A HYBRID AND RELIABLE METHOD INTEGRATING DEPTH AND TECHNICAL ANALYSIS WITH MACHINE LEARNING TECHNIQUES FOR PREDICTING STOCK PRICES

HİSSE SENEDİ FİYAT HAREKETLERİNİN TAHMİNİ İÇİN MAKİNE ÖĞRENİM TEKNİKLERİ İLE DERİNLİK VE TEKNİK ANALİZİ ENTEGRE EDEN HİBRİT VE GÜVENİLİR BİR YÖNTEM

SEÇİL TABUROĞLU

Asst. Prof. Dr. Fuat AKAL
Supervisor

Submitted to Graduate School of Science and Engineering of Hacettepe University
as a Partial Fulfillment to the Requirements
for the Award of the Degree of Master of Science
in Computer Engineering.

2019

This work titled "**A Hybrid and Reliable Method Integrating Depth and Technical Analysis with Machine Learning Techniques for Predicting Stock Prices**" by **Seçil TABUROĞLU** has been approved as a thesis for the Degree of **Master of Science in Computer Engineering** by the Examining Committee Members mentioned below.

Assoc. Prof. Dr. Osman ABUL

Head ....

Asst. Prof. Dr. Fuat AKAL

Supervisor ....

Asst. Prof. Dr. Murat AYDOS

Member ....

Asst. Prof. Dr. Engin DEMİR

Member ....

Asst. Prof. Dr. Mehmet KÖSEOĞLU

Member ....

This thesis has been approved as a thesis for the Degree of **Master of Science in Computer Engineering** by Board of Directors of the Institute for Graduate School of Science and Engineering on ...../...../........

Prof. Dr. Menemşe GÜMÜŞDERELİOĞLU

Director of the Institute of

Graduate School of Science and Engineering

To My Family and my Dearie Nephew Zeynep Helin

ETHICS

In this thesis study, prepared in accordance with the spelling rules of Institute of Graduate Studies in Science of Hacettepe University,

I declare that

- all the information and documents have been obtained in the base of the academic rules
- all audio-visual and written information and results have been presented according to the rules of scientific ethics
- in case of using others works, related studies have been cited in accordance with the scientific standards
- all cited studies have been fully referenced
- I did not do any distortion in the data set
- and any part of this thesis has not been presented as another thesis study at this or any other university.

12/09/2019

SEÇİL TABUROĞLU

YAYINLANMA FİKRİ MÜLKİYET HAKKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanması zorunlu metinlerin yazılı izin alarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayınlanan "***Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge***" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H. Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

□ Enstitü / Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.

□ Enstitü / Fakülte yönetim kurulu gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren .... ay ertelenmiştir.

□ Tezim ile ilgili gizlilik kararı verilmiştir.

12 / 09 / 2019

(İmza)

SEÇİL TABUROĞLU

ABSTRACT

A HYBRID AND RELIABLE METHOD INTEGRATING DEPTH AND TECHNICAL ANALYSIS WITH MACHINE LEARNING TECHNIQUES FOR PREDICTING STOCK PRICES

Seçil TABUROĞLU

Master of Science, Department of Computer Engineering

Supervisor: Asst. Prof. Dr. Fuat AKAL

January 2019, 70 pages

It is quite complicated and difficult to predict the price changes direction of stock, because of the price changes of stock are non-linear. Some statistical methods are used to estimate these price changes direction, but these methods are often inadequate in complex stock markets. In this thesis study, a method which tries to predict the changes direction of price in capital markets based on BIST stocks is presented. This method is also the basis for fair and safe trading of stock market transactions with depth analysis. First of all, stocks which have similar financial structure were selected by using fundamental analysis. In the long term, fundamental analysis data were used to find the best yields. Then, the depth and technical analysis data were used to construct the feature vector. These data were obtained from the order books received from Borsa Istanbul. The information in depth analysis was used to estimate the future trend in financial prices. However, in the depth analysis, the fact that the sales order is too high may not mean that the price of the stock will decrease, but the fact that the purchase orders are too high may not mean that the price will increase. It is not the right method to decide whether the stock will rise or fall just by looking at the depth analysis. Therefore, technical analysis was used as a second indicator. In the technical analysis, the historical data and price movements of the stock are analyzed, and the possible future price movements are estimated. In the price estimation, backpropagation neural networks and random forest were used. In this thesis, the combined use of fundamental analysis, technical analysis and depth analysis has contributed to studies in this area.

Keywords: Order Book, Technical Analysis, Fundamental Analysis, Machine Learning, Backpropagation Neural Network, Random Forest

ÖZET

HİSSE SENEDİ FİYAT HAREKETLERİNİN TAHMİNİ İÇİN MAKİNE ÖĞRENİM TEKNİKLERİ İLE DERİNLİK VE TEKNİK ANALİZİ ENTEGRE EDEN HİBRİT VE GÜVENİLİR BİR YÖNTEM

Seçil TABUROĞLU

Yüksek Lisans, Bilgisayar Mühendisliği Bölümü

Tez Danışmanı: Doktor Öğretim Üyesi Fuat AKAL

Haziran 2019, 70 sayfa

Hisse senetlerinin fiyat değişimleri doğrusal olmayan bir yapıda olduğundan, hissenin fiyatının değişim yönünü tahmin etmek oldukça karmaşık ve zor bir iştir. Bu fiyat değişim yönünü tahmin etmek için bazı istatiksel yöntemler kullanılmaktadır. Fakat, bu yöntemler karmaşık olan hisse senedi piyasalarında çoğu zaman yetersiz kalmaktadır. Bu tez çalışmasında Borsa Istanbul (BIST) hisselerini temel alan, sermaye piyasalarında fiyat değişimini kısa süreler için tahmin etmek üzerine çalışan bir yöntem ortaya konulmuştur. Bu yöntem ayrıca derinlik analizi ile borsa işlemlerinin adil ve güvenli bir şekilde yapılmasına temel oluşturmaktadır. İlk etapta benzer finansal yapıda olan hisseler temel analiz kullanılarak seçilmiştir. Uzun dönemde en iyi getiri sağlayacak hisseleri bulmakta temel analiz verilerinden yararlanılmıştır. Daha sonra derinlik analizi ve teknik analiz verileri ile öznitelik vektörü oluşturulmuştur. Bu veriler Borsa İstanbul'dan alınan emir defterlerinden elde edilmiştir. Derinlik analizindeki bilgi, finansal fiyatlarda gelecekteki eğilimi tahmin etmek için kullanılmıştır. Fakat derinlik analizinde satış emrinin çok olması hisse senedinin fiyatının düşeceği anlamına gelmeyebileceği gibi alış emirlerinin fazla olması da fiyatının artacağı anlamına gelmeyebilir. Sadece derinlik analizine bakarak hisse senedinin yükseleceğine ya da düşeceğine karar vermek doğru bir yöntem değildir. Dolayısıyla ikinci bir gösterge olarak da teknik analiz kullanılmıştır. Teknik analiz yönteminde ise hisse senedinin geçmiş verilerinden yararlanılarak, gelecekteki muhtemel fiyat hareketleri hakkında tahminde bulunulur. Fiyat tahmininde geri beslemeli yapay sinir ağları ve rasgele orman kullanılmıştır. Bu tez çalışmasında temel analiz, teknik analiz ve derinlik analizinin birlikte kullanılması bu alanda yapılan çalışmalara katkı sağlamıştır.

Anahtar Kelimeler: Emir Defteri, Temel Analiz, Teknik Analiz, Makine Öğrenmesi, Geri Beslemeli Yapay Sinir Ağı, Rasgele Orman

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Asst. Prof. Dr. Fuat Akal for his patience, guidance, advice, criticism, encouragements and insight throughout the research.

I wish to thank my parents Rifat and Zeynep Taburoğlu, for their love, encouragement, moral support, personal attention and care; I could not have done it without them. And a very special thanks to my dear brother Vahit Eren Taburoğlu for giving me moral and technical support. Without his help, this would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations

BIST	Borsa Istanbul
NASDAQ	American stock exchange
ANN	Artificial Neural Network
SVM	Support Vector Machine
KAP	Turkish Public Disclosure Platform - PDP
MLP	Multilayer Perceptron
SMA	Simple Moving Average
WMA	Weighted Moving Average
ROC	Rate of Change
ATR	Average True Range
EMA	Exponential Moving Average
MOM	Momentum
MACD	Moving Average Convergence Divergence
RSI	Relative Strength Index
CCI	Commodity Channel Index
MAE	Mean Absolute Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error

1. INTRODUCTION

The main objective of investors traded on the stock market is to make a profit by investing in high yielding securities. Therefore, it is very important to provide the right strategy to predict the direction in which stock prices will change. However, this task is very difficult to achieve due to non-linear, dynamic and complex nature of stock markets [1]. There many empirical researches in the literature [13, 14] but a few of them have been conducted on the Turkish stock market (BIST). Since the stock market of each country will exhibit different behavior patterns, the study will have different results. Also, depth analysis is a subjective analysis, it will differ according to each market and even every stock. Studies are mostly conducted only using either technical or fundamental analyses. In fundamental analysis, political and economic changes, financial news, financial statement, inflation, policy rate applied by the country, current account deficit / current surplus ratio data and share-price ratio etc. are used [2]. In technical analysis, market activities, past prices and values of stocks have been evaluated as a time series. Depth analysis is another type of analysis which is not found in the literature as much as technical and fundamental analysis. Depth of market data may also be called order book. The main task of this thesis is to use technical analysis and depth analysis (order book) to predict stock price index movement.

Detailed information about the order book [3] is given in the next section. Some statistical methods, machine learning and data mining have been used in this thesis. The stock exchange data, especially the non-linear modeling of stocks, has become popular use case for artificial intelligence and machine learning techniques. Artificial Neural Networks, Random Forest and Support Vector Machine are the most used algorithms for predicting stock price index movement.

Studies carried out within the scope of this thesis are listed below.

1. Fundamental analysis is used for stock selection to determine maximum yield. Stocks which have maximum yield in long term and same financial structure are tried to be selected by fundamental analysis. Fundamental valuation analysis is applied in this thesis. This analysis consists of mathematical calculation by using

some financial variables such as income statement, financial investment, investment property. These financial variables are taken from KAP (Turkish Public Disclosure Platform - PDP). Detailed information is given in the next sections.

2. Features have been identified and feature vector has been generated. At this point technical indicators and the order book data are processed. Order book is purchased as a csv file from Borsa Istanbul historical data sales for academic purpose. The most important contribution of this study is that both depth (level) and technical analysis are used together for BIST.
3. Feature vectors are generated for data at every 15 minutes and obtained prediction results are compared.

1.1. Research Contribution

There are a large number of studies in the literature about the predicting price index movement. Some of them were made using fundamental analysis [4, 5]. In the technical analysis, which is another approach, when the estimation of the metrics for the stocks was made; machine learning techniques such as signal processing methods, artificial neural networks and support vector machines were used [6, 7]. In addition to these, there are analyses made based on the order and message books [8, 9]. Depth analysis was carried out on foreign exchanges such as NASDAQ (American stock exchange) but extensive search on literature show that, there has been no specific research for prediction BIST (Borsa Istanbul) by using depth analysis (order book). Since the stock exchange of each country will exhibit different behaviors, it will have different results. Depth analysis is a subjective analysis and will vary according to each stock exchange or even every stake. The contribution of this study (the most important contribution) is depth analysis on BIST.

The second contribution is determining the stakes that will provide the best yield in the long term by using the fundamental analysis in the selection of stakes.

Another contribution of this thesis is to use both technical analysis and depth analysis together for prediction price index movement on BIST. Extensive search on literature show that, there has been no specific research by using both technical and depth analysis (order book) in any stock market. Depth analysis is very useful for events such as gathering and distributing stocks. However, this method is subject to abuse and it can be used intentionally by manipulators to influence the people who follow it [10]. It would not be very accurate to analyze only the depth (level) analysis for prediction price index movement. Using these two methods together provides more accurate results. With this approach, protection against manipulation and counteracting manipulation in stocks will be an indirect contribution of this thesis. For fair and safe stock exchange transactions artificial intelligence will be provided with depth and technical analysis.

As a result, no analysis can be found in any study with the combination of fundamental analysis, technical analysis and depth analysis. Due to the combination of these methods, it is thought that this study produces more accurate results than the available research.

1.2. Thesis Outline

This thesis consists of five sections. In the first section, aim of this study is given. The contribution of this thesis to the academy is also explained in this section. In the second section, the studies related to the prediction price index movement based on technical or depth analyses by using machine learning techniques in the literature are mentioned. The third section provides general information about the stock market exchange concept. In the fourth section, the conducted study is described in detail. In the fifth section, the results of the estimation model, visual outputs and analysis of these outputs are presented. The last section concludes the thesis and discusses possible future works.

2. LITERATURE SURVEY

There are several works in the literature considering technical and depth analyses to predict stock movements. Such works are surveyed in this section.

2.1. Machine Learning Techniques Using the Technical Analysis

In the literature there are a number of models combining technical analysis with machine learning techniques are available for prediction of stock price index movements. Technical analysis is based on historical prices and tries to predict future prices according to the past behaviors [6]. Technical indicators are generated from these historical values and used for prediction. SMA (Simple Moving Average), WMA (Weighted Moving Average), ROC (Rate of Change), ATR (Average True Range), EMA (Exponential Moving Average), MOM (Momentum), STCK% (Slow Stochastic), STCD% (Fast stochastic), MACD (Moving Average Convergence Divergence), RSI (Relative Strength Index), WILLR% (Williams), A/D Osc (Accumulation/Distribution Indicator) and CCI (Commodity Channel Index) are the most used indicators [6, 11, 12]. Detailed information regarding these indicators can be found in Preliminaries Section 3.2. Mizuno et al. conducted a study by using ANN on Tokyo stock exchange to predict trading signals and they achieve a success of 66% [11]. They used MA (moving average) indicators with several variations, RSI, Price deviation from the moving average and psychological line indicator. In this study [34], technical analysis and fundamental analysis were combined in Brazilian stock exchange (BM&FBOVESPA). Macroeconomic indicators (fundamental) and MACD, RSI, Stochastic index, OBV index, MA, Bollinger bands, MOM and WILLR% technical indicators were used for neural network algorithm. The best performance of the model is 93.62%. Khaidem et al. used random forest algorithm and used MACD, ROC, RSI, Williams %R, Stochastic index and OBV (On Balance Volume) as inputs to train their model [37]. Then, they compared their results to other studies in literature. Experiments with random forest outperformed the models as seen in various papers [6, 41, 42, 43]. [6] used ten technical indicators (SMA, WMA, MOM, STCK%, STCD%, RSI, MACD, WILLR%, A/D Osc and CCI) for prediction of price index movement in Indian Stock market. They compared Artificial Neural Network (ANN), Support Vector Machine (SVM), random forest and naive-Bayes methods. They tried to find the best prediction by best parameters for predictor models using 10 years of historical data. The experiment results showed that Random Forest outperformed the other methods.

2.1.1 Studies for BIST

Kara et al. used ten technical indicators for prediction of price index movement [12], such as the study in [6]. The same indicators as the above have been used. They compared ANN and SVM techniques. According to the results, ANN is significantly more accurate (75.74%) in terms of prediction accuracy compared to the other model. They used daily closing price movement of BIST100 data from January 1997 to January 2018.

Güreşen et al. used neural network hybrid models to forecast stock exchange movements in BIST. They used data from January 2003 to March 2008 [15].

Also Gunduz et al. predicts hourly movement of Borsa Istanbul (BIST) 100 stocks by using Convolutional Neural Network (CNN) and they obtained 0.563 F-measure rate [16]. They used data from January 2011 to December 2015. Ergür et al. used multi-layer perceptron and Support Vector Machines in BIST [35]. The estimation performance average of all stocks had been calculated as 67% as the macro-average F-score. They used 5 months data and they predicted 15 minutes price change direction.

[36] used 45 technical indicators for improved stock price forecasting performance in BIST. BIST100 index data were used between June 2005 and October 2013. They used a hybrid model based on an intuitive optimization methodology (Harmony Search and Genetic Algorithm) and ANN. Their proposed model had capability in variable selection and in determining the number of neurons in the hidden layer.

2.2. Machine Learning Techniques Using the Order Book Data

This section includes research for markets other than BIST because no study using the order book for BIST in the literature. Order book processes operate in very high frequencies such that more than one order can be entered at the same time.

Kercheval et al. used different order price levels (best ask and bid price and volume) and derivatives as features in NASDAQ [9]. They tried to predict mid-price movement and price spread crossing. They used multi-class support vector machines for a very small train and test data (4000 samples) due to limitations of the non-linear kernel-based classification model. They obtained with high accuracy F1 score. However, accuracy rate of the results can be misleading.

Ntakaris et al. made a mid-price prediction for high frequency limit order markets [24]. Limir order is the type of order in which both price and quantity are entered. If there is no partial or complete transaction at the time of entry, the unrealized part is entered into the passive orders of the stock in accordance with the price and time priority order. They used time-sensitive and time-insensitive features in NASDAQ for 400000 time series samples. They used event-based data representation because it avoids dataset imbalance. Baseline performances were evaluated by two algorithms: Ridge Regression (RR) and SLFN Network-based Nonlinear Regression. The average performance of the model is 46% (F1 score) for both methods.

Tsantekidis et al, have used deep learning methods in forecasting stock prices [28]. They propose a model based on Convolutional Neural Networks (CNNs) with both 2D and 1D convolution masks to estimate stock prices from the NASDAQ limit order book. Feature data consists of 10(n) level bid and ask prices and volumes. They compared their results with Linear SVM and MLP. CNN model performed better than other models. Maximum CNN F1 score is 59.44% when SVM is 49.42% and MLP is 55.95%.

Dixon et al. used Recurrent neural networks (RNNs) which are types of Artificial Neural Networks (ANNs) on the S&P500 E-mini futures level II data to estimate the price flip of the next event [40]. They proposed a model that solves the short sequence classification problem.

Another method used in price direction forecasting is random forest. Kanagal et al. used Random Forest and compared results with SVM and Stochastic Gradient Descent [23]. Features were the best ask and bid prices. AAPL, FB, ORCL, MSFT and GOOG stocks were used. Random forest has better results than SVM and SGD. Han, James, et al also used random forest with same feature set as [9] for price change forecast. According to their results, random forest (~85%) is more accurate than SVM (~49%) for three direction (upward, downward and stationary). Booth et al. used an ensemble of ensembles model of random forest [38]. Order book data was given from Europe Exchange (BATS Chi-X). They used price (last prices before trade), spread (bid/ask spread) and liquidity of the book (e.g. order arrival and cancelation rates) as features. They suggested a backward elimination method to determine features. The performance of the system is compared with other algorithms, such as liner regression, neural networks and support vector regression. The results indicate that Random Forest is the best algorithm (random forest produces 15% more accurate results) to predict the price.

Fletcher et. al were conducted with Multiple Kernel Learning (MKL) to predict the direction of price movements [39]. MKL provides to combine a range of different kernels for different input properties. They used SimpleMKL and LPBoostMKL to train multiclass SVM on the FX market.

In the light of previous studies and to the best knowledge of the authors, it is show that Artificial Neural Network (ANN) and Random Forest generate more accurate results.

3. PRELIMINARIES

3.1. Fundamental Analysis

Fundamental analyses is the most used technique to predict direction of the price index movement. Fundamental indicators are affected (directly or indirectly) by many different macro-economic factors such as export, import, money supply, interest rate, inflation rate, exchange rates and financial ratios calculated from financial statements of firms either. Furthermore, news is a fundamental parameter because news has effect on supply-demand relationship and market psychology.

There are many fundamental analysis methods that makes the analysis of stock value in the literature [4], [30]. That methods are Dividend Discount Models (DDM), Ratio Analysis, Discounted Cash Flow Models (DCFM) and Residual Income Valuation Model (RI). Within the scope of this thesis, fundamental valuation analysis (Buffett's Methodology [17]), which is a type of DCFM, is applied. Valuation analysis contains analysis of the company's liquidity, efficiency, profitability, and debt coverage. The aim is to estimate the future performance of the selected company to determine the true value of the stock and to decide whether or not to invest. In this method, the accounts in the financial statements are formulated by means of the company's liquidity, financial structure, profitability and activities. The financial statements used in the analysis are the balance and income statement tables. These tables are obtained from KAP (Turkish

Public Disclosure Platform - PDP) [29]. Financial tables are given in Table 1 and Table 2.

Table 1 Tables Used in Financial Analysis

Balance (Active) - Current Assets	Balance (Active) - Non-current Values	Balance (Passive) - Short Term Liabilities	Balance (Passive) - Long term liabilities	Balance (Passive)- Shareholder's equity
Cash and cash equivalents	Accounts Receivable	Financial Liabilities	Financial Liabilities	Paid Capital
Securities	Other Receivable	Trade Liabilities	Trade Liabilities	Capital Reserves
Accounts receivable		Other Liabilities	Other Liabilities	Profit Reserves
Other Receivables	Financial Real Assets	Advances Received	Progress Payments for Long-term Construction and Repair Projects	Previous Year's Profits
Stocks	Tangible Real Assets	Taxes Payable and Other Financial Liabilities	Advances Received	Previous Year's Losses
Cost of Long Term Construction Contracts	Intangible Real Assets	Provisions for Known Liabilities	Provisions for Known Liabilities	Net Income of the Period
Income and Expense Accruals for the Next Months	Other Real Assets	Income and Expense Accruals for the Future	Income and Expense Accruals for the Future	

Other Current Assets	Depletable Assets	Other Long-term Liabilities	Other Long-term Liabilities
Long Term Prepaid Expenses and Accrued Income	Long-term Prepaid Expenses and Accrued Income		
Investment Property			
Financial investment			

Table 2 Tables Used in Financial Analysis- Statement of Income

Gross Sales	Sales Deductions	Net Sales	Extraordinary Expenses and Losses
Financial Expenses	Ordinary Profit or Loss	Extraordinary Incomes and Profits	Cost of Sales
Gross Profit or Loss	Operational Expenses	Operating Profit or Loss	Income and Profit from Other Ordinary Operations
Expenses and Losses from Other Ordinary Operations	Period Income Tax Provision		

3.2. Technical Analysis

Technical analysis tries to predict the future price of stock by using historical stock prices. Technical analysis includes mathematical, statistical and financial information, formulas and graphical analysis. In the analysis, the average price and volume values are taken, and the values are compared with the current price and volume values and the direction of stock exchange or stock is estimated. These processes are carried out with the help of technical indicators. Technical indicators are the mathematical models that give an idea about the direction of the price or the continuation of the trend in technical analysis. There are many indicators used in technical analysis. Technical indicators, which are commonly used in past studies have been used to create the attributes of this thesis [6, 12, 16]. These technical indicators are detailed in the following subsections. Formulas are taken from [31].

3.2.1. Moving Average Convergence Divergence (MACD)

MACD compares the short-term price trend with the long-term price trend. The mathematical formulation can be explained as the subtraction of the 12-day exponential moving average (EMA) from the 26-day exponential moving average (EMA). If the result is greater than zero, price trends are considered to be in the up direction; if the result is less than zero, price trends are considered to be in the down direction. MACD is formulated as follows:

$$MACD = EMA_{(12)} - EMA_{(26)} \tag{1}$$

$$EMA = (Closing\ Price - EMA(prev.day)) * Multiplier + EMA(prev.day) \tag{2}$$

Where:

$$Multiplier = 2/(Number\ of\ Days\ in\ EMA + 1) \tag{3}$$

3.2.2. Simple Moving Average (SMA)

The simple moving average is calculated by dividing the total value of the closing price by the time interval. Thus, how far the current price is from the historical average can be determined. If the result above the price, trend is in the down direction; if the result is below the price, price trend is in the up direction. SMA is formulated as follows:

$$SMA = \frac{1}{t} \sum_{i=1}^{t} Closing\ Price(i) \tag{4}$$

3.2.3. Weighted Moving Average (WMA)

This indicator gives more weight to today's price than older prices. For example, when 26 periods are selected, the closing price of the current transaction is multiplied by 26 and the closing price of the previous transaction is multiplied by 25. WMA is formulated as follows:

$$WMA = Closing\ Price(t) * t + Closing\ Price(t-1) * (t-1) + \cdots + Closing\ Price(1) \tag{5}$$

3.2.4. Relative Strength Index (RSI)

RSI is the most popular indicator [44]. RSI compares the decreases and increases in the period to the current price movement. When making this calculation, RSI collects the decreases and increases of the days within the specified period separately, calculates their averages and compares with the current price level. RSI takes a value between 0 and 100.

If RSI value is less than 30, it is considered there is an oversold. If RSI value is greater than 70, it is considered there is an overbought. RSI is formulated as follows:

$$RSI = 100 - \frac{100}{1 + \frac{Average\ Gain\ of\ n\ days\ UP}{Average\ Loss\ of\ n\ days\ DOWN}} \tag{6}$$

3.2.5. Rate of Change (ROC)

ROC calculates the percentage change in price between periods. ROC is formulated as follows:

$$ROC = \frac{Closing\ Price(current) - Closing\ Price\ (n\ periods\ ago)}{Closing\ Price\ (n\ periods\ ago)} * 100 \tag{7}$$

3.2.6. Average True Range (ATR)

ATR is used to monitor volatility; this indicator is not concerned with the direction of price movements. ATR is calculated by taking the average of the highest and lowest difference of each movement in n time period. ATR is formulated as follows:

$$TR(True\ Range) = max((High\ Price - Low\ Price), abs\big(High\ Price - Closing\ Price(prev)\big), abs(Low\ Price - Closing\ Prices\ (prev))) \quad (8)$$

$$ATR = \frac{1}{n}\sum_{i=1}^{n} TR\ (i) \quad (9)$$

3.2.7. Bollinger Bands

Bollinger Bands consist of a center, upper and lower price bands. The mathematical formula of the Bollinger bands shows that how far the prices are moving away from the moving average. Bollinger Bands is formulated as follows:

$$Middle\ Band\ =\ 20 - day\ simple\ moving\ average\ (SMA) \quad (10)$$

$$Upper\ Band\ =\ 20 - day\ SMA\ +\ (20 - day\ standard\ deviation\ of\ price\ x\ 2) \quad (11)$$

$$Lower\ Band\ =\ 20 - day\ SMA\ -\ (20 - day\ standard\ deviation\ of\ price\ x\ 2) \quad (12)$$

3.2.8. Momentum (MOM)

Momentum Indicator shows the value of the price changes in % within the specified period. MOM is formulated as follows:

$$Momentum = Closing\ Price(current) - Closing\ Price(n\ days\ ago) \quad (13)$$

3.2.9. Commodity Channel Index (CCI)

CCI formula is given in below:

$$CCI = \frac{Typical\ Price - Moving\ Average}{0.015 * Mean\ Deviation} \quad (14)$$

$$\text{Typical Price (TP)}\ =\ (\text{High}\ +\ \text{Low}\ +\ \text{Close})/3 \quad (15)$$

Mean Deviation is the deviation of the average price and moving averages.

3.2.10. Stochastic K%D

Formulas for calculating %K and %D lines are given in below:

$$\%K = 100 * \frac{Closing\ Price - L(n)}{H(n) - L(n)} \quad (16)$$

$$\%D = SMA\ (\%K, 3) \quad (17)$$

H(n) = The highest price in last n samples

L(n) = The lowest price in last n samples

3.3. Order Book

Borsa Istanbul uses message book and order book information to keep the intraday data. Intraday message book contains trading information, market type information, the date of the transaction, time, session information, information about which purchase order and which sale order are met each other and the price at which the share is traded. Order book contains information such as the stock information, the price and the amount of the order, the date of the order and the validity period. The order book is the place where buyers and sellers can see clearly how much they buy or sell at a price. Orders are actions that enable buying or selling on the order book. In this thesis, limit orders are used. Limit order is the standard order type that is sent by determining a price. Limit orders are turned into trade by matching the best priced orders on the opposite side until the order amount is met. The stock orders work with the FIFO (First In First Out) algorithm. That is, first entered order is processed at first. If the orders are not processed partially or completely from the moment they are entered, the unrealized portion of the order waits in order of price and time. In order to better understand how it works, let us consider the following scenario:

Suppose that at the time T=0, all active levels seen on the board are given in the Figure 1. Provided that the orders are kept as at the time T=0 as is,

then T=1 a new order comes as:

Buy: 5.08 (50000 Lots-Active Order)

Then, the lowest price of ask is executed and this process continues until the whole order quote is satisfied. First, ask which has 5.04 price and 10000 volumes is executed, after that the ask of 5.06 price and 20000 volumes is executed. Total executed volume is 30000 at this point and third ask is executed to satisfy the remaining order quote. New situation is given in Figure 1.

Figure 1 Sample Ask Execution

Bid Volume	Price	Ask Volume
	6.00	20000 (Level 4)
	5.08	30000 (Level 3)
	5.06	20000 (Level 2)
	5.04	10000 (Level 1)
40000 (Level 4)	5.02	
20000 (Level 3)	5.0	
20000 (Level 2)	4.8	
10000 (Level 1)	4.6	

Execution Order	Execution Price	Execution Volume
1	5.04	10000
2	5.06	20000
3	5.08	20000

New Situation

Bid Volume	Price	Ask Volume
	6.00	20000
	5.08	10000
40000	5.02	
20000	5.0	
20000	4.8	
10000	4.6	

Provided that the orders are kept as at the time T=1 as is,

further at T=2 new orders come as;

Sell 5.0 (6000 Lots-Active Order)

Then, the highest price of bid is executed, and this process continues until the whole order quote is satisfied. First, bid which has 5.02 price and 40000 volumes is executed, after that ask order consists of 5.0 price and 20000 volumes is executed. New situation is given in Figure 2.

Figure 2 Sample Bid Execution

Bid Volume	Price	Ask Volume
	6.00	20000 (Level 2)
	5.08	10000 (Level 1)
40000 (Level 1)	5.02	
20000 (Level 2)	5.0	
20000 (Level 3)	4.8	
10000 (Level 4)	4.6	

Execution Order	Execution Price	Execution Volume
1	5.02	40000
2	5.0	10000

New Situation

Bid Volume	Price	Ask Volume
	6.00	20000
	5.08	10000
10000	5.0	
20000	4.8	
10000	4.6	

The main subject of this thesis is depth analysis. In order to estimate the price in the stock market, looking at buying sales analysis to predict whether the stock trend direction will rise or not is called stock depth analysis. The data used in depth analysis are kept in the order book. Order book data is reported on a monthly basis by Borsa Istanbul. Definitions of the intraday order book items which are used in this thesis are given below. All order book items are given in the Table 21.

Modified Date and Time: Displays the date and time of replacement of the order. The order in the book is made on the value in this column.

Operation Code: Stock name (code) is obtained from this field. This field is a combination of stock code and stock property code. For instance, in ADNAC.E, stock code is denoted as ADNAC, property code of stock is denoted as “E”.

Bid/Ask: [Bid: A, Ask: S]. Bid refers to the price the buyer is willing to pay. Ask refers to the price the seller wants to sell.

Order Duration ID: Specifies the date at which the entered order will take place until the end of the day in the system. This field can be monitored at the end of the day at which the order will be deleted. This field could take DAY, GTC (Good Till Cancelled), GTD (Good Till Data), GTT (Good Till Time), IMMEDIATE, and SESSION tags. In this study, immediate order is not added to depth analysis, because this order is valid only for the moment it is entered. Orders are cancelled end of the session or day, by the system. For instance, in Table 3 at first row, order with number "604FA28100463BB0" is created as daily order, then order expires end of the day as can be seen in the end of the table below.

Table 3 Sample Order Book – Order Duration ID

Order No	**Modified Date and Time**	**Operation code**	**Bid/ Ask**	**Order Duration ID**	**Change Reason**	**Session**
604FA28100 463BB0	2.01.2018 10:35	GEDIK.E	B	DAY	New	P_CONTINUOS_ OPERATION
604FA28100 3D4DCE	2.01.2018 10:22	GEDIK.E	A	DAY	New	P_CONTINUOS_ OPERATION
604FA28100 3D4DCE	2.01.2018 10:22	GEDIK.E	B	DAY	Trade	P_CONTINUOS_ OPERATION
604FA28100 453613	2.01.2018 10:33	GEDIK.E	A	DAY	Trade	P_CONTINUOS_ OPERATION
604FA28100 453614	2.01.2018 10:33	GEDIK.E	B	DAY	New	P_CONTINUOS_ OPERATION
…….............	……........... ...	……...	……...	…	..	...
604FA28100 463BB0	2.01.2018 10:35	GEDIK.E	B	DAY	Expired	P_END_OF_THE_ DAY
604FA28100 453614	2.01.2018 10:33	GEDIK.E	B	DAY	Expired	P_END_OF_THE_ DAY

Status: Provides information on the validity of the order. Whether it is active etc. Only active orders are used in this study.

Change Reason: Shows the reason for the change on the order. For example, executing, cancellation, change of quantity and so on.

Quantity: Amount of the order.

Remaining Quantity: Remaining amount of the order

Price: Price of order

Session: The trading session is the time period that matches the daytime trading hours.

Joint Contract No: It is the unique number formed by merging the member transaction numbers for the bid and ask. The moment a buy order price and volume match a sell order, the trade is executed and "Change Reason" field is set to "Trade". The contract number is also determined for the trade buy and sell orders. This field is important because trade volume and price are determined through this area. Sample scenario is given in Table 4. Firstly, a sell order is given with order id "6058662", then a buy order is given, which has "6054561" order id, and this order matches the previous order. Then, trade is executed and unique "Joint Contract No" is generated.

Table 4 Sample Order Book – Joint Contract No

Order No	**Modified Date and Time**	**Operation code**	**Bid/Ask**	**Price**	**Change Reason**	**Joint Contract No**
6058662	2018-01-03 09:40:33	GEDIK.E	B	4.8	New	
6054561	2018-01-03 10:40:43	GEDIK.E	A	4.8	New	
6058662	2018-01-03 10:40:43	GEDIK.E	B		Trade	69403400322
6054561	2018-01-03 10:40:43	GEDIK.E	A		Trade	69403400322

3.4. Machine Learning Techniques

In the light of previous studies and to the best knowledge of the authors, it is show that Artificial Neural Network (ANN) and Random Forest generate more accurate results. Therefore, ANN and RF have been used in this thesis. These techniques are detailed in the following subsections.

3.4.1. Artificial Neural Network

There are several types of artificial neural network, some of them are Feed Forward Neural Network, Back Forward Neural Network, Recurrent Neural Networks and Multilayer Perceptron. In this study, the results have been obtained using the multilayer perceptron for the prediction stock price index movement. Multilayer perceptron is the most widely used model in these models and it is efficiency in the nonlinear time series prediction process.

3.4.1.1. Multilayer Perceptron Neural Network

Artificial neural networks have emerged as a result of efforts to artificially simulate the working system of the human brain. It is a computational model consisting of several process units connected to each other. Process units are called as neurons. Neuron is a simple element that collects and transmits incoming signals to the next one. There are many descriptions of neural network [32]. In this thesis, multilayer perceptron has been used to predict the direction of price change. Therefore, in this section, a summary information is given about the multilayer perceptron.

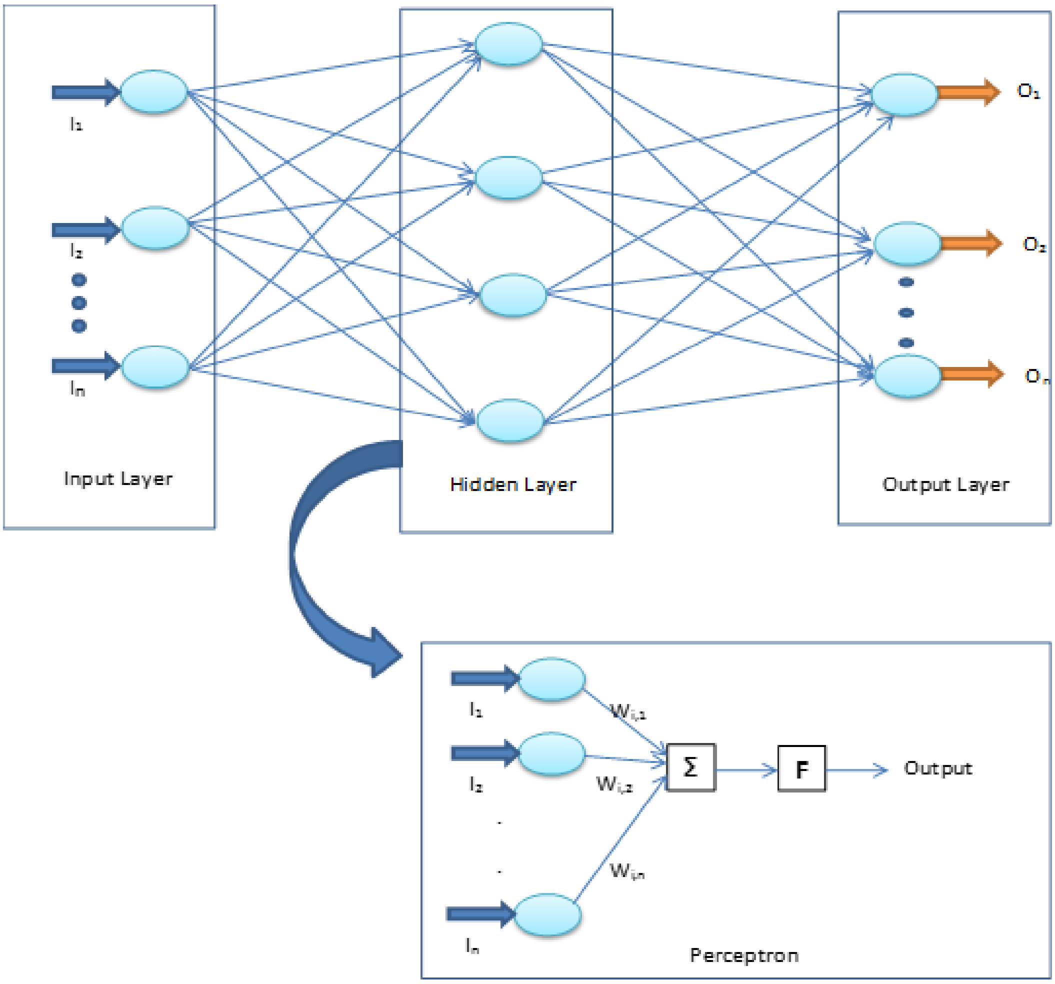
Multilayered perceptron is a model of ANN which is used in the solution of nonlinear problems, it consists of three layers named as the input layer, hidden layer and output layer as shown in the Figure 3.

Input Layer: Process elements in this layer are responsible for transferring the data from the outside world to the intermediate layers.

Hidden Layer: The information coming from the input layer is processed and sent to the output layer. There may be more than one hidden layer for a network.

Output Layer: Process elements in this layer process the input from the input layer and generates the output.

Figure 3 Multilayer Perceptron



Each layer consists of neurons. Neurons are connected with each other and each link has a numerical weight that expresses the importance of its input. A neural network performs learning by repeatedly adjusting these weights. There are transitions between layers called forward and backward propagation. In the forward propagation phase, the output and error value of the network are calculated as shown in the Figure 3. This is mathematically formulated as follows:

$$Output = F\left(\sum_{i=1}^{n} W_i\, I_i + \Phi\right) \tag{17}$$

Φ= The bias or threshold

F = Activation function (sigmoid, tanh …). Activation function determines the output value by calculating the net inputs to the cell.

W = Weighted Vectors

I = Input

In the back-propagation phase, the hidden-layer link weight values are updated to minimize the calculated error value.

$$E = \sum \frac{1}{2} (Target - Output)^2 \quad (18)$$

The gradient descent calculation can be used when calculating the weight change. Accordingly, weight change is calculated as follows:

$$\Delta W_i I_i = -\alpha \frac{\delta E}{\delta W_i I_i} + \mu W_i I_i (k-1) \quad (19)$$

k = Iteration Number

α = Learning Rate

3.4.2 Random Forest

Random forest is a model consisting of unpruned classification and regression trees created by randomly selecting samples in the training data [33]. The advantage of randomly selecting is to obtain less correlation between the trees in the forest. Random Forest can be considered as evolved from decision trees. [33] also says that it is a fast and non-specific method. According to the [33], the main idea of random forest is that: Each decision tree in the forest is created by selecting the sample from the original data set with the bootstrap. Instead of applying the decision tree algorithm over the entire data set, the data is divided into smaller subsets and the decision tree algorithms are applied to these subsets. This method is called as bagging. The decision is made according to the highest-rated subset results. Random forest receives 2 parameters: Number of tree (N) and number of variables used in the node (m). Therefore, random forest algorithm is so easy to use for the user. The working principle of random forest is as follow:

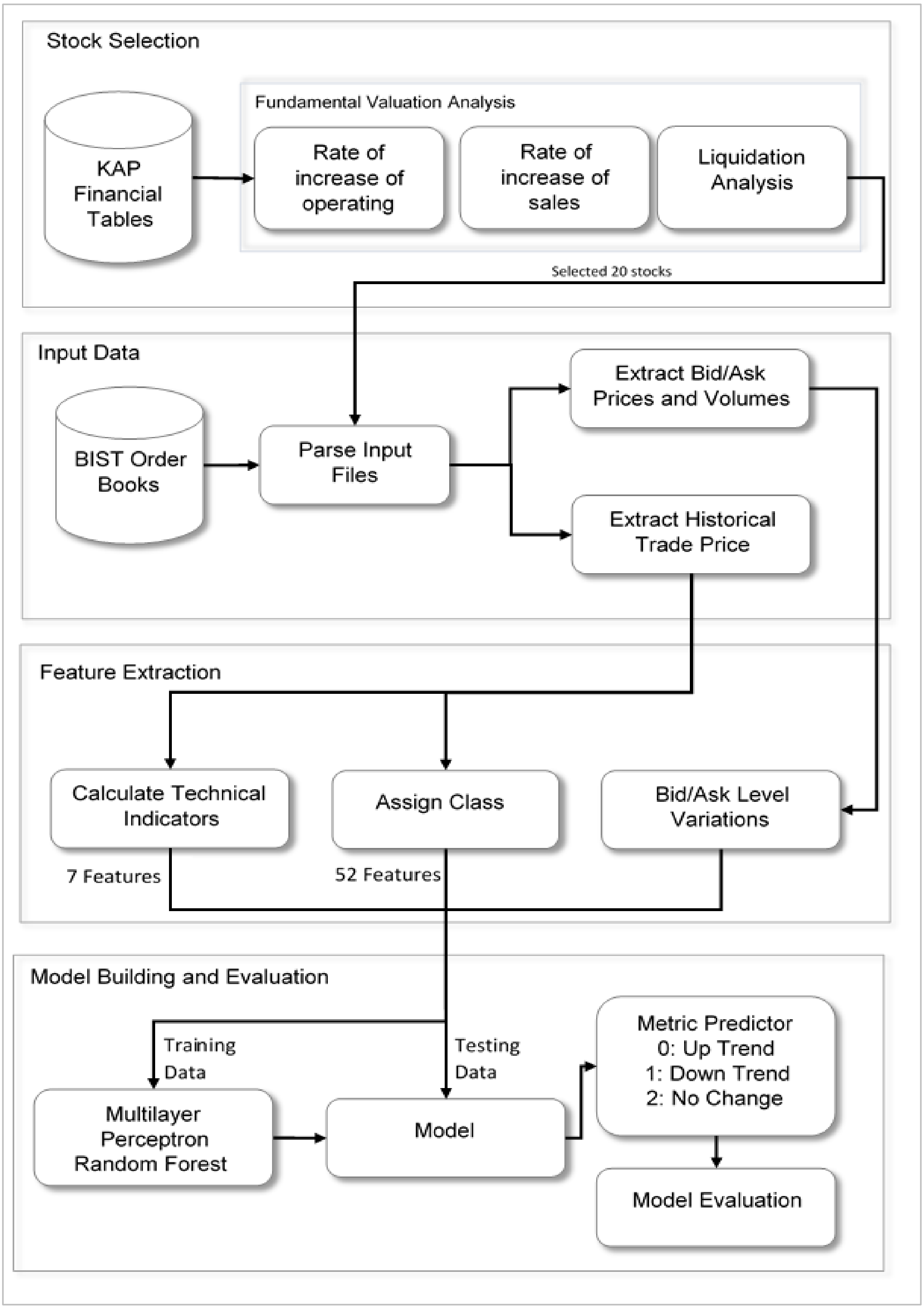
First, boot instances are generated from 2/3 of the training data. The rest of the training data set, known as out-of-bagged (OOB) data, is used to test for errors. Subsequently, the tree is grew up without pruning from each sample. At each node, m variables are randomly selected from all variables and the best branch is determined. Homogeneity of classes is calculated by gini index. If the gini index is increase, class homogeneity is decrease. When the gini index reaches zero the tree branching process ends. When all N trees are produced, the estimation class is determined based on the estimation results obtained from N trees. Random forest can also be used to feature selection.

4. METHOD

Firstly, the best 20 stocks are chosen by using fundamental analysis. In fundamental analysis, liquidity calculation, operating profit, sales are calculated and stocks in good condition are selected. Then, technical indicators and derived data from the order book are used to generating feature vector (Depth and technical analysis are used together in BIST) for selected stocks. Finally, data pre-processing and analysis, Random Forest and Neural Network algorithms are experimented. For each stock, these algorithms are used to train the models and obtained results have compared each other.

Details of the proposed model shown in Figure 4, will be given in the subsequent section.

Figure 4 Proposed Model of This Thesis



4.1. Stock Selection

Fundamental analysis defenders decide whether or not to invest in a stock by comparing the market price with the value determined by the fundamental analysis [17]. Fundamental analysis gives an idea about the company's future prospects and responds to some questions. For instance, is the revenue of the company is increasing? Or is it actually profit? Stock selection is important in terms of maximum yield. In fact, it will be used for some sort of portfolio management. The most important aim of portfolio theories is the minimum risk and the maximum yield. According to modern portfolio theories, it is ensured that the composition which has the most appropriate risk and return within the set of effective portfolios is increased to the maximum amount. For this purpose, the most successful estimation of the direction determination in long term is done by using the fundamental analysis. Stocks which have similar financial structures as selected. In another words, stocks are grouped according to fundamental analysis. This grouping is based on the value analysis of the companies, they are cheaper than the market price and the company is growing or shrinking. There may be 4 different groups as summarized in Table 5. Companies in the first group whose market price is lower than the price obtained as a result of fundamental analysis, are growing companies. The second group is that are still less than the price obtained and have a shrinking or horizontal growth course. The third group consists of growing companies and they are traded at a higher price than the price obtained. The fourth group companies are the ones which are traded at a higher price than the price obtained and are also shrinking and following a horizontal course of growth.

Table 5 Stocks Groups

Category		
1	Growing companies	Market price is less than the fundamental analysis price
2	Shrinking or horizontal growing companies	Market price is less than the fundamental analysis price
3	Growing companies	Market price is higher than the fundamental analysis price
4	Shrinking or horizontal growing companies	Market price is higher than the fundamental analysis price

The growth of the companies is decided by looking at the increasing and decreasing sales and profitability between the periods. Thanks to this grouping, it is assumed that investor profile and behavior that are interested in stock will be grouped. It is considered that the investor profile and behavior will be effect for the stock price in the short and medium term. Price fluctuations in stocks are shaped according to investor character and total sentiment. For this reason, it is known that investors with similar character are going to be oriented. For example, cheap and growing shares in price will be more popular among investors who have buying request.

Data is taken from KAP (Turkish Public Disclosure Platform - PDP) [29] under the "Financial Tables" section. The Public Disclosure Platform (KAP) is an electronic system in which the notifications required to be disclosed to the public in accordance with the capital market. Financial analysis table and balance sheet information can be accessed through KAP. In addition, KAP provides a detailed analysis of financial statements in a periodic and specific maturity range. Financial tables are helpful to investors in terms of analyzing financial situation of companies. These tables include some financial variables such as income statement, financial investment, investment property, short-term loans, long-term loans, balance sheet, cash flow statement, operating profit and cash equivalents. The data is in Excel format and are published regularly. Some financial items are selected from these Excel files for a three months periods.

Financial statement analysis is the process of reviewing and analyzing the financial tables of a company in order to make more investment decisions that will generate revenue in the future. There are many financial analysis techniques [4], [30]. These techniques analyze companies by using various indicators, measurements or making comparisons. Value analysis is one of the most reliable method. There are three pillars for value analysis; the first pillar is the rate of increase of operating profit between consecutive periods. The second pillar is based on the rate of increase of sales between consecutive periods. The third one is liquidation analysis in which company's ultimate value is calculated by using the most liquadable items which can be easily converted to cash. This information is published in the periodical financials of the company. The financial variables used in the calculation of liquidation are given in Table 6:

Table 6 Financial Variables

Operating Profit	Profit or loss after deducting operating expenses from gross sales profit or loss in the income statement. Income statement table is given in Table 2.
Cash and cash equivalents	Most liquid assets may include treasury bills, short-term deposits and fixed currencies. This variable is given in Table 1.
Accounts receivables	Receivables arising from commercial transactions of enterprises. This variable is given in Table 1.
Other Current Assets	It is a collection of stocks and bonds, receivables, inventories, cash and other revolving assets. This variable is given in Table 1
Stocks	All raw materials used, processed or semi-processed. This variable is given in Table 1.
Real Assets	They are assets that are not taken for sale, can be used for more than one period in the operating period and the benefits are distributed to more than one period. This variable is given in Table 1.
Financial investment	Stock and debt securities, a liquid market. This variable is given in Table 1.
Investment property	Investment properties are considered as held to earn rental income. This variable is given in Table 1.
Short-term loans	It must be paid within one year. This variable is given in Table 1.
Long-term loans	It is mostly non-payable loans of more than one year. This variable is given in Table 1.

Financial variables is taken from KAP. Liquidity calculation (all variables in Table 6 are used for calculation), operating profit, sales, are calculated and stocks in good condition are selected. After completing the fundamental analysis, the best 20 stocks have been chosen, which are also shown in Table 7 as an example.

Table 7 Selected Stocks

GOLTS	ALKIM	TAVHL	CIMSA
FMIZP	BOSSA	PNSUT	DOAS

ARMDA	BUCIM	HEKTS	IZOCM
AKFGY	KLMSN	AKCNS	NUGYO
YGGYO	BOLUC	EGPRO	TRGYO

4.2. Data Source

The order book is the main dataset used in this study. The order book is purchased as a csv file from Borsa Istanbul historical data sales for academic purpose. A single csv file contains one-month data. The entire data set covers the period from January 1, 2018 to May 31, 2018. Data file sizes are given in Table 8. The structure of the intra-day order book can be found in Table 21 in the Appendix section.

Table 8 Data Size

Month	**Size (Gigabytes)**
January	29,5625
February	26,80949
March	25,62054
April	27,16632
May	31,49471

Firstly, data is preprocessed. All csv files are merged. Big data solutions are applied because all csv files are merged. Apache Spark is used for processing big data in this thesis. Also using Hadoop, the data is stored in HDFS and processed with Apache Spark to make it easier and faster. Then files are parsed according to the Datastore File Format of BIST [18] for the selected stocks. This file contains bid/ask price and lot information, order change date, price type, category, validity type and session etc. Detailed information is given in Chapter 3. The modified date and time, ask/bid, order type, exchange order type, order category, order duration id, order status, reason for change, order quantity, price, session, joint contract id variables are used to extract bid/ask prices and volumes. If the following conditions are met in Table 9, modified date and time, bid/ask prices and volumes are added to new dataset (processed order book) which will be used for feature selection.

Table 9 Dataset Generation

Condition	Reason
“Order Duration ID” is not IMMEDIATE	Immediate order is not added to depth analysis, because this order is valid only for the moment it is entered
“Status” is active	Passive orders have no effect on depth analysis
“Change Reason” is not “Cancel” or “Trade”	Cancelled orders have no effect on depth analysis. Traded orders have already executed and will be considered by technical analysis.

4.3. Feature Selection

Technical indicators and derived data from the order book are used to generate feature vector. Depth (level) and technical analysis are used together in BIST. Financial indicators, which are commonly used in the past studies and observed to be successful in estimation modeling have been used to create the features used in this thesis.

4.3.1. Technical Indicators Selection

There are many empirical research studies in the literature about predicting the price index movement based on technical analysis. Technical indicators, which are commonly used in past studies [6, 12, 16] and observed to be successful in estimation modeling, have been used. Such indicators are listed below.

- **MACD (Moving Average Convergence Divergence):** This numeric attribute is calculated by subtracting the 26 periods Exponential Moving Average (EMA) from the 12 periods EMA. Each period is considered as 15 minutes, therefore long period is 15×26 minutes.
- **SMA (Simple Moving Average):** This numeric attribute is calculated by the 26 periods average price. Each period is considered as 15 minutes.
- **WMA (Weighted Moving Average):** This numeric attribute is calculated by giving more weight to the latest prices for the desired number of periods. Each period is considered as 15 minutes, therefore long period is 15×26 minutes.

- **RSI (Relative Strength Index):** This numeric attribute is calculated by comparing the closing values within the specified period of time for the instrument to be analyzed. Each period is considered as 15 minutes and calculation is made for 14 periods.
- **ROC (Rate of Change):** This numeric attribute is calculated the rate of price change between periods.
- **ATR (Average True Range):** This numeric attribute is obtained by calculating the average of the difference between the highest and the lowest in each item within the period, regardless of the trend direction, for the period followed. Each period is considered as 15 minutes and calculation is made for 14 periods.
- **Bollinger Bands:** Bollinger Bands are calculated drawn at the standard deviation level above and below the simple moving average of the price. Each period is considered as 15 minutes and calculation is made for 20 (default) periods.

4.3.2. Order Book Feature Selection

Order book or depth (level) analysis is the main concept of this study. Feature vector consists of best prices and volumes on bid and ask values for 10 different levels. Feature vector is visualized in Figure 5. The moment a buy order price and volume match a sell order, the trade is executed and “Change Reason” field is set to “Trade” in the Order Book. When “Change Reason” is “Trade”, stock price is determined according to matched bid/ask prices and added to vector. For example, in Table 10 which is a sample order book, order number 605AE28100CBD0A3 is trade, therefore new price is set to 3.22, and levels are created like at Table 11. Table 11 lists the best bid price and the best ask price first and the next best and continues sequentially.

Table 10 Sample Order Book (Only Price and Quantity)

Modified Date and Time	**Order No**	**Bid/Ask**	**Change Reason**	**Quantity**	**Price**
2018-01-10 14:56:40	605AE28100CA181D	A	New	500	3.26
2018-01-10 14:56:43	605AE28100CA1C16	A	New	1111	3.27

2018-01-10 14:56:45	605AE28100CA1F40	A	New	1111	3.24
2018-01-10 14:56:45	605AE28100CA1F67	A	New	250	3.25
2018-01-10 14:57:32	605AE28100CA74B3	B	New	100	3.20
2018-01-10 14:57:44	605AE28100CA8957	B	New	500	3.21
2018-01-10 14:57:48	605AE28100CA9244	B	New	10	3.23
2018-01-10 15:00:58	605AE28100BE797E	B	Trade	5000	3.22
2018-01-10 15:00:58	605AE28100CBD0A3	A	Trade	3	3.18

Table 11 Sample Order Book Levels

	Ask		**Bid**	
	Price	**Quantity**	**Price**	**Quantity**
Level 1	3.24	1111	3.20	10
Level 2	3.25	250	3.21	500
Level 3	3.26	500	3.23	100
….	…	….	…	…
Level 10	…	…	…	…

Also, some derived values from levels have been added to feature set. These derived values are calculated as follow:

- Average ask prices/volumes for n = 10 levels.

$$Feature\ 1 = \left\{ \frac{1}{n} \sum_{i=1}^{n} Pa(i) \right\} \tag{20}$$

$$Feature\ 2 = \left\{ \frac{1}{n} \sum_{i=1}^{n} Va(i) \right\} \tag{21}$$

- Average bid prices/volumes for n = 10 levels.

$$Feature\ 3 = \left\{ \frac{1}{n} \sum_{i=1}^{n} Pb(i) \right\} \tag{22}$$

$$Feature\ 4 = \{ \tfrac{1}{n} \textstyle\sum_{i=1}^{n} Vb(i)\ \} \tag{22}$$

- Average bid and ask prices/volumes for n = 10 and n = 5 levels.

$$Features\ 5, 6 = \left\{ \frac{\frac{1}{n} \sum_{i=1}^{n} Pa(i)\ +\ \frac{1}{n} \sum_{i=1}^{n} Pb(i)}{2} \right\} \tag{23}$$

$$Feature\ 7, 8 = \left\{ \frac{\frac{1}{n} \sum_{i=1}^{n} Va(i)\ +\ \frac{1}{n} \sum_{i=1}^{n} Vb(i)}{2} \right\} \tag{24}$$

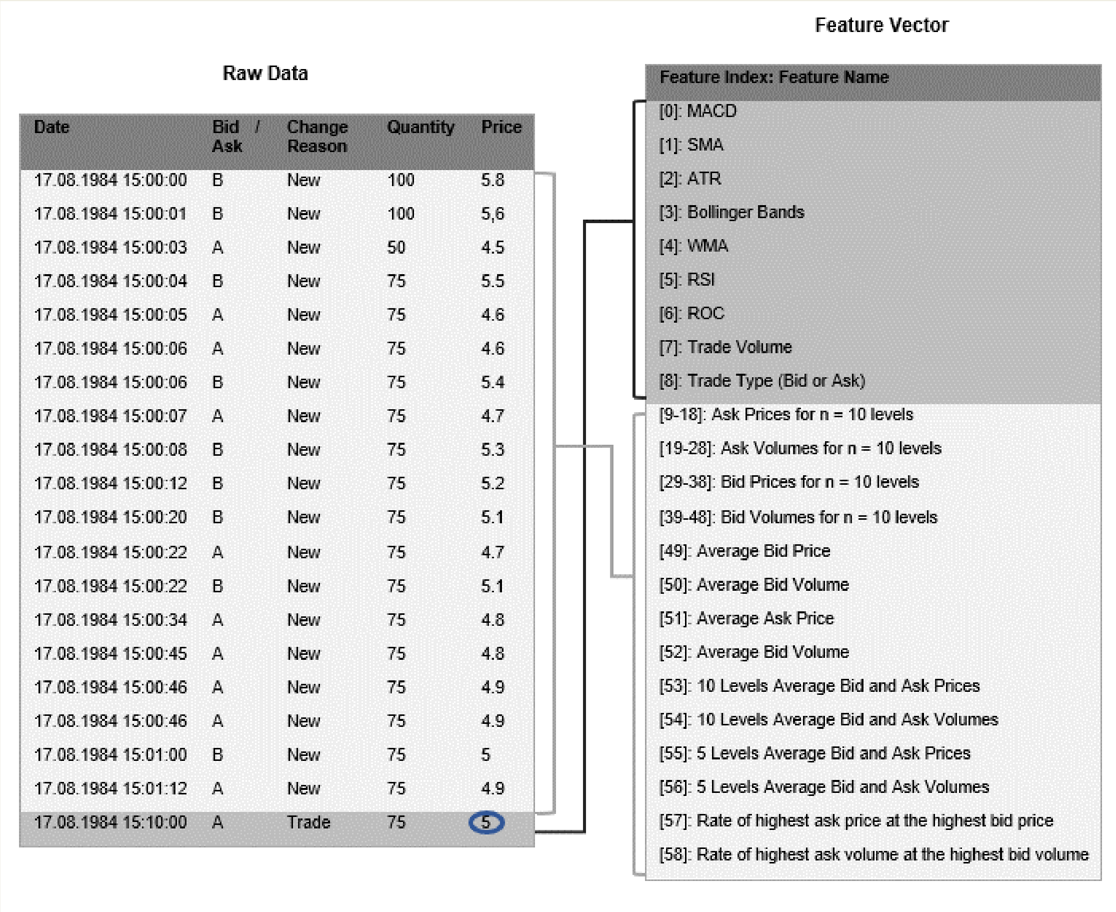
- Rate of the highest ask price/volume at the highest bid price/volume.

$$Feature\ 9 = \text{Maximum Bid Price/ Maximum Ask Price} \tag{25}$$

$$Feature\ 10 = \text{Maximum Bid Volume/ Maximum Ask Volume} \tag{26}$$

Ask/bid prices and volumes at level is denoted by Pa(i), Va(i), Pb(i), Vb(i) respectively.

Figure 5 Feature Vector



After that, among 59 features which is given in Table 12, feature selection method is used to determine the best features for each stock. For this purpose, Random Forest and its Gini importance of each feature were used as the feature selection method. selectFromModel object from scikit-learn is used to automatically select the features. According to this analysis, the technical features which are the mostly successful in predicting the direction of stock price change are MACD, SMA, WMA, RSI, ROC, ATR and BollingerBands. Selected features are given in Table 22 in the Appendix section for each stock. As an additional example, feature ranking of some stocks (TAVHL, ARMDA, and BOSSA) are given in Figure 9, Figure 10 and Figure 11. With feature selection, features that do not contain information are removed from the data sets, improved model performance and reduced the computational cost (and time) of training the model.

Table 12 Features

Feature Index	**Feature Name**	**Feature Index**	**Feature Name**
01	Ask/Bid (0 or 1)	31	bid_price_level_1

02	Volume	32	bid_volume_level_2
03	MACD	33	bid_price_level_2
04	SMA	34	bid_volume_level_3
05	WMA	35	bid_price_level_3
06	RSI	36	bid_volume_level_4
07	ROC	37	bid_price_level_4
08	ATR	38	bid_volume_level_5
09	BollingerBands	39	bid_price_level_5
10	ask_volume_level_1	40	bid_volume_level_6
11	ask_price_level_1	41	bid_price_level_6
12	ask_volume_level_2	42	bid_volume_level_7
13	ask_price_level_2	43	bid_price_level_7
14	ask_volume_level_3	44	bid_volume_level_8
15	ask_price_level_3	45	bid_price_level_8
16	ask_volume_level_4	46	bid_volume_level_9
17	ask_price_level_4	47	bid_price_level_9
18	ask_volume_level_5	48	bid_volume_level_10
19	ask_price_level_5	49	bid_price_level_10
20	ask_volume_level_6	50	average_volume_for_10_levels
21	ask_price_level_6	51	average_ask_volume_for_10_levels
22	ask_volume_level_7	52	average_bid_volume_for_10_levels
23	ask_price_level_7	53	average_volume_for_5_levels
24	ask_volume_level_8	54	max_volume_rate_10_levels(ask/bid)
25	ask_price_level_8	55	average_price_for_10_levels
26	ask_volume_level_9	56	average_ask_price_for_10_levels
27	ask_price_level_9	57	average_bid_price_for_10_levels
28	ask_volume_level_10	58	average_price_for_5_levels
29	ask_price_level_10	59	max_price_rate_10_levels(ask/bid)

30	bid_volume_level_1

After that, to predict the direction of price change, 15 minutes ($\Delta t = 15$) price vector is obtained from processed order book. Class label is set to price index movement according to the next price value in order to predict the future price direction of a stock within 15 minutes. Direction of the price is labeled as 0 for uptrend, 1 for down trend and 2 for no change. “0” means that the price of the stock will increase. “1” means that the stock price will decrease and the “2“ means that the price will not change. Class label is calculated as below:

$$cl(t) = \begin{cases} 0, & if\ p(t) < p(t + \Delta t) \\ 1, & if\ p(t) > p(t + \Delta t) \\ 2, & otherwise \end{cases} \tag{27}$$

0: Up Trend
1: Down Trend
2: No Change

cl(t) is price index movement label for time t. p(t) and p (t + Δt) denote price for a stock in time t and next price value in t + Δt

4.4. Design of the Proposed Method

4.4.1. Machine Learning Algorithms

After data pre-processing (historical data processing) and analysis, some machine learning methods are experimented. Two algorithms are used: Random Forest and Neural Network algorithms. For each stock, these algorithms are used to train the models and obtained results are compared each other. Python scikit learn library is used for machine learning methods. Some important parameters are given as follow.

Table 13 Algorithm Parameters

Algorithm	Parameter	Tested Value
Random Forest	The number of trees in the forest	64, 100,128
	Function to measure the quality of division	gini

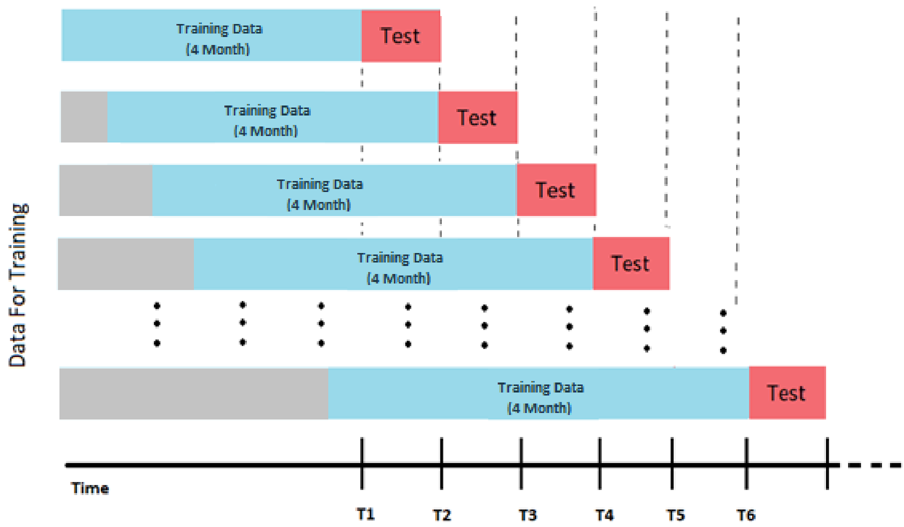
	The number of features to consider when looking for the best split	sqrt(n_features),n_features
Neural Network	Solver for weight optimization	'sgd' : stochastic gradient descent
	Hidden Layer Sizes	Karsoliya, Saurabh [20] $(2/3) * ($ Size of the input layer + size of the output layer)
	Size of minibatches	500
	Activation function for the hidden layer.	Logistic sigmoid function
	Learning rate	0.1
	Momentum	0.9

Firstly, 15 minutes features vector is trained with Neural Network algorithm. Results are generated only for one hyperparameter set (learning rate, momentum, hidden layer size etc.), but a wide range of hyperparameter values were tested in a small stock to determine neural network hyperparameters previously. Then, the same input is used for the Random Forest algorithm. The number of trees in the forest is selected as 64, 100 and 128. Oshiro et al. [22], suggest a range between 64 and 128 trees in the forest. Probst et al.[26] have achieved the biggest performance gains from the top 100 trees. Increasing the number of trees also extends the computational cost.

4.4.2. Training and Testing Dataset

In all experiments, the models are individually trained and tested on 20 stocks. The dataset is used with an 80-20 split. In the literature, there are many techniques to evaluate the performance of a machine learning model [19]. Cross validation which is a statistical method is the most used technique to overcome underfitting and overfitting. In this thesis, time series cross validation is used because stocks data does not randomly split. Time series cross validation is a type of k-fold cross validation. It is said that "If the forecaster withholds all data about events occurring after the end of the fit period, the forecast-accuracy evaluation is structurally identical to the real-world-forecasting environment, in which we stand in the present and forecast the future." [21]. For this reason, instead of using k-fold cross-validation, time series cross validation is more accurate to validate the model’s performance. This method can be imagined as sliding window. At each time testing and training data sets are rolling the forward. Visualization of the model is given Figure 6.

Figure 6 Time series cross validation



In this model, data belongs to each 15 minutes is used as the training test data. Price movement direction is predicted throughout the 15 minutes. Then test procedure is carried one step (one minute) further, the next one-minute data is taken as test data and previous test data is added to training data. Training data is rolling the forward to handle the last 120 days (4 months) dataset.

4.4.3. Evaluation Metrics

After the test results are obtained, some performance metrics are used to evaluate the classification results. Performance is measured by calculating the following metrics:

- Average accuracy: Ratio of accurate prediction to all prediction in the system.
- Recall: Indicates how successful positive states are predicted.
- Precision: A condition that indicates success in a positively predicted state.
- Macro-average F1-Score: Harmonic mean of negative and positive classes.

Although the accuracy measure is one of the most frequently used evaluation metrics, this method produces misleading results on data sets that has unbalanced distribution. If uptrend class has a high density and high accuracy, total accuracy results will be high. For example, uptrend sample size and matched number are 100 and 90 respectively, on the other hand, down trend sample size and number of matches are 10 and 1 respectively. Total accuracy value is calculated as 0.83, while the down trend accuracy value is 0.1. Therefore, this measure does not seem suitable for this thesis. So, it is more accurate than average accuracy. Evaluation and comparison are made according to Macro-average F1-Score. Average accuracy is supplied for information only. Formulas of the performance metrics are defined as follow:

$$Recall = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP)\ +\ False\ Negative\ (FN)} \tag{28}$$

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \tag{29}$$

$$F1\ Score = 2 * \frac{(Precision * Recall)}{(Precision + Recall)} \tag{30}$$

$$Average\ Accuracy = \frac{Number\ of\ Matches}{Number\ of\ Samples} \tag{31}$$

After the evaluations, if true positive and true negative rates are more than false positive and false negative areas, the model is successful.

5. EXPERIMENT RESULTS AND ANALYSIS

Experimental results are summarized in this section. All experiments are conducted on a desktop PC with a 2.20GHz Intel Core i7-8750H CPU, 16GB 1600MHz memory and 256 GB SSD running a Windows OS. The data set contains data of 20 stocks (companies) according to the results of fundamental analysis. Because of the computational cost on the parameter optimization of the training for each stock, relatively small companies are selected as experiment data. Scenario is tested as mentioned in Section 4.4.2, and results are obtained for two machine learning techniques. Table 14 shows the F1 score results. This table contains the result of a portion of the stocks for a better understanding of the results visually. All test results are available from Table 23 in appendices section.

Table 14 Experiment Results (Stock Price Index Movement)

Stock	**MLP**	**Random Forest - 64 Tree**	**Random Forest - 100 Tree**	**Random Forest - 128 Tree**
AKCNS	0.43	0.58	0.58	0.59
AKFGY	0.47	0.60	0.60	0.60

ALKIM	0.45	0.65	0.66	0.65
ARMDA	0.45	0.65	0.66	0.66
BOLUC	**0.50**	0.63	0.63	0.64
BOSSA	0.34	0.68	0.69	0.68
BUCIM	0.49	0.61	0.62	0.62
CIMSA	0.43	0.63	0.64	0.64
DOAS	0.44	0.68	0.68	0.69
EGPRO	0.42	0.64	0.65	0.65
FMIZP	0.36	0.62	0.62	0.62
GOLTS	0.43	0.65	0.65	0. 65
HEKTS	0.41	0.67	0.68	0.68
IZOCM	0.39	0.61	0.60	0.60
KLMSN	0.39	0.68	0.68	0.68
NUGYO	0.40	0.70	0.69	0.69
PNSUT	0.42	0.67	0.67	0.67
TAVHL	0.47	**0.73**	**0.73**	**0.73**
TRGYO	0.49	0.71	0.71	0.71
YGGYO	0.48	0.59	0.60	0.59
Overall	0.44	0.65	0.65	0.65

Average F1-scores are 0.65 and 0.44 for random forest and multi-layer perceptron respectively. When multi-layer perceptron is compared to random forest, the average prediction F1-scores was improved from 0.44 to 0.65. The experimental results showed that random forest outperformed neural network, because random forest is dealing with discrete/binary/continuous attributes and a nonlinear ensemble learning binary classifier and the feature is much more nonlinear. These results are more clearly in Figure 7 and Figure 8. Furthermore, a detailed view of F1 scores presented in the table tells about improvement potential in the performance. The F1 score is an average of F1 scores for uptrend (UP-F1), downtrend (DW-F1) and no-change (NC-F1) predictions. According to the results, F1 score no-change case is usually the lowest and degrading the average. This situation can be seen on Figure 8(b) more easily. This still might be due to distributions in the data set even though F1 scores are used. Therefore, if prediction rate of the no-change case was at least as good as average of the uptrend and downtrend cases, the improvement would be as high as 0.09, which yields nearly 0.75 overall performance. Experimental results show that, there is no serious difference among 64,100 and 128 trees in the forest.

Table 15 shows the prediction scores of past studies using the order book in different stock markets. Studies in Table 15 uses only depth analysis with order book while our work uses both technical and depth analyses together. According to the table, Kercheval et al. and Han et al. seem to have better results than ours at a first glance. The difference is due to some fundamental dissimilarities. First, they predicted the direction of the mid-price movement, which is the mean of best ask and bid prices. Our work, however, predicts price change directions. Since there are many parameters that affect the traded price, it is more difficult to estimate the stock price. Second, they used k-fold cross validation method in their studies. In this case, the results may be better because training dataset includes future situations. Third, especially Han et al. worked on relatively small amount of data as compared to our data set. Ntakaris et al. follows the k-fold cross validation method as well. Besides, it presents too many results regarding some selected stock. Their maximum prediction rate 65% percent for only few of the stocks. Actually, their average score is less than 65%. Thus, worse than our results. Last but not least, as a general remark, stock market of each country exhibits different behaviors. That is, depth analysis is a subjective analysis and will differ for each market and even for every stock.

Table 15 Previous Results

Study	F1 Score	Accuracy	RMSE	Technique
[24]	46%	-	-	Ridge regression (RR) and SLFN network-based nonlinear regression
[28]	57.94%	-		Convolutional Neural Networks
[9]	80.3%	-	-	Support Vector Machines, they used n-fold cross validation

[25]	~85%	-	-	Random Forest
[38]	-	-	0.40	Random Forest
[39]	-	55%	-	Multiple Kernel Learning
[45]	65%	-	-	Multilayer Perceptron
[27]	41%	-	-	Recurrent Neural Network
In this thesis	65%	70%	0.04	Random Forest

Figure 7 Performance Evaluation for MLP and RF

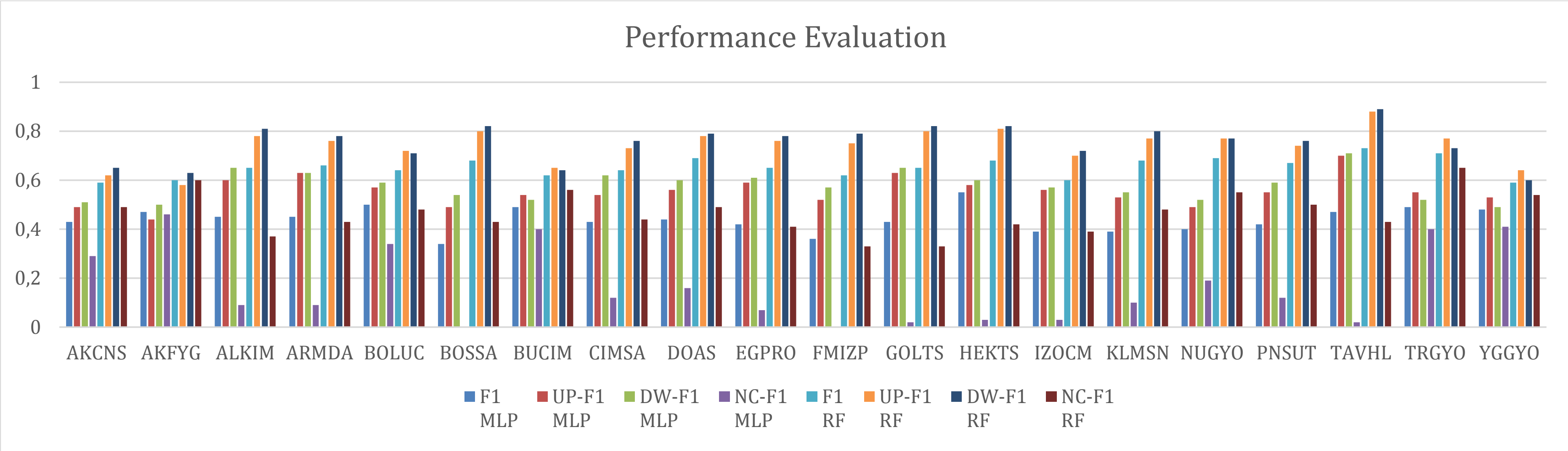
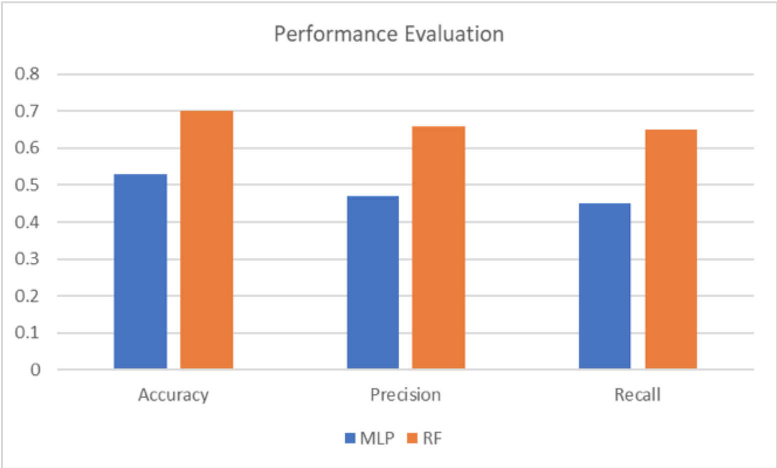
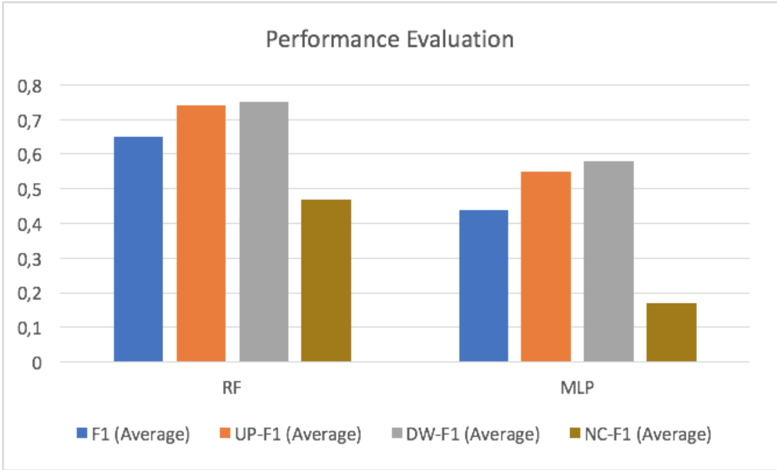


Figure 8 MLP vs. RF: Performance comparison in terms of (a) accuracy, precision and recall measures, and (b) F1 scores



(a)



(b)

5.1. Cumulative Profits

In addition to the above scenario, cumulative profits are calculated. In this scenario, the first estimate for increase in price is taken as a buy order, and similarly the first estimate for decrease is taken as a sell order. In other words, we will buy when we see the first estimate for increase, and we keep the position until the first estimate for decrease comes only then we sell and wait the next estimate for increase. At the end of the period, this period can be a month as an example, we also calculate the cumulative gain or loss as a total sum. Trading commission is accepted as 7/10000 for each transaction. The following table (Table 16) shows the cumulative gains/losses for each distinct stock or company. The percentages shown in the table are the cumulative or total gains for a month. The gains are calculated as this: as the first estimate for increase comes the stock is bought, then the position is kept until the next estimate for decrease comes, with this new estimate then the stock is sold. When it buys it uses whole capital and similarly when it is time to sell it again uses whole capital. This cycle repeats itself throughout the month provided that the market is open. This procedure is calculated for each stock which is given in the Table 16 between 01/05/2018 - 31/05/2018. The average of all the gains is about 59% and it is quite high comparing with the gain of BİST100.

Table 16 Cumulative Gain/Loss

Stock	Cumulative Return	Stock	Cumulative Return
ARMDA	146%	KLMSN	58%
TRGYO	112%	EGPRO	58%
NUGYO	88%	DOAS	55%
BOSSA	87%	BOLUC	48%
BUCIM	87%	AKFGY	45%
FMZIP	77%	YGGYO	34%
TAVHL	67%	PNSUT	18%
IZOCM	64%	AKCNS	11%
ALKIM	61%	CIMSA	9%
HEKTS	58%	GOLTS	3%

5.1. Stock Price Prediction Test

As an additional experiment it has been tried to predict the stock price. Results are given in Table 17. In this experiment, all inputs are the same, the only difference is output. Some statistical measures are used to evaluate the regression results.

In order to determine statistical error rates, MAE, MSE and RMSE are calculated for each stock.

- **Mean Absolute Error (MAE) :** This metric is used to determine the absolute error between actual values and prediction values. The closer the MAE value to zero, the better the ability of the model to predict. The MAE is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \tag{32}$$

x_i : Actual Output

y_i : Model's prediction

- **Mean Squared Error (MSE) :** This metric is used to determine the squared error between actual values and prediction values. The closer the prediction values are to the actual values, the smaller the MSE; the farther away from the actual values the larger the MSE.

$$RMSE = \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n} \tag{33}$$

x_i : Actual Output

y_i : Model's prediction

- **Root Mean Squared Error (RMSE) :** This metric is used to determine the error rate between actual values and prediction values, and when the RMSE value approaches zero, this means that the model's ability to estimate is increased. RMSE is calculated as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(x_i - y_i)^2}{n}} \tag{34}$$

x_i : Actual Output

y_i :   Model's prediction

Table 17 Experiment Results (Stock Price)

Stock	MAE	MSE	RMSE
AKCNS	0.009	0.00032	0.018
AKFGY	0.004	0.00005	0.007
ALKIM	0.020	0.00196	0.044
ARMDA	0.012	0.00065	0.025
BOLUC	0.008	0.00031	0.018
BOSSA	0.009	0.00044	0.021
BUCIM	0.007	0.00017	0.013
CIMSA	0.008	0.00022	0.015
DOAS	0.005	0.00012	0.011
EGPRO	0.011	0.00062	0.025
FMIZP	0.019	0.00153	0.039
GOLTS	0.025	0.00294	0.054
HEKTS	0.009	0.00050	0.022
IZOCM	0.034	0.00591	0.077
KLMSN	0.007	0.00022	0.015
NUGYO	0.023	0.02528	0.159
PNSUT	0.010	0.00082	0.029
TAVHL	0.013	0.00119	0.035
TRGYO	0.003	0.00005	0.007
YGGYO	0.052	0.00870	0.093

5.1. Statistical Significance Tests

Statistical Significance tests are applied to determine whether these results showed statistically significant differences. Firstly, Chi-Square Goodness of Fit Test is applied.

In the goodness of fit test one should look at whether the observed frequencies are suitable to expected frequencies. Following result is obtained. By conventional criteria, this difference is considered to be “extremely statistically significant”.

Table 18 Chi-Square Goodness of Fit Test

Chi Squared Value	P Value
189.085	P value is less than 0.0001

Secondly, chi-square test is applied in order to identify statistical significance of “order book” and “technical indicators” for each stock. Following results are obtained. From the analysis below it is evident that both order book and technical analysis are highly significant to predict stock price index movement.

Table 19 Chi-Square order book and technical analysis

Stock	**Chi Squared Value**	**P Value**	**Stock**	**Chi Squared Value**	**P Value**
AKCNS	6.737	0.0344	FMIZP	34.904	0.0001
AKFGY	10.696	0.0048	GOLTS	17.001	0.0002
ALKIM	16.113	0.0003	HEKTS	34.591	0.0001
ARMDA	21.775	0.0001	IZOCM	43.098	0.0001
BOLUC	4.415	0.1100	KLMSN	8.353	0.0154
BOSSA	35.588	0.0001	NUGYO	6.099	0.0474
BUCIM	21.258	0.0001	PNSUT	35.727	0.0001
CIMSA	14.596	0.0007	TAVHL	10.482	0.0053
DOAS	4.920	0.0854	TRGYO	22.899	0.0001
EGPRO	15.407	0.0005	YGGYO	6.548	0.0379

Additional set of experiment were conducted to see if using both technical analysis and order book features together have any advantageous over using only one of them. To do that, technical analysis and order book features were used to predict stock price index movement separately. These results were compared with the results of hybrid features.

According the results in Table 20, technical indicators and order book features are better than each other for more or less half of the cases. There is no superior among them. However, using hybrid features provides better results in 95% of the cases. There is actually just one case (i.e. AKFGY) where the hybrid features perform as good as (still not worse than) order book features only. Overall performance of technical analysis and order book features are 0.6 and 0.61, respectively. That is, using hybrid features provides 4% improvement.

Table 20 Technical Analysis vs. Order Book

Stock	**Technical Analysis F1 Score**	**Order Book F1 Score**	**Hybrid F1 Score**
AKCNS	0.58	0.55	0.59
AKFGY	0.60	0.57	0.60
ALKIM	0.64	0.62	0.65
ARMDA	0.62	0.60	0.66
BOLUC	0.62	0.58	0.64
BOSSA	0.62	0.65	0.68
BUCIM	0.58	0.58	0.62
CIMSA	0.61	0.60	0.64
DOAS	0.62	0.63	0.69
EGPRO	0.58	0.60	0.65
FMIZP	0.58	0.60	0.62
GOLTS	0.62	0.60	0.65
HEKTS	0.62	0.64	0.68
IZOCM	0.55	0.57	0.60
KLMSN	0.66	0.64	0.68
NUGYO	0.64	0.66	0.69
PNSUT	0.63	0.64	0.67
TAVHL	0.69	0.66	0.73

TRGYO	0.65	0.66	0.71
YGGYO	0.57	0.53	0.59
Average	**0.61**	**0.60**	**0.65**

6. CONCLUSION AND FURTHER SUGGESTIONS

The main objective of investors traded on the stock market is to make a profit by investing in high yielding securities. Therefore, it is very important to provide the right strategy to predict the direction in which stock prices have changed. However, it is difficult to predict the stock market because it is nonlinear, dynamic and complex. This thesis proposes a hybrid method that integrates depth analysis, fundamental and technical analysis with random forest and multilayer perceptron neural network model for predicting the moving direction of stock prices. In this thesis, 15 minutes movements of 20 stocks in Borsa Istanbul were predicted using different technical indicators and order book data. Data set includes Borsa Istanbul order book data from January 2018 to May 2018.

Firstly, stocks were determined by fundamental analysis. The fundamental analysis method is preferred by long-term investors. They perform an analysis of the general economy, then sector analysis and finally the firm analysis. The investment decision is made after these detailed analysis stages are completed. The purpose of this method is to find the real value of the stock and to compare this value with the market value. If the calculated real value is greater than the market value, the decision is made to invest. If the calculated real value is lower than the market value, it is decided not to invest for the analyzed stock. Fundamental analysis helps to select stocks with a high number of demands, this provides the selection of deeper stocks. Data needed to perform fundamental analysis are taken from KAP (Turkish Public Disclosure Platform - PDP). Such as financial analysis table and balance sheet information.

After the fundamental analysis, technical indicators and order book feature selection were conducted. The task focused in this thesis is to perform depth and technical analysis to predict direction of movement stock price. Technical analysis is a method preferred by short-term investors and aiming to gain profit from short-term price movements. This method assumes that the historical price of the stock will be repeated in the future. Depth analysis helps investors determine where the price of a stock is in the near future. Therefore, the combination of these two methods provided more accurate results.

Features variables (technical indicators) were determined as a result of literature review. Prediction stock price moving direction was categorized as either ‘uptrend’, ‘downtrend’ or “no change”. Random Forest and multilayer perceptron were applied, and experimental result were obtained on 4 months training and 1-month test set. As a result of the tests performed, it was found that the Random Forest has better. F1 score improved from 0.39 to 0.51 in first experiment. In the second experiment F1 score improved to 0.72. The macro-average F1-score was used for the TAVHL stock that yielded the best results with the Random Forest. Order book data is nonlinear therefore Random Forest was more successful than other method. High accuracy predictions were obtained by using technical and depth analysis together.

There are some limitations of the study. Firstly, 7 technical indicators as input variables were determined as a result of literature review. However, the number of technical indicators in the financial literature is quite high. Likewise, optimization of parameters could not be performed for all shares. Due to the computational cost on the parameter optimization, only the less traded stocks were selected, and the optimization method was applied. The effect of the use of other inputs on forecast performance can be examined in future studies.

In the light of this information, portfolio management can be done in future studies. In other words, control the distribution of the stocks according to the market can also be controlled by artificial intelligence.

REFERENCES

[1] Abu-Mostafa, Yaser S., and Amir F. Atiya. "Introduction to financial forecasting." Applied Intelligence 6.3 (**1996**): 205-213.

[2] Abarbanell, Jeffrey S., and Brian J. Bushee. "Fundamental analysis, future earnings, and stock prices." Journal of accounting research 35.1 (**1997**): 1-24.

[3] Biais, Bruno, Pierre Hillion, and Chester Spatt. "An empirical analysis of the limit order book and the order flow in the Paris Bourse." the Journal of Finance 50.5 (**1995**): 1655-1689.

[4] Wafi, Ahmed S., Hassan Hassan, and Adel Mabrouk. "Fundamental analysis models in financial markets–Review study." Procedia economics and finance 30 (**2015**): 939-947.

[5] Mittal, Anshul, and Arpit Goel. "Stock prediction using twitter sentiment analysis." Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf) 15 (**2012**).

[6] Patel, Jigar, et al. "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques." Expert Systems with Applications 42.1 (**2015**): 259-268.

[7] Dash, Rajashree, and Pradipta Kishore Dash. "A hybrid stock trading framework integrating technical analysis with machine learning techniques." The Journal of Finance and Data Science 2.1 (**2016**): 42-57.

[8] Brewer, Jacob Bruce. "A machine learning approach to intra-market price impact modeling using nasdaq level-2 itch data. " Diss. Brigham Young University, (**2016)**.

[9] Kercheval, Alec N., and Yuan Zhang. "Modelling high-frequency limit order book dynamics with support vector machines." Quantitative Finance 15.8 (**2015**): 1315-1329.

[10] Montgomery, John D. "Spoofing, market manipulation, and the limit-order book." Available at SSRN 2780579 (**2016**).

[11] Mizuno, Hirotaka, et al. "Application of neural network to technical analysis of stock market prediction." Studies in Informatic and control 7.3 (**1998**): 111-120.

[12] Kara, Yakup, Melek Acar Boyacioglu, and Ömer Kaan Baykan. "Predicting direction of stock price index movement using artificial neural networks and support vector

machines: The sample of the Istanbul Stock Exchange." Expert systems with Applications 38.5 (**2011**): 5311-5319.

[13] Atsalakis, George S., and Kimon P. Valavanis. "Surveying stock market forecasting techniques–Part II: Soft computing methods." Expert Systems with Applications 36.3 (**2009**): 5932-5941.

[14] Guresen, Erkam, Gulgun Kayakutlu, and Tugrul U. Daim. "Using artificial neural network models in stock market index prediction." Expert Systems with Applications 38.8 (**2011**): 10389-10397.

[15] Güreşen, Erkam, and Gülgün Kayakutlu. "Forecasting stock exchange movements using artificial neural network models and hybrid models." International Conference on Intelligent Information Processing. Springer, Boston, MA (**2008**).

[16] Gunduz, Hakan, Yusuf Yaslan, and Zehra Cataltepe. "Intraday prediction of Borsa Istanbul using convolutional neural networks and feature correlations." Knowledge-Based Systems 137 (**2017**): 138-148.

[17] Introduction To Fundamental Analysis https://www.investopedia.com/university/fundamentalanalysis/

[18] https://datastore.borsaistanbul.com/assets/files/DataStore_Veri_Bildirim_ve_Kabul_Formatlar%C4%B1.pdf (Jan 2, **2019**)

[19] James, Gareth, et al. An introduction to statistical learning. Vol. 112. New York: springer (**2013**).

[20] Karsoliya, Saurabh. "Approximating number of hidden layer neurons in multiple hidden layer BPNN architecture." International Journal of Engineering Trends and Technology 3.6 (**2012**): 714-717.

[21] Tashman, Leonard J. "Out-of-sample tests of forecasting accuracy: an analysis and review." International journal of forecasting 16.4 (**2000**): 437-450.

[22] Oshiro, Thais Mayumi, Pedro Santoro Perez, and José Augusto Baranauskas. "How many trees in a random forest?." International workshop on machine learning and data mining in pattern recognition. Springer, Berlin, Heidelberg (**2012**).

[23] Kanagal, Kapil, Yu Wu, and Kevin Chen. "Market Making with Machine Learning Methods." Available online here https://web. stanford. edu/class/msande448/2017/Final/Reports/gr4. pdf (**2017**).

[24] Ntakaris, Adamantios, et al. "Benchmark dataset for mid- price forecasting of limit order book data with machine learning methods." Journal of Forecasting 37.8 (**2018**): 852-866.

[25] Han, James, et al. Machine learning techniques for price change forecast using the limit order book data. Working Paper (**2015**).

[26] Probst, Philipp, and Anne-Laure Boulesteix. "To Tune or Not to Tune the Number of Trees in Random Forest." Journal of Machine Learning Research 18 (**2017**): 181-1.

[27] Dixon, Matthew. "Sequence classification of the limit order book using recurrent neural networks." Journal of computational science 24 (**2018**): 277-286.

[28] Tsantekidis, Avraam, et al. "Forecasting stock prices from the limit order book using convolutional neural networks." 2017 IEEE 19th Conference on Business Informatics (CBI). Vol. 1. IEEE (**2017**).

[29] https://www.kap.org.tr/tr/ (Jan 2, **2019**)

[30] Bernard, Victor Lewis. "Accounting-based valuation methods, determinants of market-to-book ratios, and implications for financial statement analysis." (**1994**).

[31] http://www.investopedia.com (Jan 2, **2019**)

[32] Fausett, Laurene. Fundamentals of neural networks: architectures, algorithms, and applications. Prentice-Hall, Inc. (**1994**).

[33] Breiman, Leo. "Random forests." Machine learning 45.1 (**2001**): 5-32.

[34] de Oliveira, Fagner A., Cristiane N. Nobre, and Luis E. Zarate. "Applying Artificial Neural Networks to prediction of stock price and improvement of the directional prediction index–Case study of PETR4, Petrobras, Brazil." Expert Systems with Applications 40.18 (**2013**): 7596-7606.

[35] Ergür, Bircan. Borsa İstanbul (BIST) Hisse Fiyat Değişim Yönünün ilişkisel Borsa Ağı Kullanılarak tahmin Edilmesi. Diss. Fen Bilimleri Enstitüsü (**2014**).

[36] Göçken, Mustafa, et al. "Integrating metaheuristics and artificial neural networks for improved stock price prediction." Expert Systems with Applications 44 (**2016**): 320-331.

[37] Khaidem, Luckyson, Snehanshu Saha, and Sudeepa Roy Dey. "Predicting the direction of stock market prices using random forest." arXiv preprint arXiv:1605.00003 (**2016**).

[38] Booth, Ash, Enrico Gerding, and Frank McGroarty. "Predicting equity market price impact with performance weighted ensembles of random forests." 2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr). IEEE (**2014)**.

[39] Fletcher, Tristan, Zakria Hussain, and John Shawe-Taylor. "Multiple kernel learning on the limit order book." Proceedings of the First Workshop on Applications of Pattern Analysis. (**2010**).

[40] Dixon, Matthew. "Sequence classification of the limit order book using recurrent neural networks." Journal of computational science 24 (**2018**): 277-286.

[41] Rodriguez, Pedro N., and Arnulfo Rodriguez. "Predicting stock market indices movements." WIT Transactions on Modelling and Simulation 38 (**2004**).

[42] Kumar, Manish, and M. Thenmozhi. "Forecasting stock index movement: A comparison of support vector machines and random forest." Indian institute of capital markets 9th capital markets conference paper. (**2006**).

[43] Basak, Suryoday, et al. "Predicting the direction of stock market prices using tree-based classifiers." The North American Journal of Economics and Finance 47 (**2019**): 552-567.

[44] Murphy, John J. "Technical Analysis of Financial Markets (New York Institute of Finance, New York, NY)." (**1999**).

[45] Ntakaris, Adamantios, et al. "Feature Engineering for Mid-Price Prediction With Deep Learning." IEEE Access 7 (2019): 82390-82412.

APPENDICES

EK 1 – Order Book

Table 21 Structure Of the Order Book

[0]Date	[1]Order No	[2]Entry Date and Time	[3]Modified Date and Time	[4]Operation code	[5]Bid/Ask	[6]Price Type	[7]Type	[8]Category	[9]Order Duration ID	[10]Status	[11] Change Reason	[12]Quantity
2018-01-02	4FA2820023377B	2018-01-02 09:40:04	2018-01-02 09:40:04	KSTUR.E	A	1	0	1	DAY	1	6	350
1.08.2018	61785B01002548D4	1.08.2018 09:40	1.08.2018 09:40	ADNAC.E	A	1	0	1	DAY	1	6	50000
2018-07-02	614E2A810023E02E	2018-07-02 09:43:44	2018-07-02 09:55:01	HUBVC.E	S	2	0	1	IMMEDIATE	2	3	10
2018-07-02	614E2A810027DE0A	2018-07-02 10:00:23	2018-07-02 10:00:23	HUBVC.E	S	1	0	1	DAY	2	6	157

[13]Remaining Quantity	[14]Visible Quantity	[15]Price	[16]Agency / Fund Code Flag	[17]Session	[18]Best Purchase Price	[19]Best Selling Price	[20]Previous Order Number	[21]Joint Contract No	[22]Number of Transactions	[23]Give up Flag	[24]Update Number	[25]Update Time
350	350	8.22	H	P_ACILIS_EMIR_TPL	8.22	-	-	-	-	H	2	2018-01-02 06:41:01
50000	50000	1.04	H	P_ACILIS_EMIR_TPL	1.04	0	-	-	-	H	2	1.08.2018 06:40

0	10	0	H	P_ESLESTIRME	1.27	1.31	-	701158844205819704	10	H	3	2018-07-02 06:56:33
157	157	1.56	H	P_SUREKLI_ISLEM	1.27	1.31	-	-	-	H	1	2018-07-02 07:01:56

EK 2 – Feature Importances

Figure 9 TAVHL

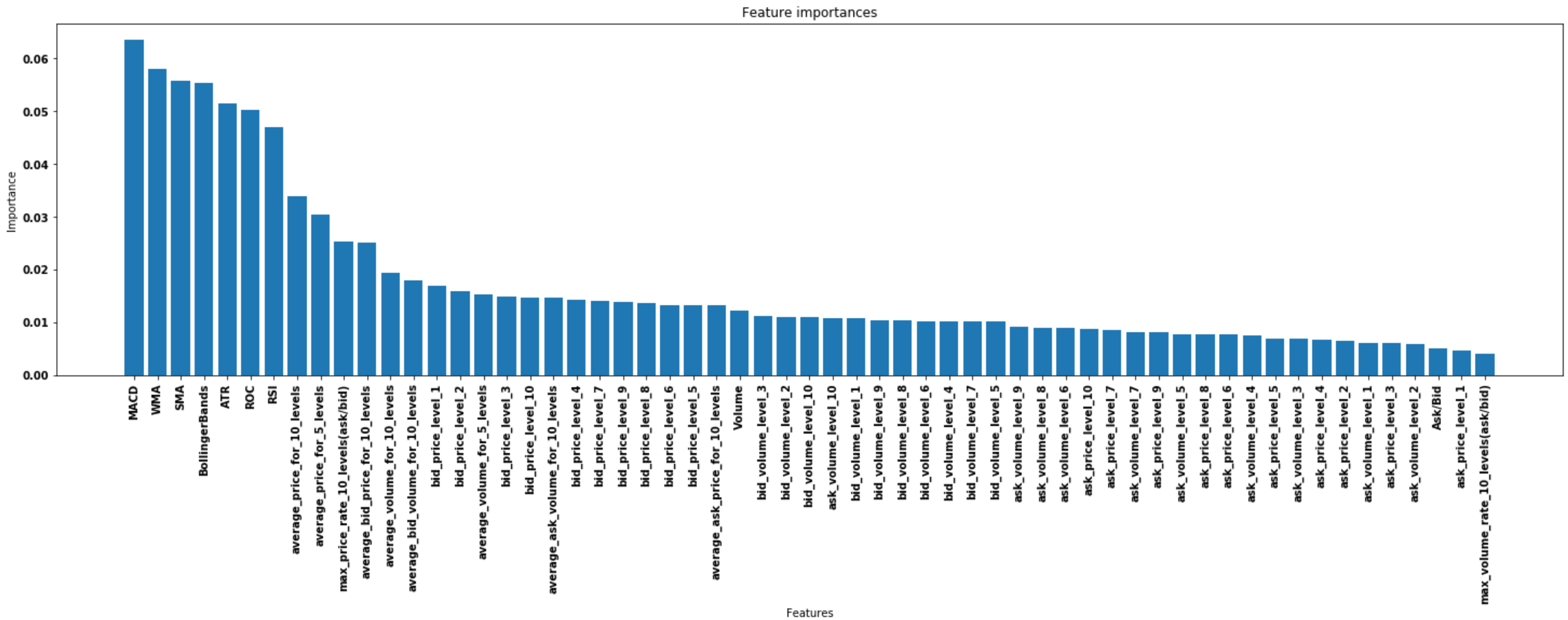


Figure 10 ARMDA

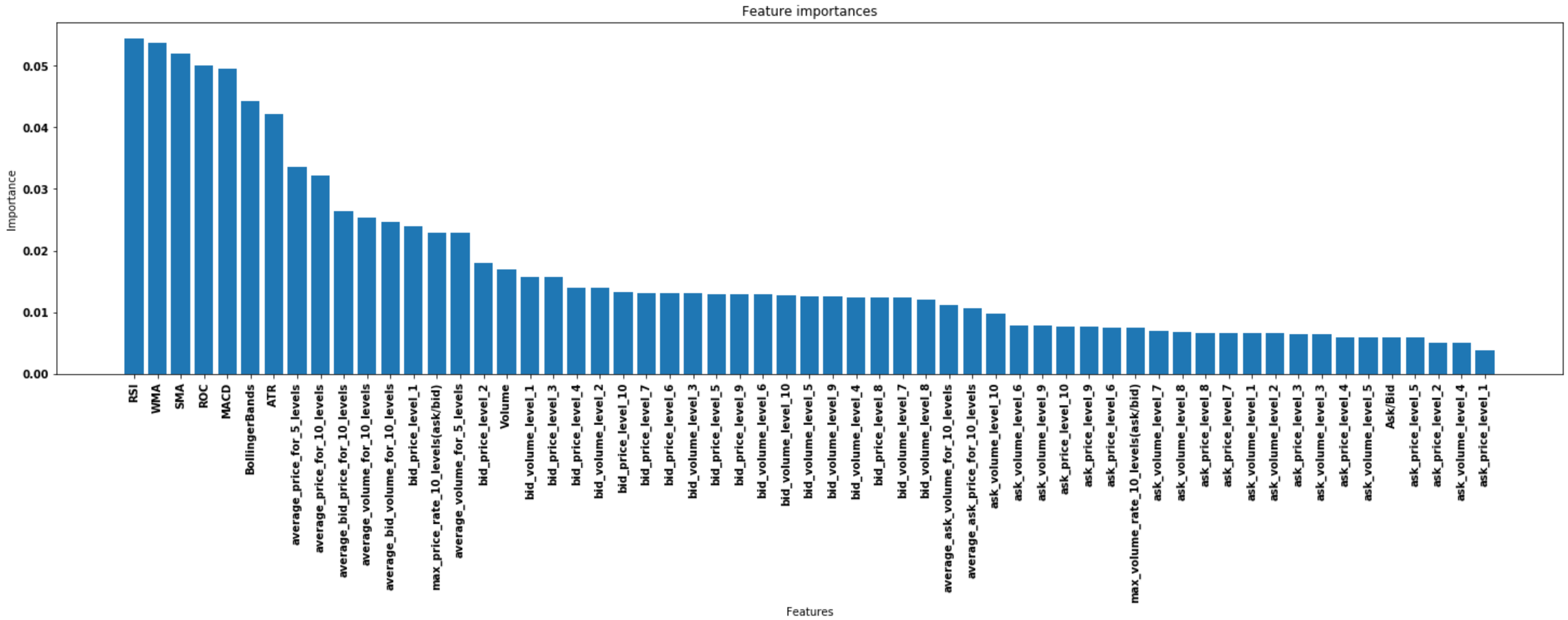


Figure 11 BOSSA

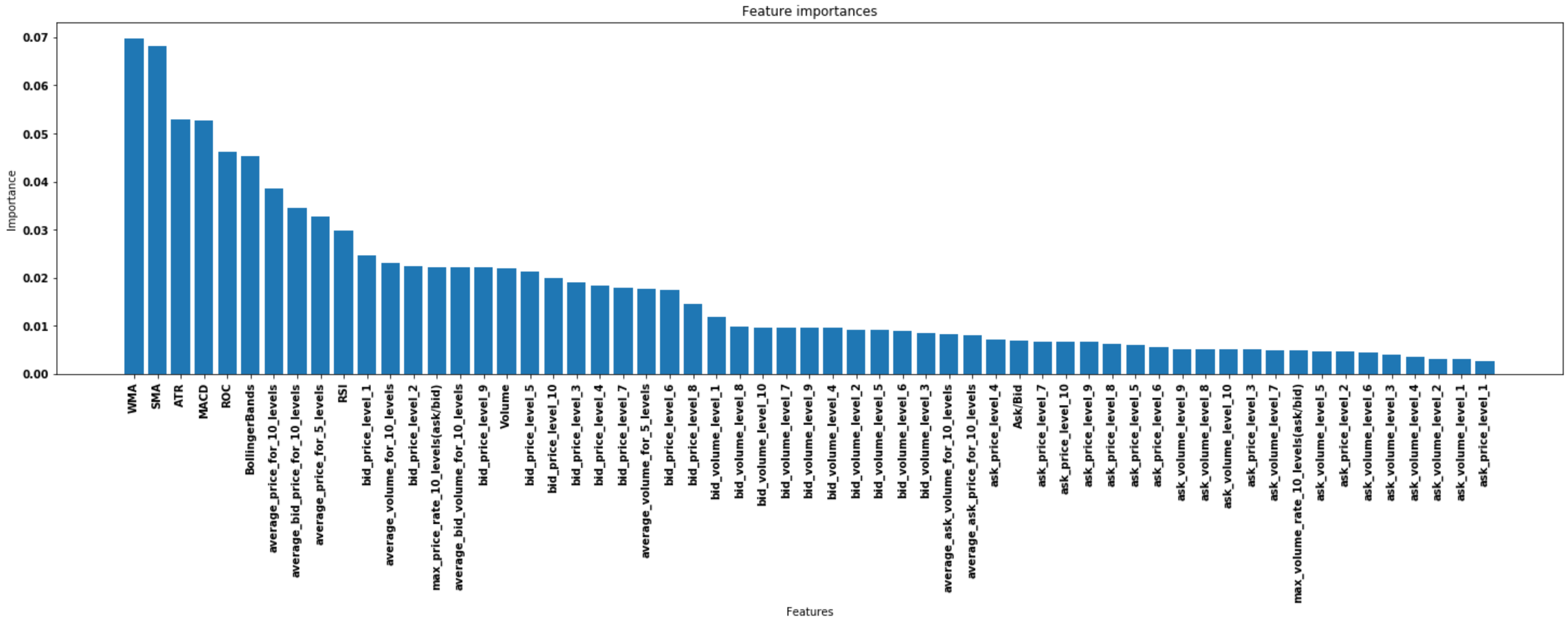


Table 22 Selected Features

Stock	Selected Features
AKCNS	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_volume_level_5, bid_volume_level_6, bid_volume_level_7, bid_volume_level_8, bid_volume_level_9, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
AKFGY	[Ask/Bid, Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_volume_level_4, bid_volume_level_5, bid_volume_level_6, bid_volume_level_7, bid_volume_level_8, bid_volume_level_9, bid_volume_level_10, average_volume_for_10_levels, average_ask_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_ask_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
ALKIM	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_volume_level_5, bid_volume_level_6, bid_volume_level_7, bid_volume_level_8, bid_volume_level_9, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
ARMDA	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_price_level_4,

	bid_volume_level_5, bid_price_level_5, bid_volume_level_6, bid_price_level_6, bid_price_level_7, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
BOLUC	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_volume_level_5, bid_volume_level_6, bid_volume_level_7, bid_volume_level_8, bid_volume_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_ask_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
BOSSA	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_price_level_2, bid_price_level_3, bid_price_level_4, bid_price_level_5, bid_price_level_6, bid_volume_level_7, bid_price_level_7, bid_volume_level_8, bid_price_level_8, bid_volume_level_9, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
BUCIM	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_volume_level_5, bid_volume_level_6, bid_volume_level_7, bid_volume_level_8, bid_volume_level_9, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]

CIMSA	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_volume_level_5, bid_volume_level_6, bid_volume_level_7, bid_volume_level_8, bid_volume_level_9, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
DOAS	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_price_level_4, bid_volume_level_5, bid_volume_level_6, bid_volume_level_7, bid_volume_level_8, bid_volume_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)
EGPRO	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_price_level_4, bid_volume_level_5, bid_price_level_5, bid_price_level_6, bid_price_level_7, bid_price_level_8, bid_price_level_9, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
FMIZP	[MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, ask_price_level_8, ask_price_level_9, ask_volume_level_10, ask_price_level_10, bid_price_level_1, bid_price_level_2, bid_price_level_3, bid_price_level_4, bid_price_level_5, bid_price_level_6, bid_price_level_7, bid_price_level_8, bid_price_level_9, bid_price_level_10, average_volume_for_10_levels, average_ask_volume_for_10_levels, average_bid_volume_for_10_levels,

	average_volume_for_5_levels, average_price_for_10_levels, average_ask_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
GOLTS	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_price_level_4, bid_price_level_5, bid_volume_level_6, bid_price_level_6, bid_price_level_7, bid_price_level_8, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
HEKTS	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_price_level_4, bid_price_level_5, bid_price_level_6, bid_price_level_7, bid_volume_level_8, bid_price_level_8, bid_volume_level_9, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
IZOCM	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, ask_volume_level_10, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_price_level_3, bid_price_level_4, bid_price_level_5, bid_price_level_6, bid_price_level_7, bid_price_level_8, bid_price_level_9, bid_price_level_10, average_volume_for_10_levels, average_ask_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_ask_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
KLMSN	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_price_level_4, bid_price_level_5, bid_price_level_6, bid_volume_level_7, bid_price_level_7, bid_price_level_8, bid_price_level_9, bid_volume_level_10,

	bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
NUGYO	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_price_level_3, bid_price_level_4, bid_price_level_5, bid_price_level_6, bid_price_level_7, bid_price_level_8, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_ask_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_ask_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
PNSUT	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_price_level_4, bid_price_level_5, bid_price_level_6, bid_price_level_7, bid_price_level_8, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
TAVHL	[Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_price_level_4, bid_price_level_5, bid_price_level_6, bid_price_level_7, bid_price_level_8, bid_price_level_9, bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_ask_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_ask_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
TRGYO	[Ask/Bid, Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_price_level_3, bid_volume_level_4, bid_volume_level_5, bid_volume_level_6,

	bid_volume_level_7, bid_volume_level_8, bid_volume_level_9, bid_volume_level_10, average_volume_for_10_levels, average_ask_volume_for_10_levels, average_bid_volume_for_10_levels, average_volume_for_5_levels, average_price_for_10_levels, average_bid_price_for_10_levels, average_price_for_5_levels, max_price_rate_10_levels(ask/bid)]
YGGYO	Ask/Bid, Volume, MACD, SMA, WMA, RSI, ROC, ATR, BollingerBands, bid_volume_level_1, bid_price_level_1, bid_volume_level_2, bid_price_level_2, bid_volume_level_3, bid_volume_level_4, bid_volume_level_5, bid_volume_level_6, bid_volume_level_7, bid_volume_level_8, bid_volume_level_9,bid_volume_level_10, bid_price_level_10, average_volume_for_10_levels, average_ask_volume_for_10_levels, average_bid_volume_for_10_levels , average_volume_for_5_levels, average_price_for_10_levels , average_bid_price_for_10_levels, average_price_for_5_levels , max_price_rate_10_levels(ask/bid)]

EK 3 – Prediction Performances

Table 23 All Experiment Results For 15 Minutes Test Data

Stock	Technique	Accuracy	Precision	Recall	F1	UP-F1	DW-F1	NC-F1
AKCNS	MLP	0.45	0.43	0.43	0.43	0.49	0.51	0.29
AKCNS	RF	0.60	0.59	0.59	0.59	0.62	0.65	0.49
AKFYG	MLP	0.47	0.47	0.47	0.47	0.44	0.50	0.46
AKFYG	RF	0.61	0.61	0.60	0.60	0.58	0.63	0.60
ALKIM	MLP	0.59	0.49	0.46	0.45	0.60	0.65	0.09
ALKIM	RF	0.75	0.66	0.65	0.65	0.78	0.81	0.37
ARMDA	MLP	0.59	0.49	0.47	0.45	0.63	0.63	0.09
ARMDA	RF	0.73	0.67	0.65	0.66	0.76	0.78	0.43
BOLUC	MLP	0.53	0.50	0.50	0.50	0.57	0.59	0.34

BOLUC	RF	0.66	0.64	0.64	0.64	0.72	0.71	0.48
BOSSA	MLP	0.48	0.32	0.37	0.34	0.49	0.54	0.00
BOSSA	RF	0.76	0.69	0.68	0.68	0.80	0.82	0.43
BUCIM	MLP	0.49	0.49	0.49	0.49	0.54	0.52	0.40
BUCIM	RF	0.62	0.62	0.62	0.62	0.65	0.64	0.56
CIMSA	MLP	0.53	0.44	0.44	0.43	0.54	0.62	0.12
CIMSA	RF	0.69	0.64	0.64	0.64	0.73	0.76	0.44
DOAS	MLP	0.52	0.46	0.45	0.44	0.56	0.60	0.16
DOAS	RF	0.72	0.69	0.68	0.69	0.78	0.79	0.49
EGPRO	MLP	0.55	0.51	0.45	0.42	0.59	0.61	0.07
EGPRO	RF	0.72	0.65	0.65	0.65	0.76	0.78	0.41
FMIZP	MLP	0.52	0.34	0.39	0.36	0.52	0.57	0.00
FMIZP	RF	0.73	0.63	0.62	0.62	0.75	0.79	0.33
GOLTS	MLP	0.60	0.48	0.45	0.43	0.63	0.65	0.02
GOLTS	RF	0.72	0.65	0.65	0.65	0.80	0.82	0.33
HEKTS	MLP	0.55	0.50	0.43	0.55	0.58	0.60	0.03
HEKTS	RF	0.76	0.69	0.68	0.68	0.81	0.82	0.42
IZOCM	MLP	0.53	0.69	0.42	0.39	0.56	0.57	0.03
IZOCM	RF	0.67	0.62	0.60	0.60	0.70	0.72	0.39
KLMSN	MLP	0.49	0.43	0.41	0.39	0.53	0.55	0.10
KLMSN	RF	0.73	0.68	0.68	0.68	0.77	0.80	0.48
NUGYO	MLP	0.45	0.41	0.41	0.40	0.49	0.52	0.19
NUGYO	RF	0.71	0.69	0.69	0.69	0.77	0.77	0.55
PNSUT	MLP	0.51	0.45	0.44	0.42	0.55	0.59	0.12
PNSUT	RF	0.70	0.67	0.67	0.67	0.74	0.76	0.50
TAVHL	MLP	0.67	0.53	0.49	0.47	0.69	0.71	0.01
TAVHL	RF	0.84	0.74	0.73	0.73	0.88	0.89	0.43
TRGYO	MLP	0.49	0.49	0.49	0.49	0.55	0.52	0.40

TRGYO	RF	0.71	0.72	0.71	0.71	0.77	0.73	0.65
YGGYO	MLP	0.48	0.48	0.48	0.48	0.53	0.49	0.41
YGGYO	RF	0.60	0.60	0.59	0.59	0.64	0.60	0.54

Figure 12 ACNS Prediction Performances – MLP and RF

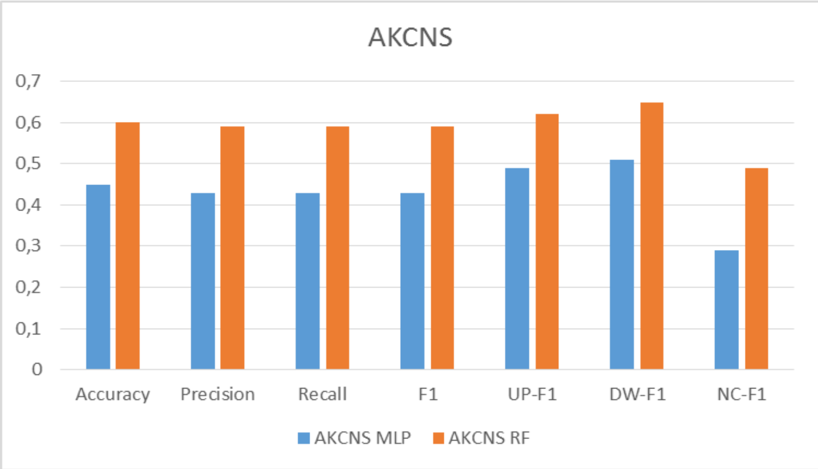


Figure 13 AKFYG Prediction Performances – MLP and RF

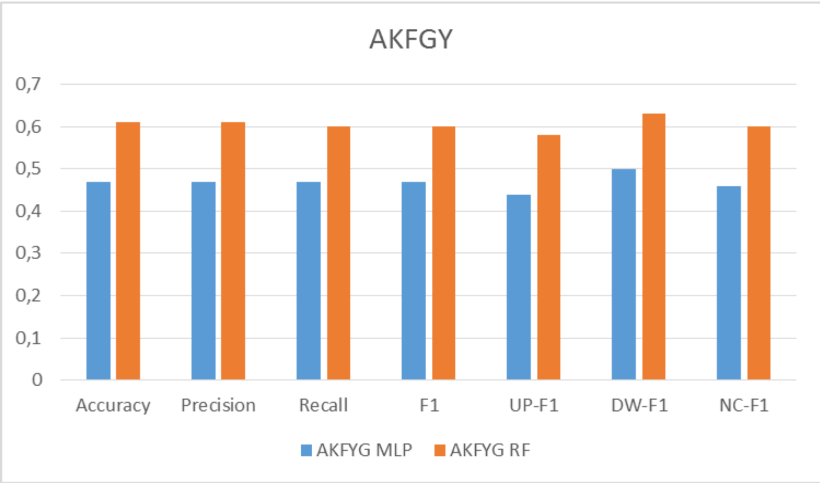


Figure 14 ALKIM - Prediction Performances– MLP and RF

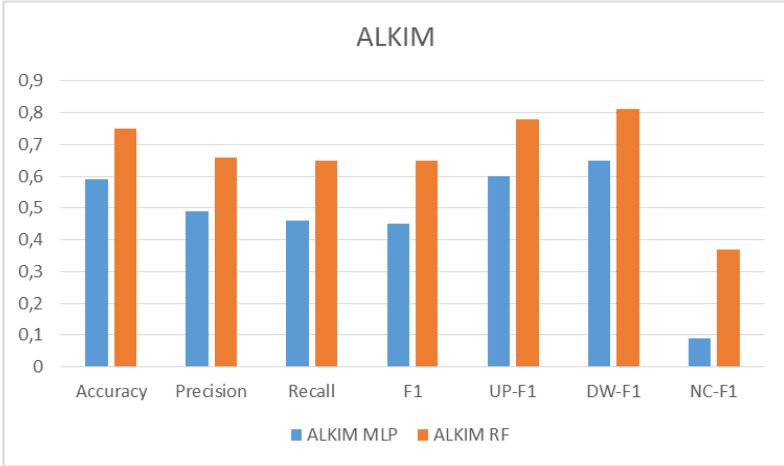


Figure 15 ARMDA - Prediction Performances– MLP and RF

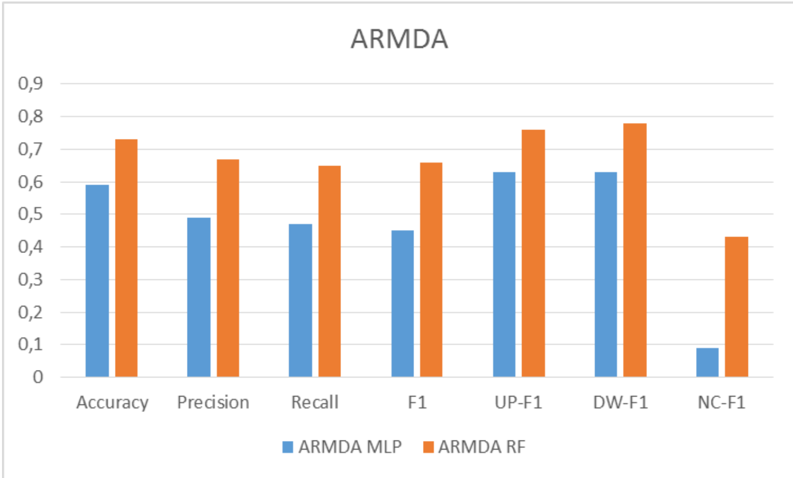


Figure 16 BUCIM - Prediction Performances– MLP and RF

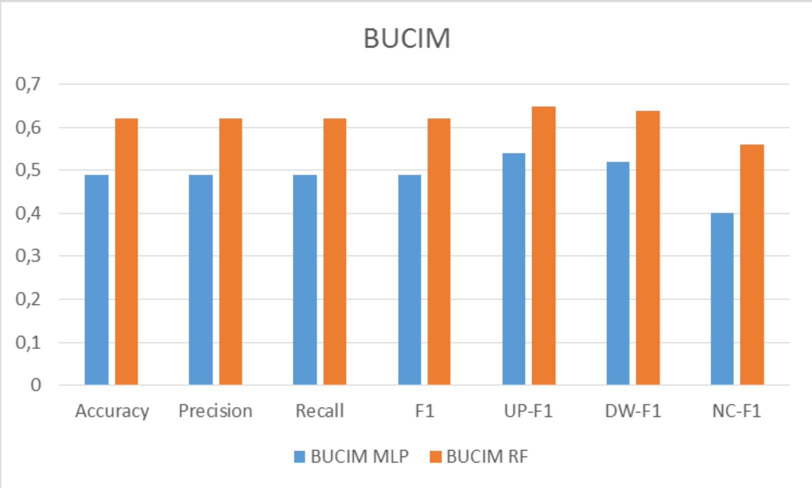


Figure 17 BOSSA - Prediction Performances– MLP and RF

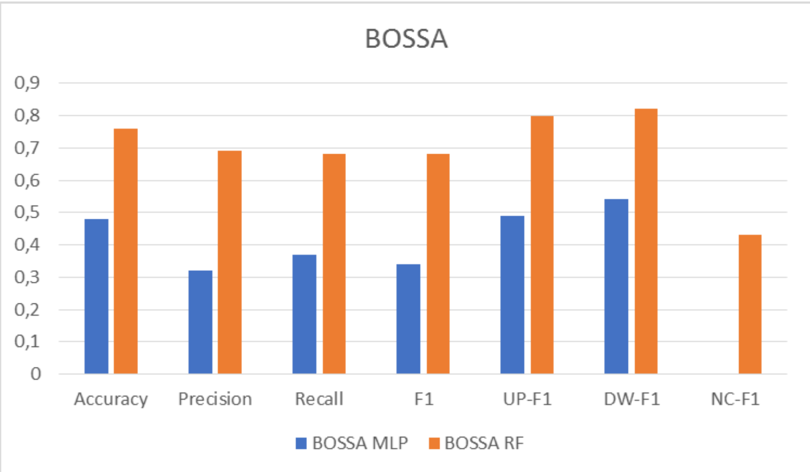


Figure 18 BOLUC - Prediction Performances– MLP and RF

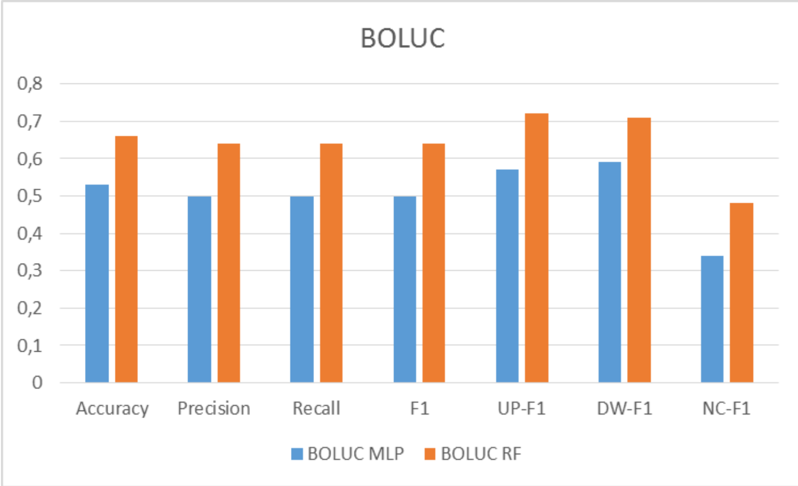


Figure 19 CIMSA - Prediction Performances– MLP and RF

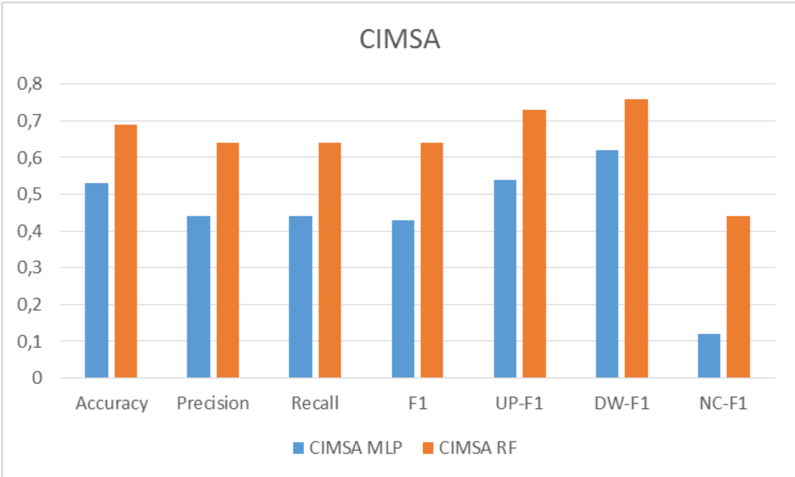


Figure 20 DOAS - Prediction Performances– MLP and RF

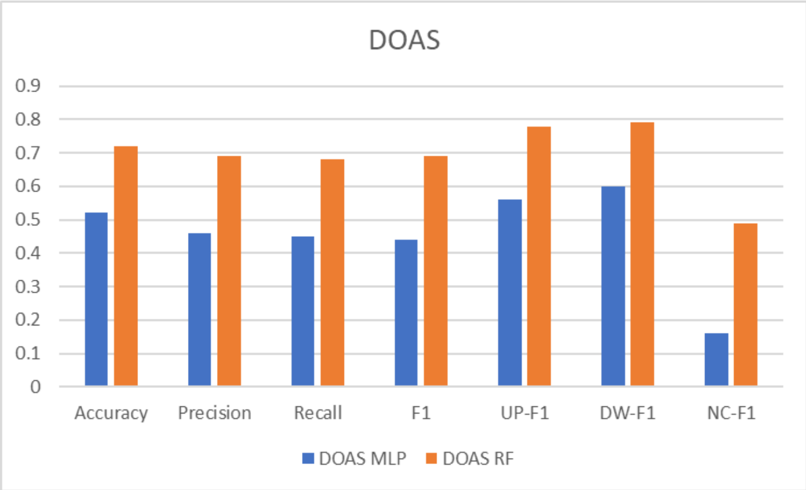


Figure 21 EGPRO - Prediction Performances– MLP and RF

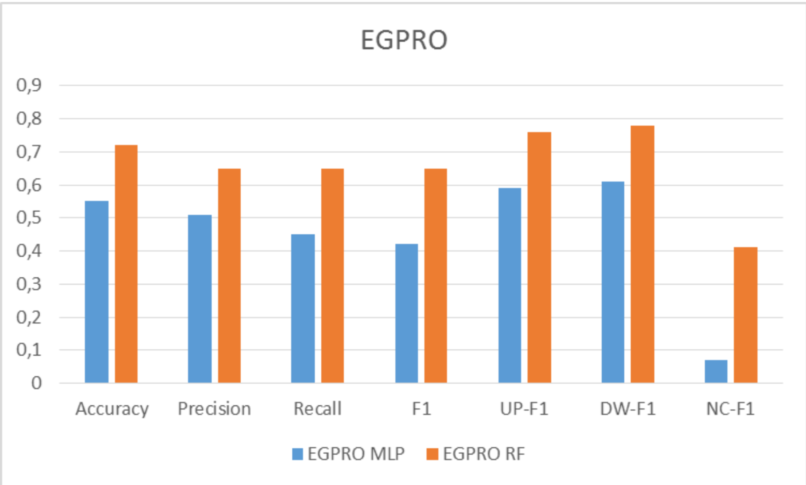


Figure 22 FMIZP - Prediction Performances– MLP and RF

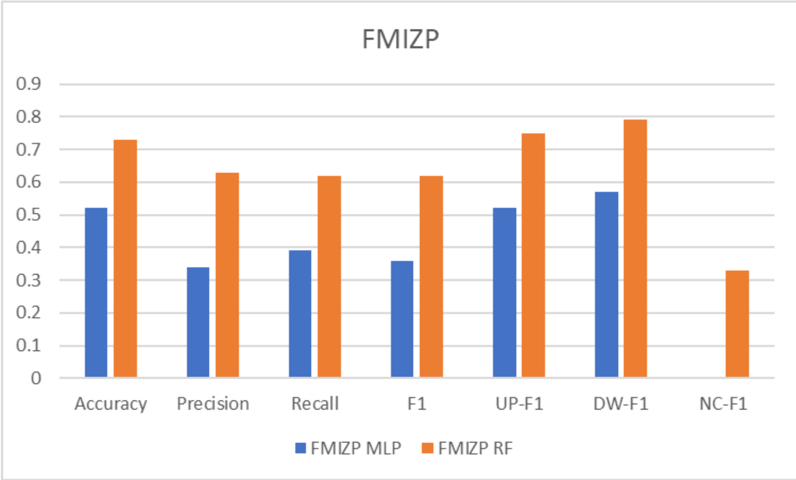


Figure 23 GOLTS - Prediction Performances– MLP and RF

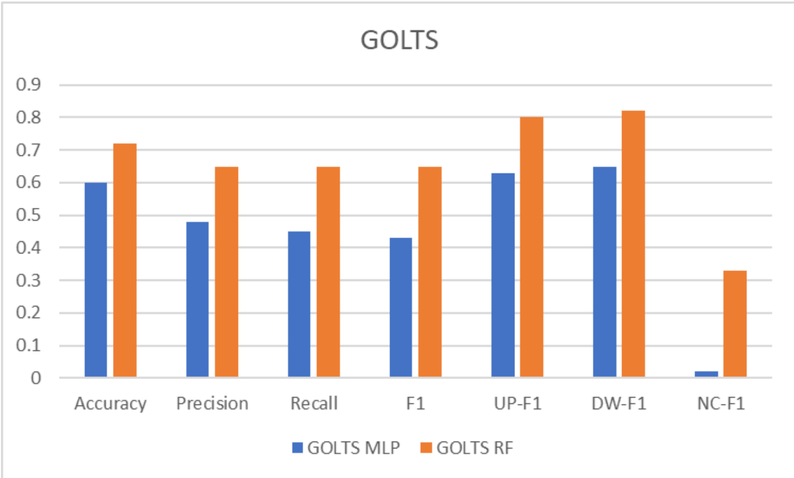


Figure 24 HEKTS - Prediction Performances– MLP and RF

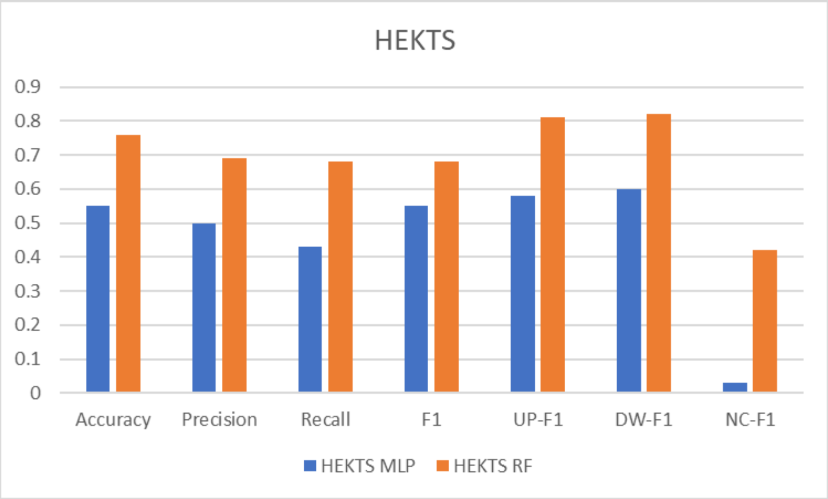


Figure 25 IZOCM - Prediction Performances– MLP and RF

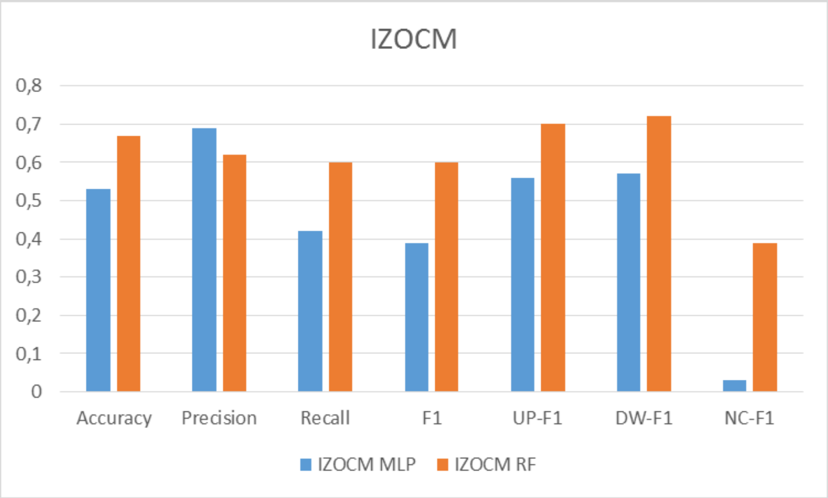


Figure 26 HEKTS - Prediction Performances– MLP and RF

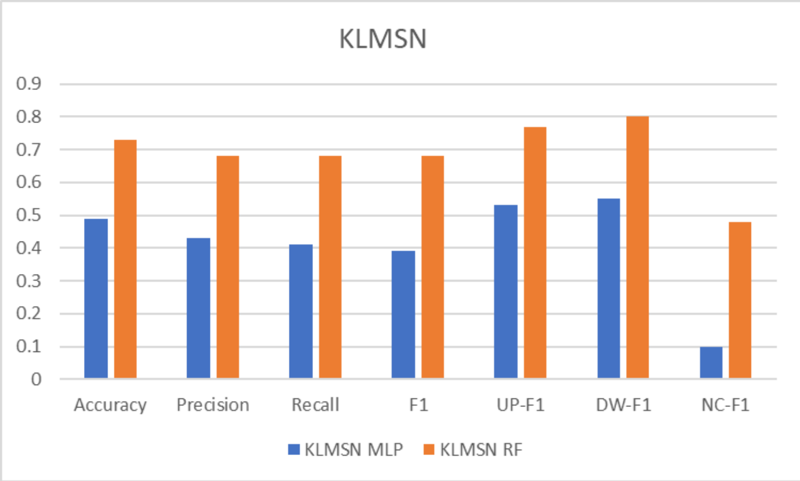


Figure 27 NUGYO - Prediction Performances– MLP and RF

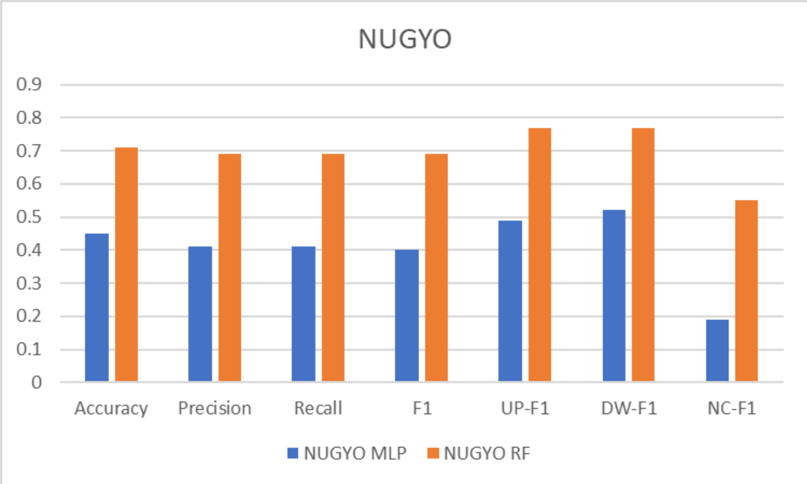


Figure 28 PNSUT - Prediction Performances– MLP and RF

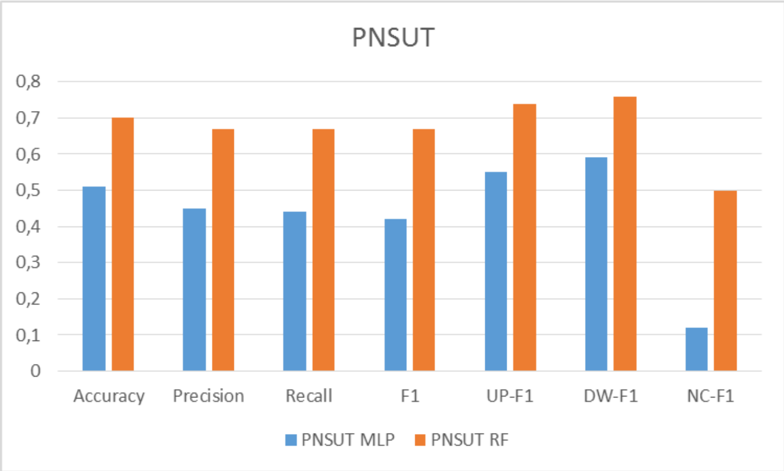


Figure 29 TAVHL - Prediction Performances– MLP and RF

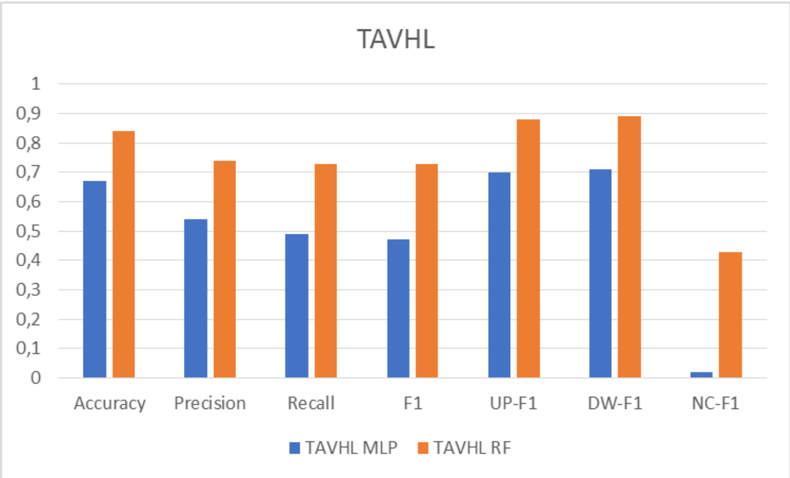


Figure 30 TRGYO - Prediction Performances– MLP and RF

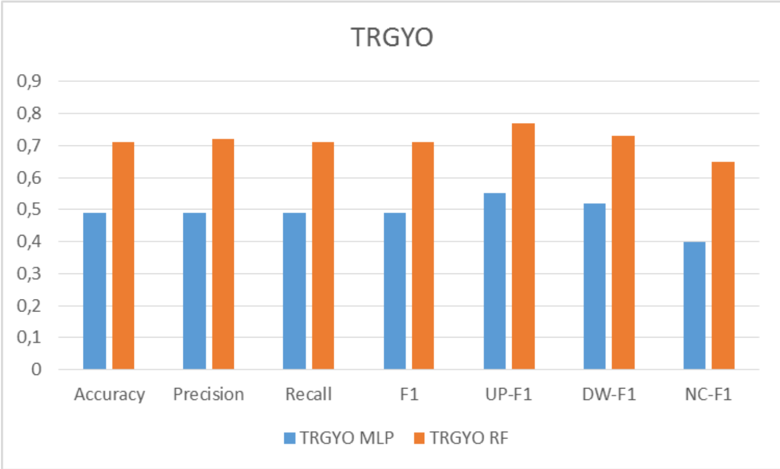
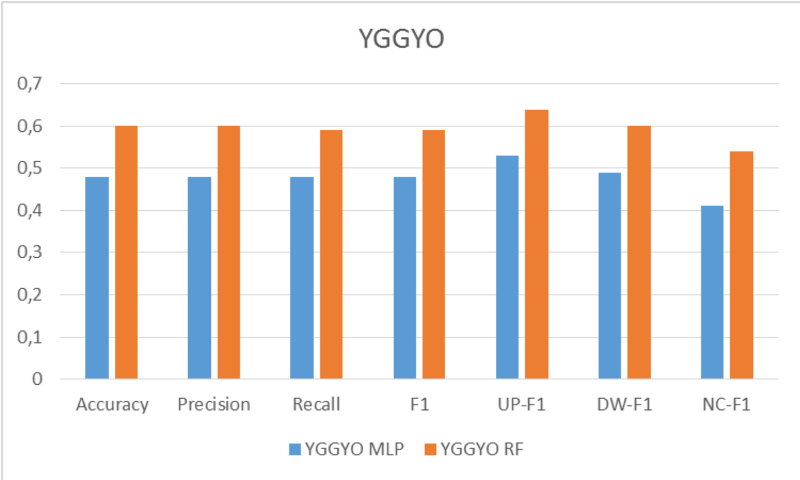


Figure 31 YGGYO - Prediction Performances– MLP and RF



HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
THESIS/~~DISSERTATION~~ ORIGINALITY REPORT

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF SCIENCE AND ENGINEERING
TO THE DEPARTMENT OF COMPUTER ENGINEERING

Date: 12/09/2019

Thesis Title / Topic: A Hybrid and Reliable Method Integrating Depth and Technical Analysis with Machine Learning Techniques for Predicting Stock Prices

According to the originality report obtained by ~~myself~~/my thesis advisor by using the *Turnitin* plagiarism detection software and by applying the filtering options stated below on 12/09/2019 for the total of 54 pages including the a) Title Page, b) Introduction, c) Main Chapters, d) Conclusion sections of my thesis entitled as above, the similarity index of my thesis is 9 %.

Filtering options applied:

1. Bibliography/Works Cited excluded
2. Quotes excluded / ~~included~~
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Sciene and Engineering Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.

12/09/2019

Name Surname: Seçil TABUROĞLU

Student No: N16128801

Department: Computer Engineering

Program: Computer Engineering

Status: ☒ Masters ☐ Ph.D. ☐ Integrated Ph.D.

ADVISOR APPROVAL

APPROVED.

Asst. Prof. Dr. Fuat AKAL

AUTHOR'S CV

Name, Surname	: SEÇİL TABUROĞLU
Place of Birth	: KIRŞEHİR
Date of Birth	: 17.08.1984
Marital Status	: Single
E-mail	: seciltaburoglu@gmail.com
Foreign Language	: English

EDUCATION

Bachelor of Science (B.Sc.)	: Computer Engineering, Hacettepe University
Master of Science (M.Sc.)	: Computer Engineering, Hacettepe University
Doktora	:

Work Experience

2015- …	Turkish Aerospace Industries, Inc. (TAI)
2011-2015	Tübitak SAGE
2007-2011	AYESAŞ

Areas of Experiences : -

Projects and Budgets : -

Publications: