



Μεταπτυχιακή Διπλωματική Εργασία  
**ΠΡΟΒΛΕΨΗ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΩΝ  
ΧΡΟΝΟΛΟΓΙΚΩΝ ΣΕΙΡΩΝ**

---

**ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΠΕΙΡΑΙΩΣ**

---

Τμήμα: Ψηφιακών Συστημάτων

Πρόγραμμα Μεταπτυχιακών  
Σπουδών: Ψηφιακά Συστήματα &  
Υπηρεσίες

Κατεύθυνση: Δικτυοκεντρικά  
Πληροφοριακά Συστήματα

Όνοματεπώνυμο: Ιωαννίδης  
Ματθαίος

A.M: 1533

Επιβλέπων: Δουλκερίδης  
Χρήστος



## Ευχαριστίες

Στους γονείς μου.

## Περίληψη

Η πρόβλεψη, ανάλυση και μελέτη των χρηματοοικονομικών μεταβλητών κυρίως μετοχών καθώς και η ανάπτυξη οικονομικής στρατηγικής για το χρηματιστήριο αλλά και γενικότερα, ως αντικείμενα είναι ιδιαίτερα δημοφιλή θέματα, όχι απλώς στον ακαδημαϊκό τομέα αλλά και στον καθένα τον οποίο ενδεχόμενος να τον ενδιαφέρει το κέρδος μέσω αυτής της πρακτικής.

Ο λόγος που είναι τόσο δημοφιλής αυτό το αντικείμενο είναι κάτι παραπάνω από προφανές, όλοι θέλουν να κατασκευάσουν ένα μοντέλο το οποίο θα «νικάει» την αγορά, θα προβλέπει πότε θα αυξηθεί μία μετοχή ή το αντίθετο, θα συμβουλεύει τον χρήστη πότε να πουλήσει, πότε να αγοράσει όπως και επίσης μακροπρόθεσμα από ποια μετοχή θα έχει περισσότερα κέρδη, κοκ.

Η ανάλυση χρονολογικών σειρών έχει πλήθος τεχνικών που ως σκοπό έχουν την πρόβλεψη τους για την επόμενη χρονική στιγμή, είτε αυτή είναι μία μέρα είτε μία ώρα είτε μία εβδομάδα, κοκ. Όμως η ανάλυση χρονολογικών σειρών δεν μπορεί να λάβει υπόψιν της πλήθος άλλων παραγόντων οι οποίοι αποδεδειγμένα επηρεάζουν την αγορά, πχ. Ειδήσεις. Έτσι, με την ευρεία χρήση της μηχανικής μάθησης, τα τελευταία χρόνια, έχουν προταθεί και δημοσιευθεί πλήθος εργασιών οι οποίες χρησιμοποιώντας κάποια από τις μεθόδους της μηχανικής μάθησης προτείνουν κάποια μέθοδο και παρουσιάζουν τα αντίστοιχα αποτελέσματα.

Προχωρώντας ένα βήμα παραπέρα, πλέον οι προβλέψεις πραγματοποιούνται με συνδυασμούς τεχνικών, πχ. νευρωνικών δικτύων με ανάλυση συναισθήματος και χρήση ιστορικών δεδομένων, με τα αποτελέσματα να βελτιώνονται συνεχώς.

Η παρούσα διατριβή διαπραγματεύεται την ανάλυση, μελέτη (χτίσιμο στρατηγικής, παρουσίαση οικονομικών μεταβλητών, κτλ.) και πρόβλεψη μετοχών με την χρήση τεχνικών όπως ανάλυση χρονολογικών σειρών και μηχανικής μάθησης.

Η χρήση της μηχανικής μάθησης για το συγκεκριμένο αντικείμενο φαίνεται πως αποδίδει αρκετά καλά, με ακρίβεια στις προβλέψεις που ξεπερνούν ακόμα και το 70% σε ορισμένα μοντέλα και απλώς ικανοποιητικά για κάποια άλλα μοντέλα, σε κάθε περίπτωση όμως οι προβλέψεις είναι άνω του 50%. Επίσης ικανοποιητικές μπορούν να θεωρηθούν και οι προβλέψεις με την μέθοδο της ανάλυσης χρονοσειρών.

## Summary

The prediction, analysis and the study of financial time series variables, mainly stocks, as well as the development of a financial strategy for the stocks market and generally, as subject matters are very popular subjects not only in the academic sector but also for everyone who has ambitions on making profit through this practice. The reason this subject matter is so popular is more than obvious to everyone who wants to build a model which will manage to "beat" the market, forecast when a stock rises or falls, when someone should buy or sell, which stock will earn him more in the long-term period, etc.

The time series analysis has plenty of techniques the purpose of which is to predict the future either concerning a day, an hour, a week, etc. However, the time series analysis cannot take a lot of other factors which have been proven to affect the market into consideration, for example the daily news. Thus, with the extensive use of machine learning techniques, a lot of new methods have been proposed and plenty of publications with relevant results have been made.

Taking a step further, the predictions are now being made with combination of techniques such as neural networks with sentiment analysis and the use of historical data, with the results being increasingly promising.

This thesis is about the analysis, study and prediction of stocks with the use of techniques, such as time series analysis and machine learning.

The use of machine learning techniques for this subject shows that the accuracy of predictions can exceed 70% in some models while for the rest their performance is just satisfactory. In either case though predictions exceed 50%. The predictions through time series analysis can also be considered as satisfactory.

## Πίνακας Περιεχομένων

Περίληψη .....	4
Summary .....	5
Πίνακας Περιεχομένων .....	6
Πίνακας Εικόνων .....	9
1 Εισαγωγή.....	11
1.1 Αντικείμενο Εργασίας .....	11
1.2 Δομή Εργασίας.....	12
2 Μετοχές Γενικά .....	13
2.1 Τι Είναι Οι Μετοχές.....	13
2.2 Αποτίμηση.....	13
2.3 Χαρακτηριστικά Μετοχών .....	14
2.4 Τύποι Μετοχών .....	14
2.5 Γνωστοί Δείκτες.....	14
2.6 Χρηματιστηριακές Στρατηγικές .....	16
2.7 Backtesting.....	17
2.8 Πώς Διαβάζεται Ένας Πίνακας Μετοχών.....	17
2.9 Τι Προκαλεί Αλλαγή Τιμής Σε Μία Μετοχή .....	18
2.10 Efficient Market Hypothesis.....	19
3 Μηχανική Μάθηση Γενικά.....	20
3.1 Τι Είναι Μηχανική Μάθηση .....	20
3.2 Εφαρμογές Μηχανικής Μάθησης .....	20
3.3 Κατηγορίες Μηχανικής Μάθησης .....	21
3.4 Προετοιμασία Δεδομένων.....	22
3.4.1 Διασταυρωμένη Επικύρωση (Cross Validation) .....	22
3.4.2 Υπέρ – εκπαίδευση – Overfitting και Υπό – εκπαίδευση - Underfitting.....	23
3.5 Δένδρα Αποφάσεων .....	24
3.5.1 Τυχαία Δάση (Random Forest) .....	25
3.6 Πιθανοτικά Δίκτυα - Bayesian Networks.....	25
3.7 Τετραγωνική Διακριτή Ανάλυση (QDA) – Γραμμική Διακριτή Ανάλυση (LDA).....	26
3.8 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines).....	27
3.8.1 Παράμετρος C .....	30
3.8.2 Παράμετρος $\gamma$ .....	31
3.8.3 Συνάρτηση Κελύφους ΜΔΥ.....	31
3.9 Παλινδρόμηση Διανυσμάτων Υποστήριξης (Support Vector Regression) .....	32
3.10 Συσταδοποίηση (Clustering) .....	32
3.11 Νευρωνικά Δίκτυα .....	33

4	Χρονολογικές Σειρές Γενικά (Time Series) .....	36
4.1	Ορισμός.....	36
4.2	Αιτιοκρατία Και Στοχαστικότητα.....	36
4.3	Στασιμότητα.....	36
4.4	Διακύμανση .....	37
4.5	Τάση .....	37
4.6	Εποχικότητα .....	39
4.7	Κυκλικότητα .....	39
4.8	Τυχαίες – Ακανόνιστες κινήσεις .....	39
4.9	Λευκός θόρυβος.....	40
4.10	Λοξότητα (Skewness).....	40
4.11	Κύρτωση.....	40
4.12	Συσχέτιση .....	41
4.12.1	Γραμμική Αυτό – Συσχέτιση Και Μερική Αυτό - Συσχέτιση.....	41
4.13	Πρόβλεψη Χρονολογικών Σειρών.....	42
4.13.1	Πρόβλεψη Χρονολογικών Σειρών Με Μεθόδους Εξομάλυνσης.....	43
4.13.2	Πρόβλεψη Χρονολογικών Σειρών Με διάσπαση .....	45
4.13.3	Ανάλυση ARIMA.....	47
4.13.4	Πρόβλεψη Με Τεχνικές Μηχανικής Μάθησης.....	50
5	Εφαρμογές .....	56
5.1	Βιβλιοθήκες Και Εργαλεία Που Χρησιμοποιήθηκαν .....	56
5.1.1	Συλλογή Δεδομένων (stock datasets).....	57
5.2	Ανάλυση, Οπτικοποίηση Και Ανάπτυξη Οικονομικής Στρατηγικής.....	58
5.2.1	Τα Δεδομένα .....	59
5.2.2	Οπτικοποίηση Δεδομένων.....	61
5.2.3	Γράφημα Απεικόνισης .....	61
5.2.4	Κεριά (Candlesticks) .....	62
5.2.5	Κινητοί Μέσοι .....	62
5.2.6	Daily Returns .....	63
5.2.7	Cumulative Returns.....	64
5.2.8	Αστάθεια - Volatility.....	65
5.2.9	Ordinary Least-Squares Regression (OLS).....	65
5.2.10	Χρηματιστηριακή Στρατηγική .....	67
5.3	Εφαρμογές – Πρόβλεψη Χρονολογικής Σειράς Με Την Τεχνική ARIMA.....	68
5.3.1	Συλλογή Δεδομένων .....	69
5.3.2	Οπτικοποίηση Δεδομένων.....	69
5.3.3	Μετατροπή Χρονολογικής Σειράς Σε Στάσιμη .....	71
5.3.4	Υπολογισμός Όρων $p$ , $d$ , $q$ .....	74

5.3.5	Πρόβλεψη .....	76
5.4	Πρόβλεψη Με Τεχνικές Μηχανικής Μάθησης.....	78
5.4.1	Εκπαίδευση Μοντέλου .....	80
5.4.2	Απόδοση – Λάθη.....	80
5.4.3	Αποτελέσματα.....	81
6	Συμπεράσματα.....	85
	Αναφορές.....	86



## Πίνακας Εικόνων

Εικόνα 1 - Οι 30 εταιρείες που περιλαμβάνονται στον δείκτη DJIA, Ιανουάριος 2017 [8].....	15
Εικόνα 2 - Παράδειγμα "reversal" , φαίνεται καθαρά η αλλαγή της τάσης [10].....	16
Εικόνα 3 - Πίνακας με οικονομικά δεδομένα [16].....	17
Εικόνα 4 - Κατηγορίες Μηχανικής Μάθησης [25]. ....	22
Εικόνα 5 - Διαδικασία διασταυρούμενης επικύρωσης [27]. ....	23
Εικόνα 6 - Παράδειγμα Δένδρου Δεδομένων [31]. ....	24
Εικόνα 7 - Παράδειγμα τυχαίων δασών [32]. ....	25
Εικόνα 8 - Τύπος δικτύου Bayes .....	26
Εικόνα 9 - Διακριτή Ανάλυση [34]. ....	26
Εικόνα 10 - Κατηγοριοποίηση SVM [35]. ....	28
Εικόνα 11 - SVM hyperplane με διανύσματα υποστήριξης A, B, Γ [36].....	28
Εικόνα 12 – SVM μεγάλο διάστημα [36]. ....	29
Εικόνα 13 - SVM μικρό διάστημα [36]. ....	29
Εικόνα 14 - Μη γραμμική παλινδρόμηση.....	30
Εικόνα 15 - Αριστερά: πολυώνυμο η βαθμού, Δεξιά: radial kernel [37]. ....	30
Εικόνα 16 - Παραδείγματα διαφορετικών παραμέτρων $\gamma$ από την ιστοσελίδα της sklearn. ....	31
Εικόνα 17 - Πυρήνες SVM όπως παρουσιάζονται στην ιστοσελίδα της sklearn.....	32
Εικόνα 18 – Διάγραμμα υποστήριξης παλινδρόμησης με τα κελύφη του [39]. ....	32
Εικόνα 19 -Έξοδος αλγορίθμου k -means της sklearn με τρεις συστάδες [40]. ....	33
Εικόνα 20 -Δομή των Νευρώνων.....	34
Εικόνα 21 - Νευρωνικό δίκτυο πολλοί είσοδοι, μία έξοδος και τα αντίστοιχα βάρη [41]. ....	34
Εικόνα 22 – Χρονολογική σειρά στάσιμη (αριστερά) και μη στάσιμη (δεξιά) [45]. ....	37
Εικόνα 23 - Χρονολογική σειρά με γραμμική τάση [46]. ....	38
Εικόνα 24 - Χρονολογική σειρά με λογαριθμική τάση [46]. ....	38
Εικόνα 25 - Χρονολογική σειρά με εκθετική τάση [46]. ....	38
Εικόνα 26 - Παράδειγμα εποχικότητας [47]. ....	39
Εικόνα 27 - Παράδειγμα κυκλικότητας [48]. ....	39
Εικόνα 28 - Παράδειγμα θετικής και αρνητικής λοξότητας [50]. ....	40
Εικόνα 29 - Παράδειγμα κύρτωσης. $\alpha = 3$ κανονική κατανομή, $\alpha < 3$ λεπτό - κυρτή, $\alpha > 3$ πλατύ - κυρτή [52].....	41
Εικόνα 30 - Γράφημα αυτό - συσχέτισης μοντέλου αυτό - παλινδρόμησης για συντελεστή $\phi$ θετικό αριστερά και $\phi$ αρνητικό δεξιά [55]. ....	42
Εικόνα 31 - Παράδειγμα εφαρμογής κινητών μέσων [57]. ....	44
Εικόνα 32 - Παράδειγμα εφαρμογής εκθετικής μεθόδου, παρατηρούμε πως η τιμή από τα βάρη μειώνεται όσο πιο παλιά είναι η παρατήρηση [57]. ....	45
Εικόνα 33 - Διαδικασίες AR [64]. ....	49
Εικόνα 34 - Διαδικασίες MA και λευκός θόρυβος [64]. ....	49
Εικόνα 35 - Διαδικασίες MA [64]. ....	50
Εικόνα 36 - Διάγραμμα ροής μεθοδολογίας sentiment analysis [65].....	51
Εικόνα 37 - Παράδειγμα κατηγοριοποίησης των tweets [65]. ....	52
Εικόνα 38 - Αρχιτεκτονική Back Propagation Neural Network [67]. ....	52
Εικόνα 39 - LSTM δίκτυο [69]. ....	53
Εικόνα 40 - Single Layer Perceptron Model.....	53
Εικόνα 41 - Multi - Layer Perceptron Model.....	54
Εικόνα 42 - Radial Basis Function.....	55
Εικόνα 43 - Παροχή ιστορικών δεδομένων από google finance.....	59
Εικόνα 44 - Παράδειγμα αρχείου csv. ....	60
Εικόνα 45 - Τιμών κλεισίματος (close) σε συνάρτηση με τον χρόνο (2010 - 2017), τριών μετοχών μαζί.....	61
Εικόνα 46 - Παράδειγμα κεριών στην εφαρμογή.....	62
Εικόνα 47 - Γράφημα απεικόνισης κινητών μέσων.....	62
Εικόνα 48 - Κινητοί μέσοι που απεικονίζουν την μετοχή της apple για παράθυρο ενός μήνα και ενός έτους.....	63
Εικόνα 49 - Γράφημα daily returns.....	64
Εικόνα 50 - Ιστόγραμμα daily returns, όπως βλέπουμε οι περισσότερες τιμές είναι γύρω από το μηδέν. ....	64
Εικόνα 51 - Γράφημα cumulative returns.....	64
Εικόνα 52 - Γράφημα Volatility, απεικονίζοντας δύο μετοχές, πράσινο την μετοχή της google και μπλε της apple. ....	65
Εικόνα 53 - Αποτελέσματα εφαρμογής OLC.....	67
Εικόνα 54 - Διάγραμμα Ordinary Least - Squares Regression.....	67
Εικόνα 55 - Γράφημα στρατηγικής. ....	68
Εικόνα 56 - Το παραπάνω γράφημα μεγεθυμένο, ώστε να φαίνονται καθαρότερα τα σήματα "πούλησε" και "αγόρασε". 68	68
Εικόνα 57 - Πορεία μετοχής S&P 500 από τον 2000 μέχρι το 2017.....	70
Εικόνα 58 - Απεικόνιση πορείας της μετοχής όπως και πριν, μαζί με κινούμενος μέσους και κινούμενη διασπορά. ....	71

Εικόνα 59 - Γράφημα για την λογαριθμημένη τιμή κλεισίματος με και τους κινούμενους μέσους.....	72
Εικόνα 60- Γράφημα απεικόνισης αρχικής πορείας της μετοχής, κινούμενων μέσων και διακύμανσης ενός μηνός. ....	72
Εικόνα 61 - Γράφημα απεικόνισης αρχικής πορείας της μετοχής, κινούμενων μέσων και διακύμανσης ενός μηνός, μετά από την εφαρμογή των διαφορών. ....	73
Εικόνα 62 - Απεικόνιση αρχικής χρονολογικής σειράς, τάσης, εποχικότητας και residuals.....	73
Εικόνα 63 - Διαγράμματα αυτό - συσχέτισης και μερικής αυτό - συσχέτισης. ....	74
Εικόνα 64- RSS μοντέλου ARIMA για τους όρους AR.....	75
Εικόνα 65 - RSS μοντέλου ARIMA για τους όρους MA. ....	75
Εικόνα 66 - Έξοδος αποτελεσμάτων ARIMA. ....	76
Εικόνα 67 - Γράφημα πυκνότητας των residuals. ....	76
Εικόνα 68 - Αποτελέσματα πρόβλεψης ARIMA. ....	77
Εικόνα 69 - Αποτελέσματα προβλέψεων ARIMA. ....	77
Εικόνα 70 - Γράφημα daily returns τελευταίων δύο μηνών. ....	79
Εικόνα 71 – Πρόβλεψη γραμμικής παλινδρόμησης για τα δεδομένα αξιολόγησης. ....	82
Εικόνα 72 - Προβλέψεις SVR Linear.....	82
Εικόνα 73 - Προβλέψεις SVR - RBF. ....	82
Εικόνα 74 - Προβλέψεις MLP.....	83

# 1 Εισαγωγή

## 1.1 Αντικείμενο Εργασίας

Η εργασία αυτή έχει ως αντικείμενο την πρόβλεψη οικονομικών χρονολογικών σειρών, με την τεχνική της ανάλυσης χρονολογικών σειρών και με μεθόδους της μηχανικής μάθησης καθώς και την κατανόηση και εξαγωγή οικονομικών στρατηγικών, μέσω των παραπάνω μεθόδων και να παρουσιάσει τα αποτελέσματα στον κάθε ενδιαφερόμενο. Έχει πραγματοποιηθεί μία εφαρμογή η οποία χρησιμοποιώντας τις παραπάνω τεχνικές πραγματοποιεί προβλέψεις της τιμής κλεισίματος των μετοχών. Επίσης έχει πραγματοποιηθεί άλλη μία εφαρμογή η οποία μέσω γραφημάτων κυρίως και γνωστών οικονομικών τεχνικών παρουσιάζει την πορεία της μετοχής στο επιλεγμένο χρονικό διάστημα, εφαρμόζει μία απλή οικονομική στρατηγική και παρουσιάζει τα αποτελέσματα στον χρήστη.

Η ανάλυση χρονολογικών σειρών γενικώς ως στόχο έχει την εξαγωγή τάσης, κυκλικότητας ή κάποιου συγκεκριμένου μοτίβου που εμπεριέχετε σε αυτήν καθώς επίσης και την πρόβλεψη των μελλοντικών τιμών της.

Πιο συγκεκριμένα όμως για χρηματοοικονομικά δεδομένα η ανάλυση είναι χρήσιμη διότι μέσω αυτής μας είναι εφικτό να παρακολουθήσουμε πως μία μετοχή, ή κάποια άλλη οικονομική μεταβλητή, εξελίσσεται σε σχέση με τον χρόνο, δηλαδή θα μπορούμε να δούμε αν έχει και πότε εποχικότητα ή υπάρχουν τάσεις σε αυτήν, πότε δεν έχει αποδώσει καλά, κτλ. Έτσι θα είναι πιο εύκολο στην κατανόηση της πορείας της μετοχής για το διάστημα που επιλέξαμε αλλά και επίσης να αναπτύξουμε οικονομικές στρατηγικές οι οποίες θα έχουν ως σκοπό να συμβουλέψουν τον επενδυτή ή τον κάθε ενδιαφερόμενο για πιθανή αγορά ή πώληση για παράδειγμα μίας μετοχής [1].

Η εφαρμογή της μηχανικής μάθησης σε χρονολογικές σειρές από την άλλη έχει μελετηθεί αρκετά τα τελευταία χρόνια και έχουν αναπτυχθεί και προταθεί πλήθος τεχνικών πρόβλεψης μετοχών για τους πιο γνωστούς αλγόριθμους. Η πρόβλεψη αυτή μπορεί να αφορά την πορεία κάποιας οικονομικής μεταβλητής σε βάθος χρόνου (long term), πχ. Αν στον επόμενο μήνα η μεταβλητή θα είναι μεγαλύτερη ή όχι, αλλά και προβλέψεις για την επόμενη μέρα, ώρα, κτλ.

Πέραν όμως των δύο παραπάνω τεχνικών, μεταξύ των πιο δημοφιλών τρόπων πρόβλεψης αποτελούν επίσης και η χρήση νευρωνικών δικτύων, η ανάλυση ειδησεογραφικών ειδήσεων και τα τελευταία χρόνια κυρίως, κάποιος συνδυασμός των ανωτέρω. Η πρόβλεψη των οικονομικών δεδομένων παραμένει ένα από τα πιο δημοφιλή αντικείμενα.

## 1.2 Δομή Εργασίας

Στο κεφάλαιο 1 περιγράφονται συνοπτικά το αντικείμενο και η δομή της εργασίας. Το κεφάλαιο 2 περιέχει μία θεωρητική παρουσίαση των μετοχών, πώς αυτές μεταβάλλονται, γιατί μεταβάλλονται, τι τις επηρεάζει καθώς και τους τύπους μετοχών. Το κεφάλαιο 3 αποτελεί μία εισαγωγή στην μηχανική μάθηση (machine learning) και τι μεθόδους και τεχνικές χρησιμοποιεί για την πρόβλεψη, παρουσιάζονται γενικά οι πιο δημοφιλείς αλγόριθμοι σε κάθε κατηγορία και σε τι εφαρμογές συναντώνται. Στο κεφάλαιο 4 παρουσιάζεται μία συνοπτική θεωρητική εισαγωγή στις χρονολογικές σειρές (time series), καθώς και στην ανάλυση και πρόβλεψη αυτών, με τις τεχνικές που αναφέρθηκαν και παραπάνω.

Στο κεφάλαιο 5 περιγράφονται 3 εφαρμογές, στην πρώτη γίνεται ανάλυση και εξαγωγή συμπερασμάτων στην πορεία μίας μετοχής και υλοποιείται μία οικονομική στρατηγική. Η δεύτερη με την χρήση τεχνικών μηχανικής μάθησης υλοποιεί προβλέψεις με είσοδο ιστορικά δεδομένα και στην τρίτη χρησιμοποιείται η ανάλυση χρονολογικών σειρών και πραγματοποιούνται μέσω αυτής οι προβλέψεις.

Τέλος, στο 6<sup>ο</sup> κεφάλαιο διατυπώνονται τα συμπεράσματα από τα αποτελέσματα των ανωτέρω εφαρμογών.

## 2 Μετοχές Γενικά

### 2.1 Τι Είναι Οι Μετοχές

Η μετοχή είναι ένα χρεόγραφο που στην ουσία αποτελεί ένα μερίδιο ιδιοκτησίας το οποίο διανεμήθηκε από μία εταιρία και αντιπροσωπεύει την αξίωση μεριδίου στα κέρδη της εταιρείας, [2]. Ο σκοπός που διανέμονται από τις εταιρίες είναι κυρίως η αύξηση του μετοχικού τους κεφαλαίου [3], αποτελώντας μία καλύτερη εναλλακτική λύση από το να δανειστούν από μία τράπεζα για τον ίδιο σκοπό.

Η μετοχή, ως αξιόγραφο, ενσωματώνει τα δικαιώματα του μετόχου που πηγάζουν από τη συμμετοχή του στην ανώνυμη εταιρία. Τα δικαιώματα αυτά, είναι ανάλογα του αριθμού μετοχών που κατέχει ο μέτοχος. Ενδεικτικά δικαιώματα που προκύπτουν από την κατοχή μετοχών είναι το ποσοστό ίσο με τον αριθμό των μετοχών που κατέχει ο μέτοχος προς το σύνολο των μετοχών της εταιρείας, του μερίσματος από τα διανεμόμενα κέρδη της εταιρείας, καθώς και αντίστοιχο ποσοστό από την περιουσία της εταιρείας, σε περίπτωση που αυτή διαλυθεί. Αντίστοιχα έχει και τον αναλογούντα αριθμό ψήφων στην Γενική Συνέλευση των μετόχων, εκτός εάν κατέχει μετοχές άνευ ψήφου.

Οι μετοχές κυρίως ανταλλάσσονται στα χρηματιστήρια, τα οποία είναι μέρη όπου εξειδικευμένοι διαμεσολαβητές αγοράζουν και πωλούν χρεόγραφα και αποφασίζουν για την τιμή, υπάρχουν τα φυσικά αλλά και τα εικονικά χρηματιστήρια.

### 2.2 Αποτίμηση

Οι κοινές μετοχές μπορούν να αποτιμηθούν, αποτιμώντας συγκριτικά παρόμοιες εταιρείες, λαμβάνοντας υπόψη μία κοινή παράμετρο όπως κέρδη, ταμιακές αγορές, λογιστική αξία, πωλήσεις, κτλ. Ο εντοπισμός συγκρίσιμων εταιρειών για την αξιολόγηση των δεδομένων τους αποτελεί βασική προϋπόθεση για την μέθοδο των πολλαπλασιαστών [5]. Μία δεύτερη μέθοδος είναι η χρησιμοποίηση του δείκτη P/E ratio ο οποίος δείχνει πόσες φορές παραπάνω οι επενδυτές είναι διατεθειμένοι να πληρώσουν για τα κέρδη της εταιρείας και σε πόσα έτη θα πάρουν πίσω τα χρήματα της επένδυσής τους. Επιπλέον τρόποι μπορούν να αποτελούν παραλλαγές των πολλαπλασιαστών οι οποίοι ενδεχομένως να υπολογίζουν διαφορετικές οικονομικές μεταβλητές όπως κέρδη πριν από τόκους, φόρους, κτλ [5].

### 2.3 Χαρακτηριστικά Μετοχών

- Ονομαστική αξία: Η δηλωθείσα αξία ενός τίτλου
- Κεφαλαιοποίηση: Ο αριθμός των μετοχών μίας εταιρείας που βρίσκονται σε κυκλοφορία επί την τιμή της μετοχής, γνωστή και ως χρηματιστηριακή αξία
- Υψηλή κεφαλαιοποίηση (large cap ή big cap): Αναφέρεται σε εταιρείες με αξίας κεφαλαιοποίησης άνω των 10 δις. δολαρίων.
- Μεσαία κεφαλαιοποίηση (mid cap): Αναφέρεται σε εταιρείες με αξίας κεφαλαιοποίησης μεταξύ 2 και 10 δις. δολαρίων.
- Μικρή κεφαλαιοποίηση (small cap): Αναφέρεται σε εταιρείες με αξίας κεφαλαιοποίησης μεταξύ 300 εκατομμυρίων και 2 δις. δολαρίων.
- Ανώνυμες μετοχές: Τίτλος μετοχής που αποτελεί ιδιοκτησία όποιου κατέχει το σχετικό πιστοποιητικό κυριότητας της μετοχής.

### 2.4 Τύποι Μετοχών

Υπάρχουν δύο τύποι μετοχών οι «κοινές μετοχές» και οι «προνομιούχες μετοχές» [5].

Κοινές μετοχές: Όταν αναφερόμαστε σε μετοχές συνήθως αναφερόμαστε σε αυτές, είναι ο πιο συνηθισμένος τύπος μετοχής και περιλαμβάνει όλα τα βασικά δικαιώματα ενός μετόχου, όπως δικαίωμα συμμετοχής στα κέρδη, στην έκδοση νέων μετοχών, στο προϊόν της εκκαθάρισης, καθώς και δικαίωμα ψήφου στη Γενική Συνέλευση της εταιρείας και συμμετοχής στη διαχείρισή της.

Προνομιούχες μετοχές: Ως λειτουργίες μοιάζουν με αυτά των ομολόγων, προσφέρει απλά ένα προβάδισμα έναντι των κατόχων κοινών μετοχών, στη λήψη μερίσματος και στη λήψη του προϊόντος της εκκαθάρισης σε περίπτωση διάλυσης της επιχείρησης, αλλά συνήθως στερείται του δικαιώματος ψήφου και συμμετοχής στη διαχείριση της επιχείρησης [4].

### 2.5 Γνωστοί Δείκτες

- Dow Jones (Industrial Average): Βασίζεται στις τιμές των 30 μετοχών των μεγαλύτερων εταιρειών στις Ηνωμένες Πολιτείες, οι οποίες συναλλάσσονται στο χρηματιστήριο της Νέας Υόρκης (New York Stock Exchange, NYSE). Ο δείκτης είναι μία σταθμισμένη μέση τιμή, το οποίο σημαίνει πως οι μετοχές με μεγαλύτερες τιμές, έχουν μεγαλύτερη βαρύτητα [6].

- Standard & Poor 500: Αυτός ο δείκτης βασίζεται στην αξία 500 εταιρειών, από την οικονομία των Ηνωμένων Πολιτειών πάλι, καταγράφεται η συνολική αξία της ιδιοκτησίας και των 500 εταιρειών. Λόγω του ότι καλύπτει ένα μεγάλο μερίδιο της αγοράς θεωρείται πως την αντιπροσωπεύει καλύτερα. Δίνει μεγαλύτερα βάρη στις μεγαλύτερες εταιρείες. [7]
- Nasdaq Composite: Εν αντίθεση με τους προηγούμενους, αυτός αποτελείται από μικρότερες και καινούργιες εταιρίες, είναι επίσης σταθμισμένος δείκτης με βάση την αξία περισσότερων των 3000 μετοχών

<i>The 30 Companies in the DJIA as of January 2017</i>		
<b>Company Name</b>	<b>Ticker</b>	<b>Exchange</b>
The 3M Company	MMM	NYSE
The American Express Company	AXP	NYSE
Apple Inc.	AAPL	NASDAQ
The Boeing Company	BA	NYSE
Caterpillar Inc.	CAT	NYSE
Chevron Corporation	CVX	NYSE
Cisco Systems, Inc.	CSCO	NASDAQ
The Coca-Cola Company	KO	NYSE
The Walt Disney Company	DIS	NYSE
E. I. du Pont de Nemours and Company	DD	NYSE
Exxon Mobil Corporation	XOM	NYSE
General Electric Company	GE	NYSE
The Goldman Sachs Group, Inc.	GS	NYSE
The Home Depot, Inc.	HD	NYSE
IBM Corporation	IBM	NYSE
Intel Corporation	INTC	NASDAQ
Johnson & Johnson	JNJ	NYSE
JPMorgan Chase & Co.	JPM	NYSE
McDonald's Corporation	MCD	NYSE
Merck & Co., Inc.	MRK	NYSE
Microsoft Corporation	MSFT	NASDAQ
Nike, Inc.	NKE	NYSE
Pfizer Inc.	PFE	NYSE
Procter & Gamble Co.	PG	NYSE
Travelers Companies Inc	TRV	NYSE
United Technologies Corporation	UTX	NYSE
UnitedHealth Group Inc.	UNH	NYSE
Verizon Communications, Inc.	VZ	NYSE
Visa Inc.	V	NYSE
Wal-Mart Stores, Inc.	WMT	NYSE

Εικόνα 1 - Οι 30 εταιρείες που περιλαμβάνονται στον δείκτη DJIA, Ιανουάριος 2017 [8].

## 2.6 Χρηματιστηριακές Στρατηγικές

Οι οικονομικές στρατηγικές έχουν και αυτές διάφορες φάσεις υλοποίησης και πολυπλοκότητας. Αναφέρονται στις προδιαγραφές τις οποίες θέτει ένας επενδυτής, ώστε να πραγματοποιήσει συναλλαγές να σταματήσει αγορές, κοκ. Γενικά χωρίζονται σε τεχνικές και θεμελιώδεις, οι οποίες και οι δύο στηρίζονται σε ποσοτικά δεδομένα. Συνήθως επαληθεύονται από την μεθοδολογία «backtesting», η οποία θα εξηγηθεί πιο κάτω.

Διασταύρωση – Crossover [9] → Για τα μοντέλα τεχνικής ανάλυσης αποτελεί στην ουσία δείκτη υπόδειξης σημάτων για «αγορά» και «πώληση». Για τους εμπόρους αποτελούν τεχνικούς δείκτες για να παρακολουθούν τις μεταβολές στις τάσεις των οικονομικών μεταβλητών, κτλ. Μία συνηθισμένη τεχνική που χρησιμοποιεί κινούμενους μέσους για βρει μία αλλαγή στην τάση μίας οικονομικής μεταβλητής «reversal».



Εικόνα 2 - Παράδειγμα "reversal", φαίνεται καθαρά η αλλαγή της τάσης [10].

Κάποιες γνωστές στρατηγικές είναι [11]:

- Χρυσός σταυρός - Golden Cross → Βασίζεται στους απλούς κινητούς μέσους και η ως στρατηγική είναι απλή. Ο χρυσός σταυρός εμφανίζεται όταν οι κινητοί μέσοι των 50 ημερών ξεπερνούν τους αντίστοιχους των 200 ημερών [12].
- 10-30 Crossover → Χρησιμοποιεί 2 σήματα απλών κινητών μέσων με παράθυρα 10 και 30 ημερών αντίστοιχα.
- 5-10-20 Crossover → Εδώ χρησιμοποιούνται 3 σήματα απλών κινητών μέσων με παράθυρα των 5, 10, 20 ημερών αντίστοιχα.



## 2.7 Backtesting

Είναι η διαδικασία της δοκιμής (testing) μίας χρηματιστηριακής στρατηγικής σε σχετικά ιστορικά δεδομένα για την εξασφάλιση της βιωσιμότητας πριν ο επενδυτής ρισκάρει κάποιο κεφάλαιο [13]. Στην πράξη αυτή η τεχνική χρησιμοποιεί τα ιστορικά δεδομένα ώστε να αξιολογήσει την απόδοση του μοντέλου που έχουμε κατασκευάσει για την πρόβλεψη μιας μετοχής, συγκρίνοντας τις προβλέψεις του μοντέλου με τα ιστορικά στοιχεία. Σχεδόν όλες οι μέθοδοι μπορούν να αξιολογηθούν με αυτήν την τεχνική. Ως τεχνική είναι χρήσιμη διότι βοηθάει τους επενδυτές και τους αναλυτές να αξιολογήσουν τις μεθόδους τους πριν τις εφαρμόσουν [14].

## 2.8 Πώς Διαβάζεται Ένας Πίνακας Μετοχών

Stock tables / stock quotes

Συνήθως οι οικονομικοί πίνακες μοιάζουν με παρακάτω [15]:

52W high	52W low	Stock	Ticker	Div	Yield %	P/E	Vol 00s	High	Low	Close	Net chg
545.39	19.75	ResMed	RMD			52.5	3831	42.00	39.51	41.50	-1.90
11.63	3.55	Revlon A	REV				162	6.09	5.90	6.09	+0.12
77.25	55.13	RioTinto	RTP	2.30	3.2		168	72.75	71.84	72.74	+0.03
31.31	16.63	RitchieBr	RBA			20.9	15	24.49	24.29	24.49	-0.01
8.44	1.75	RiteAid	RAD				31028	4.50	4.20	4.31	+0.21
538.63	18.81	RobtHalf	RHI			26.5	6517	27.15	26.50	26.50	+0.14
51.25	27.69	Rockwell	ROK	1.02	2.1	14.5	6412	47.99	47.00	47.54	+0.24

Column 1    Column 2    Column 3    Column 4    Column 5    Column 6    Column 7    Column 8    Column 9    Column 10    Column 11    Column 12

Εικόνα 3 - Πίνακας με οικονομικά δεδομένα [16].

Στήλη 1 και 2: High / Low, είναι η υψηλότερη και η χαμηλότερη τιμή της μετοχής ως αποτέλεσμα ύστερα από συναλλαγή.

Στήλη 3: Stock, το όνομα της μετοχής.

Στήλη 4: Ticker, Είναι μοναδικό όνομα το οποίο χρησιμοποιείται ως αναγνωριστικό για την εκάστοτε μετοχή.

Στήλη 5: Div., Η ετήσια πληρωμή μερισμάτων (dividend) για κάθε μερίδιο, αν το πεδίο είναι κενό η εταιρεία δεν πληρώνει μερίσματα.

Στήλη 6: Dividend Yield, Η τοις 100 επιστροφή των μερισμάτων.

Στήλη 7: Price / Earnings Ratio, Υπολογίζεται διαιρώντας την τωρινή τιμή της μετοχής με τα μερίδια των 4 προηγούμενων τριμήνων.

Στήλη 8: Trading Volume, «Δείχνει» τον συνολικό αριθμό μερισμάτων που συναλλάχθηκαν εκείνη την μέρα σε εκατοντάδες με επιπλέον «00» έχουμε τον κανονικό αριθμό.

Στήλη 9 και 10: Day High and Low, Η υψηλότερη και η χαμηλότερη τιμή που συναλλάχθηκε την εκάστοτε μέρα.

Στήλη 11: Close, Η τιμή κλεισίματος είναι η τελευταία τιμή που καταγράφηκε πριν κλείσει η αγορά. Η επόμενη μέρα δεν ξεκινάει με την τιμή κλεισίματος, διότι η τιμή συνεχώς αλλάζει ακόμα και όταν το χρηματιστήριο είναι κλειστό.

Στήλη 12: Net Change, Η διαφορά τιμής από της προηγούμενης μέρας την τιμή κλεισίματος.

## 2.9 Τι Προκαλεί Αλλαγή Τιμής Σε Μία Μετοχή

Υπάρχουν πολλοί παράγοντες οι οποίοι μπορούν να προκαλέσουν αλλαγή στην τιμή μιας μετοχής, καθώς η προσφορά και η ζήτηση αλλάζουν συνεχώς, π.χ. αν περισσότεροι άνθρωποι θέλουν να αγοράσουν μία μετοχή από αυτούς που θέλουν να την αγοράσουν τότε η τιμή της μετοχής θα ανέβει, αντιθέτως αν περισσότεροι θέλουν να πουλήσουν μία μετοχή από να την αγοράσουν τότε έχουμε μεγαλύτερη προσφορά από ότι ζήτηση, οπότε η τιμή της μετοχής θα πέσει. Αυτό που ωθεί τον κόσμο να έχει αυξημένη ζήτηση σε μία μετοχή ή το αντίθετο να την «αποφεύγει» μπορεί να είναι οι οικονομικές ειδήσεις θετικές είτε αρνητικές, επίσης σημασία έχουν τα κέρδη της εκάστοτε επιχείρησης, μία επιχείρηση χωρίς κέρδη μακροπρόθεσμα δεν μπορεί να επιβιώσει. Συνοπτικά θα μπορούσαμε να πούμε πώς τα παρακάτω επίσης επηρεάζουν την αγορά [17]:

- Ιστορικές τιμές
- Ειδήσεις
- Εξέλιξη τεχνολογίας
- Η Ψυχολογία της αγοράς
- Γενική διάθεση (των επενδυτών)

Επίσης υπάρχουν και άλλες μελέτες που αφορούν την μεταβολή των μετοχών όπως του καθηγητή Gordon Fisher ο οποίος εξέτασε τις επιδράσεις τεσσάρων μεταβλητών στις τιμές των μετοχών, όπως κυριάρχησαν στην αγορά για διαφορετικές εταιρίες. Αυτές οι τέσσερις μεταβλητές ήταν:

1. Το τελευταίο δηλωθέν μέρισμα ανά μετοχή
2. Το τελευταίο δηλωθέν παρακρατηθέν κέρδος ανά μετοχή
3. Η μέση ετήσια ανάπτυξη στα μερίσματα ανά μετοχή, και
4. Τα μεγέθη των εταιριών στις οποίες οι μετοχές αντιστοιχούν [18]

Όπως και υποθέσεις σαν αυτήν του τυχαίου περιπάτου η οποία ισχυρίζεται πως οι τιμές των μετοχών δεν εξαρτώνται στις ιστορικές τιμές τους και αυτό έχει ως αποτέλεσμα το ότι δεν μπορεί να εξαχθεί από εκεί κάποιο μοτίβο, αυτό σημαίνει πως η κίνηση των μετοχών στο μέλλον είναι τυχαίες και απρόβλεπτες [19].

## 2.10 Efficient Market Hypothesis

Εδώ, αξίζει να αναφερθεί εδώ μία οικονομική θεωρία γνωστή ως, **Efficient Market Hypothesis**, η οποία εξελίχθηκε από τον Eugene Fama, το 1960 στην οποία αναφέρεται ότι είναι αδύνατον για κάποιον να «κερδίσει» την αγορά, διότι οι μετοχές πάντα ανταλλάσσονται σε δίκαιη τιμή κάνοντας το έτσι αδύνατον για τους επενδυτές να αγοράσουν μετοχή υποτιμημένη ή το αντίθετο, αφήνοντας στους επενδυτές ως μόνη επιλογή για να αυξήσουν τα κέρδη τους, το μεγαλύτερο ρίσκο. Δηλαδή οι τιμές της «χ» μετοχής για αύριο, θα προέλθουν μόνο από τα δεδομένα του αύριο. Ωστόσο αν ίσχυε κατά απόλυτο βαθμό αυτή η θεωρία, τότε δεν θα μπορούσαμε να προβλέψουμε κάποια μελλοντική τιμή με ακρίβεια άνω του 50%. [20] Υπάρχουν 3 είδη Efficient Market Hypothesis, δυνατή (strong), ημί – δυνατή (semi - strong) και αδύνατη (weak) [21].

## 3 Μηχανική Μάθηση Γενικά

### 3.1 Τι Είναι Μηχανική Μάθηση

Ως ορισμός για την μηχανική μάθηση έχουν δοθεί πολλοί κάποιιοι από αυτούς είναι:

*Ένα πρόγραμμα υπολογιστή θεωρείται ότι μαθαίνει από την εμπειρία  $E$  σε σχέση με μια κατηγορία εργασιών  $T$  και μια μετρική απόδοσης  $P$ , αν η απόδοση του σε εργασίες της  $T$ , όπως μετριοούνται από την  $P$ , βελτιώνονται με την εμπειρία  $E$  [22]*

Ο παραπάνω ορισμός μπορεί να ερμηνευτεί ως: Ένας υπολογιστής ο οποίος επεξεργάζεται μία ή περισσότερες εργασίες, όσο αυξάνεται η «εμπειρία», τόσο πρέπει να αυξάνεται και η απόδοση.

Η μηχανική μάθηση είναι μία υποκατηγορία της τεχνητής νοημοσύνης. Τα συστήματα που σχεδιάζονται μπορούν να μάθουν από τα δεδομένα αφού εκπαιδευτούν. Αυτά τα συστήματα μπορούν να μάθουν και να βελτιωθούν με την εμπειρία και με τον χρόνο να βελτιστοποιηθούν. [23]

Επιπλέον ορισμοί:

- **Carbonell (1987)**, "...η μελέτη υπολογιστικών μεθόδων για την απόκτηση νέας γνώσης, νέων δεξιοτήτων και νέων τρόπων οργάνωσης της υπάρχουσας γνώσης".<sup>[2]</sup>
- **Witten & Frank (2000)**, "Κάτι μαθαίνει όταν αλλάζει τη συμπεριφορά του κατά τέτοιο τρόπο ώστε να αποδίδει καλύτερα στο μέλλον".

### 3.2 Εφαρμογές Μηχανικής Μάθησης

Η μηχανική μάθηση μπορεί να χρησιμοποιηθεί σε πλήθος αντικειμένων, μεταξύ των οποίων:

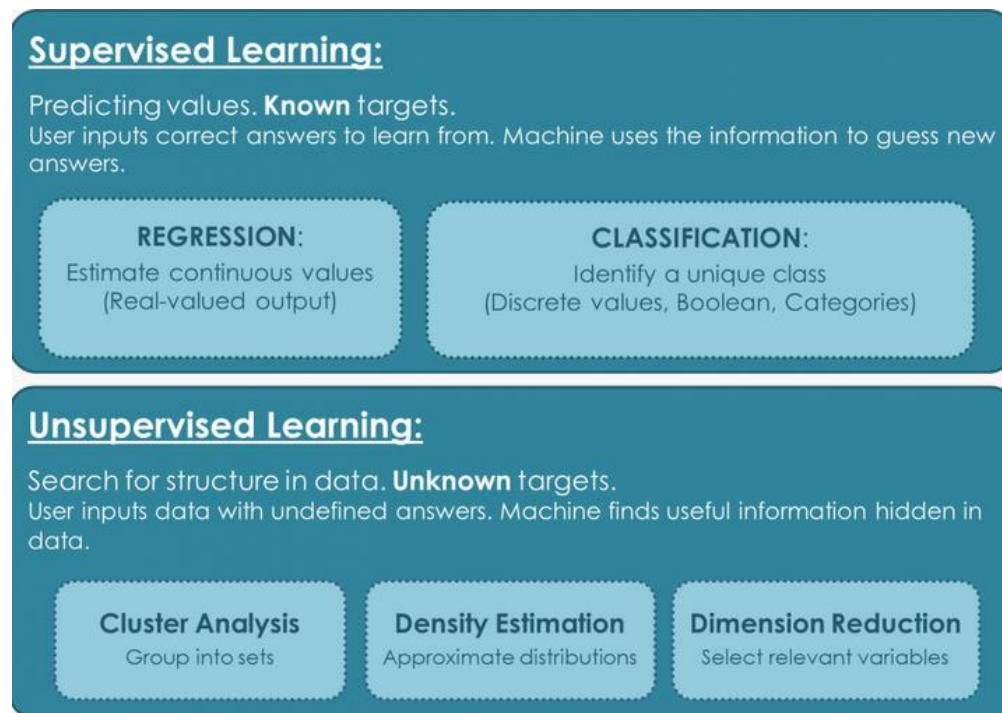
1. Λογισμικό: Αποσκοπούν στην βελτίωση εμπειρίας με τον χρήστη, προσπαθώντας να προβλέψει τι πιθανόν να ήθελε στην συνέχεια ο χρήστης
2. Ανίχνευση ανεπιθύμητων: Κλασική εφαρμογή της μηχανικής μάθησης
3. Ανίχνευση φωνής
4. Συναλλαγές μετοχών: Προσπαθεί να συμβουλέψει τον χρήστη για την αγορά ή πώληση μίας μετοχής
5. Ρομποτική
6. Διαφήμιση

## 7. Ιατρική

### 3.3 Κατηγορίες Μηχανικής Μάθησης

Υπάρχουν πολλές τεχνικές μηχανικής μάθησης, που εφαρμόζονται ανάλογα με την φύση του προβλήματος, το απαιτούμενο αποτέλεσμα είναι στη ουσία το κριτήριο που θα επιλεγεί στις ουσιαστικά δύο κατηγορίες μάθησης

- Μάθηση με επίβλεψη: Supervised learning, Μάθηση με παραδείγματα. Τα δεδομένα εκπαίδευσης συνοδεύονται από ετικέτες “labels”, για κάθε παράδειγμα στα δεδομένα εκπαίδευσης έχουμε ένα αντικείμενο εισαγωγής και ένα εξαγωγής. Τα νέα δεδομένα κατηγοριοποιούνται με βάση το σύνολο της εκπαίδευσης. Ένα από τα θέματα που θα πρέπει κάποιος να προσέξει είναι το «*bias-variance dilemma*»: δηλαδή πώς αντιδράει με μεγαλύτερη ακρίβεια το μοντέλο χρησιμοποιώντας διαφορετικά δεδομένα για εκπαίδευση [23]. Στα προβλήματα τα οποία χρησιμοποιούν μάθηση με επίβλεψη διακρίνονται δύο είδη προβλημάτων:
  - Ταξινόμησης: Classification, Αφορά στην δημιουργία μοντέλων πρόβλεψης διακριτών κλάσεων – κατηγοριών
  - Παρεμβολής: Regression, Αφορά την δημιουργία μοντέλων πρόβλεψης αριθμητικών τιμών [24]
- Μάθηση χωρίς επίβλεψη: Unsupervised learning, Μάθηση από παρατήρηση. Ο αλγόριθμος προσπαθεί να βρει ένα κρυφό μοτίβο «pattern» στα δεδομένα και να ορίσει κατηγορίες, συστάδες κτλ. Οι κατηγορίες των δεδομένων εκπαίδευσης δεν είναι γνωστές.
- Ενισχυτική μάθηση: Reinforcement Learning, ο αλγόριθμος αλληλοεπιδρά με το περιβάλλον και αναλόγως «μαθαίνει». Είναι μία υβριδική προσέγγιση παραλλαγή της μάθησης με επίβλεψη. Συναντάται σε εφαρμογές όπως έλεγχος κίνησης ρομπότ, κτλ.



Εικόνα 4 - Κατηγορίες Μηχανικής Μάθησης [25].

### 3.4 Προετοιμασία Δεδομένων

Πριν από την χρησιμοποίηση του κάθε αλγορίθμου για κατηγοριοποίηση ή παλινδρόμηση τα δεδομένα που θα χρησιμοποιηθούν στην εκάστοτε μέθοδο θα πρέπει στην αρχή να «προετοιμαστούν» ώστε να εισαχθούν στον εκάστοτε αλγόριθμο, τα βασικά βήματα αυτής της διαδικασίας είναι τα εξής:

- Καθαρισμός δεδομένων (Data cleaning) → Διαχείριση ελλিপών – εσφαλμένων τιμών και προ - επεξεργασία των δεδομένων προκειμένου να μειωθεί ο θόρυβος
- Ανάλυση σχετικότητας → Απομάκρυνση μη σχετικών τιμών(outliers)
- Μετασχηματισμός δεδομένων → Γενίκευση δεδομένων

#### 3.4.1 Διασταυρωμένη Επικύρωση (Cross Validation)

Είναι στην ουσία μία εκτίμηση για την επίδοση ενός συστήματος βασισμένο στην μηχανική μάθηση. Αυτό το σύστημα καλείται να προσαρμόσει το μοντέλο του όσο το δυνατότερο καλύτερα στα δεδομένα της εκπαίδευσης. Η εφαρμογή της έχει ως εξής: το σύνολο των δεδομένων της εκπαίδευσης χωρίζεται σε κ ισομεγέθη περίπου τμήματα και εκτελούνται κ έλεγχοι όσοι δηλαδή και

τα τμήματα, σε κάθε έλεγχο το μοντέλο εκπαιδεύεται στα  $k - 1$  σύνολα δεδομένων εκπαίδευσης και καταγράφεται η απόδοση, κοκ. Υπολογίζοντας τον μέσο όρο των αποδόσεων των και ελέγχων, υπολογίζεται η ολική απόδοση του συστήματος, σύνηθες τιμές για το  $k$  είναι οι 5, 10  $\rightarrow$  5 fold validation, 10 – fold validation, κτλ [26].

Dataset	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5
1	Test	Train	Train	Train	Train
2	Train	Test	Train	Train	Train
3	Train	Train	Test	Train	Train
4	Train	Train	Train	Test	Train
5	Train	Train	Train	Train	Test

Εικόνα 5 - Διαδικασία διασταυρούμενης επικύρωσης [27].

Επίσης αποτελεί και μία από τις μεθόδους αντιμετώπισης της υπέρ – εκπαίδευσης, που περιγράφεται παρακάτω [28].

Πιο σύνθετη διαδικασία είναι ο χωρισμός των δεδομένων με διασταυρωμένη επικύρωση στην οποία υπάρχει και πιο ειδική υλοποίηση για τον χωρισμό των χρονολογικών σειρών, η οποία αποτελεί μία παραλλαγή της “KFold” υλοποίησης της βιβλιοθήκης της sklearn. Ως υλοποίηση, παρέχει ευρετήρια για να χωρίσει τα δεδομένα εκπαίδευσης και εξέτασης τα οποία παρατηρούνται σε συγκεκριμένο χρονικό μεσοδιάστημα, σε κάθε χώρισμα τα ευρετήρια που ανήκουν στην εξέταση πρέπει να είναι υψηλότερα από πριν, συνεπώς αποφεύγεται το ανακάτεμα. Τέλος για αποτέλεσμα επιστρέφει τα πρώτα  $k$  «χωρίσματα» ως σετ εκπαίδευσης και το  $k + 1$  σετ, ως σετ εξέτασης [29].

### 3.4.2 Υπέρ – εκπαίδευση – Overfitting και Υπό – εκπαίδευση - Underfitting

Και τα δύο αυτά φαινόμενα αποτελούν έναν από τους λόγους μη καλής απόδοσης του μοντέλου της μηχανικής μάθησης.

Η υπέρ - εκπαίδευση προκαλείται όταν το μοντέλο πέραν των χαρακτηριστικών που μαθαίνει όταν εκπαιδεύεται, μαθαίνει επίσης και τον θόρυβο που περιέχουν τα δεδομένα εκπαίδευσης, αυτό θεωρητικά σημαίνει ως μεταβλητές όπως ο θόρυβος και η τυχαία διακύμανση θεωρούνται ως δεδομένα προς εκπαίδευση, αυτό έχει ως αποτέλεσμα να μην έχει το μοντέλο την δυνατότητα να

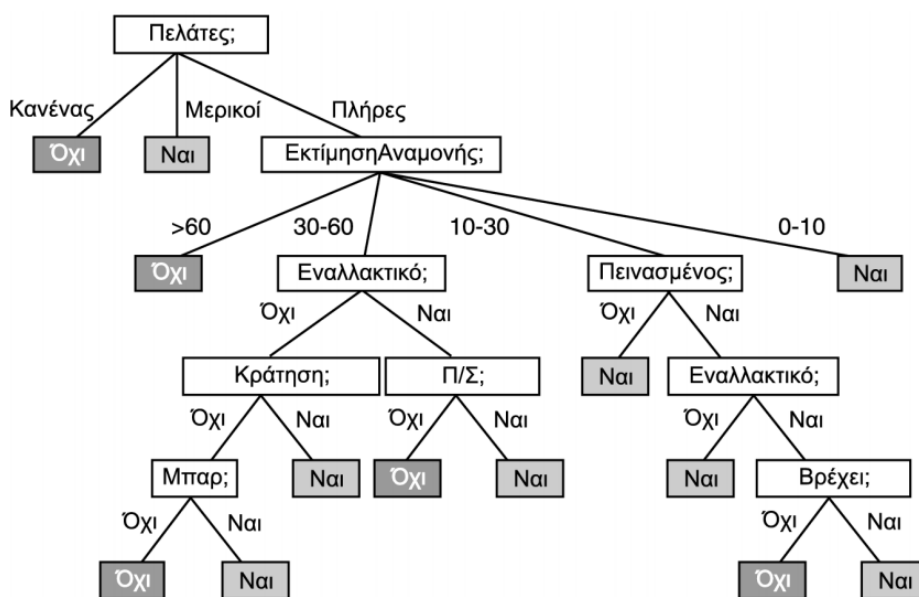
προβλέψει αποτελεσματικά όταν δέχεται καινούργια δεδομένα. Είναι πιθανότερη σε μη γραμμικά μοντέλα.

Η υπό – εκπαίδευση αναφέρεται στα μοντέλα τα οποία δεν μπορούν να μοντελοποιήσουν τα δεδομένα εκπαίδευσης αλλά ούτε και να γενικοποιήσουν νέα δεδομένα. [30]

### 3.5 Δένδρα Αποφάσεων

Τα δένδρα αποφάσεων προσπαθούν να προβλέψουν με όσο μεγαλύτερη ακρίβεια γίνεται την τιμή μεταβλητής που μοντελοποιούν με βάση τις τιμές των εισαγόμενων μεταβλητών. [24] Αποτελούν μία από τις ευρέως χρησιμοποιούμενες τεχνικές για κατηγοριοποίηση και πρόβλεψη.

Κάθε δένδρο αποτελείται από κόμβους, κάθε κόμβος είναι συνδεδεμένος με μία από τις μεταβλητές εισαγωγής, το κάθε φύλο αντιπροσωπεύει την τιμή η οποία προκύπτει από την μεταβλητή εισαγωγής. Ξεκινούν πάντα από την κορυφή και τελειώνουν σε φύλλο.



Εικόνα 6 - Παράδειγμα Δένδρου Δεδομένων [31].

Η δημιουργία ενός δένδρου απόφασης αποτελείται από:

- Κατασκευή δένδρου και
- Tree pruning

Ως πλεονεκτήματα έχουν το ότι είναι ευανάγνωστα όπως και επίσης μπορούν να εξηγηθούν εύκολα σε κάποιον τρίτο, μπορούν να διαχειριστούν εύκολα αριθμητικά δεδομένα και κατηγοριοποιημένα δεδομένα.



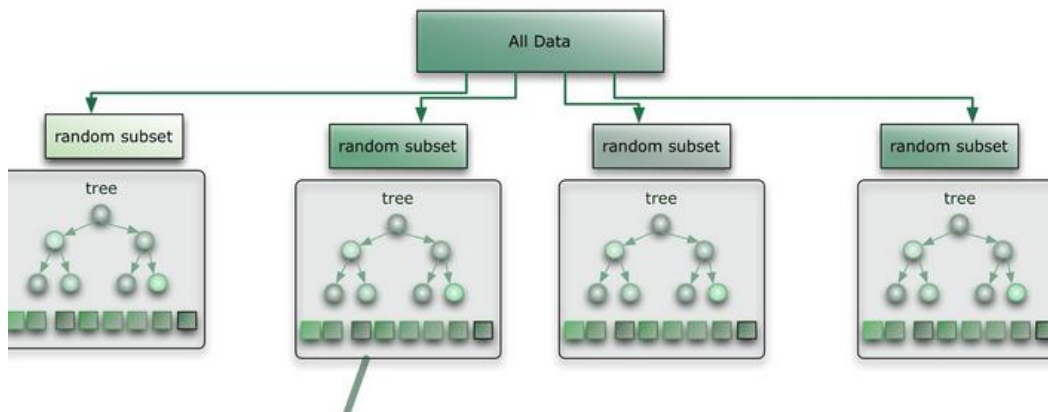
Στα μειονεκτήματά τους είναι το ότι δημιουργούν πολύπλοκα μοντέλα, μεταξύ άλλων. [23]

Οι πιο συνηθισμένοι αλγόριθμοι είναι:

- ID3
- C4.5
- CHAID
- MARS

### 3.5.1 Τυχαία Δάση (Random Forest)

Αποτελούνται από πολλά παράλληλα δέντρα απόφασης (decision trees) και συνδυάζουν τις προβλέψεις τους. Χρησιμοποιούνται τόσο για παλινδρόμηση όσο και για κατηγοριοποίηση. Συνήθως οι προβλέψεις του τυχαίου δάσους είναι καλύτερες των απλών δένδρων απόφασης [32].



Εικόνα 7 - Παράδειγμα τυχαίων δασών [32].

### 3.6 Πιθανοτικά Δίκτυα - Bayesian Networks

Τα Bayesian δίκτυα είναι τύπου πιθανολογικού γραφικού μοντέλου τα οποία μπορούν να χρησιμοποιηθούν να χτίσουν μοντέλα από δεδομένα, καθώς επίσης και συνδέουν ένα σύνολο μεταβλητών με σχέσεις πιθανοτήτων.

Ένα δίκτυο Bayes είναι μια αναπαράσταση μία δομής δεδομένων για πλήρεις συνδυασμένες κατανομές πιθανοτήτων, ένα δίκτυο Bayes για ένα σύνολο από μεταβλητές  $X = \{X_1, \dots, X_n\}$  αποτελείται από μία δομή δικτύου  $S$  η οποία υποδηλώνει το σύνολο των εξαρτήσεων μεταξύ των μεταβλητών  $X$  και από ένα σύνολο  $P$  από τοπικές κατανομές πιθανοτήτων συσχετιζόμενες με κάθε μία μεταβλητή.

Δοθέντος μιας δομής δικτύου  $S$  η συνδυασμένη κατανομή πιθανότητας για το  $X$  δίνεται από τον εξής τύπο:

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i \mid \text{parents}(X_i)).$$

Εικόνα 8 - Τύπος δικτύου Bayes

Έτσι, για μια συγκεκριμένη μεταβλητή  $X = \{X_1=A, \dots, X_n=B\}$  η πιθανότητά της ισούται με το γινόμενο της παραπάνω συνάρτησης για τις τιμές της  $X$ .

**Θεώρημα του Bayes:**

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

$P(A)$  → Πιθανότητα την οποία την γνωρίζουμε πριν την εκτέλεση του πειράματος

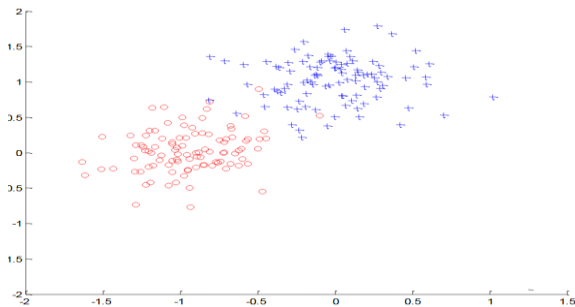
$P(A|B)$  → Δεσμευτική πιθανότητα του  $A$  δεδομένου του  $B$

$P(B|A)$  → Δεσμευτική πιθανότητα του  $B$  δεδομένου του  $A$

$P(B)$  → Παράγοντας εξομάλυνσης

### 3.7 Τετραγωνική Διακριτή Ανάλυση (QDA) – Γραμμική Διακριτή Ανάλυση (LDA)

Πρόκειται για δύο κλασικούς κατηγοριοποιητές οι οποίοι πραγματοποιούν τετραγωνική και γραμμική κατηγοριοποίηση αντίστοιχα. Η διακριτή ανάλυση γενικά είναι μια στατιστική τεχνική για να κατηγοριοποιεί αντικείμενα σε ομάδες, βασιζόμενο σε μετρήσεις χαρακτηριστικών κάποιου αντικειμένου. Αφορά κατηγοριοποίηση 2 κλάσεων για παραπάνω κλάσεις εφαρμόζεται ως πολλαπλή διακριτή ανάλυση, Multiple Discriminant Analysis (MDA) [33].



Εικόνα 9 - Διακριτή Ανάλυση [34].

- LDA → Ο LDA, κατηγοριοποιητής υπολογίζει διακριτά «σκορ» για κάθε παρατήρηση, ώστε να υπολογίσει σε ποια κλάση ανήκει η εκάστοτε μεταβλητή. Υποθέτει ότι οι παρατηρήσεις για κάθε κλάση «σχεδιάζονται» από μία πολύ μεταβλητή Γκαουσιανή κατανομή και πως οι συνδιακύμανση των μεταβλητών της πρόβλεψης είναι κοινές.
- QDA → Υποθέτει και αυτή πως οι παρατηρήσεις της κάθε κλάσης ακολουθούν Γκαουσιανή κατανομή, ωστόσο αντίθετα με το LDA υποθέτει πως κάθε κλάση περιέχει το δικό της πίνακα διακύμανσης

### 3.8 Μηχανές Διανυσμάτων Υποστήριξης (Support Vector Machines)

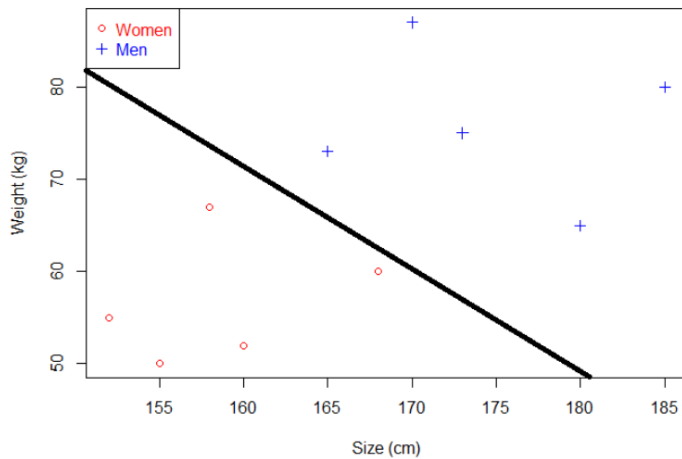
Προτάθηκε από τον Vladimir Vapnik και τους συνεργάτες του το 1992. Έχουν πλήθος εφαρμογών όπως:

- Αναγνώριση γραφής
- Ταξινόμηση κειμένων
- Αναγνώριση εικόνων
- Ιατρική
- Οικονομία

Είναι τεχνική για την ταξινόμηση τόσο γραμμικών όσο και μη γραμμικών δεδομένων και ανήκει στην μέθοδο επιβλεπόμενης μάθησης, επίσης στηρίζεται στα νευρωνικά δίκτυα τύπου Perceptron. Χρησιμοποιεί μη γραμμική απεικόνιση για να μετασχηματίσει τα αρχικά δεδομένα εκπαίδευσης σε υψηλότερες διαστάσεις, με την νέα διάσταση αναζητεί για το βέλτιστο υπέρ επίπεδο που διαχωρίζει τα δεδομένα. Με την κατάλληλη μη γραμμική απεικόνιση σε μια ικανοποιητική υψηλότερη διάσταση, δεδομένα δύο κατηγοριών μπορούν πάντα να διαχωρίζονται από ένα υπέρ επίπεδο. Το SVM βρίσκει το υπέρ επίπεδο χρησιμοποιώντας support vectors και όρια (margins), τα οποία ορίζονται από τους support vectors.

Η κατηγοριοποίηση μπορεί να είναι δυαδική (binary classification) όταν ως έξοδο μπορούμε να πάρουμε 2 τιμές, π.χ.: να θέλουμε να δούμε αν η τιμή μίας μετοχής θα είναι ανοδική ή όχι, ή παραπάνω (multiclass classification), π.χ.: αν ένα αντικείμενο είναι μήλο ή αχλάδι ή πορτοκάλι κοκ.

Υπάρχει ένα μη πεπερασμένο σύνολο γραμμών (hyperplanes) που χωρίζουν τις δύο κατηγορίες. Το SVM αναζητεί το υπέρ επίπεδο με το μεγαλύτερο margin – maximum marginal hyperplane – MMH [35].

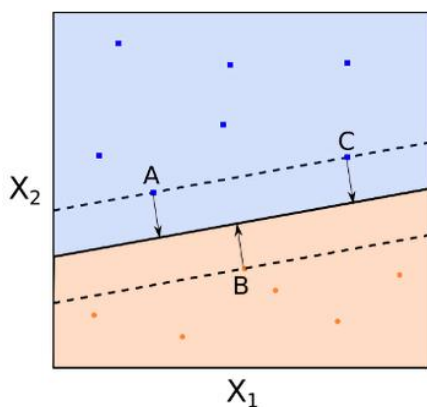


Εικόνα 10 - Κατηγοριοποίηση SVM [35].

Στην εικόνα 9 φαίνεται ένα παράδειγμα παλινδρόμησης γραμμικής, για κάθε σημείο στην γραφική παράσταση έχουμε τιμή για τον  $x$  άξονα και τον  $\psi$ . Η ευθεία  $\alpha x + \beta \psi + \gamma$  χωρίζει τα δεδομένα και στην ουσία τα κατηγοριοποιεί, οι τιμές  $\alpha$ ,  $\beta$ ,  $\gamma$  δίνονται από τον χρήστη και αλλάζουν μέχρι να επιτευχθεί μία «καθαρή ταξινόμηση». Τα SVM επεκτείνουν αυτήν την ταξινόμηση: ένα υπέρ επίπεδο διαχωρισμού μπορεί να γραφτεί ως:  $W \cdot X + b = 0$   $W \rightarrow \{w_1, w_2, \dots, w_n\}$  και  $\beta \rightarrow$  σταθερά.  $W_0 + W_1 \cdot X_1 + W_2 \cdot X_2 = 0$ , επίσης ορίζονται και οι πλευρές των ορίων:

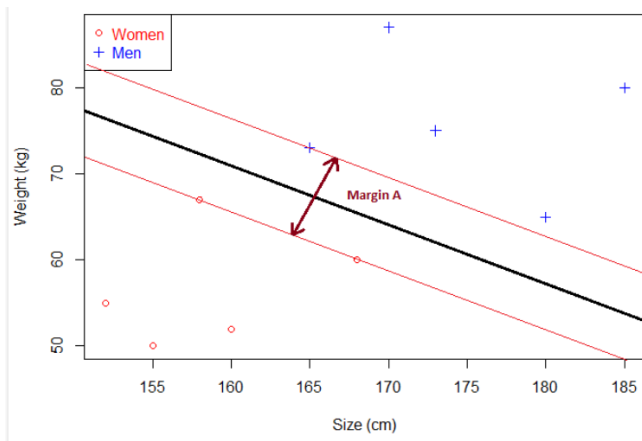
H1:  $W_0 + W_1 X_1 + W_2 X_2 \geq 1$  για  $y_i = +1$  και

H2:  $W_0 + W_1 X_1 + W_2 X_2 \leq -1$  για  $y_i = -1$ .

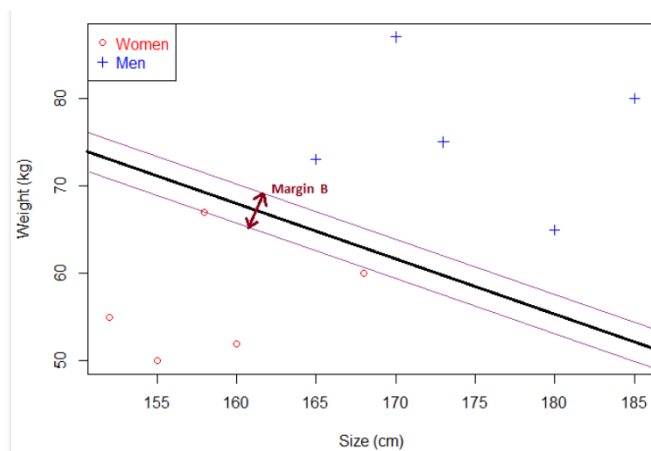


Εικόνα 11 - SVM hyperplane με διανύσματα υποστήριξης A, B, Γ [36]

Κάθε στοιχείο του συνόλου εκπαίδευσης που πέφτει πάνω στο υπέρ επίπεδο, δηλαδή τις πλευρές που ορίζονται από το margin είναι support vectors



Εικόνα 12 – SVM μεγάλο διάστημα [36].

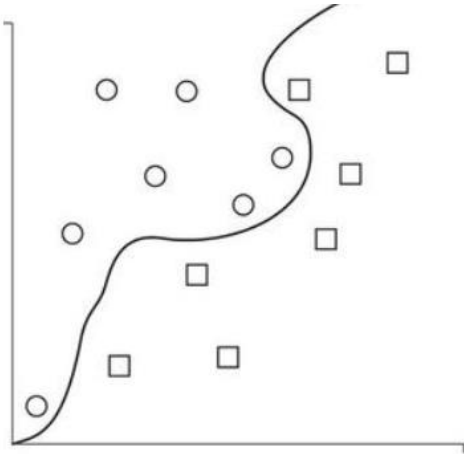


Εικόνα 13 - SVM μικρό διάστημα [36].

Εάν το hyperplane είναι αρκετά κοντά σε ένα σημείο τότε έχουμε μικρό κενό, αλλιώς αν είναι μακριά θα έχουμε μεγάλο. Όσο μεγαλύτερο το κενό μεταξύ των κατηγοριών, τότε τόσο καλύτερα αποτελέσματα θα έχουμε κατά τις προβλέψεις όπως επίσης θα έχουμε και ελάχιστο λάθος κατηγοριοποίησης για τα δεδομένα τα οποία δεν είναι σωστά [36].

Η πολυπλοκότητα του εκπαιδευόμενου κατηγοριοποιητή χαρακτηρίζεται από τον αριθμό των support vectors και όχι από τις διαστάσεις των δεδομένων.

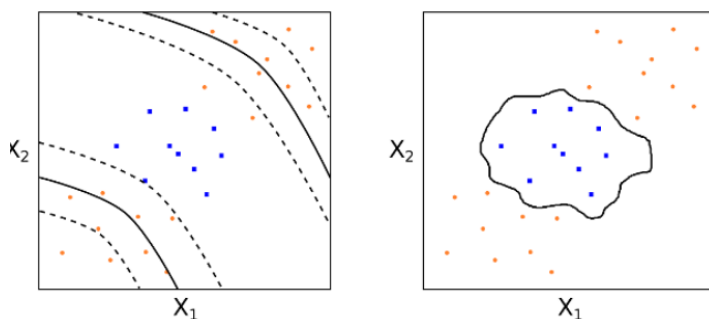
Ωστόσο τα δεδομένα δεν θα είναι τόσο ευδιάκριτα μεταξύ τους σχεδόν σε όλες τις περιπτώσεις τα δεδομένα της μίας κατηγορίας θα βρίσκονται στην μεριά της άλλης κατηγορίας όπως φαίνεται παρακάτω.



Εικόνα 14 - Μη γραμμική παλινδρόμηση

Πολλά μη γραμμικά μοντέλα μπορούν να μοντελοποιηθούν από μια πολυωνυμική συνάρτηση.

Σε αυτά τα μοντέλα χρησιμοποιείται μία τεχνική των SVM η οποία καλείται kernel trick. Οι kernels είναι συναρτήσεις οι οποίες πραγματοποιούν πολύπλοκους μετασχηματισμούς ώστε να επιτευχθεί η ταξινόμηση. Η μέθοδος αυτή αντικαθιστά την πράξη του εσωτερικού γινομένου που γίνεται για τα γραμμικά μοντέλα με τον εκάστοτε kernel function.



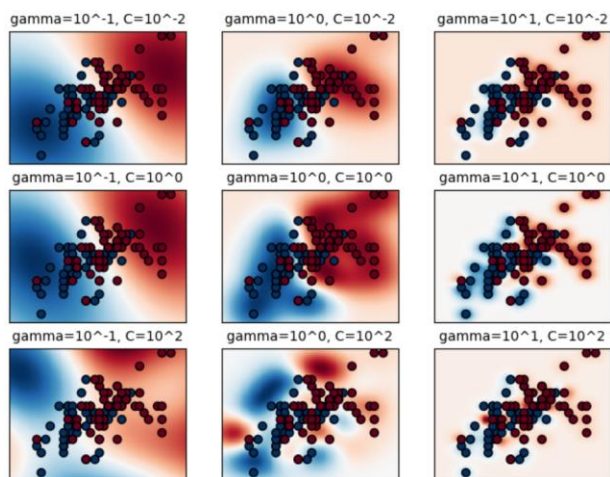
Εικόνα 15 - Αριστερά: πολυώνυμο  $n$  βαθμού, Δεξιά: radial kernel [37].

### 3.8.1 Παράμετρος C

Η παράμετρος C οποία είναι κατά ουσία η ισορροπία μεταξύ μαθησιακής ικανότητας και εκπαιδευόμενου λάθους. Για μεγάλες τιμές του C θα επιλεγεί ένα υπέρ – επίπεδο το οποίο θα παρουσιάζει μικρή ανοχή σε λανθασμένες κατηγοριοποιήσεις, το οποίο σημαίνει πως σε αντίθεση με μία μικρή τιμή του C θα έχουμε ακριβέστερη κατηγοριοποίηση, η επιλογή αυτής της παραμέτρου πάντως πρέπει να επιλεγεί αναλόγως με τα δεδομένα που έχουμε. Επίσης, πρέπει να σημειωθεί πως οι μεγάλες τιμές του C μπορούν να οδηγήσουν στο φαινόμενο της υπέρ – εκπαίδευσης.

### 3.8.2 Παράμετρος $\gamma$

Η παράμετρος  $\gamma$  η οποία αποτελεί την παράμετρο του πυρήνα και ορίζει πόση επιρροή πρόκειται να έχει ένα παράδειγμα εκπαίδευσης στο μοντέλο, υψηλότερο  $\gamma$  υψηλότερη επιρροή. Αυτό πρακτικά σημαίνει πως όταν έχω υψηλό  $\gamma$ , άρα και υψηλή επιρροή τότε τα σημεία τα οποία βρίσκονται πιο κοντά στην διαχωριστική γραμμή που δημιουργεί ο αλγόριθμος των ΜΔΥ λαμβάνονται περισσότερο υπόψιν, έχουν μεγαλύτερο βάρος, από τα απομακρυσμένα. [38]

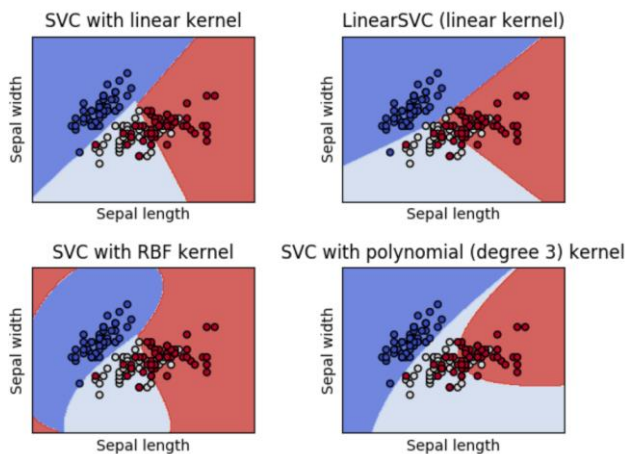


Εικόνα 16 - Παραδείγματα διαφορετικών παραμέτρων  $\gamma$  από την ιστοσελίδα της *sklearn*.

### 3.8.3 Συνάρτηση Κελύφους ΜΔΥ

Το κέλυφος μιας ΜΔΥ θα μπορούσε να είναι ένα από τα παρακάτω:

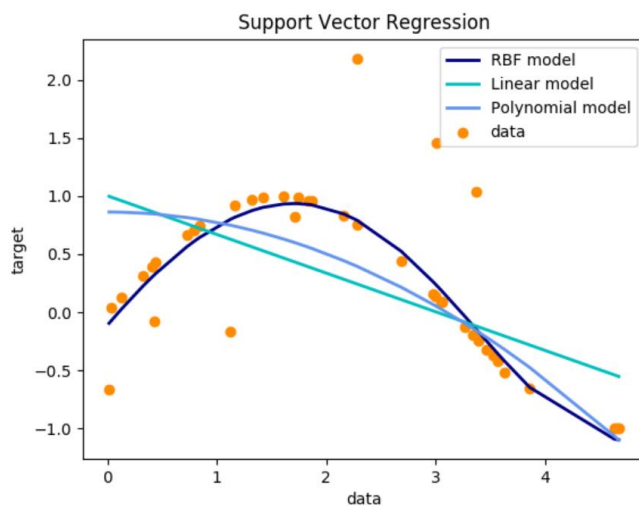
- Γραμμικό
- Πολυονομικό
- RBF
- Sigmoid



Εικόνα 17 - Πυρήνες SVM όπως παρουσιάζονται στην ιστοσελίδα της *sklearn*.

### 3.9 Παλινδρόμηση Διανυσμάτων Υποστήριξης (Support Vector Regression)

Πρόκειται για παραλλαγή των μηχανών διανυσμάτων υποστήριξης, είναι κατάλληλη για αριθμητικά χαρακτηριστικά. Εδώ αντί για προβλήματα κατηγοριοποίησης ο αλγόριθμος χρησιμοποιείται για προβλήματα παλινδρόμησης. Αποσκοπούν στο να οριστεί μία συνάρτηση  $f(x_i)$  της οποίας το αποτέλεσμα να μην αποκλίνει περισσότερο από μία ποσότητα  $\epsilon$  από τις πραγματικές τιμές  $y_i$ , για όλο το σύνολο της εκπαίδευσης.



Εικόνα 18 – Διάγραμμα υποστήριξης παλινδρόμησης με τα κελύφη του [39].

### 3.10 Συσταδοποίηση (Clustering)

Συσταδοποίηση είναι η ομαδοποίηση - συσταδοποίηση αντικειμένων τα οποία έχουν παρόμοια χαρακτηριστικά. Μία συστάδα θεωρείται καλή αν παράγει συστάδες καλής ποιότητας δηλαδή:



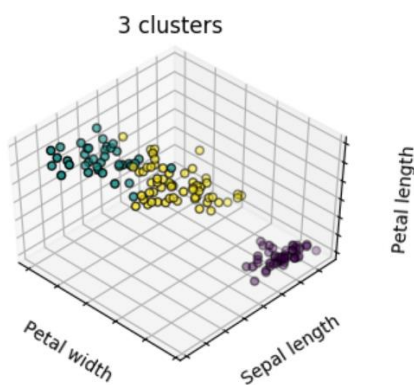
- Μεγάλη ομοιότητα εντός της συστάδας και
- Μικρή ομοιότητα ανάμεσα στις συστάδες.

Η ποιότητα εξαρτάται από την ομοιότητα και την μέθοδο υλοποίησης. Ανήκει στην μη – εποπτευόμενη μάθηση άρα και δεν χρειάζεται δεδομένα εκπαίδευσης για να εκπαιδευτεί.

Ο πιο γνωστός αλγόριθμος σε αυτήν την κατηγορία είναι k-means (διαχωριστικός αλγόριθμος). Ο αριθμός των συστάδων – ομάδων, K, εισέρχεται ως είσοδος στον αλγόριθμο, π.χ. πως έχω 1000 αντικείμενα και θέλω 4 συστάδες, κάθε μία από αυτές τις συστάδες έχει ένα centroid (το κεντρικό σημείο ενός αντικειμένου), το οποίο καλείται και mean, έτσι προκύπτει και το όνομα k-means. Άλλοι αλγόριθμοι είναι οι BFR και CURE.

Μερικές εφαρμογές της συσταδοποίησης είναι:

- Social Media Network analysis
- Market Research Companies
- Law Enforcement
- Computing

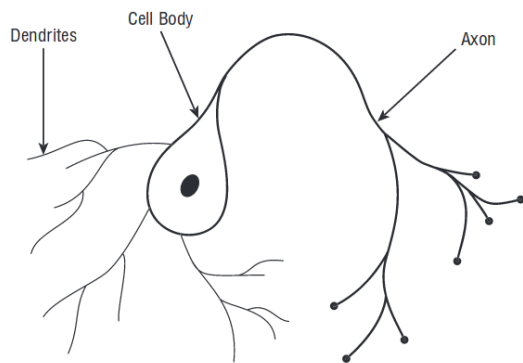


Εικόνα 19 -Εξοδος αλγορίθμου k -means της sklearn με τρεις συστάδες [40].

### 3.11 Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα είναι ουσιαστικά βασισμένα πάνω στην παράλληλη αρχιτεκτονική των, όχι απαραίτητα ανθρωπίνων, εγκεφάλων. Το δίκτυο είναι βασισμένο σε απλή μορφή εισόδων και εξόδων [23].

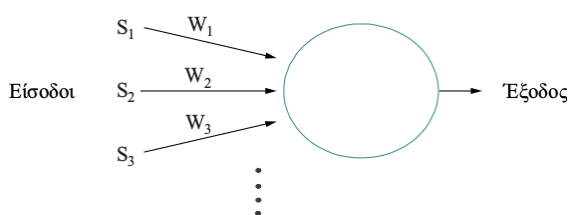
Σε βιολογικό επίπεδο ένα νευρώνας είναι ένα κελί το οποίο μπορεί να στείλει και να επεξεργαστεί ηλεκτρικά σήματα. Στον ανθρώπινο οργανισμό υπάρχουν δισεκατομμύρια νευρώνων οι οποίοι είναι συνδεδεμένοι μεταξύ τους, κάθε νευρώνας έχει μία είσοδο η οποία ονομάζεται δενδρίτης, το σώμα του κελιού και μία έξοδο η οποία ονομάζεται νευράξονας.



Εικόνα 20 -Δομή των Νευρώνων

Στα νευρωνικά δίκτυα η εκάστοτε εφαρμογή στηρίζεται στο ότι δεν ξέρουμε ακριβώς την σχέση μεταξύ εισόδου και εξόδου. Η βάση των νευρωνικών δικτύων είναι ο νευρώνας perceptron (ή αντίληπτρο) ο οποίος λαμβάνει ένα σήμα εισόδου και στην συνέχεια αφού αυτό επεξεργαστεί από μία συνάρτηση περνάει στην έξοδο, αυτή η συνάρτηση ονομάζεται activation function.

Αυτή η συνάρτηση είναι υπεύθυνη για την επεξεργασία που πραγματοποιείται αφού έχει εισαχθεί η τιμή εισόδου και ως έξοδο αποφασίζει αν η επεξεργασμένη τιμή θα περαστεί στον επόμενο νευρώνα στο δίκτυο.



Εικόνα 21 - Νευρωνικό δίκτυο πολλοί εισοδοι, μία έξοδος και τα αντίστοιχα βάρη [41].

Για την δημιουργία ενός νευρωνικού δικτύου προτιμάται η χρησιμοποίηση επιβλεπόμενης μάθησης. Τα νευρωνικά δίκτυα δουλεύουν μόνο με αριθμητικές τιμές, που σημαίνει πώς άλλου τύπου τιμές θα πρέπει να μετατραπούν.

Τα νευρωνικά δίκτυα συναντώνται σε πλήθος εφαρμογών όπως:

- Data Center Management

- High Frequency Trading
- Robotics
- Medical Monitoring
- Finance

## 4 Χρονολογικές Σειρές Γενικά (Time Series)

### 4.1 Ορισμός

Οι χρονολογικές σειρές είναι μία σειρά από παρατηρήσεις οι οποίες λαμβάνονται σε ορισμένες χρονικές στιγμές η περιόδους που απέχουν ίσα μεταξύ τους. Οι χρονικές στιγμές αν συμβολιστούν με  $X_i$  (έτη, μήνες, μέρες, ώρα, δευτερόλεπτα, κτλ.) και με  $Y_i$  οι τιμές των αντίστοιχων παρατηρήσεων, τότε δημιουργούνται ζεύγη της μορφής  $M(X_i, Y_i)$ , τα οποία μπορούν να παρασταθούν στο καρτεσιανό σύστημα αξόνων. Ενώνοντας αυτά τα σημεία δημιουργούμε ένα χρονοδιάγραμμα το οποίο θα μας δώσει την γενική εικόνα της εξέλιξης του υπό έρευνα φαινομένου ή χαρακτηριστικού.

Παραδείγματα τέτοιων χρονολογικών σειρών μπορούν να είναι ημερήσιες αφίξεις τουριστών σε μία χώρα, αριθμός πελατών σε πολυκατάστημα κατά μία χρονική στιγμή  $t$ , ημερήσια κατανάλωση ρεύματος, οικονομικές χρονολογικές σειρές, κτλ.

Για την ανάλυση των χρονολογικών σειρών συναντώνται δύο μοντέλα:

- Το προσθετικό: Το οποίο δίνεται από τον τύπο,  $Y_t = T_t + S_t + C_t + R_t$  και
- Το πολλαπλασιαστικό: Που δίνεται από τον τύπο,  $Y_t = T_t * S_t * C_t * R_t$

Όπου  $Y$ : η τιμή της μεταβλητής,  $T$ : η τάση,  $S$  η εποχιακή συνιστώσα,  $C$ : η κυκλική συνιστώσα,  $R$ : η τυχαία συνιστώσα [42].

### 4.2 Αιτιοκρατία Και Στοχαστικότητα

Όλες οι χρονολογικές σειρές από πραγματικά μεγέθη παρουσιάζουν θόρυβο, αυτές είναι και οι στοχαστικές χρονολογικές σειρές. Ο εντοπισμός αυτού του αιτιοκρατικού μέρους της σειράς είναι μεγάλη πρόκληση να πραγματοποιηθεί όταν αυτό εμπεριέχεται μέσα στον θόρυβο και τότε θεωρούμε πως το σύστημα είναι στοχαστικό [43].

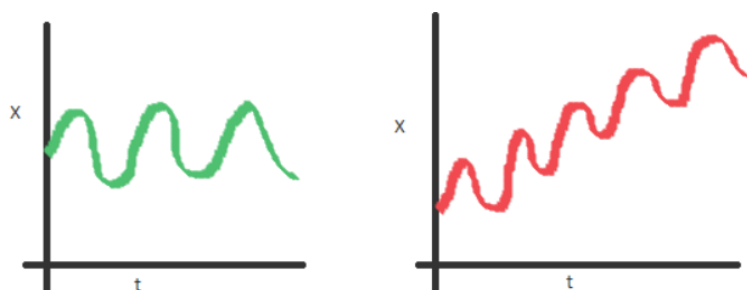
### 4.3 Στασιμότητα

Μία χρονολογική σειρά θεωρείται στάσιμη όταν σε αυτήν δεν έχουμε αλλαγή του μέσου όρου, συσχέτισης στο χρόνο και της διακύμανσης, αυτό σημαίνει ότι δεν πρέπει στην χρονολογική σειρά να παρουσιάζεται τάση. Γενικώς η μη στασιμότητα αποτελεί σοβαρό πρόβλημα για την ανάλυση και πρόβλεψη των χρονολογικών σειρών, οπότε αυτό σημαίνει πως θα πρέπει να απαλειφθεί αυτή η μη

στασιμότητα, η οποία μπορεί να προκαλείται από την τάση, την κυκλικότητα, κτλ. Υπάρχουν διάφορες τεχνικές για την αναγνώριση και απάλειψη της μη – στασιμότητας [44].

Στις χρονολογικές σειρές τέσσερις συνιστώσες διαδραματίζουν σημαντικό ρόλο στην ανάλυσή τους:

- Η τάση (trend)
- Η εποχικότητα (seasonality)
- Κυκλικότητα (Cyclical)
- Τυχαίες ή ακανόνιστες κινήσεις



Εικόνα 22 – Χρονολογική σειρά στάσιμη (αριστερά) και μη στάσιμη (δεξιά) [45].

#### 4.4 Διακύμανση

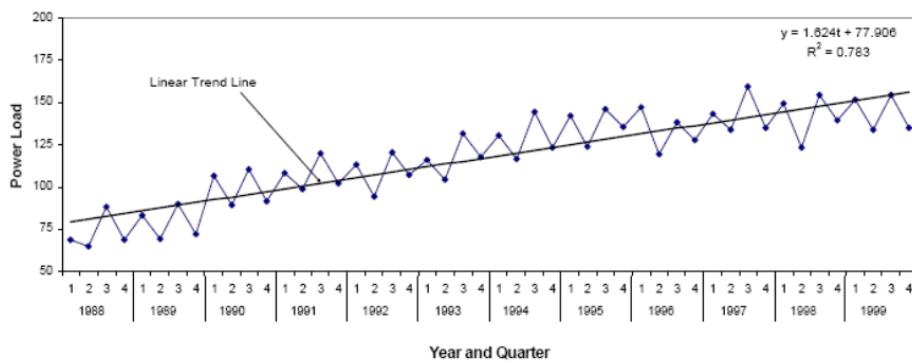
Η διακύμανση μετράει κατά πόσο απέχει κάθε τιμή από την μέση τιμή. Υπολογίζεται αφαιρώντας τον κάθε αριθμό στα δεδομένα μας με την μέση τιμή και στην συνέχεια τετραγωνίζοντας το αποτέλεσμα ώστε να είναι θετικό και στην συνέχεια διαιρούμε το άθροισμα των τετραγώνων με τον αριθμό των τιμών των δεδομένων μας.  $\sigma^2 = \frac{\sum(X-\mu)^2}{N}$ , όπου N το πλήθος των δεδομένων, X η εκάστοτε παρατήρηση και  $\mu$  η μέση τιμή. Φ.

#### 4.5 Τάση

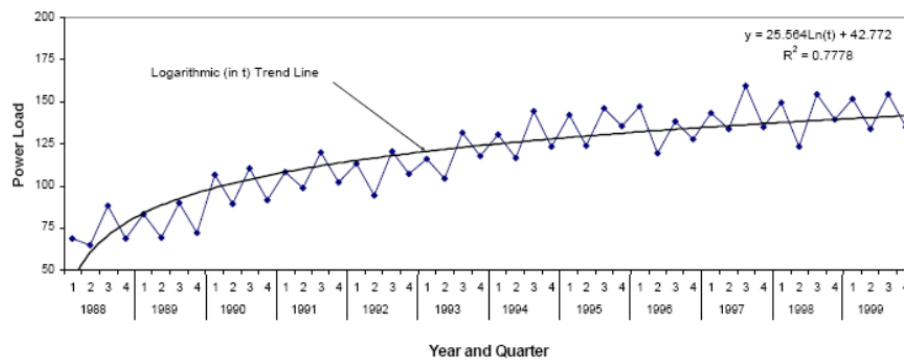
Η εμφάνιση τάσης ή περιοδικότητας στην χρονολογική σειρά υποδηλώνει ότι οι τιμές του μοντέλου αλλάζουν με τον χρόνο και δεν είναι στάσιμα για μεγάλο χρονικό διάστημα, η αλλαγή δεν είναι ανάγκη να είναι γραμμική. Σαν γενική εικόνα μπορούμε να δούμε αν η χρονολογική σειρά είναι ανοδική, πτωτική ή σταθερή. Σημαντικός παράγοντας ώστε μία τέτοια αλλαγή να θεωρηθεί τάση

είναι το ότι αυτή η αλλαγή δεν επαναλαμβάνεται και επίσης χρειαζόμαστε δεδομένα για μεγάλο διάστημα. Μπορούμε να μετρήσουμε την τάση με τις παρακάτω εξειδικεύσεις:

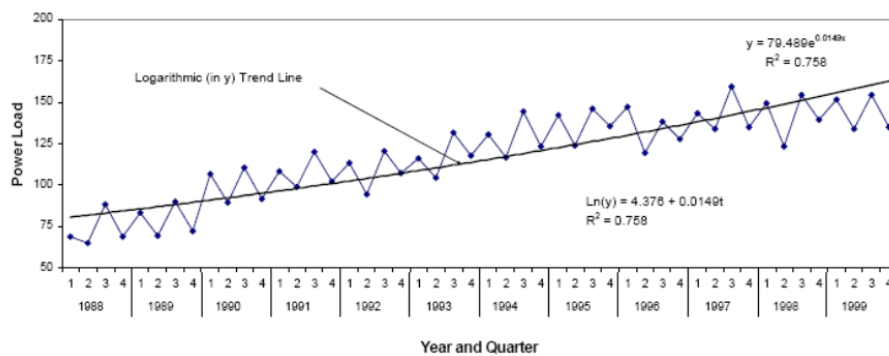
- Γραμμική  $\rightarrow Y_t = a + b t + e_t$
- Λογαριθμική  $\rightarrow Y_t = a + b \ln(t) + e_t$
- Εκθετική  $\rightarrow Y_t = \exp\{a + b t + e_t\}$
- Πολυωνυμική  $\rightarrow Y_t = a + b_1 t + b_2 t^2 + e_t$



Εικόνα 23 - Χρονολογική σειρά με γραμμική τάση [46].



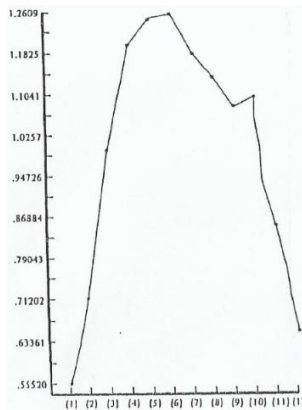
Εικόνα 24 - Χρονολογική σειρά με λογαριθμική τάση [46].



Εικόνα 25 - Χρονολογική σειρά με εκθετική τάση [46].

## 4.6 Εποχικότητα

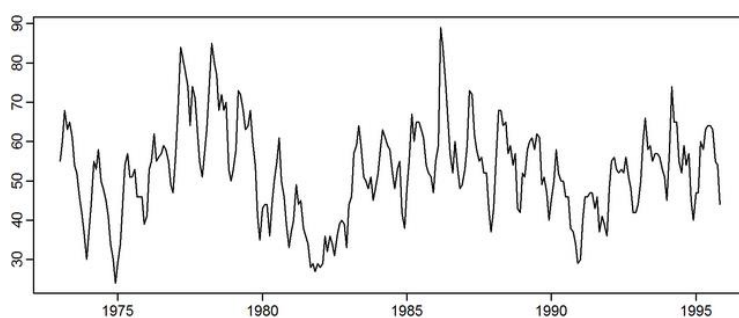
Με τον όρο εποχικότητα εννοούμε τις περιοδικές διακυμάνσεις που παρουσιάζει μία χρονολογική σειρά σε μια χρονική περίοδο, η διακύμανση αυτή επαναλαμβάνεται με τον ίδιο τρόπο. Ένα παράδειγμα μπορεί να είναι η μεγάλη αύξηση πωλήσεων πχ. Παγωτών κατά το καλοκαίρι, κτλ.



Εικόνα 26 - Παράδειγμα εποχικότητας [47].

## 4.7 Κυκλικότητα

Εμφανίζεται κατά περιόδους και οφείλεται σε ειδικές εξωγενείς συνθήκες οι περίοδοι αυτοί δεν είναι σταθερές και το μήκος τους είναι μεγαλύτερο του έτους, Συχνή κυκλικότητα παρουσιάζουν, οι χρονολογικές σειρές με δείκτες οικονομικούς, βιομηχανικούς, κτλ.



Εικόνα 27 - Παράδειγμα κυκλικότητας [48].

## 4.8 Τυχαίες – Ακανόνιστες κινήσεις

Αυτές αναφέρονται σε ακανόνιστες ή τυχαίες κινήσεις που ενδεχομένως πραγματοποιούνται σε μία χρονολογική σειρά λόγω τυχαίων γεγονότων και παραγόντων. Αυτά τα γεγονότα επηρεάζουν με ένα τρόπο ο οποίος δεν είναι συστηματικός και ούτε μπορεί να προσδιοριστεί. Κάποια παραδείγματα

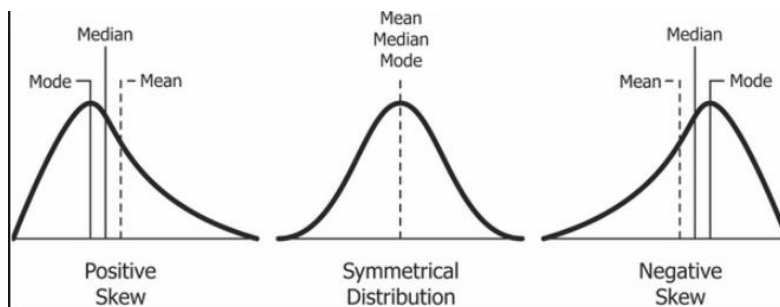
τέτοιων παραγόντων μπορεί πχ. Να είναι μία απεργία, μία αιφνίδια μεταβολή, ένα καιρικό φαινόμενο, κτλ.

#### 4.9 Λευκός θόρυβος

Αν θεωρήσουμε ως τυχαία τα διαδοχικά στοιχεία μιας χρονολογικής σειράς, τότε αυτή η χρονολογική σειρά λέγεται ότι αποτελείται από ανεξάρτητες τυχαίες μεταβλητές με ίδια κατανομή, για  $x_t, x_{t+1}, x_{(t+\tau)}$ , για  $\tau > 1$ . Μια iid χρονολογική σειρά είναι εντελώς τυχαία και δεν περιέχει αυτό – συσχετίσεις, γραμμικές ή μη – γραμμικές. Μία τέτοια χρονολογική σειρά, iid, ονομάζεται και λευκός θόρυβος (white noise) και η κατανομή της συμβολίζεται ως  $WN(0, \sigma_\varepsilon^2)$ , με μέση τιμή 0 και διασπορά  $\sigma_\varepsilon^2$ , αν αυτός ο θόρυβος επίσης, ακολουθεί γκαουσιανή κατανομή, τότε λέγεται γκαουσιανός θόρυβος.

#### 4.10 Λοξότητα (Skewness)

Η λοξότητα στατιστικά χρησιμοποιείται για να περιγράψει την ασυμμετρία από μία κανονική κατανομή. Μπορεί να είναι είτε θετική είτε αρνητική, αυτό εξαρτάται από το αν τα δεδομένα λοξεύουν πιο πολύ προς τα αριστερά, δηλαδή αρνητικά, ή προς τα δεξιά, δηλαδή θετικά. Στα χρηματοοικονομικά δεδομένα η λοξότητα μπορεί να δώσει πολλά στοιχεία σε έναν επενδυτή ο οποίος μπορεί να υπολογίσει για την λοξότητα των δεδομένων σε σχέση με τον μέσο όρο. [49]



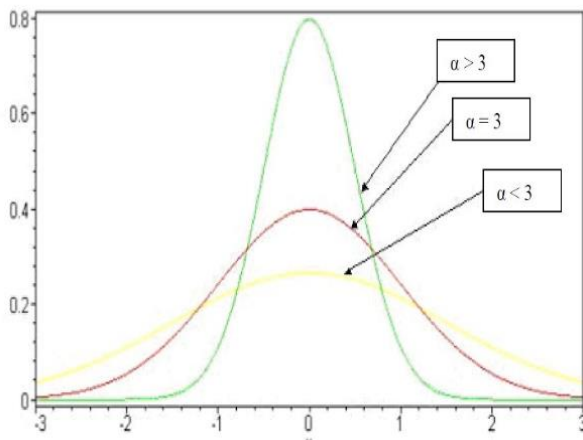
Εικόνα 28 - Παράδειγμα θετικής και αρνητικής λοξότητας [50].

#### 4.11 Κύρτωση

Η κύρτωση είναι και αυτή ένα στατιστικό μέτρο το οποίο χρησιμοποιείται για να περιγράψει την κατανομή ή και την λοξότητα που είδαμε παραπάνω, μιας παρατήρησης, γύρω από το μέσο όρο. Χρησιμοποιείται συνήθως για την περιγραφή τάσεων. Η κύρτωση υπάρχει σε γραφικές όπου



υπάρχουν «παχιές ουρές» και χαμηλής κατανομής, όπως και σε «λεπτές ουρές» και κατανομή γύρω από τον μέσο όρο. [51]



Εικόνα 29 - Παράδειγμα κύρτωσης.  $\alpha = 3$  κανονική κατανομή,  $\alpha < 3$  λεπτό - κυρτή,  $\alpha > 3$  πλατύ - κυρτή [52].

## 4.12 Συσχέτιση

### 4.12.1 Γραμμική Αυτό – Συσχέτιση Και Μερική Αυτό - Συσχέτιση

Η αυτό - συσχέτιση  $\rho_\tau$  για κάποια υστέρηση  $\tau$  είναι ο συντελεστής δύο στοιχείων της χρονοσειράς που απέχουν χρονικά  $\tau$  βήματα και υπολογίζεται ως:

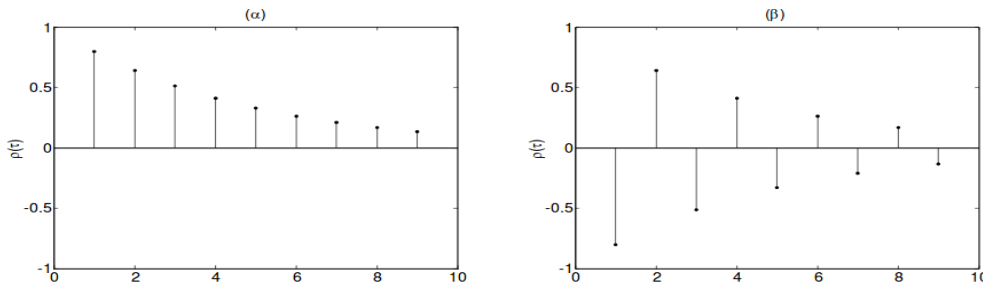
$$\rho_\tau = r_\tau = \text{Corr}(x_t, x_{t-\tau}) = \frac{\sum_{t=\tau+1}^n (x_t - \bar{x})(x_{t-\tau} - \bar{x})}{\sum_{t=\tau+1}^n (x_t - \bar{x})^2}$$

Αν υπολογιστεί η συνάρτηση αυτό – συσχέτισης ως προς την υστέρηση και επίσης όταν η χρονολογική σειρά δεν είναι στάσιμη, παρατηρούμε πως έχει πολύ υψηλές τιμές και φθίνει πολύ αργά με την υστέρηση, αυτό οφείλετε στην τάση η οποία προκαλεί ισχυρές συσχετίσεις μεταξύ κοντινών χρονικών σημείων [53].

Η συνάρτηση μερικής αυτό – συσχέτισης ορίζεται ανάμεσα στη  $Y_t$  και στην  $Y_{t-s}$  και αναφέρεται στην συσχέτιση ανάμεσά τους, όταν έχουν αφαιρεθεί οι γραμμικές επιδράσεις των ενδιάμεσων μεταβλητών,  $Y_{t-1}, Y_{t-2}, \dots$

Επίσης οι παραπάνω αυτό – συσχετίσεις αποτελούν και εργαλείο εύρεσης στασιμότητας μίας χρονολογικής σειράς, αυτή δεν είναι στάσιμη όταν οι τιμές φθίνουν γρήγορα προς το μηδέν, αν δεν φθίνουν γρήγορα υπάρχει πάλι πιθανότητα να μην είναι στάσιμη [54].

Στον  $\chi$  άξονα παρουσιάζονται οι υστερήσεις και στον  $\psi$  άξονα φαίνεται η σχέση του συντελεστή συσχέτισης, μεταξύ  $-1$  για αρνητική συσχέτιση και  $1$  για θετική. Η αυτό – συσχέτιση της χρονολογικής σειράς που περιγράφεται από στάσιμο αυτό – παλίνδρομο μοντέλο φθίνει εκθετικά στο  $0$  και ισχύει:  $\gamma_0 = \sigma_\chi^2 = \frac{\sigma^2}{1-\phi^2}$  και  $\rho_\tau = \phi^\tau$ , όπως φαίνεται και στις παρακάτω γραφικές παραστάσεις, για  $\phi > 0$  η υστερήσεις φθίνουν μονότονα προς το  $0$  ενώ για  $\phi < 0$  οι τιμές εναλλάσσονται γύρω από το  $0$ .



Εικόνα 30 - Γράφημα αυτό - συσχέτισης μοντέλου αυτό - παλινδρόμησης για συντελεστή  $\phi$  θετικό αριστερά και  $\phi$  αρνητικό δεξιά [55].

Όσο και να αυξήσουμε την τάξη του αυτό – παλίνδρομου μοντέλου, η τιμή των υστερήσεων πάντα θα φθίνει προς το  $0$  [55].

Στο γράφημα της μερικής αυτό – συσχέτισης από την άλλη παρουσιάζονται οι αυτό - συσχετίσεις για μία παρατήρηση με υστερημένες τιμές οι οποίες δεν έχουν υπολογιστεί από προγενέστερες παρατηρήσεις.

#### 4.13 Πρόβλεψη Χρονολογικών Σειρών

Η πρόβλεψη των χρονολογικών σειρών χωρίζεται στις παρακάτω κατηγορίες μεθόδου πρόβλεψων [56]:

- Μέθοδοι εξομάλυνσης
- Διάσπασης
- Ανάλυση ARIMA

Υπάρχουν επίσης κριτήρια αξιολόγησης για την επιλογή της κατάλληλης μεθόδου πρόβλεψης όπως:

- ❖ Μέση απόλυτη απόκλιση MAD (Mean Absolute Deviation):
  - Υπολογισμός απόλυτης διαφοράς μεταξύ προβλέψεων και πραγματικών τιμών
  - Εύρεση μέσης τιμής της διασποράς

- ❖ Μέσο σφάλμα τετραγώνου MSE (Mean Squared Error):
  - Υπολογισμός του τετραγώνου της διαφοράς μεταξύ προβλέψεων και πραγματικών τιμών
  - Εύρεση μέσης τιμής της διαφοράς
- ❖ Μέσο απόλυτο σφάλμα (Mean Absolute Error - MAE):
  - $MAE = \frac{1}{N} \sum_{t=1}^M |Y_t^f - Y_t^a|$
- ❖ Ρίζα του μέσου του τετραγώνου του σφάλματος (Root Mean Square Error - RMSE):
  - $RMSE = \sqrt{\frac{\sum_{t=1}^M (e_t^f)^2}{N}}$ , όπου  $Y_t^f$  είναι η προβλεπόμενη τιμή,  $Y_t^a$  είναι η παρατηρούμενη τιμή και N ο αριθμός των χρονικών περιόδων.
- ❖ Μέσο απόλυτο ποσοστιαίο σφάλμα MAPE (Mean Absolute Percentage Error):
  - $MAPE = \frac{1}{N} \sum_{t=1}^M \left| \frac{Y_t^f - Y_t^a}{Y_t^a} \right|$
- ❖ Μέσο ποσοστιαίο σφάλμα MPE (Mean Percentage Error)

#### 4.13.1 Πρόβλεψη Χρονολογικών Σειρών Με Μεθόδους Εξομάλυνσης

Αυτές οι μέθοδοι χρησιμοποιούνται για να εξομαλύνουν – στρογγυλοποιήσουν τις μετρήσεις μιας χρονολογικής σειράς, αυτές οι μέθοδοι χρησιμοποιούνται σε χρονολογικές σειρές οι οποίες δεν παρουσιάζουν τάση, κυκλικότητα και εποχικότητα, αν έχουν κάτι από τα παραπάνω θα πρέπει να εξομαλυνθούν. Αν μία χρονολογική σειρά είναι στάσιμη τότε κατάλληλη μέθοδος πρόβλεψης είναι η μέθοδος των κινητών μέσων, αν κάποια χρονολογική σειρά εμφανίζει κάποιο πρότυπο τάσης τότε χρησιμοποιούμε την μέθοδο διπλής εκθετικής εξομάλυνσης, ή την μέθοδο Brown, ή την μέθοδο Holt, αν η χρονολογική σειρά εμφανίζει εποχικότητα η πιο κατάλληλη είναι η winters. Οι πιο συνήθεις πρακτικές είναι:

- Μέθοδος κινητών μέσων (Moving average method)
- Μέθοδος Σταθμικός μέσος όρος (Weighted moving average method)
- Μέθοδος εκθετικής εξομάλυνσης (Exponential smoothing method)

#### 4.13.1.1 Μέθοδος Κινητών Μέσων

Είναι η βασικότερη μέθοδος πρόβλεψης και χρησιμοποιείται όταν η ζήτηση δεν παρουσιάζει διακυμάνσεις και η εποχικότητα δεν λαμβάνεται υπόψιν. Υπολογίζεται ο μέσος όρος της ζήτησης για ένα συγκεκριμένο αριθμό περιόδων  $t$ , αυτός ο μέσος όρος χρησιμοποιείται για την πρόβλεψη της ζήτησης της επόμενης περιόδου  $t + 1$ . Στην συνέχεια υπολογίζεται ο μέσος όρος της ζήτησης των προηγούμενων  $t$  περιόδων συνυπολογίζοντας σε αυτές την πρόβλεψη της προηγούμενης περιόδου, έτσι προστίθεται μια νέα πρόβλεψη και αφαιρείται η παλαιότερη, για αυτό και ονομάζεται η μέθοδος κινητών μέσων.  $F_{t+1} = \frac{(D_t + D_{t-1} + \dots + D_{t-n+1})}{n}$

Εβδομάδα	Παρατηρήσεις Ζήτησης	Πρόβλεψη
1	$D_1$	
2	$D_2$	
3	$D_3$	
4	$D_4$	
5	$D_5$	
6	$D_6$	
7	?	$F_7 = \frac{D_3 + D_4 + D_5 + D_6}{4}$
8	$D_8$	$F_8$
9	$D_9$	$F_9$
10	$D_{10}$	$F_{10}$
11	$D_{11}$	$F_{11}$
12	$D_{12}$	$F_{12}$

Εικόνα 31 - Παράδειγμα εφαρμογής κινητών μέσων [57].

Οι πιο σύνηθες χρονικές περίοδοι που χρησιμοποιούνται για τους κινητούς μέσους στα χρηματοοικονομικά δεδομένα είναι 15, 20, 30, 50, 100 και 200 μέρες. Όσο πιο μικρό χρονικό παράθυρο επιλέξουμε τόσο πιο ευαίσθητη θα είναι η μέθοδος στις μεταβολές των τιμών, όσο μεγαλύτερο όμως, τότε θα είναι λιγότερο ευαίσθητο και πιο εξομαλυμένο. Η επιλογή για το σωστό χρονικό παράθυρο μπορεί να γίνει αφού δοκιμαστούν διάφορα παράθυρα και στην συνέχεια επιλεγεί αυτό το οποίο ταιριάζει στην οικονομική στρατηγική που θέλουμε να ακολουθήσουμε.

#### 4.13.1.2 Μέθοδος Σταθμικού Κινητού Μέσου Όρου

Είναι παρόμοια μέθοδος με την μέθοδο του κινητού μέσου όρου, με την διαφορά ότι οι τιμές που συμμετέχουν στον υπολογισμό του μέσου όρου πλέον έχουν διαφορετική βαρύτητα. Δίνεται μεγαλύτερη βαρύτητα στις τιμές οι οποίες είναι πιο κοντά στην υπό εξέταση περίοδο πρόβλεψης και χαμηλότερη βαρύτητα στην προγενέστερη περίοδο, αυτό σημαίνει πως οι συντελεστές βαρύτητας προοδευτικά μειώνονται όσο πιο παλιές είναι οι τιμές και το αντίστροφο.

$F_{t+1} = w_1 D_1 + w_2 D_2 + \dots + w_t D_t$ , όπου  $w_t$  οι συντελεστές βαρύτητας κάθε περιόδου  $t$ , το άθροισμα των οποίων ισούται με 1.

#### 4.13.1.3 Εκθετική Μέθοδος

Η εκθετική μέθοδος θεωρεί πως οι πιο πρόσφατες τιμές της χρονολογικής σειράς έχουν μεγαλύτερη βαρύτητα από εκείνες των παλαιότερων περιόδων υποθέτοντας ότι οι συντελεστές βαρύτητας ακολουθούν μια εκθετική κατανομή. Σε σχέση με την μέθοδο του σταθμικού μέσου αυτή απαιτεί λιγότερα ιστορικά δεδομένα. Για τον υπολογισμό της ζήτησης της επόμενης περιόδου, συνυπολογίζονται μόνο η πραγματική ζήτηση της προηγούμενης περιόδου και η αντίστοιχη πρόβλεψη.  $F_{(t+1)} = F_t + a(D_t - F_t)$ , όπου  $D_t$  η πραγματική ζήτηση προηγούμενης περιόδου,  $a$  συντελεστής από 0 έως 1.

	Στάθμιση με $\alpha = 0.05$
Πιο πρόσφατη παρατήρηση = $a(1 - \alpha)^0$	0.0500
Δύο περιόδους πριν = $a(1 - \alpha)^1$	0.0475
Πριν τρεις περιόδους = $a(1 - \alpha)^2$	0.0451
Πριν τέσσερις περιόδους = $a(1 - \alpha)^3$	0.0429

Εικόνα 32 - Παράδειγμα εφαρμογής εκθετικής μεθόδου, παρατηρούμε πως η τιμή από τα βάρη μειώνεται όσο πιο παλιά είναι η παρατήρηση [57].

#### 4.13.1.4 Μέθοδος Holt

Πρόκειται για εκθετική εξομάλυνση με προσαρμογή στην τάση. Χρησιμοποιείται όταν υπάρχει τάση στην χρονολογική σειρά, ως μέθοδος έχει δύο παραμέτρους εξομάλυνσης, την σταθερή παράμετρο  $\alpha$  για τις τιμές της χρονολογική σειράς η οποία δέχεται τιμές από 0 έως και 1 και την επίσης σταθερή παράμετρο  $\beta$ , που δέχεται τιμές από 0 έως και 1 για την εξομάλυνση της τάσης.

Για την εξομάλυνση των τιμών της χρονολογικής σειράς έχουμε την σχέση:

$A_t = \alpha Y_t + (1 - \alpha)(A_{t-1} + T_{t-1})$ , όπου  $A_t$  οι εξομαλυνθείσες τιμές της χρονολογικής σειράς για  $t > 2$  και για  $t = 1$  ισχύει  $A_1 = Y_1$ .

Για την εξομάλυνση της τάσης έχουμε την σχέση:  $T_t = \beta(A_t - A_{t-1}) + (1 + \beta)T_{t-1}$  όπου  $T_t$  οι εξομαλυνθείσες τιμές της τάσης, για  $t > 2$  και για  $t = 1$  ισχύει  $T_1 = 0$ .

Η πρόβλεψη εν τέλει προσδιορίζεται ως:  $Y_{t+h} = A_t + hT_t$ , όπου  $h = 1, 2, \dots$

#### 4.13.2 Πρόβλεψη Χρονολογικών Σειρών Με διάσπαση

Οι κινήσεις μίας χρονολογικής σειράς όπως είπαμε διακρίνονται σε τέσσερα είδη τα οποία ονομάζονται συνιστώσες της χρονολογικής σειράς, αυτές οι συνιστώσες είναι οι κυκλικές κινήσεις, οι

ακανόνιστες μεταβολές (τυχαίες), η εποχικότητα και οι μακροπρόθεσμες κινήσεις. Για να προβλέψουμε μία χρονολογική σειρά με την μέθοδο αυτήν, αυτή διασπάται στις τέσσερις αυτές συνιστώσες και προσδιορίζεται η επιρροή της κάθε μιας στην τιμή της μεταβλητής

#### 4.13.2.1 Ανάλυση Εποχικότητας

Όταν στην χρονολογική σειρά εμφανίζεται κάποια εποχικότητα τότε αυτή πρέπει να αναλυθεί και να μετρηθεί με δείκτες εποχικότητας (seasonal indices) οι οποίοι θα ανιχνεύσουν τον τρόπο συμπεριφοράς της χρονολογική σειράς όπου υπάρχει το εποχικό αυτό φαινόμενο. Με την βοήθεια του προσδιορισμού αυτών των δεικτών μπορούμε να απαλλαγούμε από το φαινόμενο της εποχικότητας. Αυτοί οι δείκτες προσδιορίζονται με την εφαρμογή της μεθόδου του κεντρικού κινητού μέσου (centered moving average), μέσω αυτής της μεθόδου μπορεί να απομονωθεί η εποχικότητα από την τάση, την κυκλικότητα και την μη – κανονικότητα. Σε περίπτωση που έχουμε πολλαπλασιαστικό μοντέλο ο δείκτης εποχικότητας  $S_t$  για  $t = 1, 2, \dots, n$  προσδιορίζεται από την ακόλουθη σχέση:  $S_t = \frac{Y_t}{CA_t} = \frac{(T_t S_t C_t I_t)}{T_t C_t I_t}$ , όπου  $CA_t$  είναι η εξομαλυνθείσα τιμή της χρονολογικής σειράς που προέρχεται από την μέθοδο του κεντρικού κινητού μέσου. Θα πρέπει το άθροισμα των εποχικών δεικτών να είναι ίσο με τον αριθμό περιόδων εντός του έτους, αν αυτό δεν ισχύει τότε θα πρέπει να γίνει η κατάλληλη προσαρμογή με τον εξής τύπο:  $SAY_t = \frac{Y_t}{SA_t}$ .

#### 4.13.2.2 Ανάλυση Μακροχρόνιας Τάσης

Με τις σχέσεις:  $\beta = \frac{n \sum_{t=1}^n t Y_t - (\sum_{t=1}^n t)(\sum_{t=1}^n Y_t)}{n \sum_{t=1}^n t^2}$  και  $a = \frac{1}{n} \sum_{t=1}^n Y_t - \beta \frac{1}{n} \sum_{t=1}^n t$ , όπου  $t$  ανεξάρτητη μεταβλητή,  $\epsilon$  τυχαίο σφάλμα,  $\beta$  πόσο έχει μεταβληθεί η τιμή της χρονολογική σειράς όταν ο χρόνος  $t$  έχει μεταβληθεί κατά μία μονάδα, αν το  $\beta$  είναι θετικό τότε έχουμε ανοδική τάση αλλιώς πτωτική.

#### 4.13.2.3 Ανάλυση Κυκλικότητας Και Μη Κανονικότητας

Αφού πρώτα έχει απομονωθεί η κυκλικότητα από την τάση και την εποχικότητα έχουμε τον τύπο:  $TAY_t = \frac{SAY_t}{T_t} = C_t I_t$  όπου  $TAY_t$  είναι οι τιμές της χρονολογικής σειράς οι οποίες είναι απαλλαγμένες από εποχικότητα και από την τάση. Ο σταθμικός κεντρικός κινητός μέσος χρησιμοποιείται και εδώ για να απομονώσουμε την κυκλικότητα από τη μη – κανονικότητα,  $WA_t = \frac{Y_{t-1} + 2Y_t + Y_{t+1}}{4}$ .

Για να απομονώσουμε την μη – κανονικότητα με τον τύπο:  $CAY_t = \frac{TAY_t}{WA_t} = I_t$ , όπου  $CAY_t$ , είναι οι τιμές της χρονοσειράς απαλλαγμένες από κυκλικότητα, περιέχοντας μόνο το στοιχείο της μη κανονικότητας.

#### 4.13.3 Ανάλυση ARIMA

Τα αρχικά σημαίνουν Auto Regressive Integrated Moving Average και γράφεται συνήθως ως ARIMA(p, d, q), το p συμβολίζει τους όρους του αυτό – παλίνδρομου υποδείγματος, το q τους όρους του υποδείγματος των κινητών μέσων και το d την τάξη της ολοκληρωμένης σειράς. Αυτά τα υποδείγματα είναι στοχαστικά μαθηματικά μοντέλα και χρησιμοποιούνται για την περιγραφή της διαχρονικής εξέλιξης κάποιου φυσικού μεγέθους, περιλαμβάνουν το τυχαίο σφάλμα ή σφάλμα πρόβλεψης. Χρησιμοποιούνται όταν το μοντέλο δεν είναι στάσιμο. Επίσης, χρησιμοποιούνται ευρύτατα διότι αναλύουν πολλούς παράγοντες και έχουν καλά αποτελέσματα στην πρόβλεψη μελλοντικών τιμών της χρονολογικής σειράς. Αυτή η διαδικασία μπορεί να είναι γνωστή και ως Box – Jenkins διότι αυτοί πρότειναν πρώτοι μία ολοκληρωμένη διαδικασία εκτίμησης γραμμικών μοντέλων. Η διαδικασία Box – Jenkins έχει ως εξής:

- Ταυτοποίηση: η οποία περιλαμβάνει τον καθορισμό των τιμών p, d και q. Πρώτα υπολογίζεται το d και στην συνέχεια η τάξη της αυτό – παλίνδρομης διαδικασίας (p) και η τάξη της διαδικασίας του κινητού μέσου (q), πραγματοποιείται επίσης και έλεγχος στασιμότητας μέσω του ελέγχου της μοναδιαίας ρίζας (Dickey Fuller) είτε των γραφημάτων των αυτό - συσχετίσεων.
- Εκτίμηση και διαγνωστικός έλεγχος: Ο έλεγχος αφορά τον στατιστικό έλεγχο των σφαλμάτων, θα πρέπει τα δεδομένα του ελέγχου να ακολουθούν την διαδικασία του λευκού θορύβου, δηλαδή να μην αυτό – συσχετίζονται.
- Προβλέψεις: Διενεργείται πρόβλεψη σε βραχυχρόνια περίοδο.

Όταν η χρονολογική σειρά δεν είναι στάσιμη, τότε αυτή πρέπει να μετατραπεί σε τέτοια, υλοποιώντας την διαδικασία των διαφορών. Παίρνοντας τις πρώτες διαφορές, λέμε πως η σειρά είναι ολοκληρωμένη πρώτης τάξης (Integrated First Order), με τις δεύτερες διαφορές είναι ολοκληρωμένη δεύτερης τάξης. Γενικώς, αν μετατρέψουμε μία χρονολογική σειρά σε στάσιμη παίρνοντας τις d διαφορές τότε αυτή είναι ολοκληρωμένη d τάξης και συμβολίζεται με I(d). [56]

#### 4.13.3.1 Έλεγχος Dickey – Fuller

Αυτός ο έλεγχος πραγματοποιείται ώστε να δούμε αν κάποια χρονολογική σειρά είναι στάσιμη, χρησιμοποιεί έλεγχο μοναδιαίας ρίζας (unit root test) με τον οποίο όρο εννοούμε ότι κάποια ρίζα του πολυωνύμου  $f(x) = 1 - \rho_1 x - \rho_2 x^2 - \rho_3 x^3 - \dots - \rho_n x^n = 0$ , ισούται με την μονάδα. Πιο συγκεκριμένα ο έλεγχος Dickey – Fuller εξετάζει την συνθήκη κατά την οποία μια διαδικασία έχει μοναδιαία ρίζα και αν η οι πρώτες διαφορές απομάκρυναν αυτήν την ρίζα [58]. Επίσης απορρίπτεται αν η στατιστική τιμή  $t$  παίρνει τιμή μικρότερη της στατιστικής των Dickey Fuller.

#### 4.13.3.2 Μέθοδος Διαφορών

Η πρώτη τάξη διαφορών σε μία χρονολογική σειρά είναι οι διαφορές από την μία περίοδο σε μία άλλη.  $Y_t - Y_{(t-1)}$ , όπου  $Y_t$  η τιμή της χρονολογικής σειράς κατά την περίοδο  $t$ . [59] Το μειονέκτημα αυτής της μεθόδου είναι ότι σε κάθε εφαρμογή της χάνουμε μία παρατήρηση. [60]

Η μέθοδος διαφορών 2<sup>ης</sup> τάξης χρησιμοποιείται όταν ακόμα και μετά την εφαρμογή των διαφορών της πρώτης τάξης, υπάρχει ακόμα στασιμότητα. Σε αυτήν την περίπτωση ουσιαστικά μοντελοποιούμε τις διαφορές των διαφορών των αρχικών δεδομένων.  $y_t - 2y_{t-1} + y_t - 2$  [61]

#### 4.13.3.3 Τυχαίος Περίπατος.

Η πιο απλή μορφή ARIMA πρώτης τάξης, η οποία δεν είναι στάσιμη, είναι:  $Y_t = Y_{t-1} + \varepsilon_t$ , όπου η μεταβλητή  $\varepsilon_t$  είναι ο λευκός θόρυβος, δηλαδή το κάθε στοιχείο της προκύπτει από το προηγούμενο συν την πρόσθεση μίας τυχαίας τιμής. Αυτή η μορφή είναι γνωστή και ως τυχαία διαδρομή (ή περίπατος), random walk. Αν στο παραπάνω υπόδειγμα υπάρχει και σταθερός όρος  $\alpha$ , δηλαδή έχουμε:  $Y_t = \alpha + Y_{t-1} + \varepsilon_t$ , τότε το υπόδειγμά μας είναι τυχαίος περίπατος με περιπλάνηση (random walk with drift) [62].

#### 4.13.3.4 Εκτιμώντας Τους Όρους $p, q$ .

Αφού η χρονολογική σειρά έχει σταθεροποιηθεί, τότε θα πρέπει να βρούμε τους όρους AR και MA, δηλαδή  $p$  και  $q$  αντίστοιχα. Αυτό μπορεί να γίνει με την βοήθεια δύο διαγνωστικών γραφημάτων, τα οποία είναι: της αυτό – συσχέτισης και της μερικής αυτό – συσχέτισης. Σε αυτά τα διαγράμματα μπορούμε να παρατηρήσουμε αν το μοντέλο μας είναι AR, MA ή AR και MA. Το μοντέλο θεωρείται AR, εάν σε αυτό το γράφημα έχουμε πτώση προς το μηδέν μετά από υστέρηση και απότομη μεταβολή προς το μηδέν, επίσης μετά από υστέρηση. Αυτή η υστέρηση λαμβάνεται ως τιμή για το  $p$ . Το μοντέλο θεωρείται MA, εάν το γράφημα της μερικής αυτό – συσχέτισης φθίνει προς το μηδέν



έπειτα από υστέρηση και απότομη μεταβολή στο γράφημα της αυτό – συσχέτισης έπειτα από υστέρηση, αυτή η τιμή υπολογίζεται ως  $q$ . Και τέλος το μοντέλο θεωρείται AR και MA, εάν και τα δύο γραφήματα τείνουν στο μηδέν ασύμπτωτα. [63]

Γενικώς στον παρακάτω πίνακα οποίος εμφανίζεται στην πηγή: [64], φαίνονται ομαδοποιημένες οι διαδικασίες, με την συνάρτηση αυτό – συσχέτισης και της μερικής αυτό – συσχέτισης:

Διαδικασία	Συνάρτηση Αυτοσυσχέτισης $r_k$	Συνάρτηση Μερικής Αυτοσυσχέτισης $a_k$
AR(1)	Φθίνει προς το μηδέν. Στη θετική πλευρά εάν $a_1 > 0$ . Στην αρνητική πλευρά με εναλλαγή προσήμων εάν $a_1 < 0$ .	Μηδενίζεται μετά το $a_1$ – κορυφή στην πρώτη υστέρηση. Θετική κορυφή για $a_1 > 0$ . Αρνητική κορυφή για $a_1 < 0$ .
AR(2)	Φθίνει προς το μηδέν.	Μηδενίζεται μετά τη $2^{\text{η}}$ υστέρηση ( $a_2$ ).

Εικόνα 33 - Διαδικασίες AR [64].

Διαδικασία	Συνάρτηση Αυτοσυσχέτισης $r_k$	Συνάρτηση Μερικής Αυτοσυσχέτισης $a_k$
Αυτοπαλίνδρομη Διαδικασία $p$ τάξεως	Φθίνει γεωμετρικά ή φθίνει ακολουθώντας ημιτονοειδή συμπεριφορά.	Μηδενίζεται μετά από $p$ υστερήσεις.
Λευκός θόρυβος	Μηδέν	Μηδέν
MA(1)	Μηδενίζεται μετά την $1^{\text{η}}$ υστέρηση ( $r_1$ ). Θετική κορυφή όταν $\theta_1 < 0$ . Αρνητική κορυφή όταν $\theta_1 > 0$ .	Φθίνει γεωμετρικά στην αρνητική πλευρά για $\theta_1 > 0$ . Εναλλαγή προσήμων αρχίζοντας από τη θετική πλευρά $\theta_1 < 0$ .

Εικόνα 34 - Διαδικασίες MA και λευκός θόρυβος [64].

Διαδικασία	Συνάρτηση Αυτοσυσχέτισης $\gamma_k$	Συνάρτηση Μερικής Αυτοσυσχέτισης $\alpha_k$
Διαδικασία Κινητού Μέσου $q$ τάξεως MA( $q$ )	Μηδενίζεται μετά από $q$ υστερήσεις.	Φθίνει γεωμετρικά.
Αυτοπαλίνδρομη Διαδικασία Κινητού Μέσου ARMA( $p, q$ )	Φθίνει γεωμετρικά.	Φθίνει γεωμετρικά.

Εικόνα 35 - Διαδικασίες MA [64].

#### 4.13.3.5 Προβλέψεις ARIMA

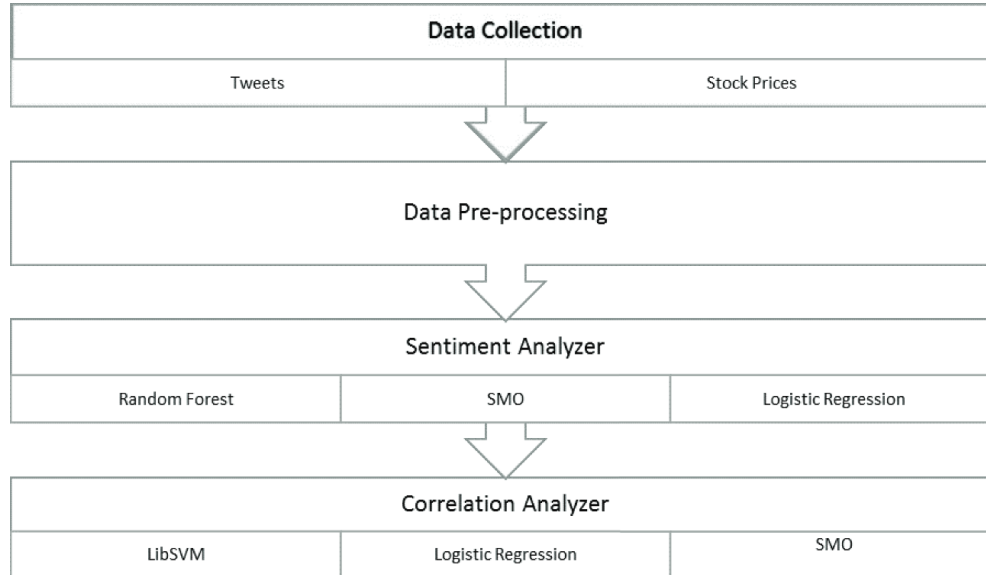
- Διαδικασίας MA( $q$ )  $\rightarrow$  Έχει μήκος  $q$ , δηλαδή περιορισμένη μνήμη, για την πρόβλεψη της τιμής  $t + 1$ , ισχύει:  $y_{t+1} = E\left(\mu + \theta_1 u_{t+1} + \frac{u_{t+1}}{t}\right) = \mu + \theta_1 u_t$  [64].
- Διαδικασία AR( $p$ )  $\rightarrow$  Απεριόριστη μνήμη και για την πρόβλεψη της τιμής  $t + 1$ , έχουμε:  $y_{t+1} = E\left(\mu + \Phi_1 y_t + \frac{u_{t+1}}{t}\right) = \mu + \Phi_1 y_t$  [64].

#### 4.13.4 Πρόβλεψη Με Τεχνικές Μηχανικής Μάθησης

Οι χρονολογικές σειρές μπορούν επίσης να προβλεφθούν με τεχνικές μηχανικής μάθησης, μπορεί να εφαρμοστεί οποιοσδήποτε αλγόριθμος μηχανικής μάθησης για την εκπαίδευση ενός μοντέλου και η χρονολογική σειρά να χρησιμοποιηθεί για εκπαίδευση, αφού έχει αφαιρεθεί η μεταβλητή του χρόνου. Η πρόβλεψη με αυτόν τον τρόπο αξιοποιεί περαιτέρω πληροφορίες που δεν θα αξιοποιούνταν από τις κλασικές τεχνικές της ανάλυσης των χρονολογικών σειρών και έτσι αποφέρει καλύτερα αποτελέσματα. Σε αυτήν την κατηγορία αναφέρονται από πολλές δημοσιεύσεις τα νευρωνικά δίκτυα, τα SVM, η αξιοποίηση του twitter, καθώς και η αξιοποίηση ειδήσεων από οικονομικές κυρίως ιστοσελίδες. Όσον αφορά τα ιστορικά δεδομένα τα οποία χρησιμοποιούνται ως εκπαίδευση στους αλγορίθμους, πολλές μελέτες έχουν δείξει ότι όντως αυτές οι μετρήσεις μπορούν να διαδραματίσουν ρόλο στην εξέλιξη της τιμής της μετοχής, όπως φυσικά και οι ειδήσεις. Πολλές δημοσιεύσεις συνδυάζουν τις παραπάνω μεθόδους ώστε να έχουν καλύτερα αποτελέσματα.

#### 4.13.4.1 Sentiment Analysis

Με την έλευση των μέσων κοινωνικής δικτύωσης, η πληροφόρηση σχετικά με το δημόσιο συναίσθημα είναι πλέον πλουσιοπάροχη. Με περισσότερα των 140 εκατομμυρίων τουϊτ κάθε μέρα από εκατομμύρια χρήστες, το τουϊτερ αποτελεί χρήσιμη πηγή πληροφοριών για τους ερευνητές. Οι πληροφορίες που μπορούν να εξαχθούν από το τουϊτερ, μπορούν να χρησιμοποιηθούν για πρόβλεψη. Οι ειδήσεις θεωρητικά μπορούν να έχουν επίδραση στην αγορά και ακόμα πιο συγκεκριμένα να έχουν θετική επίδραση ή αρνητική αν η τιμή της μετοχής στο τέλος της μέρας είναι ανοδική ή καθοδική. Άρα γενικώς ο σκοπός της συγκεκριμένης μεθόδου είναι η ανίχνευση και κατανόηση ενός συγκεκριμένου μοτίβου στις οικονομικές ειδήσεις ώστε να συνδεθεί αυτό με την άνοδο ή πτώση της μετοχής [21]. Πρόκειται για εργασία κατηγοριοποίησης συναισθήματος, αν είναι δηλαδή θετική μία δήλωση ή όχι. Όσον αφορά τα οικονομικά μέσω του τουϊτερ API μπορούν να ληφθούν και να εξαχθούν τουϊτς που αφορούν διάφορες εταιρείες λέξεις κλειδιά, πχ. Για την Microsoft #MSFT, #Windows, κτλ. Σημαντικό παράγοντα για την πρόβλεψη δεν αποτελεί μόνο η γνώμη του κόσμου για εταιρεία, αλλά επίσης πληροφορίες όπως πληροφορίες και κριτικές για προϊόντα ή υπηρεσίες που σχετίζονται με μία εταιρεία. Ως διαδικασία, αυτή η μεθοδολογία έπεται από την συλλογή δεδομένων τα τουϊτς θα πρέπει να κατηγοριοποιηθούν ως θετικά, ουδέτερα και αρνητικά, όπως φαίνεται και στο διάγραμμα ροής. [65]



Εικόνα 36 - Διάγραμμα ροής μεθοδολογίας sentiment analysis [65].

I'm really excited that today is my first day at @Microsoft as a Technical Evangelist	1
Ultrabooks in the mainstream (AAPLiPadProincontention) MSFT	1
About the Surface team, it's one of the most secret team at @Microsoft , except VPs, no one knows what they do, so don't expect leaks ;)	1
We broke all your devices, we force you to update whenever we want, and we spy on you all the time. Happy Anniversary!" - @microsoft	0
My grandfather, a depression era company man, hated that \$ MSFT didn't pay dividends (90s). I never could explain "new" idea of growth/cash.	2
Thanks @Microsoft , delivered my Surface Pro 3 power plug to a newsagents 20 miles from my home with no notification.	0
10 swift lessons. My post on @SwiftKey @Microsoft @georgewhitehead @OctopusVentures @IndexVentures @Accel	0
@Microsoft Acquires MinecraftEdu, Tailored for Schools http:// nyti.ms/1U9a1oV (via: @nytimesbusiness	0
My surface's touchscreen stopped working... I have homework... @Microsoft expand your customer service hours please	0
MSFT trailing revenues are declining since 2015, net income is up thanks to higher margins	2

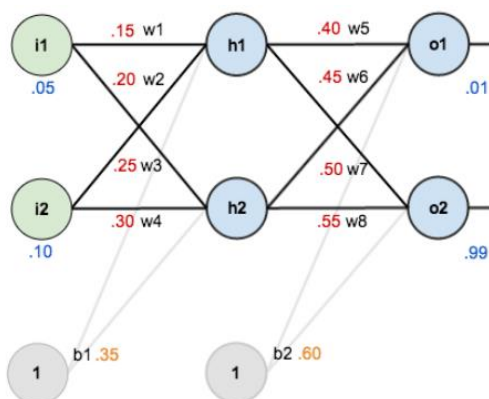
Εικόνα 37 - Παράδειγμα κατηγοριοποίησης των tweets [65].

#### 4.13.4.2 Τεχνητό Νευρωνικό Δίκτυο

Η πρόβλεψη χρονολογικών σειρών με την χρήση τεχνητών νευρωνικών δικτύων για μετοχές και γενικά οικονομικά δεδομένα, έχει αυξηθεί πλέον κατακόρυφα. Η ικανότητα της μάθησης από δεδομένα μέσω προσαρμοστικής δομής η οποία βασίζεται σε εξωτερικές ή εσωτερικές πληροφορίες, οι οποίες κατά την διάρκεια της εκπαίδευσης «κυλούν» στο δίκτυο και παράγουν μεταβλητές βασισμένες στην μάθηση είναι ένα από τα σημαντικότερα πλεονεκτήματα των τεχνητών νευρωνικών δικτύων.

#### 4.13.4.3 Back Propagation Νευρωνικό Δίκτυο

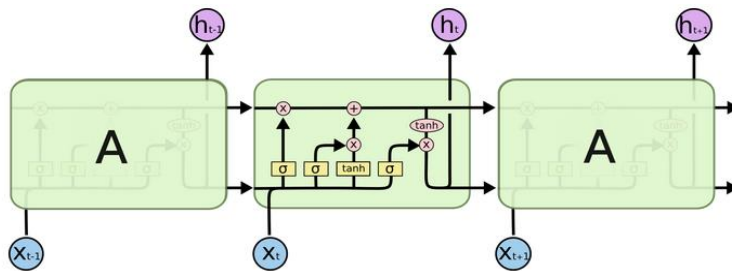
Για μη γραμμική δομή, όσον αφορά τις χρονολογικές σειρές η συνήθης δομή του είναι το τριών επιπέδων feed-forward back propagation. Τα βάρη καθορίζονται στην back propagation διαδικασία χτίζοντας συνδέσεις μεταξύ των κόμβων βασίζοντας στην εκπαίδευση δεδομένων, παράγοντας μετρήσεις για το ελάχιστο μέσο τετραγωνικό σφάλμα μεταξύ των πραγματικών ή επιθυμητών τιμών και των υπολογισμένων τιμών που εξαχθηκαν από το νευρωνικό δίκτυο. Η αρχιτεκτονική αυτού του δικτύου περιέχει ένα κρυφό επίπεδο από νευρώνια με μη γραμμική συνάρτηση μεταφοράς και ένα επίπεδο νευρώνων εξόδου με γραμμική συνάρτηση μεταφοράς. [66]



Εικόνα 38 - Αρχιτεκτονική Back Propagation Neural Network [67].

#### 4.13.4.4 Long Short-Term Memory (LSTM) Νευρωνικό Δίκτυο

Ανήκουν στα επαναλαμβανόμενα μοντέλα των νευρωνικών δικτύων, τα οποία διαφέρουν από τα παραδοσιακά feed – forward δίκτυα στο ότι δεν έχουν νευρωνικές συνδέσεις προς μία κατεύθυνση αλλά μπορούν να στείλουν δεδομένα και στο προηγούμενο η ίδιο επίπεδο. Όσον αφορά την πρόβλεψη μετοχών έχουν χρησιμοποιηθεί κυρίως αντλώντας δεδομένα κειμένου ως είσοδος, αλλά και η χρησιμοποίηση ιστορικών δεδομένων [68].

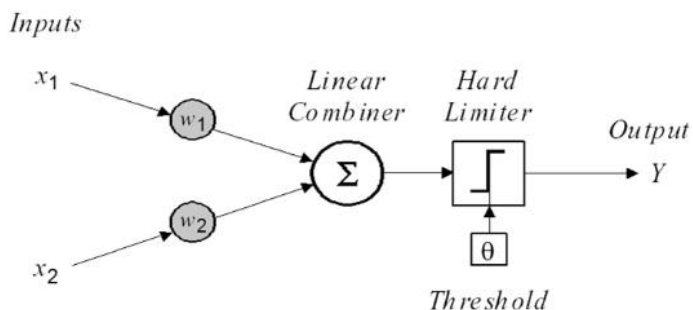


Εικόνα 39 - LSTM δίκτυο [69].

#### 4.13.4.5 Single - Layer Perceptron

Το SLP αποτελείται από ένα επίπεδο εισόδου και ένα εξόδο, οι νευρώνες στην έξοδο δέχονται το σταθμισμένο άθροισμα της εισόδου [70]. Τα βάρη σε κάθε επανάληψη υπολογίζονται σύμφωνα με τον τύπο:

- $w_i(p + 1) = w_i(p) + \Delta w_i(p)$  και
- $\Delta w_i(p) = \alpha \times X_i(p) \times e(p)$ , όπου  $\alpha$  ο ρυθμός μάθησης και  $e$  το λάθος. Η έξοδος υπολογίζεται ως:
- $Y(p) = \text{Step}[\sum_{i=1}^n X_i(p) \times w_i(p) - \theta]$ .

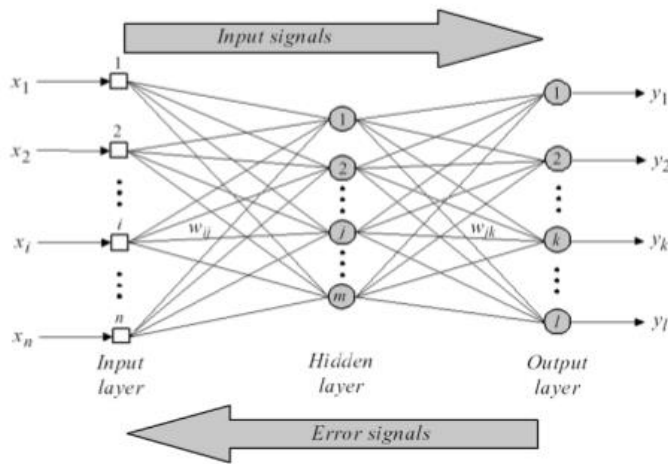


Εικόνα 40 - Single Layer Perceptron Model

#### 4.13.4.6 Multi-Layer Perceptron

Το MLP είναι τροφοδοτημένο με ένα επιπλέον επίπεδο, το οποίο ονομάζεται κρυφό επίπεδο. Αυτό το επίπεδο, το οποίο θα μπορούσε να είναι και περισσότερα, περιέχει επίσης και τους ενδιάμεσους νευρώνες. Οι νευρώνες εξόδου είναι εξαρτώμενοι από το αποτέλεσμα του κρυφού επιπέδου και το κρυφό επίπεδο εξαρτάται από τους νευρώνες εισόδου [70]. Τα βάρη στο κρυφό επίπεδο εξαρτώνται από τα βάρη του επιπέδου εξόδου, ώστε να υπολογιστεί η κλίση  $\delta$ .

- Η έξοδος των νευρώνων στο κρυφό επίπεδο υπολογίζεται από τον τύπο:  $y_j(p) = \text{Sigmoid} \left[ \sum_{i=1}^n X_{ji}(p) \times w_{ij}(p) - \theta_j \right]$
- Η κλίση  $\delta$ :  $\Delta_{w_{ij}}(p) = a \times X_i(p) \times \delta_j(p)$

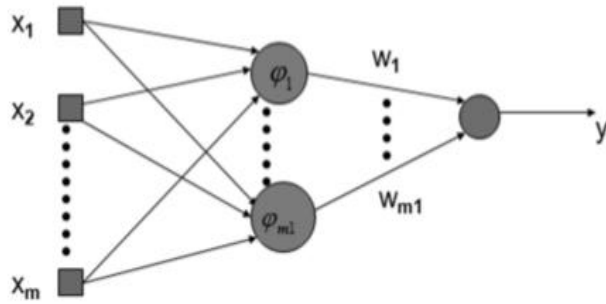


Εικόνα 41 - Multi - Layer Perceptron Model

#### 4.13.4.7 Radial Basis Function

Αυτό το είδος νευρωνικού δικτύου είναι αλγόριθμος με επίβλεψη και είναι feed – forward network. Εξαρτάται μόνο από την απόσταση της ακτίνας από ένα σημείο. Όπως και στο MLP, το RBF, έχει επίσης τρία διαφορετικά επίπεδα, δηλαδή, είσοδο, έξοδο και κρυφό επίπεδο. Κάθε κρυφός νευρώνας αναπαριστά μία radial basis function, η οποία έχει το κέντρο και το μήκος της. Ο RBF μετατρέπει τις μη – γραμμικές κλάσεις σε γραμμικές [70].

Η μη – γραμμική συνάρτηση  $\phi$  της RBF έχει ως εξής:  $\phi_{ij} = e^{-\frac{\|x-t_{-i}\|^2}{2s^2}}$



Εικόνα 42 - Radial Basis Function

#### 4.13.4.8 Bayesian Neural Network

Το Bayesian νευρωνικό δίκτυο το οποίο βασίζεται στην Bayesian πιθανολογική λογική. Τα Bayesian δίκτυα θεωρούν τις παραμέτρους του δικτύου ή τα βάρη τους ως τυχαίες μεταβλητές, οι οποίες ακολουθούν κάποια priori κατανομή:  $p(a|D) = \frac{p(D|\alpha)p(\alpha)}{p(D)}$ .

## 5 Εφαρμογές

Σε αυτήν την ενότητα εφαρμόζονται κάποιες από τις τεχνικές που μελετήθηκαν στα προηγούμενα κεφάλαια. Πιο συγκεκριμένα στην πρώτη εφαρμογή υλοποιείται στην ουσία μία παρουσίαση των μετοχών σε γραφικές παραστάσεις και γνωστούς μαθηματικούς υπολογισμούς που χρησιμοποιούνται για τέτοιου είδους δεδομένα και παρουσιάζεται μία απλή εφαρμογή μίας οικονομικής στρατηγικής. Η δεύτερη εφαρμογή πραγματοποιεί προβλέψεις με την μέθοδο ARIMA αφού πρώτα έχει αναλυθεί η χρονολογική σειρά, έχει εξαχθεί η τάση και έχουν αναγνωριστεί οι όροι  $p$ ,  $d$ ,  $q$  του μοντέλου ARIMA. Η Τρίτη εφαρμογή η οποία μπορεί να είναι και συνέχεια της πρώτης εφαρμογής ή και αυτόνομη, πραγματοποιεί προβλέψεις, αυτήν την φορά με μεθόδους και τακτικές της μηχανικής μάθησης, στην «πράξη» εφαρμόζονται οι πιο γνωστοί κατηγοριοποιητές οι οποίοι δέχονται κυρίως τις παραμέτρους οι οποίοι έχουν ήδη προταθεί, υλοποιηθεί και στις δημοσιεύσεις που κάποιες από αυτές αναφέρονται στα προηγούμενα κεφάλαια. Ή στην συνέχεια υπολογίζονται τα σφάλματα και όπου είναι εφικτό παρουσιάζονται και γραφήματα. Επιπλέον εφαρμόζεται και η μέθοδος της γραμμικής παλινδρόμησης και παρουσιάζονται τα αποτελέσματα.

### 5.1 Βιβλιοθήκες Και Εργαλεία Που Χρησιμοποιήθηκαν

Για την επίτευξη της ανάλυσης και πρόβλεψης χρονολογικών σειρών χρησιμοποιήθηκε η γλώσσα προγραμματισμού της *rython* και κάποιες βιβλιοθήκες που υπάρχουν για αυτήν την γλώσσα και αφορούν τα πεδία της μηχανικής μάθησης, στατιστικής, ανάλυσης δεδομένων, επιστημονικών υπολογισμών κτλ. Οι πιο σύνηθες και διάσημες είναι οι παρακάτω:

- **Pandas:** είναι ένα πακέτο το οποίο παρέχει γρήγορες, εκφραστικές και ευέλικτες δομές δεδομένων. Παρέχουν αρκετά πλούσιο και ανοιχτού κώδικα ανάλυση δεδομένων, τα οποία μπορούν να είναι διαφορετικού τύπου. Οι δύο βασικές δομές δεδομένων του πακέτου είναι τα "Series" και το "Dataframe", τα οποία μπορούν να διαχειριστούν πλήθος γνωστών εφαρμογών όπως οικονομικών, στατιστικών, κτλ. Μπορεί επίσης εύκολα να διαχειριστεί ελλιπή δεδομένα, χρονολογικές σειρές, είσοδο – έξοδο αρχείων κτλ. [71]. Για τις εφαρμογές έχει χρησιμοποιηθεί για να συλλέξει οικονομικά δεδομένα από ιστοσελίδες όπως *google finance* ή *yahoo finance*. Τα οποία έχουν την μορφή είτε ως Series είτε ως dataframe και παρουσιάζει τα δεδομένα σε συνεργασία και με άλλες βιβλιοθήκες. Η επεξεργασία και μετατροπή των δεδομένων από την μία μορφή σε άλλη, πχ. Από «dataframe» σε «Series», ώστε να βρίσκονται στην μορφή που χρειαζόμαστε κάθε φορά είναι σύνηθες και εύκολη.



- NumPy: Είναι πακέτο το οποίο χρησιμοποιείται για θεμελιώδη επιστημονικό υπολογισμό, περιέχει μεταξύ άλλων, N – διάστατο πίνακα, γραμμική άλγεβρα μετασχηματισμός Φουριέρ, κτλ. [72]. Χρησιμοποιείται στην εφαρμογή κυρίως για εργασίες με πίνακες και να φέρει τα δεδομένα στην μορφή που περιμένουν ως είσοδο οι συναρτήσεις της sklearn για εκπαίδευση και πρόβλεψη.
- Sklearn: Πλήρης ολοκληρωμένη βιβλιοθήκη για την rython που αφορά την μηχανική μάθηση, περιέχει εργαλεία για ανάλυση και εξόρυξη δεδομένων, ανοιχτού κώδικα, περιέχει όλους τους γνωστούς αλγορίθμους για την κατηγοριοποίηση, παλινδρόμηση, συσταδοποίηση, προ - επεξεργασία δεδομένων, κτλ. επίσης διαθέτει προ - εγκατεστημένα dataset. [73]. Στην εφαρμογή χρησιμοποιούνται για την εκπαίδευση και πρόβλεψη χρονοσειρών, καθώς και ρύθμιση παραμέτρων των αλγορίθμων, κτλ.
- Matplotlib: βιβλιοθήκη η οποία διαχειρίζεται κυρίως γραφήματα, τα οποία για τις εφαρμογές και τις οικονομικές χρονολογικές σειρές αποτελούν πολύ σημαντικό κομμάτι για την ανάλυσή τους, κατανόησή τους και τέλος την πρόβλεψή τους. [74].
- Statsmodel: Περιέχει κλάσεις και συναρτήσεις, οι οποίες χρησιμοποιούνται για την εκτίμηση πολλών διαφορετικών στατιστικών μοντέλων. [75]. Στην εφαρμογή χρησιμοποιείται για την ανάλυση και κατανόηση των χρονολογικών σειρών καθώς και για τον έλεγχο Dickey – Fuller, ο οποίος ελέγχει την στασιμότητα του μοντέλου.

### 5.1.1 Συλλογή Δεδομένων (stock datasets)

Η συλλογή δεδομένων μπορεί να γίνει από ιστοσελίδες οι οποίες περιέχουν οικονομικά δεδομένα, οι πιο γνωστές είναι της yahoo (yahoo finance, <https://finance.yahoo.com>), και της google (google, <https://www.google.com/finance>) καθώς και από άλλες κυρίως οικονομικού ενδιαφέροντος ιστοσελίδες όπως της quandl (<https://www.quandl.com/>), κτλ. Οι ανωτέρω ιστοσελίδες προσφέρουν τις χρηματοοικονομικές χρονολογικές σειρές είτε σε μορφή csv, json είτε με την μορφή api σε διάφορες γλώσσες, στην περίπτωση μας σε rython. Κυρίως από το api θα αντλούμε τα δεδομένα, καθώς αποτελεί γρήγορο και ευέλικτο τρόπο. Έχουμε χρησιμοποιήσει κυρίως την google ως πηγή δεδομένων μέσω της βιβλιοθήκης των pandas, η οποία παρέχει και αυτή την δυνατότητα τα δεδομένα αποθηκεύονται σε ένα pandas data frame, λόγω του ότι η κάθε πηγή δεδομένων δεν παρέχει ακριβώς τα ίδια δεδομένα, πχ. η google δεν παρέχει στήλη με τις τιμές προσαρμοσμένου κλεισίματος ενώ η yahoo την προμηθεύει αυτήν την στήλη, επίσης κάποιες οικονομικές μεταβλητές όπως «Split Ratio» ή προσαρμοσμένο άνοιγμα τιμής κτλ. Παρέχεται από πιο εξεζητημένες

οικονομικές ιστοσελίδες όπως της Quandl, ;άρα αναλόγως του τι θέλουμε να χρησιμοποιήσουμε και να υπολογίζουμε χρησιμοποιούμε και την κατάλληλη πηγή.

## 5.2 Ανάλυση, Οπτικοποίηση Και Ανάπτυξη Οικονομικής Στρατηγικής

Εκτός της ανάλυσης με τεχνικής ανάλυσης χρονολογικών σειρών, ειδικά για τις οικονομικές χρονολογικές σειρές, έχουν αναπτυχθεί και μελετηθεί πολλές τεχνικές όχι μόνο πρόβλεψης αλλά και ανάλυσης ανάπτυξης οικονομικών στρατηγικών, οι οποίες μπορούν να αναπαρασταθούν και να αναλυθούν μέσω κυρίως της βιβλιοθήκης matplotlib η οποία αναπαριστά στο καρτεσιανό σύστημα αξόνων την πορεία της μετοχής καθώς και άλλους υπολογισμούς που ενδεχομένως χρειαζόμαστε. Οι τύποι που χρειαζόμαστε και η στρατηγική που ακολουθείται είναι κυρίως υιοθετημένη από εκπαιδευτικό υλικό της ιστοσελίδας [www.datacamp.com](http://www.datacamp.com) η οποία εξειδικεύεται σε εκπαίδευση εξ αποστάσεως με μαθήματα κυρίως της μηχανικής μάθησης. Επίσης, χρησιμοποιήθηκε και το βιβλίο «Mastering Pandas for finance» του Michael Heydt.

Αυτή η εφαρμογή συλλέγει δεδομένα από το API της Quandl για τα έτη 2000 – 2017, υπολογίζει γνωστές οικονομικές τιμές όπως: ημερήσιες επιστροφές (daily returns), τους κινητούς μέσους όρους για 31 μέρες και ένα χρόνο, 252, τα cumulative returns, την αστάθεια (volatility),κτλ. Στην συνέχεια οπτικοποιεί τα αποτελέσματα και τέλος παρουσιάζεται μία γραφική παράσταση η οποία προτείνει μέσω της εφαρμογής μίας απλής οικονομικής στρατηγικής την αγορά ή την πώληση μίας μετοχής.

Όπως φαίνεται και παρακάτω ορίζουμε δυο μεταβλητές στις οποίες δίνουμε τα αρχικά των μετοχών που θέλουμε να εξετάσουμε και ως παραμέτρους στην συνάρτηση εισάγουμε το “ticker” κωδική ονομασία της μετοχής όπως έχει οριστεί να αναζητείται από την ιστοσελίδα της quandl, επίσης εισάγεται η ημερομηνία έναρξης δεδομένων και τέλους, ως τέλους μέσω συνάρτησης της python έχουμε ορίσει να δέχεται πάντα την ημερομηνία που τρέχει κάποιος την εφαρμογή.

```
aapl = quandl.get("WIKI/AAPL", start_date=start, end_date=end)
```

```
goog = quandl.get("WIKI/GOOG", start_date=start, end_date=end)
```

Αν διαλέξουμε να κατεβάσουμε κάποιο csv για οποιαδήποτε συμπληρωματικά δεδομένα ενδεχομένως να χρειαστούμε, αφού το κατεβάσουμε από την ιστοσελίδα, πχ. Της yahoo τότε μέσω της βιβλιοθήκης των pandas θα το διαβάσουμε ως pandas dataframe ώστε να χειριζόμαστε τα δεδομένα μας μέσω αυτού, διότι είναι αρκετά πιο εύχρηστο και άμεσο.

Daily prices					
Aug 1, 2015 - Aug 30, 2017					
Update					
Date	Open	High	Low	Close	Volume
Aug 29, 2017	905.10	923.33	905.00	921.29	1,185,564
Aug 28, 2017	916.00	919.24	911.87	913.81	1,086,484
Aug 25, 2017	923.49	925.56	915.50	915.89	1,053,376
Aug 24, 2017	928.66	930.84	915.50	921.28	1,270,306
Aug 23, 2017	921.93	929.93	919.36	927.00	1,090,248
Aug 22, 2017	912.72	925.86	911.48	924.69	1,166,737
Aug 21, 2017	910.00	913.00	903.40	906.66	943,441
Aug 18, 2017	910.31	915.28	907.15	910.67	1,342,689
Aug 17 2017	925.78	926.86	910.98	910.98	1,277,238

Εικόνα 43 - Παροχή ιστορικών δεδομένων από google finance

### 5.2.1 Τα Δεδομένα

Τα δεδομένα που αντλήθηκαν με τους παραπάνω τρόπους τα οποία είναι 1946 εγγραφές για τα δεδομένα της apple, 846 για τα δεδομένα της google για δεδομένα από τον Ιανουάριο του 2010 μέχρι την ημέρα που εκτελεί κάποιος την εφαρμογή.

Τα δεδομένα όπως είπαμε μπορούν να είναι σε οποιαδήποτε μορφή, σε κάθε περίπτωση εμείς τα προτιμούμε στην μορφή των pandas data frame τα οποία είναι ένας τρόπος αποθήκευσης δεδομένων σε πίνακες [76]. Τα δεδομένα που αντλήθηκαν περιέχουν τα ακόλουθα χαρακτηριστικά:

- Date -> Ημερομηνία,
- Open -> Η τιμή που πήρε η μετοχή όταν άνοιξε το χρηματιστήριο την ημέρα εκείνη που αναφέρεται το χαρακτηριστικό Date,
- High -> Η υψηλότερη τιμή που έφτασε η μετοχή.
- Low -> Η χαμηλότερη τιμή.
- Close -> Η τιμή κλεισίματος της μετοχής
- Volume -> Ο όγκος που συναλλάχθηκε.
- Adj. Close -> Η τιμή κλεισίματος τροποποιημένη ώστε να περιέχει διανομές και εταιρικές πράξεις, οι οποίες πραγματοποιήθηκαν πριν το άνοιγμα της τιμής της μετοχής την επόμενη μέρα.

Στις περισσότερες εφαρμογές χρησιμοποιείται κυρίως η τιμή κλεισίματος ή προσαρμοσμένου κλεισίματος (Adj Close). Επίσης να ληφθεί υπόψη ότι για τις μετοχές ο χρόνος μετράει από Δευτέρα έως Παρασκευή, δηλαδή όσο είναι ανοιχτά τα χρηματιστήρια, άρα Σάββατο και Κυριακή αν υπάρχουν στα δεδομένα, που συνήθως δεν υπάρχουν, οι τιμές δεν θα πρέπει λαμβάνονται υπόψιν.

Date	Open	High	Low	Close	Adj Close	Volume
1999-12-31	1464.469971	1472.420044	1458.189941	1469.250000	1469.250000	374050000
2000-01-03	1469.250000	1478.000000	1438.359985	1455.219971	1455.219971	931800000
2000-01-04	1455.219971	1455.219971	1397.430054	1399.420044	1399.420044	1009000000
2000-01-05	1399.420044	1413.270020	1377.680054	1402.109985	1402.109985	1085500000

Εικόνα 44 - Παράδειγμα αρχείου csv.

Επίσης από τα παραπάνω δεδομένα μπορούμε να υπολογίσουμε και επιπλέον τιμές οι οποίες είναι αρκετά χρήσιμες για την ανάλυση και την εξαγωγή συμπερασμάτων για την συμπεριφορά μιας μετοχής. Για τον υπολογισμό αυτών των τύπων μπορεί να χρησιμοποιηθεί η βιβλιοθήκη των *pandas*, που παρέχει πολλές συναρτήσεις για αυτού του είδους πράξεις είτε απλός προγραμματισμός. Τιμές όπως:

- Daily returns → Είναι στην ουσία ένας δείκτης ο οποίος δείχνει κατά πόσο μία μετοχή «ανέβηκε» ή «κατέβηκε» σε σχέση με χθες ή κάποια άλλη προηγούμενη μέρα. Υπολογίζεται ως:  $\frac{price[t]}{price[t-1]} - 1$ , όπου  $t$  η σημερινή τιμή και  $t - 1$  η χθεσινή, αντικαθιστώντας το 1 με πχ. 4 παίρνουμε ως αποτέλεσμα την μεταβολή της τιμής σε σχέση με 4 μέρες πριν, κοκ. Επίσης ένα από τα κέρδη που προσφέρεται εδώ είναι η κανονικοποίηση των δεδομένων.
- Cumulative returns → Είναι το ποσό το οποίο μία επένδυση έχει κερδίσει ή χάσει για κάποιο ορισμένο χρονικό διάστημα. Υπολογίζεται ως:  $\frac{price[t]}{price[0]} - 1$ , όπου  $t$  η χρονική στιγμή από την οποία θέλουμε να εξετάσουμε και 0, είναι η αρχική τιμή της [77].
- Total Return: Είναι τα συνολικά κέρδη που έχει κερδίσει ένας επενδυτής για κάποιο ορισμένο διάστημα, συνήθως είναι για ένα χρόνο, εκφράζεται και αυτό επί τοις εκατό (%).
- Αστάθεια - Volatility: Είναι μία στατιστική μέτρηση της μεταβολής των επιστροφών από μία μετοχή για κάποιο καθορισμένο διάστημα. Όσο ψηλότερο volatility έχουμε, τόσο πιο ριψοκίνδυνη είναι η επένδυση σε μία μετοχή, δηλαδή, χρησιμοποιείται και για την σύγκριση μετοχών, ώστε να βρεθεί ποια έχει λιγότερο ρίσκο. Υπολογίζεται παίρνοντας την κινητή απόκλιση της επί τοις εκατό αλλαγής σε μία μετοχή.

- Stock Momentum: Είναι ένας μέσος όρος για τις προηγούμενες  $n$  μέρες. Κάθε μέρα δέχεται την ένδειξη 1 ή “UP”, κτλ. Αν η εκάστοτε τιμή κλεισίματος είναι μεγαλύτερη της προηγούμενης και -1 ή “Down”, κτλ. Αν είναι μικρότερη. [78]
- Ετήσιο κέρδος (annual return): Πρόκειται για τα μέσα ετήσια κέρδη που ενδεχομένως κέρδισε ένα επενδυτής. Δίνεται από τον τύπο:  $1 + cumulative\ return^{\frac{365}{days}} - 1$ .
- Ordinary Least-Squares Regression (OLS): Είναι μία μορφή μαθηματικής παλινδρόμησης, η οποία βρίσκει την γραμμή, η οποία ταιριάζει πιο πολύ τα δεδομένα, απεικονίζοντας την μεταξύ τους συσχέτιση, μειώνει την απόκλιση. Ένα παράδειγμα θα μπορούσε να είναι να βρεθεί η σχέση μεταξύ index returns και stock returns [79].

### 5.2.2 Οπτικοποίηση Δεδομένων

Η οπτικοποίηση των δεδομένων γίνεται ώστε να μας βοηθήσει να καταλάβουμε την πορεία μίας μετοχής για ορισμένο χρονικό διάστημα, να συγκρίνουμε ενδεχομένως πολλές μετοχές σε ένα γράφημα, ώστε να βγάλουμε χρήσιμα συμπεράσματα, κτλ. Αυτό επιτυγχάνεται με την βοήθεια της βιβλιοθήκης της python, matplotlib.

### 5.2.3 Γράφημα Απεικόνισης

Η πορεία μίας μετοχή ή και περισσότερων μπορεί να παρασταθεί στο καρτεσιανό σύστημα αξόνων. Παριστάνοντας την τιμή κλεισίματος σε συνάρτηση του χρόνου.



Εικόνα 45 - Τιμών κλεισίματος (close) σε συνάρτηση με τον χρόνο (2010 - 2017), τριών μετοχών μαζί

### 5.2.4 Κεριά (Candlesticks)

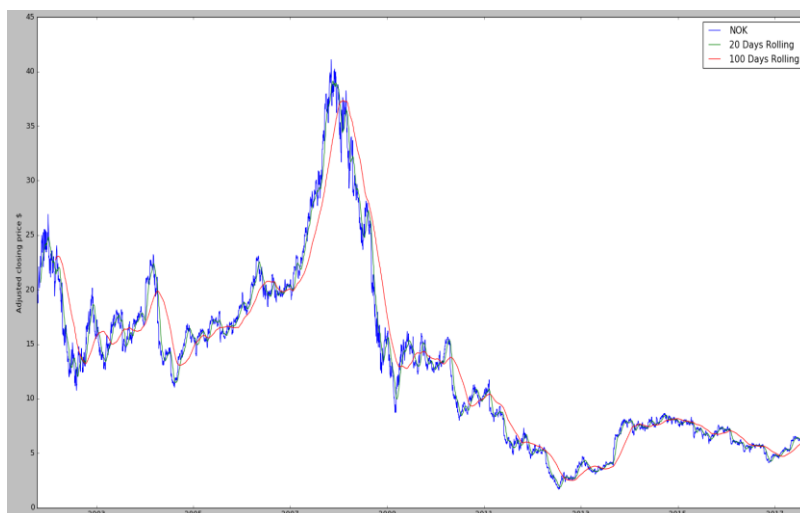
Ένας τρόπος παρουσίασης οικονομικών δεδομένων είναι τα λεγόμενα κεριά. Το κυρίως σώμα τους αντιπροσωπεύει τις τιμές μεταξύ ανοίγματος (open) και κλεισίματος (close) μίας ημέρας συναλλαγής. Όταν το κυρίως σώμα είναι μαύρο τότε αυτό σημαίνει πως η τιμή κλεισίματος είναι μικρότερη της τιμής ανοίγματος, στην αντίθετη περίπτωση το κυρίως σώμα του κεριού θα είναι άσπρο [80].



Εικόνα 46 - Παράδειγμα κεριών στην εφαρμογή

### 5.2.5 Κινητοί Μέσοι

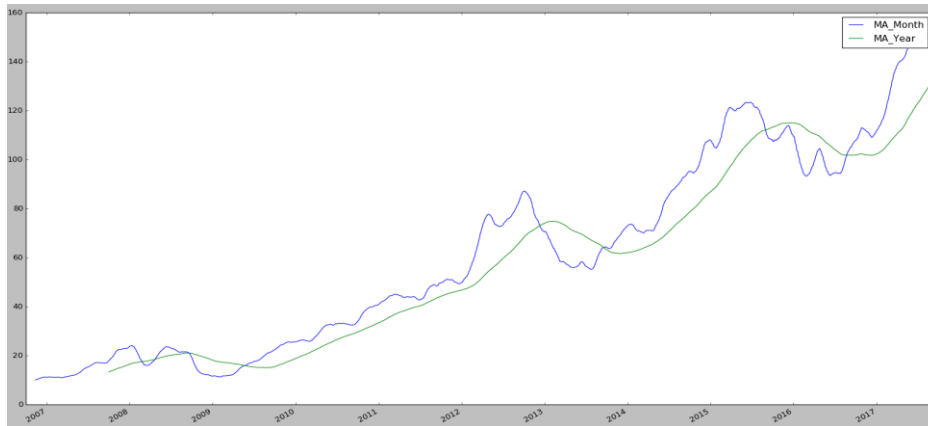
Στην παρακάτω εικόνα φαίνεται η πορεία μίας μετοχής και μαζί η εφαρμογή φίλτρου των κινητών μέσων με παράθυρο είκοσι ημερών και εκατό ημερών αντίστοιχα. Όπως φαίνεται και στο σχήμα της εφαρμογής η πράσινη και η κόκκινη γραμμή που αντιστοιχούν στα φίλτρα των κινητών μέσων, ακολουθούν πραγματική πορεία της μετοχής πιο εξομαλυμένα.



Εικόνα 47 - Γράφημα απεικόνισης κινητών μέσων

Και στην παρακάτω εικόνα φαίνεται άλλο ένα παράδειγμα κινητών μέσων για την μετοχή της apple αυτήν την φορά, έχουμε υπολογίσει κινητούς μέσους με παράθυρο 31 ημερών και 252 ημερών, το

οποίο αντιστοιχεί σε ένα έτος, θυμίζουμε ένα έτος είναι 252 διότι δεν υπολογίζονται οι ημέρες που η αγορά δεν λειτουργεί. Επίσης άλλη μία παρατήρηση είναι ότι όσο είναι το παράθυρο των κινητών μέσων που έχουμε επιλέξει τόσες τιμές δεν θα έχουμε στην αρχή των παρατηρήσεων, διότι δεν έχει υπολογιστεί ακόμα ο μέσος όρος. Άρα για ένα έτος παράθυρο μπορούμε να δούμε και στην γραφική παράσταση ότι η γραμμή που αντιστοιχεί για τους κινητούς μέσους με παράθυρο 252 ημερών ξεκινάει περίπου από το 2008, ενώ έχουμε δεδομένα από το 2007.



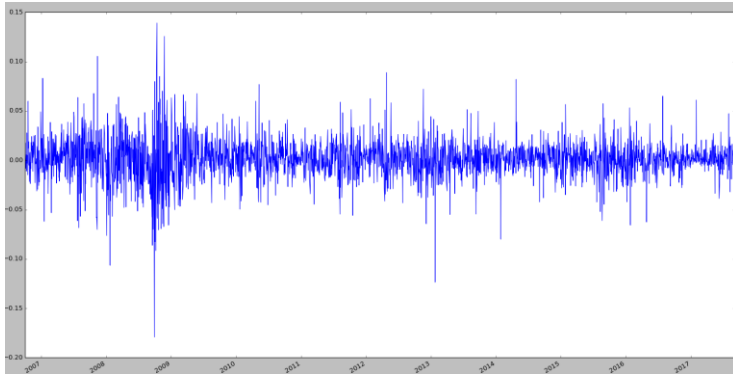
Εικόνα 48 - Κινητοί μέσοι που απεικονίζουν την μετοχή της apple για παράθυρο ενός μήνα και ενός έτους.

### 5.2.6 Daily Returns

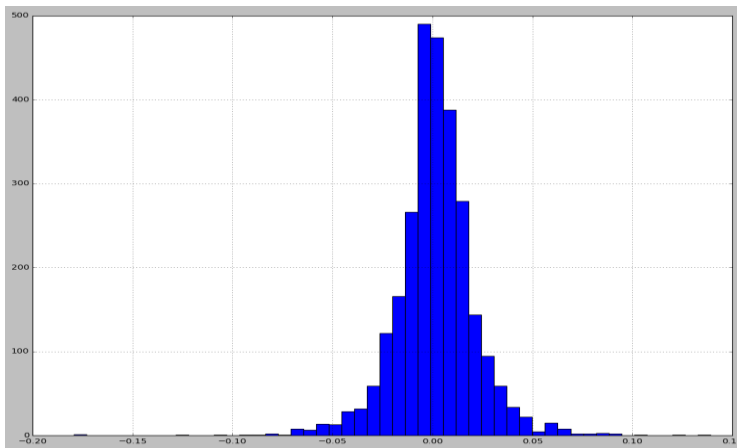
Υπολογίζουμε τα daily returns, έτσι ώστε να έχουμε μία εικόνα του πως κινείται η μετοχή για το χρονικό διάστημα που την ορίσαμε, επίσης θα δούμε και το γράφημα του ιστογράμματος των daily returns ώστε να δούμε την κατανομή των τιμών και να δούμε τον μέσο και την τιμή της διασποράς,

```
count    2749.000000
mean      0.001232
std       0.020202
min       -0.179195
25%       -0.007974
50%        0.001008
75%        0.011446
max        0.139050
Name: Adj. Close, dtype: float64
```

Τα οποία βλέπουμε πως έχουν τις τιμές 0.001 η mean και 0.020 η std αντίστοιχα. Ακόμα, όπως υπολογίσαμε τα daily returns, έτσι μπορούμε να υπολογίσουμε και returns ανά μήνα (monthly returns), quarter returns, Τα οποία μας βοηθούν να δούμε την εικόνα της μετοχής για όποια διαστήματα εμείς θέλουμε να ορίσουμε.



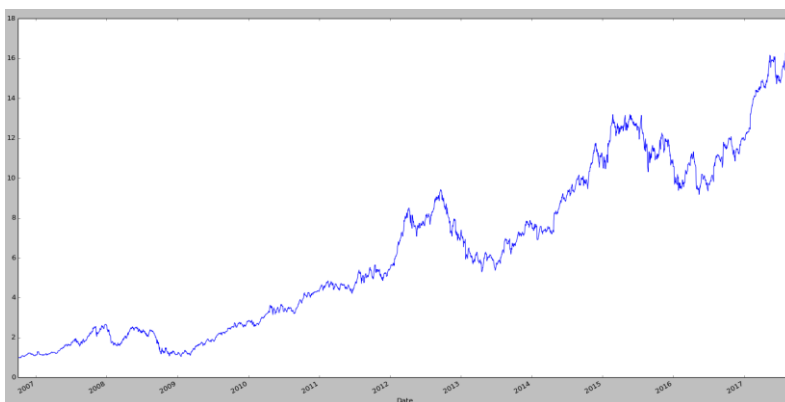
Εικόνα 49 - Γράφημα daily returns



Εικόνα 50 - Ιστόγραμμα daily returns, όπως βλέπουμε οι περισσότερες τιμές είναι γύρω από το μηδέν.

### 5.2.7 Cumulative Returns

Εδώ θα υπολογίσουμε το πόσο «κερδίσαμε» ή «χάσαμε» την εκάστοτε ημέρα. Και εδώ όπως και πριν μπορούμε να προσαρμόσουμε τα δεδομένα σε μηνιαία, τετράμηνα, κοκ.

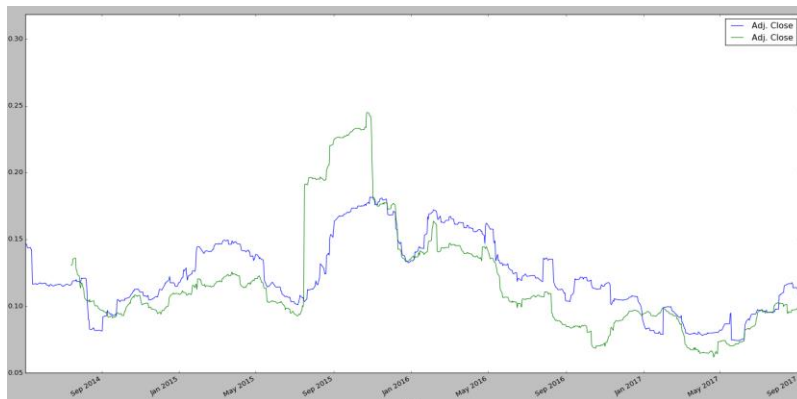


Εικόνα 51 - Γράφημα cumulative returns.



### 5.2.8 Αστάθεια - Volatility

Στο παρακάτω γράφημα φαίνεται το volatility δύο μετοχών, της apple και της google. Φαίνεται από το διάγραμμα η διαφορά στις τιμές των δύο αυτών μετοχών κατά την διάρκεια του χρόνου και κατά το 2017



Εικόνα 52 - Γράφημα Volatility, απεικονίζοντας δύο μετοχές, πράσινο την μετοχή της google και μπλε της apple.

### 5.2.9 Ordinary Least-Squares Regression (OLS)

- Η ετικέτα **Dep. Variable** παίρνει την τιμή της μεταβλητής που επιλέχθηκε ως εισοδος.
- Η ετικέτα **Model** την τιμή του μοντέλου που εφαρμόσαμε, δηλαδή, OLS.
- Παρόμοια ετικέτα με την παραπάνω είναι και η **Method** η οποία δείχνει πως υπολογίστηκαν οι τιμές του μοντέλου, δηλαδή με την εφαρμογή ελαχίστων τετραγώνων.
- Επίσης έχουμε τον αριθμό των παρατηρήσεων μέσω του **No. Observations**.
- Τον βαθμό ελευθερίας των **DF Residuals**.
- Τον αριθμό των παραμέτρων του μοντέλου, μέσω του **DF Model**, το οποίο δεν καταμετρά την σταθερά  $\chi$ .
- Συνεχίζοντας, μέσω του **R-squared** ο οποίος είναι συντελεστής της απόφασης. Αυτή η τιμή δείχνει πόσο καλά η γραμμή παλινδρόμησης προσέγγισε τα πραγματικά δεδομένα. Το 0.158 σημαίνει 16% περίπου, στο 0% έχουμε μηδενική μεταβλητότητα γύρω από το μέσο όρο και το 100% σημαίνει το αντίθετο.
- Το **Adj. R-squared** σκορ είναι η προσαρμογή της **R-Squared** τιμής με βάση τον αριθμό των παρατηρήσεων και την τιμή του **DF Residuals**.

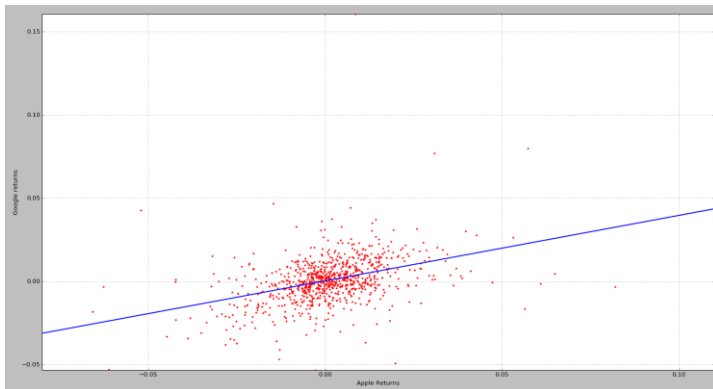
- **F-statistic**, είναι η μέτρηση του πόσο σημαντικό είναι το “fit”. Υπολογίζεται ως η διαίρεση του **mean square error** του μοντέλου με τον **residuals mean square error**.
- **Prob (F-statistic)**
- **AIC** (Akaike Information Criterion),
- **BIC** (Bayesian Information Criterion)
- Επίσης, **coef** η εκτιμώμενη τιμή του συντελεστή
- **Std, standard error** του συντελεστή **coef**.
- **T-statistic**, μετράει το πόσο στατιστικά σημαντικός είναι ο συντελεστής.
- **P > |t|**, κατά πόσο η «null - hypothesis» είναι αληθινή. Αν είναι μικρότερη ενός επιπέδου, συνήθως 0.05, τότε φαίνεται πως έχουμε στατιστικά σημαντική σχέση μεταξύ του αποτελέσματος και της παραμέτρου, σε αυτήν την περίπτωση έχουμε 0.539 για την σταθερά και 0 για τα δεδομένα.
- **Omnibus**, είναι στατιστικό τεστ για την ύπαρξη λοξότητα και κύρτωσης.
- **Prob Omnibus**
- **Skew**, μέτρηση συμμετρικότητας των δεδομένων γύρω από την μέση τιμή.
- **Kurtosis**, παρουσιάζει μία εικόνα της μορφής της κατανομής, είναι η σύγκριση των δεδομένων κοντά στην μέση τιμή και μακριά από αυτήν, δηλαδή στις ουρές.
- **Durbin – Watson**, τεστ για την ύπαρξη αυτό - συσχέτισης.
- **Jarque – Bera**, επιπλέον τεστ για την ύπαρξη λοξότητας και κύρτωσης.
- **Cond. No**, μέτρηση πολυσυγγραμμικότητας.

```

=====
                        OLS Regression Results
=====
Dep. Variable:          G00G      R-squared:                0.158
Model:                  OLS      Adj. R-squared:           0.157
Method:                 Least Squares  F-statistic:              162.4
Date:                   Thu, 07 Sep 2017  Prob (F-statistic):      3.33e-34
Time:                   22:43:51    Log-Likelihood:           2528.3
No. Observations:      867        AIC:                      -5053.
Df Residuals:          865        BIC:                      -5043.
Df Model:               1
Covariance Type:       nonrobust
=====
                    coef    std err          t      P>|t|    [0.025    0.975]
-----
const              0.0003      0.000      0.615    0.539    -0.001    0.001
AAPL               0.3934      0.031     12.742    0.000    0.333    0.454
=====
Omnibus:                 530.814    Durbin-Watson:           1.919
Prob(Omnibus):           0.000    Jarque-Bera (JB):       24152.677
Skew:                   2.113    Prob(JB):                0.00
Kurtosis:               28.509    Cond. No.                69.3
=====

```

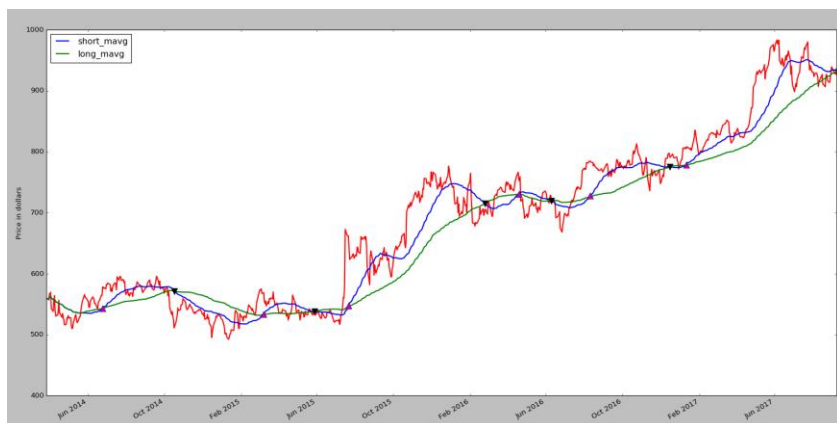
Εικόνα 53 - Αποτελέσματα εφαρμογής OLS.



Εικόνα 54 - Διάγραμμα Ordinary Least - Squares Regression.

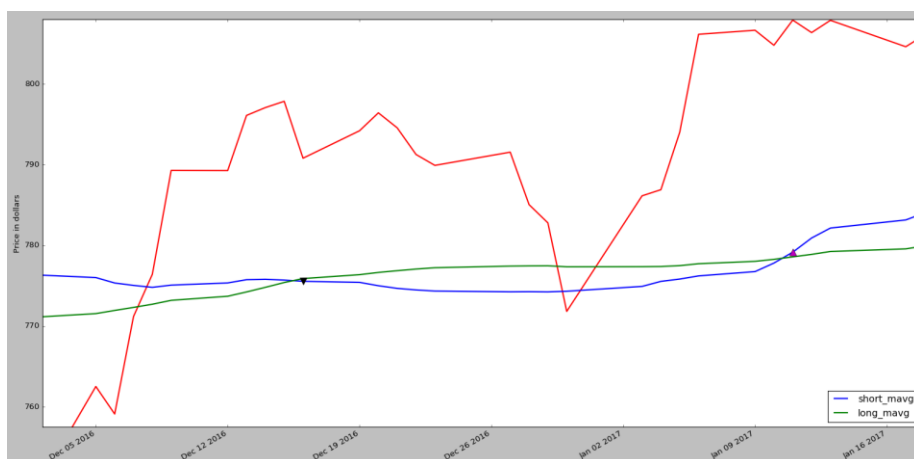
### 5.2.10 Χρηματιστηριακή Στρατηγική

Η στρατηγική η οποία θα ακολουθήσουμε είναι απλή και σύνθητες. Θα δημιουργήσουμε δύο διαφορετικούς κινητούς μέσους με διαφορετικό παράθυρο, πχ 40 και 100 ημερών. Αν το σήμα του κινητού μέσου όρου με παράθυρο 40 ημερών ξεπεράσει το αντίστοιχο των 100 ημερών τότε αυτό σημαίνει πως έχουμε σήμα να αγοράσουμε διότι η μετοχή πρόκειται να ανέβει στην αντίθετη περίπτωση που το σήμα του κινητού μέσου με 100 ημέρες ξεπεράσει το «μικρό» σήμα τότε αυτό είναι σήμα πώλησης διότι η μετοχή πρόκειται να μειωθεί



Εικόνα 55 - Γράφημα στρατηγικής.

Στο παραπάνω διάγραμμα απεικονίζεται η πορεία της μετοχής, σε αυτήν την περίπτωση της google μαζί με τους κινητούς μέσους με παράθυρα 40 ημερών, το μπλε και 100 ημερών το πράσινο. Επίσης φαίνεται ότι όπου συναντιούνται τα δύο σήματα, δηλαδή έχουμε την εφαρμογή της διασταύρωσης (crossover) έχουμε και τα αντίστοιχα σήματα «αγόρασε» και «πούλησε», στην παρακάτω εικόνα φαίνονται σε μεγέθυνση καθαρότερα δύο από τα σήματα.



Εικόνα 56 - Το παραπάνω γράφημα μεγεθυμένο, ώστε να φαίνονται καθαρότερα τα σήματα "πούλησε" και "αγόρασε".

βλέπουμε στο μεγεθυμένο γράφημα πως όπως είπαμε όταν το σήμα το «μικρό» περάσει το «μεγάλο» έχουμε σήμα «αγορά» αλλιώς έχουμε σήμα «πώληση».

### 5.3 Εφαρμογές – Πρόβλεψη Χρονολογικής Σειράς Με Την Τεχνική ARIMA

Η παρακάτω εφαρμογή στηρίζεται αρκετά στους οδηγούς των παρακάτω ιστοσελίδων, οι οποίες διαπραγματεύονται την στασιμότητα και την πρόβλεψη χρονολογικών σειρών γενικά αλλά και μέσω ARIMA.

<https://datascience.ibm.com/exchange/public/entry/view/815137c868b916821dec777bdc23013c>

<https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/>

Ο κώδικας και τα συμπεράσματα έχουν προσαρμοστεί για τα δεδομένα που θέλουμε, δηλαδή, οικονομικές χρονολογικές σειρές και είναι τροποποιημένος ανάλογα με τα αποτελέσματα που δεχόμαστε.

Η εφαρμογή αυτή αφού αντλήσει τα δεδομένα μέσω της yahoo finance για τον δείκτη της S&P 500, για τα έτη 2000 – 2017 και περίπου πάνω από 4000 εγγραφές. Παρουσιάζει μέσω γραφικών παραστάσεων την πορεία της τιμής κλεισίματος του δείκτη, των κινητών μέσω, της κινητής διακύμανσης. Από τις προαναφερόμενες γραφικές παραστάσεις και μέσω άλλων τεχνικών που περιγράφονται μπορούμε να εξάγουμε συμπεράσματα όπως για το αν υπάρχει τάση η κυκλικότητα στα δεδομένα και αν μέσω τεχνικών που εφαρμόσαμε εξαλείφθηκε. Στην συνέχεια αναγνωρίζεται η τάξη του μοντέλου ARIMA, δηλαδή οι παράμετροι  $p$ ,  $d$ ,  $q$  και αφού εφαρμοστούν πραγματοποιούνται οι προβλέψεις, όπου παρουσιάζονται γραφικά και υπολογίζεται ο ρυθμός σφαλμάτων, MSE, RMSE, κτλ.

### 5.3.1 Συλλογή Δεδομένων

Τα δεδομένα συλλέχθηκαν σε μορφή csv από την yahoo (yahoo finance), της οποίας η οικονομική ιστοσελίδα της παρέχει πλήθος οικονομικών και όχι μόνο δεδομένων και ειδήσεων, αρκετά από τα δεδομένα της τα προσφέρει δωρεάν. Πιο συγκεκριμένα κατέβηκε όπως είπαμε σε μορφή csv τα ιστορικά δεδομένα της S&P 500 (GSPC).

#### S&P 500 (^GSPC)

SNP - SNP Real Time Price. Currency in USD

☆ Add to watchlist

**2,500.23** +4.61 (+0.18%)

At close: September 15 4:38PM EDT

### 5.3.2 Οπτικοποίηση Δεδομένων

Τα ιστορικά δεδομένα επιλέχθηκαν εύκολα μέσω των επιλογών της ιστοσελίδας για το διάστημα 2000 – 2017 και συχνότητα καθημερινή (υπάρχουν επίσης επιλογές για εβδομαδιαία και μηνιαία).

Time Period: Sep 18, 2000 - Sep 18, 2017 Show: Historical Prices Frequency: Daily Apply

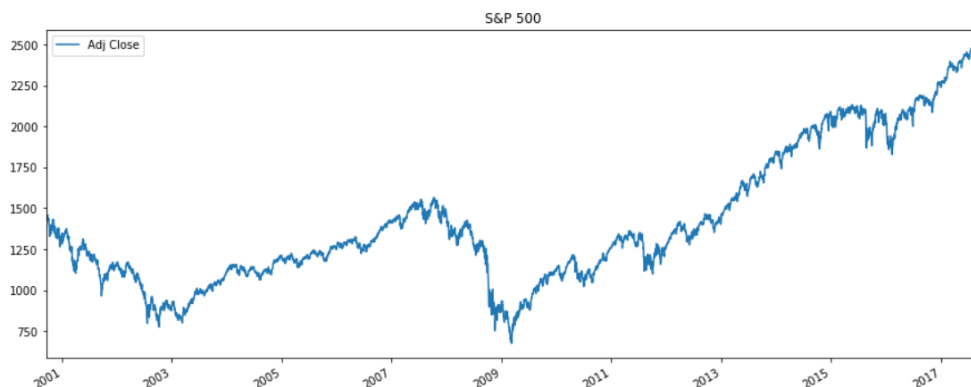
Currency in USD [Download Data](#)

Date	Open	High	Low	Close*	Adj Close**	Volume
Sep 15, 2017	2,495.67	2,500.23	2,493.16	2,500.23	2,500.23	4,853,170,000
Sep 14, 2017	2,494.56	2,498.43	2,491.35	2,495.62	2,495.62	3,414,460,000

Τα δεδομένα όπως φαίνονται και παραπάνω περιέχουν τις στήλες open, high, low, close, adj close και volume.

```
Data Types:
Date          object
Open          float64
High          float64
Low           float64
Close         float64
Adj Close     float64
Volume        int64
dtype: object
RangeIndex(start=0, stop=4276, step=1)
```

Για την ανάλυση της χρονολογικής σειράς σε αυτήν την περίπτωση μας ενδιαφέρει η στήλη «close» είτε η «adj close». Οπότε για να δούμε την συγκεκριμένη χρονολογική σειρά θα πρέπει να παρουσιάσουμε την προσαρμοσμένη τιμή κλεισίματος σε συνάρτηση με τον χρόνο.

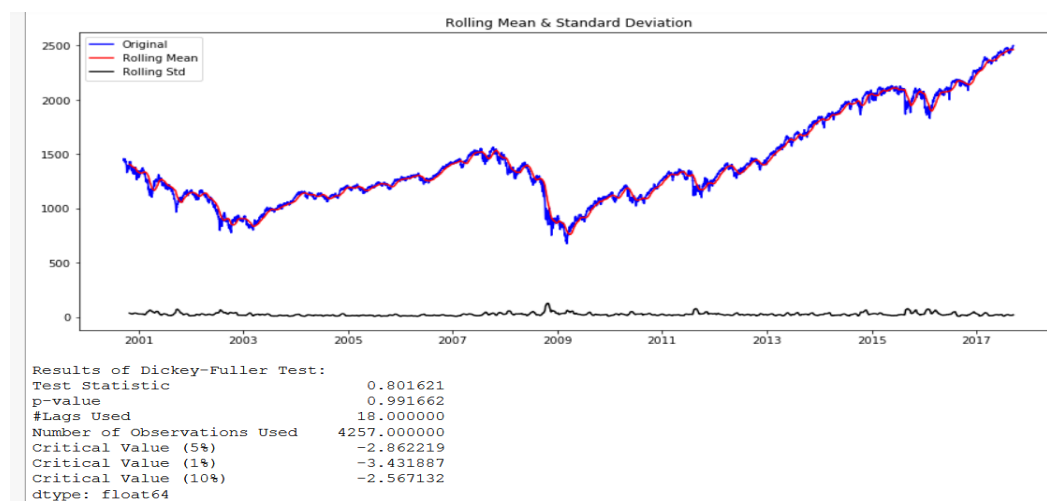


Εικόνα 57 - Πορεία μετοχής S&P 500 από τον 2000 μέχρι το 2017

Φαίνεται ξεκάθαρα από το γράφημα πως έχουμε ανοδική τάση (trend) της μετοχής, δηλαδή η χρονολογική σειρά δεν είναι στάσιμη (stationary) αυτή η τάση θα πρέπει να αφαιρεθεί για να εφαρμόσουμε την τεχνική ARIMA, αυτό μπορεί επίσης να επιβεβαιωθεί και από τον έλεγχο dickey – fuller, καθώς και να απεικονιστεί καλύτερα αν υπάρχει μέσω του γραφήματος το οποίο μετά από εφαρμογή αποσύνθεσης της χρονολογικής σειράς εμφανίζει την τάση, την εποχικότητα και τα residuals τα οποία εξηγούνται τι είναι στο γράφημα.

Από την κινούμενη διακύμανση η οποία φαίνεται στο γράφημα παρακάτω ότι δεν είναι σταθερή, μπορούμε να καταλάβουμε πως η χρονολογική σειρά δεν είναι στάσιμη. Ο έλεγχος της τάσης μπορεί μέσω του ελέγχου Dickey fuller από του οποίου τα αποτελέσματα πλέον επιβεβαιώνουμε και

καταλαβαίνουμε ότι η χρονολογική σειρά δεν είναι στάσιμη, δηλαδή η τιμή test statistic δεν είναι αρκετά χαμηλή και αρνητική, ούτε χαμηλότερη της τιμής Critical value 5% καθώς και η τιμή  $p$  δεν είναι μικρότερο του 0.05 ή μηδέν.



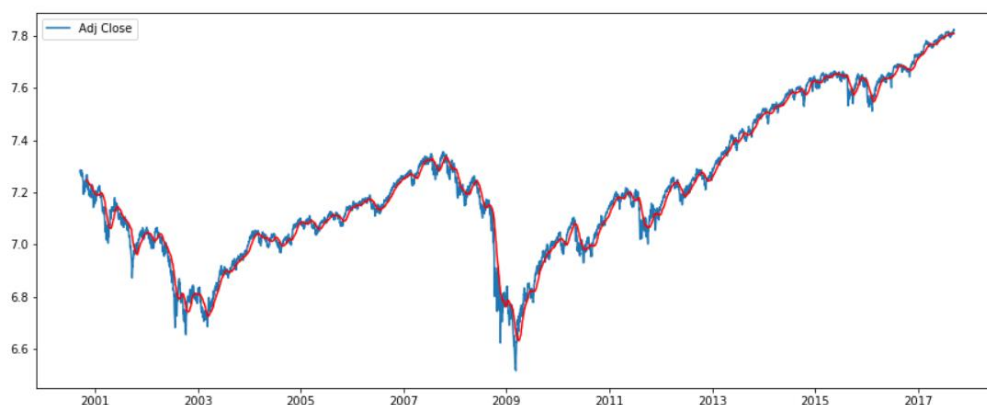
Εικόνα 58 - Απεικόνιση πορείας της μετοχής όπως και πριν, μαζί με κινούμενος μέσους και κινούμενη διασπορά.

Τα δεδομένα παρουσιάζουν τάση ή εποχικότητα τα οποία θα πρέπει να υπολογιστούν αρχικά και στην συνέχεια να αφαιρεθούν

### 5.3.3 Μετατροπή Χρονολογικής Σειράς Σε Στάσιμη

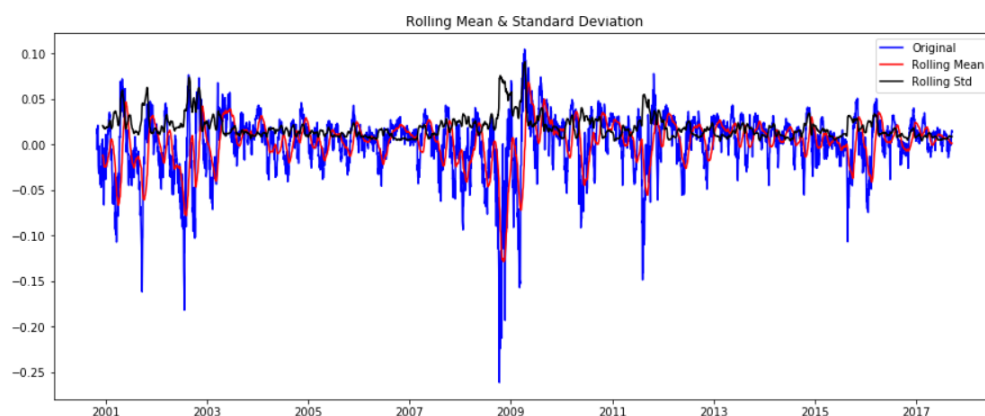
Για να γίνει η αφαίρεση της τάσης ή εποχικότητας θα πρέπει να πάρουμε τον λογάριθμο της τιμής, σε αυτήν την περίπτωση της τιμής κλεισίματος της μετοχής, ο οποίος θα μειώσει την διακύμανση της χρονολογικής σειράς. Στην συνέχεια θα πάρουμε την διαφορά 1<sup>ης</sup> τάξεως, δηλαδή σε κάθε ένα από τα δεδομένα αφαιρούμε τα αντίστοιχα της προηγούμενης μέρας.

Αυτή η διαδικασία απεικονίζεται στο ακόλουθο γράφημα δηλαδή οι κινούμενοι μέσοι και κινούμενη διακύμανση, για «παράθυρο», έχουμε πάρει την τιμή 31, δηλαδή 1 μήνα. Όπως φαίνεται και στο παρακάτω γράφημα, το οποίο απεικονίζει την λογαριθμημένη πλέον τιμή κλεισίματος και τους κινούμενους μέσους η ανοδική τάση παραμένει.



Εικόνα 59 - Γράφημα για την λογαριθμημένη τιμή κλεισίματος με και τους κινούμενους μέσους.

Άρα ως συνέχεια της διαδικασίας θα πρέπει να υπολογίσουμε τις διαφορές πρώτης τάξεως της χρονολογικής σειράς να εξετάσουμε πάλι το γράφημα και μέσω του ελέγχου Dickey Fuller να ελέγξουμε για στασιμότητα.



```
Results of Dickey-Fuller Test:|
Test Statistic      -1.090717e+01
p-value             1.120500e-19
#Lags Used          2.000000e+01
Number of Observations Used  4.225000e+03
Critical Value (5%)  -2.862224e+00
Critical Value (1%)  -3.431899e+00
Critical Value (10%) -2.567134e+00
dtype: float64
```

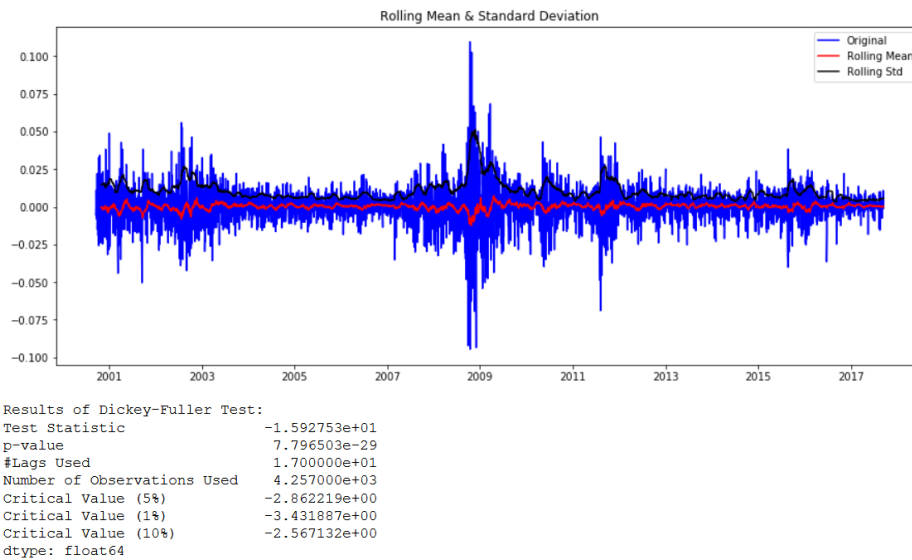
Εικόνα 60- Γράφημα απεικόνισης αρχικής πορείας της μετοχής, κινούμενων μέσων και διακύμανσης ενός μηνός.

Εξετάζοντας τις ανωτέρω τιμές φαίνεται πως έχει εξαλειφθεί αρκετά η τάση αλλά από τα αποτελέσματα του ελέγχου Dickey – Fuller συμπεραίνουμε πως η χρονολογική σειρά δεν είναι ακόμα στάσιμη λόγω του ότι η τιμή του test statistic αν και χαμηλότερη από πριν δεν είναι χαμηλότερη της τιμής του critical value 5%.

Εφαρμόζοντας πάλι τις διαφορές της λογαριθμημένης τιμής κλεισίματος, δηλαδή αφαιρώντας από την τιμή κλεισίματος της κάθε μίας μέρας την προηγούμενή της και ελέγχοντας για άλλη μία φορά

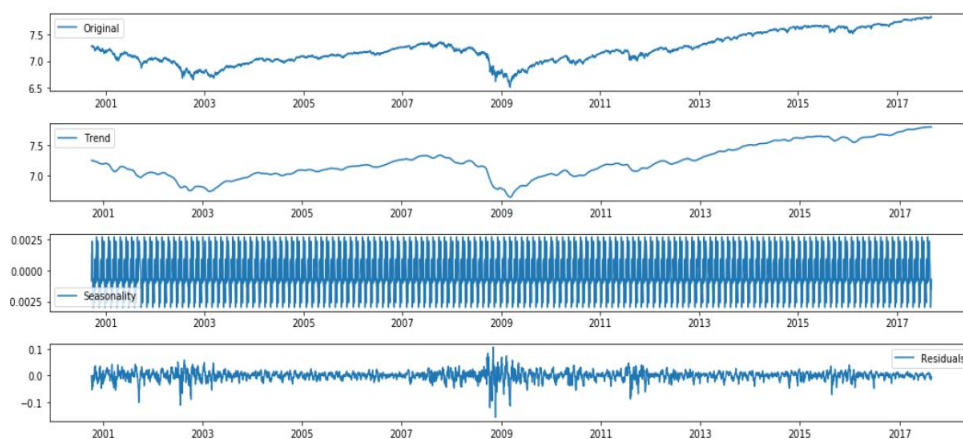


για στασιμότητα βλέπουμε πως πλέον η στασιμότητα έχει εξαλειφθεί και οι κινούμενοι μέσοι διασπορά είναι πλέον αρκετά πιο σταθεροί από όσο πριν, όπως φαίνεται και στο σχήμα:



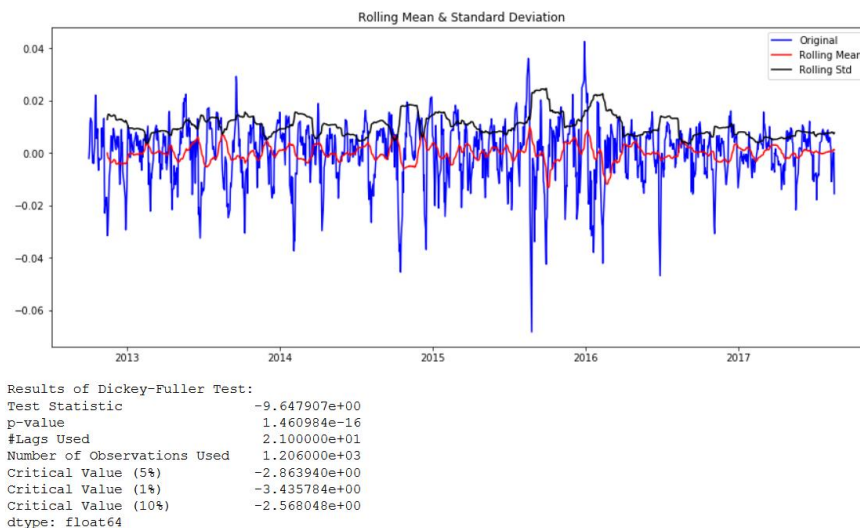
Εικόνα 61 - Γράφημα απεικόνισης αρχικής πορείας της μετοχής, κινούμενων μέσων και διακύμανσης ενός μηνός, μετά από την εφαρμογή των διαφορών.

Η τιμή του test statistic πλέον είναι χαμηλότερη του critical value 1% και η τιμή του  $p$  σχεδόν 0. Επιπλέον όμως μπορούμε να εφαρμόσουμε εποχική αποσύνθεση (seasonal decompose), μέσω της βιβλιοθήκης των stats model που παρέχεται για την rython, η οποία θα εξαλείψει επιπλέον την τάση και εποχικότητα που ενδεχομένως έχει απομείνει.



Εικόνα 62 - Απεικόνιση αρχικής χρονολογικής σειράς, τάσης, εποχικότητας και residuals.

Ελέγχοντας επιπλέον για στασιμότητα πλέον η τιμή του test statistic είναι αρκετά χαμηλή, χαμηλότερη των άλλων 3 κρίσιμων τιμών και το  $p$  – value «αγγίζει» το 0.

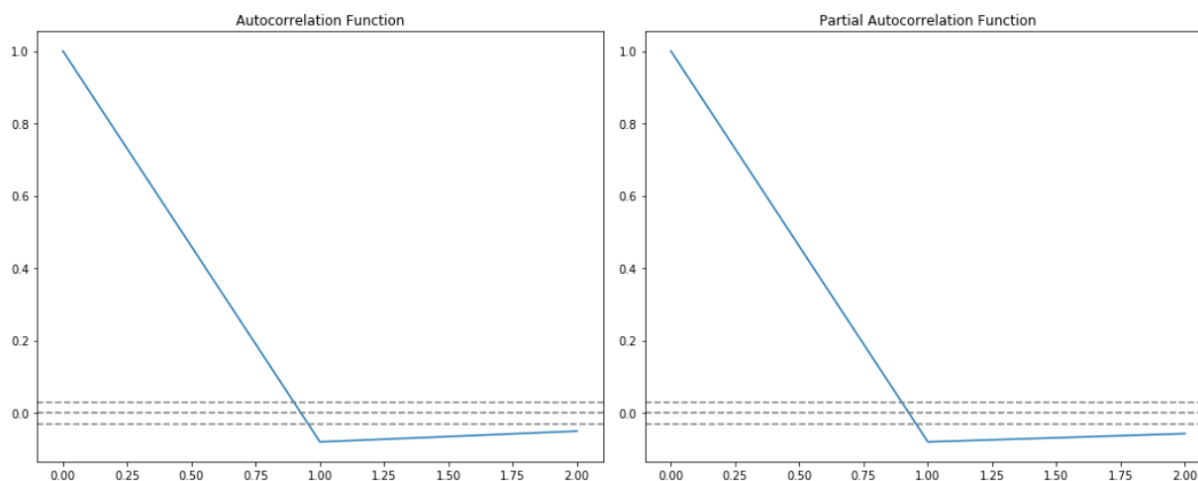


Τα residuals είναι η διαφορά της παρατηρούμενης τιμής από την πραγματική του τιμή.

Έπειτα από την εφαρμογή διαφορών και αποσύνθεσης η χρονολογική σειρά φαίνεται να έχει απαλλαγεί από τάση και εποχικότητα. Πλέον η χρονολογική σειρά είναι στάσιμη όπως επιβεβαιώνεται και από τον έλεγχο dickey fuller στο παραπάνω γράφημα.

### 5.3.4 Υπολογισμός Όρων $p$ , $d$ , $q$

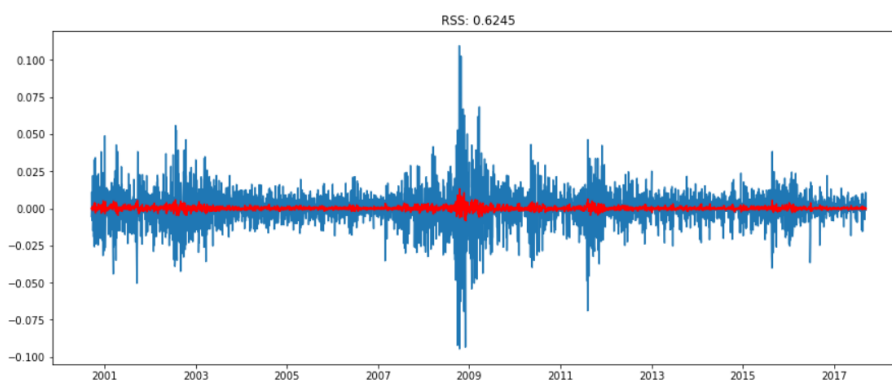
Πλέον θα πρέπει να υπολογίσουμε τους όρους  $p$ ,  $q$  οι οποίοι θα εφαρμοστούν στο μοντέλο Arima, αυτοί οι όροι μπορούν να εξαχθούν μελετώντας τα γραφήματα αυτό – συσχέτισης και μερικής αυτό – συσχέτισης.



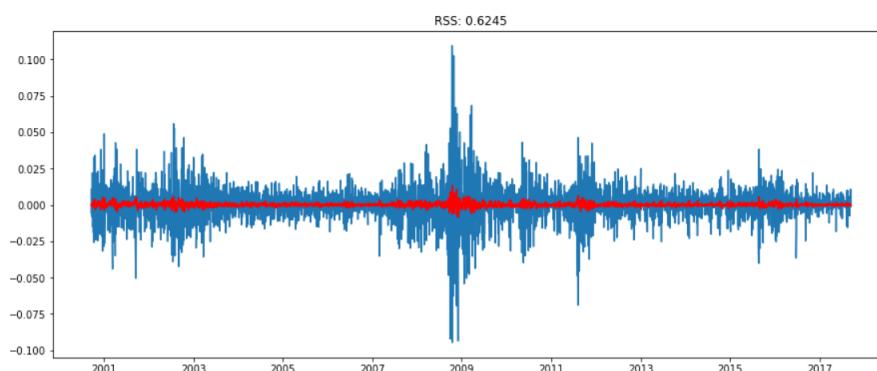
Εικόνα 63 - Διαγράμματα αυτό - συσχέτισης και μερικής αυτό - συσχέτισης.

Οι όροι  $p$  και  $q$ , είναι οι AR και MA όροι, του ARIMA μοντέλου, όπως φαίνεται και στο γράφημα της μερικής αυτό – συσχέτισης (το δεύτερο), η πρώτη φορά που η γραμμή περνάει το μηδέν έχει περίπου

την τιμή ένα, αυτή είναι η τιμή του  $\rho$ . Στο γράφημα της αυτό – συσχέτισης η γραμμή περνάει πρώτη φορά την στο άνω μεσοδιάστημα περίπου στην τιμή ένα. Άρα το μοντέλο Arima πλέον γράφεται Arima (1, 1, 1). Ωστόσο τα διαγράμματα παραπάνω είναι σχεδόν ίδια, για να επιβεβαιώσουμε το αποτέλεσμα θα χρησιμοποιήσουμε μέσω της βιβλιοθήκης «statsmodel» την συνάρτηση «arma\_order\_select\_ic» η οποία υπολογίζοντας το πληροφοριακό κριτήριο «akaike», το οποίο χρησιμοποιείται για τους υπολογισμούς στατιστικών μοντέλων και την αξιολόγηση της προσαρμογής του μοντέλου [81]. Οι τιμές που υπολογίστηκαν είναι 4 για το  $\rho$ , και 2 για το  $q$ , ενδεχομένως και άλλες τιμές να ταιριάζουν καλύτερα το μοντέλο ωστόσο να υπάρξουν μεγαλύτερες απώλειες. Θυμίζουμε ότι ο όρος  $d$  δηλαδή οι διαφορές, είναι πρώτης τάξεως άρα η τιμή είναι 1. Πλέον είμαστε έτοιμοι για τις προβλέψεις, εφαρμόζουμε όπως είδαμε παραπάνω μοντέλο (4, 1, 2). Για την αξιολόγηση του μοντέλου μπορούμε να χρησιμοποιήσουμε το άθροισμα των τετραγώνων των residuals σφαλμάτων, RSS, όσο μικρότερη τιμή τόσο το καλύτερο. Οι τιμές που υπολογίστηκαν, όπως φαίνεται και στα διαγράμματα παρακάτω είναι **0.6245**.



Εικόνα 64- RSS μοντέλου ARIMA για τους όρους AR.



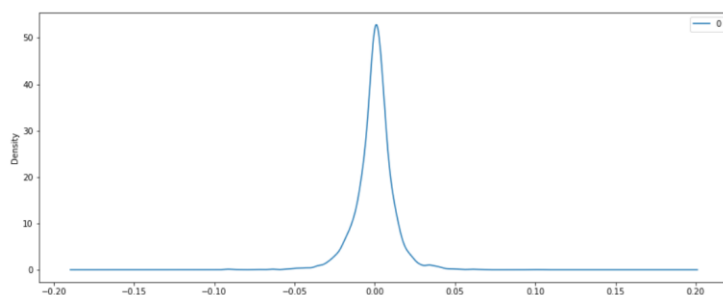
Εικόνα 65 - RSS μοντέλου ARIMA για τους όρους MA.

Το μοντέλο επίσης έχει ως έξοδο τα παρακάτω αποτελέσματα, από τα οποία πέραν των πληροφοριών όπως ημερομηνία, μέθοδος, τάξη, αριθμός παρατηρήσεων κτλ. Έχουν υπολογιστεί και τιμές όπως η «coef» η οποία «δείχνει» πόσο σημασία έχει η κάθε παράμετρος στην χρονολογική σειρά:

ARIMA Model Results						
Dep. Variable:	D.Adj Close	No. Observations:	4275			
Model:	ARIMA(4, 1, 2)	Log Likelihood	12810.858			
Method:	css-mle	S.D. of innovations	0.012			
Date:	Thu, 28 Sep 2017	AIC	-25605.716			
Time:	16:18:00	BIC	-25554.832			
Sample:	09-19-2000	HQIC	-25587.740			
	- 09-15-2017					
	coef	std err	z	P> z	[0.025	0.975]
const	0.0001	0.000	0.835	0.404	-0.000	0.000
ar.L1.D.Adj Close	1.0209	0.368	2.771	0.006	0.299	1.743
ar.L2.D.Adj Close	-0.4733	0.259	-1.824	0.068	-0.982	0.035
ar.L3.D.Adj Close	0.0323	0.025	1.295	0.195	-0.017	0.081
ar.L4.D.Adj Close	-0.0665	0.017	-4.014	0.000	-0.099	-0.034
ma.L1.D.Adj Close	-1.1067	0.369	-2.996	0.003	-1.831	-0.383
ma.L2.D.Adj Close	0.5101	0.288	1.769	0.077	-0.055	1.075
Roots						
	Real	Imaginary	Modulus	Frequency		
AR.1	1.0399	-0.8037j	1.3142	-0.1047		
AR.2	1.0399	+0.8037j	1.3142	0.1047		
AR.3	-0.7971	-2.8404j	2.9501	-0.2935		
AR.4	-0.7971	+2.8404j	2.9501	0.2935		
MA.1	1.0847	-0.8852j	1.4001	-0.1089		
MA.2	1.0847	+0.8852j	1.4001	0.1089		

Εικόνα 66 - Έξοδος αποτελεσμάτων ARIMA.

Όπως επίσης και το γράφημα της πυκνότητας των residuals, τα οποία μας υποδεικνύουν πως έχουμε μία κανονική κατανομή η οποία επικεντρώνεται γύρω από το μηδέν κατά μέσο όρο.



Εικόνα 67 - Γράφημα πυκνότητας των residuals.

### 5.3.5 Πρόβλεψη

Για τις προβλέψεις θα πρέπει να χωρίσουμε το μοντέλο σε εκπαίδευσης και εξέτασης (train, test) για το εξέτασης χρησιμοποιήσαμε 252 δείγματα δηλαδή ενός χρόνου και για το αποτέλεσμά του θα εξετάσουμε την τιμή του μέσου τετραγωνικού σφάλματος (MSE). Το μοντέλο παράγει τα ακόλουθα αποτελέσματα, στα οποία αριστερά έχουμε την τιμή που προβλέφθηκε και δεξιά την αναμενόμενη τιμή:

Printing Predicted vs Expected Values...

```

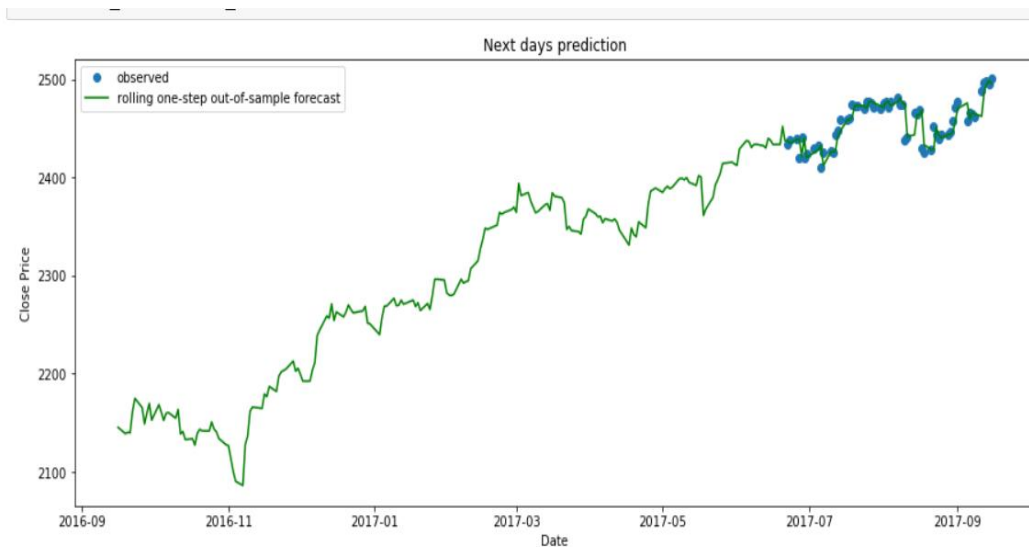
predicted=2145.184978, expected=2139.159912
predicted=2138.936074, expected=2139.120117
predicted=2140.091934, expected=2139.760010
predicted=2139.774983, expected=2163.120117
predicted=2161.378745, expected=2177.179932
predicted=2174.864912, expected=2164.689941
predicted=2165.486479, expected=2146.100098
predicted=2148.737490, expected=2159.929932
predicted=2159.895306, expected=2171.370117
predicted=2169.629358, expected=2151.129883
predicted=2152.640288, expected=2168.270020
predicted=2168.346193, expected=2161.199951
predicted=2160.746449, expected=2150.489990
predicted=2152.342100, expected=2159.729980
predicted=2159.652655, expected=2160.770020
predicted=2160.276549, expected=2153.739990
predicted=2154.663458, expected=2163.659912
predicted=2163.453909, expected=2136.729980

```

Εικόνα 68 - Αποτελέσματα πρόβλεψης ARIMA.

Και το μέσο τετραγωνικό σφάλμα είναι: **MSE: 0.000024**

Στο γράφημα στο οποίο απεικονίζουμε τις πραγματικές τιμές και αυτές που προβλέψαμε, μπορούμε να δούμε πως οι προβλέψιμες ακολουθούν σχεδόν την πορεία των αναμενόμενων τιμών όπως φαίνεται και στο παρακάτω γράφημα:



Εικόνα 69 - Αποτελέσματα προβλέψεων ARIMA.

#### 5.4 Πρόβλεψη Με Τεχνικές Μηχανικής Μάθησης

Για την πρόβλεψη οικονομικών χρονολογικών σειρών με τεχνικές μηχανικής μάθησης χρησιμοποιήθηκε η βιβλιοθήκη της sklearn, τα ιστορικά δεδομένα όπως και πριν συλλέχθηκαν μέσω του API της Quandl, για το διάστημα 2000 - 01 - 01 μέχρι την στιγμή που εκτελεί κάποιος την εφαρμογή, μέχρι την στιγμή που γράφεται αυτή η σελίδα το σύνολο των δεδομένων είναι, 4454 εγγραφές για τον δείκτη της Apple. Οι προβλέψεις θα βασίζονται στις ιστορικές τιμές του εκάστοτε δείκτη που επιλέγουμε και εφαρμόζουμε κυρίως τεχνικές κατηγοριοποίησης και παλινδρόμησης σε αυτά.

Για να εκπαιδύσουμε ένα μοντέλο που ανήκει στην κατηγορία της εποπτευόμενης μηχανικής μάθησης θα πρέπει να ετοιμάσουμε τις «ετικέτες», οι οποίες θα χρησιμεύσουν κατά την εκπαίδευση του μοντέλου και επίσης να επιλέξουμε τα χαρακτηριστικά που θα χρησιμοποιηθούν κατά την εκπαίδευση, δηλαδή για τα χαρακτηριστικά που θα χρησιμοποιήσουμε, σε αυτήν την περίπτωση είναι η τιμή ανοίγματος, η τιμή κλεισίματος, η ανώτατη τιμή που παρατηρήθηκε, η κατώτατη τιμή που παρατηρήθηκε, ο όγκος που συναλλάχθηκε, η τιμή προσαρμοσμένου κλεισίματος και ο κινητός μέσος όρος 5 ημερών δηλαδή μίας εβδομάδας. Θα πρέπει να προσαρμόσουμε αναλόγως τις «ετικέτες» ώστε να ισχύουν κάποιες τιμές με τα ανωτέρω χαρακτηριστικά «features», δηλαδή να εκπαιδεύεται τι επιρροή είχαν αυτά την επόμενη μέρα, αν η μετοχή έκλεισε με τιμή μεγαλύτερη της προηγούμενης ή όχι, αν πήγε καλύτερα από την προηγούμενη μέρα τότε η ετικέτα θα έχει την τιμή 1, αλλιώς την τιμή 0.

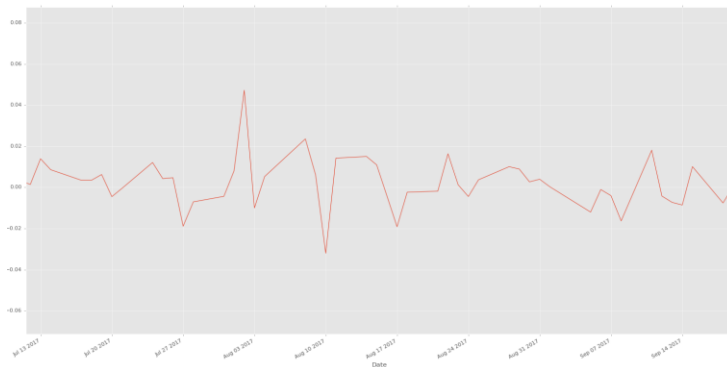
Η πορεία της μετοχής μπορεί εύκολα να υπολογιστεί με τον τύπο των daily returns, που έχουμε αναφερθεί παραπάνω, αρνητική τιμή σημαίνει η τιμή μειώθηκε σε σχέση με χθες, θετική το αντίθετο.

```

Date
2000-01-03      NaN
2000-01-04    -0.084331
2000-01-05     0.014634
2000-01-06    -0.086538
2000-01-07     0.047368
2000-01-10    -0.017588
2000-01-11    -0.051151
2000-01-12    -0.059946
[0, 0, 1, 0, 1, 0, 0, 0, 2000-01-13    0.109646

```

Παραπάνω φαίνονται οι πρώτες τιμές των daily returns, φυσικά για την πρώτη τιμή δεν έχουμε αποτέλεσμα καθώς θα έπρεπε να ξέραμε της προηγούμενης μέρας την τιμή κλεισίματος. Ως «ετικέτα» θεωρούμε το πρώτο αποτέλεσμα 0, δηλαδή σαν να είχε αρνητική πορεία, οι υπόλοιπες τιμές μπορούμε να δούμε πως ακολουθούν το μοτίβο που ορίσαμε



Εικόνα 70 - Γράφημα daily returns τελευταίων δύο μηνών.

Στην συνέχεια τα δεδομένα θα πρέπει να χωριστούν σε δεδομένα εκπαίδευσης και δεδομένα εξέτασης. Προτού όμως πραγματοποιηθεί αυτό θα πρέπει να τα επεξεργαστούμε ώστε να απαλειφούν από τιμές μη αριθμητικές, μπορούμε είτε αυτά τα δεδομένα είτε να μην τα υπολογίσουμε είτε να αντικαταστήσουμε την μη αριθμητική με κάτι άλλο, πχ. Τον μέσο όρο των δεδομένων, όπως και πραγματοποιήθηκε. Τα δεδομένα χωρίζονται με την τεχνική της διασταυρούμενης επικύρωσης που όπως είπαμε έχει ως στόχο να βρει τον καλύτερο συνδυασμό για τον χωρισμό δεδομένων εκπαίδευσης και αξιολόγησης. Πιο συγκεκριμένα για 10 «folds» η τεχνική είχε ως αποτέλεσμα τον εξής χωρισμό: 91% για εκπαίδευση και 9% για αξιολόγηση.

Με την χρήση της βιβλιοθήκης `numpy` προσαρμόζουμε τα δεδομένα ώστε να έχουν την μορφή πίνακα  $1 \times n$ , διότι σε αυτήν την μορφή δέχεται και τα απαιτεί η βιβλιοθήκη της `sklearn` για να πραγματοποιήσει την εκπαίδευση και την πρόβλεψη. Πλέον είμαστε έτοιμοι για να πραγματοποιήσουμε προβλέψεις με διάφορα μοντέλα που παρέχονται από την ανωτέρω βιβλιοθήκη και να συγκρίνουμε τα αποτελέσματα.

Για τις προβλέψεις όσον αφορά την κατηγοριοποίηση θα χρησιμοποιηθούν ο αλγόριθμος SVM με γραμμικό, πολυωνυμικό και το RBF κέλυφος, τα τυχαία δάση, και η γραμμική και ποσοτική διακριτή ανάλυση. Επίσης εφαρμόζεται και αλγόριθμος της συσταδιοποίησης των  $k$  κοντινότερων γειτόνων « $knn$ ». Και τέλος εφαρμόζεται και η παλινδρόμηση.

Τα μοντέλα παλινδρομητών που θα εφαρμοστούν για την εκπαίδευση είναι η γραμμική παλινδρόμηση και τα διανύσματα υποστήριξης παλινδρόμησης με κέλυφος το RBF, το γραμμικό και το πολυωνυμικό.

### 5.4.1 Εκπαίδευση Μοντέλου

Η εκπαίδευση όπως ειπώθηκε γίνεται με την βοήθεια της βιβλιοθήκης της sklearn για κάθε τύπο αλγόριθμου που θέλουμε να χρησιμοποιήσουμε. Αφού δηλαδή έχουμε ορίσει τις «ετικέτες» και χωρίσει τα δεδομένα σε εκπαίδευσης και εξέτασης αυτά εισέρχονται ως πίνακες  $1 \times n$  στοιχεία στην κατάλληλη συνάρτηση για εκπαίδευση και ύστερα πραγματοποιείται η εκπαίδευση, σε όλους τους τύπους αλγορίθμου η εκπαίδευση μέσω sklearn δίνεται με τον ίδιο τρόπο, δηλαδή τα «χαρακτηριστικά» features, μαζί με τις «ετικέτες»  $\rightarrow$  `clf.fit(X_train, y_train)`. Π.χ. αν για μία εγγραφή η οποία αντιστοιχεί σε μια μέρα (οποιαδήποτε) έχουμε ως τιμές των χαρακτηριστικών:

*Open: 200, Close: 280, Volume: 5000, High 350, Low: 150, Adj Close: 210* και *MA\_5* (στήλη που υπολογίσαμε εμείς, η οποία περιέχει τους κινούμενους μέσους 5 ημερών)  $\rightarrow$  τότε αναλόγως με τι αλγόριθμο και κλάση χρησιμοποιούμε αυτά τα δεδομένα θα υπολογιστούν έτσι ώστε να αντιστοιχούν με την «ετικέτα» της επόμενης μέρας 0 ή 1, που σημαίνει ότι η τιμή της μετοχής έπεσε ή ανέβηκε αντίστοιχα.

Όσον αφορά όμως τους παλινδρομητές οι οποίοι προβλέπουν αριθμητικές τιμές η εκπαίδευση θα πρέπει να πραγματοποιηθεί με διαφορετικό τρόπο. Τα χαρακτηριστικά που θα εφαρμόσουμε, features, αυτήν την φορά είναι η τιμή ανοίγματος, η μέγιστη τιμή, η ελάχιστη τιμή, ο όγκος που συναλλάχθηκε και οι κινούμενοι μέσοι 5 ημερών (1 εβδομάδα) αυτές θα πρέπει να ανταποκρίνονται στις τιμές κλεισίματος ή προσαρμοσμένου κλεισίματος, δηλαδή οι παλινδρομητές θα εκπαιδεύονται για την πρόβλεψη της τιμής κλεισίματος.

### 5.4.2 Απόδοση – Λάθη

Τα αποτελέσματα και η απόδοση της πρόβλεψης μπορούν να αξιολογηθούν με την χρήση κάποιων από των παρακάτω μετρήσεων για τις προβλέψεις για την μέθοδο της παλινδρόμησης [82]:

- *Μέσο απόλυτο σφάλμα (Mean absolute error)*  $\rightarrow$  Υπολογίζει πόσο «κοντά» «έπεσαν» οι προβλέψεις σε σχέση με τα τελικά αποτελέσματα.
- *Μέσο τετραγωνισμένο λογαριθμικό σφάλμα (Mean squared logarithmic error)*  $\rightarrow$
- *Μέσο τετραγωνικό σφάλμα (Mean squared error)*  $\rightarrow$  Υπολογίζει το μέσο τετραγωνικό σφάλμα,
- *Μέσο απόλυτο σφάλμα (Median absolute Error)*  $\rightarrow$  Είναι το μέσο όλων των απόλυτων διαφορών μεταξύ προβλέψεων και τελικών τιμών.



- *Explained variation* → Είναι η μέτρηση της αναλογίας που λογαριάζει ένα μαθηματικό μοντέλο για την διακύμανση.

Και για την μέθοδο της κατηγοριοποίησης [83]:

Classification Accuracy.

Logarithmic Loss.

Area Under ROC Curve → Είναι μέτρηση για δυαδικής κατηγοριοποίησης προβλήματα και μετράει την ικανότητα ενός μοντέλου να ξεχωρίσει τις κλάσεις, η μέτρηση 1 σημαίνει τέλεια κατηγοριοποίηση

Confusion Matrix.

Classification Report → Παρέχει μετρήσεις για την ακρίβεια, το «recall», την απόδοση «f1» και την τιμή «support».

- F1 → Πρόκειται για τον μέσο όρο της ακρίβειας και της τιμής «recall»,  

$$F1 = 2 * (precision * recall) / (precision + recall)$$
. Βέλτιστη τιμή 1

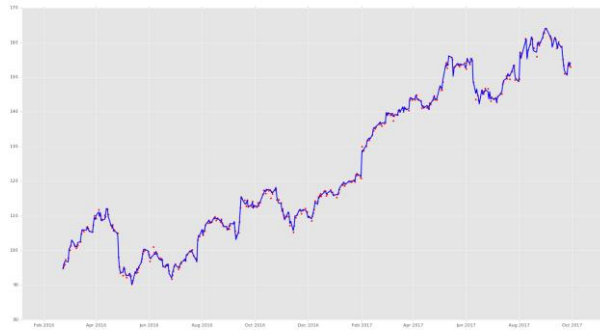
### 5.4.3 Αποτελέσματα

Όσον αφορά τις παραμέτρους που δέχεται το κάθε μοντέλο:

- Οι κατηγοριοποιητές των SVM δέχονται και τις παραμέτρους C και γ. Αυτές έχουν οριστεί ως 10000 και 0.1 αντίστοιχα, επίσης για το πολυωνυμικό κέλυφος εισήχθη η τιμή 3 για τον βαθμό του πολυωνύμου.
- Ο MLP κατηγοριοποιητής δέχεται ως παραμέτρους τα εξής: solver, alpha και hidden layers και ως τιμές τους adam ο οποίος αναφέρεται σε στοχαστικούς gradient-based βελτιστοποιητές, 0.00001 είναι το L2 πέναλτι και 30, 30, 30 τα οποία είναι τα κρυφά επίπεδα του νευρωνικού δικτύου.
- Τα τυχαία δάση δέχονται ως παράμετρο το πλήθος των «δέντρων» τα οποία έχουμε ορίσει ως 1000.
- Τέλος οι παλινδρομητές της SVR έχουν τις ίδιες τιμές με τους αντίστοιχους της κατηγοριοποίησης.

Οι τιμές κλεισίματος που προβλέφθηκαν από τους παλινδρομητές όπως φαίνεται και στα παρακάτω γραφήματα, όπου κόκκινες είναι οι προβλέψεις και μπλε η πραγματικές τιμές, φαίνεται πως οι τιμές που προβλέφθηκαν ακολουθούν σχεδόν τις πραγματικές τιμές:

- Γραμμική Παλινδρόμηση:



Εικόνα 71 – Πρόβλεψη γραμμικής παλινδρόμησης για τα δεδομένα αξιολόγησης.

- SVR Linear:



Εικόνα 72 - Προβλέψεις SVR Linear.

- SVR RBF:



Εικόνα 73 - Προβλέψεις SVR - RBF.

- MLP:



Εικόνα 74 - Προβλέψεις MLP.

Στον παρακάτω πίνακα φαίνονται συνοψισμένα όλα τα αποτελέσματα, όπως η μέτρηση ακρίβειας και σφαλμάτων, ότι ισχύει για κάθε μοντέλο.

Classifier		X	Precision	Recall	F1-Score	Support	Accuracy	MSE	MAE	Median AE
Support Vector Machines	Linear	0	0.71	0.12	0.2	185				
		1	0.56	0.96	0.71	220				
		<b>Summary</b>	0.64	0.58	0.48	405	<b>0.575</b>			
	Polynomial	0	0.68	0.76	0.72	185				
		1	0.78	0.70	0.73	220				
		<b>Summary</b>	0.55	0.55	0.55	405	<b>0.725</b>			
	RBF	0	0.5	0.53	0.52	185				
		1	0.59	0.56	0.57	220				
		<b>Summary</b>	0.55	0.55	0.55	405	<b>0.54</b>			
MLP	0	0.63	0.60	0.62	185					
	1	0.68	0.71	0.69	220					
	<b>Summary</b>	0.66	0.66	0.66	405	<b>0.66</b>				
QDA	0	0.58	0.45	0.51	185					
	1	0.61	0.73	0.67	220					

	<b>Summary</b>	0.6	0.6	0.59	405	<b>0.6</b>			
LDA	<b>0</b>	0.62	0.71	0.66	185				
	<b>1</b>	0.73	0.64	0.68	220				
	<b>Summary</b>	0.68	0.67	0.67	405	<b>0.671</b>			
Random Forest	<b>0</b>	0.52	0.38	0.44	185				
	<b>1</b>	0.58	0.71	0.64	220				
	<b>Summary</b>	0.55	0.56	0.55	405	<b>0.55</b>			
KNN	<b>0</b>	0.62	0.71	0.66	185				
	<b>1</b>	0.73	0.64	0.68	220				
	<b>Summary</b>	0.68	0.67	0.67	405	<b>0.545</b>			
Linear Regression							0.27	0.38	0.277
Support - Vector Regression									
	RBF						0.87	0.69	0.47
	Linear						0.27	0.38	0.27
MLP Regression							0.52	0.49	0.35

## 6 Συμπεράσματα

Η ανάλυση χρονολογικών σειρών με την τεχνική ARIMA φαίνεται πως μπορεί να ακολουθήσει την τάση της πορείας μίας μετοχής και οι προβλέψεις της να είναι ικανοποιητικές αν και δεν λαμβάνει υπόψιν περαιτέρω χαρακτηριστικά όπως τιμή ανοίγματος, όγκος που συναλλάχθηκε, κτλ. Παρά μόνο τον χρόνο και την επιρροή αυτού σε μία οικονομική μεταβλητή, πχ. Τιμή κλεισίματος. Επίσης η εξάλειψη της στασιμότητας και ορισμός των κατάλληλων όρων AR ( $\rho$ ), MA ( $q$ ) και  $d$  διαδραματίζουν σημαντικό ρόλο στην απόδοση του μοντέλου.

Η εφαρμογή τεχνικών μηχανικής μάθησης σε δεδομένα χρονολογικών σειρών φαίνεται πως και με μη εξειδικευμένα μοντέλα μπορεί να αποδώσει αρκετά καλά ακόμα και σε δεδομένα αρκετά πολύπλοκα όπως τα οικονομικά. Πιο συγκεκριμένα οι κατηγοριοποιητές LDA, MLP και ο SVM με πολυωνυμικό κέλυφος αποδώσανε καλύτερα με ακρίβειες, 67, 66 και 72% αντίστοιχα. Η επιλογή των χαρακτηριστικών που εισήχθησαν ως είσοδος για εκπαίδευση στους κατηγοριοποιητές και παλινδρομητές διαδραματίζει και αυτή σπουδαίο ρόλο στην απόδοση του εκάστοτε αλγορίθμου, ωστόσο στις εφαρμογές υπήρχε μικρή διαφορά στις τιμές των αποτελεσμάτων.

Ο υπολογισμός και οπτικοποίηση, τουλάχιστον των θεμελιωδών οικονομικών μεγεθών, και η ανάπτυξη οικονομικής στρατηγικής αποδεικνύονται πολύ χρήσιμες για κάποιον επενδυτή καθώς τον βοηθάει να δει γραφικά την πορεία των μετοχών και των οικονομικών μεταβλητών που τον ενδιαφέρουν και με την οικονομική στρατηγική που θα αναπτύξει να την συμβουλευτεί ώστε να «χτίσει» το χαρτοφυλάκιό του ή να πάρει επικερδής αποφάσεις όσον αφορά την αγορά και την πώληση στην οικονομική αγορά. Επίσης θα μπορούσε, αναλόγως του τι χρειαζόταν, να εισάγει αυτά τα θεμελιώδη οικονομικά μεγέθη ως είσοδο στο μοντέλο και είτε συνυπολογίζοντας και τις γνωστές ιστορικές τιμές όπως οι τιμές ανοίγματος, κλεισίματος, ανώτατη και χαμηλότερη και τον όγκο που συναλλάχθηκε, να έχει διαφορετικά αποτελέσματα στις προβλέψεις, είτε να προβλέψει άλλα μεγέθη όπως για παράδειγμα την τιμή ανοίγματος ή την αστάθεια, κτλ.

Τέλος, εφαρμόζοντας τις παραπάνω τεχνικές και μελετώντας τα αποτελέσματα, τα οποία θα πρέπει κάποιος να τα θεωρήσει ως συμβουλευτικά και σε συνεργασία με επαρκείς γνώσεις που ενδεχόμενως να έχει του πως λειτουργεί η «αγορά» και του πως μεταβάλλεται η τιμή των μετοχών κτλ. να μπορεί να έχει ενδεχόμενως καλύτερα αποτελέσματα απόδοσης από ότι θα είχε αν δεν χρησιμοποιούσε τις ανωτέρω εφαρμογές.

## Αναφορές

[1]	« <a href="http://www.investopedia.com/terms/t/timeseries.asp">http://www.investopedia.com/terms/t/timeseries.asp</a> ,» [Ηλεκτρονικό].
[2]	« <a href="http://www.investopedia.com/university/stocks/stocks1.asp?ad=dirN&amp;qo=investopediaSiteSearch&amp;qsrc=0&amp;o=40186">http://www.investopedia.com/university/stocks/stocks1.asp?ad=dirN&amp;qo=investopediaSiteSearch&amp;qsrc=0&amp;o=40186</a> ,» [Ηλεκτρονικό].
[3]	« <a href="http://www.businessdictionary.com/definition/stock.html">http://www.businessdictionary.com/definition/stock.html</a> ,» [Ηλεκτρονικό].
[4]	« <a href="http://opencourses.uom.gr/assets/site/public/847/674-Xrhmatooikonomikh-I-09-Tampakoudis.pdf">http://opencourses.uom.gr/assets/site/public/847/674-Xrhmatooikonomikh-I-09-Tampakoudis.pdf</a> ,» [Ηλεκτρονικό].
[5]	« <a href="http://www.investopedia.com/university/stocks/stocks2.asp">http://www.investopedia.com/university/stocks/stocks2.asp</a> ,» [Ηλεκτρονικό].
[6]	« <a href="https://el.m.wikipedia.org/wiki/%CE%9C%CE%B5%CF%84%CE%BF%CF%87%CE%AE_(%CE%BF%CE%B9%CE%BA%CE%BF%CE%BD%CE%BF%CE%BC%CE%AF%CE%B1)">https://el.m.wikipedia.org/wiki/%CE%9C%CE%B5%CF%84%CE%BF%CF%87%CE%AE_(%CE%BF%CE%B9%CE%BA%CE%BF%CE%BD%CE%BF%CE%BC%CE%AF%CE%B1)</a> ,» [Ηλεκτρονικό].
[7]	[Ηλεκτρονικό]. Available: <a href="http://www.investopedia.com/terms/d/djia.asp">http://www.investopedia.com/terms/d/djia.asp</a> .
[8]	« <a href="http://www.investopedia.com/terms/s/sp500.asp">http://www.investopedia.com/terms/s/sp500.asp</a> ,» [Ηλεκτρονικό].
[9]	« <a href="http://www.investopedia.com/terms/d/djia.asp">http://www.investopedia.com/terms/d/djia.asp</a> ,» [Ηλεκτρονικό].
[10]	« <a href="http://www.investopedia.com/terms/c/crossover.asp">http://www.investopedia.com/terms/c/crossover.asp</a> ,» [Ηλεκτρονικό].
[11]	« <a href="http://www.investopedia.com/terms/r/reversal.asp">http://www.investopedia.com/terms/r/reversal.asp</a> ,» [Ηλεκτρονικό].
[12]	« <a href="http://etfdb.com/etf-trading-strategies/3-simple-moving-average-etf-trading-strategies/">http://etfdb.com/etf-trading-strategies/3-simple-moving-average-etf-trading-strategies/</a> ,» [Ηλεκτρονικό].
[13]	« <a href="http://www.nasdaq.com/article/golden-cross-what-is-it-and-should-investors-watch-it-cm620734">http://www.nasdaq.com/article/golden-cross-what-is-it-and-should-investors-watch-it-cm620734</a> ,» [Ηλεκτρονικό].
[14]	« <a href="http://www.investopedia.com/terms/b/backtesting.asp">http://www.investopedia.com/terms/b/backtesting.asp</a> ,» [Ηλεκτρονικό].
[15]	« <a href="http://www.investinganswers.com/financial-dictionary/stock-market/backtesting-865">http://www.investinganswers.com/financial-dictionary/stock-market/backtesting-865</a> ,» [Ηλεκτρονικό].
[16]	« <a href="http://www.investopedia.com/university/stocks/stocks6.asp">http://www.investopedia.com/university/stocks/stocks6.asp</a> ,» [Ηλεκτρονικό].
[17]	« <a href="http://www.investopedia.com/university/stocks/stocks6.asp">http://www.investopedia.com/university/stocks/stocks6.asp</a> ,» [Ηλεκτρονικό].
[18]	« <a href="http://econ.ucsb.edu/~harmon/What%20Causes%20Stock%20Prices%20to%20Change.pdf">http://econ.ucsb.edu/~harmon/What%20Causes%20Stock%20Prices%20to%20Change.pdf</a> ,» [Ηλεκτρονικό].
[19]	« <a href="http://www.eap-forums.gr/arxeia/deo/deo31-sim.pdf">http://www.eap-forums.gr/arxeia/deo/deo31-sim.pdf</a> ,» [Ηλεκτρονικό].
[20]	« <a href="http://www.investopedia.com/terms/r/randomwalktheory.asp">http://www.investopedia.com/terms/r/randomwalktheory.asp</a> ,» [Ηλεκτρονικό].
[21]	« <a href="http://www.investopedia.com/university/beginner/">http://www.investopedia.com/university/beginner/</a> ,» [Ηλεκτρονικό].
[22]	V. H. Shah, «Machine Learning Techniques for Stock Prediction».
[23]	McGraw-Hill, «Machine Learning», 1997.
[24]	J. Bell, σε <i>Machine Learning: Hands-On for Developers and Technical Professionals</i> , 2014.
[25]	« <a href="http://aibook.csd.auth.gr/include/slides/Chap18.pdf">http://aibook.csd.auth.gr/include/slides/Chap18.pdf</a> ,» [Ηλεκτρονικό].
[26]	« <a href="https://quantdare.com/machine-learning-a-brief-breakdown/">https://quantdare.com/machine-learning-a-brief-breakdown/</a> ,» [Ηλεκτρονικό].
[27]	« <a href="https://www.cs.cmu.edu/~schneide/tut5/node42.html">https://www.cs.cmu.edu/~schneide/tut5/node42.html</a> ,» [Ηλεκτρονικό].
[28]	« <a href="https://www.packtpub.com/books/content/learning-how-classify-real-world-examples">https://www.packtpub.com/books/content/learning-how-classify-real-world-examples</a> ,» [Ηλεκτρονικό].
[29]	« <a href="http://www.cs.cmu.edu/~awm/tutorials/overfit10.pdf">http://www.cs.cmu.edu/~awm/tutorials/overfit10.pdf</a> ,» [Ηλεκτρονικό].
[30]	« <a href="http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html#sklearn.model_selection.TimeSeriesSplit">http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html#sklearn.model_selection.TimeSeriesSplit</a> ,» [Ηλεκτρονικό].
[31]	« <a href="https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/">https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/</a> ,» [Ηλεκτρονικό].
[32]	« <a href="http://www.cs.uoi.gr/~arly/courses/ai/slides/DecisionTrees.pdf">http://www.cs.uoi.gr/~arly/courses/ai/slides/DecisionTrees.pdf</a> ,» [Ηλεκτρονικό].
[33]	« <a href="http://blog.citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics">http://blog.citizennet.com/blog/2012/11/10/random-forests-ensembles-and-performance-metrics</a> ,» [Ηλεκτρονικό].

[34]	« <a href="http://uc-r.github.io/discriminant_analysis#linear">http://uc-r.github.io/discriminant_analysis#linear</a> ,» [Ηλεκτρονικό].
[35]	« <a href="http://people.cs.pitt.edu/~milos/courses/cs2750-Spring2012/Lectures/class7.pdf">http://people.cs.pitt.edu/~milos/courses/cs2750-Spring2012/Lectures/class7.pdf</a> ,» [Ηλεκτρονικό].
[36]	« <a href="https://www.svm-tutorial.com/2014/11/svm-understanding-math-part-1/">https://www.svm-tutorial.com/2014/11/svm-understanding-math-part-1/</a> ,» [Ηλεκτρονικό].
[37]	« <a href="https://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners">https://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners</a> ,» [Ηλεκτρονικό].
[38]	« <a href="https://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners">https://www.quantstart.com/articles/Support-Vector-Machines-A-Guide-for-Beginners</a> ,» [Ηλεκτρονικό].
[39]	« <a href="http://scikit-learn.org/stable/modules/svm.html#svm-classification">http://scikit-learn.org/stable/modules/svm.html#svm-classification</a> ,» [Ηλεκτρονικό].
[40]	« <a href="http://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html">http://scikit-learn.org/stable/auto_examples/svm/plot_svm_regression.html</a> ,» [Ηλεκτρονικό].
[41]	« <a href="http://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_iris.html#sphx-glr-auto-examples-cluster-plot-cluster-iris-py">http://scikit-learn.org/stable/auto_examples/cluster/plot_cluster_iris.html#sphx-glr-auto-examples-cluster-plot-cluster-iris-py</a> ,» [Ηλεκτρονικό].
[42]	Π. ΑΡΓΥΡΑΚΗΣ, Νευρωνικά Δίκτυα και Εφαρμογές, Πάτρα: ΕΛΛΗΝΙΚΟ ΑΝΟΙΚΤΟ ΠΑΝΕΠΙΣΤΗΜΙΟ, 2001.
[43]	« <a href="http://androulakis.bma.upatras.gr/mediawiki/index.php/%CE%94%CE%B9%CE%B1%CE%BC%CE%B5%CF%81%CE%B9%CF%83%CE%BC%CF%8C%CF%82_%CF%87%CF%81%CE%BF%CE%BD%CE%BF%CE%BB%CE%BF%CE%B3%CE%B9%CE%BA%CF%8E%CE%BD_%CF%83%CE%B5%CE%B9%CF%81%CF%8E%CE%BD_%CF%83%CE%B5_%CF%83%CF%8">http://androulakis.bma.upatras.gr/mediawiki/index.php/%CE%94%CE%B9%CE%B1%CE%BC%CE%B5%CF%81%CE%B9%CF%83%CE%BC%CF%8C%CF%82_%CF%87%CF%81%CE%BF%CE%BD%CE%BF%CE%BB%CE%BF%CE%B3%CE%B9%CE%BA%CF%8E%CE%BD_%CF%83%CE%B5%CE%B9%CF%81%CF%8E%CE%BD_%CF%83%CE%B5_%CF%83%CF%8</a> ,» [Ηλεκτρονικό].
[44]	« <a href="http://users.auth.gr/dkugiu/Teach/DataAnalysis/Chp6.pdf">http://users.auth.gr/dkugiu/Teach/DataAnalysis/Chp6.pdf</a> ,» [Ηλεκτρονικό].
[45]	« <a href="https://openclass.teiwm.gr/modules/document/file.php/BA-K117/4%CE%91%CE%BD%CE%AC%CE%BB%CF%85%CF%83%CE%B7%20%CE%A7%CF%81%CE%BF%CE%BD%CE%BF%CE%BB%CE%BF%CE%B3%CE%B9%CE%BA%CF%8E%CE%BD%20%CE%A3%CE%B5%CE%B9%CF%81%CF%8E%CE%BD.pdf">https://openclass.teiwm.gr/modules/document/file.php/BA-K117/4%CE%91%CE%BD%CE%AC%CE%BB%CF%85%CF%83%CE%B7%20%CE%A7%CF%81%CE%BF%CE%BD%CE%BF%CE%BB%CE%BF%CE%B3%CE%B9%CE%BA%CF%8E%CE%BD%20%CE%A3%CE%B5%CE%B9%CF%81%CF%8E%CE%BD.pdf</a> ,» [Ηλεκτρονικό].
[46]	« <a href="https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/">https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/</a> ,» [Ηλεκτρονικό].
[47]	Φ. Κουντούρη, « <a href="http://www2.aueb.gr/users/koundouri/resees/uploads/ECLC4.ppt">www2.aueb.gr/users/koundouri/resees/uploads/ECLC4.ppt</a> ,» 2008. [Ηλεκτρονικό].
[48]	« <a href="http://www.actuar.aegean.gr/notes/ΚΕΦ-4_ΕΠΟΧΙΚΗ_ΔΙΟΡΘΩΣΗ_ΚΑΙ_ΕΦΑΡΜΟΓΕΣ-23-1-16-SENT.pdf">http://www.actuar.aegean.gr/notes/ΚΕΦ-4_ΕΠΟΧΙΚΗ_ΔΙΟΡΘΩΣΗ_ΚΑΙ_ΕΦΑΡΜΟΓΕΣ-23-1-16-SENT.pdf</a> ,» [Ηλεκτρονικό].
[49]	« <a href="https://robjhyndman.com/hyndsight/cyclclcts/">https://robjhyndman.com/hyndsight/cyclclcts/</a> ,» [Ηλεκτρονικό].
[50]	« <a href="http://www.investopedia.com/terms/s/skewness.asp">http://www.investopedia.com/terms/s/skewness.asp</a> ,» [Ηλεκτρονικό].
[51]	« <a href="http://www.managedfuturesinvesting.com/managed-futures/news/aisource-news/2015/10/13/what-is-skewness">http://www.managedfuturesinvesting.com/managed-futures/news/aisource-news/2015/10/13/what-is-skewness</a> ,» [Ηλεκτρονικό].
[52]	« <a href="http://www.investopedia.com/terms/k/kurtosis.asp">http://www.investopedia.com/terms/k/kurtosis.asp</a> ,» [Ηλεκτρονικό].
[53]	« <a href="https://eclass.teicrete.gr/modules/document/file.php/DL183/%CE%A3%CF%84%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%B9%CE%BA%CE%AE%CE%99_5.pdf">https://eclass.teicrete.gr/modules/document/file.php/DL183/%CE%A3%CF%84%CE%B1%CF%84%CE%B9%CF%83%CF%84%CE%B9%CE%BA%CE%AE%CE%99_5.pdf</a> ,» [Ηλεκτρονικό].
[54]	« <a href="http://users.auth.gr/dkugiu/Teach/DataAnalysis/Chp6.pdf">http://users.auth.gr/dkugiu/Teach/DataAnalysis/Chp6.pdf</a> ,» [Ηλεκτρονικό].
[55]	« <a href="https://openclass.teiwm.gr/modules/document/file.php/BA-K117/4%CE%91%CE%BD%CE%AC%CE%BB%CF%85%CF%83%CE%B7%20%CE%A7%CF%81%CE%BF%CE%BD%CE%BF%CE%BB%CE%BF%CE%B3%CE%B9%CE%BA%CF%8E%CE%BD%20%CE%A3%CE%B5%CE%B9%CF%81%CF%8E%CE%BD.pdf">https://openclass.teiwm.gr/modules/document/file.php/BA-K117/4%CE%91%CE%BD%CE%AC%CE%BB%CF%85%CF%83%CE%B7%20%CE%A7%CF%81%CE%BF%CE%BD%CE%BF%CE%BB%CE%BF%CE%B3%CE%B9%CE%BA%CF%8E%CE%BD%20%CE%A3%CE%B5%CE%B9%CF%81%CF%8E%CE%BD.pdf</a> ,» [Ηλεκτρονικό].
[56]	« <a href="http://users.auth.gr/dkugiu/Teach/Econophysics/Chp4.pdf">http://users.auth.gr/dkugiu/Teach/Econophysics/Chp4.pdf</a> ,» [Ηλεκτρονικό].
[57]	« <a href="http://eclass.teipat.gr/eclass/modules/document/file.php/728131/%CE%98%CE%95%CE%A9%CE%A1%CE%99%CE%91/%CE%A0%CE%91%CE%A1%CE%9F%CE%A5%CE%A3%CE%99%CE%91%CE%A3%CE%95%CE%99%CE%A3%20%CE%9C%CE%91%CE%98%CE%97%CE%9C%CE%91%CE%A4%CE%A9%CE%9D%20%CE%98%CE%95%CE%A9%CE">http://eclass.teipat.gr/eclass/modules/document/file.php/728131/%CE%98%CE%95%CE%A9%CE%A1%CE%99%CE%91/%CE%A0%CE%91%CE%A1%CE%9F%CE%A5%CE%A3%CE%99%CE%91%CE%A3%CE%95%CE%99%CE%A3%20%CE%9C%CE%91%CE%98%CE%97%CE%9C%CE%91%CE%A4%CE%A9%CE%9D%20%CE%98%CE%95%CE%A9%CE</a> ,» [Ηλεκτρονικό].
[58]	« <a href="http://academics.epu.ntua.gr/LinkClick.aspx?fileticket=AsV8acFHn3A%3D&amp;tabid=380&amp;mid=838">http://academics.epu.ntua.gr/LinkClick.aspx?fileticket=AsV8acFHn3A%3D&amp;tabid=380&amp;mid=838</a> ,» [Ηλεκτρονικό].
[59]	« <a href="http://users.uom.gr/~drits/lessons/Lesson%204(MSc%20Inf).pdf">http://users.uom.gr/~drits/lessons/Lesson%204(MSc%20Inf).pdf</a> ,» [Ηλεκτρονικό].

[60]	« <a href="https://people.duke.edu/~rnau/411diff.htm">https://people.duke.edu/~rnau/411diff.htm</a> ,» [Ηλεκτρονικό].
[61]	« <a href="http://www.investopedia.com/articles/trading/07/stationary.asp">http://www.investopedia.com/articles/trading/07/stationary.asp</a> ,» [Ηλεκτρονικό].
[62]	« <a href="https://www.otexts.org/fpp/8/1">https://www.otexts.org/fpp/8/1</a> ,» [Ηλεκτρονικό].
[63]	« <a href="http://users.uom.gr/~drits/lessons/Lesson%203(MSc%20Inf).pdf">http://users.uom.gr/~drits/lessons/Lesson%203(MSc%20Inf).pdf</a> ,» [Ηλεκτρονικό].
[64]	« <a href="https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/">https://machinelearningmastery.com/gentle-introduction-box-jenkins-method-time-series-forecasting/</a> ,» [Ηλεκτρονικό].
[65]	N. Σαριαννίδης, «Οικονομετρία - Ανάλυση Χρονολογικών Σειρών», 2015.
[66]	K. N. R. C. G. P. B. M. Venkata Sasank Pagolu, «Sentiment Analysis of Twitter Data for Predicting Stock Market Movements,» <i>International conference on Signal Processing, Communication, Power and Embedded System (SCOPE5)</i> , 2016.
[67]	B. A.-h. a. M. Abbod, «Predicting Financial Time Series Data,» 2016.
[68]	« <a href="https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/">https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/</a> ,» [Ηλεκτρονικό].
[69]	A. C. M. P. R. A. d. O. David M. Q. Nelson, «Stock Market's Price Movement Prediction With LSTM Neural Networks.»
[70]	« <a href="http://colah.github.io/posts/2015-08-Understanding-LSTMs/">http://colah.github.io/posts/2015-08-Understanding-LSTMs/</a> ,» [Ηλεκτρονικό].
[71]	S. H. A. K. R. S. S. A. A. Mehak Usmani, «Stock Market Prediction Using Machine Learning Techniques,» <i>3rd International Conference On Computer And Information Sciences (ICCOINS)</i> , 2016.
[72]	« <a href="https://pandas.pydata.org/pandas-docs/stable/">https://pandas.pydata.org/pandas-docs/stable/</a> ,» [Ηλεκτρονικό].
[73]	« <a href="http://www.numpy.org/">http://www.numpy.org/</a> ,» [Ηλεκτρονικό].
[74]	« <a href="http://scikit-learn.org/stable/tutorial/basic/tutorial.html">http://scikit-learn.org/stable/tutorial/basic/tutorial.html</a> ,» [Ηλεκτρονικό].
[75]	« <a href="https://matplotlib.org/">https://matplotlib.org/</a> ,» [Ηλεκτρονικό].
[76]	« <a href="http://www.statsmodels.org/stable/index.html">http://www.statsmodels.org/stable/index.html</a> ,» [Ηλεκτρονικό].
[77]	« <a href="https://pandas.pydata.org/pandas-docs/stable/10min.html#min">https://pandas.pydata.org/pandas-docs/stable/10min.html#min</a> ,» [Ηλεκτρονικό].
[78]	« <a href="http://www.investopedia.com/terms/c/cumulativereturn.asp">http://www.investopedia.com/terms/c/cumulativereturn.asp</a> ,» [Ηλεκτρονικό].
[79]	S. Madge, «Predicting Stock Price Direction using Support Vector,» 2015.
[80]	« <a href="http://www.investopedia.com/terms/l/least-squares-method.asp">http://www.investopedia.com/terms/l/least-squares-method.asp</a> ,» [Ηλεκτρονικό].
[81]	« <a href="http://www.investopedia.com/articles/technical/02/121702.asp">http://www.investopedia.com/articles/technical/02/121702.asp</a> ,» [Ηλεκτρονικό].
[82]	« <a href="https://coolstatsblog.com/2013/08/14/using-aic-to-test-arima-models-2/">https://coolstatsblog.com/2013/08/14/using-aic-to-test-arima-models-2/</a> ,» [Ηλεκτρονικό].
[83]	« <a href="http://scikit-learn.org/stable/modules/model_evaluation.html#mean-squared-error">http://scikit-learn.org/stable/modules/model_evaluation.html#mean-squared-error</a> ,» [Ηλεκτρονικό].
[84]	« <a href="https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/">https://machinelearningmastery.com/metrics-evaluate-machine-learning-algorithms-python/</a> ,» [Ηλεκτρονικό].
[85]	« <a href="http://www.investopedia.com/university/stocks/stocks1.asp">http://www.investopedia.com/university/stocks/stocks1.asp</a> ,» [Ηλεκτρονικό].
[86]	« <a href="http://scikit-learn.org/stable/tutorial/basic/tutorial.html">http://scikit-learn.org/stable/tutorial/basic/tutorial.html</a> ,» [Ηλεκτρονικό].
[87]	« <a href="http://scikit-learn.org/stable/modules/svm.html#svm">http://scikit-learn.org/stable/modules/svm.html#svm</a> ,» [Ηλεκτρονικό].
[88]	« <a href="http://scikit-learn.org/stable/modules/svm.html">http://scikit-learn.org/stable/modules/svm.html</a> ,» [Ηλεκτρονικό].
[89]	« <a href="http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html#sklearn.model_selection.TimeSeriesSplit">http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.TimeSeriesSplit.html#sklearn.model_selection.TimeSeriesSplit</a> ,» [Ηλεκτρονικό].
[90]	« <a href="http://scikit-learn.org/stable/modules/classes.html#classification-metrics">http://scikit-learn.org/stable/modules/classes.html#classification-metrics</a> ,» [Ηλεκτρονικό].
[91]	« <a href="http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html#sphx-glr-auto-examples-model-selection-plot-precision-recall-py">http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html#sphx-glr-auto-examples-model-selection-plot-precision-recall-py</a> ,» [Ηλεκτρονικό].
[92]	R. R. Phayung Meesad, «Predicting Stock Market Price Using Support Vector Regression,» May 2013.