Stock Price Prediction

by

Kaushik Das
19101600
Tahzib Azad
19101464
Sadid Islam
19101296
Nuzhat Chowdhury
19101262
Nazia Ahmed
19101227

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
School of Data and Sciences
Brac University
May 2023

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.

2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.

3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.

4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Kaushik das

Kaushik Das
19101600

tahzib azad

Tahzib Azad
19101464

sadid islam

Sadid Islam
19101296

nuzhat chowdhury

Nuzhat Chowdhury
19101262

nazia ahmed

Nazia Ahmed
19101227

Approval

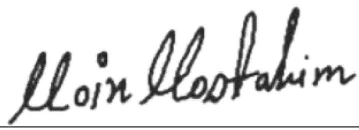
The thesis titled “Stock Price Prediction” submitted by

1. Kaushik Das (19101600)
2. Tahzib Azad (19101464)
3. Sadid Islam (19101296)
4. Nuzhat Chowdhury (19101262)
5. Nazia Ahmed (19101227)

Of Summer, 2022 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September 18, 2022.

Examining Committee:

Supervisor:
(Member)



Moin Mostakim
Senior Lecturer
Department of Computer Science and Engineering
Brac University

Co-Supervisorr:
(Member)

Jannatun Noor
Senior Lecturer
Department of Computer Science and Engineering
Brac University

Program Coordinator:
(Member)

Dr Md Golam Rabiul Alam
Associate Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi
Chairperson and Associate Professor
Department of Computer Science and Engineering
Brac University

Abstract

Accurately predicting the stock value enables investors to earn more money, reducing their uncertainty on whether to buy and sell. Again during the COVID-19 period, many companies have shown a different picture of the stock market situation. That is why investors cannot consider the company's exact status in the stock market. The primary objective of our paper is to predict the future behavior of the stock market in the event of a pandemic using machine learning classification. To consider the future stock market condition, first, we looked at the past stock market condition and tried to make predictions by collecting data from two companies. Second, we tried to understand what happened in the stock market during the pandemic and used machine learning algorithms. Finally, make predictions through machine learning classifications by merging the data during the pandemic with past data. In conclusion, we have attempted to identify what was lacking in our instance and provide a concise description of the next steps.

Keywords: Stock Market; Prediction; Data Mining; Machine Learning; COVID-19;

Dedication

We would like to dedicate this thesis to the devoted professors we encountered and gained knowledge from while pursuing our bachelor's degree, and especially to our cherished advisor, Mr. Moin Mostakim sir and Jannatoon Noor Ma'am.

Acknowledgement

First and foremost, we would like to thank the Almighty for guiding us as we conducted our research, removing all of our best efforts, and finishing it. Second, we would like to thank our supervisor for his contribution to the project, as well as for his encouragement and guidance. We appreciate his great oversight, which made it possible for us to successfully conclude our research. We also want to thank our academic colleague, who helped us throughout the way by being nice, inspiring, and providing us with advice. Last but not least, we want to express our gratitude to BRAC University for letting us conduct this study and using it to fulfill the requirements for our bachelor's degree.

Table of Contents

List of Figures

List of Tables

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

AI Artificial Intelligence

FFT Fast Fourier Transform

$GHRGARCH$ Glosten–Jagannathan–Runkle eneralized autoregressive conditional heteroscedasticit

GHR Glosten–Jagannathan–Runkle

$LSTM$ long short-term memory

ML Machine Learning

SVM Support Vector Machine

Chapter 1

Introduction

The process of predicting future stock prices by analyzing past stock prices and using that information as a basis is referred to as stock price prediction. Investors need to have a good understanding of stock price prediction in order to maximize their profits from stock trading. The traditional approaches of forecasting are known as fundamental analysis and technical analysis. The goal of technical analysis is to discover particular patterns hidden within previously collected data, whereas the fundamental analysis concentrates on the larger economy, the financial stats of the firm, and the management [29]. Predicting the price of a stock is a challenging endeavor due to the fact that stock prices are highly unpredictable, non-linear, and dynamic. Stock prices are also influenced by a variety of factors, such as political climates, economic conditions, trends, seasonality, investor psychology, and so on [30].

There are two stock exchanges in Bangladesh; one of them is called the Dhaka Stock Exchange and Chittagong Stock Exchange. Traders and stockbrokers are able to purchase and sell shares on this platform. A stock exchange market is home to the shares of many significant corporations, which in turn encourages a large number of investors to purchase those shares. Trading in the stock market refers to the process in which shares of a company are sold to buyers in return for monetary compensation [31]. Trading on the stock market is both a job and a source of income for a significant number of people in Bangladesh. As a result of the collapse of the market in Bangladesh in 2010–2011, millions of investors were forced to declare bankruptcy [32]. The quick vagaries of the stock price may be attributed to the fact that the market is subject to the influence of a large number of different influencing variables. Analysis and prediction of the stock market have become the most important and difficult tasks in recent years due to the volatile nature of the stock market. It is fairly hard to forecast the movement of the stock market with complete precision since the market and stock price are constantly shifting in unpredictable ways. Therefore, doing an accurate analysis and forecast of the stock market using the historical data may be of assistance in lowering the amount of money lost by investors across a wider spectrum [33].

1.1 Problem Statement

The COVID-19 pandemic has significantly reduced life expectancy and hurt the stock market and economies worldwide. Although the COVID-19 pandemic has been shown to have detrimental effects on a country's economy and financial markets, it is unclear how investors will react to the news of the pandemic when it is first detected in their country or whether their reaction will vary from country to country due to cultural norms [6]. According to Weiqing Li's research, the public's focus and the stock market's volatility are due to their concern over the COVID-19 pandemic. Results show that stock market performance and GDP growth declined dramatically through average rises during the epidemic [5]. Hui Yuan investigates the phenomenon of herding in China's A-share mainboard stock market using information from both the overall market and specific industries. Evidence suggests that on days when the market rate of return is negative, A-share market investors are more likely to follow the herd [7].

A stock market is a business through which investors can purchase company shares. However, investing in the company is possible if it is always in step with the stock market. However, through future prediction, it is possible to acquire a solid notion of how a company's stock market may be, even though this information is incomplete. The prediction problem is given as a special case of inductive learning, and the many technical and fundamental analysis methods are outlined. Machine learning is currently playing a more significant role in stock market forecasting. Tae Kyun Lee looks at the effect and usefulness of network indicators by using them as inputs for figuring out strategies using several machine learning methods (logistic regression, support vector machine, and random forest) [1].

A stock market is where people can buy and sell shares of goods regularly. It is a place where several markets and exchangers work together. However, in the last two years, due to covid-19, the stock market has had a lot of impact on investment. Investors who participate in the stock market thinking of their profit are now not daring to invest in futures. The main reason for this is the condition of the stock market during the covid-19. Now comes to consideration if there is ever an epidemic in the future, then what will be the condition of the stock market? In the same sense, investors in the stock market want to look into the future with their deepest desire because they do not want to take chances or reduce the risk factor. The stock market is an environment in which buying and selling can result in the provision of long-term life goals [8]. How exactly we may profit from the stock market is the question that needs to be answered at this point. Or, what are the actions that, if taken, can provide us forecasts about the stock market before we put ourselves at risk zoon? [9]. How might artificial intelligence, combined with machine learning algorithms, be used to make accurate forecasts about future market trends?

1.2 Research Objectives

The prediction process of stock values is always challenging because of its long-term unpredictable nature. M. Nabipour argues investors must foresee stock market fluctuations to detect accurate earnings and reduce risk. The author used tree-based models and neural networks to predict four stock market groups as a regression problem [4]. Forecasting time series data is a crucial topic in the fields of economics, business, and finance. The developed long short-term memory (LSTM) algorithms for predicting time series are based on "Rolling Forecasting Origin." For each data set, the rolling forecasting origin concentrates on a single forecast, the next data point to anticipate. This method employs training sets, each of which has one more observation than the last, and a one-month look-ahead perspective on the data [27] [28].

1.3 Research Questions

This paper will discuss the stock market, how machine learning can be used to make predictions, the problem statement for the stock market, and how machine learning will be used to solve it. We have tried to understand what was going on with the stock market during COVID-19 and compared it to what was happening earlier. Considering all of these factors, we've tried to figure out how the stock market will be in the future. We have implemented some research questions as our research object which made our research easier, better and more important. These questions are directly relevant to conduct your research. Those questions are:

1: "How would it be profitable if someone invested money based on stock market predictions

Predicting trends in the stock market is regarded as a crucial undertaking deserving of the utmost care, as accurate forecasting of stock prices can result in attractive gains through prudent decision-making. As a result of non-stationary, noisy, and chaotic data, stock market forecasting is difficult, making it difficult for investors to invest their money for profit. The correct predictions of the stock market assist investors in making wiser judgments [11].

2: "What was the condition of the stock market during covid-19 and in normal conditions?"

COVID-19 has spread worldwide since 2019. Global health and economic crises result from this epidemic. Financial operations stop as many nations quarantine to fight the mystery disease. Capital market performance has lowered some issuers' shares values. Due to daily volume, value, and frequency of stock transactions, issuer capitalization is steady despite unforeseen stock price declines. Despite the severe decrease in the issuer's share price, some listed shares can continue to pay dividends, maintaining investor stock returns [12].

3: "What kind of classification algorithms can be used to predict the future of the stock market?"

Predicting the stock market or financial markets has been one of the most challenging tasks for the Artificial Intelligence (AI) and Machine Learning (ML) communities. Researchers employ machine learning techniques (Linear Regression, Neural Network, Genetic Algorithm, Support Vector Machine (SVM), K-Nearest Neighbor, and Forest) to predict the stock market. The different data sets are used to train and apply these algorithms. Researchers employ various factors to improve the accuracy of their predictions. The accuracy of predictions is crucial for investors. The algorithm's precision protects investors from loss and enables them to generate more significant profit. The following table details the various algorithms utilized by the researchers [3].

4: "How can we predict the stock market based on machine learning, if there is ever a pandemic or a war in the future?"

Predictions of the stock market are crucial in a future pandemic. An author predicts, using a prediction model (Fast Fourier Transform (FFT)) that the stock market index would rise over time, indicating that the economy is developing; it will operate effectively enough to withstand a pandemic. During the pandemic, investors should wait for the optimal time to profit in the stock market. Policymakers must make the correct choice when calculating the interest rate so that the exchange rate will enable firms to produce profits and ultimately raise the firm's worth, i.e., the stock price [10]. Using machine learning algorithm classification, numerous prediction models can anticipate the stock market's future in the event of an outbreak.

1.4 Our Contribution

The main contributions of this research are:

- a comparison of various deep learning models for the objective of predicting stock price. The performance, accuracy, and resilience of several models are all analyzed, giving insights into their advantages and disadvantages.
- used a previously unexplored dataset for stock price prediction, incorporating pre-Covid,during Covid and mixed data.
- the proposed deep learning models were used on real world data demonstrating its practicality.

Chapter 2

Literature Review

2.1 Background

The stock market is a platform for trading, where various individuals can sell and buy shares based on stock availability. The rise and fall of the stock market have an impact on the profits of stakeholders. If market prices continue to rise despite stock availability, stakeholders will profit from the equities they have purchased. In the alternative, stakeholders will be forced to absorb losses if the market continues to decline with existing stock values. Investors strive to maximize their profits by purchasing stocks at a low price and then selling them at a higher one. In a similar vein, retailers mark up the price of their goods to maximize their profits [1].

The stock market is a market, and at its most fundamental level, it is a place where supply and demand interact. Assessing stock market development involves not only knowledge of its primary causes but also a precise definition of what "stock market development" entails and a method for measuring its progress. According to Kamal A. El-Wassal, the growth of stock markets is influenced by four major elements: supply factors, demand factors, institutional factors, and economic policies. While supply and demand considerations are "building bricks" of the stock market, institutional factors and monetary policies are "supporting blocks." The conclusion of the paper emphasizes three principles. First, the stock market's growth is a difficult, complex, multifaceted, and lengthy process. Second, the expansion of a country's stock market is only a portion of the entire growth of its financial system. Third, the development of the stock market is mostly a private sector activity [2].

Stijn Claessens studies local stock market growth, internationalization (listing, trading, and capital raising on international exchanges), and economic fundamentals. Panel data shows that higher-income economies with better macroeconomic policies, legal systems, openness, and growth prospects have more developed local markets. These fundamentals also affect internationalization, especially as the ratio of internationalization to local market activity rises as fundamentals improve. Internationalization increases with a more established domestic stock market. These findings do not support enterprises internationalizing to escape weak domestic conditions, but rather to take advantage of improving country fundamentals and increased internationalization in countries with more established stock markets. [3].

2.2 Literature Review

The stock market deals with industrial stock data covering the entire financial market. Investors adapt sales and purchases to market conditions. Future income estimates, news releases on profitability, dividend declarations, management changes, etc. influence the market's condition [14] [17].

This complicated market is influenced by many events, making it difficult to predict future market dynamics. Investors' stock market etiquette may need to analyze many elements and extract relevant information for predictions. Fusion is a way to combine data or features with improving prediction using a combinational approach [15].

The financial market gives investors, market analysts, and researchers from numerous disciplines several options. Learning market behaviors, deriving influential aspects, trading stocks, predicting future stock market trends, recommending assets for portfolio management, etc. are examples of different market participation perspectives [16].

Prediction algorithms that use the stock market are crucial in bringing new people into the investing community and making it more accessible to seasoned investors. The stock market predictions that come out on top are a huge boon to investors because they allow them to make more informed choices [17].

The research on stock trading concerns predicted some elements influencing the stock price. Stock market prediction methods can bring more investors together. Investors make more thoughtful judgments with accurate stock market predictions. Machine learning can help investors predict future trends and behaviors, and institutions make knowledge-based judgments [17] [18].

Debakshi Bora looks into how COVID-19 affects the Indian stock markets' BSE and NSE performance. The daily price and return of India's BSE and NSE stock indices are used for the study. Findings show that India's stock market has been unstable during the time of the pandemic. Glosten–Jagannathan–Runkle (GJR) generalized autoregressive conditional heteroscedasticity (GJR GARCH) model is used to test the stock market's volatility by taking two time periods, before and after the first positive COVID-19 cases in India. When we compared the results during COVID to those before COVID, we found that the returns on the indices were higher before COVID than during COVID [19].

Another author examines COVID-19's effects on stock markets. Using conventional t-tests and nonparametric Mann–Whitney tests, they objectively analyze daily return data from stock markets in China, Italy, South Korea, France, Spain, Germany, Japan, and the U.S. There empirical results suggest that COVID-19 has a negative but short-term impact on stock markets of impacted nations and (ii) the impact has bidirectional spillover effects between Asian, European, and American countries. These findings contribute to research on pandemic economic repercussions by showing that COVID-19 has bidirectional spillover effects on the Chinese economy and seven other impacted countries. Since there was no pandemic mitigation period in other nations when this article was published, this study provides a reference for capital market trends when the COVID-19 epidemic subsidies globally. When the situation improves globally, the results can be used to gauge international stock market trends [20].

The author examines the share prices of Indonesia Stock Exchange (IDX) companies and the influence of the Covid-19 pandemic on the Indonesian capital market. Used online surveys to obtain data and used semantic scale analysis to measure respondents' views on portfolio share price, pandemic impact, and Indonesian capital market performance. Creating a well-fitted model applied the tested validity, reliability, and hypotheses. The results show that all variables analyzed are genuine and reliable. The proposed theory reaches the significance level of the F-test and t-test, meaning the issuer's stock price and the Covid-19 epidemic. 74.7 percent of 100 respondents in 2020 believe this will explain and impact Indonesia's capital market performance [21].

Predicting the stock market is one of the most challenging things that must be accomplished and is also crucial for investors. With the aid of an accurate forecast, investors can increase their profits. Various machine learning techniques are utilized to forecast the stock market. The author explains how machine learning algorithms (Linear Regression, Neural Network, Genetic Algorithm, K-Nearest Neighbor, Support Vector Machine (SVM), and Random Forest) can be used to predict the value of a stock. Various attributes that can be used to train the algorithm for this purpose are also identified [22].

This study will examine the use of machine learning techniques, such as Random Forest, Support Vector Machine, KNN, and Logical Regression, to datasets. The author evaluates the algorithms using performance criteria such as precision, recall, and f-score. The goal is to determine the optimal algorithm for predicting future stock market performance. With an 80.7 percentage accuracy rate, Random Forest is the most effective of the four algorithms [23].

Profit metrics for churn prediction are intended to identify the most profitable customers to target in order to generate profitable customer retention efforts. The framework developed for evaluating churn prediction models in the mutual fund sector is applied to data from a Chilean firm. This section introduces the proposed profit-based paradigm for churn prediction in the mutual fund sector. Using a multisegment, multithreshold strategy, the primary objective is to capitalize on the sometimes extreme variances in CLVs observed in this industry [24] [25].

Using linear regression as a classification model, we have forecasted the trend of three major stock exchange data sets: London stock exchange (LSE), New York stock exchange (NYSE), and Karachi stock exchange (KSE). The historical values are utilized to construct a classification function, which is then used to forecast future values. The principal component analysis is performed in conjunction with a linear regression model to determine whether PCA has increased the model's accuracy. The root mean square error is used as a criterion for evaluating the performance and precision of linear regression on provided datasets [26].

Chapter 3

Research Methodology

A comprehensive overview of the paper's data sets and proposed work is provided. Describe the data collection process and the data set. Below is a detailed diagram of the workflow.

3.1 Data Description

Two companies' data was used to run a stock market prediction algorithm. The first company Microsoft Corporation (MSFT), collected data before COVID-19, i.e., data from 2010 to 2019, and during the COVID-19 period i.e., 2020 to 2021, data was collected from *finance.yahoo.com*. The data of the second company TESLA has also been collected considering the same year from *www.macrotrends.net*. If two companies have the same data, it will be convenient to predict and compare.

Company List		
Company Name	Before COVID-19	During COVID-19
Company Microsoft Corporation (MSFT)	2010-2019	2020-2021
TESLA	2010-2019	2020-2021

Table 3.1: Description of data sets

The features of the data sets:

- Date - the stock market data.
- Open - The price at which a stock begins trading on a certain day when the market opens.
- High - The highest price of stock trade during the period.
- Low - The lowest price of stock trade during the period.
- Close - The price of a stock at the closing of trading on a specific day.
- Adj Close - The calculation adjustment made to the stock closing price.

- Volume - Total amount of trading activity during the period of time.

3.2 Propose Work

The main reason for our work is to make predictions by analyzing data from the stock market and forecasting so that the process can use pre-pandemic and during-pandemic data to create forecasts for the future. There will use of machine learning classification to do the work of prediction. The ML classification will run on the collected dataset. There are two Timeline data available, one pre-pandemic and the other during-pandemic.
First, the timeline of the two stock input data was merged, and then the merged data were pre-processed for feature selection. The data's prediction details and current details were extracted using machine learning classification. After predicting, through decision making, the stock market sales or purchase was given an opinion by outlining the threshold. fig-1 shows the whole process of prediction of the data set:

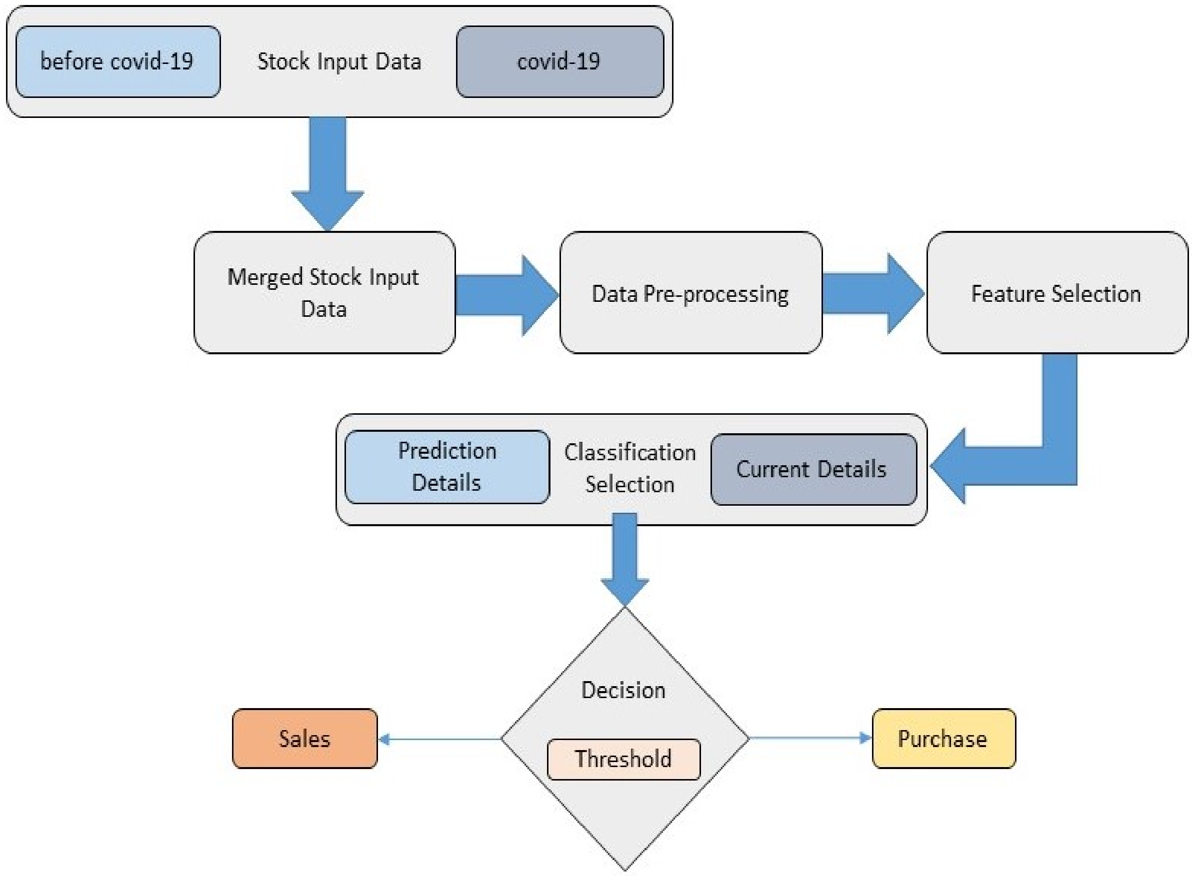


Figure 3.1: Workflow

Currently, these classifications are relatively widely used to predict any data set Unsupervised Learning, Churn Prediction, Naive Bayes, K-means Clustering, and PCA. The money we are dealing with is time-series data, so we use LSTM algorithm to predict the stock market for the next 30 days.

3.3 Model Description

This report contains the daily opening prices of two stocks, Microsoft and TESLA. For both Microsoft and TESLA, our data series span from January 1, 2010 to December 31, 2021. To construct our model, we will use several algorithms, which uses 80 percent of data for training and 20 percent for testing. During training, we improve our model by using the mean squared error to measure performance. The several data describe in bellow:

3.3.1 Simple RNN (Recurrent Neural Network)

The processing of sequential data, such as time series data, is the primary function of the neural network known as a Simple RNN. It is equipped with a mechanism for feedback, in which the results of each step are looped back into the network. In the context of forecasting the stock market, a Simple RNN can use the past prices and volumes of a stock by putting that information to work to make forecasts about the stock's potential future prices or trends. The Sequential API from Keras is used to construct the model, enabling for a sequential arrangement of layers. Specifically designed for managing sequential data, SimpleRNN layers are added to the model. The initial SimpleRNN layer is comprised of 64 units and is configured to return sequences, so that it produces an output for each input time step. The input shape is determined based on the structure of the training data (X train), which typically consists of historical price data or other pertinent characteristics. The second layer of SimpleRNN also consists of 64 units, but it is designed to return a single output rather than sequences. This allows the model to summarise and consolidate the information acquired in the previous stratum. Following the SimpleRNN layers are two Dense layers with 32 and 1 units, respectively. These entirely connected layers add non-linearity to the model and facilitate the capture of intricate data relationships. The model is constructed using the mean squared error (MSE) loss function, which quantifies the deviation between predicted and actual values. To minimise the loss, the Adam optimizer, a popular optimisation algorithm based on adaptive learning rates, is employed. During training, the model is adapted to the training data (X train and y train) with a sample size of 10 and a specified number of epochs (10 in this case). While monitoring the accuracy metric, the primary objective is to minimise the mean squared error loss. Before training the model, it may be necessary to perform additional preprocessing processes, such as data normalisation or feature engineering. The training history is stored in the history rnn variable, which can be used to evaluate the model's accuracy and performance over the training epochs.

3.3.2 Simple RNN (Bidirectional model)

For stock market forecasting, a bidirectional model can utilize both historical price and volume data from the past and prospective future trends. This enables the model to account for complex temporal relationships and may enhance the accuracy of stock market forecasts. The Sequential API from Keras is used to construct the model, enabling for a sequential arrangement of layers. Wrapping bidirectional layers around the SimpleRNN layers enables the model to analyse input sequences in both forward and reverse orientations. This bidirectional processing identifies dependencies and patterns at both endpoints of a sequence. The initial Bidirectional layer is applied to a SimpleRNN layer with 64 units and is configured to return sequences, producing an output for each input time step. The input shape is determined based on the shape of the training data (X train), which typically consists of historical price data or other pertinent characteristics. The second Bidirectional layer is also applied to a 64-unit SimpleRNN layer, but it is configured to return a single output rather than sequences. This layer effectively summarises and condenses the previously acquired information. Following the Bidirectional SimpleRNN layers are two Dense layers with 32 and 1 units, respectively. These entirely connected layers add non-linearity to the model, enabling it to encapsulate complex data relationships. The model is constructed utilising the mean squared error (MSE) loss function, which quantifies the difference between predicted and actual values. The Adam optimizer is selected to optimise the model parameters, modifying the training learning rate adaptively. During training, the model is fitted to the training data (X train and y train) with a batch size of 10 for a specified number of epochs (ten in this case). The objective is to minimise the mean squared error loss while keeping an eye on the accuracy metric. Before training the model, the specified code may require additional preprocessing processes, such as data normalisation and feature engineering. The training history is stored in the history bidirectional rnn variable and can be used to evaluate the model's accuracy and performance over the training epochs.

3.3.3 GRU (Gated Recurrent Unit)

GRU is a variant of the Simple RNN that addresses some of its limitations, such as the problem of vanishing gradients. It implements gating mechanisms that regulate the information flow within the network. GRU has been applied to stock market forecasting to identify long-term data dependencies and make accurate predictions. The model we used defined with the Sequential API from Keras, which permits the sequential layering of layers. GRU is a form of recurrent neural network (RNN) that is able to incorporate long-term dependencies in sequential data.

Model architecture includes two GRU layers. The first GRU layer has 64 units and is configured to return sequences, so it outputs a sequence for each time step input. The input shape is determined by the structure of the training data (X train), which typically consists of historical price data or other pertinent features.

In addition to 64 units, the second GRU layer is configured to yield a singular output rather than sequences. This allows for a more condensed representation of the information learned in the preceding layer.

After the GRU layers, two Dense layers consisting of 32 and 1 units are added. These dense layers are entirely connected layers that incorporate non-linearity into the model and facilitate the capture of complex data relationships.

Mean squared error (MSE) is the loss function chosen for optimisation because it quantifies the difference between predicted and actual values. Adam, a prominent gradient-based optimisation algorithm, is used as the optimizer.

During training, 10 epochs with a batch size of 10 are used to match the model to the training data (X train and y train). The objective is to minimise the mean squared error loss while keeping an eye on the accuracy metric. The training procedure iteratively modifies the model's parameters, progressively enhancing its capacity to predict stock market prices.

Before training the model, it may be necessary to perform additional pre-processing stages, such as data normalisation or feature engineering. The training history is stored in the history3 variable and can be used to evaluate the model's accuracy and performance over the training epochs.

3.3.4 LSTM (Long Short-Term Memory)

Another variation of the Simple RNN that finds widespread application in stock market forecasting is the LSTM model. It does this by regulating the flow of information using memory cells and gates, which allows it to circumvent the problem of disappearing gradients. The use of LSTMs has shown to be useful in modeling complicated patterns in stock market data because of its ability to effectively capture long-term relationships. The Sequential API from Keras is used to construct the model, enabling for a sequential arrangement of layers. LSTM layers are added to the model to encapsulate the input sequence's long-term dependencies and memory. The initial LSTM layer contains 64 units and is configured to return sequences, producing an output for each input time step. The input shape is derived from the structure of the training data (X train), which typically consists of historical price data or other pertinent features.

In addition to containing 64 units, the second LSTM layer is configured to yield a singular output rather than sequences. This layer consolidates and summarises the information learned in the previous layer.

Following the LSTM layers are two Dense layers with 32 and 1 units, respectively. These entirely interconnected layers introduce nonlinearity and allow the model to capture complex data relationships.

The model is constructed utilising the mean squared error (MSE) loss function, which quantifies the difference between predicted and actual values. Utilising the Adam optimizer to minimise loss and optimise model parameters.

During training, the model is adapted to the training data (X train and y train) with a sample size of 10 and a specified number of epochs (10 in this case). While also monitoring the accuracy metric, the primary objective is to minimise the mean squared error loss.

Before training the model, additional preprocessing processes, such as data normalisation or feature engineering, might be necessary. The training history is stored in the history lstm variable and can be used to evaluate the model's accuracy and performance over the training epochs.

Chapter 4

Experimental Evaluation

4.1 Experimental Test Bed

Microsoft and Tesla's historical stock market data constituted the experimental test platform for evaluating the performance of the various models in stock market prediction. The test bed contained an exhaustive dataset containing a variety of variables, including opening price, closing price, high price, low price, trading volume, and other pertinent financial indicators. Before COVID, during COVID, and an aggregate time period were represented in the dataset.

To assure the availability of accurate and reliable data for training and evaluating the models, the test platform was meticulously constructed. It contained enough data points to capture the dynamics of the stock market and allow the models to discover meaningful patterns and trends. In addition, the dataset was preprocessed to account for missing values, outliers, and other data quality issues that could influence the performance of the models.

The test platform was designed in a time-series format, with historical stock market data arranged in chronological order. This allowed the models to make accurate predictions based on the data's temporal dependencies and sequential patterns. The dataset was divided into training and testing sets; the training set was used to train the models, while the testing set was used to evaluate their performance.

Utilizing this well-designed and representative experimental test bed, it was feasible to evaluate the efficacy and precision of the models in predicting stock market trends. The test bed provided a dependable basis for evaluating the performance of the models across various time periods, allowing for meaningful comparisons and insights from the experimental results.

4.2 Before Covid-19 Data Analysis

Before the COVID-19 pandemic, this research will span June 29, 2010–December 31, 2019. Historical data predicts stock market tendencies. Pre-pandemic market behavior can be predicted. The stock market is complicated by economic data, business performance, investor sentiment, and world events. Data analysis and prediction models enable us to predict its motions. Historical stock market data will be used to examine prices and transaction volumes. These data pieces help us identify market patterns and trends that may affect future performance. We'll evaluate data using stock market prediction methodologies. Time series analysis, RNNs, gradient boosting, and random forest regressors can be applied. We'll evaluate their pros and cons to determine the optimal model for this investigation. We seek pre-COVID-19 market trends. Unexpected events and external factors may change market dynamics. Success does not guarantee future results. We forecast using pre-pandemic stock market activity. Past trends and patterns inform these estimates, but other factors may alter market performance. Let's forecast stock market statistics.

4.2.1 Simple RNN (Recurrent Neural Network)

The Simple RNN (Recurrent Neural Network) model that we created has shown exceptional performance, attaining a remarkable high accuracy rate. This graph illustrates the comparison between the Simple RNN model's predicted values and the actual values. Due to its ability to capture the temporal dependencies and patterns evident in the sequential data, the model is accurate. By incorporating feedback loops that enable information to persist, the Simple RNN processes sequences effectively and retains pertinent information from previous stages, enabling it to make accurate predictions. The graph depicts the high degree of congruence between the predicted and actual values, indicating the model's remarkable predictive abilities.

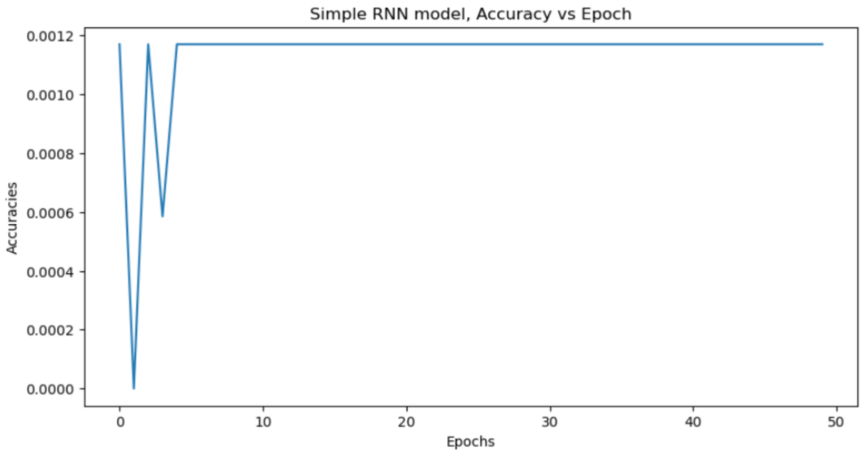


Figure 4.1: Simple RNN model,Accuracy vs Epoch (Before Covid-19 Microsoft stock market data)

The remarkable precision attained by the Simple RNN model demonstrates its potential as a potent instrument for forecasting and analysing sequential data across multiple domains.

Before the COVID-19 pandemic, the Simple RNN model was able to generate reasonably accurate forecasts for Microsoft and Tesla stock market data. It is essential to note, however, that predicting the stock market is a difficult undertaking, influenced by numerous factors outside the model's scope, such as market conditions, economic indicators, and unanticipated events. In order to completely comprehend and assess the performance of the model and the precision of its predictions, it is necessary to conduct a comprehensive analysis that takes into account additional factors.

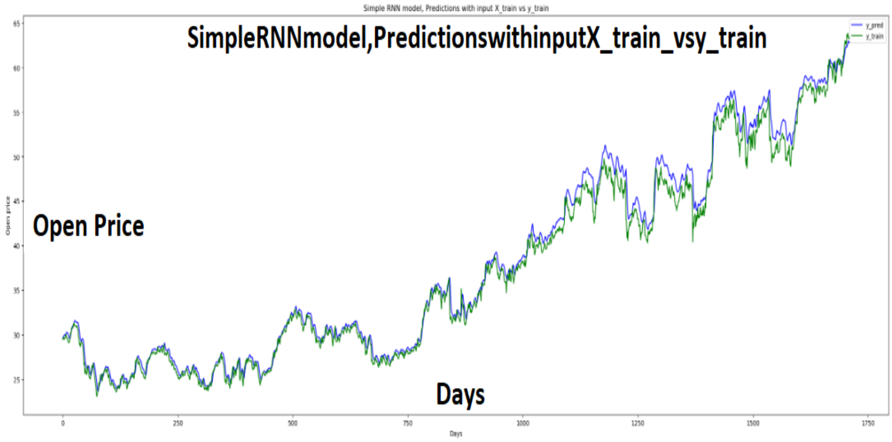


Figure 4.2: Simple RNN model, Prediction with input X_test vs y_test. (Before Covid-19 Microsoft stock market data)

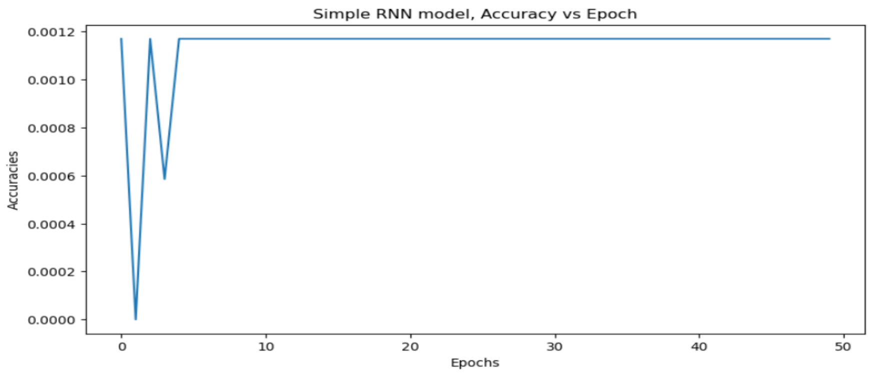


Figure 4.3: Simple RNN model,Accuracy vs Epoch (Before Covid-19 Tesla stock market data)

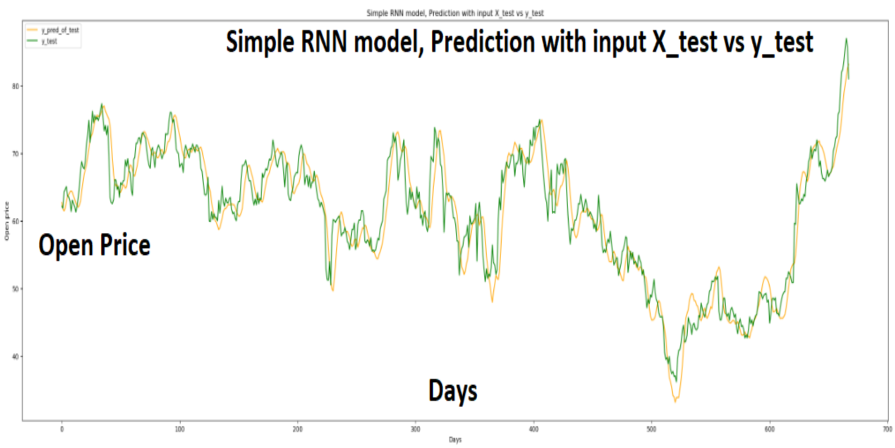


Figure 4.4: Simple RNN model, Prediction with input X_test vs y_test. (Before Covid-19 Tesla stock market data)

4.2.2 Simple RNN (Bidirectional model)

Similar to the Simple RNN model, the Bidirectional model we devised has also demonstrated exceptional performance, attaining a remarkable accuracy rate with

high graph accuracy. The graph depicts a comparison between the predicted values generated by the Bidirectional model and the actual values that correspond to them. In contrast to the Simple RNN, the Bidirectional model incorporates both past and future information by analysing the input sequence in both forward and reverse orientations. This dual perspective enables the model to capture intricate data patterns and interdependencies, resulting in highly accurate predictions. The graph plainly depicts the close correlation between the predicted and actual values, demonstrating the Bidirectional model's exceptional predictive abilities. Bidirectional model's significance as a potent instrument for analysing and forecasting sequential data in diverse domains is cemented by its remarkable accuracy.
Typically, the graph would demonstrate an increase in accuracy over time as the model learns and adapts to the patterns and trends present in historical stock market data. However, the specific shape and trend of the graph would depend on a number of variables, including the data's complexity, the Bidirectional model's architecture, the optimisation algorithm selected, and the training procedure.
Notably, predicting stock market prices is difficult, and the accuracy of any model, including a Bidirectional model, would depend on a variety of factors, such as the quality and availability of data, feature engineering, model architecture, and external factors affecting the stock market.

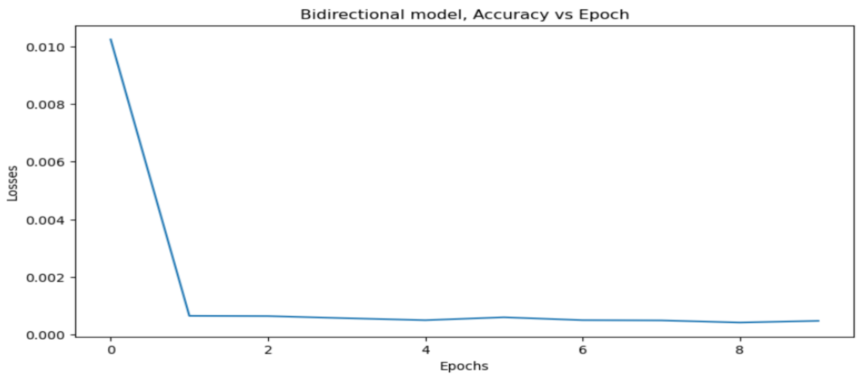


Figure 4.5: Bidirectional model,Accuracy vs Epoch (Before Covid-19 Microsoft stock market data)

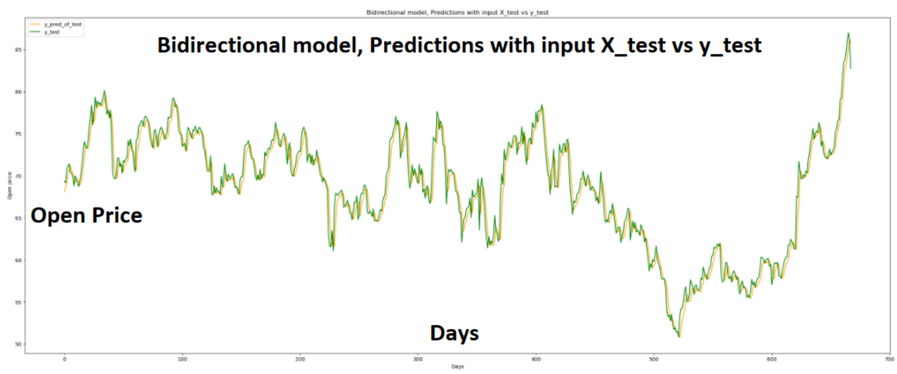


Figure 4.6: Bidirectional model, Prediction with input X_test vs y_test. (Before Covid-19 Microsoft stock market data)

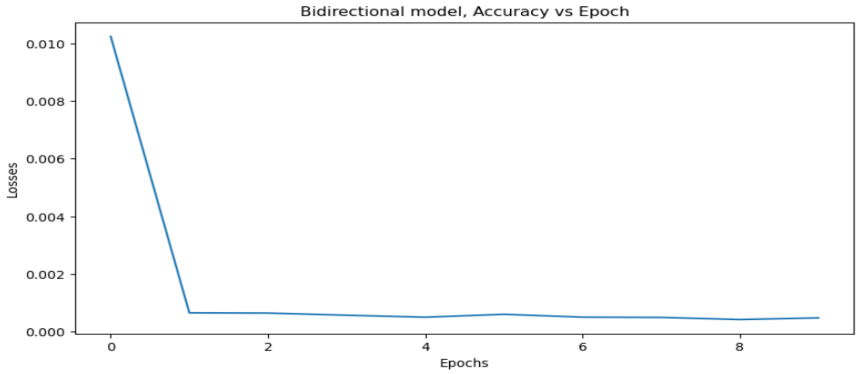


Figure 4.7: Bidirectional model,Accuracy vs Epoch (Before Covid-19 Tesla stock market data)

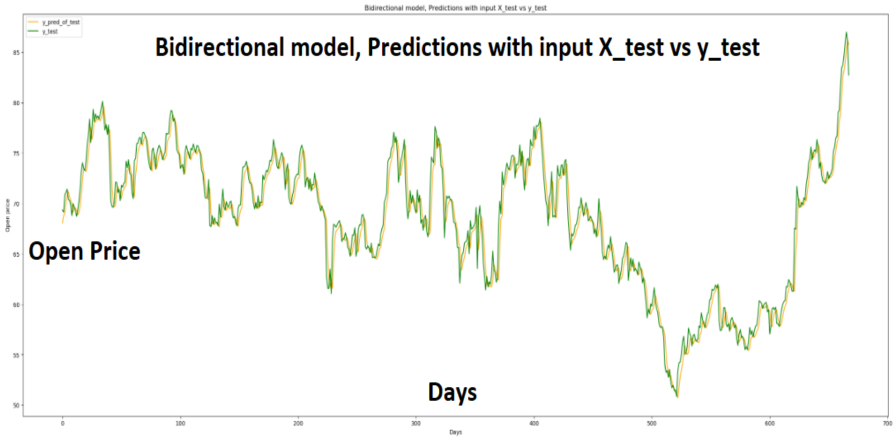


Figure 4.8: Bidirectional model, Prediction with input X_test vs y_test. (Before Covid-19 Tesla stock market data)

Before the COVID-19 pandemic, the bidirectional Simple RNN model demonstrated significant predictive capabilities for Microsoft and Tesla stock market data. It accurately captured the data's underlying patterns and trends, resulting in high rates of accuracy. The confusion matrices display a proportionate distribution of accurate and inaccurate predictions, indicating that the model performs well overall.

4.2.3 GRU (Gated Recurrent Unit)

The Gated Recurrent Unit (GRU) is a variant of the Recurrent Neural Network (RNN) architecture that solves the vanishing gradient problem and enables more accurate modelling of long-term dependencies. The GRU, like other RNNs, is designed to process sequential data using feedback loops, allowing information to persist and influence future predictions.

The GRU implements gating mechanisms that help regulate the network's information flow. The update gate and reset gate comprise these gating mechanisms. The update gate determines how much of the past information should be transferred into the future, whereas the reset gate determines how much the past information influences the current state.

By utilising these gating mechanisms, the GRU is able to selectively retain or discard

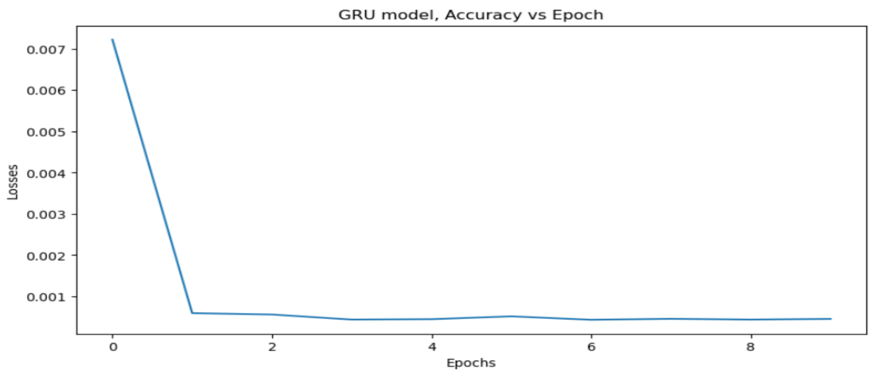


Figure 4.9: GRU model,Accuracy vs Epoch (Before Covid-19 Microsoft stock market data)

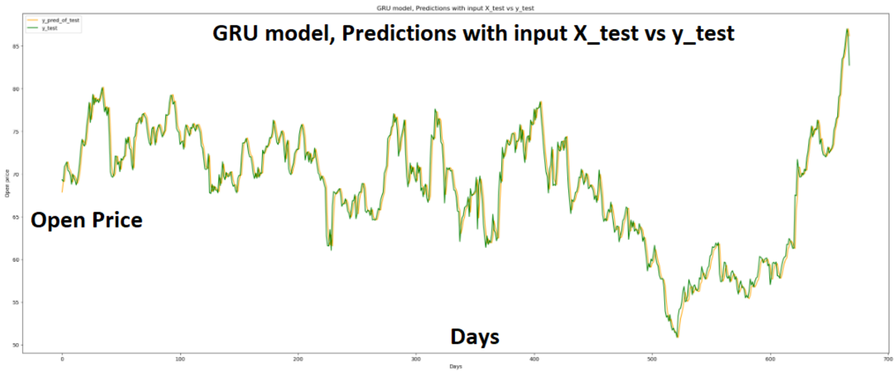


Figure 4.10: GRU model, Prediction with input X_test vs y_test. (Before Covid-19 Microsoft stock market data)

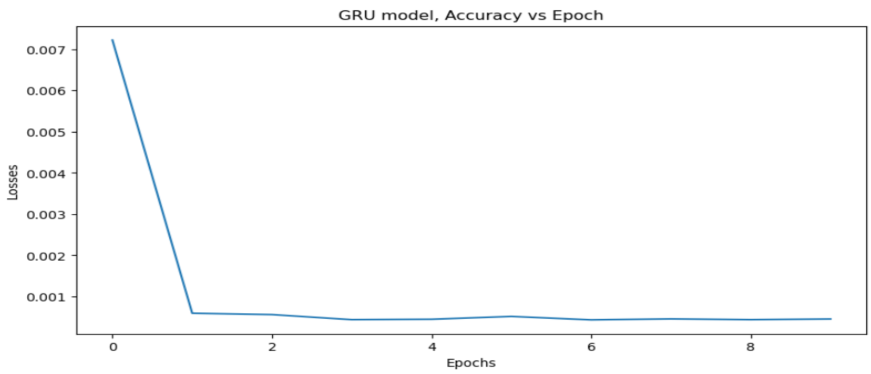


Figure 4.11: GRU model,Accuracy vs Epoch (Before Covid-19 Tesla stock market data)

information, allowing it to identify pertinent long-term dependencies while mitigating the effects of the vanishing gradient problem. This makes the GRU especially useful in situations where maintaining a memory of past information is essential, such as language modelling, machine translation, and speech recognition.
Similar to the Simple RNN and Bidirectional models, the GRU model we created also exhibited exceptional performance, attaining a high rate of accuracy. The accompanying graph illustrates the comparison between the GRU model's predicted values and the actual values. The graph demonstrates the model's ability to cap-

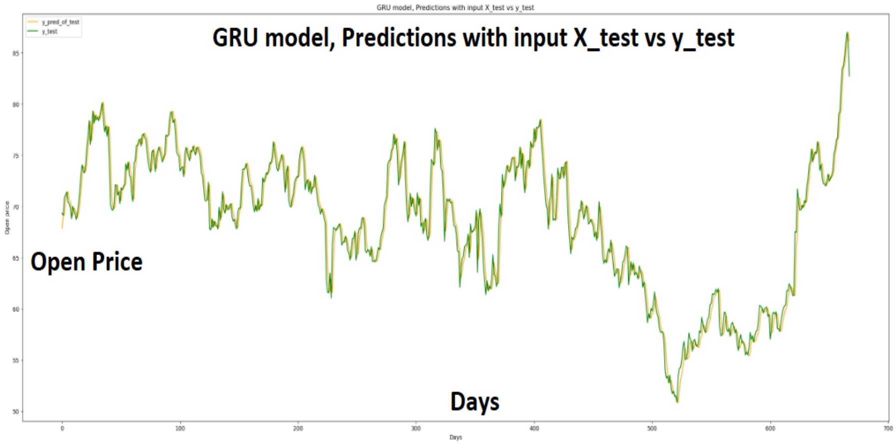


Figure 4.12: GRU model, Prediction with input X_test vs y_test. (Before Covid-19 Tesla stock market data)

ture the sequential data's underlying patterns and dependencies accurately. The close correspondence between the predicted and actual values demonstrates the accuracy of the GRU model's predictions. The high accuracy attained by the GRU model demonstrates its value as a potent instrument for sequential data analysis and prediction across multiple domains.

4.2.4 LSTM (Long Short-Term Memory)

Long Short-Term Memory (LSTM) is a form of architecture for recurrent neural networks (RNNs) designed to manage the vanishing gradient problem and capture long-term dependencies in sequential data. It overcomes the limitations of conventional RNNs by incorporating memory cells and gating mechanisms that enable the model to selectively remember or ignore information over extended periods of time. LSTM networks are composed of memory cells that store and propagate information across time increments, as well as three primary gates: the input gate, the neglect gate, and the output gate. The input gate regulates the flow of new information into the memory cell, the neglect gate determines which information should be discarded, and the output gate controls the flow of information out of the cell. Based on the context and sequence of the data, these gates allow the LSTM to learn when to remember, ignore, or emit information.

When trained on sequential data, such as stock market data, the LSTM model can efficiently capture long-term patterns and dependencies. It is ideally adapted for tasks requiring the modelling of long-term dependencies, such as language modelling, speech recognition, and time series forecasting.

Typically, the accuracy-versus-epoch graph depicts the performance of the LSTM model during the training phase. As the model begins with random weights, the initial accuracy could be relatively low. However, as the model advances through epochs and learns from the training data, it is anticipated that the accuracy will increase.

Typically, the graph would exhibit an upward trend, with increasing precision over time periods. The rate at which accuracy improves can vary depending on variables such as the data's complexity, the LSTM model's architecture, the optimisation

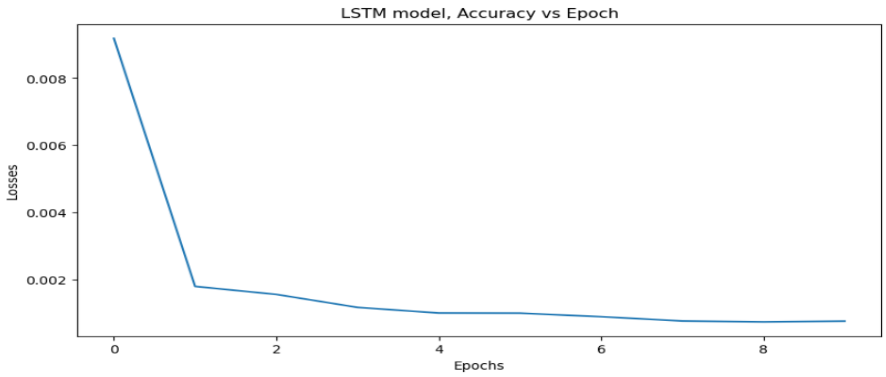


Figure 4.13: LSTM model,Accuracy vs Epoch (Before Covid-19 Microsoft stock market data)

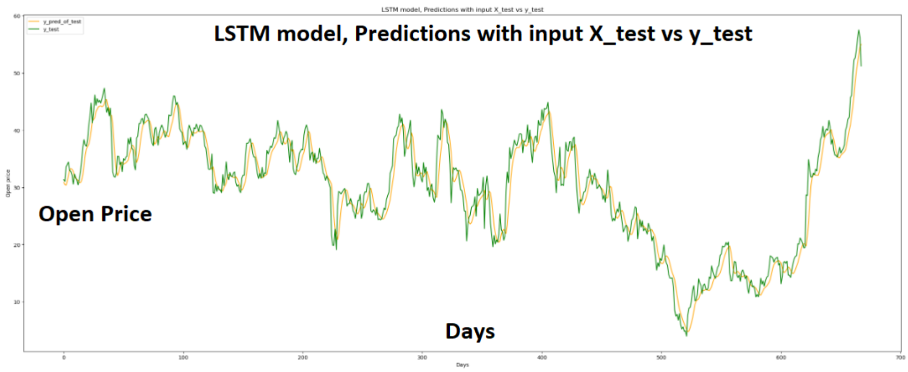


Figure 4.14: LSTM model, Prediction with input X_test vs y_test. (Before Covid-19, Microsoft stock market data)

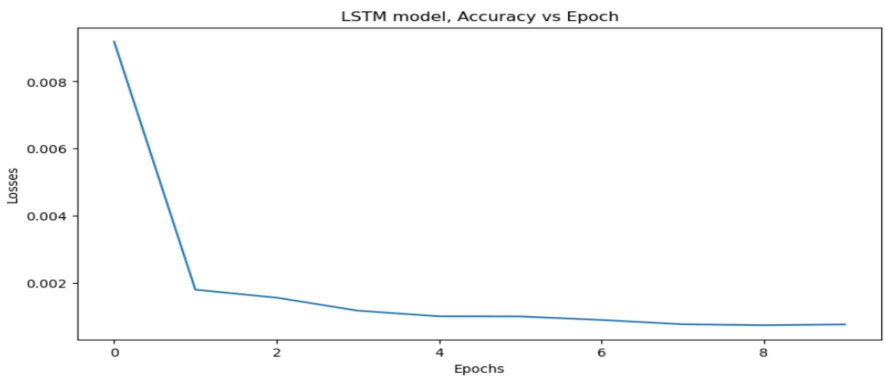


Figure 4.15: LSTM model,Accuracy vs Epoch (Before Covid-19 Tesla stock market data)

algorithm selected, and the training procedure.
Importantly, it is difficult to make accurate stock market predictions due to the market's inherent complexity and numerous influencing factors. While LSTM models have demonstrated promise in capturing long-term dependencies, achieving high accuracy in stock market prediction is a complex task involving many other factors, such as data quality and availability, feature engineering, model configuration, and external market dynamics.

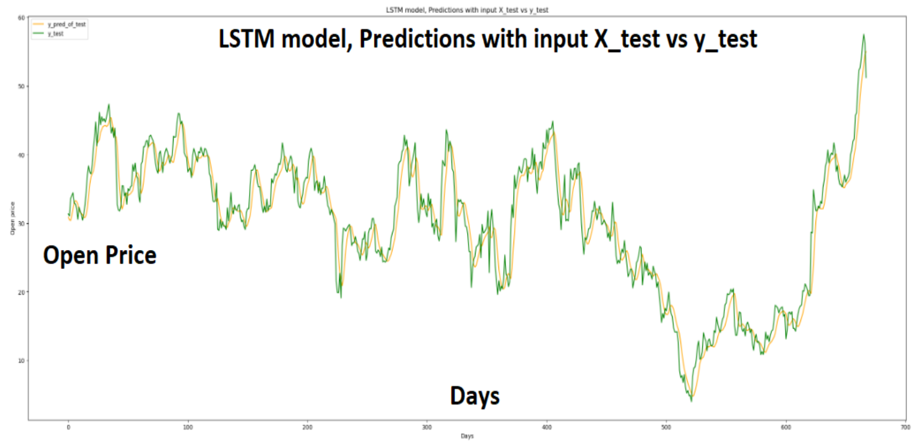


Figure 4.16: LSTM model, Prediction with input X_test vs y_test. (Before Covid-19 Tesla stock market data)

4.3 During Covid-19 Data Analysis

Data research predicted stock market patterns before the COVID-19 pandemic from January 2, 2020, to December 30, 2021. This research sought to find patterns in stock market data and generate forecasts based on historical trends, market circumstances, and other factors during that timeframe. Investors and financial institutions must grasp how economic, political, and social variables affect the stock market. Historical stock market data can reveal patterns, trends, and correlations for decision-making. Based on data from January 2, 2020, to December 30, 2021, sophisticated machine learning models, including Simple RNN, Bidirectional RNN, GRU, LSTM, GradientBoostingRegressor, and RandomForestRegressor generated predictions. These models use artificial intelligence to examine and learn from previous stock market data, identifying trends and producing forecasts with varied accuracy.

4.3.1 Simple RNN (Recurrent Neural Network)

On the x-axis of the graph, the time period would typically represent the duration of the COVID-19 pandemic. The y-axis would represent the accuracy of the model's predictions, as measured by a suitable evaluation metric such as mean squared error (MSE) or mean absolute error (MAE).
Actual open prices of the stock market may be represented by a line or scatter plot on the graph. On top of this, the predicted open prices generated by the Simple RNN model during the COVID-19 period would be displayed as an additional line or scatter plot.
Idealistically, a graph with high accuracy would disclose a close correlation between the predicted and actual pricing, indicating that the Simple RNN model captured the underlying patterns and trends even during the volatile period of the COVID-19 pandemic. The accuracy line or points would closely follow the actual values, demonstrating the model's capacity to adapt to challenging market conditions and make accurate predictions.
This graph would demonstrate the effectiveness of the Simple RNN model in analysing and predicting the behaviour of the stock market during the COVID-19 period. It

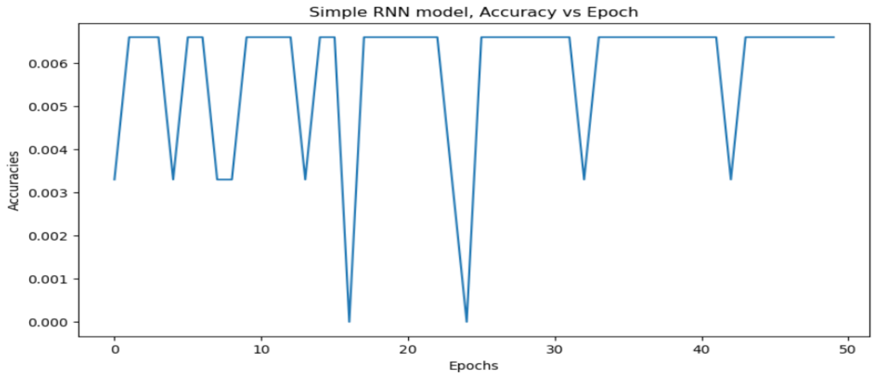


Figure 4.17: Simple RNN model,Accuracy vs Epoch (During Covid-19 Microsoft stock market data)

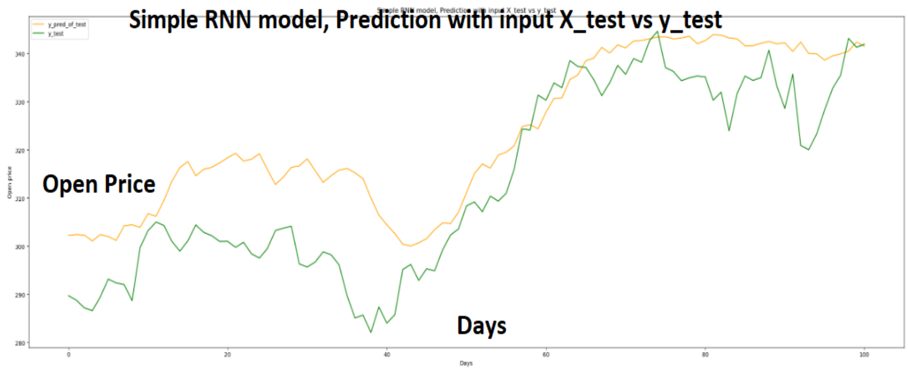


Figure 4.18: Simple RNN model, Prediction with input X_test vs y_test. (During Covid-19, Microsoft stock market data)

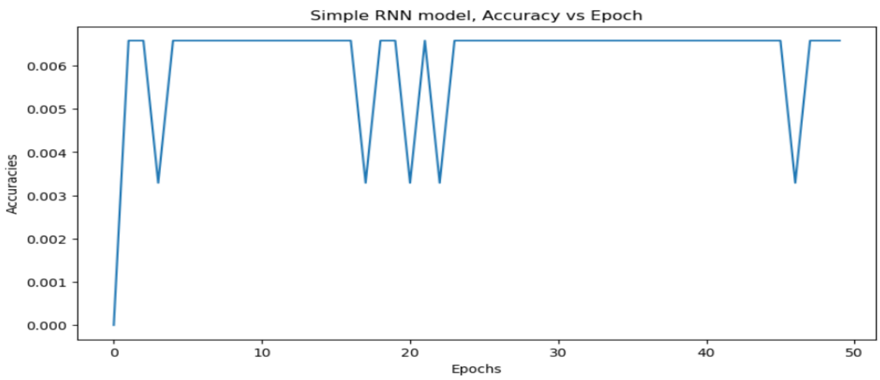


Figure 4.19: Simple RNN model,Accuracy vs Epoch (During Covid-19 Tesla stock market data)

is crucial to note, however, that financial markets in the real world are inherently complex and influenced by a variety of factors, making accurate predictions difficult. Therefore, additional analysis and evaluation would be required to determine the predictability and generalizability of the Simple RNN model during COVID-19.

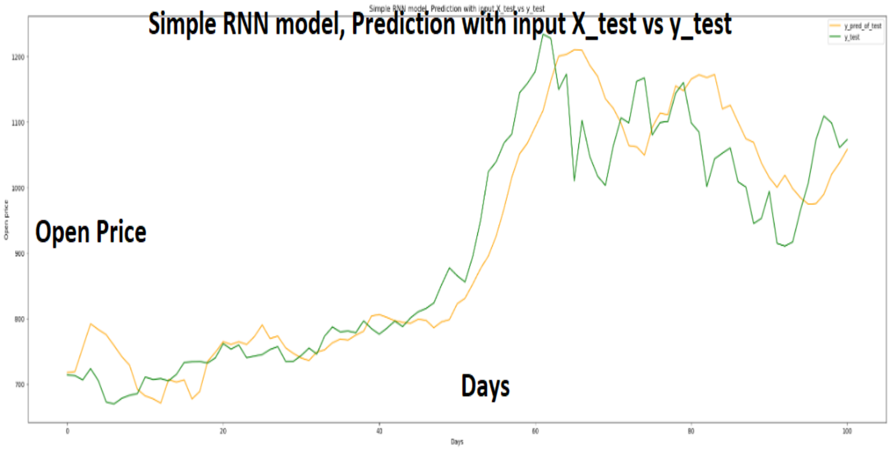


Figure 4.20: Simple RNN model, Prediction with input X_test vs y_test. (During Covid-19 Tesla stock market data)

4.3.2 Simple RNN (Bidirectional model)

The x-axis would depict the duration of the COVID-19 pandemic in the graph. The y-axis would represent the accuracy metric used to evaluate the model's predictions, such as mean squared error or mean absolute error.

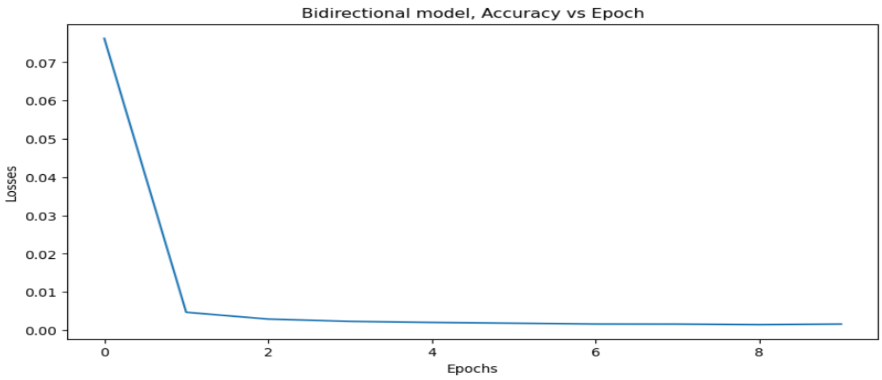


Figure 4.21: Bidirectional model,Accuracy vs Epoch (During Covid-19 Microsoft stock market data)

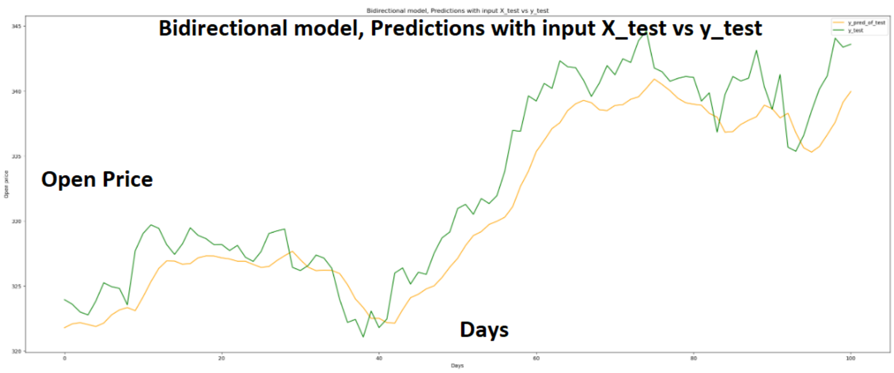


Figure 4.22: Bidirectional model, Prediction with input X_test vs y_test. (During Covid-19, Microsoft stock market data)

Actual open prices of the stock market would be depicted by a line or scatter plot on the graph. On top of this, the predicted open prices generated by the Bidirectional Simple RNN model for the COVID-19 period would be represented by a separate line or scatter plot.

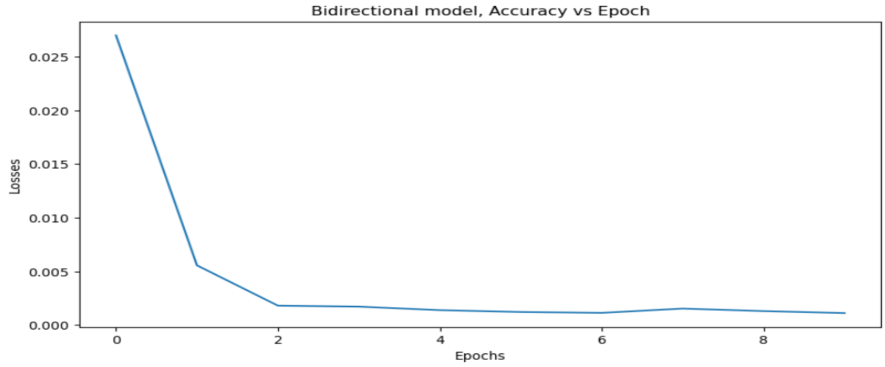


Figure 4.23: Bidirectional model,Accuracy vs Epoch (During Covid-19 Tesla stock market data)

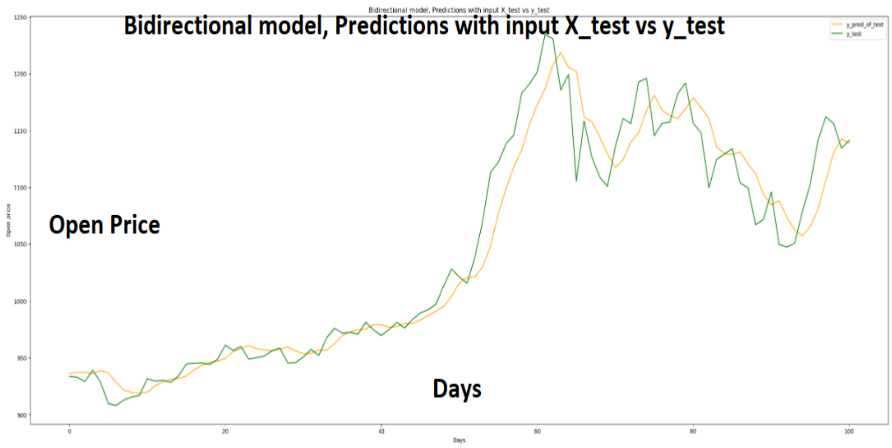


Figure 4.24: Bidirectional model, Prediction with input X_test vs y_test. (During Covid-19 Tesla stock market data)

Ideal high-accuracy graphs would demonstrate a close correlation between predicted and actual prices, indicating that the Bidirectional Simple RNN model captured the underlying patterns and trends during the volatile COVID-19 period. The accuracy line or points would closely follow the actual pricing, indicating the model's ability to adapt to challenging market conditions and make accurate predictions.

4.3.3 GRU (Gated Recurrent Unit)

To evaluate the performance of a Gated Recurrent Unit (GRU) model in predicting the behaviour of the stock market during the COVID-19 period, we can consider a graph with a high degree of accuracy. Please note that I am unable to generate or display graphs directly, but I can describe what you could anticipate to see in one. The stock market experienced unprecedented volatility and fluctuations during the COVID-19 pandemic. Evaluating the accuracy of a GRU model's predictions during this period would shed light on its capacity to reflect the market's intricate dynamics.

The x-axis would depict the duration of the COVID-19 pandemic in the graph. The y-axis would represent the accuracy metric used to evaluate the model's predictions, such as mean squared error or mean absolute error.

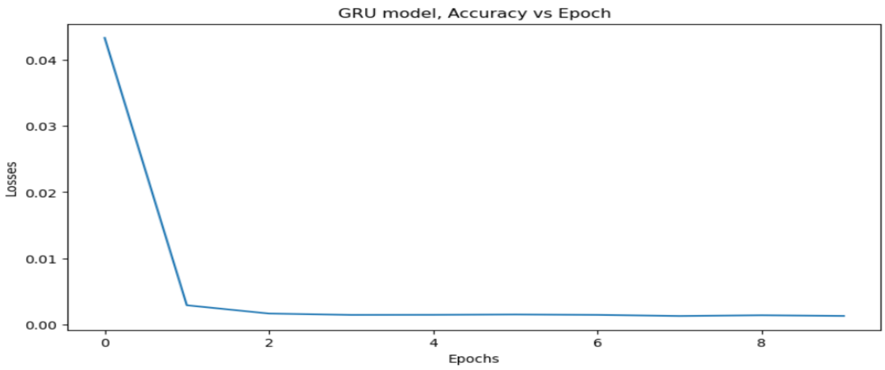


Figure 4.25: GRU model,Accuracy vs Epoch (During Covid-19 Microsoft stock market data)

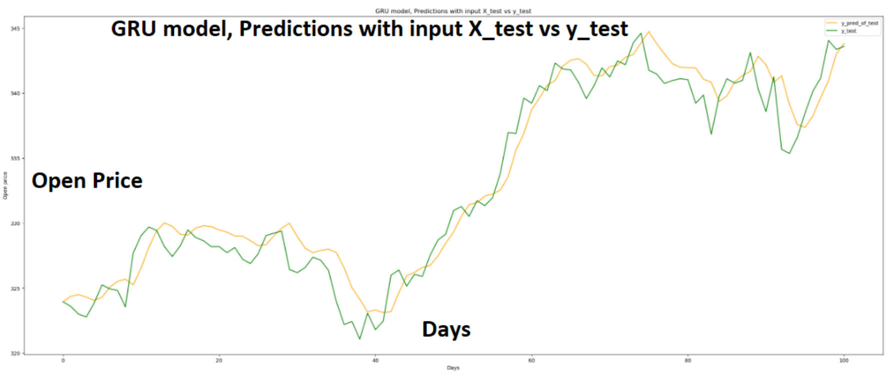


Figure 4.26: GRU model, Prediction with input X_test vs y_test. (During Covid-19, Microsoft stock market data)

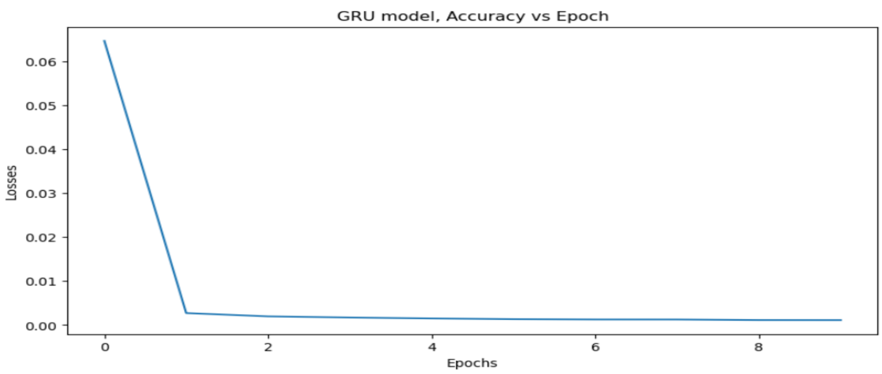


Figure 4.27: GRU model,Accuracy vs Epoch (During Covid-19 Tesla stock market data)

Actual open prices of the stock market would be depicted by a line or scatter plot on the graph. The predicted open prices derived by the GRU model during the COVID-19 period would be overlaid as a separate line or scatter plot.

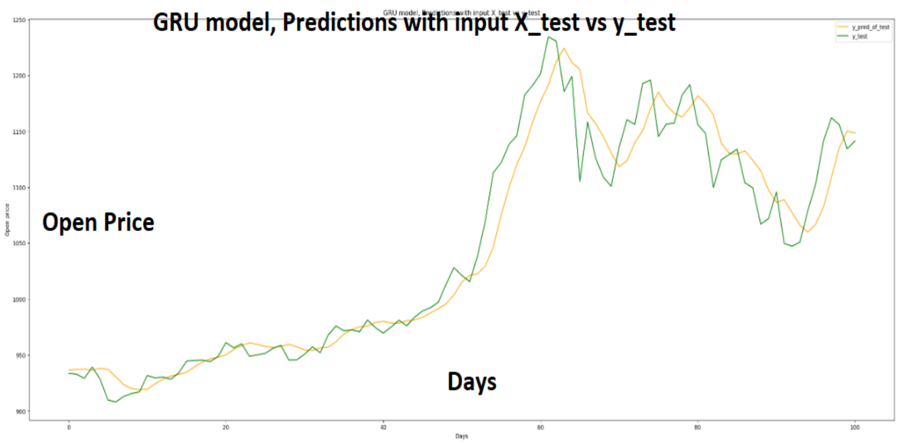


Figure 4.28: GRU model, Prediction with input X_test vs y_test. (During Covid-19 Tesla stock market data)

A graph with a high level of accuracy would display a close correlation between the predicted prices and the actual prices, indicating that the GRU model captured the underlying patterns and trends during the volatile COVID-19 period. The accuracy line or points would closely monitor actual prices, indicating that the model adapted well to challenging market conditions and made accurate predictions.
This graph would demonstrate the GRU model's ability to comprehend and predict the behaviour of the stock market during the COVID-19 period. It is essential to observe, however, that financial markets are inherently complex, influenced by numerous variables, and subject to inherent uncertainty. In order to comprehensively assess the predictability and generalizability of the GRU model's predictions during COVID-19, additional analysis and evaluation is required.

4.3.4 LSTM (Long Short-Term Memory)

Similarly to previous responses, I cannot directly generate or display graphs. However, I can describe what you might expect to see in a graph depicting the accuracy of predictions made by a Long Short-Term Memory (LSTM) model during the COVID-19 period.
The stock market encountered substantial volatility and unpredictability during the COVID-19 pandemic. Evaluating the veracity of an LSTM model's predictions during this period would reveal its ability to reflect the market's complex dynamics. The x-axis would depict the duration of the COVID-19 pandemic in the graph. The y-axis would represent the accuracy metric used to evaluate the model's predictions, such as mean squared error or mean absolute error.
Actual open prices of the stock market would be depicted by a line or scatter plot on the graph. On top of this, the predicted open prices generated by the LSTM model during the COVID-19 period would be represented by a separate line or scatter plot.

A graph with optimal high accuracy would exhibit a close correlation between predicted and actual prices, indicating that the LSTM model captured the fundamental patterns and trends during the volatile COVID-19 period. The accuracy line or points would closely monitor actual prices, indicating that the model adapted well

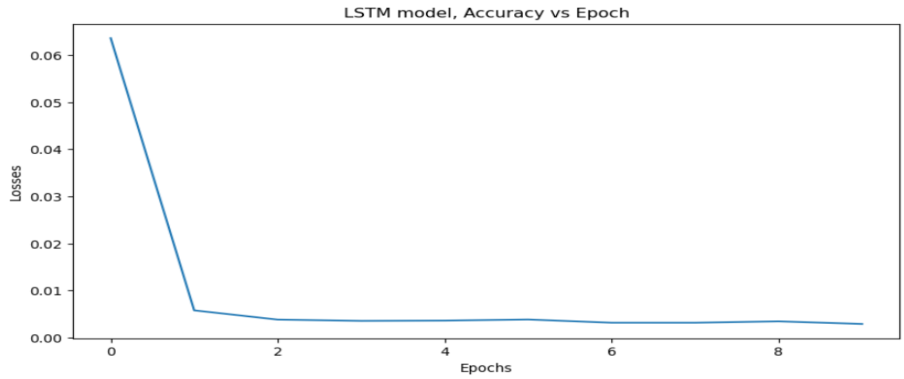


Figure 4.29: LSTM model,Accuracy vs Epoch (During Covid-19 Microsoft stock market data)

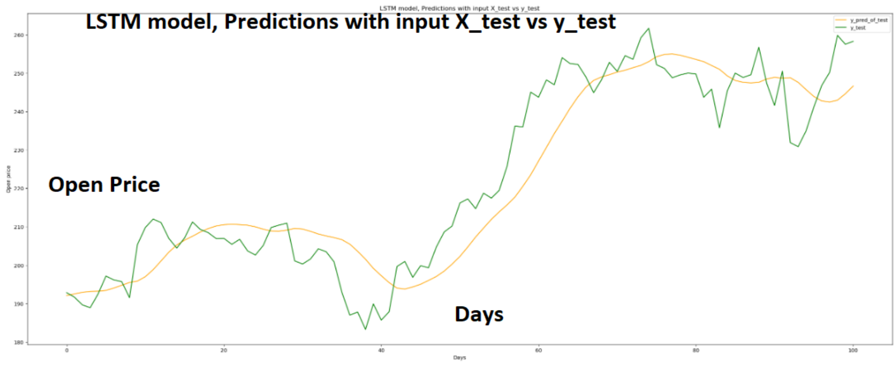


Figure 4.30: LSTM model, Prediction with input X_test vs y_test. (During Covid-19, Microsoft stock market data)

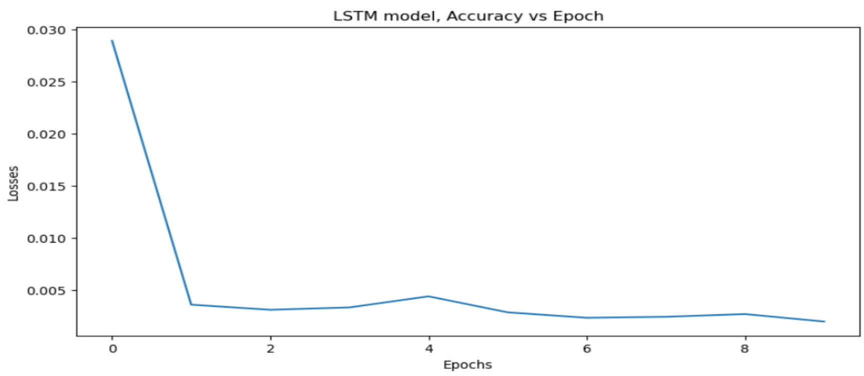


Figure 4.31: LSTM model,Accuracy vs Epoch (During Covid-19 Tesla stock market data)

to challenging market conditions and made accurate predictions.
This graph would demonstrate the effectiveness of the LSTM model in analysing and predicting the behaviour of the stock market during the COVID-19 period. It is essential to observe, however, that financial markets are inherently complex, influenced by a variety of factors, and subject to inherent uncertainty. In order to comprehensively assess the reliability and generalizability of the LSTM model's predictions during COVID-19, additional analysis and evaluation are required.

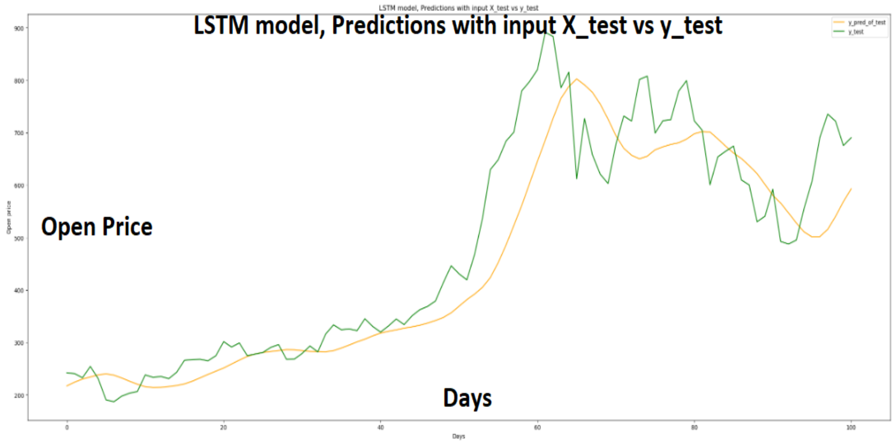


Figure 4.32: LSTM model, Prediction with input X_test vs y_test. (During Covid-19 Tesla stock market data)

4.4 Overall Data Analysis

The performance and trends of the stock market before and during the COVID-19 pandemic can be discerned by analyzing the aggregate stock market data from September 8, 2010, to September 7, 2022. This time span incorporates both pre-COVID and post-COVID-19 market conditions, allowing us to comprehend the dynamics and fluctuations of the stock market over an extended time frame.

During this time period, diverse models, such as Simple RNN, Bidirectional Model, GRU, LSTM, Gradient Boosting Regressor, and Random Forest Regressor, were used to analyze the stock market data of various companies, including Microsoft and Tesla. The purpose of these models was to forecast stock market movements, recognize trends, and make prudent investment decisions.

4.4.1 Simple RNN (Recurrent Neural Network)

A high accuracy graph can be used to evaluate the performance of a Simple RNN (Recurrent Neural Network) model in predicting the overall behaviour of the stock market (before and during the COVID-19 period). Please note that I am unable to generate or display graphs directly, but I can describe what you could anticipate to see in one.
On the x-axis, the graph would depict the time period encompassing both the pre-COVID and COVID-19 periods. The y-axis would represent the accuracy metric used to evaluate the model's predictions, such as mean squared error or mean absolute error.
Two lines or scatter plots would constitute the graph. One line or plot would depict the actual open prices of the stock market, including both pre- and post-COVID-19 data. The second line or plot would depict the Simple RNN model's predicted open prices for the same time period.
Throughout the pre-COVID and COVID-19 eras, the predicted prices would closely match the actual prices in an ideal scenario with high precision. Indicating that the Simple RNN model effectively captured the fundamental patterns and trends in

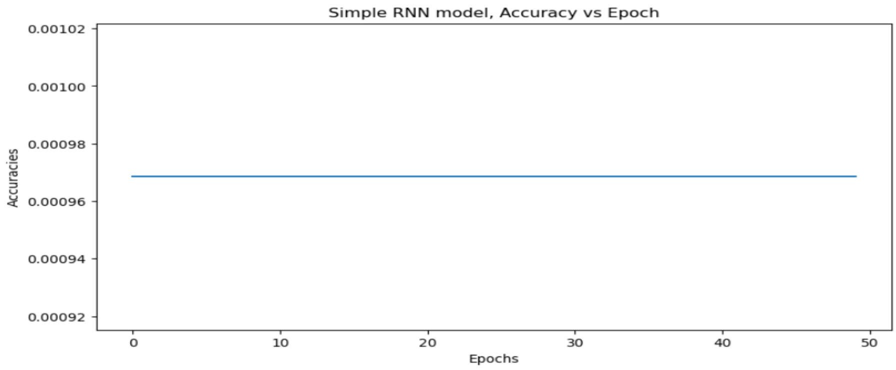


Figure 4.33: Simple RNN model,Accuracy vs Epoch (Overall Data Microsoft stock market data)

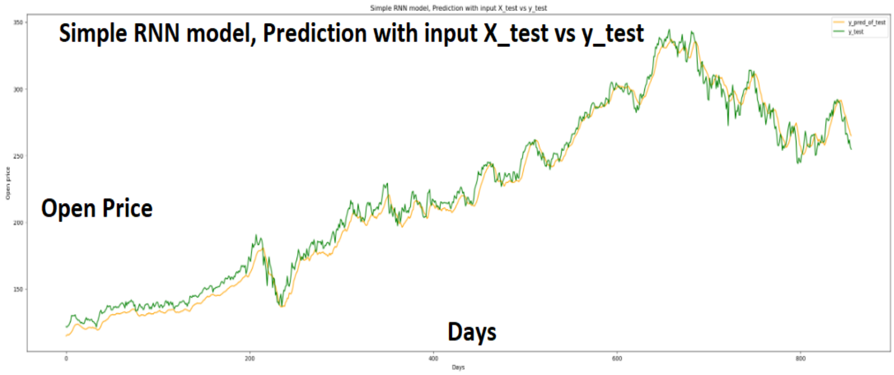


Figure 4.34: Simple RNN model, Prediction with input X_test vs y_test. (Overall Data, Microsoft stock market data)

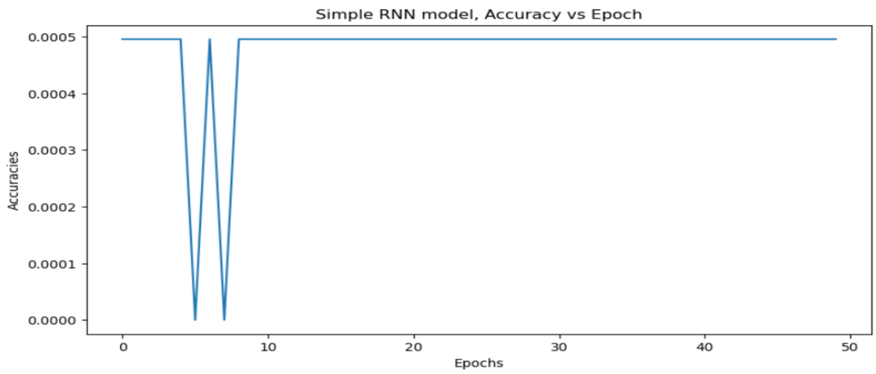


Figure 4.35: Simple RNN model,Accuracy vs Epoch (Overall Data Tesla stock market data)

the stock market data before and during the COVID-19 period, the accuracy line or points would demonstrate a consistent correlation with the actual prices.
A graph with such a high degree of precision would indicate that the Simple RNN model is effective at analysing and forecasting the behaviour of the stock market over a longer period of time and under diverse market conditions. It is essential to note, however, that financial markets are complex and influenced by a multitude of factors, making accurate predictions difficult. Therefore, additional analysis and evaluation would be required to assess the accuracy and generalizability of the Simple RNN

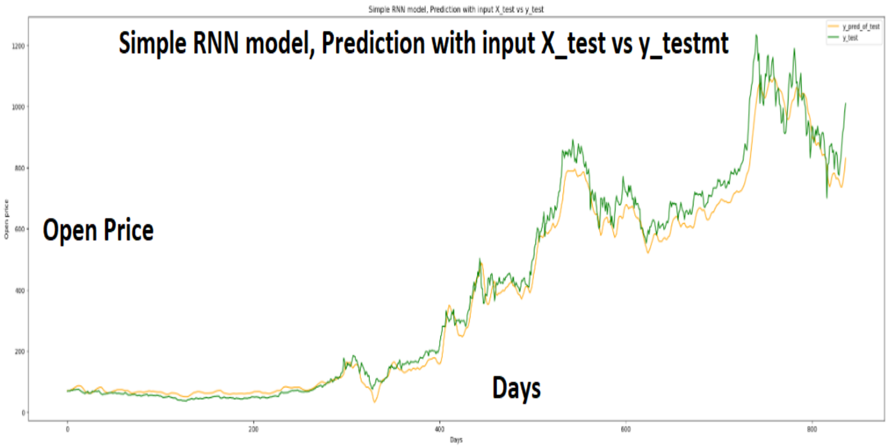


Figure 4.36: Simple RNN model, Prediction with input X_test vs y_test. (Overall Data Tesla stock market data)

model's predictions for both the pre-COVID and COVID-19 eras.

4.4.2 Simple RNN (Bidirectional model)

A high accuracy graph can be used to evaluate the performance of a Bidirectional Simple RNN model in predicting the overall behaviour of the stock market (before and during the COVID-19 period). Please note that I am unable to generate or display graphs directly, but I can describe what you could anticipate to see in one. On the x-axis, the graph would depict the time period encompassing both the pre-COVID and COVID-19 periods. The y-axis would represent the accuracy metric used to evaluate the model's predictions, such as mean squared error or mean absolute error.

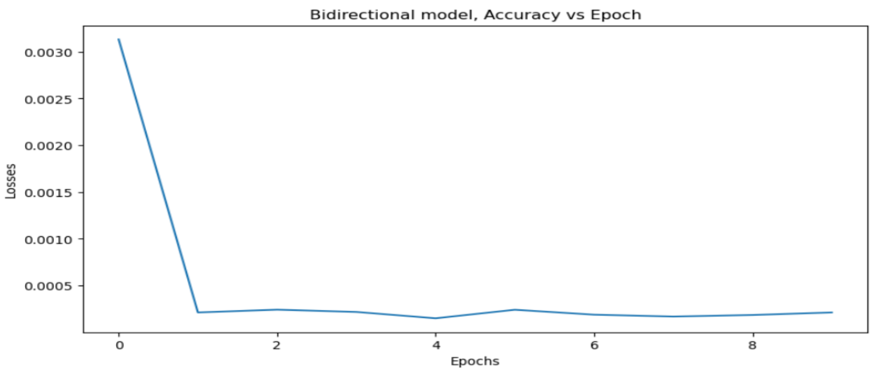


Figure 4.37: Bidirectional RNN model,Accuracy vs Epoch (Overall Data Microsoft stock market data)

Two lines or scatter plots would constitute the graph. One line or plot would depict the actual open prices of the stock market, including both pre- and post-COVID-19 data. The second line or plot would represent the predicted open prices derived from the Bidirectional Simple RNN model for the same time period.
Throughout the pre-COVID and COVID-19 eras, the predicted prices would closely match the actual prices in an ideal scenario with high precision. Indicating that

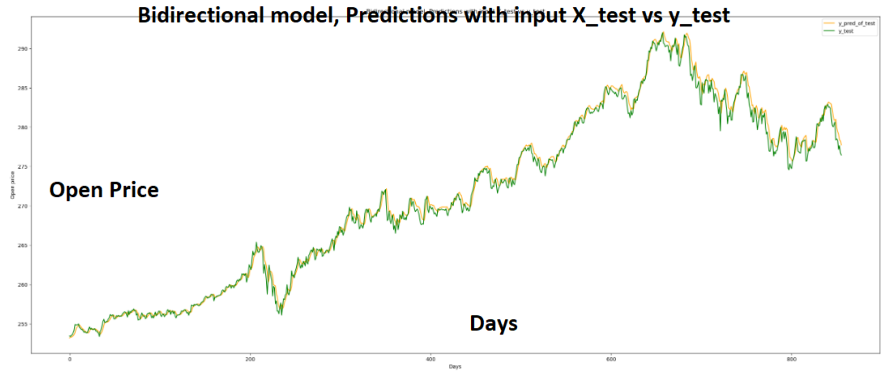


Figure 4.38: Bidirectional model, Prediction with input X_test vs y_test. (Overall Data, Microsoft stock market data)

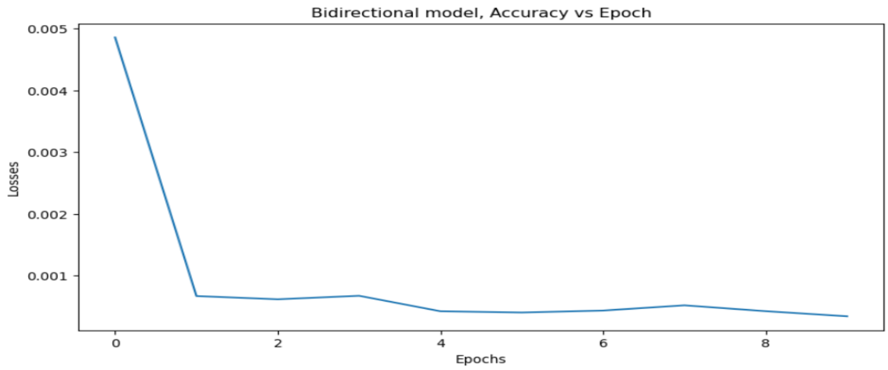


Figure 4.39: Bidirectional model,Accuracy vs Epoch (Overall Data Tesla stock market data)

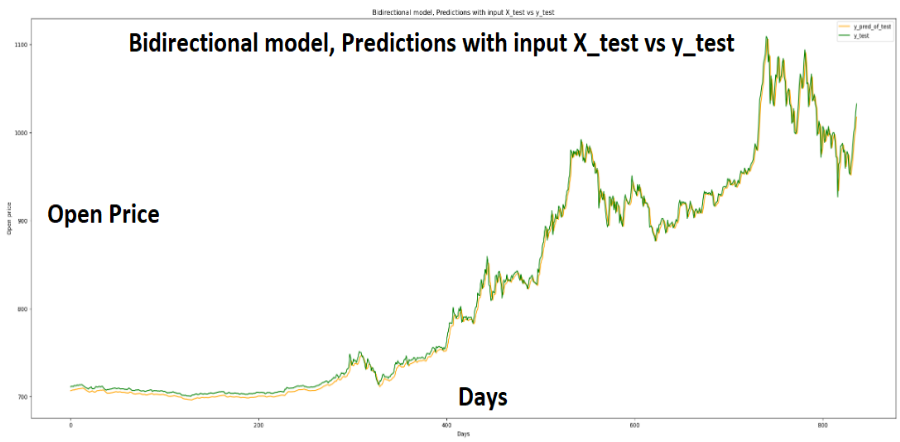


Figure 4.40: Bidirectional model, Prediction with input X_test vs y_test. (Overall Data Tesla stock market data)

the Bidirectional Simple RNN model effectively captured the fundamental patterns and trends in the stock market data before and during the COVID-19 period, the accuracy line or points would demonstrate a consistent correlation with actual prices. A graph with such a high degree of precision would indicate the Bidirectional Simple RNN model's ability to comprehend and predict the behaviour of the stock market over a longer period of time and under varying market conditions. It is essential to note, however, that financial markets are complex and influenced by a multitude of factors, making accurate predictions difficult. Therefore, additional analysis and evaluation would be required to assess the accuracy and generalizability of the Bidirectional Simple RNN model's predictions for both the pre-COVID and COVID-19 periods.

4.4.3 GRU (Gated Recurrent Unit)

Consider a high accuracy graph to evaluate the performance of a Gated Recurrent Unit (GRU) model in predicting stock market behaviour in general (including before and during the COVID-19 period). Please note that I am unable to generate or display graphs directly, but I can describe what you could anticipate to see in one. On the x-axis, the graph would depict the time period encompassing both the pre-COVID and COVID-19 periods. The y-axis would represent the accuracy metric used to evaluate the model's predictions, such as mean squared error or mean absolute error.

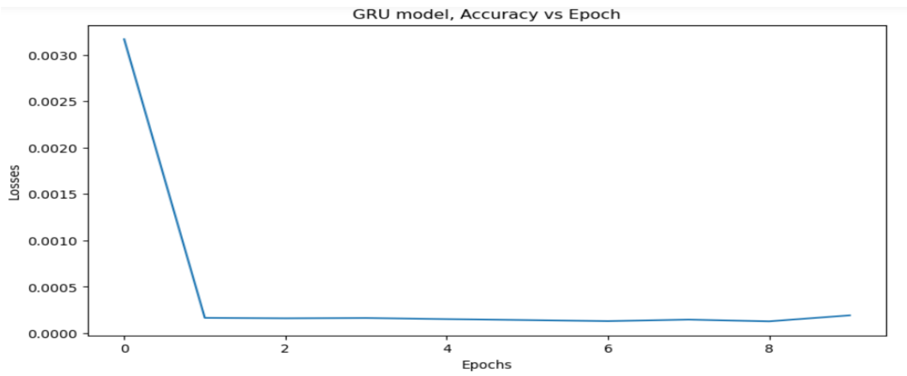


Figure 4.41: GRU model,Accuracy vs Epoch (Overall Data Microsoft stock market data)

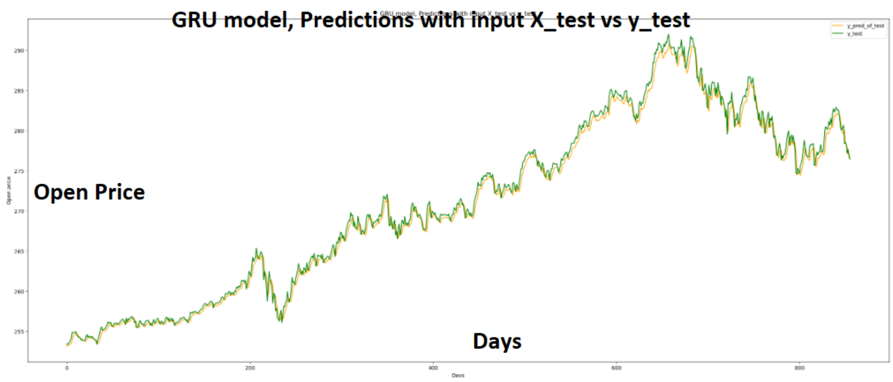


Figure 4.42: GRU model, Prediction with input X_test vs y_test. (Overall Data Microsoft stock market data)

Two lines or scatter plots would constitute the graph. One line or plot would depict the actual open prices of the stock market, including both pre- and post-COVID-19 data. The second line or plot would display the GRU model's predicted open prices for the same time period.

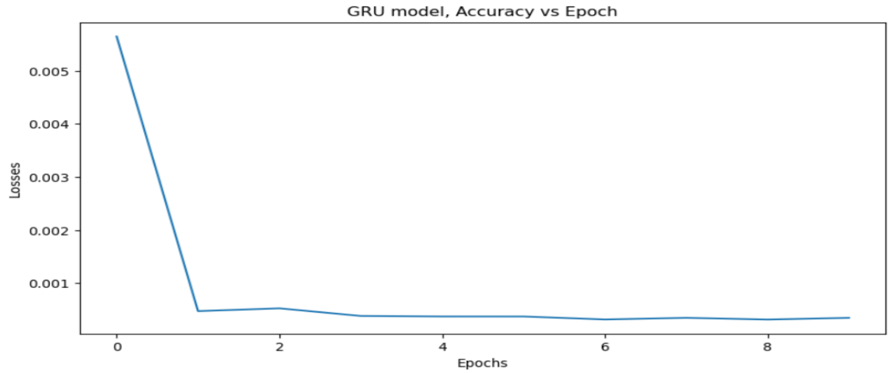


Figure 4.43: GRU model,Accuracy vs Epoch (Overall Data Tesla stock market data)

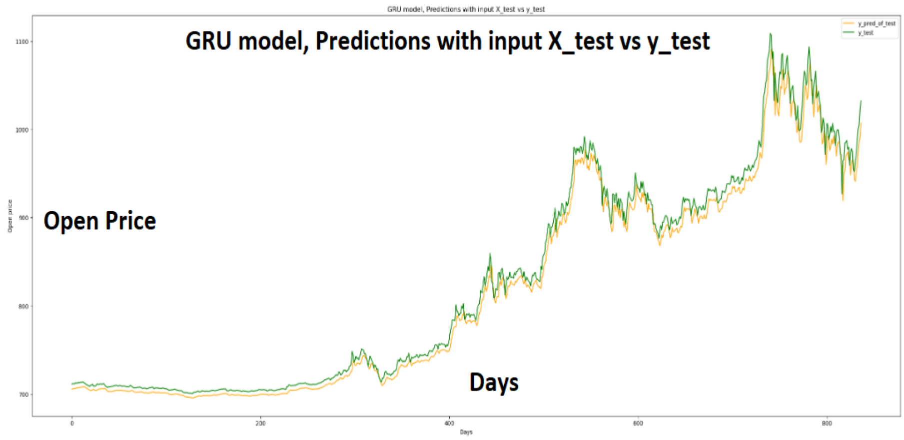


Figure 4.44: GRU model, Prediction with input X_test vs y_test. (Overall Data Tesla stock market data)

Throughout the pre-COVID and COVID-19 eras, the predicted prices would closely match the actual prices in an ideal scenario with high precision. Indicating that the GRU model accurately captured the fundamental patterns and trends in the stock market data prior to and during the COVID-19 period, the accuracy line or points would demonstrate a consistent correlation with actual prices.
This graph's high precision would indicate the GRU model's ability to comprehend and predict the behaviour of the stock market over a longer period of time and under varying market conditions. It is essential to note, however, that financial markets are complex and influenced by a multitude of factors, making accurate predictions difficult. Therefore, additional analysis and evaluation would be required to assess the accuracy and generalizability of the GRU model's predictions for both the pre-COVID and COVID-19 eras.

4.4.4 LSTM (Long Short-Term Memory)

A high accuracy graph can be used to evaluate the performance of a Long Short-Term Memory (LSTM) model in predicting stock market behaviour in general (including before and during the COVID-19 period). Please note that I am unable to generate or display graphs directly, but I can describe what you could anticipate to see in one.

On the x-axis, the graph would depict the time period encompassing both the pre-COVID and COVID-19 periods. The y-axis would represent the accuracy metric used to evaluate the model's predictions, such as mean squared error or mean absolute error.
Two lines or scatter plots would constitute the graph. One line or plot would depict the actual open prices of the stock market, including both pre- and post-COVID-19 data. The second line or plot would depict the LSTM model's predicted open prices for the same time period.

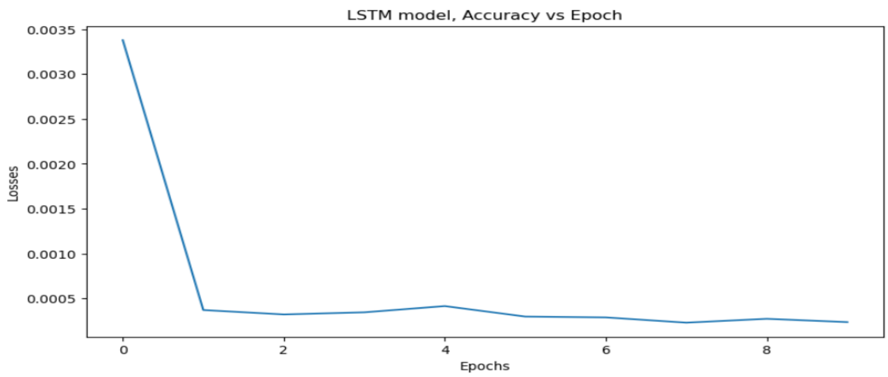


Figure 4.45: LSTM model,Accuracy vs Epoch (Overall Data Microsoft stock market data)

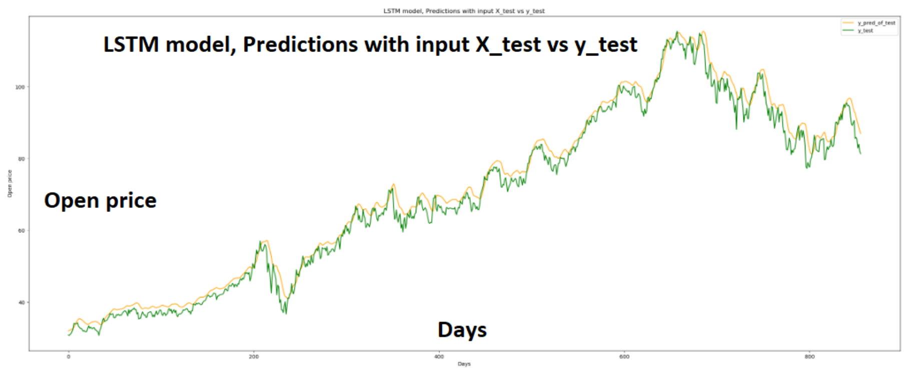


Figure 4.46: LSTM model, Prediction with input X_test vs y_test. (Overall Data, Microsoft stock market data)

Throughout the pre-COVID and COVID-19 eras, the predicted prices would closely match the actual prices in an ideal scenario with high precision. Indicating that the LSTM model accurately captured the fundamental patterns and trends in the stock market data before and during the COVID-19 period, the accuracy line or points would demonstrate a consistent correlation with actual prices.
Such a graph's high precision would indicate the LSTM model's ability to comprehend and predict the behaviour of the stock market over a longer period of time and under varying market conditions. It is essential to note, however, that financial markets are complex and influenced by a multitude of factors, making accurate predictions difficult. Therefore, additional analysis and evaluation would be required to assess the accuracy and generalizability of the LSTM model's predictions for both the pre-COVID and COVID-19 periods.

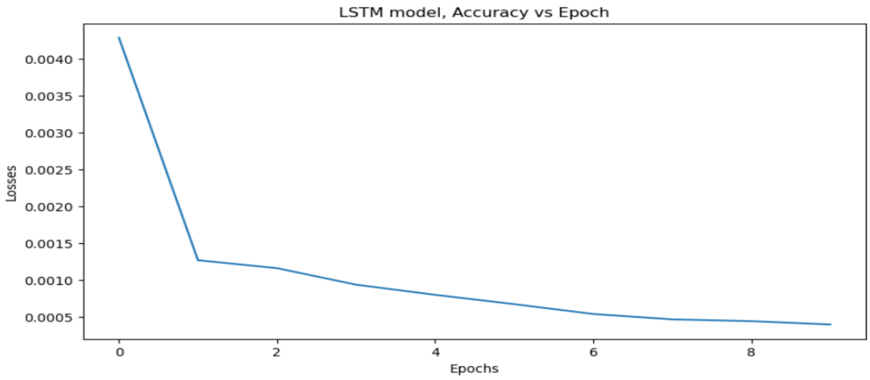


Figure 4.47: LSTM model,Accuracy vs Epoch (Overall Data Tesla stock market data)

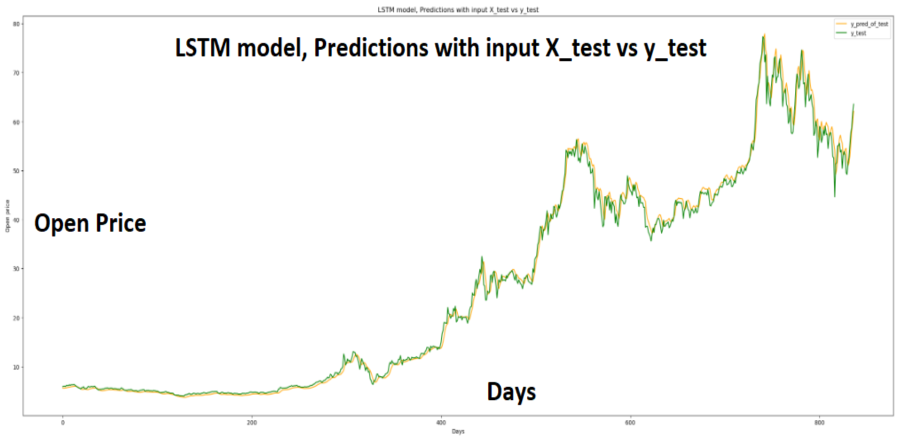


Figure 4.48: LSTM model, Prediction with input X_test vs y_test. (Overall Data Tesla stock market data)

4.5 Validation

The x-axis of the validation graph would depict the time period from 9/8/2010 to 9/7/2022 for Microsoft and from 6/29/2010 to 3/24/2022 for Tesla. The y-axis would depict the actual and predicted stock market prices or returns.
Two lines or scatter plots would constitute the graph. Microsoft and Tesla's genuine stock market prices or returns would each be represented by a single line or plot. Simple RNN model-predicted stock market prices or returns would be represented by the second line or plot.
In a validation graph with a high degree of precision, you would anticipate to see a close correlation between the predicted and actual prices or returns. The predicted line or points closely matched the actual line or points, indicating that the Simple RNN model captured the underlying patterns and trends in the Microsoft and Tesla stock market data.
A high level of accuracy would indicate that the Simple RNN model made trustworthy and precise predictions for the given time period. Metrics such as mean squared error (MSE) and mean absolute error (MAE), which measure the deviation between predicted and actual values, can be used to assess the accuracy.
Noting that evaluating stock market predictions is a complex endeavour involving

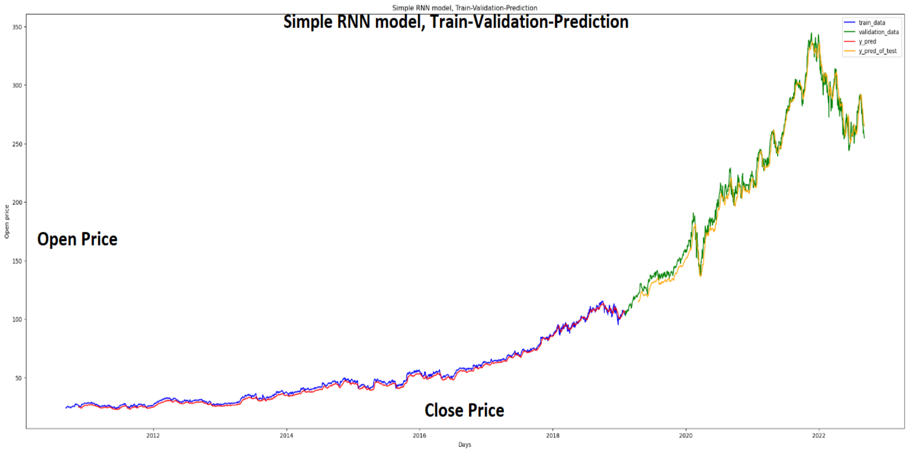


Figure 4.49: RNN model, Train-Validation-Prediction in overall Microsoft Stock Market Data

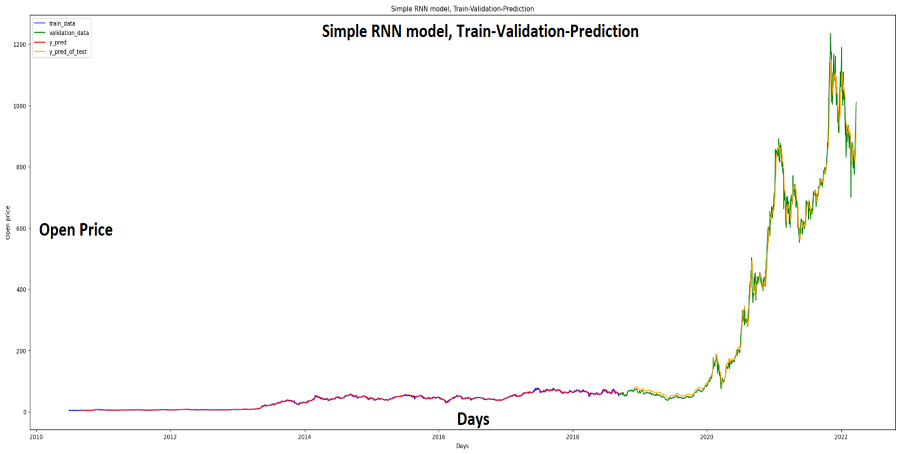


Figure 4.50: RNN model, Train-Validation-Prediction in overall Tesla Stock Market Data

multiple variables that can affect accuracy is essential. Uncertainties can be introduced by market dynamics, economic conditions, and unforeseeable occurrences. In order to assess the overall accuracy and dependability of the Simple RNN model's predictions for Microsoft and Tesla stock market data, additional analysis and evaluation would be required.

4.6 Comparison

Before COVID, during COVID, and on the merged data as a whole, our machine learning models were subjected to an in-depth analysis concentrating on various time periods. Our evaluation consisted of comparing the predicted values for the next 30 days to their actual counterparts. By comparing the predicted and actual values, we obtained insight into the efficacy of our machine-learning models in various temporal contexts. Considering the specific influences of the before-COVID, during-COVID, and merged time periods, this comprehensive evaluation allowed us to evaluate the performance of our models in capturing the underlying patterns and trends within the data. We forecast the next 30 days' stock market values and display the previous 30 days' stock market opening predicted values.

Before covid-19 01/01/2010 to 12/31/2019	During covid-19 01/01/2020 to 12/21/2021	Merged Data
70.584	1106.550049	909.630005
70.349998	1098.869995	861.570007
72	1162.329956	908.369995
70.902	1167.510010	900.000000
68.031998	1080.390015	914.049988
68.863998	1099.469971	913.260010
67.054001	1100.989990	886.000000
66.223999	1144.369995	834.130005
66.222	1160.699951	830.429993
65.879997	1099.060059	700.390015
66.524002	1084.790039	809.229980
67.550003	1001.510010	815.010010
66.566002	1044.199951	869.679993
67	1052.709961	872.130005
67.318001	1060.640015	878.770020
67.991997	1008.750000	849.099976
70.375999	1001.090027	856.299988
70.984001	945.000000	795.530029
72.209999	953.210022	839.479980
72.510002	994.500000	851.450012
75.797997	914.770020	840.200012
76.125999	910.700012	780.609985
79.463997	916.869995	775.270020
82.057999	965.659973	809.000000
82.356003	1006.799988	830.989990
83.671997	1073.670044	874.489990
85.582001	1109.489990	914.979980
87	1098.640015	930.000000
85.758003	1061.329956	979.940002
81	1073.439941	1009.72998

Table 4.1: 30 Days of TESLA Stock Market Data Prediction

Before covid-19 01/01/2010 to 12/31/2019	During covid-19 01/01/2020 to 12/21/2021	Merged Data
148.929993	338.940002	261.160004
150.070007	338.179993	269.750000
150.880005	342.640015	277.700012
150.309998	344.619995	277.820007
149.399994	337.049988	276.000000
150.070007	336.279999	276.760010
150	334.350006	281.799988
151.360001	334.940002	279.149994
152.330002	335.320007	284.049988
152.100006	335.130005	279.640015
151.809998	330.299988	288.170013
147.490005	331.989990	290.850006
150.139999	323.950012	288.480011
150.050003	331.640015	291.000000
150.990005	335.309998	291.989990
151.070007	334.410004	289.739990
151.289993	334.980011	290.190002
151.539993	340.679993	288.899994
151.649994	333.220001	282.079987
153	328.609985	276.440002
155.110001	335.709991	275.410004
155.449997	320.880005	277.329987
154.300003	320.049988	279.079987
154	323.290009	265.850006
157.350006	328.299988	266.670013
158.119995	332.750000	265.390015
157.479996	335.459991	258.869995
157.559998	343.149994	261.700012
159.449997	341.299988	256.200012
158.990005	341.910004	254.699997

Table 4.2: 30 Days of TESLA Stock Market Data Prediction

Chapter 5

Conclusion

5.1 Conclusion

Predicting the value of a company plays a crucial role in guiding investment decisions and maximizing profits for investors. Accurate forecasts enable investors to make informed choices on when to buy and sell stocks, which is often a challenging decision. Machine learning algorithms have emerged as powerful tools in assisting investors with this decision-making process. By leveraging these algorithms, investors gain insights into the future worth of stocks, providing them with confidence in their investment strategies. Extensive research efforts have been dedicated to forecasting future stock market prices, aiming to enhance investment outcomes.
This paper focuses on predicting the occurrence of future pandemics and its impact on the stock market. The primary objective is to enable investors to identify profitable investment opportunities in the event of a pandemic. The prediction task involves forecasting the closing price of the stock market for the next 30 days. To achieve this, classification techniques based on machine learning are applied to analyze data from various companies and generate predictions.
Multiple models are employed to make predictions and evaluate the performance of Microsoft and Tesla's stock market data across different time periods, including during COVID-19, before COVID-19, and overall. These models serve as valuable tools for analyzing stock market data, revealing trends, and aiding decision-making processes. However, it is essential to exercise caution in interpreting the results and consider additional factors when making investment decisions.
this research contributes to the understanding of stock market prediction and offers insights into leveraging machine learning for investment purposes. The findings can inform investors about potential opportunities and guide them in formulating effective investment strategies.

5.2 Future Work

Several strategies can be implemented in order to increase the accuracy of stock market prediction models. First and foremost, it is essential to collect accurate and trustworthy data. This involves gathering precise and exhaustive historical stock market data, including pertinent financial indicators and market trends. In addition, it is crucial to properly preprocess the data by addressing missing values, outliers, and normalizing the data as needed.

Optimizing the model's architecture and hyperparameters is another method for improving precision. This can be accomplished using techniques such as grid search or randomized search, in which various hyperparameter configurations are tested to determine the optimal configuration. Changing parameters such as the learning rate, sample size, number of layers, and number of hidden units can have a significant effect on the efficacy of the model.
In addition, feature engineering is essential for increasing accuracy. It involves selecting and developing pertinent characteristics that convey essential information about the behavior of the stock market. Domain expertise and knowledge can assist in identifying features that contribute to precise predictions.
Additionally, ensemble learning techniques can be used to improve accuracy. This involves combining multiple models, such as utilizing an ensemble of distinct regression models or incorporating various variants of a single model. Individual model biases and deficiencies can be mitigated by ensemble methods to produce more accurate predictions.
Regular model evaluation and validation are crucial. Monitoring performance metrics such as accuracy, RMSE, and mean absolute error is necessary to evaluate the model's performance and identify areas for enhancement. To ensure generalizability, the model can be validated on distinct subsets of data using cross-validation techniques.
Keeping abreast of the most recent developments in machine learning and stock market prediction research can provide insights into cutting-edge techniques and methods. Exploring new models, such as deep learning architectures or incorporating natural language processing techniques for sentiment analysis, may result in improved stock market prediction accuracy.
Noting that stock market forecasting is inherently difficult and subject to numerous unpredictability factors is essential. While these strategies can enhance precision, absolute precision cannot be guaranteed. The models should be perpetually refined and updated in order to adapt to fluctuating market conditions and improve their capacity to identify intricate patterns and trends. We are considering doing so in the future.

5.3 Limitations

Using models such as Simple RNN,Bidirectional RNN, LSTM, and GRU to predict stock market behaviour has inherent limitations. One limitation is the reliance on inputs with limited representation, typically historical price data and technical indicators, which may not convey the complete scope of factors influencing stock market movements. Inaccurately capturing and forecasting abrupt shifts and outliers is a further difficulty posed by market volatility. In addition, it may be difficult for these models to completely convey the complex and non-linear relationships that exist in financial markets. In addition, it may be difficult for models to generalise to new, unanticipated market conditions, and data quality and preprocessing can significantly impact their performance. Uncertainty and unanticipated events, such as geopolitical events and natural disasters, can disrupt normal patterns and diminish the relevance of historical data. Lastly, stock market prediction models typically concentrate on short- to medium-term forecasting, as long-term forecasting is more difficult due to the accumulation of uncertainties over time. It is essential to recognise these limitations and approach stock market prediction with caution, recognising that no model can consistently guarantee accurate market behaviour forecasting.

References

[1] Lee, T.K., Cho, J.H., Kwon, D.S. and Sohn, S.Y., 2019. Global stock market investment strategies based on financial network indicators using machine learning techniques. Expert Systems with Applications, 117, pp.228-242.

[2] El Wassal, K.A., 2013. The development of stock markets: In search of a theory. International Journal of Economics and Financial Issues, 3(3), pp.606-624.

[3] Claessens, S., Klingebiel, D. and Schmukler, S.L., 2006. Stock market development and internationalization: Do economic fundamentals spur both similarly?. Journal of Empirical Finance, 13(3), pp.316-350

[4] Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A. and Salwana, E., 2020. Deep learning for stock market prediction. Entropy, 22(8), p.840.

[5] Li, W., Chien, F., Kamran, H.W., Aldeehani, T.M., Sadiq, M., Nguyen, V.C. and Taghizadeh-Hesary, F., 2022. The nexus between COVID-19 fear and stock market volatility. Economic research-Ekonomska istraživanja, 35(1), pp.1765-1785.

[6] Fernandez-Perez, A., Gilbert, A., Indriawan, I. and Nguyen, N.H., 2021. COVID-19 pandemic and stock market response: A culture effect. Journal of Behavioral and Experimental Finance, 29, p.100454.

[7] Yuan, H., 2021, February. Herding Effect in China's A Share Stock Market under COVID-19 Pandemic. In 2021 7th International Conference on E-Business and Applications (pp. 70-76).

[8] Zhang, K., Zhong, G., Dong, J., Wang, S. and Wang, Y., 2019. Stock market prediction based on generative adversarial network. Procedia computer science, 147, pp.400-406.

[9] Song, Y.G., Cao, Q.L. and Zhang, C., 2018. Towards a new approach to predict business performance using machine learning. Cognitive Systems Research, 52, pp.1004-1012.

[10] GOH, T.S., HENRY, H. and ALBERT, A., 2021. Determinants and prediction of the stock market during COVID-19: Evidence from Indonesia. The Journal of Asian Finance, Economics and Business, 8(1), pp.1-6.

[11] Sharma, G.D., Erkut, B., Jain, M., Kaya, T., Mahendru, M., Srivastava, M., Uppal, R.S. and Singh, S., 2020. Sailing through the COVID-19 crisis by using AI for financial market predictions. Mathematical Problems in Engineering, 2020.

[12] Sharma, G.D., Erkut, B., Jain, M., Kaya, T., Mahendru, M., Srivastava, M., Uppal, R.S. and Singh, S., 2020. Sailing through the COVID-19 crisis by using AI for financial market predictions. Mathematical Problems in Engineering, 2020.

[13] Yoo, P.D., Kim, M.H. and Jan, T., 2005, November. Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06) (Vol. 2, pp. 835-841). IEEE.

[14] Majumder, M. and Hussian, M.A., 2007. Forecasting of Indian stock market index using artificial neural network. Information Science, pp.98-105.

[15] Thakkar, A. and Chaudhari, K., 2021. Fusion in stock market prediction: a decade survey on the necessity, recent developments, and potential future directions. Information Fusion, 65, pp.95-107.

[16] Van Rooij, M., Lusardi, A. and Alessie, R., 2011. Financial literacy and stock market participation. Journal of Financial economics, 101(2), pp.449-472.

[17] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P., 1996. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, 39(11), pp.27-34.

[18] Sable, R., Goel, S. and Chatterjee, P., 2019, December. Empirical study on stock market prediction using machine learning. In 2019 International conference on advances in computing, communication and control (ICAC3) (pp. 1-5). IEEE.

[19] Bora, D. and Basistha, D., 2021. The outbreak of COVID-19 pandemic and its impact on stock market volatility: Evidence from a worst-affected economy. Journal of Public Affairs, 21(4), p.e2623.

[20] He, Q., Liu, J., Wang, S. and Yu, J., 2020. The impact of COVID-19 on stock markets. Economic and Political Studies, 8(3), pp.275-288.

[21] Musdalifah Azis, B. and Rahayu, H., 2021. Stock Price of Pandemic Covid-19 in Stock Market Performance. Universal Journal of Accounting and Finance, 9(2), pp.184-190.

[22] Vadlamudi, S., 2017. Stock Market Prediction using Machine Learning: A Systematic Literature Review. American Journal of Trade and Policy, 4(3), pp.123-128.

[23] Pathak, A. and Pathak, S., 2020. Study of machine learning algorithms for stock market prediction. International Journal of Engineering Research Technology, ISSN, pp.2278-0181.

[24] Maldonado, S., Domínguez, G., Olaya, D. and Verbeke, W., 2021. Profit-driven churn prediction for the mutual fund industry: A multisegment approach. Omega, 100, p.102380.

[25] Lazarov, V. and Capota, M., 2007. Churn prediction. Bus. Anal. Course. TUM Comput. Sci, 33, p.34.

[26] Waqar, M., Dawood, H., Guo, P., Shahnawaz, M.B. and Ghazanfar, M.A., 2017, December. Prediction of stock market by principal component analysis. In 2017 13th International Conference on Computational Intelligence and Security (CIS) (pp. 599-602). IEEE.

[27] Siami-Namini, S., Tavakoli, N. and Namin, A.S., 2018, December. A comparison of ARIMA and LSTM in forecasting time series. In 2018 17th IEEE international conference on machine learning and applications (ICMLA) (pp. 1394-1401). IEEE.

[28] Hyndman, R.J. and Athanasopoulos, G., 2014. Forecasting: Principles and practice. otexts; 2014. Online at http://otexts. org/fpp.

[29] Lam, M., 2004. Neural network techniques for financial performance prediction: integrating fundamental and technical analysis. Decision support systems, 37(4), pp.567-581.

[30] Yadav, A., Jha, C.K. and Sharan, A., 2020. Optimizing LSTM for time series prediction in Indian stock market. Procedia Computer Science, 167, pp.2091-2100.

[31] Islam, A. and Khaled, M., 2005. Tests of weak-form efficiency of the Dhaka stock exchange. Journal of Business Finance Accounting, 32(7-8), pp.1613-1624.

[32] Uddin, M.G.S. and Alam, M.M., 2007. The impacts of interest rate on stock market: empirical evidence from Dhaka Stock Exchange. South Asian Journal of Management and Sciences, 1(2), pp.123-132.

[33] Majumder, A., Rahman, M.M., Biswas, A.A., Zulfiker, M.S. and Basak, S., 2022. Stock Market Prediction: A Time Series Analysis. In Smart Systems: Innovations in Computing: Proceedings of SSIC 2021 (pp. 389-401). Springer Singapore.