Stock Market Prediction Using Analytic Hierarchy Process and Support Vector Machine

Vaman Ashqi Saeed

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science
in
Computer Engineering

Eastern Mediterranean University
January 2017
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Mustafa Tümer
Director

I certify that this thesis satisfies the requirements as thesis for the degree of Master of Science in Computer Engineering.

Prof. Dr. Işık Aybay
Chair, Department of Computer Engineering

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Computer Engineering.

Asst. Prof. Dr. Adnan Acan
Supervisor

Examining Committee

1. Assoc. Prof. Dr. Mehmet Bodur
2. Asst. Prof. Dr. Adnan Acan
3. Asst. Prof. Dr. Ahmet Ünveren

ABSTRACT

Prediction of the stock market behavior has been a research topic for decades. Because it is a challenging subject both in terms of the choice of the prediction model and in terms of constructing the set of features that model will use for forecasting. In this thesis, a novel feature ranking and feature selection approach incorporation with weighted kernel least squares support vector machines (LS-SVMs) were used. We introduce the analytic hierarchy process (AHP) into the stock market and then evaluate criteria which provide the prediction model with relevant knowledge of the underlying processes of the studied stock market. The feature weights obtained by the AHP method are applied for feature ranking and selection and used with the LS-SVMs through a weighted kernel. The experimental results specify that the new model outperforms the benchmark models. Furthermore, the set of feature weights obtained by the new approach can also independently be incorporated into other kernel-based learners.

Keywords: stock market prediction, analytic hierarchy process, support vector machine, least squares support vector machines, weighted kernel.

ÖZ

Borsa davranışının tahmini çeyrek yüzyıl boyunca bir araştırma konusu olmuştur. Çünkü, tahmin modelinin seçimi ve modelin kullanacağı özellikler kümesimin inşası açılarından borsa davranışının tahmini iddialı bir konudur. Bu tezde, yeni bir özellik sıralama ve özellik seçme yöntemi, ağırlılı çekirdek en küçük kareler destek vektör makineleri (LS-SVM) beraberinde kullanılmıştır. Analitik hiyerarşi süreci (AHP) yöntemini borsa özellik seçimi için kullandık ve kullanılan borsa verileri için ilgili bilgiye dayanan tahmin modeli için kriterleri değerlendirdik. AHP ile elde edilen özellik ağırlıkları LS-SVM'in ağırlıklı çekirdek yaklaşımı aracılığıyla özellik sıralama ve seçimi için kullanıldı. Deneysel sonuçlar kullanılan modelin ölçüt modellerden daha başarılı olduğunu göstermiştir. Buna ek olarak, yeni yöntemle elde edilen özellik ağırlıkları başka çekirdek tabanlı sistemlere bağımsız olarak eklenebilir.

Anahtar kelimeler: Borsa tahmini, analitik hiyerarşi süreç, destek vector makineleri, en küçük kareler destek vector makineleri, ağırlıklı çekirdek.

DEDICATION

First I dedicate this study to my parents **Mr. Ashqi & Mrs. Fakhriya** for their understanding and for their overwhelming, support morally and financially.

To my loving husband and son **Amar Mustafa and Mohammed Mahdi**, my eternal gratitude.

To my **brothers and sisters** for their friendship and eternal love.

To my beloved uncles **Mr. Anwer Saeed and Mr. Hassen Saeed** for their help and support.

To my best friend **Heleen Amad J.** for her help and love.

ACKNOWLEDGMENT

I am sincerely thankful to my supervisor, Asst. Prof. Dr. Adnan Acan, whose encouragement and support from the beginning to the concluding enabled me to develop an understanding of the subject.

I would like to thank Asst. Prof. Dr. Amad Muhammed Saleh Jubrail (Rector of Nawroz University) for his help support and great thanks to Asst. Prof. Dr. Nilgun Hancioglu Eldridge as well for her help and advice.

Most importantly, my deep gratitude to my best friend and soul mate Ammar Mustafa who was the reason for motivating me to continue my study .

Finally, I would like to thank the members of my family especially my beloved Mother who assisted me with this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Background of the Study

Stock Market Prediction is one of the most widely studied and difficult problems, appealing to researchers from various fields such as economics, history, finance, mathematics, and computer science. The unpredictable nature of the stock market makes it tough to stratify simple time-series or regression methods. Financial establishments and tradesmen have created different proprietary forms to attempt and overcome the market for themselves or their customers, however seldom has anyone accomplished regularly higher than average incomes on investment. Still, the challenge of stock market prediction is so tempting because an enhancement of just a few percentage points can raise profit by millions of dollars for these associations [1].

The stock is a kind of safety that implies an ownership situation in a firm. A corporation can be separated into a number of shares and every share of stock is labeled to a commensurate share of gain or loss created by the corporation. It is a delegate of the pretension as member of the company’s assets and wages. There are various choices for people who desire to make investments. The buying and selling of stock is always the most accepted choice for general commerce. Once stockholders purchase stocks, they turn out to be a shareholder, which means that they possess a part of the firm. If the corporation's earnings rise, they will share those increased earnings with the firm. Likewise, if the corporation's profits drop, the stock price drops respectively and the loss in earnings will be shared with investors as well. The logic to produce money is that financier purchases the stock, holds it for a specific period, after that sells it at a higher value than the buying price. If they sell their stock at a price less than the value they have waged for it, they will lose money.

It is widely known that stock price is very changeable, even on a daily basis. The reason for that is because of supply and request. In stock markets, a large volume of stocks is dealt with each day. If there are more people who buy a stock than the people who sell it, out of the anticipation that the price will go up in the future, then the price will increase. On the contrary, if more people want to sell it than to buy it, the stock price will drop radically. Nevertheless, investors' anticipation for the market is in a permanent case of variability due to all types of information acquired over time that powerfully impact on their decision-making. That's also why the stock market has been treated so often over a short period of time [2].

Financial time sequences' forecasting has been studied since 1980s. The aim is to beat financial markets and gain higher income. Up to now, pecuniary estimating is still considered as one of the most challenging applications of modern time series forecasting. Pecuniary period sequences have very composite conduct, subsequent from a vast number of aspects which might be economic, political, or psychological. They are intrinsically noisy, non-stationary, and deterministically confused [2].

1.2 Statement of the Problem

Before a stockholder invests in any stock, he must be conscious of how the stock market acts. Participating in an upright stock but at a corrupt period can have tragic outcomes, whereas exploitation in a middling stock at the accurate period can bear incomes. Financial investors of nowadays are facing this problem of trading as they do not correctly understand as to which stocks to buy or which stocks to sell to acquire ideal earnings. Examining news and additional facts about a specific stock previous to capitalizing is fundamental. But in today's world, we are overloaded by enormous bases of evidence such as in periodicals, correspondents, accessible nourishes etc. Examining entire data exclusively or physically is extremely problematic. Hence, computerization of the progression is essential.

Intelligent Investors utilize machine learning techniques in forecasting the stock market behavior which provides more accurate results than analysis of numerical period sequences alone. This will tolerate pecuniary forecasters to anticipate the performance of the stock that they are absorbed in and consequently act in view of that [3].

Another crucial point that most of the stockholders are not conscious of is the determination of sufficient and required features that are necessary for training a good prediction model. If the number of features is insufficient, the prediction accuracy of the model will be poor, and the model may be prone to under fitting [4]. Oppositely, if we have too many features, the information that they provide for the model could be unnecessary or redundant. Consequently, the model might possibly have a poor generalization performance and may be prone to over-fitting [5]. The most significant issue in the construction of a stock market prediction model is the selection of input features for predictors, where the selection of suitable methods for feature subset selection is extremely relevant. The aim of stock market forecast is to develop a market prediction model that can successfully foresee markets trend direction which enables the individual stockholders to have a priori knowledge of the market trend so as to gain profit and reduce the risk involved [6].

In the business and economic environment, it is very important to predict different types of pecuniary variables to develop proper strategies and avoid the threat of possibly large losses. The forecast of a diversity of monetary directories has a deep influence on the development of the macro economy. Particularly, in the case of stock markets, the task becomes more significant because of the dynamic changes of the market behavior and immeasurable commercial benefits. According to the prediction of stock market indices, risk manager, and practitioners can realize whether their portfolio will decay in the future and they may want to sell it before it becomes depreciated. Consequently, the research of forecasting the future trends of pecuniary indices is important and essential for persons who are interested in the stock markets. However, the behavior of stock markets relies on several factors such as governmental, monetary, normal causes and numerous others. The stock markets are active and reveal extensive difference, and the expectation of the stock market is an extremely stimulating job because of the vastly nonlinear environment and complicated dimensionality [7].

1.3 Purpose of the Study

The aim of this study is to provide a suitable stock trend prediction model using analytic hierarchy process for feature ranking and selection integrated with support vector machines.

1.4 Significance of the Study

The stock market prediction has been studied over and over to extract useful patterns and predict their movements. The stock market prediction has always been a definite attraction for examiners and fiscal depositors. The reason is that people can beat the market, can gain additional profit. Financial analysts who invest in stock markets generally are not conscious of the stock market attitude. They are fronting the trouble of stock exchange as they do not know which shares to purchase and which to vend to earn extra incomes. If they can expect the upcoming attitude of stock prices, they can work instantly towards it and make an income. As a result of this, financial analysts and investors will have a decision support tool or as an autonomous artificial trader that can be extended with any interface to the stock exchange [2].

Chapter 2

RELATED WORKS

2.1 Introduction

The key to generating a high return on the stock market lies in how well we are able to effectively forecast the future movement of pecuniary asset prices. The stock market index as a hypothetical portfolio of selected stocks is generally utilized to measure the performance of both the overall stock market and a particular sector. Consequently, a market trading strategy can be considered effective only if it relies on the precise prediction of the trend of change of the index value of that particular market.

Stock market trend forecast represents a challenge for science both in terms of the choice of methodology and in terms of the theoretical basis of its application. To address these problems, machine learning models, among which the most popular were Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), were the most frequently applied alternatives to the classical statistical models in the area of pecuniary forecasting during the last two decades. Due to the principles of the weak form of the efficient market hypothesis (EMH), the behavior of pecuniary asset prices is often governed by a random walk process; thus, the degree of accuracy of an approximate 60 % hit rate obtained in prediction using various machine learning techniques is often considered a satisfactory result for stock market trend Prediction [8].

As explained in Chapter 1, the purpose of this study is the prediction of stock trend movement using analytic hierarchy process and support vector machine. In this chapter the major groundwork and preliminaries related to the subject of the study is going to be reviewed.

2.2 Analytic Hierarchy Process

Analytic hierarchy process (AHP) is a method of selection between sets of factors based on their relevance in terms of meeting even opposing criteria. The AHP calculation techniques are used on a designed Pairwise Comparison Matrix (PCM) to obtain the eigenvector which represents relative feature values for the obtained criterion. The pairwise comparison is represented using the Fundamental 1–9 Scale, as shown in Table 2.1.

Table 2.1: The Fundamental Scale of absolute numbers [9]

Intensity of importance	Definition	Explanation
1	Equal importance	Two activities contribute equally to the objective
3	Moderate importance	Experience and judgment slightly favour one activity over another
5	Strong importance	Experience and judgment strongly favour one activity over another
7	Demonstrated—very strong importance	An activity is favoured very strongly over another; its dominance demonstrated in practice
9	Extreme importance	The evidence favouring one activity over another is of the highest possible order of affirmation

In Table 2.1, for reciprocals of above, if activity i has one of the above non-zero numbers assigned to it when compared with activity j, then j has the reciprocal value when compared with i. Providing us a reasonable assumption, as well as when we have a criteria within this range (1.1–1.9), specifies that activities are very close, this may be tough to assign the best value but when compared with other contrasting activities the size of the small numbers would not be too noticeable, however they can still indicate the relative importance of the activities [9].

To make a decision in AHP in an organized way we need to decompose the decision into the following steps.

- Define the problem and determine the kind of knowledge sought.
- Structure the decision hierarchy from the top with the goal of the decision, then the objectives from a broad perspective, through the intermediate levels (criteria on which subsequent elements depend) to the lowest level (which usually is a set of the alternatives).
- Construct a set of pairwise comparison matrices. Each element in an upper level is used to compare the elements in the level immediately below with respect to it.
- Use the priorities obtained from the comparisons to weigh the priorities in the level immediately below. Do this for every element. Then for each element in the level below, add its weighed values and obtain its overall or global priority. Continue this process of weighing and adding until the final priorities of the alternatives in the bottom most level are obtained [9].

The successful application of AHP in various empirical data analysis, which is the result of the clarity of its underlying mathematical principles and its ability to evaluate decision-making consistency, has led to it being used on stock market data in this thesis.

2.3 Support Vector Machine

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm constructs a model that assigns new samples to one category or the other, making it a non-probabilistic binary linear classifier. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces. Moreover, when data are not labeled, supervised learning is not possible, and an unsupervised learning approach is required, which attempts to find natural clustering of the data to groups, and then map new data to these formed groups [10].

Chapter 3

THE RESEARCH PROCEDURES

In this section, we describe the feature selection procedure used in this study, the AHP basic calculations, as well as the algorithm for determining feature weights by applying AHP. In addition, a brief introduction of the technical indicators will be presented, followed by an introduction of the leading indicators in this thesis. Then, a simple description of the SVM algorithm is provided including LS-SVM. Finally, because selecting important features in non-linear kernel spaces is a difficult challenge in both classification and regression problems, therefore the basics of weighted kernels are been presented in relation to SVM and LS-SVM.

3.1 AHP Evaluation Criteria

In Chapter 1 and Chapter 2, it has been clarified that the crucial role of feature ranking and feature selection for stock market forecasting is to help the financial analysts and researchers to provide a prediction model with a priori knowledge of the underlying processes of the observed stock market. First, we introduce AHP evaluation criteria for the valuation of the relevance of technical indicators. Therefore, we suggest the construction of technical trading strategies as a measure of the success of each technical indicator relied on [11].

Second, a technical Trading Strategy (S_T) is composed of a set of trading rules that are applied to create trading signals. Overall, commonly used trading systems depend on one or two technical indicators that define the timing of trading signals.

The AHP evaluation criteria are twofold. The first group consists of two criteria utilized to measure the economic relevance of the chosen indicators: cumulative gross return, like a measure of stock market profitability, and systematic risk as a measure of market volatility. The third criterion denotes a comparison of the trading signals created with a trading strategy and the signals generated based on actual stock market index values, in relation to their achieved prediction accuracy. The mentioned criterions are illustrated below with their calculations.

3.1.1 Return Evaluation

Returns on investments in the case of a specific stock market index were calculated as the variances between daily index values presented in national currency, multiplied by the generated trading signal for the present day. Gross returns were defined as the cumulative capital earnings for a specified period of time, as follows:

$$R = \sum_{t=1}^{n} S_T * (CP_t - CP_{t-1}) \tag{1}$$

where S_T denotes the trading signals produced by the trading strategy. *CP* represents closing price and t =1, 2, …, n. and n represents number of days in a selected time period. The calculated return on investment value allows us to compare the selected set of technical indicators. For the evaluation criteria, we created a relative weighting function which ascribes AHP scale values to the obtained returns, taking into consideration the min–max range of the resulting calculations. The same function is applied in the calculations of the following two criteria [12].

3.1.2 Risk Evaluation

In addition to return, risk was introduced as one of the evaluation criteria in the AHP analysis into stock market prediction, since in stock trading the return is balanced with a proper level of risk Systematic risk, in relation to return, is defined as:

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{t=1}^{n}(R_t - \bar{R})} \tag{2}$$

where $\bar{R}$ represents the mean value of the gross return R in a selected time period t. And n represents number of days.

3.1.3 Accuracy Evaluation

For the evaluation of the prediction effect as a general measure, the Hit Ratio (HR) was used. HR was computed based on the number of properly generated trading signals within the test group:

$$HR = \frac{1}{m}\sum_{i=1}^{m} PO_i \tag{3}$$

where PO_i is the prediction output of the i th trading day. PO_i equals *1* if it is the actual value for the i th trading day; otherwise, PO_i equals *0*, and m is the number of data in the used data set [13].

3.2 Basics of AHP Calculations

The AHP calculations can be summarized as follows: compare n elements, A_1 ... A_n and define the significance of A_i with respect to A_j by p_{ij} to form a reciprocal matrix $P = (p_{ij})_{nxn}$ with the implication that $p_{ij} = 1/\ p_{ji}$ for $i \neq j$ and $p_{ii} = 1$. For accurately measured data, the P_{ij} matrix is transitive and the eigenvector ω of the order n can be calculated such that $P_\omega = \lambda_\omega$, where λ is an eigenvalue. In practice, the first step is to supply an initial matrix for the pairwise criteria comparisons to obtain an eigenvector, named as the Relative Value Vector (RVV). Next, for each criterion, we need a pairwise comparison matrix (PCM) to show the performance of each criterion. Then, the evaluation of the Option Performance Matrix (OPM) enables us to present the observed features in terms of the selected criteria. The final step is the multiplication of the RVV and the OPM, to obtain the overall ranks.

As a result of the inconsistency of the decision-making process, the ω vector generally satisfies the equation $P_{\omega} = \lambda_{max}\omega$ and $\lambda_{max} \geq n$. The relationship between λ_{max} and n determines the level of (in)consistency of the decisions, where equality between the two is an indication of consistency. A Consistency Index (CI) is calculated as $(\lambda_{max} - n)/(n - 1)$ and needs to be determined in relation to a corresponding Random consistency Index (RI), which leads to the calculation of the Consistency Ratio (CR) as follows: CI/RI. It is established that a CR exceeding 0.1 indicates inconsistent decisions, while a CR of *0* indicates perfectly consistent decisions [9].

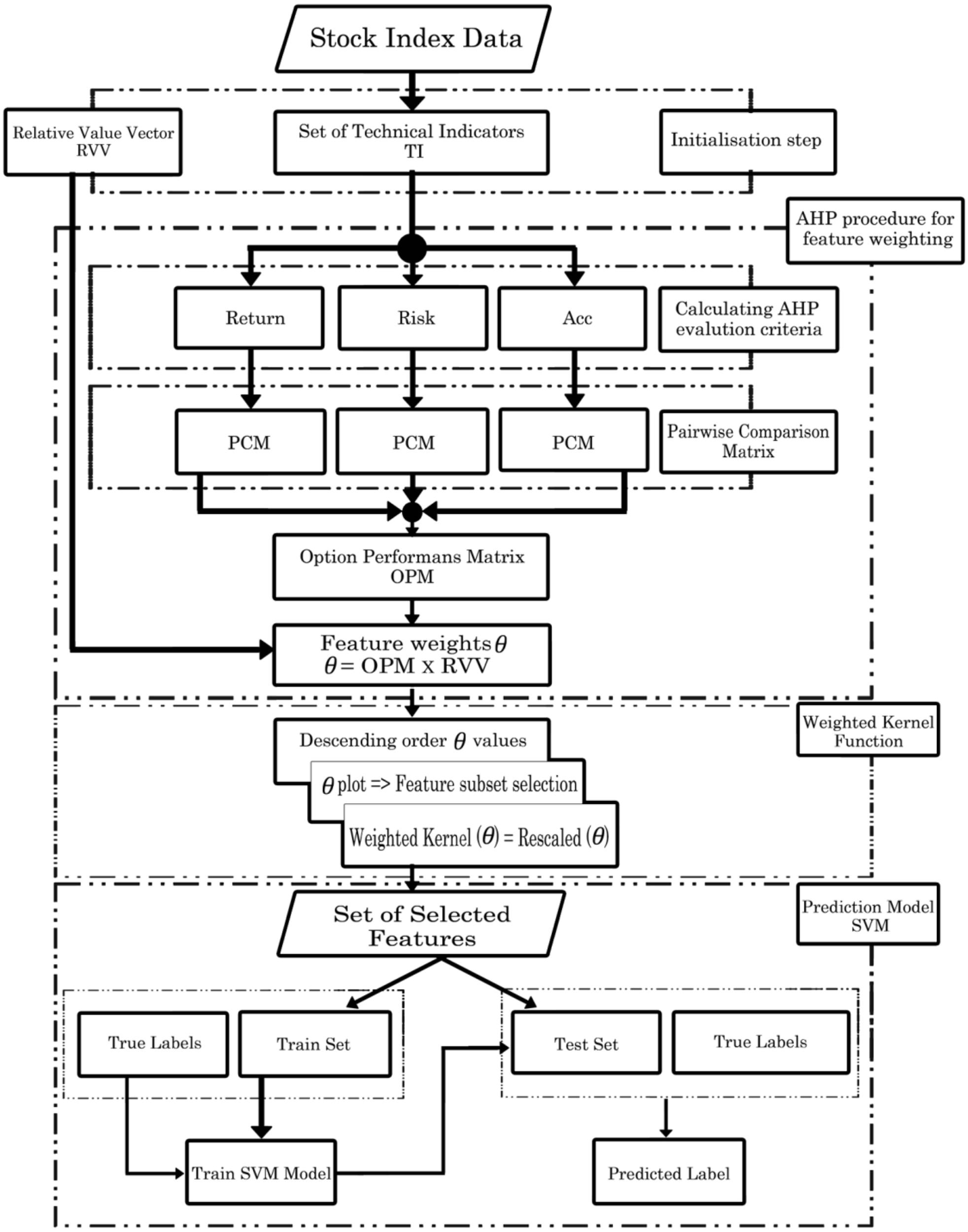


Figure 3.1: Algorithm for the Prediction Model [23]

The approach for the selection of subsets of the features in accordance with the AHP evaluations and SVM prediction model is shown in Fig. 3.1

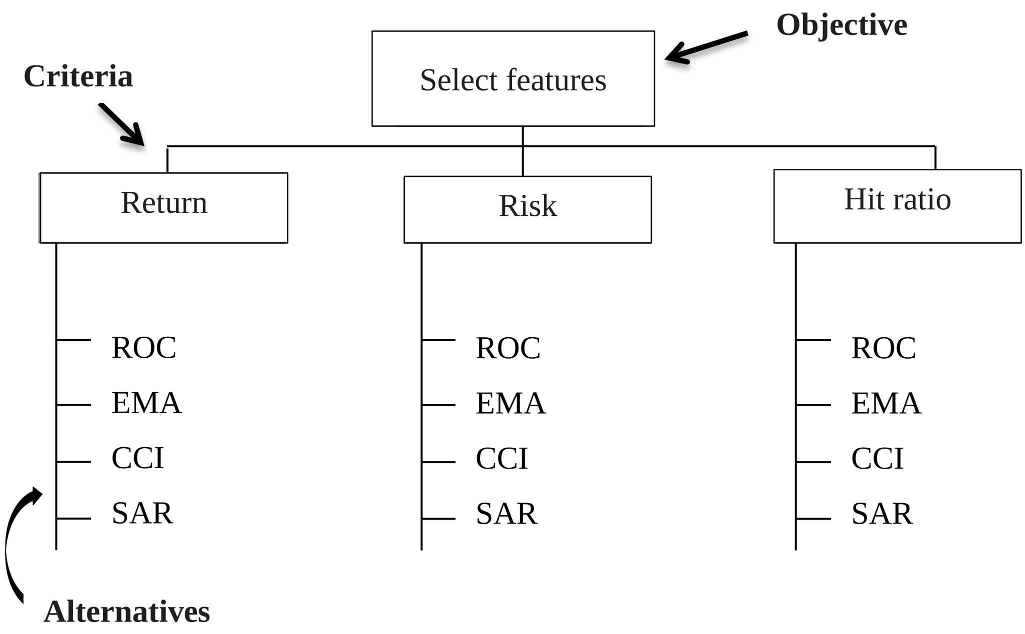
3.3 Determining feature weights by AHP

The first step in the algorithm is the calculation of the criterion values for AHP evaluation. After forming the initial set of technical indicators, for the technical indicators, calculate values of the evaluation criterion: return, systematic risk and prediction accuracy. The RVV is computed by the methods described in Sect. 3.2. Then three PCM are built. The weights in the matrices reflect how the technical indicators accomplish in terms of each criterion. According to Sect. 3.2, we then create the OPM, and in the next step multiply the RVV and the OPM to obtain the whole feature weights. The weights (θ) define the relative significance (ranking) of every input technical indicator candidate in relation to the criterion values. The next step is the ordering of the set of technical indicators in descending order according to θ values. The goal of this step is to find a feature subset that will be used for the prediction model. More precisely, if one plots the weights, the technical indicator that corresponds to the largest weight will add the most information to the prediction model. At some point the feature relevance will decrease, leading to what is known as an "angle" effect in the plot (see Fig. 4.3). The estimated feature weights for selected features should proportionally be rescaled in accordance with the constraints defined in (25). In the last step, kernel weighting is performed by feature multiplication with rescaled weights, within the input feature space.

A detailed example of AHP application based on the feature values used in this study is illustrated below. First an important part of the process is to accomplish these three steps:

- State the objective:
    - Select features
- Define the criteria:
    - Cumulative gross return, systematic risk, hit ratio,
- Pick the alternatives:
    - ROC, EMA, CCI, SAR

This information is then arranged in a hierarchical tree as bellow:



The information is then synthesized to determine relative rankings of alternatives and both qualitative and quantitative criteria can be compared using informed judgments to derive weights and priorities, in this example the judgments are as following:

- Risk is 4 times as important as return.
- Hit ratio is 4 times as important as risk.
- Hit ratio is 6 times as important as return.

Next, using pairwise comparisons matrix, the relative importance of one criterion over another can be expressed as shown in Table 2.1.

	Return	Risk	Hit ratio
Return	1/1	1/4	1/6
Risk	4/1	1/1	1/4
Hit ratio	6/1	4/1	1/1

In order to turn this matrix into ranking of criteria the eigenvector must be used as (Dr. Thomas Saaty) the developer of AHP demonstrated mathematically that the eigenvector solution was the best approach [9].

Eigenvectors are a special set of vectors associated with a linear system of equations (i.e., a matrix equation) that are sometimes also known as characteristic vectors, proper vectors, or latent vectors. Steps for solving the eigenvector:

- A short computational way to obtain this ranking is to raise the pairwise matrix to powers that are successively squared each time.
- The row sums are then calculated and normalized.
- When the difference between these sums in two consecutive calculations is smaller than a prescribed value we stop.

To illustrate the above points more clearly we solve our algebra matrix in details. By converting fractions to decimals.

Step 1: squaring the matrix

$$\begin{bmatrix} 1.0000 & 0.2500 & 0.1666 \\ 4.0000 & 1.0000 & 0.2500 \\ 6.0000 & 4.0000 & 1.0000 \end{bmatrix}$$

This time

$$\begin{bmatrix} 1.0000 & 0.2500 & 0.1666 \\ 4.0000 & 1.0000 & 0.2500 \\ 6.0000 & 40000 & 1.0000 \end{bmatrix}$$

This

The result is:

$$\begin{bmatrix} 3.0000 & 0.1750 & 8.0000 \\ 5.3332 & 3.0000 & 14.0000 \\ 1.1666 & 0.6667 & 3.0000 \end{bmatrix}$$

Step 2: computing first eigenvector (to four decimal places) first, we sum the rows.

$$\begin{bmatrix} 3.0000 & + & 0.1750 & + & 8.0000 \\ 5.3332 & + & 3.0000 & + & 14.0000 \\ 1.1666 & + & 0.6667 & + & 3.0000 \end{bmatrix} \begin{array}{ll} = 12.7500 & 0.3194 \\ = 22.3332 & 0.5595 \\ = \underline{4.8333} & \underline{0.1211} \end{array}$$

Second, we sum the row totals $\longrightarrow$ 39.9165 1.0000

Finally, we normalize by dividing the row sum by the row totals (i.e. 12.7500 divided by 39.9165 equals 0.3194).

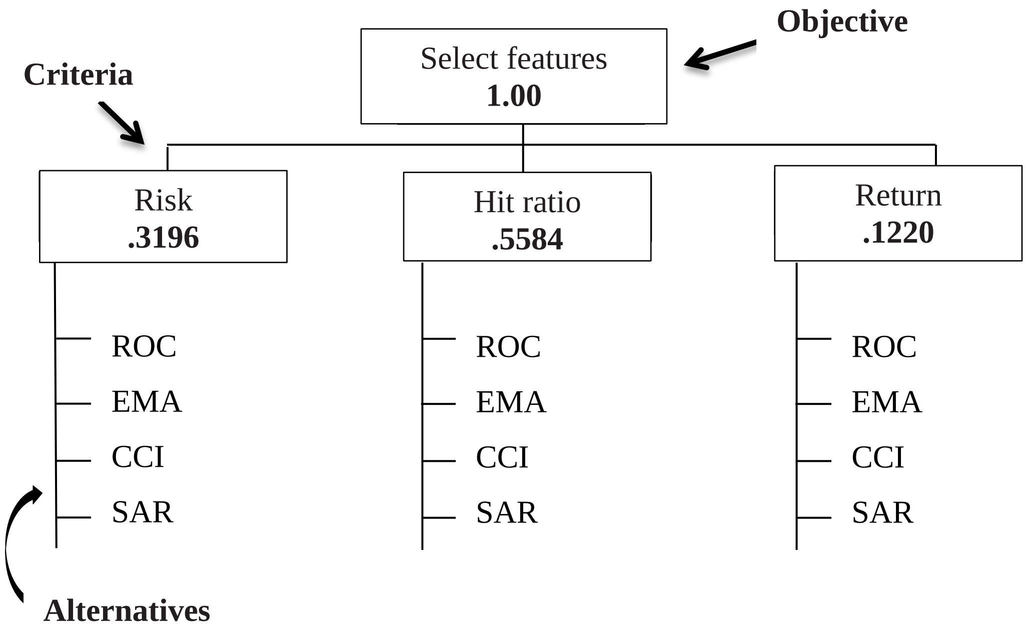
The result eigenvector:

$$\begin{bmatrix} 0.3194 \\ 0.5595 \\ 0.1211 \end{bmatrix}$$

This process must be iterated until the eigenvector solution does not change from the previous iteration. Consequently, after two iterations the comparison matrix and the computed eigenvector we obtain the relative ranking of the criteria as follows:

Risk	0.3196	← The second most important criteria
Hit ratio	0.5584	← The most important criterion
Return	0.1220	← The least important criterion

Moreover, we apply the computed criteria weights to Hierarchal tree.



In terms of return, pairwise comparisons determine the preference of each alternative over another is illustrated below.

Return

	ROC	EMA	CCI	SAR
ROC	1/1	1/4	4/1	1/6
EMA	4/1	1/1	4/1	1/4
CCI	1/4	1/4	1/1	1/5
SAR	6/1	4/1	5/1	1/1

Then in terms of Hit ratio, pairwise comparisons determine the preference of each alternative over another.

Hit ratio

$$
\begin{array}{c|cccc}
 & \text{ROC} & \text{EMA} & \text{CCI} & \text{SAR} \\
\hline
\text{ROC} & 1/1 & 2/1 & 5/1 & 1/1 \\
\text{EMA} & 1/2 & 1/1 & 3/1 & 2/1 \\
\text{CCI} & 1/5 & 1/3 & 1/1 & 1/4 \\
\text{SAR} & 6/1 & 1/2 & 4/1 & 1/1
\end{array}
$$

Accordingly the matrix algebra will be represented followed by computing the eigenvector to determine the relative ranking of alternatives under each criterion.

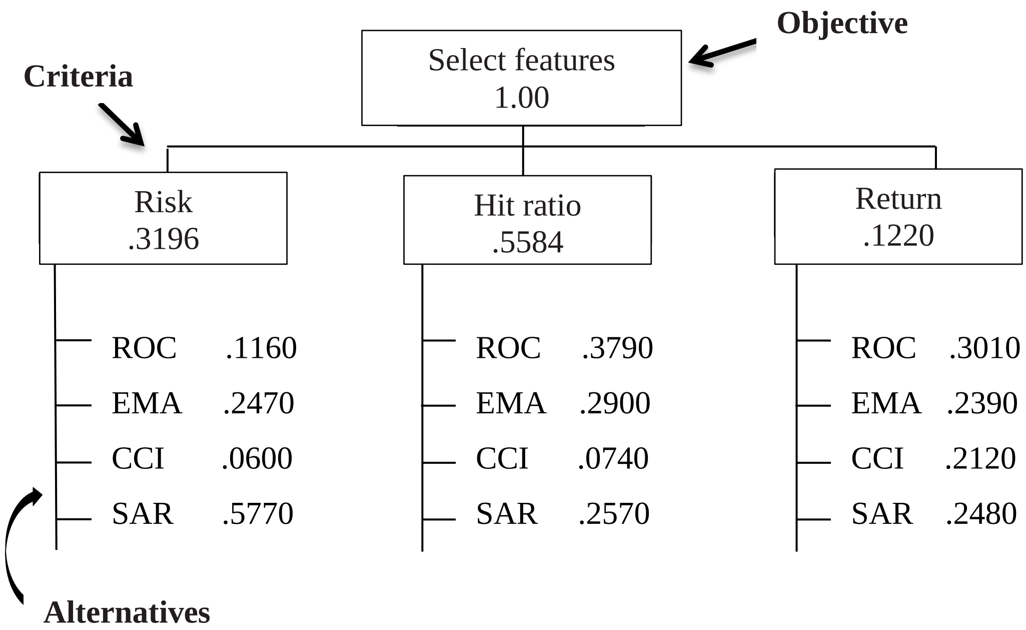
Ranking		**Return**
3	ROC	.1160
2	EMA	.2470
4	CCI	.0600
1	**SAR**	.5770

Ranking		**Hit ratio**
1	**ROC**	.3790
2	EMA	.2900
4	CCI	.0740
3	SAR	.2570

As stated earlier, AHP can combine both qualitative and quantitative information. As a result risk information is obtained for each alternative:

		Risk	
ROC	34	34/113 =	.3010
EMA	27	27/113 =	.2390
CCI	24	24/113 =	.2120
SAR	28	28/113 =	.2480
	113		**1.0000**

After normalizing the risk info allows us to use it with other rankings. Next the tree with all the weights is shown below:



Finally, the process is accomplished by multiplying the alternatives with ranking criteria to obtain the final ranking of our alternatives as shown below.

$$\begin{array}{cc} & \begin{array}{ccc} \text{Risk} & \text{Return} & \text{Hit ratio} \end{array} \\ \begin{array}{c} \text{ROC} \\ \text{EMA} \\ \text{CCI} \\ \text{SAR} \end{array} & \begin{bmatrix} .1160 & .3790 & 3010 \\ .2470 & .2900 & .2390 \\ .0600 & .0740 & .2120 \\ .5770 & .2570 & .2480 \end{bmatrix} \end{array} * \begin{array}{cc} \text{Criteria ranking} & \\ \begin{bmatrix} .3196 \\ .5584 \\ .1220 \end{bmatrix} & \begin{array}{l} \text{Risk} \\ \text{Return} \\ \text{Hit ratio} \end{array} \end{array}$$

i.e. for the roc (.1160 * .3196) + (.3790 * .5584) + (.3010 * .1220) = .3060

$$= \begin{bmatrix} .3060 \\ .2720 \\ .0940 \\ .3280 \end{bmatrix}$$

According to the outcomes the EMA is the highest ranked indicator. In summary, the AHP provides a logical framework to determine the benefits of each alternative [25].

1. EMA .3280
2. CCI .3060
3. ROC .2720
4. SAR .0940

3.4 Stock Market Indicators

3.4.1 Introduction

An indicator may be defined as a series of data points that are derived from the price of a security by applying a basic formula. Price data is a combination of open, close, high, or low over a period of time. For example, the average of 3 closing prices is one data point ((41+43+43)/3=42.33). However, one data point does not offer much information and does not make an indicator. A series of data over a period of time is required to create valid reference points to enable analysis. By creating a time series of data points, a comparison can be made between present and past levels. An indicator offers a different perspective from which to analyze the price action [14].

The function of indicators may be classified into three categories: to alert, to confirm, and to predict. An indicator can act as an alert to study price action a little more closely. If information is waning, it may be a signal to watch for a break of support. Or, if there is a large positive divergence building, it may serve as an alert to watch for a resistance break-out [15].

3.4.2 Classification of Indicators

Indicators are mathematical/ statistical functions that are applied over stock properties such as close, high, low and volume. These indicators are broadly classified into the following important categories:

- Market Momentum Indicators.
- Market Volatility Indicators.
- Market Trend Indicators.
- Broad Market indicator.
- General Momentum Indicator.

Analysts generally use at least one indicator from each of these categories for their forecasts. The indicator is generally chosen by evaluating the accuracy of the model [14].

3.4.3 Basic Indicators

The following are a set of potential input features. In this study, we rely on the most commonly used technical indicators:

3.4.3.1 Relative Strength Index

"Relative Strength Index (RSI) is a measure of the strength that is intrinsic in a field and is calculated using the amount of upward and downward changes over a given period of time. It has a range of 0 to 100 with values typically remaining between 30 and 70" [16]. Overbought conditions are indicated by higher values of the RSI while lower values indicate oversold conditions the formula for computing the RSI is as follows.

$$RSI = 100 - [\frac{100}{(1 + RS)}] \tag{4}$$

where

- *RSI* is Relative Strength Index
- *RS* is Average of *x* days' up closes Average of *x* days' down closes.

In addition the value is defined as 100 when no download price changes occur during the period of the calculation.

3.4.3.2 Stochastic Oscillator (%*K)*

"Stochastic Oscillator(SO) *%K* is an indicator that predicts the price turning points by comparing a security's closing price to its price range over a given time period" [16]. The formula for this computation is as follows.

$$\%K = 100\ (\frac{CP_t - LP}{HP - LP}) \tag{5}$$

where

- $\%K$ is Stochastic Oscillator.
- CP_t is a recent closing price.
- LP is the lowest low price during the period.
- HP is the highest high price during the period.

3.4.3.3 Stochastic Oscillator (%D)

"Stochastic Oscillator (SO) *%D* is the 3-day moving average of *%K* (the last 3 values of *%K*). Usually this is a simple moving average, but can be an exponential moving average for a less standardized weighting for more recent values" [16]. There is only one valid signal in working with *%D* alone — a divergence between *%D* and the analyzed security. And it is calculated as follow:

$$\%D = MA(\%K, s) \tag{6}$$

where

- *%D* is the 3-day moving average of *%K*.
- *MA* is the moving average like *SMA*.
- *s* is the amount of periods of calculation of the moving average.

3.4.3.4 The Exponential Moving Average (EMA)

"An exponential moving average (EMA) is one of the most used indicators in technical analysis today, and it is a type of moving average indicators that is similar to a simple moving average (SMA), except that more weight is given to the latest data" [16]. This type of moving average reacts faster to recent price changes.

$$EMA_t = P_t * k + EMA_y * (1 - k) \tag{7}$$

where P = price, t = today, y = yesterday, k = 2/ (N+1), N = number of days in EMA. As explained previously the Exponential moving averages reduce the lag by applying more weight to recent prices. The weighting applied to the most recent price depends on the number of periods in the moving average. For example, a 10-period exponential moving average applies an 18.18% weighting to the most recent price. A 10-period EMA can also be called an 18.18% EMA. A 20-period EMA applies a 9.52% weighing to the most recent price (2/ (20+1) = .0952). Notice that the weighting for the shorter time period is more than the weighting for the longer time period. In fact, the weighting drops by half every time the moving average period doubles.

The longer the moving average, the more the lag. A 10-day exponential moving average will hug prices quite closely and turn shortly after prices turn. Short moving averages are like speed boats - nimble and quick to change. In contrast, a 100-day moving average contains lots of past data that slows it down. Longer moving averages are like ocean tankers - lethargic and slow to change. It takes a larger and longer price movement for a 100-day moving average to change course. In our implementation part we used EMA_1 and EMA_{10} [17].

3.4.3.5 Moving Average Convergence-Divergence (MACD)

“Moving average convergence divergence (MACD) is a trend-following momentum indicator that shows the relationship between two moving averages of prices” [16]. The MACD is calculated by subtracting the 26-day exponential moving average (EMA) from the 12-day EMA. A nine-day EMA of the MACD, called the "signal line", is then plotted on top of the MACD, functioning as a trigger for buy and sell signals.

$$MACD_t = EMA_{12,t} - EMA_{26,t} \tag{8}$$

Accordingly, When MACD indicator increases above its signal line, a buy signal is generated. When MACD indicator decreases below its signal line, a sell signal is generated [16].

3.4.3.6 The Commodity Channel Index (CCI)

The commodity channel index (CCI) is an oscillator originally. Currently, is a very common tool for traders in identifying cyclical trends not only in commodities, but also equities and currencies. The CCI can be adjusted to the timeframe of the market traded on by changing the averaging period. CCI measures a security's variation from the statistical mean.

The CCI is calculated as the difference between the typical price of a commodity and its simple moving average (SMA), divided by the mean absolute deviation of the typical price. The index is usually scaled by an inverse factor of 0.015 to provide more readable numbers:

$$CCI_t = \frac{1}{0.015} \frac{P_t - SMA(P_t)}{\sigma(P_t)} \tag{9}$$

where

- P_t is the Typical price = $\frac{H+L+C}{3}$, And (H, L, C) are (High, Low, Close) prices.
- SMA is the simple moving average.
- σ is the mean absolute deviation.

For scaling purposes, the constant is set at 0.015 to ensure that approximately 70 to 80 percent of CCI values would fall between −100 and +100. The CCI fluctuates above and below zero. The percentage of CCI values that fall between +100 and −100 will depend on the number of periods used. A shorter CCI will be more volatile with a smaller percentage of values between +100 and −100. Conversely, the more periods used to calculate the CCI, the higher the percentage of values between +100 and −100.

3.4.3.7 Parabolic Stop and Reverse

The Parabolic Stop and Reverse (SAR) indicator combines price and time components in an attempt to generate potential buy and sell signals. The Parabolic SAR advertises itself as an effective tool to determine where to place stop loss orders. The parabolic SAR is calculated almost independently for each trend in the price. When the price is in an uptrend, the SAR emerges below the price and converges upwards towards it. Similarly, on a downtrend, the SAR emerges above the price and converges downwards. At each step within a trend, the SAR is calculated one period in advance. That is, tomorrow's SAR value is built using data available today [16]. The general formula used for this is:

$$SAR_{t+1} = SAR_t + \alpha(EP - SAR_t) \tag{10}$$

where

- SAR_t and SAR_{t+1} represent the current period and the next period's SAR values, respectively.
- *EP* (the extreme point) is a record kept during each trend that represents the highest value reached by the price during the current uptrend – or lowest

value during a downtrend. During each period, if a new maximum (or minimum) is observed, the EP is updated with that value.

- The α value represents the acceleration factor. Usually, this is set initially to a value of 0.02, but can be chosen by the trader. This factor is increased by 0.02 each time a new EP is recorded, which means that every time a new EP is observed, it will make the acceleration factor go up. The rate will then quicken to a point where the SAR converges towards the price. To prevent it from getting too large, a maximum value for the acceleration factor is normally set to 0.20. The traders can set these numbers depending on their trading style and the instruments being traded. Generally, it is preferable in stocks trading to set the acceleration factor to 0.01, so that is not too sensitive to local decreases. For commodity or currency trading, the preferred value is 0.02.

3.4.3.8 Rate of Change

"The Rate of Change indicates the margin between the current price and the previously existed one from n-time periods ago. ROC increases when the prices trend up whether it declines when they trend down" [16]. The scale of the prices changes calls the corresponding ROC change.

The ROC Indicator is calculated as a difference between the price of the current period and the price of the previous period, which is located in n periods back from the current one:

$$ROC_t = 100((CP_t - CP_{t-n})/(CP_{t-n}) \tag{11}$$

where

- CP_t is the price of the current period.
- CP_{t-n} is the price of the period, which is located in *n* periods back from the current one [17].

3.5 Support Vector Machines

Support vector machines (SVMs) use a linear model to implement nonlinear class boundaries through some nonlinear mapping input vectors into a high-dimensional feature space. The linear model constructed in the new space can represent a nonlinear decision boundary in the original space. In the new space, an optimal separating hyperplane (OSH) is constructed. Thus, SVM is known as the algorithm that finds a special kind of linear model, *the maximum margin hyperplane*. The maximum margin hyperplane gives the maximum separation between decision classes. The training examples that are closest to the maximum margin hyperplane are called support vectors. All other training examples are irrelevant for defining the binary class boundaries.

SVM is simple enough to be analyzed mathematically since it can be shown to correspond to a linear method in a high dimensional feature space nonlinearly related to input space. In this sense, SVM may serve as a promising alternative combining the strengths of conventional statistical methods that are more theory-driven and easy to analyze, and more data-driven, distribution-free and robust machine learning methods [19].

A simple description of the SVM algorithm is provided as follows. Given training set $D = \{x_i,\ y_i\}, i = 1, \ldots, N$ where N represents the overall numbers of training examples, with input of $x_i \in R^n$ and an output of $y_i \in \{-1, +1\}$, the support vector machine (SVM) classifier, according to Vanpik's original formulation, satisfies the following conditions.

$$\begin{cases} \mathrm{w}^T \emptyset(x_i) + b \geq +1, & \text{if } y_i = +1 \\ \mathrm{w}^T \emptyset(x_i) + b \geq -1, & \text{if } y_i = -1 \end{cases} \tag{12}$$

which is equivalent to

$$y_i[\mathrm{w}^T \emptyset(x_i) + b] \geq 1 \tag{13}$$

where w represents the weight vector and b the bias. Nonlinear function $\emptyset(\cdot) : \mathrm{R}^n \rightarrow \mathrm{R}^{nk}$ maps input or measurement space to a high-dimensional, and possibly infinite-dimensional, feature space. Eq. (13) then comes down to the construction of two parallel bounding hyperplanes at opposite sides of a separating hyperplane $\mathrm{w}^T \emptyset(x) + b = 0$ in the feature space with the margin width between both hyperplanes equal to $2/(||\mathrm{w}||^2)$. In primal weight space, the classifier then takes the decision function form (3)

$$\mathrm{sgn}(\mathrm{w}^T \emptyset(x) + b) \tag{14}$$

Most of classification problems are, however, linearly non-separable. Therefore, it is general to find the weight vector using slack variable (ξ_i) to permit misclassification. One defines the primal optimization problem as

$$\underset{w,b,\xi}{\text{Min}} \frac{1}{2} \mathrm{w}^T \mathrm{w} + \mathrm{C} \sum_{i=1}^{N} \xi_i \tag{15}$$

Subject to

$$\begin{cases} y_i(\mathrm{w}^T \, \emptyset(x_i) + b) \geq 1 - \xi_i, & i = 1, \ldots, N \\ \xi_i \geq 0, & i = 1, \ldots, N \end{cases} \tag{16}$$

Where ξ_i's slack variables needed to allow misclassifications in the set of inequalities, and $C \in \mathrm{R}^+$ is a tuning hyperparameter, weighting the importance of classification errors vis-à-vis the margin width. The solution of the primal problem is obtained after constructing the Lagrangian. From the conditions of optimality, one obtains a Quadratic Programming (QP) problem with Lagrange multipliers α_i's. A multiplier α_i exists for each training data instance. Data instances corresponding to non-zero α_i's are called support vectors.

On the other hand, the above primal problem can be converted into the following dual problem with objective function (17) and constraints (18). Since the decision variables are support vector of Lagrange multipliers, it is easier to interpret the results of this dual problem than those of the primal one [18].

$$\underset{\alpha}{\text{Max}} \frac{1}{2} \alpha^T \, Q\alpha - e^T \alpha \tag{17}$$

Subject to

$$\begin{cases} 0 \leq \alpha_i \leq C, & i = 1, \ldots, N \\ y^T \alpha = 0 \end{cases} \tag{18}$$

In the dual problem above, e is the vector of all ones, Q is a $N \times N$ positive semi-definite matrix, $Q_{ij} = y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) = \emptyset(x_i)^T \emptyset(x_j)$ is the kernel. Here, training vectors x_i's are mapped into a higher (maybe infinite) dimensional space by function $\emptyset$. As is typical for SVMs, we never calculate w or $\emptyset(x)$. This is made possible due to Mercer's condition, which relates mapping function $\emptyset(x)$ to kernel function $K(\cdot,\cdot)$ as follows.

$$K(x_i, x_j) = \emptyset(x_i)^T \emptyset(x_j) \tag{19}$$

For kernel function $K(\cdot,\cdot)$, one typically has several design choices such as the linear kernel of $K(x_i, x_j) = x_i^T x_j$, the polynomial kernel of degree d of $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$, the radial basis function (RBF) kernel of $K(x_i, x_j) = \exp\{-\gamma ||x_i - x_j||^2\}, \gamma > 0$, and the sigmoid kernel of $K(x_i, x_j) = \tanh\{\gamma x_i^T x_j + r\}$, where d, $r \in \mathrm{N}$ and $\gamma \in \mathrm{R}^+$are constants. Then one constructs the final SVM classifier as:

$$\mathrm{sgn}\left(\sum_{i}^{N} \alpha_i y_i K(x, x_i) + b\right) \tag{20}$$

Here the K represents a kernel function, while α_i are Lagrange multipliers. When using a Radial Basic Function (RBF) defined by:

$$K(x_i, x_j) = e^{\frac{-||x_i - x_j)||^2}{\sigma^2}} \tag{21}$$

3.5.1 Least Square Support Vector machine

Least squares are versions of support vector machines (SVM), which are a set of related supervised learning methods that analyze data and recognize patterns, and which are used for classification and regression analysis.

LS-SVM simplifies traditional SVM by introducing equality constraints instead of inequality constraints [12]. The calculations for LS-SVM are totally same as SVM classifier only in the optimization problem in the primal space:

$$\underset{w,b,\xi}{\mathrm{Min}} \frac{1}{2} \mathrm{w}^T \mathrm{w} + \frac{1}{2} \mathrm{C} \sum_{i=1}^{N} \xi_i^2 \tag{22}$$

with the following constraints:

$$y_i(\mathrm{w}^T\ \emptyset(x_i) + b) = 1 - \xi_i \tag{23}$$

3.6 Weighted kernel Function

Is a function that perform mapping from input space into higher dimension feature space. After that, a linear machine is used to classify the data in the feature space. Several kernel functions are proposed to help the SVMs in obtaining the optimal solution, but the most frequently used kernel functions are the Polynomial, Sigmoid, Gaussian and Radial Basis Function (RBF). The RBF and Gaussian kernels are frequently used by most studies in our study we are dealing with RBF [24].

Most of these, however, do not directly optimize the original regression or classification problem, but instead seek to find a good set of weights on the features for later use within the kernel of the model. In the following section, we present the basics of weighted kernels in relation to SVM and LS-SVM theory.

The weighted kernel function is defined as $K(\theta_{xi}, \theta_{xj})$ where θ is a weight vector of data set features. The classification model in dual form with feature weights is formulated in (24), with the note that feature weights were also included during the computation of α_i and b.

$$\mathrm{y(x)} = \mathrm{sgn}\left(\sum_{i=1}^{N} \alpha_i y_i K\left(\theta_{xi}, \theta_{xj}\right) + b\right) \tag{24}$$

From (24), it can be seen that the defined weighted kernel is not dependent on the type of kernel function itself. Hence, it is used to determine the weight vector $\theta = (\theta_1, \theta_2, \ldots, \theta_d)^T$ based on the AHP method, which is introduced in detail in Sect. 3.1, 3.2 and 3.3. Moreover, the elements of the feature weight vector obey the following two conditions:

$$0 \le \theta_k \le 1 \quad k = 1, \ldots, d \tag{25}$$

And

$$\sum_{k=1}^{d} \theta_k = 1$$

Accordingly the weighted RBF kernel in (21) can be rewritten as:

$$K(x_i, x_j) = e^{\frac{-||\Theta(x_i - x_j)||^2}{\sigma^2}} \tag{26}$$

where Θ= diag[θ_1, θ_2, . . . , θ_n].

Chapter 4

IMPLEMENTATION AND PERFORMANCE EVALUATION

4.1 Introduction

This section presents the experimental results and discussion of applying the new approach (AHP-WK-LS-SVM). The goal of this study is to compare the performance of the feature ranking and selection approach in combination with weighted kernel LS-SVMs with different SVM benchmark models. The section begins with a description of the datasets (Stock Market Indices) used in the experiments, following the experimental setup. Then the results are presented and been discussed. Moreover, the tool selected to design the prediction simulator was MATLAB because of its power and simplicity at the same time.

4.2 Data Description

All data used in this work was pulled from Yahoo! finance's database of historical stock data [20]. And then was imported into MATLAB as a structure, The experiments were conducted on the data for the Borsa Istanbul (BIST100), BELEX15, FTSE100 and S&P500 stock market indices. The value of indices determines the price of the most liquid stocks traded on the regulated market of the observed markets. The series consists of six time-series values which are determined for each day: the closing price, the change in the value of the index in relation to the previous trading day in percentages, the opening price, highest price, lowest price and the trading volume. The data were divided into two groups. The first group consisted of records required for the model training, from 6 October 2007 to 31 December 2014. The BIST100 index training data set consisted of 1757 samples. The FTSE100 training set consisted of 1793 data samples. The BELEX15 index training data set consisted of 1764 samples. And the S&P500 training data set consisted of 1764 samples. The Table 4.1 below is a sample of Borsa Istanbul stock index data set containing the six time-series.

Table 4.1: Sample of Borsa Istanbul Stock Index Data Set

A	B	C	D	E	F	G
			Borsa Istanbul			
Date	Open	High	Low	Close	Volume	Adj Close
4/16/2014	1846.01001	1862.310059	1846.01001	1862.310059	3155080000	1862.310059
4/15/2014	1831.449951	1844.02002	1816.290039	1842.97998	3736440000	1842.97998
4/14/2014	1818.180054	1834.189941	1815.800049	1830.609985	3111540000	1830.609985
4/11/2014	1830.650024	1835.069946	1814.359985	1815.689941	3743460000	1815.689941
4/10/2014	1872.280029	1872.530029	1830.869995	1833.079956	3758780000	1833.079956
4/9/2014	1852.640015	1872.430054	1852.380005	1872.180054	3308650000	1872.180054
4/8/2014	1845.47998	1854.949951	1837.48999	1851.959961	3721450000	1851.959961
4/7/2014	1863.920044	1864.040039	1841.47998	1845.040039	3801540000	1845.040039
4/4/2014	1890.25	1897.280029	1863.26001	1865.089966	3583750000	1865.089966
4/3/2014	1891.430054	1893.800049	1882.650024	1888.77002	3055600000	1888.77002
4/2/2014	1886.609985	1893.170044	1883.790039	1890.900024	3131660000	1890.900024
4/1/2014	1873.959961	1885.839966	1873.959961	1885.52002	3336190000	1885.52002
3/31/2014	1859.160034	1875.180054	1859.160034	1872.339966	3274300000	1872.339966
3/28/2014	1850.069946	1866.630005	1850.069946	1857.619995	2955520000	1857.619995
3/27/2014	1852.109985	1855.550049	1842.109985	1849.040039	3733430000	1849.040039

For the model testing, data from 3 January 2015 to 31 December 2015 were used, a total of 252 days of trading for all the data series. The results are obtained for one-day-ahead predictions using data over an extended period of time, 1 trading year. Next, brief descriptions of the used stock indices are given below.

4.2.1 Stock Market Index

"A stock market index is a measurement of the value of a section of the stock market. It is computed from the prices of selected stocks (typically a weighted average). It is a tool used by investors and financial managers to describe the market, and to compare the return on specific investments" [21].

An index is a mathematical construct, so it may not be invested in directly. But many mutual funds and exchange-traded funds attempt to "track" an index, and those funds that do not may be judged against those that do.

4.2.1.1 Borsa Istanbul 100 Index

The Borsa Istanbul (abbreviated as BIST100) is the sole exchange entity of Turkey combining the former Istanbul Stock Exchange (ISE) (Turkish: İstanbul Menkul Kıymetler Borsası, IMKB), the Istanbul Gold Exchange (Turkish: İstanbul Altın Borsası, İAB) and the Derivatives Exchange of Turkey (Turkish: Vadeli İşlem Opsiyon Borsası, VOB) under one umbrella. It was established as an incorporated company with a founding capital of Turkish lira symbol [22].

4.2.1.2 Standard & Poor's 500 Index

The Standard & Poor's 500 Index (S&P500) is an index of 500 stocks seen as a leading indicator of U.S. equities and a reflection of the performance of the large cap universe, made up of companies selected by economists. The S&P500 is a market value weighted index and one of the common benchmarks for the United States stock market [21].

4.2.1.3 Financial Times Stock Exchange 100 Index

The Financial Times Stock Exchange 100 (FTSE100) Index is a share index of the 100 companies listed on the London Stock Exchange with the highest market capitalization. It is seen as a gauge of prosperity for businesses regulated by United Kingdom company law. The index is maintained by the FTSE Group, a subsidiary of the London Stock Exchange Group.

4.2.1.4 Belgrade Stock Exchange

The Belgrade Stock Exchange (abbreviated as BELEX15) is a stock exchange based in Belgrade, Serbia. Currently, the Belgrade Stock Exchange is a full member of

Federation of Euro-Asian Stock Exchanges (FEAS) and an associate member of Federation of European Securities Exchanges (FESE) [21].

4.3 Experimental Results and Discussion

The stock market trend prediction problem is commonly modeled as a two-class classification problem where the classes are labeled with -1 and 1. Class -1 indicates that the closing price of the current day is higher than the closing price of the following day. The second class indicates the opposite. Figure 4.1 shows the trend fluctuations for the stock index BIST100 for a specific period of time (100 days).

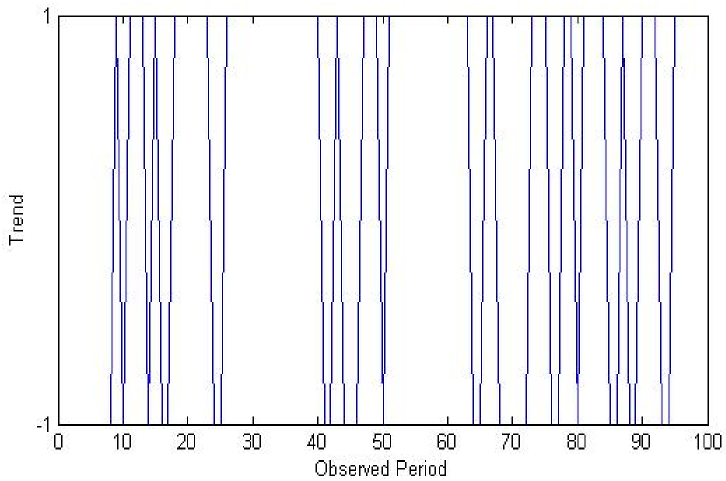


Figure 4.1: Trend Fluctuation

From Fig. 4.1, it can be noticed that the trend fluctuates up and down repeatedly, rendering it challenging for prediction.

As it was explained in section 3.4.3 the basic indicators that we are dealing with in this study, here the detailed procedure for calculating these indicators and the rules for generating trading signals are given in Table 4.2.

Table 4.2: Technical Indicators and Trading Strategies

Technical indicator	Formula	Trading strategy signals, ST
EMA	$EMA_t = P_t * k + EMA_y * (1 - k)$	$\begin{cases} 1 & \text{if } \mathrm{EMA}_{1,t} > \mathrm{EMA}_{10,t} \\ -1 & \text{if } \mathrm{EMA}_{1,t} < \mathrm{EMA}_{10,t} \end{cases}$
MACD	$MACD_t = EMA_{12,t} - EMA_{26,t}$	$\begin{cases} 1 & \text{if } \mathrm{MACD}_t > \mathrm{EMA}_{9,t} \\ -1 & \text{if } \mathrm{MACD}_t < \mathrm{EMA}_{9,t} \end{cases}$
RSI	$RSI = 100 - [\frac{100}{(1 + RS)}]$	$\begin{cases} 1 & \text{if } \mathrm{RSI}_{t-1} \geq 30 \text{ and } \mathrm{RSI}_t \geq 30 \\ -1 & \text{if } \mathrm{RSI}_{t-1} \leq 70 \text{ and } \mathrm{RSI}_t \leq 70 \end{cases}$
CCI	$CCI_t = \frac{1}{0.015} \frac{P_t - SMA(P_t)}{\sigma(P_t)}$	$\begin{cases} 1 & \text{if } \mathrm{CCI}_t > 100 \text{ or } \mathrm{CCI}_t > -100 \\ -1 & \text{if } \mathrm{CCI}_t < 100 \text{ or } \mathrm{CCI}_t < -100 \end{cases}$
SO	$\%K = 100((CP_t - LP)/(HP - LP)$ $\%D = MA(\%K, s)$	$\begin{cases} 1 & \text{if } \%\mathrm{D} < 0.2 \text{ and } \%\mathrm{K}_t > \%D \\ -1 & \text{if } \%\mathrm{D} > 0.8 \text{ and } \%\mathrm{K}_t < \%D \end{cases}$
SAR	$SAR_{t+1} = SAR_t + \alpha(EP - SAR_t)$	$\begin{cases} 1 & \text{if } \mathrm{CP}_t > \mathrm{SAR}_t \\ -1 & \text{if } \mathrm{CP}_t < \mathrm{SAR}_t \end{cases}$
ROC	$ROC_t = 100((CP_t - CP_{t-n})/(CP_{t-n})$	$\begin{cases} 1 & \text{if } \mathrm{ROC}_t > 0 \\ -1 & \text{if } \mathrm{ROC}_t < 0 \end{cases}$

Accordingly, descriptive statistics for the selected indicators based on the available data sets were calculated, and are shown in Table 4.3. Including the minimum min value, the maximum max value, mean value and the standard deviation for each stock index applied to all technical indicators. “In statistics, the standard deviation is a measure of the dispersion of a set of data from its mean. If the data points are further from the mean, there is higher deviation within the data set” [21]. Standard deviation is calculated as the square root of variance by determining the variation between each data point relative to the mean.

In Table 4.2, the difference in values depends on the calculation process of each indicator have been reported and the corresponding plots are shown in appendix section A.

Table 4.3: Descriptive Statistics for the Selected input Features

Stock indices	BIST100				FTSEl00				BELEX15				S&P500			
Technical indicator	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD	Min	Max	Mean	SD
ROC	-17.77	19.11	0.35	4.54	-14.16	8.48	0.38	2.56	-17.06	11.95	0.57	3.16	-16.30	9.25	0.43	2.71
CCI	-397.64	338.65	7.04	117.99	-427.73	287.22	19.63	118.70	-382.05	261.94	23.81	115.37	-401.40	281.29	21.78	116.81
RSI	13.45	91.26	52.60	15.12	11.04	82.88	54.59	13.92	13.73	86.85	55.48	13.99	11.39	83.98	54.99	13.45
%*K*	0.00	100.00	56.50	33.34	0.00	100.00	62.64	28.15	0.00	100.00	63.91	27.98	0.00	100.00	63.50	27.73
%*D*	0.83	99.96	56.35	31.50	3.24	99.16	62.48	26.61	2.30	99.24	63.74	26.46	2.81	98.63	63.34	26.12
EMA_1	45230.90	93178.90	69640.88	10817.03	9686.48	18636.05	14402.52	2739.17	2045.11	5339.52	3584.83	1018.34	1022.58	2190.15	1608.14	370.10
EMA_{10}	45942.84	91089.41	69569.07	10738.08	9730.22	18552.24	14380.81	2736.19	2084.58	5292.84	3576.69	1015.79	1050.69	2182.63	1605.32	369.47
MACD	-3586.37	3691.11	126.45	1111.40	-421.13	299.12	34.93	115.70	-144.29	91.06	12.47	35.81	-47.64	32.13	4.47	13.15
SAR	46120.18	91889.60	69697.87	10712.51	9604.04	18627.93	14366.91	2746.52	2080.11	5274.34	3575.68	1016.70	1029.39	2187.58	1604.58	370.62

Figure 4.2 displays the rates of SAR indicator as an example over a period of seven years approximately 1800 days as represented in x-axis for Borsa Istanbul index that are reported in Table 4.2. And the y-axis denote to the indicator rates according to the maximum max value reported for the mentioned indicator as well as minimum min value during the period.

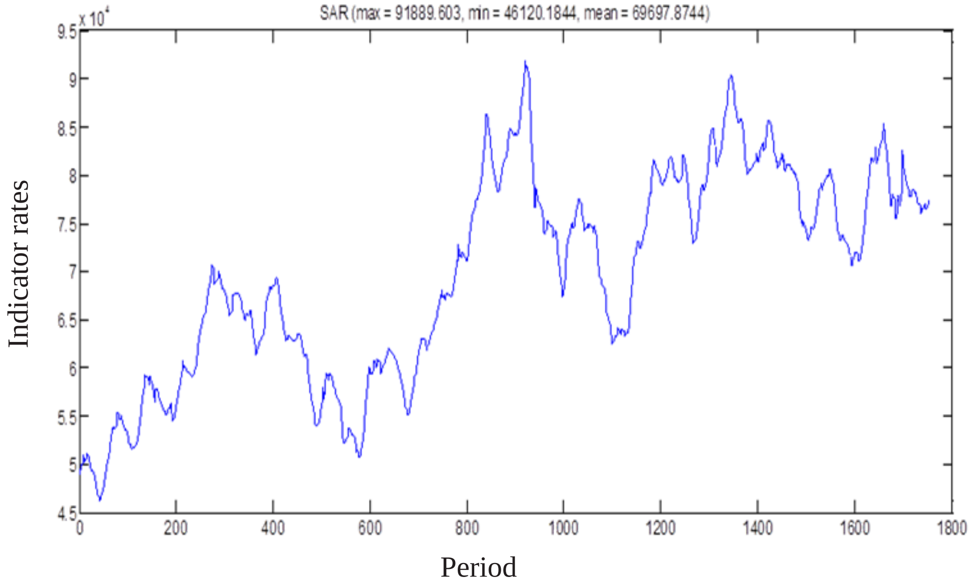


Figure 4.2: SAR Indicator for BIST100 Index

The first step for implementing AHP is to provide an initial matrix for the criterion pairwise comparisons (Table 4.4). The risk and return criteria are evaluated based on standard economic theory assumptions that investors are commonly averse to risk. Consequently, the third criterion hit ratio is evaluated as the most significant one. For our calculations, we used a 4-year trading cycle sub-sample period starting from the beginning of 2011 and lasting until the end of 2014.

Table 4.4: Pairwise Criteria Comparison Matrix

	Return	Risk	HR	RVV
Return	1	1/4	1/6	0.082
Risk	4	1	1/4	0.236
HR	6	4	1	0.682

$\lambda_{max} = 3.1078$, consistency ratio (CR) $= 0.09297$

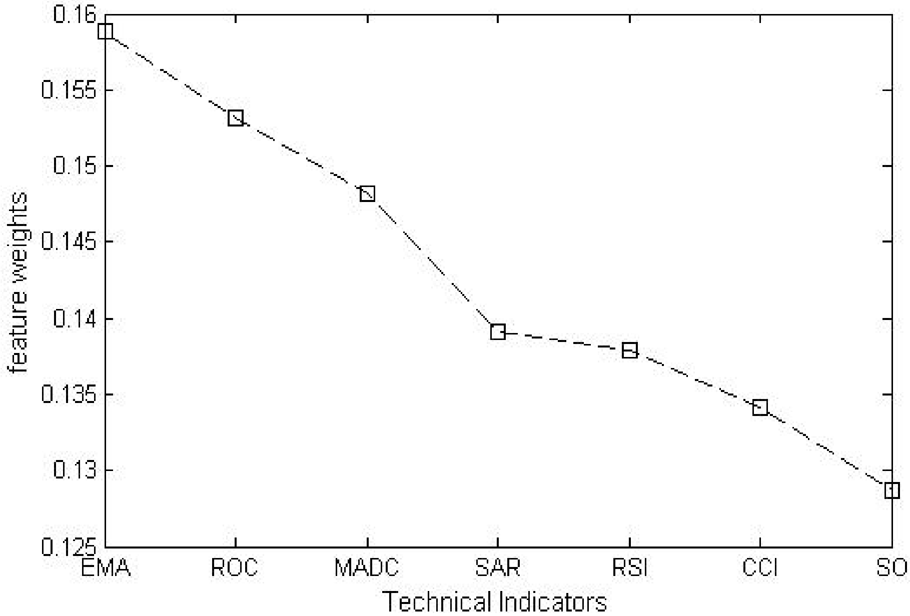
The eigenvector which is represented as a Relative Value Vector is calculated by the methods described in Sect. 3.2. As RVV = (0.082, 0.236, 0.682). These three numbers correspond, respectively, to the relative values of each criterion of return, risk and accuracy. The result 0.682 means that the model values accuracy most of all; 0.236 shows that risk is valued less; and 0.082 shows that the model values return the least. The CR value is 0.09297, which is less than the value of the critical limit 0.1, and thus the model is consistent in its choices.

In the next step using three pairwise comparisons matrices, $OPM^T \times RVV^T$ = feature weights $(\theta)^T$. we compare the selected input features in terms of the gross return, systematic risk and prediction accuracy. Table 4.5 presents the summarized option performance matrix for the observed technical indicators. And the bold typed values indicate the selected features used for the associated data sets. Furthermore, the elements of the feature weight vector in Table 4.5 obey the conditions in (25).

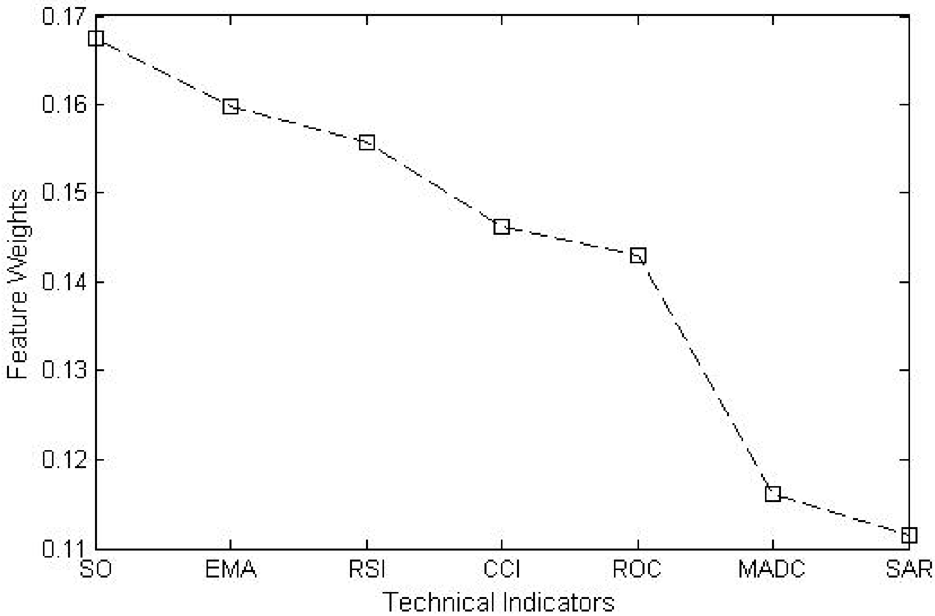
Table 4.5: Option Performance Matrix and Feature Weights

	ROC	CCI	RSI	SO	EMA	MACD	SAR
BIST100							
Return	0.158	0.120	0.077	0.021	0.207	0.253	0.164
Risk	0.194	0.125	0.135	0.103	0.167	0.138	0.138
HR	0.138	0.139	0.146	0.151	0.150	0.139	0.136
θ	**0.153**	0.134	0.138	0.129	**0.159**	**0.148**	0.139
FTSE100							
Return	0.147	0.095	0.243	0.204	0.223	0.074	0.014
Risk	0.136	0.174	0.144	0.207	0.205	0.067	0.069
HR	0.145	0.143	0.149	0.149	0.137	0.139	0.138
θ	0.143	0.146	**0.156**	**0.167**	**0.160**	0.116	0.112
BELEX15							
Return	0.134	0.086	0.164	0.257	0.130	0.031	0.197
Risk	0.109	0.148	0.116	0.220	0.146	0.074	0.187
HR	0.139	0.141	0.152	0.151	0.144	0.136	0.137
θ	0.131	0.138	0.145	**0.176**	0.144	0.113	**0.154**
S&P500							
Return	0.148	0.179	0.159	0.177	0.168	0.044	0.126
Risk	0.128	0.254	0.119	0.175	0.157	0.048	0.119
HR	0.143	0.141	0.152	0.147	0.139	0.139	0.139
θ	0.140	**0.170**	0.145	**0.156**	0.146	0.109	0.133

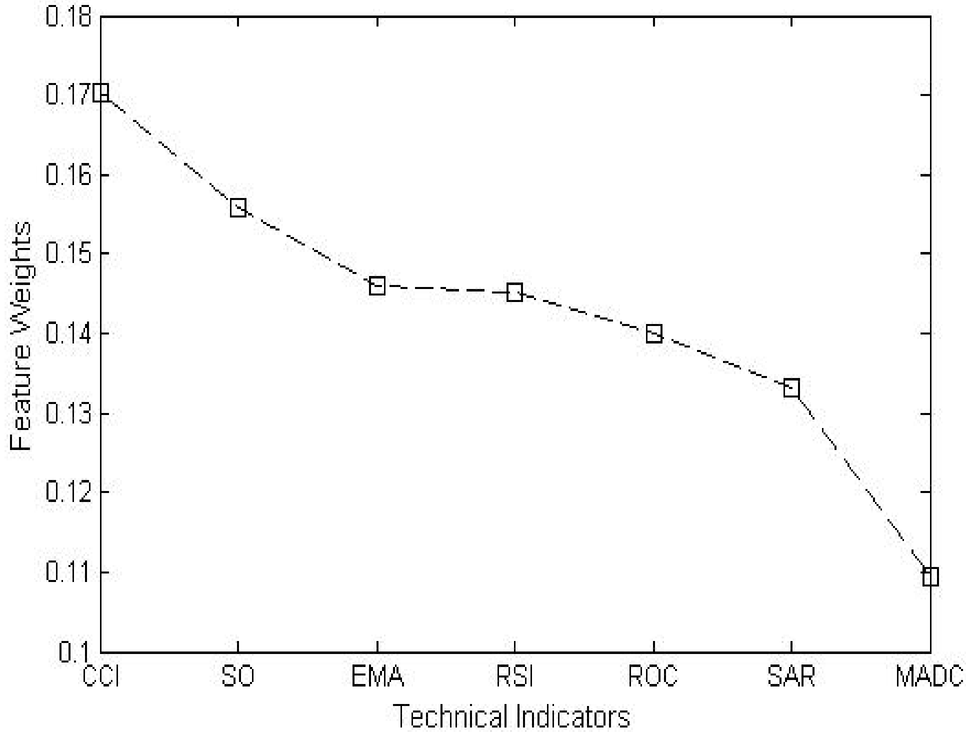
Based on the final calculations reported in Table 4.5, we obtained a decreasing order of feature weights according to the slope of the curve and Fig. 4.3 below shows a final summary of feature relevance.



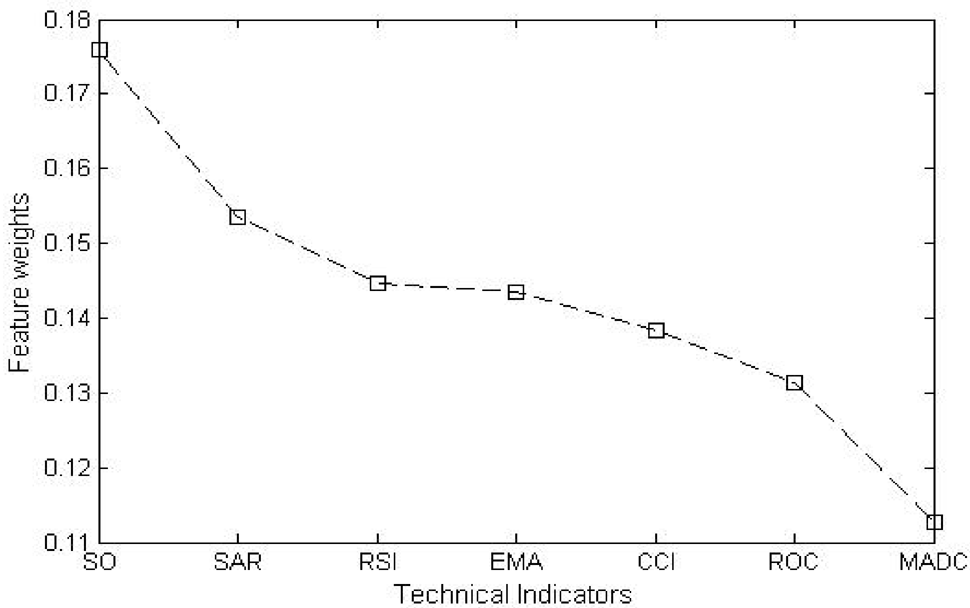
(a) BIST100



(b) FTSE100



(c) S&P500



(d) BELEX15

Figure 4.3: Description of the order of the Obtained Feature Weights

After obtaining the feature weights, we performed feature selection by analyzing the results shown in Fig. 4.3, as described in Sect. 3.3. It can be noticed from Fig. 4.3 that the indicator weights gradually decrease in the first three ranked indicator for the BIST100, S&P500 and FTSE100 index, and that for BELEX15 the decrease is significant after the first indicator. As a result, we selected the first three ranked indicators as input features for the prediction model for the BIST100 and FTSE100, and the first two ranked indicators for the S&P500 and BELEX15. Consequently, the first three rescaled weights to be incorporated into the LS-SVM kernel. For the S&P500 and BELEX15, we selected the first two ranked indicators. To form the SVM models.

4.4 Experimental Evaluation

Finally, we compared the accuracy of the new prediction model with other benchmark classifiers, for testing purposes we built 5 different SVM models and the results are shown in Table 4.6.

Because the behavior of financial asset prices is often governed by a random walk process due to the principles of the weak form of the efficient market hypothesis (EMH). Thus, the degree of accuracy of an approximate 60 % hit rate obtained in prediction using various machine learning techniques is often considered a satisfactory result for stock market trend prediction.

Furthermore, the input features for SVM are selected technical indicators. However, because they are completely in different scales we normalize them according to the constraints in eq. (25). Commonly, SVM is trained using true labels which are trading signals for closing price CP. Accordingly if the CP increase true label is *+1* unless it is *-1*. After the model is built and support vectors are found, next test data are applied to the model. Consequently SVM produces some scores for each test sample (one day). Then these scores are converted to predicted labels using a threshold (*0*) indicates that the negative scores are translated to predicted trading strategy *-1* and positive scores to *+1*. Finally for accuracy evaluation we compare true labels with predicted labels and report the obtained results according to the formula below:

$$Acc = \left(\frac{1}{N}\sum_{i=1}^{N} p_i\right) * 100 \tag{27}$$

where N is the number of test samples *i.e.* 252 days and p_i is defined as:

$$p_i = \begin{cases} 1 & \text{if true label} = \text{predicted label} \\ 0 & \text{otherwise} \end{cases}$$

Table 4.6: Accuracy Comparisons of Individual Prediction Models

Prediction model	BIST100	S&P500	FTSE100	BELEX15
SVM	55.56	56.75	57.54	56.35
AHP-SVM	58.47	58.52	58.88	56.35
AHP-WK-SVM	58.43	58.88	58.81	57.15
LS-SVM	57.14	59.52	57.54	**57.94**
AHP-LS-SVM	58.83	60.46	58.99	57.23
AHP-WK-LS-SVM*	**59.91**	**61.85**	**60.50**	57.53

From Table 4.6 it can be observed that, the AHP-WK-LS-SVM prediction model significantly outperforms all the benchmark models for the BIST100, S&P500 and FTSE100 data sets. In comparison with LS-SVM, the AHP-WK-LS-SVM is slightly lower for the BELEX15 index, around 1% less, but significantly higher for the BIST100, FTSE100 and S&P500, more than 2 and 7% respectively. Besides the AHP-WK-LS-SVM model, we tried to incorporate weights obtained from AHP into the SVM kernel. From Table 4.6, it can also be noted that the AHP-WK-SVM model significantly improves the SVM model, around 3% for the BIST100 and S&P500, and more than 1% for FTSE100 and BELEX15.

For all SVM classifiers train and test data are normalized to get zero mean and unit variance. More specifically, for each indicator mean and variance on train data is calculated and used for normalizing both train and test data.

Chapter 5

CONCLUSION

This thesis presents a novel and integrated approach to the problem of stock market forecasting. The applied methodology is relied on the concept of AHP analysis for feature ranking and selection. In addition, we used a weighted kernel to increase the generalization performance of the LS-SVM prediction model, where the kernel is weighted based on the feature relevance obtained by the conducted AHP analysis. The impact of the weighted kernel and feature selection has increased the accuracy of the prediction model. Furthermore, the set of feature weights obtained by the new approach can also independently be incorporated into other kernel-based learners, beside LS-SVMs.

The improvement in hit rates obtained on the test sets that contain data for 1 trading year can be considered a significant improvement, considering the fact that the stock market trend is predicted for the purpose of the optimization of investment strategies on the financial markets. Thus, percent increase in model precision can lead to a gain in terms of profit, since it results in greater return and a decrease in the risk involved in trading. Therefore, future improvements will focus on the study of criteria relevant to investors with different preferences regarding risk. Also, further work should include the formation of an ensemble model, where the outputs from several models would be combined into a final model by some aggregating scheme.

REFERENCES

[1] Bontempi, G., Taieb, S. B., & Le Borgne, Y. A. (2013). Machine learning strategies for time series forecasting. *In Business Intelligence* (pp. 62-77). Springer Berlin Heidelberg.

[2] Tay, F., E., H., Shen, L., & Cao, L. (2003). Ordinary Shares, Exotic Methods: Financial Forecasting Using Data Mining Techniques. *River Edge*, New Jersey (NJ). World Scientific

[3] Problem statement. (2016, Oct, 12). Retrieved from https://gks911.wordpress.com/2012/10/22/intelligentinvestor-the-problem-statement/#comment-54

[4] Stojanović, M. B., Božić, M. M., Stanković, M. M., & Stajić, Z. P. (2014). A methodology for training set instance selection using mutual information in time series prediction. *Neurocomputing*, 141, 236-245.

[5] He, Y., Fataliyev, K., & Wang, L. (2013, November). Feature selection for stock market analysis. In *International Conference on Neural Information Processing* (pp. 737-744). Springer Berlin Heidelberg.

[6] Lin, Y., Guo, H., & Hu, J. (2013, August). An SVM-based approach for stock market trend prediction. In *Neural Networks (IJCNN), The 2013 International Joint Conference on* (pp. 1-7). IEEE.

[7] Guresen, E., Kayakutlu, G., & Daim, T. U. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 38(8), 10389-10397.

[8] Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. *Computers & Operations Research*, 32(10), 2513-2522.

[9] Saaty, T. L. (2008). Decision making with the analytic hierarchy process. *International journal of services sciences*, 1(1), 83-98.

[10] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

[11] Kaufman, P. J. (2003). A short course in technical trading (Vol. 210). *John Wiley & Sons*.

[12] Inghelbrecht, K., Heyman, D., Pauwels, S., & Marius, P. (2012). Technical trading rules in emerging stock markets. *World Academy of Science, Engineering and technology*, 59, 2241-2264.

[13] Barak, S., & Modarres, M. (2015). Developing an approach to evaluate stocks by forecasting effective features with data mining methods. *Expert Systems with Applications*, 42(3), 1325-1339.

[14] Thawornwong, S., Dagli, C. H., & Enke, D. L. (2001). Using neural networks and technical analysis indicators for predicting stock trends.

[15] Saad, E. W., Prokhorov, D. V., & Wunsch, D. C. (1998). Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. *IEEE Transactions on neural networks*, 9(6), 1456-1470.

[16] Technical indicators. (2016, Nov, 2). Retrieved from https://www.forexrealm.com/technical-analysis/technical-indicators.html

[17] Technical indicators. (2016, Nov, 4). Retrieved from http://www.onlinetradingconcepts.com/TechnicalAnalysis.html

[18] Vapnik, V. N., & Vapnik, V. (1998). Statistical learning theory (Vol. 1). *New York: Wiley*.

[19] Min, J. H., & Lee, Y. C. (2005). Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications*, 28(4), 603-614.

[20] Yahoo!, "Yahoo! Finance." (2016, Oct, 11). Retrieved from http://finance.yahoo.com/

[21] Stock market indices. (2016, Oct, 6). Retrieved from http://www.investopedia.com

[22] Gunduz, H., & Cataltepe, Z. (2015). Borsa Istanbul (BIST) daily prediction using financial news and balanced feature selection. *Expert Systems with Applications*, 42(22), 9001-9011.

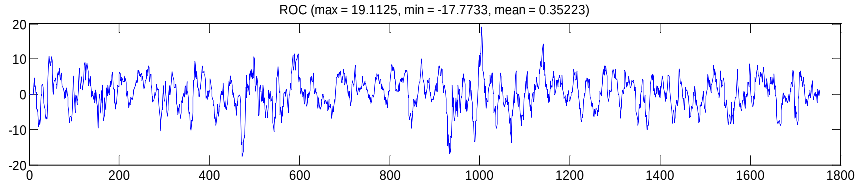
[23] Marković, I., & Stojanović, M., & Stanković, J., & Stanković, M. (2016). Stock market trend prediction using AHP and weighted kernel LS-SVM. *Soft Computing*, 1-12.

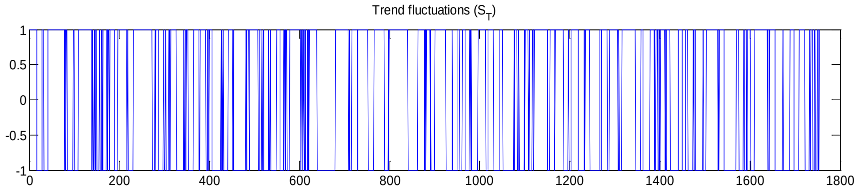
[24] Xing, H. J., Ha, M. H., Hu, B. G., & Tian, D. Z. (2009). Linear feature-weighted support vector machine. *Fuzzy Information and Engineering*, 1(3), 289-305.

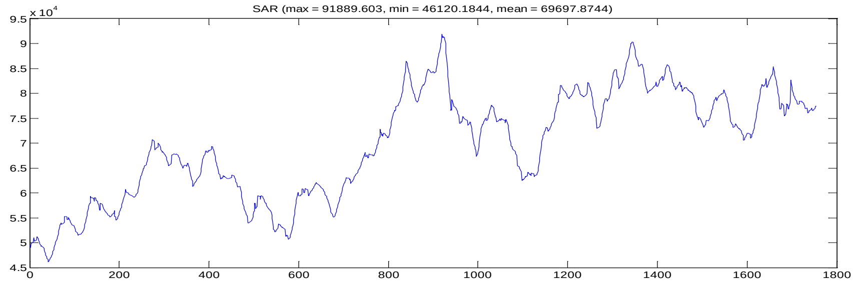
[25] Xiao, S. H., Fan, Z. P., & Wang, M. G. (2002). Uniform approach to two forms of preference information-AHP judgment matrices and fuzzy preference relation matrices in group decision making. *Journal of Systems Engineering*, 17(1), 82-86.

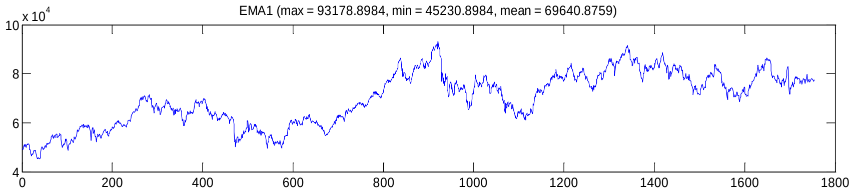
APPENDICES

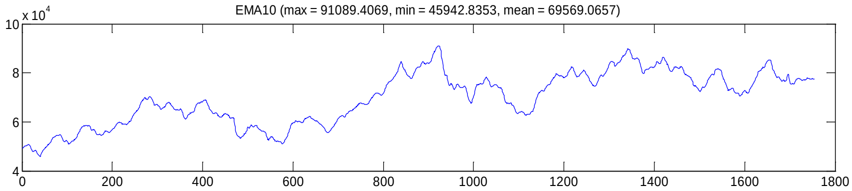
Appendix A: BIST100

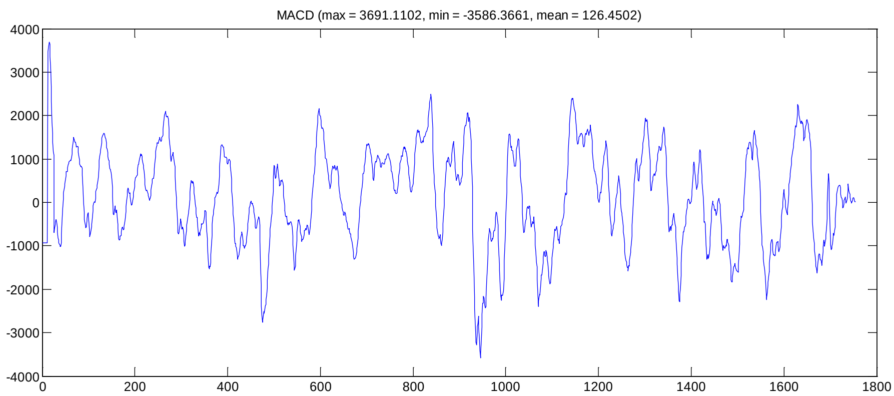


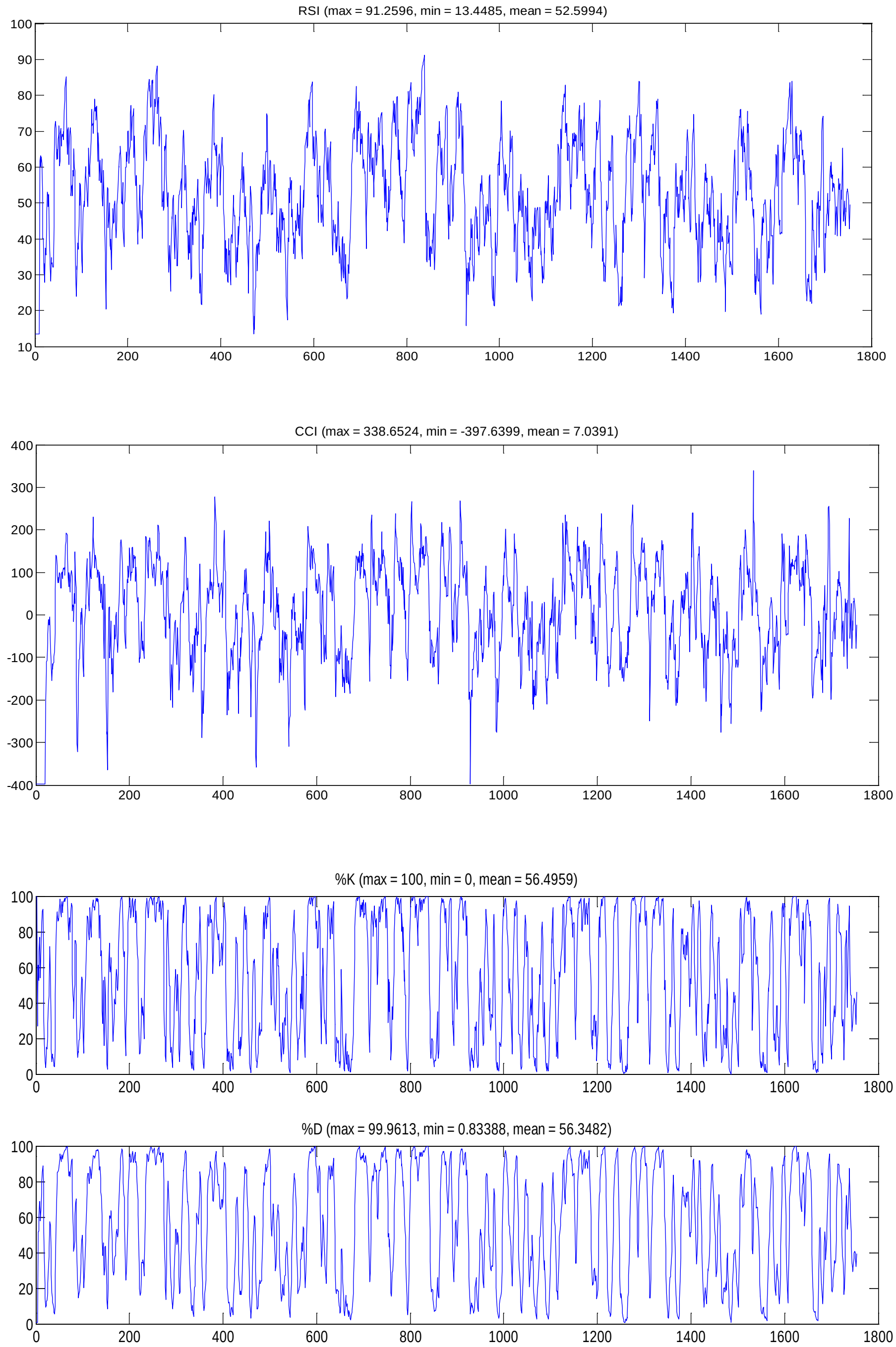




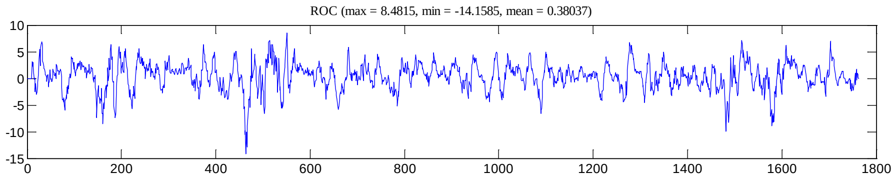


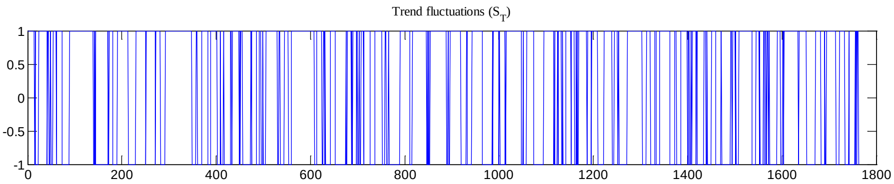


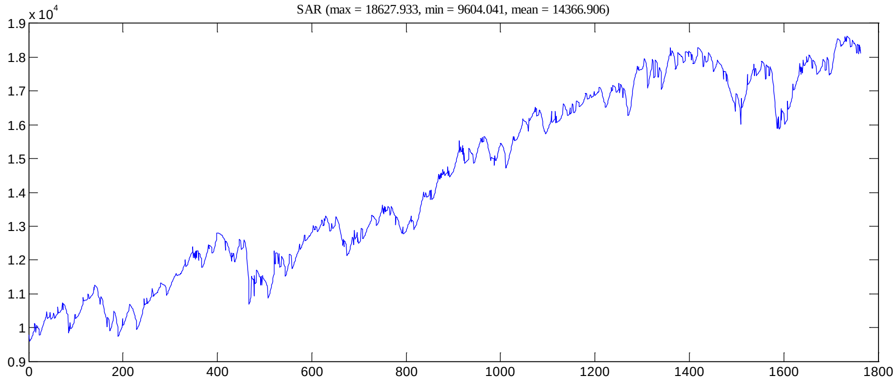


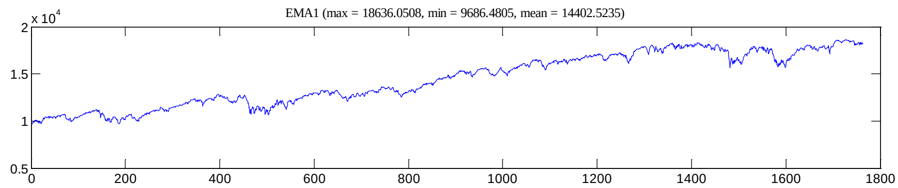

RSI (max = 91.2596, min = 13.4485, mean = 52.5994)
CCI (max = 338.6524, min = -397.6399, mean = 7.0391)
%K (max = 100, min = 0, mean = 56.4959)
%D (max = 99.9613, min = 0.83388, mean = 56.3482)

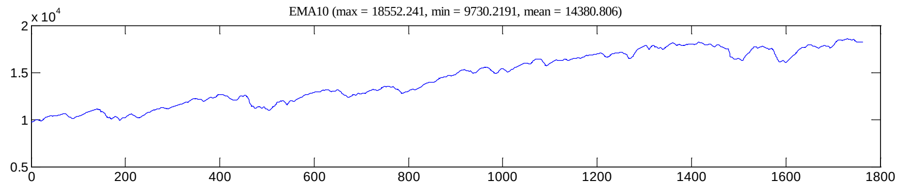
Appendix B: FTSE100

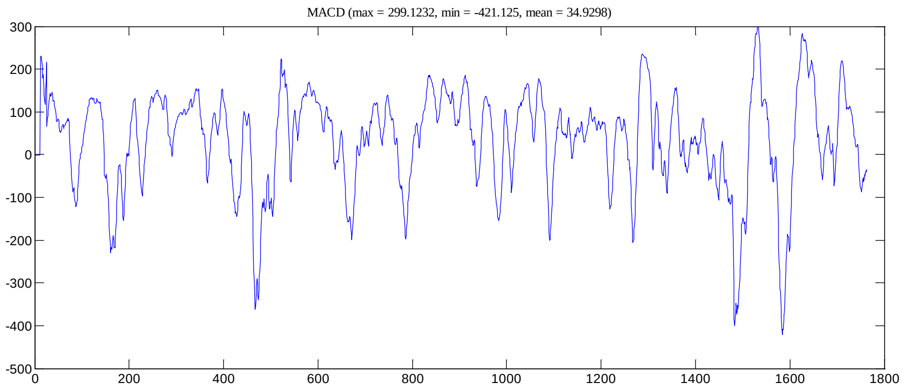


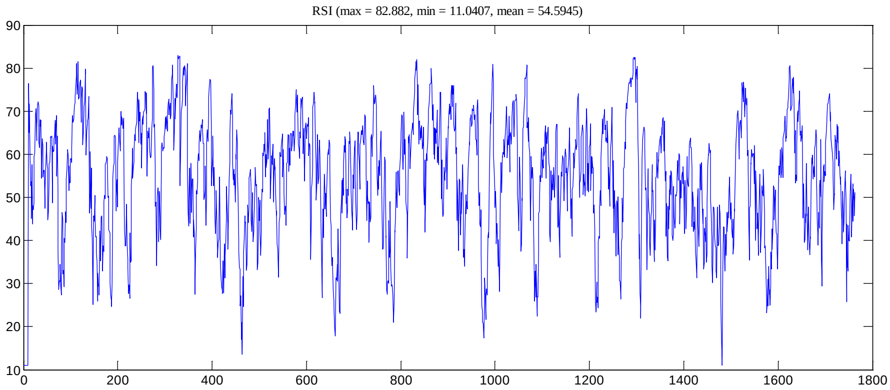


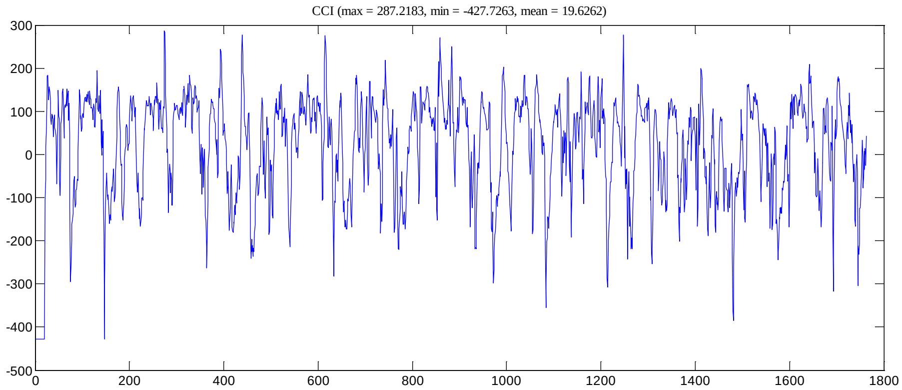


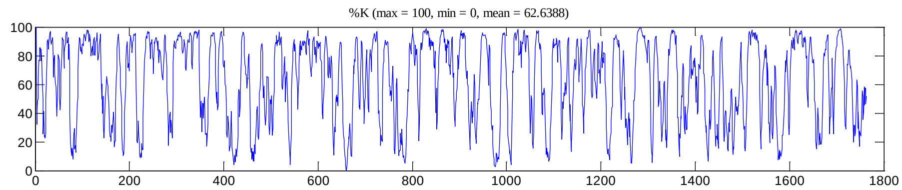


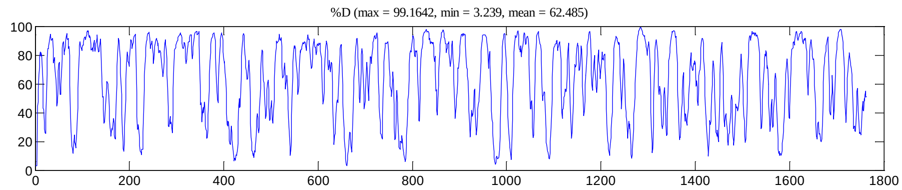




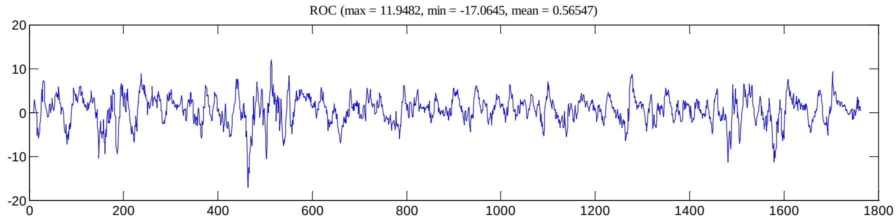

RSI (max = 82.882, min = 11.0407, mean = 54.5945)
90
80
70
60
50
40
30
20
10
0
200
400
600
800
1000
1200
1400
1600
1800

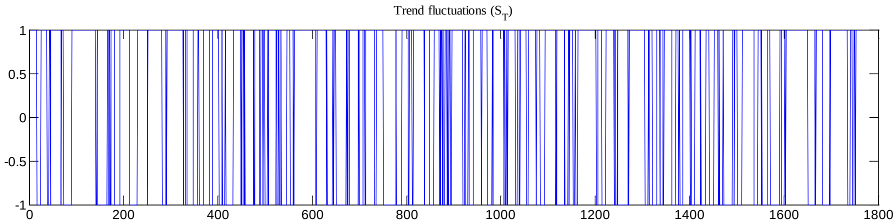

CCI (max = 287.2183, min = -427.7263, mean = 19.6262)
300
200
100
0
-100
-200
-300
-400
-500
0
200
400
600
800
1000
1200
1400
1600
1800

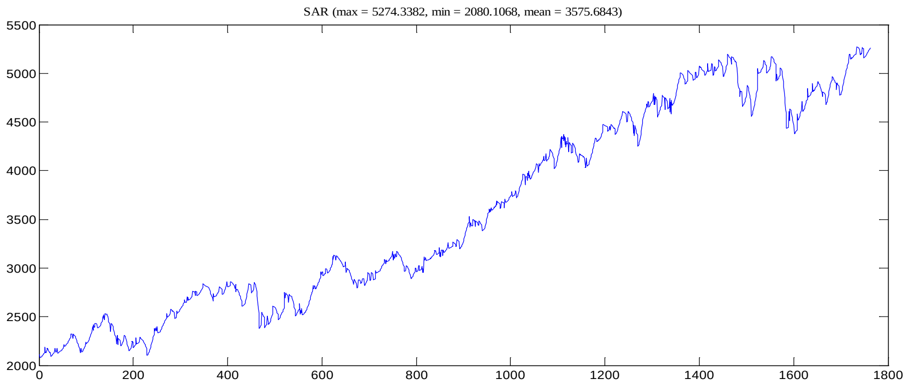

%K (max = 100, min = 0, mean = 62.6388)
100
80
60
40
20
0
0
200
400
600
800
1000
1200
1400
1600
1800

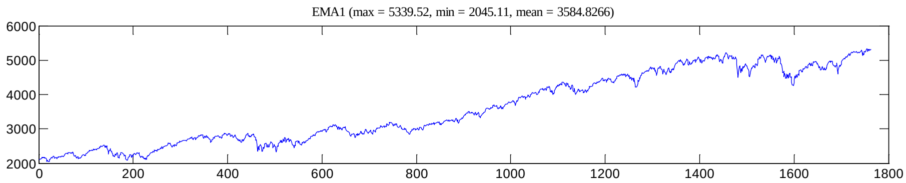

%D (max = 99.1642, min = 3.239, mean = 62.485)
100
80
60
40
20
0
0
200
400
600
800
1000
1200
1400
1600
1800

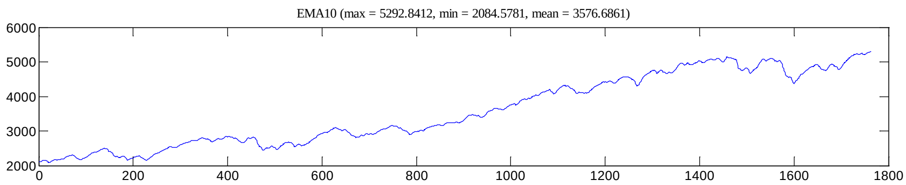
Appendix C: BELEX15

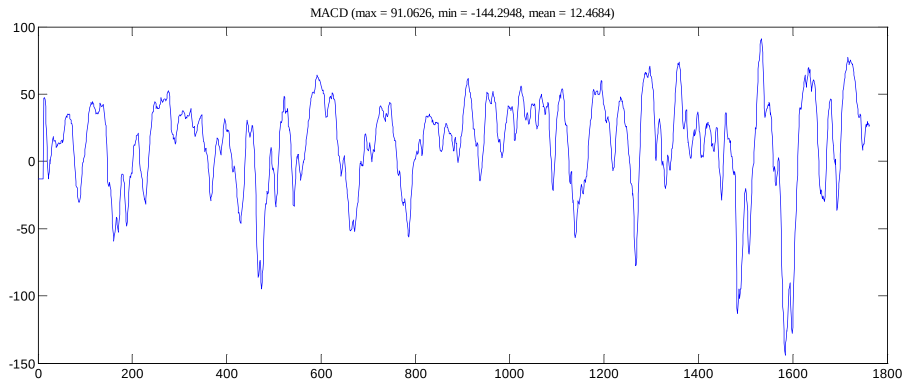


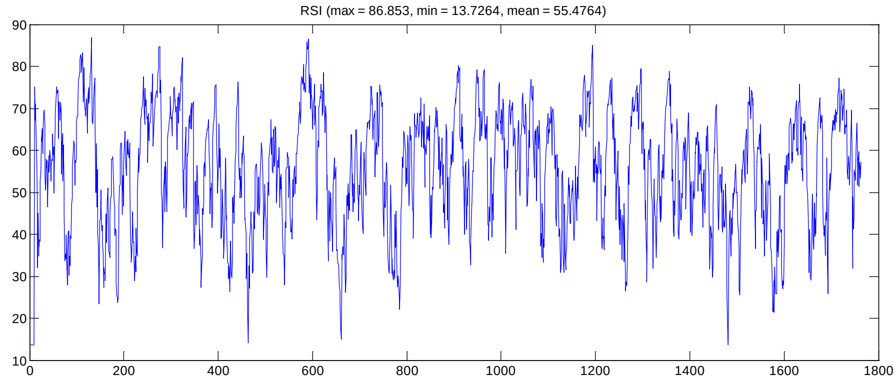


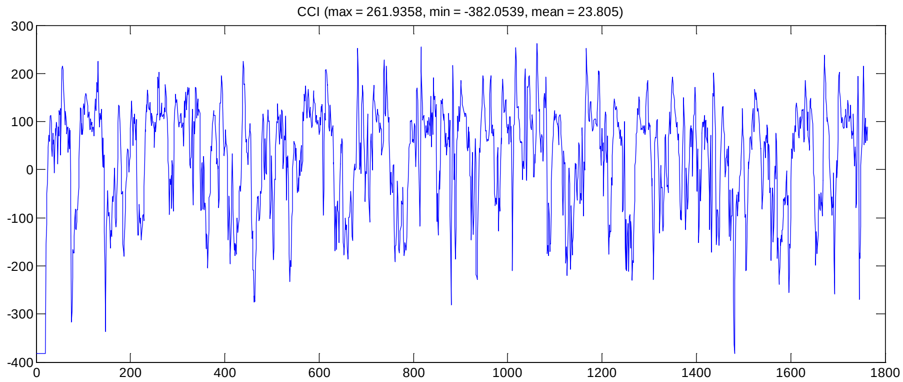


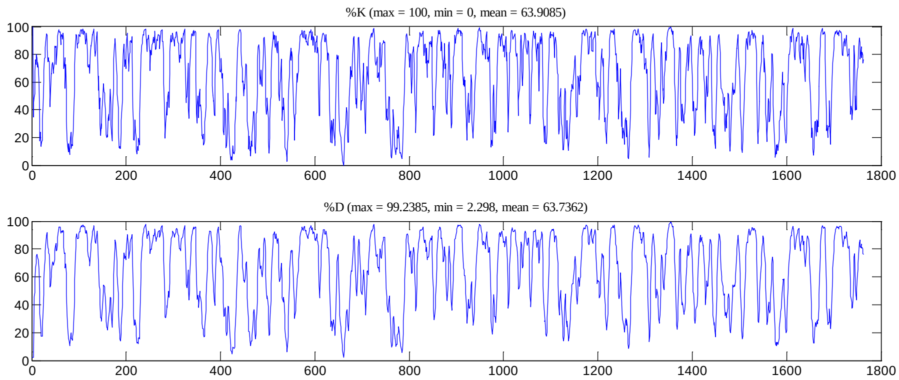




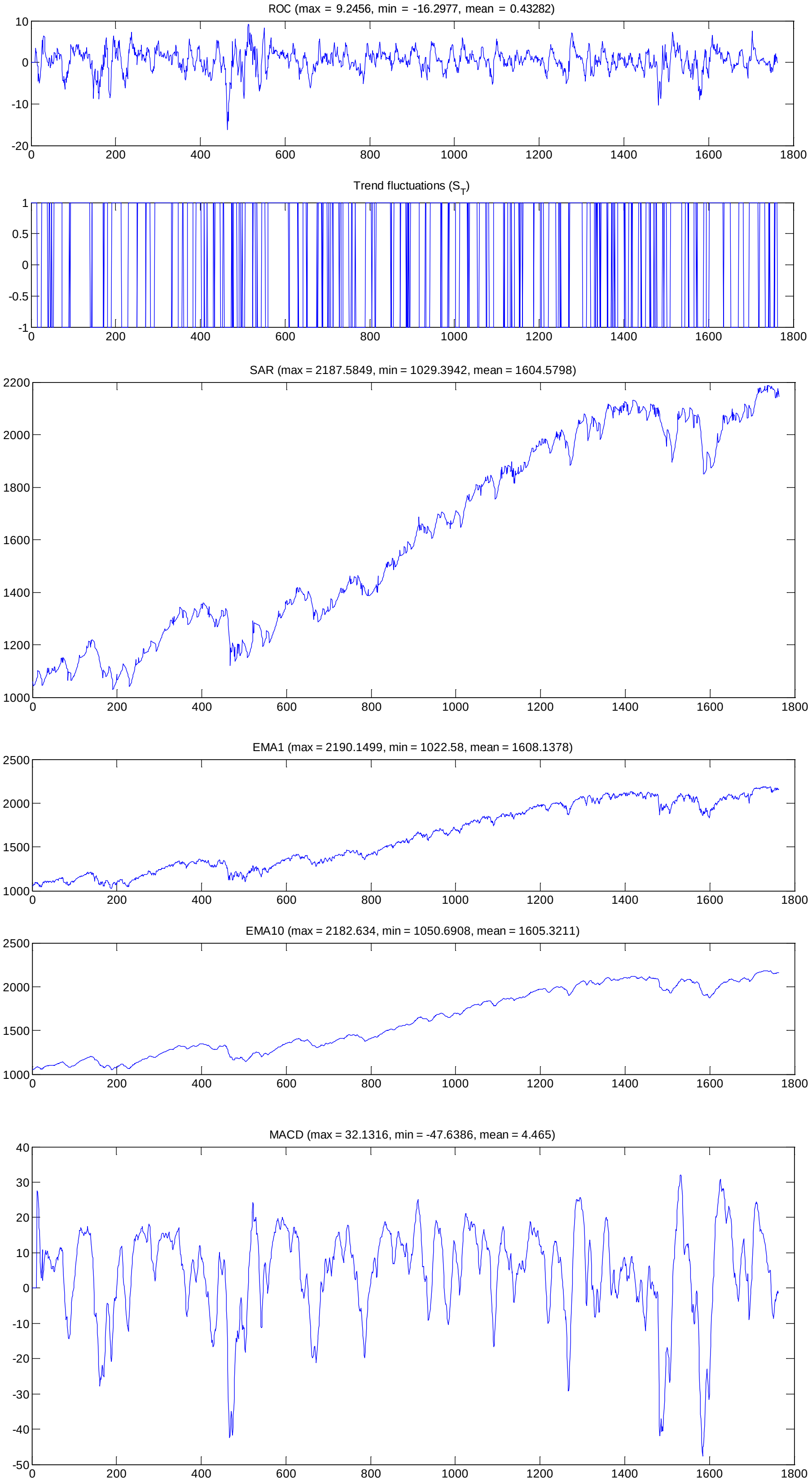


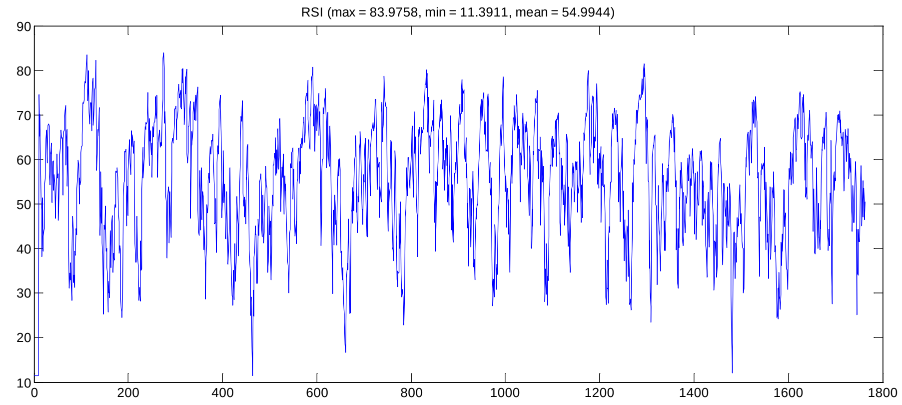

RSI (max = 86.853, min = 13.7264, mean = 55.4764)

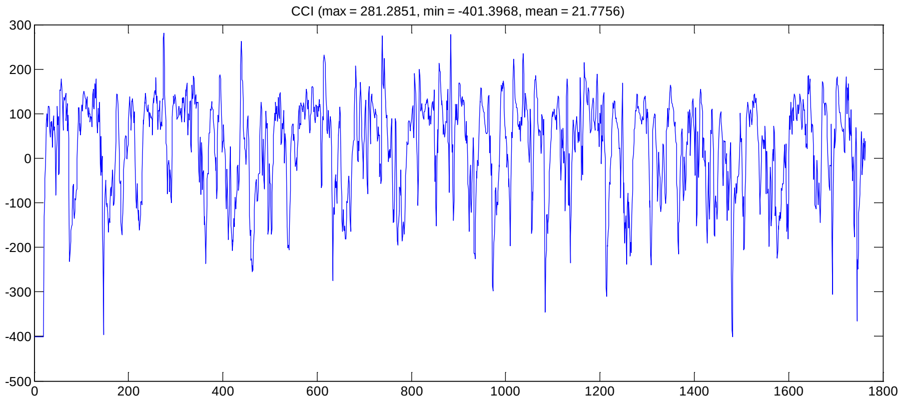

CCI (max = 261.9358, min = -382.0539, mean = 23.805)

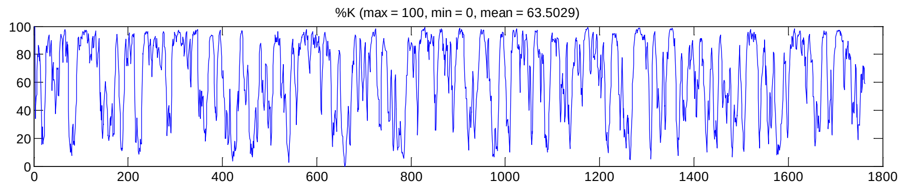

%K (max = 100, min = 0, mean = 63.9085)
%D (max = 99.2385, min = 2.298, mean = 63.7362)

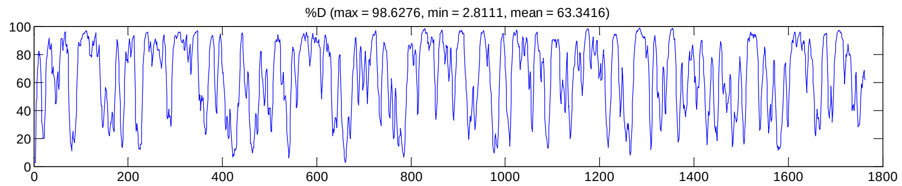
Appendix D: S&P 500




RSI (max = 83.9758, min = 11.3911, mean = 54.9944)


CCI (max = 281.2851, min = -401.3968, mean = 21.7756)


%K (max = 100, min = 0, mean = 63.5029)


%D (max = 98.6276, min = 2.8111, mean = 63.3416)