Phase-2 Submission

Student Name: GUNA M

Register Number: 412023106303

Institution: SRI KRISHNA ENGINEERING COLLEGE:4120

Department: B.E,ELECTRONICS AND COMMUNICATION ENGINEERING

Date of Submission: 09/05/2025

GitHub Repository Link:

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

---

1. Problem Statement

Stock market prediction is a complex task influenced by various factors, including economic indicators, news sentiment, and historical price trends. This project aims to solve the problem of ***predicting future stock prices using AI-based time series analysis*** *techniques. The focus is on applying machine learning and deep learning models (e.g., ARIMA, LSTM) to analyze and forecast price movements using historical stock data.*

- ***Problem Type****: Regression*
- ***Why it Matters****: Accurate stock prediction supports investors in making informed decisions, helps manage risk, and contributes to building intelligent trading systems. Automating predictions can reduce human error and lead to more data-driven financial strategies.*

2. Project Objectives

- *Apply time series analysis techniques for stock price forecasting.*
- *Compare traditional ML models (e.g., Linear Regression, Random Forest) with deep learning models (e.g., LSTM).*
- *Optimize model performance using RMSE and MAE as evaluation metrics.*
- *Explore feature engineering using technical indicators like Moving Averages and RSI.*
- *Improve model accuracy by experimenting with window sizes, normalization, and sequence lengths.*

3. Flowchart of the Project Workflow *:*

css

CopyEdit

Data Collection → Data Preprocessing → EDA → Feature Engineering →

Train-Test Split → Model Building (ML & DL) → Evaluation → Visualization → Conclusion

4. Data Description

- ***Dataset Name**: Historical Stock Price Data*
- ***Source**: Yahoo Finance (via yfinance API)*
- ***Data Type**: Time-series, structured*
- ***Number of Records**: ~5,000 rows per stock (daily data)*
- ***Features**: Date, Open, High, Low, Close, Adj Close, Volume*

- ***Target Variable***: *'Close' price (for next time step prediction)*
- ***Dataset Nature***: *Dynamic*

5. Data Preprocessing

- *Handled missing values using forward fill and interpolation*
- *Checked and removed duplicate rows (none found)*
- *Converted date column to datetime format and set as index*
- *Added moving averages (7-day, 30-day), RSI as engineered features*
- *Normalized data using MinMaxScaler*
- *Converted time series into supervised learning format using windowing*

6. Exploratory Data Analysis (EDA)

- ***Univariate Analysis***: *Line plots of stock prices over time, volume distribution*
- ***Bivariate Analysis***: *Correlation matrix between technical indicators and target*
- ***Multivariate Analysis***: *Pairplots of price with indicators (MA, RSI)*
- ***Insights Summary***:
  - *Strong autocorrelation in stock prices*
  - *Short-term moving averages show predictive signals*
  - *Volume spikes often precede volatility*

7. Feature Engineering

- Generated lagged features for Close prices
- Calculated rolling means (MA7, MA30)
- Added RSI (Relative Strength Index)
- Created trend direction (binary up/down) as an auxiliary label
- Justified each feature based on technical analysis strategies
- *Justify each feature added or removed.*

8. Model Building

- ***Models Used***:

- *Linear Regression (baseline)*
- *Random Forest Regressor*
- *LSTM (Long Short-Term Memory) network*

- ***Data Split**: 80% training, 20% testing (time-aware split)*
- ***Evaluation Metrics**:*
  - *MAE, RMSE, R^2 Score*
  - *LSTM outperformed traditional models in RMSE*

9. Visualization of Results & Model Insights

- *Line plots of predicted vs. actual stock prices*
- *Residual plots to analyze prediction errors*
- *Feature importance plot from Random Forest*
- *LSTM learning curves (loss vs. epochs)*

10. Tools and Technologies Used

- ***Language**: Python*
- ***IDE/Notebook**: Jupyter Notebook*
- ***Libraries**: pandas, numpy, matplotlib, seaborn, scikit-learn, TensorFlow/Keras, yfinance*

- ***Visualization***: *Matplotlib, Seaborn, Plotly*

11. Team Members and Contributions

KANNISH S - Data cleaning

GUNA M - Model development

HARIGURUBHARATHI - Documentation