



STOCK MARKET PREDICTION USING LINEAR REGRESSION ALGORITHM BY COMPARING WITH SUPPORT VECTOR MACHINE ALGORITHM TO IMPROVE ACCURACY

Myneni. Keerthi Pranav¹, S. Ashok Kumar^{2*}

Article History: Received: 12.12.2022

Revised: 29.01.2023

Accepted: 15.03.2023

Abstract

Aim: The aim is to improve accuracy and develop a Linear Regression model for prediction of stock market prices for better investments and huge profits using a novel method.

Materials and Methods: Supervised Machine learning techniques such as Linear Regression Algorithm is being compared with Support Vector machine algorithm for predicting Stock market values. Sample size is determined by using G power calculator and found to be 20 per group. G power is predicted to be 80%. Total of 2 groups are used. Statistical analysis is done on SPSS Software. Significance value is observed to be 1.000.

Results: Based on the analysis Linear Regression Algorithm method has an accuracy of 85% and Support vector machine has 77% and the significance value achieved is 1.000 ($p > 0.05$). It shows that two groups are statistically insignificant.

Conclusion: It is concluded that based on the execution analysis Linear regression shows the better accuracy when compared to the Support vector machine algorithm.

Keywords: Linear Regression, Support Vector Machine, Stock Market, Prediction, Data Mining, Accuracy.

¹Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India, Pincode: 602105

^{2*}Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Tamil Nadu, India, Pincode: 602105

1. Introduction

Prediction of stock values is one of the difficult tasks as it mostly deals with the financial data of a company as well as the financial market. At present, across the world, a major part of the amount is invested in stock markets as it deals with the huge profits (Mehta, Pandya, and Kotecha 2021). Dataset plays a major role in predicting the stock prices for the future period of time. Stock market prediction is used for making the investors to feel better and invest in good stocks which brings better results (Mehta, Pandya, and Kotecha 2021; Gupta et al. 2019). Stock market predictions are used to forecast the data for the future predictions which make the investors understand about the right investments using the novel method (Albahli et al. 2022). Data mining techniques and machine learning techniques are used in determining the predicted stock values along with accurate values using the novel methods (Dong 2019).

Stock market predictions using linear regression are implemented for prediction to benefit stock investors, nearly 95 in IEEE explorer and 180 in Google scholar. The articles that are mostly cited based on the most viewed are one of them depending on the study of stock market predictions. Various companies have their main role as predicting the future stock prices to make investors get clarified through analysis (Dong 2019; Kecman 2001). This analysis is done using classification and regression algorithms. This research paper states that analyzing stock prices is a difficult task as this deals with the financial data and financial markets (Hatami et al. 2022). The prediction of financial data is a dangerous task as it deals with the economy of a person and a country. Predictions of stock market prices are to be done accurately as it deals with the financial data as well as to bring out huge profits (Valencia, Gómez-Espinosa, and Valdés-Aguirre 2019). Data mining techniques and machine learning algorithms are mainly used in forecasting the outcomes in prediction based problems. These are used to bring the outcome efficiently (Ecer et al. 2020). Our team has extensive knowledge and research experience that has translated into high quality publications (Pandiyani et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022).

The limitations of stock investors is to get the values accurately and accurately predicting the stock prices using the novel method. In this research it is observed that Linear Regression has brought the outcome of more accuracy than

Support vector machine algorithm (Pyo et al. 2017). The main aim of study is to increase the performance based on the accuracy with significance value using the innovative Linear regression and Support vector machine algorithm in an innovative prediction of stock market values.

2. Materials and Methods

The research work is carried out in DBMS Laboratory, Department of Computer Science Engineering, Saveetha School of Engineering. In this study 2 sample groups were taken. Group 1 was the Linear Regression model and group 2 was the Support vector machine algorithm. Sample size is calculated using G-power software considering the pretest power to be 80% and CI of 80%. The work has been carried out with 3500 records which were taken from a kaggle data set. The accuracy in predicting loyalties was initiated by two different groups (Hatami et al. 2022; Prime 2020). Totally 10 iterations were analyzed and performed on each group to accomplish maximum accuracy. Dataset contains 3000 instances and 7 attributes named Date, High, Low, Close, Open, Volume, OpenInt. Here data is from kaggle website (Kaggle: Your Machine Learning and Data Science Community).

Linear regression model is one of the novel methods in the regression model in Data mining techniques used for stock market price prediction. This regression model mostly needs the dataset as the prerequisite for price prediction. This linear regression is classified into two types based on the explanatory variables, such as simple linear regression and multiple linear regression (R. Ho 2017). Linear regression model uses two different types such as Independent variables and Dependent variables named as x and y respectively and the equation is represented as $Y = \beta_0 + \beta_1 X$. Support vector machine is one of the classification algorithms and supervised machine learning algorithms. This SVM is classified mainly into two types such as Linear SVM and Non-Linear SVM. Hyper planes are used for determining the prices using the SVM'S. These hyper planes are supported by vectors hence it is known as support vectors. (Nabipour et al. 2020)

The tool used to execute the program is Google colab and the databases are directly imported when the commands are instructed in the command prompt. Algorithm is implemented in the python code and the accuracy is obtained based on the dataset.

Linear Regression(LR) Algorithm

Linear regression is a data mining model which shows the best relation between the independent and dependent variables for the prediction of Stock

market values. Dependent variables named as X and Y respectively and the equation is represented as $Y = \beta_0 + \beta_1 X$. Linear regression model is the best model for finding out the best relation between variables and forecasting.

Algorithm Steps

Step 1: First initialize the parameters
Step 2: Predict the value of a dependent variable by giving an independent variable
Step 3: Get the accuracy

Pseudo Code

Input: Training Data

Output: Accuracy

Begin

- For all the existing data
- Read and Load the data set.
- Extract Data features of a sound file imported.
- Add the technical indicators.
- Initialize the LR classifier.
- Train the classifier
- Predict the test set

Return the Accuracy

End for

Support Vector Machine(SVM) Algorithm

Support vector machine is a supervised machine learning algorithm that can be used for both classification and regression challenges in data mining. Hyper plane plays a major role in determining the predicted values of the stock markets. This SVM can be implemented through using the programming languages such as Python and R

Pseudo Code

Input: Training Data

Output: Accuracy

Begin

- For all the existing data
- Read and Load the data set.
- Extract Data features of a sound file imported.
- Calculate the mean and standard deviation of the predictor variables in each class.
- Initialize the SVM classifier.
- Train the classifier
- Predict the test set

Return the Accuracy

End for

Statistical Analysis

The analysis was done using IBM SPSS software. Independent sample t test is carried out for analysis. Independent variables are Date, High, Low, Close, Open, Volume, OpenInt and dependent variable is accuracy (R. Ho 2017).

3. Results

The Support vector machine algorithm shows that databases play a major role in determining the accurate predictions. Table 1 represents the difference in predicting the stock prices using Linear regression method and Support vector machine method.

In Table 2, the results achieved with $p=1.000$ ($p>0.05$) shows that two groups are statistically insignificant. Table 2 represents the group statistics analysis which include the total no. of Values, mean, standard deviation, standard error mean. Linear regression method and SVM algorithm has an accuracy of 85.60% and 77.40% respectively. Standard error mean for SVM method is 1.035 less than Linear regression method. It represents an independent sample test which provides a significance of 0.64 for both assumed and non assumed values.

Figure 1 represents the graph that explains the comparison of the accuracy value with algorithm Linear regression method and SVM where the accuracy of Linear regression method is 85% and the accuracy value of the SVM is 77%.

4. Discussion

The data evolution was performed using IBM SPSS version 21. To analyze the data, Independent sample T-test and group statistics can be carried out. In this research study it is proved that the linear regression method has got the more accurate values than the Support vector Machine algorithm..

Linear regression is a method which has lower time complexity when compared to other regression and classification methods (Ecer et al. 2020). This method takes less time for compiling to get the accurate values (Tsihrantzis and Jain 2020). Linear regression performs exceptionally well for linearly separable data. This method is easier to implement, interpret and efficiently train. Whereas in the SVM model choosing a good kernel function is not an easy task as it consumes most time (Bosco and Khan 2018). The SVM model works slow when it works on large datasets. The accuracy of output depends on the quality of the data. SVM is difficult to understand its variable weights and individual impact (Pyo et al. 2017).

The limitations of the investors are the accurate stock price prediction and less efficiency. The main aim of the study is to provide accurate stock price prediction using linear regression method (Nabipour et al. 2020). There are various applications and novel methods where linear regression can be used

such as predicting weather, forecasting results, academic reports, etc. (T.-T. Ho and Huang 2021).

5. Conclusion

In the proposed model, the accuracy percentage of predicting the stock market prices using linear regression model is 85%, whereas in the comparison model the Support vector machine algorithm has got the accuracy of 77% only.

DECLARATION

Conflict of Interests

No conflict of interest in this manuscript.

Authors Contribution

Author M.KP was involved in data collection, data analysis, and manuscript writing. Author S. AK was involved in the action process. Data verification and validation and Criteria review of manuscript.

Acknowledgement

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.

1. Oracle Tech Solutions
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

6. References

- Albahli, Saleh, Tahira Nazir, Awais Mehmood, Aun Irtaza, Ali Alkhalifah, and Waleed Albattah. 2022. "AEI-DNET: A Novel DenseNet Model with an Autoencoder for the Stock Market Predictions Using Stock Technical Indicators." *Electronics*. <https://doi.org/10.3390/electronics11040611>.
- Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." *Chemosphere*. <https://doi.org/10.1016/j.chemosphere.2022.134596>.
- Bosco, Joish, and Fateh Khan. 2018. *Stock Market Prediction and Efficiency Analysis Using Recurrent Neural Network*. GRIN Verlag.
- Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." *Bioresource Technology*. <https://doi.org/10.1016/j.biortech.2022.127234>.
- Dong, Zhengyang. 2019. "Dynamic Advisor-Based Ensemble (dynABE): Case Study in Stock Trend Prediction of Critical Metal Companies." *PloS One* 14 (2): e0212487.
- Ecer, Fatih, Sina Ardabili, Shahab S. Band, and Amir Mosavi. 2020. "Training Multilayer Perceptron with Genetic Algorithms and Particle Swarm Optimization for Modeling Stock Price Index Prediction." *Entropy* 22 (11). <https://doi.org/10.3390/e22111239>.
- Gupta, Deepak, Mahardhika Pratama, Zhenyuan Ma, Jun Li, and Mukesh Prasad. 2019. "Financial Time Series Forecasting Using Twin Support Vector Regression." *PloS One* 14 (3): e0211402.
- Hatami, Zahra, Hesham Ali, David Volkman, and Prasad Chetti. 2022. "A New Approach for Analyzing Financial Markets Using Correlation Networks and Population Analysis." *Proceedings of the 24th International Conference on Enterprise Information Systems*. <https://doi.org/10.5220/0011073800003179>.
- Ho, Robert. 2017. "Hypothesis Testing: T Test for Independent and Correlated Groups." *Understanding Statistics for the Social Sciences with IBM SPSS*. <https://doi.org/10.4324/9781315182452-13>.
- Ho, Trang-Thi, and Yennun Huang. 2021. "Stock Price Movement Prediction Using Sentiment Analysis and CandleStick Chart Representation." *Sensors* 21 (23). <https://doi.org/10.3390/s21237957>.
- Karpagam, M., R. Beulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." *Soft Computing*. <https://doi.org/10.1007/s00500-022-06931-1>.
- Kecman, Vojislav. 2001. *Learning and Soft Computing: Support Vector Machines, Neural Networks, and Fuzzy Logic Models*. MIT Press.
- Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time

- Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation.” *Journal of Thermal Analysis and Calorimetry*. <https://doi.org/10.1007/s10973-022-11298-4>.
- Mehta, Pooja, Sharnil Pandya, and Ketan Kotecha. 2021. “Harvesting Social Media Sentiment Analysis to Enhance Stock Market Prediction Using Deep Learning.” *PeerJ. Computer Science* 7 (April): e476.
- Nabipour, Mojtaba, Pooyan Nayyeri, Hamed Jabani, Shahab, and Amir Mosavi. 2020. “Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms via Continuous and Binary Data: A Comparative Analysis.” *IEEE Access: Practical Innovations, Open Solutions* 8: 150199–212.
- Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. “Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm.” *Computational Intelligence and Neuroscience* 2022 (April): 6461690.
- Nagaraju, V., B. R. Tapas Bapu, P. Bhuvanewari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. “Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands.” *Silicon*. <https://doi.org/10.1007/s12633-022-01825-1>.
- Pandiyani, P., R. Sitharhan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. “A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies.” *Sustainable Energy Technologies and Assessments*. <https://doi.org/10.1016/j.seta.2022.102155>.
- Prime, Sunantha. 2020. “Forecasting the Changes in Daily Stock Prices in Shanghai Stock Exchange Using Neural Network and Ordinary Least Squares Regression.” *Investment Management and Financial Innovations*. [https://doi.org/10.21511/imfi.17\(3\).2020.22](https://doi.org/10.21511/imfi.17(3).2020.22).
- Pyo, Sujin, Jaewook Lee, Mincheol Cha, and Huisu Jang. 2017. “Predictability of Machine Learning Techniques to Forecast the Trends of Market Index Prices: Hypothesis Testing for the Korean Stock Markets.” *PloS One* 12 (11): e0188107.
- Tsihrintzis, George A., and Lakhmi C. Jain. 2020. *Machine Learning Paradigms: Advances in Deep Learning-Based Technological Applications*. Springer Nature.
- Valencia, Franco, Alfonso Gómez-Espinosa, and Benjamín Valdés-Aguirre. 2019. “Price Movement Prediction of Cryptocurrencies Using Sentiment Analysis and Machine Learning.” *Entropy* 21 (6). <https://doi.org/10.3390/e21060589>.
- Venu, Harish, Ibhama Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. “Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends.” *Energy*. <https://doi.org/10.1016/j.energy.2022.123806>.
- Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. “Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture.” *Chemosphere* 301 (August): 134638.
- Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. “Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review.” *Chemosphere* 299 (July): 134390.

TABLES AND FIGURES

Table 1. Group Statistics results (Mean of Linear Regression is 85 which is greater when compared to SVM Algorithm is 77 and Standard error mean for Linear Regression is 1.024 and SVM algorithm is 1.035)

	Algorithms	N	Mean	Std. Deviation	Std. Error Mean
Accuracy	Linear Regression	20	85	3.239	1.024

	SVM Algorithm	20	77	3.273	1.035
--	---------------	----	----	-------	-------

Table 2. The results achieved with $p=1.000$ ($p>0.05$) shows that two groups are statistically insignificant. The below table shows the results of independent variables of two algorithms with the comparison of accuracy and T-test for Equality of Mean.

		Levene's test for equality of variables		T-test for Equality of Mean						
		F	Sig	t	df	Sig(2-tailed)	Mean difference	Std. Error difference	95% confidence interval of the difference	
									Lower	Upper
Accuracy	Equal variance assumed	.00	1.00	5.632	18	0.000	8.200	1.456	5.141	11.259
	Equal variances not assumed			5.632	17.998	0.000	8.200	1.456	5.141	11.259

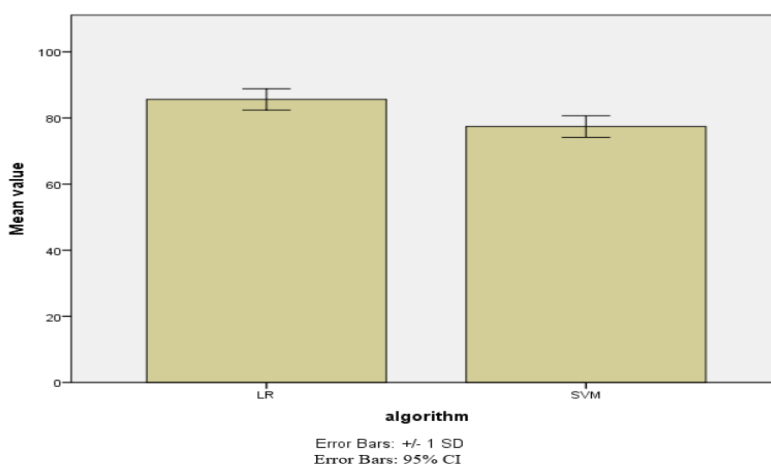


Fig. 1. Bar Chart representing the comparison of Mean Accuracy of Linear Regression and SVM Algorithm. Mean accuracy of Linear Regression is 85% appears to be better than SVM Algorithm which is 77%. The X-axis represents Linear Regression and SVM Algorithm and the Y-axis represents the mean accuracy ± 1 SD.