International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Stock Market Prediction With Risk Analysis Using Two ml Module

Sayem Patni[1], Dr. Amit R Gadekar[2]

Submitted: 04/09/2022 **Accepted:** 24/12/2022

Abstract

Attempts to predict movements in stock prices have historically proven difficult for academic researchers. Seminal publications in the literature have shown that the seemingly random movement patterns of stock price time series may be anticipated with a high degree of accuracy, contrary to the claims of proponents of the efficient market theory. Such risk-adjusted prediction models require appropriate variable selection, variable transformation processes, and model parameter adjustment. This study proposes a methodology for predicting stock prices using a combination of statistical analysis and machine learning that is both predictable and accurate in its risk analysis. We utilize five-minutely daily stock price data from a major company listed on India's National Stock Exchange (NSE). When building and training forecasting models, the granular data is aggregated into various time slots throughout the day. We propose that agglomerative model development, combining statistical and machine learning approaches, may successfully learn from unpredictable and erratic movement patterns in stock price data while mitigating risk. Using this effective learning, models can be trained to be robust and low-risk, increasing their utility for predicting stock movement patterns and short-term stock prices. Regression and classification models are built using statistical and machine-learning techniques. A large amount of data on these models' efficacy has been provided and thoroughly examined.

Keywords: Prediction, Machine learning, Stock market trend, Feature engineering, risk analysis.

I. Introduction

Predicting the behavior of stock prices in the future has been the subject of much research. Some proponents of the efficient market hypothesis claim that accurate stock price predictions are impossible, while other arguments have shown that this is not necessarily the case if the appropriate language and models are used. The second school of thought advocated using carefully selected variables and appropriate functional forms or forecasting models to build precise statistical, econometric, and machine learning models. Theories based on time series analysis and decomposition are available in the literature for projecting stock prices into the future. It's an improved and expanded version of our earlier work (Mehtab & Sen, 2019). Eight different types of models—classification, regression, a state-of-the-art deep learning model using long- and short-term memory (LSTM), and four different forecasting models using convolutional neural networks—are presented in this research as part of a prediction framework (CNN).

Investors and traders know the stock market to be highly unpredictable and complex. Numerous (macro and micro) factors, including politics, international economic conditions, unexpected events, a company's financial performance, and others, make accurate stock price predictions challenging.

This means there is a mountain of information from which to conclude. Analysts from the financial sector, academia, and the data science community all keep digging into analytics tools in search of stock market regularities. This resulted in the development of algorithmic trading, which employs automated, pre-programmed trading methods to carry out trades.

A breakdown of how the rest of the project is structured follows. Modeling and predicting stock prices are the subject of Section 2, which provides a complete literature analysis on the topic. Our methods are discussed in depth in Section 3 of this publication. In Section 4, we describe in detail the benefits of machine learning. In this part, we also compare the models' overall performance. The Methodology is Detailed Upon in Section 5.

[1]*Research Scholar, School Of Computer Sciences and Engineering, Sandip University*
Maharashtra, India
sayempatni22@gmail.com
[2]*Associate Professor, School Of Computer Sciences and Engineering, Sandip University*
Maharashtra, India
amit.gadekar@sandipuniversity.edu.in

II. Related Work

Asghar et al. argue that the predictors often employed to develop stock price prediction models are poorly picked, leading to inaccurate stock price predictions. This results in subpar performance from the models. To improve predicting accuracy, the authors offer a multivariate regression model with which they advise selecting predictors with great care. The quality of the interaction with a system is enhanced by a simple and intuitive user interface. However, time series with extremely volatile stock values are outside the scope of this approach's simplicity. Models for forecasting stock prices using multiple regressions are presented by Park et al.

Relationships are examined using several methods, including the Granger causality test, the vector error correction model (VECM), and the arbitrage pricing theory (APT). The results suggest a key conclusion: long-term investments in the Chinese stock market generate a high rate of return, while similar investments in the Hong Kong stock market do not. Bao et al. hybrid deep learning. architecture for stock price prediction consists of the wavelet transform (WT), stacked autoencoders (SAEs), and long- and short-term memory (LSTM) gates. Stock price data is first decomposed using WT for data denoising. After the data has been denoised, it is sent to the SAEs, extracting the data's deep features and sending them on to the LSTM module, which makes predictions about stock prices. It turns out that the model is remarkably precise. Instead of hourly or even daily stock price data, the model is tested using daily data. Therefore, it can't be relied on for making day-to-day trading decisions.

According to the findings, RNN performs better than CNN when it comes to identifying context and modelling temporal aspects for stock price prediction, but CNN performs better when it comes to extracting semantic meaning from text inputs to the model. The main problem with existing stock price prediction theories is that they cannot accurately predict short-term changes in stock prices. This study utilizes two deep neural networks and machine learning techniques to model and forecast stock price behavior, thereby compensating for this deficiency.

2.1 MOTIVATION

Current practices benefit major financial institutions rather than the broader public [10]. The average individual, who does not have a complete understanding of the stock market, needs a solution that teaches the investor about the firm's historical and predicted trend, as well as public opinion, and offers investments by assigning a risk percentage to each. There aren't many modest, simple smartphone methods for reading stock market data. This paper's solution provides all of these capabilities for free and is simple to use. Anyone can make an informed decision with the help of this product, which is consumer-ready and simple to use.

3. Methodology

As stated before, the primary purpose of this study is to establish a reliable methodology for making short-term predictions regarding stock price. Data on the short-term fluctuations in the prices of stocks is gathered using the Meta stock platform (Meta stock). For this report, we gathered data about Godrej Consumer Products Ltd. Data was collected at 5-minute intervals for each trading day in 2013 and 2014 when the National Stock Exchange (NSE) was operational. Date, time, open price, high price, low price, closing price, and the number of shares traded were all included in the raw data for each stock. The stock market's closing price is recorded at 5-minute intervals, hence the term "variable time." Therefore, the raw data included a 5-minute gap between each pair of entries. This raw data format was generated over two years for the stock of Godrej Consumer Products. We collected the NIFTY index at 5-minute intervals for the same two-year period as the seven parameters described above in the raw data to predict the full market sentiment at each instant using the combined data of past stock prices and the market sentiment index. This means there are seven distinguishable factors in the raw data for both stocks. We aggregate the raw data because a 5-minute timeframe is insufficient. We split the entire day into three segments: (1) the morning session, which runs from 9:15 AM to 11 AM; (2) the midday slot, which runs from 11:45 AM to 2:30 PM; and (3) the evening session, which runs from 2:35 PM till the NSE closes on that day. There are now three entries in the daily stock information, each containing stock price data for a period.

Metrics for evaluating and providing support services

Due to the nature of stock price prediction as a regression problem, we will be using RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error%) to assess the performance of our models. Both methods are helpful in assessing the precision of a forecast.

$$MAPE = \frac{1}{N} * \sum_{t=1}^{N} \left| \frac{At - Ft}{At} \right|$$

$$RMSE = \sqrt{\frac{1}{N} * \sum_{t=1}^{N} (At - Ft)^2}$$

where N = the number of time points, At = the actual / true stock price, Ft = the predicted / forecast value.

RMSE offers the discrepancies between anticipated and actual values, whereas MAPE (%) examines this difference in respect to the genuine values. For example, a MAPE score of 12% indicates a 12% average difference between forecast and actual stock prices.

4. IMPLEMENTATION

Based on a client-server architectural concept, the implementation was made. When a user uploads an input to Firebase using a mobile application, the application serves as the client in this scenario and sends a request to the server. With the aid of an event listener function, which records any change in the firebase and executes the relevant procedures once the change is logged [14], the server continuously listens for requests from the user. Because they are sequential, the implementation can be split into two modules.

Module 1: Prediction of stock values using polynomial regression

Predicting future stock market values is the subject of the first module. A machine learning model was used to do this.

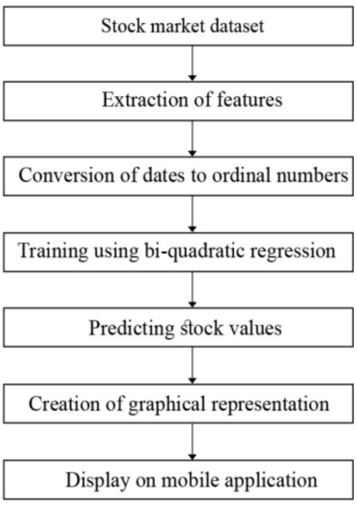

Fig 1: Block diagram for module 1 (Stock value prediction)

The opening stock prices for more than 30 firms from the dates 2006-01-01 to 29-12-2017 were used as the training data, and they were found on Kaggle in the format YYYY-MMDD. The variables for the machine learning algorithm were chosen to be the stocks' dates and opening prices. The dates had to be converted into ordinal numbers to be used as a feature, which was done using the to ordinal function and the date time package in Python. The stock value corresponding to the ordinal number was utilized as the dependent variable, and the ordinal numbers obtained as a result of these processes were used as the independent variables. Since bi-quadratic regression does not overfit or underfit the training set of data, it was chosen as the model. The regression model's training set included 2000 days. In Fig. 1, this procedure is depicted. The Python sklearn and numpy packages [11, 12] provided access to the machine learning function. The stock value for any needed date was then determined using the trained model.

The matplotlib module for Python plots the resultant graph after conducting curve fitting on the training data [11]. This computation was carried out on a remote backend Google Cloud Server in the Singapore region. The generated graphs were saved in the Google Cloud Storage Bucket and may be dynamically seen from the app by executing API queries. These graphs clearly show the trend of the stock value increase or decline for the company, enabling the customer to make an educated choice [21].

Module 2: Risk and fluctuation in the estimation of stock prices For each of the five organizations, a risk calculation is done, and the results are shown on the mobile application. Through the mobile application, users can input information such as the number of days for the investment and the amount to be invested, as shown in Fig. 02.

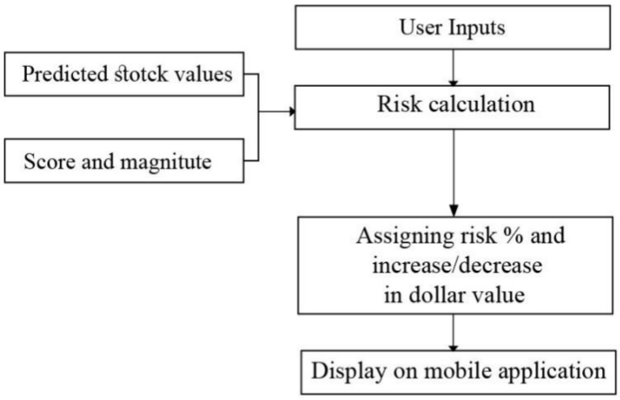

Fig 02: Block diagram for module 2 (Risk calculation)

Using the arrow function in the Python library, it was possible to determine the date for the current day, which is the day the investment period begins [11]. The trained regression model was used to find the stock values for all days from this day until the end of the investment period after converting this date to an ordinal number. These values were then compared to the stock price for the current day, and a count was made for each day on whether the stock price was higher or lower than the values anticipated for the subsequent days. All 5 companies had this surgery done. The score and magnitude values acquired by conducting sentiment analysis on each of the 5 companies and adding them together were discovered in the previously calculated product. Together with the count values, these values created a precise risk prediction metric. A risk percentage range was assigned using these two values.

The mean of the forecasted values was taken and deducted from the current day value, which will be the price at which the investor will be investing, to determine the dollar growth or reduction in value for each stock. If the difference is negative, the stock value has decreased; if it is positive, the stock value has increased. These values are uploaded to the firebase and the risk percentage range, from which the mobile application retrieves the data to show to the user.

4.1 Stock market forecast

Sensex and social media mood data are combined with risk analysis results to forecast the stock market. The method for forecasting stock market outcomes is shown in Table 1.

Table 1. Prediction of Sentiment and Sensex-Moving Average Final Result

Sentiment Analysis Result	Sensex-Moving Average Result	Final-Result Prediction
Positive	Positive	Positive
Positive	Negative	Neutral
Negative	Positive	Neutral
Negative	Negative	Negative

The final projection is also optimistic if the sentiment analysis and the Sensex Moving Average results are good. If both are negative, both combinations will yield neutrality, and the result will be erroneous.

4.2 Experimental results

This experimental study compiles Dow Jones stock market forecasts for RABK (DJ). The suggested algorithm's performance is compared to stock market forecasting algorithms based on data mining [16]. The Dow Jones (DJ) Oracle database contains historical prices for each of the 230 listed businesses on the exchange dating back to 2000. From 2005 to 2007, historical prices were computed.

For April 2006, the moving average is determined. Emotion in April 2000 is calculated in the same way. After combining the observations, a stock market forecast is made using the data in Table 1. The month and year are crucial for sentiment research. The following Fig.3, which highlights positive news outcomes, illustrates the positive and negative phrases in the news text.

The moving average is used to examine the Sensex's historical data. The 5-day, 10-day, and 15-day moving averages calculated for the Sensex point are displayed in Fig. 3. This is followed by the moving average computation.

Five-day moving averages and ten-day moving averages are displayed in the graph below in blue, red, and green, respectively.

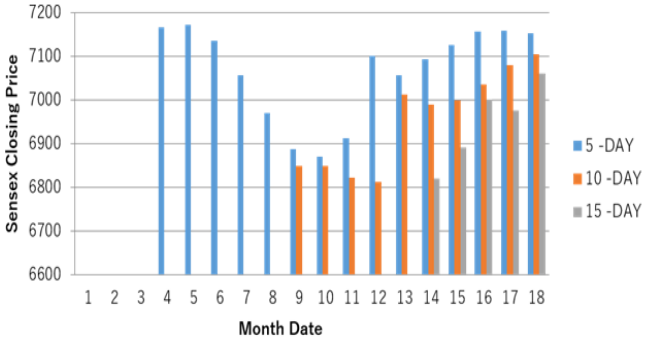

Fig 03: Result analysis of Sensex and stock market

5. CONCLUSION

The results show that, upon user input, the risk factor derived using yahoo data offers us a risk percentage range and the increase per stock in dollar value instantly and with an acceptable degree of accuracy. In conclusion, it can be said with confidence that the proposed system differs from the current system in that, in contrast to other systems, it assigns a risk percentage corresponding to each company based on the sentiment analysis and machine learning results rather than attempting to predict the accurate stock value of the company. When bi-quadratic regression was utilized, the stock value estimates had a plus or minus 1.5-dollar range. However, as the risk also depends on the sentiment analysis component, these variances have little impact on the risk percentage assignment.

The finished product is a reliable, fully functional final product prepared for consumer usage. Any regular guy with little knowledge of how the stock market operates can utilize this product and still make an informed choice.

FUTURE WORK Due to computational power limitations, the current solution is limited in terms of speed and efficiency, although better hardware or processors can be employed to handle concurrent users and large amounts of server requests. The system can be designed so that it can gradually learn the user's behaviors, such as their investing schedule, preferences, and risk tolerance, and then offer ideas that are more suited to their requirements [18]. To give the customer more alternatives, the five firms can be enlarged to include all of the companies listed on the stock exchange.

References:

[1]. M. Raddant, D.Y. Kenett, Interconnectedness in the global fnancial market. J. Int. Money Finance 110, 102280 (2021)

[2]. J. Zhao, M. Shahbaz, X. Dong et al., How does fnancial risk afect global CO2 emissions? The role of technological innovation. Technol. Forecast. Soc. Change

168, 120751 (2021) optimization algorithm for multimodal multi-objective optimization. Swarm Evolut. Comput. 62, 100843 (2021)

[3]. A. Kim, Y. Yang, S. Lessmann et al., Can deep learning predict risky retail investors? A case study in fnancial risk behavior forecasting. Eur. J. Oper. Res. 283(1), 217–234 (2020)

[4]. K.T. Chui, D.C.L. Fung, M.D. Lytras et al., Predicting at-risk university students in a virtual learning environment via a machine learning algorithm. Comput. Hum. Behav. 107, 105584 (2020)

[5]. T.S. Kumar, Data mining based marketing decision support system using hybrid machine learning algorithm. J. Artif. Intell. 2(03), 185–193 (2020)

[6]. Y. Chen, S. Hu, H. Mao et al., Application of the best evacuation model of deep learning in the design of public structures. Image Vis. Comput. 102, 103975 (2020)

[7]. Q. Wang, M. Su, Integrating blockchain technology into the energy sector—from theory of blockchain to research and application of energy blockchain. Comput. Sci. Rev. 37, 100275 (2020)

[8]. A.M. Acquah, M. Ibrahim, Foreign direct investment, economic growth and fnancial sector development in Africa. J. Sustain. Finance Invest. 10(4), 315–334 (2020)

[9]. B.M. Henrique, V.A. Sobreiro, H. Kimura, Literature review: Machine learning techniques applied to fnancial market prediction. Expert Syst. Appl. 124, 226–251 (2019)

[10]. M. Andoni, V. Robu, D. Flynn et al., Blockchain technology in the energy sector: a systematic review of challenges and opportunities. Renew. Sustain. Energy Rev. 100, 143–174 (2019)

[11]. Q. Yang, Y. Wang, Y. Ren, Research on fnancial risk management model of internet supply chain based on data science. Cogn. Syst. Res. 56, 50–55 (2019)

[12]. A.A. Nagra, F. Han, Q.H. Ling et al., An improved hybrid method combining gravitational search algorithm with dynamic multi swarm particle swarm optimization. IEEE Access 7, 50388–50399 (2019)

[13]. J.L. Speiser, M.E. Miller, J. Tooze et al., A comparison of random forest variable selection methods for classifcation prediction modeling. Expert Syst. Appl. 134, 93–101 (2019)

[14]. R. Janani, S. Vijayarani, Text document clustering using spectral clustering algorithm with particle swarm optimization. Expert Syst. Appl. 134, 192–200 (2019)

[15]. T. Fischer, C. Krauss, Deep learning with long short-term memory networks for fnancial market predictions. Eur. J. Oper. Res. 270(2), 654–669 (2018)

[16]. M. Azam, S.A. Raza, Financial sector development and income inequality in ASEAN-5 countries: does fnancial Kuznets curve exists? Global Bus. Econ. Rev. 20(1), 88–114 (2018)

[17]. C. Alexiou, S. Vogiazas, J.G. Nellis, Reassessing the relationship between the fnancial sector and economic growth: dynamic panel evidence. Int. J. Financ. Econ. 23(2), 155–173 (2018)

[18]. H. Pollitt, J.F. Mercure, The role of money and the fnancial sector in energy-economy models used for assessing climate and energy policy. Clim. Policy 18(2), 184–197 (2018)

[19]. G. Chen, B. Xu, M. Lu et al., Exploring blockchain technology and its potential applications for education. Smart Learn. Environ. 5(1), 1–10 (2018)

[20]. A.J. Asaleye, J.I. Adama, J.O. Ogunjobi, Financial sector and manufacturing sector performance: evidence from Nigeria. Invest. Manag. Financ. Innov. 15(3), 35–48 (2018)