International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 8, ISSN No. 2455-2143, Pages 134-138
Published Online December 2021 in IJEAST (http://www.ijeast.com)

PREDICTING VOLATILITY OF STOCKS FOR TIME SERIES DATA USING STATISTICAL FORECASTING MODELS

Pavan Kumar T M, Sayantan Saha, Navila, Ramya K G
Department of Computer Science and Engineering
Dayananda Sagar College of Engineering, Bengaluru

Chaitra S P
Assistant Professor
Department of Computer Science and Engineering
Dayananda Sagar College of Engineering, Bengaluru

***Abstract*—Time series data is an ordered set of data points, which are values of a variable at different points in time, spaced at equal intervals. In other words, time series represents values of a specific variable measured at a constant time frequency. An example is of temperature of a city measured at 11 AM every day for 30 consecutive days. Analysis of time series data to forecast and predict the future finds use across multiple industries. A very important application of time series analysis is to predict prices of stocks in the future. Stock markets are complicated financial structures where stocks of public companies are traded. Prices of stocks witness great deal of fluctuations and changes over time. Professional stock traders and investors analyze time series data to forecast prices of stocks in the future with reasonable accuracy, which helps them generate higher returns on their investments with minimal risk involvement. Stock prices are highly volatile and react very quickly to changes in the market. Time series forecasting is one technique that helps in developing a model that can accurately predict such drastic and quick changes. In our research, we have made an attempt to analyze the time series data of certain stocks from Indian stock market (NIFTY), using which we have developed a forecasting model to accurately and efficiently predict the prices of these stocks in the future.**

***Index Terms*—Stock price prediction, Deep-Learning, LSTM, RNN, ARIMA.**

I. INTRODUCTION

The implementations of the stock market and financial operations have changed massively due to the digitalization of financial activities, internet connectivity and support of related softwares. Speculation of stock prices in the future is fraught with uncertainties that even industry insiders find difficult. Stock market is unpredictable with sudden and massive reactions to external and internal factors. Financial environment of the country, business situation, along with other macroeconomic and microeconomic factors play a role in determining stock prices, thereby making the task of prediction a challenging one. Deep learning and machine learning can provide a solution to these challenges by analyzing data from the past and evaluating them in order identify pattern and make use of the same to predict future prices of the stocks. Quite a few methods function with just static data from the past to make future forecasts. While few other techniques make use of factors such as price, volume, financial ratio, technical factors and the "OHCLV" parameters, which are Open, High, Close, Low, & Volume respectively; with variation in these factors. Such datasets with variety along with algorithms like Support Vector Machine (SVM), Logistic Regression, Naïve Bayes and Random Forests, the forecast precision or correctness likely expected middle from two points a limited value of 45.1% to a maximum financial worth of 86.2% only. RMSE, MSE and Mean Absolute Error (MAE) exist few influential measures secondhand fashionable the judgment of reversion located question. The various methods bear an RMSE of about 21 and MAE of 4.98 to 16.68 accompanying a predictable difference of 3.7%. We bear secondhand the S&P 500 stock index for this paper. The NIFTY 50 stock exchange index exist used to calculate the stock performance of top 50 big crowd of people of differing stock exchanges across India. It exist deliberate expected a justly close likeness of the stock exchange. Our objective at this moment paper is to forecast the stock prices of the various filed party. We have in mind to take live stock exchange information in visible form from the APIs, develop in mind or physically a model establish ARIMA-LSTM for declaration made in advance, and compare the accomplishment of two together algorithms. The better taking part in a presentation invention will exist used to express an outcome in advance future stock exchange accomplishment and stock prices.

II. METHODOLOGY: ALGORITHMS

A. ARIMA

ARIMA stands for "Auto-Regressive Integrated Moving Average". It is an algorithmic model which is precisely used in statistics and econometrics for computing and analyzing events which take a period of time to occur. This algorithm is applied to analyse historical data and forecast data of the future, when considering historical data spaced by equal time intervals. The data to be analysed is collected at equal intervals, can be spaced apart by seconds, minutes, hours, or months. Such as daily temperature at 11 AM, monthly sales of a store, etc. This model falls under the kind called Box-Jenkins Method.

Two important techniques implemented in prediction of time series are univariate and multivariate analysis:

- Univariate examination and determination utilises the historic principles of time series information i.e, the data and uses it to calculate the future values
- Multivariate analysis uses data variables which are external along with the time series data.

ARIMA model analyses historic time series data to predict the values data might take in future and may be used for prediction of non-seasonal series of numbers exhibiting patterns, instead of being random values. For example, sales data of a grocery outlet will exhibit characteristics of time series since it is data collected over a period of time, at a regular frequency. In order to make predictions over multiple seasons, the model can be altered to suit our needs. When data is collected over different seasons, the data needs to be adjusted to account for seasonal variations. Additionally, holidays have an impact on data too. Holidays may lead to increase or decrease in sales depending on when the holiday falls in the calendar.

The ARIMA model is finding acceptance as a valuable tool to make predictions of future such as sales forecasts, demand of products or prices of stocks.

To understand ARIMA, one should analyze the name. The "AR" implies auto-regression, which display the model that shows a dynamic variable regressing on its own earlier or lagged data i.e, values. Which means that it predicts the upcoming values based on its own past values. The "I" implies integrated, that means it evaluates the difference between the static data values and its previous values. Lastly, "MA" indicates the moving average, the reliance between an observed value and a residual error from a moving average model which was applied to earlier observations.

There are three component functions in the ARIMA model:

- AR (p): describes the time varying process in financial data.
- I (d): shows the difference in the non-seasonal observations.
- MA (q): the size of the moving average window.

An ARIMA model order is symbolically represented as (p,d,q). These are values for the order or number of times the function occurs in running the model. Values of zero can be accepted.

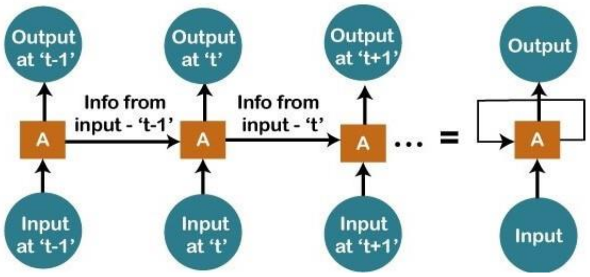
ARIMA models can be developed using both seasonal and non-seasonal data. The seasonal model should consider events occurring in each session apart from the autoregressive, differencing, and average terms of each season.

ARIMA models can be built using different software tools such as Python. Also, it must be ascertained that the process is a fit for the model. To do so, the model is built first and trained on a train dataset, before using actual data to make real forecast of stock prices.

B. Recurrent Neural Network

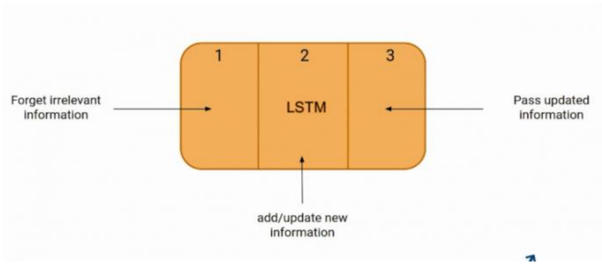
RNNs are the only one amongst neural networks with an internal memory, which makes them highly powerful and suitable for a wide variety of applications. Recurrent neural networks have been created in 1980s, which makes them relatively old like other deep learning algorithms. However, it is in recent years that they've seen rampant usage. Drastic increment of available computing resources, huge amounts of data availability, and advent of long short-term memory (LSTM) in 1990s have made application of RNNs convenient. RNNs have the potential of recalling important inputs by storing them in internal memory and using the data for predictions, thereby making predictions accurate. This capability makes recurrent neural networks more competent in analyzing sequential data and generating insights. Thus, they are the most favored algorithm for sequential data like time series, speech, text, financial data, audio, video, weather and many others.

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 8, ISSN No. 2455-2143, Pages 134-138
Published Online December 2021 in IJEAST (http://www.ijeast.com)



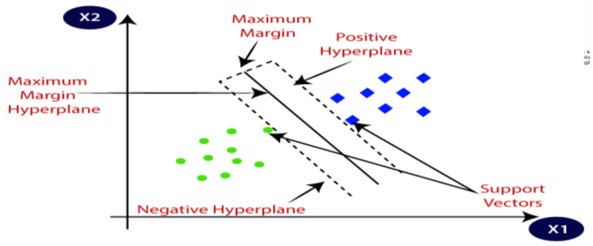
C. Long Short Term Memory

Long Short Term Memory Network is an advanced variant of RNN. It is a sequential network which allows for information storage. Long Short Term Memory Network is not affected by the vanishing gradient problem of the RNN. A recurrent neural network is used for its capability of persistent memory. An example of the same can be made of our memory, which remembers the previous scene when watching a video clip or remember details of previous chapters while reading a book. Long Short Term Memory Network type RNNs store the earlier information. They make use of the same when processing new data. Thus, they overcome the weakness of RNN which is not remembering long term dependencies due to vanishing gradient. Avoiding long-term dependency is one primary use case of Long Short Term Memory Networks.

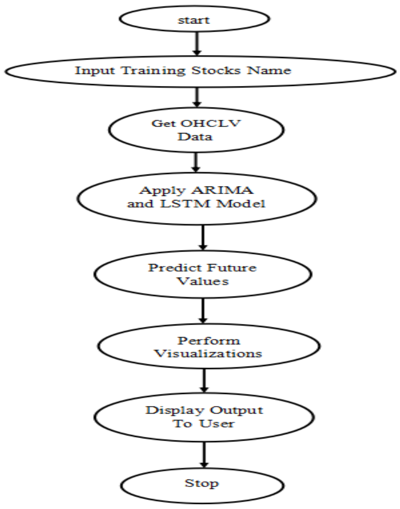


D. Support Vector Machine

Support Vector Machine or SVM is a frequently used Supervised Learning algorithm. It finds application in both Classification and Regression problems, though the former experiences more usage in Machine Learning applications. SVM algorithm attempts to create the best line or decision boundary that can segregate n-dimensional space into classes so that new data points can be put in the correct category in the future. This decision boundary which fits the best for the case is termed as the hyperplane. SVM allows the selection of the extreme points/vectors that lets creation the hyperplane. These extreme cases are termed as support vectors, which has led to the nomenclature of this algorithm. The figure displayed below shows two different categories that are classified using a decision boundary or hyperplane:



III. SYSTEM FLOWCHART



IV. CONTRIBUTION

Min Wen and others applied convolution neural networks on a reconstructed time series data. They found macroscopic of the information in visible form and caught a precision or correctness of nearly 56.14% and precision of 55.44% and recall of 74.75% utilizing motif extraction. Creighton and others used ARIMA and Backpropagation Neural Network (BNN) ahead of the S&P 500 and S&P 400 during the daily closing index so that it complements the future values. He also used the KoNstanz Information MinEr (KNIME) analytics that happens to be the best case for predicting an outcome in weekly information data. But that happens to be not proper for predicting an outcome i.e., data that occurs every day. The directional accuracy, MAE, MSE and RMSE of this model are: 45.1%, 16.68, 434.121 and around 20.836 respectively.
Hossain and others developed models that are capable of producing higher precision or correctness of MAE of approximately 0.023, compared to the individual LSTM or GRU layers, he created a composite model utilizing LSTM and GRU and applied it on S&P 500 data. Shah and others distinguished LSTM and Deep Neural Network (DNN) utilizing the Bombay Stock Exchange (BSE) data to estimate the same. Both the models gave an

RMSE of close to 1%. However, for data which required predictions every seven days, LSTM was found to be more precise. Iyer and others executed an examination and determination of miscellaneous models used to predict future values of stock and he distinguished the accomplishment of the various algorithmic models utilizing datasets from NSE, HSI, BSE and TWI. Zhang and others developed a forecasting method that could classify stocks in various terms like up, down, nil or flat depending upon the variation of price; TA- Lib open source library was used to gather data. This method displayed accuracy of 67.5% and also a SD of 3.7%. It proved to be beneficial in forecasting stock prices over a longer time period such as 30 to 40 days.

Authors	Dataset	Technique	Evaluation Metric
Wen et al.	S&P 500	CNN using Motif extraction	Accuracy- 56.14% Precision- 55.44% Recall- 74.75%
Hossain et al.	S&P 500	LSTM- GRU hybrid	MSE- 0.00098 MAE- 0.023
Shah et al.	BSE Sensex	LSTM, DNN	RMSE- 1%
Iyer et al.	NSE, Hang Seng, etc	Fuzzy System, ANN, etc	Accuracy- 88%
Zhang et al.	TA- Lib	Random Forest	Accuracy- 67.5% Std deviation- 3.7%
Creighton et al.	S&P 500 and S&P 400	ARIMA- BPNN hybrid	MAE- 16.68 MSE- 434.121 RMSE- 20.836 Accuracy- 45.1%
Selvin et al.	NSE	RNN, LSTM, CNN	MAE- 5.13% (RNN) 5.31% (LSTM) 4.98% (CNN)
Sadia et al.	Kaggle Dataset	SVM, Random Forest	Accuracy- 78.7% (SVM) 80.8% (Random forest)
Maini et al.	Dow Jones Industrial Average	SVM, Random Forest	Accuracy- 84.6% (SVM-linear) 85.18% (SVM-RBF) 86.2% (Random forest)
Deepak et al.	BSE Sensex	SVM- RBF kernel	Accuracy- 80 to 85%
Persio et al.	Google Assets	RNN, LSTM, GRU	Accuracy- 72%

V. CONCLUSION

Forecasting how market dynamics will change stock prices is a complicated, risk-intensive, and laborious task. Our proposal involves a novel ARIMA- LSTM algorithm to analyze volatile time series data of stock prices and forecast stock prices in the future. Based on the error rate of each model, we determine the better efficient one. The present work is used to identify trends in the data with the help of the techniques and statistics of the time series model using ARIMA and LSTM algorithms. As a future scope of this project, we can consider how news and current affairs impact the stock market performance. We can analyse the impact of news on the performance of the ARIMA and LSTM model and its impact on the error rate in predictions. Further, hybrid models utilizing deep learning techniques such as CNN may also be applied to measure the model's overall performance.

VI. REFERENCES

[1]. S. M. Idrees, M. A. Alam and P. Agarwal, "A Prediction Approach for Stock Market Volatility Based on Time Series Data," in IEEE Access, vol. 7, pp.17287-17298,2019,doi: 10.1109/ACCESS.2019.2895252.

[2]. M. A. Istiake Sunny, M. M. S. Maswood and A. G. Alharbi, "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model," 2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES), 2020, pp. 87-92, doi: 10.1109/NILES50944.2020.9257950.

[3]. Mahadik, D. Vaghela and A. Mhaisgawali, "Stock Price Prediction using LSTM and ARIMA," 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), 2021, pp.1594-1601, doi:10.1109/ICESC51422.2021.9532655.

[4]. K A. Althelaya, E. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short-and long-term stock market prediction,"?2018 9th International Conference on Information and Communication Systems (ICICS), 2018, pp. 151-156, doi: 10.1109/IACS.2018.8355458.

[5]. S. Siami-Namini, N. Tavakoli and A. SiamiNamin, "A Comparison of ARIMA and LSTM in Forecasting Time Series,"?2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018, pp. 1394-1401, doi:10.1109/ICMLA.2018.00227.

[6]. S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model,"?2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2017, pp. 1643-1647, doi:10.1109/ICACCI.2017.8126078.

[7]. Qian and X. Chen, "Stock Prediction Based on

LSTM under Different Stability,"?2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2019, pp. 483-486, doi: 10.1109/ICCCBDA.2019.8725709.

[8]. L. Yunpeng, H. Di, B. Junpeng and Q. Yong, "Multi-step Ahead Time Series Forecasting for Different Data Patterns Based on LSTM Recurrent Neural Network,"?2017 14th Web Information Systems and Applications Conference (WISA), 2017, pp. 305-310, doi: 10.1109/WISA.2017.25.

[9]. Sharaff and M. Choudhary, "Comparative Analysis of Various Stock Prediction Techniques,"?2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI), 2018, pp. 735-738, doi: 10.1109/ICOEI.2018.8553825.

[10]. Bathla, "Stock Price prediction using LSTM and SVR,"?2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2020, pp. 211-214, doi: 10.1109/PDGC50313.2020.9315800.

[11]. Y Mehta, A. Malhar and R. Shankarmani, "Stock Price Prediction using Machine Learning and Sentiment Analysis,"?2021 2nd International Conference for Emerging Technology (INCET), 2021, pp. 1-4, doi: 10.1109/ INCET51464.2021.9456376.

[12]. C. Li, X. Zhang, M. Qaosar, S. Ahmed, K. M. R. Alam and Y. Morimoto, "Multi-factor Based Stock Price Prediction Using Hybrid Neural Networks with Attention Mechanism,"?2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2019, pp.961-966, doi:10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00176.

[13]. R. Kumar, P. Kumar and Y. Kumar, "Analysis of Financial Time Series Forecasting using Deep Learning Model,"?2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 877-881, doi: 10.1109/Confluence51648.2021.9377158.

[14]. M. Goyal, "Financial Time Series Stock Price Prediction using Deep Learning,"?2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2020, pp. 0378-0383, doi: 10.1109/IEMCON51383.2020.9284879.

[15]. Shivani, K. S. Sandhu and A. Ramachandran Nair, "A Comparative Study of ARIMA and RNN for Short Term Wind Speed Forecasting,"?2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2019, pp. 1-7, doi: 10.1109/ICCCNT45670.2019.8944466.