A REPORT ON

Analysis and Prediction of Stock Price Returns based on Indian Market Data (2000–2020)

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

OF

BACHELOR OF ENGINEERING

(COMPUTER ENGINEERING)

SUBMITTED BY

Gayatri Kurulkar	41039
Sakshi Ingale	41021
Advait Shinde	41058
Yash Jangale	41027

UNDER THE GUIDENCE OF

Prof.Dr.Ms.Deipali Gore

DEPARTMENT OF COMPUTER ENGINEERING

P.E.S MODERN COLLEGE OF ENGINEERING

PUNE - 411005.

SAVITRIBAI PHULE PUNE UNIVERSITY

[2024- 25]

Progressive Education Society's **Modern College of Engineering**, Shivaji Nagar, Pune- 411005.

Certificate

This is to certify that the Mini-Project report entitled,

Analysis and Prediction of Stock Price Returns based on Indian Market Data (2000–2020)

Submitted by

Gayatri Kurulkar 41039
Sakshi Ingale 41021
Advait Shinde 41058
Yash Jangale 41027

Academic Year: 2024-2025

This bonafide work is carried out by the student under the supervision of . Prof.Dr.Ms.Deipali Gore and the Mini-Project report is approved for the partial fulfillment of the requirements for the degree of Bachelor of Engineering (Computer Engineering) of Savitribai Phule Pune University, Pune.

.Ms.Deipali Gore

Internal Supervisor

Prof. Dr.(Mrs.)S. A. Itkar

Head of Department

(Computer Engineering)

Contents

Introduction	1
Objectives	7
Motivation	10
Scope and rational of the study	13
Methodological details	16
Algorithm	23
Results	23
Analysis	26
Inferences and Conclusion	29
Acknowledgement	30
List of reference	33

Abstract:

This mini project focuses on analyzing and predicting stock price returns using Indian stock market data from 2000 to 2020. The goal is to explore historical trends and predict future stock movements with machine learning techniques. The project begins with Exploratory Data Analysis (EDA) to identify key patterns, including major events like the 2008 financial crisis and the 2020 COVID-19 market crash. Important features such as moving averages and volatility are extracted to be used in predictive models. LSTM, a type of recurrent neural network designed for time-series forecasting, is particularly effective for capturing long-term dependencies in the stock data. The project is visualized using Streamlit, enabling users to visualize historical data, analyze trends, and predict future stock prices interactively through a web application. The results show that simpler models like Linear Regression provide limited insights for volatile markets, whereas LSTM models deliver more accurate predictions by learning sequential patterns. However, challenges such as overfitting and sensitivity to market volatility remain. The project highlights the potential of machine learning in stock price forecasting, with future improvements possibly incorporating additional data sources like news sentiment and exploring more sophisticated prediction models.

Keywords: Stock price prediction, Machine learning, Time-series forecasting, LSTM, Exploratory Data Analysis (EDA), Feature engineering, Stock market trends, Volatility, Streamlit, Financial forecasting, Stock returns.

Chapter 1

Introduction

1.1.Introduction:

The stock market is a complex and dynamic system influenced by a wide range of factors, such as economic conditions, political events, investor sentiment, and global crises. Predicting stock price movements is a challenging task, yet it holds significant importance for investors and financial institutions seeking to make informed decisions. This mini project focuses on analyzing historical stock price data from the Indian market between 2000 and 2020 and using machine learning techniques to predict future stock price returns.

The objective of this project is twofold: first, to analyze the historical trends and fluctuations in stock prices, identifying key events that shaped the market; and second, to implement machine learning models that can predict future stock price movements. By combining Exploratory Data Analysis (EDA) with advanced prediction techniques, this project aims to provide insights into the patterns of stock price behavior and the factors influencing market volatility.

Machine learning models Long Short-Term Memory (LSTM) networks are used to develop predictive models. Among these, LSTM—a specialized deep learning model for sequential data—has been particularly effective for time-series forecasting. The project is implemented using Streamlit, Python framework for building interactive web applications, allowing users to visualize historical stock data and experiment with different prediction models.

In this report, we will explore the dataset, methodologies, and results of the project, providing a detailed evaluation of the models' performance and the challenges faced in predicting stock price returns. We also highlight areas for future improvement and discuss the limitations of the current approach in capturing stock market volatility.

Chapter 2

Objectives

2.1: Objectives of the Project

- Stock Market Analysis:
  - Perform Exploratory Data Analysis (EDA) on historical Indian stock market data (2000–2020) to identify key trends, market patterns, and significant events such as economic booms, recessions, and crashes.
  - Analyze the ups and downs of stock prices to understand factors influencing market fluctuations and investor behavior over time.

- Feature Engineering:
  - Extract important stock market indicators such as moving averages, volatility, and returns that can help in improving the prediction models.

- Predict Future Stock Price Returns:
  - Implement RNN machine learning model LSTM to predict future stock price movements based on historical data.
  - Evaluate the performance of the model and compare their accuracy in predicting stock price returns over different time horizons.

- Build an Interactive Web Application:
  - Develop a user-friendly interface using Streamlit to allow users to interact with the model, visualize historical stock data, and make predictions for future stock prices.

- Provide Insights for Future Stock Market Movements:
  - Identify strengths and limitations of each predictive model and provide recommendations for improving prediction accuracy in future research.
  - Explore the impact of external events (e.g., global crises, political changes) on stock price volatility and how predictive models can account for these factors.

- Future Enhancements:
  - Propose future improvements, such as incorporating additional data sources (e.g., news sentiment, macroeconomic indicators) and exploring more sophisticated models to improve prediction reliability.

Chapter 3
Motivation

3.1 Motivation

- Dynamic Nature of Stock Markets: The stock market's volatility and fluctuations, influenced by economic trends, political events, and investor sentiment, present opportunities for data-driven analysis.

- Significant Historical Events: Notable events such as the 2008 financial crisis and the 2020 COVID-19 pandemic have led to substantial swings in stock prices, highlighting the need for better predictive models.

- Data Availability: The increasing availability of historical financial data allows for the application of advanced machine learning techniques to analyze and predict stock price movements.

- Advancements in Machine Learning: Rapid developments in machine learning and artificial intelligence make it feasible to build sophisticated models capable of learning complex patterns from stock market data.

- Practical Application: Developing an interactive solution using Streamlit enables users to explore stock market trends, visualize data, and experiment with different predictive models.

- Risk Mitigation and Informed Decisions: Enhanced prediction capabilities can help investors and financial analysts make more informed decisions, mitigate risks, and optimize returns in a volatile market.

- Real-World Implications: The project serves as a practical application of data science in finance, demonstrating how technology can improve stock market predictions and support investment strategies.

Chapter 4

Scope and rationale of the Study

4.1:Scope and Rationale

4.1.1 : Scope:

The scope of this project encompasses both data analysis and machine learning model development for predicting stock price returns using Indian stock market data from 2000 to 2020. The first step involves Exploratory Data Analysis (EDA), where the dataset will be thoroughly examined to identify key trends, patterns, and any anomalies in stock price movements. This includes visualizing important features such as market volatility, stock returns, and moving averages. Feature engineering will be performed to extract these key financial indicators, which will serve as inputs for machine learning models.

The project will implement various machine learning models, including LSTM to predict future stock price movements. Each model will be evaluated based on its predictive accuracy using metrics like Mean Squared Error (MSE) and R-squared, allowing for a comparative analysis to identify the best-performing model. This will enable the development of a robust prediction model tailored for the complexity and volatility of the stock market.

An interactive web application will be developed using Streamlit, allowing users to interact with the data and the prediction models. This app will enable users to visualize historical stock data, analyze trends, and predict future stock price movements based on different time frames and model parameters. The application will be user-friendly, providing real-time insights into stock market trends and making it accessible for investors, financial analysts, and stakeholders.

The project's scope also includes addressing the limitations of the models and proposing future enhancements. For example, additional data sources like news sentiment analysis or macroeconomic indicators can be integrated to improve the model's predictive accuracy. Lastly, more advanced models like Transformer-based models or reinforcement learning techniques could be explored for developing automated trading strategies.

4.1.2. Rationale

- Need for Accurate Stock Market Predictions: The stock market's volatility, influenced by external factors like economic trends, political events, and global crises, makes it essential for investors and financial analysts to have accurate prediction methods.

- Limitations of Traditional Methods: Traditional stock forecasting techniques often fail to capture the non-linear and unpredictable nature of financial markets, creating a need for more advanced approaches.

- Advancement through Machine Learning: Machine learning models, especially those designed for time-series data like LSTM, can capture complex patterns and long-term dependencies in stock prices, providing more reliable and accurate forecasts.

- Bridging the Gap: The project aims to combine traditional stock market analysis with modern machine learning techniques, enabling better predictions by learning from historical data..

- Practical Application: Developing an interactive Streamlit web application makes advanced financial forecasting tools accessible to users, allowing them to visualize data and make predictions in real time.

- Democratization of Tools: The project provides a user-friendly platform that offers actionable insights for investors, financial analysts, and data enthusiasts, enhancing decision-making in stock market investments.

- Leveraging Technology for Financial Markets: The rationale is based on the belief that combining data science and machine learning can significantly improve the accuracy and timeliness of stock market predictions, which is crucial in today’s fast-moving financial landscape.

Chapter 5

Methodological details

5.1: Methodological Details for the Project:

The methodology for this project involves several key steps, from data collection to model development and deployment.

1. Data Collection & Preprocessing:
   - Historical Indian stock market data (2000–2020) is sourced and cleaned.
   - Feature engineering is applied to extract moving averages, volatility, and returns.
   - The data is split into training and testing sets for model development.

2. Exploratory Data Analysis (EDA):
   - EDA is performed to identify patterns, trends, and anomalies in stock price movements, along with visualizing market indicators and key events like the 2008 financial crisis and the 2020 pandemic.

3. Model Development:
   - Machine learning model LSTM, is implemented to predict future stock prices.
   - Models are trained on historical data and optimized using hyperparameter tuning techniques.

4. Model Evaluation:
   - Models are evaluated using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared to assess predictive accuracy.
   - The performance of each model is compared to identify the best one for stock price forecasting.

5. Deployment via Streamlit:
   - An interactive web application is developed using **Streamlit**, allowing users to upload stock data, visualize trends, and make predictions using different models.

6. Result Analysis:
   - The performance of models is analyzed, with a focus on identifying strengths and weaknesses, particularly in capturing stock market volatility.

7. Future Enhancements:

- Future improvements include incorporating sentiment analysis from news or using advanced models like Transformers for more accurate predictions and real-time trading strategies.

This methodology provides a structured approach for building stock price prediction models and delivering practical insights through a user-friendly web app.

Chapter 6

Algoritm

6.1 Algorithm

- Data Collection:
  - Input: Historical stock price data (2000-2020).
  - Output: Cleaned and structured dataset.
  - Steps:
    1. Load stock market data (Open, Close, High, Low prices, and Volume).
    2. Clean data by handling missing values and inconsistencies.

- Data Preprocessing:
  - Input: Cleaned stock data.
  - Output: Feature-engineered dataset ready for modeling.
  - Steps:
    1. Scale/normalize features to bring them within a uniform range.
    2. Perform feature engineering to create new features like moving averages, volatility, and stock returns.
    3. Split the dataset into training (80%) and testing (20%) sets.

- Exploratory Data Analysis (EDA):
  - Input: Preprocessed dataset.
  - Output: Insights and visualizations of stock trends.
  - Steps:
    1. Plot stock prices over time to identify patterns and anomalies.
    2. Visualize key indicators like volatility and moving averages.
    3. Identify key historical events (e.g., financial crises, market crashes).

- Model Development:
  - Input: Training dataset.
  - Output: Trained models.
  - Steps:
    1. Implement RNN as a baseline model.
    2. Develop ARIMA model for time-series forecasting.
    3. Create an LSTM network to capture long-term dependencies in stock price data.

- Model Training:

- Input: Training dataset.
- Output: Optimized models.
- Steps:
  1. Train each model using the training data.
  2. Optimize model parameters using techniques like grid search or random search.
  3. For LSTM, reshape the data for time-series modeling and train the network.

- Model Evaluation:

- Input: Testing dataset and trained models.
- Output: Performance metrics.
- Steps:
  1. Test the models on the test dataset.
  2. Evaluate performance using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared.
  3. Compare models to determine the most effective for stock price prediction.

- Model Deployment (Streamlit):

- Input: Trained models.
- Output: Interactive web application.
- Steps:
  1. Develop a Streamlit app that allows users to upload data, visualize trends, and select models for predictions.
  2. Provide users with real-time stock price forecasting capabilities.

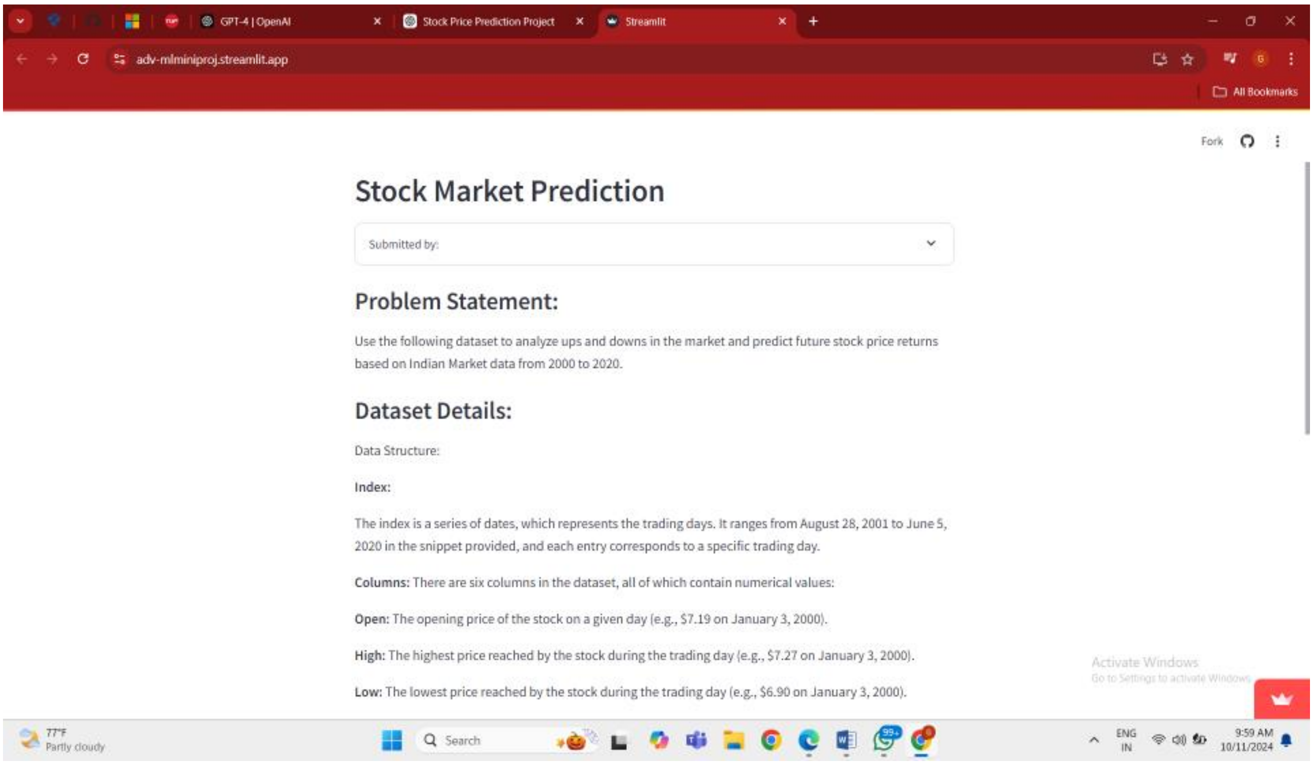
- Result Analysis:

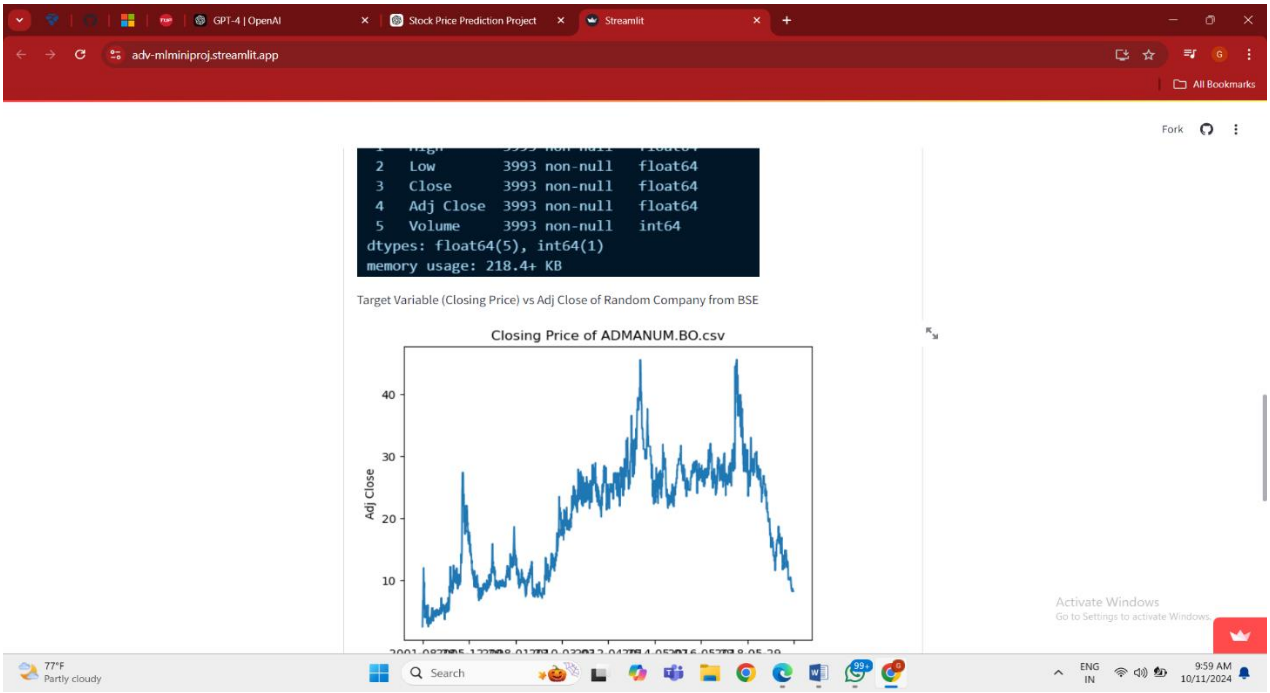
- Input: Predictions from the models.
- Output: Insights and performance comparisons.
- Steps:
  1. Analyze the accuracy of each model's predictions.
  2. Identify patterns and trends in the predicted stock prices.
  3. Highlight strengths and limitations of each approach, with a focus on LSTM's ability to handle long-term trends.

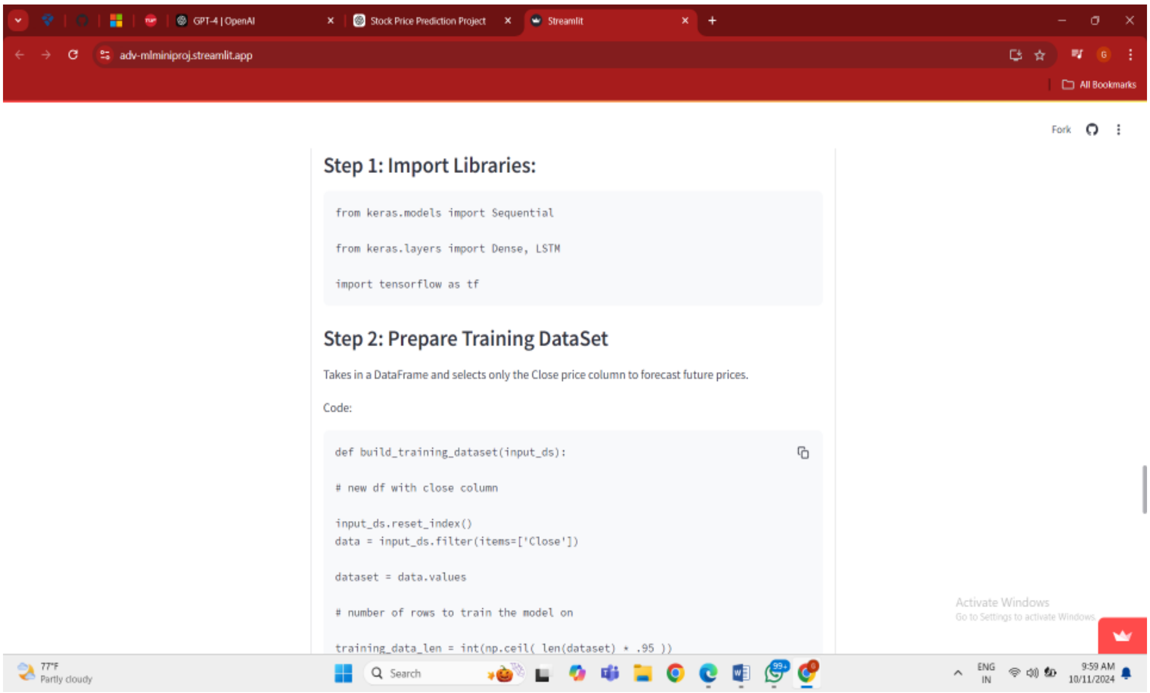
Chapter 7

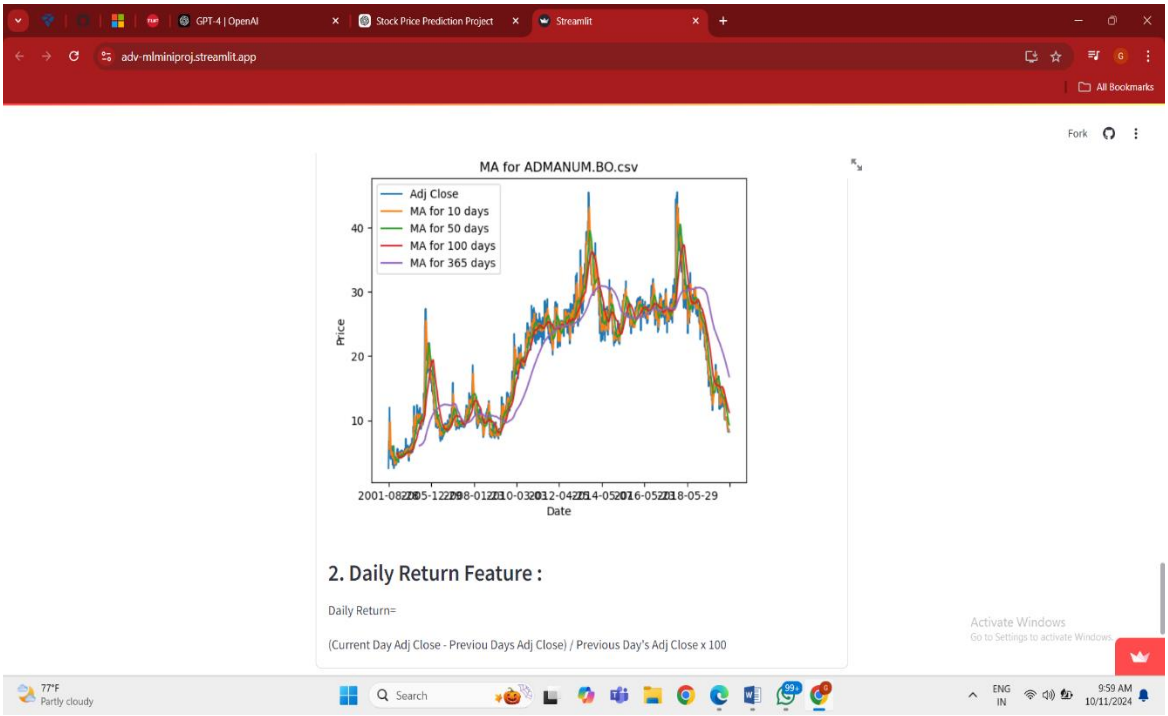
Results

7.1: Results


adv-mlminiproj.streamlit.app
Stock Market Prediction
Submitted by:
Problem Statement:
Use the following dataset to analyze ups and downs in the market and predict future stock price returns based on Indian Market data from 2000 to 2020.
Dataset Details:
Data Structure:
Index:
The index is a series of dates, which represents the trading days. It ranges from August 28, 2001 to June 5, 2020 in the snippet provided, and each entry corresponds to a specific trading day.
Columns: There are six columns in the dataset, all of which contain numerical values:
Open: The opening price of the stock on a given day (e.g., $7.19 on January 3, 2000).
High: The highest price reached by the stock during the trading day (e.g., $7.27 on January 3, 2000).
Low: The lowest price reached by the stock during the trading day (e.g., $6.90 on January 3, 2000).


adv-mlminiproj.streamlit.app
2 Low 3993 non-null float64
3 Close 3993 non-null float64
4 Adj Close 3993 non-null float64
5 Volume 3993 non-null int64
dtypes: float64(5), int64(1)
memory usage: 218.4+ KB
Target Variable (Closing Price) vs Adj Close of Random Company from BSE
Closing Price of ADMANUM.BO.csv
Adj Close
40
30
20
10


Step 1: Import Libraries:
from keras.models import Sequential
from keras.layers import Dense, LSTM
import tensorflow as tf
Step 2: Prepare Training DataSet
Takes in a DataFrame and selects only the Close price column to forecast future prices.
Code:
def build_training_dataset(input_ds):
# new df with close column
input_ds.reset_index()
data = input_ds.filter(items=['Close'])
dataset = data.values
# number of rows to train the model on
training_data_len = int(np.ceil(len(dataset) * .95))


MA for ADMANUM.BO.csv
Adj Close
MA for 10 days
MA for 50 days
MA for 100 days
MA for 365 days
Price
Date
2. Daily Return Feature :
Daily Return=
(Current Day Adj Close - Previou Days Adj Close) / Previous Day's Adj Close x 100

Chapter 8

Analysis

8.1: Analysis:

- Data Trends and Insights:
  - The stock data from 2000 to 2020 highlights market volatility during key events like the 2008 financial crisis and the 2020 pandemic. These trends are crucial for developing predictive models that can adapt to market changes.

- Model Performance:
  - LSTM excelled by learning from sequential data and predicting long-term trends.

- Challenges:

- Overfitting was a notable issue, especially with LSTM, as it tended to memorize the training data. Regularization techniques like dropout helped, but model tuning remained complex.

- Evaluation Metrics:

- Models were evaluated using MSE and RMSE. Advanced models like LSTM had the lowest RMSE, showing superior performance in handling time-dependent stock data.

- Deployment and Usability:

- The Streamlit app made the project accessible, allowing users to visualize stock data, test models, and make predictions in real-time, enhancing the practicality of the solution.

- Model Comparison:
  - LSTM was the best performer for time-series forecasting, Simpler models like Linear Regression were less effective in volatile scenarios.

- Potential for Improvement:

- Incorporating external data like economic indicators and news sentiment could enhance model accuracy. Exploring Transformer networks could also improve real-time prediction.
significant challenge.

Chapter 9

Conclusion

9.1: Conclusion:

This project demonstrates the potential of machine learning models in predicting stock price movements using historical data from the Indian stock market. Through a combination of exploratory data analysis, feature engineering, and the application of advanced machine learning techniques, the project provides insights into stock market trends and future price movements. Especially LSTM prove to be more robust and accurate, with LSTM excelling in handling sequential time-series data. The Streamlit-based web application enhances the usability of these models, allowing users to interactively visualize trends, experiment with models, and make predictions. However, the project also highlights the challenges of stock market prediction, particularly in accounting for extreme events and market shocks. Overfitting and sensitivity to sudden market changes remain significant limitations that need to be addressed. In the future, the integration of additional data sources, such as news sentiment and macroeconomic indicators, as well as the exploration of more advanced models like Transformer networks, could further improve the predictive accuracy and adaptability of the models. Overall, this project serves as a foundational step toward leveraging machine learning for more accurate stock market analysis and forecasting.

Chapter 10

Acknowledgment

10.1-Acknowledgement:

We would like to express our sincere gratitude to **Prof. Dr. Ms. Deipali Gore** whose exceptional guidance, mentorship, and unwavering support have been instrumental in the successful completion of this project on **Analysis and Prediction of Stock Price Returns based on Indian Market Data (2000–2020)**. Their invaluable insights and expertise have significantly enriched the depth and quality of this analysis. His keen understanding of Machine learning, coupled with their thoughtful guidance on Kernel programming techniques and their encouragement to explore innovative research approaches, have been invaluable in shaping the direction and outcomes of this study. Their unwavering belief in our abilities and their commitment to our success have motivated us to strive for excellence throughout this research endeavor. Our sincere thanks go to the open-source community for providing the foundational technologies and tools utilized in this project. The availability of these resources enabled us to integrate advanced encryption methods and optimize file management processes, contributing significantly to the security and efficiency of our cloud controller. We Furthermore would like to acknowledge the invaluable contributions of the open-source community. The availability of open-source software tools and libraries has significantly facilitated the data analysis and modeling processes in this research. The open-source community's commitment to sharing knowledge and resources has played a crucial role in advancing research and innovation in the field of Machine Learning. Thank you to all who contributed directly and indirectly to this project. Your support has been essential to its success.

Chapter 11

List of reference

References:

1. "Rabin-Karp Algorithm for String Matching" on GeeksforGeeks

2. Naive String Matching Algorithm" on TutorialsPoint.

3. Exact String Matching Algorithms: Theory and Practice" by Gonzalo Navarro

4. "A Fast String Matching Algorithm" by Michael O. Rabin and Richard M. Karp

5. Rabin, M. O., & Karp, R. M. (1987). *A Fast String-Matching Algorithm*. ACM Transactions on Mathematical Software (TOMS), 13(2), 127-141.
DOI: 10.1145/24475.24481

6. Citation:
Navarro, G. (2001). *A Guided Tour to Approximate String Matching*. ACM Computing Surveys (CSUR), 33(1), 31-88.
DOI: 10.1145/375360.375365

7. Knuth, D. E., Morris Jr, J. H., & Pratt, V. R. (1977). *Fast Pattern Matching in Strings*. SIAM Journal on Computing, 6(2), 323-350.
DOI: 10.1137/0206024