



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

Αξιολόγηση μεθόδων ενίσχυσης κλίσης για την εκτίμηση
κατανάλωσης ενέργειας σε κρυπτονομίσματα

Δημήτριος Σέρβος

«Εφαρμοσμένη Οικονομική και Ανάλυση Δεδομένων»
Τμήμα Οικονομικών Επιστημών

Σχολή Οικονομικών Επιστημών και Διοίκησης Επιχειρήσεων

Διατριβή που υπεβλήθη για την ικανοποίηση μέρους των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος Ειδίκευσης

09/2024

Πανεπιστήμιο Πατρών, Τμήμα Οικονομικών Επιστημών

Δημήτριος

© 2024 – Με την επιφύλαξη παντός δικαιώματος

Τριμελής Επιτροπή Επίβλεψης διπλωματικής εργασίας

Επιβλέπων: Εμμανουήλ Τζαγκαράκης Αναπληρωτής Καθηγητής

Μέλος Επιτροπής: Κωνσταντίνος Κουνετάς Αναπληρωτής Καθηγητής

Μέλος Επιτροπής: Γεώργιος Φίλης Επίκουρος Καθηγητής

Η παρούσα διατριβή με τίτλο

«Αξιολόγηση μεθόδων ενίσχυσης κλίσης για την εκτίμηση κατανάλωσης
ενέργειας σε κρυπτονομίσματα »

Εκπονήθηκε από τον **Δημήτριο Σέρβο, Α.Μ 1067563**, για τη μερική ικανοποίηση των απαιτήσεων για την απόκτηση Μεταπτυχιακού Διπλώματος Ειδίκευσης στην «Εφαρμοσμένη Οικονομική και Ανάλυση Δεδομένων» από το Πανεπιστήμιο Πατρών και εγκρίθηκε από τα μέλη της τριμελούς επιβλέπουσας επιτροπής.

Θα ήθελα να αφιερώσω τη διπλωματική μου εργασία στους γονείς μου, Θεόδωρο & Γεωργία, στην κοπέλα μου, Νάσια, και σε όλους όσους με βοήθησαν.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου, κύριο Εμμανουήλ Τζαγκαράκη, για την πολύτιμη βοήθεια του στην ολοκλήρωση της διπλωματικής μου εργασίας.

Επίσης, θα ήθελα να ευχαριστήσω τον Αναπληρωτή Καθηγητή, κύριο Κωνσταντίνο Κουνετά, και τον Επίκουρο Καθηγητή, κύριο Γεώργιο Φίλη, για τις εποικοδομητικές τους υποδείξεις στην ολοκλήρωση της εργασίας, ως μέλη της τριμελούς επιτροπής.

Επιπλέον, ευχαριστώ τον υποψήφιο διδάκτορα, Ανδρέα Ρετουινιώτη, για τη βοήθεια του καθ'όλη τη διάρκεια της ερευνητικής διαδικασίας.

Τέλος, θα ήθελα να εκφράσω την ευγνωμοσύνη μου στην οικογένεια μου και στην κοπέλα μου για την κατανόηση και τη στήριξη που μου προσέφεραν από την αρχή μέχρι το τέλος των σπουδών μου.

Περίληψη

Το Bitcoin ήρθε να φέρει επανάσταση στους τρόπους με τους οποίους μπορεί να πραγματοποιηθεί μια συναλλαγή, με τη συγκεκριμένη τεχνολογία να προσφέρει πλεονεκτήματα, όπως η ασφαλής πραγματοποίηση των συναλλαγών. Ωστόσο το Bitcoin καταναλώνει μεγάλα ποσά ενέργειας, γεγονός που έχει κινήσει το ενδιαφέρον της βιβλιογραφίας τα τελευταία χρόνια. Καθώς το μεγαλύτερο μέρος αυτής της ενέργειας που καταναλώνεται προέρχεται από άνθρακα και υδρογονάνθρακες, παράγοντας έτσι βλαβερές ουσίες. Η παρούσα διατριβή χρησιμοποίησε ένα μοντέλο μηχανικής μάθησης XGBoost για την πρόβλεψη της κατανάλωσης ενέργειας του Bitcoin, Χρησιμοποιώντας 11 ανεξάρτητες μεταβλητές, εκ των οποίων κάποιες σχετίζονται με τη λειτουργία και οι υπόλοιπες με τα οικονομικά χαρακτηριστικά του Bitcoin. Για όλες τις μεταβλητές συλλέχθηκαν ημερήσια δεδομένα από τις 10 Φεβρουαρίου 2017 έως τις 7 Ιουλίου 2020, με τη μονάδα μέτρησης της εξαρτημένης μεταβλητής να είναι τα TWh. Όσον αφορά τα αποτελέσματα της διατριβής, χρησιμοποιήθηκαν ως μέτρα αξιολόγησης το Nash-Sutcliffe Efficiency(NSE), το Mean absolute error(MAE) και το Root mean square error (RMSE). Τα μοντέλα πολλαπλής γραμμικής παλινδρόμησης και Gradient Boosting regression tree χρησιμοποιήθηκαν ως σημεία αναφοράς, ώστε να υπάρξει καλύτερη κατανόηση της καταλληλότητας του μοντέλου στο υπό εξέταση πλαίσιο. Έχοντας μια συνεχή μεταβλητή στόχο με την υποεξέταση περίοδο να παρουσιάζει μεγάλη μεταβλητότητα, το μοντέλο XGBoost κατάφερε να επιτύχει μια αποδεκτή προβλεπτική ικανότητα στο πλαίσιο ανάλυσης. Επίσης, παρατηρήθηκε ότι οι ακραίες τιμές επηρεάζουν αρνητικά την προβλεπτική ικανότητα του μοντέλου σε σύγκριση με την απόδοση του χωρίς ακραίες τιμές. Τέλος, παρατηρήθηκε ότι το μοντέλο είχε την καλύτερη απόδοση χρησιμοποιώντας το 70% του δείγματος ως σετ εκπαίδευσης.

λέξεις-κλειδιά: μηχανική μάθηση, XGBoost, Κατανάλωση ενέργειας.

Summary

Bitcoin has revolutionized transaction methods by offering benefits such as secure transactions. However, Bitcoin's significant energy consumption has attracted attention in recent literature. Since most of this energy comes from coal and hydrocarbons, it results in harmful emissions. This thesis used an XGBoost machine learning model to predict Bitcoin's energy consumption, utilizing 11 independent variables, some related to Bitcoin's operation and others to its economic characteristics. Daily data for these variables were collected from February 10, 2017, to July 7, 2020, with the dependent variable measured in TWh. Evaluation metrics included Nash-Sutcliffe Efficiency(NSE), Mean Absolute Error(MAE), and Root Mean Square Error(RMSE). The multiple linear regression model and Gradient Boosting regression tree were used as a reference point to provide a better understanding of the suitability of the model within the framework under examination. Despite the high variability of the target variable during the study period, the XGBoost model achieved acceptable predictive performance. It was also observed that outliers negatively affected the model's predictive accuracy compared to its performance without outliers. Finally, the model showed the best performance when using 70% of the sample for training.

Keywords: machine learning, XGBoost, energy consumption.

Περιεχόμενα

1	ΕΙΣΑΓΩΓΗ	1
2	ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ	6
2.1	Το Bitcoin και η κατανάλωση ενέργειας	6
2.2	Bitcoin και Ενεργειακή Κατανάλωση: Ποσοτική Ανάλυση . . .	11
2.3	Gradient Boosting	14
3	ΜΕΘΟΔΟΛΟΓΙΑ	16
3.1	Δεδομένα	16
3.2	Προεπεξεργασία των δεδομένων	26
3.3	Extreme Gradient Boosting (XGBoost)	29
3.3.1	Ανάλυση Μεθοδολογίας	31
3.3.2	Υπερπαράμετροι	33
4	ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	35
4.1	Σετ εκπαίδευσης-δοκιμής	35
4.2	Μετρικές αξιολόγησης	36
4.3	Αποτελέσματα	38
5	ΣΥΜΠΕΡΑΣΜΑΤΑ	43
6	ΒΙΒΛΙΟΓΡΑΦΙΑ	46

Κατάλογος Σχημάτων

1	Οι Χρονοσειρές των μεταβλητών energy_consumption, Average_Blocksize.	20
2	Οι Χρονοσειρές των μεταβλητών Bitcoin_Blockchain, Cost_Per_Transaction.	20
3	Οι Χρονοσειρές των μεταβλητών Bitcoin_Hashrate, Miner_Revenue.	21
4	Οι Χρονοσειρές των μεταβλητών Unique_Address, Transaction_Fee.	21
5	Οι Χρονοσειρές των μεταβλητών Transaction_Per_Block, Unique_Transaction.	22
6	Οι Χρονοσειρές των μεταβλητών Bitcoin_Market_Price, Market_Capitalization.	22
7	Παρατηρούμαι την σχέση μεταξύ του energy_consumption, Average_Blocksize.	23
8	Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Bitcoin_Blockchain, Cost_Per_Transaction.	23
9	Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Bitcoin_Hashrate, Miner_Revenue.	24
10	Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Unique_Address, Transaction_Fee.	24
11	Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Transaction_Per_Block, Unique_Transaction.	25
12	Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Bitcoin_Market_Price, Market_Capitalization.	25
13	Το θηκόγραμμα των μεταβλητών energy_consumption, Average_Blocksize, Bitcoin_Blockchain, Cost_Per_Transaction.	28
14	Το θηκόγραμμα των μεταβλητών Bitcoin_Hashrate, Miner_Revenue, Unique_Address, Transaction_Fee.	28
15	Το θηκόγραμμα των μεταβλητών Transaction_Per_Block, Unique_Transaction, Bitcoin_Market_Price, Market_Capitalization.	29

Κατάλογος Πινάκων

1	Περιγραφικά Στατιστικά.	19
2	Κάποια επιπλέον Στατιστικά Μέτρα.	19
3	Επίδοση σε Αναλογία Σετ Εκπαίδευσης/Δοκιμής (80%,20%). .	38
4	Επίδοση σε Αναλογία Σετ Εκπαίδευσης/Δοκιμής (70%,30%). .	39
5	Επίδοση σε Αναλογία Σετ Εκπαίδευσης/Δοκιμής (60%,40%). .	39

1 ΕΙΣΑΓΩΓΗ

Η συνεχής προόδος της τεχνολογίας έχει δώσει τη δυνατότητα σε πολλά μοντέλα μηχανικής μάθησης να μπορούν να πραγματοποιηθούν σε ένα αποδεκτό χρονικό διάστημα, με αποτέλεσμα να αυξάνεται η χρηστικότητα τους. Σε συνδυασμό με τις επιδόσεις που συνοδεύουν αυτή τη χρηστικότητα, έχουν αποτελέσει πόλο έλξης τα τελευταία χρόνια για ένα μεγάλο σύνολο επιστημονικών πεδίων, ένα από τα οποία είναι και τα οικονομικά. Μια από τις πιο δημοφιλείς και ταυτόχρονα επιτυχημένες τεχνικές μηχανικής μάθησης είναι η Gradient boosting. Το μοντέλο Gradient boosting προτάθηκε από τον Friedman (2001). Ο τρόπος λειτουργίας του μοντέλου παράγει μια διαδικασία η οποία είναι ερμηνεύσιμη και πολύ ανθεκτική, τόσο για παλινδρόμηση όσο και για ταξινόμηση. Επίσης, για να μπορέσει να πραγματοποιηθεί το boosting, το μοντέλο χρησιμοποιεί δέντρα για αδύναμους μαθητές. Η έννοια των αδύναμων μαθητών, ή αλλιώς weak learners, προήλθε από τον Schapire (1990). Ο οποίος απέδειξε ότι ένα μοντέλο μάθησης το οποίο ο μαθητής αποδίδει λίγο καλύτερα από μια τυχαία εικασία είναι ισοδύναμο με ένα μοντέλο μάθησης στο οποίο το σφάλμα του μαθητή είναι πολύ μικρό. Επομένως, από έναν αλγόριθμο αδύναμης μάθησης μπορείς να τον μετατρέψεις σε έναν αλγόριθμο που επιτυγχάνει υψηλή ακρίβεια. Έχοντας έτσι πρακτικές εφαρμογές, μετατρέποντας αποτελεσματικά έναν μέτριο αλγόριθμο μάθησης σε έναν αλγόριθμο που επιτυγχάνει εξαιρετικά καλή απόδοση.

Τα τελευταία χρόνια, η ερευνητική δραστηριότητα για ένα από τα πιο γνωστά και σημαντικά κρυπτονομίσματα, το Bitcoin, έχει αυξηθεί παράλληλα με την αύξηση και την προσβασιμότητα των δεδομένων. Αυτή η αύξηση ενδιαφέροντος οφείλεται σε λόγους όπως η κατανάλωση ενέργειας για την παραγωγή του Bitcoin και οι πιθανές συνέπειες που ενέχει, ενώ η αποκεντρωμένη φύση του αποτελεί επίσης έναν πόλο έλξης. Ο Nakamoto (2008) είναι το άτομο πίσω από τη δημιουργία του γνωστού σε όλους Bitcoin. Στηρίζεται σε ένα σύστημα ηλεκτρονικών πληρωμών που δεν χρειάζεται κάποιον διαμεσολαβητή, και άρα δεν βασίζεται στην εμπιστοσύνη. Επίσης, το σύστημα αυτό παρέχει πολύ καλή ασφάλεια λόγω του proof of work, το οποίο καθιστά υπολογιστικά ανέφικτη την ηλεκτρονική επίθεση. Το proof of work χρησιμοποιείται για την καταγραφή των συναλλαγών, για την οποία δίνεται μια ανταμοιβή βάσει της καλύτερης προσπάθειας.

Ο αλγόριθμος μηχανικής μάθησης Gradient boosting έχει δείξει εξαιρετικές επιδόσεις σε μια πληθώρα εφαρμογών. Παρακάτω ακολουθεί μια περιληπτική αναφορά σε κάποιες από αυτές τις εφαρμογές. Οι Basak et al. (2019) χρησιμοποιούν ένα μοντέλο XGBoost μαζί με κάποια άλλα μοντέλα για να προβλέψουν αν η τιμή της μετοχής θα αυξηθεί ή θα μειωθεί σε σχέση με την τιμή που είχε μερικές μέρες νωρίτερα. Στην ανάλυση τους χρησιμοποίησαν 10 διαφορετικές

μετοχές, εκ των οποίων οι 2 ήταν της Apple και της Facebook. Αναφέροντας την μέση ακρίβεια για διάφορες μετοχές, διαπίστωσαν ότι το μοντέλο XGBoost είχε καλύτερες επιδόσεις από άλλα μοντέλα ταξινόμησης που έχουν χρησιμοποιηθεί (Logistic Regression, SVM, ANN) με το XGBoost να πετυχαίνει ακρίβεια πάνω από 80%. Σύμφωνα με τους Derbentsev et al. (2020), χρησιμοποίησαν ένα μοντέλο μηχανικής μάθησης Gradient boosting για να προβλέψουν την τιμή τριών κρυπτονομισμάτων (Bitcoin (BTC), Ethereum (ETH) and Ripple (XRP)). Ως μέτρα αξιολόγησης του μοντέλου χρησιμοποίησαν τα MAPE, RMSE για να ποσοτικοποιήσουν την ακρίβεια του μοντέλου τους. Βρήκαν ότι το μοντέλο Gradient boosting είχε καλύτερη ακρίβεια από το μοντέλο μηχανικής μάθησης Random Forests, παρόλα αυτά, και οι δύο αλγόριθμοι κρίθηκαν ικανοί για το πλαίσιο ανάλυσης. Για ακόμα μια φορά, ο αλγόριθμος gradient boosting decision trees χρησιμοποιείται για πρόβλημα ταξινόμησης. Πιο συγκεκριμένα, οι Gavaille και Zasova (2023) χρησιμοποιούν τον συγκεκριμένο αλγόριθμο μηχανικής μάθησης για να προβλέψουν εάν μια επιχείρηση είναι φορολογικά συνεπής ή φοροδιαφεύγει, επιτυγχάνοντας μια πολύ υψηλή ακρίβεια 86.8%, υποδεικνύοντας πολύ καλή απόδοση του μοντέλου. Ένα ακόμα άρθρο των Guven και Kayalica (2023) χρησιμοποίησε ένα Gradient Boosted Regression Tree για να προβλέψει την ηλεκτρική κατανάλωση νοικοκυριών στην Τουρκία. Χρησιμοποιώντας μέτρα αξιολόγησης όπως MAE, MAPE, RMSE κατέληξαν στο συμπέρασμα ότι το μοντέλο Gradient Boosted Regression Tree έχει το μικρότερο σφάλμα από τα άλλα δύο μοντέλα που χρησιμοποιήθηκαν (Decision Tree, Random Forest). Τέλος, οι Hu et al. (2023) χρησιμοποίησαν κάποια μοντέλα, ένα από τα οποία ήταν το gradient boosting decision tree, για να προβλέψουν την απόδοση των μετοχών A της Κίνας. Για να μπορέσουν να αξιολογήσουν τις εκτός δείγματος αποδόσεις των μοντέλων, χρησιμοποίησαν το μέτρο αξιολόγησης R^2 . Κατέληξαν στο συμπέρασμα ότι το μοντέλο μηχανικής μάθησης gradient boosting decision tree είναι αποτελεσματικό στην πρόβλεψη της απόδοσης των μετοχών A της Κίνας. Ο αλγόριθμος μηχανικής μάθησης XGBoost προτάθηκε από τους Chen και Guestrin (2016). Πρόκειται για έναν πρόσφατο αλγόριθμο μηχανικής μάθησης, ο οποίος βασίζεται στο πλαίσιο gradient tree boosting με κανονικοποιημένη συνάρτηση κόστους και διάφορα χαρακτηριστικά που του επιτρέπουν να χειρίζεται δισεκατομμύρια παρατηρήσεις με πιο αποτελεσματικό τρόπο. Αυτό του επιτρέπει να εκτελείται σημαντικά πιο γρήγορα από άλλα συστήματα, ενώ προσφέρει παράλληλα επιλογές για την αντιμετώπιση της υπερπροσαρμογής. Έχει κερδίσει έδαφος στη βιβλιογραφία, καθώς παράλληλα με την αποτελεσματικότητά του, έχει δείξει εξαιρετικές επιδόσεις σε μια πληθώρα εφαρμογών. Παρακάτω παρουσιάζεται μια συνοπτική αναφορά σχετικά με τις εφαρμογές του συγκεκριμένου αλγορίθμου μηχανικής μάθησης.

Οι Dey et al. (2016) χρησιμοποίησαν ένα μοντέλο μηχανικής μάθησης, το eXtreme Gradient Boosting(XGBoost), για ένα πρόβλημα ταξινόμησης. Πιο αναλυτικά, το μοντέλο χρησιμοποιήθηκε για την πρόβλεψη της πτώσης ή της ανόδου της αξίας των μετοχών της Apple Inc και της Yahoo! Inc για περιόδους 60 και 90 ημερών. Το μοντέλο κατάφερε να επιτύχει ακρίβεια μεταξύ 87% με 99%, χρησιμοποιώντας μέτρα αξιολόγησης όπως accuracy, precision. Χρησιμοποιώντας ένα μοντέλο XGBoost, οι Touzani et al. (2018) προέβλεψαν την κατανάλωση ενέργειας για εμπορικά κτήρια και βρήκαν ότι όχι μόνο το μοντέλο απέδωσε πολύ καλά αποτελέσματα σε αυτό το πλαίσιο, αλλά βελτίωσε και την ακρίβεια πρόβλεψης σε σχέση με μοντέλα όπως το TOWT και το RF. Για την αξιολόγηση της ακρίβειας των μοντέλων, χρησιμοποιήθηκαν μέτρα όπως το R-squared και το RMSE. Η δημιουργία μοντέλων για την πρόβλεψη αν μια τράπεζα θα πτωχεύσει είναι υψίστης σημασίας. Αυτό έκαναν οι Carmona et al. (2019), οι οποίοι χρησιμοποίησαν ένα μοντέλο extreme gradient boosting για την πρόβλεψη πτώχευσης τραπεζών στον τραπεζικό τομέα των ΗΠΑ. Πρόκειται για ένα πρόβλημα ταξινόμησης στο οποίο το μοντέλο αποδείχθηκε ικανοποιητικό, επιτυγχάνοντας ακρίβεια 94.74% και ανώτερη προβλεπτική ικανότητα σε σχέση με άλλα μοντέλα, όπως η Logistic Regression και το Random Forest. Σε ακόμα ένα πρόβλημα ταξινόμησης, οι Kipkoge et al. (2021) χρησιμοποίησαν ένα μοντέλο XGBoost για να ταξινομήσουν τις επιχειρήσεις σε επιτυχείς και αποτυχημένες, δηλαδή για να προσδιορίσουν εάν έχουν κλείσει ή αν εξακολουθούν να λειτουργούν και είναι κερδοφόρες. Το μοντέλο XGBoost σε αυτό το πλαίσιο ανάλυσης κατάφερε να επιτύχει ακρίβεια 96.15%, ξεπερνώντας σε ακρίβεια μοντέλα όπως Random Forest, Decision Tree, Logistic Regression. Τέλος, ο αλγόριθμος XGBoost χρησιμοποιήθηκε σε ακόμα ένα πρόβλημα ταξινόμησης. Οι Prasad et al. (2022) χρησιμοποίησαν το προαναφερθέν μοντέλο για να προβλέψουν εάν η τιμή του VIX την επόμενη μέρα θα αυξηθεί ή θα μειωθεί σε σύγκριση με την προηγούμενη ημέρα, καθώς και σε εβδομαδιαία βάση. Το μοντέλο κατάφερε να επιτύχει ακρίβεια 62% για ημερήσια συχνότητα και 60% για εβδομαδιαία συχνότητα στο υποσύνολο δοκιμής. Συμπεράναν ότι το μοντέλο παρουσιάζει αποδεκτή απόδοση και πρέπει να ληφθεί υπόψη, καθώς αποδίδει καλύτερα από τη logistic regression. Το Bitcoin είναι μια τεχνολογία που έχει πολλά θετικά, όπως γρηγορότερες και πιο ασφαλείς συναλλαγές. Παρόλα αυτά, ο τρόπος λειτουργίας του απαιτεί τεράστια ποσά ηλεκτρικής ενέργειας, γεγονός που τα τελευταία χρόνια έχει τραβήξει το ενδιαφέρον. Παρακάτω θα γίνει μια σύντομη βιβλιογραφική ανασκόπηση σχετικά με την κατανάλωση ενέργειας του Bitcoin.

Ο Vranken (2017) μιλάει για έναν απαιτητικό ενεργειακό αλγόριθμο (proof-of-work) που χρησιμοποιεί το κρυπτονόμισμα Bitcoin, ο οποίος αποτρέπει την αλλοίωση συναλλαγών που έχουν ήδη επιβεβαιωθεί και επίσης αποτρέπει τη δι-

πλή δαπάνη του ίδιου νομίσματος. Επίσης, αναφέρει την εξέλιξη του υλικού που χρησιμοποιείται για την εξόρυξη του Bitcoin. Αναφέρει ότι, όσο αυξάνεται ο ανταγωνισμός, η εξόρυξη ενός Bitcoin θα γίνεται πιο δύσκολη, επομένως θα επιβιώσουν αυτοί που χρησιμοποιούν τον ενεργειακά πιο αποδοτικό εξοπλισμό εξόρυξης. Βάσει των παραπάνω, κατέληξε στο συμπέρασμα ότι η ενεργειακή κατανάλωση του Bitcoin είναι 100 MW. Ένας από τους προτεργάτες του συγκεκριμένου πλαισίου, ο de Vries (2018), με αρκετά άρθρα για την συγκεκριμένη θεματολογία, αναφέρει διάφορες μεθοδολογίες που έχουν χρησιμοποιηθεί για τον προσδιορισμό της τρέχουσας και μελλοντικής κατανάλωσης ηλεκτρικής ενέργειας του Bitcoin. Με βάση αυτές τις μεθοδολογίες, καταλήγει στο συμπέρασμα ότι η κατανάλωση ηλεκτρικής ενέργειας του Bitcoin είναι τουλάχιστον 2.55GW, με δυνατότητα στο μέλλον να κυμανθεί έως τα 7.67GW, νούμερο συγκρίσιμο με χώρες όπως η Ιρλανδία (3,1 GW) και η Αυστρία (8,2 GW). Επιπλέον, βασιζόμενος στις ημερήσιες συναλλαγές του δικτύου Bitcoin, καταλήγει στο ότι η μέση κατανάλωση ηλεκτρικής ενέργειας που προκύπτει από μία συναλλαγή είναι τουλάχιστον 300kWh, τονίζοντας παράλληλα ότι αυτό είναι ένα σοβαρό πρόβλημα στο οποίο πρέπει να βρεθεί λύση. Ο de Roos (2021), στο άρθρο του ανέπτυξε ένα δυναμικό μοντέλο που χρησιμοποιείται για την πρόβλεψη της κατανάλωσης ενέργειας του Bitcoin. Στο μοντέλο του χρησιμοποίησε μεταβλητές όπως τα κομμάτια που αποτελούν ένα σύστημα εξόρυξης Bitcoin, μεταβλητές που σχετίζονται με το blockchain, δηλαδή εκεί όπου αποθηκεύονται οι συναλλαγές, χρηματοοικονομικές παράμετροι όπως τα κέρδη των εξορυκτών, ο αριθμός των συστημάτων εξόρυξης και τέλος τα συστήματα ψύξης που χρησιμοποιούνται. Κατέληξε στο συμπέρασμα ότι για το 2021, η κατανάλωση ενέργειας του Bitcoin κυμαίνεται από 38.5TWh έως 166TWh. Ένα μοντέλο vector error correction χρησιμοποιήθηκε από τους Schinckus et al. (2022) για να εξετάσουν τη σχέση μεταξύ του hashrate και της κατανάλωσης ενέργειας του Bitcoin. Το hashrate είναι η δυσκολία με την οποία επιβεβαιώνεται μια συναλλαγή, και άρα η υπολογιστική ισχύς που απαιτείται αυξάνεται όσο μεγαλώνει η δυσκολία. Κατέληξαν στο συμπέρασμα ότι το hashrate του Bitcoin έχει θετική επίδραση στην ενεργειακή του κατανάλωση. Στην συνέχεια, οι Zhang et al. (2023) εξετάζουν την σχέση μεταξύ της κατανάλωσης ενέργειας του Bitcoin και των εκπομπών CO_2 , χρησιμοποιώντας την αιτιότητα Granger σε ποσοστά και τη δυναμική συνδεσιμότητα. Βρίσκουν μια σημαντική ασύμμετρη αιτιότητα Granger μεταξύ της κατανάλωσης ηλεκτρικής ενέργειας του Bitcoin και των εκπομπών CO_2 . Πιο συγκεκριμένα, διαπιστώνουν ότι οι χρονοσειρές της μεταβλητής κατανάλωσης ενέργειας του Bitcoin είναι χρήσιμες στην πρόβλεψη των εκπομπών CO_2 , και αυτό το εντόπισαν σε υψηλές τιμές των υπό εξέταση μεταβλητών.

Στη συγκεκριμένη διατριβή, χρησιμοποιούμε ένα μοντέλο extreme gradient

boost regressor για να προβλέψουμε την κατανάλωση ενέργειας του Bitcoin. Χρησιμοποιώντας 11 ανεξάρτητες μεταβλητές, εκ των οποίων κάποιες αφορούν την τεχνική φύση του δικτύου του Bitcoin, ενώ οι υπόλοιπες σχετίζονται με την οικονομική φύση του Bitcoin. Η συχνότητα με την οποία συλλέξαμε τα δεδομένα ήταν ημερήσια, και η χρονική περίοδος κατά την οποία εξετάσαμε την απόδοση του μοντέλου μηχανικής μάθησης και συλλέξαμε δεδομένα για τις μεταβλητές ήταν από τις 10 Φεβρουαρίου 2017 έως τις 7 Ιουλίου 2020.

Η συγκεκριμένη θεματολογία είναι αρκετά καινούρια και υπάρχει σημαντικός χώρος για βελτίωση του πλαισίου μελέτης. Κύριως αποτελείται από θεωρητικά πλαίσια βάσει των οποίων γίνεται εκτίμηση της κατανάλωσης ενέργειας του Bitcoin, ή χρησιμοποιούνται στατιστικά μοντέλα ή μοντέλα μηχανικής μάθησης για να μελετηθεί η σχέση διαφόρων μεταβλητών με την κατανάλωση ενέργειας του Bitcoin. Η συνεισφορά μας σε αυτό το πλαίσιο είναι η χρήση ενός νέου μοντέλου για την πρόβλεψη της κατανάλωσης ενέργειας του Bitcoin, το οποίο είναι πολλά υποσχόμενο καθώς έχει δείξει εξαιρετικά αποτελέσματα σε μια πληθώρα εφαρμογών. Το μοντέλο ονομάζεται XGBoost. Επίσης, το παραπάνω πρόβλημα πρόβλεψης αντιμετωπίζεται ως πρόβλημα παλινδρόμησης και όχι ταξινόμησης, όπως συμβαίνει στην πλειονότητα της βιβλιογραφίας. Επίσης τίθεται το πρόβλημα του κατάλληλου διαχωρισμού της μεταβλητής σε κατηγορίες, εφόσον η μεταβλητή του πλαισίου μελέτης είναι συνεχής. Επομένως, εξετάζουμε την απόδοση ενός νέου μοντέλου μηχανικής μάθησης σε πρόβλημα παλινδρόμησης σε ένα απαιτητικό πλαίσιο εφαρμογής, καθώς η εξαρτημένη μεταβλητή χαρακτηρίζεται από μεγάλη μεταβλητότητα εντός του διαστήματος στο οποίο εφαρμόζεται το μοντέλο, για το οποίο, βάσει της βιβλιογραφίας, υπάρχουν ελάχιστες έως καθόλου αναφορές. Κρίνουμε εάν είναι κατάλληλο για το συγκεκριμένο πλαίσιο, συμβάλλοντας έτσι στην εύρεση κατάλληλων μοντέλων για την αντιμετώπιση του προβλήματος της ενεργειακής κατανάλωσης του Bitcoin, το οποίο επιβαρύνει επιπλέον τον πλανήτη μας. Επιπλέον εξετάζουμε, μεταξύ άλλων, αν οι ακραίες τιμές επηρεάζουν την απόδοση και ποια αναλογία εκπαίδευσης/δοκιμής οδηγεί στην καλύτερη απόδοση. Χρησιμοποιούμε και συγκριτικά μοντέλα(μοντέλο πολλαπλής γραμμικής παλινδρόμηση, Gradient Boosting regression tree) ώστε να έχουμε μια καλύτερη κατανόηση της απόδοσης του μοντέλου στο υπό εξέταση πλαίσιο.

2 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 Το Bitcoin και η κατανάλωση ενέργειας

Το Bitcoin έχει φέρει επανάσταση σχετικά με τους ενδεδειγμένους τρόπους για την πραγματοποίηση μιας χρηματοοικονομικής συναλλαγής. Τα τελευταία χρόνια έχουν αρχίσει να προκύπτουν αμφιβολίες ανάμεσα στην επιστημονική κοινότητα σχετικά με αυτήν την νέα τεχνολογία και το αν τα θετικά που προσφέρει μπορούν να καλύψουν τα εν δυνάμει αρνητικά της. Ο Vranken (2017), εστιάζοντας στη λειτουργία μέσω της οποίας γίνεται η εξόρυξη Bitcoin και εξετάζοντας την εξέλιξη στον υπολογιστικό εξοπλισμό που χρησιμοποιείται για αυτήν, κατέληξε ότι για την συγκεκριμένη περίοδο η ενεργειακή κατανάλωση του Bitcoin είναι 100MW. Επίσης, έγινε εξέταση συστημάτων εξόρυξης Bitcoin που απαιτούν πολύ λιγότερη ενέργεια και θα μπορούσαν να περιορίσουν το πρόβλημα, όπως το proof-of-stake. Στο proof-of-stake η δυσκολία εξόρυξης προσδιορίζεται ατομικά για κάθε παίκτη, σε σύγκριση με το proof-of-work, όπου όσο αυξάνεται ο ανταγωνισμός, αυξάνεται και η δυσκολία εξόρυξης και άρα η κατανάλωση ενέργειας. Στην συνέχεια, ο de Vries (2018) περιγράφει μεθοδολογίες που στοχεύουν στην εκτίμηση της τρέχουσας και μελλοντικής κατανάλωσης ηλεκτρικής ενέργειας του δικτύου Bitcoin. Σύμφωνα με τις εκτιμήσεις, το Bitcoin καταναλώνει τουλάχιστον 2.55GW ηλεκτρικής ενέργειας, ενώ στο μέλλον θα μπορούσε να φτάσει στα 7.67GW, μέγεθος συγκρίσιμο με χώρες όπως η Ιρλανδία(3.1GW) και η Αυστρία(8.2GW). Επίσης, με βάση τις συναλλαγές ανά ημέρα, καταλήγει στο συμπέρασμα ότι η μέση κατανάλωση ενέργειας ανά συναλλαγή είναι 300kWh, με το συγκεκριμένο νούμερο να μπορεί να ξεπεράσει τα 900kWh μέχρι το τέλος του 2018. Τονίζοντας έτσι ένα μεγάλο πρόβλημα του Bitcoin που θα πρέπει να ληφθεί σοβαρά υπόψη.

Την ίδια χρονιά, μια διαφορετική οπτική του ίδιου προβλήματος δόθηκε από τον Truby (2018). Ο οποίος προτείνει δημοσιονομικές και ρυθμιστικές πολιτικές, με στόχο μια περιβαλλοντικά βιώσιμη ανάπτυξη των εφαρμογών που αφορούν τη λειτουργία του Bitcoin. Χωρίς αυτό να σημαίνει ότι θα υπάρξει κάποια ζημιά στα πλεονεκτήματα που προσφέρει μια τέτοια τεχνολογία. Μια από τις πολλές πολιτικές που προτάθηκαν είναι μέσω της χρήσης δημοσιονομικών εργαλείων να ενθαρρυνθεί η κατασκευή πιο αποτελεσματικών τεχνολογιών που βοηθούν στην επίτευξη περιβαλλοντικών στόχων. Ο Dilek και Furuncu (2019) μελέτησαν επιπτώσεις από την υψηλή κατανάλωση ενέργειας του Bitcoin. Σύμφωνα με τη μελέτη τους, η ενέργεια που απαιτείται για τις συναλλαγές και την εξόρυξη του Bitcoin προέρχεται από άνθρακα και υδρογονάνθρακες. Τέτοιες πηγές ενέργειας προκαλούν αύξηση της θερμοκρασίας, της ρύπανσης ακόμα και το ποσοστών

θνησιμότητας, με αποτέλεσμα να θέτουν σε κίνδυνο τον πλανήτη. Βλέπουμε έως τώρα, σύμφωνα με τη βιβλιογραφία, ότι ενώ το Bitcoin έχει πολλά θετικά, έχει σοβαρά μειονεκτήματα που πρέπει οπωσδήποτε να ληφθούν υπόψη. Ο Stoll et al. (2019) εκτιμούν την κατανάλωση ενέργειας της τεχνολογίας blockchain του Bitcoin χρησιμοποιώντας αναφορές από την IPO για κατασκευαστές εξοπλισμού υπολογιστών για εξόρυξη Bitcoin, πληροφορίες σχετικές με την κλίμακα εξόρυξης Bitcoin, και τέλος, πληροφορίες σχετικές με δημόσιες ομάδες εξόρυξης. Μετά τη συλλογή των παραπάνω πληροφοριών, μετέφρασαν τα αποτελέσματα κατανάλωσης ενέργειας από την εκτίμηση τους σε εκπομπές άνθρακα χρησιμοποιώντας τις διευθύνσεις IP. Με βάση τις παραπάνω πληροφορίες, εκτίμησαν ότι η ετήσια κατανάλωση ηλεκτρικής ενέργειας του Bitcoin είναι 45.8TWh μέχρι τον Νοέμβριο του 2018 και οι ετήσιες εκπομπές άνθρακα κυμαίνονται μεταξύ 22.0 και 22.9 MtCO₂. Αυτές οι μετρήσεις είναι συγκρίσιμες με χώρες όπως Ιορδανία, Σρι Λάνκα και το Kansas City.

Εν συνεχεία, ένα χρόνο μετά, οι Küfeoğlu και Özkuran (2019) εκτίμησαν την κατανάλωση ενέργειας που προέρχεται από τη διαδικασία proof-of-work του Bitcoin. Χρησιμοποίησαν 160GB δεδομένων blockchain. Στο blockchain πραγματοποιούνται και αποθηκεύονται όλες οι συναλλαγές. Επιπλέον, έλαβαν υπόψη την απόδοση 269 μοντέλων υλικού CPU, GPU, FPGA, ASIC που χρησιμοποιούνται στη διαδικασία εξόρυξης Bitcoin. Με βάση τα παραπάνω στοιχεία, υπολόγισαν δύο μετρικές: την ελάχιστη και τη μέγιστη κατανάλωση ενέργειας. Η ελάχιστη κατανάλωση προκύπτει από τη χρήση του πιο αποδοτικού μοντέλου από όλους τους συμμετέχοντες στην εξόρυξη Bitcoin. Η μέγιστη κατανάλωση προκύπτει από τη χρήση της λιγότερης αποδοτικότητας του υλικού που απαιτείται για να υπάρξει κερδοφορία από την εξόρυξη Bitcoin. Με βάση τα παραπάνω στοιχεία, βρήκαν ότι η μέγιστη κατανάλωση ενέργειας, κατά την εξεταζόμενη περίοδο (2009-2018), καταγράφηκε σε μια περίοδο δύο εβδομάδων, η οποία ξεκίνησε από τις 18 Δεκεμβρίου 2017 και κυμαίνονταν μεταξύ 1.3 και 14.8GW. Αυτά τα νούμερα συγκρίνονται με τις χώρες όπως η Δανία (~ 14GW) και η Φινλανδία (~ 16GW). Επίσης, παρατήρησαν ότι η ενεργειακή κατανάλωση του Bitcoin επηρεάζεται από την τιμή του. Με την πτώση της τιμής παρατήρησαν πτώση στη μέγιστη κατανάλωση ενέργειας. Αναφέρουν επίσης τη δυσκολία της εκτίμησης λόγω παραγόντων όπως η μεγάλη μεταβολή της τιμής και η δυσκολία αποτίμησης της αξίας του Bitcoin, οι οποίοι επηρεάζουν άμεσα την κατανάλωση ενέργειας, καθώς και την αποδοτικότητα του υλικού που χρησιμοποιούν οι συμμετέχοντες στην αγορά Bitcoin.

Αυτή η συνεχής αύξηση της δημοτικότητας του Bitcoin, κάνοντας το το πιο δημοφιλές και αποδεκτό κρυπτονόμισμα παγκοσμίως, συνέχισε να αυξάνει το ενδιαφέρον στην επιστημονική κοινότητα. Ο Mir (2020) αναλύει τα θετικά και

τα αρνητικά που συνεπάγεται μια τεχνολογία όπως το Bitcoin και αναφέρει προκλήσεις που σχετίζονται με την κατανάλωση ενέργειας του. Κάποια από τα πολλά θετικά που έχει η τεχνολογία Bitcoin σε σύγκριση με τις συμβατικές μεθόδους συναλλαγών είναι η γρήγορη επεξεργασία, κάνοντας τις συναλλαγές ταχύτερες, και η μεγάλη ασφάλεια, καθώς οι συναλλαγές δεν αποθηκεύουν τα προσωπικά στοιχεία του αποστολέα ή του παραλήπτη. Προτείνει κάποιες εναλλακτικές λύσεις σε σχέση με τη διαδικασία εξόρυξης του Bitcoin (proof of work), όπως το Proof of Authentication, Intel's Proof-of-Elapsed. Αυτές οι εναλλακτικές προτάσεις είναι πιο αποδοτικοί αλγόριθμοι επιβεβαίωσης συναλλαγών σε σύγκριση με τον αλγόριθμο που χρησιμοποιεί το Bitcoin, όπου ο ισχυρότερος υπολογιστής, και συνεπώς η μεγαλύτερη κατανάλωση ενέργειας, θα επιβεβαιώσει τη συναλλαγή και θα πάρει την ανταμοιβή του. Επιπλέον, υπάρχει πρόβλημα με τις πηγές ενέργειας, και μια λύση θα μπορούσε να είναι η αιολική ενέργεια ή η ενέργεια από φράγματα για την παραγωγή ηλεκτρικής ενέργειας. Συνεπώς, αν και τα θετικά είναι αρκετά για να μην αγνοηθούν οι δυνατότητες της τεχνολογίας, χρειάζονται ριζικές αλλαγές στον τρόπο με τον οποίο γίνονται η επιβεβαίωση των συναλλαγών και οι πηγές ενέργειας που χρησιμοποιούνται.

Για ακόμα μία φορά, ο de Vries (2020) ήρθε να συμβάλει στη βιβλιογραφία του Bitcoin προτείνοντας μια προσέγγιση δυναμικής αγοράς για την εκτίμηση της κατανάλωσης ενέργειας. Είναι σημαντικό να λαμβάνονται υπόψη οι συνθήκες που περιβάλλουν την αγορά, διότι σύμφωνα με αυτές επηρεάζονται οι συμμετέχοντες, και σύμφωνα με τους συμμετέχοντες επηρεάζεται η κατανάλωση ενέργειας του Bitcoin. Σε αντίθεση με άλλες εκτιμήσεις της κατανάλωσης ενέργειας του Bitcoin, οι οποίες υποθέτουν στατικότητα στην αγορά, παρέχοντας έτσι αισιόδοξα αποτελέσματα, η παραπάνω προσέγγιση αναγνωρίζει την επίδραση μιας δυναμικής αγοράς. Λόγω της τεράστιας ανάπτυξης και κερδοφορίας του Bitcoin, δίνεται η δυνατότητα στους συμμετέχοντες στην αγορά να μην επιλέγουν την πιο αποδοτική λύση. Με βάση τα παραπάνω και μια λίστα πωλήσεων Bitcoin, πραγματοποιήθηκε μια συντηρητική εκτίμηση της κατανάλωσης ενέργειας του Bitcoin στα 87.1TWh ετησίως έως τις 30 Σεπτεμβρίου 2019, ξεπερνώντας άλλες εκτιμήσεις που είχαν γίνει εκείνο το χρονικό διάστημα (73.1 έως 78.3TWh ετησίως). Αυτός ο αριθμός είναι ισοδύναμος με την κατανάλωση ενέργειας του Βελγίου (87.9TWh) και περίπου το μισό της κατανάλωσης ενέργειας των κέντρων δεδομένων παγκοσμίως (200TWh). Σε συνέχεια της προηγούμενης αναφοράς, ο de Vries (2021), γνωστός από τις προηγούμενες μελέτες του για την ενεργειακή κατανάλωση του Bitcoin, αναφέρει τις αρνητικές συνέπειες που έχει μια αύξηση της τιμής στην κατανάλωση ενέργειας του Bitcoin. Σύμφωνα με τον συγγραφέα, όσο τα έσοδα από ένα Bitcoin είναι υψηλότερα από τα έξοδα για την εξόρυξη του, τότε ένας συμμετέχων στην αγορά μπορεί να αυξήσει την υπολογιστική του

ισχύ και, κατά συνέπεια, να έχει περισσότερα κέρδη, και αυτό θα συνεχίζεται μέχρι να επέλθει η ισορροπία στην αγορά. Σύμφωνα με αυτό το απλό οικονομικό μοντέλο και το γεγονός ότι τα έξοδα ενός miner είναι το σύστημα του και η ηλεκτρική ενέργεια που καταναλώνει, μπορεί να γίνει μια εκτίμηση για τις πιθανές συνέπειες της κατανάλωσης ενέργειας του Bitcoin. Τέλος αξιοποιώντας αυτό το πολύ απλό μοντέλο και υποθέτοντας ότι οι miners πληρώνουν 0.05\$ ανά κιλοβατώρα, με μια τιμή Bitcoin 42.000\$, με τέλη συναλλαγών να είναι το 10% του εισοδήματος των miners και τέλος το 60% του εισοδήματος τους να πηγαίνει για ηλεκτρική ενέργεια, εκτιμάται ότι η συνολική κατανάλωση ενέργειας του Bitcoin είναι 184TWh τον χρόνο. Αυτό είναι πολύ κοντά στην ηλεκτρική κατανάλωση όλων των κέντρων δεδομένων παγκοσμίως (200TWh).

Σύμφωνα με το άρθρο των Huynh et al. (2022) χρησιμοποιώντας variance decompositions σε συνδυασμό με realized semi-variances, βρήκαν μια αμφίδρομη επίδραση μεταξύ της κατανάλωσης ενέργειας και των αποδόσεων του Bitcoin, καθώς και μεταξύ της κατανάλωσης ενέργειας και των όγκων συναλλαγών. Με βάση τα αποτελέσματα τους, κατέληξαν στο συμπέρασμα ότι ο όγκος συναλλαγών Bitcoin έχει σημασία για την κατανάλωση ενέργειας. Επίσης, βρήκαν ότι μια πτώση στην τιμή του Bitcoin δεν μειώνει μόνο την κατανάλωση ενέργειας, αλλά επηρεάζει επίσης την μακροχρόνια σύνδεση μεταξύ τιμής και ενεργειακής κατανάλωσης. Οι Meng et al. (2023) μελετούν την αλληλεπίδραση της τιμής του Bitcoin, των εκπομπών CO_2 και της κατανάλωσης ενέργειας. Χρησιμοποιώντας μια αυτοπαλίνδρομη διαδικασία με τυχαία μεταβλητότητα και χρονικές παραμέτρους, παρατηρούν την αλληλεπίδραση μεταξύ των παραπάνω μεταβλητών, η οποία εξαρτάται από τον χρόνο. Με βάση την παραπάνω μελέτη, καταλήγουν στο συμπέρασμα ότι η τιμή του Bitcoin επηρεάζει τόσο θετικά όσο και αρνητικά της εκπομπές άνθρακα και την κατανάλωση ενέργειας, με την πλειονότητα των επιδράσεων να είναι θετικές. Η θετική επίδραση μεταφράζεται σε αύξηση των εκπομπών και της κατανάλωσης ενέργειας με την αύξηση της τιμής του Bitcoin. Σε περιόδους αρνητικής επίδρασης, οι εκπομπές και η κατανάλωση ενέργειας επηρεάζονται περισσότερο από εποχιακούς παράγοντες. Επίσης, Φαίνεται να υπάρχει μια αμφίδρομη σχέση μεταξύ των μεταβλητών, προτείνοντας κάποια μέτρα για την αντιμετώπιση της κατανάλωσης ενέργειας και εκπομπών CO_2 , όπως η χρήση βιώσιμων πηγών ενέργειας και η χρήση πιο αποδοτικών μηχανημάτων, όπως τα ASIC.

Η παραπάνω έρευνα του Kohli et al. (2023), συλλέγοντας από διάφορους δείκτες οι οποίοι εκτιμούν την κατανάλωση ενέργειας του Bitcoin από το 2018-2021(Digiconomist and the Cambridge Bitcoin Energy Consumption Index (CBECI)), συγκρίνει τους δείκτες κατανάλωσης ενέργειας με χώρες και διάφορες μεθόδους συναλλαγών όπως η Visa. Αυτή η σύγκριση έδειξε ότι η κατανάλωση

ενέργειας του Bitcoin τον Ιούλιο του 2021 είναι συγκρίσιμη με χώρες όπως η Σουηδία και Ρουμανία. Επιπλέον, η ανασκόπηση έδειξε ότι η μέθοδος συναλλαγών Visa είναι πολύ πιο ενεργειακά αποτελεσματική από αυτή του Bitcoin. Εντόπισε επίσης βασικά προβλήματα που προκαλούν την υψηλή κατανάλωση ενέργειας, όπως η υπερβολική δικτύωση και οι συσκευές εξόρυξης. Προτείνει και λύσεις όπως αλλαγή του μηχανισμού επικύρωσης συναλλαγών proof-of-work με εναλλακτικές όπως PoSpace, PoST, PoA, uPoW, REM, PoS.

2.2 Bitcoin και Ενεργειακή Κατανάλωση: Ποσοτική Ανάλυση

Με την προσθήκη δεικτών που παρακολουθούν την εκτιμώμενη κατανάλωση ενέργειας του Bitcoin, σε συνδυασμό με την αυξανόμενη φήμη τους, δόθηκε η δυνατότητα σε εμπειρικές μελέτες να διεξάγονται σχετικά με την κατανάλωση ενέργειας του Bitcoin σε εμπειρικό επίπεδο. Γνωστοί ως Cambridge Bitcoin Electricity Consumption Index (CBECI) και Digiconomist Bitcoin Energy Consumption Index. Ο Sizer και Yilmaz (2022) εξετάζουν την αιτιώδη σχέση μεταξύ του δείκτη της παγκόσμιας οικονομικής και πολιτικής αβεβαιότητας και της κατανάλωσης ηλεκτρικής ενέργειας Bitcoin. Για την κατανάλωση ηλεκτρικής ενέργειας χρησιμοποίησαν τον Digiconomist Bitcoin Energy Consumption Index και δεδομένα από την περίοδο 2011-2022. Εφαρμόζοντας ανάλυση αιτιότητας Toda-Yamamoto, βρήκαν ότι υστερήσεις και των δύο μεταβλητών βοηθούν η μία στην πρόβλεψη της άλλης, έχοντας έτσι μια σχέση Granger causality σε επίπεδο σημαντικότητας 5%. Σύμφωνα με τους Calvo-Pardo et al. (2022), χρησιμοποίησαν ένα μοντέλο feed-forward neural networks για να μετρήσουν το αποτύπωμα άνθρακα από την εξόρυξη Bitcoin για τα έτη 2017-2019. Χρησιμοποιώντας 42 μεταβλητές στο μοντέλο και μεταφράζοντας τον δείκτη (CBECI) σε ανθρακικό αποτύπωμα, βρήκαν αποτυπώματα άνθρακα που ανέρχονται για το 2017 έως το 2019 σε 2.77, 16.08, 14.99 $MtCO_2$ αντίστοιχα. Τέλος, για τον τελικό υπολογισμό των εκπομπών άνθρακα στο περιβάλλον χρησιμοποιήθηκαν καταθέσεις IPO και διαστήματα εμπιστοσύνης. Για την τάξη μεγέθους, τα παραπάνω αποτελέσματα για το έτος 2018 είναι συγκρίσιμα με πολιτείες των ΗΠΑ όπως Maine, New Hampshire και μικρές χώρες όπως Βολιβία και το Σουδάν.

Η ίδια χρονία με μια ακόμα δημοσίευση ήρθε για να ενισχύσει την βιβλιογραφία από τους Schinckus et al. (2022). Χρησιμοποιώντας ένα μοντέλο vector error correction model, εξετάζουν την σχέση μεταξύ του Bitcoin hashrate και της κατανάλωσης ενέργειας για μια περίοδο από το 2016-2021. Σύμφωνα με τους συγγραφείς το Bitcoin hashrate έχει θετική σχέση με την κατανάλωση ενέργειας, η οποία είναι ισχυρότερη βραχυπρόθεσμα. Ένα έντονο ενδιαφέρον έχει κλιμακωθεί σχετικά με τη βιωσιμότητα μια τεχνολογίας όπως το Bitcoin. Χρησιμοποιώντας comprising dynamic ARDL simulations και general-to-specific VAR, οι Sarkodie et al. (2022) εξέτασαν την σχέση μεταξύ του όγκου συναλλαγών Bitcoin με την κατανάλωση ενέργειας και το ανθρακικό αποτύπωμα του Bitcoin. Ως προσέγγιση για την ετήσια κατανάλωση ενέργειας του Bitcoin χρησιμοποίησαν τον Digiconomist Bitcoin Energy Consumption Index, με την υπό εξέταση περίοδο να είναι από το 2017 έως το 2021. Με βάση τη μεθοδολογία και την υπό εξέταση περίοδο, βρήκαν ότι μια αύξηση του όγκου συναλλαγών Bitcoin προκαλεί αύξηση του ανθρακικού και του ενεργειακού αποτυπώματος Bitcoin κα-

τά 24%. Τέλος, ένα σοκ στον όγκο συναλλαγών προκαλεί αύξηση 46.54% στο αποτύπωμα ενέργειας και άνθρακα.

Ένα βασικό ερώτημα που προκύπτει από τις παραπάνω αναφορές και την γενική ανασκόπηση της βιβλιογραφίας είναι αν τα οφέλη που προσφέρει μια τεχνολογία όπως το Bitcoin αξίζουν τις συνέπειες που θα προκαλέσει. Βασισμένοι στον Cambridge Bitcoin Electricity Consumption Index (CBECI) ως εκτίμηση για την ενεργειακή κατανάλωση του Bitcoin σε καθημερινή βάση, και χρησιμοποιώντας ως ανεξάρτητες μεταβλητές τον μέσο όγκο συναλλαγών πέντε κρυπτονομισμάτων (bitcoin, ethereum, ripple, stellar, litecoin) από το 2015 έως το 2021. Ο Fahmy (2022) χρησιμοποίησε ένα μοντέλο bivariate vector autoregression (VAR) και Granger causality test. Επιβεβαιώνοντας μια μονόδρομη αιτιακή σχέση από τον όγκο συναλλαγών των κρυπτονομισμάτων στην κατανάλωση ενέργειας του Bitcoin. Επίσης, φαίνεται ότι τα σοκ στον όγκο συναλλαγών των κρυπτονομισμάτων επηρεάζουν την κατανάλωση ενέργειας. Καταλήγοντας στο συμπέρασμα ότι το 2026 το ανθρακικό αποτύπωμα του Bitcoin θα ισοδυναμεί με το 1.6% των παγκόσμιων εκπομπών. Οι Zhang et al. (2023) εξέτασαν την σχέση της εξόρυξης Bitcoin και της κλιματικής αλλαγής χρησιμοποιώντας granger causality σε ποσοστιαία τμήματα. Βρήκαν ότι όταν οι τιμές CO_2 είναι υψηλές, τότε η κατανάλωση ενέργειας που προκύπτει από την εξόρυξη Bitcoin έχει σημαντική προβλεπτική ικανότητα, υποδηλώνοντας μια ασύμμετρη σχέση μεταξύ των δύο μεταβλητών. Επίσης, βρήκαν ότι η μεταβλητότητα, απόδοση του Bitcoin, το hashrate, το μέγεθος blockchain και η δυσκολία του δικτύου έχουν θετική επίδραση στην κατανάλωση ηλεκτρικής ενέργειας του Bitcoin όταν αυτή βρίσκεται πάνω από την διάμεσο.

Οι Sapra και Shaikh (2023) χρησιμοποίησαν ένα μοντέλο OLS, ARDL με δεδομένα από το 2018 έως το 2023 και Cambridge Bitcoin Electricity Consumption Index ως εξαρτημένη μεταβλητή για την κατανάλωση ενέργειας του Bitcoin. Το μέγεθος του μπλοκ Bitcoin, ο ρυθμός κατακερματισμού στο δίκτυο Bitcoin και ο αριθμός των συναλλαγών Bitcoin βρέθηκε ότι έχουν θετική σχέση με την κατανάλωση ενέργειας και επίσης έχουν σημαντική επιρροή σε αυτήν. Επιπροσθέτως, ο CIX200, και η τιμή του Ethereum έχουν σημαντική επίδραση στην κατανάλωση ενέργειας. Συγκεκριμένα η μεταβλητή CIX200 έχει μια θετική συσχέτιση, ενώ η τιμή του Ethereum έχει αρνητική επίδραση στην κατανάλωση ενέργειας του Bitcoin βραχυπρόθεσμα και θετική επίδραση μακροπρόθεσμα. Με βάση τη μελέτη των Ben Omrane et al. (2023), χρησιμοποιώντας δεδομένα από το 2019 έως το 2020, Βρέθηκε ότι οι μακροοικονομικές ειδήσεις είναι εφικτό να αυξήσουν την κατανάλωση ενέργειας και, άρα το ανθρακικό αποτύπωμα των κρυπτονομισμάτων. Επίσης, οι αποφάσεις επιτοκίων που λαμβάνονται από τη Federal Reserve επηρεάζουν θετικά και σημαντικά τον όγκο συναλλαγών του Bitcoin,

με αποτέλεσμα την αυξημένη κατανάλωση ενέργειας. Σύμφωνα με τους Sapra et al. (2024) χρησιμοποιώντας ένα μοντέλο VAR, ένα τεστ Granger causality, τη συνδεσιμότητα Diebold-Yilmaz και στοιχεία από την περίοδο 2019-2023, ως προσέγγιση για την κατανάλωση ενέργειας του Bitcoin χρησιμοποίησαν τον Cambridge Bitcoin-based Electricity Consumption Index(CBECI). Βάσει της μελέτης τους, βρήκαν αμφίδρομη σχέση μεταξύ του δείκτη CVI και της κατανάλωσης ενέργειας του Bitcoin. Επίσης, βρήκαν ότι η τιμή του Bitcoin επηρεάζει την κατανάλωση ενέργειας και τέλος η κατανάλωση ενέργειας του Bitcoin είναι καθαρός αποδέκτης της αστάθειας.

Την ίδια χρονιά, οι Jamali et al. (2024) χρησιμοποίησαν τον δείκτη Cambridge Bitcoin Electricity Consumption Index (CBECI) για την κατανάλωση ενέργειας του Bitcoin. Για την ανάλυση τους, χρησιμοποίησαν ένα μοντέλο εμποτευόμενης μηχανικής μάθησης(Regression). Για δεδομένα εκπομπών πήραν στοιχεία από το 2017 έως το 2021, ενώ για την τιμή του Bitcoin χρησιμοποίησαν στοιχεία από το 2011 έως το 2021. Βρήκαν ότι υπάρχει μια μέτρια θετική συσχέτιση μεταξύ την κατανάλωσης ενέργειας του Bitcoin και τον εκπομπών CO_2 , χρησιμοποιώντας τον Pearson correlation coefficient. Σύμφωνα με το μοντέλο το οποίο χρησιμοποίησαν, βρήκαν στατιστικά σημαντική σχέση μεταξύ της τιμής του Bitcoin, της κατανάλωσης ενέργειας του Bitcoin, και των εκπομπών CO_2 , όπου οι εκπομπές CO_2 χρησιμοποιήθηκαν ως εξαρτημένη μεταβλητή.

2.3 Gradient Boosting

Οι Li et al. (2019) χρησιμοποιούν Tweets για ένα χρονικό διάστημα 3.5 εβδομάδων και τον όγκο συναλλαγών ώστε να προβλέψουν τη διακύμανση της τιμής ενός κρυπτονομίσματος που ονομάζεται ZClassic. Ένα μοντέλο Extreme Gradient Boosting Regression Tree φάνηκε να είχε πολύ καλά αποτελέσματα στην ανάλυση συναισθήματος, με τις προβλεπόμενες τιμές να έχουν μια συσχέτιση 0.81 με τα ιστορικά δεδομένα και στατιστικά σημαντικές τιμές προβλέψεις. Σύμφωνα με τον Yoon (2021), ο οποίος χρησιμοποίησε τη μεγέθυνση του πραγματικού ΑΕΠ της Ιαπωνίας για να κάνει μια πρόβλεψη για το διάστημα 2001 έως 2018 η πρόβλεψη αυτή συγκρίθηκε με προβλέψεις που έχουν γίνει από το Διεθνές Νομισματικό Ταμείο και την Τράπεζα της Ιαπωνίας. Διαπιστώνει ότι η μέθοδος gradient boosting επιτυγχάνει καλύτερη προβλεπτική ικανότητα και, επομένως μεγαλύτερη ακρίβεια από τα προαναφερθέντα μοντέλα, καθώς και από το μοντέλο random forest. Χρησιμοποιεί ως μέτρα αξιολόγησης μέσο απόλυτο ποσοστό σφάλματος και τη ριζική μέση τετραγωνική απόκλιση. Μια ακόμα πρόβλεψη της ανάπτυξης του ΑΕΠ, όπως στο άρθρο που αναφέρθηκε προηγουμένως, χρησιμοποιώντας αλγόριθμους μηχανικής μάθησης που βασίζονται σε δέντρα, πραγματοποιήθηκε από τους Soybilgen και Yazgan (2021). Με τη διαφορά ότι η πρόβλεψη για την περίοδο από το 2000 έως το 2018 γίνεται ανά τρίμηνο και αφορά της ΗΠΑ. Χρησιμοποιώντας ένα σύνολο μεταβλητών που υπερβαίνει τις 100, βρήκαν ότι τα μοντέλα που βασίζονται σε δέντρα υπερέχουν των γραμμικών μοντέλων, με ένα από τα μοντέλα να είναι το stochastic gradient tree boosting. Επίσης, συγκρίνανε τα μοντέλα μηχανικής μάθησης με το μοντέλο πρόβλεψης της FED, το οποίο ονομάζεται GDPNow, και βρήκαν ότι τα μοντέλα μηχανικής μάθησης υπερέχουν σε επιδόσεις πρόβλεψης στην αρχή του τριμήνου. Την επόμενη χρονιά, έγινε μια βιβλιογραφική ανασκόπηση σχετικά με τις μεθοδολογίες μηχανικής μάθησης και τους λόγους για τους οποίους χρησιμοποιούνται αυτές οι μεθοδολογίες στην οικονομία της υγείας και στην έρευνα εκβάσεων από τους Lee et al. (2022). Συλλέχθηκαν δημοσιευμένες μελέτες από το 2020 έως το 2021 και βρέθηκε ότι το 86% των δημοσιεύσεων τη χρησιμοποιούν για πρόβλεψη. Επίσης, οι πιο δημοφιλείς αλγόριθμοι μηχανικής μάθησης ήταν αλγόριθμοι που βασίζονται σε δέντρα, όπως το random forest, gradient boosting. τα παραπάνω συμπεράσματα προέκυψαν από την ανάλυση 161 άρθρων.

Στο παραπάνω άρθρο των Smith και Alvarez (2022) χρησιμοποιείται ένα μοντέλο μηχανικής μάθησης, το XGBoost, ώστε να κατηγοριοποιήσουν Ισπανικές εταιρείες σε πτωχευμένες ή μη, χρησιμοποιώντας δεδομένα από το 1992 έως το 2016, με ορίζοντα πρόβλεψης από 1 έως 4 χρόνια. Σε όλους τους ορίζοντες πρόβλεψης, το μοντέλο ήταν αρκετά ικανό στο να κατατάσσει σε ποια κατηγορία

ανήκει η εκάστοτε εταιρεία. Επίσης, έγινε σύγκριση με άλλα μοντέλα, όπως το λογιστικό μοντέλο, από το οποίο είχε καλύτερα αποτελέσματα, καθώς και με άλλα μοντέλα μηχανικής μάθησης, στα οποία παρουσίασε παρόμοια αποτελέσματα. Η μελέτη του Atalan (2023) χρησιμοποιεί μεθόδους μηχανικής μάθησης, με μία από αυτές να είναι η μέθοδος gradient boosting, για να προβλέψει την τιμή του γάλακτος στην Τουρκία, με δεδομένα από το 2014 έως το 2021. Ο αλγόριθμος random forest είχε την καλύτερη ακρίβεια, ενώ ο αλγόριθμος gradient boosting είχε μέση ακρίβεια 97.3%. Όλοι οι αλγόριθμοι που χρησιμοποιήθηκαν στην εν λόγω μελέτη κατέληξαν στο συμπέρασμα ότι είναι ικανοί να προβλέψουν την τιμή του γάλακτος. Έχοντας συλλέξει μηνιαία στοιχεία από το 1986 έως το 2019 για την τιμή του χρυσού, οι Ben Jabeur et al. (2024) χρησιμοποίησαν ένα μοντέλο eXtreme Gradient Boosting (XGBoost) καθώς και άλλα πέντε μοντέλα για να προβλέψουν την τιμή του χρυσού και στη συνέχεια να τα συγκρίνουν. Για να αξιολογηθούν οι επιδόσεις των μοντέλων στην πρόβλεψη, χρησιμοποιήθηκαν τέσσερα μέτρα απόδοσης πρόβλεψης, με δύο από αυτά να είναι το root mean square error (RMSE), mean square error (MSE). Το μοντέλο XGBoost φαίνεται, βάσει της παραπάνω έρευνας, να επιτυγχάνει την καλύτερη απόδοση, έχοντας επιλέξει το 80% των δεδομένων ως σετ εκπαίδευσης, ώστε να μπορεί να εκπαιδευτεί το μοντέλο, και το υπόλοιπο ως σετ ελέγχου, για αξιολογηθεί η απόδοση του μοντέλου σε νέα, άγνωστα δεδομένα. Τέλος οι Wijayanti και Setyoko (2024) χρησιμοποιούν τη μέθοδο XGBoost για να προβλέψουν την παραγωγή ρυζιού, συλλέγοντας στοιχεία από τον Food and Agriculture Organization (FAO). Χρησιμοποιώντας ως μέτρα αξιολόγησης τα R^2 , MAE, MSE, το μοντέλο πέτυχε εξαιρετικές επιδόσεις σε σύγκριση με τα υπόλοιπα μοντέλα μηχανικής μάθησης (Random Forest (RF), Gradient Boost (GB), Bagging Regressor (BR), K-Nearest Neighbor (KNN)), καθιστώντας το το καταλληλότερο μοντέλο για προβλέψεις παραγωγής ρυζιού.

Η παρακάτω εργασία είναι δομημένη ως εξής. Στην επόμενη ενότητα(3) αναλύουμε τα δεδομένα που θα χρησιμοποιήσουμε, εισάγοντας και κάποια περιγραφικά στατιστικά. Στη συνέχεια, εξηγούμε την προεπεξεργασία που έγινε στα δεδομένα, ώστε να είναι κατάλληλα για χρήση στο μοντέλο. Ακολούθως, εξηγούμε το μοντέλο που θα χρησιμοποιήσουμε στην ανάλυση, τους λόγους που το επιλέξαμε, καθώς και την επιλογή των υπερπαραμέτρων του μοντέλου. Έπειτα, στην ενότητα(4) μιλάμε για το σετ εκπαίδευσης/δοκιμής, τα μέτρα αξιολόγησης που θα χρησιμοποιήσουμε, καθώς και για την απόδοση του μοντέλου. Στην ενότητα (5) παρουσιάζουμε τα συμπεράσματα της μελέτης, καθώς και τους περιορισμούς της. Στη συνέχεια, κάνουμε και μια πρόταση για μελλοντική έρευνα.

3 ΜΕΘΟΔΟΛΟΓΙΑ

3.1 Δεδομένα

Στην παρακάτω υποενότητα θα μιλήσουμε για τα δεδομένα τα οποία έχουμε χρησιμοποιήσει στα επόμενα κεφάλαια, για τα περιγραφικά τους στατιστικά, καθώς και για τη γραφική απεικόνιση των μεταβλητών. Ως εξαρτημένη μεταβλητή έχουμε χρησιμοποιήσει την κατανάλωση ενέργειας του Bitcoin. Η μεταβλητή που χρησιμοποιήθηκε ως εξαρτημένη είναι ο Digiconomist Bitcoin Energy Consumption Index (energy_consumption), που αποτελεί μια εκτίμηση της κατανάλωσης ενέργειας του Bitcoin, και συλλέχθηκε από την ακόλουθη ιστοσελίδα: <https://digiconomist.net/>. Η πλατφόρμα δημιουργήθηκε από τον de Vries το 2014, τον οποίο έχουμε αναφέρει και στη βιβλιογραφική ανασκόπηση, καθώς έχει συνεισφέρει στο να γίνει γνωστό το πρόβλημα της ενεργειακής κατανάλωσης του Bitcoin. Ως εξαρτημένες μεταβλητές χρησιμοποιήσαμε τις average block size (Average_Blocksize), blockchain size (Bitcoin_Blockchain), hash rate (Bitcoin_Hashrate), unique address (Unique_Address), transaction per block (Transaction_Per_Block), unique transaction (Unique_Transaction) οι οποίες σχετίζονται με τη δομή και τη λειτουργία του Bitcoin. Επιπλέον χρησιμοποιήσαμε και τις παρακάτω μεταβλητές: (Cost_Per_Transaction), miner revenue (Miner_Revenue), transaction fee (Transaction_Fee), market price of Bitcoin (Bitcoin_Market_Price), market capitalization of Bitcoin (Market_Capitalization) οι οποίες σχετίζονται με τα οικονομικά και χρηματοοικονομικά χαρακτηριστικά του Bitcoin. Οι παραπάνω μεταβλητές μπορούν να βρεθούν στην ακόλουθη ιστοσελίδα, μαζί με σχετικές πληροφορίες: <https://data.nasdaq.com/institutional-investors>. Για όλες τις μεταβλητές συλλέξαμε ημερήσια δεδομένα από τις 10 Φεβρουαρίου 2017 έως τις 7 Ιουλίου 2020. Επίσης, η μονάδα μέτρησης της εξαρτημένης μεταβλητής είναι τα TWh.

Η επιλογή των παραπάνω μεταβλητών ως εξαρτημένων για την κατανάλωση ενέργειας του Bitcoin έγινε με βάση τη βιβλιογραφία, όπως εξηγείται και παρακάτω πιο αναλυτικά. Οι μεταβλητές Bitcoin_Market_Price, Average_Blocksize, Transaction_Fee χρησιμοποιήθηκαν από τους de Roos (2021) για τη δημιουργία ενός μοντέλου με το οποίο εκτίμησαν την κατανάλωση ενέργειας του Bitcoin. Οι Sapra και Shaikh (2023) χρησιμοποιώντας δύο μοντέλα, ένα OLS και ένα ARDL, βρήκαν ότι μεταβλητές όπως Bitcoin_Hashrate, Unique_Transaction έχουν μια σημαντική θετική σχέση με την κατανάλωση ενέργειας του Bitcoin και συνεισφέρουν σημαντικά στην αυξανόμενη κατανάλωση ενέργειας του Bitcoin. Χρησιμοποιώντας ανάλυση Directional predictability, οι Zhang et al. (2023) βρήκαν ότι η μεταβλητή Bitcoin_Blockchain είναι κατάλληλη για να χρησιμοποιείται στην

πρόβλεψη της κατανάλωσης ενέργειας του Bitcoin. Σύμφωνα με τους Sarkodie et al. (2022), βρήκαν μια μακροχρόνια σχέση μεταξύ της Market_Capitalization και της ενεργειακής κατανάλωσης του Bitcoin. Επίσης, βρέθηκε μια αρνητική σχέση μεταξύ του Miner_Revenue και της ενεργειακής κατανάλωσης του Bitcoin από τους Das και Dutta (2020), χρησιμοποιώντας quantile regression. Τέλος, οι Jana et al. (2022) χρησιμοποιούν όλες τις παραπάνω μεταβλητές που χρησιμοποιούμε και εμείς στην ανάλυση τους για να προβλέψουν την ενεργειακή κατανάλωση του Bitcoin, τις οποίες συνέλεξαν και πρόσθεσαν στο μοντέλο τους ακολουθώντας τη βιβλιογραφία. Παρακάτω, ο Πίνακας 1 περιέχει περιγραφικά στατιστικά για τις μεταβλητές που θα χρησιμοποιήσουμε στην ανάλυση μας, καθώς και διαγράμματα για την γραφική τους απεικόνιση.

Υπολογίσαμε επίσης και τον συντελεστή διακύμανσης(CV), ο οποίος επιτρέπει τη συγκρισιμότητα μεταξύ των μεταβλητών. Μεγαλύτερες τιμές αυτού του δείκτη υποδηλώνουν μεγαλύτερη μεταβλητότητα γύρω από τον μέσο όρο. Παρατηρούμε ότι οι μεγαλύτεροι συντελεστές διακύμανσης εμφανίζονται στις μεταβλητές energy_consumption και Transaction_Fee, και άρα αυτές παρουσιάζουν συγκριτικά τη μεγαλύτερη μεταβλητότητα γύρω από τον μέσο όρο σε σχέση με τις άλλες μεταβλητές. Αυτό μπορούμε να το διαπιστώσουμε και διαγραμματικά, καθώς παρατηρούμε ότι υπάρχει μεγάλη μεταβλητότητα στις χρονοσειρές τους. Παράλληλα, χρησιμοποιήσαμε τον συντελεστή συσχέτισης Pearson για να εξετάσουμε αν υπάρχει γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων. Η μηδενική υπόθεση είναι ότι δεν υπάρχει στατιστικά σημαντική γραμμική σχέση μεταξύ των δύο μεταβλητών. Μπορούμε να παρατηρήσουμε στον Πίνακα 2, όπου υπάρχει και ο συντελεστής διακύμανσης(CV), ότι υπάρχουν μεταβλητές όπως οι Bitcoin_Blockchain, Bitcoin_Hashrate, Miner_Revenue, Bitcoin_Market_Price, Market_Capitalization, οι οποίες έχουν μια στατιστικά σημαντική αρνητική γραμμική συσχέτιση με την εξαρτημένη μεταβλητή energy_consumption, καθώς και μία ανεξάρτητη μεταβλητή(Transaction_Fee), η οποία έχει στατιστικά σημαντική θετική γραμμική συσχέτιση με την εξαρτημένη μεταβλητή. Παρόλα αυτά, η τιμή του στατιστικού μέτρου του συντελεστή συσχέτισης Pearson είναι πολύ μικρή, κοντά στο μηδέν, σε όλες τις περιπτώσεις, υποδεικνύοντας ότι υπάρχει πολύ μικρή γραμμική σχέση, είτε αρνητική είτε θετική, η οποία μπορεί να θεωρηθεί έως και αμελητέα. Αυτό που αναφέραμε χρησιμοποιώντας τον συντελεστή συσχέτισης Pearson μπορούμε να το παρατηρήσουμε και στα Σχήματα 7 έως 12, όπου έχουμε χρησιμοποιήσει γραφήματα διασποράς για να οπτικοποιήσουμε την σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών. Επομένως, και από τα γραφήματα διασποράς καταλήγουμε στο ίδιο συμπέρασμα, ότι δεν υπάρχει γραμμική σχέση της εξαρτημένης μεταβλητής με τις ανεξάρτητες. Στο ίδιο συμπέρασμα καταλήγουν και οι Jana et

al. (2022), οι οποίοι χρησιμοποίησαν τα τεστ Shapiro-Wilk, Frosini για να εξετάσουν την κατανομή των μεταβλητών, τα τεστ ADF, Zivot-Andrews (ZA) για να εξετάσουν την στασιμότητα των μεταβλητών, και τέλος των Hurst&Exponent για να εξετάσουν την αυτοσυσχέτιση των χρονοσειρών. Κατέληξαν στο συμπέρασμα ότι οι μεταβλητές εμφανίζουν μη παραμετρικά μοτίβα, δηλαδή δεν ακολουθούν κάποια γνωστή η συγκεκριμένη κατανομή και τα μοτίβα αλλάζουν με την πάροδο του χρόνου, με τα μοτίβα να μην είναι γραμμικά, χρησιμοποιώντας το ίδιο σύνολο μεταβλητών που χρησιμοποιούμε στην ανάλυσή μας. Επομένως, είναι κατάλληλη η επιλογή ενός μοντέλου μηχανικής μάθησης, καθώς είναι ικανό να μοντελοποιεί περίπλοκες μη γραμμικές σχέσεις που δεν ακολουθούν κάποια γνωστή κατανομή.

Πίνακας 1. Περιγραφικά Στατιστικά.

Μεταβλητές	Μέσος όρος	Διάμεσος	T.A	Ελάχιστο	Μέγιστο
energy_consumption	$1.16e + 16$	$7.31e + 15$	$1.33e + 16$	$5.66e + 09$	$6.40e + 16$
Average_Blocksize	1.00	1.00	0.17	0.43	1.41
Bitcoin_Blockchain	191623.01	188032.22	52547.80	102104.81	286510.98
Cost_Per_Transaction	48.52	46.17	27.96	6.38	161.69
Bitcoin_Hashrate	$4.86e + 07$	43047398.76	$3.69e + 07$	$2.92e + 06$	$1.36e + 08$
Miner_Revenue	13220491.90	12546757.96	7747166.10	1696707.27	53191582.14
Unique_Address	530423.60	523827.0	104344.03	301131.00	1072861.00
Transaction_Fee	119.94	39.74	161.07	9.97	1495.95
Transaction_Per_Block	1906.03	1975.68	409.98	834.13	2762.54
Unique_Transaction	282346.30	287255.0	59593.30	131875.00	490644.00
Bitcoin_Market_Price	6806.39	6929.46	3268.75	927.93	19498.68
Market_Capitalization	$1.19e + 11$	$1.22e + 11$	$5.74e + 10$	$1.52e + 10$	$3.35e + 11$

Σημειώσεις: T.A = τυπική απόκλιση, $xe + n = x * 10^n$

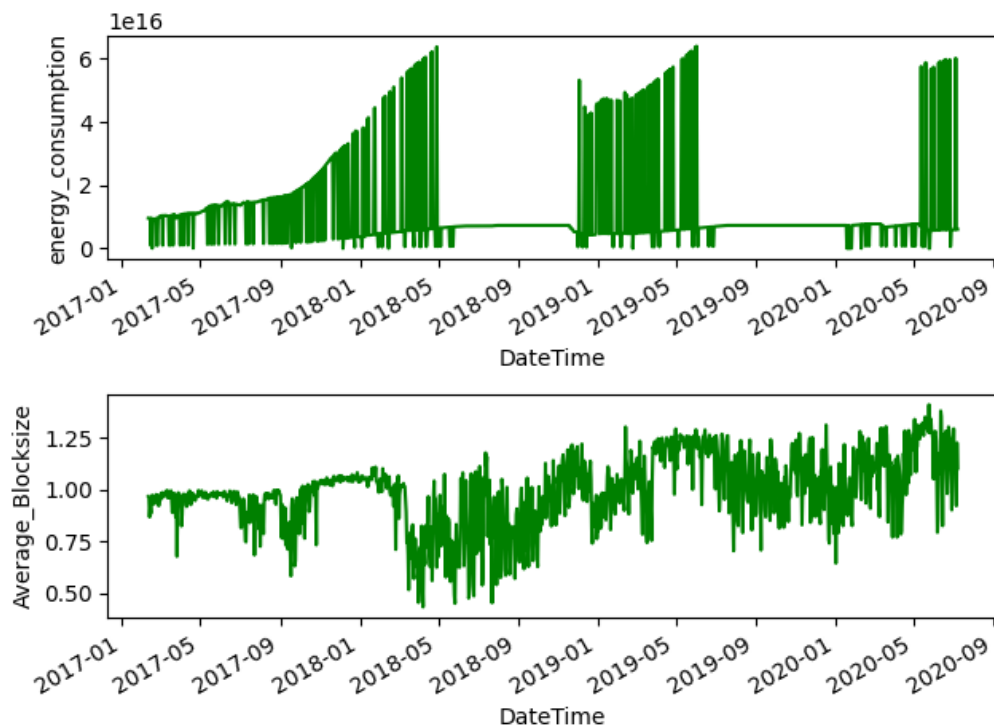
Πίνακας 2. Κάποια επιπλέον Στατιστικά Μέτρα.

Μεταβλητές	CV	PCC
energy_consumption	1.15	
Average_Blocksize	0.17	-0.005
Bitcoin_Blockchain	0.27	-0.080***
Cost_Per_Transaction	0.58	-0.033
Bitcoin_Hashrate	0.76	-0.142***
Miner_Revenue	0.59	-0.051*
Unique_Address	0.2	0.018
Transaction_Fee	1.34	0.064**
Transaction_Per_Block	0.22	0.010
Unique_Transaction	0.21	0.004
Bitcoin_Market_Price	0.48	-0.053*
Market_Capitalization	0.48	-0.062**

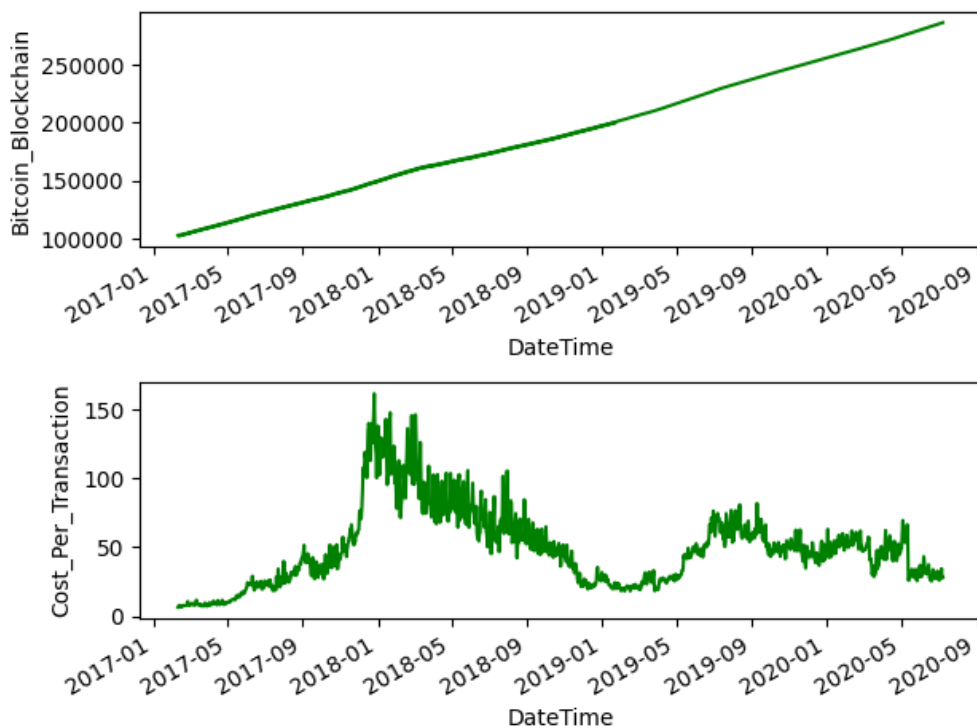
CV = Συντελεστής διακύμανσης

PCC = Συντελεστής συσχέτισης Pearson.

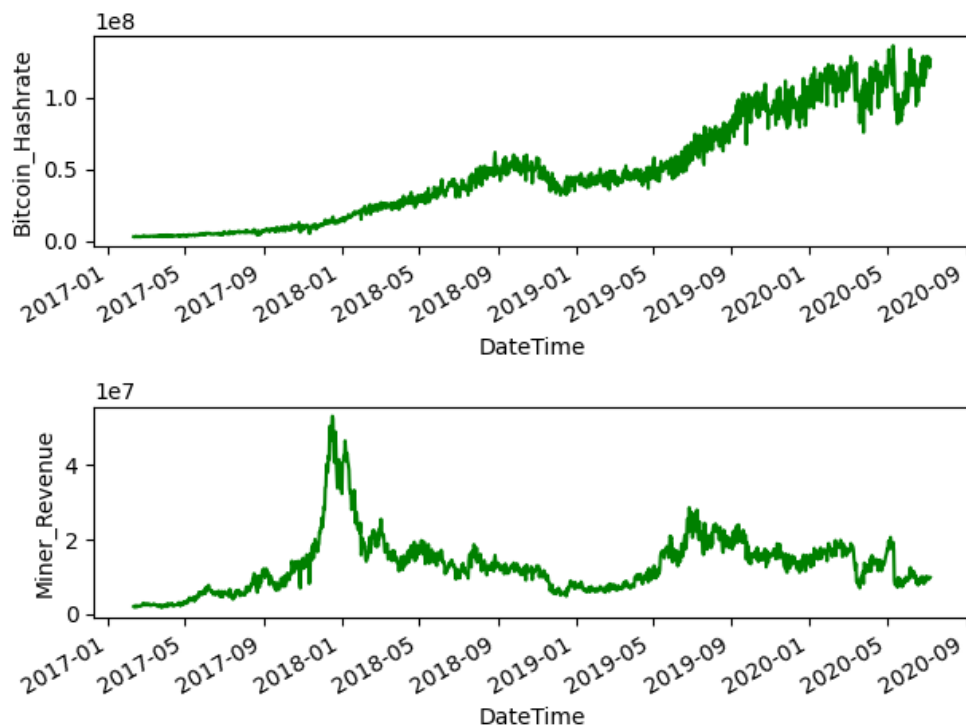
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$



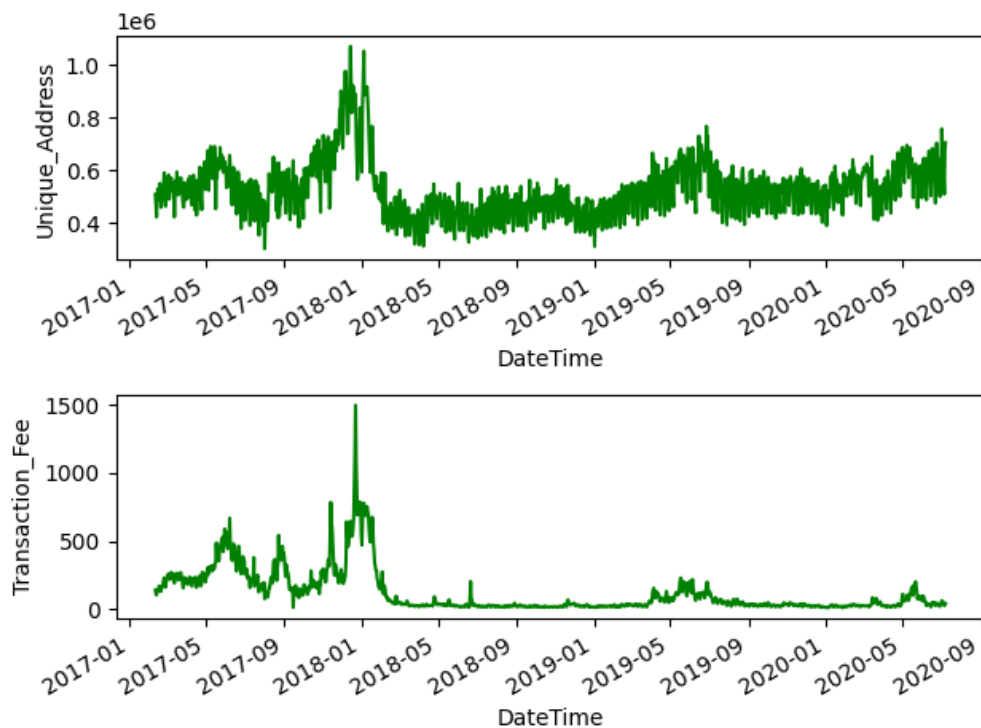
Σχήμα 1. Οι Χρονοσειρές των μεταβλητών energy_consumption, Average_Blocksize.



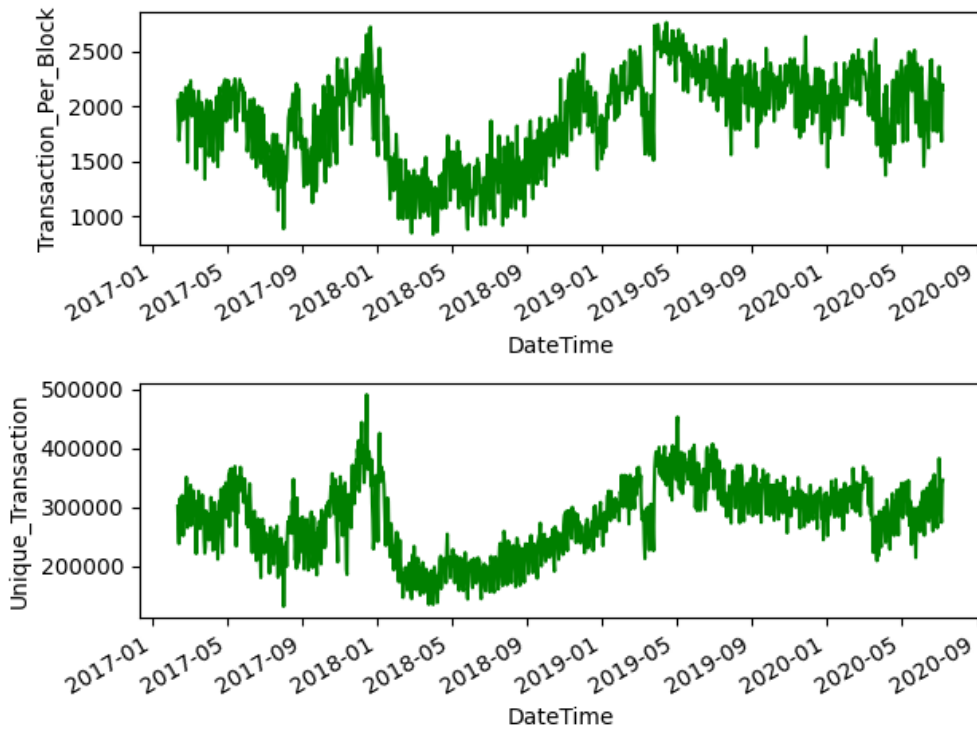
Σχήμα 2. Οι Χρονοσειρές των μεταβλητών Bitcoin_Blockchain, Cost_Per_Transaction.



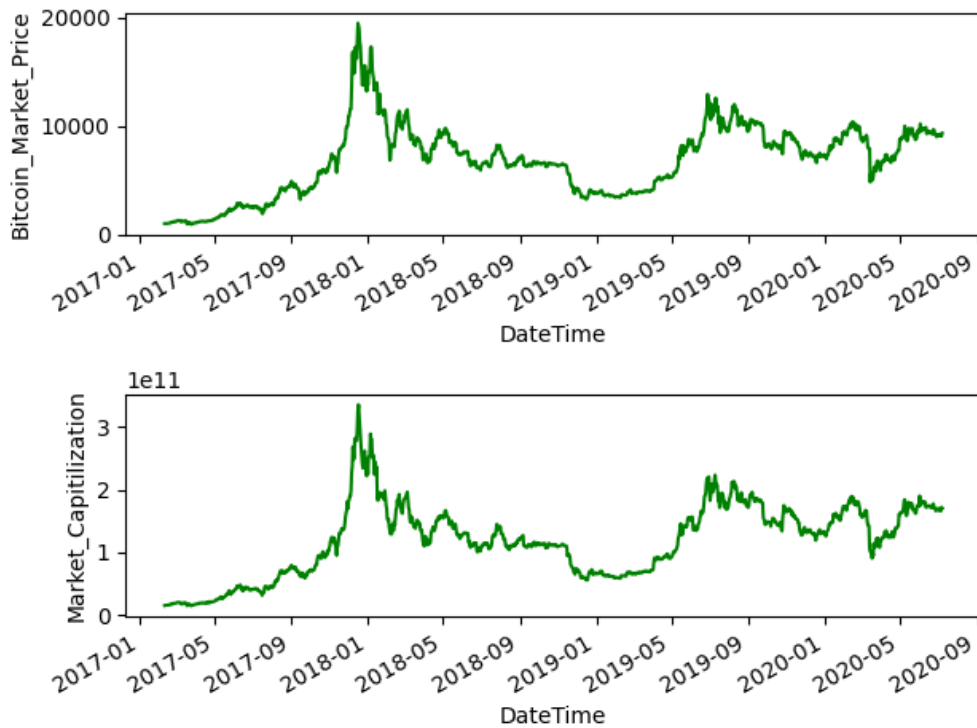
Σχήμα 3. Οι Χρονοσειρές των μεταβλητών Bitcoin_Hashrate, Miner_Revenue.



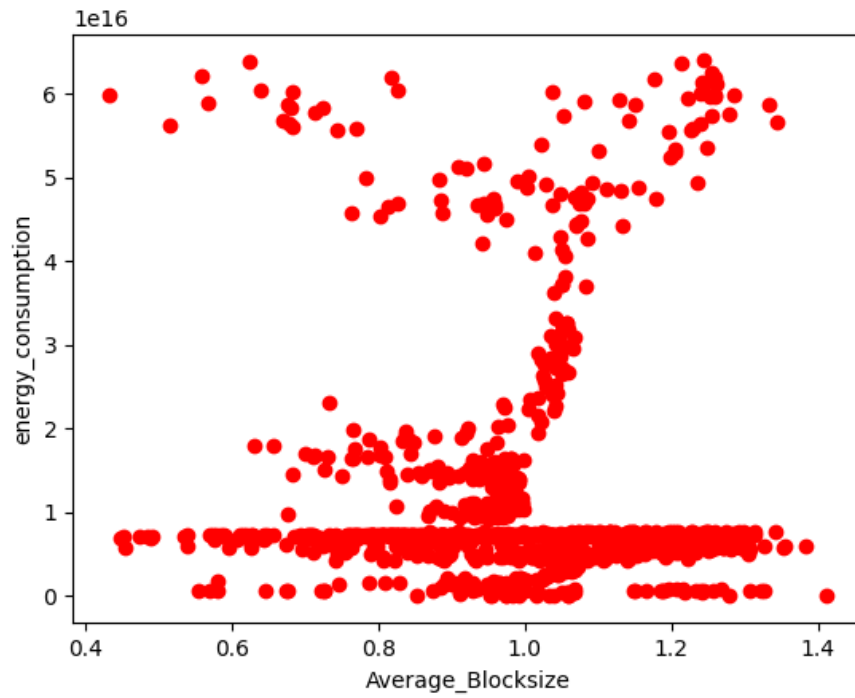
Σχήμα 4. Οι Χρονοσειρές των μεταβλητών Unique_Address, Transaction_Fee.



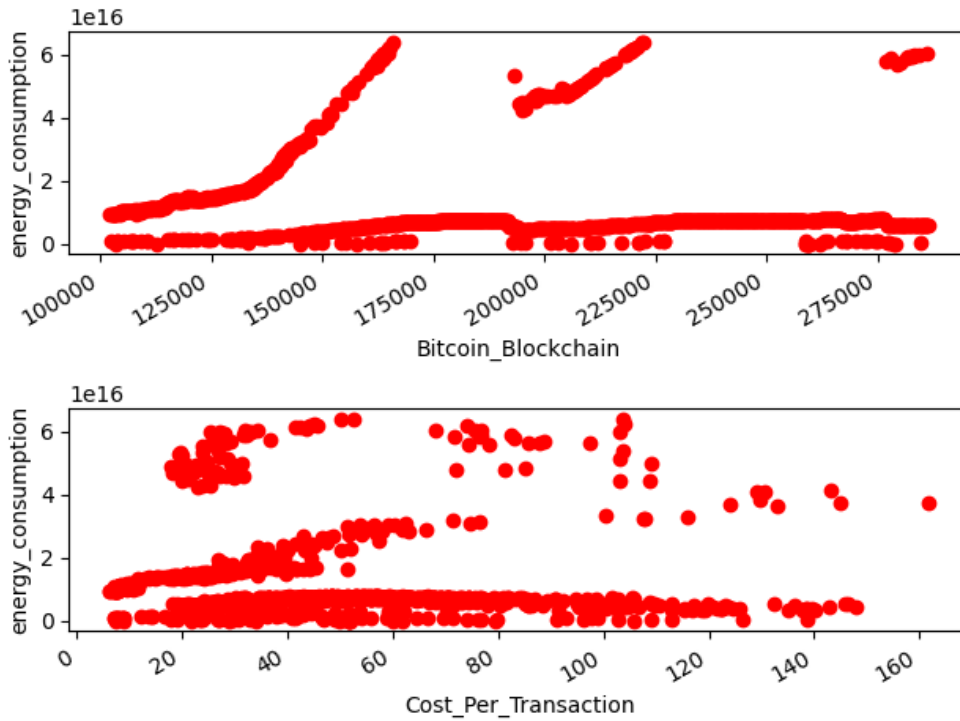
Σχήμα 5. Οι Χρονοσειρές των μεταβλητών Transaction_Per_Block, Unique_Transaction.



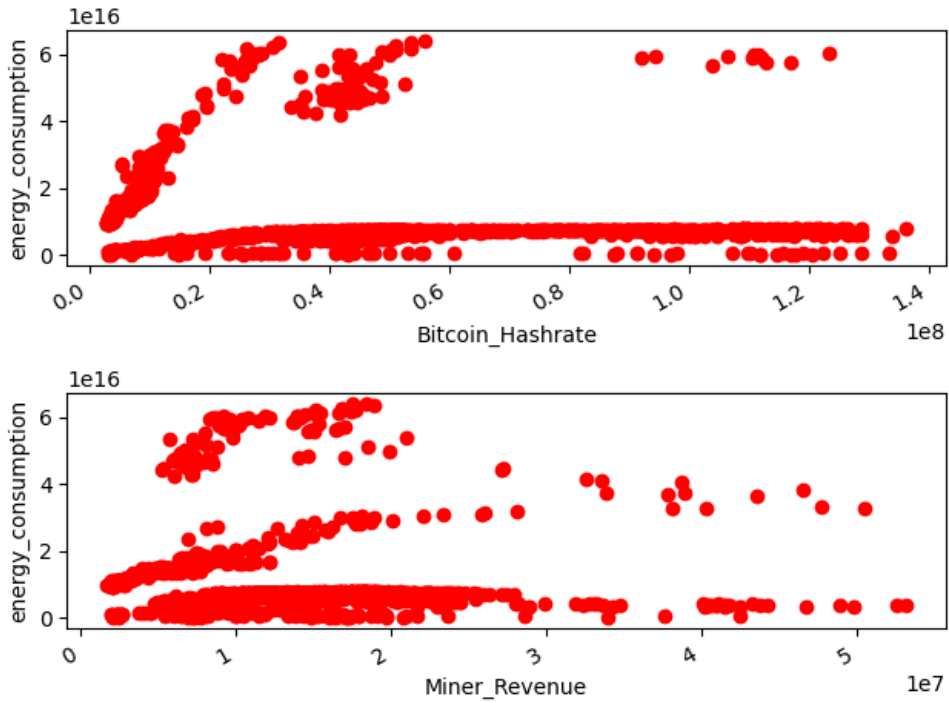
Σχήμα 6. Οι Χρονοσειρές των μεταβλητών Bitcoin_Market_Price, Market_Capitalization.



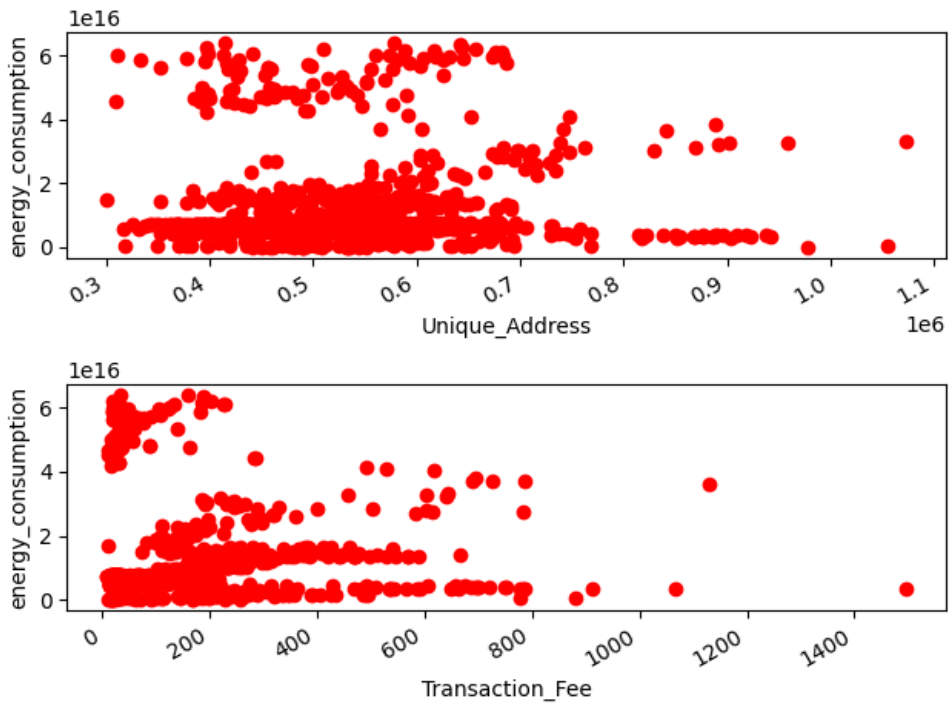
Σχήμα 7. Παρατηρούμαι την σχέση μεταξύ του energy_consumption, Average_Blocksize.



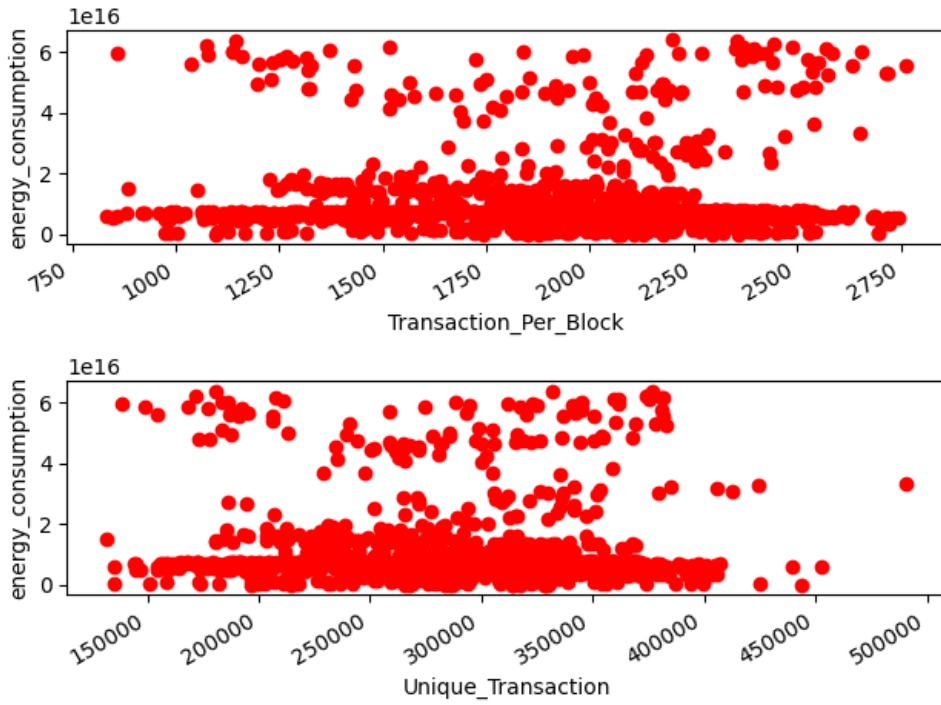
Σχήμα 8. Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Bitcoin_Blockchain, Cost_Per_Transaction.



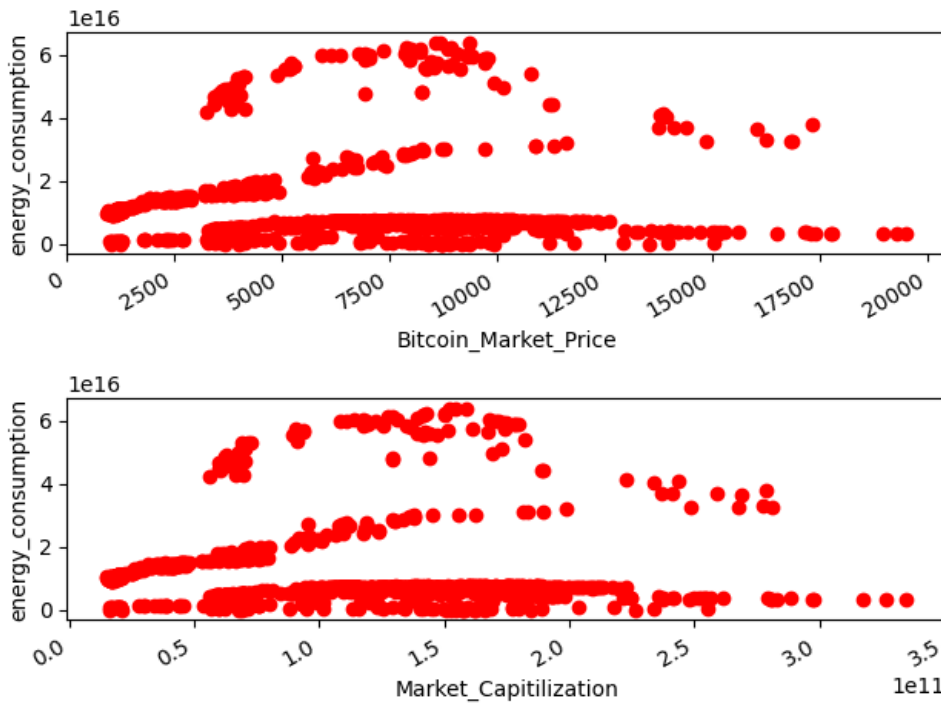
Σχήμα 9. Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Bitcoin_Hashrate, Miner_Revenue.



Σχήμα 10. Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Unique_Address, Transaction_Fee.



Σχήμα 11. Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Transaction_Per_Block, Unique_Transaction.



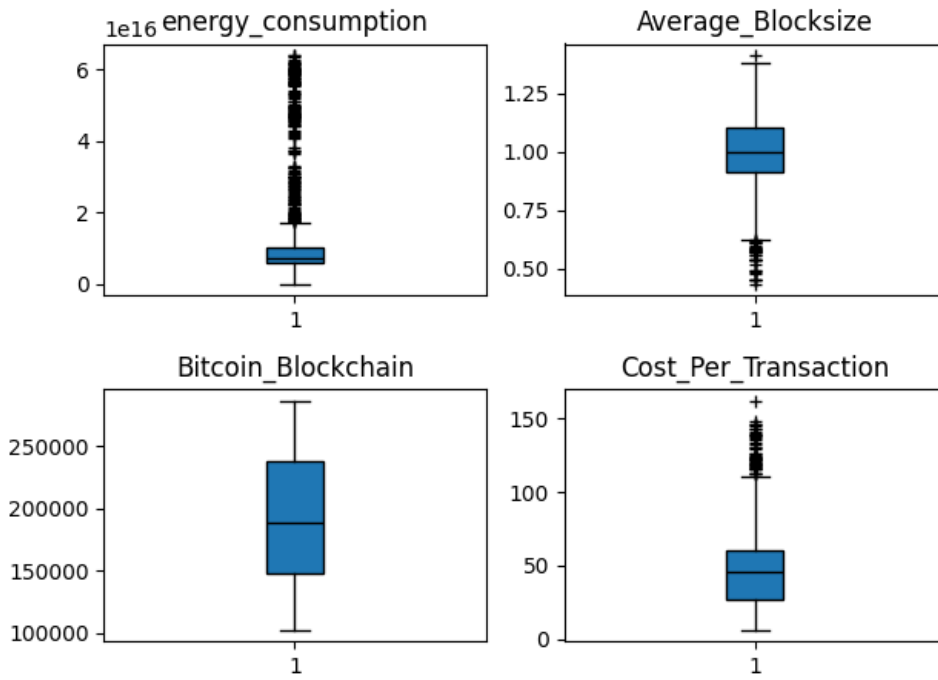
Σχήμα 12. Παρατηρούμαι την σχέση μεταξύ του energy_consumption με τις Bitcoin_Market_Price, Market_Capitalization.

3.2 Προεπεξεργασία των δεδομένων

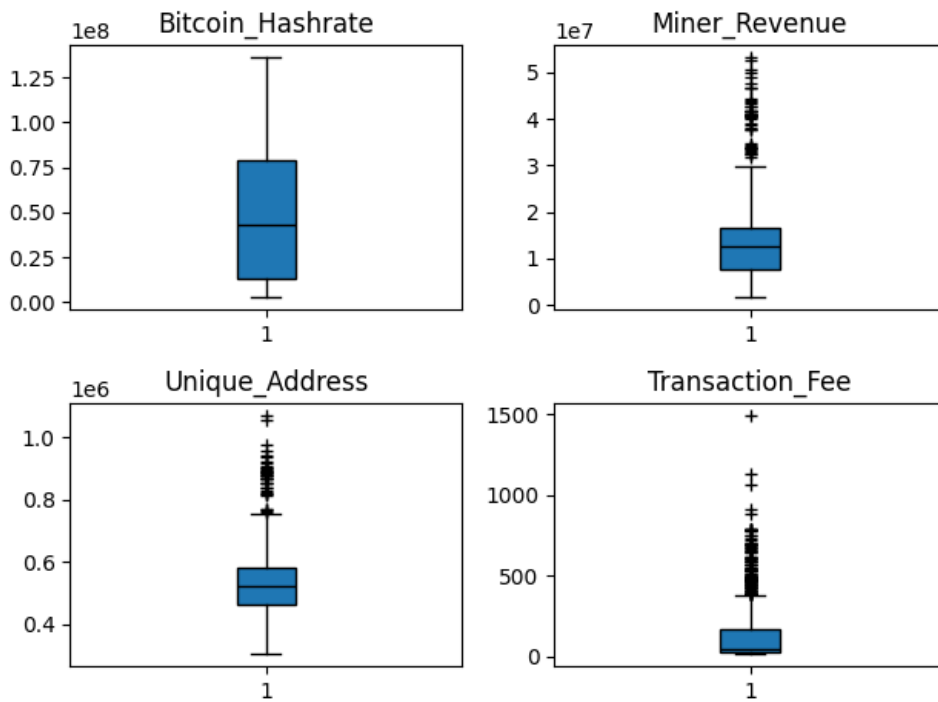
Είναι σημαντικό, προτού προχωρήσουμε στην ανάλυση του μοντέλου που θα χρησιμοποιήσουμε, να γίνει μια προεπεξεργασία των δεδομένων. Διότι χωρίς την κατάλληλη προεπεξεργασία, μπορεί να υπάρξουν σημαντικές αλλοιώσεις στα συμπεράσματα που θα προκύψουν. Για αρχή, ελέγχουμε εάν υπάρχουν ελλείπουσες τιμές, οι οποίες μπορούν να επηρεάσουν τα αποτελέσματα μας, εισάγοντας μεροληψία. Μετά από έλεγχο για ελλείπουσες τιμές, καταλήξαμε στο συμπέρασμα ότι καμία από τις μεταβλητές που χρησιμοποιούμε δεν έχει καμία ελλείπουσα τιμή. Στην συνέχεια, εξετάσαμε εάν υπάρχουν ακραίες τιμές στα δεδομένα μας, δηλαδή τιμές που διαφέρουν πολύ από το υπόλοιπο σύνολο δεδομένων. Χρησιμοποιήσαμε θηκόγραμμα για να οπτικοποιήσουμε το πρόβλημα των ακραίων τιμών, όπως μπορούμε να δούμε στα Σχήματα 13 έως 15. Επιπλέον, είναι ένας τρόπος απεικόνισης των δεδομένων χρησιμοποιώντας το μέγιστο, το ελάχιστο, τη διάμεσο, το πρώτο τεταρτημόριο και το τρίτο τεταρτημόριο. Ως ακραίες τιμές θεωρήσαμε τιμές που βρίσκονται κάτω από το όριο $Q1 - 3 * IQR$ και πάνω από το όριο $Q3 + 3 * IQR$ αντίστοιχα. Οι μεταβλητές $Q1$ και $Q3$ αντιστοιχούν στο πρώτο και τρίτο τεταρτημόριο, αντίστοιχα, ενώ η μεταβλητή IQR είναι η απόσταση μεταξύ του πρώτου από του τρίτου τεταρτημορίου. Με βάση το προαναφερθέν πλαίσιο, εντοπίστηκαν 158 παρατηρήσεις οι οποίες θεωρήθηκαν ακραίες τιμές και αφαιρέθηκαν από το σύνολο δεδομένων, το οποίο αντιστοιχεί στο 12.7% του συνολικού αριθμού παρατηρήσεων. Όσον αφορά τις ακραίες τιμές, σύμφωνα με τους Aguinis et al. (2013), οι οποίοι, μέσω ανάλυσης της βιβλιογραφίας, έχουν καταλήξει σε διάφορες μεθοδολογίες που μπορεί να χρησιμοποιήσει ένας αναλυτής για να διαχειριστεί σωστά τις ακραίες τιμές. Με βάση το παραπάνω άρθρο οι, ακραίες τιμές δημιουργούν πρόβλημα στην προσαρμογή του μοντέλου μηχανικής μάθησης που θα χρησιμοποιήσουμε για πρόβλεψη. Μία από τις προτάσεις που θα χρησιμοποιήσουμε είναι να αναφέρουμε τα αποτελέσματα του μοντέλου με και χωρίς τις ακραίες τιμές, για να δούμε κατά πόσο η αφαίρεση των ακραίων τιμών επηρεάζει την προσαρμογή του μοντέλου και, κατά συνέπεια, τα αποτελέσματα πρόβλεψης.

Οι Tian et al. (2020) χρησιμοποίησαν ένα μοντέλο μηχανικής μάθησης (Gradient Boosting Decision Tree) για να προβλέψουν αν ένα άτομο είναι αξιόπιστο για να λάβει δάνειο ή όχι. Στην ανάλυση τους, είχαν μεταβλητές που είχαν διαφορετικές μονάδες μέτρησης και διαφορετικές κλίμακες. Για να αντιμετωπίσουν αυτό το πρόβλημα και να βελτιώσουν την απόδοση και τη συγκρισιμότητα του μοντέλου μηχανικής μάθησης, χρησιμοποίησαν μια μέθοδο κανονικοποίησης που ονομάζεται κανονικοποίηση ελάχιστου-μέγιστου. Η ίδια μέθοδος θα χρησιμοποιηθεί και από εμάς για τους παραπάνω λόγους. Ποιο αναλυτικά, η συγκεκριμένη μέθοδος

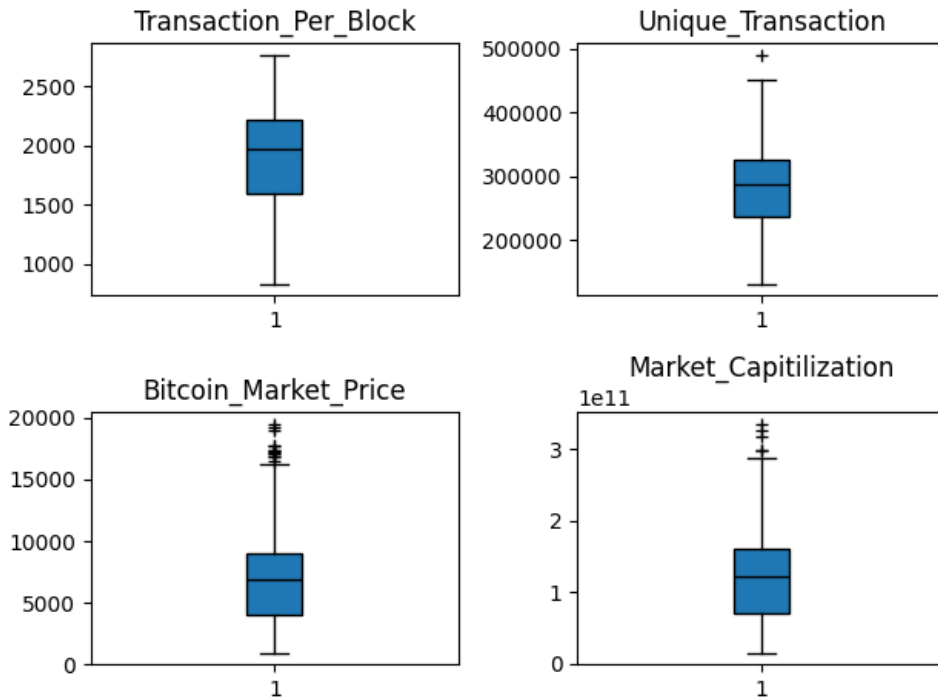
μετασχηματίζει όλα τα δεδομένα σε μια κλίμακα μεταξύ του 0 και 1, ώστε η σημασία ενός χαρακτηριστικού να μην επηρεάζεται από την κλίμακα του και, ως εκ τούτου, να μην επηρεάζει την προσαρμογή του μοντέλου.



Σχήμα 13. Το θηκόγραμμα των μεταβλητών energy_consumption, Average_Blocksize, Bitcoin_Blockchain, Cost_Per_Transaction.



Σχήμα 14. Το θηκόγραμμα των μεταβλητών Bitcoin_Hashrate, Miner_Revenue, Unique_Address, Transaction_Fee.



Σχήμα 15. Το θηκόγραμμα των μεταβλητών Transaction_Per_Block, Unique_Transaction, Bitcoin_Market_Price, Market_Capitilization.

3.3 Extreme Gradient Boosting (XGBoost)

Οι Chen και Guestrin (2016) περιγράφουν ένα σύστημα, το οποίο ονομάζεται XGBoost, το οποίο βασίζεται στη μέθοδο μηχανικής μάθησης gradient tree boosting. Προτείνουν έναν αλγόριθμο συστήματος που βελτιώνει τη διαχείριση δεδομένων, δίνοντας στο μοντέλο την δυνατότητα να διαχειρίζεται μεγάλο πλήθος δεδομένων με πολύ λιγότερους πόρους συγκριτικά με δημοφιλή υπάρχοντα συστήματα. Συγκρίνοντας το σύστημα XGBoost με άλλες υλοποιήσεις που χρησιμοποιούν δέντρα ως αδύναμους μαθητές, όπως οι βιβλιοθήκες scikit-learn, Rsgrbm, σε προβλήματα κατάταξης, τα αποτελέσματα έδειξαν ότι η μέθοδος XGBoost παράγει καλύτερα ή παρόμοια αποτελέσματα συγκριτικά με τα άλλα πακέτα σε πολύ πιο σύντομο χρόνο. Επίσης, πραγματοποιήθηκε αξιολόγηση στη διαδικασία μάθησης της κατάταξης με το καλύτερο σύστημα για αυτή την εργασία (pGBRT). Τα αποτελέσματα έδειξαν ότι είχε παρόμοιες επιδόσεις, με το XGBoost να είναι σημαντικά ταχύτερο. Τέλος, η δημιουργία του συγκεκριμένου συστήματος έδειξε ότι μπορεί να παρέχει υψηλού επιπέδου αποτελέσματα σε μια πλήθωρα εφαρμογών, όπως πρόβλεψη πωλήσεων καταστημάτων, η πρόβλεψη συμπεριφοράς πελατών, η πρόβλεψη ποσοστού κλικ σε διαφημίσεις,

πρόβλεψη κινδύνου. Όσον αφορά, παρακάτω θα δούμε επιπλέον εφαρμογές του συγκεκριμένου αλγορίθμου και αποτελέσματα που αφορούν την επίδοση του. Οι Dey et al. (2016) χρησιμοποίησαν ένα μοντέλο eXtreme Gradient Boosting (XGBoost) για να προβλέψουν την τάση της χρηματιστηριακής αγοράς. Επιτυγχάνοντας ακρίβεια 87% για περιόδους πρόβλεψης 60 και 90 ημερών, το XGBoost αποδείχθηκε ότι έχει την καλύτερη ακρίβεια από όλα τα άλλα μοντέλα που χρησιμοποιούνται στο άρθρο (logistic regression, svm, Random Forest (RF), ANN (artificial neural network)). Στην συνέχεια, οι Touzani et al. (2018) χρησιμοποίησαν ένα μοντέλο XGBoost για να προβλέψουν την κατανάλωση ενέργειας σε εμπορικά κτίρια. Χρησιμοποιώντας 410 εμπορικά κτίρια, το μοντέλο XGBoost πέτυχε τις καλύτερες αποδόσεις συγκριτικά με τα μοντέλα Time-of-Weekend-Temperature (TOWT), RF. Χρησιμοποιώντας ένα μοντέλο XGBoost, οι Carmona et al. (2019) προβλέπουν την πτώχευση τραπεζών στον τομέα των ΗΠΑ για την περίοδο από το 2001 έως το 2015. Το μοντέλο πέτυχε ακρίβεια 94.74%, επιτυγχάνοντας μεγαλύτερη προβλεπτική ικανότητα σε σύγκριση με τα άλλα δύο μοντέλα που χρησιμοποιήθηκαν (random forest, logistic regression).

Εάν μια επιχείρηση έκλεισε λόγω ζημιών, θεωρείται μη επιτυχημένη επιχείρηση, ενώ μια επιχείρηση που είναι ακόμα ανοιχτή και έχει κέρδη θεωρείται επιτυχημένη με βάση αυτή την κατηγοριοποίηση. Οι Kipkoge et al. (2021), χρησιμοποιώντας 205.245 επιχειρήσεις για ένα χρονικό διάστημα από το 1996 έως το 2020, χρησιμοποίησαν διάφορες μεθόδους μηχανικής μάθησης για να προβλέψουν τις επιχειρήσεις που είναι επιτυχημένες. Χρησιμοποιώντας διάφορα μέτρα αξιολόγησης όπως Accuracy, Precision score κατέληξαν στο συμπέρασμα ότι τα μοντέλα με τη μεγαλύτερη ακρίβεια ήταν Gradient Boosting, XGBoost, σε σύγκριση με τα υπόλοιπα μοντέλα που ήταν παρόμοια με τις προηγούμενες αναφορές. Ένα ακόμη άρθρο χρησιμοποίησε μοντέλο XGBoost με πολύ καλά αποτελέσματα. Σύμφωνα με τους Prasad et al. (2022), χρησιμοποίησαν διάφορες μεθόδους μηχανικής μάθησης για να προβλέψουν την τάση του δείκτη μεταβλητότητας του S&P 500 (VIX) σε ημερήσια και εβδομαδιαία συχνότητα. Τα μοντέλα μηχανικής μάθησης βασισμένα σε gradient boosting αποδείχθηκαν ανώτερα από τη λογιστική παλινδρόμηση. Με βάση στοιχεία από κινεζικές εταιρείες για την περίοδο από το 1998 έως το 2014, οι Liu et al. (2023) χρησιμοποίησαν διάφορα μοντέλα μηχανικής μάθησης για να προβλέψουν ποιες από τις επιχειρήσεις βρίσκονται σε κατάσταση οικονομικής δυσκολίας. Σύμφωνα με τα αποτελέσματά τους, τα μοντέλα gradient boosting που βασίζονται σε δέντρα είχαν καλύτερες επιδόσεις από τα υπόλοιπα μοντέλα πρόβλεψης, όπως τα Bag.DT, Ada.DT, LR. Με βάση τα παραπάνω, θα χρησιμοποιήσουμε το μοντέλο μηχανικής μάθησης XGBoost, διότι μπορεί να διαχειρίζεται μεγάλα σύνολα δεδομένων χρησιμοποιώντας λιγότερους πόρους, καθιστώντας το πιο γρήγορο στην εκτέλεση. Επιπλέον, όπως μπορούμε

να συμπεράνουμε από τις αναφορές, επιτυγχάνει πολύ καλά αποτελέσματα όσον αφορά την προβλεπτική του ικανότητα και έχει μεγάλη επιτυχία σε μια πληθώρα εφαρμογών. Στην συνέχεια, θα διεξάγουμε μια πιο αναλυτική ανάλυση σχετικά με το μοντέλο μηχανικής μάθησης που θα χρησιμοποιήσουμε στην ανάλυση.

3.3.1 Ανάλυση Μεθοδολογίας

Η περιγραφή του μοντέλου που θα χρησιμοποιήσουμε βασίζεται στους Chen και Guestrin (2016). Για αρχή, ας υποθέσουμε ότι έχουμε ένα σύνολο χαρακτηριστικών x_i . Με βάση τα χαρακτηριστικά, το μοντέλο προσπαθεί να προβλέψει το αποτέλεσμα, το οποίο θα το ονομάσουμε \hat{y}_i .

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i) \quad (1)$$

Το K αντιστοιχεί στο άθροισμα των συναρτήσεων που χρησιμοποιούνται για να προβλέψουν τη μεταβλητή \hat{y}_i . Η συνάρτηση $f_k(x_i)$ αντιστοιχεί σε ένα ανεξάρτητο δέντρο, το οποίο αποτελείται από φύλλα. Κάθε δέντρο χρησιμοποιεί τα χαρακτηριστικά για να διαχωρίσει τα δεδομένα και θα προτιμήσει το χαρακτηριστικό που χωρίζει τα δεδομένα πιο ξεκάθαρα στην προσπάθειά του να προβλέψει την τιμή του \hat{y}_i . Η τελική πρόβλεψη για το κάθε \hat{y}_i θα προέρχεται από το άθροισμα των προβλέψεων κάθε ανεξάρτητου δέντρου. Για να εκπαιδεύσουμε το μοντέλο μας ώστε να επιτύχει τα καλύτερα δυνατά αποτελέσματα, χρησιμοποιούμε μια κανονικοποιημένη συνάρτηση κόστους, την οποία το μοντέλο πρέπει να ελαχιστοποιήσει.

$$L(\phi) = \sum_i I(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

Το I είναι μια συνάρτηση η οποία είναι διαφορίσιμη και κυρτή και χρησιμοποιείται για να μετρήσει τη διαφορά μεταξύ της τιμής πρόβλεψης (\hat{y}_i) και της πραγματικής τιμής (y_i) της μεταβλητής στόχου. Ο δεύτερος όρος (Ω) χρησιμοποιείται ως ποινή για την πολυπλοκότητα του μοντέλου, ώστε να αποτρέψει την υπερπροσαρμογή, η οποία προκαλείται όταν το μοντέλο προσαρμόζεται πάρα πολύ καλά στα δεδομένα εκπαίδευσης, με αποτέλεσμα να μην έχει καλή απόδοση σε νέα, άγνωστα δεδομένα. Με αυτόν τον τρόπο, το μοντέλο θα τείνει να επιλέγει πιο απλές δομές, αποφεύγοντας τέτοιου είδους προβλήματα. Αξίζει να σημειωθεί ότι όταν η παράμετρος κανονικοποίησης ισούται με τον μηδέν, η συνάρτηση κόστους μετρείται στη συνάρτηση κόστους που χρησιμοποιείται στο μοντέλο gradient tree boosting. Η βάση του μοντέλου είναι ο αλγόριθμος gradient tree boosting, ο οποίος κάθε φορά κατασκευάζει ένα δέντρο, είτε ταξινόμησης είτε

παλινδρόμησης, ανάλογα με τον τύπο της εξαρτημένης μεταβλητής, π.χ. αν είναι συνεχής η κατηγορική. Κάθε δέντρο που κατασκευάζεται ονομάζεται αδύναμος μαθητής (weak learner), ο οποίος προσπαθεί να διορθώσει τα σφάλματα που προκλήθηκαν από όλα τα προηγούμενα δέντρα, με στόχο να ελαχιστοποιήσει την συνάρτηση κόστους. Άρα, κάθε αδύναμος μαθητής στην ουσία χρησιμοποιεί το σφάλμα πρόβλεψης του προηγούμενου μοντέλου, δηλαδή τη διαφορά μεταξύ της εκτιμήσεως και τις πραγματικής τιμής της μεταβλητής, για να εκπαιδευτεί με σκοπό να βελτιώσει την πρόβλεψη και να μειώσει την συνάρτηση κόστους, προσεγγίζοντας έτσι την πραγματική τιμή.

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3)$$

Το $\hat{y}_i^{(t)}$ είναι το δέντρο t , το $\hat{y}_i^{(t-1)}$ είναι το προηγούμενο δέντρο πριν το t , και το $f_t(x_i)$ είναι το δέντρο που δημιουργήθηκε τώρα, με στόχο, κατά τη διάρκεια της εκπαίδευσής του, να προσαρμοστεί στα υπολείμματα του προηγούμενου μοντέλου. Τέλος, το t είναι το σύνολο των δέντρων. Όσον αφορά την περιγραφή της βάσης του μοντέλου XGBoost, δηλαδή του gradient tree boosting χρησιμοποιήσαμε το άρθρο των Zhang et al. (2023). Χρησιμοποιείται μια προσέγγιση δεύτερης τάξης, η οποία σχετίζεται με την ακρίβεια στην προσέγγιση μιας συνάρτησης, για να γίνει γρήγορη βελτιστοποίηση της αντικειμενικής συνάρτησης σε γενικό πλαίσιο εφαρμογής.

$$\tilde{L}^{(t)} = \sum_{i=1}^n [g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t) \quad (4)$$

Όπου ο όρος g_i είναι η πρώτη παράγωγος και η h_i είναι η δεύτερη παράγωγος του πρώτου μέλους της εξίσωσης (2). Η πρώτη παράγωγος υποδεικνύει προς ποια κατεύθυνση πρέπει να αλλάξουμε τις τιμές των παραμέτρων ώστε να μειωθεί η συνάρτηση κόστους. Η δεύτερη παράγωγος μας βοηθάει ως προς την αλλαγή του ρυθμού μεταβολής της συνάρτησης κόστους. Εκτός από την κανονικοποιημένη συνάρτηση κόστους, χρησιμοποιούνται και κάποιες επιπλέον τεχνικές για να αποφευχθεί η υπερπροσαρμογή. Μία από αυτές ονομάζεται Column Subsampling, που σημαίνει ότι για κάθε δέντρο που εκπαιδεύεται με βάση τα σφάλματα του προηγούμενου, θα επιλέγεται τυχαία ένα υποσύνολο από τα συνολικά χαρακτηριστικά για να εκπαιδευτεί το μοντέλο. Τέλος, ο αλγόριθμος XGBoost χρησιμοποιεί έναν πιο αποτελεσματικό αλγόριθμο εύρεσης διάσπασης συγκριτικά με τις υπόλοιπες υλοποιήσεις, οι οποίες δοκιμάζουν όλους τους δυνατούς συνδυασμούς χαρακτηριστικών ώστε να επιλέξουν αυτήν που διαχωρίζει τα δεδομένα

πιο καθαρά, κάνοντας τους υπολογιστικά απαιτητικούς.

3.3.2 Υπερπαράμετροι

Οι υπερπαράμετροι είναι βάρη ή τιμές με βάση τις οποίες διαμορφώνονται η δομή του μοντέλου και η διαδικασία μάθησης. Επομένως, οι συγκεκριμένες παράμετροι ρυθμίζονται πριν από τη διαδικασία εκπαίδευσης του μοντέλου. Παράδειγμα τέτοιων παραμέτρων είναι ο ρυθμός μάθησης, ο οποίος ελέγχει το βήμα προς την ελαχιστοποίηση της συνάρτησης κόστους μετά από κάθε επανάληψη και, άρα ελέγχει το πόσο γρήγορα μαθαίνει ένα μοντέλο. Μια άλλη παράμετρος είναι το μέγιστο βάθος, το οποίο ρυθμίζει το μέγιστο βάθος κόμβων που μπορεί να έχει ένα δέντρο, δηλαδή την απόσταση μεταξύ του κόμβου ρίζας και του πιο απομακρυσμένου κόμβου. Αυτή η συγκεκριμένη παράμετρος χρησιμοποιείται για να αποφύγουμε την υπερπροσαρμογή, καθώς πολύ περίπλοκα δέντρα θα προσαρμοστούν υπερβολικά καλά στο σύνολο εκπαίδευσης και, ως αποτέλεσμα, δεν θα αποδίδουν σε νέα άγνωστα δεδομένα. Τέλος, οι παράμετροι α , λ είναι όροι που προστίθενται στη συνάρτηση κόστους που προσπαθούμε να ελαχιστοποιήσουμε, με σκοπό την αντιμετώπιση της υπερπροσαρμογής, βοηθώντας έτσι στην δημιουργία πιο απλών μοντέλων και στην επιλογή χαρακτηριστικών.

Για να ρυθμίσουμε τις υπερπαραμέτρους και να επιτύχουμε ένα μοντέλο με καλή προβλεπτική ικανότητα, θα χρησιμοποιήσουμε ένα μοντέλο που, μέσω του υποσυνόλου εκπαίδευσης, θα εξασφαλίσει ότι δεν υπάρχει αλληλεπίδραση στη απόδοση του μοντέλου σε άγνωστα δεδομένα. Στη συνέχεια με τη χρήση μετρικών αξιολόγησης, θα αξιολογηθεί η απόδοση του μοντέλου. Θα εντοπίσουν οι υπερπαράμετροι που επιτυγχάνουν την καλύτερη απόδοση, οι οποίες θα χρησιμοποιηθούν για την εκτίμηση των παραμέτρων του μοντέλου.

$$\theta^* = \underset{\theta}{\operatorname{argmin}} f(x|\theta) \quad (5)$$

Όπου j είναι το διάνυσμα των υπερπαραμέτρων, το x είναι το υποσύνολο εκπαίδευσης του δείγματος και η f είναι η συνάρτηση σφάλματος, η οποία στην περίπτωση μας είναι το root mean square error.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

Το y_i είναι η πραγματική τιμή και το \hat{y}_i είναι η εκτιμημένη τιμή της εξαρτημένης μεταβλητής, αλλιώς μεταβλητής στόχου. Όσο μικρότερη είναι η τιμή του μέτρου αξιολόγησης, τόσο καλύτερη είναι η προβλεπτική ικανότητα του μοντέλου.

Οι υπερπαράμετροι ρυθμός μάθησης, μέγιστο βάθος και ο αριθμός εκτιμητών(n),

ρυθμίστηκαν χρησιμοποιώντας Μπεϋζιανή βελτιστοποίηση. Η υπερπαραμέτρος n καθορίζει τον αριθμό των gradient boosted trees. Στην Μπεϋζιανή βελτιστοποίηση, η συνάρτηση κόστους που θέλουμε να ελαχιστοποιήσουμε θεωρείται τυχαία μεταβλητή και της αποδίδεται μια κατανομή πιθανοτήτων. Ο αλγόριθμος αναζητά μέσα σε αυτή την κατανομή μια παρατήρηση, η οποία είναι ένας συνδυασμός από υπερπαραμέτρους, και την εφαρμόζει στο σύνολο εκπαίδευσης για να αξιολογηθεί η απόδοση. Καθώς αυτή η διαδικασία επαναλαμβάνεται, η κατανομή βελτιώνεται και ο αλγόριθμος μαθαίνει καλύτερα σε ποιες περιοχές να αναζητά και ποιες όχι. Επομένως, ο αλγόριθμος προσπαθεί να διαμορφώσει μια κατανομή για την συνάρτηση που επιδιώκουμε να βελτιστοποιήσουμε, λαμβάνοντας υπόψη και τις προηγούμενες πληροφορίες. Η διαδικασία θα σταματήσει όταν η συνάρτηση προς ελαχιστοποίηση σταθεροποιηθεί. Για να έχουμε καλύτερη προσέγγιση όσον αφορά την επιλογή των βέλτιστων υπερπαραμέτρων, χρησιμοποιήσαμε διασταυρούμενη επικύρωση, η οποία χωρίζει το σύνολο εκπαίδευσης κατά τη διαδικασία εύρεσης υπερπαραμέτρων σε k μη επικαλυπτόμενα υποσύνολα. Όπου τα $k - 1$ χρησιμοποιούνται ως σετ εκπαίδευσης και το υπολειπόμενο ως σετ δοκιμής. Η εκτέλεση σταματά όταν όλα τα υποσύνολα έχουν χρησιμοποιηθεί ως σετ δοκιμής. Με αυτόν τον τρόπο, έχουμε μια πολύ καλύτερη εκτίμηση της συνάρτησης προς ελαχιστοποίηση και, ως αποτέλεσμα, μια πιο αξιόπιστη επιλογή υπερπαραμέτρων ($k = 10$). Υπάρχουν και άλλοι τρόποι να βρεις τις βέλτιστες υπερπαραμέτρους, όπως το grid search, το οποίο δοκιμάζει όλους τους δυνατούς συνδυασμούς υπερπαραμέτρων σε έναν δεδομένο χώρο. Επίσης, μπορούμε να θεωρήσουμε το διάνυσμα των υπερπαραμέτρων ως μια τυχαία μεταβλητή και να αναζητούμε τυχαία κάθε υπερπαραμέτρο σε μια προκαθορισμένη κατανομή. Οι Di Persio και Fraccarolo (2023) χρησιμοποιούν ένα μοντέλο Gradient Boosting Decision Tree για να προβλέψουν τη ζήτηση ενέργειας για μια εταιρεία ηλεκτρικής ενέργειας σε 7 ζώνες της ιταλικής αγοράς ενέργειας. Για να βελτιστοποιήσουν τις υπερπαραμέτρους χρησιμοποίησαν Μπεϋζιανή βελτιστοποίηση, καθώς οι άλλες δύο προαναφερθείσες είτε είναι πολύ χρονοβόρες είτε δεν λαμβάνουν υπόψη τα προηγούμενα αποτελέσματα στην επόμενη παρατήρηση. Για αυτούς τους λόγους, επιλέγουμε και εμείς αυτή τη μεθοδολογία για να βρούμε τον βέλτιστο συνδυασμό υπερπαραμέτρων για το μοντέλο μας μέσα σε ένα λογικό χρονικό πλαίσιο. Η παραπάνω περιγραφή της Μπεϋζιανή βελτιστοποίησης έγινε βασισμένη στο άρθρο που αναφέραμε.

4 ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

4.1 Σετ εκπαίδευσης-δοκιμής

Προτού αναλύσουμε τα αποτελέσματα και προχωρήσουμε στην πρόβλεψη, είναι σημαντικό το σύνολο δεδομένων να χωριστεί σε σετ εκπαίδευσης και σετ δοκιμής. Το σετ εκπαίδευσης είναι ένα υποσύνολο του συνόλου δεδομένων, το οποίο χρησιμοποιείται για να εκτιμήσουμε της παραμέτρους του μοντέλου, δηλαδή για να μάθει τις σχέσεις μεταξύ των μεταβλητών, ώστε μετά την εκπαίδευση να έχει καλή προβλεπτική ικανότητα. Μπορείς επίσης να χρησιμοποιήσεις το σετ εκπαίδευσης για πρόβλεψη, ώστε να έχεις μια εκτίμηση του πραγματικού σφάλματος πρόβλεψης. Το σετ δοκιμής είναι ένα υποσύνολο του συνόλου δεδομένων, το οποίο δεν έχει έρθει σε επαφή με το μοντέλο μας, και θα χρησιμοποιήσουμε το μοντέλο για να κάνουμε πρόβλεψη σε αυτό το σύνολο δεδομένων. Έτσι, το μοντέλο μας δοκιμάζεται σε δεδομένα που δεν έχει ξαναδεί, προκειμένου να προσεγγίσουμε την προβλεπτική του ικανότητα σε πραγματικές καταστάσεις και να πάρουμε μια καλή εκτίμηση για το σφάλμα πρόβλεψης, αλλιώς γνωστό ως σφάλμα γενίκευσης. Τα δεδομένα εκπαίδευσης και δοκιμής επιλέγονται τυχαία από το σύνολο των δεδομένων, και πριν από τον διαχωρισμό γίνεται ένα τυχαίο ανακάτεμα των δεδομένων. Με αυτόν τον τρόπο, τείνουμε να προσεγγίσουμε καλύτερα ότι κάθε σύνολο δεδομένων θα περιλαμβάνει παρατηρήσεις που αντικατοπτρίζουν όλες τις πιθανές τιμές που μπορεί να πάρει το σύνολο, παρέχοντας στο μοντέλο μεγαλύτερη ποικιλία στις καταστάσεις στις οποίες μπορεί να βρεθεί μια μεταβλητή κατά την εκπαίδευση του. Με αυτόν τον τρόπο, το μοντέλο επιτυγχάνει μια πιο αμερόληπτη εκπαίδευση και, ως αποτέλεσμα οι σχέσεις που δημιουργεί το μοντέλο μεταξύ των μεταβλητών τείνουν να είναι πιο ανθεκτικές και ακριβείς, βελτιώνοντας παράλληλα την απόδοση του σε νέα, αθέατα δεδομένα. Το μοντέλο θα δοκιμαστεί σε διαφορετικές αναλογίες σετ εκπαίδευσης και δοκιμής, ώστε να έχουμε μια καλύτερη εκτίμηση της απόδοσης του και να αξιολογηθεί η προβλεπτική ικανότητα του μοντέλου υπό πολλές συνθήκες. Θα χωρίσουμε το σετ εκπαίδευσης και δοκιμής σε 3 αναλογίες: (80%,20%), (70% 30%) και (60% 40%) αντίστοιχα. Το (60% 40%) αντικατοπτρίζει συνθήκες όπου δεν υπάρχει μεγάλο δείγμα για να εκπαιδευτεί το μοντέλο και μεγάλο δείγμα αγνώστων δεδομένων στα οποία θα πρέπει παράλληλα να αποδώσει. Το (80%,20%) ανταναχλά συνθήκες όπου υπάρχει ένα αρκετά μεγάλο σύνολο δεδομένων για την εκπαίδευση του μοντέλου. Στην συνέχεια, θα αναλύσουμε τα μέτρα αξιολόγησης που έχουμε χρησιμοποιήσει στην ανάλυση για να αξιολογήσουμε την προβλεπτική ικανότητα του μοντέλου(XGBoost).

4.2 Μετρικές αξιολόγησης

Σε ένα τέτοιο πλαίσιο ανάλυσης, μας ενδιαφέρει το μοντέλο μας να προβλέπει με υψηλή ακρίβεια την εξαρτημένη μεταβλητή σε δεδομένα που δεν έχει ξαναδεί, δηλαδή δεδομένα εκτός του συνόλου εκπαίδευσης. Για να μπορέσουμε να αξιολογήσουμε την απόδοση του μοντέλου, η βασική ιδέα είναι να συγκρίνουμε τις τιμές της εξαρτημένης μεταβλητής που προέκυψαν από την πρόβλεψη του μοντέλου στο σετ δοκιμής και να τις συγκρίνουμε με τις πραγματικές τιμές της εξαρτημένης μεταβλητής στο σετ δοκιμής. Η διαφορά μεταξύ της προβλεπόμενης τιμής από το σετ δοκιμής και της πραγματικής τιμής από το σετ δοκιμής της εξαρτημένης μεταβλητής ονομάζεται σφάλμα πρόβλεψης. Υπάρχουν πολλές παραλλαγές, καθεμία από τις οποίες δίνει έμφαση σε διαφορετική πτυχή του σφάλματος πρόβλεψης. Παρακάτω θα γίνει αναφορά στα μέτρα που χρησιμοποιήσαμε στην ανάλυση. Το πρώτο μέτρο αξιολόγησης με το οποίο θα ξεκινήσουμε την αναφορά ονομάζεται Nash-Sutcliffe Efficiency (NSE).

$$NSE = 1 - \frac{\sum_{t=1}^N (\hat{y}_t - y_t)^2}{\sum_{t=1}^N (y_t - \bar{y})^2} \quad (7)$$

Οπου y_t είναι οι πραγματικές τιμές της εξαρτημένης μεταβλητής, \hat{y}_t είναι οι προβλεπόμενες τιμές της εξαρτημένης μεταβλητής και, τέλος \bar{y} είναι ο μέσος όρος της πραγματικής εξαρτημένης μεταβλητής. Το συγκεκριμένο μέτρο κυμαίνεται από $-\infty$ έως 1. Όταν το $NSE < 0$, σημαίνει ότι ο μέσος όρος της πραγματικής εξαρτημένης μεταβλητής έχει καλύτερη προβλεπτική ικανότητα από το μοντέλο. Όταν $NSE = 0$, ο μέσος όρος της πραγματικής εξαρτημένης μεταβλητής έχει την ίδια προβλεπτική ικανότητα με το μοντέλο. Τέλος, όταν $NSE > 0$, σημαίνει ότι το μοντέλο έχει καλύτερη προβλεπτική ικανότητα από τον μέσο όρο της πραγματικής εξαρτημένης μεταβλητής. Μάλιστα, όταν $NSE = 1$ τότε το μοντέλο έχει τέλεια προβλεπτική ικανότητα. Στην συνέχεια, θα παρουσιάσουμε το μέτρο αξιολόγησης Root mean square error (RMSE).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

Το n είναι ο αριθμός των παρατηρήσεων, το y_i είναι οι πραγματικές τιμές της εξαρτημένης μεταβλητής και, τέλος, το \hat{y}_i είναι οι προβλεπόμενες τιμές της εξαρτημένης μεταβλητής. Υπολογίζει τη μέση διαφορά που έχουν οι πραγματικές από τις προβλεπόμενες τιμές της εξαρτημένης μεταβλητής. Επίσης, αυτές οι διαφορές υψώνονται στο τετράγωνο ώστε κατά τη διάρκεια του αθροίσματος να μην απαιφθούν τιμές με αντίθετο πρόσημο. Επίσης, επειδή οι διαφορές υψώνονται στο

τετράγωνο, τιμωρεί περισσότερο τα μεγάλα σφάλματα, δηλαδή τις περιπτώσεις όπου η προβλεπόμενη τιμή έχει πολύ μεγάλη διαφορά από την πραγματική τιμή. Η τιμή του RMSE είναι μεγαλύτερη ή ίση με το μηδέν. Όσο χαμηλότερη είναι η τιμή του RMSE, τόσο καλύτερα έχει προσαρμοστεί το μοντέλο στα δεδομένα, δηλαδή όσο χαμηλότερη είναι η τιμή του μέτρου αξιολόγησης RMSE, τόσο καλύτερη προβλεπτική ικανότητα έχει το μοντέλο μας. Όταν $RMSE = 0$ σημαίνει ότι το μοντέλο έχει προβλέψει ακριβώς όλες τις πραγματικές τιμές της εξαρτημένης μεταβλητής. Έστω, ένα ακόμη μέτρο που θα χρησιμοποιήσουμε στην ανάλυση μας ονομάζεται Mean absolute error (MAE).

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (9)$$

Οι μεταβλητές n , y_i και \hat{y}_i έχουν την ίδια ακριβώς ερμηνεία όπως στο μέτρο αξιολόγησης RMSE. Οι τιμές που μπορεί να πάρει το συγκεκριμένο μέτρο αξιολόγησης είναι μεγαλύτερες ή ίσες με το μηδέν. Όσο πιο κοντά είναι οι τιμές στο μηδέν, τόσο καλύτερη προβλεπτική ικανότητα έχει το μοντέλο. Το συγκεκριμένο μέτρο αξιολόγησης μετρά το απόλυτο μέσο σφάλμα πρόβλεψης μεταξύ των πραγματικών τιμών και των προβλεπόμενων τιμών της μεταβλητής στόχου. Με τον όρο σφάλμα εννοούμε τη διαφορά μεταξύ της πραγματικής τιμής και της προβλεπόμενης τιμής της εξαρτημένης μεταβλητής. Το συγκεκριμένο μέτρο, σε αντίθεση με το RMSE, δεν υψώνει τα σφάλματα στο τετράγωνο, δίνοντας σε όλα το ίδιο βάρος στην τελική τιμή. Τέλος, ένα ακόμα μέτρο απόδοσης του μοντέλου που θα χρησιμοποιήσουμε είναι το Index of Agreement (IA), το οποίο θα αναλύσουμε παρακάτω.

$$IA = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (|\hat{y}_i - \bar{y}| + |y_i - \bar{y}|)^2} \quad (10)$$

Οι παραπάνω μεταβλητές (y_i , \hat{y}_i , \bar{y}) έχουν αντίστοιχη ερμηνεία με το μέτρο NSE που αναφέραμε προηγουμένως. Ο συγκεκριμένος δείκτης κυμαίνεται από 0 έως 1. Όσο πιο κοντά στο 1 είναι ο δείκτης, τόσο περισσότερο συμφωνούν οι τιμές του μοντέλου, δηλαδή οι προβλεπόμενες τιμές, με τις παρατηρούμενες τιμές της εξαρτημένης μεταβλητής. Παρακάτω, χρησιμοποιώντας όλα τα μέτρα αξιολόγησης που αναφέραμε, θα αναλύσουμε τα αποτελέσματα του μοντέλου.

Τα μέτρα απόδοσης που αναφέρθηκαν στην συγκεκριμένη ενότητα χρησιμοποιούνται από τους Jana και Ghosh (2022) για να αξιολογήσουν ένα μοντέλο μηχανικής μάθησης random forest σε συνδυασμό με μια τεχνική αποσύνθεσης MODWT, με πλαίσιο ανάλυσης την πρόβλεψη της τιμής του φυσικού αερίου.

4.3 Αποτελέσματα

Οι διαφορετικές αναλογίες σετ εκπαίδευσης και σετ δοκιμής που αναφέρθηκαν στην ενότητα 4.1 έχουν χρησιμοποιηθεί στην αξιολόγηση του μοντέλου XGBoost χρησιμοποιώντας τα μέτρα απόδοσης που αναλύθηκαν στην ενότητα 4.2. Παρακάτω, οι Πίνακες 3, 4 και 5 απεικονίζουν την απόδοση του μοντέλου XGBoost όσον αφορά την πρόβλεψη της κατανάλωσης ενέργειας του Bitcoin, κάτω από διαφορετικές αναλογίες σετ εκπαίδευσης και δοκιμής, χρησιμοποιώντας παράλληλα 4 μέτρα αξιολόγησης. Το XGBoost είναι μια βιβλιοθήκη διαθέσιμη για το κοινό, η οποία χρησιμοποιείται από πολλές γλώσσες προγραμματισμού. Εμείς χρησιμοποιήσαμε τη γλώσσα προγραμματισμού python και τον αλγόριθμο μηχανικής μάθησης xgboost.XGBRegressor από την συγκεκριμένη βιβλιοθήκη, καθώς έχουμε συνεχή μεταβλητή ως εξαρτημένη (<https://xgboost.readthedocs.io/en/stable/>). Αξίζει να αναφέρουμε, προτού προχωρήσουμε στην ανάλυση, ότι ο χρόνος εκτέλεσης του μοντέλου XGBoost για μια αναλογία σετ εκπαίδευσης/δοκιμής είναι περίπου 3 λεπτά.

Πίνακας 3. Επίδοση σε Αναλογία Σετ Εκπαίδευσης/Δοκιμής (80%,20%).

Μέτρα αξιολόγησης	Χωρίς A.T	Με A.T	Με Σ.Ε	MLRM	GBrt
Σετ εκπαίδευσης					
NSE	0.50	0.23	0.35	0.26	0.51
IA	0.80	0.51	0.69		
MAE	0.06	0.11	0.08	0.09	0.06
RMSE	0.12	0.19	0.13	0.14	0.12
Σετ δοκιμής					
NSE	0.31	0.00	0.35	0.13	0.31
IA	0.72	0.37	0.69		
MAE	0.07	0.11	0.09	0.10	0.07
RMSE	0.14	0.18	0.15	0.15	0.14

Μεταβλητή στοχος: Κατανάλωση ενέργειας.

του Bitcoin, A.T = ακραίες τιμές.

Σ.Ε = σύνολο επιχύρωσης.

MLRM = Multiple Linear Regression Model.

GBrt = Gradient Boosting regression trees.

Πίνακας 4. Επίδοση σε Αναλογία Σετ Εκπαίδευσης/Δοκιμής (70%,30%).

Μέτρα αξιολόγησης	Χωρίς A.T	Με A.T	Με Σ.Ε	MLRM	GBrt
Σετ εκπαίδευσης					
NSE	0.47	0.25	0.36	0.25	0.49
IA	0.78	0.53	0.70		
MAE	0.07	0.12	0.07	0.09	0.07
RMSE	0.12	0.19	0.13	0.15	0.12
Σετ δοκιμής					
NSE	0.41	0.02	0.33	0.22	0.41
IA	0.76	0.40	0.67		
MAE	0.07	0.11	0.09	0.09	0.07
RMSE	0.12	0.18	0.15	0.14	0.12

Μεταβλητή στοχος : Κατανάλωση ενέργειας.

του Bitcoin, A.T = ακραίες τιμές.

Σ.Ε = σύνολο επιχύρωσης.

MLRM = Multiple Linear Regression Model.

GBrt = Gradient Boosting regression trees.

Πίνακας 5. Επίδοση σε Αναλογία Σετ Εκπαίδευσης/Δοκιμής (60%,40%).

Μέτρα αξιολόγησης	Χωρίς A.T	Με A.T	Με Σ.Ε	MLRM	GBrt
Σετ εκπαίδευσης					
NSE	0.50	0.28	0.37	0.23	0.46
IA	0.80	0.56	0.69		
MAE	0.07	0.12	0.08	0.09	0.07
RMSE	0.12	0.19	0.13	0.15	0.12
Σετ δοκιμής					
NSE	0.38	-0.00	0.33	0.26	0.40
IA	0.73	0.39	0.68		
MAE	0.07	0.12	0.08	0.09	0.07
RMSE	0.13	0.19	0.14	0.14	0.13

Μεταβλητή στοχος : Κατανάλωση ενέργειας.

του Bitcoin, A.T = ακραίες τιμές.

Σ.Ε = σύνολο επιχύρωσης.

MLRM = Multiple Linear Regression Model.

GBrt = Gradient Boosting regression trees.

Αρχικά, μπορούμε να παρατηρήσουμε στους πίνακες 3, 4 και 5 ότι, με βάση τις ακραίες τιμές που ορίστηκαν στην ενότητα 3.2, εάν δεν τις αφαιρέσουμε από το σύνολο δεδομένων, επηρεάζουν αρνητικά την προσαρμογή και, την προβλεπτική ικανότητα του μοντέλου XGBoost. Καθώς όλα τα μέτρα αξιολόγησης σε όλες τις αναλογίες σετ εκπαίδευσης και δοκιμής δείχνουν χειρότερα αποτελέσματα όταν επιλέγουμε να κρατήσουμε τις ακραίες τιμές συγκριτικά με το να τις αφαιρέσου-

με, επομένως η επιλογή να τις αφαιρέσουμε ήταν σωστή, καθώς επηρεάζει την ικανότητα πρόβλεψης της μεταβλητής στόχου. Στην τρίτη στήλη χρησιμοποιήσαμε σύνολο επικύρωσης, το οποίο αποτελεί ένα υποσύνολο που αντιστοιχεί στο 20% του συνόλου δεδομένων, από το οποίο έχουν αφαιρεθεί οι ακραίες τιμές. Το χρησιμοποιήσαμε ώστε, κατά τη διάρκεια της ρύθμισης των υπερπαραμέτρων, να μην υπάρξει κάποια αλληλεπίδραση με την προσαρμογή των παραμέτρων που πραγματοποιείται κατά την εκπαίδευση του μοντέλου. Παρατηρούμε ότι, σε όλες τις αναλογίες εκπαίδευσης/δοκιμής, η χρήση ενός συνόλου επικύρωσης μας έδωσε χειρότερα αποτελέσματα συγκριτικά με το μοντέλο χωρίς ακραίες τιμές, στο οποίο χρησιμοποιήθηκε το σύνολο εκπαίδευσης για τη ρύθμιση των υπερπαραμέτρων. Το μέτρο αξιολόγησης IA φαίνεται να ξεχωρίζει σε σχέση με τα άλλα μέτρα, παράγοντας αποτελέσματα κοντά στην μονάδα σε όλες τις περιπτώσεις χωρίς ακραίες τιμές. Αντίστοιχα, σε σύγκριση με τα άλλα μέτρα αξιολόγησης, με ακραίες τιμές, δίνει πολύ καλύτερα αποτελέσματα. Σύμφωνα με τους Legates και McCabe (1999), το μέτρο αξιολόγησης IA παρουσιάζει δυσκολίες στην ερμηνεία του και έχει παρατηρηθεί ότι εμφανίζει υψηλές τιμές σε σύγκριση με την πραγματική απόδοση του μοντέλου. Το μέτρο παρουσιάζει γενικά μεγάλη ευαισθησία στις ακραίες τιμές λόγω του τετραγώνου, γεγονός που οδηγεί σε παραπλανητικά μεγάλες τιμές. Παρόλο που χρησιμοποιείται για την αξιολόγηση μοντέλων μηχανικής μάθησης για τους παραπάνω λόγους δεν θα το λάβουμε υπόψη στα τελικά μας συμπεράσματα. Στην συνέχεια, όλοι οι παρακάτω σχολιασμοί της απόδοσης του μοντέλου θα αφορούν το σύνολο δεδομένων χωρίς ακραίες τιμές. Βλέπουμε ότι σχεδόν σε όλες τις αναλογίες σετ εκπαίδευσης και δοκιμής το μέτρο αξιολόγησης MAE είναι σταθερό και ίσο με 0.07. Αυτό σημαίνει ότι κατά μέσο όρο οι προβλεπόμενες τιμές απέχουν 0.07 από τις πραγματικές τιμές της εξαρτημένης μεταβλητής, χωρίς να λαμβάνει υπόψη την κατεύθυνση αυτής της διαφοράς, αντιμετωπίζοντας όλα τα σφάλματα με τον ίδιο συντελεστή.

Όσον αφορά το μέτρο αξιολόγησης RMSE, είναι σταθερό και ίσο με 0.12 σε όλες τις αναλογίες για το σετ εκπαίδευσης. Αυτό σημαίνει ότι οι προβλεπόμενες τιμές του μοντέλου XGBoost έχουν, κατά μέσο όρο απόκλιση 0.12 από τις πραγματικές τιμές της μεταβλητής κατανάλωση ενέργειας του Bitcoin. Στην περίπτωση του σετ δοκιμής, η απόκλιση διαφέρει ανάλογα με την αναλογία εκπαίδευσης και δοκιμής. Η καλύτερη επίδοση, όσον αφορά αυτό το μέτρο παρατηρείται με την αναλογία 70%, 30%, με μέση απόκλιση 0.12 ενώ η χειρότερη επίδοση παρατηρείται με την αναλογία 80%, 20%, με μέση απόκλιση 0.14. Στην μέση βρίσκεται η αναλογία 60%, 40%, με μέση απόκλιση 0.13. Σε όλες τις περιπτώσεις μπορούμε να παρατηρήσουμε ότι ισχύει $RMSE > MAE$. Αυτό οφείλεται στο γεγονός ότι τα μεγάλα σφάλματα έχουν πολύ μεγάλη επίδραση στο μέτρο RMSE, δηλαδή προβλεπόμενες τιμές που απέχουν πολύ από τις πραγματικές τιμές της μεταβλητής

στόχου. Τιμωρώντας έτσι το μοντέλο όταν αποκλίνει σημαντικά στις προβλέψεις του σε σύγκριση με την πραγματική τιμή της εξαρτημένης μεταβλητής. Τέλος, έχουμε το μέτρο NSE, για το οποίο θα συμβουλευτούμε τη βιβλιογραφία σχετικά με τις αποδεκτές τιμές, ώστε να θεωρείται ένα μοντέλο ικανό όσον αφορά την προβλεπτική του ικανότητα, και βάσει αυτού θα χαρακτηρίσουμε την απόδοση του μοντέλου μας. Σύμφωνα με τους Eryani et al. (2022), ένα μοντέλο είναι ικανοποιητικό εάν η τιμή του μέτρου NSE είναι μεταξύ 0.36 και 0.75. Οι Ritter και Muñoz-Carpena (2013) αναφέρουν ως αποδεκτή προβλεπτική ικανότητα του μοντέλου για τιμές $NSE \geq 0.65$. Επίσης, αναφέρουν ότι όταν $RMSE > MAE$, μπορεί να είναι μια ένδειξη παρουσίας ακραίων τιμών. Επιπλέον, αναφέρουν ότι το μέτρο NSE επηρεάζεται από ακραίες τιμές, καθώς η διαφορά της παρατηρούμενης τιμής με αυτή που υπολογίζεται από το μοντέλο τετραγωνίζεται, παρόμοια με το μέτρο RMSE. Τέλος, όσον αφορά το μέτρο αξιολόγησης NSE, οι Knoben et al. (2019) αναφέρουν ότι η προσαρμογή του μοντέλου στις παρατηρούμενες τιμές είναι καλή όταν το μέτρο είναι $NSE > 0$. Ο σχολιασμός, όπως αναφέρθηκε παραπάνω, εξακολουθεί να αφορά τα δεδομένα χωρίς ακραίες τιμές. Όσον αφορά το μέτρο NSE στο Σετ εκπαίδευσης, σύμφωνα με την πλειοψηφία των αναφορών, το μοντέλο έχει μια ικανοποιητική προβλεπτική ικανότητα, καθώς οι τιμές κυμαίνονται από 0.47 έως 0.50, ανάλογα με την αναλογία σετ εκπαίδευσης/δοκιμής. Ωστόσο, οι τιμές είναι μικρότερες από το 0.65, οπότε με βάση αυτό το πλαίσιο, κατατάσσεται ως ένα μοντέλο με μη αποδεκτή προβλεπτική ικανότητα. Αυτό μπορεί να οφείλετε είτε στην πιθανή παρουσία ακραίων τιμών, καθώς το $RMSE > MAE$ σε συνδυασμό με την ευαισθησία του μέτρου σε ακραίες τιμές. Παρόμοια συμπεράσματα μπορούν να προκύψουν και από το σετ δοκιμής, με τη διαφορά ότι στην αναλογία σετ εκπαίδευσης/δοκιμής (80%,20%), το μοντέλο, σύμφωνα με την πλειοψηφία της βιβλιογραφίας, κρίνεται ότι έχει μη ικανοποιητική προβλεπτική ικανότητα, κάνοντας παράλληλα την αναλογία με την χειρότερη απόδοση όσον αφορά το σετ δοκιμής, λαμβάνοντας υπόψη όλα τα μέτρα αξιολόγησης της απόδοσης του μοντέλου XGBoost. Στην τέταρτη στήλη έχουμε χρησιμοποιήσει ένα μοντέλο Multiple Linear Regression Model χωρίς ακραίες τιμές, για να συγκρίνουμε την απόδοση του με το μοντέλο αναφοράς. Καθώς μιλάμε για ένα από τα πιο γνωστά στατιστικά μοντέλα, το οποίο είναι εξαιρετικά εύκολο στην εφαρμογή του και έχει δείξει εξαιρετικά αποτελέσματα σε προβλήματα αιτιότητας και πρόβλεψης, παράλληλα είναι και ένα από τα πιο μελετημένα μοντέλα. Οι Hong et al. (2010) χρησιμοποιούν ένα μοντέλο Multiple Linear Regression για ερμηνευσιμότητα και πρόβλεψη της ωριαίας ζήτησης ενέργειας μιας αμερικανικής εταιρείας παροχής ηλεκτρικής ενέργειας, επιτυγχάνοντας παράλληλα μια πρόβλεψη με ικανοποιητική ακρίβεια. Στην συνέχεια, οι Hong et al. (2011) προτείνουν ένα μοντέλο αναφοράς το οποίο θα χρησιμοποιείται για σύγκριση (Benchmarking).

Αναφέρουν ένα απλό μοντέλο πολλαπλής γραμμικής παλινδρόμησης για πρόβλεψη βραχυπρόθεσμης ζήτησης, δηλαδή της κατανάλωσης ηλεκτρικής ενέργειας, βασισμένο σε μια Αμερικανική εταιρεία παροχής ενέργειας. Χρησιμοποιώντας διάφορα μοντέλα, όπως οικονομετρικά μοντέλα και μοντέλα μηχανικής μάθησης, με σκοπό την πρόβλεψη της τιμής του αργού πετρελαίου(WTI), Οι Xu et al. (2023) κατέληξαν σε ορισμένα συμπεράσματα. Ένα από αυτά ήταν ότι από τα συνηθισμένα οικονομετρικά μοντέλα, το μοντέλο Ordinary Least Square είχε την καλύτερη απόδοση σε σύγκριση με τα ARIMA και GARCH. Τέλος, παρατηρώντας τους πίνακες σε όλες τις αναλογίες σετ εκπαίδευσης/δοκιμής, σύμφωνα με τα μέτρα αξιολόγησης, το μοντέλο μας έχει καλύτερη προβλεπτική ικανότητα από το μοντέλο αναφοράς (Benchmarking). Η σύγκριση γίνεται μόνο με το σετ δοκιμής, καθώς αυτό αποτελεί και την καλύτερη εκτίμηση της αναμενόμενης απόδοσης του μοντέλου μας στο υπό εξέταση πλαίσιο. Στην πέμπτη στήλη, έχουμε χρησιμοποιήσει ένα μοντέλο Gradient Boosting regression tree, το οποίο θα χρησιμοποιηθεί ως επιπλέον μοντέλο αναφοράς. Λόγω των εξαιρετικών επιδόσεων του σε μια πληθώρα εφαρμογών, θα μας επιτρέψει να έχουμε ένα ακόμα καλύτερο σημείο αναφοράς όσον αφορά την προβλεπτική ικανότητα του μοντέλου μας. Συγκριτικά, και τα δύο μοντέλα έχουν παρόμοιες επιδόσεις στις αναλογίες σετ εκπαίδευσης/δοκιμής, με την μεγαλύτερη διαφορά να εμφανίζεται στο σετ εκπαίδευσης/δοκιμής με αναλογία 60%,40%. Σε αυτή την περίπτωση, το μοντέλο Gradient Boosting regression tree παρουσιάζει καλύτερη επίδοση στο σετ δοκιμής και μικρότερη τάση για υπερπροσαρμογή. Στις υπόλοιπες αναλογίες, το μοντέλο Gradient Boosting regression tree έχει παρόμοιες επιδόσεις, αλλά παρουσιάζει μεγαλύτερη τάση για υπερπροσαρμογή σε σχέση με το μοντέλο XGBoost. Η παραπάνω περιγραφή έγινε λαμβάνοντας υπόψη το μέτρο αξιολόγησης NSE. Σύμφωνα με τα υπόλοιπα μέτρα αξιολόγησης, οι επιδόσεις των δύο μοντέλων είναι ίσες.

5 ΣΥΜΠΕΡΑΣΜΑΤΑ

Στο παρόν άρθρο, χρησιμοποιήσαμε ένα μοντέλο μηχανικής μάθησης XGBoost για να προβλέψουμε την κατανάλωση ενέργειας του Bitcoin. Χρησιμοποιήσαμε ως εξαρτημένη μεταβλητή τον δείκτη Digiconomist Bitcoin Energy Consumption Index για μια περίοδο από τις 10 Φεβρουαρίου 2017 έως τις 7 Ιουλίου 2020 με ημερήσια συχνότητα. Χρησιμοποιήσαμε συνολικά 11 ανεξάρτητες μεταβλητές, μερικές από τις οποίες αποτελούν τη σύνθεση και τη λειτουργία του Bitcoin, ενώ οι υπόλοιπες αποτελούν τις οικονομικές και χρηματοοικονομικές ιδιότητες του Bitcoin.

Η διαδικασία απόκτησης του κρυπτονομίσματος Bitcoin απαιτεί τεράστια ποσά ηλεκτρικής ενέργειας, η οποία προέρχεται από τον άνθρακα. Ως αποτέλεσμα συμβάλλουν σε έναν υπολογίσιμο βαθμό στην παραγωγή βλαβερών ουσιών και θέτουν σε ακόμα μεγαλύτερο κίνδυνο τον πλανήτη μας. Η ανάπτυξη προβλεπτικών μοντέλων είναι σημαντική, καθώς παρέχει μια πιο ξεκάθαρη εικόνα, η οποία συνεπάγεται σε καλύτερη κατανόηση του προβλήματος, δηλαδή της κατανάλωσης ενέργειας του Bitcoin. Ως αποτέλεσμα, μπορεί να βελτιωθεί σημαντικά ο τρόπος αντιμετώπισης του. Από τη συγκεκριμένη μελέτη καταλήξαμε σε ορισμένα συμπεράσματα. Για αρχή, εξετάσαμε αν υπάρχει γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής και των ανεξάρτητων μεταβλητών. Χρησιμοποιώντας τον συντελεστή συσχέτισης Pearson και γραφήματα διασποράς, καταλήξαμε στο συμπέρασμα ότι δεν υπάρχει γραμμική σχέση μεταξύ των μεταβλητών. Τα ευρήματα φαίνεται να συμφωνούν με τη βιβλιογραφία, επιβεβαιώνοντας παράλληλα την καταλληλότητα του μοντέλου. Επιπλέον, αναφέραμε τα αποτελέσματα του μοντέλου XGBoost με και χωρίς ακραίες τιμές και διαπιστώσαμε ότι οι ακραίες τιμές επηρεάζουν αρνητικά την προβλεπτική ικανότητα του μοντέλου. Για να βελτιστοποιήσουμε την επιλογή υπερπαραμέτρων για το μοντέλο μας, χρησιμοποιήσαμε Μπεϋζιανή βελτιστοποίηση. Στην συνέχεια, χωρίσαμε τα δεδομένα μας σε υπόσυνολα εκπαίδευσης και δοκιμής σε διαφορετικές αναλογίες αντίστοιχα. Καταλήξαμε στο συμπέρασμα ότι η αναλογία σετ εκπαίδευσης/δοκιμής με τη χειρότερη συνολική επίδοση, δίνοντας μεγαλύτερο βάρος στο σετ δοκιμής, ήταν 80%/20% αντίστοιχα. Ένα αποτέλεσμα που δεν φαίνεται να συμφωνεί με τη βιβλιογραφία, καθώς το μοντέλο, υπό τις συγκεκριμένες συνθήκες, έχει ένα αρκετά μεγάλο δείγμα ώστε να εκπαιδευτεί και άρα ένα μεγαλύτερο σύνολο πληροφοριών για να προσαρμοστούν οι παράμετροι του. Αυτό φαίνεται να λειτουργεί αρνητικά στο μοντέλο XGBoost καθώς είναι και η αναλογία με την καλύτερη επίδοση σύμφωνα με τα μέτρα αξιολόγησης στο σετ εκπαίδευσης, αλλά τη χειρότερη στο σετ δοκιμής, προκαλώντας στο μοντέλο μια επιπλέον τάση για υπερπροσαρμογή συγκριτικά με τις υπόλοιπες αναλογίες σετ εκπαίδευσης/δοκιμής. Επίσης, στην περίπτωση που

χρησιμοποιούμε σύνολο επικύρωσης, το οποίο είναι θετικό, καθώς κατά τη διαδικασία βελτιστοποίησης των υπερπαραμέτρων δεν έρχεται σε επαφή με το σύνολο εκπαίδευσης και, ως αποτέλεσμα, δεν επηρεάζει τη διαδικασία προσαρμογής των παραμέτρων. Παρόλα αυτά, τα αποτελέσματα που προέκυψαν ήταν χειρότερα σε σύγκριση με την περίπτωση που χρησιμοποιήσαμε το σύνολο εκπαίδευσης στη διαδικασία προσδιορισμού των υπερπαραμέτρων. Αυτό μπορεί να οφείλεται στο μέγεθος του δείγματος, καθώς, χρησιμοποιώντας ένα μέρος των δεδομένων για το σύνολο επικύρωσης, μειώνονται οι παρατηρήσεις που θα χρησιμοποιηθούν στο σύνολο εκπαίδευσης. Ως αποτέλεσμα, το μοντέλο έχει λιγότερες πληροφορίες για τη σχέση των μεταβλητών, γεγονός που οδηγεί σε χειρότερη εκτίμηση των παραμέτρων του μοντέλου, καθιστώντας το μέγεθος του δείγματος έναν σημαντικό παράγοντα στην προβλεπτική ικανότητα του μοντέλου. Όσον αφορά μια πιο γενική εικόνα για την απόδοση του μοντέλου XGBoost και την καταλληλότητα του για το υπό εξέταση πλαίσιο, για τις αναλογίες σετ εκπαίδευσης/δοκιμής 70%,30% και 60%,40% αντίστοιχα, καταλήγουμε στο συμπέρασμα ότι το μοντέλο έχει μια αποδεκτή προβλεπτική ικανότητα. Συγκεκριμένα, με βάση τα μέτρα αξιολόγησης $NSE(0.41,0.38)$, $RMSE(0.12,0.13)$ και $MAE(0.07)$ που χρησιμοποιήθηκαν, το μοντέλο παρέχει ικανοποιητική απόδοση για το υπό εξέταση πλαίσιο. Επίσης, η απόδοση του μοντέλου μας συγκρίθηκε με ένα μοντέλο πολλαπλής γραμμικής παλινδρόμησης. Είναι ένα μοντέλο ευρέως διαδεδομένο στην οικονομική επιστήμη, με πολύ εύκολη και κατανοητή εφαρμογή, εκτενή έρευνα και καλές αποδόσεις σε μια πληθώρα εφαρμογών, καθιστώντας το κατάλληλο για σύγκριση. Το μοντέλο μας έχει καλύτερη προβλεπτική ικανότητα σύμφωνα με τα μέτρα αξιολόγησης σε όλες τις αναλογίες σετ εκπαίδευσης/δοκιμής. Όσον αφορά το σετ δοκιμής, αυτό αποτελεί την καλύτερη εκτίμηση της προβλεπτικής ικανότητας του μοντέλου μας, δίνοντας έτσι μια ακόμη επιβεβαίωση της καταλληλότητας του μοντέλου μας στο υπό εξέταση πλαίσιο. Για να έχουμε μια ακόμα καλύτερη εικόνα σχετικά με την προβλεπτική ικανότητα του μοντέλου μας, χρησιμοποιήσαμε ένα επιπλέον μοντέλο σύγκρισης, το οποίο ονομάζεται Gradient Boosting regression tree. Πρόκειται για ένα πολύ δημοφιλές μοντέλο μηχανικής μάθησης με εξαιρετικές επιδόσεις σε προβλήματα πρόβλεψης. Παρόλο που ο αλγόριθμος μηχανικής μάθησης XGBoost χρησιμοποιεί μια προσεγγιστική μέθοδο για την επιλογή των χαρακτηριστικών βάσει των οποίων θα γίνει ο διαχωρισμός, σε σύγκριση με τον αλγόριθμο μηχανικής μάθησης Gradient Boosting regression tree, ο οποίος δοκιμάζει όλους τους δυνατούς συνδυασμούς χαρακτηριστικών για να βρει την καλύτερη διαίρεση, έχουν επιτύχει παρόμοια επίδοση στο υπό εξέταση πλαίσιο. Αυτή η προσεγγιστική μέθοδος του αλγορίθμου XGBoost προσφέρει πλεονέκτηματα στον χρόνο εκτέλεσης σε σύγκριση με τον αλγόριθμο Gradient Boosting regression tree, παρέχοντας παράλληλα παρόμοιες επιδόσεις με ταχύτερη εκτέλεση. Με

έναν αλγόριθμο που έχει σημειώσει τόσο μεγάλη επιτυχία και έχει εφαρμοστεί σε ένα ευρύ φάσμα μελετών, αυτό αποτελεί ένα ακόμα θετικό στοιχείο όσον αφορά την απόδοση του αλγορίθμου στο υπό εξέταση πλαίσιο. Επιπλέον, διαθέτει την ικανότητα να δοκιμάζεται αποτελεσματικά σε πολύ μεγαλύτερα δείγματα σε σχέση με το μοντέλο Gradient Boosting regression tree, προσφέροντας ένα μεγαλύτερο εύρος πιθανών καταστάσεων στις οποίες ο αλγόριθμος XGBoost μπορεί να χρησιμοποιηθεί συγκριτικά με τον αλγόριθμο Gradient Boosting regression tree. Αυτό καθιστά σημαντική την εξέταση του στο υπό εξέταση πλαίσιο, δεδομένου ότι είναι και ένας σχετικά πιο νέος αλγόριθμος μηχανικής μάθησης, με περιθώρια για περαιτέρω μελέτη. Λαμβάνοντας παράλληλα υπόψη ότι πρόκειται για μια δύσκολη περίοδο πρόβλεψης, καθώς η εξαρτημένη μεταβλητή, ή αλλιώς μεταβλητή στόχος, παρουσίασε μεγάλη μεταβλητότητα γύρω από τον μέσο όρο, όπως προκύπτει από τον συντελεστή διακύμανσης ($CV = 1.15$). Αυτό φαίνεται επίσης και οπτικά στην χρονοσειρά της μεταβλητής. Παρατηρήσαμε επίσης ότι το $RMSE > MAE$, γεγονός που υποδηλώνει ότι για ορισμένες παρατηρήσεις της μεταβλητής στόχου, η τιμή της πρόβλεψης απέχει πολύ από την πραγματική τιμή. Αυτό είναι πιθανό να επηρεάζεται από τη μεγάλη μεταβλητότητα της εξαρτημένης μεταβλητής κατά την υπό εξέταση περίοδο. Επιπλέον, αυτή η μεγάλη μεταβλητότητα θα μπορούσε να έχει προκληθεί από τη μεγάλη μεταβλητότητα της τιμής του Bitcoin, όπου, σύμφωνα με τη βιβλιογραφία, η τιμή του Bitcoin επηρεάζει την κατανάλωση ενέργειας του Bitcoin. Παρόλα αυτά, για μια δύσκολη περίοδο, το τελικό συμπέρασμα είναι ότι το μοντέλο XGBoost κατάφερε να αποδώσει σε ικανοποιητικό βαθμό με χρόνο εκτέλεσης μόλις περίπου 3 λεπτά¹ ανά αναλογία σετ εκπαίδευσης/δοκιμής, καθιστώντας το χρήσιμο στο υπό εξέταση πλαίσιο. Επιπλέον, ο γρήγορος χρόνος εκτέλεσης του το καθιστά αποτελεσματικό, καθώς οι αλγόριθμοι μηχανικής μάθησης κατά την εκτέλεση τους απαιτούν μεγάλα ποσά ενέργειας, γεγονός που είναι σημαντικό και θα πρέπει να λαμβάνεται υπόψη στην αξιολόγηση της απόδοσης ενός μοντέλου μηχανικής μάθησης.

Ένας σημαντικός περιορισμός της παρούσας έρευνας είναι ότι η κατανάλωση ενέργειας του Bitcoin, η οποία χρησιμοποιήθηκε ως μεταβλητή στόχος στο μοντέλο και πάνω στην οποία βασίστηκαν τα συμπεράσματά μας, αποτελεί εκτίμηση και όχι πραγματική τιμή. Επίσης, χρησιμοποιήσαμε το υποσύνολο εκπαίδευσης για τη ρύθμιση των υπερπαραμέτρων, το οποίο μπορεί να έχει επηρεάσει τη διαδικασία εκτίμησης των παραμέτρων του μοντέλου. Μελλοντικές έρευνες θα μπορούσαν να εξετάσουν διαφορετικές περιόδους, για να διαπιστώσουν κατά πόσο η μεταβλητότητα της κατανάλωσης ενέργειας του Bitcoin επηρεάζει την προβλεπτική ικανότητα του μοντέλου XGBoost.

¹GPU: NVIDIA GeForce GTX 1080 Ti, CPU: AMD RYZEN 9 7900X.

6 ΒΙΒΛΙΟΓΡΑΦΙΑ

Di Persio, L., & Fraccarolo, N. (2023). Energy consumption forecasts by gradient boosting regression trees. *Mathematics*, 11(5), 1068.

Yoon, J. (2021). Forecasting of real GDP growth using machine learning models: Gradient boosting and random forest approach. *Computational Economics*, 57(1), 247-265.

Lee, W., Schwartz, N., Bansal, A., Khor, S., Hammarlund, N., Basu, A., & Devine, B. (2022). A scoping review of the use of machine learning in health economics and outcomes research: part 2—data From nonwearables. *Value in Health*, 25(12), 2053-2061.

Smith, M., & Alvarez, F. (2022). Predicting firm-level bankruptcy in the Spanish economy using extreme gradient boosting. *Computational Economics*, 59(1), 263-295.

Dey, S., Kumar, Y., Saha, S., & Basak, S. (2016). Forecasting to Classification: Predicting the direction of stock market price using Xtreme Gradient Boosting. *PESIT South Campus*, 1-10.

Jabeur, S. B., Mefteh-Wali, S., & Viviani, J. L. (2024). Forecasting gold price with the XGBoost algorithm and SHAP interaction values. *Annals of Operations Research*, 334(1), 679-699.

Gavoille, N., & Zasova, A. (2023). What we pay in the shadows: Labor tax evasion, minimum wage hike and employment. *Journal of Public Economics*, 228, 105027.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, 1189-1232.

Schapire, R. E. (1990). The strength of weak learnability. *Machine learning*, 5, 197-227.

Guven, D., & Kayalica, M. O. (2023). Analysing the determinants of the Turkish household electricity consumption using gradient boosting regression tree. *Energy for Sustainable Development*, 77, 101312.

Li, T. R., Chamrajnagar, A. S., Fong, X. R., Rizik, N. R., & Fu, F. (2019). Sentiment-based prediction of alternative cryptocurrency price fluctuations using gradient boosting tree model. *Frontiers in Physics*, 7, 98.

Tian, Z., Xiao, J., Feng, H., & Wei, Y. (2020). Credit risk assessment based on gradient boosting decision tree. *Procedia Computer Science*, 174, 150-160.

Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. *The North American Journal of Economics and Finance*, 47, 552-567.

Prasad, A., Bakhshi, P., & Seetharaman, A. (2022). The impact of the US macroeconomic variables on the CBOE VIX Index. *Journal of Risk and Financial Management*, 15(3), 126.

Soybilgen, B., & Yazgan, E. (2021). Nowcasting us gdp using tree-based ensemble models and dynamic factors. *Computational Economics*, 57(1), 387-417.

Derbentsev, V., Datsenko, N., Babenko, V., Pushko, O., & Pursky, O. (2020, October). Forecasting cryptocurrency prices using ensembles-based machine learning approach. In *2020 IEEE International Conference on Problems of Infocommunications. Science and Technology (PIC S&T)* (pp. 707-712). IEEE.

Liu, J., Li, C., Ouyang, P., Liu, J., & Wu, C. (2023). Interpreting the prediction results of the tree-based gradient boosting models for financial distress prediction with an explainable machine learning approach. *Journal of Forecasting*, 42(5), 1112-1137.

Kipkogei, F., Kabano, I. H., Murorunkwere, B. F., & Joseph, N. (2021). Business success prediction in Rwanda: a comparison of tree-based models and logistic regression classifiers. *SN Business & Economics*, 1, 1-19.

Atalan, A. (2023). Forecasting drinking milk price based on economic, social, and environmental factors using machine learning algorithms. *Agribusiness*, 39(1), 214-241.

Hu, Y., Guo, H., Huang, W., & Xu, Y. (2023). Yield Forecasting by Machine Learning Algorithm: Evidence from China's A-share Market. *Emerging Markets Finance and Trade*, 59(6), 1767-1781.

Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, 158, 1533-1543.

Nakamoto, Satoshi, Bitcoin: A Peer-to-Peer Electronic Cash System (August 21, 2008). Available at SSRN: <https://ssrn.com/abstract=3440802> or <http://dx.doi.org/10.2139/ssrn.3440802>

Huynh, A. N. Q., Duong, D., Burggraf, T., Luong, H. T. T., & Bui, N. H. (2022). Energy consumption and Bitcoin market. *Asia-Pacific Financial Markets*, 29(1), 79-93.

Vranken, H. (2017). Sustainability of bitcoin and blockchains. *Current opinion in environmental sustainability*, 28, 1-9.

Küfeoglu, S., & Özkuran, M. (2019). Energy consumption of bitcoin mining.

Truby, J. (2018). Decarbonizing Bitcoin: Law and policy choices for reducing the energy consumption of Blockchain technologies and digital currencies. *Energy research & social science*, 44, 399-410.

Wijayanti, E. B., & Setyoko, B. H. (2024). Dataset Analysis and Feature Characteristics to Predict Rice Production based on eXtreme Gradient Boosting. *Journal of Computing Theories and Applications*, 1(3), 299-310.

Mir, U. (2020). Bitcoin and its energy usage: Existing approaches, important opinions, current trends, and future challenges. *KSII Transactions on Internet and Information Systems (TIIS)*, 14(8), 3243-3256.

Dilek, Ş., & Furuncu, Y. (2019). Bitcoin mining and its environmental effects. *Atatürk üniversitesi iktisadi ve idari bilimler dergisi*, 33(1), 91-106.

De Vries, A. (2021). Bitcoin boom: What rising prices mean for the network's energy consumption. *Joule*, 5(3), 509-513.

Kohli, V., Chakravarty, S., Chamola, V., Sangwan, K. S., & Zeadally, S. (2023). An analysis of energy consumption and carbon footprints of cryptocurrencies and possible solutions. *Digital Communications and Networks*, 9(1), 79-89.

Stoll, C., Klaaßen, L., & Gallersdörfer, U. (2019). The carbon footprint of bitcoin. *Joule*, 3(7), 1647-1661.

De Vries, A. (2018). Bitcoin’s growing energy problem. *Joule*, 2(5), 801-805.

Carmona, P., Climent, F., & Momparler, A. (2019). Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 61, 304-323.

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).

Roos, M. C. (2021). Bitcoin energy index: a transparent and dynamic model (Bachelor’s thesis, University of Twente).

Sapra, N., Shaikh, I., Roubaud, D., Asadi, M., & Grebinevych, O. (2024). Uncovering Bitcoin’s electricity consumption relationships with volatility and price: Environmental Repercussions. *Journal of Environmental Management*, 356, 120528.

Jamali, A., Ali, N. I., Brohi, I. A., Kanasro, N. A., Murad, M. U., & Jamali, A. A. (2024, January). Exploring Relationships among Bitcoin’s Market Price, Energy Consumption and Carbon Dioxide Emissions: A Machine Learning Approach. In *2024 IEEE 1st Karachi Section Humanitarian Technology Conference (KHI-HTC)* (pp. 1-6). IEEE.

Sizer, L., & Yilmaz, Y. (2022). Analysis of the Relationship Between Bitcoin Electricity Consumption and the Global Economic and Political Uncertainty Index (GEPU). *International Journal of Business and Economic Studies*, 4(1), 45-52.

De Vries, A. (2020). Bitcoin’s energy consumption is underestimated: A market dynamics approach. *Energy Research & Social Science*, 70, 101721.

Calvo-Pardo, H. F., Mancini, T., & Olmo, J. (2022). Machine learning the carbon footprint of bitcoin mining. *Journal of Risk and Financial Management*, 15(2), 71.

Sapra, N., & Shaikh, I. (2023). Impact of Bitcoin mining and crypto market determinants on Bitcoin-based energy consumption. *Managerial Finance*, 49(11), 1828-1846.

Ben Omrane, W., Qi, Q., & Saadi, S. (2023). Cryptocurrency markets, macroeconomic news announcements and energy consumption. *Annals of Operations Research*, 1-18.

Fahmy, H. (2022). Assessing the carbon footprint of cryptoassets: evidence from a bivariate VAR model. In *Big Data in Finance: Opportunities and Challenges of Financial Digitalization* (pp. 207-230). Cham: Springer International Publishing.

Zhang, D., Chen, X. H., Lau, C. K. M., & Xu, B. (2023). Implications of cryptocurrency energy usage on climate change. *Technological Forecasting and Social Change*, 187, 122219.

Sarkodie, S. A., Ahmed, M. Y., & Leirvik, T. (2022). Trade volume affects bitcoin energy consumption and carbon footprint. *Finance Research Letters*, 48, 102977.

Jana, R. K., Ghosh, I., & Wallin, M. W. (2022). Taming energy and electronic waste generation in bitcoin mining: Insights from Facebook prophet and deep neural network. *Technological Forecasting and Social Change*, 178, 121584.

Schinckus, C., Nguyen, C. P., & Chong, F. H. L. (2022). Cryptocurrencies' hashrate and electricity consumption: evidence from mining activities. *Studies in Economics and Finance*, 39(3), 524-546.

Das, D., & Dutta, A. (2020). Bitcoin's energy consumption: Is it the Achilles heel to miner's revenue?. *Economics Letters*, 186, 108530.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational research methods*, 16(2), 270-301.

Zhang, W., Gu, X., Hong, L., Han, L., & Wang, L. (2023). Comprehensive review of machine learning in geotechnical reliability analysis: Algorithms, applications and further challenges. *Applied Soft Computing*, 136, 110066.

Jana, R. K., & Ghosh, I. (2022). A residual driven ensemble machine learning approach for forecasting natural gas prices: Analyses for pre-and during-COVID-19 phases. *Annals of Operations Research*, 1-22.

Legates, D. R., & McCabe Jr, G. J. (1999). Evaluating the use of “goodness-of-fit” measures in hydrologic and hydroclimatic model validation. *Water resources research*, 35(1), 233-241.

Eryani, I. G. A. P. (2022). Sensitivity analysis in parameter calibration of the WEAP Model for integrated water resources management in Unda watershed. *Civil Engineering and Architecture*, 455-469.

Ritter, A., & Muñoz-Carpena, R. (2013). Performance evaluation of hydrological models: Statistical significance for reducing subjectivity in goodness-of-fit assessments. *Journal of Hydrology*, 480, 33-45.

Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, 23(10), 4323-4331.

Hong, T., Gui, M., Baran, M. E., & Willis, H. L. (2010, July). Modeling and forecasting hourly electric load by multiple linear regression with interactions. In *Ieee pes general meeting* (pp. 1-8). IEEE.

Hong, T., Wang, P., & Willis, H. L. (2011, July). A naïve multiple linear regression benchmark for short term load forecasting. In *2011 IEEE power and energy society general meeting* (pp. 1-6). IEEE.

Xu, Z., Mohsin, M., Ullah, K., & Ma, X. (2023). Using econometric and machine learning models to forecast crude oil prices: Insights from economic history. *Resources Policy*, 83, 103614.