
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΗ ΛΟΓΙΣΤΙΚΗ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ

Διπλωματική Εργασία

# ΕΥΛΟΓΗ ΑΞΙΑ ΚΑΙ ΤΙΜΟΛΟΓΗΣΗ ΚΡΥΠΤΟΝΟΜΙΣΜΑΤΩΝ ΜΕ ΤΕΧΝΙΚΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

του

ΙΩΑΝΝΗ ΛΙΑΓΓΟΥ
Επιβλέπων Καθηγητής: Ανέστης Λαδάς

Υποβλήθηκε ως απαιτούμενο για την απόκτηση του Μεταπτυχιακού Διπλώματος στη Λογιστική και Χρηματοοικονομική

Δεκέμβριος 2023

# ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία ασχολείται με τον υπολογισμό της εύλογης αξίας και την τιμολόγηση των κρυπτονομισμάτων χρησιμοποιώντας τεχνικές Μηχανικής Μάθησης. Αρχικά παρουσιάζεται η βιβλιογραφική επισκόπηση που έχει σκοπό να παραθέσει μια ολοκληρωμένη εικόνα του τρόπου με τον οποίο πραγματοποιείται ο υπολογισμός της εύλογης αξίας και του τρόπου τιμολόγησης των κρυπτονομισμάτων εξετάζοντας ενδελεχώς μελέτες και μοντέλα που πραγματοποιούν αυτόν τον υπολογισμό κυρίως με την χρήση τεχνικών μηχανικής μάθησης. Στη συνέχεια διερευνάται η τεχνολογία της Αλυσίδας Συστοιχιών και οι εφαρμογές της, καθώς και το πεδίο της Μηχανικής Μάθησης με ιδιαίτερη αναφορά στον συνδυασμό του με την Ανάλυση Συναισθημάτων. Μετέπειτα, πραγματοποιείται πρόβλεψη τιμής του κρυπτονομίσματος «Bitcoin» με τη χρήση τριών διαφορετικών μοντέλων Μηχανικής Μάθησης, του μοντέλου Δικτύων Μακράς Βραχύχρονης Μνήμης (LSTM), του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) και ενός Υβριδικού μοντέλου που χρησιμοποιεί Ανάλυση Συναισθημάτων και αποτελείται από Δίκτυα Αναδρομικής Πύλης (GRU) και Συνελικτικά Νευρωνικά Δίκτυα (CNN). Στο τέλος, συγκρίνονται στατιστικά το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU) με το Υβριδικό μοντέλο και προκύπτει ότι το Υβριδικό μοντέλο με την χρήση της Ανάλυσης Συναισθημάτων λειτουργεί πιο αποδοτικά. Μετά από αυτό το συμπέρασμα, προτείνεται για επιπρόσθετη μελλοντική μελέτη, να μελετηθεί ο τρόπος με τον οποίο επιλέγονται τα δεδομένα της Ανάλυσης Συναισθημάτων και η αναβάθμιση των Μεταμορφωτών, με πιο προηγμένους, οι οποίοι χρησιμοποιούνται για την μετατροπή των δεδομένων κειμένου της Ανάλυσης Συναισθημάτων σε μεταβλητές.

**Λέξεις-Κλειδιά**: Αλυσίδα Συστοιχιών, Μηχανική Μάθηση, Ανάλυση Συναισθημάτων, Δίκτυα Μακράς Βραχύχρονης Μνήμης (LSTM), Δίκτυα Αναδρομικής Πύλης (GRU), Συνελικτικά Νευρωνικά Δίκτυα (CNN), Μεταμορφωτές

# ABSTRACT

The present thesis deals with the calculation of fair value and the pricing of cryptocurrencies using Machine Learning techniques. Initially, a literature review is presented, aiming to provide a comprehensive picture of how the fair value calculation and cryptocurrency pricing is carried out, examining in detail studies and models that perform this calculation primarily using machine learning techniques. Subsequently, the technology of Blockchain and its applications are explored, as well as the field of Machine Learning, with particular reference to its combination with Sentiment Analysis. Then, a price prediction for the cryptocurrency "Bitcoin" is made using three different Machine Learning models: the Long Short-Term Memory (LSTM) model, the Gated Recurrent Unit (GRU) model, and a Hybrid model that uses Sentiment Analysis and consists of Gated Recurrent Units (GRU) and Convolutional Neural Networks (CNN). In the end, the Gated Recurrent Unit (GRU) model is statistically compared with the Hybrid model, and it is concluded that the Hybrid model, using Sentiment Analysis, operates more efficiently. Following this conclusion, it is suggested for future research to further study the way in which Sentiment Analysis data is selected and to upgrade the Transformers used to convert the text data of Sentiment Analysis into variables, making them more advanced.

**Keywords**: Blockchain, Machine Learning, Sentiment Analysis, Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolutional Neural Networks (CNN), Transformers

# ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα πρώτα απ' όλα να ευχαριστήσω τον καθηγητή και επιβλέποντα μου, Ανέστη Λαδά για την βοήθεια, την καθοδήγηση και τη στήριξη που μου προσέφερε κατά τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας. Θέλω να ευχαριστήσω τους φίλους που με στήριξαν, με συμβούλεψαν και με ενθάρρυναν όταν το χρειαζόμουν. Τέλος, αλλά όχι λιγότερο σημαντικό, θέλω να εκφράσω τη βαθιά μου ευγνωμοσύνη προς την οικογένειά μου για την αγάπη, την πίστη και τη στήριξη που μου παρείχαν ανελλιπώς. Είναι το ισχυρότερο μου κίνητρο και τους ανήκει η εργασία αυτή.

# ΠΕΡΙΕΧΟΜΕΝΑ

# ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

## 1.1 Αντικείμενο της Διπλωματικής

Η παρούσα Διπλωματική εργασία διερευνά την τεχνολογία Αλυσίδας Συστοιχιών (Blockchain) και το πεδίο της Μηχανικής Μάθησης (Machine Learning) με ιδιαίτερη αναφορά στον συνδυασμό του με το πεδίο της Ανάλυσης Συναισθημάτων (Sentiment Analysis), στοχεύοντας στην δημιουργία ενός αποτελεσματικότερου μοντέλου πρόβλεψης τιμής κρυπτονομισμάτων από τα ήδη υπάρχοντα, τετριμμένα μοντέλα Μηχανικής Μάθησης. Αρχικά, περιγράφεται λεπτομερώς η τεχνολογία Αλυσίδας Συστοιχιών, παρουσιάζονται τα μοντέλα της και τονίζεται η σημαντικότητα της τεχνολογίας αυτής επισημαίνοντας τις πολλαπλές εφαρμογές της σε διάφορα πεδία, με έμφαση την εφαρμογή της στα ψηφιακά συστήματα μετρητών, που έχουν πάρει τεράστιες διαστάσεις τα τελευταία χρόνια. Συγκεκριμένα, ο υπολογισμός της εύλογης αξίας και η τιμολόγηση των κρυπτονομισμάτων έχει απασχολήσει πολλούς ερευνητές σε παγκόσμιο επίπεδο καθώς ολοένα και περισσότερο γίνεται χρήση των κρυπτονομισμάτων στην καθημερινή ζωή, τόσο από επιχειρήσεις όσο και από επενδυτές. Με αφορμή το γεγονός αυτό, έχουν γνωστοποιηθεί πολλά μοντέλα, μέθοδοι και τεχνικές στην βιβλιογραφία που παρουσιάζουν ιδιαίτερο ενδιαφέρον και χρήζουν περεταίρω διερεύνησης. Η παρούσα διπλωματική λοιπόν, αναλύει κάποια μοντέλα εξ' αυτών θέλοντας να βρει τα πιο αποδοτικά ως προς τον υπολογισμό της εύλογης αξίας και τιμολόγησης των κρυπτονομισμάτων, ώστε στη συνέχεια να παραθέσει ένα νέο μοντέλο που παρουσιάζει καλύτερες αποδόσεις, συνδυάζοντας τα υπάρχοντα μοντέλα με την Ανάλυση Συναισθημάτων και την χρήση Μεταμορφωτών (Transformers). Για τον σκοπό αυτό, γίνονται διάφορες δοκιμές και συγκρίσεις και επιλέγεται το καλύτερο εξ' αυτών, το Υβριδικό Μοντέλο Μηχανικής Μάθησης, το οποίο αποτελεί έναν συνδυασμός ενός Δικτύων Αναδρομικής Πύλης (GRU) και Συνελικτικών Νευρωνικών Δικτύων (CNN) με Ανάλυση Συναισθημάτων (Sentiment Analysis).

## 1.2 Δομή της Διπλωματικής

Η παρούσα διπλωματική εργασία αποτελείται από πέντε κεφάλαια τα οποία είναι: Το κεφάλαιο «Εισαγωγή», το κεφάλαιο «Βιβλιογραφική Επισκόπηση», το κεφάλαιο «Τεχνολογία Αλυσίδας Συστοιχιών (Blockchain Technology)», το κεφάλαιο «Μηχανική Μάθηση (Machine Learning)» και τέλος, το κεφάλαιο «Πρόβλεψη Τιμής Κρυπτονομισμάτων».

Στο κεφάλαιο «Εισαγωγή», προσδιορίζεται το θέμα που θα εξεταστεί στην παρούσα εργασία, ο βασικός σκοπός, καθώς και το πρόβλημα-θέμα που στάθηκε αφορμή για την μελέτη του συγκεκριμένου θέματος.

Στο κεφάλαιο «Βιβλιογραφική Επισκόπηση», παρατίθεται μία ολοκληρωμένη εικόνα του τρόπου με τον οποίο μοντέλα, κυρίως με χρήση τεχνικών μηχανικής μάθησης με ή χωρίς την χρήση Ανάλυσης Συναισθημάτων, υπολογίζουν την εύλογη αξία των κρυπτονομισμάτων και τα τιμολογούν, βάσει της υφιστάμενης βιβλιογραφίας.

Στο κεφάλαιο «Τεχνολογία Αλυσίδας Συστοιχιών (Blockchain Technology)», ερευνάται ενδελεχώς η τεχνολογία Αλυσίδας Συστοιχιών. Αρχικά, γίνεται μία ιστορική αναδρομή ώστε να κατανοηθεί η ιδέα της ύπαρξης της στο πέρασμα του χρόνου, παρουσιάζονται οι διάφορες κατηγορίες της και αναλύεται η αρχιτεκτονική της. Στη συνέχεια, διερευνώνται τα Μοντέλα

Συναίνεσης και η πραγματοποίηση διακλαδώσεων και τέλος, επισημαίνονται οι πολλαπλές εφαρμογές της σε πληθώρα τομέων στη σύγχρονη εποχή.

Στο κεφάλαιο «Μηχανική Μάθηση (Machine Learning)», παρουσιάζεται η ιδέα της Μηχανικής Μάθησης, τα είδη της καθώς και οι παράμετροι βάσει των οποίων κατηγοριοποιείται. Επιπρόσθετα, επισημαίνονται οι μεγαλύτερες προκλήσεις που αντιμετωπίζει, εξηγείται ο τρόπος δοκιμής και επικύρωσης συστημάτων Μηχανικής Μάθησης και τέλος, παρουσιάζεται η Ανάλυση Συναισθημάτων η οποία μπορεί να συνδεθεί άμεσα με τα συστήματα αυτά.

Το κεφάλαιο «Πρόβλεψη Τιμής Κρυπτονομισμάτων», αποτελεί το πρακτικό μέρος της διπλωματικής, όπου πραγματοποιείται η πρόβλεψη της τιμής του κρυπτονομίσματος «Bitcoin». Αρχικά παρουσιάζονται τα δεδομένα που θα χρησιμοποιηθούν για την πρόβλεψη καθώς και η προεπεξεργασία τους. Στη συνέχεια, λαμβάνει χώρα η δοκιμή «Breusch-Pagan» για να διαπιστωθεί εάν τα γραμμικά μοντέλα μπορούν να ερμηνεύσουν τα δεδομένα. Αφού διαπιστωθεί ότι δεν μπορούν, δοκιμάζεται κατά σειρά η απόδοση ενός μοντέλου Δικτύων Μακράς Βραχύχρονης Μνήμης (LSTM), ενός μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) και ενός Υβριδικού μοντέλου που χρησιμοποιεί Ανάλυση Συναισθημάτων και αποτελείται από Δίκτυα Αναδρομικής Πύλης (GRU) και Συνελικτικά Νευρωνικά Δίκτυα (CNN). Τέλος, πραγματοποιείται στατιστική σύγκριση μεταξύ του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) και του Υβριδικού μοντέλου (Hybrid Model).

# ΚΕΦΑΛΑΙΟ 2: ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ

## 2.1 Εισαγωγή

Η παρακάτω βιβλιογραφική επισκόπηση έχει σκοπό να παραθέσει μια ολοκληρωμένη εικόνα του τρόπου με τον οποίο πραγματοποιείται ο υπολογισμός της εύλογης αξίας και της τιμολόγησης των κρυπτονομισμάτων, εξετάζοντας ενδελεχώς μελέτες και μοντέλα που πραγματοποιούν αυτόν τον υπολογισμό κυρίως με την χρήση τεχνικών μηχανικής μάθησης.

## 2.2 Κρυπτονομίσματα και Αλυσίδες Συστοιχιών

Στις 31 Οκτωβρίου του 2008 δημοσιεύτηκε η εργασία «Bitcoin: A Peer-to-Peer Electronic Cash System» από τον συγγραφέα με το ψευδώνυμο Satoshi Nakamoto. Στην εργασία αυτή παρουσιάστηκε ένα πλήρως αποκεντροποιημένο σύστημα ηλεκτρονικών συναλλαγών όπου οι συναλλαγές γίνονται μεταξύ των χρηστών χωρίς να διέρχονται μέσα από κάποιο χρηματοπιστωτικό ίδρυμα. Για την υλοποίηση του εν λόγω συστήματος χρησιμοποιήθηκε η τεχνολογία των αλυσίδων συστοιχιών η οποία μέχρι τότε δεν είχε ευρεία εφαρμογή. Η συγκεκριμένη εργασία αποτέλεσε το έναυσμα για την δημιουργία και άλλων πλήρως αποκεντροποιημένων συστημάτων ηλεκτρονικών συναλλαγών καθώς και για την περαιτέρω διερεύνηση των τομέων στους οποίους μπορεί να έχει εφαρμογή η τεχνολογία των αλυσίδων συστοιχιών.

## 2.3 Υπολογισμός εύλογης αξίας Κρυπτονομισμάτων

Η περιπλοκότητα υπολογισμού της εύλογης αξίας των κρυπτονομισμάτων καθώς και οι τεράστιες μεταβολές της αξίας της, οδήγησε την επιστημονική κοινότητα να αναζητήσει πιο προηγμένες τεχνικές για τον υπολογισμό της. Αυτό γίνεται εύκολα αντιληπτό από τους Khedr, A.M. et al, (2021), όπου παρουσιάζουν περιεκτικές περιλήψεις μελετών στο πεδίο της πρόβλεψης της τιμής των κρυπτονομισμάτων από το 2010 έως και το 2020. Οι μελέτες αυτές χρησιμοποιούν για την πρόβλεψη της τιμής των κρυπτονομισμάτων παραδοσιακές στατιστικές τεχνικές πρόβλεψης καθώς και τεχνικές μηχανικής μάθησης. Από το άρθρο γίνεται σαφές ότι οι τεχνικές μηχανικής μάθησης υπερτερούν των στατιστικών, καθώς οι στατιστικές τεχνικές απαιτούνε πολλές στατιστικές παραδοχές οι οποίες δεν είναι ρεαλιστικές. Από τις τεχνικές μηχανικές μάθησης αυτή που υπερτερεί και προτείνεται για περαιτέρω διερεύνηση είναι τα Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM) και οι διάφορες παραλλαγές τους. Τέλος, από το άρθρο διαφαίνεται ότι ένας πολύ σημαντικός παράγοντας βελτίωσης της ακρίβειας των μοντέλων είναι η ανάλυση της κοινής γνώμης με τη χρήση ανάλυσης συναισθημάτων (sentiment analysis) καθώς η τιμή των κρυπτονομισμάτων φαίνεται να είναι άρρηκτα συνδεδεμένη με αυτή.

Οι τεχνικές μηχανικής μάθησης που εφαρμόζονται για τον υπολογισμό της εύλογης αξίας ανάλογα με τα δεδομένα που χρησιμοποιούν μπορούν να χωριστούν σε δύο κατηγορίες, σε αυτές που δεν χρησιμοποιούν ανάλυση συναισθημάτων (sentiment analysis) και σε αυτές που χρησιμοποιούν. Πιο αναλυτικά, οι τεχνικές που χρησιμοποιούν ανάλυση συναισθημάτων εκτός από τα κλασικά δεδομένα χρονοσειράς όπως είναι η τιμή κλεισίματος, η τιμή ανοίγματος

ο όγκος συναλλαγών κτλ., χρησιμοποιούν και δεδομένα για την ανάλυση της κοινής γνώμης όπως είναι οι δημοσιεύσεις χρηστών σε διάφορα δημόσια κοινωνικά δίκτυα.

### 2.3.1 Τεχνικές Μηχανικής Μάθησης χωρίς την χρήση Ανάλυσης Συναισθημάτων

Το 2019 δημοσιεύτηκε άρθρο από τους Ji, Kim et al, (2019), όπου πραγματοποιούν πρόβλεψη τιμής του κρυπτονομισμάτος Bitcoin με τη χρήση βαθιών νευρωνικών δικτύων (deep neural networks), δικτύων μακράς βραχύχρονης μνήμης (long short-term memory), συνελικτικών νευρωνικών δικτύων (convolutional neural networks), βαθιών υπολειμματικών δικτύων (deep residual networks), συνδυασμό συνελικτικών νευρωνικών δικτύων και δικτύων μακράς βραχύχρονης μνήμης, καθώς και με τη χρήση ενός μοντέλου συνόλου (ensemble model) το οποίο αποτελείται από ένα βαθύ νευρωνικό δίκτυο, ένα δίκτυο μακράς βραχύχρονης μνήμης και ένα συνελικτικό νευρωνικό δίκτυο. Για την εκπαίδευση και δοκιμή των μοντέλων χρησιμοποίησαν ημερήσιες τιμές των χαρακτηριστικών της αλυσίδας συστοιχιών του κρυπτονομίσματος Bitcoin από 29 Νοεμβρίου 2011 έως και 31 Δεκεμβρίου του 2018 (2590 ημέρες) όπως φαίνεται αναλυτικά στην Εικόνα 1. Από αυτά τα χαρακτηριστικά αποκλείστηκαν όλα όσα είχαν ελλιπή δεδομένα στο ανωτέρω χρονικό διάστημα, καθώς και όσα είχαν συντελεστή συσχέτισης κάτω από 0,75 και πάνω από 0,95.

| Feature | Description |
|---|---|
| avg-block-size | The 24 h average block size in MB. |
| blockchain-size | The total size of all block headers and transactions. |
| cost-per-trans | Miners revenue divided by the number of transactions. |
| cost-per-trans-pct | Miners revenue as percentage of the transaction volume. |
| difficulty | A relative measure of difficulty in finding a new block. |
| est-trans-vol | The estimated value of transactions on the Bitcoin blockchain in BTC. |
| est-trans-vol-usd | The estimated USD value of transactions. |
| hash-rate | The estimated number of tera hashes per second the Bitcoin network is performing. |
| market-cap | The total USD value of Bitcoin supply in circulation. |
| market-price | The average USD market price across major Bitcoin exchanges. |
| med-cfm-time | The median time for a transaction to be accepted into a mined block. |
| mempool-count | The number of transactions waiting to be confirmed. |
| mempool-growth | The rate of the memory pool (mempool) growth per second. |
| mempool-size | The aggregate size of transactions waiting to be confirmed. |
| miners-revenue | The total value of Coinbase block rewards and transaction fees paid to miners. |
| my-wallets | The total number of blockchain wallets created. |
| n-trans | The number of daily confirmed Bitcoin transactions. |
| n-trans-excl-100 | The total number of transactions per day excluding the chains longer than 100. |
| n-trans-excl-popular | The total number of transactions, excluding those involving any of the network's 100 most popular addresses. |
| n-trans-per-block | The average number of transactions per block. |
| n-trans-total | The total number of transactions. |
| n-unique-addr | The total number of unique addresses used on the Bitcoin blockchain. |
| output-val | The total value of all transaction outputs per day. |
| total-bitcoins | The total number of Bitcoins that have already been mined. |
| trade-vol | The total USD value of trading volume on major Bitcoin exchanges. |
| trans-fees | The total BTC value of all transaction fees paid to miners. |
| trans-fees-usd | The total USD value of all transaction fees paid to miners. |
| trans-per-sec | The number of Bitcoin transactions added to the mempool per second. |
| utxo-count | The number of unspent Bitcoin transactions outputs. |

**Εικόνα 1: Χαρακτηριστικά αλυσίδων συστοιχιών (Ji, Kim et al, 2019)**

Για την πραγματοποίηση των πειραμάτων, χρησιμοποιήθηκε το 80% των δεδομένων για την εκπαίδευση των μοντέλων και το υπόλοιπο 20% για την δοκιμή τους. Κατά την προεπεξεργασία των δεδομένων, αρχικά για 6 χαρακτηριστικά χρησιμοποιήθηκε η

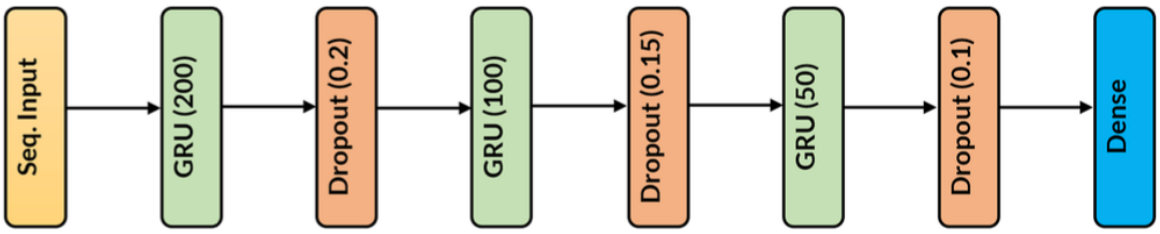
λογαριθμική τους τιμή λόγω της πολύ μεγάλης απόκλισης μεταξύ της ελάχιστης και της μέγιστης τιμής τους, και στη συνέχεια σε όλα τα χαρακτηριστικά εφαρμόστηκε έπειτα από πειραματική διαδικασία κανονικοποίηση πρώτης τιμής (first value-based normalization). Από τα πειράματα που πραγματοποιήθηκαν, αυτό που αξίζει ιδιαίτερης αναφοράς είναι η επίδραση του μεγέθους της ακολουθίας στην απόδοση των μοντέλων. Με την έννοια ακολουθία εννοείται ο αριθμός των προηγούμενων ημερών που λαμβάνεται υπόψιν για την πρόβλεψη της τιμής. Από τα αποτελέσματα που φαίνονται στην Εικόνα 2, προκύπτει ότι τα μοντέλα έχουν καλύτερη απόδοση για μέγεθος ακολουθίας 5. Ωστόσο, οι συγγραφείς προκρίνουν την επιλογή του μεγέθους 20 καθώς τα μοντέλα πρόβλεψης συμπεριφέρθηκαν σαν ένα μοντέλο μετατόπισης (shift model) και έτσι τα διακριτά χαρακτηριστικά τους φάνηκαν να εξαφανίζονται.

| Size ($m$) | DNN | LSTM | CNN | ResNet | CRNN | Ensemble | SVM |
|---|---|---|---|---|---|---|---|
| 5 | **3.61** | 3.79 | 4.27 | 4.95 | 4.12 | 4.02 | 4.75 |
| 10 | 4.00 | **3.96** | 4.88 | 7.12 | 4.26 | 4.80 | 4.88 |
| 20 | 4.81 | **4.46** | 7.93 | 8.96 | 5.90 | 6.19 | 5.19 |
| 50 | 10.88 | 6.68 | 20.00 | 16.91 | 10.11 | 11.45 | **6.34** |
| 100 | 21.44 | 27.75 | 115.22 | 52.10 | 39.13 | 48.87 | **12.77** |

**Εικόνα 2: Συγκριτικά αποτελέσματα μοντέλων πρόβλεψης (MAPE, %) (Ji, Kim et al, 2019)**

Τέλος, οι συγγραφείς λαμβάνοντας υπόψιν το σύνολο των πραγματοποιθέντων πειραμάτων κατέληξαν στο γεγονός ότι τα δίκτυα μακράς βραχύχρονης μνήμης (long short-term memory) υπερτερούν ελαφρώς σε σχέση με τα υπόλοιπα μοντέλα πρόβλεψης.

Στο άρθρο που δημοσιεύτηκε στο περιοδικό Springer (Patra & Mohanty 2022), προτείνεται η πρόβλεψη τιμών κρυπτονομισμάτων με τη χρήση ενός μοντέλου αποτελούμενου από Μονάδες Αναδρομικής Πύλης (Gated Recurrent Unit – GRU). Πιο αναλυτικά, το μοντέλο αποτελείται από τρία στρώματα Μονάδων Αναδρομικής Πύλης με αριθμό μονάδων 200, 100, και 50 αντίστοιχα. Ανάμεσα από τα στρώματα Μονάδων Αναδρομικής Πύλης, υπάρχουν τρία στρώματα Αποκλεισμού Δικτύου (Dropout Layer) με ποσοστά αποκλεισμού 20%, 15% και 10% αντίστοιχα. Στο τέλος υπάρχει ένα Πυκνό στρώμα (Dense Layer) το οποίο αποτελεί και την έξοδο του μοντέλου. Στην Εικόνα 3 φαίνεται αναλυτικά η αρχιτεκτονική του περιγραφόμενου μοντέλου.




**Εικόνα 3: Η αρχιτεκτονική του προτεινόμενου μοντέλου. (Patra & Mohanty 2022)**

Για την δημιουργία του μοντέλου χρησιμοποιήθηκαν 6 χαρακτηριστικά, η τιμή ανοίγματος της ημέρας, η τιμή κλεισίματος της ημέρας, η υψηλότερη τιμή της ημέρας, ο όγκος των κρυπτονομισμάτων και το ποσό σε δολάρια που συναλλάχτηκε την κάθε ημέρα. Για κάθε κρυπτονόμισμα χρησιμοποιήθηκαν διαφορετικές ημερομηνίες δεδομένων για εκπαίδευση και δοκιμή. Πιο αναλυτικά, για το «Bitcoin» χρησιμοποιήθηκαν δεδομένα από τις 7 Ιανουαρίου

του 2014 έως και τις 3 Ιουνίου του 2021, για το «Ethereum» από τις 9 Φεβρουαρίου του 2016 έως και τις 15 Απριλίου του 2020 και για το «Dogecoin» από τις 17 Σεπτεμβρίου του 2014 έως και τις 6 Ιουλίου του 2021. Κατά την προεπεξεργασία των δεδομένων, αρχικά στα δεδομένα εφαρμόστηκε η τεχνική της κανονικοποίησης ελαχίστου-μεγίστου (min-max normalization), και διαγράφηκαν τα δεδομένα που δεν είχαν τιμή. Στη συνέχεια, από αυτά τα δεδομένα και σύμφωνα με τον ψευδοκώδικα, για την δοκιμή του μοντέλου επιλέχτηκαν οι τιμές των τελευταίων 21 ημερών ενώ τα υπόλοιπα δεδομένα χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου. Τέλος, ως περίοδος αναδρομής επιλέχτηκε η μία ημέρα. Από τα αποτελέσματα του μοντέλου που φαίνονται στην Εικόνα 4, παρατηρείται ακραία καλή απόδοση του μοντέλου, η οποία δεν ήταν δυνατόν να επαληθευτεί πειραματικά. Παρά όμως την αδυναμία επαλήθευσης των ανωτέρω αποτελεσμάτων, το μοντέλο έδειξε πολύ καλή απόδοση καθώς στις δοκιμές που πραγματοποιήθηκαν ακόμη και με πολύ μεγαλύτερα ποσοστά δοκιμής (30%) το μοντέλο πετύχαινε απόδοση MAPE περίπου στο 0.05, φανερώνοντας την εξαιρετική απόδοση των Μονάδων Αναδρομικής Πύλης στην πρόβλεψη της τιμής των κρυπτονομισμάτων.

| Cryptocurrency | Performance measure | Proposed model |
|---|---|---|
| Bitcoin | MSE | **63,307.3120** |
| | RMSE | **251.6094** |
| | MAE | **164.4882** |
| | MAPE | **0.0031** |
| | Accuracy | **99.69** |
| Ethereum | MSE | **0.3663** |
| | RMSE | **0.6052** |
| | MAE | **0.5121** |
| | MAPE | **0.0037** |
| | Accuracy | **99.63** |
| Dogecoin | MSE | **9.5562 e-07** |
| | RMSE | **9.7758 e-04** |
| | MAE | **7.6192 e-04** |
| | MAPE | **2.8281 e-03** |
| | Accuracy | **99.99** |

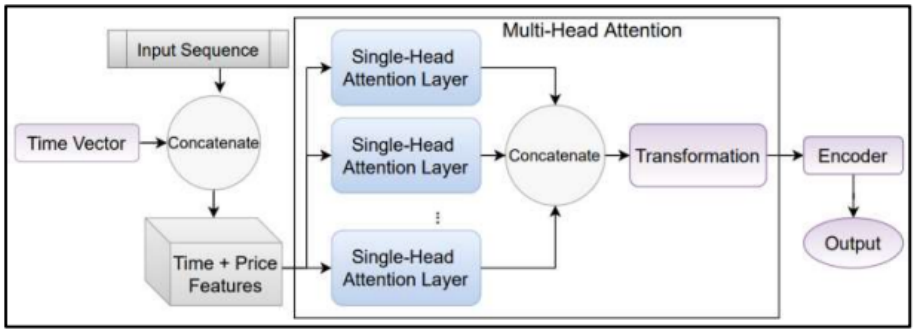
**Εικόνα 4: Τα αποτελέσματα του προτεινόμενου μοντέλου. (Patra & Mohanty 2022)**

Σύμφωνα με το άρθρο που δημοσίευσαν οι Encean & Zinca (2023), χρησιμοποιήθηκαν Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM) και Δίκτυα Αναδρομικής Πύλης (Gated Recurrent Unit – GRU) για την πρόβλεψη τιμής των κρυπτονομισμάτων «Riple (XRP)», «DOGEcoin» και «OMG Network». Για την εκπαίδευση και δομική των ανωτέρω μοντέλων χρησιμοποιήθηκαν ως δεδομένα οι τιμές κλεισίματος (Close price) των κρυπτονομισμάτων για την χρονική περίοδο από 8 Απριλίου του 2020 έως και 10 Μαρτίου του 2022. Τα δεδομένα αυτά κανονικοποιήθηκαν έτσι ώστε να έχουν τιμές από 0 έως 1. Οι δοκιμές που πραγματοποιήθηκαν σε μοντέλα LSTM και GRU αποτελούνταν από ένα στρώμα νευρώνων, ο αριθμός των οποίων διέφερε από δοκιμή σε δοκιμή. Επίσης, από δοκιμή σε δοκιμή διέφερε και το ποσοστό των δεδομένων που χρησιμοποιήθηκε για την δοκιμή των μοντέλων. Από τα αποτελέσματα των μοντέλων που φαίνονται στην Εικόνα 5, μπορούμε να συμπεράνουμε ότι τα μοντέλα έχουν πολύ καλή απόδοση αν και ως είσοδο είχαν μόνο μία μεταβλητή, την τιμή κλεισίματος.

| | DOGEcoin | | XRP | | OMG Network | |
|---|---|---|---|---|---|---|
| Model | **LSTM** | GRU | LSTM | **GRU** | **LSTM** | GRU |
| Neurons | **100** | 100 | 100 | **100** | **100** | 4 |
| Epochs | **100** | 100 | 20 | **100** | **20** | 5 |
| Test Base % | **30** | 20 | 20 | **20** | **10** | 15 |
| Price Prediction | **0.1185** | 0.1280 | 0.7883 | **0.7805** | **4.0389** | 3.7005 |
| Real price | **0.1156** | | **0.8021** | | **4.0382** | |
| MAPE % | **3.4783** | 3.8716 | 3.5464 | **3.4571** | **4.0111** | 4.7015 |

**Εικόνα 5: Τα καλύτερα αποτελέσματα των μοντέλων ανά κρυπτονόμισμα. (Encean & Zinca 2023)**

Στο άρθρο που δημοσίευσαν οι Sridhar & Sanagavarapu (2021), χρησιμοποιούν ένα μοντέλο μεταμορφωτή αυτοπροσοχής πολλαπλών κεφαλών (Multi-Head Self-Attention Transformer) για την πρόβλεψη τιμής του κρυπτονομίσματος «Dogecoin». Στην Εικόνα 6 φαίνεται η αρχιτεκτονική του μοντέλου.




**Εικόνα 6: Η αρχιτεκτονική του μοντέλου μεταμορφωτή (Transformer). (Sridhar & Sanagavarapu 2021)**

Για την δοκιμή και εκπαίδευση του μοντέλου χρησιμοποίησαν τις τιμές ανά ώρα των χαρακτηριστικών τιμής ανοίγματος (Open), υψηλότερης τιμή (High), τιμής κλεισίματος (Close) και όγκου συναλλαγών του «Dogecoin» (Volume of Dogecoin traded) για το χρονικό διάστημα από 5 Ιουλίου του 2019 έως και 28 Απριλίου του 2021. Στα δεδομένα αυτά έγινε ενσωμάτωση του χρόνου (Time Embedding) με τη χρήση της συνάρτησης που φαίνεται στην παρακάτω εικόνα.

$$t2v(\tau)[i] = \begin{cases} \omega_i\tau + \varphi_i, & i = 0 \\ F(\omega_i\tau + \varphi_i), & 1 \le i \le k \end{cases}$$

**Εικόνα 7: Η συνάρτηση ενσωμάτωσης χρόνου Time2Vec. (Sridhar & Sanagavarapu 2021)**

Στη συνέχεια στα δεδομένα δημιουργήθηκε ολισθαίνον παράθυρο (Sliding Window) 50 ωρών και τα δεδομένα χωρίστηκαν σε ποσοστό 80% για εκπαίδευση και 20% για δοκιμή. Από τα αποτελέσματα της απόδοσης του μοντέλου που φαίνονται στην Εικόνα 8, προκύπτει ότι το μοντέλο ανταποκρίνεται πολύ καλά, δείχνοντας στην επιστημονική κοινότητα ότι τα μοντέλα των Μεταμορφωτών (Transformers) εκτός από την ανάλυση κειμένου μπορούν να

χρησιμοποιηθούν επιτυχώς και στην πρόβλεψη τιμών. Αυτό έχει πολύ μεγάλο ενδιαφέρον καθώς σε θεωρητικό επίπεδο οι Μεταμορφωτές (Transformers) μπορούν να «αντιληφθούν» πολύ πιο μακροχρόνιες σχέσεις μεταξύ των μεταβλητών από τα Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM) και τα Δίκτυα Αναδρομικής Πύλης (Gated Recurrent Unit – GRU).

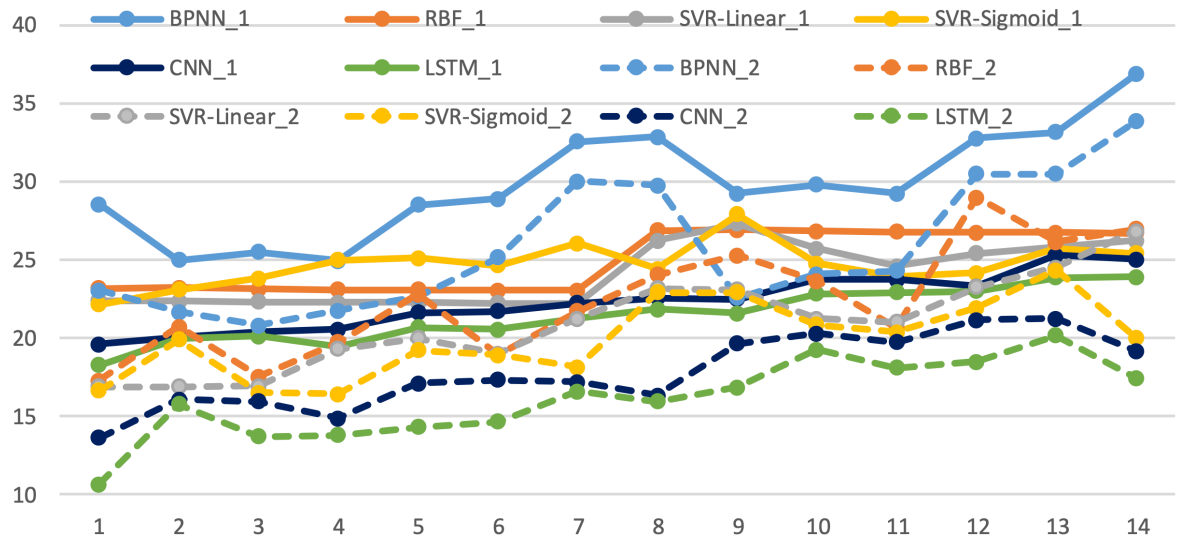
| **Model** | **Evaluation Metric** | | | |
|---|---|---|---|---|
| | ***RMSE*** | ***MAE*** | ***Predictive R-squared value*** | ***Accuracy*** |
| Multi-Head Self-Attention Transformer | 25.873 | 3.217 | 86.17 | 98.47 |

**Εικόνα 8: Η απόδοση του μοντέλου Μεταμορφωτή (Transformer). (Sridhar & Sanagavarapu 2021)**

### 2.3.2 Τεχνικές Μηχανικής Μάθησης με την χρήση Ανάλυσης Συναισθημάτων

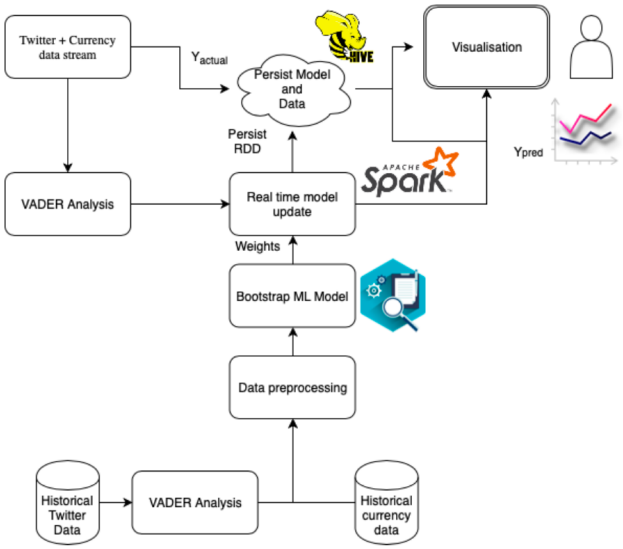
Το 2020 δημοσιεύτηκε το άρθρο από τους Wang & Chen (2020), στο οποίο παρουσιάζουν ένα μοντέλο πρόβλεψης τιμής κρυπτονομισμάτων το οποίο έχει ως είσοδο δύο τύπους δεδομένων, τα δεδομένα συναλλαγών στην αγορά και τα σχόλια των χρηστών σε μέσα κοινωνικής δικτύωσης. Τα δεδομένα συναλλαγών περιέχουν τον μέσο όρο μεταξύ τριών διαφορετικών αγορών της υψηλότερης τιμής, της χαμηλότερης τιμής, της τιμής ανοίγματος, της τιμής κλεισίματος και του όγκου των συναλλαγών. Επιπρόσθετα, περιέχουν και το επιτόκιο προσαύξησης (premium rate) μεταξύ της αγοράς χρηματιστηριακών συναλλαγών (exchange trading market) και της αγοράς συναλλάγματος νομισμάτων (fiat trading market), καθώς και μεταξύ της αγοράς διαπραγμάτευσης συμβολαίων (contract trading market) και της αγοράς χρηματιστηριακών συναλλαγών (exchange trading market). Τα δεδομένα των χρηστών από τα μέσα κοινωνικής δικτύωσης προκειμένου να μετατραπούν σε αριθμητά δεδομένα χρησιμοποιήθηκε ένα λεξικό συναισθήματος (sentiment lexicon). Για την εκπαίδευση και δοκιμή του μοντέλου χρησιμοποιήθηκαν δεδομένα από 1 Ιανουαρίου του 2019 έως και 31 Μαρτίου του 2019. Κατά την πραγματοποίηση των δοκιμών χρησιμοποιήθηκαν δύο μοντέλα, ένα μοντέλο που να μην χρησιμοποιεί ανάλυση συναισθημάτων (sentiment analysis) δηλαδή που να χρησιμοποιεί ως είσοδο μόνο τα δεδομένα της αγοράς και ένα μοντέλο που να χρησιμοποιεί ανάλυση συναισθημάτων δηλαδή να έχει ως είσοδο τόσο των δεδομένων της αγοράς όσο και των σχολίων των χρηστών. Κατά την πραγματοποίηση των δοκιμών χρησιμοποιήθηκαν διάφορες τεχνικές μηχανικής μάθησης καθώς και ολισθαίνον παράθυρο (Sliding Window) μεταξύ 1 – 7 ημερών. Οι τεχνικές που χρησιμοποιήθηκαν ήταν ένα Νευρωνικό Δίκτυο Οπισθοδιάδοσης (Propagation Neural Network – BPNN), ένα Ακτινικό Δίκτυο Βάσης (Radial Basis Function – RBF), Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machine – SVM), Νευρωνικά Δίκτυα Συνέλιξης (Convolutional Neural Networks – CNN) και Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM). Από τα αποτελέσματα που φαίνονται στην Εικόνα 9 προκύπτει ότι ανεξαρτήτως μεγέθους του ολισθαίνοντος παραθύρου, τα Δίκτυα Μακράς Βραχύχρονης Μνήμης χρησιμοποιώντας ανάλυση συναισθημάτων έχουνε την καλύτερη απόδοση. Παράλληλα, φαίνεται ξεκάθαρα από το διάγραμμα της Εικόνας 9 ότι τα Δίκτυα Μακράς Βραχύχρονης Μνήμης χωρίς την χρήση ανάλυσης συναισθημάτων αποδίδουν χειρότερα σε σχέση με Νευρωνικά Δίκτυα Συνέλιξης που χρησιμοποιούν. Αυτό φανερώνει τη μεγάλη συμβολή της ανάλυσης συναισθημάτων στην απόδοση του μοντέλου καθώς κατά κανόνα τα Δίκτυα Μακράς Βραχύχρονης Μνήμης

αποδίδουν πολύ καλύτερα στην πρόβλεψη της τιμής των κρυπτονομισμάτων από τα Νευρωνικά Δίκτυα Συνέλιξης.




**Εικόνα 9: RMSE ανά μέγεθος ολισθαίνοντος παραθύρου. (Wang & Chen 2020)**

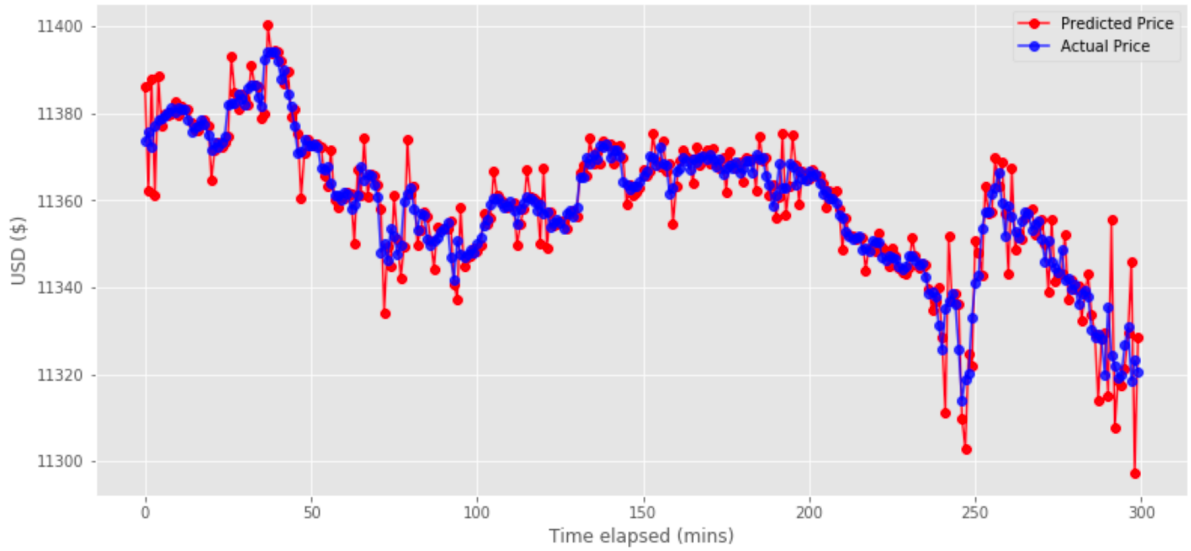
Την ίδια χρονιά, δημοσιεύτηκε επίσης ένα άρθρο από τους Mohaparta, et al, (2020), όπου παρουσίασαν μια πλατφόρμα πρόβλεψης των τιμών των κρυπτονομισμάτων σε πραγματικό χρόνο πραγματοποιώντας ανάλυση συναισθημάτων (sentiment analysis) σε δεδομένα προερχόμενα από το Twitter. Η αρχιτεκτονική της πλατφόρμας φαίνεται στην Εικόνα 10.




**Εικόνα 10: Η αρχιτεκτονική της πλατφόρμας KryptoOracle. (Mohaparta, et al, 2020)**

Κατά την λειτουργία της πλατφόρμας, αρχικά, αντλείται ανά λεπτό η τιμή ανοίγματος, η τιμή κλεισίματος, η υψηλότερη τιμή, η χαμηλότερη τιμή, η τιμή κλεισίματος του προηγούμενου λεπτού, ο κινητός μέσος όρος των τελευταίων 100 λεπτών καθώς και ο κινητός μέσος όρος των τελευταίων 100 βαθμολογιών συναισθηματικής ανάλυσης. Η βαθμολογία συναισθηματικής ανάλυσης προκύπτει με την εφαρμογή σε κάθε «tweet» που αντλείται, της βασισμένης σε λεξικό (lexicon-based) τεχνικής που ονομάζεται VADER (Valence Aware Dictionary and sentiment Reasoner) και στη συνέχεια προσδίδοντας της βάρος ανάλογα με τους ακόλουθους, τα «like» και τα «retweet». Αφού συγκεντρωθούν και υπολογιστούν τα

παραπάνω δεδομένα, στη συνέχεια εισέρχονται στο μοντέλο μηχανικής μάθησης XGBoost (eXtreme Gradient Boosting) έτσι ώστε να πραγματοποιηθεί η πρόβλεψη τιμής για το επόμενο λεπτό. Από τα αποτελέσματα που φαίνονται στην Εικόνα 11 προκύπτει ότι η πλατφόρμα ανταποκρίνεται μέτρια στην πρόβλεψη τιμής σε σχέση με τα προηγούμενα μοντέλα που εξετάστηκαν στην βιβλιογραφική ανασκόπηση. Αυτό ίσως να οφείλεται στη μη επιλογή ενός πιο αποτελεσματικού μοντέλου μηχανικής μάθησης όπως είναι τα Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM) και τα Δίκτυα Αναδρομικής Πύλης (Gated Recurrent Unit – GRU).




**Εικόνα 11: Δοκιμή πλατφόρμας KryptoOracle. (Mohaparta, et al, 2020)**

Οι Zamani, Yan, et al, (2022) στο άρθρο που δημοσίευσαν, πραγματοποίησαν πρόβλεψη της τιμής των κρυπτονομισμάτων «Bitcoin» και «Ethereum» χρησιμοποιώντας ένα μοντέλο Αμφίδρομης Αναδρομικής Πύλης (Bidirectional Gated Recurrent Unit – BiGRU) το οποίο δέχεται ως είσοδο τη βαθμολογία συναισθήματος και δεδομένα τιμών κρυπτονομισμάτων. Τα δεδομένα τιμών κρυπτονομισμάτων που χρησιμοποιούνται είναι η τιμή ανοίγματος, η υψηλότερη τιμή, η χαμηλότερη τιμή, η τιμή κλεισίματος και ο όγκος των συναλλαγών. Για τον υπολογισμό της βαθμολογίας συναισθημάτων χρησιμοποιούνται οι τίτλοι των Αγγλικών και Μαλαισιανών ειδήσεων. Πιο αναλυτικά, σε κάθε τίτλο δόθηκε από τρεις έμπειρους σχολιαστές μία βαθμολογία μεταξύ των τιμών -1 και 1 ανάλογα με το πόσο αρνητικός ή θετικός αντίστοιχα ήταν ο τίτλος. Από αυτές τις βαθμολογίες δημιουργήθηκαν τρία χαρακτηριστικά, ο μέσος όρος των θετικών βαθμολογιών κάθε ημέρας, ο μέσος όρος των αρνητικών βαθμολογιών κάθε ημέρας και ο συνολικός μέσος όρος των βαθμολογιών κάθε ημέρας. Τα παραπάνω χαρακτηριστικά αντλήθηκαν για το χρονικό διάστημα από 1 Ιανουαρίου του 2021 έως και 31 Δεκεμβρίου του 2021. Στη συνέχεια κανονικοποιήθηκαν λαμβάνοντας τιμές από 0 έως 1. Για την εκπαίδευση και δοκιμή του μοντέλου τα δεδομένα χωρίστηκαν τυχαία σε 80% δεδομένα εκπαίδευσης και 20% δεδομένα δοκιμής. Για την δοκιμή του μοντέλου εφαρμόστηκε 5 φορές διασταυρούμενη επικύρωση (5-fold cross-validation), και έγινε δοκιμή διάφορων συνδυασμών των διαθέσιμων χαρακτηριστικών. Από τα αποτελέσματα που φαίνονται στην Εικόνα 12, προκύπτει ότι η προσθήκη ανάλυσης συναισθημάτων βελτιώνει το μοντέλο αισθητά ακόμη και χρησιμοποιώντας τους Μαλαισιανούς τίτλους ειδήσεων.

| # | HISTORICAL PRICES | | | | | AVERAGE SENTIMENT SCORES | | | EVALUATIONS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Open | High | Low | Close | Volume | Overall | Pos | Neg | RMSE | MAE | Adj $R^2$ |
| ENGLISH | | | | | | | | | | | |
| 1 | x | x | x | x | x | x | x | x | 6481.10 | 4948.84 | 0.557 |
| 2 | x | x | x | x | x | x | | | 6627.19 | 4861.41 | 0.537 |
| 3 | x | x | x | x | x | | x | x | 6661.64 | 4922.41 | 0.532 |
| 4 | | | | **x** | | | **x** | **x** | **6472.38** | **4830.07** | **0.558** |
| 5 | x | x | x | x | x | | | | 7266.14 | 5597.19 | 0.443 |
| MALAY | | | | | | | | | | | |
| 1 | x | x | x | x | x | x | x | x | 7311.34 | 5363.88 | 0.436 |
| 2 | x | x | x | x | x | x | | | 6837.61 | 5045.44 | 0.507 |
| 3 | x | x | x | x | x | | x | x | 6941.56 | 5106.66 | 0.492 |
| 4 | | | | **x** | **x** | **x** | | | **6624.10** | **4884.40** | **0.537** |
| 5 | x | x | x | x | x | | | | 7521.18 | 5773.90 | 0.403 |

**Εικόνα 12: Αποτελέσματα δοκιμών μοντέλου BiGRU. (Zamani, Yan, et al, 2022)**

## 2.4 Πεδία διερεύνησης στον υπολογισμό της εύλογης αξίας Κρυπτονομισμάτων

Έπειτα από τη δημοσίευση του άρθρου από τους Vaswani, et al, (2017), επήλθε η επανάσταση στον τομέα της επεξεργασίας της φυσικής γλώσσας (natural language processing) καθώς οι Μεταμορφωτές (Transformers) μπορούν να χειρίζονται καλύτερα τις εξαρτήσεις μεγάλης απόστασης χρησιμοποιώντας τον μηχανισμό αυτοπροσοχής (self-attention mechanism), που τους επιτρέπει να συλλαμβάνουν τις εξαρτήσεις μεταξύ απομακρυσμένων λέξεων σε μια πρόταση, καθιστώντας τους καλύτερους στην κατανόηση της συνολικής σημασίας μιας πρότασης. Αυτή η επανάσταση στην επεξεργασία της φυσικής γλώσσας μπορεί να προσφέρει νέους ορίζοντες στην χρήση της ανάλυσης συναισθημάτων δίνοντας την δυνατότητα σε ήδη αποτελεσματικά μοντέλα πρόβλεψης τιμής όπως είναι τα Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM) και τα Δίκτυα Αναδρομικής Πύλης (Gated Recurrent Unit – GRU) να γίνουν ακόμα πιο αποτελεσματικά. Παράλληλα, ένα μεγάλο θέμα που χρήζει διερεύνησης είναι και ο τρόπος σύνδεσης του μοντέλου που θα πραγματοποιεί ανάλυση συναισθημάτων με το μοντέλο πρόβλεψης τιμής, καθώς οι μεταβλητές αυτών των μοντέλων διαφέρουν τόσο σε ποσότητα όσο και στον τρόπο που περιέχουν την πληροφορία.

## 2.5 Συμπεράσματα στον υπολογισμό της εύλογης αξίας

Ο υπολογισμός της εύλογης αξίας των κρυπτονομισμάτων αποτελεί ένα πολύπλοκο πρόβλημα το οποίο όπως έχει αποδειχθεί δεν μπορεί να λυθεί αποτελεσματικά με παραδοσιακές στατιστικές τεχνικές. Έτσι η ερευνητική κοινότητα έκανε στροφή προς την χρήση τεχνικών μηχανικής μάθησης και ιδιαίτερα σε τεχνικές οι οποίες μπορούν να ανταποκριθούν καλύτερα σε δεδομένα χρονοσειρών. Από τα παραπάνω εξεταζόμενα άρθρα προκύπτει ξεκάθαρα ότι οι πιο αποτελεσματικές τεχνικές για την πρόβλεψη της τιμής των κρυπτονομισμάτων είναι τα Δίκτυα Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM) και τα Δίκτυα Αναδρομικής Πύλης (Gated Recurrent Unit – GRU). Αν και αυτές οι δύο τεχνικές έχουν πολύ καλά αποτελέσματα δεν μπορούμε να πούμε ότι είναι αρκετές από μόνες τους για την αποτελεσματική πρόβλεψη τιμής. Για την βελτίωση της απόδοσης αυτών των τεχνικών δημοσιεύτηκαν άρθρα στα οποία δοκιμάστηκε η προσθήκη δεδομένων στα χαρακτηριστικά εισόδου των μοντέλων, τα οποία προήλθαν μέσω της ανάλυσης συναισθημάτων. Αυτή η προσθήκη αποδείχτηκε, με βάση τα δημοσιευμένα άρθρα, ότι μπορεί να βελτιώσει αισθητά την απόδοση των μοντέλων. Ωστόσο σε κανένα από αυτά τα άρθρα δεν χρησιμοποιήθηκαν για την ανάλυση συναισθημάτων Μεταμορφωτές (Transformers) οι οποίοι έφεραν την

επανάσταση στην επεξεργασία της φυσικής γλώσσας. Έτσι, στο πλαίσιο αυτής της διπλωματικής θα διερευνηθεί η δυνατότητα δημιουργίας ενός μοντέλου αποτελούμενου από Δίκτυα Μακράς Βραχύχρονης Μνήμης ή Δίκτυα Αναδρομικής Πύλης που στα χαρακτηριστικά εισόδου τους θα προστεθούν χαρακτηριστικά προερχόμενα από ανάλυση συναισθημάτων με την χρήση Μεταμορφωτών.

# ΚΕΦΑΛΑΙΟ 3: ΤΕΧΝΟΛΟΓΙΑ ΑΛΥΣΙΔΑΣ ΣΥΣΤΟΙΧΙΩΝ (BLOCKCHAIN TECHNOLOGY)

## 3.1 Εισαγωγή

Μία αλυσίδα συστοιχιών (blockchain) αποτελεί ένα κατανεμημένο ψηφιακό καθολικό (ledger) που βασίζεται σε τεχνικές κρυπτογράφησης και συναίνεσης για την καταγραφή και διασφάλιση των δεδομένων (Casey, 2018). Τα βασικά χαρακτηριστικά της αλυσίδας συστοιχιών είναι η αμεταβλητότητα, η δυνατότητα αποκεντροποίησης και η επεκτασιμότητα. Η αμεταβλητότητα επιτυγχάνεται με τη συνεχή αύξηση της αλυσίδας και την χρήση ψηφιακής υπογραφής στην καταγραφή των δεδομένων στο καθολικό (Casey, 2018). Η αποκεντροποίηση κατορθώνεται με την χρησιμοποίηση κατάλληλων αλγορίθμων συναίνεσης και επικύρωσης. Ενώ τέλος, η επεκτασιμότητα επιτυγχάνεται όταν υπάρχει πλήρης αποκεντροποίηση της αλυσίδας συστοιχιών λόγω της ίδιας της αρχιτεκτονικής της.

### 3.1.1 Ιστορική Αναδρομή

Η βασική ιδέα πίσω από την αλυσίδα συστοιχιών εμφανίστηκε το 1991 όταν μία υπογεγραμμένη αλυσίδα από πληροφορίες χρησιμοποιήθηκε ως ηλεκτρονικό καθολικό για ψηφιακά υπογεγραμμένα έγγραφα με τέτοιο τρόπο έτσι ώστε να είναι φανερό ότι δεν έχει τροποποιηθεί κανένα έγγραφο από την συλλογή (Yaga, et al, 2018). Ο στόχος ήταν να υπάρχει μία συλλογή εγγράφων όπου δεν θα ήταν δυνατή η αναδρομική προσθήκη και επεξεργασία εγγράφων.

Η αλυσίδα συστοιχιών ξεκίνησε να γίνεται ευρέως γνωστή από το 2008 με την χρησιμοποίηση της για πρώτη φορά σε ψηφιακά μετρητά. Πιο συγκεκριμένα, το 2008 δημοσιεύτηκε από τον συγγραφέα με το ψευδώνυμο Satoshi Nakamoto η εργασία με τίτλο ‘Bitcoin: A Peer-to-Peer Electronic Cash System’ (Nakamoto, 2008), η οποία παρουσίασε ένα ολοκληρωμένο, αποκεντροποιημένο σύστημα ψηφιακών μετρητών βασισμένο εκτός των άλλων και στην τεχνολογία αλυσίδων συστοιχιών δημιουργώντας έτσι το πρώτο κρυπτονόμισμα.

Σήμερα, η τεχνολογία των αλυσίδων συστοιχιών εκτός από τα κρυπτονομίσματα βρίσκει εφαρμογή ή γίνεται προσπάθεια να βρεθεί τρόπος εφαρμογής της σχεδόν σε όλους τους τομείς που είναι απαραίτητη η αποθήκευση ευαίσθητων πληροφοριών. Τέτοιοι τομείς για παράδειγμα, είναι οι οικονομικές και κοινωνικές υπηρεσίες, η διαχείριση κινδύνου και οι εγκαταστάσεις υγείας (Monrat, et al, 2019).

## 3.2 Κατηγοριοποίηση Αλυσίδων Συστοιχιών (Blockchain Categorization)

Τα δίκτυα αλυσίδων συστοιχιών μπορούν να κατηγοριοποιηθούν βάσει του μοντέλου άδειας (permission model) που χρησιμοποιούν, το οποίο καθορίζει και ποιος μπορεί να τα συντηρεί, π.χ. προσθέτοντας μπλοκ (Yaga, et al, 2018). Εάν μπορεί ο οποιασδήποτε να δημοσιεύσει ένα νέο μπλοκ, τότε το δίκτυο είναι χωρίς άδεια (permissionless) (Yaga, et al, 2018). Αντίθετα, όταν μπορεί μόνο μια συγκεκριμένη ομάδα χρηστών να δημοσιεύσει ένα νέο μπλοκ, τότε αυτό είναι με άδεια (permissioned) (Yaga, et al, 2018).

### 3.2.1 Αλυσίδες συστοιχιών χωρίς άδεια (Permissionless Blockchain)

Τα δίκτυα αλυσίδων συστοιχιών χωρίς άδεια αποτελούν αποκεντροποιημένες πλατφόρμες καθολικού στις οποίες μπορεί ο κάθε χρήστης να δημοσιεύει νέα μπλοκ, χωρίς να χρειάζεται την άδεια κάποιας αρχής (Yaga, et al, 2018). Οι πλατφόρμες αυτές είναι συνήθως ανοιχτού κώδικα και διαθέσιμες για λήψη από οποιονδήποτε το επιθυμεί (Yaga, et al, 2018). Έτσι εφόσον ο κάθε χρήστης μπορεί να διαβάσει και να γράψει στο καθολικό δημιουργείται το πρόβλημα ότι μπορεί κάποιος κακόβουλος χρήστης να προσπαθήσει να επιτεθεί στο δίκτυο δημοσιεύοντας κακόβουλα μπλοκ, δηλαδή μπλοκ που περιέχουν κακόβουλες συναλλαγές. Για να αντιμετωπιστεί αυτό το πρόβλημα, τα δίκτυα συστοιχιών χωρίς άδεια χρησιμοποιούν ένα σύστημα συναίνεσης (consensus system), το οποίο απαιτεί από τους χρήστες να αναλώσουν ή να διατηρήσουν σημαντικούς πόρους (π.χ. ηλεκτρικό ρεύμα).

### 3.2.2 Αλυσίδες συστοιχιών με άδεια (Permissioned Blockchain)

Τα δίκτυα αλυσίδων συστοιχιών με άδεια είναι αυτά όπου οι χρήστες για να δημοσιεύσουν πρέπει να εξουσιοδοτηθούν από κάποια αρχή η οποία μπορεί να είναι είτε κεντροποιημένη είτε αποκεντροποιημένη (Yaga, et al, 2018). Σε αυτά τα δίκτυα η αρχή καθορίζει εξολοκλήρου τον τρόπο με τον οποίο οι χρήστες αλληλοεπιδρούν με το καθολικό. Για παράδειγμα μπορεί να δίνουν την δυνατότητα σε όλους τους χρήστες του δικτύου να διαβάζουν τα δεδομένα του καθολικού αλλά να έχουν εξουσιοδοτήσει μόνο ελάχιστους να μπορούν να προσθέτουν συναλλαγές. Με αυτόν τον τρόπο μπορούν να εξοικονομηθούν σημαντικοί πόροι (π.χ. ηλεκτρικό ρεύμα) σε σχέση με τα δίκτυα συστοιχιών χωρίς άδεια καθώς συνήθως δεν υπάρχει η ανάγκη για κάποιο σύστημα συναίνεσης, ή ακόμη και να υπάρχει δεν απαιτεί σημαντική κατανάλωση πόρων καθώς υπάρχει στο δίκτυο ένα επίπεδο εμπιστοσύνης.

## 3.3 Η Αρχιτεκτονική των Αλυσίδων Συστοιχιών

Στα συστήματα αλυσίδων συστοιχιών χρησιμοποιούνται βασικοί μηχανισμοί από τον τομέα της Πληροφορικής, οι οποίοι χρησιμοποιούνται ευρέως και επιτυχώς εδώ και αρκετά χρόνια σε σημαντικές εφαρμογές της Πληροφορικής, σε συνδυασμό με τεχνικές καταγραφής αρχείων. Πιο συγκεκριμένα, οι αλυσίδες συστοιχιών χρησιμοποιούν συνδεδεμένες λίστες, κατανεμημένα δίκτυα, βασικές τεχνικές κρυπτογράφησης όπως είναι οι συναρτήσεις κατακερματισμού και τα μη συμμετρικά κρυπτοσυστήματα καθώς και καθολικά που επιτρέπουν μόνο την προσθήκη (Yaga, et al, 2018).

### 3.3.1 Συναρτήσεις Κατακερματισμού (Hash Functions)

Ένα από τα σημαντικότερα και αναπόσπαστα εργαλεία που χρησιμοποιούνται σε πληθώρα λειτουργιών από τις αλυσίδες συστοιχιών είναι οι συναρτήσεις κατακερματισμού (Yaga, et al, 2018). Οι συναρτήσεις κατακερματισμού υπολογίζουν μία σύνοψη μηνύματος (message digest) σταθερού μεγέθους από μία ακολουθία δυαδικής εισόδου με αυθαίρετο μέγεθος (Tanenbaum & Steen, 2002). Αυτό σημαίνει ότι η είσοδος μπορεί να αποτελείται από δεδομένα οποιασδήποτε μορφής αρκεί να είναι δυνατή η μετατροπή τους σε δυαδική ακολουθία. Δηλαδή μπορεί να είναι για παράδειγμα ένα αρχείο, μερικές γραμμές κειμένου ή ακόμη και μία εικόνα

(Yaga, et al, 2018). Ακόμη και η μικρότερη αλλαγή στη είσοδο της συνάρτησης (π.χ. ένα bit) αποφέρει μία τελείως διαφορετική σύνοψη μηνύματος (Yaga, et al, 2018). Στην Εικόνα 3.1 φαίνονται μερικά παραδείγματα εισόδου και εξόδου από τον αλγόριθμο κατακερματισμού SHA-256 που χρησιμοποιείται ευρέως στις αλυσίδες συστοιχιών (Yaga, et al, 2018).

| Input Text | SHA-256 Digest Value |
|---|---|
| 1 | 0x6b86b273ff34fce19d6b804eff5a3f5747ada4eaa22f1d49c01e52ddb7875b4b |
| 2 | 0xd4735e3a265e16eee03f59718b9b5d03019c07d8b6c51f90da3a666eec13ab35 |
| Hello, World! | 0xdffd6021bb2bd5b0af676290809ec3a53191dd81c7f70a4b28688a362182986f |

**Εικόνα 3.1: Παραδείγματα κειμένου εισόδου και εξόδου αλγορίθμου SHA-256 (Yaga, et al, 2018)**

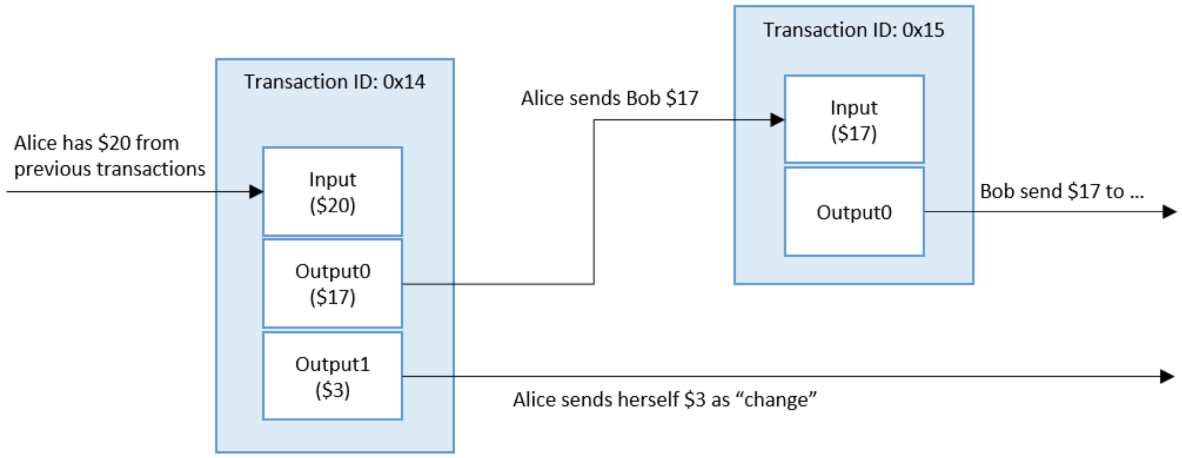
Οι συναρτήσεις κατακερματισμού χαρακτηρίζονται από τρεις βασικές ιδιότητες ασφαλείας, την αντίσταση προεικόνας (preimage resistance), την δεύτερη αντίσταση προεικόνας (second preimage resistance) και την αντίσταση σύγκρουσης (collision resistance) (Yaga, et al, 2018). Η αντίσταση προεικόνας ή αλλιώς η ιδιότητα μονόδρομου (one-way property) είναι η ιδιότητα των συναρτήσεων κατακερματισμού να είναι δύσκολη η εύρεση της αντίστροφης συνάρτησης. Αυτό σημαίνει ότι δίνοντας ένα στοιχείο στο εύρος μιας συνάρτησης κατακερματισμού είναι υπολογιστικά αδύνατο να βρεθεί η είσοδος που αντιστοιχεί σε αυτό το στοιχείο (Tilborg & Jajodia, 2011). Η δεύτερη αντίσταση εικόνας, είναι η ιδιότητα μιας συνάρτησης κατακερματισμού σύμφωνα με την οποία, δοθείσας μίας τιμής εισόδου είναι υπολογιστικά αδύνατο να βρεθεί μία δεύτερη τιμή εισόδου που να αντιστοιχεί στην ίδια έξοδο. Η αντίσταση σύγκρουσης, είναι η ιδιότητα μιας συνάρτησης κατακερματισμού κατά την οποία είναι υπολογιστικά αδύνατο να βρεθούν δύο διακριτές τιμές εισόδου που να έχουν την ίδια έξοδο (Tilborg & Jajodia, 2011). Η διαφορά μεταξύ της δεύτερης αντίστασης εικόνας και της αντίστασης σύγκρουσης είναι ότι στην πρώτη δίνεται ως δεδομένη η πρώτη τιμή εισόδου και αναζητείται η δεύτερη που να δίνει το ίδιο αποτέλεσμα, ενώ στην δεύτερη αναζητούνται και οι δύο τιμές εισόδου που να δίνουν την ίδια έξοδο.

Μία συνάρτηση κατακερματισμού που χρησιμοποιείται ευρέως σε διάφορες υλοποιήσεις είναι ο αλγόριθμος SHA-256 ο οποίος έχει έξοδο μεγέθους 256 bits (Yaga, et al, 2018). Πολλοί υπολογιστές υποστηρίζουν αυτόν τον αλγόριθμο σε επίπεδο υλικού καθιστώντας τον υπολογισμό του αρκετά γρήγορο (Yaga, et al, 2018). Η έξοδος του συνήθως αναπαρίσταται με τη χρήση μιας δεκαεξαδικής ακολουθίας χαρακτήρων όπως φαίνεται στην Εικόνα 1. Αυτό σημαίνει ότι υπάρχουν $2^{256} \approx 10^{77}$ πιθανές συνόψεις μηνυμάτων (Yaga, et al, 2018). Για να βρεθεί σε αυτόν τον αλγόριθμο μία σύγκρουση, δηλαδή δύο είσοδοι με την ίδια έξοδο θα πρέπει να εκτελεστεί ο αλγόριθμος κατά μέσο όρο $2^{128}$ φορές (Yaga, et al, 2018). Έτσι, λαμβάνοντας υπόψιν τον ρυθμό κατακερματισμού ολόκληρου του δικτύου Bitcoin το 2015 που ήταν 300 τετράκις εκατομμύρια ανά δευτερόλεπτο, θα έπαιρνε στο δίκτυο Bitcoin $3.6x10^{13}$ χρόνια για να κατασκευάσει μία σύγκρουση, δηλαδή δύο εισόδους με την ίδια έξοδο (Yaga, et al, 2018).

Οι συναρτήσεις κατακερματισμού χρησιμοποιούνται σε πολλές διαδικασίες μέσα σε ένα δίκτυο αλυσίδων συστοιχιών όπως είναι η παραγωγή διευθύνσεων, η δημιουργία μοναδικών αναγνωριστικών και η ασφάλιση των δεδομένων και της κεφαλίδας των μπλοκ (Yaga, et al, 2018).

### 3.3.2 Συναλλαγές (Transactions)

Μία συναλλαγή μπορεί να θεωρηθεί ως μία δομή δεδομένων που αντιπροσωπεύει μία καταγραφή δραστηριότητας που συμβαίνει σε ψηφιακά ή φυσικά περιουσιακά στοιχεία κατά την αλληλεπίδραση μεταξύ ομότιμων μελών (Yaga, et al, 2018) (Monrat, et al, 2019). Με τα κρυπτονομίσματα για παράδειγμα, μία συναλλαγή αντιπροσωπεύει την μεταφορά κρυπτονομισμάτων μεταξύ χρηστών του δικτύου αλυσίδων συστοιχιών (Yaga, et al, 2018). Στην Εικόνα 3.2 φαίνεται ένα παράδειγμα μιας συναλλαγής κρυπτονομισμάτων.




**Εικόνα 3.2: Παραδείγματα συναλλαγής κρυπτονομισμάτων (Yaga, et al, 2018)**

Μία τυπική συναλλαγή κρυπτονομισμάτων περιέχει τουλάχιστον τις πληροφορίες εισόδου και εξόδου (Yaga, et al, 2018). Οι πληροφορίες εισόδου περιέχουν συνήθως μία λίστα από ψηφιακά περιουσιακά στοιχεία τα οποία πρόκειται να μεταφερθούν, ενώ οι πληροφορίες εξόδου περιέχουν συνήθως τους λογαριασμούς στους οποίους θα μεταφερθούν τα ψηφιακά περιουσιακά στοιχεία, καθώς και την ποσότητα αυτών που θα μεταφερθεί (Yaga, et al, 2018). Σε περίπτωση που η ποσότητα των ψηφιακών περιουσιακών στοιχείων εισόδου είναι μεγαλύτερη από την ποσότητα εξόδου, είτε θα μείνει ως έχει και θα θεωρηθεί η επιπλέον ποσότητα ως προμήθεια συναλλαγής, είτε θα πρέπει η επιπλέον ποσότητα να γυρίσει στον αρχικό της κάτοχο προσθέτοντας στις πληροφορίες εξόδου την επιπλέον ποσότητα σε λογαριασμό του αρχικού κατόχου.

Δύο βασικά ζητήματα ασφαλείας των συναλλαγών που αποτελούν προϋπόθεση για την ομαλή λειτουργία της αλυσίδας συστοιχιών είναι η αποφυγή της διπλής δαπάνης (double spending) και ο έλεγχος γνησιότητας (authenticity). Η αποφυγή της διπλής δαπάνης πραγματοποιείται με τον έλεγχο εγκυρότητας σύμφωνα με τον οποίο ελέγχονται όλες οι προηγούμενες συναλλαγές που έχουν γίνει έτσι ώστε να βρεθεί το πραγματικό υπόλοιπο του λογαριασμού εισόδου. Αυτό συμβαίνει καθώς στις αλυσίδες συστοιχιών κατά την πραγματοποίηση μιας συναλλαγής, δεν αφαιρείται από το υπόλοιπο το ποσό εισόδου, καθώς δεν διατηρούνται υπόλοιπα, αλλά συνδέεται με την προηγούμενη συναλλαγή. Ο έλεγχος γνησιότητας αφορά την παροχή απόδειξης από τον αποστολέα της συναλλαγής ότι έχει πρόσβαση στην αναφερόμενη είσοδο. Αυτό επιτυγχάνεται με την ψηφιακή υπογραφή της συναλλαγής, δηλαδή με τη χρήση μη συμμετρικών κρυπτοσυστημάτων.

### 3.3.3 Μη συμμετρικά κρυπτοσυστήματα (Asymmetric Cryptosystems)

Τα μη συμμετρικά κρυπτοσυστήματα χρησιμοποιούν ένα ζευγάρι από κλειδιά, το δημόσιο κλειδί (public key) και το ιδιωτικό κλειδί (private key) τα οποία είναι μαθηματικά συσχετιζόμενα μεταξύ τους (Yaga, et al, 2018). Για τον λόγο αυτόν, τα μη συμμετρικά κρυπτοσυστήματα αποκαλούνται και συστήματα δημοσίου κλειδιού (public-key system) (Tilborg & Jajodia, 2011). Το δημόσιο κλειδί δημοσιοποιείται χωρίς να μειώνεται η ασφάλεια, ωστόσο το ιδιωτικό κλειδί πρέπει να διατηρηθεί κρυφό για να παραμείνουν τα δεδομένα κρυπτογραφημένα (Yaga, et al, 2018). Αν και υπάρχει μία μαθηματική σχέση μεταξύ των κλειδιών, το ιδιωτικό κλειδί δεν μπορεί να υπολογιστεί αποδοτικά βάσει του δημοσίου κλειδιού (Yaga, et al, 2018).

Τα μη συμμετρικά κρυπτοσυστήματα χρησιμοποιούνται για δύο βασικές λειτουργίες, την αυθεντικοποίηση και την ασφαλή διακίνηση δεδομένων. Για την πραγματοποίηση της αυθεντικοποίησης, ο πρώτος χρήστης χρησιμοποιεί το ιδιωτικό του κλειδί για να κρυπτογραφήσει ένα έγγραφο και ο δεύτερος χρήστης χρησιμοποιεί το δημόσιο κλειδί του πρώτου χρήστη για να αποκρυπτογραφήσει το έγγραφο πιστοποιώντας ότι το έγραψε αυτός. Η διαδικασία που πραγματοποιεί ο πρώτος χρήστης ονομάζεται και ψηφιακή υπογραφή (digital sign). Η ασφαλή διακίνηση δεδομένων συμβαίνει όταν ο πρώτος χρήστης χρησιμοποιεί το δημόσιο κλειδί του δεύτερου χρήστη για να κρυπτογραφήσει ένα έγγραφο και ο δεύτερος χρήστης χρησιμοποιεί το ιδιωτικό του κλειδί για να αποκρυπτογραφήσει το περιεχόμενο του εγγράφου. Με αυτόν τον τρόπο διασφαλίζεται η ασφαλής διακίνηση των δεδομένων, αφού με την προϋπόθεση ότι ο δεύτερος χρήστης έχει διατηρήσει το ιδιωτικό του κλειδί κρυφό, είναι ο μόνος που μπορεί να αποκρυπτογραφήσει το έγγραφο.

Τα μη συμμετρικά κρυπτοσυστήματα αποτελούν αναπόσπαστο κομμάτι στις αλυσίδες συστοιχιών. Τα ιδιωτικά κλειδιά χρησιμοποιούνται για την ψηφιακή υπογραφή των συναλλαγών καθιστώντας έτσι δυνατή την αυθεντικοποίηση τους. Από την άλλη, τα δημόσια κλειδιά χρησιμοποιούνται ως τις διευθύνσεις των χρηστών, καθώς και για την αυθεντικοποίηση των συναλλαγών.

### 3.3.4 Διευθύνσεις (Addresses)

Μία πολύ σημαντική συνιστώσα των αλυσίδων συστοιχιών είναι οι διευθύνσεις στις οποίες αποδίδονται τα ψηφιακά περιουσιακά στοιχεία. Οι διευθύνσεις συνήθως αποτελούνται από μια αλφαριθμητική συμβολοσειρά που παράγεται από το δημόσιο κλειδί ενός χρήστη χρησιμοποιώντας μία συνάρτηση κατακερματισμού μαζί με μερικά επιπλέον δεδομένα όπως είναι για παράδειγμα ο αριθμός της έκδοσης ή το άθροισμα ελέγχου (checksum) (Yaga, et al, 2018). Στις περισσότερες υλοποιήσεις αλυσίδων συστοιχιών, οι διευθύνσεις έχουν μικρότερο μήκος σε σχέση με τα δημόσια κλειδιά και είναι δημόσιες (Yaga, et al, 2018). Ένας τρόπος για να επιτευχθεί αυτό είναι να γίνει εφαρμογή μιας συνάρτησης κατακερματισμού στο δημόσιο κλειδί (Yaga, et al, 2018).

Ένα από τα βασικότερα πλεονεκτήματα που μπορεί να προσδώσει η παραπάνω μεθοδολογία απόδοσης διεύθυνσης στους χρήστες, είναι η όσο το δυνατόν περισσότερη διασφάλιση της ανωνυμίας τους. Αυτό συμβαίνει καθώς στις περισσότερες υλοποιήσεις αλυσίδων συστοιχιών οι χρήστες έχουν την δυνατότητα να δημιουργούν όσα ζευγάρια ασύμμετρων κλειδιών θέλουν άρα και κατ' επέκταση διευθύνσεων (Yaga, et al, 2018).

### 3.3.5 Αποθήκευση Ιδιωτικού Κλειδιού (Private Key Storage)

Στα δίκτυα αλυσίδων συστοιχιών οι χρήστες πρέπει να διαχειρίζονται και να αποθηκεύουν με ασφαλή τρόπο τα ιδιωτικά κλειδιά που τους ανήκουν (Yaga, et al, 2018). Τις περισσότερες φορές αντί να τα καταγράφουν μόνοι τους, χρησιμοποιούν κατάλληλο λογισμικό για την ασφαλή αποθήκευση τους (Yaga, et al, 2018). Αυτό το λογισμικό συνήθως αναφέρεται ως πορτοφόλι (wallet) (Yaga, et al, 2018). Το πορτοφόλι, εκτός από την ασφαλή αποθήκευση των ιδιωτικών κλειδιών μπορεί να εκτελεί και άλλες λειτουργίες όπως είναι ο υπολογισμός της ψηφιακής περιουσίας που κατέχει ο χρήστης (Yaga, et al, 2018).

Τα ιδιωτικά κλειδιά πρέπει να αποθηκεύονται με μεγάλη ασφάλεια καθώς στα δίκτυα αλυσίδων συστοιχιών χωρίς άδεια είναι αδύνατη η ανάκτηση τους έπειτα από απώλεια ή ακόμη και κλοπή. Στην περίπτωση της κλοπής, ο κακόβουλος χρήστης χρησιμοποιεί το ιδιωτικό κλειδί που έκλεψε για να μεταφέρει τα ψηφιακά περιουσιακά στοιχεία μέσω μιας έγκυρης συναλλαγής σε έναν δικό του λογαριασμό. Εάν συμβεί αυτό, ο μόνος τρόπος να ακυρωθεί η συναλλαγή είναι η πραγματοποίηση της σκληρής διακλάδωσης (hard fork), σύμφωνα με την οποία` το μεγαλύτερο σύνολο των χρηστών του δικτύου αποφασίζουν να αλλάξουν την ροή της αλυσίδας των συστοιχιών θεωρώντας ως μη έγκυρες όλες τις συναλλαγές που έλαβαν χώρα από την κακόβουλη συναλλαγή και μετά.

### 3.3.6 Καθολικά (Ledgers)

Ένα καθολικό είναι μία συλλογή από συναλλαγές (Yaga, et al, 2018). Τα καθολικά αποθηκεύονται ψηφιακά σε μεγάλες βάσεις δεδομένων, οι οποίες συνήθως ανήκουν και λειτουργούν από κάποιο κεντροποιημένο αξιόπιστο τρίτο μέρος (centralized trusted third party), εκ μέρους μιας κοινωνίας χρηστών (community of users) (Yaga, et al, 2018). Αυτά τα καθολικά με κεντροποιημένη ιδιοκτησία μπορούν να εφαρμοστούν και με κατανεμημένο τρόπο με τη χρήση για παράδειγμα ενός συμπλέγματος διακομιστών (cluster of servers) (Yaga, et al, 2018).

Σε διαφοροποίηση με τα παραπάνω, η τεχνολογία αλυσίδων συστοιχιών επιτρέπει μία προσέγγιση χρησιμοποιώντας αμφότερα κατανεμημένη ιδιοκτησία (distributed ownership) και κατανεμημένη φυσική αρχιτεκτονική (distributed physical architecture) (Yaga, et al, 2018). Η κατανεμημένη φυσική αρχιτεκτονική των δικτύων αλυσίδων συστοιχιών, περιλαμβάνει συνήθως πολύ μεγαλύτερο αριθμό από υπολογιστές σε σχέση με την τυπική κατανεμημένη αρχιτεκτονική που διαχειρίζεται κεντρικά (Yaga, et al, 2018). Το έντονο ενδιαφέρον που υπάρχει σχετικά με τα καθολικά κατανεμημένης ιδιοκτησίας οφείλεται στα πιθανά πλεονεκτήματα που έχουν σε σχέση με τα καθολικά κεντρικής ιδιοκτησίας, ως προς την εμπιστοσύνη, την ασφάλεια και την αξιοπιστία (Yaga, et al, 2018).

Μερικά από τα βασικότερα πλεονεκτήματα των κατανεμημένης ιδιοκτησίας καθολικών έναντι των κεντρικής ιδιοκτησίας καθολικών είναι:

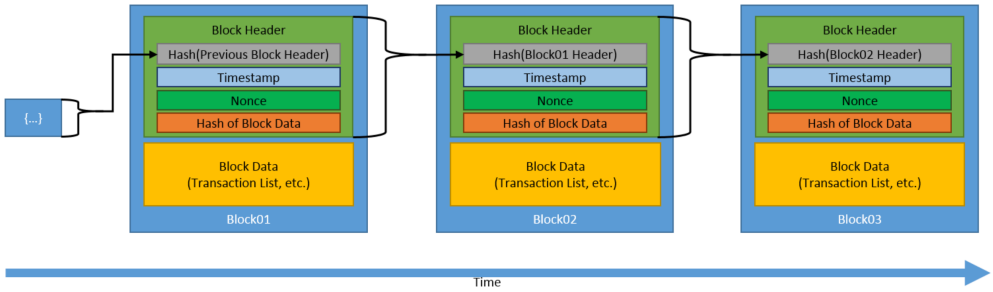
- Τα κεντρικής ιδιοκτησίας καθολικά σε περίπτωση απώλειας ή καταστροφής, βασίζονται στο γεγονός ότι ο ιδιοκτήτης τους θα έχει κάνει αντίγραφα ασφαλείας, εξασφαλίζοντας την ομαλή συνέχεια τους (Yaga, et al, 2018). Αντίθετα, στα καθολικά κατανεμημένης ιδιοκτησίας, ο κάθε χρήστης έχει την δυνατότητα να διατηρεί αντίγραφα ασφαλείας εξασφαλίζοντας στην ουσία την ψηφιακή του περιουσία.
- Τα κεντρικής ιδιοκτησίας καθολικά συνήθως βρίσκονται σε ομοιογενή δίκτυα, όπου η υποδομή λογισμικού, υλικού και δικτύου είναι συνήθως η ίδια για όλο το δίκτυο (Yaga,

et al, 2018). Εξαιτίας αυτού του χαρακτηριστικού, η ανθεκτικότητα ολόκληρου του συστήματος μπορεί να μειωθεί, καθώς σε περίπτωση επιτυχούς επίθεσης σε ένα μέρος του δικτύου, θα είναι δυνατή και στο υπόλοιπο (Yaga, et al, 2018). Αντίθετα, τα καθολικά κατανεμημένης ιδιοκτησίας είναι ετερογενή, δηλαδή η υποδομή λογισμικού, υλικού και δικτύου είναι διαφορετική, οπότε μία επιτυχής επίθεση σε ένα μέρος του δικτύου δεν εγγυάται ότι θα είναι επιτυχής και στο υπόλοιπο (Yaga, et al, 2018).

- Οι συναλλαγές που λαμβάνουν χώρα σε ένα κεντρικής ιδιοκτησίας καθολικό δεν διατρέχονται από απόλυτη διαφάνεια καθώς η εγκυρότητα, η επιτυχής ολοκλήρωση και η αμεταβλητότητα των συναλλαγών εξαρτάται αποκλειστικά και μόνο από τον ιδιοκτήτη του. Σε αντίθεση με τα κατανεμημένης ιδιοκτησίας καθολικά, όπου η διαφάνεια μπορεί να ελεγχθεί από όλους τους χρήστες του δικτύου.

### 3.3.7 Μπλοκ (Blocks)

Οι δομικές μονάδες των αλυσίδων συστοιχιών που περιέχουν τις συναλλαγές ονομάζονται μπλοκ. Τα μπλοκ είναι συνδεδεμένα μεταξύ τους αποτελώντας μία ακολουθία από μπλοκ η οποία είναι γνωστή και ως αλυσίδα (Yaga, et al, 2018). Τα μπλοκ είναι συνδεδεμένα μεταξύ τους μέσω μιας αναφοράς κατακερματισμού η οποία ανήκει στο προηγούμενο μπλοκ γνωστό και ως γονικό μπλοκ (parent block) (Monrat, et al, 2019). Το πρώτο μπλοκ της αλυσίδας συστοιχιών ονομάζεται μπλοκ γένεσης και δεν έχει γονικό μπλοκ (Monrat, et al, 2019). Κάθε μπλοκ αποτελείται από την κεφαλίδα (header) και τα δεδομένα (data). Στην Εικόνα 3 φαίνεται μία αλυσίδα συστοιχιών αποτελούμενη από τρία μπλοκ.




**Εικόνα 3: Αλυσίδα Συστοιχιών με τρία Μπλοκ (Yaga, et al, 2018)**

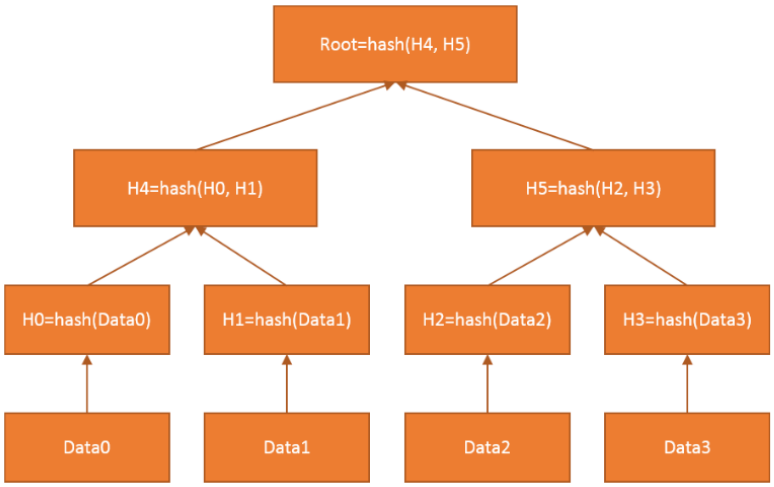
Η κεφαλίδα κάθε μπλοκ περιέχει:

- Τον αριθμό του μπλοκ, ο οποίος σε πολλά δίκτυα αλυσίδων συστοιχιών είναι γνωστός και ως το ύψος του μπλοκ (Yaga, et al, 2018).
- Την κατακερματισμένη τιμή του προηγούμενου μπλοκ (Yaga, et al, 2018).
- Μία κατακερματισμένη αναφορά των δεδομένων του μπλοκ (Yaga, et al, 2018). Για την δημιουργία αυτής της αναφοράς μπορούν να χρησιμοποιηθούν διάφορες τεχνικές, ωστόσο αυτή που χρησιμοποιείται κυρίως είναι η δημιουργία του Δέντρου Μερκλ (Merkle Tree). Το Δέντρο Μερκλ αποτελεί μία δομή δεδομένων όπου τα δεδομένα κατακερματίζονται και συνδυάζονται μέχρι να υπάρξει μία μοναδική τιμή κατακερματισμού (ρίζα) η οποία αντιπροσωπεύει ολόκληρη τη δομή (Εικόνα 4) (Yaga, et al, 2018).
- Μία χρονική σήμανση (timestamp) (Yaga, et al, 2018).
- Το μέγεθος του μπλοκ (Yaga, et al, 2018).
- Την μηδενική αξία (nonce value) (Yaga, et al, 2018). Η μηδενική αξία αποτελεί για τα δίκτυα αλυσίδων συστοιχιών που χρησιμοποιούν εξόρυξη (mining), τον αριθμό που χειραγωγείται από τον κόμβο δημοσίευσης (publishing node) για να λυθεί το παζλ

κατακερματισμού (Yaga, et al, 2018). Τα υπόλοιπα δίκτυα αλυσίδων συστοιχιών είτε δεν τον περιλαμβάνουν, είτε τον χρησιμοποιούν για κάποιον άλλον σκοπό (Yaga, et al, 2018).

Τα δεδομένα του μπλοκ περιέχουν:

- Τον αριθμό των συναλλαγών που περιέχει το μπλοκ (Monrat, et al, 2019).
- Την λίστα των καταγεγραμμένων συναλλαγών που περιέχονται στο μπλοκ (Monrat, et al, 2019).




**Εικόνα 4: Παράδειγμα Δέντρου Μερκλ (Yaga, et al, 2018)**

Ο τρόπος σύνδεσης των μπλοκ μεταξύ τους αποτελεί το βασικό χαρακτηριστικό της αλυσίδας συστοιχιών. Συνήθως η αναφορά κατακερματισμού μέσω της οποίας επιτυγχάνεται η σύνδεση περιέχει τη σύνοψη μηνύματος της κεφαλίδας του γονικού μπλοκ. Με αυτόν τον τρόπο εάν ένα από τα μπλοκ μεταβληθεί θα έχει διαφορετικό μήνυμα σύνοψης, και κατά συνέπεια θα οδηγήσει στην μεταβολή και των επόμενων μπλοκ της αλυσίδας. Έτσι μπορούν να εντοπιστούν πολύ εύκολα οι μεταβολές των προηγούμενων μπλοκ και να απορριφθούν.

## 3.4 Μοντέλα Συναίνεσης (Consensus Models)

Μία βασική πτυχή της τεχνολογίας των αλυσίδων συστοιχιών είναι η απόφαση ποιος χρήστης θα δημοσιεύσει το επόμενο μπλοκ (Yaga, et al, 2018). Αυτή η απόφαση επιτυγχάνεται με την χρησιμοποίηση των μοντέλων συναίνεσης. Τα δημοφιλέστερα μοντέλα συναίνεσης είναι το μοντέλο απόδειξης εργασίας (Proof of Work Model), το μοντέλο απόδειξης πονταρίσματος (Proof of Stake Model), το μοντέλο εξουσιοδοτημένης απόδειξης πονταρίσματος (Delegated Proof of Stake Model), το κατά σειρά μοντέλο (Round Robin Model), το μοντέλο απόδειξης ταυτότητας (Proof of Identity Model) και το μοντέλο απόδειξης του χρόνου που παρήλθε (Proof of Elapsed Time Model).

### 3.4.1 Μοντέλο απόδειξης εργασίας (Proof of Work Model)

Το μοντέλο της απόδειξης εργασίας χρησιμοποιεί μία τεχνική συναίνεσης η οποία βασίζεται στη απόδειξη εργασίας του χρήστη. Πιο αναλυτικά, η βασική ιδέα αυτής της τεχνικής είναι η ταυτοποίηση και η επιλογή του κόμβου που θα αποκτήσει το δικαίωμα να προσθέσει ένα νέο

μπλοκ στην υπάρχουσα αλυσίδα, παρέχοντας την επαρκή απόδειξη της εργασίας του (Monrat, et al, 2019). Για την απόδειξη της εργασίας του ο κάθε χρήστης προσπαθεί να λύσει έναν υπολογιστικά απαιτητικό γρίφο. Η λύση αυτού του γρίφου αποτελεί την απόδειξη εργασίας του χρήστη. Ο γρίφος είναι σχεδιασμένος με τέτοιο τρόπο έτσι ώστε η εύρεση της λύσης του να είναι υπολογιστικά δύσκολη, ενώ ταυτόχρονα ο έλεγχος ότι μια λύση είναι η σωστή να είναι υπολογιστικά πολύ εύκολος (Yaga, et al, 2018). Αυτό καθιστά εύκολη την επικύρωση των νέων μπλοκ που προτείνονται και την απόρριψη των μπλοκ που προτείνονται κακόβουλα χωρίς να ικανοποιούν τον γρίφο.

Ένας γρίφος που χρησιμοποιείται συχνά είναι η εύρεση της σύνοψης μηνύματος της κεφαλής ενός μπλοκ σε συνδυασμό με έναν αυθαίρετο αριθμό που χρησιμοποιείται μόνο μία φορά και ονομάζεται μονοεπιλογή (nonce). Η εύρεση της σύνοψης μηνύματος πραγματοποιείται με την συνεχή αλλαγή της μονοεπιλογής ώσπου η τιμή της σύνοψης μηνύματος να είναι μικρότερη ή ίση με την τιμή στόχου. Επειδή σε κάθε προσπάθεια για τον υπολογισμό της πιθανής σύνοψης μηνύματος υπολογίζεται το αποτέλεσμα μιας συνάρτησης κατακερματισμού, η όλη διαδικασία είναι υπολογιστικά πολύ δύσκολη. Σε πολλές περιπτώσεις η δυσκολία προσαρμόζεται στο πέρασμα του χρόνου, μέσω της αλλαγής της τιμής του στόχου, έτσι ώστε να παραμείνει σταθερός ο ρυθμός δημοσίευσης μπλοκ.

Προκειμένου να δοθεί κίνητρο στους χρήστες να καταναλώνουν σημαντικούς προσωπικούς τους πόρους για την εύρεση του γρίφου, ο πρώτος χρήστης που λύνει τον γρίφο λαμβάνει τις προμήθειες των συναλλαγών και ένα έπαθλο από το σύστημα. Το έπαθλο είναι αρκετά μεγάλο με αποτέλεσμα πολλοί χρήστες να οργανώνονται σε συλλογικότητες (collectives) για να λύσουν μαζί τον γρίφο και να μοιραστούν το έπαθλο. Αυτό ενέχει τον κίνδυνο της κεντρικοποίησης του συστήματος σε περίπτωση που μία συλλογικότητα έχει στην κατοχή της το μεγαλύτερο μέρος της υπολογιστικής δύναμης του δικτύου.

Το κύριο μέλημα σχετικά με το μοντέλο απόδειξης εργασίας είναι η υψηλή απαίτηση σε πόρους για την λειτουργία του. Οι πόροι που απαιτούνται είναι είτε σε υλικό (hardware), είτε σε ενέργεια, ειδικότερα σε ηλεκτρικό ρεύμα. Για να γίνει σαφές το παραπάνω, αρκεί να αναλογιστούμε ότι το κόστος κατανάλωσης ρεύματος από το δίκτυο αλυσίδων συστοιχιών ‘Bitcoin’ ξεπερνάει τα 3 δισεκατομμύρια δολάρια τον χρόνο (Casey, 2018). Αυτό αποτελεί κίνδυνο για την βιωσιμότητα των δικτύων αλυσίδων συστοιχιών που χρησιμοποιούν το μοντέλο απόδειξης εργασίας, καθώς κατά την επίλυση του γρίφου ανταμείβεται μόνο ένας χρήστης, οπότε όλοι οι υπόλοιποι χρήστες που κατανάλωσαν πόρους για την επίλυση του γρίφου βγαίνουν ζημιογόνοι.

### 3.4.2 Μοντέλο απόδειξης πονταρίσματος (Proof of Stake Model)

Το μοντέλο απόδειξης πονταρίσματος βασίζεται στην ιδέα ότι όσο περισσότερο είναι το ποντάρισμα (stake) που έχει επενδύσει ένας χρήστης σε ένα σύστημα, τόσο πιο πιθανό είναι να θέλει το σύστημα να πετύχει (Yaga, et al, 2018). Ως ποντάρισμα αναφέρεται συνήθως το ποσό των κρυπτονομισμάτων που ένας χρήστης του δικτύου της αλυσίδας συστοιχιών έχει επενδύσει στο σύστημα με διάφορα μέσα, όπως είναι το κλείδωμα τους μέσω ενός ειδικού τύπου συναλλαγής, ή στέλνοντας τα σε μία συγκεκριμένη διεύθυνση, ή κρατώντας τα μέσα σε ένα ειδικό ψηφιακό πορτοφόλι (Yaga, et al, 2018). Αφότου ένας χρήστης πραγματοποιήσει το ποντάρισμα, τα κρυπτονομίσματα του πονταρίσματος δεσμεύονται από το σύστημα και ο χρήστης λογίζεται ως επικυρωτής (validator). Για την επιλογή του επικυρωτή, δηλαδή του χρήστη που θα δημοσιεύσει το επόμενο μπλοκ, το σύστημα λαμβάνει υπόψη την αναλογία του

πονταρίσματος σε σχέση με το συνολικό ποντάρισμα στο σύστημα. Δηλαδή, όσο μεγαλύτερο είναι το ποντάρισμα ενός χρήστη στο σύστημα, τόσο πιο πιθανό είναι να επιλεχθεί ως επικυρωτής.

Το μοντέλο απόδειξης πονταρίσματος πλεονεκτεί σε σχέση με το μοντέλο απόδειξης εργασίας ως προς την κατανάλωση πόρων, καθώς δεν είναι υπολογιστικά απαιτητικό. Πιο συγκεκριμένα, για την λειτουργία του χρειάζεται ελάχιστος χρόνος, ηλεκτρικό ρεύμα και υπολογιστική ισχύ. Μία ακόμη σημαντική διαφοροποίηση, είναι ότι το μοντέλο απόδειξης πονταρίσματος δεν ανταμείβει τους χρήστες που δημοσιεύουν ένα νέο μπλοκ. Δηλαδή, οι χρήστες που δημοσιεύουν ένα νέο μπλοκ έχουν ως όφελος μόνο τις προμήθειες από τις συναλλαγές που περιέχονται στο μπλοκ.

Υπάρχουν διάφορες μέθοδοι που ένα δίκτυο αλυσίδων συστοιχιών χρησιμοποιεί το ποντάρισμα για την επιλογή του χρήστη που θα δημοσιεύσει ένα νέο μπλοκ. Όλες οι μέθοδοι, αν και διαφέρουν ως προς την χρήση του πονταρίσματος, έχουν ως κοινό στοιχείο ότι όσο μεγαλύτερο είναι το ποντάρισμα του χρήστη, τόσο περισσότερες πιθανότητες έχει να δημοσιεύσει ένα νέο μπλοκ. Μερικές από αυτές τις μεθόδους είναι η τυχαία επιλογή των χρηστών που πόνταραν (random selection of staked users), η κυκλική ψηφοφορία (multi-round voting), τα συστήματα παλαίωσης νομισμάτων (coin aging systems) και τα εξουσιοδοτημένα συστήματα (delegated systems) (Yaga, et al, 2018).

Σύμφωνα με τη μέθοδο τυχαίας επιλογής των χρηστών που πόνταραν, το δίκτυο αλυσίδας συστοιχιών επιλέγει τον χρήστη που θα δημοσιεύσει το επόμενο μπλοκ βάσει της αναλογίας του πονταρίσματος κάθε χρήστη σε σχέση με το συνολικό ποντάρισμα. Οι χρήστες που δεν έχουν ποντάρει δεν αποτελούν επιλογή. Για παράδειγμα εάν ένας χρήστης έχει ποντάρει 51% του συνόλου των πονταρισμάτων, τότε έχει 51% πιθανότητες να επιλεχθεί για την δημοσίευση του επόμενου μπλοκ.

Η μέθοδος κυκλικής ψηφοφορίας αναφέρεται και ως Βυζαντινή απόδειξη ανοχής σφαλμάτων πονταρίσματος (Byzantine fault tolerance proof of stake) (Yaga, et al, 2018). Σε αυτή τη μέθοδο, αρχικά το σύστημα αλυσίδας συστοιχιών επιλέγει μερικούς χρήστες βάσει της αναλογίας του πονταρίσματος τους σε σχέση με το συνολικό ποντάρισμα, και στη συνέχεια γίνεται ψηφοφορία, στην οποία οι χρήστες που έχουν ποντάρει ψηφίζουν ποιος από τους επιλεγμένους χρήστες θα δημοσιεύσει το επόμενο μπλοκ. Προκειμένου να αποφασιστεί ο χρήστης, μπορεί να λάβουν χώρα αρκετοί γύροι ψηφοφορίας (Yaga, et al, 2018). Από τα παραπάνω προκύπτει ότι σε αυτή τη μέθοδο έχουν λόγο όλοι οι χρήστες που πόνταραν ανεξαρτήτως μεγέθους πονταρίσματος.

Στα συστήματα παλαίωσης νομισμάτων εκτός από το μέγεθος του πονταρίσματος λαμβάνεται υπόψη και ο χρόνος δέσμευσης του. Πιο συγκεκριμένα, μετά το πέρας ενός χρονικού διαστήματος δέσμευσης του πονταρίσματος, ο χρήστης είναι ικανός να επιλεχθεί από το σύστημα για την δημοσίευση του επόμενου μπλοκ. Σε περίπτωση που επιλεχθεί ένας χρήστης, το ποντάρισμα του αποδεσμεύεται, και για να καταστεί ικανός επιλογής θα πρέπει να ποντάρει ξανά και να περάσει το απαιτούμενο χρονικό διάστημα. Αυτή η μέθοδος, αν και επιτρέπει στους χρήστες με τα μεγαλύτερα πονταρίσματα, και άρα στους χρήστες με το μεγαλύτερο μερίδιο της αγοράς, να επιλέγονται συχνότερα για τη δημοσίευση νέων μπλοκ, τους περιορίζει χρησιμοποιώντας τον μηχανισμό επαναφοράς του χρόνου δέσμευσης του πονταρίσματος που αναφέρθηκε παραπάνω.

Στα εξουσιοδοτημένα συστήματα, οι χρήστες που ποντάρουν ψηφίζουν για την επιλογή του χρήστη που θα δημοσιεύσει το νέο μπλοκ. Η ψήφος του κάθε χρήστη έχει ισχύ ανάλογα με το ύψος του ποντάρίσματός του σε σχέση με το σύνολο των ποντάρισμάτων. Οι ψήφοι εκτός από θετικές μπορεί να είναι και αρνητικές. Δηλαδή, οι χρήστες μπορούν να ψηφίσουν τον χρήστη που δεν θέλουν να δημοσιεύσει το επόμενο μπλοκ. Με αυτόν τον τρόπο διασφαλίζεται ότι οι χρήστες έχουν κίνητρο να μην λειτουργούν κακόβουλα, καθώς εάν το κάνουν, δεν πρόκειται να δημοσιεύσουν ποτέ ένα νέο μπλοκ και να λάβουν τα οφέλη που αυτό συνεπάγεται.

Δύο πολύ σημαντικά προβλήματα που μπορεί να υπάρχουν στο μοντέλο απόδειξης ποντάρίσματος ανεξάρτητα της μεθόδου επιλογής του χρήστη που θα δημοσιεύσει το επόμενο μπλοκ, είναι το πρόβλημα της έλλειψης ποντάρισμάτων και της κεντρικοποίησης του συστήματος με την συσσώρευση πλούτου. Το πρόβλημα της έλλειψης ποντάρισμάτων αναφέρεται στην κατάσταση του συστήματος όπου κανένας χρήστης δεν έχει ποντάρει. Αυτή η κατάσταση καθιστά το σύστημα αναξιόπιστο καθώς ο οποιασδήποτε κακόβουλος χρήστης μπορεί με ένα ελάχιστο ποντάρισμα να καταρρίψει το σύστημα. Το πρόβλημα της κεντρικοποίησης του συστήματος με την συσσώρευση πλούτου, εμφανίζεται όταν τα ψηφιακά περιουσιακά στοιχεία όπως είναι τα κρυπτονομίσματα συγκεντρώνονται σε λίγους χρήστες. Αυτοί οι λίγοι χρήστες, έχουν την δυνατότητα πολύ μεγαλύτερων ποντάρισμάτων σε σχέση με τους υπόλοιπους χρήστες, και άρα την δυνατότητα να ελέγχουν το σύστημα μέσω της συνεχούς επιλογής τους ως επικυρωτές.

### 3.4.3 Κατά σειρά μοντέλο (Round Robin Model)

Το κατά σειρά μοντέλο χρησιμοποιείται από μερικά δίκτυα αλυσίδων συστοιχιών με άδεια (Yaga, et al, 2018). Η επιλογή του χρήστη που θα δημοσιεύσει το επόμενο μπλοκ είναι κυκλική, δηλαδή όλοι οι χρήστες μπαίνουν σε μία ουρά και επιλέγονται με τη σειρά. Σε περίπτωση που έρθει η σειρά ενός χρήστη και δεν είναι διαθέσιμος για την υποβολή ενός νέου μπλοκ, τότε το σύστημα αναμένει ένα μικρό χρονικό διάστημα μήπως ο χρήστης καταστεί διαθέσιμος και αν στην συνέχεια δεν το κάνει, τότε προχωράει στον επόμενο. Για την λειτουργία αυτού του μοντέλου απαιτούνται ελάχιστοι πόροι από τους χρήστες, ωστόσο είναι ακατάλληλο για χρήση σε δίκτυα αλυσίδων συστοιχιών χωρίς άδεια, καθώς δίνει την δυνατότητα σε κακόβουλους χρήστες να δημοσιεύουν νέα μπλοκ.

### 3.4.4 Μοντέλο απόδειξης ταυτότητας (Proof of Identity Model)

Το μοντέλο απόδειξης ταυτότητας χρησιμοποιεί την φήμη του κάθε χρήστη ως μέτρο σύγκρισης για την επιλογή του χρήστη που θα δημοσιεύσει ένα νέο μπλοκ. Οι χρήστες που επιλέγονται στην ουσία ποντάρουν την φήμη τους. Προκειμένου να καταστεί ένας χρήστης επιλέξιμος, θα πρέπει αρχικά να έχει προχωρήσει στην ταυτοποίηση των στοιχείων του βάσει διάφορων εγγράφων ταυτοποίησης (π.χ. διαβατήριο) που δημοσιεύονται μέσα στο δίκτυο. Κάθε χρήστης που έχει ταυτοποιηθεί, όση περισσότερη φήμη έχει μέσα στο δίκτυο, τόσες περισσότερες πιθανότητες έχει να επιλεχθεί για την δημοσίευση ενός νέου κόμβου. Η φήμη του κάθε χρήστη αυξάνεται όταν ο χρήστης λειτουργεί προς το συμφέρον του δικτύου, ενώ μειώνεται όταν λειτουργεί κακόβουλα.

### 3.4.5 Μοντέλο απόδειξης του χρόνου που παρήλθε (Proof of Elapsed Time Model)

Στο μοντέλο απόδειξης του χρόνου που παρήλθε, ο κάθε χρήστης που θέλει να επιλεχθεί από το σύστημα για την δημοσίευση ενός νέου κόμβου, ζητάει ένα χρονικό διάστημα αναμονής από ένα ασφαλές υλικό (hardware) και στη συνέχεια μπαίνει σε αναμονή για αυτό το διάστημα. Ο πρώτος χρήστης που θα βγει από αυτή την αναμονή, επιλέγεται από το σύστημα να δημοσιεύσει τον νέο κόμβο. Για να λειτουργήσει αυτό το μοντέλο σωστά και προς όφελος του δικτύου, θα πρέπει να υπάρχει εμπιστοσύνη τόσο μεταξύ των χρηστών όσο και για το υλικό που χρησιμοποιούνε. Αυτό πρακτικά σημαίνει ότι το συγκεκριμένο μοντέλο μπορεί να χρησιμοποιηθεί μόνο σε δίκτυα αλυσίδων συστοιχιών με άδεια.

## 3.5 Πραγματοποίηση Διακλαδώσεων (Forking)

Η πραγματοποίηση αλλαγών και αναβαθμίσεων στις αλυσίδες συστοιχιών μπορεί να αποδειχθεί μία πολύ δύσκολη διαδικασία, ειδικά στα δίκτυα αλυσίδων συστοιχιών χωρίς άδεια όπου συμμετέχουν πολλοί και διαφορετικοί χρήστες. Οι αλλαγές και οι αναβαθμίσεις σε μία αλυσίδα συστοιχιών, ονομάζονται διακλαδώσεις (forks). Οι διακλαδώσεις ανάλογα με τον τρόπο που επηρεάζουν το δίκτυο, μπορούν να χωριστούν σε δύο κατηγορίες, την μαλακή διακλάδωση (soft fork) και την σκληρή διακλάδωση.

### 3.5.1 Μαλακή Διακλάδωση (Soft Fork)

Μία μαλακή διακλάδωση αποτελεί μία αλλαγή που συμβαίνει σε ένα δίκτυο αλυσίδας συστοιχιών όταν αυτή η αλλαγή είναι συμβατή προς τα πίσω (Yaga, et al, 2018). Αυτό σημαίνει ότι οι χρήστες που δεν υιοθέτησαν αυτή την αλλαγή, μπορούν να συνεχίσουν να συναλλάσσονται με αυτούς που την υιοθέτησαν χωρίς πρόβλημα. Κάθε αλλαγή που πραγματοποιείται, για να μονιμοποιηθεί στο δίκτυο, θα πρέπει να υιοθετηθεί από την πλειοψηφία των χρηστών που συμμετέχουν στο δίκτυο.

### 3.5.2 Σκληρή Διακλάδωση (Hard Fork)

Μία σκληρή διακλάδωση αποτελεί μία αλλαγή που συμβαίνει σε ένα δίκτυο αλυσίδας συστοιχιών όταν αυτή η αλλαγή δεν είναι συμβατή προς τα πίσω (Yaga, et al, 2018). Μία τέτοια αλλαγή αφορά συνήθως την διόρθωση συναλλαγών που δεν έπρεπε να είχαν δημοσιευτεί. Σε αυτή την περίπτωση, επιλέγεται να δημιουργηθεί μία νέα αλυσίδα από το προηγούμενο μπλοκ του προβληματικού, δηλαδή αυτού που περιέχει τις λανθασμένες συναλλαγές. Έτσι δημιουργείται στην ουσία μία νέα αλυσίδα η οποία δεν περιέχει το προβληματικό μπλοκ και όλα όσα δημοσιεύτηκαν μετά από αυτό. Για να εδραιωθεί αυτή η αλλαγή στο δίκτυο, θα πρέπει η πλειοψηφία των χρηστών να ακολουθήσει αυτή τη νέα αλυσίδα. Οι χρήστες που δεν ακολούθησαν την νέα αλυσίδα, δεν μπορούν να συνεχίσουν να συναλλάσσονται με αυτούς που την ακολούθησαν.

### 3.5.3 Διακλάδωση στο Καθολικό (Forking in Ledger)

Μία διακλάδωση στο καθολικό συμβαίνει όταν δημοσιεύονται ταυτόχρονα πολλαπλά μπλοκ. Αυτό έχει ως αποτέλεσμα, έστω και για πάρα πολύ μικρό χρονικό διάστημα να υπάρχουν δύο

αλυσίδες μέσα στο δίκτυο αλυσίδων συστοιχιών. Για την επίλυση αυτής της διακλάδωσης, το σύστημα επιλέγει την αλυσίδα που έχει το μεγαλύτερο μήκος στο άμεσο χρονικό διάστημα. Αυτή η διακλάδωση μπορεί να θεωρηθεί ως μία σκληρή διακλάδωση που πραγματοποιείται στιγμιαία από το σύστημα, και οι χρήστες δεν έχουν κανέναν έλεγχο πάνω σε αυτήν.

## 3.6 Έξυπνες Συμβάσεις (Smart Contracts)

Μία έξυπνη σύμβαση είναι μία συλλογή από κώδικα και δεδομένα (μερικές φορές αναφέρονται και ως συναρτήσεις και καταστάσεις), που έχουν αναπτυχθεί χρησιμοποιώντας κρυπτογραφικά υπογεγραμμένες συναλλαγές μέσα σε ένα δίκτυο αλυσίδας συστοιχιών (Yaga, et al, 2018). Οι έξυπνες συμβάσεις εκτελούνται συνήθως από τους χρήστες που δημοσιεύουν τα μπλοκ στο δίκτυο, και δημοσιεύονται στην αλυσίδα συστοιχιών. Οι έξυπνες συμβάσεις, αν και προϋπήρχαν σαν ιδέα των δικτύων αλυσίδας συστοιχιών, δεν μπορούσαν να βρουν πρακτική εφαρμογή λόγω της έλλειψης πλήρως αποκεντροποιημένων δικτύων όπως είναι πολλά από τα δίκτυα αλυσίδας συστοιχιών.

Οι έξυπνες συμβάσεις μπορούν να αποτελέσουν ένα πολύ σημαντικό εργαλείο για τους χρήστες των δικτύων αλυσίδας συστοιχιών, καθώς πρόκειται για συμβάσεις που υιοθετούν όλα τα χαρακτηριστικά του δικτύου αλυσίδας συστοιχιών που τις περιέχει. Για παράδειγμα, σε ένα πλήρως αποκεντροποιημένο δίκτυο αλυσίδας συστοιχιών, οι συμβάσεις δεν μπορούν να αλλάξουν μονομερώς από τη στιγμή που θα δημοσιευτούν και η εκτέλεση τους επικυρώνεται από όλο το δίκτυο. Δηλαδή το δίκτυο λειτουργεί ως ένα έμπιστο τρίτο μέρος. Με αυτό το χαρακτηριστικό οι συναλλαγές μεταξύ επιχειρήσεων θα μπορούσαν να είναι τελείως διαφανείς, πολύ γρήγορες στην εκτέλεση και απαλλαγμένες από γραφειοκρατικά κόστη.

## 3.7 Εφαρμογές των Αλυσίδων Συστοιχιών (Blockchain Applications)

Η τεχνολογία αλυσίδων συστοιχιών αν και έχει συνδεθεί άρρηκτα με τα κρυπτονομίσματα, μπορεί να βρει εφαρμογή σε πληθώρα άλλων διαφορετικών τομέων. Οι κυριότεροι από αυτούς τους τομείς είναι η διαχείριση των αρχείων υγειονομικής περίθαλψης, η ενεργειακή βιομηχανία, η διαχείριση της ταυτότητας, το διαδίκτυο της αξίας (Internet of Value), οι εφοδιαστικές αλυσίδες, το χρηματιστήριο, η χρηματοδότηση του εμπορίου και τα κρυπτονομίσματα που αποτελούν αυτή τη στιγμή την κυριότερη εφαρμογή των αλυσίδων συστοιχιών.

### 3.7.1 Διαχείριση των αρχείων υγειονομικής περίθαλψης

Η τεχνολογία αλυσίδων συστοιχιών μπορεί να συνεισφέρει δραστικά σε ένα πολύ σημαντικό πρόβλημα που αντιμετωπίζουν οι φορείς παροχής υπηρεσιών υγειονομικής περίθαλψης, την διαχείριση των δεδομένων των ασθενών. Αυτή την στιγμή, κάθε φορέας συνήθως έχει την δική του βάση δεδομένων όπου καταγράφει τα δεδομένα του ασθενή. Αυτό έχει ως αποτέλεσμα εάν ο ασθενής μεταβεί σε κάποιον άλλον φορέα, ο νέος φορέας να μην έχει πρόσβαση στα ιατρικά δεδομένα του ασθενή, και να τον περιθάλπει χωρίς να γνωρίζει το ακριβές ιστορικό του. Αυτό το πρόβλημα μπορεί να λυθεί με την χρήση ενός δικτύου αλυσίδας συστοιχιών με άδεια όπου όλοι οι φορείς θα έχουν πρόσβαση.

### 3.7.2 Ενεργειακή Βιομηχανία

Η τεχνολογία αλυσίδων συστοιχιών μπορεί να εφαρμοστεί για την δημιουργία ενός έξυπνου ηλεκτρικού δικτύου. Το έξυπνο ηλεκτρικό δίκτυο θα μπορεί να φέρει σε επαφή όλους όσους παράγουν και καταναλώνουν ηλεκτρικό ρεύμα έτσι ώστε να διαχειρίζεται πιο αποτελεσματικά το παραγόμενο ηλεκτρικό ρεύμα. Αυτό θα έχει ως αποτέλεσμα αφενός την μείωση του ηλεκτρικού ρεύματος που παράγεται με περιβαλλοντικά ακριβούς τρόπους κι αφετέρου την μείωση του κόστους της ηλεκτρικής ενέργειας. Για να συμβεί αυτό, θα πρέπει να δημιουργηθεί ένα δίκτυο αλυσίδας συστοιχιών όπου με τη βοήθεια έξυπνων συμβάσεων θα είναι δυνατή η άμεση συναλλαγή ηλεκτρικού ρεύματος.

### 3.7.3 Διαχείριση Ταυτότητας

Η έρευνα επικεντρώνεται επίσης στην βελτίωση της διαχείρισης της ταυτότητας τόσο για την παραδοσιακή λειτουργία ταυτοποίησης που εκτελείται από καθιερωμένες βιομηχανίες όπως είναι οι τράπεζες, όσο και για μια αναδυόμενη αποκεντρωμένη αυτόνομη έννοια γνωστή ως «αυτόνομη ταυτότητα» (self-sovereign identity) (Casey, 2018). Η βελτίωση της παραδοσιακής λειτουργίας της ταυτότητας μπορεί να επιτευχθεί με την μερική αποκεντροποίηση του συστήματος διαχείρισης ταυτότητας. Με αυτόν τον τρόπο, οι καθιερωμένες βιομηχανίες όπως είναι οι τράπεζες και το κράτος θα μπορούν να έχουν κοινή πρόσβαση στα πιστοποιητικά ταυτοποίησης χωρίς να είναι απαραίτητη η προσκόμιση τους από τους χρήστες. Η έννοια της «αυτόνομης ταυτότητας», αφορά σε ένα πλήρως αποκεντροποιημένο σύστημα όπου ο κάθε χρήστης θα μπορεί να ταυτοποιείται έχοντας τον έλεγχο των πληροφοριών που χρησιμοποιούνται για την ταυτοποίηση του.

### 3.7.4 Διαδίκτυο της Αξίας (The Internet of Value)

Το Διαδίκτυο της Αξίας μπορεί να οριστεί συνοπτικά ως η άμεση μεταφορά περιουσιακών στοιχείων τα οποία μπορούν να εκφραστούν σε νομισματικούς όρους μέσω του Διαδικτύου μεταξύ ομότιμων χρηστών χωρίς την ανάγκη ενδιάμεσων (Treiblmaier, 2022). Στην ουσία, αποτελεί την εξέλιξη του Διαδικτύου των Πραγμάτων (Internet of Things) όπου οι έξυπνες συσκευές θα μπορούν να ανταλλάσσουν μεταξύ τους αξίες οι οποίες θα μπορούν να είναι πολύ μικρές. Η χρήση του Διαδικτύου των Πραγμάτων θα μπορούσε στο άμεσο μέλλον να φέρει την επανάσταση σε όλους τους τομείς στους οποίους μπορούν να εφαρμοστούν προγράμματα επιβράβευσης χρησιμοποιώντας μικροσυναλλαγές, όπως είναι το μάρκετινγκ.

### 3.7.5 Εφοδιαστικές αλυσίδες

Με την εφαρμογή της τεχνολογίας των αλυσίδων συστοιχιών στις αλυσίδες εφοδιασμού μπορεί να αυξηθεί σημαντικά η διαφάνεια, η ιχνηλασιμότητα και η αποτελεσματικότητα τους. Αυτό μπορεί να συμβεί με την χρήση ενός μερικώς αποκεντροποιημένου συστήματος αλυσίδων συστοιχιών, όπου οι εμπλεκόμενοι στην αλυσίδα εφοδιασμού θα μπορούν να μοιράζονται και να παρακολουθούν με ασφάλεια το σύνολο των δεδομένων της εφοδιαστικής αλυσίδας, συμπεριλαμβανομένης της προέλευσης του προϊόντος, της διαδικασίας παραγωγής και της μεταφοράς.

Η αύξηση της διαφάνειας και της ιχνηλασιμότητας μπορεί να εξασφαλισθεί, καθώς κάθε συναλλαγή σε μια αλυσίδα συστοιχιών καταγράφεται και δεν μπορεί να τροποποιηθεί ή να

διαγραφεί, παρέχοντας ένα μόνιμο και απαραβίαστο αρχείο της διαδρομής του προϊόντος μέσω της αλυσίδας εφοδιασμού. Με αυτόν τον τρόπο μειώνονται οι πιθανότητες απάτης και διασφαλίζεται η ποιότητα και η ασφάλεια των προϊόντων.

Για την βελτίωση της αποτελεσματικότητας, μπορούν να χρησιμοποιηθούν τα έξυπνα συμβόλαια και ένα αποκεντροποιημένο λογιστικό βιβλίο. Με την χρήση των έξυπνων συμβολαίων θα είναι δυνατή η αυτόματη ενεργοποίηση πληρωμών και η αποδέσμευση αγαθών βάσει προκαθορισμένων κριτηρίων, όπως η επιβεβαίωση της παράδοσης ή οι έλεγχοι ποιοτικού ελέγχου. Ενώ παράλληλα, με την χρήση των αποκεντροποιημένων λογιστικών βιβλίων θα είναι δυνατή η ανοικοδόμηση σχέσεων εμπιστοσύνης μεταξύ των εμπλεκόμενων μερών.

### 3.7.6 Χρηματιστήριο

Η χρήση της τεχνολογίας αλυσίδων συστοιχιών μπορεί να βελτιώσει την αποτελεσματικότητα του Χρηματιστηρίου οδηγώντας σε μια πιο αξιόπιστη και προσβάσιμη αγορά για τους επενδυτές. Η βελτίωση της αποτελεσματικότητας επιτυγχάνεται με την αύξηση της ασφάλειας και της διαφάνειας, καθώς και με την μείωση του κόστους και του χρόνου διεκπεραίωσης των συναλλαγών. Η αύξηση της ασφάλειας και της διαφάνειας μπορεί να επιτευχθεί, καθώς με τη χρήση αλυσίδων συστοιχιών κάθε συναλλαγή καταγράφεται σε πραγματικό χρόνο χωρίς να μπορεί να τροποποιηθεί ή να διαγραφεί, ενώ ταυτόχρονα είναι ορατή από οποιονδήποτε έχει πρόσβαση στην εν λόγω αλυσίδα. Αυτό έχει ως αποτέλεσμα την αποτροπή των συναλλαγών εκ των έσω, την μείωση του κινδύνου απάτης και πειρατείας και γενικότερα την αύξηση της εμπιστοσύνης στο χρηματιστήριο. Η μείωση του κόστους και του χρόνου διεκπεραίωσης των συναλλαγών επιτυγχάνεται, καθώς με τη χρήση της τεχνολογίας αλυσίδων συστοιχιών δεν θα είναι απαραίτητη η ύπαρξη μεσαζόντων, όπως οι μεσίτες και τα γραφεία εκκαθάρισης. Αυτές οι μειώσεις συμβάλλουν καθοριστικά στην μείωση του κινδύνου αποτυχιών διακανονισμού και στη βελτίωση της ρευστότητας στην αγορά.

### 3.7.7 Χρηματοδότηση του εμπορίου

Με την χρήση της τεχνολογίας αλυσίδων συστοιχιών μπορεί να βελτιωθεί καθοριστικά η χρηματοδότηση του εμπορίου αυξάνοντας τη διαφάνεια, την ασφάλεια και την αποτελεσματικότητα της διαδικασίας. Η αποτελεσματικότητα της διαδικασίας μπορεί να βελτιωθεί με την χρήση έξυπνων συμβάσεων. Πιο αναλυτικά, με την χρήση έξυπνων συμβάσεων θα είναι δυνατή η αυτοματοποίηση της διαδικασίας χρηματοδότησης του εμπορίου, μειώνοντας έτσι την ανάγκη για μεσάζοντες και γραφειοκρατία. Για παράδειγμα, μπορεί να δημιουργηθεί ένα έξυπνο συμβόλαιο μεταξύ ενός αγοραστή και ενός πωλητή, όπου η πληρωμή αποδεσμεύεται μόνο όταν πληρούνται οι όροι του έξυπνου συμβολαίου. Με τον τρόπο αυτό εξαλείφεται η ανάγκη να ενεργεί μια τράπεζα ως μεσάζων στη συναλλαγή, μειώνοντας το κόστος, αυξάνοντας την ταχύτητα και κατά συνέπεια τις ταμειακές ροές των επιχειρήσεων και επομένως συνολικά την αποτελεσματικότητα. Η αύξηση της διαφάνειας και της ασφάλειας επιτυγχάνεται, καθώς με την χρήση αλυσίδων συστοιχιών κάθε εμπορικό έγγραφο καταγράφεται σε πραγματικό χρόνο χωρίς να μπορεί να τροποποιηθεί ή να διαγραφεί, ενώ ταυτόχρονα είναι ορατό από οποιονδήποτε έχει πρόσβαση στην εν λόγω αλυσίδα. Με αυτόν τον τρόπο αυξάνεται η εμπιστοσύνη μεταξύ των μερών και μειώνεται ο κίνδυνος απάτης.

### 3.7.8 Ψηφιακά Συστήματα Μετρητών (Digital Cash Systems)

Η πιο γνωστή εφαρμογή της τεχνολογίας αλυσίδων συστοιχιών υπήρξε στην δημιουργία αποκεντροποιημένων συστημάτων μετρητών, για την υλοποίηση των οποίων χρησιμοποιήθηκαν ψηφιακά νομίσματα, τα οποία ονομάστηκαν κρυπτονομίσματα «cryptocurrencies». Το πρώτο κρυπτονόμισμα που δημιουργήθηκε ήταν το «Bitcoin», και στη συνέχεια δημιουργήθηκαν πολλά ακόμη τα οποία αν και βασίστηκαν στην τεχνολογία του «Bitcoin» έχουν αρκετές διαφορές μεταξύ τους. Παρακάτω αναλύονται τα πιο σημαντικά.

#### 3.7.8.1 Bitcoin

Το κρυπτονόμισμα «Bitcoin» είναι το πρώτο κρυπτονόμισμα που δημιουργήθηκε το 2009 από ένα άγνωστο άτομο ή ομάδα ατόμων με το ψευδώνυμο «Satoshi Nakamoto». Το κρυπτονόμισμα αυτό, αποτελεί ένα πλήρως αποκεντρωμένο ψηφιακό νόμισμα το οποίο χρησιμοποιεί ένα ομότιμο δίκτυο για την πραγματοποίηση των συναλλαγών μεταξύ των χρηστών χωρίς την ανάγκη κεντρικής αρχής ή ενδιάμεσου φορέα. Στην ουσία αποτελεί το πρώτο επιτυχημένο ολοκληρωμένο σύστημα ψηφιακών μετρητών (digital cash system).

Για την υλοποίηση αυτού του συστήματος ψηφιακών μετρητών χρησιμοποιείται η τεχνολογία των αλυσίδων συστοιχιών χωρίς άδεια. Κάθε συναλλαγή επαληθεύεται από τους κόμβους του δικτύου μέσω κρυπτογραφίας κάνοντας χρήση της συνάρτησης κατακερματισμού «SHA-256» και καταγράφεται σε ένα καθολικό κατανεμημένης ιδιοκτησίας. Για την καταγραφή των συναλλαγών στο καθολικό, χρησιμοποιούνται τα μπλοκ. Κάθε μπλοκ μπορεί να έχει μέγιστη χωρητικότητα 1mb με συνέπεια να μπορεί να περιέχει περίπου 3 με 4 χιλιάδες συναλλαγές. Η καταγραφή κάθε μπλοκ στο καθολικό πραγματοποιείται κάθε 10 λεπτά με τη χρήση μοντέλου απόδειξης εργασίας (proof of work) και της προσαρμογής της μονοεπιλογής (nonce), έτσι ώστε ανεξάρτητα της υπολογιστικής ισχύς που χρησιμοποιείται για την απόδειξη εργασίας, να διατηρείται αυτό το χρονικό διάστημα.

Οι χρήστες που προσθέτουν μπλοκ στο καθολικό ανταμείβονται με έναν αριθμό κρυπτονομισμάτων «Bitcoin» καθώς και τις προμήθειες των συναλλαγών που περιέχονται στο μπλοκ που καταγράφεται στο καθολικό. Αυτή τη στιγμή, τα κόστη συναλλαγών αν και τεχνικά είναι προαιρετικά, στην πραγματικότητα είναι απαραίτητα για την πραγματοποίηση μιας συναλλαγής καθώς στα μπλοκ εισάγονται οι συναλλαγές που περιέχουν τις μεγαλύτερες προμήθειες. Αυτό θα γίνει ιδιαίτερα αισθητό με την παραγωγή 21 εκατομμυρίων «Bitcoin», όπου η ανταμοιβή για την καταγραφή ενός μπλοκ στο καθολικό θα μηδενιστεί, και η μόνη ανταμοιβή των χρηστών που προσθέτουν μπλοκ θα είναι οι προμήθειες συναλλαγών.

Το κρυπτονόμισμα «Bitcoin» αν και για τεχνικούς λόγους, με τον βασικότερο από τους οποίους να είναι ο περιορισμένος αριθμός συναλλαγών ανά μπλοκ, από μόνο του δεν μπορεί να αποτελέσει σύστημα πληρωμών του πλανήτη. Μπορεί να αποτελέσει όμως ένα πολύτιμο και άκρως αξιόπιστο περιουσιακό στοιχείο πάνω στο οποίο θα εφαρμόζονται πρωτόκολλα τα οποία θα του δίνουν την δυνατότητα να αποτελέσει μέρος ενός πλήρους αποκεντροποιημένου συστήματος πληρωμής μετρητών. Όσον αφορά την αξιοπιστία του, το κρυπτονόμισμα «Bitcoin» έχει τεράστιο πλεονέκτημα σε σχέση με τα υπόλοιπα κρυπτονομίσματα, καθώς ακόμη και σήμερα οι δημιουργοί του παραμένουν άγνωστοι με αποτέλεσμα το σύστημα να είναι απολύτως αποκεντροποιημένο χωρίς να επηρεάζεται από συγκεκριμένο πρόσωπο και ιδρύματα. Με αυτό τον τρόπο, η νομισματική πολιτική του κρυπτονομίσματος «Bitcoin» παραμένει υπόθεση των χρηστών του δικτύου χωρίς να κατευθύνονται από την άποψη των δημιουργών του, όπως συμβαίνει σε άλλα κρυπτονομίσματα όπως είναι το «Etherium».

#### 3.7.8.2 Bitcoin Cash

Το κρυπτονόμισμα «Bitcoin Cash» δημιουργήθηκε τον Αύγουστο του 2017 με την πραγματοποίηση σκληρής διακλάδωσης (hard fork) από την αλυσίδα συστοιχιών του κρυπτονομίσματος «Bitcoin». Η σκληρή διακλάδωση ξεκίνησε από μια ομάδα προγραμματιστών που ήθελαν να αυξήσουν το όριο μεγέθους μπλοκ από 1 MB σε 8 MB, έτσι ώστε να μεγαλώσουν την χωρητικότητα των μπλοκ σε συναλλαγές. Με αυτόν τον τρόπο αποσκοπούσαν στην αύξηση της ταχύτητας των συναλλαγών και στη μείωση των προμηθειών. Εκτός από την αύξηση του ορίου μεγέθους των μπλοκ, η υλοποίηση του συστήματος είναι ίδια με αυτή του κρυπτονομίσματος «Bitcoin».

#### 3.7.8.3 Litecoin

Το κρυπτονόμισμα «Litecoin» δημιουργήθηκε από τον Charlie Lee, έναν πρώην μηχανικό της Google, και σχεδιάστηκε για να είναι μια "ελαφριά" έκδοση του κρυπτονομίσματος «Bitcoin», με ταχύτερους χρόνους συναλλαγών και χαμηλότερα κόστη συναλλαγών. Πιο αναλυτικά, το κρυπτονόμισμα «Litecoin» σε επίπεδο υλοποίησης διαφέρει σε σχέση με το κρυπτονόμισμα «Bitcoin» ως προς το όριο μεγέθους μπλοκ, όπου το κρυπτονόμισμα «Litecoin» έχει όριο 4 MB ενώ το κρυπτονόμισμα «Bitcoin» έχει 1 MB, ως προς τον αλγόριθμο κατακερματισμού, όπου το κρυπτονόμισμα «Litecoin» χρησιμοποιεί την συνάρτηση κατακερματισμού «Scrypt» ενώ το κρυπτονόμισμα «Bitcoin» χρησιμοποιεί την συνάρτηση κατακερματισμού «SHA-256», ως προς το χρονικό διάστημα μεταξύ καταγραφής δύο μπλοκ στο καθολικό, όπου στο κρυπτονόμισμα «Litecoin» είναι 2,5 λεπτά ενώ στο κρυπτονόμισμα «Bitcoin» είναι 10 λεπτά και τέλος ως προς τον αριθμό των μέγιστων παραγόμενων κρυπτονομισμάτων, όπου στο κρυπτονόμισμα «Litecoin» είναι 84 εκατομμύρια ενώ στο κρυπτονόμισμα «Bitcoin» είναι 21 εκατομμύρια.

#### 3.7.8.4 Ethereum

Το κρυπτονόμισμα «Ethereum» είναι μια αποκεντρωμένη πλατφόρμα που βασίζεται στην τεχνολογία των αλυσίδων συστοιχιών και επιτρέπει τη δημιουργία και εκτέλεση έξυπνων συμβολαίων και αποκεντρωμένων εφαρμογών (dApps). Δημιουργήθηκε από τον Vitalik Buterin το 2013, ξεκίνησε το 2015 και μεταβλήθηκε ουσιαστικά τον Σεπτέμβριο του 2022 όταν και ολοκληρώθηκε η μετάβαση από μοντέλο απόδειξης εργασίας (proof of work model) σε μοντέλο απόδειξης πονταρίσματος (proof of stake model). Αυτή η αλλαγή είχε ως συνέπεια την σημαντική διαφοροποίηση του τρόπου υλοποίησης σε σχέση με τα υπόλοιπα κρυπτονομίσματα που κατά κύριο λόγο χρησιμοποιούν το μοντέλο απόδειξης εργασίας. Αυτή η διαφοροποίηση δίνει δύο βασικά πλεονεκτήματα στο κρυπτονόμισμα «Ethereum», την μείωση της κατανάλωσης των πόρων και την αύξηση της ταχύτητας πραγματοποίησης των συναλλαγών.

Σε επίπεδο υλοποίησης, εκτός από την χρήση του μοντέλου συναίνεσης «Proof of Stake» που αναλύθηκε παραπάνω, στο κρυπτονόμισμα «Ethereum» παράγονται κάθε χρόνο για εξόρυξη κρυπτονομίσματα με ποσοστό πληθωρισμού περίπου 4% χωρίς να υπάρχει κάποιο άνω όριο, όπως υπάρχει στο κρυπτονόμισμα «Bitcoin». Τέλος, ως συνάρτηση κατακερματισμού χρησιμοποιείται ο αλγόριθμος «Keccak-256» ο οποίος αποτελεί μία παραλλαγή της συνάρτησης «SHA-3».

Το κρυπτονόμισμα «Ethereum», αν και αποτελεί ένα αποκεντροποιημένο κρυπτονόμισμα, η νομισματική του πολιτική αλλά και ο τρόπος λειτουργίας καθορίζεται μέσω μιας κυβερνητικής

διαδικασίας στην οποία ρόλο «κλειδί» διαδραματίζει το ίδρυμα «Ethereum Foundation» (EF), το οποίο ιδρύθηκε από μια ομάδα ατόμων συμπεριλαμβανομένων των Vitalik Buterin, Gavin Wood και Joseph Lubin, οι οποίοι συμμετείχαν στην αρχική ανάπτυξη της πλατφόρμας «Ethereum». Το ίδρυμα «Ethereum Foundation» αποτελεί έναν μη κερδοσκοπικό οργανισμό αφιερωμένο στην υποστήριξη του «Ethereum» και των σχετικών τεχνολογιών ο οποίος κατέχει περίπου το 0,297% των κρυπτονομισμάτων «Ethereum» καθώς και άλλα περιουσιακά στοιχεία, το σύνολο των οποίων μαζί με την αξία των κρυπτονομισμάτων που κατέχει υπολογίζονται κοντά στο 1 δις δολάρια. Έτσι, το ίδρυμα έχει τους πόρους τόσο για να εξελίξει το κρυπτονόμισμα «Ethereum» όσο και για να προωθήσει τις αλλαγές που επιθυμεί στο κρυπτονόμισμα ώστε να μπορέσει να ανταγωνιστεί αποτελεσματικά στο μέλλον τα υπάρχοντα κεντροποιημένα συστήματα πληρωμής μετρητών (Centralized Cash Systems).

#### 3.7.8.5 Ethereum Classic

Το κρυπτονόμισμα «Ethereum Classic» δημιουργήθηκε το 2016 με την πραγματοποίηση σκληρής διακλάδωσης (hard fork) από την αλυσίδα συστοιχιών του κρυπτονομίσματος «Ethereum». Η διακλάδωση ήταν αποτέλεσμα μιας διαφωνίας εντός της κοινότητας του «Ethereum» σχετικά με την απόφαση για την αντιστροφή του hack DAO που είχε ως αποτέλεσμα την απώλεια κρυπτονομισμάτων «Ethereum» αξίας εκατομμυρίων δολαρίων. Αυτή την στιγμή το κρυπτονόμισμα «Ethereum Classic» συνεχίζει να υπάρχει, χωρίς όμως την υποστήριξη του ιδρύματος «Ethereum Foundation», με αποτέλεσμα να έχει σταματήσει να εξελίσσεται.

# ΚΕΦΑΛΑΙΟ 4: ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ (MACHINE LEARNING)

## 4.1 Εισαγωγή

Η Μηχανική Μάθηση είναι ένα υποπεδίο της επιστήμης των υπολογιστών που δίνει την ικανότητα στους υπολογιστές να μαθαίνουν από δεδομένα χωρίς να είναι ρητά προγραμματισμένοι γι' αυτό (Arthur Samuel, 1959). Πιο αναλυτικά, αποτελεί ένα εργαλείο εξαγωγής γνώσης από δεδομένα, ενώ παράλληλα εκτός από υποπεδίο της επιστήμης των υπολογιστών, αποτελεί υποπεδίο της τεχνητής νοημοσύνης και είναι στενά συνδεδεμένη με την στατιστική και την βελτιστοποίηση.

## 4.2 Είδη Συστημάτων Μηχανικής Μάθησης

Υπάρχουνε πολλά και διαφορετικά είδη συστημάτων μηχανικής μάθησης τα οποία μπορούν να κατηγοριοποιηθούν βάσει της ανθρώπινης εποπτείας κατά την εκπαίδευση των συστημάτων, βάσει του τρόπου ενσωμάτωσης των νέων δεδομένων που είναι διαθέσιμα για εκπαίδευση και τέλος βάσει των τρόπων εύρεσης συσχετίσεων μεταξύ των χαρακτηριστικών.

### 4.2.1 Κατηγοριοποίηση βάσει της ανθρώπινης εποπτείας

Τα συστήματα μηχανικής μάθησης μπορούν να κατηγοριοποιηθούν βάσει του τρόπου και της έντασης της ανθρώπινης εποπτείας κατά την εκπαίδευση τους σε μάθηση με επίβλεψη (supervised learning), σε μάθηση χωρίς επίβλεψη (unsupervised learning), σε μάθηση με ημι-επίβλεψη (semi-supervised learning) και σε ενισχυτική μάθηση (reinforcement learning).

#### 4.2.1.1 Μάθηση με Επίβλεψη (Supervised Learning)

Στην μάθηση με επίβλεψη τα δεδομένα εκπαίδευσης περιέχουν εκτός από τις τιμές εισόδου και τις τιμές στόχους οι οποίες ονομάζονται ετικέτες (labels). Δηλαδή, τα δεδομένα εκπαίδευσης περιλαμβάνουν τις τιμές των χαρακτηριστικών και την επιθυμητή τιμή εξόδου. Με την χρήση αυτών των δεδομένων το σύστημα προσπαθεί να «μάθει» επαγωγικά μία συνάρτηση η οποία έχει ως διάνυσμα εισόδου τις τιμές των χαρακτηριστικών και ως έξοδο την επιθυμητή τιμή εξόδου. Στην μάθηση εμφανίζονται δύο ευρείς κατηγορίες προβλημάτων, τα προβλήματα ταξινόμησης (classification) όπου γίνεται πρόβλεψη διακριτών τάξεων, και τα προβλήματα παρεμβολής (regression) όπου πραγματοποιείται πρόβλεψη μιας αριθμητικής τιμής.

#### 4.2.1.2 Μάθηση χωρίς Επίβλεψη (Unsupervised Learning)

Στην μάθηση χωρίς επίβλεψη το σύστημα δέχεται ως είσοδο δεδομένα εκπαίδευσης τα οποία δεν περιέχουν τιμές στόχους, δηλαδή ετικέτες. Αυτό συμβαίνει καθώς αυτά τα συστήματα προσπαθούν να ανακαλύψουν συσχετίσεις και ομάδες μεταξύ των δεδομένων βάσει κάποιας ομοιότητας των χαρακτηριστικών τους. Στην μάθηση χωρίς επίβλεψη τα προβλήματα μπορούν να κατηγοριοποιηθούν σε δύο κατηγορίες, στα προβλήματα συσταδοποίησης (clustering),

όπου τα δεδομένα εισόδου χωρίζονται σε άγνωστες εκ των προτέρων ομάδες, το πλήθος των οποίων καθορίζεται από τον χρήστη, και στα προβλήματα μείωσης διαστασιμότητας (dimentionality reduction), όπου τα δεδομένα εισόδου αντιστοιχίζονται σε ένα χώρο με λιγότερες διαστάσεις.

#### 4.2.1.3 Μάθηση με Ημι-επίβλεψη (Semi-supervised Learning)

Στην μάθηση με ημι-επίβλεψη το σύστημα έχει ως είσοδο δεδομένα εκπαίδευσης που περιέχουν τιμές στόχους και δεδομένα που δεν περιέχουν τιμές στόχους. Συνήθως τα δεδομένα χωρίς τιμές στόχους, δηλαδή χωρίς ετικέτες, αποτελούν την μεγαλύτερη πλειοψηφία. Αυτό συμβαίνει καθώς αυτά τα συστήματα χρησιμοποιούν τα δεδομένα εκπαίδευσης με ετικέτες για τον καθορισμό των ομάδων των δεδομένων και στη συνέχεια κατηγοριοποιούν τα δεδομένα χωρίς ετικέτες σε αυτές τις ομάδες. Δηλαδή στην ουσία η μάθηση με ημι-επίβλεψη αποτελεί τον συνδυασμό της μάθησης χωρίς επίβλεψης και της μάθησης με επίβλεψη.

#### 4.2.1.4 Ενισχυτική Μάθηση (Reinforcement Learning)

Η ενισχυτική μάθηση αποτελεί μία τελείως διαφορετική προσέγγιση από τις προηγούμενες. Το σύστημα καλείται πράκτορας (agent) ο οποίος παρατηρεί το περιβάλλον και πραγματοποιεί ενέργειες (actions). Η κάθε ενέργεια που πραγματοποιεί του επιφέρει μία ανταμοιβή ή ένα πέναλτι. Με αυτόν τον τρόπο μαθαίνει ποια είναι η καλύτερη μακροχρόνια στρατηγική, η οποία ονομάζεται και πολιτική (policy), έτσι ώστε να λάβει τις περισσότερες ανταμοιβές με την πάροδο του χρόνου. Η κάθε πολιτική καθορίζει τις ενέργειες που πρέπει να πραγματοποιήσει ένας πράκτορας σε μία δεδομένη κατάσταση.

### 4.2.2 Κατηγοριοποίηση με βάση τον τρόπο εκπαίδευσης

Τα συστήματα μηχανικής μάθησης ανάλογα με το εάν μπορούν ή όχι να μαθαίνουν αυξητικά από μία ροή εισερχόμενων δεδομένων, κατηγοριοποιούνται σε συστήματα μαζικής μάθησης (batch learning) και σε συστήματα απευθείας μάθησης (online learning).

#### 4.2.2.1 Μαζική μάθηση (Batch Learning)

Στην μαζική μάθηση το σύστημα δεν είναι ικανό να μαθαίνει αυξητικά, δηλαδή πρέπει να εκπαιδεύεται κάθε φορά με τη χρήση του συνόλου των διαθέσιμων δεδομένων. Αυτό οδηγεί το σύστημα να μαθαίνει ουσιαστικά εκτός σύνδεσης (offline), καθώς κατά κανόνα η διαδικασία της μάθησης καταναλώνει πολύ χρόνο και υπολογιστικές πηγές. Αυτό πρακτικά σημαίνει ότι εάν ένα σύστημα μαζικής μάθησης πρέπει να εκπαιδευτεί σε νέα δεδομένα θα πρέπει να εκπαιδευτεί μία νέα έκδοση του συστήματος από το μηδέν στο σύνολο των δεδομένων και στη συνέχεια να σταματήσει η παλιά έκδοση του συστήματος και να αντικατασταθεί από την καινούργια. Η εύκολη αυτοματοποίηση αυτής της διαδικασίας δίνει την δυνατότητα στα συστήματα μαζικής μάθησης να υιοθετούν τις αλλαγές όσο συχνά χρειάζεται κάθε σύστημα, με τον περιορισμό φυσικά του απαραίτητου χρόνου και υπολογιστικής ισχύς που χρειάζεται το σύστημα έτσι ώστε να εκπαιδευτεί από την αρχή. Οπότε, από τα παραπάνω γίνεται εύκολα αντιληπτό ότι τα συστήματα μαζικής μάθησης προτιμώνται όταν τα διαθέσιμα δεδομένα για εκπαίδευση δεν είναι πολλά και όταν το περιβάλλον του συστήματος δεν μεταβάλλεται συχνά.

#### 4.2.2.2 Απευθείας Μάθηση (Online Learning)

Στην απευθείας μάθηση το σύστημα εκπαιδεύεται αυξητικά τροφοδοτώντας το με τα δεδομένα εκπαίδευσης διαδοχικά, είτε μεμονωμένα είτε σε μικρές ομάδες, οι οποίες αποκαλούνται μίνι-παρτίδες (mini-batches). Με αυτόν τον τρόπο κάθε βήμα της εκπαίδευσης είναι γρήγορο και φθηνό με αποτέλεσμα το σύστημα να εκπαιδεύεται «εν κινήσει». Αυτός ο τρόπος μάθησης είναι ιδανικός για συστήματα που λαμβάνουν δεδομένα σε συνεχή ροή και πρέπει να προσαρμοστούν άμεσα σε αυτά, για συστήματα που διαθέτουν περιορισμένους υπολογιστικούς πόρους και για συστήματα που λαμβάνουν μεγάλο όγκο δεδομένων και πρέπει να τον διαχειριστούν τμηματικά. Τέλος, μία πολύ σημαντική παράμετρος των συστημάτων που χρησιμοποιούν την απευθείας μάθηση είναι η ταχύτητα προσαρμογής στα δεδομένα που αλλάζουν. Η παράμετρος αυτή ονομάζεται ρυθμός μάθησης (learning rate) και ανάλογα με την τιμή της προσαρμόζεται και η ταχύτητα. Πιο αναλυτικά, εάν ο ρυθμός μάθησης είναι υψηλός τότε το σύστημα προσαρμόζεται γρήγορα στα νέα δεδομένα «ξεχνώντας» πιο γρήγορα και τα παλιά δεδομένα, ενώ εάν ο ρυθμός μάθησης είναι χαμηλός, το σύστημα προσαρμόζεται αργά στα νέα δεδομένα και δεν ξεχνά γρήγορα τα παλιότερα. Στην περίπτωση που ο ρυθμός μάθησης είναι υψηλός, υπάρχει ο κίνδυνος το σύστημα να έχει μεγάλες διακυμάνσεις στην απόδοση του λόγω κακών νέων δεδομένων εκπαίδευσης. Για να αντιμετωπιστεί αυτή η κατάσταση, το σύστημα πρέπει να γυρίσει σε κάποια προηγούμενη έκδοση που δεν περιλαμβάνει στην εκπαίδευση τα προβληματικά δεδομένα.

### 4.2.3 Κατηγοριοποίηση με βάση τον τρόπο γενίκευσης

Ένας ακόμη τρόπος με τον οποίο μπορούν να κατηγοριοποιηθούν τα συστήματα μηχανικής μάθησης είναι με βάση τον τρόπο που γενικεύουν (Aurélien Géron, 2019). Πιο αναλυτικά, σκοπός των συστημάτων είναι να μπορούν να πραγματοποιούν ακριβείς προβλέψεις πάνω σε άγνωστα δεδομένα. Αυτό σημαίνει ότι θα πρέπει να εκπαιδεύονται πάνω στα δεδομένα εκπαίδευσης αλλά ταυτόχρονα να μπορούν να γενικεύουν σε άγνωστα δεδομένα. Για τη γενίκευση υπάρχουν δύο κύριες προσεγγίσεις, η μάθηση βασισμένη σε στιγμιότυπα (instace-based learning) και η μάθηση βασισμένη σε μοντέλο (model-based learning) (Aurélien Géron, 2019).

#### 4.2.3.1 Μάθηση βασισμένη σε στιγμιότυπα (Instance-based learning)

Τα συστήματα που χρησιμοποιούν την μάθηση βασισμένη σε στιγμιότυπα, μαθαίνουν απομνημονεύοντας συγκεκριμένα παραδείγματα δεδομένων τα οποία και χρησιμοποιούν για να πραγματοποιήσουν νέες προβλέψεις. Πιο συγκεκριμένα, χρησιμοποιούν μέτρα ομοιότητας για να συγκρίνουν το διάνυσμα των νέων δεδομένων εισόδου με τα απομνημονευμένα δεδομένα, έτσι ώστε να πραγματοποιήσουν την γενίκευση, δηλαδή την κατηγοριοποίηση ή την πρόβλεψη των δεδομένων εισόδου. Ένα πολύ χαρακτηριστικό παράδειγμα αλγορίθμου που χρησιμοποιεί την μάθηση βασισμένη σε στιγμιότυπα είναι k-κοντινότεροι γείτονες (k-NN).

#### 4.2.3.2 Μάθηση βασισμένη σε μοντέλο (Instance-based learning)

Τα συστήματα που χρησιμοποιούν τη μάθηση βασισμένη σε μοντέλο γενικεύουν από ένα σετ παραδειγμάτων χτίζοντας αρχικά ένα μοντέλο από αυτά τα δεδομένα, το οποίο στη συνέχεια το χρησιμοποιούν για να πραγματοποιήσουν προβλέψεις. Με αυτόν τον τρόπο δίνεται έμφαση στην κατανόηση της υποκείμενης δομής των δεδομένων και όχι απλώς στην εύρεση μοτίβων σε αυτά. Ειδικότερα, ο στόχος είναι η δημιουργία ενός αποτελεσματικού μοντέλου από τα

δεδομένα εκπαίδευσης που να αποτυπώνει την υποκείμενη σχέση μεταξύ των μεταβλητών εισόδου και των μεταβλητών εξόδου έτσι ώστε να πραγματοποιεί ακριβείς προβλέψεις. Η επιλογή του κατάλληλου μοντέλου παίζει καθοριστικό ρόλο στην απόδοση του συστήματος, καθώς αν επιλεχθεί για παράδειγμα μία γραμμική συνάρτηση είναι πολύ πιθανό να μην μπορεί να μοντελοποιήσει τα δεδομένα και να είναι απαραίτητη η χρήση μιας πολυωνυμικής συνάρτησης. Από αυτό προκύπτει ότι για την κρίσιμη επιλογή του μοντέλου είναι απαραίτητη, εκτός από την σχολαστική μελέτη των δεδομένων εκπαίδευσης, και η εμπειρία στον τομέα της μοντελοποίησης και της στατιστικής. Μερικές από τις πιο συχνά χρησιμοποιούμενες τεχνικές είναι η ανάλυση παλινδρόμησης, η Μπεϋζιανή μοντελοποίηση, τα δέντρα αποφάσεων και τα νευρωνικά δίκτυα.

### 4.2.4 Κατηγοριοποίηση με βάση την παρεμβατικότητα του χρήστη κατά την εκπαίδευση

Ένα άλλο κριτήριο με βάση το οποίο μπορούν να κατηγοριοποιηθούν τα συστήματα μηχανικής μάθησης είναι ανάλογα με τον τρόπο που συμμετέχει ο χρήστης στην εκπαίδευση του συστήματος. Ανάλογα με τον τρόπο αυτό, τα συστήματα μπορούν να κατηγοριοποιηθούν σε ενεργά (active) και παθητικά (passive).

#### 4.2.4.1 Ενεργά συστήματα (Active systems)

Τα συστήματα μηχανικής μάθησης στα οποία ο χρήστης παρεμβαίνει άμεσα κατά την διάρκεια της εκπαίδευσης τους ονομάζονται ενεργά. Αναλυτικότερα, σε αυτά τα συστήματα η διαδικασία της εκπαίδευσης ξεκινάει με τη χρήση ενός μικρού συνόλου επισημασμένων δεδομένων και στη συνέχεια επιλέγονται επαναληπτικά από το σύστημα πρόσθετα παραδείγματα για να επισημανθούν από τον χρήστη και να μοντελοποιηθούν από το σύστημα. Αυτός ο τύπος συστημάτων απαιτεί πολλούς υπολογιστικούς και ανθρώπινους πόρους, καθώς η διαδικασία της εκπαίδευσης είναι πολύπλοκη και απαιτεί πολύ εξειδικευμένο ανθρώπινο δυναμικό για την επισήμανση των νέων δεδομένων εκπαίδευσης.

#### 4.2.4.2 Παθητικά συστήματα (Passive systems)

Τα παθητικά συστήματα μηχανικής μάθησης σε αντίθεση με τα ενεργά, μαθαίνουν από δεδομένα χωρίς να απαιτούν την παρέμβαση από τον χρήστη. Πιο συγκεκριμένα, σε αυτά τα συστήματα ο χρήστης συλλέγει τα δεδομένα πριν την εκπαίδευση, τα προεπεξεργάζεται και τα δίνει ως είσοδο στο σύστημα προκειμένου να εκπαιδευτεί χωρίς να χρειάζεται στη συνέχεια τον χρήστη. Τα συστήματα αυτά είναι ιδανικά για τις περιπτώσεις στις οποίες ο όγκος των δεδομένων είναι πολύ μεγάλος και οι σχέσεις μεταξύ των δεδομένων πολύ πολύπλοκες με αποτέλεσμα να μην είναι εμφανείς στον χρήστη.

## 4.3 Οι μεγαλύτερες προκλήσεις στην Μηχανική Μάθηση

Στην Μηχανική Μάθηση από την στιγμή που οι βασικές εργασίες είναι η επιλογή του αλγορίθμου μάθησης και η εκπαίδευση του πάνω σε μερικά δεδομένα, τα δύο μέρη που μπορούν να καταστραθούν προβληματικά είναι «ο κακός αλγόριθμος» και «τα κακά δεδομένα» (Aurélien Géron, 2019). Οπότε γίνεται εύκολα αντιληπτό ότι οι δύο κύριοι παράγοντες γύρω από τους οποίους κινούνται οι μεγαλύτερες προκλήσεις για τα συστήματα μηχανικής μάθησης

είναι οι αλγόριθμοι μάθησης και η ποιότητα των δεδομένων που χρησιμοποιούνται στην εκπαίδευση.

### 4.3.1 Μη αντιπροσωπευτικά δεδομένα εκπαίδευσης

Προκειμένου τα συστήματα μηχανικής μάθησης να γενικεύουν αποδοτικά, είναι απαραίτητο τα δεδομένα εκπαίδευσης να αντιπροσωπεύουν το σύνολο του πληθυσμού. Για παράδειγμα, εάν ένα σύστημα αναγνώρισης προσώπου εκπαιδεύεται σε πρόσωπα μιας συγκεκριμένης εθνικότητας είναι πολύ πιθανό να μην ανταποκρίνεται αποδοτικά σε πρόσωπα μιας άλλης εθνικότητας. Για να αντιμετωπιστεί αυτή η πρόκληση πρέπει να επιλέγονται προσεκτικά οι μέθοδοι συλλογής δεδομένων και δειγματοληψίας και να διασφαλίζεται ότι τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση είναι αντιπροσωπευτικά του πληθυσμού. Σε περίπτωση έλλειψης αντιπροσωπευτικών δεδομένων μπορούν να χρησιμοποιηθούν τεχνικές όπως η αύξηση δεδομένων και η μάθηση μεταφοράς για τη συμπλήρωση τους.

### 4.3.2 Ποιότητα Δεδομένων

Η ποιότητα των δεδομένων αποτελεί ίσως τον πιο καθοριστικό παράγοντα για την απόδοση των μοντέλων μηχανικής μάθησης, καθώς όπως σε μία συνταγή μαγειρικής, όσο καλή και να είναι, αν τα υλικά που χρησιμοποιούνται για την εκτέλεση της δεν είναι ποιοτικά το αποτέλεσμα θα είναι κακό, έτσι και στα συστήματα μηχανικής μάθησης, όσο καλά και να είναι, αν τους παρέχονται κακής ποιότητας δεδομένα το αποτέλεσμα, δηλαδή οι προβλέψεις, θα είναι κακές. Η ποιότητα των δεδομένων καθορίζεται κυρίως από τρεις βασικές παραμέτρους, την ποσότητα, την πληρότητα και τον θόρυβο που περιέχουν. Η μεγάλη ποσότητα των δεδομένων είναι απαραίτητη καθώς σε αντίθεση με την ανθρώπινη μάθηση, τα συστήματα τεχνητής νοημοσύνης χρειάζονται πληθώρα δεδομένων για να επιλύσουν ακόμη κι ένα πολύ απλό για τον άνθρωπο πρόβλημα. Η πληρότητα των δεδομένων επηρεάζει σημαντικά την αποτελεσματικότητα των συστημάτων σε άγνωστα δεδομένα, καθώς εάν δεν υπάρχουν δεδομένα εκπαίδευσης από το σύνολο του πληθυσμού καθιστά το σύστημα αδύνατο να γενικεύει σε άγνωστα δεδομένα. Τέλος, ο θόρυβος στα δεδομένα μπορεί να επηρεάσει πολύ αρνητικά την απόδοση των συστημάτων, καθώς εάν στα δεδομένα εκπαίδευσης υπάρχει πληθώρα ελλείψεων, λανθασμένων και ακραίων τιμών, είναι αδύνατη η μοντελοποίηση των δεδομένων εκπαίδευσης και κατά συνέπεια η γενίκευση σε άγνωστα δεδομένα.

### 4.3.3 Σχετικότητα χαρακτηριστικών

Η σχετικότητα των χαρακτηριστικών του συνόλου των δεδομένων εκπαίδευσης διαδραματίζει καθοριστικό ρόλο στην απόδοση των συστημάτων τεχνητής νοημοσύνης καθώς είναι δεδομένο ότι εάν «ταΐσεις» ένα σύστημα με σκουπίδια τότε και αυτό θα παράγει σκουπίδια. Οπότε λόγω της σημαντικότητας της επιλογής των σχετικών χαρακτηριστικών στα δεδομένα εισόδου των συστημάτων, υπάρχει μία ξεχωριστή διαδικασία η οποία ονομάζεται μηχανική χαρακτηριστικών (feature engineering). Αυτή η διαδικασία περιλαμβάνει την επιλογή των πιο χρήσιμων χαρακτηριστικών για εκπαίδευση από το σύνολο των χαρακτηριστικών, την παραγωγή πιο χρήσιμων χαρακτηριστικών από τον συνδυασμό των υφιστάμενων χαρακτηριστικών και τέλος την δημιουργία νέων χαρακτηριστικών όταν αυτό κρίνεται απαραίτητο συλλέγοντας νέα δεδομένα.

### 4.3.4 Υπερπροσαρμογή στα δεδομένα εκπαίδευσης

Η υπερπροσαρμογή (overfitting) στα δεδομένα εκπαίδευσης συμβαίνει όταν ένα σύστημα μηχανικής μάθησης «αποστηθίζει» τα δεδομένα εκπαίδευσης και δεν μαθαίνει από αυτά. Αυτό έχει ως αποτέλεσμα το σύστημα να ανταποκρίνεται άριστα στα δεδομένα εκπαίδευσης και πολύ αναποτελεσματικά στα νέα δεδομένα καθώς λόγω της «αποστήθισης» αδυνατεί να γενικεύσει σε αυτά. Για την αποφυγή της υπερπροσαρμογής χρησιμοποιούνται συνήθως τρεις βασικές τεχνικές, η ομαλοποίηση (regularization), η πρόωρη διακοπή και η αύξηση του πλήθους των δεδομένων εκπαίδευσης. Η ομαλοποίηση έχει ως στόχο τον περιορισμό του μοντέλου κατά την εκπαίδευση του με τη χρήση διάφορων υπερπαραμέτρων όπως είναι ο ρυθμός μάθησης (learning rate). Με την πρόωρη διακοπή η εκπαίδευση του μοντέλου διακόπτεται προτού προσαρμοστεί υπερβολικά το μοντέλο στα δεδομένα εκπαίδευσης. Τέλος, με την αύξηση του πλήθους των δεδομένων εκπαίδευσης επιτυγχάνεται η καλύτερη γενίκευση στα νέα δεδομένα, καθώς το σύστημα δεν «προλαβαίνει» να προσαρμοστεί στο σύνολο των δεδομένων.

### 4.3.5 Υποπροσαρμογή στα δεδομένα εκπαίδευσης

Η υποπροσαρμογή (underfitting) αποτελεί το αντίθετο της υπερπροσαρμογής, δηλαδή το σύστημα μηχανικής μάθησης αδυνατεί να μάθει από τα δεδομένα εκπαίδευσης, με αποτέλεσμα να μην έχει την δυνατότητα να πραγματοποιήσει αποτελεσματικές προβλέψεις. Αυτό συμβαίνει συνήθως, είτε γιατί ο αλγόριθμος μάθησης είναι ακατάλληλος για τον τύπο του προβλήματος, είτε γιατί η επιλογή των χαρακτηριστικών των δεδομένων εκπαίδευσης ήταν λανθασμένη, είτε γιατί το μοντέλο έχει περιοριστεί παραπάνω απ' όσο χρειάζεται κατά την διαδικασία της ομαλοποίησης (regularization). Όπως εύκολα προκύπτει, για την αντιμετώπιση της υποπροσαρμογής αρκεί να αντιμετωπιστούν οι παραπάνω αιτίες που την δημιουργούν, είτε με την επιλογή διαφορετικού, πιο κατάλληλου αλγορίθμου εκπαίδευσης, είτε με την τροφοδότηση του συστήματος με καλύτερα χαρακτηριστικά, είτε τέλος με την αποτελεσματικότερη επιλογή των υπερπαραμέτρων ομαλοποίησης.

## 4.4 Δοκιμή & Επικύρωση Συστημάτων Μηχανικής Μάθησης

Ο μόνος αποτελεσματικός τρόπος για τον υπολογισμό της αποδοτικότητας ενός συστήματος μηχανικής μάθησης είναι η δοκιμή της απόδοσής του σε νέα δεδομένα. Προκειμένου να συμβεί αυτό, το σύνολο των δεδομένων χωρίζεται σε δύο σύνολα, στο σύνολο εκπαίδευσης (training set) και στο σύνολο δοκιμής (test set). Όπως γίνεται εύκολα αντιληπτό από την ονομασία τους, το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση του συστήματος ενώ το σύνολο δοκιμής χρησιμοποιείται για την δοκιμή του συστήματος και κατά συνέπεια για τον υπολογισμό της απόδοσης του. Η απόδοση του συστήματος υπολογίζει το ποσοστό λάθους προβλέψεων που πραγματοποίησε το σύστημα στα δεδομένα δοκιμής. Το ποσοστό αυτό ονομάζεται σφάλμα γενίκευσης (regularization error) ή σφάλμα εκτός δείγματος (out-of-sample error).

Εκτός από το σύνολο εκπαίδευσης και το σύνολο δοκιμής υπάρχει και το σύνολο επικύρωσης (validation set) το οποίο αποτελεί υποσύνολο του συνόλου εκπαίδευσης και χρησιμοποιείται κατά τη διάρκεια εκπαίδευσης του συστήματος για την αποτελεσματικότερη επιλογή των υπερπαραμέτρων ομαλοποίησης (regularization hyperparameters). Δηλαδή, το σύνολο επικύρωσης χρησιμοποιείται ως σύνολο δοκιμής κατά τη διάρκεια της εκπαίδευσης του

μοντέλου για την βέλτιστη επιλογή των υπερπαραμέτρων εκπαίδευσής όπως είναι ο ρυθμός μάθησης (learning rate), ο αριθμός των στρωμάτων (number of layers) και το πλήθος των νευρώνων κάθε στρώματος.

Όπως γίνεται εύκολα κατανοητό, η δημιουργία του συνόλου επικύρωσης αποτελεί σπατάλη του συνόλου εκπαίδευσης, από το οποίο εξαρτάται άμεσα και η αποτελεσματικότητα ολόκληρου του συστήματος. Για την αποφυγή αυτής της σπατάλης χρησιμοποιείται η τεχνική της διασταυρούμενης επικύρωσης (cross-validation). Σύμφωνα με αυτή την τεχνική, το σύνολο εκπαίδευσης χωρίζεται σε συμπληρωματικά υποσύνολα και κάθε μοντέλο εκπαιδεύεται σε διαφορετικό συνδυασμό αυτών των υποσυνόλων και επικυρώνεται στα εναπομείναντα μέρη. Όταν επιλεχθούν ο τύπος του μοντέλου και οι υπερπαράμετροι, το τελικό μοντέλο εκπαιδεύεται στο σύνολο των δεδομένων εκπαίδευσης, δηλαδή δεν χρησιμοποιείται σύνολο επικύρωσης, και το σφάλμα γενίκευσης υπολογίζεται πάνω στο σύνολο δεδομένων δοκιμής.

## 4.5 Ανάλυση Συναισθημάτων (Sentiment Analysis) και Μηχανική Μάθηση

### 4.5.1 Εισαγωγή

Η ανάλυση συναισθημάτων είναι μια τεχνική επεξεργασίας φυσικής γλώσσας που χρησιμοποιείται για τον εντοπισμό και την εξαγωγή υποκειμενικών πληροφοριών από δεδομένα κειμένου. Ο στόχος της ανάλυσης συναισθημάτων είναι να προσδιοριστεί το συναίσθημα που εκφράζεται σε ένα κομμάτι κειμένου, όπως θετικό, αρνητικό ή ουδέτερο.

### 4.5.2 Εφαρμογή Ανάλυσης Συναισθημάτων

Η ανάλυση συναισθημάτων βρίσκει εφαρμογή σε πληθώρα τομέων όπως είναι η παρακολούθηση των μέσων κοινωνικής δικτύωσης, η ανάλυση ανατροφοδότησης πελατών, η έρευνα αγοράς, η πολιτική ανάλυση, η υγειονομική περίθαλψη και τα χρηματοοικονομικά. Από τους τομείς που αναφέρθηκαν παραπάνω γίνεται εύκολα κατανοητή η σπουδαιότητα της ανάλυσης συναισθημάτων καθώς εμπλέκεται σε τομείς όπου η αποδοτική χρήση της μπορεί να αποφέρει τεράστια οικονομικά και κοινωνικά οφέλη. Ίσως το πιο χαρακτηριστικό παράδειγμα αποδοτικής χρήσης της ανάλυσης συναισθημάτων είναι η χρήση της από την εταιρία Cambridge Analytica για να επηρεάσει το αποτέλεσμα των εκλογών που πραγματοποιήθηκαν στις Ηνωμένες Πολιτείες το 2016. Σε αυτή την περίπτωση η εταιρία συνέλεξε παράνομα δεδομένα χρηστών από την πλατφόρμα κοινωνικής δικτύωσης «Facebook» και χρησιμοποιώντας τεχνικές ανάλυσης συναισθημάτων δημιούργησε ανάλογα με το προφίλ του χρήστη εξειδικευμένες πολιτικές διαφημίσεις και μηνύματα με σκοπό να επηρεάσει την ψήφο τους.

### 4.5.3 Διαδικασία Ανάλυσης Συναισθημάτων

Η διαδικασία της ανάλυσης συναισθημάτων περιλαμβάνει έξι βήματα που έχουν ως αφηρημένο σκοπό την ανάλυση των λέξεων, των φράσεων και άλλων γλωσσικών χαρακτηριστικών των δεδομένων ενός κειμένου για τον προσδιορισμό του εκφραζόμενου συναισθήματος. Στο πρώτο βήμα, συλλέγονται τα δεδομένα που θα χρησιμοποιηθούν για την ανάλυση, τέτοια δεδομένα μπορεί να είναι για παράδειγμα τα άρθρα ειδήσεων ή κριτικές στα

μέσα κοινωνικής δικτύωσης. Στο δεύτερο βήμα, τα δεδομένα που συλλέχτηκαν προεπεξεργάζονται έτσι ώστε να αφαιρεθεί τυχόν θόρυβος και να έρθουν σε κατάλληλη για το σύστημα μορφή. Στο τρίτο βήμα, πραγματοποιείται η εξαγωγή των χαρακτηριστικών από τα δεδομένα εισόδου έτσι ώστε να χρησιμοποιηθούν ως είσοδο για το μοντέλο ανάλυσης συναισθημάτων. Τέτοια χαρακτηρίστηκα μπορεί να είναι για παράδειγμα λέξεις, φράσεις ή και «n-grams». Στο τέταρτο βήμα, πραγματοποιείται η ανάλυση συναισθημάτων στα δεδομένα κειμένου με τη χρήση τεχνικών όπως είναι το σύστημα βασισμένο σε κανόνες (rule based system), οι αλγόριθμοι μηχανικής μάθησης και τα μοντέλα βαθιάς μάθησης. Στο πέμπτο βήμα πραγματοποιείται η αξιολόγηση του μοντέλου ανάλυσης συναισθημάτων εφαρμόζοντας το σε δοκιμαστικά δεδομένα. Τέλος, στο έκτο βήμα εφόσον το μοντέλο έχει αξιολογηθεί θετικά, το μοντέλο ανάλυσης συναισθημάτων αναπτύσσεται για χρήση σε πραγματικές εφαρμογές.

### 4.5.4 Προσεγγίσεις Ανάλυσης Συναισθημάτων

Υπάρχουν διάφορες προσεγγίσεις για την ανάλυση συναισθήματος, οι βασικότερες εκ των οποίων είναι η προσέγγιση βάσει κανόνων (rule-based approach), η προσέγγιση με βάση το λεξικό (lexicon-based approach), η προσέγγιση μηχανικής μάθησης (machine learning approach) και η υβριδική προσέγγιση (hybrid approach) η οποία αποτελεί συνδυασμό των δύο προηγούμενων.

#### 4.5.4.1 Προσέγγιση με βάση το λεξικό

Η ανάλυση συναισθημάτων με βάση το λεξικό είναι μια δημοφιλής προσέγγιση που χρησιμοποιεί προκαθορισμένα λεξικά λέξεων, που περιέχουν για κάθε λέξη μία βαθμολογία συναισθήματος, για την ανάλυση του συναισθήματος ενός κειμένου. Πιο αναλυτικά, η προσέγγιση αυτή αποδίδει σε κάθε λέξη ενός κειμένου μία βαθμολογία βάσει του λεξικού που χρησιμοποιεί και στο τέλος συγκεντρώνει αυτές τις βαθμολογίες για να προκύψει μια συνολική βαθμολογία συναισθήματος για το κείμενο. Οι βαθμολογίες μπορεί να είναι είτε δυαδικές (θετικές ή αρνητικές) είτε συνεχείς (που κυμαίνονται από πολύ αρνητικές έως πολύ θετικές).

Η ανάλυση συναισθήματος βάσει λεξικού αν και έχει το πλεονέκτημα ότι μπορεί να εφαρμοστεί εύκολα και γρήγορα σε κάθε τύπο κειμένου απαιτώντας μόνο ένα καλό λεξικό και τίποτα παραπάνω, έχει το μεγάλο μειονέκτημα ότι στην ουσία αντιλαμβάνεται ως προσέγγιση μόνο το νόημα ανά λέξη και όχι των φράσεων ή ακόμη και του κειμένου, με αποτέλεσμα να υστερεί σημαντικά σε περιπτώσεις κειμένου με περίπλοκο νόημα όπως είναι ο σαρκασμός. Αυτή η σημαντική αδυναμία οδήγησε την επιστημονική κοινότητα στην εύρεση πιο αποτελεσματικών προσεγγίσεων όπως είναι η μηχανική μάθηση.

#### 4.5.4.2 Προσέγγιση Μηχανικής Μάθησης

Η ανάλυση συναισθήματος με βάση την μηχανική μάθηση αποτελεί την πιο σύγχρονη προσέγγιση στην επεξεργασία φυσικής γλώσσας καθώς στις περισσότερες περιπτώσεις είναι πολύ πιο αποτελεσματική από τις υπόλοιπες προσεγγίσεις. Στην προσέγγιση αυτή χρησιμοποιούνται αλγόριθμοι μηχανικής μάθησης για την αυτόματη εκμάθηση μοτίβων και χαρακτηριστικών από μεγάλες ποσότητες δεδομένων. Στην συνέχεια, αυτά τα μοτίβα χρησιμοποιούνται είτε για την ταξινόμηση του κειμένου σε διάφορες κατηγορίες (πχ θετικές, αρνητικές και ουδέτερες) είτε για πιο περίπλοκες περιπτώσεις όπως είναι για παράδειγμα τα προβλήματα παλινδρόμησης.

Η ανάλυση συναισθήματος με βάση τη μηχανική μάθηση έχει το πλεονέκτημα έναντι των υπολοίπων μεθόδων ότι μπορεί να «αντιλαμβάνεται» το νόημα του κειμένου στα πλαίσια ενός προβλήματος και όχι απλά να λειτουργεί σαν ένα «χαζό» λογισμικό που απλά βαθμολογεί το κείμενο βάσει ενός καθορισμένου λεξικού. Αυτό το πλεονέκτημα εμφανίστηκε πιο έντονα με την χρήση μεταμορφωτών (transformers), οι οποίοι έχουν την ικανότητα να χειρίζονται καλύτερα τις εξαρτήσεις μεγάλης απόστασης, δηλαδή τις εξαρτήσεις μεταξύ απομακρυσμένων λέξεων σε μια πρόταση. Αυτό έχει ως αποτέλεσμα να κατανοούν αποτελεσματικά την συνολική σημασία μιας πρότασης ή μιας παραγράφου. Ωστόσο, για να μπορέσει να λειτουργήσει αποτελεσματικά αυτή η προσέγγιση, χρειάζεται αφενός μεγάλος όγκος επισημασμένων δεδομένων για την εκπαίδευση, τα οποία σε πολλές περιπτώσεις είναι πολύ δύσκολο να αποκτηθούν, κι αφετέρου μεγάλη υπολογιστική ισχύ και άρα μεγάλο κόστος για την δημιουργία των μοντέλων. Επιπλέον, απαιτούνται ποιοτικά δεδομένα, καθώς με κακής ποιότητας δεδομένα, είναι πολύ εύκολο τα μοντέλα που δημιουργούνται με βάση αυτήν την προσέγγιση να μεροληπτούν και να οδηγούν σε λανθασμένα μοντέλα.

# ΚΕΦΑΛΑΙΟ 5: ΠΡΟΒΛΕΨΗ ΤΙΜΗΣ ΚΡΥΠΤΟΝΟΜΙΣΜΑΤΩΝ

## 5.1 Εισαγωγή

Σε αυτό το κεφάλαιο πραγματοποιείται πρόβλεψη της τιμής του κρυπτονομίσματος «Bitcoin» με τεχνικές μηχανικής μάθησης. Πιο αναλυτικά, αρχικά πραγματοποιείται η προεπεξεργασία των δεδομένων και πραγματοποιείται η δοκιμή «Breusch-Pagan». Στη συνέχεια, αφού με την δοκιμή «Breusch-Pagan» αποδεικνύεται ότι τα δεδομένα δεν είναι δυνατόν να ερμηνευτούν από το μοντέλο γραμμικής παλινδρόμησης των συνήθων ελαχίστων τετραγώνων, επιλέγονται για την πρόβλεψη της τιμής το μοντέλο πρόβλεψη τιμής με την χρήση Δικτύων Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM), το μοντέλο πρόβλεψης τιμής με την χρήση Δικτύων Αναδρομικής Πύλης (Gated Recurrent Unit Networks – GRU) και το μοντέλο πρόβλεψης τιμής με την χρήση Υβριδικού Μοντέλου (Hybrid Model) το οποίο χρησιμοποιεί και Ανάλυση Συναισθημάτων (Sentiment Analysis). Στο τέλος, πραγματοποιείται στατιστική σύγκριση μεταξύ του Υβριδικού Μοντέλου και του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) ώστε να φανεί εάν η προσθήκη Ανάλυσης συναισθημάτων (Sentiment Analysis) βελτιώνει ουσιαστικά το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU).

## 5.2 Περιγραφή Δεδομένων

### 5.2.1 Εισαγωγή

Το σύνολο των δεδομένων πάνω στο οποίο πραγματοποιείται η εκπαίδευση και η δοκιμή των μοντέλων για την πρόβλεψη της τιμής του κρυπτονομίσματος «Bitcoin» χωρίζεται σε 2 διακριτά υποσύνολα, στα αριθμητικά δεδομένα που αφορούν τα διάφορα χαρακτηριστικά ανά ημέρα του κρυπτονομίσματος «Bitcoin» και στα δεδομένα κειμένου που χρησιμοποιούνται για την ανάλυση συναισθημάτων του εν λόγω κρυπτονομίσματος.

### 5.2.2 Αριθμητικά δεδομένα

Τα ιστορικά αριθμητικά δεδομένα αντλήθηκαν με την χρήση της βιβλιοθήκης «cryptocmd» η οποία αντλεί δεδομένα από την ιστοσελίδα «https://coinmarketcap.com». Τα δεδομένα που αντλήθηκαν αφορούν την χρονική περίοδο από 01/01/2015 έως και 01/01/2022 και περιλαμβάνουν τις τιμές των παρακάτω χαρακτηριστικών του κρυπτονομίσματος «Bitcoin» ανά ημέρα:

- Date (Ημερομηνία): Η ημερομηνία αναφέρει την ακριβή ημερομηνία στην οποία αφορούν τα υπόλοιπα χαρακτηριστικά του κρυπτονομίσματος «Bitcoin» που αντλήθηκαν.
- Open (Τιμή Ανοίγματος): Η τιμή ανοίγματος είναι η τιμή εκκίνησης που διαπραγματεύεται το κρυπτονόμισμα «Bitcoin».
- High (Υψηλότερη Τιμή): Η υψηλότερη τιμή είναι η μέγιστη ημερήσια τιμή διαπραγμάτευσης του κρυπτονομίσματος «Bitcoin».
- Low (Χαμηλότερη Τιμή): Η χαμηλότερη τιμή είναι η ελάχιστη ημερήσια τιμή διαπραγμάτευσης του κρυπτονομίσματος «Bitcoin».

- Close (Τιμή Κλεισίματος): Η τιμή κλεισίματος είναι η τιμή που διαπραγματεύεται το κρυπτονόμισμα «Bitcoin» στο τέλος της ημέρας.
- Volume (Όγκος Συναλλαγών): Ο όγκος συναλλαγών είναι το πλήθος των συναλλαγών που πραγματοποιήθηκαν κατά τη διάρκεια μιας ημέρας για το κρυπτονόμισμα «Bitcoin».
- Market Cap (Κεφαλαιοποίηση Αγοράς): Η κεφαλαιοποίηση αγοράς είναι η συνολική αξία των συναλλαγών που έλαβαν χώρα κατά τη διάρκεια μιας ημέρας για το κρυπτονόμισμα «Bitcoin».

### 5.2.3 Δεδομένα Κειμένου

Τα δεδομένα κειμένου αποτελούνται από άρθρα ειδήσεων σχετικών με το κρυπτονόμισμα «Bitcoin» τα οποία δημοσιεύτηκαν την χρονική περίοδο από 01/01/2015 έως και 01/01/2022. Για κάθε ημέρα έγινε προσπάθεια άντλησης 50 άρθρων ανά ημέρα ωστόσο αυτό δεν κατέστη δυνατό για όλες τις ημέρες, με αποτέλεσμα το πλήθος των άρθρων που αντλήθηκαν ανά ημέρα να διαφέρει από ημέρα σε ημέρα. Για την άντληση των άρθρων χρησιμοποιήθηκε η διασύνδεση προγραμματισμού εφαρμογών (API) «News API» της εταιρείας «MatcherLabs», η οποία είναι διαθέσιμη επί πληρωμή στην ιστοσελίδα «https://rapidapi.com». Για κάθε άρθρο αντλήθηκαν τα παρακάτω χαρακτηριστικά:

- Title (Τίτλος): Η τιμή του τίτλου περιλαμβάνει τον τίτλο του εν λόγω άρθρου.
- Url (Ενιαίος Εντοπιστής Πόρων): Η τιμή του ενιαίου εντοπιστή πόρων περιέχει την διεύθυνση του άρθρου στον Παγκόσμιο Ιστό.
- Published Date (Ημερομηνία Δημοσίευσης): Η ημερομηνία δημοσίευσης αναφέρεται στην ημερομηνία που δημοσιεύτηκε το εκάστοτε άρθρο.
- Keywords (Λέξεις Κλειδιά): Το χαρακτηριστικό «Keywords» περιέχει τις λέξεις κλειδιά του κάθε άρθρου.
- Author (Συγγραφέας): Το χαρακτηριστικό «Author» αναφέρεται στο όνομα ή στα ονόματα των συγγραφέων του κάθε άρθρου.
- Description (Περιγραφή): Το χαρακτηριστικό «Description» περιέχει μία σύντομη περιγραφή του άρθρου.
- Thumbnail (Μικρογραφία): Το χαρακτηριστικό «Thumbnail» περιέχει την διεύθυνση της μικρογραφίας του άρθρου στο διαδίκτυο.
- Publisher Name (Όνομα Εκδότη): Το χαρακτηριστικό «Publisher Name» περιέχει το όνομα του εκδότη του κάθε άρθρου.
- Publisher Url (Ενιαίος Εντοπιστής Πόρων Εκδότη): Το χαρακτηριστικό «Publisher Url» περιέχει την κεντρική διεύθυνση του άρθρου στον Παγκόσμιο Ιστό.

## 5.3 Προεπεξεργασία Δεδομένων

### 5.3.1 Εισαγωγή

Προκειμένου τα αριθμητικά δεδομένα και τα δεδομένα κειμένου που αντλήθηκαν να καταστούν κατάλληλα για χρήση από τις πραγματοποιθείσες υλοποιήσεις, έχρηζε επιτακτικής ανάγκης η προεπεξεργασία τους.

### 5.3.2 Προεπεξεργασία Αριθμητικών Δεδομένων

Κατά την προεπεξεργασία των αριθμητικών δεδομένων, αφαιρέθηκε από τα δεδομένα το χαρακτηριστικό «Date» και δημιουργήθηκαν δύο σύνολα δεδομένων. Το πρώτο σύνολο

δεδομένων περιέχει όλα τα εναπομείναντα χαρακτηριστικά και αποτελεί το σύνολο εισόδου ενώ το δεύτερο σύνολο περιέχει την τιμή του χαρακτηριστικού «Close» της επόμενης ημέρας και αποτελεί το σύνολο εξόδου. Στη συνέχεια και τα δύο σύνολα κανονικοποιήθηκαν ανά χαρακτηριστικό με εύρος τιμών από -1 έως και 1. Μετά την κανονικοποίηση των δεδομένων πραγματοποιήθηκε η παραθυροποίηση των δεδομένων, δηλαδή η δημιουργία παραθύρων προσαρμοζόμενου μεγέθους όπου κάθε παράθυρο περιέχει τα δεδομένα τουλάχιστον δύο ημερών.

### 5.3.3 Προεπεξεργασία Δεδομένων Κειμένου

Η προεπεξεργασία των δεδομένων κειμένου πραγματοποιήθηκε σε τέσσερα βασικά βήματα προκειμένου να μετατραπούν σε αριθμητικά δεδομένα, έτσι ώστε να μπορούν να χρησιμοποιηθούν ως δεδομένα εισόδου στις υλοποιήσεις. Το πρώτο βήμα είναι η άντληση ολόκληρου του κειμένου κάθε άρθρου με την χρήση της βιβλιοθήκης σε γλώσσα «Python» «newspaper 0.1.0.7». Πιο αναλυτικά, με την χρήση του χαρακτηριστικού «Url» και της κλάσης «Article» αντλείται το κείμενο του κάθε άρθρου και αποθηκεύεται ως χαρακτηριστικό στα δεδομένα κειμένου με την ονομασία «article_text». Το δεύτερο βήμα είναι η πραγματοποίηση περίληψης του κειμένου που περιέχεται στο χαρακτηριστικό «article_text» με την χρήση του μεταμορφωτή (transformer) «XLNet». Κατά το τρίτο βήμα γίνεται η μετατροπή του κειμένου της περίληψης σε ένα διάνυσμα 768 μεταβλητών με τιμές από -1 έως και 1. Για την μετατροπή αυτή χρησιμοποιείται ο μεταμορφωτής (transformer) «bigBird». Τέλος, στο τέταρτο βήμα σχηματίζεται ένας πίνακας διανυσμάτων ο οποίος αποτελείται από τα διανύσματα των άρθρων κάθε ημέρας. Αναλυτικότερα όπως φαίνεται και στην εικόνα 5.1, κάθε γραμμή του πίνακα διανυσμάτων περιέχει το διάνυσμα ενός άρθρου. Όλα τα άρθρα του πίνακα διανυσμάτων αφορούν την ίδια ημερομηνία και το πλήθος τους καθορίζεται από τον χρήστη ανά περίπτωση υλοποίησης. Αυτός ο πίνακας των διανυσμάτων αποτελεί και το σύνολο δεδομένων εισόδου για την ανάλυση συναισθημάτων.

| | Τιμή 1 | Τιμή 2 | Τιμή 3 | Τιμή 4 | … | Τιμή 768 |
|---|---|---|---|---|---|---|
| **Άρθρο 1** | 0,4 | 0,2 | 0,8 | -0,3 | … | -0,2 |
| **Άρθρο2** | 0,1 | -0,3 | 0,4 | -0,1 | … | 0,7 |
| **Άρθρο 3** | 0,9 | -0,3 | 0,6 | -0,4 | … | 0,4 |

**Εικόνα 5.1: Πίνακας διανυσμάτων με άρθρα ημέρας**

## 5.4 Πραγματοποίηση δοκιμής «Breusch-Pagan».

### 5.4.1 Εισαγωγή

Η δοκιμή «Breusch-Pagan» είναι μία στατιστική δοκιμή η οποία προερχόμενη από την αρχή της δοκιμής του πολλαπλασιαστή «Lagrange» ελέγχει αν η διακύμανση των σφαλμάτων μιας παλινδρόμησης (regression) εξαρτάται από τις τιμές των ανεξάρτητων μεταβλητών, δηλαδή χρησιμοποιείται για την ανίχνευση της ετεροσκεδαστικότητας (heteroscedasticity) στην ανάλυση των μοντέλων παλινδρόμησης (regression models). Η δοκιμή πραγματοποιείται με την σύγκριση των αθροισμάτων των τετραγωνικών καταλοίπων ενός μοντέλου παλινδρόμησης με μηδενική υπόθεση ομοσκεδαστικότητας (σταθερή διακύμανση) και ενός μοντέλου που επιτρέπει την ετεροσκεδαστικότητα (μεταβαλλόμενη διακύμανση). Εάν η διαφορά μεταξύ αυτών των δύο αθροισμάτων είναι στατιστικά σημαντική, τότε μπορούμε να απορρίψουμε τη

μηδενική υπόθεση και να συμπεράνουμε ότι τα σφάλματα στο υπόδειγμά μας δεν είναι ομοσκεδαστικά. Στην περίπτωση ετεροσκεδαστικότητας, τότε μπορεί να είναι απαραίτητο να τροποποιηθεί το μοντέλο ή να χρησιμοποιηθεί διαφορετικός τύπος μοντέλου.

### 5.4.2 Δεδομένα Υλοποίησης

Για την πραγματοποίηση της δοκιμής «Breusch-Pagan» χρησιμοποιήθηκαν τα προεπεξεργασμένα αριθμητικά δεδομένα όπως περιγράφονται στην υποενότητα 5.3.2 με τιμές παραθύρων 1, 7, 14, 30 και 60.

### 5.4.3 Υλοποίηση Δοκιμής

Για την υλοποίηση της δοκιμής «Breusch-Pagan», αρχικά εφαρμόστηκε το μοντέλο γραμμικής παλινδρόμησης (lineal regression) των συνήθων ελαχίστων τετραγώνων (Ordinary Least Squares – OLS) πάνω στα προεπεξεργασμένα αριθμητικά δεδομένα. Στην συνέχεια, έγινε η σύγκριση τετραγωνικών καταλοίπων και υπολογίστηκαν τα αποτελέσματα. Τα αποτελέσματα περιέχουν τον υπολογισμό των μεταβλητών «LM Statistic», «LM p-value», «F Statistic» και «F p-value». Η παραπάνω διαδικασία πραγματοποιήθηκε ξεχωριστά για κάθε τιμή παραθύρου.

### 5.4.4 Αποτελέσματα

Από τα αποτελέσματα που φαίνονται στην Εικόνα 5.2 προκύπτει ότι ανεξάρτητα από το μέγεθος του παραθύρου οι τιμές των μεταβλητών «LM p-value» και «F p-value» παραμένουν πολύ πιο χαμηλές από το κατώφλι (threshold) του 0,05. Αυτό υποδεικνύει την παρουσία ετεροσκεδαστικότητας στα κατάλοιπα, δηλαδή την αδυναμία ερμηνείας των προεπεξεργασμένων αριθμητικών δεδομένων από το μοντέλο γραμμικής παλινδρόμησης των συνήθων ελαχίστων τετραγώνων. Έτσι, γίνεται ξεκάθαρη η ανάγκη εύρεσης πιο αποτελεσματικών μοντέλων από τα γραμμικά για την αποτελεσματική ερμηνεία αυτών των δεδομένων. Τέτοια μοντέλα μπορεί να αποτελούν τα μοντέλα μηχανικής μάθησης.

| Ημέρες Παραθύρου | LM Statistic | LM p-value | F Statistic | F p-value |
|---|---|---|---|---|
| 1 | 523,15 | 8,67E-110 | 109,32 | 6,12E-123 |
| 7 | 693,28 | 7,84E-119 | 22,29 | 4,28E-141 |
| 14 | 876,34 | 3,40E-132 | 15,39 | 9,71E-168 |
| 30 | 1.107,90 | 2,81E-117 | 10,18 | 4,21E-188 |
| 60 | 1.329,61 | 4,26E-96 | 6,76 | 5,01E-181 |

**Εικόνα 5.2: Αποτελέσματα δοκιμής «Breusch-Pagan»**

## 5.5 Πρόβλεψη τιμής με την χρήση Δικτύων Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM)

### 5.5.1 Εισαγωγή

Τα δίκτυα μακράς βραχύχρονης μνήμης (LSTM) είναι ένας τύπος επαναλαμβανόμενων τεχνητών νευρωνικών δικτύων που έχουν σχεδιαστεί για να αντιμετωπίζουν το πρόβλημα των εξαφανιζόμενων κλίσεων (Vanishing Gradient Problem) που μπορεί να εμφανιστεί στα

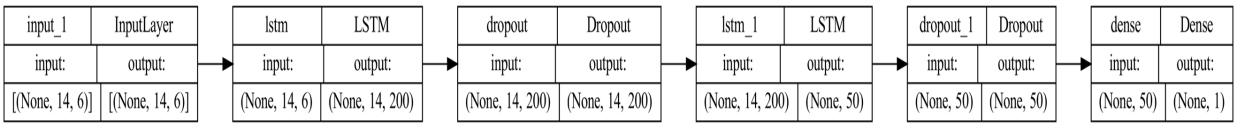
παραδοσιακά επαναλαμβανόμενα νευρωνικά δίκτυα (RNN). Τα δίκτυα μακράς βραχυπρόθεσμης μνήμης έχουν σχεδιαστεί για να επεξεργάζονται διαδοχικά δεδομένα, όπως χρονοσειρές ή κείμενα φυσικής γλώσσας έχοντας το πλεονέκτημα, έναντι των υπολοίπων επαναλαμβανόμενων τεχνητών νευρωνικών δικτύων, ότι είναι σε θέση να αποθηκεύουν και να χειρίζονται πληροφορίες για μεγάλες χρονικές περιόδους, καθιστώντας τα ιδιαίτερα κατάλληλα για εργασίες που περιλαμβάνουν μακροπρόθεσμες εξαρτήσεις. Για να επιτευχθούν αυτές οι μακροπρόθεσμες εξαρτήσεις χρησιμοποιείται μια σειρά από κύτταρα μνήμης (memory cells), καθένα από τα οποία έχει τις δικές του πύλες εισόδου (input gate), εξόδου (output gate) και λήθης (forget gate), οι οποίες ελέγχουν την ροή των πληροφοριών μέσα και έξω από το κύτταρο. Αυτές οι πύλες υλοποιούνται χρησιμοποιώντας τις σιγμοειδείς (sigmoid) και εφαπτόμενες (tanh) συναρτήσεις ενεργοποίησης, οι οποίες επιτρέπουν στο δίκτυο να θυμάται ή να ξεχνά επιλεκτικά πληροφορίες σε κάθε χρονικό βήμα.

### 5.5.2 Δεδομένα Υλοποίησης

Για την εκπαίδευση και δοκιμή του μοντέλου Δικτύων Μακράς Βραχύχρονης Μνήμης χρησιμοποιήθηκαν τα προεπεξεργασμένα αριθμητικά δεδομένα όπως περιγράφονται στην υποενότητα 5.3.2 με τιμές παραθύρων 1, 7, 14, 30 και 60. Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε το πρώτο 80% των δεδομένων ενώ για την δοκιμή του Δικτύου το υπόλοιπο 20%.

### 5.5.3 Δομή Δικτύου Μακράς Βραχύχρονης Μνήμης

Το μοντέλο του Δικτύου Μακράς Βραχύχρονης Μνήμης που δημιουργήθηκε αποτελείται από 6 στρώματα όπως φαίνονται παρακάτω στην Εικόνα 5.3. Το πρώτο στρώμα αποτελεί το στρώμα εισόδου (Input Layer) το οποίο λαμβάνει δεδομένα και τα διαβιβάζει στο επόμενο στρώμα για επεξεργασία. Το δεύτερο και τέταρτο στρώμα αποτελούν στρώματα μακράς βραχυπρόθεσμης μνήμης (LSTM) με 200 και 50 αριθμούς μονάδων αντίστοιχα. Το τρίτο και πέμπτο στρώμα αποτελούν στρώματα διακοπής (Dropout Layer) με ποσοστά 20% και 10% αντίστοιχα. Το τελευταίο -έκτο στρώμα- αποτελεί πυκνό στρώμα (Dense Layer) με μία μονάδα. Η διαμόρφωση της διαδικασίας εκμάθησης του μοντέλου πραγματοποιήθηκε με τον ορισμό του βελτιστοποιητή «adam», της συνάρτησης απώλειας του Μέσου Τετραγωνικού Σφάλματος (Mean squared error - MSE) και των μετρικών αξιολόγησης του Μέσου Απόλυτου Σφάλματος (Mean Absolute Error - MAE) και της Ακρίβειας (Accuracy - ACC). Τέλος, για την προσαρμογή του μοντέλου στα δεδομένα χρησιμοποιούνται 30 εποχές (epochs), 128 μέγεθος δέσμης (batch size) και ποσοστό διαχωρισμού επικύρωσης (validation split) 30%.




**Εικόνα 5.3: Η δομή του Δικτύου Μακράς Βραχύχρονης Μνήμης με τιμή παραθύρου 14.**

### 5.5.4 Αποτελέσματα δοκιμών

Οι δοκιμές πραγματοποιήθηκαν για τιμές παραθύρων 1, 7, 14, 30 και 60. Για κάθε τιμή παραθύρου έλαβαν χώρα 30 δοκιμές και πάρθηκε ο μέσος όρος κάθε μέτρου

αποτελεσματικότητας. Τα μέτρα αποτελεσματικότητας που χρησιμοποιήθηκαν είναι το Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error - MSE), η Μέση Ρίζα Τετραγωνικού Σφάλματος (Root Mean Squared Error – RMSE) , το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error - MAPE) και το Μέσο Απόλυτο Σφάλμα (Mean Absolute Error - MAE). Από τα αποτελέσματα των δοκιμών που φαίνονται αναλυτικά στην Εικόνα 5.4, προκύπτει ότι το μοντέλο είναι πιο αποτελεσματικό για τιμή παραθύρου 7, ενώ καθόλου αποτελεσματικό για τιμή παραθύρου μίας ημέρας.

| Error | Window | Mean | Median | Min | Max | Stdev |
|---|---|---|---|---|---|---|
| RMSE | 1 | 11.627,01 | 11.683,56 | 10.798,78 | 12.170,55 | 353,94 |
| | 7 | 3.754,67 | 3.707,69 | 3.379,35 | 4.858,19 | 324,51 |
| | 14 | 4.020,36 | 3.953,48 | 3.662,18 | 5.019,23 | 365,47 |
| | 30 | 4.039,46 | 3.959,80 | 3.684,85 | 4.686,75 | 295,24 |
| | 60 | 4.239,08 | 4.145,14 | 3.789,33 | 5.358,93 | 373,25 |
| MAPE | 1 | 20,06% | 20,15% | 18,49% | 21,25% | 0,72% |
| | 7 | 6,79% | 6,75% | 6,34% | 7,88% | 0,32% |
| | 14 | 7,52% | 7,40% | 6,90% | 8,52% | 0,45% |
| | 30 | 7,62% | 7,64% | 6,78% | 9,10% | 0,62% |
| | 60 | 7,87% | 7,83% | 7,00% | 9,56% | 0,61% |

**Εικόνα 5.4: Αποτελέσματα δοκιμών μοντέλου LSTM.**

## 5.6 Πρόβλεψη τιμής με την χρήση Δικτύων Αναδρομικής Πύλης (Gated Recurrent Unit Networks – GRU)

### 5.6.1 Εισαγωγή

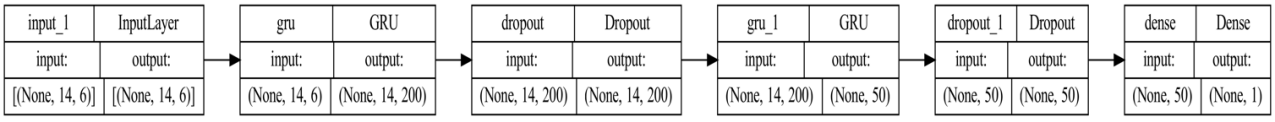
Τα δίκτυα αναδρομικής πύλης (GRU) είναι ένας τύπος αναδρομικών νευρωνικών δικτύων που χρησιμοποιείται με μεγάλη επιτυχία στην πρόβλεψη χρονομετρών. Τα δίκτυα αναδρομικής πύλης αποτελούνται από δύο ειδών πύλες, τις πύλες ενημέρωσης «update gate» και τις πύλες επαναφοράς «reset gate». Οι πύλες ενημέρωσης καθορίζουν την ποσότητα της παλιάς πληροφορίας που θα διατηρηθεί και της νέας πληροφορίας που θα ενημερωθεί, ενώ οι πύλες της επαναφοράς αποφασίζουν την ποσότητα της προηγούμενης κατάστασης που θα αγνοηθεί. Έτσι, τα δίκτυα αναδρομικής πύλης αποκτούν «μνήμη», δηλαδή την δυνατότητα να διατηρούν πληροφορίες από προηγούμενα βήματα. Αυτή η δυνατότητα δίνει το πλεονέκτημα στα δίκτυα αναδρομικής πύλης να μαθαίνουν με μεγαλύτερη ακρίβεια από άλλες αρχιτεκτονικές, όπως τα απλά επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) σε δεδομένα χρονοσειρών.

### 5.6.2 Δεδομένα Υλοποίησης

Για την εκπαίδευση και δοκιμή του μοντέλου Δικτύων Αναδρομικής Πύλης, χρησιμοποιήθηκαν τα προεπεξεργασμένα αριθμητικά δεδομένα όπως περιγράφονται στην υποενότητα 5.3.2 με τιμές παραθύρων 1, 7, 14, 30 και 60. Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε το πρώτο 80% των δεδομένων ενώ για την δοκιμή του Δικτύου το υπόλοιπο 20%.

### 5.6.3 Δομή Μοντέλου Δικτύων Αναδρομικής Πύλης

Το μοντέλο Δικτύων Αναδρομικής Πύλης που δημιουργήθηκε αποτελείται από 6 στρώματα όπως φαίνονται στην Εικόνα 5.4. Το πρώτο στρώμα αποτελεί το Στρώμα Εισόδου (Input Layer) το οποίο λαμβάνει δεδομένα και τα διαβιβάζει στο επόμενο στρώμα για επεξεργασία. Το δεύτερο και τέταρτο στρώμα αποτελούν στρώματα Αναδρομικής Πύλης (GRU) με 200 και 50 αριθμούς μονάδων (units) αντίστοιχα. Το τρίτο και πέμπτο στρώμα αποτελούν Στρώματα Διακοπής (Dropout Layers) με ποσοστά διακοπής 20% και 10% αντίστοιχα. Το τελευταίο - έκτο στρώμα- αποτελεί Πυκνό Στρώμα (Dense Layer) μιας μονάδας. Η διαμόρφωση της διαδικασίας εκμάθησης του μοντέλου πραγματοποιήθηκε με τον ορισμό του βελτιστοποιητή «adam», της συνάρτησης απώλειας του Μέσου Τετραγωνικού Σφάλματος (Mean squared error - MSE) και των μετρικών αξιολόγησης του Μέσου Απόλυτου Σφάλματος (Mean Absolute Error - MAE) και της Ακρίβειας (Accuracy - ACC). Τέλος, για την προσαρμογή του μοντέλου στα δεδομένα χρησιμοποιούνται 60 εποχές (epochs), κανένα μέγεθος δέσμης (batch size) και ποσοστό διαχωρισμού επικύρωσης (validation split) 20%.




**Εικόνα 5.4: Η δομή του Μοντέλου Δικτύων Αναδρομικής Πύλης με τιμή παραθύρου 14.**

### 5.6.4 Αποτελέσματα δοκιμών

Οι δοκιμές πραγματοποιήθηκαν για τιμές παραθύρων 1, 7, 14, 30 και 60. Για κάθε τιμή παραθύρου έλαβαν χώρα 30 δοκιμές και υπολογίστηκε ο μέσος όρος, η διάμεσος, το ελάχιστο, το μέγιστο και η τυπική απόκλιση για τις τιμές κάθε μέτρου αποτελεσματικότητας. Τα μέτρα αποτελεσματικότητας που χρησιμοποιήθηκαν είναι η Μέση Ρίζα Τετραγωνικού Σφάλματος (Root Mean Squared Error – RMSE) και το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error - MAPE). Από τα αποτελέσματα των δοκιμών που φαίνονται αναλυτικά στην Εικόνα 5.5, προκύπτει ότι το μοντέλο είναι πιο αποτελεσματικό για τιμές παραθύρου 14 και 30, ενώ καθόλου αποτελεσματικό για τιμή παραθύρου μίας ημέρας.

| Error | Window | Mean | Median | Min | Max | Stdev |
|---|---|---|---|---|---|---|
| RMSE | 1 | 3.624,60 | 3.680,66 | 2.858,83 | 4.138,00 | 304,34 |
| | 7 | 2.582,84 | 2.507,69 | 1.975,14 | 3.962,29 | 490,03 |
| | 14 | 2.200,69 | 2.089,41 | 1.934,69 | 2.872,36 | 269,17 |
| | 30 | 2.225,90 | 2.122,55 | 1.998,00 | 2.982,23 | 247,05 |
| | 60 | 2.325,97 | 2.248,85 | 1.997,25 | 4.177,25 | 405,20 |
| MAPE | 1 | 5,83% | 5,86% | 4,48% | 7,04% | 0,53% |
| | 7 | 5,18% | 5,02% | 3,75% | 7,50% | 1,00% |
| | 14 | 4,31% | 4,08% | 3,59% | 6,17% | 0,62% |
| | 30 | 4,42% | 4,33% | 3,70% | 5,50% | 0,49% |
| | 60 | 4,37% | 4,32% | 3,65% | 7,19% | 0,65% |

**Εικόνα 5.5: Αποτελέσματα δοκιμών μοντέλου GRU.**

## 5.7 Πρόβλεψη τιμής με την χρήση Υβριδικού Μοντέλου (Hybrid Model) που χρησιμοποιεί και Ανάλυση Συναισθημάτων (Sentiment Analysis)

### 5.7.1 Εισαγωγή

Τα Υβριδικά Μοντέλα (Hybrid Models) αποτελούν συνδυασμό πολλαπλών μοντέλων ή τεχνικών για την επίλυση ενός συγκεκριμένου προβλήματος. Συνδυάζουν τα πλεονεκτήματα και τις δυνατότητες διαφορετικών μοντέλων προκειμένου να βελτιώσουν την συνολική αποτελεσματικότητα και ακρίβεια. Τα μοντέλα αυτά χρησιμοποιούνται συχνά όταν ένα μόνο μοντέλο ή τεχνική δεν επαρκεί για να συλλάβει όλες τις πολυπλοκότητες ή τα πρότυπα των δεδομένων. Στην συγκεκριμένη περίπτωση χρησιμοποιείται ο συνδυασμός των Δικτύων Αναδρομικής Πύλης (GRU) με τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks – CNN). Τα Δίκτυα Αναδρομικής Πύλης (GRU) αποτελούν έναν τύπο αναδρομικών νευρωνικών δικτύων όπως αναλύονται στην υποενότητα 5.6.1, ενώ τα Συνελικτικά Νευρωνικά Δίκτυα (CNN) αποτελούν δίκτυα βαθιάς μάθησης που έχουν σχεδιαστεί ειδικά για να επεξεργάζονται δομημένα δεδομένα που μοιάζουν με πλέγμα, όπως εικόνες. Είναι εμπνευσμένα από τον μηχανισμό οπτικής επεξεργασίας του ανθρώπινου εγκεφάλου και είναι ιδιαίτερα αποτελεσματικά στην καταγραφή τοπικών μοτίβων και χωρικών εξαρτήσεων στα δεδομένα εισόδου, καθιστώντας τα κατάλληλα για την ανάλυση οπτικών πληροφοριών. Παράλληλα, αποτελούν ένα από τα πιο αποτελεσματικά εργαλεία για την μείωση διαστάσεων στα δεδομένα εισόδου.

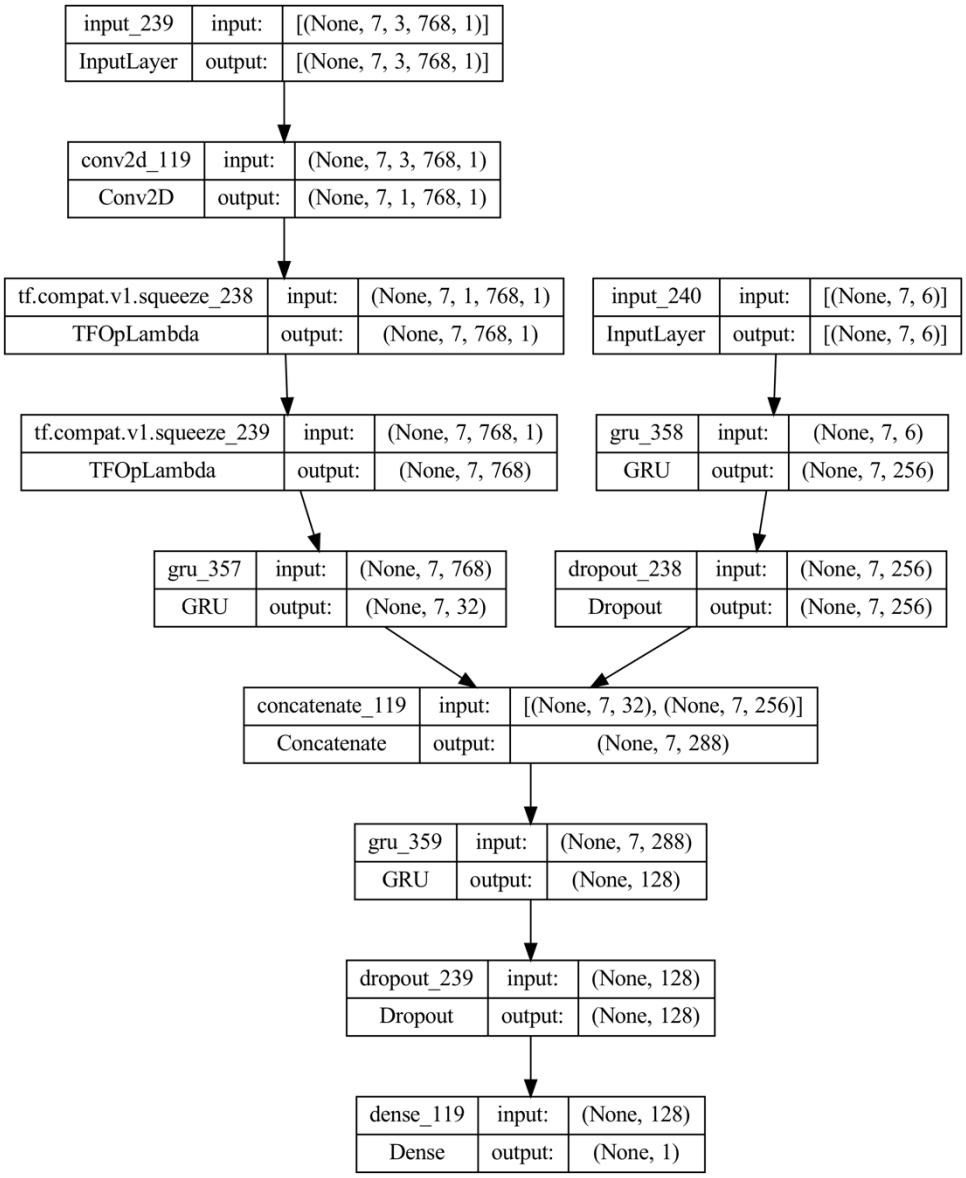
### 5.7.2 Δεδομένα Υλοποίησης

Στην συγκεκριμένη υλοποίηση, για την εκπαίδευση και δοκιμή του Υβριδικού Μοντέλου, χρησιμοποιήθηκαν τα προεπεξεργασμένα αριθμητικά δεδομένα και τα προεπεξεργασμένα δεδομένα κειμένου όπως περιγράφονται στις υποενότητες 5.3.2 και 5.3.3 αντίστοιχα. Από τα δεδομένα κειμένου, επιλέχτηκαν για κάθε ημερολογιακή ημέρα 1, 3, 5, 7 και 10 άρθρα. Η επιλογή των άρθρων έγινε με προτεραιότητα την σειρά άντλησης που αποτελεί και την σειρά σχετικότητας της είδησης με το κρυπτονόμισμα «Bitcoin». Στις περιπτώσεις όπου δεν ήταν δυνατή η άντληση 10 διαφορετικών άρθρων που να είναι σχετικά με το κρυπτονόμισμα «Bitcoin», πραγματοποιήθηκε για τα υπολειπόμενα άρθρα τυχαία επιλογή με επανατοποθέτηση. Οι περιπτώσεις στις οποίες δεν μπόρεσαν να αντληθούν 10 άρθρα ήταν ελάχιστες και αφορούσαν ημερομηνίες στις οποίες πρωτοεμφανίστηκε στην επικαιρότητα το κρυπτονόμισμα «Bitcoin». Οι τιμές παραθύρων που χρησιμοποιήθηκαν είναι 1, 7, 14 και 30. Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε το πρώτο 80% των δεδομένων, ενώ για την δοκιμή του Δικτύου το υπόλοιπο 20%.

### 5.7.3 Δομή Υβριδικού Μοντέλου

Το Υβριδικό Μοντέλο που δημιουργήθηκε όπως φαίνεται και στην Εικόνα 5.6, αποτελείται από ένα υπομοντέλο Συνελικτικών Νευρωνικών Δικτύων (CNN) και ένα υπομοντέλο Δικτύων Αναδρομικής Πύλης (GRU), τα οποία συνενώνονται με την χρήση ενός μοντέλου Δικτύων Αναδρομικής Πύλης (GRU). Το υπομοντέλο των Συνελικτικών Νευρωνικών Δικτύων (CNN) δέχεται ως είσοδο τα προεπεξεργασμένα δεδομένα κειμένου, δηλαδή το διάνυσμα των άρθρων ημέρας όπως περιγράφεται στην υποενότητα 5.3.3, και αποτελείται από 5 στρώματα. Το πρώτο στρώμα αποτελεί το Στρώμα Εισόδου (Input Layer) το οποίο δέχεται ως είσοδο τα δεδομένα

κειμένου με το σχήμα: «None», Μέγεθος Παραθύρου, Πλήθος άρθρων ανά ημέρα, Διάνυσμα άρθρου, 1. Το δεύτερο στρώμα αποτελεί ένα Συνελικτικό Στρώμα Νευρωνικών Δικτύων 2 διαστάσεων (Conv2D Layer) με τιμή φίλτρου (filter) 1, μέγεθος πυρήνα (kernel size) (Πλήθος άρθρων ανά ημέρα, 1), σχήμα βηματισμών (strides) (Πλήθος άρθρων ανά ημέρα, 1) και συνάρτηση ενεργοποίησης (activation function) την υπερβολική εφαπτομένη (hyperbolic tangent – tanh). Το τρίτο και τέταρτο στρώμα, αποτελούν Στρώματα Συμπίεσης (Squeeze Layers) με τιμές αξόνων (axis) 2 και -1 αντίστοιχα. Το πέμπτο και τελευταίο στρώμα του υπομοντέλου Συνελικτικών Νευρωνικών Δικτύων, αποτελεί ένα Στρώμα Δικτύων Αναδρομικής Πύλης (GRU Layer) με 32 μονάδες (units). Το υπομοντέλο Δικτύων Αναδρομικής Πύλης (GRU) αποτελείται από 3 στρώματα. Το πρώτο στρώμα αποτελεί το Στρώμα Εισόδου (Input Layer), το οποίο δέχεται ως είσοδο τα αριθμητικά δεδομένα με το σχήμα: «None», Μέγεθος Παραθύρου, Πλήθος αριθμητικών τιμών(6). Το δεύτερο στρώμα, αποτελεί ένα Στρώμα Δικτύων Αναδρομικής Πύλης (GRU Layer) με 256 μονάδες (units). Το τρίτο και τελευταίο στρώμα του υπομοντέλου Δικτύων Αναδρομικής Πύλης (GRU), αποτελεί ένα Στρώμα Διακοπής (Dropout Layer) με ποσοστό διακοπής 20%. Για την συνένωση των παραπάνω δύο υπομοντέλων χρησιμοποιούνται 4 επιπλέον στρώματα. Αρχικά χρησιμοποιείται ένα Στρώμα Συνένωσης (Concatenate Layer) το οποίο ενώνει τα τελευταία στρώματα των υπομοντέλων. Στη συνέχεια ακολουθεί ένα Στρώμα Δικτύων Αναδρομικής Πύλης (GRU Layer) με 128 μονάδες (units). Μετά ακολουθεί ένα Στρώμα Διακοπής (Dropout Layer) με ποσοστό διακοπής 10%. Τέλος, για την έξοδο του συνολικού μοντέλου, χρησιμοποιείται ένα Πυκνό Στρώμα (Dense Layer) μιας μονάδας.




**Εικόνα 5.6: Η δομή του Υβριδικού μοντέλου με τιμή παραθύρου 7 και χρήση 3 άρθρων ανά ημέρα.**

### 5.7.4 Αποτελέσματα δοκιμών

Οι δοκιμές πραγματοποιήθηκαν για τιμές παραθύρων 1, 7, 14 και 30 και τιμές άρθρων ανά ημέρα 1, 3, 5, 7 και 10. Για κάθε τιμή παραθύρου έλαβαν χώρα 30 δοκιμές και υπολογίστηκε ο μέσος όρος, η διάμεσος, το ελάχιστο, το μέγιστο και η τυπική απόκλιση για τις τιμές κάθε μέτρου αποτελεσματικότητας. Τα μέτρα αποτελεσματικότητας που χρησιμοποιήθηκαν είναι η Μέση Ρίζα Τετραγωνικού Σφάλματος (Root Mean Squared Error – RMSE) και το Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error - MAPE). Από τα αποτελέσματα των δοκιμών που φαίνονται αναλυτικά στην Εικόνα 5.7, προκύπτει ότι το μοντέλο είναι πιο αποτελεσματικό για τιμή άρθρων ανά ημέρα 7 και τιμή παραθύρου 7 ως προς το μέτρο αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος. Παράλληλα, είναι πιο αποτελεσματικό για τιμή άρθρων ανά ημέρα 1 και τιμή παραθύρου 7 ως προς το μέτρο αποτελεσματικότητας Απόλυτου Ποσοστιαίου Σφάλματος και τέλος, το μοντέλο έχει την

καλύτερη αποτελεσματικότητα για τιμή άρθρων ανά ημέρα 3 και τιμή παραθύρου 7 συνδυαστικά και για τα δύο μέτρα αποτελεσματικότητας.

| Articles Per Day | Window | RMSE | | | | | MAPE | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Median | Min | Max | Stdev | Mean | Median | Min | Max | Stdev |
| 1 | 1 | 2.401,07 | 2.356,37 | 1.849,02 | 3.102,46 | 372,50 | 4,18% | 4,04% | 3,46% | 5,30% | 0,52% |
| | 7 | 2.007,57 | 1.961,18 | 1.861,42 | 2.347,30 | 119,99 | 3,78% | 3,72% | 3,42% | 4,56% | 0,28% |
| | 14 | 2.062,55 | 1.994,48 | 1.866,29 | 2.632,68 | 176,57 | 4,07% | 3,89% | 3,47% | 5,92% | 0,58% |
| | 30 | 2.014,10 | 1.990,86 | 1.878,83 | 2.349,72 | 101,62 | 3,98% | 3,85% | 3,45% | 5,65% | 0,52% |
| 3 | 1 | 2.419,44 | 2.379,86 | 1.855,93 | 3.102,46 | 355,90 | 4,19% | 4,07% | 3,46% | 5,30% | 0,53% |
| | 7 | 1.996,83 | 1.956,02 | 1.866,13 | 2.260,08 | 103,56 | 3,86% | 3,85% | 3,50% | 4,76% | 0,30% |
| | 14 | 2.040,43 | 1.997,12 | 1.913,83 | 2.387,90 | 128,76 | 4,13% | 4,05% | 3,49% | 5,84% | 0,50% |
| | 30 | 2.023,42 | 1.991,31 | 1.875,30 | 2.673,64 | 151,71 | 3,90% | 3,74% | 3,47% | 4,94% | 0,39% |
| 5 | 1 | 2.432,47 | 2.420,97 | 2.090,68 | 2.964,28 | 226,18 | 4,22% | 4,23% | 3,59% | 4,86% | 0,34% |
| | 7 | 2.018,05 | 1.977,03 | 1.883,71 | 2.497,96 | 141,96 | 3,96% | 3,80% | 3,48% | 5,32% | 0,47% |
| | 14 | 2.066,95 | 1.970,09 | 1.888,06 | 2.994,53 | 234,84 | 4,06% | 3,98% | 3,47% | 5,30% | 0,46% |
| | 30 | 1.990,34 | 1.981,60 | 1.897,50 | 2.111,18 | 55,64 | 4,09% | 4,02% | 3,52% | 5,33% | 0,43% |
| 7 | 1 | 2.357,69 | 2.365,38 | 1.729,57 | 2.865,72 | 293,21 | 4,13% | 4,02% | 3,32% | 5,16% | 0,48% |
| | 7 | 1.980,60 | 1.954,37 | 1.878,88 | 2.246,48 | 82,16 | 4,01% | 3,96% | 3,44% | 5,00% | 0,39% |
| | 14 | 2.004,29 | 1.984,79 | 1.918,58 | 2.201,50 | 73,35 | 4,03% | 3,89% | 3,54% | 5,17% | 0,43% |
| | 30 | 2.042,61 | 1.962,47 | 1.860,38 | 2.866,12 | 222,94 | 4,02% | 3,94% | 3,45% | 5,41% | 0,48% |
| 10 | 1 | 2.427,29 | 2.421,12 | 1.832,28 | 3.147,96 | 283,07 | 4,24% | 4,27% | 3,34% | 4,98% | 0,44% |
| | 7 | 2.027,92 | 1.984,79 | 1.863,06 | 2.357,17 | 131,29 | 3,90% | 3,79% | 3,41% | 4,62% | 0,36% |
| | 14 | 2.028,88 | 2.005,53 | 1.900,71 | 2.506,55 | 114,26 | 4,11% | 4,07% | 3,51% | 5,17% | 0,40% |
| | 30 | 2.013,02 | 1.988,72 | 1.894,16 | 2.212,19 | 97,04 | 3,93% | 3,83% | 3,51% | 5,02% | 0,36% |

**Εικόνα 5.7: Αποτελέσματα δοκιμών Υβριδικού μοντέλου.**

Από τα αποτελέσματα, είναι ξεκάθαρο ότι το Υβριδικό μοντέλο είναι πιο αποτελεσματικό συγκριτικά με το μοντέλο Δικτύων Μακράς Βραχύχρονης Μνήμης (Long Short-term Memory – LSTM) και με το μοντέλο Δικτύων Αναδρομικής Πύλης (Gated Recurrent Unit Networks – GRU). Αυτό σημαίνει ότι η προσθήκη Ανάλυσης Συναισθημάτων (Sentiment Analysis) έχει βοηθήσει στην αύξηση της αποτελεσματικότητας, καθώς στην ουσία το Υβριδικό μοντέλο αποτελεί συνδυασμό του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) που διαχειρίζεται τα αριθμητικά δεδομένα, με ένα μοντέλο Συνελικτικών Νευρωνικών Δικτύων (CNN) που διαχειρίζεται τα δεδομένα κειμένου.

## 5.8 Στατιστική σύγκριση του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) με το Υβριδικό Μοντέλο (Hybrid Model) που χρησιμοποιεί και Ανάλυση Συναισθημάτων (Sentiment Analysis)

### 5.8.1 Εισαγωγή

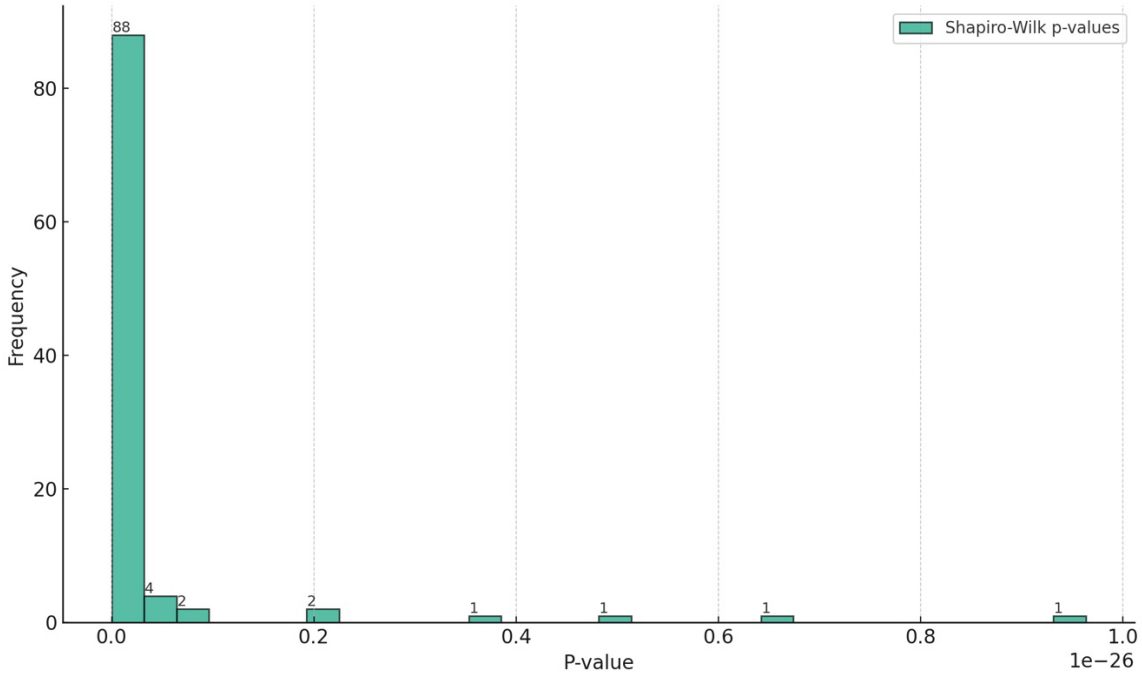
Από τις δοκιμές που προηγήθηκαν προκύπτει ότι τόσο το μοντέλο των Δικτύων Αναδρομικής Πύλης (GRU) όσο και το Υβριδικό Μοντέλο (Hybrid Model) κάνουν αποτελεσματικές προβλέψεις της τιμής του κρυπτονομίσματος «Bitcoin», με το Υβριδικό μοντέλο να φαίνεται εκ πρώτης όψεως ότι είναι πιο αποτελεσματικό. Η ουσιαστική διαφορά των μοντέλων είναι η χρησιμοποίηση Ανάλυσης Συναισθημάτων από το Υβριδικό Μοντέλο, καθώς και τα δύο μοντέλα κατά κύριο λόγο χρησιμοποιούν Δίκτυα Αναδρομικής Πύλης (GRU). Αυτό καθιστά σημαντική την στατιστική σύγκριση των δύο μοντέλων καθώς σε περίπτωση που υπάρχει σημαντική ένδειξη ότι το Υβριδικό Μοντέλο λειτουργεί πιο αποτελεσματικά, υποδεικνύεται ότι η χρήση της Ανάλυσης Συναισθημάτων με αυτόν τον τρόπο, και πιο συγκεκριμένα με την ερμηνεία ειδησεογραφικών άρθρων από μοντέλα μεταμορφωτών (transformers), βοηθάει καθοριστικά στην βελτιστοποίηση των τετριμμένων μεθόδων Μηχανικής Μάθησης.

Για την σύγκριση των δύο μοντέλων πραγματοποιούνται 100 συγκρίσεις μεταξύ του Υβριδικού μοντέλου και του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU). Για κάθε σύγκριση χρησιμοποιούνται τα προεπεξεργασμένα αριθμητικά δεδομένα και τα προεπεξεργασμένα δεδομένα κειμένου όπως περιγράφονται στις υποενότητες 5.3.2 και 5.3.3 αντίστοιχα με ποσοστό 80% για εκπαίδευση και 20% για δοκιμές. Η κάθε δοκιμή αποτελεί και μία σύγκριση ανά ζεύγος (paired comparison) των δύο μοντέλων. Το Υβριδικό μοντέλο ακολουθεί την δομή όπως περιγράφεται στην υποενότητα 5.7.3 με 3 άρθρα ανά ημέρα και μέγεθος παραθύρου 7. Το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU) ακολουθεί την δομή όπως περιγράφεται αναλυτικά στην υποενότητα 5.6.3 με τιμή παραθύρου 14. Αρχικά χρησιμοποιείται η δοκιμή «Shapiro-Wilk» ώστε να γίνει έλεγχος εάν τα δεδομένα της κάθε σύγκρισης ακολουθούν κανονική κατανομή. Στη συνέχεια, αφού αποδειχθεί ότι τα δεδομένα δεν ακολουθούν την κανονική κατανομή, πραγματοποιείται η δοκιμή «Wilcoxon signed-rank» ώστε να αποδειχθεί εάν η διαφορά μεταξύ των δύο μοντέλων στην εν λόγω σύγκριση είναι στατιστικά σημαντική.

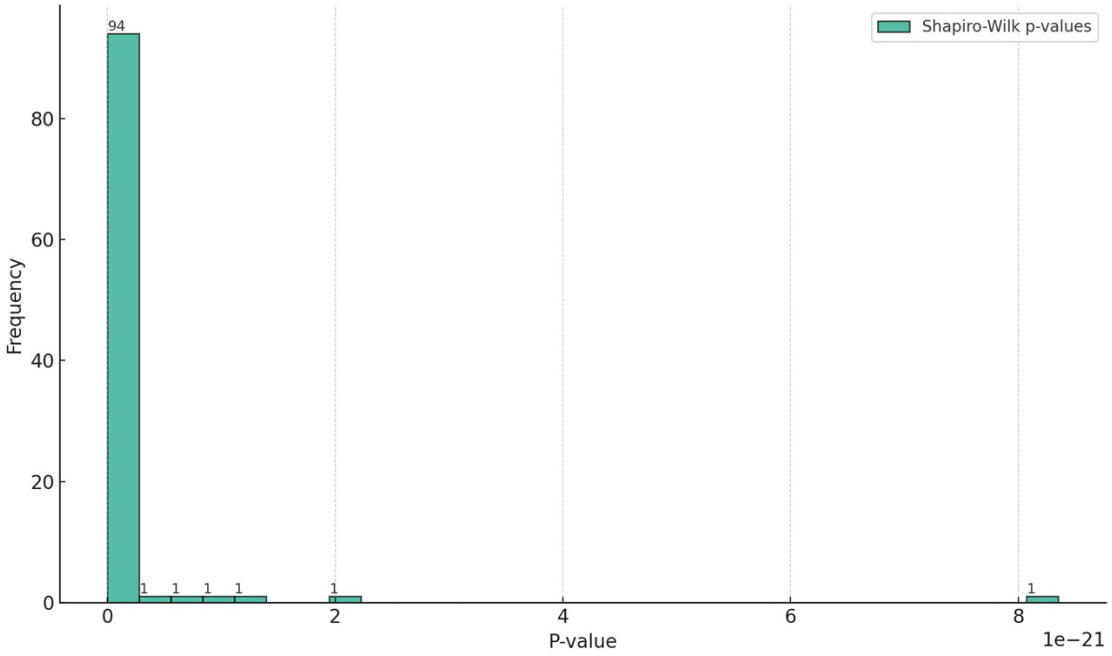
### 5.8.2 Πραγματοποίηση δοκιμής «Shapiro-Wilk».

Η δοκιμή «Shapiro-Wilk» είναι μια ευρέως χρησιμοποιούμενη μέθοδος με σκοπό να ελέγξει αν ένα δείγμα δεδομένων προέρχεται από έναν κανονικά κατανεμημένο πληθυσμό. Η δοκιμή αυτή έχει σχεδιαστεί ειδικά για μικρά μεγέθη δείγματος, συνήθως λιγότερες από 50 παρατηρήσεις, αλλά μπορεί να χρησιμοποιηθεί και για μεγαλύτερα δείγματα. Η μηδενική υπόθεση αυτής της δοκιμής είναι ότι ο πληθυσμός είναι κανονικά κατανεμημένος. Έτσι, εάν η τιμή «p» είναι μικρότερη από το επιλεγμένο επίπεδο άλφα το οποίο σε αυτήν την περίπτωση είναι 0.05, τότε η μηδενική υπόθεση απορρίπτεται και υπάρχουν ενδείξεις ότι τα δεδομένα που εξετάστηκαν δεν κατανέμονται κανονικά. Από την άλλη πλευρά, εάν η τιμή «p» είναι μεγαλύτερη από το επιλεγμένο επίπεδο άλφα (0.05), τότε η μηδενική υπόθεση δεν απορρίπτεται, και κατά συνέπεια τα δεδομένα ακολουθούν κανονική κατανομή.

Στις Εικόνες 5.8 και 5.9 παρουσιάζονται σε Ιστόγραμμα τα αποτελέσματα των 100 δοκιμών «Shapiro-Wilk» για τα μέτρα αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος (RMSE) και Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE) αντίστοιχα. Από τα Ιστογράμματα προκύπτει ότι οι τιμές «p» όλων των δοκιμών ήταν πολύ μικρότερες από 0.05, άρα στο σύνολο των συγκρίσεων τα δεδομένα δεν ακολουθούν κανονική κατανομή για όλα τα μέτρα αποτελεσματικότητας.




**Εικόνα 5.8: Ιστόγραμμα αποτελεσμάτων δοκιμών «Shapiro-Wilk» για το μέτρο αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος (RMSE).**



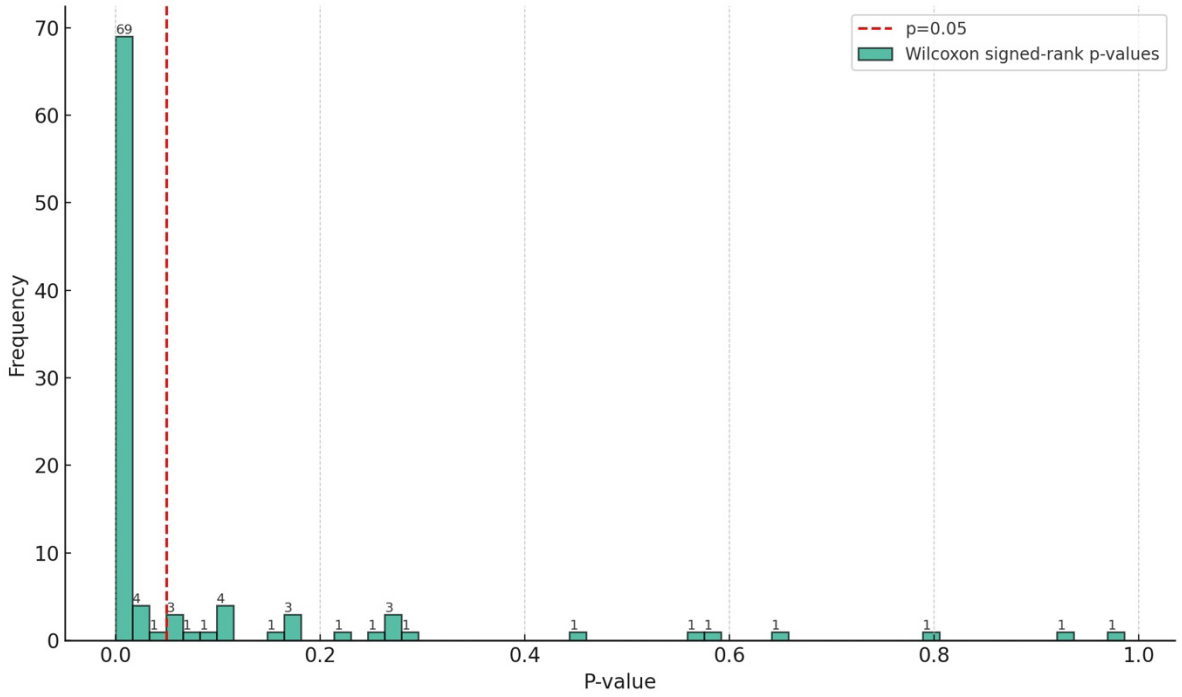
**Εικόνα 5.9: Ιστόγραμμα αποτελεσμάτων δοκιμών «Shapiro-Wilk» για το μέτρο αποτελεσματικότητας Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE).**

### 5.8.3 Πραγματοποίηση δοκιμής «Wilcoxon signed-rank».

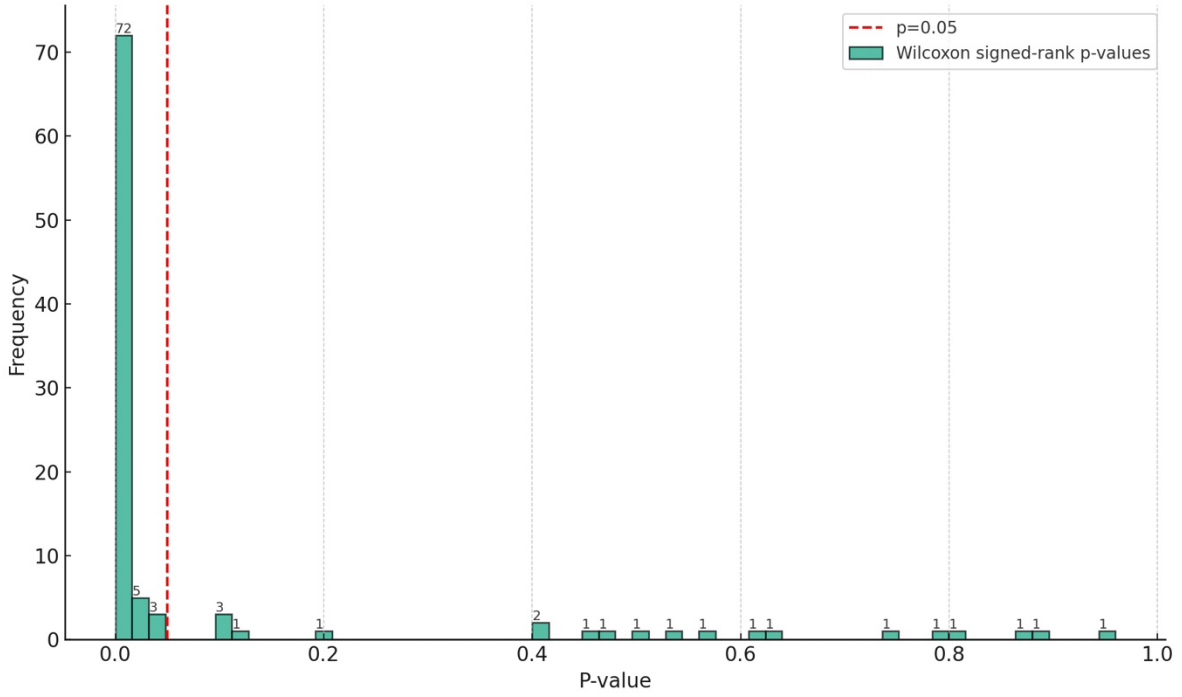
Η δοκιμή «Wilcoxon signed-rank» είναι μια μη παραμετρική στατιστική δοκιμή υποθέσεων που χρησιμοποιείται για τη σύγκριση δύο σχετικών δειγμάτων, αντιστοιχισμένων δειγμάτων ή επαναλαμβανόμενων μετρήσεων σε ένα μόνο δείγμα για να εκτιμηθεί εάν οι μέσες τάξεις του πληθυσμού τους διαφέρουν, δηλαδή με λίγα λόγια πρόκειται για μια δοκιμασία αντιστοιχισμένης διαφοράς (paired difference test). Η δοκιμή «Wilcoxon signed-rank» χρησιμοποιείται ως εναλλακτική λύση του αντιστοιχισμένου «t-test» του «Student» (paired Student's t-test) όταν ο πληθυσμός δεν ακολουθεί την κανονική κατανομή. Η μηδενική υπόθεση αυτής της δοκιμής είναι ότι η διάμεσος των διαφορών μεταξύ των αντιστοιχισμένων δειγμάτων είναι μηδέν, δηλαδή υποθέτει ότι δεν υπάρχει σταθερή διαφορά μεταξύ των αντιστοιχισμένων δειγμάτων.

Για την πραγματοποίηση της δοκιμής «Wilcoxon signed-rank» πραγματοποιήθηκαν πέντε βήματα. Στο πρώτο βήμα, υπολογίζονται οι διαφορές μεταξύ κάθε ζευγαριού τιμών. Στο δεύτερο βήμα, κατατάσσονται οι απόλυτες διαφορές, αγνοώντας τα μηδενικά. Κατά το τρίτο βήμα, αποδίδονται βαθμοί στις θετικές και στις αρνητικές διαφορές. Στο τέταρτο βήμα υπολογίζονται τα αθροίσματα των θετικών και των αρνητικών διαφορών και επιλέγεται το μικρότερο άθροισμα. Κατά το πέμπτο και τελευταίο βήμα, συγκρίνεται το επιλεγμένο άθροισμα (p-value) με μία κρίσιμη τιμή από την κατανομή «Wilcoxon signed-rank», η οποία στην πραγματοποιηθείσα δοκιμή είναι 0,05. Εάν το επιλεγμένο άθροισμα είναι μικρότερο από την κρίσιμη τιμή (0,05) τότε η διαφορά μεταξύ των ομάδων είναι στατιστικά σημαντική.

Στις Εικόνες 5.10 και 5.11 παρουσιάζονται σε Ιστόγραμμα τα αποτελέσματα των 100 δοκιμών «Wilcoxon signed-rank» για τα μέτρα αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος (RMSE) και Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE) αντίστοιχα. Από τα Ιστογράμματα προκύπτει ότι για το μέτρο αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος (RMSE) οι 74 δοκιμές ήταν μικρότερες από 0.05, που σημαίνει ότι σε 74 δοκιμές υπήρξε στατιστικά σημαντική διαφορά μεταξύ των δύο μοντέλων, ενώ για το μέτρο αποτελεσματικότητας Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE) οι 80 δοκιμές ήταν μικρότερες από 0.05, που σημαίνει αντίστοιχα ότι σε 80 δοκιμές υπήρξε στατιστικά σημαντική διαφορά μεταξύ των δύο μοντέλων.




**Εικόνα 5.10: Ιστόγραμμα αποτελεσμάτων δοκιμής «Wilcoxon signed-rank» για το μέτρο αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος (RMSE).**

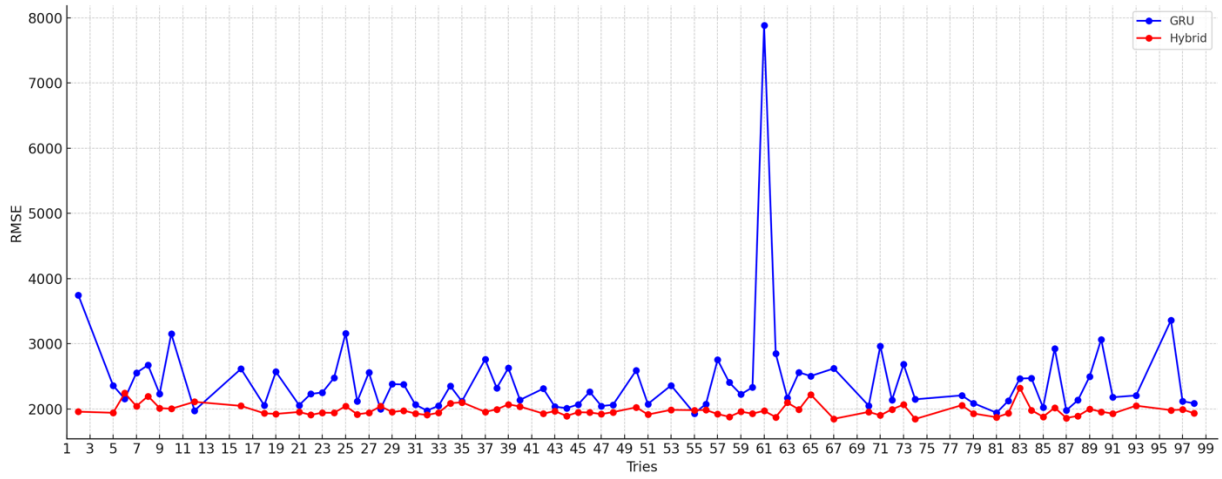



**Εικόνα 5.11: Ιστόγραμμα αποτελεσμάτων δοκιμής «Wilcoxon signed-rank» για το μέτρο αποτελεσματικότητας Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE).**

### 5.8.4 Αποτελέσματα σύγκρισης του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) με το Υβριδικό Μοντέλο (Hybrid Model) που χρησιμοποιεί και Ανάλυση Συναισθημάτων (Sentiment Analysis)

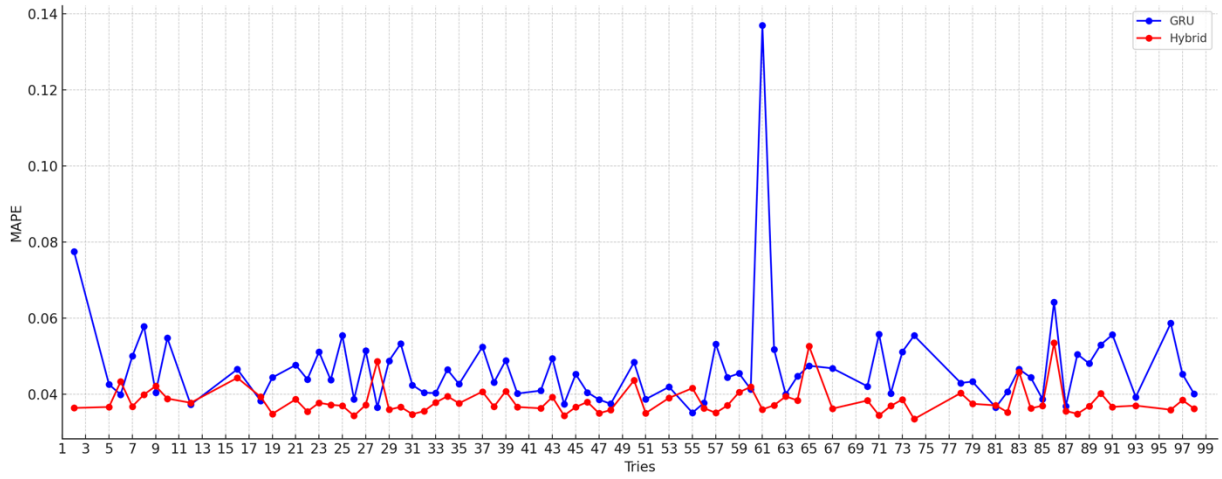
Από την πραγματοποίηση της δοκιμής «Wilcoxon signed-rank» προκύπτει ότι ένα μέρος από τις 100 συγκρίσεις μεταξύ των δύο μοντέλων ήταν στατιστικά σημαντικές για κάθε μέτρο αποτελεσματικότητας. Αυτό σημαίνει ότι προκειμένου να λάβουμε ένα σωστό αποτέλεσμα σύγκρισης μεταξύ των δύο μοντέλων θα πρέπει να λάβουμε υπόψη μόνο τις δοκιμές που υπήρξε στατιστικά σημαντική διαφορά μεταξύ τους για κάθε μέτρο αποτελεσματικότητας, καθώς στις υπόλοιπες θεωρούμε ότι τα μοντέλα ανταποκρίνονται το ίδιο.

Στην Εικόνα 5.12 παρουσιάζονται τα συγκριτικά αποτελέσματα μεταξύ του Υβριδικού μοντέλου (Hybrid) και του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) για το μέτρο αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος (RMSE). Στο διάγραμμα παρουσιάζονται μόνο οι δοκιμές στις οποίες υπήρξε στατιστικά σημαντική διαφορά σύμφωνα με τη δοκιμή «Wilcoxon signed-rank». Από το διάγραμμα προκύπτει ότι το Υβριδικό μοντέλο (Hybrid) ήταν πιο αποτελεσματικό για 70 δοκιμές, ενώ το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU) για 4. Αυτό πρακτικά σημαίνει ότι από τις δοκιμές στις οποίες υπήρξε σημαντική στατιστική διαφορά, σε ποσοστό 94,59% εξ' αυτών, το Υβριδικό μοντέλο (Hybrid) ήταν πιο αποτελεσματικό από το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU). Άρα συνολικά, όπως φαίνεται και στην Εικόνα 5.14, από τις 100 δοκιμές που πραγματοποιήθηκαν για το μέτρο αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος (RMSE), στο 70% των δοκιμών ήταν πιο αποτελεσματικό το Υβριδικό μοντέλο (Hybrid), στο 26% των δοκιμών δεν υπήρξε στατιστική διαφορά μεταξύ των μοντέλων, και τέλος στο 4% υπήρξε πιο αποτελεσματικό το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU).




**Εικόνα 5.12: Σύγκριση απόδοσης Υβριδικού μοντέλου (Hybrid) και μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) για το μέτρο αποτελεσματικότητας Μέσης Ρίζας Τετραγωνικού Σφάλματος (RMSE).**

Στην Εικόνα 5.13 παρουσιάζονται τα συγκριτικά αποτελέσματα μεταξύ του Υβριδικού μοντέλου (Hybrid) και του μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) για το μέτρο αποτελεσματικότητας Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE). Στο διάγραμμα παρουσιάζονται μόνο οι δοκιμές στις οποίες υπήρξε στατιστικά σημαντική διαφορά σύμφωνα με τη δοκιμή «Wilcoxon signed-rank». Από το διάγραμμα προκύπτει ότι το Υβριδικό μοντέλο (Hybrid) ήταν πιο αποτελεσματικό για 73 δοκιμές, ενώ το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU) για 7. Αυτό πρακτικά σημαίνει ότι από τις δοκιμές στις οποίες υπήρξε σημαντική στατιστική διαφορά, σε ποσοστό 91,25% εξ' αυτών το Υβριδικό μοντέλο (Hybrid) ήταν πιο αποτελεσματικό από το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU). Άρα συνολικά, όπως φαίνεται και στην Εικόνα 5.14, από τις 100 δοκιμές που πραγματοποιήθηκαν για το μέτρο αποτελεσματικότητας Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE), στο 73% των δοκιμών ήταν πιο αποτελεσματικό το Υβριδικό μοντέλο (Hybrid), στο 20% των δοκιμών δεν υπήρξε στατιστική διαφορά μεταξύ των μοντέλων, και τέλος στο 7% υπήρξε πιο αποτελεσματικό το μοντέλο Δικτύων Αναδρομικής Πύλης (GRU).




**Εικόνα 5.13: Σύγκριση απόδοσης Υβριδικού μοντέλου (Hybrid) και μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) για το μέτρο αποτελεσματικότητας Μέσου Απόλυτου Ποσοστιαίου Σφάλματος (MAPE).**

| **Μέτρο Αποτελεσματικότητας** | **Σύνολο δοκιμών** | **Δοκιμές όπου το Υβριδικό μοντέλο (Hybrid) ήταν πιο αποτελεσματικό** | **Δοκιμές όπου το μοντέλου Δικτύων Αναδρομικής Πύλης (GRU) ήταν πιο αποτελεσματικό** | **Δοκιμές όπου δεν υπήρξε στατιστικά σημαντική διαφορά μεταξύ των μοντέλων** |
|---|---|---|---|---|
| Μέση Ρίζα Τετραγωνικού Σφάλματος (RMSE) | 100 | 70 | 4 | 26 |
| Απόλυτο Ποσοστιαίο Σφάλμα (MAPE) | 100 | 73 | 7 | 20 |

**Εικόνα 5.14: Συνολικά αποτελέσματα δοκιμών.**

Συμπερασματικά, λαμβάνοντας υπόψη τα αποτελέσματα της Εικόνας 5.14, γίνεται εύκολα αντιληπτό ότι το Υβριδικό Μοντέλο (Hybrid Model) υπερτερεί σημαντικά έναντι του Μοντέλου Δικτύων Αναδρομικής Πύλης (GRU). Αυτή η υπεροχή μας δείχνει ότι η προσθήκη Ανάλυσης Συναισθημάτων (Sentiment Analysis) σε ένα μοντέλο Δικτύων Αναδρομικής Πύλης (GRU) βελτιώνει σημαντικά την απόδοση του. Εκτός από το απλό συμπέρασμα ότι η Ανάλυση Συναισθημάτων βελτιώνει το μοντέλο, αξίζει να αναλυθεί και ο τρόπος με τον οποίο προστέθηκε η Ανάλυση Συναισθημάτων. Πιο αναλυτικά, ο τρόπος προσθήκης της Ανάλυσης Συναισθημάτων μπορεί να χωριστεί σε δύο κατηγορίες, στον τρόπο με τον οποίο τεχνικά προστέθηκε και στον τρόπο με τον οποίο συμπεριφορικά προστέθηκε. Συμπεριφορικά, η προσθήκη της Ανάλυσης Συναισθημάτων πραγματοποιήθηκε προσομοιώνοντας τον τρόπο με τον οποίο συμπεριφέρεται ο μέσος επενδυτής κρυπτονομισμάτων. Πιο συγκεκριμένα, ο μέσος επενδυτής κρυπτονομισμάτων προκειμένου να αποφασίσει τον τρόπο που θα κινηθεί επενδυτικά, διαβάζει τις πρώτες ειδήσεις που θα του προσφέρει μία μηχανή αναζήτησης σε σχέση με το εκάστοτε κρυπτονόμισμα και με βάση αυτές συνειδητά ή ασυνείδητα αποφασίζει τον τρόπο που θα κινηθεί. Αυτό είναι λογικό να συμβαίνει, καθώς η πλειοψηφία των επενδυτών στα κρυπτονομίσματα ασχολείται ερασιτεχνικά με τις επενδύσεις οι οποίες δεν αποτελούν τον κύριο τρόπο βιοπορισμού τους. Τεχνικά, η Ανάλυση Συναισθημάτων προστέθηκε με την χρησιμοποίηση μεταμορφωτών (transformers) οι οποίοι φαίνεται πως έχουν την δυνατότητα να εξάγουν το νόημα ενός κειμένου καθώς μπορούν να ερμηνεύσουν τις σχέσεις μεταξύ πολύ μακρινών λέξεων μέσα σε ένα κείμενο.

# ΚΕΦΑΛΑΙΟ 6: ΣΥΜΠΕΡΑΣΜΑΤΑ, ΠΕΡΙΟΡΙΣΜΟΙ ΚΑΙ ΠΡΟΤΑΣΕΙΣ ΓΙΑ ΜΕΛΛΟΝΤΙΚΗ ΕΡΕΥΝΑ

## 6.1 Συμπεράσματα και περιορισμοί

Η εύλογη αξία και η τιμολόγηση των κρυπτονομισμάτων αποτελεί και θα συνεχίσει να αποτελεί ένα περίπλοκο πρόβλημα το οποίο όσο επεκτείνεται η χρήση των κρυπτονομισμάτων θα αποτελεί έναν πολύ μεγάλο πονοκέφαλο τόσο για τις επιχειρήσεις από την σκοπιά της διαχείρισης περιουσιακού στοιχείου, όσο και από τους επενδυτές που θα πρέπει να αξιολογούν την αξία τους ως περιουσιακό στοιχείο. Η πολυπλοκότητα του υπολογισμού της εύλογης αξίας και της τιμολόγησης των κρυπτονομισμάτων έχει ωθήσει την επιστημονική κοινότητα να αφήσει τις κλασικές στατιστικές μεθόδους και να στραφεί προς τις τεχνικές Μηχανικής Μάθησης οι οποίες μπορούν να ερμηνεύσουν αποτελεσματικά μη γραμμικά δεδομένα. Ωστόσο, οι τεχνικές Μηχανικής Μάθησης από μόνες τους φαίνεται πως περιορίζονται σε ένα άνω όριο απόδοσης, το οποίο αποδείχθηκε στην παρούσα εργασία ότι μπορεί να βελτιωθεί με την συνδυαστική χρήση Ανάλυσης Συναισθημάτων (Sentiment Analysis).

## 6.2 Ανοιχτά ζητήματα και προτάσεις για μελλοντική έρευνα

Τα ζητήματα που παραμένουν ανοιχτά και θα είχε πολύ ενδιαφέρον να επιλυθούν ώστε να επιτευχθεί περαιτέρω βελτίωση του Υβριδικού Μοντέλου (Hybrid Model) είναι το φιλτράρισμα τον ειδησεογραφικών άρθρων μέσα από μία διαδικασία αξιολόγησης, η προσθήκη επιπλέον παραμέτρων της Αλυσίδας Συστοιχιών (Blockchain) του εκάστοτε κρυπτονομίσματος, η χρήση πιο αναβαθμισμένων Μεταμορφωτών (Transformers) που χρησιμοποιούνται για την μετατροπή του κειμένου των άρθρων σε μεταβλητές, και τέλος η χρήση του Υβριδικού Μοντέλου μαζί με μία επενδυτική στρατηγική για την αποκόμιση κέρδους επενδύοντας σε κρυπτονομίσματα.

# ΑΝΑΦΟΡΕΣ

[1] Casey, M. (2018). *The impact of blockchain technology on Finance: A catalyst for change*. ICMB International Center for Monetary and Banking Studies.

[2] Encean, A.-A. and Zinca, D. (2022) 'Cryptocurrency price prediction using LSTM and GRU Networks', *2022 International Symposium on Electronics and Telecommunications (ISETC)* [Preprint]. doi:10.1109/isetc56213.2022.10010329.

[3] Flach, P. (2017). Machine learning: The art and science of algorithms that make sense of data. Cambridge University Press.

[4] GERON, A. (2017). Hands-on machine learning with scikit-learn and tensorflow: Concepts, tools and techniques to build Intelligent Systems. O'Reilly.

[5] Guido, S. (2016). Introduction to machine learning with python. O'Reilly Media.

[6] Ji, S., Kim, J. and Im, H. (2019) "A comparative study of bitcoin price prediction using Deep Learning," *Mathematics*, 7(10), p. 898. Available at: https://doi.org/10.3390/math7100898.

[7] Khedr, A.M. *et al.* (2021) "Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey," *Intelligent Systems in Accounting, Finance and Management*, 28(1), pp. 3–34. Available at: https://doi.org/10.1002/isaf.1488.

[8] Mitchell, T. M. (1997). *Machine learning*. MacGraw-Hill.

[9] Mohapatra, S., Ahmed, N. and Alencar, P. (2019) "Kryptooracle: A real-time cryptocurrency price prediction platform using Twitter sentiments," *2019 IEEE International Conference on Big Data (Big Data)* [Preprint]. Available at: https://doi.org/10.1109/bigdata47090.2019.9006554.

[10] Monrat, A. A., Schelen, O., & Andersson, K. (2019). A survey of blockchain from the perspectives of applications, challenges, and opportunities. *IEEE Access*, *7*, 117134–117151. https://doi.org/10.1109/access.2019.2936094

[11] Nakamoto, S. (2008). Bitcoin: A Peer-to-Peer Electronic Cash System. https://bitcoin.org/bitcoin.pdf

[12] OpenAI. (2023). ChatGPT [Large language model]. https://chat.openai.com

[13] Patra, G.R. and Mohanty, M.N. (2022) "Price prediction of cryptocurrency using a multi-layer gated recurrent unit network with multi features," *Computational Economics* [Preprint]. Available at: https://doi.org/10.1007/s10614-022-10310-1.

[14] S. Yaga, D., Mell, P., Roby, N., & Scarfone, K. (2018). Blockchain Technology Overview. https://doi.org/10.6028/nist.ir.8202

[15] Sridhar, S. and Sanagavarapu, S. (2021) "Multi-head self-attention transformer for dogecoin price prediction," *2021 14th International Conference on Human System Interaction (HSI)* [Preprint]. Available at: https://doi.org/10.1109/hsi52170.2021.9538640.

[16] Sutton, R. S., Bach, F., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press Ltd.

[17] Tanenbaum, A. S., & Steen, M. V. (2002). *Distributed systems: Principles and paradigms*. Prentice Hall.

[18] Treiblmaier, H. (2022). Defining the internet of value. *Future of Business and Finance*, 3–10. https://doi.org/10.1007/978-3-030-78184-2_1

[19] van Tilborg, H. C. A., & Jajodia, S. (2011). *Encyclopedia of cryptography and security*. Springer US.

[20] Vaswani, Ashish & Shazeer, Noam & Parmar, Niki & Uszkoreit, Jakob & Jones, Llion & Gomez, Aidan & Kaiser, Lukasz & Polosukhin, Illia, "Attention is all you need", 2017.

[21] Wang, Y. and Chen, R. (2020) "Cryptocurrency price prediction based on multiple market sentiment," *Proceedings of the Annual Hawaii International Conference on System Sciences* [Preprint]. Available at: https://doi.org/10.24251/hicss.2020.136.

[22] Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R., & Le, Q. V. (2020). XLNet: Generalized Autoregressive Pretraining for Language Understanding. arXiv [Cs.CL]. Retrieved from http://arxiv.org/abs/1906.08237

[23] Zaheer, M., Guruganesh, G., Dubey, A., Ainslie, J., Alberti, C., Ontanon, S., … Ahmed, A. (2021). Big Bird: Transformers for Longer Sequences. arXiv [Cs.LG]. Retrieved from http://arxiv.org/abs/2007.14062

[24] Zamani, N.A., Yan, J.L. and Yusof, A.M. (2022) "Cryptocurrency price prediction using bi-gru model with English and Malay news sentiment features," *2022 3rd International Conference on Artificial Intelligence and Data Sciences (AiDAS)* [Preprint]. Available at: https://doi.org/10.1109/aidas56890.2022.9918725.

[25] Ι.Βλαχάβας, Π.Κεφαλάς, Ν.Βασιλειάδης, Φ.Κόκκορας, Η.Σακελλαρίου. (2011). Τεχνητή Νοημοσύνη - Γ΄Έκδοση, ISBN: 978-960-8396-64-7