Bachelor of Science in Computer Science And Engineering

Machine Learning Approach for Stock Market Predection

Authors

Mostafa Al Mamun Fahim (20101366)

Annesha Karim (20101048)

Rakibuzzaman Rayhan(18101082)

Date: 21/12/22

1 Abstract

Forecasting stock prices is a challenging endeavor that requires an analytical basis to forecast future share prices. It is impossible to foresee the movement of stock prices due to the inherent association between them that the market's nature provides. Using machine learning techniques, such as a recurrent neural network termed Long Short Term Memory, and stochastic gradient descent to modify the weights of each data point, the proposed method predicts share prices. This strategy for predicting stock prices will generate more accurate findings than others. To produce the best visual results, the network is trained and tested with data of varying sizes.

2 Introduction

The stock exchange is where shares of publicly traded firms are traded. As we have shown, the volatility of the stock market requires thorough examination utilizing historical data. In the past, algorithms used historical time series data for stocks to forecast future stock patterns. Standard scientific methodologies for predicting stock prices need quantitative analysis of historical stock data. The objective of this project is to construct a program that can predict future stock prices using training sets comprised of historical data and prices for a certain share.

This model employs an RNN (Recurrent) approach called Long Short Term Memory to analyze past changes in a company's share price (LSTM). Predictions on a certain characteristic can be made using the suggested method since it takes into account all of the available prior data on a share. Opening price, High, Low, previous day's o price, Close price, Date of trade, Total Trade Quantity, and Turnover are all aspects of shares. In order to forecast a stock price over a specified period of time, the suggested model employs time series analysis. Microsoft is one of the companies that will be considered for the project. The YFinance, a YahooFinance exchange business, was one of the first in the world to offer cutting-edge trading technology to investors all over the world.

Long Short Term Memory (LSTM) networks are recurrent neural networks (RNNs) that can solve linear problems. LSTM is a deep learning method. Long-term Memory (LSTM) Units are compelled to learn extremely long sequences. This is a broader version of the gated recurrent system. Because LSTMs address the evanescent gradient issue, they are more benign than other deep learning methods such as RNN or traditional feed forward.

As a final step, the actual and predicted price variations of the LSTM-based model are graphed alongside the price changes over time.

3 Methodology

Predicting the stock market seems to be a challenging topic since there are several ignored factors and it first does not look statistical. Nevertheless, with the proper use of machine learning techniques, it is feasible to integrate past data with current data and teach the computer to learn from it and make appropriate assumptions. This study makes predictions using the most significant LSTM model.

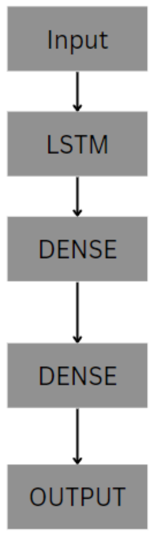

Figure 1: Proposed workflow

[h]

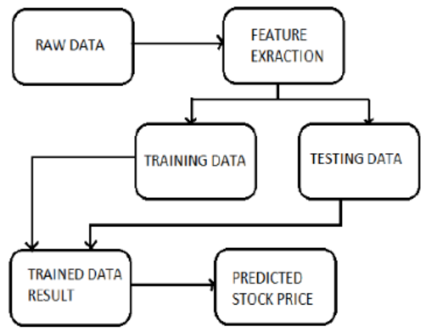

Figure 2: Proposed System Architecture

Our data collection is comprised of the following five components: the open, the close, the low, and the high values, and the volume. In order to generate an accurate prediction of the stock price, we will utilize these factors. This location contains data sets that cover the time period from 1986 to 2020. At addition to that, we have over 9000 different instances in one location.

An improved version of recurrent neural networks (RNN), the long short-term memory (LSTM) architecture, may remember data from a prior state. RNNs, on the other hand, are mostly interested in figuring out how recent and current input relate to one another. These involve enduring reliances. As a result, it appears that the information interval is significantly faster than LSTM. The main justification for using this model to forecast the stock market is that these projections are mostly based on the long-term history of the market and rely on enormous amounts of data. By helping RNNs retain information from earlier stages, LSTM manages error in this manner, ultimately leading to an increase in prediction accuracy.

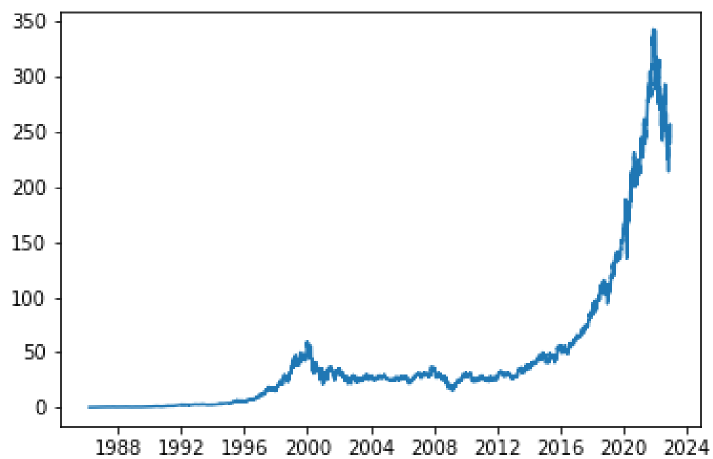

Figure 3: Plotting of the dataset (since: 1986-2020)

Date, Open, High, Low, Close, Adj Close, and Volume comprise the columns of our data set. Unfortunately, we must locate a precise stock price on a specific day. Therefore, we can utilize the date as an index and establish a cutoff time relative to that number (stock price). This is the reason why we skipped additional indexes.

A cell, an information door, an entrance door, and a view door can all be regarded as fundamental LSTM module elements. The user specifies the intervals at which the cell collects values, and inputs 1, 2, and 3 control how data enters and leaves the cell. The LSTM's main advantage is its ability to learn temporal dependence within a specific environment. Each LSTM unit gathers either long- or short-term data without explicitly applying the activation function within the recurrent components (thus the name). Due to the huge volume of data processed on the stock market, the gradients of the weight matrix may become small, hence slowing the system's rate of learning. This is identical to the issue with gradients that disappear. LSTM can, thankfully, prevent this from occurring. A forget gate, an input gate, an output gate, and a remembering cell comprise the LSTM. The cell stores values for release at a later time, and gates restrict their release. In this paper, we provide a sequential model that uses an LSTM layer to generate 64-bit output. Finally, we add a dense core layer in which every neuron is connected to every other neuron in the layer above, yielding 32 parameters as input to the subsequent dense core layer, which yields a single output value of 1. Accuracy is selected as the prediction metric, and a model is developed using a mean-square cost function to maintain constant error.

List of parameters used in the dataset follows as below:

- Date: Date of stock price.
- Open: Open price of share.
- Close: Closing price of share.
- Volume: Numbers of shares traded.
- Adj Close: Adjusted closing price of share.

Algorithm: Stock price prediction using LSTM.

Input: Historic stock data.

Output: prediction of stock price using price variation.

***Step*1**: Start.

***Step*2**: Data Preprocessing after getting the historic data from the market for a particular share.

***Step*3**: import the dataset to the data structure and read the open price.

***Step*4**: do a feature scaling on the data so that the data values will vary from 0 and 1.

***Step*5**: Creating a data structure with 60 timestamps and 1 output.

***Step*6**: Building the RNN (Recurrent neural network) for Step 5 data set and Initialize the RNN by using sequential repressor.

***Step*7**: Adding the first LSTM layer and some Dropout regularization for removing unwanted values.

***Step*8**: Adding the output layer.

***Step*9**: Compiling the RNN by adding ADAM optimization and the loss as mean absolute error.

***Step*10**: Making the predictions and visualizing the results using plotting techniques.

The suggested system uses long short-term memory and online learning to estimate stock values in the near future (LSTM). Long Short-Term Memory (LSTM) is a technique used in deep learning that simulates an intermittent neural network (RNN). LSTM is capable of forming associations between its inputs, unlike typical feed forward neural networks. In addition to being concerned with full information layouts, the approach is also concerned with relatively restricted information (For example, a speech or a video). LSTM is appropriate for applications such as speech recognition, handwriting identification without partitioning, and anomaly detection in structured traffic and intrusion detection systems.

Before doing any study, it is essential to collect market information. Importing data from a market clearing house such as YFinance and collecting data from a market clearing house are the first phases in our proposed framework. The dataset that will be used to market expectations must be examined from multiple perspectives. Another effective method for enhancing the dataset is to collect additional external data. The majority of our data consists of stock values from the previous year. The data is subsequently preprocessed, an important development in information mining consisting of reformatting unstructured data into a simpler format. It is a given that information acquired from the original source will be inaccurate, insufficient, and inconsistent. During the preprocessing phase, the data will be cleansed, followed by the use of highlights scaling to highlight the variables. Cross-approval, which is a well-founded, projected execution of the model using the preparation data, is part of the model preparation process. The purpose of training is to strengthen the models and improve the calculation itself. We ensure that our test data is immaculate, as we think that model decisions should never be based on concealed or insufficient information. Include the price of the offer in the data expansion. Using a data visualization methodology, we display the output of our method and the range of potential possibilities.

4 Results and Discussion

Python implementation of the suggested LSTM model, which uses past data to forecast Microsoft Corporation's share price in the future. The depiction of prediction is shown in the figure below. The graph below shows the anticipated share price of Microsoft Corp.; in our work, we describe the creation of an algorithm that forecasts the stock price of a specific share over a given period of time.

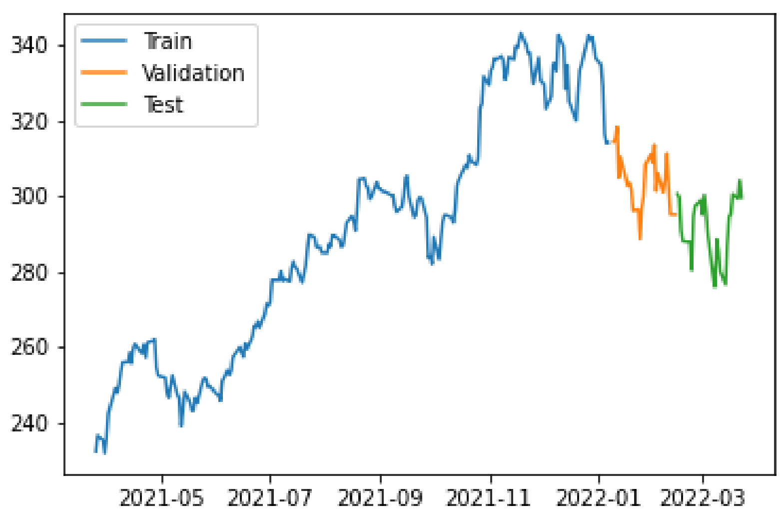

Figure 4: Training and Testing data.

After doing all of our studies, we painstakingly duplicated every graph for confirmation to assure the veracity of our findings. Therefore, both the training graph and the validation graph are displayed. By increasing the magnification, you can observe how closely these values are packed. If we applied the same methodology to the tests, we would observe that, when plotted on the same graph, the tests also tend to form clusters. Having demonstrated our ability to develop long-term projections, we can assume with confidence that the values or forecasts we generated three days ago were accurate.

Consequently, we generated the forecast using the three preceding

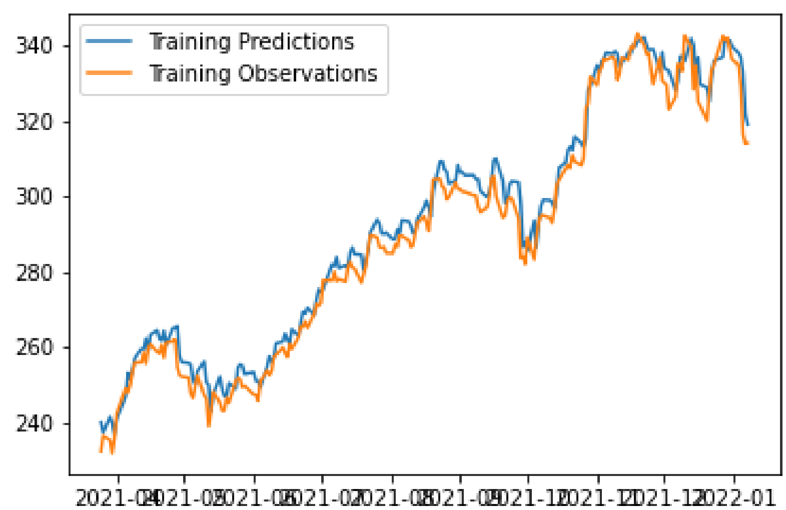

Figure 5: Observation and Prediction.

days and then applied it the following day, when we got the actual three days in question.

Under the premise that this is the only data available, training will proceed in this environment, and predictions will be created using a backpropagation technique. and see what it says; hence, we will execute a "from copy import deep copy," generate a list, and then begin constructing "recursive predictions" to create that function.

As a result, we have a T-shaped graph similar to the one we had previously, but this time we have also included the recursive dates

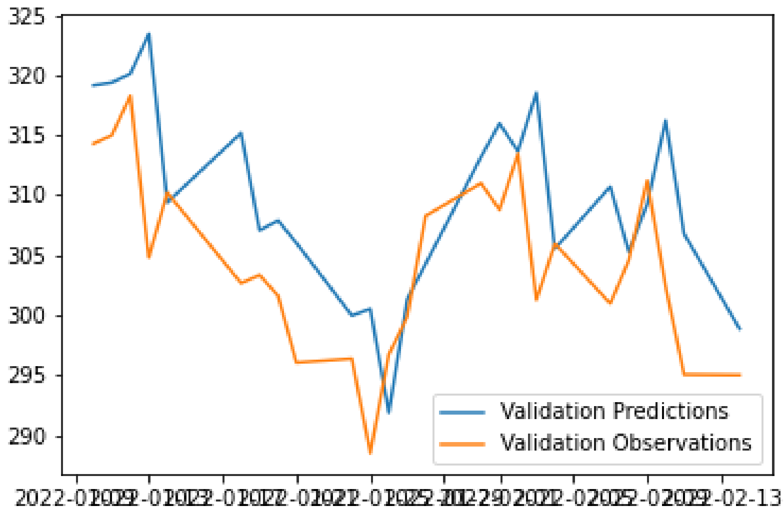

Figure 6: Validation of values.

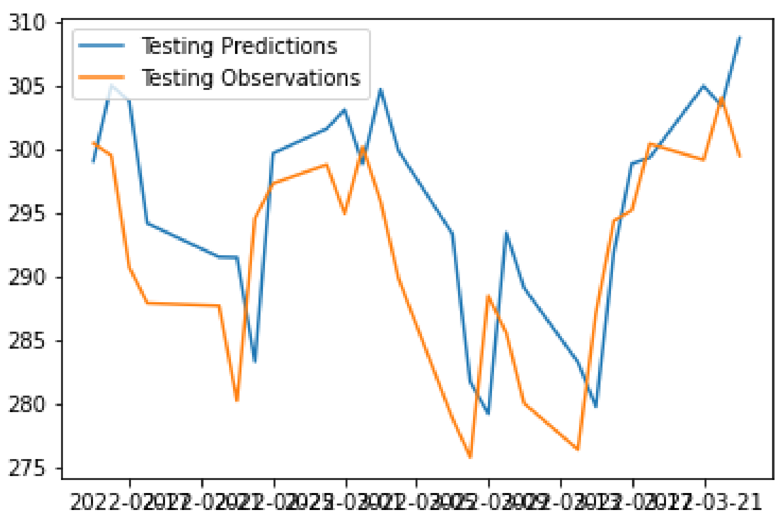

Figure 7: Testing and Observations.

and the recursive predictions, and the result is quite amusing because the model has no idea how to predict the future;

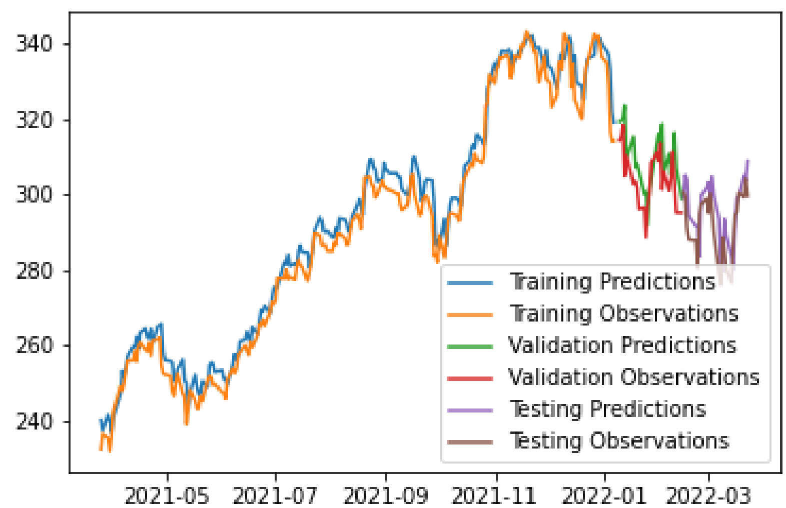

Figure 8: Total Prediction and Observations.

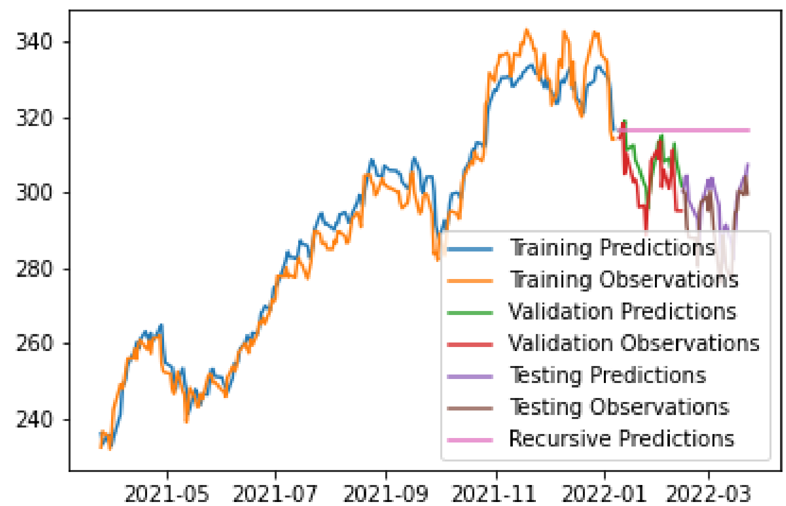

Figure 9: Recursive Training and Prediction

It simply assumes that things will remain the same, which is a reasonable assumption. We may consider the trend; we observed the graph eventually begin to rise, which leads us to feel that this stock is worth investing in. However, the dilemma resides in the difficulty of producing accurate stock market forecasts.

5 Conclusion

This article analyzes shares; it may be updated in the future to analyze additional shares. The prediction may be more accurate if the model is trained with a larger number of data sets, in addition to more powerful processing power, additional layers, and LSTM modules. In the future, we intend to apply social media sentiment analysis to determine how the market feels about share price fluctuations. This is made possible by integrating the Twitter and Facebook APIs into our program, as Facebook is a widely utilized social network with a multitude of user-submitted market trend data.