Stock market analysis and prediction using LSTM models

Rajrishi Sarkar, MEMS Dept., IIT-Bombay, EE769, 203110061@iitb.ac.in

Abstract: *The paper aims to predict the stock prices of a certain company after analysing the past performance of the stocks. The current trend in the stock market is used to predict the upcoming stock performances using machine learning. This paper uses LSTM based machine learning models to predict the same. The dataset used is "**New York stock exchange**", available in Kaggle.*

1.INTRODUCTION

Prediction of stock prices is hugely profitable to brokers and stock market investors. It is seen that such prediction is possible through careful investigation of the history of the stock market of the desired company one is interested in. Machine learning and deep learning is an efficient way to do the same. The main part of the thesis is the data used in training the model. It should be concrete, even minute deviations can cause massive changes in the output. In this project the dataset used is taken from Kaggle, which uses yahoo finance as it's source. There are certain indexes which one should be aware of when dealing with stock-based datasets. Close, open, high and low are various bidding prices at different times. Volume is defined as the number of shares passed from one owner to another during a particular time period.[1]

Several models can be used, regression based models primarily focuses on error reduction where as the LSTM remembers the past data which is used in the long run. Other RNN models have instant memory and[3] is not suitable for time series analysis, especially stock market prediction. In this project we sample the "APPLE" stocks and perform exploratory data analysis on it. We visualize the stock's performance over the years. After that LSTM based model is used.

2.PREVIOUS WORKS

It's obvious that stock market is a complex problem since there are many factors that are yet to be addressed. Previous works are reviewed and it's seen that support vector machines, artificial neural networks, genetic algorithms and several AI based techniques are used for stock market predictions. LSTM is identified as the best approach for stock market analysis and predictions because using a Long Short-Term Memory (LSTM) Model to Predict Stock Prices[6]. LSTMs are very powerful because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price.

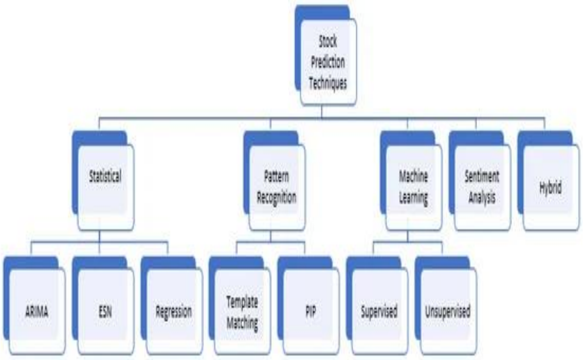

Fig1. Different stock prediction techniques.

3.DATASETS

The dataset for the project is taken from Kaggle. Prices were fetched from Yahoo Finance, fundamentals are from Nasdaq Financials, extended by some fields from EDGAR SEC databases. Dataset consists of following files:

a)*prices.csv:* raw, as-is daily prices. Most of data spans from 2010 to the end 2016, for companies new on stock market date range is shorter. There have been approx. 140 stock splits in that time, this set doesn't account for that.

b)*prices-split-adjusted.csv*: same as prices, but there have been added adjustments for splits.

c)*securities.csv:* general description of each company with division on sectors

d)*fundamentals.csv*: metrics extracted from annual SEC 10K fillings (2012-2016), should be enough to derive most of popular fundamental indicators.

There are not much challenges involved when it comes to stock market datasets because a massive chunk of data can be obtained from finance based websites like Yahoo finance, NASDAQ and also in data science platforms like Kaggle.

4.PROCEDURE AND EXPERIMENTS

After data collection and using pandas library to create a data-frame, when checked for missing values, none are found. EDA is performed on the dataset. For statistical analysis[5] scipy library and for visualizations matplotlib and seaborn libraries are used. The dataset consists data of 501 companies. Since we are interested in the stocks of "APPLE" we resample our data so that it consists of only apple's stocks. Opening, closing prices and stock volume are observed.

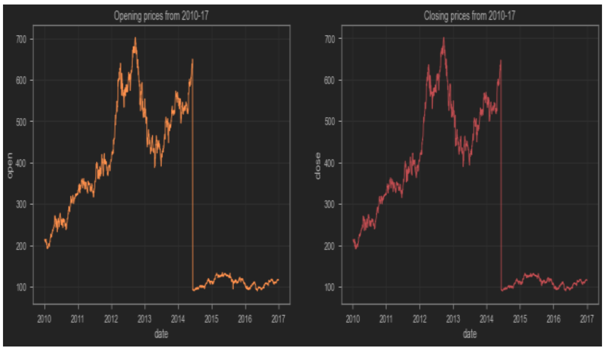

Fig2.Opening and closing prices of Apple stocks from 2010 to 2017.

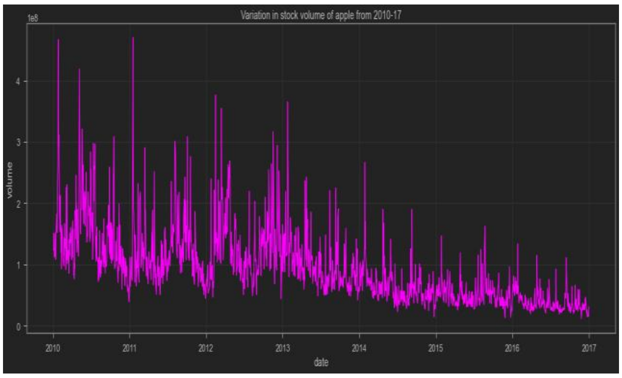

Fig3.Volume of stocks from 2010 to 2017.

It can be seen that over the period of 7 years the stock prices and volume deteriorated by a significant extent. Statistical analysis of the variables are done based on their distribution plots, probability plots and boxplots.

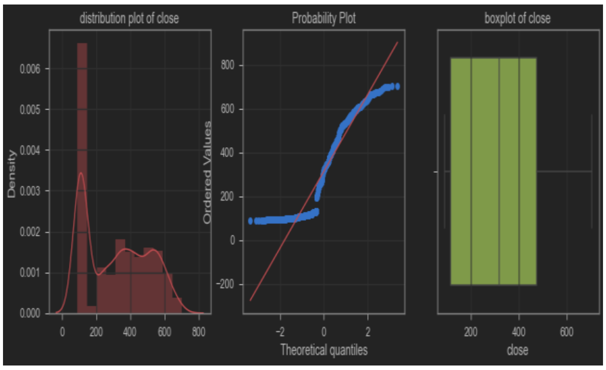

Fig4.Statistical analysis of the target variable("close") from left, distribution plot, probability plot and boxplot.

Closing prices of the stock is taken as the target variable, as we are interested in predicting the future closing prices. After exploratory data analysis feature scaling is performed, standard-scaler is used to do so. The data is created for feeding into the LSTM model.

LSTM is basically advanced version of RNN. The major advantage of LSTM over RNN is that the information belonging to the previous states persists for a longer time. RNN focuses on finding the relationship between recent and curre3nt information[7]. Predictions are usually dependant on long term history of the market, therefore LSTM regulates the error by providing aid to RNNs by retaining past information[8]. As stock market involves processing huge amount of data, the gradients w.r.t the weight matrix gradually becomes very small and therefore degrades the learning rate of the system. This is known as the vanishing gradient problem. LSTM resists this from happening. It consists of 3 gates input, output and forget gate and a remembering cell.

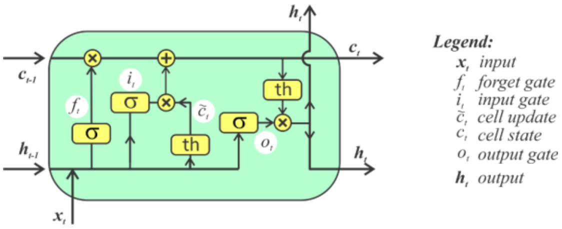

Fig6.Architecture of LSTM.

In this paper a sequential model is prepared that involves stacking of 3 LSTM layers. A dropout value of 0.2 has been fixed. This means that 0.2 nodes out of the total nodes will be frozen during the model training to avoid overfitting of the model. Finally dense layer is added which gives output as 1. The model is compiled with the help of mean square cost function. R2 and MSE are the chosen metrics for evaluation.

5.RESULTS

Training and validation losses are plotted w.r.t the epochs after training the model with 100 epochs.

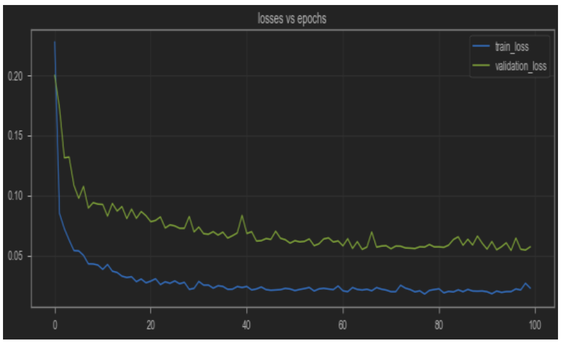

Fig7.Plot of training and validation losses w.r.t the number of epochs.

The model shows overfitting since validation loss > training loss. This particular model gave a R2 value of 0.9, & a mean squared error of 0.06.

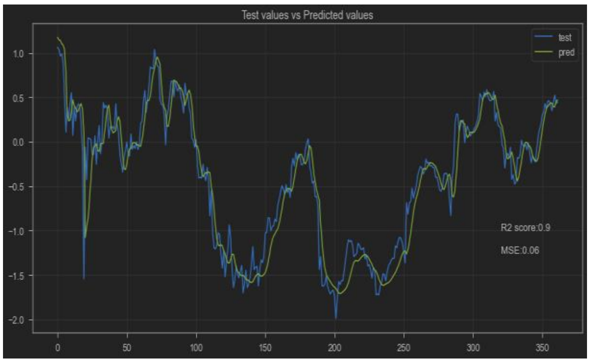

Fig8.Plot showing the actual and LSTM predicted values.

Another model is created by reducing the LSTM layers from 3 to 2. Dropout values and other parameters are not changed. Overfitting is reduced to some extent and the R2 value shot to 0.92, having MSE of 0.05.

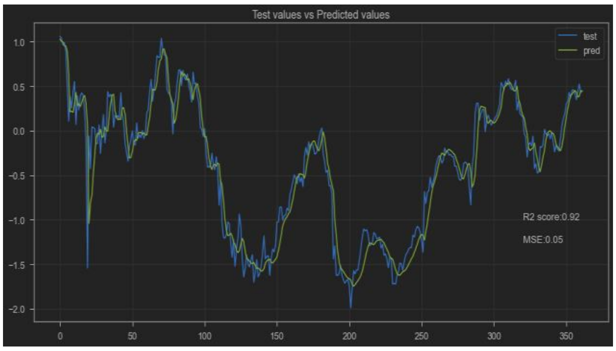

Fig8.Plot showing the actual and LSTM predicted values.

6.CONCLUSIONS

LSTM based models can be used to predict stock prices of the market. It's accuracy and efficiency is impressible. One must take into account that there is good chance of overfitting and therefore take suitable measures. Hyper-tuning the model by varying the batch sizes, number of LSTM layers can be used to increase the accuracy off the model.

7.REFERENCES

[1] M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "*Stock market prediction using machine learning techniques*," 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), Kuala Lumpur, 2016, pp. 322-327.

[2] K. Raza, "*Prediction of Stock Market performance by using machine learning techniques*," 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT), Karachi, 2017, pp. 1-1.

[3] H. Gunduz, Z. Cataltepe and Y. Yaslan, "*Stock market direction prediction using deep neural networks*," 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, 2017, pp. 1-4.

[4] M. Billah, S. Waheed and A. Hanifa, "*Stock market prediction using an improved training algorithm of neural network*," 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE), Rajshahi, 2016, pp. 1-4.

[5] H. L. Siew and M. J. Nordin, "*Regression techniques for the prediction of stock price trend*," 2012 International Conference on Statistics in Science, Business and Engineering (ICSSBE), Langkawi, 2012, pp. 1-5.

[6] K. V. Sujatha and S. M. Sundaram, "*Stock index prediction using regression and neural network models under non normal conditions*," INTERACT-2010, Chennai, 2010, pp. 59-63.

[7] S. Liu, G. Liao and Y. Ding, "*Stock transaction prediction modelling and analysis based on LSTM*," 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA), Wuhan, 2018, pp. 2787-2790.

[8] T. Gao, Y. Chai and Y. Liu, "*Applying long short term memory neural networks for predicting stock closing price*," 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, 2017, pp. 575-578.

[9] 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), Irbid, 2018, pp. 151-156