



STOCK PRICE PREDICTION USING MACHINE LEARNING

G. Santhiya¹, B. Karthika², A.S. Balaji³

Student, B.E. Computer Science and Engineering, Anand Institute of Higher Technology, Chennai, India^{1,2}

Assistant Professor, CSE, Anand Institute of Higher Technology, Chennai, India³

Abstract: Machine studying has large programs in the finance industry. danger Analytics, consumer Analytics, and inventory marketplace Predictions are a number of the domains where gadget mastering techniques can be implemented. accurate prediction of inventory marketplace returns is extremely tough because of volatility inside the marketplace. the main factor in predicting a inventory market is a excessive stage of accuracy and precision. With the creation of artificial intelligence and excessive computational potential, efficiency has increased. in the past few a long time, the surprisingly theoretical and speculative nature of the inventory market has been tested by capturing and the use of repetitive patterns. numerous gadget mastering algorithms like more than one Linear Regression, ARIMA model, Random forest set of rules, and so forth. are used here. The financial records incorporates elements like Date, volume, Open, high, Low close, and near prices. The fashions are evaluated the use of fashionable strategic signs LSTM and R2 score. lower values of those two indicators mean better efficiency of the educated fashions. various agencies hire distinct types of analysis gear for forecasting and the primary goal is the accuracy to achieve the most earnings. The a success prediction of the stock may be a useful asset for the inventory marketplace institutions and could provide real-existence solutions to the issues of the traders.

I. INTRODUCTION

Predicting stock price trend and its movement has been viewed as a standout amongst the most difficult utilizations of time arrangement expectation. Despite the fact that there has been numerous research which manage the issues of anticipating stock price trend, most exact discoveries are related with the developed financial markets. But it is difficult to predict the trend or price of the stock because of the uncertainty in the stock market. There are two types of analysis, Fundamental analysis and technical analysis.

In fundamental analysis, performance of the company, economic factors and political factors are considered. In technical factors, previous n days closing price, highest price, lowest price etc. are considered. We can predict the trend of stock or price of the stock using technical analysis. Fundamental analysis is hard to measure and hard to implement in computer language. Technical analysis does not measure the intrinsic security value of the stock, but it uses technical stock charts to predict the trend of the stock.

In initial stage of the study of the stock market prediction, classical methods were used. But as stock market is a non-stationary time series of data. It was not so effective. So non-linear machine learning techniques like Artificial neural networks (ANN) and Support Vector Machine(SVM) are used widely. In this project we have used both the techniques to predict the trend of the stock and measured the accuracy of both the techniques.

1.1 OBJECTIVE

The main objective of implementation of this project is to predict the stock market current condition with reference of last day's stock market value. It is beneficial to the use who can predict the value of the any company share value using this algorithm so it will guess that the financial condition of that company whether the share of company is satisfactory or not how it goes in increasing range or going to the decreasing.

So it is forecast the stock exchange. In market so many algorithm are present for stock market prediction but the SVM provides the stock value with efficiently within a less amount of time.



1.2 SCOPE

Stock market investment is challenging tasks. There is higher probability of losing money but also gain lots of money from it. So there should be one model which gives the gist of trends in stock market. So using machine learning techniques, we can predict the trend of the stock market and invest in it according to that.

II. ANALYSIS

2.1 SYSTEM ANALYSIS

System Analysis is a combined process dissection the system responsibilities that are based on problem domain characteristics and user requirement.

2.1.1 Problem Definition

The purpose of this thesis is to analyse if social media analysis can be used to predict a company's stock price. The following problem to be investigated is therefore can social media analysis be used solely to predict a company's stock price? We expect that social media has a strong impact on a company's stock price. However, we are unsure if the impact is strong enough to be used exclusively for stock market prediction.

2.1.2 Existing System

Time series forecasting consists of a research area designed to solve various problems, mainly in the financial area □ Support vector regression (SVR), a variant of the SVM, is typically used to solve nonlinear regression problems by constructing the input-output mapping function. □ The least squares support vector regression (LSSVR) algorithm is a further development of SVR and its use considerably reduces computational complexity and increases efficiency compared to standard SVR. □ The Firefly Algorithm (FA), which is a nature-inspired metaheuristic method, has recently performed extremely well in solving various optimization problems.

DISADVANTAGES

- The existing system focuses on the stock price market in Taiwan, but does not generalize for other markets worldwide.
- The system does not allow the import of raw data directly □ The existing system cannot be used to analyze multivariate time series □.
- Lastly, the system does not have a user-interface which can be distributed as a web app to users for personal use.

2.1.3 Proposed system

To generalize the application of the existing system, our work uses the system to estimate other stocks in similar emerging markets and mature markets. The system can be extended to analyze multivariate time series data and import raw dataset directly . Profit can be maximized even when the corporate stock market is has lower value . The development of a web- based application has been considered to improve the user-friendliness and usability of the expert system.

MODULES

- 4.1 Stock Price Data Collection
- 4.2 Data Pre-processing
- 4.3 Training
- 4.4 Stock Price Predictions

4.1 STOCK PRICE DATA COLLECTION

For the task, S&P 500 shares from exceptional industries are decided on. more than one factors are taken into consideration whilst picking the stocks, consisting of stock fee volatility, absolutely the significance of the charge, the respective industries, organization size, and so forth., and stocks displaying extraordinary traits are picked.

4.2 DATA PRE-PROCESSING

3 Python scripts are written to transform the raw stock prices (.csv files) into feature vectors, for training, predicting and testing respectively. The scripts take the input options and the raw stock prices as inputs and produce the correct features by building the lookback arrays and the moving averages. It concatenates the features into the final feature



vectors, which will be passed to the model for training or testing. The 3 scripts share common operations in building a dataset except the output size and the range of dates to build from, so common functions are written to centralize the logic instead of repeating the same index-calculation-intensive work across functions.

4.3 TRAINING

In training, a randomized initial model is first generated from the model options definition. A training set is generated by the build training dataset script, which generates the training set features from the input options and the raw stock price data. Then, the data is fed into the model for training.

4.4 STOCK PRICE PREDICTION

To predict the next-day stock price and a long-term stock price. The accuracy ratio is defined as the number of days that the model correctly classified the testing data over the total number of testing days. With the short term model predicting the next day stock price, it has very low accuracy, the Quadratic Discriminant Analysis is the best among all models, it scored a 58.2% accuracy. With the long term model predicting the next n days stock prices, the longer the time frame, the better in the accuracy for SVM.

When all of the 16 features were used, the accuracy of the model reached 79%, while it fell to 64% when only 8 features were used, and 55% if only 1 feature was used. Our project will also investigate how the timeframe would affect the accuracy of price predictions of different models. As models have to reach a certain threshold to have significance for the users to work as a reference, it is essential for us to optimize our model to figure out what the optimal parameters and model structure are for our stock price prediction purpose.

RESULTS AND DISCUSSION

In this study, for datasets (Word press) and classifiers have been used for stock market prediction. The comparison of the classifier constructed in this study with similar systems in the literature is a challenging task due to diversities in the classification techniques, and data extraction methods. All the classifiers used in this study were trained twice, once with normal data and then with leaked data. Further, results concluded that Random Forest with leaked dataset and Bagging with leaked dataset provides above satisfactory performance.

CONCLUSION

In the project, we proposed the usage of information accrued from exceptional worldwide monetary markets with system learning algorithms to expect the stock index moves. Our conclusion may be summarized into following aspects: SVM algorithm works on the massive dataset fee which collected from special international monetary markets. additionally SVM does no longer provide a problem of over fitting. Correlation analysis shows robust interconnection among the marketplace inventory index and international markets that near proper earlier than or at the very starting of buying and selling time. diverse device gaining knowledge of primarily based fashions are proposed for predicting daily fashion of marketplace stocks. Numerical results advise high efficiency a realistic trading version is built upon our nicely educated predictor. The version generates better earnings as compared to chose benchmarks.

REFERENCES

- [1]. Ehsan Hoseinzade, Saman Haratizadeh, (2019) "CNNpred: CNN-based stock market prediction using a diverse set of variables", Expert Systems with Applications, vol. 129, pp. 273-285.
- [2]. G. Liu and X. Wang, (2019) "A Numerical-Based Attention Method for Stock Market Prediction With Dual Information," IEEE Access, vol. 7, pp. 7357-7367.
- [3]. J. Lee, R. Kim, Y. Koh and J. Kang, (2019) "Global Stock Market Prediction Based on Stock Chart Images Using Deep Q-Network," in IEEE Access, vol. 7, pp. 167260-167277.
- [4]. Kang Zhang, Guoqiang Zhong, Junyu Dong, Shengke Wang, Yong Wang, (2019) "Stock Market Prediction Based on Generative Adversarial Network", Procedia Computer Science, vol. 147, pp. 400-406.
- [5]. S. M. Idrees, M. A. Alam and P. Agarwal, (2019) "A Prediction Approach for Stock Market Volatility Based on Time Series Data," IEEE Access, vol. 7, pp. 17287-17298.