LUT UNIVERSITY
School of Business and Management
Strategic Finance and Business Analytics

Juha Kapanen

DAX INDEX PRICE PREDICTION USING ARTIFICIAL NEURAL NETWORKS

Master's Thesis
2020

1st Examiner: Professor Pasi Luukka
2nd Examiner: Postdoctoral Researcher Christoph Lohrmann

Abstract

Author:	Juha Kapanen
Title:	DAX index price prediction using artificial neural networks
Faculty:	LUT School of Business and Management
Master's program:	Strategic Finance and Business Analytics
Year:	2020
Master's thesis:	46 pages, 13 tables, 14 figures
Examiners:	Professor Pasi Luukka
	Postdoctoral Researcher Christoph Lohrmann
Keywords:	ANN, technical analysis, efficient market hypothesis, weekday anomaly, DAX

Artificial neural networks (ANNs) have become widely popular in the field of machine learning and are also increasingly used in the financial field, especially in asset price prediction. This thesis studies if the price change of an asset can be predicted with ANNs that use historical prices, technical analysis indicators, and weekday information as predictive variables. Study uses Deutscher Aktien IndeX 30's (DAX) daily data between 2009 and 2018 to test if the direction of next day's return can be predicted and if the weekday variables have any effect on the prediction performance. The results show that ANNs predict the test data returns better than random guessing, but these results cannot be proven to be statistically significant. An investment strategy built on the tested model can yield better investment returns than buy and hold strategy when transaction fees are not counted for. Study also shows, that removing the weekday variables as predictors lowers the model's prediction accuracy. The study cannot provide undisputed answer whether the used model or some variation of the model could be used to predict asset price changes significantly better than random guessing.

Tiivistelmä

Tekijä:	Juha Kapanen
Aihe:	DAX indeksin arvon ennustaminen neuroverkkojen avulla
Tiedekunta:	LUT School of Business and Management
Pääaine:	Strategic Finance and Business Analytics
Vuosi:	2020
Pro Gradu:	46 sivua, 13 taulukkoa, 14 kuviota
Tarkastajat:	Professori Pasi Luukka Tutkijatohtori Christoph Lohrmann
Hakusanat:	ANN, tekninen analyysi, markkinoiden tehokkuus, viikonpäivä anomalia, DAX

Neuroverkot ovat kasvattaneet suosiotaan koneoppimisessa ja niiden käyttö on lisääntynyt myös rahoitusalalla, etenkin arvopapereiden ja muiden omaisuuserien arvon ennustamisessa. Tässä tutkielmassa testataan voiko neuroverkoilla ennustaa pörssi-indeksin arvonmuutosta käyttäen ennustavina muuttujina sen historiallisia hintoja, teknisen analyysin indikaattoreita sekä viikonpäivämuuttujia. Tutkielmassa käytetään Deutscher Aktien IndeX 30 (DAX) pörssi-indeksin hintahistoriaa vuosilta 2009-2018 selvittääkseen, onko seuraavan päivän hinnan muutoksen suuntaa mahdollista ennustaa, ja onko viikonpäivämuuttujien mukanaololla vaikutusta ennustetarkkuuteen. Tulokset osoittavat, että neuroverkko pystyy ennustamaan testiaineiston hinnanmuutoksia paremmin kuin satunnainen arvaus, mutta tulokset eivät ole tilastollisesti merkittäviä. Testatun mallin päälle rakennettu sijoitusstrategia pystyy parempiin tuottoihin kuin osta ja pidä -strategia, kun transaktiokustannuksia ei huomioida. Tulokset osoittavat myös, että viikonpäivämuuttujien poistaminen mallista vähentää sen ennustustarkkuutta. Tutkielma ei pysty kiistattomasti osoittamaan pystyvätkö käytetyt mallit tai niiden muunnelmat ennustamaan omaisuuserien arvon muutosta paremmin kuin satunnainen arvaus.

Acknowledgements

I would like to thank Pasi Luukka for the guidance and comments throughout the writing process, and Christoph Lohrmann for giving me valuable comments during the finalization of this thesis.

I want to also thank Mikael Collan, Sheraz Ahmed and other faculty members for the excellent education and variety of valuable skills I acquired at LUT.

Thank you for all my great friends at LUT. You made the time at Lappeenranta truly fun and enjoyable.

Table of contents

List of figures

List of tables

List of abbreviations

ANN	Artificial Neural Network
TP	True Positive
FP	False Positive
TN	True Negative
FN	False Negative
TI	Technical Indicator
MLP	Multilayer Perceptron

1 INTRODUCTION

Technical analysis, an attempt to predict future asset prices based on their historical prices, is widely popular and many individual and institutional investors are investing according to it (Kirkpatrick & Dahlquist, 2016). There are variety of information available about different strategies and measures that are claimed to be able to beat the market return and predict price movements. Many studies have been made on the subject with different frameworks, but there is not a clear conclusion, whether it is possible or not to use past prices as predictor for future prices (Nazario et al, 2017).

Artificial Neural Networks (ANNs), a form of machine learning algorithms that are inspired by a biological brain, have been around since mid-1900s (Kleene, 1956), but have recently grown more popular, because of the abundancy of data sources and advancements in computing technology (The Economist, 2020). They have also found their way into the field of financial analysis. ANNs' have proven their capabilities in pattern recognition (Bishop, 2006, 9), so they could possibly be used to find patterns in market prices that traditional technical analysis tools cannot recognize and are therefore a potential tool to predict market price movements and reach abnormal investment gains.

1.1 Motivation and background

Technical analysis in stock market prediction is very debated subject, and lot of research can be found supporting it or against it. Effectiveness of technical analysis depends highly on used technical indicators (TIs), studied target markets, and chosen time periods. Research reviews on the subject point out that technical analysis can be used to predict stock market movements in certain cases. Most of the studies to be found test TIs as individual indicators and do not test predictive powers of combinations of these indicators. (Nazario et al, 2017)

ANNs are a powerful tool for combining large amount of inputs to one prediction output. They are self-learning, which means that they can be used to develop predictive models by using historical data as a training data. ANN's prediction accuracy is dependent on many aspects, such as chosen input variables, network structure, learning rate, chosen activation functions and amount and quality of training data. (Bishop, 2006, 225-241)

This study was done to find out whether ANNs can be used to create model with multiple different TIs and other historical price related variables to create a predictive model that could outperform market return. Because ANNs are considered as black boxes, and it is difficult to interpret the mathematical models inside them (Medium, 2020), this study was more focused on measuring the results and performance of the ANN on the selected target market. Earlier researches have studied subjects that are closely related to this study, but this research takes a new look into the subject with more varied collection of different predictive variables.

As literature review in chapter 2.2 will show, most of the research on technical analysis and ANNs use US markets as an empirical test market. There is also research done on European markets and smaller developing markets, but they are less represented in the studies. Because of this, less studied German DAX index (Deutscher Aktienindex) was chosen as subject market in this research.

1.2 Objectives

The goal of this research is to test whether ANNs can be used to predict daily price changes of DAX index with multiple different technical analysis indicators and other variables related to historical price information, and can the model be used to gain excess returns when compared to the market return. Objectives can be divided into following research questions:

Q1: Can ANNs predict the direction of daily return of DAX index better than random guessing?

Past research has mixed results for asset price prediction with technical analysis and machine learning models. This research tries to answer that question with new selection of models and predictive variables. ANNs usually perform well in classification tasks (Bishop, 2006, 225-241). Therefore, to simplify the classification task, only binary outcome is predicted, whether the next trading day's return is positive or negative.

Q2: What are the daily returns on predicted positive and negative days?

Besides the overall binary prediction accuracy of the model, the mean returns of the predicted days can be used to evaluate its predictive power. Possible prediction inaccuracies in the days with small price movements can possibly be compensated with higher prediction accuracies in days with high positive or negative returns. Therefore, the mean returns give a better understanding on the models' overall predictive power.

Q3: Can ANN investment strategy beat buy and hold strategy for DAX index?

In addition to measuring the ANNs' ability to predict returns, ANNs' investment performance is also measured. The goal of technical analysis is to predict market movements and use that information to gain abnormal returns on the markets. Even if the model is better in prediction than random guessing, it does not guarantee abnormal returns. Market return can be achieved just by executing buy and hold strategy, so the ANNs' predictions are used as an investment strategy and benchmarked against buy and hold strategy to see whether abnormal returns can be achieved.

Q4: Is ANN's performance better with weekdays as predicting variables?

Earlier research suggests that daily market returns, and variance differ between weekdays (Fama, 1965; Cross, 1973). This research uses weekdays as a part of the ANN models' predicting variables to achieve more accurate results. Same model is also trained and tested without weekday variables to see if they cause the model to be more accurate or not.

1.3 Structure

This thesis has following structure: Chapter 2 is the theoretical framework, which first describes the key theoretical concepts regarding this research. Chapter 2 also presents some of the previous literature around the same research subject and chosen variables in this research. In chapter 3, selected data and methodology is described in more detail. Chapter 4 is the presentation of the results, which are drawn to conclusion in chapter 5. Chapter 5 also goes through the limitations of this research and possible future research that could be derived from this research.

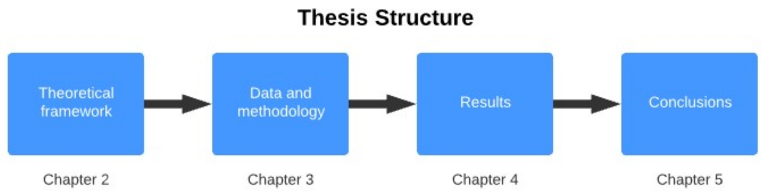

Figure 1. Thesis structure

2 THEORETICAL FRAMEWORK

Theoretical framework of this research is divided into two parts: First part presents the most important theories regarding this research and discusses the validity of those theories based on earlier research. Second part presents earlier empirical research done on technical analysis and ANNs on asset price prediction. Past research has given mixed results on efficiency of different TIs and machine learning algorithms in different financial markets, so the selected research was reduced to TIs and ANNs that have found to be most successful in asset price prediction.

2.1 Underlying theories and methods

2.1.1 Efficient market hypothesis

Efficient market hypothesis (EMH) is a widely studied financial theory first presented by Fama (1970). It states that all the available information is counted into prices of publicly traded assets. Therefore, it presents that it is impossible gain abnormal returns or "beat the market". Current price of an asset should always be a fair price, because all the available information is calculated into the price, and the price of an asset should only be affected by new information. (Fama, 1970)

EMH consists of three variants of market efficiency: weak, semi-strong and strong form. The most relevant form of market efficiency regarding this study, is the weak form efficiency. EMH states that for the markets to be weak-form efficient, market prices cannot be predicted by their historical price information, because the past prices do not affect the future prices, and all the available information is already priced in to the asset value. Therefore, the current asset price should be the best prediction for the future asset price. This theory is also called the random walk theory. (Fama, 1970).

Random walk theory is widely studied, and very debated subject. Academics and economists have not reached consensus whether the theory holds or not. Fama (1970) reviewed the contemporary empirical and theoretical studies on the subject and reached a conclusion that price history can be used predict asset price movements to some extent, but there was not enough overall evidence to disprove the random walk theory. More recent study by Titan (2015), also collects multiple empirical studies around the subject. Titan's conclusion is that precise results have not been reached and new research and models are needed to reach conclusion on the matter.

2.1.2 Weekday anomaly

Lot of research has been done around the subject of weekdays and stock returns. There is some empirical evidence that the weekday may affect the daily returns of stock markets. This effect is called weekday anomaly or the weekday effect. There is no clear explanation for this effect. Dicle & Hassan (2007) present a theory that the effect is purely psychological. Investors tend to be more optimistic on Fridays, and pessimistic on Mondays.

Fama (1965) tested the random walk theory with various ways to see if any patterns could be retracted from daily stock return data. He analyzed daily returns of all the stocks in Dow-Jones Industrial Average from the end of 1957 to September 1962. He found out that Monday's variance is 22% higher than other weekdays.

Cross (1973) showed that Standard and Poor's Composite Stock Index's returns has had considerably higher returns on Fridays than Mondays. Between the year 1953 and 1970 there were more Fridays when the index increased than Fridays when the index decreased. On Mondays results were opposite. Average returns on Fridays and Mondays were 0.12% and -0.18%, respectively. In 1981, Gibbons and Hess reached similar result when researching weekday effect on S&P 500 Index, and Center for Research in Security Prices' value and equally weighted portfolio indexes. Between 1962 and 1978 all the indexes' daily returns proved not be equally distributed by weekdays. Mean return on Mondays was negative on every tested dataset, while weekdays from Wednesday to Friday had always positive mean returns.

More recent studies of weekday anomaly have given more of a mixed result. Linden and Louhelainen (2006) did a wide study where they analyzed 18 different stock indexes' weekday returns from 1990 to mid-2003 with ordinary least squares (OLS) and minimum absolute deviation (MAD) estimators. With OLS method only two of the 18 indices showed unequal weekday returns. However, MAD method resulted in eight of 18 indices having unequal returns.

2.1.3 Technical analysis

Technical analysis, the statistical analysis of the past asset price and investment activities, is a widely used methodology when making investment decisions in financial markets (Menkhoff, 2010). Technical analysis uses variety of different indicators regarding past market activities such as trading volume, price history, volatility et cetera. Usefulness and value of technical analysis as an investment decision making tool is not undisputed. Technical analysis' ability to predict asset prices, is dependent on the validity of the EMH. Technical analysis bases its predictive power purely on past price information, so if the weak form EMH holds, technical analysis should not have any value in asset price prediction.

Nazario et al (2017) reviewed 85 technical analysis studies from years 1959 to 2014. Research was limited only to studies which covered technical analysis performance on stock markets. These studies used variety of methods from moving averages to neural networks, and covered markets all around the world. They concluded that majority of the studies supported technical analysis. However, they pointed out that many of the studied papers were researching technical analysis in emerging markets, which had major positive effect on the results. Emerging markets are usually less efficient and therefore more subject to technical analysis, which is also supported by research done by Park et al (2007). Nazario et al (2017) also considered that there might be a publishing bias, caused by researchers' unwillingness to publish results, which support EMH, and therefore do not add any new insights on the matter.

2.1.4 Artificial Neural Networks

Artificial neural networks are computing systems, which originate from attempts to model information processing of biological systems. They are based on a collection of nodes, which are loosely based on biological neurons. There are various types of neural networks, but multilayer perceptron (MLP) proved to have the greatest practical value for overall statistical pattern recognition. Multilayer perceptron (Figure 2) contains multiple layers of logistic regression models. It consists of nodes in different layers: input layer, hidden layers and output layer. These nodes are connected to other nodes in previous or following layers. Nodes take inputs and transform those inputs using activation functions. The produced outputs are fed to the nodes in the next layer. MLPs are limited to only feed outputs forwards to the next layers, unlike recurrent neural networks (RNN), which can also pass the signals sideways. (Bishop, 2006, 225-241)

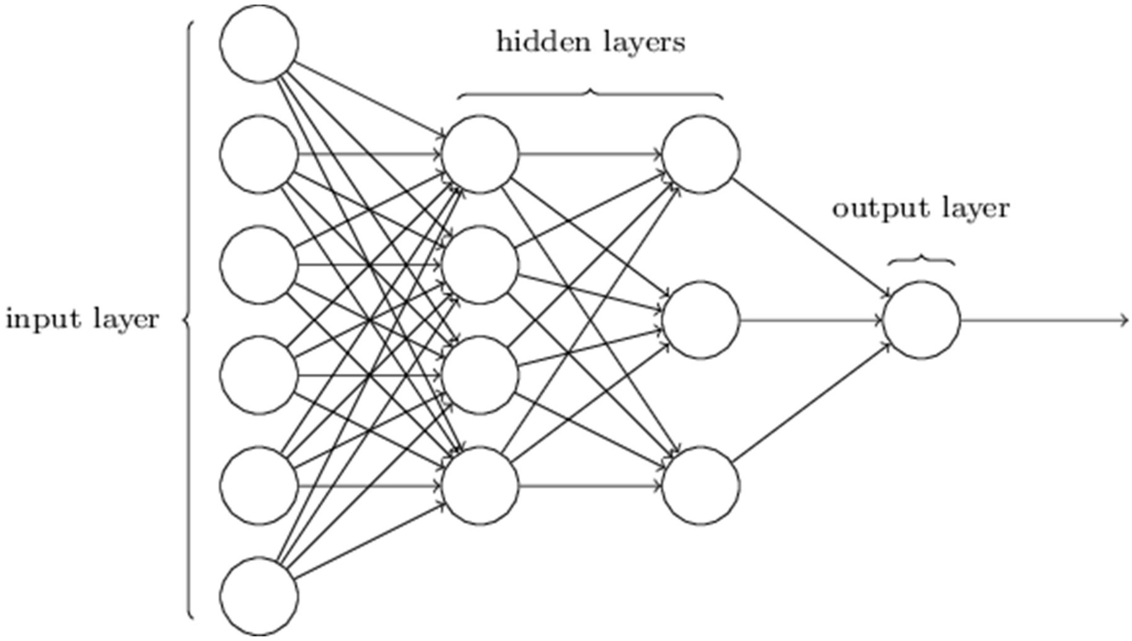

Figure 2. Four-layer artificial neural network

The basic three-layer neural network model can be described by following function (1):

$$y_k(\mathrm{x}, \mathrm{w}) = \sigma\left(\sum_{J=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)}\right) + w_{k0}^{(2)}\right) \tag{1}$$

In the first layer of the network, M linear combinations of the input variables x_i are constructed as activations with weights w_{ji} and biases w_{j0}. These activations are transformed with nonlinear activation functions h(x). Activation functions are generally sigmoidal functions such as logistic sigmoid or the tanh function. Activation functions' outputs are then fed forward as output layer activations, where w_{kj} are weights and w_{k0} biases. These activations are again transformed with activation function σ(x), to give the output or set of outputs y_k. (Bishop, 2006, 225-241)

MLP uses backpropagation to adjust the weights of the nodes. MLP is trained with training data, and the weights are adjusted to minimize the squared error function (2), where y is the network output and t_n is the target value. (Bishop, 2006, 225-241)

$$E(\mathrm{w}) = \frac{1}{2}\sum_{n=1}^{N} (y(x_n, w) - t_n)^2 \tag{2}$$

MLP adjusts the weights and goes through multiple iterations of learning until the error function is minimized. Of the more commonly used methods of backpropagation is gradient decent method, where gradient, the derivative of the squared error function with respect to the weights of the network is calculated. Gradient is then multiplied with chosen learning rate η to adjust the weights (3). (Bishop, 2006, 225-241)

$$w^{(\tau+1)} = w^{(\tau)} - \eta \nabla E\left(w^{(\tau)}\right) \tag{3}$$

Network's structure is a key part of the network's predictive capabilities. The number of layers and nodes in each layer can considerably change the outcome of the predictions. Size of the network affects the complexity, learning times and the generalization abilities of the network. More complex structures might cause the network to be too overfitted and thus be bad at predicting the results outside of the training data. Therefore, determining the correct network structure is essential for creating working predictive model. (Kavzoglu, 1999)

There are no preset rules to determine the best structure for the network. Only way to find the best performing structure is to test the network with different structures. One approach is to first create a simple network and increase its size, adding more hidden nodes and layers until the most accurate results are achieved. These are called constructive techniques (Hirose et al, 1991). Another approach is to start with a large network and make it smaller by removing nodes and interconnections between nodes that are not participating in the solution. This method is called pruning. (Kavzoglu, 1999)

2.2 Earlier Literature

This chapter presents some of the earlier studies on chosen technical analysis indicators, and on ANN based asset price prediction. Each part includes both earlier studies on the subjects, and more recent studies, as the pricing characteristics of financial markets may have changed through years due to globalization and technological advancements.

2.2.1 Trading Volume

The trading volume of a stock or an index is believed to affect the price of the asset. Common believe is that when the trading volume is high, the price of the asset usually increases. Some of the theories suggest that it is caused by optimistic investors' strong reactions to positive news, causing the demand for assets to increase strongly, while reactions to negative news are weaker. (Karpoff, 1987)

Karpoff (1987) did a literature review of earlier research on the subject. His review consisted of studies done with different markets and asset classes, while most of them being stocks or indexes in US stock markets. Time intervals of the studied price changes varied between hourly and yearly. Used sample periods varied in length from days to years, all being from period between 1939 and 1983. In his review, 12 studies showed support for positive correlation between the price change and trading volume, while 4 studies showed that there is no positive correlation.

Hutson et al. (2008) examined the price and volume correlation with more recent data. They studied 11 different international stock markets. They collected daily and monthly data from various periods between January 1980 and August 2004. They analyzed the relation between mean, variance and skewness of the returns and trading volume, using single equation and vector autoregression VAR models. Their single equation model showed that higher trading volumes cause negatively skewed returns in 8 markets.

2.2.2 Moving average oscillator

Moving average (MA) oscillator generates buy and sell signals based on two different moving averages, a long period and a short period moving average of the asset price. In this strategy, investor buys assets when the short-term moving average rises higher than the long-term moving average and sells or shorts assets when the short-term moving average dips below long-term moving average. When short-term average moves above or below the long-term average, it is seen as signal that there is an increasing or decreasing trend in an asset value. Moving average rule can vary based on the length of the short-term and long-term period. (Brock et al., 1992)

Brock et al. (1992) studied profitability of MA oscillator. Analysis was done using Dow Jones index prices from 1897 to 1986. Rules were tested using the entire period, as well as different subsamples of time. Following short-term-long-term rules for MA oscillator were tested: 1-50, 1-150, 5-150, 1-200, and 2-200. They used two variations of this strategy Variable-Length Moving Average (VMA) and Fixed-Length Moving Average. In VMA, they bought or short sold assets every time short-term average crossed long-term average, in FMA they held long or short position for fixed 10-day period when signal was created.

Results showed that these technical trading rules were effective in predicting stocks' price changes. For example, the daily average return for all VMA strategies were 0.042% for buy-signals, and –0.025% for sell signals, while unconditional return for the whole period being 0.017% per day. Six of the ten different variations were statistically significant at 5%. (Brock et al., 1992)

There have been many other studies where MA rules have been found to have predictive powers. For more recent example, Metgalchi et al (2008) tested profitability of MA rules on Swedish stock index OMX Stockholm 30. They tested various MA rules with daily logarithmic returns from 2.1.1986 to 13.9.2004. Rules were tested with different lengths of long and short periods, as well as different variations of the MA rule itself. The tested variations were standard moving average rule (SMA), increasing moving average rule (IMA), and Arnold and Rahfeldt moving average rule (ARMA).

Every variation of the rule had better daily return than buy and hold strategy, although the transaction costs were not considered in the analysis. Average daily returns of the different variations of the SMA, IMA, and ARMA strategies were 0.0012, 0.00091 and 0.00157 respectively, while buy and hold strategy had only 0.00041 average daily return. All the mentioned averages were significant at 5% level. (Metgalchi et al, 2008)

2.2.3 RSI

Relative Strength indicator (RSI) is a technical analysis indicator which calculates the relative strength of an asset price development by calculating the ratio between last 14 days' average returns for up and down periods. Higher RSI is considered as signal that asset value has increasing trend, while lower value is seen as decreasing trend. More precise definition is shown in this paper in Figure 6. (Wilder, 1978)

Chong and Ng (2008) tested profitability of RSI trading rules using data of the London Stock Exchange FT30 Index from years 1935 and 1994. They tested a trading rule where sell signal was generated when RSI moved below 50, and buy signal was generated when RSI rose above 50. Results had some variance when compared between different time periods, but overall results showed that the trading rules could be used to gain abnormal returns. In the whole period 296 buy signals and 311 sell signals were created, and the mean 10-day return after buy signals were 0.779% and -0.127% after sell signals.

Chiang et al (2012) tested the predictive power of eight different technical analysis trading strategies and compared them to buy and hold strategy. To test the strategies, they used Taiwan's TAIEX Stock Index data from 21.7.1998 to 21.7.2008. The strategies took long and short position in TAIEX stock index futures. The average returns were calculated for 1-month and 3-month investment horizons. Tested strategies were candlestick pattern, directional movement index (DMI), moving average (MA) crossover, moving average convergence divergence (MACD) crossover, parabolic, relative strength index (RSI) oscillator, trend-lines automatic and volatility expansion. 7 out of 8 strategies gained better mean returns on 1-month investments horizon, and all strategies were better than buy and hold on 3-month investment horizon. Parabolic and RSI oscillator were superior when compared to other strategies. Parabolic strategy gained 2.71% and RSI oscillator gained 2.55% 1-month mean returns in the 3-month horizon before transaction costs.

2.2.4 Neural networks in stock market prediction

In one of the earlier financial ANN studies, Kim (2003) tested support vector machine's (SVM) performance in stock price prediction against back propagating ANN and case-based reasoning. He tried to predict the direction of the daily price change using 12 different technical analysis indicators as variable. Most of the indicators were related to momentum or the moving average of the asset price. Research used data from Korea composite stock price index (KOSPI), between the years 1989 and 1998. The best performing SVM model achieved an accuracy of 57.8% when using the testing dataset, while ANN had 54.7% accuracy at best. CBR had an accuracy of 52.0%.

Guresen et al (2011) studied and compared the time-series prediction accuracy of four different types of neural network models: MLP, dynamic artificial neural network (DAN2) and hybrid models of previous two which used generalized autoregressive conditional heteroscedasticity (GARCH) to extract new input variables. They used all four models to make predictions of daily price changes of NASDAQ index and compared the mean squared errors (MSE) and mean absolute deviation (MAD) of each model. Their models used NASDAQ's price data from October 7, 2008 to June 26, 2009. First 146 days were used as training and cross-validation data, the rest 36 days were used for testing. The MLP model performed best of all four models. MLP's MAD was 2.324%. The hybrid methods added did not improve the basic MLP model's performance.

Kara et al (2011) studied ANN's and SVM's prediction capabilities by trying to predict the direction of daily price change of Istanbul Stock Exchange's National 100 (BIST 100) Index. In their research, they used data from year 1997 to 2007. Data was divided it into equal size of training and holdout datasets. They used equal amounts of data points from each year for training and holdout data. Their models used 10 different technical analysis indicators as predictors. These indicators were simple 10-day moving average, weighted 10-day moving average, momentum, stochastic K%, stochastic D%, RSI (Relative Strength Index), MACD (moving average convergence divergence), Larry William's R%, A/D (Accumulation/Distribution) Oscillator, CCI (Commodity Channel Index). Using these variables, they found that their three-layered ANN had better prediction rate than the SVM. ANN's accuracy was 75.74%, and SVM's was 71.52%.

Qiu et al (2016) used ANN to predict the direction of daily price change of Nikkei 225 index. They used genetic algorithm (GA) to optimize ANN's initial weights and bias values. Studied period was from January 2007 till December 2013. First 78.6% of data was used for training and last 21.4% was used as test data. They used two different sets of variables which each had their own type of technical input variables. First set had 13 more commonly used variables, similar to what Kara et al (2011) used. Second set had 9 technical variables, such as moving average of price and average returns from previous days. Variables were normalized so that all feature components fitted in specified range. Their best model achieved 81.27% accuracy using the second set of 9 input variables.

Previous studies that have used ANNs in stock market asset value prediction provide variety of evidence that ANNs could be used to successfully to predict changes in asset prices. Therefore, ANNs were selected as the predictive model in this research.

3 DATA AND METHODOLOGY

3.1 Data

All the data used in this research is derived from daily closing prices of Deutscher Aktienindex (DAX), which is a German blue-chip stock market index. It consists of 30 major German companies traded in the Frankfurt Stock Exchange, with total market capitalization over one trillion euros (Dax-indices.com, 2020). DAX was chosen, because DAX is very liquid and highly traded European index, and most of earlier research focusses on US indexes or other assets. DAX's daily close and volume values were obtained from Yahoo Finance, which was also used as a data source in many of the previous ANN researches. DAX price development from 30.12.1987 to 22.7.2019 can be seen on the Figure 3.

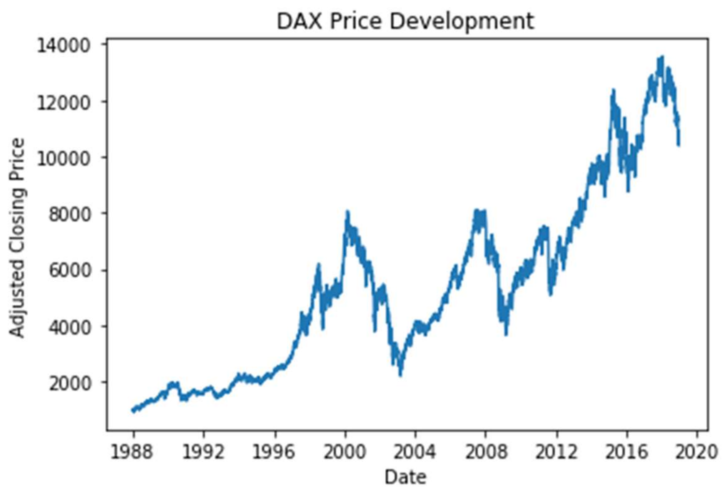

Figure 3. DAX index price development

Most of the previous researches using machine learning have used data from one to ten-year period, and the data they have used has usually been very recent. This strategy has often led to great results. Characteristics of stock market change through the bull and bear cycle, so it is important that the data captures current market characteristics. Market dynamics and reaction times have also been changing due to technical innovations, such as developments in algorithmic trading and increase of online trading (Ft.com, 2020). Therefore, old market data might not be able to show same relationships between the inputs and independent variables as contemporary data. It is also important, that the time period of the data is long enough to include multiple shorter rise and decline periods. If the ANN is trained with data which has only one long trending rise or decline, it might not be able predict longer opposite trends so well. Because of all the previously mentioned reasons, this study uses data from 1.1.2009 to 31.12.2018. This ten-year period starts at the end of the 2008 market crash and consist of long bull cycle with three shorter bear cycles in between. The bear cycles being at the second half of 2011, from the mid-2015 to the start of 2016, and the whole year 2018. Price development and trading volume of DAX in the selected period, can be seen on Figures 4 and 5, respectively.

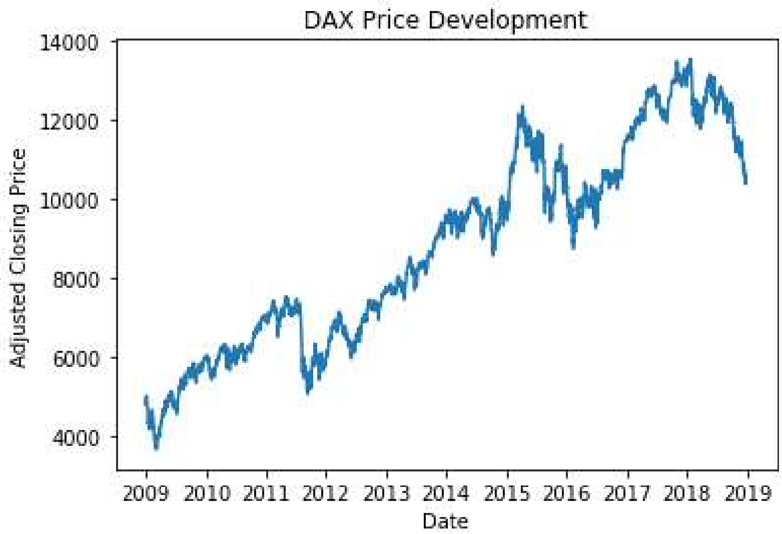

Figure 4. DAX index price development for selected period

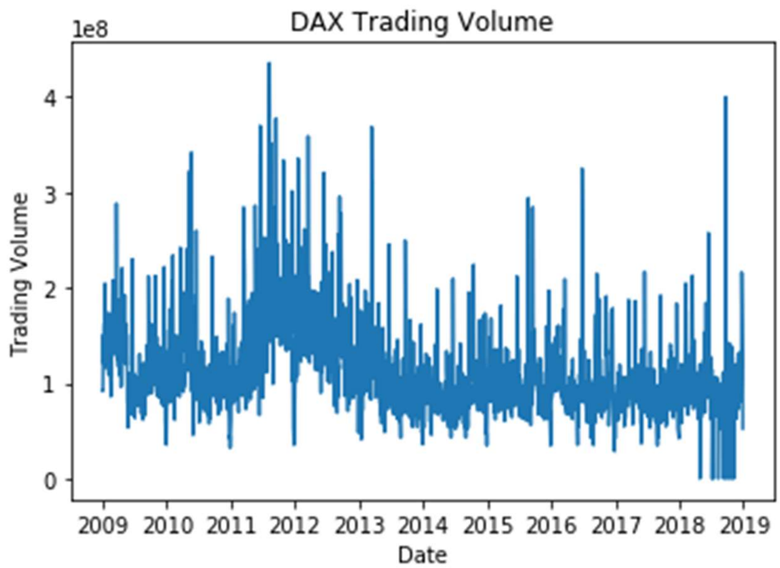

Figure 5. DAX index trading volume for selected period

Data contains 2533 observation dates, from 2.1.2009 to 27.12.2018. Observations do not include weekends or holidays, because stocks are not being traded in those day, due to Deutsche Börse being closed. First 80% (2026 observations, 2.1.2009-22.12.2016) of the data will be used for training the ANN, the last 20% (507 observations 23.12.2016-27.12.2018) will be used as test data. This allocation should have enough data to train the network adequately, while still having enough test data to minimize the effect of randomness in results. Training data is also split in to five subsamples (n=405) for cross-validation purposes. Feature statistics of the daily returns of whole dataset, training data, and test data, can be seen in Table 1. Distribution of daily returns are presented in Figure 6.

Training and testing datasets have some major differences in their feature statistics. Most importantly, their mean returns differ noticeably. Training dataset has around 0.0506% positive mean daily return, while testing dataset has negative mean return of -0.0125%. Training dataset has also 64.5% higher standard deviation, and its minimum and maximum values are about two times as large as testing dataset's comparable values. Large differences in features are not optimal, there is a chance that models will overfit the training data and are not able to make accurate predictions for the lower variance testing data.

Returns	All (n=2533)	Train (n=2026)	Test (n=507)
Positive (n)	1351	1091	260
Negative (n)	1182	935	247
Mean	0.000380	0.000506	-0.000125
Std	0.012840	0.013733	0.008348
Skew	-0.155031	-0.164706	-0.197296
Kurtosis	2.444991	2.060114	1.239210
Min	-0.068233	-0.068233	-0.034755
Max	0.060723	0.060723	0.033731

Table 1. Feature statistics of daily returns.

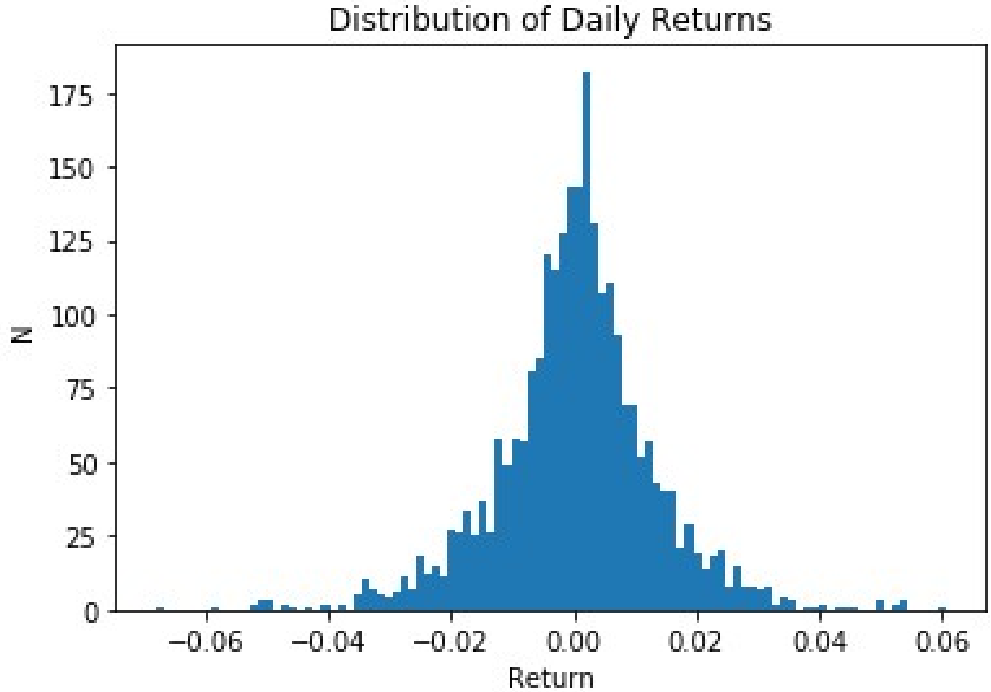

Figure 6. Distribution of daily returns

The predictive variables used in this research have some proven predictive capabilities when used with ANNs, or they correlate with daily returns in other technical analysis researches. Chosen variables were: 150 day moving average signal, 14-day RSI, last ten lags of daily returns, last 10 lags of daily volume and 5 weekday dummy variables. Variety of different types of variables was chosen to see if the combination of these variables can have better predictive capabilities than they have alone. Each of them induces different information to the model. Lags of returns and volume give unprocessed information from previous days' trades, while 150-day moving average and 14-day RSI represent the longer period price development and momentum. Day of the week dummies for trading weekdays were included to capture the possible effects from day of the week anomaly. ANN is also trained and tested without day of the week variables to see if these variables give any performance benefits. Detailed explanations of the calculations of each variable can be seen in Table 2.

Because all the input variables besides volume and weekdays are technical variables, that are based on the historical prices and price movement of DAX index, all the variable values can be created from the same dataset. This dataset contained daily historical closing prices and trading volume of DAX. The predicted variable values – directions of the price change of the DAX daily closing price – were also created from the same dataset.

Variable	**Formula**
Trading volume at t	V_t
Trading volume at t-1	V_{t-1}
…	…
Trading volume at t-9	V_{t-9}
DAX return at t	$\frac{C_t - C_{t-1}}{C_{t-1}}$
DAX return at t-1	$\frac{C_{t-1} - C_{t-2}}{C_{t-2}}$
…	…
DAX return at t-9	$\frac{C_{t-9} - C_{t-10}}{C_{t-10}}$
150-day moving average*	$\frac{\sum_{i=0}^{150} C_{t-i}}{150}$
14-day RSI**	$100 - \frac{100}{1 + \frac{\sum_{i=0}^{13} Up_{t-i}}{14} \Big/ \frac{\sum_{i=0}^{13} Dw_{t-i}}{14}}$
Monday dummy variable***	D_{mon}
…	…
Friday dummy variable	D_{fri}
C is the closing price of the DAX index at the day t. V is the total trading volume of the DAX at the day t. *150-day moving average will be converted to a buy or sell signal. Buy signal will be given when the current price is above moving average and sell signal when the current price is below the moving average. **When daily return is positive, Up=1 and Dw=0. When daily return is negative, Up=0 and Dw=1. *** If weekday at t+1 is Monday then D_{mon} = 1, otherwise D_{mon} = 0.	

Table 2. Variables

For the ANN to achieve better results, input variables are normalized. Normalization ensures that different inputs with largely different values are in order unity, which results the initial weights in network to also be in order unity. (Bishop, 1999, 298-300) Most variables were normalized to numbers between -1 and 1. Daily returns were divided by the maximum absolute daily return of the time series, so that negative returns were all between -1 and 0, and positive returns were between 0 and 1. Volumes were normalized to from 0 to 1 by dividing the volume by the maximum volume in the dataset. 150 day moving average was converted to buy or sell signals, buy signal being 1, and sell signal being 0. 14-day RSI was converted linearly to numbers between -1 and 1, so that value 0 converts to -1, 50 to 0 and 100 to 1. Weekday dummy variables were 1 when the predicted date is the weekday in question, and otherwise 0.

3.2 Neural Network Structure

The hidden layers' nodes will use Rectified Linear Unit (ReLU) functions as activation functions (Figure 7). ReLU function has become the default activation function in most types of ANNs (Brownlee, 2020). When compared to other activation functions it has many advantages, such as computational simplicity, ability to output true zero values, and linear behavior, which makes it easy to optimize (Glorot et al., 2011). Output layer will use standard logistic sigmoid activation function, which will give results between of 0 and 1. Output of 0.5 or higher will be interpreted as price increase signal, and output below 0.5 as price decrease signal.

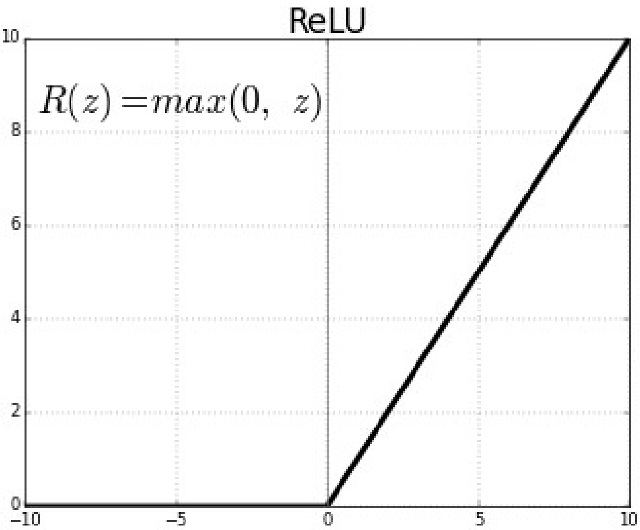

Figure 7. ReLU function.

ANN used Adam stochastic gradient-based optimizer solver proposed by Diederik Kingma and Jimmy Ba (2020). It is the default solver used in scikit learn MLP classifier in Python. Maximum number of iterations in training phase was raised from default of 200 to 10000 to avoid underfitting. Otherwise ANN settings were left on default, except for the ANN structure.

ANN's topology will a have substantial effect on its performance. To find a structure which achieves the best predictive performance, network will be trained and tested multiple times with same variables using different network structures. Previous studies suggest that a structure with only one hidden layer should be the best option. All researches referenced in the literature review Qiu et al (2016), Kara et al (2011), Guresen et al (2011) and Kim (2003) used three layered network structures. There are no certain rules for how many nodes should hidden layers in ANN have. In this research, three different three-layer fully connected feedforward structures were tested: 27-27-1, 27-50-1 and 27-100-1. These different models are from here on forward referred as ANN 1 (27-27-1), ANN 2 (27-50-1), and ANN 3 (27-100-1)

3.3 Evaluation

Model is evaluated with out-of-sample test data using usual binary classifier evaluation measures such as accuracy, sensitivity and specificity. These characteristics are reported in confusion matrix. Receiver operating characteristic (ROC) is also used to analyze results. Mean daily returns of predicted positive and negative return days are also examined. Those returns are compared to the mean daily returns of the entire test dataset. The best performing network topology's predictions is also simulated as an investment strategy against buy and hold strategy.

To test how well the results can be generalized, there is also a cross-validation. Training data is divided into five subsets, then each subset alone is used as test data, while other four are used as training data. Averages of these five runs are reported in confusion matrix and in mean daily returns. Cross-validation is done separately for all three different network topologies.

The predictive power of weekday variables is evaluated by training and testing the best performing ANN topology again without the weekday variables. The model is evaluated with all the same methods as the three baseline models. The results of this model are reported in their own section.

4 RESULTS

4.1 Confusion Matrices

First evaluation method was confusion matrix, which shows the amount true positive (TP), false positive (FP), true negative (TN) and false negative (FN) predictions. These numbers are used to calculate accuracy, precision, negative predictive value, sensitivity, and specificity of the ANNs. Formulas for these measures can be seen in Table 3. Confusion matrices for the ANNs tested with out-of-sample dataset are presented in Tables 4-6.

	Predicted Positive	**Predicted Negative**	
Actual Positive	TP	FN	**Sensitivity:** TP/(TP + FN)
Actual Negative	FP	TN	**Specificity:** TN/(TN + FP)
	Precision: TP/(TP + FP)	**Neg. Pred. Value:** TN/(TN + FN)	**Accuracy:** (TP + FN)/ (TP + FN + FP + FN)

Table 3. Confusion matrix formulas

27 Hidden Nodes	**Predicted Positive (n=336)**	**Predicted Negative (n=171)**	
Actual Positive (n=260)	TP: 184	FN: 76	Sensitivity: 70.8%
Actual Negative (n=247)	FP: 152	TN: 95	Specificity: 38.5%
	Precision: 54.8%	Neg. Pred. Value: 55.6%	Accuracy: 55.0%

Table 4. Confusion matrix for ANN 1

50 Hidden Nodes	Predicted Positive (n=323)	Predicted Negative (n=184)	
Actual Positive (n=260)	TP: 172	FN: 88	Sensitivity: 66.2%
Actual Negative (n=247)	FP: 151	TN: 96	Specificity: 38.9%
	Precision: 53.3%	Neg. Pred. Value: 52.2%	Accuracy: 52.9%

Table 5. Confusion matrix for ANN 2

100 Hidden Nodes	Predicted Positive (n=347)	Predicted Negative (n=160)	
Actual Positive (n=260)	TP: 186	FN: 74	Sensitivity: 71.5%
Actual Negative (n=247)	FP: 161	TN: 86	Specificity: 34.8%
	Precision: 53.6%	Neg. Pred. Value: 53.8%	Accuracy: 53.6%

Table 6. Confusion matrix for ANN 3

Differences between ANNs were relatively small. All models had accuracy between 52.9% and 55.0%. Precision and negative predictive value of each model were between 53.3% and 55.6%. Sensitivity and specificity had larger margin between models, although still being relatively similar. Sensitivities were between 66.2% and 71.5%, and specificities between 34.8% and 38.9%.

Best model (ANN 1) having an accuracy of 55% indicates that none of the models are very good at accurately predicting whether the next day’s return is positive or negative. As random selection should result in 50% accuracy in long term, 55% accuracy could be explained by variance resulted from limited number of observations in the testing dataset. Sensitivity and specificity values seem to be more deviated from randomness. High sensitivity and low specificity in all ANNs were a result of them predicting larger count of positive than negative returns. Count of positive predictions being between 63.7% and 68.4% from the test data sample size.

ANN 1 performed best overall from the three different models. It had the highest accuracy, precision, and negative predictive value. It did not have the best sensitivity or specificity between models but had the best combined result with 70.8% sensitivity and 38.5% specificity.

Cross-validation of the models did not provide any support for their predictive capabilities. All the models had an average accuracy under 50% in the cross-validation tests. Negative predictive values were very low, being between 44.0% and 44.4%, while the models using whole dataset and out-of-sample test data had values between 52.2% and 55.6%. Confusion matrixes of cross-validation models are shown in Tables 7-9.

Cross Validation 27 Hidden Nodes	**Predicted Positive (n=228.4)**	**Predicted Negative (n=176.6)**	
Actual Positive (n=218.2)	TP: 120	FN: 98.2	Sensitivity: 55.0%
Actual Negative (n=186.8)	FP: 108.4	TN: 78.4	Specificity: 42.0%
	Precision: 52.5%	Neg. Pred. Value: 44.4%	Accuracy: 49.0%

Table 7. Confusion matrix for cross-validation ANN 1

Cross Validation 50 Hidden Nodes	**Predicted Positive (n=227.8)**	**Predicted Negative (n=177.2)**	
Actual Positive (n=218.2)	TP: 119	FN: 99.2	Sensitivity: 54.5%
Actual Negative (n=186.8)	FP: 108.8	TN: 78	Specificity: 41.8%
	Precision: 52.2%	Neg. Pred. Value: 44.0%	Accuracy: 48.6%

Table 8. Confusion matrix for cross-validation ANN 2

Cross Validation 100 Hidden Nodes	**Predicted Positive (n=225.4)**	**Predicted Negative (n=179.6)**	
Actual Positive (n=218.2)	TP: 117.6	FN: 100.6	Sensitivity: 53.9%
Actual Negative (n=186.8)	FP: 107.8	TN: 79	Specificity: 42.3%
	Precision: 52.2%	Neg. Pred. Value: 44.0%	Accuracy: 48.5%

Table 9. Confusion matrix for cross-validation ANN 3

4.2 ROC Curves

To support the confusion matrix, each model was also analyzed with a ROC curve (Figure 8), which presents the ratio between true positive rate and false positive rate in different classification thresholds. Threshold being the probability value of positive (1.0) classification for the prediction. Red line in each curve represents the curve which would be generated by random guessing. AUC (area under ROC curve) was also calculated for each model to give a numerical evaluation for the ROC curve.

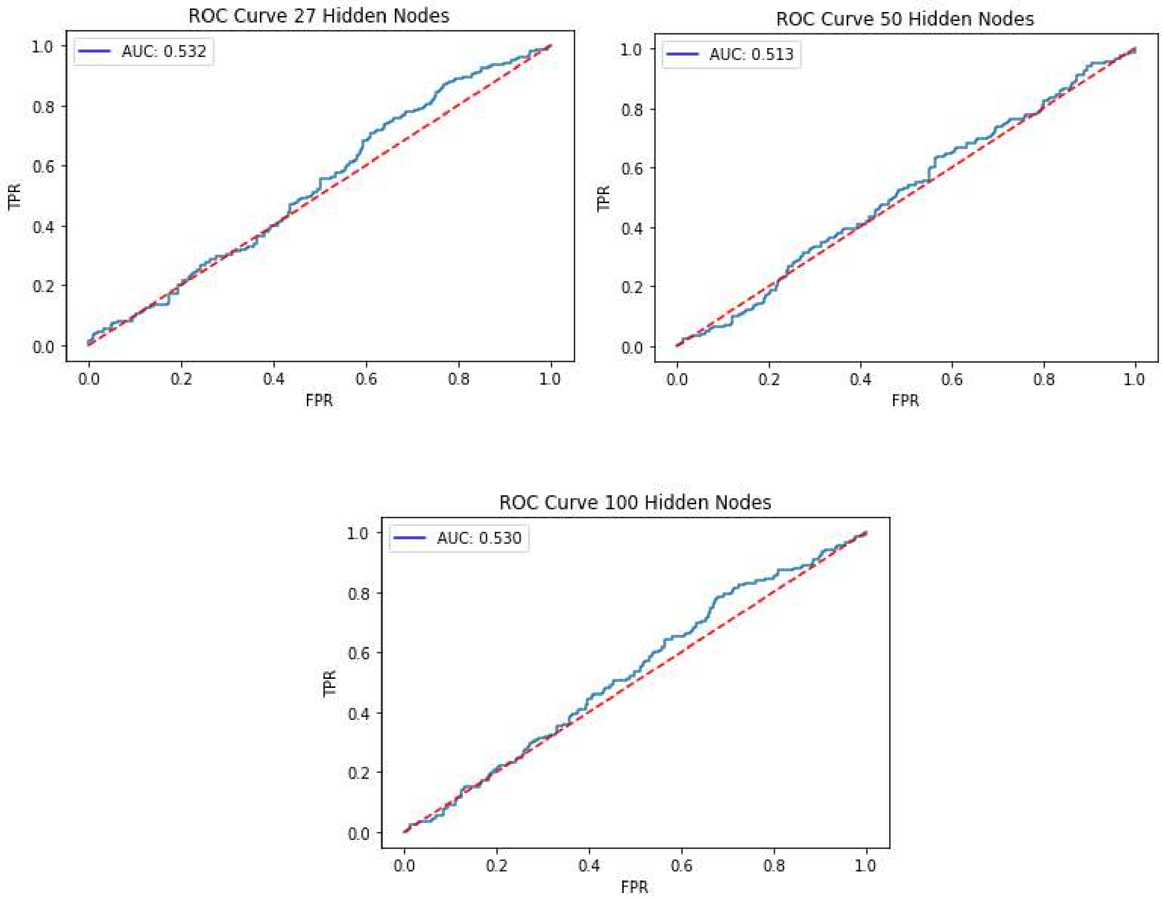

Figure 8. ANNs' ROC curves

All models' ROC curves are drawn very close to the random guessing line. They perform a little better when threshold is risen, especially ANN 1 and ANN 3. That suggests that when predictions have high positive probability value, the precision of predictions increases, but overall curves suggest that any of the models are not much better at predicting returns than random guessing. AUC values also support this conclusion, being 0.532, 0.513 and 0.530 for ANN 1, ANN 2 and ANN 3, respectively.

4.3 Daily Returns

Confusion matrix and ROC curve can be used to evaluate how well the models are able to predict whether the next day's return is positive or negative, but they do not consider whether the return is high or low. If the ANNs can predict high positive or negative returns, they might have some benefit in terms of average returns achieved for the predictions made with the model, even though the overall accuracy is low. Therefore, mean daily returns of positive and negative predictions were analyzed to get more detailed results from the predictions.

Statistical significance of the returns was tested by calculating t-values for the returns using formulas that Brock et al. (1992) used in their research. T-value was calculated for positive and negative predictions separately, and for the difference between positive and negative prediction return. T-value for separate positive and negative predictions were calculated using following formula (4):

$$\frac{\mu_r - \mu}{(\sigma^2/N + \sigma^2/N_r)^{1/2}} \tag{4}$$

where and μ_r and N_r are the mean return and count of signals for the positive or negative predictions and μ and N are the mean and count of observations in entire sample. σ^2 is the entire sample's estimated variance. For the difference of mean positive and negative prediction returns the t-value is calculated with formula (5):

$$\frac{\mu_p - \mu_n}{\left(\sigma^2/N_p + \sigma^2/N_n\right)^{1/2}} \tag{5}$$

where and μ_p and N_p are the mean return for and number of signals for positive predictions and μ_n and N_n are the mean return for and number of signals for negative predictions.

Evaluation of mean daily returns of test data predictions provided more support for ANNs' predictive capabilities than confusion matrix and ROC curve. Results can be seen in Table 10.

Mean daily returns of positive predictions for all three models were between 0.025% (6.59% annualized) and 0.032% (8.38% annualized). Negative predictions had mean returns between -0.079% (-18.6% annualized) and -0.109% (-24.4% annualized). Mean daily return of the complete test dataset was -0.012% (-3.11% annualized).

ANN 3 was the best performing model, with the highest mean return on positive prediction days, and lowest mean return on negative prediction days. Significance of results were analyzed using two tailed t-test. When significance of the positive and negative predictions' returns was analyzed separately, none of them were significant at any level. Only the return differences between positive and negative returns in ANN 1 and ANN 3 were significant at 10% level.

Model	Prediction	N	Mean Daily Return	Annualized Return [1]	T-value	Pos.-Neg.	T-value
ANN 1 (27 Hidden Nodes)	Positive	336	0.031717 %	8.35 %	0.75	0.131064 %	1.67**
	Negative	171	-0.099347 %	-22.23 %	-1.18		
ANN 2 (50 Hidden Nodes)	Positive	323	0.025228 %	6.59 %	0.63	0.103925 %	1.35
	Negative	184	-0.078697 %	-18.06 %	-0.92		
ANN 3 (100 Hidden Nodes)	Positive	347	0.031830 %	8.38 %	0.76	0.140434 %	1.76**
	Negative	160	-0.108603 %	-24.04 %	-1.27		
All Days	**-**	**507**	**-0.012488 %**	**-3.11 %**			

[1] Compound interest for 253 trading days.

** Significant at 10% level

Table 10. Mean daily returns

These results with combined with previous tests' low prediction accuracy suggest that ANNs performed better at predicting positive or negative daily returns when returns' absolute value was higher than average. Prediction accuracies analyzed in different return categories can be seen in Figure 9, which shows the number of positive and negative predictions in different categories of actual return. The predictive power on higher absolute return days can be seen especially in ANN 3. When absolute value of returns were around 1.5% and higher, it was able to predict the direction of the return correctly 62% of the times.

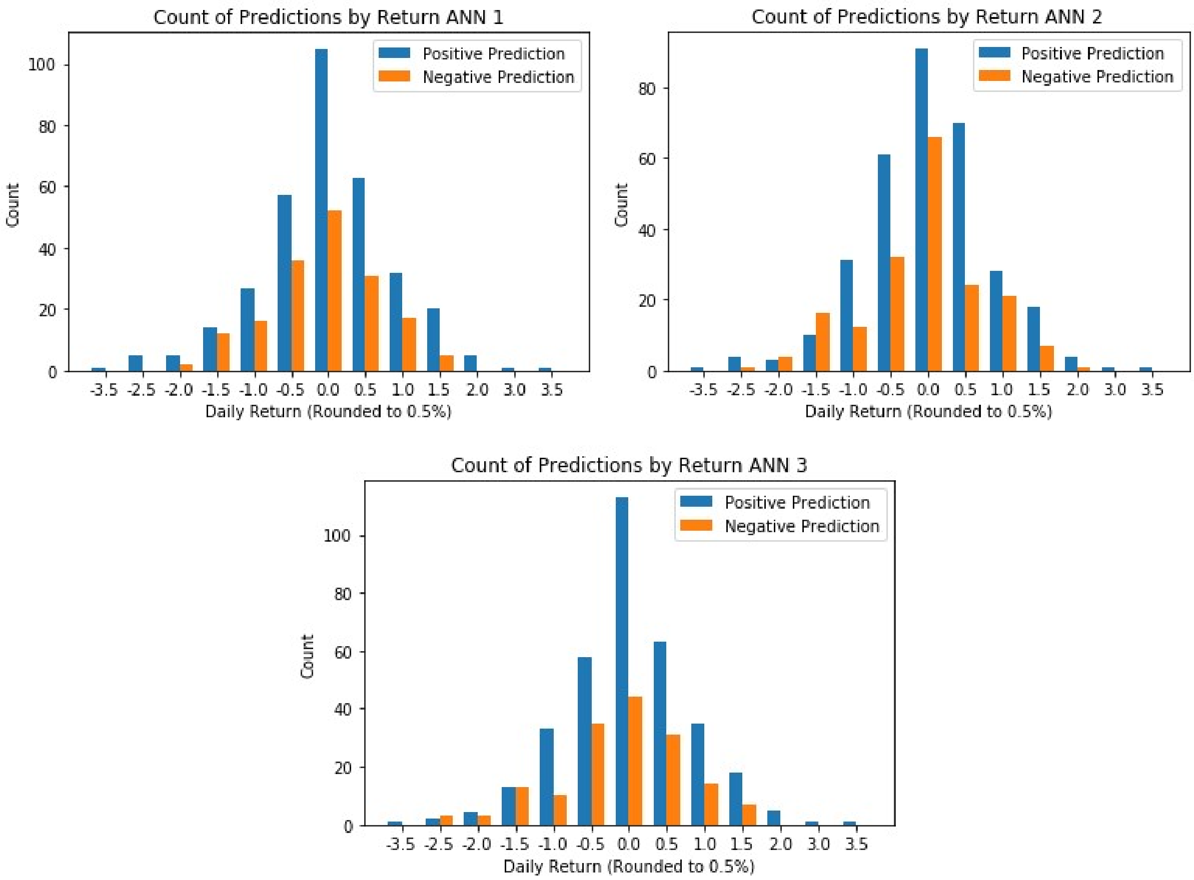

Figure 9. Predictions for different actual return categories

Again, the cross-validation models resulted in considerably worse results than the baseline models with entire training and testing data. Mean returns on negative prediction days were positive for all three models and were higher than mean returns on positive prediction days, which means that models' prediction accuracy was even worse than random guessing. This suggests that trained models are not at all generalizable. T-values for all the returns were very low, so the results did not have any statistical significance. Return statistics for cross-validation models can be seen in Table 11.

Cross Validation Model	Prediction	N	Mean Daily Return	Annualized Return [1]	T-value	Pos.-Neg.	T-value
ANN 1 (27 Hidden Nodes)	Positive	228.4	0.041067 %	11.38 %	-0.08	-0.018633 %	-0.14
	Negative	176.6	0.059700 %	17.79 %	0.07		
ANN 2 (50 Hidden Nodes)	Positive	227.8	0.039759 %	12.35 %	-0.10	-0.032389 %	-0.24
	Negative	177.2	0.072148 %	22.47 %	0.17		
ANN 3 (100 Hidden Nodes)	Positive	225.4	0.028876 %	9.07 %	-0.19	-0.050658 %	-0.37
	Negative	179.6	0.079533 %	26.61 %	0.23		
All Days	**-**	**405**	**0.050665 %**	**13.99 %**			

[1] Compound interest for 253 trading days.

** Significant at 10% level

Table 11. Mean daily returns for cross-validation models

4.4 Trading Strategy Simulation

To answer the question, whether ANNs can be used to gain extraordinary returns on stock markets, investment simulation was created using the predictions of the best performing ANN, ANN 3. Simulation was started with 100 points starting balance. Simulation predicted the direction of the next day's return in test data each day and invested the current balance on long or short position like 1x bull or bear certificate, based on the prediction. Balance would change based on investment decision and next day's return, and the total next day's ending balance would be invested again using the same strategy.

This ANN investment strategy was benchmarked against buy and hold strategy, where 100 points would be invested in a long position at the start of the simulation, and the position would stay same the entire simulation. Simulation was run without brokerage fee, and with 0.1% brokerage fee, which would be paid for the current balance every time the position would be changed from long to short or vice versa.

Like daily mean returns, simulations gave very different results from the initial confusion matrix analysis. After 507 investment days, investment strategy based on ANN 3's predictions ended up giving 30.55% positive return, while the buy and hold strategy had 7.78% negative return. Line chart of the simulation (Figure 10) shows that invested balances started separating when index downturn started around 120 days after the start. This difference grew even bigger between days 200 and 270, when index stagnated, but ended up narrowing down during next bull run. After 350 days index started to go down again, and ANN strategy gained abnormal returns on that bear market.

Figure 10. ANN investment strategy simulation without brokerage fee

When simulated with 0.1% brokerage fee (Figure 11), ANN and buy and hold strategy traded places during the whole investing period, but the downturn at the end of simulation caused ANN investment strategy ending up with 2.99% positive return, while buy and hold strategy had 7.78% negative return.

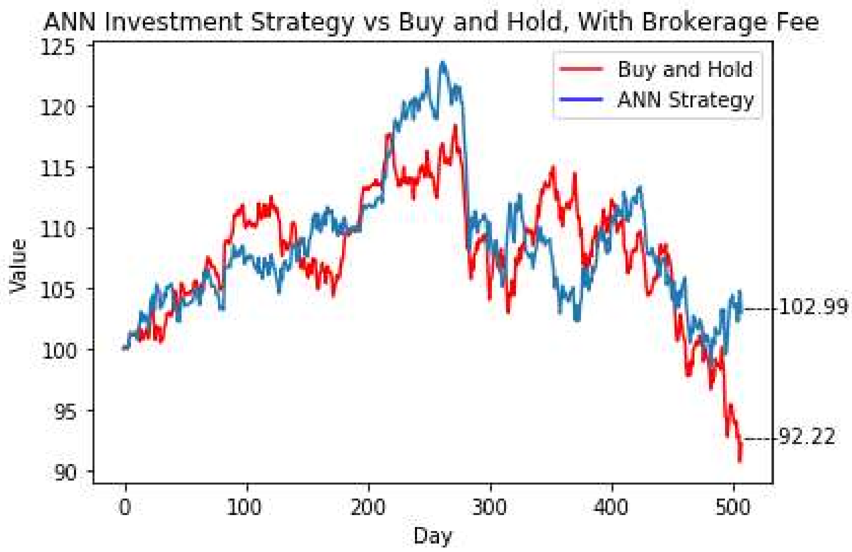

Figure 11. ANN investment strategy simulation with brokerage fee

4.5 Effect of Weekday Variables

To test the impact of the weekday variables to prediction accuracy, the best performing model, ANN3, was trained and tested without the weekday variables with the same complete training and testing datasets as the baseline ANN3.

Removing the variables resulted in worse results than the baseline model in all test. Accuracy of the model was only 49.9%, while the baseline model had an accuracy of 53.6%. Confusion matrix for the model without weekday variables can be seen in Table 12. Model's ROC curve (Figure 12) fell below the random guessing line with almost every threshold.

w/o Weekdays 100 Hidden Nodes	Predicted Positive (n=326)	Predicted Negative (n=181)	
Actual Positive (n=260)	TP: 166	FN: 94	Sensitivity: 63.8%
Actual Negative (n=247)	FP: 160	TN: 87	Specificity: 35.2%
	Precision: 50.9%	Neg. Pred. Value: 48.1%	Accuracy: 49.9%

Table 12. Confusion matrix for model without weekday variables

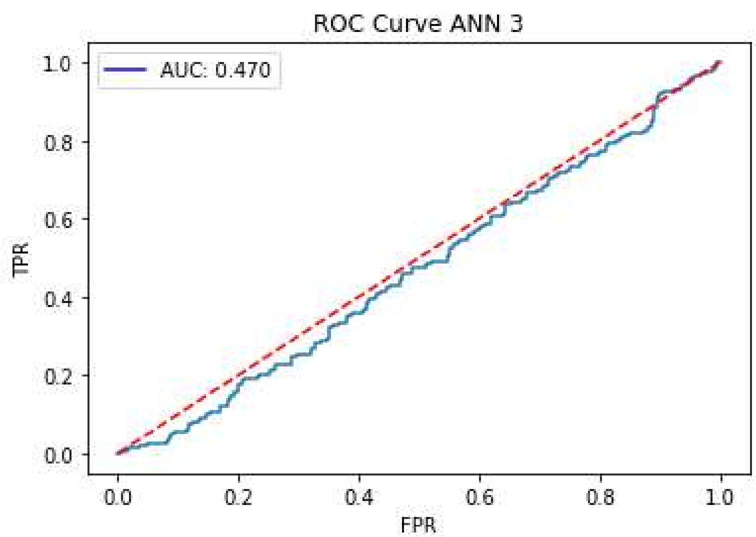

Figure 12. ROC curve for model without weekday variables

While the baseline model was able to predict positive and negative returns on average, model without weekday variables had opposite results. Mean return on positive prediction days was negative and mean return on negative prediction days was positive. These returns did not have any statistical significance alone, but when the positive and negative predictions' return difference was tested against baseline model's comparable return difference, the results were significant at 1% level. This suggests that model without weekday variables had significantly worse returns than baseline model. Mean returns of the model are presented in Table 13.

Model	Prediction	N	Mean Daily Return	Annualized Return [1]	T-value	Pos.-Neg.	T-value	Return - Baseline Return	T-value	Pos.-Neg. - Baseline Pos.-Neg.	T-value
ANN 3 (100 Hidden Nodes)	Positive	326	-0.022166 %	-5.45 %	-0.16	-0.027109 %	-0.35	-0.053996 %	-0.84	-0.167543 %	-3.192*
	Negative	181	0.004943 %	1.26 %	0.24			0.113546 %	1.25		
All Days	-	**507**	**-0.012488 %**	**-3.11 %**							

[1] Compound interest for 253 trading days.

* Significant at 1% level

Table 13. Mean daily returns for model without weekday variables

When the model's predictions were simulated as an investment strategy, it could not beat buy and hold strategy. Simulation without brokerage fees resulted in 9.32% negative return against buy and hold's 7.78% negative return. With brokerage fees the negative return rise to 28.47%. Simulations' value graphs can be seen in Figures 13 and 14.

Figure 13. ANN without weekday variables investment strategy simulation w/o brokerage fee

Figure 14. ANN without weekday variables investment strategy simulation with brokerage fee

5 CONCLUSIONS

Overall, the evaluation methods provided some support for the models' predictive capabilities, but the cross-validation undermined the significance of these results. Even though the results are conflicting, some conclusions can be drawn from the results. The results of tests do not definitively answer the research questions defined in the chapter 1.2, but some answers with limitations could be given.

Q1: Can ANNs predict the direction of daily return of DAX index better than random guessing?

In selected research frame, null hypothesis cannot be disproved. ANNs were not able to predict the direction of the DAX index price change on very high accuracy, and the cross-validation did not give any support for prediction accuracy. Prediction accuracy increased when the absolute returns were higher, but the sample size was not large enough to draw definitive conclusions.

Q2: What are the daily returns on predicted positive and negative days?

On average, the mean daily returns were positive on predicted positive days and negative predicted on negative days. Returns deviated largely from the mean daily return of the whole test dataset but were not statistically significant. Especially the mean return for negative days was significantly lower than the average return on the whole period. These results could not be repeated in cross-validation.

Q3: Can ANN investment strategy beat buy and hold strategy for DAX index?

Simulation of investment strategies on the test dataset showed that the ANN was able to predict some of the negative return days, especially when the absolute return was higher, causing it to perform much better than buy and hold strategy, when there were no brokerage fees. When brokerage fees were counted in, ANN investment strategy did not perform significantly better than buy and hold and was able gain better profits just because of the tested period ended in the longer downturn phase in the market.

Q4: Is ANN's performance better with weekdays as predicting variables?

All the evaluation methods yielded better results when weekday variables were used as predicting variables. In this case, ANN's overall performance was better with weekday variables, and differences in mean returns proved to be statistically significant.

This research, as all the researches that study financial markets' efficiency, have multiple limitations on the applicability of its results. The ANN models were trained only with data from DAX index from the selected period. ANNs are very sensitive tools and can easily overfit to the training data, therefore the data selection has great impact on the results. Selected period did not include any great modern market crashes such as early-2000 internet bubble, 2008 financial crisis, or 2020 stock market crash. Including these might have helped models to recognize market characteristics of deep bear markets.

The ANNs tested had similar topologies with same number of hidden layers, only the number of nodes in hidden layer were changed. Change in topology, for example using multiple hidden layers instead of just one, might have given different results. Other changes in other network characteristics like activation function, learning rate and solver, might also yield different results. Even if the ANNs architecture stays the same, one ANN gives different results when trained multiple times, because every time ANN is trained its initial weights are random and it converges to different local optimum or reaches iteration limit.

Technical analysis and ANNs is very broad field to study with countless of variations of data and models to test. Robustness of the ANN model developed in this research could be studied more by training and testing similar model with different financial markets and longer time periods. This would give more insight whether the current results were just caused by chance and selected dataset or could the model be usable in different markets and time periods. The number of variables could also be expanded or limited, and the model could be run with different combinations of chosen variables, to find out which are the most important variables in the model and could there be more optimal combination of variables for prediction. For example, it would be interesting to see if better accuracy could be achieved by using more than 10 days of historical returns and adding various lengths of moving averages to capture different term trends in the data.

6 REFERENCES

Bishop, C., 1999. Neural Networks for Pattern Recognition. Oxford: Oxford Univ. Press.

Bishop, C., 2006. Pattern Recognition and Machine Learning. Springer, New York, USA.

Brock, W., Lakonishok, J. and Lebaron, B., 1992. Simple Technical Trading Rules and the Stochastic Properties of Stock Returns. The Journal of Finance, 47(5), pp.1731-1764.

Brownlee, J., 2020. A Gentle Introduction to The Rectified Linear Unit (Relu). [online] Machine Learning Mastery. Available at: <https://machinelearningmastery.com/rectified-linear-activation-function-for-deep-learning-neural-networks/> [Accessed 26 March 2020].

Chiang, Y. & Ke, M. & Liao, T. & Wang, C. 2012. Are technical trading strategies still profitable? Evidence from the Taiwan Stock Index Futures Market. Applied Financial Economics, Volume 22, Issue 12, pp. 955-965.

Chong, T. and Ng, W., 2008. Technical analysis and the London stock exchange: testing the MACD and RSI rules using the FT30. Applied Economics Letters, 15(14), pp.1111-1114.

Cross, F., 1973. The Behavior of Stock Prices on Fridays and Mondays. Financial Analysts Journal, 29(6), pp. 67-69.

Dax-indices.com. 2020. DAX Digital | DAX® (TR) EUR. [online] Available at: <https://www.dax-indices.com/index-details?isin=DE0008469008> [Accessed 26 March 2020].

Dicle, M.F.; Hassan M. K. 2007. Day of the week effect in Istanbul Stock Exchange.Scientific Journal of Administrative Development, vol. 5, pp. 53-83.

Fama, E., 1965. The Behavior of Stock-Market Prices. The Journal of Business, 38(1), p.34.

Fama, E., 1970. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25(2), p.383.

Ft.com. 2020. Volatility: How ‘Algos’ Changed the Rhythm Of The Market. [online] Available at: <https://www.ft.com/content/fdc1c064-1142-11e9-a581-4ff78404524e> [Accessed 26 March 2020].

Geman, S., Bienenstock, E. and Doursat, R., 1992. Neural Networks and the Bias/Variance Dilemma. *Neural Computation*, 4(1), pp.1-58.

Gibbons, M. and Hess, P., 1981. Day of the Week Effects and Asset Returns. The Journal of Business, 54(4), p. 579.

Glorot, X., Bordes, A. and Bengio, Y., 2011. Deep Sparse Rectifier Neural Networks. In: Fourteenth International Conference on Artificial Intelligence and Statistics. [online] Proceedings of Machine Learning Research, pp.315-323. Available at: <http://proceedings.mlr.press/v15/glorot11a> [Accessed 26 March 2020].

Goodfellow, I., Bengio, Y. and Courville, A., 2017. Deep Learning. Cambridge, Mass: The MIT Press.

Guresen, E. & Kayakutlu, G. & Daim, T. 2011. Using artificial neural network models in stock market index prediction. Expert Systems with Applications, Volume 38, Issue 8, pp. 10389-10397.

Hirose, Y., Yamashita, K., and Hijiya, S., 1991, Back-propagation algorithm which varies the number of hidden units. Neural Networks, 4, pp. 61-66.

Hutson, E., Kearney, C. and Lynch, M., 2008. Volume and skewness in international equity markets. Journal of Banking & Finance, 32(7), pp.1255-1268.

Jasic, T. & Wood, D. 2004. The profitability of daily stock market indices trades based on neural network predictions: case study for the S&P 500, the DAX, the TOPIX and the FTSE in the period 1965–1999. Applied Financial Economics, Volume 14, Issue 4, pp. 285-297.

Kara, Y. & Acar Boyacioglu, M. & Baykan, Ö. 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, Volume 38, Issue 5, pp. 5311-5319.

Karpoff, J., 1987. The Relation Between Price Changes and Trading Volume: A Survey. The Journal of Financial and Quantitative Analysis, 22(1), p.109.

Kavzoglu, T., 1999. Determining Optimum Structure for Artificial Neural Networks. In: Annual Technical Conference and Exhibition of the Remote Sensing Society. Nottingham: The University of Nottingham, pp.675-682.

Kim, K., 2003. Financial time series forecasting using support vector machines. *Neurocomputing*, 55(1-2), pp.307-319.

Kingma, D. and Ba, J., 2020. Adam: A Method for Stochastic Optimization. [online] arXiv.org. Available at: <https://arxiv.org/abs/1412.6980> [Accessed 26 March 2020].

Kirkpatrick II, C. & Dahlquist, J., 2016. Technical Analysis: The Complete Resource for Financial Market Technicians. Pearson, pp. 3-7.

Kleene, S. C. (1956). Representation of Events in Nerve Nets and Finite Automata. In C. E. Shannon, J. McCarthy (Eds.), Automata Studies. (AM-34) (pp. 3–42).

Linden, M. and Louhelainen, M., 2006. Testing for weekday anomaly in international stock index returns with non-normal errors. Applied Financial Economics Letters, 2(3), pp.193-197.

Medium. 2020. MACHINE LEARNING: How Black Is This Beautiful Black Box. [online] Available at: <https://towardsdatascience.com/machine-learning-how-black-is-this-black-box-f11e4031fdf> [Accessed 23 March 2020].

Menkhoff, L., 2010. The use of technical analysis by fund managers: International evidence. Journal of Banking & Finance, 34(11), pp.2573-2586.

Metghalchi, M., Chang, Y. and Marcucci, J., 2008. Is the Swedish stock market efficient? Evidence from some simple trading rules. International Review of Financial Analysis, 17(3), pp.475-490.

Nazario, R. & Silva, J. & Sobreiro, V. 2017. Literature review of technical analysis on stock markets. The Quarterly Review of Economics and Finance, Volume 66, pp. 115-126.

Park, C.-H. & Irwin, S. 2007. What do we know about the profitability of technical analysis? Journal of Economic Surveys, Volume 21, pp. 786-826.

Principe, J. & Euliano, N & Lefebvre, W. 1999. Neural and adaptive systems: Fundamentals through simulations. John Wiley & Sons, New York, USA.

Qiu, M. & Song, Y. 2016. Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model. PLoS ONE, Volume 11, Issue 5, Article number e0155133.

Scikit-learn.org. 2020. Scikit-Learn: Machine Learning in Python — Scikit-Learn 0.22.2 Documentation. [online] Available at: <https://scikit-learn.org/stable/> [Accessed 26 March 2020].

The Economist. 2020. From Not Working to Neural Networking. [online] Available at: <https://www.economist.com/special-report/2016/06/23/from-not-working-to-neural-networking> [Accessed 19 March 2020].

Titan, A. 2015. The Efficient Market Hypothesis: Review of Specialized Literature and Empirical Research. Procedia Economics and Finance, Volume 32, pp. 442-449.

Wilder, J., 1978. New Concepts In Technical Trading Systems. Greensboro, N.C.: Trend Research, pp.63-70.