# 

**PROJECT DOCUMENT**

**STOCK MONIT**

**SUBMITTED BY**

**HARSH PRAKASH KAMBLE**

**IN PARTIAL FULFILLMENT OF**

**M.Sc. (IT) Part-II (Sem-3)**

## UNIVERSITY OF MUMBAI

## DEPARTMENT OF IT

**Parle Tilak Vidyalaya Association’s**

**MULUND COLLEGE OF COMMERCE**

**ABSTRACT**

STOCK MONIT is a Web and Android based application which helps to analyse the data of previous performance of particular stock and predicts what could be the outcoming of the performance of the stock in future. It basically predicts how a stock could perform with the help of analysed data. In this project we attempt to implement machine learning approach to predict stock prices. Machine learning is effectively implemented in forecasting stock prices. The objective is to predict the stock prices in order to make more informed and accurate investment decisions. We propose a stock price prediction system that integrates mathematical functions, machine learning, and other external factors for the purpose of achieving better stock prediction accuracy and issuing profitable trades. Stock market is place where people buy and sell shares of publicly listed companies. Every buyer and seller try to predict the stock market price movements to get maximum profits and minimum losses. Using cutting edge technology such as AI can improve prediction stock price. In the procedure of considering strategies and variables to be considered, we found ML algorithmics such as Random forest, LSTM, SVM, ANN was not fully utilized. In this model we will introduce and review more a possible way to predict stock movements with high accuracy. The first thing we considered is data of previous year's share market prices, historical prices of currency and commodity market and the historical news headlines. The datasets were pre-processed and prepared for actual analysis. Therefore, our model will also focus on preprocessing of datasets. Second, after processing the datasets earlier, we will review the use of major AI technique for that data and productive results. In addition, the proposed system evaluates the application of the forecast system to the real-world scenario and the problems associated with the accuracy of the total values provided. The high accuracy and profitability was achieved when results of all algorithms are combined and considered all factors affecting the stock prices. Successful valuation prediction of share price can become a big asset for stock market firms and provide real life solutions to the difficulties faced by stock market individual investors have.

**Objectives of the Project:**

1. Minimize manual calculations for stock performance prediction.
2. Easily carry out complex tasks to predict the stock price.
3. Analyse the data in csv format from last decade and predict the upcoming stock price.

**Hardware Requirements:**

• RAM: 8 GB or more

• Storage: 500 GB

• CPU: 2 GHz or faster

• Architecture: 32-bit or 64-bit Software Requirements:

• Python 3.5 or above in Anaconda Navigator and Jupyter Notebook is used for data pre-processing, model training and prediction.

• AlphaVantage API for fetching datasets in json and csv format.

• Flask for creating Restful API.

• Android Studio

• Operating System: windows 7 and above or Linux based OS or MAC OS.

# Table of Contents

[Table of Contents 2](#_2et92p0)

[Chapter 1: Introduction 4](#_3dy6vkm)

[1.1 Introduction 5](#_1t3h5sf)

[1.2 What is Stock Market ? 6](#_4d34og8)

[1.2 Need for stock prediction 7](#_2s8eyo1)

[1.3 Motivation to Work 8](#_17dp8vu)

[1.4 Problem Statement 9](#_3rdcrjn)

[Chapter 2: Review of Literature 10](#_26in1rg)

[2.1 Introduction 9](#_1t3h5sf)

[2.2 Existing Methods 9](#_1t3h5sf)

[Chapter 3: Methodology 1](#_1ksv4uv)9

[3.1 Proposed Systems 1](#_1t3h5sf)9

[Chapter 4: Proposed Design 30](#_1pxezwc)

[4.1 Structure Chart 2](#_1t3h5sf)9

[4.2 UML Diagrams 2](#_1t3h5sf)9

[4.2.1 Use Case Diagram 3](#_1t3h5sf)0

[4.2.2 Sequence Diagram 3](#_1t3h5sf)1

[4.2.3 Activity Diagram 3](#_1t3h5sf)1

[4.2.4 Collaboration Diagram 3](#_1t3h5sf)2

[4.2.5 Flow Chart 3](#_1t3h5sf)3

[4.2.6 Component Diagram 3](#_1t3h5sf)4

**Table of Figures**

[Figure 1 LMS Inputs and Outputs 21](#_z337ya)

[Figure 2 LMS Updating Weights 21](#_3j2qqm3)

[Figure 3 Delay System 23](#_2xcytpi)

[Figure 4 LSTM Architecture 24](#_1ci93xb)

[Figure 5 Preprocessing Data 28](#_2bn6wsx)

[Figure 6 Overall Architecture 29](#_3as4poj)

[Figure 7 Training and Predictions 30](#_2p2csry)

[Figure 8 Using LMS, LSTM and LSTM with LMS in the system 32](#_3o7alnk)

[Figure 9 Execution based on Model Selection 33](#_23ckvvd)

[Figure 10 Execution based on Algorithm Selection 34](#_ihv636)

[Figure 11 Data Transfer between modules 35](#_32hioqz)

[Figure 12 Flow of Execution 36](#_1hmsyys)

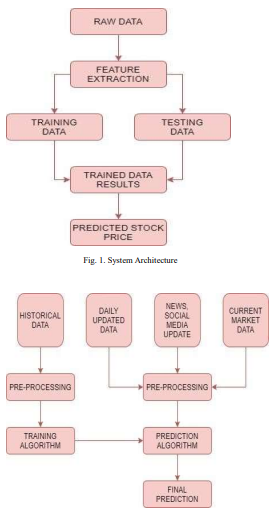
[Figure 13 Components present in the system 37](#_41mghml)

# Chapter 1: Introduction

### Introduction

Nowdays in this world, due to increasing inflation in economy and various other aspects, it is quite hard for one to earn and easily manage their income. Most of the people tend to save or create tangible assets to save money or reserve it for further use. Few people could even suggest to invest in stock market but due to incomplete information regarding the gain and losses, most people have no interest in trading their capital. However, stock market prediction and analysis are some of the most difficult jobs to complete. There are numerous causes for this, including market volatility and a variety of other dependent and independent variables that influence the value of a certain stock in the market. These variables make it extremely difficult for any stock market expert to anticipate the rise and fall of the market with great precision.

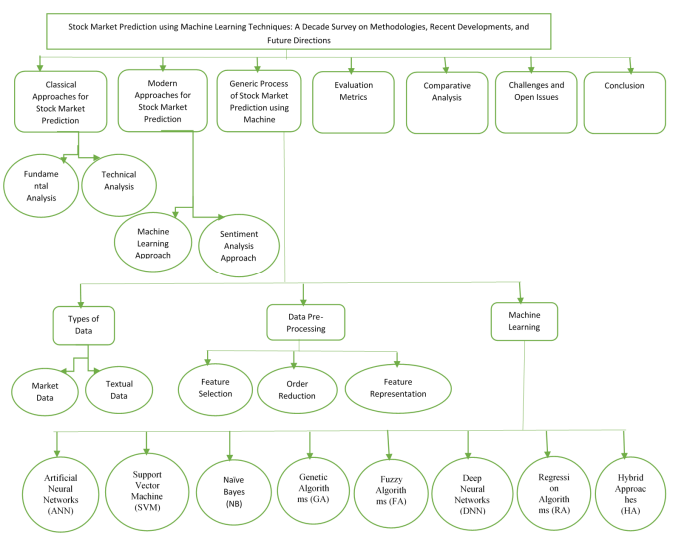
Predicting the Stock Market has been the goal of investors since its existence. Everyday billions of dollars are traded on the exchange, and behind each dollar is an investor hoping to profit in one way or another. Should an investor be able to accurately predict market movements, it offers tantalizing promises of wealth and influence. It is no wonder then that the Stock Market and its associated challenges find their way into the public imagination every time it misbehaves. The share market is a compilation of different people buying and selling the shares. Mostly known as stock (stake)which generally refers to claims of ownerships over a business by an individual or group of individuals. The way of finding the future valuation of the stock market prices is called the stock market estimate. Expected to be Strong, accurate and effective. The system should work in line with real life scenarios and be well connected to that. The movement in the stock market is usually determined by the sentiments of thousands of investors [3]. These events are political events such as the statements of a ministers or government officials, statement of government bodies such as RBI, SEBI, scandal news, etc. It can also be the global happening such as rapid movements in currencies prices and commodities prices [8]. All this thing affects the earning of companies, which ultimately affects the sentiment of stock market investors. This is beyond the reach of almost all individual to assess these accurately and consistently. This method usually requires collection various social media data, news that affect stock market investors sentiment and the feelings expressed by individuals. Other data such as last year’s stock prices are also considered. The relationship between different data points. Is considered and an estimation is done using these variety of data points.



However, with the introduction of Machine Learning and its strong algorithms, the most recent market research and Stock Market Prediction advancements have begun to include such approaches in analyzing stock market data.

In summary, Machine Learning Algorithms are widely utilized by many organizations in Stock market prediction. This article will walk through a simple implementation of analyzing and forecasting the stock prices of a Popular Worldwide Online Retail Store in Python using various Machine Learning Algorithms.

The financial market is a dynamic and composite system where people can buy and sell currencies, stocks, equities and derivatives over virtual platforms supported by brokers. The stock market allows investors to own shares of public companies through trading either by exchange or over the counter markets. This market has given investors the chance of gaining money and having a prosperous life through investing small initial amounts of money, low risk compared to the risk of opening new business or the need of high salary career. Stock markets are affected by many factors causing the uncertainty and high volatility in the market. Although humans can take orders and submit them to the market, automated trading systems (ATS) that are operated by the implementation of computer programs can perform better and with higher momentum in submitting orders than any human. However, to evaluate and control the performance of ATSs, the implementation of risk strategies and safety measures applied based on human judgements are required. Many factors are incorporated and considered when developing an ATS, for instance, trading strategy to be adopted, complex mathematical functions that reflect the state of a specific stock, machine learning algorithms that enable the prediction of the future stock value, and specific news related to the stock being analysed. Time-series prediction is a common technique widely used in many real-world applications such as weather forecasting and financial market prediction. It uses the continuous data in a period of time to predict the result in the next time unit. Many timeseries prediction algorithms have shown their effectiveness in practice. The most common algorithms now are based on Recurrent Neural Networks (RNN), as well as its special type - Long-short Term Memory (LSTM) and Gated Recurrent Unit (GRU). Stock market is a typical area that presents time-series data and many researchers study on it and proposed various models. In this project, LSTM model is used to predict the stock price.



### 1.2 What is Stock Market ?

Investing in equities is an important investment that we make in order to generate inflation beating returns. This was the conclusion we drew from the previous chapter. Having said that, how do we go about investing in equities? Clearly before we dwell further into this topic, it is extremely important to understand the ecosystem in which equities operate. Just like the way we go to the neighborhood kirana store or a super market to shop for our daily needs, similarly we go to the stock market to shop (read as transact) for equity investments. Stock market is where everyone who wants to transact in shares go to. Transact in simple terms means buying and selling. For all practical purposes, you can’t buy/sell shares of a public company like Infosys without transacting through the stock markets. The main purpose of the stock market is to help you facilitate your transactions. So if you are a buyer of a share, the stock market helps you meet the seller and vice versa. Now unlike a super market, the stock market does not exist in a brick and mortar form. It exists in electronic form. You access the market electronically from your computer and go about conducting your transactions (buying and selling of shares).

Also, it is important to note that you can access the stock market via a registered intermediary called the stock broker. We will discuss more about the stock brokers at a later point. There are two main stock exchanges in India that make up the stock markets. They are the Bombay Stock Exchange (BSE) and the National Stock Exchange (NSE). Besides these two exchanges there are a bunch of other regional stock exchanges like Bangalore Stock Exchange, Madras Stock Exchange that are more or less getting phased out and don’t really play any meaningful role anymore.

### Need for stock prediction

Before we address the above question, let us understand what would happen if one choose not to invest. Let us assume you earn Rs.50,000/- per month and you spend Rs.30,000/- towards your cost of living which includes housing, food, transport, shopping, medical etc. The balance of Rs.20,000/- is your monthly surplus. For the sake of simplicity, let us just ignore the effect of personal income tax in this discussion. 1.To drive the point across, let us make few simple assumptions. 2.The employer is kind enough to give you a 10% salary hike every year 3.The cost of living is likely to go up by 8% year on year 4. You are 30 years old and plan to retire at 50. This leaves you with 20 more years to earn 5. You don’t intend to work afer you retire 6. Your expenses are fixed and don’t foresee any other expense 7.The balance cash of Rs.20,000/- per month is retained in the form of hard cash.

Few things are quite startling from the above calculations: 1.Afer 20 years of hard work you have accumulated Rs.1.7 Crs. 2.Since your expenses are fixed, your lifestyle has not changed over the years, you probably even suppressed your lifelong aspirations – better home, better car, vacations etc 3.Afer you retire, assuming the expenses will continue to grow at 8%, Rs.1.7 Crs is good enough to sail you through roughly for about 8 years of post retirement life. 8th year onwards you will be in a very tight spot with literally no savings lef to back you up. What would you do afer you run out of all the money in 8 years time? How do you fund your life? Is there a way to ensure that you collect a larger sum at the end of 20 years? Let’s consider another scenario as per Table 1.2 in the following page where instead of keeping the cash idle, you choose to invest the cash in an investment option that grows at let’s say 12% per annum. For example – in the first year you retained Rs.240,000/- which when invested at 12% per annum for 20 years yields Rs.2,067,063/- at the end of 20th year. With the decision to invest the surplus cash, your cash balance has increased significantly. The cash balance has grown to Rs.4.26 Crs from Rs.1.7 Crs. This is a staggering 2.4x times the regular amount. This translates to you being in a much better situation to deal with your post retirement life. Now, going back to the initial question of why invest? There are few compelling reasons for one to invest.. 1. Fight Inflation – By investing one can deal better with the inevitable – growing cost of living – generally referred to as Inflation 2.Create Wealth – By investing one can aim to have a better corpus by the end of the defined time period. In the above example the time period was upto retirement but it can be anything – children’s education, marriage, house purchase, retirement holidays etc 3.To meet life’s financial aspiration.

### 1.3 Motivation to Work

Businesses primarily run over customer’s satisfaction, customer reviews about their products. Shifts in sentiment on social media have been shown to correlate with shifts in stock markets. Identifying customer grievances thereby resolving them leads to customer satisfaction as well as trustworthiness of an organization. Hence there is a necessity of an un biased automated system to classify customer reviews regarding any problem. In today’s environment where we’re justifiably suffering from data overload (although this does not mean better or deeper insights), companies might have mountains of customer feedback collected; but for mere humans, it’s still impossible to analyse it manually without any sort of error or bias. Oftentimes, companies with the best intentions find themselves in an insights vacuum. You know you need insights to inform your decision making and you know that you’re lacking them, but don’t know how best to get them. Sentiment analysis provides some answers into what the most important issues are, from the perspective of customers, at least. Because sentiment analysis can be automated, decisions can be made based on a significant amount of data rather than plain intuition.

### Problem Statement

Time Series forecasting & modelling plays an important role in data analysis. Time series analysis is a specialized branch of statistics used extensively in fields such as Econometrics & Operation Research. Time Series is being widely used in analytics & data science. Stock prices are volatile in nature and price depends on various factors. The main aim of this project is to predict stock prices using Long short term memory (LSTM).

## History of Stock Exchange

When people talk stocks, they are usually talking about companies listed on major stock exchanges like the [New York Stock Exchange](https://www.investopedia.com/terms/n/nyse.asp) (NYSE) or the Nasdaq. Many of the major American companies are [listed](https://www.investopedia.com/terms/l/listed.asp) on the NYSE, and it can be difficult for investors to imagine a time when the bourse wasn't synonymous with investing and trading stocks. But, of course, it wasn't always this way; there were many steps along the road to our current system of stock exchanges. You may be surprised to learn that the first stock exchange thrived for decades without a single stock being traded.

In this article, we will look at the evolution of [stock exchanges](https://www.investopedia.com/articles/basics/04/092404.asp), from the Venetian states to the British coffeehouses, and finally to [the NYSE and its brethren](https://www.investopedia.com/financial-edge/1212/stock-exchanges-around-the-world.aspx).

### KEY TAKEAWAYS

* While the New York Stock Exchange (NYSE) is arguably the most powerful stock exchange in the world, it was not the first exchange to exert influence on the markets.
* In the 1300s, Venetian moneylenders began to sell debt issues to other lenders and to individual investors.
* In the 1500s, Belgium's exchange dealt exclusively in promissory notes and bonds.
* In the 1600s, the emergence of various East India companies that issued stock led to a financial boom, which was followed by a bust when it was revealed some companies conducted very little actual business.
* Competition from the Nasdaq, which was formed in 1971 to trade securities electronically, has caused the NYSE to evolve and innovate in order to cement its dominance as the premier global exchange.

## The Real Merchants of Venice

The moneylenders of Europe filled important gaps left by the larger banks. Moneylenders traded debts between each other; a lender looking to unload a high-risk, high-interest loan might exchange it for a different loan with another lender. These lenders also bought government [debt issues](https://www.investopedia.com/terms/d/debt-issue.asp). As the [natural evolution](https://www.investopedia.com/articles/07/banking.asp) of their business continued, the lenders began to sell debt issues to the first individual investors. The Venetians were the leaders in the field and the first to start trading securities from other governments.

In the 1300s, Venetian lenders would carry slates with information on the various issues for sale and meet with clients, much like a [broker](https://www.investopedia.com/terms/b/broker.asp) does today.

## The First Stock Exchange—Sans the Stock

According to our [research](https://www.investopedia.com/financial-research-data-hub-5192040), Belgium boasted a stock exchange as far back as 1531 in Antwerp. Brokers and moneylenders would meet there to deal with business, government, and even individual debt issues. It is odd to think of a stock exchange that dealt exclusively in [promissory notes](https://www.investopedia.com/terms/p/promissorynote.asp) and [bonds](https://www.investopedia.com/terms/b/bond.asp), but in the 1500s there were no real stocks. There were many flavors of business-financier partnerships that produced income as stocks do, but there was no official share that changed hands.

## All Those East India Companies

In the 1600s, the Dutch, British, and French governments all gave [charters](https://www.investopedia.com/terms/c/corporatecharter.asp) to companies with East India in their names. On the cusp of imperialism's high point, it seems like everyone had a stake in the profits from the East Indies and Asia except the people living there. Sea voyages that brought back goods from the East were extremely risky—on top of Barbary pirates, there were the more common risks of weather and poor navigation.

To lessen the risk of a lost ship ruining their fortunes, ship owners had long been in the practice of seeking investors who would put up money for the voyage—outfitting the ship and crew in return for a percentage of the proceeds if the voyage was successful. These early [limited liability companies](https://www.investopedia.com/terms/l/llc.asp) often lasted for only a single voyage. They were then dissolved, and a new one was created for the next voyage. [Investors spread their risk](https://www.investopedia.com/investing/importance-diversification/) by investing in several different ventures at the same time, thereby playing the odds against all of them ending in disaster.

When the East India companies were formed, they changed the way business was done. These companies issued stock that would pay [dividends](https://www.investopedia.com/terms/d/dividend.asp) on all the proceeds from all the voyages the companies undertook, rather than going voyage by voyage. These were the first modern [joint-stock companies](https://www.investopedia.com/terms/j/jointstockcompany.asp). This allowed the companies to demand more for their shares and build larger fleets. The size of the companies, combined with royal charters forbidding competition, meant huge profits for investors.

## A Little Stock With Your Coffee?

Because the shares in the various East India companies were issued on paper, investors could sell the papers to other investors. Unfortunately, there was no stock exchange in existence, so the investor would have to track down a broker to carry out a trade. In England, most brokers and investors did their business in the various coffee shops around London. Debt issues and shares for sale were written up and posted on the shops' doors or mailed as a newsletter.

## The South Seas Bubble Bursts

The British East India Company had one of the biggest [competitive advantages](https://www.investopedia.com/terms/c/competitive_advantage.asp) in financial history—a [government-backed monopoly](https://www.investopedia.com/terms/l/legalmonopoly.asp). When the investors began to receive huge dividends and sell their shares for fortunes, other investors were hungry for a piece of the action.

The budding financial boom in England came so quickly that there were no rules or regulations for the [issuing of shares](https://www.investopedia.com/terms/i/issue.asp). The South Seas Company (SSC) emerged with a similar charter from the king and its shares, and the numerous re-issues, sold as soon as they were listed. Before the first ship ever left the harbor, the SSC had used its newfound investor fortune to open plush offices in the best parts of London.

Encouraged by the success of the SSC—and realizing that the company hadn't done a thing except for issue shares—other "businessmen" rushed in to offer new shares in their own ventures. Some of these were as ludicrous as reclaiming the sunshine from vegetables or, better yet, a company promising investors shares in an undertaking of such vast importance that they couldn't be revealed. They all sold. Before we pat ourselves on the back for how far we've come, remember that these [blind pools](https://www.investopedia.com/terms/b/blind_pool.asp) still exist today.

Inevitably, the bubble burst when the SSC failed to pay any dividends on its meager profits, highlighting the difference between these new share issues and the British East India Company. The subsequent [crash](https://www.investopedia.com/terms/c/crash.asp) caused the government to outlaw the issuing of shares—the ban held until 1825.

## The New York Stock Exchange

The first stock exchange in London was officially formed in 1773, a scant 19 years before the New York Stock Exchange. Whereas the [London Stock Exchange](https://www.investopedia.com/terms/l/lse.asp) (LSE) was handcuffed by the law restricting shares, the New York Stock Exchange has dealt in the trading of stocks, for better or worse, since its inception. The NYSE wasn't the first stock exchange in the U.S., however. That honor goes to the [Philadelphia Stock Exchange](https://www.investopedia.com/terms/p/phlx.asp), but the NYSE quickly became the most powerful.

Formed by brokers under the spreading boughs of a [buttonwood tree](https://www.investopedia.com/terms/b/buttonwoodagreement.asp), the New York Stock Exchange made its home on [Wall Street](https://www.investopedia.com/terms/w/wallstreet.asp). The exchange's location, more than anything else, led to the dominance that the NYSE quickly attained. It was in the heart of all the business and trade coming to and going from the United States, as well as the domestic base for most banks and large corporations. By setting [listing requirements](https://www.investopedia.com/terms/l/listingrequirements.asp) and demanding fees, the New York Stock Exchange became a very wealthy institution.

The NYSE faced very little serious domestic competition for the next two centuries. Its international prestige rose in tandem with the burgeoning American economy, and it was soon the most important stock exchange in the world. The NYSE had its share of ups and downs during the same period, too. Everything from the [Great Depression](https://www.investopedia.com/terms/g/great_depression.asp) to the Wall Street bombing of 1920 left scars on the exchange. The 1920 bombing, believed to have been carried out by anarchists, left 38 dead and also literally scarred many of Wall Street's prominent buildings. The less literal scars on the exchange came in the form of stricter listing and reporting requirements.

### NYSE Competitors

On the international scene, London emerged as the major exchange for Europe, but many companies that were able to list internationally still listed in New York. Many other countries including Germany, France, the Netherlands, Switzerland, South Africa, Hong Kong, Japan, Australia, and Canada developed their own stock exchanges, but these were largely seen as proving grounds for domestic companies to inhabit until they were ready to make the leap to the LSE and from there to the big leagues of the NYSE. Some of these international exchanges are still seen as a dangerous territory because of weak listing rules and less rigid [government regulation](https://www.investopedia.com/articles/economics/11/government-regulations.asp).

Despite the existence of stock exchanges in Chicago, Los Angeles, Philadelphia, and other major centers, the NYSE was the most powerful stock exchange domestically and internationally. In 1971, however, an upstart emerged to challenge the NYSE hegemony.

## The New Kid on the Block

The [Nasdaq](https://www.investopedia.com/terms/n/nasdaq.asp) was the brainchild of the National Association of Securities Dealers (NASD)—now called the [Financial Industry Regulatory Authority](https://www.investopedia.com/terms/f/finra.asp) (FINRA). From its inception, it has been a different type of stock exchange. It does not inhabit a physical space, as with 11 Wall Street. Instead, it is a network of computers that executes trades electronically.

The introduction of an electronic exchange made trades more efficient and reduced the [bid-ask spread](https://www.investopedia.com/terms/b/bid-askspread.asp)—a spread the NYSE wasn't above profiting from. The competition from Nasdaq has forced the NYSE to evolve, both by listing itself and by merging with [Euronext](https://www.investopedia.com/terms/e/euronext.asp) to form the first trans-Atlantic exchange, which it maintained until 2014 when Euronext was spun off to become an independent entity.

## The Future: World Parity?

The NYSE is still the largest and, arguably, the most powerful stock exchange in the world. The Nasdaq has more companies listed, but the NYSE has a [market capitalization](https://www.investopedia.com/terms/m/marketcapitalization.asp) that is larger than Tokyo, London, and the Nasdaq exchanges combined. The NYSE, once closely tied to the fortunes or failures of the American economy, is now global. Although the other stock exchanges in the world have grown stronger through mergers and the development of their domestic economies, it is difficult to see how any of them will dislodge the 800-pound gorilla that is the New York Stock Exchange.

1. **Aim**

To make the stock market investment process simple, less time consuming and less tedious, and make process of investing in stock market This system will help news investors understand the stock market very quickly and easily. To make profitable trade in the stock market, investors need to predict the trend of stock price, future range of stock price, which requires daily updating of market movements, and keeping eye on daily market data. This whole process makes the small investors or people with constraints due to their routine struggle in the stock market keeping a regular update of such information becomes difficult for them.

1. **Objectives** 

Increase the accuracy of price prediction  Make stock market prediction easy and simple  Provide useful insights to new investors to understand market quickly.  Reduce the time required to make prediction by providing different data analysis at one point.  To make the stock market investment process simple.

1. **Scope**

Predicting stock price range, volatility, risk, reward, trend of stocks, comparison of stock with other peer stocks, comparison with global market will be achieved using stock’s open, high, low, close, volume data, price of other commodities related to stock, forex(currency) price. Sentiment analysis of news is performed to find overall sentiment of stock market investors. Combining all these data and different algorithm’s result to obtain better accuracy of prediction.

# Chapter 2: Review of Literature

### 2.1 Introduction

"What other people think” has always been an important piece of information for most of us during the decision-making process. The Internet and the Web have now (among other things) made it possible to find out about the opinions and experiences of those in the vast pool of people that are neither our personal acquaintances nor well-known professional critics — that is, people we have never heard of. And conversely, more and more people are making their opinions available to strangers via the Internet. The interest that individual users show in online opinions about products and services, and the potential influence such opinions wield, is something that is driving force for this area of interest. And there are many challenges involved in this process which needs to be walked all over in order to attain proper outcomes out of them. In this survey we analysed basic methodology that usually happens in this process and measures that are to be taken to overcome the challenges being faced.

### 2.2 Existing Methods

##### 2.2.1 Stock Market Prediction Using Machine Learning

The research work done by V Kranthi Sai Reddy Student, ECM, Sreenidhi Institute of Science and Technology, Hyderabad, India. In the finance world stock trading is one of the most important activities. Stock market prediction is an act of trying to determine the future value of a stock other financial instrument traded on a financial exchange. This paper explains the prediction of a stock using Machine Learning. The technical and fundamental or the time series analysis is used by the most of the stockbrokers while making the stock predictions. The programming language is used to predict the stock market using machine learning is Python. In this paper we propose a Machine Learning (ML) approach that will be trained from the available stocks data and gain intelligence and then uses the acquired knowledge for an accurate prediction. In this context this study uses a machine learning technique called Support Vector Machine (SVM) to predict stock prices for the large and small capitalizations and in the three different markets, employing prices with both daily and up-to-the-minute frequencies.

##### 2.2.2 Forecasting the Stock Market Index Using Artificial Intelligence Techniques

The research work done by Lufuno Ronald Marwala A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering. The weak form of Efficient Market hypothesis (EMH) states that it is impossible to forecast the future price of an asset based on the information contained in the historical prices of an asset. This means that the market behaves as a random walk and as a result makes forecasting impossible. Furthermore, financial forecasting is a difficult task due to the intrinsic complexity of the financial system. The objective of this work was to use artificial intelligence (AI) techniques to model and predict the future price of a stock market index. Three artificial intelligence techniques, namely, neural networks (NN), support vector machines and neuro-fuzzy systems are implemented in forecasting the future price of a stock market index based on its historical price information. Artificial intelligence techniques have the ability to take into consideration financial system complexities and they are used as financial time series forecasting tools. Two techniques are used to benchmark the AI techniques, namely, Autoregressive Moving Average (ARMA) which is linear modelling technique and random walk (RW) technique. The experimentation was performed on data obtained from the Johannesburg Stock Exchange. The data used was a series of past closing prices of the All Share Index. The results showed that the three techniques have the ability to predict the future price of the Index with an acceptable accuracy. All three artificial intelligence techniques outperformed the linear model. However, the random walk method out performed all the other techniques. These techniques show an ability to predict the future price however, because of the transaction costs of trading in the market, it is not possible to show that the three techniques can disprove the weak form of market efficiency. The results show that the ranking of performances support vector machines, neuro-fuzzy systems, multilayer perceptron neural networks is dependent on the accuracy measure used.

#### 2.2.3 Indian stock market prediction using artificial neural networks on tick data

The research work done by Dharmaraja Selvamuthu, Vineet Kumar and Abhishek Mishra Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India. A stock market is a platform for trading of a company’s stocks and derivatives at an agreed price. Supply and demand of shares drive the stock market. In any country stock market is one of the most emerging sectors. Nowadays, many people are indirectly or directly related to this sector. Therefore, it becomes essential to know about market trends. Thus, with the development of the stock market, people are interested in forecasting stock price. But, due to dynamic nature and liable to quick changes in stock price, prediction of the stock price becomes a challenging task. Stock m Prior work has proposed effective methods to learn event representations that can capture syntactic and semantic information over text corpus, demonstrating their effectiveness for downstream tasks such as script event prediction. On the other hand, events extracted from raw texts lacks of common-sense knowledge, such as the intents and emotions of the event participants, which are useful for distinguishing event pairs when there are only subtle differences in their surface realizations. To address this issue, this paper proposes to leverage external common-sense knowledge about the intent and sentiment of the event. Experiments on three event-related tasks, i.e., event similarity, script event prediction and stock market prediction, show that our model obtains much better event embeddings for the tasks, achieving 78% improvements on hard similarity task, yielding more precise inferences on subsequent events under given contexts, and better accuracies in predicting the volatilities of the stock market1. Markets are mostly a nonparametric, non-linear, noisy and deterministic chaotic system (Ahangar et al. 2010). As the technology is increasing, stock traders are moving towards to use Intelligent Trading Systems rather than fundamental analysis for predicting prices of stocks, which helps them to take immediate investment decisions. One of the main aims of a trader is to predict the stock price such that he can sell it before its value decline, or buy the stock before the price rises. The efficient market hypothesis states that it is not possible to predict stock prices and that stock behaves in the random walk. It seems to be very difficult to replace the professionalism of an experienced trader for predicting the stock price. But because of the availability of a remarkable amount of data and technological advancements we can now formulate an appropriate algorithm for prediction whose results can increase the profits for traders or investment firms. Thus, the accuracy of an algorithm is directly proportional to gains made by using the algorithm.

#### 2.2.4 The Stock Market and Investment

The research work done by Manh Ha Duong Boriss Siliverstovs. Investigating the relation between equity prices and aggregate investment in major European countries including France, Germany, Italy, the Netherlands and the United Kingdom. Increasing integration of European financial markets is likely to result in even stronger correlation between equity prices in different European countries. This process can also lead to convergence in economic development across European countries if developments in stock markets influence real economic components, such as investment and consumption. Indeed, our vector autoregressive models suggest that the positive correlation between changes equity prices and investment is, in general, significant. Hence,monetary authorities should monitor reactions of share prices to monetary policy and their effects on the business cycle.

#### 2.2.5 Automated Stock Price Prediction Using Machine Learning

The research work done by Mariam Moukalled Wassim El-Hajj Mohamad Jaber Computer Science Department American University of Beirut. Traditionally and in order to predict market movement, investors used to analyse the stock prices and stock indicators in addition to the news related to these stocks. Hence, the importance of news on the stock price movement. Most of the previous work in this industry focused on either classifying the released market news as (positive, negative, neutral) and demonstrating their effect on the stock price or focused on the historical price movement and predicted their future movement. In this work, we propose an automated trading system that integrates mathematical functions, machine learning, and other external factors such as news’ sentiments for the purpose of achieving better stock prediction accuracy and issuing profitable trades. Particularly, we aim to determine the price or the trend of a certain stock for the coming end-of-day considering the first several trading hours of the day. To achieve this goal, we trained traditional machine learning algorithms and created/trained multiple deep learning models taking into consideration the importance of the relevant news. Various experiments were conducted, the highest accuracy (82.91%) of which was achieved using SVM for Apple Inc. (AAPL) stock.

#### 2.2.6 Stock Price Correlation Coefficient Prediction with ARIMALSTM Hybrid Model

The research work done by Hyeong Kyu Choi, B.A Student Dept. of Business Administration Korea University Seoul, Korea. Predicting the price correlation of two assets for future time periods is important in portfolio optimization. We apply LSTM recurrent neural networks (RNN) in predicting the stock price correlation coefficient of two individual stocks. RNN’s are competent in understanding temporal dependencies. The use of LSTM cells further enhances its long-term predictive properties. To encompass both linearity and nonlinearity in the model, we adopt the ARIMA model as well. The ARIMA model filters linear tendencies in the data and passes on the residual value to the LSTM model. The ARIMA-LSTM hybrid model is tested against other traditional predictive financial models such as the full historical model, constant correlation model, single-index model and the multi-group model. In our empirical study, the predictive ability of the ARIMA-LSTM model turned out superior to all other financial models by a significant scale. Our work implies that it is worth considering the ARIMALSTM model to forecast correlation coefficient for portfolio optimization.

#### 2.2.7 Event Representation Learning Enhanced with External Common-sense Knowledge

The research work done by Xiao Ding, Kuo Liao, Ting Liu, Zhongyang Li, Junwen Duan Research Center for Social Computing and Information Retrieval Harbin Institute of Technology, China. Prior work has proposed effective methods to learn event representations that can capture syntactic and semantic information over text corpus, demonstrating their effectiveness for downstream tasks such as script event prediction. On the other hand, events extracted from raw texts lacks of common-sense knowledge, such as the intents and emotions of the event participants, which are useful for distinguishing event pairs when there are only subtle differences in their surface realizations. To address this issue, this paper proposes to leverage external common-sense knowledge about the intent and sentiment of the event. Experiments on three event-related tasks, i.e., event similarity, script event prediction and stock market prediction, show that our model obtains much better event embeddings for the tasks, achieving 78% improvements on hard similarity task, yielding more precise inferences on subsequent events under given contexts, and better accuracies in predicting the volatilities of the stock market.

#### 2.2.8 Forecasting directional movements of stock prices for intraday trading using LSTM and random forests

The research work done by Pushpendu Ghosh, Ariel Neufeld, Jajati Keshari SahooDepartment of Computer Science & Information Systems, BITS Pilani K.K.Birla Goa campus, India bDivision of Mathematical Sciences, Nanyang Technological University, Singapore cDepartment of Mathematics, BITS Pilani K.K.Birla Goa campus, India. We employ both random forests and LSTM networks (more precisely CuDNNLSTM) as training methodologies to analyse their effectiveness in forecasting outof-sample directional movements of constituent stocks of the S&P 500 from January 1993 till December 2018 for intraday trading. We introduce a multi-feature setting consisting not only of the returns with respect to the closing prices, but also with respect to the opening prices and intraday returns. As trading strategy, we use Krauss et al. (2017) and Fischer & Krauss (2018) as benchmark and, on each trading day, buy the 10 stocks with the highest probability and sell short the 10 stocks with the lowest probability to outperform the market in terms of intraday returns – all with equal monetary weight. Our empirical results show that the multi-feature setting provides a daily return, prior to transaction costs, of 0.64% using LSTM networks, and 0.54% using random forests. Hence, we outperform the singlefeature setting in Fischer & Krauss (2018) and Krauss et al. (2017) consisting only of the daily returns with respect to the closing prices, having corresponding daily returns of 0 .41% and of 0 .39% with respect to LSTM and random forests, respectively. 1 Keywords: Random forest, LSTM, Forecasting, Statistical Arbitrage, Machine learning, Intraday trading.

**2.2.9 A Deep Reinforcement Learning Library for Automated Stock Trading in Quantitative Finance**

The research work done by Xiao-Yang Liu1 Hongyang Yang,Qian Chen4,Runjia ZhangLiuqing Yang Bowen Xiao Christina Dan Wang Electrical Engineering, 2Department of Statistics, 3Computer Science, Columbia University, 3AI4Finance LLC., USA, Ion Media Networks, USA, Department of Computing, Imperial College, 6New York University (Shanghai). As deep reinforcement learning (DRL) has been recognized as an effective approach in quantitative finance, getting hands-on experiences is attractive to beginners. However, to train a practical DRL trading agent that decides where to trade, at what price, and what quantity involves error-prone and arduous development and debugging. In this paper, we introduce a DRL library FinRL that facilitates beginners to expose themselves to quantitative finance and to develop their own stock trading strategies. Along with easily-reproducible tutorials, FinRL library allows users to streamline their own developments and to compare with existing schemes easily. Within FinRL, virtual environments are configured with stock market datasets, trading agents are trained with neural networks, and extensive back testing is analysed via trading performance. Moreover, it incorporates important trading constraints such as transaction cost, market liquidity and the investor’s degree of risk-aversion. FinRL is featured with completeness, hands-on tutorial and reproducibility that favors beginners: (i) at multiple levels of time granularity, FinRL simulates trading environments across various stock markets, including NASDAQ-100, DJIA, S&P 500, HSI, SSE 50, and CSI 300; (ii) organized in a layered architecture with modular structure, FinRL provides finetuned state-of-the-art DRL algorithms (DQN, DDPG, PPO, SAC, A2C, TD3, etc.), commonly used reward functions and standard evaluation baselines to alleviate the debugging workloads and promote the reproducibility, and (iii) being highly extendable, FinRL reserves a complete set of user-import interfaces. Furthermore, we incorporated three application demonstrations, namely single stock trading, multiple stock trading, and portfolio allocation. The FinRL library will be available on GitHub at link https://github.com/AI4Finance-LLC/FinRL-Library.

**2.2.10 An innovative neural network approach for stock market prediction**

The research work done by Xiongwen Pang, Yanqiang Zhou, Pan Wang, Weiwei Lin. To develop an innovative neural network approach to achieve better stock market predictions. Data were obtained from the live stock market for real-time and off-line analysis and results of visualizations and analytics to demonstrate Internet of Multimedia of Things for stock analysis. To study the influence of market characteristics on stock prices, traditional neural network algorithms may incorrectly predict the stock market, since the initial weight of the random selection problem can be easily prone to incorrect predictions.

Based on the development of word vector in deep learning, we demonstrate the concept of “stock vector.” The input is no longer a single index or single stock index, but multi-stock high-dimensional historical data. We propose the deep long short-term memory neural network (LSTM) with embedded layer and the long short-term memory neural network with automatic encoder to predict the stock market. In these two models, we use the embedded layer and the automatic encoder, respectively, to vectorize the data, in a bid to forecast the stock via long short-term memory neural network. The experimental results show that the deep LSTM with embedded layer is better. Specifically, the accuracy of two models is 57.2 and 56.9%, respectively, for the Shanghai A-shares composite index. Furthermore, they are 52.4 and 52.5%, respectively, for individual stocks. We demonstrate research contributions in IMMT for neural network-based financial analysis.

#### 2.2.11 An Intelligent Technique for Stock Market Prediction

The research work done by M. Mekayel Anik · M. Shamsul Arefin (B) Department of Computer Science and Engineering, Chittagong University of Engineering and Technology, Chittagong, Bangladesh. A stock market is a loose network of economic transactions between buyers and sellers based on stocks also known as shares. In stock markets, stocks represent the ownership claims on businesses. These may include securities listed on a stock exchange as well as those only traded privately. A stock exchange is a place where brokers can buy and/or sell stocks, bonds, and other securities. Stock market is a very vulnerable place for investment due to its volatile nature. In the near past, we faced huge financial problems due to huge drop in price of shares in stock markets worldwide. This phenomenon brought a heavy toll on the international as well as on our national financial structure. Many people lost their last savings of money on the stock market. In 2010–2011 financial year, Bangladeshi stock market faced massive collapse [1]. This phenomenon can be brought under control especially by strict monitoring and instance stock market analysis. If we can analyse stock market correctly in time, it can become a field of large profit and may become comparatively less vulnerable for the investors. Stock market is all about prediction and rapid decision making about investment, which cannot be done without thorough analysis of the market. If we can predict the stock market by analysing historical data properly, we can avoid the consequences of serious market collapse and to be able to take necessary steps to make market immune to such situations.

## Related Work

The paper “Machine Learning Approach in Stock Market Prediction” was written by authors Raut Sushrut Deepak, Shinde Isha Uday, Dr D. Malathi. The paper states that a high level of accuracy and precision are the key parameters to be considered while predicting the share prices. Time series analysis, fundamental analysis of companies, Technical analysis are used by many share market individual investors and institutions during the prediction. However, these methods are not completely reliable, so there is a need to provide supporting way to predict the stock market. In this paper, a Machine Learning (ML) method that will be trained in available stock data, acquire intelligence and use the information obtained to accurately predict. After extensive research of various algorithms and their robustness in the various problem domains, the Artificial Neural Network (ANN) was more appropriate than any other algorithm. The Customized neural network model which has wide number of features and parameters makes it easier to apply. An important method used in this paper is to obtain the results of predicting machine learning concepts and the results tested on the Bombay Stock Exchange index data set [7]. The paper "Stock Price Prediction Using Technology Analysis and Mechanical Learning" was written by author Jan Ivar Larsen. The paper states that historical stock prices are used to predict future stock price movements. The used stock price model uses a two-layer consultation approach that uses background information from technical analysis to the first layer of thinking to direct the second layer of thinking based on machine learning. The model is supplemented by a financial management strategy that uses the historical success of the predictions made by the model to determine the amount of money that will invest in future estimates. Based on several portfolio simulations and trading models generated by the model, they conclude that the forecast model successfully exceeds the Oslo Benchmark Index (OSEBX) [9].

The paper "A Machine Learning Model for Stock Market Prediction" was written by authors Omar S. Soliman, Osman Hegazy and Mustafa Abdul Salam. The paper states that the Share market price prediction is an attempt of estimating the future values of a stock prices and other financial instruments that are traded in different exchanges. Successful forecasting of future share prices can increase individual investor’s profits. This paper proposes a machine learning model for this assessment of the share market price. The proposed algorithm combines particle swarm optimization and random forest algorithm. The Particle swarm optimization algorithm is used to grow a random forest to estimate prices everyday. This given model works on the study of past dataset of share market and technical indicators. The Particle swarm optimization algorithm selects a combination of the most parameters for random forest integration which avoids further congestion and problem of local minima which can improve prediction accuracy. These presented model tested on multiple financial sets of databases and was compared with the Leuralberg-Marquardt neural network algorithms. The results attained indicated the model proposed here can have better accuracy of prediction and robust Particle swarm optimization algorithm [10]. The paper "Prediction Models for Indian Stock Market" was written by authors Aparna Nayak, M. M. Manohara Pai and Radhika M. Pai. The paper states that share market price data is produced in large amount and changes on every moment. The share market is a perplex and demanding system where investors can earn more money or lose all money. In this, an attempt of predicting share price trend is done. Two types are made up of one for predicting next day price prediction and one is for prediction of prices after one month. The learning algorithms of the monitored machines are used to build the models. As part of the daily predictive model, historical values are associated with emotions. Up to 70% of perceived accuracy using standardized machine learning techniques in daily forecast models. The monthly forecasting model attempts to assess whether there are any similarities between any two-month course. Tests prove that the trend for at least one month is related to the trend of other month. [11]

The paper "Study on the prediction of stock price based on associated network model of LSTM" was written by authorsGuangyu Ding and Liangxi Qin.This paper says The share market has attracted a wide range of individuals and institutions. It has always been a great junction for individuals and Investment firms, constantly adopting changes in the stock market and its trend. Right now there multiple ways to estimate the future share prices. future share price prediction methods can be divided into at least two categories: Mathematical Methods and techniques of AI. Mathematical methods include the systematic model, the ARCH model, and so on. Among the methods of AI are the MLP, the convolutional Neural Network, the Bayes Network, Backward Propagation Network, SVM, Single-Layer Long Short-Term Memory, etc. In order to estimate multiple values as output in a model, it is necessary to design a model that can various values as output parallelly and take various inputs for this purpose, an deep neural network model is proposed based on a short-term memory network with multiple inputs and outputs. This network can simultaneously estimate price, open, high and low price of shares. Convenient network model Compared to the Long Short-Term Memory model and the Deep Neural Network model. Results show that accuracy in estimating multiple values parallelly the associated model has greater accuracy than the other models and Its estimated accuracy is more than 95%. [12] The paper "Stock Market Forecasts Using Machine Learning" was written by authors Prof. S.P. Pimpalkar, Jenish Karia, Muskaan Khan, Satyamandand, Tushar Mukherjee. The paper states that the construction of the guesswork will use a variety of attributes such as input and will predict whether the market value will be positive or negative. The various attributes used in the model include oil, Foreign Exchange Rate, interest, gold and silver prices, NEWS, twitter news feeds and pattern matching. A variety of ML methods are used including Regression Support Vector Machine and Recurrent Neural Network [5].

The paper "Automated Stock Price Prediction Using Machine Learning" was written by authors Mariam Moukalled, Wassim El-Hajj, Mohamad Jaber. The paper says that traditionally, investors analyze share values and stock indices to predict market movements along with event news related to these shares. Therefore the importance of news in share market. The majority of the models created earlier focuses on categorizing news as good, bad, neutral and finding their impact on share prices or considering past price changes and estimating their upcoming price changes In this, they presented an automated trading system that combines statistical method, Artificial Intelligence and other visible factors such as media sentiment with the aim of gaining better stock forecasting and profitable trading. In particular, we aim to determine the valuation or trend of a particular share price for the future date by considering the initial market time. To obtain this objective, they trained conventional machine learning models and developed / trained multiple in-depth reading models taking into account the importance of relevant issues. Various tests were performed, with the highest accuracy (82.91%) obtained using Apple Inc.'s Support Vector Machine stock. (AAPL) stock [14]. The paper "Stock Market Forecasts Using Machine Learning" was written by author V.V Kranthi Sai Reddy. The paper states that stock trading globally is one of the most important. Share market prediction is an attempt to estimate the future prices of other financial instrument traded on a currency exchange. This paper describes stock forecasts using Learning Machines. Technical and basic or time analysis is used by most stockbrokers when making stock forecasts. The language of the system used to predict the stock market using Python learning tools. In this paper we propose a Machine Learning (ML) method that will be trained in available stock data and gain intelligence and use the information obtained to accurately predict. In this context the study uses a machine learning system called the Support Vector Machine to predict the prices of major and minor stocks and three different markets, using daily and minute wave prices [15].

## Existing Systems

In the present system the various algorithms used for forecasting can be divided into queues (AR, MA, ARIMA, ARMA) and incompatible models (ARCH, GARCH, Neural Network) and AI framework such as Naive Bayes, the closest neighbours k (k-NN), Support Vector Machine (SVM), Linear Regression, Artificial Neural Network (ANN) and Random Forest were used to advance the gauge model. Points are settled on stock values []. Current models predict that the stock market uses only one algorithm to predict different conditions and variables and also does not combine multiple algorithm results or consider multiple algorithms to accurately predict. Previous results indicate that the stock price cannot be calculated using traditional divisions. The current system does perform optimally if there is a change in the operating surrounding because it does not focus on external events occurring in the surrounding such as news events and other factors affecting prices such as the Forex and Commodity market. It uses only one data source, thus being extremely biased. The existing system requires some kind of input translation, so it needs to be measured. The existing system uses only historical data or media analysis simultaneously, not used together The current system does not take into account certain important data such as trade volume and transaction value in the trading volume and the percentage of the amount that are be delivered and the percentage of delivery that predicts an investment or investment that occurs in a particular stock by a major fund manager or large investors. This feature is often overlooked by new investors and existing algorithms, that does not take these factors into account for better market movement analysis. The combination of these factors and volume data can be an important predictor parameter that is not considered much in existing systems. Most of the existing systems uses only one algorithm and one data at a time. The existing system also does try to predict share prices in all conditions and on all days but in real world share market can not be predictable every time so certain conditions need to be checked before predicting sharemarket.

**Machine Learning Approach**

Because of global digitization, SMP has entered a technological era. Machine learning in stock price prediction is used to discover patterns in data [22]. Usually, a tremendous amount of structured and unstructured heterogeneous data is generated from stock markets. Using machine learning algorithms, it is possible to quickly analyze more complex heterogeneous data and generate more accurate results. Various machine learning methods have been used for SMP [23]. The machine learning approaches are mainly categorized into supervised and unsupervised approaches. In the supervised learning approach, named input data and the desired output are given to the learning algorithms. Meanwhile, in the unsupervised learning approach, unlabeled input data is provided to the learning algorithm, and the algorithm identifies the patterns and generates the output accordingly. Furthermore, different algorithmic approaches have been used in SMP, such as the Support Vector Machine (SVM), k Nearest Neighbors (kNN), Artificial Neural Networks (ANN), Decision Trees, Fuzzy Time-Series, and Evolutionary Algorithms. The SVM is a supervised machine learning technique that limits error and augments geometric margins, and is a pattern classification algorithm [24]. In terms of accuracy, the SVM is an important machine learning algorithm compared to the other classifiers [25]. In the kNN, stock prediction is mapped into a classification based on closeness. Using Euclidean distance, the kNN classifies the “k” nearest neighbors in the training set. The ANN is a nonlinear computational structure for various machine learning algorithms to analyze and process complex input data together. The FIS (Fuzzy Inference Systems) apply rules to fuzzy sets and then apply de-fuzzification to give crisp outputs for decision making [26]. The evolutionary algorithms include gene-inspired neuro-fuzzy and neuro-genetic algorithms, mimic the natural selection theory of species, and can give an optimal output.

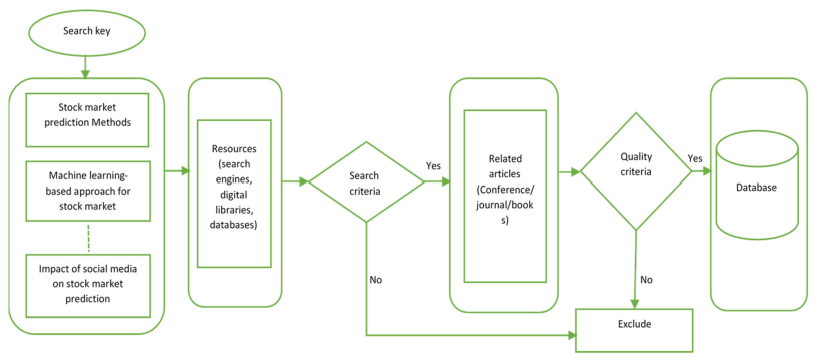
**Sentiment Analysis Approach**

One of the phenomena of current times that is changing the world is the global availability of the internet. The most-used platforms on the internet are social media. It is estimated that social media users all over the world will number around 3.07 billion [27]. There is a high association between stock prices and events related to stocks on the web. The event information is extracted from the internet to predict stock prices; such an approach is known as event-driven stock prediction [28]. Through social networks, people generate tremendous amounts of data that is filled with emotions. Much of this data is related to user perceptions and concerns [29]. Sentiment analysis is a field of study that deals with the people’s concerns, beliefs, emotions, perceptions, and sentiments towards some entity [30,31]. It is the process of analyzing text corpora, e.g., news feeds or stock marketspecific tweets, for stock trend prediction. The Stock Twits, Twitter, Yahoo Finance, and so on are well-known platforms used for the extraction of sentiments. There is a significant importance of using sentimental data for enhancing the prediction of volatility in the stock market. The ‘Wisdom of Crowds’ and sentiment analysis generate more insights that can be used to increase the performance in various fields, such as box office sales, election outcomes, SMP, and so on [32]. This suggests that a good decision can be made by taking the opinions and insights of large groups of people with varied types of information [33]. The information generated through social media allows us to explore vast and diverse opinions. Exploring sentiments from social media in addition to numeric time-series stock data would enhance the accuracy of the prediction. Using time-series data as well as social media data would intensify the prediction accuracy. Different approaches and techniques have been proposed over time to anticipate stock prices through numerous methodologies, thanks to the dynamic and challenging panorama of stock markets [34].

# Chapter 3: Methodology

**3.1 PROPOSED SYSTEMS**

The prediction methods can be roughly divided into two categories, statistical methods and artificial intelligence methods. Statistical methods include logistic regression model, ARCH model, etc. Artificial intelligence methods include multi-layer perceptron, convolutional neural network, naive Bayes network, back propagation network, single-layer LSTM, support vector machine, recurrent neural network, etc. They used Long short-term memory network (LSTM).



The system is divided into the following modules:

• Data Collection Data is collected from various sources, such as yahoo finance and google finance, which are in the form of .csv format. For news data gathering system uses Google News api.

• Analysis Manipulation and Visualization of data Data collected is cleaned and pre-processed to make it ready to use for mal algorithms and models. and along with this data visualization is also performed.

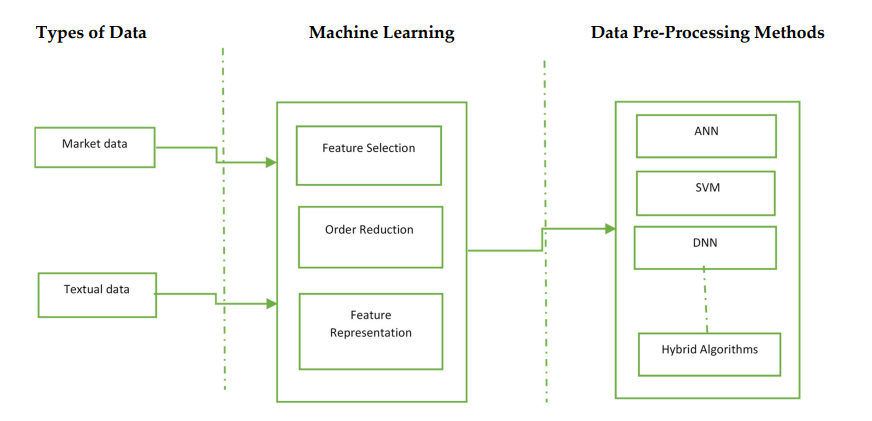
• Build a Model The cleaned and pre-processed data is used to create, build and train various machine learning algorithms which can be used in predictions.

• Predict Outcomes After the model has been built successfully, the next thing to do is predict an outcome pattern for a particular stock and check the accuracy of the predictions

• Predict combining results of all algorithms After models are built system combines their results for better prediction and high accuracy and predicts the output based on real time data provided to it.

This section explains the overall process of the literature collection on SMP using machine learning. Initially, the phrase “stock market prediction using machine learning” was keyed to various search engines, digital libraries and databases, including ‘google scholar’, ‘research gate’, ‘ACM digital library’, ‘IEEE Explore’, ‘Scopus’, and so on. During the process of literature collection, various phrases like “stock market prediction methods”, “impact of sentiments on stock market prediction”, and “machine learning-based approach for stock market prediction” were keyed. The OR and AND operators were used for the keyword searches in single and multiple classes, respectively. As a result, some of the fundamental papers in the field of stock market prediction were retrieved. By the careful analysis of a few basic papers, a primary insight into the domain was obtained. The search criteria were further modified to collect the literature of the last decade, in order to enhance and improve the domain. In addition, the literature selected was screened by applying quality criteria, where metrics such as indexing, quartiles, impact factors and publishers wer

We aim to predict the future share prices using various machine learning methods. In this proposed system, we were able to train and test the the algorithms of ML from the different data points from the past for making the future share price estimation. We considered EOD (End OF the Day) data of the past years to test and train the ML models [15]. We used various ML libraries and frameworks to obtain the goal system uses Numpy, Pandas, for data manipulation and visualization, Sklearn, Kears, Tensorflow, for machine learning models and math for mathematical operations on data. The other dataset was Historical news headlines used for sentiment analysis. The main library was NumPy library, which is commonly used to manipulate and clean and datasets, to convert it into form which can be directly used for ML models [6]. The other was sklearn library, which is used for actual calculation, estimation and prediction. The data we used was of historical share markets, and was gathered from the various public and open sources available online, and 85 or more percentage of dataset was used to training of the ML models depending upon the need and the rest was used for validation and testing purpose [4]. The common way of the supervised ML model is to analyse and remember the pattern and correlations in the dataset from the train and validation dataset and then produce result for the test dataset [13]. System does feature extraction, various pre-processing of data to make it ready for training of model [13]. The python library named pandas was used for pre-processing data such as it was used to combine various datasets into one dataframe. This prepared data frame data was ready for feature extraction. The features of these dataframe was date, close, open, high, low, volume, delivery percentage, number of trades, turnover and other derived features. All features was used by us to train machine of a random forest model and predicted the object variable, and we used some selected features to predict using the LSTM, SVM algorithm also which was the values of share of upcoming day [14]. We also measured the accuracy of prediction on the test dataset and on real values. The presented method includes different aspects of research including data manipulation and pre-processing. System uses ANN (Artificial Neural Network) for sentiment analysis of news data, and Google News API for fetching real time data of news related to stock market.



**Types of Data**

SMP systems can be classified according to the type of data they use as the input. Most of the studies used market data for their analysis. Recent studies have considered textual data from online sources as well. In this section, the studies are classified based on the type of data they use for prediction purposes. At the end of this section, Table 1 points out the comparison of the data sources, type of input and prediction duration used in the studies so far. 4.1. Market Data Market data are the temporal historical price-related numerical data of financial markets. Analysts and traders use the data to analyze the historical trend and the latest stock prices in the market. They reflect the information needed for the understanding of market behavior. The market data are usually free, and can be directly downloaded from the market websites. Various researchers have used this data for the prediction of price movements using machine learning algorithms. The previous studies have focused on two types of predictions. Some studies have used stock index predictions like the Dow Jones Industrial Average (DJIA) [35], Nifty [36], Standard and Poor’s (S&P) 500 [37], National Association of Securities Dealers Automated Quotations (NASDAQ) [38], the Deutscher Aktien Index (DAX) index [39], and multiple indices [40,41]. Other studies have used individual stock prediction based on some specific companies like Apple, Google or other group of companies.

Furthermore, the studies focused on time-specific predictions like intraday [45], daily [20], weekly [46], and monthly predictions [47], and so on. Moreover, most of the previous research is based on categorical prediction, where predictions are categorized into discrete classes like up, down, positive, or negative. Technical indicators have been widely used for SMP due to their summative representation of trends in time series data. Some studies considered different types of technical indicators, e.g., trend indicators, momentum indicators, volatility indicators and volume indicators [32,49,50]. Furthermore, numerous studies have used an amalgam of different types of technical indicators for SMP.

**Machine Learning Methods**

This section attempts to summarize the machine learning models used in previous studies for stock prediction and forecasting. After the data is pre-processed and transformed to a standard representation, it is fed to machine learning models for further processing. TThe following section briefly summarizes the different machine learning approaches presented:

• Artificial Neural Networks (ANN)

• Support Vector Machine (SVM)

• Naïve Bayes (NB)

• Genetic Algorithms (GA)

• Fuzzy Algorithms (FA)

• Deep Neural Networks (DNN)

• Regression Algorithms (RA)

• Hybrid Approaches (HA)

**Artificial Neural Networks (ANN)**

The ANN is a biological brain-inspired technique in which a large number of artificial neurons are strongly interconnected in order to solve complex problems [90]. These models understand the context of a problem by creating multiple transformations on the feature space, followed by non-linearity, to create its simplified representations [91]. Numerous studies have employed ANN models for SMP [38,40,92–95]. For example, the authors in [68] employed ANN for daily trend prediction of the S&P 500 index. Threedimensional reduction techniques—e.g., PCA, Fuzzy Robust Principal Component Analysis (FRPCA), and Kernel-based Principal Component Analysis—were applied to streamline the dataset. The results suggested that combining the ANNs with the PCAs is more efficient. Furthermore, the selection of an appropriate kernel function directly affects the performance of KPCA [68]. Multilayer perceptron (MLP) is a frequently used technique for SMP [42,43,96,97]. MLP is an ANN with one input and output layer, and one or more intermediate layers. Generally, the MLP uses the backpropagation method for training, in which predicted errors are back propagated from the output layer to the input layer to minimize the errors [98,99]. A study compared three ANN models—MLP, dynamic artificial neural network (DAN2) and autoregressive conditional heteroscedasticity (GARCH)—for NASDAQ price prediction [38]. All three models were evaluated using the Mean Absolute Deviate (MAD) and Mean Square Error (MSE). The results demonstrated that the MLP outperformed DAN2 and GARCH-MLP. Furthermore, it provides a future direction for researchers by suggesting that they focus on finding out whether GARCH has a remedying impact on forecasts or other correlated variables that have a remedying impact on forecasts.

**Support Vector Machine (SVM)**

The SVM is a supervised machine learning technique that limits error and augments geometric margins. It is a pattern classification and regression algorithm. In terms of accuracy, the SVM is an important linear separation algorithm compared to other classifiers. . The authors developed a daily and monthly SMP model using historical and sentimental data for the bank, mining, and oil sectors. The historical prices were obtained from yahoo finance, and a sentiment dataset was created by using news and tweets for one year. PCA with multiple factors was applied to the sparse dataset considered for the sentiment analysis. In this study, three algorithms—i.e., Decision-Boosted Tree, SVM, and Logistic Regression—were compared, and the accuracy was used as a performance metric. The Decision-Boosted Tree outflanked the Logistic Regression and SVM. The DecisionBoosted Tree achieved accuracies of 54.8%, 76%, and 76.9% for the bank, mining, and oil sectors, respectively. The Logistic regression attained accuracies of 65.4%, 61%, and 44.2%, respectively, and the SVM achieved accuracies of 51%, 59%, and 44.2% for the respective sectors. The study finally suggested the consideration of the impact of intra-day price movement for the next-day stock price to improve the accuracy.

**Naïve Bayes (NB)**

NB is a classification method that classifies the data points based on the Bayesian Theorem of probability. This classification method is extremely fast, and can scale over large datasets. This classification approach has been used widely for SMP . For example, the authors employed the Naïve Bayes algorithm for the sentiment analysis of textual data from multiple sources. The authors compared the effect of conventional and social media data sources on different companies and their interrelatedness.

**Genetic Algorithms (GA)**

GA are a heuristic approach to problem-solving that mimic the natural evolution process. The algorithms apply the concept of natural selection to select the optimal possible solution. In SMP, GA is used to fine-tune the parameters for the generation of the best trading rule. Numerous studies have used GA to enhance SMP accuracies. For example, the authors has developed an intelligent decision support system for stock trading. This study employed rough sets and GA for non-linear and complex stock data to find the features that can be used to generate the optimal trading rules. These rules are applied to generate optimal buy or sell strategies.

During the development of the project, we will be using Long Short Term Memory Network algorithm, since it has been used widely around the world in terms of using it for Stock Market Prediction practices.

#### Long short-term memory network:

Long short-term memory network (LSTM) is a particular form of recurrent neural network (RNN).

#### Working of LSTM:

LSTM is a special network structure with three “gate” structures. Three gates are placed in an LSTM unit, called input gate, forgetting gate and output gate. While information enters the LSTM’s network, it can be selected by rules. Only the information conforms to the algorithm will be left, and the information that does not conform will be forgotten through the forgetting gate. The experimental data in this paper are the actual historical data downloaded from the Internet. Three data sets were used in the experiments. It is needed to find an optimization algorithm that requires less resources and has faster convergence speed.

• Used Long Short-term Memory (LSTM) with embedded layer and the LSTM neural network with automatic encoder.

• LSTM is used instead of RNN to avoid exploding and vanishing gradients.

• In this project python is used to train the model, MATLAB is used to reduce dimensions of the input. MySQL is used as a dataset to store and retrieve data.

• The historical stock data table contains the information of opening price, the highest price, lowest price, closing price, transaction date, volume and so on.

• The accuracy of this LSTM model used in this project is 57%.

#### LMS Filter:

The LMS filter is a kind of adaptive filter that is used for solving linear problems. The idea of the filter is to minimize a system (finding the filter coefficients) by minimizing the least mean square of the error signal.

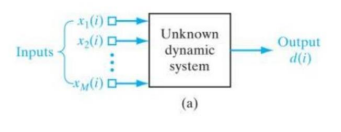


Figure 1 LMS Inputs and Outputs

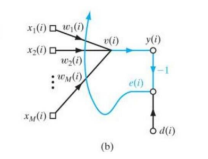
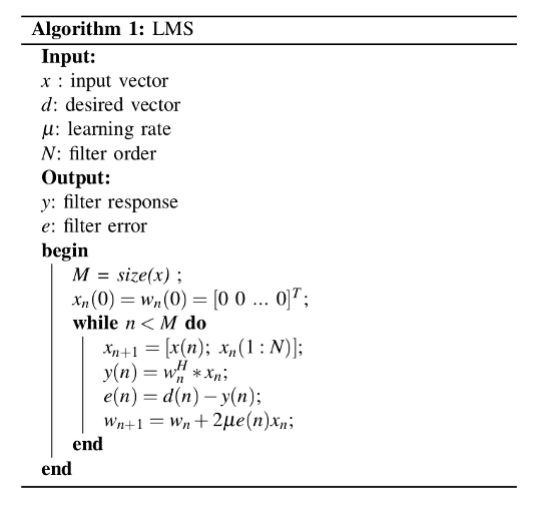


Figure 2 LMS Updating Weights



# In general, we don’t know exactly if the problem can be solved very well with linear approach, so we usually test a linear and a non-linear algorithm. Since the internet always shows non-linear approaches, we will use LMS to prove that stock market prediction can be done with linear algorithms with a good precision. But this filter mimetizes a system, that is, if we apply this filter in our data, we will have the filter coefficients trained, and when we input a new vector, our filter coefficients will output a response that the original system would (in the best case). So we just have to do a tricky modification for using this filter to predict data.

#### The system:

# First, we will delay our input vector by l positions, where l would be the quantity of days we want to predict, this l new positions will be filled by zeros.

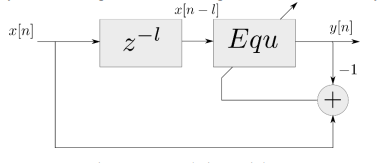
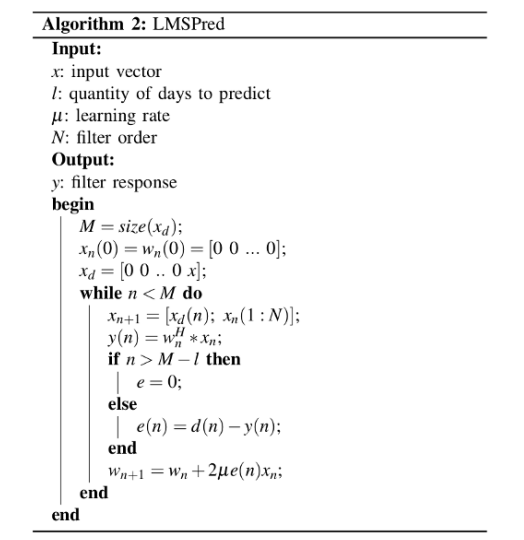


Figure 3 Delay System

When we apply the LMS filter, we will train the filter to the first 178 data. After that, we will set the error as zero, so the system will start to output the answers as the original system to the last l values. We will call the tricky modification as the LMSPred algorithm.



#### LSTM Architecture

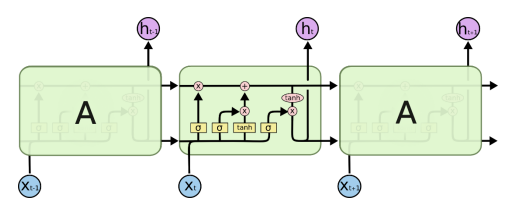


Figure 4 LSTM Architecture

**Forget Gate:**

A forget gate is responsible for removing information from the cell state.

• The information that is no longer required for the LSTM to understand things or the information that is of less importance is removed via multiplication of a filter.

• This is required for optimizing the performance of the LSTM network. • This gate takes in two inputs; h\_t-1 and x\_t. h\_t-1 is the hidden state from the previous cell or the output of the previous cell and x\_t is the input at that particular time step.

**Input Gate:**

1. Regulating what values need to be added to the cell state by involving a sigmoid function. This is basically very similar to the forget gate and acts as a filter for all the information from hi-1 and x\_t.

2. Creating a vector containing all possible values that can be added (as perceived from h\_t-1 and x\_t) to the cell state. This is done using the tanh function, which outputs values from -1 to +1.

3. Multiplying the value of the regulatory filter (the sigmoid gate) to the created vector (the tanh function) and then adding this useful information to the cell state via addition operation.

#### Output Gate:

The functioning of an output gate can again be broken down to three steps:

• Creating a vector after applying tanh function to the cell state, thereby scaling the values to the range -1 to +1.

• Making a filter using the values of h\_t-1 and x\_t, such that it can regulate the values that need to be output from the vector created above. This filter again employs a sigmoid function.

• Multiplying the value of this regulatory filter to the vector created in step 1, and sending it out as a output and also to the hidden state of the next cell.

#### Program Code:

• # LSTM

• Inputs: dataset

• Outputs: RMSE of the forecasted data

•

• # Split dataset into 75% training and 25% testing data

• size = length(dataset) \* 0.75

• train = dataset [0 to size]

• test = dataset [size to length(dataset)]

•

• # Procedure to fit the LSTM model

• Procedure LSTMAlgorithm (train, test, train\_size, epochs)

• X = train

• y = test

• model = Sequential()

• model.add(LSTM(50), stateful=True)

• model.compile(optimizer='adam', loss='mse')

• model.fit(X, y, epochs=epochs, validation\_split=0.2)

• return model

•

• # Procedure to make predictions

• Procedure getPredictonsFromModel (model, X)

• predictions = model.predict(X)

• return predictions

•

• epochs = 100

• neurons = 50

• predictions = empty

• # Fit the LSTM model

• model = LSTMAlgorithm (train, epoch, neurons)

•

• # Make predictions

• pred = model.predict(train)

•

• # Validate the model

• n = len(dataset)

•

• error = 0 • for i in range(n): error += (abs(real[i] - pred[i])/real[i]) \* 100 • accuracy = 100 - error/n

**Hardware Requirements:**

• RAM: 8 GB or more

• Storage: 500 GB

• CPU: 2 GHz or faster

• Architecture: 32-bit or 64-bit Software Requirements:

• Python 3.5 or above in Anaconda Navigator and Jupyter Notebook is used for data pre-processing, model training and prediction.

• Django Python for Restful API.

• Android Studio

• Operating System: windows 7 and above or Linux based OS or MAC OS.

**Functional requirements**

Functional requirements describe what the software should do (the functions). Think about the core operations. Because the “functions” are established before development, functional requirements should be written in the future tense. In developing the software for Stock Price Prediction, some of the functional requirements could include:

• The software shall accept the tw\_spydata\_raw.csv dataset as input.

• The software should shall do pre-processing (like verifying for missing data values) on input for model training.

• The software shall use LSTM ARCHITECTURE as main component of the software.

• It processes the given input data by producing the most possible outcomes of a CLOSING STOCK PRICE.

Notice that each requirement is directly related to what we expect the software to do. They represent some of the core functions.

**Non-Functional requirements**

**Product properties**

• Usability: It defines the user interface of the software in terms of simplicity of understanding the user interface of stock prediction software, for any kind of stock trader and other stakeholders in stock market.

• Efficiency: maintaining the possible highest accuracy in the closing stock prices in shortest time with available data.

Performance: It is a quality attribute of the stock prediction software that describes the responsiveness to various user interactions with it.

#### 3.1.1 System Architecture

1) Preprocessing of Data



Figure 5 Preprocessing Data

1. Overall Architecture

# figure 8

Figure 6 Overall Architecture

# Chapter 4: Proposed Design

### 4.1 Structure Chart

A structure chart (SC) in software engineering and organizational theory is a chart which shows the breakdown of a system to its lowest manageable levels. They are used in structured programming to arrange program modules into a tree. Each module is represented by a box, which contains the module's name.

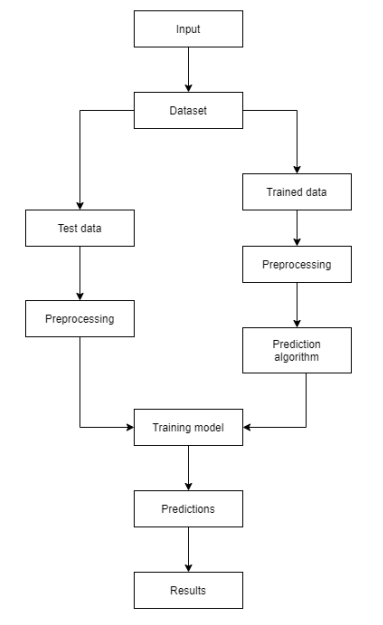


Figure 7 Training and Predictions

### 4.2 UML Diagrams

A UML diagram is a partial graphical representation (view) of a model of a system under design, implementation, or already in existence. UML diagram contains graphical elements (symbols) - UML nodes connected with edges (also known as paths or flows) - that represent elements in the UML model of the designed system. The UML model of the system might also contain other documentation such as use cases written as templated texts. The kind of the diagram is defined by the primary graphical symbols shown on the diagram. For example, a diagram where the primary symbols in the contents area are classes is class diagram. A diagram which shows use cases and actors is use case diagram. A sequence diagram shows sequence of message exchanges between lifelines. UML specification does not preclude mixing of different kinds of diagrams, e.g. to combine structural and behavioral elements to show a state machine nested inside a use case. Consequently, the boundaries between the various kinds of diagrams are not strictly enforced. At the same time, some UML Tools do restrict set of available graphical elements which could be used when working on specific type of diagram. UML specification defines two major kinds of UML diagram: structure diagrams and behavior diagrams. Structure diagrams show the static structure of the system and its parts on different abstraction and implementation levels and how they are related to each other. The elements in a structure diagram represent the meaningful concepts of a system, and may include abstract, real world and implementation concepts. Behavior diagrams show the dynamic behavior of the objects in a system, which can be described as a series of changes to the system over time.

#### 4.2.1 Use Case Diagram

In the Unified Modelling Language (UML), a use case diagram can summarize the details of your system's users (also known as actors) and their interactions with the system. To build one, you'll use a set of specialized symbols and connectors. An effective use case diagram can help your team discuss and represent:

• Scenarios in which your system or application interacts with people, organizations, or external systems.

• Goals that your system or application helps those entities (known as actors) achieve.

• The scope of your system.

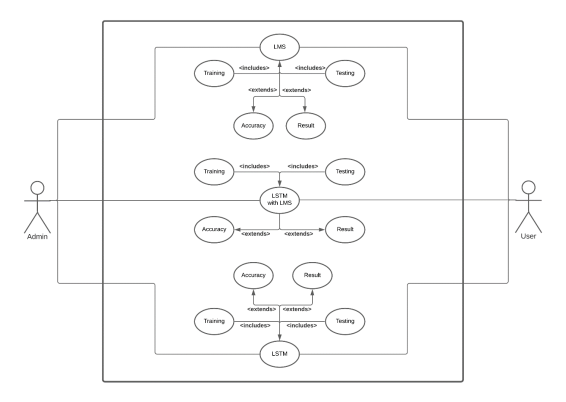


Figure 8 Using LMS, LSTM and LSTM with LMS in the system

#### 4.2.2 Sequence Diagram

A sequence diagram is a type of interaction diagram because it describes how and in what order a group of objects works together. These diagrams are used by software developers and business professionals to understand requirements for a new system or to document an existing process. Sequence diagrams are sometimes known as event diagrams or event scenarios. Sequence diagrams can be useful references for businesses and other organizations. Try drawing a sequence diagram to:

• Represent the details of a UML use case.

• Model the logic of a sophisticated procedure, function, or operation.

• See how objects and components interact with each other to complete a process.

• Plan and understand the detailed functionality of an existing or future scenario.

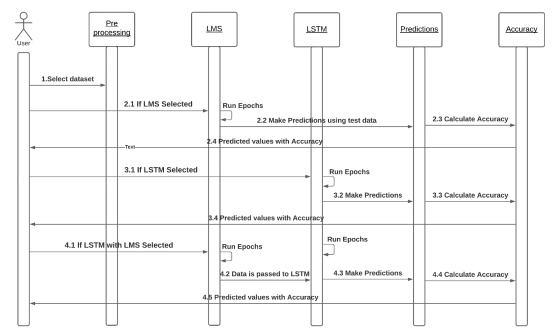


Figure 9 Execution based on Model Selection

**4.2.3 Activity Diagram**

An activity diagram is a behavioral diagram i.e. it depicts the behavior of a system. An activity diagram portrays the control flow from a start point to a finish point showing the various decision paths that exist while the activity is being executed.

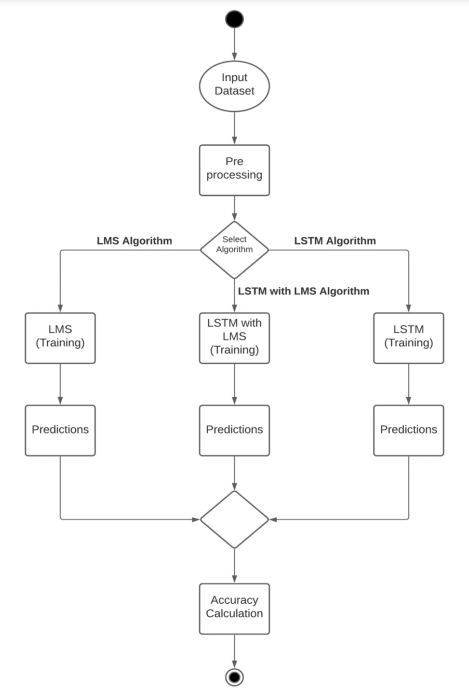


Figure 10 Execution based on Algorithm Selection

**4.2.4 Collaboration Diagram**

Collaboration diagrams are used to show how objects interact to perform the behavior of a particular use case, or a part of a use case. Along with sequence diagrams, collaboration are used by designers to define and clarify the roles of the objects that perform a particular flow of events of a use case. They are the primary source of information used to determining class responsibilities and interfaces. The collaborations are used when it is essential to depict the relationship between the object. Both the sequence and collaboration diagrams represent the same information, but the way of portraying it quite different. The collaboration diagrams are best suited for analyzing use cases.

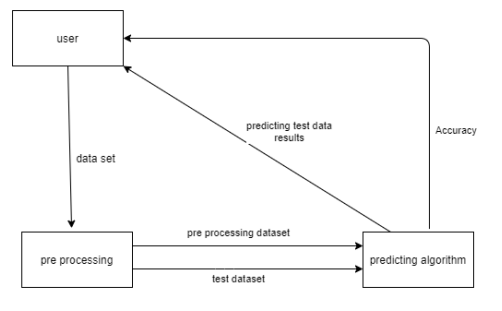


Figure 11 Data Transfer between modules

**4.2.5 Flow Chart**

A flowchart is a type of diagram that represents a workflow or process. A flowchart can also be defined as a diagrammatic representation of an algorithm, a step-by-step approach to solving a task. The flowchart shows the steps as boxes of various kinds, and their order by connecting the boxes with arrows.

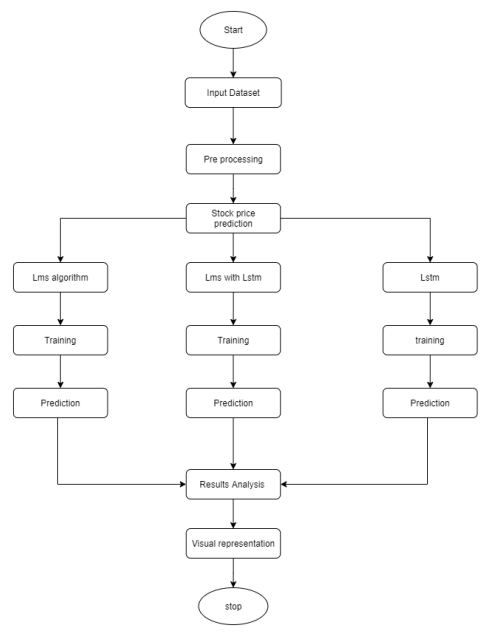


Figure 12 Flow of Execution

**4.2.6 Component Diagram**

Component diagram is a special kind of diagram in UML. The purpose is also different from all other diagrams discussed so far. It does not describe the functionality of the system but it describes the components used to make those functionalities. Component diagrams are used in modeling the physical aspects of object-oriented systems that are used for visualizing, specifying, and documenting component-based systems and also for constructing executable systems through forward and reverse engineering. Component diagrams are essentially class diagrams that focus on a system's components that often used to model the static implementation view of a system.

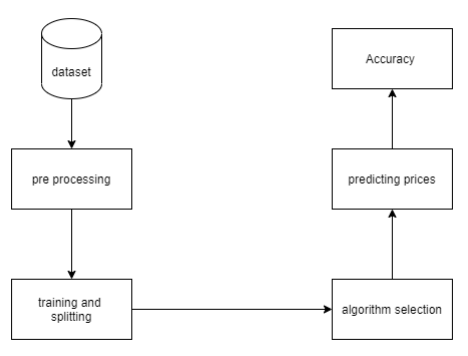


Figure 13 Components present in the system