Currency Exchange Rate prediction using LSTM and Bi-LSTM Model

AUTHOR

Syed Shah Mir Ul Hassan

MS-18-CE-00000274839

SUPERVISOR

DR ALI HASSAN

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

Feb, 2022

Currency Exchange Rate prediction using LSTM and Bi-LSTM

Author

Syed Shah Mir Ul Hassan

A thesis submitted in partial fulfillment of the requirements for the

Masters of Science Degree in Computer Engineering.

Thesis Supervisor: Dr. Ali Hassan

Thesis Supervisor's Signature: ______________________________

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING
COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING
NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,
ISLAMABAD

Feb, 2022

Declaration

I confirm that "*Currency Exchange Rate prediction using LSTM and Bi-LSTM* " is my original work. This work has not been submitted to anywhere else for review. The usage of content from other sources has been appropriately recognized and referred to.

Student's Signature

SYED SHAH MIR UL HASSAN

MS-18-CE-00000274839

Language Correctness Certificate

This thesis has been proofread and is free of typographical, syntactic, semantic, grammatical, and spelling errors. The format of the thesis is also dictated by the university.

Student's Signature

SYED SHAH MIR UL HASSAN

00000274839

Supervisor's Signature

DR. ALI HASSAN

Plagiarism Certificate (Turnitin Report)

Plagiarism has been verified on this thesis. Attached is a Turnitin report that has been approved by the supervisor.

Student's Signature

SYED SHAH MIR UL HASSAN

00000274839

Supervisor's Signature

DR. ALI HASSAN

Copyright Statement

Acknowledgements

I am grateful to Allah, my Creator, for guiding me through every stage of this project and for every new notion You implanted in my head to help me enhance it. Without Your invaluable assistance and instruction, I would have been unable to do anything. It was Your will that whomever assisted me over the duration of my thesis, whether it was my parents or anybody else, thus no one else is deserving of appreciation but You.

I am eternally grateful to my loving parents, who reared me from the time I was unable to walk and have continued to support me in every aspect of my life.

I'd also want to thank my supervisor, DR. ALI HASSAN, for his guidance during my thesis and for the Machine Learning course that he taught me. I can confidently state that I have never mastered an engineering topic as thoroughly as the one he has taught.

I'd like to express my gratitude to my co-supervisor, DR. WASI HAIDER BUTT, for his unwavering support and collaboration. He always had an answer for me when I was stuck on something. I would not have been able to finish my thesis without his help. Throughout the thesis, I appreciate his patience and help.

I'd also want to express my gratitude to DR. FARHAN RIAZ, DR. FARHAN HUSSAIN and A.P. JAHAN ZEB for serving on my thesis guidance and assessment committee.

Finally, I'd want to extend my thanks to all of the folks who have helped me with my research.

Abstract

Predicting financial data, whether stock market rates or foreign exchange rate prices, has always been tricky. Financiers all throughout the world are interested in accurate data prediction because of the possible financial rewards. According to Efficient Market theory financial market is random process and impossible to predict accurately. However, researchers are trying different methods to predict market values. Different statistical approaches like Moving Average, Smoothing Average, ARIMA and other models have shown quite better results for prediction. Now with advent of processing power machine learning algorithms have also shown great efficiency. Neural networks are quite popular due to its highest accuracies. In this work we have worked on prediction of currency exchange rates through 2 different by using a modified version of Convolution Neural Networks called Long Short Term memory (LSTM). We generated two models in this experiment, one using LSTM and the other using Bi LSTM. NASQAD Composite (IXIC) is an American Stock index, that we have tried to model and predict using both the models. For modelling we have used historical price closing data. The stock's information was acquired from Yahoo Finance. As an accuracy statistic, we applied MSE. We compared its accuracy to previously published results for the same dataset. We have noticed that models have significantly improved the accuracy. We have also applied these models to Forex data for seven different currency pairs. Daily closing exchange rates were available in the data for AUDUSD, EURUSD, GBPUSD, NZDUSD, CADUSD CHFUSD and JPYUSD. The dataset recorded values of these currencies for each 5 minutes from Oct 2020 to March 2021. We selected one instance of values per day instead of each five-minute value. The dataset for these values were obtained from State Bank of Pakistan for research. Results have shown that our Bi-LSTM Model has far better performance than LSTM model. NASQAD IXIC stock market MSE reported by researches using LSTM is 0.0022 while as proposed LSTM has achieved MSE of 0.0001 which is quite less. Forex Data set has an MSE of 0.00014 for LSTM and 3.25 x 10 ^ -5 for Bi-LSTM model. From our research, we can state that these models are crucial for correct forecasts of time series data.

Keywords— LSTM, Bi LSTM, Forex Data, financial data**.**

Table of Contents

List of Figures

Tables

CHAPTER NO. 1

INTRODUCTION

CHAPTER No. 1: INTRODUCTION

Nowadays, investor as well as other capitalists are focusing on finding out new strategies to estimate the market prices. Financial forecasting is the process of calculating, estimating, or projecting a company's future performance. A financial prognosis is a prediction of how the firm's finances will look like in the future. A typical example of economic prediction is forecasting a sales revenue. Although sales are connected to or associated to the majority of financial statement accounts, forecasting sales can help a business make other economic choices that support its goals. Nevertheless, as sales increase, so do the costs of producing those additional sales. Each estimate has an influence on the business's financial situation. In subsequent parts, we shall explore several forms of market that consumers are interested in. Apart from standard statistical models, now-a-days machine learning techniques are thought to be pretty beneficial for the task. Due to improvement in processing power, latest models of Machine Learning are being evaluated for greater accuracies.

A random walk down Wall street, revised book version published in 2017 [1] has enormous and useful information regarding dealing in stock business based on practical approach. The hypothesis states that stock prices reflect all the available information, and any market movement is a random event. Hence, it is impossible to predict future market movements. Physical vs. psychological causes, rational vs. illogical conduct, and other elements all play a role in the forecast. All of these factors combine to make stock values very volatile and difficult to anticipate accurately. That is the main reason why this task despite of its potential benefits is still a difficult task. Despite the difficulty, researchers are constantly trying to do the prediction. Author in [2] 1996, applied neural networks along with decision rule base model for prediction of S&P 500 Stock market values. Similar efforts has been made by [3], [4], [5] & [6]. Various hybrid intelligence systems have been developed in recent years to do complex automation tasks, modelling expertise, and decision guidance helped by common analytical approaches to execute stock market analysis Newly developed systems are quite help with accurate predictions. [7].

In this section, the survey of various machine learning techniques used in stock market prediction and the challenges of the existing methods are discussed. This research has resulted in the construction of two models, one based on LSTM and the other based on Bi LSTM. We have tested these models on three unique data sets and found them to be effective. The first is the NASDAQ Composite (IXIC), which is an American stock index that estimates future values based on historical data. The information regarding this stock was collected from Yahoo Finance. As an accuracy statistic, we have applied the mean square error (MSE). In this study, we looked at the validity of the data and compared it to previously published results. As a second stage, we ran the models against Google Stock Data. Yahoo Finance was also used to get the stock values for this investigation. Further for more study, both models are applied to projected values of seven other currencies rates to the United States Dollar currency exchange rate for the next day. For this reason, the State Bank of Pakistan contributed a data set that was used for study. In the results section we can see the improvements in the MSE for both the models.

1.1 Motivation

Main motivation for my research work is to find and develop models that can predict correct stock or currency exchange rates. Ultimate is goal is for financers and capitalists to make profits. As discussed the prediction is not an easy task. Market is unpredictable and a random process that can change any moment. Despite its randomness researches are constantly looking for accurate prediction because of its financial benefits. The method through which a corporation develops its financial representation is known as financial modelling. Business choices are made using the model that was developed. A company's financial models are computational models in which variables are connected together.

The modelling method is developing an Excel spreadsheet with a summary of a company's financial data. The model may be used to predict the consequences of a management choice or an upcoming event. The organization may also tweak the variables in the spreadsheet to see how the changes will influence the business.

1.2 Main Objectives

In this thesis, we have concentrated on projecting the currency exchange rates of a range of different currencies in relation to the US dollar. We have also worked on different stock market values as well. Predicting financial data has always been a challenging task, whether it be stock market rates or the worth of a currency exchange rate. The precise prediction of such data is of tremendous interest to financiers all around the world because of the probable financial rewards that may be obtained. Major objective that we have worked in this study can be highlighted as under:

- Study the already existing methods for prediction of Stocks or Forex Data
- Studying machine learning models that have already applied on time dependent data.
- Creating an LSTM model that shows better performance that existing models.
- Creating a Bi-LSTM model for same purpose.

1.3 Problem Statement

In data analysis, time series forecasting and modelling are crucial. Time series analysis is a subset of statistics that is widely employed in subjects like econometrics and operations research. Both stocks and forex markets accurate predictions is still a challenge for current day researchers. Machine and deep learning algorithms are expected to give better response.

In analytics and data science, time series is commonly employed. This is a custom-made time series issue, and our task is the accurate prediction of future values.

Rest of the report has been structured as under:

1. In Chapter No. 2 we will discussion briefly about the financial market and see how different statistical methods are being used for prediction purpose.
2. In Chapter No. 3 we will look at the Machine Learning Algorithms for same purpose.
3. In Chapter No. 4 I will discuss the different dataset that we have worked on in our work.
4. Chapter No. 5 comprises of details structure of 2 models that have been used.
5. Chapter No. 6 shows results section along with comparison with other researchers' work.
6. In this last Chapter we will see at possibilities of future work based on this research.

1.4 Chapter Summary

In this Chapter I have discussed how I have done my research. Starting from the machine learning techniques and its use in financial data prediction we have seen different authors work. Next we discussed that financial gains are main motivation behind such research. We have seen the objectives and learning outcomes of the research starting from financial market till development of the model are included in the objective. At the end we have seen the overall structure of remaining chapter in this research.

CHAPTER NO. 2

FINANCIAL MARKET

CHAPTER No. 2: FINANCIAL MARKET

2.1 FINANCIAL MARKET

Capital markets play a critical role in ensuring the smooth running of economic systems by distributing capital and producing flexibility for enterprises and entrepreneurship. The capital markets end up making asset swapping easier for sellers and buyers. Capital systems offer security products and provide a reward for those of us with additional cash (shareholders) while apparently making these money available to someone in need (borrowers).[8]

Some financial institutions are small and have limited trading activity, while others, such as the New York Stock Exchange (NYSE), handle trillions of dollars in transaction records [9] The share market (equity funds) is an investment market that allows investors to buy and sell interests in publicly traded corporations. The major stock market would be where new stock offerings, often known as initial public offerings (IPOs), are sold. Any extra securities trading takes place in the secondhand market, wherein investors buy and sell assets they already have.

Financial markets can be viewed of as conduits for available for lending money to flow from a suppliers with excess commodities to a demander in need of cash or capital.

2.2 Types of Markets:

There are numerous sorts of financial markets and their definition depends on the qualities of the financial claims being exchanged and the demands of the different market players. We distinguish different sorts of markets, which vary dependent on the type of the instruments exchanged and their age.

Some of the most popular varieties are mentioned as under: [10]

2.2.1 Stock Markets:

The stock markets have become the most widely used of all financial markets. These are venues where businesses list their stock values and also where businesspeople and customers can buy and sell them. Organizations utilize stock markets / stocks marketplaces to raise money through an initial public offering (IPO) (IPO). These securities are then sold and bought by a large number of people. The buying and selling refers to these marketplaces.

Stocks can be exchanged accented with oil - rubbed or on public exchanges like the New York Stock Exchange (NYSE) or Nasdaq (OTC). The majority of share market takes place on organized exchanges, which play an important role in economic growth as either a barometer of the economic growth overall health and a source of capital appreciation and stock dividends for shareholders, particularly those with retirement savings like IRAs and 401(k) plans.

Market makers (MMs) and experts who manage stability and create two-sided marketplace are prominent actors in the stock market, as are corporate and business professional investors. Brokers are unaffiliated third parties whose facilitate exchanges between customers and manufacturers but do not have a stock position.

2.2.2 Forex Market:

Foreign Exchange is abbreviated as Forex. The forex market (foreign exchange market) is a market where people can buy, purchase, hedging purposes, and make assumptions on the exchange rates of different currency pairs. Because cash has been the most liquid asset, the exchange rate is the most transparent market on the planet. The financial market transacts $5 trillion each day that is more than both commodities and financial markets altogether. The FX market, like OTC markets, is decentralized and comprised of a multitude of platforms and brokers. The simultaneous acquisition of one fiat exchange and marketing of another is known as currency. For illustration, foreign currencies are depicted as pairs, USD/JPY (US dollar against Japanese yen) or AUD/USD (Australian dollar versus US dollar) [11]. We will see in the next chapter where we have engaged on Forex market prices. The forex market is made up of banks, corporate houses, central banks, financial advisory organizations, financiers, and foreign exchange trading investment companies and investors.

2.2.3 Commodities Market:

Companies and consumers trade physical commodity such as grain, cattle, and soybeans, along with energy items (crude oil, gasoline, charcoal), valuable materials (gold, silver, and platinum), and "soft" products at commodity prices (such as silk, tea, and sugar). Physical commodities are swapped for cash on spot commodities marketplaces.

But at the other hand, the majority of all these commodities being transacted in derivative contracts that use spot commodities as underlying assets. Commodity contracts,

contracts, and options typically traded on organized exchanges such as the Chicago Mercantile Exchange (CME) and the Intercontinental Exchange (ICE) (ICE).

2.2.4 Cryptocurrency Market:

Bitcoin and Ethereum, for example, are decentralised digital products based on block chain technology that have grown in popularity in recent years. Multiple cryptocurrency currencies are already available and transacted on a patchwork of browser crypto exchanges all around the world. These marketplaces provide dealers with digital wallets that allow them to change cryptocurrency for fiat currency such as dollars and euros.

Users are vulnerable to hackers and fraud because of bulk of crypto exchanges are centralised. There is little need for a centralized power in a decentralised exchange. These marketplaces enable direct peer-to-peer digital currency P-P exchange without the use of a regulated market to handle the transactions. On cryptocurrencies, futures contracts trading are also accessible.

2.3 Existing Prediction Methods:

Market forecasting is a challenging yet lucrative task. Any investor should be focused on two prices: the current price of the commodity they own or want to buy, and the long term price where which it should be sold. Shareholders, on the other hand, are constantly reviewing previous price movements in attempting to manipulate their upcoming investment choices [12]. Some investors may avoid buying a company or index that has climbed too quickly because they believe it is due for a correction, while others will resist selling a stock that is sinking because they believe it will continue to decline.

Forecasting is a method of making well-informed forecasts about the direction of sustainable trends by using existing information as input. Organizations use forecasting to figure out how to distribute their budgets and budget for projected costs and higher. The projected value of a product and services performed is frequently the determining factor.

A variety of approaches for predicting financial market movements have been developed throughout the years. These patterns aid bankers and investors in choosing the correct commodities and investing in the proper direction [13]. A few of classical methods that we and still are used will be discussed below.

2.3.1 Auto Regression (AR)

The auto regression (AR) approach models the next step in a series as a linear function of previous time steps' data. This approach was used to predict stock prices by Sahoo, Pk, and their co-authors [14]. They came to the conclusion that because the model is simple, it works better in the majority of circumstances.

The AR(p) model is defined as

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t \tag{1}$$

where φ are the parameters of the model, c is a constant, and έ is white noise.

The approach works well with univariate time series that don't have any trend or seasonal components. The AR Model is depicted graphically in Figure. 1.

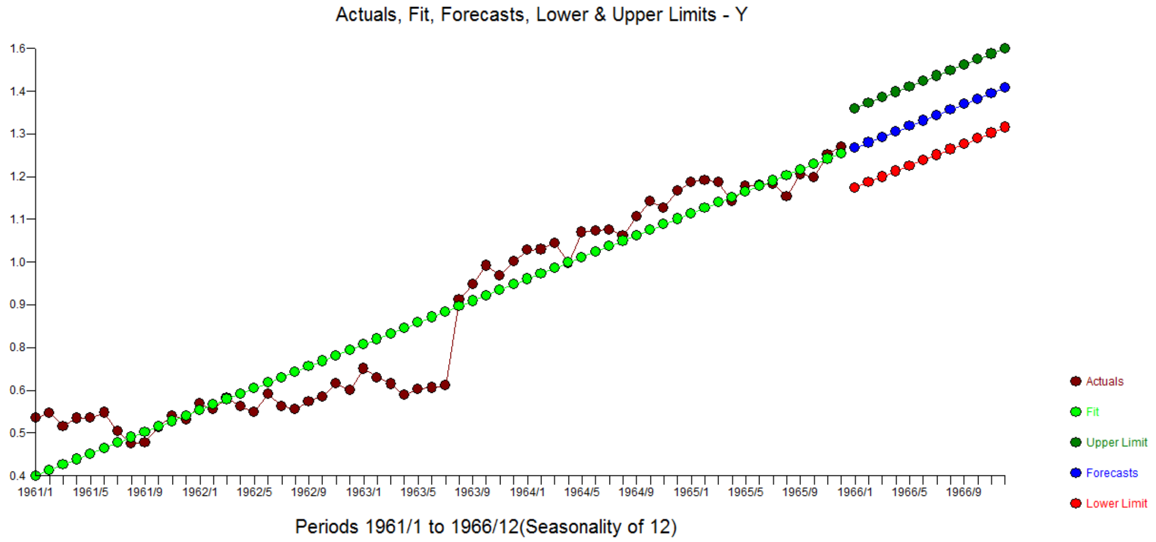


Figure. 1 Representation of AR Model

2.3.2 Moving Average (MA)

The very next step in the series is represented by the moving average (MA) technique as a linear transformation of regression from prior stages from an average mechanism. In [15] authors applied moving average approaches since it reduces noise from the data which further helps in prediction of values. In their study they have compared outcomes with different types of moving average models.

There are number of types of moving averages. Some of these are discussed below.

2.3.3 Simple Moving Average: SMA

By accumulating the prices / values from a collection of single period, this same simple OR arithmetical marching score was calculated (for instance, 12 hours). The number is then defined as the fraction of times this has happened. Figure. 2 represents graphical representation of Simple Moving Average.

Figure. 2 Simple Moving Average [16]

Equation 2 represents mathematical representation of the model.

$$SMA_i = \frac{\sum_{k=i-n}^{i} x_k}{n} \qquad (2)$$

Where:

n — number of periods in the moving average.

Lauren S. in 2015 [15] applied simple moving average along with other machine learning algorithms for prediction of stock prices by combining stock price values with news from different resources. By combining simple moving average technique and news classification, and the result indicated improved prediction responsiveness.

2.3.4 Exponential Moving Average (EMA)

By multiplying the original function of the rolling average by a proportion of the present current share price, the increasingly smoothing moving average is generated. When utilizing enormously adjusted moving averages, the most recent close pricing have much more relevance.

Figure. 3 Exponential Moving Average

$$EMA_{\text{Today}} = \left(\text{Valuetoday} * \left(\frac{\text{Smoothing}}{1+ \text{Days}} \right) \right)$$
$$+ EMA_{\text{Yesterday}} * \left(1 - \left(\frac{\text{Smoothing}}{1+ \text{Days}} \right) \right)$$

(3)

Fikri M and co-authors have used Moving average filter on data from different sensors for removing noise and discussed how Alpha parameter can adjusted for better results.[17]

2.3.5 Autoregressive Moving Average (ARMA)

The Autoregressive Moving Average (ARMA) method predicts the very next time step in a succession as a linear model of something like the observations and remaining errors from the preceding time steps.

It combines both Auto regression (AR) and Moving Average (MA) models.

$$X_t = w_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j w_{t-j} \tag{4}$$

In a study in 2020, [18] authors have applied ARMA model for forecasting US dollar and British Pound currency exchange rate prediction and shown their results using MSE, MAPE and MAE . They have also concluded that for prediction of financial values, classical approaches are better.

2.3.6 Autoregressive Integrated Moving Average (ARIMA)

The Autoregressive Integrated Moving Average (ARIMA) method models the next simulation time as a linear function of the conditional variance observed and previous time step residuals. It's a type of approach that can capture a variety of common temporal features in time series data. When you use an ARIMA model to analyse a time series, you're assuming that the theoretical underpinnings that created the data is an ARIMA process. That might seem self-evident, but it aids in motivating the need to test the model's hypotheses in raw observations and unexplained errors of process as compared.

To keep the sequencing stationary, it uses both Autoregression (AR) and Moving Average (MA) models, as well as a sequences public spending and economic pre-processing phase called integration.

ARIMA model despite a classical approach for many applications is still used. In 2020, Benvenuto applied ARIMA model for prediction of spread of Covid-19 virus [19].

2.4 Use of Apps in Trading:

In this section we will discuss that how a mobile trading app uses different indicators to help its users. Not only mobile but softwares on desktops and laptops are also very common. We are not going to recommend or discuss any specific application. Here we are only interested in the facilities provided by these apps in general. For trading purpose, a number of mobile and desktop applications are available. Some a free other are paid.

Blain Reinkensmeyer in his article published on Stockbrokers.com have recommended few mobile apps for trading based on different features offered. Almost all the application uses market movers / indicators to their users for facilitation [20].

Figure. 4 Graph from a randomly selected app showing MA

Figure. 4 is a randomly selected picture of a mobile trading app that shows values of stocks in candlestick pattern. There are a number of different indicators shown in Blue, Black lines. Each indicator provides some information that can help the user in making purchase or sell decision.

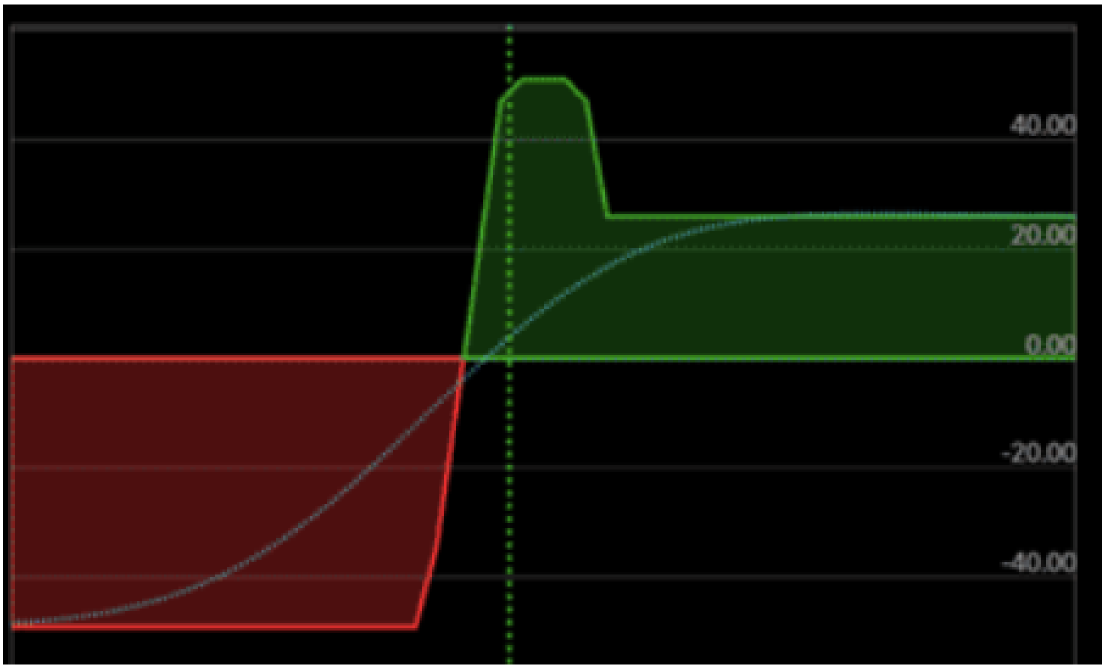


Figure. 5 Buy and Sell Recommendation

In Figure. 5 a mobile app recommends the user about buy and sell decisions depending upon the curve obtained from values of Stock.

2.5 Chapter Summary:

The discussions in this chapter demonstrate that financial markets play a crucial role in our society. They are of interest to a broader audience than the capitalists and investors who benefit from them. However, these marketplaces are critical to a country's progress. Not only does investing in such a market require capital, but it also requires expertise in order to make educated judgements. Throughout history, people have tried with a variety of tactics and devices to forecast future market positions. Due to the fact that these markets do not follow a clear trend or pattern, it has always been a difficult assignment. Numerous statistical approaches, as discussed before in this chapter, are used to help in decision-making.

In this chapter, we have discussed in detail different types of markets. Depending on the business and the firms within a certain industry, there are a range of distinct market systems. Whenever making production and investment decisions, or deciding whether one should enter or quit a certain industry, it is critical for businesses to comprehend whatever different types of market system they are working in. A number of investors and other people are associated with these markets. Further we have also discussed a number of Statistical methods that have been in use of prediction purpose. One strategy to anticipate stock prices using statistical numerical technique is to use ARIMA that is a type of linear regression that assesses the amplitude of one dependent variable compared to those other variables associated. The purpose of the model is to anticipate future stock prices by looking at the disparities in the variables' values. For moment research methodology, exponential smoothing average and basic smoothed average are often utilized. These approaches are used to make a judgement call on the user's past expectations, such as periodicity.

In the last section of the chapter we seen how these methods are practically applied and available to user that are involved in different types of market. Using these methods and many other people have developed indicator that are helpful.

CHAPTER NO. 3

MACHINE LEARNING ALGORITHMS

CHAPTER No. 3: MACHINE LEARNING METHODS

In the last chapter, we looked at how traditional/statistical methodologies are employed as indicators. With the advancement in processing power and the usage of Artificial Intelligence, we can now employ machine learning algorithms to do this work.

Machine learning algorithms are widely used for a number of tasks like Recognition, Classification, Pattern Recognition etc. [21]. The research of computer algorithms that can improve itself independently depending on knowledge & data is known as machine learning (ML). It's thought to be a part of artificial intelligence. In terms of generating forecasts or assessments without ever being explained in the previous, machine learning techniques build a methodology based on available data, often to as "dataset." Machine learning algorithms are used in a wide variety of uses, such as healthcare, spam filtering, language understanding, and machine learning, when current methods seem virtually impossible to build.

In other words, machine learning can be explained by considering a scenario. Normally we have two number and an operator between it. Let us we have two numbers. 7 and 13. Depending upon the operator we calculate the answer. Addition will result in 20 and multiplication 91.

Now consider the following data.

x	y	z
2	4	6
5	7	12
8	10	18
12	12	24

Table No. 1 Data showing 2 inputs and 1 output

Table No. 1 shows the inputs x and y, as well as the output z. The terms x, y are called attributes or features. While as z is called Labels. As the data is very simple, we can see that the result (z) is calculated simply adding the two values (x and y) together. For some complex data where there are large number of attributes it might not be possible to derive an equation which correctly gives us Label or output value. Machine learning algorithms are effective at detecting patterns when given a large amount of tagged data to work with. The pattern in the example above was addition. The mapping function is another name for this pattern.

In other cases, there may be a pattern that a person cannot comprehend simply by looking at it. In real-world challenges, we typically have tens or even hundreds of characteristics. Some classifications, such as Fish Types, may also be left out. Consider the following illustration.

Index	**lightness**	**width**	**species**
0	2.834754098360656	21.087142857142855	0
1	3.329180327868852	18.877142857142857	0
2	3.6904918032786886	19.824285714285715	0
3	4.812459016393442	17.759999999999998	0
4	8.078709677419354	19.09893238434164	1
5	8.078709677419354	18.46049822064057	1
6	7.798064516129033	17.920284697508897	1
7	8.284516129032259	19.516370106761567	1

Table No. 2 Salmon and Seabass fish dataset

Table No.2 represents a subset of Salmon and Seabass fish dataset. In this Dataset 0 represents Salmon and 1 Represents Seabass. Depending upon the lightness and length these two types of fish can be classified using machine learning[22]. Each point in the dataset represents a fish. On X – axis we have Lightness of the fish while as on y axis we have width of fish. The data is known to us. We know the species or label of the fish. We know from background information that red dots represent sea bass and black dots represents salmon. Each machine learning algorithm may classify this data in different sections. But the simplest method is to put a simple line that seperates the data. In Figure. 6 we can a linear line separating both data. Also note that some of the black dots are in red and a few red dots are also shown with black dots. These items are called miss classification or error. May be a line repressing a higher order equation correctly classifies this data.

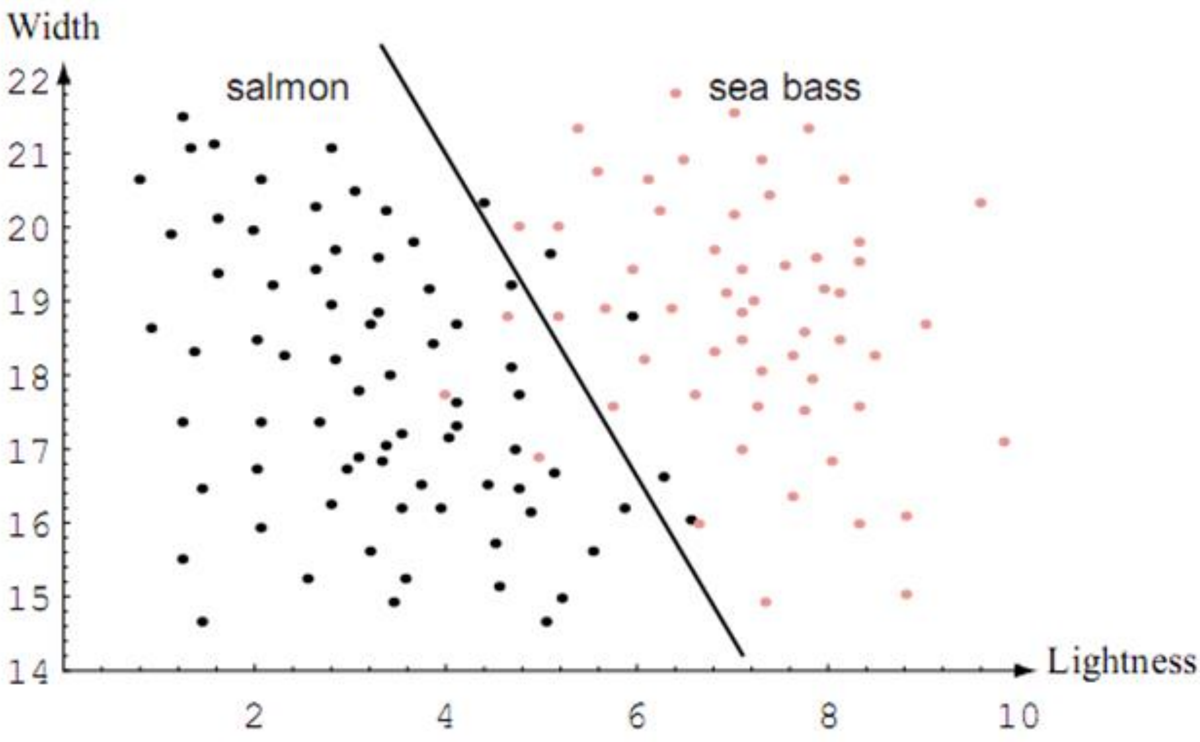


Figure. 6 Feature Classification of salmon and sea bass fish [22]

Each point represents a fish with length x1 and width x2. We construct a line dividing space into two sections, with any fish falling below it identified as salmon and any fish falling over it labelled as sea bass.

3.1 Unsupervised and supervised Learning:

These are separated into two groups based on the distinct roles of machine learning:

1. Supervised Machine Learning
2. Unsupervised Machine Learning

Machine learning approaches include administered and unsupervised knowledge. However, both strategies are employed in various contexts and with various datasets. Supervised learning is a type of machine learning that comprises using labelled data to train classifiers. In supervised learning, algorithms must find a nonlinear mapping to relate the input parameter to the output variable *(x)* to the output variable *(y).*

$$y = f(x) \tag{5}$$

Supervised learning, comparable to how a student has learned in the supervision of an instructor, is required to train your models. Logistic regression are two types of problems that can be tackled utilizing supervised methods.

Learning is a method of machine learning that discovers patterns from unsupervised learning. The goal of supervised methods is to identify structure and relationships from unstructured text. While learning individually, there is little need for supervision. Alternatively, it looks for patterns or relationships on it's own and.

Unsupervised learning model identifies hidden patterns in data whereas supervised learning model predicts output. When compared to supervised learning, the supervised learning model delivers accurate results, while the unsupervised learning approach may generate less accurate results.

Sapkal [23] in his research applied both supervised and unsupervised learning methods to classify 250 object images. The techniques used for classification are Competitive learning, Self-organizing map (SOM) and Learning Vector Quantization (LVQ) networks.

3.2 Supervised learning algorithms

There are a number of supervised learning algorithms that are widely used in different applications. Some of these are discussed below.

3.2.1 k Nearest-neighbor

K-nearest neighbours (KNN) is a simple, incredibly simple supervised machine learning algorithm that could be used to contribute to solving problems. The most used classification method is K-Nearest Neighbor (KNN), which is among the top ten highly powerful and successful algorithms. KNN is regarded for being simple and straightforward. This algorithm's goal is to categorize new objects based on their properties and training data. The Euclidean equation formula is used to categorize things based on the training data that has the closest distance to a new item. Because it is good at controlling noise, simple, uncomplicated, easy, and employs a huge network, the KNN algorithm becomes a possibility. [24]

```
Algorithm   kNN kernel Algorithm
Input  :  D, a chunk of the original distance matrix;
          n_chunksize, dimension of the chunk;
          split, index of split;
          chunk, index of chunk;
          Maxk, an array to hold the farthest neighbors for each row index in the chunks;
Output:   none, an (intermediate) kNN graph stored in Gk′;

row′ ← blockIdx.x × blockDim.x + threadIdx.x;
if row′ < n_chunksize then
    row ← split × n_chunkSize + row′ // absolute row in the distance matrix;
    for column′ ← 1 to n_chunksize do
        column ← chunk × n_chunkSize + column′ //absolute column in the distance matrix;
        if row = column or row > n_row or column > n_col then
            continue /* exclude diagonal and pad regions */;
        if D[row′ × n_chunksize + column′] < Gk′[Maxk[row′]].weight then
            Gk′[Maxk[row′]].source ← row;
            Gk′[Maxk[row′]].target ← column;
            Gk′[Maxk[row′]].weight ← D[row′ × n_chunksize + column′];
            Search the new maximum element in row′(D) and store the index in Maxk[row′];
```

Algorithm: kNN kernel Algorithm

The KNN Algorithm follows following steps:

First we put the data on the system. K value is selected for number of classes that are required. For data point in the data, distance is calculated from each K point and data point. Distance matrix is created and distance for each data point with query example is recorded. These collection of distances are ordered from smallest to largest. From sorted distance first K elements are selected. Labels for K entries is given. If there is regression. Means of the labels is returned else K labels are classified.

In KNN method, selecting value of K matters the most. The following example will clarify this.

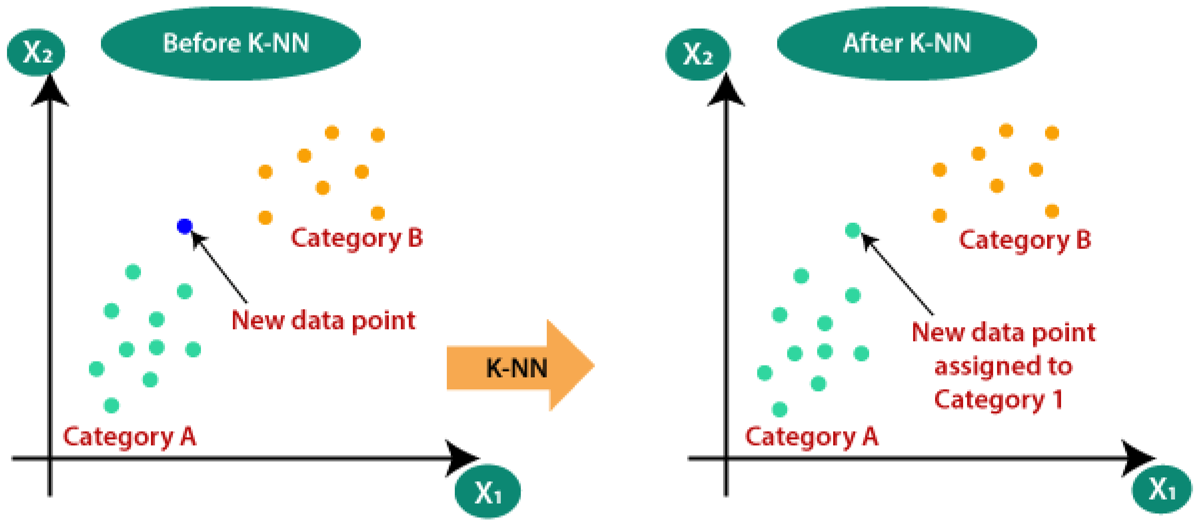


Figure. 7 KNN Classification[24]

In the Figure. 7, two classes green triangles and red stars are plotted. A new sample is being plotted. If we select k = 3, 3 nearest samples will be selected and class with maximum votes will be assigned in this case Green Triangles. For k=7, Red star class will be assigned.

Ehsani, R, Drabløs, F in 2020 [25] KNN method was used to test the algorithm's performance on four distinct types of cancer data sets. They discover that the performance of the novel distance measures is equivalent to that of more well-established distance measures, particularly for the Sobolev distance.

3.2.2 Naive Bayes

Naive Bayes is a probabilistic ML algorithm that has been used in a wide variety of classification tasks [26]. Filtering mail-spam, classifying documents, and predicting sentiment are examples of common uses. It is based on Rev.'s writings. The naive hypothesis is that the characteristics of the models are irrespective from each other. That is, adjust the position of one property has no effect on the algorithm's other properties.

The much more likely class for a particular case characterized by its feature space is assigned by Bayesian classifiers. The assumption that characteristics are agnostic of class can substantially simplify the acquisition of such classifications,

$$P(\mathbf{X} \mid C) = \prod_{i=1}^{n} P(X_i \mid C) \qquad (6)$$

where $\mathbf{X} = (X_1, \cdots, X_n)$ is a feature vector and C is a class.

The likelihood of each class is computed through this approach, and the candidate is assigned to the class with the highest probability. The eager learning classifier Naive Bayes has a fast execution time. As a result, it's possible to produce real-time forecasts with it. The Naive Bayes approach is also well-known for multi-class prediction, which involves sorting situations into many groups. When used to categorize text, a Naive Bayes classifier beats other algorithms due to its ability to perform well on multi-class issues while assuming independence. As a result, spam filtering and sentiment analysis are prominent applications.

In a study,[27] authors applied Naïve Bays Algorithm and KNN Classifier for approval of credit card application.

3.2.3 Decision Trees

By applying machine learning techniques from historical data, a Clustering Algorithm can then be used to create a training model that could be used to perform classification or quantity of target variables (training data). A logistic regression usually assesses all of the material as a base before starting. Then, situationally, it starts splitting and take judgments through branched or processing elements until it produces a leaf. Radar signal categorization, character recognition, remote sensing, medical diagnosis, expert systems, and a variety of other applications have all seen success using decision trees. [28]. Following Figure 8 clearly explains the process followed by Decision Trees.

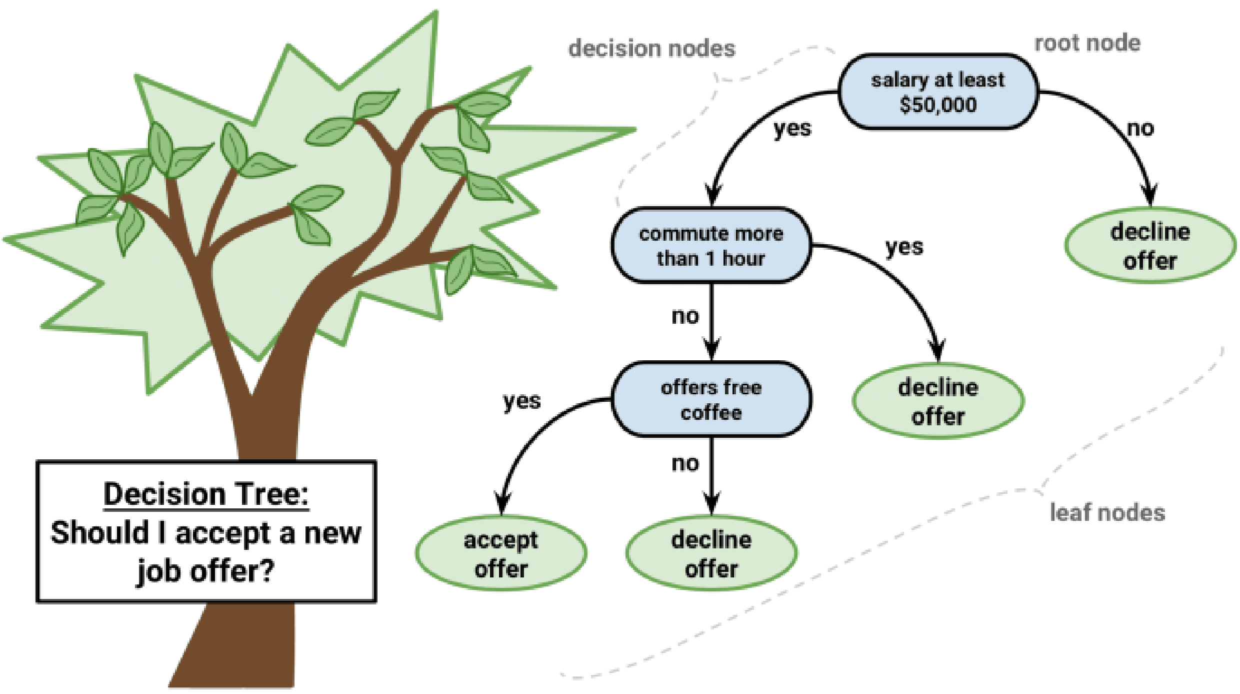


Figure. 8 Decision Tress explaining decision nodes, root node and leaf nodes

The most difficult aspect of creating a Decision tree is deciding on the best characteristic for the root node and sub-nodes. To address such circumstances, a method known as Optimization Criterion, or ASM, might be applied. And used this measurements, we can easily find the best characteristic for the tree's nodes. There seem to be two commonly used ASM techniques. These are given below:

Information Gain

Gini Index

$$Gain(S, A) \equiv Entrpoy(S) - \sum_{v \in Values(A)} rac|S_v||S| \cdot Entropy(S_v) \qquad (7)$$

Values(A) is the set of all possible values for attribute.

And Entropy (S) is

$$Entropy\ (S) = -p_+ \, log_2 \, p_+ - p_- \, log_2 \, p_- \qquad (8\ )$$

Conversely, we may calculate the classification performance for each attribute (from the collection of attributes) and choose the optimal characteristic to split on based on the maximum classification algorithm.

3.2.4 Support Vector Machines

The Support Vector Machine (SVM) is a supervised machine learning algorithm that is used to classify and/or predict data. It's a straightforward algorithm. It's more usually used for classifying, but in some cases, it can also be useful for regression. SVM is used to find a higher dimensional space, which acts as a boundary between distinct sorts of data. This high energy was nothing but a horizontal path while viewed in two dimensions.

We display each data item in the dataset in an N-dimensional space, where N is the number of features/attributes in the data. This is known as the SVM method. Following that, determine the optimum hyperplane for separating the data. This should have made it clear that SVM can only conduct binary classification by default, as previously stated (i.e., choose between two classes). There are a variety of strategies that may be used to solve multi-class issues, though.

When dealing with linearly separable data, SVM performs admirably without any adjustments. Linearly separable data is any data that can be shown on a graph and can be divided into classes by drawing a straight line between the data points. When dealing with non-linearly separable data, we employ Kernelized SVM.

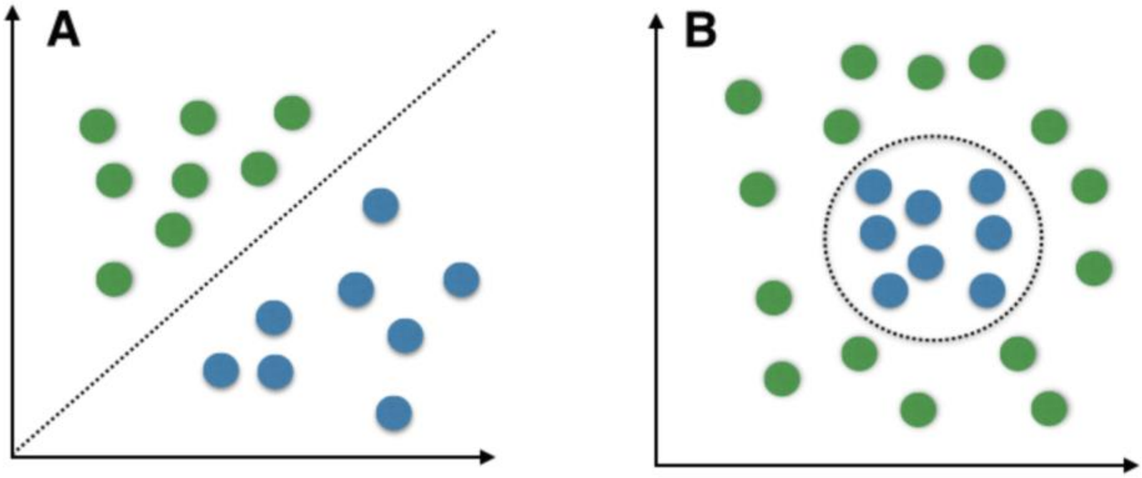


Figure. 9 Linearly Separable data v Non-linearly separable data

The support vector machine algorithm's goal is to find a hyperplane in an N-dimensional space (N — the number of features) that distinguishes between pieces of information.

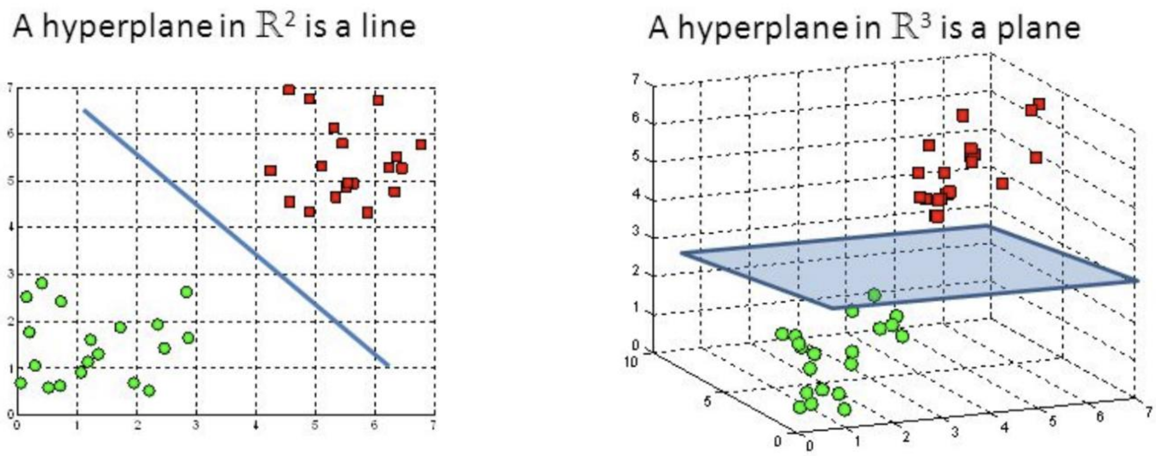


Figure. 10 Hyperplane in 2D and 3D dataset

Figure 9 we can see data can be linearly separable that can be easily separated by a linear line. However, non-linearly separable data can also be classified by calculating the non-linear line like a curve or circle that separates the data. In 2 dimensional data, this plane is a line or curve but in 3D dimensional data, it becomes a plane as shown in Figure 10. Main objective is to reduce the loss function.

Loss function of the SVM can be expressed as under:

$$\min_{w} \lambda \parallel w \parallel^2 + \sum_{i=1}^{n} (1 - y_i \langle x_i, w \rangle)_+ \qquad (Eq\ 9)$$

Speech signals are somewhat similar to financial data because of its random behavior. In 2021, authors proposed a model using SVM for classification of some neurologic diseases such as Alzheimer, depression, and Parkinson. They worked on parkison speech dataset that comprises of 34 voice notes, 14 of healthy speeches and 20 patients affected by PD. They achieved an accuracy of 91.1% [29].

3.2.5 Artificial Neural Networks

Neural Networks. Neural networks, also called Artificial Neural Networks (ANN) are inspired by the network of Neurons from Human Brains [30]. Neurons are the building blocks of ANNs.

3.2.5.1 Neurons

Neuron takes the input of data at the input layer, takes its weighted sum, passes it through an activation function, and generates an output unit. Figure.11 represents the structure of a simple neuron.

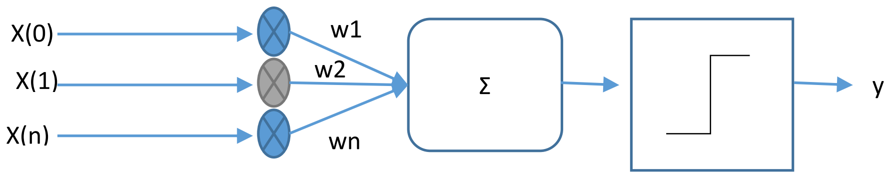


Figure. 11 Structure of a Neuron

Mathematically, neuron works as

$$y = f(\sum_{k=0}^{n} x_n w_n) \qquad (Eq\ 10)$$

3.2.5.2 Feed Forward Neural Networks:

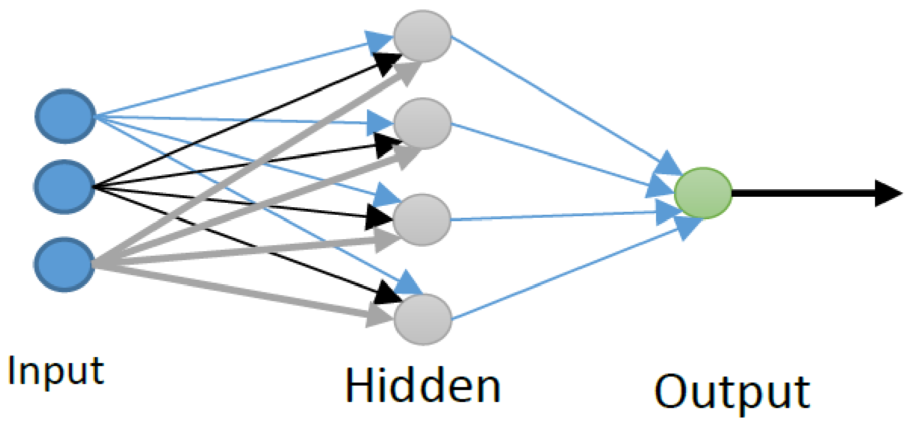
In neural networks, neurons are arranged and connected in layers to perform a difficult task, as shown in Figure.12. The input layer is the upper left level in this network, and the neurons within it are known as layers of neurons. The output neurons, or in this example, a single output neuron, are found in the constitutes an alternative or output layer. A concealed layer is the innermost part.

Figure. 12 Simple Neural Network

The above figure shows a feed-forward neural network as data flow from the input to the output layer. The architecture of a neural network can vary depending upon the number of hidden layers and input/output nodes. Depending upon the architecture, there are several feed-forward ANNs like Single-layer perceptron (SLP), Multilayer perceptron (MLP), Radial basis function network (RBFN), generative adversarial networks (GANs), sparse auto-encoder (SAE) [31].

3.2.5.3 Feed Back Neural Networks

Feedback neural networks or recurrent neural networks have feedback loops connections as shown in Figure.13. These feedback loops make the network complicated, which can perform difficult tasks. RNNs are powerful learners which can detect sequences and are recommended for speech recognition and stock market value prediction tasks [32].



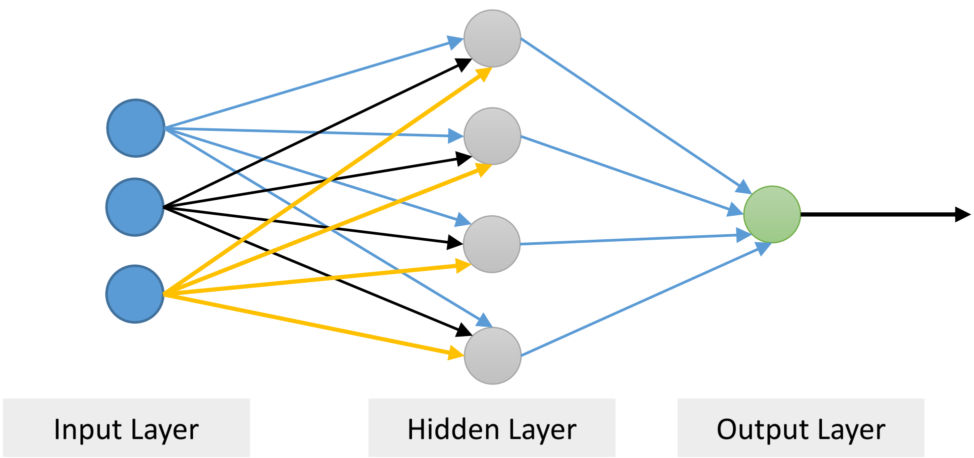


Figure. 13 Simple Architecture of recurrent neural network

Depending on the architecture of feedback RNN, there are different types, like LSTM, GRU, etc.

3.2.5.3.1 LSTM

Just like in a movie or video clip, previous patterns help us understand the current video screen. Similarly, in time-series data prediction, previous data is important to predict future values. LSTM was first introduced by Hochreiter and Schmidhuber in 1997[33]. Simple RNN has Long Term dependencies, while LSTM avoids such dependencies due to its structure and gates. Fig.13 shows a single LSTM cell.

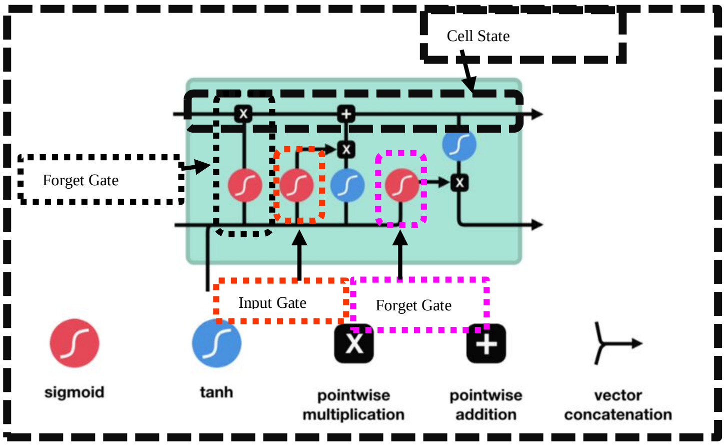


Figure. 14 A cell of LSTM along with its operations

Gates mentioned in Figure.14 are calculated by following equations

$$Input\ gate\ (I_t) = f_g(W_i X_t + R_i h_{t-1} + b_i)$$

$$\text{Forget gate } (f_t) = f_g\left(W_f X_t + R_{fht-1} + b_f\right)$$

$$\text{Cell State } (C_t) = f_c(W_c X_t + R_c h_{t-1} + b_c)$$

$$\text{Output gate } (O_t) = f_g(W_o X_t + R_o h_{t-1} + b_o)$$

3.2.5.3.2 Bi-LSTM

Bi-Directional LSTM is a modified version of LSTM. It was introduced by Schuster & Paliwal in 1997[34]. It incorporates two LSTM models that can access all the previous data and predicts future information simultaneously. The state cells of neural networks are duplicated. One acts as the forward state and the other acts as the backward state neuron as shown in Figure. 15.

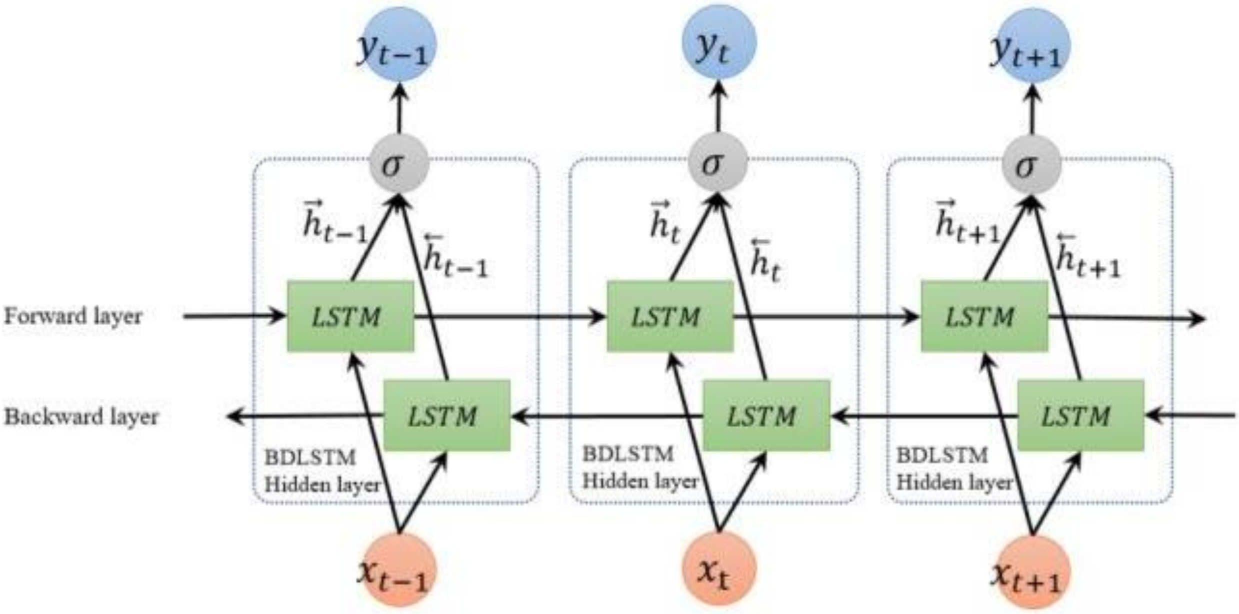


Figure. 15 Bi-LSTM Layer[35]

Zhiyong Cui and his co researchers in [35] used LSTM neural network and Bi Directional LSTM model for short term traffic forecasting. They claimed that their model is a scalable model can predict traffic speed for both freeway and complex urban traffic network. They used MAE and MAPE as accuracy metrics. For 5% of missing values MAE was 3.8 and MAPE was recorded as 9.053.

3.3 Chapter Summary

In this Chapter we have discussed a number of machine learning algorithms. For each algorithm we have seen how different researchers have used them for a variety of different task. Starting from very basic algorithms to complex Convolution Neural Networks we have seen their mathematical models. Their algorithms, their building blocks and their application by different researchers. We have used LSTM and Bi-LSTM models for our research that why these two models have been discussed in details. There also a number of other machine learning algorithms that are not mentioned here.

CHAPTER NO. 4

RELATED WORK AND DATA SETS

CHAPTER NO. 4 RELATED WORK AND DATA SETS

LSTM models have been used in a variety of applications, including the prediction of stock market values and other time series data. Models such as the LSTM and Bi LSTM are not just used for stock or financial data prediction, but also for other purposes. These models have also been used to mimic a variety of additional jobs in addition to those listed above. Chunting Zhou [36] in 2015, used Convolve Neural Network along with LSTM model for Text Classification purpose. The sentence representation is obtained by using a CNN to extract a succession of higher-level phrase representations, which are then fed into a long short-term memory recurrent neural network (LSTM) for training. They named their model as C-LSTM. In 2019, authors [37] have used LSTM model for prediction of vehicle trajectory predicting model. In order to deal with gradient vanishing, they also added shortcut connections between the inputs and outputs of two successive LSTM layers, which they called "shortcut connections." The suggested novel model was tested on the I-80 and US-101 datasets, and the findings revealed that it performed significantly better than existing models in terms of accuracy. In 2019, authors [38] have used long short term memory models for predicting air quality in the region of Madrid. The results, supported by statistical evidence, indicate that a single comprehensive model might be a better option than multiple individual models.

Researchers have applied multiple techniques for the prediction of time series and financial data. Farah Shahid and colleagues during a study on modelling of Covid-19 pandemic deaths and recoveries find out LSTM and Bi LSTM models to be the highest performance achieving models [39]. They applied ARIMA, SVR, GRU, Bi LSTM and LSTM models for forecasting of deaths and recoveries in 10 major countries across the globe. MAE and RMSE value of 0.007 and 0.0077 were recorded as the lowest for Bi-LSTM models for deaths in China amongst all other models. Zhigang Jin in 2020, used LSTM model along with sentiment index model and EMD model for prediction of Stock prices [40]. For their study they have used AAPL (stock of Apple) values.

Samuel Olusegun Ojo and his colleagues in presented a layered LSTM model for prediction of stock price of NASDAQ Composite ^IXIC Stock of New York Stock Exchange [41]. They have trained their data for 200 epochs using Spyder IDE for Python programming language. In the coming section we will discuss the data set in details. In our research we have also used same data set. They used 2-layered LSTM model with 100 and 50 neurons respectively and reported MSE of 0.0022.

Sunny [42] applied LSTM and Bi-LSTM models to predict google stock values. They collected data of Google stock from Yahoo finance from 19/08/2004 to 04/10/2019. Train test split was 88% and 12%. They reported their accuracy in RMSE. After 50 epochs of training their model gave an 0.0002 RMSE.

4.1 Google Stock Dataset

This data set comprises of stock price of Google. These values were also obtained from Yahoo finance [43]. Value from 19/08/2004 to 04/10/2019 were selected. We selected this range as in upcoming sections same range of the dataset is also used by other authors. Complete dataset consists of Open, Low, High, Close and total volume traded values. In this we were again interested in prediction Close values. Figure.16 shows plot of the data.

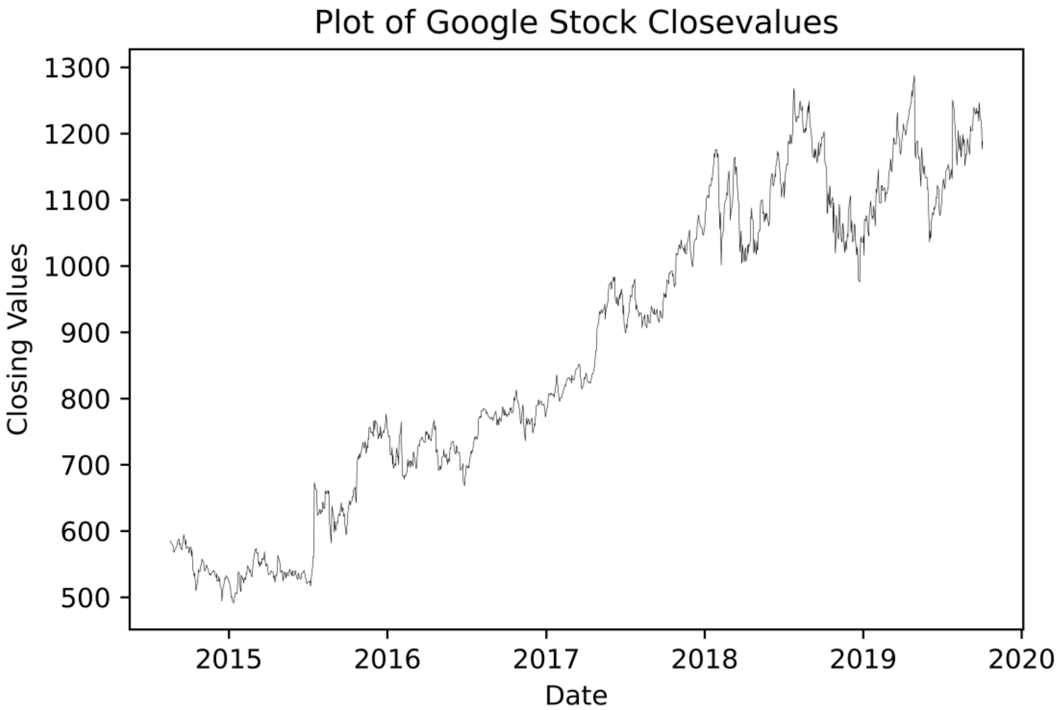


Figure. 16 Google stock closing price from 19 Aug 2004 to 04 Oct 2019

Google stock price has shown a gradual increase over the time period mentioned. However, it shows a cyclical pattern in the last quarter of the graph.

4.2 NASQAD IXIC Dataset

This data set comprises NASDAQ Composite (ˆIXIC) stock values from Jan 2009 – July 2019 [43]. Prices were obtained from yahoo finance. The data set includes Open, High, Low, Close, Adj Close, Volume columns for each day. We are interested in the prediction of the Closing value, so we extracted the closing value. Figure. 17 shows the plot of Close values for the selected period.

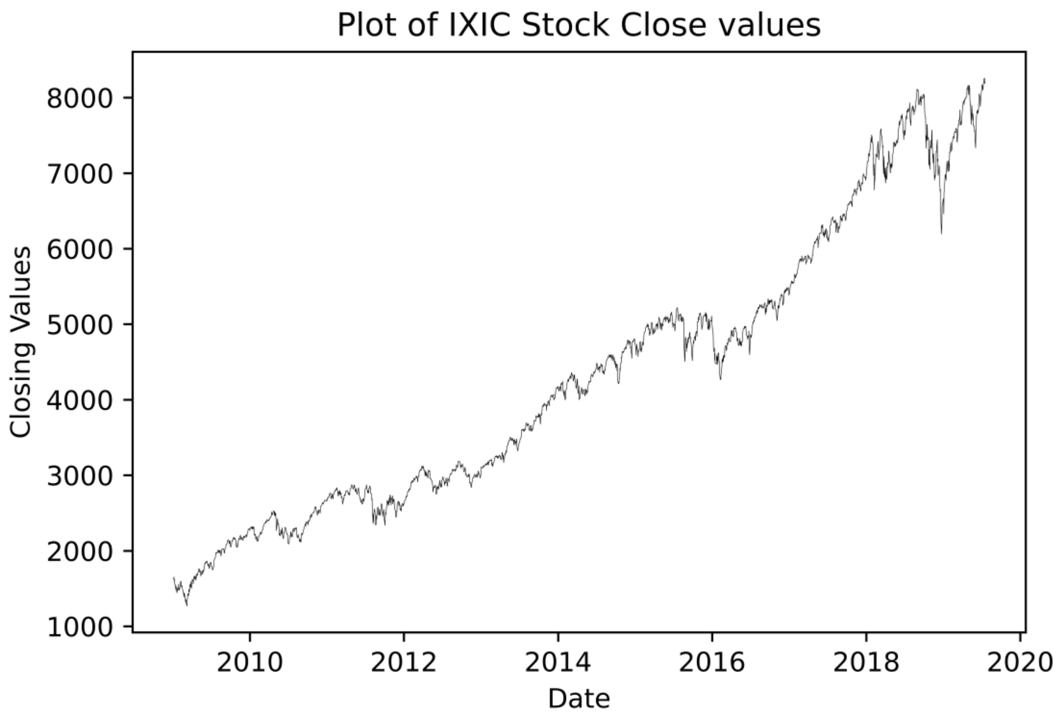


Figure. 17 NASQAD Composite Values Jan 2009 – July 2019

It is obvious from plot that price of this stock increased over the longer period. However, minor dips can be seen. Overall, the data doesn't contain a specific pattern that can be seen by naked eye. This makes the prediction a difficult task. Machine learning algorithms are used to find the pattern to find values.

4.4 State bank currency Exchange Rates

This data set comprises of hourly currency exchange rate of world-famous currencies w.r.t dollar. On each working day, 24 entries were recorded. As this a multi-dimensional data so we are going to show it in tabular form. Table No. 3 shows a tabular representation of such data.

Date Time	AUDUSD	EURUSD	GBPUSD	NZDUSD	CADUSD	CHFUSD	JPYUSD

1/2/2020 0:00	0.70188	1.121885	1.32501	0.67362	0.770639	1.033314	0.009197
1/2/2020 0:05	0.701825	1.121895	1.32497	0.67374	0.770707	1.033351	0.0092
1/2/2020 0:10	0.701875	1.121835	1.325305	0.673645	0.770651	1.033138	0.009199
1/2/2020 0:15	0.7017	1.121635	1.325055	0.67353	0.770579	1.032908	0.009199
1/2/2020 0:20	0.70174	1.12163	1.325135	0.673525	0.770573	1.032924	0.0092
1/2/2020 0:25	0.701615	1.121615	1.32519	0.67335	0.770496	1.032834	0.009201
1/2/2020 0:30	0.70151	1.121625	1.32518	0.67323	0.77049	1.032834	0.009201
1/2/2020 0:35	0.70141	1.121635	1.32526	0.67323	0.770499	1.032786	0.0092
1/2/2020 0:40	0.70142	1.1216	1.325305	0.67323	0.770538	1.032642	0.009199
1/2/2020 0:45	0.7014	1.12175	1.325695	0.67323	0.770541	1.032748	0.0092
1/2/2020 0:50	0.70161	1.121835	1.32567	0.67338	0.770591	1.032658	0.0092

Table No. 3 Currency Exchange rates w.r.t Dollar

Complete data set comprises 6 months' data (each 5-minute entry) except for weekends and public holidays. In total, there were more than 55000 entries that come out to be huge data. During preprocessing, we picked one entry per hour so that our data can be reduced a little. Our simplified data set comprised 3113 entries for all seven currency exchange rates. Our work has been done on this simplified data set for all the currencies.

In Table No. 3 we can see the complete data set. As discussed we have simplified our data to one entry per day. In Figure. 18 we can see the plot of Original values for AUDUSD values.

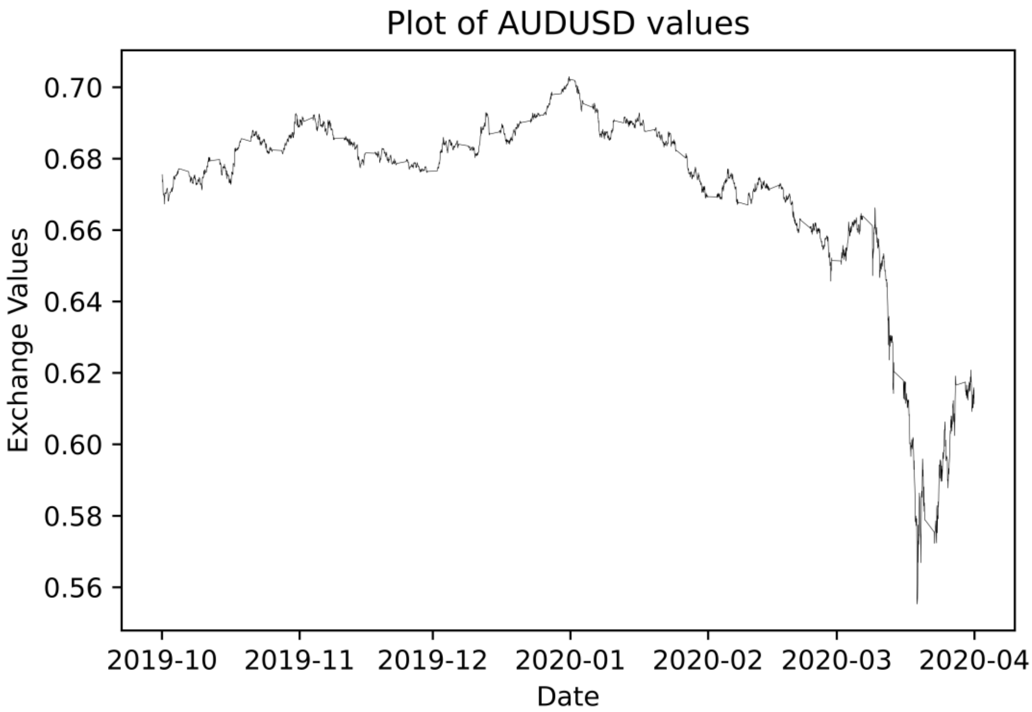


Figure. 18 Original values of AUDUSD

In Figure. 18 we can see on X – axis the time from Oct 2019 to March 2020. On Y – axis closing price of exchange values have been taken. Initially the exchange rate was about 0.67 and with gradual fluctuation between 0.66 and 0.7 it was constant till March 2020. In march 2020 we can see a deep drop in values to the lowest of 0.56 untill it recovered to 0.62 at end of March 2020.

4.2 Chapter Summary

In this Chapter we discussed the datasets that are used in further in this study. Mainly three data sets are modelled. Stock closing prices of google stock from Jan 2009 to 2019 data set has been studied and previous work results are also given. Next dataset that we work on is NASQAD IXIC stock dataset. LSTM and Bi-LSTM models after applying on these data sets were then applied to Currency exchange rates data set with 7 different exchanges rates. In further sections I reported research framework and results obtained.

CHAPTER NO. 5

RESEARCH FRAME WORK

CHAPTER NO. 5 RESEARCH FRAMEWORK

5.1 Overview

In our research, we used two different models to compare and contrast. The specifics of both models will be discussed in greater depth in the following section. As discussed in previous sections, we have run our models on three different data set. Out of these three data, two data sets were already published and authors have worked on these data set. They have announced the results of their study. We have applied our models to these two data sets and compared the accuracy of our predictions. In addition, we have also worked on third data set of state bank data. Two models are used in this piece of work. One architecture is of LSTM (Long Short Term Memory) and the other is Bi-LSTM. (Bi Directional Long Short Term Memory). Both models were written in the Python programming language.

These models take one-dimensional data as input and predict future value based on it. For both models 80 percent of data was used for training and rest was used for testing purpose. The following sections go over the structure and other specifics.

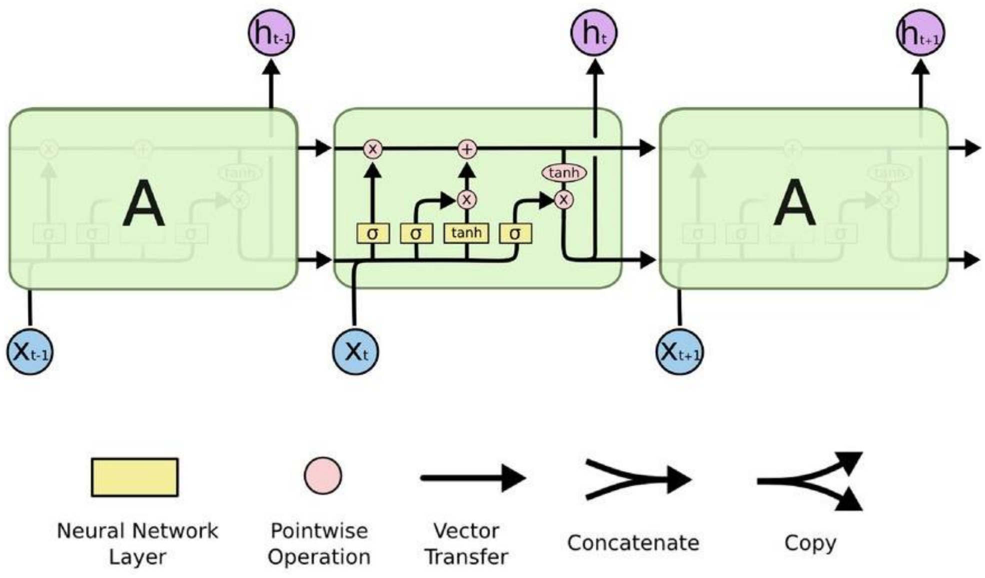


Figure. 19 Generalized structure of LSTM Network

In Figure 19 the structure of LSTM network can be seen. Figure shows a network of only 3 LSTM cells. Data points X_{t-1}, X_t and X_{t+1} are given as input and after passing through these cells we get h_{t-1}, h_t and h_{t+1} as output. Function operated at each gates were discussed in previous chapters in detail.

5.2 Structure of LSTM Model

For the LSTM model, we have used 3 LSTM layers, each having 100 input units and one dense output unit. The layers are represented in Figure. 19. The input layer takes 100 points from the data set, one at each node. The model based on these hundred values predict the 101th value. Similarly, for next iteration first value is not picked and next 100 values are fed to algorithm. Which results in forecasting of 102nd value. In this way whole data is entered to the model. So except of first 100 values all other values are predicted by system.

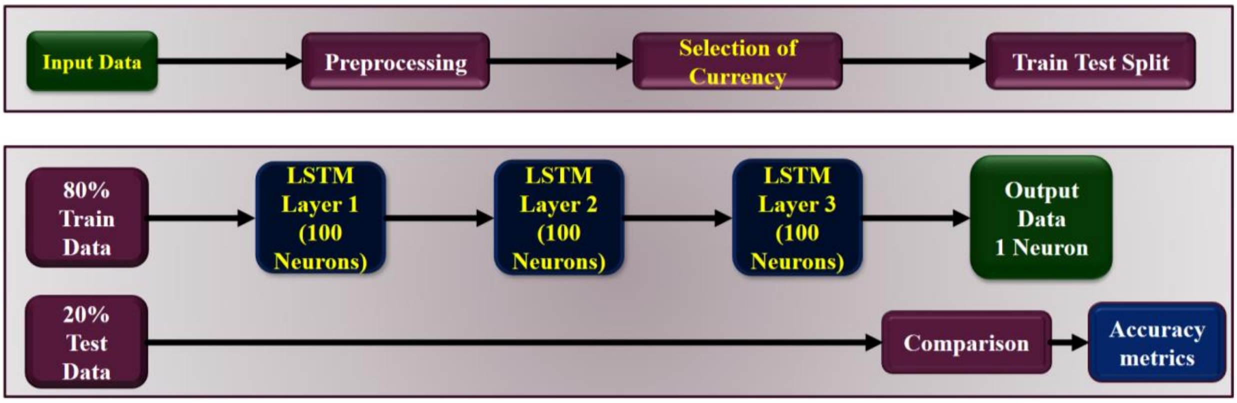


Figure. 20 Structure of LSTM Model

In Figure. 20 we can see the structure of LSTM model. We will discuss the methodology step wise in this section.

5.2.1 Preprocessing

In this step first we need to preprocess the data before feeding it to algorithms. As discussed in previous chapter we have worked on two types of data in this research. Google stock data and IXIC stock data we obtained from yahoo finance. While as currency exchange dataset is obtained from State Bank Officials. Yahoo finance provides data in xlxs format with seven columns and daily entry. Seven columns include open, low, high, close, adj close, volume values of same day. Data of IXIC stock exchange that was downloaded from Yahoo Finance contained a total of 2653 observations with 7 columns. After preprocessing the we only selected the Closing price for each day. Python pandas was used for preprocessing of data.

For Currency exchanges rates data was recorded at each 5 minute of interval for 6 months. As we know forex markets do not operate on weekends and other public holidays therefore data at such days was not provided. This data was had 7 currency exchange rate each with time stamp and a column that calculates the change in rate. So, that comes out to be 21 columns. Total number of observations in this dataset were 2,637,320 with 21 columns. After preprocessing hourly data points were selected as 5minutes data points did not had any significant change in the values. Also it added to the computational requirements. On basis of hourly selection, we had only 2653 instances with 7 columns each representing a currency pairs.

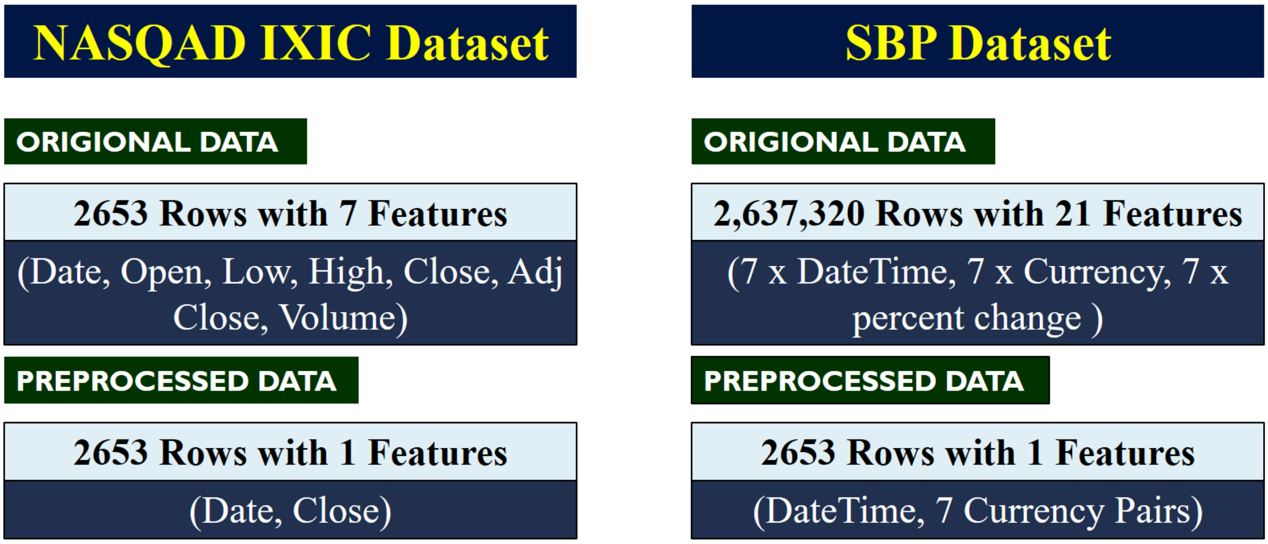


Figure. 21 Preprocessing of data

Figure. 21 represents the whole preprocessing data process for all data sets.

5.2.2 Feature Selection

In this section we selected the feature that we need to predict. In case of Yahoo Finance datasets, the closing values were selected for prediction purposes. And in case of SBP data the specific currency pair was selected for prediction purpose. The selected features in both cases made our data a one dimensional data that was further fed to model. Each feature selected for each dataset is summarized in given table.

Data Set	**Feature**
NASDAQ IXIC Stock	Close – Represents the closing value in dollars
Google Stock	Close – Represents the closing value in dollars
AUDUSD Currency Exchange Rate	AUDUSD – exchange rate
GBPUSD Currency Exchange Rate	GBPUSD – exchange rate

Table 4 Feature Selection Data set

5.2.3 Train Test Split

After feature selection our data was a simple one dimensional data indexed with date and timestamp. In this model we have selected 80-20 train test split. 80% of data was used for training purpose and 20% of data was used for testing purpose. As the data is indexed and future values depend on previous values so we didn't shuffle the data as it is done for other data sets. Starting 80% data was used for training and last 20% was used for testing evaluation purpose. We ran this model on 4 data sets i.e NASDAQ IXIC Stock, Google Stock, AUDUSD currency exchange pair and GBPUSD Currency exchange pair. The train test split duration of each dataset is summarized in given table.

Data Set	**Duration of Test Data**	**Duration of Test Data**
NASDAQ IXIC Stock	2009-01-02 to 2017-06-08	2017-06-09 to 2019-07-18
Google Stock	2014-08-19 to 2018-09-24	2018-09-25 to 2019-10-03
AUDUSD Currency Exchange Rate	2019-10-01 to 2020-02-24	2020-02-24 to 2020-03-31
GBPUSD Currency Exchange Rate	2019-10-01 to 2020-02-24	2020-02-24 to 2020-03-31

Table. 5 Test Train Split of data

In table 5 we can see that for each dataset duration of each dataset is listed. Date format listed is Year – Month – data format.

5.2.3 LSTM model layers

After train test split training data was the passed to LSTM layers. In our model we chose 3 LSTM layers each with 100 neurons as shown in Figure. 20. After these 3 LSTM layers this model has 1 output neuron. At each epoch weights of the networks were updated. In the Figure. 22 we can see the number of parameters that were trained at each epoch in order to accurately train the model. We can there is 1 input layer followed by a dropout layer. Input layer is responsible for giving data to the model. Drop out layer ignores some of the data received from previous layer. There are 2 LSTM layers that are made up of LSTM cells. Each layers contains 100 LSTM cells. Dense layers display the output value at output neuron.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 100, 100)	40800
dropout (Dropout)	(None, 100, 100)	0
lstm_1 (LSTM)	(None, 100, 100)	80400
lstm_2 (LSTM)	(None, 100)	80400
dropout_1 (Dropout)	(None, 100)	0
dense (Dense)	(None, 1)	101
Total params: 201,701 Trainable params: 201,701 Non-trainable params: 0		

Figure. 22 Total trainable and non-trainable parameters of LSTM model

We can see there are approximately 200,000 parameters that are calculated at each epoch. This model was trained for 50 epochs and Adam optimizer was used. The activation relu was used. Adam is a popular deep learning method since it produces good results quickly. There are four configuration parameters of Adam optimizer. Alpha, Beta1, Beta2 and Epsilon. Alpha is learning rate or step size. By default, it is 0.001. Larger values train the model fast but with low accuracy. Smaller values take more time to train but with better accuracy. Beta1 is the exponential decay rate for first moment approximation while as Beta 2 is rate for second moment approximation. While is epsilon is a very small number to prevent zero division. Default values that we have used are as under:

Alpha = 0.001, Beta1=0.9, Beta2=0.999, epsilon=1e-08

5.2.3 Evaluation Measure

After successfully training of the model it is now time to measure the accuracy or error of the model. For evaluation measure we have predicted the whole data through this trained model. After passing the data we received a list of predicted values. For fitting data, we used `model.predict` function to get predicted values list. At this stage we have original values and a list of predicted values. We have used accuracy metrics of Mean Squared Error as evaluation measure. We have used Python built-in function of `mean_squared_error` for calculating Mean Squared Error. Root mean squared error, percent error are some other evaluation metrics that can be used at this stage when we have actual and predicted values. In order to compare our results with already published results we have selected mean squared error metrics. The detail comparison is given in next chapter.

In the given table we will see the MSE for each dataset that was achieved by using LSTM model.

Data Set	**MSE**
NASDAQ IXIC Stock	0.00011
Google Stock	0.0009
AUDUSD Currency Exchange Rate	0.00014
GBPUSD Currency Exchange Rate	0.0003

Table 6 MSE values - LSTM Model

In Table 6 we can see the values achieved by model. These are the error values. We can the values are too low so that represents that performance of model is quite better. In next chapter we will compare our results our other published datasets.

5.3 Structure of Bi-LSTM Model

In this section we will discuss the structure of Bi-LSTM model. In Bi-Directional LSTM model, two layers each of 128 Neurons were used. First layer was the input layer and the other was hidden layer comprised of 128 Neuron. The last node was a single node that gives us next value of the series. At Input node 128 values were given through 128 nodes.

Model flow chart for both the models is described by following diagram shown in Figure.23

In figure. 23 we can see that this model also follows the same steps as in the LSTM model. In this Bi-LSTM model same preprocessed data was used a discussed in section 5.2.1. Similarly, the feature selected method was same as in previous model. Similarly, the train test split was also the same discussed in sections 5.2.3. However, the Layers in this models were only 2 each containing 128 Bi-Directional LSTM neurons. All other parameters are approximately same in this method.

In Bi-LSTM model the 2 Bi-LSTM layers were chosen each with 128 neurons followed by 1 output neuron. That means that based of 128 previous values 129th values were predicted by this model using Bi directional LSTM cells.

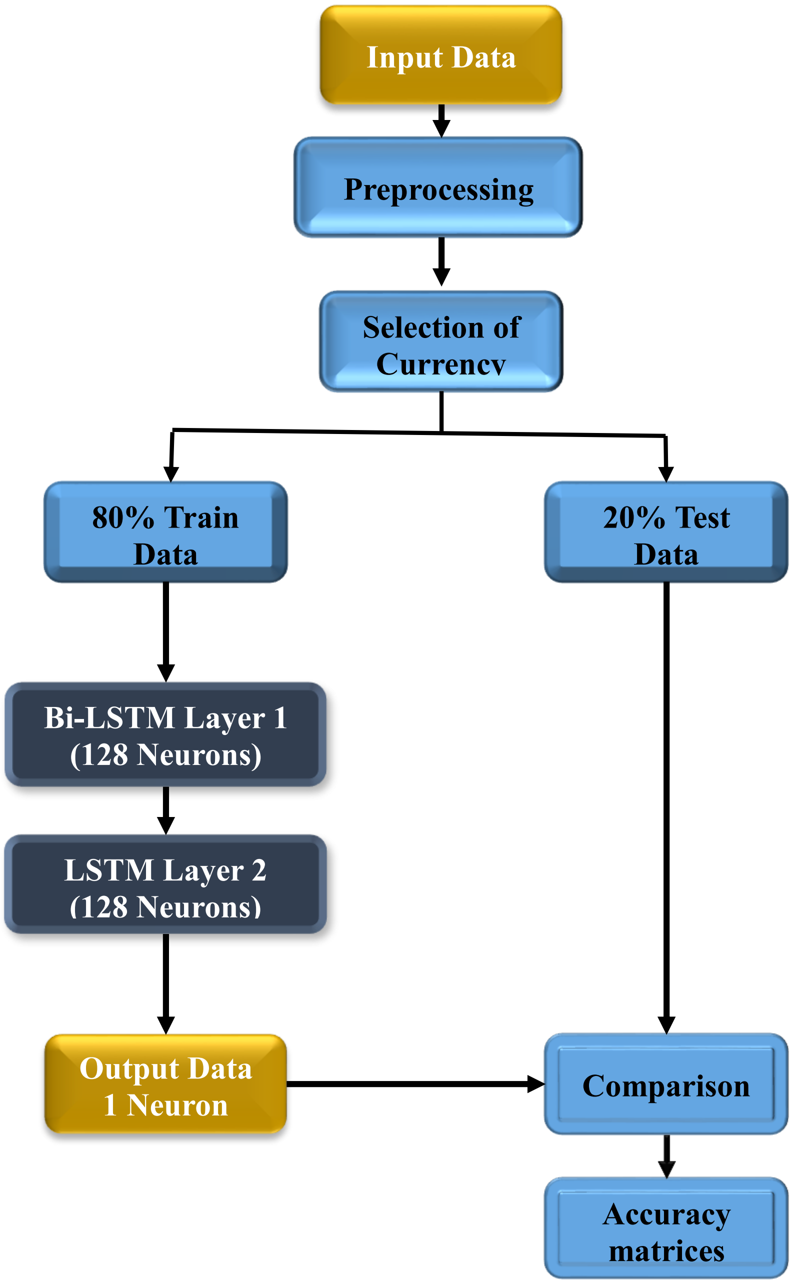


Figure. 23 Flow Chart of Bi LSTM Model

After passing the data through this models we had a set of predicted values. In order to the comparison we have calculated mean squared error of actual and predicted values. The MSE and original data and its predicted data plots will be discussed in detail in Chapter 6. As discussed in previous section we have selected MSE as evaluation metric. Bi-LSTM model was used to predicted the same 4 data sets. MSE achieved for Bi-LSTM model can be seen in following table.

Data Set	MSE
NASDAQ IXIC Stock	0.000114
Google Stock	0.0002
AUDUSD Currency Exchange Rate	3.251e-05
GBPUSD Currency Exchange Rate	8.4 x 10e-5

Table 7 MSE for Bi-LSTM model

Bi-LSTM model has shown very low MSE values for currency exchanges rates. However, stock values are also lowered. So Bi-LSTM model has shown greater accuracy and can be used confidently for prediction purposes.

In figure 24 we can see the trainable and non-trainable parameters that Bi-LSTM model had to calculate at each epoch.

Layer (type)	Output Shape	Param #
bidirectional (Bidirectional)	multiple	133,120
bidirectional_1 (Bidirectional)	multiple	394,240
dense (Dense)	multiple	257
Total params: 527,617 Trainable params: 527,617 Non-trainable params: 0		

Figure. 24 Total trainable and non-trainable parameters of Bi- LSTM model

In Figure. 24 we can see that we have approximately 5 lac parameters that are trained at each epoch based on the structure of the model.

We ran this model for 50 epochs with linear activation and Adam optimizer.

5.4 Hardware and computational requirements

Both models were trained over 50 epochs, and 80% of data was used for training purposes. These models were developed and run on Dell Notebook core i-5 with a processor of 2.8 GHz, 6GB RAM, and Windows 10 using Python 3.7 with Tensor Flow Version 2.4.1.

Training time for LSTM was approximately 60 -70 seconds per Epoch while as Bi-LSTM took on average of 3 minutes per epoch for an input of about 3000 data points.

These time calculations are approximate or average but it is clear that Bi LSTM model took approximately double the time of LSTM model for training.

Compared to the computational resources of [41] their model was developed and run on a MacBook Pro with a 2.2GHz Intel Core i7 processor, and 16 GB 1600 MHz DDR3 memory. While as Sunny [42] used Google co-laboratory with GPU and Ubuntu 18.04.3 LTS OS with 12 GB RAM as a simulation environment for the research.

5.4 Chapter Summary

In this methodology chapter we have discussed the structure of both LSTM and Bi-LSTM models in detail. We have seen how the data was preprocessed. Layers and structure of each model was discussed. Also we look at the trainable and non-trainable parameters for each model. We saw the accuracy metric of each model in this chapter. At the end we have put in the detail the hardware resources that we required to run these models.

CHAPTER NO. 6

RESULTS AND DISCUSSIONS

CHAPTER No. 6 RESULTS AND DISCUSSIONS

In our work we have LSTM and Bi-LSTM model. Using these two model a number of data sets were modelled. In this chapter we will see at the results of each model and discuss the results.

6.1 Results from LSTM Model

While training for 50 epochs, Mean Square Error was calculated after each epoch. This is referred to the cost function. Cost function for LSTM Model for Google stock Data Set is shown in Figure. 21.

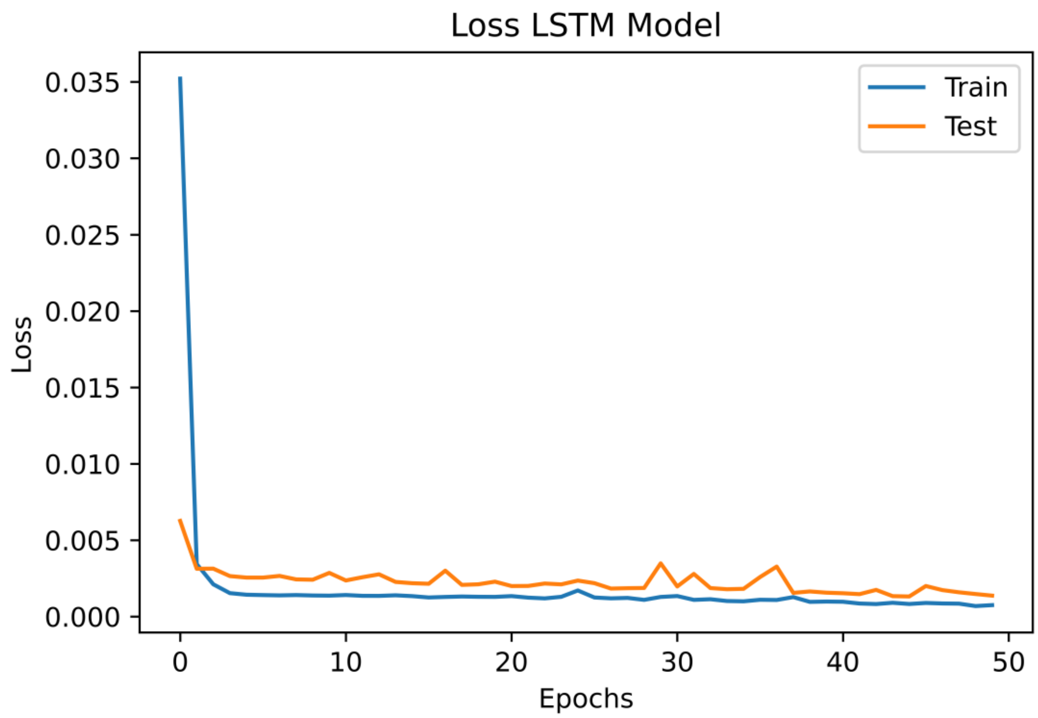


Figure. 25 Cost function of LSTM Model for 50 epochs Google Stock

Cost function shows that with each passing epoch, loss is being reduced. At 50th epoch both loss and validation loss is quite low. Graph shows that at 50th epoch both the losses are at its lowest position.

Following figures actual and predicted values of IXIC dataset, Google Data and State Bank Data state. Fig.21 shows actual vs predicted values of IXIC data set. The range for this data is the same as described before that is 19/08/2004 to 04/10/2019.

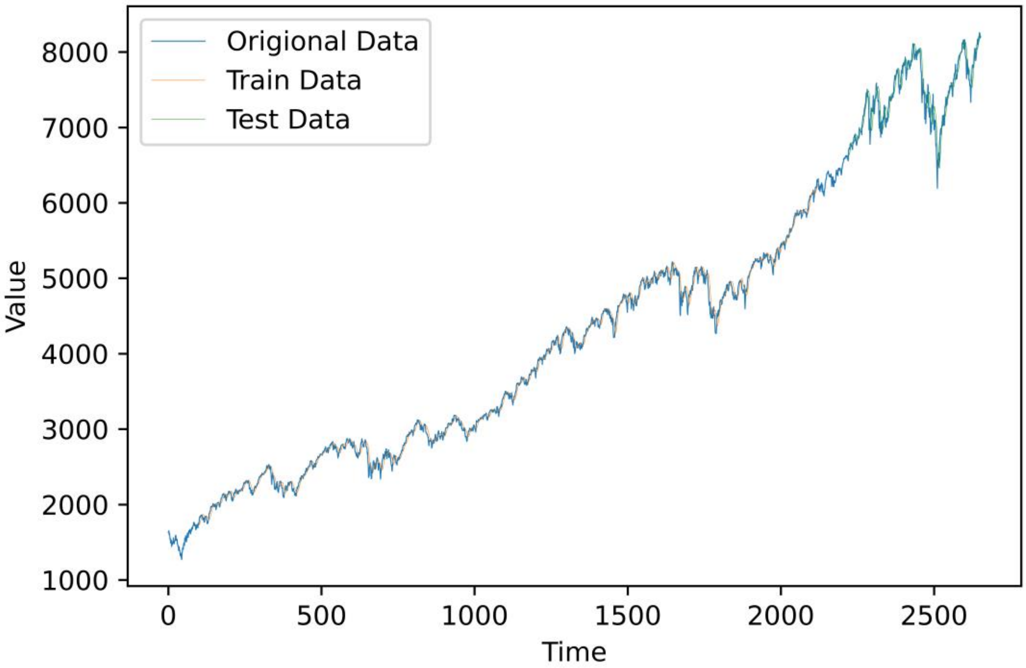


Figure. 26 Plot of Actual and Predicted values of IXIC Stock

In above plot, blue line indicates actual values NASQAD IXIC stock. Orange line shows predicted values for train data while as green line indicates predicted values of test data. From plot, we can see that both train and line closely follows the original line. Such representation shows very close and accurate prediction of the model. Further in terms of MSE we will see the results.

In Figure. 23 shows actual vs predicted plot of Google stock plotted using LSTM model.

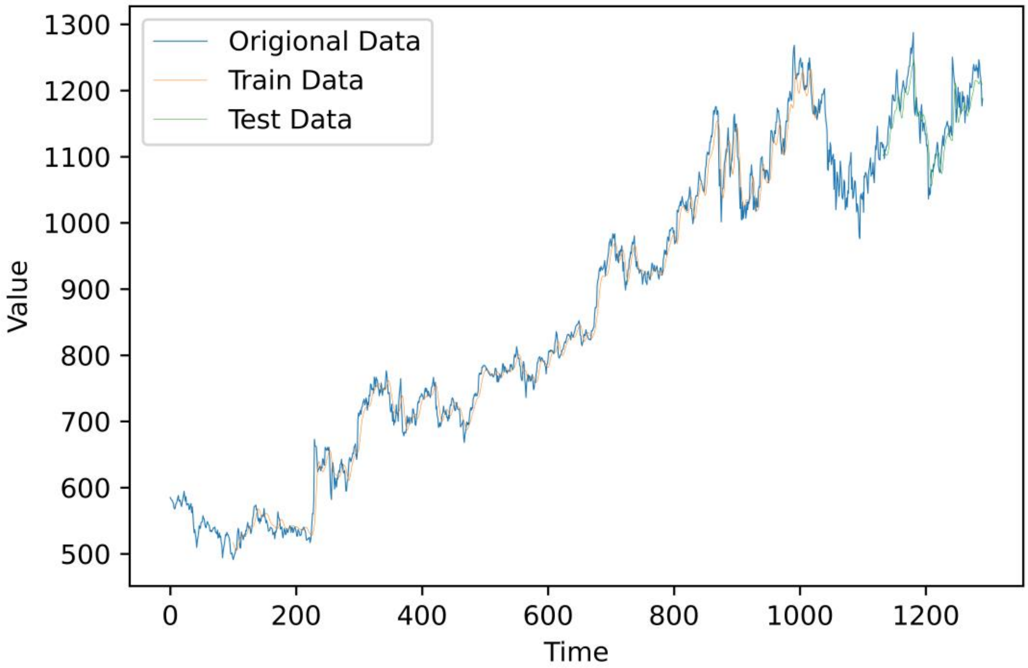


Figure. 27 Plot of Actual and Predicted values of Google Stock

In above plot, we can see that prediction points closely follow the original data points. Strong dips represent abrupt changes that were not predicted were closely. However, by looking at the graph the prediction looks fine. In this plot, blue line shows actual data points of Google stock while as blue and green lines show predicted value of test and train data set respectively.

In figure 24, 25 and 26, we will plot prediction of AUSUSD, GBPUSD and CADUSD currency exchange rates through LSTM model. These are the currency exchange rates for 3 different currencies obtained from dataset provided by State bank of Pakistan.

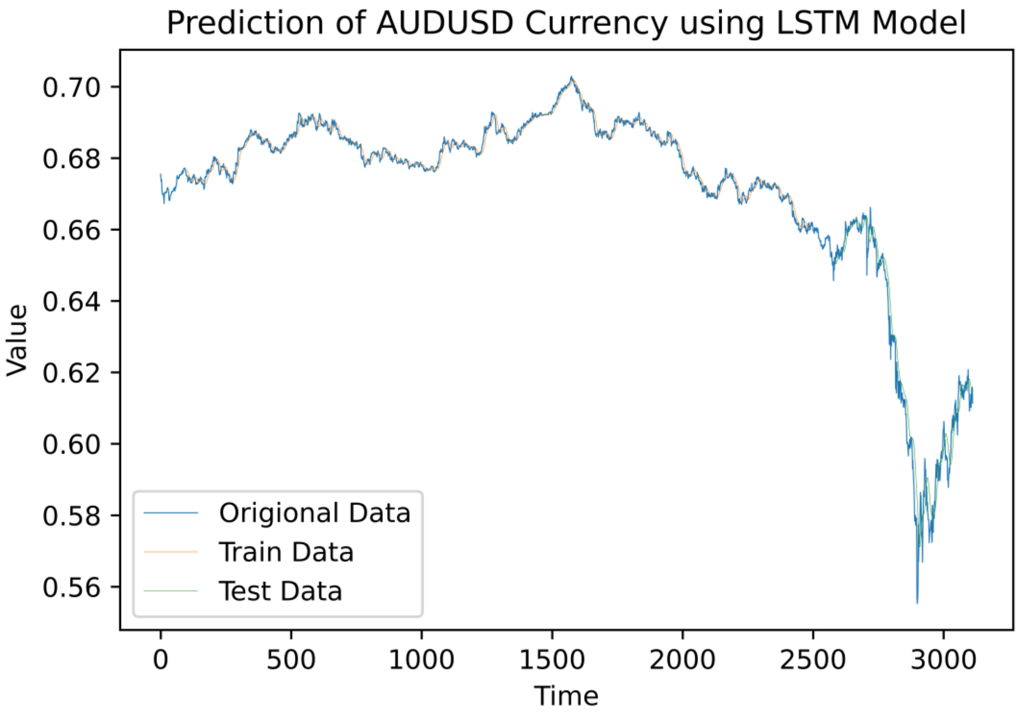


Figure. 28 Plot of Actual and Predicted values of AUDUSD Currency Exchange Rate

In Figure. 28 we can see the blue line that shows the original exchange rate value. Each point on X axis is one day on which exchange rates were recorded. On Y-axis we have values of exchange rates. Same pattern of Blue, Orange and green line for Original, train and test points have been followed here. We can see very close prediction by viewing the plot. Further by looking at Table. No 4 we will see the error in terms of MSE and RMSE for all the models and data sets. However, from plot both train and test forecasts seems quite accurate. Next we will look at other figures obtained from modelling GBPUSD and CADUSD.

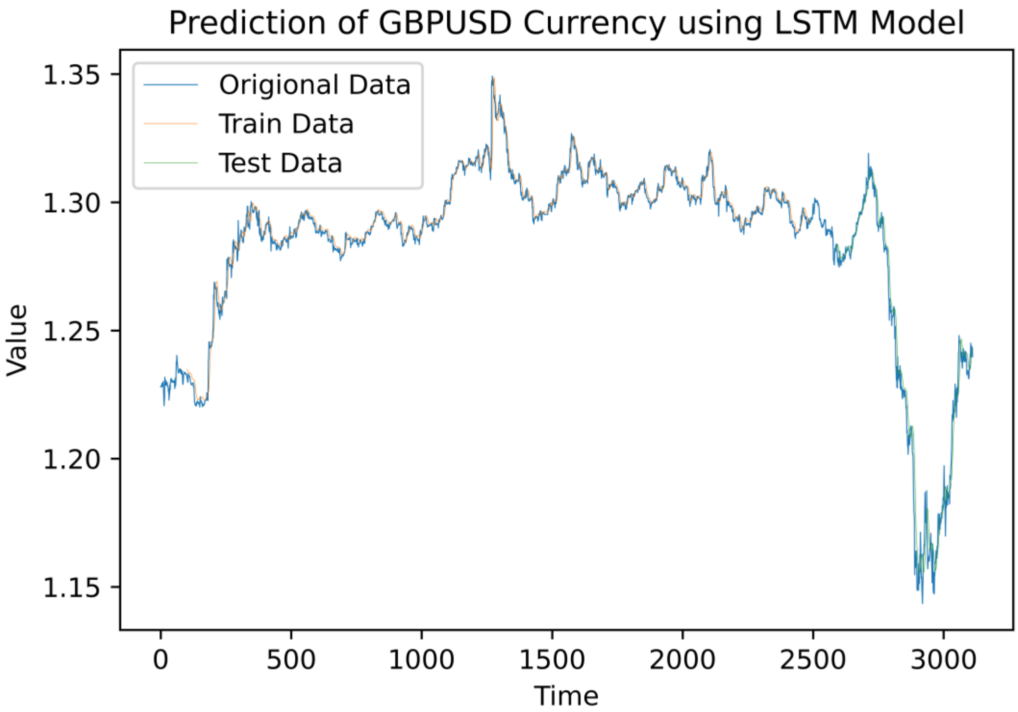


Figure. 29 Plot of Actual and Predicted values of GBPUSD Currency Exchange Rate

Figure. 29 shows results of modelling GBPUSD data on LSTM model. On x-axis we have days and on y-axis exchange values. We can see we have approximately 3300 data points out which 80% represented in Orange were used as training data and rest 20% was used for testing shown in green line. Original data is shown in blue line. Plot shows very close prediction as blue and green lines closely follows the blue line. Next we will see at another currency from same data set.

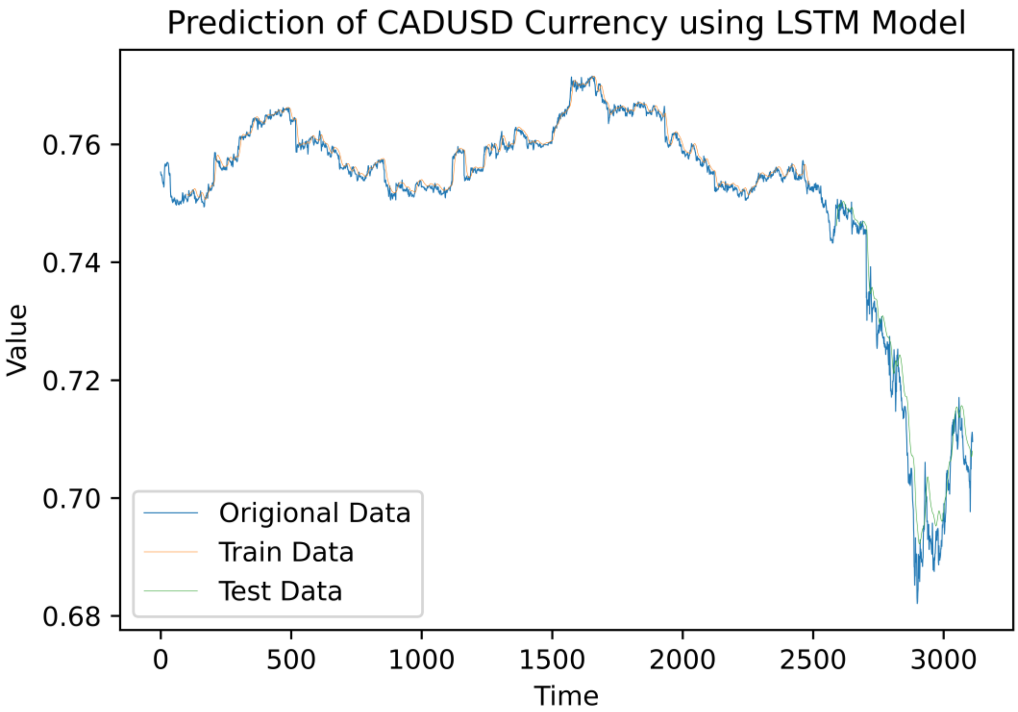


Figure. 30 Plot of Actual and Predicted values of CADUSD Currency Exchange Rate

CADUSD is exchange rate of US dollar to Canadian Dollar. In Figure. we have shown prediction plot of CADUSD data points using LSTM model.

In the next section we will see results of same data set from Bi-LSTM model.

6.2 Results from Bi-LSTM Model

In this part we will discuss and see results obtained for same data sets using Bi-LSTM model. First we will look at the loss function of the model. Bi-LSTM model was also trained for 50 epochs. In the same way as discussed earlier, loss function and predicted values are plotted so that efficiency of model can be seen. At the end of the plots, we will discuss the accuracies and see comparison with other authors results.

In Figure. 27, Loss function for AUDUSD currency exchange rate is plotted.

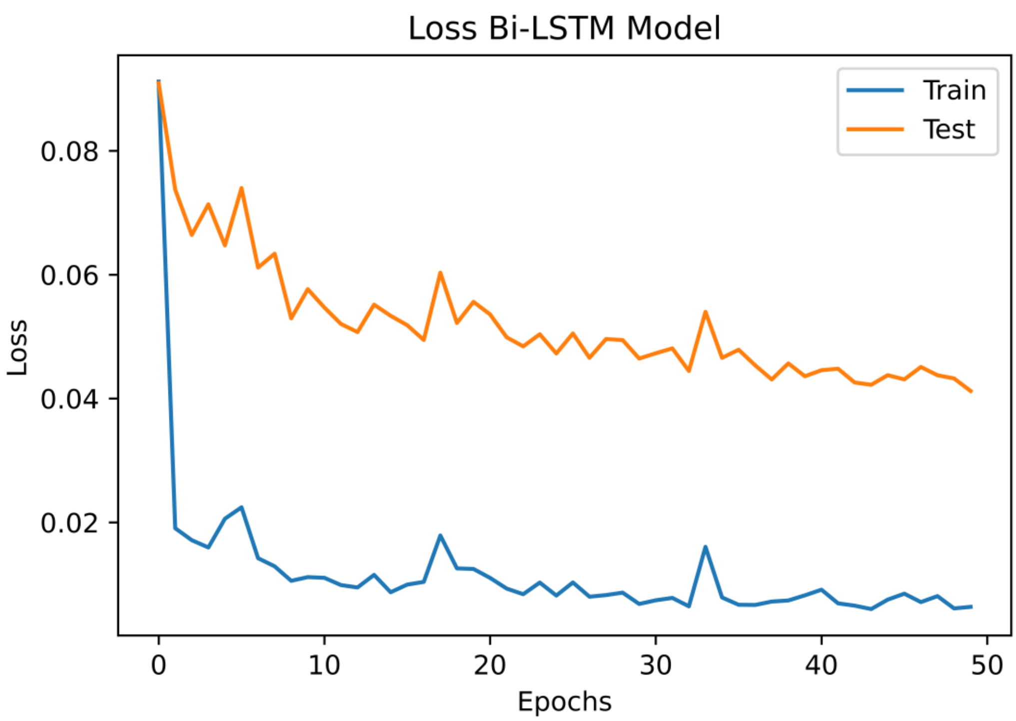


Figure. 31 Loss function of AUDUSD Currency Exchange Rate for Bi-LSTM Model

In the above Figure we can see the loss function for both train and test data. Although we have trained our model for 50 epochs for comparison purpose. Graph shows that further training of the model will improve the accuracy further. With each epoch both the train and test is loss is decreasing.

Using Bi-LSTM model I will show results of NASQAD IXIC plot. Following graph shows actual and forecasted values of NASQAD IXIC stock value for given period.

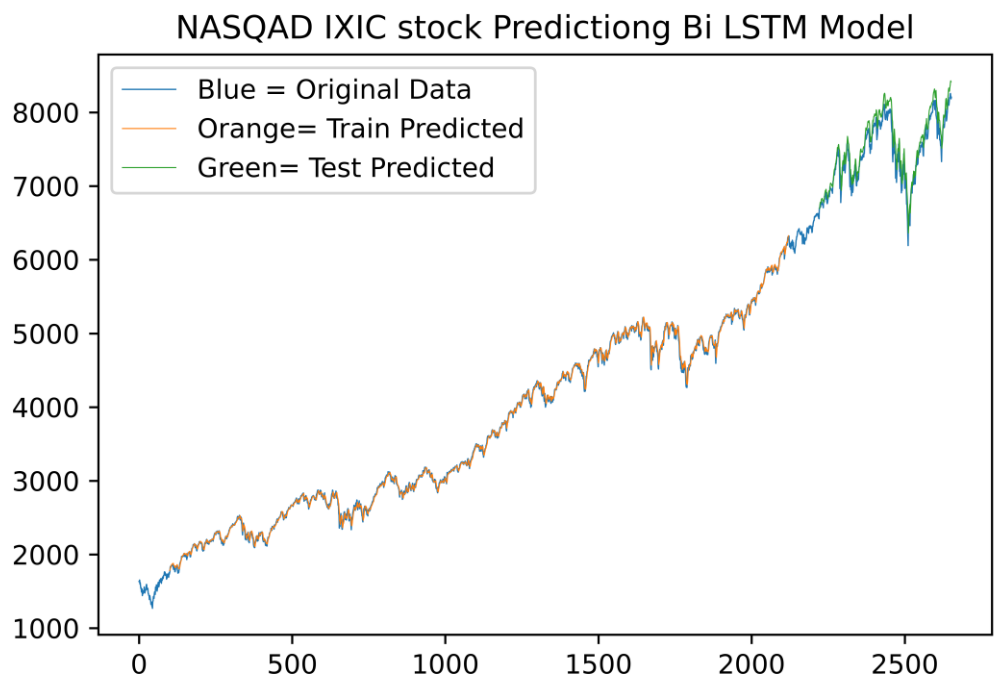


Figure. 32 Plot of Actual and Predicted values of NASQAD IXIC Stock Bi LSTM Model

Figure. 32 shows plot of Bi-LSTM model using NASQAD IXIC stock closing price. In this plot we can see the close following pattern by Orange and Green line that represents train and test data. In this plot on X-axis each point represents 1 day and on Y axis each point represent closing point of the stock on day. Although it seems a single line but in fact three different lines are plotted to see how accurate the model has produced results. In our work we have used Bi-LSTM model for all the data sets but in following figure I will represent another dataset result.

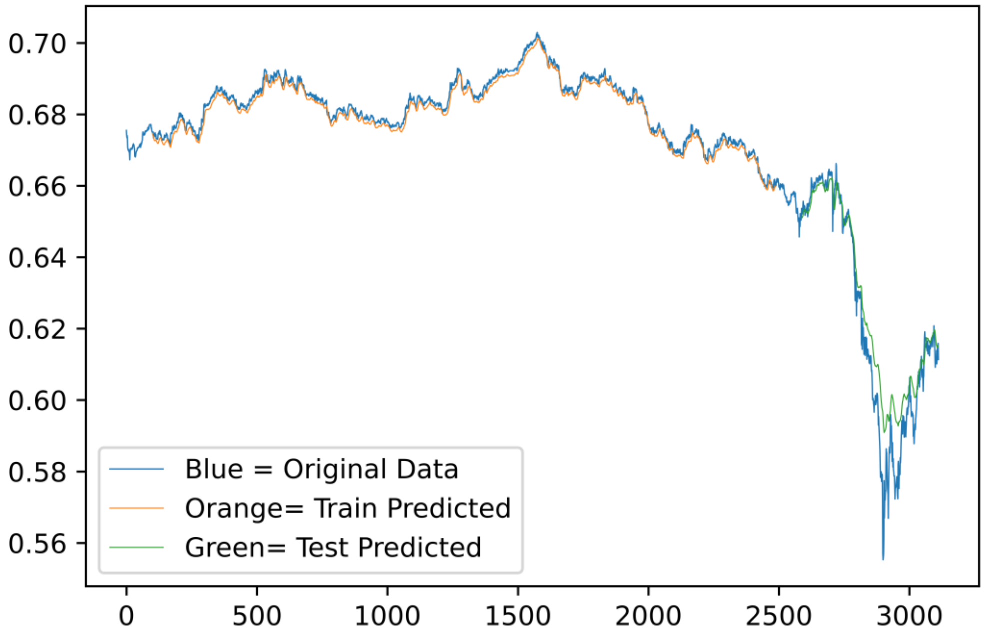


Figure. 33 AUDUSD Currency exchange rate prediction using Bi-LSTM

In Figure. 33 we can see the results of Bi-LSTM model on AUDUSD currency. This plot has also the same parameters as others have. Same Blue, Orange and Green line have been used. From plot we can see the test data was more accurately predicted if compare to test data. However, the test data prediction is also accurate despite the deep dips that ae usually difficult to follow.

Bi-LSTM model has shown better predicted as compared to LSTM model as we can see that abrupt changes were also predicted to some extent through this model. Same plots for Google Stock, other in data set were obtained.

In our study, we have used MSE and RMSE as accuracy metrics. In Table 2 we look at the results of other authors and our results on same data sets for comparison purposes.

6.3 Comparison with previous work

Data Set	Author	MSE - LSTM	MSE – Bi LSTM
IXIC Stock	[41]	0.0022	-
	Proposed Model	0.00011	0.000114
Google Stock	[42]	0.0004	0.0004
	Proposed Model	0.0009	0.0002
AUSUSD Currency Exchange rate	Proposed Model	0.00014	3.251e-05
GBPUSD Currency Exchange rate	Proposed Model	0.0003	8.4 x 10e-5

Table No. 8 Comparison of MSE for LSTM and Bi-LSTM model on different Datasets

In the Table No. 4 we can see that reported MSE value for IXIC model is 0.0022 while proposed LSTM and Bi-LSTM model have reduced the error to 0.00011. For this specific dataset, the MSE has been significantly reduced however, there is no significant difference between both the models.

If we look at the MSE of Google Stock data, LSTM model has not shown better accuracy but Bi-LSTM model has shown better efficiency.

For currency exchange rate dataset, we seen that there is significant improvement from LSTM to Bi-LSTM model.

CHAPTER NO. 7

CONCLUSION AND FUTURE WORK

CHAPTER No. 7 CONCLUSION AND FUTURE RESEARCH

In light of the findings, we may infer that Bi-LSTM models have significantly improved the outcomes. Bi-LSTM models may be used successfully for the prediction of time-series data, and they have a number of advantages. It is self-evident that the model we built generates the best-predicted results in the majority of situations. It should be noted that the model was run on a modest PC and that training does not need the use of many resources. Using this approach, it is possible to create an application that obtains real-time market values of different stocks or currency exchange rates and forecasts their future values in real time. It is possible that the same forecasted values may be valuable to financial market investors or other stakeholder groups in the future. This can assist investors in gaining a significant financial benefit while also maintaining a stable atmosphere in the stock market. In the future, we intend to study data from a larger number of stock markets from a variety of different categories in order to evaluate the success of our strategy.

As discussed in Chapter No. 2 we saw that a number of mobile and web application use indicators to help users make decisions while trading. Keeping in view the same approach we can develop a machine learning based indicators that can be used in these or another application. These newly developed indicator might be pre-trained or might have the ability to get trained from the fresh data accessed.

In our work we have developed the model that gave us considerably greater accuracy. Next we need to deploy this algorithm to actual trading applications. This work can be further extended to a mobile application as well as desktop application. It can also be integrated with an existing trading application. One of the famous trading application is IQ Option. If one uses that application, we can see it gives us a number of indicators. Most of these were discussed in earlier chapters. A new category of LSTM and Bi-LSTM indicators can be deployed under ML indicators section.

References:

[1] A. Sangvinatsos, "A random walk down wall street: The time-tested strategy for successful investing," ed: Taylor & Francis, 2017.

[2] T. Chenoweth and Z. Obradović, "A multi-component nonlinear prediction system for the S&P 500 index," *Neurocomputing,* vol. 10, no. 3, pp. 275-290, 1996.

[3] F. Fernandez-Rodrıguez, C. Gonzalez-Martel, and S. Sosvilla-Rivero, "On the profitability of technical trading rules based on artificial neural networks:: Evidence from the Madrid stock market," *Economics letters,* vol. 69, no. 1, pp. 89-94, 2000.

[4] M. Ghiassi, H. Saidane, and D. Zimbra, "A dynamic artificial neural network model for forecasting time series events," *International Journal of Forecasting,* vol. 21, no. 2, pp. 341-362, 2005.

[5] H. Niu and J. Wang, "Volatility clustering and long memory of financial time series and financial price model," *Digital Signal Processing,* vol. 23, no. 2, pp. 489-498, 2013.

[6] A. Lendasse, E. de Bodt, V. Wertz, and M. Verleysen, "Non-linear financial time series forecasting-Application to the Bel 20 stock market index," *European Journal of Economic and Social Systems,* vol. 14, no. 1, pp. 81-91, 2000.

[7] A. Abraham, B. Nath, and P. K. Mahanti, "Hybrid intelligent systems for stock market analysis," in *International Conference on Computational Science*, 2001: Springer, pp. 337-345.

[8] O. N. Uglitskikh, I. I. Glotova, Y. E. Klishina, and E. P. Tomilina, "Financial Market Development Trends," vol. 206, ed, 2021.

[9] M. O. Ahunov, L. Van Hove, and M. Jegers, "Banking Market Structure, Institutions, and SMEs' Access to Finance in Transition Economies: The Case of Ukraine," *Institutions, and SMEs' Access to Finance in Transition Economies: The Case of Ukraine (October 15, 2010),* 2010.

[10] A. Hayes. "Financial Markets." https://www.investopedia.com/terms/f/financial-market.asp (accessed on 25th Feb 2022.)

[11] B. Monisha, "A Study Of Forex Market And Analysis Of Major Currrencies At Jmarathon Advisory Services Pvt Ltd," 2020.

[12] A. B. Suthar, H. R. Patel, and S. M. Parikh, "A comparative study on financial stock market prediction models," *The International Journal of Engineering And Science (IJES),* vol. 1, no. 2, pp. 188-191, 2012.

[13] A. Czapkiewicz, P. Jamer, and J. Landmesser, "Effects of macroeconomic indicators on the financial markets interrelations," *Finance a Uver,* vol. 68, no. 3, pp. 267-292, 2018.

[14] P. Sahoo and K. Charlapally, "Stock price prediction using regression analysis," *International Journal of Scientific & Engineering Research,* vol. 6, no. 3, pp. 1655-1659, 2015.

[15] S. Lauren and S. D. Harlili, "Stock trend prediction using simple moving average supported by news classification," in *2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 2014: IEEE, pp. 135-139.

[16] A. Hayes. "Simple Moving Average (SMA)." https://www.investopedia.com/terms/s/sma.asp (accessed on 25th Feb 2022.)

[17] M. Fikri, S. Herdjunanto, and A. Cahyadi, "On the Performance Similarity Between Exponential Moving Average and Discrete Linear Kalman Filter," in *2019 Asia Pacific Conference on Research in Industrial and Systems Engineering (APCoRISE)*, 2019: IEEE, pp. 1-5.

[18] A. Isiaka, A. Isiaka, and A. Isiaka, "Forecasting with ARMA models: A case study of the exchange rate between the US Dollar and a unit of the British Pound," *International Journal of Research in Business and Social Science (2147-4478),* vol. 10, no. 1, pp. 205-234, 2021.

[19] D. Benvenuto, M. Giovanetti, L. Vassallo, S. Angeletti, and M. Ciccozzi, "Application of the ARIMA model on the COVID-2019 epidemic dataset," *Data in brief,* vol. 29, p. 105340, 2020.

[20] "https://www.stockbrokers.com/guides/mobile-trading." (accessed on 25th Feb 2022.)

[21] R. B, "Pattern Recognition and Neural Networks Cambridge University Press," 1996.
[22] S. Furui, "Generalization problem in ASR acoustic model training and adaptation," in *2009 IEEE Workshop on Automatic Speech Recognition & Understanding*, 2009: IEEE, pp. 1-10.
[23] K. S.D, S.N, Revankar, "Analysis of classification by supervised and unsupervised learning, in: Proceedings," *International Conference on Computational Intelligence and Multimedia Applications,* pp. 280–284, 2008, doi: 10.1109/ICCIMA.2007.83.
[24] A. Navlani. https://www.datacamp.com/community/tutorials/k-nearest-neighbor-classification-scikit-learn?utm_source=adwords_ppc&utm_campaignid=1455363063&utm_adgroupid=65083631748&utm_device=c&utm_keyword=&utm_matchtype=b&utm_network=g&utm_adpostion=&utm_creative=278443377095&utm_targetid=aud-392016246653:dsa-429603003980&utm_loc_interest_ms=&utm_loc_physical_ms=1011086&gclid=CjwKCAjwndCKBhAkEiwAgSDKQeV1vULVKSr4PGV1M-0ifc303HiDiOTkh2LAid1pL7VDJC0bG-vt5RoCJ94QAvD_BwE (accessed on 25th Feb 2022.)
[25] R. Ehsani and F. Drabløs, "Robust Distance Measures for k NN Classification of Cancer Data," *Cancer informatics,* vol. 19, p. 1176935120965542, 2020.
[26] S. Siledar and S. Chaudhary, "Comparative analysis of naive bays classifier and decision tree c4. 5 on credit payment data set," *International Journal of Research in Engineering and Technology,* vol. 6, no. 4, pp. 43-44, 2017.
[27] S. Ting, W. Ip, and A. H. Tsang, "Is Naive Bayes a good classifier for document classification," *International Journal of Software Engineering and Its Applications,* vol. 5, no. 3, pp. 37-46, 2011.
[28] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics,* vol. 21, no. 3, pp. 660-674, 1991.
[29] Z. Soumaya, B. D. Taoufiq, N. Benayad, K. Yunus, and A. Abdelkrim, "The detection of Parkinson disease using the genetic algorithm and SVM classifier," *Applied Acoustics,* vol. 171, p. 107528, 2021.
[30] A. K. Jain, J. Mao, and K. M. Mohiuddin, "Artificial neural networks: A tutorial," *Computer,* vol. 29, no. 3, pp. 31-44, 1996.
[31] O. I. Abiodun *et al.*, "Comprehensive review of artificial neural network applications to pattern recognition," *IEEE Access,* vol. 7, pp. 158820-158846, 2019.
[32] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd international conference on Machine learning*, 2006, pp. 369-376.
[33] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.
[34] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE transactions on Signal Processing,* vol. 45, no. 11, pp. 2673-2681, 1997.
[35] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction," *arXiv preprint arXiv:1801.02143,* 2018.
[36] C. Zhou, C. Sun, Z. Liu, and F. Lau, "A C-LSTM neural network for text classification," *arXiv preprint arXiv:1511.08630,* 2015.
[37] S. Dai, L. Li, and Z. Li, "Modeling vehicle interactions via modified LSTM models for trajectory prediction," *IEEE Access,* vol. 7, pp. 38287-38296, 2019.
[38] R. Navares and J. L. Aznarte, "Predicting air quality with deep learning LSTM: Towards comprehensive models," *Ecological Informatics,* vol. 55, p. 101019, 2020.
[39] F. Shahid, A. Zameer, and M. Muneeb, "Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM," *Chaos, Solitons & Fractals,* vol. 140, p. 110212, 2020.
[40] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Computing and Applications,* vol. 32, no. 13, pp. 9713-9729, 2020.

[41] S. O. Ojo, P. A. Owolawi, M. Mphahlele, and J. A. Adisa, "Stock market behaviour prediction using stacked LSTM networks," in *2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, 2019: IEEE, pp. 1-5.

[42] M. A. I. Sunny, M. M. S. Maswood, and A. G. Alharbi, "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model," in *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, 2020: IEEE, pp. 87-92.

[43] Yahoo Finance. https://finance.yahoo.com/ (accessed on 25th Feb 2022.)