INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

IJARIIT

ISSN: 2454-132X
Impact Factor: 6.078
(Volume 7, Issue 4 - V7I4-1715)
Available online at: https://www.ijariit.com

A speculation technique for the stock market using time series

Rajesh K. S.
rajeshks.hrr@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

B. S. Darshith
bsdarshith1106@outlook.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Amaresh A. Channad
amareshchannad@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

Amaresh R. Vasisth
amaresh.vasisth@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

G. Jayavardhan
jaivardhan.19.1999@gmail.com
Rajarajeswari College of Engineering, Bengaluru, Karnataka

ABSTRACT

Using ARIMA (Auto Regressive Integrated Moving Average) which is a model, this research study presents a method for anticipating stock market values. To begin, historical stock price data is gathered and pre-processed. Pre-processing which is a subcategory of Data Mining in which raw original data is processed into a specified format so that the model can use it. Pre-processing is done to fill in the missing values and arrange the data based on the qualities required for developing the prediction model because the raw data to be processed is partly partial or inconsistent with certain errors and missing values. After the data has been filtered and classified, it is standardized by transforming it into a common format that can be used to train the prediction model. The obtained data is then divided into two data - training and testing data, with the majority of the data being used to train the model. And using the variables linked with the stock price, the testing data is fed to the ARIMA model to predict the corresponding stock price on a particular date. The results are used to determine the model's accuracy.

Keywords *- Machine Learning, Dicky Fuller Test, NSE India, Data Extraction, L-jung Box Test.*

1. INTRODUCTION

Attempting to evaluate stock valuation and providing investors with a better grasp of business and stock price information is known as stock market prediction. Because stock market trends are dependent on a variety of variables and events, a single dataset may not be sufficient to forecast stock market trends, and a single data set may produce inaccurate results. Using a traditional classifier makes it unreliable and fails when circumstances alter in some way. Our fundamental idea was that by using machine learning and training it on historical data from websites, we could forecast the stock price closing rate. We may not always be able to discover patterns because of the complex interactions between inputs and outputs. It has a strong ability to find relationships in input data sets without making any assumptions about how the input and output data are related.

2. LITERATURE SURVEY

The impact of social media tweets on stock exchange in Jamaica is seen, according to SA Bogle and WD Potter [1]. Sentiment analysis necessitates the employment of a pre-processing technique due to the unstructured nature of tweets.[1]Different Machine learning predictors such as neural networks, support vector machines SVM, and decision trees all of them were used to construct stock projections based on news sentiment data. The accuracy of motion prediction was determined and calculated to be 87 percent, while the correlation coefficient for price prediction was 0.99, indicating that there is still space for improvement. [1]. Similarly, by analysing tweet sentiments, [2] discovered a high correlation between a company's stock price and the emotions or public opinions stated about it on Twitter. Using Ngram and Word2vec textual representations in combination with logistic regression is another technique to detect public viewpoints in tweets. The link between market sentiment and public sentiment was proven in a study that used Twitter data to gauge public mood and used previous day's values to forecast changes. The method for forecasting is based on past stock market values and sentiment analysis of financial news. The model achieved accuracy rates ranging from 72 to 86.21 percent by taking into account historical stock prices as well as various types of business and market news.[4]. Combining past stock prices with news headlines helps improve prediction accuracy. [5]. Similarly, [6] discusses strategies for predicting Indonesian stock market movement based on sentiment values in tweets, such as margin percentage prediction, stock price prediction, and price fluctuation prediction. In compared to other commonly used classification algorithms, random forest and nave bayes classification algorithms of prediction fared well. [6]. On the other hand, the prices from the prior five days are extremely useful for forecasting. Bing used a model [7] to analyse hourly stock price trends and public tweets. [7] discovers connected relationships between quantitative stock prices and public mood using data mining and natural language processing technologies. The

© 2021, www.IJARIIT.com All Rights Reserved

sentiment analysis of news articles to influence share price is reported in [8], and the news dataset was obtained using the Bing API. Using a customised sentiment lexicon to analyse stock articles, as discussed in [8, is a clever technique to handle the prediction strategy. Qing Li et al. [9] discovered that firm-specific news stories can boost investors' understanding by using a proposed quantitative media-aware trading technique to see its impact on stock markets. Emotional changes in investors cause individuals to make decisions based on their emotions [8, 9].The use of supervised machine learning algorithms to develop a daily and monthly forecast model was influenced by article content and business characteristics. For the daily prediction model, sentiment analysis was combined with historical data. For the monthly prediction model, two months' worth of trends were used to check whether there was any similarity between them. The findings revealed that one month's trend is the least related to the next month's trend.

3. PROPOSED METHODOLOGY

Fig.1: System Architecture

The proposed methodology has 4 modules

- **Data Extraction**
- **Data Pre-Processing**
- **Building a model**
- **Review Analysis**

Data Extraction

To train the model for stock market prediction, we require the previous year's stock prices, which is referred to as historical data. To acquire these statistics, we use the Nifty50 website. It consists of a compilation of stock market data from prior months or years, such as opening and closing prices, volume, and a variety of other metrics. It is stored in a csv file (excel file). To read the csv file, we'll utilise the pandas package.

Data Pre-Processing

The data is then pre- processed which means when we download the excel file there will data which are not required. Those data will be removed manually in the excel file. Then the data is tested for stationarity after it has been extracted. To acquire reliable findings, the given data must be stationary. If the data is not stationary, the ARIMA model's Integration model must be used to make it stationary. The integration model determines the degree of data differencing. The data is shown using the ACF (Auto Correlation Function) and PACF (Parametric Auto Correlation Function) after the degree of difference has been determined (Partial Auto Correlation Function). The ACF and PACF values are used to estimate the values of p, d, and q.

Fig.2: Data Pre-processing

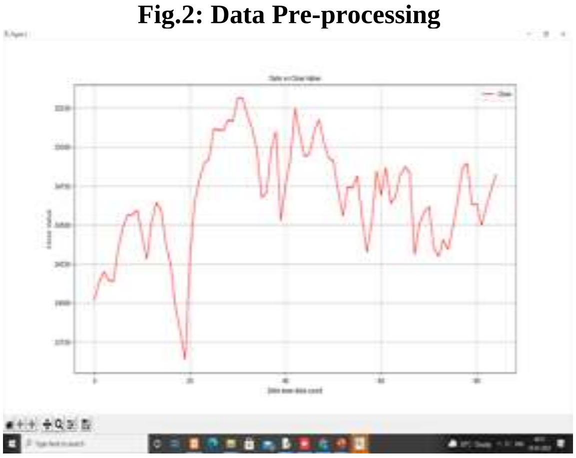

Fig.3: Date vs Close Value

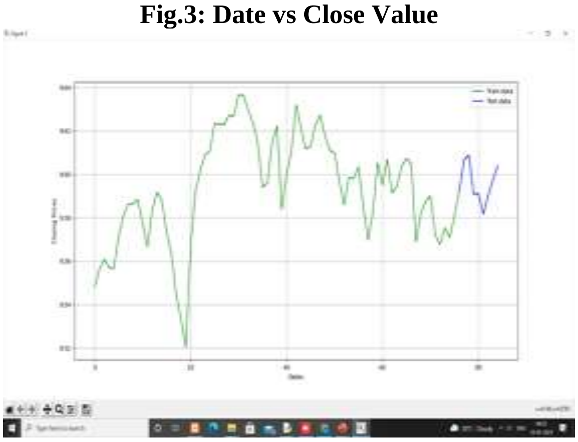

Fig.4: Train vs Test Data

Building a Model

To obtain the findings, we use auto.arima() function. To obtain the forecasted series, the auto.arima() function selects the most appropriate "ARIMA (p, d, q)" parameters. A "trace" is used by the auto.arima() method to elucidate why (p,d,q) parameters chosen for the ARIMA (p, d, q) model are the best. Essentially, the "AIC" is what determines the model's correctness. The "AIC" value with the lowest value indicates the best model for prediction.

Fig.5: Building A Model

© 2021, www.IJARIIT.com All Rights Reserved

Review Analysis

The forecast is completed once the model has been developed. The L-Jung-Box test can also be used to validate the predicted series and determine whether the residuals are random. "Whether our time series residuals follow a haphazard pattern or whether there is a considerable degree of non-randomness," this test determines. The reasoning behind this is that if the residuals are correlated, we can deduce that the model is unable to handle time series behaviour and that our time series will suffer as a result. The L-Jung Box test should yield a result of at least 0.05 to confirm that the time series residual follow a random pattern.

```
                               SARIMAX Results
==============================================================================
Dep. Variable:                      y   No. Observations:                   56
Model:               SARIMAX(0, 1, 0)   Log Likelihood                 158.502
Date:                Mon, 19 Jul 2021   AIC                           -315.005
Time:                        11:31:36   BIC                           -312.998
Sample:                             0   HQIC                          -314.225
                                 - 56
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
sigma2         0.0002   2.61e-05      7.038      0.000       0.000       0.000
===================================================================================
Ljung-Box (L1) (Q):                   1.60   Jarque-Bera (JB):                 6.49
Prob(Q):                              0.21   Prob(JB):                         0.04
Heteroskedasticity (H):               0.48   Skew:                             0.31
Prob(H) (two-sided):                  0.13   Kurtosis:                         4.56
===================================================================================
```

Fig.6: Review Analysis Results

4. RESULTS

Fig.7: Date vs Open & High Value

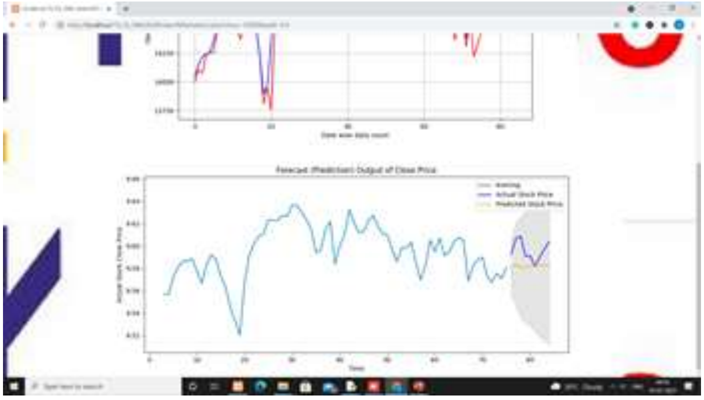

Fig.8: Closing Price Prediction Graph

The training data is shown in the diagram above, and it is fed to the model to train it.

The leftover (testing) data is utilised to determine the model's accuracy after the training data is supplied as input to the model. This data is used to see how accurately the model can estimate stock market prices; it is part of the pre-processed data, and after the testing data is entered into the model, the model analysis process begins.

The customer can see a graphical representation of the anticipated stock price compared to the actual stock price.

5. CONCLUSION

In the context of Nifty50 data, successfully taught the notion of time series analysis and forecasting. The sharp depreciation of the Indian rupee in recent years has necessitated the use of stock market forecasting to protect investors' interests. An ARIMA model was created to forecast the volatility of the Indian Nifty 50 stock market. The data for this study came from the NSE's official website. The result shows a 5% mean variation in the actual mean value of the Nifty 50 data. Future goals include effectively using the same ARIMA Model for SENSEX data and attempting to provide as much accuracy as possible in determining stock market volatility and real-time closing prices in the future.

6. REFERENCES

[1] S. A. Bogle , W.D. Potter, "SentAMaL - A Sentiment Analysis Machine Learning Stock Predictive Model," in Proc. of Int. Conf. of Data Mining and Knowledge Engineering , 2015.

[2] V. S. Pagolu, K.N. Reddy, G. Panda , B. Majhi , "Sentiment Analysis of Twitter Data for Predicting Stock Market Movements," Int. conf. on Signal Processing, Communication, Power and Embedded System(SCOPUS) 3-5 Oct. 2016.

[3] A. Mittal , A. Goel , "Stock Prediction Using Twitter Sentiment Analysis," Standford University, 2012.

[4] P. S. Michael Rechenthin ,W. N. Street, "Stock Chatter: Using Stock Sentiment to Predict Price Direction, "Algorithmic Finance, vol. 2, no. 3-4, pp. 169-196, 2013.

[5] A. E. Khedr , S.E. Salama , and N. Yaseen , "Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis," Int. J. Intelligent Systems and Applications,pp. 22-30 , 2017.

[6] B. D. Trisedya, Y. E. Cakra, "Stock Price Prediction using Linear Regression based on Sentiment Analysis", Int. Conf. Adv. Computer. Sci. Inf. Syst. , pp. 147–154, 2015.

[7] L. I. Bing , K. C. C. Chan , C. Ou, "Public Sentiment Analysis in Twitter Data for Prediction of a Company's Stock Price Movements," IEEE 11th Int. Conf. E-bus. Eng. pp. 232-239, 5-7 Nov. 2014.

[8] V. Kalyanaraman , S. Kazi , Rohan, Tondulkar , S.Oswal , " Sentiment Analysis on News Articles for Stocks," 8th Asia Modelling Symposium, 23-25 Sept. 2014.

[9] Q. Li , T. Wanga, P. Li , L. liu, Q. Gonga, and Y. Chenb, " The effect of news and public mood on stock movements ," Information Science , Vol. 278, pp. 826-840, Sep 10, 2014.

[10] A. Nayak, M. M. Manohara Pai, R.M. Pai "Prediction models for Indian stock market," Procedia Computer science, Vol. 89, pp. 441-449, 2016.

[11] Sheikh Mohammed Idrees, M. Afshar Alam, Parul Agarwal. "A Prediction Approach for Stock Market Volatility Based on Time Series Data", IEEE Access,2019

[12] González'Rivera G, Lee TH. Nonlinear time series in financial forecasting. Encyclopedia of Complexity and Systems Science. Springer, New York. 2009.

[13] P. Pai and C. Lin, "A hybrid ARIMA and support vector machines model in stock price prediction", Omega vol.33 pp. 497-505, 2005

[14] J.J. Wang, J.Z. Wang, Z.G. Zhang and S.P. Guo, "Stock index forecasting based on a hybrid model", Omega vol.40 pp.758-766, 2012.
[15] L.Y. Wei, "A hybrid model based on ANFIS and adaptive expectation genetic algorithm to forecast TAIEX", Economic Modelling vol. 33 pp. 893-899, 2013.
[16] Montgomery, Douglas C., Cheryl L. Jennings, and Murat Kulahci. *Introduction to time series analysis and forecasting*. John Wiley & Sons, 2015.
[17] Gerbing, D., 2016. Time series components. *Portland State University*, p.9.
[18] Adhikari, R., & Agrawal, R. K. (2013). An introductory study on time series modeling and forecasting. *arXiv preprint arXiv:1302.6613*.
[19] Granger, C.W.J. and Newbold, P., 2014. *Forecasting economic time series*. Academic Press.
[20] Petrevska, B., 2017. Predicting tourism demand by ARIMA models. *Economic research-Ekonomska istraživanja*, *30*(1), pp.939-950.
[21] Zhang, M. and Pi, D., 2017. A New Time Series Representation Model and Corresponding Similarity Measure for Fast and Accurate Similarity Detection. *IEEE Access*, *5*, pp.24503-24519.