Project Report

On

APPLE STOCK PRICE PREDICTION

Submitted

in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in

APPLIED STATISTICS AND DATA ANALYTICS

by

RIZWANA K J

(Reg No. SM22AS017)

(2022-2024)

Under the Supervision of

Mrs. KAVYA KISHORE

DEPARTMENT OF MATHEMATICS AND STATISTICS

ST. TERESA'S COLLEGE (AUTONOMOUS)

ERNAKULAM, KOCHI – 682011

APRIL 2024

ST. TERESA'S COLLEGE (AUTONOMOUS), ERNAKULAM

CERTIFICATE

This is to certify that the dissertation entitled, **APPLE STOCK PRICE PREDICTION** is a bonafide record of the work done by Ms. **RIZWANA K J** under my guidance as partial fulfilment of the award of the degree of **Master of Science in Applied Statistics and Data Analytics** at St. Teresa's College (Autonomous), Ernakulam affiliated to Mahatma Gandhi University, Kottayam. No part of this work has been submitted for any other degree elsewhere.

Date: 29/04/2024
Place: Ernakulam

Kavya

Mrs. Kavya Kishore

Assistant Professor,
Department of BCA (CT & ISM),
St. Teresa's College (Autonomous)
Ernakulam.

ST. TERESA'S COLLEGE (AUTONOMOUS) DEPARTMENT OF MATHEMATICS AND STATISTICS ERNAKULAM

Mrs. Nisha Oommen
Assistant Professor & HOD,
Department of Mathematics and Statistics,
St. Teresa's College (Autonomous),
Ernakulam.

External Examiners

1. ..... Chinju Joseph

2. ..... LAKSHMI SURESH 29/04/2024

DECLARATION

I hereby declare that the work presented in this project is based on the original work done by me under the guidance of **Mrs. Kavya Kishore**, Assistant Professor, Department of BCA (CT & ISM), St. Teresa's College (Autonomous), Ernakulam and has not been included in any other project submitted previously for the award of any degree.

Place: Ernakulam

Date: 09/04/2024

RIZWANA K J

SM22AS017

ACKNOWLEDGEMENT

I must mention several individuals who encouraged me to carry out this work. Their continuous invaluable knowledgeable guidance throughout this study helped me to complete the work up to this stage.

I am very grateful to my project guide Mrs. Kavya Kishore for the immense help during the period of work.

In addition, the very energetic and competitive atmosphere of the Department had much to do with this work. I acknowledge with thanks to the faculty, teaching, and non-teaching staff of the department and Colleagues.

I am also very thankful to HoD Mrs. Nisha Oommen for their valuable suggestions, critical examination of work during the progress.

Place: Ernakulam

Date: 29/04/2024

RIZWANA K J

SM22AS017

ABSTRACT

Stock market seems to be a simple platform where the shares of publicly listed companies are traded. But in reality, it is very complex. On today's economy, stock market has a dominant impact. There is huge risk involved in the investment in the stock market, but it also helps in gaining income. Investors and economists are attracted towards the gains. Thus, efficiently predicting the stock market trend helps investors to reduce the risk and increase the profit. Therefore, it is very important to analyze the stock market trend carefully. Machine learning algorithms helps in easy prediction with more accurate results. Accurate predictions help the investors to gain profits. In this project, two models were created applying the principle of ANN (Artificial Neural Network) and GRU (Gated Recurrent Unit) to predict the stock price of Apple.

ST.TERESA'S COLLEGE (AUTONOMOUS) ERNAKULAM

Certificate of Plagiarism Check for Dissertation

Author Name	RIZWANA K J
Course of Study	M.Sc. Applied Statistics & Data Analytics
Name of Guide	Ms. KAVYA KISHORE
Department	Post Graduate Mathematics & Statistics
Acceptable Maximum Limit	20%
Submitted By	library@teresas.ac.in
Paper Title	APPLE STOCK PRICE PREDICTION
Similarity	0% AI 9%
Paper ID	1689391
Submission Date	2024-04-24 10:39:15

Signature of Student

Signature of Guide

Checked By
College Librarian

* This report has been generated by DrillBit Anti-Plagiarism Software

Contents

CHAPTER – 1
INTRODUCTION

1.1 Introduction to Stock Market

The Stock Market is a place where the shares or stocks of the publicly listed companies are bought and sold. The other names of Stock Market are Equity Market or Share Market. Shares represent ownership in a company. Publicly Listed Companies issues stocks so as to increase their business. Investors invest in the stock market to create wealth or to get profit from the companies which is called the Dividend. There are many types of Investors like Individual Investors also known as Retail Investors, Institutional Investors etc. These Investors perform Fundamental Analysis, Technical Analysis about the company whose shares they are going to buy so that their investment is not wasted. Some of the factors influencing stock price are Company's performance, Market Sentiments etc. Stock Brokers facilitates the trading of the shares of the publicly listed companies in the stock market. When an Institution or a Company goes public, they issue their shares which represents the ownership in a company. Then the shares get listed in the stock exchange. National Stock Exchange (NSE) and Bombay Stock Exchange are the 2 Stock Exchanges in India. There are many factors that influence the stock market like Sentiments of the Investor, Supply, Earnings and Revenue of the Company, Market Sentiment etc. Some of the major stock exchanges all over the world includes New York Stock Exchange (NYSE), NASDAQ Stock Market, Tokyo Stock Exchange, London Stock Exchange, Hong Kong Stock Exchange etc. Companies must be more transparent regarding information related to their business so as to stay in the stock market. Nation's economic growth is indicated by the conditions of the Stock Market.

Therefore, Stock Market plays an important role to raise funds for business expansion and also serves as a platform for investors to earn money from their investment. But One Should always remember that investing in Stock Market are

subjected to Market Risks. So Careful Analysis of Market Trends are required before investing in the Stock Market.

1.2 STOCK PRICE PREDICTION

Predicting stock prices is very strenuous as it involves forecasting future values based on Historical data, Market trends. The Objective of Stock price prediction is to assist the investors to build informed decisions regarding the stocks, consequently reducing risks. There are various strategies and techniques applied for stock price prediction including Fundamental Analysis, Technical Analysis and machine learning algorithms. Fundamental Analysis provides investors the understanding about the company's performance and its potential growth. Investors use a procedure called The Technical Analysis to forecast future price movements based on historical data. Various Machine Learning algorithms are used to identify market trends and patterns and to make predictions. Unfortunately, there is no prediction model that can guarantee precise conclusions, so before making any decision its important that sufficient research is done beforehand.

This project seeks to first perform exploratory analysis on the stock data of Apple (AAPL) and then to predict the stock price of Apple (AAPL) using Gated Recurrent Unit (GRU) and Artificial Neural Network (ANN).

1.3 OBJECTIVES OF THE STUDY

1. To Perform Exploratory Data Analysis on Apple Stock Data.
2. To Predict the Closing Price of Apple using Gated Recurrent Unit.
3. To Predict the Closing Price of Apple using Artificial Neural Network.
4. To Compare the Performance between Artificial Neural Network and Gated Recurrent Unit.

CHAPTER-2

LITERATURE REVIEW

This Chapter presents the conclusions derived from related research that analyzed and predicted future stock values using different machine learning algorithms.

- Nunno (2014) predicted the stock closing price using linear regression, Stochastic Gradient Descent and Support Vector Regression. They found that Support Vector Regression was the most competent out of the models used.
- Shastri et al. (2019) predicted the stock price using a method by which the sentiment scores are calculated first through Naïve Bayes classifier and then on historical data and on sentiment scores obtained, Neural Network is utilized. An Accuracy level of more than 90% is obtained from this study.
- Jin et al. (2019) predicted the stock price based on investor's sentiment and then by adopting Empirical modal decomposition, they had slowly decomposed the complex sequence with different scale of fluctuations and the prediction accuracy got improved. Then they predicted using LSTM and revised it by adopting attention mechanism to focus more on the critical information. The Experimental results showed that the revised LSTM reduced the time delay along with better prediction accuracy.
- Zhu (2020) predicted the stock price of Apple using a deep learning technique called RNN. Prediction accuracy of the conducted Experiment was 95%.
- Aasi et al. (2021) predicted the stock price of Apple using Multivariate Multistep Output Long-Short-Term-Memory (MMLSTM). On Compared to ARIMA and Random Forest Model ,MMLSTM Model improved the MSE value of up to 65% when compared with these 2 models.

- Burra et al. (2022) attempted predicting the stock price of Twitter Data through Zero-Shot Sentiment Classification and also using Deep Learning Algorithms. The accuracy obtained by the model was of 94.21%.
- Han and Fu (2023) predicted the stock price of Apple using Bi-directional Long Short-Term Memory (Bi-LSTM). The difference between LSTM and Bi-LSTM is that Bi-LSTM processes the input sequence in both directions. So Long term dependencies and short-term dependencies are captured in the dataset. Then Mean Squared Error is evaluated. They found out that, Bi-LSTM is a good tool for predicting the stock prices.
- Zhen (2023) predicted the trend of the stock by performing LSTM of Sandp500 data. Hundred Epochs training were conducted in the dataset. Without having any delay, they predicted the learning curve. The MSE value of 0.238 was obtained.
- Wu et al. (2023) predicted the stock market using Long Short-Term Memory (LSTM) Neural Network. They have also predicted the stock price using Arima model. When both were compared , LSTM model provided more accurate predictions.
- Bin and Huang (2023) used LSTM model to predict the stock price of Apple by using quarterly as well as daily historical price. Different variations in the LSTM model are studied and evaluated. Thus, they found that, based on careful selection of the fundamental indicators for training the model based on Fundamental Analysis, the accuracy of the model can be improved.

CHAPTER-3

MATERIALS AND METHODOLOGY

3.1 DATA DESCRIPTION

Daily stock price data of Apple (AAPL) from 2012 to 2024 has been collected from the website of "Yahoo Finance"(https://wwwYahoofinance.com). This data set involves daily information on the opening, closing, highest, lowest, adj volume, volume of the Above Tech Company.

3.2 METHODOLOGY

The Preliminary and the fundamental step in this analysis was to study the data in detail. The Main purpose of the study was to gain insights about the stock data of Apple and then predicted the closing price of Apple using Artificial Neural Network (ANN) and Gated Recurrent Unit (GRU) and finally found out which model performed the best based on the RMSE values.

❖ TOOLS FOR ANALYSIS AND PREDICTION

- Artificial Neural Network (ANN)
- Gated Recurrent Unit (GRU)

❖ TOOLS FOR COMAPRISON

- Root Mean Squared Error (RMSE)

3.2.1 EXPLORATORY DATA ANALYSIS

Exploratory Data Analysis is a method of analyzing data set and gaining insights by abstracting the main characteristics about the data set. By performing EDA, Researchers can identify trends and patterns. Visualization plots such as Box Plot,

density plot are used. EDA helps us to gain the summary statistics of the dataset such as mean, median, std deviation etc. If the Data set contains outlies, then the result will not be reliable. By performing EDA, one can identify and handle the outliers. Because these outliers affect the validity of our dissection. So, by executing data cleaning process, the results obtained will be more accurate and reliable. Graphical Exploratory Data Analysis and Numerical Data Analysis are some of the types of Exploratory. Thus, Exploratory Data Analysis helps the researchers to make the better and informative decisions.

3.2.2 GATED RECURRENT UNIT (GRU)

Introduced by Kyunghyun Cho et al in 2014, Gated Recurrent Units (GRUs) are a gating mechanism in Recurrent Neural networks. It is similar to Long Short-Term Memory (LSTM), in addition to a gating mechanism but there is no output gate due to which only few parameters are considered than LSTM. Gated Recurrent Unit is a modification of the Recurrent Neural Networks (RNN) used to solve the limitations of the traditional RNN. It works by handling sequential data. GRU has fewer parameters and their architecture is simple. Traditional RNNs endures in capturing long term dependencies, this problem is solved by GRUs which regulate the movement of information through the network, by introducing gating mechanism.

There are several important components in the architecture of Gated Recurrent Unit, this includes:

1. Update Gate: It is a component which allows selection of only important information that is required in the present state of the network. It acts like a selective mechanism.

2. Reset Gate: This component decides the quantity of information ignored by GRU while processing current input.

3. Current Memory Content: This Component indicates about the current state of the GRU Unit.

4. Candidate Activation: This Component put forward updated memory content that might sometimes be added to the Unit.

5. Hidden State: It is the Output of the GRU, through which one can gain insights from the input. It helps to make predictions and informed decisions.

ADVANTAGES OF GATED RECURRENT UNIT

- Efficient Training : The Architecture of GRU is very simple making its computation less demanding when compared with LSTM. They have few parameters, so the efficiency during training is maximum and the Computation Cost is low.
- Simplicity and Elegance: Due to its simple architecture and having few parameters and operations, interpretation and analyzation becomes easier.
- Memory Efficiency: By using update and reset gates, memory efficiency is achieved by the GRUs.
- Robustness to Vanishing Gradients: Vanishing Gradient Problem is one of the obstacles in training deep neural networks, which means that the gradient decline exponentially as they transmit backward through the network layers. GRUs because of its gating mechanism, Vanishing Gradient Problem is addressed effectively, which promotes the flow of the gradients across time steps.
- Competitive Performance: GRUs captures long term dependencies, complex patterns are also found. Accurate Predictions are also generated by GRUs.

3.2.3 ARTIFICIAL NEURAL NETWORKS

An interconnected group of nodes makes an Artificial Neural Network or simply Neural network. These nodes are called artificial neurons, which is similar to the neurons constituted in our brain. The Connected neurons sends signal to the artificial neuron, then after processing it ,sends signals to other connected neurons. The signal received is a real number. The Activation Function is a nonlinear function of the sum of the inputs of each neuron which computes the output of each neuron. Neurons are aggregated into layers. So, from the first layer (The Input layer) to the last layer (The Output layer), the signals will get transmitted. There might be multiple intermediate

layers or called the hidden layers also in between the first and the last layer. A Deep Neural Network (DNN) has at least two hidden layers.

In Predictive Modelling and in Adaptive Control, Artificial Neural Networks are used. It is also used in Artificial Intelligence to solve problems. From Experience, these networks learn and also can obtain conclusions from more difficult set of datasets.

The Neurons mentioned have a weight. These weight gets adjusted as the learning proceeds and these weights represents the strength between the connections of the neurons. Feedforward Neural Networks (FNNs), Recurrent Neural Networks, Convolutional Neural Networks are some of the types of the Artificial Neural Networks.

Neural Network uses a method called Empirical Risk Minimization during training. In this method, networks parameters are optimized so that the difference between the actual target value and the predicted output is minimized in a given dataset. Every neuron contains a parameter called bias which helps the Artificial Neural Network to learn more complex functions. Optimizer is a component that reduce the loss function. When training, this reduction happens by updating the weights and biases. Some of the examples of optimization algorithm include Stochastic Gradient Descent, Adam, RMSprop etc. The Estimation of the parameters in the networks is done by using a method which is based on the Gradient. When the input data is given into the neural networks , the output gets computed layer by layer until we obtain the final output. This Process is called as The Forward propagation. An Important method which is used for adjusting the connection weights to compensate for error during learning is called Backpropagation.

ADVANTAGES OF ARTIFICIAL NEURAL NETWORK

- Non-Linearity: Artificial Neural Network is used to model complex datasets. It is used when the relationship the input and the output variables is non-linear.
- Parallel Processing: Simultaneously performs multiple operations. So, this enables handling of large sets of data.
- Fault Tolerance: Even when there is noise in the dataset, the ANN model can make accurate predictions on unseen dataset.
- Feature Learning: ANN can generalize the results without performing manual feature engineering. It can extract important features from the dataset.
- Scalability: From Complex dataset to smaller dataset, this model is used. So, this property makes the model adaptable and applicable to different range of applications.

3.2.4 Root Mean Squared Error (RMSE)

Root Mean Squared Error is a metric which is used to evaluate the accuracy of the model. A lower RMSE value indicate more accurate prediction. It is given by the Equation

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

where,

y_i is the actual value for the i_{th} observation

$y_i\hat{}$ is the predicted value for the i_{th} observation

n is the number of observations

3.2.5 Python

A high-level programming language, dynamically typed and garbage collected. It is also a multi-paradigm programming language. It is known for its simplicity, versatility. It is one of the most popular programming languages. NumPy, SciPy, Pandas etc. are some of the libraries equipped in python. These libraries help researchers to analyze large dataset. TensorFlow is a library which facilitate the execution of machine learning algorithms. Also, Libraries like Matplotlib and Seaborn help to Visualize complex data.

CHAPTER-4

RESULTS AND ANALYSIS

4.1 EXPLORATORY DATA ANALYSIS

First checked whether there is any null values so as to ensure accuracy in the model's prediction.

Table 4.1.1 shows the data type and the count of non-null values in each of the variable in the Column Head.

#	Column	Non-Null Count	Data Type
0	Open	3052	Float64
1	High	3052	Float64
2	Low	3052	Float64
3	Close	3052	Float64
4	Adj Close	3052	Float64
5	Volume	3052	Float64

Table 4.1.1

Next, Table 4.1.2 shows the descriptive statistics

	Open	High	Low	Close	Adj Close	Volume
Count	3052	3052	3052	3052	3052	3.052000e+03
Mean	69.347829	70.091298	68.640151	69.399770	67.279904	1.900940e+08
Std	56.899839	57.530737	56.316247	56.959720	57.446441	1.689772e+08
Min	13.856071	14.271429	13.753571	13.947500	11.982237	2.404830e+07
25%	24.852500	25.168125	24.667500	24.963125	22.474305	8.593780e+07
50%	41.931250	42.512750	41.546249	42.056250	39.895237	1.295468e+08
75%	126.051872	127.451252	124.730001	125.892500	123.944839	2.292135e+08
max	198.020004	199.619995	197.000000	198.110001	197.857529	1.506120e+09

Table 4.1.2

Then Plotted the closing price of Apple from 2012-01-03 to 2024-02-20

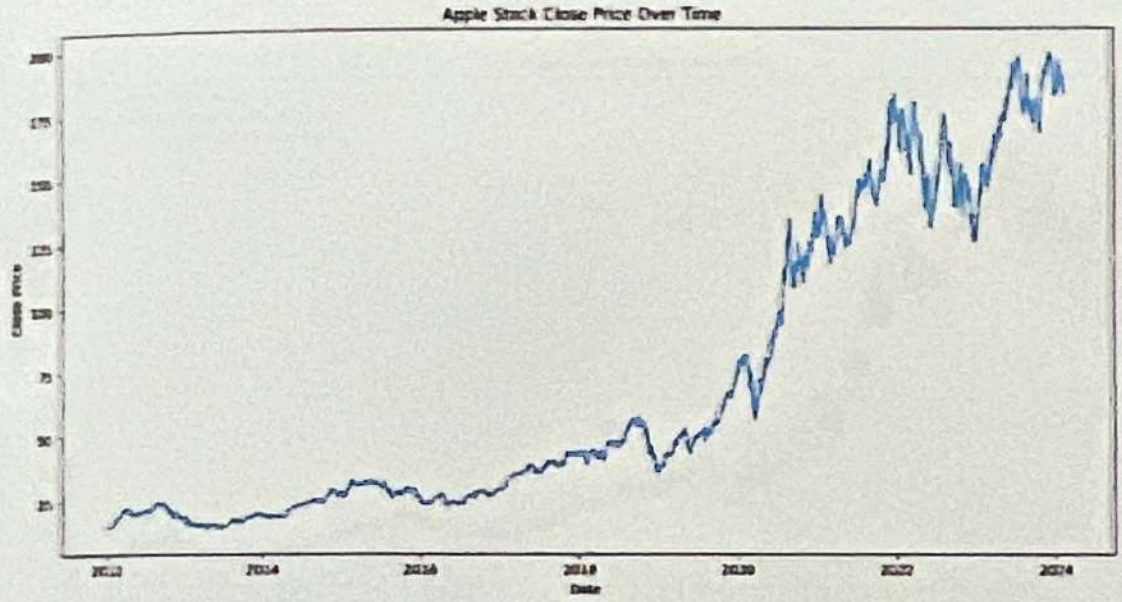

Figure 4.1.3

4.2. Predicting the Closing price using Gated Recurrent Unit

Step1 : Collecting the data. We have taken Stock quote of Apple beginning January 1st 2012.

Step 2 : Normalize the Data by using the MinMaxScaler which scales between 0 and 1. Then the scaled dataset is converted to a data frame.

Step3 : Split the Data. The Dataset is split into Training and Testing dataset using the common 80:20 ratio.

Step4 : Build GRU model. Setting up the layers of the GRU model. Patterns can be studied by adjusting these layers, Compile the model and then train the data.

Step5 : Now we calculate the testing rmse and the training rmse. As the amount of data is increased, the value of MSE decreases.

Training RMSE : 1.1709772735249386

Testing RMSE : 7.069117949973832

Step6 : Data Visualization. Now plotted the graph for the Predicted value and the Actual value.

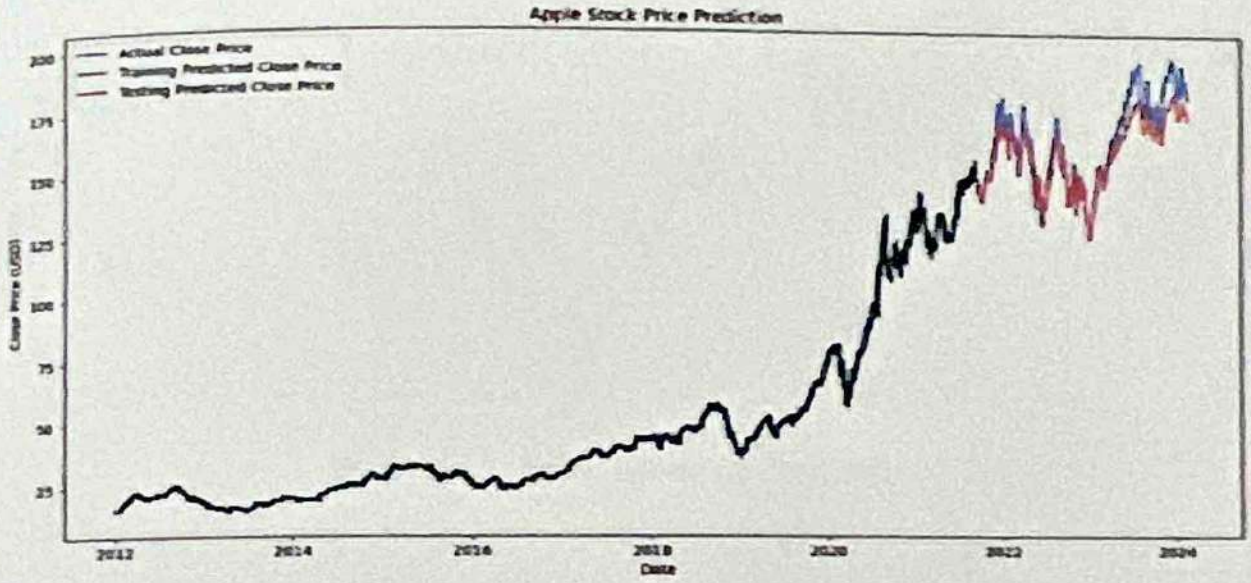

Figure 4.1.5: Apple Stock data Visualization

Now, A closer look into the Actual and Predicted values from 2021-09-20 to 2024-02-21 is given by Figure 4.1.6.

Date	Actual	Predicted
2021-09-20	142.940002	146.753021
2021-09-21	143.429993	144.072357
2021-09-22	145.850006	143.439880
2021-09-23	146.830002	144.923553
2021-09-24	146.919998	146.211975
...	...	...
2024-02-14	184.149994	175.775772
2024-02-15	183.860001	174.989243
2024-02-16	182.309998	174.514862
2024-02-20	181.559998	173.703125
2024-02-21	182.320007	173.055145

[609 rows x 2 columns]

Figure 4.1.6: Actual Close Vs Predicted Value

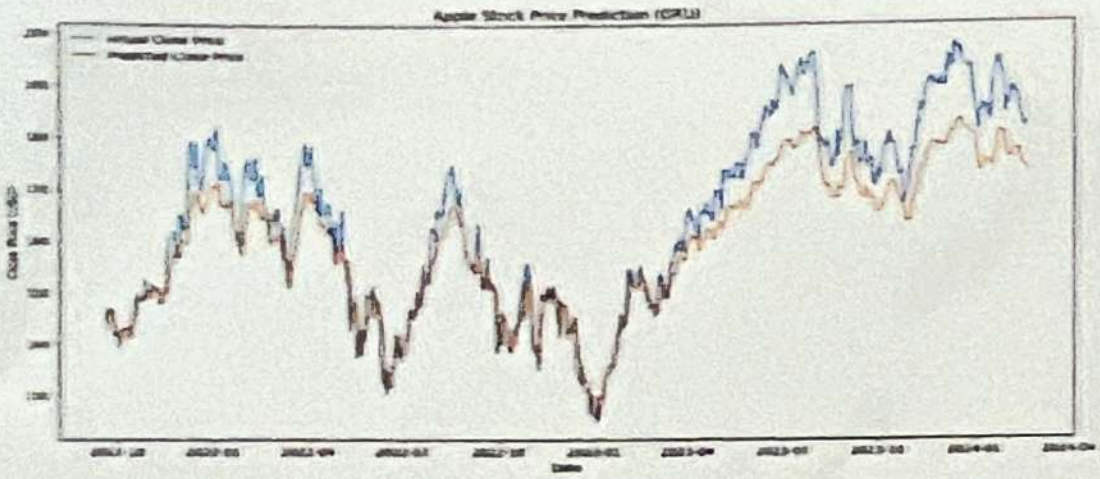

Figure 4.1.7: Apple Actual v/s Predicted

4.3 Predicting the Closing price using Artificial Neural Network

Step1 : We have taken the same dataset for prediction

Step2 : Data Preprocessing. Selected the close column which represents the stock closing price.

Step3 : Normalization. Min-Max scaling is used to normalize the close prices. This converts the values in the range (0,1).

Step4 : Sequences are created for training the model.

Step5 : Split the data. The Dataset is split into Training and Testing using the common 80:20 ratio.

Step6 : Build ANN model. Setting up the layers of the ANN model. Train the model.

Step7 : Predictions are made. Then, Evaluate the model. Calculate the testing rmse and training rmse.

Training RMSE : 1.230208451870644

Testing RMSE : 3.1263837864694923

Step8 : Data Visualization. We plot the graph for the trained and the predicted value.

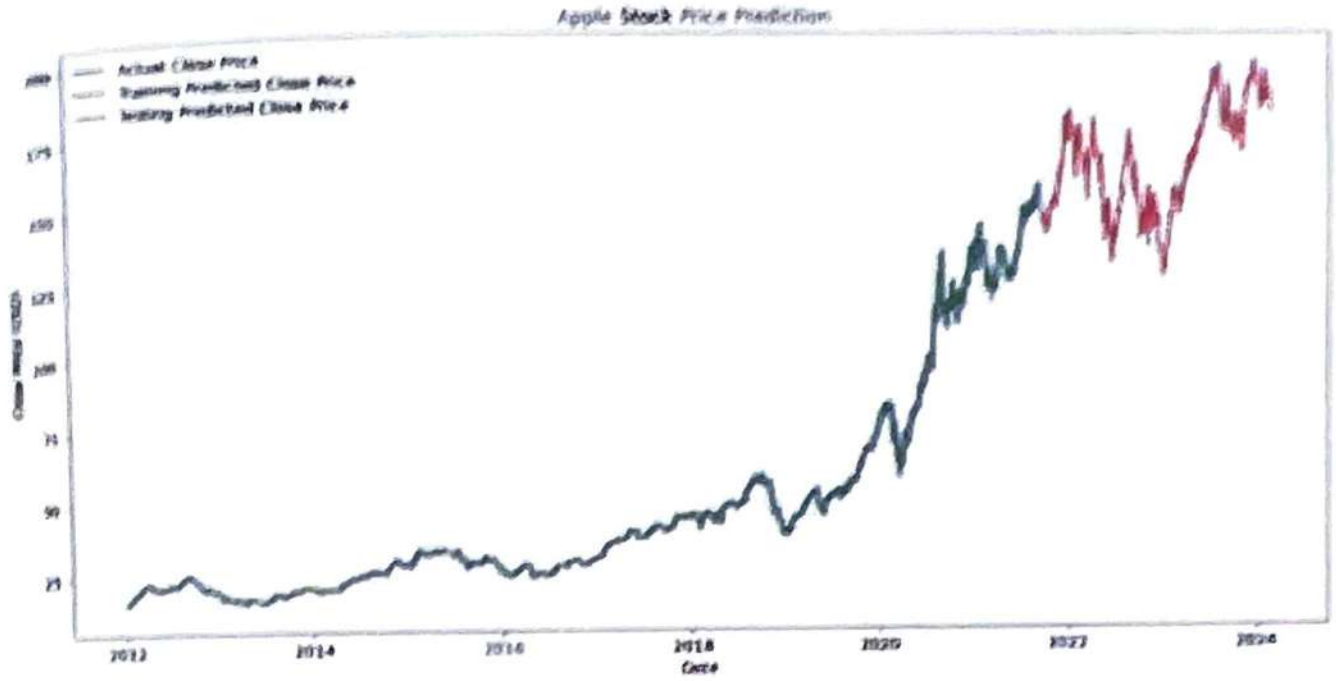

Figure 4.1.8. Apple Stock data Visualization

Now, A closer look into the Actual and predicted values from 2012-09-20 to 2024-02-21 is given by Figure 4.1.9.

Date	Actual	Predicted
2021-09-20	142.940002	146.364304
2021-09-21	143.429993	143.746002
2021-09-22	145.850006	143.174194
2021-09-23	146.830002	145.050613
2021-09-24	146.919998	146.707138
...	...	...
2024-02-14	184.149994	185.440399
2024-02-15	183.860001	183.890015
2024-02-16	182.309998	183.018356
2024-02-20	181.559998	182.002487
2024-02-21	182.320007	180.746582

[609 rows x 2 columns]

Figure 4.1.9: Actual Close vs Predicted Value

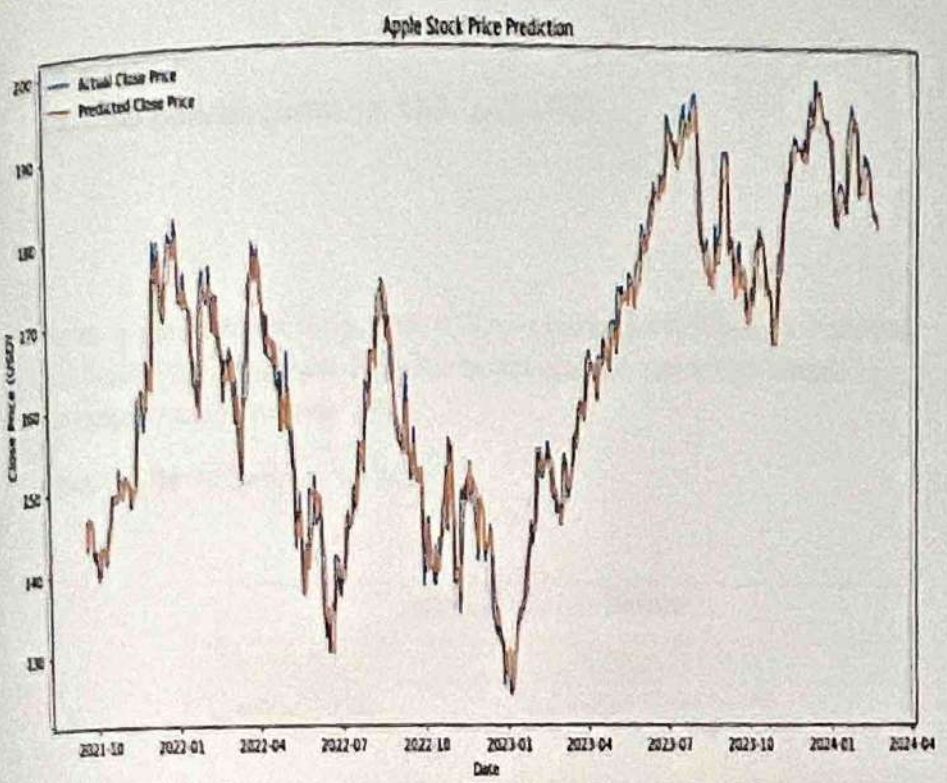

Figure 4.1.9: Apple Actual v/s Predicted

4.4 Comparison Between ANN and GRU

We have predicted the Stock price of Apple using Gated Recurrent Unit and Artificial Neural Network. Now Let us compare the above two models to determine which performs better.

From the above findings, we have

Artificial Neural Network	
Training RMSE	1.230208451870644
Testing RMSE	3.1263837864694923

Gated Recurrent Unit	
Training RMSE	1.1709772735249386
Testing RMSE	7.069117949973832

Since there is a slighter higher training rmse for the ANN model compared to the GRU model. This shows us that on the training data, the GRU performs slightly better. But on the testing dataset, The ANN model has a lower rmse on compared with the GRU model. This tells us that on the unseen testing data, The performance of the ANN model is better than the GRU model.

CHAPTER 5

CONCLUSION

The study was undertaken to perform exploratory data analysis and also to predict the stock price using Artificial Neural Network and Gated Recurrent Unit. The daily stock data of Apple from 2012 to 2024 are used for the study. We were able to draw valuable insights by performing EDA. Summary Statistics of stock data help us to comprehend the dispersion and central tendency of the stock prices. Then Successfully created a model that can predict a close value to the Actual value. The trend of the actual stock is predicted by our model. Accuracy of the model can be increased by training with more data and also increasing the layers of the models used. We have found on the basis of comparison between the ANN and the GRU that ANN performs better. Calculation of RMSE allows us to know about its accuracy.

REFERENCES

1. Aasi, B., Imtiaz, S. A., Qadeer, H. A., Singarajah, M., & Kashef, R. (2021, April). Stock Price Prediction Using a Multivariate Multistep LSTM: A Sentiment and Public Engagement Analysis Model. In 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS) (pp. 1-8). IEEE.

2. Burra, L. R., Koppolu, B. P., Karthik, B. D., Priya, B. L. N. S., Prasanthi, A., & Tumuluru, P. (2023, January). Stock Price Prediction using Zero-Shot Sentiment Classification. In 2023 5th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 741-746). IEEE.

3. Han, C., & Fu, X. (2023). Challenge and opportunity: deep learning-based stock price prediction by using Bi-directional LSTM model. Frontiers in Business, Economics and Management, 8(2), 51-54.

4. Huang, M. B. (2023). Recurrent neural networks for Apple stock price prediction.

5. Jin, Z., Yang, Y., & Liu, Y. (2020). Stock closing price prediction based on sentiment analysis and LSTM. Neural Computing and Applications, 32, 9713-9729.

6. Nunno, L. (2014). Stock market price prediction using linear and polynomial regression models. Computer Science Department, University of New Mexico: Albuquerque, NM, USA.

7. Shastri, M., Roy, S., & Mittal, M. (2019). Stock price prediction using artificial neural model: an application of big data. EAI Endorsed Transactions on Scalable Information Systems, 6(20), e1-e1.

8. Wu, H., Chen, S., & Ding, Y. (2023). Comparison of ARIMA and LSTM for Stock Price Prediction. Financial Engineering and Risk Management, 6(1), 1-7.

9. Zhu, Y. (2020, October). Stock price prediction using the RNN model. In Journal of Physics: Conference Series (Vol. 1650, No. 3, p. 032103). IOP Publishing.

10. Zhen, L. (2023). Stock Price Prediction using Deep Recurrent Model. Highlights in Science, Engineering and Technology, 44, 318-322.


ST. TERESA'S COLLEGE (AUTONOMOUS)
DEPARTMENT OF MATHEMATICS AND STATISTICS
ERNAKULAM