Performance analysis of Stock Price Prediction over multiple datasets using LSTM, RNN

Agni Otulonio Noor, *190042120* Abrar Chowdhury, *190042140* and Arafat Kabir Khan, *190042146*

Abstract—The prediction of stock value is a complex task which needs a robust algorithm background in order to compute the longer term share prices. Stock prices are correlated within the nature of market; hence it will be difficult to predict the costs. The proposed algorithm using the market data to predict the share price using machine learning techniques like recurrent neural network named as Long Short Term Memory, in that process weights are corrected for each data points using stochastic gradient descent. This system will provide accurate outcomes in comparison to currently available stock price predictor algorithms. The network is trained and evaluated with various sizes of input data to urge the graphical outcomes.

Index Terms—LSTM, RNN, Feature analysis, Dataset, Prediction.

1 Introduction

STOCK price prediction using machine learning is the process of predicting the future value of a stock traded on a stock exchange for reaping profits. With multiple factors involved in predicting stock prices, it is challenging to predict stock prices with high accuracy, and this is where machine learning plays a vital role. In this project, we built a machine learning model for predicting stock values of Google, Apple and Tesla.

1.1 Problem Statement

To predict the stock price of a company based on previous data and to compare the performance of our LSTM model on multiple datasets on the basis of comparison factors such as accuracy, f1 score etc.

1.2 Problem Domain

Broadly, stock market analysis is divided into two parts – Fundamental Analysis and Technical Analysis. Fundamental Analysis involves analyzing the company's future profitability on the basis of its current business environment and financial performance. Technical Analysis, on the other hand, includes reading the charts and using statistical figures to identify the trends in the stock market. Our focus will be on the technical analysis and visualization part. This is a Regression problem in the field of Machine Learning as we will be trying to accurately fit the predicted stock prices (from the model) to the actual stock prices.

1.3 Motivation

Learning about LSTM and wrapping our heads around RNN was a challenge. But the harder challenge was how to implement these. The motivation was to use what we learnt from our Machine Learning course in a real life scenario and we chose Stock Market prediction as it is a popular problem that involves the use of Neural Network models.

- *Agni, Abrar and Arafat were with the Department of Computer Science and Engineering, Islamic University of Technology, Board Bazar, Gazipur.*
- *E-mail: abrarchowdhury@iut-dhaka.edu, arafatkabir@iut-dhaka.edu, agniotulonio@iut-dhaka.edu*

Manuscript received on December 25, 2022

2 Background Study

2.1 Overview of the project and related terms

Based on experiments conducted in different articles, LSTMs seem to be the best initial approach in solving the stock price prediction problem. Other methods can combine features extracted from LSTM or Bi-LSTM models and fed into a classical ANN regressor. This approach might help extract information previously missed by a simple LSTM regression model. More recent research uses graph neural networks and multi-headed attention mechanisms, while others use reinforcement learning. LSTM -based approaches are also being actively used and researched lately despite such complex mechanisms.

2.2 Data collection and feature analysis

We have collected about two to three (Google, Tesla, Netflix) datasets of past few years of companies' stocks. We have collected the said datasets from WSJ, Kaggle and Yahoo Finance. The dataset contains the opening price, closing price, adjusted closing price, lowest price of the stock, highest price of the stock and the total number of sales (also called volume) corresponding to each day in a specific time period. According to []

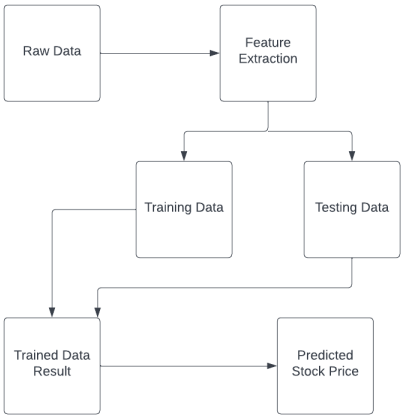

Fig: Architecture of Stock Price Prediction Model

2.2.1 Data Collection

Data collection is the primary step and initial module for the project. Here it deals with collecting the right dataset for the project. The dataset has been collected from Quandl, WSJ, Kaggle and Yahoo Finance..The right dataset is collected to predict the stock price. Our dataset collected from the historical data.

2.2.2 Feature Extraction

Feature Extraction is the system to pick the vital characteristic to be expecting the inventory price. Because the uncooked statistics are amassed they may have too many characteristics, however simplest the vital characteristic we want is to be expecting the inventory price. So the usage of the random woodland set of rules choosing the vital characteristic. The random woodland set of rules classify the characteristic just like the tree shape and set up the vital characteristic on one facet and undesirable characteristic on different facet. Based on the frequency price the vital characteristics are arranged. In that dataset they vital characteristic are Close, Open, High, Low, Volume and Adj extent those are the vital characteristics are expecting the inventory market.

2.2.3 Data Split

The dataset that used to be expecting the inventory rate is located to break up into training set and validation set. The data is generally broken up into testing data and validation data. The schooling set carries a recognized output and the version learns this information on the way to be generalized to different information later on. They break up the training data proposition and validation. The training data can be in 70 percent, wherein the validation data can be in 30Description of the Regression model:

i. Classification model used LSTM and Recurrent Neural Network Recurrent Neural Network(RNN) is a type of Neural Network where the output from the previous step are fed as input to the current step.

ii. Architectural diagram of the classifier

iii. Explanation of the parameters or hyperparameter (if any):

Different statistical values, price to book ratio, price to earning ratio,price earnings to growth, debt to equity ratio, dividend yield

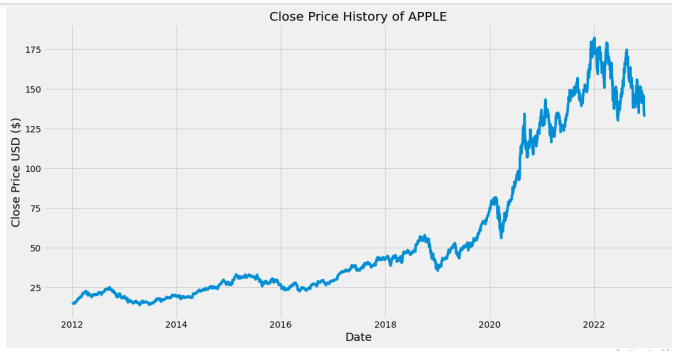

Fig. 1. Closing Price History of Apple.

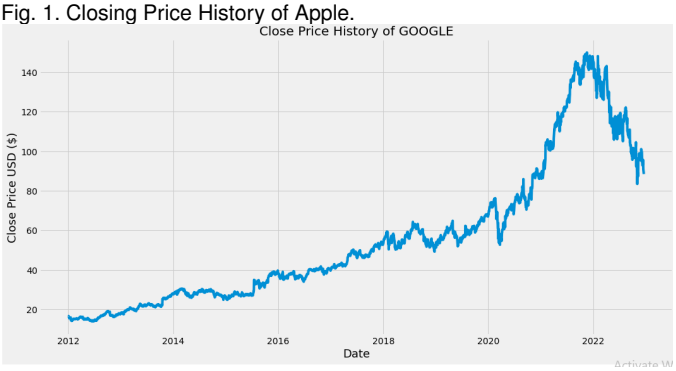

Fig. 2. Closing Price History of Google.

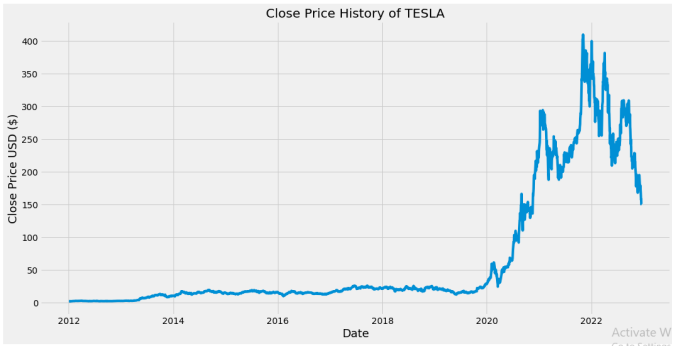

Fig. 3. Closing Price History of Tesla.

3 Result and Discussion

We used a total of three data sets of stock quotes of Apple, Google and Tesla from 2012-01-03 to 2022-12-20.

The data sets have a total of six features : Opening Price, High Price, Low Price, Closing Price, Adj. Closing Price and Volume. We visualized the closing price history of the stocks and built our model to predict the closing price of each stock at any given date.

We split our data sets into testing and training data. Training data sets consists of 80 percent and testing data sets consists of 20 percent of the whole data sets.

We built a neural model with a total of four layers. The first two layers are LSTM layers with 50 neurons each. The third layer is a dense layer with 25 neurons and the last layer is another dense layer with a single neuron.

After that we predicted the data using the trained models and compare the real data with the predicted data.

Then we calculate their root mean square errors(RMSE). RMSE for Apple is 6.85, Google is 20.70, Tesla is 134.20.

```
# Get room mean squared error (RMSE)
rmse = np.sqrt(np.mean(predictions- y_test)**2)
rmse2 = np.sqrt(np.mean(predictions2- y_test2)**2)
rmse3 = np.sqrt(np.mean(predictions3- y_test3)**2)
print(rmse)
print(rmse2)
print(rmse3)
```

```
6.85251852394878
20.706000203671664
134.20712382551554
```

Fig. 4. Root Mean Squared Errors

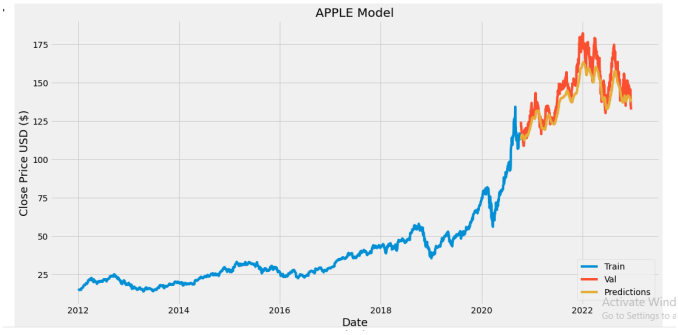

Fig. 5. Predicted value comparison of Apple

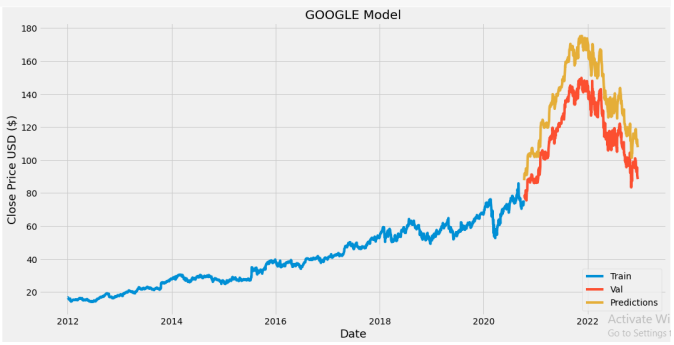

Fig. 6. Predicted value comparison of Google

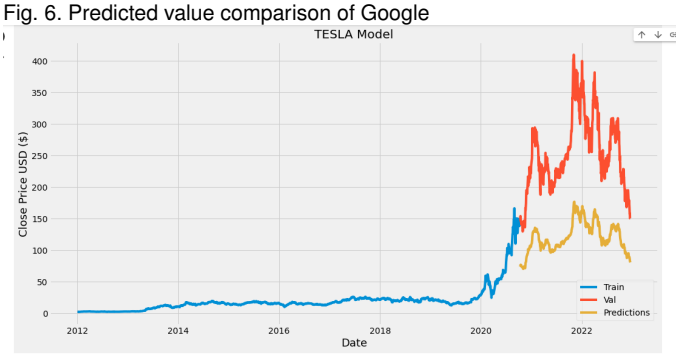

Fig. 7. Predicted value comparison of Tesla

4 Conclusion

As we can see from the figures 5,6 and 7, our model gives close predictions to the Apple and Google stock data but doesn't perform well on the Tesla stock data.

References

[1] Pramod, Pm, Mallikarjuna, *Stock Price Prediction Using LSTM*, 83rd ed. Test Engineering and Management, 2021.