Phase-3

Student Name: GUNA M

Register Number: 412023106303

Institution: SRI KRISHNA ENGINEERING COLLEGE

Department : Electronics & Communication engineering

Date of Submission: 11.05.2025

Github Repository Link:

https://github.com/GUNA9578/https-github.com-GunaM-Stock-Price-Prediction-Guna

1. ProblemStatement

- Stock market prices are highly volatile and influenced by numerous unpredictable factors.
- Traditional prediction models often fail to capture non-linear patterns and trends.
- There is a need for a smart AI-based solution that can analyze historical stock data effectively.
- This project aims to develop a time series-based prediction system using models like ARIMA and LSTM.
- The goal is to assist investors, traders, and analysts with accurate and data-driven decision-making.

2. Abstract

- This project focuses on predicting future stock prices using Artificial Intelligence and Time Series models.
- It combines both Machine Learning (ARIMA) and Deep Learning (LSTM) to improve forecasting accuracy.

- Historical stock market data is collected, preprocessed, and analyzed for hidden patterns and trends.

- Models are trained and evaluated using performance metrics such as RMSE and MAE.

- The system aims to assist investors and analysts in making accurate, data-driven financial decisions.

3. SystemRequirements

Hardware Requirements:

- Processor: Intel Core i5 or higher

- RAM: Minimum 8 GB (16 GB recommended)

- Storage: At least 10 GB of free disk space

Software Requirements:

- Operating System: Windows or Linux

- Programming Language: Python 3.8 or higher

- Development Environment: Jupyter Notebook / Google Colab

- Required Libraries:

  - pandas

  - numpy

  - matplotlib

  - seaborn

  - scikit-learn

  - keras

4. Objectives

- To predict stock price trends using AI-based time series models.

- To compare the performance of traditional models (ARIMA) and deep learning models (LSTM).

- To apply data preprocessing, EDA, and feature engineering techniques.

- To evaluate model accuracy using RMSE and MAE metrics.

- To help investors make informed decisions with visualized predictions.

5. Project Work flow (Flowchart)

1. Data Collection
2. Data Preprocessing
3. Exploratory Data Analysis (EDA)
4. Feature Engineering
5. Model Selection and Training (ARIMA, LSTM)
6. Model Evaluation (RMSE, MAE)
7. Forecast Visualization

6.DatasetDescription

- Source: Yahoo Finance / Kaggle

- Dataset Type: Time Series Data

- Features Included: Date, Open, High, Low, Close, Volume

- Target Variable: Close Price (for next-day prediction)

- Number of Records: ~5,000 daily entries per stock

- Format: CSV files used for training and testing .

6. DataPreprocessing

- - Removed null or missing values from the dataset to ensure model quality.
- - Converted date columns to standard datetime format for time series compatibility.
- - Normalized or scaled numeric features to improve model training accuracy.
- - Encoded categorical variables (if any) to numerical format.
- - Split the dataset into training and testing sets to evaluate performance.
- Removed null values
- Before:
- Date Open High Low Close Volume
- 2023-01-01 150.2 152.1 NaN 151.0 3.2M
- After:
- Date Open High Low Close Volume
- 2023-01-01 150.2 152.1 150.0 151.0 3.2M
- Converted 'Date' column to datetime format
- Before:
- '2023-01-01' → string type
- After:
- 2023-01-01 → datetime64[ns]
- Scaled features using MinMaxScaler
- Before:
- Open = 145.2
- After:

- Scaled Open = 0.67
- Final Output:
- Cleaned, structured dataset ready for EDA and model building

7. Model Building

- Used ARIMA model for linear time series forecasting.

- Trained LSTM (Long Short-Term Memory) model for deep learning-based prediction.

- Applied Random Forest Regression as a traditional ML baseline.

- Split dataset into training and testing sets (e.g., 80/20 ratio).

- Tuned model hyperparameters (epochs, learning rate, p, d, q values for ARIMA).

- Handled time series sequence preparation for LSTM using sliding window technique.

- Compared models based on prediction accuracy and evaluation metrics.

9.ModelEvaluation

- Evaluated all models using RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error).

- ARIMA Model:

- RMSE: 4.35

- MAE: 3.82

- LSTM Model:

- RMSE: 2.78

- MAE: 2.45

- Random Forest:

- RMSE: 3.94

- MAE: 3.12

- Visualized predicted vs actual prices using line plots.

- Found LSTM to perform best in capturing trends and seasonality in stock prices.

10.SourceCode:

https://github.com/GUNA9578/https-github.com-GunaM-Stock-Price-Prediction-Guna

11.FutureScope

- Integrate real-time stock data streaming for live prediction.

- Include news sentiment analysis to enhance model predictions.

- Deploy the model as a web or mobile application for user interaction.

- Expand to multi-stock or portfolio forecasting.

- Optimize models using advanced deep learning techniques like Transformers.

12.Conclusion

- The project successfully demonstrated stock price prediction using time series analysis.

- LSTM outperformed other models due to its ability to learn temporal dependencies.

- The model provided valuable insights and can assist investors with better decision-making.

- With further enhancements, the system can be scaled for real-world financial forecasting.

13.Team Members and Roles

KANNISH S - Data cleaning

GUNA M - Model development

HARIGURUBHARATHI - Documentation