

Stock Market Prediction Using Machine Learning

Himanshu Kumar, Aman Kumar, Deepak Kumar

Department of Computer Science Engineering

Dr. A. P. J. Abdul Kalam Technical University, Lucknow, UP, India

Abstract: *In Stock Market Prediction, the aim is to predict the long run worth of the monetary stocks of a company. The recent trend available market prediction technologies is that the use of machine learning that makes predictions supported the values of current stock exchange indices by coaching on their previous values. Machine learning itself employs totally different models to form prediction easier and authentic. The paper focuses on the employment of time Series and LSTM based Machine learning to predict stock values. Factors thought-about are open, close, low, high and volume.*

Keywords: Close, high, low, LSTM model, open, Time Series, Streamlit, Heroku and volume

I. INTRODUCTION

A correct prediction of stocks will cause vast profits for the vendor and therefore the broker. Frequently, it's brought out that prediction is chaotic instead of random, which implies it will be expected by fastidiously analyzing the history of various stock market. Machine learning is an economical thanks to represent such processes. It predicts a market price on the point of the tangible value, thereby increasing the accuracy. Introduction of machine learning to the world of stock prediction has appealed to several researches thanks to its efficient and correct measurements.

The important part of machine learning is that the information set used. The dataset ought to be as concrete as attainable as a result of a touch modification within the data will preserve huge changes in the outcome. During this project, statistic machine learning & Deep Learning supervised LSTM is used on a dataset obtained from Yahoo Finance. This dataset includes of following 5 variables: open, close, low, high and volume. Open, close, low and high are totally different bid costs for the stock at separate times with nearly direct names. The quantity is the range of shares that passed from one owner to a different during the time period. The model is then checked on the test data.

Time Series and LSTM models are engaged for this conjecture separately. Time series involves predicting output that relies on Time and LSTM contributes to basic cognitive process the information and results for the long run. Finally, the graphs for the fluctuation values of price with the dates (in case of time series based model) and between actual and expected price (forth LSTM based model) are plotted. The remainder of the paper consists of following: Section II discusses the connected work.

Section III puts forward the 2 models used and also the ways utilized in them in detail. Section IV discusses the results produced with totally different plots for each the models in detail. Whereas Section V consists of conclusion and also the last section involves the references.

II. RELATED WORK

From the literature survey, it absolutely was ascertained that the appliance of machine learning techniques to stock exchange prediction is being undertaken completely throughout the world. Machine Learning techniques are proving to be rather more correct and quicker as compared to modern prediction techniques.

Significant work has been done throughout the world in this field. A testament to which is the work of John Graunt. The primary actual application of autoregressive models to knowledge will be brought back to the work of G. U yule and J. Walker within the 1920s and 1930s. The ARIMA model was developed in the 1970s by George Box and Gwilym Jenkins. K. V. Sujatha and S. M. Sundaram prompt perceptive techniques on handling non-normal things which can often arise throughout the operating of the system and cause disruptions or cause inaccurate prediction Liu, G. Liao and Y. Ding conducted similar work and designed a model for applying LSTM to stock prediction with ample scope for enhancements to prediction accuracy. K. A. Althelaya, E.

M. El-Alfy and S. Mohammed any contributed to the sector by staging experiments and simulations to assess the practicability of applying deep learning techniques to prediction of stock prices.

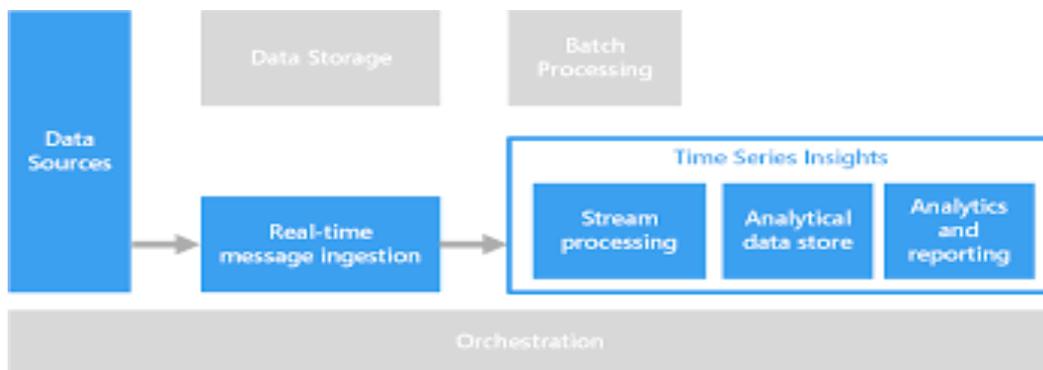
III. METHODOLOGY

Stock market prediction looks a posh downside as a result of there are several factors that have nonetheless to be self addressed and it doesn't appear applied mathematics at first. however by correct use of machine learning techniques, one will relate previous data to the present data and train the machine to be told from it and build acceptable assumptions. The information set being utilised for analysis was picked up from Yahoo Finance. The dataset consisted of roughly 9 lakh records of the desired stock costs and alternative relevant values. The data mirrored the stock prices at bound time intervals for every day of the year. It consisted of assorted sections specifically date, image, open, close, low, high and volume. For the aim of simulation and analysis, the information for under one company was considered. All the data was on the market during a file of csv format that was initial scan and reworked into a data-frame mistreatment the Pandas library in Python. From this, the data for one explicit company was extracted by segregating data on the premise of the symbol field. Following this normalisation of the data was performed through usage of the sklearn library in Python and therefore the data was divided into training and testing sets. The take a look at set was unbroken as 20% of the on the market dataset.

Although machine learning in and of itself has several models however this paper focuses on two of the foremost necessary amongst them and created the predictions mistreatment these.

3.1 Time Series Based Model

A time series is **an observation from the sequence of discrete-time of successive intervals**. A timeseries is a running chart. The time variable/feature is the independent variable and supports the target variable to predict the results.



ARIMA, short for 'Auto Regressive Integrated Moving Average' is actually a class of models that 'explains' a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that equation can be used to forecast future values.

3.2 Long Short Term Memory (LSTM) Network Based Model

LSTM is that the advanced version of Recurrent-Neural- Networks (RNN) wherever the data belonging to previous state persists. These are completely different from RNNs as they involve long run dependencies and RNNs works on finding the connection between the recent and also the current information. this means that the interval of knowledge is comparatively smaller than that to LSTM.

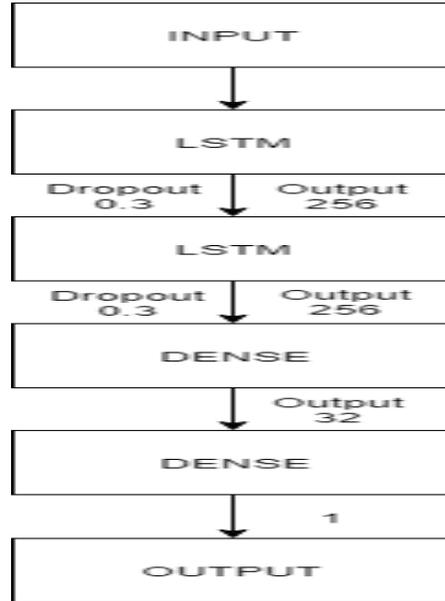
The most purpose behind victimization this model available market prediction is that the predictions depends on massive amounts of data and are usually addicted to the long term history of the market. Therefore LSTM regulates error by giving an aid to the RNNs through holding info for older stages creating the prediction additional correct .so proving itself the maximum amount more reliable compared to alternative methods.

Since stock exchange involves process of giant data, the gradients with regard to the load matrix may become terribly tiny and will degrade the training rate.

This corresponds to the matter of Vanishing Gradient. LSTM prevents this from happening. The LSTM consists of a memory cell, input gate, output gate and a forget gate. The cell remembers the worth for long term propagation and therefore the gates regulate them. In this paper, a sequential model has been created that involves stacking two LSTM layers on top of each other with the output value of 256. The input to the layer is within the kind of two layer [0] and layer[1]. A dropout



value of 0.3 has been mounted which suggests that 0.3out of total nodes can be frozen throughout the training process to avoid over-fitting of data and increase the speed of the training process. At last, the core dense layer wherever every neuron is connected to every other within the next layer is further providing input of 32 parameters to future core layer which supplies output as 1. The model is compiled with amean sq. price operate to take care of the error throughout the method and accuracy is chosen as a metric for the prediction.



IV. EXPERIMENTAL RESULTS

The planned system is trained and tested over the dataset taken from Yahoo Finance. it's split into coaching and testing sets severally and yields the subsequent results upon passing through the various models.

4.1 Time Series Based Model Results

The plot in figure3 is the result of application of Time Series algorithm on the dataset to predictvarying prices with respect to the time

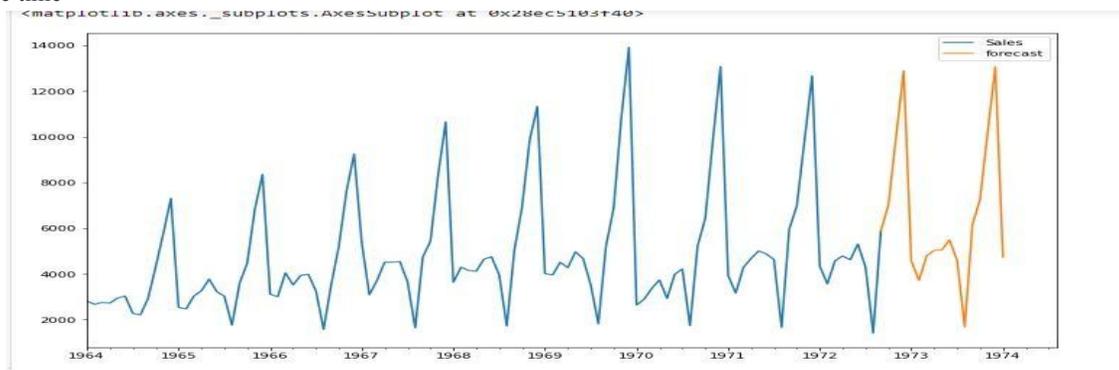


Figure 3: Plot between Price and Date Using ARIMA Model

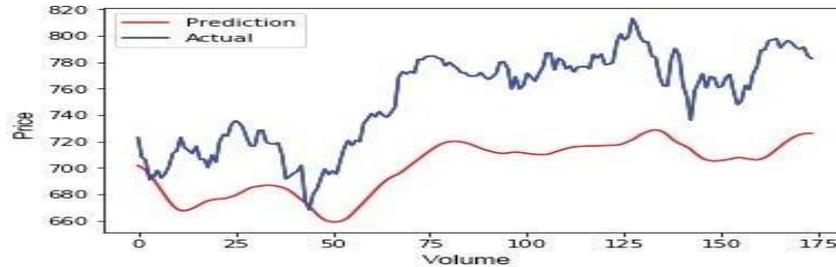
The above graph Fig. 3 is plot over the data having batch size 512 and 90 epochs. The R-square confidence test resulted in a confidence score of 0.86625.

4.2 LSTM Based Model Results

The prediction is shown by red line and therefore the actual trend shown by blue. The proximity of these two lines tells, however economical the LSTM based mostly model is. The prediction approximates real trend once a substantial quantity



of time has passed. The model resulted during a Train Score of 0.00106 MSE (0.03 RMSE) and a check Score of 0.00875 MSE (0.09 RMSE).The lot of the system is trained and the larger the scale of the dataset utilised the greater the accuracy which is able to be attained. The LSTM Model offered more accuracy than the Regression based Model.



V. CONCLUSION

This paper was a shot to determine the long run costs of the stocks of a corporation with bigger accuracy and dependability exploitation machine learning techniques. the first contribution of these researchers being the appliance of the novel LSTM Model as a way of decisive the stock prices. Each the techniques have shown an improvement within the accuracy of predictions, thereby yielding positive results with the LSTM model proving to be additional efficient. The results are quite promising and has led to the conclusion that it's doable to predict stock exchange with more accuracy and potency using machine learning techniques. Within the future, the accuracy of the stock exchange prediction system will be more improved by utilizing a far larger dataset than the one being utilised currently. Furthermore, different rising models of Machine Learning might even be studied to examine for the accuracy rate resulted by them. Sentiment analysis although Machine Learning on however news affects the stock costs of a corporation is additionally a very promising area. different deep learning primarily based models can even be used for prediction purposes

REFERENCES

- [1]. Dash, R., & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. *The Journal of Finance and Data Science*, 2(1), 42-57.
- [2]. Kumar, L., Pandey, A., Shrivastava, S., & Darbari, M. (2011). A hybrid machine learning system for stock market forecasting. *Journal of International Technology and Information Management*, 20(1), 3.
- [3]. Kim, K. J., & Lee, W. B. (2004). Stock market prediction using artificial neural networks with optimal feature transformation. *Neural computing & applications*, 13(3), 255-260.
- [4]. Kim, M. J., Min, S. H., & Han, I. (2006). An evolutionary approach to the combination of multiple classifiers to predict a stock price index. *Expert Systems with Applications*, 31(2), 241-247.
- [5]. Liao, Z., & Wang, J. (2010). Forecasting model of global stock index by stochastic time effective neural network. *Expert Systems with Applications*, 37(1), 834-841.