



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



Αξιολόγηση μοντέλων για τη βραχυπρόθεσμη πρόβλεψη της τιμής του Bitcoin

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αλεξάνδρα Καλλίρι

Επιβλέπων: Βασίλειος Ασημακόπουλος

Καθηγητής Ε.Μ.Π.

Υπεύθυνος: Ευάγγελος Σπηλιώτης

Διδάκτωρ Ε.Μ.Π.

Αθήνα, Οκτώβριος 2020



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



Αξιολόγηση μοντέλων για τη βραχυπρόθεσμη πρόβλεψη της τιμής του Bitcoin

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αλεξάνδρας Καλλίρι

Επιβλέπων: Βασίλειος Ασημακόπουλος

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την (Ημερομηνία) 2020.

(Υπογραφή)

.....

Βασίλειος
Ασημακόπουλος
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....

Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

(Υπογραφή)

.....

Χρυσόστομος Δούκας
Επικ. Καθηγητής
Ε.Μ.Π.

(Υπογραφή)

.....

ΑΛΕΞΑΝΔΡΑ ΚΑΛΛΙΡΙ

Διπλωματούχος Οργάνωσης και Διοίκησης Επιχειρήσεων Πανεπιστημίου Πειραιώς.

© 2020 – All rights reserved. Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

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

Περίληψη

Μετρώντας κάτι περισσότερο από μία δεκαετία ζωής, το Bitcoin αποτελεί αναμφισβήτητα μία από τις μεγαλύτερες καινοτομίες της δεκαετίας που πέρασε. Όντας δημιούργημα του Satoshi Nakamoto, το Bitcoin και η τεχνολογία πάνω στην οποία βασίζεται έχει πλέον αλλάξει τον τρόπο με τον οποίον ο κόσμος μετακινεί αξία. Ως χρηματοδοτικό αγαθό, το Bitcoin είναι ελεύθερο από τον έλεγχο οποιασδήποτε κεντρικής τράπεζας και δεν χρειάζεται να πιστοποιηθεί από τρίτους για να διασφαλιστεί η αξία του. Αντίθετα, παρακολουθείται και επαληθεύεται από χιλιάδες υπολογιστές σε ένα κοινό δίκτυο που ονομάζεται blockchain, το οποίο χρησιμοποιεί κρυπτογραφία για να επιτύχει συμφωνία. Ως επενδυτικό προϊόν, το Bitcoin είναι ιδιαίτερα ευμετάβλητο καθώς παρουσιάζει σημαντικές διακυμάνσεις στην αξία μονάδας του. Παρόλα αυτά, είναι ιδιαίτερα ελκυστικό προς τους επενδυτές καθώς δεν επηρεάζεται από τις διακυμάνσεις στις πιστωτικές αγορές.

Η παρούσα εργασία μελετά την τιμή του Bitcoin και προσπαθεί να προβλέψει τη διακύμανσή της σε βραχυπρόθεσμους χρονικούς ορίζοντες (από μία έως δεκατέσσερις ημέρες). Για την επίτευξη του στόχου έγινε χρήση και αξιολόγηση των σημαντικότερων μοντέλων πρόβλεψης, τόσο στατιστικών όσο και μηχανικής μάθησης. Τα ιστορικά δεδομένα που χρησιμοποιήθηκαν προήλθαν από αξιόπιστα και ανοιχτά αποθετήρια και μετασχηματίστηκαν κατάλληλα για τις ανάγκες της μελέτης. Η μελέτη πραγματοποιήθηκε με χρήση γλώσσας R, στην πλατφόρμα RStudio. Οι υλοποιήσεις των μοντέλων που χρησιμοποιήθηκαν και των γραφημάτων που δημιουργήθηκαν προήλθαν από ευρέως διαδεδομένες βιβλιοθήκες R ανοικτού κώδικα.

Λέξεις Κλειδιά: Bitcoin, Κρυπτονομίσματα, Μοντέλα Πρόβλεψης, Μηχανική Μάθηση, R

Abstract

Counting more than ten years of life, Bitcoin is arguably one of the biggest innovations of the past decade. Created by Satoshi Nakamoto, Bitcoin and the technology on which it is based have now changed the way people move value. As a financial asset, Bitcoin is free from the control of any central bank and does not need to be certified by third parties to ensure its value. Instead, it is monitored and verified by thousands of computers on a common network called blockchain, which uses cryptography to reach an agreement. As an investment product, Bitcoin is highly volatile as it fluctuates significantly in its unit value. Nevertheless, it is particularly attractive to investors as it is not affected by fluctuations in credit markets.

The present study deals with the value of Bitcoin and tries to predict its fluctuation in short term horizons (from one to fourteen days). To achieve the goal, the most significant prediction models were used and evaluated, both statistical and machine learning. The historical data used came from reliable and open repositories and were transformed appropriately for the needs of the study. The study was conducted using R language, on the RStudio platform. The implementations of the models used and the graphs generated came from well-known open source R libraries.

Key Words: Bitcoin, Cryptocurrencies, Forecasting Models, Machine Learning, R

Ευχαριστίες

Η παρούσα διπλωματική εργασία εκπονήθηκε στα πλαίσια του Προγράμματος Μεταπτυχιακών Σπουδών «Τεχνο - Οικονομικά Συστήματα» στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Εθνικού Μετσόβιου Πολυτεχνείου.

Η παρούσα εργασία αποτελεί τον επίλογο μιας προσπάθειας εκπλήρωσης ενός σημαντικού για εμένα ακαδημαϊκού στόχου. Ο στόχος αυτός δεν θα μου έδινε τόση ικανοποίηση αν η πορεία προς αυτόν δεν ήταν εξίσου ικανοποιητική. Μια πορεία που δεν θα ήταν το ίδιο όμορφη αν δεν είχα δίπλα μου ανθρώπους να με στηρίζουν και ανθρώπους να πορευτώ μαζί. Στα πλαίσια όλων αυτών θα ήθελα να ευχαριστήσω θερμά:

Τον επιβλέποντα καθηγητή κ. Βασίλειο Ασημακόπουλο για την ευκαιρία που μου έδωσε να εκπονήσω τη Διπλωματική αυτή.

Τον Καθηγητή κ. Ι. Ψαρρά και τον Αναπληρωτή Καθηγητή κ. Χ. Δούκα για την τιμή που μου έκαναν να συμμετάσχουν στην εξεταστική επιτροπή της εργασίας.

Τον Ευάγγελο Σπηλιώτη, ερευνητικό συνεργάτη της Μονάδας Προβλέψεων και Στρατηγικής, για την καθοδήγηση και τις συμβουλές, για τη συνεργασία και για την ευγενική του προθυμία να με βοηθήσει σε όλα τα στάδια της εργασίας αυτής.

Τον Σταμάτη που είναι πάντα εκεί.

Την Ελένη και τη Νικόλ που αποτέλεσαν πολλές φορές αιτία για περισσότερη προσπάθεια.

Την οικογένεια και τους φίλους μου που με στηρίζουν και με ενθαρρύνουν.

Πίνακας Περιεχομένων

Περίληψη	5
Abstract.....	7
Ευχαριστίες.....	9
Πίνακας Περιεχομένων.....	11
Πίνακας Γραφημάτων.....	15
Πίνακας Εικόνων.....	15
Ευρετήριο Πινάκων	15
1 Εισαγωγή.....	1
2 Μελετώντας τα κρυπτονομίσματα.....	5
2.1 Βασικές Έννοιες	6
2.1.1 Συναλλαγή (Transaction).....	6
2.1.2 Παραστατικό Χρήμα (Fiat money)	6
2.1.3 Αποταμίευση (Saving).....	6
2.1.4 Επένδυση (Investment).....	6
2.1.5 Χρηματιστήριο (Stock exchange)	7
2.1.6 Ηλεκτρονικό χρήμα (Electronic money)	7
2.1.7 Ψηφιακό χρήμα (Digital currency)	7
2.1.8 Εικονικό νόμισμα (Virtual currency).....	7
2.1.9 Εναλλακτικό νόμισμα (Alternative currency).....	7
2.2 Κρυπτονομίσματα.....	8
2.2.1 Βασικές έννοιες κρυπτονομισμάτων	9
2.2.2 Τα κυριότερα κρυπτονομίσματα εκτός από το Bitcoin.....	11
2.3 Επενδύσεις.....	13
2.4 Bitcoin	14
2.4.1 Ορόσημα.....	15
2.4.2 Βασικά χαρακτηριστικά	16
2.4.3 Πλεονεκτήματα.....	17
2.4.4 Μειονεκτήματα	17
2.4.5 Το Bitcoin ως επενδυτικό αγαθό.....	18
3 Τεχνικές Προβλέψεων.....	19
3.1 Γενικά περί Προβλέψεων	19
3.1.1 Πρόβλεψη	19
3.1.2 Κατηγορίες Μεθόδων Προβλέψεων.....	20
3.2 Χρονοσειρά - Χαρακτηριστικά Χρονοσειρών.....	22

3.3 Μέθοδοι Προβλέψεων	23
3.3.1 Απλοϊκή μέθοδος (Naive)	24
3.3.2 Μέθοδος Αποσύνθεσης.....	24
3.3.3 Μέθοδοι Εξομάλυνσης	24
3.3.4 Μοντέλα Παλινδρόμησης.....	27
3.4 Αξιολόγηση Μεθόδων πρόβλεψης.....	27
3.5 Επιλογή Κατάλληλης Μεθόδου Πρόβλεψης	30
4 Παράγοντες που επηρεάζουν την τιμή του Bitcoin.....	31
4.1 Συνολικά Διαθέσιμα Bitcoin.....	31
4.2 Συναλλαγματικές Ισοτιμίες	31
4.3 Μακροοικονομικά Δεδομένα	32
4.4 Δημοφιλία	32
4.5 Συναίσθημα των συναλλασσόμενων.....	32
4.6 Γενικός Δείκτης Χρηματιστηρίου της Κίνας.....	32
4.7 Hashrate.....	33
4.8 Ευγενή Μέταλλα	33
4.9 Βασικά αγαθά	33
4.10 Χρηματιστηριακοί Δείκτες.....	34
4.11 Όγκος Συναλλαγών	34
5 Επικρατέστερα μοντέλα πρόβλεψης της τιμής του Bitcoin	35
5.1 AutoRegressive Integrated Moving Average (ARIMA)	35
5.2 Gradient Boosting Machine (GBM)	36
5.3 Extreme Gradient Boosting (XGBoost)	38
5.4 Prophet.....	40
5.5 Random Forest (RF).....	41
5.6 Μοντέλα πρόβλεψης της τιμής του Bitcoin εκτός πλαισίου έρευνας της παρούσας εργασίας.....	43
5.6.1 Recurrent Neural Networks (RNN)	43
5.6.2 Long Short Term Memory (LSTM).....	44
6 Σχεδιασμός και Ανάπτυξη Μελέτης - Πρόβλεψη Τιμής του Bitcoin	47
6.1 Εργαλεία Υλοποίησης.....	47
6.1.1 R.....	47
6.1.2 RStudio	49
6.2 Δεδομένα Μελέτης.....	49
6.2.1 Πηγές Άντλησης Δεδομένων	50
6.2.2 Προεπεξεργασία.....	54

6.3 Διερευνητική Ανάλυση Δεδομένων - EDA	57
6.3.1 Στατιστικά στοιχεία	57
6.3.2 Αποσύνθεση	59
6.3.3 Πίνακας Αλληλοσυσχέτισης (Correlation Matrix).....	61
6.4 Αξιολόγηση Μοντέλων	63
6.4.3 Πίνακες Επιδόσεων ανά Μοντέλο και Χρονικό Ορίζοντα	65
6.4.4 Επιλογή Μοντέλου	71
6.4.5 Αποτελέσματα Πρόβλεψης.....	72
7 Επίλογος.....	73
7.1 Σύνοψη και συμπεράσματα	73
7.2 Μελλοντικές επεκτάσεις - βελτιώσεις	75
7.2.1 Χρήση περισσότερων και πιο απαιτητικών μοντέλων	75
7.2.2 Διερεύνηση διαφορετικού ορίζοντα πρόβλεψης.....	76
Βιβλιογραφία.....	77
Ηλεκτρονικές Πηγές	79

Πίνακας Γραφημάτων

Γράφημα 1: Ιστορικότητα τιμής Bitcoin	15
Γράφημα 2: Ιστορικότητα Κρυπτονομισμάτων	51
Γράφημα 3: Ιστορικότητα Χρηματιστηριακών Δεικτών και Bitcoin	52
Γράφημα 4: Ιστορικότητα τιμών εμπορευμάτων και Bitcoin	52
Γράφημα 5: Ιστορικότητα Επισκέψεων Wikipedia και τιμών Bitcoin	53
Γράφημα 6: Ιστορικότητα Τιμής Bitcoin ως προς το Hashrate	54
Γράφημα 7: Κατανομή Τιμών Τιμής Bitcoin	57
Γράφημα 8: Κατανομή Τιμών Βασικών Εμπορευμάτων (1/2)	58
Γράφημα 9: Κατανομή Τιμών Βασικών Εμπορευμάτων (2/2)	59
Γράφημα 10: Κατανομή Χρηματιστηριακών Δεικτών	59
Γράφημα 11: Αποσύνθεση τιμής Bitcoin	60
Γράφημα 12: Αλληλοσυσχέτιση χαρακτηριστικών dataset	62

Πίνακας Εικόπων

Εικόνα 1: Εκπαίδευση Gradient Boosting Machine	37
Εικόνα 2: Υπολογισμός όρου ρύθμισης κατά την εκτέλεση του XGBoost	39
Εικόνα 3: Ταξινόμητης Random Forest	42

Ευρετήριο Πινάκων

Πίνακας 1: Κατανομή τιμής Bitcoin	58
Πίνακας 2: Βελτιστοποίηση υπερπαραμέτρων Holt	66
Πίνακας 3: Βελτιστοποίηση υπερπαραμέτρων Random Forest	66
Πίνακας 4: Αποτελέσματα αξιολόγησης ανεξάρτητων μεταβλητών για RF	67
Πίνακας 5: Βελτιστοποίηση υπερπαραμέτρων Gradient Boosting Machine	68
Πίνακας 6: Αποτελέσματα αξιολόγησης ανεξάρτητων μεταβλητών για GBM	69
Πίνακας 7: Βελτιστοποίηση υπερπαραμέτρων ARIMA	70
Πίνακας 8: Βελτιστοποίηση υπερπαραμέτρων XGBoost	70
Πίνακας 9: Αποτελέσματα αξιολόγησης ανεξάρτητων μεταβλητών για XGBoost	71
Πίνακας 10: Αποτελέσματα αξιολόγησης μοντέλων ανά ορίζοντα	72
Πίνακας 11: Αποτελέσματα Αξιολόγησης Πρόβλεψης Τιμής Bitcoin	72

1 Εισαγωγή

Η σύγχρονη εποχή χαρακτηρίζεται από διαρκείς, έντονες και καταγιστικές εξελίξεις τόσο σε οικονομικό όσο και σε κοινωνικό επίπεδο. Στο φαινόμενο αυτό έχει συμβάλλει το Διαδίκτυο και οι τεχνολογίες που το συνοδεύουν. Στο πλαίσιο αυτό όλοι οι τομείς δραστηριότητας πρέπει να αναπροσαρμόζονται και να ακολουθούν τις τάσεις προκειμένου να μην χαθούν στον χρόνο.

Από τις προαναφερόμενες αλλαγές δεν θα μπορούσε να μείνει ανεπηρέαστη η αγορά του χρήματος και των συναλλαγών. Οι ανάγκες της παγκοσμιοποιημένης ψηφιακής οικονομίας μετατόπισαν το ενδιαφέρον πέρα από τα συμβατικά νομίσματα όπως το ευρώ ή το δολάριο και ως εκ τούτου παρατηρείται μια μετατόπιση κεφαλαίου προς αυτήν την κατεύθυνση. Αρχικά, έκανε την εμφάνισή του το ηλεκτρονικό χρήμα, ακολούθησε το εικονικό και το ψηφιακό και τέλος έκαναν την εμφάνισή τους τα κρυπτονομίσματα.

Τα κρυπτονομίσματα, είναι ψηφιακά νομίσματα, τα οποία χρησιμοποιούνται ως μέσο ανταλλαγής, αποταμίευσης και επένδυσης. Παρότι εκ πρώτης όψεως φαίνεται να παρουσιάζουν αρκετά από τα χαρακτηριστικά του χρήματος δεν είναι ευρέως αποδεκτά ως τέτοιο. Σε ότι αφορά τις συναλλαγές, το βασικό χαρακτηριστικό της αγοράς των κρυπτονομισμάτων είναι το αποκεντρωμένο δίκτυο συναλλαγών, ένα δίκτυο απελευθερωμένο από οποιονδήποτε ενδιάμεσο αντίστοιχο των κεντρικών τραπεζών. Συνεπώς, η αγορά κρυπτονομισμάτων προσφέρει μεταξύ άλλων και ανωνυμία η οποία θα μπορούσε να χρησιμοποιηθεί σε αθέμιτες ενέργειες. Τα τελευταία αποτελούν λόγους για τους οποίους οι κεντρικές τράπεζες και οι εθνικές κυβερνήσεις εναντιώνονται στην αγορά των κρυπτονομισμάτων.

Η έννοια της επένδυσης αναφέρεται σε ένα αρχικό έξοδο το οποίο ακολουθείται από μελλοντικές εισροές. Τα κλασικά επενδυτικά μοντέλα που αναφέρονται σε πραγματικά περιουσιακά στοιχεία ή χρηματοοικονομικές επενδύσεις δεν είναι αρκετά για την ολοκλήρωση ενός επιτυχημένου χαρτοφυλακίου επενδύσεων. Συνεπώς, τα κρυπτονομίσματα αν και δεν συνιστώνται ως κύρια επένδυση λόγω της υψηλής μεταβλητότητας στην τιμή τους, φαίνεται να προτιμώνται από ένα τμήμα των επενδυτών ως ένα εναλλακτικό μέσο αποθήκευσης περιουσιακών στοιχείων και ως μέσο

συναλλαγών. Το πρώτο κρυπτονόμισμα που χάραξε τον δρόμο για τα παραπάνω δεν είναι άλλο από το Bitcoin.

Το Bitcoin, αν και ξεκίνησε ως ένα πείραμα του οικονομικού συστήματος στο οποίο αρχικά δεν δόθηκε αρκετή προσοχή ούτε από τους ίδιους τους θεσμικούς επενδυτές, έχει καταφέρει στα μόλις δέκα χρόνια ύπαρξής του να κερδίσει έδαφος, τόσο, που πλέον λογίζεται ως κίνδυνος για τα κλασικά μοντέλα.

«Τι πραγματικά είναι το Bitcoin και πώς φτάσαμε σήμερα να μιλάμε για αυτό;»

Η αρθρογραφία είναι αρκετά πρόσφατη και η βιβλιογραφία που υπάρχει περιορισμένη. Ωστόσο, αρκετές μελέτες έχουν πραγματοποιηθεί γύρω από τα κρυπτονομίσματα και ιδιαίτερα το Bitcoin. Το αποτέλεσμα των μέχρι τώρα ερευνών δεν έχει υπάρξει αρκετό ώστε η επιστημονική κοινότητα να καταλήξει σε πάγια συμπεράσματα σχετικά με τα κρυπτονομίσματα καθώς οι μελέτες βρίσκονται ακόμα σε πρώιμο στάδιο.

Η παρούσα εργασία έχει στόχο να βάλει το δικό της λιθαράκι στη μελέτη του Bitcoin. Πιο συγκεκριμένα, η έρευνα της εργασίας αυτής θα εστιάσει στην πρόβλεψη της τιμής του Bitcoin. Η πρόβλεψη θα αφορά τρεις χρονικούς ορίζοντες: τον βραχυπρόθεσμο, τον μεσοπρόθεσμο και τον μακροπρόθεσμο. Για την επίτευξη του στόχου αυτού θα εξετάσουμε πως ανταποκρίνονται τα μοντέλα πρόβλεψης στους ορίζοντες που αναφέραμε φτάνοντας στο στάδιο της αξιολόγησης για να αντιστοιχίσουμε τον ορίζοντα πρόβλεψης με το καλύτερο για αυτόν μοντέλο. Πιστεύουμε πως έτσι θα βοηθήσουμε όσους ενδιαφέρονται να επενδύσουν σε Bitcoin ή όσους ενδιαφέρονται να απαλλαχθούν από τα Bitcoin που έχουν στην κατοχή τους δίνοντάς τους πληροφορίες για τρεις εναλλακτικές περιόδους στις οποίες θα μπορεί να λάβει χώρα η συναλλαγή τους.

Ακολουθεί μια σύντομη περιγραφή της εργασίας και των κεφαλαίων που ακολουθούν.

Στο κεφάλαιο 2 περιγράφονται σημαντικές έννοιες και ορισμοί σχετικοί με το χρήμα, τις συναλλαγές, τα κρυπτονομίσματα. Επιπλέον, περιλαμβάνεται μια περιεκτική αναφορά στα όσα γνωρίζουμε μέχρι σήμερα για το Bitcoin. Στο κεφάλαιο 3 γίνεται μια σύντομη και περιεκτική αναφορά γύρω από τις έννοιες και τις τεχνικές προβλέψεων. Στο κεφάλαιο 4 παρατίθεται μια ανάλυση της τιμής του Bitcoin και των παραγόντων που την επηρεάζουν. Στο κεφάλαιο 5 γίνεται αναφορά στα επικρατέστερα μοντέλα πρόβλεψης που έχουν

χρησιμοποιηθεί και μελετηθεί στη βιβλιογραφία. Στο κεφάλαιο 6 παρουσιάζεται η ανάλυση της μελέτης της εργασίας. Αρχικά, θα γίνει παρουσίαση, ανάλυση και απεικόνιση των δεδομένων που χρησιμοποιήθηκαν. Στη συνέχεια, θα γίνει αναφορά στις μεθόδους που χρησιμοποιήθηκαν για την πρόβλεψη της τιμής του Bitcoin. Τέλος, στο κεφάλαιο 7 παρατίθενται τα αποτελέσματα, συμπεράσματα και προτάσεις.

2 Μελετώντας τα κρυπτονομίσματα

Απαντώντας με μια πιο γενική εξήγηση του τι είναι κρυπτονομίσματα θα λέγαμε ότι είναι ψηφιακά (ή εικονικά) νομίσματα τα οποία βασίζονται στην τεχνολογία της κρυπτογραφίας για να διασφαλίσουν την ασφαλή επικοινωνία. Αυτή η τεχνολογία είναι ο κύριος παράγοντας που καθιστά σχεδόν αδύνατη την παραχάραξη των νομισμάτων αυτών. Το πρώτο κρυπτονόμισμα (Bitcoin) δημιουργήθηκε το 2009 όταν ο Satoshi Nakamoto εκτέλεσε για πρώτη φορά στον υπολογιστή του το πρόγραμμα που είχε δημιουργήσει για την εξόρυξη του Bitcoin. Το πρόγραμμα αυτό δοκιμάζει τυχαίους αριθμούς μέχρι να βρει τον “σωστό” αριθμό με βάση την κρυπτογραφία.

Ουσιαστικά, το Bitcoin είναι ένα διαφανές και ανοιχτό προς όλους μητρώο στο οποίο μπορεί οποιοσδήποτε να δει τα υπόλοιπα και τις συναλλαγές που γίνονται ανά πάσα ώρα και στιγμή. Για να επιβεβαιωθούν οι συναλλαγές που γίνονται στο δίκτυο πρέπει να περάσουν από τυχαίους υπολογιστές μέσα στο δίκτυο οι οποίοι πραγματοποιούν ελέγχους εγκυρότητας της συναλλαγής. Κάθε συναλλαγή που επιβεβαιώνεται μπαίνει χρονικά στην σειρά σε ένα μπλοκ μαζί με όλες τις προηγούμενες συναλλαγές. Έτσι σχηματίζεται η λεγόμενη αλυσίδα με μπλοκ (blockchain). Με την συμπλήρωση ενός μπλοκ συναλλαγών εκδίδεται ένας συγκεκριμένος αριθμός Bitcoin ο οποίος μοιράζεται στους χρήστες των υπολογιστών που βοήθησαν στην εξακρίβωση των συναλλαγών σαν αμοιβή για την προσπάθειά τους. Η ανωνυμία που προκύπτει από την διαδικασία αυτή καθώς και ο μη έλεγχος από οποιαδήποτε αρχή κάνει τα κρυπτονομίσματα και ιδιαίτερα το Bitcoin πολύ ελκυστικά στους κερδοσκόπους. Τι όμως είναι αυτό που εναντιώνεται στους κερδοσκόπους ώστε να μην στρέφονται αποκλειστικά στα κρυπτονομίσματα για επένδυση; Η απάντηση βρίσκεται στην εξαιρετικά μεγάλη μεταβλητότητα της τιμής τους.

Τον όρο κρυπτονόμισμα τον συναντάμε σε συζητήσεις στους κύκλους των ακαδημαϊκών, των τραπεζών, των επιχειρήσεων, των επαγγελματιών κερδοσκόπων ακόμα και στους κύκλους των φίλων μας. Η έντονη αυτή παρουσία των κρυπτονομισμάτων καθώς και η καινοτομία που εισάγουν στις συναλλαγές αποτέλεσαν μεγάλο κίνητρο στην εκπόνηση της παρούσας μελέτης.

Προτού εισχωρήσουμε βαθύτερα σε αυτή και αναπτύξουμε τη μεθοδολογία πρόβλεψης της τιμής του Bitcoin, κρίνεται σκόπιμο να εισαχθούν ορισμένες θεωρητικές έννοιες για την καλύτερη κατανόηση του προβλήματος που θα ακολουθήσει. Αρχικά, θα γίνει αναφορά

σε έννοιες και ορισμούς σχετικές με τις συναλλαγές, το χρήμα και τις επενδύσεις. Θα ακολουθήσει μια θεωρητική αναφορά για τα κρυπτονομίσματα η οποία θα καταλήξει τελικά στο θεωρητικό πλαίσιο που περιγράφει το υπό μελέτη κρυπτονόμισμα, το Bitcoin.

2.1 Βασικές Έννοιες

2.1.1 Συναλλαγή (Transaction)

Με τον όρο συναλλαγή γίνεται αναφορά στην εναλλακτική δραστηριότητα μεταξύ δύο ή περισσότερων προσώπων. Δηλαδή, τη συμφωνία των μερών για ανταλλαγή αγαθών, υπηρεσιών ή χρηματικού αντιτίμου για την κάλυψη των αναγκών αμφοτέρων των μερών (Wikipedia, 2016).

2.1.2 Παραστατικό Χρήμα (Fiat money)

Ως χρήμα ορίζεται οποιοδήποτε ανταλλακτικό μέσο το οποίο είναι αποδεκτό από όλα τα μέλη μιας κοινωνίας και χρησιμοποιείται για το αντάλλαγμα οποιουδήποτε αγαθού (Σπύρου Σ., 2003). Ως χρήμα έχει επικρατήσει να λέμε τα χαρτονομίσματα και τα νομίσματα που εκδίδονται από το κράτος ωστόσο υπάρχουν και άλλες μορφές, όπως πιστωτικές κάρτες και επιταγές (ΕΚΤ, 2017). Το χρήμα λειτουργεί ως:

- Μέσο συναλλαγής για την αγορά αγαθών και υπηρεσιών
- Λογιστική μονάδα για τον καθορισμό των τιμών
- Μέσο αποθήκευσης αξίας για αποταμίευση

2.1.3 Αποταμίευση (Saving)

Είναι το μέρος του εισοδήματος μιας οικονομικής μονάδας (άτομο, επιχείρηση) που δεν δαπανάται για την κάλυψη των τρεχουσών αναγκών αλλά για μελλοντική χρήση (Σπύρου Σ., 2009).

2.1.4 Επένδυση (Investment)

Ορίζεται κάθε υλικό, διαρκές, παραγωγικό αγαθό το οποίο συμβάλλει στην αύξηση της παραγωγής από την μη κατανάλωση με τη χρήση του. Παράδειγμα επένδυσης αποτελεί η επέκταση ενός κτιρίου.

2.1.5 Χρηματιστήριο (Stock exchange)

Είναι μια επίσημη, οργανωμένη και ελεγχόμενη αγορά κινητών αξιών οι τιμές των οποίων προσδιορίζονται από τις δυνάμεις της προσφοράς και της ζήτησης. Αποτελεί μια εναλλακτική μορφή τοποθέτησης των χρημάτων που έχουν αποταμιευτεί. Για παράδειγμα, ο κάτοχος μιας μετοχής προσπαθεί να πουλήσει στη μεγαλύτερη δυνατή τιμή ενώ ο ενδιαφερόμενος να αγοράσει στη χαμηλότερη δυνατή τιμή. Έτσι, για μία δεδομένη χρονική στιγμή, η προσφορά και η ζήτηση ισορροπούν και διαμορφώνουν την τιμή της μετοχής.

2.1.6 Ηλεκτρονικό χρήμα (Electronic money)

Με τον όρο ηλεκτρονικό χρήμα περιγράφουμε κάθε μορφή μεταφοράς κεφαλαίου μεταξύ δύο ή περισσότερων μερών που γίνεται με ψηφιακό τρόπο και χωρίς τη μεσολάβηση υλικού μέσου.

2.1.7 Ψηφιακό χρήμα (Digital currency)

Ο όρος ψηφιακό και εικονικό χρησιμοποιούνται εναλλακτικά. Τα ψηφιακά νομίσματα αποτελούν έννοια ευρύτερη των εικονικών νομισμάτων. Παραδείγματα ψηφιακών νομισμάτων εκτός από τα κρυπτονομίσματα είναι τα αεροπορικά μίλια και οι μάρκες ηλεκτρονικών παιχνιδιών.

2.1.8 Εικονικό νόμισμα (Virtual currency)

Ως εικονικό νόμισμα νοείται μια ψηφιακή αναπαράσταση αξίας που ούτε εκδίδεται από κεντρική τράπεζα ή δημόσια αρχή ούτε συνδέεται κατ' ανάγκη με παραστατικό νόμισμα. Γίνεται όμως αποδεκτή από φυσικά ή νομικά πρόσωπα ως μέσο πληρωμής και μπορεί να μεταφέρεται, να αποθηκεύεται ή να διακινείται ηλεκτρονικά (ΕΚΤ, 2015). Τα εικονικά νομίσματα δεν αποτελούν νομίμως αναγνωρισμένα νομίσματα και έτσι θα πρέπει να χαρακτηρίζονται μόνο ως μέσα ανταλλαγών και όχι ως μέσα πληρωμών. Ορισμένα από τα δημοφιλέστερα εικονικά νομίσματα είναι το Bitcoin και το Ethereum, τα γνωστά μας κρυπτονομίσματα.

2.1.9 Εναλλακτικό νόμισμα (Alternative currency)

Ο όρος αυτός αναφέρεται σε νομίσματα τα οποία μπορεί να χρησιμοποιηθούν αντί του κυρίαρχου εθνικού νομίσματος υπό ορισμένες συνθήκες και για συγκεκριμένες συναλλαγές. Μπορούν να δημιουργηθούν από εταιρείες, ιδιώτες, ιδιωτικούς ή δημόσιους οργανισμούς. Ένα παράδειγμα είναι η επιβράβευση με μίλια από τις αεροπορικές εταιρείες.

2.2 Κρυπτονομίσματα

Με τον όρο κρυπτονομίσματα αναφερόμαστε σε ψηφιακά (ή εναλλακτικά) νομίσματα τα οποία βασίζονται σε αποκεντρωμένο λογιστικό σύστημα όπου οι συναλλαγές επαληθεύονται και εκτελούνται μέσω του δικτύου των χρηστών που το αποτελούν. Τα κρυπτονομίσματα δεν ορίζονται ως χρήμα εφόσον δεν πληρούν όλα τα χαρακτηριστικά του χρήματος. Κύριο χαρακτηριστικό τους είναι η χρήση μιας αλυσίδας “μπλοκ” η οποία αποθηκεύει τις συναλλαγές μέσω της τεχνολογίας της κρυπτογράφησης. Η κρυπτογραφία χρησιμοποιεί τεχνολογία η οποία διασφαλίζει ότι η πληροφορία είναι αδιάβλητη από οποιονδήποτε μη μέτοχο στη δημιουργία και τη μεταφορά του.

Με τον όρο κρυπτονομίσμα (Cryptocurrency) περιγράφεται μια νέα μορφή χρήματος η οποία χρησιμοποιεί την κρυπτογραφία τόσο για τη δημιουργία του όσο και τις συναλλαγές με αυτό. Το 2008, ο δημιουργός του Bitcoin με το ψευδώνυμο Satoshi Nakamoto δημοσίευσε στο λεγόμενο white paper τις πρώτες προδιαγραφές του Bitcoin το οποίο λίγο αργότερα κατατέθηκε δημόσια (Nakamoto, 2008). Κάπως έτσι ορίστηκαν οι βάσεις για τη δημιουργία ενός χρηματικού εργαλείου ανθεκτικού και αδιάβλητου. Αυτός είναι ένας από τους λόγους για τους οποίους τα κρυπτονομίσματα κερδίζουν όλο και περισσότερο έδαφος.

Σήμερα παρατηρείται μια ραγδαία αύξηση στη δημιουργία κρυπτονομισμάτων. Αυτό ξεκίνησε μετά την εξαιρετικά μεγάλη αύξηση της τιμής του Bitcoin το 2013 και ενέπνευσε πολλούς προγραμματιστές να αναζητήσουν την τύχη τους στη δημιουργία κρυπτονομισμάτων. Είναι αλήθεια ότι δεν λαμβάνουν όλα τα κρυπτονομίσματα την ίδια απήχηση με αποτέλεσμα αρκετά από αυτά να οδηγούνται στον θάνατο. Ωστόσο, αρκετά είναι αυτά τα οποία κατάφεραν να αυξήσουν την αξία τους και να αντέξουν τις συνθήκες της νέας, παγκόσμιας, ανταγωνιστικής και απολύτως ελεύθερης αυτής αγοράς. Η εξέλιξη της πορείας τους αποτελεί αντικείμενο προς μελέτη.

Ορισμένα από τα κύρια χαρακτηριστικά των κρυπτονομισμάτων είναι τα ακόλουθα:

- Χρησιμοποιούν αλγορίθμους ανοιχτού κώδικα οι οποίοι είναι προσβάσιμοι για μελέτη και επέκταση από όλους. Η παραπάνω δυνατότητα επιτρέπει σε οποιονδήποτε έχει τη γνώση και την υπολογιστική ισχύ να εκδώσει νομίσματα (Mining)
- Ο όγκος των νομισμάτων είναι εκ των προτέρων γνωστός και περιορισμένος

- Χρησιμοποιούν την τεχνολογία Blockchain για την επιβεβαίωση των συναλλαγών με αποτέλεσμα να μην ρυθμίζονται αλλά ούτε να ελέγχονται από κάποια κεντρική αρχή (κράτος ή τράπεζα)
- Διατηρούν ανωνυμία των συναλλαγών καθώς δεν είναι απαραίτητο να υποβάλλονται προσωπικά στοιχεία κατά την πραγματοποίησή τους
- Μπορούν να χρησιμοποιηθούν από οποιονδήποτε, οπουδήποτε μέσω μη ανιχνεύσιμων συναλλαγών
- Οι συναλλαγές οποιουδήποτε όγκου δεν υπόκεινται σχεδόν σε κανένα κόστος και τέλος
- Μπορούν να μετατραπούν σε φυσικό νόμισμα

2.2.1 Βασικές έννοιες κρυπτονομισμάτων

Λογιστικό Βιβλίο (Block)

Τα μπλοκ ή λογιστικό βιβλίο είναι ένα αρχείο σε μια σειρά από κοινοποιημένες και επικυρωμένες συναλλαγές που έχουν γίνει στο δίκτυο. Στο block περιέχονται δεδομένα που σχετίζονται τόσο με το δίκτυο όσο και με το περιεχόμενο της κάθε συναλλαγής. Τα blocks προστίθενται σε μία αλυσίδα μπλοκ (blockchain) με συχνότητα 10 περίπου λεπτών.

Αλυσίδα Μπλοκ (Blockchain)

Μπορεί να χαρακτηριστεί ως ένα δημόσιο μητρώο συναλλαγών που περιέχει δημόσια αρχεία – blocks με χρονολογική σειρά. Για την επιβεβαίωση μιας συναλλαγής με την τεχνολογία blockchain αρχικά η συναλλαγή μετατρέπεται σε μπλοκ, κωδικοποιείται και κρυπτογραφείται, έπειτα μοιράζεται μεταξύ όλων των χρηστών του δικτύου και μόνο όταν γίνει αποδεκτό από όλο το δίκτυο (με κριτήρια στα οποία δεν θα επεκταθούμε) εισάγεται στην αλυσίδα συναλλαγών με τρόπο αδιάβλητο.

Ανταλλακτήριο Κρυπτονομισμάτων (Cryptocurrency exchange)

Τα ανταλλακτήρια κρυπτονομισμάτων είναι ιστοσελίδες ή υπηρεσίες που επιτρέπουν την ανταλλαγή ψηφιακών κρυπτο-αγαθών με κρυπτονομίσματα και αντιστρόφως. Δύναται επίσης η ανταλλαγή παραστατικού χρήματος με κρυπτονομίσματα βάσει της ισοτιμίας που διαμορφώνεται από την αγορά.

Πορτοφόλι Κρυπτονομισμάτων (Cryptocurrency wallet)

Το πορτοφόλι κρυπτονομισμάτων είναι ένα πρόγραμμα λογισμικού που αποθηκεύει ιδιωτικά και δημόσια κλειδιά και αλληλοεπιδρά με διάφορες αλυσίδες από blocks (blockchains) για να επιτρέψει στον κάτοχό του να πραγματοποιεί συναλλαγές κρυπτονομισμάτων αλλά και να επιβεβαιώνει το υπόλοιπο που έχει στην κατοχή του.

Σε αντίθεση με ένα παραδοσιακό πορτοφόλι, το πορτοφόλι κρυπτονομισμάτων δεν αποθηκεύει νομίσματα αφού από τη φύση τους αυτά δεν έχουν φυσική υπόσταση και άρα δεν έχουν γεωγραφική τοποθεσία. Αντί νομισμάτων, το πορτοφόλι κρυπτονομισμάτων περιέχει εγγραφές συναλλαγών που είναι αποθηκευμένες στο blockchain. Προκειμένου οι συναλλαγές να θεωρούνται έγκυρες και η κυριότητα να καταγράφεται με ακρίβεια, το πορτοφόλι κρυπτονομισμάτων αποθηκεύει και δημόσια και ιδιωτικά κλειδιά λογισμικού τα οποία είναι απαραίτητα στη διαδικασία συναλλαγών.

Αποκέντρωση (Decentralization)

Αναφέρεται στον τρόπο με τον οποίο είναι δομημένο ένα δίκτυο και στον τρόπο με τον οποίο κατανέμονται και διακινούνται τα δεδομένα μεταξύ των κόμβων του δικτύου.

Εξόρυξη (Mining)

Εξαιτίας της φύσης των κρυπτονομισμάτων, η επιβεβαίωση των συναλλαγών απαιτεί ένα τεράστιο ποσό υπολογιστικής ισχύος και εξειδικευμένου υλισμικού. Ως αντάλλαγμα για την προσφερόμενη υπολογιστική ισχύ, οι κάτοχοι των συστημάτων που επιβεβαιώνουν (και συνεπώς εγκρίνουν) μια συναλλαγή πληρώνονται σε κρυπτονομίσματα. Αυτή η διαδικασία ονομάζεται εξόρυξη. Η εξόρυξη είναι η διεργασία που δημιουργεί το υλισμικό υπολογιστή (computer hardware) για να κάνει μαθηματικούς υπολογισμούς ώστε το δίκτυο ενός κρυπτονομίσματος να επιβεβαιώνει τις συναλλαγές και να διατηρεί ασφάλεια μεταξύ τους. Ως αμοιβή για τις υπηρεσίες τους, όσοι πραγματοποιούν εξόρυξη μπορούν να συλλέγουν τα τέλη συναλλαγών για τις συναλλαγές που επιβεβαιώνουν, μαζί με τα πρόσφατα δημιουργημένα κρυπτονομίσματα. Η εξόρυξη είναι μια εξειδικευμένη και ανταγωνιστική αγορά όπου οι αμοιβές διανέμονται ανάλογα με τον συνολικό παραγόμενο υπολογισμό.

Κόμβος (Node)

Ο κόμβος είναι μια υπολογιστική μονάδα στο blockchain. Είναι υπεύθυνος για την πιστοποίηση των συναλλαγών αλλά και για τη διατήρηση και επικαιροποίησή της διανεμημένης καθολικής αλυσίδας συναλλαγών.

Ομότιμη σύνδεση (Peer-to-Peer - P2P)

Η ομότιμη σύνδεση ή P2P αναφέρεται στην επικοινωνία ανάμεσα σε δύο ή περισσότερα μέλη ενός έντονα διασυνδεδεμένου δικτύου. Οι συμμετέχοντες ενός τέτοιου δικτύου επικοινωνούν απευθείας μεταξύ τους, μέσα από ένα μοναδικό σημείο διαμεσολάβησης. Η ομότιμη σύνδεση αναφέρεται σε συστήματα που λειτουργούν όπως μια οργανωμένη συλλογικότητα, επιτρέποντας στο κάθε άτομο να αλληλοεπιδρά άμεσα με τα άλλα.

Απόδειξη εργασίας (Proof-of-work)

Το Proof-of-work είναι ο αρχικός αλγόριθμος συναίνεσης ή η απόδειξη εργασίας σε ένα δίκτυο blockchain. Ο αλγόριθμος αυτός χρησιμοποιείται για την επιβεβαίωση συναλλαγών και την παραγωγή νέων blocks στην αλυσίδα. Με το Proof-of-work, οι υπολογιστικές μηχανές που πραγματοποιούν εξόρυξη ανταγωνίζονται μεταξύ τους για να επιβεβαιώσουν συναλλαγές στο δίκτυο και να ανταμειφθούν.

Απόδειξη κυριότητας (Proof-of-stake)

Η απόδειξη κυριότητας (Proof-of-stake) είναι ένας τύπος αλγόριθμου συναίνεσης με τον οποίο ένα δίκτυο blockchain στοχεύει στην επίτευξη κατανεμημένης συναίνεσης. Στα κρυπτονομίσματα που βασίζονται σε απόδειξη κυριότητας, ο δημιουργός του επόμενου block επιλέγεται μέσω διαφόρων συνδυασμών τυχαίας επιλογής και πλούτου ή ηλικίας.

2.2.2 Τα κυριότερα κρυπτονομίσματα εκτός από το Bitcoin

Ethereum

Το (Ethereum, 2020) είναι μια πλατφόρμα ανοιχτού κώδικα δεύτερης γενιάς η οποία έχει το δικό της blockchain. Το Ethereum είναι το κρυπτονόμισμα με την αμέσως μεγαλύτερη κεφαλαιοποίηση μετά το Bitcoin, με το οποίο παρουσιάζει αρκετές ομοιότητες. Πρόκειται για ένα εξ ολοκλήρου ψηφιακό νόμισμα η διαχείριση του οποίου γίνεται αποκεντρωμένα, γεγονός που ενισχύει την αξιοπιστία. Η ποσότητα του είναι περιορισμένη. Πιο συγκεκριμένα το 2014 προσφέρθηκαν 66 εκατομμύρια μονάδες Ethereum σε όσους έλαβαν μέρος στην αγοραπωλησία, ενώ 12 εκατομμύρια μονάδες δόθηκαν στους ιδρυτές του κρυπτονομίσματος. Είναι εμφανές ότι παρά το αποκεντρωμένο χαρακτήρα του ένα μεγάλο ποσοστό των διαθέσιμων μονάδων ανήκει στους ιδρυτές του πράγμα που σημαίνει ότι έστω και κατ' ελάχιστον οι τελευταίοι θα μπορούσαν να επηρεάσουν την τιμή του.

Στα τέλη Ιουλίου 2014 υπήρξε η πρώτη καταγεγραμμένη αποτίμηση της αξίας του Ethereum όταν μια μονάδα του ήταν ίση με 0.235 ως 0.316 δολάρια ενώ τον Αύγουστο του 2015 σημειώθηκε η πρώτη διαπραγμάτευση. Αντίστοιχα με το Bitcoin, η αξία του Ethereum αυξήθηκε, φτάνοντας στα τέλη του 2016 να είναι 40 φορές μεγαλύτερη από την αρχική. Από το σημείο αυτό και έπειτα ο ταχύς ρυθμός αύξησης έχει περιοριστεί.

Το Ethereum δημιουργείται τόσο μέσω εξόρυξης όσο και μέσω μίας πλατφόρμας έξυπνου συμβολαίου. Η δυσκολία πρόβλεψης της ζήτησης για τα τελευταία κάνει την τιμή του Ethereum ακόμα πιο ευμετάβλητη με αποτέλεσμα να το καθιστά ακατάλληλο και μη ελκυστικό ως μέσο αποθήκευσης αξίας.

Litecoin (LTC)

Το (Litecoin, 2020) εμφανίστηκε για πρώτη φορά τον Οκτώβριο του 2011 ως αποτέλεσμα μικρών μεταβολών στο λογισμικό του Bitcoin για αυτό και αναφέρεται ως συμπληρωματικό κρυπτοαγαθό του Bitcoin. Είναι ένα ακόμα ψηφιακό νόμισμα το οποίο επιτρέπει άμεσες και χωρίς κόστος συναλλαγές οπουδήποτε στον κόσμο. Το blockchain του Litecoin είναι ικανό να χειρίζεται μεγαλύτερο όγκο συναλλαγών από αυτό του Bitcoin. Διαθέτει ταχύτερους χρόνους επιβεβαίωσης συναλλαγών και βελτιωμένη απόδοση αποθήκευσης.

Η παραγωγή του Litecoin είναι επίσης προκαθορισμένη, ωστόσο, είναι 4 φορές μεγαλύτερη από αυτή του Bitcoin και ανέρχεται σε 84 εκατομμύρια μονάδες. Η πρώτη καταγεγραμμένη τιμή του Litecoin υπήρξε τον Ιούλιο του 2012 με αξία 0.035 δολάρια ανά μονάδα και έκτοτε οι αποδόσεις του έχουν υπερ-εκατονταπλασιαστεί. Το δίκτυο του Litecoin είναι πολύ λιγότερο ασφαλές και ανθεκτικό σε επιθέσεις με αποτέλεσμα να μη θεωρείται πολύ πιο αξιόπιστο. Παρ' όλα αυτά θεωρείται ότι είναι μία πολύ καλή επενδυτική επιλογή η οποία όμως πρέπει να καταφέρει γιγαντιαία βήματα ώσπου να φτάσει αντίστοιχα υψηλή ταχύτητα συναλλαγών, ισχύ επεξεργασίας και ισχύ στην αγορά με το Bitcoin.

Monero (XMR)

Το (Monero, 2020) είναι ένα κρυπτονόμισμα το οποίο διαφοροποιείται από τα άλλα ως προς τον τρόπο με τον οποίο επαληθεύονται οι συναλλαγές δίνοντας έμφαση σε ιδιωτικές και ανθεκτικές στη λογοκρισία συναλλαγές. Ενώ στα υπόλοιπα κρυπτονομίσματα τα blockchain είναι εμφανή σε όλους, δηλαδή οι συναλλαγές είναι ανοιχτά επαληθεύσιμες και

ανιχνεύσιμες από οποιονδήποτε το Monero χρησιμοποιεί μεθόδους κρυπτογράφησης για να διασφαλίσει την ανωνυμία της διεύθυνσης αποστολής και λήψης και τα ποσά των συναλλαγών. Μιλάμε για ένα πάντα ενεργό απόρρητο διασφάλισης συναλλαγών.

Μερικά από τα πλεονεκτήματα που προβάλλονται για την προτίμησή του είναι:

- Ασφάλεια από Capital controls
- Μικρός χρόνος αναμονής επαλήθευσης συναλλαγών
- Γρήγορες και αξιόπιστες συναλλαγές από οπουδήποτε στον κόσμο οποιαδήποτε στιγμή

Dash

Το (Dash, 2020) είναι ένα ακόμα κρυπτονόμισμα, το οποίο ξεκίνησε ως εναλλακτικό προϊόν από τον πηγαίο κώδικα του Bitcoin. Στη διάρκεια της ιστορίας του από το 2014 και μέχρι να λάβει το τελικό του όνομα συναντάται επίσης με τα ονόματα Xcoin και Darkcoin. Κύριο χαρακτηριστικό του είναι πως οι συναλλαγές εντός του δικτύου του μπορούν να επαληθευτούν άμεσα. Επιπλέον, και σε σχέση με το κρυπτονόμισμα αναφοράς Bitcoin, η διαχειριστική αρχή είναι ένας αποκεντρωμένος αυτόνομος οργανισμός.

Μερικά από τα μοναδικά χαρακτηριστικά του εκτός των παραπάνω είναι τα ακόλουθα:

- Το πρωτόκολλο blockchain του Dash είναι αυτοχρηματοδοτούμενο και αυτοδιαχειριζόμενο
- Διαθέτει διευθύνσεις φιλικές στην ανάγνωση για καλύτερη κατανόηση ως προς τον προορισμό της κάθε συναλλαγής
- Διαθέτει ιδιωτικές συναλλαγές, οι οποίες εκτός από ανώνυμες είναι και μη ανιχνεύσιμες
- Οι δημιουργοί αλλά και οι διαφημιστές του Dash εξασφαλίζουν πληρωμή μέσω των επιτυχών συναλλαγών που πραγματοποιούνται

2.3 Επενδύσεις

Επένδυση για τις επιχειρήσεις “καλείται η δέσμευση κεφαλαίου μιας οικονομικής μονάδας για την απόκτηση περιουσιακών στοιχείων, τα οποία προορίζονται να παραμείνουν στην επιχείρηση για μεγάλο χρονικό διάστημα και συνδέονται με την ομαλή λειτουργία αυτής”

(Αλεξανδρίδης, 2005). Επένδυση είναι επίσης οποιαδήποτε οικονομική δραστηριότητα που χρησιμοποιεί πόρους για την παραγωγή αγαθών και υπηρεσιών. Μπορεί να είναι κερδοσκοπικού ή μη κερδοσκοπικού χαρακτήρα όπως, για παράδειγμα, τοποθετήσεις σε μετοχές ή δέσμευση κεφαλαίου για την αγορά κατοικίας. Οι επενδύσεις κατηγοριοποιούνται σε ιδρυτικές επενδύσεις (επενδύσεις σε μία νέα εταιρεία), τρέχουσες επενδύσεις (αντικατάστασης εξοπλισμού) κ.ά.

Ο άνθρωπος αρέσκεται να επενδύει χρήματα και να λαμβάνει ρίσκο με σκοπό το κέρδος. Η αγορά των κρυπτονομισμάτων έχει έρθει για να συμπληρώσει τις κλασικές μορφές επενδύσεων και να ικανοποιήσει στο μέγιστο αυτήν την ανάγκη. Τα κρυπτονομίσματα έχουν θεωρηθεί ως ένας νέος τρόπος δημιουργίας και αποθήκευσης χρημάτων και μάλιστα αρκετά προσοδοφόρος.

Πολλοί λόγοι οδηγούν όσους θέλουν να επενδύσουν στα κρυπτονομίσματα. Αρχικά, ο επενδυτής μπορεί να επενδύσει σε υποδιαίρεση της μονάδας και όχι αναγκαστικά σε ολόκληρη τη μονάδα ενός κρυπτονομίσματος. Η έλλειψη μηχανισμού ελέγχου δίνει τη δυνατότητα η αγορά να είναι ανοιχτή 24/7 σε αντίθεση με τις κλασικές αγορές και να διαμορφώνεται η τιμή αποκλειστικά και μόνο από τον νόμο της προσφοράς και της ζήτησης την εκάστοτε στιγμή. Επιπλέον, σύμφωνα με τους (Liu & Tsyvinski, 2018) η επένδυση σε κρυπτονομίσματα επιλέγεται εναλλακτικά για να αντισταθμιστούν οι ζημιές από τις παραδοσιακές επενδύσεις αφού τα χαρακτηριστικά της αγοράς αυτής είναι πολύ διαφορετικά από τα αντίστοιχα χαρακτηριστικά άλλων περιουσιακών στοιχείων όπως μετοχές και νομίσματα. Επηρεάζονται δηλαδή από παράγοντες διαφορετικούς από αυτούς των μετοχών ή των αποδόσεων των νομισμάτων.

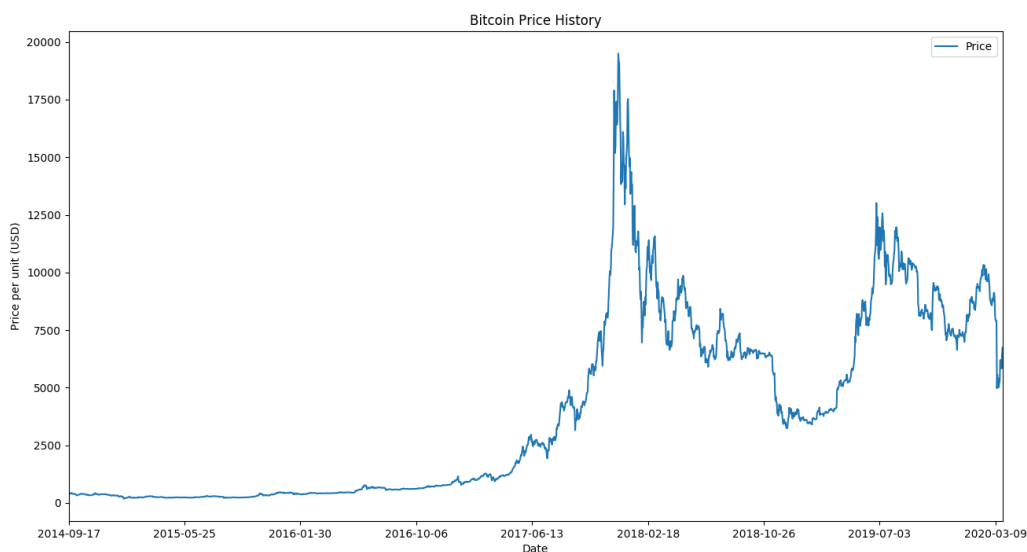
2.4 Bitcoin

Τον Οκτώβριο του 2008 ο δημιουργός του πρώτου κρυπτονομίσματος όρισε το Bitcoin ως μία P2P έκδοση των ηλεκτρονικών συναλλαγών που επιτρέπει στις πληρωμές να αποστέλλονται απευθείας από το ένα μέρος στο άλλο χωρίς να περάσουν από ένα χρηματοπιστωτικό ίδρυμα. Στόχος του ήταν να φέρει μια καινοτόμα λύση στο πρόβλημα των διπλών δαπανών κατά την εκτέλεση συναλλαγών με τη διαμεσολάβηση χρηματοπιστωτικών ιδρυμάτων (Nakamoto, 2008).

Στο ίδιο άρθρο ο δημιουργός του Bitcoin εισήγαγε την έννοια του Blockchain ως μια μέθοδο επαλήθευσης των συναλλαγών. Πρότεινε η κάθε χρονική στιγμή κατά την οποία πραγματοποιείται μια συναλλαγή να καταγράφεται σε μια αλυσίδα με συνεχή και αδιάβλητο τρόπο. Η αλυσίδα αυτή μπορεί οποιαδήποτε στιγμή να επαληθευτεί από όλους τους συμμετέχοντες στο ομότιμο δίκτυο του Bitcoin και είναι μαθηματικά αδύνατο να παραποιηθεί.

Τέλος το ανώτατο πλήθος των Bitcoin σε κυκλοφορία είναι προκαθορισμένο, ενώ η παραγωγή τους γίνεται μέσω της διαδικασίας εξόρυξης. Κατά τη διαδικασία αυτή, υπολογιστικοί κόμβοι μεγάλης ισχύος πραγματοποιούν σύνθετους μαθηματικούς υπολογισμούς απαραίτητους για τη διαδικασία παραγωγής Bitcoin. Οι ιδιοκτήτες τους αμείβονται μέσω του πρωτοκόλλου με τρόπο ανάλογο της συνολικής παραγόμενης υπολογιστικής ισχύος.

2.4.1 Ορόσημα



Γράφημα 1: Ιστορικότητα τιμής Bitcoin

Όπως έχει αναφερθεί σε προηγούμενο σημείο της εργασίας η έννοια Bitcoin εμφανίστηκε για πρώτη φορά το 2008. Ο αρχικός σκοπός του ήταν να χρησιμοποιηθεί ως υποκατάστατο του χρήματος και ως τέτοιο χρησιμοποιήθηκε για πρώτη φορά (αυτή ήταν η πρώτη επίσημη καταγραφή) τον Μάιο του 2010 όπου με 10,000 Bitcoin αγοράστηκαν 2 πίτσες οι οποίες σήμερα θα άξιζαν περίπου 792 εκατομμύρια δολάρια. Δύο μήνες

αργότερα, λειτούργησε το πρώτο ανταλλακτήριο Bitcoin (Mt. Gox) το οποίο έμελλε να πτωχεύσει το 2014 (Brandvold κ.ά., 2015). Η αξία του Bitcoin άγγιξε για πρώτη φορά το ένα δολάριο τον Απρίλιο του 2011 ενώ τον Φεβρουάριο του 2013 ξεπέρασε την ουγκιά ασημιού. Το ίδιο έτος είχαμε το πρώτο ATM Bitcoin. Η πρώτη ενέργεια που έδωσε τη δυνατότητα σε όσους αντιτίθενται στο Bitcoin να στηρίζουν περισσότερο την απόρριψη του πραγματοποιήθηκε τον Φεβρουάριο του 2014 με την πρώτη επίθεση hacker κατά την οποία κλάπηκαν 850,000 Bitcoin αξίας μισού δισεκατομμυρίου δολαρίων και 28 εκατομμύρια μετρητά (Lea, 2016). Τον Μάρτιο του 2017 η αξία του ξεπέρασε αυτή της μίας ουγκιάς χρυσού ενώ τον Δεκέμβριο του ίδιου έτους έφτασε στο υψηλότερο σημείο που έχει φτάσει ποτέ (19,666 δολάρια ανά μονάδα) σημειώνοντας αύξηση 1,300% μέσα σε ένα έτος, ενώ η πορεία του ακολούθησε πτωτική πορεία μέχρι τον Δεκέμβριο του 2018 που κατέγραψε την χαμηλότερη τιμή μετά το προηγούμενο πικ με αξία περίπου 3,236 δολάρια ανά μονάδα. Τον Ιούλιο του 2019 ανέβηκε στο υψηλότερο σημείο του έτους με αξία περίπου 11,959 δολάρια ανά μονάδα. Έκτοτε σημειώνει πτωτική πορεία μέχρι τον Μάρτιο του 2020, εν μέσω πανδημίας, με την τιμή του να έχει πέσει στα 5,200 δολάρια ανά μονάδα και μέχρι τη στιγμή σύνταξης αυτής της εργασίας η τιμή του είναι 9,300 δολάρια ανά μονάδα.

2.4.2 Βασικά χαρακτηριστικά

Τα βασικά χαρακτηριστικά του Bitcoin είναι τα παρακάτω:

- Αποτελεί αποκεντρωμένο ψηφιακό νόμισμα ως προς τον έλεγχο. Αυτό σημαίνει ότι χρησιμοποιεί τεχνολογία μεταξύ ομότιμων κόμβων (peer-to-peer) χωρίς κεντρική εξουσία ή παρουσία τραπεζών. Η διαχείριση συναλλαγών και η έκδοση των Bitcoin διεξάγεται συλλογικά από το δίκτυο.
- Υλοποιείται με ανοικτό κώδικα ή αλλιώς Open source. Ο όρος αυτός αναφέρεται στο γεγονός ότι ο πηγαίος κώδικας του Bitcoin είναι προσβάσιμος από όλους δίνοντας τους παράλληλα τη δυνατότητα να συμμετέχουν στην ανάπτυξη και τη βελτίωσή του.
- Διαθέτει αποκεντρωμένο P2P δίκτυο. Αυτό σημαίνει ότι δεν υπάρχει κάποιος κεντρικός Server στον οποίο συνδέονται τα μέλη του δικτύου. Αντιθέτως όλα τα μέλη του δικτύου είναι ομότιμα και μπορούν να πραγματοποιήσουν αποκλειστική επικοινωνία μεταξύ τους.

- Πραγματοποιεί χρήση πρωτοκόλλου Blockchain για την πραγματοποίηση και επαλήθευση των συναλλαγών, μέσω του οποίου διασφαλίζεται η εγκυρότητα και η καθολική αποδοχή των συναλλαγών.

2.4.3 Πλεονεκτήματα

Μερικά από τα πλεονεκτήματα του Bitcoin είναι τα παρακάτω:

- Το μεγαλύτερο όφελος που ενδέχεται να αποκομίσει η οικονομία από το Bitcoin και τα κρυπτονομίσματα γενικότερα είναι η υποκείμενη τεχνολογία Blockchain, εφόσον αυτή χρησιμοποιηθεί όχι για την υποκατάσταση των κατεστημένων νομισμάτων, αλλά για άλλες εφαρμογές
- Μπορεί να χρησιμοποιηθεί σαν αποθήκη αξίας
- Αποτελεί επενδυτικό στοιχείο υψηλού κέρδους
- Το κόστος των συναλλαγών σε Bitcoin είναι πολύ χαμηλό καθώς δεν υπάρχουν μεσάζοντες άρα και ύπαρξη προμηθειών
- Οι συναλλαγές γίνονται σχεδόν σε μηδενικό χρόνο εν αντιθέσει με τις συναλλαγές των τραπεζών όπως εμβάσματα στο εξωτερικό τα οποία μπορεί να χρειαστούν κάποιες μέρες για τη διεκπεραίωση της συναλλαγής
- Το Bitcoin εξασφαλίζει ανωνυμία συναλλαγών σε σχέση με τα τραπεζικά συστήματα

2.4.4 Μειονεκτήματα

Τέλος, παρακάτω καταγράφονται τα κύρια μειονεκτήματα του Bitcoin:

- Η τιμή μονάδας παρουσιάζει υψηλή μεταβλητότητα
- Το Bitcoin παρουσιάζει υψηλό ρίσκο για τους επενδυτές λόγω της αδιαφάνειας που το χαρακτηρίζει και της έλλειψης ρυθμιστικού πλαισίου
- Αποτελεί πρόσφορο έδαφος για χρηματοδότηση οργανωμένου εγκλήματος και τρομοκρατίας, εξαιτίας της ανωνυμίας που διασφαλίζει
- Η πιστοποίηση μίας και μόνης συναλλαγής στο σύστημα Bitcoin με τον τρόπο που περιεγράφηκε παραπάνω συνεπάγεται την κατανάλωση δυσανάλογα μεγάλων ποσών ηλεκτρικής ενέργειας. Για να τελειοποιηθεί μια συναλλαγή, δαπανάται ηλεκτρική ενέργεια ισοδύναμη με εκατοντάδες (Mora κ.ά., 2018).

- Από τη σκοπιά του μεμονωμένου επενδυτή, οι τοποθετήσεις σε Bitcoin ενέχουν ιδιαίτερος αυξημένο κίνδυνο απώλειας του κεφαλαίου του καθώς δεν υπάρχει ασφάλεια των καταθέσεων όπως στα τραπεζικά συστήματα

2.4.5 Το Bitcoin ως επενδυτικό αγαθό

Το Bitcoin έκανε την πρώτη του εμφάνιση το 2009 μετά το ξέσπασμα της παγκόσμιας οικονομικής κρίσης του 2008. Αυτό το χρονικό ορόσημο θα μπορούσε να υποστηρίξει κανείς ότι συντέλεσε καθοριστικά στην αποδοχή παντός είδους νέων αγαθών και υπηρεσιών τα οποία ήταν ανεξάρτητα από τα τεκταινόμενα της περιόδου και μάλιστα θα μπορούσαν να αποτελέσουν εναλλακτική επενδυτική επιλογή για όσους κινδύνευαν να ζημιωθούν από την αβεβαιότητα των παραδοσιακών επενδυτικών αγορών.

Τα πρώτα έτη της εμφάνισής του όποιος επένδυε είχε είτε συνειδητά αναλάβει έναν μεγάλο κίνδυνο είτε διέθετε πλήρη άγνοια. Πλέον, 12 χρόνια μετά, όλα δείχνουν πως το Bitcoin ήρθε για να μείνει. Το ερώτημα είναι αν θα καταφέρει να αλλάξει ριζικά τον τρόπο που λειτουργεί σήμερα το χρηματοπιστωτικό σύστημα ή θα παραμείνει μια ανεξέλεγκτη ανεξάρτητη παράλληλη αγορά.

3 Τεχνικές Προβλέψεων

Αναμφισβήτητα, η ανάγκη της πρόβλεψης στις μέρες μας, εντοπίζεται σε όλες τις δραστηριότητες, ιδιαίτερα δε στις επιχειρήσεις οι οποίες λειτουργούν σε ένα περιβάλλον ασάφειας και αβεβαιότητας. Τόσο το άτομο όσο και η επιχείρηση χρησιμοποιούν άμεσες και έμμεσες προβλέψεις προσπαθώντας να σχεδιάσουν το μέλλον. Για παράδειγμα, η απάντηση στην ερώτηση “θα έχει κίνηση σήμερα” μπορεί να ενδιαφέρει ένα άτομο σε προσωπικό επίπεδο υπό την έννοια του αν θα χρειαστεί να ξεκινήσει νωρίτερα για τον προορισμό του και σε επαγγελματικό επίπεδο για το αν μία εταιρεία θα μπορέσει να τροφοδοτήσει εγκαίρως τους πελάτες της. Σε ευρύτερο επίπεδο ενδιαφέρει τον ακαδημαϊκό κόσμο ως αποτέλεσμα της ανάγκης που γεννάται σε άλλες ομάδες, όπως οι επιχειρήσεις ή οι κυβερνήσεις, να αντιμετωπίσουν την αβεβαιότητα του μέλλοντος. Με την πάροδο του χρόνου και ιδιαίτερα με την ανάπτυξη της επιστήμης των υπολογιστών και τις γρήγορες ταχύτητες του διαδικτύου η μεταβλητότητα του περιβάλλοντος έχει ενταθεί. Η μεταβλητότητα αυτή συνοδεύεται από μεγαλύτερη αβεβαιότητα και συνεπώς επιβάλλεται η πιο συστηματική και προσεκτική έρευνα του μέλλοντος. Οι προβλέψεις που παράγονται ως αποτέλεσμα αυτής της έρευνας χρησιμοποιούνται σαν δεδομένα σε διαδικασίες προγραμματισμού, πολιτικού και κοινωνικού σχεδιασμού, ελέγχου αναφοράς και σε πλήθος άλλων δραστηριοτήτων.

3.1 Γενικά περί Προβλέψεων

3.1.1 Πρόβλεψη

Η πρόβλεψη αποτελεί μια εκτίμηση για ένα μελλοντικό γεγονός. Για την παραγωγή προβλέψεων σε επίπεδο απλών καθημερινών δραστηριοτήτων βασιζόμαστε συνήθως στην παρατήρηση, την εμπειρία ή ακόμα και τη διαίσθηση. Οι προβλέψεις μπορεί να είναι μακροπρόθεσμες όπως για παράδειγμα ο σχεδιασμός της στρατηγικής μιας επιχείρησης ή βραχυπρόθεσμες όπως το κόστος λειτουργίας του τμήματος πωλήσεων για το επόμενο έτος. Η παραγωγή προβλέψεων σε επιχειρησιακό επίπεδο βασίζεται συνήθως σε συνδυασμό επιστημονικών και εμπειρικών μεθόδων. Σε κάθε μία από τις παραπάνω περιπτώσεις η ανάγκη της πρόβλεψης είναι σημαντική. Ακόμα πιο σημαντική όμως, είναι η ανάγκη η πρόβλεψη να είναι αποτελεσματική. Αυτό σημαίνει ότι θέλουμε να έχουμε τη μικρότερη δυνατή απόκλιση από το πραγματικό γεγονός.

3.1.2 Κατηγορίες Μεθόδων Προβλέψεων

Προτού γίνει η επιλογή της μεθόδου πρόβλεψης ακολουθείται μια διαδικασία στην οποία μπορούμε να διακρίνουμε τα ακόλουθα βήματα:

1. Προσδιορισμός του προβλήματος και συλλογή δεδομένων
2. Επεξεργασία και καθορισμός των δεδομένων
3. Κατασκευή και αξιολόγηση του μοντέλου
4. Εφαρμογή του μοντέλου
5. Πρόβλεψη με το μοντέλο που επιλέχθηκε
6. Αξιολόγηση της πρόβλεψης

Οι τεχνικές προβλέψεων ταξινομούνται σε τρεις μεγάλες κατηγορίες:

1. Ποσοτικές (Quantitative)
2. Κριτικές (Judgmental)
3. Τεχνολογικές (Technological)

Σε κάθε μία από τις παραπάνω κατηγορίες εμπεριέχονται επιμέρους μοντέλα. Ένα μοντέλο αντιπροσωπεύει μία διαδικασία που ακολουθείται για να παραχθούν προβλέψεις και έτσι το κάθε ένα από αυτά αντιστοιχεί σε μία συγκεκριμένη τεχνική. Ακολουθεί σύντομη περιγραφή των χαρακτηριστικών των παραπάνω μεθόδων.

Ποσοτικές μέθοδοι

Οι ποσοτικές μέθοδοι στηρίζονται στην ποσοτικοποίηση της διαθέσιμης πληροφορίας σε αριθμητικά δεδομένα. Για να μπορεί να λειτουργήσει αυτή η μέθοδος απαιτεί τα δεδομένα αυτά να ακολουθούν το ίδιο πρότυπο συμπεριφοράς και στο μέλλον, γεγονός το οποίο δεν μπορεί να διασφαλιστεί. Αυτό αποτελεί και το μεγαλύτερο μειονέκτημα αυτών των μεθόδων. Τα μοντέλα που αναφέρονται στις ποσοτικές μεθόδους πρόβλεψης διακρίνονται σε μοντέλα χρονοσειρών (time series models) και αιτιοκρατικά μοντέλα (explanatory models).

Τα μοντέλα χρονοσειρών βασίζονται στην υπόθεση ότι η μεταβολή της τιμής του προς πρόβλεψης μεγέθους ακολουθεί ένα συγκεκριμένο λανθάνον πρότυπο το οποίο επαναλαμβάνεται στο χρόνο και παραμένει σταθερό (Πετρόπουλος κ.ά., 2013). Οι πιο δημοφιλείς μέθοδοι χρονοσειρών είναι η μέθοδος της αποσύνθεσης (decomposition), η

μέθοδος εξομάλυνσης (smoothing) και οι αυτοπαλινδομικές (ARMA). Ορισμένες από τις μεθόδους αυτές θα περιγράψουν παρακάτω.

Στα αιτιοκρατικά μοντέλα η βασική υπόθεση είναι ότι υπάρχει μια σχέση μεταξύ του υπό πρόβλεψη μεγέθους (εξαρτημένη μεταβλητή) και ορισμένων παραμέτρων (ανεξάρτητες μεταβλητές) που το επηρεάζουν (Πετρόπουλος κ.ά., 2013).

Κριτικές μέθοδοι

Αυτές οι μέθοδοι αποτελούν μονόδρομο όταν δεν υπάρχουν αρκετά δεδομένα για να εφαρμοστούν ποσοτικές μέθοδοι και χρησιμοποιούνται υποστηρικτικά σε άλλες μεθόδους πρόβλεψης. Σε αυτές τις μεθόδους προβλέψεων η πρόβλεψη γίνεται από έναν ή περισσότερους ειδικούς με βάση τη γνώμη, την εμπειρία ή και το ένστικτό τους. Το βασικό μειονέκτημα αυτών των μεθόδων είναι η υποκειμενικότητα ή αλλιώς προκατάληψη (bias). Οι βασικές μέθοδοι κριτικών προβλέψεων είναι:

- Η **απλή κρίση (unaided judgment)**, η οποία αποτελεί και μέθοδο αναφοράς (benchmark) στις κριτικές προβλέψεις, αναφέρεται στην πρόβλεψη από τους ειδικούς χωρίς αυτοί να ακολουθούν κάποια συνεργασία ή μεθοδολογία. Χρησιμοποιείται σε προβλήματα απόφασης όταν δεν υπάρχει σύγκλιση απόψεων.
- Η μέθοδος **Delphi** για την εφαρμογή της οποίας συγκροτείται μια ομάδα ειδικών εξωτερικών συνεργατών για την οποία ομάδα διατηρείται η ανωνυμία των συμμετεχόντων ώστε να μην μπορεί να τους ασκηθεί οποιαδήποτε πίεση από τη διοίκηση και έτσι οι απόψεις τους να είναι ανεπηρέαστες και έγκυρες. Με τη μέθοδο αυτή γίνεται συλλογή δεδομένων από τα μέλη της ομάδας με τη μορφή ερωτηματολογίων. Η διαδικασία επαναλαμβάνεται περισσότερες από μία φορές και δεν απαιτείται ομοφωνία για το τελικό αποτέλεσμα αλλά αρκεί η μέγιστη δυνατή συναίνεση.
- Η μέθοδος **Αναλογιών** βασίζεται στην ανάκληση γεγονότων και καταστάσεων του παρελθόντος, όμοιων με το τρέχον πρόβλημα, με σκοπό να χρησιμοποιηθούν ως καθοδήγηση για την ερμηνεία και την πρόβλεψη τρεχουσών καταστάσεων. Χρησιμοποιείται για παράδειγμα, για τον σχεδιασμό ενός προϊόντος όπου η πρόβλεψη πραγματοποιείται με το ιστορικό ενός παρόμοιου προϊόντος.
- Η μέθοδος **Δομημένων Αναλογιών** βασίζεται στη χρήση μίας δομημένης διαδικασίας κατά την οποία οι ειδικοί ανακαλούν στη μνήμη τους παρόμοιες καταστάσεις του παρελθόντος και έπειτα αξιολογούν τη σχέση τους με την

τρέχουσα κατάσταση. Η μέθοδος αυτή χρησιμοποιείται για να βοηθήσει τους ειδικούς να επεξεργαστούν καλύτερα τις πληροφορίες που μπορούν να παράγουν. Εφαρμόζεται στις περιπτώσεις όπου δύναται να υφίσταται αμεροληψία για αμφιλεγόμενα ή συναισθηματικά φορτισμένα θέματα. Η διαδικασία που ακολουθείται στην πιο απλή της μορφή περιλαμβάνει τα ακόλουθα στάδια:

1. Περιγραφή της κατάστασης
2. Επιλογή Ειδικών
3. Προσδιορισμός αναλογιών από τους ειδικούς
4. Αποτίμηση ομοιότητας Αναλογιών
5. Παραγωγή Προβλέψεων

Τεχνολογικές μέθοδοι προβλέψεων

Οι μέθοδοι αυτές βασίζονται στις απόψεις των ειδικών σχετικά με επερχόμενες τάσεις και αλλαγές. Χρησιμοποιούνται συνήθως για μακροπρόθεσμες προβλέψεις όταν τα δεδομένα για τα πρότυπα του παρελθόντος είναι ελλιπή. Διακρίνονται στις ακόλουθες δύο κατηγορίες με βάση τον χρόνο των δεδομένων:

- **Διερευνητικές (exploratory)** οι οποίες βασίζονται στην επέκταση των δεδομένων του παρελθόντος, διαμέσου του παρόντος, στο μέλλον, όπως για παράδειγμα η μέθοδος Delphi, οι S- καμπύλες και οι αναλογίες.
- **Δεοντολογικές (normative)** χαρακτηρίζονται οι τεχνικές στις οποίες μία επιθυμητή ή πιθανή κατάσταση γεγονότων προβάλλεται στο μέλλον. Συγκεκριμένα, προσδιορίζεται μία αναδρομική σειρά βημάτων που εκτιμάται ότι θα οδηγήσουν στην τρέχουσα κατάσταση και αξιολογείται η πιθανότητα πραγματοποίησής τους. Παραδείγματα τέτοιων μεθόδων είναι οι μήτρες αποφάσεων, τα δέντρα συσχετίσεων και η ανάλυση συστημάτων.

3.2 Χρονοσειρά - Χαρακτηριστικά Χρονοσειρών

Η χρονοσειρά αποτελείται από δεδομένα που συλλέγονται σε διαδοχικές χρονικές στιγμές. Παρατηρώντας τη γραφική απεικόνιση μιας χρονοσειράς μπορούμε να παρατηρήσουμε τέσσερα βασικά ποιοτικά χαρακτηριστικά (στην ιδανική περίπτωση όπου εμφανίζονται όλα). Αυτά είναι: η **τάση**, η **κυκλικότητα**, η **εποχικότητα** και οι **τυχαίες διακυμάνσεις**. Πιο συγκεκριμένα, αν σε μία χρονοσειρά τα δεδομένα μεταβάλλονται γύρω από μία σταθερή τιμή τότε λέμε ότι έχουμε μία οριζόντια επανάληψη και άρα η χρονοσειρά είναι

στάσιμη. Στην περίπτωση που μία χρονοσειρά εμφανίζει τάση σημαίνει ότι τα δεδομένα αυξάνονται ή μειώνονται με την πάροδο του χρόνου. Η κυκλικότητα μπορεί να εμφανίζεται στην περίπτωση όπου τα δεδομένα έχουν καθόδους και ανόδους με διαφορετικές περιόδους οι οποίες δεν είναι απαραίτητα σταθερές ενώ συνήθως είναι μεγαλύτερες του έτους. Επίσης, οι παρατηρήσεις μιας χρονοσειράς μπορεί να επηρεάζονται από εποχικούς παράγοντες, οπότε σε αυτή την περίπτωση λέμε ότι έχουμε εποχικότητα. Η διακύμανση που οφείλεται στην εποχικότητα είναι συνήθως σταθερή και μικρότερη του έτους. Τέλος, σε μια χρονοσειρά μπορούν να παρατηρηθούν πολύ απότομες αλλαγές στο επίπεδο των τιμών της, κάτι που μπορεί να οφείλεται σε ένα έντονο και απρόβλεπτο γεγονός.

Πολύ σημαντικό στάδιο στη διαδικασία πρόβλεψης είναι η διερευνητική ανάλυση των δεδομένων (Exploratory Data Analysis) κατά την οποία πραγματοποιείται τόσο η ανάλυση των ποιοτικών χαρακτηριστικών όσο και η εφαρμογή στατιστικής ανάλυσης στα δεδομένα μιας χρονοσειράς προκειμένου να γίνει καλύτερη κατανόηση των δεδομένων και τελικά να πραγματοποιηθεί η επιλογή του καλύτερου μοντέλου με περισσότερη ασφάλεια.

Η στατιστική ανάλυση περιλαμβάνει ορισμένους βασικούς στατιστικούς δείκτες όπως τους ακόλουθους:

- Μέση τιμή (Average)
- Μέγιστη και ελάχιστη τιμή (Maximum και Minimum)
- Τυπική απόκλιση (Standard Deviation)
- Διακύμανση (Variance)
- Συνδιακύμανση (Covariance)
- Συντελεστής γραμμικής συσχέτισης (Linear Correlation Coefficient)
- Συντελεστής αυτοσυσχέτισης (Autocorrelation Coefficient)
- Συντελεστής μεταβλητότητας (Coefficient of variation)
- Μέση τιμή διαστήματος μεταξύ ζητήσεων (Intermittent Demand)

3.3 Μέθοδοι Προβλέψεων

Το κεφάλαιο αυτό παραθέτει μερικές από τις πιο σημαντικές μεθόδους προβλέψεων.

3.3.1 Απλοϊκή μέθοδος (Naive)

Η μέθοδος αυτή δίνει ως πρόβλεψη για την επόμενη χρονική περίοδο την ίδια τιμή με την πραγματική τιμή της παρατήρησης την ακριβώς προηγούμενη χρονική περίοδο. Αποτελεί την πιο απλή μέθοδο πρόβλεψης και χρησιμοποιείται συνήθως ως σημείο αναφοράς (benchmark) για άλλες μεθόδους. Η πρόβλεψη για μια χρονική στιγμή t δίνεται από τον τύπο:

$$F_t = Y_{t-1}$$

Όπου:

F_t = Πρόβλεψη

Y_{t-1} = Παρατήρηση

3.3.2 Μέθοδος Αποσύνθεσης

Αντικείμενο των μεθόδων αποσύνθεσης είναι ο διαχωρισμός των κύριων χαρακτηριστικών των χρονοσειρών και η απομόνωσή τους. Τα χαρακτηριστικά των χρονοσειρών είναι: η τάση, ο κύκλος, η εποχικότητα και η τυχαιότητα. Η τάση που αντιπροσωπεύει τη γενική εικόνα της χρονοσειράς μπορεί να είναι ανοδική, πτωτική ή μηδενική. Ο κυκλικός παράγοντας αντιπροσωπεύει τις ανόδους ή τις πτώσεις λόγω ειδικών οικονομικών συνθηκών. Η εποχικότητα αντιπροσωπεύει τις περιοδικές διακυμάνσεις. Η τυχαιότητα αναφέρεται σε τυχαίες και μη προβλέψιμες διακυμάνσεις στα δεδομένα της χρονοσειράς. Η μαθηματική διατύπωση της αποσύνθεσης για μία δεδομένη χρονική στιγμή t είναι η εξής (Πετρόπουλος κ.ά., 2013):

$$Y_t = f(S_t, T_t, C_t, R_t)$$

Όπου:

Y_t = Παρατήρηση

S_t = Συνιστώσα εποχικότητας

T_t = Συνιστώσα τάσης

C_t = Συνιστώσα κύκλου

R_t = Συνιστώσα τυχαιότητας

3.3.3 Μέθοδοι Εξομάλυνσης

Οι μέθοδοι αυτές είναι καταλληλότερες για βραχυπρόθεσμες και μεσοπρόθεσμες προβλέψεις. Βασικά πλεονεκτήματα των μεθόδων αυτών είναι η απλότητά τους, η

εφαρμογή τους σε μεγάλου όγκου χρονοσειρές και οι μικρές απαιτήσεις σε υπολογιστικό χρόνο. Η πρώτη μέθοδος που χρησιμοποιείται για την εξομάλυνση των τιμών μίας χρονοσειράς είναι η μέθοδος κινητών μέσων όρων που χρησιμοποιούν έναν σταθερό αριθμό δεδομένων με ίσα βάρη για τον υπολογισμό του μέσου όρου. Στην περίπτωση αυτή η πρόβλεψη δίνεται από τον ακόλουθο τύπο:

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

Όπου:

F_{t+1} = Πρόβλεψη

k = Αριθμός παρατηρήσεων για τον υπολογισμό του μέσου όρου

Y_i = Παρατηρήσεις

Η μέθοδος εκθετικής εξομάλυνσης αναπτύχθηκε ως επέκταση της προηγούμενης μεθόδου. Η πρόβλεψη με τη μέθοδο αυτή προεκτείνει τα στοιχεία του προτύπου που ακολουθούν τα δεδομένα (τάση, κυκλικότητα) στο μέλλον. Στη μέθοδο αυτή τα ιστορικά δεδομένα συμμετέχουν όλα αλλά με άνισα βάρη και συγκεκριμένα οι πιο πρόσφατες παρατηρήσεις έχουν τη μεγαλύτερη βαρύτητα άρα και τη μεγαλύτερη επίδραση στην πρόβλεψη. Τα μοντέλα εκθετικής εξομάλυνσης χωρίζονται σε κατηγορίες ανάλογα με τη γενική μορφή της γραφικής παράστασης της χρονοσειράς. Σύμφωνα με την κατηγοριοποίηση αυτή προκύπτουν τέσσερα μοντέλα τάσης: τα σταθερού επιπέδου, γραμμικής τάσης, εκθετικής τάσης και φθίνουσας τάσης (Πετρόπουλος κ.ά., 2013).

Μοντέλο σταθερού επιπέδου (Simple Exponential Smoothing)

Αναφέρεται ως μοντέλο σταθερού επιπέδου διότι στη χρονοσειρά δεν παρατηρείται τάση. Περιγράφεται από τις εξισώσεις:

$$e_t = Y_t - F_t$$

$$S_t = S_{t-1} + a * e_t$$

$$F_{t+1} = S_t$$

Όπου:

e_t = Σφάλμα

S_t = Επίπεδο

F_t = πρόβλεψη

a = Συντελεστής εξομάλυνσης για το επίπεδο και λαμβάνει τιμές στο διάστημα $[0,1]$

Μοντέλο γραμμικής τάσης (Holt Exponential Smoothing)

Πρόκειται για ένα μοντέλο όμοιο με το προηγούμενο με το επιπλέον χαρακτηριστικό ότι μπορεί να διαχειριστεί τη συνιστώσα της τάσης. Περιγράφεται από τις παρακάτω εξισώσεις:

$$\begin{aligned}e_t &= Y_t - F_t \\S_t &= S_{t-1} + T_{t-1} + a * e_t \\T_t &= T_{t-1} + a * \beta * e_t \\F_{t+m} &= S_t + m * T_t\end{aligned}$$

Όπου:

e_t = Σφάλμα

S_t = Επίπεδο

T_t = Τάση

F_t = πρόβλεψη

a = Συντελεστής εξομάλυνσης για το επίπεδο και λαμβάνει τιμές στο διάστημα $[0,1]$

β = Συντελεστής εξομάλυνσης τάσης και λαμβάνει τιμές στο διάστημα $[0,1]$

m = Χρονικός ορίζοντας πρόβλεψης

Μοντέλο μη γραμμικής τάσης (Damped Exponential Smoothing)

Αποτελεί μια ακόμα πιο βελτιωμένη διατύπωση του παραπάνω μοντέλου το οποίο έχει μια τάση για υπερεκτίμηση (θετική προκατάληψη). Συγκεκριμένα είναι εμπλουτισμένο με την παράμετρο διόρθωσης της τάσης, την παράμετρο φ , η οποία ελέγχει τον ρυθμό αύξησης των τιμών. Οι εξισώσεις που περιγράφουν το μοντέλο αυτό είναι:

$$\begin{aligned}e_t &= Y_t - F_t \\S_t &= S_{t-1} + \varphi * T_{t-1} + a * e_t \\T_t &= \varphi * T_{t-1} + T_{t-1} + a * \beta * e_t \\F_{t+m} &= S_t + \sum_{i=1}^m \varphi^i * T_t\end{aligned}$$

Όπου:

e_t = Σφάλμα

S_t = Επίπεδο

T_t = Τάση

F_t = πρόβλεψη

a = Συντελεστής εξομάλυνσης για το επίπεδο και λαμβάνει τιμές στο διάστημα $[0, 1]$

β = Συντελεστής εξομάλυνσης τάσης και λαμβάνει τιμές στο διάστημα $[0, 1]$

m = Χρονικός ορίζοντας πρόβλεψης

Η παράμετρος φ μπορεί να λάβει τιμή 0 οπότε τότε μιλάμε για απλή εκθετική εξομάλυνση, τιμή από 0 έως 1 οπότε προκύπτει το μοντέλο φθίνουσας τάσης και τιμή μεγαλύτερη του 1 οπότε προκύπτει το μοντέλο της εκθετικής τάσης. Τέλος για τιμή ίση με 1 έχουμε το μοντέλο της γραμμικής τάσης.

3.3.4 Μοντέλα Παλινδρόμησης

Τα μοντέλα παλινδρόμησης εστιάζουν στην εύρεση συσχέτισης μεταξύ μίας εξαρτημένης και μίας ή περισσότερων ανεξάρτητων μεταβλητών. Χρησιμοποιείται τόσο για πρόβλεψη όσο και για την ανάλυση των μεταβλητών που επηρεάζουν τη μεταβλητή που θέλουμε να προβλέψουμε. Η ανάλυση αυτή είναι πολύ σημαντική κατά την επεξεργασία των δεδομένων για να αποφανθούμε για τις μεταβλητές που θα λάβουμε υπόψη στη μελέτη. Διακρίνονται ανάλογα με τον αριθμό των μεταβλητών που συσχετίζονται με την υπό πρόβλεψη μεταβλητή σε απλή παλινδρόμηση (για μία) και πολλαπλή παλινδρόμηση (για περισσότερες). Στην περίπτωση της απλής γραμμικής παλινδρόμησης η σχέση μεταξύ των μεταβλητών εκφράζεται από την εξίσωση της ευθείας γραμμής ως ακολούθως:

$$Y = a + b * X$$

Όπου:

a = Σημείο τομής της ευθείας με τον άξονα των εξαρτημένων

b = Κλίση της ευθείας

X = Ανεξάρτητη μεταβλητή

3.4 Αξιολόγηση Μεθόδων πρόβλεψης

Πολύ σημαντικό στάδιο της διαδικασίας των προβλέψεων, αυτό που θα μας αποκαλύψει την ποιότητα της ανάλυσης, είναι αυτό της αξιολόγησης. Στο στάδιο της αξιολόγησης χρειάζονται οι προβλεφθείσες και οι πραγματικές τιμές προκυμμένου να υπολογιστεί η

απόκλιση (σφάλμα). Ως σφάλμα ορίζεται η διαφορά μεταξύ της πραγματικής και προβλεφθείσας τιμής. Για μία περίοδο i το σφάλμα δίνεται από τον τύπο:

$$e_i = Y_i - F_i$$

Όπου:

e_i = Σφάλμα

Y_i = Πραγματική τιμή

F_i = Πρόβλεψη

Για τις ανάγκες (την αξιολόγηση) της μελέτης της παρούσας εργασίας αποφασίστηκε να χρησιμοποιηθεί η ρίζα του μέσου τετραγωνικού σφάλματος, το οποίο θα αναλυθεί εκτενέστερα σε επόμενο κεφάλαιο. Κρίνεται, ωστόσο, σκόπιμο να γίνει μια αναφορά σε ορισμένους από τους σημαντικότερους δείκτες της ανάλυσης ακρίβειας της πρόβλεψης.

Ορισμένοι από τους δείκτες υπολογισμού του σφάλματος είναι οι παρακάτω:

Μέσο σφάλμα (Mean Error)

Το μέσο σφάλμα αναφέρεται στον απλό μέσο όρο των σφαλμάτων και για πραγματικές τιμές Y και προβλεφθείσες F δίνεται από τον τύπο:

$$ME = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)$$

Όπου:

Y_i = Πραγματική τιμή

F_i = Πρόβλεψη

n = Πλήθος τιμών πραγματικής χρονοσειράς για τις οποίες έχουμε κάνει πρόβλεψη

Για θετικές τιμές του δείκτη λέμε ότι έχουμε αισιόδοξες προβλέψεις ενώ για αρνητικές τιμές του δείκτη έχουμε αισιόδοξες προβλέψεις.

Μέσο απόλυτο σφάλμα (Mean Absolute Error)

Ο δείκτης αυτός είναι ένα μέτρο που υποδεικνύει την ακρίβεια της πρόβλεψης σε σχέση με τις πραγματικές τιμές. Υψηλή τιμή του δείκτη σημαίνει ότι η πρόβλεψη αποκλίνει πολύ από την πραγματική τιμή και άρα είναι άστοχη. Δίνεται από τον τύπο:

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - F_i|$$

Όπου:

Y_i = Πραγματική τιμή

F_i = Πρόβλεψη

n = Πλήθος τιμών πραγματικής χρονοσειράς για τις οποίες έχουμε κάνει πρόβλεψη

Μέσο τετραγωνικό σφάλμα (Mean Squared Error)

Είναι επίσης ένα μέτρο της ακρίβειας της πρόβλεψης και δεδομένου ότι το σφάλμα τετραγωνίζεται προκύπτει εύλογα το συμπέρασμα ότι δίνει μεγαλύτερη βαρύτητα σε μεγάλες αποκλίσεις από την πραγματική τιμή και μικρότερη σε μικρές αποκλίσεις. Δίνεται από τον ακόλουθο τύπο:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - F_i)^2$$

Όπου:

Y_i = Πραγματική τιμή

F_i = Πρόβλεψη

n = Πλήθος τιμών πραγματικής χρονοσειράς για τις οποίες έχουμε κάνει πρόβλεψη

Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error)

Πρόκειται για έναν δείκτη ιδιαίτερα χρήσιμο στην αξιολόγηση μίας πρόβλεψης που έχει προκύψει από διαφορετικές χρονοσειρές και διαφορετικό επίπεδο μέσης τιμής. Δεδομένου του ότι εκφράζεται σε ποσοστό επί τοις εκατό παίρνει τιμές μεγαλύτερες ή ίσες του μηδενός. Όσο μικρότερη η τιμή αυτού του δείκτη τόσο πιο ακριβής είναι η πρόβλεψη. Δίνεται από τον τύπο:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - F_i}{Y_i} \right| * 100 (\%)$$

Όπου:

Y_i = Πραγματική τιμή

F_i = Πρόβλεψη

n = Πλήθος τιμών πραγματικής χρονοσειράς για τις οποίες έχουμε κάνει πρόβλεψη

3.5 Επιλογή Κατάλληλης Μεθόδου Πρόβλεψης

Η σημαντικότητα της ανάγκης για πρόβλεψη σε όλους τους τομείς δραστηριότητας, επιχειρησιακούς ή όχι, έχει οδηγήσει στην ανάπτυξη των τεχνικών προβλέψεων από διαφορετικούς κλάδους και διαφορετικούς σκοπούς. Για να αποτελέσει μία μέθοδος πρόβλεψης εργαλείο στα χέρια ενός μελετητή θα πρέπει να έχει επιλεγεί έπειτα από προσεκτική ανάλυση και κατανόηση του προβλήματος. Η κατανόηση του προβλήματος μπορεί να επιτευχθεί αν απαντηθούν ερωτήματα όπως το πώς θα χρησιμοποιηθούν οι προβλέψεις και από ποιους. Στη συνέχεια θα πρέπει να συγκεντρωθούν και να αναλυθούν τα διαθέσιμα δεδομένα έτσι ώστε να έρθουν σε μορφή χρήσιμη για την κατασκευή ενός μοντέλου. Η ανάλυση των δεδομένων αναφέρεται στη γραφική απεικόνιση των δεδομένων για την οπτική επιθεώρηση των στοιχείων, στη στατιστική ανάλυση αυτών με τον υπολογισμό στατιστικών δεικτών κ.ά. Τα τελευταία βοηθούν στην εισήγηση των καταλληλότερων μοντέλων πρόβλεψης και στην τελική πρόβλεψη.

4 Παράγοντες που επηρεάζουν την τιμή του Bitcoin

Το Bitcoin παρ' ότι εμφανίζεται στην αγορά περισσότερο από δέκα χρόνια αυτά δεν είναι αρκετά για να μπορέσουν να εξαχθούν πάγια συμπεράσματα για τη συμπεριφορά του σαν νόμισμα. Η ραγδαία αύξηση της τιμής του, αναμφίβολα, έχει οδηγήσει το Bitcoin στο να αποτελέσει πεδίο ενδιαφέροντος για την επιστημονική κοινότητα παγκοσμίως. Από τα πρώτα χρόνια της εμφάνισής του διάφορες επιστημονικές ομάδες έχουν προσπαθήσει να αποσαφηνίσουν το φαινόμενο Bitcoin, τα χαρακτηριστικά του και τους παράγοντες που το επηρεάζουν. Στο παρόν κεφάλαιο θα γίνει αναφορά στους παράγοντες εκείνους που βάσει βιβλιογραφίας φαίνεται να επηρεάζουν την τιμή του Bitcoin αλλά και σε εκείνους που δεν το επηρεάζουν παρ' ότι κάτι τέτοιο θα αναμενόταν (Ciaian κ.ά., 2014).

4.1 Συνολικά Διαθέσιμα Bitcoin

Βασικό χαρακτηριστικό του Bitcoin είναι ο προκαθορισμένος αριθμός μονάδων που θα προσφερθούν συνολικά. Πλέον, τα διαθέσιμα Bitcoin αυξάνονται βραδέως. Το έτος 2140 έχει οριστεί ως το τέλος της διαδικασίας εξόρυξης Bitcoin. Το έτος αυτό ο ρυθμός έκδοσής του θα έχει συγκλίνει στο μηδέν ενώ οι διαθέσιμες μονάδες θα φτάσουν στον μέγιστο αριθμό, ο οποίος είναι 21 εκατομμύρια, και δεν θα μπορούν να εκδοθούν νέες μονάδες. Για να γίνει αντιληπτή η βραδεία ανάπτυξη του αρκεί να σημειώσουμε ότι μέχρι το τέλος αυτού του έτους (2020) ο όγκος των Bitcoin που θα έχει εξορυχθεί αναμένεται να έχει φτάσει στο 90% της συνολικής παραγωγικής δυναμικότητας.

4.2 Συναλλαγματικές Ισοτιμίες

Παρά το γεγονός ότι ο δημιουργός του Bitcoin οραματιζόταν ένα ψηφιακό νόμισμα η λειτουργία του οποίου θα ήταν αντίστοιχη με αυτή των γνωστών αποδεκτών νομισμάτων δύσκολα μπορεί να πει κανείς ότι έχει επιτελέσει τον αρχικό του σκοπό. Υποστηρίζεται ότι το Bitcoin δεν συμπεριφέρεται όπως ένα κοινό νόμισμα, σύμφωνα με τα κριτήρια που διακρίνουν τα νομίσματα. Αντί αυτού το Bitcoin μοιάζει περισσότερο κερδοσκοπικό μέσο επένδυσης όμοιο με τις μετοχές. Από έρευνα έχει προκύψει ότι η ημερήσια συναλλαγματική ισοτιμία του Bitcoin με το δολάριο παρουσιάζει σχεδόν μηδενική συσχέτιση σε σχέση με τις συναλλαγματικές ισοτιμίες του δολαρίου έναντι άλλων νομισμάτων όπως το ευρώ (Yermack, 2015).

4.3 Μακροοικονομικά Δεδομένα

Σύμφωνα με την έρευνα των (Ciaian κ.ά., 2014) η τιμή του Bitcoin δεν επηρεάζεται από μακροοικονομικά δεδομένα αφού δεν εκδίδεται από καμία κυβέρνηση. Για τον παραπάνω ισχυρισμό έγινε η θεώρηση ότι η προσφορά και η ζήτηση του Bitcoin είναι εξίσου σημαντικά για τη διαμόρφωση της τιμής του. Η προσφορά ορίστηκε ως τα διαθέσιμα Bitcoin σε μια συγκεκριμένη ημέρα επί την τιμή μονάδας εκείνης της ημέρας. Η ζήτηση ορίστηκε ως η τιμή του συναλλάγματος δολαρίου σε ευρώ επί το πλήθος των συναλλαγών ανά ημέρα διά τα διαθέσιμα συνολικά Bitcoin την ημέρα εκείνη επί το Bitcoin velocity. Με τον όρο velocity περιγράφεται η αξία των συναλλαγών μίας ημέρας διά τη συνολική αξία όλου του διαθέσιμου Bitcoin.

4.4 Δημοφιλία

Η δημοφιλία του Bitcoin στο ευρύ κοινό είναι ένας παράγοντας που φαίνεται να επηρεάζει την τιμή του. Ένας τρόπος να μετρηθεί η δημοφιλία του Bitcoin είναι μέσα από τις καθημερινές επισκέψεις στο λήμμα Bitcoin του Wikipedia (Kristoufek, 2013). Οι επισκέψεις αυτές έχουν συμπεριληφθεί στα δεδομένα για τη μελέτη της εργασίας αυτής. Η δημοφιλία μπορεί εξίσου να μετρηθεί και μέσω του αριθμού των ερωτημάτων στο Google Search που περιέχουν τον όρο Bitcoin (Bouoiyour κ.ά., 2014).

4.5 Συναίσθημα των συναλλασσόμενων

Μια έρευνα που πραγματοποιήθηκε το 2016 έρχεται να ενισχύσει την άποψη ότι η τιμή του Bitcoin επηρεάζεται περισσότερο από μη ορθολογικούς παράγοντες όπως είναι η ελκυστικότητά του ως προς τις κερδοσκοπικές επενδύσεις. Προκύπτει δηλαδή ότι το συναίσθημα των συναλλασσόμενων επηρεάζεται από το πως έχουν διαμορφωθεί προηγούμενες συναλλαγές στην αγορά με αποτέλεσμα να καθοδηγούν τις προτιμήσεις στο μέλλον και άρα να επηρεάζουν τη μεταβλητότητα του Bitcoin. Πιο συγκεκριμένα, υπάρχει σχέση μεταξύ θετικού συναισθήματος και μεταβολής στην τιμή του Bitcoin. Επιπλέον, το θετικό συναίσθημα έχει μεγαλύτερη επιρροή σε σύγκριση με το αρνητικό. Σε περιόδους υπερβολικής μεταβλητότητας η αξία της ερμηνείας του συναισθήματος αυξάνεται, ειδικά του θετικού (Bukovina, 2016).

4.6 Γενικός Δείκτης Χρηματιστηρίου της Κίνας

Μελετώντας τους δείκτες του χρηματιστηρίου που επηρεάζουν την τιμή του Bitcoin έχει διαπιστωθεί ότι ο δείκτης του Κινέζικου χρηματιστηρίου παίζει μεγάλο ρόλο στην τιμή του

Bitcoin και αυτό γιατί η Κίνα είναι η μεγαλύτερη αγορά Bitcoin παγκοσμίως (Βουοιγour κ.ά., 2014). Η επίδραση της Κινεζικής αγοράς φαίνεται εξίσου και σε άλλα περιστατικά όπως αυτό της χαρακτηριστικής αύξησης της τιμής του Bitcoin μόλις η Κινεζική εταιρεία Baidu, αντίστοιχη της Google, ανακοίνωσε ότι δέχεται το Bitcoin ως μέσο για πληρωμές.

4.7 Hashrate

Με τον όρο Hashrate περιγράφεται η συνολική υπολογιστική ισχύς του δικτύου εκφρασμένη ως το πλήθος των hashes που είναι ικανό να παράξει το συνολικό δίκτυο του Bitcoin. Σύμφωνα με μελέτη που πραγματοποιήθηκε το 2014 η τιμή του Bitcoin φαίνεται να επηρεάζεται σημαντικά από το υπάρχον Hashrate (Βουοιγour κ.ά., 2014).

4.8 Ευγενή Μέταλλα

Ένα ερώτημα που έχει απασχολήσει αρκετές έρευνες είναι η επιρροή που μπορεί να έχουν ή μη τα ευγενή μέταλλα στην τιμή του Bitcoin. Δεδομένου του έντονου επενδυτικού ενδιαφέροντος για το Bitcoin αλλά και της ριζοσπαστικής φύσης του εμφανίζεται στη βιβλιογραφία και ως νέος χρυσός. Ως προς την επιρροή που έχει σε αυτό ο πραγματικός χρυσός, αυτή φαίνεται να είναι ελάχιστη σύμφωνα με την ακόλουθη έρευνα (Βουοιγour κ.ά., 2014). Σε μία ακόμα έρευνα, οι (Klein κ.ά., 2018) αναλύοντας και συγκρίνοντας τις ιδιότητες της υπό συνθήκη διακύμανσης του Bitcoin και άλλων περιουσιακών στοιχείων κι έπειτα χρησιμοποιώντας ένα BEKK-GARCH μοντέλο για να προσδιορίσουν τις χρονικά μεταβαλλόμενες υπό συνθήκη συσχετίσεις, έδειξαν ότι ο χρυσός και το Bitcoin είναι δύο πολύ διαφορετικά περιουσιακά στοιχεία. Τα αποτελέσματα δείχνουν πως το Bitcoin συμπεριφέρεται ακριβώς αντίθετα από τον χρυσό, δηλαδή συσχετίζεται θετικά με τις καθοδικές αγορές. Τέλος, η εν λόγω έρευνα μελετά το Bitcoin και ως συστατικό χαρτοφυλακίου, το οποίο σημειώνεται ως περιουσιακό στοιχείο χωρίς καμία δυνατότητα αντιστάθμισης. Το τελικό συμπέρασμα στο οποίο η έρευνα καταλήγει είναι ότι το Bitcoin και ο χρυσός έχουν εντελώς διαφορετικές ιδιότητες ως περιουσιακά στοιχεία και ότι το Bitcoin δεν αντικατοπτρίζει καμία διακριτή ιδιότητα του χρυσού.

4.9 Βασικά αγαθά

Η συσχέτιση της διακύμανσης της τιμής του Bitcoin με τις αγορές βασικών αγαθών έχει υπάρξει ένα ακόμη βασικό ερώτημα διάφορων ερευνών. Σύμφωνα με τα ευρήματα της μελέτης των (Ji κ.ά., 2019), το Bitcoin και τα κρυπτονομίσματα γενικότερα είναι ενσωματωμένα μέσα σε καθορισμένες αγορές βασικών προϊόντων. Η μελέτη στηρίχθηκε

σε δεδομένα για τα κρυπτονομίσματα Bitcoin, Ethereum, Ripple, Stellar και Litecoin καθώς και για προϊόντα άλλων ομάδων όπως το φυσικό αέριο, το πετρέλαιο θέρμανσης, το αργό πετρέλαιο, ο χρυσός, το ασήμι, το αλουμίνιο, ο χαλκός και το νικέλιο. Επιπλέον έκανε χρήση δεδομένων για το κακάο, τον καφέ, το καλαμπόκι, το βαμβάκι και τη ζάχαρη. Τα συμπεράσματα στα οποία καταλήγει η εν λόγω έρευνα είναι ότι παρόλο που το Bitcoin δεν είναι έξω από το σύστημα, δεν επηρεάζει απαραίτητα τις τιμές των βασικών αγαθών με την αυστηρή έννοια. Αντιθέτως, φαίνεται να σχετίζεται με τα γεωργικά προϊόντα (αν και με κάποια καθυστέρηση).

4.10 Χρηματιστηριακοί Δείκτες

Οι (Panagiotidis κ.ά., 2018) μελέτησαν την επιρροή διαφόρων περιστατικών σοκ στις αποδόσεις του Bitcoin. Χρησιμοποιώντας μοντέλα VAR, FAVAR και Impulse Responses Functions αξιολόγησαν τις επιδράσεις των χρηματιστηριακών αγορών, των συναλλαγματικών ισοτιμιών, των αποδόσεων χρυσού και αργού πετρελαίου, των επιτοκίων των Τραπεζών FED (Ομοσπονδιακή Τράπεζα των ΗΠΑ) και EKT (Ευρωπαϊκή Κεντρική Τράπεζα) καθώς και των διαδικτυακών τάσεων. Συμπέραναν ότι οι αποδόσεις του Bitcoin αλληλοεπιδρούν σημαντικά με τις παραδοσιακές χρηματιστηριακές αγορές και ασθενέστερα με τις αγορές συναλλάγματος (FOREX). Αξιοσημείωτο εύρημα της εργασίας είναι επίσης το αντίκτυπο των ασιατικών αγορών στη διαμόρφωση των τιμών του Bitcoin. Βάσει της παραπάνω έρευνας αλλά και της γενικότερης βιβλιογραφίας η συσχέτιση των χρηματιστηριακών δεικτών με τον καθορισμό της τιμής του Bitcoin χρήζει η επιβεβαίωσης.

4.11 Όγκος Συναλλαγών

Ο (Sovbetov, 2018) χρησιμοποιώντας εβδομαδιαία δεδομένα έδειξε πως παράγοντες που σχετίζονται με κρυπτονομίσματα, όπως ο όγκος συναλλαγών φαίνεται να είναι καθοριστικοί για τη διαμόρφωση της τιμής τους τόσο βραχυπρόθεσμα όσο και μακροπρόθεσμα. Συμπέρασμα της παραπάνω έρευνας είναι ότι το Bitcoin, ως ο σημαντικότερος εκπρόσωπος των κρυπτονομισμάτων, φαίνεται να παρουσιάζει συσχέτιση της τιμής του με τον όγκο των συναλλαγών που διενεργείται.

5 Επικρατέστερα μοντέλα πρόβλεψης της τιμής του Bitcoin

Μελετώντας κανείς τη διαθέσιμη βιβλιογραφία σχετικά με την πρόβλεψη της τιμής του Bitcoin θα συναντήσει αρκετές διαφορετικές προσεγγίσεις όσον αφορά τις μεθόδους που χρησιμοποιούνται για το πρόβλημα αυτό. Τα μοντέλα αυτά επιτυγχάνουν περισσότερο ή λιγότερο καλά αποτελέσματα υπό συγκεκριμένες προϋποθέσεις ενώ εφαρμόζονται σε διαφορετικά σύνολα δεδομένων. Για τις ανάγκες της εργασίας έγινε χρήση των μοντέλων, η περιγραφή των οποίων ακολουθεί, και σύγκριση των επιδόσεών τους σε πιο πρόσφατα και ολοκληρωμένα δεδομένα του Bitcoin. Δεδομένων των υψηλών υπολογιστικών απαιτήσεων που απαιτούνται για την εκτέλεση κάποιων μοντέλων και της αδυναμίας πρόσβασης σε υπολογιστικούς πόρους δεν κατέστη εφικτή η εκτέλεση ορισμένων άλλων μοντέλων που περιγράφονται στο υποκεφάλαιο 5.6. Με δεδομένο ότι στο επόμενο κεφάλαιο ακολουθεί η λεπτομερής ανάλυση της μελέτης της εργασίας, κρίνεται σκόπιμο για την καλύτερη κατανόησή της να γίνει μια πρώτη αναφορά στα μοντέλα και τις βασικές τους παραμέτρους.

5.1 AutoRegressive Integrated Moving Average (ARIMA)

Με τον όρο ARIMA γίνεται αναφορά στα αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου (AutoRegressive Integrated Moving Average). Τα μοντέλα ARIMA είναι στοχαστικά μαθηματικά μοντέλα με τα οποία μπορεί να περιγραφεί η διαχρονική εξέλιξη κάποιου φυσικού μεγέθους.

Στη βιβλιογραφία συναντώνται διάφορα στοχαστικά μοντέλα αλλά ως προς τη μελέτη της τιμής του Bitcoin τα μοντέλα ARIMA τυγχάνουν προτίμησης (Bakar κ.ά., 2017). Εκτός από την πρόβλεψη της τιμής του Bitcoin χρησιμοποιούνται ευρύτατα γιατί βρίσκουν εφαρμογή στη μελέτη πολλών άλλων μεγεθών και φαίνεται να δίνουν μια "καλή" εικόνα της διαχρονικής τους συμπεριφοράς, καθώς και ικανοποιητικά αποτελέσματα στην πρόβλεψη των μελλοντικών τιμών του μεγέθους.

Πρωτοπόροι στη μελέτη των μοντέλων ARIMA είναι οι (Box κ.ά., 2015), οι οποίοι πρότειναν ένα σύνολο αλγεβρικών μοντέλων πρόβλεψης από τα οποία θα μπορεί κανείς να επιλέξει το καταλληλότερο για την πρόβλεψη μιας χρονοσειράς. Στα μοντέλα αυτά η πρόβλεψη στηρίζεται στα ιστορικά δεδομένα και τα πρότυπα που ακολουθούν. Επιπλέον,

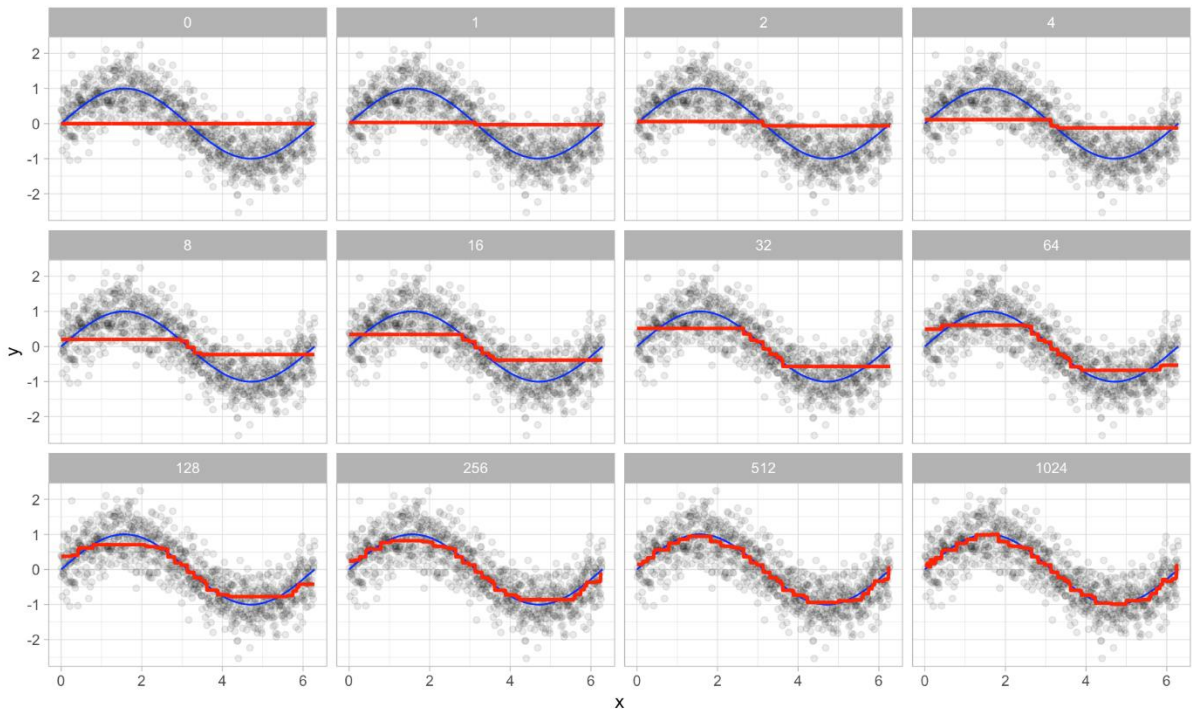
οι παραπάνω προτείνουν για την ανάλυση των χρονοσειρών μια διαδικασία τριών σταδίων: ταυτοποίηση, εκτίμηση και έλεγχος, πρόβλεψη.

Η ταυτοποίηση αναφέρεται στον καθορισμό των τιμών p , d και q ενός μοντέλου ARIMA (p , d , q). Ο αριθμός d υποδηλώνει τον αριθμό των διαφορών των παρατηρήσεων που χρησιμοποιήθηκαν για να μετατραπεί η σειρά σε στάσιμη. Έπειτα, το p αναφέρεται στους όρους αυτοπαλινδρόμησης του υποδείγματος και τέλος το q υποδηλώνει την τάξη της διαδικασίας του κινητού μέσου. Στο επόμενο στάδιο της εκτίμησης περιλαμβάνεται η εκτίμηση των p παραμέτρων της AR διαδικασίας και των q παραμέτρων της διαδικασίας MA. Ταυτόχρονα, μπορεί να διενεργείται και έλεγχος τάξεως συγκρίνοντας το υπόδειγμα με άλλο μεγαλύτερης τάξης. Τελικό στάδιο της διαδικασίας είναι η διενέργεια πρόβλεψης σε βραχυχρόνια περίοδο, όπου είναι και ο κύριος σκοπός της εξειδίκευσης και εκτίμησης ενός υποδείγματος ARIMA. Στα πλαίσια της εργασίας θα εξετάσουμε το μοντέλο ARIMA ελέγχοντας τα αποτελέσματά του σε μία σειρά από συνδυασμούς των παραμέτρων του.

5.2 Gradient Boosting Machine (GBM)

Το Gradient Boosting Machine (GBM) είναι μια μέθοδος μηχανικής μάθησης που εφαρμόζεται σε προβλήματα ταξινόμησης και παλινδρόμησης. Καταλήγει σε ένα μοντέλο πρόβλεψης χρησιμοποιώντας κυρίως δέντρα αποφάσεων. Τα δέντρα αυτά σε κάθε τους στάδιο προσπαθούν να διορθώσουν το λάθος της πρόβλεψης του προηγούμενου σταδίου. Λέγεται ότι η μέθοδος προσπαθεί να μετατρέψει τους ασθενείς μαθητές σε ισχυρούς. Σε κάθε επανάληψη προστίθεται ένα νέο δέντρο που προκύπτει ως μια τροποποιημένη έκδοση του αρχικού συνόλου δεδομένων. Έτσι εκπαιδεύονται πολλά μοντέλα με σταδιακό, προσθετικό και διαδοχικό τρόπο. Για να το πετύχει, το GBM χρησιμοποιεί την τεχνική της «καθόδου κλίσης» ώστε να βρει ελλείψεις και να τις διορθώσει στην επόμενη επανάληψη.

Στην παρακάτω εικόνα παρατηρούμε την εξέλιξη της εκπαίδευσης του μοντέλου GBM μετά από 1, 2, 4, ..., 1024 επαναλήψεις. Όσο αυξάνεται ο αριθμός των επαναλήψεων τόσο καλύτερα προσαρμόζεται το μοντέλο στις πραγματικές τιμές.



Εικόνα 1: Εκπαίδευση Gradient Boosting Machine

Παρά την σημαντικότητά και την αξία της μεθόδου αυτής αναπόφευκτα προκύπτουν κάποια προβλήματα από την εφαρμογή της ιδιαίτερα δε σε εξελιγμένες μορφές της όπως για παράδειγμα το XGBoost. Η κατανάλωση μνήμης είναι ένα από αυτά. Τα μοντέλα αυτά είναι αρκετά χρονοβόρα και απαιτούν μεγάλη υπολογιστική ισχύ. Είναι ωστόσο χρήσιμα και βρίσκουν ιδιαίτερη εφαρμογή σε προβλήματα βελτιστοποίησης μιας συνάρτησης κόστους.

Κατά την εφαρμογή του μοντέλου αυτού και για τον σκοπό της βελτίωσης των αποτελεσμάτων μπορούμε να ελέγξουμε τις υπερπαραμέτρους του μοντέλου. Από προηγούμενες μελέτες (Sun κ.ά., 2020) έχει προκύψει ότι κατά την πρόβλεψη της τιμής του Bitcoin με τον αλγόριθμο GBM έχει νόημα να βελτιστοποιηθούν οι εξής υπερπαραμέτροι (CRAN.R, 2020):

- **n.trees:** Η παράμετρος αυτή αναφέρεται στον μέγιστο ακέραιο αριθμό των δέντρων που θα χρησιμοποιήσει ο αλγόριθμος κατά την εκπαίδευσή του. Ξεκινάει από την προκαθορισμένη τιμή 100. Ανάλογα με το πρόβλημα ο αριθμός αυτός μπορεί να αυξηθεί για χάρη της βελτιστοποίησης του αποτελέσματος.
- **shrinkage:** Αναφέρεται ως παράμετρος συρρίκνωσης ή ποσοστό εκμάθησης. Η προκαθορισμένη τιμή της είναι 0,1 ενώ μπορεί να πάρει τιμές 0,01 ή 0,001.

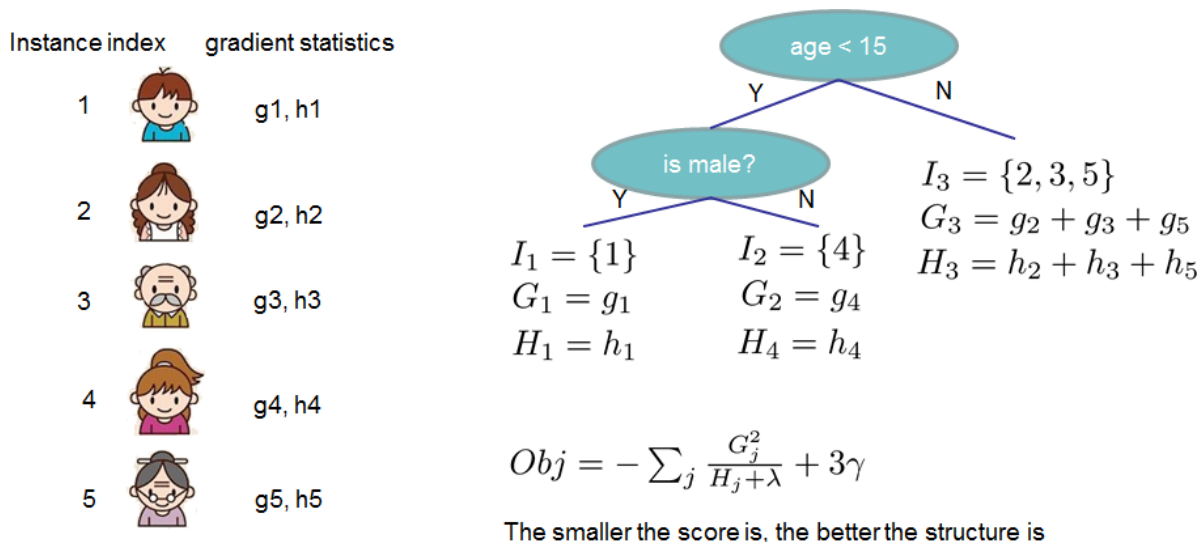
- **interaction.depth:** Πρόκειται για έναν ακέραιο αριθμό ο οποίος καθορίζει το μέγιστο επίπεδο μεταβλητών που μπορούν να αλληλοεπιδρούν ή αλλιώς το βάθος του δέντρου. Η προεπιλογή είναι το 1 ωστόσο μπορεί να λάβει και μεγαλύτερες τιμές.

5.3 Extreme Gradient Boosting (XGBoost)

Το μοντέλο XGBoost ή αλλιώς Ακραία Βαθμιδωτή Ενίσχυση (Extreme Gradient Boosting) αποτελεί μία υποκατηγορία των μοντέλων πρόβλεψης Βαθμιδωτής Ενίσχυσης (Gradient Boosting). Προτάθηκε από τους (Chen κ.ά., 2016), στα πλαίσια μίας έρευνας για το Deep Machine Learning Community (DMLC) Group. Ανάμεσα σε άλλες εφαρμογές του, επιτυγχάνει πολύ καλά αποτελέσματα και για την πρόβλεψη της τιμής του Bitcoin (Λιναρδάτος κ.ά.,2020).

Η Βαθμιδωτή Ενίσχυση είναι μια μέθοδος που προσθέτει διαδοχικά τα δεδομένα εκπαίδευσης και τους αποδίδει έναν συντελεστή βαρύτητας. Ωστόσο, αντί να αναθέσει διαφορετικά βάρη στους ταξινομητές μετά από κάθε επανάληψη, η μέθοδος αυτή προσαρμόζει το νέο μοντέλο στα υπολείμματα της προηγούμενης πρόβλεψης και στη συνέχεια ελαχιστοποιεί την απώλεια προσθέτοντας την τελευταία κάθε φορά πρόβλεψη.

Στόχος της μεθόδου είναι η εύρεση των βέλτιστων παραμέτρων που οδηγούν στη μεγαλύτερη μείωση απώλειας. Μέσω αυτού του κριτηρίου η μέθοδος Βαθμιδωτής Ενίσχυσης επιχειρεί να ελαχιστοποιήσει τα σφάλματα. Τελικά, το μοντέλο ενημερώνεται χρησιμοποιώντας Βαθμιδωτή Μείωση, αιτιολογώντας την ονομασία Gradient Boosting. Το άθροισμα των προβλέψεων γίνεται ολοένα και ισχυρότερο μετά από κάθε βήμα ενώ αυτή η διαδικασία επαναλαμβάνεται μέχρι να υπολογιστεί η τελική πρόβλεψη.



Εικόνα 2: Υπολογισμός όρου ρύθμισης κατά την εκτέλεση του XGBoost

Στην περίπτωση του XGBoost, ο αλγόριθμος για την ενίσχυση των δέντρων αποφάσεων υλοποιείται μέσω ενός πρόσθετου σημαντικού όρου ρύθμισης, την ποινή (ή Penalty). Ο όρος αυτός προστίθεται στη συνάρτηση υπολογισμού απώλειας. Ένα μειονέκτημα του μοντέλου XGBoost είναι ότι δέχεται μόνο αριθμούς ως δεδομένα εισόδου, μην μπορώντας να επεξεργαστεί οποιασδήποτε δεδομένα άλλης φύσης, όπως ημερομηνίες ή λέξεις (Raghu, 2018).

Στο μοντέλο XGBoost της μελέτης της εργασίας αναμένεται να εξεταστούν οι εξής υπερπαράμετροι (XGBoost, 2020):

- **eta:** Πρόκειται για τον ρυθμό εκμάθησης (learning rate ή shrinkage). Αντιστοιχεί στη συρρίκνωση των βαρών που σχετίζονται με τα χαρακτηριστικά μετά από κάθε γύρο. Με άλλα λόγια καθορίζει το ποσοστό της «διόρθωσης» που κάνουμε σε κάθε βήμα. Στην πράξη, η ύπαρξη χαμηλότερου eta καθιστά το μοντέλο μας πιο ανθεκτικό στο overfitting έτσι, συνήθως, όσο χαμηλότερο είναι το ποσοστό εκμάθησης, τόσο καλύτερο. Αλλά με χαμηλότερο eta, χρειαζόμαστε περισσότερους γύρους ενίσχυσης, που απαιτούν περισσότερο χρόνο για να εκπαιδευσουμε, μερικές φορές μόνο για οριακές βελτιώσεις.
- **max_depth:** Είναι ο μέγιστος αριθμός κόμβων που επιτρέπονται από τη ρίζα έως το πιο μακρινό φύλλο ενός δέντρου. Τα πιο βαθιά δέντρα μπορούν να μοντελοποιήσουν πιο πολύπλοκες σχέσεις προσθέτοντας περισσότερους

κόμβους, αλλά καθώς πηγαίνουμε βαθύτερα, οι διαχωρισμοί γίνονται λιγότερο σχετικοί και μερικές φορές οφείλονται μόνο σε θόρυβο, δημιουργώντας overfitting.

- **subsample:** Αντιστοιχεί στο κλάσμα των παρατηρήσεων που χρησιμοποιούνται για την επιλογή υποδείγματος σε κάθε βήμα.

5.4 Prophet

Η μέθοδος Prophet πρόκειται για μία διαδικασία πρόβλεψης χρονοσειρών η οποία κυκλοφόρησε σε πακέτο ανοικτού κώδικα από το Facebook το 2017 (Taylor κ.ά., 2017). Η μέθοδος αυτή στηρίζεται σε ένα προσθετικό μοντέλο εποχικότητας στο οποίο οι μη γραμμικές τάσεις προσαρμόζονται στην ετήσια, εβδομαδιαία αλλά και καθημερινή εποχικότητα καθώς και στην επίδραση των αργιών. Σύμφωνα με τις προδιαγραφές του, το μοντέλο Prophet παρουσιάζει καλύτερη επίδοση σε χρονοσειρές οι οποίες εμφανίζουν ισχυρή εποχικότητα και περιέχουν αρκετές εποχές ιστορικών δεδομένων. Τέλος, το μοντέλο Prophet έχει καλή συμπεριφορά ως προς τα outliers, τις αλλαγές τάσης και στις περιπτώσεις απουσίας δεδομένων.

Η μέθοδος Prophet είναι μία διαδικασία αθροιστικής παλινδρόμησης τεσσάρων παραγόντων, η οποία περιγράφεται από τον ακόλουθο μαθηματικό τύπο:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon$$

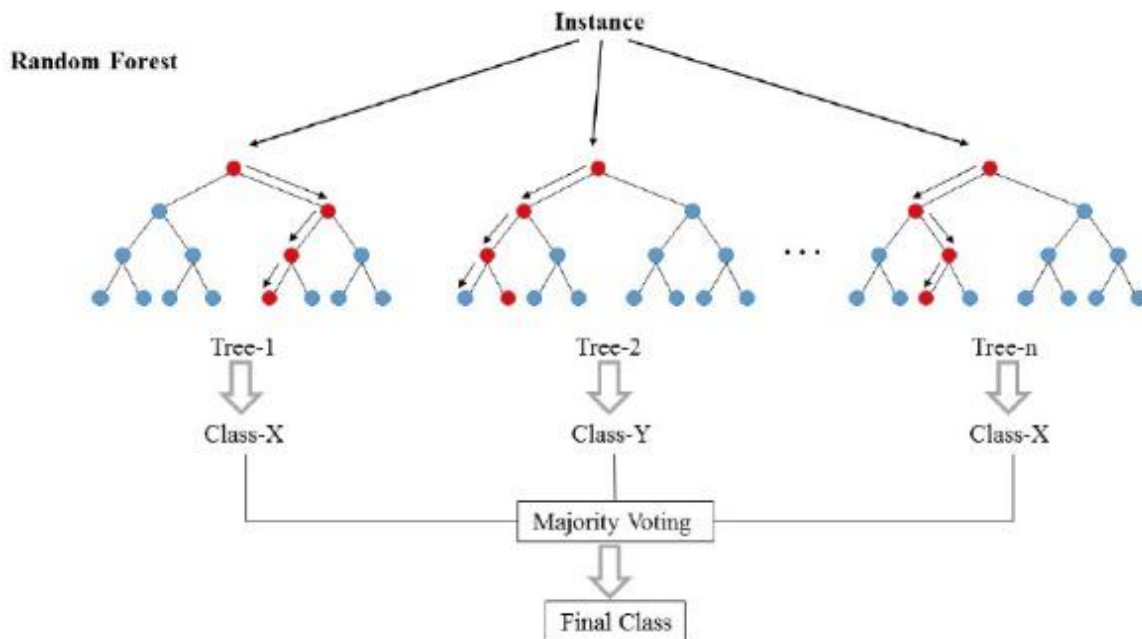
- $g(t)$ - Ο παράγοντας αυτός μοντελοποιεί την τάση, αφού περιγράφει μία μακροπρόθεσμη αύξηση ή μείωση των δεδομένων. Το Prophet ενσωματώνει δύο μοντέλα τάσεων: ένα μοντέλο κορεσμού και ένα μερικώς γραμμικό μοντέλο, ανάλογα με το είδος του προβλήματος πρόβλεψης. Επίσης, εντοπίζει αυτόματα τις αλλαγές της τάσης, εντοπίζοντας και επιλέγοντας τα σημεία αλλαγής από τα συνολικά δεδομένα.
- $s(t)$ - Ο παράγοντας αυτός μοντελοποιεί την εποχικότητα με τη χρήση σειρών Fourier, αφού περιγράφει τον τρόπο με τον οποίο τα δεδομένα επηρεάζονται από εποχικούς παράγοντες. Ένας τέτοιος παράγοντας είναι η χρονική στιγμή του έτους.
- $h(t)$ - Ο παράγοντας αυτός μοντελοποιεί τις επιπτώσεις των διακοπών ή των μεγάλων γεγονότων που επηρεάζουν τις χρονοσειρές των επιχειρήσεων. Παραδείγματα τέτοιων γεγονότων είναι τα Χριστούγεννα, το Πάσχα, η Black Friday. Η λίστα των σημαντικών αργιών ορίζεται χειροκίνητα από τον χρήστη.

- ε - Ο παράγοντας αυτός αντιπροσωπεύει τον όρο του μη αναστρέψιμου σφάλματος.

Τέλος, η μέθοδος Prophet αξιολογεί αυτόματα τις επιδόσεις πρόβλεψης και προειδοποιεί όπου απαιτείται χειροκίνητη παρέμβαση. Η αξιολόγηση της απόδοσης πραγματοποιείται μέσω μια διαδικασίας που ονομάζεται προσομοίωση ιστορικών προβλέψεων (SHF). Η διαδικασία SHF πραγματοποιείται μέσω της παραγωγή K προγνώσεων σε διάφορα χρονικά σημεία της ιστορίας, τα οποία στη συνέχεια ταιριάζουν στο μοντέλο του αναμενόμενου σφάλματος σε διαφορετικούς ορίζοντες πρόβλεψης (Liu, 2018). Αναφορικά με την πρόβλεψη της τιμής του Bitcoin, την καλύτερη επίδοση της Prophet έναντι της ARIMA έρχεται να επιβεβαιώσει η έρευνα των (Yenidoğan κ.ά., 2018).

5.5 Random Forest (RF)

Το μοντέλο Random Forest (RF) είναι μια τεχνική μηχανικής μάθησης που συνδυάζει πολλαπλά δέντρα αποφάσεων (δάσος) κατά τον χρόνο εκπαίδευσης, προκειμένου να καταλήξει στην τελική πρόβλεψη (γνώμη της πλειοψηφίας) ως αποτέλεσμα του μέσου όρου των προβλέψεων των μεμονωμένων δέντρων (Kursa κ.ά., 2011). Για την κατασκευή κάθε δέντρου χρησιμοποιείται διαφορετικό υποσύνολο δεδομένων και σε κάθε βήμα της κατασκευής χρησιμοποιείται τυχαία διαφορετικό σύνολο χαρακτηριστικών. Κάθε κόμβος του δέντρου αποτελεί έναν σταθμό ελέγχου κάποιου χαρακτηριστικού των δεδομένων εκπαίδευσης, έτσι ο διαχωρισμός γίνεται με βάση το καλύτερο χαρακτηριστικό του τυχαίου υποσυνόλου. Τα δυνατά αποτελέσματα αυτού του ελέγχου αποτελούν τα κλαδιά του κόμβου ενώ τα φύλλα του δέντρου αποτελούν καταστάσεις τερματισμού. Ο αλγόριθμος τερματίζει μόλις δοθεί τιμή στο κύριο χαρακτηριστικό (ετικέτα).



Εικόνα 3: Ταξινομητής Random Forest

Το μοντέλο Random Forest μπορεί να χρησιμοποιηθεί τόσο για κατηγοριοποίηση όσο και για παλινδρόμηση, κάνοντας χρήση κάθε φορά διαφορετικού κριτηρίου. Στην παραπάνω εικόνα απεικονίζεται η κατηγοριοποίηση δεδομένων σε δύο κλάσεις X και Y με χρήση του μοντέλου Random Forest. Ιδιαίτερα κατά την πρόβλεψη της τιμής του Bitcoin έχει χρησιμοποιηθεί σε μελέτες όπως των (Madan κ.ά., 2015) και (Velankar κ.ά., 2018).

Στον αλγόριθμο RF της μελέτης της εργασίας αναμένεται να εξεταστούν οι εξής υπερπαράμετροι (CRAN.R, 2020):

- **ntree**: Πρόκειται για τον μέγιστο αριθμό των δέντρων που θα αναπτυχθούν
- **mtry**: Ο αριθμός των μεταβλητών που θεωρήθηκαν τυχαίες σε κάθε διαίρεση.
- Σημειώνεται η δυνατότητα **pruning** του δέντρου όπου στην μελέτη δεν θα χρησιμοποιηθεί.

Στο σημείο αυτό αξίζει να σημειωθεί ότι η τυχαιότητα με την οποία εκπαιδεύονται τα δέντρα του μοντέλου Random Forest το καθιστά ιδιαίτερα αξιόπιστο για δύο λόγους. Αρχικά, είναι πιο ισχυρό από ένα μόνο δέντρο απόφασης και δεύτερον γιατί αποφεύγεται το overfitting. Αντίθετα, τα μοντέλα Gradient Boosting, στην περίπτωση δεδομένων με μεγάλο θόρυβο, είναι ιδιαίτερα ευαίσθητα στο overfitting κατά τη διαδικασία εκμάθησης. Επιπλέον, το Random Forest απαιτεί μεγάλο υπολογιστικό κόστος και χρόνο για μεγάλο αριθμό

δέντρων γεγονός που το κάνει δύσχρηστο για προβλέψεις σε πραγματικό χρόνο. Ωστόσο, είναι κατάλληλο για πολυδιάστατα προβλήματα με έντονα συσχετισμένα χαρακτηριστικά καθώς δίνει ακρίβεια πρόβλεψης. Στην περίπτωση του Gradient Boosting, αν οι παράμετροι οριστούν σωστά, μπορεί να δώσει εξίσου αξιόπιστα ή και καλύτερα αποτελέσματα από το Random Forest και ιδιαίτερα όταν τα δεδομένα είναι μη ισορροπημένα. Τέλος, στα πλαίσια έρευνας της εργασίας αυτής επιλέχθηκε να εξεταστούν όλες οι παραπάνω μέθοδοι αρχικά για εκπαιδευτικούς λόγους και έπειτα γιατί το σύνολο των διαθέσιμων δεδομένων, ως προς τον όγκο και τη μορφή, είναι αρκετά μικρό για να αποτελέσει κριτήριο ξεκάθαρης επιλογής ή απόρριψης ενός ή περισσοτέρων από τα μοντέλα.

5.6 Μοντέλα πρόβλεψης της τιμής του Bitcoin εκτός πλαισίου έρευνας της παρούσας εργασίας

Τα μοντέλα που παρουσιάζονται στο παρόν υποκεφάλαιο, εντοπίζονται σε πληθώρα ερευνών στη διαθέσιμη βιβλιογραφία με εξαιρετικά αποτελέσματα. Παρά την αρχική πρόθεση για την αξιολόγησή τους αυτό δεν κατέστη εφικτό. Ο κυριότερος λόγος υπήρξε η αδυναμία πρόσβασης σε ισχυρούς υπολογιστικούς πόρους. Επιπρόσθετα, λόγω των μικρών σε όγκο δεδομένων η βελτίωση που πιθανόν να εμφάνιζαν δεν θα ήταν τόσο αισθητή. Ο λόγος είναι ότι η επίδοση των νευρωνικών δικτύων είναι στενά συνδεδεμένη με τον μεγάλο όγκο διαθέσιμων δεδομένων. Στην περίπτωση που η μελέτη αφορούσε ωριαία δεδομένα και άρα 24 φορές περισσότερα στο σύνολο, η αξιολόγηση των νευρωνικών δικτύων θα ήταν επιβεβλημένη. Συνεπώς, σκοπός του υποκεφαλαίου είναι μία περισσότερο θεωρητική αναφορά σε αυτά τα μοντέλα ώστε να γίνει καλύτερα κατανοητή η λειτουργία τους.

5.6.1 Recurrent Neural Networks (RNN)

Τα αναδρομικά νευρωνικά δίκτυα (RNN) είναι μία υποκατηγορία νευρωνικών δικτύων με πληθώρα αναφορών στη βιβλιογραφία. Η ιδέα πίσω από τα RNN είναι η χρήση διαδοχικών πληροφοριών. Πιο συγκεκριμένα, τα RNN ονομάζονται αναδρομικά ή επαναλαμβανόμενα επειδή εκτελούν την ίδια εργασία για κάθε στοιχείο μιας ακολουθίας, με την έξοδο να εξαρτάται από τους προηγούμενους υπολογισμούς. Ένας άλλος τρόπος περιγραφής των RNN είναι ως μία κατηγορία νευρωνικών δικτύων με μνήμη ικανή να διατηρεί πληροφορίες για τους υπολογισμούς που έχουν πραγματοποιήσει μέχρι κάποια δεδομένη χρονική στιγμή. Θεωρητικά, τα RNN μπορούν να χρησιμοποιήσουν

πληροφορίες από αυθαίρετα μεγάλες αλληλουχίες, αλλά στην πράξη περιορίζονται στην επιστροφή μόνο μερικών βημάτων.

Ένα νευρωνικό δίκτυο για να λειτουργήσει αποδοτικά είναι ανάγκη να εκπαιδευτεί σωστά. Για να χαρακτηριστεί η εκπαίδευση επιτυχής ή μη πρέπει πρώτα να οριστούν οι παράμετροι που την επηρεάζουν. Στην περίπτωση των NN οι βασικότερες παράμετροι είναι ο αριθμός των κρυφών επιπέδων του δικτύου, το πλήθος των κόμβων του κάθε επιπέδου και ο ρυθμός εκμάθησης.

Στην έρευνα των (McNally κ.ά., 2016) έγινε χρήση του μοντέλου RNN για την πρόβλεψη της τιμής του Bitcoin, όπου μετά την εφαρμογή μεθόδου αυτοσυσχέτισης με μήκος παραθύρου 20 για την εκπαίδευση του μοντέλου, 2 κρυφά επίπεδα 20 νευρώνων και με εφαρμογή έγκαιρου τερματισμού (early stop), κατέληξαν σε υψηλά ποσοστό ακρίβειας της πρόβλεψης έναντι άλλων μοντέλων που εξέτασαν όπως το ARIMA.

5.6.2 Long Short Term Memory (LSTM)

Τα Long Short Term Memory (LSTM) νευρωνικά δίκτυα αποτελούν μια ειδική κατηγορία των νευρωνικών δικτύων RNN τα οποία ειδικεύονται στην εκμάθηση μακροσκελών ακολουθιών. Τα LSTM νευρωνικά δίκτυα προτάθηκαν αρχικά από τους (Hochreiter κ.ά., 1997) ενώ χρησιμοποιήθηκαν και εξελίχθηκαν από πολλούς ερευνητές στην πάροδο των χρόνων. Στις μέρες μας χρησιμοποιούνται σε μία πληθώρα εφαρμογών, μία εκ των οποίων είναι η πρόβλεψη της τιμής του Bitcoin. Τα LSTM νευρωνικά δίκτυα είναι ειδικά σχεδιασμένα για να αποφεύγουν την εξάρτηση που δημιουργείται ανάμεσα στις μακροσκελείς ακολουθίες κατά τη διαδικασία εκμάθησης. Πιο συγκεκριμένα, έχουν σχεδιαστεί ώστε να είναι ικανά να διαθέτουν μεγάλη μνήμη και να διατηρούν μεγάλο όγκο πληροφορίας, η οποία εντοπίζεται σε πολύ μακρινές χρονικές στιγμές. Αυτό αποτελεί και το συγκριτικό τους πλεονέκτημα έναντι των RNN.

Στην έρευνα των (McNally κ.ά., 2016) έγινε χρήση του μοντέλου LSTM για την πρόβλεψη της τιμής του Bitcoin, όπου μετά την εφαρμογή μεθόδου αυτοσυσχέτισης με μήκος παραθύρου 100 για την εκπαίδευση του μοντέλου, 2 κρυφά επίπεδα 20 νευρώνων και με εφαρμογή έγκαιρου τερματισμού (early stop), κατέληξαν σε σφάλμα ελάχιστα μεγαλύτερο από το αντίστοιχο του RNN μοντέλου που είχαν εφαρμόσει. Παρόλα αυτά, το πλεονέκτημα του μοντέλου αυτού ήταν ότι χρειάστηκε λιγότερο χρόνο εκπαίδευσης έναντι του RNN. Στο συμπέρασμα πως το μοντέλο LSTM επιτυγχάνει καλύτερα αποτελέσματα στην πρόβλεψη

της τιμής του Bitcoin έναντι άλλων, όπως το ARIMA, καταλήγει και η έρευνα των (Karakoyun κ.ά., 2018).

6 Σχεδιασμός και Ανάπτυξη Μελέτης - Πρόβλεψη Τιμής του Bitcoin

Το παρόν κεφάλαιο περιλαμβάνει αναλυτική περιγραφή της μελέτης που πραγματοποιήθηκε για την πρόβλεψη της τιμής του Bitcoin. Αρχικά, παρουσιάζονται τα εργαλεία με τα οποία πραγματοποιήθηκε η μελέτη. Στη συνέχεια ακολουθούν τα δεδομένα των χρονοσειρών που ελήφθησαν υπόψη καθώς και η προεπεξεργασία που χρειάστηκε για την ομογενοποίησή τους σε ενιαίο πίνακα δεδομένων. Αμέσως μετά καταγράφονται τα αποτελέσματα της διερευνητικής ανάλυσης των δεδομένων (Exploratory Data Analysis - EDA) πάνω στο σύνολο των δεδομένων και τέλος παρουσιάζονται οι επιδόσεις των μοντέλων πρόβλεψης που συμπεριλήφθηκαν.

6.1 Εργαλεία Υλοποίησης

Για την εκπόνηση της εργασίας έγινε χρήση της γλώσσα προγραμματισμού R και η ανάπτυξη κώδικα πραγματοποιήθηκε μέσω του γραφικού περιβάλλοντος R. Η επιλογή των εργαλείων στηρίχθηκε στην ευκολία που προσφέρουν μέσω των διαθέσιμων βιβλιοθηκών και λειτουργιών στην επεξεργασία δεδομένων και στην παραγωγή αποτελεσμάτων. Στα υποκεφάλαια που ακολουθούν παρουσιάζονται τόσο η γλώσσα προγραμματισμού R όσο και το γραφικό περιβάλλον RStudio.

6.1.1 R

Η γλώσσα προγραμματισμού R είναι μία γλώσσα ανοικτού κώδικα η οποία χρησιμοποιείται κατά κόρον για στατιστικούς σκοπούς και για εφαρμογές γραφικών. Η γλώσσα R είναι αρκετά δημοφιλής στους επιστήμονες που ειδικεύονται στη στατιστική και την εξόρυξη δεδομένων. Μέσω της R μπορεί να αναπτυχθεί λογισμικό στατιστικής και ανάλυσης δεδομένων (CRAN.R, 2020). Στη βιβλιογραφία καταγράφεται συχνή χρήση της R για όλους τους παραπάνω σκοπούς και μαζί με την Python είναι οι de facto γλώσσες προγραμματισμού για ανάλυση δεδομένων.

Η πρώτη ώριμη έκδοση της γλώσσας κυκλοφόρησε το 2000 από τους Ross Ihaka και Robert Gentleman, ερευνητές του Πανεπιστημίου του Auckland. Εκτός των δημιουργών υπήρξε και ένας μεγάλος αριθμός συντελεστών που συνεισέφεραν στην ανάπτυξή της μέσω κώδικα ή δημιουργίας αναφορών σφαλμάτων. Από τον Αύγουστο του 2018 υπεύθυνος για την ανάπτυξη της R είναι η ομάδα R Development Core Team.

Η σχεδίαση της R έχει βαθιές επιρροές από άλλες δύο γλώσσες προγραμματισμού: την S και την Scheme. Ως προς τον τρόπο σύνταξης παραπέμπει στην S αλλά ως προς την αρχιτεκτονική σχεδίαση ακολουθεί αυτή της Scheme.

Η γλώσσα προγραμματισμού R είναι υψηλού επιπέδου και υλοποιείται με τις γλώσσες προγραμματισμού C και Fortran. Παρ' ότι μπορεί να χρησιμοποιηθεί μέσω φλοιού γραμμής εντολών, η συνηθέστερη χρήση της γίνεται μέσω γραφικού περιβάλλοντος. Το πιο δημοφιλές γραφικό περιβάλλον για R είναι το RStudio. Εκτός του RStudio η R εκτελείται πολύ συχνά και μέσω Jupyter Notebooks, μίας ακόμα γραφικής απεικόνισης με μορφή ηλεκτρονικού σημειωματάριου.

Η διανομή της R περιέχει βιβλιοθήκες για ένα μεγάλο πλήθος στατιστικών διαδικασιών. Ανάμεσά τους είναι τα γενικευμένα γραμμικά μοντέλα, τα μη γραμμικά παλινδρομικά μοντέλα, η ανάλυση χρονοσειρών, τα νευρωνικά δίκτυα και άλλα. Τέλος ανάμεσα σε άλλα η R διαθέτει βιβλιοθήκες για γραφική αναπαράσταση δεδομένων. Ορισμένες πολύ χρήσιμες βιβλιοθήκες που χρησιμοποιήθηκαν στην εργασία αυτή τόσο για την ομογενοποίηση των δεδομένων όσο και για την αναπαράσταση και την υλοποίηση ήταν:

- **ggplot2:** Πρόκειται για μία πολύ δημοφιλή βιβλιοθήκη που χρησιμοποιείται για την οπτικοποίηση δεδομένων εκφράζοντας σχέσεις μεταξύ των χαρακτηριστικών των δεδομένων και της γραφικής παράστασης τους
- **forecast:** Περιλαμβάνει μεθόδους και εργαλεία για την προβολή και την ανάλυση προβλέψεων χρονοσειρών
- **gbm:** Βιβλιοθήκη της γλώσσας R για Gradient Boosting Machine
- **xgboost:** Βιβλιοθήκη της γλώσσας R για XGBoost
- **randomForest:** Βιβλιοθήκη της γλώσσας R για Random Forest. Υλοποιεί ταξινόμηση και παλινδρόμηση με βάση ένα δάσος δέντρων χρησιμοποιώντας τυχαίες εισόδους
- **prophet:** Βιβλιοθήκη της γλώσσας R για το μοντέλο Prophet. Χρησιμοποιείται για την πρόβλεψη χρονοσειρών όπου οι μη γραμμικές τάσεις ταιριάζουν με τις ετήσιες, εβδομαδιαίες και την καθημερινή εποχικότητα

6.1.2 RStudio

Το RStudio είναι ένα ολοκληρωμένο περιβάλλον ανάπτυξης (IDE) ανοικτού κώδικα για την ανάπτυξη προγραμμάτων σε κώδικα R. Όπως κάθε IDE, έτσι και το RStudio, είναι μία εφαρμογή που προσφέρει απλοποιημένες λειτουργίες στους προγραμματιστές προκειμένου να μπορούν να αναπτύξουν κώδικα με ευκολία.

Το RStudio αποτελείται από ένα περιβάλλον προβολής και επεξεργασίας πηγαίου κώδικα, αυτοματοποιημένα εργαλεία για μεταγλώττιση και εκτέλεση του κώδικα αλλά και ένα εργαλείο ανίχνευσης και επιδιόρθωσης σφαλμάτων του κώδικα. Επίσης περιλαμβάνει και έναν φλοιό εκτέλεσης εντολών στον διερμηνέα της R.

Το RStudio διαθέτει διάφορες ενδιαφέρουσες λειτουργίες οι οποίες απευθύνονται τόσο σε νέους όσο και σε έμπειρους προγραμματιστές της γλώσσας προγραμματισμού R. Ανάμεσα σε αυτές περιλαμβάνονται η αυτόματη συμπλήρωση κώδικα, η απευθείας εκτέλεση από τον πηγαίο κώδικα, η ιστορικότητα στην επεξεργασία κώδικα αλλά και στην εκτέλεση εντολών φλοιού. Επιπλέον, διαθέτει ευρετήριο για τις βιβλιοθήκες της R και των οδηγιών χρήσης των συναρτήσεών τους ενώ επιτρέπει και την αναζήτηση και εγκατάσταση διαθέσιμων βιβλιοθηκών από το διαδίκτυο (Rstudio.com, 2020).

Τέλος, το RStudio μπορεί να εγκατασταθεί και να χρησιμοποιηθεί σε όλες τις ευρέως διαδεδομένες διανομές Λειτουργικών Συστημάτων (Windows, Mac OS X, Ubuntu, Fedora) ή ακόμα να χρησιμοποιηθεί ως εξυπηρετητής. Στην περίπτωση αυτή το RStudio εγκαθίσταται απομακρυσμένα και οι προγραμματιστές μπορούν να συνδεθούν σε αυτό μέσω του φυλλομετρητή διαδικτύου της αρεσκείας τους.

6.2 Δεδομένα Μελέτης

Σε κάθε μελέτη πρόβλεψης ένα από τα σημαντικότερα στάδια, μετά από αυτό της κατανόησης του προβλήματος, είναι η συλλογή, η προετοιμασία και η οργάνωση των δεδομένων σε ένα ενιαίο σύνολο δεδομένων. Δεδομένα ελλιπή ή κακής ποιότητας τα οποία θα τροφοδοτήσουν ένα πολύ καλό προγραμματιστικά μοντέλο μπορεί να οδηγήσουν σε αποτελέσματα χειρότερα από αυτά ενός απλού μοντέλου που έχει δουλέψει σε καλύτερα δεδομένα. Με γνώμονα ότι τα δεδομένα που θα χρησιμοποιήσουμε στην μελέτη μας πρέπει να συγκεντρωθούν προσεκτικά προχωρήσαμε στη συλλογή των δεδομένων. Στην εργασία αυτή συλλέξαμε δεδομένα από πολλαπλές πηγές. Σχεδόν στο

σύνολο τους τα δεδομένα αποτελούνται από ιστορικές τιμές – χρονοσειρές των παραγόντων εκείνων που βάσει βιβλιογραφίας επηρεάζουν περισσότερο την τιμή του υπό μελέτη κρυπτονομίσματος Bitcoin.

Στο σημείο αυτό είναι σημαντικό να αναφερθούμε στο χρονικό εύρος των δεδομένων που συλλέχθηκαν. Στόχος της μελέτης αυτής ήταν να χρησιμοποιήσει όσο το δυνατόν περισσότερα ιστορικά στοιχεία υπήρχαν διαθέσιμα στο Διαδίκτυο. Ωστόσο, σημαντικός περιοριστικός παράγοντας για το χρονικό εύρος που επιλέχθηκε ήταν και η διάρκεια ζωής των διαθέσιμων παρατηρήσεων. Παραδείγματος χάριν, ενώ για το πετρέλαιο ή τα πολύτιμα μέταλλα τα διαθέσιμα ιστορικά δεδομένα μπορεί να εκτείνονται δεκαετίες πίσω στον χρόνο, για τα κρυπτονομίσματα οι παρατηρήσεις αυτές είναι πολύ πιο λίγες. Έτσι το χρονικό εύρος ορίστηκε ως το σημείο τομής αυτών των παρατηρήσεων. Έχοντας ως κριτήριο την ύπαρξη δεδομένων για όλα τα μεγέθη για όλες τις παρατηρήσεις που θα χρησιμοποιηθούν, η **07/08/2015** επιλέχθηκε ως η εκκίνηση του χρονικού διαστήματος του υπό μελέτη dataset. Εξαιτίας της μεταβλητότητας και της αβεβαιότητας που εισήχθη κατά το τρέχον έτος από το ξέσπασμα της πανδημίας του κορωνοϊού αποφασίστηκε το dataset να μην περιλαμβάνει δεδομένα του 2020 (WHO, 2020). Ως εκ τούτου, η **31/12/2019** επιλέχθηκε ως η τελευταία ημερομηνία παρατηρήσεων του υπό μελέτη dataset. Τα δεδομένα μας αφορούν σχεδόν μια ολόκληρη πενταετία με συνολικά **1.608 ημερήσιες** παρατηρήσεις. Οι τιμές του Bitcoin και των λοιπών κρυπτονομισμάτων καθώς και των χρηματιστηριακών δεικτών αναφέρονται σε **τιμές κλεισίματος ανά ημέρα**.

Τα δεδομένα αυτά χρησιμοποιήθηκαν, όπως θα αναλυθεί παρακάτω, τόσο για την αξιολόγηση των μοντέλων όσο και για την εκπαίδευσή τους. Στην περίπτωση που η μελέτη πραγματοποιούνταν σε πραγματικό χρόνο δεν θα είχαμε μελλοντικά διαθέσιμα ούτε για το Bitcoin ούτε για τις υπόλοιπες μεταβλητές εισόδου των μοντέλων μας και άρα θα έπρεπε να κάνουμε πρόβλεψη για όλα αυτά.

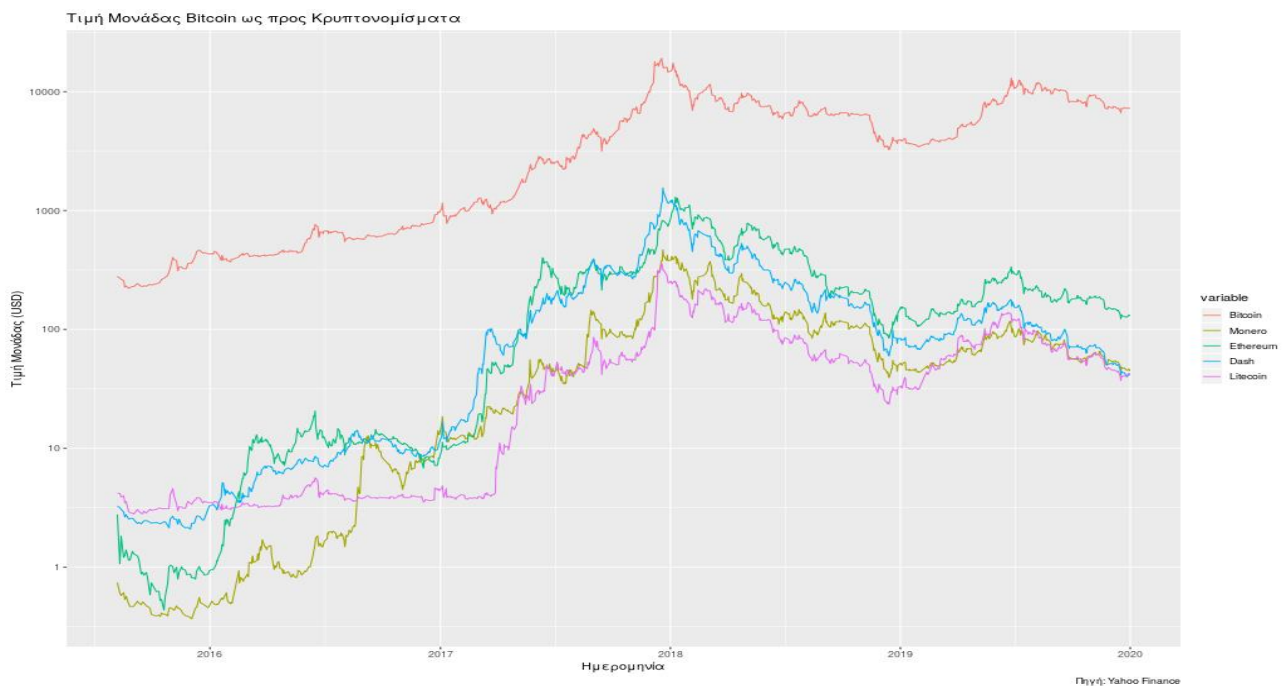
6.2.1 Πηγές Άντλησης Δεδομένων

Βασική πηγή άντλησης δεδομένων υπήρξε η οικονομική σελίδα της Yahoo (Yahoo! Finance, 2020). Αν και το Yahoo Finance διαθέτει πληθώρα δεδομένων δεν διατίθενται όλα δωρεάν. Πιο συγκεκριμένα, για τις τιμές του Κινεζικού Δείκτη τα δωρεάν δεδομένα του Yahoo Finance είναι αρκετά ελλιπή και παρουσιάζουν κενές τιμές. Για τον δείκτη αυτό τα δεδομένα αντλήθηκαν από την ιστοσελίδα Investing.com (Shanghai Composite, 2020).

Από τον ιστότοπο της ηλεκτρονικής εγκυκλοπαίδειας αντλήθηκαν οι ημερήσιες επισκέψεις χρηστών στο λήμμα Bitcoin μέσω της ανοικτής διεπαφής που διαθέτει για άντληση τέτοιων δεδομένων (Wikipedia Bitcoin Views, 2020).

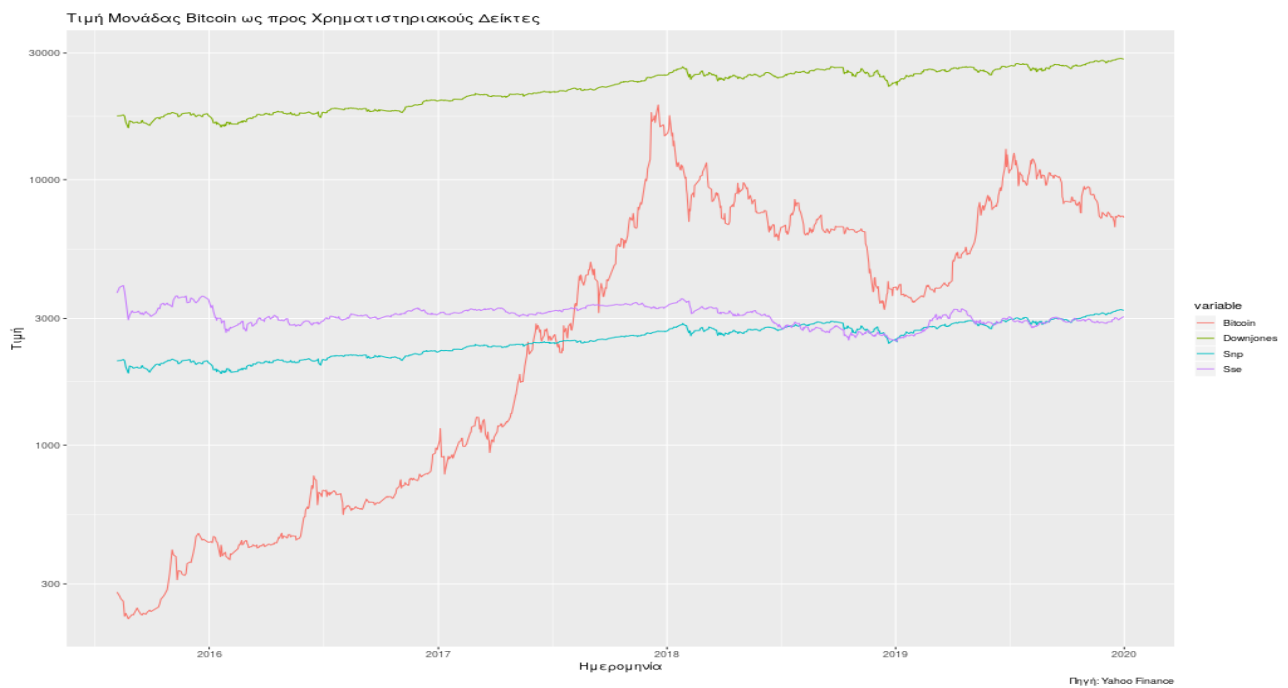
Από την ιστοσελίδα Quandl.com (Hashrate, 2020) αντλήθηκαν δεδομένα για το ημερήσιο Hashrate του δικτύου Bitcoin, τα οποία και είναι αρκετά σημαντικά καθώς επηρεάζουν κατά πολύ την τιμή μονάδας του. Ανάμεσα στο πλήθος οικονομικών δεδομένων που διαθέτει το Quandl, περιέχει και στατιστικά για την τρέχουσα κατάσταση του δικτύου του Bitcoin όπως το διαθέσιμο Market Cap, το τελευταίο hash που παράχθηκε και άλλα.

Στο παρακάτω γράφημα παρατηρούμε τη συσχέτιση μεταξύ της τιμής του Bitcoin και των τιμών των άλλων κρυπτονομισμάτων. Όπως μπορούμε να παρατηρήσουμε οι τιμές των άλλων κρυπτονομισμάτων φαίνεται να ακολουθούν τις έντονες διακυμάνσεις του Bitcoin, αν και είναι πολύ χαμηλότερες σε αξία.



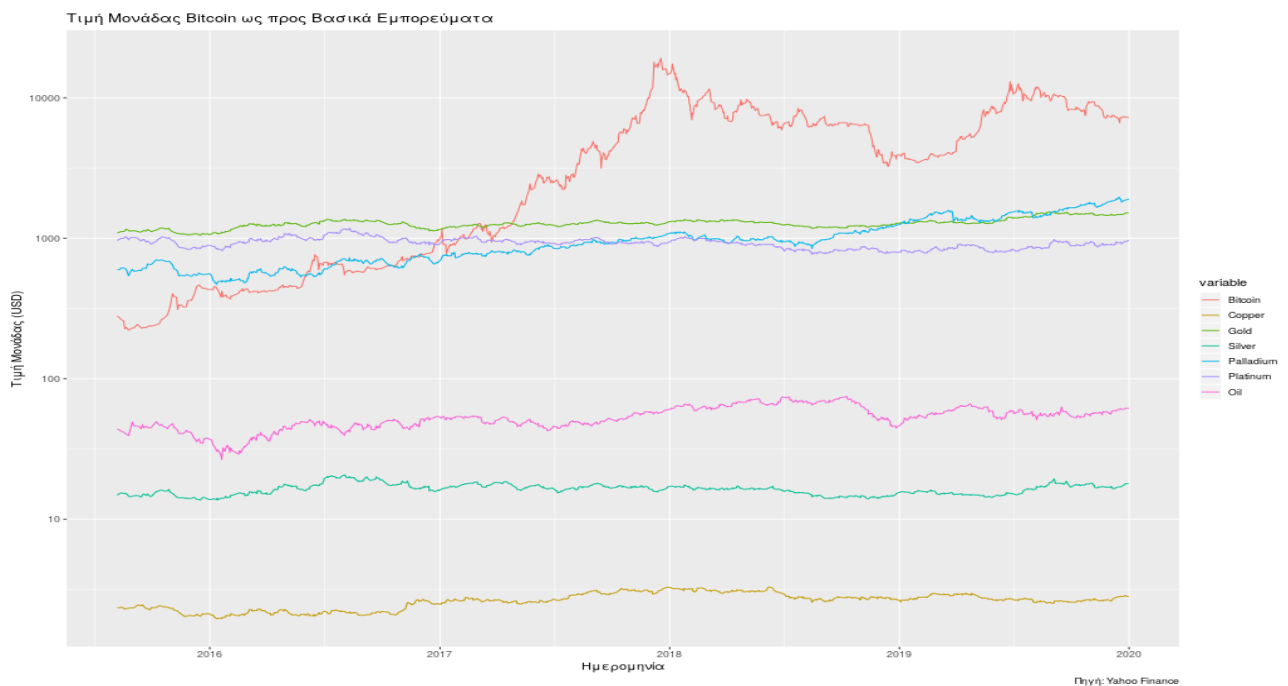
Γράφημα 2: Ιστορικότητα Κρυπτονομισμάτων

Στο παρακάτω γράφημα παρατηρούμε τη συσχέτιση μεταξύ της τιμής του Bitcoin και των τιμών κλεισίματος των χρηματιστηριακών δεικτών. Η μόνη ίσως εμφανής συσχέτιση φαίνεται να είναι με τον δείκτη Down Jones (πράσινη γραμμή) όπου οι διακυμάνσεις τους ακολουθούν όμοια πορεία.



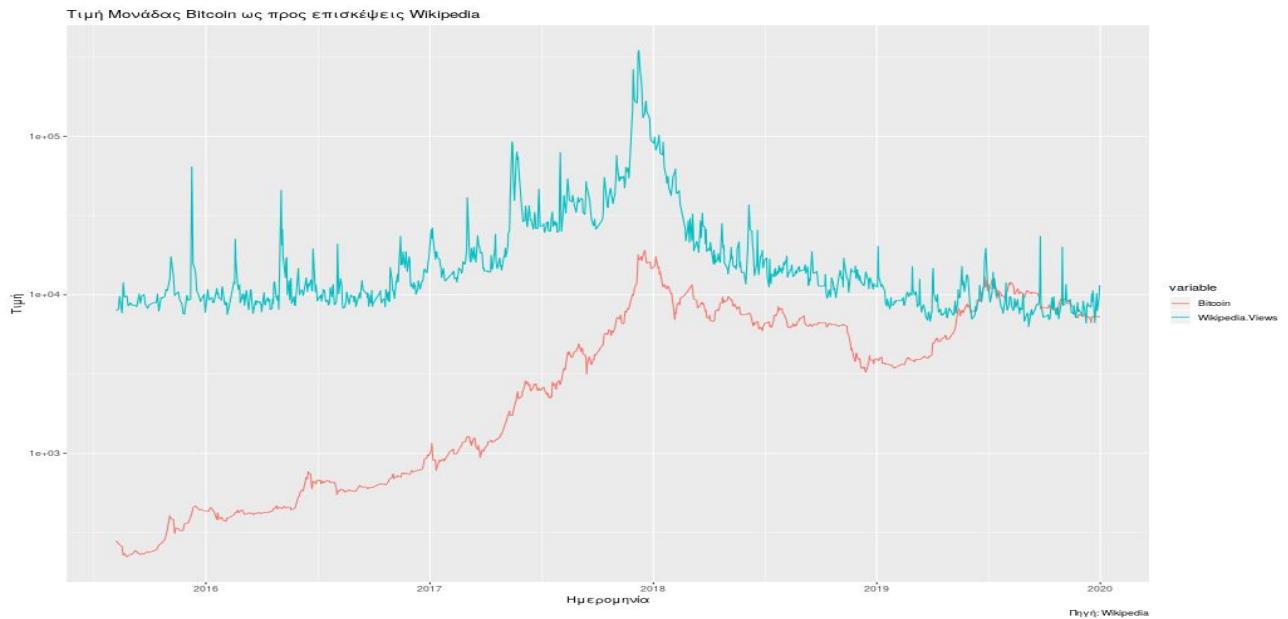
Γράφημα 3: Ιστορικότητα Χρηματιστηριακών Δεικτών και Bitcoin

Στο παρακάτω γράφημα απεικονίζεται η συσχέτιση μεταξύ της τιμής του Bitcoin και των τιμών κλεισίματος των βασικών εμπορευμάτων και πολύτιμων μετάλλων, παρόλο που δεν προκύπτει με μια πρώτη ματιά εμφανής συσχέτιση.



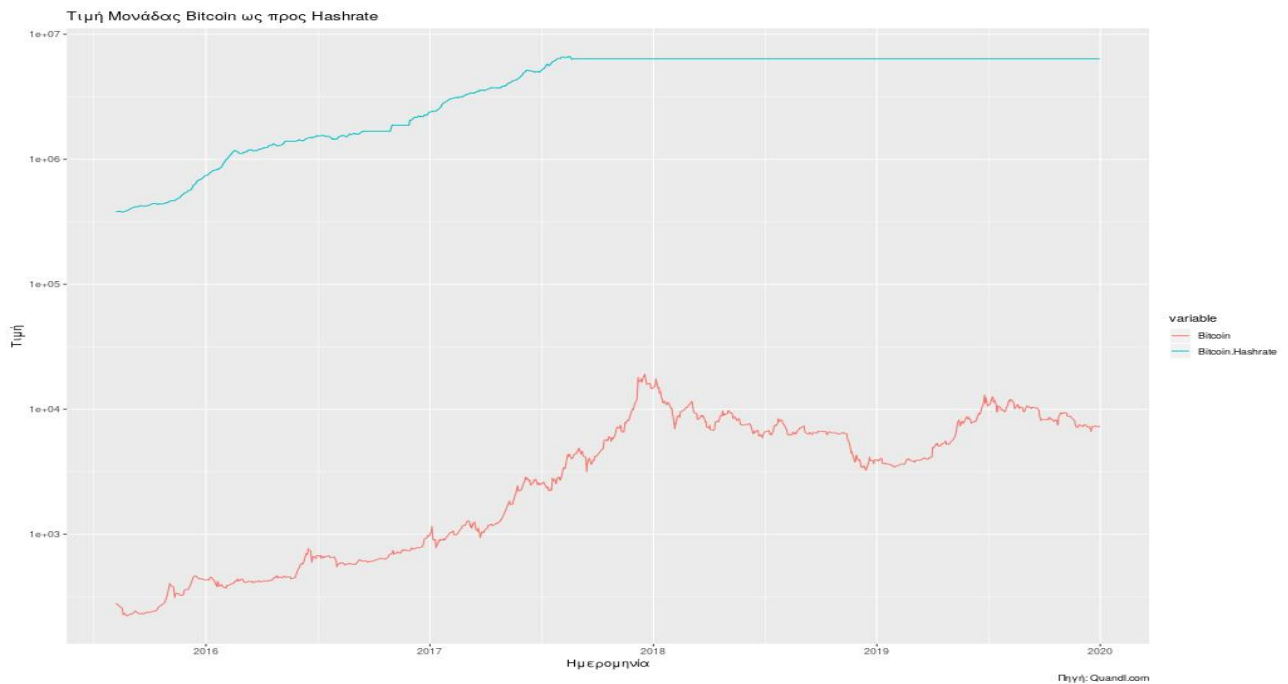
Γράφημα 4: Ιστορικότητα τιμών εμπορευμάτων και Bitcoin

Στο παρακάτω γράφημα παρατηρούμε τη συσχέτιση μεταξύ της τιμής του Bitcoin και των ημερήσιων επισκέψεων στο λήμμα Bitcoin στο Wikipedia. Η συσχέτιση μεταξύ των δύο φαίνεται να είναι έντονη. Η γραφική απεικόνιση των τιμών του Bitcoin φαίνεται να είναι μετατοπισμένη προς τα δεξιά, το οποίο είναι εύλογο αφού προηγείται η επίσκεψη και διαμόρφωση γνώμης στο Wikipedia από την απόφαση για αγορά ή πώληση μονάδων Bitcoin.



Γράφημα 5: Ιστορικότητα Επισκέψεων Wikipedia και τιμών Bitcoin

Στο παρακάτω γράφημα παρατηρούμε τη συσχέτιση μεταξύ της τιμής του Bitcoin και του συνολικού ημερήσιου Hashrate του δικτύου του Bitcoin.



Γράφημα 6: Ιστορικότητα Τιμής Bitcoin ως προς το Hashrate

6.2.2 Προεπεξεργασία

Έχοντας συλλέξει τις χρονοσειρές στην πρωτότυπη μορφή τους το επόμενο βήμα είναι η προεπεξεργασία των δεδομένων ώστε να μπορούν να χρησιμοποιηθούν για τη μελέτη της τιμής του Bitcoin. Η προεπεξεργασία πραγματοποιήθηκε μέσω της πλατφόρμας RStudio και με κώδικα γραμμένο σε γλώσσα προγραμματισμού R.

Δεδομένα Yahoo Finance

Τα δεδομένα από το Yahoo Finance αποτέλεσαν την πλειοψηφία των δεδομένων του κοινού dataset. Πέραν λίγων χρονοσειρών, όλες οι υπόλοιπες αντλήθηκαν από το Yahoo Finance. Στη συγκεκριμένη ιστοσελίδα παρέχεται η δωρεάν δυνατότητα εξαγωγής χρονοσειρών διάφορων οικονομικών δεικτών σε μορφή CSV και για προκαθορισμένο χρονικό διάστημα. Το αρχικό στάδιο προεπεξεργασίας ήταν για κάθε μία από τις χρονοσειρές που χρησιμοποιήθηκαν για την εργασία η εξαγωγή τους σε CSV και η αρχειοθέτησή τους σε κατάλληλη δενδρική δομή.

Στη συνέχεια και σαν δεύτερο στάδιο προεπεξεργασίας, με επαναληπτική διαδικασία σε κάθε ένα από τα αρχεία έγινε εξαγωγή της στήλης **τιμής κλεισίματος** για κάθε ημερήσια παρατήρηση.

Το τελευταίο στάδιο της προεπεξεργασίας ήταν η δημιουργία ενός ενιαίου αρχείου όπου η πρώτη στήλη αναπαριστά την ημερομηνία της εκάστοτε παρατήρησης και η κάθε επόμενη αντιπροσωπεύει την τιμή κλεισίματος για κάθε ένα από τα οικονομικά μεγέθη. Σε αυτό το κοινό αρχείο το όνομα της στήλης κάθε ενός αγαθού προέκυψε από το όνομα του αρχείου, αφαιρώντας τις κάτω παύλες και μετατρέποντας το πρώτο γράμμα σε κεφαλαίο.

Δεδομένα Κινεζικού Δείκτη

Όπως προαναφέρθηκε, τα δεδομένα του Κινεζικού Δείκτη προήλθαν από άλλη πηγή και πιο συγκεκριμένα την ιστοσελίδα Investing.com. Αν και η συγκεκριμένη χρονοσειρά ήταν ετήσια αυτή είχε διαφορετική μορφή.

Προκειμένου η επαναληπτική διαδικασία για τα μεγέθη των χρονοσειρών από το Yahoo Finance να συμπεριλάβει με ενιαίο τρόπο και τον Κινεζικό Δείκτη, τα δεδομένα του δεύτερου χρειάστηκαν μία επιπλέον προεπεξεργασία. Πιο συγκεκριμένα, δύο χαρακτηριστικά χρειάστηκαν τροποποίηση. Αρχικά, σε αντίθεση με τα δεδομένα του Yahoo Finance, αυτά του Κινεζικού Δείκτη ήταν αποθηκευμένα από το πιο πρόσφατο στο πιο παλιό. Συνεπώς, η σειρά τους αντιστράφηκε. Κατά δεύτερον, η στήλη της ημερομηνίας ήταν αποθηκευμένη σε διαφορετική αναπαράσταση. Με τη βοήθεια των έτοιμων βιβλιοθηκών της R, η ημερομηνία μετατράπηκε στην ίδια μορφή με τα δεδομένα του Yahoo Finance.

Έχοντας πραγματοποιήσει την παραπάνω προεπεξεργασία και τοποθετώντας το αρχείο του δείκτη εντός της ιεραρχίας των χρονοσειρών από το Yahoo Finance στη θέση `stock_exchange/sse.csv` πραγματοποιήθηκε ενιαία επεξεργασία βάσει των όσων περιγράφονται στο προηγούμενο υποκεφάλαιο.

Δεδομένα Επισκέψεων Wikipedia

Για την εξαγωγή των επισκέψεων στο λήμμα Bitcoin της ηλεκτρονικής εγκυκλοπαίδειας Wikipedia χρειάστηκε μία τελείως διαφορετική προσέγγιση αναφορικά με τους οικονομικούς δείκτες. Πιο συγκεκριμένα, οι ημερήσιες επισκέψεις όπως και άλλα στατιστικά για το εκάστοτε λήμμα, είναι διαθέσιμα από το Wikipedia δωρεάν, μέσω ειδικής διεπαφής.

Η διεπαφή του Wikipedia για τη λήψη ημερήσιων δεδομένων επισκεψιμότητας απαιτεί επικοινωνία μέσω πρωτοκόλλου HTTP και με την εισαγωγή του συνδυασμού λήμματος

(Bitcoin), ημερομηνίας Έναρξης (StartDate) και ημερομηνίας λήξης (EndDate) ο χρήστης λαμβάνει ημερήσια δεδομένα επισκεψιμότητας.

Η μορφή των αποτελεσμάτων που είναι διαθέσιμα μέσω της επαφής είναι μορφότυπος JSON και η μετατροπή τους σε dataframe δομή δεδομένων έγινε με χρήση της συνάρτησης `Idply` της R. Ύστερα από την επιτυχή μετατροπή τους και κατάλληλο μετασχηματισμό της στήλης ημερομηνίας (Date), η στήλη που χρησιμοποιήθηκε ήταν αυτή των επισκέψεων (views).

Τελευταίο βήμα της προεπεξεργασίας των δεδομένων επισκεψιμότητας ήταν η εισαγωγή τους στο κοινό αρχείο του dataset ως ανεξάρτητη στήλη με τίτλο Wikipedia Views.

Δεδομένα Bitcoin Hashrate

Όπως και με τα δεδομένα επισκεψιμότητας στο Wikipedia, έτσι και με το Hashrate του δικτύου του Bitcoin, η προεπεξεργασία ήταν διαφορετική. Αρχικά, πραγματοποιήθηκε αύξουσα ταξινόμηση βάσει ημερομηνίας και μετατροπή της αναπαράστασής της στην ενιαία μορφή των υπόλοιπων δεδομένων.

Έχοντας επεξεργαστεί κατάλληλα τα δεδομένα από το Quandl, η στήλη που χρησιμοποιήθηκε ήταν η Network.Hashrate.Terhashs, η οποία περιλαμβάνει τα ημερήσια THash που παράγονται από το συνολικό δίκτυο Bitcoin. Ως τελευταίο βήμα προεπεξεργασίας, το ημερήσιο Hashrate εισάχθηκε στο κοινό αρχείο του dataset με τίτλο Bitcoin Hashrate.

Διαχείριση κενών τιμών

Έχοντας δημιουργήσει ένα κοινό dataset με όλες τις χρονοσειρές που θα χρησιμοποιηθούν για τις ανάγκες πρόβλεψης της τιμής του Bitcoin και προτού προχωρήσουμε στη μελέτη χρειάστηκε να γίνει συμπλήρωση των κενών τιμών που παρουσίαζαν ορισμένες από τις χρονοσειρές.

Πιο συγκεκριμένα, τόσο οι τιμές των commodities όσο και οι τιμές των χρηματιστηριακών δεικτών παρουσιάζουν εκ φύσεως κενές τιμές. Εν αντιθέσει με τα κρυπτονομίσματα, η παραγωγή και η ανταλλαγή των οποίων πραγματοποιείται 24/7, οι μετοχές αλλά και οι τίτλοι κτήσης των commodities ανταλλάσσονται εντός του χρόνου λειτουργίας των χρηματιστηριακών αγορών. Όπως είναι φυσικό τα δεδομένα αυτών των χρονοσειρών

εμφανίζουν κενές τιμές κάθε Σάββατο και Κυριακή όπως και τις ημέρες των Εθνικών Αργιών. Οι μέρες στις οποίες δεν καταγράφονται παρατηρήσεις ήταν περίπου το **30%** των παρατηρήσεων. Για τη συμπλήρωσή τους έγινε χρήση της τελευταίας καταγεγραμμένης παρατήρησης πριν την εκάστοτε κενή τιμή. Κατά αυτό τον τρόπο, οι κενές τιμές του Σαββάτου όπως και της Κυριακής συμπληρώθηκαν με την τιμή της Παρασκευής που προηγείται αυτών. Αντίστοιχα, για τις εθνικές αργίες, γίνεται η χρήση της πιο πρόσφατης προηγούμενης τιμής.

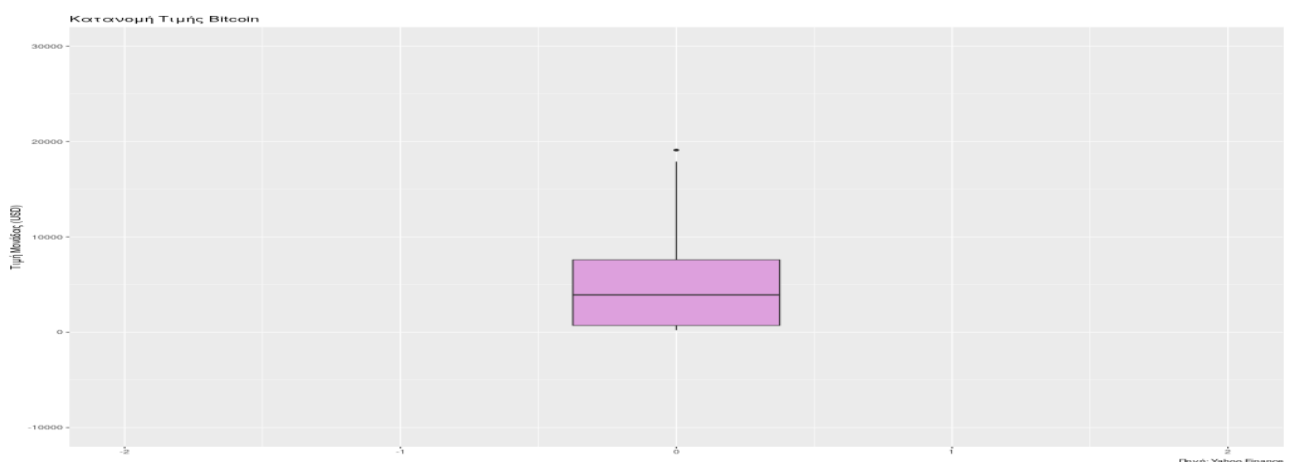
6.3 Διερευνητική Ανάλυση Δεδομένων - EDA

Η διερευνητική ανάλυση δεδομένων αλλιώς Exploratory Data Analysis (EDA) αποτελεί βασικό κομμάτι της μελέτης ενός μεγέθους. Μέσα από την EDA, μπορεί κανείς να κατανοήσει καλύτερα το υπό μελέτη μέγεθος, να εντοπίσει τα χαρακτηριστικά που το αποτελούν και να μελετήσει την αλληλεπίδρασή του με άλλα μεγέθη.

Η διερεύνηση ξεκινάει με τη γραφική απεικόνιση των πραγματικών τιμών ως προς τον χρόνο. Από τη γραφική απεικόνιση είναι πιθανόν ακόμα και ένας μη ειδικός να παρατηρήσει τα ποιοτικά χαρακτηριστικά της χρονοσειράς όπως τάση, εποχικότητα, τυχαιότητα, ασυνέχειες ή ακραίες μεταβολές. Στις υποενότητες που ακολουθούν παρουσιάζονται κάποια βασικά χαρακτηριστικά που προκύπτουν από την EDA.

6.3.1 Στατιστικά στοιχεία

Πρώτο βήμα στην κατανόηση της τιμής κλεισίματος του Bitcoin σαν οικονομικό μέγεθος είναι η μελέτη της κατανομής της. Με τη βοήθεια της R, προκύπτει το παρακάτω γράφημα:



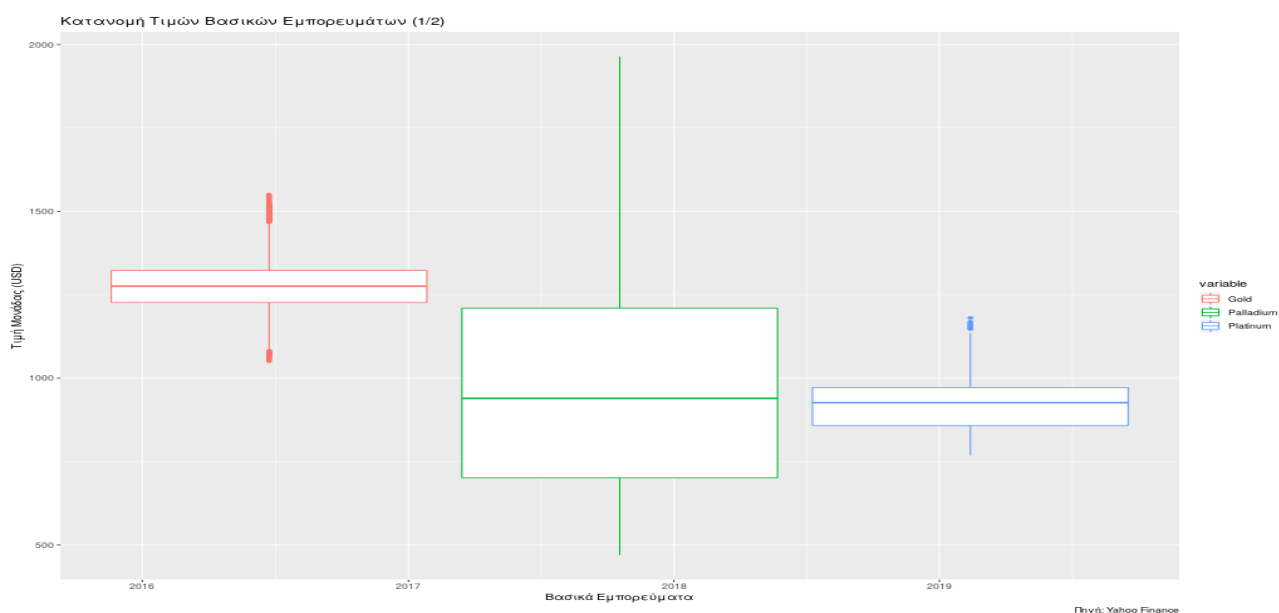
Γράφημα 7: Κατανομή Τιμών Τιμής Bitcoin

Όπως μπορεί να γίνει εύκολα αντιληπτό η τιμή του Bitcoin παρουσιάζει μία σχετική συγκέντρωση μεταξύ ~700 και ~7,500 δολάρια. Επιπλέον όμως, ενώ η ελάχιστη τιμή δεν απέχει πολύ από το πρώτο τεταρτημόριο, η μέγιστη τιμή είναι υπερδιπλάσια από το τρίτο. Τα παραπάνω μπορούν να γίνουν αντιληπτά στην έκφραση του γραφήματος σε αριθμούς, όπως αυτοί παρουσιάζονται στον παρακάτω πίνακα:

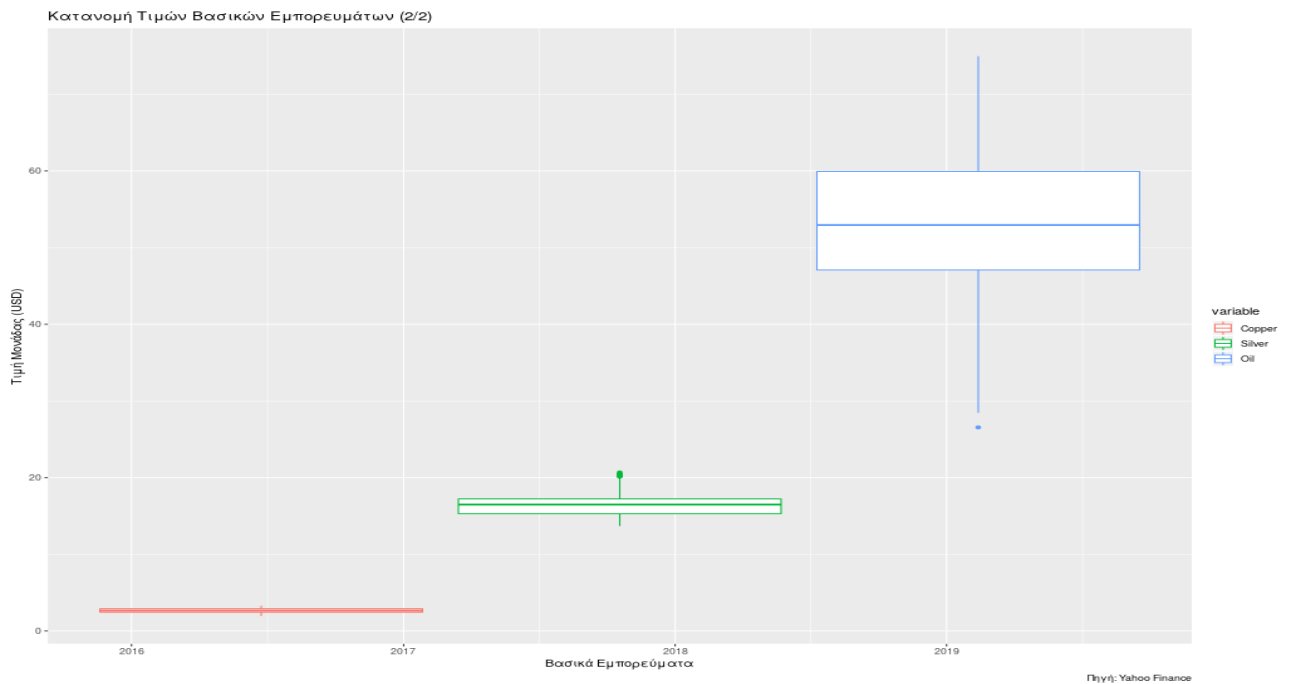
Ελάχιστη Τιμή	Πρώτο Τεταρτημόριο	Διάμεσος Τιμή	Μέση Τιμή	Τρίτο Τεταρτημόριο	Μέγιστη Τιμή
221,6	719,4	3924,4	4647,1	7604,3	19114,2

Πίνακας 1: Κατανομή τιμής Bitcoin

Στα παρακάτω γραφήματα παρουσιάζονται οι κατανομές των τιμών κλεισίματος των βασικών εμπορευμάτων που μελετήθηκαν. Μεταξύ τους παρατηρείται αρκετή ανομοιογένεια.

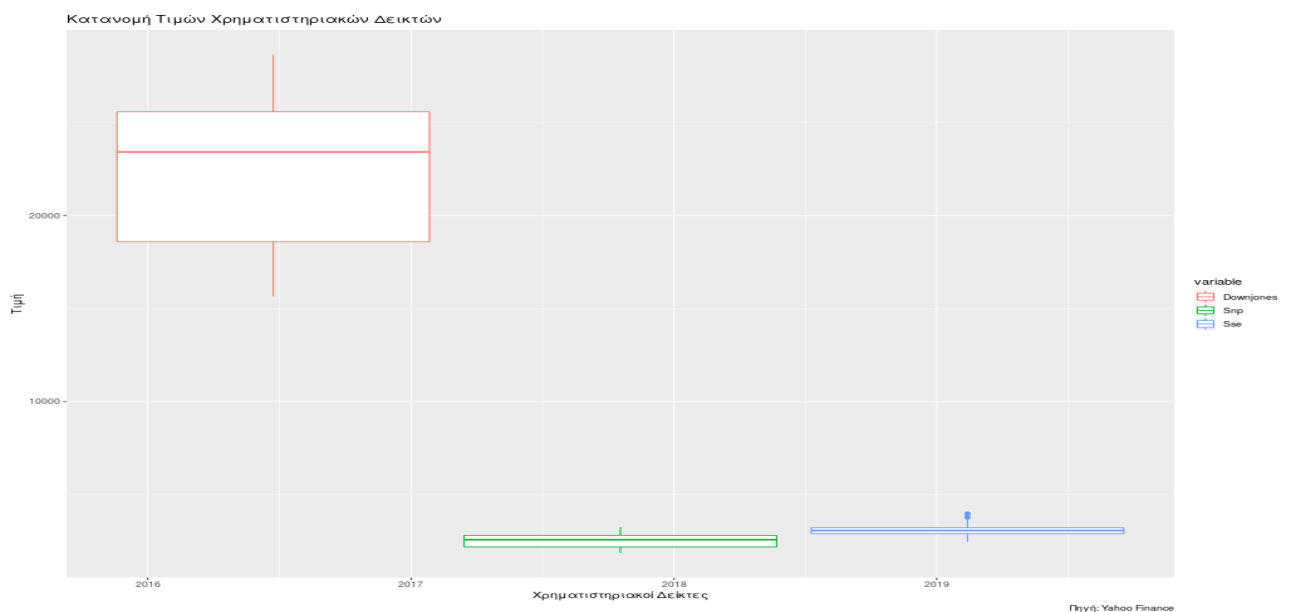


Γράφημα 8: Κατανομή Τιμών Βασικών Εμπορευμάτων (1/2)



Γράφημα 9: Κατανομή Τιμών Βασικών Εμπορευμάτων (2/2)

Τέλος, στο παρακάτω γράφημα παρουσιάζονται οι κατανομές των τιμών κλεισίματος των χρηματιστηριακών δεικτών που μελετήθηκαν:

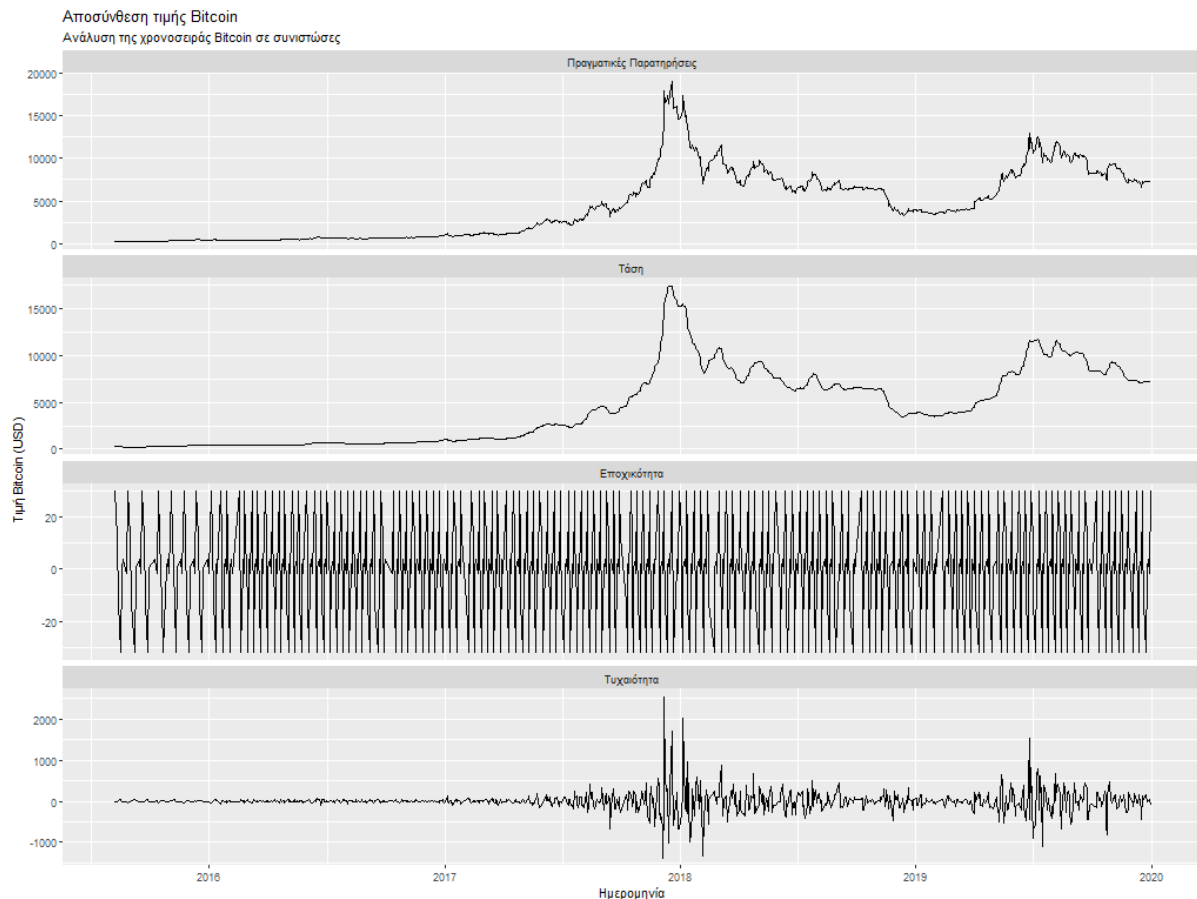


Γράφημα 10: Κατανομή Χρηματιστηριακών Δεικτών

6.3.2 Αποσύνθεση

Επόμενο στοιχείο που έρχεται να μελετήσει η EDA είναι οι συνιστώσες της χρονοσειράς ημερήσιων δεδομένων τιμής του Bitcoin. Μέσω της μεθόδου αποσύνθεσης η χρονοσειρά

του Bitcoin, μπορεί να αναλυθεί περαιτέρω και να υπάρξει καλύτερη κατανόηση ως προς τα χαρακτηριστικά της. Παρακάτω παρατίθενται η αποσύνθεση της χρονοσειράς της τιμής κλεισίματος του Bitcoin στα χαρακτηριστικά της με συχνότητα 7 ημερών.



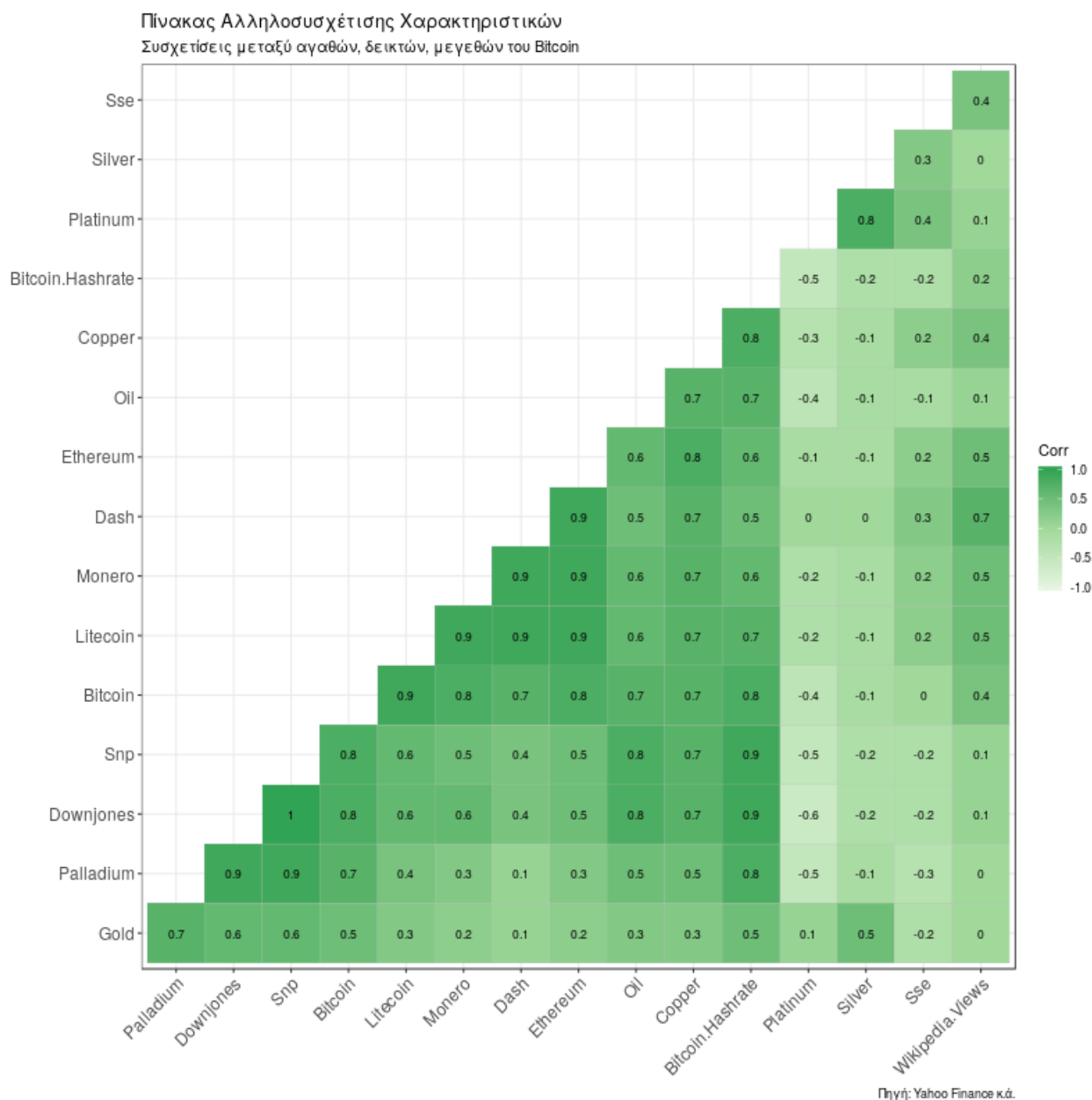
Γράφημα 11: Αποσύνθεση τιμής Bitcoin

Από το γράφημα προκύπτουν τα ακόλουθα συμπεράσματα:

- **Τάση:** Παρατηρείται μία εκθετική τάση, η οποία ξεκινάει στις αρχές του 2017 και σταθεροποιείται το πρώτο τρίμηνο του 2018. Στη συνέχεια η χρονοσειρά εμφανίζει φθίνουσα τάση για περίπου έναν χρόνο και στο τελευταίο τμήμα του γραφήματος είναι πάλι ανοδική.
- **Εποχικότητα:** Από την αποσύνθεση δεν προκύπτει κάποια εποχικότητα.
- **Τυχασιότητα:** Παρατηρούνται ανωμαλίες ως προς την τυχασιότητα της εξέλιξης της τιμής του Bitcoin. Πιο συγκεκριμένα, το τελευταίο τρίμηνο του 2018 έως το πρώτο τρίμηνο του 2019 παρατηρείται μία εκρηκτική αύξηση. Επίσης σημαντική αύξηση της τυχασιότητας παρατηρείται στα μέσα του 2019 η οποία στη συνέχεια εξομαλύνεται.

6.3.3 Πίνακας Αλληλοσυσχέτισης (Correlation Matrix)

Πολύ σημαντικός για την επιλογή των κατάλληλων παραμέτρων εισόδου των μοντέλων πρόβλεψης είναι ο πίνακας αλληλοσυσχέτισης. Ο πίνακας αυτός παρουσιάζει τη συσχέτιση μεταξύ δύο χαρακτηριστικών εντός του συνόλου των παρατηρήσεων ενός πίνακα δεδομένων (dataframe). Προκειμένου να διερευνηθούν οι συσχετίσεις των ανεξάρτητων μεταβλητών (αναφερόμαστε στο σύνολο των διαθέσιμων ανεξάρτητων μεταβλητών που αρχικά εξετάστηκαν), οι οποίες αναφέρθηκαν σε προηγούμενο κεφάλαιο, με την τιμή του Bitcoin, δημιουργήθηκε με τη βοήθεια της R ο πίνακας αλληλοσυσχέτισης για το υπό μελέτη dataframe. Για την παραγωγή του πίνακα που ακολουθεί έγινε χρήση της μεθόδου Pearson.



Γράφημα 12: Αλληλοσυσχέτιση χαρακτηριστικών dataset

Οι τιμές των κελιών του διαγράμματος αποτελούν τους συντελεστές συσχέτισης μεταξύ των μεταβλητών. Όσο μεγαλύτερος ο συντελεστής τόσο μεγαλύτερη η συσχέτιση. Βάσει του γραφήματος, παρατηρούμε πως ορισμένα χαρακτηριστικά παρουσιάζουν ισχυρή αλληλοσυσχέτιση με την τιμή του Bitcoin και άλλα όχι. Πολύ υψηλή είναι η συσχέτιση του Bitcoin με όλα τα κρυπτονομίσματα και ιδιαίτερα με το Litecoin (**0.9**). Επιπλέον, πολύ ισχυρή είναι και η συσχέτιση με το Hashrate (**0.8**), κάτι απολύτως αναμενόμενο βάσει της βιβλιογραφίας. Ενδιαφέρον παρουσιάζουν τα αναπάντεχα αποτελέσματα σχετικά τόσο με το Πετρέλαιο και τον Χαλκό όσο και με τις επισκέψεις στο λήμμα της Wikipedia. Αναφορικά με τα πρώτα, παρατηρείται μία καλή συσχέτιση (**0.7**) ενώ με τις επισκέψεις στο Wikipedia

η συσχέτιση είναι αρκετά μικρή (0.4).

Ειδικά η συσχέτιση με τις επισκέψεις στο Wikipedia αναμενόταν ιδιαίτερα μεγάλη βάσει της βιβλιογραφίας, κάτι που δεν εντοπίστηκε. Μία πιθανή εξήγηση για το παραπάνω είναι με την πάροδο των ετών και με τα περισσότερα διαθέσιμα δεδομένα, η άλλοτε ισχυρή συσχέτιση να έχει αρχίσει πλέον να φθίνει.

6.4 Αξιολόγηση Μοντέλων

Όπως έχει ήδη αναφερθεί στόχος της παρούσας εργασίας είναι τόσο η αξιολόγηση των μοντέλων πρόβλεψης και η επιλογή του καλύτερου από αυτά με βάση τη μετρική επίδοσης όσο και η ίδια η πρόβλεψη της τιμής του Bitcoin στους τρεις ορίζοντες που έχουμε αναφέρει.

Για τις ανάγκες αυτής της μελέτη θα χρησιμοποιήσουμε τον δείκτη της ρίζας του μέσου τετραγωνικού σφάλματος (Root Mean Squared Error) η οποία δίνεται από τον ακόλουθο τύπο:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} * \sum_{i=1}^n (Y_i - F_i)^2}$$

Όπου:

Y_i = Πραγματική τιμή

F_i = Πρόβλεψη

n = Πλήθος τιμών πραγματικής χρονοσειράς για τις οποίες έχουμε κάνει πρόβλεψη

Πρόκειται για ένα μέτρο ακρίβειας της πρόβλεψης και δεδομένου ότι το σφάλμα τετραγωνίζεται προκύπτει εύλογα το συμπέρασμα ότι δίνει μεγαλύτερη βαρύτητα σε μεγάλες αποκλίσεις από την πραγματική τιμή και μικρότερη σε μικρές αποκλίσεις. Η διαφορά του από το μέσο τετραγωνικό σφάλμα είναι ότι είναι εκφρασμένο στις ίδιες μονάδες μέτρησης με την αρχική χρονοσειρά.

Στην περίπτωση της αξιολόγησης χρησιμοποιήθηκε η μέθοδος της αλυσιδωτής επαλήθευσης. Έστω $train=100$ παρατηρήσεις και $validation=20$ παρατηρήσεις με ορίζοντα πρόβλεψης 4. Σύμφωνα με τη μέθοδο αυτή το μοντέλο εκπαιδεύεται στις

παρατηρήσεις του train και προβλέπει για το μήκος (4) του ορίζοντα που έχει επιλεγεί. Έπειτα, έχοντας τις τιμές της πρόβλεψης και τις τιμές του validation υπολογίζει το σφάλμα RMSE1 της πρώτης επανάληψης. Στη δεύτερη επανάληψη προσθέτει στο train τις πρώτες 4 παρατηρήσεις του validation, εκπαιδεύεται στο νέο train (104 παρατηρήσεις) και προβλέπει για τις επόμενες 4 παρατηρήσεις υπολογίζοντας το RMSE2 κοκ. Οι επαναλήψεις θα είναι τόσες ώστε να ενσωματωθεί όλο το validation στο train (εδώ 5). Το τελικό σφάλμα RMSE προκύπτει ως ο μέσος όρος των επιμέρους σφαλμάτων των επαναλήψεων.

Σε ότι αφορά τα μοντέλα που αξιολογήθηκαν αρχικός στόχος ήταν να αξιολογηθούν τα περισσότερα από αυτά που αναφέρονται στη βιβλιογραφία για την πρόβλεψη τιμής του Bitcoin. Ωστόσο λόγω αδυναμίας εύρεσης υπολογιστικών πόρων, η εκτέλεση κάποιων εξ αυτών, όπως το RNN, δεν ήταν εφικτή. Προκειμένου να υπάρχει σχετικά μεγάλο εύρος αποτελεσμάτων διάφορων μοντέλων, αποφασίστηκε να χρησιμοποιηθούν και μοντέλα που δεν εμφανίζονται ιδιαίτερα στη βιβλιογραφία για τη μελέτη της τιμής του Bitcoin. Οι τελικές επιλογές που χρησιμοποιήθηκαν και συγκρίθηκαν τα αποτελέσματά τους ήταν τα ακόλουθα μοντέλα:

1. Naive
2. Holt
3. Random Forest
4. Gradient Boosting Machine (GBM)
5. ARIMA
6. Prophet
7. XGBoost
8. Συνδυαστική Μέθοδος (Comb)

Κατά την έναρξη της διαδικασίας αξιολόγησης των μοντέλων έγινε διαχωρισμός των δεδομένων σε τρία μέρη. Οι συνολικές παρατηρήσεις, όπως έχει ήδη αναφερθεί, είναι 1,608 στον αριθμό. Το 80% αυτών χρησιμοποιήθηκε για την προσαρμογή των μοντέλων (forecast model fitting), το επόμενο 10% αφορά τις παρατηρήσεις στις οποίες έγινε ο υπολογισμός του σφάλματος (in-sample error) κατά την αξιολόγηση των μοντέλων και τέλος το υπόλοιπο 10% για την αξιολόγηση της πρόβλεψης (out-sample error). Συγκεκριμένα έχουμε $1,608 * 0.8 = 1,286$ (ακέραιο αποτέλεσμα) για το train $1,608 * 0.1 = 161$ παρατηρήσεις για το validation και **161** για το αξιολόγηση της τελικής πρόβλεψης.

Οι ανεξάρτητες μεταβλητές που αρχικά επιλέχθηκαν να μελετηθούν, όπως έχουμε αναφέρει, ήταν οι: Bitcoin, Dash, Ethereum, Litecoin, Monero, Copper, Gold, Oil, Palladium, Platinum, Silver, Downjones, Snp, Sse, Bitcoin.Hashrate και Wikipedia.Views. Από τις μεταβλητές αυτές μετά την ανάλυση του πίνακα αλληλοσυσχέτισης (correlation matrix) επιλέχθηκαν να χρησιμοποιηθούν στην εκπαίδευση των μοντέλων μόνον αυτές που εμφάνιζαν αλληλοσυσχέτιση μεγαλύτερη από **0.7**. Συνεπώς, χρησιμοποιήσαμε τις **Litecoin, Monero, Dash, Ethereum, Oil, Copper, Bitcoin.Hashrate**.

Στην επόμενη ενότητα παρατίθενται αναλυτικά τα αποτελέσματα της διαδικασίας αξιολόγησης των μοντέλων.

6.4.3 Πίνακες Επιδόσεων ανά Μοντέλο και Χρονικό Ορίζοντα

Holt

Κατά την αξιολόγηση του αποτελεσματικότερου μοντέλου Holt αναζητήσαμε να βρούμε τις καλύτερες υπερπαραμέτρους. Έτσι, για την υπερπαραμέτρο alpha δοκιμάστηκαν τιμές στο διάστημα **[0.5, 0.9]** με βήμα **0.1** ενώ για την υπερπαραμέτρο beta τιμές στο διάστημα **[0.1,0.4]** επίσης με βήμα **0.1**. Στον παρακάτω πίνακα παρουσιάζονται τα αποτελέσματα αυτής της αναζήτησης και για τους τρεις ορίζοντες. Στην περίπτωση του ορίζοντα τριών και επτά ημερών βλέπουμε ότι για μικρές τιμές του alpha (συντελεστής εξομάλυνσης) και μικρές του beta το σφάλμα είναι συγκριτικά μεγαλύτερο από συνδυασμούς με μεγάλες τιμές στο alpha και το beta μαζί. Στην περίπτωση των δεκατεσσάρων ημερών βλέπουμε ότι μια μέση τιμή λίγο πάνω από το μέσο του διαστήματος τιμών των δύο παραμέτρων δίνει μικρότερο σφάλμα.

3 DAYS			7 DAYS			14 DAYS		
alpha	beta	RMSE	alpha	beta	RMSE	alpha	beta	RMSE
0,5	0,1	65,42	0,5	0,1	109,85	0,5	0,1	121,31
0,5	0,2	65,30	0,5	0,2	90,48	0,5	0,2	47,81
0,5	0,3	65,90	0,5	0,3	73,53	0,5	0,3	33,44
0,5	0,4	69,48	0,5	0,4	49,57	0,5	0,4	33,29
0,6	0,1	62,34	0,6	0,1	100,89	0,6	0,1	114,35
0,6	0,2	61,62	0,6	0,2	78,91	0,6	0,2	49,33
0,6	0,3	60,36	0,6	0,3	55,43	0,6	0,3	25,17
0,6	0,4	60,62	0,6	0,4	32,87	0,6	0,4	28,54
0,7	0,1	59,36	0,7	0,1	94,28	0,7	0,1	110,22
0,7	0,2	57,85	0,7	0,2	71,15	0,7	0,2	50,29
0,7	0,3	55,47	0,7	0,3	48,36	0,7	0,3	29,22

0,7	0,4	53,48	0,7	0,4	28,72	0,7	0,4	29,23
0,8	0,1	56,34	0,8	0,1	89,49	0,8	0,1	107,36
0,8	0,2	54,03	0,8	0,2	66,24	0,8	0,2	50,50
0,8	0,3	50,75	0,8	0,3	45,16	0,8	0,3	30,94
0,8	0,4	47,29	0,8	0,4	28,70	0,8	0,4	28,76
0,9	0,1	53,36	0,9	0,1	86,04	0,9	0,1	105,34
0,9	0,2	50,36	0,9	0,2	63,12	0,9	0,2	50,51
0,9	0,3	46,33	0,9	0,3	43,65	0,9	0,3	31,61
0,9	0,4	42,06	0,9	0,4	29,51	0,9	0,4	28,11

Πίνακας 2: Βελτιστοποίηση υπερπαραμέτρων Holt

Random Forest

Στην περίπτωση του Random Forest δοκιμάστηκαν συνδυασμοί των υπερπαραμέτρων ntree και mtry για τιμές **500** και **1,000** για την πρώτη και **2, 3, 4** για τη δεύτερη. Όπως φαίνεται και στον ακόλουθο πίνακα ο συνδυασμός **{1,000, 4}** δίνει τα καλύτερα αποτελέσματα με το σφάλμα να αυξάνεται ελάχιστα όσο αυξάνεται ο ορίζοντας πρόβλεψης γεγονός που δεν παρατηρείται στα υπόλοιπα μοντέλα που μελετήθηκαν. Αυτό μας δείχνει ότι το Random Forest είναι ανεπηρέαστο από την τάση των παρατηρήσεων.

3 DAYS			7 DAYS			14 DAYS		
ntree	mtry	RMSE	ntree	mtry	RMSE	ntree	mtry	RMSE
500	4	3,55	500	4	3,75	500	4	3,85
500	3	3,73	500	3	4,05	500	3	4,19
500	2	3,99	500	2	4,40	500	2	4,66
1000	4	3,36	1000	4	3,70	1000	4	3,78
1000	3	3,71	1000	3	4,01	1000	3	4,22
1000	2	4,15	1000	2	4,43	1000	2	4,51

Πίνακας 3: Βελτιστοποίηση υπερπαραμέτρων Random Forest

Την επιλογή των υπερπαραμέτρων ακολούθησε η επιλογή των ανεξάρτητων μεταβλητών. Ενώ από τον πίνακα αλληλοσυσχέτισης είχαμε ήδη περιορίσει τον αριθμό των εξαρτημένων μεταβλητών σε 7 (Litecoin, Monero, Dash, Ethereum, Oil, Copper, Bitcoin.Hashrate) στο σημείο αυτό πραγματοποιήθηκε μια πιο ενδελεχής εξέταση. Έχοντας ως στόχο την εύρεση του μικρότερου δυνατού σφάλματος εκτελέσαμε το μοντέλο εννέα φορές (τρεις επαναλήψεις ανά χρονικό ορίζοντα) απομονώνοντας κάθε φορά μία ή περισσότερες μεταβλητές. Τα αποτελέσματα φαίνονται στον ακόλουθο πίνακα:

	3 DAYS			7 DAYS			14 DAYS		
	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη
Litecoin									
Monero									
Dash									
Ethereum									
Oil									
Copper									
Bitcoin. Hashrate									
RMSE	1,01	0,19	12,61	1,32	1,00	7,22	1,32	1,00	7,22
Συνολικός Χρόνος	6,85	6,57	6,09	6,78	6,50	5,42	6,56	6,54	5,22

Πίνακας 4: Αποτελέσματα αξιολόγησης ανεξάρτητων μεταβλητών για RF

Είναι φανερό πως οι τιμές των άλλων κρυπτονομισμάτων επηρεάζουν κατά πολύ την τιμή του Bitcoin αρκεί μόνο να παρατηρήσουμε πόσο μεγάλη είναι η αύξηση του RMSE στις περιπτώσεις όπου αφαιρέθηκαν οι μεταβλητές αυτές ιδιαίτερα στην περίπτωση του ορίζοντα των τριών ημερών.

Gradient Boosting Machine (GBM)

Στην εξέταση αυτού του μοντέλου εξετάστηκαν οι υπερπαραμέτροι (οι έννοιες των οποίων έχουν αναλυθεί σε προηγούμενο κεφάλαιο):

- n.trees με τιμές **500, 1000** και **2000**
- interaction.depth με τιμές **2, 3** και **4**
- shrinkage με τιμές **0.1, 0.01** και **0.001**

Στον παρακάτω πίνακα παρουσιάζονται, με αύξουσα κατάταξη ως προς το σφάλμα, τα αποτελέσματα της ανάλυσης μας για τους τρεις ορίζοντες που εξετάζονται. Παρατηρούμε ότι ο συνδυασμός των τιμών **{2000, 4, 0.1}**, αντίστοιχα, δίνει το καλύτερο αποτέλεσμα και στις τρεις περιπτώσεις.

3 DAYS				7 DAYS				14 DAYS			
n.trees	interaction .depth	shrinkage	RMSE	n.trees	interaction .depth	shrinkage	RMSE	n.trees	interaction .depth	shrinkage	RMSE
500	4	0,1	21,45	500	4	0,1	23,59	500	4	0,1	22,38

500	3	0,1	25,51	500	3	0,1	26,60	500	3	0,1	25,18
500	2	0,1	32,07	500	2	0,1	32,75	500	2	0,1	31,80
500	4	0,01	89,11	500	4	0,01	89,99	500	4	0,01	83,39
500	3	0,01	106,70	500	3	0,01	107,30	500	3	0,01	99,24
500	2	0,01	149,15	500	2	0,01	149,72	500	2	0,01	139,40
500	4	0,001	1988,49	500	4	0,001	1988,18	500	4	0,001	1973,82
500	3	0,001	1991,17	500	3	0,001	1991,13	500	3	0,001	1976,68
500	2	0,001	2155,18	500	2	0,001	2155,58	500	2	0,001	2140,20
1000	4	0,1	16,51	1000	4	0,1	17,45	1000	4	0,1	17,09
1000	3	0,1	19,12	1000	3	0,1	20,73	1000	3	0,1	20,00
1000	2	0,1	23,66	1000	2	0,1	23,80	1000	2	0,1	23,74
1000	4	0,01	44,34	1000	4	0,01	44,57	1000	4	0,01	42,98
1000	3	0,01	52,12	1000	3	0,01	52,91	1000	3	0,01	50,91
1000	2	0,01	65,33	1000	2	0,01	65,90	1000	2	0,01	62,35
1000	4	0,001	1307,71	1000	4	0,001	1307,14	1000	4	0,001	1294,54
1000	3	0,001	1324,07	1000	3	0,001	1323,64	1000	3	0,001	1310,77
1000	2	0,001	1496,91	1000	2	0,001	1496,27	1000	2	0,001	1483,12
2000	4	0,1	12,88	2000	4	0,1	13,94	2000	4	0,1	14,82
2000	3	0,1	14,87	2000	3	0,1	16,25	2000	3	0,1	16,51
2000	2	0,1	17,72	2000	2	0,1	18,22	2000	2	0,1	20,44
2000	4	0,01	30,94	2000	4	0,01	32,03	2000	4	0,01	31,14
2000	3	0,01	36,50	2000	3	0,01	37,82	2000	3	0,01	35,78
2000	2	0,01	45,95	2000	2	0,01	47,75	2000	2	0,01	44,73
2000	4	0,001	579,16	2000	4	0,001	578,56	2000	4	0,001	567,38
2000	3	0,001	627,67	2000	3	0,001	626,89	2000	3	0,001	615,90
2000	2	0,001	756,01	2000	2	0,001	755,72	2000	2	0,001	743,51

Πίνακας 5: Βελτιστοποίηση υπερπαραμέτρων Gradient Boosting Machine

Κατά όμοιο τρόπο με το παραπάνω μοντέλο (Random Forest) επιλέχθηκε να “τρέξουμε” το μοντέλο (3 επαναλήψεις ανά χρονικό ορίζοντα) απομονώνοντας στην πρώτη επανάληψη το Bitcoin.Hashrate στη δεύτερη τα εμπορεύματα και στη τρίτη επανάληψη τα άλλα κρυπτονομίσματα. Τα αποτελέσματα συνοψίζονται στον πίνακα που ακολουθεί:

	3 DAYS			7 DAYS			14 DAYS		
	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη
Litecoin									
Monero									
Dash									
Ethereum									
Oil									
Copper									
Bitcoin. Hashrate									
RMSE	18,84	9,02	8,67	16,44	20,75	27,11	16,44	20,75	27,11
Συνολικός Χρόνος	1,62	1,34	1,07	1,52	1,35	1,03	1,36	1,35	1,02

Πίνακας 6: Αποτελέσματα αξιολόγησης ανεξάρτητων μεταβλητών για GBM

ARIMA

Κατά την αναζήτηση του αποτελεσματικότερου μοντέλου ARIMA αναζητήσαμε να βρούμε τις καλύτερες υπερπαραμέτρους. Έτσι, για τις υπερπαραμέτρους p, d, q δοκιμάστηκαν οι τιμές 1, 2, 3. Στον παρακάτω πίνακα παρουσιάζονται τα αποτελέσματα αυτής της αναζήτησης και για τους τρεις ορίζοντες.

p	3 DAYS			p	7 DAYS			p	14 DAYS		
	d	q	RMSE		d	q	RMSE		d	q	RMSE
1	1	1	14,19	1	1	1	17,46	1	1	1	22,02
1	1	2	12,09	1	1	2	17,73	1	1	2	22,07
1	1	3	12,01	1	1	3	17,74	1	1	3	22,08
1	2	1	12,43	1	2	1	17,79	1	2	1	22,87
1	2	2	17,52	1	2	2	26,05	1	2	2	40,71
1	2	3	16,49	1	2	3	24,69	1	2	3	37,03
1	3	1	16,81	1	3	1	27,67	1	3	1	59,50
1	3	2	59,10	1	3	2	102,37	1	3	2	164,50
1	3	3	58,62	1	3	3	111,50	1	3	3	194,17
2	1	1	12,00	2	1	1	17,74	2	1	1	22,08
2	1	2	14,25	2	1	2	17,46	2	1	2	22,02
2	1	3	14,25	2	1	3	17,52	2	1	3	21,79
2	2	1	15,10	2	2	1	23,27	2	2	1	37,49
2	2	2	16,68	2	2	2	23,84	2	2	2	36,26
2	2	3	16,28	2	2	3	23,01	2	2	3	33,77
2	3	1	46,42	2	3	1	21,35	2	3	1	34,95
2	3	2	55,78	2	3	2	104,10	2	3	2	171,52
2	3	3	54,65	2	3	3	108,99	2	3	3	224,89
3	1	1	13,47	3	1	1	17,85	3	1	1	22,31
3	1	2	14,81	3	1	2	18,80	3	1	2	24,12

3	1	3	30,07	3	1	3	34,66	3	1	3	36,35
3	2	1	16,36	3	2	1	27,12	3	2	1	44,63
3	2	2	15,93	3	2	2	19,90	3	2	2	29,03
3	2	3	24,09	3	2	3	26,71	3	2	3	32,03
3	3	1	47,94	3	3	1	20,66	3	3	1	30,86
3	3	2	58,28	3	3	2	112,36	3	3	2	210,95
3	3	3	60,50	3	3	3	115,64	3	3	3	216,00

Πίνακας 7: Βελτιστοποίηση υπερπαραμέτρων ARIMA

XGBoost

Στην περίπτωση του XGBoost εξετάστηκαν οι υπερπαραμέτροι eta και max_depth για τιμές **0.05**, **0.1**, **0.2** για την πρώτη και **2**, **4**, **6** και **10** για την δεύτερη. Το καλύτερο αποτέλεσμα και για τους τρεις ορίζοντες δίνεται από τον συνδυασμό **{0.2, 10}**.

3 DAYS			7 DAYS			14 DAYS		
eta	max_depth	RMSE	eta	max_depth	RMSE	eta	max_depth	RMSE
0,05	2	143,58	0,05	2	143,68	0,05	2	133,92
0,05	4	27,96	0,05	4	29,42	0,05	4	31,17
0,05	6	26,96	0,05	6	27,63	0,05	6	29,34
0,05	10	26,67	0,05	10	26,93	0,05	10	28,14
0,1	2	75,00	0,1	2	75,17	0,1	2	69,96
0,1	4	18,97	0,1	4	19,94	0,1	4	19,53
0,1	6	9,03	0,1	6	10,23	0,1	6	10,65
0,1	10	3,11	0,1	10	3,40	0,1	10	3,69
0,2	2	37,36	0,2	2	37,80	0,2	2	36,45
0,2	4	17,06	0,2	4	17,88	0,2	4	17,83
0,2	6	9,29	0,2	6	10,79	0,2	6	11,30
0,2	10	2,55	0,2	10	2,57	0,2	10	2,58

Πίνακας 8: Βελτιστοποίηση υπερπαραμέτρων XGBoost

Επιπλέον της αναζήτησης του καλύτερου συνδυασμού υπερπαραμέτρων προχωρήσαμε σε έλεγχο εκτενέστερης βελτιστοποίησης πειραματιζόμενοι με τις ανεξάρτητες μεταβλητές. Τα αποτελέσματα των δοκιμών που πραγματοποιήθηκαν φαίνονται παρακάτω:

	3 DAYS			7 DAYS			14 DAYS		
	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη	1η Επαν/ψη	2η Επαν/ψη	3η Επαν/ψη
Litecoin									
Monero									
Dash									
Ethereum									
Oil									
Copper									
Bitcoin. Hashrate									
RMSE	3,11	3,15	5,65	3,24	3,35	5,47	3,34	3,45	5,70
Συνολικός Χρόνος	12,32	11,45	9,65	5,48	4,96	4,11	3,01	2,71	2,43

Πίνακας 9: Αποτελέσματα αξιολόγησης ανεξάρτητων μεταβλητών για XGBoost

Παρατηρούμε ότι, σε όλες τις περιπτώσεις, το σφάλμα μεγαλώνει όσο μεγαλώνει ο χρονικός ορίζοντας ενώ ο χρόνος εκτέλεσης μειώνεται. Στην περίπτωση της αφαίρεσης της μεταβλητής Bitcoin.Hashrate και στην περίπτωση των εμπορευμάτων παρατηρούμε μια αύξηση του σφάλματος ωστόσο το σφάλμα αυτό είναι συγκριτικά μεγαλύτερο όταν αφαιρούμε τα άλλα κρυπτονομίσματα. Η αφαίρεση μίας μόνο μεταβλητής δίνει το μικρότερο δυνατό σφάλμα αν και μεγαλύτερο από το αρχικό όπου στην περίπτωση του ορίζοντα τριών ημερών για παράδειγμα ήταν RMSE ίσο με **2.54956**. Σε κάθε περίπτωση όμως είναι μεγαλύτερο από το σφάλμα του αρχικού μοντέλου χωρίς της αφαίρεση των μεταβλητών.

6.4.4 Επιλογή Μοντέλου

Μετά την αξιολόγηση όλων των μοντέλων και την επιλογή των υπερπαραμέτρων και των ανεξάρτητων μεταβλητών ακολουθεί η παρουσίαση του τελικού πίνακα αποτελεσμάτων των μοντέλων, με ανάλυση ανά ορίζοντα. Παρατηρούμε ότι οι μέθοδοι με το μικρότερο σφάλμα πρόβλεψης είναι κατά αύξουσα σειρά η XGBoost, Random Forest, Combination και ARIMA στην περίπτωση του ορίζοντα τριών ημερών ή GBM στην περίπτωση των άλλων δύο.

Στο σημείο αυτό σημειώνεται ότι η μέθοδος Combination προκύπτει ως ο μέσος όρος της πρόβλεψης των τριών μοντέλων που επιλέχθηκαν. Τα αποτελέσματα της αξιολόγησης όλων των μοντέλων φαίνονται στον παρακάτω πίνακα.

Method	3 DAYS	7 DAYS	14 DAYS
Naive	12,01	17,68	21,87
HOLT	42,06	28,70	25,17
Random Forest	3,36	3,70	3,78
GBM	12,88	13,94	14,82
ARIMA	12,00	17,46	21,79
Prophet	40,92	40,34	42,02
XGBoost	2,55	2,57	2,58
Comb	5,68	6,24	6,11

Πίνακας 10: Αποτελέσματα αξιολόγησης μοντέλων ανά ορίζοντα

6.4.5 Αποτελέσματα Πρόβλεψης

Για τις ανάγκες της τελικής πρόβλεψης αποφασίστηκε να εξεταστούν τα μοντέλα Random Forest, XGBoost, Combination καθώς και το ARIMA, το οποίο παρόλο που δεν εμφανίστηκε στα τρία καλύτερα της αξιολόγησης είναι αυτό που είχε τελικά το μικρότερο αποτέλεσμα κατά την πρόβλεψη.

	3 Ημέρες	7 Ημέρες	14 Ημέρες
Random Forest	266,81	209,65	1420,74
XGBoost	323,19	410,25	1239,88
Comb	275,05	285,95	1230,54
ARIMA	208,49	239,97	1027,00

Πίνακας 11: Αποτελέσματα Αξιολόγησης Πρόβλεψης Τιμής Bitcoin

Παρατηρούμε ότι το σφάλμα πρόβλεψης της τιμής του Bitcoin στον ορίζοντα των δεκατεσσάρων ημερών εκτοξεύεται. Είναι περίπου πέντε φορές μεγαλύτερο. Η πρόβλεψη αφορά το διάστημα **25/07/2019** έως **07/08/2019**. Τις πρώτες επτά ημέρες αυτού του διαστήματος η τιμή του Bitcoin ήταν κατά μέσο όρο **10,000\$** ενώ στις επόμενες επτά αυξήθηκε κατά περίπου **2,000\$**. Είναι γνωστό (αρκεί να παρατηρήσουμε το γράφημα) ότι η χρονοσειρά της τιμής του Bitcoin παρουσιάζει πολύ θόρυβο. Ο θόρυβος αυτός είναι που καθιστά δύσκολη την πρόβλεψη για μεγάλο ορίζοντα.

7 Επίλογος

Ολοκληρώνοντας την παρουσίαση της μελέτης αξίζει να γίνει αναφορά στη σημασία των αποτελεσμάτων που επιτεύχθηκαν ως προς την πρόβλεψη της τιμής του Bitcoin. Η ενότητα αυτή περιλαμβάνει μία σύνοψη των όσων υλοποιήθηκαν και αναδεικνύει τη χρησιμότητά τους σε ένα περιβάλλον γεμάτο προκλήσεις όπως αυτό των επενδύσεων. Τέλος, περιγράφονται ιδέες για πιθανές επεκτάσεις της μελέτης που αποσκοπούν στην όσο το δυνατόν μεγαλύτερη κατανόηση του ενδιαφέροντος αυτού κρυπτονομίσματος.

7.1 Σύνοψη και συμπεράσματα

Μέσω της εργασίας αυτής έγινε μία ολοκληρωμένη μελέτη γύρω από την ιστορικότητα της τιμής κλεισίματος του Bitcoin καθώς επίσης και ως προς τα χαρακτηριστικά που το επηρεάζουν. Στη συνέχεια, έχοντας ληφθεί υπόψη η βιβλιογραφία, έγινε προσπάθεια μελέτης και πρόβλεψης της τιμής του. Ζητούμενο δεν ήταν μόνο η εξεύρεση του πιο αποδοτικού μοντέλου πρόβλεψης για τους τρεις ορίζοντες που εξετάστηκαν αλλά και η τελική πρόβλεψη της τιμής του. Εξαιτίας της πολύ σύντομης ιστορίας του Bitcoin και των κρυπτονομισμάτων γενικότερα, τα διαθέσιμα δεδομένα διαρκώς αυξάνονται. Όπως είναι λογικό, έρευνες που έχουν στηριχθεί πάνω σε παρελθόντα δεδομένα είναι αρκετά πιθανό να θεωρούνται πλέον παρωχημένες. Η μελέτη αυτή, κάνοντας χρήση τόσο των δεδομένων των προηγούμενων ερευνών όσο και των πιο πρόσφατων διαθέσιμων, προχώρησε στην πρόβλεψη της τιμής του Bitcoin και τα κατάφερε σε ικανοποιητικό βαθμό.

Ανακεφαλαιώνοντας, μπορούμε να πούμε ότι τα συμπεράσματα που προκύπτουν από την παρούσα μελέτη είναι ποικίλα, τόσο σε σχέση με τα μοντέλα όσο και σε σχέση με τους ορίζοντες πρόβλεψης.

- Αρχικά, κατά το στάδιο της αξιολόγησης των μοντέλων προκύπτει ότι το μοντέλο με το μικρότερο RMSE είναι το XGBoost. Το επόμενο στη σειρά καλύτερο μοντέλο είναι το Random Forest με σφάλμα περίπου 30% μεγαλύτερο από το πρώτο. Ακολουθεί το GBM και το ARIMA το οποίο στην περίπτωση του ορίζοντα των τριών ημερών είχε καλύτερο αποτέλεσμα. Σε ότι αφορά τις μεθόδους Holt και Prophet παρατηρούμε σφάλμα κατά πολύ μεγαλύτερο από αυτό που παρουσιάζει η απλοϊκή μέθοδος Naive γεγονός που επιβεβαιώνει την εξάρτησή τους από την τάση, η οποία είναι έντονη ανά διαστήματα (ανοδική ή πτωτική). Συμπερασματικά,

κατά την αξιολόγηση τα μοντέλα μηχανικής μάθησης είχαν με διαφορά καλύτερο αποτέλεσμα.

- Ένας από τους στόχους της εργασίας ήταν να εντοπίσει το καταλληλότερο μοντέλο για κάθε έναν από τους τρεις ορίζοντες μελέτης. Το αποτέλεσμα της μελέτης ανέτρεψε τις αρχικές προσδοκίες καθώς καταλήξαμε στο συμπέρασμα ότι σε βραχυχρόνιους ορίζοντες τα μοντέλα έχουν το ίδιο αποτέλεσμα.
- Επιβεβαιώθηκε η βιβλιογραφία ως προς την υψηλή συσχέτιση της τιμής του Bitcoin με αυτές των άλλων κρυπτονομισμάτων. Αντίθετα, οι τιμές των βασικών εμπορευμάτων και ο ρυθμός παραγωγής Hashes του συνολικού δικτύου του Bitcoin δεν επηρέασε σημαντικά το αποτέλεσμα του μοντέλου. Στην περίπτωση του XGBoost η αφαίρεση των μεταβλητών των κρυπτονομισμάτων οδήγησε σε μια αύξηση του σφάλματος του μοντέλου για τον ορίζοντα 3 ημερών από 2,55 σε 5,65, δηλαδή περίπου 120%. Το σφάλμα του Random Forest για την αφαίρεση των ίδιων μεταβλητών και ορίζοντα 7 ημερών αυξήθηκε από 3,7 σε 7,22, ήτοι 95%, ενώ το GBM στον ορίζοντα των 14 ημερών αυξήθηκε από 14,82 σε 27,11, ήτοι 82%. Από το παράδειγμα γίνεται φανερό ότι η αξιολόγηση των μεταβλητών εισόδου και με τα τρία μοντέλα μηχανικής μάθησης επιβεβαίωσε την υψηλή συσχέτιση που έχει η τιμή του Bitcoin με τις τιμές των άλλων κρυπτονομισμάτων.
- Προχωρώντας, στο στάδιο της πρόβλεψης της τιμής του Bitcoin και της αξιολόγησής της με τη χρήση του RMSE προκύπτει ότι το μοντέλο που έδωσε τελικά την καλύτερη πρόβλεψη ήταν το ARIMA και αυτό γιατί σαν αυτοπαλινδρομικό μοντέλο ερμηνεύει καλύτερα την εξέλιξη μιας τιμής στον χρόνο σε αντίθεση με τα μοντέλα μηχανικής μάθησης τα οποία σε λίγα δεδομένα μπορεί να οδηγηθούν σε όχι τόσο ακριβή αποτελέσματα. Το σφάλμα της πρόβλεψης ήταν περίπου 18 φορές μεγαλύτερο από αυτό της αξιολόγησης στην περίπτωση των 3 ημερών και 14 φορές μεγαλύτερο στην περίπτωση των 7 ημερών. Ο ορίζοντας των 14 ημερών είχε εξαιρετικά μεγάλο σφάλμα τόσο σε σχέση με τις πραγματικές τιμές όσο και σε σχέση με αυτό της αξιολόγησης γεγονός που μας οδηγεί στο συμπέρασμα ότι η τιμή του Bitcoin είναι τόσο ευμετάβλητη και έχει τόσο μεγάλες διακυμάνσεις που είναι σχεδόν αδύνατο να προβλέψεις για διάστημα μεγαλύτερο από 7 ημέρες.

- Για ημερήσια δεδομένα, όπως αυτά που μελετήθηκαν στην παρούσα εργασία, με τόσο μεγάλο θόρυβο και μικρό όγκο, όσο πιο λίγες περιόδους έχει ο ορίζοντας πρόβλεψης τόσο πιο ασφαλής είναι η πρόβλεψη.
- Εν κατακλείδι, διαπιστώθηκε ότι τα μοντέλα μηχανικής μάθησης δεν έφεραν σημαντικά καλύτερα αποτελέσματα κατά την πρόβλεψη. Η χρήση ενός στατιστικού μοντέλου όπως το ARIMA είναι ίσως η καλύτερη επιλογή καθώς είναι πιο απλό και απαιτεί λιγότερους πόρους και πολύ λιγότερο χρόνο στην παραμετροποίηση και στην τελική εφαρμογή του.

7.2 Μελλοντικές επεκτάσεις - βελτιώσεις

Η παρούσα μελέτη εστίασε στον εντοπισμό του καταλληλότερου μοντέλου πρόβλεψης για τους ορίζοντες τριών, επτά και δεκατεσσάρων ημερών, θέλοντας με αυτό τον τρόπο ένα χρήσιμο εργαλείο που απευθύνεται σε όποιον ενδιαφέρεται για το Bitcoin ως επενδυτικό αγαθό.

Αν και αποτελεί αυτοτελές έργο, στο μέλλον θα μπορούσε να επεκταθεί ώστε να παρέχει μία ακόμα πιο ολοκληρωμένη εικόνα για το Bitcoin και τη διακύμανση της τιμής του. Οι υποενότητες που ακολουθούν περιγράφουν το πως.

7.2.1 Χρήση περισσότερων και πιο απαιτητικών μοντέλων

Ανάμεσα στις μελλοντικές επεκτάσεις πρώτη και κύρια είναι η εκτέλεση μοντέλων νευρωνικών δικτύων, κάτι το οποίο δεν έγινε εφικτό με το διαθέσιμο υλισμικό. Βάσει της βιβλιογραφίας, τα αποτελέσματα πρόβλεψης της τιμής του Bitcoin αναμένονται να είναι ακόμα καλύτερα κάνοντας χρήση βαθιών νευρωνικών δικτύων όπως τα RNN, LSTM.

Φυσικά, εκτός των παραπάνω, υπάρχουν και άλλα μοντέλα, μερικά εκ των οποίων έχουν αναπτυχθεί αρκετά πρόσφατα και τα οποία απουσιάζουν από τη βιβλιογραφία. Ενδεικτική είναι η περίπτωση του μοντέλου Catboost, το οποίο ισχυρίζεται ότι μπορεί να επιτύχει υπό προϋποθέσεις καλύτερα αποτελέσματα από το XGBoost. Μία μελλοντική επέκταση της παρούσας μελέτης θα ήταν η πιθανή χρήση του μοντέλου και διερεύνηση των αποτελεσμάτων του.

7.2.2 Διερεύνηση διαφορετικού ορίζοντα πρόβλεψης

Για τις ανάγκες της παρούσας διπλωματικής επιλέχθηκε ορίζοντας πρόβλεψης ίσος με τρεις, επτά και δεκατέσσερις ημερήσιες παρατηρήσεις. Μία ενδιαφέρουσα επέκταση της παρούσας μελέτης θα ήταν η εξέταση και άλλου (ή άλλων) ορίζοντα πρόβλεψης.

Ενδεικτικά, ένας αρκετά σύντομος ορίζοντας πρόβλεψης ενός μήνα ή αντίστοιχα ένας μεγάλος ορίζοντας τριών ή έξι μηνών, είναι πολύ πιθανόν να δώσουν διαφορετικά αποτελέσματα, τόσο ως προς την ορθότητα των προβλέψεων όσο και ως προς την επιμέρους επίδοση των μοντέλων.

Βιβλιογραφία

Σπύρου, Σ. (2003). Αγορές χρήματος και κεφαλαίου. Ευγ. Μπένου.
Ευρωπαϊκή Κεντρική Τράπεζα, (2017). Τι είναι το χρήμα;

Zentralbank, E. (2015). Virtual currency schemes-A further analysis. Frankfurt am Main, Februar.

Σπύρου, Σ. (2009). Εισαγωγή στη Συμπεριφορική Χρηματοοικονομική. Κεφαλαιαγορές και επενδυτική ψυχολογία. Εκδ. Γ. Μπένου.

Αλεξανδρίδης, Μ., & Επενδύσεων, Δ. (2005). Εκδόσεις Σύγχρονη Εκδοτική ΕΠΕ.

Liu, Y., & Tsyvinski, A. (2018). Risks and returns of cryptocurrency (No. w24877). National Bureau of Economic Research.

Brandvold, M., Molnár, P., Vagstad, K., & Valstad, O. C. A. (2015). Price discovery on Bitcoin exchanges. *Journal of International Financial Markets, Institutions and Money*, 36, 18-35. -
Lea, T. (2016). *Down the Rabbit Hole: (Discover the Power of the Blockchain) (Vol. 1)*. Tim Lea. –

Mora, C., Rollins, R. L., Taladay, K., Kantar, M. B., Chock, M. K., Shimada, M., & Franklin, E. C. (2018). Bitcoin emissions alone could push global warming above 2 C. *Nature Climate Change*, 8(11), 931-933.

Yermack, D. (2015). Is Bitcoin a real currency? An economic appraisal. In *Handbook of digital currency* (pp. 31-43). Academic Press.

Ciaian, P., Rajcaniova, M., & Kancs, A. The economics of bitcoin price formation (2014). ArXiv preprint arXiv: 1405.4498.

Kristoufek, L. (2013). BitCoin meets Google Trends and Wikipedia: Quantifying the relationship between phenomena of the Internet era. *Scientific reports*, 3, 3415.

Bouoiyour, J., Selmi, R., & Tiwari, A. (2014). Is Bitcoin business income or speculative bubble? Unconditional vs. conditional frequency domain analysis.

Bukovina, J., & Martiček, M. (2016). Sentiment and bitcoin volatility. *MENDELU Working Papers in Business and Economics*, 58.

Klein, T., Thu, H. P., & Walther, T. (2018). Bitcoin is not the New Gold—A comparison of volatility, correlation, and portfolio performance. *International Review of Financial Analysis*, 59, 105-116.

Ji, Q., Bouri, E., Roubaud, D., & Kristoufek, L. (2019). Information interdependence among energy, cryptocurrency and major commodity markets. *Energy Economics*, 81, 1042-1055.

Panagiotidis, T., Stengos, T., & Vravosinos, O. (2018). On the determinants of bitcoin returns: A LASSO approach. *Finance Research Letters*, 27, 235-240.

Ciaian, P., Rajcaniova, M., & Kancs, A. The economics of bitcoin price formation (2014). arXiv preprint arXiv:1405.4498.

Sovbetov, Y. (2018). Factors influencing cryptocurrency prices: Evidence from bitcoin, ethereum, dash, litecoin, and monero. *Journal of Economics and Financial Analysis*, 2(2), 1-27.

Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). Time series analysis: forecasting and control. John Wiley & Sons.

Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.

Kursa, B. M., and Rudnicki, W. R. (2011), "The All Relevant Feature Selection using Random Forest", ResearchGate, 1-16

Makridakis, S., Wheelwright, S.C., Hyndman, R.J, (1998) "Forecasting: methods and applications", New York, John Wiley & Sons, Inc

Nelder, J. A., & Wedderburn, R. W. (1972). Generalized linear models. Journal of the Royal Statistical Society: Series A (General), 135(3), 370-384.

Hochreiter, S., & Schmidhuber, J. (1997). LSTM can solve hard long time lag problems. In Advances in neural information processing systems (pp. 473-479).

World Health Organization. (2020). WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020.

Bakar, N. A., & Rosbi, S. (2017). Autoregressive integrated moving average (ARIMA) model for forecasting cryptocurrency exchange rate in high volatility environment: A new insight of bitcoin transaction. International Journal of Advanced Engineering Research and Science, 4(11), 237311.

Sun, X., Liu, M., & Sima, Z. (2020). A novel cryptocurrency price trend forecasting model based on LightGBM. Finance Research Letters, 32, 101084.

Linardatos P., Kotsiantis S. (2020) Bitcoin Price Prediction Combining Data and Text Mining. In: Hatzilygeroudis I., Perikos I., Grivokostopoulou F. (eds) Advances in Integrations of Intelligent Methods. Smart Innovation, Systems and Technologies, vol 170. Springer, Singapore

S. McNally, J. Roche and S. Caton, "Predicting the Price of Bitcoin Using Machine Learning," 2018 26th Euromicro International Conference on Parallel, Distributed and Network-based Processing (PDP), Cambridge, 2018, pp. 339-343

Karakoyun, Ebru Şeyma, and Ali Osman Çıbıkdiken. "Comparison of ARIMA Time Series Model and LSTM Deep Learning Algorithm for Bitcoin Price Forecasting." ECONOMICS, MANAGEMENT AND MARKETING (MAC-EMM 2018): 171.

I. Yenidoğan, A. Çayır, O. Kozan, T. Dağ and Ç. Arslan, "Bitcoin Forecasting Using ARIMA and PROPHET," 2018 3rd International Conference on Computer Science and Engineering (UBMK), Sarajevo, 2018, pp. 621-624

Madan, Isaac, Shaurya Saluja, and Aojia Zhao. "Automated bitcoin trading via machine learning algorithms." URL: <http://cs229.stanford.edu/proj2014/Isaac%20Madan> 20 (2015).
Velankar, Siddhi, Sakshi Valecha, and Shreya Maji. "Bitcoin price prediction using machine learning." 2018 20th International Conference on Advanced Communication Technology (ICTACT). IEEE, 2018.

Ηλεκτρονικές Πηγές

Bitcoin. CCC - CoinMarketCap, 2015-2020. Yahoo! Finance. Retrieved, from <https://finance.yahoo.com/quote/BTC-USD>

Ethereum. CCC - CCC - CoinMarketCap.2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/ETH-USD>

Monero. CCC - CoinMarketCap.2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/XMR-USD>

Litecoin. CCC - CoinMarketCap.2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/LTC-USD>

Dash. CCC - CoinMarketCap.2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/DASH-USD>

Dow Jones Industrial Average (^DJI)

DJI - DJI Real Time Price. Currency in USD. JI - DJI Real Time Price. 2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/%5EDJI>

S&P 500 (^GSPC). SNP - SNP Real Time Price. 2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/%5EGSPC>

Crude Oil Jul 20 (CL=F). NY Mercantile - NY Mercantile Delayed Price. 2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/CL%3DF>

Gold Aug 20 (GC=F). COMEX - COMEX Delayed Price. 2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/GC%3DF>

Silver Jul 20 (SI=F). COMEX - COMEX Delayed Price. 2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/SI%3DF>

Copper Jul 20 (HG=F). COMEX - COMEX Delayed Price. 2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/HG=F>

Palladium Sep 20 (PA=F). NY Mercantile - NY Mercantile Delayed Price. 2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/PA%3DF>

Platinum Jul 20 (PL=F). NY Mercantile - NY Mercantile Delayed Price. 2015-2020. Yahoo! Finance. Retrieved from <https://finance.yahoo.com/quote/PL=F>

Shanghai Composite. 2015-2020. Investing.com. Retrieved from <https://www.investing.com/indices/shanghai-composite-historical-data>

Wikipedia Bitcoin Views. 20105-2020. Retrieved from https://wikimedia.org/api/rest_v1/metrics/pageviews/per-article/en.wikipedia/all-access/all-agents/Bitcoin/daily/2015010100/2020010100

Raghu Raj Rai, "XGBoost: The Excalibur for Everyone". 2018. Retrieved from <https://towardsdatascience.com/xgboost-the-excalibur-for-everyone-8009bd015f1e>

Taylor, S.J. and Letham, B., "Prophet: forecasting at scale". 2017. Retrieved from <https://research.fb.com/prophet-forecasting-at-scale>

Liu, S., "Forecasting with Prophet". 2018. Retrieved from <https://towardsdatascience.com/forecasting-with-prophet-d50bbfe95f91>

Nakamoto, S., & Bitcoin, A peer-to-peer electronic cash system. 2008. Retrieved from <https://bitcoin.org/bitcoin.pdf>

Συναλλαγή. Βικιπαίδεια, Η Ελεύθερη Εγκυκλοπαίδεια. 2016. Retrieved from <https://el.wikipedia.org/w/index.php?title=%CE%A3%CF%85%CE%BD%CE%B1%CE%BB%CE%BB%CE%B1%CE%B3%CE%AE>

CRAN. R. 2020. Retrieved from https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-is-R_003f

Rstudio. 2020. Retrieved from <https://blog.rstudio.com/2011/02/28/rstudio-new-open-source-ide-for-r>

Hashrate. 2020. Quandl. Retrieved from <https://www.quandl.com/data/BITCOINWATCH-Bitcoin-Watch>

Ethereum. 2020. Retrieved from <https://ethereum.org/en/what-is-ethereum/>

Litecoin. 2020. Retrieved from <https://litecoin.org/>

Monero. 2020. Retrieved from <https://web.getmonero.org/get-started/what-is-monero/>

Dash. 2020. Retrieved from <https://www.dash.org/faq/>

XGBoost. 2020. Retrieved from <https://xgboost.readthedocs.io/en/latest/parameter.html>