Phase -2 Submission Template

Student Name: Logeshwari S

Register Number: 623323104023

Institution: Vetri Vinayaha College of Engineering and Technology

Department: B.E (cse)

Date of Submission: 07/05/2025

Problem Statement:

Stock markets are inherently volatile and influenced by numerous macro and microeconomic factors. Investors and analysts face significant challenges in predicting stock prices accurately. This project aims to leverage AI and time series analysis to forecast stock prices, enabling informed decision-making and reducing investment risk.

Project Objectives:

* To collect and preprocess historical stock market data.
* To perform time series analysis and uncover trends, seasonality, and anomalies.
* To engineer features that enhance model performance.
* To build and compare different AI models for stock price prediction.
* To visualize predictions and derive actionable insights.
* To develop an interactive dashboard for real-time predictions and analysis.

Flowchart of Project Workflow:

Start

|

Data Collection

|

Data Preprocessing

|

Exploratory Data Analysis

|

Feature Engineering

|

Model Building (ARIMA, LSTM, Prophet, etc.)

|

Model Evaluation & Tuning

|

Result Visualization & Insights

|

Dashboard Integration (Optional)

|

End

Data Description:

# Source:

Yahoo Finance, Alpha Vantage, or Quandl APIs.

# Attributes:

1. Date
2. Open
3. High
4. Low
5. Close
6. Adjusted Close
7. Volume

Data Preprocessing:

* Handling missing values.
* Normalizing/Standardizing features.
* Log transformation for stationarity.
* Date formatting and time-indexing.
* Outlier detection and treatment.

Exploratory Data Analysis (EDA):

* Time series decomposition (trend, seasonality, residual).
* Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots.
* Moving averages (SMA, EMA).
* Volatility analysis.
* Correlation with external economic indicators.

Feature Engineering:

* Lag features (e.g., previous day prices).
* Rolling statistics (mean, std).
* Technical indicators (RSI, MACD, Bollinger Bands).
* Categorical time-based features (Day, Month, Year, Weekday).
* Event-based signals (earnings, dividends).

Model Building:

* Statistical Models: ARIMA, SARIMA
* Machine Learning Models: XGBoost, Random Forest
* Deep Learning Models: LSTM, GRU, Transformer
* Hybrid Models: Ensemble of above techniques

Visualization of Results & Model Insights:

* Actual vs Predicted price plots.
* Forecasting intervals and confidence bands.
* Feature importance (ML models).
* Model performance metrics (RMSE, MAE, MAPE).
* Interactive dashboard using Streamlit/Power BI/Tableau.

Tools and Technologies Used:

1. Languages: Python
2. Libraries: Pandas, NumPy, Scikit-learn, TensorFlow/Keras, Statsmodels, Prophet, Matplotlib, Seaborn, Plotly
3. APIs: Yahoo Finance, Alpha Vantage
4. Visualization: Power BI, Streamlit, Tableau
5. Version Control: Git, GitHub
6. Notebook: Jupyter

Team Members & Contributions

1. Project Lead & Data Analyst –Jaya sree S
2. AI/ML /data Engineer – Atchaya K
3. Visualization & Presentation Specialist –Logeshwari S
4. Documentation & Report Writer – Kaviya. S