ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΠΡΟΗΓΜΕΝΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΉΜΑΤΑ ΚΑΙ ΥΠΗΡΕΣΙΕΣ

Γραμμική παλινδρόμηση για την πρόβλεψη της ανόδου ή πτώσης του χρηματιστηρίου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΔΗΜΑΝΟΠΟΥΛΟΥ ΜΙΧΑΗΛ(ΜΕ1830)

Επιβλέπων : Γεώργιος Βασιλακόπουλος
Καθηγητής ΠΑ.ΠΕΙ.

Αθήνα, 2020

(Η σελίδα σκοπίμως αφέθηκε κενή)

Διπλωματική Εργασία

Γραμμική παλινδρόμηση για την πρόβλεψη της ανόδου ή πτώσης του χρηματιστηρίου

Δημανόπουλος Μιχαήλ

(Η σελίδα σκοπίμως αφέθηκε κενή)

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΑ
ΣΧΟΛΗ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ
ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΠΡΟΗΓΜΕΝΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΥΠΗΡΕΣΙΕΣ

Γραμμική παλινδρόμηση για την πρόβλεψη της ανόδου ή πτώσης του χρηματιστηρίου

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΔΗΜΑΝΟΠΟΥΛΟΥ ΜΙΧΑΗΛ

Επιβλέπων : Γεώργιος Βασιλακόπουλος
Καθηγητής ΠΑ.ΠΕΙ.

Εγκρίθηκε από επιτροπή Φεβρουαρίος 2020.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Πρέντζα Ανδριάνα Καθηγητής ΠΑ.ΠΕΙ.	Κυριαζής Δημοσθένης Αναπλ. Καθηγητής ΠΑ.ΠΕΙ.	Όνομα Επώνυμο Καθηγητής ΠΑ.ΠΕΙ.

Αθήνα, 2020

(Υπογραφή)

.....

ΔΗΜΑΝΟΠΟΥΛΟΣ ΜΙΧΑΗΛ

Μεταπτυχιακός φοιτητής του τμήματος Ψηφιακών Συστημάτων ΠΑ.ΠΕΙ.

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέποντα καθηγητή μου κ. Γεώργιο Βασιλακόπουλο, για τη καθοδήγηση που μου παρείχε σε όλα τα στάδια εκπόνησης της διπλωματικής μου εργασίας.

Επίσης θα ήθελα να ευχαριστήσω την οικογένεια μου για την στήριξη που μου παρείχε σε όλα τα στάδια του Μεταπτυχιακού Προγράμματος.

Ολοκληρώνοντας, θα ήθελα να ευχαριστήσω το Ίδρυμα Κρατικών Υποτροφιών (ΙΚΥ) για την υποτροφία που μου παρείχε.

Περίληψη

Σκοπός της παρούσας διπλωματικής εργασίας ήταν η ανάπτυξη και αξιολόγηση μοντέλου γραμμικής παλινδρόμησης για την πρόβλεψη της ανόδου ή πτώσης του παγκόσμιου χρηματιστηριακού δείκτη MSCI. Συγκεκριμένα, θεωρήσαμε ότι υπάρχει γραμμική συσχέτιση μεταξύ του παγκόσμιου χρηματιστηριακού δείκτη με τους πιο σημαντικούς χρηματιστηριακούς δείκτες ανά τον κόσμο και την πολικότητα των ειδήσεων (δηλαδή αν οι ειδήσεις ήταν θετικές ή αρνητικές). Οι δείκτες που μελετήθηκαν περιλαμβάνουν το Χρηματιστήριο Φρανκφούρτης (DAX), το Χρηματιστήριο Παρισιού (CAC 40), το Daw Jones, το Χρηματιστήριο Λονδίνου (FTSE 100), το NASDAQ και το S&P 500.

Τα ευρήματα των πειραμάτων που πραγματοποιήθηκαν στα πλαίσια της εργασίας είναι ενδιαφέροντα τόσο για τους διάφορους χρηματιστηριακούς δείκτες που μελετήθηκαν όπως επίσης και για την συνεισφορά της επιπρόσθετης πληροφορίας από τα κείμενα που συλλέχθηκαν.

Τέλος, την εργασία την συνοδεύει μία γραφική διεπαφή χρήστη, η οποία δίνει την δυνατότητα σε ενδιαφερόμενους να εκπαιδεύσουν τα δικά τους μοντέλα γραμμικής παλινδρόμησης, φορτώνοντας σύνολα εκπαίδευσης. Επίσης, οι ενδιαφερόμενοι μπορούν να προβλέψουν την άνοδο ή την πτώση του χρηματιστηρίου κάποια χρονική στιγμή δίνοντας πραγματικές τιμές στις ανεξάρτητες μεταβλητές.

Λέξεις Κλειδιά: μηχανική μάθηση, γραμμική παλινδρόμηση, πρόβλεψη χρηματιστηρίου

Abstract

The scope of this thesis was the development and the evaluation of linear regression for the prediction of the international stock market MSCI rise/decline. Particularly, we assumed that there is a linear correlation between the MSCI index and the most important indices around the world along with the polarity of the news (i.e. if the news are positives or negatives). The indices we used are, the stock market of Frankfort (DAX), the stock market of Paris (CAC 40), the Dow Jones, the stock market of London (FTSE 100), the NASDAQ and the S&P 500.

The findings of the experiments, conducted in the context of the thesis, are interesting both for the various stock market indices studied as well as for the contribution of additional information from the collected texts.

Finally, the thesis is accompanied by a graphical user interface, which enables people who are interested to this topic to train their own linear regression models, loading training sets. People can also predict the stock market rise or decline by giving real time prices to the independent variables.

Keywords: machine learning, Linear Regression, Stock Market Prediction

Πίνακας περιεχομένων

Ευρετήριο Εικόνων

(Η σελίδα σκοπίμως αφέθηκε κενή)

Ευρετήριο Πινάκων

(Η σελίδα σκοπίμως αφέθηκε κενή)

1

Εισαγωγή

1.1 Πρόβλεψη Ανόδου ή Πτώσης Διεθνών Χρηματιστηρίων

Το χρηματιστήριο νοείται μία οργανωμένη αγορά, κατά την οποία τα ενδιαφερόμενα μέρη συμμετέχουν στη λειτουργία του, η οποία αποτελεί τον τόπο συνάντησης τους, ώστε να διεκπεραιωθεί η διενέργεια της αγοραπωλησίας κινητών αξιών ή και εμπορευμάτων. Η αγορά αυτή του χρηματιστηρίου κατά των πλείστων οργανώνεται από το ίδιο το κράτος ενώ καθορίζονται και νομοθετικά μέτρα για την ομαλή λειτουργία του κατά τη διάρκεια κλεισίματος και ανοίγματος του, προστατεύοντας με αυτόν τον τρόπο τα εμπλεκόμενα μέρη.

Η δομή του χρηματιστηρίου αποτελείται από τις λεγόμενες χρηματιστηριακές μετοχές τις οποίες ένας οργανισμός, όπου είναι εισηγμένος στο χρηματιστήριο του κράτους, εκδίδει με στόχο το κέρδος. Ο αριθμός των μετοχών που εκδίδεται από μία επιχείρηση είναι συγκεκριμένος κι ορίζεται ανάλογα με το καταστατικός της. Οι μετοχές είναι αυτές όπως παίζουν τον κύριο ρόλο στην διαμόρφωση των τιμών ανοίγματος και κλεισίματος του χρηματιστηρίου.

Ο γενικός δείκτης ενός χρηματιστηρίου, μεταβάλλεται ανάλογα με τις αυξομειώσεις των μετοχών κατά τη διάρκεια της ημέρας. Ο μηχανισμός που εξάγει τη τελική τιμή του γενικού δείκτη, διαμορφώνεται από τη σημαντικότητα που χαρακτηρίζουν τις συγκεκριμένες μετοχές, όπου μία αύξηση ή μία μείωση τους θα κρίνει την τελική τιμή του δείκτη κατά το κλείσιμο του χρηματιστηρίου. Αυτό συνεπάγεται ότι μεταξύ των σημαντικών δεικτών

υπάρχει σχέση αλληλεξάρτησης η οποία καθορίζει κατά πόσο ανοδική ή πτωτική πορεία μπορεί να έχει το χρηματιστήριο σε μία χώρα.

Με την δραματική αύξηση των τιμών των μετοχών από της μεγάλες χρηματιστηριακές αγορές όπως των ΗΠΑ, στη δεκαετία του 90' και η ακόλουθη συντριβή του χρηματιστηρίου που ξεκίνησε το 2000, αποτέλεσε μία τρανή απόδειξη της ισχυρής συσχέτισης και αλληλεξάρτησης μεταξύ των χρηματιστηριακών αγορών παγκοσμίως [4].

Ένας από τους μεγαλύτερους συντελεστές στην εξέλιξη της πορείας των χρηματιστηριακών αγορών, για πολλά χρόνια, ήταν η αγορά των ΗΠΑ. Το γεγονός αυτό μπορούσε ο καθένας να το διακρίνει από τα διάφορα περιστατικά που συνέβησαν στο παρελθόν και αφορούσαν σημαντικές πτώσεις των χρηματιστηριακών δεικτών των ΗΠΑ. Ένα από τα τελευταία και πιο καταστροφικά για την παγκόσμια αγορά ήταν η πτώση του χρηματιστηρίου (όπως έχει μείνει στην ιστορία το χρηματιστηριακό κραχ του 1929), όπου είχαν κλονιστεί όλες οι αγορές του κόσμου εξαιτίας της ραγδαίας εξάπλωσής του. Το κραχ σηματοδότησε την αρχή της παγκόσμιας ύφεσης που επηρέασε πολλά βιομηχανικά δυτικά κράτη [21].

Καθώς οι χρηματιστηριακές αγορές της Ευρωπαϊκής Ένωσης (ΕΕ) αναπτύχθηκαν, πλέον μπορούσαν να θωρηθούν ως σημαντικές στον παγκόσμιο οικονομικό χάρτη. Αυτό σήμαινε ότι οι απότομες αυξομειώσεις του χρηματιστηριακού δείκτη, ενός μεγάλου βιομηχανικού κράτους της ΕΕ (όπως είναι η Γερμανία), θα επηρέαζε σημαντικά την παγκόσμια κοινότητα του χρηματιστηρίου φτάνοντας ακόμα και στις αγορές της Κίνας ή των ΗΠΑ.

Η ενσωμάτωση των ευρωπαϊκών χρηματοπιστωτικών αγορών[1], σύμφωνα με ερευνητές, μπορεί πιθανόν να οδηγήσει σε ακόμη πιο ισχυρή συσχέτιση μεταξύ των διάφορων τιμών των χρηματιστηριακών δεικτών στις ευρωπαϊκές χώρες. Το φαινόμενο αυτό έχει ως αποτέλεσμα την σύγκλιση των οικονομικών δραστηριοτήτων σε όλες τις χώρες της Ευρωπαϊκής Ένωσης. Οι εξελίξεις των χρηματιστηριακών αγορών επηρεάζονται από πραγματικές μεταβλητές όπως είναι οι επενδύσεις που πραγματοποιούνται στην εκάστοτε χώρα αλλά και ο δείκτης της κατανάλωσης από την πλευρά των πολιτών της ΕΕ [15].

[1] Χρηματοπιστωτική αγορά είναι το σύνολο των αγορών όπου οι επενδυτές αγοράζουν και πουλάνε φυσικά και χρηματιστικά περιουσιακά στοιχεία ή χρηματοπιστωτικές απαιτήσεις.

Η παραπάνω λογική λειτουργίας του χρηματιστηρίου της ΕΕ, μας δείχνει ότι πλέον οι οικονομίες που στηρίζονται στο χρηματιστήριο μπορούν να επηρεάσουν τους χρηματιστηριακούς δείκτες κι άλλων χωρών εκτός ΕΕ. Επίσης στο πλαίσιο πιθανής επίδρασης της χρηματιστηριακής αγοράς στη μακροοικονομική πολιτική που χρησιμοποιείται από την ΕΕ, οι τιμές των μετοχών μπορεί να αποτελέσουν έναν από τους σημαντικότερους και καθοριστικούς παράγοντες της νομισματικής πολιτικής που ακολουθούν τα κράτη σε παγκόσμιο επίπεδο [15].

Αν αναλογιστούμε την επίδραση της παγκοσμιοποίησης στις χρηματαγορές της κάθε χώρας, διαπιστώνουμε ότι πλέον μία δραματική αύξηση ή αντίστοιχα μείωση των τιμών ισχυρών χρηματιστηριακών δεικτών όπως των ΗΠΑ ή της Αγγλίας, μπορούν να επηρεάσουν έντονα και τους υπόλοιπους χρηματιστηριακούς δείκτες άλλων χωρών. Έτσι τέθηκε το ερώτημα από την κοινότητα των επενδυτών αν θα μπορούσε να υπάρξει ποτέ μηχανισμός ο οποίος να μπορεί να προβλέψει την τιμή του ανοίγματος και του κλεισίματος του χρηματιστηριακού δείκτη, ώστε να μπορέσουν έστω και κατά προσέγγιση να διαμορφώσουν μία συνολική εικόνα διακύμανσης των χρηματιστηριακών τιμών, για τα μετέπειτα επενδυτικά τους σχέδια.

Μόλις τέθηκε ο παραπάνω προβληματισμός σχετικά με την πρόβλεψη ανόδου ή πτώσης των χρηματιστηριακών δεικτών, ερευνητές έσπευσαν να δώσουν λύσεις οι οποίες μπορούσαν να εφαρμοστούν για την πρόβλεψη βραχυχρόνιων τιμών χρηματιστηριακών δεικτών. Οι λύσεις που είχαν δώσει για την μελλοντική πρόβλεψη των τιμών δε μπορούσε να θεωρηθεί αξιόπιστη διότι τα μαθηματικά μοντέλα σε συνδυασμό με την τεχνολογία εκείνης της εποχής, καθιστούσαν αδύνατη την ακριβής προσέγγιση των μακροχρόνιων χρηματιστηριακών δεικτών όπου ήταν και το ζήτημα του προβληματισμού που τέθηκε από τους επενδυτές.

Κατά την εξέλιξη της τεχνολογίας και την εισαγωγή των ηλεκτρονικών υπολογιστών στις εργασίες των ερευνητών, μπορούσαν πλέον να αξιοποιηθούν πολλά μαθηματικά μοντέλα τα οποία ήταν ικανά να προβλέψουν τις τιμές των χρηματιστηριακών τιμών σε μακροχρόνιο επίπεδο. Ο συνδυασμός αυτών των μοντέλων με την τεχνητή νοημοσύνη είχε σαν στόχο την ακριβή προσέγγιση των χρηματιστηριακών δεικτών, λαμβάνοντας δεδομένα τιμών των δεικτών από το ιστορικό του κάθε χρηματιστηρίου.

Οι αυτοματοποιημένες λύσεις που μας δίνουν τα προγράμματα ηλεκτρονικών υπολογιστών, χρησιμοποιούν τεχνολογίες εξόρυξης δεδομένων και τις συνδυάζουν με τεχνολογίες πρόβλεψης, αποδίδοντας έτσι ένα σημαντικό μελλοντικό στίγμα των συναλλαγών στις αγορές. Η εξόρυξη δεδομένων βασίζεται στη θεωρεία ότι τα ιστορικά δεδομένα των χρηματιστηριακών δεικτών κατέχουν την απαραίτητη γνώση για την πρόβλεψη μελλοντικών κατευθύνσεων της τιμής του δείκτη. Αυτού του είδους οι τεχνολογίες σχεδιαστήκαν για να βοηθήσουν τους επενδυτές να προβλέψουν την άνοδο ή την πτώση των χρηματιστηριακών δεικτών κι άλλων χρηματοοικονομικών φαινομένων. Η ανάλυση των δεδομένων είναι ένας τρόπος πρόβλεψης των μελλοντικών τιμών των μετοχών και σχετίζεται με την άνοδο ή την πτώση τους στο χρηματιστήριο [6].

Στη παρούσα εργασία πραγματοποιείται η χρήση μίας μεθοδολογίας της μηχανικής μάθησης η οποία είναι παρακλάδι της τεχνητής νοημοσύνης. Η μηχανική μάθηση έχει σαν στόχο να ορίσει το πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί. Στόχος της είναι η χρήση αλγορίθμων οι οποίοι έχουν τη δυνατότητα να μπορούν να μάθουν από τα δεδομένα που εισάγει ο χρήστης, κάνοντας με αυτόν τον τρόπο σενάρια πρόβλεψης σχετικά με τον τύπο των δεδομένων που τους είχε δοθεί [10].

Οι αλγόριθμοι μηχανικής μάθησης λειτουργούν κατασκευάζοντας μοντέλα, εισάγοντας σε αυτούς έναν τεράστιο όγκο δεδομένων, τα οποία θεωρούνται πειραματικά δεδομένα, προκειμένου να κάνουν όσο το δυνατόν πιο ακριβείς προβλέψεις βασιζόμενες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως αποτέλεσμα. Το πιο σημαντικό στοιχείο για τη σωστή λειτουργία τέτοιου είδους αλγορίθμων είναι η εισαγωγή των δεδομένων του προβλήματος που μελετούν [9], όπου στην προκειμένη περίπτωση είναι δεδομένα τιμών από το ιστορικό των χρηματιστηριακών δεικτών. Έτσι οι προβλέψεις που διαμορφώνουν αυτοί οι αλγόριθμοι αφορούν τα χρηματιστηριακά δεδομένα και πως αυτά μεταβάλλονται κατά την πάροδο του χρόνου ανάλογα με το χρονικό διάστημα που επιλέξαμε να εξετάσουμε.

Κατά τα επόμενα χρόνια κι ενώ οι ερευνητές παρατήρησαν ότι το φαινόμενο της παγκοσμιοποίησης των αγορών είναι ένας από τους σημαντικότερους παράγοντες που επηρεάζουν τους χρηματιστηριακούς δείκτες κάθε χώρας, πρότειναν την διαμόρφωση ενός

παγκόσμιου δείκτη ο οποίος θα μπορούσε να εξάγει αποτελέσματα τα οποία θα ήταν χρήσιμα στη μελέτη των πιο ισχυρών οικονομιών σε διεθνές επίπεδο. Ο παγκόσμιος αυτός χρηματιστηριακός δείκτης ονομάζεται MSCI World και περιλαμβάνει στη λίστα του περίπου 1.655 μετοχές εταιρειών οι οποίες παίζουν σημαντικό ρόλο στη διαμόρφωση των χρηματιστηριακών δεικτών.

Η χρήση ενός παγκόσμιου χρηματιστηριακού δείκτη είχε σαν στόχο τη μελέτη οικονομικών μεγεθών που σχετίζονται με το χρηματιστήριο και πως οι τιμές των μετοχών μεγάλων εταιρειών ή χρηματοπιστωτικών ιδρυμάτων ανά χώρα μπορούν να επηρεάσουν πολλούς δείκτες ανεξαρτήτως έδρας χρηματιστηρίου. Αυτή η εκμετάλλευση του δείκτη μας δείχνει κατά πόσο η παγκόσμια οικονομία έχει ανοδική ή πτωτική τάση κατά τη χρονική περίοδο που μελετάμε [2].

Ο χρηματιστηριακός δείκτης MSCI World περιλαμβάνει μία λίστα χωρών όπου το επίπεδο των χρηματαγορών έχει μεγάλο αντίκτυπο στην παγκόσμια οικονομία. Κάποιες από τις χώρες που εξετάζει αυτός ο δείκτης είναι οι ΗΠΑ, η Αυστραλία, η Αυστρία, η Αγγλία, η Γερμανία, ο Καναδάς, το Ισραήλ, η Ισπανία, η Σουηδία, η Ελβετία, η Γαλλία, η Ιαπωνία, η Ιταλία κ.ά. [5].

Με την πάροδο των χρόνων κρίθηκε αναγκαία η δυνατότητα πρόβλεψης του παγκόσμιου χρηματιστηριακού δείκτη ώστε οι επενδυτές να σχεδιάζουν τις οικονομικές τους βλέψεις κατά τα επόμενα χρόνια. Μια τέτοια πρόβλεψη δε θα βοηθούσε μόνο τους επενδυτές αλλά και τις διάφορες χώρες οι οποίες παίζουν σημαντικό ρόλο στον παγκόσμιο οικονομικό χάρτη. Έτσι οι χώρες αυτές θα μπορούσαν να προσαρμόσουν την νομισματική τους πολιτική αλλά και τις επενδυτικές τους κινήσεις για την αποφυγή ακραίων οικονομικών φαινομένων.

Στόχος της εργασίας είναι η δυνατότητα πρόβλεψης του παγκόσμιου χρηματιστηριακού δείκτη σε μακροχρόνιο επίπεδο με τεχνικές μηχανικής μάθησης. Η χρήση του αλγορίθμου για τη πρόβλεψη των μελλοντικών τιμών των υπό εξέταση δεικτών σε συνδυασμό με την εισαγωγή ιστορικών δεδομένων αυτών, θα μας βοηθούσε να εξάγουμε σημαντικά αποτελέσματα σχετικά με τους χρηματιστηριακούς δείκτες οι οποίοι παίζουν κρίσιμο ρόλο στη διαμόρφωση του παγκόσμιου δείκτη. Κατ’ αυτόν τον τρόπο θα έχουμε μια συνολική

εικόνα του δείκτη σε διεθνές επίπεδο βγάζοντας κρίσιμα συμπεράσματα σχετικά με την επίδραση των ισχυρών χρηματιστηριακών δεικτών στον παγκόσμιο δείκτη.

1.2 Αντικείμενο διπλωματικής

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η πρόβλεψη του παγκόσμιου χρηματιστηριακού δείκτη MSCI. Συγκεκριμένα, δοθέντος ενός συνόλου από τους πιο σημαντικούς χρηματιστηριακούς δείκτες και την πολικότητα των ειδήσεων (δηλαδή αν οι ειδήσεις είναι θετικές ή αρνητικές) για μια συγκεκριμένη χρονική στιγμή, στόχος μας είναι να προβλέψουμε την άνοδο ή την πτώση του παγκόσμιου χρηματιστηριακού δείκτη. Για την επίτευξη του στόχου αυτού, υποθέτουμε ότι υπάρχει γραμμική συσχέτιση μεταξύ των χρηματιστηριακών δεικτών και της πολικότητας των ειδήσεων με τον παγκόσμιο χρηματιστηριακό δείκτη.
Έτσι, πειραματιστήκαμε με την εκπαίδευση ενός μοντέλου γραμμικής παλινδρόμησης το οποίο θα έχει την δυνατότητα να μαθαίνει από ιστορικά δεδομένα των χρηματιστηριακών δεικτών και των ειδήσεων την άνοδο ή την πτώση του παγκόσμιου δείκτη.

Επίσης, ένας δεύτερο στόχος της συγκεκριμένης εργασίας, είναι οπτικοποίηση των αποτελεσμάτων και η δημιουργία μίας φιλικής διεπαφής για τον χρήστη. Με αυτόν τον τρόπο, ενδιαφερόμενοι μπορούν να χρησιμοποιήσουν την εφαρμογή και να φορτώσουν δικά τους σύνολα δεδομένων που απαρτίζονται από ιστορικά δεδομένα. Με αυτά τα δεδομένα θα έχουν την δυνατότητα να εκπαιδεύσουν το δικό τους μοντέλο γραμμικής παλινδρόμησης και να εξάγουν συμπεράσματα για την κίνηση του παγκόσμιου δείκτη για διάφορες χρονικές στιγμές.

1.2.1 Συνεισφορά

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Μελετήσαμε συστήματα πρόβλεψης της τιμής του χρηματιστηρίου
2. Υλοποιήσαμε ένα μοντέλο γραμμικής παλινδρόμησης για την επίλυση του προβλήματος πρόβλεψης του χρηματιστηριακού δείκτη MSCI.
3. Επεκτείναμε το σύνολο χαρακτηριστικών με κείμενα παρμένα από καθημερινά νέα.

4. Δημιουργήσαμε γραφική διεπαφή χρήστη για την εύκολη εκπαίδευση και αξιολόγηση των μοντέλων.

1.3 Οργάνωση κειμένου

Το κείμενο της διπλωματικής εργασίας απαρτίζεται από τα παρακάτω κεφάλαια.

Στο κεφάλαιο 2 παρουσιάζουμε την σχετική βιβλιογραφία. Αναλυτικά, καταγράφουμε μεθοδολογίες που έχουν προταθεί από διάφορους ερευνητές για την επίλυση του προβλήματος πρόβλεψης του χρηματιστηρίου. Στην συνέχεια (κεφάλαιο 3) αναφέρουμε το θεωρητικό υπόβαθρο της εργασίας. Σε αυτό το σημείο, ο αναγνώστης αναμένει να δει την ορολογία και τους ορισμούς που χρησιμοποιήθηκαν για την επίλυση του προβλήματος. Επίσης, αναφέρουμε και τις μεθοδολογίες που έχουν αναπτύξει άλλοι ερευνητές ώστε να υπάρχει μια έμμεση σύγκριση της μεθοδολογίας που επιλέξαμε σε σύγκριση με τις υπόλοιπες. Στο κεφάλαιο 4, ορίζουμε το πρόβλημα που έχουμε κληθεί να αντιμετωπίσουμε ενώ στο κεφάλαιο 5 παρουσιάζουμε την επίλυση του προβλήματος με την μεθοδολογία που επιλέξαμε. Στο κεφάλαιο 6, παρουσιάζονται οι παράμετροι που εξετάστηκαν καθώς και τα αποτελέσματα που παράχθηκαν από την μεθοδολογία που ακολουθήθηκε. Εν συνεχεία (κεφάλαιο 7) καταγράφουμε όλες τις τεχνικές λεπτομέρειες που σχετίζονται με την ανάπτυξη και αξιολόγηση της μεθοδολογίας μας όπως και της ανάπτυξης της γραφικής διεπαφής. Στο σημείο αυτό, ο αναγνώστης θα πληροφορηθεί σχετικά με τις πλατφόρμες και τα εργαλεία που χρησιμοποιήθηκαν ενώ θα είναι σε θέση να εκτελέσει την εγκατάσταση των προγραμμάτων που υλοποιήθηκαν. Τέλος (κεφάλαιο 8) συνοψίζουμε τα συμπεράσματα της εργασίας και κάποιες μελλοντικές επεκτάσεις που θα μπορούσαν να γίνουν.

2

Σχετικές εργασίες

Όπως αναφέρθηκε, αντικείμενο της παρούσας εργασίας είναι η πρόβλεψη ανόδου ή πτώσης του παγκόσμιου χρηματιστηριακού δείκτη με την χρήση γραμμικής παλινδρόμησης. Επομένως, είναι άρρηκτη ανάγκη να παρουσιάσουμε εκείνες τις εργασίες που προσπάθησαν να λύσουν αυτό το πρόβλημα με διάφορες τεχνικές είτε με παραδοσιακές τεχνικές μηχανικές μάθησης ή με νευρωνικά δίκτυα (για εκμετάλλευση της μη γραμμικότητας που τα δίκτυα αυτά προσφέρουν).

2.1 Πρόβλεψη τιμών χρηματιστηρίου

Την ανάγκη για σωστές επενδύσεις σε μετοχές ώστε να μην υπάρχει απώλεια χρημάτων αντιλήφθηκαν οι [19] για την επενδυτική δραστηριότητα στην Ταιβάν. Αναλυτικά, πρότειναν ένα υβριδικό μοντέλο πρόβλεψης των τιμών μετοχών βιομηχανιών ηλεκτρικών. Το μοντέλο αυτό αποτελείται από ένα νευρωνικό δίκτυο συγκεκριμένα ένα πολυστρωματικό εμπροσθοτροφοδοτούμενο νευρωνικό δίκτυο (Multilayer Feedforward Networks) το οποίο εκπαιδεύεται με τον αλγόριθμο πίσω διάδοσης τους λάθους και ένα δέντρο απόφασης. Τα αποτελέσματα έδειξαν ότι η από κοινού χρήση αυτών των μοντέλων έδωσε 77% ακρίβεια, όπου είναι υψηλότερη έναντι του αν χρησιμοποιούνταν τα μοντέλα ξεχωριστά για την πρόβλεψη των τιμών. Η διαφορά αυτής της εργασίας με την προτεινόμενη από εμάς είναι το γεγονός ότι η δεύτερη χρησιμοποιεί μη γραμμικά μοντέλα για την εκπαίδευση του μοντέλου πρόβλεψης. Επίσης, η δική μας μελέτη εστιάζει σε επίπεδο χρηματιστηρίου και όχι σε επίπεδο μετοχών συγκεκριμένης περιοχής (π.χ. βιομηχανίας ηλεκτρικών). Τέλος, είναι

σημαντικό ότι λαμβάνουμε υπόψιν και άλλους χρηματιστηριακούς δείκτες, που εν μέρει δικαιολογείται καθώς οι τιμές του ενός χρηματιστηρίου επηρεάζουν και τις τιμές των υπολοίπων. Την από κοινού χρήση μοντέλων πρόβλεψης χρησιμοποίησαν και οι [13]. Αναλυτικά, χρησιμοποιώντας τεχνητά νευρωνικά δίκτυα, δέντρα απόφασης και την μέθοδο των κ πλησιέστερων γειτόνων κατάφεραν να πετύχουν 65% ακρίβεια για τον χρηματιστηριακό δείκτη Dow Jones.

Χρήση εξωτερικών παραγόντων πραγματοποιήθηκε από τους [3] για την πρόβλεψη της κίνησης του χρηματιστηρίου Dow Jones. Συγκεκριμένα, ο στόχος της έρευνας ήταν να χρησιμοποιηθούν εξωτερικοί δείκτες όπως οι τιμές των βασικών εμπορευμάτων και η συναλλαγματική ισοτιμία για την πρόβλεψη του χρηματιστηρίου. Η υπόθεση που πραγματοποιήθηκε ήταν ότι τα κέρδη των επιχειρήσεων επηρεάζονται και από εσωτερικούς παράγοντες όπως η ανάπτυξη ενός προϊόντος ή μίας υπηρεσίας αλλά και από εξωτερικούς παράγοντες όπως το κόστος ενέργειας και η συναλλαγματική ισοτιμία ξένων χωρών. Στα πλαίσια της εργασίας αυτής, δημιουργήθηκε επίσης και ένα σύνολο δεδομένων από καθημερινές τιμές του Dow Jones, που προέρχονται από τεχνικούς εξωτερικούς δείκτες. Ο στόχος του μοντέλου μάθησης που υλοποίησαν ήταν να προβλέψουν την τιμή του κλεισίματος της τρέχουσας μέρας δίνοντας το άνοιγμα της ημέρας αυτής και άλλες εισόδους όπως τις τιμές των ανοιγμάτων κάποιες μέρες πριν. Όπως και στην προηγούμενη εργασία, εφαρμόστηκε το πολυστρωματικό εμπροσθοτροφοδοτούμενο νευρωνικό δίκτυο (Multilayer Feedforward Networks) με τον ίδιο αλγόριθμο διάδοσης λάθους. Τα αποτελέσματα έδειξαν ότι προσθέτοντας εξωτερικούς δείκτες στο διάνυσμα που δίνεται ως είσοδος στο νευρωνικό δίκτυο, η συνολική απόδοση ως προς την κερδοφορία που επιφέρει η πρόβλεψη του μοντέλου έχει βελτιωθεί σημαντικά. Η διαφορά έναντι της εργασίας που παρουσιάζουμε έγκειται στο γεγονός ότι λαμβάνουμε υπόψιν ως εξωτερικούς παράγοντες για έναν δείκτη τόσο κείμενα τα οποία προέρχονται από καθημερινά νέα όσο και τιμές από άλλα διεθνή χρηματιστήρια.

Μία πηγή εξωτερικών παραγόντων που μπορεί να επηρεάσει την κίνηση του χρηματιστηρίου στις μέρες μας είναι τα μέσα κοινωνικής δικτύωσης. Συγκεκριμένα, γίνεται η υπόθεση ότι η τα μέσα κοινωνικής δικτύωσης αναπαριστούν την κοινή γνώμη και το συναίσθημα σχετικά με τα τρέχοντα γεγονότα, ειδικότερα, μέσα κοινωνικής δικτύωσης που σχετίζονται περισσότερο με τα κοινωνικό πολιτικά ζητήματα όπως είναι το twitter. Οι [12]

χρησιμοποιήσαν τα τιτιβίσματα από το twitter που σχετίζονται με μία εταιρεία για να δείξουν αν υπάρχει συσχέτιση μεταξύ αυτών των σχολίων και την πτώση ή άνοδο της εταιρείας στο χρηματιστήριο του Dow Jones. Για να επιλύσουν το πρόβλημα αυτό, θεώρησαν ότι έπρεπε να εφαρμοστεί ανάλυση συναισθημάτων για τα τιτιβίσματα που προήλθαν για μία εταιρεία. Ουσιαστικά, με την ανάλυση αυτή, μπορεί να διαπιστωθεί αν τα τιτιβίσματα αυτά είναι θετικά ή αρνητικά. Έτσι, γίνεται η υπόθεση ότι με βάση τα θετικά σχόλια οι άνθρωποι είναι πιο διατεθειμένοι να επενδύσουν για την συγκεκριμένη εταιρεία. Έτσι, χρησιμοποιήθηκαν μοντέλα αναπαράστασης γλώσσας, το Word2Vec και Ngrams και μετέτρεψαν το πρόβλημα της ανάλυσης συναισθήματος σε πρόβλημα ταξινόμησης. Τα αποτελέσματα έδειξαν ότι η μεθοδολογία που ακολούθησαν όντως υποδηλώνει ισχυρή συσχέτιση μεταξύ της πορείας της εταιρείας στο χρηματιστήριο με βάση τα σχόλια από το twitter. Σε αντίθεση με αυτή την εργασία, εμείς προτείνουμε την από κοινού εκπαίδευση ενός μοντέλου για την πρόβλεψη κίνησης του χρηματιστηρίου. Ουσιαστικά, υποθέτουμε ότι τα κείμενα των ειδήσεων επηρεάζουν την κίνηση του χρηματιστηρίου και προσπαθούμε να χτίσουμε ένα γραμμικό μοντέλο που θα παίρνει αποφάσεις με βάση τα κείμενα και τους διάφορους χρηματιστηριακούς δείκτες που προέρχονται από έναν δείκτη. Επίσης, η εργασία μας μελετάει την άνοδο ή την πτώση του χρηματιστηρίου και όχι μία συγκεκριμένη εταιρεία. Μια επέκταση του μοντέλου πρόβλεψης θα ήταν η δημιουργία ενός προτασιακού συστήματος το οποίο θα προτείνει σε χρηματιστές που να επενδύσουν ώστε να έχουν το μέγιστο δυνατό κέρδος. Στην κατεύθυνση αυτή οι [11] προτείνουν ένα προτασιακό σύστημα βασισμένο σε κανόνες συσχέτισης που αναλύουν τα χρηματιστηριακά δεδομένα και προτείνουν μετοχές. Χαρακτηριστικό αυτής της εργασίας είναι ότι το σύστημα που προτείνουν βρίσκει συσχετίσεις μεταξύ των μετοχών και προτείνει ένα χαρτοφυλάκιο ενώ οι υπάρχουσες τεχνικές προτείνουν την αγορά ή την πώληση μίας συγκεκριμένης μετοχής. Η μεθοδολογία που προτείνουν είναι η χρήση ασαφής λογικής. Σε αντίθεση με την εργασία που μελετάμε εμείς, η εργασία τους στηρίζεται σε σχέσεις μεταξύ μετοχών ενώ εμείς αντιθέτως ψάχνουμε τις σχέσεις μεταξύ δεικτών χρηματιστηρίου. Επίσης, για την συσχέτιση μεταξύ των χρηματιστηριακών δεικτών χρησιμοποιούμε γραμμικά μοντέλα παλινδρόμησης που διαφέρουν από τους κανόνες συσχέτισης αφού τα πρώτα ανήκουν στην κατηγορία επιβλεπόμενης μάθησης (άρα απαιτούνται οι κλάσεις των δεδομένων) ενώ οι κανόνες συσχέτισης ανήκουν στην μη επιβλεπόμενη μάθηση. Αυτό έχει ως αποτέλεσμα, η δική μας μεθοδολογία να αξιοποιεί περισσότερη πληροφορία για την δημιουργία ενός μοντέλου

πρόβλεψης. Αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks) χρησιμοποιήθηκαν από τους [14] για τον σχεδιασμό ενός χαρτοφυλακίου επενδύσεων όπως αναφέρθηκε παραπάνω, με την διαφορά ότι λαμβάνονται υπόψιν πολλαπλοί χρηματιστηριακοί δείκτες. Η υπόθεση που κάνουν στην συγκεκριμένη εργασία ταιριάζει με την δική μας, δηλαδή ότι η τάση ενός χρηματιστηριακού δείκτη μπορεί καλύτερα να προβλεφθεί με βάσει πολλούς άλλους δείκτες αντί για έναν. Ωστόσο η διαφορά έγκειται στο γεγονός ότι εμείς μελετάμε γραμμικές σχέσεις μεταξύ των χρηματιστηριακών δεδομένων και των δεικτών έναντι αυτών που στηρίζονται σε μη γραμμικά μοντέλα.

Την διαφορά της απόδοσης μεταξύ νευρωνικών δικτύων και γραμμικής παλινδρόμησης την μελέτησαν οι [1] στο πρόβλημα πρόβλεψης της τιμής του χρηματιστηρίου. Τα αποτελέσματα έδειξαν ότι τα νευρωνικά δίκτυα προβλέπουν σωστά τους δείκτες χρηματιστηρίου μέχρι 58%, 67% και 78% για καθημερινά, εβδομαδιαία και μηνιαία δεδομένα αντίστοιχα. Επίσης, έδειξαν ότι τα νευρωνικά δίκτυα γενικεύουν καλύτερα από τα μοντέλα γραμμικής παλινδρόμησης όταν χρησιμοποιούνται ως στρατηγικές συναλλαγών. Ωστόσο, δεν μελετήθηκε εάν οι χρηματιστηριακοί δείκτες επηρεάζουν ο ένας τον άλλον ούτε κατά πόσο εξωτερικοί παράγοντες (όπως καθημερινά νέα) μπορούν να επηρεάσουν την απόδοση των νευρωνικών δικτύων.

Τα οικονομικά νέα είναι ένας παράγοντας που μπορεί να επηρεάσει την πρόβλεψη των τιμών στον χρηματιστήριο. Οι [17] χρησιμοποιήσαν διάφορες αναπαράστασης κειμένου (σάκος λέξεων, ονομαστικές φράσεις και ονόματα οντοτήτων) για να αναπαραστήσουν 9 χιλιάδες οικονομικά κείμενα. Χρησιμοποιώντας τις αναπαραστάσεις κειμένου ως χαρακτηριστικά, δημιούργησαν ένα μοντέλο πρόβλεψης το οποίο προβλέπει την τιμή του χρηματιστηρίου 20 λεπτά μετά από την δημοσίευση κάποιου άρθρου ειδήσεων. Αντιθέτως με αυτήν την εργασία, εμείς χρησιμοποιούμε την πολικότητα των κειμένων μαζί με τις τιμές άλλων χρηματιστηριακών δεικτών.

3

Θεωρητικό υπόβαθρο

Το πρόβλημα που κλήθηκε να μελετήσει η συγκεκριμένη εργασία είναι ένα πρόβλημα επιβλεπομένης μάθησης (Supervised Machine learning). Επομένως, είναι σημαντικό να αναφερθεί το θεωρητικό πλαίσιο της κατηγορίας αυτής. Έπειτα, θα εξεταστούν κάποια μοντέλα μηχανικής μάθησης που έχουν προταθεί από την βιβλιογραφία που εξετάστηκε στο κεφάλαιο 2 και θα αναφερθεί το μοντέλο μάθησης στο οποίο θα στηριχτεί αυτή η εργασία. Στην συνέχεια, θα γίνει μία ανάλυση στο πρόβλημα ανάλυσης συναισθήματος από κείμενα καθώς η πολικότητα των κειμένων ως θετικά ή αρνητικά θα είναι μία παράμετρος του μοντέλου που πραγματεύεται η εργασία. Τέλος, θα αναφέρουμε την έννοια της χρονοσειράς καθώς το πρόβλημα σχετίζεται άμεσα με αυτόν τον όρο.

3.1 Κατηγορίες Μηχανικής Μάθησης

Η περιοχή της μηχανικής μάθησης ασχολείται με την δημιουργία μοντέλων μάθησης τα οποία είναι ικανά από δεδομένα που δίνονται από το περιβάλλον στο οποίο δραστηριοποιούνται να κωδικοποιούν την γνώση για μελλοντικές προβλέψεις. Συγκεκριμένα οι κατηγορίες που συναντώνται πιο συχνά είναι οι εξής:

1. Επιβλεπόμενη Μηχανική Μάθηση (Supervised Machine Learning)
2. Μη Επιβλεπόμενη Μηχανική Μάθηση (Unsupervised Machine Learning)
3. Ενισχυτική Μηχανική Μάθηση (R*einforcement* Machine Learning)

3.1.1 Επιβλεπόμενη Μηχανική Μάθηση (Supervised Machine Learning)

Στην επιβλεπόμενη μηχανική μάθηση ο στόχος είναι το μοντέλο να μάθει μία μεταβλητή στόχο. Συγκεκριμένα, δίνονται ως είσοδο σε ένα μοντέλο επιβλεπόμενης μάθησης κάποια χαρακτηρίστηκα και γι' αυτά τα χαρακτηριστικά αντιστοιχεί μία κατηγορία (classification task) ή μία συνεχόμενη τιμή (regression task). Το μοντέλο μαθαίνει την κατανομή της μεταβλητή στόχου και με αυτόν τον τρόπο είναι σε θέση να προβλέψει την μελλοντική τιμή. Η πρόβλεψη μελλοντικών τιμών αναφέρεται στην διαδικασία όπου δίνουμε στο μοντέλο ως είσοδο τις τιμές των χαρακτηριστικών αλλά σε αυτήν την περίπτωση δεν υπάρχει τιμή για την μεταβλητή στόχο, έτσι, το μοντέλο θα είναι σε θέση να προβλέψει την τιμή με βάση τις τιμές των χαρακτηριστικών. Κάποια παραδείγματα που μπορούν να θεωρηθούν ως προβλήματα ταξινόμησης (classification) είναι η σημασιολογική δεικτοδότηση κειμένου, όπου ως είσοδο δίνονται τα κείμενα και προσπαθούμε να προβλέψουμε σε ποια κατηγορία αυτά τα κείμενα ανήκουν, η κατηγοριοποίηση ενός email αν είναι spam ή όχι κ.α. Από, την άλλη ως προβλήματα παλινδρόμησης μπορεί να θεωρηθεί η πρόβλεψη του τζίρου μιας επιχείρησης για το 2021 δεδομένου ότι έχουμε τους τζίρους των προηγούμενων ετών [20].

3.1.2 Μη Επιβλεπόμενη Μηχανική Μάθηση (Unsupervised Machine Learning)

Σκοπός της μη επιβλεπόμενης μάθησης είναι η ομαδοποίηση δεδομένων ώστε δεδομένα που ανήκουν στην ίδια ομάδα να είναι πιο κοντά (με βάση κάποια συνάρτηση ομοιότητας ή απόστασης) ενώ οι ομάδες θα πρέπει να είναι διακριτές και να απέχουν αρκετά η μία από την άλλη. Η κατηγορία αυτή λέγεται μη επιβλεπόμενη καθώς δεν υπάρχει περεταίρω γνώση για τον διαχωρισμό δεδομένων παρά μόνο τα ίδια τα δεδομένα και μία συνάρτηση που εφαρμόζεται σε αυτά.

3.1.3 Ενισχυτική Μηχανική Μάθηση (Reinforcement Machine Learning)

Από την άλλη ο ρόλος της ενισχυτικής μάθησης είναι να μιμηθεί την διαδικασία μάθησης με ενίσχυση. Δηλαδή, ο πράκτορας (το μοντέλο μάθησης) προσπαθεί να μάθει μία συγκεκριμένη πολιτική με βάση τις ενέργειες που πράττει. Για παράδειγμα, κατά την εκπαίδευση ενός μοντέλου που μαθαίνει να παίζει σκάκι, το μοντέλο θα παίρνει μία

ανταμοιβή κάθε φορά που μία ενέργεια του είναι σωστή (π.χ. προκαλεί ματ) ενώ αντιθέτως θα λαμβάνει μία ποινή.

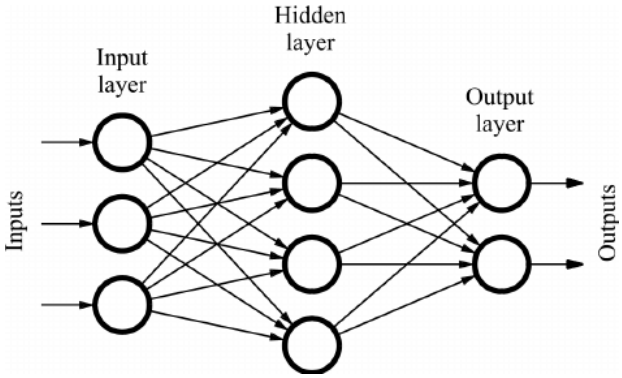
3.2 Τεχνικές Μηχανικής Μάθησης

Στην ενότητα αυτή θα αναφερθούμε στα μοντέλα παλινδρόμησης ή ταξινόμησης που έχουν προταθεί για την επίλυση προβλημάτων παλινδρόμησης ή ταξινόμησης και θα επικεντρώσουμε την ανάλυσή μας στο γραμμικό μοντέλο παλινδρόμησης που είναι και ο στόχος της συγκεκριμένης εργασίας.

3.2.1 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks)

Μία κατηγόρια αλγορίθμων μηχανικής μάθησης είναι τα τεχνητά νευρωνικά δίκτυα. Το χαρακτηριστικό αυτών των δικτύων είναι ότι η συνάρτηση που μαθαίνουν από τα δεδομένα εισόδου μπορεί να «αιχμαλωτίζει» μη γραμμικές σχέσεις. Με αυτόν τον τρόπο, προβλήματα τα οποία δεν μπορούν να λυθούν με τους παραδοσιακούς τρόπους μάθησης εξαιτίας της δομής των δεδομένων θα μπορούσαν να λυθούν με κάποια αρχιτεκτονική νευρωνικού δικτύου. Συγκεκριμένα, ένα τεχνητό νευρωνικό δίκτυο αποτελείται από απλές μονάδες επεξεργασίας (νευρώνες) που επιδρούν τα δεδομένα εισόδου και αυτές οι μονάδες επικοινωνούν μεταξύ τους. Το κίνητρο των νευρωνικών δικτύων ήταν η μίμηση της δομής ενός πραγματικού εγκεφάλου, παρόλα αυτά πρέπει να σημειωθεί ότι οι αρχιτεκτονικές που προτείνονται διαφέρουν σημαντικά από αυτές ενός βιολογικού οργανισμού. Σημαντικό ζήτημα που απασχολεί την κοινότητα των ερευνητών που ασχολούνται με αυτό το ζήτημα είναι η ύπαρξη δεδομένων, καθώς για κάποια προβλήματα δεν υπάρχουν πολλά δεδομένα, ενώ ένα δεύτερο ζήτημα είναι η υπερμοντελοποίηση (overfitting) όπου το μοντέλο έχει εκπαιδευτεί και προσαρμοστεί στα εκάστοτε δεδομένα με αποτέλεσμα να μην μπορεί να γενικεύσει σε νέα δεδομένα. Διακρίνουμε τρία είδη νευρωνικών δικτύων. Τα δίκτυα πρόσθιας τροφοδότησης, τα ανατροφοδοτούμενα νευρωνικά δίκτυα και τα βαθιά νευρωνικά.

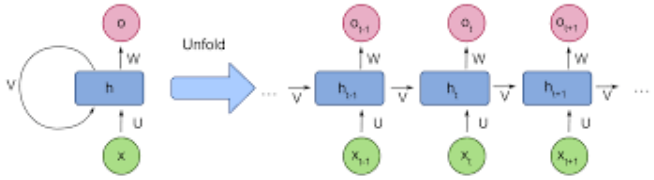
3.2.1.1 Πρόσθιας Τροφοδότησης(Feedforward neural network)



Εικόνα 1: Νευρωνικό Δίκτυο Πρόσθιας Τροφοδότησης

Τα νευρωνικά μοντέλα αυτού του τύπου εκπαιδεύονται μέσω της τεχνικής διάδοσης του λάθους προς τα πίσω. Συγκεκριμένα ένα πρόσθιας τροφοδότησης νευρωνικό δίκτυο αποτελείται από νευρώνες που είναι διατεταγμένοι σε επίπεδα (εικόνα 1). Το πρώτο επίπεδο καλείται επίπεδο εισόδου και το τελευταίο επίπεδο, επίπεδο εξόδου, ενώ τα μεταξύ των δύο επίπεδα καλούνται κρυφά. Για περαιτέρω μελέτη ο αναγνώστης μπορεί να ανατρέξει στην εργασία [18], όπου αναφέρεται όλο το πλαίσιο μάθησης ενός τέτοιου είδους νευρωνικού δικτύου.

3.2.1.2 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (Recurrent neural network)

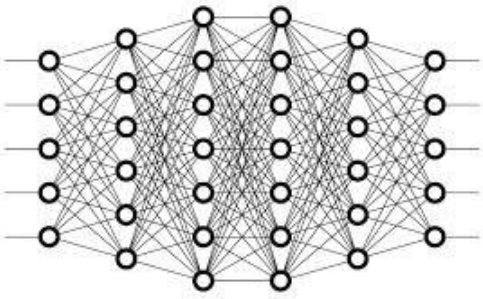


Εικόνα 2: Ανατροφοδοτούμενα Νευρωνικά Δίκτυα

Σε αντίθεση με τα νευρωνικά δίκτυα πρόσθιας τροφοδότησης, τα ανατροφοδοτούμενα νευρωνικά δίκτυα, μπορούν να χρησιμοποιήσουν την εσωτερική τους μνήμη για να επεξεργαστούν ακολουθίες εισόδων. Για παράδειγμα, το γνωστό πρόβλημα της επεξεργασίας φυσικής γλώσσας που σχετίζεται με την πρόβλεψη της επόμενης λέξης, δοθέντος ενός παραθύρου *κ λέξεων* θα μπορούσε να λυθεί με την χρήση ανατροφοδοτούμενων νευρωνικών δικτύων θεωρώντας τις λέξεις ως χρονικές στιγμές (timestamps) το μοντέλο θα είναι σε θέση να μάθει το πλαίσιο (context) της λέξης που θέλουμε να προβλέψουμε δεδομένων των *κ λέξεων* που προηγούνται αυτής της λέξης. Στη

εικόνα 2 αριστερά βλέπουμε την αρχιτεκτονική του δικτύου και στα δεξιά τον τρόπο όπου αυτό «ξεδιπλώνεται» και λαμβάνει την είσοδο ως χρονικές στιγμές. Ένα άλλο παράδειγμα είναι η κίνηση ενός χρηματιστηριακού δείκτη. Στο μοντέλο δίνονται τα κλεισίματα των προηγούμενων ημερών ενός δείκτη και το μοντέλο πρέπει να μάθει να προβλέπει το κλείσιμο της επόμενης μέρας [22].

3.2.1.3 Βαθιά Νευρωνικά Δίκτυα (Deep neural networks)



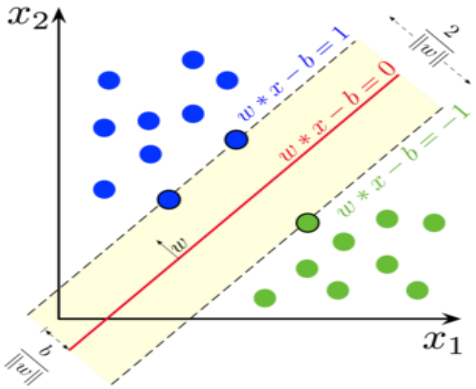
Εικόνα 3:Βαθιά Νευρωνικά Δίκτυα

Η επιστημονική κοινότητα αντιλήφθηκε ότι όσα περισσότερα κρυφά επίπεδα έχει ένα νευρωνικό δίκτυο, τόσο πιο πολύ πληροφορία μπορεί να μάθει. Έτσι εισάγονται τα βαθιά νευρωνικά τα οποία είναι μία γενίκευση των απλών νευρωνικών δικτύων με πολλαπλά κρυφά επίπεδα. Με βάση την μοντέρνα αντίληψη για τα νευρωνικά δίκτυα κάθε επίπεδο του δικτύου μαθαίνει και κάτι διαφορετικό. Για παράδειγμα, θεωρώντας ένα μοντέλο αναπαράστασης γλώσσας όπως είναι το BERT, έρευνες έχουν δείξει ότι σε κάθε επίπεδο του νευρωνικού, το μοντέλο είναι ικανό να μάθει διάφορου τύπου πληροφορίες, όπως η σύνταξη της γλώσσας, η σημασιολογία και άλλες πιο λεπτές έννοιες που δεν μπορούσαν να κωδικοποιήσουν άλλα μοντέλα. Ωστόσο, πρέπει να σημειωθεί ότι τα βαθιά νευρωνικά δίκτυα είναι μια οικογένεια νευρωνικών δικτύων και όχι ένα συγκεκριμένο δίκτυο [7].

3.2.2 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machine)

Οι μηχανές διανυσμάτων υποστήριξης [16] είναι μία ομάδα αλγορίθμων που ανήκουν στην επιβλεπόμενη μάθηση για προβλήματα ταξινόμησης αλλά και παλινδρόμησης. Σκοπός των αλγορίθμων αυτών είναι η αρχή του κατασκευαστικού ρίσκου (SRM) που έχει αποδειχθεί ότι υπερτερεί έναντι της ελαχιστοποίησης του εμπειρικού ρίσκου (ERM) στην οποία στηρίζονται

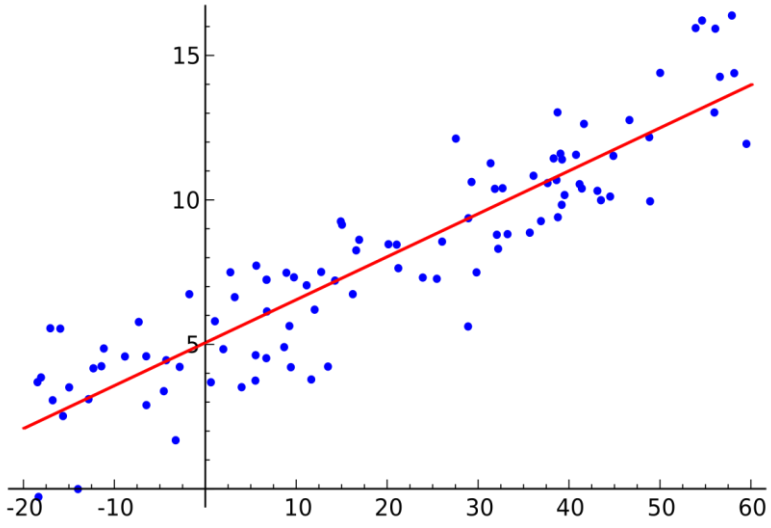
τα νευρωνικά δίκτυα. Συγκεκριμένα, σκοπός των μηχανών διανυσμάτων υποστήριξης είναι η εύρεση ενός υπερεπιπέδου που διαχωρίζει τα δεδομένα δημιουργώντας το μέγιστο περιθώριο. Μία τεχνική που εφαρμόζεται για την δημιουργία του υπερεπιπέδου είναι το τέχνασμα του πυρήνα.



Εικόνα 4:Ταξινόμηση με χρήση Μηχανών Διανυσμάτων Υποστήριξης

3.2.3 Γραμμική Παλινδρόμηση (Linear Regression)

Η χρήση της γραμμικής παλινδρόμησης έχει μεγάλη σημασία στην δημιουργία μοντέλων πρόβλεψης και η σημασία είναι αρκετά μεγάλη στην επιστήμη των υπολογιστών. Όπως υποδηλώνει και το όνομα του τίτλου της μεθόδου, η γραμμική παλινδρόμηση χρησιμοποιείται για προβλήματα επιβλεπόμενης μάθησης όπου η μεταβλητή στόχος είναι συνεχόμενη μεταβλητή.



Εικόνα 5: Μοντέλο Γραμμικής Παλινδρόμησης

Η σημαντική υπόθεση που κάνει η μέθοδος είναι ότι υπάρχει μια γραμμική σχέση μεταξύ των ανεξάρτητων μεταβλητών (τα χαρακτηριστικά) και της εξαρτημένης μεταβλητής στόχου. Δίνοντας έναν αυστηρό ορισμό του προβλήματος έχουμε:

Δεδομένου ενός συνόλου δεδομένων:

$$\{y_i, x_{i1}, \dots, x_{ip}\}^n_{i=p}$$

ένα μοντέλο γραμμικής παλινδρόμησης υποθέτει ότι υπάρχει γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής y και του p-διανύσματος των ανεξάρτητων μεταβλητών x. Η σχέση αυτή μοντελοποιείται θεωρώντας και έναν όρο ε που αναπαριστά το σφάλμα που προκύπτει κατά την διαδικασία της μάθησης και προσθέτει θόρυβο στην γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής με τις ανεξάρτητες μεταβλητές. Το μοντέλο που προκύπτει ορίζεται ως:

$$y_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} + \varepsilon_i = x_i^T b + \varepsilon_i \quad \text{όπου} \quad i = 1, \dots, n$$

Όπου το T ορίζει την αντιστροφή του διανύσματος του x ενώ ο όρος $x_i^T b$ συμβολίζει το εσωτερικό γινόμενο μεταξύ του διανύσματος και των βαρών.

Θέλοντας να ακολουθήσουμε την σημειογραφία των πινάκων μπορούμε να ορίσουμε το πρόβλημα ως:

$$\boldsymbol{y} = \mathbf{X}\mathrm{b} + \varepsilon$$

Όπου

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} x_1^T \\ x_2^T \\ \dots \\ x_n^T \end{pmatrix} \qquad b = \begin{pmatrix} b_0 \\ b_1 \\ \dots \\ b_p \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots \\ \varepsilon_n \end{pmatrix}$$

Έστω ότι έχουμε ένα σύνολο δεδομένων που αποτελείται από τις ανεξάρτητες μεταβλητές μας και την τιμή της εξαρτημένης μεταβλητής y. Χωρίζουμε τις παρατηρήσεις μας σε δύο ομάδες (train set, test set) και δημιουργούμε την καμπύλη παλινδρόμησης βάσει των δεδομένων που περιέχονται στο train set και βάσει αυτής προβλέπουμε την έξοδο των παρατηρήσεων του test set. Με αυτό τον τρόπο μπορούμε να εξετάσουμε την απόδοση του μοντέλου μας και το κατά πόσο μπορεί να προβλέψει την έξοδο μελλοντικών παρατηρήσεων.

Να σημειωθεί ότι η γραμμική παλινδρόμηση:

- Περιορίζεται σε γραμμικά προβλήματα, διότι από την φύση της εξετάζει γραμμικές σχέσεις.
- Εστιάζει στη μέση τιμή της εξαρτημένης μεταβλητής.

- Επηρεάζεται από ακραίες τιμές παρατηρήσεων.
- Τα δεδομένα πρέπει να είναι ανεξάρτητα μεταξύ τους.

3.3 Ανάλυση Συναισθήματος

Το πρόβλημα ανάλυσης συναισθήματος [8] είναι ένα πρόβλημα επεξεργασίας φυσικής γλώσσας, ανάλυσης κειμένου και υπολογιστικής γλωσσολογίας και ο σκοπός είναι η αναγνώριση, ποσοτικοποίηση και μελέτη συναισθηματικών καταστάσεων και υποκειμενικότητας. Ένα παράδειγμα που σχετίζεται με αυτό το πρόβλημα είναι η ταξινόμηση σχολίων χρηστών ως θετικά ή αρνητικά. Με αυτόν τον τρόπο για παράδειγμα, μία επιχείρηση μπορεί να γνωρίζει την απόδοση ενός προϊόντος στην αγορά. Για την επίλυση αυτού του προβλήματος έχουν προταθεί πολλοί αλγόριθμοι ενώ κυριαρχούν αυτοί που προέρχονται από την επιστημονική περιοχή της μηχανικής μάθησης.

3.4 Χρονολογικές Σειρές

Σε διάφορες επιστημονικές περιοχές όπως είναι η σεισμολογία, η εφαρμοσμένη ιατρική και η οικονομική επιστήμη απαιτείται η πρόγνωση μελλοντικών καταστάσεων για την λήψη αποφάσεων. Έτσι, εισάγεται ο όρος μέθοδοι πρόβλεψης και διακρίνουμε δύο κατηγορίες (1) ποιοτικές μεθόδους πρόβλεψης και (2) ποσοτικές μεθόδους πρόβλεψης. Στην συγκεκριμένη εργασία, το ενδιαφέρον μας στρέφεται στην δεύτερη κατηγορία εφόσον στόχος μας είναι η πρόβλεψη κίνησης του χρηματιστηρίου.

Οι ποσοτικές μέθοδοι πρόβλεψης κάνουν χρήση ιστορικών δεδομένων. Συγκεκριμένα στόχος των μοντέλων πρόβλεψης της κατηγόριας αυτής είναι η μελέτη παρελθοντικών στοιχείων για την βαθύτερη κατανόηση της δομής των δεδομένων, η οποία με την σειρά της θα οδηγήσει στα απαραίτητα μέσα για την πρόβλεψη μελλοντικών ενδεχομένων. Διακρίνουμε δύο είδη τεχνικών πρόβλεψης που βασίζονται σε μοντέλα ανάλυσης χρονολογικών σειρών και σε αιτιώδη μοντέλα.

Τα αιτιώδη μοντέλα πρόβλεψης, στοχεύουν στον προσδιορισμό παραγόντων που επιδρούν στις διακυμάνσεις της τυχαίας μεταβλητής, για την οποία επιθυμούμε πρόβλεψη. Ως παράδειγμα τυχαίας μεταβλητής μπορούμε να θεωρήσουμε την άνοδο ή την πτώση του

χρηματιστηρίου. Συνεπώς, στα αιτιώδη μοντέλα επιτυγχάνεται πρόβλεψη των τιμών μιας τυχαίας μεταβλητής, έχοντας αποκτήσει προηγουμένως γνώση για τις αιτίες διακύμανσης αυτής.

Τα μοντέλα ανάλυσης χρονολογικών σειρών βασίζονται στην υπόθεση ότι τα ιστορικά στοιχεία συνιστούν τιμές εξαρτώμενες από το χρόνο και καλούνται χρονοσειρές. Δηλαδή, οι τιμές της χρονοσειράς εξαρτώνται από το χρόνο με την έννοια ότι επηρεάζονται από τις τιμές των παρατηρήσεων που προηγούνται χρονικά.

3.4.1 Μέθοδοι πρόβλεψης

Οι μέθοδοι πρόβλεψης που συναντώνται στις χρονολογικές σειρές συνοψίζονται παρακάτω:

- Κινητός Μέσος όρος
- Εκθετική Εξομάλυνση
- Αποσύνθεση Χρονοσειρών
- Ανάλυση Παλινδρόμησης
- Τεχνικές Box-Jenkins
- Οικονομετρικά Μοντέλα

3.4.2 Μοτίβα Δεδομένων

Σε μια χρονολογική σειρά αναγνωρίζονται τέσσερα βασικά μοτίβα:

- Τάση → οι τιμές της μεταβλητής μεταβάλλονται, τείνοντας εν γένει να αυξάνουν ή να φθίνουν κατά τη διάρκεια ολόκληρης της χρονικής περιόδου
- Εποχικά → οι τιμές της μεταβλητής αυξομειώνονται και οι μεταβολές αυτές εξαρτώνται από ημερολογιακές περιόδους. Για παράδειγμα, η κατανάλωση αερίου θα είναι υψηλότερη τους χειμερινούς μήνες έναντι των μηνών του καλοκαιριού
- Κυκλικά → Οι ανοδικές και καθοδικές κυμάνσεις είναι μεταβλητού χρονικού και ποσοτικού εύρους.
- Άρρυθμα → Οι τιμές της μεταβλητής παρουσιάζουν αυξομειώσεις που δεν είναι δυνατόν να ερμηνευθούν ή να προβλεφθούν.

4

Πρόβλεψη Ανόδου ή Πτώσης Χρηματιστηρίου

Η πρόβλεψη του χρηματιστηρίου όπως αναφέρθηκε στην εισαγωγή αποτελεί μία πρόκληση ενώ ένα καλά εκπαιδευμένο μοντέλο πρόβλεψης θα βοηθούσε σημαντικά στις αποφάσεις που παίρνουν καθημερινά επενδυτές ανά τον κόσμο. Στον κεφάλαιο αυτό, εισάγουμε την έννοια του χρηματιστηριακού δείκτη που θα μας απασχολήσει στην μοντελοποίηση. Στην συνέχεια, εισάγουμε το πρόβλημα πρόβλεψης της ανόδου ή πτώσης του χρηματιστηρίου ως πρόβλημα μηχανικής μάθησης και δίνουμε βασικούς ορισμούς.

4.1 Χρηματιστηριακοί Δείκτες

Ως χρηματιστηριακό δείκτη εννοούμε μία μέτρηση της συνολικής αξίας του χρηματιστηρίου με βάση τις μέσες τιμές. Αναλυτικά, αποτελεί έναν σταθμισμένο μέσο όρο, επομένως οι πιο ακριβές μετοχές τον επηρεάζουν περισσότερο από ό,τι οι μετοχές με πιο χαμηλές τιμές. Κατά το κλείσιμο του χρηματιστηρίου ο χρηματιστηριακός δείκτης διαμορφώνεται ανάλογα με το βάρος των τιμών που διαθέτουν κάποιες σημαντικές μετοχές μέσα στο χρηματιστήριο.

Στην ουσία ένας χρηματιστηριακός δείκτης αποτελείται από ένα σύνολο μετοχών από εταιρείας οι οποίες είναι εισηγμένες σε μία ή περισσότερες χρηματιστηριακές αγορές σύμφωνα πάντα με την ικανοποίηση σημαντικών κριτηρίων (η μελέτη τους δεν είναι της παρούσης). Οι καθοδικές κι ανοδικές τάσεις των μετοχών αυτών αντικατοπτρίζουν την τιμή του χρηματιστηριακού δείκτη.

Οι χρηματιστηριακοί δείκτες είναι ένα εργαλείο που χρησιμοποιούν οι επενδυτές για να περιγράψουν την αγορά και ορίζονται διάφοροι τύποι εξ' αυτών. Οι τύποι δεικτών μπορούν να ταξινομηθούν με διάφορους τρόπους. Για παράδειγμα, ένας παγκόσμιος δείκτης χρηματιστηρίου περιλαμβάνει μετοχές από πολλές περιοχές. Όπου περιοχή μπορεί να θεωρηθεί γεωγραφικά (π.χ. Ευρώπη, Ασία, Αμερική) ή σε επίπεδα εισοδήματος (π.χ. Αναπτυγμένες αγορές).

Πολλοί επενδυτές παρακολουθούν και χρησιμοποιούν τα επίπεδο τιμών ενός χρηματιστηριακού δείκτη ώστε να διαμορφώσουν μία εικόνα σχετικά με τις τάσεις της αγοράς. Από τις διακυμάνσεις του δείκτη αξιολογείται η εμπιστοσύνη του επενδυτικού κοινού ως προς τις τιμές των μετοχών. Οι χρηματιστηριακοί δείκτες χωρίζονται σε δύο βασικές κατηγορίες οι οποίες είναι:

1. **Οι Πολυκλαδικοί Δείκτες (Board Based Indices):** Οι δείκτες αυτοί, οι οποίοι είναι και το αντικείμενο μελέτης της παρούσας εργασίας, παρακολουθούν τη συμπεριφορά συγκεκριμένων μετοχών και κατ' επέκταση εταιρειών διαφορετικών κλάδων. Τέτοιοι δείκτες είναι για παράδειγμα ο Dow Jones Industrial 30 (ΗΠΑ), ο FTSE 100 (Αγγλία), ο CAC 40 (Γαλλία), S&P 500 (ΗΠΑ), κ.ά.
2. **Οι Κλαδικοί ή Συγκεντρωτικοί Δείκτες (Narrow Based Indices):** Οι δείκτες αυτοί μετράνε και παρακολουθούν τη συμπεριφορά συγκεκριμένων μετοχών ενός κλάδου όπως για παράδειγμα είναι ο τραπεζικός, ο κατασκευαστικός, ο επενδυτικός, κ.ά.

Από την άλλη ένας εθνικός δείκτης αναπαριστά την απόδοση της χρηματαγοράς ενός έθνους και κατά προσέγγιση αντανακλά την θέληση για επενδύσεις μίας εθνικής οικονομίας. Κάποιοι από τους πιο γνωστούς δείκτες που θα μελετηθούν στην συγκεκριμένη εργασία είναι οι εξής:

1. Χρηματιστηριακός δείκτης της Γερμανίας (DAX)
2. Χρηματιστηριακός δείκτης της Γαλλίας (CAC 40)
3. Χρηματιστηριακός δείκτης της Αγγλίας (FTSE 100)
4. Dow Jones
5. NASDAQ
6. S&P 500

Ένας από τους σημαντικότερους δείκτες όπου αποτελεί πεδίο έρευνας πολλών οικονομολόγων είναι ο παγκόσμιος χρηματιστηριακός δείκτης MSCI World. Ο δείκτης MSCI World πρωτοεμφανίστηκε το 1969 ενώ το 1970 έγινε η πρώτη μέτρησή του. Η διαμόρφωση ενός τέτοιου δείκτη, είχε ως στόχο την ανάλυση αλλά και την επίβλεψη των παγκόσμιων χρηματαγορών ανάλογα με τις διακυμάνσεις των σημαντικότερων μετοχών που επηρεάζουν την παγκόσμια αγορά του χρηματιστηρίου. Αυτός ο δείκτης διαμορφώνεται από μετοχές οι οποίες έχουν σημαντική βαρύτητα τιμών στην εγχώρια οικονομία των διαφόρων κρατών που περιλαμβάνονται στην λίστα διαμόρφωσης του παγκόσμιου χρηματιστηριακού δείκτη.

Ο παγκόσμιος χρηματιστηριακός δείκτης μας δίνει μία συνολική εικόνα της παγκόσμιας οικονομίας σε περιόδους οικονομικής ύφεσης ή ανάπτυξης. Έτσι μελετώνται οι παράγοντες εκείνοι οι οποίοι θεωρούνται ότι επηρέασαν την πτώση ή την άνοδο των μετοχών του χρηματιστηρίου σε παγκόσμιο επίπεδο. Κατ’ αυτόν τον τρόπο οικονομολόγοι αλλά και επενδυτές μπορούν να αξιολογήσουν την κατάσταση της παγκόσμιας αγοράς και να κρίνουν σε βραχυχρόνιο επίπεδο την πορεία του χρηματιστηριακού δείκτη. Το πρόβλημα που προκύπτει όμως από τις βραχυχρόνιες αξιολογήσεις είναι η μη αποτελεσματική πρόβλεψη της απόδοσης των χρηματιστηρίων με αποτέλεσμα πολλοί επενδυτές αλλά και διάφοροι οικονομολόγοι να κρίνουν λανθασμένα την πορεία του εκάστοτε χρηματιστηριακού δείκτη που μελετούν.

Συνεπώς κρίθηκε απαραίτητη η δυνατότητα πρόβλεψης ενός χρηματιστηριακού δείκτη σε μακροχρόνιο επίπεδο διακύμανσης των τιμών του. Η πρόβλεψη των τιμών του δείκτη, μας δίνει μία εικόνα των μεταβολών των τιμών του ανοίγματος αλλά και του κλεισίματος των χρηματιστηριακών δεικτών με αποτέλεσμα να μας γνωστοποιείται η μεταβολή του σε συγκεκριμένη μελλοντική στιγμή.

Το πρόβλημα που έχουμε κληθεί να λύσουμε σχετίζεται με την πρόβλεψη του παγκόσμιου χρηματιστηριακού δείκτη MSCI. Χρησιμοποιώντας το μοντέλο της γραμμικής παλινδρόμησης, θα μπορέσουμε να εξάγουμε αποτελέσματα προβλέψεων των χρηματιστηριακών δεικτών για συγκεκριμένο χρονικό διάστημα κι έπειτα να διαμορφώσουμε τον παγκόσμιο χρηματιστηριακό δείκτη εξάγοντας σημαντικά συμπεράσματα ως προς την ακεραιότητα των τιμών σε μακροχρόνιο επίπεδο.

4.2 Μοντέλο για την Πρόβλεψη Ανόδου ή Πτώσης του Χρηματιστηρίου.

Σκοπός της εργασίας είναι η δημιουργία ενός μοντέλου που θα προβλέπει την άνοδο ή πτώση του παγκόσμιου δείκτη MSCI. Συγκεκριμένα, τίθεται το ερώτημα αν μπορούν οι δείκτες που εισαγάγαμε πρωτύτερα να χρησιμοποιηθούν για την πρόβλεψη του δείκτη MSCI. Πέρα από αυτό, είναι σημαντικό να εξεταστεί αν επιπλέον η χρήση κειμένων από καθημερινά νέα μπορεί να συμβάλει στην δημιουργία ενός έγκυρου εκτιμητή του παγκόσμιου δείκτη.
Επομένως το πρόβλημα ορίζεται ως εξής:

"Δοθέντος ενός συνόλου δεικτών και την εκτίμηση των ειδήσεων ως θετικές ή αρνητικές για την πορεία του χρηματιστηρίου, επιζητούμε ένα μοντέλο για την πρόβλεψη του παγκόσμιου δείκτη MSCI"

Άρα, επιζητούμε ένα μοντέλο το οποίο να μπορεί να βρει τις σχέσεις μεταξύ των δεικτών και του παγκόσμιου δείκτη. Επίσης, σημαντικό είναι να βρεθούν εκείνοι οι δείκτες που επηρεάζουν περισσότερο την κίνηση του παγκόσμιου δείκτη και να βγουν χρήσιμα συμπεράσματα για την σχέση των δεικτών.

5

Περιγραφή Μοντέλου Μηχανικής Μάθησης

Εισάγοντας το πρόβλημα πρόβλεψης ανόδου ή πτώσης του χρηματιστηρίου ως πρόβλημα μηχανικής μάθησης, μπορούμε να εφαρμόσουμε πολλές τεχνικές για την εκπαίδευση ενός καλού εκτιμητή. Στην συγκεκριμένη εργασία, μελετάμε την χρήση της γραμμικής παλινδρόμησης που λαμβάνει υπόψιν τους μεγαλύτερους χρηματιστηριακούς δείκτες ανά τον κόσμο όπως επίσης και κείμενα προερχόμενα από καθημερινά νέα. Επομένως, στο κεφάλαιο αυτό, εισάγουμε την επίλυση του προβλήματος πρόβλεψης με την εφαρμογή γραμμικής παλινδρόμησης.

5.1 Εφαρμογή γραμμικής παλινδρόμησης για την Πρόβλεψη Ανόδου ή Πτώσης του Χρηματιστηρίου

Ορίζοντας το πρόβλημα ως πρόβλημα μηχανικής μάθησης και συγκεκριμένα επιβλεπομένης μάθησης μπορούμε να ορίσουμε την σχέση (αν αυτή υπάρχει) μεταξύ των δεικτών που παρουσιάσθηκαν στο προηγούμενο κεφάλαιο και του παγκόσμιου δείκτη. Θα προσπαθήσουμε να λύσουμε το πρόβλημα με την χρήση γραμμικής παλινδρόμησης. Έτσι το πρόβλημα ορίζεται ως:

“Δοθέντος των τιμών κλεισίματος των X1,X2,…,XN δεικτών, δηλαδή των ανεξάρτητων μεταβλητών XN, την χρονική στιγμή t-1 θέλουμε να προβλέψουμε τον παγκόσμιο δείκτη, δηλαδή την εξαρτημένη μεταβλητή, την χρονική στιγμή t.”

Με την υπόθεση της γραμμικής παλινδρόμησης ορίζεται το πρόβλημα ως μαθηματικό πρόβλημα με τις εξής σχέσεις:

Δεδομένου ενός συνόλου:

$$\{y_i, x_{i1}, \dots, x_{ip}, z_{i1}\}^n_{i=p}$$

ένα μοντέλο γραμμικής παλινδρόμησης υποθέτει ότι υπάρχει γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής y (παγκόσμιος δείκτης) και του p-διανύσματος των ανεξάρτητων μεταβλητών x και της μεταβλητής z που παίρνει τιμές 0 ή 1 ανάλογα με το αν τα σχόλια για την συγκεκριμένη μέρα ήταν θετικά ή όχι. Η σχέση αυτή μοντελοποιείται θεωρώντας και έναν όρο ε που αναπαριστά το σφάλμα που προκύπτει κατά την διαδικασία της μάθησης και προσθέτει θόρυβο στην γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής y με τις ανεξάρτητες μεταβλητές x που μελετάμε. Το μοντέλο που προκύπτει ορίζεται ως:

$$y_i = b_0 + b_1 x_{i1} + \cdots + b_p x_{ip} + b_{p+1} x_{i(p+1)} + \varepsilon_i = x_i^T b + \varepsilon_i \quad \text{όπου} \quad i = 1, \dots, n$$

Όπου το T ορίζει την αντιστροφή του διανύσματος του x ενώ ο όρος $x_i^T b$ συμβολίζει το εσωτερικό γινόμενο μεταξύ του διανύσματος και των βαρών.
Επομένως το πρόβλημα μπορεί να λυθεί ως:

$$\boldsymbol{y} = \mathbf{X}\mathrm{b} + \varepsilon$$

Όπου

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} x_1^T \\ x_2^T \\ \cdots \\ x_n^T \end{pmatrix} \qquad b = \begin{pmatrix} b_0 \\ b_1 \\ \cdots \\ b_p \end{pmatrix} \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdots \\ \varepsilon_n \end{pmatrix}$$

Ωστόσο ο παραπάνω ορισμός του προβλήματος διαπιστώνουμε ότι δεν αντιστοιχεί ακριβώς στο πρόβλημα που προσπαθούμε να λύσουμε. Το πρόβλημα έγκειται στο γεγονός ότι για να προβλέψουμε την τιμή της εξαρτημένη μεταβλητή y (παγκόσμιο δείκτης) σήμερα θα πρέπει να έχουμε τις τιμές των ανεξάρτητών μεταβλητών x (δείκτες και τα κείμενα των ειδήσεων) χθες. Επομένως, ο χρόνος θα πρέπει να εισαχθεί στο μοντέλο ώστε να οριστεί ακριβώς το πρόβλημα.
Έτσι,

Δεδομένου ενός συνόλου:

$$\{y_t, x_{(t-1)1}, \dots, x_{(t-1)p}, z_{(t-1)1}\}_{i=p}^{n}$$

ένα μοντέλο γραμμικής παλινδρόμησης υποθέτει ότι υπάρχει γραμμική σχέση μεταξύ της εξαρτημένης μεταβλητής y την χρονική στιγμή t και του p-διανύσματος των ανεξάρτητων μεταβλητών x και της ανεξάρτητης μεταβλητής z που παίρνει τιμές 0 ή 1 ανάλογα με το αν τα σχόλια για την συγκεκριμένη μέρα ήταν θετικά ή όχι την χρονική στιγμή t-1. Το μοντέλο που προκύπτει ορίζεται ως:

$$y_t = b_0 + b_1 x_{(t-1)1} + \cdots + b_p x_{(t-1)p} + b_{p+1} x_{(t-1)(p+1)} + \varepsilon_{(t-1)}$$
$$= x_{(t-1)}^T b + \varepsilon_{(t-1)} \quad \text{όπου} \quad t = 1, \dots, n$$

Καθώς το πρόβλημα που έχουμε κληθεί να αντιμετωπίσουμε είναι η πρόβλεψη της ανόδου ή πτώσης χρηματιστηρίου και όχι η τιμή του παγκόσμιου δείκτη, η μεταβλητή στόχος ορίζεται ως:

$$y_t = \frac{(y_t - y_{t-1})}{y_t} * 100$$

Με αυτόν τρόπο ορίζουμε την άνοδο ή την πτώση του χρηματιστηρίου. Για παράδειγμα έστω ότι την χθεσινή μέρα η τιμή κλεισίματος του παγκόσμιου δείκτη ήταν 2151,82 και η τιμή η σημερινή 2153,83. Τότε υπολογίζοντας με τον παραπάνω τύπο προκύπτει +0.0933. Δηλαδή υπάρχει άνοδος του χρηματιστηρίου από χθες μέχρι σήμερα κατά +0.0933.

6

Αξιολόγηση

Στο κεφάλαιο αυτό, θα γίνει παρουσίαση των αποτελεσμάτων της εργασίας. Αρχικά θα ορίσουμε τις παραμέτρους που μελετήσαμε και θα εξηγήσουμε τον λόγο για τον οποίο επιλέξαμε τις παραμέτρους αυτές. Στην συνέχεια θα αναφερθούμε στις μετρικές που χρησιμοποιήθηκαν για να αξιολογήσουμε την τεχνική που υλοποιήσαμε. Έπειτα, αναφέρουμε τα πειράματα που πραγματοποιήσαμε συμπεριλαμβάνοντας το σύνολο δεδομένων για την εκπαίδευση του μοντέλου μάθησης. Εν συνεχεία, θα παρουσιαστούν τα αποτελέσματα της εργασίας και τέλος θα γίνει μία σύνοψη των αποτελεσμάτων.

6.1 Παράμετροι αξιολόγησης

Σκοπός μας είναι η δημιουργία ενός μοντέλου μάθησης το οποίο θα έχει υψηλότερη ακρίβεια. Επομένως, θα πρέπει να εξεταστούν ποιες μεταβλητές (χρηματιστηριακοί δείκτες και κείμενο) συνεισφέρουν περισσότερο στην ακρίβεια. Συγκεκριμένα, οι παράμετροι που θα εξεταστούν είναι η ύπαρξη όλων ή μερικών χρηματιστηριακών δεικτών κατά την διαδικασία μάθησης. Επίσης, η συνεισφορά των κειμένων ανάλογα με την πολικότητά τους (θετικά ή αρνητικά) θα πρέπει να εξεταστεί καθώς είναι ένας παράγοντας που επηρεάζει την εικόνα του χρηματιστηρίου.

6.2 Σύστημα αξιολόγησης

Για την αξιολόγηση του μοντέλου μάθησης θα χρησιμοποιηθούν δύο μετρικές.

6.2.1 *Ακρίβεια (Accuracy)*

Ο τύπος της ακρίβειας δίνεται από τον παρακάτω τύπο:

$$acc = \frac{TP + TN}{TP + TN + FP + FN}$$

Όπου TP και TN είναι οι προβλέψεις θετικές και αρνητικές που όντως είναι σωστές ενώ FP και FN είναι οι προβλέψεις που λανθασμένα θεωρήθηκαν θετικές ή αρνητικές αντίστοιχα. Η μετρική αυτή θα μας δείξει κατά πόσο το μοντέλο πρόβλεψης μπορεί να προβλέψει την άνοδο ή πτώση του παγκόσμιου χρηματιστηρίου.

6.2.2 *Ελάχιστο Τετραγωνικό Σφάλμα (Mean squared error)*

Εφόσον η μεταβλητή στόχος του μοντέλου γραμμικής παλινδρόμησης είναι ποσοτική, χρειαζόμαστε ένα μέτρο που να λαμβάνει αυτόν τον παράγοντα υπόψιν. Κατάλληλο μέτρο αξιολόγησης είναι η μέθοδος του ελάχιστου τετραγωνικού σφάλματος που ορίζεται ως ακολούθως:

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

Όπου Y είναι οι πραγματικές τιμές για n τιμές ενώ Z είναι οι εκτιμήσεις του μοντέλου πρόβλεψης.

6.2.3 *Συντελεστής Προσδιορισμού – R^2 (Coefficient of determination)*

Ο συντελεστής Προσδιορισμού ή αλλιώς R^2 είναι το ποσοστό της διακύμανσης της εξαρτημένης μεταβλητής, που θα γίνεται πρόβλεψη από τις ανεξάρτητες μεταβλητές. Η μέθοδος του R τετραγώνου ορίζεται ως εξής:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - Z_i)^2}{\sum_{i=1}^{n}(Y_i - K_i)^2}$$

Όπου Y είναι οι πραγματικές τιμές, Z είναι οι εκτιμήσεις του μοντέλου πρόβλεψης ενώ K είναι το άθροισμα των τετραγωνικών αποκλίσεων από το Y.

6.3 Οργάνωση πειραμάτων

6.3.1 Συλλογές δεδομένων

Για την συλλογή των δεδομένων ακολουθήσαμε την εξής διαδικασία:

1) Αρχικά κατεβάσαμε από την ιστοσελίδα investing.com τις τιμές κλεισίματος των χρηματιστηριακών δεικτών και του παγκόσμιου δείκτη στο διάστημα 04/2012 με 07/2016.
2) Έπειτα, από την ιστοσελίδα kaggle.com/aaron7sun/stocknews κατεβάσαμε το σύνολο δεδομένων με τις ειδήσεις που πραγματοποιήθηκαν και προσαρμόσαμε το σύνολο ώστε να συμβαδίζει με το σύνολο των χρηματιστηριακών δεικτών στο διάστημα 04/2012 μέχρι 07/2016

Εφόσον συλλέξαμε το σύνολο δεδομένων το χωρίσαμε σε δύο μέρη. Το σύνολο εκπαίδευσης αποτελείται από τις ημερομηνίες μέχρι τον 10/2015 ενώ το σύνολο δοκιμών από τις ημερομηνίες 10/2015 και έπειτα.

6.3.2 Περιγραφή δείγματος δεδομένων

Σε συνέχεια της προηγούμενης ενότητας, η μορφή των δεδομένων μας ήταν εξής:

1) Το σύνολο των δεδομένων των δεικτών περιείχε:
    - Ημερομηνία γεγονότος
    - Τιμή κλεισίματος
    - Τιμή ανοίγματος
    - Υψηλότερη τιμή
    - Χαμηλότερη τιμή
    - Όγκος συναλλαγών
    - Ποσοστιαία μεταβολή
2) Το σύνολο των δεδομένων για τις ειδήσεις περιείχε:
    - Ημερομηνία γεγονότος
    - Σύνολο ειδήσεων (κειμένων)
    - Ετικέτα με την ένδειξη 1 ή 0

6.3.3 Διεξαγωγή πειραμάτων

Τα πειράματα που πραγματοποιήθηκαν προσπαθούν να απαντήσουν τα παρακάτω ερευνητικά ερωτήματα:

1) Η πολικότητα των κειμένων μπορεί να συνεισφέρει στην εκπαίδευση ενός μοντέλου με μεγαλύτερη ακρίβεια;
2) Ποιοι και πόσοι δείκτες συμβάλλουν στην πρόβλεψη της ανόδου ή πτώσης του χρηματιστηρίου;

6.4 Αποτελέσματα

Αρχικά μελετάμε την επίδραση της πολικότητας των κειμένων δεδομένου ότι όλοι οι υπόλοιποι δείκτες λαμβάνονται υπόψιν κατά την διαδικασία εκπαίδευσης. Στον πίνακα 1 συνοψίζονται τα αποτελέσματα.

Πίνακας 1: Πολικότητα κειμένων στην πρόβλεψη (ακρίβεια)

Παράμετροι	Accuracy
Χρηματιστηριακοί Δείκτες	53,21%
Χρημ/κοί Δείκτες + Πολικότητα Κειμένων	51,50%

Όπως φαίνεται, η πολικότητα των κειμένων δεν έχει οδηγήσει σε καλύτερα αποτελέσματα. Αυτό μπορεί να οφείλεται στο γεγονός ότι τα κείμενα που λήφθηκαν υπόψιν δεν αντικατοπτρίζουν την πραγματική άποψη του κόσμου σχετικά με την πορεία του χρηματιστηρίου. Στο πίνακα 2 συνοψίζουμε τα αποτελέσματα με βάση την μετρική του ελάχιστου τετραγωνικού σφάλματος.

Πίνακας 2: Πολικότητα κειμένων στην πρόβλεψη (MSE & R^2)

Παράμετροι	MSE	R^2
Χρηματιστηριακοί Δείκτες	0,5582792828290312	-0,071199
Χρημ/κοί Δείκτες + Πολικότητα Κειμένων	0,8748465877068812	-0,0107646

Παρατηρούμε και στον πίνακα 2 ότι το ελάχιστο τετραγωνικό σφάλμα είναι μεγαλύτερο όταν εισάγουμε κείμενο.

Στην συνέχεια εξετάζουμε μεμονωμένα κάθε δείκτη και κατά πόσο αυτός συμβάλει στην πρόβλεψη της ανόδου ή πτώσης του χρηματιστηρίου. Σε αυτό το σημείο εφαρμόσαμε απλά γραμμικά μοντέλα παλινδρόμησης με ένα δείκτη ως ανεξάρτητη μεταβλητή.

Πίνακας 3: Εξέταση κάθε δείκτη στην εκπαίδευση του μοντέλου πρόβλεψης (Accuracy)

Παράμετροι	Accuracy
DAX	53,57%
NASDAQ	54,46%
DOW_JONES	54,46%
CAC 40	53,81%
S&P 500	54,46%
FTSE 100	53,76%

Στους πίνακες 3 και 4 συνοψίζονται τα αποτελέσματα. Παρατηρούμε ότι αν χρησιμοποιήσουμε μεμονωμένα του χρηματιστηριακούς δείκτες, δεν απέχουμε πολύ από το εάν χρησιμοποιήσουμε όλους τους δείκτες μαζί στην διαδικασία εκπαίδευσης. Επίσης, παρατηρούμε ότι ο NASDAQ, ο DOW JONES και ο S&P 500 φαίνεται να επιδρούν περισσότερο από τους άλλους δείκτες. Η ακρίβεια είναι η ίδια και αυτό μπορεί να οφείλεται στο γεγονός ότι αυτοί οι τρεις δείκτες είναι δείκτες του αμερικάνικου χρηματιστηρίου και πολλές εταιρείες είναι διπλοεισεγμένες. Ωστόσο, ο FTSE παρόλο που έχει μικρότερη ακρίβεια, το ελάχιστο τετραγωνικό σφάλμα είναι μικρότερο.

Πίνακας 4: Εξέταση κάθε δείκτη στην εκπαίδευση του μοντέλου πρόβλεψης (MSE & R^2)

Παράμετροι	MSE	R^2
DAX	0,5303705655103875	-0,01000
NASDAQ	0,5478937541475412	-0,05127
DOW JONES	0,5564200354507725	-0,06763
CAC 40	0,529509257859955	-0,0082

S&P 500	0,5416463501510586	-0,0392
FTSE 100	0,5288513532414462	-0.0071

Σχετικά με το R^2, και στις δύο κατηγορίες πειραμάτων, παρατηρούμε ότι οι τιμές που λαμβάνει είναι κάτω από το μηδέν, πράγμα που μπορεί να οφείλεται στο γεγονός ότι το μοντέλο δεν ακολουθεί την τάση των δεδομένων.

Τέλος χρησιμοποιώντας όλους τους δείκτες θα βρούμε την γραμμική συνάρτηση παλινδρόμησης καθώς και τους συντελεστές βαρύτητας για κάθε έναν δείκτη. Ο πίνακας συνοψίζει τους συντελεστές βαρύτητας.

Πίνακας 5: Συντελεστές βαρύτητας του μοντέλου γραμμικής παλινδρόμησης

Παράμετροι	Συντελεστής
DAX	-0,0001373
NASDAQ	-0,0002185
DOW JONES	0,0000005027
CAC 40	0,000004987
S&P 500	0,000005032
FTSE 100	-0,000001938

6.5 Σύνοψη συμπερασμάτων αξιολόγησης

Έπειτα από την διεξαγωγή των πειραμάτων συμπεραίνουμε ότι η χρήση κειμένου δεν συνείσφερε στην ακρίβεια του μοντέλου θετικά. Επίσης, διαπιστώνουμε ότι η χρήση όλων των δεικτών κατά την διαδικασία της εκπαίδευσης οδήγησε σε χειρότερα αποτελέσματα. Σε αυτό μπορεί να οφείλεται το γεγονός ότι ο παγκόσμιος δείκτης επηρεάζεται περισσότερο από μεγάλες εταιρείες και αφού πολλές από αυτές εδρεύουν στην Αμερική, συνεπάγεται ότι οι δείκτες της Αμερικής θα επηρεάζουν περισσότερο τον παγκόσμιο δείκτη. Τέλος, θα πρέπει να σημειωθεί ότι τα αποτελέσματα ταιριάζουν με αυτά της υπόθεσης για το χρηματιστήριο. Δηλαδή, οι χρηματιστηριακές τιμές ακολουθούν ένα τυχαίο μοτίβο και γι’ αυτό δεν μπορεί να είναι προβλέψιμη η κίνηση πάνω από 50% [13].

7

Τεχνικές λεπτομέρειες

Στην ενότητα αυτή θα αναφερθούν οι τεχνικές λεπτομέρειες του μοντέλου πρόβλεψης που υλοποιήσαμε καθώς και της γραφικής διεπαφής. Έπειτα θα αναφερθούμε σε εκείνες τις πλατφόρμες και εργαλεία που βοήθησαν στην επίτευξη του στόχου μας.

7.1 Λεπτομέρειες υλοποίησης

Το τεχνικό κομμάτι της εργασίας χωρίζεται σε δύο μέρη. Αφενός, ορίζεται το μοντέλο γραμμικής παλινδρόμησης για την επίλυση του προβλήματος και αφετέρου η γραφική διεπαφή χρήστη που χρησιμοποιεί το μοντέλο και προσφέρει μία φιλική διεπαφή για τον χρήστη.

7.1.1 Μοντέλο Γραμμικής Παλινδρόμησης

Τα στάδια που ακολουθήθηκαν για την υλοποίηση του μοντέλου γραμμικής παλινδρόμησης είναι η προ επεξεργασία των δεδομένων (χρηματιστηριακοί δείκτες και πολικότητα κειμένου). Έπειτα, έγινε εκπαίδευση του μοντέλου χρησιμοποιώντας το σύνολο δεδομένων προερχόμενο από την προηγούμενη φάση και τέλος, έγινε η αξιολόγηση του μοντέλου με βάση τις μετρικές που αναλύθηκαν στο κεφάλαιο 6.

7.1.1.1 Προ επεξεργασία

Οι τιμές κλεισίματος των χρηματιστηριακών δεικτών που χρησιμοποιήθηκαν από την εργασία προήλθαν από το investing.com όπως αναφέρθηκε πρωτύτερα ενώ η πολικότητα των κειμένων από ένα σύνολο δεδομένων που παρέχεται από το Kaggle. Το πρόβλημα που αντιμετωπίσαμε κατά την διαδικασία δημιουργίας του συνόλου ήταν ότι οι ημερομηνίες κάθε παραδείγματος ενός συνόλου δεδομένων δεν ταίριαζε ακριβώς με τις ημερομηνίες ενός άλλου συνόλου δεδομένων. Για παράδειγμα, υπήρχαν κάποιες τιμές κλεισίματος από κάποιους χρηματιστηριακούς δείκτες για συγκεκριμένες χρονικές περιόδους, όπου οι συγκεκριμένες χρονικές περίοδοι δεν υπήρχαν στο σύνολο δεδομένων των καθημερινών νέων. Έτσι, έπρεπε με κάποιον τρόπο να κρατήσουμε εκείνες τις ημερομηνίες που ήταν κοινές σε όλα τα σύνολα δεδομένων. Έτσι εφαρμόσαμε τα παρακάτω βήματα.

1. Αρχικά αποφασίστηκε να χρησιμοποιηθεί η δομή του λεξικού για την μοντελοποίηση των δεδομένων. Πρακτικά αυτό σημαίνει ότι κάθε εγγραφή που αποθηκεύεται σε αυτήν την δομή αποτελείται από ζεύγη κλειδιών και τιμών. Με αυτόν τον τρόπο μπορεί να γίνει εύκολα αναζήτηση με βάση το κλειδί το οποίο δεν είναι αναγκάστηκα αριθμητική τιμή (π.χ. αν χρησιμοποιούσαμε πίνακες δεν θα μπορούσε να εφαρμοστεί αυτό). Στην περίπτωσή μας, τα κλειδιά είναι οι ημερομηνίες ενώ ως τιμές έχουμε πίνακες. Κάθε πίνακας περιέχει τις τιμές κλεισίματος κάθε δείκτη και την πολικότητα κειμένου για εκείνη την χρονική στιγμή του αντίστοιχου κλειδιού. Για παράδειγμα μία εγγραφή στο λεξικό θα μπορούσε να είναι «23/10/2019» : [4534,43 , 3425,2 … , 0]. Η εγγραφή αυτή αποθηκεύει την εξής σχέση: στις 23/10/2019 η τιμή κλεισίματος του πρώτου δείκτη είναι 4534,43 , η τιμή του δεύτερου δείκτη 3425,2 ενώ η πολικότητα των κειμένων ήταν αρνητική (εξού και το 0 στον παραπάνω πίνακα).
2. Έχοντας την παραπάνω δομή γνωρίζουμε τις τιμές των χαρακτηριστικών του μοντέλου που θα υλοποιηθεί για κάθε χρονική στιγμή. Έτσι, η διάσταση του πίνακα μιας εγγραφής είναι ίση με τον αριθμό των χαρακτηριστικών. Σε περίπτωση επομένως, που η διάσταση ενός πίνακας, μιας εγγραφής, δεν είναι ίση με τον αριθμό των χαρακτηριστικών, αυτό υποδηλώνει ότι για ένα ή παραπάνω χαρακτηριστικό δεν υπάρχει τιμή. Δηλαδή, για ένα χαρακτηριστικό δεν έχουμε την τιμή του εκείνη την χρονική στιγμή. Έτσι, μπορούμε να απαλείψουμε τις εγγραφές

όπου ο αντίστοιχος πίνακάς τους δεν είναι ίσης διάστασης με τον αριθμό των χαρακτηριστικών πετυχαίνοντας έτσι, την δημιουργία ενός κοινού συνόλου δεδομένων για τις συγκεκριμένες ημερομηνίες.

Με τα παραπάνω δύο βήματα εγγυηθήκαμε την δημιουργία ενός συνόλου δεδομένων όπου για την χρονική στιγμή t έχουμε τις τιμές κλεισίματος των χρηματιστηριακών δεικτών και την πολικότητα κειμένου. Ωστόσο, στους χρηματιστηριακούς δείκτες έχουμε και τον παγκόσμιου δείκτη την χρονική στιγμή t. Όμως, με βάση την μεθοδολογία που θέλαμε να ακολουθήσουμε θα έπρεπε την χρονική στιγμή t να έχουμε τις τιμές των χαρακτηριστικών την χρονική στιγμή t -1, ενώ για μεταβλητή στόχο θέλαμε την ποσοστιαία μεταβολή του κλεισίματος την χρονική στιγμή t -1 με την χρονική στιγμή t. Έτσι, διατρέξαμε την δομή του λεξικού ταξινομημένη σε αύξουσα διάταξη παίρνοντας 2 εγγραφές την φορά. Σε μία νέα δομή λεξικού αποθηκεύαμε τις τιμές χαρακτηριστικών από την πρώτη εγγραφή ενώ υπολογίζαμε την ποσοστιαία μεταβολή μεταξύ των δύο εγγραφών. Με αυτόν τον τρόπο η νέα δομή λεξικού περιέχει τις ημερομηνίες πλην της τελευταίας από την προηγούμενη δομή λεξικού πετυχαίνοντας τον στόχο μας.

Εν συνεχεία, χωρίσαμε την δομή λεξικού σε δύο επι μέρους δομές ώστε να έχουμε ένα σύνολο εκπαίδευσης και ένα σύνολο δοκιμής. Από την προηγούμενη φάση, αποκτήσαμε ένα σύνολο δεδομένων σε μορφή λεξικού, όπου ως κλειδιά έχει τις ημερομηνίες του συνόλου δεδομένων και ως τιμές τους πίνακες των χαρακτηριστικών. Καθώς, για την εκπαίδευση του μοντέλου αλλά και για την αξιολόγηση δεν χρειαζόμαστε τις ημερομηνίες αλλά μόνο τις τιμές των χαρακτηριστικών, διατρέξαμε τη δομή του λεξικού και αποθηκεύσαμε την τιμή του λεξικού σε έναν νέο πίνακα. Με αυτόν τον τρόπο δημιουργήσαμε έναν 2-διαστάσεων πίνακα, όπου το σύνολο των σειρών είναι ο αριθμός των παραδειγμάτων που θα χρησιμοποιήσουμε στην εκπαίδευση ενώ το σύνολο των στηλών είναι τα χαρακτηριστικά. Στο σύνολο των στηλών όπως αναφέραμε υπάρχει και η ποσοστιαία μεταβολή του παγκόσμιου δείκτη που όμως δεν ανήκει στο σύνολο των χαρακτηριστικών αλλά είναι η μεταβλητή στόχος για το μοντέλο παλινδρόμησης. Έτσι, κάθε τιμή του παγκόσμιου δείκτη προστέθηκε σε νέο πίνακα. Ο νέος πίνακας είναι μίας διάστασης και κάθε κελί αντιστοιχεί σε ένα παράδειγμα του συνόλου δεδομένων. Συνοψίζοντας, έχουμε:

$$X = \{x_{i1}, x_{i2}, \dots, x_{in}\}_1^m$$

όπου n ο αριθμός των χαρακτηριστικών και m ο αριθμός των παραδειγμάτων.

Ο πίνακας για την μεταβλητή στόχο ορίζεται ως:

$$Y = \{y_i\}_1^m$$ όπου m ο αριθμός των παραδειγμάτων.

Επαναφέροντας, την μεθοδολογία που αναφέραμε στο κεφάλαιο 5, παρατηρούμε ότι το σύνολο δεδομένων έχει την ίδια μορφή, απλώς έχουμε χωρίσει το σύνολο σε δύο διαφορετικούς πίνακες

7.1.1.2 Εκπαίδευση μοντέλου γραμμικής παλινδρόμησης.

Έχοντας τους πίνακες από την προηγούμενη φάση, αναλυτικά:

$$X_{train} = \{x_{i1}, x_{i2}, \dots, x_{in}\}_1^m$$
$$Y_{train} = \{y_i\}_1^m$$
$$X_{test} = \{x_{i1}, x_{i2}, \dots, x_{in}\}_1^r$$
$$Y_{test} = \{y_i\}_1^r$$

Όπου r το σύνολο παραδειγμάτων δοκιμών, χρησιμοποιήσαμε το αντικείμενο της κλάσης LinearRegression() της βιβλιοθήκης scikit-learn που εισάγεται παρακάτω σε αυτό το κεφάλαιο.

Συγκεκριμένα, χρησιμοποιήθηκε η συνάρτηση fit ως εξής:

$$fit(X_{train}, Y_{train})$$

Η συνάρτηση αυτή χρησιμοποιεί τα δεδομένα εισόδου και εξόδου (δηλαδή τους δύο πίνακες) για την εκπαίδευση του μοντέλου γραμμικής παλινδρόμησης.

Εφόσον το μοντέλο έχει εκπαιδευτεί, μπορούμε να χρησιμοποιήσουμε την συνάρτηση predict ως εξής:

$$predict(X_{test})$$

Αυτή η συνάρτηση επιστρέφει όλα τα $Y_{predictions}$ για το σύνολο δοκιμών που έχουμε περάσει ως είσοδο. Επομένως, με αυτές τις δύο συναρτήσεις πετυχαίνουμε την εκπαίδευση του μοντέλου και μπορούμε να πάρουμε προβλέψεις από αυτό δοθέντος ενός συνόλου δοκιμής.

7.1.1.3 Αξιολόγηση μοντέλου γραμμικής παλινδρόμησης.

Για τον υπολογισμό της ακρίβειας έπρεπε να μεταβούμε σε μία «μετά» επεξεργασία των προβλέψεων καθώς η ακρίβεια απαιτεί ακριβή ταίριασμα των τιμών πρόβλεψης με αυτών του συνόλου δοκιμών. Αναλυτικά δημιουργήσαμε έναν νέο πίνακα $Y_{prediction2}$ όπου κάθε κελί έχει τιμή 1 σε περίπτωση που η πρόβλεψη είναι μεγαλύτερη του μηδενός, ενώ έχει τιμή 0 στην αντίθετη περίπτωση. Το ίδιο εφαρμόσαμε στο σύνολο δοκιμών και δημιουργήσαμε τον πίνακα Y_{test2}. Επομένως, ο υπολογισμός της ακρίβειας γίνεται ως:

$$\text{accuracy\_score}(Y_{predictions2}, Y_{test2})$$

Από την άλλη, από την προηγούμενη φάση πήραμε όλες τις προβλέψεις του συνόλου δοκιμών X_{test} στον πίνακα $Y_{predictions}$. Για την αξιολόγηση του μοντέλου απλώς χρησιμοποιήσαμε τις μετρικές του πακέτου metrics από το scikit-learn. Αναλυτικά, για την μέθοδο ελάχιστων τετραγώνων και για την μέθοδο R^2 καλέσαμε αντίστοιχα:

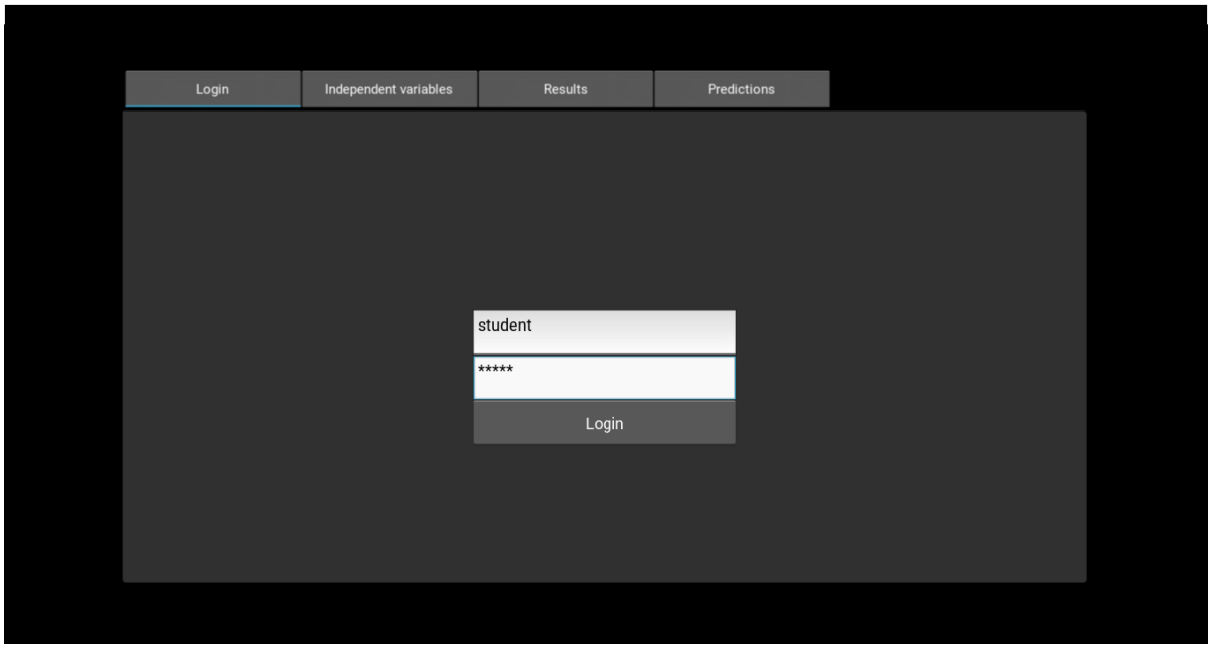
$$\text{mean\_squared\_error}(Y_{predictions}, Y_{test})$$

και

$$\text{r2\_score}(Y_{predictions}, Y_{test})$$

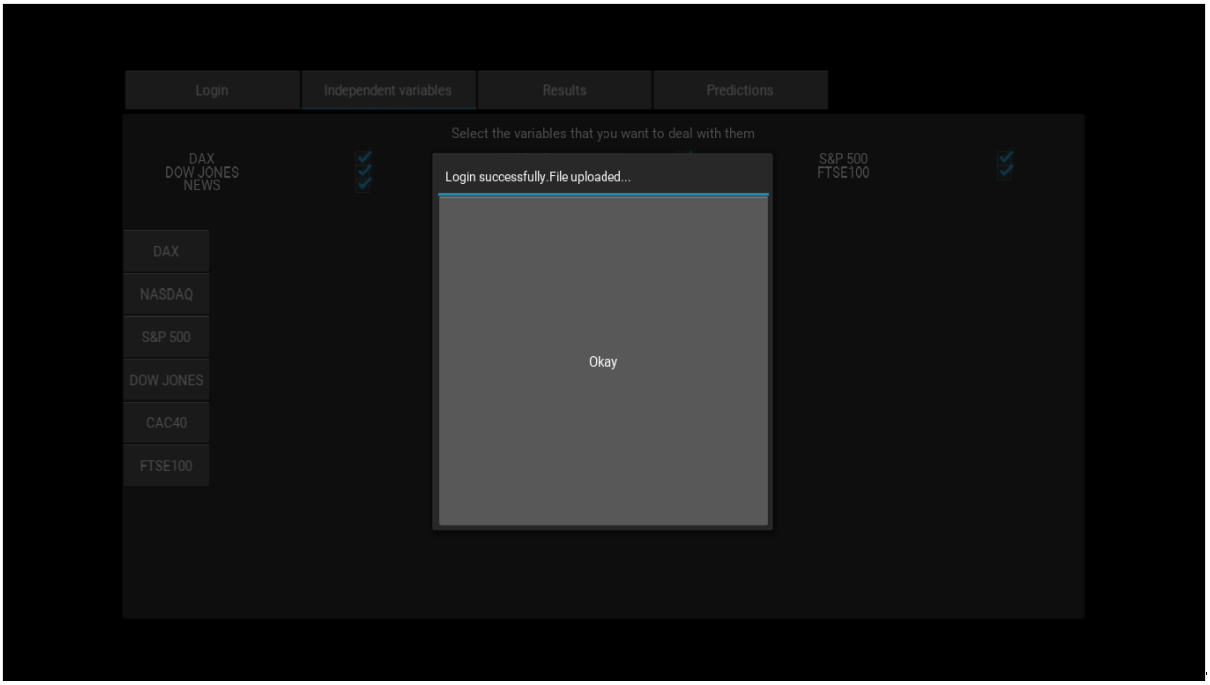
7.1.2 Γραφική Διεπαφή χρήστη

Η γραφική διεπαφή υλοποιήθηκε με τη χρήση του kivy που αναλύεται παρακάτω. Συγκεκριμένα, η διεπαφή είναι ένα πρόγραμμα υλοποιημένο με χρήση της γλώσσας προγραμματισμού Python και περιέχει τέσσερις καρτέλες. Οι καρτέλες έχουν δημιουργηθεί μέσω του αντικειμένου της κλάσης TabbedPanel ενώ κάθε καρτέλα είναι στιγμιότυπο του αντικειμένου TabbedPanelItem Στην πρώτη καρτέλα, ο χρήστης κάνει είσοδο με τα στοιχεία που επαληθεύουν την ύπαρξή του μέσα στο σύστημα. Στην εικόνα 6 παρουσιάζεται ένα παράδειγμα. Ο χρήστης αφού συμπληρώσει το «Username» και το «Password» κάνει είσοδο στο σύστημα.



Εικόνα 6: Σύνδεση στο Σύστημα

Στην συνέχεια το σύστημα ενημερώνει τον χρήστη για την επιτυχή είσοδο του καθώς παράλληλα φορτώνεται το σύνολο δεδομένων που υπάρχει στο φάκελο όπου ανήκει το εκτελέσιμο, όπως φαίνεται στην εικόνα 7.



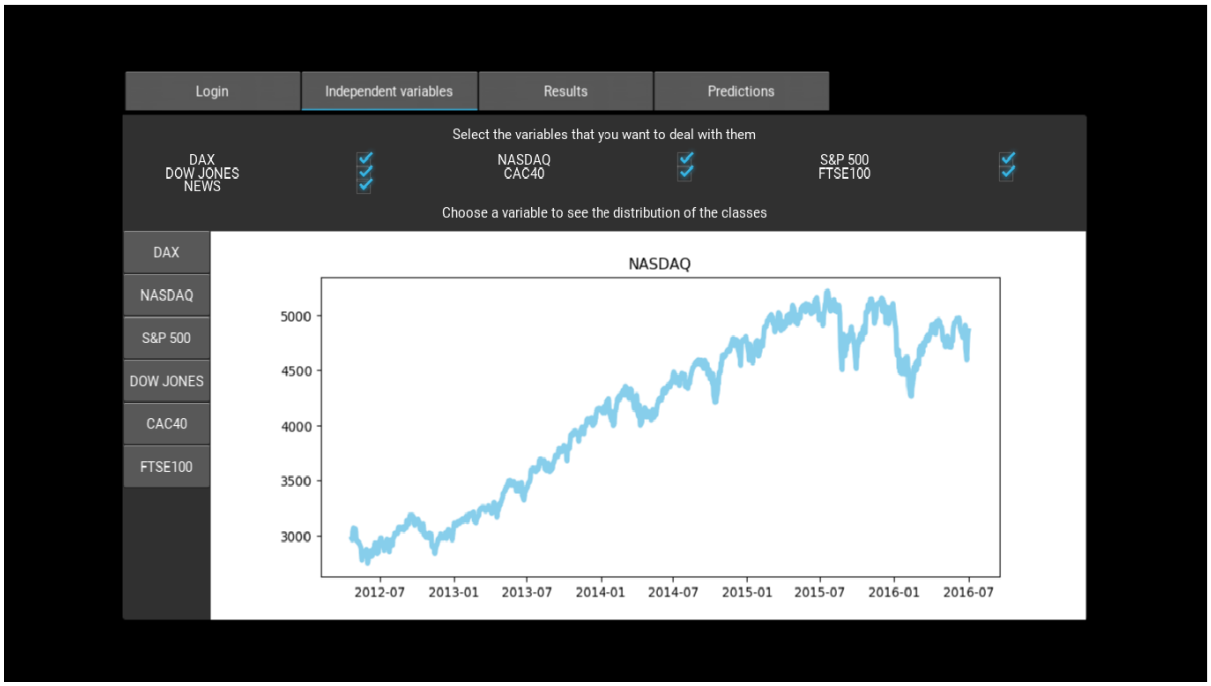
Εικόνα 7:Ενημέρωση επιτυχής σύνδεσης και φόρτωσης δεδομένων στο σύστημα

Εφόσον φορτωθεί το σύνολο δεδομένων στην μνήμη μέσω της επιλογής του χρήστη, μεταβαίνουμε στην δεύτερη καρτέλα της εφαρμογής που φαίνεται στην εικόνα 8.

Η δεύτερη καρτέλα χωρίζεται σε δύο τμήματα. Στο πρώτο τμήμα ο χρήστης επιλέγει ποιες ανεξάρτητες μεταβλητές θα χρησιμοποιήσει για την διαδικασία μάθησης και αξιολόγησης. Το συστατικό που χρησιμοποιήθηκε για την επιλογή είναι το CheckBox ενώ για να πετύχουμε την σωστή παρουσίαση των CheckBox στην καρτέλα, επιλέξαμε το GridLayout. Λειτουργικά στο σημείο αυτό, η εφαρμογή άνοιξε το αρχείο εισόδου που περιέχει το σύνολο δεδομένων και φόρτωσε τα χαρακτηριστικά που παρουσιάζονται στην πρώτη γραμμή του αρχείου. Τα χαρακτηριστικά αυτά, παρουσιάσθηκαν ως CheckBoxes και παρουσιάσθηκαν στο πάνω μέρος της περιοχής της καρτέλας.

Εικόνα 8: Επιλογή Ανεξάρτητων Μεταβλητών

Στο κάτω μέρος, ο χρήστης μπορεί να επιλέξει έναν από τους δείκτες/χαρακτηριστικά και να παρουσιάσει τα αποτελέσματα στο κέντρο της εφαρμογής όπως φαίνεται στην εικόνα 9.



Εικόνα 9: Παρουσίαση Κίνησης Κάποιας Ανεξάρτητης Μεταβλητής με Βάση το Χρόνο

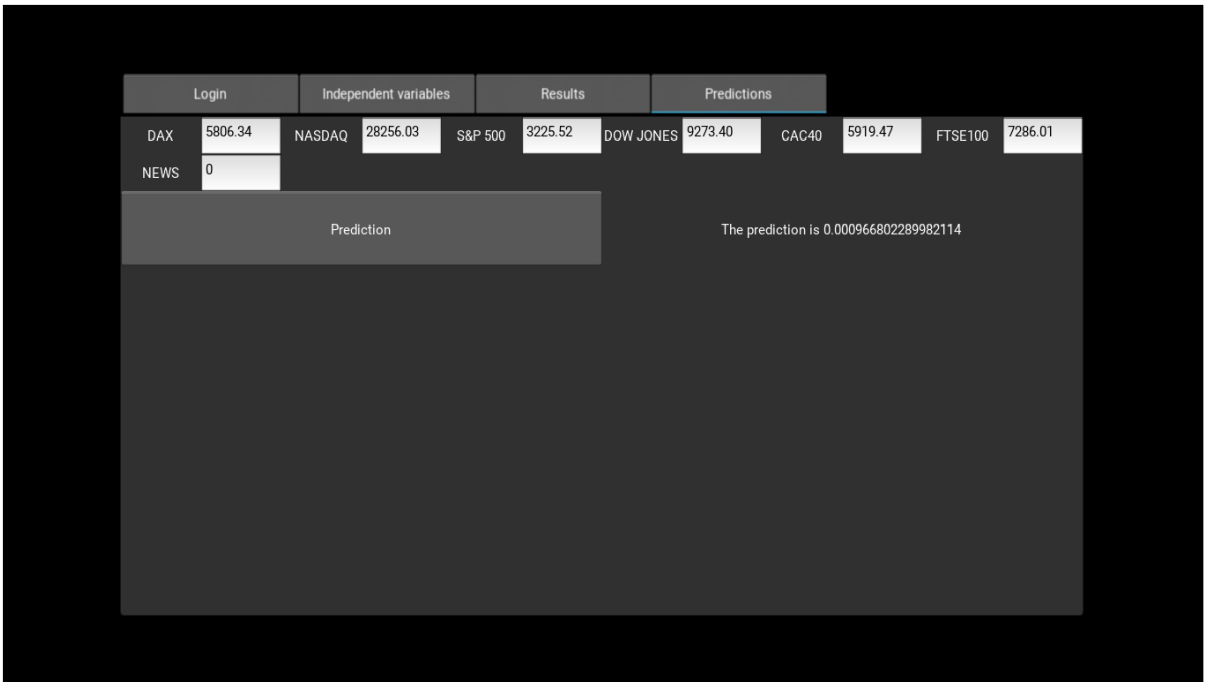
Το παράδειγμα της εικόνας 8 μας δείχνει ότι έχει επιλεγεί ό δείκτης NASDAQ για παρουσίαση των δεδομένων του. Ο άξονας των Χ παριστάνει τις ημερομηνίες, ενώ στον άξονα Υ βλέπουμε τις τιμές κλεισίματος του δείκτη ανάλογα με κάποια ημερομηνία. Για την υλοποίηση του μενού επιλογών αριστερά χρησιμοποιήθηκε αντικείμενο της κλάσης DropDown ενώ για την παρουσίαση του γραφήματος, αρχικά δημιουργήθηκε το γράφημα με την βοήθεια της βιβλιοθήκης matplotlib και έπειτα εισάχθηκε σε αντικείμενο της κλάσης FigureCanvasKivyAgg που διαχειρίζεται τα γραφήματα και τις εικόνες της βιβλιοθήκης του Kivy.

Τέλος, αφού έχουν επιλεχθεί τα χαρακτηριστικά από την δεύτερη καρτέλα μεταβαίνουμε στην τρίτη καρτέλα, το περιεχόμενο της οποίας είναι υπεύθυνο για την δημιουργία του μοντέλου πρόβλεψης. Στην εικόνα 10 παρουσιάζεται το αποτέλεσμα εκτέλεσης του αλγορίθμου γραμμικής παλινδρόμησης.

Εικόνα 10: Εκπαίδευση και Αξιολόγηση του Μοντέλου Γραμμικής Παλινδρόμησης

Το περιεχόμενο της καρτέλας αυτής χωρίζεται εξίσου σε δύο μέρη. Στο πάνω μέρος, ο χρήστης πρέπει να πατήσει το κουμπί για την εκτέλεση του αλγορίθμου γραμμικής παλινδρόμησης. Εφόσον συμβεί αυτό, παρακάτω φαίνονται τα αποτελέσματα κατά την αξιολόγηση του αλγορίθμου. Συγκεκριμένα, από την εκπαίδευση μαθαίνουμε το συνολικό αριθμό των συντελεστών, ενώ κατά την αξιολόγηση παρατηρούμε ότι υλοποιήθηκαν οι δύο συναρτήσεις αξιολόγησης που προτάθηκαν πρωτύτερα στο κεφάλαιο 6 (MSE, Accuracy). Τέλος παρουσιάζεται ένα γράφημα αποτελούμενο από δύο γραμμές. Η πρώτη (με το μπλε χρώμα) αντιστοιχεί στην πραγματική τιμή της μεταβλητή στόχου ενώ η δεύτερη (με το λαδί χρώμα) αντιστοιχεί στην πρόβλεψη. Στο λειτουργικό τμήμα της εφαρμογής, εφόσον ο χρήστης έχει φορτώσει το σύνολο δεδομένων επιλέγεται το 80% αυτού του συνόλου ως σύνολο εκπαίδευσης, ενώ το 20% ως σύνολο δοκιμής. Θα πρέπει να αναφερθεί ότι εφόσον αναφερόμαστε σε χρονοσειρές το 20% βρίσκεται μετά από το 80% χρονολογικά. Επομένως, προβλέπουμε το 20% των τιμών που βρίσκονται χρονολογικά ύστερα από το 80% των δεδομένων.

Τέλος, στην εικόνα 11 βλέπουμε την τέταρτη καρτέλα που σχετίζεται με την πρόβλεψη τις ανόδου/πτώσης χρηματιστηρίου, όπου ο χρήστης μπορεί να εισάγει τις τιμές των χαρακτηριστικών για την πρόβλεψη της αυριανής μέρας.



Εικόνα 11: Πρόβλεψη Ανόδου ή Πτώσης για την Αυριανή Μέρα

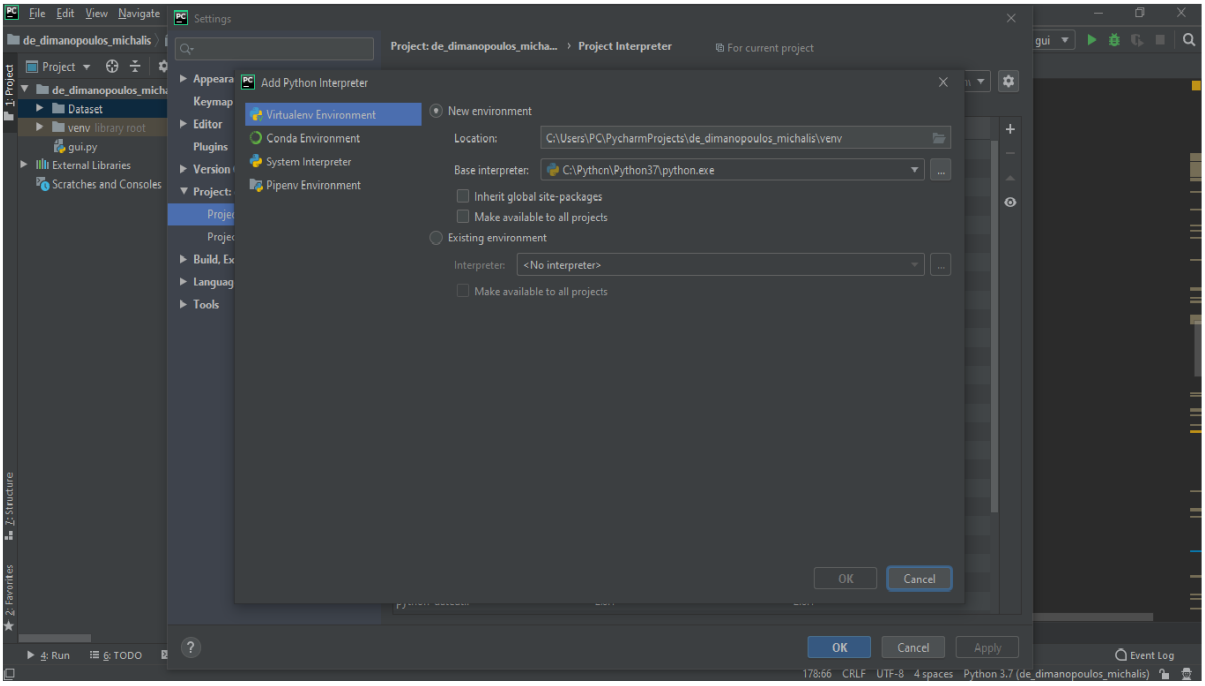
Όπως φαίνεται, υπάρχουν TextInputs() στα οποία μπορεί ο χρήστης να εισάγει τις τιμές των χαρακτηριστικών για την σημερινή μέρα. Εφόσον, συμπληρωθούν τα πεδία, πατώντας το κουμπί Prediction, χρησιμοποιείται το μοντέλο που εκπαιδεύτηκε στην προηγούμενη καρτέλα για την πρόβλεψη της σημερινής. Εφόσον, η τιμή είναι αρνητική, συμπεραίνουμε ότι θα έχουμε πτώση του παγκόσμιου χρηματιστηριακού δείκτη.

7.2 Πλατφόρμες και προγραμματιστικά εργαλεία

Για την συγγραφή του πηγαίου κώδικα χρησιμοποιήθηκε η πλατφόρμα PyCharm 2019 ενεργοποιώντας το ακαδημαϊκό πακέτο που προσφέρει η JetBrains και η γλώσσα προγραμματισμού Python. Για την εγκατάσταση της πλατφόρμας ο αναγνώστης μπορεί να επισκεφτεί την σελίδα https://www.jetbrains.com/pycharm/ . Από προεπιλογής, το PyCharm χρησιμοποιεί τον διερμηνευτή της Python που είναι εγκατεστημένος στον ηλεκτρονικό υπολογιστή. Η έκδοση που επιλέξαμε είναι η 3.7, ωστόσο, ο κώδικάς μας είναι συμβατός για όλες τις εκδόσεις Python > 3.0.

Κατά την δημιουργία του project της εργασίας στο Pycharm δημιουργήσαμε μία εικονική μηχανή για τον διερμηνευτή της Python. Με αυτόν τον τρόπο, τα πακέτα που έγιναν εγκατάσταση δεν εγκαταστάθηκαν στον διερμηνευτή του συστήματος αλλά σε ένα

αποθετήριο που σχετίζεται μόνο με το project της εργασίας. Με αυτόν τον τρόπο, τα πακέτα που έγιναν εγκατάσταση δεν εγκαταστάθηκαν στον διερμηνευτή του συστήματος αλλά σε ένα αποθετήριο που σχετίζεται μόνο με το project της εργασίας, όπως φαίνεται στην εικόνα 12.



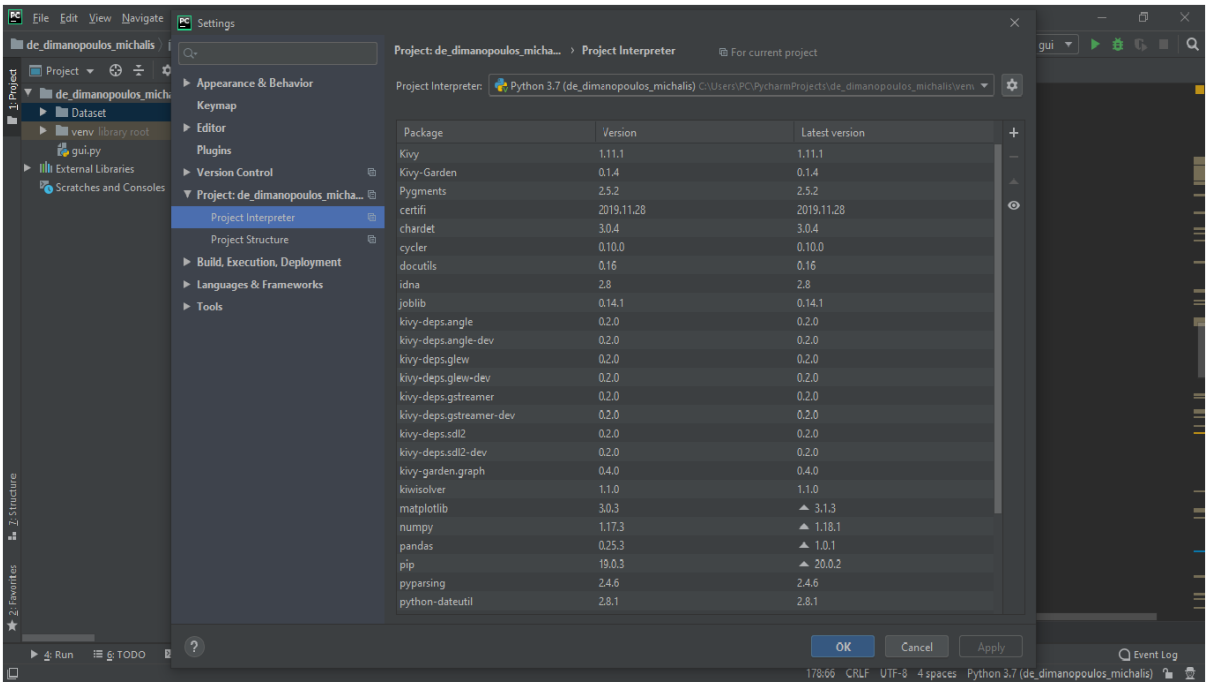
Εικόνα 12: Δημιουργία Διερμηνευτή

7.2.1 Πακέτα εγκατάστασης στην Python

Τα πακέτα που χρησιμοποιήθηκαν για την συγκεκριμένη εργασία συνοψίζονται παρακάτω:

- Matplotlib 3.1.2
- Numpy 1.17.3
- Pandas 0.25.3
- Scikit-learn 0.21.3
- Kivy 1.11.1
- Kivy-garden 0.1.4
- Kivy-garden.graph 0.4.0

Για την εγκατάσταση των πακέτων μπορούμε να χρησιμοποιήσουμε είτε το πακέτο pip της Python. Για παράδειγμα, pip install numpy. Ωστόσο, από το γραφικό περιβάλλον του Pycharm, ένα πακέτο μπορεί να εγκατασταθεί από τις ρυθμίσεις, όπως φαίνεται στην εικόνα 13.



Εικόνα 13: Εγκατάσταση Πακέτων

Για να παρασταθούν τα γραφήματα στην γραφική διεπαφή θα πρέπει μέσω του πακέτου garden να γίνει εγκατάσταση του matplotlib (garden install matplotlib).

7.2.2 Μοντέλο πρόβλεψης με χρήση του scikit-learn

Για την δημιουργία του μοντέλου γραμμικής παλινδρόμησης χρησιμοποιήθηκε η βιβλιοθήκη της python, scikit-learn [23].

Η βιβλιοθήκη αυτή αποτελεί ένα απλό και αποδοτικό εργαλείο για μοντέλα πρόβλεψης τόσο αυτά που σχετίζονται με ταξινόμηση όσο και εκείνα με παλινδρόμηση. Είναι ανοιχτού κώδικα κάτω από την άδεια BSD και έχει υλοποιηθεί με χρήση των βιβλιοθηκών numpy, scipy και matplotlib [23].

Για την εκπαίδευση, αξιολόγηση του μοντέλου χρησιμοποιήθηκε η κλάση LinearRegression() η οποία δημιουργεί ένα αντικείμενο μοντέλου. Το μοντέλο εκπαιδεύεται μέσω της συνάρτησης fit(X,Y) όπου X είναι το σύνολο των τιμών των χαρακτηριστικών του μοντέλου πρόβλεψης ενώ το Y είναι το σύνολο των τιμών πρόβλεψης.

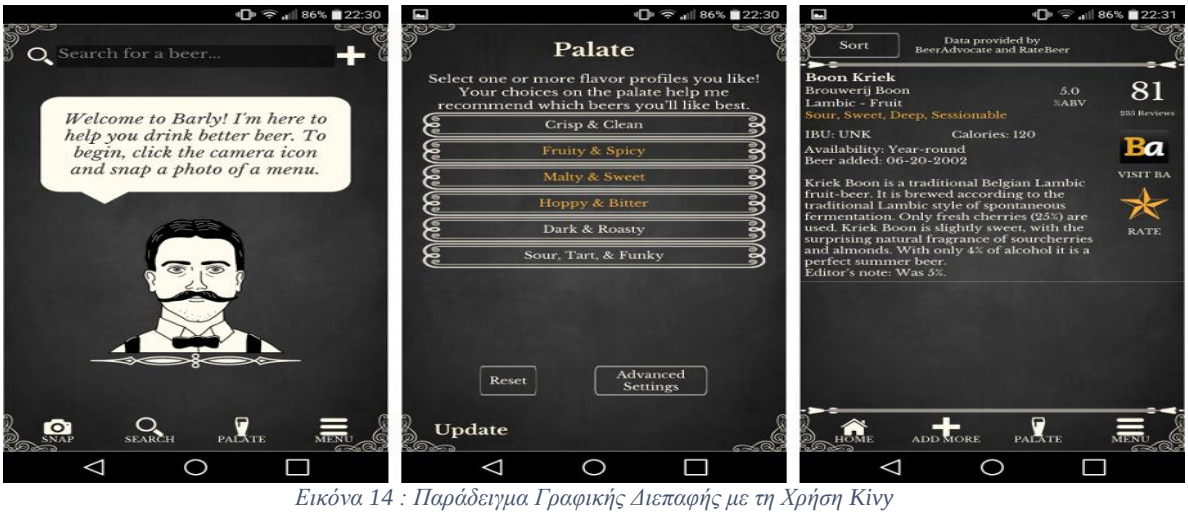
7.2.3 Γραφική διεπαφή με την χρήση του Kivy

Η βιβλιοθήκη της Python Kivy είναι μία ανοιχτού κώδικα βιβλιοθήκη για την γρήγορη ανάπτυξη εφαρμογών χρησιμοποιώντας καινοτόμες γραφικές διεπαφές όπως είναι οι εφαρμογές multi-touch. Το θετικό αυτής της βιβλιοθήκης είναι ότι οι εφαρμογές είναι συμβατές με διάφορα λειτουργικά συστήματα όπως Linux, Winodws, OS X, Android, IOS και Raspberry Pi. Επίσης, υποστηρίζει τις περισσότερες συσκευές εισόδου και πρωτοκόλλα συμπεριλαμβανομένων των WM_Touch, WM_Pen, Mac OS X Trackpad και Magic Mouse, Mtdev, Linux Kernel HID, TUIO.

Το Kivy είναι ανοιχτού κώδικα κάτω από την άδεια του MIT και LGPL 3. Επίσης, μπορεί να χρησιμοποιηθεί για εμπορικά προϊόντα. Η βιβλιοθήκη είναι σταθερή και περιέχει πολλές πηγές πληροφόρησης και εκμάθησής της.

Τέλος, ο μηχανισμός γραφικών χτίστηκε πάνω στο OpenGL ES2 και περιέχει πάνω από 20 μικρό-εφαρμογές. Τα περισσότερα κομμάτια του κώδικα είναι γραμμένα στην C χρησιμοποιώντας Cython.

Για την ανάπτυξη κώδικα με χρήση της συγκεκριμένης βιβλιοθήκης αφενός μπορείς είτε να χρησιμοποιήσει την python ως γλώσσα αντικειμενοστραφούς προγραμματισμού δεδομένου ότι όλα τα συστατικά μιας εφαρμογής είναι αντικείμενα κλάσεων ή χρησιμοποιώντας την γλώσσα kv που είναι αποκλειστικά υλοποιημένη από τους δημιουργούς και μπορεί να χρησιμοποιηθεί σε συνδυασμό με την Python για τον διαχωρισμό του γραφικού τμήματος της διεπαφής από το λειτουργικό. Στην εικόνα 12 φαίνεται ένα παράδειγμα χρήσης του Kivy για την υλοποίηση εφαρμογής android.



Εικόνα 14 : Παράδειγμα Γραφικής Διεπαφής με τη Χρήση Kivy

8

Επίλογος

8.1 Σύνοψη και συμπεράσματα

Ανακεφαλαιώνοντας, η παρούσα εργασία τέθηκε με σκοπό την επίλυση του προβλήματος πρόβλεψης της ανόδου ή πτώσης του παγκόσμιου χρηματιστηριακού δείκτη MSCI χρησιμοποιώντας τους πιο σημαντικούς χρηματιστηριακούς δείκτες ανά τον κόσμο. Το μοντέλο που υλοποιήθηκε και αξιολογήθηκε, είναι ένα μοντέλο επιβλεπόμενης μηχανικής μάθησης, γραμμική παλινδρόμηση. Οι δυνατότητες που έχει το μοντέλο είναι να εκπαιδεύεται με δεδομένα χρηματιστηρίου μία μέρα πριν την πρόβλεψη που θέλουμε να πετύχουμε. Δηλαδή, έχοντας τιμές για τις ανεξάρτητες μεταβλητές (χρηματιστηριακοί δείκτες αλλά και την πολικότητα των κειμένων) για κάποια χρονική στιγμή t-1, θέλουμε να προβλέψουμε την άνοδο ή πτώση του παγκόσμιου χρηματιστηριακού δείκτη (εξαρτημένη μεταβλητή) την χρονική στιγμή t. Πέρα από την ανάπτυξη του μοντέλου μηχανικής μάθησης, αναπτύχθηκε στα πλαίσια της εργασίας και μία γραφική διεπαφή, φιλική προς τον χρήστη, ώστε ο κάθε ενδιαφερόμενος να μπορεί να φορτώσει τα δικά του σύνολα δεδομένων (δηλαδή χρηματιστηριακούς δείκτες και κείμενα για συγκεκριμένες χρονικές στιγμές) και να προβλέψει την κίνηση του χρηματιστηρίου. Έτσι, θα έχει την δυνατότητα κάποιος να βγάλει χρήσιμα συμπεράσματα για την κίνηση του χρηματιστηρίου καθώς παράγονται και γραφήματα για ευκολότερη «ανάγνωση» των αποτελεσμάτων.

Τα αποτελέσματα έδειξαν ότι η χρήση της πολικότητας των κειμένων όχι μόνο δεν ωφέλησε το μοντέλο πρόβλεψης από την σκοπιά της ακρίβειας αλλά έριξε την ακρίβεια του μοντέλου για 2%. Αυτό όπως συζητήθηκε και παραπάνω μπορεί να οφείλεται στο γεγονός ότι τα

κείμενα δεν είναι αντιπροσωπευτικά για την εξέταση του παγκόσμιου δείκτη χρηματιστηρίου. Στη συνεχεία, παρατηρήσαμε ότι χρησιμοποιώντας μόνο έναν δείκτη πετυχαίναμε καλύτερα αποτελέσματα αντί να χρησιμοποιήσουμε όλους τους δείκτες μαζί. Κυρίως, αυτό παρατηρήθηκε στους Αμερικάνικους δείκτες. Αυτό μπορεί να οφείλεται στο γεγονός ότι κάποιες εταιρείες είναι διπλοεισαγμένες. Δηλαδή είναι εισαγμένες και στον παγκόσμιο χρηματιστηριακό δείκτη αλλά και τους υπόλοιπους που μελετάμε.

8.2 Μελλοντικές επεκτάσεις

Οι μελλοντικές επεκτάσεις γι' αυτήν την εργασία θα μπορούσαν να χωριστούν στις παρακάτω τρείς κατηγορίες:

1) Στο μεθοδολογικό κομμάτι της.
2) Σε σχέση με την γραφική διεπαφή.
3) Με την ενσωμάτωση του μοντέλου μηχανική μάθησης σε ένα ευρύτερο πληροφοριακό σύστημα.

8.2.1 Μεθοδολογικό κομμάτι

Αναλυτικά, θα μπορούσε κάποιος να πειραματιστεί και με άλλα μοντέλα μηχανικής μάθησης όπως παραδείγματος χάριν αυτά που ανήκουν στην οικογένεια των νευρωνικών δικτύων και όπως φαίνεται επιδρούν στα αποτελέσματα της πρόβλεψης. Επίσης, θα μπορούσαν να χρησιμοποιηθούν και άλλοι εξωτερικοί παράγοντες εκτός από τις ειδήσεις, όπως είναι τα κοινωνικά μέσα δικτύωσης. Να σημειωθεί ότι εστιάσαμε μόνο στην πολικότητα του κειμένου, ενώ αντιθέτως κάποιος θα μπορούσε να χρησιμοποιήσει αναπαράσταση κειμένου και να δοκιμάσει αυτά ως είσοδο στο μοντέλο πρόβλεψης.

8.2.2 Γραφική διεπαφή

Από την σκοπιά της γραφικής διεπαφής, θα μπορούσαν να προστεθούν περισσότεροι αλγόριθμοι μηχανικής μάθησης ώστε ο ενδιαφερόμενος να έχει την δυνατότητα να

πειραματιστεί με διάφορους εξ' αυτών και να διαλέξει τον καλύτερο με βάση τα δεδομένα που διαθέτει.

8.2.3 Ενσωμάτωση σε ένα ευρύτερο πληροφοριακό σύστημα

Τέλος, καθότι το μοντέλο μας χρησιμοποιεί τεχνικές μηχανικής μάθησης, δηλαδή την εφαρμογή της γραμμική παλινδρόμηση, θα μπορούσε να χρησιμοποιηθεί σε επιπλέων περιπτώσεις χρήσης που αφορούν συναφή ή ετερογενή συστήματα της σύγχρονης ψηφιακής εποχής.

Πιο συγκεκριμένα, εστιάζοντας στο πεδίο των χρηματοοικονομικών, το μοντέλο το οποίο χρησιμοποιήσαμε θα μπορούσε να εφαρμοστεί ως επέκταση ή/και βασικό δομικό μπλοκ σε ένα ευρύτερο πληροφοριακό σύστημα προβλέψεων σε τραπεζικούς και επενδυτικούς οργανισμούς.

9

Βιβλιογραφία

[1] Altay, E., & Satman, M. H. (2005). Stock market forecasting: artificial neural network and linear regression comparison in an emerging market. Journal of Financial Management & Analysis, 18(2), 18.

[2] Bekaert, G., & Harvey, C. R. (2017). Emerging equity markets in a globalizing world. Available at SSRN 2344817.

[3] O'Connor, N., & Madden, M. G. (2005, December). A neural network approach to predicting stock exchange movements using external factors. In International Conference on Innovative Techniques and Applications of Artificial Intelligence (pp. 64-77). Springer, London.

[4] Duong, Manh Ha, and Boriss Siliverstovs. "The stock market and investment." (2006).

[5] Heaney, R., & Hooper, V. (1999). World, regional and political risk influences upon Asia Pacific equity market returns. Australian Journal of Management, 24(2), 131-142.

[6] Kannan, K. S., Sekar, P. S., Sathik, M. M., & Arumugam, P. Financial stock market forecast using data mining techniques. In Proceedings of the International Multiconference of Engineers and computer scientists (2010, March): Vol. 1, p. 4.

[7] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.

[8] Liu, B., & Zhang, L. (2012). A survey of opinion mining and sentiment analysis. In Mining text data (pp. 415-463). Springer, Boston, MA.

[9] Mallet, V., Stoltz, G., & Mauricette, B. (2009). Ozone ensemble forecast with machine learning algorithms. Journal of Geophysical Research: Atmospheres, 114(D5).

[10] Michie, D., Spiegelhalter, D. J., & Taylor, C. C. (1994). Machine learning. Neural and Statistical Classification, 13.

[11] Paranjape-Voditel, P., & Deshpande, U. (2013). A stock market portfolio recommender system based on association rule mining. Applied Soft Computing, 13(2), 1055-1063.

[12] Pagolu, V. S., Reddy, K. N., Panda, G., & Majhi, B. (2016, October). Sentiment analysis of Twitter data for predicting stock market movements. In 2016 international conference on signal processing, communication, power and embedded system (SCOPES) (pp. 1345-1350). IEEE.

[13] Qian, B., & Rasheed, K. (2007). Stock market prediction with multiple classifiers. Applied Intelligence, 26(1), 25-33.

[14] Roman, J., & Jameel, A. (1996, January). Backpropagation and recurrent neural networks in financial analysis of multiple stock market returns. In Proceedings of HICSS-29: 29th Hawaii International Conference on System Sciences (Vol. 2, pp. 454-460). IEEE.

[15] Rigobon, R. and B. Sack (2003). Measuring the Response of Monetary Policy to the Stock Market, Quarterly Journal of Economics, 639-669.

[16] Shawe-Taylor, J., & Cristianini, N. (2000). Support vector machines. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods, 93-112.

[17] Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The AZFin text system. ACM Transactions on Information Systems (TOIS), 27(2), 12.

[18] Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. Chemometrics and intelligent laboratory systems, 39(1), 43-62.

[19] Tsai, C. F., & Wang, S. P. (2009, March). Stock price forecasting by hybrid machine learning techniques. In Proceedings of the International MultiConference of Engineers and Computer Scientists (Vol. 1, No. 755, p. 60).

[20] I . Vlahavas , P . Kefalas , N . Bassiliades , F . Kokkoras , I. Sakellariou. Artificial Intelligence - 3rd Edition, ISBN: 978-960-8396-64-7, Publisher: University of Macedonia Press / Greece, 2011.

[21] White, Eugene N. "The stock market boom and crash of 1929 revisited." Journal of Economic perspectives 4.2 (1990): 67-83.

[22] Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. arXiv preprint arXiv:1409.2329.

[23] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot and ´Edouard Duchesnay (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12 2825-2830