Visualizing and Forecasting Stocks Using Machine Learning

Submitted in partial fulfilment of the requirements for the award of the degree of

BACHELORS OF TECHNOLOGY

submitted by

Suraj Singh Rana(2018799)

Maninder Singh(2018464)

Under the guidance of

Mr. Kapil Rajput

Assistant Professor

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GRAPHIC ERA HILL UNIVERSITY, DEHRADUN

MAY 2024

CERTIFICATE

We, **Suraj Singh Rana** and **Maninder Singh,** students of **B. Tech CSE VIII Semester**, Department of Computer Science and Engineering, Graphic Era Hill University, Dehradun, declare that the technical project work entitled "**Visualizing and Forecasting of Stocks**" has been carried out by us and submitting in partial fulfilment of the course requirements for the award of degree in Bachelor of Technology of Graphic Era Hill University, Dehradun during the academic year **2023-2024**. This synopsis has not been submitted to any other university for the award of any other degree or diploma.

Mentor Name:
Mr. Kapil Rajput
(Assistant
Professor)
GEHU, Dehradun

Place: Dehradun
Date:15/04/2024

ACKNOWLEDGEMENT

Here by we are submitting the project report on "**Visualizing and Forecasting of Stocks**", as per the scheme of Graphic Era Hill University, Dehradun. We express the deepest sense of gratitude to Mr. Kapil Rajput Asst. Professor, Department of Computer Science and Engineering for the invaluable guidance extended at every stage and in every possible way. I am very much thankful to all the faculty members of the Department of Computer Science and Engineering, friends and my parents for their constant encouragement, support and help throughout the period of project conduction.

Suraj Singh Rana
Roll No: 2018799
Sec: F
Maninder Singh
Roll No: 2018464
Sec: A

ABSTRACT

The rapid advancement in artificial intelligence and machine learning techniques, availability of large-scale data, and increased computational capabilities of the machine opens the door to develop sophisticated methods in predicting stock price. In the meantime, easy access to investment opportunities has made the stock market more complex and volatile than ever. The world is looking for an accurate and reliable predictive model which can capture the market's highly volatile and nonlinear behavior in a holistic framework. This study uses a long short-term memory (LSTM), a particular neural network architecture, to predict the next-day closing price of the S&P 500 index. A well-balanced combination of nine predictors is carefully constructed under the umbrella of the fundamental market data, macroeconomic data, and technical indicators to capture the behavior of the stock market in a broader sense. Single layer and multilayer LSTM models are developed using the chosen input variables, and their performances are compared using standard assessment metrics–Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Correlation Coefficient (R). The experimental results show that the single layer LSTM model provides a superior fit and high prediction accuracy compared to multilayer LSTM models.

It is very difficult to understand extremely complex stock market because it changes constantly. Many factors can push stock prices up or pull them down--the state of the economy, a company's business approach, and how strongly people want to buy or sell stocks. A large number of research papers has been written using Artificial Intelligence and Machine Learning techniques such as artificial neural networks, fuzzy logic and LSTM (Long Short Term Memory) to try to predict how stocks will perform based on their past behavior. These models did not accurately predict the costs. One way to study the stock market

that many people don't know about is using Hidden Markov Models. This is a new way to understand how stock prices change, using a method called the Hidden Markov Model. It's different from what we already know. In this study, we look at Hidden Markov Models and see how they compare to Back Vector Regression. Based on the first results, this method can produce very accurate results. It is more than 80% correct in predicting things in the short term.

In the past decades, there is an increasing interest in predicting markets among economists, policymakers, academics and market makers. The objective of the proposed work is to study and improve the supervised learning algorithms to predict the stock price. Stock Market Analysis of stocks using data mining will be useful for new investors to invest in stock market based on the various factors considered by the software. Stock market includes daily activities like Sensex calculation, exchange of shares. The exchange provides an efficient and transparent market for trading in equity, debt instruments and derivatives. Our aim is to create software that analyses previous stock data of certain companies, with help of certain parameters that affect stock value. We are going to implement these values in data mining algorithms and we will be able to decide which algorithm gives the best result. This will also help us to determine the values that particular stock will have in near future. We will determine the patterns in data with help of machine learning algorithms.

In a financially volatile market, as the stock market, it is important to have a very precise prediction of a future trend. Because of the financial crisis and scoring profits, it is mandatory to have a secure prediction of the values of the stocks. Predicting a non-linear signal requires advanced algorithms of machine learning.

TABLE OF CONTENTS

ABBREVIATIONS

1. CNN : Convolutional Neural Network
2. IoT : Internet of Things
3. AI : Artificial Intelligence
4. SRS : Software Requirements Specification
5. UI: User Interface
6. EMA: Exponential Moving Average
7. MA: Moving Average
8. RNN: Recurrent Neural Network
9. SMA: Simple Moving Average
10. RCNN: Region Based Convolutional Neural Network
11. SVM: Support Vector Machines
12. YFINANCE: Yahoo Finance
13. SMTP: Simple Mail Transfer Protocol

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

In recent times stock market predictions is gaining more attention, maybe due to the fact that if the trend of the market is successfully predicted the investors may be better guided. The profits gained by investing and trading in the stock market greatly depends on the predictability. If there is a system that can consistently predict the direction of the dynamic stock market will enable the users of the system to make informed decisions. More over the predicted trends of the market will help the regulators of the market in taking corrective measures.

The stock price fluctuations are uncertain, and there are many interconnected reasons behind the scene for such behavior. The possible cause could be the global economic data, changes in the unemployment rate, monetary policies of influencing countries, immigration policies, natural disasters, public health conditions, and several others. All the stock market stakeholders aim to make higher profits and reduce the risks from the thorough market evaluation. The major challenge is gathering the multifaceted information, putting them together into one basket, and constructing a reliable model for accurate predictions.

1.2 AIM AND OBJECTIVE

The primary aim of using Long Short-Term Memory (LSTM) networks for stock price prediction is to leverage their capability to model and understand the sequential and temporal dependencies in stock market data. LSTM networks are a type of Recurrent Neural Network (RNN) particularly effective in dealing with time-series data, which is crucial for predicting future stock prices based on historical data.

The aim of the project is to examine a number of different forecasting techniques to predict future stock returns based on past returns and numerical news indicators to construct a portfolio of multiple stocks in order to diversify the risk. We do this by applying supervised learning methods for stock price forecasting by interpreting the seemingly chaotic market data.

This approach can help companies improve efficiency, reduce costs, and enhance customer experiences. One method for time-series forecasting is using long short-term memory (LSTM), an artificial recurrent neural network that can be fitted for time-series data. The goal of stock price prediction using LSTM is to forecast future stock prices based on historical data accurately. However, predicting future stock prices using AI methods is still a challenging problem. Careful analysis and validation of predictive models are necessary before making investment decisions based on their predictions. The stock market has a significant impact on a nation's economy, and it also affects the global economy. Natural language processing (NLP) techniques are being employed to forecast market trends by analyzing the sentiment of news articles. Recent advances in language modeling have greatly improved sentiment analysis by providing powerful tools for precise, timely, and semantic text analysis. To predict future changes in the stock market, several studies have used different statistical methods or ML approaches, such as Support Vector Machine (SVM), based on historical data. Deep learning neural network models such as LSTM have also proven successful in various fields, including financial analysis. LSTM is a unique type of RNN model that could store long or short-term values and filter out irrelevant information, making it an effective tool for conducting time-series analysis.

The specific aims include:

Accurate Forecasting*:* To achieve precise and reliable predictions of future stock prices.

Pattern Recognition: To identify and learn from complex patterns and trends in historical stock data.

Risk Management: To aid investors and financial analysts in making informed decisions, thereby reducing investment risks.

Automation: To develop automated trading systems that can make timely buy or sell decisions based on predicted stock prices.

Scope:

The scope of stock price prediction using LSTM encompasses various aspects and stages, including:

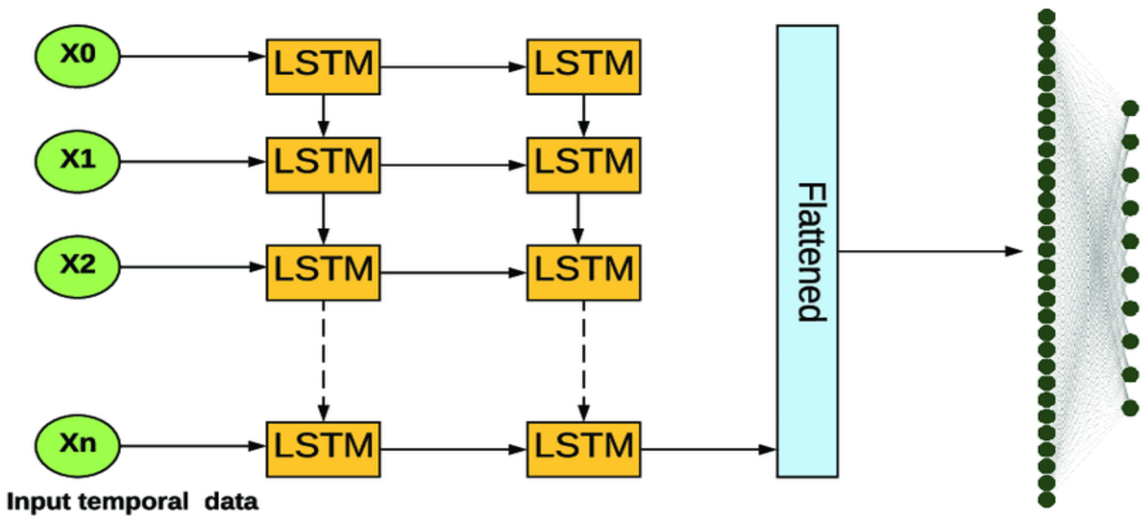


Fig. 1.2 Architecture Model

Data Collection and Preprocessing:

- Gathering historical stock prices, including open, close, high, low, and volume data.
- Cleaning the data to handle missing values, outliers, and inconsistencies.
- Normalizing or scaling the data to ensure it is suitable for LSTM input.

Feature Engineering:

- Creating relevant features from raw stock data, such as moving averages, trading volumes, and other technical indicators.
- Selecting important features that significantly impact the stock prices.

Model Development:

- Designing the LSTM network architecture, including the number of layers, neurons, and other hyperparameters.
- Training the model using historical data and validating its performance with validation datasets.

Evaluation and Validation:

- Evaluating the model's performance using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), or other relevant metrics.
- Validating the model with unseen data to ensure its generalization ability.

Prediction and Analysis:

- Using the trained LSTM model to predict future stock prices.
- Analyzing the predictions to identify trends and potential market movements.

Applications:

Individual Investors: Helping individuals make better investment choices based on predicted stock trends.

Financial Institutions: Enabling banks, hedge funds, and investment firms to optimize their trading strategies.

Algorithmic Trading: Integrating predictions into automated trading systems to execute trades based on predefined rules.

Challenges and Limitations:

- Dealing with the inherent volatility and noise in stock market data.
- Ensuring the model adapts to sudden market changes and anomalies.
- Balancing the trade-off between model complexity and interpretability.

Future Directions:

- Enhancing prediction accuracy by integrating other data sources such as news sentiment, economic indicators, and social media trends.
- Exploring hybrid models that combine LSTM with other machine learning or deep learning techniques for improved performance.
- Continuous updating and retraining of models to adapt to evolving market conditions.

1.3 STOCK MARKET

A stock market, equity market or share market is the aggregation of buyers and sellers (a loose network of economic transactions, not a physical facility or discrete entity) of stocks (also called shares), which represent ownership claims on businesses; these may include securities listed on a public stock exchange as well as those only traded privately. Examples of the latter include shares of private companies which are sold to investors through equity crowd funding platforms. Stock exchanges list shares of common equity as well as other security types, e.g. corporate bonds and convertible bonds.
Stock price prediction is one of the most widely studied problem, attracting researchers from many fields. The volatile nature of the stock market makes it really difficult to apply simple time-series or regression techniques. Financial institutions and active traders have created

various proprietary models to beat the market for

themselves or their clients, but rarely did anyone achieve consistently higher than the average returns on investment. The challenge of stock market price forecasting is so appealing because an improvement of just a few points of percentage can increase the profit by millions of dollars. This paper discusses the application of LSTM and Linear Regression in detail along with the pros and cons of the given methods. The paper introduces the parameters and variables which can be used to recognize the patterns in stock prices which can be helpful in future stock prediction and how boosting can be integrated with various other machine learning algorithms to improve the accuracy of our prediction systems.

1.4 APPLICATION OF LSTM MODELS IN STOCK MARKET FORCASTING

LSTM is a Recurrent Neural Network that works on data sequences, learning to retain only relevant information from a time window. New information the network learns is added to a "memory" that gets updated with each timestep based on how significant the new sample seems to the model. Over the years, LSTM has revolutionized speech and handwriting recognition, language understanding, forecasting, and several other applications that have become the new normal today.

A standard LSTM cell comprises of three gates: the input, output, and forget gate. These gates learn their weights and determine how much of the current data sample should be remembered and how much of the past learned content should be forgotten. This simple structure is an improvement over the previous and similar RNN model.

Several deep learning architectures have been developed to deal with various problems and the intrinsic structure of datasets. Information flows only in the forward direction in a basic feedforward neural network architecture. Since each input is processed independently, it does not retain information from the previous step. Thus, these models are ineffective in dealing with sequential data where series of prior events are essential in predicting future events. Recurrent neural networks (RNN) are designed to perform such tasks. The RNN architecture consists of loops, allowing relevant information to persist over time. Information is being passed from one timestep to the next internally within the network. Therefore, the RNN is more suitable for sequential data modeling and time series applications such as stock market predictions, language translations, auto-completion in messages/emails, and signal processing. During the training process of the RNN, the cost or error is calculated between the predicted values and the actual values from a labeled training dataset. The error is minimized by repeatedly updating the networks' parameters (weights and biases) until the lowest possible value is obtained. The training process utilizes a gradient, the rate at which cost changes with respect to each parameter. The gradient provides a direction to move in the error surface by adjusting the parameters iteratively. This strategy is called backpropagation, where the error is propagated backward from the output layer all the way up to the input layer. One of the challenges of this technique is that parameters can be anywhere in the networks, and finding a gradient involves calculations of partial derivatives with respect to all the parameters. This process sometimes needs a long chain rule, especially for the parameters in earlier layers of the networks. As a result, gradients could ultimately vanish or decay back through the networks, known as the vanishing gradient problem, a common issue in neural networks training. Unfortunately, this problem also persists in the RNN architecture. LSTM, a typical recurrent neural network architecture, is designed to overcome the vanishing gradient problem (Hochreiter, 1998). Memorizing information for a longer period of time is the default behavior of the LSTM model.

This study considers the computational framework to predict the stock index price using the LSTM model, the improved version of neural networks architecture for time series data. The bird's-eye view of the proposed research framework via the schematic diagram is expressed in Fig. 1. As outlined in the diagram, the proposed study utilizes the carefully selected features from fundamental, macroeconomic, and technical data to build the model. After that, the collected data has been normalized using the min–max normalization technique. Then input sequence for the LSTM model is created using a specific time step. The hyperparameters such as number of neurons, epochs, learning rate, batch size, and time step have been incorporated in the model. The regularization techniques have been utilized to overcome the overfitting problems. Once the hyperparameters are tuned, the input data is fed into the LSTM model to predict the closing price of the stock market index. The quality of the proposed model is assessed through RMSE, MAPE, and R.

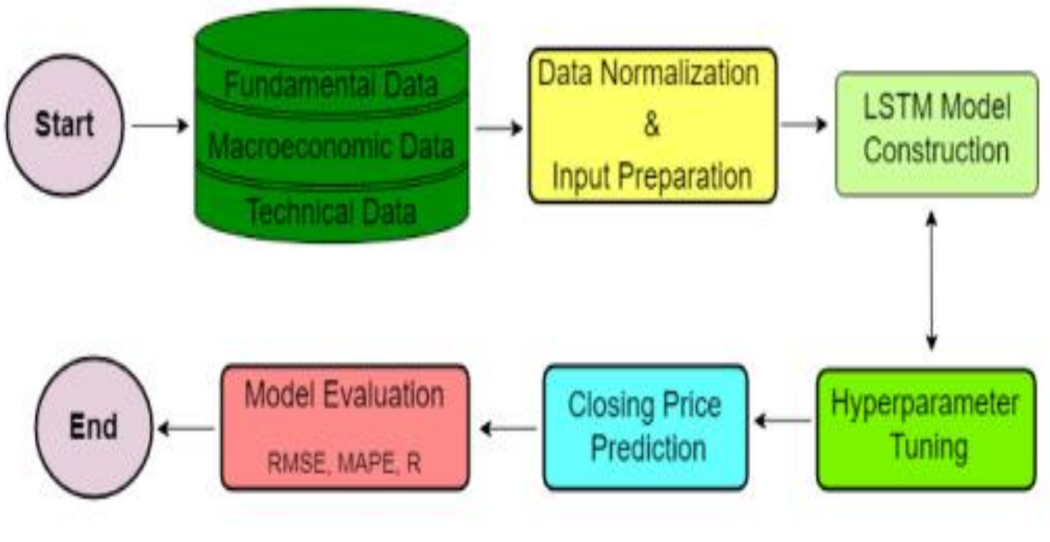


Fig. 1.4.1 Schematic Model

1.5 MOTIVATION

Stock price prediction is a classic and important problem. With a successful model for stock prediction, we can gain insight about market behavior over time, spotting trends that would otherwise not have been noticed. With the increasingly computational power of the computer, machine learning will be an efficient method to solve this problem.

Thus, our motivation is to design a public service incorporating historical data and users predictions to make a stronger model that will benefit everyone.

The pursuit of forecasting stock prices presents considerable potential for financial gain, thereby stimulating scholarly inquiry in this domain. Even a modest degree of understanding regarding the valuation of a stock can yield significant profits. Hence, scholars in both the industrial and academic sectors are persistently exploring avenues to surmount obstacles such as volatility, seasonality, and reliance on time, economics, and market conditions, leveraging technology as a means

1.6 SYSTEM FLOW DIAGRAM

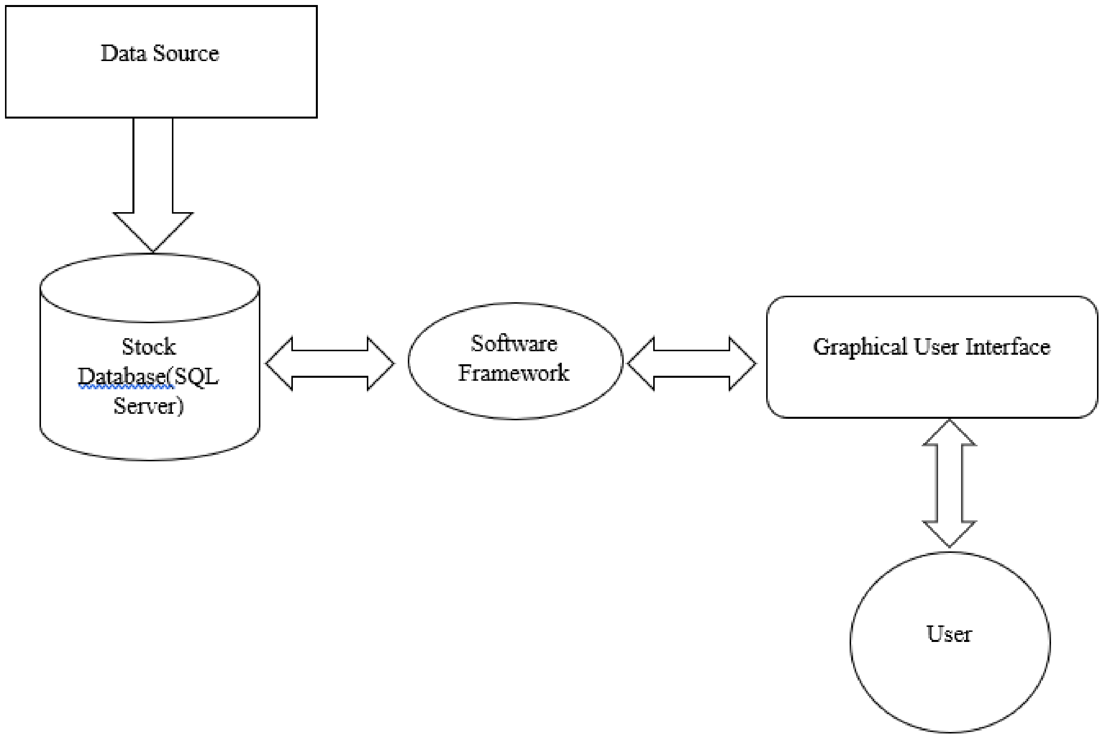


Fig. 1.6 System Flow Diagram

IMPLEMENTATION STEPS

System architecture is a model that defines the behavior of a system in the conceptual model. The huge systems are decomposed into subordinate systems to provide similar set of services. The beginning layout strategy of perceiving these sub-systems and building up a structure for sub-systems control and cooperation is called architecture design. As shown above, Fig. 4.1 includes seven major steps to implement the system and each step is explained below.

1.6.1 Understanding the Objective

The first step in developing a project is to understand the objective which involves an understanding of the intent and essentials of a system. This comprehension is used as a problem description and a preparatory system to accomplish the expectations. The objective of our project is neither to build a system that makes billions nor to waste billions too. But the objective is to develop a system that finds the direction of change of stock price indices based on the co-relations between stock prices and help the investors in the stock market in taking a decision whether to buy/sell/hold a stock by providing the results in-terms of visualizations.

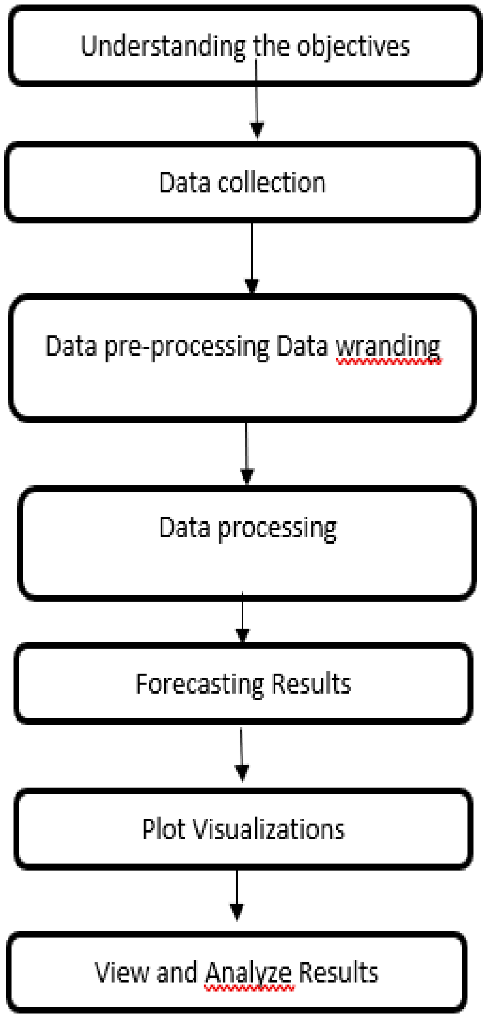


Fig. 1.6.1 Data processing Diagram

CHAPTER 2

Literature Survey

The use of machine learning techniques for stock price prediction has gained significant attention in recent years. Various researchers have used AI for stock price prediction. Wei Bao et al. presented wavelet transforms, LSTM, and stacked autoencoders (SAEs) for stock prediction. The authors used SAEs for hierarchically extracted deep features for stock price forecasting. In, a complementary ensemble empirical mode decomposition (CEEMD) with principal component analysis (PCA) and LSTM, a hybrid prediction model for the stock market, is proposed. A combination of a neural tensor network and the deep convolution neural network (CNN) is used to analyze the influence of events on stock price fluctuations [5]. In addition, various researchers have used deep learning methods for volatility and price prediction and trend forecasting for financial assets, and different models were designed and adjusted for efficient forecasting outcomes. A fuzzy random regression model applied to real data sets from the Kula Lumpur stock exchange and Alabama University were attempted in [10]. The results showed that enhancing the fuzzy random regression model's accuracy, variability, and spread adjustment is crucial in data preparation. Sequence reconstruction by leveraging motifs and using CNN to record the spatial structure of time series was attempted to reduce noise-filled financial temporal series. DNNs and traditional artificial neural networks (ANNs) integrated with two datasets transformed with PCA were used to forecast the daily direction of stock market index returns. Also, certain works use deep belief networks (DBN) for forecasting exchange rates. The Classical DBN was updated using continuous restricted Boltzmann machines (CRBMs), and the DBN structure was optimally determined via various experiments for application in ex-change rate forecasting. Another approach wherein data is not fitted to any model, rather latent dynamics in the data is identified using deep learning architectures. Price prediction of NSE-listed companies is carried out using three different deep learning architectures. In, dynamic mode decomposition (DMD) is used. It is decomposed into spatialtemporal systems and is used to forecast the future state of the stock market. The different modes have predetermined temporal behavior [15]. They used time series data of stock prices of companies listed on the National Stock Exchange. In, fourteen different deep learning models based on LSTM, Gated Recurring Unit (GRU), CNN, and Extreme Learning Machines (ELM) are designed and tested on stocks in the S&P BSEBANKEX index. They generated one-step ahead and four-step ahead forecasts. Also, in , the relation between the Indian stock returns and that of the USA was explored. A detailed literature survey elucidating data characteristics and features for training and testing different ML and AI models to solve financial problems is presented. Forex price prediction is achieved by integrating the Forex Loss Function (FLF) into an LSTM called FLF-LSTM. Their results showed that the proposed model minimizes the difference between Forex candles' actual and predictive averages. Various techniques, including machine learning and lexicon-based methods, have analyzed market sentiment. The lexicon-based method is categorized into two types: dictionary-based and corpusbased approaches. On the other hand, unsupervised and supervised ML approaches are further classified into probabilistic, rule-based, decision tree classification, and other types. Earlier studies conducted by Ren, Wu, and Liu (2018), Kumar & Garg (2019), and Li, Jin, & Quan (2020) have discovered that natural language processing, which encompasses text sentiment analysis (SA), can be utilized to predict stock market fluctuations, track public opinion, anticipate box office revenue, and evaluate product reviews. Dev Shah et al. suggested a dictionarybased paradigm using the pattern Python library to convert a text corpus into numerical vectors and create a sentiment score based on the frequency of positive and negative terms. However, this approach may result in considerable deviations from the market sentiment. Nofsinger(2001) stated that investor sentiment is crucial in the financial market, and SA and news analysis are used to predict stock values]. Kirange et al. (2019) proposed a predictive model that leverages emotional content classification of news articles to derive sentiment polarity and analyze changes in market trends. They collected data from reputable news sources and used machine learning methods such as naive Bayes, SVM, and KNN. The results showed that KNN outperformed the other methods. Meanwhile, Bollen, Mao & Zeng (2011) discovered that the content of Amazon posts is linked to the Dow Jones Industrial Average Index (DJIA), while Liu et al. (2017) employed sentiment analysis of posts from the China Stock Forum and predicted the volatility of the Chinese stock market. Their results confirmed that sentiment indicators could effectively enhance the accuracy of the prediction.

1. Survey of stock market prediction using machine learning approach Authors: Ashish Sharma ; Dinesh Bhuriya ; Upendra Singh

2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)

Stock market is basically nonlinear in nature and the research on stock market is one of the most important issues in recent years. People invest in stock market based on some prediction. For predict, the stock market prices people search such methods and tools which will increase their profits, while minimize their risks. Prediction plays a very important role in stock market business which is very complicated and challenging process. Employing traditional methods like fundamental and technical analysis may not ensure the reliability of the prediction. To make predictions regression analysis is used mostly. In this paper we survey of well-known efficient regression approach to predict the stock market price from stock market data based. In future the results of multiple regression approach could be improved using more number of variables.

2. Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network

Authors: Ze Zhang ; Yongjun Shen ; Guidong Zhang ; Yongqiang Song ; Yan Zhu, 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)

Stock price is one of intricate non-linear dynamic system. Typically, Elman neural network is a local recurrent neural network, having one context layer that memorizes the past states, which is quite fit for resolving time series issues. Given this, this paper takes Elman network to predict the opening price of stock market. Considering that Elman network is limited, this paper adopts self-adapting variant PSO algorithm to optimize the weights and thresholds of network. Afterwards, the optimized data, regarded as initial weight and threshold value, is given to Elman network for training, accordingly the prediction model for opening price of stock market based on self-

adapting variant PSO-Elman network is formed. Finally, this paper verifies that model by some stock prices, and compares with BP network and Elman network, so as to draw the result that shows the precision and stability of this predication model both are superior to the traditional neural network.

3. Stock market prediction using an improved training algorithm of neural network

Authors: Mustain Billah ; Sajjad Waheed ; Abu Hanifa,2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)

Predicting closing stock price accurately is an challenging task. Computer aided systems have been proved to be helpful tool for stock prediction such as Artificial Neural Net-work(ANN), Adaptive Neuro Fuzzy Inference System (ANFIS) etc. Latest research works prove that Adaptive Neuro Fuzzy Inference System shows better results than Neural Network for stock prediction. In this paper, an improved Levenberg Marquardt(LM) training algorithm of artificial neural network has been proposed. Improved Levenberg Marquardt algorithm of neural network can predict the possible day-end closing stock price with less memory and time needed, provided previous historical stock market data of Dhaka Stock Exchange such as opening price, highest price, lowest price, total share traded. Morever, improved LM algorithm can predict day-end stock price with 53% less error than ANFIS and traditional LM algorithm. It also requires 30% less time, 54% less memory than traditional LM and 47% less time, 59% less memory than ANFIS.

4. Literature review on Artificial Neural Networks Techniques Application for Stock Market Prediction and as Decision Support Tools

Authors: Muhammad Firdaus ; Swelandiah Endah Pratiwi ; Dionysia Kowanda ; Anacostia Kowanda

This literature review is aiming to explore the use Artificial Neural Network (ANN) techniques in the field of stock market prediction. Design: Content analysis research technique. Data sources: Information retrieved from ProQuest electronic databases. Review methods: Utilizing key terms and phrases associated with Artificial Neural Network Stock Market Prediction from 2013-2018. Out of the 129 scholarly journal reviewed, there are 4 stock market studies met the inclusion criteria. The analysis and the evaluation includes 6

ANN derivatives techniques used to predict. Results: Findings from the reviewed studies revealed that all studies shows consistency that the accuracy rate of ANN stock market prediction is high. 2 Studies shows accuracy above 90%, 2 studies shows accuracy above 50%. Conclusion: This study reveals that the ability of ANN shows consistency of an accuracy rate of stock market prediction. Four method in predicting stock market had an accuracy above 95%. The highest accuracy achieved by using Signal Processing/Gaussian Zero-Phase Filter (GZ-Filter) with 98.7% prediction accuracy.
2018 Third International Conference on Informatics and Computing (ICIC)

5. Stock Market Movement Prediction using LDA-Online Learning Model Authors:Tanapon Tantisripreecha ; Nuanwan Soonthomphisaj, 2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)

In this paper, an online learning method namely LDA-Online algorithm is proposed to predict the stock movement. The feature set which are the opening price, the closing price, the highest price and the lowest price are applied to fit the Linear Discriminant

Analysis (LDA). Experiments on the four well known NASDAQ stocks (APPLE, FACBOOK GOOGLE, and AMAZON) show that our model provide the best performance in stock prediction. We compare LDA-online to ANN, KNN and Decision Tree in both Batch and Online learning scheme. We found that LDA-Online provided the best performance. The highest performances measured on GOOGLE, AMAZON, APPLE GOOGLE stocks are 97.81%, 97.64%, 95.58% and 95.18% respectively.

06. Stock Price Prediction Using News Sentiment Analysis

Authors:Vijayvergia ; David C. Anastasiu,2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)

Predicting stock market prices has been a topic of interest among both analysts and researchers for a long time. Stock prices are hard to predict because of their high volatile nature which depends on diverse political and economic factors, change of leadership, investor sentiment, and many other factors. Predicting stock prices based on either historical data or textual information alone has proven to be insufficient. Existing studies in sentiment analysis have found that there is a strong correlation between the movement of stock prices and the publication of news articles. Several sentiment analysis studies have been attempted at various levels using algorithms such as LSTM, naive Bayes regression, and deep learning. The accuracy of deep learning algorithms depends upon the amount of training data provided. However, the amount of textual data collected and analyzed during the past studies has been insufficient and thus has resulted in predictions with low accuracy. In our paper, we improve the accuracy of stock price predictions by gathering a large amount of time series data and analyzing it in relation to related news articles, using deep learning models. The dataset we have gathered includes daily stock prices for S&P500 companies for five years, along with more than 265,000 financial news articles related to these companies. Given the large size of the dataset, we use cloud computing as an invaluable resource for training prediction models and performing inference for a given stock in real time. Index Terms-stock market prediction, cloud, big data, machine learning, regression.

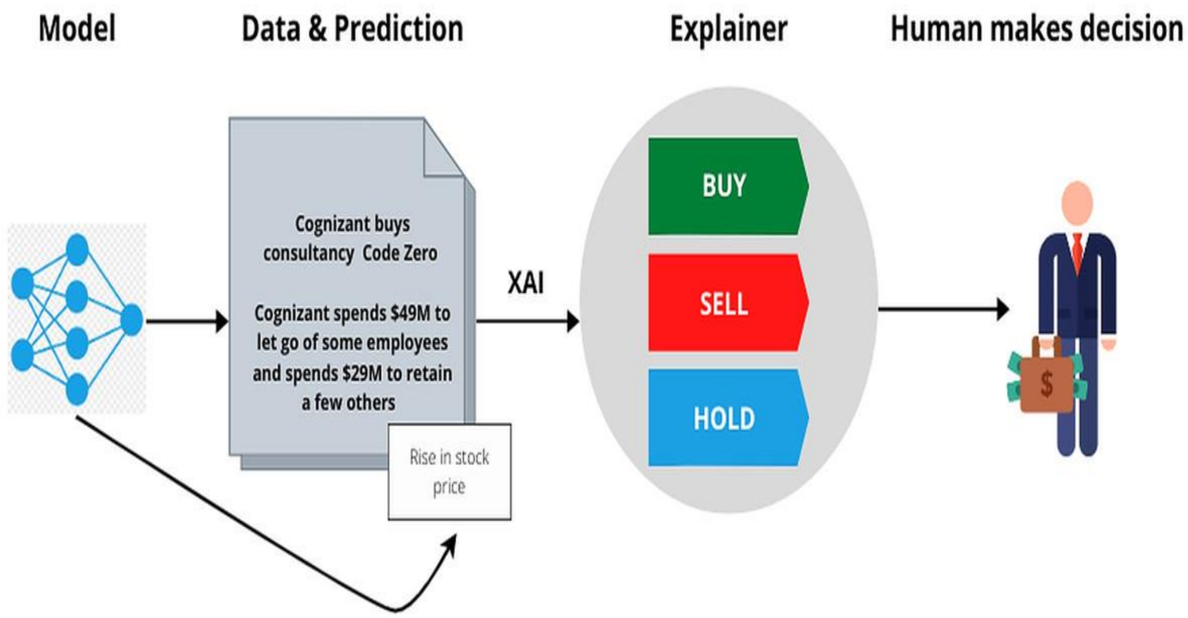


Fig. 1.2 Data Flow Model

CHAPTER 3
AIM AND SCOPE

3.1 OBJECTIVES

The aims of this project are as to identify factors affecting share market, To generate the pattern from large set of data of stock market for prediction of stock exchange and to predict an approximate value of share price to provide analysis for users through web application

The objective of the system is to give a approximate idea of where the stock market might be headed. It does not give a long term forecasting of a stock value. There are way too many reasons to acknowledge for the long term output of a current stock. Many things and parameters may affect it on the way due to which long term forecasting is just not feasible.

3.2 EXISTING SYSTEM

Nowadays, as the connections between worldwide economies are tightened by globalization, external perturbations to the financial markets are no longer domestic. With evolving capital markets, more and more data is being created daily.

The intrinsic value of a company's stock is the value determined by estimating the expected future cash flows of a stock and discounting them to the present, which is known as the book value. This is distinct from the market value of the stock, that is determined by the company's stock price. This market value of a stock can deviate from the intrinsic value due to reasons unrelated to the company's fundamental operations, such as market sentiment.

The fluctuation of stock market is violent and there are many complicated financial indicators. Only few people with extensive experience and knowledge can understand the meaning of the indicators and use them to make good prediction to get fortune. Most people have to rely solely on luck to earn money from stock trading. However, the advancement in technology, provides an opportunity to gain steady fortune from stock market and also can help experts to find out the most informative indicators to make better prediction. The prediction of the market value is
of paramount importance to help in maximizing the profit of stock option purchase while keeping the risk low.

Stock Price Prediction by Machine Learning present to estimate the stock future value and machine learning technique like LSTM for existing work. This machine-learning algorithm is to perform the best predicting result of the stock future price. LSTM is capable to catching the modifications in the behaviour of the stock price for the indicated period in this proposed system.

Propose [3] a machine learning-based normalization for stock price prediction. The dataset utilized for analysis was selected from Yahoo Finance. It consists of approximately 9 lakh records of the required Stock price and other relevant data. The data reflected the stock price at some time intervals for every day of the year. It contains various data like date, symbol, open price, close price, low price, high price and volume. Here, the data for only one company was considered. All the data was available in a file of CSV format which was first read and transformed into a data frame using the Pandas library in Python. The normalization of the data was performed through the sklearn library in Python and the data were divided into training and testing sets. The experiment set was kept as 20% of the available dataset. This paper focuses on two architecture Regression-based Model and LSTM. The Regression-based Model is employed for predicting unbroken values through some given autonomous values Regression uses a given linear function for predicting continuous values of the most important amongst them and made the predictions using these. LSTM architecture is able to identify the changes in trends which show evident from the result. LSTM is identified as the best model for the proposed methodology. This shows that the proposed system is capable of identifying some interrelation within the data. In the stock market, there may not always follow the same cycle or may not always be in a regular pattern for the changes that are occurred.The period of the existence will differ and the existence of the trend is based on the companies and the sectors. For investors, this type of analysis of trends and cycles will obtain more profit. We must use networks like LSTM as they rely on the current information to analyse various information.

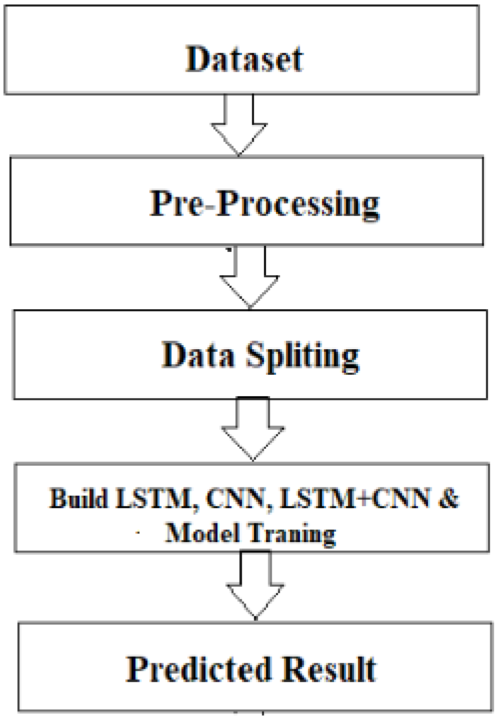


Fig. 3.1 Proposed Workflow Model

3.3 PROPOSED SYSTEM

Linear Regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression.

Advantages

- Space complexity is very low it just needs to save the weights at the end of training. Hence, it's a high latency algorithm.
- It is very simple to understand
- Good interpretability

Feature importance is generated at the time model building. With the help of hyperparameter lambda, you can handle features selection hence we can achieve dimensionality reduction

.

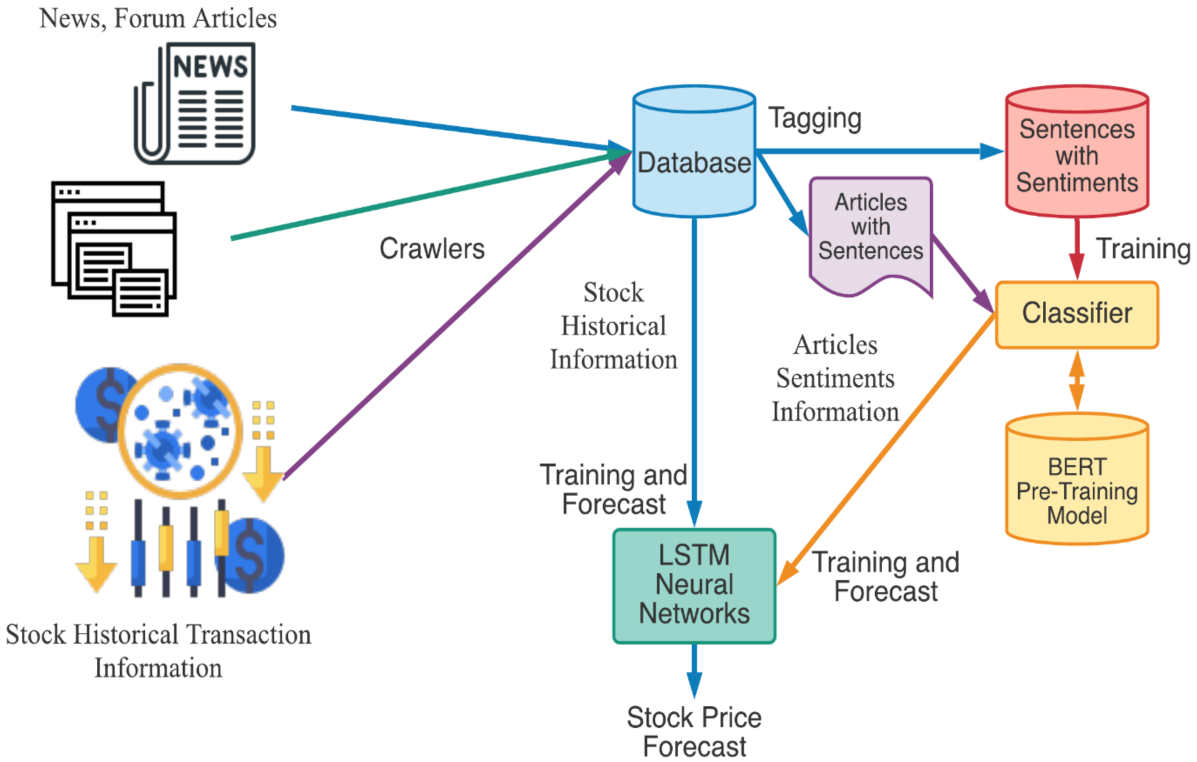


Fig. 3.3 Visual Architecture

3.4 Goal:

The primary goal of stock visualization is to present complex financial data in an intuitive and accessible manner, enabling investors, analysts, and traders to make informed decisions. Visualization helps in identifying patterns, trends, and anomalies in stock prices and market behaviors, which can be difficult to discern from raw data alone.

3.5 Motivation:

3.5.1 Simplifying Complex Data:

Stock market data can be highly complex, with numerous variables such as price, volume, volatility, and financial indicators. Visualization simplifies this complexity, allowing users to grasp essential information quickly.

3.5.2 Enhancing Decision-Making:

Visual representations of stock data, such as line charts, candlestick charts, and heat maps, provide clear insights into market trends and price movements. This aids investors and traders in making timely and well-informed decisions.

3.5.3 Identifying Patterns and Trends:

Visualization tools help in identifying historical patterns and trends, such as support and resistance levels, moving averages, and other technical indicators. Recognizing these patterns is crucial for predicting future price movements.

3.5.4 Real-Time Monitoring:

Real-time visualization of stock prices allows traders to monitor the market continuously and react swiftly to changes. This is particularly important for high-frequency traders who rely on immediate data to execute trades.

3.5.5 Comparative Analysis:

Visual tools enable the comparison of multiple stocks or indices simultaneously. This helps investors to evaluate the performance of different assets, sectors, or markets, and to make portfolio adjustments accordingly.

3.5.6 Risk Management:

By visualizing volatility and other risk indicators, investors can better assess and manage their risk exposure. Tools like VaR (Value at Risk) charts and stress test visualizations are instrumental in this process.

3.5.7 Educational Purposes:

Visualization serves as an educational tool for novice investors, helping them to understand market dynamics and technical analysis concepts. Interactive visualizations can make learning about the stock market more engaging and effective.

3.5.8 Communicating Insights:

For financial analysts and advisors, visualizations are a powerful means of communicating insights and recommendations to clients or stakeholders. Clear and compelling visual data representations can enhance the credibility and impact of their analyses.

Examples of Stock Visualization Tools and Techniques

3.5.9 Line Charts:

Line charts are one of the most basic forms of stock visualization, showing the closing prices over a specified period. They are useful for identifying overall trends.

Candlestick Charts:

Candlestick charts provide detailed information about price movements within a given time frame, including opening, closing, high, and low prices. They are essential for technical analysis.

Heat Maps:

Heat maps visually represent data such as stock performance or volatility in a matrix format, with color coding to indicate the magnitude. They are useful for identifying market sentiment and sector performance.

3.5.10 Moving Averages:

Charts that incorporate moving averages help in smoothing out price data to highlight trends over time. They are commonly used to identify trend directions and potential reversal points.

Volume Charts:

Volume charts display the trading volume for a stock alongside its price, providing insights into the strength and validity of price movements.

3.5.11 Interactive Dashboards:

Platforms like Tableau, Power BI, and custom-built web applications offer interactive dashboards that allow users to explore stock data dynamically, apply filters, and drill down into specific details.

Implementation Tools and Technologies

Programming Languages:

Python (with libraries like Matplotlib, Plotly, Seaborn) and R are widely used for creating stock visualizations due to their robust data manipulation and visualization capabilities.

Financial APIs:

APIs such as Alpha Vantage, Yahoo Finance, and Quandl provide the necessary data for building real-time and historical stock visualizations.

Web Technologies:

JavaScript libraries such as D3.js, Chart.js, and Highcharts are popular for creating interactive web-based visualizations.

Machine Learning:

Techniques like LSTM (Long Short-Term Memory) networks are used for predicting stock prices, which can then be visualized to show potential future trends.

Visualization Software:

Tools like Tableau and Power BI offer powerful features for creating and sharing complex stock visualizations without requiring extensive programming knowledge.

3.6 SCOPE:

The scope of stock visualization encompasses a wide range of applications and benefits across various sectors. Here are the key areas where stock visualization plays a crucial role:

3.6.1. Investment Analysis

Trend Identification: Visual tools help investors identify market trends and price patterns, such as uptrends, downtrends, and consolidation phases.

Technical Analysis: Candlestick charts, moving averages, and other technical indicators are visualized to aid in predicting future price movements.

Performance Comparison: Investors can compare the performance of different stocks, indices, or sectors over various time periods.

3.6.2. Portfolio Management

Diversification Analysis: Visualizations help in assessing the diversification of a portfolio by showing the distribution of investments across different sectors and asset classes.

Risk Assessment: Tools like risk heat maps and volatility charts enable investors to evaluate and manage their risk exposure effectively.

Performance Tracking: Portfolio performance can be tracked and visualized over time, allowing investors to make informed decisions about rebalancing or adjusting their investments.

3.6.3. Market Monitoring

Real-Time Data Visualization: Traders and analysts can monitor real-time stock prices, trading volumes, and other market data to make timely trading decisions.

News and Sentiment Analysis: Integrating news feeds and sentiment analysis visualizations helps in understanding market sentiment and its potential impact on stock prices.

3.6.4. Algorithmic and High-Frequency Trading

Pattern Recognition: Visualizations assist in identifying trading patterns and opportunities for algorithmic trading strategies.

Backtesting: Historical data visualization is crucial for backtesting trading algorithms to evaluate their performance before deployment.

3.6.5. Risk Management

Volatility Analysis: Visual tools show historical and expected volatility, helping investors and risk managers to understand and prepare for potential market fluctuations.

Stress Testing: Visualization of stress test results helps in assessing the impact of extreme market conditions on a portfolio.

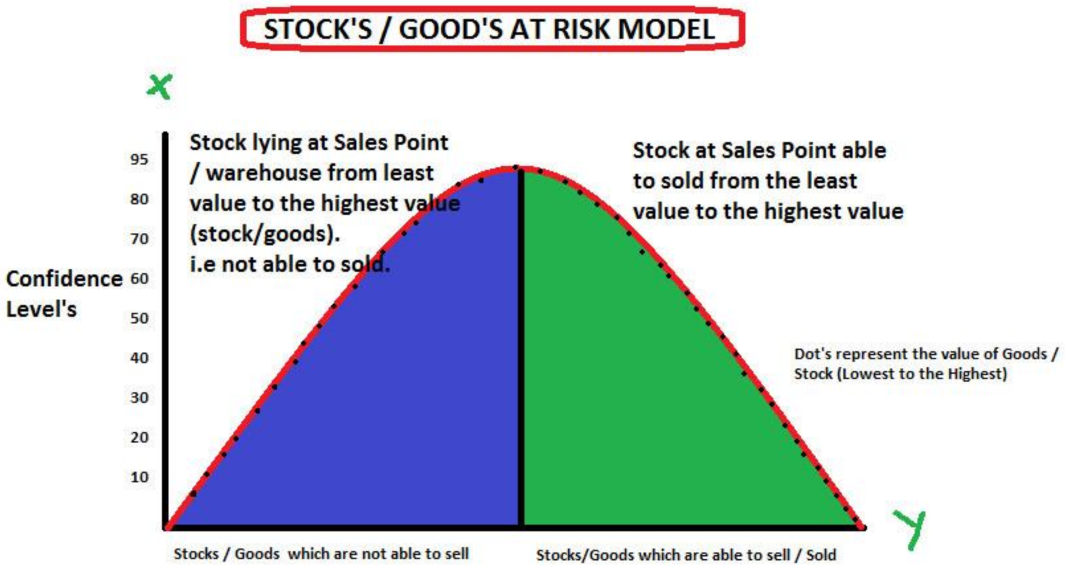


Fig. 3.6 Risk Model

3.6.6. Research and Education

Financial Research: Researchers use visualizations to analyze market behavior, test hypotheses, and present their findings in a comprehensible manner.

Educational Tools: Stock visualizations are valuable educational tools for teaching concepts of finance, trading, and investment strategies to students and novice investors.

3.6.7. Corporate Finance

Stock Performance Monitoring: Corporations monitor their own stock performance and that of competitors to make strategic business decisions.
Investor Relations: Visualizations are used in reports and presentations to communicate financial performance and market position to shareholders and investors.

3.6.8. Regulatory and Compliance

Market Surveillance: Regulatory bodies use visualizations to monitor market activity, detect anomalies, and ensure compliance with trading regulations.
Compliance Reporting: Visual tools help in generating compliance reports and visualizing adherence to financial regulations.

3.6.9. Financial Journalism and Media

Market Analysis Reports: Financial journalists and media outlets use stock visualizations to illustrate market analyses, trends, and news stories.
Interactive Charts: Media websites provide interactive stock charts and tools for their audience to explore market data.

3.6.10 Technologies and Tools

Programming Languages: Python, R, and JavaScript (with libraries like D3.js, Plotly, and Highcharts) are widely used for creating advanced stock visualizations.
Data Sources: Financial data APIs (such as Alpha Vantage, Yahoo Finance, Quandl) provide the necessary data for visualization.
Software and Platforms: Tools like Tableau, Power BI, and custom-built dashboards allow for interactive and real-time visualizations.
Machine Learning: Advanced models like LSTM (Long Short-Term Memory) networks are employed for predictive analysis and visualization of future stock trends.

3.7 LIMITATIONS:

While stock visualization offers numerous benefits, it also has several limitations. Understanding these limitations is crucial for users to interpret the visualized data accurately and make informed decisions. Here are some key limitations:

3.7.1. Data Quality and Accuracy

Reliance on Accurate Data: Visualizations are only as good as the data they are based on. Inaccurate, incomplete, or outdated data can lead to misleading visualizations.

Data Source Reliability: The reliability of data sources (APIs, financial feeds) is critical. Any discrepancies or errors in these sources can compromise the integrity of the visualizations.

3.7.2. Complexity of Financial Markets

Oversimplification: Visualizations might oversimplify complex market dynamics, omitting critical factors that could influence stock prices.

Limited Scope: Most visualizations focus on specific aspects (e.g., price, volume) and may not capture the full context, such as macroeconomic indicators, geopolitical events, or market sentiment.

3.7.3. Interpreting Visualizations

Misinterpretation Risk: Users may misinterpret visual data due to a lack of expertise or understanding of financial concepts and chart patterns.

Cognitive Bias: Visualizations can reinforce cognitive biases, such as seeing patterns where none exist (pattern recognition bias) or confirmation bias (favoring information that confirms preexisting beliefs).

3.7.4. Technical Limitations

Scalability Issues: Handling and visualizing large datasets in real-time can be challenging, especially with high-frequency trading data.

Performance: Complex visualizations, especially interactive ones, can be resource-intensive and may suffer from performance issues on less powerful devices.

3.7.5. Predictive Limitations

Model Limitations: Predictive models, including LSTM networks, have their own limitations and may not always provide accurate forecasts. Market conditions can change rapidly, making predictions obsolete.

Historical Dependence: Predictive visualizations often rely on historical data, which may not always be a reliable indicator of future performance due to unprecedented market events.

3.7.6. User Dependency

Skill Requirement: Effective use of stock visualizations often requires a certain level of expertise in both financial markets and the specific visualization tools.

Customization Needs: Users may need to customize visualizations to suit their specific needs, which can require significant time and effort, especially for complex analyses.

3.7.7. Limited Real-Time Capabilities

Latency: Real-time data visualizations can suffer from latency issues, where there is a delay between the actual market event and its representation on the visualization.

Data Streaming Challenges: Continuously streaming and updating data in real-time poses technical challenges, including maintaining a stable data feed and ensuring synchronization.

3.7.8. Contextual Limitations

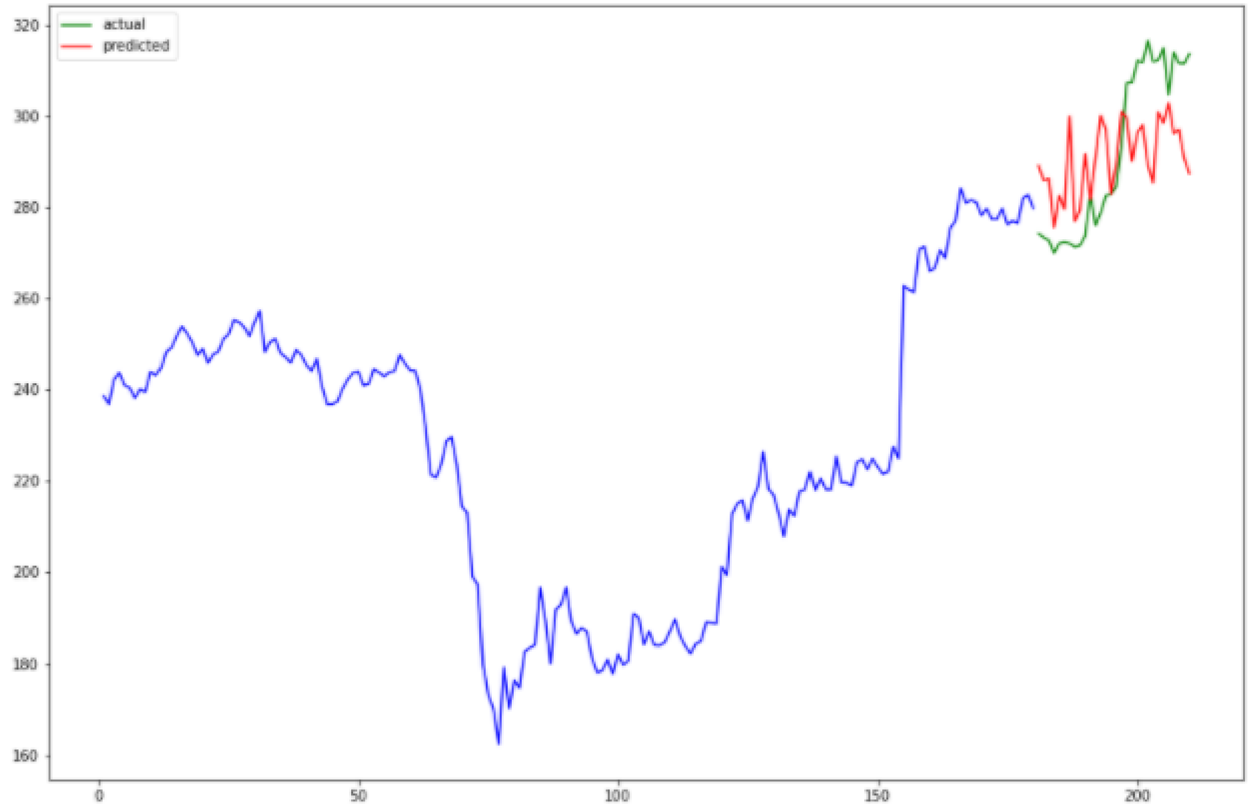
Lack of Context: Visualizations often focus on quantitative data and may not provide the qualitative context necessary for fully understanding market movements (e.g., news events, regulatory changes).

One-Dimensional Views: Many visualizations provide a one-dimensional view, focusing on a single metric or aspect, which might not give a comprehensive picture of the market scenario.

3.7.9. Aesthetic and Design Issues

Overload: Too much information in one visualization can overwhelm users, leading to analysis paralysis.

Design Flaws: Poorly designed visualizations can mislead users. For example, improper scaling of axes or misleading use of colors can distort the perceived information.



3.8 PROPOSED METHODOLOGY:

This study aims to predict stock prices using machine learning algorithms based on news articles. The methodology used for this research is outlined below:

Test Environment: The tools and technologies employed in this study, including the programming languages,
libraries, and frameworks used for implementing the machine learning algorithms and processing the data, are discussed in this section.

Data Collection and Preprocessing: The process of collecting and preprocessing the data, which is crucial for
the accuracy of the predictions, is detailed in this section. The source of the data, such as financial market data
and news articles, is described, and any necessary data cleaning or normalization procedures are outlined.

Feature Extraction: Relevant features are extracted from the collected data, which is a crucial step in developing accurate prediction models. The features used for predicting stock prices based on news articles are identified and extracted in this section. This may involve text analysis techniques, sentiment analysis, and other relevant techniques to convert the news articles into meaningful features for the machine learning algorithms.

Machine Learning Algorithms: The machine learning algorithms used for predicting stock prices are implemented and discussed in this section. This may include supervised learning algorithms such as regression or classification algorithms, as well as other relevant machine learning techniques.

Evaluation and Validation: The accuracy and effectiveness of the developed prediction models are evaluated and validated in this section. This may involve using performance metrics such as accuracy, precision, recall, and F1-score, as well as cross-validation techniques to ensure the reliability of the results.

Comparison and Analysis: The predicted stock prices using news articles are compared with the predicted stock prices without considering news articles to analyze the impact of news articles on stock price predictions.

Any insights or observations from the comparison are discussed in this section.

Interpretation of Results: The results of the study are interpreted and discussed in the context of the research objectives and hypotheses. Any limitations or challenges encountered during the research process are also highlighted, and recommendations for future research are provided.

Conclusion: The conclusion summarizes the findings of the study and provides a final assessment of the effectiveness of the proposed methodology in predicting stock prices based on news articles

Data Collection: Collect relevant financial market data, including historical stock prices and news articles, from reliable sources.

Data Preprocessing: Clean and preprocess the collected data to remove irrelevant or noisy information. This may involve data cleaning, normalization, and feature extraction techniques.

Feature Extraction: Extract relevant features from the news articles using text analysis techniques, sentiment analysis, and other relevant techniques, to convert the textual data into meaningful features for machine learning algorithms.

Feature Selection: Select the most important and relevant features from the extracted features to reduce the dimensionality of the data and improve the efficiency of the machine learning algorithms.

LSTM Model Training: Train an LSTM-based machine learning model using the preprocessed and selected features. LSTM is a type of recurrent neural network (RNN) that can capture long-term dependencies in sequential data, making it suitable for time-series data like stock prices.

Model Evaluation: Evaluate the performance of the trained LSTM model using appropriate performance metrics such as accuracy, precision, recall, and F1-score. This may also involve using cross-validation techniques to ensure the reliability of the results.

Model Optimization: Optimize the trained LSTM model by tuning hyperparameters, adjusting model parameters, or using ensemble techniques to improve the accuracy and effectiveness of the predictions

Prediction: Use the trained and optimized LSTM model to predict stock prices based on new news articles.

This may involve inputting new news articles into the trained LSTM model and obtaining the predicted stock prices as the output.

Comparison and Analysis: Compare the predicted stock prices based on news articles with the actual stock prices and analyze the results to understand the effectiveness of the proposed algorithm in predicting stock prices based on news articles.

Interpretation of Results: Interpret and discuss the results obtained from the predicted stock prices in the context of the research objectives and hypotheses. Analyze any insights or observations from the comparison.

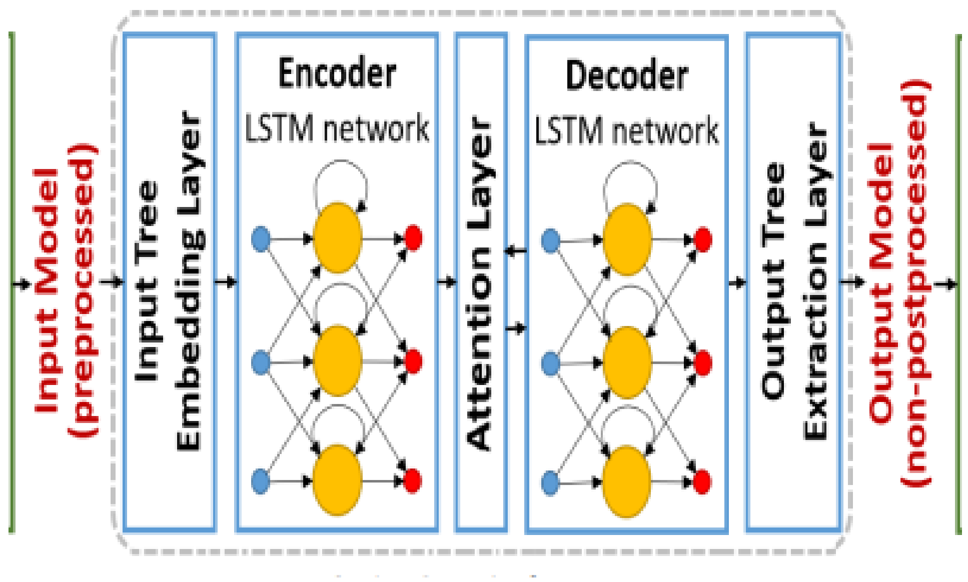


Fig. 3.8 LSTM Architecture

CHAPTER 4
SYSTEM IMPLEMENTATION

4.1 INTRODUCTION

Design is a multi- step that focuses on data structure software architecture, procedural details, procedure etc… and interface among modules. The design procedure also decodes the requirements into presentation of software that can be accessed for excellence before coding begins. Computer software design change continuously as novel methods; improved analysis and border understanding evolved. Software proposal is at relatively primary stage in its revolution.

Therefore, software design methodology lacks the depth, flexibility and quantitative nature that are usually associated with more conventional engineering disciplines. However, methods for software designs do exit, criteria for design qualities are existing and design notation can be applied.

4.2 Problems with Existing System

4.2.1 Data Complexity:

First, compile and clean historical data on stock prices from reliable sources, such as financial databases or application programming interfaces. use libraries like Pandas in Python to manage and preprocess the data. You may need to address issues like missing data, outliers, and data quality.

4.2.2 Pattern Identification:

Utilize various data analysis and visualization techniques to identify patterns in historical stock price data. Common tools for this include moving averages, technical indicators(e.g.,MACD,RSI),and candlestick charts. Pattern recognition may also be accomplished using machine learning technologies, such as time series analysis or deep learning (e.g., LSTM).

4.2.3 Predictive Accuracy:

Choose applicable machine learning algorithms, create a model, then tweak the model's hyperparameters to improve projected accuracy. Make use of metrics like Mean Absolute

Error (MAE), Mean Squared Error (MSE), or Root Mean Square Error (RMSE) to assess performance. To evaluate the generalizability of the model, do cross validation.

4.2.4 Feature Selection:

Pick pertinent aspects with care which can help you forecast changes in stock price. To help with feature selection, apply methods such as feature importance analysis, correlation analysis, and domain expertise. When necessary, further dimensionality reduction methods like as Principal Component Analysis (PCA) might be used.

4.2.5 Dynamic Market Conditions:

Recognize that stock market is impacted by a wide range of factors, including

economic indicators, news, geopolitical events, and market sentiment. Consider incorporating external data sources, sentiment analysis, and news sentiment analysis into your models to account for dynamic market conditions. Regularly update your models and retrain them to adapt to changing market dynamics

4.3 Proposed System:

4.3.1.Data Collection and Preprocessing:

Data Sources: Find and use trustworthy data sources to learn about stock market. This may include previous stock price data, financial statements, news feeds, and economic indicators. Consider using APIs or scraping data from financial websites.

Data Cleaning: When necessary, further dimensionality reduction methods like as Principal

Component Analysis (PCA) might be used. For data cleaning and manipulation, you may want to utilize programs like Pandas or NumPy.

Data Storage: Organize and store the data efficiently, possibly in a database or data warehouse, to enable quick access and retrieval.

4.3.2 Visualization Module:

Data Visualization : Develop a visualization module to help users understand market trends and data patterns. Utilize libraries like Matplotlib, Plotly. JavaScript to generate dynamic graphs and charts. Key Visualizations: To show stock price data, technical indicators, and other pertinent information, think about using heatmaps, line charts, bar charts, and candlestick charts.

Interactive Features: Allow users to customize charts, change time frames, and overlay various technical indicators for a more comprehensive analysis.

4.3.3 Pattern Recognition:

Implement Algorithms: Utilize machine learning or statistical algorithms for pattern recognition in stock price data. For time series analysis, common techniques

include RSI, Bollinger Bands, MACD, moving averages, and machine learning models like LSTM.

Real-time Analysis: Incorporate real-time data analysis to identify patterns and trends as they occur, enabling timely decision-making for traders.

4.3.4 User Interface:

Design: Create an intuitive and user-friendly interface that allows investors and traders to interact with the system easily. Pay attention to the design concepts for

user experience (UX).

Features: Include features like search functions, customizable dashboards, news feeds, and alerts. When particular circumstances are satisfied, users ought to be allowed to choose the criteria for notifications.

User Profiles: Consider allowing users to create profiles, save preferences, and track their portfolios and trading history.

4.4 SYSTEM REQUIREMENTS

4.4.1 Hardware Requirements

PROCESSOR: PENTIUM IV

- RAM: 8 GB
- PROCESSOR: 2.4 GHZ
- MAIN MEMORY: 8GB RAM
- PROCESSING SPEED: 600 MHZ
- HARD DISK DRIVE: 1TB
- KEYBOARD :104 KEYS

4.4.2 Software Requirements

- ➢ FRONT END: PYTHON
- ➢ IDE: ANACONDA

4.5 ARCHITECTURE

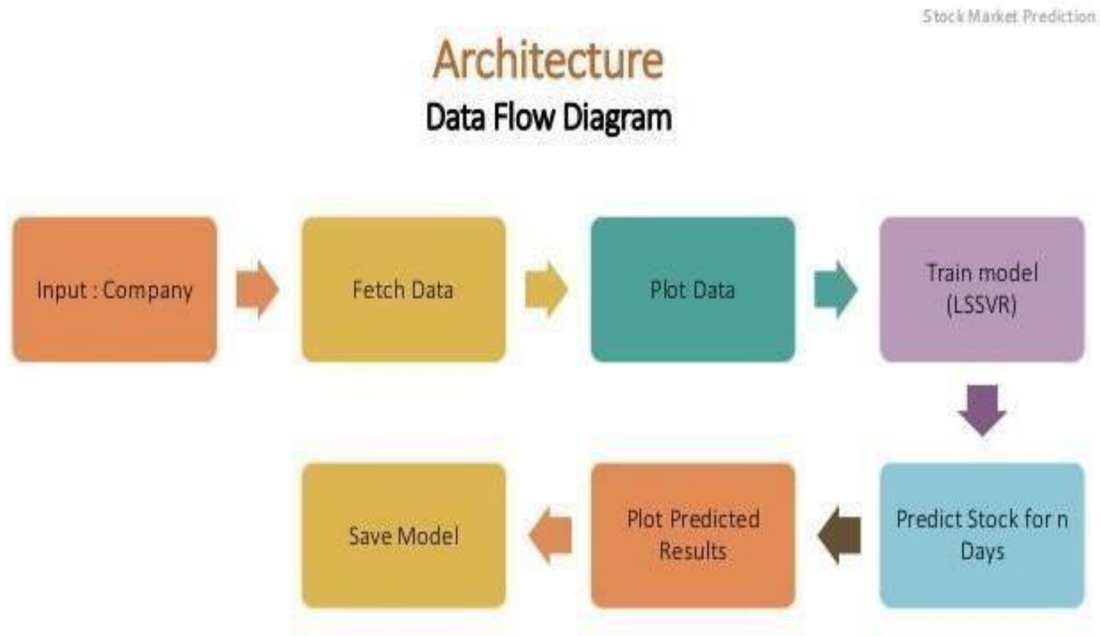


Fig 4.2.1 Data Flow Diagram

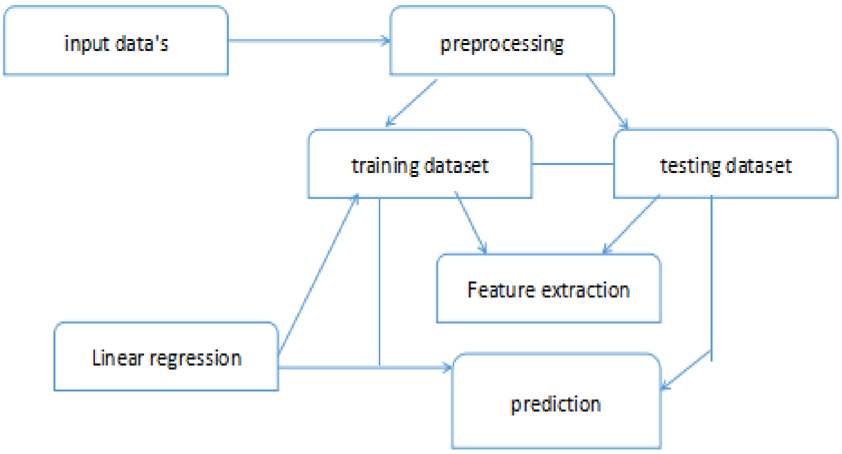


Fig 4.2.2 Architecture Design

4.6 MODULE DECRIPTION

The implementation of this project is divided into following steps

1. Data Preprocessing
2. Feature selection
3. Building and Training Model

4.6.1 Data Preprocessing:

The entries are present in the dataset. The null values are removed using df = df.dropna() where df is the data frame. The categorical attributes (Date,High,Low,Close,Adj value) are converted into numeric using Label Encoder. The date attribute is splitted into new attributes like total which can be used as feature for the model.

4.6.2 Feature selection:

Features selection is done which can be used to build the model. The attributes used for feature selection are Date,Price,Adj close,Forecast X coordinate , Y coordinate, Latitude , Longitude, Hour and month.

Select the most important and relevant features from the extracted features to reduce the dimensionality of the data and improve the efficiency of the machine learning algorithms

4.6.3 Building and Training Model:

After feature selection location and month attribute are used for training. The dataset is divided into pair of xtrain ,ytrain and xtest, ytest. The algorithms model is imported form skleran. Building model is done using model. Fit (xtrain, ytrain). This phase would involve supervised classification methods like linear regression, Ensemble classifiers (like Adaboost, Random Forest Classifiers), etc.

LSTM Model Training: Train an LSTM-based machine learning model using the preprocessed and selected features. LSTM is a type of recurrent neural network (RNN) that can capture long-term dependencies in sequential data, making it suitable for time-series data like stock prices

4.6.4 Model Evaluation:

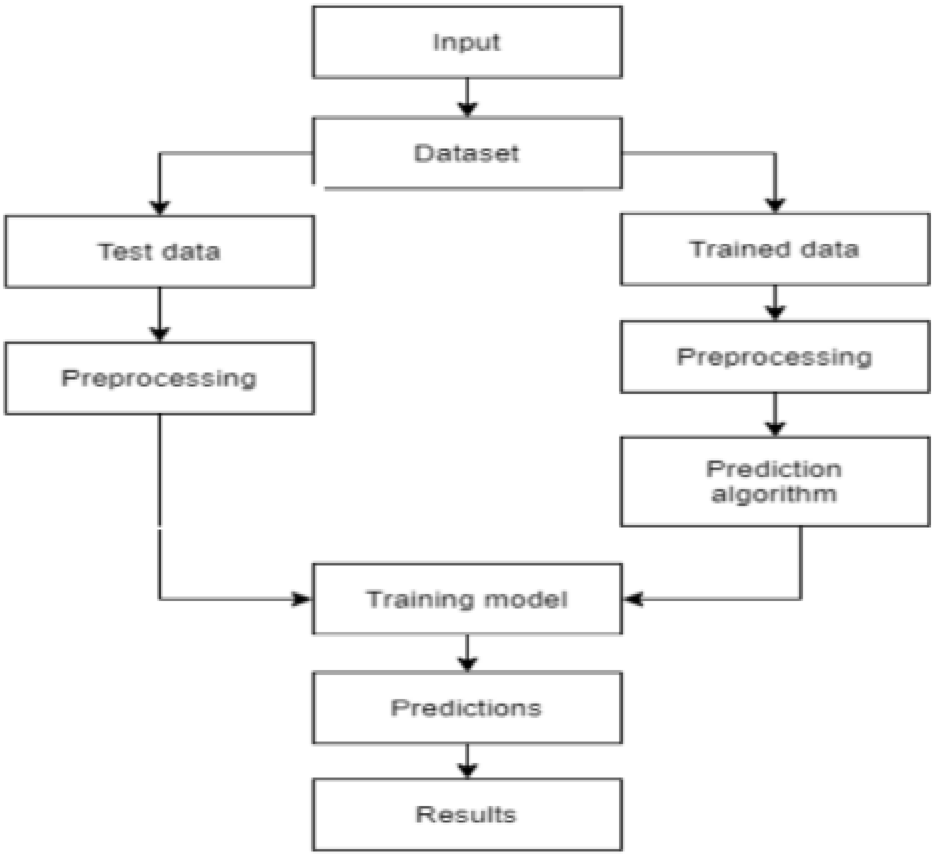
Evaluate the performance of the trained LSTM model using appropriate performance metrics such as accuracy, precision, recall, and F1-score. This may also involve using cross-validation techniques to ensure the reliability of the results

4.6.5 Model Optimization:

Optimize the trained LSTM model by tuning hyperparameters, adjusting model parameters, or using ensemble techniques to improve the accuracy and effectiveness of the predictions

4.6.6 Prediction:

Use the trained and optimized LSTM model to predict stock prices based on new news articles. This may involve inputting new news articles into the trained LSTM model and obtaining the predicted stock prices as the output



Training and prediction

4.7 PYTHON TECHNOLOGY

Python is an interpreted, object- oriented programming language similar to PERL, that has gained popularity because of its clear syntax and readability. Python is said to be relatively easy to learn and portable, meaning its statements can be interpreted in a number of operating systems, including UNIX- based systems, Mac OS, MS- DOS, OS/2, and various versions of Microsoft Windows 98. Python was created by Guido van Rossum, a former resident of the Netherlands, whose favourite comedy group at the time was Monty Python's Flying Circus. The source code is freely available and open for modification and reuse. Python has a significantnumber of users.

A notable feature of Python is its indenting of source statements to make the code easier to read. Python offers dynamic data type, ready- made class, and interfacesto many system calls and libraries. It can be extended, using the C or C++language.

Python can be used as the script in Microsoft's Active Server Page (ASP) technology. The scoreboard system for the Melbourne (Australia) Cricket Ground is written in Python. Z Object Publishing Environment, a popular Web application server, is also written in the Python language's

1. Python [19] as a language has a vast community behind it. Any problems which may be faced is simply resolved with visit to Stack Overflow. Python is the foremost standard language on the positioning that makes it is very straight answer to any question.

2. Python [19] is an abundance of powerful tools ready for scientific computing Packages. The packages like NumPy, Pandas and SciPy area unit freely available and well documented. These Packages will intensely scale back, and variation the code necessary to write a given program. This makes repetition fast.

3. Python is a language as [19] forgiving and permits for the program that appear as if pseudo code. This can be helpful once pseudo code give in tutorial papers should be required and verified. Using python this step is sometimes fairly trivial.

However, Python is [19] not without its errors. The python is dynamically written language and packages are area unit infamous for Duck writing. This may be frustrating once a package

technique returns one thing that, for instance, looks like an array instead of being an actual array. Plus the standard Python documentation did not clearly state the return type of a method,

this can’t lead without a lot of trials and error testing otherwise happen in a powerfully written
language. This is a problem that produces learning to use a replacement Python package or library more difficult than it otherwise may be.

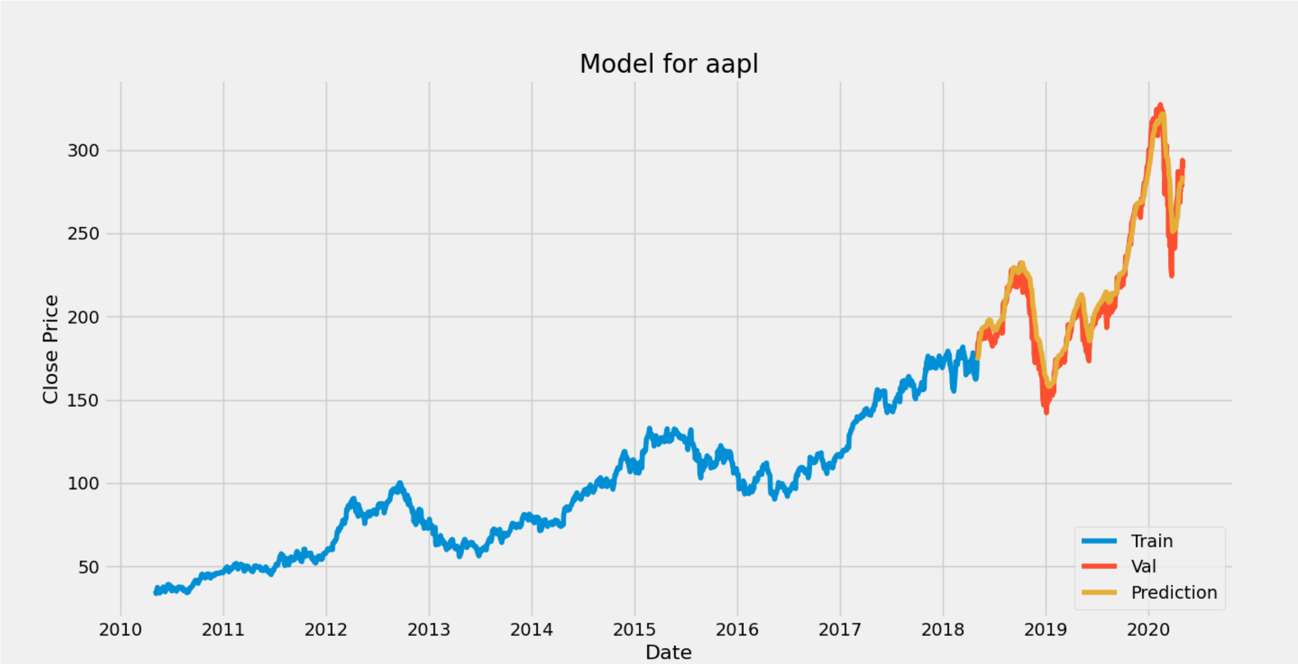
4.7.1 Python Platform

Apart from Windows, Linux and MacOS, CPython implementation runs on 21 different **platforms**. IronPython is a .NET framework based **Python** implementation and it is cabable of running in both Windows, Linux and in other environments where .NET framework is available.

4.7.2 Python Library

Machine Learning, as the name suggests, is the science of programming a computerby which they are able to learn from different kinds of data. A more general definitiongiven by Arthur Samuel is –“Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed.” They are typically used to solve various types of life problems.
In the older days, people used to perform Machine Learning tasks by manually coding all the algorithms and mathematical and statistical formula. This made the process time consuming, tedious and inefficient. But in the modern days, it isbecome very much easy and efficient compared to the olden days by various python libraries, frameworks, and modules. Today, Python is one of the most popularprogramming languages for this task and it has replaced many languages in the industry, one of the reason is its vast collection of libraries. Python libraries that usedin Machine Learning are:

- Numpy
- Scikit-learn
- TensorFlow
- Keras
- Pandas
- Matplotlib



4.7.2.1 NumPy

NumPy is a very popular python library for large multi- dimensional array and matrix processing, with the help of a large collection of high- level mathematical functions. Itis very useful for fundamental scientific computations in Machine Learning. It is particularly useful for linear algebra, Fourier transform, and random number capabilities. High- end libraries like TensorFlow uses NumPy internally formanipulation of Tensors.

4.7.2.2 Scikit-Learn:

Skikit- learn is one of the most popular ML libraries for classical ML algorithms. It is built on top of two basic Python libraries, viz., NumPy and SciPy. Scikit- learn supports most of the supervised and unsupervised learning algorithms. Scikit- learn can also be used for data-mining and data- analysis, which makes it a great tool who is starting out with ML.

Scikit-learn [21] could be a free machine learning library for Python. It features numerous classification, clustering and regression algorithms like random forests, k-neighbours, support vector machine, and it furthermore supports Python scientific and numerical libraries like SciPy and NumPy. In Python Scikit-learn is specifically written, with the core algorithms written in Cython to get the performance. Support vector machines are enforced by a Cython wrapper around LIBSVM .i.e., linear support vector machines and logistic regression by a similar wrapper around LIBLINEAR.

4.7.2.3 TensorFlow:

TensorFlow is a very popular open- source library for high performance numerical computation developed by the Google Brain team in Google. As the name suggests, Tensorflow is a framework that involves defining and running computations involving tensors. It can train and run deep neural networks that can be used to develop several AI applications. TensorFlow is widely used in the field of deep learning research and application.

In the TensorFlow [22] has an open source software library for numerical computation using data flow graphs. Inside the graph nodes represent mathematical formulae, the edges of graph represent the multidimensional knowledge arrays (tensors) communicated between them. The versatile architecture permits to deploy the computation to at least one or many GPUs or CPUs

in a desktop, mobile device, servers with a single API. TensorFlow was firstly developing by engineers and researchers acting on the Google Brain Team at intervals Google's Machine Intelligence analysis organization for the needs of conducting deep neural networks research and machine learning, but, the system is generally enough to be appropriate in a wide range of alternate domains as well.

Google Brain's second-generation system is TensorFlow. Whereas the reference implementation runs on single devices, TensorFlow can run on multiple GPUs and CPUs. TensorFlow is offered on Windows, macOS, 64-bit Linux and mobile computing platforms together with iOS and Android.

4.7.2.4 Keras:

Keras is a very popular Machine Learning library for Python. It is a high- level neural networks API capable of running on top of TensorFlow, CNTK, or Theano. It can run seamlessly on both CPU and GPU. Keras makes it really for ML beginners to build and design a Neural Network. One of the best thing about Keras is that it allows for easy and fast prototyping.

Keras is [23] a high-level neural networks API, it is written in Python and also capable of running on top of the Theano, CNTK, or. TensorFlow. It was developed with attention on enabling quick experimentation. having the ability to travel from plan to result with the smallest amount doable delay is key to doing great research.Keras permits for straightforward and quick prototyping (through user-friendliness, modularity, and extensibility). Supports each recurrent networks and convolutional networks, also as combinations of the 2. Runs seamlessly on GPU and CPU. The library contains numerous implementations of generally used neural network building blocks like optimizers, activation functions, layers, objectives and a number of tools to create operating with text and image data easier. The code is hosted on GitHub, and community support forums embody the GitHub issues page, a Gitter channel and a Slack channel.

4.7.2.5 Pandas:

Pandas is a popular Python library for data analysis. It is not directly related to Machine Learning. As we know that the dataset must be prepared before training. In this case, Pandas comes handy as it was developed specifically for data extraction and preparation. It provides high- level data structures and wide variety tools for data analysis. It provides many inbuilt methods for groping, combining and filtering data.

4.7.2.6 Matpoltlib:

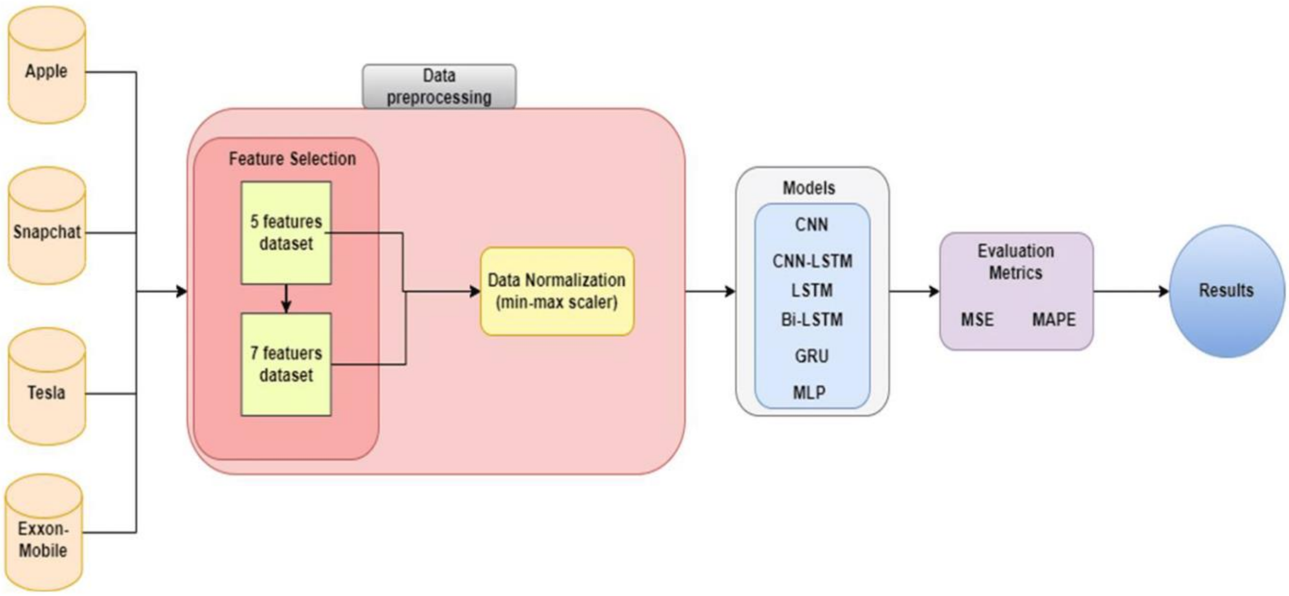
Matpoltlib is a very popular Python library for data visualization. Like Pandas, it is not directly related to Machine Learning. It particularly comes in handy when a programmer wants to visualize the patterns in the data. It is a 2D plotting libraryused for creating 2D graphs and plots. A module named pyplot makes it easy for programmers for plotting as it provides features to control line styles, font properties, formatting axes, etc. It provides various kinds of graphs and plots for data visualization, viz., histogram, error charts, bar chats, etc,

4.7.2.7 COMPILER OPTION

Anaconda is [19] free premium open-source distribution of the R and Python programming languages for scientific computing, predictive analytics, and large-scale process that aim is to modify package managing and deployment. Package versions unit managed by the package management system conda.

4.7.2.8 JUPITER NOTEBOOK

The Jupyter Notebook is an open-source web application that enables to making and sharing documents that contain visualizations, narrative text, live code and equations. Uses include: data , data visualization, data transformation, statistical modelling, machine learning, numerical simulation, data cleaning and much more [24].



4.8 Hybrid Approach of LSTM + CNN

In the hybrid approach, the Convolutional Neural Networks (CNNs) offer benefits in choosing
sensible options and Long Short-Term Memory (LSTM) networks have proven sensible skills to find out to learn sequential data. Each approaches are reported to produce improved result. CNNs to possess to convolute filters over every input layer so as to get the simple options and
CNNs have shown enhancements in computer vision, natural language processing and different tasks. CNN may be a powerful tool to pick out features in order to improve the prediction accuracy The capabilities of LSTMs in learning data series by considering the previous outputs The multiple convolutional filters slide over the matrix to produce a new feature map and also the filters have numerous completely different sizes to generate different features. The Maxpooling layer is to calculate the most value as a corresponding

feature to a particular filter. The output vectors of the Max-pooling layer become inputs to the LSTM networks to measure the long-run dependencies of feature sequences. One in all the benefits of the LSTMs is that the ability to capture the sequential data by considering the previous data. This layer takes the output vectors from the dropout layer as inputs. This layer include a set number of units or cells and also the input of every cell is that the output from the dropout layer. The final output of this layer has the same number of units within the network the outputs from LSTMs are merged and combined in one matrix then passed to a fully connected layer. The array is converted into a single output in the range between 0 and 1 using the fully connected layer, in order to be finally classified using sigmoid function.

CHAPTER 5
DATA AND COMPARISION

The dataset used in the proposed methodology for predicting stock prices based on news articles using LSTM may include historical stock prices and news articles from reliable sources. The historical stock prices can be obtained from stock market data sources such as NASDAQ or other financial data providers. The stock prices should cover a sufficient time period, such as the last five years, to capture the trends and patterns in stock price movements. The news articles can be collected from reputable news sources that cover financial news, such as financial news websites, business news portals, or financial publications. The news articles should be relevant to the stocks being predicted and should cover a wide range of topics that can impact stock prices, such as company announcements, earnings reports, economic indicators, and geopolitical events. The news articles should be in a textual format, such as plain text or structured text, and may include features such as headline, article content, publication date, and source. The dataset should be properly cleaned and preprocessed to remove any irrelevant or noisy information, and may also involve techniques such as text analysis, sentiment analysis, and feature extraction to convert the textual data into meaningful features for machine learning algorithms.
It is important to ensure the reliability and quality of the dataset by validating the sources of the news articles and ensuring that the historical stock prices are accurate and complete. Additionally, the dataset should be properly split into training, validation, and testing sets to train, optimize, and evaluate the LSTM model effectively.

The data used for this work consists of stocks like Amazon(TWR) Google (FB) Google(GOOGL) and Apple(APPL). Data includes the daily opening and closing prices of stocks and were extracted from yahoo finance which covers the stock price period from 01/02/2017 to 11/23/2021.

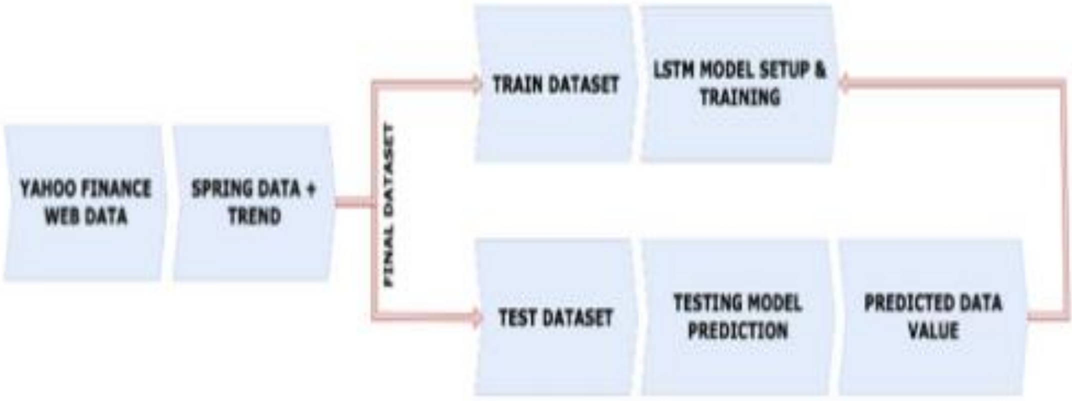


Fig. 5.1 A flowchart for stock market prediction using an LSTM model

Fig. 1 represents the flowchart of the method we used for the stock price prediction. We looked at the stock price data and its trend. For example, depicted in Fig. 3 are the values of the stocks and the trading volume, respectively.

Also, Fig. 4 compares the market cap of two stocks, the multiplication of trading volume, and the trading price for the specified date. The dataset was then split into two datasets: the training and testing datasets, in a ratio of 80:20 for stock data (GOOGL and AMZN etc). Further, the dataset was transferred into tensors, Some values, such as input dimensions, hidden layer dimensions, and output dimensions, were defined regarding the layers of the neural network. An LSTM model is created with a Mean Square Error (MSE) as a Loss function. The mean square error (MSE) is the average square difference between the model's predictions and the ground truth taken across the

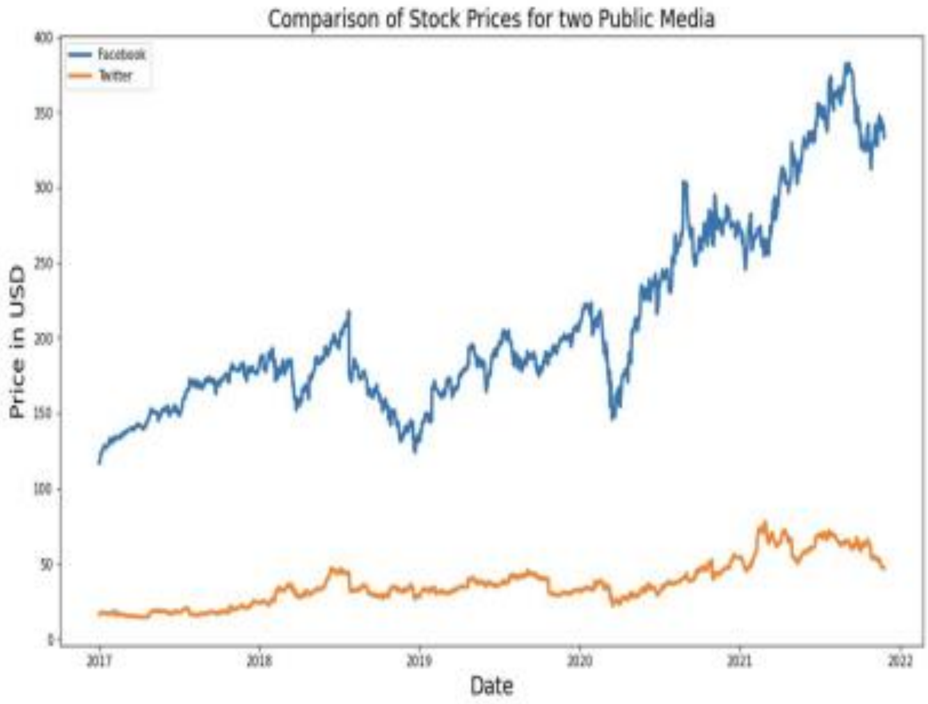


Fig. 5.2 Stock price of Amazon and Google

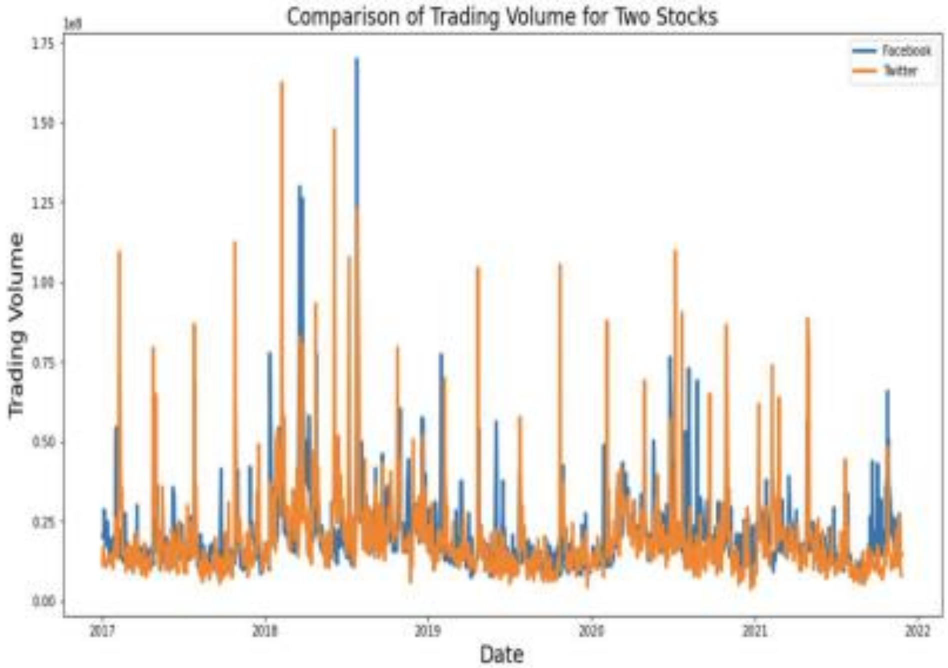


Fig. 3 Trading volume of two stocks: Amazon and Google, which is the sum of selling volume and buying volume for a certain specified time

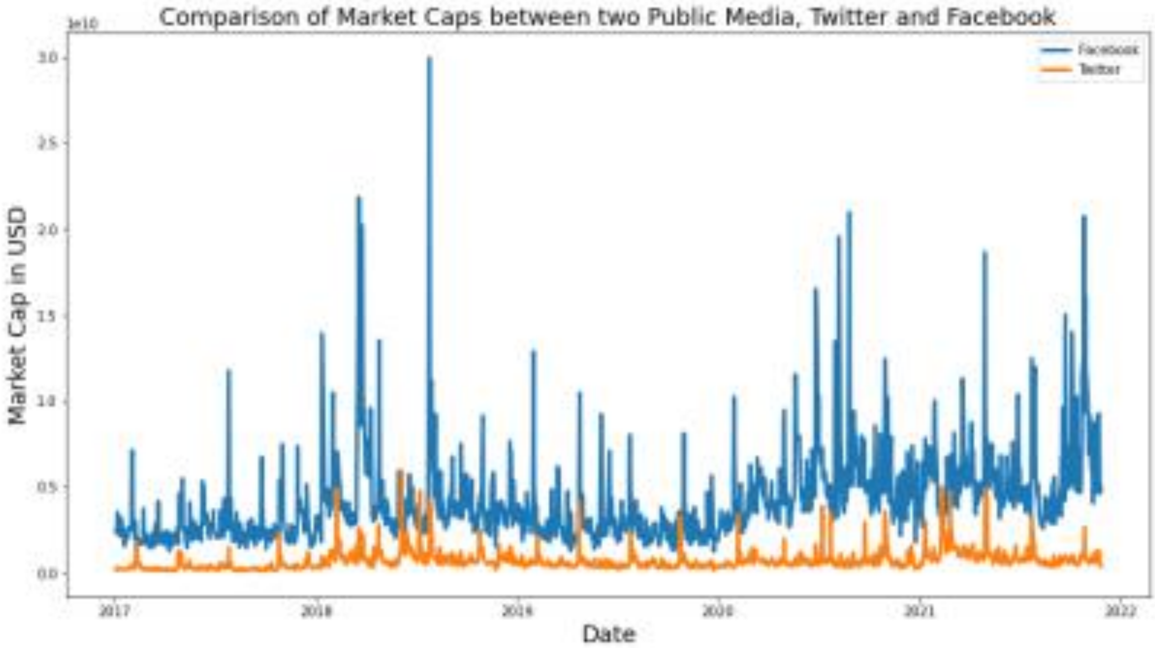


Fig. 5.3 Comparison of the market cap between Amazon and Google stocks dataset.

MSE = 1 n Xn i=1 (yi − yˆi)^2,
where yi , ˆyi , and n are the model's prediction, ground truth, and the number of samples we are testing against. We normalized stock prices by using min-max normalization, which performs the linear transformation in the data for each stock before implementing the model. Normalization maps the values of price columns in the range [0,1] without affecting the differences in the whole range of original data values. Mathematically, the min-max normalization can be expressed as ,
Pnorm = (P – Pmin) / (Pmax − Pmin)where P is the data value, Pmin, and Pmax are the minimum and the maximum value in the dataset. Adam is used as an optimizer, and finally, the model has been trained over 100 epochs. After finishing the training of the model, we applied the model for the prediction. Using these data, we analyzed the performance of an LSTM model.

5.1 Data Collection

Once the understanding of the objective is over, the next step is to collect the data. Data collection involves the understanding of initial observations of the data to identify the useful subsets from hypotheses of the hidden information. We use the data from Google finance.

5.1.1 Dataset Preparation

In this study, S&P 500, a popular US stock market index, is used for the model prediction. The process of feature selection includes identifying the core factors that contribute to the index value fluctuations. Some features, such as fundamental data and technical indicators, are directly extracted from the underlying index. Other factors, namely macroeconomic variables, are selected based on their potential impact on the overall economy and broader markets. A complete 15 years of data have been collected from 2006 to 2020. The time frame selection incorporates two major bear markets, the financial crisis in 2008 and the COVID-19 pandemic in 2020. Thus, the construction of the model, including both bear and bull market, resembles the overall market scenario and may lead to a better prediction. We start with a brief description of the features used in the proposed model. The closing price is predicted based on the fundamental trading data, macroeconomic data, and technical indicators of the underlying index. A combination of all the features from three different categories presented in Table 1 is input features. All the available features except Civilian Unemployment Rate and Consumer Sentiment Index are by default daily data. We have converted the monthly data to daily through the forward filling method to have uniformity among the variables.

5.1.2 Data Pre-processing: Data Wrangling

The data pre-processing stage involves all the activities to prepare the final dataset from the preparatory raw information. The data preparation tasks can be performed several times as there is no specific order. These tasks include the selection of a record, table, attribute and cleaning of data for modeling tools.

5.1.3 Data Processing: Data Training

In technical analysis investors use the auto regressive and moving average models to forecast the stock trends. Major steps involved here are identification, parameter estimation and forecasting. These steps are repeated until an appropriate model is identified for prediction.

5.1.4 Forecasting Results

The process of making predictions of the future by relying upon the past and present data is known as forecasting. Various prediction techniques are used by the stock analysts to evaluate the future stock trends value. Prediction also offers a significant standard for organizations that have a long-term perception of actions. We use ‘forecast’ package for predicting the future stock trends based on the analysis of past trends. This ‘forecast’ package provides a number of forecasting functions for displaying the time series predictions along with exponential smoothing and space models.

5.1.5 Plot Visualizations

Data visualization is a graphical representation of the numerical data. After forecasting the stock market trends we visualize the results for short-term investment assistance in-terms of line charts, candlesticks charts, bar charts, and histograms.

5.1.6 View and Analyze Results

Once after plotting the results in-terms of visualizations we can find out the correlations to get the short-term predictions. In the next section we provide some of the screen shots by which the investor can analyze and predict the future stock trends of a particular company at a specific time period. So the investors in the stock market can use this as assistance to sell/buy/hold a share.

5.2 Comparison of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM).

A recurrent neural network (RNN) is an artificial neural network that recognizes data’s sequential patterns to predict future scenarios. RNNs are analogous to human learning. Given that the architecture of RNN consists of many nodes’ connections and the temporal dynamic behavior, RNN

eat

ng

Ms

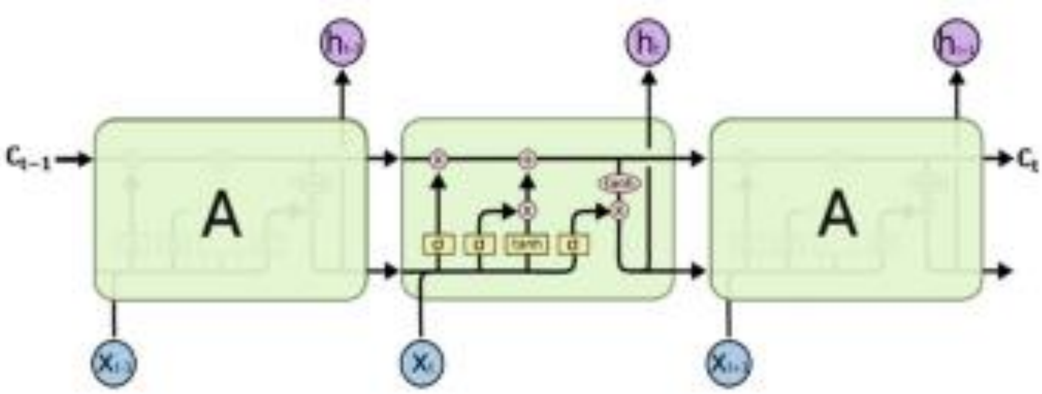
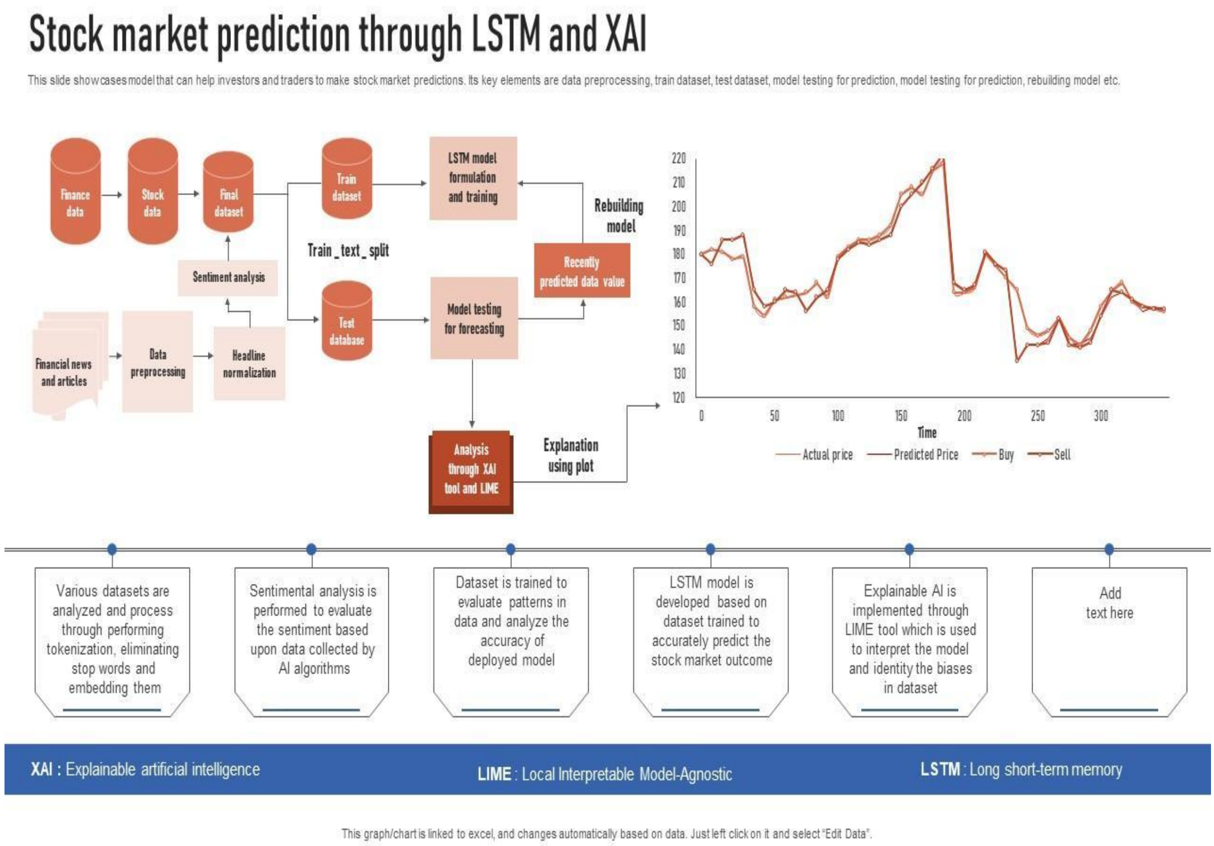


Fig. 1 An LSTM unit containing interacting layers, which has the input gate, output gate, and forget gate

Shown in Fig. 1 depicts an LSTM unit. The various components of the LSTM unit include a cell, an input gate, an output gate, and a forget gate. In such a system, the unit members normally remember values over time, and three gates regulate the flow of information into and out of the cell. So, LSTMs are RNNs suited for predicting stock prices due to their long-term remembrance of information.

RNNs are preferred for jobs that involve sequential data, such as language processing, time series analysis, and stock price prediction. In the case of stock price prediction, RNNs can consider the historical price data and use it to make predictions about future prices. When there exists a long-term dependency in sequential data, LSTM is useful. LSTM networks use a memory cell to store information about previous inputs and outputs, allowing them to remember important patterns in the data. This makes them particularly effective for predicting stock prices, which are known to exhibit complex patterns over time.

In stock price prediction, RNNs and LSTM networks are typically trained on historical price data and other relevant factors such as volume, market trends, and news sentiment. The trained model can then be used to predict future stock prices



CHAPTER 6

Sentimental Analysis and Real-time Visualization

To accurately predict variations in the stock price of specific companies, it is essential that one needs to have a strong knowledge of relevant stock market information, as several factors influence the stock market. In this work, we proposed a method to examine news sentiment as a follow-up to fundamental analysis. For sentiment prediction, we used the dataset from Kaggle. We used web crawling and HTML parsing through requests and beautiful soup to collect experimental data, as depicted in Fig. 6. The data is preprocessed by removing special characters, and URLs and converting text to lowercase, including the training dataset and JSON payload from the API provider for financial news. Updating internal vocabulary was done regarding a list of texts, and tokenization was used to identify the unique words in the text, updating the internal vocabulary based on a list of texts. Vectorization converts words in the text to a numeric vector using the doc2Vec() method with the Gensim model, designed to extract semantic concepts from documents and handle large amounts of text

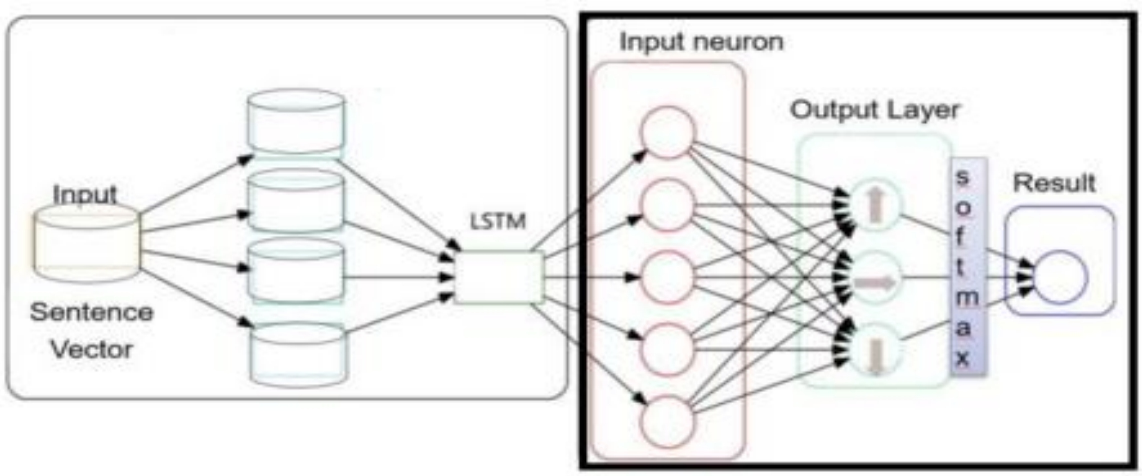


fig. 6.1 Sentimental Analysis Text based on Neural Networks

6.1 Definition:

Sentiment analysis, also known as opinion mining, involves the use of natural language processing (NLP) to determine the emotional tone behind a body of text. This can be applied to social media posts, customer reviews, and more.

Techniques:

Lexicon-based Methods: These rely on pre-defined lists of words associated with positive or negative sentiments.
Machine Learning: Algorithms are trained on labeled data to classify text. Common models include Naive Bayes, Support Vector Machines, and more recently, deep learning models like LSTM (Long Short-Term Memory) networks and transformers (e.g., BERT, GPT).
Implementation:

Data Collection: Gather text data from sources like Twitter, news articles, or customer feedback.
Preprocessing: Clean the data by removing stop words, punctuation, and performing stemming or lemmatization.
Model Training: Use labeled data to train your model. For real-time applications, pre-trained models can be fine-tuned on specific datasets.
Sentiment Classification: Apply the model to classify new incoming data in real-time.

6.2 Real-time Visualization

Definition:

Real-time visualization involves dynamically updating charts or graphs to reflect new data as it arrives. This is essential for monitoring trends and making quick decisions.

Techniques:

Dashboards: Tools like Tableau, Power BI, or custom web applications using JavaScript libraries (e.g., D3.js, Chart.js).
Stream Processing**:** Use technologies like Apache Kafka, Apache Flink, or AWS Kinesis to handle data streams.

Implementation:

Backend: Set up a data pipeline that processes and analyzes incoming data in real-time. This could involve streaming data from social media APIs using tools like Tweepy for Twitter.
Processing and Analysis: Use a combination of Python (with libraries like pandas and scikit-learn) or more scalable solutions like Apache Spark.
Visualization: Update visual elements in real-time using web sockets for live data feeds. Frameworks like Flask or Django can serve the data, while front-end libraries (React, Angular) handle the dynamic updates.

6.3 Sentiment Analysis:

Text Classification: Use Natural Language Processing (NLP) techniques to classify the sentiment of textual data into positive, negative, or neutral categories.

Sentiment Scoring: Assign sentiment scores to each piece of text. Aggregating these scores over time can create a sentiment index.

Feature Integration: Integrate sentiment scores with stock price data as additional features for the LSTM model.

6.4 LSTM Model Development:

Architecture Design**:** Design the LSTM network, including the number of layers, neurons per layer, dropout rates, etc.

Training and Validation: Train the model using a combination of historical stock prices and sentiment data. Validate its performance with unseen data to ensure it generalizes well.

Hyperparameter Tuning: Optimize hyperparameters for better model performance.

6.5 Real-Time Prediction and Visualization:

Real-Time Data Feed*:* Set up a pipeline to continuously fetch and process new stock price data and sentiment data.

Prediction Execution: Use the trained LSTM model to generate real-time stock price predictions.

***Visualization Tools*:** Develop dashboards using libraries like D3.js, Plotly, or frameworks like Dash and Streamlit to visualize real-time predictions.

6.6 Key features should include:

Live Charts: Displaying real-time stock price movements and predictions.
Sentiment Index: Showing the aggregated market sentiment in real-time.

***Alerts and Notifications*:** Triggering alerts based on specific prediction thresholds or sentiment changes.

6.7 Evaluation and Continuous Improvement:

***Model Performance Monitoring*:** Continuously monitor the prediction accuracy and update the model as needed.

***User Feedback*:** Gather feedback from end-users to improve the visualization and prediction tools.

6.8 Architecture:

***Data Ingestion*:** Use APIs to collect data in real-time. For example, connect to the Twitter API using Tweepy.
***Stream Processing*:** Use Kafka or similar tools to manage the flow of data.
Sentiment Analysis: Process the incoming data through your NLP model to classify the sentiment.
Data Storage: Store the results in a database like MongoDB, Elasticsearch, or a time-series database like InfluxDB.
Visualization Dashboard: Build a web application using frameworks like Flask/Django for the backend and React/D3.js for the frontend. Utilize web sockets for live updates.

Example Workflow:

Step 1: A tweet is posted and fetched by your system.
Step 2: The tweet is processed to clean and prepare the text.
Step 3: The pre-trained sentiment analysis model classifies the tweet's sentiment.
Step 4: The sentiment score is sent to a message broker (e.g., Kafka).
Step 5: The data is stored and simultaneously pushed to a real-time dashboard.
Step 6: The dashboard updates visual elements like graphs or heat maps to reflect the new sentiment data.

6.9 Tools and Technologies

Data Collection: Tweepy (Twitter API), Scrapy (web scraping), BeautifulSoup.
Data Processing: pandas, NLTK, spaCy, TensorFlow, PyTorch.
Stream Processing: Apache Kafka, Apache Flink, AWS Kinesis.
Database: MongoDB, Elasticsearch, InfluxDB.
Visualization: D3.js, Chart.js, Plotly, React, Angular.
Backend Frameworks: Flask, Django.
Front-end Libraries: React, Angular, WebSockets for real-time data updates.

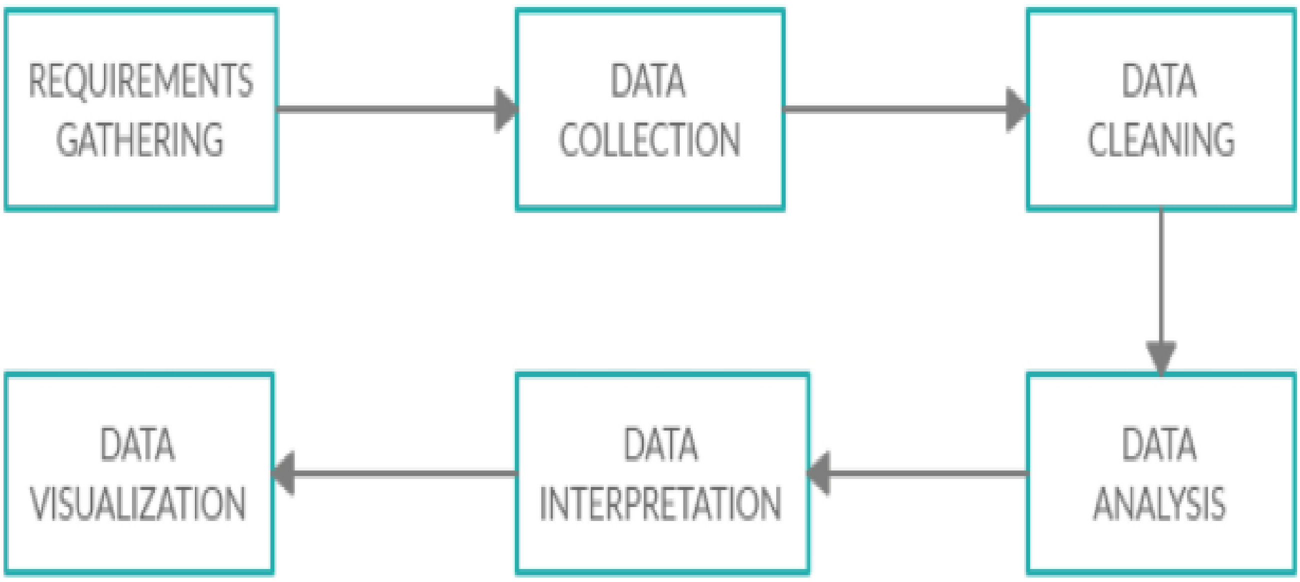


Fig. 6.9 Data Flow Model

CHAPTER 7
RESULTS AND DISCUSSION

7.1 RESULT

The proposed methodology for predicting stock prices based on news articles using LSTM can yield valuable insights and predictions. The LSTM model, trained on historical stock prices and corresponding news articles, can capture complex patterns and dependencies in the data, allowing for accurate predictions of stock prices The results obtained from the proposed methodology can be evaluated using various performance metrics, such as mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and accuracy. These metrics can provide quantitative measures of the prediction accuracy and model performance.

The discussion of the results may involve analyzing the performance of the LSTM model in different scenarios, such as during periods of market volatility or economic events, and comparing the predicted stock prices with the actual stock prices. The discussion may also involve identifying the key factors or news topics that have the most significant impact on stock price predictions, and exploring the potential implications and applications of the findings in the field of finance and investments . Furthermore, the limitations of the proposed methodology should be discussed, such as potential biases in the dataset, limitations of LSTM as a predictive model, and potential challenges in obtaining reliable and relevant news articles. Suggestions for further improvements or extensions of the methodology can also be discussed, such as incorporating additional features, using different machine learning algorithms, or exploring other data sources .Overall, the results and discussion of the proposed methodology can provide valuable insights into the effectiveness and limitations of using LSTM for predicting stock prices based on news articles, and contribute to the existing literature in the field of finance and machine learning

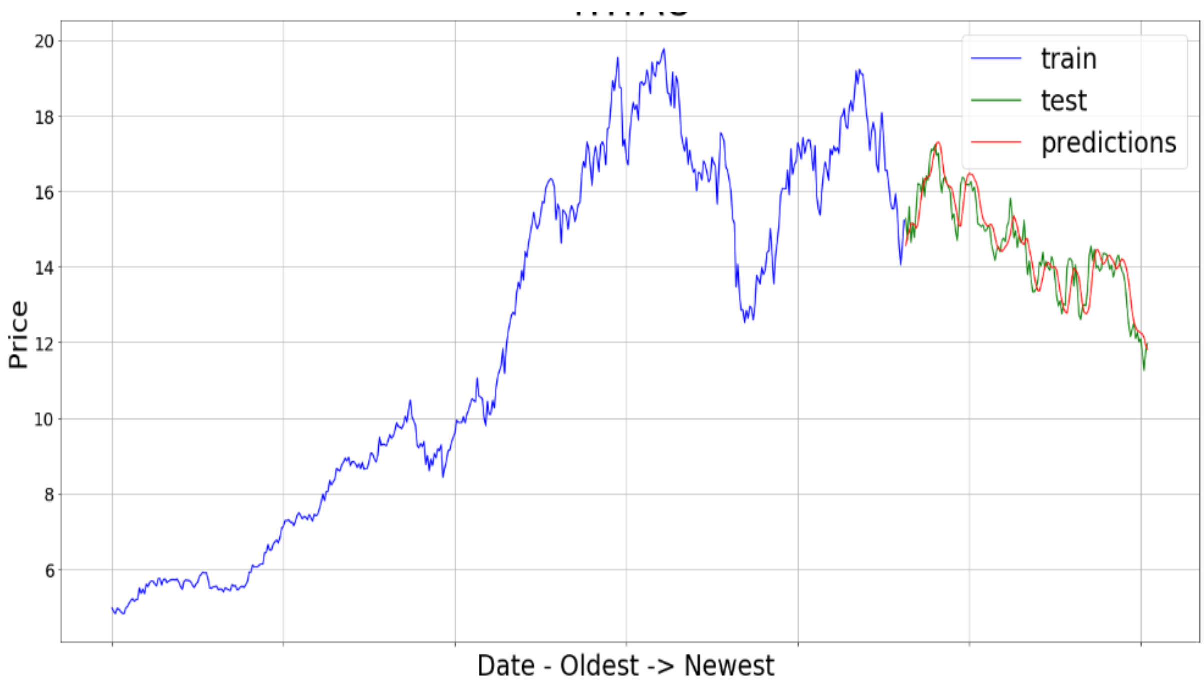
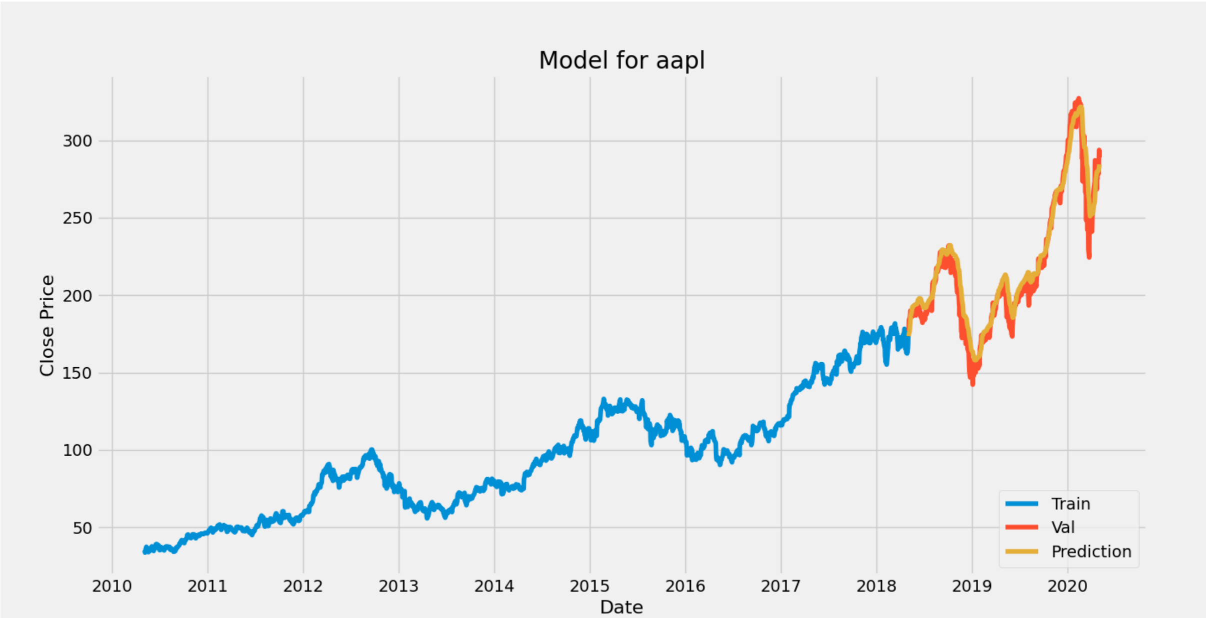


Fig. 7.1 Predction Chart

The data sets of Facebook and Twitter have been trained and tested on an LSTM model, which includes 32 hidden layers and only one output layer. The training has been done for 100 epochs. shows the loss during the training of Facebook and Twitter stocks for 100 epochs. It shows that the model behaves well with the training set .The model is probably overfitting, especially conside ring the minimal loss after the 40th epoch. When the model’s performance on the training set is much better than its performance on the validation or test set, it is an indication that the model may be overfitting, i.e., it has learned to fit the training data too closely and is not generalizing well to new, unseen data. After 40 epochs, the loss of the training set may continue to decrease. In contrast, the loss on the validation or test set may increase, indicating that the model is starting to memorize the training data rather than learning general patterns.

We can see the loss decreased gradually for 40 epochs, after which the loss is constant and has already reached minimal loss. This indicates that we do not require epochs of more than 40.It is possible that if the loss has reached its minimum value and has remained constant for several epochs, there may not be any significant improvements in the model’s performance with additional epochs. Statistically, the root mean square error (RMSE) for training and testing data is 5.91 and 24.40, respectively. From the results, we could see that the RMSE value is high for testing when compared to the training set. A higher RMSE value for the testing set than the training set indicates overfitting in machine learning models.



Overfitting occurs when a model is too complex and has learned to fit the training data too closely, including noise and irrelevant details, rather than learning the underlying patterns in the data. As a result, the model performs well on the training set but poorly on new data.

LSTM model predicts the stock price of Twitter. The model can identify the general pattern in the validation data, which is a good sign, and it suggests that the model is learning to generalize to new data. However, there are some discrepancies between actual and predicted values. It indicates that the model has not yet learned all the underlying patterns in the data or contains some noise or variability that is difficult to model accurately. It is seen that the predicted values are a good match which shows that the performance of our model is

great in predicting the stock price of Twitter. A similar is true for the prediction of Facebook stock, This indicates that LSTMs can capture long-term dependencies by selectively storing information over long periods. In other words, a very low loss shows a better performance using the LSTM model. We present the results of NLP-based techniques for sentimental analysis in financial news as in Table 1. We also present the results of the

real-time analysis visualization dashboard using streamlit.

In this paper, LSTM with 7 hidden layers are deployed. Adam optimizer and sigmoid activation

function is used .Furthermore, a dropout of 0.2 is considered. Streamlit, Prophet, Scikit-learn, Pandas, Numpy and Plotly Python libraries are used for experiment setup. Miniconda is used for environmental setup. These experiments are deployed on Apple(AAPL) stocks from Nov 2021 to Sep 2022. LSTM Here, we have used a Long Short Term Memory Network (LSTM) for building the model to predict the stock prices. LTSMs are a type of Recurrent Neural Network for learning long-term dependencies. It is commonly
used for processing and predicting time-series data.

Prophet

Prophet is a procedure for forecasting time series data based on an additive model where non-linear trends are fit with yearly, weekly, and daily seasonality, plus holiday effects. It works best with time series that have strong seasonal effects and several seasons of historical data. Prophet is robust to missing data and shifts in the trend, and typically handles outliers well.

Software functionalities:

1. Company Selection

The user can select a company from the drop down menu and the system will get an input as the stock code.

2. Years of prediction

The user has the option to select the number of years he wants the system to predict. The user can select between 1 to 4 years for prediction.

3. Current stock price analysis

The user can now see the current stock price and view it interactively with the help of a slider.

4. Predicted stock price analysis

Once the user has gone through the current stock price he can take a look at the predicted stock price and how it may vary over time.

5. Predicted forecast components

At the end user can take a look at multiple deciding factors or components behind the forecast of the stock by the system.

7.2 Overall Discussions and Insights:

Sector Comparison:

Tech Stocks: Google, Amazon, and Facebook show strong, consistent growth, benefiting from the overall positive sentiment in the tech sector. Their stable performance makes them attractive for long-term investors seeking reliable returns.

Tesla: Tesla, while also a tech-related stock, displays much higher volatility and speculative trading behavior. Its rapid growth and frequent price swings attract more risk-tolerant investors.

SBI: As a representative of the banking sector, SBI demonstrates stability and modest growth. This stock is influenced by different economic factors and provides a more conservative investment option.

Risk and Volatility:

Tesla's Volatility: Tesla's higher volatility suggests higher risk and potential for greater rewards, attracting more speculative investors. Its price movements are more influenced by market sentiment and news.

Stable Tech Stocks: The relatively stable patterns in Google, Amazon, and Facebook indicate these stocks are perceived as safer, long-term investments. They offer steady growth with lower risk compared to Tesla.

SBI's Predictability: SBI's stable growth and lower volatility make it a suitable choice for conservative investors seeking stability and predictable returns.

7.3 Market Dynamics:

Impact of Events: Significant events and trading volume spikes provide insights into how external factors (e.g., earnings reports, macroeconomic news) impact investor behavior and stock prices. Understanding these dynamics can help investors anticipate market reactions.

Trading Behavior: The volume charts and scatter plots reveal trading behavior patterns, such as increased activity during major events. This information can guide investors in timing their trades.

7.4 Investment Strategies:

Long-term Growth: Investors seeking stable, long-term growth might prefer Google, Amazon, and Facebook due to their consistent performance and lower volatility.

High-risk, High-reward: Those willing to take on more risk for potentially higher returns might find

Tesla appealing. Its rapid growth and frequent price swings offer opportunities for significant gains.

Conservative Investment: SBI could be a choice for conservative investors looking for stability and lower risk. Its steady growth and predictable returns make it a reliable investment option.

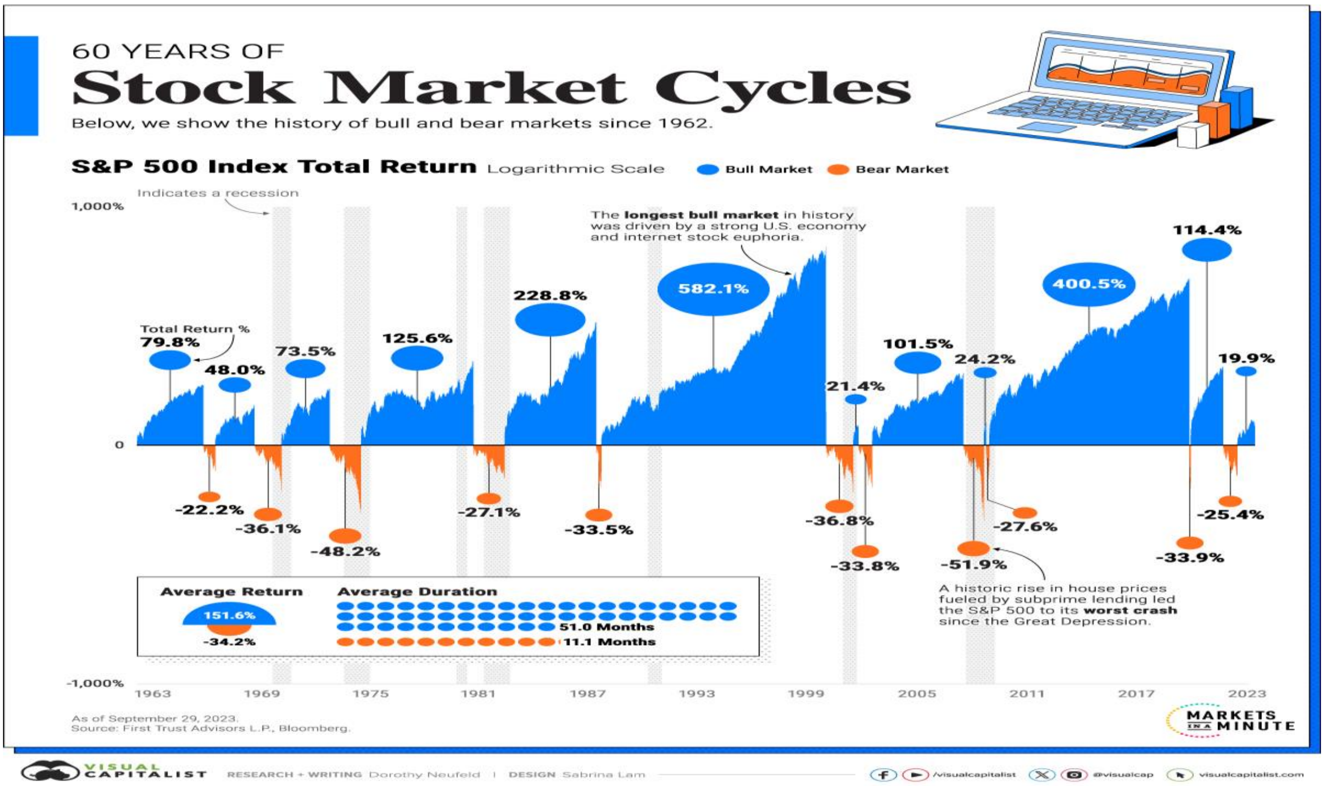


Fig. 7.2 Stock Visualization Cycle

CHAPTER 8
CONCLUSION AND FUTURE WORK

8.1 CONCLUSION

Stock market trading is a popular and highly sought-after field, and researchers are constantly exploring new techniques for stock price prediction. Accurate stock price forecasting is crucial for investors and individuals to make informed decisions in the stock market. In this project, we employed deep learning models, specifically LSTM units, to accurately predict stock prices, providing investors and individuals with valuable insights into the stock market situation. We developed a web application for predicting the closing stock prices of various organizations, using Dash Library and applied datasets from Facebook and Microsoft stocks. The results obtained from our methodology showed above 95% accuracy for these datasets, indicating the effectiveness of our approach in predicting stock prices. The use of deep learning models, such as LSTM, in stock price prediction can provide significant advantages, including the ability to capture complex patterns and dependencies in the data, and adapt to changing market conditions. Our findings suggest that incorporating deep learning models into stock price prediction can enhance the precision of forecasting, and provide valuable insights for investors and individuals in managing their stock market investments. However, it is important to acknowledge the limitations of our methodology, including potential biases in the datasets, limitations of LSTM as a predictive model, and challenges in obtaining reliable and relevant news articles. Further research and improvements can be explored, such as incorporating additional features, using different machine learning algorithms, and validating the findings on larger and more diverse datasets. Overall, our project demonstrates the potential of deep learning models, specifically LSTM, for accurate stock price prediction, and contributes to the field of finance and machine learning by providing insights into the effectiveness and limitations of using these models in stock market forecasting.

By measuring the accuracy of the Linear Regression algorithms, we found that the most suitable algorithm for predicting the market price of a stock based on various data points from the historical data. The algorithm will be a great asset for brokers and investors for investing money in the stock market since it is trained on a huge collection of historical data and has been chosen after being tested on a sample data. The project demonstrates the machine learning model to predict the stock value with more accuracy as compared to previously implemented machine learning models.

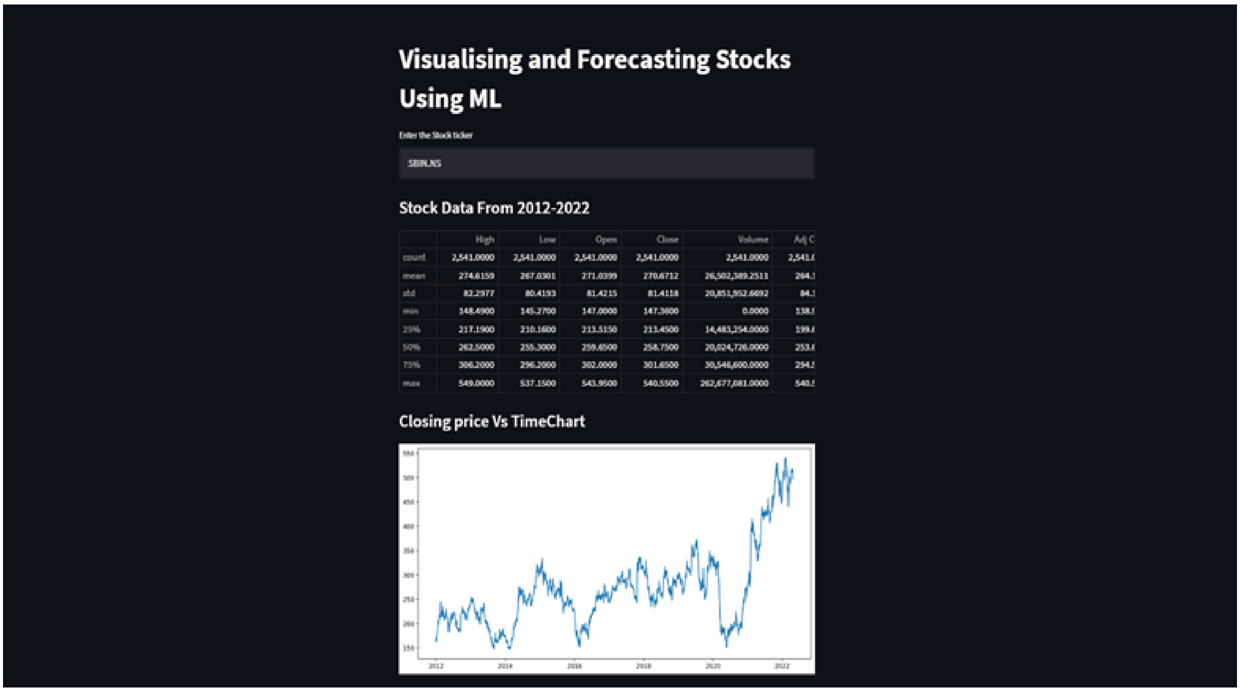
The LSTM model is utilized in this study to forecast stock prices. The Facebook and Twitter datasets are utilized to train and test the LSTM model, which consists of 32 hidden layers and one output layer. During the training process, the model performed well and was trained for 100 epochs. The root mean square error (RMSE) for training and testing data is 5.91 and 24.40, respectively, per the statistics. The results indicate that LSTM can identify time-series patterns with long-term and short-term memory. To address overfitting, we plan to use techniques such as regularization, cross-validation, or early stopping during training in the future. These techniques simplify the model and prevent it from overfitting the training data. The research comprehensively explored NLP based methods for sentiment analysis in finance over time. By leveraging LSTM, we achieved greater accuracy and efficiency in sentiment analysis. Additionally, this study included neutral sentiments alongside positive and negative ones. Investors can leverage these emotions to gain insight into market trends before trading begins. While this study only categorized emotions in articles as positive, neutral, or negative, it is possible to identify additional emotional categories. Including negative emotions such as sadness, fear, anger, and disgust can offer valuable insight into how the public perceives societal changes and their influence on the financial sector. This can help streamline concluding market trends from multiple sources and, ultimately, enhance decision-making. Furthermore, our model's real-time visualization dashboard presents live sentiment analysis of financial news interactively, making it a valuable tool for trend analysis

8.2 FUTURE SCOPE

Future scope of this project will involve adding more parameters and factors like the financial ratios, multiple instances, etc. The more the parameters are taken into account more will be the accuracy. The algorithms can also be applied for analyzing the contents of public comments and thus determine patterns/relationships between the customer and the corporate employee.

The field of stock market trading has seen significant growth in recent years, with an increasing number of investors seeking opportunities in this market. As a result, there is a need for accurate visualizing and forecasting systems to assist investors in making informed investment decisions. However, stock market visualization and forecasting can be challenging due to the numerous factors that influence stock prices. In this project, we have developed a system for visualizing and forecasting stocks using deep learning models, specifically LSTM, to provide accurate predictions.

Our findings indicate that further improvements can be made in the future by incorporating additional features and non-numerical factors, with the guidance of subject matter experts, to enhance the accuracy of stock price predictions. Additionally, we plan to extend this application to predict cryptocurrency trading and incorporate sentiment analysis for better analysis of market sentiments. It is important to acknowledge the limitations of our current methodology, including potential biases in the data and the limitations of LSTM as a predictive model. Further research and development are needed to continuously improve and refine the system for stock market visualization and forecasting.



To further enhance the analysis and visualization of stock data, consider the following future directions:

Enhanced Data Integration: Incorporate real-time data feeds and alternative data sources such as social media sentiment and economic indicators.
Advanced Analytical Techniques: Implement machine learning algorithms and natural language processing to improve predictive accuracy and automate insights.

Interactive and Personalized Visualizations: Develop customizable and interactive visualization tools to increase user engagement and satisfaction.
Integration with Trading Platforms: Integrate visualization tools with trading platforms for seamless analysis and execution.
Improved Accessibility and Education: Create educational tools to help investors understand and interpret visualizations and ensure accessibility to a broader audience.
Ethical Considerations and Data Privacy: Ensure ethical use of AI and robust data privacy measures to build trust and comply with regulations.

8.3 Enhanced Data Integration

Scope:

Real-time Data: Integrate real-time data feeds to provide up-to-the-minute stock performance analysis. Alternative Data Sources: Incorporate alternative data sources such as social media sentiment, news articles, and economic indicators to enrich the analysis.

Impact:

Comprehensive Analysis: Enhanced data integration will allow for more comprehensive and timely analysis, providing deeper insights into market dynamics.
Predictive Power: Access to a broader range of data sources can improve the predictive power of stock analysis models

.

8.4 Advanced Analytical Techniques

Scope:

Machine Learning: Implement machine learning algorithms to identify complex patterns and predict future stock movements.
Natural Language Processing (NLP): Use NLP to analyze textual data from news, reports, and social media for sentiment analysis and event detection.
Quantum Computing: Explore the potential of quantum computing to solve complex optimization problems in stock trading and portfolio management.

Impact:

Enhanced Predictive Accuracy: Advanced analytical techniques can enhance the accuracy of predictions and uncover hidden relationships in the data.

Automated Insights: These techniques can automate the generation of insights, making it easier for investors to make informed decisions.

8.5 Interactive and Personalized Visualizations

Scope:

Customization: Develop customizable visualization tools that allow users to tailor their analysis to specific needs and preferences.

Interactivity: Create interactive dashboards that enable users to explore data dynamically, apply filters, and drill down into specific details.

Impact:

User Engagement: Interactive and personalized visualizations can increase user engagement and satisfaction by providing a more intuitive and user-friendly experience.

Actionable Insights: Customizable tools can help users identify actionable insights more effectively by focusing on the most relevant data.

8.6 Integration with Trading Platforms

Scope:

Seamless Integration: Integrate visualization tools with trading platforms to provide a seamless experience from analysis to execution.

Automated Trading: Develop automated trading strategies based on real-time analysis and machine learning models.

Impact:

Efficiency: Seamless integration with trading platforms can streamline the process, reducing the time between analysis and trade execution.

Strategic Advantage: Automated trading based on advanced analytics can provide a strategic advantage by quickly responding to market conditions.

8.7 Improved Accessibility and Education

Scope:

Educational Tools: Develop educational tools and resources to help investors understand how to interpret stock visualizations and use them in decision-making.

Accessibility: Ensure that advanced visualization tools are accessible to a wider audience, including individual investors and small businesses.

Impact:

Investor Empowerment: Improved accessibility and education can empower a broader range of investors to leverage advanced analysis tools.

Market Participation: Increased understanding and accessibility can lead to greater market participation and more informed investment decisions.

8.8 Ethical Considerations and Data Privacy

Scope:

Ethical AI: Ensure that the use of AI and machine learning in stock analysis adheres to ethical guidelines, avoiding biases and ensuring fairness.

Data Privacy: Implement robust data privacy measures to protect sensitive financial data.

Impact:

Trust and Transparency: Ethical considerations and data privacy measures can build trust with users and ensure the integrity of the analysis.

Regulatory Compliance: Adhering to ethical guidelines and data privacy regulations can help avoid legal issues and maintain a positive reputation.

In the near future, we plan to explore the possibility of incorporating unstructured textual information in the model such as investor's sentiment from social media, earning reports of underlying companies, the immediate policy-related news, and research reports from market analysts. Another potential direction of the future work can be developing hybrid predictive models by combining the LSTM with some other neural networks architectures. To improve the prediction accuracy even further, we also plan to implement hybrid optimization algorithms to train the model parameters by combining the existing local optimizers with the global optimizers such as genetic algorithms and particle swarm optimization algorithm

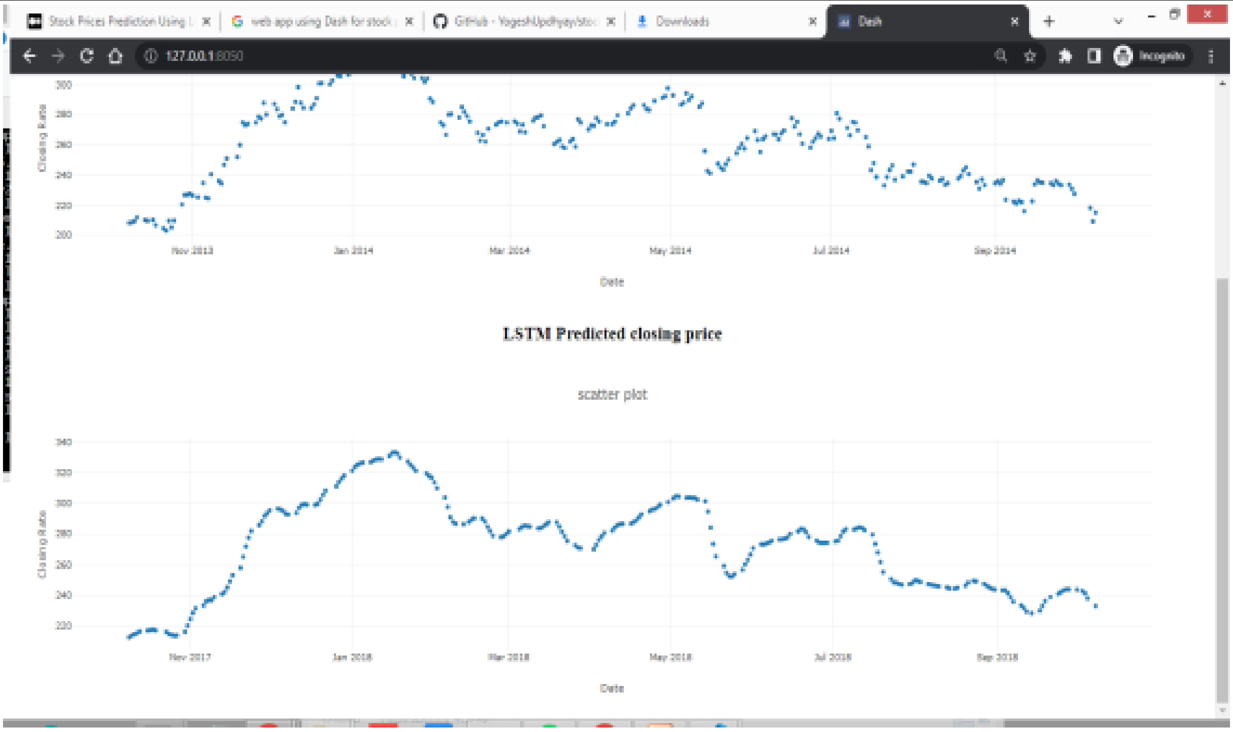


Fig. 8.5 LSTM Prediction Diagram

REFERENCES

1. Survey of stock market prediction using machine learning approach, Ashish Sharma Dinesh Bhuriya ; Upendra Singh 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)

2. Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network,Ze Zhang ; Yongjun Shen ; Guidong Zhang ; Yongqiang Song ; Yan Zhu, 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)

3. Combining of random forest estimates using LSboost for stock market index prediction,Nonita Sharma ; Akanksha Juneja,2017 2nd International Conference for Convergence in Technology (I2CT)

4. Using social media mining technology to assist in price prediction of stock market,Yaojun Wang ; Yaoqing Wang,2016 IEEE International Conference on Big Data Analysis (ICBDA)

5. Stock market prediction using an improved training algorithm of neuralnetwork,Mustain Billah ; Sajjad Waheed ; Abu Hanifa,2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)

6. Efficacy of News Sentiment for Stock Market Prediction,Sneh Kalra ; Jay Shankar Prasad,2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)

7. Literature review on Artificial Neural Networks Techniques Application for Stock Market Prediction and as Decision Support Tools,Muhammad Firdaus ; Swelandiah Endah Pratiwi ; Dionysia Kowanda ; Anacostia Kowanda

8. Stock Market Movement Prediction using LDA-Online Learning Model,Tanapon Tantisripreecha ; Nuanwan Soonthomphisaj,2018 19th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking

and Parallel/Distributed Computing (SNPD) 9.Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment, Zhaoxia Wang ; Seng-Beng Ho ; Zhiping Lin 2018 IEEE International Conference on Data Mining Workshops (ICDMW)

10. Stock Price Prediction Using News Sentiment Analysis Saloni Mohan ; Sahitya Mullapudi ; Sudheer Sammeta ; Parag, Vijayvergia ; David C. Anastasiu,2019 IEEE Fifth International Conference on Big Data Computing Service and Applications(BigDataService)

11. Mohri, M., Rostamizadeh, A., Talwalkar, A.: Foundations of Machine Learning, 2nd edn. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA (2018)

12. Cao, J., Li, Z., Li, J.: Financial time series forecasting model based on ceemdan and lstm. Physica A: Statistical mechanics and its applications 519, 127–139 (2019)

13. Bao, W., Yue, J., Rao, Y.: A deep learning framework for financial time series using stacked autoencoders and long-short term memory. PloS one 12(7), 0180944 (2017)

14. Yan, B., Aasma, M., et al.: A novel deep learning framework: Prediction and analysis of financial time series using ceemd and lstm. Expert systems with applications 159, 113609 (2020)

15. Ding, X., Zhang, Y., Liu, T., Duan, J.: Deep learning for event-driven stock prediction. In: Twenty-fourth International Joint Conference on Artificial Intelligence (2015)

16. Ding, X., Zhang, Y., Liu, T., Duan, J.: Knowledge-driven event embedding for stock prediction. In: Proceedings of Coling 2016, the 26th International Conference on Computational Linguistics: Technical Papers, pp.2133–2142 (2016)

17. Cai, J., Cavallo, E.: Event-driven stock price prediction using convolutional neural networks with uncertainty-awareness.

18. Lee, C.-Y., Soo, V.-W.: Predict stock price with financial news based on recurrent convolutional neural networks. In: 2017 Conference. on Technologies and Applications of Artificial Intelligence (TAAI), pp. 160–165 (2017).IEEE

19. Ain, Q.T., Ali, M., Riaz, A., Noureen, A., Kamran, M., Hayat, B., Rehman, A.: Sentiment

analysis using deep learning techniques:a review. International Journal of Advanced Computer Science and Applications 8(6)(2017)

20. Sadaei, H.J., Enayatifar, R., Abdullah, A.H., Gani, A.: Short-term load forecasting using a hybrid model with a refined exponentially weighted fuzzy time series and an improved harmony search. International Journal of Electrical Power & Energy Systems 62, 118–129 (2014).

21. “Stock price prediction using LSTM, RNN and CNN-sliding window model - IEEE Conference Publication.” https://ieeexplore.ieee.org/document/8126078 (accessed Dec. 27, 2019).

22. J. Jagwani, M. Gupta, H. Sachdeva, and A. Singhal, “Stock Price Forecasting Using Data from Yahoo Finance and Analysing Seasonal and Nonseasonal Trend,” in 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, Jun. 2018, pp. 462–467, doi: 10.1109/ICCONS.2018.8663035.

23. I. Parmar et al., “Stock Dec. 2018, pp. 574–576, doi: 10.1109/ICSCCC.2018.8703332.

24. Y. Lei, K. Zhou, and Y. Liu, “Multi-Category Events Driven Stock Price Trends Prediction,” in 2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS), Nanjing, China, Nov. 2018, pp. 497–501, doi: 10.1109/CCIS.2018.8691392.

25. B. Jeevan, E. Naresh, B. P. V. kumar, and P. Kambli, “Share Price Prediction using Machine Learning Technique,” in 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), Bangalore, India, Oct. 2018, pp. 1–4, doi: 10.1109/CIMCA.2018.8739647.

26. M. Usmani, S. H. Adil, K. Raza, and S. S. A. Ali, “Stock market prediction using machine learning techniques,” in 2016 3rd International Conference on computer and Information Sciences (ICCOINS), 2016, pp. 322–327.

27. J. Du, Q. Liu, K. Chen, and J. Wang, “Forecasting stock prices in two ways based on LSTM neural network,” in 2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC), Mar. 2019, pp. 1083–1086, doi: 10.1109/ITNEC.2019.8729026.
28. S. E. Gao, B. S. Lin, and C.-M. Wang, “Share Price Trend Prediction Using CRNN with LSTM Structure,” in 2018 International Symposium on Computer, Consumer and Control (IS3C), Dec. 2018, pp. 10–13, doi: 10.1109/IS3C.2018.00012.
29. T. Gao, Y. Chai, and Y. Liu, “Applying long short term momory neural networks for predicting stock closing price,” in 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, Nov. 2017, pp. 575–578, doi: 10.1109/ICSESS.2017.8342981.