A RESEARCH REVIEW

on

Apple Stock Price Prediction Using Machine Learning

GALGOTIAS UNIVERSITY

(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

Submitted in partial fulfillment of the

requirement for the award of the degree of

Master of Computer Applications

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING /

DEPARTMENT OF COMPUTER APPLICATION

GALGOTIAS UNIVERSITY, GREATER NOIDA

INDIA

Under The Supervision Of

Dr. Bharat Bhushan Naib

Submitted By

Sarvesh Kr. Singh (20SCSE2030046)

Trideev Ganguly (20SCE2030048)

DECEMBER 2021

TABLE OF CONTENTS

ABSTRACT

Predicting the stock market's performance is a challenging task. Physical vs. psychological causes, rational vs. irrational behaviour, and so on are all components in the prediction. All of these factors combine to make stock values extremely volatile and difficult to anticipate accurately. We will use historical data on the stock prices of an **Apple** company in this study. We'll use a combination of machine learning algorithms to forecast this company's future stock price, starting with simple algorithms like averaging and regression analysis and progressing to more complicated techniques like **LSTM**.

Submitted By: Supervisor:

Saurabh Kr. singh

INTRODUCTION

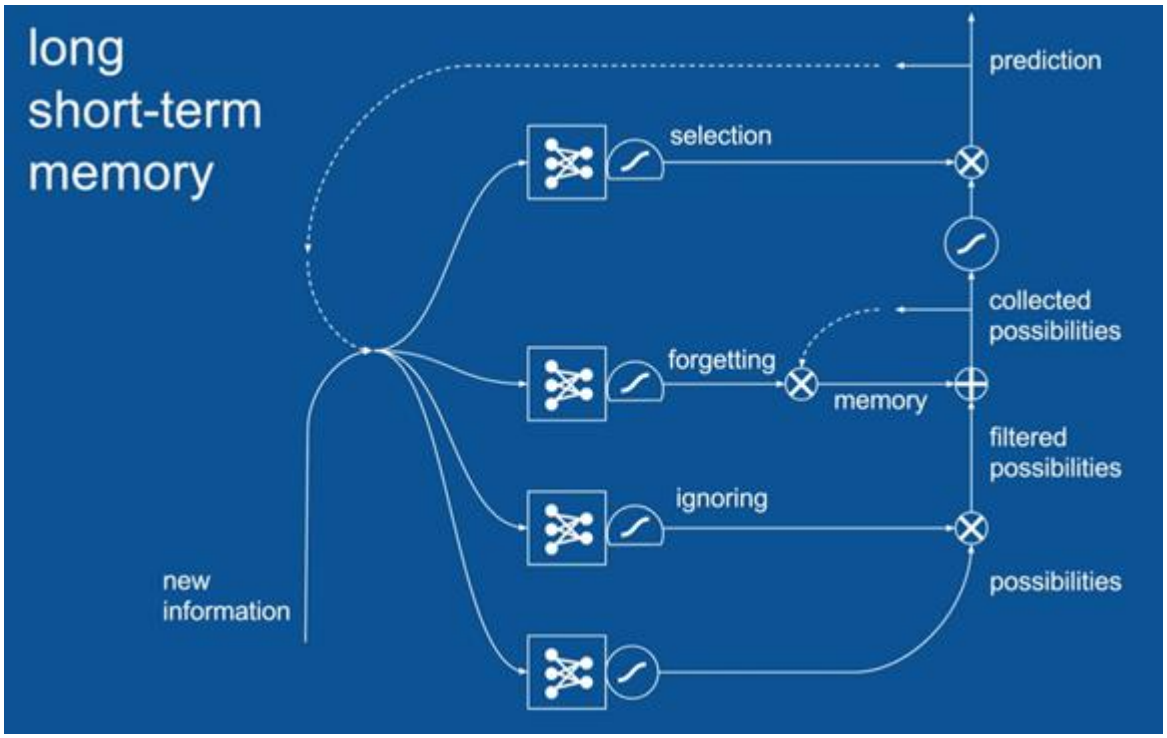
Recurrent Neural Networks are hardly used today. Although all recurrent neural networks feature feedback loops in their recurrent layer, which allows them to keep information in 'memory,' training them is problematic for RNNs that must learn long-term temporal connections. The vanishing gradient problem occurs because the gradient of the specified loss function decays exponentially with time.

Memory cells that keep information in memory for long periods of time are included in **LSTM** networks in addition to conventional units. Multiple gates are used in this model to govern when information enters the memory, when it is outputted, and when it is forgotten, boosting the network's ability to learn longer-term dependencies

LITERATURE REVIEW

The RNN (Recurrent Neural Network) LSTM (Long Short Term Memory) can be used to predict stock values using historical data. Because it retains an internal state to keep track of the data it has already seen, LSTM is well suited to modelling sequence data. Time series analysis and natural language processing are two common uses of LSTMs.

LSTM networks have hidden states and cell states that can delete and add information with gate regulation, but vanilla RNN networks simply have hidden states for memory.



In terms of stock predictions, the vanilla RNN would forget early-stage stock price data as the years passed, whereas the **LSTM** could use historical trends and data to make predictions that are more precise. The memory gate gathers potential outputs and stores the ones that are relevant; the selection gate chooses the final output from the memory gate's possible outputs; and the forget and ignore gate decides which data memories aren't relevant and discards them.

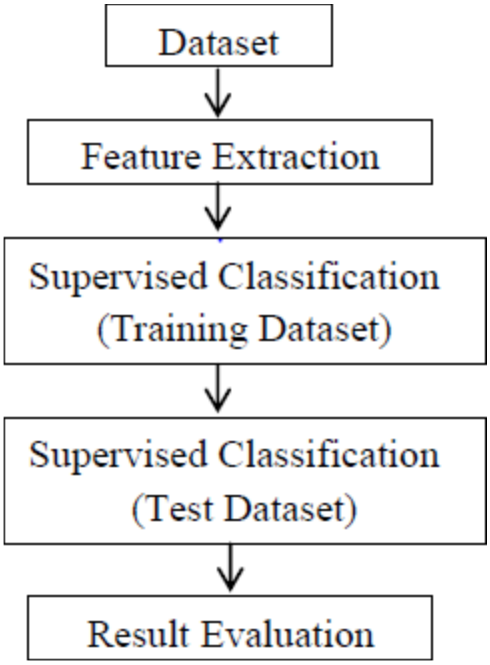
METHODOLOGY

Using LSTM layers, we will attempt to forecast the price at time t using historical prices from the previous 90 days (i.e. from t-90 to t-1). Please note that we will attempt to capture the trend rather than exact costs. The input must be a 3D tensor with shape (batch size, timesteps, input dim) for LSTM to work.

The number of timesteps is 90 since we are analysing 90 days of past data (t-90 to t-1) to create a forecast at time t. To make the prediction, we only utilise the "Adj Close" price, therefore the input dim is 1. Additional features can be added to

improve input dim. The value of a competitor's shares, for example, may have an impact on Apple's stock price. If we do

the prediction with a second variable, the input dim will be 2. The **batch_size** specifies how many observations should be fed into the LSTM layers before the weights are updated.



THE SYSTEM FLOW

Due to the extensive computations performed in a neural network, it is also preferable to normalise the data. Normalization can be accomplished with a few simple mathematical procedures or by utilising pre-defined library functions. I'll use scikit-MinMaxScaler. learn's It normalises numbers into a range of [0,1] by default.

We're working on an assignment that requires supervised learning. We'll use some data to train the model and then use previously unseen data to test its performance. As a result, we must divide the data into training and test sets. The

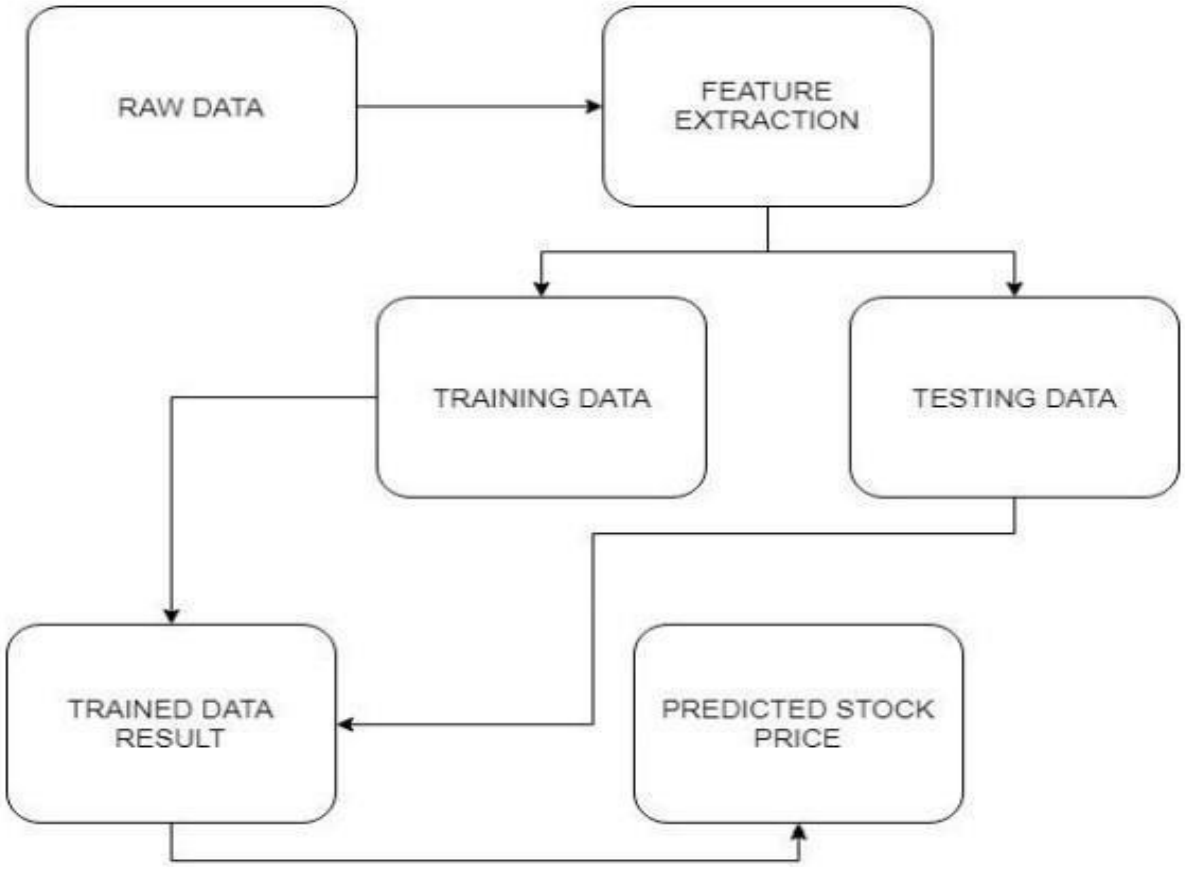
model should have no knowledge of the data in the test set in order to obtain reliable performance results. When normalising the data, we should keep this in mind as well. First, we must divide the data. Then, separately, apply normalisation.

Many people have dedicated their time and effort to bringing global trade closer to people and making it more dependable in order to carry out resources and make their lifestyles more intentional than before. Over the last few years, various approaches and ideas have been devised and executed, and the topic is still a point of research where people are coming up with answers.

Human intellect fascinates humanity, and putting it in a machine and integrating it is a hot research issue. The same scientific project is being worked on by several persons. ASheta developed TS, a fuzzy set model, after testing its invention on two nonlinear processes. Existing models are unable to cope with the vulnerabilities and delete the rarest data that they cannot interpret, resulting in severe data loss and a forecasting issue.

The importance of observation in resource and forecast management cannot be overstated. The point of time estimation is affected if the result cannot be seen, making it less reliable in the market. The current technology does not allow for this type of monitoring. The existing system in stock market predictions looks to be biassed because it uses a single source point for data. Before the data set is anticipated, a simple data retrieval should be created and tested on the training data set, which is more adaptable and versatile.

Because the stock varies every day and the loss margin can increase with time, the current system has a significant problem



with sight loss. An initial instance is utilised for prediction.

THE DATA FLOW

SOFTWARE USED

Jupyter Notebook is a web-based interactive computational tool for creating Jupiter Notebook documents, also known as IPython Notebook. The term "notebook" refers to a large entity that encapsulates the integration of many different entity sets. The basic document format for the execution is JSON, which is dependent on the schema as well as input and output methods. It has a high level of integration with a number of different language sets and provides a number of different possibilities.

This is done via the ".ipynb" extension, which is supported by this platform. It's an open-source piece of software that enables interactive communication. For this, it has its own set of open standards. It's a friendly environment that's perfect for ambitious programmers.

The same's versatility is phenomenal, and its setup and integration are simple and straightforward, resulting in no prior distortion, and its efficiency can be examined across any system of choice. It is one of the most popular software packages, with widespread use for product design and development as well as substantial support.

Not only that, but it also ensures the scalability of the code and its deployment. Various languages can be utilised to complete the project, and the project can be completed in any of them. The notebook files that are created can be shared and saved for later use in a variety of ways. It enables the production of interactive and evolved output sets. The elements can be graphed, charted, and shown in a variety of ways. The same data integration is at its most effective level.

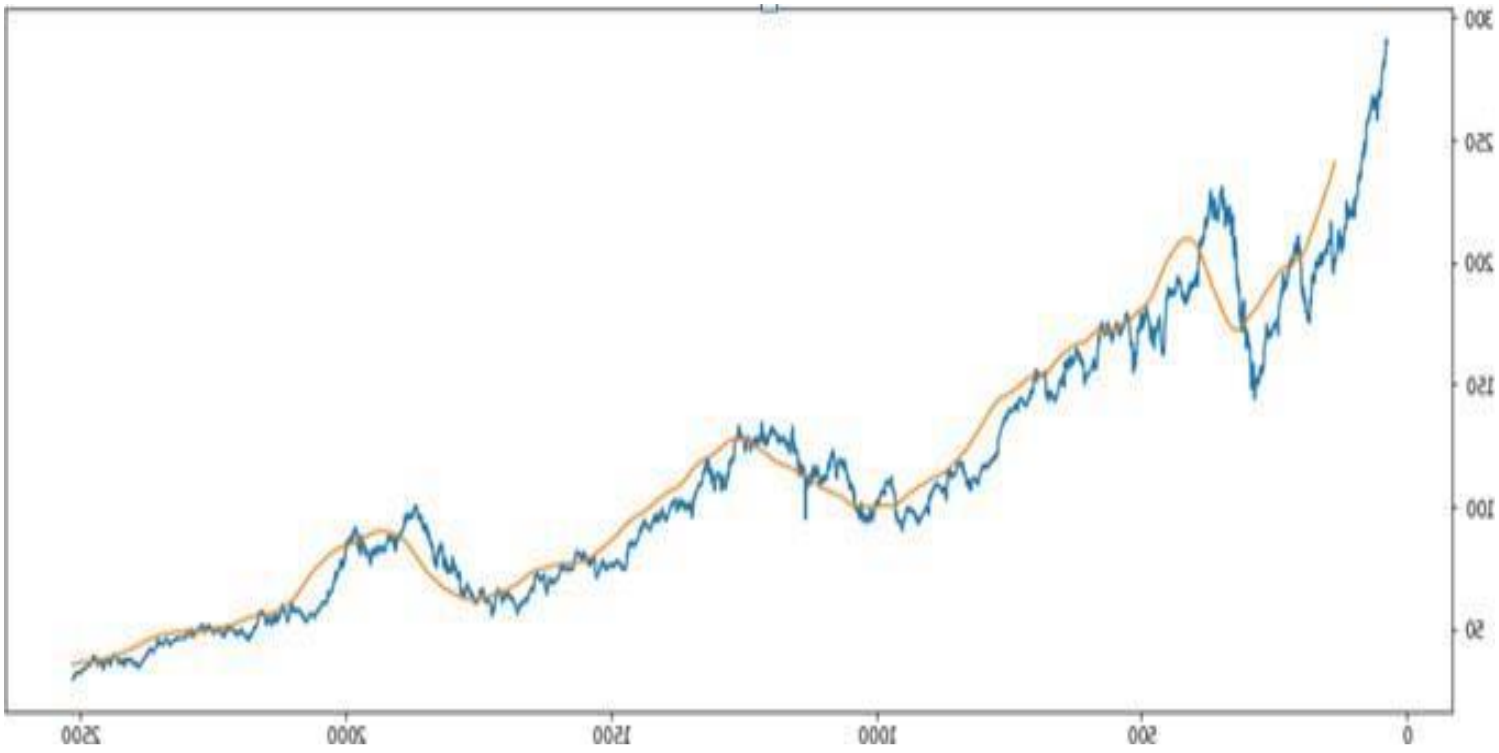
RESULT & DISCUSSION

The suggested LSTM model is written in Python and uses historical data to forecast the future price of Apple stock. The figure below shows a representation of Apple stock forecasting. The construction of an algorithm that forecasts a stock's price for a given period of time; the graph below depicts the projected price of an apple stock based on our method.

The graph below shows the plotted outcome of our technique employing 96 LSTM units to attain precision.

There are numerous strategies to try to improve this model's performance. You can improve the outcome by adjusting:

- Count of layers
- A layer's node count is the total number of nodes in the layer.
- The total number of timesteps
- Ratio of dropouts
- The total number of epochs
- Size of the batch



Predicted Stock Price

The approximate value is taken into account, and the hit, profit, or gain rate for the same is calculated.

CONCLUSION

To summarise, stock is an unpredictably unexpected system that follows chain parts and their interdependencies. It's characterised as a curve that changes over time, shifting the price from low to high and vice versa.

Because the level of integration is higher with other dependencies, leaving one out compromises the level of precision. Accuracy is not a phrase used in the stock market since actual forecasting for any fiscal day is impossible because the market is always shifting and turning the tables day and night. With more component assets and dependencies, it becomes more possible and adaptable, making it even more difficult to anticipate.

SCOPE FOR FUTURE STUDY

The stock market is the finest way for a business to expand, and it also provides a side income for those who are willing to invest and profit from it. Since then, the term "stock" has been in use, and it continues to expand in popularity. Thousands of investors have put their money into it and profited handsomely.

Middle-level agents and stock sellers are among those who study and invest in the same. The expense of the stock consultation is substantial and costly. So, when it comes to individuals, they think a lot and invest a lot, but there's no guarantee that the same would deliver a profitable result.

As a result, stocks are more volatile than ever, and the likelihood of their growth is higher than ever. If the stock market and its forecasting can be done correctly, it will be beneficial to both individuals and organisations. The risk factor must be reduced so that the system's efficiency may be maximised and individuals can be confident in their long-term investment.

REFERENCES

[1] K. Senthamarai Kannan, P. Sailapathi Sekar, M.Mohamed Sathik and P. Arumugam, “Financial stock market forecastusing data mining Techniques", 2010, Proceedings of the international multiconference of engineers and computerscientists.

[2] Tiffany Hui-Kuang yu and Kun-Huang Huarng, “A Neural network-based fuzzy time series model to improveforecasting”, Elsevier, 2010, pp: 3366-3372.

[3] Md. Rafiul Hassan and Baikunth Nath, “Stock Market forecasting using Hidden Markov Model: A New Approach”, Proceeding of the 2005 5th International conference on intelligent Systems Design and Application 0-7695-2286-06/05, IEEE 2005.

[4] Bonde, Ganesh, and Rasheed Khaled. “Extracting the best features for predicting stock prices using machine learn- ing.” Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), 2012.

[5] P. Hajek, Forecasting Stock Market Trend using Prototype Generation Classifiers, WSEAS Transactions on Systems, Vol.11, No. 12, pp. 671-80, 2012.

[6] Hagenau, Michael, Michael Liebmann, Markus Hedwig, and Dirk Neumann. "Automated news reading: Stock price prediction based on financial news using context- specific features." InSystem Science (HICSS), 2012 45th Hawaii International Conference on, pp. 1040-1049. IEEE, 2012.

[7] Kyoung-jae Kim, Ingoo Han. “Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index”. Expert Systems with Applications, Volume 19, Issue 2, 2000, Pages 125-132, ISSN 0957-4174.

[8] Leung, Carson Kai-Sang, Richard Kyle MacKinnon, and Yang Wang. “A machine learning approach for stock price prediction.” Proceedings of the 18th International Database Engineering & Applications Symposium. ACM, 2014.

Bonde, Ganesh, and Rasheed Khaled. “Extracting the best features for predicting stock prices using machine learn- ing.” Proceedings on the International Conference on Artificial Intelligence (ICAI). The Steering Committee of The World Congress in Computer