

DOI: 10.37943/AITU.2021.47.22.009

V. Serbin

Associate professor of Department of Information Systems
v_serbin@mail.ru, orcid.org/0000-0002-5807-3873
International Information Technology University, Kazakhstan

U. Zhenisserov

Master of Science in "Management of IT-projects"
zheniserov@gmail.com, orcid.org/0000-0003-2948-1189
International Information Technology University, Kazakhstan

ANALYSIS OF MACHINE LEARNING METHODS FOR PREDICTIONS OF STOCK EXCHANGE SHARE PRICES

Abstract: Since the stock market is one of the most important areas for investors, stock market price trend prediction is still a hot subject for researchers in both financial and technical fields. Lately, a lot of work has been analyzed and done in the field of machine learning algorithms for analyzing price patterns and predicting stock prices and index changes. Currently, machine-learning methods are receiving a lot of attention for predicting prices in financial markets. The main goal of current research is to improve and develop a system for predicting future prices in financial markets with higher accuracy using machine-learning methods. Precise predicting stock market returns is a very difficult task due to the volatile and non-linear nature of financial stock markets. With the advent of artificial intelligence and machine learning, forecasting methods have become more effective at predicting stock prices. In this article, we looked at the machine learning techniques that have been used to trade stocks to predict price changes before an actual rise or fall in the stock price occurs. In particular, the article discusses in detail the use of support vector machines, linear regression, and prediction using decision stumps, classification using the nearest neighbor algorithm, and the advantages and disadvantages of each method. The paper introduces parameters and variables that can be used to recognize stock price patterns that might be useful in future stock forecasting, and how the boost can be combined with other learning algorithms to improve the accuracy of such forecasting systems.

Keywords: Neural networks, stock market, price prediction, machine learning.

Introduction

Formulation of the problem. Everyone wants to be rich in their life without making a lot of effort and having great advantages. Likewise, we want to look into our future with the innermost desire, since we do not want to take risks or want to reduce the risk factor. The stock market is a place where buying and selling can secure future life goals. [12] Now the question is, how can we benefit from the stock market? Alternatively, what are the steps that can give us predictions in the stock markets before you get into the risk zone [13]. How can artificial intelligence with machine learning algorithms help predict future market trends.

By predicting the trend of the stock market, we can avoid wasting money. Stock market forecasting is the process of predicting the future based on past data. Forecasting reduces the level of risk for investors and increases the level of confidence for investments. If they predicted goals before they reached them, they could avoid losing money. All of these considerations

work like forecasting the stock market. Based on trends in historical data, we propose a future trend called the stock market forecast.

According to an international Microsoft survey, 94% of CEOs believe that artificial intelligence technologies are important for the strategic goals of their organizations. At the same time, 27% of respondents have already implemented the relevant technologies in key business processes; another 46% are conducting pilot projects.

Stock prices are considered highly dynamic and subject to rapid changes due to the underlying nature of the financial industry and in part due to a combination of known parameters and external unknown factors.

An intelligent trader predicts the price of a stock and buys a stock before the price rises, or sell it before its value declines. Although it is very difficult to replace the experience gained by an experienced trader, an accurate forecasting algorithm can directly lead to high returns for investment companies, which indicates a direct relationship between the accuracy of the forecasting algorithm and the profit obtained from its use [8]. In practice, there are 2 methodologies for forecasting reserves:

Fundamental Analysis: This method, performed by basic fundamental analysts, is more about the company than the actual supplies. analysts accept conclusions based on past performance of the company, profit forecast, etc.

Technical analysis. This method, produced by technical analysts, deals with determining the value of stocks based on past stock models (with the introduction of time series analysis) [11].

In the case of using machine-learning methods for processing trade data, the most often used method is technical analysis - the task is to understand whether the algorithm can accurately determine the patterns of stock behavior over time. Nevertheless, machine learning can also be used to assess and predict the results of a company for further use in fundamental analysis. In the end, the most effective method for automated prediction of stock prices and generation of investment recommendations is a hybrid approach that combines the approaches of fundamental and technical analysis [3,4].

In early research related to stock market prediction, Fama, E.F. (1970) proposed the efficient market hypothesis (EMH), and Horn, J.S. and Parker, J.G. (1967) proposed the theory of a random walk. These theories assumed that market prices were influenced by information other than historical prices, and therefore the market price could not be predicted.

The EMH theory assumes that the price of a share is completely dependent on market information and therefore any new information will cause the price to change in response to recently released information. This theory also argued that stocks always trade at their fair value when traders cannot buy or sell stocks at a special price that is undervalued or inflated, and therefore the only way a trader can increase their profits is to increase their risk. The EMH discusses three different options that affect the market price: the Weak Form, where only historical data is considered, the Semi-Strong Form, which includes current public data in addition to historical data, and the Strong form, which goes further to include private data. The EMH states that any price movement is either the result of new published information or random movement that impedes the success of forecasting models [1,8,9].

The random walk hypothesis of Horn, J.S. and Parker, J.J. (1967) argues that stock prices move at random and argue that past price movements are independent of current movements. This differs slightly from EMH as it focuses on the short-term structure of the stock market [4].

The purpose and objectives of the study. In this dissertation, we aim to evaluate machine-learning methods for long-term forecasting of reserves based on fundamental analysis. We do this by comparing the prediction performance of several advanced machine-learning techniques based on fundamental analysis using fundamental functions.

Indicator functions. We will now take a quick look at the attributes and indicators that are commonly used in technical analysis of stock prices.

Indicators can be any of the following:

Moving Average (MA) – displays the average of n past values up to the current moment. This algorithm is one of the most important algorithms in use today. From financial time series and signal processing to neural networks, it is widely used. Any data in sequence. The algorithm is mainly used to reduce noise in noisy time series data. The term we use for this is data smoothing. We do this by weighting the number of observations and their mean.

Exponential Moving Average (EMA) – gives more weight to the most recent values, but does not completely discard the old values. An exponential moving average is also called an exponentially weighted moving average. An exponentially weighted moving average is more responsive to recent price changes than a simple moving average that applies the same weight to all observations over a period.

Rate of Change (RoC) is one of the simplest technical indicators calculated as the ratio or difference between the current price and the price n periods ago.

The Relative Strength Index (RSI) is a momentum indicator used in technical analysis that measures the magnitude of recent price changes to assess overbought or oversold conditions in the price of a stock or other asset. The RSI is displayed as an oscillator (a line chart that moves between two extreme points) and can range from 0 to 100. The indicator was originally developed by J. Wells Wilder Jr. and presented in his seminal 1978 book *New Concepts in technical trading systems* [5].

For the described project, EMA was chosen as the main indicator - it allows you to process an almost unlimited amount of historical data, which is very important for analysis using time series. However, it should be noted that the use of other indicators can bring greater accuracy of forecasts of the analyzed stocks [2,7].

$$EMA(t)=EMA(t-1)+\alpha*(Price(t)-EMA(t-1)) \quad (1)$$

where, $\alpha=2/(N+1)$, so for $N=9$, $\alpha=0.20$

In theory, the problem of predicting the price of a stock can be considered as evaluating the function F over time T based on the previous values of F at times $t-1, t-2 \dots t-n$, assigning the corresponding weighting function w to at each moment F :

$$F(t)=w_1*F(t-1)+w_2*F(t-2)+\dots+w_n*F(t-n) \quad (2)$$

Learning environment. During the project, Weka and YALE data mining environments were used. The configuration looked like this:

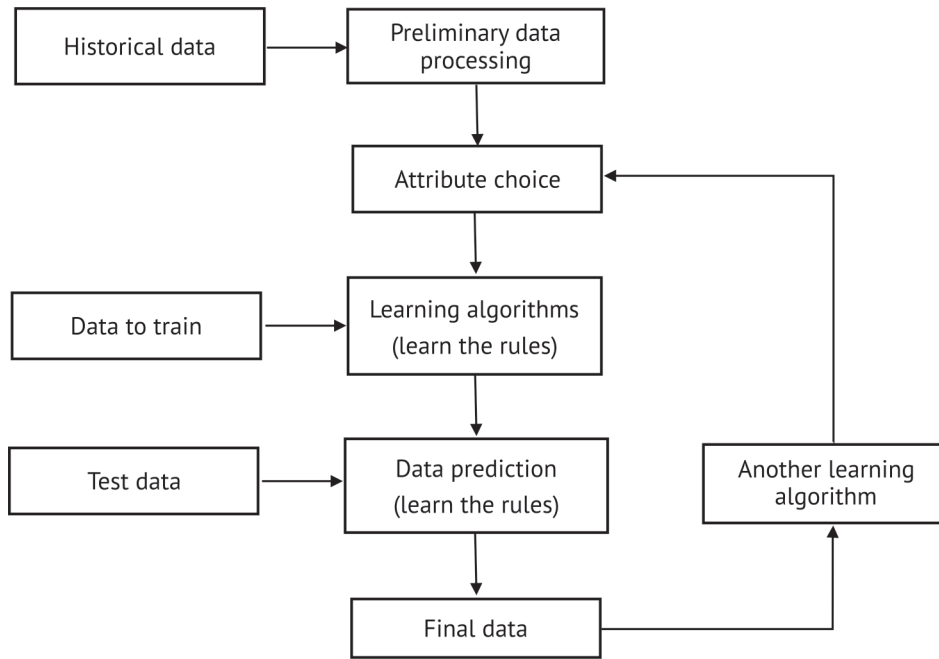


Fig. 1. Algorithm configuration for data mining.

Since the system for predicting the prices of stocks of stock exchanges decides to choose an algorithm for the method with input data on template business processes, import and use of BPMN processes comes in handy, instead of simulating them independently.

It is also worth noting that in order to allow business analysts to make the necessary edits to the processes “on the fly”. This platform is needed on its own due to the fact that it belongs to the Low-code type, which ensures the minimum participation of the IT department in the creation and launch of work of business processes [6].

The general structure of the business process model is as follows:

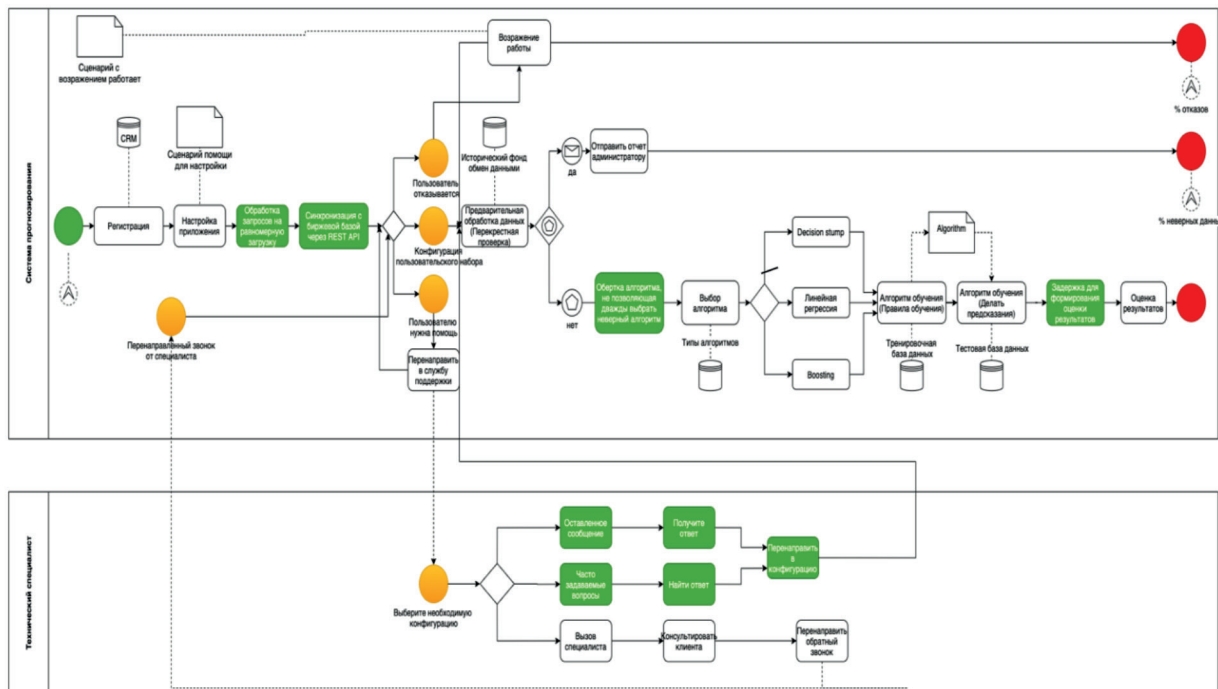


Fig. 2. TO-BE model

These BPMN processes illustrate the important benefits of solutions that can be created for a stock price prediction system. In addition, such solutions are beneficial in that they require minimal costs at the implementation stage and make it possible to implement them gradually, expanding the scope of application from one area (for example, CRM) to the entire work of the enterprise. This allows you to avoid shocks during a sharp change in software, gradually train the staff and, most importantly, first assess the real advantages of the platform, and then deepen the use of solutions created on its basis [7,10].

Study results.

This model deals with three different algorithms Decision stump, Linear regression and Boosting. Below are the results of applying machine learning algorithms.

Decision Stump Algorithm

The application of a simple algorithm for predicting EMA allowed us to achieve the following results:

- Correlation coefficient 0.8597
- Average absolute error 46.665
- Root mean square error 57.8192
- Relative absolute error 46.8704%
- Root mean square relative error 50.9763%
- Total number of periods 681

Linear regression

Applying simple linear regression to predict the EMA yielded the following results:

- Correlation coefficient 0.9591
- Average absolute error 12.9115
- Root mean square error 32.0499
- Relative absolute error 12.9684%
- Root root mean square relative error 28.2568%
- Total number of periods 681

Linear regression

Applying simple linear regression to predict the EMA yielded the following results:

- Correlation coefficient 0.9591
- Average absolute error 12.9115
- Root mean square error 32.0499
- Relative absolute error 12.9684%
- Root root mean square relative error 28.2568%
- Total number of periods 681

Boosting

After using the C-SVC algorithm, the AdaBoostM1 boosting algorithm was applied to the dataset – this resulted in a significant improvement in accuracy.

Root mean square error: 0.467 ± 0.008

Accuracy: $64.32\% \pm 3.99\%$

Conclusion

From all the algorithms we have applied, we have seen that only support vector machines combined with acceleration and boosting give us satisfactory results. Linear regression gave lower mean square errors when predicting the EMA pattern.

Selecting the indicator function can significantly improve / reduce the accuracy of the forecasting system. In addition, a particular machine learning algorithm may be better suited for a certain type of stock, say, Technology Stocks, while the same algorithm may be less accurate when predicting some other types of stocks, such as Energy Stocks.

In addition, we should also note that when applying machine-learning algorithms for technical analysis, we assumed that the influence of unknown factors (election results, rumors, political effects, etc.) was already built into the historical model of stocks. Commercial trading systems may have a more sophisticated mechanism for accounting for unknowns. Although we studied the algorithms discretely, most often the hybrid algorithm is used to predict reserves.

References

1. Weng, B., Ahmed, M.A., & Megahed, F.M. (2017). Stock market one-day ahead movement prediction using disparate data sources. *Expert Systems with Applications*, 79, 153-163. <https://www.sciencedirect.com/science/article/abs/pii/S0957417417301331>
2. Li, X., Huang, X., Deng, X., & Zhu, S. (2014). Enhancing quantitative intra-day stock return prediction by integrating both market news and stock prices information. *Neurocomputing*, 142, 228–238. <https://www.sciencedirect.com/science/article/abs/pii/S092523121400647X>
3. Arévalo, A., Niño, J., Hernández, G., & Sandoval, J. (2016, August). High-frequency trading strategy based on deep neural networks. In *International conference on intelligent computing* (pp. 424-436). Springer, Cham. https://www.researchgate.net/publication/305214717_High_Frequency_Trading_Strategy_Based_on_Deep_Neural_Networks.
4. Machine learning techniques for predicting price indicators: indicator functions and news analysis. <http://elitetrader.ru/index.php?newsid=279203>
5. Van Horne, J. C., & Parker, G. G. (1967). The random-walk theory: an empirical test. *Financial Analysts Journal*, 23(6), 87-92. https://econpapers.repec.org/article/asiaeafjr/2014_3ap_3a1840-1848.htm
6. Schumaker, R. P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. *Information Processing & Management*, 45(5), 571-583. <https://www.sciencedirect.com/science/article/abs/pii/S0306457309000478>.
7. Ding, X., Zhang, Y., Liu, T., & Duan, J. (2014, October). Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1415-1425). https://www.researchgate.net/publication/301405008_Using_Structured_Events_to_Predict_Stock_Price_Movement_An_Empirical_Investigation.
8. Shah, V.H. (2007). Machine learning techniques for stock prediction. *Foundations of Machine Learning* Spring, 1(1), 6-12. <https://evilinside.ru/texniki-mashinnogo-obucheniya-dlya-prognozirovaniya-cen-akcij-funkcii-indikatorov-i-analiz-novostej/>
9. Bollen, J., & Mao, H. (2011). Twitter mood as a stock market predictor. *IEEE Annals of the History of Computing*, 44(10), 91-94. <https://www.sciencedirect.com/science/article/abs/pii/S187775031100007X>
10. Hellström, T. (2001). *Techniques and software for development and evaluation of trading strategies*. Univ. <http://www.cs.umu.se/~thomash/reports/phdthesis.pdf>
11. Pahwa, K., & Agarwal, N. (2019, February). Stock market analysis using supervised machine learning. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 197-200). IEEE. <https://ieeexplore.ieee.org/document/8862225>
12. Zhang, K., Zhong, G., Dong, J., Wang, S., & Wang, Y. (2019). Stock market prediction based on generative adversarial network. *Procedia computer science*, 147, 400-406. <https://www.sciencedirect.com/science/article/pii/S1877050919302789>
13. Song, Y. G., Cao, Q. L., & Zhang, C. (2018). Towards a new approach to predict business performance using machine learning. *Cognitive Systems Research*, 52, 1004-1012. <https://www.sciencedirect.com/science/article/abs/pii/S1389041718304376>

Главный редактор:
Белощицкий А.А.
Ответственный редактор:
Амиргалиев Б.Е.

Подписано в печать 30.03.2021 г.
Формат 60x84 1/8. Усл. п.л 6,25.
Тираж 300 экз. Заказ № _____.
Отпечатано в ТОО «Шаңырақ-Медиа».
г. Нур-Султан, ул. Кокарал, 2/1, тел. 87077770066.
www.smedia.kz