



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ  
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
ΣΤΑ ΠΡΟΗΓΜΕΝΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

Διπλωματική Εργασία

*Ανάλυση και Πρόβλεψη της Τιμής των Κρυπτονομισμάτων (Bitcoin) – Με  
Χρήση Αλγορίθμων Μηχανικής Μάθησης και Τεχνητών Νευρωνικών  
Δικτύων*

**Κουτσαβδής Ραφαήλ**

**A.M.: ME1907**

**Επιβλέπων: Αν. Καθηγητής κ. Φιλιππάκης Μιχαήλ**

*Πειραιάς, 2022*

## Περιεχόμενα

<b>Ευχαριστίες.....</b>	<b>5</b>
<b>Περίληψη .....</b>	<b>6</b>
<b>Abstract.....</b>	<b>7</b>
<b>Εισαγωγή .....</b>	<b>8</b>
<b>Κεφάλαιο 1<sup>ο</sup>: Κρυπτονομίσματα .....</b>	<b>10</b>
1.1. Ορισμός.....	10
1.2. Ιστορική αναδρομή .....	10
1.3. Blockchain.....	12
1.3.1. Τύποι Blockchain.....	14
1.4. Συναλλαγές κρυπτονομισμάτων με τεχνολογία Blockchain.....	15
1.5. Πλεονεκτήματα και μειονεκτήματα κρυπτονομισμάτων.....	16
1.5.1. Πλεονεκτήματα και δυνατότητες.....	17
1.5.2. Μειονεκτήματα και κίνδυνοι .....	18
1.6. Ανταλλακτήρια Κρυπτονομισμάτων.....	19
<b>Κεφάλαιο 2<sup>ο</sup>: Μηχανική Μάθηση .....</b>	<b>22</b>
2.1. Ορισμός Μηχανικής Μάθησης .....	23
2.2. Σκοπός της Μηχανικής Μάθησης .....	25
2.3. Είδη Μηχανικής Μάθησης.....	26
2.3.1. Μάθηση με επίβλεψη (supervised learning).....	27
2.3.2. Μάθηση χωρίς επίβλεψη (unsupervised learning) .....	29
2.3.3. Μάθηση με ενίσχυση (reinforcement learning).....	30
2.4. Αλγόριθμοι Μηχανικής Μάθησης .....	31
2.4.1. Αλγόριθμος Naive Bayes.....	31
2.4.2. Αλγόριθμος K-means .....	32
2.4.3. Διαδικασίες αποφάσεων Markov .....	34
<b>Κεφάλαιο 3<sup>ο</sup>: Χρονοσειρές .....</b>	<b>35</b>
3.1. Ορισμός Χρονοσειράς.....	36
3.2. Κατηγοριοποίηση Χρονοσειρών .....	36
3.3. Ποιοτικά χαρακτηριστικά χρονοσειρών .....	41
3.4. Αποσύνθεση Χρονοσειρών .....	43
3.5. Ανάλυση Χρονοσειρών (Time Series Analysis).....	45
3.5.1. Μελέτη και ανάλυση των δεδομένων.....	45
3.5.2. Προεπεξεργασία δεδομένων.....	46

3.5.3. Μοντελοποίηση και Πρόβλεψη .....	46
3.6. Στασιμότητα Χρονοσειράς.....	47
3.6.1. Έλεγχος ADF (Augmented Dickey-Fuller test – ADF) .....	47
3.6.2. Έλεγχος KPPS (Kwiatkowski–Phillips–Schmidt–Shin) .....	48
3.7. Στατιστικά μεγέθη χρονοσειράς.....	49
3.7.1. Μέση τιμή (mean value).....	49
3.7.2. Συνδιακύμανση (Covariance) και Αυτοσυνδιακύμανση (Autocovariance).....	49
3.7.3. Αυτοσυσχέτιση (Autocorrelation) .....	50
3.7.4. Μερική Αυτοσυσχέτιση (Partial autocorrelation) .....	51
<b>Κεφάλαιο 4<sup>ο</sup>: Μοντέλα Πρόβλεψης Χρονοσειρών.....</b>	<b>52</b>
4.1. Στοχαστικά μοντέλα.....	52
4.1.1. Λευκός θόρυβος (White noise) .....	52
4.1.2. Τυχαίος περίπατος (Random walk) .....	53
4.2. Γραμμικά μοντέλα.....	54
4.2.1. Γραμμική Παλινδρόμηση (Linear Regression) .....	54
4.2.3. Απλή Γραμμική Παλινδρόμηση ( Simple Linear Regression).....	55
4.2.4. Αυτοπαλίνδρομο Μοντέλο (Auto Regressive Model) AR (p) .....	55
4.2.5. Μοντέλο Κινητού Μέσου (Moving Average).....	56
4.2.6. Αυτοπαλίνδρομο Μοντέλο Κινητού Μέσου Όρου (ARMA).....	57
4.2.7. Μικτό ολοκληρωμένο μοντέλο ARIMA (Autoregressive Integrated Moving Average model) .....	58
4.3. Τεχνητά Νευρωνικά Δίκτυα (ANN) .....	59
4.3.1. Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks).....	62
4.3.2. Νευρωνικά Δίκτυα LSTM (Long Short-Term Memory) .....	63
4.4. Μηχανές Διανυσματικής Υποστήριξης (Support Vector Machines).....	66
4.5. Αξιολόγηση Μοντέλων Πρόβλεψης .....	68
4.5.1. Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Square Error, RMSE).....	69
4.5.2 Μέσο Απόλυτο Σφάλμα (Mean Absolute Error, MAE).....	69
4.5.3 Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error, MAPE) .	69
4.5.4. Μέσο Τετραγωνικό Σφάλμα (Mean Square Error, MSE) .....	70
<b>Κεφάλαιο 5<sup>ο</sup>: Ερευνητικό Μέρος.....</b>	<b>71</b>
5.1. Βασική Ιδέα.....	71
5.2. Δεδομένα.....	71
5.3. Πλαίσιο Ανάπτυξης.....	72
5.4. Μεθοδολογία.....	72

5.5. Αρχική επεξεργασία των δεδομένων .....	73
5.5.1. Έλεγχος στασιμότητας χρονοσειράς .....	75
5.5.2. ACF και PACF .....	77
5.5.3. Μετασηματισμός χρονοσειράς .....	78
5.6. Μοντέλο ARIMA και αποτελέσματα.....	81
5.6.1 ARIMA (1,0,1) .....	83
5.6.2. ARIMA (0,1,0) .....	86
5.6.3. ARIMA (2,0,1) .....	87
5.7. Μοντέλο LSTM και αποτελέσματα .....	88
5.8. Μοντέλο SVM και αποτελέσματα .....	93
<b>Συμπεράσματα και σχόλια .....</b>	<b>96</b>
<b>Προτάσεις για περαιτέρω μελέτη .....</b>	<b>97</b>
<b>Βιβλιογραφικές αναφορές .....</b>	<b>99</b>

### Κατάλογος πινάκων

Εικόνα 1 Δομή και τρόπος λειτουργίας Blockchain .....	13
Εικόνα 2 Τύποι Blockchain .....	15
Εικόνα 3 Τρόπος συναλλαγής Blockchain.....	16
Εικόνα 4 3-month and 6-month Bitcoin volatilities.....	19
Εικόνα 5 Ποσοστά συναλλαγών ανά ανταλλακτήριο.....	21
Εικόνα 6 Εξόρυξη δεδομένων.....	24
Εικόνα 7 Είδη Μηχανικής Μάθησης .....	26
Εικόνα 8 Μάθηση με επίβλεψη (supervised learning).....	27
Εικόνα 9 Μάθηση χωρίς επίβλεψη (unsupervised learning) .....	29
Εικόνα 10 Μάθηση με ενίσχυση (reinforcement learning).....	30
Εικόνα 11 Αλγόριθμος K-means .....	33
Εικόνα 12 Διαδικασίες αποφάσεων Markov .....	35
Εικόνα 13 Ετήσιες ποσοστιαίες μεταβολές του παγκόσμιου ΑΕΠ .....	38
Εικόνα 14 Μεταβολή τιμής Bitcoin χρονικής διάρκειας 2018-2021 .....	38
Εικόνα 15 Ποσοστό ανεργίας στην Ελλάδα για το μήνα Ιούνιο για χρονική διάρκεια 2009-2021.....	39
Εικόνα 16 Αύξηση τιμής φυσικού αερίου .....	39
Εικόνα 17 Πωλήσεις υβριδικών-ηλεκτρικών αυτοκινήτων (2011-2015).....	40
Εικόνα 18 Μεταβολή της τιμής της ατμοσφαιρικής πίεσης στη Θεσσαλονίκη.....	40
Εικόνα 19 Τάση .....	41
Εικόνα 20 Εποχικότητα .....	42
Εικόνα 21 Κυκλικότητα.....	42
Εικόνα 22 Τυχαία σφάλματα .....	43
Εικόνα 23 Αυτοσυσχέτιση.....	50
Εικόνα 24 Μερική Αυτοσυσχέτιση.....	51
Εικόνα 25 Διάγραμμα Λευκού Θορύβου.....	53
Εικόνα 26 Διάγραμμα Τυχαίου Περίπατου.....	54
Εικόνα 27 Απλή Γραμμική Παλινδρόμηση .....	55
Εικόνα 28 Δομή Τεχνητού Νευρωνικού Δικτύου.....	60

Εικόνα 29 Τρόπος λειτουργίας Τεχνητού Νευρωνικού Δικτύου.....	62
Εικόνα 30 Τρόπος λειτουργίας Αναδρομικών Νευρωνικών Δικτύων.....	63
Εικόνα 31 Δίκτυο LSTM.....	64
Εικόνα 32 Διάγραμμα SVM.....	66
Εικόνα 33 Διάγραμμα επιπέδων SVM.....	68
Εικόνα 34 Φόρτωση Δεδομένων στο προγραμματιστικό περιβάλλον.....	73
Εικόνα 35 Έλεγχος δεδομένων για ακραίες, λανθασμένες ή εκλιπόντες τιμές.....	74
Εικόνα 36 Οπτικοποίηση Δεδομένων.....	74
Εικόνα 37 Αποσύνθεση χρονοσειράς.....	75
Εικόνα 38 Έλεγχος ADF.....	76
Εικόνα 39 Έλεγχος KPSS.....	76
Εικόνα 40 Διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης.....	77
Εικόνα 41 Αρχική μορφή δεδομένων.....	78
Εικόνα 42 Διαφοροποιημένη μορφή δεδομένων.....	79
Εικόνα 43 Έλεγχος ADF.....	79
Εικόνα 44 Έλεγχος KPSS.....	80
Εικόνα 45 Διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης δεδομένων.....	81
Εικόνα 46 Αποτελέσματα ARIMA.....	82
Εικόνα 47 Σφάλματα Πρόγνωσης.....	83
Εικόνα 48 Κατανομή σφαλμάτων.....	84
Εικόνα 49 Γραφικές παραστάσεις ACF και PACF.....	84
Εικόνα 50 Διάγραμμα πραγματικών και προβλεπόμενων τιμών Bitcoin.....	85
Εικόνα 51 Πραγματικές και προβλεπόμενες τιμές.....	85
Εικόνα 52 Διάγραμμα πραγματικών και προβλεπόμενων τιμών Bitcoin.....	86
Εικόνα 53 Πραγματικές και προβλεπόμενες τιμές.....	87
Εικόνα 54 Διάγραμμα πραγματικών και προβλεπόμενων τιμών Bitcoin.....	87
Εικόνα 55 Πραγματικές και προβλεπόμενες τιμές.....	88
Εικόνα 56 Διάγραμμα απώλειας εκπαίδευσης- ελέγχου.....	90
Εικόνα 57 Παραμετροποίηση LSTM.....	91
Εικόνα 58 Εκπαίδευση Δεδομένων.....	91
Εικόνα 59 Διάγραμμα πραγματικών και προβλεπόμενων τιμών LSTM.....	92
Εικόνα 60 Πραγματικές και προβλεπόμενες τιμές για 7 ημέρες.....	93
Εικόνα 61 Παραμετροποίηση SVM.....	94
Εικόνα 62 Δείκτης ακρίβειας SVM.....	94
Εικόνα 63 Προβλεπόμενες τιμές SVM.....	95
Εικόνα 64 Διάγραμμα πραγματικών και προβλεπόμενων τιμών SVM.....	95
Εικόνα 65 Δείκτης αξιολόγησης μοντέλου SVM.....	96

## Ευχαριστίες

Στη διαδρομή της συγγραφής της παρούσας διπλωματικής εργασίας είχα καθοδηγητή τον Καθηγητή μου, Μιχαήλ Φιλιππάκη, τον οποίο θα ήθελα να ευχαριστήσω θερμά, για τις καίριες υποδείξεις του και για όλα όσα μου έμαθε στην πορεία μου στο Μεταπτυχιακό Πρόγραμμα Σπουδών.

Ευχαριστώ τη γυναίκα μου, τους γονείς μου και τους φίλους μου για την ψυχολογική συμπαράσταση και την υπομονή τους.

## Περίληψη

Στα πλαίσια της παρούσας διπλωματικής εργασίας γίνεται βιβλιογραφική ανασκόπηση ερευνών, οι οποίες χρησιμοποιούν αλγορίθμους μηχανικής μάθησης και τεχνητά νευρωνικά δίκτυα για την πρόβλεψη του δείκτη τιμών κρυπτονομισμάτων και συγκεκριμένα του Bitcoin.

Γίνεται αναφορά για την ιστορική αναδρομή των κρυπτονομισμάτων, για τον τρόπο λειτουργίας τους, καθώς και για την τεχνολογία και τους μηχανισμούς που κρύβονται πίσω από αυτήν. Περιγράφονται με κάθε λεπτομέρεια οι παράγοντες που επηρεάζουν την πορεία των κρυπτονομισμάτων και τις συνέπειες που έχουν αυτές οι μεταβολές στην έντονη διακύμανση της τιμής τους.

Αναλύεται ο τρόπος λειτουργίας των αλγορίθμων μηχανικής μάθησης και των τεχνητών νευρωνικών δικτύων που χρησιμοποιούνται για την πρόβλεψη του συγκεκριμένου δείκτη. Στην συνέχεια επεξηγείται ο τρόπος με τον οποίο γίνεται η συλλογή, η προεπεξεργασία και η μοντελοποίηση των δεδομένων με σκοπό την χρήση τους στους αλγόριθμους μηχανικής μάθησης που εξετάζονται.

Τέλος γίνεται σύγκριση των αποτελεσμάτων που παράγονται από τα μοντέλα πρόβλεψης και παραθέτονται τα συμπεράσματα που προκύπτουν για συζήτηση.

**Λέξεις – Κλειδιά:** Κρυπτονομίσματα, Μηχανική Μάθηση, Χρονοσειρές, Πρόβλεψη, Τεχνητά Νευρωνικά Δίκτυα

## **Abstract**

In the framework of this Thesis, a bibliographic review of researches, machine learning algorithm devices and artificial neural networks for the prediction of the cryptocurrency price index, specifically Bitcoin, is carried out.

Reference is made to the historical background of cryptocurrencies, how they work as well as the technology and mechanisms behind it. The factors that influence the course of cryptocurrencies and the consequences that these changes have on the sharp fluctuation of their price are described in detail.

The operation of machine learning algorithms and artificial neural networks used to predict this indicator is analyzed. The following explains how data is collected, pre-processed and modeled for use in the machine learning algorithms under consideration.

Finally, the results produced by the forecasting models are compared and the conclusions that emerge for discussion are presented.

**Keywords:** Cryptocurrencies, Machine Learning, Time Series, Prediction, Artificial Neural Networks



## Εισαγωγή

Η παρούσα διπλωματική εργασία έχει ως αντικείμενο μελέτης και ανάλυσης την πρόβλεψη χρονοσειρών με μεθόδους μηχανικής μάθησης και νευρωνικά δίκτυα, με απώτερο σκόπο την εφαρμογή των διάφορων μοντέλων στην πρόβλεψη μελλοντικών τιμών του κρυπτονομίσματος Bitcoin.

Στο δεύτερο κεφάλαιο γίνεται μια λεπτομερής αναφορά για την ιστορική αναδρομή των κρυπτονομισμάτων, τον τρόπο λειτουργίας και τους μηχανισμούς που συμβάλουν σε αυτή. Περιγράφεται αναλυτικά η τεχνολογία Blockchain και αναλύεται ο τρόπος με τον οποίο γίνονται οι συναλλαγές με κρυπτονομισμάτα. Επίσης γίνεται λόγος για τους παράγοντες που μπορεί να επηρεάσουν την τιμή τους είτε μακροπρόθεσμα είτε βραχυπρόθεσμα.

Στο τρίτο κεφάλαιο μελετώνται οι βασικές αρχές της Μηχανικής Μάθησης. Αναλύονται όλα τα είδη της Μηχανικής Μάθησης και παρουσιάζονται παραδείγματα αλγορίθμων που χρησιμοποιούνται στην καθημερινότητα για τη λήψη σημαντικών αποφάσεων. Πλέον η Μηχανική Μάθηση παίζει καθοριστικό ρόλο στην λειτουργία πολλών επαγγελματικών κλάδων, όπως Στατιστική, Ιατρική κτλ.

Στο τέταρτο κεφάλαιο δίνεται ιδιαίτερη βαρύτητα στο θεωρητικό πλαίσιο γύρω από τις χρονοσειρές, οι οποίες αποτελούν την κύρια μέθοδο πρόβλεψης. Γίνεται λεπτομερής αναφορά στα πλεονεκτήματα των χρονοσειρών καθώς επίσης και στα μοντέλα πρόβλεψης χρονοσειρών που πρόκειται να χρησιμοποιηθούν για τους σκοπούς της διπλωματικής εργασίας. Τα μοντέλα που χρησιμοποιούνται είναι το στατιστικό μοντέλο ARIMA, το τεχνητό νευρωνικό δίκτυο LSTM και το μοντέλο SVM.

Στο πέμπτο κεφάλαιο γίνεται εφαρμογή των μοντέλων αυτών πάνω στα δεδομένα της χρονοσειράς που εξετάζεται με σκοπό την πρόβλεψη μελλοντικών τιμών. Περιγράφεται αναλυτικά η συλλογή, η προεπεξεργασία και η μοντελοποίηση των δεδομένων καθώς και τη διαδικασία που ακολουθείται μέχρι την παραγωγή και εξήγηση των αποτελεσμάτων.

Στο έκτο και έβδομο κεφάλαιο παρουσιάζονται τα συμπεράσματα και σχόλια που προκύπτουν από την ανάλυση των αποτελεσμάτων και παραθετονται προτάσεις για μελλοντική μελέτη και ανάλυση.

Οι κύριοι στόχοι της παρούσας διπλωματικής εργασίας είναι οι εξής:

- ✓ Η μελέτη των βασικών εννοιών που αφορούν τη Μηχανική Μάθηση, τους αλγόριθμους και τα χαρακτηριστικά τους. Η ανάλυση του δείκτη που θα εξεταστεί καθώς επίσης και ο τρόπος και τα εργαλεία που θα χρησιμοποιηθούν με σκοπό την πρόβλεψη του. Με αυτόν

τον τρόπο προκύπτουν όλα τα στοιχεία που χρειάζονται για την ανάλυση των χρονοσειρών.

✓ Η μελέτη των βασικών εννοιών που αφορούν τις χρονοσειρές και τα μοντέλα πρόβλεψης που χρησιμοποιούνται για την ερευνητική διαδικασία.

✓ Η εφαρμογή των μοντέλων χρονοσειρών και τεχνητών νευρωνικών δικτύων στο πρακτικό κομμάτι της εργασίας.

Τα παραπάνω θα συμβάλουν σημαντικά στην απόκτηση εξειδικευμένης γνώσης και στη δημιουργία ισχυρών βάσεων για την συνέχεια.

## **Κεφάλαιο 1<sup>ο</sup>: Κρυπτονομίσματα**

### **1.1. Ορισμός**

Τα ψηφιακά νομίσματα, γνωστά και ως κρυπτονομισμάτα, είναι πλέον από τα πιο διαδεδομένα στη σημερινή εποχή. Κάθε κρυπτονόμισμα είναι μια peer to peer αποκεντρωμένη ηλεκτρονική μορφή περιουσιακού στοιχείου που χρησιμοποιείται ως μέσο ανταλλαγής (νόμισμα) και ελέγχεται εξ ολοκλήρου από τους ιδιοκτήτες του. Βασίζεται στις αρχές της κρυπτογραφίας, δηλαδή δημιουργείται μέσω λογισμικού που κρυπτογραφεί δεδομένα συναλλαγών για τη διασφάλιση του δικτύου στο οποίο πραγματοποιούνται αυτές καθώς και για τον έλεγχο και την επιβεβαίωση της μεταφοράς νομισμάτων. Τα περισσότερα είδη κρυπτονομισμάτων κάνουν χρήση μιας κατανεμημένης βάσης δεδομένων ως τον πυλώνα του συστήματός τους, το γνωστό Blockchain. Βρίσκονται πέρα από την εμβέλεια των κυβερνήσεων και των διαμεσολαβητών, δεν εκδίδονται από κάποια κεντρική τράπεζα, επιτρέπουν το ομότιμο εμπόριο, τις άμεσες πληρωμές, τις αποταμιεύσεις και πολλά άλλα. Έχει τη μορφή χρήματος για τους ανθρώπους με αποτέλεσμα όσοι περισσότεροι άνθρωποι το χρησιμοποιούν, τόσο πιο χρήσιμο γίνεται.

### **1.2. Ιστορική αναδρομή**

Πριν εμφανιστούν οι σύγχρονες μορφές κρυπτονομισμάτων, έγιναν πολλές απόπειρες να δημιουργηθεί κάτι αντίστοιχο κατά τη δεκαετία του 1980 στην Ολλανδία και στις Ηνωμένες Πολιτείες, όπως τα B-Money, Hashcash, Flooz και Bit Gold, χωρίς όμως να κυκλοφορήσουν επίσημα. Το 1990, ο Αμερικανός κρυπτογράφος David Chaum, ένας επιστήμονας υπολογιστών και μαθηματικός, δημιούργησε το πρώτο αξιοσημείωτο ψηφιακό νόμισμα ως επέκταση ενός αλγόριθμου κρυπτογράφησης που θεωρήθηκε δημοφιλής κατά τη διάρκεια εκείνης της εποχής, το Digicash, το οποίο όμως δεν διαδόθηκε. Στη συνέχεια, η γνωστή σε όλους εταιρεία PayPal, όπως και οι ανταγωνιστές της που εμφανίστηκαν αργότερα, κατάφεραν να εξελίξουν τις ψηφιακές συναλλαγές χρησιμοποιώντας τα ήδη υπάρχοντα νομίσματα. Οι εταιρείες αυτές, εξακολουθούν να διαδραματίζουν πρωταγωνιστικό ρόλο στις διαδικτυακές συναλλαγές και στο διεθνές εμπόριο. Η αξιοπιστία είναι εκείνη που καθορίζει την επιτυχία ενός κρυπτονομίσματος, δηλαδή να υπάρχει ασφάλεια για τον κάτοχο του σε περιπτώσεις παραποιήσεων και υπεξαίρέσεων των νομισμάτων. Τα ψηφιακά περιουσιακά στοιχεία είναι πολύ δύσκολο να ασφαλιστούν στον ψηφιακό κόσμο και αυτό συμβαίνει διότι μία πληροφορία

είναι πολύ εύκολο να αντιγραφεί και να αναπαραχθεί. Για το λόγο αυτό η δημιουργία ενός ψηφιακού νομίσματος απαιτεί και την δημιουργία μιας ψηφιακής ταυτότητας η οποία δεν μπορεί να αντιγραφεί και ανήκει σε ένα μόνο πιστοποιημένο κάτοχο. Όπως μια τράπεζα φυλάσσει τα χρηματικά ποσά που διαθέτει ένας πολίτης, είτε σε τραπεζικές θυρίδες είτε σε τραπεζικούς λογαριασμούς, τα οποία δεν μπορεί να τα αντιγράψει για να χρησιμοποιήσει τα αντίγραφα, το ίδιο πρέπει να συμβεί και με τα ψηφιακά νομίσματα. Να δημιουργηθούν δηλαδή συστήματα υψίστης ασφαλείας για να γίνεται χρήση των ψηφιακών νομισμάτων.

Με την πάροδο του χρόνου και την ραγδαία εξέλιξη της τεχνολογίας, το 2008 δημιουργείται το κρυπτονόμισμα Bitcoin, από τον Satoshi Nakamoto, πιθανόν ψευδώνυμο το οποίο αναφέρεται είτε σε ένα άτομο είτε σε μια ομάδα ανθρώπων. Η προέλευσή του παραμένει ακόμα μυστηριώδης. Ο Satoshi Nakamoto δημοσίευσε τη λευκή βίβλο, η οποία περιλαμβάνει μία περιγραφή υλοποίησης του συγκεκριμένου κρυπτονομίσματος. Το Bitcoin λειτουργεί μια νέα τεχνολογία που ονομάζεται Blockchain, η οποία είναι γνωστή και ως σύστημα «τριπλής λογιστικής καταχώρησης». Σε κάθε συναλλαγή που πραγματοποιείται θα πρέπει ο αποστολέας, ο παραλήπτης και ένας τρίτος να επιβεβαιώνουν και να συμφωνούν σχετικά με τα στοιχεία της συγκεκριμένης συναλλαγής. Αυτό προσφέρει εμπιστοσύνη σε επίπεδο ανωνυμίας, καθώς όλες οι συναλλαγές σε ένα συγκεκριμένο πορτοφόλι Bitcoin, μπορεί να εντοπιστεί, χωρίς όμως να γίνεται γνωστό το όνομα κατόχου του. Η τεχνολογία Blockchain έπαιξε σημαντικό ρόλο στην ανάπτυξη πολλών κρυπτονομισμάτων και εφαρμογών, τα οποία με τη σειρά τους κατάφεραν να κάνουν έντονα αισθητή την παρουσία τους στην αγορά. Η συνολική αξία των Bitcoin, γνωστή και ως «market cap», ξεπέρασε το 1 τρισεκατομμύριο δολάρια τον Μάρτιο του 2021. Το Bitcoin είναι το πιο διαδεδομένο κρυπτονόμισμα παγκοσμίως, η εμπορική του αξία αν και εξαιρετικά ασταθής λόγω σημαντικών μεταβολών, παραμένει σε εξαιρετικά υψηλά επίπεδα.

Παράλληλα με την αύξηση της δημοτικότητας του Bitcoin, δημιουργήθηκαν και άλλα κρυπτονομισμάτα που κατέλαβαν και αυτά σημαντικό μερίδιο της αγοράς. Σύμφωνα με τις τελευταίες μεγαλύτερες κεφαλαιοποιήσεις για συναλλαγές κρυπτονομισμάτων στο δημοφιλές χρηματιστήριο Coinbase, τα κρυπτονομισμάτα αυτά είναι τα εξής:

- Ethereum
- Binance Coin
- Tether
- Cardano
- Polkadot

- XRP
- Uniswap
- Litecoin
- THETA3

Στα κρυπτονομισμάτα αυτά παρατηρείται συχνή μεταβολή στην εμπορική αξία τους και αυτό οφείλεται στη ζήτηση της αγοράς, όπως το Bitcoin και νομίσματα που συνδέονται με το δολάριο ΗΠΑ. Πολλοί αναλυτές θεωρούν ότι η ξαφνική άνοδος των κρυπτονομισμάτων αποτελεί απειλή για τα συμβατικά νομίσματα και την παγκόσμια οικονομία. Ουσιαστικά χρησιμοποιούν έναν peer-to-peer μηχανισμό (δίκτυο συνδεδεμένων-συνεργαζόμενων υπολογιστών) που καθιστούν μη αναγκαία την ύπαρξη του μεσάζοντα (τράπεζες ή άλλα χρηματοπιστωτικά ιδρύματα) για την διεκπεραίωση χρηματοοικονομικών συναλλαγών. Δεν χρειάζεται κάποιος τραπεζικός λογαριασμός ή πιστωτική κάρτα για να πραγματοποιηθούν συναλλαγές με κρυπτονομισμάτα.

### **1.3. Blockchain**

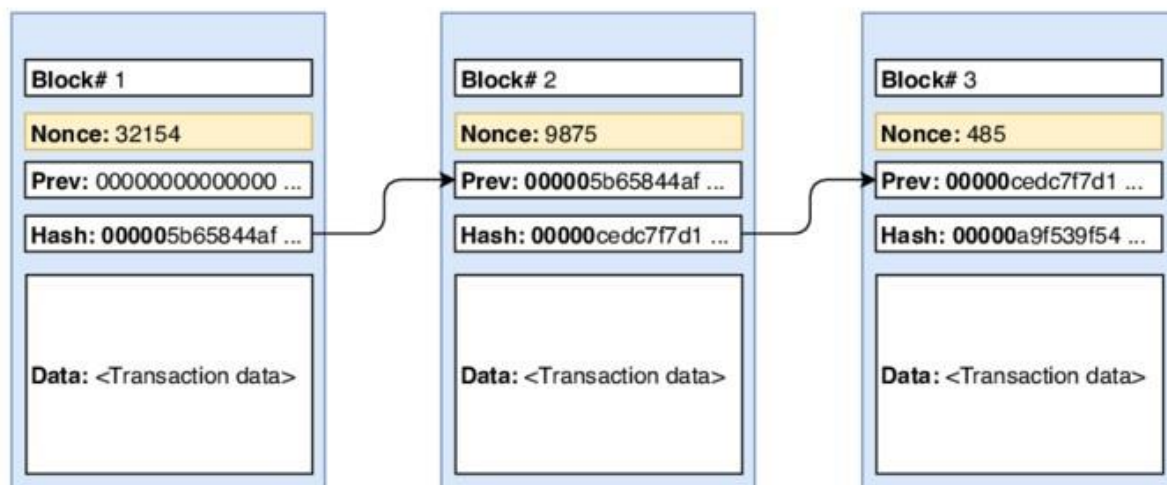
Τα παράγωγα τεχνολογικών εξελίξεων τα οποία σηματοδοτούν καθημερινά έντονες και καινοτόμες διαφοροποιήσεις αποτελούν αντικείμενο συζήτησης και ερευνών στην επιστημονική κοινότητα. Η αποδοχή και εφαρμογή τους από την κοινωνία επισημαίνουν τον σπουδαίο ρόλο που πρόκειται να διαδραματίσουν στις τεχνολογικές εξελίξεις. Ένα τέτοιο παράγωγο της εξέλιξης της επιστήμης των Ηλεκτρονικών Υπολογιστών αποτελεί η τεχνολογία του Blockchain.

Η τεχνολογία Blockchain, όπως και οι περισσότερες τεχνολογικές εξελίξεις, αναπτύχθηκε με γνώμονα τις ήδη υπάρχουσες τεχνολογίες. Ανακαλύφθηκε από τον David Chaum το 1982 (Sherman et al., 2019) σε πολύ πρώιμο στάδιο. Το 1991 μελετήθηκε και εξελίχθηκε περαιτέρω από τους επιστήμονες Haber και Stornetta (Narayanan et al., 2016), με αποτέλεσμα να τελειοποιηθεί και να εφαρμοστεί στην βελτιωμένη του έκδοση σήμερα από τον Satoshi Nakamoto το 2008 (Nakamoto, 2008). Ο τελευταίος κατάφερε να βελτιώσει αρκετά το σχεδιασμό του χρησιμοποιώντας τη δική του τεχνική στην εφαρμογή των ψηφιακών νομισμάτων και πιο συγκεκριμένα του Bitcoin.

Πιο αναλυτικά, η τεχνολογία Blockchain είναι μια αποκεντρωμένη, κατακεντρωμένη βάση δεδομένων στην οποία, με τη βοήθεια των μαθηματικών και της κρυπτογραφίας, καταγράφεται και διατηρείται μία συνεχώς αυξανόμενη λίστα καταχωρήσεων, που

ονομάζονται μπλοκ (συστοιχίες) και συνδέονται μεταξύ τους σαν μια αλυσίδα (chain). Επομένως, όταν αναφερόμαστε στις παραπάνω έννοιες, στην πραγματικότητα μιλάμε για ψηφιακές πληροφορίες συναλλαγών (μπλοκ) οι οποίες είναι αποθηκευμένες σε μια δημόσια βάση δεδομένων (αλυσίδα).

Κάθε μπλοκ περιέχει ένα ψηφιακό δακτυλικό αποτύπωμα ή μοναδικό αναγνωριστικό (hash), τη χρονική σφραγίδα των πρόσφατων συναλλαγών και το αναγνωριστικό (hash) του προηγούμενου μπλοκ. Έτσι τα μπλοκ συνδέονται μεταξύ τους με τέτοιο τρόπο ώστε κάθε επόμενο μπλοκ να επαληθεύει το προηγούμενο και επομένως ολόκληρου του δικτύου και έτσι εμποδίζεται η οποιαδήποτε αλλαγή. Μόνο τότε οι συναλλαγές θεωρούνται έγκυρες. Οι συνεχόμενες λοιπόν καταχωρήσεις δεδομένων αφορούν συναλλαγές μεταξύ χρηστών (χρήματα, αγαθά, ακίνητη περιουσία, ψηφιακά νομίσματα, εργασία κτλ.) σε δημόσια η ιδιωτικά (peer-to-peer) δίκτυα, όπου κάθε φορά μία καινούρια ομάδα καταχωρήσεων συνδέεται με τις προηγούμενες, δημιουργώντας έτσι μια αλυσίδα δεδομένων από την αρχική συναλλαγή έως την τρέχουσα. Όλες οι συναλλαγές που αφορούν ένα μπλοκ αποστέλλονται σε όλους τους κόμβους (miners) του δικτύου, το οποίο επικυρώνει τη δημιουργία του νέου μπλοκ (Laurence, 2017).



Εικόνα 1 Δομή και τρόπος λειτουργίας Blockchain

Όταν πραγματοποιείται μια συναλλαγή μεταξύ δύο χρηστών, τότε οι διαδικασίες απαιτούν την επίβλεψη κάποιου τρίτου για την εδραίωση της εμπιστοσύνης, όπως για παράδειγμα μία τράπεζα ή ένα δικηγορικό γραφείο, με αποτέλεσμα το υψηλό κόστος, την απάτη, την αναξιοπιστία και την ανασφάλεια. Η επαναστατική τεχνολογία Blockchain αλλάζει ριζικά το περιβάλλον των συναλλαγών. Προσφέρει ασφαλείς, άμεσες συναλλαγές, χωρίς την

διαμεσολάβηση τρίτων, συνδυάζοντας απλά την κρυπτογραφία και τα κατανεμημένα υπολογιστικά συστήματα.

### 1.3.1. Τύποι Blockchain

Ένα από τα πιο βασικά χαρακτηριστικά δικτύου Blockchain είναι ο τύπος του. Υπάρχουν δύο τύποι δικτύου, ανάλογα με την άδεια που έχουν οι χρήστες. Χωρίζονται στα δίκτυα που λειτουργούν χωρίς κάποια συγκεκριμένη άδεια (public) και σε αυτά που λειτουργούν με άδεια (private) (Jayachandran, 2017).

Αναλυτικότερα:

**1. Public – Δημόσια:** Τα δημόσια δίκτυα Blockchain είναι τα δίκτυα που λειτουργούν χωρίς άδεια. Διαθέτουν ένα αποκεντρωμένο ψηφιακό δημόσιο σημειωματάριο, το οποίο είναι προσβάσιμο για κάθε χρήστη (προγραμματιστές, miners κτλ.) που θέλει να πραγματοποιήσει μια συναλλαγή χωρίς να χρειάζεται η άδεια και η εποπτεία από κάποια ελεγκτική αρχή. Το δημόσιο δίκτυο Blockchain είναι συνήθως ανοιχτού κώδικα (open-source) και ελεύθερος σε όποιον επιθυμεί να συμμετάσχει. Μπορεί οποιοσδήποτε να δει το ιστορικό των συναλλαγών του δικτύου και να προχωρήσει στην έκδοση μπλοκ. Παρόλο όμως που η πρόσβαση είναι ελεύθερη, υπάρχει μεγάλος κίνδυνος να πραγματοποιηθούν απόπειρες αλλοίωσης των δεδομένων των συναλλαγών ώστε να επωφελούνται συγκεκριμένοι χρήστες. Για το λόγο αυτό δημιουργούνται διάφορα πρωτόκολλα συναίνεσης μεταξύ των χρηστών για την πιο ομαλή λειτουργία του δικτύου (Jayachandran, 2017).

**2. Private – Ιδιωτικά:** Τα ιδιωτικά δίκτυα Blockchain είναι αυτά που λειτουργούν με άδεια. Η συμμετοχή των χρηστών στα συγκεκριμένα δίκτυα απαιτεί την παροχή άδειας από κάποια κεντρική αρχή, συγκεντρωμένη ή αποκεντρωμένη. Πρόκειται για κλειστού τύπου δίκτυα στα οποία δίνονται δικαιώματα στους χρήστες για τις ενέργειες που θέλουν να κάνουν, όπως να μπορούν να δουν τις συναλλαγές που έχουν ήδη πραγματοποιηθεί ή να δημιουργήσει καινούρια μπλοκ. Όπως και στα δημόσια δίκτυα, έτσι και εδώ υπάρχουν πρωτόκολλα συναίνεσης μεταξύ των χρηστών, τα οποία είναι διαφοροποιημένα μεταξύ τους. Τα ιδιωτικά δίκτυα Blockchain εφαρμόζονται κυρίως σε επιχειρήσεις που διαχειρίζονται ευαίσθητα δεδομένα, όπως για παράδειγμα κατασκευαστικές επιχειρήσεις, η δημόσια ασφάλεια, ο στρατός, κρατικές επιχειρήσεις που σχετίζονται με την φορολογία αλλά και με θέματα επιβολής του νόμου, και με αυτόν τον τρόπο αποφεύγεται η διαχείρισή τους από τρίτους. Αν και υπάρχει εμπιστοσύνη μεταξύ των χρηστών, πολλές φορές γίνονται προσπάθειες αλλοίωσης των δεδομένων των συναλλαγών χωρίς αποτέλεσμα, λόγω της υψηλής

ασφάλειας, της ιδιωτικότητας, της συμμόρφωσης, της απόδοσης και των κανόνων που προσφέρονται από αυτά τα δίκτυα (Jayachandran, 2017).

Public	Private
Open, anyone can join the network	Restricted and permissioned, a new member joins the network via invitation
Each node has equal transmission power (Distributed)	Only certain nodes can create new transactions
Low speed of transaction accomplishment	Fast speed of transaction accomplishment
Long transaction approval frequency	Short transaction approval frequency
High cost of each transaction	Comparatively cheap cost of each transaction
Proof-of-work, proof-of-stake consensus protocols for the adding on a new block	Pre-approved participants initiate adding of a new block
Anonymous	Nonymous
Requires no trust among the members	Members need to trust each other
Large energy consumption	Low energy consumption

Εικόνα 2 Τύποι Blockchain

#### 1.4. Συναλλαγές κρυπτονομισμάτων με τεχνολογία Blockchain

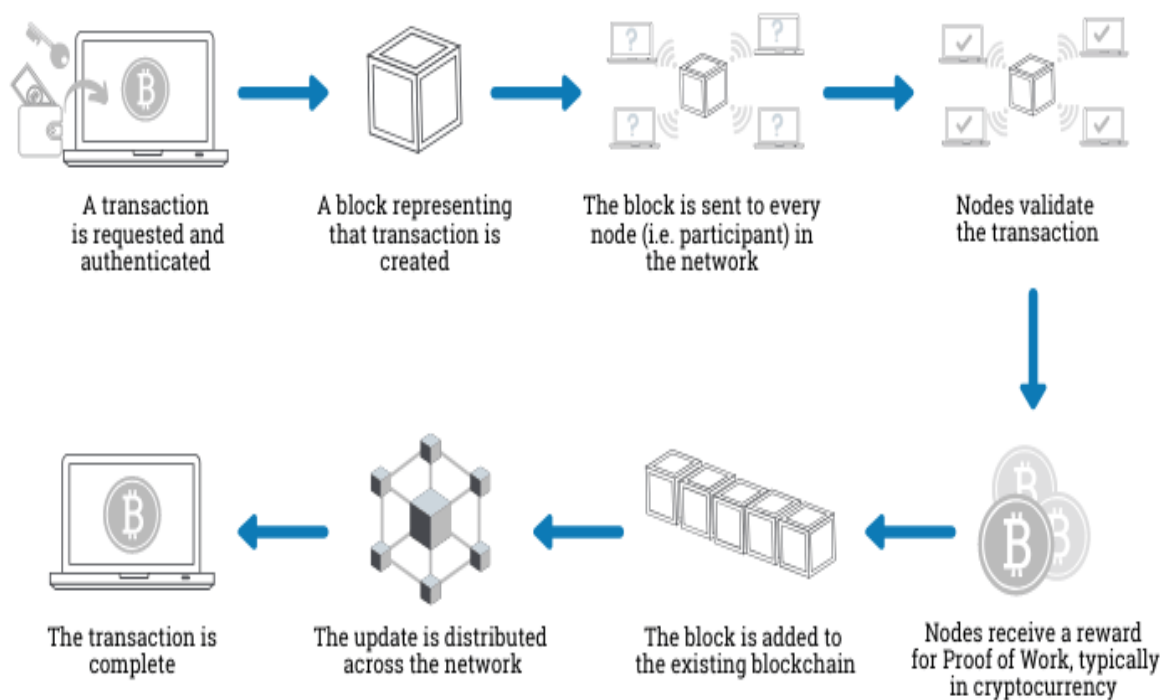
Ο τρόπος με τον οποίο πραγματοποιείται μια συναλλαγή με Blockchain είναι ο εξής:

- Ένα δίκτυο Blockchain χρησιμοποιεί δημόσια και ιδιωτικά κλειδιά για να δημιουργήσει μια ψηφιακή υπογραφή. Έτσι εξασφαλίζει την μοναδικότητα, την ασφάλεια και την εμπιστοσύνη.
- Γίνεται η ταυτοποίηση μέσω της ψηφιακής υπογραφής και στην συνέχεια απαιτείται η ανάγκη της εξουσιοδότησης.
- Όλα τα μέλη του δικτύου που συμμετέχουν πραγματοποιούν ένα μαθηματικό έλεγχο και συμφωνούν σε μία συγκεκριμένη τιμή.
- Κατά τη διάρκεια της μεταφοράς, ο αποστολέας χρησιμοποιεί το ιδιωτικό κλειδί του και γνωστοποιεί τις πληροφορίες της συναλλαγής στο δίκτυο. Η συναλλαγή που προκύπτει, καταχωρείται σε ένα μπλοκ και περιλαμβάνει τις εξής πληροφορίες : την ψηφιακή υπογραφή, τη χρονική σήμανση και το δημόσιο κλειδί του παραλήπτη.
- Το συγκεκριμένο μπλοκ πληροφοριών μεταδίδεται μέσω του δικτύου και ξεκινά η διαδικασία επικύρωσης.
- Οι ανθρακωρύχοι σε όλο το δίκτυο αρχίζουν να αναλύουν και να επεξεργάζονται τις πληροφορίες της συναλλαγής, βασισμένοι πάντα στην υπολογιστική τους δύναμη.



- Πρόκειται για ένα μαθηματικό παζλ που καλούνται να επιλύσουν. Στην πρώτη επίλυση του παζλ, οι ανθρακωρύχοι ανταμείβονται με την μορφή κρυπτονομισμάτων, ως απόδειξη εργασίας (Proof of Work).
- Στη συνέχεια, όταν οι περισσότεροι κόμβοι καταλήξουν σε συναίνεση, συμφωνούν σε μια κοινή λύση και έτσι το μπλοκ σφραγίζεται με το χρόνο και την ψηφιακή υπογραφή και προστίθεται στο ήδη υπάρχον μπλοκ.
- Μόλις γίνει η προσθήκη του νέου μπλοκ στην αλυσίδα, όλα τα δεδομένα του blockchain ενημερώνονται το ίδιο για όλους τους κόμβους στο δίκτυο. (Natarajan, 2018)

## How does a transaction get into the blockchain?



Εικόνα 3 Τρόπος συναλλαγής Blockchain

### 1.5. Πλεονεκτήματα και μειονεκτήματα κρυπτονομισμάτων

Σκοπός του συγκεκριμένου κεφαλαίου είναι να αναλύσει τα πλεονεκτήματα και τα μειονεκτήματα που απορρέουν από την ανάπτυξη και χρήση των κρυπτονομισμάτων. Όπως αναφέρθηκε σε προηγούμενο κεφάλαιο το ψηφιακό νόμισμα είναι πιο αποτελεσματικό από το φυσικό χρήμα που υποστηρίζουν οι κυβερνήσεις. Οι λόγοι που συμβαίνει αυτό ποικίλουν,

όπως υποστηρίζουν οι χρήστες των ψηφιακών νομισμάτων. Παρόλα αυτά όμως υπάρχουν και αρκετά μειονεκτήματα, τα οποία θα παρουσιάσουμε και θα αναλύσουμε παρακάτω.

### 1.5.1. Πλεονεκτήματα και δυνατότητες

Τα πλεονεκτήματα και οι δυνατότητες που προσφέρει η χρήση των κρυπτονομισμάτων είναι τα εξής:

- **Ελευθερία πληρωμών:** Οι χρήστες ψηφιακών νομισμάτων μπορούν να πραγματοποιήσουν τις συναλλαγές τους άμεσα χωρίς περιορισμούς και επιβαλλόμενα όρια. Με αυτόν τον τρόπο έχουν τον πλήρη έλεγχο των περιουσιακών τους στοιχείων.
- **Χαμηλό κόστος:** Το κόστος των συναλλαγών είναι είτε μηδενικό είτε πολύ χαμηλό από τη στιγμή που δεν υπάρχει κάποιος μεσάζων που πρέπει να ελέγχει και να διεκπεραιώνει κάθε φορά την οποιαδήποτε συναλλαγή.
- **Έλεγχος πληθωρισμού:** Πρόκειται για ένα φαινόμενο που εμφανίζεται όταν αυξάνεται το γενικό επίπεδο τιμών και υπηρεσιών σε μία οικονομία. Η αύξηση αυτή οδηγεί σε μείωση της αγοραστικής δυνατότητας του φυσικού χρήματος το οποίο ελέγχεται από μία κεντρική αρχή. Στην περίπτωση του ψηφιακού νομίσματος που δεν ελέγχεται, ο πληθωρισμός είναι αδύνατος.
- **Αποκεντρωμένη διαχείριση:** Το δίκτυο συναλλαγών κρυπτονομισμάτων είναι αποκεντρωμένο και δεν ελέγχεται από κάποια κεντρική αρχή με αποτέλεσμα να είναι προσβάσιμο σε όλους τους χρήστες, χωρίς κανόνες και περιορισμούς.
- **Ασφάλεια:** Όπως αναφέρθηκε προηγουμένως, όλοι οι χρήστες έχουν τον πλήρη έλεγχο των συναλλαγών τους. Κανένας χρήστης δεν μπορεί να προχωρήσει σε ανεπιθύμητες ενέργειες όπως συμβαίνει συχνά με άλλες με άλλες μεθόδους πληρωμής. Όλες οι πληρωμές με κρυπτονομισμάτα πραγματοποιούνται χωρίς να μεταφέρουν προσωπικά δεδομένα κατά τη συναλλαγή, με αποτέλεσμα να επιτυγχάνεται υψηλή προστασία και ασφάλεια. Οι χρήστες προστατεύουν τα περιουσιακά τους στοιχεία με αντίγραφα ασφαλείας και κρυπτογράφηση, αποκρύπτοντας τα πραγματικά τους στοιχεία.
- **Διαφάνεια:** Οι πληροφορίες που μεταφέρονται κατά τις συναλλαγές των κρυπτονομισμάτων είναι άμεσα διαθέσιμες στο δίκτυο για όλους τους χρήστες σε πραγματικό χρόνο. Κανένας όμως δεν μπορεί να παραποιήσει το πρωτόκολλο του δικτύου διότι είναι κρυπτογραφικά ασφαλές με αποτέλεσμα να ενισχύεται η αξιοπιστία και η διαφάνεια.

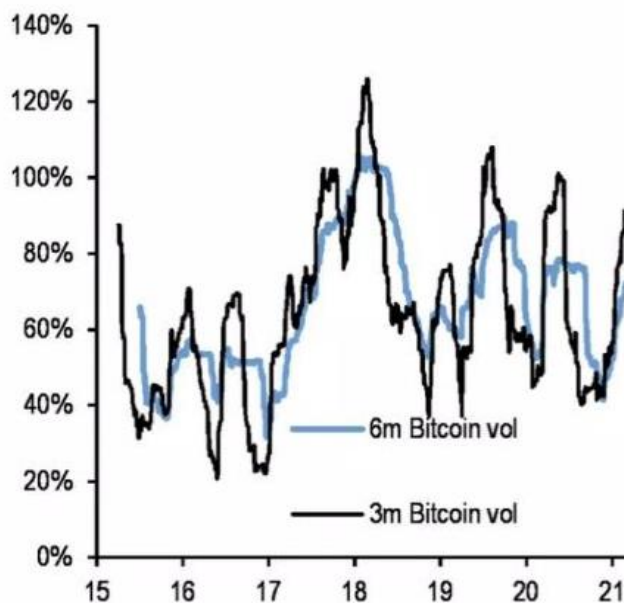
- **Μείωση του κινδύνου απάτης:** Πολλοί χρήστες προσπαθούν να εξαπατήσουν το σύστημα με διάφορους τρόπους προς όφελός τους. Η τεχνολογία που κρύβεται πίσω από την κρυπτογράφηση εξαλείφει κατά ένα μεγάλο ποσοστό τον κίνδυνο απάτης με αθέμιτα μέσα, επειδή όλες οι συναλλαγές που πραγματοποιούνται σε ένα δίκτυο blockchain είναι διαφανείς και αμετάβλητες.
- **Ταχύτητα συναλλαγών:** Παρατηρείται αύξηση της ταχύτητας διεκπεραίωσης των συναλλαγών, συνεπώς και μείωση του χρόνου αναμονής.
- **Υψηλά κέρδη:** Η χρήση κρυπτονομισμάτων καθώς και η επένδυση σε αυτά αποφέρουν σε όλους τους χρήστες υψηλά κέρδη.

### 1.5.2. Μειονεκτήματα και κίνδυνοι

Μπορεί τα πλεονεκτήματα των κρυπτονομισμάτων να είναι πολλά αλλά υπάρχουν φορές που η χρήση τους μπορεί να από φέρει κινδύνους και αξιολογούνται ως μία κακή εναλλακτική του φυσικού χρήματος. Παρακάτω ακολουθούν και αναλύονται τα μειονεκτήματα που εμφανίζονται κατά τη χρήση των κρυπτονομισμάτων:

- **Αποδοχή:** Η διάδοση των κρυπτονομισμάτων δεν έχει γίνει ακόμα γνωστή σε μεγάλο μερίδιο του παγκόσμιου πληθυσμού. Ολοένα και περισσότερες επιχειρήσεις δέχονται τα κρυπτονομισμάτα ως μέσο συναλλαγών, για να επωφεληθούν από τα πλεονεκτήματά τους.
- **Παράνομες δραστηριότητες:** Πολλές φορές τα κρυπτονομισμάτα έχουν χρησιμοποιηθεί για εγκληματικές δραστηριότητες, όπως διακίνηση μαύρου χρήματος, ναρκωτικών, κυβερνοεπιθέσεις και τρομοκρατικές οργανώσεις.
- **Αβεβαιότητα:** Η τιμή των κρυπτονομισμάτων διαμορφώνεται από η ζήτηση του συγκεκριμένου νομίσματος. Έχει μεγάλες διακυμάνσεις και χαρακτηρίζεται από έντονη μεταβλητότητα για κάποιο μέσο καταναλωτή που θέλει να χρησιμοποιήσει τα κρυπτονομισμάτα αντί του φυσικού χρήματος. Υπάρχουν πολλοί προσδιοριστικοί παράγοντες που επηρεάζουν την τιμή των κρυπτονομισμάτων. Η μεταβλητότητα αυτή, καθιστά αδύναμα τα κρυπτονομισμάτα να διατηρήσουν σταθερή την αξία τους και επομένως να αποτελέσουν σε βάθος χρόνων την κύρια μορφή χρήματος.
- **Περιβαλλοντικές συνέπειες:** Η τεχνολογία Blockchain που χρησιμοποιείται για την παραγωγή κρυπτονομισμάτων απαιτεί τόσο τη χρήση διαδικτύου όσο και τη χρήση υπολογιστών. Οι ανάγκες για πρόσβαση στο διαδίκτυο δεν είναι σημαντικές. Οι χρήστες

όμως τόσο για τη διαδικασία εξόρυξης κρυπτονομισμάτων, όσο και για την πραγματοποίηση των συναλλαγών τους, χρειάζονται υψηλή υπολογιστική ισχύ. Επομένως οι ενεργειακές ανάγκες μεγαλώνουν καθώς αυξάνεται ο αριθμός χρηστών που συμμετέχουν στο mining. Χρησιμοποιούν όλους τους ενεργειακούς πόρους για να παράγουν κρυπτονομισμάτα, των οποίων τα οφέλη απολαμβάνονται από τους ίδιους, ενώ παράλληλα η ατμοσφαιρική ρύπανση που δημιουργείται από τις εκπομπές διοξειδίου του άνθρακα προκαλεί ζημία για όλους τους κατοίκους.



Εικόνα 4 3-month and 6-month Bitcoin volatilities

## 1.6. Ανταλλακτήρια Κρυπτονομισμάτων

Οι χρήστες μπορούν να αποκτήσουν τα κρυπτονομισμάτα με διάφορους τρόπους όπως με ανταλλαγή αγαθών και υπηρεσιών, ανταλλαγή φυσικού χρήματος ή ακόμα και με ανταλλαγή άλλων κρυπτονομισμάτων. Οι συναλλαγές αυτές πραγματοποιούνται μέσω μεσιτών και χρηματιστηρίων. Οι μεσίτες είναι τρίτα πρόσωπα που πωλούν και αγοράζουν κρυπτονομισμάτα και τα ανταλλακτήρια είναι σαν τα online χρηματιστήρια για κρυπτονομισμάτα.

Τα ανταλλακτήρια κρυπτονομισμάτων, ή αλλιώς ανταλλακτήρια ηλεκτρονικών νομισμάτων, είναι επιχειρήσεις που επιτρέπουν στους πελάτες τους τις ανταλλαγές κρυπτονομισμάτων για άλλα ηλεκτρονικά νομίσματα.

Οι επιχειρήσεις αυτές διακρίνονται σε δύο κατηγορίες:

- επιχειρήσεις με φυσική παρουσία
- επιχειρήσεις που λειτουργούν με ηλεκτρονική μορφή

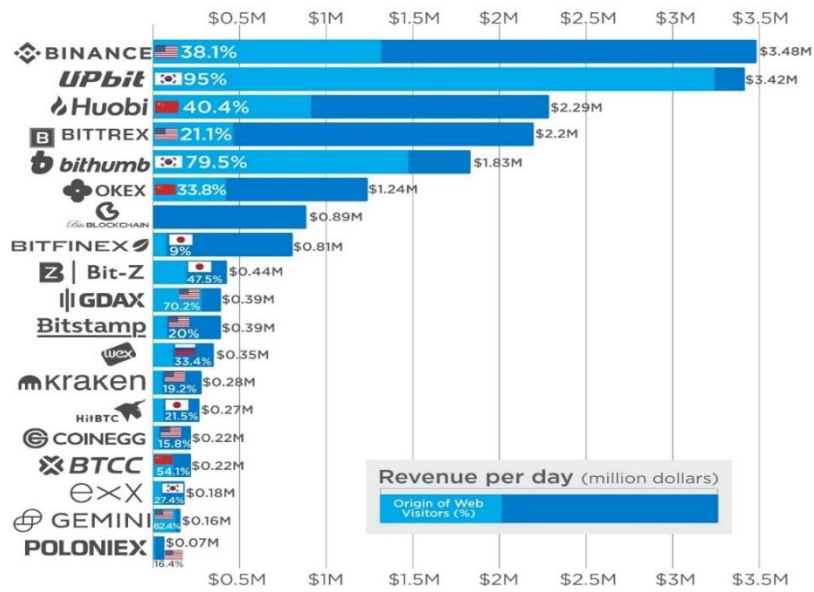
Στις επιχειρήσεις με φυσική παρουσία, οι ανταλλαγές κρυπτονομισμάτων γίνονται με παραδοσιακούς τρόπους πληρωμής, ενώ οι ηλεκτρονικές επιχειρήσεις ανταλλάσσουν χρήματα με ηλεκτρονικές πληρωμές και ηλεκτρονικά νομίσματα.

Τα ανταλλακτήρια δέχονται πληρωμές με πιστωτική κάρτα, μεταφορά χρημάτων και άλλους τρόπους πληρωμής με αντάλλαγμα ηλεκτρονικά νομίσματα ή κρυπτονομισμάτα.

Η λειτουργία των συγκεκριμένων επιχειρήσεων απαιτεί την εφαρμογή θεσμικών κανόνων. Το 2016 οι επιχειρήσεις που λειτούργησαν στον ευρωπαϊκό χώρο απέκτησαν άδειες λειτουργίας από τις αρχές της Ευρωπαϊκής Ένωσης. Το 2018 η επιτροπή της κεφαλαιαγοράς των Η.Π.Α. ανακοίνωσε πως αν μια τέτοιá επιχείρηση προσφέρει ανταλλαγή ηλεκτρονικών περιουσιακών στοιχείων θα πρέπει να εγγραφεί στην Επιτροπή Κεφαλαιαγοράς. Στις ανατολικές χώρες οι κανονισμοί προβλέπουν τη χορήγηση αδειών λειτουργίας ανταλλακτηρίων κρυπτονομισμάτων εκτός από την Κίνα και την Κορέα που έχουν απαγορεύσει την εξόρυξη κρυπτονομισμάτων.

Σύμφωνα με το Bloomberg, τα μεγαλύτερα και καλύτερα ανταλλακτήρια με βάση τον όγκο των συναλλαγών το 2021 είναι τα εξής :

- Binance
- Coinbase
- Crypto.com
- Huobi
- Kraken
- Ftx



Εικόνα 5 Ποσοστά συναλλαγών ανά ανταλλακτήριο

## Κεφάλαιο 2<sup>ο</sup>: Μηχανική Μάθηση

Ο άνθρωπος, από την αρχαιότητα, προσπαθούσε να προβλέψει ή να εκτιμήσει διάφορες καταστάσεις για το μέλλον χρησιμοποιώντας ως γνώμονα παρατηρήσεις είτε σε παρελθοντικά γεγονότα είτε σε επαναλαμβανόμενα φαινόμενα του περιβάλλοντός του.

Η ανάγκη αυτή του ανθρώπου να προβλέψει μελλοντικές καταστάσεις είναι ακόμα μεγαλύτερη στις μέρες μας. Η σωστή πρόβλεψη είναι καθοριστική για τη λήψη σημαντικών αποφάσεων, όπως πχ. οι προβλέψεις που γίνονται για τους χρηματιστηριακούς δείκτες που ορίζουν τις αγορές, οι προβλέψεις για τα καιρικά φαινόμενα, για ασθένειες κ.α.

Η διαδικασία της πρόβλεψης σε πραγματικά φαινόμενα, βασίζεται κατά κύριο λόγο στα δεδομένα που έχουμε στην κατοχή μας τη συγκεκριμένη χρονική στιγμή. Δηλαδή, στα δεδομένα του παρόντος και τα παρελθοντικά δεδομένα που υπάρχουν για το εκάστοτε φαινόμενο, όπως πχ. ο καιρός. Η ανάλυση και η επεξεργασία των δεδομένων αυτών με απώτερο σκοπό την επιτυχία μιας πρόβλεψης με όσο το δυνατόν μεγαλύτερα ποσοστά επιτυχίας, οδήγησε στην ανάπτυξη της Μηχανικής Μάθησης (Machine Learning). Η μηχανική μάθηση αποτελεί έναν από τους πιο σημαντικούς πυλώνες της Τεχνητής Νοημοσύνης και η λειτουργία της στοχεύει στη δημιουργία ενός μοντέλου του φαινομένου μέσω του συνόλου δεδομένων εκπαίδευσης που παρατηρήθηκαν σε παρελθοντικό χρόνο, ώστε να καταφέρει επιτυχώς να εκτιμήσει την πορεία του φαινομένου σε μελλοντικό χρόνο. (I.Βλαχάβας, 2006).

Με άλλα λόγια η Μηχανική Μάθηση έχει σκοπό τη δημιουργία μηχανών ικανών να μαθαίνουν, να βελτιώνουν την απόδοσή τους σε πολλούς τομείς με χρήση και αξιοποίηση προηγούμενης γνώσης και εμπειρίας.

Παρόλο που είναι πολύ δύσκολο έως και ακατόρθωτο να δημιουργηθούν μηχανές ικανές να μαθαίνουν εύκολα και γρήγορα μεγάλη ποικιλία πραγμάτων όπως ο άνθρωπος, έχουν αναπτυχθεί σημαντικοί αλγόριθμοι μηχανικής μάθησης οι οποίοι σε πολλούς εμπορικούς τομείς έχουν μεγάλα ποσοστά επιτυχίας. Η Μηχανική Μάθηση συνδέεται και με άλλους επιστημονικούς κλάδους όπως για παράδειγμα η Εξόρυξη Δεδομένων, Ιατρική, Στατιστική, Πιθανότητες, Νευροβιολογία κ.α.

Στηρίχθηκε στο θεώρημα πιθανοτήτων του Bayes (A History of Bayes' Theorem) και χρονολογείται περίπου από το 1700. Τα τελευταία χρόνια, διανύοντας την εποχή της έκρηξης των Μεγάλων Δεδομένων, η επιστημονική κοινότητα με τη χρήση της ολοένα και αυξανόμενης υπολογιστικής ισχύς που μας προσφέρει ο τεχνολογικός κλάδος της νεφοϋπολογιστικής (cloud computing), η οποία αναπτύσσεται ραγδαία, και του διαδικτύου,

έχει προχωρήσει σημαντικά στην εξέλιξή της, που πλέον θεωρείται σαν ένα ξεχωριστό επιστημονικό πεδίο.

Χρησιμοποιείται σε πολλές σύγχρονες δραστηριότητες όπως στις μηχανές αναζήτησης, στον τομέα υγείας, στο ηλεκτρονικό εμπόριο, στην εξερεύνηση του διαστήματος, στη ρομποτική, στη διαχείριση πληροφοριών, στον τομέα της ναυτιλίας, σε οικονομικά μεγέθη κ.α. με στόχο την επεξεργασία δεδομένων, τις προβλέψεις σε πραγματικό χρόνο, την πολυδιάστατη οπτικοποίηση των δεδομένων και τέλος τον καθοριστικό ρόλο στη λήψη σημαντικών αποφάσεων. (Ian Goodfellow, 2016)

## 2.1. Ορισμός Μηχανικής Μάθησης

Η Μηχανική Μάθηση είναι κλάδος της Τεχνητής Νοημοσύνης που ασχολείται με τη μελέτη μαθηματικών αλγορίθμων που χρησιμοποιούνται για την εκπαίδευση υπολογιστικών συστημάτων, τα οποία επιφορτίζονται με τις καθημερινές εργασίες που μέχρι πρότινος εκτελούνταν συνήθως από ανθρώπινο δυναμικό. Το επίκεντρο της Μηχανικής Μάθησης είναι ο αλγόριθμος εκπαίδευσης που έχει την ικανότητα να επεξεργάζεται διάφορες πηγές δεδομένων και να ‘μαθαίνει’ από αυτά. Η ικανότητά του αυτή αυξάνεται με την πάροδο του χρόνου μέσω επαναλαμβανόμενων διεργασιών. (B. Βερούκιος, 2015)

Έχουν δοθεί διάφοροι ορισμοί από την επιστημονική κοινότητα για την εννοιολογική σημασία της Μηχανικής Μάθησης (Machine Learning). Πιο συγκεκριμένα:

- «Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί.» *Arthur Samuel (IBM, 1959)*
- «Η μελέτη υπολογιστικών μεθόδων για την απόκτηση νέας γνώσης, νέων δεξιοτήτων και νέων τρόπων οργάνωσης της υπάρχουσας γνώσης.» *Carbonell (1987)*
- «Ένα πρόγραμμα υπολογιστή θεωρείται ότι μαθαίνει από την εμπειρία E ως προς μια κλάση εργασιών T και μέτρο απόδοσης P, αν η απόδοσή του σε εργασίες από το T, όπως μετριέται το P, βελτιώνεται μέσω εμπειρίας E.» *Mitchell (1997)*
- «Κάτι μαθαίνει όταν αλλάζει η συμπεριφορά του κατά τέτοιο τρόπο ώστε να αποδίδει καλύτερα στο μέλλον.» *Witten & Frank (2000)*

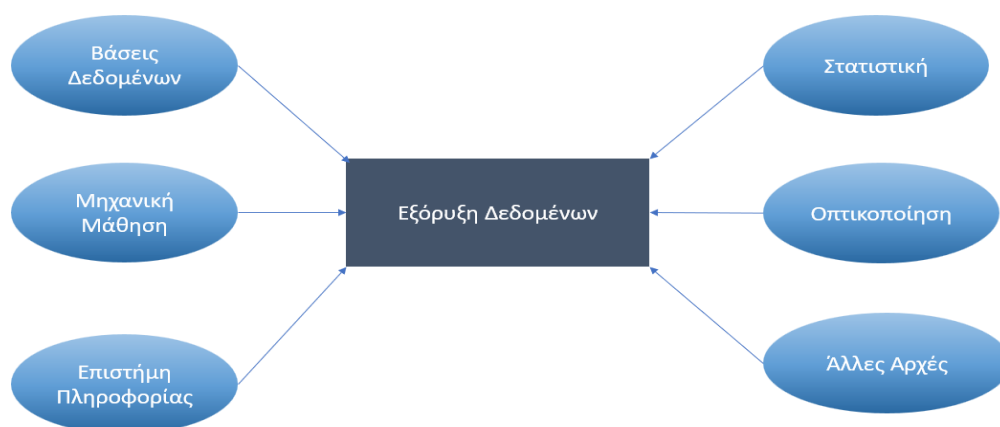
Η Μηχανική Μάθηση τέθηκε σε εφαρμογή για πρώτη φορά από τον Arthur Samuel στην IBM το 1952, ο οποίος ανέπτυξε ένα υπολογιστικό πρόγραμμα το οποίο παρατηρούσε και πρόβλεπε τις βέλτιστες κινήσεις για το παιχνίδι της ντάμας. Το συγκεκριμένο πρόγραμμα, με την πάροδο του χρόνου, γινόταν καλύτερο στις προβλέψεις με μεγάλα ποσοστά επιτυχίας.



Τις επόμενες δεκαετίες δημιουργήθηκαν νέα μοντέλα Μηχανικής Μάθησης, όπως ο αλγόριθμος του Rosenblatt Perceptron το 1957, ο αλγόριθμος Backpropagation το 1981, τα δέντρα απόφασης και ο αλγόριθμος ID3 το 1986, ο αλγόριθμος Support Vector Machines το 1995, καθώς και πολλά άλλα μοντέλα, τα οποία διαδραματίζουν καθοριστικό ρόλο στην εξέλιξη της τεχνολογίας.

Η Μηχανική Μάθηση βασίζεται κατά κύριο λόγο στη στατιστική, εφόσον χρησιμοποιεί ανάλυση δεδομένων, καθώς επίσης και στην εξόρυξη δεδομένων (data mining) που εστιάζει περισσότερο στην διερευνητική ανάλυση δεδομένων.

Η εξόρυξη δεδομένων (data mining) είναι η διαδικασία που περιλαμβάνει τον εντοπισμό καίριων και καινοτόμων μοτίβων ή προτύπων τα οποία παρουσιάζουν μεγάλο ενδιαφέρον, καθώς και τη δημιουργία περιγραφικών, κατανοητών και προβλεπτικών μοντέλων από δεδομένα μεγάλης κλίμακας. Συνδυάζει τις παραδοσιακές μεθόδους ανάλυσης δεδομένων με περίπλοκους αλγόριθμους για την επεξεργασία των δεδομένων αυτών. Τέτοιες τεχνικές (data mining) αναπτύσσονται για να διατρέξουν μεγάλες βάσεις δεδομένων με σκοπό την ανακάλυψη χρήσιμης πληροφορίας που σε άλλες περιπτώσεις θα ήταν αδύνατη. Ένα άλλο πλεονέκτημά τους είναι η δυνατότητα πρόβλεψης των αποτελεσμάτων μελλοντικών παρατηρήσεων. Η εξόρυξη δεδομένων (data mining) αποτελεί ένα από τα σημαντικότερα κομμάτια της επιστήμης των υπολογιστών με αντικείμενο την ανακάλυψη λειτουργικής γνώσης, μέσω της ανάλυσης δεδομένων από μεγάλες αποθήκες δεδομένων (data warehouses) καθώς επίσης και την εύρεση των μοντέλων που αναδεικνύουν γνώση κρυμμένη μέσα στα δεδομένα και δεν μπορεί να εξαχθεί με ευκολία από τον άνθρωπο (Mohammed, J. Z. & Wagner, 2017). Είναι ο κυρίαρχος συνδετικός κρίκος πολλών επιστημονικών αρχών.



Εικόνα 6 Εξόρυξη δεδομένων

## 2.2. Σκοπός της Μηχανικής Μάθησης

Όπως αναφέραμε πιο πάνω ως Μηχανική Μάθηση ορίζουμε τη διαδικασία κατά την οποία βελτιώνεται η επίδοση ενός συστήματος με τη χρήση αλγορίθμων σε μία συγκεκριμένη εργασία μετά από επαναλαμβανόμενες φορές εκτέλεσης της εργασίας. Η διαδικασία αυτή απαιτεί:

- περιβάλλον που προσφέρει δεδομένα υπό μορφή παραδειγμάτων στο σύστημα
- ένα κριτήριο αξιολόγησης με βάση το οποίο θα αξιολογείται η επίδοση του συστήματος
- μια εργασία την οποία το σύστημα θα εκτελέσει πολλές φορές.

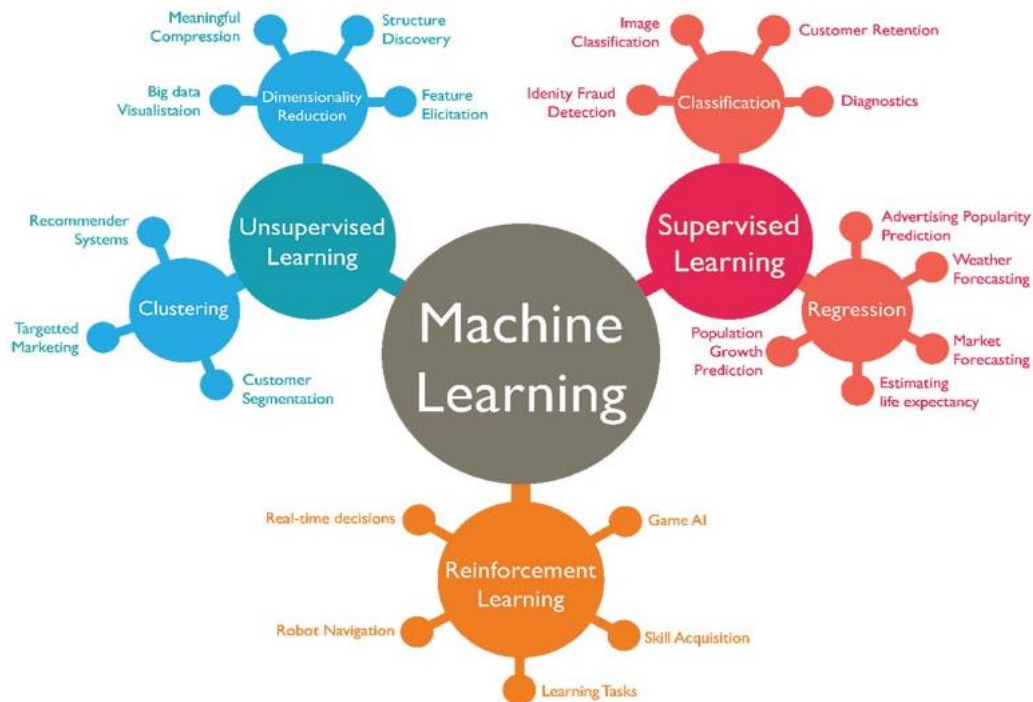
Υπάρχουν πάρα πολλά παραδείγματα τέτοιων εργασιών που βελτιώνονται μέσα από τη Μηχανική Μάθηση όπως:

- **Η αναγνώριση** διάφορων αντικειμένων, π.χ. ρούχα, παπούτσια, ζώνες κτλ. Όλα τα αντικείμενα χαρακτηρίζονται από πληθώρα χρωμάτων, σχημάτων και μεγεθών, τα οποία το σύστημα μπορεί να αναγνωρίσει μετά από πολλές παρατηρήσεις και αναλύσεις. Το σύστημα μαθαίνει να αναγνωρίζει με μεγαλύτερη ακρίβεια, όταν θα μπορεί να αντιστοιχίζει τα κατάλληλα γνωρίσματα στο σωστό αντικείμενο.
- **Η πρόβλεψη** της τιμής μετρήσιμων ποσοτήτων, όπως π.χ. η θερμοκρασία σε κάποιο σημείο στον πλανήτη, όταν γνωρίζουμε άλλες σχετικές με τη θερμοκρασία τιμές (πίεση, υγρασία κτλ.) στο ίδιο σημείο. Παρατηρώντας και αναλύοντας μεγάλο δείγμα των μετρήσεων αυτών των ποσοτήτων σε διάφορα μέρη του πλανήτη και βρίσκοντας τη σχέση μεταξύ τους, τότε λέμε ότι το σύστημα επιτυγχάνει με μεγάλη ακρίβεια την πρόβλεψή του.
- **Η συσταδοποίηση** συγγενών αντικειμένων, όπως π.χ. η ομαδοποίηση γραπτών κειμένων τα οποία έχουν το ίδιο θέμα. Το σύστημα μπορεί να μάθει να ομαδοποιεί τα δεδομένα με βάση κάποιο κριτήριο ομοιότητας.
- **Η ανάλυση** δεδομένων, όπως π.χ. η εύρεση παραγόντων που επηρεάζουν τις τιμές των μετοχών στο χρηματιστήριο ή των κρυπτονομισμάτων. Το σύστημα σε αυτή την περίπτωση μαθαίνει αυτούς τους παράγοντες παρατηρώντας και αναλύοντας μεγάλο πλήθος διακυμάνσεων των τιμών σε ένα μεγάλο χρονικό διάστημα.
- **Η ανάπτυξη** συγκεκριμένων στρατηγικών σε παιχνίδια, όπως π.χ. η εύρεση των βέλτιστων κινήσεων σε μία παρτίδα σκάκι. Το σύστημα το πετυχαίνει αυτό, παρατηρώντας και αναλύοντας πολλές παρτίδες σκάκι.

Από τα παραπάνω παραδείγματα, διαπιστώνουμε ότι στόχος της Μηχανικής Μάθησης είναι η δυνατότητα παραγωγής σωστών εκτιμήσεων και προβλέψεων σχετικά με δεδομένα τα οποία αντιμετωπίζονται για πρώτη φορά από το σύστημα (Κ. Διαμαντάρας- Δ. Μπότσης, 2019).

### 2.3. Είδη Μηχανικής Μάθησης

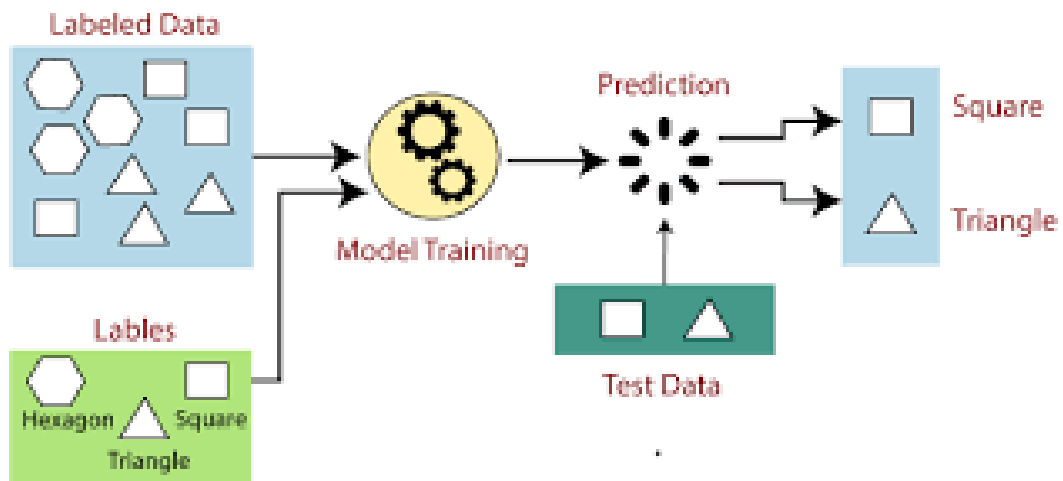
Η Μηχανική Μάθηση από δεδομένα, διακρίνεται σε κατηγορίες, ώστε ένα υπολογιστικό σύστημα να μάθει να υλοποιεί επιτυχώς μια συγκεκριμένη λειτουργία. Στα επόμενα υποκεφάλαια αναλύονται τα είδη μηχανικής μάθησης



Εικόνα 7 Είδη Μηχανικής Μάθησης

### 2.3.1. Μάθηση με επίβλεψη (supervised learning)

Πρόκειται για μια διαδικασία μηχανικής μάθησης κατά την οποία το μοντέλο δέχεται μία σειρά από πρότυπα εισόδου, γνωστή και ως σύνολο εκπαίδευσης (training data) από το χρήστη, με σκοπό να εξάγει ένα συμπέρασμα-αποτέλεσμα. Η είσοδος αυτή έχει τη μορφή ενός πίνακα δύο διαστάσεων, όπου η μία μεταβλητή είναι το δεδομένο ( $X_1, X_2, X_3, \dots$ ) και η άλλη μεταβλητή χαρακτηρίζει τον επιθυμητό στόχο ( $Y_1, Y_2, Y_3, \dots$ ). Έτσι, το σύστημα μαθαίνει τη γνώση που δέχεται, εκπαιδεύεται, για να μπορεί στο μέλλον να προβλέψει μόνο του το αποτέλεσμα από αγνώστα δεδομένα.



Εικόνα 8 Μάθηση με επίβλεψη (supervised learning)

Για τη λύση προβλήματος μάθησης με επίβλεψη απαιτούνται τα παρακάτω βήματα:

- ✓ Ορισμός τύπου δεδομένων.
- ✓ Συλλογή δεδομένων (dataset).
- ✓ Δημιουργία συνόλου εκπαίδευσης (training data).
- ✓ Ορισμός του τρόπου αναπαράστασης των δεδομένων που εισάγονται στο σύστημα (π.χ. η διαστάσεων).
- ✓ Επιλογή του κατάλληλου αλγορίθμου μηχανικής μάθησης.
- ✓ Συνεχόμενη εκμάθηση του συστήματος μέσω του συνόλου εκπαίδευσης (training data).
- ✓ Σύγκριση και αξιολόγηση των αποτελεσμάτων και της ακρίβειας του συγκεκριμένου αλγορίθμου με χρήση αγνώστων δεδομένων στο σύστημα.

Τα προβλήματα μάθησης με επίβλεψη είναι τα συνηθέστερα προβλήματα μηχανικής μάθησης. Χωρίζονται σε δύο υποκατηγορίες:

- **Προβλήματα ταξινόμησης (classification problems)**, όπου σε αυτά τα προβλήματα οι στόχοι είναι διακριτές τιμές και αντιστοιχούν σε κλάσεις αντικειμένων. Η βασική ιδέα είναι ότι έχοντας ένα σύνολο κατηγοριών (class) και ένα σύνολο δεδομένων, τα οποία είναι γνωστό σε ποια κατηγορία ανήκουν (training data), ο στόχος της κατηγοριοποίησης είναι η δημιουργία ενός μοντέλου που να μπορεί να αντιστοιχίσει μόνο του σε αυτές τις κατηγορίες νέα δεδομένα. Οι τύποι αλγορίθμων που χρησιμοποιούνται για την επίλυση αυτών των προβλημάτων περιλαμβάνουν δέντρα απόφασης, μηχανές διανύσματος υποστήριξης και πολλοί άλλοι.
- **Προβλήματα παλινδρόμησης (regression problems)**, όπου σε αυτά τα προβλήματα οι στόχοι είναι συνεχείς τιμές ή απεριόριστο πλήθος διακριτών τιμών και αντιστοιχούν σε τιμές κάποιων ποσοτήτων. Η βασική ιδέα εδώ είναι η δημιουργία ενός μοντέλου, μέσω της εκπαίδευσης της συνάρτησης των δεδομένων (training data), όπου θα εμφανίζεται η είσοδος του συστήματος σε μια πραγματική μεταβλητή. Είναι μια προγνωστική μέθοδος, σύμφωνα με την οποία προβλέπονται οι τιμές μια εξαρτημένης μεταβλητής από κάποιες ανεξάρτητες. Η παλινδρόμηση χρησιμοποιείται σε προβλήματα πρόβλεψης τιμών, όπως θερμοκρασία, χρηματιστηριακοί δείκτες κτλ. Οι τύποι αλγορίθμων που χρησιμοποιούνται σε αυτή την κατηγορία προβλημάτων περιλαμβάνουν γραμμική παλινδρόμηση, πολυδιαφασική παλινδρόμηση κτλ. (Κ. Διαμαντάρας- Δ. Μπότσης, 2019).

Υπάρχουν πολλοί αλγόριθμοι που υποστηρίζουν τη μάθηση με επίβλεψη και οι πιο γνωστοί είναι οι παρακάτω:

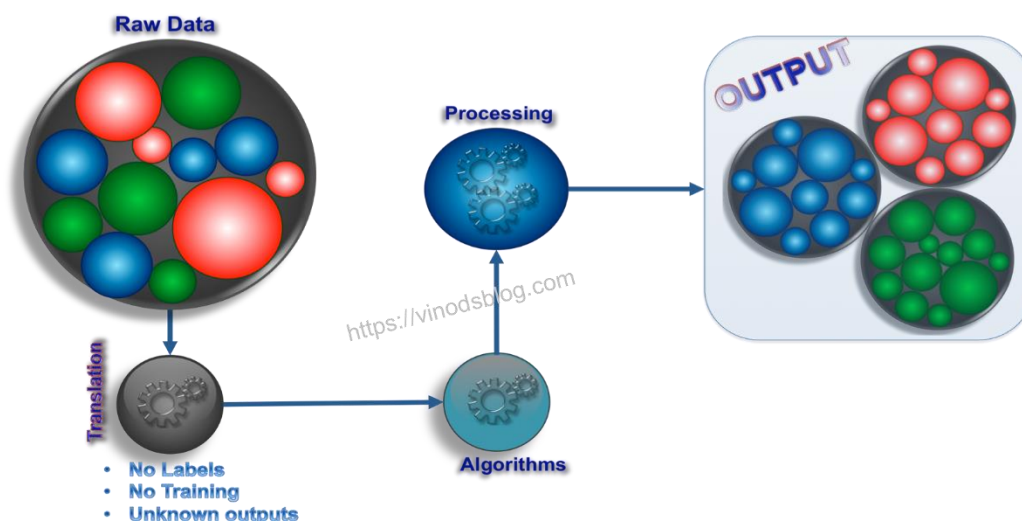
- Naives Bayes
- Decision Trees
- Linear regression
- Logistic regression
- K-nearest neighbours
- SVM (Support vector machines)
- Neural Networks

Δεν υπάρχει ιδιαίτερη διάκριση μεταξύ των αλγορίθμων που χρησιμοποιούμε για τη λύση των προβλημάτων. Το μόνο που μπορούμε να πούμε είναι ότι υπάρχουν αλγόριθμοι που ταιριάζουν καλύτερα ανάλογα με τη φύση του προβλήματος που καλούνται να επιλύσουν. Οι

αλγόριθμοι που θα χρησιμοποιήσουμε για το σκοπό της συγκεκριμένης διπλωματικής εργασίας θα αναλυθούν περισσότερο στα επόμενα κεφάλαια.

### 2.3.2. Μάθηση χωρίς επίβλεψη (unsupervised learning)

Πρόκειται για μία διαδικασία μηχανικής μάθησης κατά την οποία το μοντέλο χρησιμοποιεί τα πρότυπα εισόδου  $X_1, X_2, X_3, \dots$ , χωρίς όμως να διαθέτει συγκεκριμένες πληροφορίες σχετικά με τους στόχους. Σκοπός του συστήματος είναι η δημιουργία συνόλου κατηγοριών (συστάδων), οι οποίες θα περιέχουν δείγματα δεδομένων με τα ίδια ή παρόμοια χαρακτηριστικά.



Εικόνα 9 Μάθηση χωρίς επίβλεψη (unsupervised learning)

Τα προβλήματα μάθησης χωρίς επίβλεψη χωρίζονται σε δύο υποκατηγορίες:

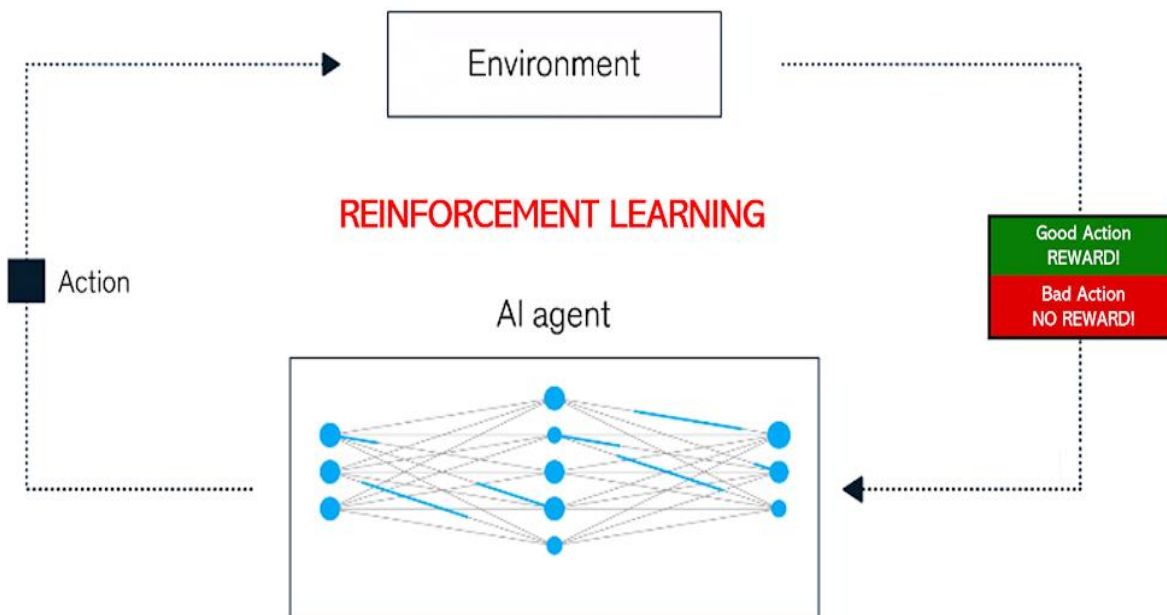
**Προβλήματα ομαδοποίησης (clustering).** Ο κύριος στόχος των αλγόριθμων στα προβλήματα αυτά είναι η ομαδοποίηση των δεδομένων εισόδου σε διαφορετικές κατηγορίες χρησιμοποιώντας μόνο τα χαρακτηριστικά που προέρχονται από τα δεδομένα εισόδου και καμία άλλη εξωτερική πληροφορία. Η ομαδοποίηση χρησιμοποιείται στην στατιστική ανάλυση δεδομένων, στην εξόρυξη δεδομένων κτλ. Μερικοί από τους αλγόριθμους που χρησιμοποιούνται για την επίλυση προβλημάτων ομαδοποίησης είναι ο k-means, k-medoids και ιεραρχικοί αλγόριθμοι.

**Προβλήματα ανακάλυψης κανόνων συσχέτισης (association rules),** όπου οι αλγόριθμοι που χρησιμοποιούνται σε αυτά τα προβλήματα εξάγουν κανόνες και μοτίβα από σύνολα

δεδομένων. Οι συγκεκριμένοι κανόνες εξηγούν τις σχέσεις μεταξύ των μεταβλητών και απεικονίζουν σύνολα στοιχείων που εμφανίζονται στα δεδομένα. Επίσης βοηθούν στην ανακάλυψη νέων χρήσιμων πληροφοριών για επιχειρήσεις και οργανισμούς λόγω του μεγάλου όγκου δεδομένων τους. Αυτοί οι αλγόριθμοι χρησιμοποιούνται στον τομέα των πωλήσεων και σε άλλες δραστηριότητες και οι πιο γνωστοί είναι ο Apriori και ο FP Growth.

### 2.3.3. Μάθηση με ενίσχυση (reinforcement learning)

Στη συγκεκριμένη μάθηση το μοντέλο χρησιμοποιεί μία ακολουθία προτύπων εισόδου  $X_1, X_2, X_3, \dots$ , και οι στόχοι είναι συνήθως τιμές ανταμοιβής ή τιμωρίας. Ο αλγόριθμος που χρησιμοποιείται στην ενισχυτική μάθηση ονομάζεται πράκτορας και έχει στόχο την



Εικόνα 10 Μάθηση με ενίσχυση (reinforcement learning)

μεγιστοποίηση κάποιων μεταβλητών (ανταμοιβή) ή την ελαχιστοποίηση αυτών (τιμωρία) μέσα από την αλληλεπίδραση με το περιβάλλον σε πραγματικό χρόνο. Δεν υπάρχει κάποιος εξωτερικός παράγοντας που να ελέγχει και να καθοδηγεί το σύστημα για τις ενέργειες που θα ακολουθήσει. Πρέπει να ανακαλύψει μόνο του τις ενέργειες εκείνες που θα του αποφέρουν το βέλτιστο κέρδος.

Η μάθηση με ενίσχυση περιγράφεται στο παραπάνω σχήμα και τα βήματα που ακολουθούνται είναι τα εξής:

1. Ο αλγόριθμος (agent) και το περιβάλλον (environment) αλληλεπιδρούν συνεχώς. Ο αλγόριθμος επιλέγει μόνος του τις ενέργειες και το περιβάλλον ανταποκρίνεται σε αυτές και του παρουσιάζει νέες καταστάσεις.
2. Το περιβάλλον επιστρέφει στον αλγόριθμο κάποιος μεταβλητές, θετικές η αρνητικές.
3. Ο αλγόριθμος προσπαθεί να μεγιστοποιήσει αυτές τις μεταβλητές.

Οι κεντρικές ιδέες που χαρακτηρίζουν τη μάθηση με ενίσχυση είναι δύο. Συγκεκριμένα:

**Εξερεύνηση(exploration):** όπου πρόκειται για δοκιμή διαφορετικών επιλογών, ώστε να διερευνηθεί η αντίδραση του περιβάλλοντος και η ανταμοιβή από κάθε ενέργεια.

**Εκμετάλλευση(exploitation):** όπου πρόκειται για τη χρήση της γνώσης που αποκτούμε από την εξερεύνηση των διάφορων επιλογών ώστε να επιλέξουμε την καλύτερη ενέργεια κάθε φορά.

Τέτοιου είδους αλγόριθμοι χρησιμοποιούνται σε ρομποτικές εφαρμογές, για τη βελτιστοποίηση εργασιών σε επιχειρήσεις και για την μάθηση επιτραπέζιων παιχνιδιών με υπολογιστή (π.χ. σκάκι).

## 2.4. Αλγόριθμοι Μηχανικής Μάθησης

Υπάρχουν πολλοί αλγόριθμοι που έχουν αναπτυχθεί για την επίλυση πολλών και διαφορετικών προβλημάτων. Στην ενότητα αυτή θα περιγράψουμε κάποιους εξ αυτών για κάθε κατηγορία μηχανικής μάθησης. Συγκεκριμένα θα περιγράψουμε τον αλγόριθμο του Naives Bayes (*supervised learning*), τον αλγόριθμο k-means (*unsupervised learning*) και τον αλγόριθμο MDP μαρκοβιανής διαδικασίας απόφασης (*reinforcement learning*). Τους αλγόριθμους που θα χρησιμοποιήσουμε για την ανάλυση, θα τους περιγράψουμε αναλυτικότερα στα επόμενα κεφάλαια.

### 2.4.1. Αλγόριθμος Naive Bayes

Ο συγκεκριμένος αλγόριθμος βασίζεται στο θεώρημα πιθανοτήτων Bayes. Κύριος στόχος του είναι να προβλέψει αποτελέσματα από μη επισημασμένα δεδομένα, δηλαδή από δεδομένα που δεν έχουν επισημανθεί με ετικέτες που προσδιορίζουν χαρακτηριστικά και



ιδιότητες. Η μέθοδος ταξινόμησης στο θεώρημα Bayes, προϋποθέτει την ανεξαρτησία των χαρακτηριστικών των δεδομένων και της κλάσης κατηγοριοποίησής τους.

Τα μοντέλα Bayes διακρίνονται σε τύπους ανάλογα με τον τρόπο χειρισμού των χαρακτηριστικών. Για παράδειγμα στο μοντέλο Bernoulli οι τιμές των χαρακτηριστικών πρέπει να είναι δυαδικές (0-1, T-F, κ.α.).

Ο αλγόριθμος Naïve Bayes είναι ένας χρήσιμος αλγόριθμος ταξινόμησης και χρησιμοποιείται για να χαρακτηρίσει σύνολα δεδομένων με μεγάλο όγκο πληροφοριών. Είναι εύκολος στην εκτέλεση και εφαρμογή του και όπως αναφέρθηκε παραπάνω βασίζεται στο θεώρημα πιθανοτήτων Bayes και χαρακτηρίζεται από τον παρακάτω μαθηματικό τύπο:

$$P(A|B) = P(B|A)*P(A)/ P(B)$$

- $P(A|B)$  = η ζητούμενη πιθανότητα
- $P(B|A)$  = η δεσμευμένη πιθανότητα
- $P(A)$  = η αρχική πιθανότητα κλάσης
- $P(B)$  = η αρχική πιθανότητα ταξινομητή

$$P(A|B) = P(B1|A)* P(B2|A) *...* P(Bn|A) * P(A)$$

Η ζητούμενη πιθανότητα  $P(A|B)$  υπολογίζει την πιθανότητα του αποτελέσματος που προκύπτει από την εισαγωγή μιας νέας πληροφορίας. Επομένως το A αναφέρεται στην κλάση που ταξινομούνται τα δεδομένα και το B τον ταξινομητή. Η δεσμευμένη πιθανότητα χαρακτηρίζει την πιθανότητα να βρίσκεται ο ταξινομητής (χαρακτηριστικά δεδομένων) μέσα στην αντίστοιχη κλάση Heckerman, D.(1999).

## 2.4.2. Αλγόριθμος K-means

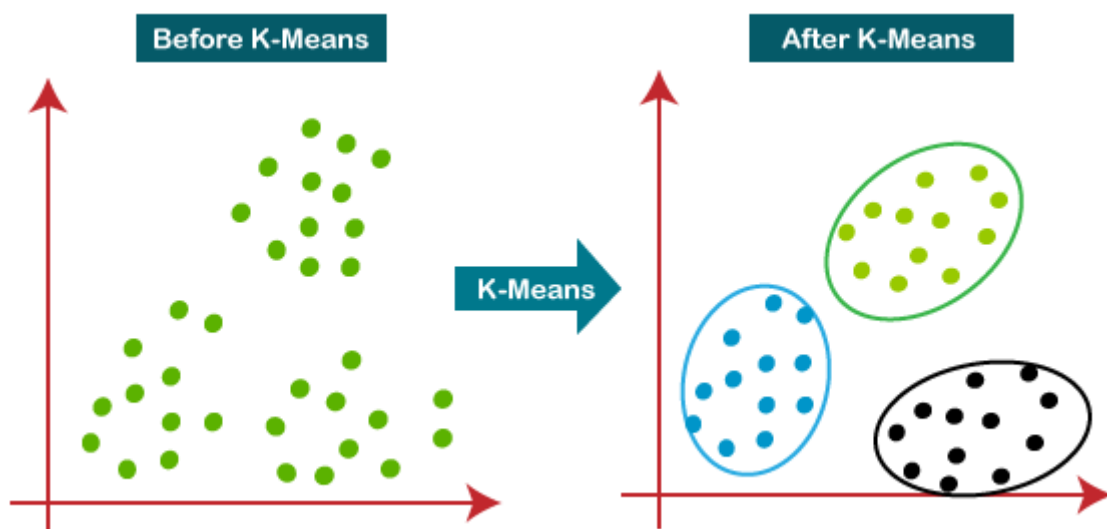
Πρόκειται για έναν επαναληπτικό αλγόριθμο που χωρίζει το σύνολο των δεδομένων σε K διακριτές υποομάδες (συστάδες), όπου κάθε σημείο δεδομένων ανήκει σε μία μόνο ομάδα, δηλαδή ξεκινά με έναν αρχικό διαχωρισμό του συνόλου δεδομένων και επανατοποθετεί τα στοιχεία σε ομάδες με βασικό κριτήριο την ομοιότητα των στοιχείων και τα κέντρα βάρους των ομάδων μέχρι να ικανοποιηθεί κάποια συνθήκη (όπως η ελαχιστοποίηση του τετραγωνικού σφάλματος που είναι και η συχνότερη). Το τετραγωνικό σφάλμα είναι η πιο συχνή συνάρτηση που χρησιμοποιείται σε αλγόριθμους διαμέρισης όπως ο k-means και αποδίδει σωστά σε συμπαγείς ομάδες δεδομένων. Η συνάρτηση του τετραγωνικού σφάλματος για μια ομαδοποίηση L ενός συνόλου δεδομένων X με K ομάδες ορίζεται ως εξής:

$$e^2(X, L) = \sum_{j=1}^k \sum_{i=1}^{n_j} \|x_i^{(j)} - c_j\|^2$$

όπου  $X$  είναι το σύνολο των δεδομένων,  $L$  η ομαδοποίηση των δεδομένων,  $K$  οι ομάδες διαχωρισμού του συνόλου των δεδομένων,  $x_i^{(j)}$  είναι το  $i$  στοιχείο που ανήκει στην  $j$  ομάδα και  $c_j$  είναι το κέντρο βάρους της  $j$  ομάδας.

Ο αλγόριθμος k-means ακολουθεί συνοπτικά τα εξής βήματα:

1. Επιλέγει  $k$  τυχαία σημεία από το αρχικό σύνολο δεδομένων, τα οποία ονομάζονται κεντροειδή της συστάδας (ομάδας) και δηλώνουν το κέντρο βάρους της συστάδας. Το  $k$  υποδηλώνει τον αριθμό των συστάδων που θέλουμε ο αλγόριθμος να δημιουργήσει.
2. Αναθέτει κάθε στοιχείο του συνόλου των δεδομένων στην κοντινότερη του συστάδα με κριτήριο την μικρότερη απόστασή του από το κέντρο βάρους της συστάδας.
3. Επανυπολογίζει και επαναπροσδιορίζει τα κέντρα βάρους των συστάδων χρησιμοποιώντας τα στοιχεία που είχαν τοποθετηθεί σε αυτές στο βήμα 2.
4. Επαναλαμβάνει τα βήματα 2 και 3 μέχρις ότου ο αλγόριθμος να τερματίσει με κάποιο κριτήριο τερματισμού, δηλαδή να μην επηρεάζονται τα κέντρα βάρους των συστάδων ή να υπάρχει σημαντική ελαχιστοποίηση του τετραγωνικού σφάλματος.
5. Υπολογίζει τα κεντροειδή για τις συστάδες λαμβάνοντας υπόψη το μέσο όρο όλων των σημείων των δεδομένων που ανήκουν σε κάθε συστάδα.



Εικόνα 11 Αλγόριθμος K-means

Στο παραπάνω διάγραμμα διακρίνουμε πως είναι τα δεδομένα πριν και μετά την εκτέλεση του αλγορίθμου k-means.

### 2.4.3. Διαδικασίες αποφάσεων Markov

Όπως αναφέρθηκε παραπάνω η μάθηση με ενίσχυση είναι ένα σημαντικό μοντέλο μάθησης και χρησιμοποιείται όταν ένα σύστημα που ονομάζεται πράκτορας (**agent**) αλληλεπιδρά με το περιβάλλον (**environment**) εκτελώντας συγκεκριμένες ενέργειες (**actions**) οι οποίες επιφέρουν την ανταμοιβή (είτε είναι θετική και ονομάζεται κέρδος, είτε είναι αρνητική και ονομάζεται ποινή).

Ένα συνηθισμένο πρόβλημα μάθησης με ενίσχυση είναι όταν το σύστημα-πράκτορας αλληλεπιδρά με ένα περιβάλλον που έχει μνήμη. Οι βασικές υποθέσεις τέτοιου είδους προβλημάτων είναι οι ακόλουθες:

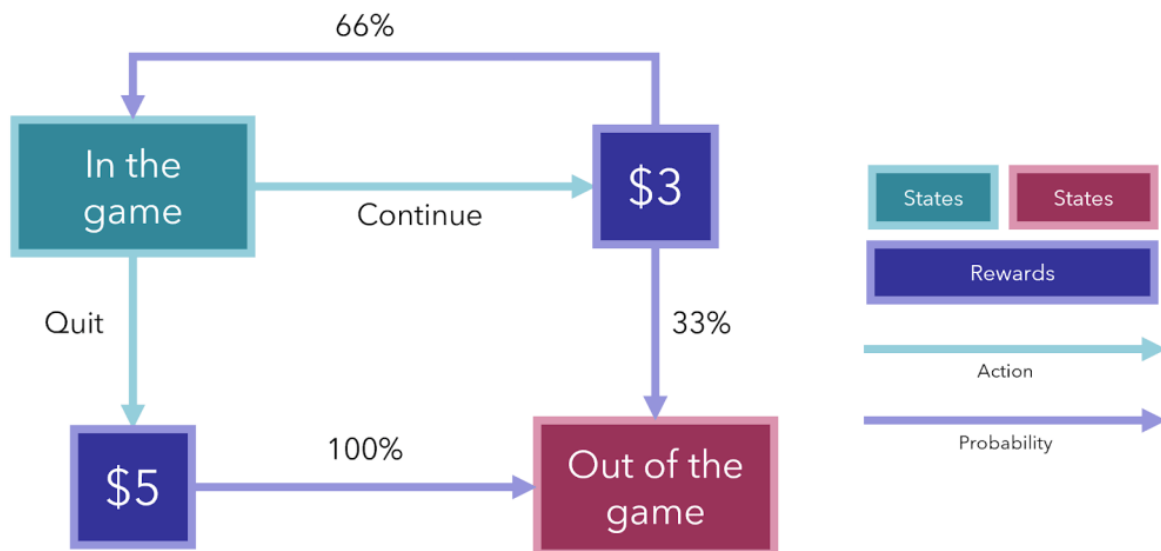
Το σύστημα-πράκτορας αλληλεπιδρά με το περιβάλλον σε διακριτές χρονικές στιγμές  $t = 1, 2, \dots$ . Για οποιαδήποτε χρονική στιγμή  $t$ :

- Το περιβάλλον βρίσκεται σε μια κατάσταση  $S(t)$  από ένα πλήθος καταστάσεων  $N$ .
- Το σύστημα-πράκτορας επιλέγει μια ενέργεια  $A(t)$  από ένα πλήθος ενεργειών  $K$ .
- Στην συνέχεια το σύστημα λαμβάνει μία ανταμοιβή  $R(t)$ , η οποία εξαρτάται και από την κατάσταση  $S(t)$  και από την ενέργεια  $A(t)$ .
- Αλλάζει η κατάσταση του περιβάλλοντος σε  $S(t+1)$ , η οποία με τη σειρά της εξαρτάται από την προηγούμενη κατάσταση  $S(t)$  καθώς επίσης και από την ενέργεια  $A(t)$ .

Τέτοια προβλήματα που διατυπώνονται με βάση τις παραπάνω υποθέσεις είναι γνωστά ως διαδικασίες αποφάσεων Markov (Markov Decisions Process – MDP). Σε αυτές τις περιπτώσεις προβλημάτων η κατάσταση του περιβάλλοντος εξαρτάται μόνο από την κατάσταση στην αμέσως προηγούμενη χρονική στιγμή. Χαρακτηρίζονται από τον παρακάτω μαθηματικό τύπο:

$$p(s', r | s, a) = \Pr(S(t+1) = s', R(t+1) = r | S(t) = s, A(t) = a)$$

όπου περιγράφεται η πιθανότητα το σύστημα να βρίσκεται στην κατάσταση  $s'$  με ανταμοιβή  $r$  στη χρονική στιγμή  $t+1$ , δοθέντος ότι στη χρονική στιγμή  $t$  το περιβάλλον βρισκόταν στην κατάσταση  $s$  και επιλέχθηκε η ενέργεια  $a$ .



Εικόνα 12 Διαδικασίες αποφάσεων Markov

### Κεφάλαιο 3<sup>ο</sup>: Χρονοσειρές

Στο κεφάλαιο αυτό θα αναφέρουμε και θα αναλύσουμε την πρόβλεψη χρονοσειρών. Θα ορίσουμε μια χρονοσειρά και θα περιγράψουμε μερικές από τις κυριότερες μεθόδους πρόβλεψης. Οι μέθοδοι πρόβλεψης χρονοσειρών χωρίζονται σε δύο κατηγορίες:

- μέθοδοι πρόβλεψης ειδικές για χρονοσειρές
- μέθοδοι πρόβλεψης χρονοσειρών με τη χρήση αλγορίθμων μηχανικής μάθησης.

Στην ενότητα αυτή θα επεκταθούμε στη δεύτερη κατηγορία και θα περιγράψουμε αναλυτικά τους συγκεκριμένους αλγόριθμους που χρησιμοποιούμε στην συγκεκριμένη κατηγορία.

Η ανάλυση χρονοσειρών αποτελεί μία από τις πιο βασικές λειτουργίες της διοίκησης πολλών επιχειρήσεων και η πρόβλεψη χρονοσειρών αποτελεί μια πολύ σημαντική πηγή πληροφόρησης, η οποία υποστηρίζει τη λήψη σημαντικών και πολύπλοκων αποφάσεων. Για κάθε επιχείρηση ή οργανισμό, η πρόβλεψη σημαντικών οικονομικών μεγεθών παίζει καθοριστικό ρόλο στις μελλοντικές εξελίξεις της. Αυτό επιτυγχάνεται μέσω της επεξεργασίας των διαθέσιμων δεδομένων με την εφαρμογή των κατάλληλων μεθόδων πρόβλεψης. Για παράδειγμα, οι επιχειρήσεις παρακολουθώντας τους δείκτες του χρηματιστηρίου, μπορούν να προβλέψουν τις μεταβαλλόμενες διακυμάνσεις της αγοράς. Επομένως με την χρήση των κατάλληλων εφαρμογών μπορούν να προβλέψουν τις τιμές των χρηματοοικονομικών δεικτών και επομένως είτε να μειώσουν το ρίσκο και τον κίνδυνο είτε να αυξήσουν το κέρδος τους.

### 3.1. Ορισμός Χρονοσειράς

Μία χρονοσειρά ορίζεται ως ένα σύνολο παρατηρήσεων  $X_t$  που συλλέγονται κατά την πάροδο του χρόνου και κάθε μία παρατήρηση καταγράφεται σε μία συγκεκριμένη χρονική στιγμή  $t$ . Μερικά σημαντικά και επίκαιρα παραδείγματα χρονοσειρών που χρήζουν συστηματικής έρευνας και μελέτης είναι η ετήσια μέση τιμή θερμοκρασίας του πλανήτη, ο ετήσιος αριθμός ηλιακών κηλίδων, το Α.Ε.Π μιας χώρας, η τιμή μιας μετοχής στο χρηματιστήριο, η τιμή ενός κρυπτονομίσματος που αποτελεί και το ερευνητικό αντικείμενο της εργασίας.

Μαθηματικά, μία χρονοσειρά ορίζεται ως ένα σύνολο από  $p$  τιμές:  $\langle t_1, x_1 \rangle, \langle t_2, x_2 \rangle, \langle t_3, x_3 \rangle, \dots, \langle t_p, x_p \rangle$ , όπου σε κάθε μία χρονική στιγμή  $p$  αντιστοιχεί μία τιμή  $x$  (θερμοκρασία, τιμή κλεισίματος μετοχής κτλ.). Οι χρονικές στιγμές είναι συγκεκριμένες και ισαπέχουν μεταξύ τους και μπορούν να χαρακτηριστούν σαν ένα διάνυσμα  $\langle x_1, x_2, x_3 \dots x_p \rangle$ . Επομένως το  $X$  είναι μία συνάρτηση του  $t$  που συμβολίζεται ως  $X=F(t)$ . Η γραφική παράσταση της συνάρτησης  $X=F(t)$  δείχνει την εξέλιξη της μεταβλητής  $X$  στο χρόνο. (C. Chatfield, 1984)

### 3.2. Κατηγοριοποίηση Χρονοσειρών

Οι χρονοσειρές ανάλογα με τα διαφορετικά γνωρίσματα που την χαρακτηρίζουν κατηγοριοποιούνται στις εξής κατηγορίες:

- **Μονομεταβλητές και πολυμεταβλητές (Univariate and Multivariate)**

Οι χρονοσειρές οι οποίες περιέχουν μία μόνο μεταβλητή κατά τη διάρκεια του χρόνου ονομάζονται μονομεταβλητές. Υπάρχουν περιπτώσεις που σε μία χρονοσειρά παρακολουθούνται παραπάνω από μία μεταβλητές. Οι χρονοσειρές αυτές ονομάζονται πολυμεταβλητές. Ένα τέτοιο παράδειγμα είναι η μελέτη της θερμοκρασίας και υγρασίας σε μία περιοχή και πως αυτές οι δύο μεταβλητές μεταβάλλονται στη διάρκεια μια συγκεκριμένης χρονικής περιόδου.

- **Στάσιμες και μη-στάσιμες (Stationary and Non-stationary)**

Οι χρονοσειρές μπορεί να είναι στάσιμες και μη-στάσιμες. Η στασιμότητα των χρονοσειρών εξαρτάται από το αν μεταβάλλεται ο μέσος όρος των τιμών της μεταβλητής, η διακύμανσή της, και η αυτοδιακύμανσή της σε διακριτές περιόδους. Οι χρονοσειρές που διατηρούν σταθερή τη μέση τιμή της μεταβλητής τους χαρακτηρίζονται ως στάσιμες, ενώ

εκείνες των οποίων η μέση τιμή της μεταβλητής τους εμφανίζει διακυμάνσεις στο χρόνο χαρακτηρίζονται ως μη-στάσιμες .(Montgomery, 2015)

- **Συνεχείς και διακριτές (Continuous and Discrete)**

Οι χρονοσειρές διακρίνονται σε συνεχείς και διακριτές. Στη συνεχή χρονοσειρά οι παρατηρήσεις συλλέγονται καθ' όλη τη διάρκεια του χρόνου ενώ στη διακριτή χρονοσειρά οι παρατηρήσεις συλλέγονται σε τακτά χρονικά διαστήματα τα οποία ισαπέχουν χρονικά μεταξύ τους όπως για παράδειγμα κάθε ώρα, κάθε μήνα, κάθε χρόνο κτλ.

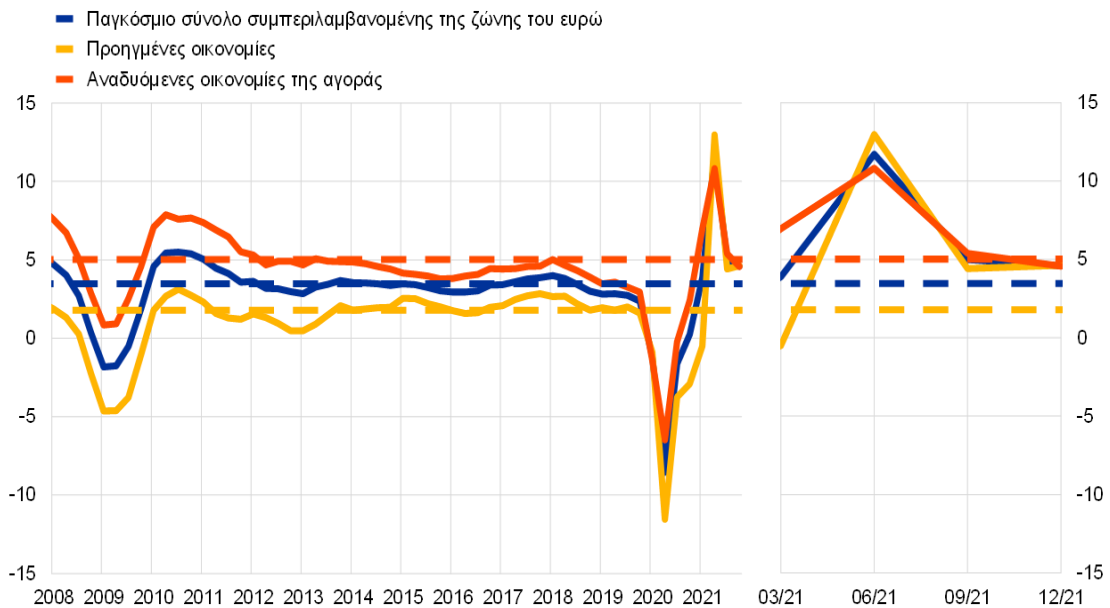
- **Γκαουσιανές και μη Γκαουσιανές (Gaussian and Non-Gaussian)**

Σε αυτή την κατηγορία μια χρονοσειρά χαρακτηρίζεται ως γκαουσιανή αν η κατανομή των τιμών της μεταβλητής της ακολουθεί την κανονική κατανομή. Αντίθετα αν δεν ακολουθείται η κανονική κατανομή η χρονοσειρά χαρακτηρίζεται ως μη γκαουσιανή. Στην δεύτερη περίπτωση υπάρχουν αρκετοί τρόποι μετασχηματισμού των δεδομένων έτσι ώστε να ακολουθήσουν την κανονική κατανομή, όπως λογαριθμικός, εκθετικός, κτλ.

### **3. Παραδείγματα χρονοσειρών**

Οι χρονοσειρές αναπαρίστανται με διαγράμματα τα οποία βοηθούν τις επιχειρήσεις να οπτικοποιήσουν τα αποτελέσματα των μετρήσεων που διαθέτουν. Μερικά παραδείγματα διαγραμμάτων είναι τα εξής:

Στο εικόνα 13 απεικονίζονται οι ετήσιες ποσοστιαίες μεταβολές του παγκόσμιου πραγματικού ΑΕΠ καθώς επίσης και τις μεταβολές του σε τριμηνιαία βάση για το έτος 2021(Haver Analytics).



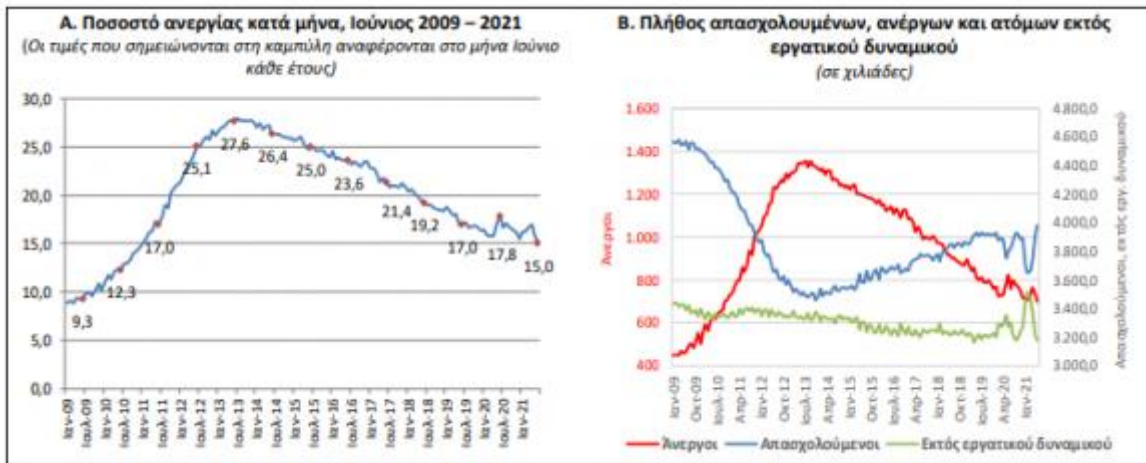
Εικόνα 13 Ετήσιες ποσοστιαίες μεταβολές του παγκόσμιου ΑΕΠ

Στο διάγραμμα της εικόνας 14 διακρίνονται όλες οι διακυμάνσεις τις τιμές του BTC από το 2018 έως το 2021.



Εικόνα 14 Μεταβολή τιμής Bitcoin χρονικής διάρκειας 2018-2021

Στο διάγραμμα της εικόνας 15 απεικονίζεται το ποσοστό ανεργίας στη χώρα μας για τον μήνα Ιούνιο για τη χρονική διάρκεια 2009-2021, σύμφωνα με τα αναθεωρημένα στοιχεία της Ελληνικής Στατιστικής Αρχής.



Εικόνα 15 Ποσοστό ανεργίας στην Ελλάδα για το μήνα Ιούνιο για χρονική διάρκεια 2009-2021

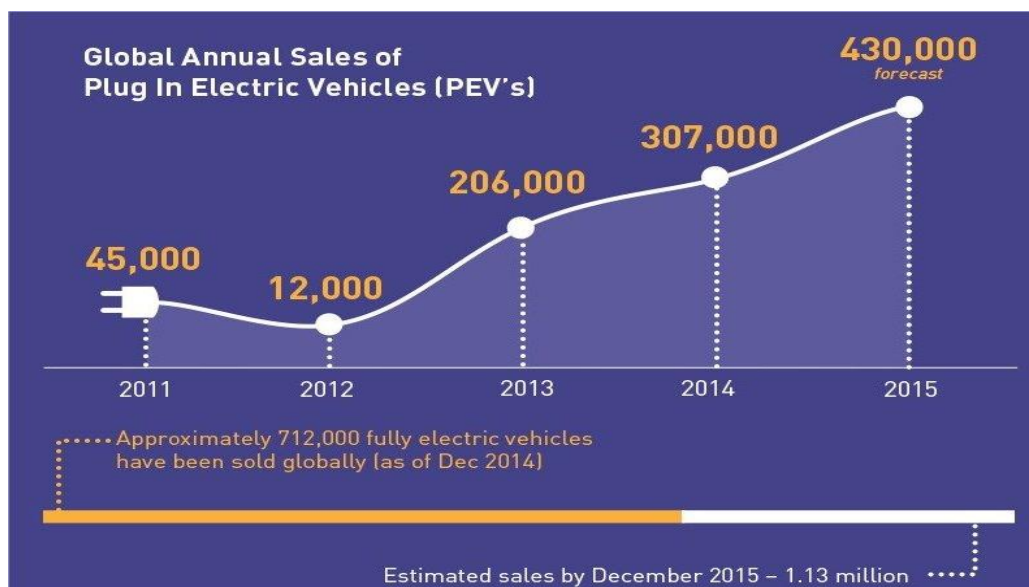
Στην εικόνα 16 διακρίνεται η εκρηκτική αύξηση της τιμής του φυσικού αερίου στην Ευρώπη για τους μήνες Απρίλιος-Σεπτέμβριος 2021.



Εικόνα 16 Αύξηση τιμής φυσικού αερίου

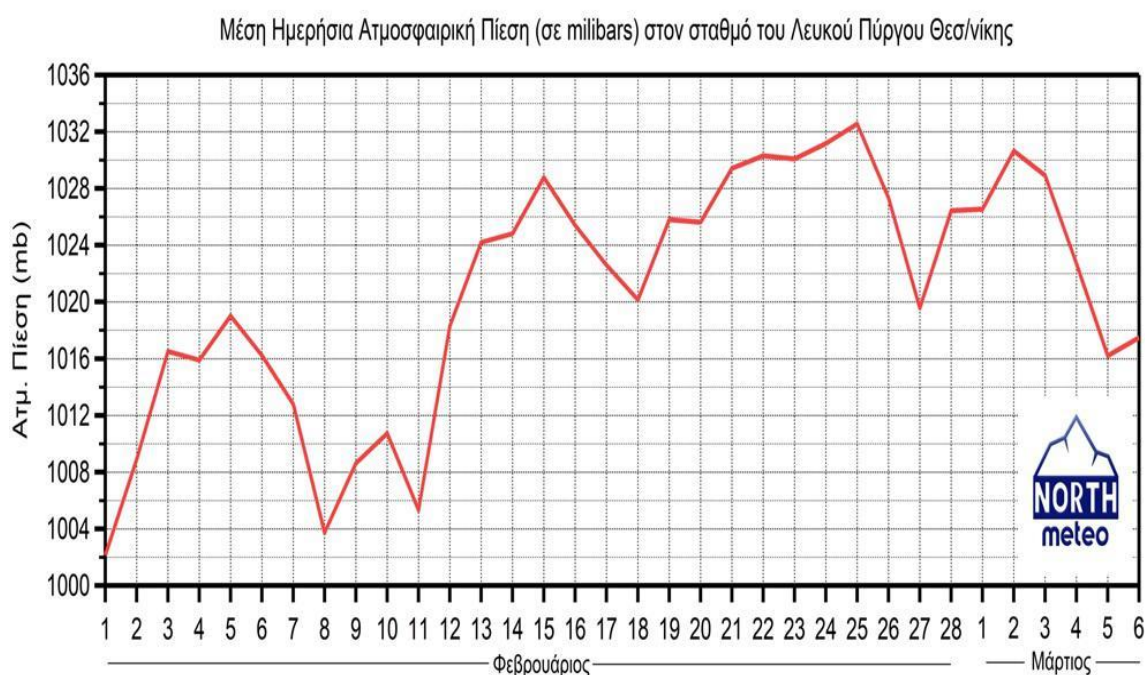


Στο διάγραμμα της εικόνας 17 απεικονίζονται οι πωλήσεις των υβριδικών-ηλεκτρικών αυτοκινήτων από το 2011 έως το 2015.



Εικόνα 17 Πωλήσεις υβριδικών-ηλεκτρικών αυτοκινήτων (2011-2015)

Στο διάγραμμα της εικόνας 18 διακρίνονται οι διακυμάνσεις της τιμής της ατμοσφαιρικής πίεσης στη Θεσσαλονίκη για τον Φεβρουάριο και τις πρώτες ημέρες του Μαρτίου 2021.

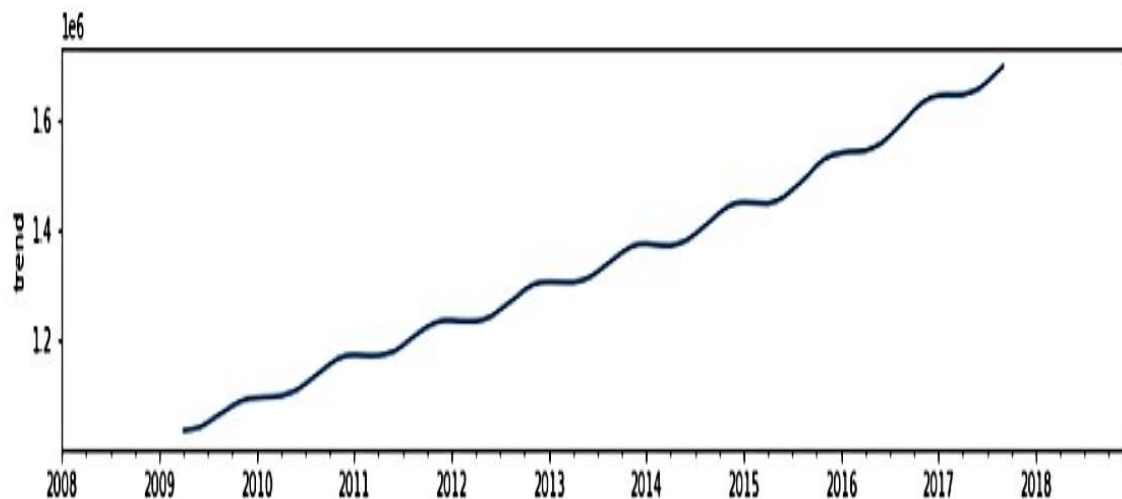


Εικόνα 18 Μεταβολή της τιμής της ατμοσφαιρικής πίεσης στη Θεσσαλονίκη

### 3.3. Ποιοτικά χαρακτηριστικά χρονοσειρών

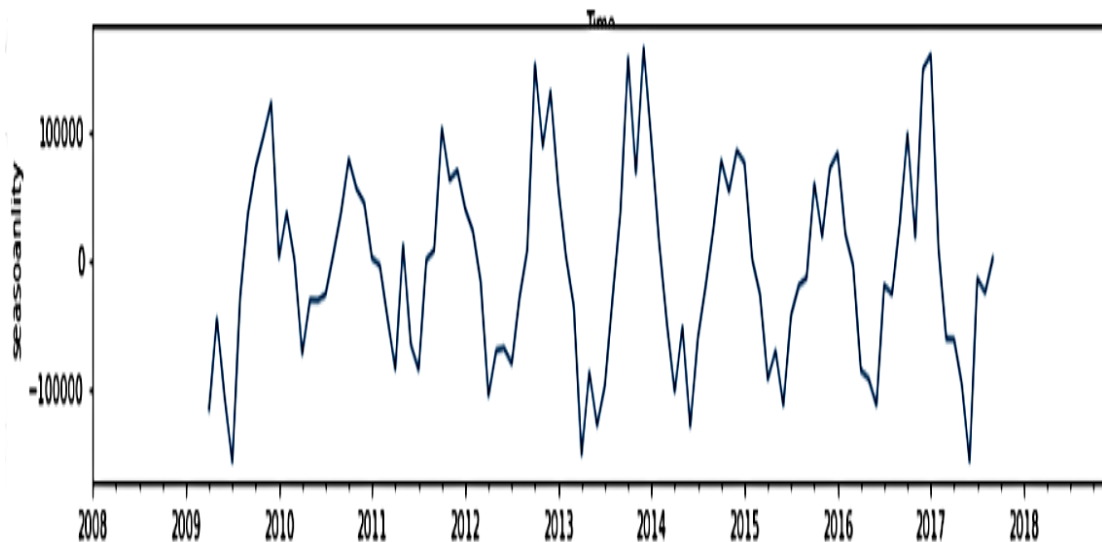
Όπως αναφέρθηκε προηγουμένως οι χρονοσειρές είναι ιστορικά δεδομένα που χαρακτηρίζονται από παρατηρήσεις μέσα σε ένα συγκεκριμένο χρονικό διάστημα. Οι παρατηρήσεις αυτές μπορεί να είναι μηνιαίες, τριμηνιαίες, ετήσιες, εβδομαδιαίες κτλ. Τα βασικά χαρακτηριστικά των χρονοσειρών είναι τα εξής:

✓ **Τάση (Trend)**, η οποία είναι μία συστηματική (γραμμικών ή μη γραμμικών στοιχείων) διαδικασία. Αλλάζει κατά την πάροδο του χρόνου και δεν επαναλαμβάνεται μέσα στα χρονικά όρια που έχουμε παρατηρήσεις. Δηλαδή, είναι η τάση που έχει μια χρονοσειρά να αυξάνεται ή να μειώνεται ή να παραμένει στάσιμη για μεγάλο χρονικό διάστημα



Εικόνα 19 Τάση

✓ **Εποχικότητα (Seasonal)**, όπου σε αυτή την περίπτωση οι τιμές της χρονοσειράς επηρεάζονται από εποχικούς παράγοντες, σε συγκεκριμένα χρονικά πλαίσια (1 έτος περίπου). Τέτοια παραδείγματα χρονοσειρών που παρουσιάζουν εποχικότητα είναι τα έσοδα από τουριστικές δραστηριότητες, αγροτικές δραστηριότητες που επηρεάζονται από την αλλαγή του κλίματος κτλ.



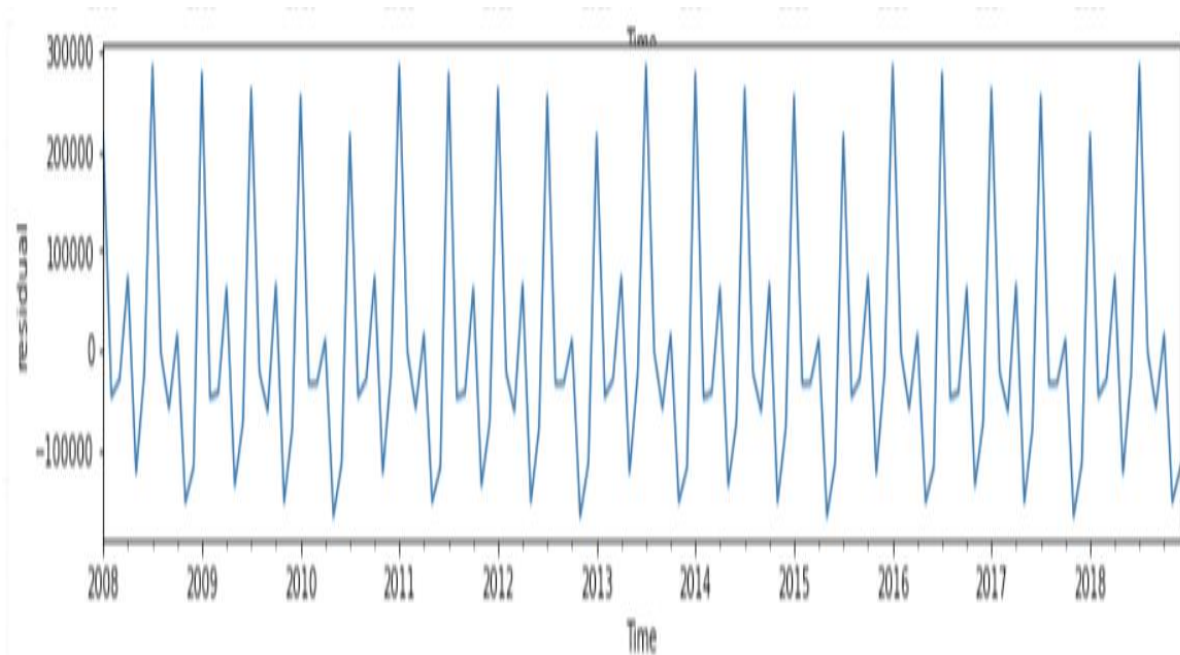
Εικόνα 20 Εποχικότητα

✓ **Κυκλικότητα (Cyclical)**, κατά την οποία οι τιμές αυξομειώνονται από περιστάσεις που προκλήθηκαν αλλά όχι σε σταθερές χρονικές περιόδους. Οι περιστάσεις αυτές επαναλαμβάνονται σε κύκλους. Η χρονική διάρκεια ενός κύκλου είναι μεγαλύτερη του ενός έτους (περίπου 10 χρόνια) και απαρτίζεται από τις διακυμάνσεις στο επίπεδο της συνολικής οικονομίας όπως πχ. περίοδοι οικονομικής ύφεσης και οικονομικής ανάπτυξης. Στις γραφικές παραστάσεις των χρονοσειρών αυτών, η κυκλικότητα παρουσιάζεται ως μία κυματοειδής γραμμή που κινείται ανάμεσα στην υψηλότερη και χαμηλότερη τιμή.



Εικόνα 21 Κυκλικότητα

✓ **Τυχαία σφάλματα (Outliers or residuals)**, όπου πρόκειται για τυχαίες τιμές που προκαλούνται από απρόβλεπτα γεγονότα τα οποία δεν μπορούν να προβλεφθούν και δεν επαναλαμβάνονται σε συγκεκριμένα χρονικά διαστήματα (πόλεμος, πανδημία, φυσικές καταστροφές κτλ.). Στην ανάλυση μιας χρονοσειράς που ανήκει στην συγκεκριμένη κατηγορία, τα τυχαία σφάλματα θεωρούνται ως θόρυβος και αφαιρούνται πριν από την ανάλυση για να την καλύτερη κατανόηση των υπόλοιπων χαρακτηριστικών της χρονοσειράς.



Εικόνα 22 Τυχαία σφάλματα

### 3.4. Αποσύνθεση Χρονοσειρών

Κάθε χρονοσειρά διασπάται στα τέσσερα προαναφερόμενα συνθετικά συστατικά της και έτσι προσδιορίζεται η επιρροή και η έκταση που έχει το καθένα από αυτά στη διαμόρφωση των τιμών της μεταβλητής. Είναι αδύνατο από τη φύση του να προβλεφθεί το μέλλον με μεγάλη ακρίβεια και επομένως πάντα θα υπάρχει ως ένα βαθμό κάποιο σφάλμα στην πρόβλεψη.

Για την διάσπαση των χρονοσειρών αποδεχόμαστε την παραδοχή ότι τα δεδομένα ακολουθούν την παρακάτω λογική:

$$\text{Δεδομένα} = \text{πρότυπο(υπόδειγμα)} + \text{σφάλμα} = f(\text{συνιστώσες}) + \text{σφάλμα}$$

Το σφάλμα που εμφανίζεται, εκφράζεται συνήθως ως η διαφορά μεταξύ των συνιστωσών της χρονοσειράς. Οι τέσσερις συνιστώσες αλληλεπιδρούν μεταξύ τους και

επηρεάζουν τόσο τις τιμές των δεδομένων όσο και την συμπεριφορά των χρονοσειρών κατά την πάροδο του χρόνου.

Με τον όρο αποσύνθεση χρονοσειρών (decomposition of timeseries) εννοούμε τον τρόπο απαλοιφής κάποιων από τα βασικά χαρακτηριστικά μια χρονοσειράς. Στηρίζεται στην υπόθεση ότι οι τιμές της χρονοσειράς δημιουργούνται από τα τέσσερα χαρακτηριστικά που τη συνθέτουν και τα οποία είναι αναφέραμε: η τάση, η εποχικότητα (περιοδικότητα), η κυκλικότητα και τα τυχαία σφάλματα. Η αποσύνθεση σε αυτές τις τέσσερις συνιστώσες γίνεται με δύο μαθηματικά μοντέλα (Yaffee,2000):

- το προσθετικό μοντέλο (addictive model)
- το πολλαπλασιαστικό μοντέλο (multiplicative model)

Έστω λοιπόν ότι:

- $Y_t$  = πραγματική τιμή της χρονοσειράς
- $T_t$  = Τάση
- $S_t$  = Εποχικότητα
- $C_t$  = Κυκλικότητα
- $E_t$  = Τυχαίο σφάλμα

όπου  $t = 1, 2, 3, \dots, n$ .

Τότε το προσθετικό μοντέλο δίνεται από την παρακάτω συνάρτηση:

$$Y_t = T_t + S_t + C_t + E_t \text{ (addictive model)}$$

Στην περίπτωση αυτή, βασικό χαρακτηριστικό του συγκεκριμένου μοντέλου είναι η ανεξαρτησία μεταξύ των συνιστωσών. Η συγκεκριμένη υπόθεση όμως είναι σχεδόν αδύνατο να πραγματοποιηθεί παρά μόνο όταν πρόκειται για μία πρόβλεψη για ένα πολύ σύντομο χρονικό διάστημα. Το μοντέλο αυτό εφαρμόζεται σε περιπτώσεις που οι εποχικές διακυμάνσεις ή η διακύμανση της τάσης είναι ανεξάρτητες από το επίπεδο της χρονοσειράς.

Το πολλαπλασιαστικό μοντέλο δίνεται από την παρακάτω συνάρτηση:

$$Y_t = T_t \cdot S_t \cdot C_t \cdot E_t \text{ (multiplicative model)}$$

Στην περίπτωση αυτή, το βασικό χαρακτηριστικό του συγκεκριμένου μοντέλου είναι η συσχέτιση μεταξύ των συνιστωσών της χρονοσειράς, δηλαδή τα  $T$ ,  $S$ ,  $C$  και  $E$  σε επίπεδο τιμών και δεικτών συσχετίζονται. Αυτό σημαίνει ότι τα  $S$ ,  $C$  και  $E$  είναι δείκτες που μπορούν να αυξομειώνουν την τάση σε οποιοδήποτε σημείο της χρονοσειράς. Αυτό το σενάριο φαντάζει πιο ρεαλιστικό με καλύτερη εφαρμογή στην πραγματικότητα. Το συγκεκριμένο μοντέλο εφαρμόζεται σε περιπτώσεις που οι εποχικές διακυμάνσεις ή η διακύμανση της τάσης εξαρτώνται από το επίπεδο της χρονοσειράς.

### 3.5. Ανάλυση Χρονοσειρών (Time Series Analysis)

Η ανάλυση των δεδομένων μιας χρονοσειράς είναι πολύ σημαντική και έχει σκοπό την ανακάλυψη των χαρακτηριστικών εκείνων που βοηθούν στην κατανόηση της ιστορικής συμπεριφοράς των δεδομένων και συμβάλουν στην πρόβλεψη των μελλοντικών τιμών τους. Η ανάλυση χρονοσειρών ακολουθείται από τρία στάδια.

#### 3.5.1. Μελέτη και ανάλυση των δεδομένων

Πρόκειται για το πρώτο βήμα των αναλυτών στην ανάλυση της χρονοσειράς. Κύριος στόχος είναι η μελέτη και η περιγραφή των δεδομένων με την βοήθεια μετρήσεων και οπτικοποιήσεων. Υπάρχουν δύο τρόποι περιγραφής μιας χρονοσειράς:

- Απλή περιγραφή που γίνεται με την μελέτη των βασικών μεγεθών μιας χρονοσειράς (τάση, εποχικότητα, κυκλικότητα, τυχαία σφάλματα). Στην περιγραφή αυτή σημαντικό ρόλο παίζουν οι γραφικές απεικονίσεις των δεδομένων, καθώς από εκεί ο αναλυτής καταλαβαίνει πολλά για την συμπεριφορά της χρονοσειράς.
- Προχωρημένη περιγραφή που γίνεται με τη μελέτη στοχαστικών μοντέλων (στοχαστικές διαδικασίες) όπως είναι ο λευκός θόρυβος (white noise), ο τυχαίος περίπατος (random walk), η συνάρτηση συνδιασποράς κ.α. Υπάρχουν ειδικά εργαλεία τα οποία μπορούν να προσφέρουν στον αναλυτή σημαντικές πληροφορίες για τα χαρακτηριστικά των δεδομένων (μέση τιμή, τυπική απόκλιση, διακύμανση κτλ.). Με τον τρόπο αυτό εξετάζεται η ύπαρξη λευκού θορύβου στη χρονοσειρά, όπου θα πρέπει η μέση διακύμανση της χρονοσειράς να είναι σταθερή σε οποιαδήποτε χρονική στιγμή της χρονοσειράς. Ο λευκός θόρυβος δηλώνει ότι οι μεταβλητές είναι ανεξάρτητες μεταξύ τους.

Ένας άλλος σημαντικός έλεγχος που γίνεται είναι για τη στασιμότητα των δεδομένων, η οποία είναι εξαιρετικά σημαντική για την προβλεπτική ικανότητα και το ποσοστό επιτυχίας πρόβλεψης του μοντέλου που θα χρησιμοποιηθεί. Με τον όρο στασιμότητα εννοούμε κατά πόσο μεταβάλλεται η όχι η κοινή κατανομή των μεταβλητών μιας χρονοσειράς κατά την πάροδο του χρόνου. Στο στάδιο αυτό γίνεται επίσης ενδελεχής έλεγχος για λάθος ή ελλειπείς τιμές, καθώς τέτοιες τιμές δυσκολεύουν τις διαδικασίες ανάλυσης και πρόβλεψης. (Bashirahamad Fardin Momin 2017)

### 3.5.2. Προεπεξεργασία δεδομένων

Στο στάδιο αυτό, εφόσον έχει αποκτηθεί η κατάλληλη γνώση για τα χαρακτηριστικά των δεδομένων και τη συμπεριφορά τους κατά τη διάρκεια της χρονοσειράς, είναι απαραίτητο να υποστούν κάποια επεξεργασία πριν την τελική τους μορφή. Πολλές φορές τα δεδομένα περιλαμβάνουν λευκό θόρυβο, ή μπορεί να έχουν ελλειπίες τιμές. Η πρώτη ενέργεια των αναλυτών σε αυτό το στάδιο είναι η ανίχνευση λευκού θορύβου και η εφαρμογή κατάλληλων μεθόδων για την απομάκρυνση του. Αν δεν γίνει αντιληπτός ο λευκός θόρυβος τα αποτελέσματα της ανάλυσης ενδέχεται να είναι ανακριβή και αναξιόπιστα. Το ίδιο συμβαίνει και με τις ελλειπίες τιμές, δηλαδή εφαρμόζονται ειδικές τεχνικές με σκοπό την απομάκρυνσή τους. Τέτοιες τεχνικές είναι η απόδοση τιμής (imputation), η παρεμβολή τιμής (Interpolation) και η απαλοιφή τιμής.

Οι αλγόριθμοι που χρησιμοποιούνται για την ανάλυση χρονοσειρών είναι διαφορετικοί μεταξύ τους και έχουν τα δικά τους χαρακτηριστικά. Για το λόγο αυτό η χρονοσειρά πρέπει να είναι αντίστοιχα κατάλληλα διαμορφωμένη, όπως για παράδειγμα κάποια στατιστικά μοντέλα διαχειρίζονται καλύτερα χρονοσειρές που εμφανίζουν τάση, οι αλγόριθμοι νευρωνικών δικτύων αποδίδουν καλύτερα σε χρονοσειρές που τα δεδομένα ακολουθούν την κανονική κατανομή  $[0, 1]$ . Επομένως, είναι πιθανός ο μετασχηματισμός των δεδομένων μιας χρονοσειράς με κατάλληλες μεθόδους, για να υπάρχει μεγαλύτερη ακρίβεια στα αποτελέσματα. (Ramakrishnan 2003)

### 3.5.3. Μοντελοποίηση και Πρόβλεψη

Η πρόβλεψη μας δίνει σημαντικές πληροφορίες για τη συμπεριφορά μιας χρονοσειράς στο μέλλον που είναι και ο απώτερος σκοπός της ανάλυσης χρονοσειρών. Πρακτικά αυτό σημαίνει ότι τα στοιχεία της χρονοσειράς που προέρχονται από τις παρελθοντικές τιμές της μεταβλητής, μοντελοποιούνται με ειδικές τεχνικές και μεθόδους. Με βάση αυτή τη διαδικασία, παράγεται κάποια εκτίμηση για την τιμή της χρονοσειράς σε μελλοντικό χρόνο. Δηλαδή παράγεται νέα γνώση, πολύ χρήσιμη για τον άνθρωπο, ώστε να μπορέσει να την αξιοποιήσει για να πετύχει τους στόχους του. Η πρόβλεψη παίζει καθοριστικό ρόλο στη λήψη στρατηγικών αποφάσεων σε πολλούς τομείς (οικονομία, εμπόριο, δημογραφία, ιατρική, φυσική, κτλ.), λαμβάνοντας υπόψη τα αποτελέσματα που προκύπτουν από την παραπάνω διαδικασία.

### 3.6. Στασιμότητα Χρονοσειράς

Πολλές φορές οι χρονοσειρές παρουσιάζουν έντονες διακυμάνσεις που την χαρακτηρίζουν μη-στάσιμη χρονοσειρά. Τα περισσότερα μοντέλα πρόγνωσης απαιτούν στασιμότητα χρονοσειράς για να μπορέσουν να αποδώσουν όσο δυνατόν πιο ακριβή αποτελέσματα, διαφορετικά τα αποτελέσματα θα είναι ανακριβή. Οι γραφικές παραστάσεις της χρονοσειράς αποτελούν τον πιο απλό τρόπο για να καταλάβει κάποιος αν η χρονοσειρά είναι στάσιμη ή όχι. Υπάρχουν περιπτώσεις όμως που ο χαρακτηρισμός μιας χρονοσειράς από τις γραφικές παραστάσεις δεν επαρκεί. Για τον λόγο αυτό έχουν αναπτυχθεί κάποιες εξειδικευμένες μέθοδοι, που ελέγχουν αρκετά γρήγορα τη στασιμότητα μιας χρονοσειράς. Οι πιο σημαντικές και διαδεδομένες μέθοδοι είναι ο αυξημένος έλεγχος Dickey – Fuller (Dickey – Fuller 1981) και ο έλεγχος Kwiatkowski–Phillips–Schmidt–Shin. (Kwiatkowski -1992)

#### 3.6.1. Έλεγχος ADF (Augmented Dickey-Fuller test – ADF)

Ο αυξημένος έλεγχος ADF είναι μία στατιστική διαδικασία κατά την οποία ελέγχεται η αρχική υπόθεση του αναλυτή, δηλαδή ότι η χρονοσειρά παρουσιάζει στασιμότητα. Ο έλεγχος αυτός υπολογίζει το μέτρο σύμφωνα με το οποίο ο αναλυτής θα εγκρίνει ή θα απορρίψει την αρχική υπόθεση. Η διαδικασία αυτή στηρίζεται σε ένα μοντέλο αυτό-παλινδρόμησης και ελέγχει την συγκεκριμένη υπόθεση για διαφορετικά χρονικά βήματα μεταξύ των μεταβλητών της χρονοσειράς.

- $H_0$ : είναι η αρχική υπόθεση (μηδενική), η οποία αν εγκριθεί σημαίνει ότι η χρονοσειρά δεν είναι στάσιμη.
- $H_1$ : είναι η εναλλακτική υπόθεση κατά την οποία εάν απορριφθεί η μηδενική υπόθεση τότε η χρονοσειρά θεωρείται στάσιμη.

Κατά την εφαρμογή της συγκεκριμένης διαδικασίας παράγονται κάποια σημαντικά στοιχεία που αποτελούν τα κριτήρια αξιολόγησης του ελέγχου και είναι η ADF τιμή και η τιμή p-value. Για την τιμή p-value ισχύουν τα εξής:

- $p > 0,05$  : όπου ισχύει η μηδενική υπόθεση και η χρονοσειρά δεν θεωρείται στάσιμη.
- $p < 0,05$  : όπου δεν ισχύει η μηδενική υπόθεση και η χρονοσειρά θεωρείται στάσιμη.

Ο έλεγχος ADF λειτουργεί με μοντέλα που εκφράζονται από την παρακάτω μαθηματική σχέση:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta I \Delta y_{t-1} + \dots + \delta p - I \Delta y_{t-p+1} + \varepsilon_t$$

όπου,



- $\alpha$ , μια σταθερά
- $\beta$ , ο συντελεστής της τάσης
- $\rho$ , ο συντελεστής χρονικού βήματος της αυτοπαλινδρόμησης
- $\varepsilon$ , ο θόρυβος της χρονοσειράς
- $y$ , η τιμή της μεταβλητής
- $\Delta y$ , η διαφορά της τιμής της μεταβλητής με την προηγούμενη τιμή

Σε περίπτωση που οι συντελεστές  $\alpha$  και  $\beta$  είναι μηδέν, τότε η χρονοσειρά είναι τυχαίος περίπατος. Σύμφωνα με τον έλεγχο, ο συντελεστής  $\gamma$  είναι 0 ή εναλλακτικά ο συντελεστής  $\gamma$  είναι μικρότερος του 0.

### 3.6.2. Έλεγχος KPPS (Kwiatkowski–Phillips–Schmidt–Shin)

Ο έλεγχος KPPS είναι ένας συμπληρωματικός έλεγχος που γίνεται μετά τον ADF έλεγχο για διαπιστωθεί η στασιμότητα μιας χρονοσειράς. Σε αντίθεση με τον ADF η αρχική υπόθεση στηρίζεται στην ύπαρξη μοναδιαίας ρίζας στη χρονοσειρά και επομένως θεωρείται στάσιμη. Δηλαδή:

- $H_0$ : είναι η αρχική υπόθεση (μηδενική), η οποία αν εγκριθεί σημαίνει ότι η χρονοσειρά είναι στάσιμη.
- $H_1$ : είναι η εναλλακτική υπόθεση κατά την οποία εάν απορριφθεί η μηδενική υπόθεση τότε η χρονοσειρά δεν θεωρείται στάσιμη.

Από τα αποτελέσματα του ελέγχου KPPS προκύπτουν τα ακόλουθα συμπεράσματα:

- Όταν οι δυο έλεγχοι καταλήγουν στο συμπέρασμα ότι η χρονοσειρά δεν είναι στάσιμη, τότε ισχύει η μη-στασιμότητα της χρονοσειράς.
- Όταν οι έλεγχοι καταλήγουν στο συμπέρασμα ότι η χρονοσειρά είναι στάσιμη, τότε ισχύει η στασιμότητα της χρονοσειράς.
- Όταν ο έλεγχος KPPS δηλώνει στασιμότητα και ο ADF δηλώνει μη στασιμότητα, τότε η χρονοσειρά έχει στάσιμη τάση η οποία πρέπει να αφαιρεθεί για να υπάρξει στασιμότητα.

- Όταν ο έλεγχος KPPS δηλώνει μη στασιμότητα και ο ADF έλεγχος δηλώνει στασιμότητα, τότε η χρονοσειρά θεωρείται διαφορικά στάσιμη και πρέπει να γίνει διαφορικός μετασχηματισμός (differencing) για να είναι στάσιμη.

### 3.7. Στατιστικά μεγέθη χρονοσειράς

Στην ενότητα αυτή θα παρουσιαστούν τα πιο σημαντικά στατιστικά μεγέθη μιας χρονοσειράς, οι τιμές των οποίων έχουν κυρίαρχο ρόλο στην ανάλυση. Τα μεγέθη αυτά είναι τα εξής (Robert H. Shumway, 2011):

#### 3.7.1. Μέση τιμή (mean value)

Η μέση τιμή (mean value) μιας χρονοσειράς  $X$  δίνεται από την παρακάτω σχέση:

$$\mu_t = E[X_t]$$

η οποία δηλώνει την προσδοκώμενη τιμή μιας μεταβλητής σε χρόνο  $t$ . Η συγκεκριμένη τιμή σχετίζεται απόλυτα με την τάση της χρονοσειράς. Αν δηλαδή μια χρονοσειρά παρουσιάζει αύξηση ή μείωση της τάσης σε ένα συγκεκριμένο χρονικό διάστημα, τότε αυτή η μεταβολή θα επηρεάσει και τη μέση τιμή.

#### 3.7.2. Συνδιακύμανση (Covariance) και Αυτοσυνδιακύμανση (Autocovariance)

Πρόκειται για ένα στατιστικό μέγεθος που εκφράζει το βαθμό συσχέτισης μεταξύ δύο τυχαίων μεταβλητών  $Y$  και  $W$ . Η συνάρτηση συνδιακύμανσης ορίζεται ως εξής:

$$\gamma_{Y,W} = Cov(Y, W) = E[(Y - \mu_Y)(W - \mu_W)]$$

και εκφράζει τη γραμμική σχέση δύο οποιονδήποτε τιμών της χρονοσειράς σε διαφορετικούς χρόνους. Επίσης για τη χρονοσειρά  $X$ , η συνάρτηση αυτοσυνδιακύμανσης καθορίζει την αλληλεπίδραση μεταξύ δύο τιμών της χρονοσειράς,  $X_t$  και  $X_{t-j}$ , οι οποίες καταγράφηκαν σε αντίστοιχες περιόδους  $t$  και  $t-j$ . Η συνάρτηση αυτοσυνδιακύμανσης ορίζεται ως εξής:

$$\gamma_j = Cov(X_t, X_{t-j}) = E[(X_t - E(X_t))(X_{t-j} - E(X_{t-j}))]$$

Για  $j=0$ , η αυτοσυνδιακύμανση ταυτίζεται με τη διακύμανση, δηλαδή

$$\gamma_0 = \text{Var}(X_t) = \sigma^2$$

### 3.7.3. Αυτοσυσχέτιση (Autocorrelation)

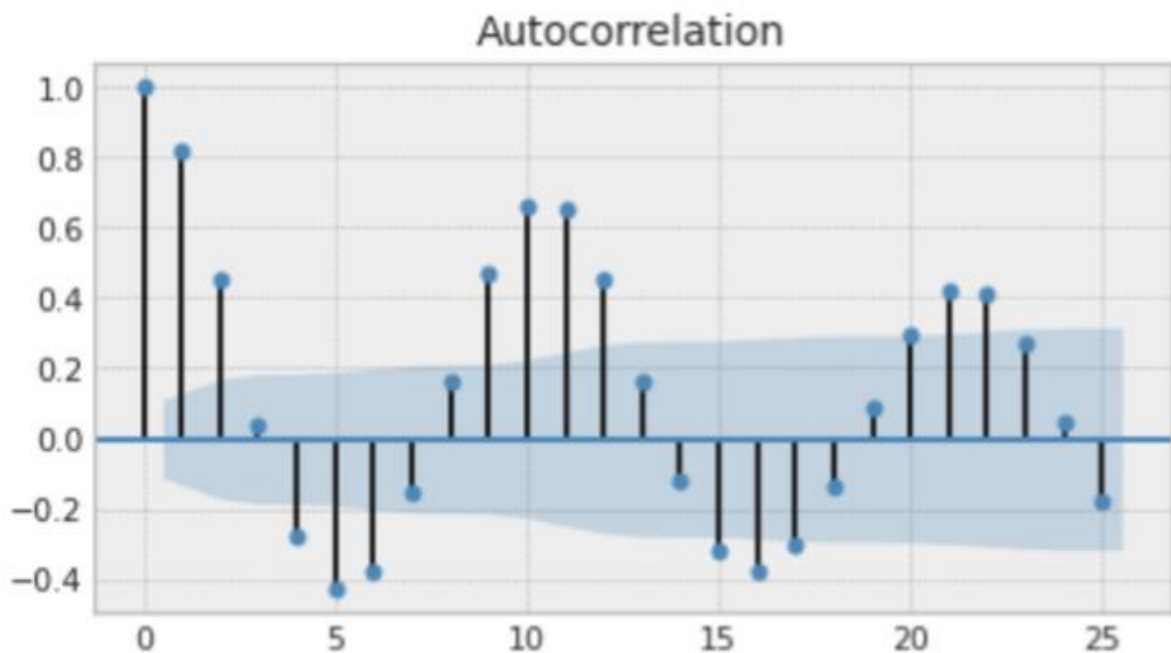
Ο συντελεστής αυτοσυσχέτισης είναι ένας σημαντικός δείκτης που χρησιμοποιείται στην ανάλυση χρονοσειρών και εκφράζει τον βαθμό αλληλεξάρτησης μεταξύ των διαδοχικών παρατηρήσεων για τον έλεγχο τυχαιότητας των χρονοσειρών. Ο συντελεστής αυτοσυσχέτισης ορίζεται ως:

$$\rho_j = \frac{\gamma_j}{\gamma_0} = \frac{E[(X_t - \mu_t)(X_{t-j} - \mu_{t-j})]}{E[(X_t - \mu_t)^2]}$$

Η συνάρτηση αυτοσυσχέτισης εκφράζει την αυτοσυσχέτιση της  $j$ -στής τάξης  $\rho_j$  της τυχαίας μεταβλητής  $X_t$  με μια προγενέστερη εκδοχή της  $X_{t-j}$ . Ο συντελεστής  $\rho$  λαμβάνει ένα εύρος τιμών που κυμαίνεται εντός του διαστήματος  $[-1,1]$ . Επομένως:

- Αν  $\rho \approx \pm 1$ , τότε ο συντελεστής φανερώνει ισχυρή γραμμική συσχέτιση μεταξύ των μεταβλητών.
- Αν  $\rho \approx 0$ , τότε ο συντελεστής φανερώνει μη γραμμική συσχέτιση των μεταβλητών.

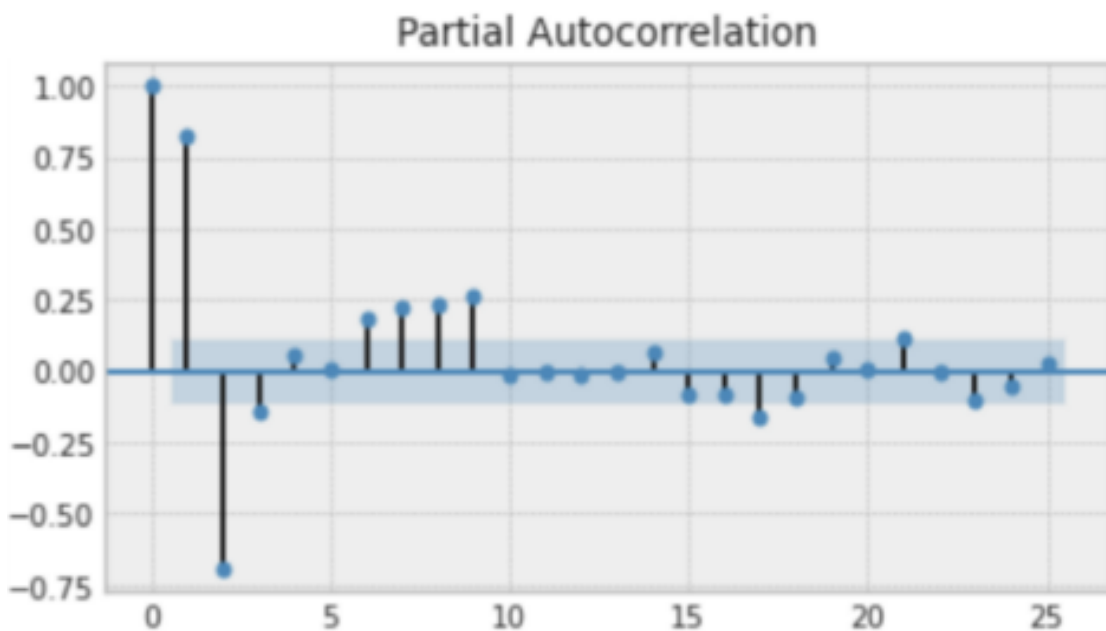
Η γραφική παράσταση της αυτοσυσχέτισης μιας χρονοσειράς ονομάζεται ACF (Autocorrelation Function), απεικονίζει χρήσιμες πληροφορίες για βασικά χαρακτηριστικά της χρονοσειράς, όπως της τάσης, της εποχικότητας και της στασιμότητας των χρονοσειρών.



Εικόνα 23 Αυτοσυσχέτιση

### 3.7.4. Μερική Αυτοσυσχέτιση (Partial autocorrelation)

Η αυτοσυσχέτιση μεταξύ δύο παρατηρήσεων καθορίζεται τόσο από την άμεση συσχέτιση όσο και από τις έμμεσες συσχετίσεις. Η μερική αυτοσυσχέτιση PACF (Partial Autocorrelation), είναι θεωρητικά μία επέκταση της αυτοσυσχέτισης και περιγράφει τη σχέση που υπάρχει μεταξύ δύο παρατηρήσεων  $x$  και  $y$  σε μια χρονοσειρά, με την προϋπόθεση ότι αφαιρούνται οι υπόλοιπες παρατηρήσεις ( $z_1, \dots, z_n$ ) που επιδρούν πάνω σε αυτές. Η μερική αυτοσυσχέτιση δείχνει πιο σημαντικές πληροφορίες από την αυτοσυσχέτιση για χρονοσειρές που συνδυάζουν περισσότερες από μια συνιστώσες, όπως για παράδειγμα μια χρονοσειρά που εμφανίζει ταυτόχρονα εποχικότητα και κυκλικότητα. Επομένως η συνάρτηση μερικής αυτοσυσχέτισης παρέχει μία πιο ολοκληρωμένη και καθαρή εικόνα για τις σχέσεις εξάρτησης μεταξύ των μεταβλητών μιας χρονοσειράς που δεν επηρεάζονται από άλλες.



Εικόνα 24 Μερική Αυτοσυσχέτιση

## Κεφάλαιο 4<sup>ο</sup>: Μοντέλα Πρόβλεψης Χρονοσειρών

### 4.1. Στοχαστικά μοντέλα

Με τον όρο στοχαστική διαδικασία εννοούμε ένα σύνολο τυχαίων μεταβλητών που εξαρτάται από μία παράμετρο ή ένα πόρισμα. Συνήθως στην ανάλυση χρονοσειρών η παράμετρος αυτή είναι ο χρόνος.

Όπως αναφέρθηκε ο λευκός θόρυβος (white noise) και ο τυχαίος περίπατος (random walk) είναι στοχαστικά μοντέλα χρονοσειρών που παρουσιάζουν ιδιαίτερο ενδιαφέρον.

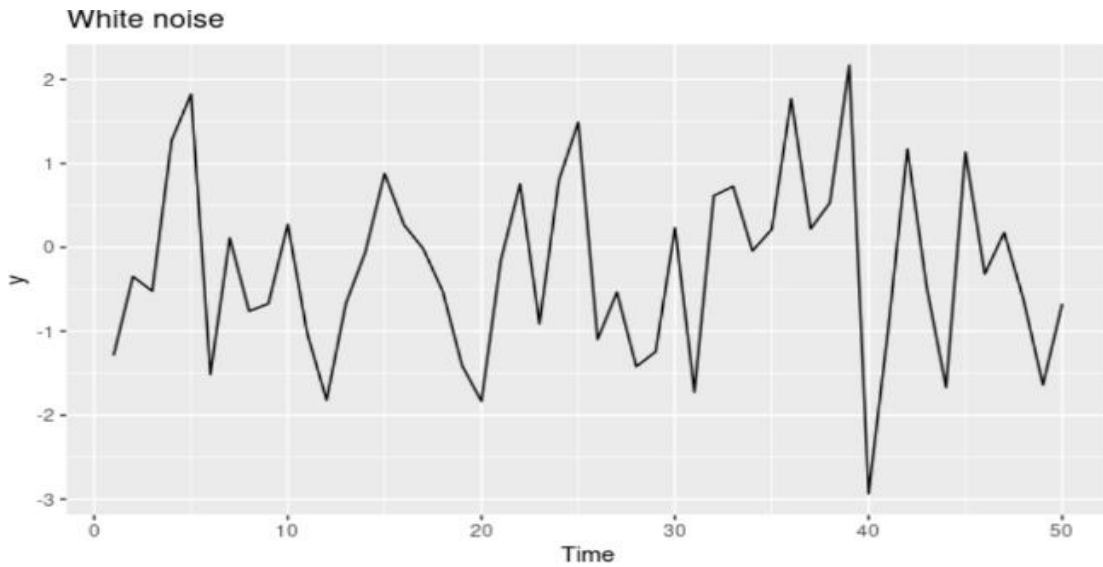
#### 4.1.1. Λευκός θόρυβος (White noise)

Λευκός θόρυβος ορίζεται ως η απλούστερη μορφή χρονοσειράς η οποία δεν εμφανίζει τάση ή περιοδικότητα. Όλες οι παρατηρήσεις της συγκεκριμένης χρονοσειράς είναι ανεξάρτητες και ομοιόμορφα κατανομημένες στο χώρο. Για το λόγο αυτό η χρονοσειρά εκτός από white noise είναι γνωστή και ως iid noise (independent, identically distributed).

Μαθηματικά και με τη βοήθεια των πιθανοτήτων, αν  $Y_1, Y_2, \dots, Y_n$  είναι τυχαίες μεταβλητές και  $y_1, y_2, \dots, y_n$  είναι πραγματικοί αριθμοί τότε η χρονοσειρά λευκός θόρυβος ορίζεται ως εξής:

$$P(Y_1 \leq y_1, Y_2 \leq y_2, \dots, Y_n \leq y_n) = P(Y_1 \leq y_1) \times P(Y_2 \leq y_2) \times \dots \times P(Y_n \leq y_n)$$

Από αυτή την μαθηματική συνάρτηση αποδεικνύεται ότι η μέση τιμή του λευκού θορύβου είναι μηδενική  $E(Y_t) = 0$  και η διασπορά του παραμένει σταθερή κατά την πάροδο του χρόνου  $E(Y_i Y_j) = \sigma_w^2$ . Τα δύο αυτά χαρακτηριστικά περιγράφουν το συγκεκριμένο μοντέλο ως μια στάσιμη χρονοσειρά (Brockwell-Davies 2016).



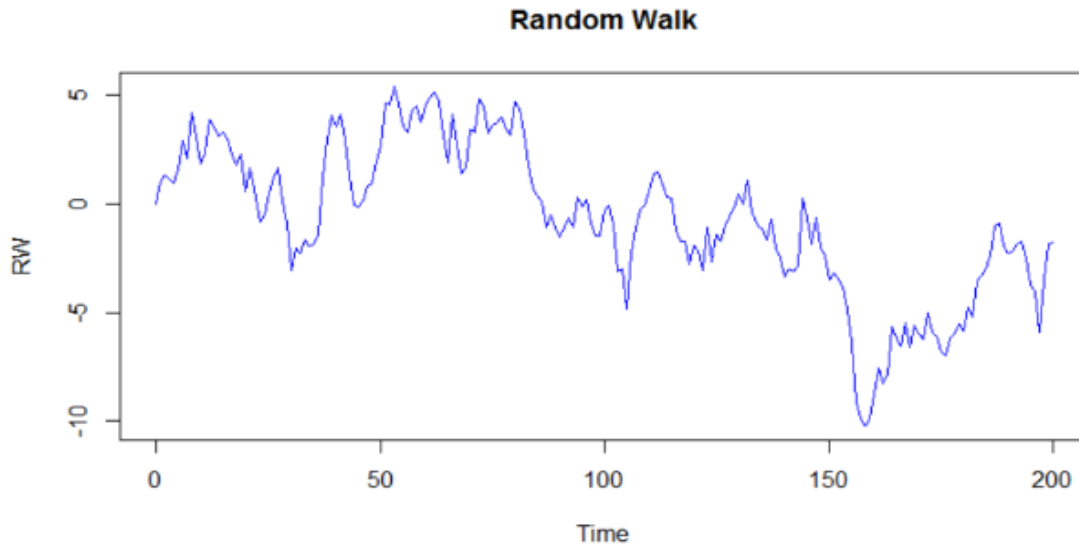
Εικόνα 25 Διάγραμμα Λευκού Θορύβου

#### 4.1.2. Τυχαίος περίπατος (Random walk)

Τυχαίος περίπατος ορίζεται ως μια μη στάσιμη χρονοσειρά όπου κάθε τυχαία μεταβλητή  $Y$  σε μία χρονική στιγμή  $t$  είναι αποτέλεσμα της προηγούμενης τυχαίας μεταβλητής στην χρονική στιγμή  $t-1$ , με την προϋπόθεση ότι σε αυτή την περίπτωση προστίθεται μία τυχαία μεταβλητή  $X$  (λευκός θόρυβος). Δηλαδή:

$$Y_t = Y_{t-1} + X_t$$

Επομένως από την παραπάνω μαθηματική συνάρτηση αποδεικνύεται ότι για τον τυχαίο περίπατο η μέση τιμή μιας μεταβλητής είναι μηδενική  $E(Y_t) = 0$  και η διασπορά της προκύπτει από το γινόμενο του χρόνου  $t$  και της διασποράς του λευκού θορύβου  $E(Y_t^2) = t \times \sigma_w^2$ .



Εικόνα 26 Διάγραμμα Τυχαίου Περιπάτου

Τα δύο αυτά στοχαστικά μοντέλα, ο λευκός θόρυβος και ο τυχαίος περίπατος, είναι τα πιο απλά μοντέλα περιγραφής και πρόβλεψης χρονοσειρών. Δεν έχουν κάποιο ιδιαίτερο ενδιαφέρον από μόνα τους, αλλά αποτελούν τη βάση για τη δημιουργία πιο σύνθετων μοντέλων πρόβλεψης.

## 4.2. Γραμμικά μοντέλα

### 4.2.1. Γραμμική Παλινδρόμηση (Linear Regression)

Η γραμμική παλινδρόμηση είναι μία διαδικασία που γίνεται για τη μελέτη και μοντελοποίηση του βαθμού συσχέτισης μεταξύ μιας εξαρτημένης μεταβλητής  $Y$  και μίας ανεξάρτητης μεταβλητής  $X$  και περιγράφεται μέσω μιας απλής συνάρτησης. Η συνάρτηση αυτή περιγράφει τον τρόπο με τον οποίο οι μεταβολές της ανεξάρτητης μεταβλητής  $X$  επηρεάζει την εξαρτημένη μεταβλητή  $Y$ . Στην περίπτωση που υπάρχουν περισσότερες από μία ανεξάρτητες μεταβλητές, η διαδικασία ονομάζεται πολλαπλή γραμμική παλινδρόμηση.

### 4.2.3. Απλή Γραμμική Παλινδρόμηση ( Simple Linear Regression)

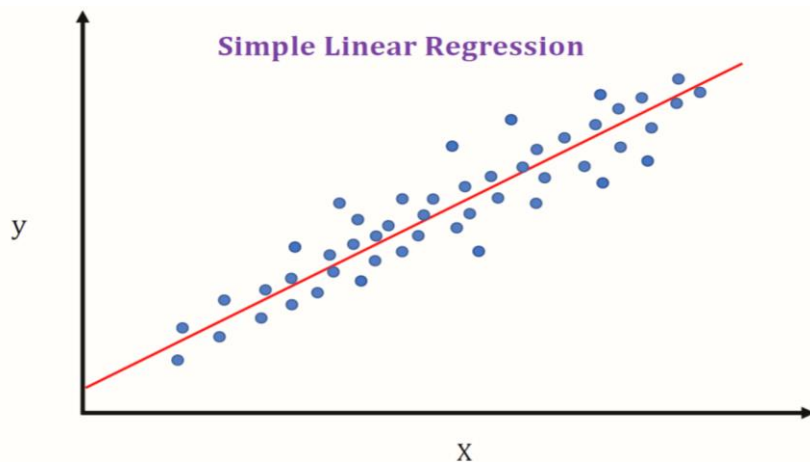
Η απλή γραμμική παλινδρόμηση (simple linear regression) είναι η πιο απλή μορφή γραμμικής παλινδρόμησης και εκφράζει την γραμμική σχέση μεταξύ δύο μεταβλητών  $Y$  και  $X$ . Η μεταβλητή  $Y$  αφορά την εξαρτημένη μεταβλητή και η  $X$  την ανεξάρτητη μεταβλητή (Montgomery, 2015). Η απλή γραμμική παλινδρόμηση ορίζεται μαθηματικά ως εξής:

$$Y_t = a_0 + a_1 X_t + \varepsilon_t$$

όπου,

- $Y_t$  είναι η εξαρτημένη μεταβλητή
- $a_0, a_1$  είναι οι άγνωστοι συντελεστές του μοντέλου
- $X_t$  είναι η ανεξάρτητη μεταβλητή
- $\varepsilon_t$  είναι το τυχαίο σφάλμα

Διαγραμματικά διακρίνεται ότι η απλή γραμμική παλινδρόμηση είναι μία ευθεία γραμμή:



Εικόνα 27 Απλή Γραμμική Παλινδρόμηση

### 4.2.4. Αυτοπαλίνδρομο Μοντέλο (Auto Regressive Model) AR (p)

Τα αυτοπαλίνδρομα μοντέλα (AR) είναι μοντέλα γραμμικής παλινδρόμησης και ορίζονται ως εξής: (Brockwell 2002)

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t$$

Όπου,

- $X_t$  είναι η τυχαία μεταβλητή της χρονοσειράς  $X$  σε μια χρονική στιγμή  $t$



- $X_{t-p}$  είναι η ανεξάρτητη μεταβλητή της χρονοσειράς  $X$  σε προηγούμενες χρονικές στιγμές, δηλαδή  $X_{t-1}, \dots, X_{t-p}$
- Ο δείκτης  $p$  δηλώνει την τάξη (order) του μοντέλου και αναφέρεται στο μέγεθος των χρονικών βημάτων, δηλαδή AR(1): Αυτοπαλίνδρομο 1ης τάξης, AR(2): Αυτοπαλίνδρομο 2ης τάξης, κτλ.
- $a_0, a_1, a_p$  είναι οι παράμετροι του μοντέλου
- $\varepsilon_t$  εκφράζει το λευκό θόρυβο και ακολουθεί την κανονική κατανομή  $N(0, \sigma^2)$ .

Πρόκειται για μία παλινδρόμηση της ίδιας μεταβλητής μεταξύ παροντικών και παρελθοντικών χρονικών στιγμών  $t$ . Οι παράμετροι  $a_0, a_1, a_p$  και η διακύμανση του λευκού θορύβου είναι οι παράγοντες που επηρεάζουν και καθορίζουν το μοντέλο (AR), επομένως και την πρόβλεψη.

Για παράδειγμα το αυτοπαλίνδρομο μοντέλο 1<sup>ης</sup> τάξης AR(1) ορίζεται από την παρακάτω μαθηματική σχέση:

$$X_t = a_0 + a_1 X_{t-1} + \varepsilon_t$$

Εδώ έχουμε τις εξής περιπτώσεις:

- $a_1 = 0$ , όπου προκύπτει η συνάρτηση του λευκού θορύβου
- $a_1 = 1$  και  $a_0 = 0$ , όπου προκύπτει η συνάρτηση του τυχαίου περιπάτου
- $a_1 = 1$  και  $a_0 \neq 0$ , όπου προκύπτει η συνάρτηση του τυχαίου περιπάτου (έχοντας μετακινηθεί κατά  $a_0$ )
- $a_1 < 0$ , όπου φθίνει προς τη μέση τιμή.

Για τις στάσιμες χρονοσειρές ισχύει ότι  $-1 < a_1 < 1$ .

#### 4.2.5. Μοντέλο Κινητού Μέσου (Moving Average)

Όπως και το αυτοπαλίνδρομο μοντέλο (AR), έτσι και το μοντέλο κινητού μέσου (MA) ανήκει στην κατηγορία των γραμμικών μοντέλων πρόβλεψης. Η διαφορά μεταξύ των δυο μοντέλων είναι ότι για το μοντέλο (AR) εξετάζεται η μελλοντική τιμή της ίδιας μεταβλητής που προκύπτουν από την ανάλυση και επεξεργασία των παρελθοντικών τιμών της, ενώ για το μοντέλο κινητού μέσου (MA) βασικό αντικείμενο ανάλυσης και επεξεργασίας είναι οι παρελθοντικές στιγμές του τυχαίου σφάλματος (λευκός θόρυβος).

Επομένως, μια χρονοσειρά  $X$  είναι χρονοσειρά κινητού μέσου MA( $q$ ), όταν κάθε παρατήρηση  $X_t$  της χρονοσειράς είναι αποτέλεσμα ενός σταθμισμένου αθροίσματος μιας

σταθεράς  $\mu$ , του λευκού θορύβου και των  $q$  παρελθοντικών εκδοχών του λευκού θορύβου. Η συνάρτηση που περιγράφει αυτή τη σχέση ορίζεται ως εξής: (Brockwell 2002)

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

όπου,

- $X_t$  είναι η εξαρτημένη μεταβλητή
- $\mu$  και  $\theta_1, \theta_2, \dots, \theta_q$  είναι η παράμετροι του μοντέλου
- $q$  οι παρελθοντικές τιμές του λευκού θορύβου
- $\varepsilon_t$  είναι ο λευκός θόρυβος και ακολουθεί την κανονική κατανομή  $N(0, \sigma^2)$ .

Για παράδειγμα το μοντέλο κινητού μέσου MA(1), ορίζεται από την παρακάτω μαθηματική σχέση:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Και εδώ, όπως και για τα αυτοπαλίνδρομα μοντέλα (AR), ισχύει ότι για τις στάσιμες χρονοσειρές  $-1 < \theta_1 < 1$ .

#### 4.2.6. Αυτοπαλίνδρομο Μοντέλο Κινητού Μέσου Όρου (ARMA)

Υπάρχουν κάποιες στάσιμες χρονοσειρές που μπορεί να παρουσιάζουν χαρακτηριστικά και ιδιότητες και από τις δύο κατηγορίες AR ή MA. Στην περίπτωση αυτή, αυτές οι χρονοσειρές μοντελοποιούνται σε ένα γενικότερο μικτό μοντέλο που προκύπτει από τον συνδυασμό ενός AR(p) μοντέλου και ενός MA(q) μοντέλου και ονομάζεται αυτοπαλίνδρομο μοντέλο κινητού μέσου όρου τάξης (p, q). Επομένως κάθε παρατήρηση  $X_t$  μιας ARMA(p, q) χρονοσειράς ορίζεται ως εξής: (Brockwell 2002)

$$X_t = a_0 + a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}$$

όπου  $-1 < a_i < 1$  και  $-1 < \theta_i < 1$  για κάθε  $i = 1, 2, \dots, n$ .

Από αυτή την συνάρτηση, γίνεται κατανοητό ότι η στασιμότητα μιας χρονοσειράς ορίζεται από το AR μέρος της και η αντιστρεψιμότητα της χρονοσειράς ορίζεται από το MA μέρος της. Αν ένας από τους δύο δείκτες είναι μηδενικός τότε έχουμε τις εξής περιπτώσεις:

- ARMA(1,0) και αντιστοιχεί σε μοντέλο AR(1) που είναι το αυτοπαλίνδρομο μοντέλο AR(p)

- ARMA(0,1) και αντιστοιχεί σε μοντέλο MA(1) που είναι το μοντέλο κινητού μέσου MA(q).

#### 4.2.7. Μικτό ολοκληρωμένο μοντέλο ARIMA (Autoregressive Integrated Moving Average model)

Στο σημείο αυτό θα αναλύσουμε το μικτό ολοκληρωμένο μοντέλο ARIMA, το οποίο θα χρησιμοποιηθεί και θα αξιολογηθεί στο ερευνητικό μέρος της παρούσας διπλωματικής εργασίας. Οι περισσότερες χρονοσειρές δεν χαρακτηρίζονται ως στάσιμες. Μη στάσιμη χρονοσειρά χαρακτηρίζεται μια χρονοσειρά όταν δεν έχει σταθερή τάση και διακύμανση. Για την καλύτερη δυνατή ανάλυση και την αποφυγή διάφορων προβλημάτων που μπορεί να προκύψουν στην συνέχεια, το πρώτο βήμα που πρέπει να γίνει, είναι η μετατροπή των μη στάσιμων χρονοσειρών σε στάσιμες.

Το μοντέλο ARIMA είναι το πιο γνωστό μοντέλο ανάλυσης χρονοσειρών και προκύπτει από το συνδυασμό δύο διαφορετικών μοντέλων, της αυτοπαλινδρόμησης (AR), του κινητού μέσου όρου (MA) και του συντελεστή διαφοροποίησης (Integration) που εφαρμόζεται για το μετασχηματισμό των μη στάσιμων χρονοσειρών σε στάσιμες. (Nelson, 2008)

Έτσι, όταν μια χρονοσειρά μετατρέπεται σε στάσιμη, έχοντας εφαρμόσει τις πρώτες διαφορές και τον τελεστή υστέρησης B, η χρονοσειρά ονομάζεται ολοκληρωμένη πρώτης τάξης και συμβολίζεται με I(1), δηλαδή:

$$X_t - X_{t-1} = (1 - B) X_t = \Delta X_t$$

Αν μια χρονοσειρά μετατρέπεται σε στάσιμη χρησιμοποιώντας τις δεύτερες διαφορές, ονομάζεται ολοκληρωμένη δεύτερης τάξης και συμβολίζεται με I(2). Επομένως μια χρονοσειρά ονομάζεται ολοκληρωμένη σειρά d τάξης και συμβολίζεται με I(d), όπου d ο αριθμός των διαφορών που εφαρμόζεται για την μετατροπή της χρονοσειράς από μη στάσιμη σε στάσιμη.

Όπως αναφέρθηκε, το μοντέλο ARIMA(p, d, q) προκύπτει από το συνδυασμό των μοντέλων AR και MA με την προσθήκη του συντελεστή διαφοροποίησης I(d) και μαθηματικά ορίζεται ως εξής:

$$\Phi(B)(1 - B)^d X_t = c + \Theta(B) \varepsilon_t$$

όπου το αριστερό μέρος της εξίσωσης δηλώνει το μοντέλο  $AR(p)$  και το βαθμό διαφοροποίησης της χρονοσειράς και το δεξί μέρος δηλώνει το μοντέλο  $MA(q)$ .

Χαρακτηριστικές περιπτώσεις για το μοντέλο  $ARIMA(p, d, q)$  είναι οι εξής:

- αν  $d=0$ , τότε το μοντέλο χαρακτηρίζεται ως  $ARIMA(p, 0, q)$  και είναι ίδιο με το μοντέλο  $ARMA(p,d)$
- αν  $d=0, q=0$ , τότε το μοντέλο χαρακτηρίζεται ως  $ARIMA(p, 0, 0)$  και είναι ίδιο με το αυτοπαλίνδρομο μοντέλο  $AR(p)$
- αν  $d=0, p=0$ , τότε το μοντέλο χαρακτηρίζεται ως  $ARIMA(0, 0, q)$  και είναι ίδιο με το μοντέλο  $MA(q)$ .
- αν  $d=1, p=0, q=0$ , τότε το μοντέλο χαρακτηρίζεται ως  $ARIMA(0, 1, 0)$  και είναι ίδιο με τον τυχαίο περίπατο.
- αν  $d=0, p=0, q=0$ , τότε το μοντέλο χαρακτηρίζεται ως  $ARIMA(0, 0, 0)$  και είναι ίδιο με το λευκό θόρυβο.

### 4.3. Τεχνητά Νευρωνικά Δίκτυα (ANN)

Η βαθιά μάθηση είναι μία μορφή μηχανικής μάθησης που χρησιμοποιεί τεχνητά νευρωνικά δίκτυα για τη μάθηση και πρόβλεψη. Τα ΤΝΔ είναι ένα σύνολο στατιστικών μοντέλων μάθησης και η ορολογία τους είναι εμπνευσμένη από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου. Βασικός στόχος των ΤΝΔ είναι η κατασκευή ενός μοντέλου το οποίο θα μπορεί να μιμηθεί την ανθρώπινη νοημοσύνη και να αποκτήσει γνώση από αυτή την εμπειρία, η οποία με τη σειρά της θα παρέχει χρήσιμες πληροφορίες

Στον ανθρώπινο εγκέφαλο υπάρχουν οι βιολογικοί νευρώνες, δηλαδή νευρικά κύτταρα που επικοινωνούν μεταξύ τους κατά μήκος δεσμών που ονομάζονται συνάψεις. Ένα απλό παράδειγμα για τον τρόπο λειτουργίας των βιολογικών νευρώνων είναι ο άνθρωπος όταν μαθαίνει να περπατάει. Κατά τη διάρκεια της μάθησης, οι νευρώνες αυτοί που είναι υπεύθυνοι για αυτή τη λειτουργία, επικοινωνούν όλο και καλύτερα κάθε φορά και ενεργοποιούνται κάθε φορά που ο άνθρωπος χρειάζεται να περπατήσει.

Ένα ΤΝΔ αποτελείται από ένα μεγάλο αριθμό διασυνδεδεμένων στοιχείων επεξεργασίας που ονομάζονται νευρώνες, τα οποία επικοινωνούν και εργάζονται από κοινού για την επίλυση συγκεκριμένων προβλημάτων.

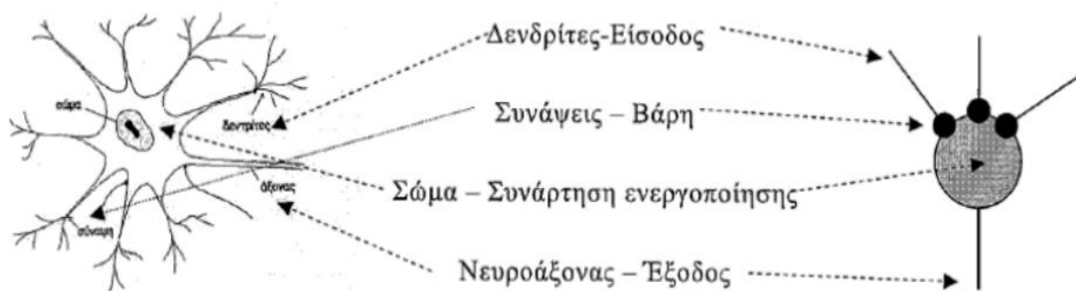
Ο τρόπος με τον οποίο συνδέονται αυτοί οι νευρώνες ενός δικτύου ονομάζεται τοπολογία ή αρχιτεκτονική του τεχνητού νευρωνικού δικτύου. Τα ΤΝΔ υιοθετούν την

αρχιτεκτονική του ανθρώπινου εγκεφάλου και η εκπαίδευσή τους γίνεται όπως και στα βιολογικά νευρωνικά δίκτυα, δηλαδή μέσα από ένα σύνολο παραδειγμάτων. Η αρχιτεκτονική των ΤΝΔ καθορίζει και την κατηγορία στην οποία ανήκουν. Τα ΤΝΔ διακρίνονται σε δυο κατηγορίες:

- πρόσθιας τροφοδότησης (feedforward)
- οπίσθιας τροφοδότησης (feedbackward)

Όπως αναφέρθηκε παραπάνω, το βασικό δομικό στοιχείο του εγκεφάλου είναι ο νευρώνας. Τόσο στα ΒΝΔ όσο και στα ΤΝΔ ο νευρώνας αποτελείται από τα εξής τμήματα (Κ. Διαμαντάρας – Δ. Μπότσης 2019):

- τον πυρήνα (σώμα-συνάρτηση ενεργοποίησης)
- τους δενδρίτες, που αποτελούν τις πύλες εισόδου του νευρώνα. Δέχονται πληροφορίες από άλλους νευρώνες.
- τον άξονα, που αποτελεί την πύλη εξόδου του νευρώνα και ο οποίος στέλνει σήματα σε άλλους νευρώνες.
- τις συνάψεις, που αποτελούν τα σημεία ένωσης του άξονα ενός νευρώνα και των δενδριτών από άλλους νευρώνες. Το ποσοστό της πληροφορίας που μεταδίδεται από τον άξονα στον δενδρίτη ονομάζεται συναπτικό βάρος. Τα σημεία συνάψεων των ΒΝΔ όπως και τα συναπτικά βάρη για τα ΤΝΔ, είναι υπεύθυνα για την αποθήκευση της γνώσης που παράγεται.



Εικόνα 28 Δομή Τεχνητού Νευρωνικού Δικτύου

Τα ΤΝΔ επεξεργάζονται πληροφορίες που προέρχονται από τις εισόδους. Κάθε τεχνητός νευρώνας αποτελείται από πολλές εισόδους και μία μόνο έξοδο. Κάθε είσοδος

καθορίζεται από το συναπτικό βάρος και τα αποτελέσματα των εισόδων αθροίζονται μέσω της μαθηματικής συνάρτησης αθροίσματος:

$$F = \sum_i^n x_i w_i$$

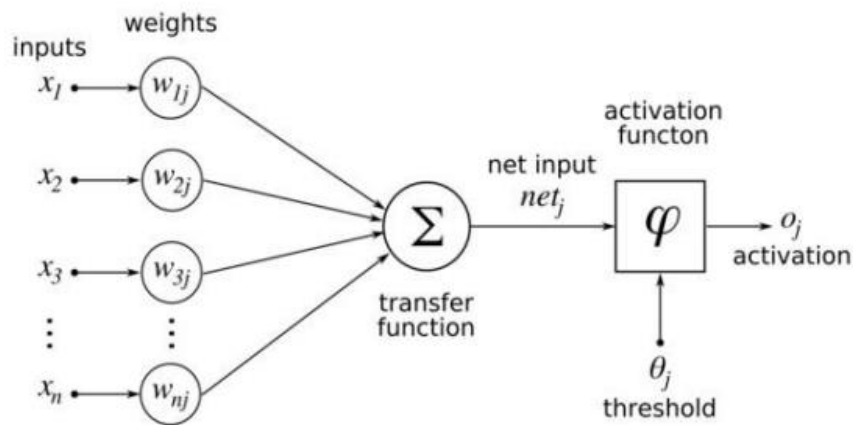
Όπου,

- $x_i$  είναι ο αριθμός των εισόδων
- $w_i$  είναι το συναπτικό βάρος των εισόδων

Ο τεχνητός νευρώνας δίνει μία μόνο έξοδο, μόνο όταν το άθροισμα των εισόδων είναι μεγαλύτερο μιας ορισμένης τιμής κατωφλίου  $\theta$  και εκφράζεται μέσω της παρακάτω συνθήκης:

$$\sum_i^n x_i w_i - \theta > 0$$

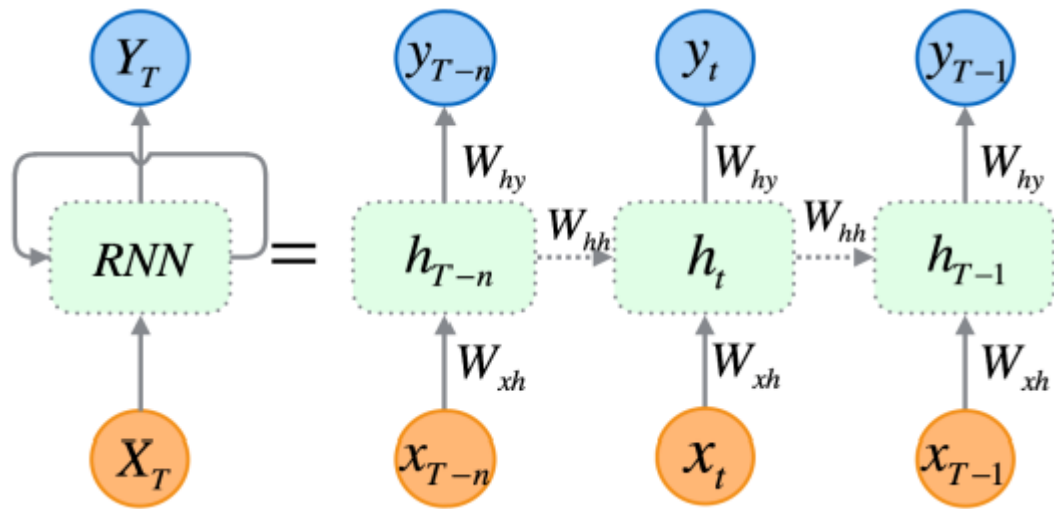
Τα ΤΝΔ αποτελούνται από επίπεδα (layers). Υπάρχουν περιπτώσεις που μπορεί να υπάρχουν και ενδιάμεσα επίπεδα τα οποία ονομάζονται κρυμμένα επίπεδα (hidden layers). Κάθε επίπεδο αποτελείται από έναν αριθμό κόμβων (nodes) που συνδέονται με τέτοιο τρόπο ώστε ένας κόμβος να συνδέεται με πολλούς άλλους κόμβους του ίδιου ή άλλου επιπέδου. Όπως αναφέρθηκε κάθε κόμβος λαμβάνει το σταθμισμένο άθροισμα όλων των εισόδων που καταλήγουν σε αυτόν και στη συνέχεια δημιουργείται μία μοναδική έξοδος, εάν το άθροισμα είναι μεγαλύτερο από μία τιμή κατωφλίου. Τέλος οι εισοδοί (inputs) επικοινωνούν με τα κρυμμένα επίπεδα, τα οποία με τη σειρά τους συνδέονται με την έξοδο (output). (Georgouli, A. 2015).



Εικόνα 29 Τρόπος λειτουργίας Τεχνητού Νευρωνικού Δικτύου

### 4.3.1. Αναδρομικά Νευρωνικά Δίκτυα (Recurrent Neural Networks)

Τα αναδρομικά νευρωνικά δίκτυα (RNN) είναι μια κατηγορία νευρωνικών δικτύων και παρουσιάζει μια σημαντική διαφορά σε σχέση με τα άλλα νευρωνικά δίκτυα. Τα RNN δίκτυα έχουν την ιδιότητα να αποθηκεύουν συνεχώς παρελθοντικές καταστάσεις του συστήματος. Τα υπόλοιπα νευρωνικά δίκτυα επεξεργάζονται όλο το σύνολο των πληροφοριών μαζί, ενώ τα RNN δίκτυα επεξεργάζονται τα δεδομένα με τη σειρά διατηρώντας πάντα στην μνήμη παρελθοντικές καταστάσεις του συστήματος. Η μνήμη αυτή δημιουργείται μέσω ενός βρόχου σε κάθε κόμβο του δικτύου, ο οποίος με τη σειρά του διατηρεί και μεταφέρει πληροφορίες στον ίδιο κόμβο από την μία χρονική στιγμή στην άλλη. Η ιδιότητα αυτή καθιστούν τα RNN δίκτυα τα ιδανικά μοντέλα για προβλήματα όπου η χρονική σειρά των δεδομένων είναι πολύ σημαντική όπως για παράδειγμα η πρόβλεψη των τιμών των κρυπτονομισμάτων, όπου σε αυτή την περίπτωση πρέπει να γνωρίζουμε όχι μόνο τα σημερινά στοιχεία, αλλά και τα παρελθοντικά.



Εικόνα 30 Τρόπος λειτουργίας Αναδρομικών Νευρωνικών Δικτύων

Η μαθηματική συνάρτηση των εξισώσεων ενός αναδρομικού νευρωνικού δικτύου είναι η εξής:

$$y_t = \phi(W x_t + U h_{t-1} + b)$$

Όπου,

- $y_t$ , η κρυφή αναπαράσταση τη χρονική στιγμή  $t$
- $x_t$ , το διάνυσμα του στοιχείου της ακολουθίας τη χρονική στιγμή  $t$
- $\Phi$ , είναι η μη γραμμική συνάρτηση ενεργοποίησης
- $W$ , είναι ο πίνακας βαρών για την είσοδο  $x_t$
- $U$ , είναι ο πίνακας βαρών που επιδρούν στην έξοδο του δικτύου την προηγούμενη χρονική στιγμή
- $b$ , είναι ένα διάνυσμα πόλωσης

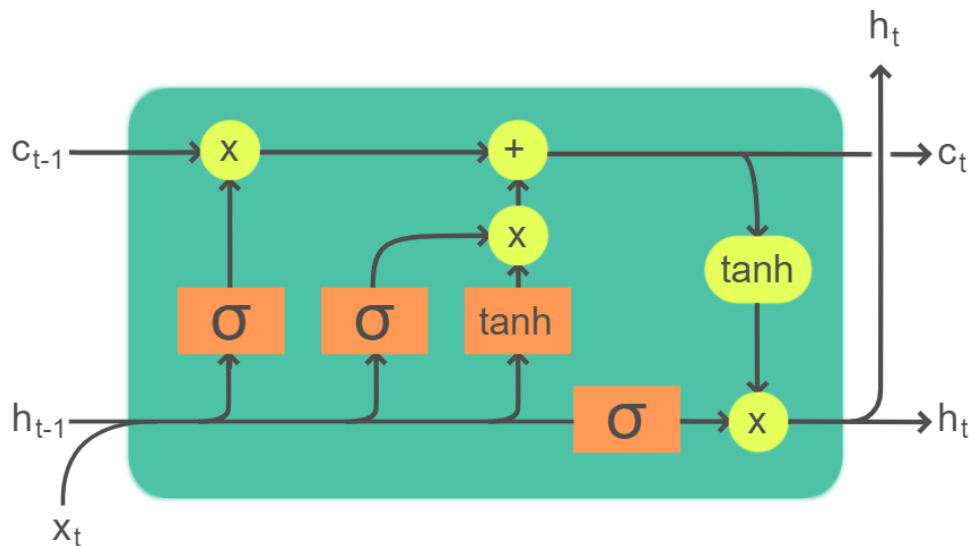
#### 4.3.2. Νευρωνικά Δίκτυα LSTM (Long Short-Term Memory)

Στο σημείο αυτό θα περιγράψουμε τα μοντέλα LSTM καθώς αποτελούν μέρος της ερευνάς την παρούσας διπλωματικής εργασίας. Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (LSTM) ανήκουν στην κατηγορία των ανατροφοδοτούμενων νευρωνικών δικτύων (RNN) και μπορούν να μάθουν μακροσκελείς ακολουθίες. Τα νευρωνικά δίκτυα τέτοιου τύπου αναπτύχθηκαν το 1997 από τους Hochreiter και Schmidhuber και εξελίχθηκαν από πολλούς



επιστήμονες στην συνέχεια. Στις μέρες μας χρησιμοποιούνται για την επίλυση πολλών προβλημάτων σε όλους τους επιστημονικούς κλάδους. (S. Hochreiter 1997)

Τα μοντέλα των LSTM δικτύων είναι σχεδιασμένα με τέτοιο τρόπο ώστε να μπορούν να αποθηκεύουν πληροφορία που βρίσκεται σε πολύ μακρινές χρονικές αποστάσεις. Και αυτή είναι η πιο ουσιαστική διαφορά που έχουν με τα υπόλοιπα RNN δίκτυα. Τα δίκτυα αυτά έχουν εισάγει μία ένωση που επεκτείνεται αλυσιδωτά σε όλη τη δομή του νευρωνικού δικτύου χωρίς να επηρεάζει τα χαρακτηριστικά σε κάθε κόμβο. Αυτή η αλυσιδωτή ένωση ονομάζεται κατάσταση κελιού (Cell state). Επίσης διαθέτουν κελιά μνήμης (cell) πύλες εισόδου, πύλες εξόδου και πύλες ανατροφοδότησης μνήμης. Τα κελιά μνήμης είναι υπεύθυνα να επιλέγουν τις πληροφορίες που θα αποθηκεύσουν συνδυάζοντας τη προηγούμενη κατάσταση, την μνήμη και την είσοδο. (S. Hochreiter, 1997)



Εικόνα 31 Δίκτυο LSTM

- **Κατάσταση κελιού (Cell state)**

Στο σημείο αυτό ανανεώνεται η κατάσταση της προηγούμενης μνήμης  $c_{t-1}$  στην καινούρια  $c_t$ . Πολλαπλασιάζεται η πύλη ανατροφοδότησης μνήμης με τις τιμές της προηγούμενης κατάστασης  $c_{t-1}$ . Στην συνέχεια προστίθεται ο όρος  $i_t \times \vec{c}_t$ , όπου συμβολίζει νέες υποψήφιας τιμές που πρόκειται να ενημερώσουν την τρέχουσα κατάσταση.

$$c_t = f_t \times c_{t-1} + i_t \times \vec{c}_t$$

- **Πύλη ανατροφοδότησης μνήμης (Forget gate)**

Είναι η πύλη που καθορίζει για την μεταφορά των πληροφοριών από την προηγούμενη κατάσταση στην επόμενη καθώς επίσης και για την απομάκρυνση των περισσευόμενων πληροφοριών. Δέχεται ως είσοδο την  $x_t$  την έξοδο του προηγούμενου χρονικού βήματος  $h_{t-1}$  και παράγει στην έξοδο έναν αριθμό που ανήκει στο εύρος  $[0,1]$ . Στην συνέχεια η έξοδος αυτή πολλαπλασιάζεται με κάθε αριθμό του διανύσματος  $c_{t-1}$  της προηγούμενης κατάστασης, απομακρύνοντας έτσι κάθε περισσευόμενη πληροφορία.

$$f_t = \sigma (W_f x_t + U_f h_{t-1} + b_f )$$

- **Πύλη εισόδου (Input Gate)**

Είναι η πύλη που αποφασίζει ποια νέα πληροφορία θα αποθηκευτεί στη μνήμη. Η πύλη εισόδου δέχεται ως είσοδο την  $x_t$  και την έξοδο του προηγούμενου χρονικού βήματος  $h_{t-1}$ . Στην συνέχεια η  $x_t$  και η  $h_{t-1}$  περνάνε σε ένα μόνο-επίπεδο νευρωνικό δίκτυο, και συνάρτηση ενεργοποίησης την υπερβολική εφαπτομένη παράγει τις νέες τιμές  $c_t$ , που θα προστεθούν και θα αποθηκευθούν στη μνήμη.

$$i_t = \sigma (W_i x_t + U_i h_{t-1} + b_i )$$

$$c_t = \tanh (W_c x_t + U_c h_{t-1} + b_c )$$

- **Πύλη εξόδου (Output Gate)**

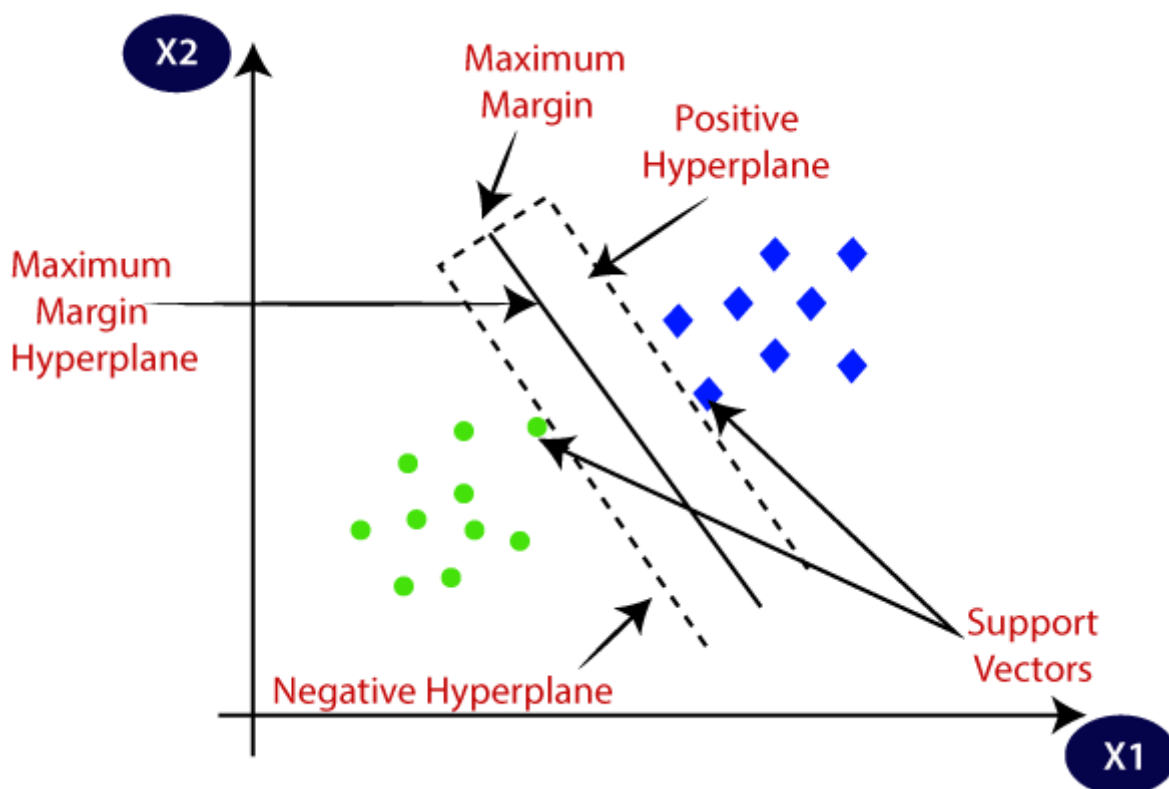
Η έξοδος βασίζεται κυρίως στην κατάσταση της μνήμης. Η είσοδος  $x_t$  και η έξοδος του προηγούμενου χρονικού βήματος  $h_{t-1}$  περνάνε σε ένα μόνο-επίπεδο νευρωνικό δίκτυο με συνάρτηση ενεργοποίησης τη σιγμοειδή, με σκοπό να αποφασιστεί ποια μέρη της μνήμης θα συμμετέχουν στην τελική έξοδο. Αυτό γίνεται με την βοήθεια της συνάρτησης  $\tanh$ , η οποία συμπίεζει την μνήμη. Η μνήμη πολλαπλασιάζεται με την πύλη εξόδου και προκύπτει η τελική έξοδος.

$$o_t = \sigma (W_o x_t + U_o h_{t-1} + b_o )$$

$$h_t = o_t \times \tanh c_t$$

#### 4.4. Μηχανές Διανυσματικής Υποστήριξης (Support Vector Machines)

Οι μηχανές διανυσματικής υποστήριξης (SVM) αποτελούν ένα είδος επιβλεπόμενης μηχανικής μάθησης. Περιγράφουν με γραφικές απεικονίσεις διάφορα στοιχεία τα οποία διαχωρίζονται ανάλογα με την κλάση στην οποία ανήκουν και παρέχουν μεγάλη ακρίβεια στην πρόγνωση. Στηρίζονται στην ιδέα του Vapnik με σκοπό την επίλυση προβλημάτων κατάταξης προτύπων όπως αναγνώριση προσώπου και κατάταξη κειμένου. Βέβαια, με την πάροδο του χρόνου, τα μοντέλα SVM χρησιμοποιήθηκαν και σε άλλα επιστημονικά πεδία όπως στην παλινδρόμηση και στις προβλέψεις χρονοσειρών. Οι SVM προσαρμόζονται ανάλογα με την περίπτωση και τη φύση των δεδομένων. (C. Cortes and V. Vapnik, 1995)



Εικόνα 32 Διάγραμμα SVM

Στόχος του αλγόριθμου SVM είναι να δημιουργήσει την βέλτιστη απόσταση που μπορεί να διαχωρίσει το χώρο  $n$ -διαστάσεων σε κλάσεις, ώστε να γίνεται πιο εύκολη η ταξινόμηση των στοιχείων στη σωστή κατηγορία στο μέλλον. Αυτή η βέλτιστη απόσταση ή

αλλιώς όριο απόφασης ονομάζεται υπερεπίπεδο μέγιστου περιθωρίου (maximum-margin hyperplane). Τα πιο κοντινά στοιχεία των  $n$ -κλάσεων που εξετάζονται ονομάζονται διανύσματα υποστήριξης (support vectors). Έστω λοιπόν ότι για ένα σύνολο σημείων  $(x_i, y_i)$  όπου  $(x_i) \in R^n$  είναι το σύνολο χαρακτηριστικών που αντιστοιχεί στο στοιχείο  $i$ , και  $y_i \in (-1,1)$  ένα σημείο της κλάσης, τότε ο αλγόριθμος SVM θα δημιουργήσει έναν υπερεπίπεδο που θα διαχωρίζει κατά τον βέλτιστο δυνατό τρόπο τις κλάσεις στο  $(-1, 1)$ . Η μαθηματική συνάρτηση που εκφράζει το υπερεπίπεδο είναι η εξής: (Mohammed, J. Z. & Wagner, M. Jr., 2017)

$$w * x_i - b = 0$$

Όπου  $w$  είναι τα διανύσματα βαρών του υπερεπιπέδου. Τα υπερεπίπεδα που δημιουργούνται μεταξύ των κλάσεων και διαχωρίζουν τα στοιχεία ορίζονται ως εξής: (Κ. Διαμαντάρας – Δ. Μπότσης 2019)

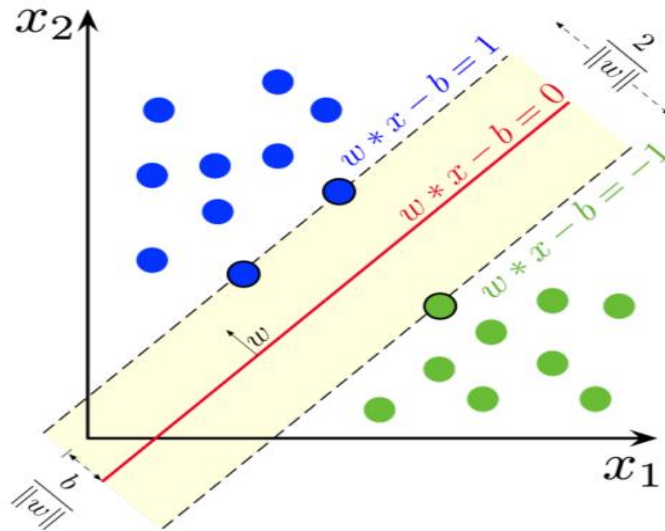
$$w * x_i - b \geq -1 \text{ για } y_i = 1$$

$$w * x_i - b \leq -1 \text{ για } y_i = -1$$

Έχοντας ως δεδομένα τα διανύσματα βαρών  $w$ , η πόλωση  $b$  επιλέγεται με τέτοιο τρόπο ώστε η διαχωριστική επιφάνεια να βρίσκεται στη μέση μεταξύ των κοντινότερων σημείων των κλάσεων. Η απόσταση των δύο παραπάνω υπερεπιπέδων ονομάζεται περιθώριο (margin) και ισούται με  $\frac{2}{\|w\|}$ . Οπότε άλλο ένα πρόβλημα που προκύπτει είναι αυτό της μεγιστοποίησης της απόστασης μεταξύ των δύο επιπέδων, τέτοια ώστε να ισχύει:

$$y_i(w * x_i - b) \geq -1$$

Η λύση στο συγκεκριμένο πρόβλημα αλλά και να ισχύει η παραπάνω συνθήκη, είναι η ελαχιστοποίηση της Ευκλείδειας νόρμας του διανύσματος των βαρών  $w$ .



Εικόνα 33 Διάγραμμα επιπέδων SVM

Τα μοντέλα SVM μπορούν να χρησιμοποιηθούν σε:

- γραμμικά μοντέλα κατηγοριοποίησης, όπου τα δεδομένα κατηγοριοποιούνται μέσω ενός υπερεπιπέδου.
- μη γραμμικά μοντέλα κατηγοριοποίησης, όπου η λειτουργία τους βασίζεται καθαρά στον προσδιορισμό του μη γραμμικού υπερεπιπέδου.

#### 4.5. Αξιολόγηση Μοντέλων Πρόβλεψης

Τα σφάλματα (residuals) παίζουν καθοριστικό ρόλο στην μελέτη και ανάλυση χρονοσειρών. Ως σφάλμα ορίζεται η διαφορά που προκύπτει ανάμεσα στην πραγματική τιμή μιας μεταβλητής και την πρόβλεψη της μεταβλητής και εκφράζεται από την παρακάτω συνάρτηση:

$$e_t = X_t - \hat{X}_t$$

όπου,

- $e_t$ , το σφάλμα που προκύπτει από τη διαφορά πραγματικής τιμής και πρόβλεψης τη χρονική στιγμή  $t$
- $X_t$ , η πραγματική τιμή
- $\hat{X}_t$ , η τιμή πρόβλεψης

Για τη σύγκριση ανάμεσα στην πραγματική και την προβλεπόμενη τιμή καθώς επίσης και για την αξιολόγηση της προβλεπτικής ικανότητας του μοντέλου πρόβλεψης που χρησιμοποιείται ανάλογα την περίπτωση, έχουν αναπτυχθεί διάφορα μαθηματικά κριτήρια. Τα πιο σπουδαία αναλύονται παρακάτω.

#### **4.5.1. Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Square Error, RMSE)**

Είναι από τις πιο γνωστές μεθόδους αξιολόγησης της ακρίβειας των μοντέλων πρόβλεψης και χρησιμοποιείται στον τομέα της μηχανικής μάθησης και μαθηματικά εκφράζεται ως εξής:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2}$$

Όπου,

- $x_t$ , η πραγματική τιμή
- $\hat{x}_t$ , η προβλεπόμενη τιμή
- $n$ , ο αριθμός καταχωρήσεων

#### **4.5.2 Μέσο Απόλυτο Σφάλμα (Mean Absolute Error, MAE)**

Η συγκεκριμένη μέθοδος αξιολόγησης χρησιμοποιείται για να υπολογίζει το μέσο σφάλμα και εκφράζεται από την παρακάτω σχέση:

$$MAE = \frac{1}{n} \sum_{t=1}^n |x_t - \hat{x}_t|$$

#### **4.5.3 Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error, MAPE)**

Η μέθοδος αυτή χρησιμοποιείται για να εκφράσει την μέθοδο MAE σε ποσοστό και εκφράζεται από την παρακάτω σχέση:

$$MAPE = \left( \frac{1}{n} \sum_{t=1}^n \frac{|x_t - \hat{x}_t|}{|x_t|} \right) 100\%$$

#### 4.5.4. Μέσο Τετραγωνικό Σφάλμα (Mean Square Error, MSE)

Το μέσο τετραγωνικό σφάλμα είναι η μέθοδος που υπολογίζει το τετράγωνο της διαφοράς της πραγματικής και της προβλεπόμενης τιμής. Είναι παρόμοια με την MAE αλλά η διαφορά είναι ότι υπολογίζεται πιο εύκολα η κλίση της από αυτή της MAE. Μαθηματικά εκφράζεται από την παρακάτω εξίσωση:

$$MSE = \frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_t)^2$$

## Κεφάλαιο 5<sup>ο</sup>: Ερευνητικό Μέρος

### 5.1. Βασική Ιδέα

Στο κεφάλαιο αυτό θα αναλυθεί ο τρόπος με τον οποίο τα μοντέλα πρόβλεψης λειτουργούν και διαχειρίζονται τα δεδομένα καθώς επίσης και τα αποτελέσματα που προκύπτουν από κάθε περίπτωση ξεχωριστά. Πρόκειται για ανάλυση χρονοσειρών που οι παρατηρήσεις τους μεταβάλλονται στο χρόνο και αντιπροσωπεύουν την εξέλιξη της τιμής του Bitcoin, με τη χρήση στατιστικών μοντέλων και τεχνητών νευρωνικών δικτύων.

Το ερευνητικό κομμάτι της διπλωματικής εργασίας χωρίζεται σε τρία μέρη. Το πρώτο μέρος απευθύνεται στην παρουσίαση του συνόλου δεδομένων (dataset) που θα χρησιμοποιηθεί για την ανάλυση, την επιλογή των χρονοσειρών και τον έλεγχο των χαρακτηριστικών της. Το δεύτερο μέρος αντιπροσωπεύει την επιλογή των μοντέλων που θα χρησιμοποιηθούν, την κατασκευή τους, τα αποτελέσματα που θα προκύψουν και την αξιολόγηση τους ως προς την ακριβή πρόβλεψη και την απόδοσή τους. Στο τρίτο μέρος θα γίνει ανάλυση των συμπερασμάτων που προκύπτουν από τα αποτελέσματα που θα προκύψουν.

### 5.2. Δεδομένα

Σύμφωνα με το σκοπό της εργασίας θέλουμε να προσεγγίσουμε το πρόβλημα της πρόβλεψης της τιμής του κρυπτονομίσματος Bitcoin (BTC). Το σύνολο δεδομένων (dataset) που έχουμε στη διάθεση μας για τη συγκεκριμένη ανάλυση είναι σε μορφή csv. Η συχνότητα καταγραφών των παρατηρήσεων είναι ημερήσια και αφορά ιστορικά δεδομένα για τη χρονική περίοδο 17/09/2014 έως 05/05/2022 και συλλέχθηκαν από την ιστοσελίδα <https://www.kaggle.com/>, η οποία που παρέχει πολλά datasets για όλους τομείς για σκοπούς ανάλυσης ιστορικών δεδομένων. Πρόκειται για 2788 παρατηρήσεις, οι οποίες περιλαμβάνουν τα παρακάτω χαρακτηριστικά:

- Date, η ημερομηνία κάθε παρατήρησης
- Open, η τιμή ανοίγματος
- High, η πιο υψηλή τιμή του Bitcoin μέσα στη μέρα
- Low, η πιο χαμηλή τιμή του Bitcoin μέσα στη μέρα
- Close, η τιμή κλεισίματος



- Adj Close, η προσαρμοσμένη τιμή κλεισίματος
- Volume, ο όγκος των συναλλαγών μέσα στην μέρα

### 5.3. Πλαίσιο Ανάπτυξης

Για την προεπεξεργασία των δεδομένων, την ανάλυση, την μοντελοποίηση και την πρόβλεψη της τιμής του Bitcoin, χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python (έκδοση 3.9), η οποία θεωρείται από τις πιο σημαντικές γλώσσες προγραμματισμού για σκοπούς ανάλυσης, πρόγνωσης και ανάπτυξης νευρωνικών δικτύων, λόγω των εξελιγμένων εργαλείων και βιβλιοθηκών που προσφέρει για επεξεργασία και διαχείριση μεγάλου όγκου δεδομένων.

Η ανάπτυξη όλου του κώδικα έγινε στο περιβάλλον IDE της πλατφόρμας Anaconda και συγκεκριμένα στο Jupyter Notebook, το οποίο καλύπτει την ανάγκη της συγκεκριμένης διπλωματικής εργασίας για απεικόνιση πινάκων και γραφικών παραστάσεων και γενικότερα ενδείκνυται για την ανάλυση δεδομένων. προγραμματισμού για σκοπούς ανάλυσης, πρόγνωσης και ανάπτυξης νευρωνικών δικτύων, λόγω των εξελιγμένων εργαλείων που προσφέρει για επεξεργασία και διαχείριση μεγάλου όγκου δεδομένων.

### 5.4. Μεθοδολογία

Κύριος στόχος είναι η κατανόηση των δεδομένων. Αυτό επιτυγχάνεται με την περιγραφική ανάλυση της τιμής του Bitcoin. Με την βοήθεια της Python της βιβλιοθήκης matplotlib.pyplot παρουσιάζονται τα δεδομένα της χρονοσειράς σε γραφικές παραστάσεις καθώς και τα στατιστικά στοιχεία της χρονοσειράς όπως η μέση τιμή της και η διασπορά. Μελετώντας αυτά τα στοιχεία μπορούμε να διαπιστώσουμε σημαντικές πληροφορίες για την μορφή και τη δομή της χρονοσειράς. Προχωράμε με τον έλεγχο της χρονοσειράς για στασιμότητα ή μη-στασιμότητα με τον έλεγχο υπόθεσης Augmented Dickey Fuller. Εξετάζεται η αυτοσυσχέτιση μεταξύ των δεδομένων με την απεικόνιση των γραφικών παραστάσεων των συντελεστών αυτοσυσχέτισης και μερικής αυτοσυσχέτισης. Σε περίπτωση μη στασιμότητας της χρονοσειράς, χρησιμοποιούνται διάφοροι τρόποι μετασχηματισμού διαφοροποίησης της χρονοσειράς σε στάσιμη.

Το επόμενο βήμα περιλαμβάνει την εφαρμογή των κατάλληλων μοντέλων πρόγνωσης, σύμφωνα με τα οποία μοντελοποιούνται τα δεδομένα. Το πρώτο μοντέλο που εφαρμόζεται είναι το στατιστικό μοντέλο ARIMA, εφόσον οριστεί ο πιο κατάλληλος συνδυασμός των τιμών  $p$ ,  $d$ ,  $q$  των του μοντέλου. Στην συνέχεια παρουσιάζονται οι προβλέψεις και τα αποτελέσματα του μοντέλου.

Το δεύτερο μοντέλο που εφαρμόζεται είναι το τεχνητό νευρωνικό δίκτυο LSTM, όπου τα δεδομένα τροποποιούνται με τις κατάλληλες εντολές σε τέτοια μορφή ώστε να ακολουθούν την κανονική κατανομή. Το μοντέλο θα κατασκευαστεί με την βοήθεια της βιβλιοθήκων Keras και TensorFlow που είναι κατάλληλες για τη χρήση νευρωνικών δικτύων. Το τρίτο και τελευταίο μοντέλο που εφαρμόζεται είναι οι μηχανές διανυσμάτων υποστήριξης (Support Vector Machine) SVM.

Ο δείκτης που χρησιμοποιείται για την αξιολόγηση και την απόδοση των τριών μοντέλων είναι η ρίζα τετραγωνικού μέσου τετραγωνικού σφάλματος RMSE, όπου συνήθως το μοντέλο που παρουσιάζει το μικρότερο δείκτη RMSE είναι το πιο αξιόλογο τόσο στην πρόγνωση όσο και στην ακρίβεια. Το τελευταίο βήμα είναι ο σχολιασμός των αποτελεσμάτων που παράγονται καθώς επίσης και τα συμπεράσματα που προκύπτουν.

## 5.5. Αρχική επεξεργασία των δεδομένων

Ξεκινώντας την ανάλυση, το πρώτο βήμα που γίνεται είναι η φόρτωση των δεδομένων στο προγραμματιστικό περιβάλλον, εφόσον φορτωθούν και οι κατάλληλες βιβλιοθήκες. Με την βοήθεια της βιβλιοθήκης Pandas φορτώθηκε το αρχείο BTC-USD.csv στο περιβάλλον όπως φαίνεται στην εικόνα 34.

```
In [2]: #Load Data
df = pd.read_csv(r'C:\Users\rkout\Desktop\BTC-USD.csv', usecols =['Date', 'Open', 'High', 'Low', 'Close', 'Adj Close'])
df=df.dropna()
print('Shape of data',df.shape)
df.head()
print(df)

Shape of data (2788, 6)
```

Εικόνα 34 Φόρτωση Δεδομένων στο προγραμματιστικό περιβάλλον

Στην συνέχεια έγινε έλεγχος για τυχόν ακραίες, λανθασμένες ή εκλιπόντες τιμές και διαπιστώθηκε ότι δεν υπάρχουν τέτοιες τιμές.

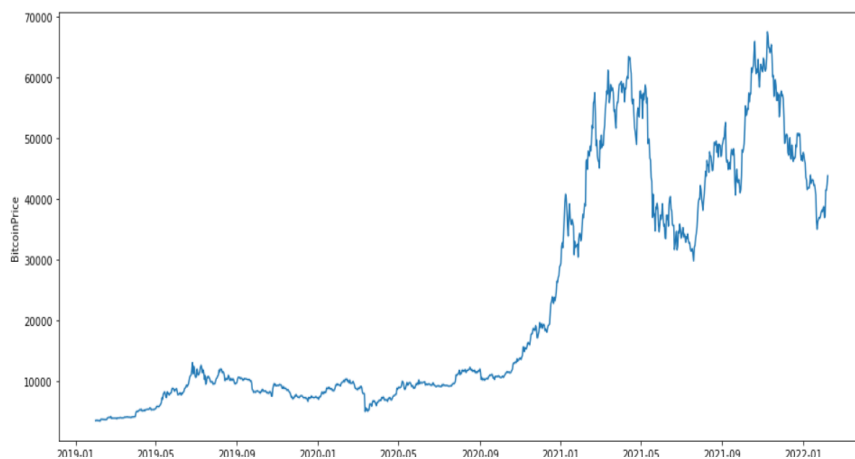
```
In [4]: Data = df.dropna()
Data.index = pd.to_datetime(Data.Date)
Data = Data['2019-01-31':'2022-02-07']['Adj Close']
data=Data.copy()
Data=Data[:-7]
print(Data)
print('Null Values:', df.isnull().values.sum())
print('NA Values:', df.isnull().values.any())
```

```
Date
2019-01-31    3457.792725
2019-02-01    3487.945313
2019-02-02    3521.060791
2019-02-03    3464.013428
2019-02-04    3459.154053
...
2022-01-27    37138.234375
2022-01-28    37784.332031
2022-01-29    38138.179688
2022-01-30    37917.601563
2022-01-31    38483.125000
Name: Adj Close, Length: 1097, dtype: float64
Null Values: 0
NA Values: False
```

Εικόνα 35 Έλεγχος δεδομένων για ακραίες, λανθασμένες ή εκλιπόντες τιμές

Οι μόνες στήλες από το σύνολο δεδομένων που ενδιαφέρουν την ερευνητική διαδικασία είναι αυτές με τα ονόματα 'Date' και 'Adj Close', όπου αντιπροσωπεύουν την ημερομηνία καταχώρησης και την προσαρμοσμένη τιμή κλεισίματος αντίστοιχα. Οι υπόλοιπες στήλες διαγράφηκαν από τον πίνακα. Για ερευνητικούς κυρίως λόγους και λόγω των απρόβλεπτα μεγάλων διακυμάνσεων στην τιμή του Bitcoin από το 2019 και έπειτα, όπου η τιμή του Bitcoin αυξήθηκε κατά πολύ μεγάλα ποσοστά, το σύνολο των δεδομένων που χρησιμοποιείται για την ανάλυση είναι από 31-01-2019 έως 07/02/2022. Αυτό έγινε για να ελαχιστοποιήσουμε τον RMSE, διότι με τη χρήση ολόκληρου του dataset, ο δείκτης RMSE ήταν εξαιρετικά μεγάλος.

Όπως αναφέρθηκε, στην καλύτερη κατανόηση των δεδομένων, σημαντικό ρόλο παίζει και η γραφική τους απεικόνιση. Από το παρακάτω γράφημα φαίνεται η έντονη διακύμανση της τιμής του Bitcoin για το συγκεκριμένο χρονικό διάστημα.

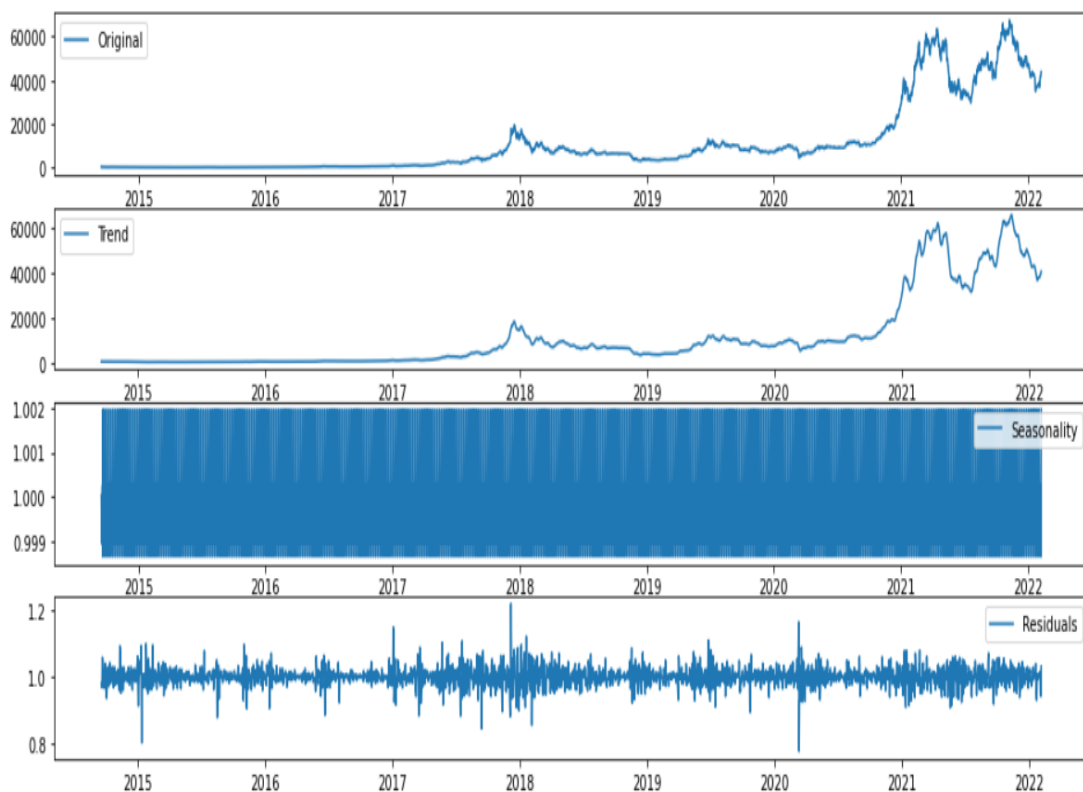


Εικόνα 36 Οπτικοποίηση Δεδομένων

Οι τιμές αντιπροσωπεύουν ημερήσιες καταχωρήσεις. Στο γράφημα της εικόνας 36 φαίνεται ότι η τιμή Bitcoin χαρακτηρίζεται από έντονες διακυμάνσεις οι οποίες ακολουθούν μία απρόβλεπτη πορεία.

### 5.5.1. Έλεγχος στασιμότητας χρονοσειράς

Το επόμενο βήμα που πραγματοποιείται στην διερευνητική διαδικασία είναι ο έλεγχος στασιμότητας της χρονοσειράς. Για να έχουμε όσο το δυνατόν πιο ακριβή αποτελέσματα θα πρέπει η χρονοσειρά που χρησιμοποιείται να είναι στάσιμη. Αυτό επιτυγχάνεται με την διαδικασία της αποσύνθεσης στοιχείων της χρονοσειράς. Στο γράφημα της εικόνας 37 απεικονίζονται ξεχωριστά τα στοιχεία που συνθέτουν την χρονοσειρά.



Εικόνα 37 Αποσύνθεση χρονοσειράς

Από τις γραφικές αυτές παραστάσεις, παρατηρεί κανείς την έντονη ανοδική τάση που παρουσιάζει η χρονοσειρά, η οποία ωστόσο μακροπρόθεσμα μπορεί γίνει καθοδική, διότι η πορεία της τιμής των κρυπτονομισμάτων είναι πολύ απρόβλεπτη. Επίσης παρατηρείται έντονη και ξεκάθαρη εποχικότητα η οποία εμφανίζεται και στη γραφική παράσταση των residuals,

τα οποία είναι πολύ σημαντικά για την ανάλυση της ολικής εικόνας της χρονοσειράς. Για να εξεταστεί η στασιμότητα χρονοσειράς, πραγματοποιήθηκε αυξημένος έλεγχος Dickey-Fuller (ADF).

```
In [6]: #ADF test for checking stationarity
from statsmodels.tsa.stattools import adfuller
adftest = adfuller(Data)
print ('pvalue of adfuller test is: ', adftest[1]) # p-value = 0.74 > 0.05 The timeseries is non-stationary.

pvalue of adfuller test is: 0.760933298413905
```

Εικόνα 38 Έλεγχος ADF

Από τα αποτελέσματα του ελέγχου φαίνεται ότι η τιμή p-value είναι 0,76, η οποία δηλώνει ότι η χρονοσειρά δεν είναι στάσιμη. Για να είναι στάσιμη θα πρέπει η τιμή p-value να είναι μικρότερη από 0,05. Υπάρχουν περιπτώσεις όμως που ο έλεγχος ADF αφήνει περιθώρια αμφισβήτησης για το αν μια χρονοσειρά παρουσιάζει στασιμότητα ή όχι. Για τον λόγο αυτό πραγματοποιείται ένας ακόμη έλεγχος, ο έλεγχος KPSS. Από τον έλεγχο KPSS προκύπτουν τα παρακάτω αποτελέσματα:

```
import statsmodels.api as sm

#perform KPSS test
sm.tsa.stattools.kpss(Data, regression='ct')

C:\Users\rkout\anaconda1\lib\site-packages\statsmodels\tsa\stattools.py:2018: InterpolationWarning: The test statisti
c is outside of the range of p-values available in the
look-up table. The actual p-value is smaller than the p-value returned.

warnings.warn(

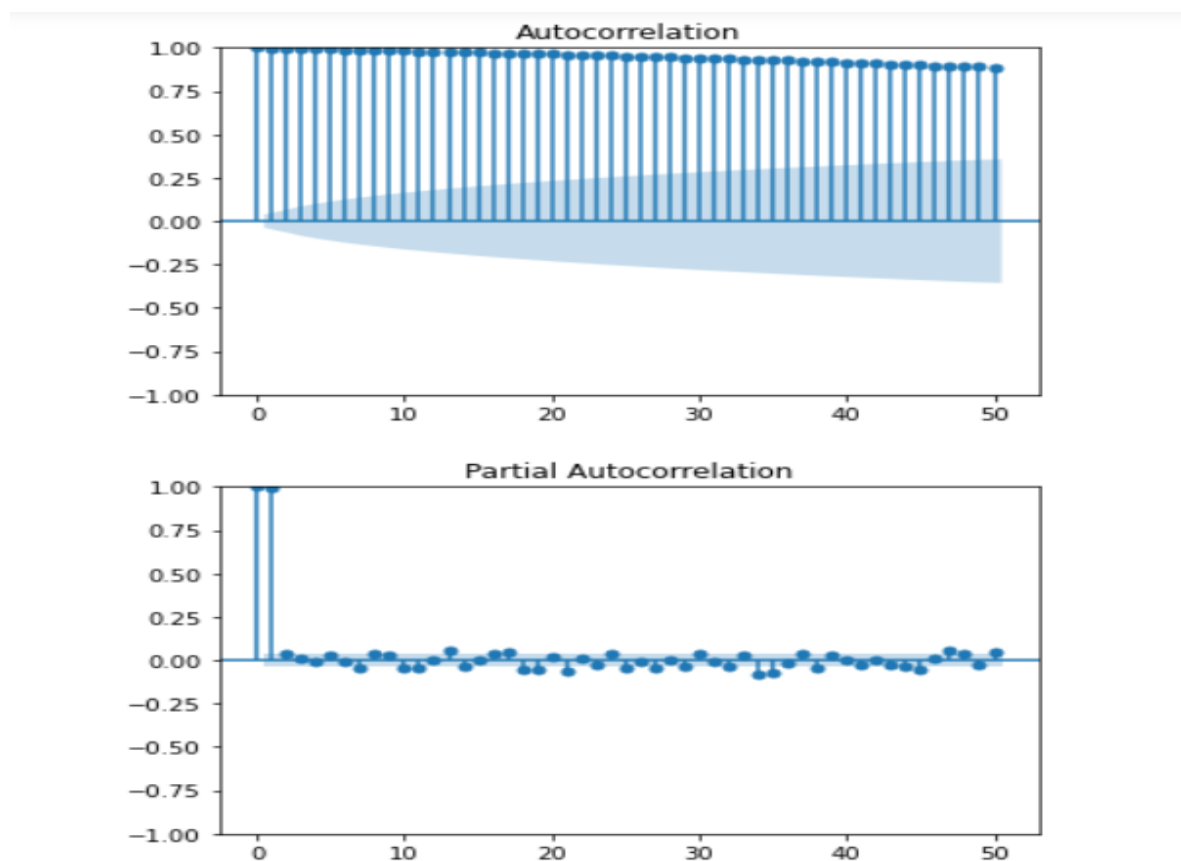
(1.1197330303415853,
 0.01,
 31,
 {'10%': 0.119, '5%': 0.146, '2.5%': 0.176, '1%': 0.216}))
```

Εικόνα 39 Έλεγχος KPSS

Η τιμή p-value είναι 0,01. Η αρχική υπόθεση του ελέγχου KPSS είναι ότι η χρονοσειρά είναι στάσιμη σε αντίθεση με την αρχική υπόθεση του ελέγχου ADF. Παρατηρείται ότι η τιμή του ελέγχου είναι 1,12 που σημαίνει ότι η αρχική υπόθεση KPSS απορρίπτεται και η χρονοσειρά θεωρείται μη-στάσιμη.

### 5.5.2. ACF και PACF

Ένας ακόμη έλεγχος που γίνεται για την τιμή της χρονοσειράς του Bitcoin, είναι ο έλεγχος της αυτοσυσχέτισης (ACF) και της μερικής αυτοσυσχέτισης (PACF). Στο γράφημα της εικόνας 40 παρουσιάζονται οι γραφικές παραστάσεις των συντελεστών αυτοσυσχέτισης και μερικής αυτοσυσχέτισης σε συνάρτηση με τα χρονικά βήματα.



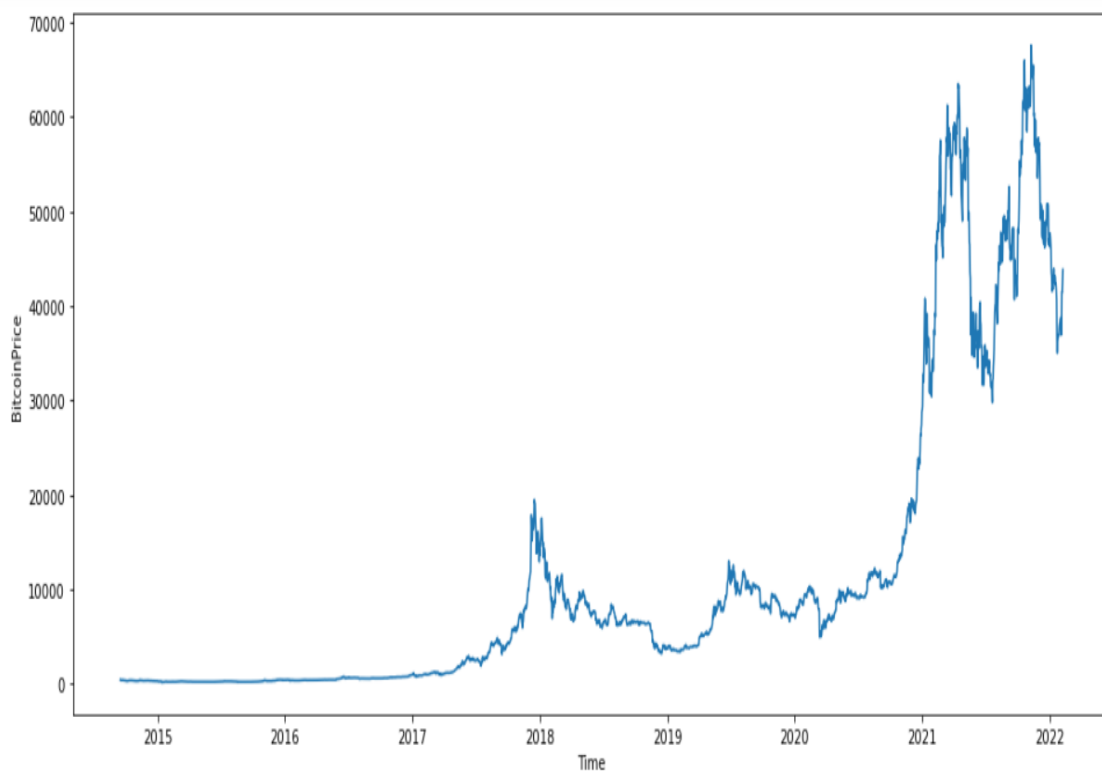
Εικόνα 40 Διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης

Παρατηρείται ότι υπάρχει αυτοσυσχέτιση από τις παραπάνω γραφικές παραστάσεις. Υπάρχει έντονη διαφορά ανάμεσα στους δύο συντελεστές, η οποία δηλώνει ότι πρέπει να γίνει μετασχηματισμός διαφοροποίησης της χρονοσειράς καθώς ο συντελεστής μερικής αυτοσυσχέτισης δείχνει πως οι τιμές των δύο προηγούμενων βημάτων εμφανίζουν πραγματική αυτοσυσχέτιση το τρέχον χρονικό βήμα. Είναι σημαντικό σε αυτό το σημείο να τονιστεί ότι η στασιμότητα σε μια χρονοσειρά αποτελεί πολύ σημαντικό χαρακτηριστικό και παίζει καθοριστικό ρόλο στην διαδικασία της πρόβλεψης. Αν ένα μοντέλο εκπαιδευτεί σε κάποια δεδομένα που έχουν διαφορετικά χαρακτηριστικά, τότε τα αποτελέσματα της πρόβλεψης δεν θα είναι ακριβή, ενώ σε περίπτωση που το μοντέλο εκπαιδευτεί με δεδομένα που

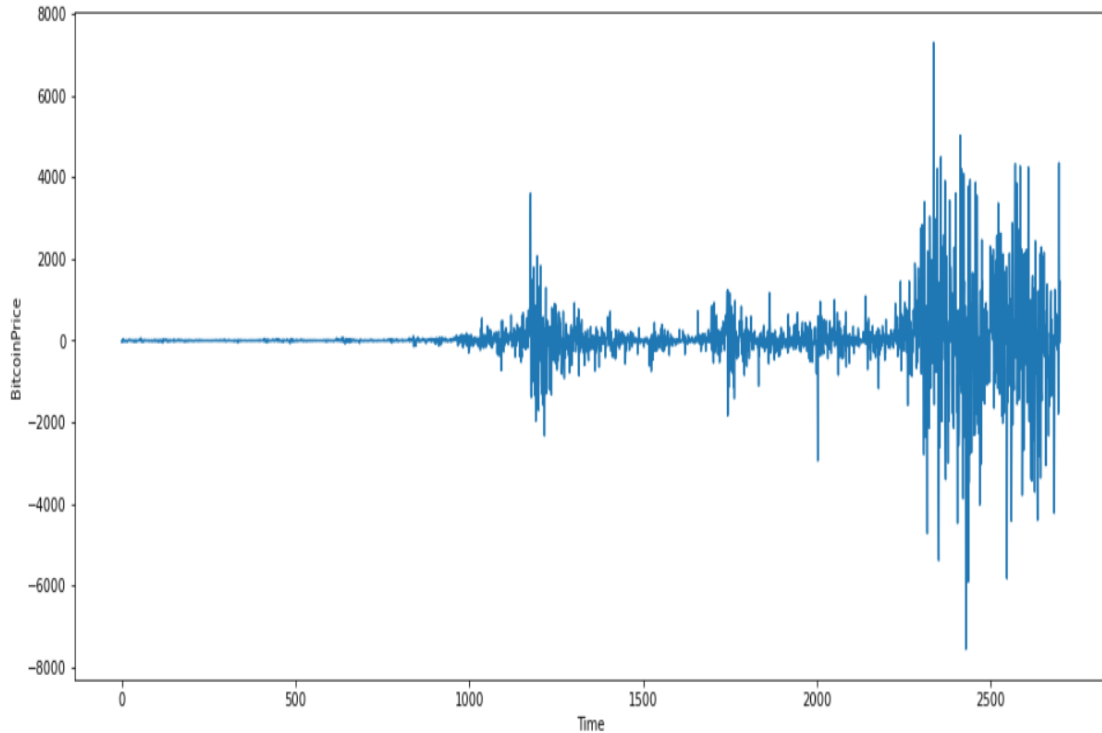
παρουσιάζουν παρόμοια χαρακτηριστικά, τότε τα αποτελέσματα της πρόβλεψης θεωρούνται αξιόπιστα. Για το λόγο αυτό πριν την πρόβλεψη θα πρέπει η χρονοσειρά να είναι στάσιμη.

### 5.5.3. Μετασχηματισμός χρονοσειράς

Στα πλαίσια της παρούσης διπλωματικής εργασίας, ο μετασχηματισμός της χρονοσειράς γίνεται διαφοροποίηση των τιμών, ο οποίος είναι από τους πιο διαδεδομένους μηχανισμούς μετασχηματισμού στην ερευνητική βιβλιογραφία για τη χρονοσειρά του Bitcoin.



Εικόνα 41 Αρχική μορφή δεδομένων



Εικόνα 42 Διαφοροποιημένη μορφή δεδομένων

Στα παραπάνω γραφήματα διακρίνεται η χρονοσειρά στην αρχική της μορφή και η χρονοσειρά μετά τον μετασχηματισμό της σε στάσιμη. Η διαφοροποιημένη χρονοσειρά φαίνεται ότι εμφανίζει μεγάλη σταθερότητα. Τα αποτελέσματα των ελέγχων KPSS και ADF πάνω στην διαφοροποιημένη χρονοσειρά είναι τα εξής:

```
In [45]: #ADF test for checking stationarity
from statsmodels.tsa.stattools import adfuller
adftest = adfuller(diff)
print ('pvalue of adfuller test is: ', adftest[1]) # p-value = 0.74 > 0.05 The timeseries is non-stationary.
plt.figure(figsize = (16,7))
fig = plt.figure(1)
ax1 = fig.add_subplot(111)
ax1.set_xlabel('Time')
ax1.set_ylabel('BitcoinPrice')
ax1.plot(diff)
```

pvalue of adfuller test is: 1.0098121743821113e-13

Εικόνα 43 Έλεγχος ADF

Η τιμή p-value για τον έλεγχο ADF είναι 1,0098121743821113e-13 που σαφώς είναι μικρότερη από την τιμή 0,05 , επομένως η χρονοσειρά είναι στάσιμη.



```

In [46]: import statsmodels.api as sm

#perform KPSS test
sm.tsa.stattools.kpss(diff, regression='ct')

C:\Users\rkout\anaconda1\lib\site-packages\statsmodels\tsa\stattools.py:2022: InterpolationWarning: The test statistic is outside of the range of p-values available in the look-up table. The actual p-value is greater than the p-value returned.

warnings.warn(

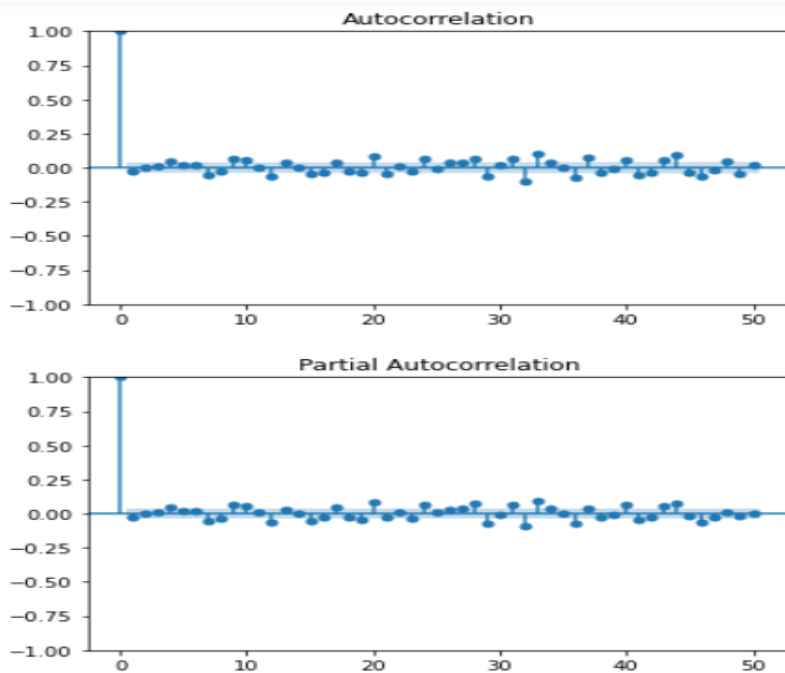
Out[46]: (0.03775923907376182,
0.1,
10,
{'10%': 0.119, '5%': 0.146, '2.5%': 0.176, '1%': 0.216})

```

Εικόνα 44 Έλεγχος KPSS

Ακολούθως από τον έλεγχο KPSS προκύπτει ότι η τιμή p-value είναι 0,1 και η τιμή του ελέγχου είναι 0,03 , που δηλώνει στασιμότητα της χρονοσειράς.

Διασταυρώντας τα αποτελέσματα των ελέγχων ADF και KPSS προκύπτει ότι οι τιμές p-value δείχνουν ότι η διαφοροποίηση τιμών της χρονοσειράς του Bitcoin έχει ως αποτέλεσμα την εμφάνιση στασιμότητας στη χρονοσειρά. Επίσης στο γράφημα της εικόνας 45, παρουσιάζονται οι γραφικές παραστάσεις των συντελεστών συσχέτισης και μερικής αυτοσυσχέτισης για την διαφοροποιημένη χρονοσειρά, όπου διακρίνεται η διαφοροποίηση στα χαρακτηριστικά της χρονοσειράς.



Εικόνα 45 Διαγράμματα αυτοσυσχέτισης και μερικής αυτοσυσχέτισης δεδομένων

## 5.6. Μοντέλο ARIMA και αποτελέσματα

Μέχρι στιγμής είναι γνωστό ότι η χρονοσειρά του Bitcoin δεν παρουσιάζει Στασιμότητα και πως η κατάλληλη μέθοδος για το μετασχηματισμό της σε στάσιμη είναι η διαφοροποίηση των τιμών της. Το επόμενο βήμα που πραγματοποιείται είναι η εφαρμογή του στατιστικού μοντέλου ARIMA για μοντελοποίηση της χρονοσειράς αλλά και πρόγνωση της τιμής του Bitcoin για τις επόμενες 7 ημέρες. Με βάση την ερευνητική βιβλιογραφία, τα τεχνητά νευρωνικά δίκτυα θεωρούνται πιο αξιόπιστα μοντέλα πρόγνωσης και παρουσιάζουν πιο ακριβή αποτελέσματα. Παρόλα αυτά, στην παρούσα διπλωματική εργασία, θεωρήθηκε απαραίτητο να χρησιμοποιηθεί και το στατιστικό μοντέλο ARIMA για να μπορέσουμε να συγκρίνουμε τα αποτελέσματα του με τα αποτελέσματα των άλλων μοντέλων που θα χρησιμοποιηθούν.

Το μοντέλο ARIMA, μετά την προεπεξεργασία των δεδομένων και την διαφοροποίηση της χρονοσειράς σε στάσιμη, δέχεται σας αρχικές παραμέτρους τρεις τιμές τις  $p$ ,  $d$ ,  $q$ , οι οποίες δηλώνουν τους συντελεστές της αυτοπαλινδρόμησης ( $p$ ), το βαθμό της διαφοροποίησης της χρονοσειράς ( $d$ ) και του κινητού μέσου ( $q$ ). Πολλές φορές η επιλογή των παραμέτρων γίνεται με τη χρήση των γραφικών παραστάσεων της αυτοσυσχέτισης και της μερικής

αυτοσυσχέτισης, ωστόσο για την συγκεκριμένη εργασία κρίθηκε απαραίτητο να γίνει χρήση διαφόρων μεθόδων για να βρεθεί ο κατάλληλος και πιο ακριβής συνδυασμός. Για το λόγο αυτό χρησιμοποιήθηκαν τρεις συνδυασμοί ARIMA:

- ARIMA(1,0,1), ο οποίος είναι θεωρητικός μιας και προκύπτει από τις γραφικές παραστάσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης
- ARIMA(0,1,0), ο οποίος είναι πρακτικός και προκύπτει από τον μηχανισμό εύρεσης του κατάλληλου συνδυασμού
- ARIMA (2,0,1) , ο οποίος συναντάται αρκετά στην ερευνητική βιβλιογραφία για χρονοσειρές κρυπτονομισμάτων

Σύμφωνα λοιπόν με τις γραφικές παραστάσεις συσχέτισης και αυτοσυσχέτιση, ο καλύτερος συνδυασμός παραμέτρων (p, d, q) που μπορούν να εφαρμοστούν στο μοντέλο είναι ο ARIMA(1,0,1). Ο συντελεστής d είναι 0 διότι η χρονοσειρά είναι ήδη στάσιμη και δεν χρειάζεται διαφοροποίηση. Στην συνέχεια δοκιμάζονται και άλλοι συνδυασμοί για να το μοντέλο Arima για να διαπιστωθεί ποιος συνδυασμός δίνει τα καλύτερα αποτελέσματα.

Όπως αναφέρθηκε στην αρχή του κεφαλαίου, το κριτήριο αξιολόγησης της ακρίβειας του μοντέλου προσδιορίζεται από την μέτρηση του δείκτη RMSE. Τα αποτελέσματα των δοκιμών παρουσιάζονται στην εικόνα 46. Για τις δοκιμές τα δεδομένα χωρίστηκαν σε δύο σετ:

σετ εκπαίδευσης (train) και περιλαμβάνει το 80% των παρατηρήσεων και σε σετ ελέγχου (test) που περιλαμβάνει το 20% των παρατηρήσεων. Η διαδικασία διήρκησε περίπου 6 ώρες.

```
ARIMA(0, 1, 0) RMSE=1624.784
ARIMA(0, 1, 1) RMSE=1627.523
ARIMA(0, 1, 2) RMSE=1631.693
ARIMA(1, 1, 0) RMSE=1627.741
ARIMA(1, 1, 1) RMSE=1638.718
ARIMA(1, 1, 2) RMSE=1635.967
ARIMA(3, 1, 0) RMSE=1635.401
ARIMA(3, 1, 1) RMSE=1639.232
ARIMA(3, 1, 2) RMSE=1655.972
ARIMA(4, 1, 0) RMSE=1637.897
ARIMA(4, 1, 1) RMSE=1639.913
ARIMA(4, 1, 2) RMSE=22403.345
ARIMA(5, 1, 0) RMSE=1642.431
ARIMA(5, 1, 1) RMSE=1646.814
ARIMA(6, 1, 0) RMSE=1647.401
ARIMA(6, 1, 1) RMSE=1652.535
ARIMA(6, 1, 2) RMSE=1653.688
ARIMA(8, 1, 0) RMSE=1655.333
ARIMA(8, 1, 1) RMSE=1657.614
ARIMA(8, 1, 2) RMSE=1652.767
ARIMA(10, 1, 0) RMSE=1657.922
ARIMA(10, 1, 1) RMSE=1659.920
ARIMA(10, 1, 2) RMSE=1666.865
Best ARIMA(0, 1, 0) RMSE=1624.784
```

Εικόνα 46 Αποτελέσματα ARIMA

Από την αναζήτηση των βέλτιστων συνδυασμό στο περιβάλλον Jupyter-Notebook προκύπτει ότι ο καλύτερος συνδυασμός είναι αυτός που έχει τον χαμηλότερο δείκτη RMSE. Στην συγκεκριμένη περίπτωση ο καλύτερος συνδυασμός είναι ο ARIMA(0,1,0) , ο οποίος παραπέμπει σε λευκό θόρυβο. Σε σχετική βιβλιογραφία, ένας από τους καλύτερους συνδυασμούς ARIMA για την πρόγνωση κρυπτονομισμάτων είναι ο ARIMA(2,0,1).

Στην διαδικασία της πρόβλεψης, τα δεδομένα χωρίζονται πάλι σε 80% σετ εκπαίδευσης και βάσει αυτού οι συνδυασμοί του μοντέλου ARIMA θα κάνουν πρόγνωση για το υπόλοιπο 20% του σετ. Στην συνέχεια οι πραγματικές τιμές του 20% αποκρύπτονται από το μοντέλο και θα χρησιμοποιηθούν για σύγκριση με τις προβλεπόμενες τιμές του μοντέλου. Από το σύνολο των δεδομένων που εξετάζεται, αφαιρούνται οι τελευταίες 7 ημέρες, δηλαδή από 01-02-2022 έως 07-02-2022, για τις οποίες το μοντέλο θα προβλέψει τις τιμές τους. Η διαδικασία της πρόγνωσης θα πραγματοποιηθεί και για τους τρεις συνδυασμούς ARIMA (1,0,1) , ARIMA(0,1,0) και ARIMA(2,0,1).

### 5.6.1 ARIMA (1,0,1)

Για τον συγκεκριμένο συνδυασμό προκύπτει ο παρακάτω πίνακας που εμφανίζει τα σφάλματα πρόγνωσης:

```

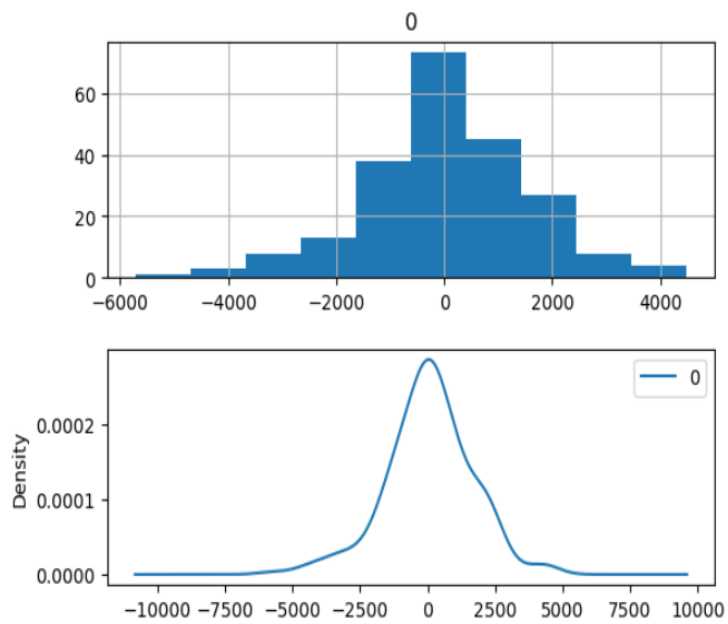
count      220.000000
mean        59.860209
std         1605.515263
min         -5721.712909
25%         -860.484551
50%          61.794776
75%          997.264254
max          4494.251099

```

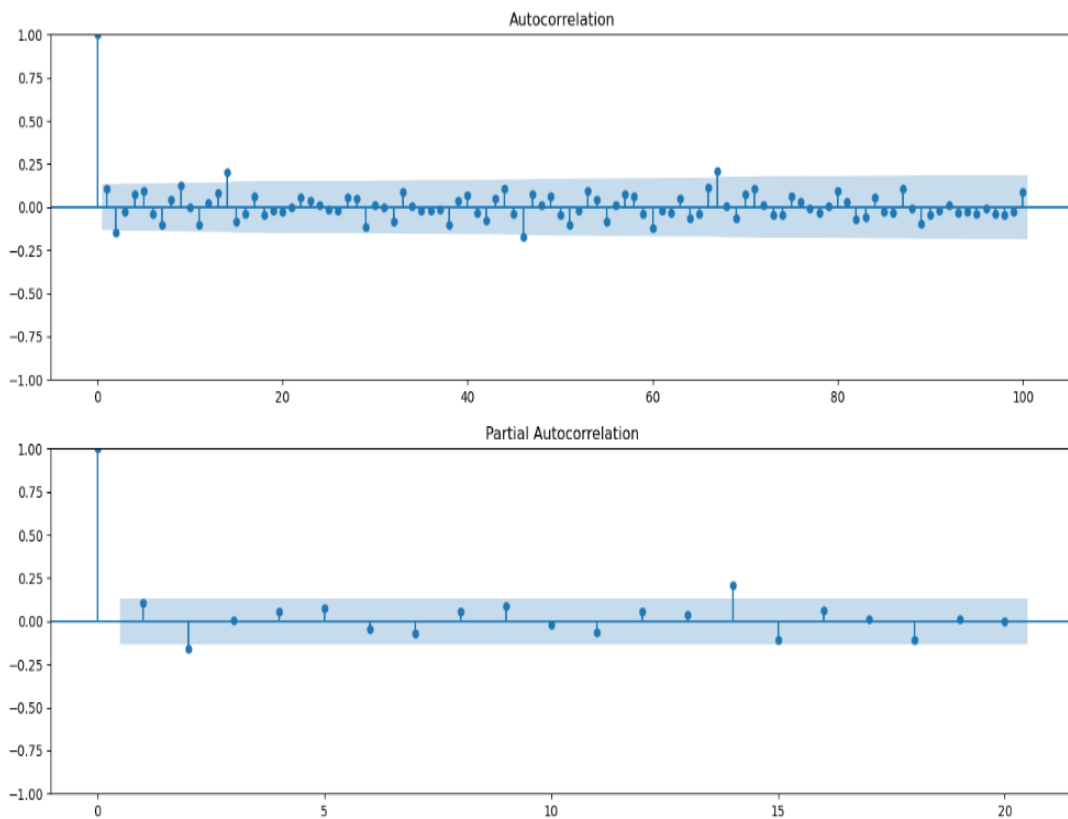
Εικόνα 47 Σφάλματα Πρόγνωσης

Ο πίνακας αυτός εμφανίζει τα αποτελέσματα της αφαίρεσης των πραγματικών τιμών από τις τιμές που προβλέφθηκαν, οι οποίες θα έπρεπε να εμφανίζουν μέση τιμή 0. Η μέση τιμή τους φαίνεται ότι είναι 59.86 θεωρείται τιμή bias για τη χρονοσειρά , η οποία προστίθεται στις προβλέψεις για να εξαλειφθεί από τα σφάλματα. Στις εικόνες 48 και 49 φαίνονται οι

κατανομές των σφαλμάτων καθώς επίσης οι γραφικές παραστάσεις των συντελεστών συσχέτισης και αυτοσυσχέτισης.

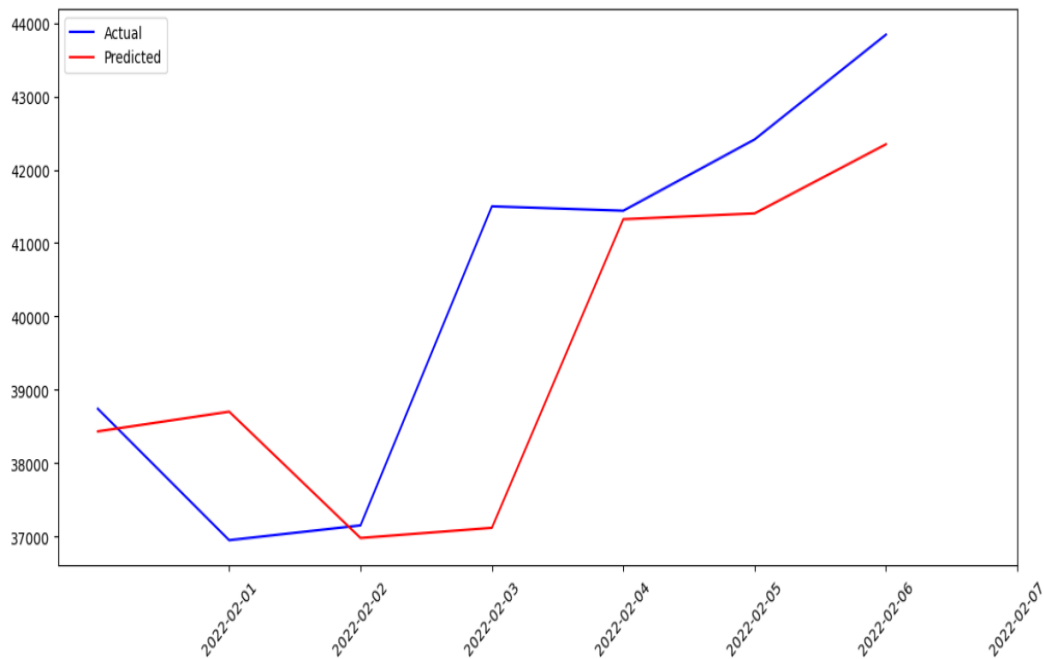


Εικόνα 48 Κατανομή σφαλμάτων



Εικόνα 49 Γραφικές παραστάσεις ACF και PACF

Μετά την διόρθωση με την προσθήκη της bias, επαναλαμβάνεται η διαδικασία για το μοντέλο ARIMA(1,0,1), το οποίο θα χρησιμοποιηθεί για την πρόγνωση των 7 ημερών που έχουν αφαιρεθεί από την χρονοσειρά. Τα αποτελέσματα της πρόβλεψης φαίνονται στην εικόνα 50 όπου φαίνεται να είναι ικανοποιητικά. Ακολουθούν περίπου την ίδια πορεία με τις πραγματικές τιμές, με μια εξαίρεση μεταξύ δεύτερης και τρίτης ημέρας όπου φαίνεται να έχει μεγάλη απόκλιση σε σχέση με την πραγματική τιμή.



Εικόνα 50 Διάγραμμα πραγματικών και προβλεπόμενων τιμών Bitcoin

Οι πραγματικές τιμές και οι προβλεπόμενες παρουσιάζονται στον παρακάτω πίνακα στην εικόνα 51, αριστερά οι προβλεπόμενες και δεξιά οι πραγματικές.

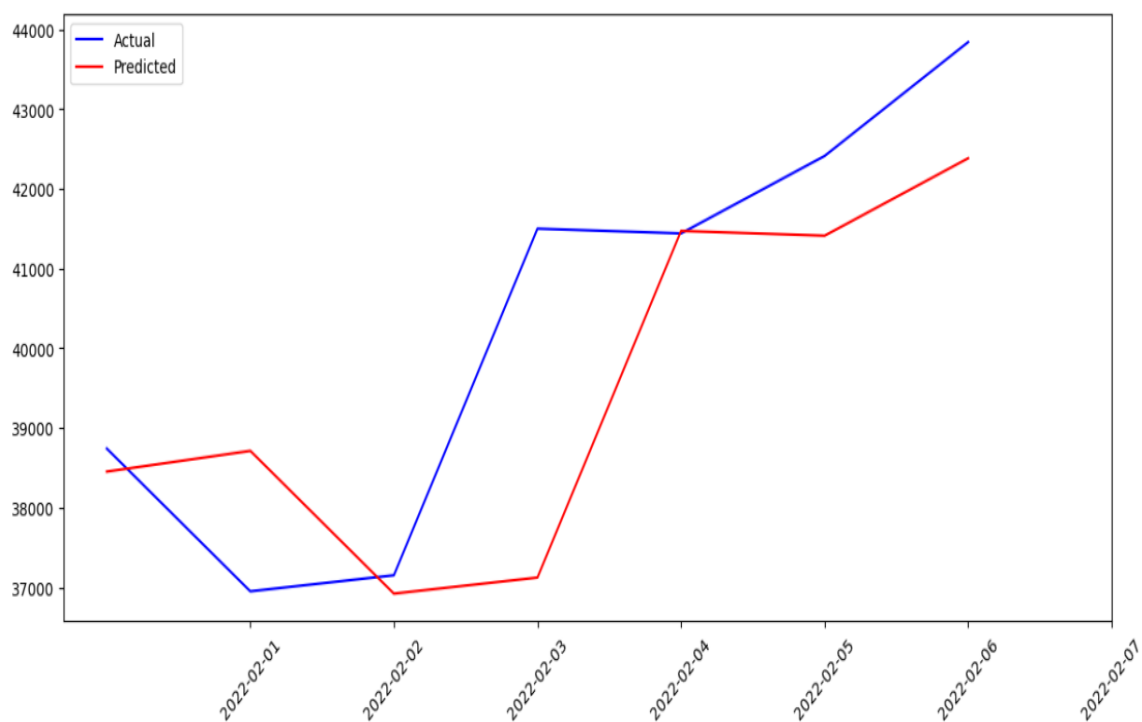
```
>Predicted=38434.999, Expected=38743
>Predicted=38703.917, Expected=36953
>Predicted=36982.889, Expected=37155
>Predicted=37121.735, Expected=41501
>Predicted=41325.122, Expected=41441
>Predicted=41404.697, Expected=42412
>Predicted=42346.375, Expected=43840
RMSE: 1913.408
```

Εικόνα 51 Πραγματικές και προβλεπόμενες τιμές

Στην συνέχεια της διπλωματικής εργασίας θα επαναλάβουμε την ίδια διαδικασία για τα μοντέλα ARIMA(0,1,0) και ARIMA(2,0,1).

### 5.6.2. ARIMA (0,1,0)

Για το μοντέλο ARIMA(0,1,0) έχουμε τα εξής αποτελέσματα:



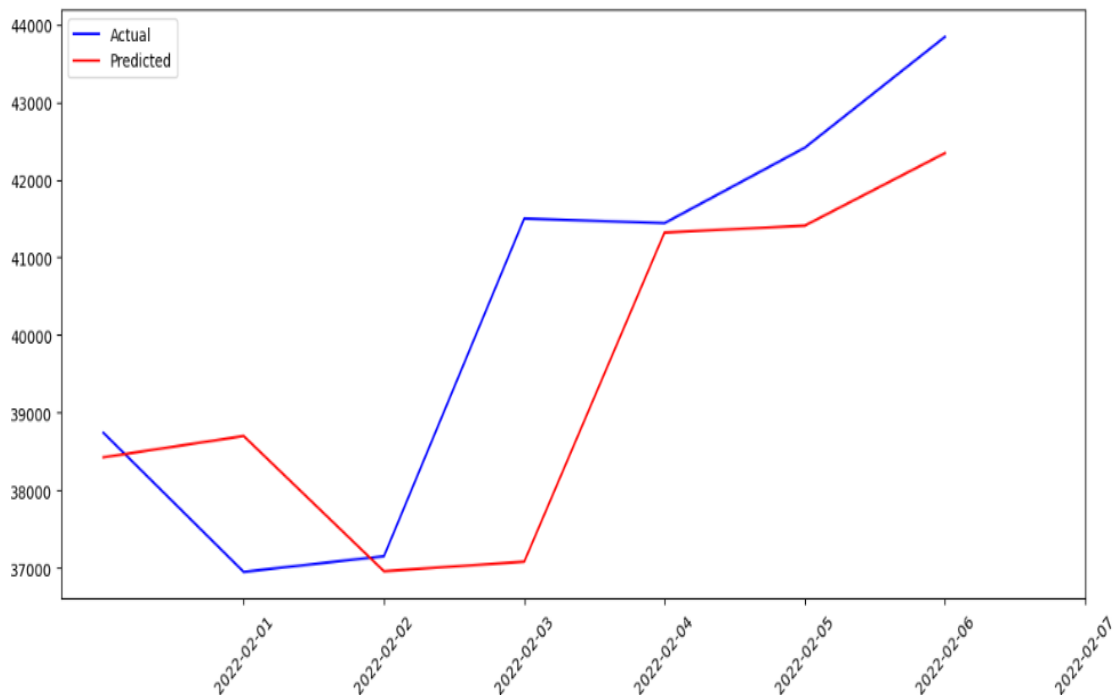
Εικόνα 52 Διάγραμμα πραγματικών και προβλεπόμενων τιμών Bitcoin

```
>Predicted=38714.528, Expected=36953
>Predicted=36924.239, Expected=37155
>Predicted=37125.857, Expected=41501
>Predicted=41472.130, Expected=41441
>Predicted=41412.419, Expected=42412
>Predicted=42383.689, Expected=43840
RMSE: 1908.740
```

Χρησιμοποιώντας τον συγκεκριμένο συνδυασμό διαπιστώνει κανείς ότι ο δείκτης  $RMSE=1908,74$  είναι ικανοποιητικός και η πρόγνωση δίνει πολύ καλά αποτελέσματα. Οι προβλεπόμενες τιμές ακολουθούν με μικρές αποκλίσεις την πορεία των πραγματικών τιμών. Επομένως ο συγκεκριμένος συνδυασμός ενδείκνυται για πρόγνωση

### 5.6.3. ARIMA (2,0,1)

Για το μοντέλο ARIMA(2,0,1) έχουμε τα εξής αποτελέσματα:



Εικόνα 54 Διάγραμμα πραγματικών και προβλεπόμενων τιμών Bitcoin



```
>Predicted=38427.138, Expected=38743
>Predicted=38701.565, Expected=36953
>Predicted=36961.336, Expected=37155
>Predicted=37084.099, Expected=41501
>Predicted=41319.737, Expected=41441
>Predicted=41409.474, Expected=42412
>Predicted=42343.571, Expected=43840
RMSE: 1925.897
```

*Εικόνα 55 Πραγματικές και προβλεπόμενες τιμές*

Το μοντέλο ARIMA (2,0,1) εμφανίζει παρόμοια αποτελέσματα με το ARIMA(1,0,1). Ο δείκτης RMSE=1925,897 και έχει μικρή διαφορά από το δείκτη του μοντέλου ARIMA(1,0,1) που είναι RMSE=1913,408 καθώς και από τον δείκτη RMSE = 1908,74 του μοντέλου ARIMA(0,1,0). Όπως φαίνεται και σε αυτήν την περίπτωση οι προβλεπόμενες τιμές ακολουθούν τις πραγματικές τιμές. Και οι τρεις δείκτες ενδείκνυνται για πρόγνωση χρονοσειρών εφόσον είναι αποδεκτοί από την ερευνητική βιβλιογραφία.

## **5.7. Μοντέλο LSTM και αποτελέσματα**

Στην ενότητα αυτή θα χρησιμοποιηθεί το τεχνητό νευρωνικό δίκτυο LSTM για να γίνει πρόγνωση των ίδιων τιμών της χρονοσειράς. Τα RNN δίκτυα, όπως το LSTM μπορούν να διαχειριστούν μη στάσιμες χρονοσειρές, επομένως δεν χρειάζεται να γίνει κάποιος μετασχηματισμός της χρονοσειράς. Σύμφωνα με την βιβλιογραφία, τα τεχνητά νευρωνικά δίκτυα παράγουν αποτελέσματα με μεγάλη ακρίβεια και με μεγάλο ποσοστό επιτυχίας στην πρόβλεψη.

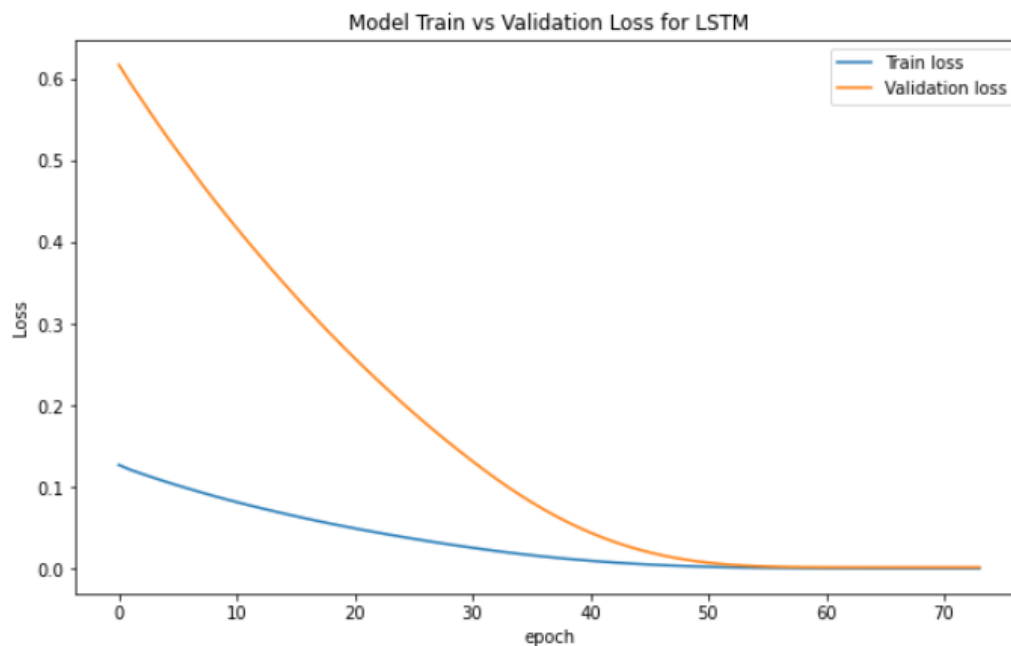
Το πρώτο βήμα που γίνεται στην συγκεκριμένη περίπτωση είναι η απομάκρυνση των τελευταίων παρατηρήσεων του συνόλου δεδομένων για να γίνει η σύγκρισή τους με τις προβλεπόμενες τιμές, όπως ακριβώς συνέβη και στο μοντέλο ARIMA. Το σύνολο των δεδομένων χωρίζεται πάλι σε 80% σετ δεδομένων εκπαίδευσης και 20% σετ δεδομένων ελέγχου. Είναι σημαντικό να αναφερθεί, ότι τα δίκτυα LSTM λειτουργούν καλύτερα με δεδομένα που ακολουθούν την κανονική κατανομή, επομένως γίνεται ένας μικρός μετασχηματισμός στα δεδομένα τέτοιος ώστε το εύρος τιμών να λαμβάνουν τιμές από 0 έως 1, τόσο για τα δεδομένα εκπαίδευσης όσο και για τα δεδομένα ελέγχου.

Το επόμενο βήμα που πραγματοποιείται, είναι η μετατροπή της μορφής των δεδομένων σε πίνακα, τέτοιος ώστε να αποτελείται από μία στήλη που περιλαμβάνει τις

μετασχηματισμένες τιμές και να έχει εύρος όσες είναι οι παρατηρήσεις της χρονοσειράς. Με αυτόν τον τρόπο γίνεται κατανοητό ποιες είναι οι τιμές των δεδομένων για πρόβλεψη και ποιες είναι οι τιμές των δεδομένων για εκπαίδευση.

Στο πλαίσιο ανάπτυξης του τεχνητού νευρωνικού δικτύου LSTM χρησιμοποιήθηκαν βελτιστοποιητές και συναρτήσεις ενεργοποίησης, μεταβλητές δηλαδή που χρησιμοποιούνται συχνά στην βιβλιογραφία για την πρόγνωση χρηματιστηριακών δεικτών, δεικτών κρυπτονομισμάτων κτλ. Εξετάστηκαν αρκετοί συνδυασμοί τέτοιων μεταβλητών, διάφορα μεγέθη για το νευρωνικό δίκτυο για να παράξει το καλύτερο δυνατό αποτέλεσμα. Στην παρούσα διπλωματική εργασία αξίζει να σημειωθεί ότι το νευρωνικό δίκτυο πρέπει να δομηθεί σε μικρές διαστάσεις και αριθμό νευρώνων, λόγω του μικρού αριθμού δεδομένων που είναι για επεξεργασία. Για το κατάλληλο στήσιμο του νευρωνικού δικτύου επιλέχθηκε μεγάλος αριθμός επαναλήψεων της εκπαίδευσης (epochs=10000). Έγινε η κατάλληλη ρύθμιση έτσι ώστε οι επαναλήψεις να τερματίζονται όταν η συνάρτηση απώλειας δεν μειώνεται σημαντικά για κάποια χρονικά βήματα. Χρησιμοποιήθηκαν αρκετοί συνδυασμοί διαφορετικών μεταβλητών, συναρτήσεων ενεργοποίησης και κόστους, διαφορετικές τιμές για batch size, ρυθμό εκμάθησης και αριθμό νευρώνων, έτσι ώστε να βρεθεί ο κατάλληλος συνδυασμός που θα αποδώσει τα καλύτερα αποτελέσματα.

Ένα άλλο σημαντικό στοιχείο που πρέπει να εξετάζεται κατά την παραμετροποίηση του τεχνητού νευρωνικού δικτύου είναι η αποφυγή του overfitting, δηλαδή της υπερπροσαρμογής. Σε αυτή την περίπτωση το μοντέλο εξειδικεύεται παραπάνω από ότι πρέπει στα δεδομένα εκπαίδευσης και εμφανίζει καλή προβλεπτική ικανότητα σε αυτά. Το μειονέκτημά του όμως είναι ότι δεν παράγει καλά αποτελέσματα στην πρόβλεψη τιμών μετά την εκπαίδευση. Για το λόγο αυτό δοκιμάστηκε σε κάθε πιθανό συνδυασμό και το φαινόμενο της υπερπροσαρμογής, το οποίο φαίνεται καλύτερα στο γράφημα της εικόνας 56.



Εικόνα 56 Διάγραμμα απώλειας εκπαίδευσης- ελέγχου

Στο σχεδιάγραμμα αυτό φαίνεται η τιμή της συνάρτησης απώλειας σε κάθε επανάληψη της εκπαιδευτικής διαδικασίας. Το καλύτερο σενάριο είναι η γραφική παράσταση της συνάρτησης απώλειας του σετ ελέγχου να είναι πάνω από την γραφική συνάρτηση απώλειας του σετ εκπαίδευσης και η διαδικασία να τερματίζει πριν συναντηθούν. Όταν αυτές οι δύο γραφικές παραστάσεις συναντηθούν τότε συμβαίνει το φαινόμενο overfitting, ενώ αν η γραφική παράσταση της συνάρτησης απώλειας του σετ ελέγχου είναι πιο πάνω από την γραφική παράσταση της συνάρτησης απώλειας του σετ εκπαίδευσης, τότε συμβαίνει το φαινόμενο underfitting ή αλλιώς υποπροσαρμογή και θα πρέπει να ρυθμιστούν κατάλληλα οι παράμετροι.

Στην έρευνα της διπλωματικής εργασίας χρησιμοποιήθηκαν οι παράμετροι που εμφανίζονται στην εικόνα 57.

<b>Αριθμός νευρώνων</b>	64
<b>Ρυθμός εκμάθησης</b>	0.00001
<b>Βελτιστοποιητής</b>	Adam
<b>Συνάρτηση απώλειας</b>	MSE
<b>Επαναλήψεις</b>	1000
<b>Batch size</b>	7
<b>Συναρτηση ενεργοποίησης</b>	tanh

Εικόνα 57 Παραμετροποίηση LSTM

Εισάγοντας αυτές τις παραμέτρους και έχοντας προσαρμόσει κατάλληλα το μοντέλο πρόβλεψης, ξεκινάει το σετ επαναλήψεων για τα δεδομένα εκπαίδευσης, όπως φαίνεται στην εικόνα 58.

```
Epoch 999/1000
123/123 [=====] - 1s 5ms/step - loss: 5.2240e-04 - val_loss: 0.0013
Epoch 1000/1000
123/123 [=====] - 1s 4ms/step - loss: 5.2206e-04 - val_loss: 0.0013
Model: "sequential_15"

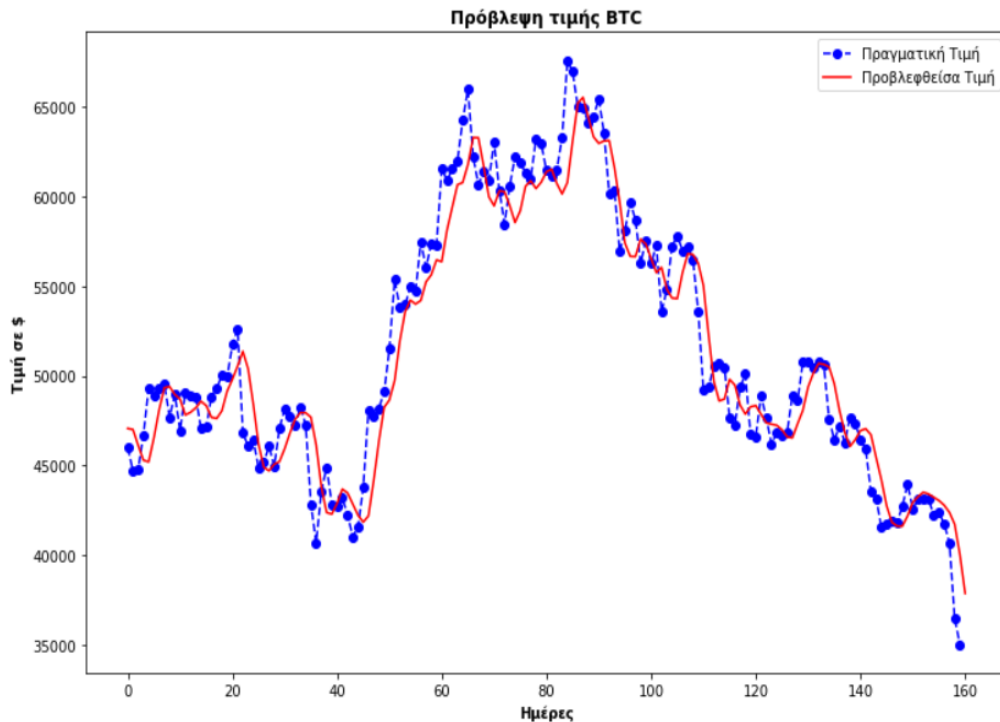
Layer (type)                 Output Shape              Param #
=====
lstm_15 (LSTM)                (None, 64)                16896
dense_15 (Dense)              (None, 1)                  65
=====

Total params: 16,961
Trainable params: 16,961
Non-trainable params: 0

7/7 [=====] - 0s 3ms/step
RMSE = 2167.370628257379
```

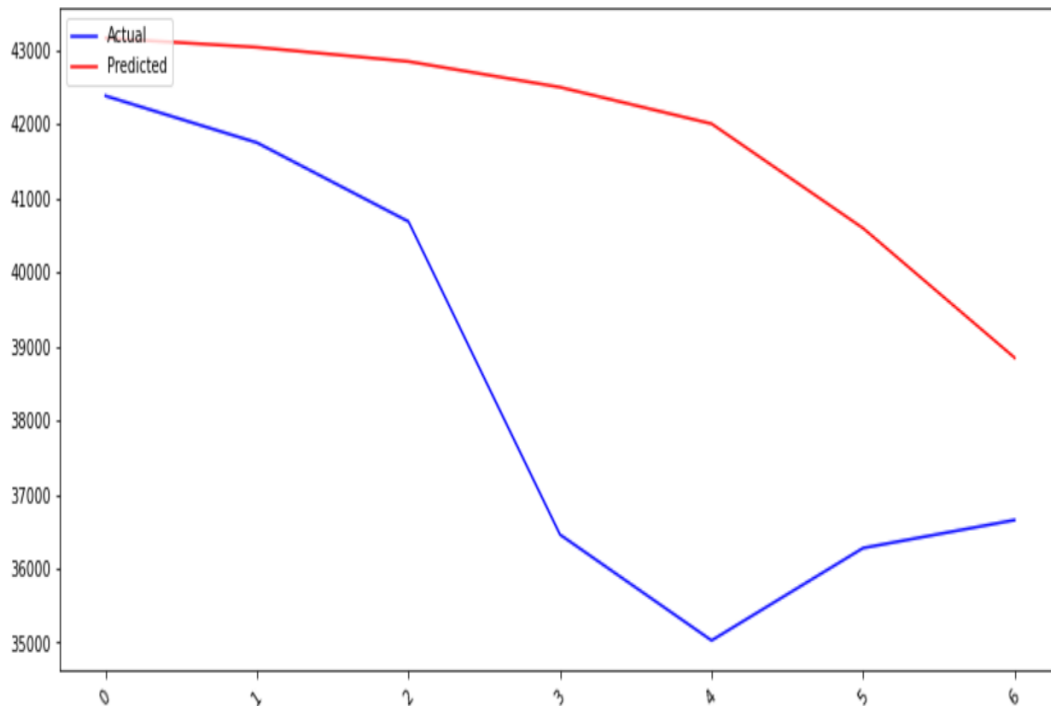
Εικόνα 58 Εκπαίδευση Δεδομένων

Φαίνεται ότι οι επαναλήψεις εκμάθησης τερματίζουν μετά από 1000 επαναλήψεις και ο δείκτης RMSE παίρνει την τιμή 2167.37 που είναι αρκετά ικανοποιητικό σχετικά με την βιβλιογραφία. Στην συνέχεια γίνεται πρόγνωση για τα δεδομένα ελέγχου και όπως φαίνεται στην εικόνα 59 οι προβλεφθείσες τιμές ακολουθούν τις πραγματικές τιμές.



Εικόνα 59 Διάγραμμα πραγματικών και προβλεπόμενων τιμών LSTM

Η γραφική παράσταση της εικόνας 60 δείχνει με μπλε χρώμα τις πραγματικές τιμές των τελευταίων 7 ημερών και με κόκκινο χρώμα τις προβλεπόμενες τιμές. Το μοντέλο προέβλεψε με σφάλμα  $RMSE = 4058,37$  το οποίο δεν είναι πολύ ικανοποιητικό. Η διαφορά ανάμεσα στο  $RMSE$  σφάλμα της εκπαίδευσης και το αντίστοιχο σφάλμα της πρόβλεψης είναι σχετικά μεγάλο και αυτό φαίνεται και από την διαφορά της τιμής μεταξύ τρίτης και τέταρτης ημέρας, κατά τις οποίες υπάρχει μεγάλη απόκλιση της προβλεπόμενης τιμής από την πραγματική. Επομένως η ακρίβεια του μοντέλου δεν κρίνεται ικανοποιητική.



Εικόνα 60 Πραγματικές και προβλεπόμενες τιμές για 7 ημέρες

## 5.8. Μοντέλο SVM και αποτελέσματα

Στην ενότητα αυτή θα ασχοληθούμε με το μοντέλο SVM και ειδικότερα με μία υποκατηγορία του το μοντέλο SVR που εξειδικεύονται στην παλινδρόμηση, το οποίο ταιριάζει απόλυτα στο πρόβλημα της πρόβλεψης κρυπτονομισμάτων. Οι παράμετροι που θα εξεταστούν είναι η kernel function, όπου μπορεί να είναι linear (γραμμική), rbf (Gaussian or Radial Basis Function) ή polynomial (πολυώνυμο τάξης  $p$ ). Στην συγκεκριμένη εργασία θα χρησιμοποιηθεί η rbf. Το σύνολο των δεδομένων χωρίζεται σε 90% δεδομένα για εκπαίδευση και σε 10% δεδομένα για έλεγχο. Αυτό γίνεται λόγω του μικρού αριθμού παρατηρήσεων που εξετάζονται και σύμφωνα με την ερευνητική βιβλιογραφία είναι αποδεκτό.

Στην συνέχεια ακολουθεί η επιλογή των δύο μεταβλητών  $C$  και  $\gamma$  του αλγορίθμου SVM. Σε κάθε υλοποίηση ενός τέτοιου μοντέλου συνιστάται η αναζήτηση για την εύρεση των τιμών των μεταβλητών  $C$  και  $\gamma$ , οι οποίες βελτιστοποιούν την ακρίβεια του μοντέλου. Πιο απλά, για την επιλογή των βέλτιστων τιμών αυτών των μεταβλητών είναι απαραίτητη η επανάληψη της διαδικασίας εκπαίδευσης και ελέγχου του μοντέλου χρησιμοποιώντας κάθε φορά διαφορετικούς συνδυασμούς, έτσι ώστε να καταλήξει στο βέλτιστο δυνατό. Οι τιμές που

βρέθηκαν για αυτές τις μεταβλητές είναι η  $C=1e3$  (1000) και  $\gamma=1e-06$  (0,000001) όπως φαίνεται στην εικόνα 61.

```
In [17]: # SVM Model
from sklearn.svm import SVR
# Create and train the Support Vector Machine (Regression) using the radial
svr = SVR( kernel="rbf", C=1e3, gamma=0.000001)
svr.fit(x_train, y_train)

Out[17]: SVR(C=1000.0, gamma=1e-06)
```

Εικόνα 61 Παραμετροποίηση SVM

Στην συνέχεια εξετάζεται η ακρίβεια του μοντέλου. Η ιδανική τιμή που θα μπορούσε να έχει ο δείκτης ακρίβειας για το SVM μοντέλο είναι 1, ενώ η τιμή του δείκτη ακρίβειας που προκύπτει είναι 0,902, η οποία είναι ικανοποιητική.

```
In [18]: # Testing Model: Score returns the accuracy of the prediction.
# The best possible score is 1.0
svm_confidence = svr.score(x_test, y_test)
print("svm confidence: ", svm_confidence)

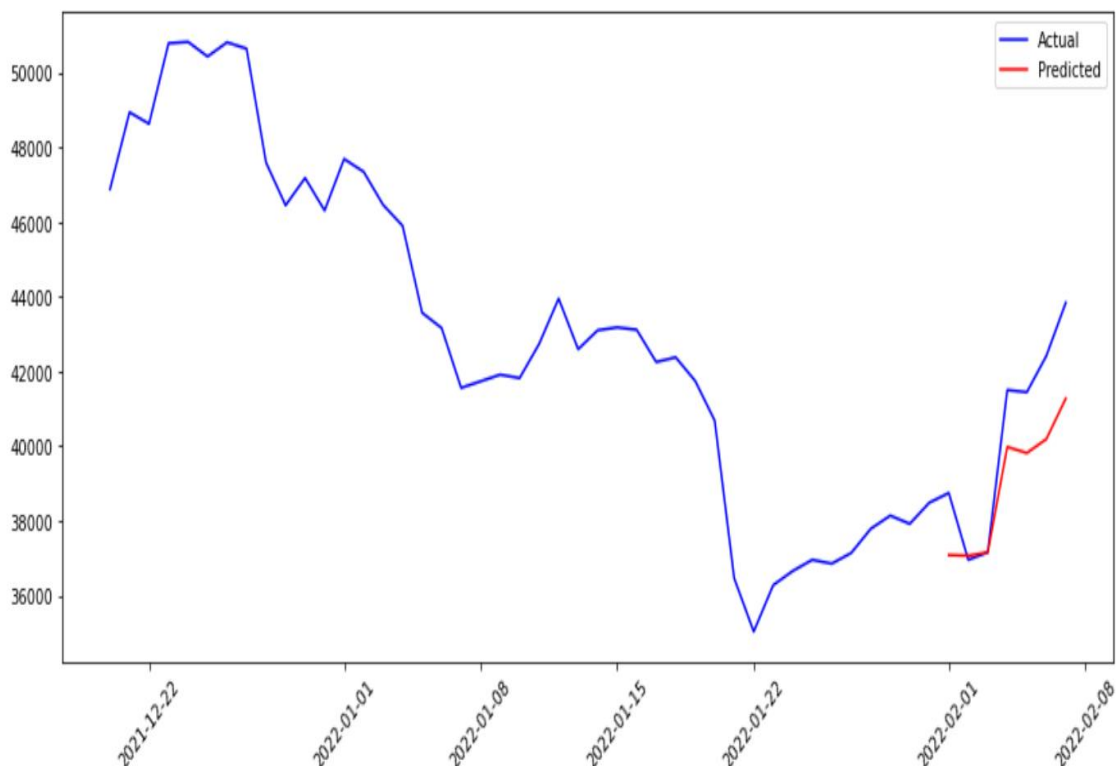
svm confidence: 0.9022885309135328
```

Εικόνα 62 Δείκτης ακρίβειας SVM

Πρέπει και εδώ να σημειωθεί ότι οι τελευταίες 7 ημέρες του συνόλου δεδομένων αποκρύπτονται, με σκοπό να προβλεφθούν και να συγκριθούν με τις πραγματικές τιμές. Από την διαδικασία της πρόβλεψης προκύπτει ότι οι προβλεπόμενες τιμές φαίνονται την εικόνα 63 και ακολουθούν τις πραγματικές όπως φαίνεται στο γράφημα της εικόνας 64.

2022-02-01	37081.776580
2022-02-02	37064.871323
2022-02-03	37155.269122
2022-02-04	39977.206563
2022-02-05	39813.803908
2022-02-06	40185.173998
2022-02-07	41274.792974

Εικόνα 63 Προβλεπόμενες τιμές SVM



Εικόνα 64 Διάγραμμα πραγματικών και προβλεπόμενων τιμών SVM



Ο δείκτης πρόβλεψης του μοντέλου είναι ο  $RMSE=1659,83$  , ο οποίος είναι πολύ ικανοποιητικός σύμφωνα με την βιβλιογραφία και το μοντέλο SVM ενδείκνυται για πρόβλεψη.

```
In [24]: from sklearn.metrics import mean_squared_error

rmse = mean_squared_error(actual, svm_prediction, squared=False)
rmse

Out[24]: 1659.8361782301288
```

*Εικόνα 65 Δείκτης αξιολόγησης μοντέλου SVM*

## **Συμπεράσματα και σχόλια**

Αξιολογώντας τα αποτελέσματα της ερευνητικής διαδικασίας της διπλωματικής εργασίας, καταλήγουμε στο συμπέρασμα ότι ο συνδυασμός αποτελεσμάτων από διάφορα στατιστικά μοντέλα, μοντέλα μηχανικής μάθησης και τεχνητών νευρωνικών δικτύων προσδίδει μεγαλύτερη ακρίβεια στην πρόβλεψη των τιμών των κρυπτονομισμάτων. Όπως είναι γνωστό η φύση των κρυπτονομισμάτων είναι τέτοια που καθιστά αρκετά δύσκολη την πρόβλεψη των τιμών τους. Το στατιστικό μοντέλο ARIMA που χρησιμοποιήθηκε, έδωσε σημαντικές πληροφορίες για τα χαρακτηριστικά της χρονοσειράς που εξετάζεται, τα οποία δεν είναι και πολύ κατανοητά από τα τεχνητά νευρωνικά δίκτυα. Σύμφωνα με την βιβλιογραφία το μοντέλο ARIMA προσδίδει μεγάλη ακρίβεια στην πρόβλεψη της χρονοσειράς. Ο δείκτης αξιολόγησης της πρόβλεψης για το μοντέλο ARIMA(1,0,1) που προέκυψε από τις γραφικές παραστάσεις αυτοσυσχέτισης και μερικής αυτοσυσχέτισης είναι  $RMSE=1913,4$  και τα αποτελέσματα δείχνουν να ακολουθούν τα πραγματικά αποτελέσματα. Επίσης η διαδικασία εύρεσης του κατάλληλου συνδυασμού ARIMA με το μικρότερο RMSE έβγαλε καλά

αποτελέσματα καθώς ο βέλτιστος συνδυασμός που έδειξε ήταν ο ARIMA(0,1,0) με δείκτη πρόβλεψης  $RMSE=1908,74$ , ο οποίος είναι ικανοποιητικός και ενδείκνυται για πρόβλεψη. Τέλος δοκιμάστηκε και ένας συνδυασμός που συναντάται στην βιβλιογραφία ο ARIMA(2,0,1) με δείκτη πρόβλεψης  $RMSE=1925,897$  που ακολουθεί και αυτός τις πραγματικές τιμές του Bitcoin.

Στην συνέχεια επιλέχθηκε το τεχνητό νευρωνικό δίκτυο LSTM. Είναι γνωστό ότι τα τελευταία χρόνια αυξάνεται η ανάγκη για χρήση ανατροφοδοτούμενων νευρωνικών δικτύων, λόγω της ικανότητας τους στην πρόγνωση βάσει της βιβλιογραφίας. Στην συγκεκριμένη εργασία ο δείκτης αξιολόγησής του σετ εκπαίδευσης είναι ο  $RMSE = 2167,37$ , ο οποίος είναι ικανοποιητικός αλλά ο δείκτης πρόβλεψης που έδωσε το συγκεκριμένο μοντέλο είναι ο  $RMSE=4058.37$ . Η απόκλιση μεταξύ  $RMSE$  εκπαίδευσης και πρόβλεψης είναι αρκετά μεγάλη, γεγονός που καθιστά το μοντέλο λιγότερο αξιόπιστο. Ένα σχόλιο που πρέπει να γίνει εδώ είναι ότι πρέπει να δίνεται μεγάλη προσοχή στα αποτελέσματα που παράγονται από τέτοιου είδους μοντέλα και κρίνεται σκόπιμο η χρήση διάφορων συνδυασμών παραμετροποίησης, οι οποίοι μπορούν να παρουσιάσουν μεγαλύτερα ποσοστά ακρίβειας και να εμφανίσουν καλύτερα αποτελέσματα πρόβλεψης.

Τέλος χρησιμοποιήθηκε το μοντέλο SVM (Support Vector Machine). Το μοντέλο αυτό εμφάνισε δείκτη ακρίβειας με τιμή 0,9. Η ιδανική τιμή ακρίβειας για το συγκεκριμένο μοντέλο είναι 1. Ο δείκτης πρόβλεψης είναι ο  $RMSE= 1659.83$ , που είναι πολύ ικανοποιητικό και καθιστά το μοντέλο ιδανικό για πρόβλεψη.

Σύμφωνα με τη βιβλιογραφία, όσον αφορά την ακρίβεια του μοντέλου στην πρόγνωση, τα τεχνητά νευρωνικά δίκτυα δίνουν τα καλύτερα αποτελέσματα και τα μεγαλύτερα ποσοστά ακρίβειας. Η συγκεκριμένη εργασία έδειξε ότι τα καλύτερα δυνατά αποτελέσματα δόθηκαν από το μοντέλο SVM, μετά το στατιστικό μοντέλο ARIMA και τέλος το τεχνητό νευρωνικό δίκτυο LSTM.

Συνοψίζοντας, κατά την παρούσα διπλωματική εργασία επετεύχθη ο στόχος της. Τα μοντέλα πρόγνωσης που χρησιμοποιήθηκαν για την πρόβλεψη τιμών χρονοσειράς Bitcoin πέτυχαν τα καλύτερα δυνατά αποτελέσματα και κατά τη διάρκεια της έρευνας αποκτήθηκε εξειδικευμένη γνώση πάνω στο αντικείμενο.

## **Προτάσεις για περαιτέρω μελέτη**

Οι στόχοι της συγκεκριμένης διπλωματικής εργασίας σε γενικές γραμμές επετευχθησαν. Σίγουρα όμως υπάρχουν ορισμένα πράγματα που μπορούν να μελετηθούν

πραιτέρω σε μελλοντικό χρόνο. Ένα σημαντικό βήμα που θα μπορούσε να αποτελέσει τη συνέχεια της έρευνας, είναι η χρήση και άλλων μεταβλητών που επηρεάζουν το δείκτη τιμών Bitcoin, όπως για παράδειγμα στις μέρες οι υπόλοιποι χρηματιστηριακοί δείκτες δεν βιώνουν καλή περίοδο και οι έντονες μεταβολές που έχουν επηρεάζουν εξίσου και τα κρυπτονομισμάτα. Με βάση αυτή την παραδοχή θα μπορούσαμε να προχωρήσουμε στην κατασκευή μοντέλων πρόβλεψης για την ανάλυση πολυμεταβλητών (multivariate) χρονοσειρών και ίσως να είχαμε και καλύτερα αποτελέσματα.

## Βιβλιογραφικές αναφορές

- Bashirahamad F. M. & Gaurav C., «Univariate Time Series Models for Forecasting Stationary and Non-stationary Data: A Brief Review». *Smart Innovation*, March 2017.
- Βερούκιος, Β., Καγκλής, Β. & Σταυρόπουλος, Η. (2015). «*Η επιστήμη των δεδομένων μέσα από την γλώσσα R*». [Προπτυχιακό εγχειρίδιο]. Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις. κεφ 5.
- Βλαχάβας, Ι., Κεφαλάς, Π., Βασιλειάδης, Ν., Κόκκορας, Φ. & Σακελλαρίου, Η. (2006). «*Τεχνητή Νοημοσύνη*», 2<sup>η</sup> έκδοση. Αθήνα: Γκιούρδας.
- Box, G. E. P. & Pierce, D.A. (1970). Distribution of residual autocorrelation in autoregressive-integrated moving average time-series models. *J. Amer. Statist. Ass.*, 65, 1509-26.  
<https://doi.org/10.2307/2284333>
- Box, G.E.P., Jenkins, G. M., Reinsel, G.C. and Ljung G.M. (2015). «*Time series analysis: Forecasting and control*», 5th edition. New Jersey: Wiley.
- Brockwell, P. J. , Davis, R. A. , (2016), «*Introduction to Time Series and Forecasting*», 3rd edition, NY: Springer.
- Brockwell, P.J., Davis, R.A. (2002). «*Introduction to time series and forecasting*», 2nd edition. USA. Springer-Verlag.
- Chatfield, C. (1984). «*The Analysis of time series: An Introduction*», 3rd Edition, NY: Springer.
- Cochrane, J.H. (1997). «*Time Series for Macroeconomics and Finance*». Graduate School of Business, University of Chicago.
- Cortes, C. & Vapnik, V. (1995). «Support-vector networks». *Machine learning*, vol. 20, no. 3, pp. 273– 297.

- Διαμαντάρας, Κ. & Μπότσης, Δ. (2019). «Μηχανική Μάθηση». Αθήνα: Κλειδάριθμος.
- Dickey, D.A. and Fuller, W.A. (1981). «Likelihood Ratio Statistics for Autoregressive Time Series with Unit Root». *Econometrica*, 49, pp.1057-1072.
- Faraway, J. & Chatfield, C. (1998). «Time series forecasting with neural networks: a comparative study using the airline data». *Applied Statistics*, 47:2, pp. 231-250.
- Farooq T., Guergachi A., Krishnan S.(2007). «Chaotic time series prediction using knowledge based Green's Kernel and least-squares support vector machines. Systems, Man and Cybernetics». *IEEE*, <https://doi.org/10.1109/ICSMC.2007.4414023>.
- Georgouli, A. (2015). *Μηχανική Μάθηση* [Chapter]. In Georgouli, A. 2015. *Τεχνητή νοημοσύνη*. [Chapter] Athens: Kallipos, Open Academic Editions. chapter 4. <http://hdl.handle.net/11419/3382>
- Goodfellow, I., Bengio, Y. & Courville, A. (2016). «*Deep Learning*». [Chapter 5]. The MIT Press.
- Hamilton J.D. (1994). «*Times Series Analysis*». Princeton University Press.
- Heckerman, D.(1999). Learning in graphical models, [Chapter A] *Tutorial on learning with Bayesian networks* pp. 301–354. MIT Press
- Hochreiter, S. & Schmidhuber, J. (1997). «Long short-term memory». *Neural computation*, vol. 9, no. 8, pp. 1735–1780.
- Huisu, J. & Jaewook, L., (2017). «An Empirical Study on Modeling and Prediction of Bitcoin Prices With Bayesian Neural Networks Based on Blockchain Information». *Access*, 6:2018.pp. 5427-5437.
- Jayachandran, P. (2017). In: [Why so Delirious about Blockchain? A Technical View](#).

- Jayachandran, P. (2017). In: [The difference between public and private blockchain](#),
- Kihoro, J.M., Otieno, R.O. & Wafula, C. (2006). «Seasonal Time Series Forecasting: A Comparative Study of ARIMA and ANN Models». *African Journal of Science and Technology (AJST) Science and Engineering Series*, 5:2, pp. 41-49.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., Shin, Y. (1992). «Testing the null hypothesis of stationarity against the alternative of a unit root». *Journal of Econometrics*. 54 (1–3), pp. 159–178.
- Mitchell, T. (1997). *Machine Learning*. NY: McGraw Hill.
- Mohammed, J. Z. & Wagner, M. Jr., (2017). *Εξόρυξη και ανάλυση δεδομένων – Βασικές έννοιες και αλγόριθμοι*. Αθήνα: Κλειδάριθμος.
- Montgomery D.C., Jennings, C.L. & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting*. New Jersey: Wiley.
- Montgomery, D. C., Jennings, C. L. & Kulahci, M., (2015). *Introduction to Time Series Analysis and Forecasting*, 2nd edition, John Wiley & Sons, Inc.
- Nelson, B.K. (2008). «Time series analysis using Autoregressive Integrated Moving Average (ARIMA) models». *Academic Emergency Medicine* <https://doi.org/10.1111/j.1553-2712.1998.tb02493>.
- Ντούνης, Χ. (1995). *«Μελέτη και αγορά κρυπτονομισμάτων και η σημασία τους για την παγκόσμια οικονομία. Προβλέψεις για την εξέλιξή τους»*. Διπλωματική εργασία. Πάτρα: Ελληνικό Ανοικτό Πανεπιστήμιο.
- Paolella, M.S. (2019). *Linear models and Time Series Analysis, Regression, ANOVA, ARMA and GARCH*. WILEY Series in probability and Statistics.

Πασπάλας, Β. (2018). *Ανάπτυξη και αξιολόγηση αλγορίθμων μηχανικής μάθησης και νευρωνικών δικτύων σε δεδομένα κοινωνικών δικτύων με εφαρμογή στις μεταβολές των κρυπτονομισμάτων*. [Διπλωματική εργασία]. Αθήνα: Ε.Μ.Π.

Peter J., Brockwell R. & Davis A.(1997). «Introduction to Time Series and Forecasting», Second Edition. Springer Texts in Statistics.

Raicharoen T., Lursinsap C., Sanguanbhoki P.(2003). «*Application of critical support vector machine to time series prediction*». Proceedings of the 2003 International Symposium on Volume 5.

Ramakrishnan, R. & Gehrke, J. (2003). «*Database Management Systems*», 3rd Edition. NY: Mcgraw hill.

Rojas, R. (1996). «*Neural Networks*». Berlin: Springer-Verlag.

Sherman, A.T. & Javani, F. Zhang, H. & Golaszewski, E. (2019). «On the Origins and Variations of Blockchain Technologies». *IEEE*. 17:1, pp. 72-77.

Shumway, R. & Stoffer, D. (2011). «*Time Series Analysis and Its Applications With R Examples*» 3<sup>rd</sup> Edition. Springer.

Sodhi, P., Awasthi, N. & Sharma, V. (2019). «*Introduction to Machine Learning and Its Basic Application in Python*». Proceedings of 10th International Conference on Digital Strategies for Organizational Success.

Stutz, D. (2014). «*Introduction to Neural Networks*». Germany: RWTH Aachen University.

[Types of machine learning algorithms | en.proft.me](https://en.proft.me).

Yaffee,R.A. (2000). «*Introduction to time series analysis and forecasting (with applications of SAS and SPSS)*». New York: Academic Press, INC.

Φιλιππάκης, Ε. Μ., (2019). *Θεωρία Πιθανοτήτων και Στοιχεία Στατιστικής Ανάλυσης*. Αθήνα:  
Τσότρας.

### **Ιστοσελίδες**

<https://www.euro2day.gr/investments/crypto/article/2072469/jpmorgan-ypohorei-h-metavlhthta-toy-bitcoin-anoi.html>

<https://www.lesswrong.com/posts/RTt59BtFLqObsSiqd/a-history- of-Bayes-theorem>

<https://www.ecb.europa.eu/pub/annual/html/ecb.ar2021~14d7439b2d.el.html>

<https://www.moneyminority.com/analisi-bitcoin/>

<https://www.ot.gr/2021/08/12/oikonomia/macro/elstat-sto-15-i-anergia-ton-iounio-tou-2021/>

<https://www.euro2day.gr/news/world/article/2098900/nea-istorika-rekor-stis-times-fysikoy-aerioyreyamat.html>

<https://getelectric.gr/ilektrokinisi-ena-akrivo-paichnidi-i-to-mellon-tis-aytokinisis/>

<https://parallaximag.gr/pou-ofeiletai-chamili-stathmi-tis-thalassas-tis-teleftaies-imeres-105763>

<https://otexts.com/fpp2/wn.html>

<https://financetrain.com/simulate-random-walk-rw-in-r>

<https://towardsdatascience.com/multi-step-time-series-forecasting-with-arima-lightgbm-and-prophet-cc9e3f95dfb0>

<https://medium.com/@dhaval.sony.504/everything-in-short-about-simple-linear-regression-633fc9f8dd65>

<https://www.i2tutorials.com/what-is-the-difference-between-bidirectional-rnn-and-rnn/>

<https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>