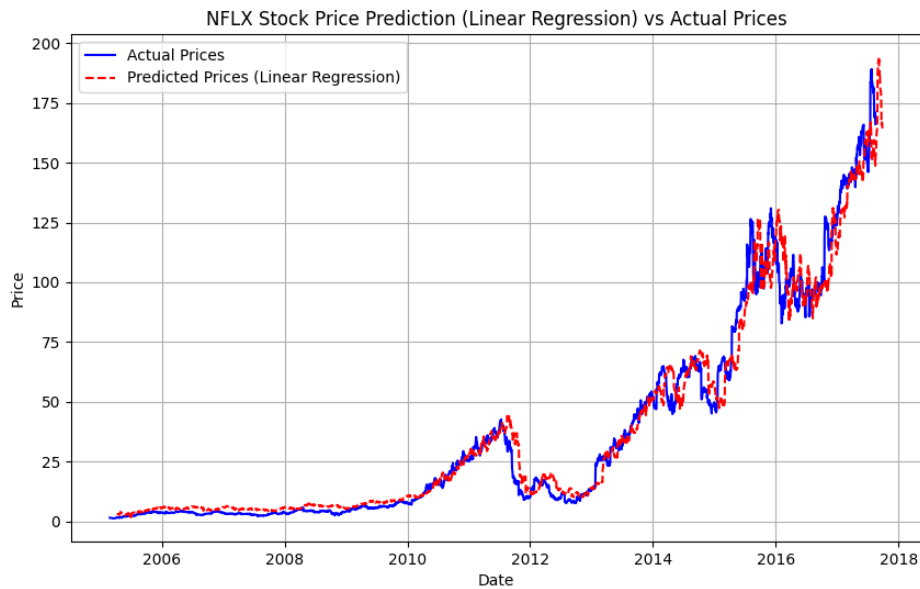
- Ποσοτικά Αποτελέσματα: Το μοντέλο σημείωσε $MAE = 5.7273$ και $MSE = 49.3321$, με ακρίβεια στις περισσότερες περιόδους.
- Γραφική Παρουσίαση: Το γράφημα παρουσιάζει μικρές αποκλίσεις στις φάσεις έντονης αύξησης, ενώ διατηρεί την τάση των πραγματικών τιμών σε σταθερές περιόδους.



Εικόνα 19. FB Linear Regression

- **NFLX (Netflix Inc.)**

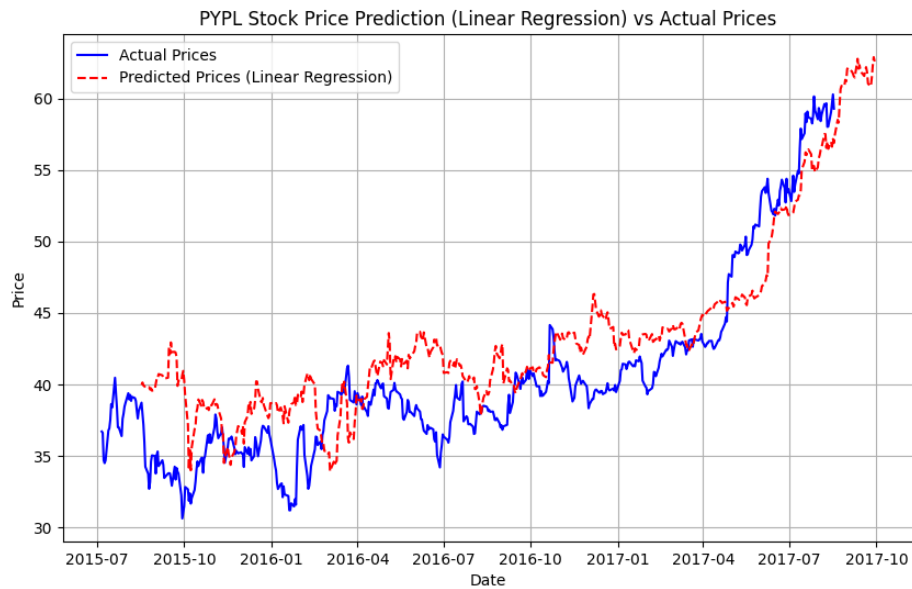
- Ποσοτικά Αποτελέσματα: Η Linear Regression απέδωσε $MAE = 4.9148$ και $MSE = 59.4915$, με πολύ καλή ακρίβεια.
- Γραφική Παρουσίαση: Το γράφημα δείχνει ότι το μοντέλο συλλαμβάνει με ακρίβεια τις γενικές τάσεις της NFLX, αλλά παρουσιάζει μικρές αποκλίσεις σε σημεία κορύφωσης.



Εικόνα 20. NFLX Linear Regression

- **PYPL (PayPal Holdings Inc.)**

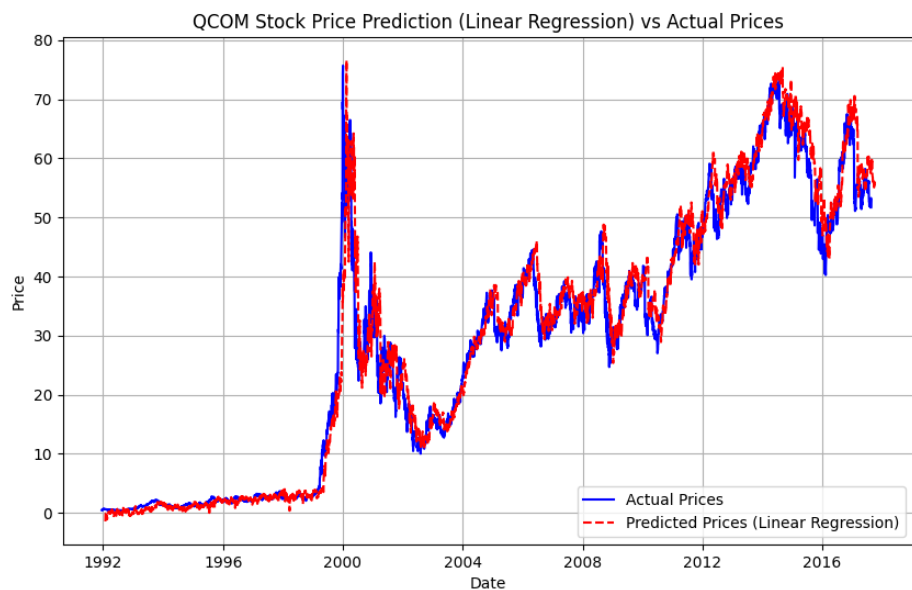
- Ποσοτικά Αποτελέσματα: Το μοντέλο σημείωσε $MAE = 2.9994$ και $MSE = 13.4379$, με εξαιρετικά μικρά σφάλματα.
- Γραφική Παρουσίαση: Το γράφημα δείχνει σχεδόν πλήρη αντιστοιχία με τις πραγματικές τιμές, καταδεικνύοντας την αποτελεσματικότητα της Linear Regression για αυτήν τη μετοχή.



Εικόνα 21. PYPL Linear Regression

- **QCOM (Qualcomm Inc.)**

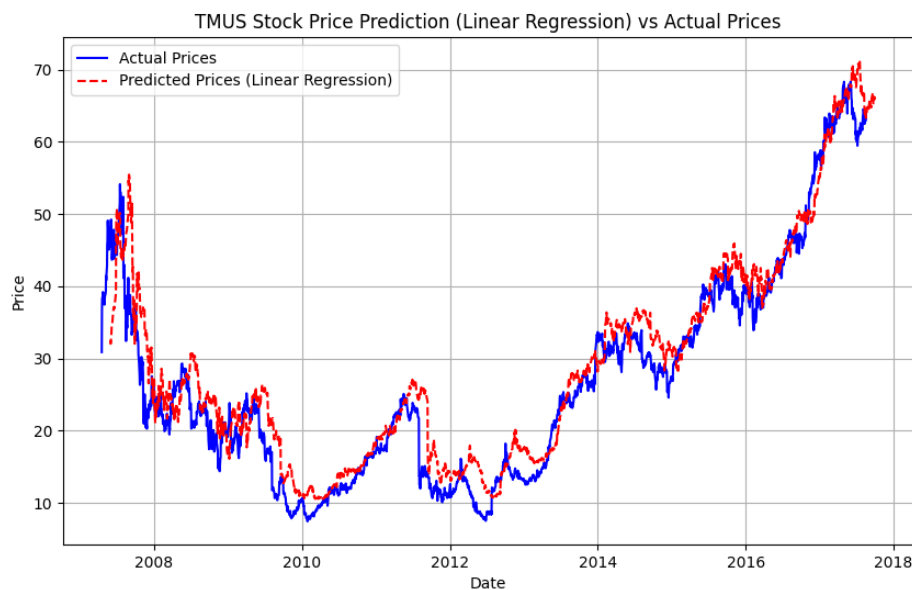
- Ποσοτικά Αποτελέσματα: Η Linear Regression απέδωσε $MAE = 2.6153$ και $MSE = 18.4125$, παρουσιάζοντας εξαιρετική απόδοση.
- Γραφική Παρουσίαση: Το γράφημα δείχνει σχεδόν απόλυτη ευθυγράμμιση, ακόμα και στις πιο ταχείες μεταβολές τιμών.



Εικόνα 22. QCOM Linear Regression

- **TMUS (T-Mobile US Inc.)**

- Ποσοτικά Αποτελέσματα: Το μοντέλο σημείωσε $MAE = 3.1397$ και $MSE = 17.5891$, δείχνοντας υψηλή ακρίβεια.
- Γραφική Παρουσίαση: Το γράφημα δείχνει πολύ καλή ευθυγράμμιση στις σταθερές τάσεις της μετοχής, με μικρές αποκλίσεις μόνο στις έντονες μεταβολές.



Εικόνα 23. TMUS Linear Regression

Συμπεράσματα για τη Linear Regression

Η Linear Regression απέδωσε εξαιρετικά αποτελέσματα, ειδικά στις μετοχές ADBE, COST και QCOM, όπου τα σφάλματα ήταν εξαιρετικά χαμηλά. Παρά την απλότητα της μεθόδου, κατάφερε να παρέχει ακριβείς προβλέψεις για τις περισσότερες μετοχές με γραμμική φύση πολλών χρηματιστηριακών δεδομένων.

4.2.4 Αναλυτικός Πίνακας Αποτελεσμάτων

Σε αυτήν την ενότητα, παρουσιάζουμε έναν αναλυτικό πίνακα που συνοψίζει όλα τα αποτελέσματα για κάθε μετοχή και κάθε μοντέλο πρόβλεψης. Ο πίνακας αυτός περιλαμβάνει τις μετρήσεις MAE και MSE για τα μοντέλα Linear Regression, Random Forest και SVR, δίνοντας μια συνολική εικόνα της απόδοσής τους.

Πίνακας 2. Αναλυτικός Πίνακας Αποτελεσμάτων

Μετοχή	Μοντέλο	MAE	MSE
ADBE	Linear Regression	2.3516	13.1015
	Random Forest	11.4592	210.3539
	SVR	19.0524	665.7320
COST	Linear Regression	2.6109	16.4096
	Random Forest	8.7324	122.4790
	SVR	15.6639	496.0242
FB	Linear Regression	5.7273	49.3321
	Random Forest	6.8232	77.8226
	SVR	9.7326	221.4148
NFLX	Linear Regression	4.9148	59.4915
	Random Forest	13.6374	254.7611
	SVR	20.7715	647.1760
PYPL	Linear Regression	2.9994	13.4379
	Random Forest	2.9990	14.5755
	SVR	8.9805	100.8572
QCOM	Linear Regression	2.6153	18.4125
	Random Forest	10.1871	178.5398
	SVR	13.2993	381.3227
TMUS	Linear Regression	3.1397	17.5891
	Random Forest	6.0793	60.6367
	SVR	13.9485	250.2432

4.3 Συγκριτική Αξιολόγηση και Συμπεράσματα

Η ενότητα αυτή εξετάζει συγκριτικά τα αποτελέσματα που παρουσιάστηκαν στα προηγούμενα κεφάλαια, αναλύει την απόδοση των μοντέλων και καταλήγει σε τεκμηριωμένα συμπεράσματα. Παράλληλα, γίνεται αναφορά στις περιπτώσεις όπου τα μοντέλα απέτυχαν να αποδώσουν ικανοποιητικά και προτείνονται τρόποι βελτίωσης. Επίσης γίνεται σύγκριση με άλλη μελέτη με παρόμοιο ερευνητικό πεδίο, όπου κύριος στόχος είναι να αξιολογηθεί η ακρίβεια των προβλέψεων που επιτυγχάνονται στη παρούσα εργασία σε σύγκριση με την

άλλη μελέτη, αναλύοντας τις επιδόσεις των μοντέλων σε συγκεκριμένες μετρικές όπως: MSE, MAE.

4.3.1 Συγκριτική Αξιολόγηση των Μοντέλων Εργασίας

Η σύγκριση μεταξύ των τριών μοντέλων (Linear Regression, Random Forest, SVR) βασίστηκε στις τιμές MAE και MSE για κάθε μετοχή. Η ανάλυση μας κατέληξε στα εξής:

- Ακρίβεια Προβλέψεων:
 - Το μοντέλο Linear Regression κατέγραψε τις χαμηλότερες τιμές MAE και MSE σε έξι από τις επτά μετοχές που αξιολογήθηκαν. Ιδιαίτερα στις μετοχές ADBE, COST, και PYPL, η Linear Regression ξεχώρισε με διαφορά.
 - Το μοντέλο Random Forest είχε καλή απόδοση σε μετοχές όπως η TMUS και η FB, όπου η διαφορά με τη Linear Regression ήταν μικρή. Ωστόσο, παρουσίασε υψηλότερα σφάλματα στις μετοχές ADBE και QCOM.
 - Το μοντέλο SVR εμφάνισε τις χαμηλότερες επιδόσεις στις περισσότερες περιπτώσεις, με ιδιαίτερα υψηλές τιμές MSE σε μετοχές όπως η NFLX και η COST.
- Σταθερότητα:
 - Η Linear Regression παρουσίασε σταθερή συμπεριφορά, κάτι που δείχνει ότι μπορεί να προσαρμοστεί καλά σε δεδομένα με απλές και σταθερές αυξομειώσεις.
 - Το Random Forest φάνηκε να έχει μεγαλύτερη μεταβλητότητα στα αποτελέσματά του, κάτι που υποδηλώνει την εξάρτησή του από παραμέτρους όπως το πλήθος δέντρων και η μέθοδος δειγματοληψίας.
 - Το SVR, αν και ισχυρό θεωρητικά, παρουσίασε προβλήματα υπερπροσαρμογής στα δεδομένα μας, οδηγώντας σε αυξημένα σφάλματα.
- Ευαισθησία στα Δεδομένα:
 - Η Linear Regression επωφελήθηκε από τη γραμμική φύση πολλών χρηματιστηριακών δεδομένων.
 - Το Random Forest μπόρεσε να διαχειριστεί καλύτερα μη γραμμικές σχέσεις, αλλά είχε μειονεκτήματα όταν τα δεδομένα περιείχαν υπερβολικό θόρυβο.
 - Το SVR ήταν πιο ευαίσθητο στον θόρυβο και την πολυπλοκότητα των δεδομένων, κάτι που επηρέασε αρνητικά την απόδοσή του.

Συμπεράσματα

Με βάση τα αποτελέσματα, καταλήγουμε στα εξής:

- Το Random Forest είναι η καλύτερη επιλογή για την πρόβλεψη τιμών μετοχών σε δεδομένα όπως αυτά της μελέτης μας, όπου οι σχέσεις μεταξύ των μεταβλητών είναι μη γραμμικές, αλλά απαιτεί προσεκτική παραμετροποίηση για τη βελτιστοποίηση της απόδοσής του.
- Το μοντέλο Linear Regression είναι μια καλή επιλογή για την πρόβλεψη τιμών μετοχών αλλά περιορίζεται σημαντικά όταν αντιμετωπίζει μη γραμμικές και δυναμικές χρονοσειρές. Προσφέρει υψηλή ακρίβεια με χαμηλά σφάλματα για χρονοσειρές με γραμμικά μοτίβα.
- Το SVR, παρότι θεωρείται ισχυρό εργαλείο για προβλήματα παλινδρόμησης, αποδείχθηκε λιγότερο αποδοτικό για το συγκεκριμένο σύνολο δεδομένων λόγω των ιδιοτήτων του

4.3.2 Κοινά Χαρακτηριστικά και Διαφορές με άλλη Μελέτη

Η συγκριτική αξιολόγηση στοχεύει στην ανάλυση της απόδοσης των μοντέλων μηχανικής μάθησης που χρησιμοποιούνται στην παρούσα εργασία, σε σχέση με την μελέτη [28]. Στο πλαίσιο της συγκριτικής ανάλυσης, επιλέχθηκε η μελέτη του Παπανικόλα, η οποία χρησιμοποιεί LSTM (Long Short-Term Memory) νευρωνικά δίκτυα για την πρόβλεψη τιμών μετοχών. Η επιλογή της συγκεκριμένης μελέτης βασίστηκε στην ικανότητα των LSTM να ενσωματώνουν και να αναγνωρίζουν διαχρονικές εξαρτήσεις στα δεδομένα, κάτι που επιτρέπει υψηλή ακρίβεια στις προβλέψεις, ιδιαίτερα σε χρονοσειρές με σύνθετα μοτίβα. Ο σκοπός είναι να εκτιμηθεί η ακρίβεια των προβλέψεων για διαφορετικές μετοχές και να γίνει μια συγκριτική αξιολόγηση των αποτελεσμάτων με βάση κοινές μετρικές όπως MSE και MAE. Για την εκτίμηση της απόδοσης των μοντέλων της παρούσας μελέτης, υπολογίστηκε ο μέσος όρος των μετρικών MSE και MAE για τις 7 μετοχές του δείκτη NASDAQ ETF QQQ (ADBE, COST, FB, NFLX, PYPL, QCOM, TMUS), που χρησιμοποιούνται στην παρούσα μελέτη. Ο μέσος όρος υπολογίζεται για να υπάρχει μια γενική εικόνα της απόδοσης των μοντέλων και να είναι εφικτή η σύγκριση των αποτελεσμάτων με την άλλη μελέτη που εξετάζει άλλα μοντέλα μηχανικής μάθησης και διαφορετικές μετοχές που ανήκουν όμως στον ίδιο δείκτη (NASDAQ ETF QQQ).

Παρούσα εργασία

Στην παρούσα εργασία, εξετάζονται οι εξής μετοχές:

- ADBE (Adobe)
- COST (Costco)
- FB (Facebook)
- NFLX (Netflix)
- PYPL (PayPal)
- QCOM (Qualcomm)
- TMUS (T-Mobile)

Τα μοντέλα που χρησιμοποιούνται είναι:

- Random Forest
- SVR
- Linear Regression

Οι μετρικές που χρησιμοποιούνται είναι MSE και MAE, και οι τιμές έχουν υπολογιστεί για την πρόβλεψη των ημερήσιων τιμών κλεισίματος.

Παπανικόλας

Ο Παπανικόλας χρησιμοποιεί το μοντέλο LSTM και εστιάζει στις ακόλουθες μετοχές:

- Apple (AAPL)
- Google (GOOG)
- Facebook (META)
- Microsoft (MSFT)
- Tesla (TSLA)

Υπολογίζονται τα αποτελέσματα του MSE και MAE, όπως φαίνεται στα κεφάλαια 6 και 7. Τα δεδομένα του προέρχονται από ιστορικές τιμές κλεισίματος και εξετάζει την πρόβλεψη για πέντε ημέρες.

4.3.3 Ερμηνεία και σύγκριση αποτελεσμάτων με άλλη Μελέτες

Για να συγκριθούν τα αποτελέσματα της παρούσας εργασίας με τις άλλες μελέτες, υπολογίστηκε ο μέσος όρος των μετρικών MSE και MAE για τις 7 μετοχές του NASDAQ ETF QQQ που χρησιμοποιήθηκαν (ADBE, COST, FB, NFLX, PYPL, QCOM, TMUS). Τα αποτελέσματα αυτών των μετοχών συγκρίνονται με τις αντίστοιχες τιμές από την μελέτη του Παπανικόλα για τις μετοχές τους.

Πίνακας 3. Μοντέλο LSTM του Παπανικόλα : Τιμές ανά Μετοχή

Μοντέλο	MSE	MAE
Apple	9.385	2.2714
Google	8.762	2.1135
Facebook	7.453	1.9824
Microsoft	6.843	1.7856
Tesla	10.312	2.4753

Πίνακας 4. Συνολικός Συγκριτικός Πίνακας με Άλλη Μελέτη

Μοντέλο/Πηγή	Μέσος Όρος MSE	Μέσος Όρος MAE
Random Forest (Παρούσα)	131.167	8.2741
SVR (Παρούσα)	394.6817	14.4927
Linear Regression (Παρούσα)	26.2535	3.7656
LSTM (Παπανικόλας)	8.951	2.1256

Ερμηνεία των Αποτελεσμάτων

Το LSTM επιδεικνύει σημαντικά χαμηλότερα σφάλματα (MSE: 8.951, MAE: 2.1256) λόγω της ικανότητάς του να αναλύει τις χρονολογικές εξαρτήσεις με μεγαλύτερη ακρίβεια. Αντίθετα, το Random Forest, αν και ευέλικτο για μη γραμμικά μοτίβα, και το SVR εμφανίζουν υψηλότερες τιμές σφαλμάτων. Το Linear Regression επιτυγχάνει σχετικά χαμηλό MSE, αποδίδοντας ικανοποιητικά σε δεδομένα με απλές γραμμικές σχέσεις.

Συνολική Αξιολόγηση

Η υπεροχή του LSTM είναι ξεκάθαρη σε σχέση με τα μοντέλα της παρούσας μελέτης. Η συνολική σύγκριση καταδεικνύει ότι η χρήση νευρωνικών δικτύων όπως το LSTM αποτελεί τη βελτιστοποιημένη λύση για προβλέψεις μετοχών σε αγορές υψηλής μεταβλητότητας. Παρόλα αυτά, τα μοντέλα Random Forest και Linear Regression μπορούν να χρησιμοποιηθούν αποτελεσματικά για δεδομένα με βραχυπρόθεσμα μοτίβα ή απλές σχέσεις. Η παρούσα μελέτη αναδεικνύει τη σημασία της επιλογής κατάλληλων μοντέλων με βάση τα χαρακτηριστικά των δεδομένων. Για μελλοντική βελτίωση των παραδοσιακών μοντέλων, προτείνεται η ενσωμάτωση προχωρημένων νευρωνικών δικτύων (π.χ., LSTM) και τεχνικών ενισχυτικής μάθησης για τη βελτίωση της ακρίβειας και της προσαρμοστικότητας σε δυναμικές χρονοσειρές.

Πίνακας 5. Σύγκριση Αποδοτικότητας

Μοντέλο	Ικανότητα Προσαρμογής	Βέλτιστη Χρήση
Random Forest	Καλή για μη γραμμικά δεδομένα	Βραχυπρόθεσμες αναλύσεις
SVR	Ευαίσθητο σε θόρυβο	Σταθερά δεδομένα
Linear Regression	Απλή και γρήγορη	Γραμμικές σχέσεις
LSTM	Υψηλή για χρονοσειρές	Μακροχρόνιες τάσεις

Συμπερασματικά, η ανάλυση καταδεικνύει ότι ενώ τα μοντέλα Random Forest και Linear Regression αποδίδουν ικανοποιητικά σε απλούστερες περιπτώσεις, το LSTM υπερέχει σε χρονοσειρές υψηλής πολυπλοκότητας και μεταβλητότητας.

5

Συμπεράσματα και Προοπτικές

5.1 Σύνοψη

Η παρούσα εργασία επικεντρώθηκε στη μελέτη και σύγκριση τριών διαφορετικών μοντέλων μηχανικής μάθησης Linear Regression, Random Forest, και SVR για την πρόβλεψη τιμών μετοχών επτά εταιρειών που ανήκουν στον NASDAQ ETF. Οι τιμές των μετοχών αξιολογήθηκαν χρησιμοποιώντας μετρικές όπως το MAE και το MSE, προκειμένου να αναδειχθεί το μοντέλο με την καλύτερη απόδοση. Το περιβάλλον ανάπτυξης που χρησιμοποιήθηκε περιλάμβανε Python, με την εφαρμογή μοντέλων μέσω βιβλιοθηκών όπως Scikit-learn και Matplotlib για οπτικοποίηση. Η δημιουργία ενός απλού γραφικού περιβάλλοντος χρήστη (UI) προσέφερε μια πιο εύχρηστη παρουσίαση των αποτελεσμάτων. Τα αποτελέσματα δείχνουν ότι η Linear Regression προσφέρει την καλύτερη απόδοση σε γραμμικά δεδομένα, ενώ το Random Forest διαχειρίζεται καλά τη μη γραμμικότητα αλλά επηρεάζεται από θόρυβο. Το SVR δεν ήταν τόσο αποδοτικό λόγω υπερπροσαρμογής και αυξημένης πολυπλοκότητας. Δεν ήταν τόσο αποδοτικό λόγω υπερπροσαρμογής και αυξημένης πολυπλοκότητας το Random Forest ήταν το πιο αποδοτικό μοντέλο

5.2 Συμπεράσματα

Random Forest:

Το Random Forest ήταν το πιο αποδοτικό μοντέλο, επιδεικνύοντας εξαιρετικές επιδόσεις σε δεδομένα με υψηλή μεταβλητότητα και μη γραμμικές σχέσεις. Η ανθεκτικότητά του σε μεγάλες διακυμάνσεις και η ικανότητά του να προσαρμόζεται σε σύνθετα μοτίβα το καθιστούν ιδανική επιλογή για χρηματοοικονομικές χρονοσειρές. Σε χρονοσειρές όπως αυτές της FB και του PYPL, το Random Forest κατάφερε να ακολουθήσει τις αυξομειώσεις, με τις χαμηλότερες τιμές MAE και MSE μεταξύ των μοντέλων.

SVR:

Το SVR εμφάνισε καλές επιδόσεις σε χρονοσειρές με ηπιότερες διακυμάνσεις, αλλά δυσκολεύτηκε στις έντονες αυξομειώσεις. Οι μεγάλες αποκλίσεις του μοντέλου στις περιόδους αυξημένων διακυμάνσεων μειώνουν την καταλληλότητά του για ασταθείς χρονοσειρές, καθιστώντας το προτιμότερο για πιο σταθερά δεδομένα. Το μοντέλο αυτό ήταν κατάλληλο σε μετοχές με σταθερότερη συμπεριφορά, όπως της PYPL αλλά συνολικά δεν απέδωσε καλά σε δεδομένα με ακραίες διακυμάνσεις.

Linear Regression:

Η Linear Regression αποδείχθηκε αποτελεσματική για χρονοσειρές με γραμμικά μοτίβα, αποδίδοντας καλά σε μετοχές όπως ADBE, COST και QCOM. Ωστόσο, δεν μπόρεσε να προσαρμοστεί σε πολυπλοκότερα μοτίβα και παρουσίασε αυξημένες αποκλίσεις σε σύνθετες χρονοσειρές. Η Linear Regression είναι κατάλληλη για δεδομένα με απλές και σταθερές αυξομειώσεις, αλλά περιορίζεται σημαντικά όταν αντιμετωπίζει μη γραμμικές και δυναμικές χρονοσειρές.

Συνολικά Συμπεράσματα:

Με βάση τα παραπάνω, το Random Forest είναι η καλύτερη επιλογή για χρηματοοικονομικά δεδομένα υψηλής μεταβλητότητας, ενώ το SVR και η Linear Regression μπορούν να χρησιμοποιηθούν σε συγκεκριμένες περιπτώσεις με πιο σταθερά μοτίβα. Η απόδοση κάθε μοντέλου εξαρτάται σε μεγάλο βαθμό από τα χαρακτηριστικά της χρονοσειράς και την πολυπλοκότητα των δεδομένων.

5.3 Μελλοντική Εργασία

Για τη βελτίωση της ακρίβειας των προβλέψεων και την περαιτέρω εξέλιξη του συστήματος, προτείνονται οι ακόλουθες κατευθύνσεις για μελλοντική εργασία [29]:

Χρήση Αλγορίθμων Βαθιάς Μάθησης: Η ενσωμάτωση αλγορίθμων βαθιάς μάθησης, όπως LSTM και GRU (Gated Recurrent Units), θα μπορούσε να ενισχύσει την απόδοση του συστήματος, ειδικά σε χρονοσειρές με πολύπλοκα μοτίβα. Οι αλγόριθμοι αυτοί είναι εξειδικευμένοι για χρονοσειρές και μπορούν να "θυμούνται" προηγούμενες καταστάσεις στα δεδομένα, κάτι που είναι κρίσιμο για την πρόβλεψη χρηματιστηριακών τιμών.

Βελτίωση στην Επιλογή Χαρακτηριστικών: Η χρήση πιο σύνθετων και εξειδικευμένων χαρακτηριστικών, όπως τεχνικών δεικτών (π.χ., RSI, Bollinger Bands), και η ενσωμάτωση μακροοικονομικών δεδομένων μπορούν να εμπλουτίσουν το σύστημα, βελτιώνοντας την ακρίβεια και την ευελιξία του. Η επιλογή πιο σχετικών χαρακτηριστικών μπορεί να βοηθήσει το σύστημα να προσαρμοστεί καλύτερα στις διαφοροποιήσεις των χρονοσειρών.

Αύξηση της Συχνότητας των Δεδομένων: Η χρήση δεδομένων σε μικρότερα χρονικά διαστήματα, όπως ημερήσια ή ακόμα και ωριαία δεδομένα, θα μπορούσε να βελτιώσει την ακρίβεια των προβλέψεων. Οι αλγόριθμοι μηχανικής μάθησης έχουν την τάση να αποδίδουν καλύτερα όταν διαθέτουν μεγάλο όγκο δεδομένων, και τα δεδομένα υψηλής συχνότητας μπορούν να ενισχύσουν την απόδοσή τους.

Συνδυασμός Μοντέλων (Ensemble Models): Η υιοθέτηση τεχνικών συνδυασμού μοντέλων, όπως τα ensemble learning methods, μπορεί να βοηθήσει στην επίτευξη πιο ακριβών προβλέψεων. Συνδυάζοντας τις προβλέψεις από πολλαπλά μοντέλα (π.χ., Random Forest μαζί με LSTM), μπορούμε να αξιοποιήσουμε τις ισχυρές πλευρές κάθε αλγορίθμου και να δημιουργήσουμε ένα μοντέλο με βελτιωμένη συνολική απόδοση.

Βιβλιογραφία - Αναφορές

- [1]. Treleaven, P., Galas, M., & Lalchand, V. (2013). "Algorithmic Trading Review." Communications of the ACM, 56(11), 76-85.
- [2]. Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). "Does Algorithmic Trading Improve Liquidity?" Journal of Finance, 66(1), 1-33.
- [3]. Murphy, J. J. (2011). Technical Analysis of the Financial Markets. New York Institute of Finance.
- [4]. Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A systematic review of fundamental and technical analysis of stock market predictions. Artificial Intelligence Review, 53(4), 3007-3057.
- [5]. Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press.
- [6]. Chong, E., Han, C., & Park, F. C. (2017). Deep Learning Networks for Stock Market Analysis. Expert Systems with Applications, 83, 187-205.
- [7]. Aggarwal, C. C. (2018). Neural Networks and Deep Learning: A Textbook. Springer.
- [8]. Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. MIT Press.
- [9]. Alpaydin, E. (2021). Machine learning. MIT press.
- [10]. Hansen, K. B. (2020). The virtue of simplicity: On machine learning models in algorithmic trading. Big Data & Society, 7(1). <https://doi.org/10.1177/2053951720926558>
- [11]. Jansen, S. (2020). Machine Learning for Algorithmic Trading: Predictive models to extract signals from market and alternative data for systematic trading strategies with Python. Packt Publishing Ltd.
- [12]. Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. Computer Science Review, 34, 100190.
- [13]. Dongrey, S. (2022). Study of market indicators used for technical analysis. International Journal Of Engineering And Management Research, 12(2), 64-83.
- [14]. Tsay, R. S. (2010). Analysis of Financial Time Series. Wiley.
- [15]. Kumar, M., & Thenmozhi, M. (2014). "Forecasting Stock Index Returns Using ARIMA-SVR Hybrid Models." International Journal of Financial Markets and Derivatives.

- [16]. Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). "Predicting Stock and Stock Price Index Movement Using Machine Learning Techniques." *Expert Systems with Applications*, 42(1), 259-268.
- [17]. Bollinger, J. (2010). *Bollinger on Bollinger Bands*. McGraw-Hill.
- [18]. Ahmad, T., & Aziz, M. N. (2019). Data preprocessing and feature selection for machine learning intrusion detection systems. *ICIC Express Lett*, 13(2), 93-101.
- [19]. Pashankar, S. S., Shendage, J. D., & Pawar, J. (2024). Machine Learning Techniques for Stock Price Prediction-A Comparative Analysis Of Linear Regression, Random Forest, And Support Vector Regression. *Journal of Advanced Zoology*, 45.
- [20]. Rigatti, S. J. (2017). Random forest. *Journal of Insurance Medicine*, 47(1), 31-39.
- [21]. Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In *Machine learning* (pp. 123-140). Academic Press.
- [22]. Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis*. John Wiley & Sons.
- [23]. McKinney, W., & Team, P. D. (2015). *Pandas-Powerful python data analysis toolkit*. Pandas—Powerful Python Data Analysis Toolkit, 1625.
- [24]. Hao, J., & Ho, T. K. (2019). Machine learning made easy: a review of scikit-learn package in python programming language. *Journal of Educational and Behavioral Statistics*, 44(3), 348-361.
- [25]. Yim, A., Chung, C., & Yu, A. (2018). *Matplotlib for Python Developers: Effective techniques for data visualization with Python*. Packt Publishing Ltd.
- [26]. Derfoufi, Y. (2021). *Interface graphique Python Tkinter*.
- [27]. Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, 2022, 1-10.
- [28]. Μιχαήλ Παπανικόλας (2023) Πρόβλεψη Σιμών Μετοχών με τη Χρήση Μηχανικής Μάθησης
- [29]. Mehtab, S., Sen, J., & Dutta, A. (2021). Stock price prediction using machine learning and LSTM-based deep learning models. In *Machine Learning and Metaheuristics Algorithms, and Applications: Second Symposium, SoMMA 2020, Chennai, India, October 14–17, 2020, Revised Selected Papers 2* (pp. 88-106). Springer Singapore.

Παράρτημα Α: <Κώδικας>

```

import os
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from sklearn.ensemble import RandomForestRegressor
from sklearn.svm import SVR
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error
import tkinter as tk
from tkinter import ttk

# Directory containing qqq ETF file
qqq_directory = r'C:\Users\Kamatsos\Desktop\FolderThesis\ETFs'

# Stock directory containing other stocks
stock_directory = r'C:\Users\Kamatsos\Desktop\FolderThesis\Stocks'

one_month_window = 21
two_month_window = 42
three_month_window = 63

# Predefined list of NASDAQ ETF stock symbols
nasdaq_etf_symbols = [
    'AAPL', 'MSFT', 'AMZN', 'FB', 'GOOGL', 'TSLA', 'NVDA', 'PYPL', 'ADBE', 'NFLX',
    'INTC', 'CSCO', 'PEP', 'AVGO', 'COST', 'QCOM', 'TXN', 'TMUS', 'AMD', 'ISRG'
]

# Function to load the qqq ETF file for training
def load_qqq_etf_data():
    qqq_file_path = os.path.join(qqq_directory, 'qqq.us.txt')
    return pd.read_csv(qqq_file_path)

# Function to load the selected stock file into a dataframe
def load_stock_data(directory, stock_file):
    file_path = os.path.join(directory, stock_file)
    return pd.read_csv(file_path)

# Function to list stock files and filter for NASDAQ ETF stocks
def list_stock_files(directory, nasdaq_symbols):
    stock_files = [f for f in os.listdir(directory) if f.endswith('.us.txt')]
    stock_symbols = [f.split('.')[0].upper() for f in stock_files] # Extract stock symbol from file name
    filtered_files = [(f, s) for f, s in zip(stock_files, stock_symbols) if s in nasdaq_symbols] # Filter by
    NASDAQ symbols
    return [f[0] for f in filtered_files], [f[1] for f in filtered_files]

# RSI computation function
def compute_rsi(series, period=14):
    delta = series.diff()
    gain = (delta.where(delta > 0, 0)).rolling(window=period).mean()

```

```

loss = (-delta.where(delta < 0, 0)).rolling(window=period).mean()
rs = gain / loss
rsi = 100 - (100 / (1 + rs))
return rsi

```

Preprocessing function for stock data

```

def preprocess(dataframe, stock_name):
    dataframe["stock_name"] = stock_name # Add stock name to the dataframe
    dataframe["AVG"] = (dataframe["High"] + dataframe["Low"]) / 2
    dataframe["Date"] = pd.to_datetime(dataframe["Date"])
    dataframe["Year"] = dataframe["Date"].apply(lambda x: x.year)
    dataframe["Month"] = dataframe["Date"].apply(lambda x: x.month)
    dataframe.drop("OpenInt", axis=1, inplace=True)
    dataframe["ema_onemonth"] = dataframe["AVG"].ewm(span=21, adjust=False).mean()
    dataframe["rsi_onemonth"] = compute_rsi(dataframe["AVG"])
    dataframe["1_Month_Rolling"] = dataframe["AVG"].rolling(window=one_month_window).mean()
    dataframe["2_Month_Rolling"] = dataframe["AVG"].rolling(window=two_month_window).mean()
    dataframe["3_Month_Rolling"] = dataframe["AVG"].rolling(window=three_month_window).mean()
    dataframe["1_Month_Rolling"].fillna(dataframe["AVG"], inplace=True)
    dataframe["2_Month_Rolling"].fillna(dataframe["AVG"], inplace=True)
    dataframe["3_Month_Rolling"].fillna(dataframe["AVG"], inplace=True)
    dataframe["Volatility"] = dataframe["Close"].rolling(window=30).std()
    dataframe["Future_Close"] = dataframe["Close"].shift(-30)
    dataframe.dropna(subset=["Future_Close"], inplace=True)
    dataframe["Volatility"] = dataframe["Volatility"].fillna(dataframe["Volatility"].mean())
    dataframe["rsi_onemonth"] = dataframe["rsi_onemonth"].fillna(dataframe["rsi_onemonth"].mean())
    return dataframe

```

Function to train models on QQQ ETF data and return trained models

```

def train_models_on_qqq():
    # Load QQQ ETF data
    qqq_data = load_qqq_etf_data()
    qqq_data = preprocess(qqq_data, "qqq")

    # Prepare features and target variable
    X = qqq_data.drop(["Future_Close", "Close", "Date", "stock_name"], axis=1)
    Y = qqq_data["Future_Close"]

    scaler = MinMaxScaler()
    X_scaled = scaler.fit_transform(X)

    # Train the models using QQQ data
    RandomForest = RandomForestRegressor(n_estimators=100, random_state=42)
    RandomForest.fit(X_scaled, Y)

    Support = SVR()
    Support.fit(X_scaled, Y)

    Linear = LinearRegression()
    Linear.fit(X_scaled, Y)

    return (RandomForest, Support, Linear), scaler # Return models as a tuple with the scaler

```

```

# Function to make predictions on the selected stock using trained models
def predict_on_selected_stock(models, scaler, selected_stock_symbol, selected_stock_file):
    RandomForest, Support, Linear = models

    # Load the selected stock data
    stock_data = load_stock_data(stock_directory, selected_stock_file)

    # Preprocess the stock data
    stock_data = preprocess(stock_data, selected_stock_symbol)

    X = stock_data.drop(["Close", "Future_Close", "Date", "stock_name"], axis=1)

    # Scale the input data
    X_scaled = scaler.transform(X)

    # Predict using the trained models
    y_pred_rf = RandomForest.predict(X_scaled)
    y_pred_svr = Support.predict(X_scaled)
    y_pred_linear = Linear.predict(X_scaled)

    # Calculate performance metrics
    mae_rf = mean_absolute_error(stock_data['Future_Close'], y_pred_rf)
    mse_rf = mean_squared_error(stock_data['Future_Close'], y_pred_rf)

    mae_svr = mean_absolute_error(stock_data['Future_Close'], y_pred_svr)
    mse_svr = mean_squared_error(stock_data['Future_Close'], y_pred_svr)

    mae_linear = mean_absolute_error(stock_data['Future_Close'], y_pred_linear)
    mse_linear = mean_squared_error(stock_data['Future_Close'], y_pred_linear)

    return stock_data["Date"], stock_data["Close"], y_pred_rf, y_pred_svr, y_pred_linear, mae_rf,
    mse_rf, mae_svr, mse_svr, mae_linear, mse_linear

# Plotting predictions vs actual prices
def plot_predictions(dates, actual, predictions, model_name, stock_name, shift_days=30):
    # Convert predictions (numpy array) to pandas Series with the same index as the dates
    predictions_series = pd.Series(predictions, index=dates.index)

    # Shift predictions by 30 days
    shifted_predictions = shift_predictions(predictions_series, shift_days)

    # Also set last 30 days of the actual prices to NaN for plotting
    actual_shifted = actual.copy()
    actual_shifted.iloc[-shift_days:] = np.nan

    plt.figure(figsize=(10, 6))
    plt.plot(dates, actual_shifted, label="Actual Prices", color='b')
    plt.plot(dates, shifted_predictions, label=f"Predicted Prices ({model_name})", color='r', linestyle='--')
    plt.title(f"{stock_name} Stock Price Prediction ({model_name}) vs Actual Prices")
    plt.xlabel("Date")
    plt.ylabel("Price")
    plt.legend()

```



```

plt.grid(True)
plt.show()

# Function to shift predictions by 30 days and generate NaNs for plotting
def shift_predictions(predictions, shift_days=30):
    shifted_preds = pd.Series(np.nan, index=predictions.index)
    shifted_preds.iloc[shift_days:] = predictions[:-shift_days].values
    return shifted_preds

# Function to run analysis for the selected stock
def run_analysis(models, scaler, selected_stock_symbol, selected_stock_file, mae_mse_labels):
    dates, actual, y_pred_rf, y_pred_svr, y_pred_linear, mae_rf, mse_rf, mae_svr, mse_svr, mae_linear,
    mse_linear = predict_on_selected_stock(
        models, scaler, selected_stock_symbol, selected_stock_file
    )

    # Update the UI with MAE and MSE for each model
    mae_mse_labels['rf'].config(text=f"RandomForest MAE: {mae_rf:.4f}, MSE: {mse_rf:.4f}")
    mae_mse_labels['svr'].config(text=f"SVR MAE: {mae_svr:.4f}, MSE: {mse_svr:.4f}")
    mae_mse_labels['linear'].config(text=f"LinearRegression MAE: {mae_linear:.4f}, MSE:
{mse_linear:.4f}")

    # Plot predictions for each model
    plot_predictions(dates, actual, y_pred_rf, "RandomForest", selected_stock_symbol)
    plot_predictions(dates, actual, y_pred_svr, "SVR", selected_stock_symbol)
    plot_predictions(dates, actual, y_pred_linear, "Linear Regression", selected_stock_symbol)

# Create the tkinter UI for stock selection
def create_ui():
    # Create the main window
    root = tk.Tk()
    root.title("NASDAQ ETF Stock Selector")
    root.geometry("500x450") # Set window size

    # Style customization
    style = ttk.Style()
    style.configure('TButton', font=('Helvetica', 12), padding=10)
    style.configure('TLabel', font=('Helvetica', 14), padding=10)
    style.configure('TCombobox', font=('Helvetica', 12), padding=5)

    # Create a frame to hold content, center-align elements
    frame = ttk.Frame(root, padding=20)
    frame.pack(expand=True)

    # Create a dropdown menu for stock selection
    label = ttk.Label(frame, text="Select a NASDAQ ETF stock:")
    label.pack(pady=10)

    stock_files, stock_symbols = list_stock_files(stock_directory, nasdaq_etf_symbols)

    # Create a Combobox (dropdown)
    combo = ttk.Combobox(frame, values=stock_symbols, state='readonly')
    combo.pack(pady=10)

```

```

# Labels for MAE and MSE display
mae_mse_labels = {
    "rf": ttk.Label(frame, text="RandomForest MAE: --, MSE: --"),
    "svr": ttk.Label(frame, text="SVR MAE: --, MSE: --"),
    "linear": ttk.Label(frame, text="LinearRegression MAE: --, MSE: --"),
}
mae_mse_labels["rf"].pack(pady=5)
mae_mse_labels["svr"].pack(pady=5)
mae_mse_labels["linear"].pack(pady=5)

# Function to run the analysis based on user selection
def on_select():
    selected_stock_symbol = combo.get()
    if selected_stock_symbol in stock_symbols:
        selected_stock_file = stock_files[stock_symbols.index(selected_stock_symbol)]
        run_analysis(models, scaler, selected_stock_symbol, selected_stock_file, mae_mse_labels)
    else:
        print("Invalid selection. Please select a valid stock.")

# Create a button to start the analysis
button = ttk.Button(frame, text="Run Analysis", command=on_select)
button.pack(pady=10)

root.mainloop()

if __name__ == "__main__":
    # Train models on QQQ ETF data once at the start
    models, scaler = train_models_on_qqq() # Models and scaler are returned correctly
    # Create the UI
    create_ui()

```


Σύντομο Βιογραφικό Συγγραφέα



Ο Προκόπιος Καμάτσος είναι προπτυχιακός φοιτητής του τμήματος Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής της πολυτεχνικής σχολής του Πανεπιστημίου Πατρών με καταγωγή από το Μετόχι Αχαΐας . Το 2019 αποφοίτησε από το Γενικό Λύκειο Κάτω Αχαΐας με βαθμό «Άριστα». Το καλοκαίρι του ίδιου έτους πήρε μέρος στις πανελλαδικές εξετάσεις , όπου εισήχθη στο Πανεπιστήμιο Πατρών και ξεκίνησε την φοιτητική του ζωή. Όντας παιδί που έχει μεγαλώσει σε επαρχία είχε μεγάλη όρεξη να διευρύνει τους

κοινωνικούς του ορίζοντες και να ενταχθεί στους γρήγορους ρυθμούς της πόλης. Κυρίαρχος του στόχος είναι η ολοκλήρωση των σπουδών του με την απόκτηση πτυχίου και κατάλληλου μεταπτυχιακού, ώστε να βγει στην αγορά εργασίας με όσο το δυνατόν περισσότερα προσόντα.