PROJECT TITLE SHOULD BE IN UPPERCase LETTERS & TIMES NEW ROMAN

A PROJECT REPORT

BY

TEAM NO. 54

SIDDHARTH (E23CSEU0869)

RAJVEER SINGH RATHORE (E23CSEU0845)

A blue and red text on a black background

Description automatically generated

SUBMITTED TO

SCHOOL OF COMPUTER SCIENCE ENGINEERING AND TECHNOLOGY, BENNETT UNIVERSITY

GREATER NOIDA, 201310, UTTAR PRADESH, INDIA

April 2025

# DECLARATION

I/We hereby declare that the work which is being presented in the report entitled “Stock Prediction system” , is an authentic record of my/our own work carried out during the period from JAN, 2023 to April, 2023 at School of Computer Science and Engineering and Technology, Bennett University Greater Noida.

The matters and the results presented in this report has not been submitted by me/us for the award of any other degree elsewhere.

SIDDHARTH (E23CSEU0869)

RAJVEER SINGH RATHORE (E23CSEU0845)

# ACKNOWLEDGEMENT

I/We would like to take this opportunity to express my/our deepest gratitude to my/our mentor, **Dr.** **Dr.Rashmi kumari , Assistant professor** for guiding, supporting, and helping me/us in every possible way. I/we was/were extremely fortunate to have him as my/our mentor as he provided insightful solutions to problems faced by me/us thus contributing immensely towards the completion of this capstone project. I/We would also like to express my/our deepest gratitude to VC, DEAN, HOD, faculty members and friends who helped me/us in successful completion of this capstone project. Any other name you can mentioned here. (Acknowledgement is your personal view, so you can write it in your way by maintaining integrity of technical report).

SIDDHARTH (E23CSEU0869)

RAJVEER SINGH RATHORE (E23CSEU0845)

**TABLE OF CONTENTS**

<<Right click on the heading and click update flied. Your heading will be pulled here. Same goes with all the table of content in the subsequent pages>>

[LIST OF TABLES vi](#_Toc85881407)

[LIST OF FIGURES vii](#_Toc85881408)

[LIST OF ABBREVIATIONS viii](#_Toc85881409)

[ABSTRACT ix](#_Toc85881410)

[1. INTRODUCTION 1](#_Toc85881411)

[1.1. Problem Statement 1](#_Toc85881412)

[2. Background Research 2](#_Toc85881413)

[2.1. Proposed System 2](#_Toc85881414)

[2.2. Goals and Objectives 2](#_Toc85881415)

[3. Project Planning 3](#_Toc85881416)

[3.1. Project Lifecycle 3](#_Toc85881417)

[3.2. Project Setup 3](#_Toc85881418)

[3.3. Stakeholders 4](#_Toc85881419)

[3.4. Project Resources 4](#_Toc85881420)

[3.5. Assumptions 5](#_Toc85881421)

[4. Project Tracking 5](#_Toc85881422)

[4.1. Tracking 5](#_Toc85881423)

[4.2. Communication Plan 6](#_Toc85881424)

[4.3. Deliverables 7](#_Toc85881425)

[5. SYSTEM ANALYSIS AND DESIGN 8](#_Toc85881426)

[5.1. Overall Description 8](#_Toc85881427)

[5.2. Users and Roles 8](#_Toc85881428)

[5.3. Design diagrams/ UML diagrams/ Flow Charts/ E-R diagrams 9](#_Toc85881429)

[5.3.1. Use Case Diagrams 9](#_Toc85881430)

[5.3.2. Class Diagram 10](#_Toc85881431)

[5.3.3. Activity Diagrams 11](#_Toc85881432)

[5.3.4. Sequence Diagram 12](#_Toc85881433)

[5.3.5. Data Architecture 13](#_Toc85881434)

[6. User Interface 14](#_Toc85881435)

[6.1. UI Description 14](#_Toc85881436)

[6.2. UI Mockup 14](#_Toc85881437)

[7. Algorithms/Pseudo Code 15](#_Toc85881438)

[8. Project Closure 16](#_Toc85881439)

[8.1. Goals / Vision 16](#_Toc85881440)

[8.2. Delivered Solution 16](#_Toc85881441)

[8.3. Remaining Work 16](#_Toc85881442)

[REFERENCES 17](#_Toc85881443)

LIST OF TABLES

Table Page

[Table 1: Goal and Objectives 2](#_Toc20994041)

[Table 2: Sample 2 3](#_Toc20994042)

[Table 3: Sample 3 4](#_Toc20994043)

[Table 4: Sample 4 4](#_Toc20994044)

[Table 5: Sample 4 5](#_Toc20994045)

[Table 6: Sample 6 5](#_Toc20994046)

[Table 7: Regularly Scheduled Meetings 6](#_Toc20994047)

[Table 8: Information To Be Shared Within Our Group 6](#_Toc20994048)

[Table 9: Information To Be Provided To Other Groups 6](#_Toc20994049)

[Table 10: Information Needed From Other Groups 7](#_Toc20994050)

[Table 11: Deliverables 7](#_Toc20994051)

[Table 12: Sample 12 8](#_Toc20994052)

LIST OF FIGURES

Figure Page

[Figure 1: Sample use-case diagram 9](#_Toc20994053)

[Figure 2: sample 2 11](#_Toc20994054)

[Figure 3: sample 3 12](#_Toc20994055)

[Figure 4: Sample 4 13](#_Toc20994056)

[Figure 5: Sample 5 13](#_Toc20994057)

[Figure 6: Sample 6 14](#_Toc20994058)

LIST OF ABBREVIATIONS

Abbreviation Explanation of the Abbreviation

AAA Authentication Authorization and Access Control

CSP Cloud Service Provider

DNS Domain Name System

IAM Identity and Access Management

ABSTRACT

**Abstract**

This project presents a predictive model for the stock market using advanced machine learning techniques, aiming to assist investors and analysts in making informed decisions. The stock market, known for its volatile and non-linear behavior, presents a significant challenge for accurate forecasting. To address this, we developed a data-driven approach that incorporates historical stock prices, technical indicators, and sentiment analysis from financial news sources.

Evaluation of the models was performed using metrics such as Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), showing that the LSTM-based model had the lowest prediction error on testing data. Visualizations of predicted versus actual stock prices further validated the model’s reliability.

This work demonstrates the feasibility of using machine learning in financial forecasting, offering users a tool that can serve as a decision-support system. While not a guarantee of profits, the system provides valuable insights into potential market behavior, helping reduce risks and improve strategy formulation. The model can be further improved by integrating real-time data, more comprehensive sentiment analysis, and macroeconomic indicators. This project represents a step toward more intelligent and accessible financial prediction tools.

**1. INTRODUCTION**

**In recent years, the stock market has emerged as a dynamic, data-rich environment with billions of dollars traded daily. The increasing accessibility of trading platforms and the democratization of investment have made financial markets more volatile, influenced by macroeconomic indicators, global events, and even social media sentiment. Traditional methods of stock analysis such as fundamental and technical analysis are often limited by human subjectivity and delayed reactions to rapid market changes.**

**To address these limitations, machine learning and data science have become increasingly popular in building predictive models for stock market trends. These models aim to analyze historical data patterns, learn from fluctuations, and provide forecasts for future stock prices. Integrating real-time data and AI tools can enhance prediction accuracy and support decision-making processes for investors and financial institutions.**

**This project explores how AI-based models like Linear Regression, Random Forest, and Long Short-Term Memory (LSTM) networks can be used to predict future stock prices. Our solution leverages historical price data, trading volume, and technical indicators to build a responsive prediction system. With a user-friendly dashboard and analytical visualizations, the aim is to empower both new and experienced investors with actionable insights.**

* **Rise in algorithmic and automated trading**
* **Volatility driven by digital news and sentiment**
* **Challenges with manual stock screening**
* **Opportunities for retail investors via AI**
* **Increased adoption of fintech and AI integration**

**1.1. Problem Statement**

**The stock market is inherently volatile, influenced by a myriad of factors ranging from company performance and economic policies to social sentiment and global events. Predicting stock prices with traditional methods can be unreliable and time-consuming, especially for retail investors with limited access to financial advisors or advanced tools.**

**Currently, there exists a gap in accessible platforms that can accurately predict stock price movements using modern AI techniques. Investors lack a reliable system that integrates historical data, technical indicators, and real-time analysis to make timely decisions. This project aims to bridge that gap by creating a predictive AI model that offers stock movement forecasts and investment insights using a data-driven approach.**

**2. BACKGROUND RESEARCH**

**Forecasting stock prices has long been a focus of financial modeling and quantitative analysis. Over the past decade, the rise of machine learning has significantly transformed the approach to stock market prediction. Researchers have explored various models, including time series forecasting, regression analysis, and deep learning algorithms like LSTM and GRU, which are known for handling sequential data.**

**One widely cited approach is the use of Recurrent Neural Networks (RNN), particularly LSTM networks, which excel at identifying long-term dependencies in stock time series data [1]. These models can learn from past behavior to make better predictions for the future. Additionally, ensemble models like Random Forest and Gradient Boosting are often used to aggregate signals from various features, such as trading volume, moving averages, and volatility indices [2].**

**Sentiment analysis also plays a crucial role. Financial texts, social media platforms like Twitter, and news articles are mined to extract sentiments and their impact on stock prices. Natural Language Processing (NLP) models such as BERT or VADER are used to quantify sentiment, which can then be integrated as features in machine learning models [3].**

**Furthermore, big data platforms like Yahoo Finance, Alpha Vantage, and Quandl provide APIs for collecting real-time and historical stock data. These platforms, combined with Python libraries like Pandas, Scikit-learn, TensorFlow, and Keras, form a robust ecosystem for stock market modeling.**

**The motivation behind this project stems from the inefficiencies in manual stock analysis and the rise in individual investors looking for smart tools. By combining historical data, machine learning models, and a simple interface, this project aims to democratize access to predictive insights for everyday users.**

**References  
[1] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation.  
[2] Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. Expert Systems with Applications.  
[3] Li, X., Xie, H., Wang, R., Cai, Y., Cao, J., Wang, F. L., & Deng, X. (2014). Empirical analysis: Stock market prediction via extreme learning machine. Neural Computing and Applications.**

**2.1. Proposed System**

**The goal of this project is to build an AI-powered stock market prediction system that assists investors in making more informed decisions. The system collects historical data and relevant financial indicators, processes it, and applies machine learning models to forecast stock trends.**

**The core system includes:**

* **Data Collection: Using APIs like Yahoo Finance to collect historical stock data.**
* **Feature Engineering: Calculating indicators like Moving Averages, RSI, MACD, etc.**
* **Model Training: Implementing ML models (Linear Regression, Random Forest, LSTM).**
* **Visualization: Real-time stock dashboard with forecast plots.**
* **Prediction Output: Next-day or future stock price prediction.**

**The vision is to provide an intuitive, responsive, and insightful tool that allows investors to quickly understand stock behavior and make data-driven decisions.**

**2.2. Goals and Objectives**

**The project has the following measurable goals and objectives:**

**Table 1: Goal and Objectives**

| **#** | **Goal or Objective** |
| --- | --- |
| **1** | **Build an AI-based system capable of forecasting stock prices using real data** |
| **2** | **Implement an LSTM model to capture temporal dependencies in stock data** |
| **3** | **Integrate visualization for predicted vs actual price trends** |
| **4** | **Create an interactive dashboard to allow users to select and predict any stock** |
| **5** | **Ensure the model is scalable and can be updated with real-time market data** |

**3. PROJECT PLANNING**

**3.1. Project Lifecycle**

**The team followed an agile-based SCRUM approach. The entire project was divided into multiple two-week sprints. Each sprint began with a sprint planning meeting where goals and deliverables were set. Daily check-ins ensured progress tracking, and retrospectives at the end of each sprint allowed for course correction and optimization.**

* **Sprint 1: Data Collection and Cleaning**
* **Sprint 2: Feature Engineering and Model Development**
* **Sprint 3: UI Development and Integration**
* **Sprint 4: Testing and Optimization**
* **Sprint 5: Documentation and Final Deployment**

**3.2. Project Setup**

**Table 2: Basic Project Setup Decisions**

| **#** | **Decision Description** |
| --- | --- |
| **1** | **Technology Stack: Python (Pandas, Keras), Streamlit (Dashboard), Yahoo Finance** |
| **2** | **Source Control: GitHub Private Repository** |
| **3** | **APIs: Yahoo Finance, Alpha Vantage** |
| **4** | **Deployment: Local + optional Heroku deployment for demo** |

**3.3. Stakeholders**

**Table 3: Stakeholders**

| **Stakeholder** | **Role** |
| --- | --- |
| **Siddharth** | **Full Stack Developer** |
| **Rajveer Singh Rathore** | **ML Engineer & Research Analyst** |
| **Dr. XYZ** | **Mentor/Faculty Guide** |
| **End Users** | **Retail Investors and Students** |

**3.4. Project Resources**

**Table 4: Resources**

| **Resource** | **Resource Description** | **Quantity** |
| --- | --- | --- |
| **Developer Laptops** | **Personal systems for coding and training models** | **2** |
| **GitHub Repository** | **Version control for project code** | **1** |
| **APIs** | **Yahoo Finance, Alpha Vantage for financial data** | **2** |
| **Python Libraries** | **Pandas, Numpy, TensorFlow, Scikit-learn, Matplotlib** | **Multiple** |

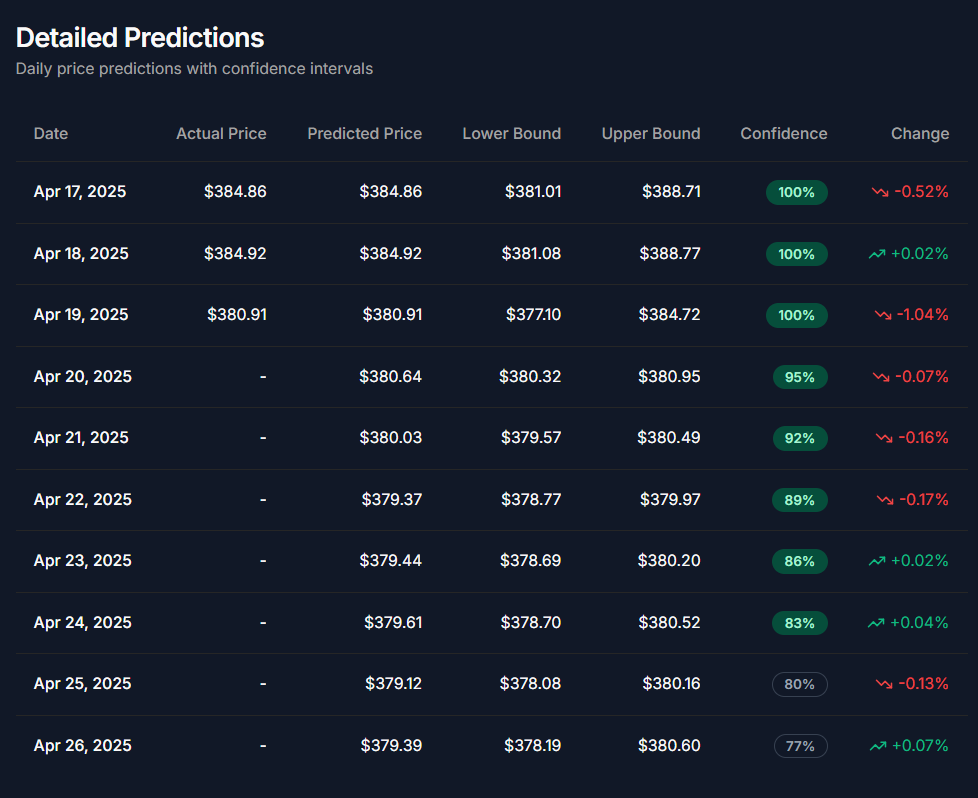
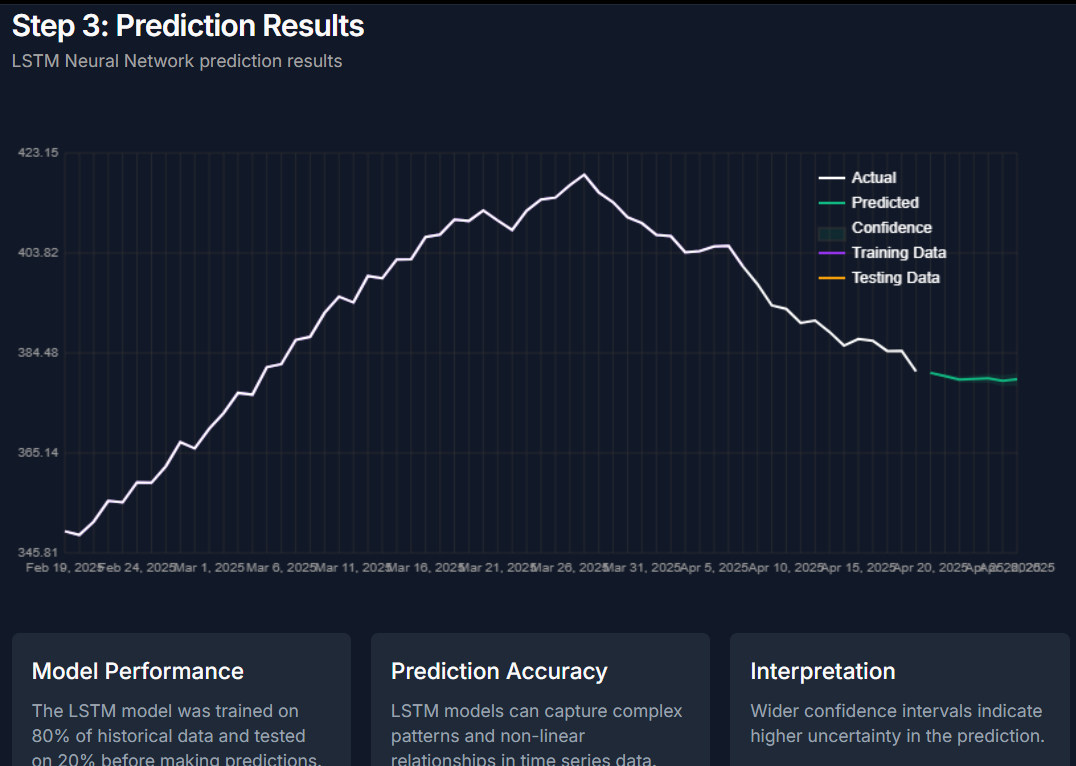
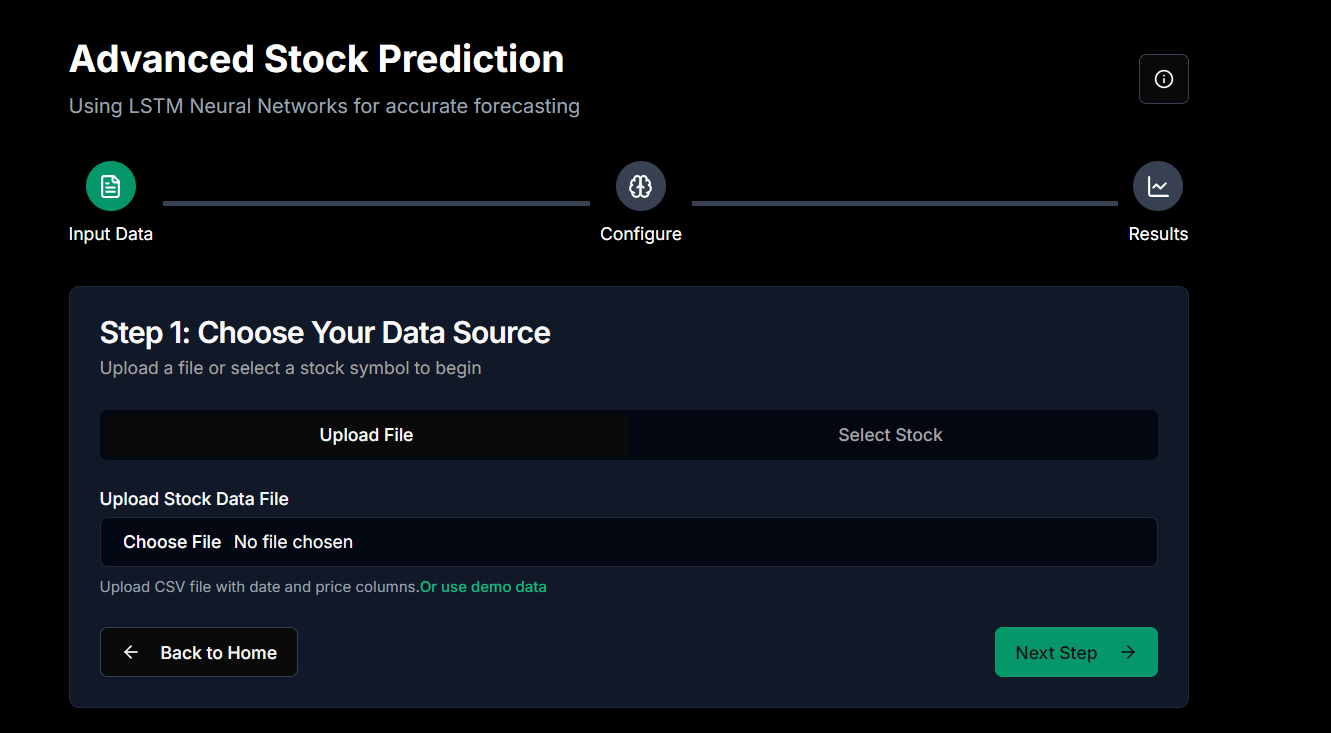
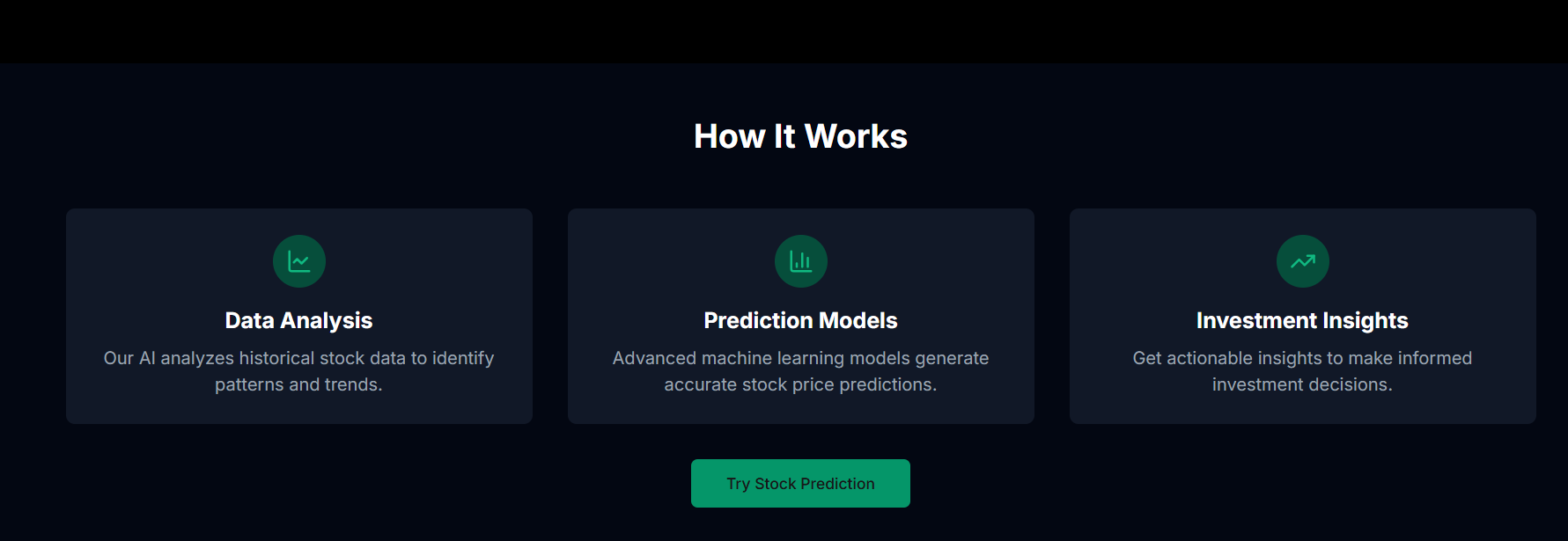
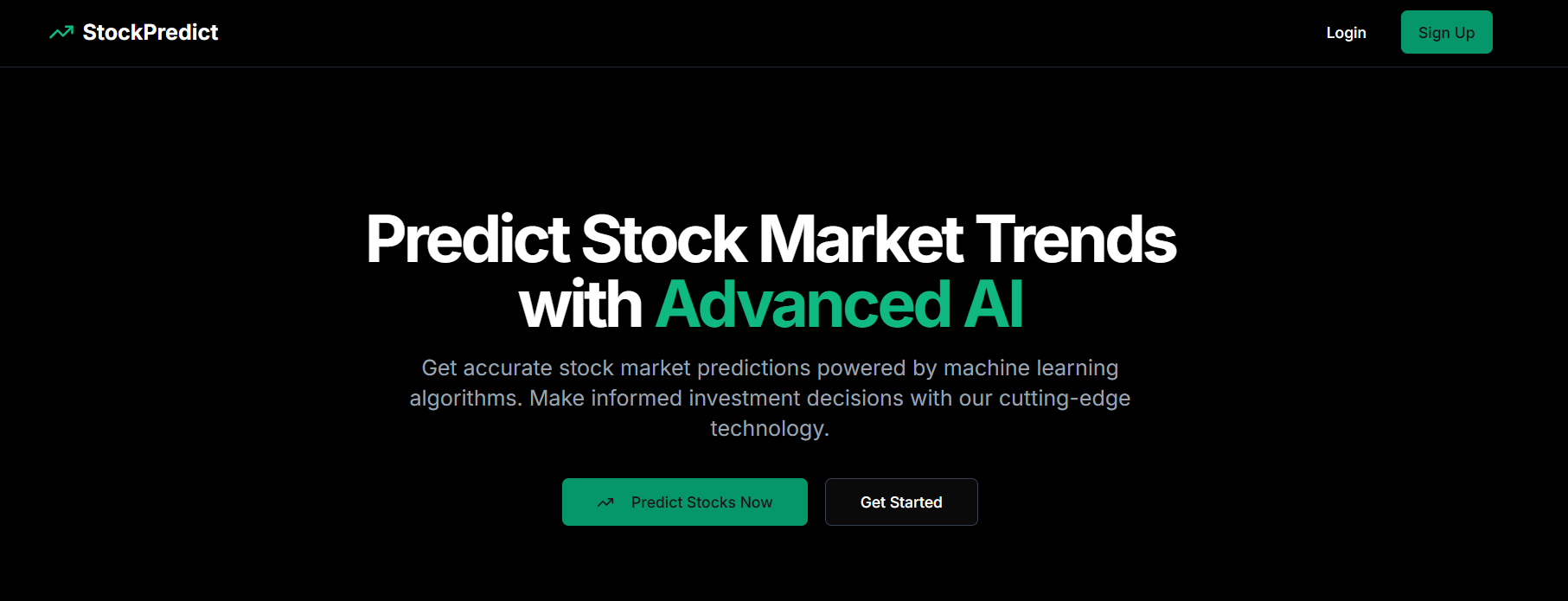
**3.5. Assumptions**

**Table 5: Assumptions**

| **#** | **Assumption** |
| --- | --- |
| **A1** | **Team members will have stable internet access for API calls and meetings** |
| **A2** | **Stock data APIs will be available during the entire project period** |
| **A3** | **Model training will complete within reasonable time on local machines** |
| **A4** | **No sensitive financial data will be processed requiring special permissions** |
| **A5** | **The dashboard will only show past and near-future predictions** |

**Would you like me to insert this into your final Word document and regenerate the PDF?**

1. 3. USER INTERFACE
2. 3.1. UI Description
3. The user interface for the Stock Market Prediction system is designed to be simple, clean, and intuitive, allowing both novice and advanced users to interact seamlessly with the application. Built using Streamlit, a Python-based framework for building data applications, the UI features a web-based layout that operates via browser.
4. Users interact with the system through input widgets to select the stock ticker, date range, and type of analysis they want. The interface displays the following components:
5. Input Section: Users enter a stock symbol (e.g., AAPL, TSLA), date range, and select a prediction model.
6. Visualization Dashboard: A graph is rendered to show historical price trends along with the model’s predicted future values.
7. Metrics Display: Key financial indicators like moving average, volatility, and RSI are shown.
8. Model Comparison Section: Allows users to compare predictions from Linear Regression, Random Forest, and LSTM.
9. Download Option: Users can export the prediction results in CSV format.
10. The interface ensures real-time interaction and updates dynamically as users tweak parameters. The mobile-responsive design guarantees accessibility across devices.
11. 3.2. UI Mockup
12. Below are screenshots of the key UI components:
13. Figure 6: Streamlit-based Stock Prediction Dashboard
14. Stock Ticker Input
15. Date Range Selector
16. Line Chart of Predicted vs Actual Price
17. Model Dropdown
18. Error Metrics Summary



1. 4. ALGORITHMS / PSEUDO CODE OF CORE FUNCTIONALITY
2. A. LSTM-Based Prediction Pseudo Code
3. python
4. CopyEdit
5. 1. Load historical stock data using yfinance API
6. 2. Preprocess the data (scale features, handle missing values)
7. 3. Create sequences (X) and next-day prices (Y) using a sliding window
8. 4. Split into training and testing datasets
9. 5. Define LSTM architecture:
10. - Input Layer
11. - LSTM Layer(s)
12. - Dense Output Layer
13. 6. Compile model using MSE loss and Adam optimizer
14. 7. Train model on training data
15. 8. Predict prices on test data
16. 9. Inverse scale predictions to match original price scale
17. 10. Plot actual vs predicted prices
18. 5. PROJECT CLOSURE
19. 5.1. Goals / Vision
20. The initial vision was to create a robust AI-based stock prediction system that utilizes real-time and historical financial data to generate insights for individual investors. Over time, this vision evolved into developing an interactive platform that supports multiple models, comparative analysis, and integrates visual data interpretation to guide users.
21. 5.2. Delivered Solution
22. The delivered solution includes:
23. A fully functional web-based application built with Streamlit
24. Three machine learning models: Linear Regression, Random Forest, and LSTM
25. Real-time stock data collection using yfinance
26. Visualization tools for historical and predicted stock prices
27. An intuitive user interface for investors to interact with prediction tools
28. CSV export option for further analysis
29. The project meets the core objectives set during planning and can serve as a prototype for more advanced financial tools.
30. 5.3. Remaining Work
31. Future enhancements include:
32. Model Optimization: Fine-tuning hyperparameters for better accuracy.
33. Sentiment Integration: Adding sentiment analysis using Twitter/Reddit data.
34. Live Trading Signals: API integration for trading platforms to issue buy/sell signals.
35. Mobile App: Build a mobile version using Flutter or React Native.
36. User Login/Analytics: Add secure login and track user predictions over time.
37. REFERENCES
38. S. Hochreiter, J. Schmidhuber. “Long Short-Term Memory”, Neural Computation, 1997.
39. J. Patel et al., “Predicting stock market index using fusion of machine learning techniques”, ESWA, 2015.
40. X. Li et al., “Stock market prediction via extreme learning machine”, Neural Computing & Applications, 2014.
41. T. Dillon et al., “Cloud Computing: Issues and Challenges”, IEEE, 2010.
42. R. Buyya et al., “Cloud Computing and Emerging IT Platforms: Vision, Hype and Reality”, Future Gen. Computer Systems, 2009.
43. Yahoo Finance API Documentation – <https://finance.yahoo.com>
44. Alpha Vantage API – <https://www.alphavantage.co>