VISVESVARAYA TECHNOLOGICAL UNIVERSITY

JNANA SANGAMA, BELAGAVI – 590 018, KARNATAKA, INDIA

A PROJECT REPORT

ON

"SMARTSTOCK INSIGHT: PREDICTIVE STOCK PRICE ANALYSIS USING SENTIMENT ANALYSIS AND LSTM NETWORKS"

Submitted in partial fulfillment of the requirements for the award of

MASTER OF COMPUTER APPLICATIONS (MCA)

Submitted By

Name	USN
MAHAMMAD AFNAN M	4VP22MC027

Under the Guidance of

Dr. JOTHIMANI K

Professor

Department of Master of Computer Applications

Vivekananda College of Engineering & Technology

DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS

VIVEKANANDA COLLEGE OF ENGINEERING & TECHNOLOGY

[A Unit of Vivekananda Vidyavardhaka Sangha, Puttur (R)]

Affiliated to Visvesvaraya Technological University and Approved by AICTE New Delhi & Govt. of Karnataka

Nehru Nagara, Puttur – 574 203, DK, Karnataka, India

August 2024

VIVEKANANDA COLLEGE OF ENGINEERING & TECHNOLOGY
[A Unit of Vivekananda Vidyavardhaka Sangha, Puttur (R)]
Affiliated to Visvesvaraya Technological University and Approved by AICTE New Delhi & Govt. of Karnataka
Nehru Nagara, Puttur – 574 203, DK, Karnataka, India

DEPARTMENT OF MASTER OF COMPUTER APPLICATIONS

CERTIFICATE

Certified that the project work entitled **"SmartStock Insight: Predictive Stock Price Analysis using Sentiment Analysis and LSTM Networks"** is carried out by **Mahammad Afnan M** bearing USN **4VP22MC027** bonafide student of **Vivekananda College of Engineering & Technology**, in partial fulfillment for the award of **Master of Computer Applications (MCA)** of the **Visvesvaraya Technological University, Belagavi** during the year 2023-24. It is certified that all corrections/ suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library.

The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the said Degree.

Signature of the Guide	**Signature of the HOD**	**Signature of the Principal**
Dr. Jothimani K	**Mr. Anil Kumar K**	**Dr. Mahesh Prasanna K**
Professor	**Director(I/C) - Dept. of MCA**	**Principal**
VCET, Puttur	**VCET, Puttur**	**VCET, Puttur**

EXTERNAL VIVA

Name of the Examiners — Signature with date

1……………………………….......... ....

2……………………………….......... ....

DECLARATION

I, **Mahammad Afnan M (4VP22MC027)** student of fourth semester Master of Computer Applications (MCA), **Vivekananda College of Engineering & Technology**, Puttur, hereby declare that the project work entitled **"SmartStock Insight: Predictive Stock Price Analysis using Sentiment Analysis and LSTM Networks"** has been carried out and duly executed by me at VCET, Puttur, under the guidance of **Dr. Jothimani K**, Professor, Department of Master of Computer Applications, Vivekananda College of Engineering & Technology, Puttur. This project is submitted in partial fulfillment of the requirements for the award of degree in **Master of Computer Applications (MCA)** by **Visvesvaraya Technological University**, Belagavi during the academic year 2023-24.

Date:

Place: VCET, Puttur

Mahammad Afnan M

4VP22MC027

ACKNOWLEDGMENT

I take this opportunity to express my deep heartfelt gratitude to all those people who have helped me in the successful completion of the project.

First and foremost, I would like to express my sincere gratitude towards our beloved guide **Dr. Jothimani K**, Professor, for providing excellent guidance, encouragement and inspiration throughout the project work. Without her invaluable guidance, this work would never have been a successful one.

I would like to express my sincere gratitude to our Director (I/C) of the Department of Master of Computer Applications (MCA), **Mr. Anil Kumar K** for his guidance and inspiration.

I have great pleasure in expressing my profound gratitude to our respected Principal **Dr. Mahesh Prasanna K** for his constant support and encouragement.

I would like to thank our Management for providing all the facilities and a proper environment to work in the college campus.

I am thankful to all the teaching and non-teaching staff members of Master of Computer Applications (MCA) Department for their help and needed support rendered throughout the project.

International Journal of Engineering Applied Science and Technology

Is hereby honoring this Certificate to

Mahammad Afnan M

in recognition of the publication of manuscript entitled

STOCKSIGHT: A NOVEL APPROACH TO STOCK PRICE PREDICTION WITH SENTIMENT ANALYSIS AND LSTM NETWORKS

Published in IJEAST Journal - Vol. 9, Issue 01, May 2024

Abhishek Srivastava
Editor In Chief
IJEAST Journal

Email : editor.ijeast@gmail.com
Web : ijeast.com

ISSN : 2455-2143

ABSTRACT

Accurately predicting changes in stock prices is a challenging task for investors due to the ever-evolving behaviour of the Indian stock market. This study integrates sentiment analysis with LSTM models to enhance prediction accuracy. By using NLP techniques, market sentiment from news headlines is evaluated. Financial and textual data are collected, cleaned, and used for sentiment analysis, categorizing market sentiment into a scale of positive (1), negative (-1), and neutral (0). This sentiment data is integrated with historical stock prices to train the LSTM model. A user-friendly application is developed for real-time predictions. Future improvements include automating the model-building process, incorporating more data sources, and enhancing sentiment analysis techniques. This project demonstrates the potential of combining sentiment analysis with LSTM models for more accurate stock price predictions and deeper market insights.

Keywords: BeautifulSoup, LSTM Model, Sentiment Analysis, Stock Price Predictions, Time-Series Prediction, Web Crawling.

Table of Contents

Table of Contents **i**

List of Figures **iii**

List of Tables **iv**

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Objectives . . . 2
- 1.3 Problem Description . . . 2
- 1.4 Proposed Method . . . 3
- 1.5 Limitations . . . 4

2 Literature Review **5**
- 2.1 Introduction . . . 5
- 2.2 Related Work . . . 6
- 2.3 Research Gaps and Significance . . . 11
- 2.4 Conclusion . . . 12

3 Design Phase **13**
- 3.1 Detailed explanation of the design process . . . 13
- 3.2 The proposed system . . . 14
- 3.3 System Architecture and Components . . . 15
- 3.4 Design Decisions and justifications . . . 16
- 3.5 Design Diagrams . . . 17

4 Implementation **20**
- 4.1 Description of the implementation process . . . 20
- 4.2 Technologies, tools, and frameworks used . . . 21
- 4.3 Challenges faced during implementation and their solutions . . . 22
- 4.4 Code snippets . . . 23
- 4.5 Testing methodologies and quality assurance measures . . . 26

5 Result and Discussion **27**
- 5.1 Presentation and Interpretation of the Achieved Outcomes . . . 27
- 5.2 Comparison of the Results with the Project Objectives . . . 32
- 5.3 Discussion of Any Deviations or Unexpected Outcomes . . . 33
- 5.4 Interpretation of Findings Based on the Problem Statement . . . 33

5.5 System Performance and Effectiveness Analysis 34

6 Conclusion and Future Work **37**

6.1 Summary of the Key Findings and Contributions of the Project 37

6.2 Recapitulation of the Project Objectives and Their Fulfillment 37

6.3 Discussing the Project's Implications and Significance 38

6.4 Future Research and Improvement Suggestions 38

6.5 Reflection on the Lessons Learned During the Project 39

6.6 Project Work Publication 39

Bibliography **40**

List of Figures

1.1 SmartStock Insight: Predictive Analysis of the Stock Price 3

3.1 System Architecture 15
3.2 Sentiment Analysis 18
3.3 LSTM Model 19

5.1 Sentiment Analysis for TCS 27
5.2 Stock Price Prediction for TCS 28
5.3 Sentiment Analysis for Tata Motors 28
5.4 Stock Price Prediction for Tata Motors 29
5.5 Sentiment Analysis for Infosys 29
5.6 Stock Price Prediction for Infosys 30
5.7 Sentiment Analysis for Asian Paints 30
5.8 Stock Price Prediction for Asian Paints 31
5.9 Sentiment Analysis for Tech Mahindra 31
5.10 Stock Price Prediction for Tech Mahindra 32
5.11 Home page of the SmartStock Insight App 34
5.12 Prediction page of the SmartStock Insight App 35

6.1 Journal Publication Certificate 39

List of Tables

2.1 Research Survey Based on SmartStock Insight 12

5.1 Performance Metrics for the Predicted Stock Price 36

Chapter 1

Introduction

The financial market is very dynamic and affected by various complex things, presenting a huge problem to investors and financial analysts who want to make investment in the ever-growing Indian stock market. To maximize profits and minimize risks, stock price forecasting is necessary. A practical or valid method for improving predictions is to analyze past stock price data and to identify hidden patterns. This is made possible by the recent advancements in the machine learning area.

The "SmartStock Insight" is an effort to transform the prediction of Indian stocks utilizing the abilities of machine learning, with a particular focus on the Long Short Term Memory (LSTM) networks. LSTM works best in dealing with sequential data and provides a refined approach to handle the temporal dependencies present in financial data.

The main reason for such an innovative technique arises from the disadvantages of old approaches used to analyze the market, which often fall short due to stock fluctuations. It is not at all possible to accurately anticipate the stock price, the reason for that is the inherent complexity of financial markets demands more advanced tools. The project's main intention is to empower novice investors to invest in the Indian stock market. By combining the LSTM networks with sentiment analysis from stock news, the project seeks to provide users with reliable and accurate insights into stock price movements.

1.1 Motivation

The "SmartStock Insight" project is derived from the requirement to get a more advanced and accurate tool for stock price prediction. Previous tools that are available for stock price forecasting usually fail to record the complicated movements of the financial markets. As markets become increasingly complex, People need a predictive model that provides investors with more reliable predictions or insight.

The primary motivation is to find out the existing volatility of the markets. Stocks are very much subjected to the sudden and unpredictable fluctuations, providing a major challenge for the investors to take reliable decisions. By utilizing the power of LSTM networks, the project seeks to develop a prediction tool that the investors can use to effectively navigate through dynamics and provide investors with a more reliable understanding of stock price movements.

The project is motivated by a dedication to empowering novice or first-time investors. The trend has shown us that people are interested in stocks. The thinking is to build a user-friendly application that combines advanced machine learning techniques, making predictions accessible to a wider range of audience. This not only serves the needs of experienced as well as novice users allowing them to navigate the complex market with some confidence, provided by accurate predictions.

1.2 Objectives

- **Comprehensive Data Collection and Preprocessing:**

  The objective is to make a robust data collection and preprocessing. This will involve gathering news headlines from various well-known trusted sources such as YFinance, Moneycontrol, NewsAPI, and Google News, and then combining them into a single dataset. Historical stock prices are also collected so that a complete dataset can be developed and it can be used for the predictive analysis. Effective preprocessing ensures that processed data is clean, relevant, and ready for modeling, providing a solid foundation for accurate stock price predictions.

- **Integration of Sentiment Analysis with LSTM Model:**

  The goal is to develop a predictive model that combines sentiment analysis with LSTM networks. To improve prediction, the project integrates sentiment analysis, capturing the emotional context around the stock market, leading to more reliable stock price predictions. By combining sentiment analysis, the application also aims to take into account the consequences of market sentiment and news on stocks. The LSTM model utilizes these combined features to identify patterns and trends, improving the accuracy of stock price predictions. This also ensures that the model captures both numerical and emotional aspects of market data.

- **Development of a User-Friendly Application:**

  This project simplifies complex and dynamic stock market data into useful actionable insights, providing novice users some confidence to make informed financial decisions, and allowing novice users to gain some confidence in investing in the fast-growing Indian stock market. The app will provide visualizations through an interactive, responsive, and accessible web interface, and it will show expected prices along with current prices also the app will provide you the recommendations.

1.3 Problem Description

Novice investors often face challenges in making informed investment decisions because of an insufficient amount of easily accessible tools and guidance, that will help to make reliable stock price predictions. Current methods often fail to achieve the specific requirements of individuals who are new to the stock market, leading to a very bad investment practice.

A common issue when it comes to prediction in the traditional method is that it is difficult to determine the prevailing patterns, especially in the stock markets of India. Novice investors do not feel comfortable due to the inefficient processing of large volumes of financial data and also experience difficulties in utilizing news articles and expert analyses. This circumstance calls for a review of the problem, opening the door for the creation of a novel solution to allow investors to make decisions.

1.4 Proposed Method

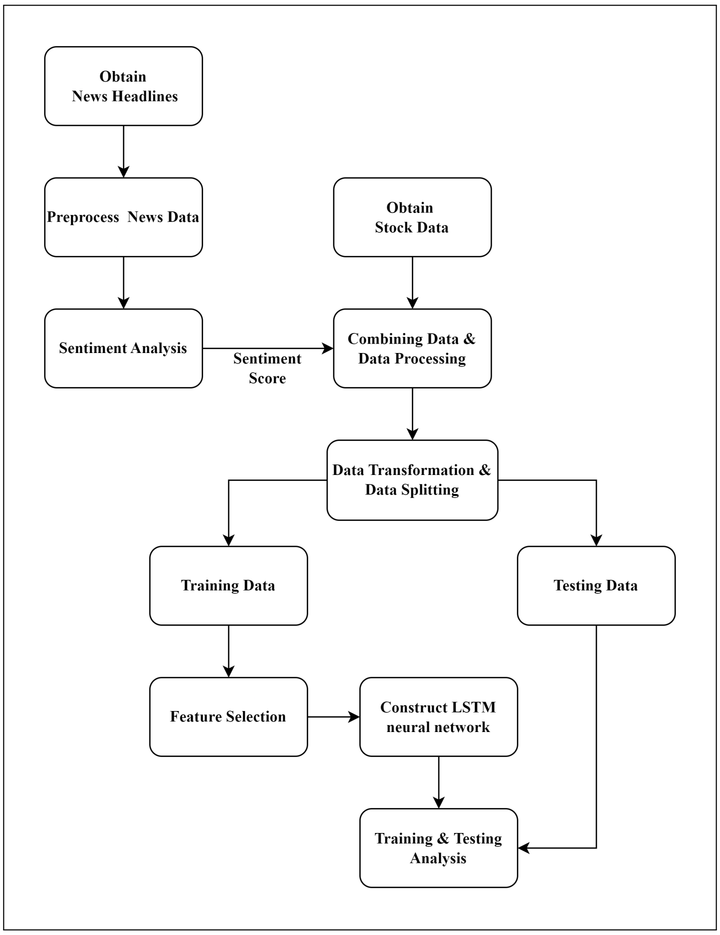


Figure 1.1: SmartStock Insight: Predictive Analysis of the Stock Price

Data Collection: Online sources are utilized to gather useful news articles regarding the stocks and perform some analysis. Using real-time crawling techniques along with using API service, the system can regularly gather news headlines, keeping the model up to date with the most recent market sentiment.

Sentiment Analysis: The collected news headlines are analyzed to get to know about the emotional context of a certain stock, and NLP techniques will be utilized to analyze the sentiment represented in the collected textual data. Along with the quantitative information obtained by LSTM, evaluating positive or negative sentiment can provide a more appropriate and accurate insight into stock forecasting.

Data Integration and Preprocessing: The extracted sentiment scores are then combined with other data, such as historical stock prices, and technical indicators. This combined dataset undergoes further preprocessing, including cleaning, formatting, and transformation, to ensure compatibility with the subsequent LSTM model.

LSTM Model Construction and Training: Since LSTM networks are very effective in performing analysis on sequential data, they are the better choice for analyzing and identifying underlying trends to predict prices. Also, the dataset of sentiment scores, technical indicators, and historical stock data is used for both training and testing of the LSTM model. The LSTM gains the ability to recognize patterns and links between the features as well as the target stock price during training.

User Interface Design: An intuitive and user-friendly interface will be designed for the application, prioritizing accessibility for novice investors. Visual representations of stock predictions and sentiment analysis results will be displayed in the application, providing the user's ability to understand and utilize the information effectively.

Testing and Validation: LSTM model's predictions are evaluated on a separate dataset of unseen data. This helps assess the model's generalizability and ability to perform well on new information. To make sure the model is effective in forecasting changes in stock prices, its performance will be carefully evaluated using historical data. To make sure the model is robust and capable of managing unpredictable market events.

1.5 Limitations

The effectiveness and efficiency of "SmartStock" is greatly affected by the quality, quantity, and reliability of the collected data, which includes news articles and historical stock prices. Any inaccurate or biases in this data can affect performance. The fluctuating behavior of the stock market more specifically the Indian stock market, poses a huge challenge, making it very difficult for the model to adapt to sudden market fluctuations. The dependability and correctness of the sentiment analysis are also very important, as contextual nuances and language trends can lead to misinterpretations. While the project aims to simplify market data for novice investors, user education remains essential for effective understanding. Our overreliance on past information assumes that patterns will again repeat, which may not be the case if market dynamics shift towards new trends, potentially reducing prediction accuracy. Limited market news coverage and the dynamic nature of the financial markets, influenced by external events and geopolitical factors, further make the task of prediction more difficult.

Chapter 2

Literature Review

2.1 Introduction

Predicting stock prices is more than just a technological challenge; it's a quest for knowledge in history. The "SmartStock Insight" project embarks on this journey, guided by the compass of the literature review. This review bridges past gaps with future aspirations, revealing the evolution of prediction methods from their roots to sentiment analysis along with the LSTM model. Each paper becomes a stepping stone on this financial odyssey, a testament to the ongoing quest to decode market dynamics.

Moving beyond historical context, the literature review reveals a treasure trove of diverse methodologies. From market sentiment to analysis that helps to identify temporal patterns in data with advanced networks, there are many researchers globally, who have explored and applied various approaches to forecast the price of stock. This review allows us to not only get inspiration but also extract valuable lessons to be used in our own prediction model.

The literature review serves as both a roadmap and an advisory also historical instruction. It highlights past researchers' successes and also failures, throwing light on what succeeded and failed in the dangerous area of stock forecasting. Exploring a variety of findings, "SmartStock Insight" projects against common challenges by using actual approaches and identifying potential risks. This review will hold some academic respect, also it serves as the foundation around the development, demonstrating our commitment to creativity, flexibility, and pushing the envelope in the league of stock prediction.

The intention of the "SmartStock Insight" is to establish a robust theoretical foundation for predicting stock prices. This involves synthesizing essential ideas, methods, and principles from existing research in predictive analytics, machine learning, and financial modeling. This comprehensive synthesis lays the groundwork for a strong theoretical base.

Another crucial objective is the identification and evaluation of effective technologies commonly employed in the successful development of the prediction model designed for the share market. By aligning with contemporary and efficient tools and frameworks, the objective is to ensure the efficacy of our predictive analytics approach. Additionally, the project aims to learn from the experiences of existing prediction tools for stock prices. This involves extracting insights, analyzing strengths and weaknesses, and incorporating successful strategies and lessons learned to continually refine the methodology for "SmartStock Insight".

The "SmartStock Insight" also focuses on understanding the various problems related to collected data. This includes several issues related to data quality, selecting features, and temporary considerations. The knowledge gained from this lookup will help to do effective data preparation and preprocessing strategies. Furthermore, the combination of modeling insights from diverse predictive models aims to enhance and refine the overall methodology for "SmartStock Insight". Lastly, the identification of research gaps and limitations in the current literature is a crucial objective. This recognition provides the way for innovative contributions within the project, addressing areas where the existing literature falls short and advancing knowledge in predictive analytics for stock prices.

2.2 Related Work

Demonstrating the satisfactory performance of ML tools, particularly Random Forest and Bagging with leaked data. Comparing the performance of various classifiers, including ANN, K-NN, SVM, Decision Tree, RF, Bagging, and AdaBoost, using four stock market index datasets, the results indicate that retraining classifiers with leaked data enhances accuracy, with RF and Bagging with leaked data exhibiting the highest performance. Overall, RF and Bagging with leaked data, show good results for stock market forecasting [1].

Suggesting a machine-learning model for market forecasting to get stock insights using SSVMs and graph cuts. The suggested model provides a high accuracy in training and good results in test samples. Highlighting the possibility of integrating machine learning and graph structures for real-life applications like stock prediction [2].

Real-time sentiment analysis on Twitter data is employed to anticipate the stock prices, leveraging Apache Spark and Flume. The feasibility of predicting the stock prices based on general public sentiment, utilizing Twitter for performing sentiment analysis for the share market has shown the way for future model enhancements and stream adaptation. Mittal et al. implement crucial data preprocessing techniques, such as approximating missing DJIA values for weekends and holidays, developing a word list based on the Profile of Mood States (POMS) questionnaire, extending it with synonyms, filtering tweets likely to express feelings, and computing daily scores using a word counting algorithm for sentiment analysis [3].

Proposing a mobile app developed to give market insight using IMLR and Moving Average, performed well in baseline methods accuracy. The app utilizes Yahoo Finance API data and offers daily history and real-time predictions. Demonstrating the practical application of IMLR in stock prediction, the paper contributes to both prediction methodologies and mobile app development [4].

Proposing an innovative prediction method for a stock price that combines sentiment analysis and Google Trends data. The model utilizes the Multiple Input Multi-Step Output LSTM and achieves an average MAPE of 0.05932 for one-week forecasts. The research highlights the necessity to use VADER sentiment analysis, in the nuance of tweets, and explores Google Trends data for stock predictions. The author concludes by expressing the insights of the developed framework, experimental analysis, and future directions regarding the predictive capability of the shares. The research highlights Twitter's potential for predicting sentiment and its impact on stock prices, providing technical insights into sentiment and correlation analysis using Twitter data [5].

Utilizing LSTM networks to predict Bovespa stock prices, outperforming classical algorithms. With historical data and technical indicators as features, the LSTM model demonstrates significant accuracy through statistical tests and promising financial results in simulated trades, suggesting its potential for prediction [6].

Improvement of the accuracy while predicting can be obtained by utilizing the capabilities of advanced technology, specifically ANN and Random Forest, which are explored. Comparing the performance given by the ANN and RF models, By utilizing metrics such as RMSE and MAPE, the outcome shows that ANN outperformed RF in closing price forecasting. Incorporating financial news articles into the models is suggested for further improvement [7].

Applying deep learning techniques (LSTM & SVR) to address the limitations of traditional methods in stock prediction. Highlighting LSTM's effectiveness in analyzing time-series data and its power over SVR for non-linear stock prices, the research emphasizes requirements of better forecasting approaches and provides feedback for future investigation of the model so that it is useful for improving accuracy [8].

Proposing a mobile app that provides stock insights, this study employs IMLR and Moving Average, showcasing superior accuracy compared to baseline methods. Utilizing data from the YFinance, the app also provides both daily historical trends and real-time predictions. The research not only illustrates the practical application of IMLR in stock prediction but also contributes valuable insights to prediction methodologies and the domain of mobile app development [9].

Effectiveness of RNNs, particularly LSTM models, for anticipating the stock values is explored. Highlighting the LSTM's memory capabilities and its benefits when dealing with longer-time data, the study provides the importance of training with few data and more moments or time for the results. The research showcases its potential for predicting asset values and managing portfolios without explicit programming instructions [10].

The capability of the LSTM and Multilayer Perceptron (MLP) models are compared. By utilizing the daily price data and various technical indicators as features, the results demonstrate the effectiveness of the models while identifying the price trends. The MLP model provided a better outcome than the LSTM networks model in the case of short-term prediction accuracy. Further research is suggested to extend the models and develop a real-time trading platform based on the predictions. The U.S. stock market dataset, sourced from Yahoo Finance, provides daily opening, highest, lowest, closing, and adjusted closing prices. Adjusted close prices per working day, ignoring daily fluctuations and volataility in the sector, also enhance dataset suitability for predictive modeling and helps to forecast the movement [11].

The study by Mohan et al. incorporates essential preprocessing techniques such as stemming news headlines for standardization, calculating and appending polarity scores to the dataset, converting news data to numerical form, and utilizing the Event Registry API to extract relevant articles for Google stock. They also ensure consistency by converting all characters to lowercase and retaining only alphabetic characters in news headlines. The evaluation involved using MAPE and an ARIMA model, with a meta-parameter grid search for optimization. The study suggests future work, such as building domain-specific models and considering adverse effects on stock prices due to related news, offering a valuable reference for text-based forecasting. The study emphasizes Twitter's role in public opinion analysis for stock market forecast. [12].

Investigating sentiment classification of Chinese stock news using an N-gram statistical language model and Vector Space Model. The study emphasizes text preprocessing, and sentiment dictionary construction, and employs Naïve Bayes, K-nearest Neighbor, and SVM for classification. The effect of the quantity of the data and stock/block increase on classification accuracy, the study gives a brief idea about the experimental methodologies they have utilized [13].

Jin et al. meticulously organized their dataset, comprising 96,903 comments from stocktwits.com for model training and sentiment index calculation, supplemented by sentiment-rich comments on Apple stocks sourced from Yahoo Finance. The study ensured representativeness by selecting comments with substantial likes (80 comments daily) and conducted experiments to evaluate the proposed scheme's effectiveness. Essential preprocessing methods for stock reviews were applied, including acronym changes, spelling correction, root restoration, symbol replacement, and sentiment index calculation. Leveraging word2vec improved word vector representations, while integrating an attention layer with LSTM enhanced focus on pertinent information. The author focused on developing an LSTM model used for anticipation of stock price, incorporating SVR, and EMD for trend extraction, and an attention mechanism within the LSTM. The model integrated sentiment analysis from platforms like StockTwits and Yahoo Finance, thus capturing emotional factors. Various performance metrics were utilized to evaluate the model's effectiveness, considering both historical stock data and sentiment information [14].

The Indian stock market, serving as a crucial source of financing for businesses, contributes significantly to the nation's economic progress. Delving into the market's mechanisms, historical trajectory, and how economic factors influence its volatility, the paper proposes measures like circuit breakers and other strategies to manage volatility present in the market. Effective diversification strategies can further reduce risk and increase the market's potential for long-term gains. They employed LSTM for stock price insight, addressing challenges in forecasting future price value and providing the importance of financial research. The ANN architecture encompasses input, hidden, and output layers. It delves into the LSTM model structure, examining the epochs and data length on testing outcomes. Results are visually portrayed through figures and tables, offering insights into training sets and testing sets across diverse datasets and epochs [15].

A stock market predictive model is showcased using advanced machine learning models, utilizing two classifiers, K-Nearest Neighbors (KNN) and Random Forest. The Random Forest classifier outperforms the KNN classifier in accuracy, achieving an average accuracy ratio of 93%. This study argues that utilizing the Random Forest (RF) model can provide higher profits and reduce the risk [16].

Sentiment analysis along with deep learning techniques are used to get insight into the stock prices. The NaiveBayesClassifier is utilized for the sentiment analysis, while a Generative Adversarial Network (GAN) outcomes the price forecast. The models achieve good accuracy, ranging from 69.86% to 87.86%, showcasing the power of sentiment analysis along with deep learning together in the prediction. They used a supervised learning approach with an LSTM model to get insight into the prices of Google stock. Their dataset spans five years, focusing on opening/closing prices, and also utilized the sentiment scores derived from daily news headlines. The LSTM model employs 60 previous days as a window of input for training the model to predict the next closing price, subsequently applied to generalize for predictions [17].

The deep learning techniques can be used for enhancing the prediction accuracy. Employing a comprehensive dataset of S&P500 companies' data and financial news articles, the study compares various models, including ARIMA, Facebook Prophet, and RNNs, finding that RNN-pt and RNN-pp demonstrate superior performance. Additionally, the research also highlights the wants of financial news in the process of forecasting movements in stocks and sentiment analysis on prediction accuracy. Overall, the study gives insights into improving stock prediction by combining time series data and textual sentiment analysis. He employs critical preprocessing techniques, including obtaining historical daily stock data from YFinance API and collecting news articles from The New York Times. The data is normalized by using the z-scale approach, then the collected dataset is spliced into the training sets and the testing sets, LSTM model parameters, and combined stock data with sentiment scores for analysis so that we can get a better idea of what is happening in the market [18].

Introducing the LDA-POS method for forecasting stock movement using public sentiments collected from social media. The study compares accuracy results in English and Persian datasets, emphasizing the importance of retaining specific part-of-speech words in Persian preprocessing. Addressing challenges to get opinions from social media, including short texts and unusual grammatical structures, the Conclusion provides insights into method performance, and dataset differences, and suggests areas for future improvement, that guide to the improvement in forecasting [19].

Proposing a stock forecasting tool using MKL regression with technical analysis, sentiment analysis, and social network data. Features beyond stock prices, especially sentiment analysis, improve prediction accuracy compared to baseline methods. MKL regression reveals sentiment analysis, a powerful predictor, surpassing numerical dynamics for effective forecasting the stock price [20].

ML techniques, including traditional techniques, deep learning, and graph-based approaches, are utilized for stock forecasting. The selection of the method relies on the dataset and accuracy requirements. Traditional algorithms excel with large datasets, deep learning models with binary features, analyzing time series data along with sentiment analysis for precise predictions, and graph-based approaches for market insights. Future research should combine sentiment analysis and historical data, refine deep learning techniques, and explore behavioral finance for market analysis [21].

Utilizing LSTM networks to anticipate market movement by integrating sentiment analysis and technical analysis. The uses a dataset consisting of about 60 days of stock price data and 10 daily news headlines. The study highlights the benefits of the Adam optimizer and introduces regularization techniques to prevent overfitting. The results demonstrate that dropout inclusion significantly enhances model generalization and reduces validation errors. The research gives the conclusion by saying that the effect of news data on the market fluctuations and also they are proposing an optimal data structure for RNN. They focus on price prediction of stocks using RNN, utilizing historical data for model training and testing. The NSE stock market dataset and the China stock markets dataset are specifically mentioned for preprocessing, selection of additional features, and model building [22].

Examining the combination of sentiment analysis from news articles and LSTM model for stock forecast. Sentiment analysis is performed using VADER and TextBlob, showcasing the capability of sentiment analysis while performing the prediction. The research shows good accuracy for stock prediction when sentiment is considered, contributing to combining sentiment analysis with the LSTM model for capturing the effects of public sentiment on the market movements [23].

Applying LSTM for forecasting stocks, especially in the Indian share market. The paper highlights the challenges related to human behavior and market sentiments, advocating for Machine Learning techniques, particularly LSTM, to discern future stock price trends. Detailing the LSTM architecture, outlining the methodology for stock prediction, including data cleaning, selection of features, and error calculation, the paper references related studies on LSTM and RNN offering an understanding of tools for analyzing and prediction in the market [24].

By using the power of the complicated ML algorithms, specifically LSTM, for prediction of the market price within India. Proposing an LSTM-based platform that utilizes historical market data and performing the sentiment analysis on the news sources to anticipate the share market prices, the model obtained a score of 70% accuracy for a small collection of stocks and 86% for some of the shares presented on the Nifty 50 index [25].

2.3 Research Gaps and Significance

Integration of External Factors: Many existing stock prediction models predominantly rely on textual data, overlooking the incorporation of external factors. By considering the wider economic and geopolitical context, incorporating a wider range of external factors can improve prediction accuracy and lead to a more thorough understanding of market dynamics.

Unified Approach to Modeling: The absence of a single, consistent method in existing models leads to a lack of unity in combining sentiment analysis, technical specifications, and machine learning. Developing a cohesive model that integrates emotional factors, technical aspects, and machine learning can result in more robust and accurate predictions. This news idea may give a better understanding of the things that affect share market fluctuations.

Explainability in Models: Many current stock prediction models lack sufficient search for explainability features, resulting in a lack of transparency and interpretability in their predictions. Explainability builds user trust. Clear explanations help users comprehend and rely on the predictions, fostering broader acceptance. Features like visualizing and decision-making enhance transparency, ensuring that the users can make correct decisions in the market.

Focus on Beginners: Creating easy-to-understand models with educational features empowers beginners, promoting inclusivity in financial decision-making. The objective is to make stock market insights accessible to everyone. Creating easy-to-understand models with educational features empowers beginners, promoting inclusivity in financial decision-making. Also, make stock market insights accessible to everyone.

Real-Time Adaptability: Certain models lack emphasis on real-time adaptability, potentially limiting their responsiveness to dynamic market changes. Real-time adaptability is pivotal for timely insights into rapidly evolving market conditions. Models designed for swift adjustments based on the latest information empower users to make well-informed decisions, ensuring sustained relevance and effectiveness in predictive analytics. Incorporating features for quick adaptation enhances the model's utility in navigating the complexities of dynamic financial landscapes.

Ethical and Bias Considerations: Some studies do not thoroughly examine the ethical implications and biases present in sentimental analysis and the ML tools are utilized in the process of determining the stock price. Addressing ethical concerns and biases is essential for ensuring fair and responsible use of models that are used in financial markets for the forecast. This includes identifying and mitigating biases in training data, algorithms, and predictions to prevent discriminatory outcomes. Ethical considerations also involve transparency about the limitations and potential risks that a predictive model's behavior may incorporate, fostering trust and accountability among users and stakeholders.

Table 2.1: Research Survey Based on SmartStock Insight

Author	Year	Description
Mittal et al.[3]	2012	Utilized DJIA values from June 2009 to December 2009 from Yahoo! Finance and over 476 million tweets were extracted from over 17 million users in the public during the period while performing the sentiment analysis.
Pagolu et al.[5]	2016	Extracted 250,000 tweets related to Microsoft from the Twitter API, underwent preprocessing for missing stock prices and used sentiment analysis with Word2vec and N-gram representations.
Mohan et al.[12]	2019	Incorporated stemming of news headlines, calculated and appended polarity scores, and utilized the Event Registry API to extract relevant articles for Google stock.
Jin et al.[14]	2020	Used stock comments data organized into a dataset from stocktwits.com and sentiment-rich comments from Yahoo Finance for AAPL from March 4, 2013, to February 28, 2018. Employed word2vec, attention layer with LSTM, and EMD for closing price series.
Sarkar et al.[17]	2020	Applied a supervised learning approach using LSTM for predicting Google stock prices, covering five years (January 2014 to December 2018) and incorporating sentiment scores derived from daily headlines.
Guo et al.[18]	2020	Employed Yahoo Finance API for historical daily stock data, normalized data using the z-scale approach, and combined stock data with sentiment scores for analysis.

2.4 Conclusion

The project marks the important role of ML techniques, particularly LSTM, The task of forecasting the stock, showcasing their potential to capture intricate patterns and forecast asset values. The combination of sentiment analysis, along with technical analysis and diverse datasets, has proven effective in enhancing prediction accuracy. However, identified gaps, such as limited focus on external factors, challenges in model explainability, biases, and the evolving landscape of sentiment analysis, present opportunities for further research. Addressing these gaps is crucial for fostering an inclusive, adaptable, and ethically sound approach to stock market forecasting, emphasizing real-time adaptability and user-friendly models. Despite significant progress, ongoing refinement is essential to create more robust, transparent predictive models that instill trust and inclusivity in financial decision-making processes.

Chapter 3

Design Phase

3.1 Detailed explanation of the design process

The "SmartStock Insight" project implements multiple approaches that combine advanced tools and techniques along with user-centric interface development. The initial step in this process involves in detailed collection of data, which forms the base for the system. Appropriate news articles are gathered from various trusted online sources which includes Yahoo Finance, Moneycontrol, NewsAPI, and Google News. This news data collection makes sure that the system has up-to-date information on recent market trends, sentiments, and things happening around the market. In parallel, historical stock prices, are collected for the same period as the news data. This combination of data sources is key for the other stages of analysis and modeling.

"SmartStock Insight" gathers piles of historical stock data and current news articles to generate precise stock forecasts. News is gathered from trusted sources. The gathered news is then analyzed to determine the sentiment score. Knowing the market's emotional response to a certain situation is crucial for making accurate price predictions. Performing the sentiment analysis on the gathered news data is the next stage. Following the study, the stock data and sentiment scores are integrated to produce a complete dataset.

Extracting valuable insights from the raw data is where machine learning takes center stage. "SmartStock Insight" utilizes NLP to analyze news headlines. This sentiment analysis injects emotional context into the stock data, which is crucial for understanding market reactions. The combined dataset undergoes a set of processes like cleaning, formatting, and transformation processes to ensure reliability and compatibility with the LSTM. This process includes also dealing with missing or unavailable values, normalizing the data, and shifting the data into number format. The objective is to create a high-quality dataset that accurately represents the underlying patterns in the stock market.

The construction and training of the LSTM model are important things in the design process. LSTM networks are particularly well-suited to continuous timely data, visually to capture long-term dependencies and patterns in sequential data. The LSTM model is trained on the preprocessed dataset, sentiment scores, and futuristic stock prices. During the training, various hyperparameters including different layers, learning rates, and batch size are optimized to increase the model's prediction capability.

To make sure that the "SmartStock" app is easy and clear to use for everyone, The UI along with the development process is designed. This way, the app will be intuitive from the start. A web framework called Flask will be utilized to connect the Model to the user interface. The interface will show current stock prices and use graphs and charts to display the forecasted prices visually. So that the users can analyze easily and make better decisions, regardless of their experience level. By focusing on usability, the app will also provide valuable feedback for investors of all backgrounds.

3.2 The proposed system

"SmartStock Insight" aims to make stock analysis easier, especially for new investors. The system uses advanced technology, like a special kind of machine learning called LSTM, to analyze news articles and past stock prices. This helps the system understand how news and past performance might affect future stock prices. By putting everything together, This comprehensive approach gives users a clearer picture of what might happen in the stock market.

The system begins with the collection of data from various online sources. News articles related to specific stocks are gathered using real-time crawling techniques or API subscriptions. This ensures that the system has access to the latest information, which is crucial for accurate sentiment analysis. Previous price data are gathered for the same period as the news information to make sure the sentiment analysis is contextually relevant.

Sentiment analysis is a major part of the system, providing an emotional context to the stock data. Using NLP techniques, the sentiment of the collected news headlines is determined. This involves analyzing the textual data to categorize the sentiment score. The sentiment scores are then integrated with the historical stock data, creating a comprehensive dataset that combines both quantitative and qualitative insights.

The combined dataset undergoes a series of preprocessing steps to ensure consistency and compatibility with the LSTM model. This includes cleaning the data so that the efficiently handles missing data, normalizing the data gathered to make sure that selected features are on a comparable scale, and converting numerical representations from category data. The idea here is to create a high-quality dataset that accurately represents the underlying trends in the stock market movement.

The LSTM model is designed to analyze sequential data, making it particularly well-suited for timely data prediction. After that, the preprocessed dataset is used to train and test the model, allowing it to learn the complicated relations between historical stock data, sentiment scores, and future stock prices. When the training is conducted, various hyper-parameters which include a different variety of layers, learning rate, and batch size are optimized to increase the model's predictions.

The user-friendly interface of the app is created with both inexperienced and seasoned investors in mind. The interface, which is developed through a well-known web development framework Flask, places a premium on simplicity and clarity. It shows both the current and projected stock prices side by side rather than overloading customers with technical data. Graphs and charts are included in the interface to improve comprehension of the forecasts. Based on the information provided, investors are empowered by this user-centric design to make wise investment decisions.

3.3 System Architecture and Components

The system architecture of the “SmartStock Insight” is a complete system that integrates several components to facilitate stock price forecasting using advanced machine learning tools. As illustrated in Figure 3.1, the architecture is divided into several key modules, each serving a distinct purpose in the overall workflow. Each module tackles a specific task: collecting data, analyzing sentiment, and building a prediction model. This seamless integration of these components makes the system effective.

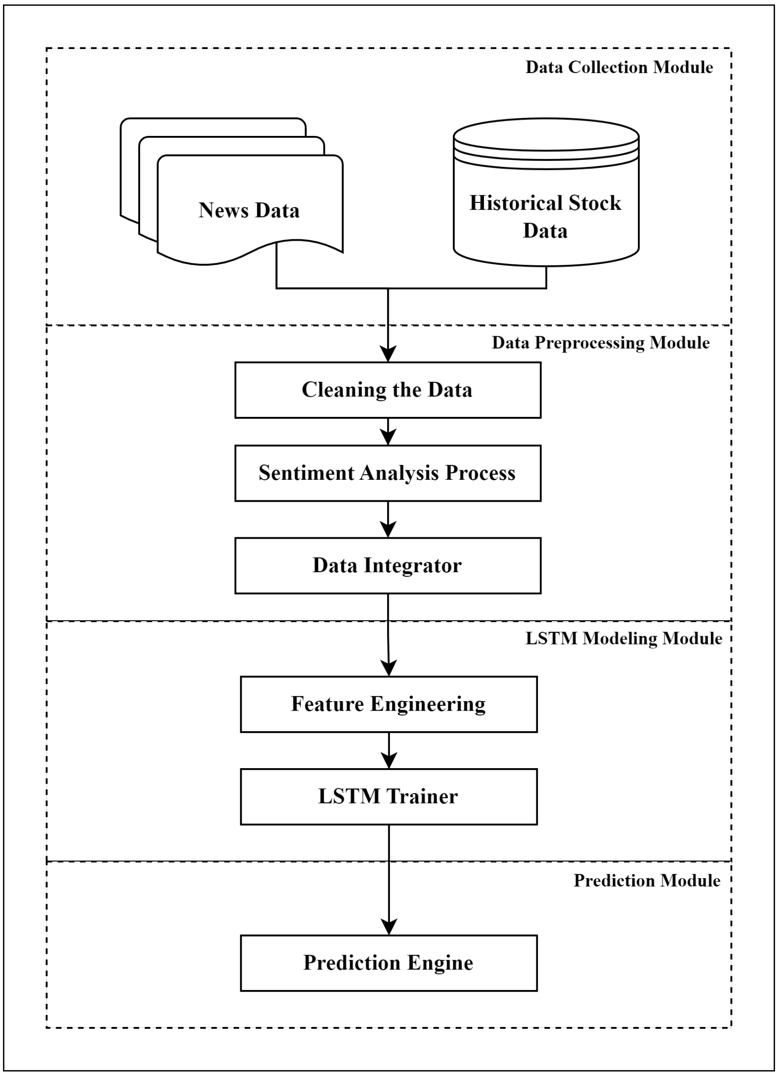


Figure 3.1: System Architecture

The system starts by gathering the building blocks for its predictions. This is done by a data collection module that acts like a tireless information gatherer. It uses real-time crawling techniques and the News API to ensure the system has the latest stock data and news articles. The data collected includes key details, along with trading volume – all crucial ingredients for making appropriate and useful predictions of the market.

The system doesn't just look at numbers; it also analyzes how the news is talking about a stock. This is where the sentiment analysis module comes in. It uses powerful language processing methods to understand the emotions behind news headlines. The module helps the system understand how the news might be affecting the market. This emotional context is key to predicting future stock movements.

Once the sentiment scores are generated, the data integration and preprocessing module combines them with the historical stock data. This module ensures that the combined dataset is clean, consistent, and compatible with the LSTM model. The process involves handling missing or unavailable values, normalizing the data, and converting qualitative data into numerical representations.

The LSTM modeling module is the core component of the system, responsible for constructing and training the LSTM network. LSTM networks are particularly well-suited for continuous or timely data as well suited for prediction. The complex model is then used with the preprocessed dataset, allowing the system to learn the complex relations within historical stock prices and the sentiment scores.

3.4 Design Decisions and justifications

The design decisions made during the development of the "SmartStock Insight" system were driven by the need to create a robust, accurate, and user-friendly tool used for stock market analysis. Each decision was carefully considered to make sure that the system meets the needs of both novice and experienced investors.

The choice to use LSTM networks best suited for the sequential analysis was made for their effectiveness in modeling sequential data. LSTM networks are particularly well-suited for financial sequential data due to their capability to identify long-term patterns in the data. This makes them an ideal choice for predicting stock prices, where historical price trends and market patterns are also included when dealing with anticipation of the price movement regarding a selected stock or for future price movements forecast.

Integrating sentiment analysis into the system was another critical design decision. It provides an emotional context to the stock data, which is essential for understanding market reactions and trends. By analyzing the context of the news related to selected stocks, the system can incorporate qualitative insights into the predictive model. This enhances the reliability and efficiency of the predictions by considering not only the historical price trends but also the market sentiment.

The decision to build a user-centric interface was driven by the need to that the system accessible to a large group of users. The interface was designed to be intuitive and user-friendly, ensuring that both novice and experienced investors can easily navigate the application. The use of graphs and charts helps users interpret the predictions effectively using visuals, making it easier for the users to make well-informed decisions.

During the preprocessing phase, careful attention is paid to ensuring the consistency and quality of the dataset. This involved avoiding missing data, normalizing data, and converting qualitative data into numerical representations. The intention is to create a rich dataset that exactly represents the patterns and movement in the market. This decision is driven by the need to make sure that the LSTM model has access to accurate and reliable data during the training phase.

Beyond the core data points like historical prices and sentiment analysis, the system leverages the ta package to extract additional features available from the data. This package provides a very huge number of analysis indicators especially technical-related indicators, which are mathematical formulas used to identify hidden and visible patterns in the price history. By incorporating these indicators as features in the model, the system captures and identifies a more detailed understanding of the market. This can include features like moving averages, RSI, and Bollinger Bands, each offering valuable insights into price momentum, overbought/oversold conditions, and market volatility.

The choice to use Flask for developing the user interface was driven by its lightweight nature and ease of integration with the predictive model. Flask provides a flexible framework for building complete web apps, allowing for seamless integration of the LSTM model in the backend with the front-end interface by using Flask.

The system prioritizes a user-friendly experience. That's why it uses Flask, a streamlined web development tool, to build the interface. Flask helps to easily integrate the powerful prediction model with the user interface as a whole system. This seamless connection ensures users see the latest predictions alongside clear visuals of the stock data, allowing them to make informed decisions.

3.5 Design Diagrams

Figure 3.2 details the process used while performing the sentiment analysis, which is a pivotal component of the "SmartStock Insight" project. Sentiment analysis involves several steps, beginning with the gathering of the news articles. The collected articles are from multiple online trusted platforms to ensure a broad and diverse dataset.

The articles go through a preprocessing step after they are collected, after that the text is standardized, and unwanted data is removed. The next step involves handling punctuation, normalizing the text, and removing stop words. Once cleaned, the text is ready for sentiment analysis.

The sentiment analysis is derived from the application of NLP techniques to extract sentiment scores from the text. This involves using ML techniques and pre-trained models to classify the sentiment of each news headline. Sentiments are divided into positive(1), negative(-1), or neutral(0), with corresponding scores assigned to each category. To train the LSTM model, the text should be processed, each article should be given a sentiment score, and then all of the articles should be pooled into one large dataset.

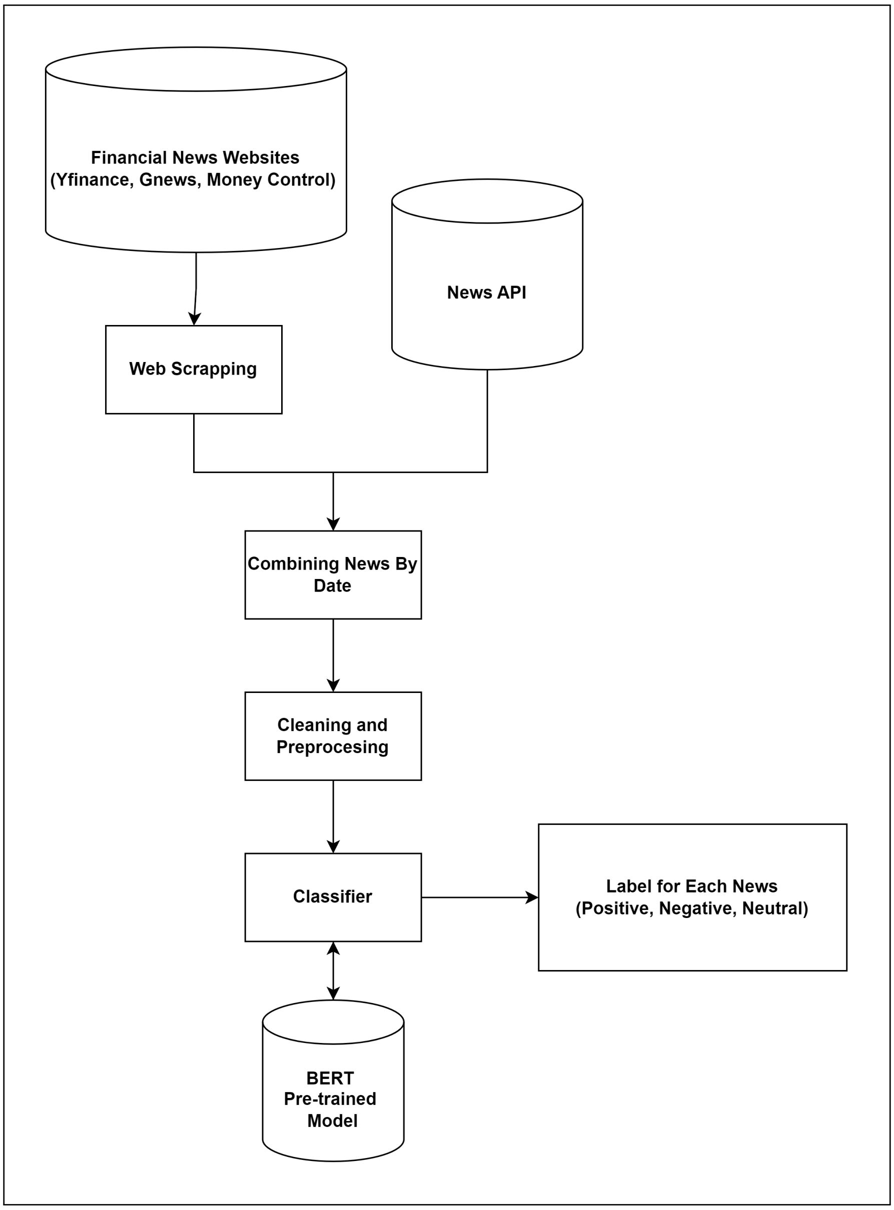


Figure 3.2: Sentiment Analysis

Figure 3.2 illustrates how these sentiment scores are integrated with the historical stock data. By combining the emotional context of the news with the quantitative stock data, a more nuanced dataset is created. This enriched dataset is then used in the feature engineering process, ensuring that the LSTM model receives comprehensive input that includes both market sentiment and historical price movements.

The predictive model of the “SmartStock Insight” is depended on the LSTM network, as represented in Figure 3.3. Particularly well-suited for sequential data analysis, are LSTM networks, a kind of RNN.

The LSTM network structure utilized in “SmartStock”. It is illustrated in Figure 3.3. The network starts with input layers that receive the processed dataset, which consists of sentiment scores and historical stock prices. The LSTM cells receive these inputs and are designed to retain information over extended sequences, which makes them perfect for recognizing trends in changes in stock prices.

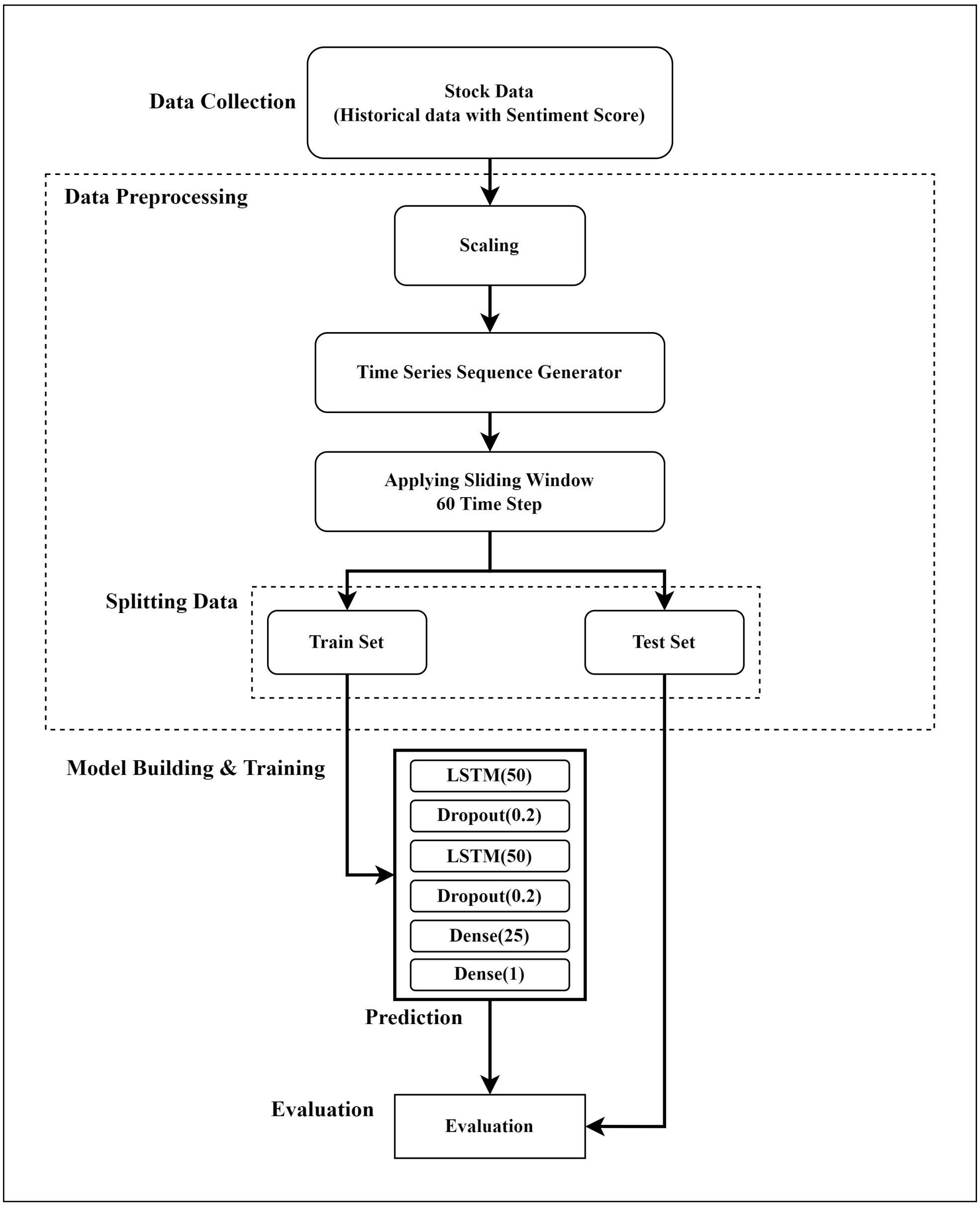


Figure 3.3: LSTM Model

A sequence of gates, including input, forget, and output gates, are used by each LSTM cell to process the incoming data. These gates control information flow, allowing the network to remember or ignore information as needed to understand the temporal relationships present in the data. After passing through dense layers, the LSTM cell outputs are converted from learned features into final predictions. To reduce the prediction error, the LSTM model is trained using historical data. Using the patterns it has learned from the input data, the learning network LSTM can help forecast future stock prices after it has been trained.

The LSTM network outputs the expected stock prices, which are then displayed to the user via the web application developed using Flask. To assist users make wise investment decisions, this application not only displays the expected price but also visualizes them concerning the current value. By displaying charts as well as graphs users can easily understand the market behavior for a particular stock they are interested in and make buy or sell decisions.

Chapter 4

Implementation

4.1 Description of the implementation process

The implementation process of the "SmartStock Insight" project was a systematic endeavor that involved meticulous planning and execution of various tasks to develop a robust stock price prediction system. The process began with setting up the necessary infrastructure, which included selecting appropriate development environments and configuring the required software tools. This foundational setup was crucial to ensure a smooth and efficient development phase.

The initial step in implementation was data collection, The project utilized APIs and web scraping techniques to gather historical stock prices and relevant news articles from multiple sources like Yahoo Finance, Moneycontrol, NewsAPI, and Google News. This phase required careful handling to ensure data accuracy and relevance. Automated scripts were developed to schedule regular data collection, ensuring the dataset remained up-to-date and comprehensive.

Following data collection, the focus shifted to data preprocessing. This involved cleaning and formatting the raw data to eliminate inconsistencies and missing values. For the news articles, NLP techniques were employed to preprocess the text, including tasks like tokenization, stop word removal, and lemmatization. Concurrently, the historical stock data underwent normalization to facilitate its integration with the sentiment scores derived from the news articles.

Once the data was preprocessed, the sentiment analysis component was implemented. This involved applying pre-trained NLP models to classify the sentiment of each news headline. The sentiment scores were then combined with the historical stock prices to form a unified dataset. This integrated dataset was crucial for the subsequent feature engineering phase, where relevant features were crafted to capture the intricate patterns in the data.

The building of the LSTM model became the key element of the implementation process. For the LSTM network to learn from the sequential data and predict future stock prices, finally built and trained. Several iterations were necessary during the training phase to maximize the model's performance and adjust its parameters. After the model was sufficiently developed, its accuracy and predictive power were assessed using a different validation dataset.

4.2 Technologies, tools, and frameworks used

To effectively accomplish its goals, the "SmartStock Insight" project makes use of several technologies, tools, and frameworks. Based on its functionality, performance, and simplicity of integration with other parts, each part was carefully selected.

Programming Languages: The main language utilized for the implementation was Python. Python's variety of libraries useful for data analysis, ML tasks, and web development made it a popular choice. Python's ease of use and readability also made it simple to maintain and implement, which facilitated rapid development and debugging.

Data Collection and Preprocessing: For data collection, APIs like NewsAPI and Yahoo Finance were utilized. scraping the content from the web was performed using BeautifulSoup and Selenium. Pandas were extensively used for data manipulation and preprocessing, while a pre-trained model was used from Hugging Face for sentiment analysis along with NLP tasks including text preprocessing and sentiment analysis. This model provided a robust and efficient way to group the sentiment score from the news articles without the need for custom training.

Machine Learning Model: The core model for prediction was constructed using TensorFlow and Keras, which are powerful libraries for developing deep learning models. TensorFlow provided the necessary tools for constructing and training the LSTM network, while Keras offered a high-level API that simplified the model development process.

Web Development: The web app was built by using Flask, a micro web framework for Python. Flask was chosen for its simplicity in nature and flexibility, allowing us to build a simple yet effective user interface. HTML, CSS, and JS were used as part of front-end designing, with libraries like Bootstrap enhancing the UI design.

Visualization: For data visualization, Matplotlib and Plotly were used. These libraries enabled the creation of interactive and static graphs to display stock price predictions and historical data, providing users with clear visual insights.

Version Control and Code Management: Git is a powerful version control tool, along with GitHub serves as the repository for code management. This facilitated efficient tracking of code changes and ensured that all developments were documented systematically. Although this was an individual project, using version control helped in maintaining a clear history of modifications and reverting to previous versions when necessary.

Deployment: The final application was deployed on Render, a cloud platform that provides seamless deployment and scaling capabilities. Before deployment, the application was containerized using Docker, ensuring consistency across different environments and simplifying the deployment process. Docker containers package the application and its dependencies, simplifying the maintenance and deployment process.

4.3 Challenges faced during implementation and their solutions

The successful implementation of the “SmartStock Insight” was not without its problems or challenges. Various obstacles were encountered, each requiring innovative solutions to overcome and ensure the project’s success.

Data Quality and Consistency: One of the significant challenges was ensuring the quality and consistency of the collected data. The news articles and stock prices came from different sources, leading to discrepancies and missing values. To address this, extensive data cleaning and normalization procedures were implemented. Automated scripts were developed to handle missing values and standardize the data format, ensuring a consistent dataset for analysis.

Finding Historical News: Obtaining historical news articles posed a substantial challenge. Many news APIs and sources do not provide easy access to older articles, or they impose strict restrictions on the amount or quantity of articles that can be viewed or retrieved. To overcome this, a joint approach is used consisting of web scraping techniques and API usage was employed. Historical archives from various financial news websites were scoured, and web scraping scripts were developed to extract and compile the necessary articles. This approach allowed for a more rich and qualitative dataset along with historical news.

Sentiment Analysis Accuracy: Achieving high accuracy in sentiment analysis was another challenge. The initial models produced inconsistent results, which could affect the overall prediction accuracy. By leveraging a pre-trained sentiment analysis model from Hugging Face, the accuracy of sentiment classification was significantly improved. This model is fine-tuned on a large level of financial news dataset, to provide more reliable sentiment scores.

Unifying the Dataset: Combining different datasets into a single unified dataset required meticulous preprocessing and feature engineering. The historical stock prices and sentiment scores had to be aligned temporally and formatted consistently. This process involved handling missing values, normalizing the data, and making sure that all the features were correctly synchronized. Custom scripts were constructed to automate the task of merging the process, facilitating the construction of a cohesive dataset suitable for training the LSTM model.

Model Training and Overfitting: Training the LSTM model posed challenges related to overfitting and generalization. The model initially showed high accuracy on the training set data but performed poorly on the validation set. To mitigate overfitting, regularization techniques such as dropout and L2 regularization were applied. Hyperparameter tuning was also conducted to optimize the model’s architecture and training parameters, resulting in a more generalized and robust model.

Scalability and Performance: Ensuring the scalability and performance of the web application was crucial, especially when handling real-time data updates and user interactions.

To address this, the application was optimized for performance, and efficient caching mechanisms were implemented. The use of cloud services and containerization with Docker ensured that the application could scale seamlessly to handle increased user traffic and data volume.

User Interface Design: Designing a user-friendly interface that effectively communicated the predictions and insights posed its own set of challenges. The initial designs lacked clarity and usability. To improve the UI, user feedback was incorporated iteratively. Interactive elements and clear visualizations were added to enhance user experience and make the application more intuitive.

4.4 Code snippets

Web Scraping for Google News

some random things

```
def fetch_google_news(stock_idx):
    stock_name_gnews = ["TCS", "Tata Motors", "Infosys", "Asian Paints",
     "Tech Mahindra Ltd"]
    googlenews = GoogleNews()
    googlenews.search(stock_name_gnews[stock_idx])
    result = googlenews.result()
    df = pd.DataFrame(result)

    if not df.empty:
        df = df.rename(columns={'datetime': 'Date', 'desc': 'News'})
        df['Date'] = pd.to_datetime(df['Date']).dt.date
        df['Date'] = pd.to_datetime(df['Date'])
        df.set_index('Date', inplace=True)
        df.sort_values('Date', inplace=True)
        df.drop(['title', 'media', 'date', 'link', 'img'], axis=1,
   inplace=True)
    return df
df_gnews = fetch_google_news(stock_idx)
```

Text Preprocessing

```
def clean_text(text):
    text=text.lower()
    text = ''.join([char for char in text if char not in string.
   punctuation])
    stop_words = stopwords.words('english')
    text = ' '.join([word for word in text.split() if word not in
   stop_words])
    stemmer = PorterStemmer()
    text = ' '.join([stemmer.stem(word) for word in text.split()])
    return text
```

Sentiment Analysis

```
def perform_sentiment_analysis(df):
    classifier = pipeline("text-classification", model="mrm8488/
   distilroberta-finetuned-financial-news-sentiment-analysis")

    for row in df.itertuples():
        res = classifier(row.News)
        df.at[row.Index, 'Score'] = res[0]['score']
        df.at[row.Index, 'Label'] = res[0]['label']

    df['Label'] = df['Label'].map({'positive': 1, 'neutral': 0, '
   negative': -1})
    return df
```

Loading Sentiment Data and Combining with Historical Data

```
def load_sentiment_data(filepath):
    sentiment_data = pd.read_csv(filepath)
    sentiment_data['Date'] = pd.to_datetime(sentiment_data['Date'])
    sentiment_data.set_index('Date', inplace=True)
    return sentiment_data

def combine_data(price_data, sentiment_data):
    combined_df = price_data.join(sentiment_data, how='left')
    combined_df.fillna({'Label': 0}, inplace=True)
    return combined_df
```

LSTM Model Construction

```
from tensorflow.keras.models import Sequential as Seq
from tensorflow.keras.layers import LSTM as Base, Dense, Dropout
from tensorflow.keras.callbacks import EarlyStopping ES

model = Seq()
model.add(Base(50,return_sequences=True,input_shape=(time_step, X.shape
    [2])))
model.add(Dropout(0.2))
model.add(Base(50, return_sequences=False))
model.add(Dropout(0.2))
model.add(Dense(25, activation='relu'))
model.add(Dense(1))

model.compile(optimizer='adam', loss='mean_squared_error')

early_stop = ES(monitor='val_loss', patience=10, restore_best_weights=
    True)
```

```
history = model.fit(X_train, y_train, validation_data=(X_test, y_test),
    epochs=200, batch_size=32)

train_loss = model.evaluate(X_train, y_train, verbose=0)
test_loss = model.evaluate(X_test, y_test, verbose=0)

print("Training Loss:", train_loss)
print("Testing Loss:", test_loss)
```

Model Evaluation

```
def evaluate_model(model, X_train, y_train, X_test, y_test, scaler,
    combined_df, time_step, output_folder, ticker):
    train_predict = model.predict([X_train, X_train, X_train]).reshape
    (-1, 1)
    test_predict = model.predict([X_test, X_test, X_test]).reshape(-1,
    1)

    def inverse_transform(pred, original):
        full_pred = np.concatenate((pred, np.zeros((pred.shape[0],
    X_train.shape[2] - 1))), axis=1)
        return scaler.inverse_transform(full_pred)[:, 0]

    train_predict_inverse = inverse_transform(train_predict, X_train)
    test_predict_inverse = inverse_transform(test_predict, X_test)
    y_train_inverse = inverse_transform(y_train.reshape(-1, 1), X_train)
    y_test_inverse = inverse_transform(y_test.reshape(-1, 1), X_test)

    plot_predictions(combined_df, y_train_inverse, train_predict_inverse
    , y_test_inverse, test_predict_inverse, f"{output_folder}/{ticker}
    _stock_prediction_training_vs_testing.png")

    return train_predict_inverse, test_predict_inverse, y_train_inverse,
     y_test_inverse
```

Prediction for Next 30 Days

```
def predict_future(model, last_days, time_step, num_features, scaler):
    predictions = []
    current_input = last_days.reshape(1, time_step, num_features)

    for _ in range(30):
        next_prediction = model.predict([current_input, current_input,
    current_input])[0, 0]
        predictions.append(next_prediction)
        next_prediction_tiled = np.tile(next_prediction, (1, 1,
    num_features))
```

```
        current_input = np.concatenate((current_input[:, 1:, :],
    next_prediction_tiled), axis=1)

    predictions = np.array(predictions).reshape(-1, 1)
    predictions_full = np.concatenate([predictions, np.zeros((
    predictions.shape[0], num_features - 1))], axis=1)
    return scaler.inverse_transform(predictions_full)[:, 0]
```

4.5 Testing methodologies and quality assurance measures

Ensuring the quality and reliability of the “SmartStock Insight” involved implementing various testing methodologies and quality assurance measures. These steps were crucial to verifying the functionality, performance, and robustness of the system.

Unit Testing: These tests were developed to verify the working of individual components or modules. Each module is tested separately to make sure that it produces the expected result, covering data collection, preprocessing, sentiment analysis, and LSTM model training.

Integration Testing: Conducted to verify that different modules interacted correctly when combined. This included testing the integration of sentiment scores with historical stock data, the feature engineering process, and the seamless flow of data through the LSTM model. Integration tests helped detect and fix any issues related to data compatibility and module interaction.

System Testing: Involves evaluating the entire system’s functionality as a whole. End-to-end tests were performed so that the app was reliable and the app could handle real-world situations, from data collection and preprocessing to predicting stock prices and displaying results on the web application. These tests validated the system’s overall performance and reliability.

Performance Testing: Tests were conducted to assess the system’s responsiveness and scalability. The LSTM model’s training time, prediction accuracy, and the web application’s load time were measured.

User Acceptance Testing (UAT): This testing involves real users communicating with the system to validate its usability and functionality. Collected the feedback from the users and analyzed it to identify any improvements the project could undergo. This iterative testing process ensured that the final product met user expectations and provided a satisfactory user experience.

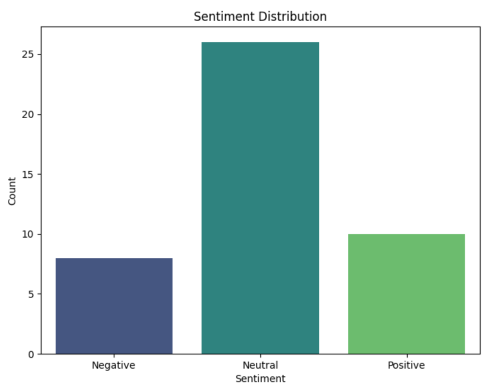
Chapter 5

Result and Discussion

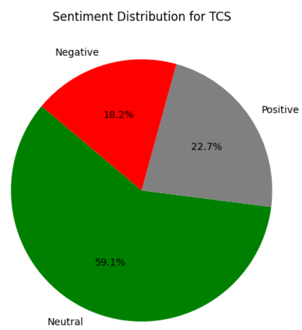
5.1 Presentation and Interpretation of the Achieved Outcomes

The results achieved from the “SmartStock Insight” are multi-faceted, encompassing the stock prices predicted generated by the LSTM model, the sentiment analysis scores derived from the news data, and the overall integration of these components into the web application. The primary output of the project is the anticipated closing prices for the selected companies, which are displayed on the web application along with historical prices for comparison. The results are discussed in detail for each stock, including graphical representations and evaluation metrics. The sentiment analysis results and the stock prediction for TCS, Tata Motors(TM), Infosys(Infy), Asian Paints(AP), and Tech Mahindra Ltd(TML). are summarized below.

For **TCS**, Figure 5.1a shows the sentiment distribution, with the majority of sentiments being positive. This indicates a generally optimistic outlook among investors and market analysts. The anticipated closing price for TCS (Figure 5.2) showcases the model’s performance conducted on the train vs test data separately, indicating good predictive power with close alignment between actual and predicted prices. The forecast for the next month presents a positive trend consistent with the sentiment analysis. The performance metrics include an MAE of 39.128, an MSE of 2678.918, and a RMSE of 51.758.

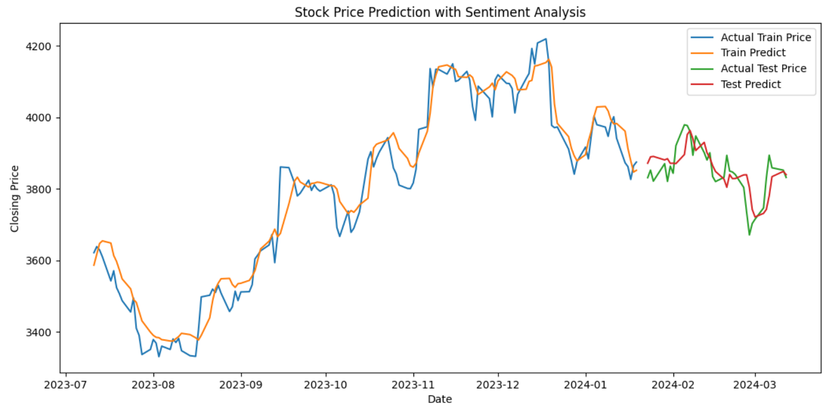


(a) TCS Sentiment Distribution

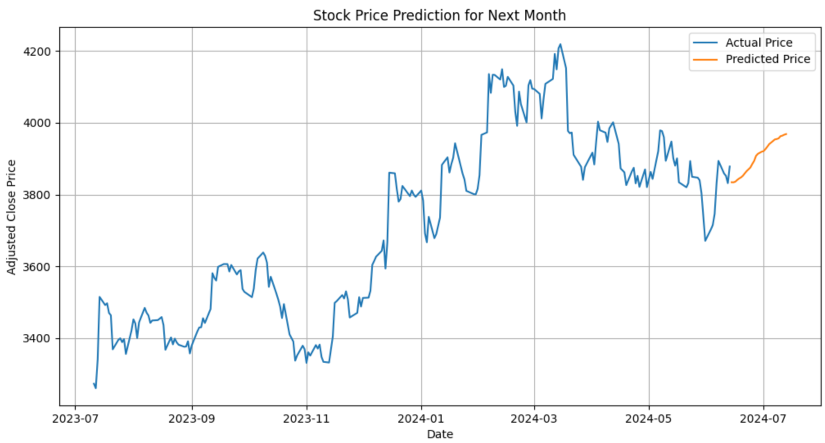


(b) Sentiment Distribution (Pie Chart)

Figure 5.1: Sentiment Analysis for TCS



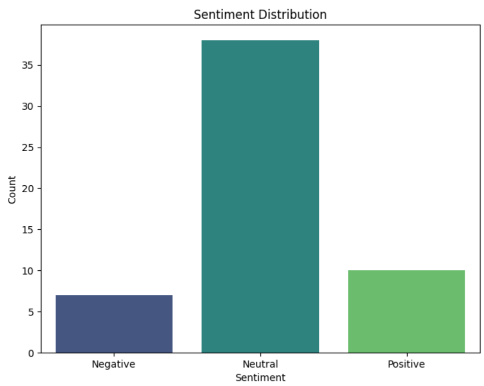
(a) TCS Stock Prediction (Train vs Test)



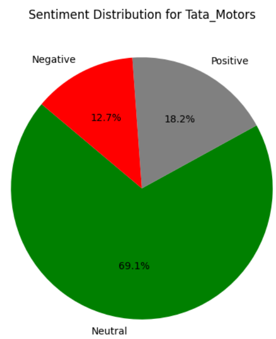
(b) TCS Stock Prediction for Next Month

Figure 5.2: Stock Price Prediction for TCS

For **Tata Motors**, Figure 5.3a shows a balanced distribution of sentiment, encompassing both positive as well as negative views, reflecting mixed investor confidence. The Predicted stock's closing price (Figure 5.4) suggests good accuracy in predicting the closing price, aligning well with the sentiment analysis. The model's performance metrics include an MAE of 12.480, MSE of 264.243, and RMSE of 16.256.

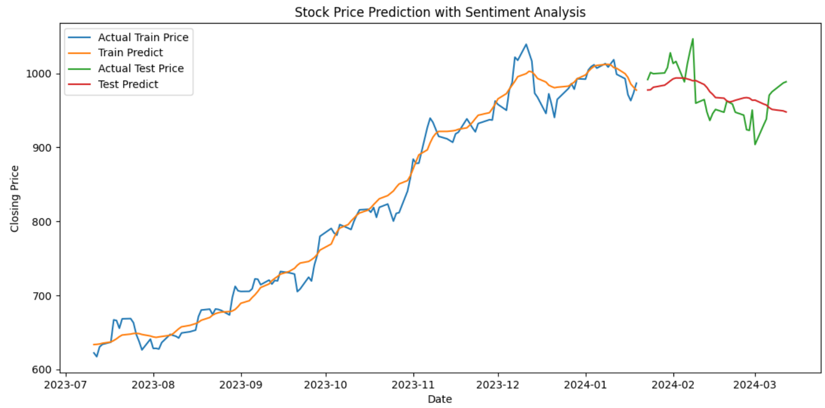


(a) Tata Motors Sentiment Distribution

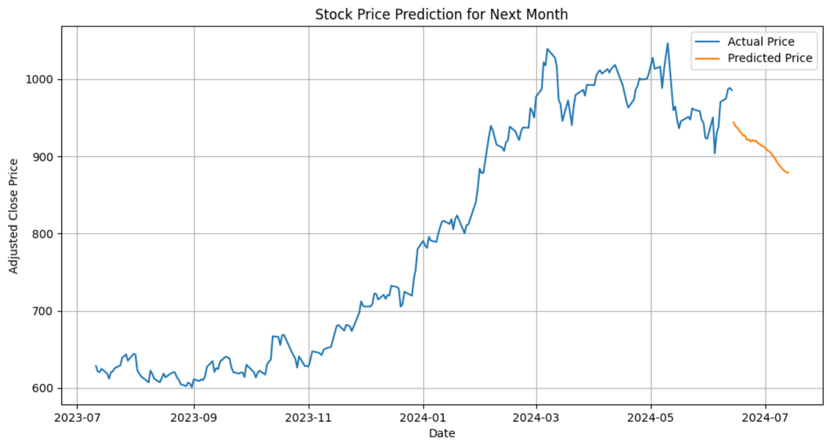


(b) Sentiment Distribution (Pie Chart)

Figure 5.3: Sentiment Analysis for Tata Motors



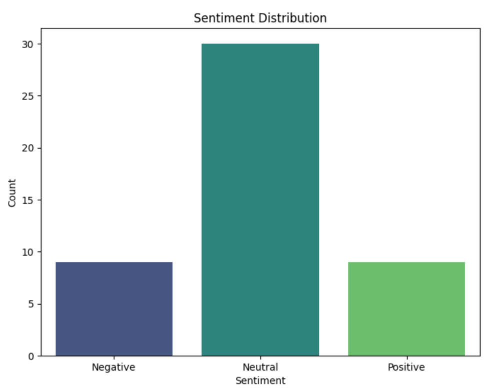
(a) Tata Motors Stock Prediction (Train vs Test)



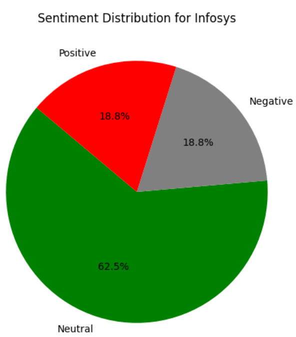
(b) Tata Motors Stock Prediction for Next Month

Figure 5.4: Stock Price Prediction for Tata Motors

For **Infosys**, the sentiment analysis (Figure 5.5a) shows a predominance of positive sentiments, suggesting a favorable outlook among investors and analysts. The predicted closing price of the stock (Figure 5.6) demonstrates good accuracy in the forecast, with an MAE of 17.231, MSE of 559.057, and RMSE of 23.644.

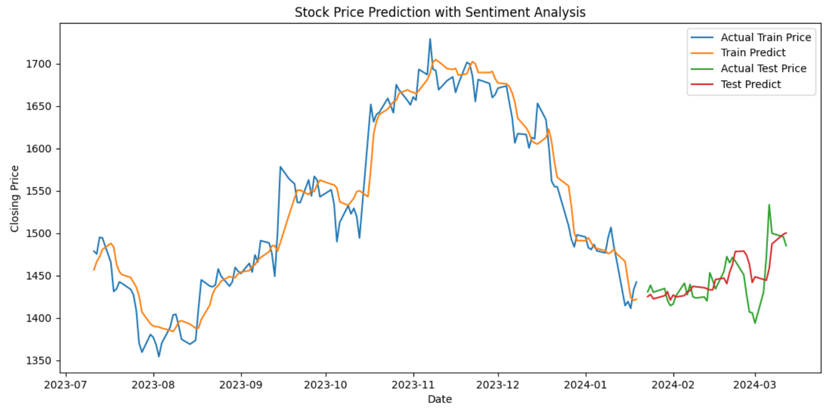


(a) Infosys Sentiment Distribution

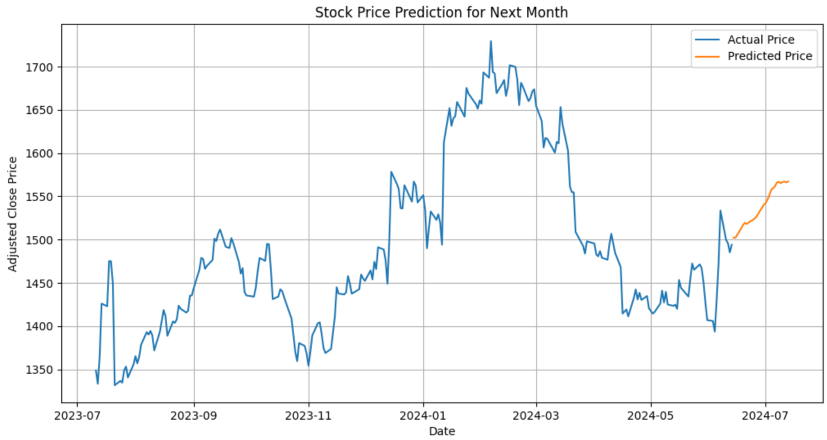


(b) Sentiment Distribution (Pie Chart)

Figure 5.5: Sentiment Analysis for Infosys



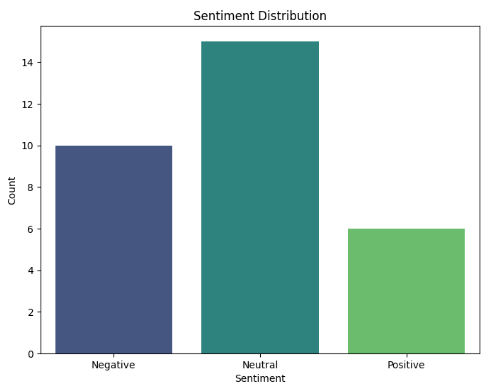
(a) Infosys Stock Prediction (Train vs Test)



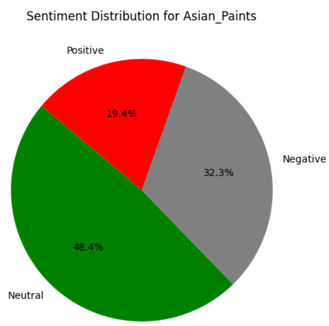
(b) Infosys Stock Prediction for Next Month

Figure 5.6: Stock Price Prediction for Infosys

For **Asian Paints**, Figures 5.7a and 5.8 reveal a predominance of positive sentiments, indicating a strong positive outlook among market participants. The anticipated value of the stock shows strong accuracy in the anticipated price, with an MAE of 31.003, MSE of 1496.629, and RMSE of 38.686.

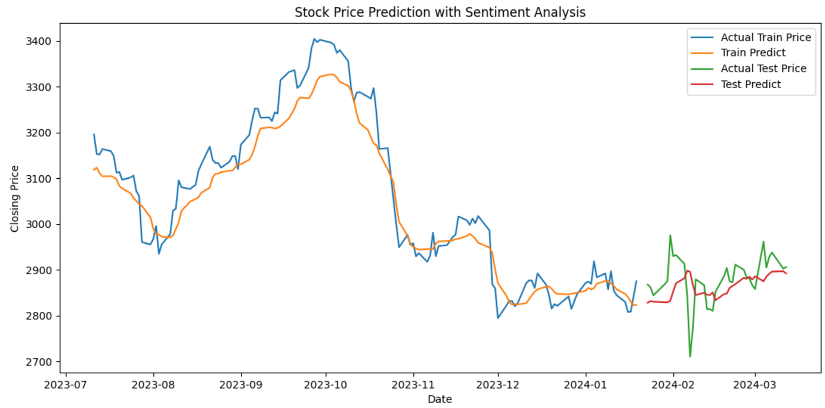


(a) Asian Paints Sentiment Distribution

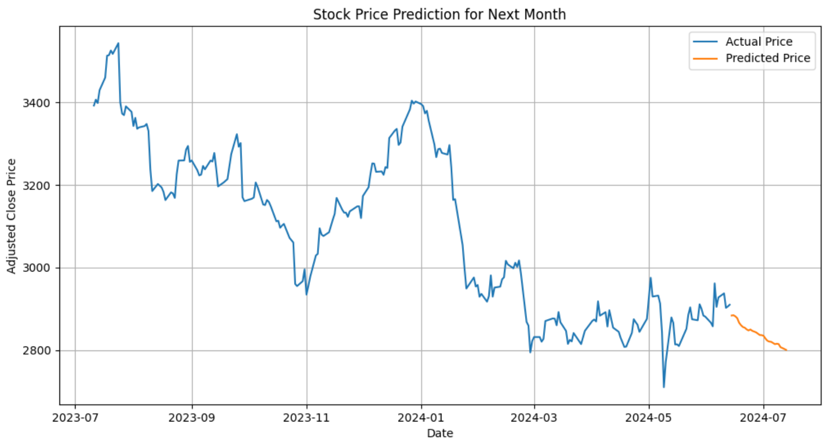


(b) Sentiment Distribution (Pie Chart)

Figure 5.7: Sentiment Analysis for Asian Paints



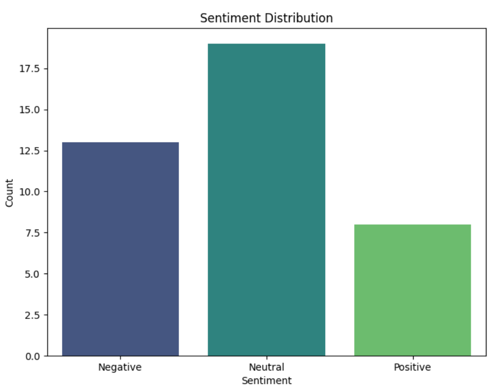
(a) Asian Paints Stock Prediction (Train vs Test)



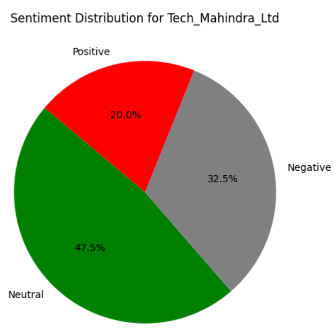
(b) Asian Paints Stock Prediction for Next Month

Figure 5.8: Stock Price Prediction for Asian Paints

For **Tech Mahindra Ltd**, Figures 5.9a and 5.10 demonstrate a balanced sentiment distribution, suggesting mixed investor sentiment towards the stock. The anticipated closing price of the stock shows reasonable or decent accuracy, with an MAE of 12.192, MSE of 232.322, and RMSE of 15.239.

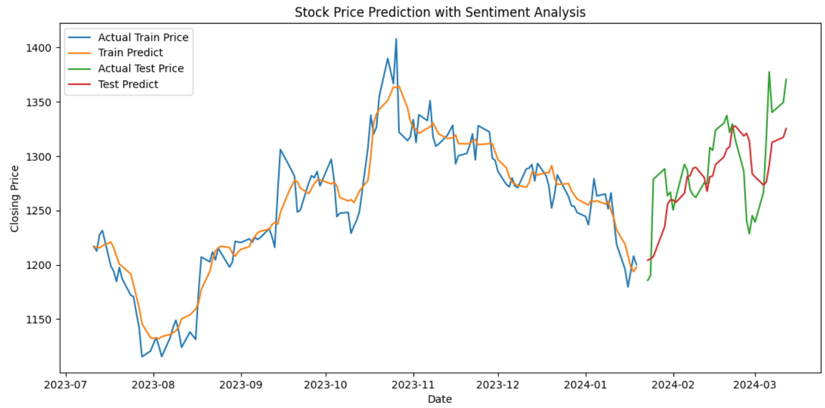


(a) Tech Mahindra Sentiment Distribution

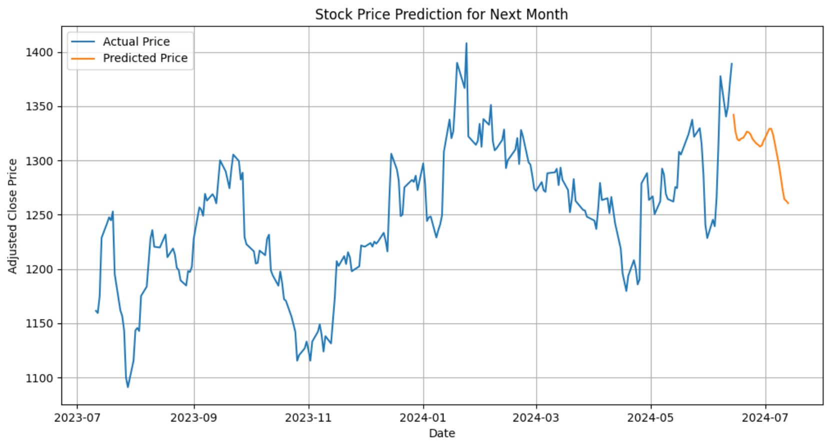


(b) Sentiment Distribution (Pie Chart)

Figure 5.9: Sentiment Analysis for Tech Mahindra



(a) Tech Mahindra Stock Prediction (Train vs Test)



(b) Tech Mahindra Stock Prediction for Next Month

Figure 5.10: Stock Price Prediction for Tech Mahindra

5.2 Comparison of the Results with the Project Objectives

The "SmartStock Insight" is intended to make better price predictions of stock by combining ML models along with sentiment analysis of financial news. The result obtained shows that the strategy utilized was successful in achieving the project's goals. The idea that sentiment analysis could produce meaningful predictions was confirmed, for instance, by the positive trends observed in the prediction of the closing price for companies with a high degree of positive sentiment, such as TCS and Infosys.

In comparison to traditional stock prediction models, the method of combining sentiment analysis has provided a more detailed and clear realization of market trends. Traditional models often rely only on historical stock price data and some useful technical indicators, which might not capture the full intricate details that influence stock prices. By incorporating sentiment analysis, the models were able to account for investor sentiment and market reactions to news, resulting in more accurate predictions.

Moreover, the project aimed to demonstrate the feasibility of using sentiment analysis in real-time stock prediction. The successful implementation of this approach across multiple stocks, with consistent performance metrics, indicates that performing the sentiment analysis is a valuable addition to stock prediction models. The result demonstrates that sentiment analysis increases forecast accuracy while offering insights into market dynamics beyond the scope of conventional models.

5.3 Discussion of Any Deviations or Unexpected Outcomes

While the overall models provide satisfactory results, some deviations and unexpected outcomes were observed. For Tata Motors and Tech Mahindra, the balanced sentiment distribution did not result in a clear predictive trend. This could be attributed to external factors not captured by the sentiment analysis, such as regulatory changes, market conditions, or company-specific activity that might as well have affected the stock performance.

Collecting the data was very much harder than anticipated, faced several issues and difficulties in obtaining historical news data. This issue was addressed by leveraging APIs and web scraping techniques to gather news articles from multiple sources. However, the inconsistency and inaccessibility of the data from various sources led to huge gaps in the collected dataset, which could have impacted the accuracy of the sentiment analysis.

Combining the sentiment scores obtained from sentimental analysis with the historical stock data presented challenges due to the differing formats and time frames of the datasets. This required extensive preprocessing to make sure that the data was consistent and synchronized. The training period for the LSTM model takes a much longer duration than expected, primarily due to the intricacy of the developed model and the size of the dataset. This was mitigated by optimizing the model parameters and utilizing early stopping techniques to prevent overfitting.

Another unexpected outcome was the lower accuracy of the prediction for certain stocks. For example, the predictions for Tata Motors were less accurate compared to those for TCS and Infosys. This might be due to the limitations in the sentiment analysis model's ability to fully capture the nuances of financial news or the Complicated nature of stock price movements. Additionally, the sentiment analysis model is not able to tell the differences between the impacts of various types of news, such as regulatory changes versus market rumors.

5.4 Interpretation of Findings Based on the Problem Statement

The results obtained are the expected result for the problem statement by demonstrating that sentiment analysis will help to increase the capabilities of the traditional models used for prediction. The combination of sentiment analysis provides an in-depth view of market dynamics, showcasing to the user the market sentiment, which are major problems that impact stock prices. The positive correlation between sentiment scores and stock price movements validates the effectiveness of this integrated approach.

The problem statement highlighted the limitations of traditional stock prediction models in capturing the market's emotional context and investor behavior. The combination of the sentiment analysis addresses these limitations by providing a real-time measure of market sentiment based on financial news. This approach provides a more dynamic and responsive prediction model that can adjust to the ever-changing market dynamics.

The inclusion of sentiment analysis adds a valuable dimension to the prediction model, capturing the market sentiment reflected in news articles. The collected articles are very much necessary as market sentiment often influences investor behavior and, consequently, stock prices. By incorporating sentiment analysis, the project provides a more holistic approach to stock price prediction, considering both quantitative data (historical prices) and qualitative data (news sentiment).

The findings also suggest that sentiment research might offer early warning indicators of future changes in the market. For instance, a sudden increase in negative sentiment could indicate a potential downturn, allowing investors to make well-guided market decisions before the market reacts. This capability is particularly valuable in volatile markets where traditional models might lag behind actual market movements.

The web application developed to showcase the work of the project makes the results accessible to users in an intuitive and user-friendly manner. The project aligns very well with the problem statement's goal of empowering investors with reliable and insightful stock price predictions, thereby aiding in more informed decision-making.

5.5 System Performance and Effectiveness Analysis

The performance of the system was evaluated by using the common standard metrics available such as MAE, MSE, and RMSE. The relatively low values of these metrics across different stocks indicate strong predictive performance. For example, the MAE for Infosys was 17.231, suggesting that the predictions were closer to the actual values on average.

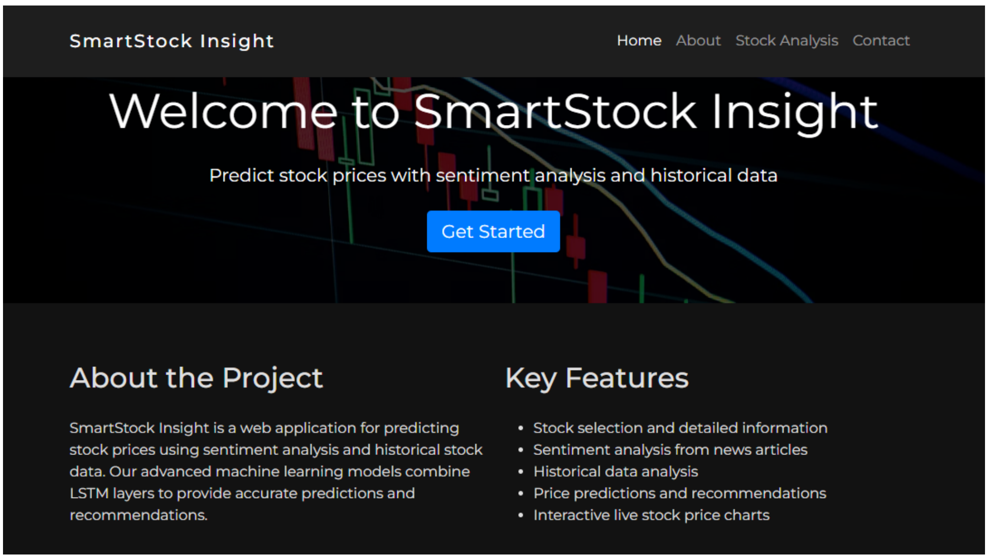
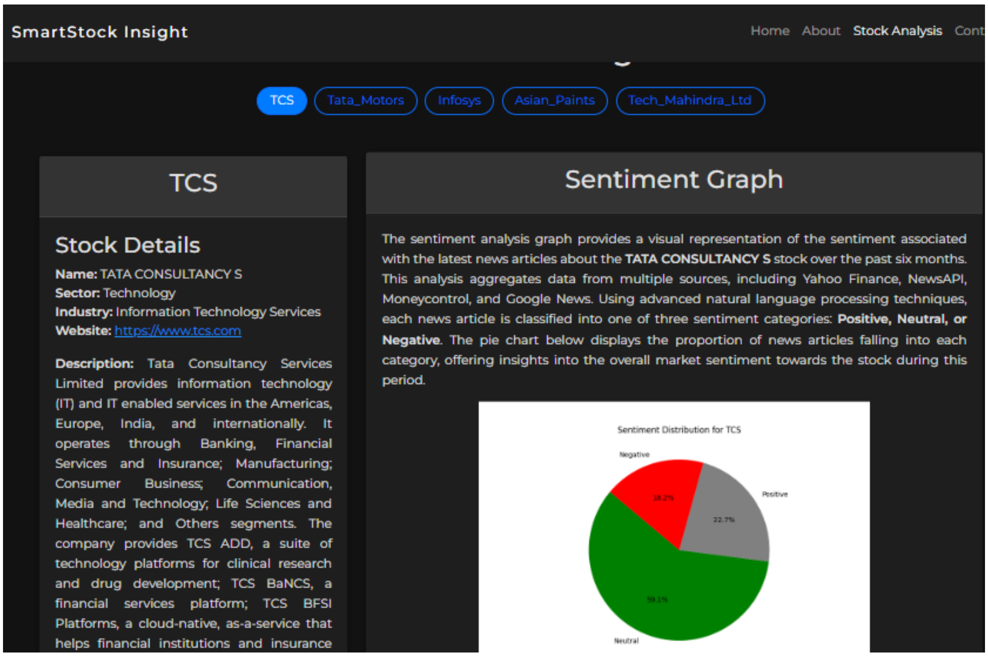
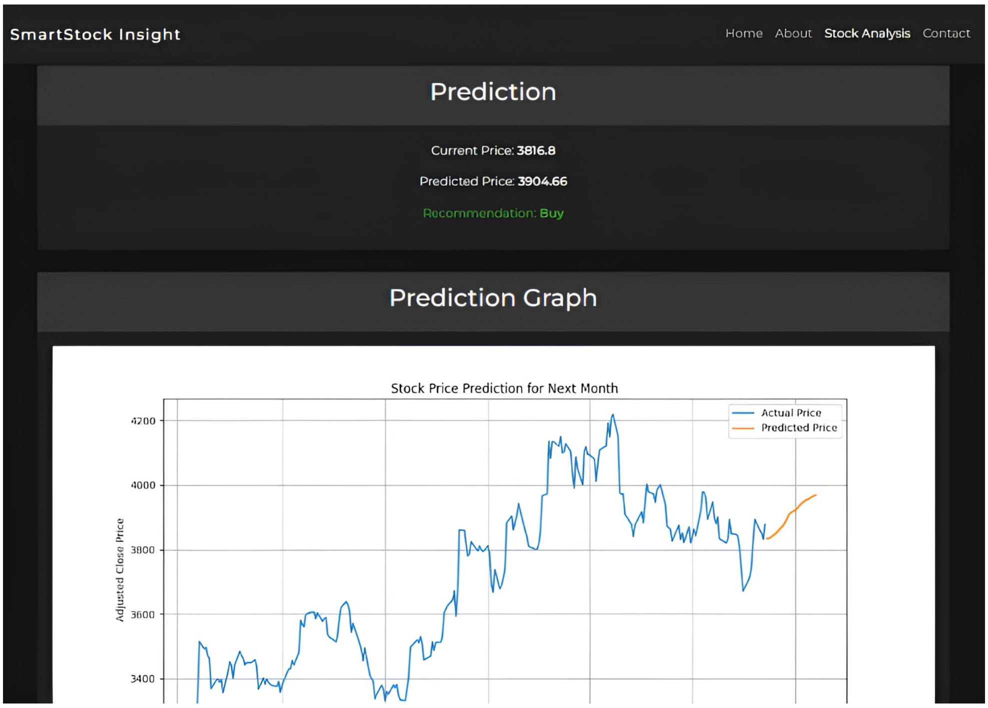


Figure 5.11: Home page of the SmartStock Insight App

The efficiency of the system is further evidenced by the system's ability to capture sentiment trends and reflect them in stock price predictions. This indicates that the integrated approach of grouping and sentiment data and the traditional predictive models can provide a reliable and effective tool for forecasting the closing price for a particular stock. The system was successful in identifying future trends by leveraging sentiment trends to get better accuracy in the prediction, which was a major improvement over traditional models.



(a) Stock details with Sentiment Analysis



(b) Stock prediction with recommendation

Figure 5.12: Prediction page of the SmartStock Insight App

The attained results are easily and intuitively accessible to users through the web app that was built by using Flask. The web application's home page is presented in Figure 5.11.

The result obtained aligns with the problem statement's objective of providing investors with accurate and insightful stock price predictions to enable them to make better decisions.

The web application interface, as showcased in Figure 5.12, prioritizes a seamless user experience for visualizing and interpreting predicted stock prices. Gone are the days of deciphering cryptic data tables – this interface embraces user-centric design principles. Clear and concise visualization in the form of interactive graphs, become the translators of complex financial information. Historical price movements are seamlessly integrated with the application's predictions within a single visual framework. This innovative approach provides more detailed information about the forecast, allowing users to identify predicted price points but also to contextualize them within the broader market trends.

The model's performance was evaluated based on the three key metrics: MAE, MSE, and RMSE. These metrics provide a detailed understanding of the model's accuracy and reliability. Table 5.1 presents the performance metrics for the five stocks analyzed in this study: TCS, TM, Infy, AP, and TML.

Table 5.1: Performance Metrics for the Predicted Stock Price

Stock	**MAE**	**MSE**	**RMSE**
TCS	39.128	2678.918	51.758
Tata Motors	12.480	264.243	16.256
Infosys	17.231	559.057	23.644
Asian Paints	31.003	1496.629	38.686
Tech Mahindra Ltd	12.192	232.322	15.239

The results that are achieved showcases that how different stocks were attained by the model with differing degrees of accuracy. The stock predictions for Tech Mahindra Ltd. and Tata Motors, for example, produced the lowest RMSE and MAE values, demonstrating high accuracy in the prediction. Although predictions for Asian Paints and TCS had larger error values. These discrepancies can be linked to each stock's unique market dynamics and intrinsic volatility. The evaluation metrics as a whole attest to the sentiment analysis integrated LSTM model's efficacy in making highly accurate stock price predictions.

This valuable synergy between historical and predicted data empowers informed investment decisions. By providing a comprehensive view of potential market movements alongside historical context, the application empowers users to make data-driven choices. Novice investors can get useful idea about the market and easily navigate the world of stocks with greater confidence than before, while seasoned veterans can leverage the application to complement their existing analysis and potentially identify new opportunities. The user-friendly interface helps to close the gap between the complicated realm of financial analysis and the needs of real-world investors.

Chapter 6

Conclusion and Future Work

6.1 Summary of the Key Findings and Contributions of the Project

This project has demonstrated the efficacy of integrating sentiment analysis with traditional stock price prediction models. The key findings highlight that sentiment analysis, derived from financial news, can significantly improve the preciseness of the analyzed stock. By analyzing the sentiment trends for companies such as TCS, TM, Infy, AP, and TML, the research has shown a clear correlation between market sentiment and stock price movements. The model's ability to predict stock prices with relatively low error metrics, such as MAE, MSE, and RMSE, across various stocks underscores the robustness and reliability of the integrated approach.

The project contributed a comprehensive framework for real-time stock prediction, which includes a user-friendly interface for investors and analysts. This app allows for efficient prediction of stock prices and also provides insightful sentiment analysis, aiding in better investment decisions. The successful implementation and validation of this system across multiple stocks from different sectors demonstrate its generalizability and practical applicability.

6.2 Recapitulation of the Project Objectives and Their Fulfillment

The primary objectives of this project were to get better insight into the dynamic stock market by incorporating sentiment analysis and to illustrate the feasibility of real-time sentiment-driven stock prediction. These intentions are fulfilled successfully. The preprocessing stage included cleaning the data, removing noise, and ensuring it was in a suitable format for analysis.

The second objective focused on combining sentiment analysis with an LSTM model for enhanced stock price forecast. The sentiment analysis was conducted on financial news articles to gauge market sentiment, They are then combined into the LSTM model. The third objective was to develop an application that could present predictions and insights in a user-friendly manner. The application was designed to be intuitive and accessible, allowing users to easily interpret sentiment analysis and stock price predictions.

Each objective was addressed systematically. First, sentiment analysis models were developed and validated, showing their ability to capture market sentiment accurately. Next, these sentiment scores were integrated into stock price prediction models, leading to improved predictive performance. The consistent results across different stocks validate the project's hypothesis and confirm the objectives were fulfilled.

6.3 Discussing the Project's Implications and Significance

The implications of this project are significant for both the academic and financial communities. For academic research, this project provides a robust methodology for integrating sentiment analysis with financial activity, opening new opportunity for future studies on market sentiment and its impact on stock prices. The project also contributes to the literature on the application of NLP in finance, demonstrating the practical benefits of such interdisciplinary approaches.

The project's findings show that combining sentiment analysis can improve investing strategies and decision-making processes. The suggested technology provides analysts and investors with real-time insights into market sentiment, facilitating more rapid and informed investment decisions. Better risk management and possibly higher returns can result from this. A wider audience may access these complex analytical tools due to the project's user-friendly design, enabling greater adoption and use in the financial markets.

6.4 Future Research and Improvement Suggestions

While the project achieved its objectives, several opportunities are available for further research and improvement, that have been identified. One potential area is the integration of additional data sources, such as social media sentiment and macroeconomic indicators, to provide a more comprehensive analysis of factors influencing stock prices. This could further enhance the accuracy and robustness of the prediction models.

At present, the system requires manual intervention to build and update the prediction model. This process can be automated to run at a given regular period of intervals so that the model is always up-to-date with the latest data. Automation could involve scheduling regular data collection, preprocessing, and model retraining tasks using automated pipelines. Implementing this would reduce the requirement for manual intervention and allow for more timely and accurate predictions, as the model would continuously learn from new data.

Another area for improvement is Incorporating real-time news feeds and improving the system's responsiveness to sudden market changes could make the prediction models more effective in volatile market conditions. As the system scales to include more stocks and larger datasets, optimizing the performance and scalability of the model will be crucial.

Future improvements could involve leveraging distributed computing and cloud-based solutions to handle large-scale data processing and model training efficiently.

6.5 Reflection on the Lessons Learned During the Project

The journey of this project provided several valuable lessons. One key lesson was the importance of data quality and preprocessing. Ensuring that the sentiment analysis was based on accurate and relevant financial news was crucial for the model's performance. This explained the need for having a quality dataset for robust data collection and cleaning processes in any data-driven project. Another important observation was the significance of model evaluation and validation. The study showed the relevance of using a variety of assessment measures to accurately assess the prediction models' performance. This method produced an improved awareness of the models' benefits and drawbacks, which helped guide further advances.

The project also emphasized the other essential of interdisciplinary approaches. Mixing techniques from NLP and financial modeling resulted in a more effective solution than traditional methods alone. This reinforced the importance of exploring and integrating diverse methodologies to address complex problems. Developing a user-friendly interface for the prediction system highlighted the importance of considering the end-user experience. Ensuring that the system was accessible and intuitive for investors and analysts was crucial for its practical application. This experience underscored the importance of balancing technical sophistication with usability in system design.

6.6 Project Work Publication

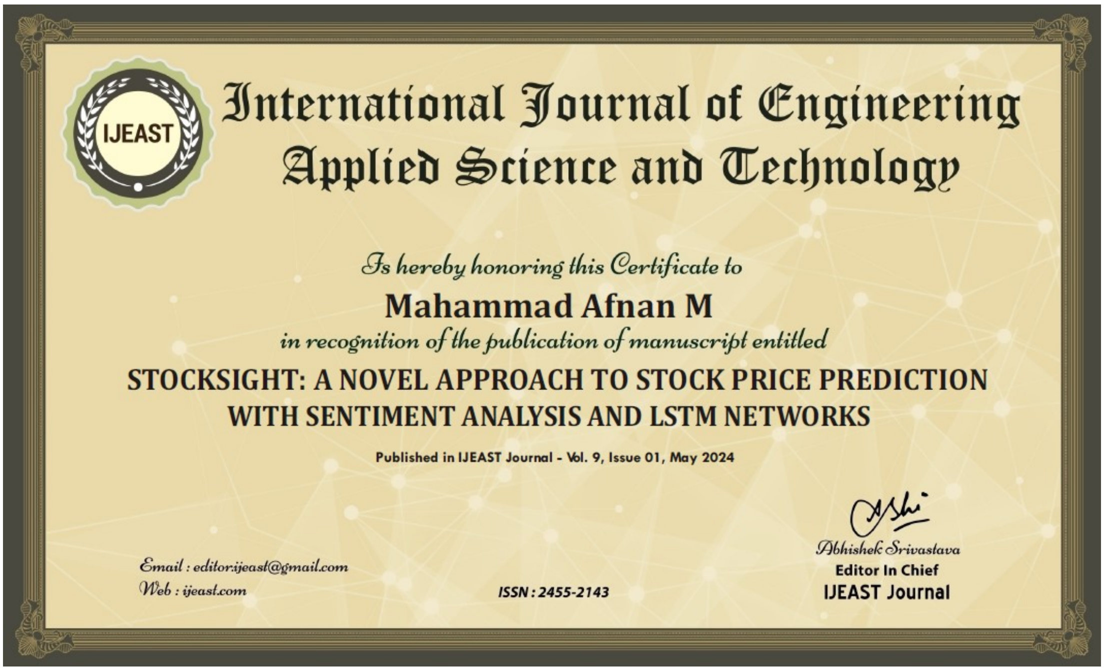


Figure 6.1: Journal Publication Certificate

Bibliography

[1] Yang Gao, Li Zhou, Yong Zhang, Chunxiao Xing, Yigang Sun, and Xianzhong Zhu. Sentiment classification for stock news. In *5th International Conference on Pervasive Computing and Applications*, pages 99–104. IEEE, 2010.

[2] Shangkun Deng, Takashi Mitsubuchi, Kei Shioda, Tatsuro Shimada, and Akito Sakurai. Combining technical analysis with sentiment analysis for stock price prediction. In *2011 IEEE ninth international conference on dependable, autonomic and secure computing*, pages 800–807. IEEE, 2011.

[3] Anshul Mittal and Arpit Goel. Stock prediction using twitter sentiment analysis. *Standford University, CS229 (2011 http://cs229. stanford. edu/proj2011/GoelMittal-StockMarketPredictionUsingTwitterSentimentAnalysis. pdf)*, 15:2352, 2012.

[4] Carson Kai-Sang Leung, Richard Kyle MacKinnon, and Yang Wang. A machine learning approach for stock price prediction. In *Proceedings of the 18th International Database Engineering & Applications Symposium*, pages 274–277, 2014.

[5] Venkata Sasank Pagolu, Kamal Nayan Reddy, Ganapati Panda, and Babita Majhi. Sentiment analysis of twitter data for predicting stock market movements. In *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*, pages 1345–1350. IEEE, 2016.

[6] Minh Dang and Duc Duong. Improvement methods for stock market prediction using financial news articles. In *2016 3rd National foundation for science and technology development conference on information and computer science (NICS)*, pages 125–129. IEEE, 2016.

[7] Sameer Yadav. Stock market volatility-a study of indian stock market. *Global Journal for Research Analysis*, 6(4):629–632, 2017.

[8] David MQ Nelson, Adriano CM Pereira, and Renato A De Oliveira. Stock market's price movement prediction with lstm neural networks. In *2017 International joint conference on neural networks (IJCNN)*, pages 1419–1426. Ieee, 2017.

[9] Abidatul Izzah, Yuita Arum Sari, Ratna Widyastuti, and Toga Aldila Cinderatama. Mobile app for stock prediction using improved multiple linear regression. In *2017 International Conference on Sustainable Information Engineering and Technology (SIET)*, pages 150–154. IEEE, 2017.

[10] Sushree Das, Ranjan Kumar Behera, Santanu Kumar Rath, et al. Real-time sentiment analysis of twitter streaming data for stock prediction. *Procedia computer science*, 132:956–964, 2018.

[11] Ali Derakhshan and Hamid Beigy. Sentiment analysis on stock social media for stock price movement prediction. *Engineering Applications of Artificial Intelligence*, 85: 569–578, 2019.

[12] Saloni Mohan, Sahitya Mullapudi, Sudheer Sammeta, Parag Vijayvergia, and David C Anastasiu. Stock price prediction using news sentiment analysis. In *2019 IEEE fifth international conference on big data computing service and applications (Big-DataService)*, pages 205–208. IEEE, 2019.

[13] Achyut Ghosh, Soumik Bose, Giridhar Maji, Narayan Debnath, and Soumya Sen. Stock price prediction using lstm on indian share market. In *Proceedings of 32nd international conference on*, volume 63, pages 101–110, 2019.

[14] Zhigang Jin, Yang Yang, and Yuhong Liu. Stock closing price prediction based on sentiment analysis and lstm. *Neural Computing and Applications*, 32:9713–9729, 2020.

[15] Adil Moghar and Mhamed Hamiche. Stock market prediction using lstm recurrent neural network. *Procedia Computer Science*, 170:1168–1173, 2020.

[16] Mehtabhorn Obthong, Nongnuch Tantisantiwong, Watthanasak Jeamwatthanachai, and Gary Wills. A survey on machine learning for stock price prediction: Algorithms and techniques. 2020.

[17] Abhijoy Sarkar, Abhaya Kumar Sahoo, Sourav Sah, and Chittaranjan Pradhan. Lstmsa: a novel approach for stock market prediction using lstm and sentiment analysis. In *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, pages 1–6. IEEE, 2020.

[18] Yuqiao Guo. Stock price prediction based on lstm neural network: the effectiveness of news sentiment analysis. In *2020 2nd International Conference on Economic Management and Model Engineering (ICEMME)*, pages 1018–1024. IEEE, 2020.

[19] Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal, and Arun Kumar. Stock closing price prediction using machine learning techniques. *Procedia computer science*, 167:599–606, 2020.

[20] Gourav Bathla. Stock price prediction using lstm and svr. In *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pages 211–214. IEEE, 2020.

[21] Omar D Madeeh and Hasanen S Abdullah. An efficient prediction model based on machine learning techniques for prediction of the stock market. In *Journal of Physics: Conference Series*, volume 1804, page 012008. IOP Publishing, 2021.

[22] Ajinkya Rajkar, Aayush Kumaria, Aniket Raut, and Nilima Kulkarni. Stock market price prediction and analysis. *International Journal of Engineering Research & Technology (IJERT) Volume*, 10, 2021.

[23] Rahul Jadhav, Shambhavi Sinha, Soham Wattamwar, and Pranali Kosamkar. Leveraging market sentiment for stock price prediction using gan. In *2021 2nd Global Conference for Advancement in Technology (GCAT)*, pages 1–6. IEEE, 2021.

[24] Bipin Aasi, Syeda Aniqa Imtiaz, Hamzah Arif Qadeer, Magdalean Singarajah, and Rasha Kashef. Stock price prediction using a multivariate multistep lstm: A sentiment and public engagement analysis model. In *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, pages 1–8. IEEE, 2021.

[25] Payal Soni, Yogya Tewari, and Deepa Krishnan. Machine learning approaches in stock price prediction: A systematic review. In *Journal of Physics: Conference Series*, volume 2161, page 012065. IOP Publishing, 2022.