Stock Market Models using the Trend Using Support Vector Machines

Ivana Markovic[1], Jelena Stankovic[2], Miloš Stojanovic[3], Miloš Bozic[4]
[1,2] University of Nis
Trg Kralja Aleksandra Ujedinitelja 11
Niš, Serbia
{ivana.markovic@eknfak.ni.ac.rs} {jelenas@eknfak.ni.ac.rs}

[3,4] Aleksandra Medvedeva 20
18000 Niš, Serbia
{milosstojanovic10380@yahoo.com} {milos1bozic@gmail.com}

ABSTRACT: *For generating model for stock markets, we have basically deployed the Least Squares Support Vector Machines (LS-SVMs) through which the classification is made. It also supports the trend prediction. We used the web based technical indications for the feature selection. We have conducted the experimentation and the results indicate that the suggested model is suitable for short-term prediction of changes in the stock market trend index.*

Keywords: Stock Market Trend Prediction, The Least Squares Support Vector Machines (LS-SVMS), Classification

Received: 3 October 2020, Revised 19 Janaury 2021, Accepted 27 February 2021

DOI: 10.6025/jio/2021/11/2/41-47

1. Introduction

Mining the stock market tendency is a challenging task, taking into consideration the fact that the financial market is a complex, evolving and dynamic system whose behavior is pronouncedly non-linear [1].

Predicting the direction of the movement of the price of financial instruments is a current area in academic research where the algorithms for machine learning have proven to be quite effective. In [2] it was indicated that the Least Squares Support Vector Machines (LS-SVMs), and SVMs - Support Vector Machines outperform other machine learning methods, since in theory they do not require any previous a priori assumptions regarding data properties. Moreover, they guarantee the global optimal solution. Although some researchers have suggested that there is evidence that stock prices are not purely random, the general consensus still is that their behavior is approximately close to the random walk process. Degrees of accuracy of an approximate 60% hit rate in predictions are often considered satisfactory results for stock market trend prediction. [3]

The value of the Belex15 index determines the price of the most liquid stocks which are traded on the regulated market of the Belgrade Stock Exchange. The index is not itself a trading commodity, its role is to measure the changes in the price of the stock which is being traded using the continuous trading method and which previously satisfied the criteria for participating in the index basket.

According to [4] one of the most widely adopted economic methods for trend prediction applied at world stock markets is a technical analysis. Considering the fact that the basic assumption of the technical analysis is that the market discounts all the relevant information, the aim of the technical analysis is to use an analysis of price movement and the volume of trading in securities to create a basis for the prediction of future price changes of financial instruments. The key role in the prediction of the price trend is played by technical indicators, which can be divided into the following groups: trending indicators, volume indicators and oscillators. Trending indicators identify and monitor the securities trends, while the Volume indicators are based on the change in the volume of trading in securities and complete the information which is offered by the trending indicators in forming trading strategies. Oscillating indicators or oscillators are the leading indicators which generate early warning signals of changes in the securities trend and determine the strength of the current trend, as well as the moment when a change in the trend occurs.

In this study the trending indicators and oscillators will be used as prediction features in forming the LS-SVM model for predicting the value trend of the Belex15 index. The study included only those predictive features for which strong evidence of a causal relationship with return series could be determined. Furthermore, the study also included the statistically determined lagged returns as input features. The problem of predicting the direction of the change in value of the stock index is then modeled as a problem of a binary classification. The proposed model represents an improvement of the model outlined in [5].

The rest of the study is organized as follows: the second part of the paper presents the theoretical basis of the LS–SVM method of binary classification. The third section describes the proposed prediction model, while the results of the testing are shown in section four. In section five some of the conclusions and ideas for further research are presented.

2. The Basic Theory of Least Squares Support Vector Machines for Binary Classification

The Least Squares Support Vector Machines, proposed by Suykens in [6], includes a set of linear equations which are solved instead of a Quadratic Programming (QP) for classical SVMs. Therefore, LS-SVMs are more time-efficient than standard SVMs.

Let's study a training group of a total of N examples $T = \{x_i, y_i\}_{i=1}^{N}$. In the learning phase, the model is formed based on the known training data (x_1, y_1), (x_2, y_2), …, (x_N, y_N), where x_i are the input vectors, and y_i are the labels of binary classes that were assigned to them. Each input vector consists of numeric features, while $y_i \in \{-1, +1\}$.

According to [6] LS-SVMs for binary classification were defined as follows:

$$\min_{w,b,e} J_{LS}(w,b,e) = \tfrac{1}{2} w^T w + C \tfrac{1}{2} \sum_{k=1}^{N} e_k^2 \qquad (1)$$

with the equality conditions:

$$y_k [w^T \varphi(x_k) + b] = 1 - e_k, \quad k = 1,..., N \qquad (2)$$

where φ is a non-linear function that maps input vectors in some higher dimensional feature space. The weight vector of the hyper plane is marked by a w, while b is the scalar shift, that is, weight threshold. The variable e_k represents the allowed errors of classification, while the parameter C controls the process, that is, the relationship between the complexity of the model and the accepted error of classification. After solving the optimization problem defined by (1) and (2), a solution can be found in [6], the function of the separation of LS-SVM classifications is defined as:

$$y(x) = sign\left[\sum_{k=1}^{N} \alpha_k y_k K(x, x_k) + b\right] \qquad (3)$$

where α_k represent the support vectors (Lagrange multipliers), and b is a constant. $K(x, x_k)$ represents the Kernel function, which is defined by the scalar product between x and x_k. In this study, the RBF kernel was used, defined by:

$$K(x, x_k) = e^{\frac{-\|x-x_k\|^2}{\sigma^2}} \quad (4)$$

When training the LS-SVM model it is necessary to determine the value of parameter C, as well as the parameters of the selected kernel, in this case the width σ. One of the ways to determine these parameters is the k fold Cross – Validation procedure in combination with a Grid – Search. More about parameter selection techniques can be found in [7].

3. Feature Selection

The selection of input features is crucial for defining an accurate prediction model. An arbitrary application of a large number of explanatory features to LS-SVMs could lead to low prediction accuracy. Carefully monitoring data on the other hand would

Indicators	**Formula**
Closing price	CP_t, $t = 1, 2, \ldots N$
Lowest price	LP_N - Lowest price in the past N days
Highest price	HP_N - Highest price in the past Ndays
Logarithmic return	$r_t = logCP_t - logCP_{t-1}$
Trend indicators	
Moving Average	$MA = (1/N) * \sum_{i=0}^{t} CP_t$
Exponential Moving Average	$EMA_N = r_t * k + EMA_{t-1} * (1-k)$; k=2/(N+1)
Oscillating indicators	
Relative Strength Index	$RSI=100-(100/(1+\frac{\frac{1}{T}\sum_{t=o}^{T} CP_t^+}{\frac{1}{T}\sum_{t=o}^{T} CP_t^-}))$
Relative Volatility Index	$RVI = 100 * \frac{\frac{1}{T}\sum_{t=o}^{T} StDev(CP_{t-9})^+}{\frac{1}{T}\sum_{t=o}^{T} StDev(CP_{t-9})^+ + \frac{1}{T}\sum_{t=o}^{T} StDev(CP_{t-9})^-}$
Moving Average Convergence	$MACDt = EMA_{12} - EMA_{26}$ Signal Line = Simple 9-day moving average of MACD
Momentum	$MoM_t = CP_t / CP_{t-n}$,
Rate of Change	$ROCt = ((CP_t - CP_{t-n})/CP_{t-n}) * 100$,
Stochastic Oscillator	$\%K = 100*((C - LP_{14})/(HP_{14} - LP_{14}))$ $\%D$ = average of the last three $\%K$

Table 1. The List of Technical Indicators

lead to higher method accuracy. This process is of great importance but there is no general rule that can be followed. Table 1 shows selected list of the technical indicator based on their frequent use in the literature [1-4].

Different technical indicators are analyzed, and the most suitable ones are chosen for the model. First, because the response variable predict the stock market trend (either an increase or decrease), the explanatory features need to measure changes as well. In effect, observing feature changes over time is more significant for prediction than the absolute value of each feature.

Second, it is necessary to evaluate the level of importance of each individual indicator. Feature evaluation is used to measure the relative importance and weights of stock indicators. The sensitivity analysis is the process which determines whether an input variable influences the output of the method or not. There are several ways to determine this, but basically the input variable can be omitted if there are no noticeable changes between running the model with and without it.

Trend Indicators. On the basis of the aforementioned criteria, from the group of trend indicators, the EMA was selected. It assigns greater significance to the more recent changes in prices and enables the calculation of an almost infinite number of steps (for example EMA150, EMA250) which additionally recommends it for modeling time series. The EMA can smooth the price data and filters out the noise. In [5] it was shown that the analysis of the value trend of the stock index and securities over shorter intervals (5 to 25 days) results in indicators which appropriately measure the sensitivity of the change in value, and thus the selected period for calculating the EMA transformation was the previous 10 days.

Oscillating Indicators. In the group of technical indicators of oscillations which are usually used to create predictive models according to [8] MACD and RVI, give better results in optimizing the investment strategies on emerging markets. The analysis of sensitivity has confirmed that the MACD has a greater significance, and it will thus be used in this study. As was shown in Table I, MACD is obtained through a combination of three movement averages. The standard combination of movement averages which determine the MACD is 12-26-9, where the first line is obtained as the difference between the 12-period EMA and 26-period EMA, and the other line, which represents the signal line, approximates a 9-period EMA of the first line. Figure 1 shows the modeling of the time series using the MACD indicators. It can clearly be seen that the abovementioned transformation contributes to the stationarity of the series, which additionally increases the effectiveness of the algorithm of machine learning.

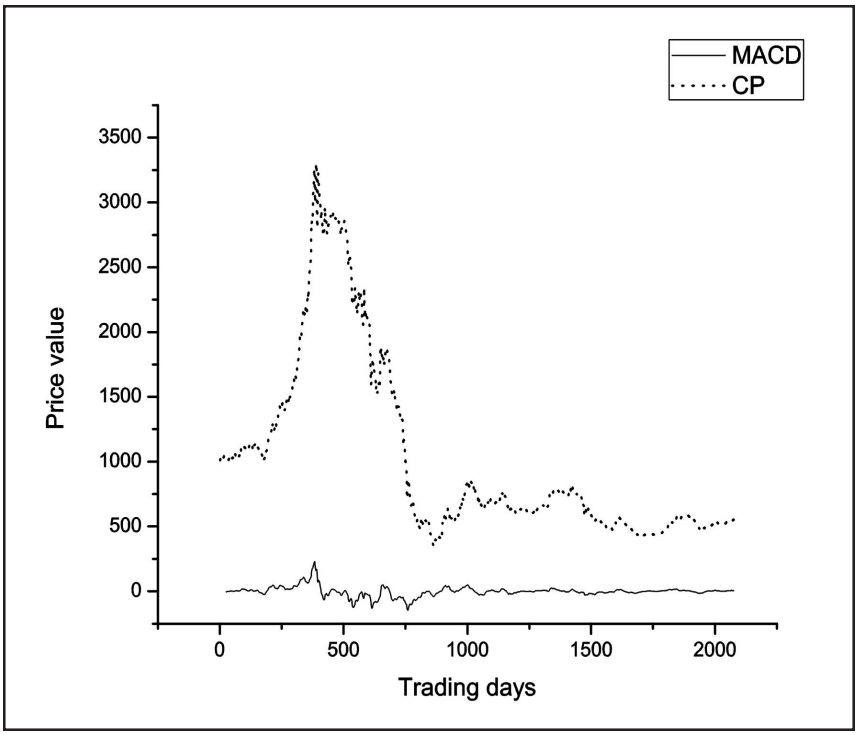

Figure 1. The relation between the MACD and the stock price

Time Series Characteristics. Taking into consideration the time series recency effect in building time series models, that is, that using data in time closer to the forecasted data produces better models, a statistical analysis of the autocorrelations was carried out on logarithmic returns so as to determine which variables have the strongest influence on the prediction of trends of change in the value of the index. The obtained values of the autocorrelation coefficients indicate that the values of the autocorrelation function decrease significantly during the first three steps and that the first and second previous value of the logarithmic return prevail. That is why the values of the logarithmic returns were added for the first and second previous value.

Trend Modeling: The variable to be predicted are the future trends of the stock market. The attribute which serves as a label for the class is a categorical variable used to indicate the movement direction of the Belex15 index over time *t*. If the logarithmic return over time *t* is larger than zero, the indicator is 1. Otherwise, the indicator -1 The continuously compounded rate of return (r_t) is computed as shown in Table 1. Figure 2 shows the changes in trend in the period between January and March 2013.

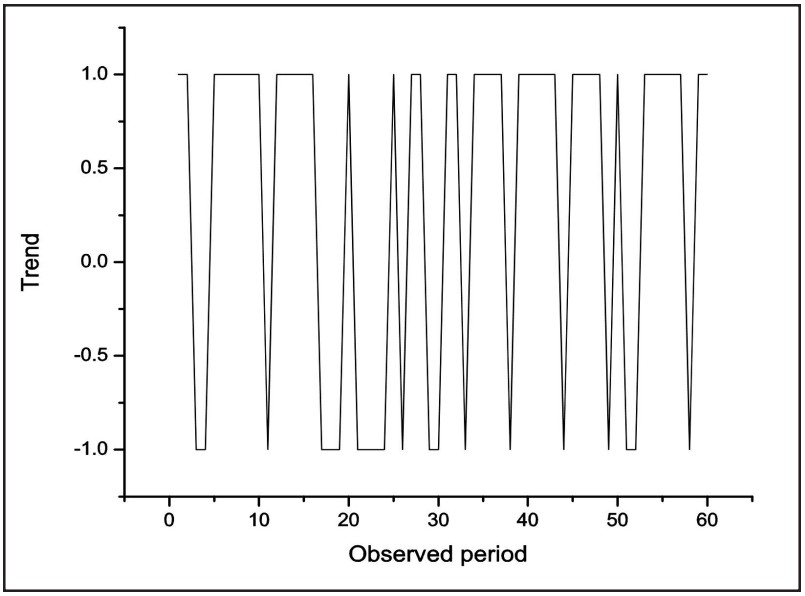

Figure 2. Trend fluctuations

It can be determined that in reality the market price trend does not constantly follow a straight line; it is volatile, and the line fluctuates up and down repeatedly. On the basis of pervious analysis the following prediction model was created:

$$y_t = LS - SVM(r_{t-1}, r_{t-2}, EMA_{10t-1}, MACD_{t-1}) \quad (5)$$

The model was formed based on four input features and the original data are scaled using min-max normalization:

$$f_i(\text{norm.}) = \frac{f_i - \min F}{\max F - \min F} \quad (6)$$

where *min F* and *max F* are the minimum and maximum values of the feature *F* and f_i is the variable in the i^{th} row. In order to form the model, LS-SVMlab [9] was used.

4. Experimental Results

The study relied on data taken from the website of the Belgrade Stock Exchange (www.belex.rs). The available data were divided into two groups. The first group consisted of 1811 records required for the training model, from October 26, 2005 to December 31, 2012. For the second group of data, from January 3, 2013 to October 1, 2013, a total of 187 days of trading were selected.

As a general measure for the evaluation of the prediction effect it is used the Hit Ratio (HR) which was calculated on the basis of the number of properly classified results within the test group:

$$HR = \frac{1}{m}\sum_{i=1}^{m} PO_i \text{ for } PO_i = \begin{cases} 1 & PV_i = AV_i \\ 0 & PV_i \neq AV_i \end{cases} \quad (7)$$

Where PO is the prediction output of the i trading day, AV_i is the actual value for the i training day and PI_i is the predicted value for the i trading day and m is the number of data in the test group [10].

Table 2 represents the hit-rate of the proposed model based on temporal sequences which correspond the real frameworks of trading on the Belgrade stock exchange, the weekly, biweekly, monthly and quarter work regime The table also shows the results obtained using the random walk model (RW) as a benchmark. The RW uses the current value to predict the future value, assuming that the latter in the following period ($y_t + 1$) will be equal to the current value (y_t).

Time	Sequence	RW LS-SVM
0-5	0.6000	0.6000
0-10	0.8000	0.8000
0-20	0.7000	0.7000
0-40	0.6000	0.6500
0-60	0.6000	0.6500
0-80	0.6125	0.6375
0-100	0.6100	0.6600
0-120	0.6083	0.6750
0-140	0.5714	0.6500
0-160	0.5438	0.6063
0-180	0.5389	0.6000
0-187	0.5348	0.5829

Table 2. A Comparison of the Models

It can be noted that in the first trading month, the rate of the hits is identical for both models, which is in favor of the previously noted strong correlation in the available data series. The longer the time period, the more dominant the prediction based on LS-SVMs.

The hit rate of the proposed predictor at the level of the entire set of the group is 0.5828 and represents an increase of approximately 2% in comparison to the model proposed in [5] which did not include the selection of technical indicators of oscillations. The number of days follows an increasing trend in the training set is 886, while in the decreasing trend it is 939. The ratio of increasing/ decreasing trading days in the training and test data set is approximately the same. At the level of the test group the hit rate in records with an increasing change in trend is 0.591 (58/98), while the hit rate in the days with a decreasing change in the trend is 0.573 (51/87). The calculated hit rates in the borderline cases remain within the range of the previously defined values of the predictors and indicate their stable characteristics.

The obtained values of the hit rates are within the expected range of precision and are comparable to the results obtained in other studies [1], [3], [10].

The computation speed for model training and the time needed to obtain the predictions is approximately 100 seconds, thus

the model is able to provide a response to the dynamic changes in the stock market movements.

5. Conclusion

Most studies in this field deal with the prediction of market indices and the price of financial instruments on developed markets. It is important to note that the studied prediction rate of the price index in this study belongs to the emerging market of the Republic of Serbia, and that it led to competitive results.

In the remainder of the paper we could define several directions of study which could lead to an improvement in the model precision. The first could be to adjust the model parameters by means of a more sensitive and comprehensive parameter setting. In the second, based on technical indicators, further studies could offer an improvement through the introduction of macroeconomic indicators, including foreign exchange rates. In addition, there is the possibility of designing a hybrid prediction model, where the outputs from more models would be combined in a final model.

Since the prediction of the movement of stock market indices plays an important role in the development of effective market trading strategies, it is important to point out that every increase in precision is considered an exceptional contribution since it leads to an increase in the return and the decrease in the risk involved in trading. Finally, it would be beneficial to test the proposed model for profitability using the currently available tools and trading strategies in economic sciences.

References

[1] Huang, W., Nakamori, Y., Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine, *Computers & Operations Research,* 32 (10) 2513– 2522.

[2] Phichhang, O., Wang, H. (2009). Prediction of Stock Market Index Movement by Ten Data Mining Techniques, *Modern Applied Science*, 3 (12) 28-42.

[3] Lahmiri, S. (2011). A Comparison of PNN and SVM for Stock Market Trend Prediction using Economic and Technical Information, *International Journal of Computer Applications*, 29 (3).

[4] Chsherbakov, V. (2010). Efficiency of Use of Technical Analysis: Evidences from Russian Stock Market, *Ekonomika a management*, volume 4.

[5] Marković, I., Stanković, J., Stojanović, M., Božić, M. Predviđanje promene trenda vrednosti berzanskog indeksa Belex15 pomoću LS-SVM klasifikatora, simpozijum INFOTEH Jahorina 2014.

[6] Suykens, J., Vandewalle, J. (1999). Least Squares Support Vector Machines, Neural processing letters, 9 (3) 293-300.

[7] Božić, M. Stajić, Z., Stojanović, M., Kratkoročno predviđanje električnog opterećenja primenom metoda podržavajućih vektora, Infoteh Jahorina, 10, 326-329, 2010.

[8] Eric, D., Andjelic, G., Redzepagic, S. (2009). Application of MACD and RVI indicators as functions of investment strategy optimization on the financial market, *In*: Proceedings of the Faculty of Economics of Rijeka, 27 (1) 171-196.

[9] De Brabanter, K., Karsmakers, P., Ojeda, F., Alzate, C., De Brabanter, J., Pelckmans, K., De Moor, B., Vandewalle, J., Suykens, J. A. K. (2011). LS-SVMlab Toolbox User's Guide, ESAT-SISTA Technical Report 10-146, 2011, http://www.esat.kuleuven.be/sista/lssvmlab/

[10] Yuling, L., Guo, H., Hu, J. (2013). An SVM-based Approach for Stock Market Trend Prediction, *Neural Networks (IJCNN), IEEE Press*, 1-7.