University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 23/2021 (371)

APPLICATION OF MACHINE LEARNING IN QUANTITATIVE INVESTMENT STRATEGIES ON GLOBAL STOCK MARKETS

Jan Grudniewicz
Robert Ślepaczuk

Warsaw 2021

WORKING PAPERS 23/2021 (371)

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES

WORKING PAPERS

Application of machine learning in quantitative investment strategies on global stock markets

Jan Grudniewicz[a], **Robert Ślepaczuk** [b]*

[a] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group*
[b] *University of Warsaw, Faculty of Economic Sciences, Quantitative Finance Research Group, Department of Quantitative Finance*
* *Corresponding author: rslepaczuk@wne.uw.edu.pl*

Abstract: The thesis undertakes the subject of machine learning based quantitative investment strategies. Several technical analysis indicators were employed as inputs to machine learning models such as Neural Networks, K Nearest Neighbor, Regression Trees, Random Forests, Naïve Bayes classifiers, Bayesian Generalized Linear Models and Support Vector Machines. Models were used to generate trading signals on WIG20, DAX, S&P500 and selected CEE indices in the period between 2002-01-01 to 2020-10-30. Strategies were compared with each other and with the benchmark buy-and-hold strategy in terms of achieved levels of risk and return. Quality of estimation was evaluated on independent subsets and with the use of sensitivity analysis. The research results indicated that quantitative strategies generate better risk adjusted returns than passive strategies and that for the analysed indices predominantly Bayesian Generalized Linear Model and Naïve Bayes were the best performing models. More comprehensive rank approach based on the results for all analysed models and indices allowed to select Bayesian Generalized Linear Model as the model which on average generated the best results.

Keywords: quantitative investment strategies, machine learning, neural networks, regression trees, random forests, support vector machine, technical analysis, equity stock indices, developed and emerging markets, information ratio

JEL codes: C4, C14, C45, C53, C58, G13

Notes: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

Introduction

Investing is the process of allocating capital to obtain future financial benefits. The investor expects the future cash flows to exceed the initial value of the investment. Investments in financial instruments are one of the methods of investing used by both individuals and institutional entities. One type of these instruments are securities traded on stock exchanges. Since the inception of stock exchanges, an increase in the value of equity instruments has been observed in the long run, which is associated with continuous economic growth.

The main stock exchange index is an indicator measuring the value of shares of the largest companies traded on a given stock exchange. In Poland, WIG20 is such an index and it is valued as a synthetic portfolio of 20 largest companies listed on the Warsaw Stock Exchange (subject to the representation levels of individual industries in the index).

Construction of investment portfolios reflecting the behavior of stock exchange indices is a method of passive capital management. In order to construct such a portfolio, it is necessary to purchase an appropriate number of shares included in the index portfolio and to buy or sell securities on an ongoing basis to best reflect its current composition. There are investment companies that create this type of index based instruments and offer them to both individuals and institutional entities. Investors may purchase investment certificates of closed-end ETFs (exchange-traded funds) traded on the stock exchange or purchase participation units of open-end funds on the OTC market.

An alternative to passive investing is active capital management. Investors can actively select equity instruments in their investment portfolio in order to achieve returns that exceed those achieved in the passive strategies. The selection can be made on the basis of the methods of fundamental analysis and technical analysis. Fundamental analysis is the classic form of assessing attractiveness of a company performed in order to assess the potential future value of its shares. Among others it consists of the analysis of the company's historical and current financial statements, its long-term development strategy, the situation of the industry and economy in which the company operates analyzed in the context of their development potential and possible threats, and analysis of possible impacts on the company's results caused by the geopolitical situation.

The second family of stock selection methods is technical analysis, main assumption of which is the possibility of forecasting the behavior of the stock price based on its historical changes. Recurring patterns can be observed in the price charts, the ongoing identification of which is the basis for making investment decisions (Beechey et al. 2000). There are investors on the capital markets who use the information obtained from the fundamental and technical analysis separately, as well as concomitantly. The way to assess whether a given active strategy is in fact profitable is to compare its results with the strategy of passive capital management. One method of benchmarking is by relating the results to a buy-and-hold strategy on an index portfolio.

This thesis describes the selection of the best performing investment strategy in which investment decisions are based on signals generated by technical analysis indicators and machine learning techniques. Technical indicators such as Simple Moving Average, Moving Average Convergence Divergence, Stochastic Oscillator, Relative Strength Index and Williams' Percent Range served as inputs to the machine learning models. The main aim of the thesis is to choose the best performing strategy among the strategies constructed using various machine learning techniques such as Neural Networks, K Nearest Neighbor, Regression Trees, Random Forests, Naïve Bayes classifiers, Bayesian Generalized Linear Models and Support Vector Machines in both Linear and Polynomial form. The second goal of the thesis is to compare the strategies with the method of passive capital management based on the buy-and-hold mechanism. The strategies were compared using risk and return measures, such as the annualized rate of return, standard deviation of returns, maximum capital drawdown, Sharpe Ratio and Information Ratio.

Thesis extends the current achievements of scientific research by employing the existing methods of machine learning and technical analysis to construct quantitative investment strategies, the profitability of which has been examined on the stock market indices of Poland (WIG20), two highly developed countries: Germany (DAX) and USA (S&P500) as well as six countries from Central and Eastern Europe: Bulgaria (SOFIX), Czech Republic (PX), Estonia (OMXT), Hungary (BUX), Latvia (OMXR) and Lithuania (OMXV). Data used in the research consisted of High, Low and Close daily prices of the indices in the period from 2002 to 2020, thus the scope of the research covers the periods of the great financial crisis of 2007-2009 and COVID-19 pandemic crisis.

The main research hypothesis (RH1) *states that active quantitative investment strategies based on the signals generated by machine learning models result in higher risk adjusted returns than buy-and-hold benchmark strategy*. The additional research hypotheses were as follows: (RH2) *Neural Networks generate the best (with regard to risk adjusted returns) investment signals compared to other machine learning techniques used in the research*; (RH3) *the very same machine learning strategy is considered best performing for all analyzed stock market indices*; (RH4) *returns obtained from signals generated by machine learning techniques* *are resistant to changes in hyperparameters underlying the models and to changes in parameters underlying the technical analysis indicators*. Based on the conclusions from the leading article developed by Dash and Dash (2016) which induced this thesis, the intuition behind the results was that machine learning based strategies can generate returns above benchmark with lower amount of investment risk. Moreover, Neural Networks strategies were suspected of being the best performing for every analyzed stock index and in every sensitivity analysis scenario.

The structure of the thesis is composed of four chapters. The first chapter describes the subject of quantitative investment strategies and presents an overview of existing research in this field. The second chapter presents the description of the data along with the method of dividing it into subsets. The third chapter describes the research methodology, including the technical analysis indicators used, machine learning techniques, the strategy construction and measures of risk and return, which were used to compare the strategies. The fourth chapter contains results of the research, including selection of the best performing strategy for each of the analyzed indices and assessment of the quality of estimation with sensitivity analysis of the selected model parameters.

1. Investment management

1.1. Historical overview

Technical analysis is a tool for detecting recurring patterns in the prices of financial instruments and making investment decisions based on them. It includes many different theories and approaches such as candlestick patterns, Elliot wave theory and Fibonacci retracement levels (Reuters 1999). All these tools are based on separate assumptions and provide different output information. However, most of them have one feature in common - they are aimed at identifying a trend prevailing on the market and forecasting further price behavior. The vast majority of professional investors use technical analysis tools of their choice, but in the academic world it is considered unrelated to science, as it does not result directly from economic theory (Tian et al. 2002). A much more popular method among scientists is analysis of the macroeconomic environment known as fundamental analysis (Tian et al. 2002). In practice, investors combine these two methods, thus obtaining a whole picture of the market situation.

Fama (1970) formulated his efficient market hypothesis claiming that prices of securities reflect all available information about them. In his work, he distinguished three types of market efficiency: weak, semi-strong and strong. The weak hypothesis states that the price of an instrument contains all information about its past, and therefore it is not possible to forecast the future direction of price changes on the basis of historical data. The conclusion from this hypothesis is unequivocal - technical analysis is not able to bring benefits to investors. The semi-strong hypothesis states that price reflects all information from history as well as information in public circulation such as company's financial statements. This means that neither technical nor fundamental analysis is helpful in forecasting and generating positive return on investment. According to the last hypothesis, a strong hypothesis, price includes information from the past, public information as well as non-public information. It implies that, in addition to the conclusions from the semi-strong hypothesis, it is not possible to gain benefits from the so-called insider trading, i.e. basing investment decisions on hard-to-reach internal information about companies.

Many scientific studies conducted before the formulation of efficient market hypothesis, have not rejected its weak form, among others Larson (1960), Osborne (1962), Mandelbrot (1963), Alexander (1964) and Fama (1965). Moreover, the assumptions of the theory of efficient markets were already discussed by Bachelier (1900) and the theory itself is in some manner a development of the theory about the behavior of prices being similar to the random walk process, and therefore their unpredictability, which was widely recognized in the financial world.

Despite the situation in academia at the time, Brock et al. (1992), in their empirical study on the US stock index, showed positive financial results coming from a set of three strategies based on technical analysis indicators. Their work started a new wave of scientific research aimed to determine the effectiveness of price forecasting based on historical data.

1.2. Review of the existing research

The subject of effectiveness of investment strategies based on technical analysis and machine learning techniques has been widely analyzed in scientific research. As part of the literature review, several papers fundamental to this thesis were discussed in the following paragraphs.

The foremost paper underlying this thesis was an article by Dash and Dash (2016) in which the authors discussed the profitability of investment strategy constructed by the computational efficient functional link artificial neural network (CEFLANN) with Extreme Learning Machine (ELM) learning approach and technical analysis indicators such as Moving Average, Moving Average Convergence Divergence, Stochastic Oscillator, Relative Strength Index and Williams' Percent Range. Models generated three classes of trading signals i.e. buy, hold and sell. Signals were then used to identify the trend on the analyzed instruments which along with several technical rules allowed for trading decision generation. Data used in the research consisted of daily quotes of two indices: BSE SENSEX and S&P 500 from 2010-2014 period which was divided into training and testing subsets. Achieved returns were compared with those generated by alternative models such as Support Vector Machines, Naive Bayesian model, K Nearest Neighbor model and Decision Tree. Results showed that CEFLANN model produced the highest returns compared to the other models.

In the paper by Jiang et al. (2012) Support Vector Machine model and Multiple Additive Regression Trees were used to solve classification problem of trend prediction in NASDAQ, DJIA and S&P 500 US stock indices daily prices. Additionally, the authors employed SVM model, linear regression and generalized linear model (GLM) as regression techniques aimed to predict the value of price movements. Research discussed the correlations between target US stock indices and other financial instruments such as Nikkei 225, Hang Seng, FTSE100, DAX and ASX indices, EURUSD, AUDUSD and USDJPY currency pairs and silver, platinum, palladium, oil and gold commodities which served as predictor variables in the models. Results showed a relatively high accuracy of trend prediction achieved bythe classification techniques. In case of regression techniques, strategies constructed in the research produced on average returns higher than the analyzed benchmarks.

Huang et al. (2005) analyzed the predictive ability of Support Vector Machine models as well as other classification techniques such as Linear Discriminant Analysis, Quadratic Discriminant Analysis and Elman Backpropagation Neural Networks. Models were applied on weekly NIKKEI 225 stock index data and incorporated several macroeconomic variables as model inputs. Performance of the models was measured by the hit ratio calculated as percentage of correct predictions of price movement direction. SVM achieved the best results compared to the remaining models. Additionally, a model combining predictions from all of the analyzed techniques with corresponding weights estimated on training subset was analyzed with the resulting hit ratio exceeding that achieved by SVM solely.

In the article written by Gerlein et al. (2016) six machine learning based models including the Naïve Bayes classifier were used to produce profitable quantitative strategies on the USDJPY, EURUSD and EURGBP currency pairs. Models incorporated technical analysis tools and a set of attributes related to among others price, seasonality and lagged price values. Several setups related to periodic retraining, training subset size and selected attributes were analyzed in order to calculate the model parameters as well as to measure the accuracy and returns. Despite the accuracy of the models being relatively low, models allowed to generate positive cumulative returns in several setups.

Madan et al. (2015) applied Generalized Linear Model, Support Vector Machine and Random Forest techniques to predict the Bitcoin price change in daily as well as high frequency intervals. Authors focused on models' accuracy measurement which was relatively high for daily price change prediction in case of GLM and Random Forest models with the highest result obtained by the GLM. Random Forest model achieved higher accuracy on 10 minute price quotes.

In an article by Chen et al. (2006), authors discussed the application of Support Vector Machines and Back Propagation Neural Networks on daily close prices of six Asian stock indices: Nikkei 225, All Ordinaries, Hang Seng, Straits Times, Taiwan Weighted and KOSPI. Models were constructed based on Exponential Moving Average and Relative Difference in Percentage of the price. Predictions were then evaluated using statistics such as Mean Squared Error, Normalized Mean Squared Error and Mean Absolute Error measuring the deviation of predicted price from the actual price as well as Directional Symmetry and Weighted Directional Symmetry measuring prediction accuracy of price movement direction with the latter also incorporating deviation component in its weights. Results showed that the analyzed models behaved better than benchmark with regard to predicted price deviation measures.

Leigh et al. (2002) described a novel approach to technical analysis bull-flag pattern recognition aiming to predict price changes. Technical indicators served as inputs to Neural Network model which was then altered with genetic algorithm in order to improve the model's coefficient of determination. Techniques were applied on New York Stock Exchange Composite Index. Calculated returns indicated the superiority of analyzed methods compared to buy-and-hold benchmark strategy.

In a paper by Lin et al. (2006), authors investigated the performance of decision trees deployed on the 'electronic stocks' of Taiwan stock market and 'technology stocks' of NASDAQ market. Prices of selected stocks were adjusted to exclude effects of dividends. Filtering rule based on moving averages was applied to the dataset. Decision Tree algorithms were then used in order to cluster the information from four input variables: money supply, inflation rate, revenues and stock market index futures' prices. Model performance was evaluated using averaged compound annual rate of return. Predictions yielded positive returns in case of both indices.

Colianni et al. (2015) discussed construction of trading strategies based on qualitative data concerning Bitcoin cryptocurrency observed on the Twitter portal. Linear Regression models, Support Vector Machines as well as Bernoulli and Multinomial Naïve Bayes classifiers were used in two approaches: text classification and sentiment analysis, both aimed to predict

Bitcoin price change directions. Bernoulli Naïve Bayes classifier achieved the highest accuracy in the text classification approach while Linear Regression resulted in the highest accuracy in the sentiment analysis approach compared to the remaining techniques.

The aforementioned studies present a scientifically and financially interesting problem of the effectiveness of technical analysis and machine learning techniques in the development of investment strategies. An interesting question is how the strategies constructed by the researchers would behave on stock markets in less developed countries and what is the impact of the great financial crisis of 2007-2009 as well as COVID-19 pandemic crisis on the final result.

In this thesis, the research objective is to investigate predictive ability of a set of machine learning techniques proposed by Dash and Dash (2016), namely Neural Networks, K Nearest Neighbor, Naïve Bayes, Regression Tree, Support Vector Machines and additionally Random Forest and Bayesian Generalized Linear Model discussed in other papers. Technical analysis indicators proposed by Dash and Dash (2016) were used as inputs to aforementioned machine learning models. Analyzed set of indicators consisted of Simple Moving Average (SMA), Moving Average Convergence Divergence (MACD), Stochastic Oscillator (STOCH), Relative Strength Index (RSI) and Williams' Percent Range (WPR) with the underlying parameters similar to those proposed in the paper.

One of the newest research (Kijewski and Ślepaczuk, 2020) testing the validity and efficiency of ML techniques compared the performance of classical techniques with LSTM model for S&P500 index on daily frequency for the last 20 years. They showed that the combination of strategies (ML and classical techniques) outperformed market significantly and that the combination of signals gave the best results by diversifying the risk of single strategy mistake. Finally, they showed that LSTM with selected hyperparameters outperformed ARIMA model but at the same time LSTM model results were not robust to initial hyperparameters assumptions.

Thesis extends current scientific research by evaluation of strategies on stock markets of both developed (Germany, USA) and less developed countries from Central and Eastern Europe including Poland and by inclusion of the great financial crisis of 2007-2009 and COVID-19 pandemic crisis data. Additionally, employment of dynamic estimation windows (periodical redevelopment of underlying parameters of the models to reflect current market behaviors) was introduced. Empirical study compared the generated set of strategies with the benchmark buy-and-hold strategy. This was to check how much an investor would gain in excess of the market if he employed the tools discussed in this thesis. The study assumed the absence of transaction costs as in the work of Brock et al. (1992).

1.3. Research hypotheses

The main objective of this study is to verify the profitability of investment strategies based on the technical analysis indicators and machine learning techniques. Previous studies on similar matters discussed the above-average profitability of such strategies. The authors additionally wanted to transfer existing scientific findings into their domestic Polish capital market, as well as the markets of Central and Eastern Europe. For this purpose, an attempt was made to verify the following research hypotheses (RH):

- RH1: Active quantitative investment strategies based on signals generated by machine learning models result in higher risk adjusted returns than buy-and-hold benchmark strategy,
- RH2: Neural Networks generate the best (with regard to risk adjusted returns) investment signals compared to other machine learning techniques used in the research,
- RH3: The very same machine learning strategy is considered best performing for all analyzed stock market indices,
- RH4: Returns obtained from signals generated by machine learning techniques are resistant to changes in hyperparameters underlying the models and to changes in parameters underlying the technical analysis indicators.

Verification of the research hypotheses should demonstrate the profitability of the quantitative investment strategies constructed using analyzed methods and deployed on the capital markets discussed in the study. Obtaining the instrument-robust answers would require research on all available stock indices as well as on other groups of financial instruments, which may become the subject of future research.

2. Data

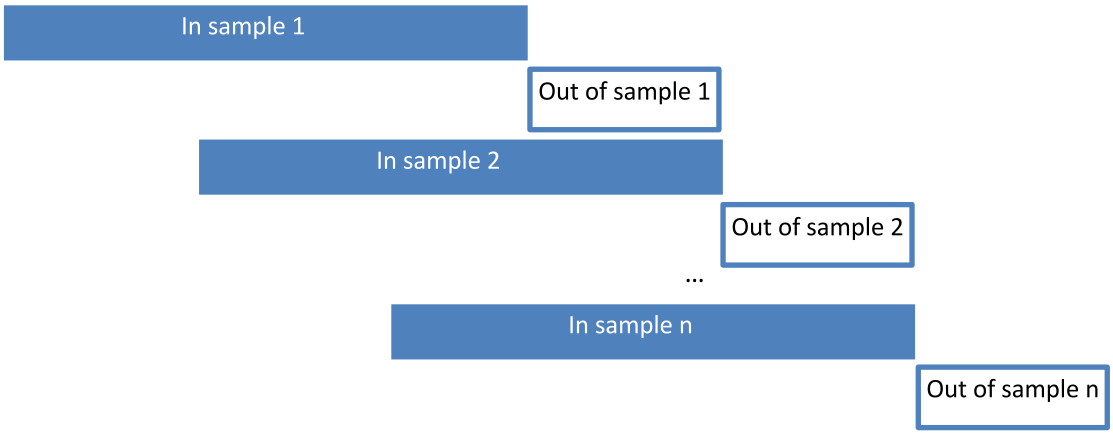
2.1. Data description

Data used in the research was downloaded from the website *http://www.stooq.pl/* and contains information about HLC (High Low Close) prices of selected stock indices. For the analysis, authors chose the stock market index of their domestic market – WIG20 (Poland), two highly liquid global indices from developed countries – DAX (Germany) and S&P500 (USA) as well as a group of less liquid Central and Eastern European countries' indices: SOFIX (Bulgaria), PX (Czech Republic), OMXT (Estonia), BUX (Hungary), OMXR (Latvia) and OMXV (Lithuania). Data for each index contained quotes from 2002-01-01 to 2020-10-30. Due to suspension of quotation observed for certain indices as well as differing holiday calendars, there were dates with no quotes available. This limitation was remediated by omitting those dates in the analysis resulting in number of observations differing among indices.

2.2. Sampling

Research employed dynamic estimation windows which means that the underlying parameters of the models were periodically recalibrated to reflect current market behaviors. Observations from the beginning of the available period were initially trimmed in order for the overall number of observations for each index to be easily divisible into equal subsets. Calibration of models' parameters was conducted on 200 trading day window (in sample) and then model predictions were applied onto next 20 trading day window (out of sample). For each subsequent dynamic window iteration, in sample and out of sample moved by 20 trading days. The process is shown on the Figure 2.1.

Figure 2.1. Sampling process overview



Note: Figure illustrates the sampling process showing how in sample and out of sample subsets are derived from the overall dataset.

Table 2.1 presents the number of observations after aforementioned trimming process for each of analyzed indices with the corresponding number of in sample and out of sample subsets created as well as the start and end date of the overall sample period.

Table 2.1. Data sampling overview

Index	Observations	Subsets	Start date	End date
WIG20	4680	224	2002-02-22	2020-10-30
DAX	4740	227	2002-03-01	2020-10-30
S&P500	4700	225	2002-03-05	2020-10-30
SOFIX	4600	220	2002-03-06	2020-10-30
PX	4680	224	2002-03-06	2020-10-30
OMXT	4680	224	2002-03-18	2020-10-30
BUX	4660	223	2002-03-04	2020-10-30
OMXR	4680	224	2002-02-27	2020-10-30
OMXV	4660	223	2002-03-06	2020-10-30

Note: Table presents descriptive statistics for each stock index: number of observations, number of subsets, start date and end date of observation period in the dataset.

2.3. Initial data analysis

Initial data analysis was conducted in order to assess the distributions of input variables (High, Low and Close prices of indices) used for the calculation of technical analysis indicators and target variable (discrete returns derived from Close prices - target variable in the models described in detail at the beginning of the next section) used for modeling purposes. Table 2.2 to Table 2.6 present mean, minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum values of those variables for each of the analyzed indices. No outliers or data quality issues were identified.

Table 2.2. Descriptive statistics of WIG20 and DAX data.

Index	WIG20				DAX			
Measure	High	Low	Close	Return	High	Low	Close	Return
Mean	2263.68	2227.01	2245.56	0.0001	7830.33	7714.99	7774.54	0.0003
Minimum	1043.70	1026.65	1039.20	-0.1328	2319.65	2188.75	2202.96	-0.1224
25th percentile	1856.62	1823.44	1841.15	-0.0073	5398.34	5290.23	5347.39	-0.0062
50th percentile	2308.79	2269.50	2290.88	0.0002	7151.88	7047.24	7093.06	0.0008
75th percentile	2499.71	2468.12	2482.79	0.0076	10654.58	10492.26	10577.92	0.0072
Maximum	3940.53	3910.96	3917.87	0.0850	13795.24	13754.04	13789.00	0.1140

Note: Table presents descriptive statistics of High, Low, Close prices and returns for WIG20 and DAX indices: mean, minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum.

Table 2.3. Descriptive statistics of S&P500 and SOFIX data.

Index	S&P500				SOFIX			
Measure	**High**	**Low**	**Close**	**Return**	**High**	**Low**	**Close**	**Return**
Mean	1686.39	1666.72	1677.18	0.0003	587.82	582.27	585.03	0.0004
Minimum	695.27	666.79	676.53	-0.1198	114.22	114.22	114.22	-0.1074
25th percentile	1169.14	1151.57	1161.99	-0.0043	399.55	394.25	397.39	-0.0043
50th percentile	1407.82	1394.69	1402.52	0.0007	490.20	485.38	488.26	0.0003
75th percentile	2104.06	2087.14	2098.15	0.0056	661.65	655.49	658.53	0.0049
Maximum	3588.11	3535.23	3580.84	0.1158	1981.80	1952.28	1952.40	0.0875

Note: Table presents descriptive statistics of High, Low, Close prices and returns for S&P500 and SOFIX indices: mean, minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum.

Table 2.4. Descriptive statistics of PX and OMXT data.

Index	PX				OMXT			
Measure	**High**	**Low**	**Close**	**Return**	**High**	**Low**	**Close**	**Return**
Mean	1055.47	1044.99	1050.06	0.0002	742.21	735.31	738.75	0.0005
Minimum	391.20	391.20	391.20	-0.1494	155.34	152.84	154.25	-0.1006
25th percentile	905.06	891.68	898.33	-0.0054	543.49	535.05	539.70	-0.0033
50th percentile	1018.08	1006.94	1012.86	0.0006	721.59	711.61	717.11	0.0005
75th percentile	1169.33	1156.43	1162.40	0.0066	974.87	965.55	970.74	0.0043
Maximum	1944.30	1918.10	1936.10	0.1316	1376.93	1370.27	1374.35	0.1286

Note: Table presents descriptive statistics of High, Low, Close prices and returns for PX and OMXT indices: mean, minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum.

Table 2.5. Descriptive statistics of BUX and OMXR data.

Index	BUX				OMXR			
Measure	**High**	**Low**	**Close**	**Return**	**High**	**Low**	**Close**	**Return**
Mean	22721.77	22349.25	22537.90	0.0004	550.79	544.58	547.79	0.0004
Minimum	6765.02	6546.35	6589.76	-0.1188	184.94	184.94	184.94	-0.1507
25th percentile	17437.59	17129.42	17310.16	-0.0072	375.58	368.97	372.44	-0.0046
50th percentile	21621.64	21232.95	21403.55	0.0005	447.62	442.84	445.08	0.0002
75th percentile	27163.14	26796.30	27004.28	0.0080	694.65	685.17	689.44	0.0054
Maximum	46476.20	45904.82	46230.22	0.1408	1143.82	1125.14	1138.74	0.1285

Note: Table presents descriptive statistics of High, Low, Close prices and returns for BUX and OMXR indices: mean, minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum.

Table 2.6. Descriptive statistics of OMXV data.

Index	OMXV			
Measure	**High**	**Low**	**Close**	**Return**
Mean	421.03	418.47	419.83	0.0005
Minimum	82.87	82.87	82.87	-0.1125
25th percentile	312.23	309.27	310.71	-0.0029
50th percentile	419.19	417.52	418.09	0.0005
75th percentile	534.40	532.02	532.97	0.0041
Maximum	792.74	788.64	790.25	0.1163

Note: Table presents descriptive statistics of High, Low, Close prices and returns for OMXV index: mean, minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum.

3. Research methodology

3.1. General model formula and target variable

One of the general goals of the quantitative investment strategies development is to construct models with price predictive ability. Machine learning models developed in this research belong to the supervised models family which means that they are fed with pairs of input (technical analysis indicators) and target (stock indices returns) variables. Information coming from the input and target pairs was used to calibrate each models' coefficients in each of the in sample periods. Those coefficients were then applied to the inputs in the following out of sample periods in order to predict the target variable in those periods. The general modeling matter can be described using Formula (1).

$$f(Y) = f(X) + \varepsilon \tag{1}$$

where: Y – vector of target variable (returns),

X – matrix of independent variables (set of technical indicators),

f(.) – function applied to transform the variables (depending on the model),

ε – vector of random errors.

Target (dependent) variable in this research is defined as a discrete return on the asset calculated from the observed Close prices as described by Formula (2).

$$r_t = \frac{C_t - C_{t-1}}{C_{t-1}} \tag{2}$$

where: r_t – discrete return in period t,

C_t – close price in period t.

3.2. Technical analysis indicators

Thesis focuses on strategies based on a set of 5 technical analysis indicators: Simple Moving Average (SMA), Moving Average Convergence Divergence (MACD), Stochastic Oscillator (STOCH), Relative Strength Index (RSI) and Williams' Percent Range (WPR) as proposed by Dash and Dash (2016). Technical indicators were then used as an input to machine learning models. The following section describes formulas used for calculation of each analyzed indicators.

3.2.1. Simple Moving Average (SMA)

Simple Moving Average is an average price of an instrument calculated on historical observations up to the reference date. Formula (3) represents the method of SMA calculation (Feller 1950).

$$SMA_t = \frac{1}{n}\sum_{i=0}^{n-1} P_{t-i} \tag{3}$$

where: SMA – Simple Moving Average,

n – number of periods included in calculation,

P_t – price on reference date.

Base level analyzed for parameter n = 15 as proposed by Dash and Dash (2016) and for the purpose of sensitivity analysis, n = {14;16} were used in order to verify robustness of the models. Input used later in the models is derived with the usage of Formula (4). It is a measure of how distant the current price is from its SMA:

$$SMA_{signal} = P_t - SMA \tag{4}$$

3.2.2. Moving Average Convergence Divergence (MACD)

MACD is an indicator developed by Appel (2005) which incorporates several Exponential Moving Averages (EMA) into its derivation. EMA is described as a moving average which assigns exponentially decreasing weights (older the observation, lower the weight) to each of the historical observations. Weights are dependent on a number of assumed periods which impacts the smoothing factor α derived using Formula (5). EMA is then calculated according to Formula (6).

$$\alpha = \frac{2}{n+1} \tag{5}$$

where: α – smoothing factor,

n – number of periods included in calculation.

$$EMA_t = \begin{cases} P_t & t = 1 \\ \alpha * P_t + (1-\alpha) * EMA_{t-1} & t > 1 \end{cases} \tag{6}$$

where: EMA – Exponential Moving Average,

P_t – price on reference date.

MACD indicator is composed of 2 distinct time series: MACD line and signal line. MACD line is defined as the difference between the long EMA and short EMA which are as proposed by Appel (2005) calculated using n = 26 and n = 12 periods. Signal line is defined as EMA with parameter n = 9. Formulas (7) and (8) describe derivation of both time series.

$$MACD\ line = EMA(n = 26) - EMA(n = 12) \tag{7}$$

$$signal\ line = EMA(n = 9) \tag{8}$$

Parameters proposed by Appel (2005) are considered base parameters in this thesis. For the purpose of sensitivity analysis, n = {25;27} for long EMA, n = {11;13} for short EMA and n = {8;10} for signal EMA were used in order to verify robustness of the models.

Input used later in the models is derived with the usage of Formula (9). It is a measure of how distant the MACD line is from the signal line:

$$MACD_{signal} = MACD\ line - signal\ line \tag{9}$$

3.2.3. Stochastic Oscillator (STOCH)

Invented by Lane (1984), Stochastic Oscillator is incorporating HLC (High Low Close) data into its calculation. STOCH comprises of 3 time series derived using Formulas (10), (11) and (12).

$$fast\ \%K = 100 * \frac{P_t - \min(L_{t-n+1}, .., L_t)}{\max(H_{t-n+1}, .., H_t) - \min(L_{t-n+1}, .., L_t)} \tag{10}$$

where: fast %K – unsmoothed %K indicator,

n – number of periods included in calculation,

P_t – close price on reference date,

H_t – highest price on reference date,

L_t – lowest price on reference date.

$$fast\ \%D = SMA_t(fast\ \%K) \tag{11}$$

where: fast %D – simple moving average of fast %K.

$$slow\ \%D = SMA_t(fast\ \%D) \tag{12}$$

where: slow %D – simple moving average of fast %D.

Lane proposed that fast %K should be calculated with parameter n = 14 while fast %D and slow %D as SMAs with n =3 periods, those levels are treated as base levels in this thesis. Sensitivity analysis was conducted using n = {13;15} for fast %K and n = {2;4} for fast %D and slow %D.

3.2.4. Relative Strength Index (RSI)

RSI was developed by Wilder (1978) as a ratio incorporating average upward (U) and downward (D) movements of the close price tracked for n historical observations. U and D are calculated with Formulas (13) and (14), while averages of U and D are derived with Formulas (15) and (16).

$$U_t = \begin{cases} P_t - P_{t-1} & P_t > P_{t-1} \\ 0 & P_t \leq P_{t-1} \end{cases} \tag{13}$$

$$D_t = \begin{cases} 0 & P_t \geq P_{t-1} \\ P_{t-1} - P_t & P_t < P_{t-1} \end{cases} \tag{14}$$

where: U_t – upward movement in period t,

D_t – downward movement in period t,

P_t – price in period t.

$$avgU_t = \begin{cases} \frac{1}{n}\sum_{i=0}^{n-1} U_{t-i} & t = n \\ \frac{avgU_{t-1} * (n-1) + U_t}{n} & t > n \end{cases} \tag{15}$$

$$avgD_t = \begin{cases} \frac{1}{n}\sum_{i=0}^{n-1} D_{t-i} & t = n \\ \frac{avgD_{t-1} * (n-1) + D_t}{n} & t > n \end{cases} \tag{16}$$

where: $avgU_t$ – average of upward movements in n periods,

$avgD_t$ – average of downward movements in n periods,

n – number of periods included in calculation.

Relative Strength Index is then calculated as a ratio of $avgU_t$ and $avgD_t$ using Formula (17).

$$RSI_t = 100 * \frac{avgU_t}{avgU_t + avgD_t} \tag{17}$$

Base level parameter n = 14 was employed as proposed by Wilder (1978). In sensitivity analysis, n = {13;15} were used in order to verify robustness of the models.

3.2.5. Williams' Percent Range (WPR)

Developed by Williams (1979), WPR is a form of a price oscillator using HLC (High Low Close) data. It is calculated similarly to fast %K in Stochastic Oscillator. Formula (18) represents calculation process for WPR.

$$WPR = 100 * \frac{\max(H_{t-n+1}, .., H_t) - P_t}{\max(H_{t-n+1}, .., H_t) - \min(L_{t-n+1}, .., L_t)} \tag{18}$$

where: WPR – Williams' Percent Range,

n – number of periods included in calculation,

P_t – close price on reference date,

H_t – highest price on reference date,

L_t – lowest price on reference date.

Similarly to RSI indicator, base level analyzed for parameter n = 14 as proposed by Williams (1979) while sensitivity analysis is conducted using n = {13;15}.

All analyzed technical indicators are lagged by one period before being used as predictors for returns in the models in order to avoid the so-called look ahead bias involving making decisions in the same period for which the given signal was generated.

3.3. Machine learning techniques

Thesis analyzed eight supervised machine learning models with majority of them proposed in the paper by Dash and Dash (2016) and other discussed in the remaining papers. Employed techniques included Neural Networks, K Nearest Neighbor, Random Forest, Regression Tree, Naïve Bayes, Bayesian Generalized Linear Model and Support Vector Machines in both Linear and Polynomial form. Following subsections describe each of the models and discuss the hyperparameters used to conduct one of the sensitivity analysis exercises.

3.3.1. Neural Networks (NN)

Models and corresponding strategies referred in this research as Neural Networks were developed using the Extreme Learning Machine (ELM) approach as proposed by Dash and Dash (2016). ELM in the configuration applied is a feedforward neural network with only one hidden layer which is the reason why this approach is computationally efficient compared to other neural network related techniques. Number of neurons in input layer is equal to the number of input technical analysis indicators. In each of the in sample estimations, model is trained using a number of neurons in hidden layer varying from 1 to twice the size of input layer and the best performing in sample variant is then chosen. Activation function in a tansig (tangent-sigmoid transfer function) form producing continuous values in the range from -1 to 1 (intuitive for return prediction) was applied to compute the output trading signal in the output layer consisting of one neuron. Model was implemented in the form discussed by Huang et al. (2006). Output from Extreme Learning Machine algorithm has a form described by Formula (19).

$$f_n(X) = \sum_{i=1}^{n} \beta_i h_i(X) \tag{19}$$

where: f_n – predicted output from the model,

X – matrix of independent variables,

n – number of neurons in hidden layer,

β_i – weight of hidden neuron i,

h_i – output of hidden neuron i.

The base activation function analyzed was the tansig transformation and, for the purpose of sensitivity analysis, two alternative functions producing outputs in the same range were chosen: sin (sine transfer function) and satlins (symmetric saturating linear transfer function).

3.3.2. K Nearest Neighbor (KNN)

Research employs K Nearest Neighbor model in its regression version. The output prediction of the model is the average value of observed target variable for k nearest neighbors identified based on the levels of input independent variables. Models were implemented in a form proposed by Altman (1991). KNN algorithm is described with Formula (20).

$$p = \frac{\sum_{i=1}^{k} y_i}{k} \tag{20}$$

where: p – predicted output from the model,

y_i – observed target variable for nearest neighbor observation i,

k – number of nearest neighbors included in the calculation.

Identification of k nearest neighbors is based on determination of high dimensional Euclidean distance between independent variables of analyzed observations as described in Formula (21).

$$d_{i,j}^2 = \sum_{l=1}^{n} (x_{l,i} - x_{l,j})^2 \tag{21}$$

where: $d^2_{i,j}$ – high dimensional Euclidean distance between observations i and j,

$x_{l,i}$ – independent variable l for observation i,

n – number of independent variables.

The hyperparameter chosen for sensitivity analysis was an optimization metric with the Root Mean Square Error (RMSE) as the base metric and alternative metrics being the coefficient of determination (Rsquared) and Mean Absolute Error (MAE).

3.3.3. Random Forest (RF)

Random Forest model in a regression form used in this thesis is a statistical modelling framework consisting of random generation of multiple decision trees with each of the trees producing a distinct prediction for target variable. Those predictions are then averaged to calculate the final output. Models were implemented in a form discussed by Breiman (2001). Process of the output generation is described by Formula (22).

$$p = \frac{\sum_{j=1}^{m} \sum_{i=1}^{n} W(x_i, x')y_i}{m} \tag{22}$$

where: p – predicted output from the model,

y_i – observed target variable for observation i,

x_i – vector of independent variables for observation i,

x' – vector of independent variables for observation in testing sample,

$W(x_i, x')$ – weight function of x_i relative to x',

n – number of observations in training sample,

m – number of generated random trees.

The hyperparameter chosen for sensitivity analysis was an optimization metric with the RMSE as the base metric and alternative metrics being the Rsquared and MAE.

3.3.4. Regression Tree (RT)

Recursive partitioning Regression Trees are a version of Decision Trees from the Classification and Regression Trees (CART) family with continuous target variable. Data is split in recursive manner in order to generate optimal decision algorithm for target variable prediction. Model inputs (independent variables) are reflected in the tree branches from which, after a set of recursive partitioning, final leaves with the computed target variable are produced. Models were implemented in a form proposed by Breiman et al. (1984). Fundamental algorithms of Regression Tree are similar to those of Random Forest model. The hyperparameter chosen for sensitivity analysis was an optimization metric with the RMSE as the base metric and alternative metrics being Rsquared and MAE.

3.3.5. Naïve Bayes (NB)

Naïve Bayes is a probabilistic classifier incorporating the assumption of naïve independence between input variables. As a classification method, it produces binary outputs (classes) computed from conditional a posteriori probabilities. Thesis used {-1;1} classes representing buy and sell trading signals thus the Naïve Bayes model was the single model implemented without an additional 'neutral' signal. Model was implemented in a form proposed by Narasimha Murty and Susheela Devi (2011). Output from the model is described by Formula (23).

$$p = \underset{k \in \{1,..,K\}}{\operatorname{argmax}}\, q(C_k) \prod_{i=1}^{n} q(x_i | C_k) \tag{23}$$

where: p – predicted output from the model, q(.) – probability of (.),

C_k – class k,

x_i – independent variable i,

n – number of independent variables.

The hyperparameter chosen for sensitivity analysis was an optimization metric with Accuracy (number of correct predictions divided by the total number of predictions) as the base metric and one alternative metric being the Cohen's kappa (Kappa).

3.3.6. Bayesian Generalized Linear Model (BGLM)

Generalized Linear Model is a generalization of linear regression models which among others allows for target variable transformations via a link function e.g. logit which was used in this thesis. Target variable prediction is computed as a linear combination of input variables. BGLM uses the Bayesian approach to model fitting instead of the Frequentist approach. A priori distributions of inputs and the likelihood function are used for a posteriori estimation of model parameters. Models were implemented in forms discussed by Nelder and Wedderburn (1972) and Dempster et al. (1977). Model output is described by Formula (24).

$$p = l^{-1}(X\beta) \tag{24}$$

where: p – predicted output from the model,

l – link function,

X – matrix of independent variables,

β – vector of coefficients fitted by the model using the Bayesian approach.

The hyperparameter chosen for sensitivity analysis was an optimization metric with RMSE as the base metric and alternative metrics being Rsquared and MAE.

3.3.7. Support Vector Machine Linear (SVML)

Regression version of Support Vector Machine models was analyzed in this research. SVM models generate multiple hyperplanes aiming to separate input independent variables and search for the most optimal solution allowing for the best prediction of the continuous target variable. Linear type of SVM was implemented in a form discussed by Cortes and Vapnik (1995). The objective function of the model is described by Formula (25). It has to be determined by identification of the optimal hyperplane using the minimization problem presented in Formula (26). Additionally, all of the model residuals have to fulfill the condition described by Formula (27).

$$f(X) = X'\beta + b \tag{25}$$

$$\frac{1}{2}\beta'\beta \tag{26}$$

$$\forall i : |y_i - (X_i{}'\beta + b)| \leq \varepsilon \tag{27}$$

where: f(X) – function of X, target to determination,

X – matrix of independent variables,

β – vector of coefficients,

b – offset intercept,

y_i – observed dependent variable for observation i,

X_i – vector of independent variables for observation i,

ε – random error.

The hyperparameter chosen for sensitivity analysis was an optimization metric with RMSE as the base metric and alternative metrics being Rsquared and MAE.

3.3.8. Support Vector Machine Polynomial (SVMP)

Polynomial Support Vector Machine models employed in the research are a version of SVM models incorporating the polynomial kernel function transforming model inputs and computing high-dimensional hyperplanes. Polynomial type of SVM was implemented in a form discussed by Boser et al. (1992). Polynomial kernel function has a form described by Formula (28).

$$k\left(x_i, x_j\right) = (1 + x_i' x_j)^d \tag{28}$$

where: k(.) – kernel function,

x_i, x_j – vectors of independent variables,

d – degree of the polynomial.

The hyperparameter chosen for sensitivity analysis was an optimization metric with RMSE as the base metric and alternative metrics being Rsquared and MAE.

3.4. Investment strategies construction

Construction of quantitative investment strategies required multiple computational steps and definition of trading rules. This section describes the process of inputting technical analysis indicators into the machine learning models as well as the process of generation of trading signals from the models' outputs.

3.4.1. Extended model formula

The general model Formula (1) can be further extended to Formula (29) to present each particular independent variable described in detail in the Technical Analysis Indicators section.

$$f(y) = f(SMA_{signal}) + f(MACD_{signal}) + f(fast\ \%K) + f(fast\ \%D) + f(slow\ \%D) + f(RSI) + f(WPR) + \varepsilon \quad (29)$$

where: y – vector of target variable (returns),

f(.) – function applied to transform the variables (depending on the model),

ε – vector of random errors.

3.4.2. Model inputs transformation

Input independent variables (technical analysis indicators) were rescaled before being fed to the models. Process was conducted using a version of min-max normalization technique which produces outputs in range from -1 to 1. This technique was chosen for two reasons: it is intuitive as the machine learning models produce output variable that is also ranging from -1 to 1 and because it causes the input data to be more comparable. Process of min-max normalization (rescaled to range from -1 to 1) is described by Formula (30) as proposed by Han et al. (2011).

$$x_t' = \frac{x_t - \min(x)}{\max(x) - \min(x)} * 2 - 1 \quad (30)$$

where: x_t' – transformed value of variable in period t,

x_t – original value of variable in period t,

min(x)/max(x) – minimum/maximum value of the variable in all analyzed periods.

3.4.3. Model outputs transformation (investment signals creation)

Machine learning models used in this research can be divided into two groups: classification models (Naïve Bayes) and regression models (remaining techniques). Outputs (returns predictions) and corresponding trading signals for each of the incorporated models constitute a distinct investment strategy. Classification models produce a binary output {-1;1} while regression models produce continuous output ranging from -1 to 1. The question to be answered is how to translate the outputs into trading signals.

For classification models a simplistic approach was undertaken: -1 output translates to sell signal while +1 output translates to buy signal. Due to the binary output, this technique does not allow to produce neutral investment signals.

In case of regression models, the outputs were highly dispersed and non-comparable among the models in the sense of distribution measures therefore not allowing to set a fixed signal thresholds based on absolute values of the outputs. The decision was made that the most universal approach to signal generation will be to calculate quantiles of the output distributions for each of the analyzed models. 40^{th} quantile and 60^{th} quantile were applied as the thresholds for buy, sell and neutral signals. Signal +1 translates to buy signal, -1 to sell signal and 0 to neutral signal. Process of signal generation for regression models is described by Formula (31).

$$signal_t = \begin{cases} 1 & y_t \geq q_{0.6} \\ 0 & q_{0.4} < y_t < q_{0.6} \\ -1 & y_t \leq q_{0.4} \end{cases} \tag{31}$$

where: $signal_t$ – trading signal in period t,

y_t – output (prediction) generated by the model for period t,

q_α – quantile of particular out of sample outputs corresponding to probability α.

The process of entering a financial position was based on buy, sell and neutral signals. Neutral signal is interpreted as not taking a position or exiting an existing one. To calculate the return from a given strategy for each date, signal was multiplied by the observed discrete return of a given financial instrument which is described by Formula (32).

$$r_t^{strategy} = r_t^{index} * signal_{t-1}^{strategy} \tag{32}$$

where: $r_t^{strategy}$ – discrete return from the strategy in period t,

r_t^{index} – discrete return from stock index (financial instrument analyzed) in period t,

$signal_t^{strategy}$ – signal generated by the strategy in period t.

Returns from the strategies were aggregated for every out of sample period in order to produce return time series and compare the strategies among each other and additionally with the benchmark strategy. Research investigates which of the models (strategies) gives the most desired results. The effectiveness of the quantitative strategies was compared to that of a buy-and-hold strategy which is based on a market portfolio (benchmark). The buy-and-hold strategy involves buying an instrument at the beginning of the period under analysis and selling at the end of the period, so it can be interpreted as an absolute measure of market movements.

3.5. Risk and return measures

Thesis incorporates a wide range of performance indicators used to assess the quality of developed investment strategies. In order to appropriately compare the strategies, not only the accumulated profits but also the risks should be taken into account. Measures and ratios used in the analysis included the compound annual growth rate, standard deviation of returns, maximum capital drawdown, Sharpe Ratio and Information Ratio.

3.5.1. Compound Annual Growth Rate (CAGR)

The rate of return is the most frequently used measure in portfolio efficiency studies. When choosing an investment, it is important to analyze not only the rate of return from the last period, but from the entire history of the strategy. For this purpose, the compound annual growth rate is used which can be interpreted as an annualized rate of return. It is a measure illustrating how much on average capital has grown in each year of the investment. To calculate it, the following Formula (33) is used (Anson et al. 2010):

$$R = CAGR(t_0, t_n) = (\frac{V(t_n)}{V(t_0)})^{\frac{1}{t_n - t_0}} - 1 \tag{33}$$

where: $CAGR(t_0,t_n)$ – Compound Annual Growth Rate,

$V(t_0)$ – initial value of an investment,

$V(t_n)$ – closing value of an investment,

t_0 – calculation start year,

t_n – calculation end year.

3.5.2. Adjusted Sharpe Ratio (SR)

The rate of return itself does not contain any information about risk. The solution to this problem was proposed by Sharpe (1966) who introduced Sharpe's coefficient. This thesis used its simplified version, which does not contain information about the risk-free rate. It is calculated by dividing the annulized rate of return by the annualized standard deviation of rates of return in a given period. The standard deviation illustrates volatility of returns and is considered as a risk measure in which greater volatility indicates a higher investment risk. Considering the aforementioned information, when comparing strategies, the better performing one is the one with the higher Adjusted Sharpe Ratio. For the purpose of this thesis, the measure was floored at 0 as the negative values are often deemed meaningless in the scientific world. It is calculated with Formula (34).

$$SR = \max\{\frac{R}{\sigma R}; 0\} \tag{34}$$

where: SR – Adjusted Sharpe Ratio,

R – annualized rate of return,

σR – annualized standard deviation of returns.

3.5.3. Maximum drawdown (MDD)

The maximum drawdown represents the maximum decrease in accumulated capital over the entire investment horizon. When analyzing the rate of return, it is worth investigating whether the portfolio has not recorded significant drops in value in the analyzed horizon, which would indicate its instability. The most frequently used measure for this purpose is the maximum drawdown which describes that risk. It is a difference between the value of capital at the lowest point and the value at the previous highest peak divided by the value at that peak. The final value is usually shown as a percentage. In this thesis, the measure is always presented as positive value, so in superior investment strategies the maximum decline should be as low as possible. This situation will represent a lower risk of a managed portfolio. Measure is calculated as follows (Magdon-Ismail et al. 2003) using Formula (35):

$$MDD = -\frac{TMin - PMax}{PMax} \tag{35}$$

where: MDD – maximum drawdown,

TMin – minimum price level,

PMax – previous maximum price level.

3.5.4. Calmar Ratio (CR)

The Calmar Ratio is a very useful extension of the maximum drawdown described before. It is another risk and return measure that results from dividing the annualized rate of return by the maximum drawdown expressed in absolute value. Due to its structure, it is a measure that has a priority in application before the maximum drawdown (Bacon 2012). As in the case of the Sharpe Ratio, the better performing strategies are those with the higher value of the Calmar Ratio. The applied formula (Young 1991) is described by Formula (36).

$$CR = \frac{R}{MDD} \tag{36}$$

where: R – annualized rate of return,

MDD – maximum drawdown.

3.5.5. Information Ratio (IR*)*

For the purpose of this thesis, an adjusted Information Ratio definition was introduced. Risk weighted performance measured by both Sharpe Ratio and Calmar Ratio are combined into one ratio described in Formula (37) as proposed by Kość et al. (2019). The measure was floored at 0 as the negative values are often deemed meaningless in the scientific world.

$$IR^* = SR * CR = \frac{(\max\{R; 0\})^2}{\sigma R * MDD} \tag{37}$$

where: IR* – Information Ratio*,

R – annualized rate of return,

σR – annualized standard deviation of returns,

MDD – maximum drawdown.

There are many measures of risk and return used by researchers to compare the perfromance of investment strategies, but each of them is suitable for analyzing different types of instruments contained in a portfolio (Bacon 2010). The aforementioned measures and ratios fully cover the needs of this thesis and will allow for an objective assessment of the quality of discussed strategies.

4. Empirical Research

4.1. Identification of the best performing strategy

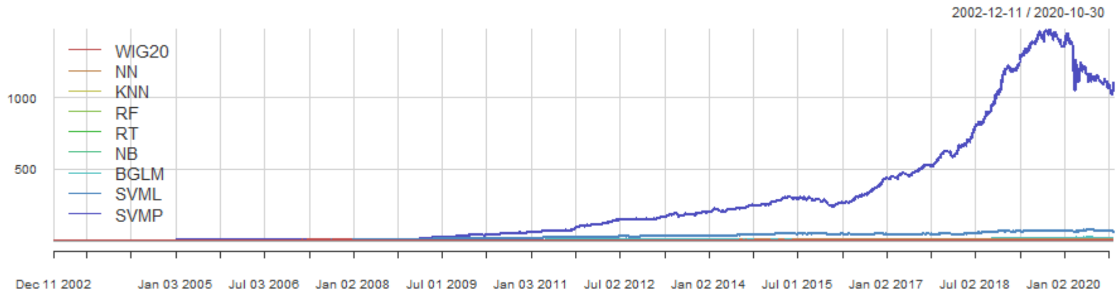
In order to determine the best performing strategy for a given stock index, the out of sample results produced by eight analyzed machine learning models were translated into trading signals and compared using risk and return measures. Trading signals produced by each machine learning model constitute a separate investment strategy. This section describes the results obtained by the strategies for each analyzed stock index. Results for the most liquid indices in the analyzed group i.e. WIG20 (Poland), DAX (Germany) and S&P500 (USA) are described in detail whereas the results from remaining CEE stock indices are presented in an aggregated approach. The evaluation of the obtained results was based on a graphical analysis i.e. equity lines (presenting how 1$ of initial investment would grow over the analyzed time period), daily returns and drawdown lines as well as the calculated values of risk and return measures. The main measure selected for strategy comparison was IR* as it contains the highest amount of information about the performance (it combines information about returns, standard deviation and maximum drawdown). Eight strategies built from eight machine learning models are also compared with the benchmark strategy i.e. buy-and-hold which in the following sections is always presented with the name of the relevant index. Results presented in this section were referred to as a base scenario in the following sensitivity analysis sections.

4.1.1. Investment strategies comparison for WIG20 (Poland)

As described in the data section, in case of WIG20 index, 224 subsamples were created with dates ranging from 2002-02-22 to 2020-10-30. Out of sample results were aggregated into a time series of discrete returns for every analyzed strategy and compared with each other as well as with the benchmark.

Figure 4.1 presents the equity lines for all analyzed strategies. Support Vector Machine strategies are dominant in case of WIG20 index with its Polynomial version outperforming the rest of the strategies significantly.

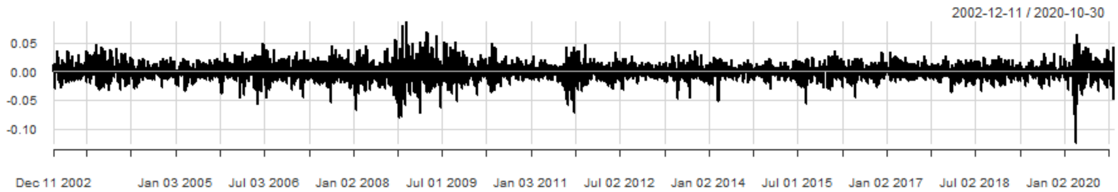
Figure 4.1. Equity lines for WIG20 (Poland)



Note: Figure shows equity lines for every strategy constructed on WIG20 index in the period from 2002-12-11 to 2020-10-30.

Comparing Figure 4.1 to Figure 4.2, an interesting relationship was observed - in most of the periods of increased index volatility, the equity lines are also increasing (e.g. in 2007-2009 great financial crisis period) while in COVID-19 pandemic crisis period this relationship is contradictory.

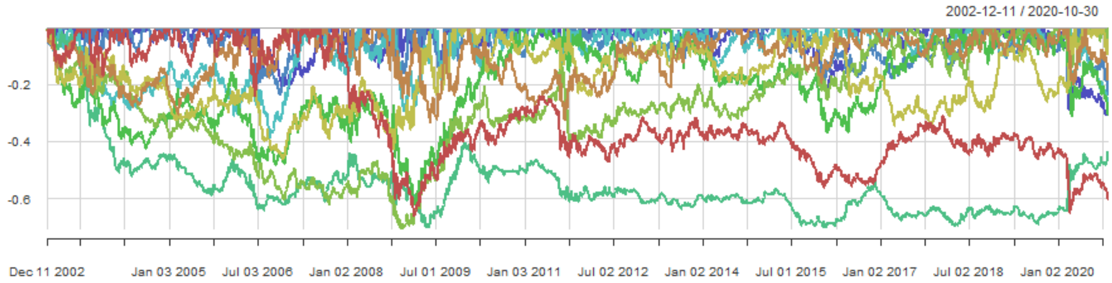
Figure 4.2. Daily returns for WIG20 (Poland)



Note: Figure shows daily returns of WIG20 index in the period from 2002-12-11 to 2020-10-30.

Another measure worth graphical analysis is the drawdown lines chart, i.e. the decline in the value of the portfolio calculated from the previous peak to the current trough achieved in the analyzed periods. Figure 4.3 presents those lines. The red line represents the buy-and-hold strategy which achieved the highest equity drops from all analyzed strategies mostly in the 2007-2009 great financial crisis period and COVID-19 pandemic period. Naïve Bayes and Random Forest strategies obtained the worst drawdowns from the group of machine learning strategies. The remaining strategies behaved in a more efficient manner. Portfolio value drops are inevitable in active investment management, but it is always worth to minimize them.

Figure 4.3. Drawdown lines for WIG20 (Poland)



Note: Figure shows drawdown lines for every strategy constructed on WIG20 index in the period from 2002-12-11 to 2020-10-30. Legend is inherited from Figure 4.1.

In the further step of the analysis, risk and return measures obtained by each investment strategy were compared among each other. The summary of those measures is presented in Table 4.1. In case of annualized returns (CAGR), Support Vector Machine models obtained the highest results with 47.96% for the Polynomial model and 26.19% for the Linear model. All remaining machine learning strategies besides Naïve Bayes and Regression Tree obtained CAGR higher than the benchmark (1.88%). The highest standard deviation of returns was observed for the benchmark strategy (22.67%) while the lowest was observed for Linear Support Vector Machine strategy (19.92%) which suggests that benchmark strategy is characterized by the highest risk. Results for Adjusted Sharpe Ratio are in line with those observed for CAGR measure. Highest Adjusted Sharpe was obtained by Polynomial Support Vector Machine (2.37) whereas NB and RT models scored lower than the benchmark strategy (0.08). In case of maximum drawdowns of capital, highest MDD was obtained by Random Forest model (70.45%) while the benchmark reported 65.75%. The lowest MDD was observed for Linear SVM model amounting to 23.85%. For the IR* measure, the best score of 3.69 was obtained by Polynomial SVM model whereas the result for the benchmark strategy was 0.002. As IR* was chosen as the most decisive in this research, Polynomial SVM was considered the best performing from all analyzed strategies. In the following sections strategies will be compared for the remaining indices.

Table 4.1. Risk and return measures for WIG20 (Poland)

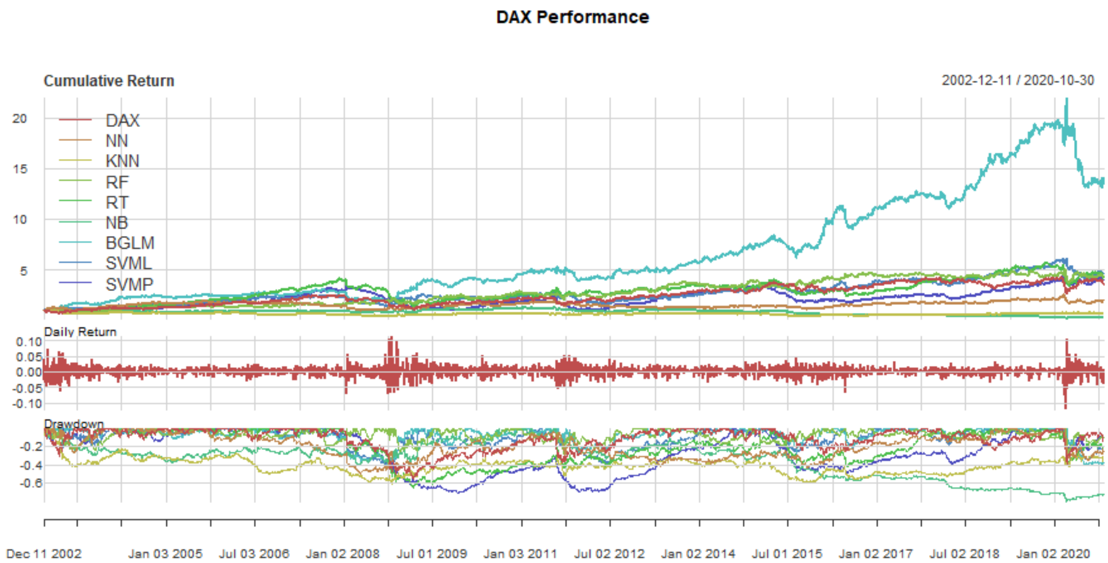
Measure	WIG20	NN	KNN	RF	RT	NB	BGLM	SVML	SVMP
CAGR	1.88%	8.85%	2.61%	3.94%	0.15%	-2.99%	17.02%	26.19%	**47.96%**
Annual. Std Dev	22.67%	20.00%	20.42%	20.42%	22.57%	21.04%	20.26%	**19.92%**	20.26%
Adj Sharpe	0.0830	0.4427	0.1278	0.1931	0.0065	0.0000	0.8398	1.3149	**2.3672**
MDD	65.75%	33.14%	46.62%	70.45%	62.18%	70.11%	38.65%	**23.85%**	30.79%
IR*	0.0024	0.1183	0.0072	0.0108	0.0000	0.0000	0.3698	1.4440	**3.6864**

Note: Table shows risk and return measures for strategies constructed on WIG20 index. The first column represents buy-and-hold strategy. Presented measures include: CAGR, annualized standard deviation, adjusted Sharpe Ratio, Maximum Drawdown and IR*. Bolded font indicates the best performance measure for all tested methods.

4.1.2. Investment strategies comparison for DAX (Germany)

For DAX index, 227 subsamples were created with dates ranging from 2002-03-01 to 2020-10-30. Figure 4.4 presents equity lines, daily returns and drawdown lines for all analyzed strategies. Bayesian Generalized Linear Model produced equity line with the highest overall return. The next best strategies were those produced by Linear Support Vector Machine and Random Forest models.

Figure 4.4. Equity lines, daily returns and drawdown lines for DAX (Germany)



Note: Figure shows equity lines, daily returns and drawdown lines for every strategy constructed on DAX index in the period from 2002-12-11 to 2020-10-30.

As in case of WIG20 index, in most of the periods of increased volatility of returns, the equity lines are also increasing e.g. in 2007-2009 great financial crisis period while in COVID-19 pandemic crisis period, this relationship is contradictory.

Highest equity drops from all analyzed strategies are observed mostly in the 2007-2009 great financial crisis period and COVID-19 pandemic period. Naïve Bayes, Polynomial Support Vector Machine, Regression Tree and K Nearest Neighbor models produced drawdowns worse than the benchmark strategy.

Table 4.2 presents risk and return measures calculated for all analyzed DAX strategies. In case of annualized returns (CAGR), Bayesian Generalized Linear Model obtained the highest results (15.53%). Second best model was the Linear Support Vector Machine (8.79%) and third – Random Forest model (8.75%). Benchmark buy-and-hold strategy obtained 7.42% CAGR. The highest standard deviation of returns was observed for the benchmark strategy (22.05%) while the lowest was observed for Naïve Bayes strategy (19.00%) which suggests that benchmark strategy is characterized by the highest risk as in case of WIG20 index. Results for Adjusted Sharpe Ratio were again in line with those observed for CAGR measure. Highest Adjusted Sharpe Ratios were obtained by BGLM (0.80), SVML (0.46) and RF (0.38) models, benchmark strategy scored 0.37. In case of maximum drawdowns, highest MDD was obtained by Naïve Bayes model (79.91%) while the benchmark reported 55.08%. The lowest MDD was observed for Random Forest model amounting to 27.72%. The best score of 0.31 for IR* measure was obtained by BGLM model whereas the result for the benchmark strategy was 0.05, therefore BGLM model was considered best performing from all analyzed strategies.

Table 4.2. Risk and return measures for DAX (Germany)

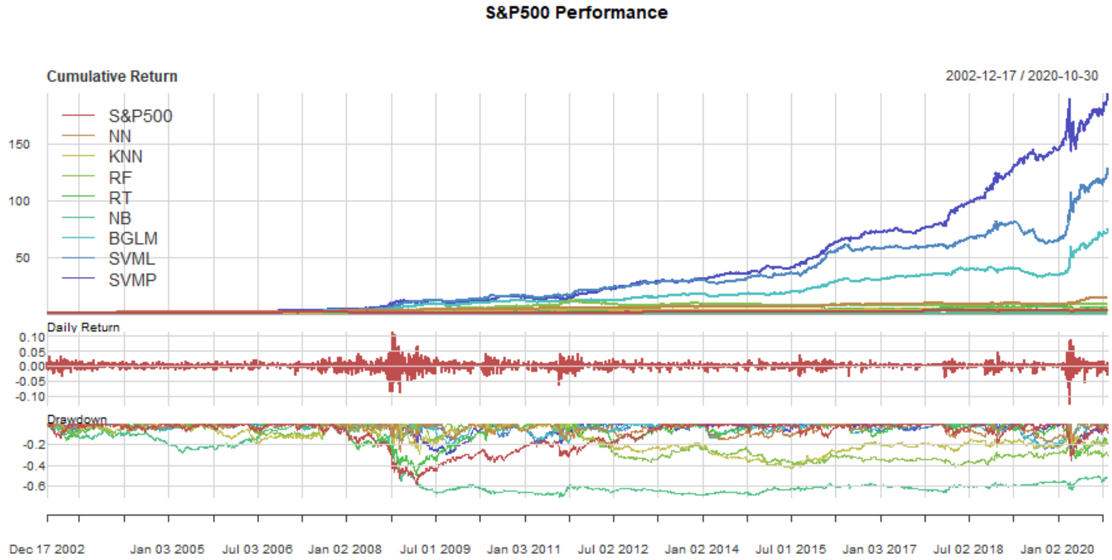
Measure	**DAX**	**NN**	**KNN**	**RF**	**RT**	**NB**	**BGLM**	**SVML**	**SVMP**
CAGR	7.42%	3.66%	-2.01%	8.75%	8.26%	-5.60%	**15.53%**	8.79%	7.88%
Annual. Std Dev	22.04%	19.89%	19.76%	19.82%	21.98%	**19.00%**	19.48%	19.23%	19.78%
Adj Sharpe	0.3368	0.1841	0.0000	0.4416	0.3759	0.0000	**0.7973**	0.4569	0.3986
MDD	55.08%	49.70%	60.40%	**27.72%**	64.73%	79.91%	40.48%	42.60%	70.39%
IR*	0.0454	0.0136	0.0000	0.1395	0.0480	0.0000	**0.3059**	0.0943	0.0446

Note: Table shows risk and return measures for strategies constructed on DAX index. The first column represents buy-and-hold strategy. Presented measures include: CAGR, annualized standard deviation, adjusted Sharpe Ratio, Maximum Drawdown and IR*. Bolded font indicates the best performance measure for all tested methods.

4.1.3. Investment strategies comparison for S&P500 (USA)

In case of S&P500 index, 225 subsamples were created with dates ranging from 2002-03-05 to 2020-10-30. Figure 4.5 presents equity lines, daily returns and drawdown lines for all analyzed strategies. Situation was similar to that observed for WIG20, both variations of Support Vector Machine strategies were dominant also for S&P500 index. The next best overall return was observed for Bayesian Generalized Linear Model which was also dominant for DAX index.

Figure 4.5. Equity lines, daily returns and drawdown lines for S&P500 (USA)



Note: Figure shows equity lines, daily returns and drawdown lines for every strategy constructed on S&P500 index in the period from 2002-12-17 to 2020-10-30.

As in case of WIG20 and DAX indices, in most of the periods of increased volatility of returns, the equity lines are also increasing. For S&P500 index it was observed not only in 2007-2009 great financial crisis period but also in COVID-19 pandemic crisis period. The worst capital drawdown was observed for Naïve Bayes model, while the remaining models behaved better than benchmark in that matter.

Risk and return measures for S&P500 strategies are presented in Table 4.3. In case of annualized returns (CAGR), SVMP model obtained the highest result (34.29%). The second best model was SVML (31.29%) and the third – BGLM model (27.39%). Benchmark buy-and-hold strategy obtained 6.94% CAGR. The highest standard deviation of returns was again observed for the benchmark strategy (19.21%) while the lowest was observed for Neural Networks strategy (17.01%). Results for Adjusted Sharpe Ratio were again in line with those observed for CAGR measure. The highest Adjusted Sharpe Ratios were obtained by SVMP (1.97), SVML (1.84) and BGLM (1.58) models whereas benchmark strategy scored 0.36. In case of maximum drawdowns, the highest MDD was observed for NB model (70.38%) while the benchmark reported 58.02%. The lowest MDD was observed for BGLM model amounting to 23.09%. The best score of 2.38 for IR* measure was obtained by SVML and the next best was observed for SVMP (2.32) whereas the result for the benchmark strategy was 0.04. Linear Support Vector Machine model was therefore considered best performing from all analyzed strategies.

Table 4.3. Risk and return measures for S&P500 (USA)

Measure	S&P500	NN	KNN	RF	RT	NB	BGLM	SVML	SVMP
CAGR	6.94%	16.01%	7.70%	12.62%	10.02%	-3.39%	27.39%	31.29%	**34.29%**
Annual. Std Dev	19.21%	**17.01%**	17.49%	17.37%	19.17%	18.04%	17.32%	17.03%	17.44%
Adj Sharpe	0.3613	0.9411	0.4406	0.7266	0.5226	0.0000	1.5815	1.8381	**1.9657**
MDD	58.02%	26.77%	42.92%	41.85%	49.71%	70.38%	**23.09%**	24.12%	29.02%
IR*	0.0432	0.5628	0.0791	0.2191	0.1053	0.0000	1.8759	**2.3847**	2.3224

Note: Table shows risk and return measures for strategies constructed on S&P500 index. The first column represents buy-and-hold strategy. Presented measures include: CAGR, annualized standard deviation, adjusted Sharpe Ratio, Maximum Drawdown and IR*. Bolded font indicates the best performance measure for all tested methods.

4.1.4. Investment strategies comparison for CEE indices

Results calculated for remaining CEE indices are described in this section in an aggregated manner. For the purpose of this thesis, strategies were compared only by the usage of the most decisive risk and return measure (IR*). Table 4.4 presents the strategies that received the highest IR* measure for each of the analyzed indices.

Naïve Bayes model was considered best performing for three out of six CEE indices with IR* value of 0.36 for SOFIX index (Bulgaria), 0.49 for OMXT index (Estonia) and 0.93 for OMXV index (Lithuania). BGLM model which was dominant in case of DAX index was also considered best performing for PX index (Czech Republic) with the score of 0.20 and OMXR index (Latvia) with the score of 0.82. SVML strategy obtained the highest IR* value (0.12) in case of BUX index (Hungary). SVML was also the best performing model for S&P500 index.

Table 4.4. Best performing models and corresponding IR* measures for CEE indices

Index	SOFIX	PX	OMXT	BUX	OMXR	OMXV
Strategy	NB	BGLM	NB	SVML	BGLM	NB
IR*	0.3594	0.2019	0.4898	0.1153	0.8185	0.9250

Note: Table shows IR* measure for the best performing strategies constructed on CEE indices: SOFIX, PX, OMXT, BUX, OMXR and OMXV.

The next section summarizes the results from all analyzed indices and aims to propose a robust approach for the best performing model selection.

4.1.5. Summary of investment strategies comparison for all indices

As shown in the previous sections, results varied across all analyzed indices. There was no strategy that could be considered the best performing for all instruments. Table 4.5 similarly to Table 4.4 presents the strategies for which the highest IR* score was observed and the corresponding names of the models (strategies) for every index analyzed in this research. It is worth noting that the IR* values for benchmark buy-and-hold strategy were in all cases lower than those obtained by best performing machine learning models.

Bayesian Generalized Linear Model performed best for three indices i.e. DAX, PX and OMXR and the same situation applies to Naïve Bayes model which performed best for SOFIX, OMXT and OMXV indices. Linear Support Vector Machine model received the best IR* score for two indices i.e. S&P500 and BUX. Polynomial variation of SVM was considered the best for WIG20 index.

Table 4.5. Best performing models and corresponding IR* measures for all analyzed indices

Index	**WIG20**	**DAX**	**S&P500**	**SOFIX**	**PX**	**OMXT**	**BUX**	**OMXR**	**OMXV**
Strategy	SVMP	BGLM	SVML	NB	BGLM	NB	SVML	BGLM	NB
IR*	3.6864	0.3059	2.3847	0.3594	0.2019	0.4898	0.1153	0.8185	0.9250

Note: Table shows IR* measure for the best performing strategies constructed on all analyzed indices: WIG20, DAX, S&P500, SOFIX, PX, OMXT, BUX, OMXR and OMXV.

In order to compare 9 analyzed strategies (8 machine learning models strategies plus the benchmark buy-and-hold strategy) across indices, a rank approach was introduced. Alternative comparison approaches were initially analyzed including different implementations of the ranking process (e.g. focusing on absolute differences between the values before the ranking process) but were rejected due to the outcomes not providing any additional information and being more biased than in the proposed approach. For each index and for each risk and return measure, strategies were ranked from 1 to 9 where 9 constitutes the highest score. For example in case of WIG20, SVMP strategy had the highest CAGR measure and received score equal to 9 in CAGR category whereas for MDD measure SVMP received score equal to 8 as NN strategy had the lowest maximum drawdown. Ranks were then averaged across all analyzed indices and presented in Table 4.6. The second column of the Table corresponds to benchmark buy-and-hold strategy (B&H).

Table 4.6. Ranked risk and return measures averaged across all analyzed indices

Measure	**B&H**	**NN**	**KNN**	**RF**	**RT**	**NB**	**BGLM**	**SVML**	**SVMP**
CAGR	5.33	4.56	3.44	5.78	5.22	3.78	**6.44**	6.11	4.33
Annual. Std Dev	1.06	5.89	5.61	6.50	1.94	3.78	6.50	**7.33**	6.39
Adj Sharpe	5.33	4.33	3.17	6.00	4.89	3.83	**6.44**	6.28	4.72
MDD	2.44	6.00	5.89	6.44	4.56	4.67	**6.78**	5.33	2.89
IR*	5.22	4.56	3.39	6.22	4.89	3.83	**6.44**	6.28	4.17

Note: Table shows ranked risk and return measures: CAGR, annualized standard deviation, adjusted Sharpe Ratio, Maximum Drawdown and IR* averaged across all analyzed indices. Bolded font indicates the best performance ranked measure for all tested methods.

In case of annualized returns (CAGR), BGLM model obtained the highest averaged rank (6.44). The second best model was SVML (6.11) and the third – RF model (5.78). Benchmark buy-and-hold strategy obtained 5.33 averaged rank for CAGR. The worst averaged ranked score for standard deviation of returns was observed for the benchmark strategy (1.06) while the best was observed for SVML strategy (7.33). Results for the Adjusted Sharpe Ratio averaged ranks were in line with those observed for CAGR measure. The highest Adjusted Sharpe Ratio averaged ranks were obtained by BGLM (6.44), SVML (6.28) and RF (6.00) models whereas benchmark strategy scored 5.33. In case of maximum drawdowns, the worst MDD averaged rank was observed for the benchmark strategy (2.44) while the best result was obtained by BGLM model (6.78). The best score of 6.44 for IR* averaged rank was obtained by BGML, the second best was observed for SVML (6.28) and the third best for RF (6.22) whereas the result for the benchmark strategy was 5.22. Based on this analysis Bayesian Generalized Linear Model (BGLM) was considered producing the most robust results across all analyzed indices. The following sections describe sensitivity analysis performed to investigate if this conclusion changes when underlying models' parameters are altered.

4.2. Sensitivity to technical analysis indicators

Technical analysis indicators served as inputs for machine learning models and therefore their levels are impacting trading signal generation in the analyzed strategies. Each technical indicator was calculated based on its underlying parameters which in all cases determine how many periods of observable stock index quotes were included in the calculation. Base parameters used in previous sections as well as parameters used for sensitivity analysis were described in Technical Analysis Indicators section. Sensitivity analysis was performed in two scenarios. In the first scenario each base parameter was decreased by 1 which means that one period less was considered in computing technical indicators. In the second scenario each base parameters was increased by 1, which means that one period more was considered in computing the indicators. In order for the results to be comparable with the previous section, similar approach to presentation of the results was implemented.

4.2.1. Scenario with decreased technical indicators' parameters

Table 4.7 presents the strategies for which the highest IR* score was observed and corresponding names of the models (strategies) for each analyzed index. Bayesian Generalized Linear Model strategy performed best for four indices i.e. DAX, PX, BUX and OMXR. SVML model was considered the best for BUX. Naïve Bayes strategy once more performed the best for three indices i.e. SOFIX, OMXT and OMXV. Polynomial Support Vector Machine strategy received the best IR* score for two indices i.e. WIG20 and S&P500.

Table 4.7. The best performing models and corresponding IR* measures for all analyzed indices in scenario with decreased technical indicators' parameters

Index	WIG20	DAX	S&P500	SOFIX	PX	OMXT	BUX	OMXR	OMXV
Strategy	SVMP	BGLM	SVMP	NB	BGLM	NB	BGLM	BGLM	NB
IR*	3.2691	0.2845	2.2332	0.5500	0.3012	0.8015	0.2214	0.9658	1.0345

Note: Table shows IR* measure for the best performing strategies constructed on all analyzed indices in the sensitivity analysis scenario with decreased technical indicators' parameters.

Ranks averaged across all indices analogous to those described for the base scenario are presented in Table 4.8. In case of annualized returns (CAGR), BGLM model once again obtained the highest averaged rank (6.78). The second best model was SVML (5.78) and the third – the benchmark strategy (5.67) which performed better compared to base scenario. The worst averaged ranked score for standard deviation of returns was observed for the benchmark strategy (1.06) while the best was observed for SVML strategy (7.44). Results for Adjusted Sharpe Ratio averaged ranks were in line with those observed for CAGR measure. Highest Adjusted Sharpe Ratio averaged ranks were obtained by BGLM (6.61), SVML (5.61) and the benchmark strategy (5.56). In case of maximum drawdowns, the worst MDD averaged rank was observed once again for the benchmark strategy (2.67) while the best result was obtained by RF model (6.78). In the base scenario BGLM model received the best score for MDD averaged rank. The best score of 6.61 for IR* averaged rank was obtained by BGML, the second best was observed for SVML (5.61) and the third best for NN (5.56) whereas the result for the benchmark strategy was 5.44. The final conclusion from the base scenario also applies to this scenario as the Bayesian Generalized Linear Model (BGLM) received the most robust results across all analyzed indices.

Table 4.8. Ranked risk and return measures averaged across all analyzed indices in scenario with decreased technical indicators' parameters

Measure	B&H	NN	KNN	RF	RT	NB	BGLM	SVML	SVMP
CAGR	5.67	5.33	2.89	5.00	5.33	3.78	**6.78**	5.78	4.44
Annual. Std Dev	1.06	6.22	6.33	6.11	1.94	3.89	5.89	**7.44**	6.11
Adj Sharpe	5.56	5.33	3.11	5.11	5.17	3.72	**6.61**	5.61	4.78
MDD	2.67	6.33	5.44	**6.78**	4.56	4.56	6.56	5.00	3.11
IR*	5.44	5.56	3.11	5.33	5.17	3.72	**6.61**	5.61	4.44

Note: Table shows ranked risk and return measures: CAGR, annualized standard deviation, adjusted Sharpe Ratio, Maximum Drawdown and IR* averaged across all analyzed indices in the sensitivity analysis scenario with decreased technical indicators' parameters. Bolded font indicates the best performance ranked measure for all tested methods.

4.2.2. Scenario with increased technical indicators' parameters

Strategies for which the highest IR* score was observed in this scenario and the corresponding names of the models (strategies) are presented in Table 4.9. Bayesian Generalized Linear Model strategy performed the best for two indices i.e. DAX and S&P500 which is a downgrade from results obtained in base and decreased parameter scenarios. Naïve Bayes strategy again performed the best for three indices i.e. SOFIX, OMXT and OMXV. Linear Support Vector Machine strategy received the best IR* score for two indices i.e. BUX and OMXR. Polynomial variation of SVM was considered the best for WIG20 index as in the base scenario. Neural Networks strategy was for the first time identified as the best performer for PX index.

Table 4.9. The best performing models and corresponding IR* measures for all analyzed indices in scenario with increased technical indicators' parameters

Index	**WIG20**	**DAX**	**S&P500**	**SOFIX**	**PX**	**OMXT**	**BUX**	**OMXR**	**OMXV**
Strategy	SVMP	BGLM	BGLM	NB	NN	NB	SVML	SVML	NB
IR*	3.0824	0.1668	2.3904	0.3986	0.1331	0.5323	0.1555	0.9080	0.7854

Note: Table shows IR* measure for the best performing strategies constructed on all analyzed indices in the sensitivity analysis scenario with increased technical indicators' parameters.

Averaged ranks for increased parameters scenario are presented in Table 4.10. In case of annualized returns (CAGR), SVML (instead of BGLM in the previous scenarios) strategy obtained the highest averaged rank (6.11). The second best model was BGLM (6.00) and the third –NN strategy (5.89). Benchmark buy-and-hold strategy obtained 5.56 averaged rank for CAGR. The worst averaged ranked score for standard deviation of returns was observed for the benchmark strategy (1.11), while the best was observed for RF (instead of SVML in the previous scenarios) strategy with the score equal to 7.00. Results for Adjusted Sharpe Ratio averaged ranks in this scenario were not in line with those observed for CAGR measure. The highest Adjusted Sharpe Ratio averaged ranks were obtained by BGLM (6.17), SVML (6.06) and RT (5.67), while the benchmark strategy scored 5.44. In case of maximum drawdowns, the worst MDD averaged rank was observed for SVMP (instead of the benchmark in the previous scenarios) strategy with the score of 2.44 while the best result was obtained by RF model (7.00) similarly to the decreased parameters scenario. In the base scenario BGLM model received the best score for MDD averaged rank. The best score of 6.17 for IR* averaged rank was obtained by both BGML and SVML strategies and the third best amounting to 5.67 for both NN and RT strategies whereas the result for the benchmark strategy was 5.11. Despite BGLM and SVML models receiving equal IR* scores, after a comparison of adjusted Sharpe Ratio averaged ranks for these models (6.17 for BGLM vs 6.06 for SVML), it can be concluded that BGLM was the most robust strategy as in previous scenarios.

Table 4.10. Ranked risk and return measures averaged across all analyzed indices in scenario with increased technical indicators' parameters

Measure	B&H	NN	KNN	RF	RT	NB	BGLM	SVML	SVMP
CAGR	5.56	5.89	3.22	4.78	5.67	4.00	6.00	**6.11**	3.78
Annual. Std Dev	1.11	5.78	5.89	**7.00**	1.89	3.67	6.56	6.89	6.22
Adj Sharpe	5.44	5.56	3.11	4.67	5.67	4.28	**6.17**	6.06	4.06
MDD	2.67	6.67	4.67	**7.00**	5.00	4.44	6.33	5.78	2.44
IR*	5.11	5.67	3.22	4.89	5.67	4.28	**6.17**	**6.17**	3.83

Note: Table shows ranked risk and return measures: CAGR, annualized standard deviation, adjusted Sharpe Ratio, Maximum Drawdown and IR* averaged across all analyzed indices in the sensitivity analysis scenario with increased technical indicators' parameters. Bolded font indicates the best performance ranked measure for all tested methods.

4.3. Sensitivity to machine learning optimization metrics

Every machine learning technique used in this research has a broad set of underlying hyperparameters which differs across the models. As there was no universal hyperparameter to be altered, the models were divided into three categories i.e. Neural Networks (considered a distinct category due to the implemented model algorithm not allowing for alteration of optimization metrics contrary to the other regression models), classification models (comprising of NB model) and regression models (comprising of the remaining six models). This sensitivity analysis exercise focused on assessing each model category separately in order to investigate if the strategies constructed from machine learning models' outputs are prone to changes in hyperparameters. Sensitivity assessment was based solely on IR* risk and return measure results as it was considered the most decisive in this thesis.

Model outputs as well as the corresponding trading signals were computed for all forms of the altered hyperparameter. It allowed for calculation of IR* measure in multiple scenarios for each of the analyzed stock indices. IR* was then compared and scenarios were ranked, with the highest score regarded as the best. Rank was then averaged across all analyzed indices.

4.3.1. Sensitivity of Neural Networks

In case of NN models, hyperparameter altered in sensitivity analysis was the activation function which was investigated in three different forms (scenarios) i.e. tansig (base function used in this thesis), sin and satlins as described in Machine Learning Techniques section, results of which are presented in Table 4.11.

Table 4.11. Ranked IR* measure averaged across all analyzed indices for Neural Networks in three activation function scenarios

Activation Function	Neural Networks
Tansig	1.72
Sin	**2.33**
Satlins	1.94

Note: Table shows ranked IR* measure for tansig, sin and satlins activation functions applied in NN models averaged across all analyzed indices. Bolded font indicates the best performance ranked measure for all tested methods.

Activation function sin received the best averaged rank (2.33) for all tested indices, satlins received 1.94 and tansig received 1.72 score. Those results can be interpreted in the following manner: by altering the activation function to sin, on average the NN models produce returns with higher IR* measure than those obtained from the employment of satlins and tansig functions.

4.3.2. Sensitivity of classification models

Naïve Bayes model is the only classification model described in this thesis. Hyperparameter that was altered in sensitivity analysis was the optimization metric investigated in two different versions (scenarios) i.e. accuracy (base metric used in this thesis) and kappa as described in Machine Learning Techniques section, results of which are presented in Table 4.12.

Table 4.12. Ranked IR* measure averaged across all analyzed indices for classification models in two optimization metric scenarios

Optimization metric	Naive Bayes
Accuracy	**1.50**
Kappa	**1.50**

Note: Table shows ranked IR* measure for accuracy and kappa optimization metrics applied in NB models averaged across all analyzed indices. Bolded font indicates the best performance ranked measure for all tested methods.

Although the metric ranks differed on index levels, both optimization metrics received the same averaged rank (1.50) which means that Naïve Bayes model is on average resistant to optimization metrics alteration.

4.3.3. Sensitivity of regression models

Regression models category comprises of six models i.e. KNN, RF, RT, BGLM, SVML and SVMP. As in case of classification models the hyperparameter altered in sensitivity analysis was the optimization metric. As described in Machine Learning Techniques section, optimization metrics for regression models differed from those that could be applied for Neural Networks and classification models. Metrics were therefore investigated in three different versions (scenarios) i.e. RMSE (root mean square error – base metric used in this thesis), Rsquared (coefficient of determination) and MAE (mean absolute error), results of which are presented in Table 4.13.

Table 4.13. Ranked IR* measure averaged across all analyzed indices for regression models in three optimization metric scenarios

Optimization metric	KNN	RF	RT	BGLM	SVML	SVMP
RMSE	**2.22**	**2.22**	**2.22**	**2.00**	**2.00**	1.89
Rsquared	1.83	1.78	1.72	**2.00**	**2.00**	**2.44**
MAE	1.94	2.00	2.06	**2.00**	**2.00**	1.67

Note: Table shows ranked IR* measure for RMSE, Rsquared and MAE optimization metrics applied in regression models averaged across all analyzed indices. Bolded font indicates the best performance ranked measure for all tested methods.

Models computed with the RMSE metric generated on average the best IR* values for K Nearest Neighbor, Random Forest and Regression Tree models (2.22 score). In case of Bayesian Generalized Linear Model and Linear Support Vector Machine, models obtained the same averaged rank (2.00) for every metric which means that they are resistant to optimization metric alteration. Polynomial Support Vector Machine model obtained the highest averaged rank of 2.44 when computed with Rsquared metric which means that on average SVMP models produce higher IR* when Rsquared metric is applied.

4.4. Summary of the empirical research

Researching the set of nine different investment strategies allowed to obtain answers to the research hypotheses stated in the thesis. With the employment of risk and return measures such as CAGR, standard deviation, maximum drawdown, adjusted Sharpe and adjusted Information Ratio as a the selection criteria, it was possible to select best performing strategy. Eight machine learning based strategies served as the subject of the research along with the benchmark buy-and-hold strategy. It was assessed whether it is possible to generate excess profits by investing using technical analysis indicators and machine learning methods.

Models were fitted and applied on dynamic windows of in sample and out of sample subsets with the dates ranging from 2002 to 2020 for multiple global and CEE stock indices. For WIG20 index (Poland) the Polynomial Support Vector Machine strategy achieved a 47.96% annualized rate of return, which is materially superior to 1.88% obtained by the benchmark strategy. Moreover, the main risk and return measure – adjusted Information Ratio also showed better results for SVMP strategy with the score of 3.69 compared to 0.002 achieved by the benchmark.

The results for DAX (Germany) and S&P500 (USA) machine learning strategies were also above the benchmark. For DAX, the best performing strategy was based on Bayesian

Generalized Linear Model with 15.53% CAGR and 0.31 IR* while benchmark achieved 7.42% CAGR and 0.05 IR*. In case of S&P500, the Linear Support Vector Machine strategy obtained the best results with CAGR and IR* amounting to 31.29% and 2.38 respectively while benchmark buy-and-hold strategy resulted in 6.94% CAGR and 0.04 IR*.

As part of the research, results for six CEE stock indices were also analyzed. In all cases, a strategy superior to the benchmark was identified. Naïve Bayes strategy was considered as the best performing for three indices with IR* value of 0.36 for SOFIX index (Bulgaria), 0.49 for OMXT index (Estonia) and 0.93 for OMXV index (Lithuania). BGLM strategy was considered the best performing for PX index (Czech Republic) with the IR* of 0.20 and OMXR index (Latvia) with the IR* of 0.82. SVML strategy obtained the highest IR* value (0.12) in case of BUX index (Hungary).

Strategies were ranked according to the value of IR* for each index separately and the ranks were then averaged among indices. The best score of 6.44 for IR* averaged rank was obtained by BGML, the second best was observed for SVML (6.28) and the third best for RF (6.22) whereas the result for the benchmark strategy was 5.22.

Based on the evaluation of achieved returns over time for each analyzed index, it was observed that in periods of the highest market volatility i.e. 2007-2009 great financial crisis and COVID-19 pandemic crisis periods; returns obtained by the best performing strategies tended to fluctuate significantly. In case of WIG20 (Poland), DAX (Germany) and SOFIX (Bulgaria) cumulative returns from the strategies increased during 2007-2009 great financial crisis and decreased during COVID-19 pandemic crisis. Opposite situation was observed for BUX (Hungary) and OMXR (Latvia). The best performing strategies for S&P500 (USA), OMXT (Estonia), OMXV (Lithuania) and PX (Czech Republic) indices recorded cumulative return increase in both 2007-2009 great financial crisis and COVID-19 pandemic crisis.

Sensitivity analysis was conducted in order to investigate whether a change (decrease and increase) in the parameters underlying technical analysis indicators which serve as inputs to the models would impact the results. In case of the decreased parameters scenario, the best score of 6.61 for IR* averaged rank was obtained by BGML, the second best was observed for SVML (5.61) and the third best for NN (5.56) whereas the result for the benchmark strategy was 5.44. In the increased parameters scenario, the best score of 6.17 for IR* averaged rank was obtained by both BGML and SVML strategies and the third best amounting to 5.67 for both NN and RT strategies whereas the result for the benchmark strategy was 5.11. Based on this analysis Bayesian Generalized Linear Model (BGLM) was considered to be producing the most robust results across all analyzed indices in all scenarios.

The second part of sensitivity analysis investigated the impact of changing the hyperparameters underlying machine learning models. As there was no universal hyperparameter to be altered, activation function was chosen for Neural Networks strategies and optimization metric was chosen for the remaining strategies. Results were again ranked and averaged across indices. By altering the activation function to sin, on average, NN models produced returns with higher IR* measure than those obtained from the employment of tansig (base function used in this research) and satlins functions. Naïve Bayes models were on average resistant to optimization metric (Accuracy and Kappa) alteration. In case of the remaining strategies, models computed with the RMSE metric generated on average the best IR* values for K Nearest Neighbor, Random Forest and Regression Tree models. Polynomial Support Vector Machine model obtained the highest averaged rank when computed with Rsquared metric which means that on average SVMP models produce higher IR* when Rsquared metric is applied. In case of Bayesian Generalized Linear Model and Linear Support Vector Machine, models obtained the same averaged rank for every metric which means that they are resistant to metric alteration.

Results obtained for all of the analyzed stock market indices indicated that quantitative investment strategies achieved better results measured by adjusted Information Ratio (IR*) than the benchmark buy-and-hold strategies and therefore the first research hypothesis cannot be rejected. Neural Networks strategies were not considered as the best performing for any of the analyzed indices thus the second research hypothesis was rejected. Analysis of each index allowed to select the best performing strategies which were not consistent among indices and therefore the third research hypothesis was rejected, although at the same time analysis indicated that BGLM model on average produced the best results. Fourth research hypothesis was also rejected as the results changed for most of the analyzed machine learning techniques when computed in different sensitivity analysis scenarios. On average however, the Bayesian Generalized Linear Model generated the best results in all sensitivity analysis scenarios in which the technical analysis indicators' parameters were altered. In case of altering the machine learning models' hyperparameters, the BGLM model was considered to be resistant to changes.

Conclusion

New technologies allowed for transfer of stock exchange trading from the trading floor to the computer screen, and also opened the possibility of automating the investment process i.e. concluding transactions with limited or even without human intervention. Automated investment strategies are now widely used by hedge funds and rely among others on rules derived from technical analysis. Technical analysis is a rich set of tools supporting investment decision making. An investor can use them directly in the construction of buy and sell signals as well as indirectly by treating them as inputs to more sophisticated models. Employment of machine learning techniques allows to generate weights for every technical analysis indicator used and therefore producing trading signals based on both the levels and the weights estimated for the technical indicators.

A set of five technical analysis indicators was analyzed in this thesis and consisted of: Simple Moving Average (SMA), Moving Average Convergence Divergence (MACD), Stochastic Oscillator (STOCH), Relative Strength Index (RSI) and Williams' Percent Range (WPR) as proposed by Dash and Dash (2016). Technical indicators were then used as an input to eight machine learning models. The techniques analyzed were: Neural Networks (NN), K Nearest Neighbor (KNN), Regression Tree (RT), Random Forest (RF), Naïve Bayes (NB), Bayesian Generalized Linear Model (BGLM), Linear Support Vector Machine (SVML) and Polynomial Support Vector Machine (SVMP).

The purpose of this study was to investigate the profitability of machine learning-based quantitative investment strategies. Performance of the strategies was determined by comparison of the risk and return measures among models and the benchmark buy-and-hold returns. Annualized rate of return (CAGR), annualized standard deviation of returns, adjusted Sharpe Ratio (SR), maximum drawdown (MDD) and the adjusted Information Ratio (IR*) were used as the aforementioned risk and return measures with the IR* considered as the most decisive.

The thesis was based on the current scientific achievements describing the mechanisms of generating trading signals from machine learning models employing technical indicators as inputs. As part of the extension of this branch of science, this research was conducted on the Polish stock market index WIG20, two highly liquid equity indices: DAX (Germany) and S&P500 (USA) as well as the indices of six Central and Eastern European countries: SOFIX (Bulgaria), PX (Czech Republic), OMXT (Estonia), BUX (Hungary), OMXR (Latvia) and OMXV (Lithuania). Data used for the calculations included the daily High, Low and Close prices of the indices in 2002-2020 period thus including the 2007-2009 great financial crisis and COVID-19 pandemic crisis. Models were fitted in dynamic window in sample and applied on out of sample subsets separately for each of the analyzed indices.

All calculations underlying the results presented in this thesis were produced in R statistical software. Each scenario (base scenario and scenarios representing sensitivity analyses) took ca. 5 hours of computation on 2.8GHz processor with 2 cores.

Results showed that in case of each index, machine learning techniques-based strategies achieved better returns than the benchmarks. The best performing (based on IR* measure) strategy was constructed from Polynomial Support Vector Machine model in case of WIG20 index (Poland), Bayesian Generalized Linear Model for DAX (Germany) and Linear Support Vector Machine for S&P500 (USA). Comparison of those results with the ones obtained for six CEE indices showed that on average the BGLM strategy generated the best risk adjusted returns.

Evaluation of cumulative returns over time indicated that the impact of increased market volatility periods i.e. 2007-2009 great financial crisis and COVID-19 pandemic crisis varied between analyzed indices. Returns increased during 2007-2009 great financial crisis and decreased during COVID-19 pandemic crisis for WIG20, DAX and SOFIX strategies whereas opposite situation was observed for BUX and OMXR. In case of S&P500, OMXT, OMXV and PX strategies, returns increased in both 2007-2009 great financial crisis and COVID-19 pandemic crisis.

In order to investigate the robustness of analyzed quantitative strategies, the sensitivity analysis was conducted. It was checked whether the change (decrease and increase) in the parameters underlying technical analysis indicators which served as inputs to the models would impact the results. Altering the parameters in both directions showed that on average the Bayesian Generalized Linear Model (BGLM) produced the most robust results across analyzed indices. The second part of sensitivity analysis investigated the impact of changing the hyperparameters underlying the machine learning models. In case of Bayesian Generalized Linear Model, results showed that returns generated by the model are resistant to optimization metric alteration.

The returns obtained from the trading signals generated by machine learning models indicated that quantitative investments strategies achieved better performance measured by adjusted Information Ratio than the benchmark buy-and-hold strategies for all of the analyzed stock market indices and therefore the main research hypothesis cannot be rejected. It should be stated that the range of financial instruments available for investment is very wide and in order to unequivocally assess this hypothesis, a larger number of instruments could be tested. The additional research hypotheses stated in this thesis were formulated with the aim to assess the best performing machine learning techniques across the analyzed indices. The analysis of each index allowed to select the best performing strategy which differed among indices but it was observed that on average the Bayesian Generalized Linear Model produced the best risk adjusted returns thus the second and the third hypothesis were rejected. The sensitivity analysis results showed that the fourth hypothesis also needs to be rejected. On a strategy level, results changed in each analyzed scenario for most of the analyzed models. On average however, the Bayesian Generalized Linear Model generated the best results in all sensitivity analysis scenarios in which the technical analysis indicators' parameters were altered. In case of altering the machine learning models' hyperparameters, the BGLM model was considered to be resistant to changes.

In order to obtain a comprehensive overview of the topics presented in this thesis, future research could be expanded to include financial instruments from other categories e.g. stocks, bonds, options or currency futures. An interesting research direction would also be to investigate how the results would change if an assumption of the existence of transaction costs was taken, which has not been the case in this thesis.

References

Altman N. S. *An Introduction to Kernel and Nearest Neighbor Nonparametric Regression.* Cornell University 1991.

Anson M., Fabozzi F., Jones F. *The Handbook of Traditional and Alternative Investment Vehicles: Investment Characteristics and Strategies*. Hoboken 2010.

Appel G. *Technical Analysis Power Tools for Active Investors*. New Jersey 2005, p. 166.

Bacon C. *Practical Risk-adjusted Performance Measurement.* Manchester 2010.

Beechey M., Gruen, D., Vickery J. *The Efficiency Market Hypothesis: A Survey*. Sydney 2000.

Boser B. E., Guyon I. M., Vapnik V. N. A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory – COLT '92,* p. 144.

Breiman L. Random Forests. *Machine Learning.* Vol. XLV 2001, pp. 5-32.

Breiman L., Friedman J. H., Olshen R. A., Stone C. J. *Classification and Regression Trees.* Wadsworth 1984.

Brock W., Lakonishok J., LeBaron B. Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance.* Vol. XLVII 1992, pp. 1731-1764.

Chen W-H., Shih J-Y., Wu S. Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. *International Journal of Electronic Finance.* Vol. I 2006, pp. 49-67.

Colianni S., Rosales S., Signorotti M. *Algorithmic Trading of Cryptocurrency Based on Twitter Sentiment Analysis.* Stanford University 2015.

Cortes C., Vapnik V. N. Support-vector networks. *Machine Learning.* Vol. XX 1995, pp. 273–297.

Dash R., Dash P. K. A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science.* Vol. II 2016, pp. 42-57.

Dempster A. P., Laird N. M., Rubin D. B. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society: Series B.* Vol. XXXIX 1977, pp. 1–38.

Fama E. Efficient Capital Markets: A review of theory and empirical work. *Journal of Finance*. Vol. XXV 1970, pp. 383-417.

Fama E. The Behavior of Stock Market Prices. *Journal of Business*. Vol. XXXVIII 1965, pp. 34-105.

Feller W. *Introduction to Probability Theory and its Applications.* Princeton 1950.

Gerlein E., McGinnity M., Belatreche A., Coleman S. Evaluating machine learning classification for financial trading: An empirical approach. *Expert Systems with Applications.* Vol. LIV 2016, pp. 193-207.

Han J., Kamber M., Pei J. *Data Mining: Concepts and Techniques.* Amsterdam 2011.

Huang G., Zhu Q., Siew C. Extreme learning machine: Theory and applications. *Neurocomputing.* Vol. LXX 2006, pp. 489-501.

Huang W., Nakamori Y., Wang S. Forecasting stock market movement direction with support vector machine. *Computers & Operations Research.* Vol. XXXII 2005, pp. 2513-2522.

Jiang H., Shen S., Zhang T. *Stock Market Forecasting Using Machine Learning Algorithms.* Stanford University 2012.

Kijewski M., Ślepaczuk R., 2020, *Predicting prices of S&P500 index using classical methods and recurrent neural networks*, Working Papers of Faculty of Economic Sciences, University of Warsaw, WP 27/2020 (333).

Kość K., Sakowski P., Ślepaczuk R. Momentum and contrarian effects on the cryptocurrency market. *Physica A: Statistical Mechanics and its Applications.* Vol. DXXIII 2019, pp. 691-701.

Lane G. Lane's Stochastics. *Technical Analysis of Stocks and Commodities.* Vol. II 1984, pp. 87-90.

Leigh W., Purvis R., Ragusa J. M. Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: a case study in romantic decision support. *Decision Support Systems.* Vol. XXXII 2002, pp. 361-377.

Lin C-H., Lin S-Y., Wu M-C. An effective application of decision tree to stock trading. *Expert Systems with Applications.* Vol. XXXI 2006, pp. 270-274.

Madan I., Saluja S., Zhao A. *Automated Bitcoin Trading via Machine Learning Algorithms.* Stanford University 2015.

Magdon-Ismail M., Pratap A., Atiya A., Abu-Mostafa Y. *On the Maximum Drawdown of a Brownian Motion.* Pasadena 2003.

Narasimha Murty M., Susheela Devi V. *Pattern Recognition: An Algorithmic Approach.* Bangalore 2011.

Nelder J., Wedderburn R. Generalized Linear Models. *Journal of the Royal Statistical Society: Series A (General).* Vol. CXXXV 1972, pp. 370–384.

Reuters Limited. *Introduction to Technical Analysis.* London 1999.

Tian G., Wan G., Guo M. Market efficiency and the returns to simple technical trading rules: new evidence from U.S. equity market and Chinese equity markets. *Asia-Pacific Financial Market.* Vol. IX 2002, pp. 241-258.

Young T. Calmar Ratio: A Smoother Tool. *Futures*. Vol. XVI 1991, pp. 5-21.

Wilder J. *New Concepts in Technical Trading Systems*. Greensboro 1978.

Williams L. *How I Made One Million Dollars Last Year Trading Commodities*. Babylon 1979.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl