VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Project Report on

INVEST IQ: Smart Stock Market Analysis and Recommendation System

In partial fulfilment of the Fourth Year, Bachelor of Engineering (B.E.) Degree in Computer Engineering at the University of Mumbai

Academic Year 2022-23

Submitted by
Sachin Kundal - D17A - 34
Tarun Sharma - D17C - 57
Sunny Bhatia - D17C - 72

Project Mentor
Mrs.Priyanka Shah

(2024-25)

VIVEKANAND EDUCATION SOCIETY'S INSTITUTE OF TECHNOLOGY
Department of Computer Engineering

Certificate

This is to certify that **Sachin Kundal (D17A, 34), Tarun Sharma (D17C, 57), Sunny Bhatia (D17C, 72),** of Fourth Year Computer Engineering studying under the University of Mumbai have satisfactorily completed the project on "**INVEST IQ: Smart Stock Market Analysis and Recommendation System** " as a part of their coursework of PROJECT-II for Semester-VIII under the guidance of their mentor **Mrs. Priyanka Shah** in the year 2024-25 .

This project report entitled **INVEST IQ: Smart Stock Market Analysis and Recommendation System** by ***Sachin Kundal, Tarun Sharma, Sunny Bhatia*** is approved for the degree of **B.E. Computer Engineering.**

Programme Outcomes	Grade
PO1,PO2,PO3,PO4,PO5,PO6,PO7, PO8, PO9, PO10, PO11, PO12 PSO1, PSO2	

Date:
Project Guide:

Project Report Approval
For
B. E (Computer Engineering)

This project report entitled **INVEST IQ: Smart Stock Market Analysis and Recommendation System** by ***Sachin Kundal, Tarun Sharma, Sunny Bhatia*** is approved for the degree of **B.E. Computer Engineering.**

Internal Examiner

External Examiner

Head of the Department

Principal

Date:
Place: Mumbai

Declaration

We declare that this written submission represents our ideas in our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Sachin Kundal (34)

Sunny Bhatia (72)

Tarun Sharma (57)

Date:

ACKNOWLEDGEMENT

We are thankful to our college Vivekanand Education Society's Institute of Technology for considering our project and extending help at all stages needed during our work of collecting information regarding the project.

It gives us immense pleasure to express our deep and sincere gratitude to Assistant Professor **Mrs. Priyanka Shah** (Project Guide) for her kind help and valuable advice during the development of project synopsis and for her guidance and suggestions.

We are deeply indebted to Head of the Computer Department **Dr. (Mrs.) Nupur Giri** and our Principal **Dr. (Mrs.) J. M. Nair ,** for giving us this valuable opportunity to do this project.

We express our hearty thanks to them for their assistance without which it would have been difficult in finishing this project synopsis and project review successfully.

We convey our deep sense of gratitude to all teaching and non-teaching staff for their constant encouragement, support and selfless help throughout the project work. It is a great pleasure to acknowledge the help and suggestion, which we received from the Department of Computer Engineering.

We wish to express our profound thanks to all those who helped us in gathering information about the project. Our families too have provided moral support and encouragement several times.

Computer Engineering Department
COURSE OUTCOMES FOR B.E PROJECT

Learners will be to,

Course Outcome	Description of the Course Outcome
CO 1	Able to apply the relevant engineering concepts, knowledge and skills towards the project.
CO2	Able to identify, formulate and interpret the various relevant research papers and to determine the problem.
CO 3	Able to apply the engineering concepts towards designing solutions for the problem.
CO 4	Able to interpret the data and datasets to be utilised.
CO 5	Able to create, select and apply appropriate technologies, techniques, resources and tools for the project.
CO 6	Able to apply ethical, professional policies and principles towards societal, environmental, safety and cultural benefit.
CO 7	Able to function effectively as an individual, and as a member of a team, allocating roles with clear lines of responsibility and accountability.
CO 8	Able to write effective reports, design documents and make effective presentations.
CO 9	Able to apply engineering and management principles to the project as a team member.
CO 10	Able to apply the project domain knowledge to sharpen one’s competency.
CO 11	Able to develop a professional, presentational, balanced and structured approach towards project development.
CO 12	Able to adopt skills, languages, environment and platforms for creating innovative solutions for the project.

Index

List of Figures:

List of Tables:

Abstract

This report presents a hybrid stock recommendation system that integrates traditional machine learning and deep learning models to forecast stock price movements and provide intelligent investment advice. Using historical stock data sourced from Yahoo Finance, including features like opening/closing prices, highs, lows, and volume, the system applies a thorough preprocessing pipeline to clean and engineer data for optimal performance. The predictive framework utilizes five models—XGBoost, Gradient Boosting, Random Forest, CNN, and LSTM—selected for their strengths in handling structured and sequential data.

Ensemble models excel in capturing short-term trends and non-linear patterns, while deep learning models, particularly LSTM, effectively model long-term dependencies in financial time-series data. These models are evaluated using standard metrics such as Mean Squared Error (MSE), Root Mean Square Error (RMSE), and R-squared (R^2) to ensure reliable forecasting.

The system features a user-friendly interface that enables users to input stock symbols, select forecasting models, and visualize trends. Based on model predictions, it provides clear recommendations—Buy, Sell, or Hold—bridging the gap between complex analytical processes and user-friendly investment tools.

By combining multiple modeling techniques, this approach enhances the accuracy and interpretability of stock predictions. Future enhancements include integrating sentiment analysis from news and social media sources, and applying reinforcement learning to refine the decision-making logic. This work contributes to the development of scalable, modular, and intelligent financial platforms that support informed investment decisions through advanced analytics.

Chapter 1: Introduction

1.1. Introduction:

The stock market is a dynamic and complex environment, heavily influenced by a wide range of economic, political, and psychological factors. Predicting market trends and making informed investment decisions in such a volatile setting is a significant challenge. However, with the rapid advancement of artificial intelligence (AI), machine learning (ML), and deep learning (DL), it has become increasingly possible to analyze vast amounts of financial data and extract meaningful patterns that can support accurate forecasting and strategic recommendations.

This project aims to develop a hybrid stock recommendation system that utilizes both traditional machine learning models and advanced deep learning architectures to analyze historical financial data and forecast future stock price movements. The system provides intelligent investment suggestions such as "Buy," "Sell," or "Hold," based on the predictive insights generated by multiple models. By combining different analytical approaches, the system achieves improved accuracy and interpretability, offering investors a reliable tool for decision-making.

The stock data used in this study is sourced from Yahoo Finance, which provides comprehensive historical records for various stocks, including attributes like opening price, closing price, highest and lowest prices of the day, trading volume, and other technical indicators. Preprocessing techniques are employed to clean the data, handle missing values, normalize numerical values, and engineer features that reflect market behavior more accurately.

The core of the system involves the implementation of five powerful models: XGBoost, Gradient Boosting, Random Forest, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks. These models were selected due to their unique strengths—ensemble models for structured and tabular data, and deep learning models for capturing complex temporal patterns and long-term dependencies in time-series data.

An important feature of this project is the inclusion of a graphical user interface (GUI) that allows users to interact with the system without needing in-depth technical knowledge. Users can input a stock ticker, choose the desired predictive model, and receive visualized forecasts and actionable investment suggestions.

By merging the strengths of traditional and modern AI techniques, this project not only enhances the reliability of stock market predictions but also makes advanced financial analytics accessible to a wider audience. The system lays the foundation for further innovations in AI-powered financial technology.

1.2. Motivation:

The motivation behind the development of our stock recommendation system stems from the growing need for accessible, data-driven financial tools in an increasingly complex and volatile market. As more individuals turn to investing—especially in the digital age—there is a rising demand for systems that can offer intelligent, real-time guidance without requiring deep financial or technical expertise. However, predicting stock movements remains a daunting task due to the dynamic interplay of economic indicators, geopolitical events, and human behavior that drives market fluctuations.

Traditional methods of stock analysis often rely on manual interpretation of charts and financial statements, which can be overwhelming and error-prone for average investors. Our aim is to bridge this gap by harnessing the power of machine learning and deep learning to automate prediction and recommendation processes. By doing so, we seek to democratize access to financial insight, enabling everyday users to make informed investment decisions.

We view technology not just as a tool, but as an enabler of financial literacy and empowerment. Through our hybrid approach and user-friendly interface, we hope to provide a reliable decision-support system that enhances confidence, reduces risk, and fosters smarter, more inclusive participation in the financial markets.

1.3. Problem Definition:

The stock market is inherently volatile and influenced by a wide array of factors such as economic indicators, political events, and investor sentiment. Accurately predicting stock price movements is a complex challenge that requires the analysis of large volumes of time-series financial data. While expert investors may have the skills and tools to interpret market trends, average retail investors often lack the resources or technical knowledge to make informed decisions. Traditional forecasting methods are limited in their ability to adapt to dynamic market conditions and may fail to capture the nonlinear relationships present in financial data.

Moreover, existing stock analysis platforms often provide raw data without actionable insights or user-friendly interfaces, making it difficult for non-technical users to interpret and act on market information. There is a clear need for an intelligent, accurate, and accessible system that can analyze historical stock data, forecast future trends, and offer clear investment recommendations.

This project addresses that need by developing a hybrid stock recommendation system that integrates both machine learning and deep learning models, supported by an intuitive user interface. The goal is to enhance predictive accuracy while making the insights understandable and usable for a broader audience of investors.

1.4. Existing Systems:

In the current landscape, several stock prediction and analysis platforms are available, ranging from basic tools to advanced trading software. Most existing systems rely on traditional statistical methods such as moving averages, linear regression, or technical indicators like RSI and MACD to interpret market trends. While these tools are useful for experienced traders, they often lack adaptability to the complex, nonlinear patterns of modern financial markets. Furthermore, many of these systems are either too simplistic or overly complex, offering little middle ground for retail investors seeking reliable and user-friendly solutions.

Some advanced platforms incorporate machine learning models, but they typically focus on a single algorithm or limited data features, resulting in less accurate or generalized predictions. Additionally, these systems rarely combine both machine learning and deep learning models, and few offer an integrated interface that allows users to visualize data, interact with models, and receive direct investment recommendations.

The majority of existing tools also do not support personalized analysis or model selection based on user preference, limiting their flexibility. Consequently, there is a significant gap between the analytical capabilities of current systems and the growing need for intelligent, intuitive, and accessible stock prediction platforms for all types of investors.

1.5. Lacuna of the Existing System:

Despite the availability of various stock market analysis tools, existing systems exhibit several critical limitations that reduce their effectiveness and accessibility. Most traditional platforms rely heavily on technical indicators and statistical models that assume linear market behavior. These approaches often fail to capture the nonlinear, time-dependent, and complex patterns that characterize real-world stock movements, resulting in inaccurate or delayed predictions.

Furthermore, many existing systems are model-specific, meaning they use only one algorithm for forecasting without leveraging the strengths of multiple techniques. This restricts their adaptability and reduces prediction robustness across different market conditions. Advanced tools that do employ machine learning are often black-box models with limited interpretability, and are not user-friendly for non-expert investors.

Another major drawback is the lack of an integrated, interactive interface. Most tools present raw data and charts without actionable insights, making them difficult to use for beginners. They also lack personalized recommendations or customizable forecasting options. In addition, these systems generally do not support real-time analysis or incorporate external influences such as market sentiment from news or social media.

These limitations highlight the need for a more comprehensive, intelligent, and user-centric system that can bridge the gap between advanced analytics and practical investment decision-making.

1.6. Relevance of the Project:

In today's fast-evolving financial markets, the ability to make accurate and timely investment decisions is more important than ever. With the increasing participation of retail investors and the vast amount of real-time market data available, there is a growing demand for intelligent systems that can analyze complex financial information and generate reliable predictions. Traditional investment strategies often fall short in adapting to dynamic market trends and require expert knowledge, creating a barrier for average investors.

This project is highly relevant as it addresses these challenges by combining the strengths of machine learning and deep learning to develop a predictive stock recommendation system. By integrating models such as XGBoost, Gradient Boosting, Random Forest, CNN, and LSTM, the system leverages both structured financial indicators and time-series forecasting capabilities. This hybrid approach significantly improves prediction accuracy and enhances the interpretability of results.

Moreover, the use of historical data from Yahoo Finance ensures that the system is grounded in real-world, up-to-date information. The inclusion of a graphical user interface (GUI) further enhances the relevance of the project by making complex financial analysis accessible to users without technical backgrounds.

Ultimately, this project contributes to the field of financial technology (FinTech) by providing a practical, scalable, and intelligent solution for modern stock market forecasting and decision-making.

Chapter 2: Literature Survey

A.<u>Overview of literature survey:</u>

The literature survey highlights various studies applying machine learning and deep learning techniques for stock market prediction. Research compares models like SVM, Random Forest, LSTM, ARIMA, and ANN in terms of prediction accuracy and performance. Surveys emphasize the role of deep learning architectures such as RNN, GNN, and Transformers. Other studies focus on real-time forecasting using streaming data and hybrid models, while some integrate sentiment analysis from news and social media to enhance predictions. Additionally, portfolio optimization using machine learning is explored, showing how data-driven strategies can improve investment outcomes and risk-adjusted returns.

2.1. Research Papers :

1. **Jena, M., Behera, R.K., Rath, S.K. (2020). Machine Learning Models for Stock Prediction Using Real-Time Streaming Data. In: Dehuri, S., Mishra, B., Mallick, P., Cho, SB., Favorskaya, M. (eds).**

a) **Abstract:** In recent years stock prediction has attracted a lot of attention to the researchers in financial sectors. Apart from the static log data, streaming data has also been proven to be a perennial source of data analysis collected in real-time, which basically deals with the continuous flow of data carrying information from sources like websites, mobile phone applications, server logs, social websites, trading floors, etc. The classifying model made out of historical data can be relentlessly honed to give even more accurate results since its outcome is always compared to the next tick of the clock. In this study, an attempt is made to develop machine learning models to predict the potential prices of a company's stock which helps in making financial decisions. Spark streaming has been considered for the processing of humongous data and data ingestion tools like NodeJS have been further used for analysis. Earlier researches are made on the same concept but the present goal of the study is to develop such a model that is scalable, fault-tolerant and has a lower latency. The model rests on a distributed computing architecture called the Lambda Architecture which helps in attaining the goals as intended. Upon analysis, it is found that prediction of stock values is more accurate when support vector regression is applied. The historical stock values are considered as supervised datasets for training the models.

b) **Inference:** This paper focuses on the implementation of machine learning models for real-time stock prediction, emphasizing the complexities involved in processing streaming financial data. It addresses key challenges such as high data velocity and large volumes, which are critical in ensuring timely and accurate predictions. The study compares the performance of models like Support Vector Machine (SVM), Random Forest, and Long Short-Term Memory (LSTM), highlighting how each model handles real-time constraints and their respective strengths in forecasting dynamic stock market behavior.

2. Singh, R., & Goyal, N.: Stock Price Prediction using Machine Learning and Deep Learning. In: 2023 IEEE 5th International Conference on Recent Advances in Information Technology (RAIT), pp. 234-239. IEEE (2023)

a) **Abstract:** The application of machine learning in stock market forecasting is a new trend, which produces forecasts of the current stock market prices by training on their prior values. This paper aims to implement Machine learning and Deep learning algorithms in real-time situations like stock price forecasting and prediction. The focus of this project is to forecast the stock price of Reliance Industries Limited (RELIANCE.NS) using the ARIMA model for up to 2 years and to predict the next day's stock price using Random Forest and LSTM to predict stock prices for test data. The following parameters were used to train the models: open, close, low, high, volume, and adjusted close.

b) **Inference:** This IEEE paper provides a comparative analysis of multiple machine learning and deep learning models, such as ARIMA, Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks, for stock price prediction. It critically evaluates each model's performance, highlighting their respective advantages and limitations. The study emphasizes how ARIMA performs well with linear data patterns, while ANN and LSTM excel in capturing complex, nonlinear, and sequential dependencies, making them more suitable for dynamic financial time-series forecasting tasks.

3. Zhang, J., & Hu, Y.: Stock Market Prediction via Deep Learning Techniques: A Survey. arXiv preprint arXiv:2212.12717 (2023).

a) **Abstract:** Existing surveys on stock market prediction often focus on traditional machine learning methods instead of deep learning methods. This motivates us to provide a structured and comprehensive overview of the research on stock market prediction. We present four elaborated subtasks of stock market prediction and propose a novel taxonomy to summarize the state-of-the-art models based on deep neural networks. In addition, we also provide detailed statistics on the datasets and evaluation metrics commonly used in the stock market. Finally, we point out several future directions by sharing some new perspectives on stock market prediction.

b) **Inference:** This survey offers an extensive evaluation of various deep learning models applied to stock market prediction, including Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), Graph Neural Networks (GNN), and Transformer architectures. It provides a thorough comparison of how each model processes sequential and structured financial data, their suitability for capturing temporal dependencies, and their effectiveness in different market scenarios. The paper serves as a valuable resource for understanding the strengths, limitations, and practical applications of deep learning in financial forecasting.

4. **T. Sugadev, N. S. Hameed, S. Vijayakumar, P. Tamilarasan and M. S. Islam, "Portfolio Optimization Using Machine Learning Techniques,"** ***2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM)*****, Dubai, United Arab Emirates, 2023.**

a) **Abstract:** A crucial component of financial decision-making is portfolio optimisation, which aims to minimise risk and maximise profits. The mean-variance model is one of the traditional methods that struggles to capture the dynamic character of financial markets. The use of machine learning approaches to improve portfolio optimisation is examined in this research article. We start by going over fundamental ideas like asset allocation and selection before gradually getting into the intricacies of machine learning techniques. We go over how to forecast asset returns and correlations using clustering, regression, and classification techniques. We also investigate portfolio rebalancing techniques using reinforcement learning. The paper reviews real-world data sources, data preprocessing, and model evaluation techniques, highlighting the importance of backtesting and risk management. Furthermore, we examine recent advancements in deep learning and neural networks to model non-linear relationships and assess their impact on portfolio optimization.

b) **Inference:** This paper explores the application of machine learning models in portfolio optimization, emphasizing their ability to enhance investment strategies by improving risk-adjusted returns. It examines how predictive models can forecast stock price movements with greater accuracy, enabling more informed asset allocation and diversification decisions. The study highlights the advantages of data-driven approaches over traditional methods, demonstrating how machine learning can balance risk and return more effectively, ultimately leading to better portfolio performance in dynamic and uncertain market environments.

5. **Adlakha, Naman & Ridhima, & Katal, Avita. (2021). Real Time Stock Market Analysis. 1-5. 10.1109/ICSCAN53069.2021.9526506.**

a) **Abstract:** The stock market is extremely important in the financial component of the country's growth, but it is also extremely dynamic and unpredictable. Significant political concerns, investor calls, journal stories, the company's restructuring and development strategies, and many other factors will quickly affect it. Single business day, thousands of investors invest in the stock market. Some of these investors either lose or gain money. However, the loss or benefit on any given trading day is unpredictable. The demand for stock price predictions is incredibly strong, necessitating the use of stock market research. Effective forecast tools assist traders indirectly by supplying supportive knowledge such as price position in the future. In this paper, the system for forecasting stock is developed that makes use of stacked LSTM, linear regression, random forest and K-nearest neighbours neural network algorithm, to forecast the stock trends on the basis of the price history. The proposed framework is designed to analyse every company's stock using mathematical technical metrics. The remaining metrics are probabilistic, although others are deterministic. The aim of this paper is to reduce the risk of failure in each exchange while increasing benefit.

b) **Inference:** This paper presents a stock market forecasting system that integrates multiple machine learning algorithms, including Stacked LSTM, Linear Regression, Random Forest, and K-Nearest Neighbors. The system utilizes historical stock price data to analyze trends and predict future movements, assisting traders in making informed decisions. By combining these diverse algorithms, the approach aims to capture various aspects of market behavior, improving the accuracy and robustness of stock trend predictions. This multi-model strategy enhances decision-making in dynamic trading environments.

6. Y. Mehta, A. Malhar and R. Shankarmani, "Stock Price Prediction using Machine Learning and Sentiment Analysis," 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, 2021, pp. 1-4, doi: 10.1109/INCET51464.2021.9456376.

a) **Abstract:** The stock market is a very dynamic market where nothing is as stable as a rock but as the technology is upgrading there are many ways and methods one can try to learn this dynamic change and be prepared accordingly. This paper focuses on such different methods of dynamically learning the market and its trends. We have used three different models for this paper and have also performed sentiment analysis on the tweets regarding the company or the stock, the model with the least error is the ideal and the most preferred method for prediction. The results of this classification have given a clear and insightful idea about the random ups and downs of the market and also a new approach for investors so that they know where they can bet their money. The ARIMA model is giving the best accuracy for every stock.

b) **Inference:** This research explores the integration of sentiment analysis from news articles and social media with machine learning models to forecast stock market trends. It emphasizes the significant impact of public sentiment on stock prices, acknowledging that investor emotions and perceptions, often reflected in news and social media, play a crucial role in market movements. By combining these sentiment signals with predictive models, the study aims to provide more accurate stock predictions, offering valuable insights for traders and investors in dynamic markets.

2.2. Patent Search :

1. Stock market prediction using natural language processing.

Inventor: Frederick S. M. HerzLyle, H. UngarJason, M. Eisner, Walter Paul Labys

A method of using natural language processing (NLP) techniques to extract information from online news feeds and then using the information so extracted to predict changes in stock prices or volatilities. These predictions can be used to make profitable trading strategies. Company names can be recognized and simple templates describing company actions can be automatically filled using parsing or pattern matching on words in or near the sentence containing the company name. These templates can be clustered into groups which are statistically correlated with changes in the stock prices. The system is composed of two parts: message understanding component that automatically fills in simple templates and a statistical correlation component that tests the correlation of these patterns to increases or decreases in the stock price. The methods can be applied to a broad range of text, including articles in online newspapers such as the Wall Street Journal, financial newsletters, radio &TV transcripts and annual reports. In an enhanced embodiment of the system statistical patterns in Internet usage data and Internet data such as newly released textual information on Web pages are further leveraged.

2. Stock market trading systems creation algorithm

Inventor: Vincent Serpico, Matthew Brunner

The invention is an algorithm that creates a technical analysis based trading system for a given market security. The trading system created provides a market trader with signals of when to buy and sell the given security. The algorithm searches for the most profitable indicators by testing them against quantifiable data of the selected security. It also factors in the user's risk management rules or trading style. Each indicator is tested using a series of three filters, each filter increasing in complexity. As an indicator fails filter one or filter two, it is deemed unprofitable and discarded from the search. This filtering approach allows a computer to perform a smart exhaustive search, using brains and not just brawn. Performing this type of search, the algorithm comes to the same conclusion as a normal exhaustive search, but uses much less computer processing time and power.

3. Method for Trading Stocks

Inventor: Mageshkumar VenkatesanBalamurugan Vellore GangadharanSatyanarayana GuptaRomeo QuiochoManit Kumar GomberJai Chandra YadavGregory S. Farber

The present invention provides a trading system with a very high percent success rate in the market by means of minimizing the risk. This invention provides an easy and simple-to-use system for swing trading. This system is automated and based on an algorithm which uses the end of the day data for a particular financial instrument to generate buy/sell signals for selected stocks. These signals provide the precise entry and exit points in the markets for those particular stocks. The signals come with a small stop loss to ensure safety of the investment. The system provides three exit levels to ensure that the investment risk is minimized and profit is reaped.

2.3. Inference Drawn:

1. **Stock Market Prediction using NLP**: This method uses NLP to analyze online news and extract key information to predict stock price movements. It involves recognizing company names, filling templates based on company actions, and correlating these patterns with stock price changes to assist in developing trading strategies.

2. **Stock Market Trading Systems Creation Algorithm**: This algorithm creates a technical analysis-based trading system, filtering profitable indicators through three stages to optimize decision-making. It factors in risk management rules, helping traders identify when to buy or sell securities with reduced computational resources.

3. **Method for Trading Stocks**: This system focuses on swing trading with high success rates by providing automated buy/sell signals based on end-of-day data. It minimizes risk through stop-loss measures and offers multiple exit levels to ensure profitable and safe investments.

Chapter 3: Requirement Gathering for the Proposed System

In this chapter we are going to discuss the resources we have used and how we analysed what the user actually needs and what we can provide. We will also discuss the functional and non-functional requirements and finally the software and hardware used.

3.1. Introduction to Requirement Gathering:

The Requirement Gathering is a process of requirements discovery or generating list of requirements or collecting as many requirements as possible by end users. It is also called as requirements elicitation or requirement capture.

The requirements gathering process consists of six steps :

- Identify the relevant stakeholders
- Establish project goals and objectives
- Elicit requirements from stakeholders
- Document the requirements
- Confirm the requirements
- Prioritise the requirements

USE CASE	DESCRIPTION
Register and Login	Users (investors/analysts) register and login to access the stock advisor app
Data Fetching	Historical and financial data of NSE stocks is fetched using Yahoo Finance
Recommend Stocks	Stocks are recommended based on calculated Buy Scores using financial metrics
Forecast Stock Price	Predict future stock prices using models like XGBoost, LSTM, and Random Forest
View Analysis	Users can view fundamental & technical indicators used in prediction
Add/Compare Stocks	Users can add or compare multiple stocks for analysis
Model Evaluation	Compare accuracy of models (MAE, RMSE, R^2) used for stock price forecasting
Display Performance	Show model results and Buy Scores in tabular and graphical format
Interactive UI	User-friendly web interface for stock exploration and recommendations

3.2. Functional Requirements:

1. **User Management**:
- Users should be able to create an account, log in, and log out securely.
- Users should be able to update their profiles, including personal information and investment preferences.

2. **Data Collection**:
- The system must gather real-time market data (stock prices, trading volumes, etc.) using APIs.
- The system must collect sentiment data from news articles and social media platforms.
- The system must retrieve relevant economic indicators from reliable sources.

3. **Data Processing**:
- The system should clean and normalize collected data to ensure consistency and accuracy.
- The system must perform sentiment analysis on gathered news and social media content to classify sentiment as positive, negative, or neutral.

4. **Predictive Analytics**:
- The system must implement various machine learning algorithms for stock price prediction and trend analysis.
- The system should support time series analysis techniques to forecast stock movements based on historical data.

5. **User Interface**:
- The dashboard should display real-time stock data, predictive analytics, and sentiment analysis results.
- Users should be able to customize their dashboard to focus on specific stocks or metrics of interest.
- The system should provide visualizations (charts, graphs) to represent data clearly and interactively.

6. **Recommendations**:
- The system must generate personalized investment recommendations based on user profiles, risk tolerance, and investment goals.
- Users should receive notifications or alerts about significant market changes or recommendations.

7. **Feedback Mechanism**:
- The system should allow users to provide feedback on the accuracy of predictions and recommendations.
- The system must implement a mechanism to update models based on user feedback and performance metrics.

8. **Reporting**:
- Users should be able to generate reports summarizing their investment performance, including historical data, predictions, and sentiment analysis.

3.3.Non-Functional Requirements:

1. **Performance**:
   - The system should process and display real-time data updates within a specified time frame (e.g., within seconds).
   - The predictive models should provide results efficiently, with a maximum response time for predictions (e.g., under 5 seconds).
2. **Scalability**:
   - The system must handle a growing number of users and data sources without degradation in performance.
   - The architecture should allow for easy addition of new data sources and functionalities.
3. **Security**:
   - The system must implement strong authentication and authorization measures to protect user accounts and sensitive data.
   - User data, including personal information and investment history, must be encrypted both in transit and at rest.
4. **Reliability**:
   - The system should ensure high availability, with an uptime of at least 99.5%.
   - The system must implement backup and recovery procedures to prevent data loss.
5. **Usability**:
   - The user interface must be intuitive and easy to navigate, enabling users of varying expertise to use the platform effectively.

- The system should provide help documentation and tutorials to assist users in understanding features and functionalities.

6. **Maintainability**:
   - The system's architecture should allow for easy updates and modifications to the codebase.
   - Documentation should be provided for all components to facilitate maintenance and future enhancements.
7. **Compliance**:
   - The system must comply with relevant financial regulations and data protection laws (e.g., GDPR) to ensure user privacy and data security.
8. **Interoperability**:
   - The system should be able to integrate with third-party applications and services (e.g., other financial tools or databases) through APIs.

3.4.Hardware, Software, Technology and Tools Utilised:

Hardware Requirements:

1. **Processor:** Intel i3 or AMD equivalent
2. **Disk Space:** < 500 MB
3. **RAM:** 4 GB
4. **OS:** Windows 10 32 bit or higher
5. **GPU:** Nvidia GPU or intel integrated graphics

Software Requirements:

Python Libraries:

1. **scikit-learn**: A Python library for machine learning, including feature extraction and model building.
2. **Matplotlib:** A popular Python library for creating static, animated, and interactive visualizations.
3. **Seaborn:** Built on top of Matplotlib, Seaborn provides a higher-level interface for creating statistical visualizations.

3.5. Constraints:

Technical Constraints:

- **API Limitations**: Restrictions on data retrieval from APIs (e.g., rate limits, data availability) may hinder the frequency and volume of data collection.
- **Data Quality**: Inconsistencies, inaccuracies, or incomplete data from various sources may affect the reliability of analyses and predictions.
- **Algorithm Complexity**: Some machine learning models may require extensive computational resources, which could limit performance on standard hardware.

Resource Constraints:

- **Budget**: Financial limitations may restrict the choice of data sources, technology stack, and tools, impacting overall project scope and quality.
- **Time**: Limited time for development can restrict the depth of features implemented, model training, and thorough testing, potentially affecting the final product quality.

Human Resource Constraints:

- **Skill Set**: Team members may have varying levels of expertise in machine learning, data analytics, and software development, which could impact project execution.
- **Availability**: The availability of team members for collaboration, feedback, and troubleshooting may affect project timelines and quality.

Chapter 4: Proposed Design

4.1. Block Diagram of the proposed system:

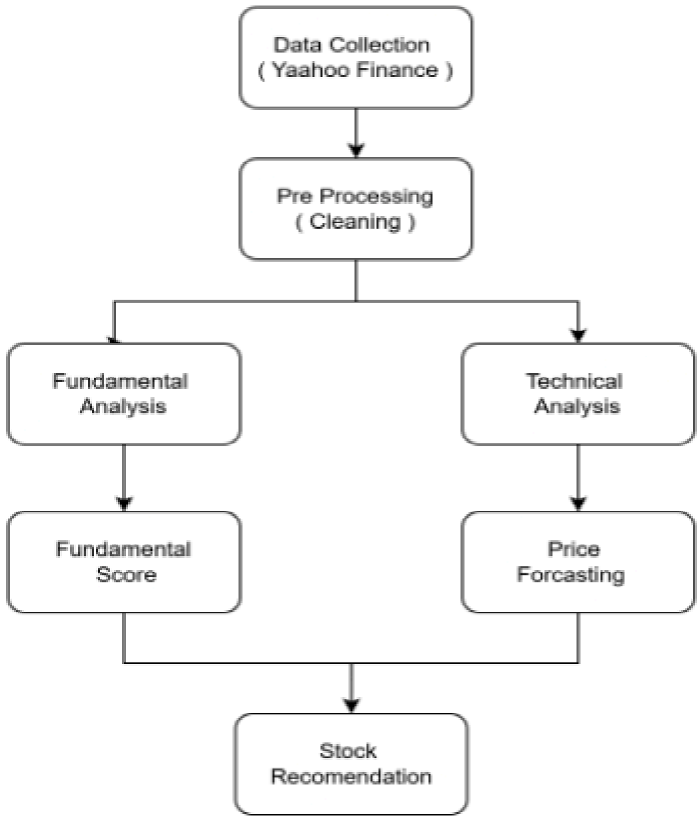


Fig 4.1: Block Diagram

The proposed system effectively integrates financial data analysis and machine learning for stock recommendation. It begins with data collection from Yahoo Finance, followed by thorough preprocessing to ensure data quality. Fundamental analysis is used to calculate financial scores based on metrics like P/E ratio, EPS, and book value. Simultaneously, technical indicators are leveraged to forecast stock prices using models such as XGBoost, Random Forest, CNN, and LSTM. These models are trained on historical data and evaluated using metrics like RMSE and MAE. The combined output of both analyses is used to generate stock recommendations. Finally, results are delivered to the user through an intuitive and interactive interface, providing investors with reliable and data-driven decision support for informed trading and investments.

4.2. Modular diagram of the system:

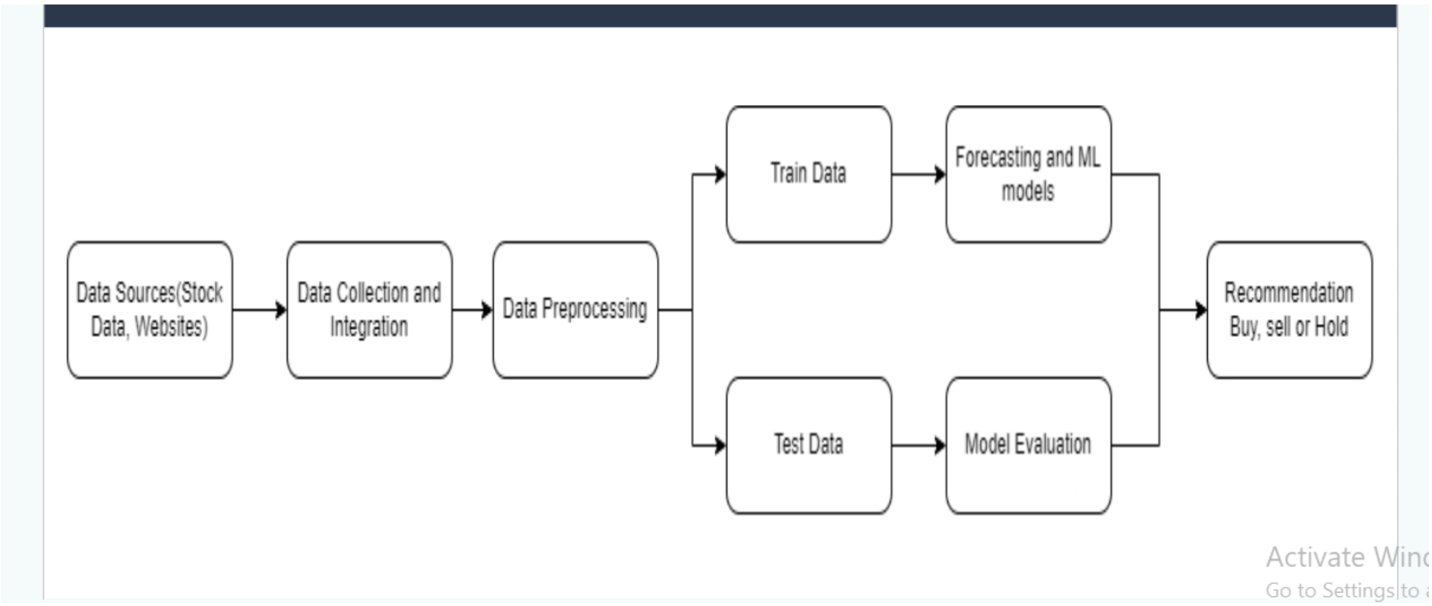


Fig 4.2: Modular Diagram

1. **Data Sources (Stock Data, Websites):**
   The system starts by gathering stock market data from various sources such as Yahoo Finance, online APIs, or financial websites.

2. **Data Collection and Integration:**
   The collected data is organized and combined into a structured format for further processing.

3. **Data Preprocessing:**
   Raw data is cleaned, missing values are handled, and technical indicators (like SMA, RSI, MACD) are added to prepare the dataset for modeling.

4. **Train Data:**
   The preprocessed data is split, and one part is used to train various forecasting and machine learning models.

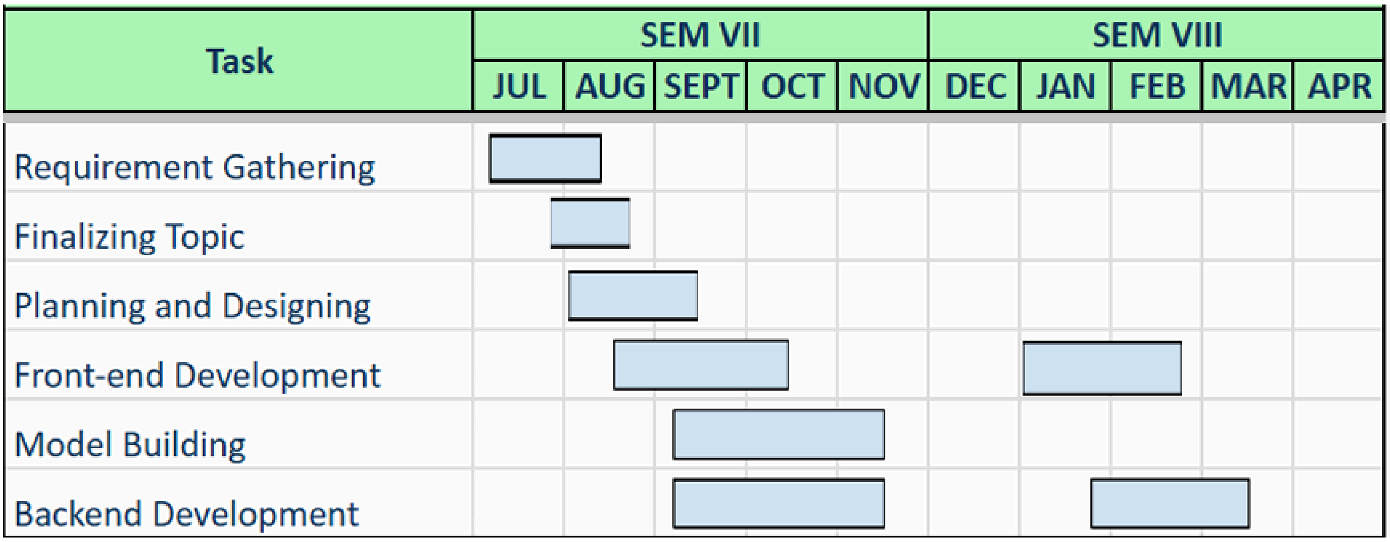
5. **Forecasting and ML Models:**
   Models such as XGBoost, Random Forest, CNN, and LSTM are applied to learn trends and predict future stock prices.

6. **Test Data:**
   The remaining part of the data is used to evaluate how well the trained models perform on unseen data.

7. **Model Evaluation:**
The models' accuracy is measured using metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.

8. **Recommendation (Buy, Sell, or Hold):**
Based on the model predictions and financial analysis, the system provides a recommendation to the user—whether to buy, sell, or hold the stock.

4.3. Project Scheduling & Tracking using Timeline / Gantt Chart :



Chapter 5: Implementation of the Proposed System

5.1.Methodology employed for development:

1. **Data Acquisition:**
   - Collected historical stock data and financial metrics from Yahoo Finance using the `yfinance` API.
2. **Data Preprocessing:**
   - Cleaned and structured the raw data by handling missing values, encoding categorical variables, and generating technical indicators (SMA, RSI, MACD, etc.).
   - Created lag features and normalized the dataset for modeling.
3. **Feature Engineering:**
   - Extracted relevant features such as P/E Ratio, EPS, Book Value, and technical indicators to enhance model performance.
4. **Model Development:**
   - Applied multiple machine learning and deep learning models including:
     - **XGBoost**
     - **Gradient Boosting**
     - **Random Forest**
     - **CNN**
     - **LSTM**
5. **Training and Validation:**
   - Split the dataset into training, validation, and test sets.
   - Trained models on the training set and tuned hyperparameters using the validation set.
6. **Model Evaluation:**
   - Evaluated model performance using MAE, RMSE, and R^2 metrics.
   - Compared model outputs to identify the most accurate approach.

7. **Stock Scoring and Ranking:**

   ○ Generated a "Buy Score" based on financial strength and model predictions to rank stocks for investment.

8. **Recommendation System:**

   ○ Integrated model outputs to provide a final recommendation (Buy, Sell, or Hold).

9. **User Interface Development:**

   ○ Built a simple and interactive web interface to allow users to input/select stocks and receive recommendations.

5.2.Algorithms for the respective modules developed:

1. **Start**
2. **Input** stock symbol from the user through the UI.
3. **Fetch** historical stock data (past 5–10 years) from Yahoo Finance API.
4. **Preprocess** the data:
5. Handle missing values, normalize data, and create technical indicators (SMA, RSI, MACD, Volatility).
6. **Model Selection**:
7. Let the user select one of the models: XGBoost, Random Forest, Gradient Boosting, CNN, or LSTM.
8. **Train** the selected model using historical data (if not already trained).
9. **Predict** next-day or future stock prices based on the trained model.
10. **Generate Recommendation**:
11. If predicted price increases significantly → **Recommend "Buy"**
12. If predicted price decreases → **Recommend "Sell"**
13. If minimal change → **Recommend "Hold"**
14. **Display** prediction, trend graph, and recommendation to the user via UI.
15. **End**

5.3.Datasets source and utilisation:

The dataset was created by fetching one year of historical stock data for 1,992 NSE-listed companies using the Yahoo Finance API through the `yfinance` Python library. For each stock and each trading day, core features like Open, High, Low, Close, Volume, Dividends, and Stock Splits were collected. Additional technical indicators such as the 50-day and 200-day Simple Moving Averages (SMA_50, SMA_200), Relative Strength Index (RSI), Moving Average Convergence Divergence (MACD), Volatility, and Capital Gains were calculated using rolling window methods and standard technical analysis formulas. Minor cleaning steps were applied to handle missing values, especially in technical indicators, while preserving essential financial information. The final dataset contains 489,656 records with 15 features, providing a comprehensive foundation for stock forecasting and analysis tasks.

Fig 5.4: Stock Dataset

Date	Open	High	Low	Close	Volume	Dividends	Stock Split	Symbol	SMA_50	SMA_200	RSI	MACD	Volatility	Capital Gains
2024-02-2	143.5845	150.1473	142.1925	148.6557	97538	0	0	20MICRON	165.2654	134.1172	41.4744	-4.82792	18.15772	
2024-03-0	149.0535	149.6998	145.2252	148.3077	51899	0	0	20MICRON	164.6648	134.4363	41.15122	-4.69276	16.42212	
2024-03-0	146.7168	149.0535	142.5902	143.5845	68411	0	0	20MICRON	164.2591	134.748	36.9438	-4.91017	17.64352	
2024-03-0	145.7224	145.7721	142.5405	144.4297	51432	0	0	20MICRON	163.757	135.0643	38.16224	-4.95712	18.58774	
2024-03-0	144.4298	173.3157	143.2863	156.5609	3078390	0	0	20MICRON	163.1206	135.4332	52.38383	-3.96969	17.51457	
2024-03-0	154.6716	154.8705	146.1699	147.8603	329960	0	0	20MICRON	162.3201	135.7627	44.48224	-3.84489	16.10969	
2024-03-1	148.2083	149.1529	142.2919	143.8828	137485	0	0	20MICRON	161.4978	136.0695	41.40734	-4.02058	16.45422	
2024-03-1	145.971	145.971	140.8501	143.7834	141801	0	0	20MICRON	160.8067	136.3736	41.33043	-4.12035	17.3041	
2024-03-1	142.1925	144.1314	131.2546	136.3258	216352	0	0	20MICRON	160.0292	136.6219	35.93825	-4.74647	20.88045	
2024-03-1	136.3258	150.2964	134.5857	147.562	185966	0	0	20MICRON	159.4703	136.9061	47.13015	-4.28659	20.34158	
2024-03-1	146.1699	154.0252	143.9823	150.2467	168550	0	0	20MICRON	159.0746	137.1741	49.40467	-3.66327	20.01156	
2024-03-1	151.1416	158.1021	149.4015	149.849	166659	0	0	20MICRON	158.5068	137.4621	49.06787	-3.1649	19.54657	
2024-03-1	147.6117	148.7055	142.7891	143.7834	68165	0	0	20MICRON	157.8883	137.7191	44.12758	-3.22223	19.24318	
2024-03-2	145.8716	148.954	143.1868	144.9766	52701	0	0	20MICRON	157.3026	137.9943	45.29445	-3.13524	17.40615	
2024-03-2	144.9766	149.0535	143.2862	143.8828	50222	0	0	20MICRON	156.7011	138.2582	44.37948	-3.11861	16.90425	
2024-03-2	143.8829	149.1529	143.3857	144.2309	132503	0	0	20MICRON	156.1015	138.5218	44.76184	-3.04228	16.99424	
2024-03-2	144.1812	149.1529	141.2478	142.1925	119407	0	0	20MICRON	155.4641	138.7861	42.90161	-3.11041	15.9303	
2024-03-2	144.6286	145.6727	139.0603	142.1925	83203	0	0	20MICRON	154.773	139.0385	42.90161	-3.12835	16.07043	
2024-03-2	143.684	147.4128	137.9167	142.8885	69677	0	0	20MICRON	154.0879	139.2663	43.82611	-3.05122	16.0768	
2024-04-0	143.1868	149.0535	134.0388	143.5845	99088	0	0	20MICRON	153.4465	139.5062	44.78879	-2.9005	16.14274	
2024-04-0	149.1529	160.7868	149.1529	151.7879	551143	0	0	20MICRON	153.19	139.7975	54.65278	-2.09496	16.85719	
2024-04-0	151.9371	154.9202	151.6885	154.075	78590	0	0	20MICRON	152.8967	140.0906	56.96135	-1.25752	18.28851	

Chapter 6: Testing of the Proposed System

6.1.Introduction to Testing :

Testing is a critical phase in the development of the proposed stock market analysis and recommendation system. It ensures that the system meets its functional and non-functional requirements, delivers accurate predictions, and provides a seamless user experience. The primary objective of testing is to validate the performance of machine learning and deep learning models, evaluate the reliability of the forecasting algorithms, and verify the accuracy of the investment recommendations generated.

Additionally, testing the user interface ensures that the platform is intuitive, responsive, and user-friendly. Various types of testing, including unit testing, integration testing, system testing, and user acceptance testing, are conducted to identify and resolve errors, enhance system robustness, and guarantee overall reliability. Through rigorous testing, the system aims to achieve high prediction accuracy, efficient processing of real-time data, and secure, error-free functionality, ultimately ensuring the platform's success for end users.

6.2.Types of tests Considered:

1. **Unit Testing**
   Unit testing focuses on testing individual components or modules of the system, such as data preprocessing functions, model training scripts, and prediction functions. Each function is tested independently to ensure it performs correctly and handles edge cases efficiently.

2. **Integration Testing**
   Integration testing is conducted to verify that different modules of the system—such as data collection, processing, model prediction, and the user interface—work together as expected. It ensures smooth data flow between components and identifies any interface or communication errors.

3. **System Testing**
   System testing evaluates the complete and integrated application to verify that the system meets its specified requirements. It includes testing real-time stock data fetching, model performance on unseen data, recommendation generation, and UI responsiveness.

4. **Performance Testing**
   Performance testing measures the system's speed, scalability, and stability under different workloads. It checks how quickly the models can make predictions, how the system handles multiple user queries, and whether real-time data is processed within acceptable time limits.

5. **User Acceptance Testing (UAT)**
   UAT ensures that the final system meets the expectations of the end-users. Real users interact with the application to validate ease of use, clarity of recommendations (Buy/Sell/Hold), and overall satisfaction with the system's functionality.

6. **Regression Testing**
   Whenever updates or new features are added, regression testing is performed to ensure that the existing functionality continues to work correctly and that new changes do not introduce new bugs.

Chapter 7: Results and Discussions

7.1.Screenshot of Use Interface(UI) for the system:

Fig 7.1: Screenshot for Login page

Fig 7.2: Screenshot for Register page

Fig 7.3: Screenshot for Stock Selection

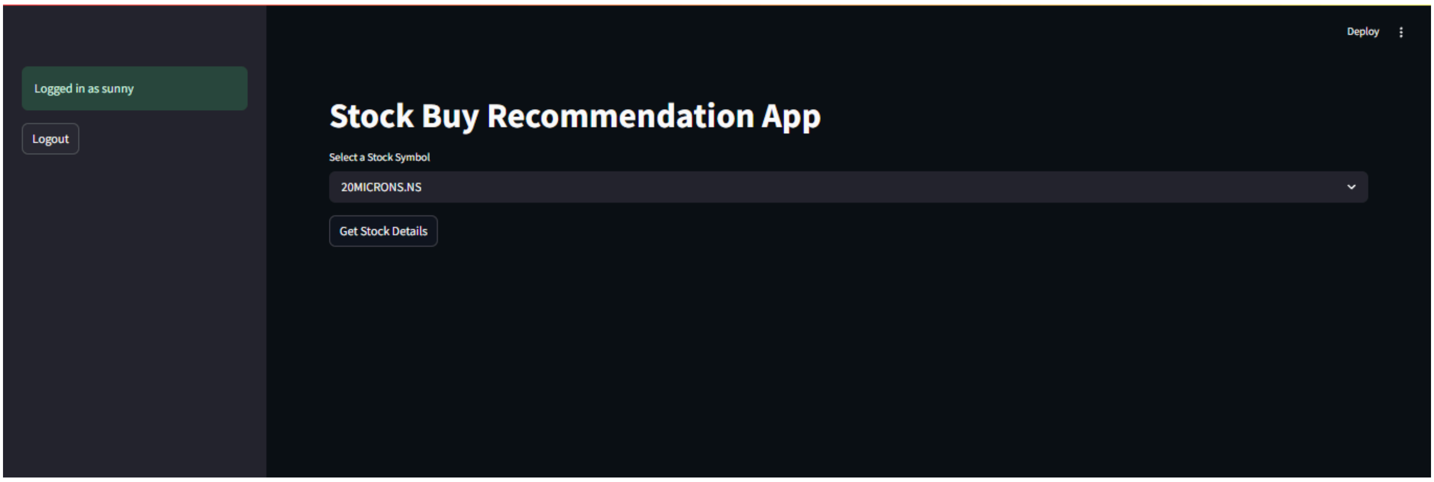


Fig 7.4: Screenshot for Stock Page 1

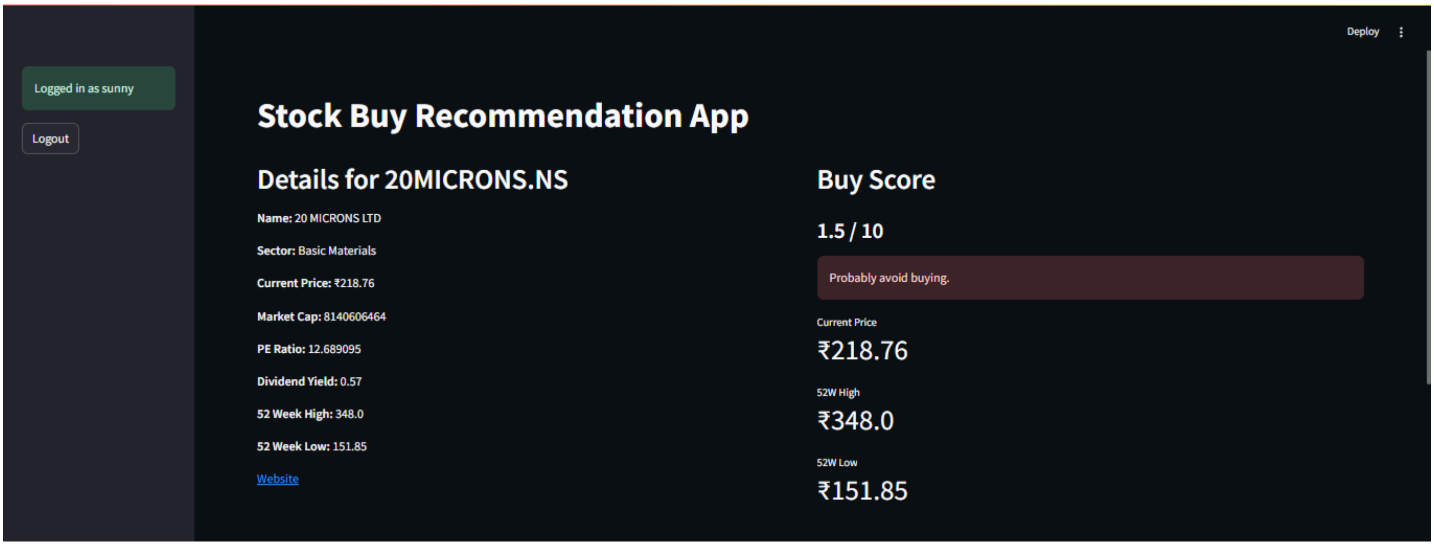
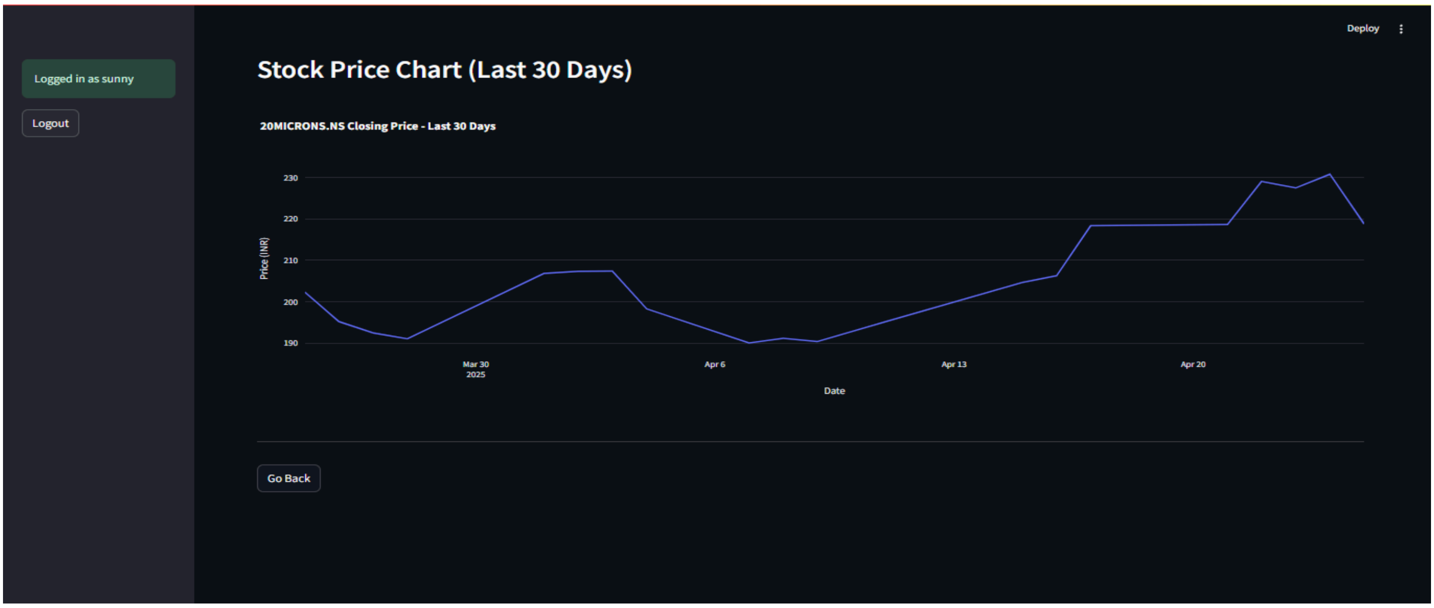


Fig 7.5: Screenshot for Stock Page 2



7.2. Performance Evaluation Measures:

1. Mean Absolute Error (MAE)

- **Interpretation**: Average absolute difference between actual and predicted values.
- **Pro**: Easy to understand.
- **Con**: Doesn't penalize large errors much.

2. Mean Squared Error (MSE)

- **Interpretation**: Average of squared differences.
- **Pro**: Penalizes larger errors more.
- **Con**: Units are squared.

3. R-squared (Coefficient of Determination)

- **Interpretation:** Proportion of variance in the target explained by the model.
- **Range: $[-\infty, 1]$;** closer to 1 is better.
- **Con:** Can be misleading for non-linear models.

7.3. Input Parameters/Features considered:

To effectively forecast stock prices and generate accurate investment recommendations, the following input parameters and features were considered in the system:

1. **Historical Stock Prices**
   - Opening Price
   - Closing Price
   - High Price
   - Low Price
   - Adjusted Closing Price
   - Trading Volume

2. **Technical Indicators** (Calculated using historical prices)
   - **SMA 50** (Simple Moving Average over 50 days)
   - **SMA 200** (Simple Moving Average over 200 days)
   - **RSI** (Relative Strength Index, 14-day window)
   - **MACD** (Moving Average Convergence Divergence)
   - **Volatility** (Difference between Bollinger High Band and Low Band)
3. **Financial Ratios** (Fetched from company financials)
   - Price to Earnings (P/E) Ratio
   - Earnings Per Share (EPS)
   - Book Value
4. **Time-Based Features**
   - Date/Time (to capture trends based on time of year, quarterly patterns)
5. **Company Financial Data**
   - Total Revenue
   - Net Income

  - Total Assets
  - Total Liabilities
  - Operating Cash Flow

6. **Sentiment Score** (Optional future enhancement)
    - Sentiment extracted from news headlines or social media about the company or stock market conditions.

7.4. Inference Drawn:

The stock recommendation app provides a clean interface where users first log in or register securely, then select a stock from a dropdown list populated from a local CSV file. After selecting, they are shown live stock details fetched via Yahoo Finance, a recommendation score out of 10 (generated by an ML model), and an interactive stock price chart. The system is simple, intuitive, and ideal for personal use, demos, or early-stage stock analysis platforms, with easy room for future upgrades like advanced analytics or automated alerts.

Chapter 8: Conclusion

8.1.Limitations:

1. **Market Volatility:**

- The system may not react accurately to sudden market crashes or news-driven events as it relies on historical data patterns.

2. **Data Dependency:**

- Model accuracy is highly dependent on the quality and completeness of stock data from sources like Yahoo Finance. Missing or outdated data can affect predictions.

3. **Overfitting Risk:**

- Complex models like CNN and LSTM can overfit the training data if not carefully tuned, leading to poor performance on unseen data.

4. **Limited Macroeconomic Context:**

- The system does not consider macroeconomic factors like interest rates, inflation, or geopolitical events which also impact stock prices.

5. **No Real-Time Updates:**

- Unless integrated with a live data pipeline, recommendations may not reflect real-time market changes.

8.2.Conclusion:

This project successfully developed a comprehensive stock recommendation system that combines machine learning and deep learning techniques for effective market forecasting. Utilizing financial and historical data from Yahoo Finance, the system integrates both fundamental and technical analysis to provide insightful investment guidance. Models such as XGBoost, Gradient Boosting, Random Forest, CNN, and LSTM were employed to improve prediction accuracy and deliver reliable "Buy," "Sell," or "Hold" recommendations.

A notable outcome of this work is the development of a user-friendly interface, enabling investors—regardless of technical background—to interact with the system and make informed decisions. The results highlight the strength of deep learning models in identifying long-term trends, while ensemble methods proved efficient for short-term predictions.

Looking ahead, the system could be enhanced by incorporating sentiment analysis from financial news and social media, as well as reinforcement learning for adaptive recommendation strategies. Overall, this solution demonstrates the potential of AI in financial decision-making and offers a scalable, intelligent tool for both individual and institutional investors.

8.3.Future Scope:

1. **Sentiment Analysis Integration:**
   Incorporate real-time sentiment analysis from news articles and social media to improve prediction accuracy.

2. **Reinforcement Learning:**
   Apply reinforcement learning to dynamically adapt and optimize the stock recommendation strategy over time.

3. **Real-Time Data Streaming:**
   Upgrade the system to handle live data feeds for real-time forecasting and recommendations.

4. **Mobile Application Development:**
   Extend the system to a mobile platform for wider accessibility and on-the-go decision-making.

5. **Portfolio Optimization:**
   Add features for automated portfolio management based on user preferences and risk tolerance.

References:

- M. Usmani, S. H. Adil, K. Raza and S. S. A. Ali, "Stock market prediction using machine learning techniques", 2016 3rd International Conference on Computer and Information Sciences (ICCOINS), pp. 322-327, 2016.

- K. V. Sujatha and S. M. Sundaram, "Stock index prediction using regression and neural network models under non normal conditions", INTERACT-2010, pp. 59-63, 2010.

- Bhandare, Y., Bharsawade, S., Nayyar, D., Phadtare, O., & Gore, D. (2020, June).

- SMART: Stock Market Analyst Rating Technique Using Naive Bayes Classifier. In 2020 International Conference for Emerging Technology (INCET) (pp. 1-4). IEEE.

- Arora, K., Aggarwal, A., & Gola, K. K. (2024). Predicting Stock Market Prices and Provide Recommendations. International Journal of Computer Information Systems and Industrial Management Applications, 16(3), 16-16.

- Grinblatt, M., Keloharju, M., & Linnainmaa, J. (2011). IQ and stock market participation. *The Journal of Finance*, *66*(6), 2121-2164.

- Stalovinaitė, I., Maknickienė, N., & Martinkutė-Kaulienė, R. (2020). Investigation of decision making support in digital trading.

- Guo, H., Liu, M., Yang, B., Sun, Y., Qu, H., & Shi, L. (2022). Rankfirst: Visual analysis for factor investment by ranking stock timeseries. *IEEE Transactions on Visualization and Computer Graphics*.

Appendix

a) **Research paper**

Invest IQ: Smart Stock Market Analysis and Recommendation System

Mr. Sunny Bhatia, Mr. Sachin Kundal, Mr.Tarun Sharma,

Prof. Priyanka Shah

Department of Computer Engineering, Vivekanand Education Society's Institute of Technology, Chembur, Mumbai, Maharashtra.

Keywords-Stock Price Prediction, Stock Market, Time-Series Prediction

Abstract: The dynamic nature of the stock market, driven by countless economic, political and psychological factors, represents a challenging yet attractive problem for predicting analysis. This study introduces a robust and hybridized inventory system that uses both traditional machine learning and deep learning methods. The main goal of this study is to provide intelligent investment proposals through a comprehensive system that predicts stock price movements and integrates several models with an interactive user interface. We focus on historical stock price data, including characteristics such as opening and final price, high and low prices, volume and other time series indicators. This data record was widely processed to handle the lack of value and to perform normalization and engineering-related functions to effectively record market trends and behavior. Each of these models was selected for their specific strengths in dealing with structured data, time series sequences, and complex nonlinear relationships. Ensemble models such as Xgboost, gradient reinforcement, and random forests are known for their structured scenario accuracy and robustness. In contrast, deep learning models such as CNN and LSTM are particularly effective when extracting temporal patterns and spatial dependencies from financial series. On the other hand, LSTM networks are specially designed to capture long-term dependencies, making them suitable for time-dependent financial forecasting. The models were trained and evaluated using a combination of mid-directional error (MSE), MESE metric

(RMSE), and R-squared metrics to evaluate performance and generalization functions. This interface allows users to enter shared symbols, select predictive models, and visualize predicted trends. Based on model results, the system provides recommendations such as "purchase", "sell", "maintenance", and provides clear instructions supported by data-controlled knowledge. The UI fills the gap between complex machine learning processes and investor ease of use, improving the decision-making process without the need for users of technical know-how. By using the strength of each model type, our hybrid approach increases the reliability and interpretability of stock predictions. Future work will include integrating mood analysis from news or social media sources and enhancing learning techniques to further improve recommended logic. Our systems form the basis of a scalable, modular, intelligent platform that enables investors through advanced data analytics and predictive modeling..

Introduction: In recent years, the fiscal and artificial intelligence interface has launched new restrictions to forecast stock market trends and improve investment strategies. Stock markets with inherent volatility and dependence on a variety of external and internal factors represent key challenges for accurate prognosis. Despite its unpredictable nature, patterns and correlations often appear in financial time series data that can be used in

advanced machine learning and deep learning techniques. The availability of historical financial data is increasing, and the progress of arithmetic tools, financial prognosis and how to control data for decision-making - making - is increasingly profitable. The goal is to not only predict share trends, but to implement these forecasts in investment recommendations that can be implemented. The models implemented in this project include XGboost, gradient boost, random forest, folding network (CNN), and long-term short-term memory (LSTM) networks. These were selected for complementary intensities to process structured data and time-scale sequences. The data records include key stock market indicators such as opening and final prices, daily highs and lows, trading volumes and other technical features. Preliminary processing procedures such as normalization, lack of value processing, and functional engineering were implemented to ensure data quality and model compatibility. The user interface allows users to enter a shared ticker, select a predictive model, and view predicted trends along with investment suggestions. These proposals, created as "buy, sell" or "buy, sell", are based on a predicted version of the model and provide a simplified decision-making framework. A hybrid modeling approach improves prediction accuracy and system reliability. Furthermore, UI integration creates a gap between advanced analytics and everyday user familiarity, making the financial prognosis more accessible to more viewers. This article examines the methods, implementation and results of stock recommendation systems, contributing to the growing landscape of AI-controlled financial products.

Literature Survey:

The implementation described in [1] focuses on using Reinforcement Learning (RL) and Q-Learning algorithms to analyze Egyptian stock market data. By automating buying and selling decisions, the system aims to enhance stock trading strategies. However, the study highlights a research gap in the practical availability of this RL-based model.

Another approach [2] utilizes Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNN) to predict short-term price trends of the Chinese stock market. The study emphasizes the role of feature engineering and financial domain knowledge in improving prediction accuracy. However, it mainly compares stocks with their historical data and does not incorporate external influencing factors.

The study in [3] reviews systematic stock market prediction techniques applied to NASDAQ, TAIEX, and National Bank of Greece stock indexes. It employs SVM for forecasting and suggests that a comprehensive analysis of stock prediction methodologies is necessary for refining existing models.

In [4], various statistical models such as ARIMA, STAR, ARMA, and PIP are combined with SVM, Decision Tree, K-Nearest Neighbors (KNN), and Logistic Regression to enhance prediction accuracy. However, the research indicates that hybrid approaches integrating statistical data with machine learning techniques may yield better predictive performance in financial forecasting.

Author/Year	Dataset	Methods	Results	Evaluation Measures
Abdulrahman A. Ahmed, Ayman Ghoneim, Mohamed Saleh (2021)	Egyptian Stock Market Data	Focusing on the digital records and using automated agents to buy and sell the stocks	There is no practical availability of this project based on RL and Q model	Accuracy = 96.25%

Jingyi Shen and M. Omair Shafq (2020)	3558 stocks of Chinese Stock Market, API called Tushare	Analyze the best approach for predicting short term price trends through feature engineering, financial domain knowledge, and prediction algorithm	This research just compares the stocks to its previous data and uses LSTM model to predict the future	Accuracy = 96.9%
Dattatray P. Gandhmal and K.kumar (2019)	NASDAQ, TAIEX and National Bank of Greece index stock	Techniques utilized for stock prediction ANN , SVM	Systematic analysis and review of stock market prediction techniques	Accuracy = 93%,
Dev Shah, Haruna Isah and Farhana Zulkernine (2019)	Self prepared	Models like ARIMA ,STAR , ARMA along with PIP technique using decision tree and combing with statistical data	Hybrid approaches that combine statistical data and machine learning techniques will probably prove more useful.	Accuracy = 57.7%
Srinath Ravikumar & Prasad Saraf (2020)	Yahoo Finance	Analysis of Deep learning, stock market, factors affecting stock market.and prediction of stock market considering these factors.	Though the prediction was quite accurate up to 96% there was no real time update of stocks. Also there were no real time graph and other pictorial format for the stock that are studied	Accuracy = 82.9%

Proposed System:

The proposed system is a comprehensive, hybrid stock recommendation platform that combines prediction capabilities of both machine learning and deep learning techniques. The aim is to analyze historical stock data and predict future price movements to provide implementation-able investment recommendations. The system architecture is divided into several core modules: data collection and preprocessing, model training and evaluation, prediction, and user interaction via custom interfaces. The data includes a variety of stock attributes, such as openings, closing prices, high, low, adaptive closures, and trading volumes. The preprocessing procedure is used to clean the data by processing missing values and normalizing functional scales, and by normalizing new features such as movement and everyday returns to enrich the dataset. This preprocessing pipeline ensures that your data is suitable for recording your model, reducing noise for improved performance. Based on capabilities, these models are chosen to effectively process structured and sequential data. Ensemble methods such as random forest, xgboost, and gradient reinforcement are particularly powerful in the absorption of nonlinear relations in tabular data, but CNNs can extract local patterns in sequences. The strengths of modeling temporary dependencies and long-range sequence relationships have led to special use of LSTM networks. This is extremely important for forecasting financial time series. These expenses are processed to sell intuitive recommendations to "buy," sell, sell, or hold, based on the expected trends and threshold values. The final component is a web-based graphical user interface in which users enter inventory lickers, select prediction visualize prediction results, and receive TaylorMade recommendations. This user interface should make the system accessible and practical,

especially for retail investors who may not have a technical background. With a powerful Model and integration of real financial data With a user-friendly interface, the proposed System provides reliable and efficient equipment for stock exchange forecasting and decision making.

Methodology:-

The methodology for this study involves multiple stages, including data collection, preprocessing, feature selection, model implementation, and evaluation.

1. Data Collection:

The dataset is sourced from Yahoo Finance,coveri stock price data of Adani Ports and Special Zone, HDFCLife, and Bombay Dyeing from 2007 to the present. The dataset consists of key financial indicators such as Open, High, Low, Close, Adjacent Close, and Volume, which are essential for stock trend analysis.

2. Data Preprocessing:

Handling missing values to ensure data completeness. date formats and aligning stock prices to a uniform time frame.Checking for outliers and anomalies using statistical techniques. Normalizing stock prices to ensure consistency across different scales.

3. Feature Selection & Engineering:

Identifying key stock market indicators that influence price movements. Calculating moving averages and volatility indicators. Applying differencing techniques to achieve stationarity in time-series data.

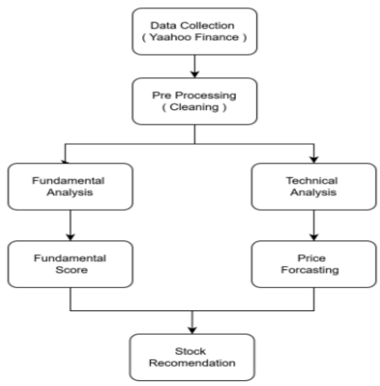
4. Model Implementation:

The proposed system begins with Yahoo Finance Data collection followed by preprocessing to clean and build the data records. Basic analysis calculates financial value, while technical analysis uses models such as Xgboost, Random Forest, CNN, LSTM and more to predict prices. The output is integrated to generate inventory recommendations through a user-friendly interface.

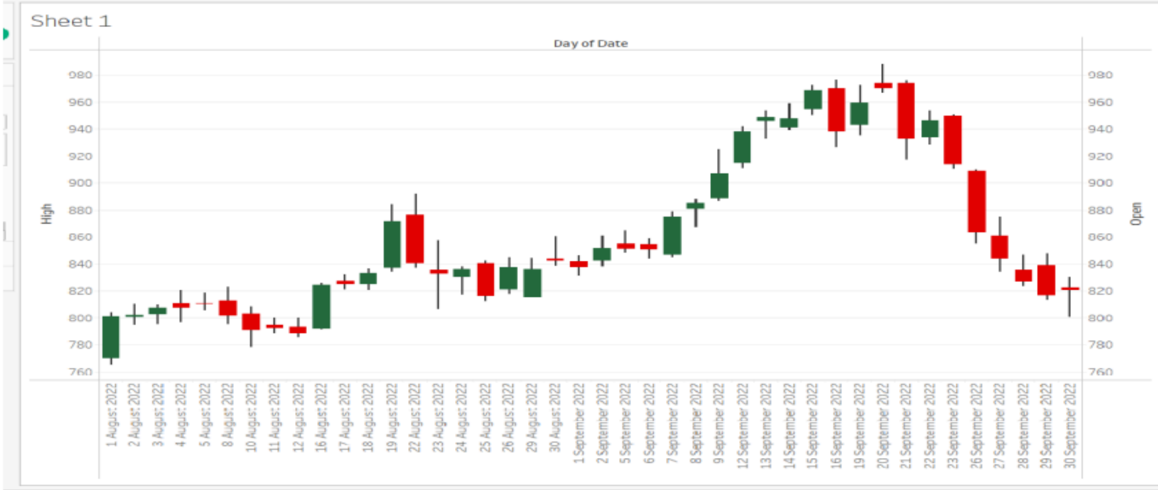
5. Model Evaluation:

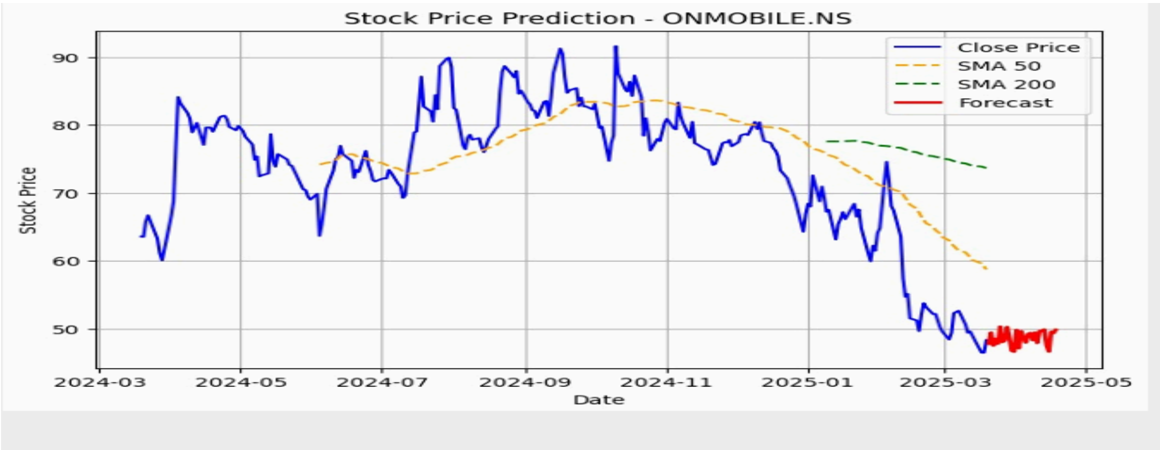
Measuring model performance using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) to assess forecast accuracy.

Fig. block Diagram for Model Evaluation



Results:


Sheet 1
Day of Date
High
Open
980
960
940
920
900
880
860
840
820
800
780
760
1 August 2022
2 August 2022
3 August 2022
4 August 2022
5 August 2022
8 August 2022
10 August 2022
11 August 2022
12 August 2022
16 August 2022
17 August 2022
18 August 2022
19 August 2022
22 August 2022
23 August 2022
24 August 2022
25 August 2022
26 August 2022
29 August 2022
30 August 2022
1 September 2022
2 September 2022
5 September 2022
6 September 2022
7 September 2022
8 September 2022
9 September 2022
12 September 2022
13 September 2022
14 September 2022
15 September 2022
16 September 2022
19 September 2022
20 September 2022
21 September 2022
22 September 2022
23 September 2022
26 September 2022
27 September 2022
28 September 2022
29 September 2022
30 September 2022


Stock Price Prediction - ONMOBILE.NS
Close Price
SMA 50
SMA 200
Forecast
Stock Price
90
80
70
60
50
2024-03
2024-05
2024-07
2024-09
2024-11
2025-01
2025-03
2025-05
Date


Server
Window
Help
Entire View
Columns
DAY(Date)
Rows
SUM(High)
SUM(Low)
Sheet 7
Day of Date
High
Low
1900
1800
1700
1600
1500
1400
1300
1200
1100
Day of Date: 26 April 2022
High: 1,563.3

Conclusion:

This study presents a robust stock recommendation system, integrated machine learning, and deep learning models to improve stock market forecasts. By using Yahoo Finances financial data, the system processes and analyzes historical trends through both basic and technical approaches. Including ensemble models such as XGBoost, gradient hosting, and random forests, as well as deep learning architectures such as CNN and LSTM, improves prediction accuracy and reliability. The developed system not only predicts stock price movements, but also provides insights that can be implemented in the form of recommendations for "buy", "selling", or "hold" recovery. By filling the gap between financial data analysis and investment decisions, the system provides practical solutions for retail and institutional investors. The results show that while long-term learning techniques are effective in short-term predictions, deep learning models are well cut when recording long-term dependencies. Furthermore, the inclusion of reinforcement learning techniques may improve the recommended logic over time. Overall, this study contributes to the growth field of AI-controlled financial technology by providing a scalable, accurate and accessible inventory forecasting system.

References:-

1. Selvamuthu, D., Mishra, A., & Kumar, V. (2019). Indian stock market prediction using artificial neural networks on tick data. *Finance Innovation, 5(16)*. https://jfin-swufe.springeropen.com/articles/10.1186/s40854-019-0131-7.

2. Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2015). Stock Price Prediction Using the ARIMA Model. *Cambridge, UK*. https://ieeexplore.ieee.org/document/7046047.

3. Vijh, M., Chandola, D., Anand, V., & Kumar, A. (2020). Stock Closing Price Prediction using Machine Learning Techniques. *Elsevier, Dept. Comp Sci, NIT Rourkela*. https://doi.org/10.1016/j.procs.2020.03.326.

4. Soni, P., Tewari, Y., & Krishnan, D. (2022). Machine Learning Approaches in Stock Price Prediction: A Systematic Review. *Journal of Physics*. https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012065.

5. Ravikumar, S., & Saraf, P. (2020). Prediction of Stock Prices using Machine Learning (Regression, Classification) Algorithms. *IEEE*. https://ieeexplore.ieee.org/document/9154061.

6. Ahmed, A. A., Ghoneim, A., & Saleh, M. (2020). Optimizing stock market execution costs using reinforcement learning. *IEEE Symposium Series on Computational Intelligence (SSCI), 1083-1090*. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9308153.

7. Shen, J., & Shafiq, M. O. (2020). Short-term stock market price trend prediction using a comprehensive deep learning system. *Journal of Big Data, 7(66)*. https://doi.org/10.1186/s40537-020-00333-6.

8. Ji, X., Wang, J., & Yan, Z. (2021). A stock price prediction method based on deep learning technology. *International Journal of Crowd Science, 5(1), 55-72*. https://doi.org/10.1108/IJCS-05-2020-0012.

9. Chandola, D., Mehta, A., & Singh, S. (2022). Forecasting Directional Movement of Stock Prices using Deep Learning. *Annals of Data Science*. https://doi.org/10.1007/s40745-022-00432-6.

10. Dopi, G., Hartanto, R., & Fauziati, S. (2021). Systematic Literature Review: Stock Price Prediction Using Machine Learning and Deep Learning. *Springer Nature*. https://dx.doi.org/10.2991/aebmr.k.211117.008.

11. Gandhmal, D. P., & Kumar, K. (2019). Systematic analysis and review of stock market prediction techniques. *Vellore Institute of Technology, Vellore, India*. https://www.sciencedirect.com/science/article/abs/pii/S157401371930084.

12. Shah, D., & Isah, H. (2019). Stock Market Analysis: A Review and Taxonomy of Prediction Techniques. *MDPI, 7(2), 26*. https://www.mdpi.com/2227-7072/7/2/26.

b.Plagiarism Report

Smart Stock Market Analysis and Recommendation System

ORIGINALITY REPORT

9%	4%	9%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	R. N. V. Jagan Mohan, B. H. V. S. Rama Krishnam Raju, V. Chandra Sekhar, T. V. K. P. Prasad. "Algorithms in Advanced Artificial Intelligence - Proceedings of International Conference on Algorithms in Advanced Artificial Intelligence (ICAAAI-2024)", CRC Press, 2025 Publication	2%
2	mpra.ub.uni-muenchen.de Internet Source	1%
3	journalarticle.ukm.my Internet Source	1%
4	journalofbigdata.springeropen.com Internet Source	1%
5	S. Prasad Jones Christydass, Nurhayati Nurhayati, S. Kannadhasan. "Hybrid and Advanced Technologies", CRC Press, 2025 Publication	1%
	Submitted to University of Greenwich	1

c.Project review sheet ;

Project review sheet 1:

Inhouse/ Industry _Innovation/Research:

Sustainable Goal:

Class: D17 A/B/C

Group No.: 53

Project Evaluation Sheet 2024 - 25

itle of Project: Invest IQ: Smart Stock Market Analysis and Recommendation System

roup Members: Sachin Kundal (D17A/34), Tarun Sharma (D17C/57), Sunny Bhatia (D17C/72)

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (2)	Applied Engg&M gmt principles (3)	Life - long learning (3)	Profess ional Skills (3)	Innov ative Appr oach (3)	Resear-ch Paper (5)	Total Marks (50)
05	05	05	03	05	02	02	02	02	02	02	02	02	02	04	45

omments:

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environ ment Friendly (2)	Ethics (2)	Team work (2)	Presentati on Skills (2)	Applied Engg&M gmt principles (3)	Life - long learning (3)	Profess ional Skills (3)	Innov ative Appr oach (3)	Resear -ch Paper (5)	Total Marks (50)
05	05	05	03	05	02	02	02	02	02	02	02	02	02	04	45

omments:

45

ate: 1st March, 2025

Priyanka Bagde

Name & Signature Reviewer 2

Project review sheet 2

Project Evaluation Sheet 2024-25

Title of Project (Group no): Invest IQ: Smart Stock Analysis and Recommendation System (53)

Group Members: Sachin Kundal (D17A, 34), Tarun Sharma (D17C, 57), Sunny Bhatia (D17C, 72)

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Research Paper (5)	Total Marks (50)
04	04	04	02	04	01	02	01	01	03	03	03	05	05	05	42

Comments:

Inhouse / Industry_Innovation / Report

Priyanka Shah

Name & Signature Reviewer1

Engineering Concepts & Knowledge (5)	Interpretation of Problem & Analysis (5)	Design / Prototype (5)	Interpretation of Data & Dataset (3)	Modern Tool Usage (5)	Societal Benefit, Safety Consideration (2)	Environment Friendly (2)	Ethics (2)	Team work (2)	Presentation Skills (3)	Applied Engg & Mgmt principles (3)	Life-long learning (3)	Professional Skills (5)	Innovative Approach (5)	Research Paper (5)	Total Marks (50)
Review of Project Stage 05	05	04	02	04	01	02	01	01	03	03	03	05	5	05	40

Comments:

Date: 1st April 2025

Richard Joseph

Name & Signature Reviewer2