



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

**Δημιουργία Πληροφοριακού Συστήματος για την
Πρόβλεψη Τιμών σε Προϊόντα του Πρωτογενή
Τομέα με χρήση γνώσεων Ανάλυσης Δεδομένων
και Μηχανικής Μάθησης.**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

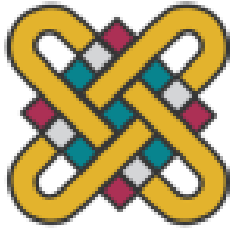
του

ΠΑΠΑΓΕΩΡΓΙΟΥ ΝΙΚΟΛΑΟΣ – ΔΟΜΙΝΙΚΟΣ

(ΑΕΜ: 3109)

Επιβλέπων : Δημόκας Νικόλαος
Επίκουρος Καθηγητής

Καστοριά Οκτώβριος - 2024



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

**Δημιουργία Πληροφοριακού Συστήματος για την
Πρόβλεψη Τιμών σε Προϊόντα του Πρωτογενή
Τομέα με χρήση γνώσεων Ανάλυσης Δεδομένων
και Μηχανικής Μάθησης.**

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

του

ΠΑΠΑΓΕΩΡΓΙΟΥ ΝΙΚΟΛΑΟΣ – ΔΟΜΙΝΙΚΟΣ

(ΑΕΜ: 3109)

**Επιβλέπων : Δημόκας Νικόλαος
Επίκουρος Καθηγητής**

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 03.10.2024

.....
Δημόκας Νικόλαος
Επίκουρος Καθηγητής

.....
Τουλόπουλος Ιωάννης
Επίκουρος Καθηγητής

.....
Σπυρίδων Νικολάου
Λέκτορας

Καστοριά Οκτώβριος - 2024

Copyright © 2024 – ΠΑΠΑΓΕΩΡΓΙΟΥ ΝΙΚΟΛΑΟΣ ΔΟΜΙΝΙΚΟΣ

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

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν αποκλειστικά τον συγγραφέα και δεν αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Δυτικής Μακεδονίας.

Ως συγγραφέας της παρούσας εργασίας δηλώνω πως η παρούσα εργασία δεν αποτελεί προϊόν λογοκλοπής και δεν περιέχει υλικό από μη αναφερόμενες πηγές.

Ευχαριστίες

Περίληψη

Η πρόβλεψη μπορεί να χρησιμοποιηθεί για την εκτίμηση μελλοντικών συμβάντων καθώς και της πιθανότητας εμφάνισης ενός αποτελέσματος. Μπορεί επίσης να χρησιμοποιηθεί για την εκτίμηση μελλοντικών απαιτήσεων ή για την εκτέλεση μίας ανάλυσης των επιπτώσεων διαφόρων σεναρίων. Όσον αφορά τον πρωτογενή τομέα, οι προβλέψεις για την παραγωγή και τις τιμές είναι χρήσιμες για τους παραγωγούς, τις κυβερνήσεις καθώς και τις βιομηχανίες πρωτογενών προϊόντων. Λόγω της ειδικής θέσης της πρωτογενούς παραγωγής στην ασφάλεια ενός έθνους, οι κυβερνήσεις έχουν γίνει και οι κύριοι προμηθευτές και χρήστες των σχετικών προβλέψεων. Χρειάζονται εσωτερικές προβλέψεις για την εφαρμογή πολιτικών που παρέχουν τεχνική υποστήριξη και υποστήριξη της αγοράς στον πρωτογενή τομέα. Η σημασία του πρωτογενούς τομέα για τη διασφάλιση της αναπτυξιακής πορείας, αλλά και την κάλυψη βασικών αναγκών μίας χώρας αποτέλεσε τον κύριο λόγο για τον οποίο επιλέχθηκε ως το αντικείμενο εφαρμογής της παρούσας μελέτης. Η εργασία διαρθρώνεται σε επτά κεφάλαια. Αρχικά, λαμβάνει χώρα μία εισαγωγή στο προς διερεύνηση ζήτημα, μέσα από την παράθεση των βασικών εννοιών και των στόχων της παρούσας έρευνας. Στη συνέχεια λαμβάνει χώρα μία βιβλιογραφική ανασκόπηση, στην οποία περιγράφεται μία σειρά συστημάτων που έχουν σχεδιαστεί για την πραγματοποίηση επιχειρηματικών προβλέψεων όπως επίσης γίνεται και παρουσιάζει μία σειράς αλγορίθμων μηχανικής μάθησης οι οποίοι χρησιμοποιούνται για την πραγματοποίηση προβλέψεων. Στο τρίτο κεφάλαιο γίνεται παρουσίαση και ανάλυση των δεδομένων που χρησιμοποιήθηκαν, εξηγείτε ο λόγος που επιλέχθηκαν καθώς επίσης τα προβλήματα που διαπιστώθηκαν κατά την εύρεση τους. Στο τέταρτο κεφάλαιο γίνεται αναφορά και επεξήγηση του αλγορίθμου πρόβλεψης που επιλέχθηκε και υλοποιήθηκε στα πλαίσια της παρούσας εργασίας. Στο πέμπτο κεφάλαιο αξιολογείται ο αλγόριθμος μέσα από την περιγραφή των αποτελεσμάτων της εκτέλεσης του. Στο έκτο κεφάλαιο περιγράφεται το πληροφοριακό σύστημα που αναπτύχθηκε μέσα από ένα αντίστοιχο εγχειρίδιο. Τέλος, στο έβδομο κεφάλαιο παρατίθενται τα συμπεράσματα που εξήχθησαν από την παρούσα εργασία, ενώ λαμβάνουν χώρα και προτάσεις αναφορικά με πιθανές μελλοντικές επεκτάσεις του συστήματος.

Λέξεις Κλειδιά: Πρόβλεψη, Ανάλυση, Πρωτογενής τομέας, Μηχανική μάθηση, Αλγόριθμοι πρόβλεψης, Ανάλυση δεδομένων, Συστήματα επιχειρηματικής πρόβλεψης, Προβλέψεις τιμών.

Abstract

Forecasting can be used to estimate future events as well as the likelihood of an outcome occurring. It can also be used to assess future demands or to perform an impact analysis of various scenarios. In terms of the primary sector, forecasts regarding production and prices are useful for producers, governments, and primary product industries. Due to the unique position of primary production in a nation's security, governments have become the main providers and users of relevant forecasts. They require internal forecasts to implement policies that provide technical and market support to the primary sector. The importance of the primary sector for ensuring developmental progress and meeting a country's basic needs was the main reason it was chosen as the subject of application for this study. The thesis is structured into seven chapters. Initially, an introduction to the issue under investigation is provided by presenting the key concepts and objectives of the current research. Subsequently, a literature review is conducted, describing a series of systems designed for business forecasting and presenting a range of machine learning algorithms used for forecasting. The third chapter presents and analyzes the data used, explaining the reasons for their selection and the issues identified during their collection. The fourth chapter discusses and explains the forecasting algorithm chosen and implemented in the context of this thesis. The fifth chapter evaluates the algorithm by describing the results of its execution. The sixth chapter describes the information system developed through a corresponding manual. Finally, the seventh chapter presents the conclusions drawn from this study and offers suggestions regarding potential future extensions of the system.

Key Words: Forecasting, Analysis, Primary Sector, Machine Learning, Forecasting Algorithm, Data Analysis, Business Forecasting System, Price Forecasting.

Πίνακας Περιεχομένων

1. Εισαγωγή	2
1.1	
21.2	
21.3	
51.462.	Βιβλιογραφική
ανασκόπηση	9
2.1	82.
2	102.2.
1	102.2.
2	112.2.
3	122.2.
4	152.
3	152.3.
1	162.3.
2	172.3.
3	182.
4	212.4.
1	242.4.
2	242.
5	283. Ανάλυση
Δεδομένων	313.
1	323.
2	333.
3	333.3.
1	343.3.
2	363.3.
3	384. Αλγόριθμοι
Πρόβλεψης	414.
1	414.1.
1	414.1.
2	414.
2	414.2.
1	424.2.
2	424.2.
3	424.2.
4	434.2.
5	434.2.
6	444.2.
7	454.2.
8	454.
3	464.3.
1	464.3.

2	474.3.
3	474.3.
4	494.3.
5	494.3.
6	494.3.
7	504.
4	514.4.
1	514.4.
2	524.4.
3	524.4.
4	534.4.
5	534.4.
6	544.4.
7	545. Αξιολόγηση
Αποτελεσμάτων	565.
1	565.
2	565.
3	575.
4	575.
5	576. Πληροφοριακό
Σύστημα	586.
1	596.
2	596.2.
1	596.2.
2	616.2.
3	626.2.
4	637. Συμπεράσματα και Μελλοντικές
Επεκτάσεις	657.
1	657.
2	66Αναφορέ
ς	68Παράρτημα
Κώδικα	70

Λίστα Εικόνων

Εικόνα 1. Χρονοσειρά: Τυχαία δεδομένα συν τάση, με την ευθεία που αντιστοιχεί καλύτερα στα δεδομένα και διαφορετικά εφαρμοσμένα φίλτρα. Πηγή: “ https://en.wikipedia.org/wiki/Time_series ”	7
Εικόνα 2. (α) Σχηματικό διάγραμμα της παραδοσιακής διαδικασίας συνδυαστικών μοντέλων πρόβλεψης και (β) ένα παράδειγμα εφαρμογής ενός παραδοσιακού συνδυαστικού μοντέλου πρόβλεψης Πηγή: “ https://www.mdpi.com/2077-0472/13/9/1671 ”	25
Εικόνα 3. Σχηματικό διάγραμμα μοντέλων πρόβλεψης ανάλυσης – ολοκλήρωσης Πηγή: “ https://www.mdpi.com/2077-0472/13/9/1671 ”	27
Εικόνα 4. Διάγραμμα Μήτρας για Μήλα	36
Εικόνα 5. Διάγραμμα Μήτρας για Αχλαδιά	37
Εικόνα 6. Μέση τιμή Μήλων ανά μήνα για κάθε χρόνο	38
Εικόνα 7. Μέση τιμή Μήλων ανά μήνα (Όλα τα χρόνια μαζί)	38
Εικόνα 8. Μέση ποσότητα παραγωγής Μήλων ανά μήνα για κάθε χρόνο	39
Εικόνα 9. Μέση ποσότητα παραγωγής Μήλων ανά μήνα (Όλα τα χρόνια μαζί)	39
Εικόνα 10. Μέση τιμή Αχλαδιών ανά μήνα για κάθε χρόνο	40
Εικόνα 11. Μέση τιμή Αχλαδιών ανά μήνα (Όλα τα χρόνια μαζί)	40
Εικόνα 12. Μέση ποσότητα παραγωγής Αχλαδιών ανά μήνα για κάθε χρόνο	41
Εικόνα 13. Μέση ποσότητα παραγωγής Αχλαδιών ανά μήνα (Όλα τα χρόνια μαζί)	41
Εικόνα 14. LSTM_1	43
Εικόνα 15. LSTM_2	43
Εικόνα 16. LSTM_3	44
Εικόνα 17. LSTM_4	44
Εικόνα 18. LSTM_5	44
Εικόνα 19. LSTM_6	45
Εικόνα 20. LSTM_7	45
Εικόνα 21. LSTM_8	46
Εικόνα 22. LSTM_9	46
Εικόνα 23. LSTM_Prediction_Results	47

Εικόνα 24. SARIMA_1	48
Εικόνα 25. SARIMA_2	48
Εικόνα 26. SARIMA_3	49
Εικόνα 27. SARIMA_4	50
Εικόνα 28. SARIMA_5	50
Εικόνα 29. SARIMA_6	51
Εικόνα 30. SARIMA_Prediction_Results	51
Εικόνα 31. PROPHET_1	52
Εικόνα 32. PROPHET_2	53
Εικόνα 33. PROPHET_3	54
Εικόνα 34. PROPHET_4	54
Εικόνα 35. PROPHET_5	55
Εικόνα 36. PROPHET_6	55
Εικόνα 37. PROPHET_Prediction_Results	56
Εικόνα 38. Ενότητα: Αρχική Σελίδα	61
Εικόνα 39. Ενότητα: Σχετικά	62
Εικόνα 40. Ενότητα: Μήλα	63
Εικόνα 41. Ενότητα: Αχλάδια	64

Λίστα Πινάκων

Πίνακας 1. Κοινές μέθοδοι ανάλυσης χρονοσειρών	14
Πίνακας 2. Παραδείγματα εφαρμογής της πρόβλεψης τιμών με τη χρήση νευρωνικών δικτύων για αγροτοκτηνοτροφικά προϊόντα	21
Πίνακας 3. Παραδείγματα εφαρμογών συνδυασμών μεθόδων	23
Πίνακας 4. Μέθοδοι ανάλυσης χρονοσειρών για πρόβλεψη τιμών αγροτικών προϊόντων.	26
Πίνακας 5. Σύνοψη μεθόδων βελτιστοποίησης και χαρακτηριστικών τους για ανάλυση και πρόβλεψη δεδομένων	30
Πίνακας 6. Παράδειγμα μορφής δεδομένων για τις τιμές μήλων, υγραερίου κίνησης και της ποσότητας παραγωγής από το 2007 έως το 2022.	33
Πίνακας 7. Παράδειγμα μορφής δεδομένων για τις τιμές αχλαδιών, υγραερίου κίνησης και της ποσότητας παραγωγής από το 2007 έως το 2022.	33
Πίνακας 8. Αποτελέσματα προβλέψεων για τους επόμενους 12 μήνες με χρήση LSTM	47
Πίνακας 9. Αποτελέσματα προβλέψεων για τους επόμενους 12 μήνες με χρήση SARIMA	51
Πίνακας 10. Αποτελέσματα προβλέψεων για τους επόμενους 12 μήνες με χρήση PROPHET	55

1. Εισαγωγή

1.1 Πρόβλεψη

Η πρόβλεψη είναι η διαδικασία των εκτιμήσεων που βασίζονται σε δεδομένα του παρελθόντος και του παρόντος. Τα αποτελέσματα των προβλέψεων σε μεταγενέστερο χρόνο μπορούν να συγκριθούν (να επιλυθούν) με την κατάσταση που συμβαίνει στην πραγματικότητα. Για παράδειγμα, μία εταιρεία μπορεί να εκτιμήσει τα έσοδά της το επόμενο έτος και στη συνέχεια να τα συγκρίνει με τα πραγματικά αποτελέσματα δημιουργώντας μία πραγματική ανάλυση διακύμανσης. Η πρόβλεψη μπορεί να αναφέρεται επίσης σε συγκεκριμένες επίσημες στατιστικές μεθόδους που χρησιμοποιούν χρονοσειρές, συγχρονικά ή διαχρονικά δεδομένα ή εναλλακτικά σε λιγότερο επίσημες μεθόδους κρίσης ή την ίδια τη διαδικασία πρόβλεψης και επίλυσης.

Ο κίνδυνος και η αβεβαιότητα παίζουν καθοριστικό ρόλο στις προβλέψεις. Γενικά, είναι καλή πρακτική να αναφέρεται το επίπεδο της αβεβαιότητας που συνοδεύει τις προβλέψεις. Σε κάθε περίπτωση, τα δεδομένα θα πρέπει να είναι ενημερωμένα ώστε να εξασφαλίζεται η μεγαλύτερη δυνατή ακρίβεια. Σε ορισμένες περιπτώσεις, τα δεδομένα που χρησιμοποιούνται για την πρόβλεψη της υπό εξέταση μεταβλητής μπορεί να αποτελούν τα ίδια αντικείμενο πρόβλεψης (French, 2017). Μια πρόβλεψη δεν πρέπει να συγχέεται με έναν προϋπολογισμό. Οι προϋπολογισμοί είναι πιο συγκεκριμένα, χρονικά σταθερά οικονομικά σχέδια που χρησιμοποιούνται για την κατανομή πόρων και τον έλεγχο, ενώ οι προβλέψεις προσφέρουν εκτιμήσεις για τη μελλοντική οικονομική πορεία, δίνοντας τη δυνατότητα ευελιξίας και προσαρμογής σε μεταβαλλόμενες συνθήκες. Και τα δύο αυτά εργαλεία είναι σημαντικά για τον οικονομικό σχεδιασμό και τη λήψη αποφάσεων, αλλά έχουν διαφορετικούς ρόλους.

1.2 Η πληροφορία στην κοινωνία

Η πληροφορία είναι μια αφηρημένη έννοια που σχετίζεται με οτιδήποτε μπορεί να προσφέρει ενημέρωση. Στο πιο βασικό της επίπεδο, αφορά την ερμηνεία (ενδεχομένως με τυπικό τρόπο) όσων μπορούν να γίνουν αντιληπτά ή τις αντίστοιχες αφαιρέσεις αυτών. Οποιαδήποτε φυσική διαδικασία που δεν είναι απολύτως τυχαία και κάθε παρατηρήσιμο μοτίβο σε οποιοδήποτε περιβάλλον μπορεί να θεωρηθεί ότι μεταφέρει μια ποσότητα πληροφοριών. Ενώ τα ψηφιακά και άλλα δεδομένα χρησιμοποιούν διακριτά σήματα για τη μετάδοση πληροφοριών, άλλες μορφές, όπως αναλογικά σήματα, ποιητικά κείμενα, εικόνες, μουσικά κομμάτια ή ήχοι, και ρεύματα, μεταφέρουν πληροφορίες σε μια πιο συνεχόμενη μορφή (Anderson & Johnsson,

1996). Η πληροφορία δεν ισοδυναμεί με τη γνώση, αλλά αντιπροσωπεύει το νόημα που προκύπτει από την αναπαράστασή της μέσω της ερμηνείας (Yockey, 2005).

Η έννοια της πληροφορίας είναι σχετική ή συνδέεται με διάφορες έννοιες (Floridi, 2010), όπως περιορισμός, επικοινωνία, έλεγχος, δεδομένα, μορφή, εκπαίδευση, γνώση, νόημα, κατανόηση, νοητικά ερεθίσματα, μοτίβο, αντίληψη, πρόταση, αναπαράσταση και εντροπία.

Οι πληροφορίες συχνά υφίστανται διαδοχική επεξεργασία: Τα δεδομένα που είναι διαθέσιμα αρχικά μετατρέπονται σε πληροφορίες, οι οποίες στη συνέχεια υπόκεινται σε περαιτέρω ανάλυση σε επόμενα στάδια. Για παράδειγμα, σε ένα γραπτό κείμενο, κάθε σύμβολο ή γράμμα περιέχει πληροφορίες σχετικές με τη λέξη στην οποία ανήκει, κάθε λέξη μεταφέρει πληροφορίες για τη φράση που αποτελεί μέρος της, κάθε φράση περιέχει πληροφορίες για την πρόταση στην οποία ανήκει, και αυτό συνεχίζεται μέχρις ότου οι πληροφορίες να ερμηνευθούν πλήρως και να μετατραπούν σε γνώση σε ένα συγκεκριμένο πλαίσιο. Σε ένα ψηφιακό σήμα, τα bits μπορούν να ερμηνευτούν ως σύμβολα, γράμματα, αριθμοί ή δομές, οι οποίες μεταφέρουν τις σχετικές πληροφορίες στο επόμενο επίπεδο. Το ουσιώδες χαρακτηριστικό της πληροφορίας είναι ότι χρειάζεται ερμηνεία και ανάλυση.

Η λήψη πληροφοριών από ένα σήμα ή μήνυμα μπορεί να θεωρηθεί ως η διαδικασία επίλυσης της ασάφειας ή της αβεβαιότητας που εμφανίζεται κατά την ερμηνεία των μοτίβων του σήματος ή του μηνύματος (Webler, 2022). Οι πληροφορίες μπορούν να οργανωθούν σε μορφή δεδομένων. Τα περιττά δεδομένα μπορούν να συμπιεστούν στο μικρότερο δυνατό μέγεθος, το οποίο αποτελεί το θεωρητικό όριο της συμπίεσης. Οι πληροφορίες που προκύπτουν από μια συλλογή δεδομένων μπορούν να αναλυθούν. Για παράδειγμα, ένα εστιατόριο μπορεί να συγκεντρώσει δεδομένα από τις παραγγελίες των πελατών και να τα αναλύσει για να αποκτήσει γνώση, όπως ο προσδιορισμός των πιο δημοφιλών ή λιγότερο δημοφιλών πιάτων.

Οι πληροφορίες μπορούν να μεταδοθούν διαχρονικά μέσω της αποθήκευσης δεδομένων και να διανεμηθούν στον χώρο μέσω των επικοινωνιών και των τηλεπικοινωνιών. Εκφράζονται είτε ως το περιεχόμενο ενός μηνύματος είτε μέσω άμεσης ή έμμεσης παρατήρησης. Αυτό που αντιλαμβανόμαστε μπορεί να θεωρηθεί ως ένα μήνυμα από μόνο του, και έτσι, όλες οι πληροφορίες μεταφέρονται πάντα ως περιεχόμενο ενός μηνύματος.

Οι πληροφορίες είναι δυνατόν να κωδικοποιηθούν σε διάφορες μορφές για να μεταδοθούν και να ερμηνευτούν, όπως μια σειρά συμβόλων ή ένα σήμα. Μπορούν επίσης να κρυπτογραφηθούν για να εξασφαλιστεί η ασφαλής αποθήκευση και επικοινωνία.

Η αβεβαιότητα ενός γεγονότος μετρείται μέσω της πιθανότητας της εμφάνισής του και είναι αντιστρόφως ανάλογη με αυτήν την πιθανότητα. Η θεωρία της πληροφορίας αξιοποιεί αυτή την ιδιότητα, καταλήγοντας στο ότι γεγονότα με μεγαλύτερη αβεβαιότητα χρειάζονται περισσότερες πληροφορίες για να επιλυθεί αυτή η αβεβαιότητα. Το bit είναι η βασική μονάδα μέτρησης των πληροφοριών, και ορίζεται ως "η ποσότητα που μειώνει την αβεβαιότητα στο μισό". Άλλες μονάδες όπως το nat μπορούν επίσης να χρησιμοποιηθούν. Για παράδειγμα, η πληροφορία από μια δίκαιη ρίψη νομίσματος είναι $\log_2(2/1) = 1$ bit, ενώ από δύο δίκαιες ρίψεις νομίσματος είναι $\log_2(4/1) = 2$ bit.

Σε ένα άρθρο του περιοδικού Science το 2011, εκτιμήθηκε ότι το 97% των τεχνολογικά αποθηκευμένων πληροφοριών είχε ήδη μετατραπεί σε ψηφιακά bit μέχρι το 2007, και ότι το 2002 σηματοδότησε την αρχή της ψηφιακής εποχής για την αποθήκευση πληροφοριών, καθώς η ψηφιακή αποθήκευση ξεπέρασε για πρώτη φορά την αναλογική (Hilbert & Lopez, 2011).

Στη σύγχρονη παγκοσμιοποιημένη οικονομία, η τεχνολογία της πληροφορίας είναι ο σημαντικότερος μοχλός τόσο της οικονομικής ανάπτυξης όσο και της βελτίωσης της ποιότητας ζωής. Το Αμερικανικό Ίδρυμα Πληροφορικής και Καινοτομίας (ITIF) στην έκθεσή του Digital Prosperity: Understanding the Economic Benefits of the Information Technology Revolution κατέγραψε τον τρόπο με τον οποίο η πληροφορική, από τα μέσα της δεκαετίας του 1990, ήταν ο κύριος μοχλός της αύξησης της οικονομικής ανάπτυξης όχι μόνο στις Ηνωμένες Πολιτείες αλλά και σε πολλά άλλα έθνη. Οι Atkinson & Castro (2008) έδειξαν ότι η πληροφορική τοποθετείται στον πυρήνα των δραματικών βελτιώσεων στην ποιότητα ζωής των ατόμων σε όλο τον κόσμο: Η πληροφορική είναι ο βασικός παράγοντας για πολλές, αν όχι τις περισσότερες, από τις βασικές καινοτομίες και βελτιώσεις του σήμερα τόσο στη ζωή όσο και στην κοινωνία — από τη βελτιωμένη εκπαίδευση και υγειονομική φροντίδα, σε ένα καθαρότερο και περισσότερο ενεργειακά αποδοτικό περιβάλλον - σε ασφαλέστερες κοινότητες και έθνη.

Οι ανθρωπίνι φραγμοί στην υιοθέτηση κρίσιμων τεχνολογιών παραμένουν μία πρόκληση για τις επιχειρήσεις που αναζητούν την επίτευξη ενός ανταγωνιστικού πλεονεκτήματος μέσω της εφαρμογής συστημάτων στρατηγικών πληροφοριών. Για να αποκτήσουν ανταγωνιστικό πλεονέκτημα σε έναν κλάδο, οι ενδιαφερόμενοι πρέπει να μειώσουν το πλεονέκτημα των πελατών και προμηθευτών μέσω της εφαρμογής τεχνολογιών πληροφορικής με μοναδικά χαρακτηριστικά λειτουργικότητας για τη διαμόρφωση της ανταγωνιστικής στρατηγικής ενός οργανισμού. Οι αλλαγές στις διαδικασίες που επιτρέπονται από τις καινοτόμες τεχνολογίες προκαλούν θεμελιώδεις αλλαγές στον τρόπο με τον οποίο μία βιομηχανία πραγματοποιεί τις λειτουργίες της (Lippert & Forman, 2005).

1.3 Ο ρόλος της πρόβλεψης στην γεωργία και το όφελος που προκύπτει.

Οι οικονομικές προβλέψεις στη γεωργία έχουν ορισμένα κοινά χαρακτηριστικά με την επιχειρηματική πρόβλεψη και τις μακροοικονομικές προβλέψεις. Αλλά κατά τη διάρκεια του χρόνου, έχουν αναπτύξει μία δική τους προοπτική. Το ζήτημα της πραγματοποίησης συστάσεων προς τους αγρότες και διαχειριστές των αγροκτημάτων για την αύξηση των κερδών εμφανίζεται κατά το πρώτο τέταρτο του 20^{ου} αιώνα. Κατά το δεύτερο τέταρτο του αιώνα, το αγροτικό επάγγελμα πραγματοποίησε μία σημαντική στροφή προς την πρόβλεψη, ευρέως καθορισμένη, συμπεριλαμβανομένης της χρήσης οικονομετρικών τεχνικών για την εκτίμηση των ελαστικότητας και την πρόβλεψη των τιμών. Κατά το τρίτο τέταρτο του 20^{ου} αιώνα, από το 1975 και έπειτα, έλαβε εξέχουσα σημασία η έρευνα για την πολιτική, το εμπόριο και την παγκόσμια οικονομία και η επέκταση σε περιβαλλοντικά προβλήματα και προβλήματα διαθεσιμότητας πόρων. Καθ' όλη τη διάρκεια της περιόδου, και πιο έντονα κατά τα τελευταία χρόνια, η εξήγηση της προηγούμενης συμπεριφοράς ήταν η κύρια εστίαση της μοντελοποίησης της γεωργικής προσφοράς, η οποία αποτελεί το αντικείμενο σχετικά με το οποίο λαμβάνουν χώρα οι περισσότερες γεωργικές προβλέψεις (Allen, 1994).

Επειδή η εξασφαλισμένη προμήθεια τροφίμων είναι σημαντική για την εθνική ασφάλεια, οι κυβερνήσεις έχουν προσπαθήσει να ποσοτικοποιήσουν τη γεωργική παραγωγή και να ασκήσουν κάποιον έλεγχο πάνω της. Στην αρχή, απλά η συλλογή και καταχώρηση σε πίνακες δεδομένων για την τρέχουσα γεωργική κατάσταση ήταν μία σημαντική πρόκληση, και οι γεωργικοί στατιστικοί διαδραμάτισαν έναν σημαίνοντα ρόλο στην ανάπτυξη των στατιστικών μεθόδων. Η αναθεώρηση των δεδομένων ήταν συχνή. Οι εκτιμήσεις της παραγωγής, για παράδειγμα, υποβλήθηκαν σε αναθεώρηση, μετά από την καταχώρηση των στοιχείων μιας νέας απογραφής. Ο σημαντικός αριθμός σχετικών εκθέσεων υποδεικνύει το ενδιαφέρον των γεωργικών στατιστικολόγων σχετικά με την εκτίμηση της τρέχουσας κατάστασης μιας σειράς δεδομένων (Allen, 1994).

Οι περισσότεροι γεωργικοί αναλυτές δεδομένων εκπαιδεύτηκαν είτε ως στατιστικοί είτε ως αγροτικοί οικονομολόγοι. Τα δύο επαγγέλματα έχουν διαμορφώσει κάτι το οποίο υπήρξε, κατά καιρούς, μία αντισυμβατική συμμαχία. Οι στατιστικοί ήταν σε μεγάλο βαθμό υπεύθυνοι για την ανάπτυξη της πρόβλεψης προοπτικών βάσει της ανάλυσης δεικτών. Οι αγροτικοί οικονομολόγοι τείνουν να τονίζουν όλο και περισσότερο περίπλοκα οικονομετρικά μοντέλα. Δίδουν ιδιαίτερη έμφαση στην παροχή πειστικών εξηγήσεων για οικονομικά

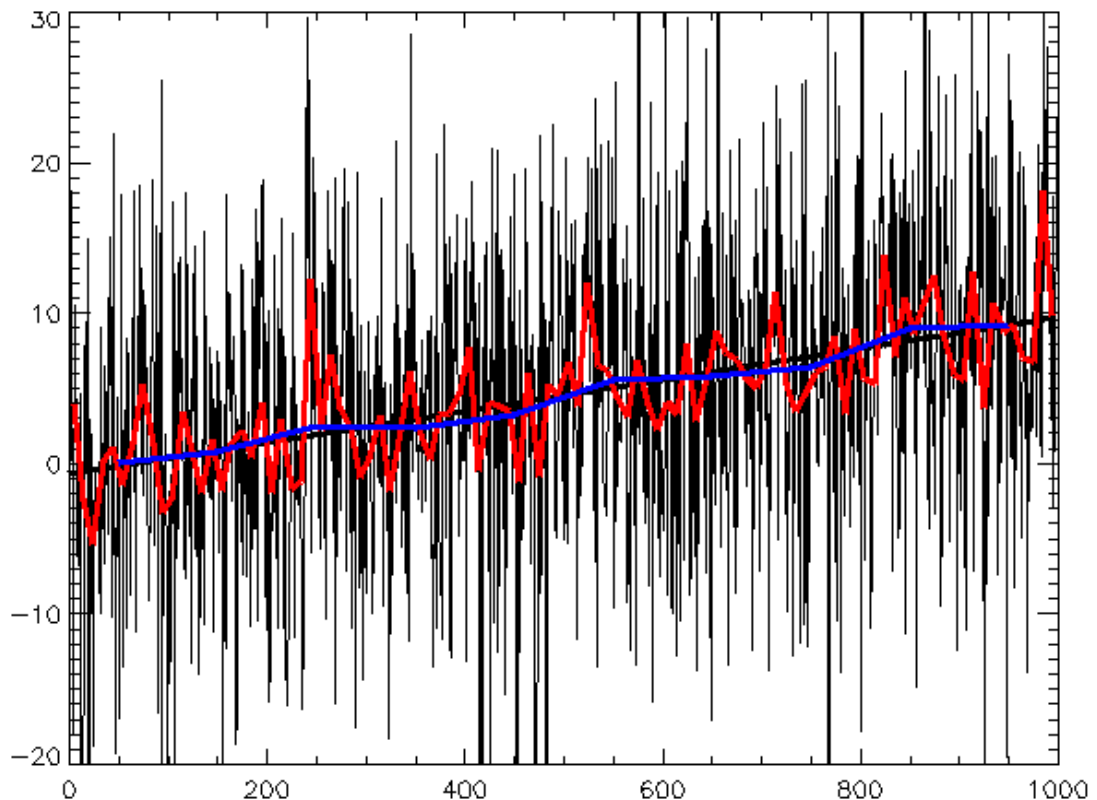
φαινόμενα, με την υπόθεση ότι αυτό δεν θα ήταν χρήσιμο μόνο για τη λήψη αποφάσεων αλλά και για την πρόβλεψη (Allen, 1994).

1.4 Χρονοσειρές σειρές

Στα μαθηματικά, μία χρονοσειρά είναι μία σειρά σημείων δεδομένων που διατάσσονται (ή παρατίθενται ή σχηματίζονται γραφικά) με χρονική σειρά. Συνηθέστερα, μία χρονοσειρά είναι μία ακολουθία που λαμβάνεται σε διαδοχικά ίσα χρονικά σημεία. Έτσι είναι μία ακολουθία δεδομένων διακριτού χρόνου. Παραδείγματα χρονοσειρών είναι τα ύψη της παλίρροιας των ωκεανών, ο αριθμός των ηλιακών κηλίδων και η ημερήσια τιμή κλεισίματος του χρηματιστηριακού δείκτη Dow Jones.

Μία χρονοσειρά σχεδιάζεται πολύ συχνά μέσω ενός γραφήματος εκτέλεσης (το οποίο είναι ένα διάγραμμα χρονικής γραμμής). Οι χρονολογικές σειρές χρησιμοποιούνται στη στατιστική, την επεξεργασία σήματος, την αναγνώριση προτύπων, την οικονομετρία, τα οικονομικά μαθηματικά, την πρόγνωση καιρού, την πρόβλεψη σεισμών, την ηλεκτροεγκεφαλογραφία, τη μηχανική ελέγχου, την αστρονομία, τη μηχανική επικοινωνιών και σε μεγάλο βαθμό σε οποιονδήποτε τομέα της εφαρμοσμένης επιστήμης και μηχανικής που περιλαμβάνει χρονικές μετρήσεις.

Η ανάλυση χρονοσειρών περιλαμβάνει μεθόδους για την ανάλυση δεδομένων χρονοσειρών προκειμένου να εξαχθούν σημαντικά στατιστικά στοιχεία και άλλα χαρακτηριστικά των δεδομένων. Η πρόβλεψη χρονοσειρών είναι η χρήση ενός μοντέλου για την πρόβλεψη μελλοντικών τιμών με βάση τις προηγουμένως παρατηρηθείσες τιμές. Ενώ η ανάλυση παλινδρόμησης χρησιμοποιείται συχνά με τέτοιο τρόπο ώστε να ελέγχονται οι σχέσεις μεταξύ μίας ή περισσότερων διαφορετικών χρονοσειρών, αυτός ο τύπος ανάλυσης δεν ονομάζεται συνήθως «ανάλυση χρονοσειρών», η οποία αναφέρεται ειδικότερα σε σχέσεις μεταξύ διαφορετικών χρονικών σημείων μέσα σε μία ενιαία σειρά.



Εικόνα 1. Χρονοσειρά: Τυχαία δεδομένα συν τάση, με την ευθεία που αντιστοιχεί καλύτερα στα δεδομένα και διαφορετικά εφαρμοσμένα φίλτρα.

Πηγή: “ https://en.wikipedia.org/wiki/Time_series ”

Τα δεδομένα χρονοσειρών έχουν μία φυσική χρονική σειρά. Αυτό κάνει την ανάλυση χρονοσειρών διαφορετική από τις συγχρονικές μελέτες, στις οποίες δεν υπάρχει φυσική σειρά των παρατηρήσεων (π.χ. εξηγώντας τους μισθούς των ατόμων με αναφορά στο αντίστοιχο επίπεδο εκπαίδευσης, όπου τα δεδομένα των ατόμων θα μπορούσαν να εισαχθούν με οποιαδήποτε σειρά). Η ανάλυση χρονοσειρών διαφέρει επίσης από την ανάλυση χωρικών δεδομένων όπου οι παρατηρήσεις σχετίζονται συνήθως με γεωγραφικές τοποθεσίες (π.χ. ένα στοχαστικό μοντέλο για μία χρονοσειρά θα αντικατοπτρίζει γενικά το γεγονός ότι οι παρατηρήσεις κοντά στον χρόνο θα συνδέονται στενότερα από τις παρατηρήσεις που βρίσκονται σε μεγαλύτερη απόσταση). Επιπλέον, τα μοντέλα χρονοσειρών χρησιμοποιούν συχνά τη φυσική μονόδρομη ταξινόμηση του χρόνου, έτσι ώστε οι τιμές για μία δεδομένη περίοδο να εκφράζονται ως προερχόμενες κατά κάποιον τρόπο από προηγούμενες τιμές και όχι από μελλοντικές τιμές (βλ. αναστρεψιμότητα χρόνου).

Η ανάλυση χρονοσειρών μπορεί να εφαρμοστεί σε δεδομένα με πραγματική αξία, συνεχή, διακριτά αριθμητικά δεδομένα ή διακριτά συμβολικά δεδομένα (δηλαδή σε ακολουθίες χαρακτήρων, όπως γράμματα και λέξεις στην αγγλική γλώσσα) (Lin et al., 2003).

2. Βιβλιογραφική ανασκόπηση

2.1 Η σημασία της πρόβλεψης στην γεωργία

Οι πληροφορίες σχετικά με τις τιμές παρέχουν σημαντική εικόνα για τις διακυμάνσεις στην αγορά γεωργικών προϊόντων. Οι συχνές και έντονες μεταβολές των τιμών επηρεάζουν σε μεγάλο βαθμό το βιοτικό επίπεδο των χωρών και την κοινωνική συνοχή. Η πρόβλεψη των τιμών των γεωργικών προϊόντων δεν σχετίζεται μόνο με την οικονομική σταθερότητα μιας συγκεκριμένης χώρας ή περιοχής, αλλά επίσης επηρεάζει την παγκόσμια ισορροπία μεταξύ προσφοράς και ζήτησης τροφίμων. Με τη συνεχή αύξηση του παγκόσμιου πληθυσμού, η επισιτιστική ασφάλεια έχει εξελιχθεί σε ένα από τα κύρια ζητήματα σε παγκόσμιο επίπεδο. Η ακριβής πρόβλεψη των τιμών των γεωργικών προϊόντων μπορεί να βοηθήσει διεθνείς οργανισμούς, κυβερνήσεις και αγροτικές επιχειρήσεις να αντιδράσουν έγκαιρα, εξασφαλίζοντας επαρκή παροχή τροφίμων και διατηρώντας την παγκόσμια επισιτιστική σταθερότητα. Συνεπώς, η μελέτη των μεθόδων πρόβλεψης για τις τιμές των γεωργικών προϊόντων είναι κρίσιμης σημασίας για τη βελτίωση της ασφάλειας των γεωργικών αγαθών, τόσο από την πλευρά της ποσότητας όσο και για την προώθηση της οικονομικής και κοινωνικής ανάπτυξης (Zheng, 2013).

Σε σύγκριση με τις γενικές τιμές των εμπορευμάτων, οι τιμές των γεωργικών προϊόντων επηρεάζονται από πιο πολύπλοκους παράγοντες και εμφανίζουν ακανόνιστες διακυμάνσεις, όπως μη στάσιμες και μη γραμμικές τάσεις. Οι συχνές και απότομες αλλαγές στις τιμές των γεωργικών αγαθών μπορούν να έχουν σημαντικές επιπτώσεις στην εθνική και παγκόσμια επισιτιστική ασφάλεια. Οι ερευνητές έχουν εντοπίσει ότι η προσφορά και η ζήτηση διαδραματίζουν καίριο ρόλο στη διαμόρφωση των γεωργικών τιμών. Η παραγωγή επηρεάζει την προσφορά και τη ζήτηση, με αποτέλεσμα την αστάθεια των τιμών. Επιπλέον, οι τιμές των γεωργικών προϊόντων επηρεάζονται από διάφορους παράγοντες, όπως το κόστος εργασίας, το αυξανόμενο λειτουργικό κόστος και τις συνθήκες της διεθνούς αγοράς. Οι ερευνητές έχουν επίσης μελετήσει τον τρόπο με τον οποίο οι τιμές μεταδίδονται στην αγορά, ανακαλύπτοντας ότι η μετάδοση αυτή παρουσιάζει ασυμμετρίες. Παράλληλα, παράγοντες όπως οι κλιματικές συνθήκες και οι πολιτικές αποφάσεις επηρεάζουν τις τιμές των γεωργικών προϊόντων σε διαφορετικό βαθμό. Οι Gu et al. (2023) πραγματοποίησαν μελέτη για τους παράγοντες που επηρεάζουν τις γεωργικές τιμές, διαπιστώνοντας ότι η θερμοκρασία, οι ώρες ηλιοφάνειας και οι επιδημίες έχουν σημαντικό αντίκτυπο. Τα ευρήματα και οι συνεισφορές των διάφορων μελετητών σχετικά με τις διακυμάνσεις των τιμών και τους παράγοντες που τις επηρεάζουν,

έχουν θέσει μια σταθερή βάση για την επίτευξη ακριβέστερων προβλέψεων των τιμών των γεωργικών προϊόντων.

Η πρόβλεψη των τιμών των γεωργικών προϊόντων περιλαμβάνει την εφαρμογή επιστημονικών μεθόδων για την εκτίμηση της τάσης και του επιπέδου των αλλαγών στις τιμές αυτών των προϊόντων κατά μια μελλοντική χρονική περίοδο, βασιζόμενη σε ιστορικά δεδομένα και τρέχουσες πληροφορίες. Οι μέθοδοι πρόβλεψης διακρίνονται σε ποιοτική και ποσοτική ανάλυση. Η ποιοτική ανάλυση εστιάζει στην πλήρη κατανόηση των πληροφοριών της αγοράς, χρησιμοποιώντας την εμπειρία για να διαμορφώσει μια γενική εκτίμηση της κατεύθυνσης της τάσης των τιμών. Από την άλλη, η ποσοτική ανάλυση συγκεντρώνει τις διαθέσιμες πληροφορίες σχετικά με τις τιμές της αγοράς και χρησιμοποιεί συγκεκριμένες μεθόδους πρόβλεψης για να καθορίσει έναν αριθμητικό υπολογισμό των αλλαγών στις τιμές των προϊόντων. Η ποσοτική ανάλυση αποτελεί την κύρια μέθοδο που εφαρμόζεται στην πρόβλεψη των γεωργικών τιμών σήμερα. Αυτή περιλαμβάνει κυρίως την ανάλυση παλινδρόμησης (γνωστή και ως αιτιώδης ανάλυση), την ανάλυση χρονοσειρών, τις μεθόδους μηχανικής μάθησης, καθώς και τα συνδυασμένα μοντέλα. Επίσης, από την άποψη των μεταβλητών, η πρόβλεψη μπορεί να είναι είτε μονομεταβλητή είτε πολυμεταβλητή, ανάλογα με τον αριθμό των παραγόντων που λαμβάνονται υπόψη.

Οι τιμές των γεωργικών προϊόντων επηρεάζονται από πλήθος παραγόντων, όπως η προσφορά και η ζήτηση, οι κλιματικές αλλαγές, οι πολιτικές παρεμβάσεις, ο ανταγωνισμός στην αγορά, καθώς και το διεθνές εμπόριο. Οι σχέσεις ανάμεσα σε αυτούς τους παράγοντες και τις τιμές συχνά είναι μη γραμμικές, δυναμικές και αβέβαιες, κάτι που καθιστά δύσκολη την ποσοτική περιγραφή τους μέσω απλών μαθηματικών μοντέλων. Παραδοσιακές μέθοδοι ανάλυσης είναι απλές, εύκολες στην κατανόηση και στην εφαρμογή, αλλά συνήθως δεν προσφέρουν ακριβείς προβλέψεις όταν αντιμετωπίζουν μη γραμμικά, μη ομαλά ή υψηλής διάστασης δεδομένα. Επιπλέον, απαιτούν συνήθως πολλές προκαταρκτικές γνώσεις και υποθέσεις. Από την άλλη πλευρά, οι ευφυείς μέθοδοι μπορούν να επεξεργάζονται πολύπλοκα δεδομένα με υψηλή ακρίβεια και δυνατότητα γενίκευσης, αλλά χρειάζονται μεγάλες ποσότητες δεδομένων και υπολογιστικών πόρων και στερούνται ερμηνευτικότητας και σταθερότητας. Ως εκ τούτου, είναι σημαντικό να κατανοούνται τα χαρακτηριστικά κάθε μεθόδου πρόβλεψης και να επιλέγεται ο κατάλληλος αλγόριθμος για τη δημιουργία ενός αποτελεσματικού μοντέλου πρόβλεψης των τιμών των γεωργικών προϊόντων. Η επιλογή της σωστής μεθόδου αποτελεί βασικό ζήτημα για την επίτευξη της πιο αποτελεσματικής πρόβλεψης των γεωργικών τιμών.

2.2 Στατιστικά Μοντέλα Προβλέψεις

2.2.1 Μέθοδος προβλέψεις ανάλυσης παλινδρόμησης

Η ανάλυση παλινδρόμησης επινοήθηκε από τον Galton, έναν διάσημο Βρετανό ανθρωπολόγο και στατιστικολόγο, όταν μελέτησε τη σχέση του πατρικού ύψους στο Ηνωμένο Βασίλειο. Το 1917, ο Moore έκανε το πρώτο βήμα προς τη χρήση ποσοτικών μεθόδων στην πρόβλεψη των γεωργικών τιμών, δημιουργώντας ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης για να προβλέψει την παραγωγή και τις τιμές του βαμβακιού. Η ανάλυση παλινδρόμησης βασίζεται στη δημιουργία ενός μαθηματικού μοντέλου που συνδέει τις τιμές με τους παράγοντες που τις επηρεάζουν. Η μέθοδος ανάλυσης παλινδρόμησης περιλαμβάνει κυρίως το μοντέλο γραμμικής παλινδρόμησης (LR – Linear Regression), το μοντέλο γενικευμένης γραμμικής παλινδρόμησης (GLR – Generalized Linear Regression), το μοντέλο μη γραμμικής παλινδρόμησης (NLR – Non Linear Regression), τις πολλαπλές προσαρμοστικές στροφές παλινδρόμησης (MARS – Multiple Adaptive Regression Splines), το γενικευμένο προσθετικό μοντέλο (GAM – Generalized Additive Model) κλπ. Τα μοντέλα απαιτούν σημαντική ποσότητα δεδομένων για την ακριβή και αξιόπιστη εκτίμηση των παραμέτρων. Υποθέτουν επίσης ότι τα δεδομένα στερούνται σφαλμάτων, ακραίων τιμών και πολυσυγγραμμικότητας (υψηλή συσχέτιση μεταξύ των εξηγηματικών μεταβλητών). Σε περιπτώσεις όπου τα δεδομένα εμφανίζουν μη γραμμικότητα, εποχιακές τάσεις, κυκλικά μοτίβα ή επηρεάζονται από δομικές αλλαγές και εξωτερικούς παράγοντες, τα μοντέλα παλινδρόμησης μπορεί να δυσκολεύονται να προσφέρουν ακριβείς προβλέψεις. Επίσης, τα μοντέλα αυτά μπορεί να υποφέρουν από υπερπροσαρμογή (όπου προσαρμόζονται σε τυχαίο θόρυβο στα δεδομένα) ή υποπροσαρμογή (όπου δεν καταφέρνουν να καταγράψουν την πολυπλοκότητα των δεδομένων), επηρεάζοντας αρνητικά την ικανότητά τους για πρόβλεψη και γενίκευση.

Η χρήση μοντέλων παλινδρόμησης είναι ιδιαίτερα κατάλληλη για προβλήματα πρόβλεψης που αφορούν βραχυπρόθεσμους ή μεσοπρόθεσμους χρονικούς ορίζοντες, εφόσον οι εξηγητικές μεταβλητές είναι γνωστές ή μπορούν να εκτιμηθούν με σχετική ακρίβεια. Επιπλέον, είναι ιδανικά για καταστάσεις όπου η εξαρτημένη μεταβλητή διατηρεί γραμμική ή σχετικά απλή μη γραμμική σχέση με τις εξηγητικές μεταβλητές, και όπου τα δεδομένα παραμένουν σχετικά σταθερά και συνεπή κατά τη διάρκεια μιας χρονικής περιόδου.

Οι Ma et al. (2022) ανέπτυξαν ένα μοντέλο VAR για να προβλέψουν τις βραχυπρόθεσμες τιμές των χοίρων, με βάση την ανάλυση των παραγόντων που επηρεάζουν αυτές τις τιμές. Τα ευρήματά τους έδειξαν ότι το μοντέλο VAR είχε καλή απόδοση για βραχυπρόθεσμες προβλέψεις. Ωστόσο, όταν εφαρμόστηκε για μεσοπρόθεσμες και μακροπρόθεσμες προβλέψεις, τα

αποτελέσματα ήταν λιγότερο ικανοποιητικά. Παράλληλα, οι Ge et al. (2020) διερεύνησαν τις διακυμάνσεις των τιμών του καλαμποκιού και τους παράγοντες που τις επηρεάζουν. Στη μελέτη τους, ανέπτυξαν δύο διαφορετικά μοντέλα για την πρόβλεψη των τιμών του καλαμποκιού: ένα μονομεταβλητό μη γραμμικό μοντέλο παλινδρόμησης, το οποίο χρησιμοποιεί τον χρόνο ως ανεξάρτητη μεταβλητή, και ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης που ενσωματώνει τους όγκους παραγωγής, κατανάλωσης, εισαγωγών και εξαγωγών ως ανεξάρτητες μεταβλητές. Ενώ το μονομεταβλητό μοντέλο μη γραμμικής παλινδρόμησης παρείχε λογικές προβλέψεις για τις τιμές του καλαμποκιού, δεν μπορούσε να εξετάσει διεξοδικά όλους τους περίπλοκους εσωτερικούς παράγοντες που επηρεάζουν τις διακυμάνσεις των τιμών. Αυτό επηρεάζει την ακρίβεια των προβλέψεων, καθιστώντας το μοντέλο κατάλληλο μόνο για αρχικές εκτιμήσεις. Επιπλέον, η βασική υπόθεση της ανεξαρτησίας στο τοπικό μοτίβο προκαλεί αποκλίσεις όταν εφαρμόζονται εξισώσεις πρόβλεψης παλινδρόμησης σε μεσοπρόθεσμες και μακροπρόθεσμες προβλέψεις. Η πολυπλοκότητα των παραγόντων που διαμορφώνουν τις τιμές των γεωργικών εμπορευμάτων παρουσιάζει προκλήσεις για τη συμπερίληψη όλων των σχετικών μεταβλητών στη διαδικασία μοντελοποίησης. Παρ' όλα αυτά, η ανάλυση παλινδρόμησης είναι εξαιρετική στην αποκάλυψη των εσωτερικών προτύπων, σχέσεων και συσχετισμών μεταξύ των διαφόρων παραγόντων, γεγονός που συμβάλλει στη σχετικά υψηλή ακρίβειά της. Η απλότητα της εφαρμογής της και η δυνατότητα βελτίωσης των βασικών μοντέλων την καθιστούν δημοφιλή επιλογή για βραχυπρόθεσμες προβλέψεις τιμών γεωργικών προϊόντων.

2.2.2 Μέθοδος πρόβλεψης του μοντέλου Gray

Το μοντέλο Gray αποτελεί μια μέθοδο πρόβλεψης για συστήματα που περιέχουν αβεβαιότητες. Πρόκειται για ένα ημιπαραμετρικό μοντέλο που χρησιμοποιεί έναν περιορισμένο αριθμό δεδομένων για να διαμορφώσει διαφορικές εξισώσεις, οι οποίες περιγράφουν τις τάσεις που εμφανίζονται στα δεδομένα. Ο κύριος στόχος είναι η εκτίμηση των παραμέτρων της εξίσωσης μέσω της μεθόδου των ελαχίστων τετραγώνων, ώστε να χρησιμοποιηθούν για την πρόβλεψη των μελλοντικών τιμών της μεταβλητής. Ορισμένα παραδείγματα των μοντέλων Gray περιλαμβάνουν τα GM (1, 1), GM (2, 1) και DGM (2, 1). Τα μοντέλα αυτά απαιτούν από τα δεδομένα να παρουσιάζουν ένα επίπεδο κανονικότητας και μονοτονίας, δηλαδή να αυξάνονται ή να μειώνονται σταθερά με την πάροδο του χρόνου. Επιπλέον, τα δεδομένα υποτίθεται ότι ακολουθούν εκθετική κατανομή, κάτι που υποδηλώνει ότι η μεταβολή τους με τον χρόνο είναι εκθετικά αυξανόμενη ή φθίνουσα. Τα μοντέλα Gray ενδέχεται να μην είναι αποδοτικά όταν τα δεδομένα παρουσιάζουν ακανόνιστα, μη μονοτονικά ή μη εκθετικά μοτίβα, καθώς και όταν υπάρχουν απότομες αλλαγές ή διακυμάνσεις. Επιπλέον, αυτά τα μοντέλα μπορεί να αντιμετωπίζουν προβλήματα με τη χαμηλή ακρίβεια και την προσαρμοστικότητά τους, κάτι που

μπορεί να επηρεάσει την προβλεπτική τους απόδοση και την ανθεκτικότητά τους. Ωστόσο, τα μοντέλα Gray είναι κατάλληλα για μακροπρόθεσμες προβλέψεις σε περιπτώσεις όπου τα διαθέσιμα δεδομένα είναι σπάνια ή ελλιπή. Είναι επίσης κατάλληλα για προβλήματα όπου οι μεταβλητές εμφανίζουν ομαλές και μονότονες τάσεις, καθώς και για δεδομένα με σχετικά απλά και σταθερά μοτίβα που διατηρούνται με την πάροδο του χρόνου.

Το Luo Han Guo είναι ένα είδος φρούτου που παράγεται στο Guangxi της Κίνας. Οι Feng et al. (2012) κατασκεύασαν ένα μοντέλο GM (1, 1) για να προβλέψουν την τιμή του. Διαπιστώθηκε ότι το μοντέλο GM (1, 1) θα μπορούσε να απεικονίσει καλύτερα το μοτίβο μεταβολής της τιμής του λόγω της υψηλής ακρίβειας προσομοίωσης στο δείγμα. Συνάγεται το συμπέρασμα ότι το μοντέλο GM (1, 1) έχει τα πλεονεκτήματα του ότι απαιτεί λιγότερα δεδομένα, υψηλή ακρίβεια προσαρμογής και πρόβλεψης και εύκολη εφαρμογή προγραμματισμού στο πρόβλημα πρόβλεψης για την τιμή του Luo Han Guo, σε σχέση με τα μοντέλα πρόβλεψης, όπως π.χ. μοντέλα παλινδρόμησης και μοντέλα χρονοσειρών και μπορεί να παρέχουν μία επιστημονική αναφορά για την πρόβλεψη της τιμής του Luo Han Guo.

2.2.3 Μέθοδος πρόβλεψης με την χρήση των χρονοσειρών

Η ανάλυση χρονοσειρών αποτελεί μια ευρέως χρησιμοποιούμενη μέθοδο μονομεταβλητής πρόβλεψης. Πρόκειται για μια στατιστική τεχνική που επικεντρώνεται στη μοντελοποίηση και ανάλυση των τιμών γεωργικών προϊόντων, λαμβάνοντας υπόψη την κανονικότητα που παρουσιάζουν οι τιμές με την πάροδο του χρόνου, ώστε να προβλεφθούν μελλοντικά δεδομένα από τα υπάρχοντα. Οι μέθοδοι ανάλυσης χρονοσειρών περιλαμβάνουν κυρίως την αυτοπαλινδρομική κίνηση μέσου όρου (ARMA), τον αυτοπαλινδρομικό ολοκληρωμένο κινητό μέσο όρο (ARIMA), τον εποχικό αυτοπαλινδρομικό ολοκληρωμένο κινητό μέσο όρο (SARIMA), την αυτοπαλινδρομική υπό συνθήκη ετεροσκεδαστικότητα (ARCH) τη γενικευμένη αυτοπαλινδρομική υπό συνθήκη ετεροσκεδαστικότητα (GARCH) κτλ.

Τα πλεονεκτήματα της ανάλυσης χρονοσειρών είναι η απλότητά της και η άμεση εφαρμογή της. Βασίζεται αποκλειστικά σε ιστορικά δεδομένα και είναι αρκετά εύελικτη, γεγονός που την καθιστά κατάλληλη για βραχυπρόθεσμες προβλέψεις. Η μέθοδος αποδίδει ιδιαίτερα καλά όταν τα δεδομένα παρουσιάζουν σαφή εποχικά, τάσης και κυκλικά μοτίβα. Η δημιουργία των μοντέλων και των προβλέψεων γίνεται χωρίς την ανάγκη ενσωμάτωσης άλλων παραγόντων που επηρεάζουν τις τιμές. Οι τεχνικές πρόβλεψης χρονοσειρών υποθέτουν ότι το μελλοντικό μοτίβο μεταβολών θα ακολουθήσει το ιστορικό μοτίβο. Ωστόσο, στην πράξη, οι εξωτερικοί παράγοντες μπορούν να επηρεάσουν τις τιμές, με αποτέλεσμα να προκύπτουν μεροληπτικές ή ανακριβείς

προβλέψεις. Η κλιματική αλλαγή, οι αλλαγές στις εφαρμοζόμενες πολιτικές και τα απρόβλεπτα γεγονότα μπορούν να οδηγήσουν σε δομικές αλλαγές στις χρονοσειρές, με αποτέλεσμα τα ιστορικά δεδομένα να μην αποτελούν μία αντιπροσωπευτική αντανάκλαση του μέλλοντος. Η μέθοδος απαιτεί πολύπλοκα βήματα, όπως ο έλεγχος ομαλότητας των δεδομένων, η εκτίμηση παραμέτρων και η επιλογή μοντέλου, που συχνά απαιτούν εξειδικευμένες γνώσεις και δεξιότητες και μπορεί να είναι υποκειμενικές και αβέβαιες. Για παράδειγμα, ένα μοντέλο ARIMA απαιτεί τον προσδιορισμό των τιμών των παραμέτρων p , d και q . Αυτές οι τιμές ενδέχεται να επηρεάσουν την αποτελεσματικότητα προσαρμογής του μοντέλου και την ακρίβεια της πρόβλεψης. Επιπλέον, οι μέθοδοι πρόβλεψης χρονοσειρών συχνά υποφέρουν από σφάλματα συσσώρευσης κατά την εκτέλεση προβλέψεων πολλαπλών βημάτων, με αποτέλεσμα κακές μακροπρόθεσμες προβλέψεις. Για παράδειγμα, εάν χρησιμοποιείται μία μέθοδος κινητού μέσου όρου για την πρόβλεψη δεδομένων σε πολλά σημεία στο μέλλον, τα δεδομένα από προηγούμενες προβλέψεις θα πρέπει να χρησιμοποιηθούν ως είσοδος, κάτι που θα μετατοπίσει το σφάλμα από προηγούμενες προβλέψεις αργότερα, καθιστώντας τις προβλέψεις όλο και περισσότερο ανακριβείς. Ο Πίνακας 1 συνοψίζει τις κυριότερες εφαρμογές των μεθόδων πρόβλεψης με τη χρήση της ανάλυσης χρονοσειρών.

Πίνακας 1. Κοινές μέθοδοι ανάλυσης χρονοσειρών

Μοντέλο	Αρχή	Χαρακτηριστικά	Πηγή
AR (Αυτοπαλινδρομικό μοντέλο – Autoregressive model)	Το μοντέλο AR υποθέτει ότι οι τρέχουσες παρατηρήσεις είναι ένας γραμμικός συνδυασμός των παρατηρήσεων του παρελθόντος και χρησιμοποιεί ιστορικές παρατηρήσεις για να προβλέψει τις μελλοντικές τιμές.	Αποτύπωση των δυναμικών ιδιοτήτων και εξελικτικές τάσεις χρονοσειρών. Χειρίζεται δεδομένα χρονοσειρών με μεγάλη μνήμη.	Cuaresma et al., 2004
MA (Μοντέλο κινητού μέσου όρου – Moving average model)	Αντικατοπτρίζει νέες παρατηρήσεις με συνεχή ενημέρωση του κινούμενου μέσου όρου. Περιλαμβάνει έναν σταθμισμένο μέσο όρο της μέσης τιμής, των	Ικανό να συλλαμβάνει την τυχαιότητα και αβεβαιότητα των διακυμάνσεων των τιμών, ειδικά όταν οι συνθήκες προσφοράς και ζήτησης	Sun et al., 2023

	σφαλμάτων λευκού θορύβου και των καθυστερημένων τιμών.	της αγοράς είναι ασταθείς.	
ARIMA (Αυτοπαλινδρομικό ολοκληρωμένο μοντέλο κινητού μέσου όρου – Autoregressive Integrated Moving Average Model)	Η βασική αρχή του μοντέλου ARIMA είναι η χρήση των παρελθόντων σημειακών δεδομένων, σφαλμάτων και διαφορών για την πρόβλεψη μελλοντικών σημείων δεδομένων. Το μοντέλο ελαχιστοποιεί το σφάλμα πρόβλεψης προσαρμόζοντας τις παραμέτρους των αυτοπαλινδρομικών συντελεστών, των συντελεστών του κινούμενου μέσου όρου και των πράξεων διαφοράς.	Δυνατότητα λήψης τάσεων χρονοσειρών και εποχικότητα. Ουσιαστικά συλλαμβάνει μόνο γραμμικές σχέσεις. Απαιτείται τα δεδομένα χρονισμού να είναι σταθερά ή να έχουν γίνει σταθερά με διαφορική διαφοροποίηση.	Jadhav et al., 2017
ES (Εκθετική Εξομάλυνση – Exponential Smoothing)	Βασίζεται στην αρχή της χρήσης ιστορικών δεδομένων για να προβλέψει τις μελλοντικές τάσεις προσαρμόζοντας συνεχώς τα βάρη ώστε τα πλέον πρόσφατα στοιχεία να έχουν μεγαλύτερη επίδραση στα αποτελέσματα της πρόβλεψης της τάσης και της περιοδικότητας των δεδομένων χρονοσειρών.	Μπορεί να μειώσει τον θόρυβο και την εποχικότητα της χρονοσειράς, βελτιώνοντας έτσι την ακρίβεια πρόβλεψης και τη σταθερότητα. Η πρόβλεψη νέων τάσεων και κύκλων απαιτεί συνεχή ενημέρωση του μοντέλου. Η επιλογή των σταθερών εξομάλυνσης είναι ευαίσθητη και λιγότερο αποτελεσματική για χρονοσειρές με έντονες περιοδικές διακυμάνσεις.	Wu et al., 2016

2.2.4 Prophet

Το Prophet αποτελεί ένα πλαίσιο ανοιχτού κώδικα, το οποίο αναπτύχθηκε από το Facebook για τη μοντελοποίηση και πρόβλεψη χρονοσειρών. Βασίζεται σε ένα προσθετικό μοντέλο, το οποίο έχει τη δυνατότητα να χειρίζεται μη γραμμικές σχέσεις και παρουσιάζει ευελιξία στον χειρισμό δεδομένων που λείπουν, καθώς και στις μεταβολές τάσεων. Το μοντέλο Prophet είναι εξαιρετικά ισχυρό στην αντιμετώπιση ακραίων τιμών και επιτρέπει την ενσωμάτωση εξωγενών μεταβλητών για τη βελτίωση των προβλέψεων.

Το Prophet επιχειρεί να επιλύσει δύο βασικά ζητήματα που προκύπτουν με τα αυτόματα και τα πιο παραδοσιακά εργαλεία πρόβλεψης χρονοσειρών. Πρώτον, πολλά αυτόματα μοντέλα παρουσιάζουν περιορισμούς, καθώς είναι ακατάλληλα για την εισαγωγή επιπρόσθετων υποθέσεων ή δεν έχουν τη δυνατότητα να προσαρμόζονται σε εξωτερικές συνθήκες. Δεύτερον, τα ισχυρότερα εργαλεία συχνά απαιτούν την εξειδικευμένη γνώση ενός αναλυτή, ο οποίος πρέπει να έχει βαθιά εμπειρία στις μεθόδους πρόβλεψης.

Το Prophet έχει τη δυνατότητα να δουλεύει με δεδομένα που είναι διαθέσιμα σε διάφορες χρονικές κλίμακες, όπως ωριαία, μηνιαία ή ετήσια δεδομένα, και μπορεί να αποδώσει αποτελεσματικά, υπό την προϋπόθεση ότι υπάρχει τουλάχιστον ένα έτος ιστορικών δεδομένων. Είναι ιδανικό για δεδομένα με έντονη εποχικότητα, για δεδομένα με εξαιρετικά γεγονότα που δεν ακολουθούν τακτική περιοδικότητα, όπως είναι η περίοδος των Χριστουγέννων, και για περιπτώσεις όπου λείπουν δεδομένα ή υπάρχουν ακραίες τιμές. Επιπλέον, το Prophet διαχειρίζεται αποτελεσματικά σημαντικές αλλαγές τάσεων, όπως λανσαρίσματα νέων προϊόντων ή υπηρεσιών, καθώς και τάσεις που τείνουν ασυμπτωτικά σε ανώτερα ή κατώτερα όρια.

2.3 Έξυπνες μέθοδοι προβλέψεις

Οι παραδοσιακές μέθοδοι πρόβλεψης των γεωργικών τιμών, όπως η ανάλυση παλινδρόμησης, η ανάλυση χρονοσειρών και τα μοντέλα Gray, είναι γενικά αποτελεσματικές όταν οι μεταβλητές είναι ανεξάρτητες, τα δεδομένα ακολουθούν μια κανονική κατανομή και υπάρχει γραμμική ή απλή μη γραμμική σχέση. Ωστόσο, στην πράξη, η πρόβλεψη των γεωργικών τιμών συχνά δεν πληροί αυτές τις προϋποθέσεις και παρουσιάζει προκλήσεις, όπως υψηλές διαστάσεις, περιορισμένα δείγματα και σύνθετες μη γραμμικές σχέσεις.

Σε σύγκριση με τις οικονομετρικές και τις μαθηματικές-στατιστικές μεθόδους, οι μέθοδοι έξυπνης πρόβλεψης έχουν λιγότερους περιορισμούς και υποθέσεις κατά τη μοντελοποίηση και μπορούν να αποτυπώσουν αποτελεσματικά μη γραμμικές σχέσεις σε χρονολογικές σειρές τιμών.

Οι παραδοσιακές μέθοδοι μηχανικής μάθησης, όπως τα δέντρα απόφασης, οι μηχανές διανυσμάτων υποστήριξης (SVM) και η Bayes, διαθέτουν πλεονεκτήματα όπως η απλότητα, η γρήγορη εκπαίδευση και η ανθεκτικότητα, αλλά έχουν περιορισμένη ικανότητα να διαχειρίζονται περίπλοκες μη γραμμικές σχέσεις. Επιπλέον, αυτές οι μέθοδοι απαιτούν χειροκίνητη επιλογή και εξαγωγή χαρακτηριστικών και έχουν περιορισμένη ικανότητα γενίκευσης. Τα μοντέλα βαθιάς μάθησης, χάρη στις ισχυρές εκφραστικές δυνατότητές τους και την αυτόματη εξαγωγή χαρακτηριστικών, μπορούν να αντλήσουν χρήσιμες πληροφορίες από τις αρχικές ακολουθίες δεδομένων χωρίς να βασίζονται στη μηχανική εξαγωγή χαρακτηριστικών. Αυτό τα καθιστά ικανά να διαχειριστούν καλύτερα μη γραμμικές σχέσεις, αρκεί να υπάρχει επαρκής εποπτεία, επαρκής όγκος δεδομένων και υψηλή ποιότητα αυτών των δεδομένων. Ωστόσο, οι μέθοδοι βαθιάς μάθησης έχουν και τα μειονεκτήματά τους, όπως η ανάγκη για μεγάλα σύνολα δεδομένων, οι δυσκολίες στη ρύθμιση των παραμέτρων, ο κίνδυνος υπερπροσαρμογής και η δυσκολία στην ερμηνεία των αποτελεσμάτων. Στην παρούσα παράγραφο περιγράφονται δύο παραδοσιακές μέθοδοι μηχανικής μάθησης, οι μηχανές διανυσμάτων υποστήριξης και η απλή Bayes, καθώς και η εφαρμογή μεθόδων νευρωνικών δικτύων για την πρόβλεψη των τιμών των γεωργικών προϊόντων.

2.3.1 Μηχανές διανυσμάτων υποστήριξης

Η μηχανή διανύσματος υποστήριξης (support vector machine - SVM) είναι μία προσέγγιση μηχανικής μάθησης που βασίζεται στη στατιστική θεωρία μάθησης (Zhang, 2000). Βασίζεται στη θεωρία διαστάσεων Vapnik – Chernovenkis (VC), την αρχή της δομικής ελαχιστοποίησης κινδύνου (de Mello & Ponti, 2018) και αντιπροσωπεύει τον πρωτοποριακό αλγόριθμο που βασίζεται στη γεωμετρική απόσταση (Andrew, 2000). Χρησιμοποιεί ως τεχνική εκμάθησης μικρού δείγματος με ισχυρή θεωρητική θεμελίωση, ενώ η συνάρτηση τελικής απόφασης ενός SVM επηρεάζεται μόνο από έναν μικρό αριθμό υποστηρικτικών διανυσμάτων. Η υπολογιστική πολυπλοκότητα της συγκεκριμένης προσέγγισης εξαρτάται από αυτά τα διανύσματα και όχι από τη διάσταση του χώρου του δείγματος, παρακάμπτοντας τη λεγόμενη «διαστατική καταστροφή». Οι Wang et al. (2018) αξιοποίησαν το SVM για να προβλέψουν τη μη γραμμική πτυχή των τιμών του σκόρδου, συνδέοντάς το με την ARIMA για τη γραμμική πρόβλεψη τιμών, με ακριβή αποτελέσματα. Ωστόσο, το SVM έχει μειονεκτήματα, συμπεριλαμβανομένης της μειωμένης απόδοσης όταν τα χαρακτηριστικά δεδομένων (διαστάσεις) ξεπερνούν το μέγεθος του δείγματος, την ευαισθησία σε παραμέτρους και τις λειτουργίες του πυρήνα του. Κατά συνέπεια, προσεγγίσεις όπως η βελτιστοποίηση παραμέτρων χρησιμοποιούνται συχνά για τη βελτίωση της απόδοσης πρόβλεψης του SVM. Οι Duan et al. (2017) χρησιμοποίησαν έναν γενετικό αλγόριθμο για τον εντοπισμό των βέλτιστων συνδυασμών παραμέτρων για ένα μοντέλο υποστηρικτικής

παλινδρόμησης. Με αυτές τις βελτιστοποιημένες παραμέτρους, κατασκεύασαν ένα μοντέλο υποστηρικτικής παλινδρόμησης διανυσμάτων για την πρόβλεψη των τιμών των ψαριών, που αποδίδει ακριβή αποτελέσματα με ελάχιστα σφάλματα. Η αξιοσημείωτη ικανότητα της Παλινδρόμησης Διανύσματος Υποστήριξης (Support Vector Regression – SVR) να διαχειρίζεται θέσεις δεδομένων πολλών διαστάσεων, μη γραμμικής φύσεως και μικρού μεγέθους δείγματος, είναι μία ιδιαιτέρως σημαντική τεχνική για την πρόβλεψη των τιμών των γεωργικών προϊόντων.

2.3.2 Μέθοδος πρόβλεψης του Bayes

Ένα δίκτυο Bayes είναι ουσιαστικά ένα κατευθυνόμενο άκυκλο γράφημα που χρησιμοποιεί πιθανοτικά δίκτυα για την εξαγωγή συμπερασμάτων αβεβαιότητας. Η άριστη απόδοση των μπεϊσιανών δικτύων στην επίλυση προβλημάτων πρόβλεψης των γεωργικών τιμών καθώς και άλλων αγροτικών προβλημάτων πηγάζουν κυρίως από τα ακόλουθα βασικά χαρακτηριστικά: (1) Τα δίκτυα Bayes μπορούν να χειριστούν ελλιπή σύνολα δεδομένων, (2) Τα δίκτυα Bayes επιτρέπουν να γίνουν κατανοητές οι σχέσεις μεταξύ των μεταβλητών και να ποσοτικοποιηθεί η δύναμη αυτών των σχέσεων, (3) έχουν την ικανότητα συνδυασμού ποσοτικών και ποιοτικών δεδομένων, (4) έχουν την ικανότητα να συνδυάζουν γνώσεις και δεδομένα εμπειρογνομόνων στο δίκτυο και (5) οι Μπεϊσιανές μέθοδοι μπορούν σχετικά εύκολα να αποφύγουν την υπερπροσαρμογή δεδομένων κατά τη διάρκεια της μαθησιακής διαδικασίας. Οι Nuvaisyah et al. (2018) χρησιμοποίησαν αλγόριθμους δικτύων Bayes ως μέθοδο ταξινόμησης εξόρυξης δεδομένων για την πρόβλεψη των τιμών των προϊόντων πιπεριού στην περιοχή Bandung με βάση τις καιρικές πληροφορίες. Ένα μειονέκτημα των δικτύων Bayes είναι το ότι δεν υποστηρίζουν δίκτυα δακτυλίου (Ticehurst et al., 2008), το οποίο θα αποδυνάμωνε την ισχυρή ικανότητα συμπερασμάτων του δικτύου, και αυτός ο περιορισμός δεν είναι φιλικός προς τα στατικά δίκτυα Bayes. Το δυναμικό δίκτυο Bayes (Dynamic Bayesian Network - DBN) είναι ένα δυναμικό μοντέλο συγχώνευσης της θεωρίας πιθανοτήτων και του διαγράμματος επιρροής. Συνδυάζει ένα χρονικά μεταβαλλόμενο κρυφό μοντέλο Markov με ένα παραδοσιακό στατικό δίκτυο Bayes, αποσπώντας οφέλη και από τα δύο ενώ παρακάμπτει τους περιορισμούς τους μέσω της δυναμικής προσαρμοστικότητας με την πάροδο του χρόνου και της ενσωμάτωσης νέων καταστάσεων (Pearl, 1998). Ο Ma (2019) χρησιμοποίησε τον αλγόριθμο PC για να μάθει από τα δεδομένα, να κατασκευάσει σύμφωνα με τις γνώσεις των ειδικών και να συνδυάσει τις γνώσεις των ειδικών και του αλγορίθμου PC για την εκτέλεση της δομικής μάθησης. Αφού απέκτησε την αρχική δομή, προσάρμοσε την αποκτηθείσα αρχική δομή για να ληφθεί η δομή δικτύου του μοντέλου, και στη συνέχεια χρησιμοποίησε τον αλγόριθμο EM για την εκτέλεση της εκμάθησης των παραμέτρων. Επιπλέον, απέκτησε ένα πλήρες δυναμικό μοντέλο δικτύου Bayes για την πρόβλεψη τιμών και επέλεξε το καλύτερο μοντέλο με βάση τα αποτελέσματα της πρόβλεψης για την πρόβλεψη της

τιμής και της παραγωγής ζωντανών χοίρων. Τα αποτελέσματα δείχνουν ότι το αποτέλεσμα της πρόβλεψης είναι καλύτερο από το ARIMA, το SVM και το BP της ομάδας ελέγχου των μοντέλων δικτύου.

2.3.3 Μέθοδος πρόβλεψης νευρωνικών δικτύων

Τα νευρωνικά δίκτυα, γνωστά και ως τεχνητά νευρωνικά δίκτυα (ΤΝΔ), είναι πολύπλοκα μη γραμμικά συστήματα. Αποτελούνται από πολλές μονάδες επεξεργασίας που είναι διασυνδεδεμένες, μιμούμενες τον τρόπο με τον οποίο λειτουργούν οι βιολογικοί νευρώνες. Τα νευρωνικά δίκτυα παρουσιάζουν ισχυρές μη γραμμικές δυνατότητες προσαρμογής, δίνοντάς τη δυνατότητα χαρτογράφησης περίπλοκων μη γραμμικών σχέσεων. Επιπλέον, οι κανόνες εκμάθησης των νευρωνικών δικτύων είναι σχετικά απλοί, καθιστώντας τα εύκολα στην εφαρμογή σε υπολογιστικά συστήματα. Διαθέτουν υψηλή ανθεκτικότητα, μνήμη, και ικανότητα μη γραμμικής χαρτογράφησης, ενώ παράλληλα έχουν ισχυρή ικανότητα αυτό-μάθησης. Αυτά τα χαρακτηριστικά τους δίνουν μοναδικά πλεονεκτήματα στην αντιμετώπιση των προκλήσεων που σχετίζονται με την πρόβλεψη των τιμών των γεωργικών προϊόντων. Το 1987, οι Lapedes και Farber έθεσαν τις βάσεις για τη χρήση των νευρωνικών δικτύων στην πρόβλεψη, ανοίγοντας τον δρόμο για την εφαρμογή τους σε προβλέψεις αυτού του είδους. Το 1993, οι Kohzadi et al. (1996) ήταν από τους πρώτους που χρησιμοποίησαν τα νευρωνικά δίκτυα feedforward για την πρόβλεψη των τιμών του σιταριού και των βοοειδών στις ΗΠΑ. Συνέκριναν τα προγνωστικά αποτελέσματα με αντίστοιχα που προέκυψαν από τη μέθοδο ARIMA, καταλήγοντας στο συμπέρασμα του ότι τα νευρωνικά δίκτυα εμφάνισαν ανώτερες δυνατότητες πρόβλεψης του σημείου καμπής και επιτεύχθηκε ακριβέστερη πρόβλεψη τιμών.

Καθώς προχωρούν οι εξελίξεις στα μεγάλα δεδομένα και η τεχνολογία της τεχνητής νοημοσύνης, τα νευρωνικά δίκτυα βρίσκουν ολοένα και περισσότερο ευρεία εφαρμογή στον γεωργικό τομέα (Liakos et al., 2018). Στον τομέα της πρόβλεψης τιμών, τα πλέον διαδεδομένα μοντέλα νευρωνικών δικτύων στις εφαρμογές των νευρωνικών δικτύων στην πρόβλεψη τιμών γεωργικών εμπορευμάτων είναι τα εξής (ο Πίνακας 2, μαζί με παραδείγματα, τα συνοψίζει):

- Δίκτυα backpropagation (BP) (Gao & An, 2021): Τα δίκτυα BP είναι εύκολο να εφαρμοστούν και να γίνουν κατανοητά. Ωστόσο, είναι εύκολο να εντοπιστούν τοπικές βέλτιστες λύσεις και στην εκπαίδευση τους η ταχύτητα είναι σχετικά αργή.
- Νευρωνικά δίκτυα ακτινικής συνάρτησης βάσης (Radial Base Function Neural Networks - RBFNN) (Zhang et al., 2018): Ένα δίκτυο BP αποτελεί μία καθολική προσέγγιση μίας μη γραμμικής χαρτογράφησης, ενώ ένα δίκτυο RBF είναι μία τοπική προσέγγιση μίας μη

γραμμικής χαρτογράφησης και εκπαιδεύεται περισσότερο γρήγορα. Το RBF μπορεί να χειριστεί πολύπλοκες μη γραμμικές σχέσεις ενώ έχει καλή ικανότητα γενίκευσης. Ωστόσο, είναι ευαίσθητο στη δομή του δικτύου και οι υπερπαραμέτροι, καθώς και η εκπαίδευση και ο συντονισμός του είναι σχετικά περίπλοκοι. Όταν ένα πρόβλημα περιέχει πολύπλοκες μη γραμμικές σχέσεις και υπάρχει επαρκής ποσότητα δεδομένων για εκπαίδευση, μπορεί να επιχειρηθεί η χρήση του νευρωνικού δικτύου RBF (Radial Basis Function).

- Δίκτυα μακράς βραχυπρόθεσμης μνήμης (Long Short Term Memory - LSTM) (Fang et al., 2021): Το νευρωνικό δίκτυο LSTM είναι ένα ειδικό είδος επαναλαμβανόμενου νευρωνικού δικτύου που λύνει τα προβλήματα της μακροπρόθεσμης εξάρτησης και την εξαφάνιση της κλίσης με την εισαγωγή δομών, όπως οι πύλες που ξεχνούν, οι πύλες εισόδου, και οι πύλες εξόδου, για τον έλεγχο της ροής των πληροφοριών μέσω των μοναδιαίων καταστάσεων. Τα νευρωνικά δίκτυα LSTM έχουν τη δυνατότητα να διατηρούν και να καταγράφουν μακροχρόνιες εξαρτήσεις. Συνεπώς, το LSTM αποτελεί μια αποτελεσματική επιλογή για προβλήματα πρόβλεψης που περιλαμβάνουν δεδομένα χρονοσειρών, ειδικά όταν υπάρχουν μακροπρόθεσμες εξαρτήσεις.
- Συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks - CNN) (Cheung et al., 2023): Τα δίκτυα CNN είναι νευρωνικά δίκτυα τροφοδοσίας πολλαπλών επιπέδων που εξαγωγή τοπικά και καθολικά χαρακτηριστικά από δεδομένα μέσω δομών, όπως συνελικτικά, ομαδικά και πλήρως συνδεδεμένα επίπεδα για την ενεργοποίηση της αυτόματης λειτουργίας μάθησης και αφαίρεσης. Στις εργασίες πρόβλεψης τιμών, τα CNN έχουν την ικανότητα να μαθαίνουν και να αποτυπώνουν σημαντικά χαρακτηριστικά, όπως χρονοσειρές, τάσεις, και περιοδικότητες στα δεδομένα εισόδου. Οι τιμές της αγοράς επηρεάζονται συχνά από έναν συνδυασμό πολλών παραγόντων, και τα CNN είναι σε θέση να διαχειριστούν αυτές τις πολύπλοκες μη γραμμικές σχέσεις πιο αποτελεσματικά.
- Χαοτικά Νευρωνικά Δίκτυα (Chaos Neural Networks - CNN) (Li et al., 2015): Τα χαοτικά νευρωνικά δίκτυα είναι μια μορφή έξυπνου συστήματος επεξεργασίας πληροφοριών που συνδυάζει την αρχή της θεωρίας του χάους με τα νευρωνικά δίκτυα. Αυτά τα δίκτυα αξιοποιούν την ευαισθησία και το απρόβλεπτο των χαοτικών φαινομένων για να ενισχύσουν τις δυνατότητες μάθησης και γενίκευσης, γεγονός που οδηγεί σε βελτιωμένη ακρίβεια στην πρόβλεψη και τη μοντελοποίηση. Με την εισαγωγή μεθόδων, όπως ο χαοτικός θόρυβος ή οι λογιστικοί χάρτες, τα χαοτικά νευρωνικά δίκτυα είναι σε θέση να αποφύγουν σε κάποιο βαθμό τα νευρωνικά δίκτυα να πέσουν στα τοπικά ελάχιστα,

επιταχύνοντας έτσι την αύξηση της εκπαιδευτικής διαδικασίας και του ποσοστού σύγκλισης.

- Μηχανές Ακραίας Μάθησης (Extreme Learning Machines – ELM) (Guo & Chen, 2015): Οι μηχανές ακραίας μάθησης είναι ένα feedforward νευρωνικό δίκτυο που προτάθηκε για πρώτη φορά από τον καθηγητή Huang Guangbin του Τεχνολογικού Πανεπιστημίου Nanjang στη Σιγκαπούρη το 2006. Το δίκτυο ELM παρουσιάζει τα πλεονεκτήματα της γρήγορης εκπαίδευσης, υψηλής ικανότητας γενίκευσης και απλής εφαρμογής.
- Νευρωνικά Δίκτυα Κυματιδίων (Wavelet Neural Network - WNN) (Puchalsky et al., 2018): Τα νευρωνικά δίκτυα κυματιδίων (WNN) είναι μία μέθοδος η οποία βασίζεται στον μετασχηματισμό κυματιδίων και τα νευρωνικά δίκτυα. Με την αποσύνθεση των αρχικών δεδομένων σε συντελεστές κυματιδίων σε διαφορετικές κλίμακες, η μέθοδος βρίσκεται σε θέση να εξάγει αποτελεσματικά διάφορα χαρακτηριστικά των δεδομένων, όπως η τάση, ο κύκλος, η εποχικότητα κ.λπ. Το δίκτυο WNN συνδυάζει την ισχυρή ικανότητα προσαρμογής των νευρωνικών δικτύων, η οποία είναι ικανή για μη γραμμική χαρτογράφηση, επιτυγχάνοντας έτσι την ακριβή πρόβλεψη των μελλοντικών τιμών. Ωστόσο, η υψηλή πολυπλοκότητα και οι υψηλές απαιτήσεις δεδομένων είναι δύο σημαντικά μειονεκτήματα αυτής της μεθόδου.

Πίνακας 2. Παραδείγματα εφαρμογής της πρόβλεψης τιμών με τη χρήση νευρωνικών δικτύων για αγροτοκτηνοτροφικά προϊόντα

Πηγή	Μοντέλο - Αλγόριθμος	Χαρακτηριστικά
Gao & An (2021)	BPNN	Ισχυρή μη γραμμική χαρτογράφηση, αυτομάθηση, αυτοπροσαρμογή, ανοχή σε σφάλματα. Μακροπρόθεσμη πρόβλεψη στην προθεσμιακή αγορά.
Zhang et al. (2018)	RBFNN	Καλύτερη προσέγγιση, ταξινόμηση, ταχύτητα εκμάθησης από BP. Απλή δομή, γρήγορη σύγκλιση, ξεπερνά τοπικά ελάχιστα. Χρήσιμο για πρόβλεψης με διαφορετικά επίπεδα τιμών.
Fan et al. (2021)	LSTM	Ξεπερνά την εξαφάνιση κλίσης σε RNN. Ιδανικό για μεγάλα χρονικά διαστήματα και καθυστερήσεις. Ο σωστός συντονισμός παραμέτρων (χρόνοι επανάληψης, ρυθμός μάθησης, μέγεθος παραθύρου, επίπεδα δικτύου) βελτιώνει την πρόβλεψη. Υψηλότερη ακρίβεια από ARIMA, MLP, SVR.

Cheung et al. (2023)	C N N	Αποτελεσματικό στην εξαγωγή χαρακτηριστικών και αυτόνομη εκμάθηση για ταξινόμηση εικόνας και ήχου. Χρησιμοποιείται για πρόβλεψη μελλοντικών τιμών καλλιεργειών, βοηθώντας στη λήψη αποφάσεων, στρατηγικό σχεδιασμό, επιλογή εταίρων, μείωση κόστους και αντιμετώπιση επισιτιστικής ανασφάλειας.
Li et al. (2015)	Χαοτικά Νευρωνικά Δίκτυα	Εξαρτώνται από τρέχουσα και προηγούμενη έξοδο, προσαρμόζονται σε μη γραμμικά δεδομένα. Ιδανικά για πρόβλεψη σύνθετων, μη γραμμικών χρονοσειρών. Υπερέχουν του ARMA σε ακρίβεια και απόδοση πρόβλεψης.
Guo & Chen (2015)	ELM	Δημιουργεί τυχαία βάρη εισόδου και κρυφά όρια, απλή ρύθμιση παραμέτρων, γρήγορη εκπαίδευση, υψηλή ακρίβεια πρόβλεψης. Υπερτερεί των παραδοσιακών αλγορίθμων σε αποφυγή τοπικών μεγίστων. Χρησιμοποιείται για πρόβλεψη τιμών σιταριού με καλή ακρίβεια.
Wang & Lu (2013)	WNN	Συνδυάζει πλεονεκτήματα νευρωνικών δικτύων και wavelet, χρησιμοποιώντας Morlet wavelet για εξαγωγή τοπικών χαρακτηριστικών και βελτίωση ακρίβειας πρόβλεψης. Χρησιμοποιείται για πρόβλεψη τιμών Radix Codonopsis και Angelica sinensis με υψηλή ακρίβεια και χαμηλό σφάλμα.

2.4 Συνδυασμός Μεθόδων

Σε πρακτικές εφαρμογές πρόβλεψης, λόγω διαφορετικών μηχανισμών μοντελοποίησης και εκκίνησης στα σημεία, συνήθως το ίδιο πρόβλημα μπορεί να έχει διαφορετικές μεθόδους πρόβλεψης. Η χρήση διαφορετικών μεθόδων πρόβλεψης παρέχει διαφορετικές χρήσιμες πληροφορίες, ενώ κάθε μέθοδος έχει τα δικά της πλεονεκτήματα και μειονεκτήματα, και δεν είναι αμοιβαία αποκλειόμενες, αλλά αλληλένδετες και συμπληρωματικές η μία στην άλλη. Μια πιο επιστημονική προσέγγιση στην πρόβλεψη είναι η χρήση της μεθόδου συνδυασμένης πρόβλεψης, όπου συνδυάζονται διάφορες μεθόδους πρόβλεψης. Σε αυτή την περίπτωση, το μοντέλο πρόβλεψης αποτελείται από δύο ή περισσότερα επιμέρους μοντέλα που συνεργάζονται για την πρόβλεψη μεταβλητών. Αυτή η μέθοδος μπορεί να αξιοποιήσει σε μεγαλύτερο βαθμό τις πληροφορίες των δεδομένων, ξεπερνώντας τους περιορισμούς ενός και μόνο μοντέλου, το οποίο είναι πιο επιρρεπές σε τυχαίους παράγοντες. Ο συνδυασμός των διαφόρων πληροφοριών από

πολλαπλές μεθόδους οδηγεί σε μια πιο ολοκληρωμένη και ακριβή πρόβλεψη. Η συνδυαστική πρόβλεψη αποτελεί σημαντικό πεδίο έρευνας στον τομέα της πρόβλεψης και από τότε που οι Bates & Granger πρότειναν το θεωρητικό της πλαίσιο το 1969, έχει προσελκύσει έντονο ερευνητικό ενδιαφέρον. Υπάρχουν διάφοροι τρόποι ταξινόμησης των συνδυαστικών μοντέλων πρόβλεψης. Προκειμένου να διευκρινιστούν τα συνδυαστικά μοντέλα πρόβλεψης περισσότερο αναλυτικά, μπορούν να χωριστούν σε δύο κύριες κατηγορίες: τα παραδοσιακά συνδυαστικά μοντέλα πρόβλεψης και τα μοντέλα πρόβλεψης που βασίζονται στην ανάλυση και τον συνδυασμό. Ο Πίνακας 3 συνοψίζει αρκετά παραδείγματα συνδυαστικών μοντέλων.

Πίνακας 3. Παραδείγματα εφαρμογών συνδυασμών μεθόδων

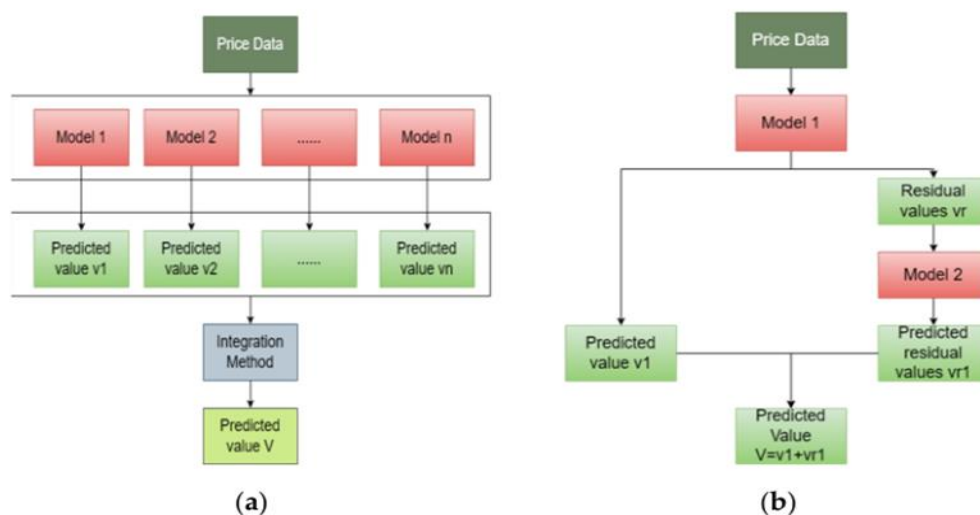
Πηγή	Μέθοδος ολοκλήρωσης	Παραδείγματα εφαρμογών	Μέθοδος αξιολόγησης
Lun et al. (2021)	Ανατίθενται διαφορετικά βάρη βάσει σφάλματος πρόβλεψης του μοντέλου και στη συνέχεια συνοψίζονται.	Χρησιμοποιήθηκαν τα μοντέλα ARIMA, GM(1,1) και συνδυαστικό μοντέλο για πρόβλεψη τιμών πατάτας το 2020. Τα αποτελέσματα ARIMA και GM συνδυάστηκαν γραμμικά για την τελική πρόβλεψη.	Η μελέτη χρησιμοποίησε το απόλυτο σφάλμα και το ποσοστό απόλυτου σφάλματος για αξιολόγηση. Το συνδυαστικό μοντέλο έδειξε καλύτερη πρόβλεψη, υψηλότερη ακρίβεια και σταθερότητα σε σχέση με τα μεμονωμένα μοντέλα.
Wang et al. (2018)	Μέθοδος ίσων βαρών	Το μοντέλο ARIMA-SVM χρησιμοποιείται για την πρόβλεψη τιμών σκόρδου. Τα αποτελέσματα πρόβλεψης από ARIMA και SVM συνδυάζονται για την τελική πρόβλεψη.	Χρησιμοποιήθηκε το σφάλμα ρίζας για αξιολόγηση. Τα αποτελέσματα υποδεικνύουν ότι το υβριδικό μοντέλο ARIMA-SVM παρουσιάζει υψηλότερη ακρίβεια στην πρόβλεψη των τιμών του σκόρδου σε σύγκριση με τα μεμονωμένα μοντέλα ARIMA και SVM, καθιστώντας το μια αποτελεσματική επιλογή για βραχυπρόθεσμες προβλέψεις.
Guo et al.	Μη γραμμικός συνδυασμός διαφορετικών αποτελεσμάτων πρόβλεψης	Ένας συνδυασμός των μοντέλων πρόβλεψης LSTM – ARIMA – BP για την πρόβλεψη τις τιμής του καλαμποκιού.	Η απόδοση αξιολογήθηκε με το μέσο ποσοστό απόλυτου σφάλματος, το σφάλμα ρίζας, και το μέσο απόλυτο σφάλμα. Τα αποτελέσματα δείχνουν ότι το μοντέλο είναι κατάλληλο για

(2022)	του μοντέλου BP		προβλέψεις τόσο σε σταθερές όσο και σε μεγάλες αλλαγές δεδομένων, παρέχοντας ακριβείς προβλέψεις.
Yin et al. (2020)	Μέθοδος ίσων βαρών	Οι τιμές των λαχανικών αναλύθηκαν σε εποχικές, τάσης, και υπόλοιπα συνιστώσες μέσω STL. Οι εξαγόμενες μεταβλητές προέκυψαν από τα υπόλοιπα. Οι μεταβλητές εισόδου μαθαίνονται με επίπεδο προσοχής και βάρη, και προβλέπονται με LSTM για τις τιμές του επόμενου μήνα.	Η αξιολόγηση με σφάλμα ρίζας και μέσο ποσοστό απόλυτου σφάλματος έδειξε ότι το μοντέλο STL-LSTM βελτιώνει την ακρίβεια πρόβλεψης κατά 12% σε σχέση με το LSTM μόνο και επιλύει την καθυστέρηση πρόβλεψης λόγω υψηλής εποχικότητας.
Cao (2015)	Μέθοδος ίσων βαρών	Η χρονοσειρά τιμών αναλύεται σε μη γραμμική ανοδική τάση, εποχική τάση, και κυκλική διακύμανση με κυματίδια. Η μη γραμμική τάση προβλέπεται με SVM, ενώ οι υπόλοιπες τάσεις με ARIMA. Οι προβλεπόμενες τιμές αθροίζονται για την τελική πρόβλεψη.	Η αξιολόγηση με μέσο ποσοστό απόλυτου σφάλματος και σφάλμα ρίζας δείχνει ότι το συνδυαστικό μοντέλο αναλύει με ακρίβεια τις τάσεις στη χρονοσειρά τιμών και υπερτερεί σε πρόβλεψη σε σύγκριση με τα μεμονωμένα μοντέλα ARIMA και SVM.
Cai et al. (2016)	Μέθοδος ίσων βαρών	Χρησιμοποιείται η μέθοδος EMD για ανάλυση των μηνιαίων τιμών χοιρινού σε υψηλές, χαμηλές συχνότητες και υπόλοιπα για επίλυση αστάθειας. Η SVM εφαρμόζεται για πρόβλεψη κάθε ενότητας. Τα αποτελέσματα συνδυάζονται για την τελική πρόβλεψη τιμών.	Η αξιολόγηση με σφάλμα ρίζας, μέσο απόλυτο ποσοστό σφάλματος, και κατευθυντική συμμετρία δείχνει ότι το μοντέλο EMD-SVM λαμβάνει υπόψη τυχαιότητα, περιοδικότητα, και τάση τιμών χοιρινού. Παρέχει υψηλή ακρίβεια πρόβλεψης και καλύτερη κατεύθυνση τάσης, προσφέροντας νέες μεθόδους για πρόβλεψη τιμών χοιρινού τοπικών μεγίστων.

			Χρησιμοποιείται για πρόβλεψη τιμών σιταριού με καλή ακρίβεια.
--	--	--	---

2.4.1 Παραδοσιακά συνδυαστικά μοντέλα πρόβλεψης

Η αρχή των παραδοσιακών συνδυαστικών μοντέλων περιλαμβάνει τη χρήση διαφορετικών μοντέλων πρόβλεψης για την πρόβλεψη των τιμών των γεωργικών εμπορευμάτων χωριστά. Τελικά, οι προβλέψεις συνδυάζονται χρησιμοποιώντας συγκεκριμένες μεθόδους ολοκλήρωσης για να αποδώσουν το αποτέλεσμα της συνολικής πρόβλεψης (Εικόνα 2α). Η Εικόνα 2β παρέχει ένα παράδειγμα παραδοσιακού συνδυαστικού μοντέλου πρόβλεψης. Σε αυτήν την περίπτωση, χρησιμοποιούνται δύο διαφορετικά μοντέλα για την πρόβλεψη των δεδομένων τιμών και των υπολοίπων χωριστά. Στη συνέχεια, οι προβλέψεις τους αθροίζονται για να παραχθεί το τελικό αποτέλεσμα της πρόβλεψης (Wang et al., 2018).



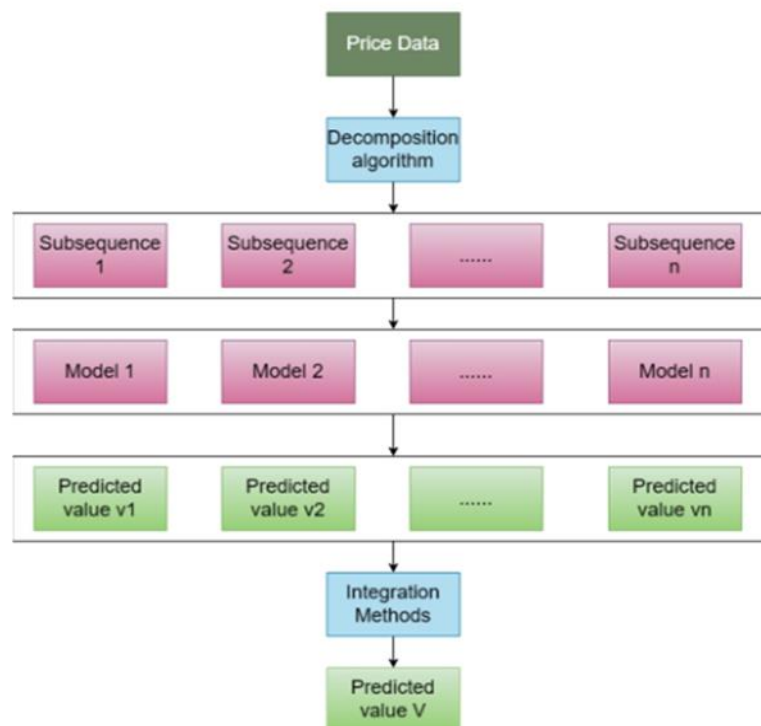
Εικόνα 2. (α) Σχηματικό διάγραμμα της παραδοσιακής διαδικασίας συνδυαστικών μοντέλων πρόβλεψης και (β) ένα παράδειγμα εφαρμογής ενός παραδοσιακού συνδυαστικού μοντέλου πρόβλεψης

Πηγή: “ <https://www.mdpi.com/2077-0472/13/9/1671> ”

2.4.2 Μέθοδος πρόβλεψης ανάλυσης – συνδυασμού

Η μέθοδος πρόβλεψης ανάλυσης - συνδυασμού βασίζεται στην ανάλυση σε πολλαπλές κλίμακες αρχικών δεδομένων σύνθετων χρονοσειρών. Αναλύει τα μοτίβα διακυμάνσεων και τάσεων σε περίπλοκα συστήματα μέσα από διαφορετικές κλίμακες. Με την κατανόηση των βασικών λειτουργικών προτύπων του συστήματος, διεξάγεται προγνωστική έρευνα, οδηγώντας σε σημαντική βελτίωση της ακρίβειας της πρόβλεψης. Το αναλυμένο συνδυαστικό συνολικό μοντέλο αναφέρεται στον διαχωρισμό της περίπλοκης ακολουθίας τιμών σε πολλές απλούστερες

υποακολουθίες. Κάθε υπο-σειρά προβλέπεται ξεχωριστά χρησιμοποιώντας τα κατάλληλα μοντέλα, και στη συνέχεια οι προβλέψεις συνδυάζονται για να προκύψουν οι προβλεπόμενες τιμές της αρχικής σειράς (όπως φαίνεται στην Εικόνα 3). Ο βασικός στόχος αυτής της προσέγγισης είναι η επιλογή αποτελεσματικών εργαλείων για την ανάλυση των δεδομένων. Οι συνήθεις μέθοδοι ανάλυσης περιλαμβάνουν την εποχική ανάλυση (Yin et al. 2020), την ανάλυση κυματιδίων (Cao & He, 2015), την εμπειρική ανάλυση (empirical modal decomposition) (Cai et al., 2016) και την ανάλυση διακύμανσης (variational modal decomposition) (Sun & Huang, 2020). Το πλεονεκτήματα των συνδυαστικών μοντέλων ανάλυσης – συνόλου έγκειται στην ικανότητά τους να αξιοποιούν πληροφορίες σε διαφορετικές κλίμακες. Μειώνουν τον αντίκτυπο παραγόντων όπως ο θόρυβος, οι τάσεις και οι κύκλοι που υπάρχουν σε σύνθετα δεδομένα, βελτιώνοντας έτσι την ακρίβεια της πρόβλεψης και την ανθεκτικότητα του μοντέλου. Ωστόσο, ένα μειονέκτημα του συνδυαστικού μοντέλου πρόβλεψης ανάλυσης - συνόλου είναι η ανάγκη προσδιορισμού κατάλληλων εργαλείων ανάλυσης δεδομένων και μεθόδων ολοκλήρωσης. Διαφορετικά, μπορεί να επηρεαστεί η διαδικασία εξαγωγής και ανακατασκευής των χαρακτηριστικών των δεδομένων. Ο Πίνακας 4 παρουσιάζει μια σύνοψη των πιο συχνά χρησιμοποιούμενων μεθόδων ανάλυσης.



Εικόνα 3. Σχηματικό διάγραμμα μοντέλων πρόβλεψης ανάλυσης – ολοκλήρωσης
 Πηγή: “ <https://www.mdpi.com/2077-0472/13/9/1671> ”

Πίνακας 4. Μέθοδοι ανάλυσης χρονοσειρών για πρόβλεψη τιμών αγροτικών προϊόντων.

Μέθοδος Ανάλυσης	Χαρακτηριστικά
STL Εποχική Ανάλυση	Αναλύει τις χρονοσειρές τιμών σε τάση, εποχική συνιστώσα και υπόλοιπα, μπορεί να χειριστεί μη-στάσιμες χρονοσειρές και είναι κατάλληλη για την πρόβλεψη τιμών αγροτοκτηνοτροφικών προϊόντων με σημαντική εποχικότητα.
Ανάλυση κυματιδίων	Η ανάλυση των χρονοσειρών τιμών σε πολλαπλές κλίμακες με τη χρήση της συνάρτησης κυματιδίων μπορεί να εξάγει χαρακτηριστικά με διαφορετικές συχνότητες, κάτι που την καθιστά κατάλληλη για την πρόβλεψη τιμών αγροτοκτηνοτροφικών προϊόντων με πολλαπλή περιοδικότητα και απότομα σημεία καμψής.
Εμπειρική Ανάλυση	Αναλύει τη χρονοσειρά τιμών σε διάφορες συναρτήσεις ιδιοτιμών και όρους υπολοίπων, οι οποίες μπορούν να χειριστούν μη γραμμικές και μη στάσιμες χρονοσειρές προσαρμοστικά και είναι κατάλληλη για την πρόβλεψη τιμών αγροτοκτηνοτροφικών προϊόντων με πολύπλοκα χαρακτηριστικά μεταβλητότητας.
Ανάλυση Διακύμανσης	Αναλύει τη χρονοσειρά τιμών σε διάφορες συναρτήσεις ιδιοτιμών, οι οποίες μπορούν να αποφύγουν αποτελεσματικά το φαινόμενο της ανάμειξης προτύπων της μεθόδου Εμπειρικής Ανάλυσης και να βελτιώσουν την επίδραση ανάλυσης, ενώ είναι κατάλληλη για την πρόβλεψη των τιμών αγροτοκτηνοτροφικών προϊόντων με συνιστώσες υψηλής και χαμηλής συχνότητας.

Είτε πρόκειται για ένα παραδοσιακό μοντέλο πρόβλεψης είτε για ένα συνδυαστικό μοντέλο πρόβλεψης ανάλυσης - συνδυασμού, η αποτελεσματικότητα της συνδυαστικής πρόβλεψης βασίζεται σε κάποιο βαθμό στην επιλεγμένη μέθοδο ολοκλήρωσης, δηλαδή σε διαφορετικά σχήματα σχεδιασμού βάρους. Η εφαρμογή αποτελεσματικών μεθόδων συνδυασμού μπορεί να οδηγήσει σε καλύτερα αποτελέσματα προβλέψεων από τις βέλτιστες μεμονωμένες προβλέψεις. Ο Πίνακας 5 παρουσιάζει μια σύνοψη διαφόρων ευρέως χρησιμοποιούμενων μεθόδων συνδυασμού. Όλες αυτές οι μέθοδοι ανήκουν στις προσεγγίσεις γραμμικής ολοκλήρωσης, όπου τα αποτελέσματα των προβλέψεων από διαφορετικά μοντέλα πολλαπλασιάζονται με συντελεστές βαρύτητας και αθροίζονται για να δώσουν το τελικό αποτέλεσμα. Αν και οι γραμμικές μέθοδοι είναι κοινές τεχνικές συνδυασμού, δεν είναι οι μόνες διαθέσιμες. Υπάρχουν

επίσης μη γραμμικές μέθοδοι συνδυασμού, όπως τα νευρωνικά δίκτυα, οι μηχανές διανυσμάτων υποστήριξης και η ασαφής λογική, που χρησιμοποιούν πιο πολύπλοκες συναρτήσεις για να ενοποιήσουν τις προβλέψεις από διάφορα μοντέλα. Σύμφωνα με τους Han et al. (2007), οι μη γραμμικές μέθοδοι συνδυασμού συχνά υπερτερούν των γραμμικών μεθόδων. Μεταξύ αυτών των μεθόδων, οι μη γραμμικής βάσης συνδυαστικές μέθοδοι πρόβλεψης νευρωνικών δικτύων παρουσιάζουν υψηλότερη προγνωστική ακρίβεια σε σχέση με άλλες ισχυρές συνδυαστικές μεθόδους. Οι Guo et al. (2022) κατασκεύασαν ένα συνδυαστικό μοντέλο LSTM-ARIMA-BP για την πρόβλεψη των τιμών του καλαμποκιού. Αυτό το μοντέλο χρησιμοποιεί ένα μοντέλο BP για την εκπαίδευση των αποτελεσμάτων πρόβλεψης των μοντέλων LSTM και ARIMA για τον υπολογισμό της τελικής τιμής πρόβλεψης και τα αποτελέσματα επιδεικνύουν μία ικανοποιητική προγνωστική απόδοση.

Πηγή	Μέθοδος ολοκλήρωσης	Αρχή	Χαρακτηριστικά
Del Sole et al. (2013)	Μέθοδος ίσης βαρύτητας	Η μέθοδος ίσης βαρύτητας αναφέρεται στην ανάθεση ίσων βαρών στα αποτελέσματα της πρόβλεψης όλων των μοντέλων και στη συνέχεια στον υπολογισμό του μέσου όρου τους για την εκτίμηση του τελικού αποτελέσματος πρόβλεψης.	Το πλεονέκτημα αυτής της προσέγγισης είναι το ότι η εφαρμογή της είναι απλή και εύκολη. Το μειονέκτημα είναι ότι δεν αντανακλά την προβλεπτική δύναμη και ακρίβεια των διαφορετικών μοντέλων και ενδέχεται να οδηγήσει σε αναποτελεσματικούς συνδυασμούς.
Takeyasu & Nagao (2008)	Μέθοδος ελάχιστης διακύμανσης	Η μέθοδος ελάχιστης διακύμανσης αναφέρεται στην ανάθεση μίας μεγαλύτερης βαρύτητας στο μοντέλο με την ελάχιστη διακύμανση βάσει της διακύμανσης των ιστορικών σφαλμάτων πρόβλεψης του κάθε μοντέλου, και στη συνέχεια στην εύρεση του σταθμισμένου μέσου όρου για τον υπολογισμό των τελικών αποτελεσμάτων πρόβλεψης.	Το πλεονέκτημα αυτής της προσέγγισης είναι ότι μπορεί να περιορίσει τη διακύμανση των συνδυασμένων προβλέψεων και να βελτιώσει την ευστάθεια. Το μειονέκτημα είναι ότι δεν μπορεί να λάβει υπόψη τη συσχέτιση ανάμεσα στα μοντέλα και ενδεχομένως να οδηγήσει σε εξάρτηση από συγκεκριμένα μοντέλα.

<p>Hou & Pelillo (2013)</p>	<p>Μέθοδος πίνακα κυριαρχίας</p>	<p>Η μέθοδος πίνακα κυριαρχίας αναφέρεται στην κατασκευή ενός πίνακα κυριαρχίας βάσει του βαθμού της κυριαρχίας πρόβλεψης του κάθε μοντέλου σε διαφορετικές χρονικές περιόδους και στη συνέχεια στην ανάθεση αντίστοιχης βαρύτητας σε κάθε στοιχείο του πίνακα βάσει του μεγέθους του και έπειτα στον υπολογισμό του σταθμισμένου μέσου όρου για την εξαγωγή του τελικού αποτελέσματος πρόβλεψης.</p>	<p>Το πλεονέκτημα αυτής της μεθόδου είναι ότι μπορεί να συνθέσει την απόδοση διαφορετικών μοντέλων σε διαφορετικές χρονικές περιόδους και το μειονέκτημα είναι ότι η κατασκευή του πίνακα κυριαρχίας απαιτεί σημαντική υποκειμενική κρίση και εμπειρία.</p>
<p>Ding (2014)</p>	<p>Μέθοδος εκτίμησης ελαχίστων τετραγώνων</p>	<p>Η μέθοδος εκτίμησης ελαχίστων τετραγώνων αναφέρεται στην εκτίμηση των βέλτιστων συντελεστών βαρύτητας βάσει της σχέσης ελαχίστων τετραγώνων ανάμεσα στα ιστορικά αποτελέσματα πρόβλεψης και τις πραγματικές τιμές του κάθε μοντέλου και στη συνέχεια στον υπολογισμό του σταθμισμένου μέσου για την εξαγωγή των τελικών αποτελεσμάτων πρόβλεψης.</p>	<p>Το πλεονέκτημα αυτής της μεθόδου είναι ότι η βέλτιστη λύση μπορεί να αποκτηθεί με τη χρήση στατιστικών μεθόδων και το μειονέκτημα είναι ότι πρέπει να ικανοποιούνται συγκεκριμένες υποθέσεις όπως η γραμμική σχέση και η κανονική κατανομή.</p>

2.5 Βελτιστοποιήσεις παραμέτρων

Κατά την καθιέρωση μοντέλων πρόβλεψης των τιμών των γεωργικών εμπορευμάτων, οι αρχικά ορισμένες ή οι λαμβανόμενες παράμετροι πιθανότατα δεν είναι βέλτιστες ή σχεδόν βέλτιστες. Σε τέτοιες περιπτώσεις, η βελτιστοποίηση των παραμέτρων είναι κρίσιμη για τη βελτίωση του μοντέλου πρόβλεψης. Συνηθισμένες τεχνικές για την παραμετροποίηση περιλαμβάνουν τη διασταυρούμενη επικύρωση, την αναζήτηση πλέγματος, τους γενετικούς αλγόριθμους, τη βελτιστοποίηση σμήνους σωματιδίων και την προσομοίωση απόπτωσης. Επιπλέον, αλγόριθμοι που εμπνέονται από συμπεριφορές κοινωνικών εντόμων ή ομάδων ζώων,

όπως οι αλγόριθμοι αποικιών μελισσών, μυρμηγκιών και σμήνους ψαριών, συχνά εφαρμόζονται για την παραμετροποίηση μοντέλων βασισμένων σε συλλογικές βιολογικές συμπεριφορές. Για παράδειγμα, οι Lu et al. (2015) αξιοποίησαν τη βελτιστοποίηση σμήνους σωματιδίων (Particle Swarm Optimization – PSO) για την ανάπτυξη ενός μοντέλου πρόβλεψης PSO-BP για τις λιανικές τιμές λαχανικών. Οι Zhang et al. (2018) πρότειναν έναν υβριδικό αλγόριθμο με την ονομασία GDGA, ο οποίος συνδυάζει τα καλύτερα στοιχεία των μεθόδων παγκόσμιας και τοπικής αναζήτησης. Τα αποτελέσματα δείχνουν ότι ο προτεινόμενος υβριδικός αλγόριθμος GDGA υπερτερεί της πολυμεταβλητής γραμμικής παλινδρόμησης και των καθαρών μεθόδων των γενετικών αλγορίθμων υπό τους όρους της προγνωστικής απόδοσης και συγκλίνει ταχύτερα από τους καθαρούς γενετικούς αλγόριθμους. Πειραματικά ευρήματα δείχνουν ότι σε σύγκριση με τις παραδοσιακές μεθόδους BP, η προσέγγιση της PSO-BP μπορεί να ξεπεράσει τα ζητήματα της υπερβολικής προσαρμογής και των τοπικών ελαχίστων, μειώνοντας αποτελεσματικά τα σφάλματα εκπαίδευσης και ενισχύοντας την προγνωστική ακρίβεια.

Με την αύξηση της σχετικής έρευνας, οι ερευνητές ανακάλυψαν σταδιακά ότι τα μοντέλα πρόβλεψης που χρησιμοποιούν τους συνδυαστικούς αλγορίθμους βελτιστοποίησης τείνουν να εμφανίζουν ανώτερη απόδοση πρόβλεψης σε σύγκριση με τους απλούς αλγόριθμους βελτιστοποίησης. Οι συνδυαστικοί αλγόριθμοι βελτιστοποίησης παραμέτρων προσφέρουν σημαντικά πλεονεκτήματα. Είναι πιο κατάλληλοι για την αντιμετώπιση πολύπλοκων προβλημάτων βελτιστοποίησης που περιλαμβάνουν διακριτές, μη γραμμικές και πολυτροπικές συναρτήσεις, σε σύγκριση με τους αλγορίθμους βελτιστοποίησης μίας μόνο παραμέτρου. Αυτοί οι τελευταίοι συχνά απαιτούν συγκεκριμένες συνθήκες, όπως η διαφορικότητα και η κυρτότητα. Επιπλέον, οι συνδυαστικοί αλγόριθμοι βελτιστοποίησης είναι πιο αποτελεσματικοί στην αποφυγή των τοπικών μεγίστων. Σε αντίθεση, οι αλγόριθμοι που βασίζονται σε μία παράμετρο είναι συχνά ευαίσθητοι στις αρχικές τιμές, γεγονός που μπορεί να οδηγήσει σε αργή σύγκλιση ή σε επιλογή μη βέλτιστων λύσεων. Επιπλέον, οι συνδυαστικοί αλγόριθμοι βελτιστοποίησης παραμέτρων μπορούν να προσαρμοστούν ευέλικτα σε διάφορα χαρακτηριστικά και απαιτήσεις του προβλήματος. Για παράδειγμα, μπορούν να εφαρμόσουν διάφορες συναρτήσεις εκπαίδευσης, στρατηγικές διασταύρωσης και μετάλλαξης, καθώς και διαφορετικές δομές γειτονίας. Σε αντίθεση, οι αλγόριθμοι βελτιστοποίησης μεμονωμένων παραμέτρων τείνουν να είναι πιο άκαμπτοι και ομοιόμορφοι, καθιστώντας την προσαρμογή και βελτίωσή τους πιο δύσκολη. Βέβαια, οι συνδυαστικοί αλγόριθμοι βελτιστοποίησης έχουν και μειονεκτήματα, όπως αυξημένη υπολογιστική πολυπλοκότητα, θεωρητικές δυσκολίες στην ανάλυση και ευαισθησία στις επιλογές των παραμέτρων. Συνεπώς, στις πρακτικές εφαρμογές, η επιλογή του κατάλληλου αλγορίθμου βελτιστοποίησης πρέπει να βασίζεται στα συγκεκριμένα χαρακτηριστικά του

προβλήματος και τους στόχους, προσαρμόζοντας και βελτιώνοντας τη διαδικασία όπου χρειάζεται. Ο Πίνακας 6 παρέχει μια συνοπτική επισκόπηση των διαφόρων κλασικών μεθόδων βελτιστοποίησης παραμέτρων, συνοδευόμενη από παραδείγματα.

Πίνακας 5. Σύνοψη μεθόδων βελτιστοποίησης και χαρακτηριστικών τους για ανάλυση και πρόβλεψη δεδομένων

Πηγή	Μέθοδος Βελτιστοποίησης	Χαρακτηριστικά
Arlot & Celisse (2010)	Διασταυρωμένη επικύρωση	<ul style="list-style-type: none"> • Αξιολογεί την απόδοση ενός μοντέλου με υποσύνολα δεδομένων. • Αποτελεσματική χρήση περιορισμένων δεδομένων. • Αποφεύγει προβλήματα υπερβολικής ή ανεπαρκούς αντιστοίχισης. • Χρήση για επιλογή βέλτιστων παραμέτρων ή χαρακτηριστικών μοντέλου.
Fayed & Atiya (2019)	Αναζήτηση πλέγματος	<ul style="list-style-type: none"> • Εξερευνά συστηματικά τον χώρο των παραμέτρων. • Εντοπίζει την ολική βέλτιστη λύση. • Υπολογιστικά ακριβή, δεν χειρίζεται συνεχείς παραμέτρους. • Κύρια χρήση: εύρεση βέλτιστων παραμέτρων.
Guo et al. (2012)	Γενετικός αλγόριθμος	<ul style="list-style-type: none"> • Προσομοιώνει βιολογική εξέλιξη για βελτιστοποίηση. • Ικανός για μη γραμμικά και συνεχή προβλήματα. • Δυνατότητα ολικής αναζήτησης, αργή σύγκλιση. • Ευρέως χρησιμοποιείται στη συνδυαστική βελτιστοποίηση και μηχανική μάθηση.
Wang et al. (2018)	Βελτιστοποίηση σωματιδίων	<ul style="list-style-type: none"> • Προσομοιώνει συμπεριφορά σμήνους για βελτιστοποίηση. • Γρήγορη σύγκλιση, απλή εφαρμογή. • Προβλήματα με υιοθέτηση τοπικών βέλτιστων λύσεων. • Χρησιμοποιείται για συνεχή προβλήματα βελτιστοποίησης.

Chen et al. (2014)	Προσομοιούμενη Ανόπτηση	<ul style="list-style-type: none"> • Προσομοιώνει διαδικασία ψύξης για βελτιστοποίηση. • Ξεπερνά τοπικά βέλτιστα, εντοπίζει ολική βέλτιστη λύση. • Κατάλληλη για συνδυαστικά προβλήματα, απαιτεί χρόνο προσαρμογής. • Χρήσεις στη μηχανική, νευρωνικά δίκτυα, επεξεργασία σήματος.
--------------------	-------------------------	--

3. Ανάλυση Δεδομένων

Για να επιτευχθεί μια επιτυχημένη ανάλυση δεδομένων, είναι κρίσιμο να διαθέτουμε καλά δομημένα σύνολα δεδομένων (datasets) που αντανακλούν με ακρίβεια τις διακυμάνσεις των τιμών των προϊόντων, καθώς και τους διάφορους παράγοντες που επηρεάζουν αυτές τις διακυμάνσεις. Ωστόσο, κατά την εκπόνηση της παρούσας έρευνας, αντιμετωπίσαμε σημαντικές προκλήσεις σχετικά με την ανεύρεση των κατάλληλων δεδομένων, ιδιαίτερα όσον αφορά τη συλλογή δεδομένων από την Ελλάδα. Αν και η αρχική πρόθεση ήταν να χρησιμοποιηθούν ελληνικά δεδομένα, αυτό αποδείχθηκε εξαιρετικά δύσκολο να πραγματοποιηθεί στην πράξη. Η απουσία κεντρικών συστημάτων καταγραφής, ο κατακερματισμός της αγοράς, καθώς και η έλλειψη επαρκών ψηφιακών υποδομών αποτέλεσαν σημαντικά εμπόδια. Επιπλέον, στην Ελλάδα δεν υφίσταται κάποιο ενιαίο και ολοκληρωμένο σύστημα που να συγκεντρώνει και να δημοσιεύει τακτικά δεδομένα τιμών για αγροτικά προϊόντα σε εθνικό επίπεδο. Οι πληροφορίες σχετικά με τις τιμές είναι συχνά αποσπασματικές και διασκορπισμένες μεταξύ διαφορετικών οργανισμών και φορέων, γεγονός που καθιστά τη συγκέντρωση αξιόπιστων και συνεκτικών δεδομένων εξαιρετικά δύσκολη και χρονοβόρα.

Με βάση τους παραπάνω περιορισμούς και κατόπιν διεξοδικής μελέτης και έρευνας, αποφασίστηκε η συλλογή των δεδομένων να πραγματοποιηθεί από την Ιταλία. Η επιλογή της Ιταλίας δεν ήταν τυχαία, καθώς στηρίχθηκε σε διάφορους παράγοντες. Οι δύο χώρες μοιράζονται παρόμοιο μεσογειακό κλίμα, το οποίο επηρεάζει την παραγωγή αγροτικών προϊόντων με συγκρίσιμο τρόπο. Επιπλέον, οι συνθήκες παραγωγής και η εποχικότητα των προϊόντων είναι παρόμοιες και στις δύο χώρες, καθιστώντας τα δεδομένα από την Ιταλία κατάλληλα για τη συγκεκριμένη ανάλυση και προσφέροντας μια συγκρίσιμη βάση για την έρευνα.

3.1 Σύνολο Δεδομένων

Τα δεδομένα συλλέχθηκαν σε μηνιαία βάση, από τον Ιανουάριο του 2007 έως τον Δεκέμβριο του 2022 και αφορούν τις τιμές για τα προϊόντα των μήλων, των αχλαδιών, του πετρελαίου κινήσεις (τα οποία συλλέχθηκαν από την [Ευρωπαϊκή Επιτροπή](#)), και τέλος, την ποσότητα παραγωγής των συγκεκριμένων προϊόντων τα οποία συλλέχθηκαν από το [Εθνικό Ινστιτούτο Στατιστικής Ιταλίας](#).

Δημιουργήθηκαν δύο datasets, ένα με τις τιμές των μήλων (DATASET_APPLE_AUTOGAS_PRICE_FINAL), και ένα με τις τιμές των αχλαδιών (DATASET_PEAR_AUTOGAS_PRICE_FINAL). Η δομή του συνόλου δεδομένων για το dataset των μήλων περιλαμβάνει την ημερομηνία καταγραφής σε μηνιαία συχνότητα (σε μορφή MM/DD/YYYY), την τιμή μήλων ανά κιλό σε ευρώ (APPLE_PRICE), την τιμή υγραερίου ανά λίτρο σε ευρώ (AUTOGAS_PRICE) και την ποσότητα παραγωγής μήλων σε χιλιάδες τόνους (PRODUCTION_QUANTITY). Αντίστοιχα το dataset των αχλαδιών περιλαμβάνει την ημερομηνία καταγραφής σε μηνιαία συχνότητα (σε μορφή MM/DD/YYYY), την τιμή μήλων ανά κιλό σε ευρώ (PEAR_PRICE), την τιμή υγραερίου ανά λίτρο σε ευρώ (AUTOGAS_PRICE) και την ποσότητα παραγωγής μήλων σε χιλιάδες τόνους (PRODUCTION_QUANTITY).

Πίνακας 6. Παράδειγμα μορφής δεδομένων για τις τιμές μήλων, υγραερίου κίνησης και της ποσότητας παραγωγής από το 2007 έως το 2022.

DATE	APPLE_PRICE	AUTOGAS_PRICE	PRODUCTION_QUANTITY
1/1/2007	0,561	1,099	0,210806852
1/2/2007	0,562	1,082	0,210431751
---	---	---	---
1/11/2022	0,764	1,81	0,192881091
1/12/2022	0,802	1,727	0,183742087

Πίνακας 7. Παράδειγμα μορφής δεδομένων για τις τιμές αχλαδιών, υγραερίου κίνησης και της ποσότητας παραγωγής από το 2007 έως το 2022.

DATE	PEAR_PRICE	AUTOGAS_PRICE	PRODUCTION_QUANTITY
1/1/2007	0,9	1,099	0,735787602
1/2/2007	0,92	1,082	0,719792219
---	---	---	---
1/11/2022	1,683	1,81	0,480195478
1/5/2007	1,664	1,727	0,490165278

3.2 Προεπεξεργασία των Δεδομένων

Πριν την εκτέλεση των αναλύσεων, πραγματοποιήθηκε προεπεξεργασία των δεδομένων για να διασφαλιστεί η ποιότητά τους και η καταλληλότητά τους για τις προβλεπόμενες αναλύσεις. Η διαδικασία προεπεξεργασίας περιλάμβανε αρκετά στάδια για την προετοιμασία των δεδομένων.

Αρχικά, πραγματοποιήθηκε έλεγχος για την ύπαρξη ελλিপών τιμών σε όλα τα πεδία των δεδομένων. Από τον έλεγχο διαπιστώθηκε ότι δεν υπήρχαν κενά δεδομένα στις μεταβλητές των μήλων και των αχλαδιών, γεγονός που επιβεβαίωσε την πληρότητα των δεδομένων για την εξεταζόμενη περίοδο. Επιπλέον, διενεργήθηκε έλεγχος για την ύπαρξη εξωγενών τιμών (outliers) στις μεταβλητές **APPLE_PRICE**, **PEAR_PRICE** και **AUTOGAS_PRICE**, **PRODUCTION_QUANTITY**. Ωστόσο, δεν εντοπίστηκαν τιμές που να αποκλίνουν σημαντικά από τις τυπικές διακυμάνσεις των δεδομένων.

Επίσης, οι ημερομηνίες καταγραφής μετατράπηκαν σε ενιαία μορφή (ISO format) για να εξασφαλιστεί η συμβατότητα μεταξύ των δύο συνόλων δεδομένων και να διευκολυνθεί η ανάλυση των χρονικών σειρών. Αυτό ήταν απαραίτητο για την εξάλειψη πιθανών διαφορών στη μορφή της ημερομηνίας μεταξύ των δύο συνόλων δεδομένων.

Η προεπεξεργασία των δεδομένων αποτέλεσε κρίσιμο βήμα για την εξασφάλιση της ακρίβειας και της αξιοπιστίας των αναλύσεων που ακολουθούν. Με τη διαδικασία αυτή, τα δεδομένα κατέστησαν έτοιμα για την εφαρμογή προβλεπτικών μοντέλων και στατιστικών αναλύσεων που θα παρουσιαστούν στα επόμενα κεφάλαια.

3.3 Οπτικοποίηση των Δεδομένων

Η οπτικοποίηση δεδομένων αποτελεί βασικό κομμάτι της ανάλυσης δεδομένων, καθώς επιτρέπει την άμεση και εύκολη κατανόηση σύνθετων πληροφοριών. Σε έναν κόσμο όπου η ποσότητα των διαθέσιμων δεδομένων αυξάνεται ραγδαία, η ανάγκη για γρήγορη και αποτελεσματική κατανόηση των δεδομένων είναι πιο επείγουσα από ποτέ. Οι άνθρωποι έχουν

την τάση να κατανοούν και να επεξεργάζονται καλύτερα την οπτική πληροφορία συγκριτικά με τα αριθμητικά ή τα κειμενικά δεδομένα. Με την σωστή οπτικοποίηση, μπορούμε να προβάλουμε τάσεις, μοτίβα, ανωμαλίες και συσχετίσεις που δεν είναι άμεσα εμφανείς όταν τα δεδομένα παρουσιάζονται σε ακατέργαστη μορφή.

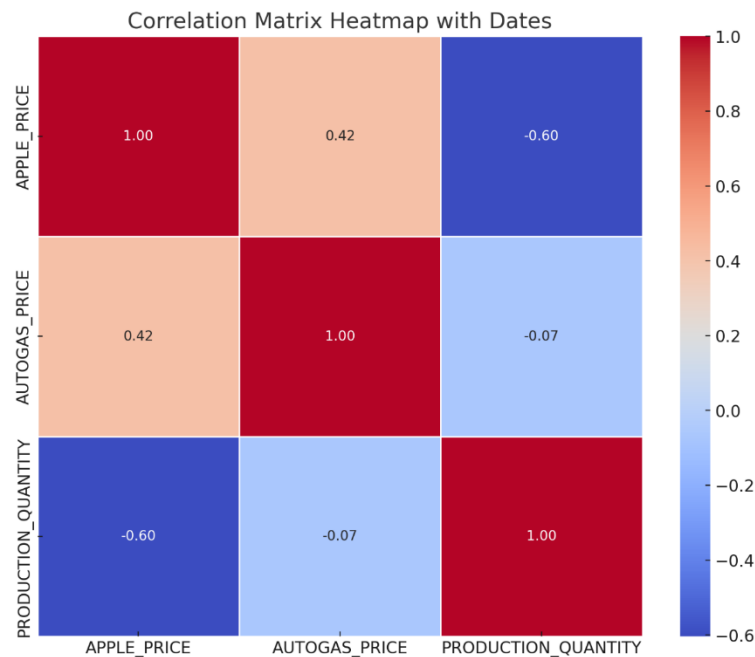
Η οπτικοποίηση δεδομένων διευκολύνει την κατανόηση της πολυπλοκότητας των δεδομένων με έναν πιο προσιτό και διαισθητικό τρόπο. Μπορεί να χρησιμοποιηθεί για την επιβεβαίωση υποθέσεων, τη διερεύνηση νέων υποθέσεων, και την επικοινωνία των ευρημάτων προς ένα ευρύτερο κοινό, συμπεριλαμβανομένων αυτών που μπορεί να μην έχουν εξειδικευμένες γνώσεις στατιστικής ή ανάλυσης δεδομένων. Οι αποτελεσματικές οπτικοποιήσεις μπορούν να αποκαλύψουν τις αφανείς σχέσεις μεταξύ των μεταβλητών, να προσδιορίσουν κατηγορίες δεδομένων, να δείξουν τη συχνότητα εμφάνισης ορισμένων φαινομένων, και να προβλέψουν μελλοντικές τάσεις.

Στην παρούσα πτυχιακή εργασία, η οπτικοποίηση δεδομένων έχει ιδιαίτερη σημασία, καθώς μέσω της οπτικοποίησης, μπορούμε να εξετάσουμε τη διακύμανση των τιμών σε βάθος χρόνου, να αναγνωρίσουμε μοτίβα και τάσεις που σχετίζονται με εποχιακές μεταβολές, κ.α.

Η επιλογή των κατάλληλων τεχνικών και εργαλείων για την οπτικοποίηση δεδομένων είναι κρίσιμη για την επιτυχία της ανάλυσης. Καθώς προχωράμε σε περαιτέρω ανάλυση, θα εξεταστούν οι μέθοδοι και οι τεχνικές που χρησιμοποιήθηκαν για τη δημιουργία γραφημάτων και διαγραμμάτων, με σκοπό να προσφέρουν σαφή και ουσιαστική κατανόηση των δεδομένων και των ευρημάτων που προέκυψαν από αυτά.

3.3.1 Διαγράμματα Μήτρας (Matrix Diagram)

Ο θερμικός χάρτης συσχέτισης αναφέρει τους συντελεστές συσχέτισης Pearson μεταξύ τριών βασικών χαρακτηριστικών του dataset: **APPLE_PRICE**, **AUTOGAS_PRICE** και **PRODUCTION_QUANTITY**.

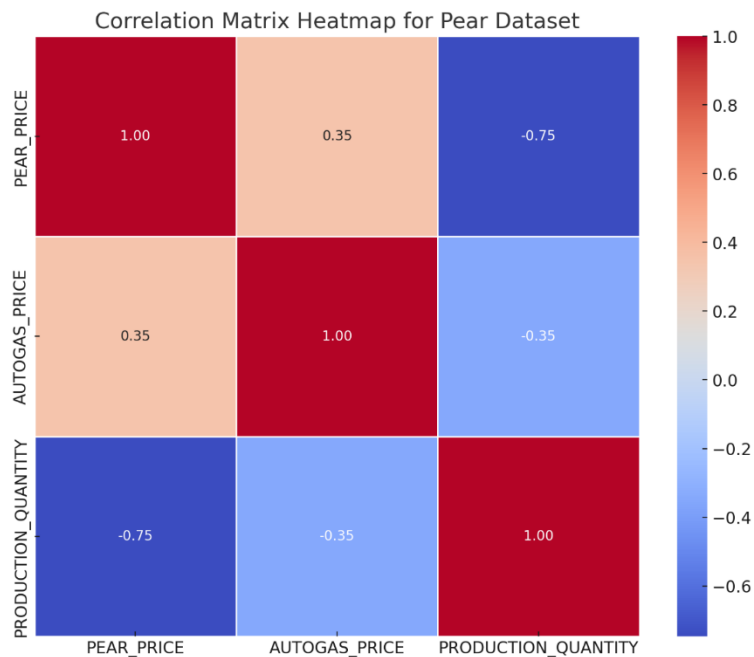


Εικόνα 4. Διάγραμμα Μήτρας για Μήλα

Με βάση τον παραπάνω θερμικό χάρτη (Εικόνα 4.) παρατηρούμε τα εξής:

- **APPLE_PRICE & AUTOGAS_PRICE:** Παρουσιάζουν μια ήπια θετική συσχέτιση (0.42), υποδεικνύοντας ότι οι τιμές των μήλων τείνουν να αυξάνονται, όταν αυξάνονται οι τιμές του υγραερίου.
- **APPLE_PRICE & PRODUCTION_QUANTITY:** Παρουσιάζουν μια ήπια προς ισχυρή αρνητική συσχέτιση (-0.60). Ωστόσο παρόλο που ο συντελεστής είναι αρνητικός, ουσιαστικά αυτό σημαίνει ότι όταν η παραγωγή αυξάνεται, η τιμή των μήλων τείνει να μειώνεται. Αυτό εν ολίγοις οφείλεται στην αυξημένη προσφορά που πιέζει τις τιμές προς τα κάτω. Οπότε, αν και η συσχέτιση φαίνεται αρνητική, στην πραγματικότητα αντικατοπτρίζει μια αντιστρόφως ανάλογη σχέση όπου η ποσότητα παραγωγής επηρεάζει άμεσα την τιμή του προϊόντος.
- **AUTOGAS_PRICE & PRODUCTION_QUANTITY:** Παρουσιάζουν πολύ ασθενής αρνητική συσχέτιση (-0.07) δείχνοντάς ότι δεν υπάρχει ουσιαστική σχέση μεταξύ της τιμής του υγραερίου και της ποσότητας παραγωγής μήλων.

Ο θερμικός χάρτης συσχέτισης παρουσιάζει τους συντελεστές συσχέτισης Pearson μεταξύ τριών βασικών χαρακτηριστικών του dataset: **PEAR_PRICE** (τιμή μήλων), **AUTOGAS_PRICE** (τιμή υγραερίου) και **PRODUCTION_QUANTITY** (ποσότητα παραγωγής μήλων).



Εικόνα 5. Διάγραμμα Μήτρας για Αχλαδιά

Με βάση τον παραπάνω θερμικό χάρτη (Εικόνα 5) παρατηρούμε τα εξής:

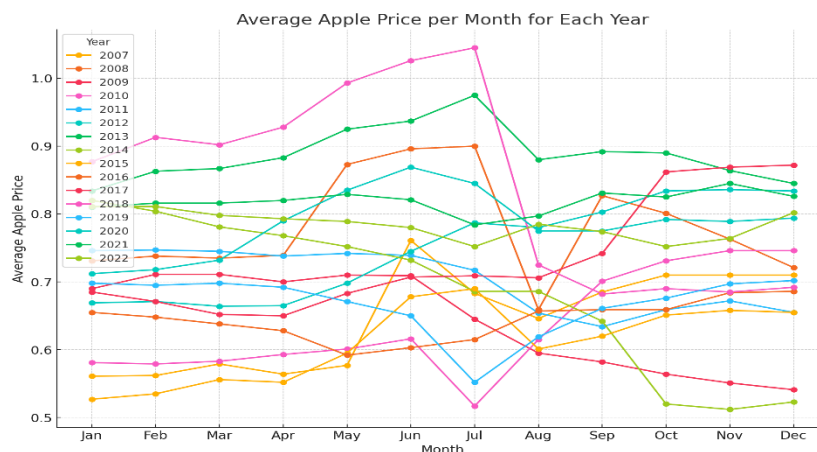
- **PEAR_PRICE & AUTOGAS_PRICE:** Παρουσιάζουν μια ήπια θετική συσχέτιση (0.35), υποδεικνύοντας ότι οι τιμές των αχλαδιών τείνουν να αυξάνονται, όταν αυξάνονται οι τιμές του υγραερίου.
- **PEAR_PRICE & PRODUCTION_QUANTITY:** Παρουσιάζουν μια ήπια προς ισχυρή αρνητική συσχέτιση (-0.75). Ωστόσο παρόλο που ο συντελεστής είναι αρνητικός, ουσιαστικά αυτό σημαίνει ότι όταν η παραγωγή αυξάνεται, η τιμή των μήλων τείνει να μειώνεται. Αυτό εν ολίγοις οφείλεται στην αυξημένη προσφορά που πιέζει τις τιμές προς τα κάτω. Οπότε, αν και η συσχέτιση φαίνεται αρνητική, στην πραγματικότητα αντικατοπτρίζει μια αντιστρόφως ανάλογη σχέση όπου η ποσότητα παραγωγής επηρεάζει άμεσα την τιμή του προϊόντος.
- **AUTOGAS_PRICE & PRODUCTION_QUANTITY:** Παρουσιάζουν πολύ ασθενής αρνητική συσχέτιση (-0.35) δείχνοντάς ότι δεν υπάρχει ουσιαστική σχέση μεταξύ της τιμής του υγραερίου και της ποσότητας παραγωγής μήλων

3.3.2 Γραφικά Διαγράμματα Μήλων

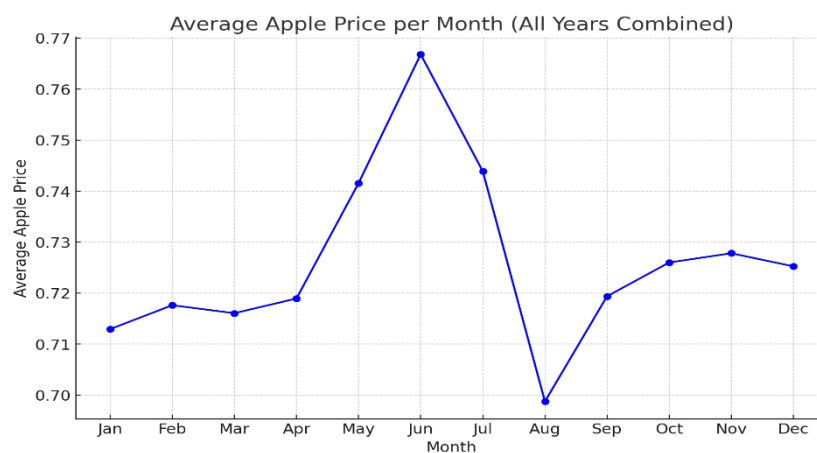
Το γραφικό διάγραμμα της **Εικόνας 6**, απεικονίζει ένα διάγραμμα πολλαπλών σειρών που παρουσιάζει τις μέσες τιμές των μήλων για κάθε μήνα ανεξάρτητα για όλα τα έτη που συμπεριλαμβάνονται στο Dataset (DATASET_APPLE_AUTOGAS_PRICE_FINAL). Κάθε γραμμή στο

διάγραμμα εκπροσωπεί τα δεδομένα για ένα συγκεκριμένο έτος επιτρέποντας με αυτόν τον τρόπο την σύγκριση των τάσεων των μήλων μέσο διαφορετικών χρονολογικών σημείων.

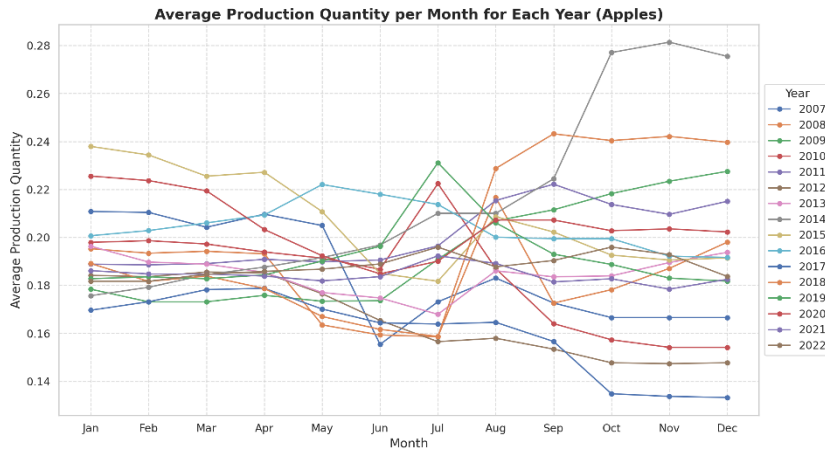
Το γραφικό διάγραμμα της **Εικόνας 7**. απεικονίζει ένα διάγραμμα που παρουσιάζει τη μέση τιμή των μήλων για κάθε μήνα έχοντας συνδυάσει τα δεδομένα από όλα τα έτη που συμπεριλαμβάνονται στο Dataset (DATASET_APPLE_AUTOGAS_PRICE_FINAL). Κάθε σημείο στο διάγραμμα εκπροσωπεί την μέση τιμή για έναν συγκεκριμένο μήνα, και όπως διαπιστώνετε από το διάγραμμα παρατηρείται μια σχετική σταθερότητα στις τιμές των μήλων κατά τους πρώτους χειμερινούς μήνες του έτους, ωστόσο από τον Απρίλιο οι τιμές αρχίζουν να αυξάνονται έχοντας κορυφωθεί τον Ιούνιο. Αυτή η απότομη αύξηση μπορεί να δικαιολογηθεί από την αυξημένη ζήτηση και την μειωμένη παραγωγή τους κατά τους καλοκαιρινούς μήνες που παρατηρείται στα γραφήματα της **Εικόνας 8**. και **Εικόνας 9**.



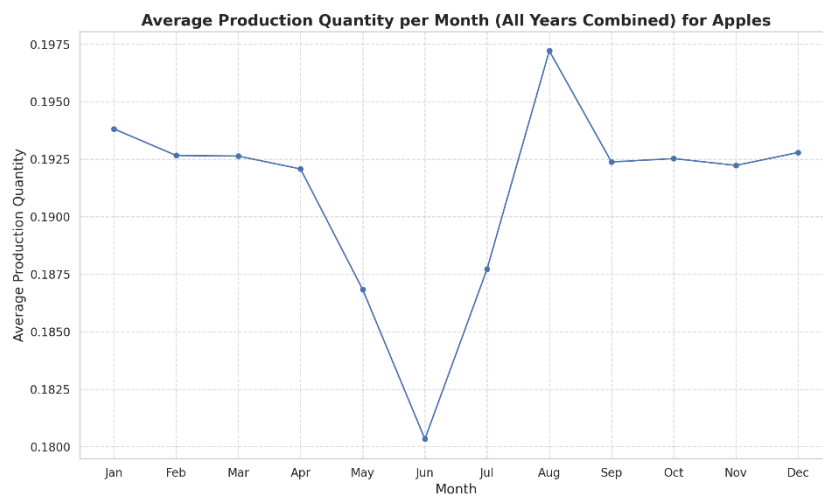
Εικόνα 6. Μέση τιμή Μήλων ανά μήνα για κάθε χρόνο



Εικόνα 7. Μέση τιμή Μήλων ανά μήνα (Όλα τα χρόνια μαζί)



Εικόνα 8. Μέση ποσότητα παραγωγής Μήλων ανά μήνα για κάθε χρόνο



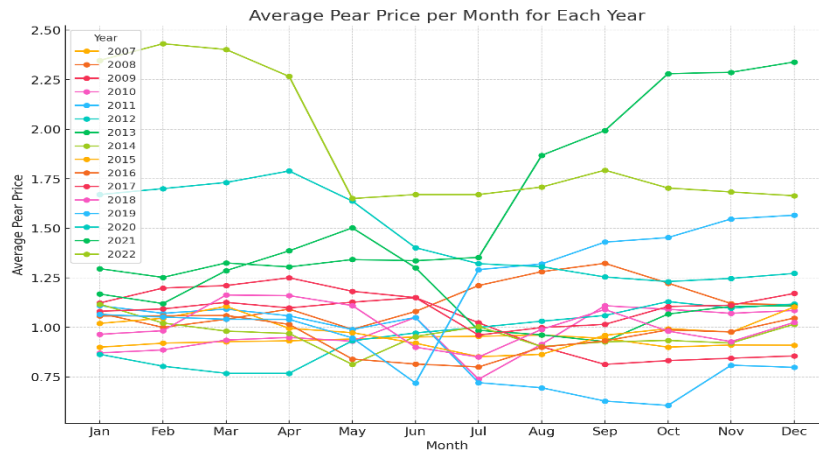
Εικόνα 9. Μέση ποσότητα παραγωγής Μήλων ανά μήνα (Όλα τα χρόνια μαζί)

3.3.3 Γραφικά Διαγράμματα Αχλαδιών

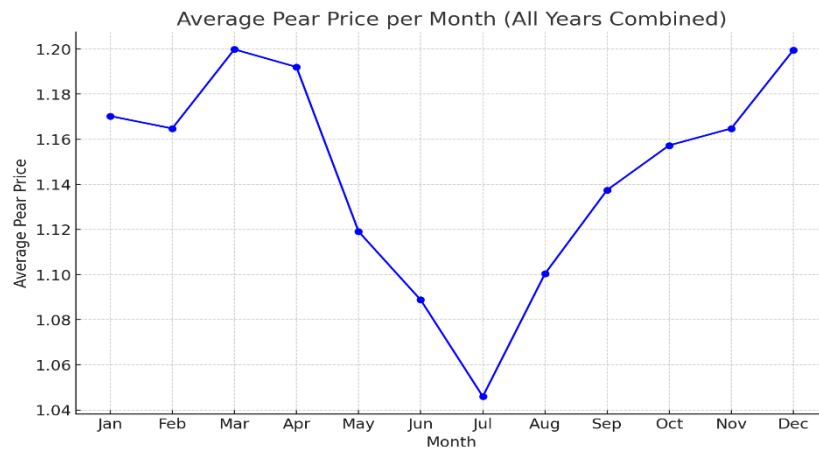
Αντίστοιχα το γραφικό διάγραμμα της **Εικόνας 10**, απεικονίζει ένα διάγραμμα πολλαπλών σειρών που παρουσιάζει τις μέσες τιμές των αχλαδιών για κάθε μήνα ανεξάρτητα για όλα τα έτη που συμπεριλαμβάνονται στο Dataset (DATASET_PEAR_AUTOGAS_PRICE_FINAL). Κάθε γραμμή στο διάγραμμα εκπροσωπεί τα δεδομένα για ένα συγκεκριμένο έτος επιτρέποντας με αυτόν τον τρόπο την σύγκριση των τάσεων των αχλαδιών μέσω διαφορετικών χρονολογικών σημείων.

Το γραφικό διάγραμμα της **Εικόνας 11**, απεικονίζει ένα διάγραμμα που παρουσιάζει τη μέση τιμή των αχλαδιών για κάθε μήνα έχοντας συνδυάσει τα δεδομένα από όλα τα έτη που συμπεριλαμβάνονται στο Dataset (DATASET_PEAR_AUTOGAS_PRICE_FINAL). Κάθε σημείο στο διάγραμμα εκπροσωπεί την μέση τιμή για έναν συγκεκριμένο μήνα, και όπως διαπιστώνετε από

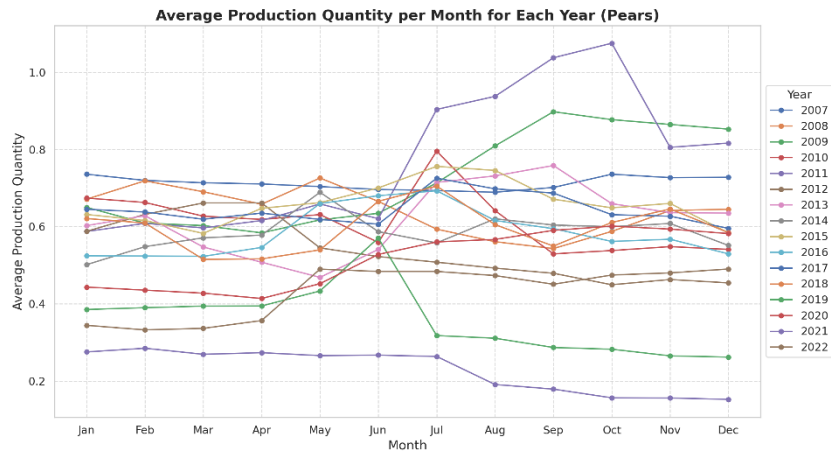
το διάγραμμα παρατηρείται μια σχετική σταθερότητα στις τιμές των μήλων κατά τους πρώτους χειμερινούς μήνες του έτους, ωστόσο από τον Απρίλιο οι τιμές αρχίζουν να μειώνονται έχοντας κατακρημνιστεί τον Ιούλιο. Αυτή η απότομη μείωση μπορεί να δικαιολογηθεί από την μειωμένη ζήτηση και την αυξημένη παραγωγή τους κατά τους καλοκαιρινούς μήνες που παρατηρείται στα γραφήματα της **Εικόνας 12.** και **Εικόνας 13.**



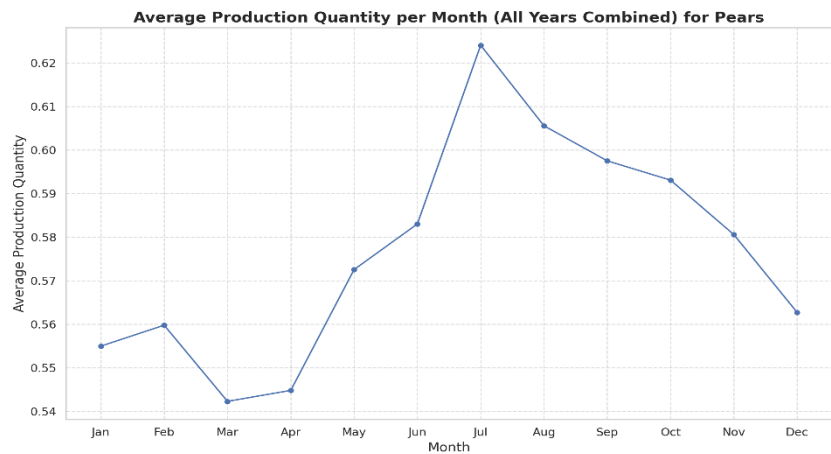
Εικόνα 10. Μέση τιμή Αχλαδιών ανά μήνα για κάθε χρόνο



Εικόνα 11. Μέση τιμή Αχλαδιών ανά μήνα (Όλα τα χρόνια μαζί)



Εικόνα 12. Μέση ποσότητα παραγωγής Αχλαδιών ανά μήνα για κάθε χρόνο



Εικόνα 13. Μέση ποσότητα παραγωγής Αχλαδιών ανά μήνα (Όλα τα χρόνια μαζί)

4. Αλγόριθμοι Πρόβλεψης

Στο συγκεκριμένο κεφάλαιο παρουσιάζεται η ανάπτυξη των μοντέλων πρόβλεψης που χρησιμοποιήθηκαν, δηλαδή των μοντέλων SARIMA, LSTM και Prophet τα οποία αναλύθηκαν εκτενέστερα στο Κεφάλαιο 2. με σκοπό να επιτευχθεί η εκτίμηση μηνιαίων τιμών των μήλων και των αχλαδιών στην Ιταλία για το χρονικό διάστημα 01.01.2023 έως 12.01.2023.

4.1 Εργαλεία

4.1.1 Python

Στην παρούσα έρευνα όλες οι μελέτες, οι στατιστικές αναλύσεις και τα υπολογιστικά μοντέλα υλοποιήθηκαν με την γλώσσα προγραμματισμού Python. Η Python επιλέχθηκε καθώς είναι μια από τις πιο εύχρηστες γλώσσες στον τομέα της επιστήμης των δεδομένων και της στατιστικής ανάλυσης, παρέχοντας ένα πολύ εύρη οικοσύστημα βιβλιοθηκών που διευκολύνουν την επεξεργασία, την ανάλυση και την οπτικοποίηση των δεδομένων.

4.1.2 Jupiter Notebook

Το Jupiter Notebook επιλέχθηκε διότι είναι ένα από τα ισχυρότερα και δημοφιλέστερα εργαλεία για την ανάλυση δεδομένων, παρέχοντας την δυνατότητα στους χρήστες να χωρίζουν τον κώδικα τους σε ανεξάρτητα κομμάτια και να οπτικοποιούν τα αποτελέσματα διαδραστικά.

4.2 LSTM Model

Ακολουθεί αναλυτική περιγραφή των βημάτων που ακολουθήθηκαν, των παραμέτρων που χρησιμοποιήθηκαν, καθώς και οι λόγοι επιλογής τους με σκοπό την πρόβλεψη των τιμών των μήλων.

4.2.1 Εισαγωγή Βιβλιοθηκών και Προεπεξεργασία Δεδομένων

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout, Input
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.regularizers import l2
```

Εικόνα 14. LSTM_1

- **NumPy** για τους αριθμητικούς υπολογισμούς και την διαχείριση πολυδιάστατων πινάκων.
- **Matplotlib** για τη δημιουργία γραφημάτων και απεικονίσεων.
- **Scikit-Learn's MinMaxScaler** για την κανονικοποίηση των δεδομένων στην κλίμακα [0, 1].
- **TensorFlow Keras API** για την κατασκευή και εκπαίδευση του νευρωνικού δικτύου LSTM.

4.2.2 Φόρτωση και Προετοιμασία Δεδομένων

```
filepath = "DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx"
data = pd.read_excel(filepath)
data.columns = data.columns.str.strip()
data['DATE'] = pd.to_datetime(data['DATE'], dayfirst=True)
data.set_index('DATE', inplace=True)
```

Εικόνα 15. LSTM_2

Γίνεται φόρτωση των δεδομένων του Dataset "DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx". Γίνεται καθαρισμός των στηλών του συνόλου δεδομένων από πιθανά κενά διαστήματα, και η στήλη ημερομηνιών μετατρέπεται σε μορφή datetime, ορίζοντας την ως δείκτη (index) του DataFrame. Αυτή η διαδικασία επιτρέπει την άμεση πρόσβαση και ανάλυση των δεδομένων με βάση τις χρονοσειρές.

4.2.3 Υπολογισμός Διαφορών και Κανονικοποίηση

```
data['Diff_APPLE_PRICE'] = data['APPLE_PRICE'].diff()
data.dropna(inplace=True)

scaler = MinMaxScaler()
data['Scaled_Diff_APPLE_PRICE'] = scaler.fit_transform(data[['Diff_APPLE_PRICE']])
```

Εικόνα 16. LSTM_3

Υπολογίζονται οι διαφορές τιμών για να μοντελοποιηθεί η διακύμανση στις τιμές των μήλων από μήνα σε μήνα και στη συνέχεια εφαρμόζεται η κανονικοποίηση χρησιμοποιώντας το MinMaxScaler. Επιπλέον, για την βελτιστοποίηση της διαδικασίας εκπαίδευσης του νευρωνικού δικτύου είναι απαραίτητη η κλιμάκωση των τιμών σε ένα εύρος [0, 1], καθώς έτσι αποφεύγεται η προκατάληψη προς μεγαλύτερες τιμές.

4.2.4 Δημιουργία Συνόλου Δεδομένων Εκπαίδευσης και Δοκιμής

```
def create_dataset(X, look_back=1):
    Xs, ys = [], []
    for i in range(len(X) - look_back):
        v = X[i:(i + look_back)]
        Xs.append(v)
        ys.append(X[i + look_back])
    return np.array(Xs), np.array(ys)
```

Εικόνα 17. LSTM_4

Η συνάρτηση create_dataset δημιουργεί το σύνολο εισόδων και εξόδων που θα χρησιμοποιηθούν για την εκπαίδευση του LSTM μοντέλου, και επιπλέον χρησιμοποιείται ένα παράθυρο οπισθοδρόμησης (look-back window) 60 μηνών (δηλαδή πέντε (5) χρόνων), καθώς με αυτό τον τρόπο το μοντέλο εκπαιδεύεται για να προβλέψει την τιμή στον επόμενο μήνα με μεγαλύτερη βάση στις τιμές των προηγούμενων 60 μηνών.

4.2.5 Ρύθμιση Παραμέτρων και Διαχωρισμός Συνόλου Δεδομένων

```
train_size = int(len(X) * 0.60)
X_train, X_test = X[:train_size], X[train_size:]
y_train, y_test = y[:train_size], y[train_size:]
```

Εικόνα 18. LSTM_5

Οι υψηλότερες παράμετροι (hyperparameters) που χρησιμοποιούνται για την κατασκευή του μοντέλου είναι οι εξής:

- look_back = 60: Η παράμετρος καθορίζει το μήκος της ακολουθίας εισόδου για το μοντέλο.
- train_size = 0.60: Καθορίζει το 60% των δεδομένων ως εκπαιδευτικό σύνολο και το υπόλοιπο 40% ως δοκιμαστικό. Η συγκεκριμένη αναλογία δίνει την δυνατότητα στο μοντέλο να εκπαιδευτεί με ένα αντιπροσωπευτικό δείγμα δεδομένων, ενώ αφήνει αρκετά δεδομένα για την αξιολόγηση της απόδοσης του.

4.2.6 Κατασκευή και Εκπαίδευση του Μοντέλου LSTM

```
model = Sequential([
    Input(shape=(look_back, 1)),
    LSTM(150, return_sequences=True, kernel_regularizer=l2(0.01)),
    Dropout(0.2),
    LSTM(150, return_sequences=True, kernel_regularizer=l2(0.01)),
    Dropout(0.2),
    LSTM(150, kernel_regularizer=l2(0.01)),
    Dropout(0.2),
    Dense(1)
])
```

Εικόνα 19. LSTM_6

Γίνεται δημιουργία τριών επαναληπτικών επιπέδων, με 150 νευρώνες το καθένα προκειμένου να μην να μπορέσει το δίκτυο να μάθει πολύπλοκα μοτίβα, αλλά όχι τόσο μεγάλα ώστε να αυξηθεί υπερβολικά η πολυπλοκότητα του καθώς κάτι τέτοιο θα μπορούσε να οδηγήσει σε υπερπροσαρμογή. Επίσης γίνεται χρήση κανονικοποίησης l2 με σκοπό να αποφευχθεί η οποιαδήποτε υπερπροσαρμογή και τέλος, χρησιμοποιείται το Dropout το οποίο τυχαία απενεργοποιεί το 20% (0.2) των νευρώνων σε κάθε στρώμα κατά τη διάρκεια τη εκπαίδευσης με σκοπό επίσης την αποφυγή υπερπροσαρμογής.

```
model.compile(optimizer='adam', loss='mean_squared_error')

early_stopping = EarlyStopping(monitor='val_loss', patience=100, restore_best_weights=True)
model.fit(X_train, y_train, sample_weight=weights_train, epochs=150, batch_size=50, validation_split=0.1, callbacks=[early_stopping], verbose=1)
```

Εικόνα 20. LSTM_7

Γίνεται εκπαίδευση του μοντέλου με χρήση του βελτιστοποιητή adam ο οποίος είναι ένας από τους πιο δημοφιλείς αλγόριθμους βελτιστοποιήσεις λόγο της αποδοτικότητας του και της ικανότητας του να προσαρμόζει τις παραμέτρους του ρυθμού εκμάθησης κατά την διάρκεια εκπαίδευσης του. Γίνεται επίσης χρήση και της συνάρτησης απώλειας mean_squared_error που είναι ιδανική για προβλήματα παλινδρόμησης. Τέλος, χρησιμοποιείται το EarlyStopping με υπομονή 100 εποχών για να αποτραπεί η υπερπροσαρμογή του μοντέλου.

4.2.7 Πρόβλεψη Μελλοντικών Τιμών

```
for i in range(12):
    next_pred = model.predict(input_seq)
    noise = np.random.normal(0, 0.01)
    next_pred = next_pred + noise
    predictions_diff_future.append scaler.inverse_transform(next_pred).ravel()[0])
    next_input = next_pred.reshape(1, 1, 1)
    input_seq = np.append(input_seq[:, 1:, :], next_input, axis=1)
```

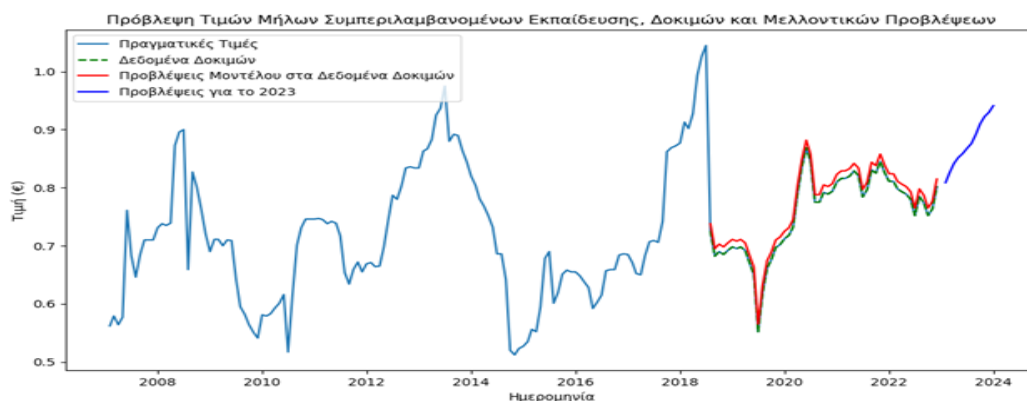
Εικόνα 21. LSTM_8

Οι προβλέψεις για τους επόμενους 12 μήνες βασίζονται στις τελευταίες γνωστές τιμές και το μοντέλο εκπαιδεύεται να προβλέπει τις διαφορές τιμών.

```
plt.figure(figsize=(12, 6))
plt.plot(data.index, data['APPLE_PRICE'], label='Πραγματικές Τιμές')
plt.plot(test_dates, data['APPLE_PRICE'][train_size + look_back:], 'g--', label='Δεδομένα Δοκιμών')
plt.plot(test_dates, data['APPLE_PRICE'][train_size + look_back:] + scaler.inverse_transform(model.predict(X_test)).ravel(), 'r-', label='Προβλέψεις Μοντέλου στα Δεδομένα Δοκιμών')
plt.plot(months_future, predictions_future, 'b-', label='Προβλέψεις για το 2023')
plt.title('Πρόβλεψη Τιμών Μήλων Συμπεριλαμβανομένων Εκπαίδευσης, Δοκιμών και Μελλοντικών Προβλέσεων')
plt.xlabel('Ημερομηνία')
plt.ylabel('Τιμή (€)')
plt.legend()
plt.show()
```

Τέλος τα αποτελέσματα των προβλέψεων απεικονίζονται έχοντας δημιουργήσει γραφήματα για την ανάλυση της ακρίβειας του μοντέλου και τη σύγκριση με τα πραγματικά δεδομένα.

4.2.8 Αποτελέσματα Μοντέλου



Εικόνα 23. LSTM_Prediction_Results

Πίνακας 8. Αποτελέσματα προβλέψεων για τους επόμενους 12 μήνες με χρήση LSTM

ΜΗΝΕΣ	ΠΡΟΒΕΠΟΜΕΝΗ ΤΙΜΗ
January 2023	0,81€
February 2023	0,83€
March 2023	0,84€
April 2023	0,85€
May 2023	0,86€
June 2023	0,87€
July 2023	0,88€
August 2023	0,89€
September 2023	0,91€
October 2023	0,92€
November 2023	0,93€
December 2023	0,94€
MSE	0,0004
MAE	0,0203

4.3 SARIMA Model

Ακολουθεί αναλυτική περιγραφή των βημάτων που ακολουθήθηκαν, των παραμέτρων που χρησιμοποιήθηκαν, καθώς και οι λόγοι επιλογής τους με σκοπό την πρόβλεψη των τιμών των μήλων.

4.3.1 Εισαγωγή Βιβλιοθηκών και Προεπεξεργασία Δεδομένων

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.model_selection import TimeSeriesSplit
from sklearn.metrics import mean_squared_error
```

Εικόνα 24. SARIMA_1

- **NumPy** για αριθμητικούς υπολογισμούς και την διαχείριση πολυδιάστατων πινάκων.
- **Pandas** για την ανάγνωση και επεξεργασία δεδομένων.
- **Matplotlib** για τη δημιουργία γραφημάτων και απεικονίσεων.
- **Statsmodels** για στατιστική μοντελοποίηση, και πιο συγκεκριμένα για την χρήση του SARIMAX.
- **Scikit-Learn TimeSeriesSplit** για τη μέθοδο διασταυρούμενης επικύρωσης (cross-validation) σε χρονοσειρές.

4.3.2 Φόρτωση και Προετοιμασία Δεδομένων

```
filepath = 'DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx'
data = pd.read_excel(filepath)

data.columns = data.columns.str.strip()
data['DATE'] = pd.to_datetime(data['DATE'], dayfirst=True)
data.set_index('DATE', inplace=True)
data.index = data.index.to_period('M')
```

Εικόνα 25. SARIMA_2

Αρχικά γίνεται φόρτωση των δεδομένων του Dataset “DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx”. Γίνεται καθαρισμός των στηλών του συνόλου δεδομένων από πιθανά κενά διαστήματα και η στήλη των ημερομηνιών μετατρέπεται σε μορφή datetime, ορίζοντας την ως δείκτης (index) του DataFrame. Στη συνέχεια, ο δείκτης μετατρέπεται σε χρονική περίοδο μήνα, διευκολύνοντας έτσι την εποχιακή ανάλυση των δεδομένων.

4.3.3 Διαχωρισμός Δεδομένων, Διαδικασία Cross-Validation και Ρυθμίσεις Παραμέτρων Μοντέλου SARIMAX

```

tscv = TimeSeriesSplit(n_splits=5)
mse_scores = []

for train_index, test_index in tscv.split(data):
    train, test = data.iloc[train_index], data.iloc[test_index]

    model = SARIMAX(train['APPLE_PRICE'],
                    order=(0, 1, 0),
                    seasonal_order=(5, 1, 0, 12),
                    enforce_stationarity=False,
                    enforce_invertibility=False)
    model_fit = model.fit(dispatch=False)

    predictions = model_fit.forecast(steps=len(test))
    mse = mean_squared_error(test['APPLE_PRICE'], predictions)
    mse_scores.append(mse)

mean_mse_score = np.mean(mse_scores)
print(f'Mean MSE Score from Time Series Cross-Validation: {mean_mse_score}')

```

Εικόνα 26. SARIMA_3

Αρχικά πραγματοποιείται διαχωρισμός των δεδομένων και εφαρμόζεται η διαδικασία διασταυρούμενης επικύρωσης (cross-validation) με τη μέθοδο TimeSeriesSplit ειδικά προσαρμοσμένη για χρονοσειρές, με 5 υποδιαιρέσεις ($n_splits=5$) καθώς προσφέρουν αρκετά διαφορετικά υποσύνολα προκειμένου να μειωθεί ο κίνδυνος υπερπροσαρμογής στα δεδομένα. Επιπλέον αυτή η προσέγγιση επιτρέπει την προσθετική αύξηση του εκπαιδευομένου συνόλου δεδομένων και την αξιολόγηση του μοντέλου σε διαδοχικά υποσύνολα του αρχικού συνόλου δεδομένων.

Επιπλέον γίνεται διαμόρφωση του μοντέλου SARIMAX με τις ακόλουθες παραμέτρους:

- **Order (p, d, q):** Καθορίζονται οι αυτοπαλίνδρομοι όροι (p), η διαφορική τάξη για σταθεροποίηση (d), και οι όροι κινούμενου μέσου (q). Για το παρόν μοντέλο, οι τιμές είναι (0, 1, 0), που σημαίνει ότι το μοντέλο έχει μία διαφορική τάξη για την αντιμετώπιση της μη στασιμότητας, χωρίς αυτοπαλινδρομικούς ή όρους κινούμενου μέσου.
- **Seasonal Order (P, D, Q, s):** Καθορίζονται οι εποχιακοί αυτοπαλίνδρομοι όροι (P), οι εποχιακές διαφορές (D), οι εποχιακοί όροι κινούμενου μέσου (Q) και η εποχιακή περίοδος (s). Στο παρόν μοντέλο, οι τιμές είναι (5, 1, 0, 12), δηλαδή εποχιακό μοντέλο με πέντε αυτοπαλίνδρομους όρους, μία εποχιακή διαφορά, χωρίς εποχιακούς όρους κινούμενου μέσου και περίοδο 12 μηνών, που αντιπροσωπεύει εποχιακές διακυμάνσεις σε ετήσια βάση.

- **enforce_stationarity** και **enforce_invertibility**: Οι παράμετροι αυτές τέθηκαν σε False για να επιτραπεί στο μοντέλο να εξετάσει ευρύτερη περιοχή λύσεων κατά την εκτίμηση των παραμέτρων, ακόμη και αν αυτές δεν είναι ακριβώς στατικές ή αναστρέψιμες.

4.3.4 Εκπαίδευση του Τελικού Μοντέλου και Πρόβλεψη

```
sarima_model = SARIMAX(data['APPLE_PRICE'],
                        order=(0, 1, 0),
                        seasonal_order=(5, 1, 0, 12),
                        enforce_stationarity=False,
                        enforce_invertibility=False)
sarima_fit = sarima_model.fit(dispatch=False)

future_steps = 12
sarima_forecast = sarima_fit.get_forecast(steps=future_steps)
sarima_forecast_index = pd.date_range(start=data.index[-1].to_timestamp(), periods=future_steps + 1, freq='M')[1:]

predicted_prices = sarima_forecast.predicted_mean
```

Εικόνα 27. SARIMA_4

Γίνεται εκπαίδευση του μοντέλου SARIMA σε ολόκληρο το σύνολο δεδομένων για την τελική πρόβλεψη των επόμενων 12 μηνών. Οι προβλέψεις αυτές απεικονίζονται για οπτική κατανόηση των μελλοντικών τιμών και η διαδικασία αυτή επιτρέπει την πρόβλεψη μελλοντικών τιμών και την οπτική παρουσίαση των προβλέψεων για την καλύτερη κατανόηση των τάσεων στις τιμές των μήλων.

4.3.5 Εκτύπωση Προβλέψεων για τους Επόμενους 12 Μήνες

```
months = ['January', 'February', 'March', 'April', 'May', 'June',
          'July', 'August', 'September', 'October', 'November', 'December']

print("Προβλέψεις για τους επόμενους 12 μήνες:")

for i, (date, price) in enumerate(zip(sarima_forecast_index, predicted_prices)):
    month_name = months[date.month - 1]
    year = date.year
    print(f"{month_name} {year}: Προβλεπόμενη τιμή {price:.2f}€")
```

Εικόνα 28. SARIMA_5

Γίνεται εκτύπωση των προβλέψεων για τους επόμενους 12 μήνες σε κατανοητή μορφή και δημιουργείται μια λίστα με τα ονόματα των μηνών όπου με τη βοήθεια ενός βρόχου for, οι προβλέψεις για τους επόμενους 12 μήνες εκτυπώνονται μαζί με το αντίστοιχο όνομα του μήνα και το έτος.

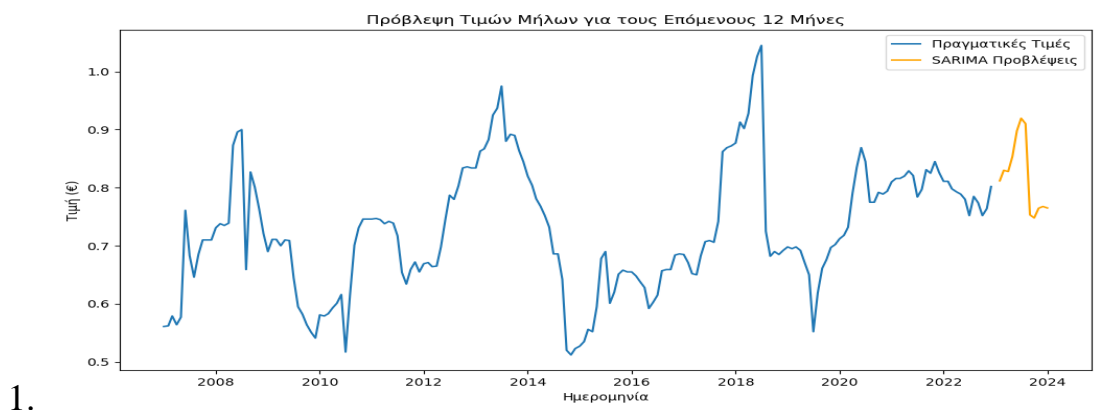
4.3.6 Οπτικοποίηση Αποτελεσμάτων

```
plt.figure(figsize=(12, 6))
plt.plot(data.index.to_timestamp(), data['APPLE_PRICE'], label='Πραγματικές Τιμές')
plt.plot(sarima_forecast_index, predicted_prices, label='SARIMA Προβλέψεις', color='orange')
plt.title('Πρόβλεψη Τιμών Μήλων για τους Επόμενους 12 Μήνες')
plt.xlabel('Ημερομηνία')
plt.ylabel('Τιμή (€)')
plt.legend()
plt.show()
```

Εικόνα 29. SARIMA_6

Τέλος τα αποτελέσματα των προβλέψεων απεικονίζονται έχοντας δημιουργήσει γραφήματα για την ανάλυση της ακρίβειας του μοντέλου και τη σύγκριση με τα πραγματικά δεδομένα.

4.3.7 Αποτελέσματα Μοντέλου



1.

Εικόνα 30. SARIMA_Prediction_Results

Πίνακας 9. Αποτελέσματα προβλέψεων για τους επόμενους 12 μήνες με χρήση SARIMA

ΜΗΝΕΣ	ΠΡΟΒΕΠΟΜΕΝΗ ΤΙΜΗ
January 2023	0,81€
February 2023	0,83€

March 2023	0,83€
April 2023	0,85€
May 2023	0,90€
June 2023	0,92€
July 2023	0,91€
August 2023	0,75€
September 2023	0,75€
October 2023	0,76€
November 2023	0,77€
December 2023	0,77€
MSE	0,0517
MAE	0,1724

4.4 Prophet Model

Ακολουθεί αναλυτική περιγραφή των βημάτων που ακολουθήθηκαν και των παραμέτρων που χρησιμοποιήθηκαν με την χρήση του μοντέλου Prophet.

4.4.1 Εισαγωγή Βιβλιοθηκών και Προεπεξεργασία Δεδομένων

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from prophet import Prophet
from sklearn.model_selection import TimeSeriesSplit
from sklearn.metrics import mean_squared_error
import logging

logging.getLogger('cmdstanpy').setLevel(logging.WARNING)
```

Εικόνα 31. PROPHET_1

- **NumPy** για αριθμητικούς υπολογισμούς και την διαχείριση πολυδιάστατων πινάκων.
- **Pandas** για την ανάγνωση και επεξεργασία δεδομένων.
- **Matplotlib** για τη δημιουργία γραφημάτων και απεικονίσεων.
- **Prophet** για την μοντελοποίηση και πρόβλεψη χρονοσειρών.
- **Scikit-Learn TimeSeriesSplit** για τη μέθοδο διασταυρούμενης επικύρωσης (cross-validation) σε χρονοσειρές.
- **Scikit-Learn mean_squared_error** για τον υπολογισμό του μέσου τετραγωνικού σφάλματος.
- **Logging** για να απενεργοποιηθεί η εμφάνιση προειδοποιήσεων που αφορούν την εσωτερική λειτουργία του “cmdstanpy” που χρησιμοποιείται από το Prophet.

4.4.2 Φόρτωση και Προετοιμασία Δεδομένων

```
filepath = 'DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx'
data = pd.read_excel(filepath)

data.columns = data.columns.str.strip()
data['DATE'] = pd.to_datetime(data['DATE'], dayfirst=True)
df_prophet = data[['DATE', 'APPLE_PRICE']].rename(columns={'DATE': 'ds', 'APPLE_PRICE': 'y'})
```

Εικόνα 32. PROPHET_2

Αρχικά γίνεται φόρτωση των δεδομένων του Dataset “DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx”. Επιπλέον γίνεται επεξεργασία και προετοιμασία των δεδομένων προκειμένου να ξεκινήσει η ανάλυση. Καθαρίζονται τα ονόματα των στηλών του συνόλου δεδομένων από τυχόν περιττά κενά διαστήματα προκειμένου να αποφευχθούν πιθανά λάθη κατά την αναφορά τους και η στήλη των ημερομηνιών μετατρέπεται σε μορφή datetime. Τέλος, τα δεδομένα προετοιμάζονται για χρήση με το μοντέλο Prophet, με τη στήλη των ημερομηνιών να μετονομάζεται σε 'ds' και τη στήλη με τις τιμές σε 'y'.

4.4.3 Διαχωρισμός Δεδομένων, Διαδικασία Cross-Validation Ρυθμίσεις και Ρυθμίσεις Παραμέτρων Μοντέλου Prophet


```

tscv = TimeSeriesSplit(n_splits=5)
mse_scores = []

for train_index, test_index in tscv.split(df_prophet):
    train, test = df_prophet.iloc[train_index], df_prophet.iloc[test_index]

    model = Prophet(yearly_seasonality=True, daily_seasonality=False, weekly_seasonality=False)
    model.fit(train)

    future = model.make_future_dataframe(periods=len(test), freq='M')
    forecast = model.predict(future)

    predictions = forecast['yhat'].iloc[-len(test):].values
    mse = mean_squared_error(test['y'], predictions)
    mse_scores.append(mse)

mean_mse_score = np.mean(mse_scores)
print(f'Mean MSE Score from Time Series Cross-Validation: {mean_mse_score:.4f}')

```

Εικόνα 33. PROPHET_3

Αρχικά γίνεται αξιολόγηση απόδοσης του μοντέλου και εφαρμόζεται η διαδικασία διασταυρούμενης επικύρωσης (cross-validation) με τη μέθοδο TimeSeriesSplit ειδικά προσαρμοσμένη για χρονοσειρές. Επιπλέον πραγματοποιείται διαχωρισμός των δεδομένων σε εκπαιδευόμενα και δοκιμαστικά. Το μοντέλο Prophet δημιουργείτε και εκπαιδεύεται με τις παραμέτρους “yearly_seasonality=True” που δίνει την δυνατότητα να επιτραπεί στο μοντέλο να ενσωμάτωση τις ετήσιες τάσης. Και με τις παραμέτρους “daily_seasonality=False” και “weekly_seasonality=False” που έχουν ως σκοπό να απενεργοποιήσουν τις ημερήσιες και εβδομαδιαίες εποχικές τάσεις, καθώς τα δεδομένα είναι σε μηνιαία βάση και δεν επηρεάζονται από μικρότερες χρονικές διακυμάνσεις.

4.4.4 Εκπαίδευση Τελικού Μοντέλου και Πρόβλεψη

```

prophet_model = Prophet(yearly_seasonality=True, daily_seasonality=False, weekly_seasonality=False)
prophet_model.fit(df_prophet)

future_steps = 12
future = prophet_model.make_future_dataframe(periods=future_steps, freq='M')
forecast = prophet_model.predict(future)

```

Εικόνα 34. PROPHET_4

Γίνεται εκπαίδευση του μοντέλου σε ολόκληρο το σύνολο δεδομένων για την τελική πρόβλεψη των επόμενων 12 μηνών χρησιμοποιώντας την συνάρτηση “make_future_dataframe”, η οποία δημιουργεί ένα νέο Dataframe για τις μελλοντικές τιμές, και τέλος χρησιμοποιείται η συνάντηση “predict” η οποία υπολογίζει τις προβλεπόμενες τιμές στις μελλοντικές ημερομηνίες.

4.4.5 Εκτύπωση Προβλέψεων για τους Επόμενους 12 Μήνες

```
predicted_prices = forecast['yhat'].iloc[-future_steps:].values
forecast_dates = forecast['ds'].iloc[-future_steps:].values

print("\nΠροβλέψεις για τους επόμενους 12 μήνες:")
for date, price in zip(forecast_dates, predicted_prices):
    print(f"{pd.Timestamp(date).strftime('%B %Y')}: Προβλεπόμενη τιμή {price:.2f}€")
```

Εικόνα 35. PROPHET_5

Γίνεται εκτύπωση των προβλέψεων για τους επόμενους 12 μήνες σε κατανοητή μορφή και δημιουργείται μια λίστα με τα ονόματα των μηνών όπου με τη βοήθεια ενός βρόχου for, οι προβλέψεις για τους επόμενους 12 μήνες εκτυπώνονται μαζί με το αντίστοιχο όνομα του μήνα και το έτος.

4.4.6 Οπτικοποίηση Αποτελεσμάτων

```
plt.figure(figsize=(12, 6))
plt.plot(df_prophet['ds'], df_prophet['y'], label='Πραγματικές Τιμές', color='blue') |
plt.plot(forecast_dates, predicted_prices, label='Prophet Προβλέψεις', color='orange')
plt.title('Πρόβλεψη Τιμών Μήλων για τους Επόμενους 12 Μήνες με Prophet')
plt.xlabel('Ημερομηνία')
plt.ylabel('Τιμή (€)')
plt.legend()
plt.show()
```

Εικόνα 36. PROPHET_6

Τέλος τα αποτελέσματα των προβλέψεων απεικονίζονται έχοντας δημιουργήσει γραφήματα για την ανάλυση της ακρίβειας του μοντέλου και τη σύγκριση με τα πραγματικά δεδομένα.

4.4.7 Αποτελέσματα Μοντέλου



Εικόνα 37. PROPHET_Prediction_Results

Πίνακας 10. Αποτελέσματα προβλέψεων για τους επόμενους 12 μήνες με χρήση PROPHET

ΜΗΝΕΣ	ΠΡΟΒΕΠΟΜΕΝΗ ΤΙΜΗ
January 2023	0,77€
February 2023	0,78€
March 2023	0,78€
April 2023	0,80€
May 2023	0,83€
June 2023	0,79€
July 2023	0,76€
August 2023	0,78€
September 2023	0,79€
October 2023	0,79€
November 2023	0,78€
December 2023	0,78€
MSE	0,0374
MAE	0,1542

5. Αξιολόγηση Αποτελεσμάτων

Στο παρόν κεφάλαιο γίνεται ανάλυση της απόδοσης των τριών μοντέλων πρόβλεψης που χρησιμοποιήθηκαν για την πρόβλεψη των τιμών ενδεικτικά των μήλων για το έτος 2023: LSTM, SARIMA και Prophet. Η αξιολόγηση βασίστηκε στους στατιστικούς δείκτες MSE (Mean Squared Error) και MAE (Mean Absolute Error).

Ο δείκτης MSE ή Μέσο Τετραγωνικό Σφάλμα, είναι ένα στατιστικό μέτρο χρησιμοποιείται για την ακριβή αξιολόγηση ενός μοντέλου πρόβλεψης. Ο MSE υπολογίζει τη μέση τιμή των τετραγωνικών διαφορών μεταξύ των προβλεπόμενων και πραγματικών τιμών.

Ο δείκτης MAE ή Μέσο Απόλυτο Σφάλμα, είναι ένα στατιστικό μέτρο που χρησιμοποιείται και αυτό για την ακριβή αξιολόγηση ενός μοντέλου πρόβλεψης, ωστόσο υπολογίζει τον μέσο όρο των απόλυτων διαφορών μεταξύ των προβλεπόμενων και πραγματικών τιμών.

5.1 Αποτελέσματα LSTM

Το μοντέλο LSTM αξιοποίησε επαναλαμβανόμενα νευρωνικά δίκτυα για να μάθει από τις χρονοσειρές των δεδομένων των τιμών των μήλων. Οι προβλέψεις του μοντέλου όπως παρουσιάζει το διάγραμμα “ Εικόνα 23. LSTM_Prediction_Result” έδειξαν ομαλή αυξητική τάση από τον Ιανουάριο έως και τον Δεκέμβριο του 2023, με τις προβλεπόμενες τιμές να κυμαίνονται μεταξύ 0,81€ έως 0,94€.

Οι Δείκτες Απόδοσης του μοντέλου ήταν για τον MSE: **0.0004**, και για τον MAE: **0.0203**. Ωστόσο παρόλο που οι δείκτες σφάλματος ήταν εξαιρετικά χαμηλοί, υποδηλώνοντας υψηλή ακρίβεια προβλέψεις, το μοντέλο δεν μπόρεσε να συλλάβει τις εποχικές διακυμάνσεις που είναι ιστορικά παρατηρούμενες, με αποτέλεσμα το μοντέλο να αναπαράγει απλά την γενική αυξητική τάση.

5.2 Αποτέλεσμα SARIMA

Το μοντέλο SARIMA αξιοποίησε την ανάλυση χρονοσειρών προκειμένου να βρει έντονα εποχιακά μοτίβα. Οι προβλέψεις του μοντέλου όπως παρουσιάζει το διάγραμμα “ Εικόνα 30. SARIMA_Prediction_Result” έδειξαν με ακρίβεια τις εποχιακές διακυμάνσεις καθώς και τις απότομες μεταβολές στις τιμές, με τις προβλεπόμενες τιμές να κυμαίνεται μεταξύ 0,75€ και 0,92€.

Οι Δείκτες Απόδοσης του μοντέλου ήταν για τον MSE: **0.0517**, και για τον MAE: **0.1724**. Ωστόσο παρόλο που οι δείκτες σφάλματος ήταν εξαιρετικά υψηλοί, η δεινότητα του να καταγραφεί με υψηλή ακρίβεια τις εποχιακές διακυμάνσεις το καθιστά πιο κατάλληλο συγκριτικά με το μοντέλο LSTM.

5.3 Αποτελέσματα Prophet

Το μοντέλο Prophet αξιοποίησε επίσης την ανάλυση χρονοσειρών προκειμένου να εμφανίσει σαφείς τάσης και εποχικότητα. Οι προβλέψεις του μοντέλου όπως παρουσιάζει το διάγραμμα “ Εικόνα 37. PROPHET_Prediction_Result” έδειξαν μια ομαλή πρόβλεψη, με τις προβλεπόμενες τιμές να κυμαίνεται μεταξύ 0,77€ και 0,80€.

Οι Δείκτες Απόδοσης του μοντέλου ήταν για τον MSE: **0.0374**, και για τον MAE: **0.1542**. Ωστόσο παρόλο που οι δείκτες σφάλματος ήταν χαμηλότερη σε σχέση με του SARIMA , η ομαλότητα των προβλέψεων του Prophet υποδεικνύει μια πολύ περιορισμένη ικανότητα στην αποτύπωση των σύνθετων εποχικών διακυμάνσεων.

5.4 Συσχέτιση Προβλεπόμενων και Ιστορικών Τιμών

Η αξιολόγηση της συσχέτισης μεταξύ των προβλεπόμενων και των ιστορικών τιμών είναι κρίσιμη για την εκτίμηση της ικανότητας κάθε μοντέλου να αποτυπώσει την πραγματική συμπεριφορά των τιμών των μήλων με την πάροδο του χρόνου. Η ιστορική ανάλυση δείχνει ότι οι τιμές των μήλων παρουσιάζουν σημαντικές εποχικές διακυμάνσεις.

- **SARIMA:** Οι προβλεπόμενες τιμές του SARIMA εμφανίζουν υψηλή συσχέτιση με τις ιστορικές διακυμάνσεις των τιμών, αποτυπώνοντας εποχικές αυξομειώσεις και απότομες αλλαγές. Αυτό εξηγεί ότι το SARIMA έχει καταφέρει να ενσωματώσει τις εποχικές συνιστώσες στη διαδικασία πρόβλεψης, παρέχοντας μια πιο ακριβή εκτίμηση των μελλοντικών τιμών με βάση τις ιστορικές τάσεις.
- **LSTM και Prophet:** Τα μοντέλα LSTM και Prophet από την άλλη εμφανίζουν χαμηλότερη συσχέτιση με τις ιστορικές τιμές σε σχέση με το SARIMA. Ωστόσο οι προβλέψεις του LSTM δείχνουν μια γραμμική και συνεχώς αυξανόμενη τάση, ενώ το Prophet παράγει πιο ομαλές προβλέψεις που δεν αποτυπώνουν πλήρως τις ιστορικές διακυμάνσεις. Αυτή η συμπεριφορά υποδηλώνει ότι τα δύο αυτά μοντέλα δεν να μην είναι τόσο αποτελεσματικά στην αποτύπωση της εποχικότητας και της μεταβλητότητας που χαρακτηρίζουν τα ιστορικά δεδομένα.

5.5 Συμπέρασμα

Συνοψίζοντας αντιλαμβανόμαστε ότι παρά τα χαμηλά αριθμητικά σφάλματα των μοντέλων **LSTM** και **Prophet**, το **SARIMA** θεωρείται το πιο κατάλληλο για την πρόβλεψη των τιμών των μήλων. Το SARIMA παρουσιάζει καλύτερη απόδοση στην αποτύπωση των εποχικών

διακυμάνσεων και των απότομων αλλαγών που είναι κρίσιμες για την ακρίβεια των προβλέψεων σε μια αγορά όπου οι τιμές επηρεάζονται έντονα από εποχικούς παράγοντες.

6. Πληροφοριακό Σύστημα

Η υλοποίηση του πληροφοριακού συστήματος για την παρούσα πτυχιακή εργασία έγινε εξ ολοκλήρου με την χρήση της πλατφόρμας Wix.com. Η ιστοσελίδα έχει έως σκοπό την παροχή πληροφορίας αναφορικά με τις τιμές των αχλαδιών και των μήλων, παρέχοντας πιο συγκεκριμένα την δυνατότητα να παρακολουθούν οι χρήστες τις μελλοντικές τιμές των προϊόντων ανά μήνα για την τρέχουσα χρονιά. Επιπλέον, είναι σημαντικό να σημειωθεί ότι στην συγκεκριμένη υλοποίηση η ιστοσελίδα αναπτύχθηκε ως **στατική** με περιορισμένες λειτουργίες διαδραστικότητας και δυναμικής ενημερώσεις. Ωστόσο, η παρούσα υλοποίηση καλύπτει τις βασικές ανάγκες παρουσίασης δεδομένων, ενώ υπάρχει δυνατότητα για μελλοντικές βελτιώσεις και επέκτασης της λειτουργικότητας στις οποίες θα γίνει αναφορά στο Κεφάλαιο 7. Τέλος, στο συγκεκριμένο κεφάλαιο περιγράφεται η διαδικασία ανάπτυξης της ιστοσελίδας και ένα εγχειρίδιο χρήσης που παρέχει οδηγίες για την χρήση του από τους τελικούς χρήστες.

6.1 Υλοποίηση του Πληροφοριακού Συστήματος μέσω Wix

Η υλοποίηση της πλατφόρμας πραγματοποιήθηκε μέσω της πλατφόρμας Wix.com, δημιουργώντας για αρχή ένα στατικό πρότυπο για την εμφάνιση των τιμών των προϊόντων και λυτών πληροφοριών. Το πρότυπο δημιουργήθηκε με γνώμονα να είναι όσο πιο user friendly γίνεται, καθώς θέλαμε να είναι απλό και ευανάγνωστο αλλά παράλληλα και ελκυστικό.

Για την ανάπτυξη της ιστοσελίδας, χρησιμοποιήθηκαν τα εργαλεία **drag-and-drop** του Wix, τα οποία επιτρέπουν την εύκολη διάταξη και προσαρμογή των στοιχείων χωρίς προγραμματισμό. Η σελίδα αναπτύχθηκε ως στατική πλατφόρμα, κάτι που σημαίνει ότι τα δεδομένα που παρουσιάζονται, όπως οι τιμές και τα γραφήματα, εισήχθησαν χειροκίνητα κατά την κατασκευή της ιστοσελίδας και δεν ενημερώνονται δυναμικά από εξωτερικές πηγές.

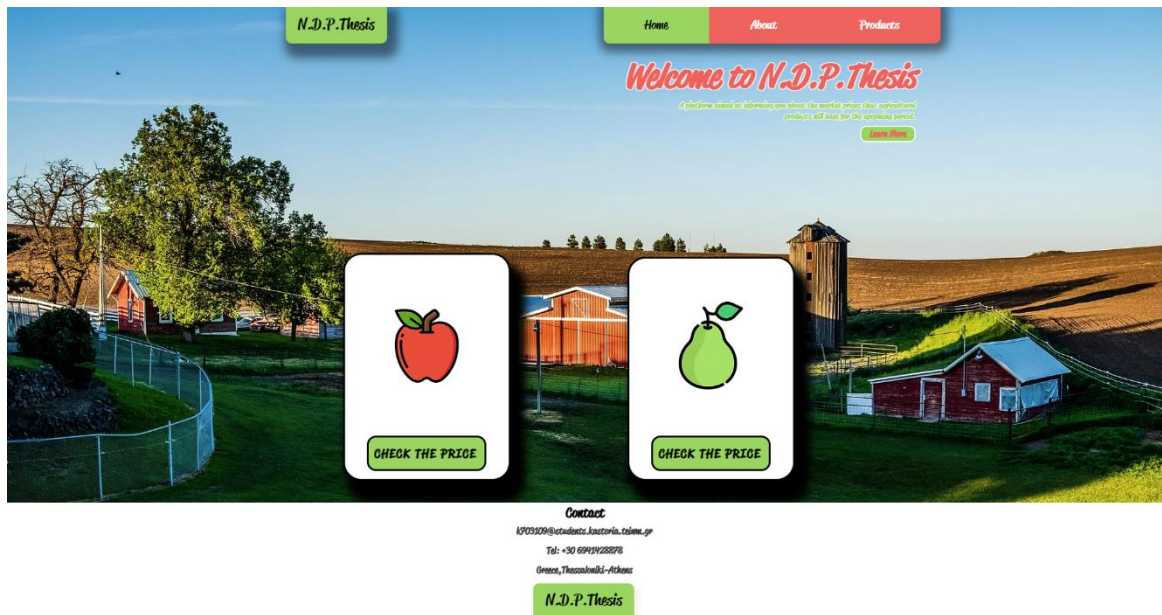
6.2 Περιγραφή του Οδηγού Χρήστη

Το User Manual παρέχει λεπτομερείς οδηγίες για τους χρήστες της ιστοσελίδας, περιγράφοντας τις βασικές λειτουργίες και τη διαδικασία πλοήγησης.

Η πλοήγηση στην ιστοσελίδα γίνεται μέσω του μενού πλοήγησης που βρίσκεται στην κορυφή κάθε σελίδας εντός του Header, και αποτελείται από τις παρακαλώ βασικές σελίδες.

- Home (Αρχική Σελίδα)
- About (Σχετικά με Εμάς)
- Product: Apples, Pears

6.2.1 Ενότητα: Αρχική Σελίδα



Εικόνα 38. Ενότητα: Αρχική Σελίδα

Ο χρήστης:

- Εισερχόμενος στο Home διακρίνει ένα Welcome μήνυμα, κάτω από το οποίο κάνοντας “κλικ” στο button “Learn More” έχει την δυνατότητα να μεταφερθεί στην σελίδα “About”.
- Διακρίνει τις διαθέσιμες επιλογές των φρούτων (μήλα & αχλαδιά) για τα οποία μπορεί να ενημερωθεί.
- Κάνοντας “κλικ” στα buttons “CHECK THE PRICE” έχει την δυνατότητα να μεταφερθεί στην υποσελίδα του κάθε προϊόντος. Δηλαδή εάν κάνει “κλικ” στο button των μήλων τότε θα μεταφερθεί στην υποσελίδα “Apples”, ειδάλλως εάν επιλέξει στο button των αχλαδιών θα μεταφερθεί στην υποσελίδα “Pears”.
- Στο κάτω μέρος τις σελίδας διακρίνει πληροφορίες επικοινωνίας εντός του Footer.

6.2.2 Ενότητα: Σχετικά

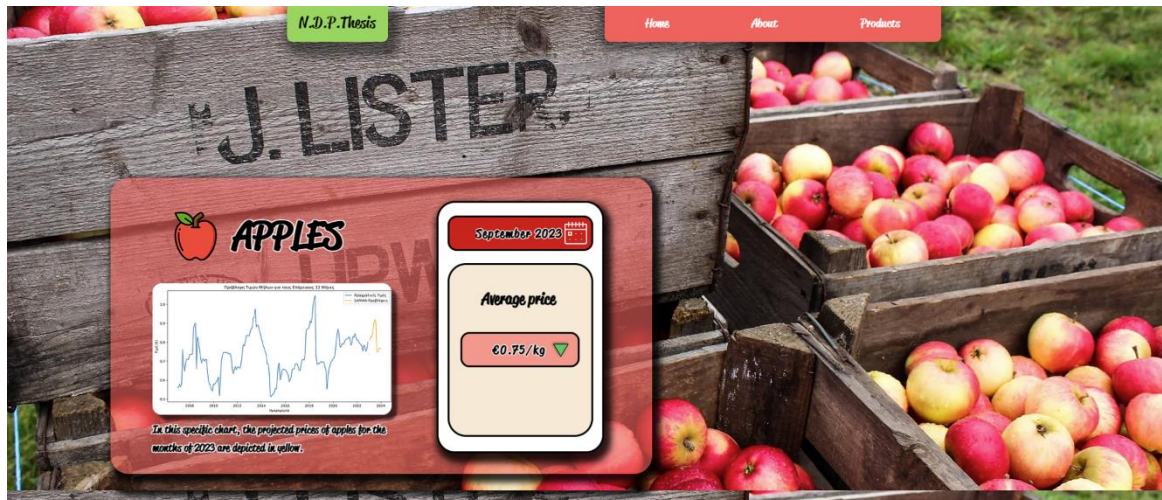


Εικόνα 39. Ενότητα: Σχετικά

Ο χρήστης:

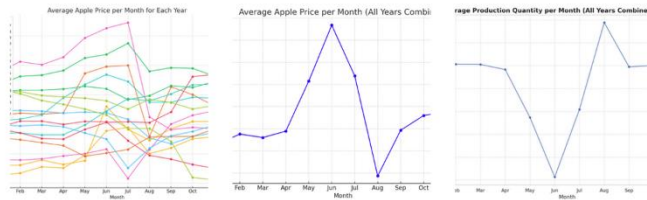
- Εισερχόμενος στην σελίδα About είτε από την μπάρα πλοήγησης, είτε από το button “Learn More” μέσω του Home, έχει την δυνατότητα να πληροφορηθεί σχετικά την πλατφόρμα.
- Έχει την δυνατότητα κάνοντας “κλικ” πάνω αριστερά στο Logo της σελίδας να μεταφέρει στο Home.
- Στο κάτω μέρος τις σελίδας διακρίνει πληροφορίες επικοινωνίας εντός του Footer.

6.2.3 Ενότητα: Μήλα



View The Line Graphs

The specific graphs depict the prices of apples from January 2007 to December 2022, the average price per month for all years combined, as well as the average production quantity per month for all years combined.



Contact

1925109@stud.ntua.gr, laourina.telmi.gr

Tel: +30 6949128278

Graeco, Thessaloniki - Athens

N.D.P.Thesis

Εικόνα 40. Ενότητα: Μήλα

Ο χρήστης:

- Εισερχόμενος στην σελίδα Apples είτε από την μπάρα πλοήγησης Products -> Apples, είτε από το button "CHECK THE PRICE" μέσω του Home, έχει την δυνατότητα να πληροφορηθεί σχετικά τις τρέχουσες ή μελλοντικές τιμές των μήλων.
 - Εντός του κόκκινου παραλληλογράμμου πλαισίου διακρίνει:
 - Ένα γράφημα στα αριστερά με τις προβλεπόμενες μελλοντικές τιμές μήλου για το χρονικό διάστημα 2023 το οποίο έχει την δυνατότητα κάνοντας "κλικ" επάνω του να μεγεθύνει.
- Σημείωση:** Οι προβλεπόμενες μελλοντικές τιμές παρουσιάζονται στο γράφημα με κίτρινο χρώμα, ενώ με μπλε παρουσιάζονται οι τιμές για το χρονικό διάστημα 2007 – 2022.
- Ένα πλαίσιο στα δεξιά το οποίο παρουσιάζει τον τρέχοντα μήνα και την τιμή τρέχουσα τιμή του προϊόντος.

- Στο σημείο “View The Line Graphs” έχει την δυνατότητα επιλέγοντας τα διαγράμματα να πληροφορηθεί αναφορικά με τις τιμές και την ποσότητα παραγωγής των μήλων.
- Έχει την δυνατότητα κάνοντας “κλικ” πάνω αριστερά στο Logo της σελίδας να μεταφέρει στο Home.
- Στο κάτω μέρος τις σελίδας διακρίνει πληροφορίες επικοινωνίας εντός του Footer.

6.2.4 Ενότητα: Αχλαδιά



View The Line Graphs

The specific graphs depict the prices of pears from January 2009 to December 2022, the average price per month for all years combined, as well as the average production quantity per month for all years combined.



Contact

19203102@studinfo.karolm.tolma.gr

Tel: +30 6941928298

Greece, Thessaloniki - Athens

N.D.P. Thesis

Εικόνα 41. Ενότητα: Αχλαδιά

Ο χρήστης:

- Εισερχόμενος στην σελίδα Pears είτε από την μπάρα πλοήγησης Products -> Pears, είτε από το button “CHECK THE PRICE” μέσω του Home, έχει την δυνατότητα να πληροφορηθεί σχετικά τις τρέχουσες ή μελλοντικές τιμές των αχλαδιών.
- Εντός του πράσινου παραλληλογράμμου πλαισίου διακρίνει:

- Ένα γράφημα στα αριστερά με τις προβλεπόμενες μελλοντικές τιμές αχλαδιού για το χρονικό διάστημα 2023 το οποίο έχει την δυνατότητα κάνοντας “κλικ” επάνω του να μεγεθύνει.
Σημείωση: Οι προβλεπόμενες μελλοντικές τιμές παρουσιάζονται στο γράφημα με κίτρινό χρώμα, ενώ με μπλε παρουσιάζονται οι τιμές για το χρονικό διάστημα 2007 – 2022.
- Ένα πλαίσιο στα δεξιά το οποίο παρουσιάζει τον τρέχοντα μήνα και την τιμή τρέχουσα τιμή του προϊόντος.
- Στο σημείο “View The Line Graphs” έχει την δυνατότητα επιλέγοντας τα διαγράμματα να πληροφορηθεί αναφορικά με τις τιμές και την ποσότητα παραγωγής των αχλαδιών.
- Έχει την δυνατότητα κάνοντας “κλικ” πάνω αριστερά στο Logo της σελίδας να μεταφέρει στο Home.
- Στο κάτω μέρος τις σελίδας διακρίνει πληροφορίες επικοινωνίας εντός του Footer.

7. Συμπεράσματα και Μελλοντικές Επεκτάσεις

7.1 Συμπεράσματα

Η παρούσα πτυχιακή εργασία είχε ως βασικό στόχο την ανάπτυξη ενός πληροφοριακού συστήματος για την πρόβλεψη τιμών αγροτικών προϊόντων, εστιάζοντας κυρίως στις τιμές μήλων και αχλαδιών. Το σύστημα αναπτύχθηκε ως μια στατική ιστοσελίδα μέσω του **Wix.com**, με κύρια λειτουργία την προβολή ιστορικών δεδομένων και προβλέψεων τιμών χρησιμοποιώντας τα μοντέλα μηχανικής μάθησης και στατιστικής ανάλυσης χρονοσειρών **LSTM**, **Prophet**, και **SARIMA**.

Κατά την υλοποίηση, το μοντέλο **SARIMA** υπερείχε στην καταγραφή εποχικών διακυμάνσεων, καθιστώντας το καταλληλότερο μοντέλο για την πρόβλεψη τιμών αγροτικών προϊόντων. Ωστόσο παρότι το **LSTM** και **Prophet** παρουσίασαν χαμηλότερα σφάλματα, δεν κατάφεραν να αποτυπώσουν επαρκώς τις εποχικές διακυμάνσεις που παρατηρήθηκαν στις τιμές.

Η ανάπτυξη της ιστοσελίδας ως στατική πλατφόρμα μέσω του **Wix.com** είχε ορισμένους περιορισμούς όσον αφορά τη δυναμική ενημέρωση των δεδομένων και της διαδραστικότητας της. Τα δεδομένα και οι προβλέψεις εισήχθησαν χειροκίνητα, χωρίς δυνατότητα αυτόματης ενημέρωσης σε πραγματικό χρόνο, γεγονός που περιόρισε τη χρηστικότητα και την αποτελεσματικότητα του συστήματος. Παράλληλα, η διεπαφή χρήστη ήταν απλή και λειτουργική, αλλά η απουσία πρόσθετων δυνατοτήτων, όπως φίλτρα και διαδραστικά γραφήματα, περιόρισε την αλληλεπίδραση των χρηστών με τα δεδομένα.

Παρόλο που η παρούσα υλοποίηση παρουσιάζει ορισμένες αδυναμίες, οι στόχοι της εργασίας καλύφθηκαν ικανοποιητικά. Το πληροφοριακό σύστημα που αναπτύχθηκε παρέχει μια καλή βάση για μελλοντικές επεκτάσεις και βελτιώσεις, με τους αλγορίθμους πρόβλεψης, όπως το **SARIMA**, να αποδεικνύουν την αξία τους στην ανάλυση τιμών που εμφανίζουν εποχικά μοτίβα. Κλίνοντας, αυτή η εργασία έθεσε τα θεμέλια για μια αποτελεσματική και λειτουργική πλατφόρμα, με προοπτικές ανάπτυξης που θα μπορούσαν να βελτιώσουν την εμπειρία των χρηστών και να ενισχύσουν τη χρησιμότητα της πλατφόρμας για τη γεωργική βιομηχανία

7.2 Μελλοντικές Επεκτάσεις

Για να ενισχυθεί η χρησιμότητα του πληροφοριακού συστήματος και να βελτιωθεί η εμπειρία του χρήστη, προτείνονται κάποιες μελλοντικές επεκτάσεις που θα μπορούσαν να βελτιώσουν τη λειτουργικότητα της πλατφόρμας και να παρέχουν μεγαλύτερη ευελιξία και διαδραστικότητα στους χρήστες:

2. **Δυναμική Ενημέρωση Δεδομένων:** Η προσθήκη ενός μηχανισμού αυτόματης ενημέρωσης των δεδομένων θα ήταν καθοριστική βελτίωση. Αυτό θα μπορεί να επιτευχθεί μέσω **APIs** που θα παρέχουν σε πραγματικό χρόνο ενημερώσεις για τις τιμές των προϊόντων. Με αυτόν τον τρόπο, οι χρήστες θα έχουν πρόσβαση στις πιο πρόσφατες τιμές και προβλέψεις χωρίς την ανάγκη χειροκίνητης παρέμβασης.
3. **Διαδραστική Πλατφόρμα και Φιλτράρισμα Δεδομένων:** Η δυνατότητα στους χρήστες να φιλτράρουν δεδομένα ανά χρονική περίοδο, γεωγραφική περιοχή ή τύπο προϊόντος θα βελτιώσει την ευελιξία του συστήματος. Επιπλέον, η ενσωμάτωση διαδραστικών γραφημάτων, όπου οι χρήστες θα μπορούν να πειραματίζονται με τα δεδομένα και να δουν διαφορετικά σενάρια τιμών, θα δώσει μεγαλύτερη αξία στο σύστημα.
4. **Προσθήκη Συστήματος Ειδοποιήσεων:** Ένα σύστημα ειδοποιήσεων που θα ενημερώνει τους χρήστες όταν οι τιμές ενός προϊόντος φτάνουν σε ορισμένα επίπεδα (π.χ., υψηλή ζήτηση ή απότομες αυξομειώσεις) θα προσφέρει πρακτική λειτουργικότητα. Οι ειδοποιήσεις θα μπορούσαν να αποστέλλονται μέσω email ή εφαρμογών για κινητά, καθιστώντας την πλατφόρμα πιο χρήσιμη για τους παραγωγούς και εμπόρους αγροτικών προϊόντων.
5. **Επέκταση σε Περισσότερα Αγροτικά Προϊόντα:** Επί του παρόντος, η ιστοσελίδα επικεντρώνεται μόνο στις τιμές των μήλων και αχλαδιών. Η επέκταση σε περισσότερα προϊόντα, όπως ελαιόλαδο, σταφύλια ή σιτηρά, θα καθιστούσε την πλατφόρμα πιο χρήσιμη για μεγαλύτερο αριθμό παραγωγών και επιχειρήσεων στον αγροτικό τομέα.
6. **Προσωποποιημένη Εμπειρία Χρήστη:** Η προσθήκη εξατομικευμένων δυνατοτήτων, όπως η δημιουργία προσωπικών προφίλ για χρήστες, όπου θα μπορούν να αποθηκεύουν τα προϊόντα που τους ενδιαφέρουν, να παρακολουθούν τις τιμές που έχουν σημασία για αυτούς και να δημιουργούν εξατομικευμένες αναφορές, θα βελτιώσει την αλληλεπίδραση και την εμπειρία χρήστη.

7. **Υποστήριξη Πολλαπλών Γλωσσών και Γεωγραφικών Περιοχών:** Η δυνατότητα υποστήριξης πολλαπλών γλωσσών και γεωγραφικών περιοχών θα καταστήσει την πλατφόρμα πρόσβαση σε ευρύτερο κοινό. Μελλοντική υποστήριξη δεδομένων για αγροτικά προϊόντα από διαφορετικές περιοχές του κόσμου θα προσθέσει αξία σε χρήστες από διάφορες αγροτικές κοινότητες.
8. **Αναλυτικά Στοιχεία και Εκθέσεις:** Ένα ακόμη χρήσιμο χαρακτηριστικό θα ήταν η δυνατότητα παροχής αναλυτικών στοιχείων και αναφορών, όπως προβλέψεις για τις τάσεις της αγοράς, εξαγωγές, εισαγωγές, και συνδυαστική ανάλυση δεδομένων από διάφορους αγροτικούς κλάδους. Αυτό θα επιτρέψει στους χρήστες να κατανοήσουν καλύτερα τις τάσεις και τις μεταβολές στην αγορά

Αναφορές

- Allen, P.G. (1994). Economic forecasting in agriculture. *International Journal of Forecasting*, 10, pp. 81-135.
- Anderson, J.B., & Johnnesson, R. (1996). *Understanding Information Transmission*. Ieee Press. ISBN 978-0471711209.
- Andrew, A.M. (2000). *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge: Cambridge University Press.
- Atkinson, R.D., & Castro, D.D. (2008). *Digital Quality of Life: Understanding the Personal and Social Benefits of the Information Technology Revolution*.
- Cuaresma, J.C., Hlouskova, J., Kossmeier, S., & Obersteiner, M. (2004). Forecasting electricity spot-prices using linear univariate time-series models. *Applied Energy*, 77, pp. 87–106.
- de Mello, R.F., & Ponti, M.A. (2018). Statistical learning theory. *Machine Learning*, 2018, pp. 75–128.
- Duan, Q., Zhang, L., Wei, F., Xiao, X., & Wang, L. (2017). Time Series GA-SVR based Fish Price Prediction Model and Validation. *Transactions of the Chinese Society of Agricultural Engineering*, 33, pp. 308–314.
- Feng, F., Wei, F., & Miu, J.H. (2012). Price Prediction of Traditional Chinese Medicine *Siraitia grosvenorii* Based on Grey System GM(1,1) Model. *Guangxi Science*, 19, pp. 15–20.
- Floridi, L. (2010). *Information – A Very Short Introduction*. Oxford University Press. ISBN 978-0-19-160954-1.
- French, J. (2017). The time traveller's CAPM. *Investment Analysts Journal*, 46(2), pp. 81–96. doi:10.1080/10293523.2016.1255469.
- Ge, Y., & Wu, H. (2020). Prediction of corn price fluctuation based on multiple linear regression analysis model under big data. *Neural Computational Applications*, 32, pp. 16843–16855.
- Gu, Z., & Zhang, Y. (2023). A study on the Influence Factors of Agricultural Prices based on Machine Learning—Taking oilseeds as an example. *Price Theory Practice*, 4, pp. 122–126.
- Hilbert, M., & López, P. (2011). The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025), pp. 60–65. doi:10.1126/science.1200970.

- Jadhav, V., Chinnappa, R.B., & Gaddi, G. (2017). Application of ARIMA model for forecasting agricultural prices. *Journal of Agricultural Science & Technology*, 9, pp. 981–992.
- B V Vishwas Ashish Patel, “Hands-on Time Series Analysis with Python: From Basics to Bleeding Edge Techniques”, <https://doi.org/10.1007/978-1-4842-5992-4>, 2020
- Lin, J., Keogh, E., Lonardi, S., & Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery. New York: ACM Press. pp. 2–11. doi:10.1145/882082.882086.
- Lippert, S.K., & Forman, H. (2005). Utilization of Information Technology: Examining Cognitive and Experiential Factors of Post-Adoption Behavior. *IEEE Transactions on Engineering Management*, 52(3), pp. 363-381.
- Ma, C., Tao, J.P., & Liu, W. (2022). Pig Epidemic Network Concerns and Pork Price Volatility: Exacerbating or Curbing? *Journal of Huazhong Agricultural University*, 6, pp. 22–34.
- Moore, H.L. (1917). *Forecasting the Yield and the Price of Cotton*. New York, NY: Macmillan.
- Nuvaisyah, P., Nhita, F., & Saepudin, D. (2018). Price prediction of chili commodities in Bandung regency using Bayesian Network. *International Journal on Information and Communication Technology*, 4, pp. 19–32.
- Pearl, J. (1998). Graphical models for probabilistic and causal reasoning. *Quantified Represent. Uncertain. Imprecision*, 1, pp. 367–389.
- Sun, F., Meng, X., Zhang, Y., Wang, Y., Jiang, H., & Liu, P. (2023). Agricultural Product Price Forecasting Methods: A Review. *Agriculture*, 13, 1671.
- Ticehurst, J.L., Letcher, R.A., & Rissik, D. (2008). Integration modelling and decision support: A case study of the Coastal Lake Assessment and Management (CLAM) Tool. *Mathematics and Computers in Simulation*, 78, pp. 435–449.
- Wang, B., Liu, P., Chao, Z., Junmei, W., Chen, W., Cao, N., O’Hare, G.M., & Wen, F. (2018). Research on Hybrid Model of Garlic Short-term Price Forecasting based on Big Data. *Computers, Materials and Continua*, 57, pp. 283–296.
- Webler, F. (2022). Measurement in the Age of Information. *Information*, 13 (3), pp. 111. doi:10.3390/info13030111.
- Wu, L., Liu, S., & Yang, Y. (2016). Grey double exponential smoothing model and its application on pig price forecasting in China. *Applied Soft Computing*, 39, pp. 117–123.

- Yockey, H.P. (2005). *Information Theory, Evolution, and the Origin of Life*. Cambridge University Press. ISBN 978-0511546433.
- Zhang, X.G. (2000). Introduction to Statistical Learning Theory and Support Vector Machines. *Journal of Automatica Sinica*, 26, pp. 32–42.
- Zhang, L. (2013). *Model Construction and Empirical Study on Production and Consumption Forecasting of Major Agricultural Products in China*. Master's Thesis. Beijing, China: University of Chinese Academy of Sciences.

Παραρτήμα Κώδικα

Κώδικας μοντέλου SARIMA:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from statsmodels.tsa.statespace.sarimax import SARIMAX
from sklearn.model_selection import TimeSeriesSplit
from sklearn.metrics import mean_squared_error

filepath = 'DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx'
data = pd.read_excel(filepath)

data.columns = data.columns.str.strip()
data['DATE'] = pd.to_datetime(data['DATE'], dayfirst=True)
data.set_index('DATE', inplace=True)
data.index = data.index.to_period('M')

tscv = TimeSeriesSplit(n_splits=5)
mse_scores = []

for train_index, test_index in tscv.split(data):
    train, test = data.iloc[train_index], data.iloc[test_index]

    model = SARIMAX(train['APPLE_PRICE'],
                    order=(0, 1, 0),
                    seasonal_order=(5, 1, 0, 12),
                    enforce_stationarity=False,
                    enforce_invertibility=False)
    model_fit = model.fit(dispatch=False)

    predictions = model_fit.forecast(steps=len(test))
```

```
mse = mean_squared_error(test['APPLE_PRICE'], predictions)
mse_scores.append(mse)

mean_mse_score = np.mean(mse_scores)
print(f'Mean MSE Score from Time Series Cross-Validation: {mean_mse_score}')

sarima_model = SARIMAX(data['APPLE_PRICE'],
                        order=(0, 1, 0),
                        seasonal_order=(5, 1, 0, 12),
                        enforce_stationarity=False,
                        enforce_invertibility=False)
sarima_fit = sarima_model.fit(dispatch=False)

future_steps = 12
sarima_forecast = sarima_fit.get_forecast(steps=future_steps)
sarima_forecast_index = pd.date_range(start=data.index[-1].to_timestamp(), periods=future_steps
+ 1, freq='M')[1:]

predicted_prices = sarima_forecast.predicted_mean

months = ['January', 'February', 'March', 'April', 'May', 'June',
          'July', 'August', 'September', 'October', 'November', 'December']
print("Προβλέψεις για τους επόμενους 12 μήνες:")

for i, (date, price) in enumerate(zip(sarima_forecast_index, predicted_prices)):
    month_name = months[date.month - 1]
    year = date.year
    print(f"{month_name} {year}: Προβλεπόμενη τιμή {price:.2f}€")

plt.figure(figsize=(12, 6))
plt.plot(data.index.to_timestamp(), data['APPLE_PRICE'], label='Πραγματικές Τιμές')
plt.plot(sarima_forecast_index, predicted_prices, label='SARIMA Προβλέψεις', color='orange')
plt.title('Πρόβλεψη Τιμών Μήλων για τους Επόμενους 12 Μήνες')
plt.xlabel('Ημερομηνία')
plt.ylabel('Τιμή (€)')
plt.legend()
plt.show()
```

Κώδικας μοντέλου LSTM:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.preprocessing import MinMaxScaler
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout, Input
from tensorflow.keras.callbacks import EarlyStopping
from tensorflow.keras.regularizers import l2

filepath = "DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx"
data = pd.read_excel(filepath)
data.columns = data.columns.str.strip()
data['DATE'] = pd.to_datetime(data['DATE'], dayfirst=True)
data.set_index('DATE', inplace=True)

data['Diff_APPLE_PRICE'] = data['APPLE_PRICE'].diff()
data.dropna(inplace=True)

scaler = MinMaxScaler()
data['Scaled_Diff_APPLE_PRICE'] = scaler.fit_transform(data[['Diff_APPLE_PRICE']])

def create_dataset(X, look_back=1):
    Xs, ys = [], []
    for i in range(len(X) - look_back):
        v = X[i:(i + look_back)]
        Xs.append(v)
        ys.append(X[i + look_back])
    return np.array(Xs), np.array(ys)

look_back = 60
X, y = create_dataset(data['Scaled_Diff_APPLE_PRICE'].values, look_back)
X = X.reshape((X.shape[0], X.shape[1], 1))

train_size = int(len(X) * 0.60)
X_train, X_test = X[:train_size], X[train_size:]
y_train, y_test = y[:train_size], y[train_size:]

weights = np.where(data.index[look_back:].year >= 2018, 0.8, 0.2)
```

```

weights_train = weights[:len(y_train)]

model = Sequential([
    Input(shape=(look_back, 1)),
    LSTM(150, return_sequences=True, kernel_regularizer=l2(0.01)),
    Dropout(0.2),
    LSTM(150, return_sequences=True, kernel_regularizer=l2(0.01)),
    Dropout(0.2),
    LSTM(150, kernel_regularizer=l2(0.01)),
    Dropout(0.2),
    Dense(1)
])

model.compile(optimizer='adam', loss='mean_squared_error')

early_stopping = EarlyStopping(monitor='val_loss', patience=100, restore_best_weights=True)
model.fit(X_train, y_train, sample_weight=weights_train, epochs=150, batch_size=50,
validation_split=0.1, callbacks=[early_stopping], verbose=1)

input_seq = X_test[-1].reshape(1, look_back, 1)
predictions_diff_future = []
months_future = pd.date_range(start='2023-01-01', periods=12, freq='M')

for i in range(12):
    next_pred = model.predict(input_seq)
    noise = np.random.normal(0, 0.01)
    next_pred = next_pred + noise
    predictions_diff_future.append(scaler.inverse_transform(next_pred).ravel()[0])
    next_input = next_pred.reshape(1, 1, 1)
    input_seq = np.append(input_seq[:, 1:, :], next_input, axis=1)

last_known_price = data['APPLE_PRICE'].iloc[-1]
predictions_future = []
for diff in predictions_diff_future:
    next_price = last_known_price + diff
    predictions_future.append(next_price)
    last_known_price = next_price

for date, price in zip(months_future, predictions_future):

```

```
print(f"{date.strftime('%B %Y')}: Προβλεπόμενη τιμή {price:.2f}€")

test_dates = data.index[train_size + look_back:]
plt.figure(figsize=(12, 6))
plt.plot(data.index, data['APPLE_PRICE'], label='Πραγματικές Τιμές')
plt.plot(test_dates, data['APPLE_PRICE'][train_size + look_back:], 'g--', label='Δεδομένα Δοκιμών')
plt.plot(test_dates, data['APPLE_PRICE'][train_size + look_back:] +
scaler.inverse_transform(model.predict(X_test)).ravel(), 'r-', label='Προβλέψεις Μοντέλου στα
Δεδομένα Δοκιμών')
plt.plot(months_future, predictions_future, 'b-', label='Προβλέψεις για το 2023')
plt.title('Πρόβλεψη Τιμών Μήλων Συμπεριλαμβανομένων Εκπαίδευσης, Δοκιμών και Μελλοντικών
Προβλέψεων')
plt.xlabel('Ημερομηνία')
plt.ylabel('Τιμή (€)')
plt.legend()
plt.show()
```

Κώδικας μοντέλου Prophet:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from prophet import Prophet
from sklearn.model_selection import TimeSeriesSplit
from sklearn.metrics import mean_squared_error
import logging

logging.getLogger('cmdstanpy').setLevel(logging.WARNING)

filepath = 'DATASET_APPLE_AUTOGAS_PRICE_FINAL.xlsx'
data = pd.read_excel(filepath)

data.columns = data.columns.str.strip()
data['DATE'] = pd.to_datetime(data['DATE'], dayfirst=True)
df_prophet = data[['DATE', 'APPLE_PRICE']].rename(columns={'DATE': 'ds',
'APPLE_PRICE': 'y'})

tscv = TimeSeriesSplit(n_splits=5)
mse_scores = []
```

```

for train_index, test_index in tscv.split(df_prophet):
    train, test = df_prophet.iloc[train_index], df_prophet.iloc[test_index]

    model = Prophet(yearly_seasonality=True, daily_seasonality=False,
weekly_seasonality=False)
    model.fit(train)

    future = model.make_future_dataframe(periods=len(test), freq='M')
    forecast = model.predict(future)

    predictions = forecast['yhat'].iloc[-len(test):].values
    mse = mean_squared_error(test['y'], predictions)
    mse_scores.append(mse)

mean_mse_score = np.mean(mse_scores)
print(f'Mean MSE Score from Time Series Cross-Validation: {mean_mse_score:.4f}')

prophet_model = Prophet(yearly_seasonality=True, daily_seasonality=False,
weekly_seasonality=False)
prophet_model.fit(df_prophet)

future_steps = 12
future = prophet_model.make_future_dataframe(periods=future_steps, freq='M')
forecast = prophet_model.predict(future)

predicted_prices = forecast['yhat'].iloc[-future_steps:].values
forecast_dates = forecast['ds'].iloc[-future_steps:].values

print("\nΠροβλέψεις για τους επόμενους 12 μήνες:")
for date, price in zip(forecast_dates, predicted_prices):
    print(f'{pd.Timestamp(date).strftime('%B %Y')}: Προβλεπόμενη τιμή {price:.2f}€")

plt.figure(figsize=(12, 6))
plt.plot(df_prophet['ds'], df_prophet['y'], label='Πραγματικές Τιμές', color='blue')
plt.plot(forecast_dates, predicted_prices, label='Prophet Προβλέψεις', color='orange')
plt.title('Πρόβλεψη Τιμών Μήλων για τους Επόμενους 12 Μήνες με Prophet')

```



```
plt.xlabel('Ημερομηνία')
```

```
plt.ylabel('Τιμή (€)')
```

```
plt.legend()
```

```
plt.show()
```