



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και
Υπολογιστών

Μηχανική Μάθηση και Ανάλυση Συναισθήματος για την πρόβλεψη της τιμής του Bitcoin

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΝΙΚΟΛΑΣ ΔΕΡΒΕΝΗΣ

Επιβλέπων : Ανδρέας-Γεώργιος Σταφυλοπάτης

Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων : Γεώργιος Αλεξανδρίδης

Ε.ΔΙ.Π. Ε.Μ.Π.

Αθήνα, Οκτώβριος 2018



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών

Τομέας Τεχνολογίας Πληροφορικής και
Υπολογιστών

Μηχανική Μάθηση και Ανάλυση Συναισθήματος για την πρόβλεψη της τιμής του Bitcoin

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΝΙΚΟΛΑΣ ΔΕΡΒΕΝΗΣ

Επιβλέπων : Ανδρέας-Γεώργιος Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.

Συνεπιβλέπων : Γεώργιος Αλεξανδρίδης
Ε.ΔΙ.Π. Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 31η Οκτωβρίου 2018.

.....
Ανδρέας-Γεώργιος Σταφυλοπάτης Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π.

.....
Νικόλαος Παπασπύρου
Αναπληρωτής Καθηγητής Ε.Μ.Π. Αναπληρωτής Καθηγητής Ε.Μ.Π.

Αθήνα, Οκτώβριος 2018

.....
Νικόλας Δερβένης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών
Ε.Μ.Π.

Copyright © Νικόλας Δερβένης, 2018.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

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

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

Περίληψη

Στην παρούσα διπλωματική εργασία εξετάστηκε αν η πληροφορία που υπάρχει στο Διαδίκτυο μπορεί να χρησιμοποιηθεί για την πρόβλεψη της τιμής των κρυπτονομισμάτων. Συγκεκριμένα, χρησιμοποιώντας αναδρομικά νευρωνικά δίκτυα, στόχος ήταν να προβλεφθεί η κίνηση (προς τα πάνω ή προς τα κάτω) της τιμής του Bitcoin, σε περίοδο μιας ημέρας. Για το σκοπό αυτό, τα επιμέρους μοντέλα που εκπαιδεύτηκαν δεχόντουσαν ως είσοδο πληροφορία από τις παλαιότερες τιμές του Bitcoin, την ανάλυση συναισθήματος της καθημερινής ειδησεογραφίας (θετική ή αρνητική), τα Google Trends και τα tweets σχετικά με το συγκεκριμένο κρυπτονόμισμα. Τα αποτελέσματα έδειξαν πως τα πιο απλά μοντέλα, που βασίζονται μόνο στις παλαιότερες τιμές του Bitcoin και στην ανάλυση συναισθήματος της καθημερινής ειδησεογραφίας, έδωσαν την καλύτερη απόδοση με ποσοστό επιτυχούς πρόβλεψης 84.21%. Τέλος, τα αποτελέσματα οδηγούν στην παραδοχή ότι η πληροφορία από τα Google Trends και τα tweets, στην περίοδο μιας ημέρας, δεν φαίνεται να συμβάλλει θετικά στην πρόβλεψη, τουλάχιστον στο βαθμό που αναλύθηκε.

Λέξεις κλειδιά

Μηχανική Μάθηση, Bitcoin, Κρυπτονόμισμα, Ανάλυση Συναισθήματος, Twitter, Google Trends

Abstract

The current thesis examines whether the information available online can be used to predict the price of cryptocurrencies. More specifically, using recurrent neural networks, the objective has been to predict the price movement (upwards or downwards) of Bitcoin, on a daily basis. For this purpose, various models have been trained on input based on previous Bitcoin prices, sentiment analysis of the daily news (positive or negative), Google Trends and tweets regarding this specific cryptocurrency. The results showed that the simpler models, based on the previous Bitcoin prices and the sentiment analysis of the daily news have been the most successful, with a prediction accuracy of 84.21%. Finally, the results may imply that the information from Google Trends and the tweets, on a daily basis, does not seem to positively affect prediction accuracy, at least to the extent it has been analyzed.

Key words

Machine Learning, Bitcoin, Cryptocurrency, Sentiment Analysis, Twitter, Google Trends

Ευχαριστίες

Ο χρόνος περνάει και εμείς δεν το καταλαβαίνουμε. Αλλά ίσως έτσι πρέπει να είναι. Ίσως έτσι καταλαβαίνεις ότι κάνεις κάτι σημαντικό.. ή αν όχι κάτι σημαντικό έστω κάτι που σε γεμίζει. Φτάνοντας λοιπόν στο τέλος αυτής της ακαδημαϊκής πορείας και λίγο πριν πάρω το δίπλωμά που (ωιμέ!) όλοι τόσο λαχταρούμε και λίγο πριν ξεκινήσω έναν νέο κύκλο που η αλήθεια είναι ότι νοιώθω ενθουσιασμό και άγχος· μέσα σε αυτή τη γλυκόπικρη αίσθηση θα πάρω λίγο χρόνο να ευχαριστήσω τους ανθρώπους που νοιώθω ότι θέλω να ευχαριστήσω.

Αρχικά θέλω να ευχαριστήσω την οικογένειά μου γιατί χωρίς αυτήν δεν θα είχα μπει στο πολυτεχνείο οπότε δεν θα άνοιγε καν αυτός ο κύκλος. Για να μην φλυαρώ θα επιλέξω ένα πράγμα που θεωρώ πως μου έχει δώσει ο καθένας, και είναι σχετικό. Ευχαριστώ την μητέρα μου, που μου έμαθε να έχω θάρρος, τον πατέρα μου, που μου μετέδωσε την αγάπη για τις θετικές επιστήμες και τον αδερφό μου, που μου έμαθε να πειραματίζομαι.

Εν συνεχεία, θα ήθελα να ευχαριστήσω το Εργαστήριο Ευφυών Συστημάτων, Περιεχομένου και Αλληλεπίδρασης και την εξεταστική επιτροπή που δέχθηκε να κάνω την διπλωματική αυτή και τελικά να μπορέσω να κλείσω αυτόν τον κύκλο. Συγκεκριμένα θέλω να ευχαριστήσω τον κ. Σταφυλοπάτη Ανδρέα-Γεώργιο, καθηγητή Ε.Μ.Π. και τους κ.κ. Στάμου Γεώργιο και Παπασπύρου Νικόλαο, Αναπληρωτές Καθηγητές Ε.Μ.Π. Επίσης, θέλω να ευχαριστήσω τον κ. Αλεξανδρίδη Γεώργιο, Ε.ΔΙ.Π. Ε.Μ.Π. που ήταν δίπλα μου στην πορεία αυτής της διπλωματικής και αφιέρωσε χρόνο και γνώση για να με βοηθήσει.

Τέλος, θα ήθελα να ευχαριστήσω τους φίλους μου. Με την λέξη φίλους εννοώ όλους τους κοντινούς μου ανθρώπους, εκτός οικογένειας. Τους ευχαριστώ γιατί με αγκαλιάζουν και με δοκιμάζουν. Το βλέπω στα μάτια τους. Αυτό με βοηθάει να εξελίξομαι και τελικά με οδηγεί να γίνω αυτός που είμαι. Καλύτερος η χειρότερος, θα φανεί στην πορεία. Μάλλον και τα δύο. Επίσης χωρίς αυτούς δύσκολα θα ήμουν διπλωματούχος αυτή τη στιγμή, οπότε τους ευχαριστώ και γι' αυτό! Κλείνοντας, τους ζητώ ένα συγνώμη αν τους έχω ταλαιπωρήσει αλλά τι να κάνουμε έτσι είναι αυτά.

Ο χρόνος περνάει και εμείς δεν το καταλαβαίνουμε. Αλλά ίσως έτσι πρέπει να είναι.

Υγεία και Αγάπη!

Νικόλας Δερβένης,

Αθήνα, 31η Οκτωβρίου 2018

Περιεχόμενα

Περίληψη	5
Abstract	7
Ευχαριστίες	9
Περιεχόμενα	11
Κατάλογος πινάκων	15
Κατάλογος σχημάτων	17
1. Εισαγωγή	19
1.1 Τι είναι το Bitcoin	19
1.2 Μηχανική Μάθηση	19
1.3 Χρονοσειρές	21
1.3.1 Χρήση τεχνικών μηχανικής μάθησης για την πρόβλεψη χρονοσειρών	22
1.4 Ανάλυση συναισθήματος	22
1.5 Κίνητρο	23
1.5.1 Google Trends	24
1.5.2 Μέσα Κοινωνικής Δικτύωσης	26
1.5.3 Twitter	27
1.5.4 Hashtag	28
1.6 Δομή της Διπλωματικής Εργασίας	30
2. Θεωρητικό Υπόβαθρο	33
2.1 Bitcoin	33
2.1.1 Εισαγωγή	33
2.1.2 Παραγωγή του Bitcoin (mining)	34
2.1.3 Αγορά ξένου συναλλάγματος	34
2.2 Νευρωνικά Δίκτυα	36

2.3	Αναδρομικά Νευρωνικά Δίκτυα	39
2.3.1	Εισαγωγή	39
2.3.2	Περιγραφή της Αρχιτεκτονικής των Αναδρομικών Νευρωνικών Δικτύων	40
2.3.3	Τεχνικές Εκμάθησης στα Αναδρομικά Νευρωνικά Δίκτυα	42
2.3.4	Backpropagation Through Time (BPTT)	45
2.4	Επιλογή υπερπαραμέτρων στα αναδρομικά νευρωνικά δίκτυα	49
2.4.1	Αριθμός κρυφών επιπέδων	49
2.4.2	Πλήθος κόμβων κρυφών επιπέδων	49
2.5	Νευρωνικά Δίκτυα Elman και Jordan	51
2.5.1	Δίκτυα Elman	51
3.	Συλλογή Δεδομένων	53
3.1	Εισαγωγή	53
3.2	Συλλογή Δεδομένων	54
3.2.1	Προηγούμενες τιμές του Bitcoin	54
3.2.2	Ανάλυση συναισθήματος από την ειδησιογραφία της ημέρας	56
3.2.3	Google Trends	58
3.2.4	Tweets	59
3.2.5	Τελική συλλογή δεδομένων	59
3.3	Χρήση τεχνικών ανάλυσης συναισθήματος στην ειδησιογραφία της ημέρας	60
3.3.1	Μηχανική Μάθηση στην Ανάλυση Συναισθήματος	60
3.4	Διαδικασία μεταβολής της τιμής του Bitcoin	61
4.	Υλοποίηση, εργαλεία και αποτελέσματα	65
4.1	Επιλογή Νευρωνικού Δικτύου	65
4.2	Εργαλεία Υλοποίησης	66
4.2.1	Γλώσσα προγραμματισμού	66
4.2.2	Βιβλιοθήκη Μηχανικής Μάθησης	67
4.3	Πειράματα και Αποτελέσματα	68
4.3.1	Πειραματική Διαδικασία	68
4.3.2	Αποτελέσματα	70
5.	Επίλογος	75
5.1	Συνεισφορά	75
5.2	Μελλοντική Έρευνα	76

Βιβλιογραφία 77

Κατάλογος πινάκων

3.1	Πρώτες τέσσερις σειρές του πίνακα της ιστορικής τιμής του Bitcoin. . . .	54
3.2	Νέος πίνακας της ιστορικής τιμής του Bitcoin.	55
3.3	Klout Score για τις μεγαλύτερες σε κυκλοφορία ειδησεογραφικές πηγές. . .	57
3.6	Συλλογή δεδομένων με την τιμή του Bitcoin και την αποτίμηση του συναισθήματος	58
3.7	Συλλογή δεδομένων με την τιμή του Bitcoin, την αποτίμηση συναισθήματος και το Google Trend της ημέρας	59
3.8	Συλλογή δεδομένων με την τιμή του Bitcoin, την αποτίμηση συναισθήματος και τον ημερήσιο αριθμό tweets που εμπεριέχουν τον όρο «#Bitcoin»	59
3.9	Η συνολική συλλογή δεδομένων	60
3.4	Παράδειγμα του συνόλου των ειδησεογραφικών τίτλων.	62
3.5	Παράδειγμα του συνόλου των ειδησεογραφικών τίτλων, εμπλουτισμένου με ανάλυση συναισθήματος	63
4.1	Εύρος τιμών στην αναζήτηση πλέγματος.	69
4.2	Παράμετροι που έμειναν σταθεροί κατά την διαδικασία εκπαίδευσης. . . .	69
4.3	Τελική απόδοση των μοντέλων στο σύνολο ελέγχου	70
4.4	Βέλτιστη αρχιτεκτονική για το μοντέλο Bitcoin	71
4.5	Βέλτιστη αρχιτεκτονική για το μοντέλο Bitcoin & Sentiment	71

Κατάλογος σχημάτων

1.1	Δεδομένα εκπαίδευσης, επαλήθευσης και δοκιμής μιας τυχαίας χρονοσειράς	21
1.2	Η χρονοσειρά της μετοχής της Apple για διάστημα ενός έτους	21
1.3	Ισοτιμία ευρώ δολαρίου (EUR/USD) για διάρκεια ενός έτους	22
1.4	Ισοτιμία Bitcoin δολαρίου (BTC/USD) για διάρκεια ενός έτους	23
1.5	Μέσος καθημερινός χρόνος στο διαδίκτυο για κάθε γενιά. (Πηγή: <i>GlobalWebIndex, Q2 2017, παγκόσμιο δείγμα, συμμετέχοντες: 43.592, ηλικίες: 16-34</i>)	24
1.6	Google Trend για συχνότητα αναζήτησης του όρου «Summer».	25
1.7	Google Trend για συχνότητα αναζήτησης του όρου «Jingle Bells»	26
1.8	Google Trend για συχνότητα αναζήτησης του όρου «Bitcoin»	27
1.9	Η εξέλιξη της τιμής του Bitcoin (σε δολάρια) σε διάστημα τριών χρόνων. . .	28
1.10	Η ομοιότητα μεταξύ του Google Trend «Bitcoin» και τιμής του Bitcoin . . .	29
1.11	Ενεργοί χρήστες (σε εκατομμύρια) μέχρι τον Ιούλιο 2018, για μερικά από τα μεγαλύτερα παγκόσμια μέσα κοινωνικής δικτύωσης (Πηγή: <i>Statista</i>)	30
1.12	Αριθμός των Tweets ημερησίως σε χιλιάδες, που συμπεριλαμβάνουν το #Bitcoin. .	31
2.1	Ιατρικό μοντέλο ενός νευρώνα	37
2.2	Μαθηματικό μοντέλο ενός νευρώνα του νευρωνικού δικτύου.	37
2.3	Δύο βασικές κατηγορίες νευρωνικών δικτύων.	38
2.4	Δίκτυο πρόσθιας τροφοδότησης με 2 κρυφά επίπεδα	44
2.5	Αναπαράσταση αναδρομικών δικτύων «ξεδιπλωμένων» στο χρόνο.	46
2.6	Νευρωνικά Δίκτυα Elman και Jordan	52
4.1	Γραφική αναπαράσταση των παραπάνω μοντέλων	72
4.2	Απόδοση μοντέλων στο σύνολο ελέγχου	73

Κεφάλαιο 1

Εισαγωγή

1.1 Τι είναι το Bitcoin

Το Bitcoin είναι το πρώτο και δημοφιλέστερο κρυπτονόμισμα. Τα κρυπτονομίσματα είναι μια νέα – πολλά υποσχόμενη – τεχνολογία που επιτρέπει έναν εναλλακτικό τρόπο ψηφιακής πληρωμής, δηλαδή μια εναλλακτική μορφή χρημάτων. Πιο συγκεκριμένα, το Bitcoin, παρέχει ένα σύστημα πληρωμών μέσω ενός συναινετικού δικτύου. Είναι το πρώτο αποκεντρωμένο δίκτυο πληρωμής μεταξύ *ομότιμων* (peer-to-peer), δηλαδή λειτουργεί εξολοκλήρου μεταξύ των χρηστών χωρίς την ανάγκη κάποιας κεντρική αρχής ή μεσάζοντα. Το Bitcoin και η απόδειξη της λειτουργίας του δημοσιεύτηκε το 2009 από τον Satoshi Nakamoto [Naka09]. Ιδιαίτερο χαρακτηριστικό του Bitcoin είναι ότι κανείς δεν είναι ιδιοκτήτης του δικτύου. Αντιθέτως το Bitcoin ελέγχεται από τους χρήστες σε όλο τον κόσμο με τις συναλλαγές που καταγράφονται στο δίκτυο να επαληθεύονται από έναν αλγόριθμο που λέγεται «*proof of work*». Η διαδικασία επαλήθευσης είναι αυτή που δημιουργεί και νέα Bitcoin, μια διαδικασία που αποκαλείται *εξόρυξη* (mining).

1.2 Μηχανική Μάθηση

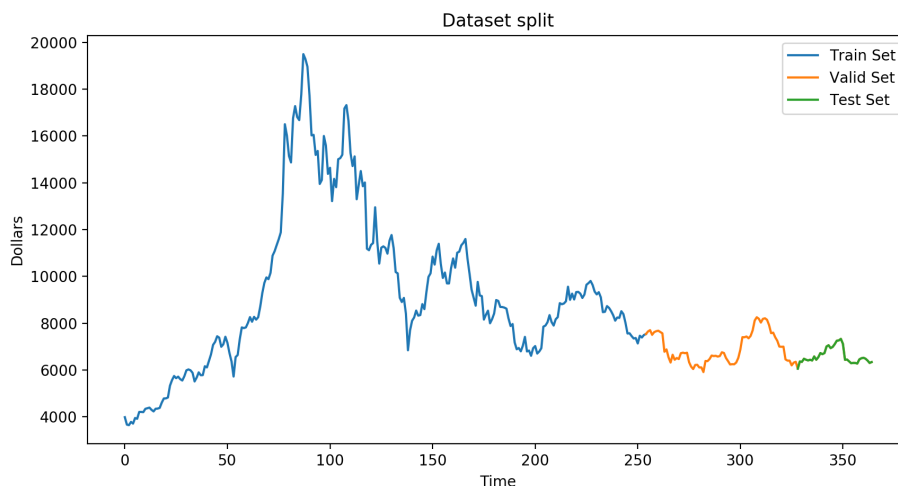
Η πρόβλεψη της τιμής του Bitcoin είναι υψίστης σημασίας και στην διπλωματική αυτή εργασία τα εργαλεία και οι μέθοδοι που θα χρησιμοποιηθούν θα έχουν ως βάση τεχνικές μηχανικής μάθησης.

Η *μηχανική μάθηση* (machine learning) είναι ένα πεδίο της επιστήμης των υπολογιστών που χρησιμοποιεί τεχνικές στατιστικής ώστε να δώσει σε υπολογιστικά συστήματα την δυνατότητα να «μάθουν» από δεδομένα χωρίς να χρειαστεί να ακολουθήσουν κάποιον ντετερμινιστικό αλγόριθμο [Samu59]. Μια διαδικασία μηχανικής μάθησης αποτελείται, σε γενικές γραμμές, από τα παρακάτω βήματα:

1. **Συλλογή δεδομένων**, που θα χρησιμοποιηθούν για την εκπαίδευση και την δοκιμή του μοντέλου

2. **Προετοιμασία δεδομένων.** Εν συνεχεία, τα δεδομένα θα πρέπει να «καθαριστούν», δηλαδή να απομακρυνθούν δεδομένα κακής ποιότητας, να ταξινομηθούν σε μια επιθυμητή σειρά και εν τέλει να τροποποιηθούν, ώστε να έχουν μια ενιαία μορφή. Τα δυο αυτά πρώτα βήματα είναι συνήθως χρονοβόρα και πολύ μεγάλης σημασίας, καθώς είναι η βάση πάνω στην οποία αργότερα στήνονται τα μοντέλα πρόβλεψης
3. **Επιλογή μοντέλου μηχανικής μάθησης.** Σε αυτό το στάδιο γίνεται η δοκιμή και τέλος η επιλογή των μοντέλων που θα χρησιμοποιηθούν. Τα μοντέλα μηχανικής μάθησης είναι πολλά και δεν υπάρχει σαφής κατηγοριοποίηση ποιο μοντέλο ταιριάζει σε ποια εργασία. Ωστόσο υπάρχουν αποδεδειγμένες ομάδες μοντέλων που προτιμώνται για συγκεκριμένες διεργασίες όπως οι *μηχανές διανυσμάτων υποστήριξης* (support vector machines) ή τα *δέντρα αποφάσεων* (decision trees) που ταιριάζουν σε προβλήματα κατηγοριοποίησης, ενώ τα συνελκτικά δίκτυα ταιριάζουν σε προβλήματα επεξεργασίας εικόνας. Στην διπλωματική αυτή θα χρησιμοποιηθούν αναδρομικά νευρωνικά δίκτυα και θα αναλυθούν εκτενέστερα σε επόμενο Κεφάλαιο.
4. **Εκπαίδευση.** Το στάδιο αυτό μπορεί από πολλούς να θεωρηθεί η ουσία της μηχανικής μάθησης. Σε αυτό το σημείο ένα κομμάτι (περίπου το 60-80%) από τα καλώς οργανωμένα δεδομένα που έχουν συλλεχθεί στα προηγούμενα στάδια περνούν από το μοντέλο μηχανικής μάθησης με σκοπό αυτό να «εκπαιδευτεί» και να μάθει να αναλύει τέτοιου τύπου δεδομένα.
5. **Εκτίμηση.** Σε αυτό το σημείο ένα μικρότερο κομμάτι από τα δεδομένα (10-20%) χρησιμοποιείται με σκοπό να δοκιμαστεί αν το εκπαιδευμένο πλέον μοντέλο μπορεί να ανταποκριθεί σε νέες εισόδους και έτσι να γίνουν οι τελικοί υπολογισμοί και οι διορθώσεις ανάλογα με την απόδοση του.
6. **Δοκιμή.** Σε αυτό το επίπεδο, το τελευταίο κομμάτι των δεδομένων θα χρησιμοποιηθεί ώστε να μετρηθεί η τελική απόδοση του μοντέλου.
7. **Βελτίωση μοντέλου.** Τις περισσότερες φορές μετά την δοκιμή θα χρειαστεί να γίνουν αλλαγές τόσο στην αρχιτεκτονική του μοντέλου όσο και στις υπερπαραμέτρους μέχρι να υπάρχει το επιθυμητό αποτέλεσμα. Σε αυτή την περίπτωση η διαδικασία επιστρέφει στα αρχικά βήματα και δοκιμάζεται ξανά το καινούριο μοντέλο.

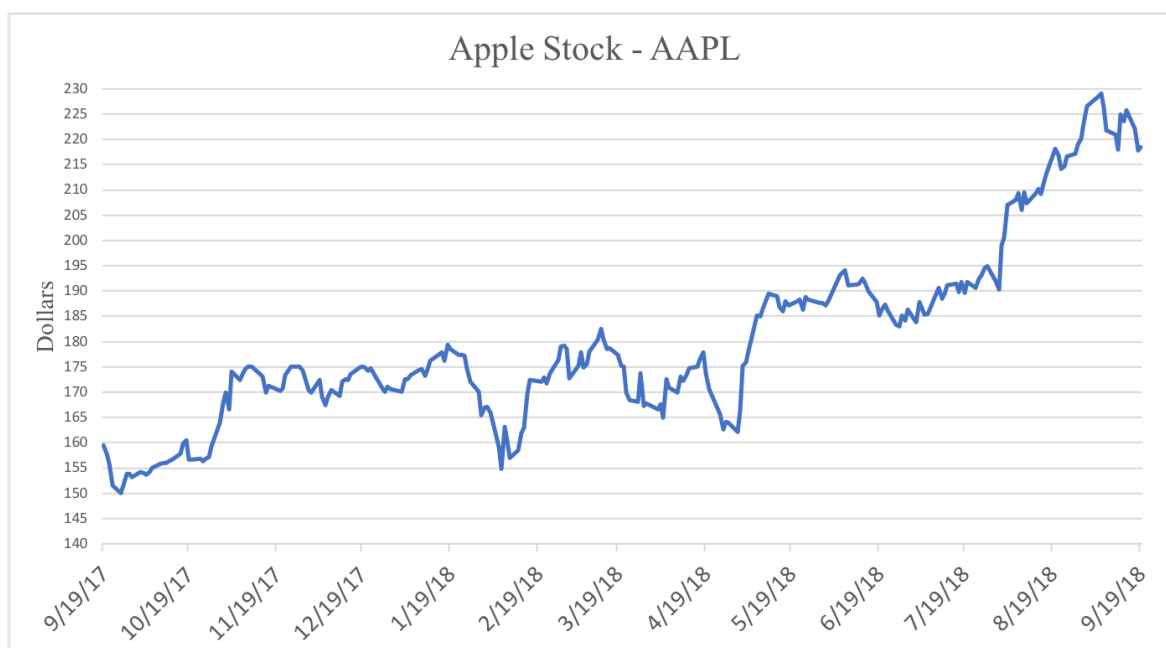
Στο Σχήμα 1.1 απεικονίζεται πως μπορούν να χωριστούν οι τιμές μιας χρονοσειράς σε τρία μέρη: εκπαίδευση, επαλήθευση και δοκιμή.



Σχήμα 1.1: Δεδομένα εκπαίδευσης, επαλήθευσης και δοκιμής μιας τυχαίας χρονοσειράς

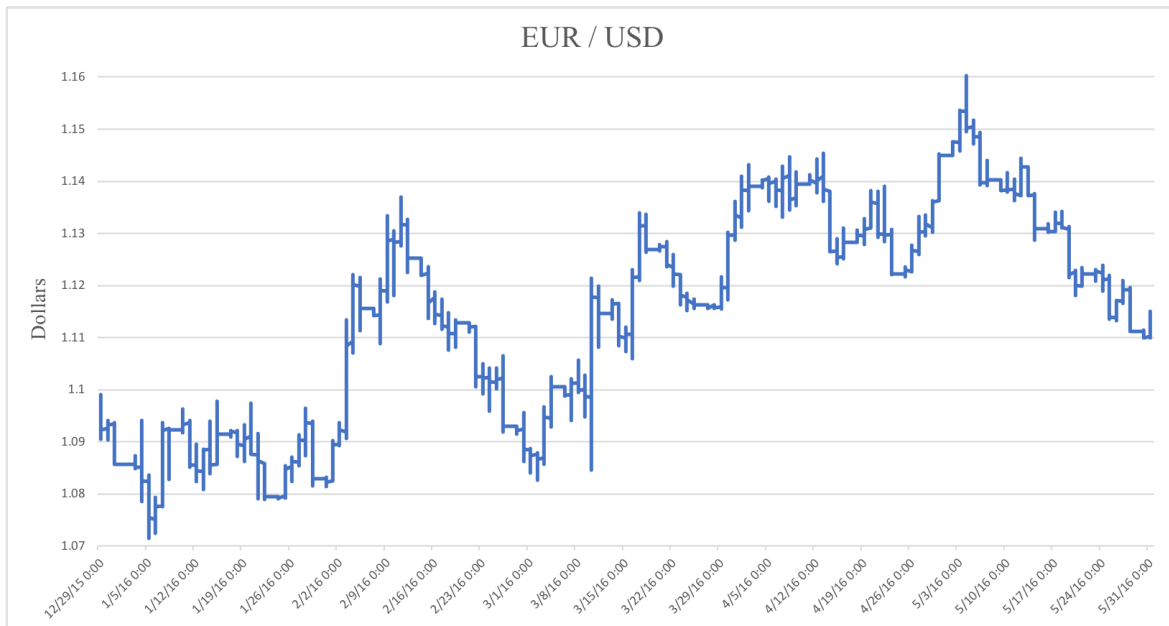
1.3 Χρονοσειρές

Η *χρονοσειρά* (time series) είναι μια σειρά από σημεία δεδομένων τοποθετημένα σε μια χρονική σειρά. Συνήθως η χρονοσειρά αποτελείται από σημεία που ισαπέχουν χρονικά και έτσι είναι μια σειρά διακριτών χρονικά σημείων. Στο Σχήμα 1.2 παρουσιάζεται η χρονοσειρά μιας μετοχής ενώ στα Σχήματα 1.3 και 1.4 οι χρονοσειρές από την ισοτιμία ευρώ δολαρίου και Bitcoin δολαρίου αντίστοιχα¹



Σχήμα 1.2: Η χρονοσειρά της μετοχής της Apple για διάστημα ενός έτους

¹ Πηγή: Yahoo Finance



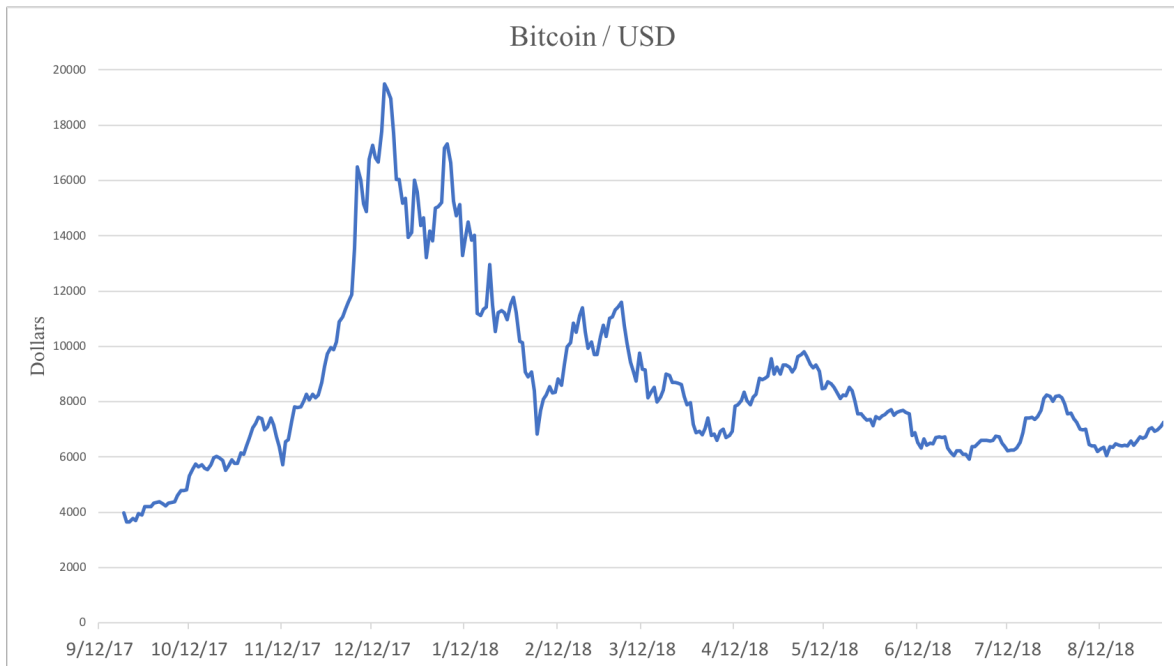
Σχήμα 1.3: Ισοτιμία ευρώ δολαρίου (EUR/USD) για διάρκεια ενός έτους

1.3.1 Χρήση τεχνικών μηχανικής μάθησης για την πρόβλεψη χρονοσειρών

Τόσο στον ακαδημαϊκό χώρο όσο και στην βιομηχανία, η *πρόβλεψη χρονοσειράς* (time-series prediction) είναι ένα μείζον θέμα που ερευνάται συνεχώς με τις πιο σύγχρονες τεχνικές και μεθόδους. Τις τελευταίες δεκαετίες η μηχανική μάθηση έχει γίνει μια πολύ δημοφιλής τεχνολογία λόγω την μεγάλης υπολογιστικής δύναμης και της δυνατότητας εύρεσης μεγάλου όγκου δεδομένων, ενώ ταυτόχρονα έχει σημειώσει πολλές επιτυχίες. Συνεπώς, στην πρόβλεψη χρονοσειράς έχουν δοκιμαστεί και δημοσιευθεί διάφορες τεχνικές που χρησιμοποιείται η μηχανική μάθηση [Sapa09, Chen06]. Η χρονοσειρά του Bitcoin είναι χρηματοοικονομικής φύσης και σε τέτοιες χρονοσειρές η μηχανική μάθηση έχει επίσης παίξει πολύ σημαντικό ρόλο [Cao03, jKim03, Gest01, Lu09].

1.4 Ανάλυση συναισθήματος

Η *ανάλυση συναισθήματος* (sentiment analysis) [Pang08] ή αλλιώς *εξόρυξη γνώμης* (opinion mining) βασίζεται στην επεξεργασία φυσικής επεξεργασίας και στην ανάλυση κειμένου, ώστε συστηματικά να εντοπίζει, να εξάγει, να ποσοτικοποιεί και να μελετά συναισθηματικές καταστάσεις και υποκειμενικές πληροφορίες. Στην βιομηχανία, η ανάλυση συναισθήματος έχει βρει ευρεία εφαρμογή σε περιοχές όπως οι κριτικές προϊόντων, οι απαντήσεις ερωτηματολογίων, τα κοινωνικά δίκτυα, ενώ οι εφαρμογές είναι πολλές και πολύ διαφορετικές μεταξύ τους, από την εξωστρεφή προώθηση προϊόντων μέχρι λ.χ. την ανάλυση της



Σχήμα 1.4: Ισοτιμία Bitcoin δολαρίου (BTC/USD) για διάρκεια ενός έτους

άποψης των καταναλωτών.

Γενικότερα και με πιο απλά λόγια, η ανάλυση συναισθήματος στοχεύει στον προσδιορισμό της στάσης ενός ανθρώπου στη βάση νοήματος που έχει εξαχθεί από κάποιο κείμενο ή γενικότερα από κάποιο γεγονός.

1.5 Κίνητρο

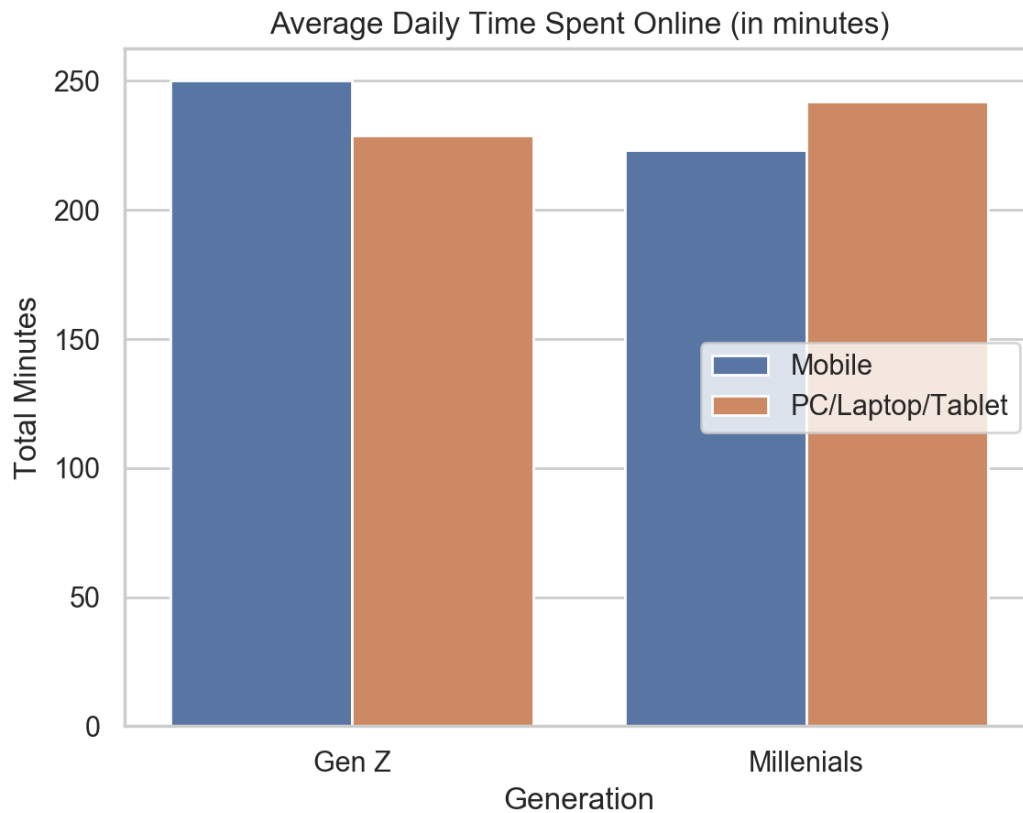
Στόχος της εργασίας είναι η πρόβλεψη της μελλοντικής τιμής του Bitcoin με βασικό εργαλείο τη μηχανική μάθηση. Ωστόσο η καινοτομία και η βασική ιδέα που προσφέρει αυτή η εργασία μπορεί να αποτυπωθεί με μία ερώτηση:

Μήπως τα αποτυπώματα που αφήνουμε στο διαδίκτυο δίνουν αρκετή πληροφορία για να προβλέψουμε την αγορά;

Σύμφωνα με την εταιρία GlobalWebIndex², η γενιά των millennials (ημερομηνία γέννησης στις αρχές του 1980 έως και μέσα του 1990 ή αρχές του 2000) και η επόμενη γενιά γνωστή ως Generation Z (μέσα του 1990 και αρχές του 2000 μέχρι και σήμερα χωρίς να έχει γίνει ξεκάθαρο το άνω φράγμα), με άλλα λόγια οι ηλικίες έως και 35 χρόνων, περνούν σημαντικά μεγάλο χρονικό διάστημα στο Διαδίκτυο, τόσο μέσω της χρήσης κινητού τηλεφώνου όσο και μέσω φορητού υπολογιστή ή tablet. Στο σχήμα 1.5 αναπαρίσταται το μέσο χρονικό

² <https://www.globalwebindex.com/>

διάστημα που αφιερώνει η κάθε γενιά στο Διαδίκτυο (σε λεπτά της ώρας).



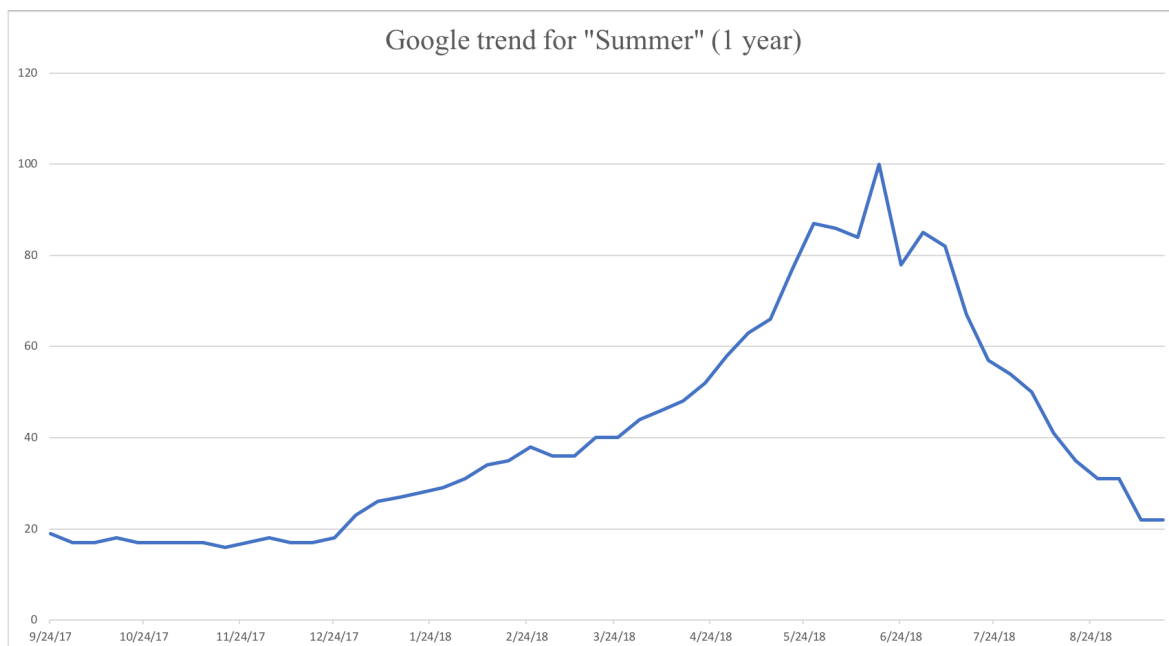
Σχήμα 1.5: Μέσος καθημερινός χρόνος στο διαδίκτυο για κάθε γενιά. (Πηγή: *GlobalWebIndex, Q2 2017, παγκόσμιο δείγμα, συμμετέχοντες: 43.592, ηλικίες: 16-34*)

1.5.1 Google Trends

Η αρχική ιδέα ήταν να μελετηθεί η κίνηση της πληροφορίας που αφορά το Bitcoin ήταν μέσω των *Google Trends*³. Τα *Google Trends* είναι ένας ιστότοπος της Google που αναλύει την δημοτικότητα των αναζητήσεων που γίνονται στην μηχανή αναζήτησης της Google, σε διάφορες περιοχές και γλώσσες. Στον ιστότοπο, ο χρήστης μπορεί να θέσει τον όρο για τον οποίο θα πάρει τα στατιστικά αναζήτησής του. Στο Σχήμα 1.6 παρουσιάζεται ένα παράδειγμα από τα *Google Trends* για συχνότητα αναζήτησης του όρου «Summer».

Στον οριζόντιο άξονα βρίσκονται οι ημερομηνίες ενώ στον κατακόρυφο άξονα είναι η συχνότητα αναζήτησης του όρου, κανονικοποιημένη στο διάστημα [0, 100]. Δηλαδή όταν η συχνότητα φτάνει στο 100 σημαίνει ότι τη συγκεκριμένη μέρα ο όρος είχε την μεγαλύτερη αναζήτηση. Αυτό το απλό εργαλείο παρουσιάζει μεγάλες δυνατότητες καθώς το «πόσο» και

³ <https://trends.google.com/trends/>



Σχήμα 1.6: Google Trend για συχνότητα αναζήτησης του όρου «Summer».

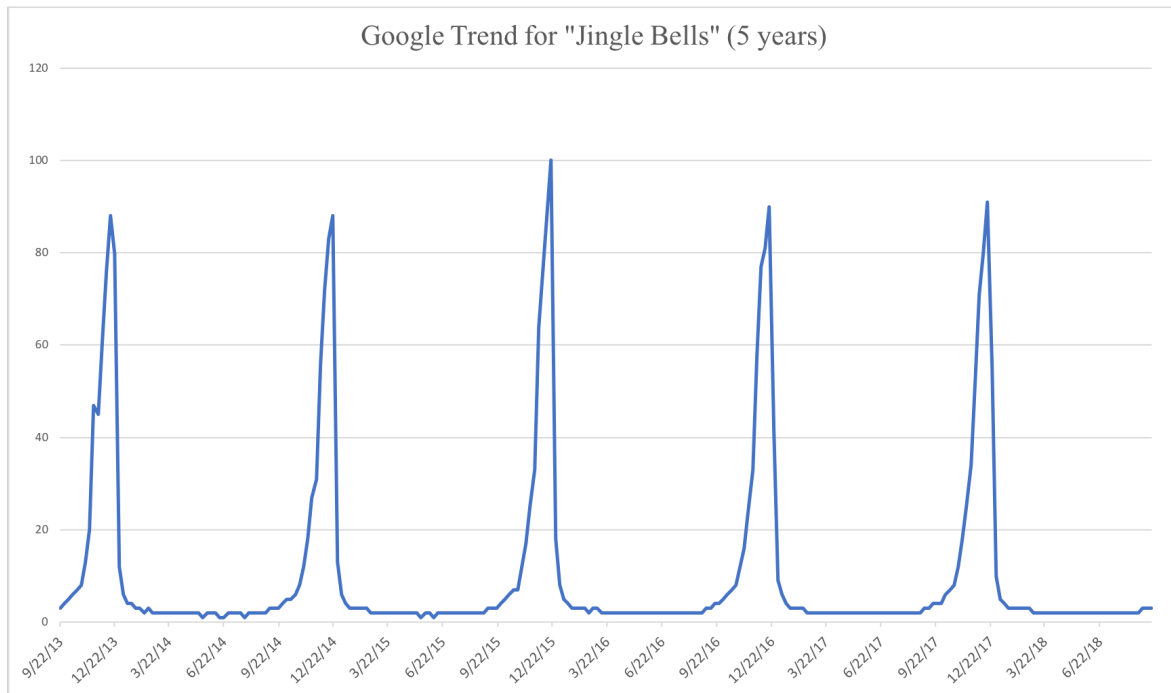
το «τι» ψάχνουν οι χρήστες μπορεί να δώσει αρκετή πληροφορία ώστε να προβλέψουμε και να υποθέσουμε καταστάσεις με μεγάλη επιτυχία. Για παράδειγμα στο Σχήμα 1.6 είναι φανερό ότι η αναζήτηση του όρου «Summer» κορυφώνεται στα μέσα Ιουνίου οπότε ακόμα και αν δεν ξέραμε πότε είναι το καλοκαίρι, το γράφημα το προδίδει. Κάποιες φορές μπορεί κανείς να δει περιοδικότητες στους όρους αναζήτησης. Στο Σχήμα 1.7 παρουσιάζεται το Google Trend για τον όρο αναζήτησης «Jingle Bells», το γνωστό χριστουγεννιάτικο τραγούδι.

Γενικότερα, μπορούμε να υποθέσουμε ότι οι απότομες κορυφές προδίδουν πότε το κοινό ενδιαφέρεται ιδιαίτερος να ψάξει για κάποιον όρο. Συνεπώς, κίνητρο της εργασίας αποτέλεσε η ιδέα ότι η συχνότητα αναζήτησης ενός όρου δείχνει το ενδιαφέρον της κοινής γνώμης για αυτόν και άρα έχει κάποια επίδραση στον όρο καθ' εαυτόν, θετική ή αρνητική. Στη περίπτωση της παρούσας διπλωματικής εργασίας μας ενδιαφέρει να ανακαλύψουμε αυτή ακριβώς την επίδραση. Στο Σχήμα 1.8 αποτυπώνεται το Google Trend του όρου «Bitcoin» για τρία χρόνια.

Όπως είναι φανερό, η αναζήτηση του Bitcoin κορυφώθηκε τον Δεκέμβρη του 2017 ενώ πριν δύο χρόνια το ενδιαφέρον για αυτό ήταν μηδαμινό. Έτσι το ερώτημα που τίθεται είναι κατά πόσο η μορφή του Google Trend «Bitcoin» συσχετίζεται με την τιμή του Bitcoin. Με σκοπό να απαντηθεί το παραπάνω ερώτημα, ακολουθεί, στο Σχήμα 1.9 η εξέλιξη της τιμής του Bitcoin σε δολάρια, για το ίδιο χρονικό διάστημα⁴

Η ομοιότητα που παρουσιάζουν τα δύο γραφήματα είναι πολύ μεγάλη. Αυτή η ομοιότητα αποτυπώνεται στο Σχήμα 1.10, καθώς έχουν συνδυαστεί και οι δύο γραφικές σε ένα διάγραμμα. Οι τιμές των δύο γραφημάτων έχουν κανονικοποιηθεί στο διάστημα [0, 1] έτσι

⁴ Πηγή: Πλατφόρμα Bitfinex (<https://www.bitfinex.com/>)



Σχήμα 1.7: Google Trend για συχνότητα αναζήτησης του όρου «Jingle Bells»

ώστε να είναι συγκρίσιμες.

Συνεπώς είναι βάσιμη η υπόθεση ότι τα Google Trends μπορούν να χρησιμοποιηθούν για την πρόβλεψη της τιμής του Bitcoin.

1.5.2 Μέσα Κοινωνικής Δικτύωσης

Με την δημοτικότητα που απέκτησε το Διαδίκτυο τις τελευταίες δεκαετίες, πολλές από τις νέες υπηρεσίες που δημιουργήθηκαν με βάση αυτό, έγιναν απαραίτητες τόσο για επαγγελματική όσο και για προσωπική χρήση. Μια από τις σχετικές υπηρεσίες είναι και τα μέσα κοινωνικής δικτύωσης.

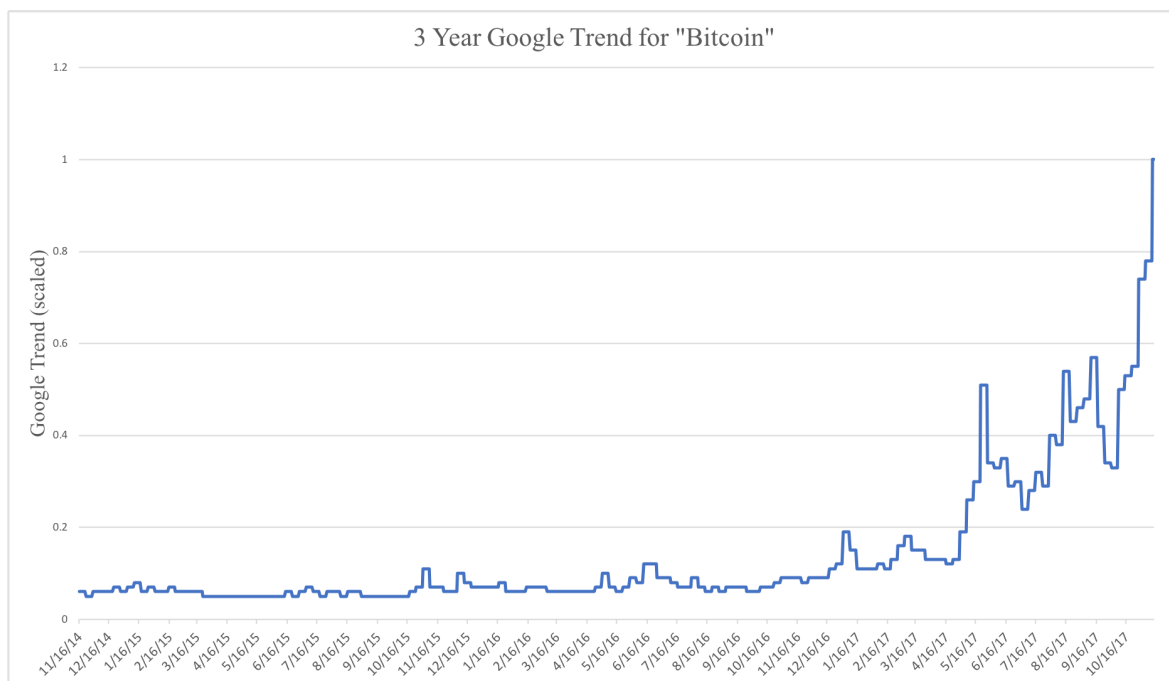
Όπως είναι γνωστό, τα μέσα κοινωνικής δικτύωσης είναι πλατφόρμες όπου οι χρήστες μπορούν να κοινοποιούν, να επικοινωνούν και να μοιράζονται υλικό. Ανάλογα με την πλατφόρμα αλλάζει και η φύση του υλικού. Για παράδειγμα το Facebook⁵ είναι μια πλατφόρμα κοινωνικής δικτύωσης φίλων ενώ το Twitter⁶ και το LinkedIn⁷ είναι σαφώς επαγγελματικά προσανατολισμένα. Το Instagram⁸ έχει έναν δυικό χαρακτήρα καθώς ξεκίνησε ως μια κοινωνική, με καλλιτεχνικό χαρακτήρα, πλατφόρμα, ωστόσο τα τελευταία χρόνια έχει σημαντικό επαγγελματικό χαρακτήρα με την παρουσία μιας νέας τάσης marketing, τους λεγόμενους «*influencers*». Για να μπορέσει κανείς να αντιληφθεί το μέγεθος της πληροφορίας που κατα-

⁵ <https://www.facebook.com/>

⁶ <https://www.twitter.com/>

⁷ <https://www.linkedin.com/>

⁸ <https://www.instagram.com/>



Σχήμα 1.8: Google Trend για συχνότητα αναζήτησης του όρου «Bitcoin»

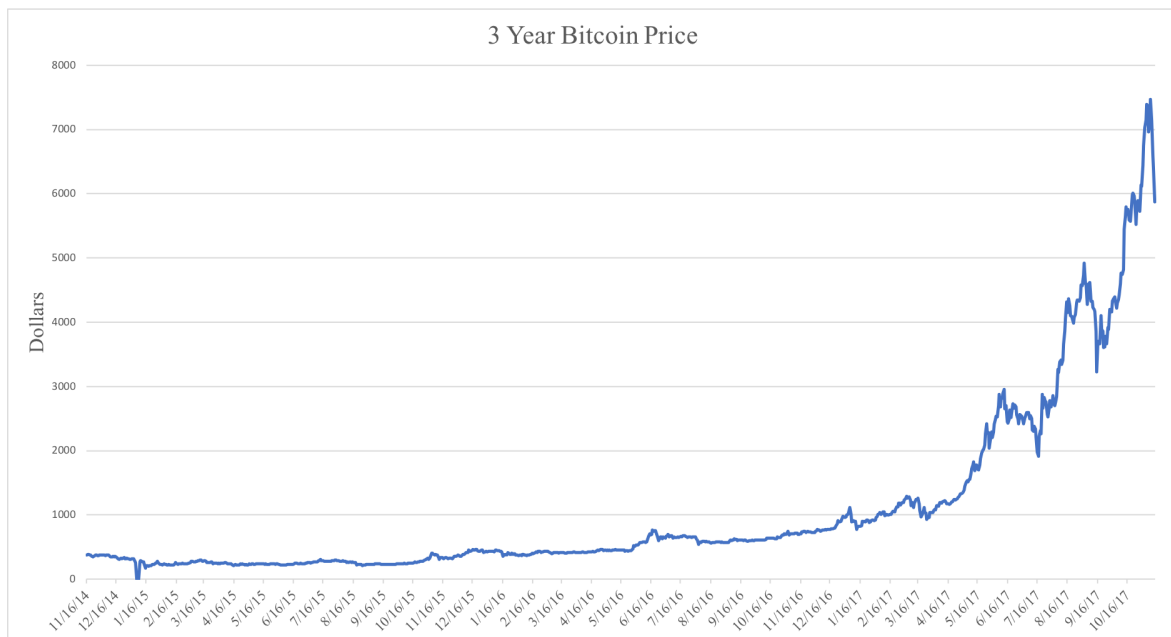
γράφεται και διακινείται μέσω των κοινωνικών δικτύων στο Σχήμα 1.11 φαίνονται οι ενεργοί χρήστες για το 2018⁹.

Το γεγονός ότι το Facebook έχει 2.196.000.000 ενεργούς χρήστες, σημαίνει ότι περίπου το 29%, ή αλλιώς το ένα τρίτο του πλανήτη, είναι ενεργός χρήστης του Facebook. Επίσης το Instagram έχει 1 δισεκατομμύριο ενεργούς χρήστες, δηλαδή για κάθε δύο χρήστες του Facebook έχουμε έναν στο Instagram. Είναι προφανές λοιπόν ότι η πληροφορία που διακινείται είναι τεράστια και άρα είναι λογικό να υποθέσουμε πως υπάρχει αρκετή πληροφορία για να αναλύσουμε πως επηρεάζεται η αγορά κρυπτονομισμάτων.

1.5.3 Twitter

Το Twitter είναι ένα μέσω κοινωνικής δικτύωσης που επιτρέπει στους χρήστες του να στέλνουν και να διαβάζουν σύντομα μηνύματα (μέχρι 280 χαρακτήρες), τα οποία ονομάζονται tweets. Ο όρος tweet προέρχεται από το χαρακτηριστικό «τιτίβισμα» που κάνει ένα πουλί, σαν σύντομο μήνυμα. Τα μηνύματα μπορούν να αναγνωστούν και από μη συνδεδεμένους χρήστες, αλλά μόνο οι συνδεδεμένοι μπορούν να δημοσιεύσουν κείμενα. Η εταιρία ιδρύθηκε το 2006 και έγινε γρήγορα δημοφιλής και σήμερα έχει 334 εκατομμύρια ενεργούς χρήστες, αποτελώντας έναν από τους δέκα πιο συχνά προσπελάσιμους ιστοτόπους του Διαδικτύου. Βασικό εργαλείο στα tweets είναι να χρησιμοποιεί κανείς λέξεις-κλειδιά γνωστές και ως *hashtags*.

⁹ Πηγή: Statista (<https://www.statista.com/>)



Σχήμα 1.9: Η εξέλιξη της τιμής του Bitcoin (σε δολάρια) σε διάστημα τριών χρόνων.

1.5.4 Hashtag

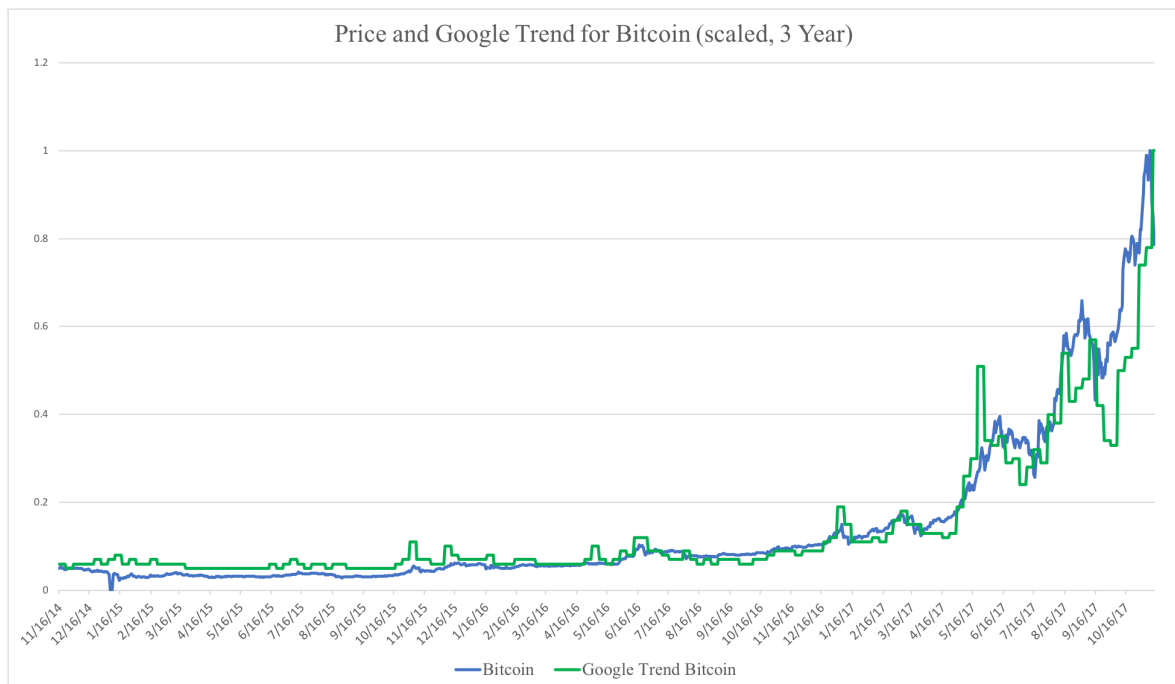
Η λέξη Hashtag προέρχεται από τον συνδυασμό των λέξεων «hash» και «tag», δηλαδή μια λέξη που ακολουθείται από το σύμβολο #. Η λέξη «hash» περιγράφει το σύμβολο # που εμείς ονομάζουμε δίσωση, ενώ η λέξη «tag» περιγράφει τον όρο ετικέτα, δηλαδή έναν όρο. Παράδειγμα #Whiskey

Η χρήση του Hashtag είναι πολύ σημαντική καθώς, οποιοσδήποτε χρήστης μπορεί να πληκτρολογήσει μέσα σε μια δημοσίευση του (ή σχόλιο) ένα ή περισσότερα. Αυτά στη συνέχεια μετατρέπονται αυτόματα σε συνδέσμους και εντάσσουν τη δημοσίευση ανάμεσα σε άλλες δημοσιεύσεις που περιέχουν τουλάχιστον ένα όμοιο hashtag. Με αυτόν τον τρόπο αλλάζει η κλίμακα του κοινού που ένας χρήστης μπορεί να προσεγγίσει. Έτσι γεννιέται το ερώτημα για το αν τα αποτυπώματα που αφήνουν οι χρήστες των μέσων κοινωνικής δικτύωσης περιέχουν αρκετή πληροφορία για να προβλεφθούν γεγονότα και καταστάσεις.

Αφού το ενδιαφέρον της εργασίας είναι η πρόβλεψη της τιμής του Bitcoin, θα εστιάσουμε στο hashtag που αφορά το Bitcoin, και το βασικότερο όλων είναι το #Bitcoin. Ακολουθεί το γράφημα που δείχνει τον αριθμό των Tweets ημερησίως σε χιλιάδες, που συμπεριλαμβάνουν το #Bitcoin.

Αν και κάπως πιο «θορυβώδης», επίσης η τάση της κυματομορφής του σχήματος 1.12 μοιάζει πολύ με την κυματομορφή της τιμής του Bitcoin στο Σχήμα 1.9.

Εν κατακλείδι, τα παραπάνω στοιχεία ήταν αυτά που έδωσαν το βασικό κίνητρο να γίνει αυτή η εργασία. Τα Google Trends και το Twitter έδωσαν μια οπτική αναπαράσταση της πληροφορίας. Εκτός όμως από την προσφορά που θα μπορούσαν να δώσουν αυτοί οι δύο



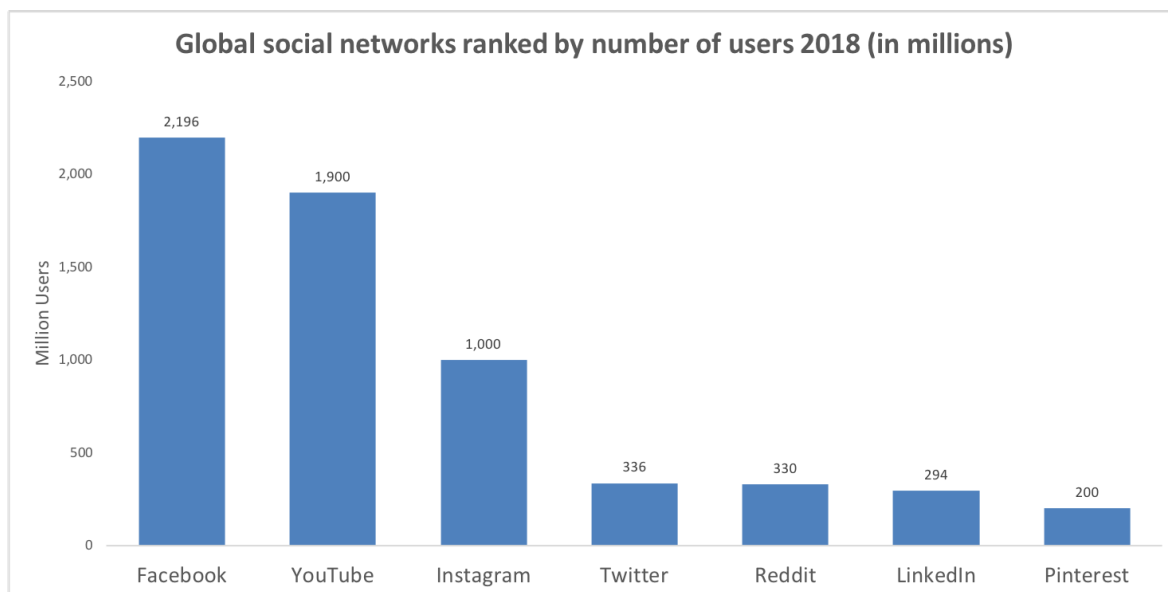
Σχήμα 1.10: Η ομοιότητα μεταξύ του Google Trend «Bitcoin» και τιμής του Bitcoin

δείκτες στην πρόβλεψη της τιμής, η παραπάνω μελέτη οδήγησε στην ιδέα να συμπεριληφθεί επίσης και η ανάλυση συναισθήματος της ειδησεογραφίας της ημέρας, κάτι που θα αναλυθεί στο Κεφάλαιο 3.

Παραδείγματα και προκλήσεις

Η εφαρμογή των τεχνικών ανάλυσης συναισθήματος είναι μεγάλη και τα πεδία, όπως αναφέρθηκε παραπάνω, είναι πολύ διαφορετικά. Ενδεικτικά ακολουθούν μερικά πραγματικά παραδείγματα:

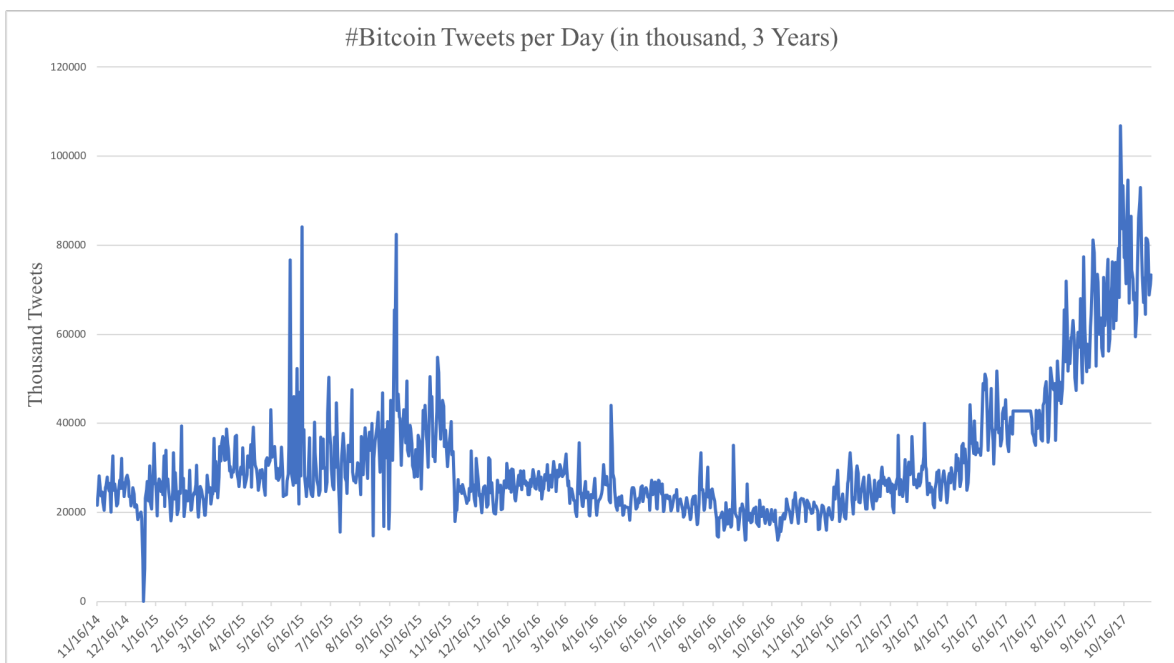
- Στο Marketing, η ανάλυση συναισθήματος χρησιμοποιείται για να μελετηθούν οι κριτικές που κάνει το κοινό για διάφορα προϊόντα, ώστε να μπορούν να τα βελτιώσουν και να τα προωθήσουν καλύτερα.
- Στην πολιτική, για να εκτιμηθεί η κλήση του κοινού προς συγκεκριμένες καταστάσεις, για παράδειγμα θετική, χαρούμενη, θυμωμένη.
- Στον χρηματοοικονομικό τομέα, προσπαθούν να βρουν την άποψη για μία φίρμα ώστε να προβλέπουν τις κινήσεις της αγοράς.
- Στην πρόληψη κινδύνου, εντοπίζουν αν υπάρχουν άνθρωποι δέχονται επίθεση, παράδειγμα στα μέσα κοινωνικής δικτύωσης.



Σχήμα 1.11: Ενεργοί χρήστες (σε εκατομμύρια) μέχρι τον Ιούλιο 2018, για μερικά από τα μεγαλύτερα παγκόσμια μέσα κοινωνικής δικτύωσης (Πηγή: Statista)

1.6 Δομή της Διπλωματικής Εργασίας

Αφού έγινε μια σύντομη εισαγωγή με σκοπό να εξηγηθούν οι βασικοί όροι και έννοιες καθώς και το κίνητρο που οδήγησε σε αυτή τη διπλωματική, ακολουθεί μια σύντομη περιγραφή της εργασίας και των Κεφαλαίων που έπονται. Αρχικά, στο Κεφάλαιο 2, περιγράφονται τα νευρωνικά δίκτυα και συγκεκριμένα τα αναδρομικά νευρωνικά δίκτυα, τα οποία και επιλέχθηκαν. Επίσης σε αυτή την ενότητα παρουσιάζονται και αιτιολογούνται τα μοντέλα μηχανικής μάθησης που θα χρησιμοποιηθούν. Εν συνεχεία, στο Κεφάλαιο 3 παρουσιάζεται αναλυτικά η διαδικασία συλλογής των δεδομένων. Στην εργασία αυτή θα περιληφθούν διάφορων τύπων δεδομένα από χρηματιστηριακού τύπου μέχρι tweets και τίτλους ειδήσεων. Ακολουθεί το Κεφάλαιο 4, όπου παρουσιάζονται και εξηγούνται τα πειράματα και τα αποτελέσματα της εργασίας. Σε αυτό το κεφάλαιο γίνεται παρουσίαση και σύγκριση των αποτελεσμάτων για όλα τα μοντέλα που εξετάστηκαν. Τέλος, η εργασία ολοκληρώνεται στο Κεφάλαιο 5, όπου πραγματοποιείται η αποτίμησή της ενώ παρουσιάζονται και ιδέες για μελλοντικές επεκτάσεις και έρευνα.



Σχήμα 1.12: Αριθμός των Tweets ημερησίως σε χιλιάδες, που συμπεριλαμβάνουν το #Bitcoin.

Κεφάλαιο 2

Θεωρητικό Υπόβαθρο

Στο Κεφάλαιο πραγματοποιείται ανάλυση του θεωρητικού υπόβαθρου της διπλωματικής εργασίας και των τεχνικών που θα χρησιμοποιηθούν. Στην πρώτη ενότητα θα γίνει αναφορά στο Bitcoin και τα κρυπτονομίσματα. Εν συνεχεία θα μελετηθεί η αγορά ξένου συναλλάγματος (**foreign exchange - forex**) με σκοπό να γίνει η σύγκρισή της με την αγορά κρυπτονομισμάτων ώστε να φανούν οι ομοιότητες και οι διαφορές μεταξύ τους. Στην δεύτερη ενότητα θα αναλυθεί θεωρητικά η λειτουργία και η εφαρμογή των νευρωνικών δικτύων ξεκινώντας με την περιγραφή του απλού νευρωνικού δικτύου, και στην πορεία θα μελετηθούν τα αναδρομικά νευρωνικά δίκτυα εστιάζοντας περισσότερο στα δίκτυα Elman και Jordan. Κλείνοντας, θα γίνει αναφορά και αιτιολόγηση για το ποια δίκτυα και βιβλιοθήκες επιλέχθηκαν για αυτήν την εργασία.

2.1 Bitcoin

2.1.1 Εισαγωγή

Το Bitcoin [Naka09] είναι μία peer-to-peer αποκεντρωμένη ηλεκτρονική μορφή χρήματος, η οποία βασίζεται πάνω στις αρχές της κρυπτογραφίας για την διασφάλιση του δικτύου και την επαλήθευση των συναλλαγών. Με απλά λόγια, είναι μια μορφή *ηλεκτρονικών μετρητών* (digital cash) όπου η ασφάλεια των συναλλαγών δεν γίνεται από έναν έμπιστο μεσάζοντα (πχ τράπεζα) αλλά από μεθόδους κρυπτογραφίας στις οποίες συμμετέχουν οι ίδιοι οι χρήστες του δικτύου. Άρα οι χρήστες του δικτύου εν δυνάμει μπορούν να επιβεβαιώσουν μια συναλλαγή π.χ. μεταφοράς bitcoin από τον έναν χρήστη στον άλλον χωρίς την ανάγκη κάποιου είδους οργανισμού, όπως μια τράπεζα.

Λόγω της ανοικτής φύσης του λογισμικού του, επετράπη σε πολλούς προγραμματιστές να πειραματιστούν με τον κώδικά του και να τον τροποποιήσουν με αποτέλεσμα την δημιουργία μιας πληθώρας νέων κρυπτονομισμάτων, στα οποία έχουν γίνει προσπάθειες για να βελτιωθούν ή και να προστεθούν λειτουργίες όπως ταχύτερες συναλλαγές, μεγαλύτερη ανωνυμία και άλλα. Πλέον υπάρχουν εταιρείες, υπηρεσίες και αγαθά τα οποία μπορεί να αγοράσει κανείς με bitcoin, καθώς και δυνατότητα ανταλλαγής του με συμβατικό χρήμα.

2.1.2 Παραγωγή του Bitcoin (mining)

Τα Bitcoin δημιουργούνται από μια ανταγωνιστική και αποκεντρωμένη διαδικασία που ονομάζεται «εξόρυξη» ή bitcoin mining. Οι «μεταλλωρύχοι» (miners) επεξεργάζονται συναλλαγές και ασφαλίζουν το δίκτυο χρησιμοποιώντας εξειδικευμένο υλικό και λογισμικό σε αντάλλαγμα συλλέγουν νέα bitcoin. Οι miners είναι οι χρήστες που διασφαλίζουν τις συναλλαγές μέσω των δύσκολων μαθηματικών και της κρυπτογραφίας. Η διαδικασία της εξόρυξης στην ουσία της μοιάζει με ένα είδος λοταρίας όπου οι bitcoin miners επενδύοντας «φυσικά» χρήματα ανταγωνίζονται όλους αυτούς που κάνουν την ίδια δουλειά με σκοπό να βρουν πρώτοι το υπόλοιπο από τα διαθέσιμα 21 εκατομμύρια «ψηφιακά» bitcoin. Αυτή την στιγμή υπολογίζεται χοντρικά ότι έχουν βρεθεί τα 16 εκατομμύρια και ότι μέχρι το 2033 θα έχει βρεθεί και το τελευταίο. Στην διπλωματική αυτή θα γίνεται αναφορά στο Bitcoin σαν ένα απλό νόμισμα.

2.1.3 Αγορά ξένου συναλλάγματος

Πριν αναλυθεί η αγορά κρυπτονομισμάτων, θα γίνει μια σύντομη περιγραφή στην αγορά ξένου συναλλάγματος, ως δημοφιλέστερη και παρόμοιας φύσης αγορά. Το forex είναι μια παγκόσμια χρηματοοικονομική αγορά στην οποία συμμετέχουν όλα τα χρηματοοικονομικά κέντρα του κόσμου και στην οποία γίνεται η αγοραπωλησία όλων των εθνικών νομισμάτων. Η αγοραπωλησία αυτή μπορεί να γίνει με δύο τρόπους, την χονδρική αγορά και την λιανική αγορά. Στην περίπτωση της χονδρικής αγοράς ο όγκος των συναλλαγών είναι πολύ μεγάλος και οι διαφορές τιμής αγοράς-πώλησης εξαιρετικά μικρές έως μηδενικές. Παράδειγμα τέτοιας συναλλαγής μπορεί να λάβει χώρα μεταξύ μιας τράπεζας και ενός άλλου οργανισμού ενώ είναι δυνατό να έχει ύψος δισεκατομμυρίων δολαρίων ακόμα και σε μία ημέρα. Στην περίπτωση της λιανικής αγοράς (retail market) μπορούν και συμμετέχουν ιδιώτες - μέσω τραπεζών, διαπραγματευτών λιανικής (retail traders) και μεσιτών (brokers) - οι οποίοι αγοράζουν και πωλούν συνάλλαγμα σε μια καθορισμένη τιμή, με σκοπό την κερδοσκοπία. Σε ένα τέτοιο «παιχνίδι» είναι προφανές ότι η πρόβλεψη της τιμής που θα έχει ένα συνάλλαγμα στο μέλλον είναι υψίστης σημασίας καθώς επιτρέπει στον ιδιώτη να κάνει βέλτιστες κινήσεις αγοράς και πώλησης.

Αγορά κρυπτονομισμάτων

Όπως με τις αγορές μετοχών (stock market), τις αγορές ξένων νομισμάτων έτσι με την δημιουργία των κρυπτονομισμάτων εμφανίστηκε και η αγορά κρυπτονομισμάτων (cryptocurrency market). Σε τέτοιες αγορές μπορεί κανείς να αγοράσει και να πουλήσει κρυπτονομίσματα με σκοπό να βγάλει κέρδος από την διαφορά (διακύμανση) της τιμής τους.

Η αγορά κρυπτονομισμάτων παρουσιάζει μερικές διαφορές σε σχέση με τις αγορές ξένων νομισμάτων. Αυτό δημιουργεί μερικά προβλήματα και δοκιμασίες, αλλά ταυτόχρονα διαφοροποιεί την φύση της διπλωματικής εργασίας από άλλες που αφορούν το forex. Οι κυριότερες διαφορές μεταξύ του forex και των αγορών κρυπτονομισμάτων παρουσιάζονται στην συνέχεια.

Άνοιγμα λογαριασμού

Για να μπορέσει να συμμετάσχει στην αγορά κρυπτονομισμάτων αρκεί να ανοίξει λογαριασμό σε κάποια πλατφόρμα (Bitfinex, Coinbase, KoineX κ.α.), στην συνέχεια να αγοράσει μερικά κρυπτονομίσματα ή *υποδιαίρεσεις* (tokens) και να ξεκινήσει κατευθείαν. Στο forex η διαδικασία έναρξης είναι πολύ μεγαλύτερη καθώς χρειάζεται να χρησιμοποιήσει κανείς ένα μεσίτη, είτε αυτό είναι κάποιο φυσικό πρόσωπο είτε μια πλατφόρμα, όπου χρειάζεται να συμπληρωθούν και κατατεθούν πολλά πιστοποιητικά και έγγραφα.

Μεταβλητότητα

Η αγορά κρυπτονομισμάτων παρουσιάζει μεγάλη *μεταβλητότητα* (volatility). Αυτό την κάνει μια επένδυση υψηλού ρίσκου. Ως τέτοια, είναι δυνατόν κανείς σε μια μέρα να δει εξαιρετικά κέρδη και αντίστοιχα να έχει εξαιρετικές απώλειες στον ίδιο ή και λιγότερο χρόνο. Το forex από την άλλη μεριά είναι πολύ πιο στιβαρή αγορά. Αυτό συμβαίνει γιατί οι παράγοντες που επηρεάζουν forex μπορούν να προκαθοριστούν. Με την μελέτη διάφορων τεχνικών παραγόντων, καλή γνώση τεχνικής ανάλυσης (η μέθοδος μελέτης διαγραμμάτων) και ακολουθώντας εργαλεία δείκτες και στρατηγικές μπορεί κανείς να έχει μεγάλες πιθανότητες επιτυχίας στην αγορά συναλλάγματος.

Κανονισμός Λειτουργίας

Οι περισσότεροι μεσίτες forex ρυθμίζονται από τις εκάστοτε χρηματοοικονομικούς κανονισμούς κάτι που καθιστά την αγορά συναλλάγματος ένα αρκετά ασφαλές τρόπο επένδυσης, καθώς είναι αρκετά δύσκολο και σχετικά απίθανο να κλαπούν ή χαθούν τα κεφάλαια του χρήστη. Η αγορά κρυπτονομισμάτων δεν ελέγχεται από καμία οντότητα και η αξία τους καθορίζεται κάθε φορά από τους ιδιοκτήτες των ίδιων των κρυπτονομισμάτων. Ως εκ τούτου δεν υπάρχει κάποια διασφάλιση των κεφαλαίων του χρήστη, με αποτέλεσμα να είναι πιθανό να υπάρχουν απώλειες, για παράδειγμα σε περίπτωση hacking, που δεν μπορούν με κανένα τρόπο να ασφαλιστούν.

Ωράριο Αγοράς

Η αγορά forex είναι ανοιχτή πέντε μέρες την εβδομάδα ενώ τα σαββατοκύριακα παραμένει κλειστή, σε αντίθεση με τα την αγορά κρυπτονομισμάτων που είναι ανοιχτή όλο το εικοσιτετράωρο, κάθε μέρα του χρόνου.

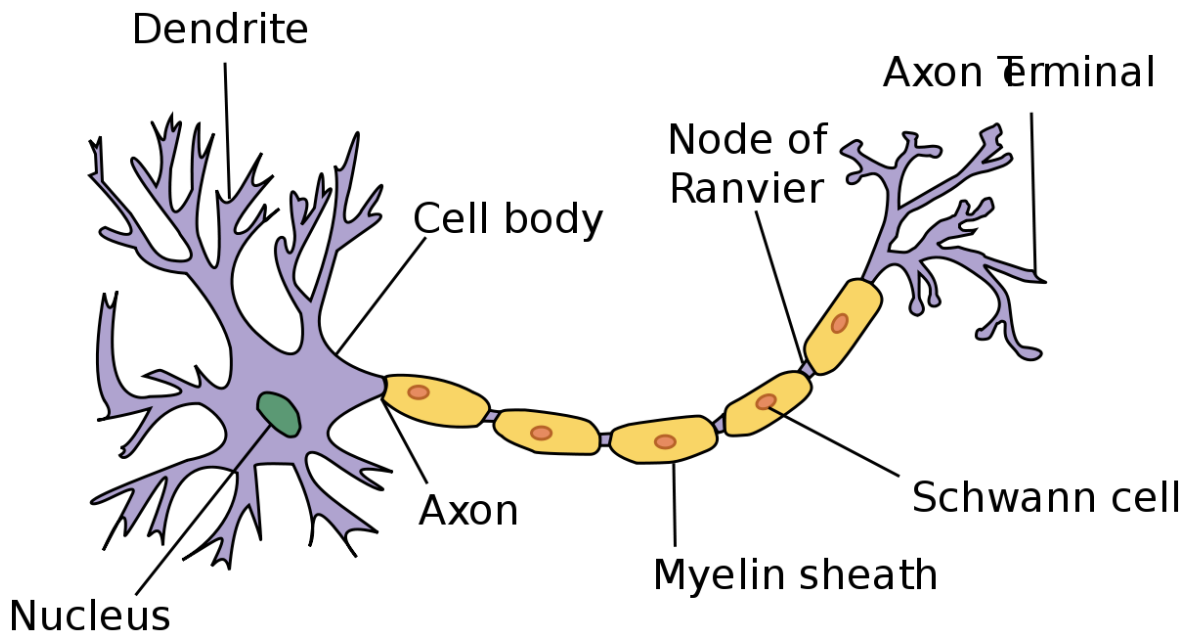
Συμπερασματικά, η αγορά κρυπτονομισμάτων μπορεί να αποφέρει κέρδος αλλά ταυτόχρονα είναι μεγαλύτερου ρίσκου.

2.2 Νευρωνικά Δίκτυα

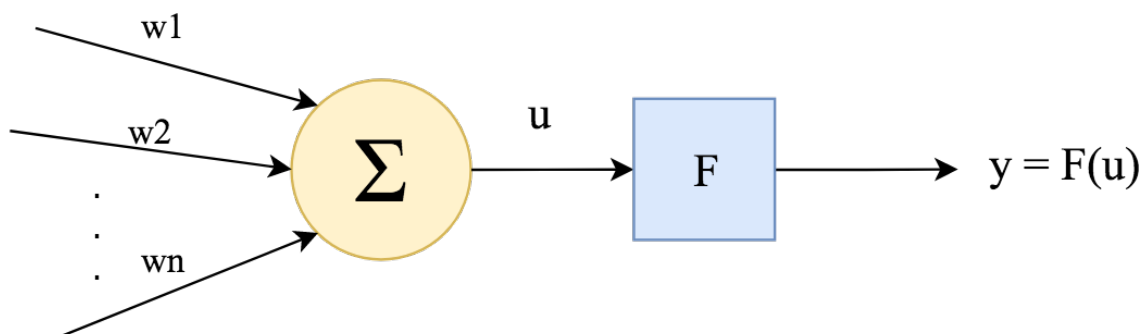
Τα τεχνητά νευρωνικά δίκτυα ή απλώς *νευρωνικά δίκτυα* (Artificial Neural Networks) είναι ένα αφηρημένο αλγοριθμικό κατασκεύασμα, το οποίο ανήκει στην οικογένεια της μηχανικής μάθησης, είναι ικανό να «μαθαίνει» να επιτελεί συγκεκριμένες λειτουργίες με το να διαβάζει παραδείγματα αυτών, και χωρίς να είναι ειδικά προγραμματισμένο. Η δομή του συστήματος έχει εμπνευστεί από τους νευρώνες του εγκεφάλου και γι' αυτό πήρε τη συγκεκριμένη ονομασία. Με τα τεχνητά νευρωνικά δίκτυα η επιστημονική κοινότητα προσπαθεί να μιμηθεί και να μοντελοποιήσει τη μη γραμμική αυτή συμπεριφορά του εγκεφάλου που σχετίζεται άμεσα με τη λήψη των αποφάσεων. Οι μέθοδοι που χρησιμοποιούνται σε αυτά τα δίκτυα παρέχουν ένα απελευθερωμένο μοντέλο, το οποίο μπορεί να προσαρμοστεί σε ένα μεγάλο εύρος εφαρμογών, έχοντας ανοχή στα σφάλματα και την ικανότητα να χρησιμοποιεί παράλληλα και διανεμημένα συστήματα για την ταχύτερη εξαγωγή της λύσης του προβλήματος. Πρόκειται στην ουσία για «μαύρα κουτιά» τα οποία «μαθαίνουν» την εσωτερική συσχέτιση ενός αγνώστου συστήματος χωρίς να προσπαθούν να «μαντέψουν» τις συναρτήσεις που διέπουν τις διασυνδέσεις του εν λόγω συστήματος. Χρησιμοποιούνται ευρέως σε θέματα όπως η προσέγγιση μαθηματικών συναρτήσεων, προβλήματα κατάταξης δεδομένων σε κλάσεις, εξαγωγής χαρακτηριστικών από μοντέλα, συσταδοποίησης, κβαντοποίησης διανυσμάτων, βελτιστοποίησης και αλλά. Αποτελούν μια από τις βασικές μεθοδολογίες στην επεξεργασία των πληροφοριών και βρίσκουν εφαρμογή σχεδόν σε ούλες τις περιοχές της επιστήμης και της μηχανικής.

Μέσα από τα παραδείγματα θα γίνει αναλυτικότερη περιγραφή των μαθηματικών και της αρχιτεκτονικής που διέπουν τα νευρωνικά δίκτυα. Όπως ειπώθηκε παραπάνω ο όρος «νευρωνικά δίκτυα» προέρχεται από τους νευρώνες του εγκεφάλου, καθώς οι αλγόριθμοι αυτοί «μιμούνται» κατά κάποιο τρόπο τις λειτουργίες τους. Ένα ιατρικό μοντέλο ενός νευρώνα του εγκεφάλου φαίνεται στο Σχήμα 2.1. Η βασική ιδέα των νευρώνων είναι ότι μεταδίδουν τα μηνύματα με τους γειτονικούς τους νευρώνες έχουν τοπικό έλεγχο είναι πολλοί σε αριθμό (περίπου 10^{11}) και με ακόμα περισσότερες συνάψεις (περίπου 10^{15}).

Στο Σχήμα 2.1 φαίνεται ότι ο νευρώνας έχει τον πυρήνα και τις συνάψεις. Δηλαδή η πληροφορία που «τρέχει» στο νευρικό σύστημα, περνάει από κάθε νευρώνα, όπου ο πυρήνας



Σχήμα 2.1: Ιατρικό μοντέλο ενός νευρώνα

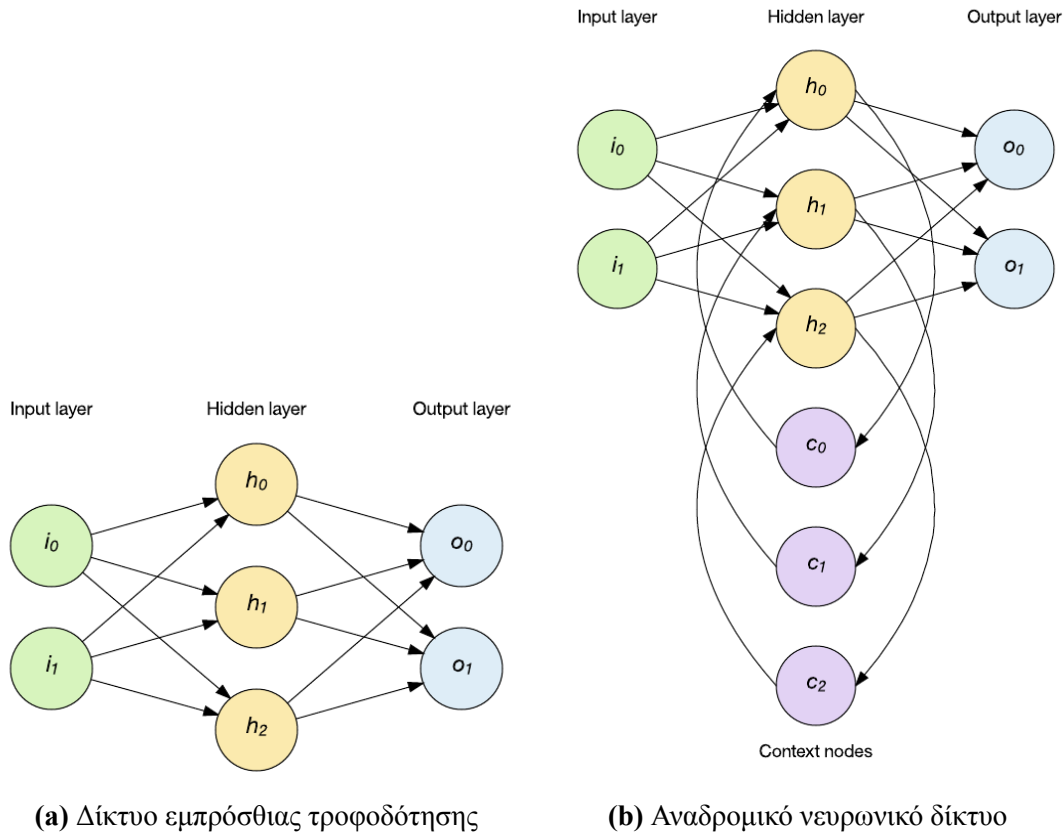


Σχήμα 2.2: Μαθηματικό μοντέλο ενός νευρώνα του νευρωνικού δικτύου.

αποφασίζει αν θα περάσει το σήμα ή όχι στους επόμενους. Έτσι μεταφέρονται τα μηνύματα και τελικά ο εγκέφαλος παίρνει τις αποφάσεις που παίρνει. Σε αυτή την ιδέα βασίστηκε και ο αλγόριθμος των νευρωνικών δικτύων. Στο Σχήμα 2.2 βλέπουμε το μαθηματικό μοντέλο ενός νευρώνα σε ένα νευρωνικό δίκτυο.

Τα νευρωνικά δίκτυα διακρίνονται με βάση την αρχιτεκτονική τους, τα χαρακτηριστικά των κόμβων τους (nodes) και τους κανόνες εκμάθησης που ακολουθούν. Αντιπροσωπευτικό της αρχιτεκτονικής τους είναι ο πίνακας $W = [w_{ij}]$ που περιέχει τις τιμές των βαρών που συνδέουν τον κόμβο i με τον κόμβο j , για όλους τους διασυνδεδεμένους μεταξύ τους κόμβους (Σχήμα 2.2). Εάν βάρος με τιμή $w_{ij} = 0$ σημαίνει ότι δεν υπάρχει κάποια σύνδεση μεταξύ των κόμβων i και j και μπορεί να χρησιμοποιηθεί για τη δημιουργία νέων τοπολογιών στο δίκτυο. Ο βασικός διαχωρισμός που μπορεί να γίνει στα νευρωνικά δίκτυα είναι σε εμπρόσθια τροφοδότησης (feedforward - Σχήμα 2.3a), αναδρομικά νευρωνικά δίκτυα (recurrent neural networks - Σχήμα 2.3b), καθώς και σε υβριδικά αυτών των δυο. Άλλες δημοφιλείς τοπολο-

γίες των νευρωνικών δικτύων είναι τα *δίκτυα πλέγματος* (lattice networks), τα *πολυεπίπεδα δίκτυα εμπρόσθιας τροφοδότησης με πλευρικές διασυνδέσεις* (layered feedforward networks with lateral connections) και τα *κυψελοειδή δίκτυα* (cellular networks).



Σχήμα 2.3: Δύο βασικές κατηγορίες νευρωνικών δικτύων.

- Τα δίκτυα πρόσθιας τροφοδότησης αποτελούνται από ένα ή περισσότερα *επίπεδα* (layers) και οι συνδέσεις μεταξύ των νευρώνων είναι προς μια μονό κατεύθυνση. Τα επίπεδα αποτελούνται από αυτό της εισόδου, τα εσωτερικά ή αλλιώς κρυφά επίπεδα και το επίπεδο εξόδου. Σε αυτού του είδους δίκτυα δεν υπάρχει καμιά σύνδεση μεταξύ των νευρώνων που βρίσκονται στο ίδιο επίπεδο καθώς και καμιά ανάδραση μεταξύ των επιπέδων. Σε ένα πλήρως διασυνδεδεμένο δίκτυο πρόσθιας τροφοδότησης, όπως αυτό του Σχήματος 2.3a, ο κάθε κόμβος ενός επιπέδου συνδέεται με όλους τους κόμβους στο επόμενο επίπεδο.
- Στα αναδρομικά νευρωνικά δίκτυα (Σχήμα 2.3b) υπάρχει τουλάχιστον μια ανάδραση, είτε ανάμεσα σε κόμβους του ίδιου επιπέδου, είτε σε κάποιο κόμβο με κόμβο ενός προηγούμενου επιπέδου. Αυτά τα δίκτυα θα μελετηθούν διεξοδικότερα στη συνέχεια

Η λειτουργία των νευρωνικών δικτύων χωρίζεται σε δυο στάδια, αυτά της *εκμάθησης* (learning) και της *ανάκλησης* (recalling). Η διαδικασία της εκμάθησης του συστήματος γίνε-

ται συνήθως με τη χρησιμοποίηση δειγμάτων που εμπίπτουν στο γενικό κανόνα πάνω στον οποίο θέλουμε να γίνει η εκπαίδευση. Με αυτό τον τρόπο οι παράμετροι του δικτύου (ο πίνακας βαρών W) προσαρμόζονται κατάλληλα με τη χρήση κάποιων κανόνων εκμάθησης. Η εκπαίδευση περιλαμβάνει την πλήρη σάρωση των δεδομένων για κάποιο πεπερασμένο αριθμό βημάτων, τα οποία ονομάζονται *εποχές* (epochs) και ο έλεγχος γίνεται με τη χρήση κάποιων κριτηρίων. Μόλις το δίκτυο φτάσει να έχει την επιθυμητή απόδοση ή έχει ολοκληρωθεί ένας αριθμός από εποχές που έχει οριστεί κατά την αρχή της εκπαίδευσης, το στάδιο της εκμάθησης τερματίζεται και το νευρωνικό δίκτυο μπορεί να χρησιμοποιηθεί πλέον για τη διαδικασία της ανάκλησης. Σε αυτό το δεύτερο και τελευταίο στάδιο το νευρωνικό δίκτυο είναι σε θέση να αντικαταστήσει πλήρως το σύστημα στη δυναμική του οποίου γινόταν η εκμάθηση και να χρησιμοποιηθεί πάνω σε νέα δείγματα.

Οι μέθοδοι εκμάθησης χωρίζονται σε **επιβλεπόμενη μάθηση** (supervised), **μη επιβλεπόμενη** (unsupervised) και **ενισχυτική** (reinforcement learning) μάθηση. Από την πλευρά της στατιστικής, η μη επιβλεπόμενη μάθηση μαθαίνει την συνάρτηση πυκνότητας πιθανότητας του συνόλου εκμάθησης, ενώ η επιβλεπόμενη μάθηση την συνάρτηση πυκνότητας πιθανότητας του $p(y|x)$. Με απλά λόγια, στην επιβλεπόμενη μάθηση τα δεδομένα που δίνουμε στο σύστημα για εκπαίδευση είναι ταξινομημένα, και χρησιμοποιείται ευρέως σε προβλήματα κατάταξης σε κλάσεις, προσέγγισης, ελέγχου, αναγνώρισης και μοντελοποίησης, επεξεργασίας σημάτων και βελτιστοποίησης. Στη μη επιβλεπόμενη μάθηση τα δεδομένα είναι αταξινομήτα, ενώ καλείται ο αλγόριθμος να βρει την σχέση που έχουν μεταξύ τους. Είναι, δηλαδή, μια τεχνική που αποσκοπεί αποκλειστικά στην εύρεση συσχέτισης ανάμεσα στα δεδομένα εισόδου και χρησιμοποιείται για να εντοπίσει τα σημαντικότερα μοτίβα ή χαρακτηριστικά αυτής, χωρίς τη βοήθεια που θα προσέφεραν τα δεδομένα εξόδου. Χρησιμοποιείται σε θέματα όπως η *ομαδοποίηση* (clustering), η κβαντοποίηση διανυσμάτων, η εξαγωγή χαρακτηριστικών, η κωδικοποίηση σήματος και η ανάλυση δεδομένων. Τέλος, η *ενισχυόμενη μάθηση* (reinforcement learning) βρίσκει εφαρμογές σε τομείς που αφορούν μηχανισμούς ελέγχου και τεχνητής νοημοσύνης.

2.3 Αναδρομικά Νευρωνικά Δίκτυα

2.3.1 Εισαγωγή

Αυτή η ενότητα ασχολείται με τα αναδρομικά νευρωνικά δίκτυα και περιγράφει τις αρχές και τη λειτουργία τους, τις αρχιτεκτονικές διάφορων αναδρομικών μοντέλων, τις συνήθεστερες μεθόδους εκμάθησης που χρησιμοποιούνται, διάφορες τεχνικές *οπίσθιας διάδοσης του σφάλματος* (backpropagation) και παραλλαγές αυτών. Τα αναδρομικά νευρωνικά δίκτυα είναι αυτά που η δομή τους μοιάζει περισσότερο με τη λειτουργία του ανθρώπινου εγκε-

φάλου, όπου οι διεργασίες της αντίληψης, της λήψης αποφάσεων και της εκμάθησης είναι αμιγώς μη γραμμικές. Το κύριο χαρακτηριστικό τους που τα διαφοροποιεί από τα δίκτυα εμπρόσθιας τροφοδότησης είναι ότι περιέχουν τουλάχιστον μια ανάδραση ανάμεσα στους κόμβους του ίδιου επιπέδου ή μεταξύ κόμβων διαφορετικών επιπέδων, όπως αναφέρθηκε στην προηγούμενη ενότητα. Η ιδιότητά αυτή τους προσδίδει δυο πολύ σημαντικά χαρακτηριστικά έναντι των δικτύων πρόσθιας τροφοδότησης. Το πρώτο είναι αυτό της αντίληψης του χρόνου, δηλαδή σε εφαρμογές με δυναμικά συστήματα μπορούν να χρησιμοποιηθούν για να προβλέψουν την έξοδο της στιγμής $t + 1$ έχοντας δεδομένα της στιγμής t ή και προηγούμενων. Το επόμενο χαρακτηριστικό είναι η εισαγωγή της μνήμης στα νευρωνικά δίκτυα. Τέτοιου είδους δίκτυα μπορούν να χρησιμοποιηθούν εγγενώς σε προβλήματα στα οποία οι εισοδοί δεν είναι ανεξάρτητες μεταξύ τους, αλλά υπάρχει κάποια συσχέτιση με προηγούμενα δείγματα.

Τα αναδρομικά νευρωνικά δίκτυα είναι δυναμικά συστήματα που μπορούν να μοντελοποιήσουν και να αναπαραστήσουν χρονικά μεταβαλλόμενα μοντέλα. Το γεγονός ότι χρησιμοποιούνται για να προσομοιώσουν συστήματα με τη χρήση κοινών διαφορικών εξισώσεων, τα κάνει κατάλληλα για ψηφιακή υλοποίηση και έχουν επιπλέον ισχυρές υπολογιστικές δυνατότητες όταν χρησιμοποιούνται σε αυτά τα μοντέλα. Τέτοιου είδους μοντέλα είναι και όλα τα βιολογικά νευρωνικά δίκτυα. Μπορούν, επίσης, να χρησιμοποιηθούν αποδοτικά σε θέματα που περιλαμβάνουν την αναγνώριση και αντίστροφη αναγνώριση συστημάτων, το φιλτράρισμα και την πρόβλεψη πληροφοριών, την κατάταξη δεδομένων σε κλάσεις, την μοντελοποίηση στοχαστικών ακολουθιών, την συμπίεση δεδομένων κ.α.. Κάποιοι από τους τομείς στους οποίους χρησιμοποιούνται αποδοτικά ή έχουν καταδείξει πολλά υποσχόμενα αποτελέσματα είναι οι τηλεπικοινωνίες, ο έλεγχος χημικών εγκαταστάσεων, ο έλεγχος μηχανών και γεννητριών, η ρομποτική, η αναγνώριση ομιλίας, η παραγωγή meta πληροφοριών από εικόνες, η ανάλυση δεδομένων σε βίντεο, η μετάφραση και η αυτόματη συγγραφή κειμένου ή κώδικα.

2.3.2 Περιγραφή της Αρχιτεκτονικής των Αναδρομικών Νευρωνικών Δικτύων

Το βασικό δομικό στοιχείο ενός αναδρομικού νευρωνικού δικτύου είναι οι νευρώνες (μονάδες ή κόμβοι του συστήματος). Οι νευρώνες συνδέονται μεταξύ τους με συναπτικούς δεσμούς οι οποίοι περιγράφονται από το βάρος τους. Διαχωρίζονται συνήθως σε *κόμβους εισόδου* (input), *κρυφούς* (hidden) και *κόμβους εξόδου* (output). Κάθε χρονική στιγμή οι κόμβοι έχουν μια τιμή ενεργοποίησης, η οποία αναπαρίσταται με $u(n)$ για τους κόμβους εισόδου, $x(n)$ για τους κρυφούς και $y(n)$ για τους κόμβους εξόδου. Πολλές φορές αναφερόμαστε σε όλες τις τιμές ενεργοποίησης με τον όρο $x(n)$.

Τα αναδρομικά νευρωνικά δίκτυα κατηγοριοποιούνται σε πλήρως αναδρομικά δίκτυα, στα οποία επιτρέπεται η σύνδεση όλων των νευρώνων μεταξύ τους και στα τοπικά αναδρομικά δίκτυα. Στα πλήρως αναδρομικά δίκτυα δεν υπάρχει διάκριση μεταξύ των κόμβων εισόδου και κάθε κόμβος μπορεί να αποτελέσει είσοδο για οποιονδήποτε άλλο κόμβο, συμπεριλαμβανομένου και του εαυτού του. Τα δεύτερα περιέχουν, συνήθως, μόνο αναδρομικές συνδέσεις μεταξύ των νευρώνων που βρίσκονται ίδιο επίπεδο και η διαδικασία της διάδοσης προς τα εμπρός κατά την εκμάθηση μοιάζει πολύ με αυτή των δικτύων πρόσθιας τροφοδότησης. Συγκριτικά με τα τοπικά αναδρομικά δίκτυα, τα πλήρως αναδρομικά εμφανίζουν μεγαλύτερη αστάθεια κατά την εκπαίδευσης και απαιτούν τη χρήση περίπλοκων και χρονοβόρων αλγορίθμων εκμάθησης. Τα τοπικά αναδρομικά δίκτυα έχουν, αντιθέτως, πιο απλή δομή που τα κάνει πιο αποδοτικά κατά τη διαδικασία εκμάθησης και παρέχουν τη δυνατότητα για έλεγχο της ευστάθειας στους εσωτερικούς τους κόμβους.

Μια ακόμα διάκριση των αναδρομικών δικτύων που χρησιμοποιούνται σε εφαρμογές διακριτού χρόνου είναι σε αναδρομικά δίκτυα με χρονοκαθυστέρηση και σε σύγχρονα αναδρομικά δίκτυα. Τα δίκτυα που λειτουργούν με χρονοκαθυστέρηση εκπαιδεύονται με στόχο τη μείωση του σφάλματος πρόβλεψης, ενώ τα σύγχρονα δίκτυα δε στοχεύουν στην ιδιότητα του να έχουν μνήμη ή καλύτερη πρόβλεψη όσο εκπαιδεύεται το δίκτυο, αλλά κάνουν χρήση των αναδράσεων με στόχο να προσφέρουν καλύτερη δυνατότητα προσέγγισης συναρτήσεων.

Ακολουθεί η περιγραφή και ο ορισμός των παραμέτρων ενός απλού αναδρομικού δικτύου. Το συγκεκριμένο μοντέλο αποτελείται από μονάδες εισόδου που περιγράφονται από το διάνυσμα ενεργοποίησης εισόδου [Hayk94]:

$$u(n) = (u_1(n), \dots, u_K(n))^t \quad (2.1)$$

υπάρχουν εσωτερικοί κόμβοι με διάνυσμα ενεργοποίησης:

$$x(n) = (x_1(n), \dots, x_K(n))^t \quad (2.2)$$

και L κόμβοι εξόδου με διάνυσμα ενεργοποίησης:

$$y(n) = (y_1(n), \dots, y_K(n))^t \quad (2.3)$$

όπου με t ορίζεται ο ανάστροφος πίνακας.

Τα βάρη των συνδέσεων των κόμβων εισόδου, εσωτερικών και εξόδου συλλέγονται σε τρεις διαφορετικούς πίνακες βαρών, διαστάσεων $N \times K$, $N \times N$ και $L \times (K + N)$ αντίστοιχα. Οι πίνακες αυτοί ορίζονται ως:

$$W^{in} = [w^{in}(i, j)], W = [w(i, j)], W^{out} = [w^{out}(i, j)] \quad (2.4)$$

Οι κόμβοι εξόδου μπορεί να συνδέονται ξανά με τους εσωτερικούς κόμβους με βάρη που αποθηκεύονται σε έναν ακόμη πίνακα:

$$W^{back} = [w^{back}(i, j)] \quad (2.5)$$

Οι συνδέσεις που έχουν βάρος 0 θεωρούνται ως μη υπαρκτές και δεν επηρεάζουν το σύστημα, ενώ μπορούν να υπάρχουν συνδέσεις με τους κόμβους εξόδου όχι μόνο από τους εσωτερικούς κόμβους αλλά και κατευθείαν από τους κόμβους εισόδου ή αναδρομικές από αυτές που βρίσκονται στο ίδιο το επίπεδο της εξόδου.

Η ενεργοποίηση των εσωτερικών κόμβων γίνεται σύμφωνα με τον τύπο:

$$x(n+1) = f(W^{in}u(n+1) + Wx(n) + W^{back}y(n)) \quad (2.6)$$

όπου $u(n+1)$ είναι η είσοδος του παρόντος δείγματος και f είναι η συνάρτηση μεταφοράς του συγκεκριμένου κόμβου ή αλλιώς συνάρτηση ενεργοποίησης. Τις περισσότερες φορές αυτή η συνάρτηση έχει τη μορφή κάποιας σιγμοειδούς συναρτήσεως, όπως αυτή της υπερβολικής εφαπτομένης ($f = \tanh$). Μπορεί όμως να είναι μια οποιαδήποτε συνάρτηση.

Η έξοδος του δικτύου υπολογίζεται από τον τύπο:

$$y(n+1) = f^{out}(W^{out}(u(n+1) + x(n+1) + y(n))) \quad (2.7)$$

όπου $(u(n+1), x(n+1), y(n))$ είναι το συγχωνευμένο διάνυσμα των διανυσμάτων εισόδου, εσωτερικών και εξόδου και η συνάρτηση f^{out} είναι συνήθως η ίδια συνάρτηση με την f , αλλά δεν απαγορεύεται και να διαφέρουν μεταξύ τους.

2.3.3 Τεχνικές Εκμάθησης στα Αναδρομικά Νευρωνικά Δίκτυα

Αναμφίβολα, ο πιο δημοφιλής κανόνας εκμάθησης στα νευρωνικά δίκτυα είναι αυτός της οπίσθιας διάδοσης του σφάλματος και έχει κατά συνέπεια επεκταθεί και στα αναδρομικά νευρωνικά δίκτυα με την ονομασία (backpropagation through time - BPTT). Αποτελεί μια γενίκευση του αλγορίθμου ελαχίστων μέσων τετραγώνων (Least Mean Squares - LMS) αλγορίθμου και ονομάζεται, επίσης, γενικευμένος κανόνας δέλτα (generalized delta rule). Χρησιμοποιεί μια τεχνική αναζήτησης της κλίσης (gradient) με στόχο να ελαχιστοποιήσει μια συνάρτηση κόστους που έχει οριστεί για την αξιολόγηση της απόκλισης μεταξύ του επιθυμητού σήματος και της εξόδου του νευρωνικού δικτύου. Η συνάρτηση κόστους που αξιολογείται είναι συνήθως το μέσο τετραγωνικό σφάλμα (mean square error - MSE).

Οπίσθια διάδοση του σφάλματος

Ο αλγόριθμος της οπίσθιας διάδοσης του σφάλματος [Hayk94] χρησιμοποιεί, για κάθε ακολουθία εισόδου, την έξοδο του νευρωνικού δικτύου ώστε να τη συγκρίνει με την επιθυ-

μητή ακολουθία εξόδου και έπειτα σχηματίζει το σφάλμα για κάθε κόμβο του τελευταίου επιπέδου του δικτύου (επίπεδο εξόδου) κάνοντας σύγκριση αυτών των δυο. Ακολούθως, το σφάλμα της εξόδου διαδίδεται προς τα πίσω περνώντας από όλους τους κόμβους του δικτύου σε μια διαδικασία κλειστής επανάληψης. Με βάση αυτό το σφάλμα διορθώνονται τα βάρη του πίνακα W με τη χρήση κάποιου αλγόριθμου μείωσης της κλίσης. Για παράδειγμα ένας δημοφιλής αλγόριθμος μείωσης κλίσης είναι η *στοχαστική κατάβαση κλίσης* (stochastic gradient descent - SGD) [Ketk17]. Προκειμένου να είναι εφικτή η υλοποίηση του αλγορίθμου οπίσθιας διάδοσης του σφάλματος απαιτείται η χρήση κάποιας συνεχούς, μη γραμμικής, αυξανόμενης μονοτονικά και διαφορίσιμης συνάρτησης ενεργοποίησης (συνηθέστερα σιγμοειδής συναρτήσεις όπως αυτή της υπερβολικής εφαπτομένης).

Σε αυτό το σημείο θα γίνει μια σύντομη περιγραφή της διαδικασίας που ακολουθεί η εκμάθηση μέσω της οπίσθιας διάδοσης του σφάλματος στα δίκτυα πρόσθιας τροφοδότησης. Αρχικά, θεωρούμε ένα *πολυεπίπεδο perceptron* (MLP) με πλήθος k κρυφών επιπέδων. Μαζί με το επίπεδο εισόδου και αυτό της εξόδου έχουμε συνολικά $k + 2$ επίπεδα στο δίκτυο, τα οποία αριθμούμε από 0 ως $k + 1$. Το Σχήμα 2.4 απεικονίζει ένα τέτοιο δίκτυο με δυο κρυφά επίπεδα. Ο αριθμός των κόμβων εισόδου είναι K , των κόμβων εξόδου L και των κόμβων των εσωτερικών επιπέδων $1, \dots$, είναι N . Η τιμή που έχει το βάρος της σύνδεσης του i κόμβου στο m επίπεδο και του κόμβου j στο επίπεδο $m + 1$ συμβολίζεται με $w^m(i, j)$. Η ενεργοποίηση του i κόμβου στο m επίπεδο είναι $x^m(i)$, για $m = 0$ αποτελεί κόμβο εισόδου και για $m = k + 1$ κόμβο εξόδου.

Τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση του δικτύου αποτελούνται από ζεύγη εισόδου-εξόδου.

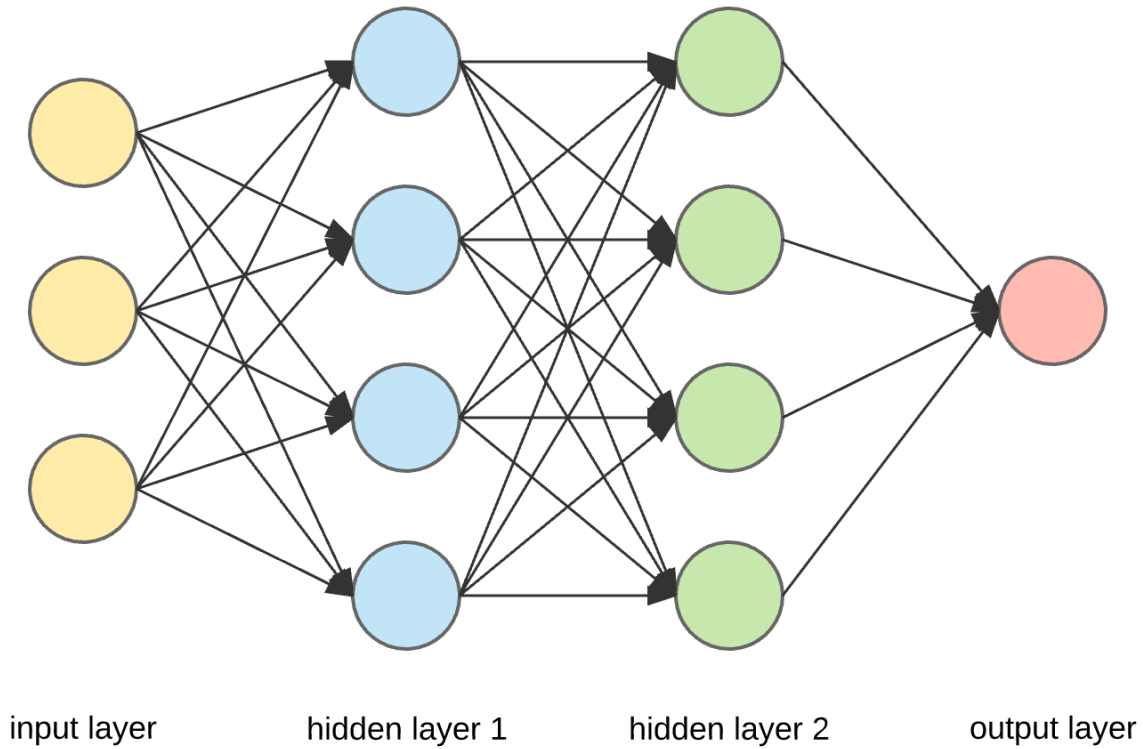
$$u(n) = (x_1^0(n), \dots, x_K^0(n))^t \quad (2.8)$$

$$d(n) = (d_1^{k+1}(n), \dots, x_L^{k+1}(n))^t \quad (2.9)$$

όπου $u(n)$ τα δεδομένα εισόδου και $d(n)$ τα δεδομένα εξόδου. Η παράμετρος n αναφέρεται στον αύξοντα αριθμό του δείγματος και όχι στο χρόνο. Ακολουθεί η ενεργοποίηση όλων των κόμβων του δικτύου, εκτός από αυτούς της εισόδου, σύμφωνα με τον τύπο:

$$x_i^{m+1}(n) = f\left(\sum_{j=1}^{N^m} w_{ij}^m x_j^m(n)\right) \quad (2.10)$$

Πολλές φορές χρησιμοποιείται και ένας όρος *πόλωσης* (bias) (θ_j^m), ο οποίος λειτουργεί αθροιστικά στον παραπάνω τύπο αλλά παραλείπεται σε αυτή την προσπάθεια περιγραφής του αλγορίθμου. Με αυτή τη μέθοδο και με δεδομένα τα στοιχεία εισόδου γίνεται η ενεργοποίηση όλων των κόμβων που βρίσκονται στα εσωτερικά επίπεδα, μέχρι τα στοιχεία του τελευταίου επιπέδου (εξόδου) που έχουν τη μορφή



Σχήμα 2.4: Δίκτυο πρόσθιας τροφοδότησης με 2 κρυφά επίπεδα

$$y(n) = (x_1^{k+1}(n), \dots, x_L^{k+1}(n))^t \quad (2.11)$$

Ο σκοπός της εκπαίδευσης του δικτύου είναι η εύρεση ενός συνόλου βαρών που να ελαχιστοποιεί το μέσο τετραγωνικό σφάλμα μεταξύ της εξόδου του νευρωνικού δικτύου και της επιθυμητής εξόδου $d(n)$.

$$E = \sum_{n=1}^T \|d(n) - y(n)\|^2 = \sum_{n=1}^T E(n) \quad (2.12)$$

Αυτό επιτυγχάνεται προσαρμόζοντας ανάλογα τα βάρη των συνδέσεων προς την αντίθετη κατεύθυνση της κλίσης του σφάλματος, με τη χρήση μια σταθεράς εκμάθησης γ (learning rate) που έχει συνήθως κάποια μικρή τιμή.

$$\frac{\partial E}{\partial w_{ij}^m} = \sum_{t=1, \dots, T} \frac{\partial E(n)}{\partial w_{ij}^m} \quad (2.13)$$

$$\text{νέο } w_{ij}^m = w_{ij}^m - \gamma \frac{\partial E}{\partial w_{ij}^m} \quad (2.14)$$

Οι παραπάνω τύποι αναφέρονται στη μέθοδο της *ομαδικής εκμάθησης* (batch learning). Η ομαδική εκμάθηση έχει σκοπό τη βελτιστοποίηση (ελαχιστοποίηση) του , όπως αναφέρθηκε παραπάνω και οι προσαυξήσεις των βαρών συσσωρεύονται μέχρι να γίνει η τελική τροποποίησή τους στο τέλος κάθε εποχής. Υπάρχει, επίσης και η *αυξητική εκμάθηση* (incremental

learning), όπου η μεταβολή των βαρών γίνεται σταδιακά στο τέλος κάθε βήματος του αλγορίθμου εκμάθησης. Η αυξητική εκμάθηση είναι μια στοχαστική μέθοδος βελτιστοποίησης των παραμέτρων του δικτύου, ενώ η ομαδική εκμάθηση ντετερμινιστική. Στην περίπτωση της ομαδικής εκμάθησης, οι παραπάνω τύποι αντικαθίστανται από τον εξής

$$\text{νέο } w_{ij}^m = w_{ij}^m - \gamma \frac{\partial E(n)}{\partial w_{ij}^m} \quad (2.15)$$

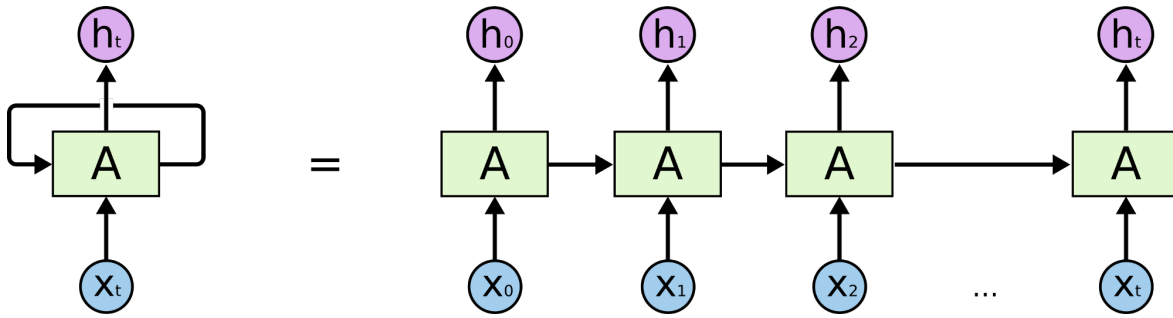
Ο αλγόριθμος της οπίσθιας διάδοσης του σφάλματος κάνει επαναληπτικά την εφαρμογή των παραπάνω σχέσεων για κάθε εποχή μέχρι να επιτύχει μείωση του σφάλματος σε τιμή μικρότερη από κάποιο προκαθορισμένο *κατώφλι* (threshold) ή μέχρι η μεταβολή του σφάλματος να πέσει κάτω από ένα άλλο προκαθορισμένο κατώφλι ή μέχρι να ολοκληρωθεί ο αριθμός των εποχών που έχει τεθεί οριστεί στην αρχή της εκπαίδευσης. Συνήθως απαιτείται κάποιος αρκετά μεγάλος αριθμός εποχών για να εκπληρωθεί κάποιος από τους παραπάνω περιορισμούς.

Η κάθε εποχή έχει πολυπλοκότητα $O(TM)$, όπου είναι ο συνολικός αριθμός των συνδέσεων μεταξύ των κόμβων του δικτύου. Η προσέγγιση μέσω της μεθόδου κατάβασης κλίσης και του αλγορίθμου οπίσθιας διάδοσης του σφάλματος έχει το μειονέκτημα της αργής σύγκλισης, διότι τυπικά επιλέγεται κάποια μικρή σταθερά εκμάθησης για να αποφευχθεί η αποσταθεροποίηση του δικτύου. Υπάρχουν, όμως, τρόποι για την επίτευξη της σύγκλισης, όπως η χρήση δευτεροβάθμιων τεχνικών κατάβασης κλίσης που εκμεταλλεύονται την καμυλότητα της κλίσης του σφάλματος, αλλά έχουν πολυπλοκότητα $O(TM^2)$. Ένα ακόμη αρνητικό που έχουν οι τεχνικές κατάβασης κλίσης είναι ότι γίνεται αναζήτηση μόνο του τοπικού ελαχίστου του σφάλματος. Το πρόβλημα αυτό μπορεί να αντιμετωπιστεί με διάφορους τρόπους, όπως η προσθήκη θορύβου κατά την εκμάθηση του δικτύου, η επανάληψη όλης της διαδικασίας με διαφορετική αρχικοποίηση των βαρών ή με τη χρήση υπαρχόντων πληροφοριών των δεδομένων εισόδου - εξόδου που μπορούν να φανούν χρήσιμες κατά την εκπαίδευση.

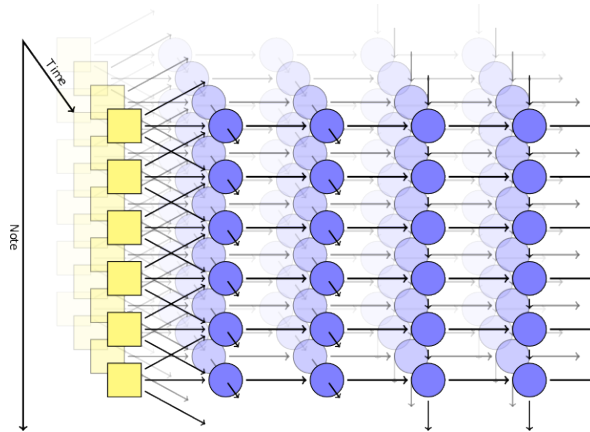
2.3.4 Backpropagation Through Time (BPTT)

Ο αλγόριθμος της οπίσθιας διάδοσης του σφάλματος που χρησιμοποιείται στα δίκτυα πρόσθιας τροφοδότησης δεν μπορεί να μεταφερθεί ως έχει στα αναδρομικά, επειδή προϋποθέτει την ύπαρξη αποκλειστικά ακυκλικών συνδέσεων μεταξύ των κόμβων του δικτύου. Η λύση δίνεται με την εισαγωγή του backpropagation through time αλγορίθμου [Werb90], που «ξεδιπλώνει» τις συνδέσεις μεταξύ των κόμβων σε ξεχωριστά χρονικά βήματα, δημιουργώντας πανομοιότυπα αντίγραφα και ανακατευθύνει τις συνδέσεις μεταξύ αυτών ώστε να προκύψει ένα δίκτυο πρόσθιας τροφοδότησης όπως αυτό του Σχήματος 2.5a και 2.5b.

Οι πίνακες των βαρών του συστήματος w_{ij}^{in} , w_{ij} , w_{ij}^{out} , w_{ij}^{back} παραμένουν οι ίδιοι σε



(a) Το δίκτυο απλοποιείται με τα εσωτερικά επίπεδα είναι κλειστά στο A



(b) Αναπαράσταση ολοκληρωμένου δικτύου.

Σχήμα 2.5: Αναπαράσταση αναδρομικών δικτύων «ξεδιπλωμένων» στο χρόνο.

όλα τα αντίγραφα των επιπέδων. Τα δεδομένα αποτελούνται από μια χρονοσειρά δειγμάτων εισόδου-εξόδου που έχουν την παρακάτω μορφή:

$$u(n) = (u_1(n), \dots, u_K(n))^t \quad (2.16)$$

$$d(n) = (d_1(n), \dots, d_L(n))^t \quad (2.17)$$

Η διαδικασία της εκμάθησης ξεκινάει από το πρώτο επίπεδο και προχωράει σταδιακά στα επόμενα επίπεδα της στοίβας που έχει δημιουργηθεί από το ξεδίπλωμα των επιπέδων στο χρόνο. Σε κάθε αντίγραφο των επιπέδων τη χρονική στιγμή n διαβάζεται η είσοδος $u(n)$, υπολογίζεται το $x(n)$ των ενδιάμεσων επιπέδων με βάση τα $u(n)$, $x(n-1)$ και $y(n-1)$ (όταν το τελευταίο δεν είναι θ) και τέλος υπολογίζεται η έξοδος $y(n)$. Η συνάρτηση του σφάλματος που ελαχιστοποιείται είναι και πάλι:

$$E = \sum_{n=1}^T \|d(n) - y(n)\|^2 = \sum_{n=1}^T E(n) \quad (2.18)$$

με τη διαφορά ότι η έννοια του t έχει μετατραπεί από τον αύξοντα αριθμό του δείγματος εκπαίδευσης σε χρονική στιγμή.

Ο αλγόριθμος που ακολουθεί το *BPTT* έχει ως εξής:

- **Είσοδος:** η χρονοσειρά των δεδομένων εκπαίδευσης και τα βάρη w_{ij} που αντιστοιχούν στην συγκεκριμένη χρονική στιγμή.
- **Έξοδος:** τα νέα βάρη των συνδέσεων.

Υπολογιστικά βήματα του αλγορίθμου:

- Εμπρόσθιο πέρασμα, όπως περιγράφηκε παραπάνω μέχρι την έξοδο $y(n)$.
- Υπολογισμός από το τέλος προς την αρχή των επιπέδων (για $n = T, \dots, 1$) και για την κάθε ενεργοποίηση των κόμβων $x_i(n), y_j(n)$ ενός όρου της διάδοσης του σφάλματος $\delta_i(n)$, ο οποίος δίνεται από τους τύπους:

$$\delta_j(T) = (d_j(T) - y_j(T)) \frac{\partial f(u)}{\partial u} \Big|_{u=z_j(T)} \quad (2.19)$$

για τους κόμβους εξόδου στο χρονικό επίπεδο ,

$$\delta_i(T) = \left[\sum_{j=1}^L \delta_j(T) w_{ji}^{out} \right] \frac{\partial f(u)}{\partial u} \Big|_{u=z_i(n)} \quad (2.20)$$

για τους εσωτερικούς κόμβους $x_i(T)$ στο χρονικό επίπεδο ,

$$\delta_j(n) = [(d_j(n) - y_j(n)) + \sum_{i=1}^N \delta_i(n+1) w_{ji}^{back}] \frac{\partial f(u)}{\partial u} \Big|_{u=z_j(n)} \quad (2.21)$$

για τους κόμβους εξόδου των προηγούμενων χρονικών επιπέδων και

$$\delta_i(n) = \left[\sum_{j=1}^N \delta_j(n+1) w_{ji} + \sum_{j=1}^L \delta_j(n) w_{ji}^{out} \right] \frac{\partial f(u)}{\partial u} \Big|_{u=z_i(n)} \quad (2.22)$$

για τους εσωτερικούς κόμβους των προηγούμενων χρονικών επιπέδων. Το $z_i(n)$ είναι η μέγιστη τιμή που μπορούν να πάρουν οι συγκεκριμένοι κόμβοι.

- Εκ νέου υπολογισμός των βαρών σύμφωνα με τις σχέσεις:

$$\text{νέο } w_{ij} = w_{ij} + \gamma \sum_{n=1}^T \delta_i(n) x_j(n-1) \quad (2.23)$$

υποθέτοντας $x_j(n-1) = 1$ για $n = 1$,

$$\text{νέο } w_{ij}^{in} = w_{ij}^{in} + \gamma \sum_{n=1}^T \delta_i(n) u_j(n) \quad (2.24)$$

$$\text{νέο } w_{ij}^{out} = w_{ij}^{out} + \gamma * \begin{cases} \sum_{n=1}^T \delta_i(n) u_j(n), & \text{αν το } j \text{ είναι κόμβος της εισόδου} \\ \sum_{n=1}^T \delta_i(n) x_j(n-1), & \text{αν το } j \text{ είναι εσωτερικός κόμβος} \end{cases} \quad (2.25)$$

$$\text{νέο } w_{ij}^{back} = w_{ij}^{back} + \gamma \sum_{n=1}^T \delta_i(n) y_j(n-1) \quad (2.26)$$

υποθέτοντας $y_j(n-1) = 1$ για $n = 1$.

Το θέμα της αργής σύγκλισης που αναφέρθηκε για την οπίσθια διάδοση του σφάλματος στα δίκτυα πρόσθιας τροφοδότησης εξακολουθεί να υπάρχει και στο BPTT και η πολυπλοκότητα του αλγορίθμου που περιγράφηκε είναι $O(TN^2)$, με τον αριθμό των εσωτερικών κόμβων. Συνήθως, χρειάζονται πολλές εποχές για να ολοκληρωθεί η διαδικασία της εκπαίδευσης. Η συνεχόμενη εκτέλεση αυτών των εποχών έχει ως αποτέλεσμα τη δημιουργία ενός πολύπλοκου δυναμικού συστήματος που συχνά μπορεί να αποκλίνει της επιθυμητής συμπεριφοράς. Επομένως, είναι πιθανό να δημιουργηθούν διακλαδώσεις όταν οι τιμές αρχικοποίησης των βαρών του δικτύου είναι αρκετά διαφορετικές από τη δυναμική του συστήματος που προσπαθούμε να μοντελοποιήσουμε. Αποτέλεσμα αυτών των διακλαδώσεων μπορεί να είναι η αλλοίωση των πληροφοριών της κλίσης και η εκτόξευση του σφάλματος σε μη αποδεκτές (πολύ υψηλές) τιμές, το οποίο στην περίπτωση του BPTT δεν εγγυάται τη σύγκλιση σε κάποια κοντινή περιοχή του ελαχίστου του. Τα προβλήματα αυτά δεν συναντώνται στα δίκτυα πρόσθιας τροφοδότησης επειδή μοντελοποιούν μόνο απλές συναρτήσεις και όχι δυναμικά συστήματα. Τέλος, δεν υπάρχει κάποια συγκεκριμένη τεχνική για να υπερκεραστούν αυτά τα προβλήματα και συνήθως χρειάζονται αρκετά πειράματα και υπολογιστικός χρόνος για να επιτευχθεί ένα ικανοποιητικό αποτέλεσμα. Για τους λόγους αυτούς, η τεχνική του BPTT χρησιμοποιείται σχεδόν αποκλειστικά σε μικρά δίκτυα μεγέθους 3 – 20 κόμβων ανά επίπεδο και η χρήση μεγαλύτερων δικτύων αποδεικνύεται πολύ δαπανηρή από άποψη χρήσης υπολογιστικού εξοπλισμού και χρόνου.

Ένα ακόμη μειονέκτημα που αφορά το ομαδική εκμάθηση που γίνεται στο BPTT, αλλά και την απλή περίπτωση της οπίσθιας διάδοσης του σφάλματος στα δίκτυα πρόσθιας τροφοδότησης είναι ότι η μεταβολή των βαρών γίνεται αποκλειστικά στο τέλος κάθε εποχής, μετά από ένα πλήρες πέρασμα των δεδομένων εκμάθησης. Το γεγονός αυτό καθιστά την τεχνική του της οπίσθιας διάδοσης του σφάλματος μη κατάλληλη για εφαρμογές πραγματικού χρόνου που απαιτούν τη συνεχή ενημέρωση των βαρών. Σε τέτοιου είδους εφαρμογές χρησιμοποιείται συνήθως ένας άλλος αλγόριθμος εκμάθησης, ο *real time recurrent learning* (RTRL) [WILL89], που πραγματοποιεί την ενημέρωση των βαρών στο τέλος κάθε βήματος του αλγορίθμου.

2.4 Επιλογή υπερπαραμέτρων στα αναδρομικά νευρωνικά δίκτυα

Εκτός από την επιλογή του καλύτερου μοντέλου αναδρομικών νευρωνικών δικτύων για κάθε εφαρμογή, υπάρχουν κάποιες βασικές παράμετροι που πρέπει να οριστούν και είναι αυτές που καθορίζουν σε μεγαλύτερο βαθμό την επιτυχία ή όχι της εκπαίδευσης. Στις περισσότερες από αυτές έχουμε ήδη αναφερθεί στην παρούσα διπλωματική εργασία, δεν έχουμε όμως αναλύσει πως γίνεται η επιλογή των κατάλληλων τιμών για αυτές. Οι βασικότερες είναι ο αριθμός των κρυφών επιπέδων του δικτύου, το πλήθος των κόμβων του κάθε επιπέδου, ο ρυθμός εκμάθησης και άλλα.

2.4.1 Αριθμός κρυφών επιπέδων

Ο αριθμός των κρυφών επιπέδων του δικτύου εξαρτάται βασικά από το πρόβλημα που θέλουμε να προσομοιώσουμε. Υπάρχει μια γενική αρχή που λέει ότι αν τα δεδομένα της λύσης του προβλήματος είναι γραμμικά διαχωριζόμενα, τότε δεν υπάρχει η ανάγκη για χρήση κανενός κρυφού επιπέδου, αφού η συνάρτηση ενεργοποίησης μπορεί να εφαρμοστεί στο επίπεδο της εισόδου που, σε συνδυασμό με αυτό της εξόδου, είναι αρκετά για να μοντελοποιήσουν τέτοιου είδους προβλήματα. Στην περίπτωση, όμως, προβλημάτων που έχουν να κάνουν με περίπλοκα μη γραμμικά συστήματα χρησιμοποιούνται συνήθως, επιπλέον από αυτό της εισόδου, ένα ή δυο κρυφά επίπεδα. Πιο συγκεκριμένα, ένα κρυφό επίπεδο χρησιμοποιείται στην περίπτωση κάποιας μη γραμμικής συνάρτησης που συνδέει δεδομένα από ένα πεπερασμένο χώρο σε έναν άλλο, ενώ δυο κρυφά επίπεδα μπορούν να αναπαραστήσουν οποιοδήποτε αυθαίρετο κριτήριο απόφασης με αυθαίρετη ακρίβεια και οποιαδήποτε γραμμική ή μη γραμμική σχέση με οποιαδήποτε ακρίβεια (όσο μικρή επιθυμούμε). Αν η ακρίβεια είναι το μόνο ζητούμενο της προσέγγισης του προβλήματος μπορεί να χρησιμοποιηθεί και τρίτο κρυφό επίπεδο, αλλά η τεχνική αυτή θα αυξήσει κατά πολύ το χρόνο της εκπαίδευσης του συστήματος και δεν ενδείκνυται γενικότερα, αφού ένα ή δυο κρυφά επίπεδα είναι επαρκή για να επιλύσουν οποιοδήποτε σύνθετο μη γραμμικό πρόβλημα. Καταλήγοντας, οι περισσότεροι ερευνητές προτείνουν τη χρήση όσο το δυνατόν λιγότερων κρυφών επιπέδων, σύμφωνα πάντα με την εφαρμογή, για την εξοικονόμηση υπολογιστικών πόρων, χρόνου εκπαίδευσης και πολυπλοκότητας της λύσης [Kars].

2.4.2 Πλήθος κόμβων κρυφών επιπέδων

Υπολογισμός χρειάζεται, επίσης, και για την επιλογή του αριθμού των νευρώνων των κρυφών επιπέδων του δικτύου. Όταν πρόκειται για τα επίπεδα της εισόδου και της εξόδου, ο αριθμός των κόμβων καθορίζεται από τα δεδομένα εισόδου-εξόδου της εφαρμογής, αλλά

όταν πρόκειται για την καλύτερη επιλογή του πλήθους των κόμβων των εσωτερικών επιπέδων δεν υπάρχει κάποια πάγια τακτική που να εγγυάται το βέλτιστο αποτέλεσμα. Βασικό μέλημα είναι η αποφυγή γεγονότων όπως η *υπερπροσαρμογή* (overfitting) και η *υποπροσαρμογή* (underfitting) στα δεδομένα εκπαίδευσης, που συμβαίνει όταν ο αριθμός των κόμβων είναι πολύ μεγάλος ή πολύ μικρός αντίστοιχα σε σχέση με το μέγεθος και την πολυπλοκότητα του προβλήματος. Άλλοι γνώμονες που πρέπει να ληφθούν υπόψη κατά την επιλογή του πλήθους των κόμβων είναι η ελαχιστοποίηση του χρόνου της εκπαίδευσης και η βελτίωση της ακρίβειας της εξόδου. Στη συνέχεια αναφέρονται οι δημοφιλέστερες μέθοδοι που χρησιμοποιούνται για την επιλογή του αριθμού των κόμβων των κρυφών επιπέδων [Stat09].

Η πιο γνωστή, ίσως, μέθοδος είναι αυτή του Kolmogorov, που υποστηρίζει ότι οποιαδήποτε συνεχής συνάρτηση μπορεί να αναπαρασταθεί από ένα νευρωνικό δίκτυο που περιέχει ακριβώς $2n + 1$ κόμβους σε ένα μόνο κρυφό επίπεδο, με το n να αποτελεί τον αριθμό των κόμβων εισόδου. Η θεωρία, όμως, αυτή έχει αποδειχθεί μη ακριβής για το λόγο ότι ισχύει μόνο για μια συγκεκριμένη συνάρτηση ενεργοποίησης των κόμβων που είναι αρκετά πιο πολύπλοκη από τις σιγμοειδείς που χρησιμοποιούνται τις περισσότερες φορές στην πράξη. Σε διαφορετική περίπτωση, το πλήθος των κόμβων για να διατηρηθεί μόνο ένα κρυφό επίπεδο μπορεί να φτάσει ακόμα και τον αριθμό των δειγμάτων εκπαίδευσης. Σαν επέκταση αυτού του θεωρήματος, ο *Huang* απέδειξε το 2003 ότι με δυο κρυφά επίπεδα, m νευρώνες στο επίπεδο της εξόδου και αριθμό δειγμάτων για εκπαίδευση, ο αριθμός των κόμβων στα κρυφά επίπεδα δεν πρέπει να ξεπερνά τους $2\sqrt{(m + 2)N}$ κόμβους. Πιο συγκεκριμένα, προτείνει ο αριθμός των κόμβων στο πρώτο κρυφό επίπεδο να μην ξεπερνά τους:

$$\sqrt{(m + 2)N} + 2\sqrt{\frac{N}{m + 2}} \quad (2.27)$$

και του δεύτερου επιπέδου τους:

$$m\sqrt{\frac{N}{m + 2}} \quad (2.28)$$

Τρόποι επιλογής του βέλτιστου αριθμού κόμβων στα εσωτερικά επίπεδα αποτελούν, ακόμα, η εξαντλητική αναζήτηση που ελέγχει όλες τις πιθανές τοπολογίες του συστήματος, αλλά δεν είναι καθόλου πρακτική εξαιτίας του χρόνου που απαιτεί η εφαρμογή της και η δοκιμή τυχαίου αριθμού από κόμβους και η μετέπειτα διόρθωση σύμφωνα με το σφάλμα. Άλλες μέθοδοι που χρησιμοποιούνται είναι οι ευριστικές, που κάνουν χρήση της εμπειρικής γνώσης που έχει συσσωρευθεί από την επίλυση προηγούμενων προβλημάτων με νευρωνικά δίκτυα και χρησιμοποιούνται κυρίως σαν σημείο εκκίνησης της αναζήτησης για τις προηγούμενες τεχνικές. Κάποιες από αυτές είναι ότι ο αριθμός των κόμβων των κρυφών επιπέδων πρέπει να κυμαίνεται στο 70 – 90% του πλήθους των κόμβων εισόδου, ότι ο συνολικός αριθμός τους δεν πρέπει να ξεπερνάει το διπλάσιο του πλήθους των κόμβων εισόδου ή ότι το μέγεθος κάθε εσωτερικού επιπέδου πρέπει να είναι ανάμεσα σε αυτά της εισόδου και εξόδου.

Μια, ακόμα, μέθοδος που αξίζει να σημειωθεί είναι αυτή που έχοντας ένα δεδομένο αριθμό κόμβων προσθέτει διαδοχικά ή αφαιρεί συνδέσεις μεταξύ τους μέχρι να επιτύχει το καλύτερο αποτέλεσμα. Τέτοιου είδους τεχνικές δουλεύουν πολύ αποδοτικά, αλλά η πολυπλοκότητά τους συμβάλλει στο να μη χρησιμοποιούνται συχνά. Τέλος, μπορεί να μην υπάρχει ξεκάθαρος τρόπος για τη βέλτιστη επιλογή των κόμβων των εσωτερικών επιπέδων, αλλά έχει φανεί πειραματικά ότι καλύτερα αποτελέσματα έχουν οι μέθοδοι που χρησιμοποιούν τον ίδιο ή περίπου τον ίδιο αριθμό κόμβων σε όλα τα κρυφά επίπεδα.

2.5 Νευρωνικά Δίκτυα Elman και Jordan

Τα δίκτυα Elman και Jordan είναι δύο συγκεκριμένοι τύποι αναδρομικών δικτύων. Είναι αρκετά δημοφιλή καθώς είναι σχετικά απλή η αρχιτεκτονική τους ενώ η δομή τους, όπως και των περισσότερων αναδρομικών δικτύων, τους επιτρέπει να έχουν μνήμη και τα κάνει κατάλληλα για επίλυση προβλημάτων που εμπεριέχουν πρόβλεψη χρονοσειράς.

2.5.1 Δίκτυα Elman

Το δίκτυο Elman [Crus06] είναι ένα αναδρομικό νευρωνικό δίκτυο τριών επιπέδων (Σχήμα 2.6) με επιπλέον μια ομάδα από «νευρώνες περιεχομένου» (context units), όπου στο Σχήμα 2.6 συμβολίζονται με c . Το μεσαίο (εσωτερικό) επίπεδο είναι συνδεδεμένο στους νευρώνες περιεχομένου με κάποιο βάρος w . Σε κάθε χρονικό βήμα, η είσοδος τροφοδοτείται με δεδομένα, τα οποία περνούν μια φορά από το δίκτυο και μετά εφαρμόζεται ο κανόνας εκμάθησης. Ωστόσο οι νευρώνες περιεχομένου έχουν ήδη κρατήσει ένα αντίγραφο από τα δεδομένα που περάσανε από το εσωτερικό επίπεδο γιατί η πληροφορία πηγαίνει εκεί προτού εφαρμοστεί ο κανόνας εκμάθησης. Αυτό δίνει στα δίκτυα Elman την ικανότητα να «θυμούνται» παρελθοντικές καταστάσεις, και έτσι να έχουν καλή απόδοση σε προβλήματα πρόβλεψης χρονοσειράς.

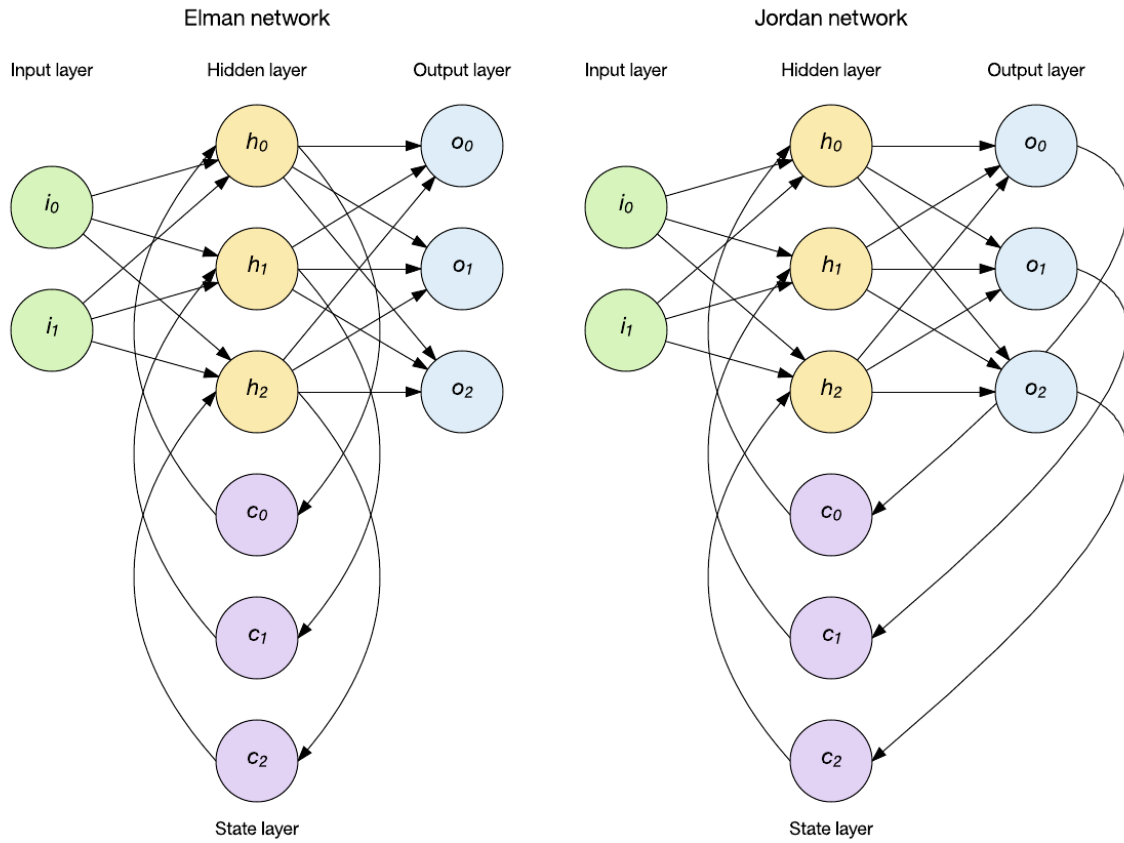
Τα νευρωνικά δίκτυα Jordan [Crus06] μοιάζουν πολύ με τα Elman. Είναι επίσης ένα δίκτυο τριών επιπέδων (Σχήμα 2.6) με την διαφορά ότι οι νευρώνες περιεχομένου τροφοδοτούνται από το επίπεδο εξόδου (Σχήμα 2.6), σε αντίθεση με τα Elman που τροφοδοτούνται από το εσωτερικό επίπεδο.

Τα δίκτυα Elman και Jordan είναι επίσης γνωστά ως απλά αναδρομικά δίκτυα (simple recurrent networks - SRN). Οι μαθηματικές εξισώσεις που διέπουν τα δύο δίκτυα περιγράφονται παρακάτω [Elma90, Jord97]:

Elman:

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h) \quad (2.29)$$

$$y_t = \sigma_y(W_y h_t + b_y) \quad (2.30)$$



Σχήμα 2.6: Νευρωνικά Δίκτυα Elman και Jordan

Jordan:

$$h_t = \sigma_h(W_h x_t + U_h y_{t-1} + b_h) \quad (2.31)$$

$$y_t = \sigma_y(W_y h_t + b_y) \quad (2.32)$$

όπου:

x_t : διάνυσμα εισόδου

h_t : διάνυσμα εσωτερικού επιπέδου

y_t : διάνυσμα εξόδου

W, U και b : πίνακες παραμέτρων και διάνυσμα *bias*

σ_h και σ_y : Συναρτήσεις ενεργοποίησης

Κεφάλαιο 3

Συλλογή Δεδομένων

3.1 Εισαγωγή

Η συλλογή δεδομένων για τα προβλήματα και τις εφαρμογές της μηχανικής μάθησης είναι υψίστης σημασίας. Θα μπορούσε να πει κανείς ότι σε πολλές, αν όχι σε όλες, τις εφαρμογές μηχανικής μάθησης η δυσκολότερη δουλειά και αυτή που έχει μεγαλύτερο βάρος είναι η συλλογή, προετοιμασία και οργάνωση των δεδομένων. Άλλωστε όπως εξηγήθηκε και στο Κεφάλαιο 2 στην μηχανική μάθηση τροφοδοτούμε ένα μοντέλο με δεδομένα για να «μάθει» από αυτά. Εάν τα δεδομένα είναι κακής ποιότητας, ακόμα και να έχει στηθεί το πιο καλό προγραμματιστικά μοντέλο, είναι πολύ πιθανό να λειτουργεί πολύ χειρότερα από ένα πολύ πιο απλό μοντέλο που έχει «μάθει» από πολύ καλύτερα δεδομένα. Με άλλα λόγια η «εξυπνάδα» του αλγορίθμου θα προέλθει από την καλή σε ποιότητα γνώση που θα του δώσουμε. Σε αυτήν την εργασία θα συλλεχθούν πέντε διαφορετικά *σύνολα δεδομένων* (datasets). Κάθε σύνολο δεδομένων θα είναι και η «τροφή» καθενός από τα πέντε μοντέλα που θα εξεταστούν πειραματικά στο Κεφάλαιο 4. Με την σειρά τα σύνολα δεδομένων που θα εξεταστούν είναι:

1. Η ιστορική τιμή του Bitcoin
2. Τίτλοι από την καθημερινή ειδησεογραφία που αφορούν το Bitcoin, με ανάλυση συναισθήματος πραγματοποιούμενη από τον άνθρωπο (sentiment master)
3. Αριθμός των ημερησίων tweets με hashtag #Bitcoin)
4. Ημερήσιος αριθμός Google Trend για τον όρο «Bitcoin»
5. Τίτλοι από την καθημερινή ειδησεογραφία που αφορούν το Bitcoin, με ανάλυση συναισθήματος πραγματοποιούμενη αλγόριθμο (sentiment analysis)

Μια σημαντική σημείωση που πρέπει να γίνει έχει να κάνει με το χρονικό εύρος των δεδομένων προς συλλογή. Όλες οι παραπάνω κατηγορίες αφορούν χρονοσειρές, οπότε έχει σημασία να οριστεί το χρονικό εύρος. Αυτή η εργασία έχει σκοπό την πρόβλεψη της τιμής του Bitcoin με βάση την πληροφορία που υπάρχει στο Διαδίκτυο. Ωστόσο μελλοντική επέκταση

(Κεφάλαιο 5) θα μπορούσε να είναι η πρόβλεψη σε πραγματικό χρόνο. Επειδή η εργασία αυτή είναι η βάση της όλης ιδέας, θεωρήθηκε ορθό η πρώτη μελέτη να έχει σχετικά λίγα δεδομένα, με σκοπό σε περίπτωση που τα αποτελέσματα δεν είναι ικανοποιητικά, ο δρόμος της βελτίωσης να είναι προς μια κατεύθυνση αύξησης των δεδομένων. Έτσι τα δεδομένα που θα συλλέχθηκαν για όλες τις παραπάνω χρονοσειρές είναι 100 σημείων. Το φυσικό νόημα του συγκεκριμένου αριθμού είναι ότι παριστάνει περίπου τρεις ημερολογιακούς μήνες. Έτσι το ερώτημα που τίθεται στην εργασία αφορά το αν είναι αρκετά τα δεδομένα των τελευταίων 3 μηνών για να γίνει πρόβλεψη της τιμής του Bitcoin.

3.2 Συλλογή Δεδομένων

3.2.1 Προηγούμενες τιμές του Bitcoin

Το Bitcoin όπως και κάθε κρυπτονόμισμα αλλά και οποιοδήποτε χρηματοοικονομικό προϊόν με την ευρύτερη έννοια, έχει μια τιμή που καθορίζεται από την προσφορά και την ζήτηση. Η ειδοποιός διαφορά είναι ότι επηρεάζεται πολύ έντονα τις τάσεις και την δημοσιότητα. Γι' αυτό και οι αλλαγές στην τιμή του είναι πολύ μεγάλες και παρουσιάζει σημαντικά μεγαλύτερη μεταβλητότητα σε σχέση, για παράδειγμα, με μετοχές όπως της Apple και της Amazon. Επίσης, όπως εξηγήθηκε στο Κεφάλαιο 2, η αγορά των κρυπτονομισμάτων έχει πολύ λίγη ρύθμιση οπότε υπάρχει περίπτωση διάφοροι ιστότοποι και πλατφόρμες εμπορίας να έχουν ψεύτικες πληροφορίες. Είναι σημαντικό λοιπόν ως πηγή για την ιστορική τιμή του Bitcoin να επιλεγεί ένα γνωστό και αποδεδειγμένα έντιμο μέσο. Στην παρούσα εργασία η ιστορική τιμή του Bitcoin συλλέχθηκε από τον ιστότοπο Bitfinex¹. Παράδειγμα των δεδομένων που συλλέχθηκαν δίνεται στον Πίνακα 3.1.

Bitcoin Dataset						
Date	Open	High	Low	Close	Volume (BTC)	Volume (Currency)
11/11/2017	6588.18	6820	6218	6355.13	14942.70256	96800268.09
10/11/2017	7150.61	7343.18	6429.44	6588.18	26614.27605	181284361.1
09/11/2017	7150.61	7490	7061.2	7148	14799.26577	106985408.5
08/11/2017	7131.38	7888	7080.01	7450.32	24679.72055	183501327.1

Πίνακας 3.1: Πρώτες τέσσερις σειρές του πίνακα της ιστορικής τιμής του Bitcoin.

με τις τιμές των στηλών να σημαίνουν,

- **Date:** Ημερομηνία καταγραφής

¹ <https://www.bitfinex.com/>

- **Open:** Τιμή του Bitcoin στο ημερήσιο άνοιγμα της αγοράς (δολάρια)
- **High:** Η μέγιστη τιμή που καταγράφηκε εντός της ημέρας
- **Low:** Η ελάχιστη τιμή που καταγράφηκε εντός της ημέρας
- **Close:** Τιμή του Bitcoin στο κλείσιμο της αγοράς
- **Volume (BTC):** Όγκος συναλλαγής σε Bitcoin
- **Volume (Currency):** Όγκος συναλλαγής σε δολάρια

Μια παρατήρηση που πρέπει να γίνει αφορά τις τιμές ανοίγματος και κλεισίματος. Στο χρηματιστήριο, για παράδειγμα στο New York Stock Exchange υπάρχουν συγκεκριμένες ώρες συναλλαγής και συγκεκριμένα από 9:30 το πρωί μέχρι 16:00 το απόγευμα. Οι τιμές των μετοχών σε αυτές τις ώρες αντιστοιχούν στις τιμές Open και Close ενώ τα σαββατοκύριακα το NYSE παραμένει κλειστό. Στην περίπτωση της αγοράς κρυπτονομισμάτων, όπως εξηγήθηκε στο Κεφάλαιο 2, η συναλλαγή είναι ανοιχτή όλες τις ώρες της ημέρας, κάθε μέρα του χρόνου. Έτσι οι τιμές ανοίγματος και κλεισίματος αφορούν την έναρξη και το τέλος της ημερολογιακής μέρας και συγκεκριμένα η Open αντιστοιχεί στην τιμή του Bitcoin την ώρα 00:00 της συγκεκριμένη μέρας, ενώ η τιμή Close αντιστοιχεί στην τιμή την ώρα 23:59.

Όπως περιγράφηκε στο Κεφάλαιο 2, η ιδέα της συγκεκριμένης διπλωματικής είναι να χρησιμοποιήσουμε την πληροφορία που υπάρχει στο Διαδίκτυο για να προβλέψουμε την αγορά την επόμενη ημέρα. Ωστόσο όπως εξηγήθηκε στην παραπάνω παράγραφο η «επόμενη μέρα» στην αγορά κρυπτονομισμάτων ορίζεται σαν την τιμή Open στις 00:00, με αποτέλεσμα να αφαιρεθούν όλες οι άλλες στήλες από το πρώτο dataset και να κρατηθεί μόνο η ημερομηνία και η τιμή Open. Παράδειγμα του ανανεωμένου dataset φαίνεται στον πίνακα 3.2:

Bitcoin Dataset	
Date	Open
11/11/2017	6588.18
10/11/2017	7150.61
09/11/2017	7150.61
08/11/2017	7131.38
⋮	⋮

Πίνακας 3.2: Νέος πίνακας της ιστορικής τιμής του Bitcoin.

3.2.2 Ανάλυση συναισθήματος από την ειδησιογραφία της ημέρας

Η δεύτερη συλλογή δεδομένων που συγκεντρώθηκε περιέχει τους τίτλους ειδήσεων που αφορούν το Bitcoin, πάνω στους οποίους εφαρμόζονται τεχνικές ανάλυσης συναισθήματος. Με άλλα λόγια σε πρώτο βήμα συλλέγονται οι τίτλοι ειδήσεων και σε δεύτερο βήμα αποφασίζεται αν αυτό οι τίτλοι αναφέρονται με θετικό τρόπο στο Bitcoin ή όχι. Στην δημιουργία της συγκεκριμένης συλλογής δεδομένων υπάρχουν δύο προκλήσεις. Η επιλογή των ειδησιογραφικών πηγών, με κάποιο αντικειμενικό κριτήριο και η εξαγωγή του σωστού συναισθήματος. Στην δεύτερη περίπτωση η λύση είναι απλή: στο συγκεκριμένο στάδιο την ανάλυση συναισθήματος στους τίτλους την πραγματοποιούμε εμείς, ως άνθρωποι, χωρίς την χρήση κάποιου αλγόριθμου. Ο άνθρωπος θεωρείται sentiment master καθώς μπορεί να αντιληφθεί και να εντοπίσει το συναίσθημα καλύτερα από την τεχνική νοημοσύνη (μέχρι στιγμής). Στην πρώτη περίπτωση, που αφορά την αντικειμενικότητα των ειδησιογραφικών πηγών, η λύση δεν είναι τόσο προφανής και έγινε έρευνα για το πως θα επιλεγούν οι πηγές. Η διαδικασία περιγράφεται παρακάτω.

Επιλογή ειδησιογραφικών πηγών και Klout Score

Για να επιλεγούν οι καλύτερες ειδησιογραφικές πηγές θα χρειαστεί να βρεθεί μια μετρική που επιτρέπει την αντικειμενική σύγκριση μεταξύ τους. Θεωρητικά θα μπορούσαν να υπάρξουν πολλοί τρόποι να αξιολογηθεί μια πηγή (όγκος ειδήσεων, κέρδος, κοινό κλπ.). Το να επιλεγεί όμως κάτι από αυτά θα ήταν αρκετά αυθαίρετο. Ωστόσο πολλές φορές ως άνθρωποι επιλέγουμε να κάνουμε «αυθαίρετα» πολύ καθοριστικές κινήσεις. Οπότε αποφασίστηκε να μαζευτούν ειδήσεις από δύο πηγές: μια που θα επιλεγεί αυθαίρετα (επειδή θα την διαλέγαμε σαν χρήστες) και μία που θα επιλεγεί με κάποιο αυστηρό κριτήριο που στηρίζεται ακαδημαϊκά. Ως αυθαίρετη πηγή επιλέχθηκε το Yahoo Finance², ένας από τους δημοφιλέστερους ιστότοπους για χρηματοοικονομικά. Για την περίπτωση της δεύτερης πηγής θα χρησιμοποιηθεί το Klout Score [Rao15], ως μετρική επιλογής.

Το Klout Score είναι ένα σύστημα βαθμολογίας επιρροής (influence scoring system), που αναθέτει βαθμούς σε 750 εκατομμύρια χρήστες σε 9 διαφορετικά μέσα κοινωνικής δικτύωσης, σε καθημερινή βάση. Στην δημοσίευσή του [Rao15] προτείνεται ένα ιεραρχικό πλαίσιο για να μετρείται η βαθμολογία επιρροής για κάθε χρήστη, χρησιμοποιώντας πληροφορία για το χρήστη από διάφορα δίκτυα και κοινότητες. Οπότε το Klout Score θα χρησιμοποιηθεί για να μετρηθεί η επιρροή που έχουν διάφορες ειδησιογραφικές πηγές στο Twitter.

Συνεπώς σκοπός είναι να βρεθούν οι πηγές με τη μεγαλύτερη επιρροή. Ως συνέπεια θα επιλεγούν οι «μεγαλύτερες» πηγές στο χώρο των χρηματοοικονομικών και μετά θα εφαρμοστεί το Klout Score. Έτσι συλλέχθηκαν³ οι μεγαλύτερες πηγές με βάση την κυκλοφορία

² <https://finance.yahoo.com>

³ σύμφωνα με το <https://www.investopedia.com/>

τους (circulation) και στη συνέχεια υπολογίστηκε το Klout Score για κάθε μία από αυτές. Το αποτέλεσμα φαίνεται στον Πίνακα 3.3.

Klout Score - Biggest News Sites	
News Site	Klout Score
MarketWatch	90
Bloomberg	91
Reuters	98
The Wall Street Journal	99
The Financial Times	95
CNBC	94

Πίνακας 3.3: Klout Score για τις μεγαλύτερες σε κυκλοφορία ειδησεογραφικές πηγές.

Έτσι, το μέσο με την μεγαλύτερη επιρροή, σύμφωνα με το Klout Score, είναι η The Wall Street Journal⁴ και αυτή αποτέλεσε τη δεύτερη πηγή για τα δεδομένα που συλλέξαμε.

Συλλογή ειδησεογραφικών τίτλων

Αφού έγινε η επιλογή των ειδησεογραφικών πηγών, θα πρέπει με κάποιο τρόπο να συλλεχθούν οι τίτλοι ειδήσεων, που αφορούν το Bitcoin, για 100 ημέρες. Για τον σκοπό αυτό αναπτύχθηκε λογισμικό, στο πλαίσιο της διπλωματικής εργασίας, το οποίο πραγματοποιούσε αυτοματοποιημένες αναζητήσεις στο Google με την καθεμία από τις ειδησεογραφικές πηγές μαζί με τον όρο «Bitcoin». Παράδειγμα τέτοιας αναζήτησης ήταν «*The Wall Street Journal Bitcoin*».

Επόμενο βήμα της συλλογής των τίτλων, αφού είχαν γίνει αυτοματοποιημένα οι αναζητήσεις, ήταν να αποθηκευτούν σε κάποιο αρχείο. Αυτό έγινε με ενσωματωμένο εργαλείο (plug-in) του φυλλομετρητή Google Chrome⁵. Συγκεκριμένα το εργαλείο είχε προγραμματιστεί να εντοπίζει τους τίτλους των ειδήσεων και τους αποθήκευε σε αρχείο τύπου CSV (Comma-Separated Values). Έτσι δημιουργήθηκε η βάση δεδομένων με τους τίτλους ειδήσεων για 100 μέρες, παράδειγμα της οποίας φαίνεται στον Πίνακα 3.4.

Εξαγωγή sentiment

Το τελευταίο βήμα είναι η εξαγωγή του συναισθήματος από τους τίτλους ειδήσεων. Όπως αναφέρθηκε και παραπάνω στη συγκεκριμένη περίπτωση αυτό προκύπτει από τον άνθρωπο (sentiment master). Η κωδικοποίηση γίνεται στο πεδίο τιμών {0, 0.5, 1}, όπου 0 είναι το

⁴ <https://www.wsj.com/>

⁵ <https://www.google.com/chrome/>

αρνητικό συναίσθημα, 0.5 το ουδέτερο και 1 το θετικό. Για παράδειγμα ο τίτλος «SEC warns bitcoin, cryptocurrency investors at risk» θα έχει αποτίμηση συναισθήματος ίση με 0. Ενώ ο τίτλος «Lightning Labs CEO: We are back to a 'bitcoin, not blockchain' world» θα έχει αντίστοιχη αποτίμηση ίση με 1. Αφού ανατεθεί το συναίσθημα με τον παραπάνω τρόπο και στους δύο τίτλους για κάθε ημέρα, στην συνέχεια θα βγει το ολικό της ημέρας ως μέσος όρος των επιμέρους. Έτσι, επειδή οι πηγές είναι δύο, οι πιθανοί μέσοι όροι που μπορεί να προκύψουν ανήκουν στο σύνολο $\{0, 0.25, 0.5, 0.75, 1\}$. Παράδειγμα της συμπληρωμένης, κατά τα παραπάνω, βάσης δεδομένων φαίνεται στον πίνακα 3.5.

Αφού ολοκληρώθηκε η παραπάνω διαδικασία, όσον αφορά το αναδρομικό νευρωνικό δίκτυο, οι τίτλοι των ειδήσεων δεν χρειάζονται πλέον (αλλά μόνο η τιμή που προκύπτει από την ανάλυση συναισθήματος). Οπότε η τελική μορφή της συλλογής δεδομένων φαίνεται στον Πίνακα 3.6.

Bitcoin Sentiment Dataset		
Date	Bitcoin	Sentiment
1/12/2017	10859.56	0.25
2/11/2017	10895.01	0.75
3/11/2017	11180.89	1
4/11/2017	11616.85	0.25
5/11/2017	11696.06	0.5
⋮	⋮	⋮

Πίνακας 3.6: Συλλογή δεδομένων με την τιμή του Bitcoin και την αποτίμηση του συναισθήματος

3.2.3 Google Trends

Σε αυτό το επίπεδο θα προσθέσουμε στη συλλογή δεδομένων το Google Trend της ημέρας για τον όρο αναζήτησης «Bitcoin». Τα Google Trends αναλύθηκαν και εξηγήθηκαν στο Κεφάλαιο 2. Έτσι τα δεδομένα πλέον παίρνουν την μορφή που φαίνεται στον Πίνακα 3.7.

Bitcoin Sentiment Trends Dataset			
Date	Bitcoin	Sentiment	Google Trend
1/12/2017	10859.56	0.25	37
2/12/2017	10895.01	0.75	29
3/12/2017	11180.89	1	29
4/12/2017	11616.85	0.25	34
5/12/2017	11696.06	0.5	34
⋮	⋮	⋮	⋮

Πίνακας 3.7: Συλλογή δεδομένων με την τιμή του Bitcoin, την αποτίμηση συναισθήματος και το Google Trend της ημέρας

3.2.4 Tweets

Σε αυτό το επίπεδο θα αντικατασταθεί η στήλη του Google Trend με αυτή που θα δείχνει τον αριθμό των ημερήσιων tweet που εμπεριέχουν το «#Bitcoin» (σε χιλιάδες). Η συλλογή των αριθμών των Tweet έγινε από την ιστοσελίδα BitInfoCharts⁶ που παρέχει διάφορα δωρεάν δεδομένα που αφορούν το Bitcoin. Η νέα συλλογή δεδομένων φαίνεται στον Πίνακα 3.8.

Bitcoin Sentiment Tweets Dataset			
Date	Bitcoin	Sentiment	Tweets
1/12/2017	10859.56	0.25	105017
2/12/2017	10895.01	0.75	88292
3/12/2017	11180.89	1	83022
4/12/2017	11616.85	0.25	97945
5/12/2017	11696.06	0.5	103919
⋮	⋮	⋮	⋮

Πίνακας 3.8: Συλλογή δεδομένων με την τιμή του Bitcoin, την αποτίμηση συναισθήματος και τον ημερήσιο αριθμό tweets που εμπεριέχουν τον όρο «#Bitcoin»

3.2.5 Τελική συλλογή δεδομένων

Το πιο σύνθετο αναδρομικό νευρωνικό δίκτυο που θα δοκιμαστεί σε αυτήν την εργασία είναι αυτό που θα εκπαιδευτεί στην τελική συλλογή δεδομένων. Σε αυτήν την περίπτωση

⁶ <https://bitinfocharts.com/>

έχουν συνδυαστεί όλες οι παραπάνω πηγές πληροφορίας και έχει προκύψει η συγκεκριμένη συλλογή δεδομένων, παράδειγμα της οποίας φαίνεται στον Πίνακα 3.9. Η συνολική συλλογή δεδομένων περιλαμβάνει 100 ημέρες, κάθε μία για την οποία υπάρχει: η τιμή του Bitcoin σε δολάρια, η αποτίμηση συναισθήματος της ειδησεογραφίας στο σύνολο $\{0, 0.25, 0.5, 0.75, 1\}$, το Google Trend για τον όρο «Bitcoin» κανονικοποιημένο στην κλίμακα $[0, 100]$ και ο αριθμός των Tweets που περιέχουν τον όρο «#Bitcoin», σε χιλιάδες.

Bitcoin Sentiment Trends Tweets Dataset				
Date	Bitcoin	Sentiment	Google Trend	Tweets
1/12/2017	10859.56	0.25	37	105017
2/12/2017	10895.01	0.75	29	88292
3/12/2017	11180.89	1	29	83022
4/12/2017	11616.85	0.25	34	97945
5/12/2017	11696.06	0.5	34	103919
⋮	⋮	⋮	⋮	⋮

Πίνακας 3.9: Η συνολική συλλογή δεδομένων

3.3 Χρήση τεχνικών ανάλυσης συναισθήματος στην ειδησιογραφία της ημέρας

Σε αυτήν την ενότητα, στις ουσία δεν θα δημιουργηθεί κάποια καινούργια συλλογή δεδομένων αλλά θα αλλάξει αυτή της Ενότητας 3.2.2. Συγκεκριμένα, στην Ενότητα 3.2.2 η αποτίμηση του συναισθήματος είχε πραγματοποιηθεί από τον άνθρωπο ως sentiment master. Στην παρούσα ενότητα η αποτίμηση συναισθήματος θα προκύψει από αλγόριθμο μηχανικής μάθησης και τεχνική ανάλυσης συναισθήματος. Με πιο απλά λόγια, θα εκπαιδευτεί ένας αλγόριθμος μηχανικής μάθησης και μετά κάθε τίτλος ειδήσεων θα δοθεί στον αλγόριθμο, ο οποίος θα κάνει και την τελική αποτίμηση.

3.3.1 Μηχανική Μάθηση στην Ανάλυση Συναισθήματος

Η μηχανική μάθηση χρησιμοποιείται ιδιαίτερος σε εφαρμογές ανάλυσης συναισθήματος. Στην παρούσα εργασία θα χρησιμοποιηθεί η βιβλιοθήκη TextBlob [Lori14]. Η TextBlob είναι μία Python βιβλιοθήκη για επεξεργασία γραπτών δεδομένων. Παρέχει ένα απλό API που χρησιμοποιείται για διάφορες διεργασίες επεξεργασίας φυσικής γλώσσας (natural language processing – NLP) όπως η κατηγοριοποίηση της ομιλίας, η εξαγωγή φράσεων, η ταξινόμηση, η μετάφραση και η ανάλυση συναισθήματος.

Απλός μπεϋζιανός ταξινομητής

Ο αλγόριθμος ταξινόμησης που θα χρησιμοποιηθεί είναι ο *απλός μπεϋζιανός ταξινομητής* [Murp06]. Όπως προδίδει το όνομα του, αλγόριθμος βασίζεται στην θεωρία πιθανοτήτων και το γνωστό θεώρημα του Bayes [Koch90].

Θεώρημα 3.1. Θεώρημα του Bayes Έστω δύο γεγονότα A και B . Τότε η πιθανότητα να είναι αληθές το A δεδομένου ότι το B είναι αληθές είναι:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (3.1)$$

όπου $P(B|A)$ είναι η πιθανότητα το B να είναι αληθές δεδομένου ότι το A είναι αληθές, $P(A)$ η πιθανότητα το A να είναι αληθές και $P(B)$ η πιθανότητα το B να είναι αληθές.

Συνοψίζοντας λοιπόν, ένας τέτοιος αλγόριθμος εκπαιδεύτηκε σε μια μεγάλη συλλογή δεδομένων εκπαίδευσης από το πανεπιστήμιο του Michigan. Στην συνέχεια στον αλγόριθμο δόθηκε ένας πίνακας παρόμοιος με τον Πίνακα 3.4 και ο αλγόριθμος για κάθε έναν από τους τίτλους ειδήσεων έγραφε στις στήλες Sentiment 1 και Sentiment 2 τις αντίστοιχες αποτιμήσεις συναισθήματος που έκανε. Τέλος βγήκε ο μέσος όρος των δύο αποτιμήσεων και προέκυψε ένας καινούριος πίνακας αντίστοιχος του Πίνακα 3.5.

3.4 Δυαδική μεταβολή της τιμής του Bitcoin

Το μοντέλο μηχανικής μάθησης θα εκπαιδευτεί ώστε να προβλέπει αν η τιμή του Bitcoin την επόμενη μέρα θα ανέβει ή θα κατέβει. Συγκεκριμένα αν η τιμή σήμερα είναι μεγαλύτερη από χθες, αυτό συμβολίζεται με 1, ενώ αν η τιμή σήμερα είναι μικρότερη ή ίση με χθες αυτό συμβολίζεται με 0. Δηλαδή,

$$up_down(price) = \begin{cases} 1 & \text{if } price_t > price_{t-1} \\ 0 & \text{if } price_t \leq price_{t-1} \end{cases}$$

Daily Headlines Dataset				
Date	Source 1	Headline 1	Source 2	Headline 2
04/01/2018	Yahoo Finance	SEC warns bitcoin, cryptocurrency investors at risk	Wall Street Journal	Bitcoin: The Rise of the Regulators
05/01/2018	Yahoo Finance	Bitcoin Rises to Break Above \$16000, and Another Alt-Coin Is Soaring	Wall Street Journal	Ripple Steals Bitcoin's Thunder, Surges 1135% in a Month
::	::	::	::	::
05/04/2018	Yahoo Finance	Lightning Labs CEO: We are back to a 'bitcoin, not blockchain' world	Wall Street Journal	South Korea Detains Cryptocurrency Executives on Embezzlement Allegations

Πίνακας 3.4: Παράδειγμα του συνόλου των ειδησεογραφικών τίτλων.

Daily Headlines Dataset							
Date	Source 1	Headline 1	Sentiment 1	Source 2	Headline 2	Sentiment 2	Total Sentiment
04/01/2018	Yahoo Finance	SEC warns bitcoin...	0	Wall Street Journal	Bitcoin: The Rise...	0.5	0.25
05/01/2018	Yahoo Finance	Bitcoin Rises...	1	Wall Street Journal	Ripple Steals...	0	0.5
:	:	:	:	:	:	:	:
05/04/2018	Yahoo Finance	Lightning Labs...	1	Wall Street Journal	South Korea...	0	0.5

Πίνακας 3.5: Παράδειγμα του συνόλου των ειδησεογραφικών τίτλων, εμπλουτισμένου με ανάλυση συναισθήματος

Κεφάλαιο 4

Υλοποίηση, εργαλεία και αποτελέσματα

Σε αυτό το κεφάλαιο θα αναλυθούν τα εργαλεία και οι μέθοδοι που χρησιμοποιήθηκαν για να υλοποιηθούν τα αναδρομικά δίκτυα. Θα εξηγηθεί γιατί χρησιμοποιήθηκαν αναδρομικά νευρωνικά δίκτυα Elman και Jordan και όχι Long Short Term Memory – LSTM, ένα άλλο επίσης δημοφιλές μοντέλο. Θα γίνει αναφορά στην γλώσσα προγραμματισμού που χρησιμοποιήθηκε, στο προγραμματιστικό περιβάλλον, στις βιβλιοθήκες που επιλέχθηκαν καθώς ποιες και γιατί απορρίφθηκαν. Τέλος θα παρουσιαστούν και εξηγηθούν τα αποτελέσματα των πειραμάτων.

4.1 Επιλογή Νευρωνικού Δικτύου

Όπως εξηγήθηκε στο προηγούμενο κεφάλαιο, η χρήση ενός αναδρομικού νευρωνικού δικτύου, είναι απαραίτητη. Αυτό γιατί κορμός του προβλήματος που καλείται να λύσει αυτή η διπλωματική εργασία είναι η πρόβλεψη της τιμής του Bitcoin, το οποίο είναι υποσύνολο του μεγαλύτερου προβλήματος που ονομάζεται πρόβλεψη χρονοσειράς. Όπως λοιπόν αναλύθηκε στο προηγούμενο κεφάλαιο, η δομή των αναδρομικών νευρωνικών δικτύων είναι τέτοια, λόγω των διασυνδέσεων αναδρομής, που έχουν ικανότητα «μνήμης» και αυτό τα καθιστά ειδικά για τα προβλήματα πρόβλεψης χρονοσειράς.

Ένα πολύ δημοφιλές νευρωνικό που ανήκει στην κατηγορία των αναδρομικών νευρωνικών δικτύων, είναι το Long Short Term Memory – LSTM. Για την ακρίβεια, τα LSTM είναι μονάδες (κόμβοι) και ο συνδυασμός τέτοιων κόμβων δίνει ένα αναδρομικό νευρωνικό δίκτυο το οποίο ονομάζεται LSTM δίκτυο. Μια LSTM μονάδα συνήθως αποτελείται από το κελί, την πύλη εισόδου, την πύλη εξόδου και την πύλη απώλειας μνήμης. Το κελί θυμάται τιμές ενός απροσδιορίστου αριθμού κύκλων και οι τρεις πύλες καθορίζουν την ροή της πληροφορίας μέσα και έξω από το κελί. Όπως αναφέρθηκε, τα συγκεκριμένα δίκτυα είναι πολύ κατάλληλα για ταξινόμηση και πρόβλεψη βασισμένη σε δεδομένα χρονοσειράς, καθώς σε μια χρονοσειρά μπορεί να υπάρχουν «κενά» αγνώστου μήκους μεταξύ σημαντικών γεγονότων. Τέλος τα LSTM έχουν αποδειχθεί πετυχημένα σε περιοχές όπως: έλεγχος ρομπότ [Maye08], αναγνώριση φωνής [Grav05, Fern07, Grav13], εκμάθηση ρυθμού [Gers02],

σύνθεση μουσικής [Eck02], εκμάθηση γραμματικής [Schm02], αναγνώριση χειρόγραφου κειμένου [Grav08] πρόβλεψη χρονοσειράς [Schm05] και άλλα.

Είναι ξεκάθαρο ότι πρόκειται για ένα επιτυχημένο μοντέλο. Έτσι όταν έγινε η πρώτη προσέγγιση του προβλήματος η επιλογή ενός δικτύου LSTM έμοιαζε πολύ καλή ιδέα. Ωστόσο από την αρχή της εφαρμογής ενός LSTM μοντέλου παρατηρήθηκε ότι, παρά την αρχιτεκτονική και την επιλογή των υπερπαραμέτρων, το LSTM μοντέλο παρουσίαζε πολλά προβλήματα υπερπροσαρμογής. Έγινε εκτενής αναζήτηση στις πιθανές αρχιτεκτονικές του μοντέλου και εν συνεχεία έγινε εξαντλητική αναζήτηση για την εύρεση των υπερπαραμέτρων. Επίσης για την δημιουργία του χρησιμοποιήθηκε η βιβλιοθήκη keras [Chol18] ενώ παράλληλα σχεδιάστηκε και από την αρχή σε απλό κώδικα. Σε καμία από αυτές τις περιπτώσεις δεν ήταν εφικτό να αποφευχθεί η υπερπροσαρμογή.

Τελικά η έρευνα για την λύση του προβλήματος της υπερπροσαρμογής έδειξε ότι το πρόβλημα ήταν ο αριθμός των δεδομένων. Όπως αναφέρθηκε στο Κεφάλαιο 3, το μήκος της εξεταζόμενης χρονοσειράς είναι 100 τιμές (τρεις ημερολογιακοί μήνες περίπου). Ωστόσο, το δίκτυο LSTM προορίζεται για προβλήματα που τα lookback/time-steps είναι τάξης μεγέθους 10^3 [Hoch97], ενώ στα απλά αναδρομικά νευρωνικά δίκτυα είναι 10^2 . Προφανώς λοιπόν σε ένα πρόβλημα όπως της συγκεκριμένης εργασίας, με δεδομένα μιας τάξης μεγέθους μικρότερα από το αναμενόμενο, είναι λογικό να υπάρχει υπερπροσαρμογή. Άρα τελικά πρέπει να χρησιμοποιηθούν απλά αναδρομικά νευρωνικά δίκτυα, όπως τα δίκτυα Elman και Jordan που αναλύθηκαν στο προηγούμενο Κεφάλαιο.

4.2 Εργαλεία Υλοποίησης

Σε συνέχεια της θεωρητικής εξήγησης για τα μοντέλα που επιλέχθηκαν, ακολουθεί η περιγραφή και θεμελίωση των εργαλείων που θα χρησιμοποιηθούν για την υλοποίηση τους.

4.2.1 Γλώσσα προγραμματισμού

Ως γλώσσα προγραμματισμού επιλέχθηκε η Python¹. Η Python τα τελευταία χρόνια έχει γίνει από τις δημοφιλέστερες γλώσσες για μηχανική μάθηση. Μερικοί από τους κυριότερους λόγους που η Python έχει αποκτήσει τέτοια δημοσιότητα στον χώρο της μηχανικής μάθησης συνοψίζονται παρακάτω:

- Η Python προσφέρει μεγάλη ποικιλία βιβλιοθηκών με τις οποίες μπορεί κανείς με σχετική ευκολία να φτιάξει και να δοκιμάσει πρωτότυπα, κόμα και σε έναν απλό οικιακό

¹ <https://www.python.org/>

υπολογιστή. Παραδείγματα τέτοιων βιβλιοθηκών είναι `scikit-learn`², `keras`³, `tensorflow`⁴.

- Χρήση APIs. Η Python προσφέρει μια σειρά από APIs που μπορούν να βοηθήσουν στην συλλογή διάφορων δεδομένων που θα χρησιμοποιηθούν αργότερα στην μηχανική μάθηση.
- Η Python είναι μία γλώσσα προγραμματισμού που είναι κατάλληλη για διαχείριση δεδομένων. Με βιβλιοθήκες όπως τα `pandas`⁵, `NumPy`⁶, `SciPy`⁷ μπορεί κανείς να διαχειριστεί εύκολα δεδομένα και να κάνει υπολογισμούς.
- Η δομή της συγκεκριμένης είναι αρκετά εύκολη που την κάνει πολύ καλή για γρήγορη υλοποίηση ιδεών. Έχει πολύ εύκολη σύνταξη σε σχέση με άλλες (C, C++, Java) κάτι που την κάνει εξαιρετική για πρωτυποποίηση.

4.2.2 Βιβλιοθήκη Μηχανικής Μάθησης

Για να υλοποιηθεί το αναδρομικό νευρωνικό δίκτυο, θα χρησιμοποιηθεί μια βιβλιοθήκη που να υποστηρίζει την υλοποίηση του. Ο λόγος που δεν προτιμάται να γραφτεί το αναδρομικό νευρωνικό δίκτυο από την αρχή σε απλό κώδικα, είναι γιατί υπάρχουν βιβλιοθήκες που είναι βελτιστοποιημένες για να υλοποιούν τέτοια μοντέλα, καθώς έχουν προκύψει μετά από χρόνια σχεδιασμού και δοκιμών.

Πρώτη επιλογή ήταν να χρησιμοποιηθεί η βιβλιοθήκη `keras` [Chol18], μια βιβλιοθήκη για μοντέλα βαθιάς μάθησης. Το `keras` είναι μια πολύ δημοφιλής βιβλιοθήκη με πολλές αναφορές σε δημοσιεύσεις και πολλά project έχουν στηθεί εξολοκλήρου πάνω της. Επίσης υποστηρίζει την σχεδίαση αναδρομικών νευρωνικών δικτύων. Ωστόσο όταν χρησιμοποιήθηκε για πρώτη φορά στο παρόν πρόβλημα παρουσιάστηκε πρόβλημα υπερπροσαρμογής. Εν συνεχεία έγινε εκτενής αναζήτηση της αρχιτεκτονικής και ακολούθησε *εξαντλητική αναζήτηση* (exhaustive search) των υπερπαραμέτρων, αλλά το πρόβλημα της υπερπροσαρμογής εξακολουθούσε να παραμένει. Να σημειωθεί ότι το `keras` προορίζεται κυρίως για *βαθιά μάθηση* (deep-learning), σε εφαρμογές όπως όραση υπολογιστών και αναγνώριση φωνής όπου τα δεδομένα είναι πολλά (τάξη μεγέθους 10^3 , 10^4 ή και παραπάνω). Επίσης η σχεδίαση στο `keras` δεν είναι εντελώς ελεύθερη καθώς χρησιμοποιούνται έτοιμες μονάδες και αυτό είχε ως αποτέλεσμα τα δεδομένα της παρούσας εργασίας να υφίσταται υπερπροσαρμογή πολύ εύκολα, ακόμα και με το πιο απλό μοντέλο. Συνεπώς το `keras` δεν μπορεί να χρησιμοποιηθεί σε αυτή τη

² <https://www.python.org/>

³ <https://keras.io/>

⁴ <https://www.tensorflow.org/>

⁵ <http://pandas.pydata.org/>

⁶ <http://www.numpy.org/>

⁷ <https://www.scipy.org/>

μελέτη. Μετά από αναζήτηση των πιθανών εργαλείων, και με γνώμονα πάντα την απλότητα του μοντέλου, αποφασίστηκε πως βέλτιστη επιλογή ήταν να χρησιμοποιήσουμε την βιβλιοθήκη Pyrenn [Atab16]. Η βιβλιοθήκη αυτή παρέχει στον χρήστη απόλυτη ελευθερία στην σχεδίαση του μοντέλου ενώ διατηρεί υψηλά επίπεδα υπολογιστικής απόδοσης καθώς είναι σχεδιασμένη για την υλοποίηση αναδρομικών νευρωνικών δικτύων. Αυτό έχει ως αποτέλεσμα να είναι δυνατόν να σχεδιαστεί ένα απλό νευρωνικό δίκτυο (τύπου Elman και Jordan), ώστε να αποφύγουμε την υπερπροσαρμογή, και τελικά να έχουμε ικανοποιητικά αποτελέσματα. Το μειονέκτημα της βιβλιοθήκης αυτής είναι ότι δεν έχει πολλές υπερπαραμέτρους αλλά θα σχεδιαστούν διάφορες αρχιτεκτονικές για να υπερκεραστεί αυτό το πρόβλημα.

4.3 Πειράματα και Αποτελέσματα

4.3.1 Πειραματική Διαδικασία

Τα μοντέλα που έχουν αναφερθεί στο Κεφάλαιο 1 ονομάστηκαν και εκπαιδεύτηκαν με την σειρά ως εξής:

1. **Bitcoin.** Αυτό είναι το πιο απλό μοντέλο με έναν νευρώνα εισόδου που δέχεται τις προηγούμενες τιμές του Bitcoin.
2. **Bitcoin & Sentiment.** Το μοντέλο αυτό έχει δύο νευρώνες εισόδου. Έναν για τις προηγούμενες τιμές του Bitcoin και έναν για την αποτίμηση του συναισθήματος της ειδησεογραφίας της ημέρας, όπως προκύπτει από τον άνθρωπο (sentiment master).
3. **Bitcoin & Sentiment & Trends.** Το μοντέλο αυτό έχει τρεις νευρώνες στην είσοδο. Οι δύο είναι όπως πριν και ο τρίτος δέχεται τον αριθμό του Google Trend των προηγούμενων ημερών για τον όρο αναζήτησης Bitcoin.
4. **Bitcoin & Sentiment & Tweets.** Παρόμοιο με το παραπάνω, με την διαφορά ότι ο τρίτος νευρώνας αντί να δέχεται το Google Trend δέχεται τον αριθμό των Tweets με hashtag #Bitcoin.
5. **Bitcoin & Sentiment & Trends & Tweets.** Το μοντέλο αυτό έχει τέσσερις νευρώνες στην είσοδο και συνδυάζει όλα τα προηγούμενα μοντέλα.
6. **Bitcoin & Sentiment Automated.** Το τελευταίο μοντέλο έχει δύο νευρώνες εισόδου. Έναν για τις προηγούμενες τιμές του Bitcoin και έναν για την αποτίμηση συναισθήματος της ειδησεογραφίας της ημέρας που πραγματοποιείται από τον αλγόριθμο ανάλυσης συναισθήματος (απλός μεϋζιανός ταξινομητής, όπως αναφέρθηκε στο Κεφάλαιο 3)

Στην Εικόνα 4.1 φαίνεται η σχηματική αναπαράσταση των παραπάνω μοντέλων. Για την υλοποίηση των μοντέλων, όπως εξηγήθηκε στις παραπάνω υποενότητες, θα χρησιμοποιηθούν δίκτυα Elman και Jordan ενώ σαν βιβλιοθήκη θα χρησιμοποιηθεί η pyrenn. Επίσης στην παραπάνω υποενότητα έγινε αναφορά στον λόγο χρήσης αυτής της βιβλιοθήκης που είναι η απλότητα (ελευθερία) σχεδίασης του μοντέλου. Γι' αυτό το λόγο, οι υπερπαραμέτροι, όπου ελέγχθηκαν, είναι λίγοι καθώς περισσότερη έμφαση δόθηκε στην αρχιτεκτονική του μοντέλου. Το πρώτο βήμα είναι να καθοριστεί το πλήθος κρυφών επιπέδων. Πειραματικά παρατηρήθηκε πως μοντέλα με μόνο ένα κρυφό επίπεδο είχαν πολύ χειρότερη απόδοση σε σχέση με μοντέλα δύο κρυφών επιπέδων. Τρίτο κρυφό επίπεδο οδηγούσε σε μοντέλα με μεγάλο ποσοστό υπερπροσαρμογής οπότε ο αριθμός κρυφών επιπέδων ορίστηκε να είναι δυο. Στον Πίνακα 4.1 φαίνεται το πλήθος των νευρώνων καθώς και το εύρος των υπερπαραμέτρων που αναζητήθηκαν.

Αναζήτηση Πλέγματος	
Νευρώνες πρώτου κρυφού επιπέδου	[1, 5]
Νευρώνες δεύτερου κρυφού επιπέδου	[1, 5]
Χρονική καθυστέρηση πρώτου κρυφού επιπέδου	[1, 5]
Χρονική καθυστέρηση δεύτερου κρυφού επιπέδου	[1, 5]
Χρονική καθυστέρηση επιπέδου εξόδου	[1, 5]
Σταθερό βήμα εκπαίδευσης	10, 20, 30

Πίνακας 4.1: Εύρος τιμών στην αναζήτηση πλέγματος.

Φυσικά οι νευρώνες εισόδου εξαρτώνται από το αντίστοιχο μοντέλο ενώ στην έξοδο υπάρχει ένας νευρώνας (η πρόβλεψη της τιμής). Το σταθερό βήμα εκπαίδευσης χρησιμοποιείται στον αλγόριθμο εκπαίδευσης που είναι ο Levenberg Marquardt [More78]. Στον Πίνακα 4.2 φαίνονται οι υπερπαραμέτροι που παρέμειναν σταθερές κατά την διάρκεια της εκπαίδευσης.

Σταθερές Παράμετροι	
Μέγιστος αριθμός εποχών	100
Επιθυμητό σφάλμα	10^{-5}
Παράγοντας εκπαίδευσης	3
Αλγόριθμος εκπαίδευσης	Levenberg Marquardt

Πίνακας 4.2: Παράμετροι που έμειναν σταθεροί κατά την διαδικασία εκπαίδευσης.

Το επιθυμητό σφάλμα υποδηλώνει ότι αν σε μια εποχή το σφάλμα είναι μικρότερο η

ίσο του επιθυμητού, η εκπαίδευση σταματάει. Διαφορετικά θα σταματήσει όταν ο αριθμός των εποχών φτάσει τον μέγιστο. Ο παράγοντας εκπαίδευσης καθορίζει το πόσο αλλάζει το σταθερό βήμα εκπαίδευσης ενώ αυτοί οι δύο παράμετροι μαζί καθορίζουν τον ρυθμό εκπαίδευσης.

4.3.2 Αποτελέσματα

Η πρόβλεψη της μελλοντικής τιμής ανήκει στο διάστημα 0, 1 με την τιμή 0 να σημαίνει ότι η τιμή του Bitcoin ήταν μικρότερη από την χθεσινή, ενώ η τιμή 1 σημαίνει ότι η τιμή ήταν μεγαλύτερη από την χθεσινή. Με άλλα λόγια 0 σημαίνει "ανέβηκε" και 1 σημαίνει "έπεσε". Η μέτρηση της απόδοσης του συστήματος θα γίνεται μετρώντας το ποσοστό των επιτυχημένων προβλέψεων.

Μετά την εκπαίδευση, τα μοντέλα στο σύνολο ελέγχου είχαν τις αποδόσεις που φαίνονται στον Πίνακα 4.3 και στην Εικόνα 4.2.

Αποδόσεις Μοντέλων	
Bitcoin	84.21%
Bitcoin & Sentiment	84.21%
Bitcoin & Sentiment & Trends	75.86%
Bitcoin & Sentiment & Tweets	82.76%
Bitcoin & Sentiment & Trends & Tweets	79.31%
Bitcoin & Sentiment Automated	75.86%

Πίνακας 4.3: Τελική απόδοση των μοντέλων στο σύνολο ελέγχου

Την καλύτερη απόδοση είχαν τα μοντέλα Bitcoin και Bitcoin & Sentiment με ποσοστό επιτυχίας 84.21% στο σύνολο ελέγχου και 83.33% και 88.46% αντίστοιχα στο σύνολο εκπαίδευσης. Παρατηρώντας τα αποτελέσματα πρέπει να γίνουν οι εξής παρατηρήσεις:

- Τα ποσοστά επιτυχίας των Bitcoin και Bitcoin & Sentiment (84.21%) είναι πολύ υψηλά. Ταυτόχρονα με τα αντιστοίχως υψηλά επίπεδα στο σύνολο εκπαίδευσης είναι βέβαιο πως δεν πρόκειται για περίπτωση υπερπροσαρμογής. Συνεπώς, τα αποτελέσματα είναι ενθαρρυντικά για μελλοντική συνέχεια της έρευνας.
- Τα πιο απλά μοντέλα έδωσαν και τις καλύτερες αποδόσεις. Αυτό πιθανόν να υποδηλώνει, ότι η πληροφορία των υπόλοιπων μεταβλητών, σε συχνότητα μιας μέρας, εισάγει πιο πολύ θόρυβο από ότι χρήσιμη πληροφορία.

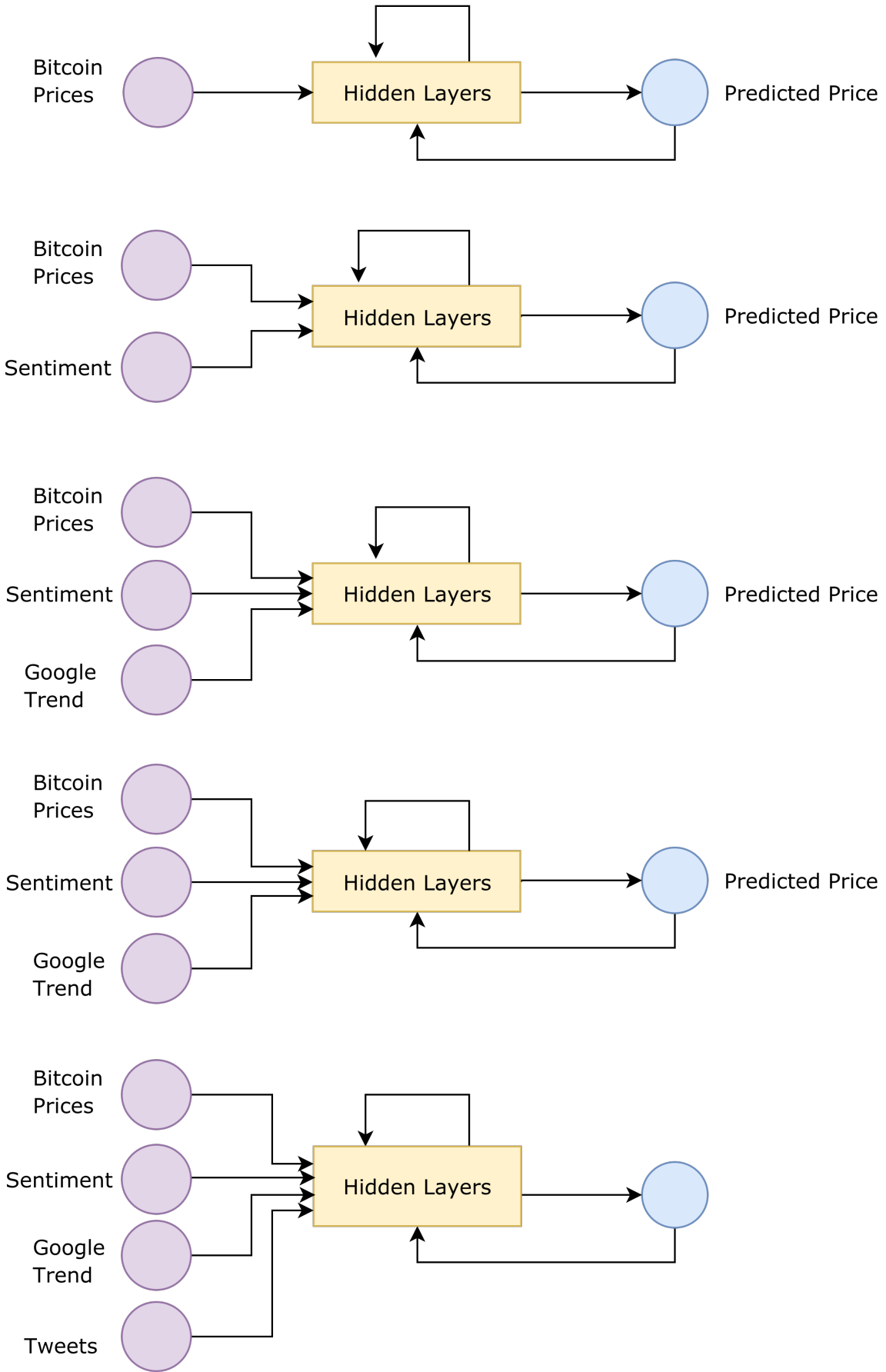
Στους Πίνακες 4.4-4.5 φαίνονται οι τελικές αρχιτεκτονικές των δύο μοντέλων που επέτυχαν τα μεγαλύτερα ποσοστά.

Bitcoin model	
Νευρώνες πρώτου κρυφού επιπέδου	2
Νευρώνες δεύτερου κρυφού επιπέδου	2
Χρονική καθυστέρηση πρώτου κρυφού επιπέδου	4
Χρονική καθυστέρηση δεύτερου κρυφού επιπέδου	–
Χρονική καθυστέρηση επιπέδου εξόδου	–
Σταθερό βήμα εκπαίδευσης	20

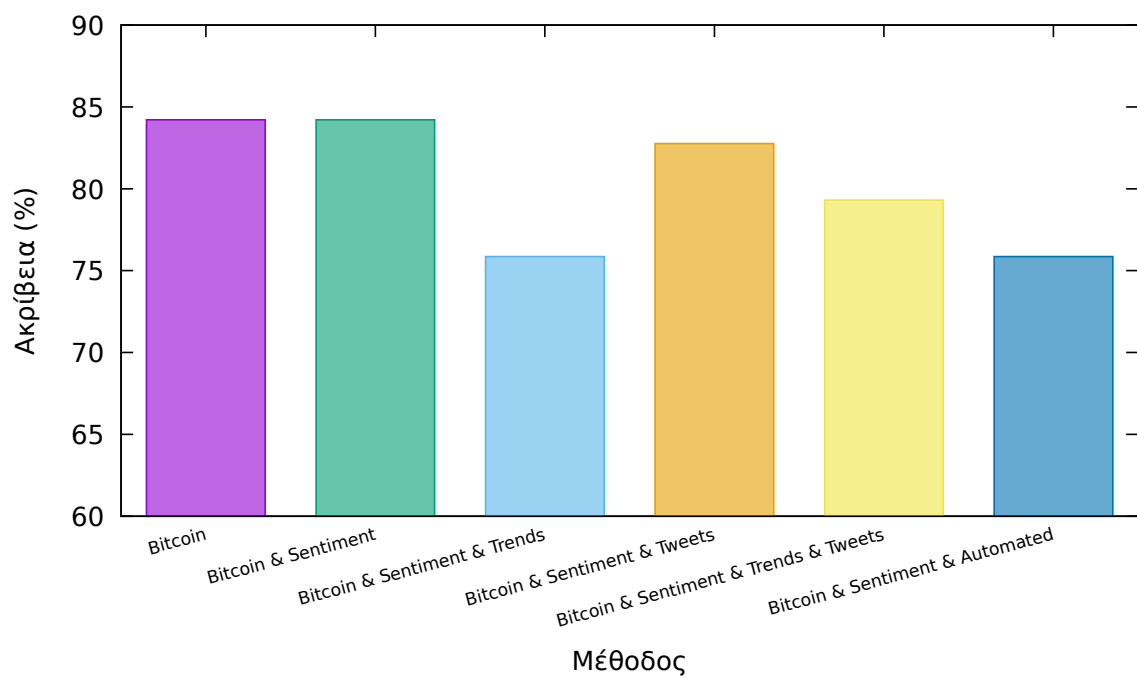
Πίνακας 4.4: Βέλτιστη αρχιτεκτονική για το μοντέλο Bitcoin

Bitcoin & Sentiment model	
Νευρώνες πρώτου κρυφού επιπέδου	1
Νευρώνες δεύτερου κρυφού επιπέδου	3
Χρονική καθυστέρηση πρώτου κρυφού επιπέδου	2
Χρονική καθυστέρηση δεύτερου κρυφού επιπέδου	1
Χρονική καθυστέρηση επιπέδου εξόδου	–
Σταθερό βήμα εκπαίδευσης	20

Πίνακας 4.5: Βέλτιστη αρχιτεκτονική για το μοντέλο Bitcoin & Sentiment



Σχήμα 4.1: Γραφική αναπαράσταση των παραπάνω μοντέλων



Σχήμα 4.2: Απόδοση μοντέλων στο σύνολο ελέγχου

Κεφάλαιο 5

Επίλογος

5.1 Συνεισφορά

Η παρούσα εργασία ασχολήθηκε με την εφαρμογή τεχνικών μηχανικής μάθησης όσο και στον χώρο των χρηματοοικονομικών και ειδικότερα της αγοράς κρυπτονομισμάτων. Πιο συγκεκριμένα, δημιουργήθηκαν και προετοιμάστηκαν σχετικές συλλογές δεδομένων, με την καινοτομία να έγκειται στον τρόπο δημιουργίας αυτών των συλλογών. Στην ουσία αυτό που έγινε είναι μια αυτοματοποιημένη διαδικασία που κατασκευάζει μια συλλογή δεδομένων ανάλογα με έναν όρο προτίμησης. Στο Κεφάλαιο 3 αναλύθηκαν τα βήματα δημιουργίας των δεδομένων. Στην συγκεκριμένη διαδικασία ο όρος προς μελέτη είναι «Bitcoin». Συνολικά, οι συλλογές δεδομένων που δημιουργήθηκαν ήταν: προηγούμενες τιμές του Bitcoin, αποτίμηση συναισθήματος της καθημερινής ειδησεογραφίας για το Bitcoin, ημερήσια tweets για το hashtag #Bitcoin), ημερήσια Google Trends για τον όρο «Bitcoin» και αποτίμηση συναισθήματος (από αλγόριθμο) της καθημερινής ειδησεογραφίας. Σε περίπτωση λοιπόν που αλλάξει ο όρος αναζήτησης, για παράδειγμα έστω ότι χρειάζεται να μελετηθεί η μετοχή της Apple, τότε απλώς θα επαναληφθεί η ίδια διαδικασία με την διαφορά ότι θα αντικατασταθεί ο όρος «Bitcoin» με τον όρο «APPL», με την υπόλοιπη διαδικασία να μένει ως έχει. Επίσης ακόμα και εκτός χρηματοοικονομικής πρόβλεψης, η δομή δημιουργίας της συγκεκριμένης συλλογής δεδομένων μπορεί να χρησιμοποιηθεί και για συλλογή δεδομένων διαφορετικού ενδιαφέροντος, όπως για παράδειγμα ο όρος «climate change» για την περίπτωση της κλιματικής αλλαγής. Αυτό λοιπόν που στην ουσία έχει δημιουργηθεί είναι μια γενική μέθοδος συλλογής δεδομένων που βασίζονται σε έναν όρο αναζήτησης, με τα δεδομένα αυτά να μπορούν να χρησιμοποιηθούν από αλγόριθμους μηχανικής μάθησης για χρηματοοικονομική πρόβλεψη αλλά και άλλες εφαρμογές. Η συνεισφορά από χρηματοοικονομικής πλευράς εστιάζεται στον χώρο της αγοράς κρυπτονομισμάτων. Η αγορά κρυπτονομισμάτων είναι μια σχετικά καινούρια αγορά (των τελευταίων χρόνων) με αποτέλεσμα να υπάρχει πολύς χώρος για έρευνα και πειραματισμό. Η ιδέα πως οι προηγούμενες τιμές του Bitcoin σε συνδυασμό με την ειδησεογραφία και τους δείκτες των Google Trends και tweets μπορούν να προβλέψουν την μελλοντική τιμή του είναι καινοτόμα και προσφέρει έδαφος για μελλοντική

έρευνα και προβληματισμό στον χώρο της αγοράς κρυπτονομισμάτων.

5.2 Μελλοντική Έρευνα

Τα αποτελέσματα έδειξαν πως η πρόβλεψη της αγοράς σε ημερήσια συχνότητα, και συγκεκριμένα η πρόβλεψη της τιμής του Bitcoin την επόμενη ημέρα, δεν είναι εύκολη. Μάλλον θα έλεγε κανείς πως μια αρχιτεκτονική ίδια ή παρόμοια με αυτή που παρουσιάζεται σε αυτή την διπλωματική μοιάζει αδύνατη να λύσει το παραπάνω πρόβλημα. Πιθανή αιτία για τον προηγούμενο ισχυρισμό βρίσκεται στα δεδομένα και ότι ως «επόμενη μέρα» στην αγορά του Bitcoin θεωρείται η τιμή στις δώδεκα τα ξημερώματα (00:00). Κάτι το οποίο είναι χαοτικό και κανείς δεν εγγυάται ότι η διάθεση, το συναίσθημα, τα Google Trends και οι υπόλοιποι δείκτες που μελετούνται σε αυτήν την εργασία έχουν κάποια υποχρέωση να αποτυπωθούν στην τιμή του Bitcoin συγκεκριμένα στις 00:00. Βέβαια σε αυτό το σημείο πρέπει να αναφερθεί πως κάποιοι δείκτες όπως το Google Trend δεν υπάρχουν σε μικρότερη συχνότητα από την ημέρα ενώ κάποιοι άλλοι δείκτες είναι εκ φύσεως ημερήσιοι, όπως «ημερήσιος αριθμός tweets με #Bitcoin». Με εξαίρεση τα Google Trends, οι υπόλοιποι δείκτες μπορούν να τροποποιηθούν ώστε να παρέχουν πληροφορία για πολύ πιο μικρή συχνότητα. Θα μπορούσε για παράδειγμα τα δεδομένα να αφορούν την ειδησεογραφία και τον αριθμό tweets για κάθε ώρα ή για κάθε λεπτό αντί για κάθε μέρα. Το εμπόδιο για αυτού του τύπου την ανάλυση είναι ότι οι περισσότερες πλατφόρμες που παρέχουν δυνατότητα πρόσβασης στα δεδομένα τους μέσω APIs, όταν πρόκειται για τέτοιες μικρές συχνότητες, τα APIs δεν διατίθενται δωρεάν. Αυτός ήταν και ο βασικότερος λόγος που σε αυτήν την εργασία επιλέχθηκε η συχνότητα της μιας μέρας. Σε μελλοντική έρευνα λοιπόν μια πρόταση είναι να συλλεχθούν ίδιου τύπου δεδομένα όμως με συχνότητα μιας ώρας ή ενός λεπτού και να επαναληφθεί η ίδια διαδικασία. Να τονιστεί πως όλη η διάταξη και η αρχιτεκτονική των μοντέλων είναι εντελώς ανεξάρτητη από την συχνότητα των δεδομένων οπότε και για τα καινούρια πειράματα το μόνο που απαιτείται είναι η ανανέωση των δεδομένων, λ.χ. ανά ώρα ή ανά λεπτό.

Βιβλιογραφία

- [Atab16] Dennis Atabay, “pyrenn: First release”, January 2016.
- [Cao03] L. J. Cao and F. E. H. Tay, “Support vector machine with adaptive parameters in financial time series forecasting”, *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1506–1518, Nov 2003.
- [Chen06] Yuehui Chen, Bo Yang and Jiwen Dong, “Time-series prediction using a local linear wavelet neural network”, *Neurocomputing*, vol. 69, no. 4, pp. 449 – 465, 2006.
- [Chol18] François Chollet et al., “Keras: The python deep learning library”, *Astrophysics Source Code Library*, 2018.
- [Crus06] Holk Cruse, “Neural networks as cybernetic systems”, *Brain, minds, and media*. See <http://www.brains-minds-media.org/archive/289>, 2006.
- [Eck02] Douglas Eck and Jürgen Schmidhuber, “Learning the long-term structure of the blues”, in *International Conference on Artificial Neural Networks*, pp. 284–289, Springer, 2002.
- [Elma90] Jeffrey L Elman, “Finding structure in time”, *Cognitive science*, vol. 14, no. 2, pp. 179–211, 1990.
- [Fern07] Santiago Fernández, Alex Graves and Jürgen Schmidhuber, “An application of recurrent neural networks to discriminative keyword spotting”, in *International Conference on Artificial Neural Networks*, pp. 220–229, Springer, 2007.
- [Gers02] Felix A Gers, Nicol N Schraudolph and Jürgen Schmidhuber, “Learning precise timing with LSTM recurrent networks”, *Journal of machine learning research*, vol. 3, no. Aug, pp. 115–143, 2002.
- [Gest01] T. Van Gestel, J. A. K. Suykens, D. . Baestaens, A. Lambrechts, G. Lanckriet, B. Vandaele, B. De Moor and J. Vandewalle, “Financial time series prediction using least squares support vector machines within the evidence framework”, *IEEE Transactions on Neural Networks*, vol. 12, no. 4, pp. 809–821, July 2001.

- [Grav05] Alex Graves and Jürgen Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures”, *Neural Networks*, vol. 18, no. 5-6, pp. 602–610, 2005.
- [Grav08] Alex Graves, Marcus Liwicki, Horst Bunke, Jürgen Schmidhuber and Santiago Fernández, “Unconstrained on-line handwriting recognition with recurrent neural networks”, in *Advances in neural information processing systems*, pp. 577–584, 2008.
- [Grav13] Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton, “Speech recognition with deep recurrent neural networks”, in *Acoustics, speech and signal processing (icassp), 2013 IEEE international conference on*, pp. 6645–6649, IEEE, 2013.
- [Hayk94] Simon Haykin, *Neural Networks: A Comprehensive Foundation*, Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 1994.
- [Hoch97] Sepp Hochreiter and Jürgen Schmidhuber, “Long short-term memory”, *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [jKim03] Kyoung jae Kim, “Financial time series forecasting using support vector machines”, *Neurocomputing*, vol. 55, no. 1, pp. 307 – 319, 2003. Support Vector Machines.
- [Jord97] Michael I Jordan, “Serial order: A parallel distributed processing approach”, in *Advances in psychology*, vol. 121, pp. 471–495, Elsevier, 1997.
- [Kars] Saurabh Karsoliya, “Approximating Number of Hidden layer neurons in Multiple Hidden Layer BPNN Architecture”.
- [Ketc17] Nikhil Ketkar, *Stochastic Gradient Descent*, pp. 113–132, Apress, Berkeley, CA, 2017.
- [Koch90] Karl-Rudolf Koch, “Bayes’ theorem”, in *Bayesian Inference with Geodetic Applications*, pp. 4–8, Springer, 1990.
- [Lori14] Steven Loria, P Keen, M Honnibal, R Yankovsky, D Karesh, E Dempsey et al., “Textblob: simplified text processing”, *Secondary TextBlob: Simplified Text Processing*, 2014.
- [Lu09] Chi-Jie Lu, Tian-Shyug Lee and Chih-Chou Chiu, “Financial time series forecasting using independent component analysis and support vector regression”, *Decision Support Systems*, vol. 47, no. 2, pp. 115 – 125, 2009.

- [Maye08] Hermann Mayer, Faustino Gomez, Daan Wierstra, Istvan Nagy, Alois Knoll and Jürgen Schmidhuber, “A system for robotic heart surgery that learns to tie knots using recurrent neural networks”, *Advanced Robotics*, vol. 22, no. 13-14, pp. 1521–1537, 2008.
- [More78] Jorge J Moré, “The Levenberg-Marquardt algorithm: implementation and theory”, in *Numerical analysis*, pp. 105–116, Springer, 1978.
- [Murp06] Kevin P Murphy et al., “Naive bayes classifiers”, *University of British Columbia*, vol. 18, 2006.
- [Naka09] Satoshi Nakamoto, “Bitcoin: A peer-to-peer electronic cash system”, 2009.
- [Pang08] Bo Pang and Lillian Lee, “Opinion Mining and Sentiment Analysis”, *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008.
- [Rao15] Adithya Rao, Nemanja Spasojevic, Zhisheng Li and Trevor Dsouza, “Klout score: Measuring influence across multiple social networks”, in *Big Data (Big Data), 2015 IEEE International Conference on*, pp. 2282–2289, IEEE, 2015.
- [Samu59] A. L. Samuel, “Some Studies in Machine Learning Using the Game of Checkers”, *IBM Journal of Research and Development*, vol. 3, no. 3, pp. 210–229, July 1959.
- [Sapa09] N. I. Sapankevych and R. Sankar, “Time Series Prediction Using Support Vector Machines: A Survey”, *IEEE Computational Intelligence Magazine*, vol. 4, no. 2, pp. 24–38, May 2009.
- [Schm02] Jürgen Schmidhuber, F Gers and Douglas Eck, “Learning nonregular languages: A comparison of simple recurrent networks and LSTM”, *Neural Computation*, vol. 14, no. 9, pp. 2039–2041, 2002.
- [Schm05] Jürgen Schmidhuber, Daan Wierstra and Faustino J Gomez, “Evolino: Hybrid neuroevolution/optimal linear search for sequence prediction”, in *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI)*, 2005.
- [Stat09] D Stathakis, “How many hidden layers and nodes?”, *International Journal of Remote Sensing*, vol. 30, no. 8, pp. 2133–2147, 2009.
- [Werb90] P. J. Werbos, “Backpropagation through time: what it does and how to do it”, *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, Oct 1990.

[WILL89] RONALD J. WILLIAMS and DAVID ZIPSER, “Experimental Analysis of the Real-time Recurrent Learning Algorithm”, *Connection Science*, vol. 1, no. 1, pp. 87–111, 1989.