

Stock Market Prediction Using Machine Learning Techniques

Manjunath S H¹, Pradhosh K², Praveen Kumar S³, Kanimozhiveena E⁴

^{1,2,3} Dept of Computer Science and Engineering ⁴Assistant Professor, Dept
of Computer Science and Engineering

^{1,2,3,4} Adhiyamaan College of Engineering, Hosur, India.

Abstract- In Stock Market Prediction, the aim is to predict the future value of the financial stocks of a company. Stock market is totally uncertain as the prices of stocks keep fluctuating on a daily basis because of numerous factors that influence it. The recent trend in stock market prediction technologies is the use of Machine Learning Algorithms. These have been used to devise new techniques to build prediction models that can forecast the prices of stock and predict about the market trend with good accuracy. It makes predictions based on the values of current stock market indices by training on their previous values. Machine learning itself employs different models to make prediction easier and authentic. The paper focuses on the use of Super Vector Regressor and Logistic Regression based Machine learning to predict stock values. Factors considered are open, close, low, high. Also, the paper calculates the closing amount price for each of the algorithms used.

Keywords- Super Vector Regressor, Logistic Regression, Stock Market Prediction, Mean-Squared Error

I. INTRODUCTION

In the last few decades forecasting of stock return has become an important field of research. In most of the cases, the researchers had attempted to establish a linear relationship between the input macroeconomic variables and the stock returns. After the discovery of nonlinearity in the stock market index returns, many literatures have come up in a nonlinear statistical modelling of the stock returns, most of them required that the nonlinear model be specified before the estimation is done. But since stock market return is noisy, uncertain, chaotic and nonlinear in nature, ANN has evolved out to be better technique in capturing the structural relationship between a stock's performance and its determinant factors more accurately than many other statistical techniques.

In this project, different sets of input variables are used to predict stock's closing price. In fact, different input variables are to predict the same set of stock return data. It makes use of the Machine Learning Algorithms such as Linear Regression,

Support Vector Regression, Long-Short Term Memory to predict the desired closing price of a company's stock.

II. LITERATURE REVIEW

For Predicting the stock price trend by interpreting the seemingly chaotic market data has always been an attractive topic to both investors and researchers. Among those popular methods that have been employed, Machine Learning techniques are very popular due to the capacity of identifying stock trend from massive amounts of data that capture the underlying stock price dynamics. In this project, we applied supervised learning methods to stock price trend forecasting. According to market efficiency theory, US stock market is semi-strong efficient market, which means all public information is calculated into a stock's current share price, meaning that neither fundamental nor technical analysis can be used to achieve superior gains in a short-term (a day or a week). Indeed, our initial next-day predication has very low accuracy around 50%. However, as we tried to predict long-term stock price trend, our models achieved a high accuracy (79%). Based on our prediction result, we built a trading strategy on the stock, which significantly outran the stock performance itself.

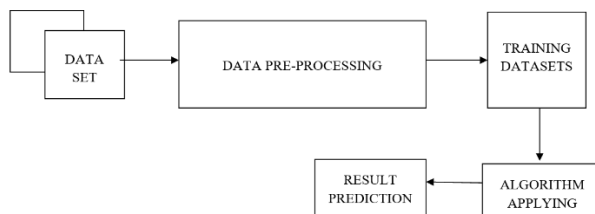
Market manipulation remains the biggest concern of investors in today's securities market, despite fast and strict responses from regulators and exchanges to market participants that pursue such practices. The existing methods in the industry for detecting fraudulent activities in securities market rely heavily on a set of rules based on expert knowledge. The securities market has deviated from its traditional form due to new technologies and changing investment strategies in the past few years. The current securities market demands scalable machine learning algorithms supporting identification of market manipulation activities. In this paper we use supervised learning algorithms to identify suspicious transactions in relation to market manipulation in stock market. We use a case study of manipulated stocks during 2003. We adopt CART, conditional inference trees, C5.0, Random Forest, Naïve Bayes, Neural Networks, SVM and KNN for classification of manipulated samples. Empirical results show that Naïve Bayes

outperform other learning methods achieving F2 measure of 53% (sensitivity and specificity are 89% and 83% respectively).

Market manipulation remains the biggest concern of investors in today's securities market, despite fast and strict responses from regulators and exchanges to market participants that pursue such practices. The existing methods in the industry for detecting fraudulent activities in securities market rely heavily on a set of rules based on expert knowledge. The securities market has deviated from its traditional form due to new technologies and changing investment strategies in the past few years. The current securities market demands scalable machine learning algorithms supporting identification of market manipulation activities.

III. METHODOLOGY

Initially, a dataset is collected for the company to be monitored. After the cleaning and pre-processing, the dataset is used for both training and testing. The algorithms are applied to get the desired result.



A. DATA COLLECTION:

It's time for a data analyst to pick up the baton and lead the way to machine learning implementation. The job of a data analyst is to find ways and sources of collecting relevant and comprehensive data, interpreting it, and analysing results with the help of statistical techniques.

B. DATA PRE-PROCESSING:

The purpose of pre-processing is to convert raw data into a form that fits machine learning. Structured and clean data allows a data scientist to get more precise results from an applied machine learning model. The technique includes data formatting, cleaning, and sampling.

C. DATA SPLITTING:

A dataset used for machine learning should be partitioned into three subsets — training, test, and validation sets.

Training set: - A data scientist uses a training set to train a model and define its optimal parameters — parameters it has to learn from data.

Test set: - A test set is needed for an evaluation of the trained model and its capability for generalization. The latter means a model's ability to identify patterns in new unseen data after having been trained over a training data. It's crucial to use different subsets for training and testing to avoid model over fitting, which is the incapacity for generalization we mentioned above.

Validation set: - A validation set is a set of data used to train artificial intelligence (AI) with the goal of finding and optimizing the best model to solve a given problem. Validation sets are also known as dev sets.

D. MODEL TRAINING:

After a data scientist has pre-processed the collected data and split it into three subsets, he or she can proceed with a model training. This process entails “feeding” the algorithm with training data. An algorithm will process data and output a model that is able to find a target value (attribute) in new data — an answer you want to get with predictive analysis. The purpose of model training is to develop a model. Two model training styles are most common — supervised and unsupervised learning. The choice of each style depends on whether you must forecast specific attributes or group data objects by similarities.

Applying Machine Learning:

Support Vector Regressor: Support Vector Regression as the name suggests is a regression algorithm that supports both linear and non-linear regressions. This method works on the principle of the Support Vector Machine. SVR differs from SVM in the way that SVM is a classifier that is used for predicting discrete categorical labels while SVR is a regressor that is used for predicting continuous ordered variables.

Logistic Regression: Logistic regression is one of the most popular Machine Learning algorithms, which comes under the Supervised Learning technique. It is used for predicting the categorical dependent variable using a given set of independent variables.

E. DATA TESTING:

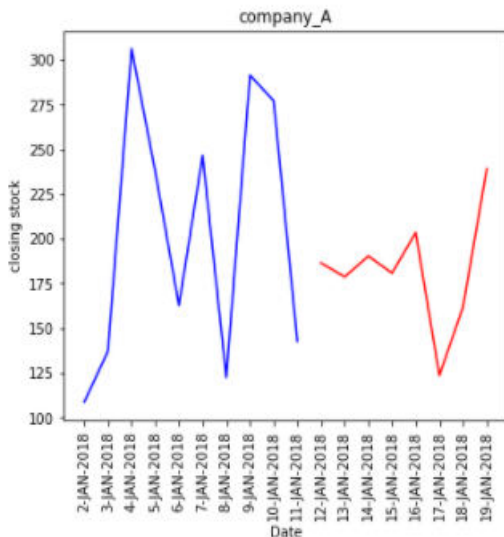
The goal of this step is to develop the simplest model able to formulate a target value fast and well enough. A data

scientist can achieve this goal through model tuning. That's the optimization of model parameters to achieve an algorithm's best performance.

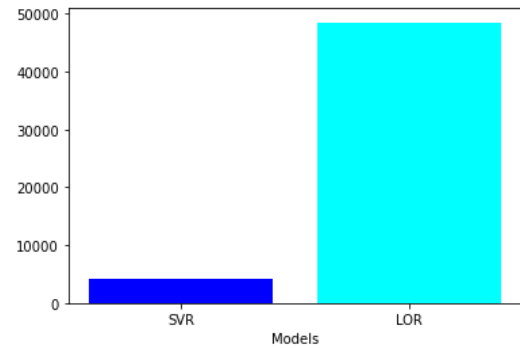
IV. CONCLUSION

Many researchers have tried to devise prediction models using machine learning algorithms to predict the accurate prices of stocks using various tools and techniques, but have yet not been able to come up with the best possible solution. The paper overall summarizes a few of the machine learning techniques that have been used by the research scholars to forecast the stock market trend and prices using machine learning and artificial intelligence algorithms, keeping in mind the extensive detailing, features and parameters involved. Even though after analyzing the major affecting factors and incorporating the social reviews related to stock, the accurate prediction of stock price is not possible. There are some techniques that have been able to get a really close approximation. The major techniques which have been reviewed in this paper are Support Vector Regression (SVR) and Logistic Regression (LOR). It is possible to utilize any of the techniques and develop a hybrid system model for the stock market price prediction, but on the other hand, it is important to design the system in which the accuracy and performance can be increased with less computational complexity.

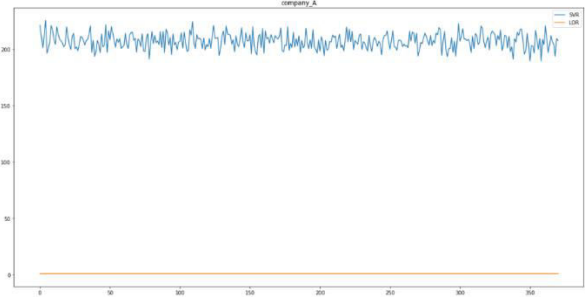
V. RESULTS



Text(0, 0.5, '')



Text(0.5, 1.6, 'company_A')



REFERENCES

- [1] Aparna Nayak, M. M. ManoharaPai and Radhika M. Pai, "Prediction Models for Indian Stock Market", Procedia Computer Science, vol. 89, 2016.
- [2] D. Yan, G. Zhou, X. Zhao, Y. Tian and F. Yang, "Predicting stock using microblog moods", China Communications, vol. 13, no. 8, pp. 244-257, Aug. 2016.
- [3] Jigar Patel, Sahil Shah, Priyank Thakkar and K Kotecha, "Predicting stock market index using fusion of machine learning techniques", Expert Systems with Applications, vol. 42, no. 4, 2015.
- [4] Jitendra Kumar and Ashutosh Kumar Singh, "Workload prediction in cloud using artificial neural network and adaptive differential evolution", Future Generation Computer Systems, vol. 81, pp. 41-52, 2018, ISSN 0167-739X.
- [5] L. Chen, Z. Qiao, M. Wang, C. Wang, R. Du and H. E. Stanley, "Which Artificial Intelligence Algorithm Better Predicts the Chinese Stock Market?", IEEE Access, vol. 6, pp. 48625-48633, 2018.
- [6] Michel Ballings, Dirk Van den Poel, Nathalie Hespeels and Ruben Gryp, "Evaluating multiple classifiers for stock price direction prediction", Expert Systems with Applications, vol. 42, no. 20, 2015.
- [7] N. Sharma and A. Juneja, "Combining of random forest estimates using LSboost for stock market index

prediction", 2017 2nd International Conference for Convergence in Technology (I2CT), pp. 1199-1202, 2017.

- [8] Q. LI, L. JIANG, P. LI and H. CHEN, "Tensor-Based Learning for Predicting Stock Movements", AAAI Conference on Artificial Intelligence, feb. 2015.
- [9] Xi Zhang, Jiawei Shi, Di Wang and Binxing Fang, "Exploiting investors social network for stock prediction in China's market", Journal of Computational Science, vol. 28, no. 2018.