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

Jan Grudniewicz and Robert Ślepaczuk
QFRG monthly meeting

Quantitative Finance Research Group
Faculty of Economic Sciences at University of Warsaw

11th April, 2022

Agenda I

- Motivation, hypotheses and research questions
- Literature review
- Data
  - Data description
  - Sampling
  - Initial data analysis
- Research methodology
  - General model formula and target variable
  - Technical analysis indicators
  - Machine learning techniques
  - Investment strategies construction
  - Risk and return measures

Agenda II

- Empirical results
  - Identification of the best performing strategy
  - Sensitivity to technical analysis indicators
  - Sensitivity to machine learning optimization metrics
  - Summary of the empirical research
- Conclusions
- Research extensions

Motivation-1

The main aims are:

- Determine the best performing strategy among the strategies constructed using 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,
- Compare the strategies with the method of passive capital management based on the buy-and-hold mechanism using risk and return measures, such as the annualized rate of return (CAGR), standard deviation of returns (ASD), maximum drawdown (MD), Sharpe Ratio (SR) and Information Ratio* (IR*),
- To test the robustness of presented strategies with regards to various parameters.

Motivation-2

Research extends the current achievements of scientific research by:

- Deploying algorithmic investment strategies on the wide range of stock market indices:
    - on domestic market: Poland (WIG20),
    - two highly developed countries: Germany (DAX) and USA (S&P500),
    - six emerging countries from Central and Eastern Europe: Bulgaria (SOFIX), Czech Republic (PX), Estonia (OMXT), Hungary (BUX), Latvia (OMXR) and Lithuania (OMXV).
- Covering periods of:
    - the great financial crisis of 2007-2009,
    - COVID-19 pandemic crisis.

Hypotheses and research questions

First Hypothesis:

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.

Second Hypothesis:

Neural Networks generate the best (in terms of risk adjusted returns) investment signals compared to other machine learning techniques used in the research.

Third Hypothesis:

The very same machine learning strategy is considered the best performing for all analyzed stock market indices.

Fourth Hypothesis:

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.

Literature overview

Literature overview

- **Dash and Dash (2016)** discussed the profitability of investment strategy constructed with Extreme Learning Machine (ELM) and technical analysis indicators such as SMA, MACD, SO, RSI and WPR. Data used in the research consisted of daily quotes of two indices: BSE SENSEX and S&P 500 from 2010-2014 period. Results showed that ELM model produced the highest returns compared to Support Vector Machines, Naive Bayesian model, K Nearest Neighbor model and Decision Tree models.
- **Jiang et al. (2012)** predicted trends using Support Vector Machine, Multiple Additive Regression Trees, linear regression and generalized linear model (GLM) on NASDAQ, DJIA and S&P 500 indices' daily prices. Results showed a relatively high accuracy of trend prediction achieved by the utilized techniques.
- **Huang et al. (2005)** analyzed the predictive ability of Support Vector Machine, Linear Discriminant Analysis, Quadratic Discriminant Analysis and Elman Backpropagation Neural Networks. Models were applied on weekly NIKKEI 225 index data and incorporated several macroeconomic variables as model inputs. A model combining predictions from all of the analyzed techniques brought the best results.

Literature overview

- **Gerlein et al. (2016)** used six ML models including the Naïve Bayes classifier to produce profitable quantitative strategies on the USDJPY, EURUSD and EURGBP currency pairs. 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.
- **Chen et al. (2006)** 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. Results showed that the analyzed models behaved better than benchmark with regard to predicted price deviation measures.
- **Lin et al. (2006)** investigated the performance of decision trees deployed on the 'electronic stocks' of Taiwan stock market and 'technology stocks' of NASDAQ market. Predictions yielded positive returns in case of both indices.

Literature overview

- **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.
- **Colianni et al. (2015)** discussed construction of trading strategies based on qualitative data concerning Bitcoin observed on the Twitter portal. Linear Regression models, Support Vector Machines as well as Bernoulli and Multinomial Naïve Bayes classifiers were used. 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.
- **Kijewski and Ślepaczuk (2020)** proposed a method which combines signals from several strategies to diversify the risk of wrong predictions by a single strategy. They showed that it is possible to double the compounded returns of S&P 500 index with the same level of risk.

Data

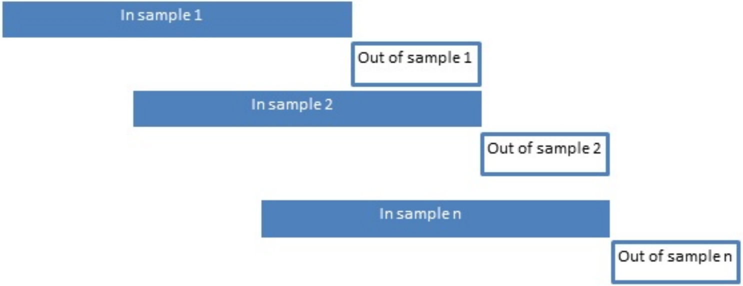
Data. Data description

- Source: http://www.stooq.pl/
- Period: HLC (High Low Close) prices with daily frequency from 2002-01-01 to 2020-10-30
- Stock indices: WIG20 (Poland), DAX (Germany), S&P500 (USA), SOFIX (Bulgaria), PX (Czech Republic), OMXT (Estonia), BUX (Hungary), OMXR (Latvia) and OMXV (Lithuania)
- Limitation: There are dates with no quotes available due to suspension of quotation observed for certain indices as well as differing holiday calendars

Data. Sampling

- Dynamic estimation windows - the underlying parameters of the models were periodically recalibrated to reflect current market behaviors,
- 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, this process is shown on 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.

Data. Sampling

Table 2.1 presents the number of observations for each of the analyzed indices with the corresponding number of in sample and out of sample subsets created as well as the start and the 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.

Data. Initial data analysis

Table 2.2 presents mean, minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum values of the variable for each of the analyzed indices. No outliers or data quality issues were identified.

Table 2.2. Descriptive statistics of all analyzed indices.

Measure / Index	WIG20	DAX	S&P500	SOFIX	PX	OMXT	BUX	OMXR	OMXV
Mean	0.0001	0.0003	0.0003	0.0004	0.0002	0.0005	0.0004	0.0004	0.0005
Minimum	-0.1328	-0.1224	-0.1198	-0.1074	-0.1494	-0.1006	-0.1188	-0.1507	-0.1125
25th percentile	-0.0073	-0.0062	-0.0043	-0.0043	-0.0054	-0.0033	-0.0072	-0.0046	-0.0029
50th percentile	0.0002	0.0008	0.0007	0.0003	0.0006	0.0005	0.0005	0.0002	0.0005
75th percentile	0.0076	0.0072	0.0056	0.0049	0.0066	0.0043	0.0080	0.0054	0.0041
Maximum	0.0850	0.1140	0.1158	0.0875	0.1316	0.1286	0.1408	0.1285	0.1163

Note: Table presents descriptive statistics for returns of all analyzed indices: mean, minimum, 25th percentile, 50th percentile (median), 75th percentile and maximum.

General model formula and target variable

Research methodology. General model formula and target variable

- Supervised machine learning models 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,
- Target (dependent) variable in this research is defined as a discrete return on the asset calculated from the observed Close prices:

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

Technical analysis indicators

Research methodology. Technical analysis indicators-1

Research focuses on strategies based on the set of 5 technical analysis indicators:

- Simple Moving Average (SMA),
- Moving Average Convergence Divergence (MACD),
- Stochastic Oscillator (STOCH),
- Relative Strength Index (RSI),
- Williams' Percent Range (WPR).

Technical indicators were then used as the input to machine learning models.

Simple Moving Average (SMA)

is an average price of an instrument calculated on historical observations up to the reference date. Base level was analyzed for the parameter representing the number of periods equals to 15 (n = 15) while for the purpose of sensitivity analysis, n = {14; 16} were used in order to verify the robustness of the models.

Input used later in the models is derived as a measure of how distant the current price is from its SMA:

$$SMA_{signal} = P_t - SMA \tag{2}$$

Research methodology. Technical analysis indicators-3

Moving Average Convergence Divergence (MACD)

is an indicator 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. 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 with length of the periods set on the level of n = 26 and n = 12 periods. Signal line is defined as EMA with parameter n = 9. Above mentioned parameters are considered base parameters in this research. 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 as a measure of how distant the MACD line is from the signal line:

$$MACD_{signal} = MACDline - SIGNALline \tag{3}$$

Research methodology. Technical analysis indicators-4

Stochastic Oscillator (STOCH)

is incorporating HLC (High Low Close) data into its calculation. STOCH comprises of 3 time series: fast %K, fast %D and slow %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 research. Sensitivity analysis was conducted using $n = \{13; 15\}$ for fast %K and $n = \{2; 4\}$ for fast %D and slow %D.

Relative Strength Index (RSI)

is reflecting current strenght or weekness of the market, base level parameter $n = 14$ was used. In sensitivity analysis, $n = \{13; 15\}$ were used in order to verify the robustness of the models.

Research methodology. Technical analysis indicators-5

Williams' Percent Range (WPR)

is a form of a price oscillator using HLC (High Low Close) data. It is calculated similarly to fast %K in Stochastic Oscillator. Similarly to RSI indicator, base level is analyzed for parameter n = 14 while sensitivity analysis is conducted using n = $\{13; 15\}$.

TA indicators summary

All analyzed TA 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.

Machine learning techniques

Research methodology. Machine learning techniques

Research analyzed eight supervised machine learning models:

- Neural Networks,
- K Nearest Neighbor,
- Random Forest,
- Regression Tree,
- Naïve Bayes,
- Bayesian Generalized Linear Model,
- Support Vector Machines (in two versions: linear and polynomial).

Neural Networks (NN)

- Developed using the Extreme Learning Machine (ELM) approach - a feedforward neural network with one hidden layer making it faster in computation,
- 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 the input layer (14),
- Activation function used was a tansig (tangent-sigmoid transfer function) form producing continuous values in the range from -1 to 1 (intuitive for the return prediction).

K Nearest Neighbor (KNN)

- Regression version,
- Output prediction is the average value of the observed target variable for k nearest neighbors,
- Identification of k nearest neighbors is based on determination of high dimensional Euclidean distance between independent variables of analyzed observations.

Research methodology. Machine learning techniques-3

Random Forest (RF)

- Regression form,
- Modelling framework consisting of random generation of multiple decision trees with each of the trees producing a distinct prediction for target variable,
- Multiple predictions are then averaged to calculate the final output.

Research methodology. Machine learning techniques-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,
- Fundamental algorithms of Regression Tree are similar to those of Random Forest model.

Research methodology. Machine learning techniques-5

Naïve Bayes (NB)

- Probabilistic classifier incorporating the assumption of naïve independence between input variables,
- It produces binary outputs (classes) computed from conditional *a posteriori* probabilities,
- Research used {-1;1} classes representing buy and sell trading signals.

Bayesian Generalized Linear Model (BGLM)

- Generalization of linear regression models which among others allows for target variable transformations via a link function e.g. logit,
- 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.

Research methodology. Machine learning techniques-7

Support Vector Machine Linear (SVML)

- Regression version,
- 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,
- The objective function of the model has to be determined by identification of the optimal hyperplane using the minimization problem.

Support Vector Machine Polynomial (SVMP)

- Version of SVM models incorporating the polynomial kernel function transforming model inputs and computing high-dimensional hyperplanes.

Research methodology. Machine learning techniques-8

Hyperparameters tunning

- Neural Networks - base activation function was tansig (tangent-sigmoid transfer function) and for the purpose of sensitivity analysis, two alternative functions were chosen: sin (sine transfer function) and satlins (symmetric saturating linear transfer function),
- Classification models - 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),
- Other regression models - the hyperparameter chosen for sensitivity analysis was an optimization metric with RMSE as the base metric and alternative metrics being Rsquared and MAE.

Investment strategies construction

Research methodology. Investment strategies construction-1

The general model formula can be extended to present each particular independent variable:

$$f(y) = f(SMA_{signal}) + f(MACD_{signal}) + f(fast\%K) + f(fast\%D) + f(slow\%D) + f(RSI) + f(WPR) + \epsilon \tag{4}$$

- Input independent variables (technical analysis indicators) were rescaled before being fed to the models 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.

Research methodology. Investment strategies construction-2

- Machine learning models used in this research can be divided into two groups:
  - classification models (Naïve Bayes),
  - 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,

Research methodology. Investment strategies construction-3

- Continous 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
- 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,
- 40th quantile and 60th 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,

Research methodology. Investment strategies construction-4

- 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
$$r_t^{strategy} = r_t^{index} * signal_{t-1}^{strategy} \tag{5}$$
- Returns from the strategies were aggregated for every out of sample period in order to compare the strategies among each other and with the buy-and-hold benchmark strategy,
- Buy-and-hold strategy involves buying an instrument at the beginning and selling at the end of the period, so it can be interpreted as an absolute measure of market movements.

Risk and return measures

Research 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:

- Compound annual growth rate,
- Maximum capital drawdown,
- Sharpe Ratio,
- Information Ratio*.

- **Compound Annual Growth Rate (CAGR)** is a measure illustrating how much on average capital has grown in each year of the investment.

$$R = CAGR(t_0, t_n) = \left(\frac{V(t_n)}{V(t_0)}\right)^{\frac{1}{t_n - t_0}} - 1 \tag{6}$$

- **Maximum drawdown (MDD)** represents the maximum decrease in accumulated capital over the entire investment horizon. It allows for investigation whether the portfolio has not recorded significant drops in value, which would indicate its instability. MDD 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 research, the measure is always presented as positive value, so in superior investment strategies the maximum decline should be as low as possible.

$$MDD = -\frac{T_{Min} - P_{Max}}{P_{Max}} \tag{7}$$

Research methodology. Risk and return measures

- **Adjusted Sharpe Ratio (SR)** is a simplified version of SR with risk-free rate equal to zero. 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. When comparing strategies, the better performing one is the one with the higher Adjusted Sharpe Ratio. Measure was floored at 0 as the negative values are often deemed meaningless in the scientific world.

$$SR = \max\left(\frac{R - R_f}{\sigma}; 0\right); R_f = 0 \tag{8}$$

- **Information Ratio* (IR*)** is an adjusted Information Ratio 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.

$$IR^* = \frac{\max(R;0)^2}{\sigma * MDD} \tag{9}$$

There are many measures of risk and return used by researchers to compare the performance of investment strategies, but each of them is suitable for analyzing different types of instruments contained in a portfolio **(Bacon 2010)** or **(Kość et al. 2019)** .

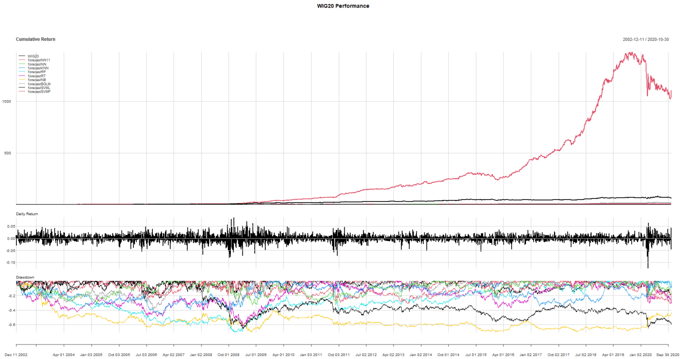
Identification of the best performing strategy

Empirical results. Identification of the best performing strategy

- OOS results for WIG20, DAX, S&P500 and other CEE indices,
- 8 various ML models,
- Trading signals from ML models separately transformed into Buy/Sell signals,
- Figures with equity lines, drawdowns, and daily returns will be analyzed,
- IR* as the main performance metric,
- Buy&Hold strategy as the main benchmark,

Empirical results. WIG20

Figure 4.1. Equity lines, daily returns and drawdown lines for WIG20



Note: Figure shows equity lines, daily returns and drawdown lines for every strategy constructed on WIG20 index in the period from 2002-02-22 to 2020-10-30.

Empirical results. WIG20

Table 4.1. Risk and return measures for WIG20

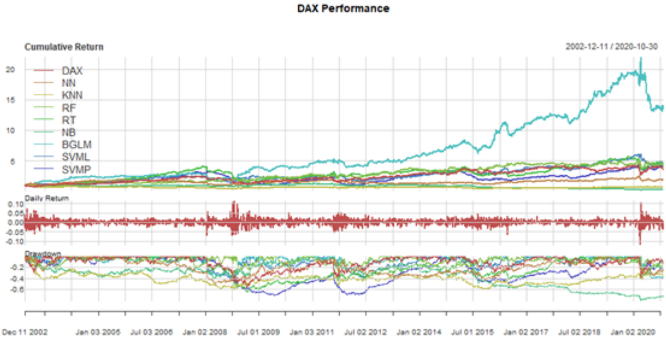
Measure	WIG20	NN	KNN	RF	RT	NB	BGLM	SVML	SVMP
CAGR	1.88%	8.85%	2.61%	3.94%	0.15%	-2.99%	17.02%	26.19%	**47.96%**
Annual. Std Dev	22.67%	20.00%	20.42%	20.42%	22.57%	21.04%	20.26%	**19.92%**	20.26%
Adj Sharpe	0.0830	0.4427	0.1278	0.1931	0.0065	0.0000	0.8398	1.3149	**2.3672**
MDD	65.75%	33.14%	46.62%	70.45%	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.

Support Vector Machine strategies are dominant in case of WIG20 index with its Polynomial version outperforming the rest of the strategies significantly.

Empirical results. DAX

Figure 4.2. Equity lines, daily returns and drawdown lines for DAX



Note: Figure shows equity lines, daily returns and drawdown lines for every strategy constructed on DAX index in the period from 2002-02-22 to 2020-10-30.

Empirical results. DAX

Table 4.2. Risk and return measures for DAX

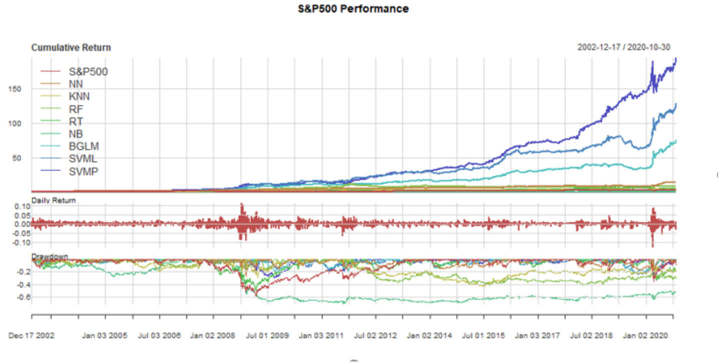
Measure	DAX	NN	KNN	RF	RT	NB	BGLM	SVML	SVMP
CAGR	7.42%	3.66%	-2.01%	8.75%	8.26%	-5.60%	**15.53%**	8.79%	7.88%
Annual. Std Dev	22.04%	19.89%	19.76%	19.82%	21.98%	**19.00%**	19.48%	19.23%	19.78%
Adj Sharpe	0.3368	0.1841	0.0000	0.4416	0.3759	0.0000	**0.7973**	0.4569	0.3986
MDD	55.08%	49.70%	60.40%	**27.72%**	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 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.

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.

Empirical results. S&P500

Figure 4.3. Equity lines, daily returns and drawdown lines for SPX



Note: Figure shows equity lines, daily returns and drawdown lines for every strategy constructed on S&P500 index in the period from 2002-02-22 to 2020-10-30.

Empirical results. S&P500

Table 4.3. Risk and return measures for S&P500

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 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.

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.

Empirical results. CEE indices

Table 4.4. Risk and return 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.

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.

Empirical results. Summary for all indices

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.

Bayesian Generalized Linear Model performed the best for three indices i.e. DAX, PX and OMXR and the same situation applies to Naïve Bayes model which performed the 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.

Empirical results. Summary for all indices. Rank approach

For each index, strategies were ranked from 1 to 9 where 9 constitutes the highest score. For example in case of WIG20, SVMP strategy had the highest IR* measure and received score equal to 9 in IR* category. 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	RT	NB	BGLM	SVML	SVMP
IR*	5.22	4.56	3.39	6.22	3.83	**6.44**	6.28	4.17

Note: Table shows ranked risk and return measure IR* averaged across all analyzed indices. Bolded font indicates the best performance ranked measure for all tested methods.

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.

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.

BGLM strategy performed the best for 4 indices i.e. DAX, PX, BUX and OMXR. SVML model was considered the best for BUX. NB strategy once more performed the best for 3 indices i.e. SOFIX, OMXT and OMXV. SVMP strategy received the best IR* score for 2 indices i.e. WIG20 and S&P500.

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
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 measure IR* averaged across all analyzed indices in the sensitivity analysis scenario with decreased technical indicators' parameters. Bolded font indicates the best performance metrics for all tested methods.

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 BGLM received the most robust results across all analyzed indices.

Empirical results. Sensitivity to technical analysis indicators

Table 4.9. The best models and IR* measures for **increased** TA 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.

BGLM strategy performed the best for 2 indices i.e. DAX and S&P500 which is a downgrade from results obtained in the base and decreased parameter scenarios. NB strategy again performed the best for 3 indices i.e. SOFIX, OMXT and OMXV. SVML strategy received the best IR* score for 2 indices i.e. BUX and OMXR. SVMP was considered the best for WIG20 index as in the base scenario. NN strategy was as the best performer for PX index.

Table 4.10. Ranked performance measures for **increased** TA parameters

Measure	B&H	NN	KNN	RF	RT	NB	BGLM	SVML	SVMP
Adj Sharpe	5.44	5.56	3.11	4.67	5.67	4.28	**6.17**	6.06	4.06
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 measure IR* averaged across all analyzed indices in the sensitivity analysis scenario with increased technical indicators' parameters. Bolded font indicates the best performance metrics for all tested methods.

The best score of 6.17 for IR* averaged rank was obtained by both BGML and SVML strategies. 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.

Empirical results. Sensitivity to ML optimization metrics. Sensitivity of Neural Networks

In case of NN models, hyperparameter altered in the sensitivity analysis was the activation function which was investigated in three different forms (scenarios) i.e. tansig (base function used in this research), 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.

Empirical results. Sensitivity to ML optimization metrics. Sensitivity of classification models

Naïve Bayes model is the only classification model described in this research. 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 research) 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

Activation Function	Naive Bayes
Accuaracy	**1.5**
Kappa	**1.5**

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.

Empirical results. Sensitivity to ML optimization metrics. 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 research), Rsquared (coefficient of determination) and MAE (mean absolute error), results of which are presented in 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 BGLM and SVML, models obtained the same averaged rank (2.00) for every metric which means that they are resistant to optimization metric alteration. SVMP 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.

Conclusions

RH1: *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.*

–> cannot be rejected

RH2: *Neural Networks generate the best (in terms of risk adjusted returns) investment signals compared to other machine learning techniques used in the research.*

–> is rejected because BGLM, NB, SVML, and SVMP produced the best results.

RH3 *The very same machine learning strategy is considered the best performing for all analyzed stock market indices.*

–> is rejected. BGLM, NB, SVML, and SVMP were the best for various indices separately.

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.*

–> is rejected. On a strategy level, results changed in each analyzed scenario for most of the analyzed models.
–> On average however, the BGLM generated the best results in all sensitivity analysis scenarios in which the TA indicators' parameters were altered. In case of altering the ML models' hyperparameters, the BGLM model was considered to be resistant to changes.

Further research

- broader set of ML models,
- a larger number of instruments from various set of asset classes will be tested,
- more detailed sensitivity analysis of the best models,
- creation of ensemble models built for various ML models within the given assets and various assets within the given ML models and all of them together,
- analysis on higher frequencies, at least hourly, if possible 1-minute,

Thank you for your attention!

Jan Grudniewicz, jangrudniewicz@gmail.com

Quantitative Finance Research Group

and

Robert Ślepaczuk, rslepaczuk@wne.uw.edu.pl,

Faculty of Economic Sciences,

University of Warsaw,

Department of Quantitative Finance

Quantitative Finance Research Group