Hacettepe University Graduate School of Social Sciences
Department of Economics

CONSTRUCTING TRADING STRATEGIES USING ARTIFICIAL INTELLIGENCE BASED MODELS: AN APPLICATION FOR BORSA ISTANBUL

Berçim BERBEROĞLU

Master's Thesis

Ankara, 2020

CONSTRUCTING TRADING STRATEGIES USING ARTIFICIAL INTELLIGENCE BASED MODELS: AN APPLICATION FOR BORSA ISTANBUL

Berçim BERBEROĞLU

Hacettepe University Graduate School of Social Sciences
Department of Economics

Master's Thesis

Ankara, 2020

To my parents

ACKNOWLEDGEMENTS

First and foremost, I would like to convey my deepest gratitude to my advisor, Dr. Başak Dalgıç for supporting me with her deep knowledge, close attention and continuous guidance during this thesis. She treates me not only as a mentor but also like a sister and guide me in my academic journey as well as my personel life. It has been a honour to work under her supervision.

I would also like to thank to the rest of my thesis comittee: Dr. Füsun Küçükbay and Dr. Özge Kandemir Kocaaslan for their valuable contributions to my thesis and taking out the time to be a part of my thesis committee.

I owe my thanks to Mustafa Kemal Binli, Dr. Selim Baha Yıldız, Dr. Zeynep Çipiloğlu Yıldız, Dr. Deniz Kılınç and Murat Şahin who encourage me to dare to work in the field of artificial intelligence. I am very grateful them for spending their precious time for answering my questions with great patience and wisdom. I am also thankful to Dr. Orhan Erdem for taking his time to help me in news sentiment data collecting stage.

I also deeply thankful to Dr. Hatice Yurtsever for always and in all conditions offering me her sincere support and guidance without any hesitation.

I want to express my sincere gratitude to my best friend Aslıhan Gül, for being always there for me. I cannot thank you enough. I also want to express my appreciation to my colleagues Ebru Sürücü Balcı and Ökkeş Çelen for always taking out the time to discuss the ideas with me.

I would like to express my profound gratitude to my dear aunt Zehra Hastürk, my dear uncle Hüseyin Demirel and my all family members who I can not count their names, for all their valuable support. I will always be grateful to you.

Lastly, I am deeply thankful more than anyone to my endless supporters, my wingless angels: my mom Alev Berberoğlu and my dad Birol Berberoğlu. No words can adequately express my gratitude to you. I love you dearly.

ABSTRACT

BERBEROĞLU, Berçim. *Constructing Trading Strategies Using Artificial Intelligence Based Models: An Application For Borsa Istanbul*, Master's Thesis, Ankara, 2020.

The aim of this thesis is to test the Efficient Market Hypothesis using artificial intelligence-based techniques. In this regard, we utilize artificial intelligence based models that have both deep and shallow architectures which are Long Short Term Memory (LSTM) Networks and Support Vector Regression (SVR) to predict the next day's close price of the selected stocks from BIST30 Index. Next, we construct trading strategies by making use of the predictions produced by the forecasting models. We feed these models using a comprehensive dataset including technical analysis indicators and investor sentiment variables. Thus, we predict the following day's close prices both by using historical price data which is accessible without any costs and the investor sentiment containing market's non-rational components. In order to proxy investor sentiment, we use Bloomberg's news sentiment data which is developed to imitate a human in processing financial news. We show the superior performance of our trading strategies that are constructed using both LSTM and SVR models compared to simply buy and hold market index in terms of all performance metrics. Moreover, we reach similar results when transaction costs are considered. Our findings reveal that successful predictions can be made and trading strategies can be built using publicly available information and artificial intelligence-based models. Moreover, investing with these strategies, above-average risk-adjusted return can be yielded. Thus, we provide contradictory evidence to EMH's negatory arguments about the asset price predictability.

Keywords: financial forecasting, machine learning, deep learning, trading strategies, investor sentiment, technical analysis.

TABLE OF CONTENTS

List of Abbreviations

AD Accumulation/Distribution Indicator

Adagrad Adaptive Gradient Algorithm

Adam Adaptive Moment Estimation

ADOSC Chaikin A/D Oscillator

AKBNK Akbank

ARIMA Autoregressive Integrated Moving Average

BIST Borsa Istanbul

BPTT Back Propagation Through Time

CCI Commodity Channel Index

CNN Convolutional Neural Network

DD Drawdown

DNN Deep Neural Network

DOHOL Doğan Holding

EMA Exponential Moving Average

EMH Efficient Market Hypothesis

EREGL Ereğli Demir ve Çelik Fabrikaları

FEARS Financial and Economic Attitudes Re-vealed by Search

GA Genetic Algorithm

GA-LSTM Genetic Algorithm-Long Short Term Memory

GARAN Türkiye Garanti Bankası

GARCH Generalized Autoregressive Conditional Heteroskedasticity

GPOMS Google-Profile of Mood States

ISCTR Türkiye İş Bankası

LSTM Long Short Term Memory

MACD Moving Average Convergence/Divergence

MAE Mean Absolute Error

MAPE Mean Absolute Percentage Error

MDD Maximum Drawdown

MLP Multi Layer Perceptron

MSE Mean Squared Error

NNs Neural Networks

OBV On Balance Volume

OHLCV Open, High,Low, Close, Volume

ReLU Rectified Linear Unit

RMSE Root Mean Squared Error

RMSProp Root Mean Square Propagation

RNN Recurrent Neural Network

ROC Rate of Change

RSI Relative Strength Index

RWH Random Walk Hypothesis

SAHOL Sabancı Holding

SGD Stochastic GradientDescent

SMA Simple Moving Average

SODA Soda Sanayii

SOFNN Self-organizing Fuzzy Neural Network

SVR Support Vector Regression

THYAO Türk Hava Yolları

TS-LSTM Trading Strategy of Long Short Term Memory

TS-SVR Trading Strategy of Support Vector Regression

TSKB Türkiye Sınai ve Kalkınma Bankası

TTS-LSTM Trading Strategy of Long Short Term Memory with Threshold

TTS-SVR Trading Strategy of Support Vector Regression with Threshold

VAKBN Türkiye Vakıflar Bankası

WMA Weighted Moving Average

YKBNK Yapı ve Kredi Bankası

List of Tables

List of Figures

INTRODUCTION

Efficient Market Hypothesis (EMH) which was introduced by Nobel Laureate Eugene Fama (1970) is considered to be the ultimate foundation to explain financial markets. The main arguments of EMH include the unpredictability of asset prices and rationality of the market participants. According to EMH, in efficient markets, when new information arrives, it spreads and incorporates into the prices immediately. Therefore, current prices reflect all available information. Trading on publicly available information, it is not possible for investors to gain above the average risk-adjusted returns. Moreover, EMH emphasize that the investors are entirely rational at their decision-making. When new information arrives, agents update their beliefs and make ”normatively acceptable decisions” (Barberis and Thaler, 2003). Even if some investors act irrationally, rational investors balance their movements, thus any significant effect of irrational investors cannot be observed in the market.

Over time, financial markets have faced the events that cannot be explained by rational facts many times and the validity of EMH has begun to be questioned. Until these experiences, researchers have accepted Jensen's 1978 view about EMH: *"There is no other proposition in economics which has more solid empirical evidence supporting it than the Efficient Markets Hypothesis."* However, they show how traditional models have become insufficient to explain the chaotic framework of the financial phenomena. Further, it is shown that such models are constructed on over-simplified assumptions expressing how investors should behave rather than how they behave (Prosad et al., 2015). Numerous attempts have been made to question the validity of EMH in terms of its arguments whether of that asset prices are unpredictable and investors are rational. In our view, even the success of investors as Warren Buffett, George Soros, Peter Lynch provides evidence that EMH is no longer valid in today's world.

Put differently, in today's increasingly competitive markets, in order to gain competitive advantage against the other market participants, investors need to make complex financial decisions by analysing the big and rapidly accumulating data. Given the dramatically growing datasets, now, investors' decision-making process requires more sophisticated analyses with more advanced methods.

When an individual intends to invest in a financial instrument, he/she should have an optimistic view of this instrument's future. In this step, investor tries to make an appropriate prediction for the following periods. However, making accurate forecasting is an appealing but very challenging task. For example, there are various parameters affecting stock prices and some of these parameters may not be considered while constructing forecasting models. Moreover, the financial system has a dynamic structure and parameters that involved in this dynamic system are continually changing and transforming. Another reason stems from the inherent in financial time series which are "dynamic, chaotic, noisy, non-linear" and have a nonstationary structure (Henrique et al., 2019). Thus, catching their pattern is quite difficult. In this context, the introduction of artificial intelligence techniques in the finance has revolutionized financial forecasting. Given their capabilities in capturing complex patterns and strong capacity to handle noisy and time-dependent data, artificial intelligence-based models are promising and exhibit superior performance in asset price predictions.

The most commonly used artificial intelligence-based models in financial forecasting consists of machine learning algorithms such as Support Vector Machine (SVM), Support Vector Regression (SVR), Decision Trees, Naive Bayes, Gaussian Naive Bayes, Nearest Neighbours and deep learning models such as Deep Multi-Layer Perceptron, Convolutional Neural Networks, Recurrent Neural Networks, Long-Short Term Memory, Deep Reinforcement Learning, etc. Considering the financial forecasting literature, one can see that LSTM and its variations along with some hybrid models dominate this domain (Sezer et al., 2020). LSTM is one of the most advanced deep learning models for sequence learning tasks including natural language processing, handwriting recognition and time series prediction. It is inherently suitable especially for financial forecasting and exhibits superior performance in such field thanks to its internal feedback mechanism between neurons that allows to memorization of important past information (Chung and Shin, 2018). On the other hand, SVR is another model that exhibit superior performance in time series forecasting. SVR is robust to the outliers, has high generalization capability, very easy to implement and can give good results with high accuracy based on a sparse subset of the whole training set (Sundaram et al., 2006; Awad and Khanna, 2015). Even though SVR has a shallow architecture compared to deep learning models, in financial forecasting literature there are many studies showing its superior performance against the models with deep architectures (Ince and Trafalis, 2008; Maqsood et al., 2020).

In this study, our aim is to test EMH by using artificial intelligence-based models. In this context, predicting asset prices to some extent does not provide contradictory evidence about the validity of EMH. One should take the advantage of this degree of predictability and construct the trading strategies to yield above-average risk-adjusted returns. Accordingly, our tests for the validity of EMH include two stages as: forecasting stock prices and applying these predictions through trading strategies. In this regard, we utilize artificial intelligence-based models to predict the next day's close price of the stocks. Thereafter, we construct trading strategies by making use of the predictions produced by the forecasting models.

We feed artificial intelligence-based models by using historical price data which is information accessible without any costs and by using the investor sentiment that contains the market's non-rational components. We choose this feature set in response to the EMH's arguments about weak-form efficiency and investor's pure rationality. In order to proxy investor sentiment, we use Bloomberg's news sentiment scores. Our choice follows the idea that the information extracted from news stories form the basis of investors' sentiment about the relevant instrument and news sentiment carries the irrational components of the markets. We choose our sample from Borsa Istanbul, utilizing the literature which argues that emerging markets are less efficient, therefore more predictable and more exposed to investor sentiment. We randomly select 10 of 45 stocks included in the BIST30 index which are among to the most capitalized and the most actively traded stocks in the past few years following the view that investors closely follow these most popular stocks' stories.

We construct our trading strategies by using LSTM and SVR models. One of these strategies mainly focuses on the direction of the price movements, while the other considers a certain threshold value in generating buy and sell signals. We take the advantage of dealing with regression instead of classification although we are interested in movement direction. This allows us to flexibility to differentiate our strategies. Then, in order to make sure whether the transaction costs will cause any disadvantage for us, we add them our trading system by making a fixed commission cut for every transaction. Finally we evaluate the performance of all trading strategies in terms of return, volatility, Sharpe ratio and Maximum Drawdown (MDD). We show the superior performance of our trading strategies compared to simply buy and hold market index in terms of all performance metrics. We reach similar results even after transaction costs are considered. Moreover, we determine the backtest period as 309 days which corresponds to more than one trading year. During this period, we consistently beat the market. This reinforces the evidence that we have not achieved this success by chance, which contradicts one of the main propositions of EMH advocating that no one can consistently outperform the market except by taking more risk, or by chance. It is all about the success of our forecasting models and trading systems. We show that we can make successful forecastings and construct trading strategies based on these predictions using artificial intelligenc- based models and publicly available information which is the fundamental purpose of this study. Thus, we provide contradictory evidence with the negatory argument about the asset price predictability of EMH.

Our contribution in this study is manifold. While the literature on stock price forecasting is still developing, many studies focus on the developed financial markets. However, there are limited number of studies in emerging markets, especially for Borsa Istanbul. Most of the studies for the sample of Borsa Istanbul focus on the direction of the market index rather than examine the stocks individually. To the best of our knowledge, this is the first study predicting close prices of the Borsa Istanbul stocks individually using LSTM and SVR models. Apart from the existing studies that do not produce practical implications from prediction results or making this for BIST index, this study is the first attempt to construct trading strategies based on the predictions for BIST stocks. Our focusing on forecasting a continuous target rather than the direction of the price movements allows us the flexibility to differentiate trading strategies. This enables us to generate an infinite number of trading strategies. Further, existing studies feed their models mostly with technical analysis indicators and do not consider non-rational components of the market. Although there are studies using news sentiment as a proxy for investor sentiment, this study is the first endeavour that proxies investor sentiment utilizing Bloomberg's news sentiment data which is developed to imitate a human in processing financial news for the sample of Borsa Istanbul stocks.

The rest of this study is organized as follows. In Chapter 1, we present a theoretical background for our investigation. In this regard, we explain EMH and the conflicting arguments with EMH. In Chapter 2, we present the existing literature about investor sentiment and search for an answer to the question of why news sentiment is an appropriate measure to proxy investor sentiment? Moreover, we review the studies that forecast asset prices using artificial intelligence-based models. In Chapter 3, we introduce our dataset including technical analysis indicators and news sentiment scores. In Chapter 4, we explain our forecasting and trading strategy construction methodology. In Chapter 5, our findings of forecasting models and backtests are presented. Finally, Chapter 6 concludes the study.

Chapter 1

THEORETICAL BACKGROUND

1.1 EFFICIENT MARKET HYPOTHESIS

Efficient Market Hypothesis (EMH) which was introduced by Nobel Laureate Eugene Fama (1970) is considered to be the ultimate foundation to explain financial markets. EMH suggests that markets are fully efficient in reflecting information and everyone has same degree of accessibility to this information. In efficient markets, when new information arrives, it spreads and incorporates into the prices immediately. Therefore, current prices reflect all available information. It is not possible for investors to gain above the average risk adjusted returns trading on publicly available information. Since stocks' are correctly priced and financial markets are informationally efficient, it is suggested that investors should adopt a passive strategy which implies buy and hold the investment portfolio rather than an active strategy involving high-frequency buy-sell transactions.

Tests in EMH contains three levels of market efficiency as weak, semi-strong and strong form. The situation where all historical information of an asset is reflected in the prices implies weak form efficiency whereas the situation that all publicly available information reflected in the prices indicate semi-strong form efficiency. Lastly, the situation that both public and private information reflected in the prices points out strong form efficiency. If a market is semi-strong form efficient, then it is also weak form efficient since historical prices are also publicly available information. Moreover, a strong form efficient market is also both weak and semi-strong form efficient, since all private and publicly available information covers the information sets in such efficiency forms.

If a market is in weak form efficiency, then technical analysis will not work in predicting asset prices. Historical asset price data is an information that can be accessed by anyone without any cost. If reliable signals about the future performance of the stock can be obtained using historical data, then all traders will learn to use such signals and it will not matter to use this data. Moreover, in semi-strong form efficient markets, neither technical analysis nor fundamental analysis will not allow investor to beat the market. In other words, trading activities that are executed using technical analysis and fundamental analysis will not be profitable if the markets are in semi-strong form efficient. Consequently, investors can not consistently beat the market with their skills and all of their efforts are in vain (Park and Irwin, 2007). Then, the market can only be beaten when the inefficiency occurs.

EMH is associated with the Random Walk Hypothesis (RWH) which simply implies asset prices follow random walk and past values of the prices cannot be used to predict future movements. Tomorrow's price movements are completely independent from today, i.e. they are only related to tomorrow's information flows. As it is not possible to predict tomorrow's events before they occur, future prices cannot be predicted.

EMH emphasize that the investors are fully rational at decision making. Arbitrageurs or informed investors have the characteristics of the homoeconomicus which is a model of humans' economic behaviour under the assumptions of perfect self-interest, perfect rationality, and perfect information. They hold all existing information and use them with pure rationality. When new information arrives, agents update their beliefs and make ”normatively acceptable decisions” (Barberis and Thaler, 2003).

According to EMH, in an efficient market, profit opportunities do not exist. When mispricing occurs, arbitrageurs bring prices to their fundamental values immediately and ensure that the markets are efficient. Moreover, even if some investors act irrationally, rational investors balance their movements. Thus any significant effect of irrational investors cannot be observed in the market. Accordingly, EMH does not assign any role for irrational investors (i.e noise traders) to create significant effects on asset prices.

1.2 THE CONFLICTING ARGUMENTS WITH EMH

1.2.1 Predictability of Asset Prices

After the financial crashes experienced that could not be explained by traditional theories, the validity of EMH and RWH began to be questioned. Numerous attempts have been made to test the validity of EMH and RWH. According to Qian and Rasheed (2007) while early works support the random walk behavior of the prices in general, more recent studies reject RWH. Even Fama (1991) said *"the extreme version of the market efficiency hypothesis is surely false."*

The negatory arguments of EMH and RWH regarding asset price predictability contradict with the ideas that advocate that the technical analysis, fundamental analysis or modern financial forecasting methods have power on predicting asset prices and provide profitable trading strategies.

Fundamental analysis is the examination of a set of information about the general situation of a country, sector and company, financial status of related and competing companies, sectors etc. Namely, fundamental analysis utilizes factors that are related to a company, the industry and the country of that the company, to predict the intrinsic values of its securities. It is an attempt to determine the discounted present value of all payments that the shareholder will receive from the stock. Fundamental analysts usually start their analyses by investigating past dividend payments and financial statements of a company. Then they enrich their analyses with more detailed economic investigations about the country and the industry as well as using the information on the company's management quality (Bodie et al., 2000). While fundamental analysis provides good solutions in long-term predictions, in short-term asset price forecasting technical analysis is found more useful (Khan et al., 2011; Nti et al., 2019).

Technical analysis is a state-of-art tool to predict future prices by taking the advantage of the asset's past and present price movements (Stanković et al., 2015). Technical analysts suggest that the asset prices have recurrent patterns and thus they are predictable. Moreover, the duration of these patterns is long enough to get recognized and to use them as signals providing profitable strategies to investors. All the technical indicators and oscillators such as Relative Strength Index (RSI), Commodity Channel Index (CCI), Moving Average Convergence/Divergence (MACD), William's R, momentum, Rate of Change

(ROC), Stochastic K, Stochastic D can be calculated by using raw price data (OHLCV namely Open, High, Low, Close, Volume). Although, some people see technical analysis as *‘voodoo finance”*, technical analysis can be an effective way in extracting useful information from historical price data (Lo et al., 2000).

In today’s increasingly competitive markets, in order to gain competitive advantage against the other market participants, investors need to make complex financial decisions by analysing the big and rapidly accumulating data. Given the dramatically growing datasets, now, investors’ decision-making process requires more sophisticated analyses with more advanced methods. In this context, the introduction of artificial intelligence applications in finance has revolutionized financial forecasting. Given their capabilities in capturing complex patterns and strong capacity to handle with noisy and time-dependent data, artificial intelligence-based models exhibit promising and superior performance in forecasting stock movements.

1.2.2 Standard Finance versus Behavioural Finance

Considering the history of financial markets, it is difficult to be a strict supporter of traditional theories. Beyond the rationality argument advocated by standard theories, financial markets have experienced the events that cannot be explained by rational facts many times such as Great Depression in 1929, the Tronics Boom, the Go-Go Years in 1960s, the Nifty Fifty bubble in 1970s, the Black Monday in 1987, Dotcom bubble in 1990s, subprime mortgage crisis in 2008. These examples show how traditional models have become insufficient to explain the chaotic framework of the financial phenomena. Until these crashes are experienced, researchers have accepted Jensen’s (1978) view about EMH: *“There is no other proposition in economics which has more solid empirical evidence supporting it than the Efficient Markets Hypothesis.”* However, later on, they began to question the fundamental assumptions of traditional models and presented contradictory evidence towards these models. It is clear that these models are constructed on over-simplified assumptions expressing how investors should behave rather than how they behave (Prosad et al., 2015).

While EMH advocates a strict market efficiency, investors’ behaviours often diverge from most of the assumptions of EMH. In this respect, the field of behavioural finance tries to capture the effect of investors’ irrationality by making use of the models incorporating both finance and psychology. Behavioural theories permit inefficiencies caused by market participants’ irrational actions and mistakes in their valuations. Moreover, they define agents as “normal” rather than “rational” (Statman, 1999). These theories accept that investors do not act perfectly rational and subject to sentiment. The consequences of irrational behaviours can be quite costly and should not be ignored. Behavioural finance models try to avoid this ignorance by taking the advantage of the approaches from both finance and psychology fields.

The emergence of the main ideas of behavioural finance traces back to the 1700s. In The Theory of Moral Sentiments, Adam Smith (1759) suggests that humans behaviours are determined with the competition between the passion and the impartial spectator. Ashraf et al. (2005) describe this relationship from Smith’s point of view as follows:

> *‘‘Smith viewed behavior as under the direct control of the passions, but believed that people could override passion-driven behavior by viewing their own behavior from the perspective of an outsider—the impartial spectator—a “moral hector who, looking over the shoulder of the economic man, scrutinizes every move he makes.””*

Moreover, Smith emphasizes the origins of the behavioural concepts which are referred as loss aversion, intertemporal choice and overconfidence today (Ashraf et al., 2005). Nearly two centuries after Smith, Keynes (1936) has introduced the term “animal spirits” in order to express the idea of how emotions impact on economic behaviours:

> *“... There is the instability due to the characteristic of human nature that a large proportion of our positive activities depend on spontaneous optimism rather than mathematical expectations, whether moral or hedonistic or economic. Most, probably, of our decisions to do something positive, the full consequences of which will be drawn out over many days to come, can only be taken as the result of animal spirits—a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities.”*

Keynes criticizes the homoeconomicus and suggests that most of the economic phenomena are driven by rational frameworks. Further, he thinks that they are also affected by animal spirits which corresponds to irrational aspects of individuals and frequently overlooked by standard theories. The concept of ”animal spirits” coined by Keynes which refers to the instabilities of the capitalist economies and can be counted as one of the first steps of emotions’ involvement in the economy.

Later on, Simon (1955) introduced the *"Bounded Rationality"* approach which is one of the first criticisms of pure rationality. Bounded rationality approach attempts to model humans' economic decision making more realistically by relaxing the strict assumptions of the expected utility theory. In bounded rationality, people are rational but they are subject to some restrictions in decision making such as knowledge, cognitive limitations and a finite amount of time. Afterwards, Festinger (1957) proposed the *"Cognitive Dissonance Theory"*. According to this theory, when a new information that conflicts with the pre-existing understandings arrives, people often experience mental discomfort, which is called cognitive dissonance and, are motivated to reduce such dissonance. They have a tendency to consider only the information that confirms their existing beliefs and, ignore or modify the information that conflicts with those beliefs. Both Bounded Rationality and Cognitive Dissonance approaches paved the way for the development of many theories and many commonly used concepts in behavioural finance.

In 1979, Kahneman and Tversky have published their seminal papers, *"Prospect Theory: An Analysis of Decision Under Risk"* and behavioural finance literature is officially born. Prospect Theory is a critique of expected utility theory in behavioural finance. It provides a novel explanation to investors' decision-making behaviour under uncertainty.

According to Prospect Theory, people might exhibit risk-seeking and risk-averse behaviour depending on their prospects. When it comes to gains and losses, people will be more sensitive to losses. In other words, people give more weight to their losses than gains. This is called loss aversion. That is why the value function is steeper in the losses area than the gains area in Figure 1. People get utility which corresponds to value function from gains and losses, measured with respect to the reference point.

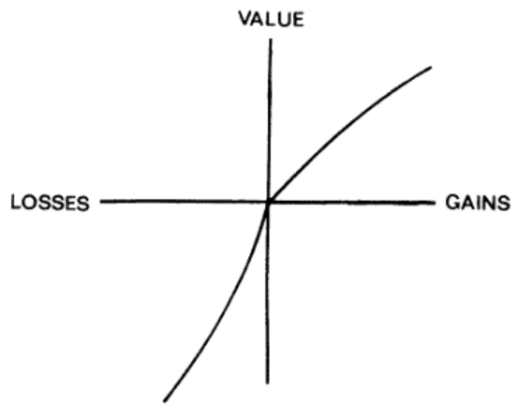

Figure 1: Value Function in Prospect Theory
Source: Kahneman and Tversky (1979)

Prospect Theory also suggests that people give more weight to the outcomes that are seen as definitive, than to outcomes which are perceived as likely to happen. This is referred as the *"certainty effect"*. Certainty effect prompts people to exhibit risk-averse behaviour towards the choices with gains while they exhibit risk-loving behaviour for those with sure loses.

The concepts mentioned so far constituted the cornerstone of behavioural finance and inspired many studies that contributed significantly to this literature.

1.2.3 Noise Trader Theory and Investor Sentiment

Noise Trader Theory is another cornerstone of behavioural finance. It is important in terms of officially introducing the presence of the impact of investors which are called noise traders who do not act with pure rationality in the market. While traditional models disregards noise traders and assumes the predominance of rational investors, Noise Trader Theory suggests that in addition to informed investors, there are also noise traders (i.e. liquidity traders, speculators, unsophisticated investors or uninformed investors) among the market participants.

The "noise trader" term takes its roots from the term *noise* which is first voiced by Kyle (1985) and Black (1986) in order to express the opposite concept of knowledge. In Noise Trader Theory, different kind of investors act with a different kind of information. While informed investors conduct information-based trading, noise traders trade on noisy signals, as if they are information (Black, 1986). In other words, they form their financial decisions with noise instead of using fundamentals. Noise is associated with the sources subject to trading activity except the information such as investors' feelings about the assets they invest in, following thoughts of financial gurus, following discussions in investor forums, sentiment that investors get from recently published news or social media posts (Shleifer and Summers, 1990; Alfano et al., 2015). One can infer from that the noise trader characterizes the irrational kind of investors who have mostly non-rational patterns in their behaviours. In order to explain this aspect of investors, Keynes (1936) referred to the term *"animal spirits"*, Kyle (1985) and Black (1986) referred to *"noise"*, more recent studies of De Long et al. (1990), Tetlock (2007) referred to *"investor sentiment"*.

According to the noise trader theory, investors are not entirely rational and they are subject to sentiment (Shleifer and Summers, 1990). Behavioural finance literature has comprehensively studied the investor sentiment and provided many definitions to this concept. These include *"irrational shifts in investor demand"* (Shleifer and Summers, 1990), *"the expectations of market participants relative to a norm: a bullish (bearish) investor expects returns to be above (below) average"* Brown and Cliff (2004), *"a belief about future cash flows and investment risks that is not justified by the facts at hand"* (Baker and Wurgler, 2006), *"level of noise traders' beliefs relative to Bayesian beliefs"* (Tetlock et al., 2008).

Investor sentiment plays a significant role in affecting price formation. A change in an investor's sentiment has an impact on the investor's decision-making process and his/her judgements about future actions. If investors have optimistic feelings about a financial instrument, they tend to buy or hold it longer. This lead a good signal for the other investors to demand more of this instrument which may cause price increases. On the other hand, if investors are pessimistic, they have tendency to sell the asset. This may generate a signal for other investors not to buy this asset and cause price decreases (Al Nasseri, 2016).

In financial markets, noise traders and arbitrageurs are in competition to set prices and expected returns (Baker and Wurgler, 2007). The actions of noise traders and their interactions with arbitrageurs have an essential role in determining the prices. Noise trading activities can deviate the prices from their fundamental values (De Long et al., 1990; Da et al., 2015). In such cases, assets may exhibit mispricing patterns due to change in sentiment of noise traders and limited arbitrage. Thus, markets can remain inefficient (Shleifer and Vishny, 1997; Baker and Wurgler, 2007).

Behavioural finance literature have attempted to investigate the impact of investor sentiment on market movements for a long time. As Baker and Wurgler (2007) pointed out, in order to accurately test whether investor sentiment causes significant movements in financial markets, it is crucial to find a proper proxy to express it:

> *Now, the question is no longer, as it was a few decades ago, whether investor sentiment affects stock prices, but rather how to measure investor sentiment and quantify its effects."*

Chapter 2

RELATED LITERATURE

2.1 MEASURING INVESTOR SENTIMENT

Investor sentiment can not be directly observed in the markets. It is an ambiguous concept and has not clear boundaries. Recent literature put a lot of effort to proxy investor sentiment in appropriate ways. In this part of the study, we gather these attempts in four categories as market-based proxies, market-independent proxies, survey-based proxies and web-based proxies. The following section presents a brief review of these measures.

2.1.1 Potential Investor Sentiment Measures

Subsequent efforts attempt to model investor sentiment with market movements which we call market-based measures. They include closed-end fund discount, mutual funds, the volatility index, the ratio of odd-lot sales to purchases, stock turnover, the share of equity issues in total equity and debt issues, the dividend premium, initial public offerings (IPO), put–call ratio etc. Some of these measures are not readily available for emerging markets. Moreover, these measures are usually reported at low frequency which can be evaluated as an obstacle to observe the short-term effects of investor sentiment. In addition, market-based measures cannot properly capture firm-level investor sentiment which has quite different characteristics (Seok et al., 2019). In order to overcome these shortcomings, researchers try to construct investor sentiment proxies using different measures.

Another kind of endeavours include expressing investor sentiment by using market independent factors. An interesting study of Hirshleifer and Shumway (2003) investigates the link between weather and stock returns. They show that returns are higher on sunnier days. Edmans et al. (2007) use sports results in order to capture investors' mood and find significant a market decline after soccer losses. Kamstra et al. (2003) associate investor mood with seasonal affective disorder and investigate its relationship between stock market returns. They report that the stock returns are significantly related to the amount of daylight through the fall and winter. Although these proxies are interesting in terms of the topics they investigate, they reduce investor sentiment to very specific events.

As for the third category of sentiment measures, some studies benefit from the surveys as a direct measure of investor sentiment. Some sources of survey data include Mexican Manufacturing Business Confidence Index (Bayram, 2017; Liston-Perez et al., 2018), Consumer Confidence Index (De Bondt, 1993; Bayram, 2017), Business Confidence Index (Antoniou et al., 2013; Bayram, 2017), American Association of Individual Investors' survey (Fisher and Statman, 2000; Brown and Cliff, 2004; Verma and Soydemir, 2006; Sayim et al., 2013) and Turkish Consumer Confidence Index (Sayim and Rahman, 2015). In survey-based proxies, investor sentiment is tried to be determined simply by using questionnaires. This approach is very laborious and expensive to carry out. Besides, surveys provide general information about investor sentiment rather than focusing on stock/currency/commodity specific characteristics.

The last category of sentiment proxies utilize web sources such as social media sharings, posts of blog and discussion board, financial news and search engine queries. One of the pioneering works relying on this approach and utilizing social media posts is conducted by Bollen et al. (2011). They use nearly one million tweets from Twitter and extract sentiments by using Opinion Finder and Google-Profile of Mood States (GPOMS) tools. Their aim is to capture the public mood and explain its relationship with changes in Dow Jones Industrial Average closing prices. Bollen et al. test this relationship considering both linear (with Granger Causality Test) and nonlinear (with Self-organizing Fuzzy Neural Network-SOFNN) characteristics. Results obtained with GPOMS are found to be more successful than those obtained with OpinionFinder. Considering both Granger Causality and SOFNN model, they reach the best results with "calm" component, which is extracted with GPOMS in terms of improving the accuracy of the predictions. Antweiler and Frank (2004) investigate the relationship between message board posts and market movements. In this regard, they use Yahoo! Finance and Raging Bull messages regarding the companies from Dow Jones Industrial Average and the Dow Jones Internet Index. Results show that message board posts help to predict both stock returns and volatility.

Based on the approach that investors now place more emphasis on peer opinions rather than financial professionals' opinions and this also plays an essential role in financial markets, Chen et al. (2013) investigate the impact of Seeking Alpha [1] articles which allow investors to share their opinions about financial instruments on the market. They find that sentiment extracted from Seeking Alpha articles and commentaries have predictive power on future stock returns and earnings surprises. Da et al. (2015) suggest a unique proxy based on Google Trends data. They construct the "Financial and Economic Attitudes Revealed by Search" (FEARS) index by aggregating the Google search queries for a list of 118 economic keywords indicating positive or negative sentiments selected from Harvard IV-4 Dictionary and the Lasswell Value Dictionary. Results validate FEARS' predictive power on short-term return reversals, temporary increases in volatility and mutual fund flows.

In our opinion, among all these investor sentiment proxies, news sentiment from web-based measures are the most appropriate way to express investor sentiment. To put it briefly, in today's world, thanks to the internet, information is spreading very quickly. Since the publishing frequency is very high, it is very convenient to observe the short term effects of investor sentiment. Moreover, thanks to the variety of resources, it is very easy to reach market/stock/currency/commodity level sentiments. Lastly, news media is the source that is followed by investors most. This makes news sentiment the fundamental proxy to express investor sentiment. The following section provides more detailed explanations about the underlying reasons for the suitability of news sentiment proxy.

2.1.2 News Sentiment

News media is one of the most important data sources that shape investors' decision-making process. Hundreds of news stories spread every day and upgrade investors judgements. It provides information about financial instruments and plays an prominent role in facilitating the diffusion of information. Thus, it continually updates the portrayed performance of the regarding instrument for investors.

News media provides real-time information. It can be obtained using many channels including newspapers, newswires, social media. These sources publish the news as soon as the events occured or publicly shared (Schumaker and Chen, 2009). By utilizing these

[1] Seeking Alpha is one of the most popular investment based social platform in the US providing valuable financial articles.

sources, investors can upgrade their trading process in time.

Financial news not only provides information about the current situation of the market, but also shapes investor sentiment. The information extracted from these news stories forms the basis of investors' sentiment about the relevant instrument (Allen et al., 2019). Market participants digest the news immediately, revise and rebalance their positions accordingly (Mitra and Mitra, 2012). Then, market aggregates the actions resulting from investor sentiment and reflect them in the price movements (Li et al., 2014). By this channel news sentiment can lead to significant market movements.

Van de Kauter et al. (2015) argues that news media have a substantial role in creating market noise which can be considered as the main source of instability of the markets. It broadens the characteristics of information with emotions thanks to its content or presentation styles. At this stage, the characteristics of information are transformed into something that affects the perception of the investor. Thus, beyond the abstract perspective of efficient markets, news media is a concept that makes investors irrational rather than an impartial information source for investors.

Today, natural language processing tools allow the news stories to be automatically understood and interpreted by computers. By utilizing rule-based or machine learning-oriented approaches, they can easily be translated from natural language to numerical data. Sentiment analysis is the task of natural language processing methods. It aims to detect positive, negative and neutral characteristics from given texts (Liu, 2015). There are two main approaches that can be adopted in order to extract sentiment in texts as machine learning oriented and rule-based methods. In machine learning-based approaches, text contents are classified according to their sentiment characteristics with the help of machine learning algorithms, while rule-based methods are based on the use of predetermined dictionaries formed with polarity values or features determined according to the semantic orientation of words (Chopra et al., 2016).

With the increasing credibility of natural language processing methods, well-known financial data vendors such as Bloomberg and Thomson Reuters have started to take the advantage of these tools. Now, they present a wide variety of materials including the sentiment measures of financial news and social media posts to their users (Feuerriegel and Prendinger, 2016). For example, Bloomberg's news sentiment scores are obtained through a automatic system which is trained to imitate a human in processing financial news, and they are presented to their users within a few milliseconds (Li et al., 2014).

The concept of news sentiment did not get enough attention in finance until Tetlock (2007) published his pioneering paper. In this work, Tetlock extracts a pessimism factor, namely low news sentiment from Wall Street Journal's Abreast of Market column and investigate its relationship with price movements. He adopts a rule-based sentiment analysis approach and utilizes the General Inquirer's Harvard-IV-4 classification dictionary. He finds that high pessimism factor predicts the downward pressure on market prices and unusually high or low pessimism predicts high trading volume. Tetlock's work is very important for the news sentiment literature since it is one of the first studies to show the media's substantial effect on market movements. Following this pioneering study, news sentiment and its applications to finance have begun to be considered by both academics and practitioners. Now, there is a growing body of literature showing the influence of news sentiment on market movements.

Similar to Tetlock (2007), Ferguson et al.'s (2015) work is one of the studies that adopt rule based sentiment analysis approach. They utilize Loughran and McDonald (2011) words list to examine the tone and volume of firm-specific news providing a valuable information about future stock returns. They find that both tone and volume of news have the power to predict future abnormal returns. More specifically, they show that positive words forecast higher returns, while negative words forecast lower returns for the next trading period. They also conclude that the impact of the volume is greater than the tone.

The performance of rule-based methods closely relates with the efficiency of the dictionary and its compatibility with the domain. Some dictionaries that are used in rule-based sentiment analysis approach do not focus on a specific domain. For example, words such as 'tax' and 'liability' are on the negative word list of the General Inquirer dictionary, but these words may not indicate any negativity in the financial context. In order to overcome this problem researchers construct domain-specific dictionaries. For example Henry (2008) and Loughran and McDonald (2011) dictionaries are created specifically for finance domain (Kearney and Liu, 2014).

In the machine learning based sentiment analysis approach, data should be labelled manually or automatically. Koppel and Shtrimberg (2006) introduce a novel methodology which relies on an idea of labelling financial news according to price changes of stocks. They reach 70 percent accuracy on average in distinguishing good news from bad news.

Following Koppel and Shtrimberg, Généreux et al. (2011) also adopt automatic labelling approach based on contemporaneous price changes and obtain around 70 percent accuracy.

Some studies have preferred to use sentiment data provided by reliable financial data vendors, instead of extracting sentiment scores from financial news using machine learning or rule-based methods. For example, Vanstone et al. (2019) and Dunham and Garcia (2020) utilize Bloomberg's sentiment scores in order to proxy investor sentiment. Vanstone et al. (2019) construct neural network autoregressive models to evaluate the predictive power of both sentiment obtained from news stories and Twitter posts. They utilize Root Mean Square Error (RMSE) as performance evaluation metric and find that sentiment measures can significantly improve the performance of stock price predictions. Dunham and Garcia (2020) examine the impact of firm-level news and social media sentiment on companies' share liquidity. They find a negative relationship between the firm-level sentiment extracted from Twitter posts and share liquidity and a positive link between the news sentiment and share liquidity. This implies, improvements in investor sentiment extracted only from Twitter cause a decrease in the firm's average share liquidity. Ho et al. (2013) investigate the relationship between news sentiment (firm-level and macroeconomic) and intraday volatility by using Fractionally Integrated GARCH and two-state Markov Regime-Switching GARCH models for the sample of Dow Jones Composite Average. They reveal a significant impact of news sentiment on volatility persistence for most of the DJN 65 stocks. Moreover, they validate a greater impact of firm-specific news sentiment (compared to macroeconomic news sentiment) and negative news sentiment (compared to positive news sentiment).

2.2 FINANCIAL FORECASTING USING ARTIFICIAL INTELLIGENCE BASED MODELS

In today's increasingly competitive markets, in order to gain competitive advantage against the other market participants, investors need to make complex financial decisions by analysing the big and rapidly accumulating data. Given the dramatically growing datasets, now, investors' decision-making process requires more sophisticated analyses with more advanced methods.

When an individual intends to invest in a financial instrument, he/she should have an optimistic view of this instrument's future. In this step, investor tries to make an appropriate prediction for the following periods. However, making accurate forecasting is an appealing but very challenging task. For example, there are various parameters affecting stock prices and some of these parameters may not be considered while constructing forecasting models. Moreover, the financial system has a dynamic structure and parameters that involved in this dynamic system are continually changing and transforming. Another reason stems from the inherent in financial time series which are "dynamic, chaotic, noisy, nonlinear" and have a nonstationary structure (Henrique et al., 2019) where catching their pattern is quite difficult. In this context, the introduction of artificial intelligence applications in the finance has revolutionized financial forecasting. Given their capabilities in capturing complex patterns and strong capacity to handle with noisy and time-dependent data, artificial intelligence-based models are promising and exhibit superior performance in forecasting stock movements.

There are two main approaches in financial forecasting literature. First one is estimating the direction of the movement by solving a classification problem in which the prediction target represent whether the movement is upwards or downwards for the determined period. The second approach is predicting a continuous target in which the aim is directly forecasting the close price or return of the stock by attempting to solve a regression problem. In predicting the direction of the movement approach, the task is converting the estimated up/down movements into simply buy/hold/sell signals, rather than presenting accurate predictions about the value of stock prices or returns. For example, while direction forecasting can provide information about whether the stock price will increase or decrease, yet, it does not give an idea of the amount of change. On the other hand, the prediction of a continuous target provides more accurate information by directly estimating the value of the target (Weng et al., 2018). Moreover, point estimates can be easily converted to direction estimates.

Gudelek et al. (2017) utilize 2-D Convolutional Neural Network to predict the stock prices of most commonly traded ETFs in the New York Stock Exchange. They use technical analysis indicators such as relative strength index (RSI), simple moving average (SMA), the moving average convergence/divergence oscillator (MACD), the Williams R%, the stochastic oscillator as input of their forecasting models. They evaluate the model on both classification and regression tasks using 2 class regression, 3 class regression, 2 class classification and 3 class classification models. They reach 72% accuracy with 2 class regression in which target values that are greater than 0 are considered as "buy" signals and values that are less than or equal to 0 are considered as "sell" signals. Moreover, this model yields more profit even after commissions are considered. Liu et al. (2017)

make use of a hybrid model which consists of CNN and LSTM networks to make stock selection and timing decisions. Then, they establish a trading strategy based on the results of the forecasting model. Their trading strategy based on the idea that (i) if forecasting model predicts +1, then they buy and hold this stock for 5 days, if a position has already been taken, then update the number of holding days as 5 and continue to hold, (ii) if forecasting model predicts -1, if a short position has already been taken or stocks are already held, then the number of holding days will be decreased by one, and if the number of holding days is equal to 0, then sell the stock. They conclude that the proposed system outperforms to benchmark index in terms of the net portfolio value, annualized rate of return and maximum retracement.

Chung and Shin (2018) adopt Genetic Algorithm-Long Short Term Memory (GA-LSTM) approach by making use of simple moving average, weighted moving average, RSI, Stochastic K% and D% to predict KOSPI Index. This approach rests on the idea of searching best hyper parameters for LSTM model by using genetic algorithm. The prediction performance of GA-LSTM approach is compared with benchmark model that predicts no day-to-day change. They show that the proposed approach has lower MSE, MAE, and MAPE the compared to benchmark model. This implies that using genetic algorithm in hyper parameter optimization can improve the performance of the forecasting model. Zhang and Tan (2018) focus on stock selection problem and propose DeepStockRanker approach to predict stocks' future return ranking. In this approach, first, a RNN model with LSTM cell is fed with data to obtain deep representation of time series information, then a feedforward neural network is connected with the RNN and convert the deep representation of time series information to stock ranking score. Finally, they construct a trading strategy that consists of the stock selection process based on the outputs that are produced by the DeepStockRanker approach. They compare results of DeepStockRanker approach several models including SVR, SVR with technical indicators, RBM, MLP and LSTM with technical indicators and conclude that the proposed approach outperforms to all methods in terms of the information coefficient, active return and information ratio.

Ince and Trafalis (2008) use Support Vector Regression (SVR), Multi Layer Perceptron (MLP) and Autoregressive Integrated Moving Average (ARIMA) to predict the short term price movements of selected 10 companies traded in NASDAQ by using technical indicators including exponential moving average, RSI and Bollinger bands, MACD. They evaluate the forecasting models according to the results of the trading system that they established using the predictions obtained from these models and RMSE. Their results show that SVR outperforms MLP and ARIMA in terms of RMSE and the risk-adjusted performance metric. Nikou et al. (2019) utilize Artificial Neural Network, SVR, Random Forest (RF), and LSTM to forecast the close price of iShares MSCI United Kingdom. They show that LSTM and SVR outperforms other models, in terms of MAE, MSE and RMSE. Maqsood et al. (2020) use Linear Regression, SVR and CNN to forecast stock prices with a feature set including raw price and investor sentiment data for the sample of selected companies from US, Hong Kong, Turkey and Pakistan. The main aim of the paper is to show the predictive performance of locally and globally important events in stock price prediction. In order to proxy these events, they utilize Twitter posts. Results show that public sentiment improves forecasting performance and should be considered in stock prediction. Moreover, compared to global events, local events have more potential to affect the performance of prediction models. Lastly, Support Vector Regression outperforms other models according to performance metrics such as RMSE and MAE.

Considering the financial forecasting literature [2], one can see that LSTM and its variations along with some hybrid models dominate this domain. (Sezer et al., 2020) LSTM is one of the most advanced deep learning models for sequence learning tasks including natural language processing, handwriting recognition and time series prediction. It is inherently suitable especially for financial forecasting and exhibits superior performance in this field thanks to its internal feedback mechanism between neurons that allows to memorization of important past information (Chung and Shin, 2018).

On the other hand, SVR is another model that exhibit superior performance in time series forecasting. SVR is robust to the outliers, has high generalization capacity, very easy to implement and can give good results with high accuracy based on a sparse subset of the whole training set (Sundaram et al., 2006; Awad and Khanna, 2015). Even though SVR has shallow structure compared to deep learning models, in financial forecasting literature, many studies are showing its superior performance against the models with deep architectures. (Ince and Trafalis, 2008; Maqsood et al., 2020)

In the financial forecasting literature, most of the studies focus on the developed financial markets. There are very limited number of studies for emerging markets, especially for Borsa Istanbul. Firstly, many studies for the sample of Borsa Istanbul focus on the direction of the market index rather than examine the stocks individually. In line with the main objectives of portfolio management, investors intend to diversify their portfolio to

[2]See Gandhmal and Kumar (2019), Henrique et al. (2019), Ozbayoglu et al. (2020) and Sezer et al. (2020) surveys for more detailed information on the use of artificial intelligence based approaches in financial applications.

get maximum profit at the same time exposing minimum risk. Although investing in index funds is an option for this purpose, investors may prefer to diversify their portfolios based on their own decisions. In this case, it is necessary to examine the stocks individually. Secondly, considering the stock price forecasting literature, most of the studies do not produce practical applications including designing trading strategies, generating transaction signals etc. Lastly, their feature set mostly contain technical analysis indicators and do not consider non-rational components of the market except for a few studies.

Among the studies that predict the direction of BIST index, Kara et al. (2011) use technical analysis indicators including simple 10 day moving average, weighted 10 day moving average, momentum, stochastic K%, stochastic D%, RSI, MACD, Larry William's R%, A/D Oscillator, CCI to predict the direction of BIST100 index. They utilize Artificial Neural Betworks and Support Vector Machines and compare the predictive performances of these models. They reach 75% accuracy with ANN and 71% accuracy with SVM. Gunduz and Cataltepe (2015) use financial news (including internet and official news) and historical price data to predict the direction of the BIST100 Index. They state that the use of news articles is a possible alternative to historical prices for predicting the direction of BIST100 Index. Moreover, they show that internet news exhibit better performance than the official news. Similarly, Sakarya et al. (2015), Halil and Demirci (2019) focus on predicting the direction of BIST index.

Gunduz et al. (2017) utilize 25 technical indicators including EMA, momentum, MACD, RSI to predict the intraday price direction of 100 stocks from Borsa Istanbul. They use correlations between instances and features to order the features before they are fed into CNN as inputs. They compare the proposed approach using CNN with randomly ordered features and Logistic Regression and show the superior performance of proposed methodology. Gunduz et al. (2017) utilize technical indicators and dollar-gold prices to predict the daily movement of selected three stocks (GARAN, THYAO and ISCTR) from Borsa Istanbul using Deep Neural Networks. They state that using dollar-gold features as input improve the prediction performance. Özçalıcı and Ayrıçay's (2016) investigation is also among the studies that focus on predicting stocks individually. Although these studies focus on the close price predictions of Borsa Istanbul stocks individually, they do not offer any financial implications based on these results.

To the best of our knowledge, this is the first study predicting close prices of the Borsa Istanbul stocks individually by making use of LSTM and SVR models. Apart from the existing studies that do not produce practical implications from prediction results or making this for BIST index, this study is the first attempt to construct trading strategies based on the predictions for BIST stocks. Although there are studies using news sentiment as a proxy for investor sentiment, this is the first endeavour that proxies investor sentiment utilizing Bloomberg's news sentiment data which is developed to imitate a human in processing financial news for the sample of Borsa Istanbul stocks.

Chapter 3

DATA

In this section, we briefly introduce the dataset that we used in this thesis. When the empirical literature is overviewed, one can see what the studies attempt to do is simply to detect the most successful factors in predicting stock price movements. These endeavours include using fundamental analysis and technical analysis indicators as well as modelling behavioural patterns.

Following previous studies, we have created a wide feature set including historical price data, technical analysis indicators and news sentiment as can be seen in Table 1. Our aim is to identify the best features that improve the prediction performance of our model and utilize them. In this respect, we run several experiments by adding or removing some of these features and reach best results in terms of MSE and MAE with the feature set which is marked in Table 1:

Feature	Is it selected?	Feature	Is it selected?	Feature	Is it selected?	Feature	Is it selected?
High	x	EMA 30		WMA 40		Bollinger Bands	x
Low	x	EMA 40		WMA 50	x	CCI	
Open	x	EMA 50	x	WMA 60		Momentum	x
Close	x	EMA 60		Return	x	ROC	
Volume	x	SMA 10	x	Risk 10		RSI	x
Adjusted Close		SMA 20		Risk 20		Stochastic K%	x
Positive Sentiment	x	SMA 30		Risk 30		Stochastic D%	x
Negative Sentiment	x	SMA 40		Risk 40		William's R%	
Net Sentiment		SMA 50	x	Risk 50		AD	
EMA 10	x	SMA 60		Risk 60		ADOSC	
EMA 12		WMA 10	x	MACD	x	OBV	
EMA 20		WMA 20		MACD Signal	x		
EMA 26		WMA 30		MACD Histogram			

Table 1: Selected Feature Set Based on the Results of the Initial Experiments

According to Aktan et al. (2018), emerging economies are less efficient compared to developed markets and thus in such markets stock forecasting methods have stronger power. Moreover, Brzeszczyński et al. (2015) argue that emerging markets are more exposed to investor sentiment. In such markets, the number of individual investors is relatively high among the market participants and they suffer from lack of high-quality information and professional services. Therefore, these markets tend to be affected by investor sentiment more.

Following these arguments, we choose an emerging stock market as our sample: Borsa Istanbul. Then we randomly select 10 of 45 stocks included in the BIST30 index which are among to the most capitalized and the most actively traded stocks in the past few years: Akbank (AKBNK), Doğan Holding (DOHOL), Ereğli Demir ve Çelik Fabrikaları (EREGL), Türkiye Garanti Bankası (GARAN), Türkiye İş Bankası (ISCTR), Sabancı Holding (SAHOL), Soda Sanayii (SODA), Türkiye Sınai Kalkınma Bankası (TSKB), Türkiye Vakıflar Bankası (VAKBN), Yapı ve Kredi Bankası (YKBNK).[3] This selection is particularly based on the view that investors closely follow such most popular stocks' stories.

Our timespan covers the period between 11 May 2015 and 10 March 2020 due to news sentiment data availability. We select 2-Year Bond Yield of Turkey as the risk-free rate in calculating Sharpe Ratio. We utilize Yahoo! Finance for historical stock price data, Bloomberg Terminal for news sentiment data and Investing.com for 2-Year Bond Yield data of Turkey. We execute all implementations with Python 3.7. We use Numpy and Pandas libraries for data obtaining, pre-processing, backtest simulation and performance evaluation; Ta-Lib for calculating technical analysis indicators; Matplotlib for data visualization; Keras for developing LSTM networks and; Sci-kit learn for constructing SVR models.

3.1 FINANCIAL DATA

Our financial data consists of raw price data and technical analysis indicators. We obtain historical stock price data from Yahoo! Finance using Python's pandas data reader module. This data contains Open, Close, High, Low, Volume columns. All technical analysis indicators are computed using raw price data. At this stage we utilize Python's TA-Lib which is one of the most commonly used libraries by trading software developers. TA-Lib library includes 200 indicators such as moving averages, William's %R, MACD, RSI, CCI, Stochastic K%, Stochastic D% etc. Among them we select the following indicat-

[3] In this study, we assign to IDs these stocks as 1, 7, 11, 13, 17, 26, 29, 36, 41, 42, respectively.

Data	Description
Raw price data	Raw price data contains Open, Close, High, Low and Volume columns.
Simple Moving Average (SMA)	SMA is the simple mean of stock's price over a certain period of time. SMA smooths the trend of stock price and filters out the noise from the stock price. In extremely volatile conditions, it may not be useful to apply SMA.
Weighted Moving Average (WMA)	WMA is the weighted mean of stock's price over a certain period of time. It gives more weight to more recent stock prices.
Exponential Moving Average (EMA)	EMA is the exponential mean of stock's price over a certain period of time. Similar to WMA, EMA gives more weight to more recent stock prices. In EMA, weighting decreases exponentially with each previous period. EMA considers not only the changes in certain time periods, but also the price changes over the entire period.
Moving Average Convergence Divergence (MACD) and MACD Signal	MACD is commonly calculated by subtracting the 26-period EMA from the 12-period EMA. MACD generates buy-sell signals with the help of signal line which is calculated taking a n-period exponential moving average of the MACD line and histogram, which is calculated by subtracting MACD signal from MACD line. There are many ways to use the MACD such as crossovers, overbought/oversold conditions, divergences.
Relative Strength Index (RSI)	RSI is a momentum indicator that evaluates overbought or oversold conditions. It takes values between 0 and 100. It is mostly used with reference values of 30 and 70. The values of 30 or below indicate that the stock is in oversold area, while the values of 70 or above indicate that the stock is overvalued or is in the overbought area.
Stochastic K% and Stochastic D%	Similar to RSI, Stochastic oscillator is also used to evaluate overbought and oversold conditions and generate trading signals. It takes values 0-100 bounded range of values. The Stochastic oscillator is displayed as two lines which are %K and %D. Transaction signals are generated when the %K crosses through %D.
Bollinger Bands	Bollinger Bands consist of three lines: The middle band which is n period simple moving average of the stock prices for the given period of time, the upper and lower bands are placed two standard deviations above and below the middle band. When the volatility is high (low), then the bands widen (narrow).
Momentum	Momentum is measured by taking price differences for the given period. It is a measurement of the speed of the price changes.
Return	Return is the price change between current price and the previous period's price.

Table 2: Selected Features for Financial Data and Their Descriptions

ors according to the results of our initial experiments: simple moving average, weighted moving average, exponential moving average, return, MACD, MACD signal, momentum, RSI, Stochastic K%, Stochastic D%, Bollinger Bands. Table 2 shows the financial variables and their descriptions[4] that are used in this thesis.

3.2 NEWS SENTIMENT DATA

Our news sentiment data contains positive and negative news sentiment scores. We obtain this data from Bloomberg which is one of the most known financial data vendors. Bloomberg utilizes artificial intelligence-based models to extract the sentiment of the

[4]One can find more detailed information about technical analysis indicators and strategies in Pring (1980), Murphy (1999), Achelis (2001).

financial news. These models are trained to imitate a human in processing financial news. Before the model is trained, financial news data should be labelled. In this stage Bloomberg adopts the idea that: *"When the investor who holds an asset in his/her portfolio reads the news about pertinent asset, what will his/her attitude be this the stock? Will it be bullish, bearish or neutral?"* (Cui et al., 2016) Then, the labelled data is trained on artificial intelligence-based models such as Support Vector Machines. Once the model is trained, when new information arrives, the model will automatically determine the positive, negative or neutral characteristics of each financial news.

Bloomberg presents two types of sentiment scores to their users which are (i) story level and (ii) company level:

- Story-level sentiment is produced in real-time since the arrival of news. It consists of score and confidence. Score is a categorical value that takes values +1, 0 and -1 which demonstrates positive, neutral and negative sentiments respectively. On the other hand, confidence is a numerical value that takes values ranging between 0 to 100, which can be seen as the probability of being positive, negative and neutral of the news stories.

- Company-level daily sentiment is the confidence-weighted average of the past 24 hours' story-level sentiment. They are published every morning approximately 10 minutes before the market opens. For intraday frequency, company-level sentiment is recomputed every two minutes with an eight-hour rolling window. It is delivered as a score that ranges from -1 to +1, indicating negative level and positive level.

Chapter 4

METHODOLOGY

In order to forecast stock prices of the next trading day, we utilize artificial intelligence-based models including Long Short-Term Memory (LSTM) Networks and Support Vector Regression (SVR). In this section, we first examine neural networks in order to provide a better understanding for LSTM networks. Then, we introduce LSTM and SVR models. Each part includes both the examination of the models and the implementations including data pre-processing and hyperparameter optimization. Finally, we design trading strategies in order to practically make use of the predictions produced by the forecasting models.

4.1 FORECASTING STOCK PRICES

4.1.1 Long Short Term Memory (LSTM)

Long short-term memory (LSTM) is one of the most advanced deep learning architecture for sequence learning tasks as a type of Recurrent Neural Networks (RNNs). It was introduced by Hochreiter and Schmidhuber (1997). Given its capabilities for learning complex sequences, it is commonly used for advanced applications such as grammar learning, handwriting recognition, robot control, rhythm learning and time series prediction etc. Before examining LSTM in detail, it is useful to explain the basic concepts of neural networks. The descriptions of DNNs and LSTM networks follow Britz (2015), Karpathy (2015), Olah (2015), Fischer and Krauss (2018), Lærum (2018) and Güdelek (2019).

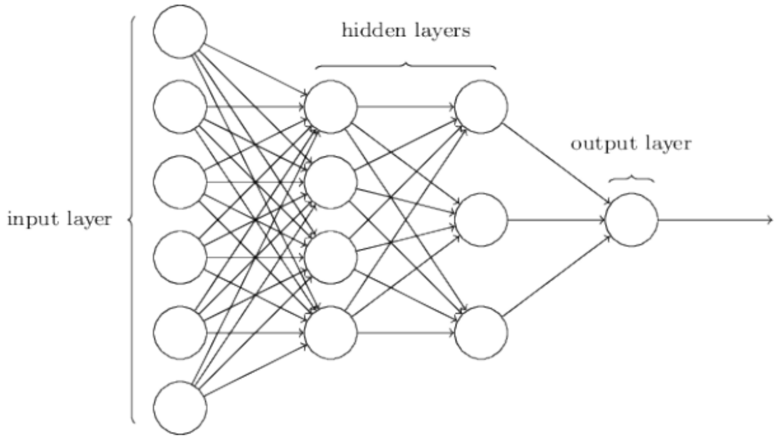

Figure 2: An Illustration of Deep Neural Networks
Source: Nielsen (2015)

Neural Networks (NNs) were first introduced by McCulloch and Pitts (1943). They are very good at defining linear and nonlinear relationships between inputs and outputs in a given dataset. Neural Networks are developed with the inspiration from the functioning of neurons in our brain.

Figure 2 demonstrates a deep neural network model which includes three sets of layers namely input layer, hidden layer(s) and output layer. Each layer consists of a given number of neurons. Also, there are weights w which are inter neuron connection strengths and biases b that are assigned to neurons that have an important role in the functioning of neural networks.

The inputs coming from the input layers, then pass through the activation function and reach the first hidden layer. Each neuron in the first hidden layer calculates the output and passes it through the activation function and then transfers output to the next hidden layer. This process continues until the output layer is reached. The process that goes from input layer to the output is called forward propagation.

The neuron assigns weights to each one of the inputs. The inputs are multiplied by the weights and summed up with the bias. The result of the calculations feeds into an activation function to obtain the output values.

$$z = w^T * x + b \tag{4.1}$$

$$y = Activation function(z) = \sigma(z) \tag{4.2}$$

Where, w is the weight vector, x is the input vector, b is the bias, z is the given value of neuron, σ is the activation function and y is the value obtained from the activation function.

Activation function determines whether the output of the neuron will be activated. Moreover, activation functions compress the outputs to a certain range of values. For instance, $(0, +1)$ for sigmoid, $(-1, +1)$ for tanh, $(0, +)$ for Rectified Linear Units (ReLU). The logic of using the activation function during this process is to prevent the output values from going to the infinity and to provide nonlinearity. Figure 3 shows the most commonly used activation functions.

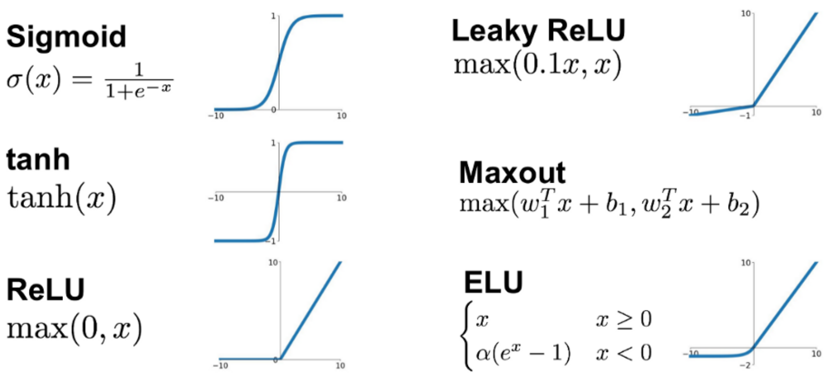

Figure 3: Most Commonly Used Activation Functions in NNs
Source: Jain (2019)

In order to evaluate the performance of the model, it is important to select the appropriate loss (i.e. cost) function which calculates how close the produced outputs is to actual outputs. There are several loss functions that are used for this purpose. Mean Squared Error (MSE) and Mean Absolute Error (MAE) are the most common ones that are used for regression problems. They are calculated as follows:

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{4.3}$$

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{4.4}$$

The training process is all about setting these weights and biases according to input and output values. Once NNs are trained, the system produces outputs according to given inputs. Then it compares the produced outputs with actual outputs with respect to loss function. The main goal is to minimize this error. To ensure this, from the last layer to the first layer, the weights and biases are adjusted. This process is executed based on backward propagation and repeated until the error is minimized.

The optimization process mentioned above is executed with the gradient descent algorithm. The gradient shows the maximum increase direction of the function at that point and is calculated by taking partial derivatives of the function. Consider that the target function is our loss function $L(W)$ which is parameterized by the model's weights. $\nabla_W L(W)$ is expressed as the gradient of this function and is a derivative vector in d dimension. Gradient descent tries to minimize the objective function by updating the weights in the opposite direction of the gradients. By using the chain rule, gradients are calculated as follows:

$$\nabla w L(W) = \langle \frac{\partial(W)}{\partial W_1}, \frac{\partial(W)}{\partial W_2}, ..., \frac{\partial(W)}{\partial W_{d-1}}, \frac{\partial(W)}{\partial W_d} \rangle \tag{4.5}$$

Gradients contribute the training process by updating weights:

$$W = W - \eta . \nabla_W L(W) \tag{4.6}$$

Here, η is the learning rate parameter that determines the size of the steps is taken to reach the valley. In other words, learning rate determines the learning speed of the model. A high learning rate allows the model to learn faster while low learning rate can cause the process to take place more slowly. Considering the extreme levels, the extremely large learning rate will cause the gradient descent algorithm to oscillate around the global minima and perhaps never reach that point as can be seen in Figure 4. On the other hand, the extremely low learning rate will cause the steps to be so low that the global minima will never be reached. Accordingly, it should be determined very carefully. Goodfellow et al. (2016) point out the importance of learning rate as follows: *"The learning rate is perhaps the most important hyperparameter. If you have time to tune only one hyperparameter, tune the learning rate."*

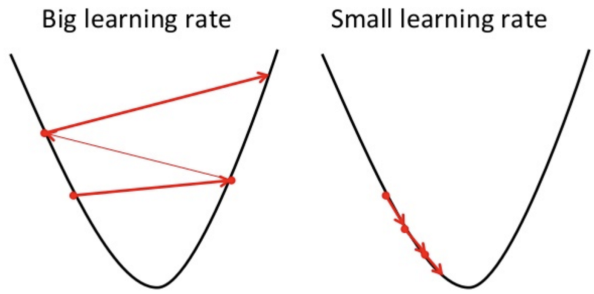

Figure 4: An Illustration to the Difference Between Big and Small Learning Rates
Source: Prasad (2018)

Figure 5 graphically shows the functioning of the gradient descent algorithm. Gradient descent selects an initial point and then takes a step forward in the steepest downhill direction. This process continues until algorithm reaches the global minima.

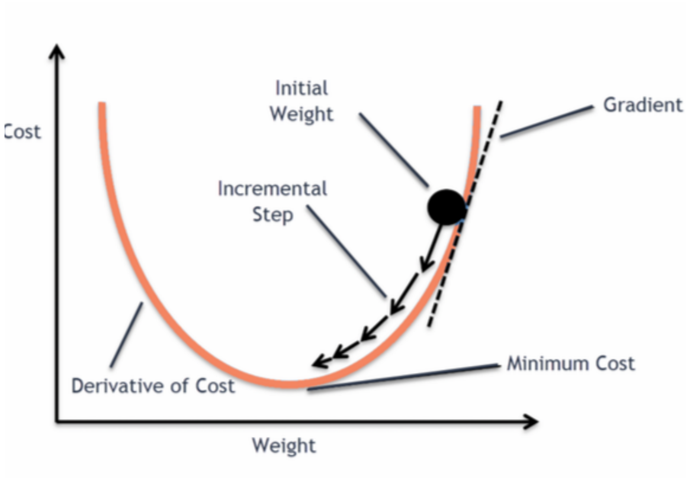

Figure 5: An Illustration of the Gradient Descent Algorithm
Source: Dawar (2020)

In order to reduce the loss, optimizers determine the neural network's attributes including weights and learning rate according to the response to the output of the loss function. At this point, loss function undertakes the task that guide the optimizer whether it move in the right or wrong direction. Adaptive Moment Estimation (Adam), Adaptive Gradient Algorithm (Adagrad), Root Mean Square Propagation (RMSProp), Stochastic Gradient Descent (SGD) are the most commonly used optimizers in neural networks. Using momentum in the optimization algorithm ensures that learning continues in cases where the gradient is vanished. In such cases, while calculating the gradient at time $t + 1$, learning is ensured by using momentum with the effect of the gradient at time t. Momentum parameter takes a value between 0 and 1.

Most of the NNs can only learn spatial patterns from given data (Kim and Shin, 2007; Chung and Shin, 2018). Moreover, in traditional neural networks, all inputs and outputs are treated as they are independent of each other, and the order of inputs to the network does not matter. The neural network forgets this input after the output is produced. However, this approach does not suit very well for all tasks. For example, in processing sequential tasks such as handwriting, speech recognition, financial forecasting etc., inputs are related to each other. For example, in a financial time series problem, the previous data can be important in predicting the next day's closing prices. The advantages of RNN over these models are emerged at this point. RNN has a memory feature that can be used to extract temporal patterns in the data. In a simple expression, RNNs are the neural networks with memory. In RNNs, there are internal feedback mechanisms between the neurons that allow the memorization of significant events. Figure 6 and Figure 7 show the functioning of simple RNN.

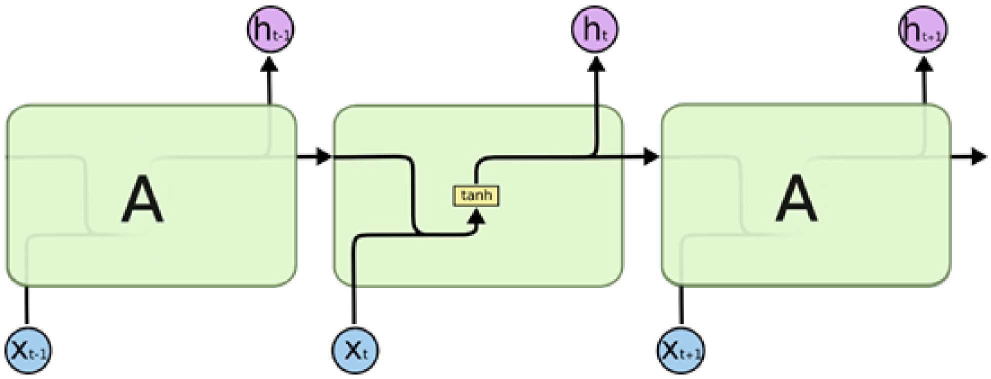

Figure 6: An Illustration of RNN
Source: Olah (2015)

In Figure 8, x_t is the input, s_t is the hidden state namely, the memory of the network, o_t is the output at the time step t, U and W are the weight vectors and f is the activation function which provides nonlinearity. In RNNs, mostly sigmoid and tanh activation functions are preferred. Hidden state is calculated based on the output of the previous hidden states and the input of the current timestep.

$$s_t = f(Ux_t + Ws_{t-1}) \tag{4.7}$$

RNN is a recurrent model since it has loops that allows the use of historical information by executing the same task for each element of the sequence with the help of previous outputs. In RNN, once the output is produced, it is copied and returned back into the network. Thus, the network considers both the current input and the output that learns from the previous timesteps. Theoretically, RNN can utilize the information in long sequences, but in practice they have some limitations and can look back only a few periods. (Britz, 2015) In RNN, while time steps increases, the information coming from the past events disappears. Thus, RNN has difficulty in capturing long time-dependencies due to vanishing gradient problem.

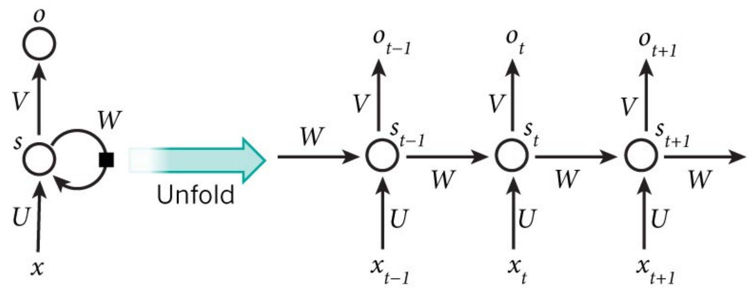

Figure 7: An Illustration of Unfolded RNN
Source: LeCun et al. (2015)

RNN handles with sequence data and has timesteps. Therefore, it uses back propagation through time (BPTT) algorithm in adjusting model parameters. BPTT unrolls all input timesteps, accumulates the errors across each timestep, then rolls up the timesteps and updates the weights. One can remember from the above-examined back propagation example that the gradients carry information to update the model's parameters. If the gradient approaches very small values, this implies that parameter updates become insignificant from layer to layer. Thus, initial layers which are very critical to the functioning of the whole system thanks to their substantial role in recognizing the essential characteristics of the input data cannot be efficaciously updated. This is called vanishing gradient problem.

There are several ways to deal with the vanishing gradient problem such as accurately initialization of the weights, regularization, using ReLU which does not cause a small derivative as an activation function instead of using sigmoid or tanh, etc. One another and even more popular solution to handle such problem is to use Long Short-Term Memory networks. As a class of RNNs LSTM is developed to overcome and address the drawbacks of simple RNNs.

Following notations are used in order to describe the functioning of LSTM networks (Fischer and Krauss, 2018):

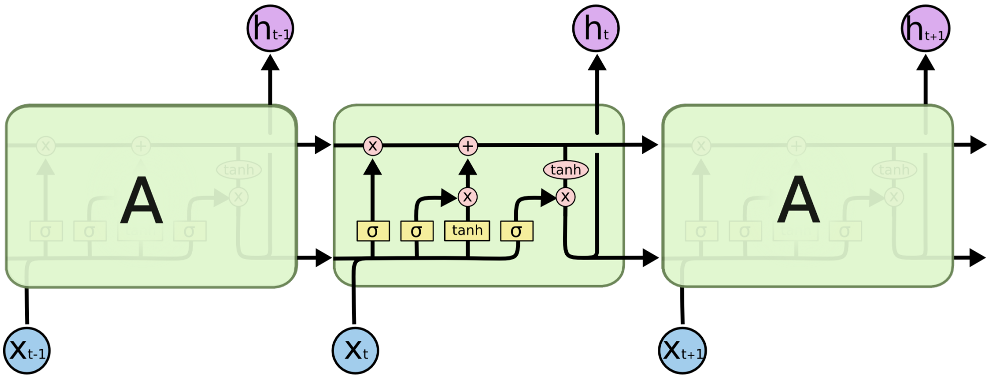

Figure 8: An Illustration of LSTM
Source: Olah (2015)

- b_f, $b_{\tilde{s}}$, b_i and b_o are the bias vectors,
- f_t, i_t and o_t are the vectors representing the values of the regarding gates,
- h_t is the output vector of the LSTM layer,
- s_t and $\tilde{s}_t$ are the vectors for cell states and candidate values,
- $W_{f,h}$, $W_{f,x}$, $W_{i,h}$, $W_{i,x}$, $W_{o,h}$, $W_{o,x}$, $W_{\tilde{s},x}$ and $W_{\tilde{s},h}$ are the weight matrices,
- x_t is the input vector,
- $\circ$ represents the elementwise product.

The functioning of the memory cell starts with determining how much information the cell state will forget by using the input at time t and the hidden state at time $t-1$. This determination is executed by the forget gate.

$$f_t = (W_{f,x}\, x_t + W_{f,h}\, h_{t-1} + b_f) \tag{4.8}$$

In the next stage, the input gate decides which information will be added to the cell state. In this regard, candidate values that could potentially be added is calculated by using the input at time t and the hidden state at time $t-1$.

$$\tilde{s}_t = tanh(W_{\tilde{s},x}\, x_t + W_{\tilde{s},h}\, h_{t-1} + b_{\tilde{s}}) \quad (4.9)$$

Then, output of the input gates is calculated.

$$i_t = sigmoid(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i) \tag{4.10}$$

In the next stage, the new cell states are computed based on the results of the previous steps.

$$s_t = f_t \circ s_{t-1} + i_t \circ \tilde{s}_t \tag{4.11}$$

In the last stage, the output of the memory cells is produced:

$$o_t = sigmoid(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o) \tag{4.12}$$

$$h_t = o_t \circ tanh(s_t) \tag{4.13}$$

Our dataset covers the period between 11 May 2015 and 10 March 2020. We split our dataset as train and test datasets according to approximately 7:3 ratio. This means that nearly 70% of data contributes to the train set and 30% of data contributes to the test set. Training set covers the period of 11 May 2015 to 31 December 2018 and consists of 922 days. Test set includes the period of 1 January 2019 to 10 March 2020 and has 309 days.

In order to rescale the features such that all values have a mean of zero and unit variance, we standardize our dataset:

$$z = \frac{X - \mu}{\sigma} \tag{4.14}$$

Like in the other artificial intelligence-based models, in LSTM, hyperparameters including number of hidden layers, number of neurons, regularization rate, learning rate, type of optimizer, window size should be carefully tuned. This stage includes trying all combinations and observing their performance over 30 epochs and choosing the best ones. Table 3 and Table 4 shows the combinations of hyperparameters that we try for hyperparameter optimization as well as the best hyperparameters that we select among them in terms of MSE and MAE.

Stock ID	**Learning Rate**	**Window Size**	**Momentum**	**L1 Regularization Rate**
1	Learning rate ∈	Windows size ∈	Momentum ∈	Regularization rate ∈
7	{0.01	{20,	{0.1,	{0,
11	0.001,	30,	0.2,	0.01,
13	0.003,	60,	0.3,	0.05,
17	0.0001,	90,	0.4,	0.001,
26	0.0003,	120,	0.5,	0.005,
29	0.0005, 0.0007,	150,	0.6,	0.0001,
36	0.0009, 0.00001,	210,	0.7,	0.00005,
41	0.00002, 0.00005,	240,	0.8,	0.000001,
42	0.000001, 0.000005}	300}	0.9}	0.0000001}

Table 3: The Parameters that are Used in Hyperparameter Optimization for LSTM

Stock ID	**Learning Rate**	**Window Size**	**Momentum**	**L1 Regularization Rate**	**Number of epochs**
1	0.000005	30	0.9	0	102
7	0.0005	30	0.7	0	22
11	0.0001	30	0.1	0	64
13	0.0001	30	0.9	0.000001	51
17	0.00005	30	0.6	0	55
26	0.00002	30	0.1	0.000001	167
29	0.0001	30	0.9	0	36
36	0.0003	30	0.1	0	60
41	0.00005	30	0.7	0.00001	82
42	0.0005	30	0.1	0	52

Table 4: Best Parameter Combination Reported by Parameter Tuning Experiments for LSTM

Window size is one of the most important hyperparameters in LSTM to tune. It determines the degree of past information to be considered. Further, it impacts on the shape of input that feeds into the model. If the window size is much lower than the optimal, there is a possibility that the model may miss important information, on the other hand, if window size is much higher, then the model becomes likely to be over-fitted and computational costs are increased. In our initial experiments, we try the following values: 20, 30, 60, 90, 120, 150, 210, 240, 300 (see Table 3). We reach the best results with window size 30.

We select RMSprop as an optimizer motived from the studies arguing that it works well with Recurrent Neural Networks (Chollet, 2016; Fischer and Krauss, 2018).

In order to determine the optimal epoch number for each stock, we use early stopping. If the validation loss does not improve in the selected patience periods, early stopping provide that the training process is stopped and the parameters of the model are adjusted with the lowest validation loss. Thus, we individually determine the number of epochs for each stock and reduce the risk of over-fitting. We restrict the patience with 15 epochs and determine maximum number of epochs as 1000.

Moreover, in order to avoid from over-fitting, we utilize regularization. Regularization is a method used to reduce the complexity of the model. In this implementation, the regularization term is added to the loss function that allows reducing the effects of variables which have high weight values. We use L1 regularization. L1 regularization ensures sparsity which means that the majority of weights are zero. This enables us to avoid from over-fitting problem by reducing model complexity.

Finally, the determined configuration of LSTM network is as follows: input layer with 23 features and 30 timesteps. LSTM layer with 100 neurons and L1 regularization for some of the stocks, output layer with one neuron.

4.1.2 Support Vector Regression (SVR)

Support Vector Regression (SVR) is a supervised machine learning method to solve regression problems and is a generalized version of Support Vector Machines (SVM) (Vapnik, 1995) which is one of the most used machine learning algorithms used in supervised classification tasks. Although not as popular as SVM, SVR is a quite effective method in regression tasks and it has received increasing attention (Liang et al., 2013). The descriptions of SVR follow Smola and Schölkopf (2004), Hsu et al. (2009), Yeh et al. (2011), Guo (2014) and Awad and Khanna (2015).

In SVR, the main goal is to find a function $f(x)$ such that the predicted response values of the training samples have at most an ϵ deviation from their observed response values. The area created by this deviation is called ϵ-tube. ϵ-tube *"represents the degree of precision at which the bounds on generalization ability apply."* (Hsu et al., 2009) ϵ-tube should be as flat as possible centred around the hyperplane, and also to contain most of the training instances. At the same time, model complexity and prediction error should be balanced.

SVR does not penalize the errors as long as they are in the ϵ-tube. Training vectors that lie within the ϵ-tube are considered as correct, while those outside the ϵ-tube are considered as incorrect and contribute to the loss function (See Figure 9). Therefore, training vectors within the ϵ-tube are not important in terms of the regression function (Hsu et al., 2009). The vectors lying on and outside of the boundary lines are the support vectors. Support vectors are the subset of the training data and are used to determine the regression surface (Awad and Khanna, 2015).

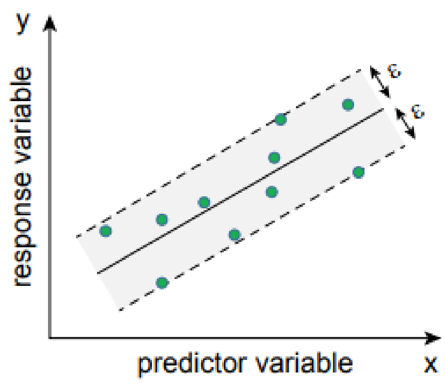

Figure 9: An Illustration of Support Vector Regression
Source: Zhang and O'Donnell (2020)

Suppose that we have the training data:

$$D = (x_1, y_1), ..., (x_l, y_l), \text{ where } x_i \in IR^n, \ i = 1, ..., l, \text{ and } y_i \in IR \tag{4.15}$$

Where each y_i is the target variable value for the n dimensional input vector x_i.

Following equation depicts the $f(x)$ function:

$$y = f(x) = \langle w, x \rangle + b = \sum_{j=1}^{M} w_j x_j + b \text{ where } y, b \in IR, \ x, w \in IR^M \tag{4.16}$$

Where w represents the weight vector and x represents the input vectors and b represents the bias.

Our aim is to find a function $f(x)$ that has the most ϵ deviation and to be as flat as possible. To ensure the flatness condition, we should minimize the norm of w. We can write this problem as a convex optimization problem:

$$\text{minimize } \frac{1}{2}||w||^2$$

$$\text{subject to: } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon \\ \langle w, x_i \rangle + b - y_i \leq \epsilon \end{cases}$$

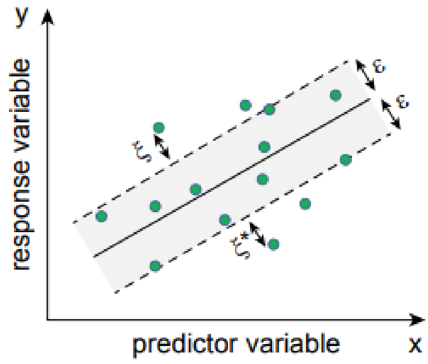

Figure 10: An Illustration of Support Vector Regression with Slack Variables
Source: Zhang and O'Donnell (2020)

So far, we have examined the case of function $f(x)$ actually exist that approximates all input data points (x_i, y_i) with ϵ precision. However, this situation may not hold all the time where outliers may deviate from most of the data points (Zhang and O'Donnell, 2020).

To handle this issue, we can add ξ, ξ^* slack variables to account for the noisy data which exceed the ϵ-tube (Cortes and Vapnik, 1995).

$$\text{minimize } \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$\text{subject to: } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \epsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

Where the constant $C > 0$ is the regularization parameter determining the trade-off between the flatness condition and the amount of deviation exceeding the ϵ is tolerated for this multi-objective optimization problem. Larger C gives more weight to minimizing the prediction errors. (Awad and Khanna, 2015).

So far, we have explained the case where $f(x)$ are linear. In order to deal with nonlinear problems, one can use a kernel function that transforms the original inputs to a higher dimensional space. The most commonly used kernel functions in SVR are linear kernel, polynomial kernel and radial basis kernel.

Where $\varphi(.) : R^d \to F$ is the transformation from original inputs to kernel space. We can re-arrange linear function $f(x)$ in terms of $\varphi(.)$ as follows:

$$y = f(x) = \langle w, \varphi(x) \rangle + b = w^T\varphi(x) + b \tag{4.17}$$

Then the optimization problem can be written as:

$$\text{minimize } \frac{1}{2}||w||^2 + C\sum_{i=1}^{l}(\xi_i + \xi_i^*)$$

$$\text{subject to:} \begin{cases} y_i - w^T\varphi(x_i) - b \leq \epsilon + \xi_i \\ w^T\varphi(x_i) + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases}$$

In this thesis, we utilize linear SVR to predict the subsequent close prices of selected stocks. As in LSTM, we split our dataset into train and test sets, employ standardization and carefully tune hyperparameters. We try 0.001, 0.01, 0.1, 1, 10, 100 and 1000 values for tuning the C parameter. Table 5 shows the C values that we reached the best results in terms of MAE and MSE.

ID	Selected parameters (C)
1	0.01
7	0.01
11	0.1
13	1
17	1
26	1
29	1
36	0.01
41	1
42	0.01

Table 5: Best Parameter Combination Reported by Parameter Tuning Experiments for SVR

4.1.3 Some Challenges in Artificial Intelligence Based Models

In artificial intelligence-based models, there is always risk to encounter with under-fitting and over-fitting. While training the model, we want the model to learn from the training data, but we do not want it to learn too small or too much since in such cases we may encounter underfitting or over-fitting problems. Both of these problems lead to low generalization and unreliable predictions on out-of-sample sets.

Consider that, we have the dataset as in Figure 11. We try to construct a machine learning model for this dataset.

Under-fitting occurs when the model cannot capture the underlying characteristics of the data. In this case, the model does not fit the data very well and does not understand the

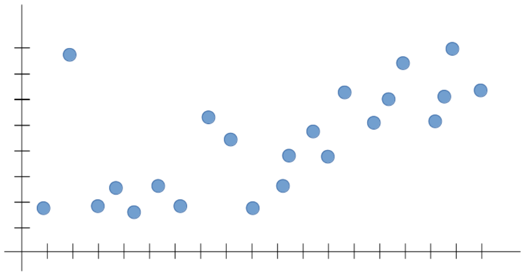

Figure 11: An Illustration of the Sample Dataset
Source: Al-Masri (2019)

relationship between the features and target variable. In simple terms, when the model cannot learn it refers to the under-fitting problem. Some possible solutions to handle with under-fitting problem include increasing model complexity, adding new features, performing feature engineering, increasing number of epochs and so on. Suppose that we face an under-fitting problem with the above dataset, then we are likely to encounter a graph like Figure 12.

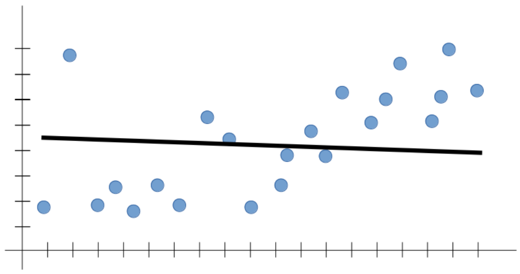

Figure 12: An Illustration of the Under-fitting Problem
Source: Al-Masri (2019)

Over-fitting means that the model learns very well on the training data but performs poorly on out-of-sample data. In this case, the model learns too much on the training data which results in learning from the noise and inaccurate points in the dataset. In other terms, the model has memorized the situations in the training set. Since the model looks for these situations in the out-of-sample set, it exhibits poor generalization performance. Possible solutions to handle with the over-fitting problem include reducing model complexity, increasing training data, setting early stopping, using drop out and regularization. Suppose that we face an over-fitting problem with the above dataset, then we have a possibility to face a graph like Figure 13.

In this study, in order to reduce the risk of encountering the under-fitting and over-fitting problems we execute several implementations. First of all, we use a comprehensive feature set in order to provide the model to understand the relationship between the input and

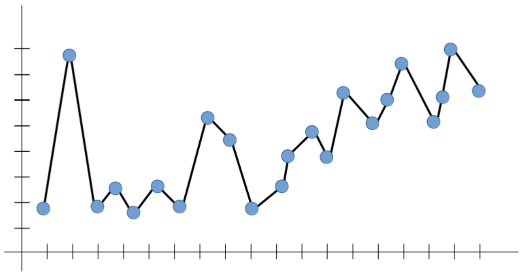

Figure 13: An Illustration of the Over-fitting Problem
Source: Al-Masri (2019)

output vectors very well. Secondly, we carefully tune the hyperparameters.

- **LSTM:** In LSTM, one of the most important hyperparameters is the window size. Window size determines the degree of past information to be considered. If window size is much lower than the optimal, there is a possibility that the model may miss important information, on the other hand if window size is much higher, then the model becomes likely to be over-fitted. Next, we use L1 regularization in order to avoid from over-fitting for some stocks' predictions. Regularization is a method used to reduce the complexity of the model. In this implementation, the regularization term is added to the loss function that allows reducing the effects of variables which have high weight values. L1 regularization ensures sparsity which means that the majority of weights are zero and enables to avoid from over-fitting. Lastly, in order to determine the optimal epoch number for each stock, we use early stopping. If the validation loss does not improve in the selected patience periods, early stopping provide that the training process is stopped and the weights of the model is adjusted with the lowest validation loss. Thus, we individually determine the number of epochs for each stock and reduce the risk of over-fitting.

- **SVR:** In SVR, it is very critical to tune C, which is the regularization parameter determining the trade-off between the flatness condition and the amount of deviation exceeding the ϵ is tolerated. Since big values of C will increase the model complexity, it is likely to encounter over-fitting. Taking this into consideration, we carefully tuned the C parameter by observing MAE and MSE values for both train and test data.

4.2 CONSTRUCTING TRADING STRATEGIES

We design two trading strategies based on the close price predictions we obtain LSTM networks (TS-LSTM) and SVR (TS-SVR). Then we compare these strategies with buy and hold strategy using BIST100 Index which simply rests on the idea that buy BIST100 Index at the beginning of the period and sell it at the end of the period. By this way, we will be able to understand how we perform if we had adopted a passive trading strategy instead of TS-LSTM (TTS-LSTM) and TS-SVR (TTS-SVR). Our simulation period covers the timespan between 02 January 2019 and 10 March 2020. We assume that stocks can be bought and sold instantly.

In our strategy, firstly, all stocks are equally weighted. We allocate a total of 1000 TL for the portfolio and 100 TL for each stock. These weights then change dynamically according to the buy and sell signals produced by the system. All transactions are carried out in the closing session (18:08-18:10) from which the transactions can be made at the closing price. This is because, the system generates the signal when the close price of today is determined.

Our trading strategy basically rests on the idea of that, buy the stocks if the prediction of the close price for the next day is higher than the close price of today, hold them until the predicted value is lower than the actual value and sell the stocks if predicted value is lower than the actual value. That is, if the predicted value of the next day is higher (lower) than today's actual value, the system will generate a buy (sell) signal. This strategy corresponds to buy the stocks, which provides a positive expectation about the next day and sell it before the expectation turns negative. So, based on the prediction results, we will be able to obtain returns as long as the price increase continues and prevent from the loss. If the system does not produce a buy signal for a period of time after generating a sell signal for a stock, we evaluate this period by investing in a risk-free interest rate. So, we run our capital every trading day whether the system has produced a signal or not.

Since our aim is to make a comparison with passive strategy, we have to make sure whether the transaction costs will cause any disadvantage for us. Therefore, the next stage is to consider transaction costs. Accordingly, we make a commission cut, fixed with the rate of 0.002 for each transaction.

We establish another kind of trading strategy that based on the idea that the trading system should generate signals to provide enough return to compensate at least transaction cost. In this stage, we're taking the advantage of focusing on the regression problem instead of a classification problem. In this strategy, our trading system will not generate signals unless the amount of return exceeds 0.002. Therefore, at least we get enough returns as much as to compensate transaction costs. We call these strategies TTS-LSTM and TTS-SVR, extra T comes from the threshold.

In order to evaluate the performance of each trading strategy we utilize performance metrics such as return, volatility, sharpe ratio and maximum drawdown. Although return and volatility provide information about the success of the trading strategy, using the risk-adjusted measures, in which the risk and return are considered together, will give a more accurate idea of the performance of the strategies. Sharpe ratio which was introduced by Nobel Laureate William Sharpe is a good and very simple metric to evaluate risk-adjusted return of a trading strategy. It scales excess return by standard deviation and expresses the return per unit of risk. (Fischer and Krauss, 2018) It is calculated by subtracting the risk-free rate (R_f) from the return of the trading strategy (R_p) and dividing by the standard deviation σ_p of the trading strategy's excess return:

$$SharpeRatio = \frac{R_p - R_f}{\sigma_p} \tag{4.18}$$

Sharpe ratio gives an idea about whether a trading strategy's excess returns are due to a result of right decisions or a result of bearing too much risk. Note that, a trading strategy is said to be successful when it yields higher returns without carrying the additional risk. The higher Sharpe ratio means a better risk-adjusted performance.

In investing operations, in addition to the risk-adjusted performance, preservation of the capital and steadiness of the strategy are also important. Therefore, the Maximum Drawdown (MDD) is a very critical metric to consider, since it focuses on capital protection which is one of the key concerns for investors. MDD is expressed in percentage terms and captures the largest percentage loss that an investor could have faced. Thus, it gives an idea about the worst scenario that an investor can experience. However, it only considers the size of the largest decline and does not provide any information about how many times this decline occurs. Calculation of MDD is as follows:

$$DD_t = min(0, \frac{p_t - p_{max}}{p_{max}}) \tag{4.19}$$

$$MDD_t = max(DD_t) \tag{4.20}$$

Where DD_t is the drawdown at the time t, p_{max} is the peak of the portfolio and pt is the current value of the portfolio.

Chapter 5

RESULTS

In this chapter, we present the results of both forecasting models and trading strategies that we construct based on the predictions of LSTM and SVR. First of all, we examine the prediction results of LSTM and SVR models in terms of Mean Absolute Error (MAE), Mean Squared Error (MSE) and directional accuracy. Moreover, in order to provide a better understanding of our prediction results, we also utilize MAE and MSE values of stock returns. Secondly, we investigate the results of trading strategies in terms of return, volatility, Sharpe ratio and maximum drawdown. In the next step, in order to make sure whether transaction costs cause any disadvantage for us we incorporate commission cuts to our analyses. Finally, we summarize and interpret the overall results.

5.1 PREDICTION RESULTS

In order to forecast the close prices of the next trading day of the selected 10 stocks from BIST30 Index, we utilize Long Short-Term Memory Networks and Support Vector Regression. In this section, we present the prediction results of LSTM networks and SVR in terms of Mean Absolute Error, Mean Squared Error and directional accuracy.

Comparison of the errors produced by the LSTM and SVR with respect to MAE, MSE and direction accuracy is shown in the Table 6. Table 6 reveals that the linear SVR model has superior predictive power compared to LSTM in terms of MAE and MSE. Except for the MAE of TSKB, the MAE of all other stocks is lower in SVR. Moreover, MSE of all stocks are also lower in SVR. However, the differences between MAE and MSE values of LSTM and SVR are quite low. Although SVR makes more closer predictions to

ID	Stock	LSTM			SVR		
		MAE	MSE	Direction Accuracy	MAE	MSE	Direction Accuracy
1	AKBNK	0,1264	0,0283	55,19%	0,1195	0,0269	53,90%
7	DOHOL	0,0405	0,0027	49,35%	0,0313	0,0015	42,86%
11	EREGL	0,1421	0,0391	50,32%	0,1220	0,0302	50,65%
13	GARAN	0,1753	0,0551	55,84%	0,1646	0,0519	54,22%
17	ISCTR	0,1320	0,0296	51,95%	0,1031	0,0191	47,73%
26	SAHOL	0,1488	0,0377	52,27%	0,1338	0,0329	55,19%
29	SODA	0,1140	0,0215	50,97%	0,0807	0,0121	47,73%
36	TSKB	0,0365	0,0031	52,92%	0,0381	0,0018	39,61%
41	VAKBN	0,1063	0,0258	49,03%	0,0949	0,0164	48,05%
42	YKBNK	0,0515	0,0045	52,60%	0,0426	0,0032	47,73%

Table 6: Close Price Prediction Results of LSTM and SVR

real values, LSTM performs better in capturing direction of price movements. As can be seen from the Table 6, considering the directional accuracy, LSTM outperforms SVR. The highest difference is for TSKB with 13.09%. This is followed by DOHOL with 6.49%, and YKBNK with 4.87%.

Figure 14 shows the prediction results of ten stocks obtained from LSTM networks and SVR along with their actual values. As can be clearly seen, the predictions of both models are very close to the actual values. However, looking at the graph of SODA and TSKB, one can see that the predicted values by LSTM are slightly away from the actual values. Nevertheless, the direction movements have been caught up better compared to SVR.

Although directional accuracy gives an idea of which stock is better predicted, using MAE and MSE without taking any action do not provide the appropriate evaluation. For example, if we look at the results of LSTM, it looks like the best predicted stock is DOHOL with 0.27% MSE value. However, in reality, we cannot understand it only by looking at the MAE and MSE values of close price predictions. There might be huge differences between close prices of stocks. In order to make a comparison between the prediction performance of stocks more precisely, we rearrange MAE and MSE values by making use of stock returns. In this direction, first, we calculate stock returns by using both predicted and actual values of close prices as follows:

$$Rp_t = \frac{Pp_t - Pa_{t-1}}{Pa_{t-1}} \tag{5.1}$$

Figure 14: Comparison of the Close Price Predictions Based on LSTM and SVR along with Actual Values

Where Rp_t denotes predicted return, Pp_t is predicted price by forecasting model, Pa_{t-1} is actual price of the stock at time $t-1$.

Thereafter, we compute MAE and MSE values using these stock returns. Thus, we obtain an idea about which stock is better predicted in more accurate way. Table 7 shows the MAE and MSE values of stock returns for both models.

ID	Stock	LSTM		SVR	
		MAE	MSE	MAE	MSE
1	AKBNK	0,0181	0,0005	0,0172	0,0005
7	DOHOL	0,0296	0,0013	0,0231	0,0008
11	EREGL	0,0177	0,0005	0,0152	0,0004
13	GARAN	0,0188	0,0006	0,0177	0,0006
17	ISCTR	0,0229	0,0009	0,0177	0,0005
26	SAHOL	0,0171	0,0005	0,0154	0,0004
29	SODA	0,0173	0,0004	0,0123	0,0002
36	TSKB	0,0373	0,0027	0,0445	0,0026
41	VAKBN	0,0218	0,0008	0,0197	0,0007
42	YKBNK	0,0224	0,0008	0,0184	0,0006

Table 7: Return prediction Results Based on the Close Price Predictions of LSTM and SVR

Looking at the MSE values of both models, it is seen that the best predicted stock is SODA. In LSTM, it is followed by SAHOL, EREGL, AKBNK, GARAN, VAKBN, YKBNK, ISCTR, DOHOL and TSKB respectively. In SVR, the prediction performance order in terms of MSE is as follows: SODA, EREGL, SAHOL, AKBNK, ISCTR, YKBNK, GARAN, VAKBN, DOHOL, TSKB. In terms of MAE, both models have similar orders with MSE as expected. In LSTM stocks can be ordered from the best predicted to the worse as follows: SAHOL, SODA, EREGL, AKBNK, GARAN, VAKBN, YKBNK, ISCTR, DOHOL and TSKB. In SVR, SODA is the best predicted stock in terms of MAE. It is followed by EREGL, SAHOL, AKBNK, ISCTR, GARAN, YKBNK, VAKBN, DOHOL and TSKB. Along with the directional accuracy, these orders are important in terms of giving an idea about which stocks are the most contributed the success of the trading strategies. In other words, one can say that the trading system generates more accurate signals about the stocks with the best rankings.

5.2 BACKTEST RESULTS

5.2.1 Results of TS- LSTM and TS- SVR

As mentioned above, we construct trading strategies based on the close price predictions obtained with LSTM networks and SVR. We call these strategies Trading Strategy of Long Short Term Memory (TS-LSTM) and Trading Strategy of Support Vector Regression (TS-SVR). If the predicted value of the next day is higher (lower) than today's actual value, the system generates a buy (sell) signal. Thus, we buy the stocks if the prediction of the close price for the next day is higher than the close price of today; hold them until the predicted value is lower than the actual value and; sell the stocks if predicted value is lower than the actual value. This strategy corresponds to buy the stocks which provides the expectation that the next day will yield positive returns and, sell it before the expectation turns to negative. Accordingly, we will able to obtain returns as long as the price increase continues and avoid from the loss. If the system has not produced a buy signal for a period of time after generating a sell signal for a stock, we evaluate this period by investing in a risk-free interest rate. Therefore, we run our capital every trading day whether the system has produced a signal or not.

We select buy and hold BIST100 Index as our benchmark strategy which simply rests on the idea that buy BIST100 Index at the beginning of the period and sell it at the end of the period. By this way, we will be able to comprehend how we perform if we had adopted a passive trading strategy instead of TS-LSTM and TS-SVR. In order to justify the performance of active strategy against the passive strategy, we have to make sure that the transaction costs do not cause any disadvantage for us in terms of performance metrics. Therefore, in the next stage of our analysis, we consider transaction costs. We apply a commission cut, fixed with the rate 0.002 for each transaction. Then we evaluate the performance of each trading strategy in terms of the annualized return, volatility, Sharpe ratio and maximum drawdown.

Table 8 shows the backtest results of our trading strategies in terms of return, volatility, Sharpe ratio and maximum drawdown without any consideration of transaction costs. As it is clear from the table, TS-LSTM and TS-SVR yield higher returns and have lower volatility than buy-and-hold BIST100 Index. TS-LSTM has the best performance with 29.71% return and 0.1603 volatility. It is followed by TS-SVR with 26.81% return and 0.1823 volatility and BIST100 14.28% return and 0.2207 volatility, respectively.

Strategy	Annualized Performance Measures			
	Return	Volatility	Sharpe Ratio	Maximum Drawdown
TS-LSTM	29.71%	0.1603	1.2031	-10.62%
TS-SVR	26.81%	0.1823	0.8987	-14.78%
BIST100	14.28%	0.2207	0.1748	-20.37%

Table 8: Backtest Results of TS-LSTM and TS-SVR without Transaction Costs

Evaluating the performance of a trading strategy just by examining its return and volatility does not provide accurate information. Therefore, we also utilize Sharpe ratio which gives us an idea about whether a trading strategy's excess returns result from right decisions or from bearing too much risk. As expected, TS-LSTM and TS-SVR generate significantly higher Sharpe ratio than the buy-and-hold BIST100 Index since they have higher return and lower volatility. TS-LSTM exhibits superior performance with 1.20 Sharpe ratio, it is followed by TS-SVR with 0.89. Buy and hold BIST100 Index strategy has the lowest Sharpe ratio with 0.17. It is seen that in terms of Sharpe ratio TS-LSTM and TS-SVR exhibit superior performance compared to benchmark strategy. Sharpe ratio of the TS-LSTM is seven times as much as the BIST100 index and Sharpe ratio of the TS-SVR is five times as much as the BIST100 index. All of these explanations indicate that TS-LSTM and TS-SVR yield higher returns relative to the amount of risk that is exposed than the passive BIST100 strategy.

We also utilize maximum drawdown, which is a measure capturing the largest percentage loss that an investor can face with. As can be seen from Table 8, TS-LSTM and TS-SVR have much lower maximum drawdown than BIST100 index with -10.62%, -14.78% and -20.37% respectively. That is, if an investor had invested with the TS-LSTM strategy, he/she would face the largest decline with -10.62%, whereas; had he/she invested in BIST100 index and adopt a passive strategy, he/she would experience the largest decline with -20.37%.

In order to justify the superior performances of our active strategies, we have to make sure that the transaction costs do not cause any disadvantage for us in terms of performance metrics. In addition, evaluating the backtest results without considering transaction costs does not provide a real-life analysis. Therefore, in the next stage of our analysis, we consider transaction costs. We apply a commission cut fixed with the rate of 0.002 for each transaction. Table 9 shows the backtest results of our trading strategies with the transaction costs.

Strategy	Annualized Performance Measures			
	Return	Volatility	Sharpe Ratio	Maximum Drawdown
TS-LSTM with commission	20.16%	0.1560	0.6238	-10.93%
TS-SVR with commission	14.05%	0.1820	0.1987	-16.56%
BIST100 with commission	13.90%	0.2211	0.1575	-20.37%

Table 9: Backtest Results of TS-LSTM and TS-SVR with Transaction Costs

As can be seen, even if transaction costs are considered, TS-LSTM and TS-SVR yield higher returns and have lower volatility compared with buy-and-hold BIST100 Index. TS-LSTM has the best performance with 20.16% return and 0.1560 volatility. It is followed by TS-SVR with 14.05% return and 0.1820 volatility and, by BIST100 with 13.90% return and 0.2211 volatility.

Moreover, TS-LSTM and TS-SVR generate higher Sharpe ratio than the buy-and-hold BIST100 Index. TS-LSTM exhibit superior performance with 0.62 Sharpe ratio, followed by TS-SVR with 0.19. Buy and hold BIST100 Index strategy has the lowest Sharpe ratio with 0.15. It is seen that in terms of Sharpe ratio, TS-LSTM and TS-SVR exhibit superior performance. Sharpe ratio of the TS-LSTM is more than four times as much as the BIST100 index and Sharpe ratio of the TS-SVR is 1.26 times as much as the BIST100 index. Moreover, TS-LSTM and TS-SVR have much lower maximum drawdown than BIST100 index with -10.93%, -16.56% and -20.37%, respectively.

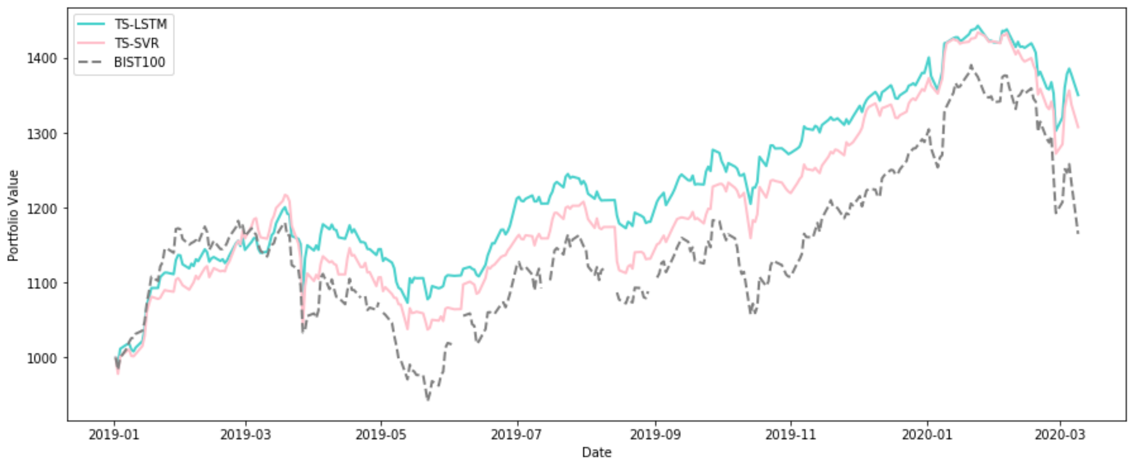

Figure 15: Portfolio Values of Each Trading Strategy During the Backtest Period without Considering Commission

Figure 15 and Figure 16 show the portfolio value of each trading strategy during backtest period. Except for TS-SVR with the commission, all these strategies, especially the TS-LSTM and TS-LSTM with commission outperform the benchmark strategy according to

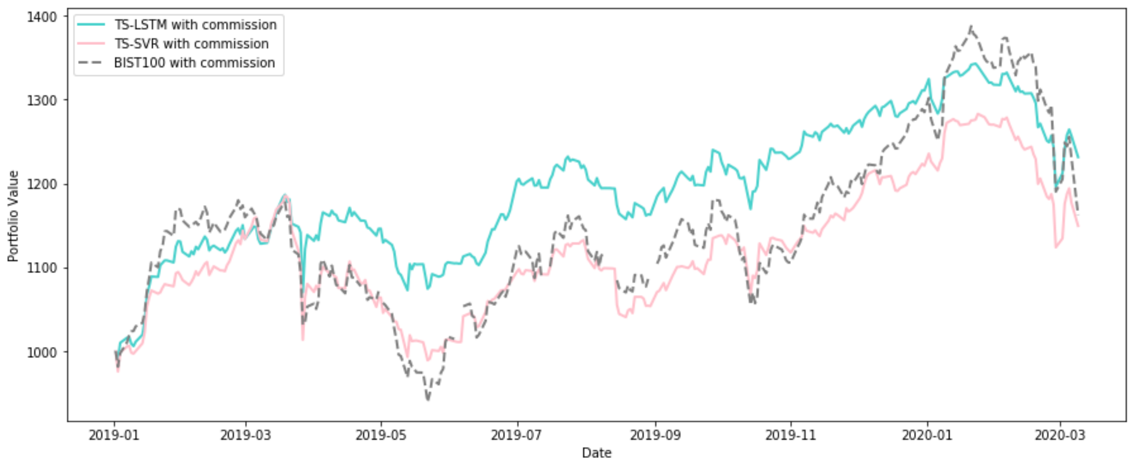

Figure 16: Portfolio Values of Each Trading Strategy During the Backtest Period with Commission Cuts

their portfolio values. Although, at the end of the period, the portfolio value of LSTM-SVR with commission is lower than the passive strategy, it performs better than the benchmark in terms of all performance metrics.

5.2.2 Results of TTS- LSTM and TTS- SVR

We additionally establish another trading strategy which is based on the idea that the trading system should generate signals to provide enough return to compensate at least the transaction costs. In this respect, we determine the threshold as 0.002, which equals to transaction cost. In this strategy, our trading system will not generate signals unless the amount of return exceeds 0.002. Table 10 shows the backtest results of trading strategies with a threshold.

Strategy	Annualized Performance Measures			
	Return	**Volatility**	**Sharpe Ratio**	**Maximum Drawdown**
TTS-LSTM	30.86%	0.1405	1.4545	-9.36%
TTS-SVR	22.01%	0.1506	0.7686	-12.46%
BIST100	14.28%	0.2207	0.1748	-20.37%

Table 10: Backtest Results of TTS-LSTM and TTS-SVR without Transaction Costs

As can be seen from Table 10, similar with TS-LSTM and TS-SVR, TTS-LSTM and TTS-SVR yield higher returns and have lower volatility than buy-and-hold BIST100 Index. TTS-LSTM has the best performance in terms of all performance metrics. Then it is followed by TTS-SVR. Compared to TS-LSTM, TTS-LSTM yields approximately 1% higher returns and 0.02 lower volatility. For TTS-SVR, it has 4% higher returns and

0.03 lower volatility compared to those of TS-SVR. As a result, one can say that in both strategies Sharpe ratio is improved. The setting threshold is also improved maximum drawdown. For TTS-SVR, although there is an improvement in terms of risk metrics, return and the Sharpe ratio is decreased compared to TS-SVR.

Then, we incorporate transaction costs to this strategy:

Strategy	**Annualized Performance Measures**			
	Return	**Volatility**	**Sharpe Ratio**	**Maximum Drawdown**
TTS-LSTM with commission	23.14%	0.1043	0.9135	-10.64%
TTS-SVR with commission	13.36%	0.1491	0.1971	-13.29%
BIST100 with commission	13.90%	0.2211	0.1575	-20.37%

Table 11: Backtest Results of TTS-LSTM and TTS-SVR with Transaction Costs

Similar to TS-LSTM and TS-SVR, TTS-LSTM and TTS-SVR exhibit superior performance in terms of nearly all performance metrics even after transaction costs are considered. As expected, TTS-LSTM has the best performance in terms of all metrics and it is followed by TTS-SVR. Compared to TS-LSTM, TTS-LSTM yields approximately 3% higher returns and has 0.05 lower volatility. For TTS-SVR, it has 1% higher returns and 0.03 lower volatility compared to TS-SVR. Considering the return and volatility results, it is seen that setting the threshold improves the Sharpe Ratio of TTS-LSTM and worsens the Sharpe Ratio results of TTS-SVR. The maximum drawdown values of both TTS-LSTM and TTS-SVR decreased and they are much better than the benchmark strategy.

5.2.3 Evaluation of the Overall Results

The success of trading strategies is directly related to the success of our forecasting models. TS-LSTM (TTS-LSTM) performs better compared to TS-SVR (TTS-SVR) with and without transaction costs. The reason behind TS-LSTM's (TTS-LSTM's) better performance in terms of backtest results, can be explained with our dealing with the direction of stock price movements and LSTM's superior performance at direction accuracy. Another reason is that TS-SVR generates more buy and sell signals compared to TS-LSTM, thus it is more exposed to transaction costs that result in lower sharpe ratio and portfolio value. Moreover, setting the threshold increases performances of trading strategies in terms of almost all performance metrics.

Remember that, in our trading strategy, the predicted value of the next day is higher (lower) than today's actual value where our trading system generates a buy (sell) signal. Also, we don't invest in every stock every day. If the system did not generate a buy signal on the day after a sell signal is produced, then we continue to run our capital by investing in the risk-free rate in the days that we don't invest in that stock. This is how the system tells us "do not invest in that stock in those days". This property of the trading system might have provided us to yield better returns and also to be exposed to less risk. For example, our trading system may not have generated a buy signal in high volatility period. Therefore, it is possible that in such periods investing in the risk free rate might have protected us from bearing the high risk. Consequently, considering the overall performance of our trading strategies, TS-LSTM (TTS-LSTM) and TS-SVR (TTS-SVR) outperform the benchmark strategy in terms of most of the performance metrics even after considering the transaction costs.

We determine the backtest period as 309 days which corresponds to more than one trading year. During this period, we consistently beat the market. This reinforces the evidence that we have not achieved this success by chance, which contradicts one of the main propositions of EMH advocating that no one can consistently outperforms the market except by taking more risk, or by chance. It's all about the success of our forecasting models and trading systems.

Chapter 6

CONCLUSION

In this study, we utilize artificial intelligence-based models to predict the next day's close price of the stocks. We choose our sample as Borsa Istanbul, inspired by the literature arguing that emerging markets are less efficient, therefore more predictable and more exposed to investor sentiment. We randomly select 10 of 45 stocks included in the BIST30 index which are among to the most capitalized and the most actively traded stocks in the past few years following the approach that investors closely follow these most popular stocks' stories.

In this respect, we focus on a regression problem and make use of LSTM and SVR models which exhibit superior performance in financial forecasting tasks. We feed these models using a comprehensive dataset, including the technical analysis indicators together with the investor sentiment variables. Thus, we make predictions by using historical price data and technical analysis indicators which are the informations that everyone can access without any costs and as well as using the investor sentiment that contains the market's non-rational components. We choose this feature set in response to the EMH's arguments about weak-form efficiency and investor's pure rationality. We use Bloomberg's news sentiment scores as a proxy for investor sentiment following the fact that the information extracted from news stories form the basis of investors' sentiment about the relevant instrument. Moreover, news sentiment carries the irrational components of the markets. Another reason of including the sentiment factor in our model is that in parallel with the studies advocating that news sentiment has predictive power on future close prices, we see news sentiment as an essential element to be taken into consideration in the prediction models.

Next, we search for an answer to the question of how an investor can take the advantage of these predictions in practice and, construct trading strategies by making use of the predictions produced by the forecasting models. In this respect, we design trading strategies using the results from both LSTM and SVR models and, compare the performance of these strategies with buy and hold BIST100 Index strategy. To do so, we establish two kind of trading strategies. One of these strategies mainly focuses on the direction of the price movements while the other considers a certain threshold value in generating buy and sell signals. We take the advantage of dealing with regression instead of classification, although we are interested in the movement direction. This allows us the flexibility to differentiate our strategies. Then, in order to make sure that transaction costs will not cause any disadvantage for us, we incorporate them into our trading system by making a fixed commission cut for every transaction. Finally, we evaluate the performance of all trading strategies in terms of return, volatility, Sharpe ratio and Maximum Drawdown.

Our results show the superior performance of our trading strategies that are constructed using both LSTM and SVR models compared to simply buy and hold market index in terms of all performance metrics. We reach similar results even after transaction costs are considered. This implies active trading strategies outperforms passive strategy in our case. Further, we determine the backtest period as 309 days which corresponds to a period more than one trading year. During this period, we consistently beat the market. This reinforces the evidence that we have not achieved this success by chance, which contradicts one of the main propositions of EMH advocating that no one can consistently outperforms the market except by taking more risk, or by chance. It is all about the success of our forecasting models and trading systems. We show that we can make successful forecastings and construct trading strategies based on these predictions using artificial intelligence-based models and publicly available information which is the fundamental purpose of this study. Thus, we provide contradictory evidence with the negatory argument about the asset price predictability of EMH.

In this study, we construct trading strategies based on the buy and sell signals generated from the predictions of forecasting models. However, we do not execute a stock selection process except for choosing the most popular stocks from BIST30 Index. Future studies can be performed to manage stock selection processes by focusing on the ranking of the stocks using sophisticated methods such as learning to rank algorithms, and synchronous approaches can be adopted for both stock selection and market timing processes. Secondly, in order to expand the feature set, fundamental analysis indicators, central bank statements, global and political news sentiment can be incorporated to the model. In addition, periods utilized in technical analysis indicators can be tuned by making use of the heuristic algorithms such as genetic algorithm. Moreover, firm-level news sentiment scores can be re-arranged considering the inter-relation of the stocks. Lastly, in order to enrich the comparison of the performance of the trading strategies, in addition to the passive strategy, technical analysis and sentiment aware strategies can be involved.

Emerging markets share some common properties including high volatility, rapid growth, low maturity etc. In addition, emerging markets are less efficient compared to developed markets which points out that in these markets, stock prices and returns are more predictable. These characteristics makes these markets special for investors, also give more chance for them to evaluate these opportunities and increase the potential of making profitable transactions. As an emerging market Borsa Istanbul carries these attributes. At this point, using artificial intelligence based techniques stand out as one effective way to capture the patterns ready to be discovered in these markets. Investors should explore and take the advantage of these methods. In our opinion, increasing usage of these methods will enhance the market liquidity and enhance financial development through financial markets' depth channel. Moreover, this will contribute to economic growth in line with the phenomenon that advocates that financial development promotes economic growth.

Today, the predictability of asset prices is still open for debate. Our humble endeavor is all about demonstrating that the stock prices can be forecasted to some degree by making use of artificial intelligence-based models and one can take the advantage of these predictions. The interaction between finance and computer science is still in its infancy. In our opinion, the success of financial forecasting efforts will improve as artificial intelligence techniques evolve and financial data accumulates. We hope this study inspires the researchers and finance practitioners who want to engage in this field.

Bibliography

Achelis, S. B. (2001). *Technical Analysis from A to Z.* McGraw Hill New York.

Aktan, C., Sahin, E. E., and Kucukkaplan, I. (2018). Testing the information efficiency in emerging markets. In Kucukkocaoglu, G. and Gokten, S., editors, *Financial Management from an Emerging Market Perspective*, chapter 3. IntechOpen, Rijeka.

Al-Masri, A. (2019). What are overfitting and underfitting in machine learning? *Available at:https://towardsdatascience.com/what-are-overfitting-and-underfitting-in-machine-learning-a96b30864690/ Accessed March 20, 2020.*

Al Nasseri, A. A. M. (2016). *The predictive power of stock micro-blogging sentiment in forecasting stock market behaviour*. PhD thesis, Brunel University London.

Alfano, S. J., Feuerriegel, S., and Neumann, D. (2015). Is news sentiment more than just noise? In *ECIS*.

Allen, D. E., McAleer, M., and Singh, A. K. (2019). Daily market news sentiment and stock prices. *Applied Economics*, 51(30):3212–3235.

Antoniou, C., Doukas, J. A., and Subrahmanyam, A. (2013). Cognitive dissonance, sentiment, and momentum. *Journal of Financial and Quantitative Analysis*, 48(1):245–275.

Antweiler, W. and Frank, M. Z. (2004). Is all that talk just noise? the information content of internet stock message boards. *The Journal of finance*, 59(3):1259–1294.

Ashraf, N., Camerer, C. F., and Loewenstein, G. (2005). Adam smith, behavioral economist. *Journal of Economic Perspectives*, 19(3):131–145.

Awad, M. and Khanna, R. (2015). Support vector regression. In *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, pages 67–80. Apress, Berkeley, CA.

Baker, M. and Wurgler, J. (2006). Investor sentiment and the cross-section of stock returns. *The journal of Finance*, 61(4):1645–1680.

Baker, M. and Wurgler, J. (2007). Investor sentiment in the stock market. *Journal of economic perspectives*, 21(2):129–152.

Barberis, N. and Thaler, R. (2003). A survey of behavioral finance. *Handbook of the Economics of Finance*, 1:1053–1128.

Bayram, S. G. (2017). Rational–irrational investor sentiments and emerging stock market returns: A comparison from turkey. *Journal of Emerging Market Finance*, 16(3):219–245.

Black, F. (1986). Noise. *The journal of finance*, 41(3):528–543.

Bodie, Z., Kane, A., and Marcus, A. (2000). *Essentials of Investments*. Irwin/McGraw-Hill series in finance, insurance, and real estate. McGraw-Hill Higher Education.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

Britz, D. (2015). Recurrent neural networks tutorial, part 1–introduction to rnns. *Available at: http://www. wildml. com/2015/09/recurrent-neural-networkstutorial-part-1-introduction-to-rnns/ Accessed 19 May 2020*.

Brown, G. W. and Cliff, M. T. (2004). Investor sentiment and the near-term stock market. *Journal of empirical finance*, 11(1):1–27.

Brzeszczyński, J., Gajdka, J., and Kutan, A. M. (2015). Investor response to public news, sentiment and institutional trading in emerging markets: A review. *International Review of Economics & Finance*, 40:338–352.

Chen, H., De, P., Hu, Y. J., and Hwang, B.-H. (2013). Customers as advisors: The role of social media in financial markets. Technical report, Working paper.

Chollet, F. (2016). Keras deep learning library.

Chopra, D., Joshi, N., and Mathur, I. (2016). *Mastering natural language processing with python*. Packt Publishing Ltd.

Chung, H. and Shin, K. (2018). Genetic algorithm-optimized long short-term memory network for stock market prediction. *Sustainability*, 10(10):3765.

Cortes, C. and Vapnik, V. (1995). Support vector machine. *Machine learning*, 20(3):273–297.

Cui, X., Lam, D., and Verma, A. (2016). Embedded value in bloomberg news and social sentiment data. *Bloomberg LP*.

Da, Z., Engelberg, J., and Gao, P. (2015). The sum of all fears investor sentiment and asset prices. *The Review of Financial Studies*, 28(1):1–32.

Dawar, H. (2020). Stochastic gradient descent. *URL:https://medium.com/analytics-vidhya/stochastic-gradient-descent-1ab661fabf89/ accessed May 20, 2020]*.

De Bondt, W. P. (1993). Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of forecasting*, 9(3):355–371.

De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738.

Dunham, L. M. and Garcia, J. (2020). Measuring the effect of investor sentiment on liquidity. *Managerial Finance*.

Edmans, A., Garcia, D., and Norli, Ø. (2007). Sports sentiment and stock returns. *The Journal of finance*, 62(4):1967–1998.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Fama, E. F. (1991). Efficient capital markets: Ii. *The journal of finance*, 46(5):1575–1617.

Ferguson, N. J., Philip, D., Lam, H., and Guo, J. M. (2015). Media content and stock returns: The predictive power of press. *Multinational Finance Journal*, 19(1):1–31.

Festinger, L. (1957). *A theory of cognitive dissonance*, volume 2. Stanford university press.

Feuerriegel, S. and Prendinger, H. (2016). News-based trading strategies. *Decision Support Systems*, 90:65–74.

Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2):654–669.

Fisher, K. L. and Statman, M. (2000). Investor sentiment and stock returns. *Financial Analysts Journal*, 56(2):16–23.

Gandhmal, D. P. and Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Computer Science Review*, 34:100190.

Généreux, M., Poibeau, T., and Koppel, M. (2011). Sentiment analysis using automatically labelled financial news items. In *Affective computing and sentiment analysis*, pages 101–114. Springer.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Güdelek, M. U. (2019). Zaman serisi analiz ve tahmini: Derin öğrenme yaklaşımı. Master's thesis, TOBB ETÜ Fen Bilimleri Enstitüsü.

Gudelek, M. U., Boluk, S. A., and Ozbayoglu, A. M. (2017). A deep learning based stock trading model with 2-d cnn trend detection. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE.

Gunduz, H. and Cataltepe, Z. (2015). Borsa istanbul (bist) daily prediction using financial news and balanced feature selection. *Expert Systems with Applications*, 42(22):9001–9011.

Gunduz, H., Cataltepe, Z., and Yaslan, Y. (2017). Stock market direction prediction using deep neural networks. In *2017 25th Signal Processing and Communications Applications Conference (SIU)*, pages 1–4.

Gunduz, H., Yaslan, Y., and Cataltepe, Z. (2017). Intraday prediction of borsa istanbul using convolutional neural networks and feature correlations. *Knowledge-Based Systems*, 137:138–148.

Guo, G. (2014). Soft biometrics from face images using support vector machines. In *Support Vector Machines Applications*, pages 269–302. Springer.

Halil, R. and Demirci, M. (2019). Predicting the turkish stock market bist 30 index using deep learning. *Uluslararası Mühendislik Araştırma ve Geliştirme Dergisi*, 11(1).

Henrique, B. M., Sobreiro, V. A., and Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems with Applications*, 124:226–251.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? *The Journal of Business Communication (1973)*, 45(4):363–407.

Hirshleifer, D. and Shumway, T. (2003). Good day sunshine: Stock returns and the weather. *The Journal of Finance*, 58(3):1009–1032.

Ho, K.-Y., Shi, Y., and Zhang, Z. (2013). How does news sentiment impact asset volatility? evidence from long memory and regime-switching approaches. *The North American Journal of Economics and Finance*, 26:436–456.

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.

Hsu, S.-H., Hsieh, J. P.-A., Chih, T.-C., and Hsu, K.-C. (2009). A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression. *Expert Systems with Applications*, 36(4):7947–7951.

Ince, H. and Trafalis, T. B. (2008). Short term forecasting with support vector machines and application to stock price prediction. *International Journal of General Systems*, 37(6):677–687.

Jain, P. (2019). A practical guide related to benefits, problems, and comparison of activation functions like sigmoid, tanh, relu, leaky relu and maxout. *Available at: https://mc.ai/complete-guide-of-activation-functions/ Accessed 19 May 2020.*

Jensen, M. C. (1978). Some anomalous evidence regarding market efficiency. *Journal of financial economics*, 6(2/3):95–101.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2):263–291.

Kamstra, M. J., Kramer, L. A., and Levi, M. D. (2003). Winter blues: A sad stock market cycle. *American Economic Review*, 93(1):324–343.

Kara, Y., Boyacioglu, M. A., and Baykan, Ö. K. (2011). 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):5311–5319.

Karpathy, A. (2015). The unreasonable effectiveness of recurrent neural networks.[online] karpathy. github. io. *http://karpathy.github.io/2015/05/21/rnn-effectiveness/ accessed March 20, 2020.*

Kearney, C. and Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33:171–185.

Keynes, J. M. (1936). The general theory of interest, employment and money.

Khan, Z. H., Alin, T. S., and Hussain, M. A. (2011). Price prediction of share market using artificial neural network (ann). *International Journal of Computer Applications*, 22(2):42–47.

Kim, H.-j. and Shin, K.-s. (2007). A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets. *Applied Soft Computing*, 7(2):569–576.

Koppel, M. and Shtrimberg, I. (2006). Good news or bad news? let the market decide. In *Computing attitude and affect in text: Theory and applications*, pages 297–301. Springer.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.

Lærum, K. H. (2018). A study of machine learning for predictive maintenance-a topic and programming guidance. Master's thesis, NTNU.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553):436–444.

Li, Q., Wang, T., Gong, Q., Chen, Y., Lin, Z., and Song, S.-k. (2014). Media-aware quantitative trading based on public web information. *Decision support systems*, 61:93–105.

Liang, X., Chen, R.-C., He, Y., and Chen, Y. (2013). Associating stock prices with web financial information time series based on support vector regression. *Neurocomputing*, 115:142–149.

Liston-Perez, D., Torres-Palacio, P., and Bayram, S. G. (2018). Does investor sentiment predict mexican equity returns? *International Journal of Managerial Finance*.

Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.

Liu, S., Zhang, C., and Ma, J. (2017). Cnn-lstm neural network model for quantitative strategy analysis in stock markets. In *International Conference on Neural Information Processing*, pages 198–206. Springer.

Lo, A. W., Mamaysky, H., and Wang, J. (2000). Foundations of technical analysis: Computational algorithms, statistical inference, and empirical implementation. *The journal of finance*, 55(4):1705–1765.

Loughran, T. and McDonald, B. (2011). When is a liability not a liability? textual analysis, dictionaries, and 10-ks. *The Journal of Finance*, 66(1):35–65.

Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Selim, M. M., and Muhammad, K. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management*, 50:432–451.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

Mitra, L. and Mitra, G. (2012). Applications of news analytics in finance: A review. In *The Handbook of News Analytics in Finance*, pages 1–39. John Wiley Sons.

Murphy, J. J. (1999). *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin.

Nielsen, M. A. (2015). *Neural networks and deep learning*, volume 2018. Determination press San Francisco, CA.

Nikou, M., Mansourfar, G., and Bagherzadeh, J. (2019). Stock price prediction using deep learning algorithm and its comparison with machine learning algorithms. *Intelligent Systems in Accounting, Finance and Management*, 26(4):164–174.

Nti, I. K., Adekoya, A. F., and Weyori, B. A. (2019). A systematic review of fundamental and technical analysis of stock market predictions. *Artificial Intelligence Review*, pages 1–51.

Olah, C. (2015). Understanding lstm networks. *URL: https://colah.github.io/posts/2015-08-Understanding-LSTMs/ accessed March 30, 2020*.

Ozbayoglu, A. M., Gudelek, M. U., and Sezer, O. B. (2020). Deep learning for financial applications: A survey. *Applied Soft Computing*, page 106384.

Özçalıcı, M. and Ayrıçay, Y. (2016). Bilgi işlemsel zeka yöntemleri ile hisse senedi fiyat tahmini: Bist uygulaması. *Pamukkale Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, (Ek1):274–298.

Park, C.-H. and Irwin, S. H. (2007). What do we know about the profitability of technical analysis? *Journal of Economic Surveys*, 21(4):786–826.

Prasad, H. (2018). Interesting deep learning techniques — optimal learningrate finder and reducing model training time. *https://medium.com/@hariprasad0994/interesting-deep-learning-techniques-lr-366f013d959a/ accessed March 30, 2020]*.

Pring, M. J. (1980). *Technical analysis explained: An illustrated guide for the investor*. McGraw-Hill.

Prosad, J. M., Kapoor, S., and Sengupta, J. (2015). Theory of behavioral finance. In

Handbook of research on behavioral finance and investment strategies: decision making in the financial industry, pages 1–24. IGI Global.

Qian, B. and Rasheed, K. (2007). Stock market prediction with multiple classifiers. *Applied Intelligence*, 26(1):25–33.

Sakarya, S., Yavuz, M., Karaoglan, A. D., Özdemir, N., et al. (2015). Stock market index prediction with neural network during financial crises: a review on bist-100. *Financial Risk and Management Reviews*, 1(2):53–67.

Sayim, M., Morris, P. D., and Rahman, H. (2013). The effect of us individual investor sentiment on industry-specific stock returns and volatility. *Review of Behavioural Finance*.

Sayim, M. and Rahman, H. (2015). The relationship between individual investor sentiment, stock return and volatility. *International Journal of Emerging Markets*.

Schumaker, R. P. and Chen, H. (2009). Textual analysis of stock market prediction using breaking financial news: The azfin text system. *ACM Transactions on Information Systems (TOIS)*, 27(2):1–19.

Seok, S. I., Cho, H., and Ryu, D. (2019). Firm-specific investor sentiment and daily stock returns. *The North American Journal of Economics and Finance*, 50:100857.

Sezer, O. B., Gudelek, M. U., and Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90:106181.

Shleifer, A. and Summers, L. H. (1990). The noise trader approach to finance. *Journal of Economic perspectives*, 4(2):19–33.

Shleifer, A. and Vishny, R. W. (1997). The limits of arbitrage. *The Journal of finance*, 52(1):35–55.

Simon, H. A. (1955). A behavioral model of rational choice. *The quarterly journal of economics*, 69(1):99–118.

Smith, A. (1759). The theory of moral sentiments, ed. *DD Raphael & AL Macfie. Liberty Fund.(Original work published in 1759.)[ELK]*.

Smola, A. J. and Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222.

Stanković, J., Marković, I., and Stojanović, M. (2015). Investment strategy optimization using technical analysis and predictive modeling in emerging markets. *Procedia Economics and Finance*, 19:51–62.

Statman, M. (1999). Behaviorial finance: Past battles and future engagements. *Financial Analysts Journal*, 55(6):18–27.

Sundaram, B., Palaniswani, M., Shilton, A., and Begg, R. (2006). Computational intelligence techniques. In *Computational Intelligence for Movement Sciences: Neural Networks and Other Emerging Techniques*, pages 139–171. IGI Global.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance*, 62(3):1139–1168.

Tetlock, P. C., Saar-Tsechansky, M., and Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. *The Journal of Finance*, 63(3):1437–1467.

Van de Kauter, M., Breesch, D., and Hoste, V. (2015). Fine-grained analysis of explicit and implicit sentiment in financial news articles. *Expert Systems with applications*, 42(11):4999–5010.

Vanstone, B. J., Gepp, A., and Harris, G. (2019). Do news and sentiment play a role in stock price prediction? *Applied Intelligence*, 49(11):3815–3820.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer-Verlag New York, Inc.

Verma, R. and Soydemir, G. (2006). The impact of us individual and institutional investor sentiment on foreign stock markets. *The Journal of Behavioral Finance*, 7(3):128–144.

Weng, B., Lu, L., Wang, X., Megahed, F. M., and Martinez, W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Systems with Applications*, 112:258–273.

Yeh, C.-Y., Huang, C.-W., and Lee, S.-J. (2011). A multiple-kernel support vector regression approach for stock market price forecasting. *Expert Systems with Applications*, 38(3):2177–2186.

Zhang, F. and O'Donnell, L. J. (2020). Support vector regression. In *Machine Learning*, pages 123–140. Elsevier.

Zhang, X. and Tan, Y. (2018). Deep stock ranker: A lstm neural network model for stock selection. In *International Conference on Data Mining and Big Data*, pages 614–623. Springer.