



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΠΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Ανάλυση συναισθήματος σε tweets με σκοπό την πρόβλεψη κίνησης της τιμής του Bitcoin

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΤΟΥ

ΛΕΝΑΚΑΚΗ Κ. ΙΑΣΟΝΑ



Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής ΕΜΠ

Συνεπιβλέπουσα: Παρασκευή Τζούβελη
Μέλος ΕΔΙΠ ΕΜΠ

Αθήνα, Οκτώβριος 2022



Ανάλυση συναισθήματος σε tweets με σκοπό την πρόβλεψη κίνησης της τιμής του Bitcoin

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΛΕΝΑΚΑΚΗ Κ. ΙΑΣΟΝΑ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής ΕΜΠ

Συνεπιβλέπουσα: Παρασκευή Τζούβελη
Μέλος ΕΔΙΠ ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 20η Οκτωβρίου 2022.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Στέφανος Κόλλιας
Καθηγητής ΕΜΠ

.....
Γεώργιος Στάμου
Καθηγητής ΕΜΠ

.....
Αθανάσιος Βουλόδημος
Επίκουρος Καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος 2022



Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.
ΛΕΝΑΚΑΚΗΣ ΙΑΣΟΝΑΣ, 2022.

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.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....
ΛΕΝΑΚΑΚΗΣ ΙΑΣΟΝΑΣ

20 Οκτωβρίου 2022

Περίληψη

Η παρούσα εργασία επικεντρώνεται στην μελέτη της πρόβλεψης της τιμής του Bitcoin, και της εκτίμησης της κίνησης της τιμής, ενός κρυπτονομίσματος το οποίο έχει ραγδαία αύξηση τα τελευταία χρόνια. Ο λόγος που έχει ιδιαίτερο ενδιαφέρον μία τέτοια μελέτη είναι ότι η τιμή αυτού του κρυπτονομίσματος, και γενικότερα των κρυπτονομισμάτων, έχει έντονες αυξομειώσεις.

Για την πρόβλεψη της τιμής του κρυπτονομίσματος γίνεται χρήση ενός συνδυασμού δεδομένων, με μεγαλύτερη βαρύτητα στην εξαγωγή συναισθήματος από Tweets, δηλαδή δημοσιεύσεις στην πλατφόρμα του Twitter. Επίσης, από την συγκεκριμένη πλατφόρμα χρησιμοποιείται η πληροφορία του πλήθους των δημοσιεύσεων, και ακόμα χρησιμοποιούνται τεχνικές πληροφορίες της τιμής του κρυπτονομίσματος, όπως ιστορικό των τιμών. Ακόμα, για την πρόβλεψη της τιμής με τα παραπάνω δεδομένα γίνεται χρήση κλασικών βαθιών νευρωνικών δικτύων για πρόβλεψη χρονοσειρών όπως επαναληπτικά νευρωνικά δίκτυα.

Λέξεις Κλειδιά

Χρονοσειρές, Επεξεργασία Φυσικής Γλώσσας, Κρυπτονομίσματα, Επαναληπτικά Νευρωνικά Δίκτυα, Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης, Φραγμένα Επαναληπτικά Δίκτυα, Τουίτερ, Πρόβλημα Παλινδρόμησης, Πρόβλημα Κατηγοριοποίησης

Abstract

In this diploma thesis, we are trying to work on the prediction of a cryptocurrency's price, Bitcoin, and also the prediction of it's trend, positive or negative. This task is challenging due to the high volatility of cryptocurrencies' prices in general.

In order to achieve our goal, we combined some sources of information, with most important source the sentiment extraction from tweets. We also used other information from Twitter, such as the number of tweets per hour, and also technical information about the price of crypto as historical prices. For the prediction Recurrent Neural Networks were used, specifically, LSTM's and GRU's.

Keywords

Time Series, Natural Language Processing, Cryptocurrencies, RNN, LSTM, GRU, Twitter, Regression Task, Classification Task

Ευχαριστίες

Θα ήθελα καταρχάς να ευχαριστήσω τον καθηγητή κύριο Στέφανο Κόλλια, για την ευκαιρία να εκπονήσω την συγκεκριμένη εργασία στο εργαστήριο Τεχνητής Νοημοσύνης και Συστημάτων Γνώσης του Εθνικού Μετσόβιου Πολυτεχνείου, καθώς και την κυρία Παρασκευή Τζούβελη, μέλος ΕΔΙΠ του ΕΜΠ, για την εξαιρετική συνεργασία, το ενδιαφέρον και την άμεση βοήθεια της όπου χρειάστηκε. Επίσης θα ήθελα να ευχαριστήσω τους κυρίους Ανδρέα-Γεώργιο Σταφυλοπάτη και Γιώργο Στάμου οι οποίοι δέχθηκαν να συμμετέχουν στην τριμελή επιτροπή.

Αθήνα, Οκτώβριος 2022

ΛΕΝΑΚΑΚΗΣ ΙΑΣΟΝΑΣ

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	5
Πρόλογος	13
1 Εισαγωγή	15
1.1 Αντικείμενο της διπλωματικής	15
1.2 Σχετικές Εργασίες	16
1.3 Οργάνωση του τόμου	17
I Θεωρητικό Μέρος	19
2 Θεωρητικό υπόβαθρο	21
2.1 Bitcoin: Ιστορική Αναδρομή - Τεχνικά Στοιχεία - Ιδιαιτερότητες	21
2.1.1 Τεχνικά Στοιχεία	22
2.2 Ανάλυση συναισθήματος	23
2.3 Επεξεργασία Φυσικής Γλώσσας - Κλασικά προβλήματα	23
2.3.1 Το πρόβλημα της εξαγωγής συναισθήματος από κείμενο	24
2.3.2 Κλασικές μέθοδοι εξαγωγής συναισθήματος από κείμενο	25
2.4 Κυλιώμενα Παράθυρα σε Χρονοσειρές	26
3 Νευρωνικά Δίκτυα	29
3.1 Εισαγωγή στη Μηχανική Μάθηση - Είδη μάθησης	29
3.2 Είδη Νευρωνικών Δικτύων για χρονοσειρές	30
3.2.1 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks-ANN)	30
3.2.2 Επαναληπτικά Νευρωνικά Δίκτυα	31
3.2.3 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης - LSTM	32
3.2.4 Φραγμένα Επαναληπτικά Δίκτυα -GRU	33
3.2.5 Αρχιτεκτονικές Βαθιών Νευρωνικών Δικτύων	34
3.3 Είδη Νευρωνικών Δικτύων για επεξεργασία φυσικής γλώσσας	35
3.3.1 Η μέθοδος των transformers στην εξαγωγή συναισθήματος	35
3.3.2 Word Embeddings	36
3.3.3 BERT	37

3.3.4 RoBERTa	39
II Πρακτικό Μέρος	41
4 Πειραματικό Μέρος	43
4.1 Dataset	43
4.1.1 Twitter - Χαρακτηριστικά	43
4.1.2 Dataset με Tweets	44
4.1.3 Dataset για fine tuning του BERT	45
4.1.4 Historical prices dataset	46
4.1.5 Εξαγωγή τελικού Dataset	46
4.2 Ροή Εργασιών	47
5 Πειραματικά Αποτελέσματα	49
5.1 Μετρικές Αξιολόγησης	49
5.1.1 Μέσο τετραγωνικό σφάλμα MSE	49
5.1.2 Ρίζα μέσου τετραγωνικού σφάλματος RMSE	50
5.1.3 Μέσο Απόλυτο Σφάλμα MAE	50
5.1.4 Μέσο Απόλυτο Ποσοστιαίο Σφάλμα MAPE	50
5.2 Αποτελέσματα και αξιολόγηση Μοντέλων και επιλογών δεδομένων	50
5.2.1 Συσχέτιση των δεδομένων με την τιμή κλεισίματος	50
5.2.2 Πειραματικά Αποτελέσματα	52
III Επίλογος	63
6 Επίλογος	65
6.0.1 Μελλοντικές Επεκτάσεις	65
Βιβλιογραφία	73

Κατάλογος Σχημάτων

2.1	Χρονικό παράθυρο σε χρονοσειρά[1]	27
3.1	Απλός Νευρώνας[2]	30
3.2	Απλός Νευρώνας[3]	31
3.3	δίκτυο RNN[4]	32
3.4	LSTM CELL[5]	32
3.5	GRU CELL[6]	34
3.6	Δομή Transformers[7]	36
3.7	Ο μηχανισμός του Self-Attention[7]	36
3.8	Η δομή του BERT[8]	38
4.1	Τελική μορφή του Dataset με συναίσθημα	44
4.2	Εξαγωγή μέσου συναισθήματος ανά ώρα μαζί με πλήθος tweets	45
4.3	Dataset με τεχνικούς δείκτες της μετοχής	46
4.4	Dataset με συναίσθημα και πλήθος tweets για κάθε ώρα της μέρας	47
4.5	Dataset με συναίσθημα και πλήθος tweets για κάθε ώρα της μέρας (μέρος 2)	47
4.6	Διάγραμμα ροής ενεργειών μέχρι την τελική πρόβλεψη με χρήση RoBERTa για εξαγωγή συναισθήματος και χρήση πλήθους των tweets	48
5.1	Γραφική παράσταση τιμής κλεισίματος και συναισθήματος	51
5.2	Γραφική παράσταση πλήθους tweets συναρτήση του χρόνου	52
5.3	Γραφική παράσταση τιμής κλεισίματος συναρτήση του χρόνου	52
5.4	Ακρίβεια μοντέλων πίνακα 5.1 συναρτήσε ημερών	54
5.5	MAPE μοντέλων πίνακα 5.2 συναρτήσε ημερών	55
5.6	Διάγραμμα πραγματικών και προβλεπόμενων τιμών για 7 μέρες μετά με συναίσθημα 3 ημερών RoBERTa σε GRU	56
5.7	Διάγραμμα πραγματικών και προβλεπόμενων τιμών για 7 μέρες μετά με συνδυασμό συναισθήματος, πλήθους και τεχνικών δεικτών RoBERTa σε GRU	56
5.8	Διάγραμμα πραγματικών και προβλεπόμενων τιμών για 7 μέρες μετά με συναίσθημα 3 ημερών Vader σε GRU	57
5.9	Διάγραμμα πραγματικών και προβλεπόμενων τιμών για 7 μέρες μετά με συνδυασμό συναισθήματος με Vader, πλήθους και τεχνικών δεικτών σε GRU	58
5.10	Accuracy μοντέλων πίνακα 5.3 συναρτήσε ημερών	58
5.11	MAPE μοντέλων πίνακα 5.4 συναρτήσε ημερών	59
5.12	Δομή δικτύου πρόβλεψης με μονάδες GRU	59

5.13	Διάγραμμα πραγματικών και προβλεπμένων τιμών 7 ημερών με χρήση συναισθήματος με Vader 3 ημερών σε LSTM	60
5.14	Διάγραμμα πραγματικών και προβλεπμένων τιμών 7 ημερών με χρήση συνδυασμού συναισθήματος 3 ημερών με Vader, πλήθους δημοσιεύσεων και τεχνικών δεικτών σε LSTM	61
5.15	Διάγραμμα του accuracy σε συνάρτηση των ημερών πρόβλεψης με Vader σε LSTM	61
5.16	Διάγραμμα του MAPE score σε συνάρτηση των ημερών πρόβλεψης με Vader σε LSTM	62

Κατάλογος Πινάκων

5.1	Πίνακας με ποσοστά ακριβείας με χρήση RoBERTa σε GRU	53
5.2	Πίνακας με την μετρική MAPE με χρήση RoBERTa σε GRU	53
5.3	Πίνακας με την μετρική Accuracy με χρήση Vader σε GRU	56
5.4	Πίνακας με την μετρική MAPE με χρήση Vader σε GRU	57
5.5	Πίνακας με την μετρική Accuracy με χρήση Vader σε LSTM	60
5.6	Πίνακας με την μετρική MAPE με χρήση Vader σε LSTM	60

Πρόλογος

Η παρούσα διπλωματική εργασία εκπονήθηκε στο εργαστήριο Τεχνητής Νοημοσύνης και Συστημάτων Γνώσης της σχολής ΗΜΜΥ του ΕΜΠ και σαν βάση είχε το μάθημα των Νευρωνικών Δικτύων του 9ου εξαμήνου. Το αντικείμενο της εργασίας προσπαθεί να προσεγγύσει ένα ρεαλιστικό πρόβλημα, την πρόβλεψη της τιμής ενός κρυπτονομίσματος, το οποίο με την ραγδαία εξέλιξη του κλάδου της βαθιάς μάθησης αποκτά νέες προοπτικές στην προσέγγιση του. Θεωρώ σημαντικό πλεονέκτημα της σχολής μας, ότι μας δίνεται η δυνατότητα στα πλαίσια αυτής της εργασίας να ασχοληθούμε με τόσο απτά πράγματα από τον πραγματικό κόσμο.

Κεφάλαιο 1

Εισαγωγή

Η μελέτη της αγοράς και η πρόβλεψη της κίνησης της απασχολεί τους ανθρώπους από την δημιουργία του χρηματιστηρίου και έπειτα, καθώς η εμπιστοσύνη για το μέλλον ήταν σημαντικό κομμάτι για την ευημερία. Λόγω της ανάγκης για εμπιστοσύνη και πίστη για το μέλλον, η γενικότερη εντύπωση που επικρατεί γύρω από ένα προϊόν στην αγορά έχει άμεση σχέση με το πως θα εξελιχθεί η τιμή του, και η ποσοτικοποίηση αυτής της εντύπωσης μέσω μοντέλων για εξαγωγή συναισθήματος θα ήταν ένα σημαντικό εργαλείο για την εκτίμηση της μελλοντικής του πορείας.

Όπως ακριβώς λοιπόν με το χρηματιστήριο και τα προϊόντα, έτσι και στην εκτίμηση της τιμής των κρυπτονομισμάτων, είναι πολύ χρήσιμο να έχουμε εικόνα, και αυτή η εικόνα να μπορεί να ποσοτικοποιηθεί, σχετικά με το γενικότερο αίσθημα που επικρατεί στην κοινή γνώμη γύρω από το κρυπτονόμισμα. Επίσης, τα κρυπτονομίσματα παρουσιάζουν ιδιαίτερο ενδιαφέρον, καθώς η τιμή τους παρουσιάζει πολύ πιο έντονες και συχνές διακυμάνσεις σε σχέση με πιο παραδοσιακές μετοχές, λόγω και του ότι η εμπιστοσύνη του κόσμου προς αυτά ακόμα δοκιμάζεται. Ένα ακόμα στοιχείο των κρυπτονομισμάτων που τα καθιστά ενδιαφέροντα, είναι ότι φέρνουν την επανάσταση στον τρόπο που γίνονται οι συναλλαγές και έτσι πιθανόν στο μέλλον να είναι το κυρίαρχο μέσο συναλλαγής στην αγορά.

Άλλοι παράγοντες που μπορούν να καθορίσουν την τιμή ενός κρυπτονομίσματος, πέρα από την κοινή γνώμη που μεταφράζεται σε μεγαλύτερη ή μικρότερη ζήτηση, είναι και τεχνικά τους χαρακτηριστικά όπως η δυσκολία στην δημιουργία τους, καθώς συγκεκριμένα κρυπτονομίσματα όπως το Bitcoin έχουν γίνει εξαιρετικά περίπλοκα στην δημιουργία τους.

1.1 Αντικείμενο της διπλωματικής

Η παρούσα διπλωματική εργασία πραγματεύεται το πρόβλημα της εκτίμησης της κίνησης της τιμής του κρυπτονομίσματος Bitcoin, αλλά και της πρόβλεψης ακριβούς τιμής, μέσω της χρήσης διαφόρων πηγών πληροφορίας, με κύρια πηγή το συναίσθημα που επικρατεί στην πλατφόρμα του Twitter .

Η πλατφόρμα του Twitter έχει συγκεκριμένα χαρακτηριστικά, όπως το συγκεκριμένο μέγιστο μήκος μιας δημοσίευσης, που την κάνουν πιο συνοπτική, άμεση και πυκνή σε πληροφορία με αποτέλεσμα ο κόσμος να στρέφεται σε αυτήν και να περιέχει αρκετή πληροφορία σχετικά με το συναίσθημα που επικρατεί. Επιπλέον, για την εξαγωγή του συναισθήματος από αυτή την πλατφόρμα γίνεται δοκιμή και προηγμένων τεχνικών εξαγωγής συναισθήμα-

τος, όπως το μοντέλο των μετασχηματιστών (Transformers), τα οποία έχουν αναπτυχθεί τα τελευταία χρόνια και αποτελούν αιχμή στον κλάδο της επεξεργασίας φυσικής γλώσσας. Επίσης πέρα από αυτά τα μοντέλα γίνεται χρήση και πιο παραδοσιακών και θεμελιωμένων τεχνικών εξαγωγής συναισθήματος, και τελικά γίνεται σύγκριση αυτών των μεθόδων ως προς την αποτελεσματικότητά τους σε σχέση με την πρόβλεψη της τιμής του Bitcoin.

Άλλες πηγές πληροφορίας που χρησιμοποιούνται είναι το ιστορικό των τιμών του κρυπτονομίσματος, αλλά και άλλες τιμές όπως τιμή ανοίγματος και όγκος συναλλαγών, ενώ κατά την διάρκεια των πειραμάτων προέκυψε σαν χρήσιμη πληροφορία η χρήση του πλήθους των δημοσιεύσεων στο Twitter.

1.2 Σχετικές Εργασίες

Ασφαλώς, το ζήτημα της πρόβλεψης της τιμής κάποιας μετοχής, αν και το Bitcoin δεν είναι μετοχή αλλά η διακύμανση της τιμής του έχει αρκετά κοινά με τις μετοχές, έχει απασχολήσει στο παρελθόν και συνεχίζει να απασχολεί πολλούς ερευνητές.

Έρευνες που εξετάζουν διακύμανση τιμών βασισμένες σε διάφορα χαρακτηριστικά παρουσιάζουν μεγάλο ενδιαφέρον καθώς και έρευνες που έχουν γίνει σχετικά με την εξαγωγή συναισθήματος από πλατφόρμες κοινωνικής δικτύωσης. Αναλυτικότερα, οι Jianfeng Si, Arjun Mukherjee στην εργασία τους [9], προσπαθούν να δημιουργήσουν μία πρόβλεψη για την τιμή του Bitcoin χρησιμοποιώντας ανάλυση συναισθήματος με μαθηματικές μεθόδους επάνω σε διαφορετικά θέματα, δηλαδή την γνώμη του κοινού ως προς διάφορες εκδοχές ενός προϊόντος όπως πχ την τιμή, την ποιότητα και άλλα. Μία επίσης σημαντική εργασία είναι [10] του Pratikkumar Prajapati ο οποίος εφάρμοσε και αυτός τεχνικές εξαγωγής συναισθήματος από κείμενο, αλλά αυτή την φορά επάνω σε νέα από ειδησεογραφικούς ιστότοπους, τα οποία θεωρούνται και πιο έγκυρα, και έκανε και αυτός ανάλυση συναισθήματος με στατικές μεθόδους ανάλυσης, με σκοπό την πρόβλεψη της τιμής. Άλλες προσεγγίσεις είχαν σαν στοιχεία πιο τεχνικά δεδομένα, δηλαδή δεν χρησιμοποίησαν συναισθημα, και φάνηκε ότι και αυτό είχε αποτέλεσμα καθώς οι τιμές των προηγούμενων ημερών περιέχουν σημαντική πληροφορία για μια μετοχή [11].

Σημείο αναφοράς για την παρούσα εργασία αποτελεί η εργασία των Jacques Vella Critien, Albert Gatt, Joshua Ellul [12] οι οποίοι έκαναν ανάλυση συναισθήματος με την μέθοδο Vader επάνω σε Tweets και στη συνέχεια, δημιουργώντας κάποιο μέσο συναίσθημα ανά ώρα, προσπάθησαν να προβλέψουν την τιμή του Bitcoin ως χρονοσειρά. Επίσης, η εργασία τους ασχολήθηκε και με την πρόβλεψη της τάσης της τιμής, δηλαδή για το αν θα κινηθεί θετικά ή αρνητικά. Παρόλα αυτά δεν έγινε χρήση κάποιων βαθιών νευρωνικών για εξαγωγή συναισθήματος. Επίσης σημαντικό ερέθισμα ήταν και η εργασία των Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, Babita Majhi [13] οι οποίοι εφάρμοσαν τεχνικές ανίχνευσης συναισθήματος στο Twitter με σκοπό την πρόβλεψη της τιμής μετοχών. Σημαντική παρατήρηση σε αυτή την εργασία είναι η ισχυρή συσχέτιση που υπάρχει μεταξύ του γενικού συναισθήματος που επικρατεί στο Twitter και της κίνησης των μετοχών, κυρίως των πιο γνωστών ενώ και εδώ η πληροφορία του συναισθήματος ήταν σε μορφή χρονοσειράς, έχοντας κάποια τιμή για κάθε δεδομένη χρονική στιγμή, ενώ υπήρξαν και άλλοι οι οποίοι ασχολήθηκαν με την εξαγωγή συναισθήματος από το Twitter για αυτό

τον σκοπό[14].

Το κομμάτι της ανάλυσης συναισθήματος σε πλατφόρμες blogging όπως το Twitter έχει ευρεία εφαρμογή πέραν της πρόβλεψης της τιμής μετοχών. Έτσι είχαμε την δυνατότητα να δούμε πως προσεγγίζουν το θέμα και τι πληροφορίες αντλούν από την πλατφόρμα ερευνητές οι οποίοι έψαξαν σε βάθος το κομμάτι της ανάλυσης συναισθήματος και έχουν καταλήξει στο ποιές μέθοδοι εξαγωγής συναισθήματος φέρνουν καλύτερα αποτελέσματα[15][16].

1.3 Οργάνωση του τόμου

Ο παρακάτω τόμος οργανώνεται ως εξής: Πρώτα, στο κεφάλαιο 2 παρουσιάζονται θεωρητικά θέματα του ευρύτερου ενδιαφέροντος για την εργασία, όπως και λεπτομέρειες για το Bitcoin, το πρόβλημα επεξεργασίας φυσικής γλώσσας, αναγνώρισης συναισθήματος και κυλιόμενων παραθύρων σε χρονοσειρές. Στην συνέχεια, στο κεφάλαιο 3 παρουσιάζονται τα νευρωνικά δίκτυα, τα οποία χρησιμοποιούνται τόσο για την εξαγωγή του συναισθήματος όσο και για την πρόβλεψη της τιμής. Στο κεφάλαιο 4 γίνεται εκτενής αναφορά στα επιμέρους σύνολα δεδομένων (Datasets) που χρησιμοποιήθηκαν και στον τρόπο που αυτά επεξεργάστηκαν και συγχωνεύτηκαν για να προκύψει το τελικό Dataset. Στο κεφάλαιο 5 αναφέρονται τα αποτελέσματα των πειραμάτων με τους διάφορους συνδυασμούς δεδομένων, μοντέλων και γίνεται και αναλυτικός σχολιασμός. Τέλος στο κεφάλαιο 6 υπάρχει ο επίλογος όπου αναφέρονται τα τελικά συμπεράσματα από την εργασία και προτάσεις για επέκταση στο μέλλον.

Μέρος I

Θεωρητικό Μέρος

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

Στο κεφάλαιο αυτό παρουσιάζεται το θεωρητικό υπόβαθρο της εργασίας αυτής. Αναλυτικότερα, επικεντρωνόμαστε στα κύρια τεχνικά στοιχεία καθώς και σε λοιπές πληροφορίες για το Bitcoin, σε μεθόδους εξαγωγής συναισθήματος από κείμενο, σε χρονοσειρές, καθώς και τα μοντέλα νευρωνικών δικτύων που χρησιμοποιούνται για τα κομμάτια αυτά.

2.1 Bitcoin: Ιστορική Αναδρομή - Τεχνικά Στοιχεία - Ιδιαιτερότητες

Το Bitcoin είναι ένα κρυπτονόμισμα το οποίο δημιουργήθηκε το 2008 από κάποιο άτομο ή ομάδα που αποκαλείται Shatoshi Nakamoto και το οποίο κρατάει κρυφή την πραγματική του ταυτότητα. Κύρια χαρακτηριστικά των κρυπτονομισμάτων, άρα και του Bitcoin, είναι η τεχνική του Blockchain στην οποία βασίζονται και η οποία τους επιτρέπει να μην υπόκεινται σε κεντρική διαχείριση από τρίτους, αλλά κάθε συναλλαγή να πραγματοποιείται άμεσα μεταξύ δύο μερών, κάτι που δεν συμβαίνει στις τραπεζικές συναλλαγές. Αυτό οφείλεται στο γεγονός ότι κάθε Bitcoin αποτελείται από ένα μπλοκ το οποίο είναι μοναδικό και το οποίο βρίσκεται καταχωρημένο στο δίκτυο για βεβαίωση εγκυρότητας[17]. Το λογισμικό υλοποίησης του Bitcoin είναι σε μορφή ανοιχτού κώδικα, κάτι που σημαίνει ότι καθένας μπορεί να δει την υλοποίηση του και να την χρησιμοποιήσει για δικές του υλοποιήσεις.

Το πρώτο μπλοκ του Bitcoin δημιουργήθηκε στις 3 Ιανουαρίου 2009 και μέσα σε επτά μήνες ο δημιουργός του, Shatoshi Nakamoto ανέφερε ότι είχε δημιουργήσει 1.1 εκατομμύρια Bitcoin, κάτι που σε σημερινά δεδομένα θα ξεπερνούσε τα 22 δισεκατομμύρια δολάρια. Βέβαια στην αρχή της πορείας του το κρυπτονόμισμα δεν είχε κάποια τιμή, καθώς δεν ήταν ανταλλάξιμο, ώσπου το 2010 στην πολιτεία της Φλόριδα, έγινε η πρώτη συναλλαγή πίστας αξίας 25 δολαρίων για περίπου 10000 Bitcoin[18], κάτι που όρισε την αρχική τιμή του Bitcoin. Η καινοτομία του Bitcoin ενέπνευσε και άλλα κρυπτονομίσματα και έτσι το 2011 δημιουργήθηκε το Litecoin (LTC)[19] και μερικά χρόνια αργότερα, το 2015, το Ethereum[20], κάτι που δημιούργησε ανταγωνισμό στην αγορά των κρυπτονομισμάτων.

Η τιμή του Bitcoin έκτοτε ξεκίνησε να ανεβαίνει φτάνοντας για πρώτη φορά την τιμή των 1000 δολαρίων τον Νοέμβριο του 2013. Ένα πρώτο σημείο σταθμός που συνέβαλε στην εκ-

Θετική αύξηση της τιμής, ήταν όταν το 2017 η εταιρεία CME ανακοίνωσε ότι θα υπογράψει δεσμευτικά συμβόλαια μελλοντικής αγοράς των Bitcoin σε συγκεκριμένη τιμή[21]. Αυτό είχε ως αποτέλεσμα η τιμή να φτάσει το Νοέμβριο του ίδιου έτους τις 10000 δολάρια και έως τον Δεκέμβριο να αγγίξει τις 20000. Το bitcoin, όπως και ανάλογες τεχνολογικές καινοτομίες στο παρελθόν, μετά από ένα διάστημα δημιούργησε πολύ μεγαλύτερες προσδοκίες στο κοινό από αυτές που μπορούσε πραγματικά να εκπληρώσει και έτσι η τιμή του ξεκίνησε να παρουσιάζει διακυμάνσεις. Χαρακτηριστικό παράδειγμα είναι ότι την επόμενη χρονιά, στα τέλη του 2018, η τιμή βρέθηκε ξανά λίγο κάτω από τις 4000 δολάρια.

Η επόμενη μεγάλη έκρηξη για το Bitcoin έγινε εν μέσω της πανδημίας του Κορονοϊού καθώς ο κόσμος επικεντρώθηκε σε αυτό σαν ασχολία τις ελεύθερες ώρες του αφού δεν υπήρχαν πολλές διαφορετικές δραστηριότητες. Αυτό οδήγησε σε νέα ιστορικά υψηλά της τιμής, φτάνοντας τον Νοέμβριο του 2021 έως και τις 68000 δολάρια. Όμως αυτή η τρελή άνοδος προκάλεσε παρεμβάσεις και περιορισμούς στις επενδύσεις μεγάλων κεφαλαίων στην αγορά κρυπτονομισμάτων ώστε να μην υπάρξει μελλοντική αστάθεια στην αγορά, με αποτέλεσμα να έρθει νέα μείωση στην τιμή. Τον Αύγουστο του 2022 η τιμή είχε φτάσει ξανά κοντά στα 22000 δολάρια ανά νόμισμα.

2.1.1 Τεχνικά Στοιχεία

Για την δημιουργία και επικύρωση των κρυπτονομισμάτων του Bitcoin απαιτείται μία διαδικασία που ονομάζεται εξόρυξη Bitcoin (mining Bitcoin)[22]. Όμως τι εννοούμε με τον όρο εξόρυξη; Όταν λοιπόν αναφερόμαστε σε εξόρυξη Bitcoin, μιλάμε για μία υπολογιστική διαδικασία, εξαιρετικά πολύπλοκη για τις συνηθισμένες υπολογιστικές δυνατότητες, η οποία δημιουργεί ένα κώδικα κρυπτογραφίας ο οποίος επιβεβαιώνει μία λίστα από έγκυρες συναλλαγές μέσω Bitcoin και επίσης εξασφαλίζει ότι κάποιο κρυπτονόμισμα ανήκει μόνο σε έναν νόμιμο ιδιοκτήτη. Επίσης με την διαδικασία του mining υποστηρίζεται και η βάση δεδομένων του Bitcoin η οποία ονομάζεται Blockchain. Η διαδικασία αυτή είναι τρομερά ενεργοβόρα, καθιστώντας οικονομικά μη βιώσιμη την διαδικασία για πολλούς ενδιαφερόμενους, αλλά και για το περιβάλλον.

Οι συναλλαγές με Bitcoin οργανώνονται σε Blocks όπου κάθε Block περιέχει πληροφορία για την προέλευση του κάθε Bitcoin αλλά και τον προορισμό του σε κάποια συναλλαγή. Για αυτό τον λόγο είναι σημαντική η επιβεβαίωση ενός μπλοκ τόσο πριν όσο και μετά από μία συναλλαγή. Κάθε μπλοκ αποτελείται από κάποιους κρυπτογραφικούς κανόνες, οι οποίοι πρακτικά είναι υπολογιστικά σύνθετες μαθηματικές συναρτήσεις, που κάθε εξορύκτης θέλει να λύσει για να επιβεβαιώσει το Block. Ο πρώτος που θα το καταφέρει έχει μία αμοιβή σε Bitcoin και έτσι προκύπτει το κέρδος από το mining. Η ανάγκη για την ύπαρξη των κρυπτογραφικών κανόνων (hashes) είναι για να μην μπορεί να υπάρξει απάτη από κάποιον που θα θέλει να αποκτήσει πρόσβαση στο μπλοκ και να περάσει στο όνομά του κάποια συναλλαγή που δεν προοριζόταν για αυτόν. Μετά την ενημέρωση ενός block για την συναλλαγή, αυτό αντιγράφεται και σε άλλα-μπλοκ κόμβους και έτσι όλοι είναι ενήμεροι για τις τελευταίες συναλλαγές. Με τον τρόπο αυτό είναι σχεδόν αδύνατο για κάποιον να εξαπατήσει το

σύστημα, καθώς οι κρυπτογράφηση είναι αρκετά ισχυρή για τους διαθέσιμους μέχρι σήμερα πόρους για να σπάσει από κάποιον εξωτερικό παράγοντα, ενώ επίσης η πληροφορία δεν συγκεντρώνεται σε ένα μπλοκ αλλά έχει διαμοιραστεί.

2.2 Ανάλυση συναισθήματος

Το πρόβλημα της εξαγωγής συναισθήματος είναι ένα πρόβλημα που απασχολεί την επιστημονική κοινότητα για αρκετά χρόνια και έχει συναντήσει σημαντική πρόοδο με την ραγδαία εξέλιξη των νευρωνικών δικτύων και της μηχανικής μάθησης γενικότερα.

Το πρόβλημα αυτό, αφορά την εξαγωγή συναισθήματος από κάθε πηγή που μπορεί να είναι διαθέσιμη, όπως ήχος, εικόνα, βίντεο, κείμενο, μουσική καθώς και συνδυασμός των παραπάνω πηγών. Για παράδειγμα, στην εξαγωγή συναισθήματος με νευρωνικά δίκτυα βασισμένα σε εικόνα, πολλές φορές γίνεται προσπάθεια να γίνει εστίαση σε ανθρώπινα χαρακτηριστικά που εκφράζουν συναίσθημα, όπως το πρόσωπο[23], καθώς και την μίξη διαφορετικών πηγών πληροφορίας όπως εικόνες από πρόσωπο και χέρια[24]. Αυτό επιτυγχάνεται βρίσκοντας συγκεκριμένα σημεία του προσώπου τα οποία εκφράζουν συναίσθημα, όπως τα μάτια και το στόμα, και ανάλογα την έκφραση που παίρνουν τους αποδίδεται και το αντίστοιχο συναίσθημα, ενώ αντίστοιχα συμβαίνει και με εικόνες από διαφορετικά μέρη του σώματος. Επίσης σημαντική πρόοδος γίνεται και στο κομμάτι της εξαγωγής συναισθήματος από βίντεο όπου εκεί δημιουργούνται χαρακτηριστικά όχι σε μία στατική εικόνα, αλλά σε διαδοχικές, οι οποίες έχουν χρονική συσχέτιση μεταξύ τους, δηλαδή συνήθως βρίσκονται σε διαδοχικές χρονικές στιγμές. Τις περισσότερες φορές στα βίντεο, τα χαρακτηριστικά αυτά αφορούν κινήσεις ή στάσεις σώματος χαρακτηριστικές ή εκφράσεις χαρακτηριστικές κάποιας συναισθηματικής κατάστασης, οι οποίες πολλές φορές δεν είναι εμφανής σε κάποιον άνθρωπο αλλά είναι ανιχνεύσιμες από κάποιο νευρωνικό δίκτυο.

Ενδιαφέρον επίσης παρουσιάζουν μέθοδοι εξαγωγής συναισθήματος από πηγές όπως η μουσική, όπου εκεί γίνεται ανάλυση σήματος, και αντιστοιχούνται συγκεκριμένες νότες, δηλαδή συχνότητες, και συνδυασμοί τους, οι κλίμακες, σε συγκεκριμένα συναισθήματα. Αυτό, σε συνδυασμό με την εξαγωγή συναισθήματος από τους στίχους ενός τραγουδιού μπορεί να μας δώσει το γενικό συναίσθημα που εκφράζει ένα τραγούδι[25]. Επίσης, ανάλυση συναισθήματος γίνεται και σε κείμενο, και είναι το είδος ανάλυσης συναισθήματος που θα μας απασχολήσει στην παρούσα εργασία. Σε αυτό το πρόβλημα γίνεται εξαγωγή συναισθήματος με βάση θετικές και αρνητικές λέξεις, αλλά και σε πιο προχωρημένα μοντέλα, με βάση των συνδυασμό των λέξεων σε μία πρόταση.

2.3 Επεξεργασία Φυσικής Γλώσσας - Κλασικά προβλήματα

Στην εργασία αυτή, μία κύρια μέθοδος που χρησιμοποιήσαμε για την πρόβλεψη της τιμής του Bitcoin ήταν η εξαγωγή συναισθήματος από κείμενο, ένα πρόβλημα που κατατάσσεται στα προβλήματα επεξεργασίας φυσικής γλώσσας.

Ο τομέας επεξεργασίας φυσικής γλώσσας ή Natural Language Processing (NLP) μαζί με τον κλάδο της υπολογιστικής όρασης είναι οι δύο κλάδοι οι οποίοι έχουν γνωρίσει την μεγαλύτερη και ταχύτερη πρόοδο λόγω της προόδου της μηχανικής μάθησης και των βαθιών

νευρωνικών δικτύων. Κύρια προβλήματα στα οποία αυτός ο κλάδος προσπαθεί να δώσει λύση είναι προβλήματα μετατροπής φωνής σε κείμενο ή κειμένου σε φωνή, δημιουργία κειμένου σχετικό με εικόνα με είσοδο μόνο την εικόνα, εξαγωγή θέματος μιας λέξης, δηλαδή να διώχνουμε την κατάληξη μια λέξης (Lemmatization), μορφολογικός διαχωρισμός των λέξεων, κάτι που χρησιμεύει σε γλώσσες όπως η ελληνική που υπάρχουν πολλές διαφορετικές μορφές μιας λέξης πχ(ανοίγω, άνοιξα, ανοιχτά κτλ) και επίσης υπάρχει και το stemming το οποίο κρατάει το θέμα μιας λέξης όπως το Lemmatization αλλά όχι σε μορφή λεξικού αλλά βασισμένο σε κανόνες. [26]

Στον τομέα αυτό ανάγονται και προβλήματα όπως η σημασία μιας λέξης μέσα σε κείμενο, η σημασία μιας πρότασης βάση του συνδιασμού των λέξεων, το συναίσθημα μιας πρότασης, δηλαδή αν χαρακτηρίζεται θετική ή αρνητική (το πρόβλημα που μελετήσαμε), ο διαχωρισμός της έννοιας λέξεων με παραπάνω από μία σημασία κτλ.

2.3.1 Το πρόβλημα της εξαγωγής συναισθήματος από κείμενο

Το πρόβλημα της εξαγωγής συναισθήματος από κείμενο έχει ως σκοπό να κατατάξει προτάσεις ως θετικές, αρνητικές ή ουδέτερες ή να τους δώσει μία δεκαδική τιμή από -1 έως 1 με την τιμή -1 να συμβολίζει εντελώς αρνητική πρόταση και 1 εντελώς θετική. Υπάρχουν επίσης και παραλλαγές αυτού του προβλήματος όπου υπάρχει μία αντίστοιχη κλίμακα από -1 έως 1 για συγκεκριμένα συναισθήματα όπως χαρά, λύπη και θυμός.

Το συγκεκριμένο πρόβλημα συναντάται σε πολλές περιπτώσεις στον σύγχρονο κόσμο. Βέβαια, υπάρχουν διάφορες πιο συγκεκριμένες περιπτώσεις στις οποίες βρίσκεται εφαρμογή και οι οποίες χρειάζονται μία περαιτέρω εξειδίκευση πέραν της κλασικής ανάλυσης συναισθήματος ως προς την ένταση[27]. Μία τέτοια περίπτωση είναι η ανάλυση πρόθεσης (intent analysis). Σε αυτή την περίπτωση, σκοπός είναι να γίνει κατανοητό από ένα κείμενο ποιός είναι ο σκοπός του, όπως για παράδειγμα αν κάνει μία πρόταση, ένα παράπονο, αναφέρει μια πληροφορία ή ασκεί κριτική. Αυτού του είδους η ανάλυση είναι χρήσιμη σε οποιονδήποτε απευθύνεται στην αγορά και θέλει να ξέρει τι συζητιέται γύρω από το όνομα του και για ποιον σκοπό, όπως μεγάλες εταιρείες.

Ένα ακόμα πιο εξειδικευμένο πρόβλημα της ανάλυσης συναισθήματος σε κείμενο, είναι η ανάλυση περιεχομένου από τα συμφραζόμενα (contextual semantic search)[28]. Εδώ κύριος σκοπός είναι η ανακάλυψη του θέματος γύρω από το οποίο περιστρέφεται μία πρόταση, όπως αν αφορά προϊόν, υπηρεσία, εξυπηρέτηση πελατών, τιμές και άλλα. Η εκπαίδευση για αυτό το πρόβλημα γίνεται με την μέθοδο της επιβεπόμενης μάθησης, καθώς δίνονται στο σύστημα δεδομένα που έχουν χαρακτηριστεί ως σχετικά με κάποιο συγκεκριμένο θέμα, πχ τιμή, και έπειτα προσπαθεί να προβλέψει άλλες προτάσεις γύρω από ποιο θέμα κινούνται χωρίς να γνωρίζει την πραγματική τους κατάσταση. Η πολυπλοκότητα του προβλήματος έγκυται και στο γεγονός ότι κάποια πρόταση μπορεί να περιστρέφεται γύρω από παραπάνω από ένα πρόβλημα, με αποτέλεσμα να δημιουργείται σύγχυση γύρω από το πλήθος των θεμάτων που υπάρχουν.

Αφού γίνει αυτή η κατηγοριοποίηση των προτάσεων, οι ενδιαφερόμενοι διεξάγουν ανάλυ-

ση συναισθήματος επάνω στις ίδιες προτάσεις και ανάλογα με τα αποτελέσματα εστιάζουν εκεί που βλέπουν ότι το κοινό δεν έχει μείνει ικανοποιημένο. Αυτά τα εργαλεία είναι πολύ χρήσιμα σε εταιρείες κολοσσοί οι οποίοι λαμβάνουν χιλιάδες μηνύματα με πολλούς διαφορετικούς τρόπους από τους χρήστες τους και δεν έχουν την δυνατότητα να τα επεξεργαστούν όλα με φυσικό τρόπο.

2.3.2 Κλασικές μέθοδοι εξαγωγής συναισθήματος από κείμενο

Σε αυτή την ενότητα θα εξετάσουμε τις διαφορετικές τεχνικές με τις οποίες μπορεί να γίνει εξαγωγή συναισθήματος από κάποιο κείμενο. Πριν την μεγάλη ανάπτυξη των βαθιών νευρωνικών δικτύων είχαν δημιουργηθεί τεχνικές βασισμένες σε κανόνες (rule based) για εξαγωγή συναισθήματος από κείμενο[29]. Ενδιαφέρον έχουν οι περιπτώσεις μοντέλων εξαγωγής συναισθήματος όπως τα LIWC, ANEW, the General Inquirer, SentiWordNet, τα οποία κατά βάση χρησιμοποιούσαν λεξικά με θετικές και αρνητικές λέξεις και στην συνέχεια μετρούσαν πόσες από αυτές περιέχονταν σε κάθε πρόταση για να αποφανθούν αν μία πρόταση είναι θετική ή αρνητική.

Η τεχνική της γλωσσικής εξερεύνησης και μετρήματος λέξεων (linguistic inquiry and word count)(LIWC)[30][31] προσπαθεί να εξάγει συμπεράσματα σχετικά με το συναίσθημα σε κάποιο κείμενο βασιζόμενη σε έρευνες πολλών ετών πάνω στον κλάδο της ψυχολογίας. Η τεχνική έχει σαν κύριο χαρακτηριστικό πάνω από 100 λεξικά που το κάθε ένα έχει χαρακτηριστεί ως θετικό ή αρνητικό, με λέξεις που περιέχουν ή όχι συναίσθημα, ενώ υπάρχουν και κατηγορίες όπως θυμός, χαρά και άλλα. Έτσι, για κάθε κείμενο που δίνεται στο σύστημα, το μοντέλο μετράει πόσες από τις λέξεις που περιέχονται στο κείμενο βρίσκονται σε κάποιο από αυτά τα λεξικά, και έτσι προκύπτει ένα ποσοστό συμμετοχής του κειμένου σε κάθε λεξικό (πόσες από τις συνολικές λέξεις του κειμένου ανήκουν σε κάποιο λεξικό). Επειδή τα λεξικά έχουν χαρακτηριστεί το καθένα ως προς μία κατηγορία, στο τέλος προκύπτει ένα ποσοστό για κάθε κατηγορία που χαρακτηρίζει το κείμενο, πχ χαρά 5%, θετικό συναίσθημα 25% κλπ.

Μία άλλη μέθοδος που βασίζεται και αυτή σε μέτρηση λέξεων είναι η Affective Norms for English Words (ANEW)[32], η οποία βασίζεται σε λεξικά τα οποία περιέχουν σκορ σχετικά με 3 συναισθηματικές κατηγορίες. Αυτές οι κατηγορίες είναι η ένταση του συναισθήματος (valence), το πόσο επιδραστικό είναι το συναίσθημα (arousal) και πόσο κυρίαρχο είναι (dominance). Αντίστοιχα λοιπόν, και αυτή η τεχνική μετράει τι ποσοστό των λέξεων του κειμένου ανήκουν σε κάθε μία από τις τρεις κατηγορίες και έτσι προκύπτει ένα ποσοστό συμμετοχής για κάθε κατηγορία. Στην ίδια λογική κινούνται και το General Inquirer και το SentiWordNet τα οποία είναι συναισθηματικά λεξικά βασισμένα σε έρευνες από το κλάδο της ψυχολογίας.

Vader

Ανάμεσα σε όλες αυτές τις στατικές τεχνικές εξαγωγής συναισθήματος, στατικές αφού το σκορ κάθε λέξης είναι ήδη καθορισμένο από το λεξικό, έχει ξεχωρίσει η τεχνική Vader καθώς έχει πετύχει καλύτερα σκορ σε σημαντικά προβλήματα[29].

Σημαντικές επιτυχίες του παραπάνω μοντέλου είναι ότι έχει καταφέρει να συνδυάσει συναισθηματικά λεξικά από διάφορες πηγές, βασισμένα και σε συναίσθημα στα μέσα κοινωνικής δικτύωσης, ενώ επίσης εφάρμοσε βασικούς κανόνες γραμματικής οι οποίοι δείχνουν

την διάθεση που υπάρχει στο κείμενο και συνδυαστικά, τα αποτελέσματα είναι είναι καλύτερα συγκριτικά με τα προαναφερθέντα μοντέλα. Σημαντικό και για την παρούσα εργασία είναι ότι το εργαλείο Vader έχει δημιουργηθεί κατάλληλα για να αναγνωρίζει το συναίσθημα σε κείμενα που συναντώνται σε social media[33], τα οποία φέρουν ιδιαίτερα χαρακτηριστικά όπως ιδιοματισμούς αργκό, λανθασμένη γραμματική και λέξεις, ενώ γενικεύει καλά και σε άλλες περιπτώσεις.

Ο τρόπος λειτουργίας του χαρακτηρίζεται από τα εξής βήματα.

- Δημιουργία λεξικού όπου κάθε λέξη έχει αποκτήσει ένα σκορ θετικότητας ή αρνητικότητας, ανάλογο του νοήματός της. Το λεξικό προκύπτει δίνοντας στο σύστημα κατηγοριοποιημένες προτάσεις οι οποίες περιέχουν λέξεις, και αυτές οι λέξεις χαρακτηρίζονται θετικές ή αρνητικές με κάποιο σκορ, ανάλογα με την συχνότητα που είναι σε θετικές και αρνητικές προτάσεις.
- Δημιουργία γραμματικών και συντακτικών κανόνων έτσι ώστε η τιμή του σκορ να λαμβάνει υπόψιν και την γραμματική και συντακτική μορφή της λέξης.
- Δημιουργία του τελικού λεξικού με το συναίσθημα για κάθε λέξη.

Μετά από αυτό το βήμα, κάθε πρόταση χαρακτηρίζεται ως θετική ή αρνητική ανάλογα με το πλήθος των θετικών και των αρνητικών όρων που περιέχει. Ένα επίσης χαρακτηριστικό του Vader είναι ότι έχει σχεδιαστεί για τον χαρακτηρισμό προτάσεων από το Twitter.

Η επανάσταση στο πρόβλημα του sentiment analysis ήρθε με την χρήση των transformers, και του attention mechanism, τα οποία βασίζονται σε βαθιά νευρωνικά δίκτυα και τα οποία έχουν καταφέρει να έχουν πολλή υψηλή επίδοση σε διάφορους τύπους προβλημάτων στο NLP, και πέρα από το πρόβλημα του sentiment analysis. Ένα χαρακτηριστικό παράδειγμα είναι το BERT [34] το οποίο αναπτύχθηκε από την google και έχει πολλές ενδιαφέρουσες παραλλαγές. Περισσότερες πληροφορίες για τους transformers αλλά και το BERT δίνονται στο αντίστοιχο υποκεφάλαιο.

2.4 Κυλιώμενα Παράθυρα σε Χρονοσειρές

Το πρόβλημα που πραγματεύεται η παρούσα διπλωματική αφορά την επεξεργασία χρονοσειρών, καθώς όλα τα δεδομένα δίνονται σε αυτή την μορφή. Στο επόμενο κεφάλαιο γίνεται ανάλυση των νευρωνικών δικτύων που θα χρησιμοποιηθούν για την επεξεργασία αυτών των χρονοσειρών, όμως σε αυτό το υποκεφάλαιο θα γίνει αναφορά σε μία τεχνική που εφαρμόζεται στις χρονοσειρές και έχει να κάνει με την παραθύρωση (windowing).

Όταν αναφερόμαστε σε χρονοσειρές, εννοούμε διακριτές συναρτήσεις οι οποίες αλλάζουν με την πάροδο του χρόνου, και ξέρουμε τις τιμές τους για συγκεκριμένες χρονικές στιγμές, που συνήθως ισαπέχουν χρονικά.

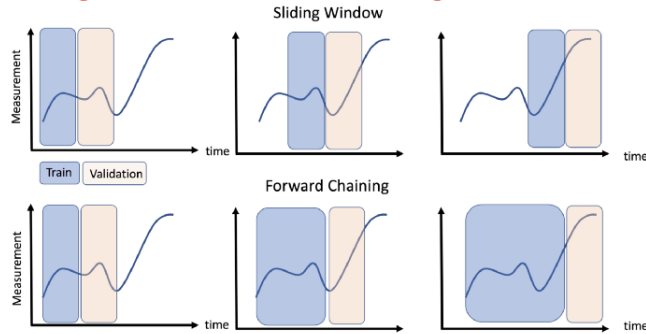
$$X_t = [X_t(0), X_t(1), \dots, X_t(n)], n \in Z$$

Το θέμα της παραθύρωσης[1] αφορά σε ποιο παράθυρο χρόνου πριν την χρονική στιγμή ενδιαφέροντος θα ψάξουμε για να δώσουμε απάντηση στο πρόβλημά μας. Όταν μιλάμε για

παράθυρο εννοούμε τις K χρονικές στιγμές των οποίων την τιμή θα συμπεριλάβουμε στην εκτίμηση της επόμενης χρονικής στιγμής, δηλαδή για την $X_t(n)$ θα εξετάσουμε τις χρονικές στιγμές από $X_t(n - k)$ έως $X_t(n - 1)$.

Σχήμα 2.1: Χρονικό παράθυρο σε χρονοσειρά[1]

Sliding window vs. forward chaining cross validation



Όπως παρατηρείται και στο σχήμα 2.1 το μπλε πλαίσιο δείχνει το χρονικό παράθυρο της χρονοσειράς το οποίο λαμβάνεται υπόψιν για να γίνει η πρόβλεψη για το κομμάτι που καλύπτει το ανοιχτόχρωμο πλαίσιο. Αυτό το κομμάτι είναι σημαντικό, καθώς πρέπει να πειραματιστούμε με τα δεδομένα που θα δίνουμε στο μοντέλο για να εξάγει την πρόβλεψη, καθώς δεν γίνεται κάθε είσοδος να λαμβάνει σαν πληροφορία οτιδήποτε υπάρχει σε μία χρονοσειρά πριν την ζητούμενη χρονική στιγμή. Η τεχνική αυτή, ονομάζεται τεχνική κυλιόμενου παραθύρου (sliding window) και έχει την έννοια ότι το μέγεθος του παραθύρου παραμένει σταθερό και ακολουθεί την χρονοσειρά.

Έχει αποτελέσει αντικείμενο έρευνας, το απαιτούμενο μέγεθος που πρέπει να έχει το παράθυρο ώστε να επιτυγχάνονται τα μέγιστα αποτελέσματα στην εκτίμηση των μελλοντικών τιμών της χρονοσειράς. Ένα σύννηθες πρόβλημα που βρίσκει εφαρμογή αυτή η μέθοδος, είναι η πρόβλεψη του φόρτου εργασίας κάποιου συστήματος, όπως για παράδειγμα κάποιο δίκτυο ενέργειας ή κάποιο Data center[35]. Ενδιαφέρον έχουν και νέες προσεγγίσεις βασισμένες σε μοντέλα βαθιάς μάθησης, όπου προσπαθούν να προσαρμόσουν το μέγεθος του παραθύρου ανάλογα με την εκάστοτε περίπτωση, ώστε να μην είναι απαραίτητα μεγάλα τα παράθυρα, αλλά και σε ορισμένες περιπτώσεις να περιέχουν όλη την απαραίτητη πληροφορία. Στην παρούσα εργασία ο πειραματισμός θα περιοριστεί στο σταθερό μέγεθος ενός παραθύρου.

Νευρωνικά Δίκτυα

Με τον όρο νευρωνικά δίκτυα, αναφερόμαστε σε μαθηματικά μοντέλα τα οποία προσπαθούν να μιμηθούν την λειτουργία μάθησης του νευρικού συστήματος του ανθρώπινου εγκεφάλου, μέσω νευρικών συνάψεων μεταξύ νευρώνων. Τα νευρωνικά δίκτυα αποτελούν μία από τις τεχνικές μηχανικής μάθησης.

3.1 Εισαγωγή στη Μηχανική Μάθηση - Είδη μάθησης

Αν έπρεπε να δώσουμε έναν ορισμό για την μηχανική μάθηση τότε ο πιο σαφώς ορισμένος είναι εκείνος του Tom M. Mitchell στο βιβλίο του Machine Learning[36] όπου ουσιαστικά αναφέρει το εξής:

Ένα πρόγραμμα μαθαίνει όταν η επίδοσή του σε μία εργασία με βάση κάποιο σκορ, βελτιώνεται όταν λαμβάνει υπόψιν μια προηγούμενη εμπειρία πάνω σε αυτή την εργασία. Αυτό υλοποιείται μοντελοποιώντας μαθηματικά το πρόβλημα και αντιισιχίζοντας την εμπειρία σε αριθμούς.

Κατά κανόνα τα είδη της μηχανικής μάθησης χωρίζονται σε τρεις κατηγορίες

- **Επιβλεπόμενη Μάθηση:**

Εδώ το πρόγραμμα προσπαθεί με βάση μία είσοδο να προβλέψει μία άγνωστη για αυτό, αλλά γνωστή σε εμάς έξοδο και προσπαθεί να την προσεγγίσει όσο καλύτερα γίνεται. Έδώ υπάρχουν και δύο υποκατηγορίες. Το πρόβλημα αυτό το χωρίζουμε σε classification όπου ο τελικός στόχος είναι μία ή περισσότερες κατηγορίες και η έξοδος του δικτύου είναι μία δακρυτή τιμή, και σε regression, "όπου εδώ και ο στόχος αλλά και η έξοδος του δικτύου είναι δεκαδική τιμή.

- **Μη Επιβλεπόμενη Μάθηση:**

Σε αυτή την κατηγορία προβλημάτων, έχουμε δεδομένα τα οποία δεν ανήκουν σαφώς σε κάποια κατηγορία και σκοπός της μάθησης είναι να βρει συσχετίσεις μεταξύ των διαφόρων εισόδων και να τις κατατάξει σε ομάδες ανάλογα με την εγγύτητά τους. Αυτό το πρόβλημα λέγεται clustering.

- **Ενισχυτική Μάθηση:**

Σε αυτό το είδος μάθησης, επίσης δεν υπάρχουν κατηγορίες για τα δεδομένα, όμως διαφέρει και από την μη επιβλεπόμενη μάθηση. Σε αυτή την περίπτωση ο αλγόριθμος

προσπαθεί να μάθει μέσω ενός συστήματος επιβράβευσης και τιμωρίας μέσα από κινήσεις σε κάποιο περιβάλλον. Πιο συγκεκριμένα, προσπαθεί σε κάθε βήμα να επιλέξει την κίνηση η οποία θα του μεγιστοποιήσει την επιβράβευση και θα ελαχιστοποιήσει την τιμωρία.

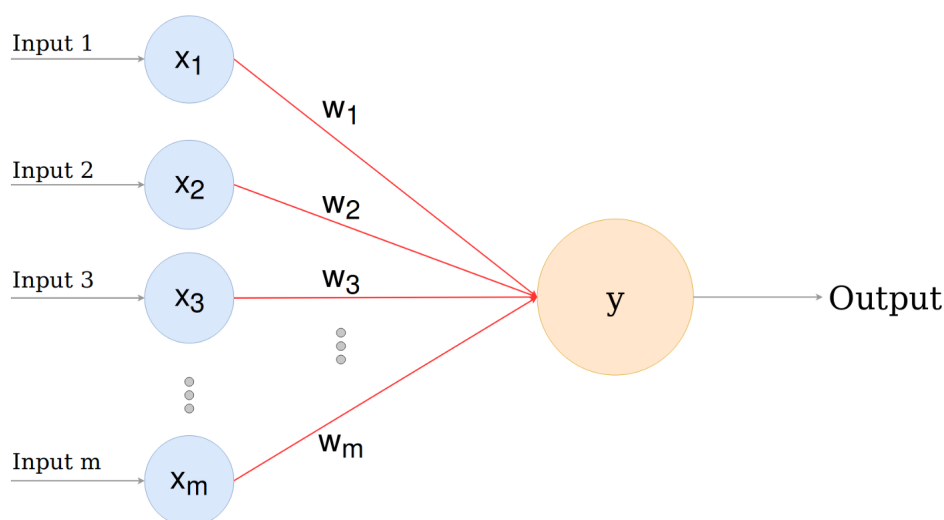
3.2 Είδη Νευρωνικών Δικτύων για χρονοσειρές

3.2.1 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks-ANN)

Πριν μπούμε σε πιο λεπτομερή δίκτυα που αφορούν την πρόβλεψη χρονοσειρών, θα περιγράψουμε την βασική δομή ενός απλού νευρωνικού δικτύου, του Multilayer perceptron (MLP).

Βασικό στοιχείο του MLP αποτελεί ο νευρώνας(perceptron), ο οποίος χαρακτηρίζεται από κάποια βάρη για το διάνυσμα εισόδου και μία συνάρτηση ενεργοποίησης για την έξοδο, όπως βλέπουμε στην παρακάτω εικόνα:

Σχήμα 3.1: Απλός Νευρώνας[2]



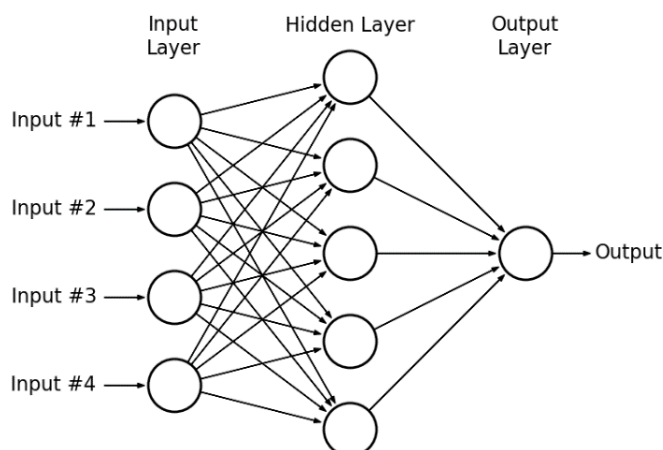
όπου η έξοδος y δίνεται από τον τύπο :

$$y = \begin{cases} 1 & \text{for } w * x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \quad (3.1)$$

Η παραπάνω σχέση ισχύει όταν έχουμε σαν συνάρτηση ενεργοποίησης [37] την βηματική συνάρτηση και χρησιμοποιείται σε περιπτώσεις δυαδικής ταξινόμησης [38]. Έχοντας σαν βάση την αρχή λειτουργίας του perceptron, ορίζονται και τα πολυστρωματικά perceptron[39], όπου η είσοδος δίνεται σε ένα στρώμα από νευρώνες και η έξοδός τους αποτελεί είσοδο για επόμενο στρώμα όπως την εικόνα που ακολουθεί:

Το perceptron έχει την δυνατότητα επίλυσης γραμμικά διαχωρίσιμων προβλημάτων, η τάξη των οποίων αυξάνεται όσο αυξάνεται το πλήθος των νευρώνων. Το MLP θεωρητικά προσεγγίζει την συμπεριφορά οποιασδήποτε συνάρτησης με σφάλμα μικρότερο από κάποιο ϵ , στην πράξη όμως δεν είναι βέβαιο ότι αυτό θα επιτευχθεί, και εξαρτάται από άλλα χαρακτη-

Σχήμα 3.2: Απλός Νευρώνας[3]



ριστικά του προβλήματος. Χαρακτηριστικό των πολυστρωματικών δικτύων είναι το επίπεδο εισόδου, στο οποίο δίνεται η είσοδος για το δίκτυο, το επίπεδο εξόδου το οποίο είναι το τελευταίο επίπεδο και δίνεται η έξοδος του δικτύου και τέλος τα ενδιάμεσα επίπεδα τα οποία αποτελούν τα κρυφά στρώματα (hidden layers).

Την ίδια δομή έχουν και τα βαθιά νευρωνικά δίκτυα (RNN, CNN) με την διαφορά ότι χρησιμοποιούν διαφορετικές δομικές μονάδες αντί του perceptron. Πρακτικά, θα λέγαμε ότι σε κάθε κρυφό επίπεδο γίνεται και μια εξαγωγή χαρακτηριστικών από το προηγούμενο επίπεδο και έτσι τελικά το δίκτυο καταφέρνει να δώσει μία έξοδο έχοντας λάβει υπ' όψιν πολλά κρυφά χαρακτηριστικά της εισόδου.

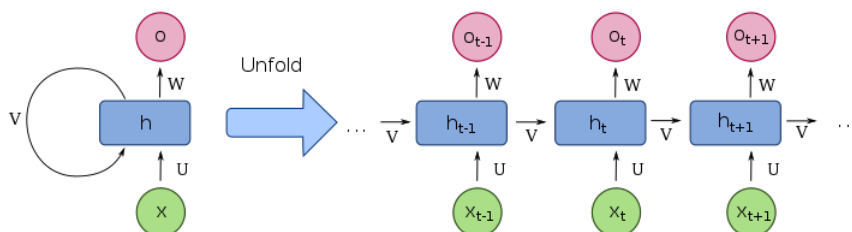
Για να καταφέρουν τα βαθιά νευρωνικά δίκτυα να μάθουν από τα δεδομένα εισόδου, χρησιμοποιούν την τεχνική backpropagation [40]. Με αυτή τη μέθοδο επί της ουσίας υπολογίζονται οι μερικές παράγωγοι σε κάθε επίπεδο του δικτύου και με βάση αυτές ο αλγόριθμος gradient decent [41] ανανεώνει τα βάρη του δικτύου. Η ενημέρωση αυτή των βαρών εξαρτάται και από άλλα χαρακτηριστικά των δικτύων που ορίζονται κατά την εκπαίδευση όπως ο ρυθμός μάθησης (learning rate).

3.2.2 Επαναληπτικά Νευρωνικά Δίκτυα

Τα επαναληπτικά νευρωνικά δίκτυα [42] αποτελούν και αυτά ένα είδος τεχνητών νευρωνικών δικτύων και είναι είδος βαθιών νευρωνικών. Δημιουργήθηκαν από την ανάγκη εξαγωγής πληροφορίας από ακολουθιακά δεδομένα όπως χρονοσειρές. Για να το επιτύχουν αυτό, κάθε δομικό στοιχείο τους λαμβάνει υπόψιν εκτός από την είσοδό του και ένα εσωτερικό διάνυσμα κρυφής κατάστασης το οποίο περιέχει πληροφορία από παρελθούσες καταστάσεις. Η δομή του φαίνεται στο σχήμα 3.3:

Όπως φαίνεται στο σχήμα 3.3, η έξοδος O_t προκύπτει από την h_t η οποία έχει λάβει υπόψιν την προηγούμενη κρυφή κατάσταση h_{t-1} . Το πρόβλημα που αντιμετωπίζουν τα επαναληπτικά δίκτυα είναι ότι επηρεάζονται το ίδιο από όλες τις παρελθοντικές τιμές καθώς και ότι οι μικρές τιμές χάνονται σε επόμενες κρυφές καταστάσεις λόγω της παραγωγίσιμης που γίνεται σε κάθε στάδιο κρυφής κατάστασης (vanishing gradient). Για αυτό τον λόγο υπήρξε η ανάγκη δημιουργίας δικτύων που θα αντιμετωπίζουν τέτοιου είδους προβλήματα.

Σχήμα 3.3: δίκτυο RNN[4]

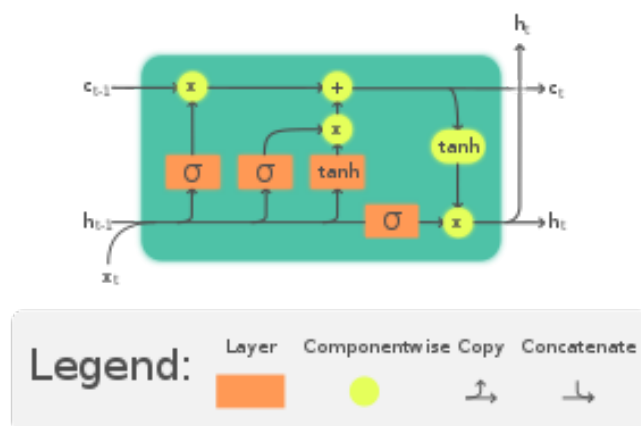


3.2.3 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης - LSTM

Ένα από τα δίκτυα που αντιμετωπίζουν το πρόβλημα που περιγράψαμε παραπάνω είναι τα δίκτυα μακράς βραχυπρόθεσμης μνήμης (long short-term memory (LSTM))[43]. Η διαφορά τους από τα απλά αναδρομικά δίκτυα έγκυται στο memory cell που έχουν στην δομή τους πέρα από τις πύλες εισόδου και εξόδου. Έτσι καταφέρνει ένα lstm δίκτυο να αντιμετωπίζει τα προβλήματα του vanishing και exploding gradient.

Πιο συγκεκριμένα, μπορούμε να δούμε την δομή του lstm παρακάτω:

Σχήμα 3.4: LSTM CELL[5]



Όπως παρατηρούμε, το hidden memory cell του lstm έχει τρεις διαφορετικές πύλες από τις οποίες περνάει η είσοδος x_t , ενώ εκτός από την είσοδο έχει και μία κρυφή κατάσταση h_t η οποία και αυτή ενημερώνεται κατάλληλα σε κάθε επόμενο κελί.

Ας μελετήσουμε λιγάκι πιο αναλυτικά την λειτουργία του παραπάνω κελιού.

Η πρώτη πύλη από τα αριστερά, η οποία ονομάζεται forget gate, έχει ως είσοδο την είσοδο x_t συν την κρυφή κατάσταση h_{t-1} του προηγούμενου κελιού. Η εξίσωση που την περιγράφει είναι η εξής:

$$F_t = \sigma(X_t W_{xf} + H_{t-1} W_{hf} + b_f)$$

Η συνάρτηση σ είναι η σιγμοειδής συνάρτηση και κάθε μία από τις τρεις πύλες που αναφέρθηκαν παραπάνω έχει μία τέτοια συνάρτηση με τα κατάλληλα κάθε φορά ορίσματα. Τα W_{xf} , W_{hf} είναι διανύσματα βαρών με τα οποία πολλαπλασιάζονται η είσοδος και η κρυφή

κατάσταση του προηγούμενου κελιού αντίστοιχα, ενώ το b_i είναι διάνυσμα σταθερών όρων.

Αντίστοιχα και για τις άλλες δύο πύλες έχουμε τις εξής συναρτήσεις ενεργοποίησης: Πρώτα για την δεύτερη πύλη η οποία είναι η πύλη εισόδου (input gate):

$$I_t = \sigma(X_t W_{xi} + H_{t-1} W_{hi} + b_i)$$

όπου ξανά τα W αποτελούν βάρη με τα οποία πολλαπλασιάζονται η προηγούμενη κρυφή κατάσταση και η τρέχουσα είσοδος, ενώ το b_i αποτελεί σταθερά. Τέλος και για την τρίτη πύλη ισχύουν αντίστοιχα :

$$O_t = \sigma(X_t W_{xo} + H_{t-1} W_{ho} + b_o)$$

Αυτό που παρατηρούμε και από το σχήμα είναι ότι η έξοδος της πύλης εισόδου (input gate) πολλαπλασιάζεται στον επόμενο κόμβο με την \tanh και αυτό δίνει ως έξοδο το εξής:

$$C'_t = \tanh(X_t W_{xc} + H_{t-1} W_{hc} + b_c)$$

Αυτός ο κόμβος ονομάζεται κόμβος εισόδου (input node).

Γνωρίζοντας τις σχέσεις που χαρακτηρίζουν τις πύλες ενός δομικού κελιού του LSTM μπορούμε να εξηγήσουμε τον τρόπο λειτουργίας του κελιού. Η πύλη forget gate καθορίζει την βαρύτητα που θα έχει στην τρέχουσα κατάσταση η κρυφή κατάσταση του προηγούμενου κελιού και έτσι καθορίζει τι ρόλο θα έχει η παρελθούσα γνώση ενώ αντίστοιχα το input gate καθορίζει το πόσο σημαντική είναι η τρέχουσα είσοδος αφού η έξοδος της σιγμοειδούς είναι μεταξύ 0 και 1. Τέλος η πύλη εξόδου output gate διαμορφώνει την τελική κρυφή κατάσταση του τρέχοντος κελιού, δηλαδή τί θα κρατηθεί από το παρελθόν και τι από το παρόν σαν πληροφορία για το επόμενο κελί. Με αυτό τον τρόπο και εκπαιδευοντας τα βάρη των τριών πυλών, το LSTM καταφέρνει να κρατάει βαθύτερες χρονικές συσχετίσεις χωρίς να αλλοιώνεται πλήρως η πληροφορία και να σταματάει να είναι πια χρήσιμη.[44]

3.2.4 Φραγμένα Επαναληπτικά Δίκτυα -GRU

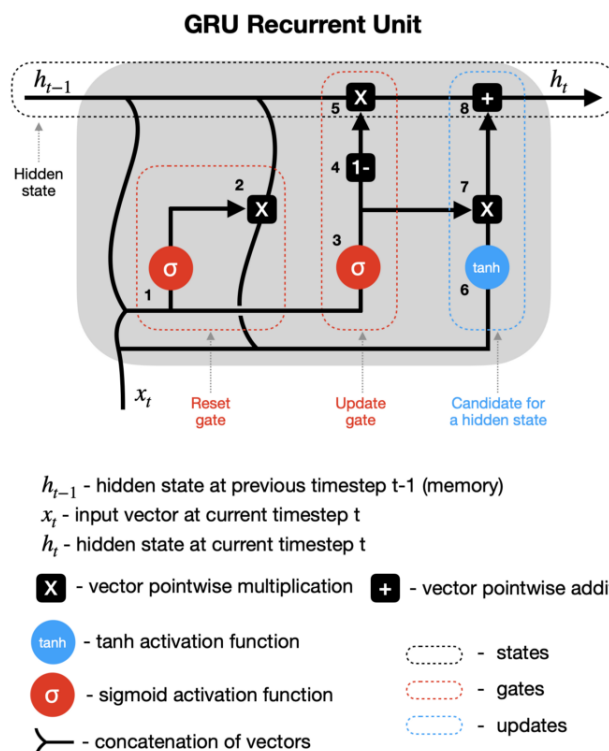
Σε παρόμοια λογική με το LSTM έχουν διαμορφωθεί και τα φραγμένα επαναληπτικά δίκτυα (Gated Recurrent Units ή GRU) το οποίο κινείται στην λογική της εσωτερικής κατάστασης όμως προσπαθώντας να κάνει λίγο πιο απλή την δομή.

Σε αντίθεση με το LSTM εδώ οι τρεις πύλες δίνουν την θέση τους σε 2, στην πύλη επαναφοράς (Reset Gate) και την πύλη ενημέρωσης (Update Gate), όπως βλέπουμε στο σχήμα 2.5.

Αρχικά, για το reset gate βλέπουμε ότι συνδυάζονται το h_{t-1} και το x_t δηλαδή η κρυφή προηγούμενη κατάσταση και η τωρινή είσοδος και περνάνε μέσα από αυτή την πύλη. Πιο συγκεκριμένα, αφού η σιγμοειδής συνάρτηση δίνει έξοδο μεταξύ του 0 και 1, στο πρώτο βήμα του reset gate ανατίθεται τιμή στο ποιές τιμές πρέπει να πεταχτούν (έξοδος 0) ποιες πρέπει να θυμάται το μοντέλο (έξοδος 1) και ποιες πρέπει να κρατηθούν εν μέρει (έξοδος μεταξύ 0 και 1). Στο βήμα δύο ανανεώνεται η προηγούμενη κρυφή κατάσταση με βάση το αποτέλεσμα του βήματος 1.

Στο Update gate περνάει ξανά ένας συνδυασμός της προηγούμενης κρυφής κατάστασης και

Σχήμα 3.5: GRU CELL[6]



της τρέχουσας εισόδου από μία σιγμοειδή συνάρτηση όμως τώρα τα βάρη για αυτό το συνδυασμένο διάνυσμα είναι διαφορετικά οπότε αυτό το βήμα είναι διαφορετικό από το βήμα ένα του reset gate. Έπειτα το αποτέλεσμα αυτής της πράξης αφαιρείται από την μονάδα αφού η τιμή της σιγμοειδούς είναι μεταξύ 0 και 1, και το αποτέλεσμα αυτό πολλαπλασιάζεται επίσης με το διάνυσμα της προηγούμενης κατάστασης. Η τελική κρυφή κατάσταση του GRU διαμορφώνεται αθροίζοντας την ανανεωμένη κατάσταση h_{t-1} με τον συνδυασμό προηγούμενης κατάστασης και τρέχουσας εισόδου πολλαπλασιασμένης με τα αντίστοιχα βάρη και περνώντας από την συνάρτηση ενεργοποίησης \tanh και πολλαπλασιασμού αυτού του διανύσματος με το διάνυσμα του βήματος 3 της Update gate [6].

3.2.5 Αρχιτεκτονικές Βαθιών Νευρωνικών Δικτύων

Αρχιτεκτονικές βαθιών νευρωνικών δικτύων έχουν υλοποιηθεί και χρησιμοποιηθεί σε διάφορες εφαρμογές από μέλη του Εργαστηρίου Συστημάτων Τεχνητής Νοημοσύνης και Μάθησης του ΕΜΠ. Ειδικότερα έχουν αναπτυχθεί τεχνικές CNN και CNN-RNN [45][46][47], ενώ έμφαση έχει δοθεί στην διαφάνεια και στην προσαρμογή των μοντέλων [48][49][50] αλλά και στην ανάπτυξη πλέον σύνθετων αρχιτεκτονικών, μπαϋεσιανών, με κάψουλες και αβεβαιότητα [51][52][53]. Βαθιές 3-Δ νευρωνικές αρχιτεκτονικές έχουν εφαρμοστεί στην αναγνώριση και σύνθεση συναισθήματος [54][55][56][57], ενώ άλλες εφαρμόζονται σε προβλήματα ανάλυσης σημάτων και εικόνων και αλληλεπίδρασης ανθρώπου-υπολογιστή [58][59][60][61].

3.3 Είδη Νευρωνικών Δικτύων για επεξεργασία φυσικής γλώσσας

3.3.1 Η μέθοδος των transformers στην εξαγωγή συναισθήματος

Η μέθοδος των transformers[62] [63] έχει προκύψει τα τελευταία χρόνια από την ανάγκη για ακόμα καλύτερη αντιμετώπιση των προβλημάτων που έχουν χρονική συσχέτιση μεταξύ τους, όπως το πρόβλημα της αναγνώρισης φυσικής γλώσσας και η επεξεργασία κειμένου, ενώ έχει βοηθήσει στην ανάπτυξη και ο μεγάλος όγκος δεδομένων που υπάρχει. Συγκεκριμένα, τα κλασικά μοντέλα για χρονοσειρές που συναντήσαμε παραπάνω, όπως το RNN και το LSTM αντιμετωπίζουν κάποια προβλήματα, όπως ότι ο υπολογισμός τους δεν μπορεί να παραλληλοποιηθεί, ότι δεν υπάρχει συγκεκριμένο μοντέλο για πιο μακροπρόθεσμες και πιο βραχυπρόθεσμες εξαρτήσεις μεταξύ των εισόδων καθώς και ότι το μοντέλο "ξεχνάει" μακρινά δεδομένα σε μια σειρά, δηλαδή δεν μπορεί να κρατήσει την πληροφορία ικανοποιητικά.

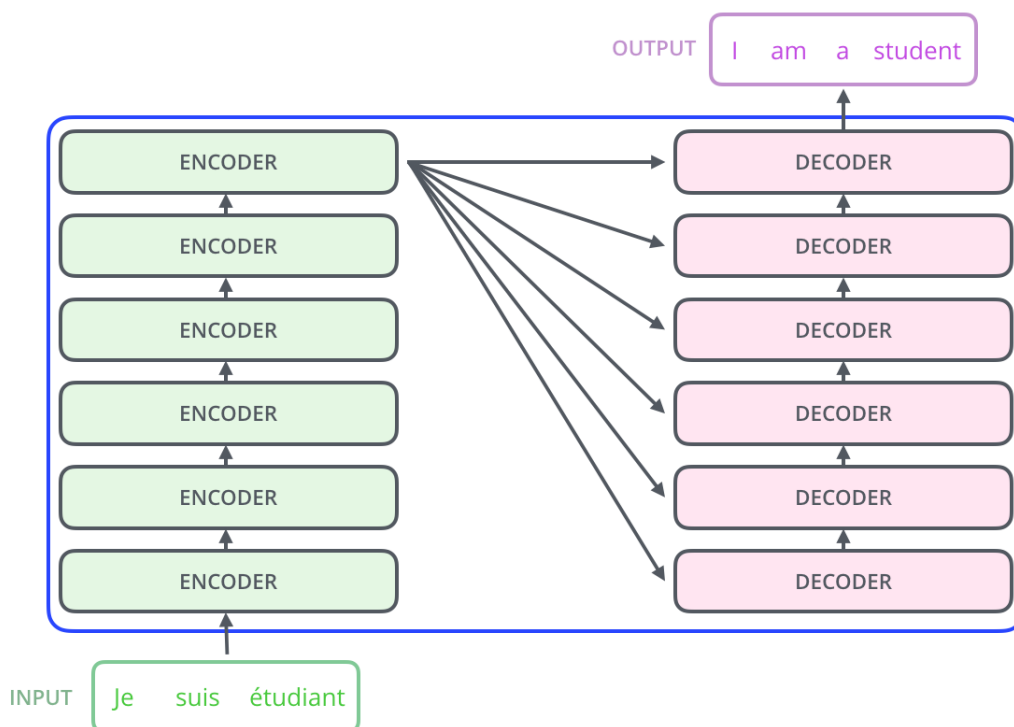
Για να αντιμετωπίσουν τα προβλήματα των κλασικών αυτών δικτύων, οι ερευνητές δημιούργησαν τους transformers, η δομή των οποίων βασίζεται στους μηχανισμούς προσοχής ή attention mechanisms[7], αλλά και σε συνελκτικά δίκτυα. Πιο συγκεκριμένα, ο μηχανισμός attention βασίζεται στο σύστημα κωδικοποιητή - αποκωδικοποιητή (encoder-decoder) και δημιουργεί μία "κρυφή", ανάλογη των επαναληπτικών νευρωνικών, κατάσταση για κάθε σημείο προσοχής της εισόδου, όπως μία λέξη, ή και όχι συγκεκριμένα λέξη όπως θα δούμε στην συνέχεια, στο επίπεδο του κωδικοποιητή (encoder), και στην συνέχεια από αυτό το σύνολο των κρυφών καταστάσεων ο αποκωδικοποιητής (decoder) μπορεί να εστιάσει καλύτερα στα κυριότερα σημεία της εισόδου και όχι απλώς να βλέπει μια μονοδιάσταση προηγούμενη κατάσταση. Για να αντιμετωπιστεί το πρόβλημα της παράλληλης επεξεργασίας της εισόδου, το μοντέλο χρησιμοποιεί συνελκτικά νευρωνικά δίκτυα, τα οποία έχουν την δυνατότητα παράλληλου υπολογισμού.

Η ακριβής δομή ενός transformer αποτελείται από έξι επίπεδα encoders και έξι επίπεδα decoders όπως βλέπουμε στο σχήμα.

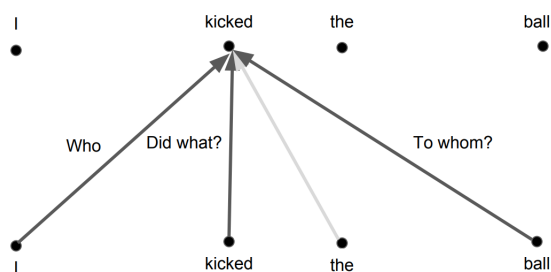
Το κάθε επίπεδο των encoders αποτελείται από ένα self attention layer και ένα απλό fully connected layer. Το self attention layer έχει την δομή του μηχανισμού προσοχής που περιγράψαμε, και στην ουσία αυτό που πετυχαίνει είναι να δημιουργεί συσχετίσεις κάθε στοιχείου της εισόδου με κάθε άλλο στοιχείο, σε επίπεδο κρυφών καταστάσεων, όπως βλέπουμε παρακάτω. Έτσι από επίπεδο σε επίπεδο γίνεται κωδικοποίηση σε διαφορετικές τέτοιες καταστάσεις και έτσι το μοντέλο καταφέρνει να μαθαίνει πολλές περισσότερες συσχετίσεις. Αντίστοιχα, σε κάθε επίπεδο decoder ανιχνεύονται αυτές οι συσχετίσεις από τις πιο ειδικές στις πιο γενικές.

Ένα επίσης σημαντικό στοιχείο των transformers, είναι ότι καταφέρνουν και κωδικοποιούν την θέση της κάθε λέξης σε ένα πρόβλημα επεξεργασίας κειμένου, και έτσι καταφέρνουν να αποκτούν πληροφορία όχι μόνο από το ποιές λέξεις χρησιμοποιούνται, αλλά και από τον τρόπο με τον οποίο συνδυάζονται.

Σχήμα 3.6: Δομή Transformers[7]



Self-Attention



Σχήμα 3.7: Ο μηχανισμός του Self-Attention[7]

3.3.2 Word Embeddings

Τα Word Embeddings[64] είναι μία εκπαιδευόμενη αναπαράσταση λέξεων όπου παρόμοιες λέξεις έχουν παρόμοιες αναπαραστάσεις. Χαρακτηρίζουμε ως εκπαιδευόμενη την αναπαράσταση καθώς, δεν είναι στατική για κάθε λέξη, αλλά αλλάζει μέσω εκπαίδευσης πάνω σε σύνολο προτάσεων και έτσι καταφέρνει να έχει την κατάλληλη αναπαράσταση η κάθε λέξη. Αυτή η αναπαράσταση είναι ουσιαστικά ένας μονοδιάστατος πίνακας αποτελούμενος από πραγματικές τιμές, η διάσταση του οποίου μπορεί να φτάνει τις δεκάδες ή και εκατοντάδες ψηφία. Με αυτό τον τρόπο μειώνεται αισθητά το μήκος αυτού του πίνακα-διάνυσμα, καθώς σε σύγκριση με την κωδικοποίηση one-hot encoding χρησιμοποιεί τον

λογάριθμο του πλήθους που θα χρησιμοποιούσε αλλιώς.

Υπάρχουν τρεις κύριοι τρόποι μάθησης για ένα word embedding.

- Ο πρώτος τρόπος είναι η εκπαίδευση ενός embedding layer μαζί με κάποιο νευρωνικό δίκτυο όπου στην ουσία εκπαιδεύουμε ένα δίκτυο, και μέσω ενός dense layer να δίνει κάθε φορά μία αναπαράσταση. Η εκπαίδευση αυτή γίνεται είτε σε προβλήματα supervised όπως η κατηγοριοποίηση κειμένου, είτε σε unsupervised όπου εκεί εξετάζονται συσχετίσεις μεταξύ κειμένων.
- Ένας δεύτερος, στατικός τρόπος μάθησης για Word Embeddings, είναι η αναπαράσταση Word2Vec[65]. Αυτή η μέθοδος παράγει embeddings τα οποία μπορούν να είναι ανεξάρτητα για κάθε λέξη. Η μάθηση επιτυγχάνεται μέσω πρόβλεψης της αναπαράστασης της λέξης για την οποία ενδιαφερόμαστε, από τις γειτονικές τις λέξεις στην ίδια πρόταση.
- Ένας τρίτος τρόπος για εκμάθηση των Word embeddings είναι οι GloVe ή Global Vectors for Word Representation [66], μία μέθοδος που έχει προκύψει από επέκταση της Word2Vec αναπαράστασης. Η μέθοδος αυτή είναι μία μίξη κλασικών τεχνικών αναπαράστασης λέξεων οι οποίες χρησιμοποιούσαν μεθόδους γραμμικής άλγεβρας όπως παραγοντοποίηση πινάκων Latent Semantic Analysis (LSA) και της τεχνικής Word2Vec η οποία εξετάζει τις συσχετίσεις με τις γειτονικές λέξεις και έτσι επιτυγχάνει την τελική αναπαράσταση των λέξεων. Συγκεκριμένα αντί να εξετάζει συσχετίσεις λέξεων μόνο σε συγκεκριμένη γειτονία, εκείνο κρατάει πίνακα συσχετίσεων για το σύνολο των λέξεων του κειμένου και αυτό έχει αποδειχθεί ότι μπορεί να οδηγήσει σε καλύτερα και πιο ισχυρά Embeddings.

3.3.3 BERT

Το Bidirectional Encoder Representations from Transformers (BERT)[34] είναι ένα μοντέλο που αναπτύχθηκε από ερευνητές της Google AI Language το 2018 το οποίο έφερε επανάσταση σε προβλήματα επεξεργασίας φυσικής φωνής και γλώσσας. Η κύρια καινοτομία του BERT είναι η αμφίδρομη εκπαίδευση των transformer η οποία όπως φάνηκε είχε εντυπωσιακά αποτελέσματα. Το μοντέλο χρησιμοποιείται κυρίως σε προβλήματα επεξεργασίας κειμένου και φυσικής γλώσσας, με τις διάφορες παραλλαγές του να εστιάζουν σε συγκεκριμένες πηγές κειμένου όπως για παράδειγμα από μέσα κοινωνικής δικτύωσης, ενώ ενδιαφέρον έχει η εφαρμογή του και σε αυτόματα διαλογικά συστήματα(chatbots).

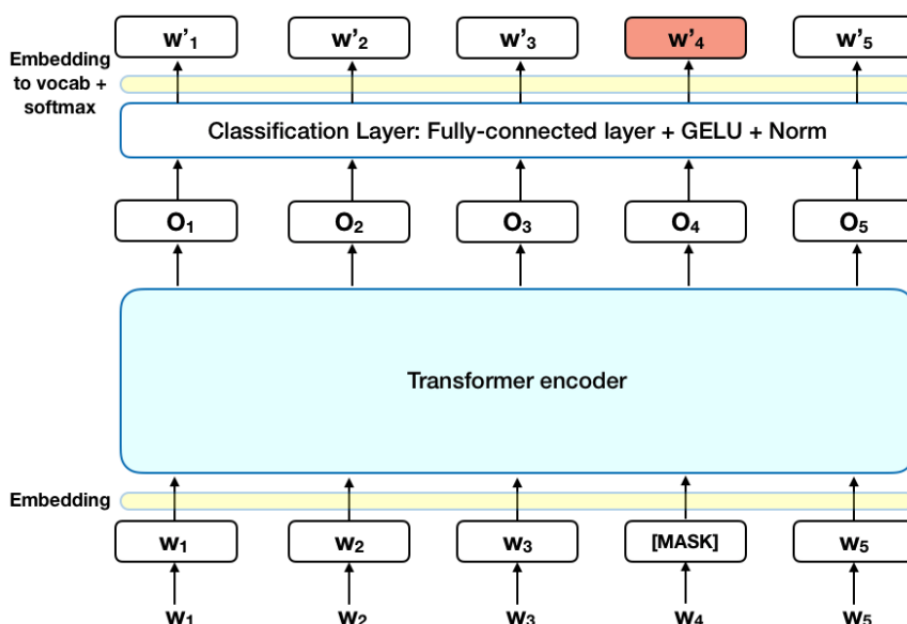
Τρόπος Λειτουργίας

Το BERT χρησιμοποιεί προεκπαιδευμένα word embeddings τα οποία μετατρέπουν την λέξη σε αναπαράσταση πινάκων και αυτοί οι πίνακες δίνονται ως είσοδος σε ένα σύστημα transformers οι οποίοι χρησιμοποιούν τους μηχανισμούς προσοχής. Συγκεκριμένα, το BERT χρησιμοποιεί μόνο τον μηχανισμό encoder των transformers αφού σκοπεύει στην δημιουργία ενός γλωσσικού μοντέλου. Έπειτα, αφού δημιουργήσει την πρόβλεψη που θέλουμε για το πρόβλημα (task) που έχουμε επιλέξει, επιλέγει σε ποιο embedding αντιστοιχεί η έξοδος του δικτύου με βάση μία softmax συνάρτηση, και αυτό είναι η έξοδος του δικτύου.

Η καινοτομία έγκυται στην αμφίδρομη εκπαίδευση ή καλύτερα στην μη κατευθυνόμενη εκπαίδευση, καθώς το μοντέλο προβλέπει την λέξη που λείπει ή την επόμενη πρόταση με βάση τι βλέπει τόσο πριν όσο και μετά το κενό που πρέπει να συμπληρωθεί. Αντίθετα, στα μοντέλα που προουπήρχαν του BERT γινόταν είτε σειριακή επεξεργασία από τα αριστερά προς τα δεξιά ή το ανάστροφο, είτε πρώτα αριστερά προς δεξιά και μετά δεξιά προς αριστερά, χωρίς να λαμβάνει σε καμία χρονική στιγμή ταυτόχρονα και τις δύο πληροφορίες υπόψιν[67]. Αυτό οδήγησε στην καλύτερη εκμετάλευση των προεκπαιδευμένων embeddings καθώς τα μέχρι τώρα μοντέλα δεν μπορούσαν να εκπαιδευτούν περαιτέρω το ίδιο καλά με το BERT σε ειδικές κατηγορίες κειμένων με συγκεκριμένα χαρακτηριστικά.

Τρόπος Εκπαίδευσης

Το BERT εκπαιδεύεται επάνω σε δύο βασικά task. Το πρώτο είναι ότι, με βάση μία είσοδο κειμένου, κρύβονται από το μοντέλο τυχαία το 15% των λέξεων του dataset και εκείνο προσπαθεί να μαντέψει τις κρυμμένες λέξεις. Ο αλγόριθμος μάθησης Gradient Decent χρησιμοποιεί τα αποτελέσματα πρόβλεψης μόνο των κρυμμένων λέξεων και έτσι η διαδικασία εκπαίδευσης είναι χρονοβόρα και απαιτεί μεγάλο όγκο πληροφορίας. Στο σχήμα 3.8 φαίνεται η δομή του BERT και η διαδικασία εκπαίδευσης με κρυμένο mask το 15% των λέξεων.



Σχήμα 3.8: Η δομή του BERT[8]

Μία δεύτερη μέθοδος εκπαίδευσης είναι η πρόβλεψη σε ένα κείμενο της επόμενης πρότασης που ακολουθεί. Έτσι αυτή την φορά το μοντέλο έχει σαν στόχο ολόκληρη πρόταση και όχι λέξη επομένως μαθαίνει συσχετίσεις τόσο σε επίπεδο λέξεων όσο και προτάσεων, αποκτώντας έτσι καλύτερη εικόνα για το νόημα των κειμένων,

Τρόπος Χρήσης του BERT

Το BERT κατά κύριο λόγο χρησιμοποιείται σαν ένα προεκπαιδευμένο δίκτυο στο οποίο εφαρμόζουμε λεπτό συντονισμό (fine tuning). Με την μέθοδο fine tuning παίρνουμε τα βάρη ενός ήδη εκπαιδευμένου δικτύου και συνεχίζουμε την εκπαίδευση με κάποιο άλλο Dataset αλλάζοντας αυτή την φορά τα βάρη μόνο στα τελευταία στρώματα του δικτύου. Με αυτόν τον τρόπο κρατάμε τις γενικές συσχετίσεις που μπορεί να βρει το μοντέλο μέσα σε προτάσεις και του δείχνουμε εμείς τον τρόπο που θα τις χρησιμοποιεί σε πιο συγκεκριμένα προβλήματα (tasks) όπως για παράδειγμα σε ένα πρόβλημα αναγνώρισης συναισθήματος σε ένα κείμενο. Λόγω αυτής της ευελιξίας του δικτύου έχει πολλές εφαρμογές και σε μέσα κοινωνικής δικτύωσης, όπως για ανίχνευση αρνητικών συναισθημάτων στην κοινή γνώμη, προβλήματα υγείας που συζητούνται, ακραία αισθήματα και άλλα[68][69][70][71][72]. Ενδιαφέρον παρουσιάζουν και οι διάφοροι συνδυασμοί που γίνονται στην αρχιτεκτονική του μοντέλου για να προαρμοστεί στα διάφορα προβλήματα, όπως η χρήση διαφορετικών word embeddings σε σχέση με τα προεκπαιδευμένα του κλασσικού μοντέλου, αλλά και αλλαγές στην δομή των τελευταίων στρωμάτων, καθώς επίσης και διαφορετικά δεδομένα για εκπαίδευση.

3.3.4 RoBERTa

Το RoBERTa αφορά μία προσέγγιση στην προεκπαίδευση και εκπαίδευση του BERT η οποία είναι πιο αποδοτική σε διάφορες ξεχωριστές περιπτώσεις[73].

Συγκεκριμένα, οι αλλαγές που δοκιμάστηκαν για καλύτερα αποτελέσματα είναι οι εξής:

- Για την εκπαίδευση με μάσκα σε τυχαίες λέξεις, το κλασσικό BERT εφαρμόζει μόνο στο κομμάτι του pretraining αυτή την διαδικασία και έτσι για κάθε εποχή της εκπαίδευσης το μοντέλο μαθαίνει επάνω στις ίδιες κρυμμένες λέξεις. Αντίθετα στο RoBERTa δημιουργούνται 10 διαφορετικά τυχαία datasets που σε κάθε ένα από αυτά υπάρχουν διαφορετικές κρυμμένες λέξεις.
- Μία δεύτερη βελτίωση που παρατηρήθηκε σε τεχνικό επίπεδο ήταν ότι βελτιωνόταν ελαφρώς η επίδοση του μοντέλου σε συγκεκριμένα dataset όταν δεν λαμβανόταν υπόψιν το loss από την πρόβλεψη της επόμενης πρότασης (αλλά μόνο των κρυμμένων λέξεων).
- Μία ακόμα βελτίωση ήρθε στην επίδοση με την αύξηση του batch size στην εκπαίδευση ενώ επίσης έτσι βελτιώθηκε και η δυνατότητα παραλληλοποίησης της διαδικασίας εκπαίδευσης και μείωσης του χρόνου που απαιτείται.

Το μοντέλο RoBERTa έχει εφαρμογές και σε πρόβλεψη συναισθήματος επάνω σε κείμενα που αφορούν οικονομικά νέα[74].

Μέρος 

Πρακτικό Μέρος

Κεφάλαιο 4

Πειραματικό Μέρος

Σε αυτό το κεφάλαιο θα μελετήσουμε τόσο τα δεδομένα (datasets) που χρησιμοποιήσαμε για την εκπαίδευση των δικτύων όσο και τις αρχιτεκτονικές αυτών.

4.1 Dataset

Τα dataset είναι ένα πολύ σημαντικό κομμάτι του προβλήματος καθώς πάνω σε αυτά προσαρμόζονται τα μοντέλα και όχι το αντίστροφο. Επομένως πρέπει να βρούμε τον κατάλληλο τρόπο να τα χρησιμοποιήσουμε, κρατώντας τα σημαντικά στοιχεία από αυτά και διώχνοντας τα περιττά, ενώ θα πρέπει και να γίνουν κατάλληλες αλλαγές για να είναι πιο εύκολη η επεξεργασία τους.

4.1.1 Twitter - Χαρακτηριστικά

Το Twitter είναι ένα μέσο κοινωνικής δικτύωσης, όπου κάθε χρήστης μπορεί να ακολουθεί άλλους χρήστες και να βλέπει τα tweets που δημοσιεύουν. Τα tweets είναι μικρά κείμενα, το μέγιστο 140 χαρακτήρων, μέσα στα οποία υπάρχουν συντομογραφίες, λέξεις αργκό αλλά και διάφορα emoticons (εικονίδια με γκριμάτσες). Σε κάθε Tweet κάθε χρήστης μπορεί να κάνει κάποιο σχόλιο το οποίο είναι εμφανές σε κάθε άλλο χρήστη ενώ μπορεί να δηλώσει και ότι του αρέσει η δημοσίευση, κάτι το οποίο είναι επίσης φανερό. Όλο αυτό καθιστά τον κόσμο του Twitter πολύ ζωντανό, άμεσο και γεμάτο πληροφορία καθώς οι χρήστες πρέπει να είναι λιτοί και περιεκτικοί στις δημοσιεύσεις τους.

Χαρακτηριστικό του Twitter είναι ότι το επιλέγουν τα πλέον επίσημα πρόσωπα, όπως πρόεδροι χωρών, επιφανείς δημοσιογράφοι αλλά και επιχειρηματίες, για να εκφράσουν τις απόψεις τους και να συμμετέχουν στον δημόσιο διάλογο, με αποτέλεσμα να αποκτά η πλατφόρμα κάποιο κύρος. Δεν είναι λίγες οι περιπτώσεις όπου tweets έχουν αποτελέσει σημείο διαμάχης σε κεντρικές πολιτικές σκηνές ενώ είναι εμφανής η επιρροή του στον κόσμο των επιχειρήσεων, με πρόσφατο παράδειγμα τα tweets του Elon Musk τα οποία έχουν άμεσο αντίκτυπο σε μετοχές κρυπτονομισμάτων όπως του Bitcoin[75].

Τα παραπάνω αναφέρονται για να γίνει κατανοητό ότι το Twitter μπορεί να έχει επιρροή και στην τιμή του Bitcoin, καθώς η γνώμη των επενδυτών επηρεάζεται από τα ποσταρίσματα των πιο γνωστών επιχειρηματιών[13].

4.1.2 Dataset με Tweets

Για να εκπαιδεύσουμε τα μοντέλα στην αναγνώριση συναισθήματος πάνω σε Tweets σχετικά με το Bitcoin, χρησιμοποιήσαμε ένα Dataset από το kaggle (<https://www.kaggle.com/datasets/kaushikuresh147/bitcoin-tweets>) το οποίο περιείχε 16 εκατομμύρια tweets που περιείχαν τον όρο "Bitcoin" ή "BTC". Το dataset αυτό δημιουργήθηκε μέσω του Twitter API το οποίο δίνει την δυνατότητα για ανάκτηση Tweets τα οποία περιέχουν συγκεκριμένες λέξεις-κλειδιά.

Το dataset αυτό περιείχε tweets σχετικά με το bitcoin από 1.1.2009 έως και 22.11.2019 αλλά με αυξημένη συχνότητα μετά την 1.1.2018, και για τον λόγο αυτό απορίψαμε όσα tweets ήταν προηγούμενων ετών. Για να χρησιμοποιήσουμε τα tweets για το δικό μας πείραμα χρειάστηκε να γίνει κάποιο καθάρισμα σε αυτά. Έτσι για αρχή επιλέχθηκαν μόνο τα tweets που είναι γραμμένα στην αγγλική γλώσσα, κάνοντας χρήση του εργαλείου `whatthelang`. Στην συνέχεια στα εναπομείναντα 1.600.000 περίπου τιτθίσματα, εφαρμόσαμε κάποιες άλλες τεχνικές επεξεργασίας κειμένου και συγκεκριμένα αφαιρέσαμε ονόματα χρηστών από αυτά, σημεία στίξης, μετατρέψαμε όλα τα γράμματα σε πεζά και αφαιρέσαμε τα emoticons. Σε αυτό πλέον το επεξεργασμένο σύνολο δεδομένων εφαρμόσαμε τα μοντέλα εξαγωγής συναισθήματος και πήραμε μία τιμή -1, 0 ή 1 για κάθε tweet με -1 αρνητικό, 0 ουδέτερο και 1 θετικό.

Επίσης αξίζει να αναφερθούμε και στις υπόλοιπες πληροφορίες που περιείχε το dataset από το Twitter. Έτσι λοιπόν, για κάθε tweet ξέραμε την ακριβή ώρα και ημερομηνία για το πότε έγινε η δημοσίευση, το πλήθος των likes, replies, retweets για κάθε δημοσίευση καθώς και όνομα χρήστη. Με βάση αυτές τις πληροφορίες το κάθε tweet έπαιρνε ένα ιδιαίτερο βάρος καθώς άλλη βαρύτητα έχει κάποιο tweet με πολλά likes σε σχέση με κάποιο που δεν είχε καμία τέτοια αντίδραση.

Για να αποφύγουμε να χρησιμοποιήσουμε tweets τα οποία δεν θα είχαν επίδραση στην τελική πρόβλεψη, κρατήσαμε μόνο όσα tweets είχαν μη μηδενικό πλήθος σε δύο από τα τρία εξής χαρακτηριστικά: likes, user followers και user favourites, δηλαδή δίνεται βαρύτητα στην επιρροή του χρήστη που δημοσιεύει και στην απήχηση του εκάστοτε tweet. Η τρέχουσα μορφή του dataset είναι αυτή που φαίνεται στο σχήμα 4.1.

	text	replies	likes	retweets	date	aux_date	sentiment
0	running bitcoin	790	14470	5542	2009-01-11	2009-01-11 03:33:52	1.0
1	looking at ways to add more anonymity to bitcoin	55	1544	392	2009-01-21	2009-01-21 17:29:40	1.0
2	thinking about how to reduce co2 emissions fro...	44	1042	277	2009-01-27	2009-01-27 20:14:10	1.0
3	from satoshi nakamoto 2009 01 11 22 32 bitc...	0	28	16	2009-01-29	2009-01-29 13:37:53	1.0
4	just wrote bitcoin new open source p2p e cas...	5	64	42	2009-02-18	2009-02-18 05:57:24	1.0

Σχήμα 4.1: Τελική μορφή του Dataset με συναισθημα

Στη συνέχεια, με βάση τα tweets που υπήρχαν ανά ώρα, έγινε εξαγωγή συναισθήματος ανά ώρα. Για να εξαγάγουμε το μέσο συναισθημα, για κάθε tweet εξαγόταν ένα βάρος που πολλαπλασιαζόταν με την τιμή του συναισθήματος. Το βάρος για κάθε tweet i ήταν:

$$W_i = \frac{\text{likes}_i + \text{retweets}_i + \text{comments}_i}{3}$$

και το τελικό σταθμισμένο συναίσθημα για κάθε tweet

$$sentiment_i = W_i S_i$$

όπου S_i είναι το αρχικό συναίσθημα χωρίς βάρους που είχε προκύψει από το μοντέλο εξαγωγής συναισθήματος. Το τελικό συναίσθημα ανά ώρα προέκυπτε με το άθροισμα των tweet μίας ώρας δια το πλήθος τους.

$$Sent_{hour}^i = \frac{\sum_{i \in hour} sentiment_i}{n}$$

Επειδή δεν είχαμε tweets για κάθε ώρα κάθε ημέρας όσες τιμές έλειπαν συμπληρώθηκαν με τον μέσο όρο των τιμών της υπόλοιπης μέρας. Η μορφή αυτή φαίνεται στο σχήμα 4.2. Αφού συγκεντρώσαμε το συναίσθημα σε ώρες, μοιράσαμε το dataset σε 24 στήλες μία για κάθε ώρα της ημέρας.

	time	average_sentiment	num_of_tweets
8100	2019-01-04 19	32.334106	5
8101	2019-01-04 20	3.084298	3
8102	2019-01-04 21	93.325665	4
8103	2019-01-04 22	57.828672	6
8104	2019-01-04 23	1.617497	3
8105	2019-01-05 00	1.770327	4
8106	2019-01-05 01	0.807234	1
8107	2019-01-05 02	14.067563	6
8108	2019-01-05 03	0.370965	1
8109	2019-01-05 04	10.692845	2
8110	2019-01-05 05	0.217171	1
8111	2019-01-05 06	22.107970	2
8112	2019-01-05 07	0.744890	2
8113	2019-01-05 08	0.217171	1
8114	2019-01-05 09	4.202632	1
8115	2019-01-05 10	5.712809	3
8116	2019-01-05 11	17.394600	1
8117	2019-01-05 12	8.482255	4
8118	2019-01-05 13	264.419125	2
8119	2019-01-05 14	0.216138	1
8120	2019-01-05 15	143.575517	4
8121	2019-01-05 17	-38.803871	3

Σχήμα 4.2: Εξαγωγή μέσου συναισθήματος ανά ώρα μαζί με πλήθος tweets

4.1.3 Dataset για fine tuning του BERT

Ένα άλλο dataset που χρειάστηκε για την ολοκλήρωση της εργασίας, ήταν το dataset για το fine tuning του μοντέλου RoBERTa το οποίο χρησιμοποιήθηκε για την εξαγωγή συναισθήματος στα Tweets. Όπως αναφέραμε στο αντίστοιχο κεφάλαιο για το BERT, το συγκεκριμένο μοντέλο είναι προεκπαιδευμένο πάνω σε μεγάλα dataset σχετικά με την αγγλική γλώσσα, και επάνω στα δύο συγκεκριμένα tasks που περιγράψαμε, επομένως είναι χρήσιμη και κάποια περαιτέρω εκπαίδευση για να προσαρμοστεί το μοντέλο στα ζητούμενα του κάθε task.

Έτσι, βρήκαμε ένα dataset[76] το οποίο περιέχει προτάσεις από άρθρα σχετικά με οικονομικές ειδήσεις, και κάθε πρόταση έχει επισημανθεί ως θετική, αρνητική ή ουδέτερη από ειδικούς πάνω στον οικονομικό κλάδο. Το dataset ονομάζεται financial phrase bank. Το dataset συνολικά περιέχει 4846 προτάσεις, αριθμός ικανοποιητικός για το fine tuning του RoBERTa. Μετά την προσαρμογή των στρωμάτων ταξινόμησης του RoBERTa πάνω σε αυτό το dataset η ακρίβεια του μοντέλου RoBERTa ήταν 86.7% στο test set. Το dataset ονομάζεται financial phrase bank και βρίσκεται στην διεύθυνση <https://www.kaggle.com/datasets/ankurzing/sentiment-analysis-for-financial-news>. Παρακάτω παραθέτουμε μία ενδεικτική γραμμή του συγκεκριμένου dataset:

neutral,"According to Gran , the company has no plans to move all production to Russia , although that is where the company is growing ."

4.1.4 Historical prices dataset

Ένα ακόμα dataset που χρησιμοποιήσαμε ήταν αυτό με ιστορικές τιμές του bitcoin. Οι συγκεκριμένες πληροφορίες ανακτήθηκαν από το yahoo finance[77] και περιείχαν όλες τις τιμές ανοίγματος και κλεισίματος της μετοχής του bitcoin, τον όγκο των μετοχών που συναλλάχθηκαν, την υψηλότερη και χαμηλότερη τιμή της μέρας και τον σταθμισμένο μέσο της τιμής κλεισίματος μίας μέρας.

	Date	Open	High	Low	Close	Adj Close	Volume
2787	2022-05-05	39695.746094	39789.281250	35856.515625	36575.140625	36575.140625	43106256317
2788	2022-05-06	36573.183594	36624.359375	35482.132813	36040.921875	36040.921875	37795577489
2789	2022-05-07	36042.503906	36129.925781	34940.824219	35501.953125	35501.953125	24375896406
2790	2022-05-08	35502.941406	35502.941406	33878.964844	34059.265625	34059.265625	36763041910
2791	2022-05-09	34039.796875	34210.621094	32771.890625	32771.890625	32771.890625	47362785280

Σχήμα 4.3: Dataset με τεχνικούς δείκτες της μετοχής

4.1.5 Εξαγωγή τελικού Dataset

Για το τελικό πείραμα έγιναν διάφορες δοκιμές σε πολλούς διαφορετικούς συνδυασμούς των πληροφοριών των dataset, με σκοπό να βρούμε ποιες πληροφορίες είναι πιο χρήσιμες στην πρόβλεψη της τιμής του Bitcoin.

Αρχικά προσπαθήσαμε να κάνουμε πρόβλεψη της τιμής για την επόμενη μέρα, για τρεις μέρες μπροστά και επτά μέρες μπροστά. Έτσι κυλίσσαμε (shift) τόσες θέσεις προς τα πάνω την στήλη με την τιμή κλεισίματος, όσες μέρες μπροστά θέλαμε να προβλέψουμε την τιμή κλεισίματος, και αυτό ήταν ο στόχος y των πειραμάτων μας.

Ένα άλλο χαρακτηριστικό που χρησιμοποιήσαμε για την πρόβλεψη της τιμής πέρα από το συναίσθημα ανά ώρα ήταν το πλήθος των tweet ανά ώρα. Αυτό το χαρακτηριστικό το αντήσαμε μαζί με το μέσο συναίσθημα κάθε ώρας. Επίσης στο τελικό dataset έπρεπε να γίνει αντιστοίχιση των τεχνικών δεικτών σε κάθε μέρα για να έχουμε τις τιμές ανοίγματος, κλεισίματος κτλ. Ασφαλώς πειραματιστήκαμε σε πολλές διαφορετικές εκδοχές του dataset πάνω σε ίδια κάθε φορά μοντέλα νευρωνικών δικτύων για να καταλήξουμε στο ποιές πληροφορίες

έχουν μεγαλύτερη συσχέτιση με την μελλοντική τιμή του Bitcoin. Ένα παράδειγμα dataset με τιμή συναισθήματος και πλήθους tweets για κάθε ώρα της μέρας είναι στα σχήματα 4.4 και 4.5, όπου οι στήλες 01, 02, ..., 22 είναι οι στήλες με το μέσο συναίσθημα για κάθε ώρα, υπολογισμένο όπως αναφέρθηκε παραπάνω, οι στήλες n1, n2, ..., n21, n22 είναι οι στήλες με το πλήθος των tweet ανά ώρα, και τέλος υπάρχει η στήλη prices η οποία έχει τις τιμές που θέλουμε να προβλέψουμε, καθώς και οι στήλες open, low, high, volume, close, οι οποίες περιέχουν τεχνικούς δείκτες.

	date	00	n0	01	n1	02	n2	03	n3	04	...	19	n19	20	n20	21	n21	22
0	2018-01-01	3.625599	3.0	14.405082	4.0	4.776748	0.0	0.838406	2.0	0.465444	...	7.471285	1.0	-1.399500	2.0	0.061827	1.0	-63.77
1	2018-01-02	0.559924	2.0	45.352557	5.0	1.492878	2.0	0.898684	4.0	12.763351	...	92.992344	8.0	257.449127	6.0	0.123655	1.0	31.417
2	2018-01-03	19.533672	7.0	0.773101	4.0	0.405097	2.0	0.281442	1.0	0.373030	...	0.374958	1.0	15.906827	1.0	0.154827	1.0	7.5467
3	2018-01-04	38.040373	4.0	9.940021	3.0	0.982087	3.0	2.331215	3.0	23.806435	...	5.223766	2.0	1.281629	2.0	57.750016	2.0	12.247
4	2018-01-05	-0.093516	2.0	3.379123	4.0	2.044549	3.0	1.302439	1.0	6.798120	...	4.387030	6.0	455.948557	3.0	17.577871	6.0	168.90

Σχήμα 4.4: Dataset με συναίσθημα και πλήθος tweets για κάθε ώρα της μέρας

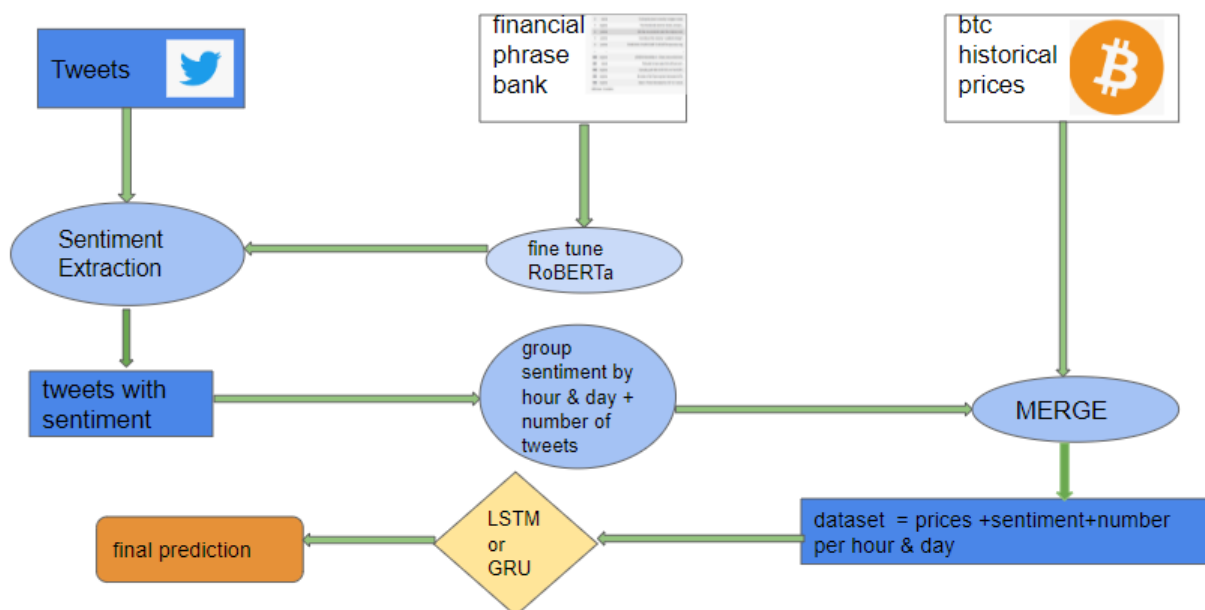
n22	23	n23	Open	Low	High	Volume	Close	prices
167.0	212.680138	137.0	8573.980469	8273.573242	8653.280273	21579470673	8309.286133	8206.145508
136.0	188.607911	126.0	8305.134766	8099.963379	8408.516602	21083613816	8206.145508	8027.268066
157.0	177.863179	123.0	8203.613281	8010.511719	8237.240234	20764300437	8027.268066	7642.750000
191.0	607.278951	149.0	8023.644531	7597.381836	8110.098145	22514243371	7642.750000	7296.577637
166.0	216.259312	117.0	7643.569336	6936.706543	7697.382813	34242315785	7296.577637	7397.796875

Σχήμα 4.5: Dataset με συναίσθημα και πλήθος tweets για κάθε ώρα της μέρας (μέρος 2)

Τέλος, επειδή το dataset με το συναίσθημα περιείχε τιμές σε διαφορετικές κλίμακες, χρειάστηκε να γίνει robust scaling όλων των δεδομένων ώστε να βρίσκονται στην κλίμακα μεταξύ 0 και 1, κάτι που όπως αποδείχθηκε βοηθάει καλύτερα τα μοντέλα να αντιλαμβάνονται τις συσχετίσεις μεταξύ των διάφορων τιμών. Με τον όρο Robust scaling αναφερόμαστε σε κανονικοποίηση των τιμών των δεδομένων σε μία κλίμακα που ορίζουμε εμείς, συνήθως από 0 έως 1, έτσι ώστε το μοντέλο να μπορεί πιο εύκολα να εκπαιδευτεί. Η κανονικοποίηση γίνεται με τέτοιο τρόπο ώστε να διατηρούνται οι αναλογίες μεταξύ των τιμών.

4.2 Ροή Εργασιών

Για να καταλήξουμε στην πρόβλεψη της τιμής του Bitcoin τις επόμενες ημέρες κατά βάση ακολουθήσαμε τα βήματα που φαίνονται στο διάγραμμα 4.6. Από εκεί παρατηρούμε ότι χρειαστήκαμε τρία ξεχωριστά datasets. Το ένα, εκείνο με τις κατηγοριοποιημένες προτάσεις σχετικά με το συναίσθημα που αφορούσαν οικονομικά άρθρα(financial phrase bank), το χρησιμοποιήσαμε με σκοπό να μας βοηθήσει στην καλύτερη εξαγωγή συναισθήματος από τα tweets, και συγκεκριμένα εξειδικεύοντας το μοντέλο RoBERTa επάνω στο πρόβλημα εξαγωγής συναισθήματος σε tweets με συγκεκριμένο περιεχόμενο. Το δεύτερο ήταν το κύριο dataset το οποίο περιείχε tweets με λέξεις κλειδιά σχετικά με το bitcoin και από το οποίο πήραμε τον μεγαλύτερο όγκο πληροφοριών για το γενικό συναίσθημα που επικρατεί ανά ώρα στο Twitter σχετικά με το bitcoin, καθώς και το πλήθος των δημοσιεύσεων. Το τρίτο dataset που χρησιμοποιήσαμε αφορούσε τεχνικούς δείκτες όπως η τιμή κλεισίματος της μετοχής και



Σχήμα 4.6: Διάγραμμα ροής ενεργειών μέχρι την τελική πρόβλεψη με χρήση RoBERTa για εξαγωγή συναισθήματος και χρήση πλήθους των tweets

μας χρειάστηκε τόσο ως τιμή στόχος για το μοντέλο, όσο και για καλύτερη πρόβλεψη του μοντέλου χρησιμοποιώντας ως πληροφορία προηγούμενες τέτοιες τεχνικές τιμές.

Έπειτα, επόμενο σημείο ενδιαφέροντος αποτελεί το μοντέλο βαθιάς μάθησης που χρησιμοποιήθηκε για την πρόβλεψη της τιμής του bitcoin με βάση τις εισόδους που του δίναμε. Έτσι, βασιστήκαμε κυρίως σε 2 κλασικά μοντέλα, το LSTM και το GRU. Όπως θα δούμε στο επόμενο κεφάλαιο, για το συγκεκριμένο πρόβλημα φαίνεται να μην έχουν μεγάλη διαφορά στην απόδοσή τους. Τα χαρακτηριστικά τους είναι 2 επίπεδα από αυτά τα δίκτυα (LSTM ή GRU) με συγκεκριμένο πλήθος μονάδων (από 32 έως 256 μονάδες βλ σχήμα 5.12) και ένα τελικό fully connected επίπεδο για την τελική κατηγοριοποίηση.

Ακόμα, αξίζουν προσοχής τα ενδιαμέσα βήματα εξαγωγής συναισθήματος ανά ώρα και κατηγοριοποίηση ανά μέρα, εκεί όπου γίνεται σύμπτυξη πληροφορίας, για να μπορεί να δοθεί πιο εύκολα σαν είσοδος στα δίκτυα που αναφέραμε στην προηγούμενη παράγραφο. Συγκεκριμένα, μετά την εξαγωγή συναισθήματος από το κάθε tweet ξεχωριστά, είτε με την μέθοδο Vader είτε με το RoBERTa, εξάγαμε το μέσο συναίσθημα ανά ώρα καθώς και το πλήθος των Tweets ανά ώρα. Στην συνέχεια σε κάθε γραμμή του Dataset προσθέσαμε και τους τεχνικούς δείκτες και αυτό αποτέλεσε το τελικό dataset που δινόταν σαν είσοδος στο δίκτυο, και μία εκδοχή του φαίνεται στα σχήματα 4.4 και 4.5.

Κεφάλαιο 5

Πειραματικά Αποτελέσματα

Σε αυτό το κεφάλαιο θα ασχοληθούμε με τα πειράματα που διεξήχθησαν για την εκτίμηση της κίνησης της τιμής του bitcoin και θα σχολιάσουμε τα αποτελέσματα με βάση κάποιες μετρικές. Επειδή τα αποτελέσματα αξιολογήθηκαν με βάση δύο συγκεκριμένες λειτουργίες, την πρόβλεψη της ακριβούς τιμής του bitcoin για κάποια επόμενη μέρα, ή αλλιώς πρόβλημα παλλινδρόμησης (regression task), καθώς και την πρόβλεψη της τάσης για κάποια επόμενη μέρα, ή πρόβλημα κατηγοριοποίησης (classification task), δηλαδή αν κινηθεί θετικά ή αρνητικά η τιμή, πρώτα θα εξηγήσουμε τις μετρικές με τις οποίες έγινε η αξιολόγηση και στην συνέχεια θα σχολιάσουμε τα ίδια τα αποτελέσματα.

5.1 Μετρικές Αξιολόγησης

Για την αξιολόγηση των αποτελεσμάτων των μοντέλων χρησιμοποιήσαμε διάφορες μετρικές. Συγκεκριμένα για το πρόβλημα εύρεσης της τιμής του Bitcoin χρησιμοποιήσαμε τις μετρικές του μέσου τετραγωνικού σφάλματος (MSE), της τετραγωνικής ρίζας του μέσου τετραγωνικού σφάλματος (RMSE), καθώς επίσης και του μέσου απόλυτου ποσοστιαίου σφάλματος (MAPE) και του μέσου απόλυτου σφάλματος (MAE). Επίσης, για το πρόβλημα εκτίμησης της κίνησης του κρυπτονομίσματος, χρησιμοποιήσαμε το ποσοστό ακρίβειας των προβλέψεων.

5.1.1 Μέσο τετραγωνικό σφάλμα MSE

Το μέσο τετραγωνικό σφάλμα αποτελεί το άθροισμα των διαφορών των επιθυμητών τιμών που πρέπει να προβλέψει το μοντέλο σε σχέση με την πραγματική τιμή που πρόέβλεψε και όλο αυτό διαιρεμένο με το πλήθος όλων των παρατηρήσεων. Ο τύπος που το χαρακτηρίζει είναι

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

Όπου \hat{Y}_i η τιμή που πρόέβλεψε το μοντέλο, Y_i η πραγματική τιμή και n το πλήθος των παρατηρήσεων.

5.1.2 Ρίζα μέσου τετραγωνικού σφάλματος RMSE

Αντίστοιχα, η ρίζα του μέσου τετραγωνικού σφάλματος, είναι η τετραγωνική ρίζα της παραπάνω μετρικής και χαρακτηρίζεται από τον τύπο:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$$

5.1.3 Μέσο Απόλυτο Σφάλμα MAE

Το μέσο απόλυτο σφάλμα είναι μία μετρική που χρησιμοποιείται σε προβλήματα παλινδρόμησης και αποτυπώνει τον μέσο όρο των πραγματικών αποστάσεων των προβλέψεων σε σχέση με τις πραγματικές τιμές. Ο τύπος είναι ο εξής:

$$MAE = \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{n}$$

όπου Y_i είναι η πραγματική τιμή και \hat{Y}_i η τιμή που έδωσε το μοντέλο.

5.1.4 Μέσο Απόλυτο Ποσοστιαίο Σφάλμα MAPE

Η συγκεκριμένη μετρική στην ουσία μετατρέπει το παραπάνω σφάλμα σε ποσοστό επί τοις εκατό, ώστε να γνωρίζουμε σε τι βαθμό σφάλει το σύστημα και να έχουμε μια πιο ολοκληρωμένη εικόνα. Ο τύπος είναι ο εξής:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| 100\%$$

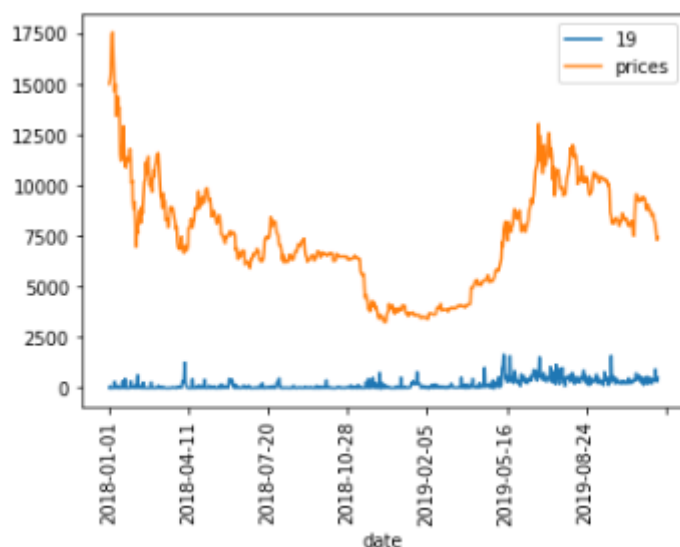
5.2 Αποτελέσματα και αξιολόγηση Μοντέλων και επιλογών δεδομένων

Στην παρούσα ενότητα θα παρουσιάσουμε τα αποτελέσματα των πειραμάτων μας επάνω σε διαφορετικούς συνδυασμούς μοντέλων εξαγωγής συναισθήματος, μοντέλων πρόβλεψης της τιμής, αλλά και διαφορετικών πληροφοριών στην είσοδο των μοντέλων. Με βάση αυτά τα πειράματα εξάγονται χρήσιμα συμπεράσματα για την επιλογή μοντέλων στην εξαγωγή συναισθήματος, στην πρόβλεψη της τιμής αλλά και στην εξαγωγή χρήσιμης πληροφορίας από το Twitter για την επίλυση του προβλήματος πρόβλεψης της τιμής του Bitcoin.

5.2.1 Συσχέτιση των δεδομένων με την τιμή κλεισίματος

Πριν ξεκινήσουμε την ανάλυση των αποτελεσμάτων των πειραμάτων επάνω στην πρόβλεψη της τιμής, κρίνεται σκόπιμο να μελετήσουμε την συσχέτιση διάφορων στηλών του dataset με την τιμή που θέλουμε να προβλέψουμε, δηλαδή την τιμή κλεισίματος κάποιας μελλοντικής μέρας.

Αρχικά, θα εξετάσουμε την περίπτωση συσχέτισης της 19ης ώρας της μέρας με την τιμή κλεισίματος της επόμενης ημέρας. Ο δείκτης συσχέτισης correlation είναι 0.315, κάτι που δείχνει μία χαλαρή συσχέτιση. Να υπενθυμίσουμε εδώ ότι η τιμή της συσχέτισης κυμαίνεται μεταξύ -1 και 1, με 1 να σημαίνει ισχυρή συσχέτιση δηλαδή οι τιμές των δύο μεγεθών αυξάνονται και μειώνονται με τον ίδιο τρόπο στο πέρασμα του χρόνου, ενώ -1 σημαίνει ότι όταν η μία τιμή αυξάνεται με κάποιο ρυθμό, η άλλη μειώνεται με τον ίδιο ρυθμό. Στο σχήμα 5.1 φαίνεται η γραφική της τιμής κλεισίματος και του συναισθήματος στις 7μμ σαν συνάρτηση της ημερομηνίας. Όπως παρατηρούμε από την γραφική, μετά την περίοδο του 2018 αρχίζει

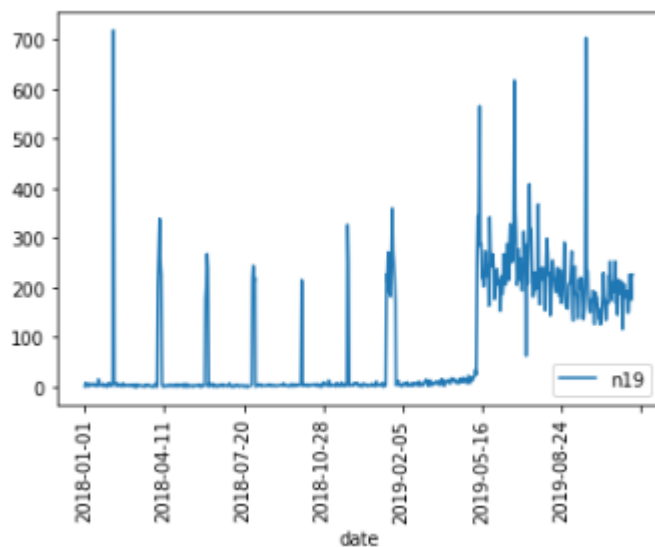


Σχήμα 5.1: Γραφική παράσταση τιμής κλεισίματος και συναισθήματος

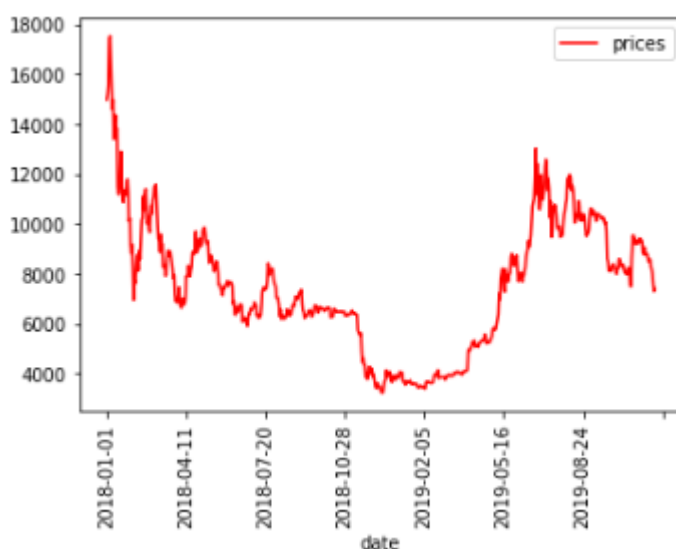
να φαίνεται μία διακύμανση στην τιμή του συναισθήματος ανάλογη με την διακύμανση της τιμής του κρυπτονομίσματος. Μία ερμηνεία για τον λόγο που συμβαίνει αυτό μόνο την συγκεκριμένη περίοδο και όχι από την αρχή των ημερομηνιών του dataset είναι ότι σε εκείνο το σημείο η πληροφορία είναι περισσότερη καθώς παρατηρείται αισθητά μεγαλύτερος αριθμός tweets σε σχέση με προηγούμενες περιόδους. Αυτό μας κάνει αισιόδοξους να πιστέψουμε ότι με dataset με περισσότερα tweets η πληροφορία θα έχει αρκετά μεγαλύτερη συσχέτιση με την τελική τιμή-στόχο.

Μία ακόμα παράμετρος της οποίας η συσχέτιση με την τιμή-στόχο έχει αξία, είναι το πλήθος των tweets ανά ώρα της ημέρας. Έτσι λοιπόν παρατηρούμε ότι το πλήθος των tweets στις 7μμ έχει μία συσχέτιση 0.35 με την τιμή κλεισίματος της επόμενης ημέρας και η γραφική της σε σχέση με την τιμή φαίνεται στα σχήματα 5.2 και 5.3.

Από αυτές τις γραφικές επιβεβαιώνεται και η υπόθεση που κάναμε, ότι για την μεγαλύτερη συσχέτιση μετά από μία συγκεκριμένη ημερομηνία ευθύνεται και το πλήθος των ποσταρισμάτων στο Twitter καθώς φαίνεται να αυξάνονται και μάλιστα να έχουν κοινά στοιχεία οι διακυμάνσεις του πλήθους των tweets και της τιμής κλεισίματος. Αυτές οι παρατηρήσεις είναι πολύ σημαντικές καθώς φαίνεται ότι υπάρχει πληροφορία μέσα από την πλατφόρμα του Twitter γενικότερα, την οποία μπορούμε να εκμεταλλευτούμε για την πρόβλεψη της τιμής του Bitcoin.



Σχήμα 5.2: Γραφική παράσταση πλήθους tweets συναρτήση του χρόνου



Σχήμα 5.3: Γραφική παράσταση τιμής κλεισίματος συναρτήση του χρόνου

5.2.2 Πειραματικά Αποτελέσματα

Για την εξαγωγή των αποτελεσμάτων δοκιμάστηκαν 2 κλασικά μοντέλα πρόβλεψης χρονοσειρών. Το πρώτο αποτελείται από δύο επίπεδα LSTM με 32 μονάδες το κάθε ένα και ένα πλήρως συνδεδεμένο επίπεδο με μία έξοδο για την τελική πρόβλεψη. Στο δεύτερο μοντέλο χρησιμοποιήθηκαν 2 επίπεδα GRU με 256 μονάδες και ένα πλήρως συνδεδεμένο επίπεδο το οποίο είχε μία έξοδο (βλ σχήμα 5.12).

Επίσης, για την εξαγωγή συναισθήματος χρησιμοποιήθηκαν δύο μοντέλα. Το πρώτο ήταν με την στατική μέθοδο του Vader και το δεύτερο με χρήση της βελτιωμένης εκδοχής του BERT το RoBERTa.

Πίνακας 5.1: Πίνακας με ποσοστά ακριβείας με χρήση RoBERTa σε GRU

Δεδομένα/accuracy	1 ημέρα	3 ημέρες	7 ημέρες
μόνο technical	0.615	0.558	0.6346
μόνο συναίσθημα 1ης μέρας	0.59	0.586	0.548
Συναίσθημα + πλήθος 1ης μέρας	0.577	0.596	0.548
Συναίσθημα 3 ημερών + πλήθος	0.587	0.606	0.572
Συναίσθημα + Πλήθος 3 ημερών + technical	0.587	0.577	0.644

Πίνακας 5.2: Πίνακας με την μετρική MAPE με χρήση RoBERTa σε GRU

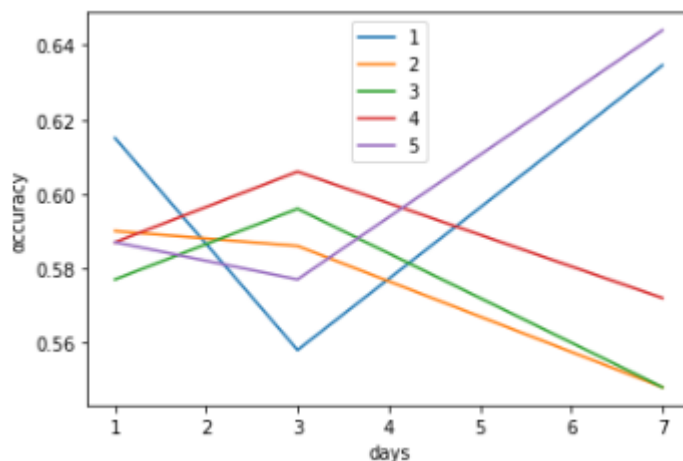
Δεδομένα/MAPE	1 ημέρα	3 ημέρες	7 ημέρες
μόνο technical	2.71	4.881	7.48
μόνο συναίσθημα 1ης μέρας	26.28	26.45	26.29
Συναίσθημα + πλήθος 1ης μέρας	25.55	26.567	26.36
Συναίσθημα 3 ημερών + πλήθος	25.47	25.78	25.00
Συναίσθημα + Πλήθος 3 ημερών + technical	2.78	6.67	7.39

Στον πίνακα 5.1 βλέπουμε την ακρίβεια που επιτυγχάνουν τα μοντέλα πρόβλεψης για 1,3 και 7 μέρες μετά, για το πρόβλημα της κατηγοριοποίησης σε θετική ή αρνητική κίνηση, και σε όποιες περιπτώσεις γίνεται χρήση συναισθήματος, η εξαγωγή από τα tweets γίνεται με χρήση του μοντέλου RoBERTa. Στις περιπτώσεις που αναφέρουμε "μόνο technical" σημαίνει ότι χρησιμοποιήθηκαν μόνο οι τεχνικοί δείκτες της τιμής κλεισίματος, τιμή ανοίγματος, του όγκου συναλλαγών και όλων των άλλων στηλών του historical prices data. Στα σημεία που αναφέρεται συναίσθημα 1ης μέρας, έχουμε 24 στήλες με το μέσο συναίσθημα για κάθε ώρα της ημέρας, συνολικά 24 στήλες, ενώ στο συναίσθημα 3 ημερών έχουμε τις ίδιες 24 στήλες και επιπλέον το συναίσθημα για τις ώρες 11, 13, 15, 17, 18 των προηγούμενων δύο ημερών. Επιλέξαμε μόνο τις συγκεκριμένες ώρες από τις προηγούμενες ημέρες καθώς ήταν οι ώρες με την μεγαλύτερη συσχέτιση με την τιμή κλεισίματος. Αντίστοιχα το ίδιο ισχύει και για τα σημεία που έχουμε σημειώσει πλήθος, δηλαδή 24 στήλες με το πλήθος των tweets ανά ώρα για την πρώτη ημέρα, και επιλογή συγκεκριμένων ωρών για τις προηγούμενες ημέρες. Ο πίνακας 5.2 περιέχει τον δείκτη MAPE για το πρόβλημα της παλινδρόμησης για προσέγγιση της τιμής του Bitcoin για τα ίδια πειράματα με τον πίνακα 5.1. Τα πειράματα έγιναν με νευρωνικό που είχε δύο στρώματα GRU και ένα τελικό στρώμα πλήρως συνδεδεμένο με μία έξοδο. Επίσης να αναφέρουμε ότι η πρόβλεψη έγινε σε δύο βήματα. Αρχικά το μοντέλο είχε ως στόχο την πρόβλεψη της ακριβούς τιμής της μετοχής για την ημέρα στόχο και αυτό ήταν το πρόβλημα της παλινδρόμησης, και στην συνέχεια ανάλογα με το αν η διαφορά της προβλεπόμενης τιμής σε σχέση με την τρέχουσα τιμή κλεισίματος ήταν θετική ή αρνητική, γινόταν και η αντίστοιχη πρόβλεψη για το αν η τάση θα είναι θετική ή αρνητική. Αντίστοιχα η πραγματική τιμή της τάσης της τιμής εξαγόταν από την πραγματική διαφορά της μέρας στόχου και της τρέχουσας ημέρας.

Παρατηρήσεις

Αρχικά, από τον πίνακα 5.1 παρατηρούμε τα εξής:

- Αρχικά, παρατηρούμε ότι παρόλο που θα περιμέναμε πως τα καλύτερα αποτελέσματα στις προβλέψεις θα ερχόντουσαν στην πρόβλεψη της επόμενης μέρας, αντίθετα η ακρίβεια παρουσιάζει διακυμάνσεις σε σχέση με το χρονικό διάστημα πρόβλεψης, και παρατηρούμε μείωση στην ακρίβεια από την πρώτη στην τρίτη μέρα και έπειτα αύξηση στην 7η. Η διακύμανση αυτή φαίνεται στο σχήμα 5.4. Αυτό πιθανόν οφείλεται στο γεγονός ότι η αίσθηση που επικρατεί στο ευρύ κοινό χρειάζεται κάποιο χρόνο για να αποτυπωθεί σε νούμερα στην αγορά των κρυπτονομισμάτων και των μετοχών, όμως υπάρχει σίγουρα κάποια συσχέτιση.



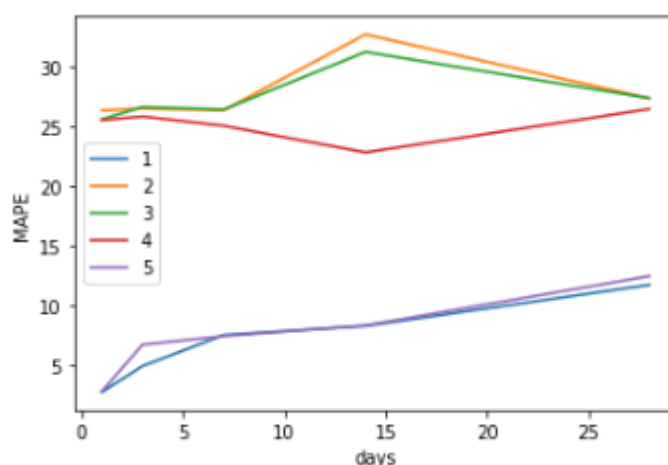
Σχήμα 5.4: Ακρίβεια μοντέλων πίνακα 5.1 συναρτήσει ημερών

- Επίσης παρατηρείται ότι για διαφορετικά χρονικά διαστήματα πχ για 3 και για 7 ημέρες ο συνδυασμός πληροφορίας που δίνει το καλύτερο αποτέλεσμα αλλάζει. Αυτό πιθανόν οφείλεται στο γεγονός ότι τα χαρακτηριστικά έχουν διαφορετική πληροφορία μέσα τους σε σχέση με την μελλοντική τιμή, και δεν είναι όλα συσχετισμένα το ίδιο σε σχέση με το μέλλον.
- Σημαντική παρατήρηση είναι ότι οι τεχνικοί δείκτες περιέχουν πολύ κρίσιμη πληροφορία για την κίνηση της τιμής, ειδικά για πιο βραχυχρόνιες προβλέψεις, όμως αφού σκοπός της διπλωματικής είναι η χρήση πληροφορίας από το Twitter και ιδιαίτερα η εξαγωγή συναισθήματος, επικεντρωνόμαστε περισσότερο σε αυτό. Έτσι, παρατηρούμε ότι με χρήση των στηλών του συναισθήματος συμπληρωματικά με τους τεχνικούς δείκτες, η επίδοση του μοντέλου βελτιώνεται, όπως φαίνεται σε όλα τα μοντέλα με χρήση συναισθήματος σε πρόβλεψη για τρεις και δεκατέσσερες μέρες.
- Ένα ακόμα σημαντικό συμπέρασμα που εξάγεται από αυτά τα αποτελέσματα είναι ότι πέρα από το μέσο συναίσθημα των tweets, σχετίζεται και το ίδιο το πλήθος των δημοσιεύσεων με την πρόβλεψη της τιμής και βλέπουμε ότι με την χρήση αυτής της πληροφορίας αυξάνεται και η επίδοση στην ακρίβεια (η βέλτιστη τιμή ακριβείας 0.75, επιτυγχάνεται με χρήση τόσο συναισθήματος όσο και πλήθους των tweets αλλά και τεχνικών δεικτών).

Στην συνέχεια, για τον πίνακα 5.2 παρατηρούμε:

- Σε ό,τι αφορά την μετρική MAPE που αφορά το πρόβλημα της παλλινδρόμησης, παρατηρούμε ότι σημειώνει σταθερή άνοδο με την αύξηση του χρονικού διαστήματος της πρόβλεψης, ιδιαίτερα όταν περιέχονται οι τεχνικοί δείκτες στο σύνολο των δεδομένων. Αυτό μας δείχνει ότι παρόλο που το μοντέλο μπορεί να βρίσκει σωστά την τάση, δηλαδή αν κινηθεί θετικά ή αρνητικά η τιμή του κρυπτονομίσματος, όσο μεγαλώνει το διάστημα τόσο αδυνατεί να προσεγγίσει την τιμή που θα έχει το Bitcoin. Μία ερμηνεία είναι ότι στο συναίσθημα μπορεί να υπάρξει πληροφορία για την γενικότερη τάση, αλλά το μοντέλο δεν είναι σε θέση να ποσοτικοποιήσει πλήρως αυτές τις μεταβολές.
- Μια ακόμα παρατήρηση, είναι ότι σαφώς όταν περιέχεται σαν μόνη πληροφορία το συναίσθημα το σκορ του MAPE είναι πολύ χειρότερο, καθώς δεν υπάρχουν οι προηγούμενες τιμές κλεισίματος, οι οποίες είχαν συσχέτιση με την μελλοντική τιμή έως και 0.99%. Παρόλα αυτά, παρατηρείται οριακή βελτίωση σε αυτή την μετρική με την χρήση του πλήθους των tweets αυτής της συγκεκριμένης ώρας, κάτι που δείχνει ξανά την χρησιμότητα αυτής της πληροφορίας (βλ. καμπύλες 2,3,4 στο σχήμα 5.5).

Η διακύμανση των τιμών της μετρικής MAPE φαίνεται στο σχήμα 5.5.

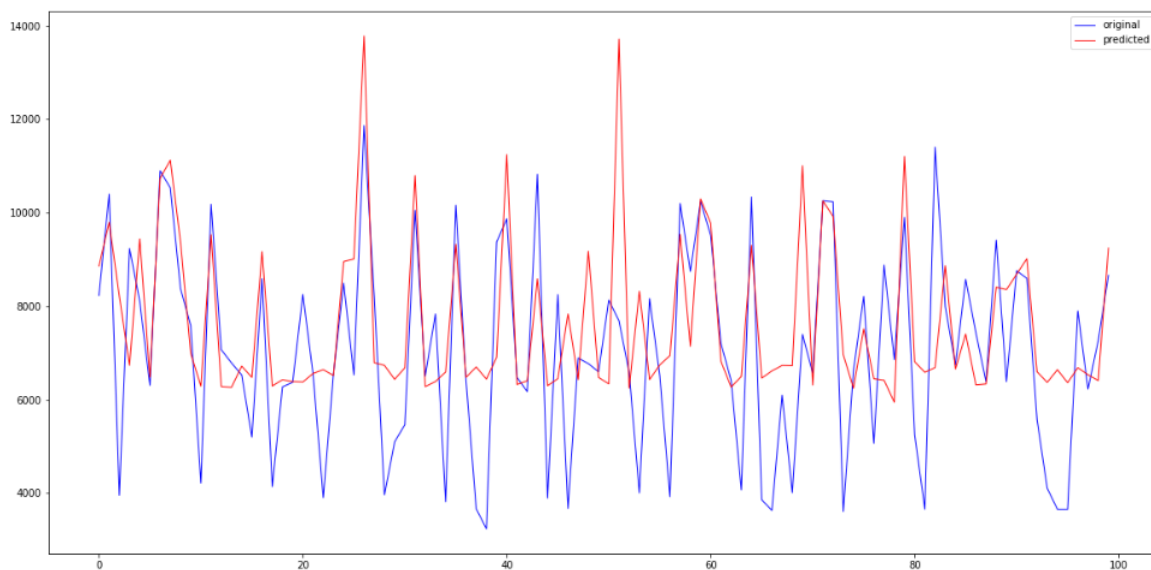


Σχήμα 5.5: MAPE μοντέλων πίνακα 5.2 συναρτήσει ημερών

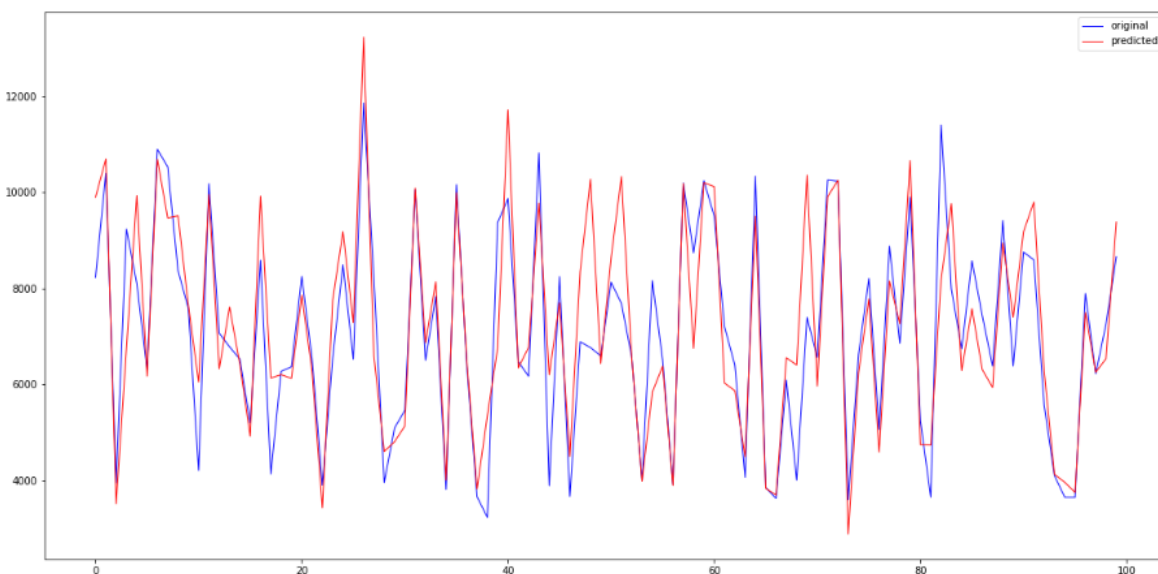
Στα σχήματα 5.6 και 5.7 φαίνονται οι γραφικές παραστάσεις για τις πραγματικές και τις προβλεπόμενες τιμές του μοντέλου για τις περιπτώσεις των συνδυασμών συναίσθημα 3 ημερών και συναίσθημα μαζί με πλήθος και τεχνικούς δείκτες για διάστημα 28 μέρες μετά.

Vader

Μία άλλη παράμετρος που εξετάσαμε στο αν επηρεάζει ή όχι την επίδοση των μοντέλων είναι ο τρόπος εξαγωγής του συναισθήματος από τα Tweets. Έτσι, εκτός από τα βαθιά νευρωνικά δίκτυα που βασίζονται στο σύστημα των transformers, πειραματιστήκαμε και με εξαγωγή συναισθήματος με κλασσικούς μεθόδους όπως το Vader, που είναι ένα στατικό μέσο εξαγωγής συναισθήματος. Στους πίνακες 5.3 και 5.4 παραθέτονται τα αποτελέσματα αυτών των πειραμάτων.



Σχήμα 5.6: Διάγραμμα πραγματικών και προβλεπμένων τιμών για 7 μέρες μετά με συναίσθημα 3 ημερών RoBERTa σε GRU



Σχήμα 5.7: Διάγραμμα πραγματικών και προβλεπμένων τιμών για 7 μέρες μετά με συνδυασμό συναισθηματος, πλήθους και τεχνικών δεικτών RoBERTa σε GRU

Πίνακας 5.3: Πίνακας με την μετρική Accuracy με χρήση Vader σε GRU

Δεδομένα/Accuracy	1 ημέρα	3 ημέρες	7 ημέρες
μόνο συναίσθημα 1ης μέρας	0.49	0.567	0.592
Συναίσθημα + πλήθος 1ης μέρας	0.569	0.55	0.611
Συναίσθημα 3 ημερών + πλήθος	0.586	0.596	0.611
Συναίσθημα + Πλήθος 3 ημερών + technical	0.549	0.538	0.543

Πίνακας 5.4: Πίνακας με την μετρική MAPE με χρήση Vader σε GRU

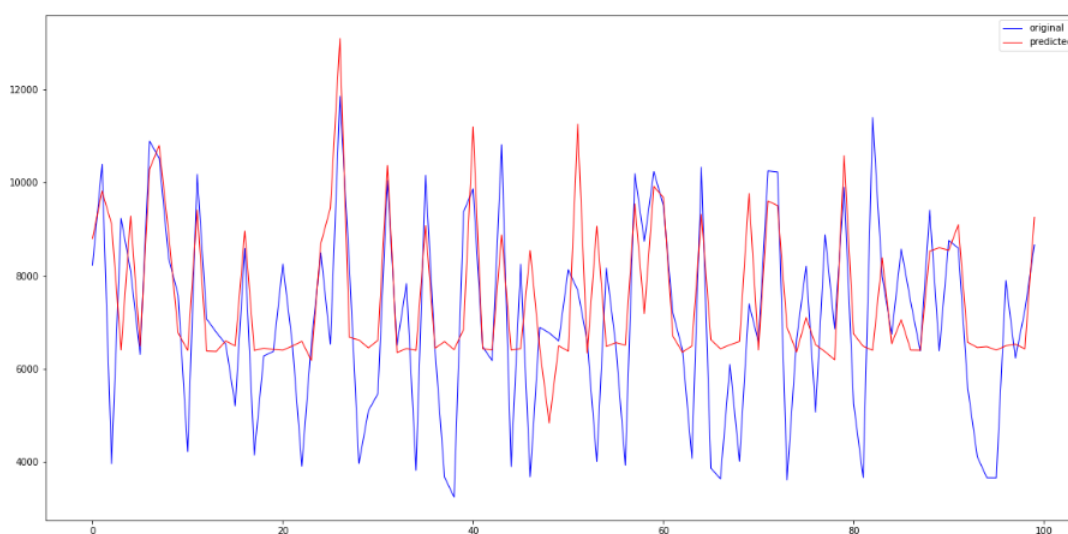
Δεδομένα/MAPE	1 ημέρα	3 ημέρες	7 ημέρες
μόνο συναίσθημα 1ης μέρας	29.89	30.502	27.18
Συναίσθημα + πλήθος 1ης μέρας	25.04	27.38	26.30
Συναίσθημα 3 ημερών + πλήθος	27.17	27.00	25.39
Συναίσθημα + Πλήθος 3 ημερών + technical	2.63	5.31	8.50

Συγκρίνοντας τα αποτελέσματα μεταξύ Vader και RoBERTa δεν μπορούμε να αποφανθούμε ξεκάθαρα υπέρ της υπερίσχυσης κάποιου σε κανένα από τα δύο προβλήματα που εξετάζουμε. Εξετάζοντας επί μέρους τα αποτελέσματα, μπορούμε να πούμε ότι σε βραχυχρόνιες περιόδους το Vader έχει πιο μικρή επίδοση και στις δύο μετρικές, ενώ το ίδιο ισχύει στον συνδυασμό με τους τεχνικούς δείκτες.

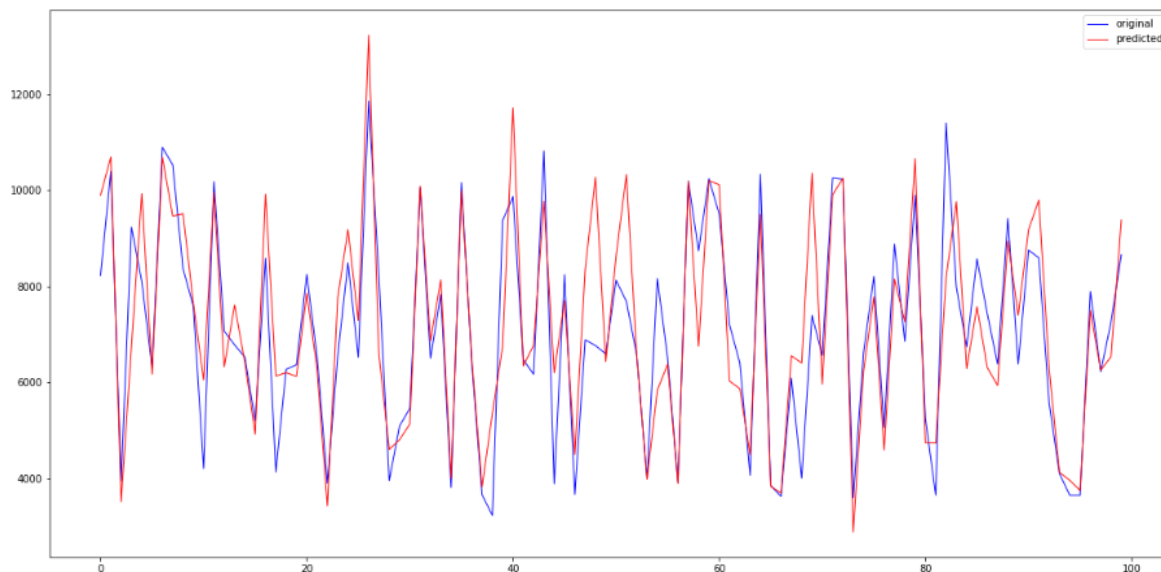
Αντίθετα, όταν χρησιμοποιείται μόνο συναίσθημα σαν πληροφορία για πρόβλεψη της τιμής, το Vader φαίνεται να επιτυγχάνει αρκετά υψηλά σκορ στο κομμάτι της ακρίβειας αν και υστερεί στην ακριβή πρόβλεψη της τιμής (βλ. πρόβλεψη για 7 μέρες με συναίσθημα + πλήθος).

Επομένως, δεν θα μπορούσαμε να αποφανθούμε γενικά ότι για το πρόβλημα πρόβλεψης της τιμής του Bitcoin η χρήση των transformers εξασφαλίζει καλύτερα αποτελέσματα, παρόλο που σε συγκεκριμένα πειράματα κατάφερε μεγαλύτερο σκορ σε σχέση με το Vader, δεν υπερέχει σε όλες τις περιπτώσεις.

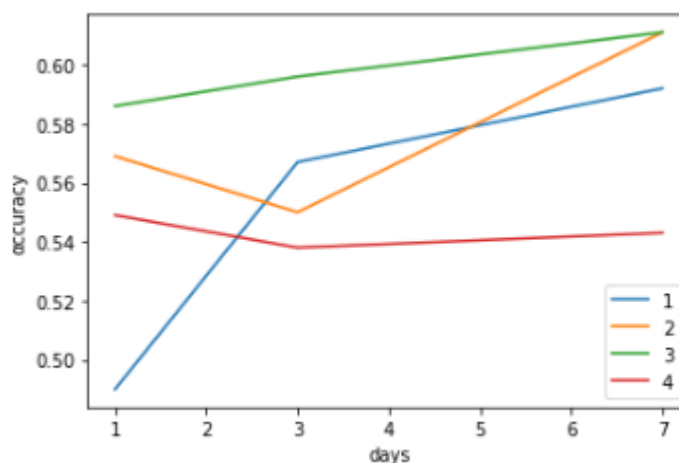
Στα σχήματα 5.8 και 5.9 φαίνονται οι γραφικές παραστάσεις για τις πραγματικές και τις προβλεπόμενες τιμές για 28 μέρες μπροστά με χρήση του συναισθήματος 3 ημερών στη μία και συνδυασμού συναισθήματος, πλήθους και τεχνικών δεικτών στην άλλη.



Σχήμα 5.8: Διάγραμμα πραγματικών και προβλεπόμενων τιμών για 7 μέρες μετά με συναίσθημα 3 ημερών Vader σε GRU



Σχήμα 5.9: Διάγραμμα πραγματικών και προβλεπμένων τιμών για 7 μέρες μετά με συνδυασμό συναισθήματος με Vader, πλήθους και τεχνικών δεικτών σε GRU

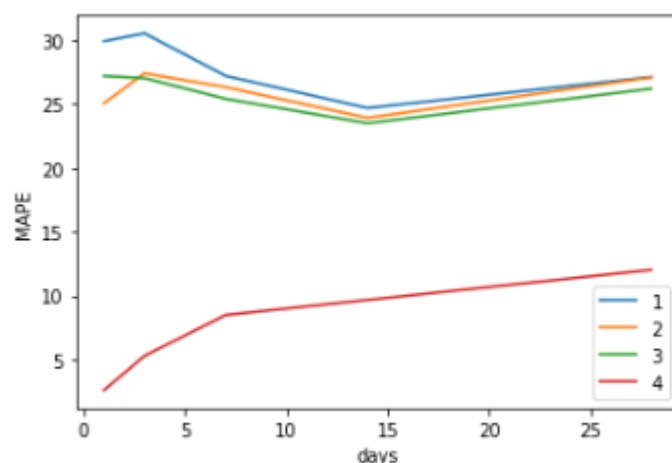


Σχήμα 5.10: Accuracy μοντέλων πίνακα 5.3 συναρτήσει ημερών

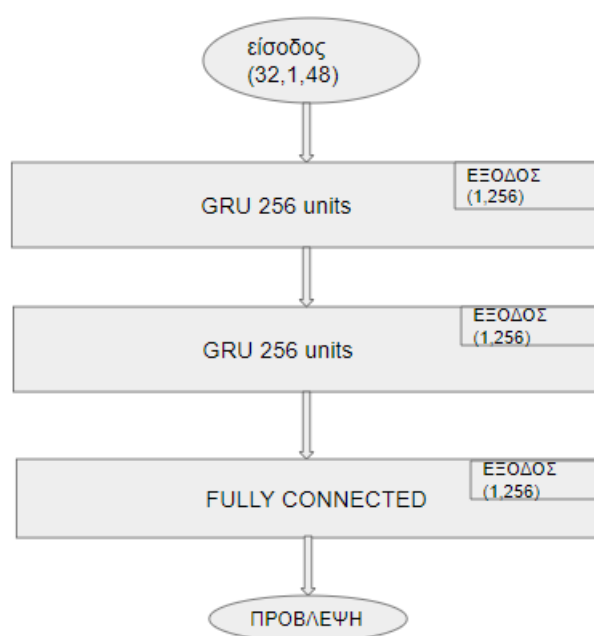
Σύγκριση LSTM με GRU

Όπως αναφέραμε στην αρχή του κεφαλαίου, πειραματισμός έγινε και επάνω στο μοντέλο πρόβλεψης το οποίο χρησιμοποιείται για την εξαγωγή της τελικής τιμής. Έτσι, εκτός από τα δίκτυα με 2 επίπεδα GRU αποτελούμενα από 256 μονάδες το κάθε επίπεδο, κάναμε και κάποιες δοκιμές με δίκτυα αποτελούμενα από LSTM επίπεδα, και 32 μονάδες το κάθε επίπεδο. Ο λόγος που δοκιμάσαμε λιγότερες μονάδες σε αυτά τα επίπεδα ήταν ότι πιθανόν τα μεγαλύτερα δίκτυα να μην μπορούσαν να εκπαιδευτούν επαρκώς λόγω του μικρού όγκου των δεδομένων και της μεγάλης χωρητικότητάς τους. Η δομή του δικτύου φαίνεται στο σχημα 5.12, με είσοδο δεδομένων με 48 στήλες.

Με βάση την παραπάνω δομή προσαρμοσμένη στο δίκτυο LSTM που περιγράψαμε, παραθέτουμε τα αποτελέσματα από τα πειράματα στους πίνακες 5.5 και 5.6:



Σχήμα 5.11: *MAPE* μοντέλων πίνακα 5.4 συναρτήσει ημερών



Σχήμα 5.12: *Δομή δικτύου πρόβλεψης με μονάδες GRU*

Όπως και στο GRU, παρατηρούμε ότι το MAPE score αυξάνεται όσο αυξάνεται το διάστημα στο οποίο κάνουμε πρόβλεψη, ενώ και πάλι είναι αισθητά μικρότερο όταν χρησιμοποιούμε τεχνικούς δείκτες. Οι τιμές του φαίνεται να κυμαίνονται στο ίδιο επίπεδο με τα GRU

Στο κομμάτι του Accuracy, και συγκρίνοντας τα αποτελέσματα με τα πειράματα του GRU που είχαν ίδιες στήλες δεδομένων, φαίνεται ότι το GRU έχει ελαφρώς καλύτερη επίδοση, όμως υπάρχουν ερωτηματικά για το αν είναι χρήσιμη η έξτρα χωρητικότητα που έχει. Ενδιαφέρουσα παρατήρηση ωστόσο είναι ότι η καλύτερη επίδοση στην ακρίβεια επιτυγχάνεται με το δίκτυο LSTM όταν σαν είσοδο έχει το συναίσθημα και το πλήθος των τριών τελευταίων ημερών μαζί με τους τεχνικούς δείκτες. Άρα, ακόμα και ένα αρκετά μικρότερο δίκτυο καταφέρνει να έχει πολύ ικανοποιητικά αποτελέσματα.

Στα σχήματα 5.13 και 5.14 φαίνονται ξανά τα διαγράμματα των πραγματικών και προ-

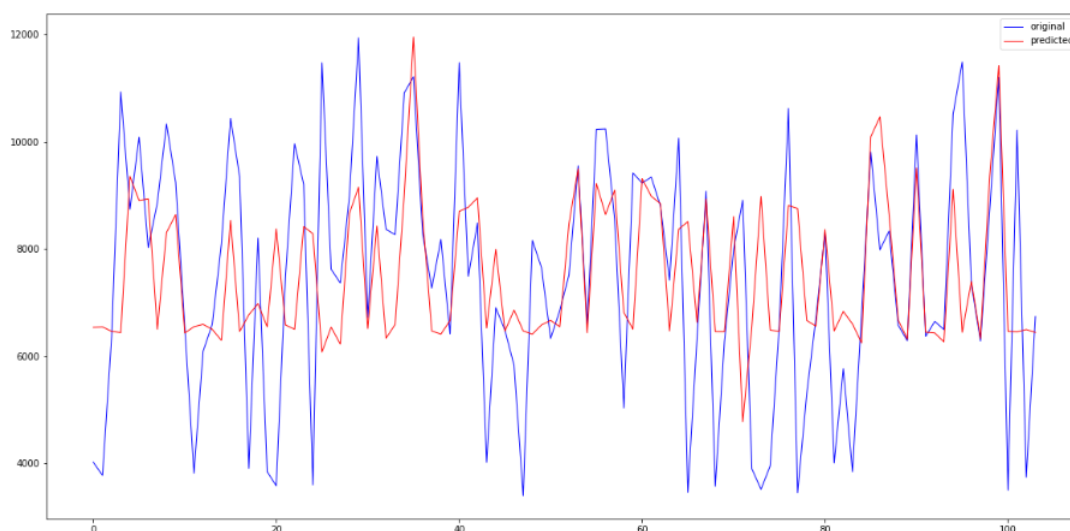
Πίνακας 5.5: Πίνακας με την μετρική Accuracy με χρήση Vader σε LSTM

Δεδομένα/Accuracy	1 ημέρα	3 ημέρες	7 ημέρες
μόνο συναίσθημα 1ης μέρας	0.62	0.538	0.537
Συναίσθημα 3 ημερών + πλήθος	0.61	0.60	0.586
Συναίσθημα + Πλήθος 3 ημερών + technical	0.55	0.586	0.577

Πίνακας 5.6: Πίνακας με την μετρική MAPE με χρήση Vader σε LSTM

Δεδομένα/Accuracy	1 ημέρα	3 ημέρες	7 ημέρες
μόνο συναίσθημα 1ης μέρας	28.66	27.61	27.26
Συναίσθημα 3 ημερών + πλήθος	26.71	27.07	26.99
Συναίσθημα + Πλήθος 3 ημερών + technical	2.76	6.35	7.69

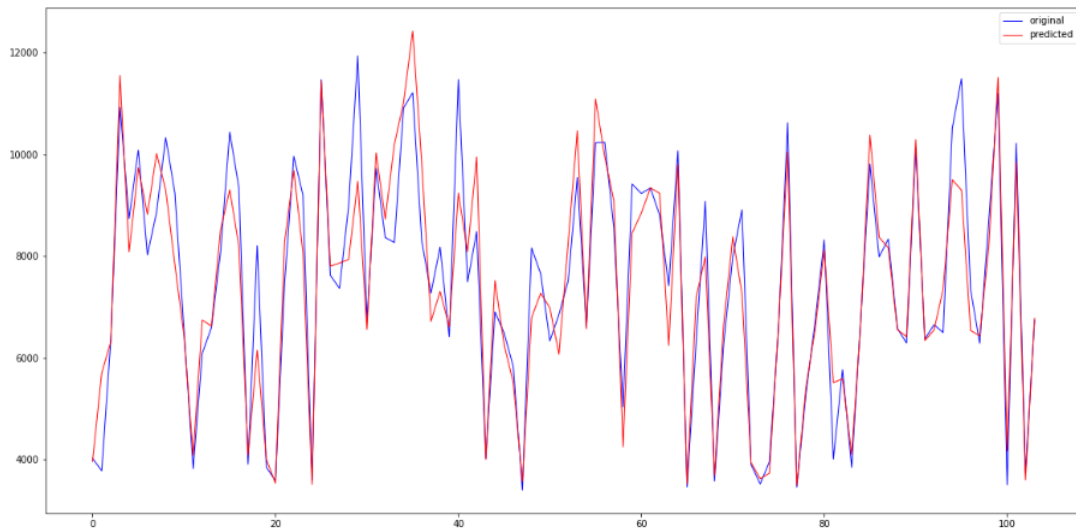
βλευμένων τιμών, ενώ στα σχήματα 5.15 και 5.16 οι γραφικές της ακρίβειας και του MAPE score αντίστοιχα σε συνάρτηση των ημερών.



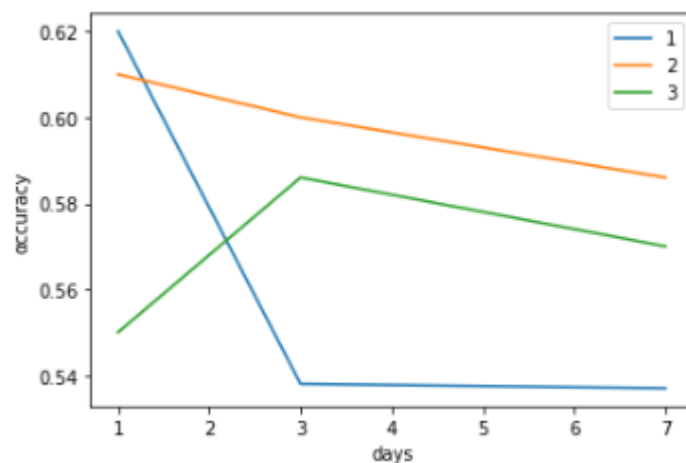
Σχήμα 5.13: Διάγραμμα πραγματικών και προβλεπόμενων τιμών 7 ημερών με χρήση συναίσθηματος με Vader 3 ημερών σε LSTM

Γενικά Συμπεράσματα

Σε γενικές γραμμές μπορούμε να πούμε ότι σε ό,τι αφορά το πρόβλημα της πρόβλεψης της τάσης (trend) της τιμής, το συναίσθημα σε συνδυασμό με άλλα δεδομένα όπως οι τεχνικοί δείκτες και ο όγκος των αναρτήσεων, έχει θετική επίδραση και μπορεί να αυξήσει την ακρίβεια. Επίσης σημαντικό είναι ότι επιτυγχάνεται ακρίβεια σε χρονικό ορίζοντα πιο μεγάλο απ' ό,τι στο πρόβλημα της παλινδρόμησης (7 μέρες), και έτσι μπορούμε να έχουμε μία εικόνα της κίνησης της τιμής πιο μεσοπρόθεσμη. Το γεγονός ότι η ακρίβεια δεν μειώνεται πάντα καθώς αυξάνονται οι ημέρες πρόβλεψης οφείλεται στο γεγονός ότι το συναίσθημα που κυριαρχεί στον κόσμο χρειάζεται κάποιες μέρες να αποτυπωθεί, όπως παρατηρούν και στην έρευνά τους οι Jacques Vella Critien, Albert Gatt, Joshua Ellul [12].



Σχήμα 5.14: Διάγραμμα πραγματικών και προβλεπμένων τιμών 7 ημερών με χρήση συνδυασμού συναισθήματος 3 ημερών με Vader, πλήθους δημοσιεύσεων και τεχνικών δεικτών σε LSTM

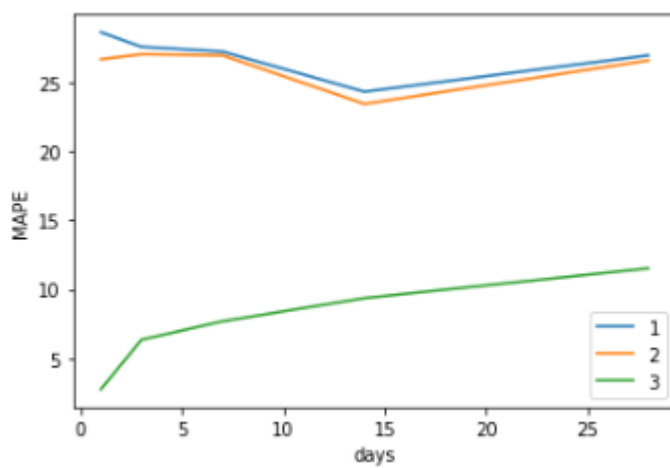


Σχήμα 5.15: Διάγραμμα του accuracy σε συνάρτηση των ημερών πρόβλεψης με Vader σε LSTM

Επίσης, μπορεί στο πρόβλημα της παλινδρόμησης οι τεχνικοί δείκτες να έχουν πολύ μεγάλη επίδραση δεν παύει όμως να παρατηρείται και εκεί μικρή βελτίωση στο MAPE score με χρήση του συναισθήματος.

Ακόμα, μπορεί το υψηλότερο ποσοστό ακρίβειας να είναι λίγο παραπάνω από το 64%, το οποίο δεν είναι ευκαταφρόνητο αλλά μπορεί να μην ξεπερνάει και σε επιτυχία κάποιον έμπειρο άνθρωπο επάνω στον κλάδο, όμως συνδυαστικά, ίσως το δίκτυο να μπορεί να προβλέπει καταστάσεις που η ανθρώπινη εμπειρία να μην μπορεί, καθώς και το αντίστροφο. Έτσι, καταλήγουμε ότι είναι χρήσιμη η ανάπτυξη και βελτίωση αυτών των μοντέλων.

Τέλος πειράματα έγιναν και για 14 και 28 ημέρες όπου παρατηρήθηκαν ακανόνιστες μεταβολές ως προς το κομμάτι της πρόβλεψης της κίνησης, ενώ στο κομμάτι της πρόβλεψης της τιμής παρατηρήθηκε συνεχιζόμενη αύξηση του MAPE score.



Σχήμα 5.16: Διάγραμμα του MAPE score σε συνάρτηση των ημερών πρόβλεψης με Vader σε LSTM

Μέρος **III**

Επίλογος

Κεφάλαιο 6

Επίλογος

Η παραπάνω εργασία σαν σκοπό είχε την μελέτη του συνδυασμού της πληροφορίας που μπορεί να αντληθεί από την πλατφόρμα του Twitter μαζί με τεχνικές πληροφορίες γύρω από το Bitcoin με τελικό σκοπό την πρόβλεψη της τιμής και της κίνησης της τιμής του κρυπτονομίσματος.

Μερικά χρήσιμα συμπεράσματα που προκύπτουν είναι ότι το συναίσθημα σε συνδυασμό με τους τεχνικούς δείκτες μπορεί να δώσει αρκετά ικανοποιητικά αποτελέσματα τόσο στο πρόβλημα της εκτίμησης της τάσης όσο και στο πρόβλημα της πρόβλεψης της τιμής. Ιδιαίτερα για το δεύτερο πρόβλημα, η βελτίωση παρατηρείται σε πιο βραχυπρόθεσμες προβλέψεις, ενώ για το πρώτο σε πιο μεσοπρόθεσμες. Επίσης, συμπεράναμε ότι και η χρήση άλλων πληροφοριών από την πλατφόρμα του Twitter, όπως το πλήθος των δημοσιεύσεων ανά ώρα, μπορεί να φανεί χρήσιμη στα προβλήματα που θέλαμε να επιλύσουμε. Οι επιπλέον πληροφορίες του συνόλου των δεδομένων από το Twitter όπως ο αριθμός των likes κάθε δημοσίευσης, είχαν και αυτές την χρησιμότητά τους, καθώς οδήγησαν στο να ξεχωρίσουμε τα δεδομένα σε πιο σημαντικά και λιγότερο σημαντικά.

Σε γενικές γραμμές μέσα από αυτή την εργασία, αναδύκνείται η δύναμη που έχει η πληροφορία στην μηχανική μάθηση και ότι σημασία δεν έχουν μόνο τα μοντέλα που θα χρησιμοποιηθούν για την πρόβλεψη, αλλά και η μορφή της πληροφορίας και με ποιον τρόπο θα την αξιοποιήσουμε στο έπακρο. Πρόκειται επομένως για μία μελέτη με επίκεντρο περισσότερο τα δεδομένα και λιγότερο τα μοντέλα βαθιάς μάθησης που χρησιμοποιήθηκαν, χωρίς να υποτιμάται η χρησιμότητά τους.

Ένα δεύτερο σκέλος στο οποίο εστίασε η εργασία ήταν το κομμάτι της εξαγωγής συναισθήματος, το οποίο παρόλο που τελικά δεν είχε την δραστική αλλαγή που θα θέλαμε στην πρόβλεψη της τιμής, δεν παύει να αποτελεί σημείο ενδιαφέροντος και καινοτομίας, καθώς αυτή την στιγμή μέθοδοι όπως οι Transformers είναι στην αιχμή του δόρατος της προόδου στον κλάδο της τεχνητής νοημοσύνης.

6.0.1 Μελλοντικές Επεκτάσεις

Σαφώς, η παραπάνω εργασία δεν καλύπτει όλους τους τρόπους πρόβλεψης της τιμής ενός κρυπτονομίσματος ή μετοχής, και αντίθετα θα λέγαμε ότι δημιουργεί περισσότερα αντικείμενα για μελέτη. Έτσι μερικές από τις προτάσεις που θα αφορούσαν σε μελλοντική επέκταση της εργασίας θα ήταν η χρήση και άλλων πληροφοριών στην πρόβλεψη της τιμής του κρυπτονομίσματος πέρα από το ιστορικό των τιμών και το συναίσθημα, όπως θα ήταν στοιχεία

για την φύση του κρυπτονομίσματος, ο καθημερινός ρυθμός δημιουργίας νέων κρυπτονομισμάτων και παρεμφερή στοιχεία, ενώ επίσης θα είχε ενδιαφέρον και η επιροή που έχει στην τιμή άλλων κρυπτονομισμάτων το αίσθημα γύρω από το Bitcoin καθώς είναι το κυρίαρχο κρυπτονόμισμα και όλα τα υπόλοιπα φαίνεται να ακολουθούν την πορεία του. Ακόμα, σε ότι αφορά στο κομμάτι της εξαγωγής συναισθήματος, το Twitter αποτελεί μία από τις πολλές πλατφόρμες στις οποίες θα μπορούσε κάποιος να ψάξει για να βρει το γενικό συναίσθημα που επικρατεί, ενώ ενδιαφέρουσα θα ήταν και η μελέτη επάνω σε πιο έγκυρα μέσα, όπως πρακτορεία ειδήσεων, για να μην υπάρχει σύγχυση μεταξύ έγκυρων ή όχι χρηστών. Τέλος, σε ό,τι αφορά το κομμάτι πρόβλεψης μετοχών, θα είχε ενδιαφέρον μία μελέτη για την εκτίμηση πιο άγνωστων μετοχών, οι οποίες δεν σχολιάζονται τόσο πολύ σε έγκυρα μέσα και έτσι θα ήταν δύσκολη η απόκτηση εικόνας για την τάση τους από αυτά, για να δούμε αν μπορεί το Twitter να δώσει μία λύση σε αυτά. Η ιδέα πίσω από αυτό είναι ότι στο Twitter υπάρχουν πολύ περισσότεροι χρήστες που δημοσιεύουν απ' ότι ειδησεογραφικά μέσα, και έτσι αποτελεί μία πολυποίκιλη πηγή πληροφοριών.

Βιβλιογραφία

- [1] Pradip, Samuel. *CROSS-VALIDATION IN TIME SERIES MODEL*. <https://medium.com/@pradip.samuel/cross-validation-in-time-series-model-b07fba65db7>[Online; accessed 5-October-2022].
- [2] *The Perceptron*. <https://towardsdatascience.com/the-perceptron-3af34c84838c>,[Online; accessed 2-August-2022].
- [3] Mohamed Zahran. *A hypothetical example of mlp*, 2022. https://www.researchgate.net/figure/A-hypothetical-example-of-Multilayer-Perceptron-Network_fig4_303875065,[Online; accessed 10-August-2022].
- [4] fdeloche. *File:Recurrent neural network unfold.svg*, 2017. https://commons.wikimedia.org/wiki/File:Recurrent_neural_network_unfold.svg,[Online; accessed 22-August-2022].
- [5] Guillaume Chevalier. *By Guillaume Chevalier - File:The LSTM Cell.svg*. <https://commons.wikimedia.org/w/index.php?curid=109362147>,[Online; accessed 21-August-2022].
- [6] Saul Dobilas. *GRU Recurrent Neural Networks – A Smart Way to Predict Sequences in Python*. <https://towardsdatascience.com/gru-recurrent-neural-networks-a-smart-way-to-predict-sequences-in-python-80864e4fe9f6>,[Online; accessed 26-September-2022].
- [7] Giuliano Giacaglia. *How Transformers Work*. <https://towardsdatascience.com/transformers-141e32e69591>,[Online; accessed 27-September-2022].
- [8] Rani Horev. *BERT Explained: State of the art language model for NLP*. <https://towardsdatascience.com/bert-explained-state-of-the-art-language-model-for-nlp-f8b21a9b6270>,[Online; accessed 27-September-2022].
- [9] Jianfeng Si, Arjun Mukherjee, Bing Liu, Qing Li, Huayi Li και Xiaotie Deng. *Exploiting Topic based Twitter Sentiment for Stock Prediction. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, σελίδες 24–29, Sofia, Bulgaria, 2013. Association for Computational Linguistics. <https://aclanthology.org/P13-2005>.
- [10] Pratikkumar Prajapati. *Predictive analysis of Bitcoin price considering social sentiments*, 2020. <https://arxiv.org/abs/2001.10343>.

- [11] Mohammed Mudassir. *Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach*. <https://link.springer.com/article/10.1007/s00521-020-05129-6#citeas>, [Online; accessed 3-October-2022].
- [12] Critien Jacques Vella. *Bitcoin price change and trend prediction through twitter sentiment and data volume*. <https://jfin-swufe.springeropen.com/articles/10.1186/s40854-022-00352-7>, [Online; accessed 3-October-2022].
- [13] Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda και Babita Majhi. *Sentiment Analysis of Twitter Data for Predicting Stock Market Movements*, 2016. <https://arxiv.org/abs/1610.09225>.
- [14] Anshul Mittal και Arpit Goel. *Stock prediction using twitter sentiment analysis*. Stanford University, CS229 (2011 <http://cs229.stanford.edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>), 15:2352, 2012.
- [15] Fajri Koto και Mirna Adriani. *A comparative study on twitter sentiment analysis: Which features are good? International Conference on Applications of natural language to information systems*, σελίδες 453–457. Springer, 2015.
- [16] Efthymios Kouloumpis, Theresa Wilson και Johanna Moore. *Twitter sentiment analysis: The good the bad and the omg! Proceedings of the international AAAI conference on web and social media*, τόμος 5, σελίδες 538–541, 2011.
- [17] Wikipedia contributors. *Bitcoin – Wikipedia, The Free Encyclopedia*, 2022. <https://en.wikipedia.org/w/index.php?> [Online; accessed 26-August-2022].
- [18] Wayne Duggan. *The history of bitcoin, the first cryptocurrency*. <https://money.usnews.com/investing/articles/the-history-of-bitcoin>, [Online; accessed 5-October-2022].
- [19] Wikipedia contributors. *Litecoin – Wikipedia, The Free Encyclopedia*, 2022. <https://en.wikipedia.org/w/index.php?title=Litecoin&oldid=1108634804> [Online; accessed 5-October-2022].
- [20] Hooson Mark και Leggatt Johanna. *Bitcoin vs Ethereum*. <https://www.forbes.com/advisor/au/investing/cryptocurrency/bitcoin-vs-ethereum/>, [Online; accessed 5-October-2022].
- [21] Wayne Duggan . *Futures Definition*. <https://money.usnews.com/investing/term/futures> [Online; accessed 5-October-2022].
- [22] Eric Rosenberg. *How Does Bitcoin Mining Work?* <https://www.thebalancemoney.com/how-does-bitcoin-mining-work-5088328> [Online; accessed 5-October-2022].
- [23] George Caridakis, Kostas Karpouzis και Stefanos Kollias. *User and context adaptive neural networks for emotion recognition*. *Neurocomputing*, 71(13):2553–2562, 2008. Αρτιφισιαλ Νευραλ Νετωρκς (Γ'ANN 2006) / Ενγινεερινγ οφ Ιντελλιγεντ Σψοτεμς (Γ'ΕΙΣ 2006).

- [24] T. Balomenos, A. Raouzaiou, S. Ioannou, A. Drosopoulos, K. Karpouzis και S. Kollias. *Emotion Analysis in Man-Machine Interaction Systems. Machine Learning for Multimodal Interaction* Samy Bengio και Hervé Bourlard, επιμελητές, σελίδες 318–328, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.
- [25] Xiao Hu, Kahyun Choi και J. Stephen Downie. *A framework for evaluating multimodal music mood classification. Journal of the Association for Information Science and Technology*, 68(2):273–285, 2017. <https://asistdl.onlinelibrary.wiley.com/doi/abs/10.1002/asi.23649>.
- [26] Wikipedia contributors. *Natural language processing – Wikipedia, The Free Encyclopedia*, 2022. https://en.wikipedia.org/w/index.php?title=Natural_language_processing&oldid=1100496063, [Online accessed 2-August-2022].
- [27] Shashank, Gupta. *Sentiment Analysis: Concept, Analysis and Applications*. <https://towardsdatascience.com/sentiment-analysis-concept-analysis-and-applications-6c94d6f58c17> [Online; accessed 5-October-2022].
- [28] Geir Solskinnsbakk και others. *Contextual Semantic Search Navigation*. Διδακτορική Διατριβή, Norwegian University of Science and Technology, Trondheim, Norway, 2013.
- [29] C. Hutto και Eric Gilbert. *VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media*, 8(1):216–225, 2014. <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>.
- [30] Pollyanna Gonçalves, Matheus Araújo, Fabrício Benevenuto και Meeyoung Cha. *Comparing and Combining Sentiment Analysis Methods. Proceedings of the First ACM Conference on Online Social Networks, COSN '13*, σελίδα 27–38, New York, NY, USA, 2013. Association for Computing Machinery. url-<https://doi.org/10.1145/2512938.2512951>.
- [31] Ryan L. Boyd και H. Andrew Schwartz. *Natural Language Analysis and the Psychology of Verbal Behavior: The Past, Present, and Future States of the Field. Journal of Language and Social Psychology*, 40(1):21–41, 2021. ΠΜΙΔ: 34413563.
- [32] Finn Årup Nielsen. *A new ANEW: Evaluation of a word list for sentiment analysis in microblogs*. 2011. <https://arxiv.org/abs/1103.2903>.
- [33] Shihab Elbagir και Jing Yang. *Twitter sentiment analysis using natural language toolkit and VADER sentiment. Proceedings of the international multiconference of engineers and computer scientists*, τόμος 122, σελίδα 16, 2019.
- [34] Jacob Devlin, Ming Wei Chang, Kenton Lee και Kristina Toutanova. *Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805*, 2018. [Online; accessed 22-August-2022].

- [35] Shujaur Rehman Baig, Waheed Iqbal, Josep Lluís Berral και David Carrera. *Adaptive sliding windows for improved estimation of data center resource utilization*. *Future Generation Computer Systems*, 104:212–224, 2020.
- [36] Tom M. Mitchell. *Machine Learning*. McGraw-Hill, New York, 1997. <https://www.bibsonomy.org/bibtex/23e79734ee1a6e49aee02ffd108224d1c/flint63>, [Online; accessed 22-August-2022].
- [37] SAGAR SHARMA. *Activation Functions in Neural Networks*. <https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d67>, [Online; accessed 4-October-2022].
- [38] Wikipedia contributors. *Perceptron – Wikipedia, The Free Encyclopedia*, 2022. <https://en.wikipedia.org/w/index.php?title=Perceptron&oldid=1102767784>, [Online; accessed 9-August-2022].
- [39] Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola. *Dive into Deep Learning*, 2022. <https://d2l.ai/index.html>, [Online; accessed 10-August-2022].
- [40] Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola. *Dive into Deep Learning*, 2022. https://d2l.ai/chapter_multilayer-perceptrons/backprop.html, [Online; accessed 12-August-2022].
- [41] Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola. *Dive into Deep Learning*, 2022. https://d2l.ai/chapter_optimization/gd.html, [Online; accessed 12-August-2022].
- [42] Afshine Amidi και Shervine Amidi. *Recurrent Neural Networks cheatsheet*. <https://stanford.edu/~shervine/teaching/cs-230/cheatsheet-recurrent-neural-networks>, [Online; accessed 4-October-2022].
- [43] Michael Phi. *Illustrated Guide to LSTM's and GRU's: A step by step explanation*. <https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21>, [Online; accessed 4-October-2022].
- [44] Aston Zhang, Zack C. Lipton, Mu Li, Alex J. Smola. *Dive into Deep Learning*, 2022. https://d2l.ai/chapter_recurrent-modern/lstm.html, [Online; accessed 26-September-2022].
- [45] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias και Georgios Tagaris. *Deep neural architectures for prediction in healthcare*. *Complex & Intelligent Systems*, 4(2):119–131, 2018.
- [46] Ilianna Kollia, Andreas Georgios Stafylopatis και Stefanos Kollias. *Predicting Parkinson's disease using latent information extracted from deep neural networks*. *2019 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1–8. IEEE, 2019.
- [47] Anastasios Arsenos, Dimitrios Kollias και Stefanos Kollias. *A Large Imaging Database and Novel Deep Neural Architecture for Covid-19 Diagnosis*. *2022 IEEE 14th Image*,

- Video, and Multidimensional Signal Processing Workshop (IVMSP)*, σελίδες 1–5. IEEE, 2022.
- [48] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis και Stefanos Kollias. *Adaptation and contextualization of deep neural network models*. 2017 IEEE symposium series on computational intelligence (SSCI), σελίδες 1–8. IEEE.
- [49] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate και S Kollias. *Transparent adaptation in deep medical image diagnosis*. *International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*, σελίδες 251–267. Springer, 2020.
- [50] D Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, I Kollia, L Sukissian, J Wingate και S Kollias. *Deep Transparent Prediction through Latent Representation Analysis*. *arXiv preprint arXiv:2009.07044*, 2020.
- [51] Fabio De Sousa Ribeiro, Francesco Calivá, Mark Swainson, Kjørtan Gudmundsson, Georgios Leontidis και Stefanos Kollias. *Deep bayesian self-training*. *Neural Computing and Applications*, 32(9):4275–4291, 2020.
- [52] Fabio De Sousa Ribeiro, Georgios Leontidis και Stefanos Kollias. *Capsule routing via variational bayes*. *Proceedings of the AAAI Conference on Artificial Intelligence*, τόμος 34, σελίδες 3749–3756, 2020.
- [53] Fabio De Sousa Ribeiro, Georgios Leontidis και Stefanos Kollias. *Introducing routing uncertainty in capsule networks*. *Advances in Neural Information Processing Systems*, 33:6490–6502, 2020.
- [54] Andreas Psaroudakis και Dimitrios Kollias. *MixAugment & Mixup: Augmentation Methods for Facial Expression Recognition*. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, σελίδες 2367–2375, 2022.
- [55] Dimitrios Kollias και Stefanos Zafeiriou. *Training deep neural networks with different datasets in-the-wild: The emotion recognition paradigm*. 2018 International Joint Conference on Neural Networks (IJCNN), σελίδες 1–8. IEEE, 2018.
- [56] Dimitrios Kollias και Stefanos Zafeiriou. *Va-stargan: Continuous affect generation*. *International Conference on Advanced Concepts for Intelligent Vision Systems*, σελίδες 227–238. Springer, 2020.
- [57] G Caridakis, A Raouzaiou, K Karpouzis και S Kollias. *Synthesizing Gesture Expressivity Based on Real Sequences*. *Workshop Programme*, τόμος 10, σελίδα 19.
- [58] Bashar Alhnaity, Stefanos Kollias, Georgios Leontidis, Shouyong Jiang, Bert Schamp και Simon Pearson. *An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth*. *Information Sciences*, 560:35–50, 2021.

- [59] Phivos Mylonas, Evaggelos Spyrou, Yannis Avrithis και Stefanos Kollias. *Using visual context and region semantics for high-level concept detection*. *IEEE Transactions on Multimedia*, 11(2):229–243, 2009.
- [60] Stefanos Kollias και Dimitris Anastassiou. *A unified neural network approach to digital image halftoning*. *IEEE Transactions on signal processing*, 39(4):980–984, 1991.
- [61] Paraskevi Tzouveli, Andreas Schmidt, Michael Schneider, Antonis Symvonis και Stefanos Kollias. *Adaptive reading assistance for the inclusion of students with dyslexia: The AGENT-DYSL approach*. *2008 Eighth IEEE International Conference on Advanced Learning Technologies*, σελίδες 167–171. IEEE, 2008.
- [62] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser και Illia Polosukhin. *Attention Is All You Need*, 2017. <https://arxiv.org/abs/1706.03762>.
- [63] Tianyang Lin, Yuxin Wang, Xiangyang Liu και Xipeng Qiu. *A Survey of Transformers*, 2021. <https://arxiv.org/abs/2106.04554>.
- [64] Jason Brownlee. *What are word embeddings*. <https://machinelearningmastery.com/what-are-word-embeddings>, [Online; accessed 27-September-2022].
- [65] Yoav Goldberg και Omer Levy. *word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method*, 2014. <https://arxiv.org/abs/1402.3722>.
- [66] Jeffrey Pennington, Richard Socher και Christopher Manning. *GloVe: Global Vectors for Word Representation*. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, σελίδες 1532–1543, Doha, Qatar, 2014. Association for Computational Linguistics. "<https://aclanthology.org/D14-1162>".
- [67] Samia Khalid. *BERT Explained: A Complete Guide with Theory and Tutorial*. <https://medium.com/@samia.khalid/bert-explained-a-complete-guide-with-theory-and-tutorial-3ac9ebc8fa7c>, [Online; accessed 28-September-2022].
- [68] Dat Quoc Nguyen, Thanh Vu και Anh Tuan Nguyen. *BERTweet: A pre-trained language model for English Tweets*, 2020. <https://arxiv.org/abs/2005.10200>.
- [69] Jamil M. Luqman, Pais Sebastião, Cordeiro João και Dias Gaël. *Detection of extreme sentiments on social networks with BERT*. <https://link.springer.com/article/10.1007/s13278-022-00882-z>, [Online; accessed 4-October-2022].
- [70] Israa Alghanmi, Luis Espinosa Anke και Steven Schockaert. *Combining BERT with Static Word Embeddings for Categorizing Social Media*. *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, σελίδες 28–33, Online, 2020. Association for Computational Linguistics.
- [71] Sidharth R, Abhiraj Tiwari, Parthivi Choubey, Saisha Kashyap, Sahil Khose, Kumud Lakara, Nishesh Singh και Ujjwal Verma. *BERT based Transformers lead the way*

- in Extraction of Health Information from Social Media*, 2021. <https://arxiv.org/abs/2104.07367>.
- [72] Marzieh Mozafari, Reza Farahbakhsh και Noel Crespi. *A BERT-Based Transfer Learning Approach for Hate Speech Detection in Online Social Media*, 2019. <https://arxiv.org/abs/1910.12574>.
- [73] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer και Veselin Stoyanov. *RoBERTa: A Robustly Optimized BERT Pretraining Approach*, 2019. <https://arxiv.org/abs/1907.11692> [Online; accessed 28-September-2022].
- [74] Elvys Linhares Pontes και Mohamed Benjannet. *Contextual Sentence Analysis for the Sentiment Prediction on Financial Data*, 2021. <https://arxiv.org/abs/2112.13790>.
- [75] Rani Molla. *When Elon Musk tweets, crypto prices move*. <https://www.vox.com/recode/2021/5/18/22441831/elon-musk-bitcoin-dogecoin-crypto-prices-tesla>, [Online; accessed 29-September-2022].
- [76] Pekka Malo, Ankur Sinha, Pyry Takala, Pekka Korhonen και Jyrki Wallenius. *Good Debt or Bad Debt: Detecting Semantic Orientations in Economic Texts*, 2013. <https://arxiv.org/abs/1307.5336>.
- [77] yahoo team. *Bitcoin USD (BTC-USD)*. <https://finance.yahoo.com/quote/BTC-USD?p=BTC-USD>, [Online; accessed 4-October-2022].