

Phase-2 Submission Template

Student Name: ROHITH S

Register Number: 422223243048

Institution: SURYA GROUP OF INSTITUTIONS

Department: B.Tech/ARTIFICIAL INTELLIGENCE & DATA
SCIENCE

Date of Submission: 04/05/2025

Github Repository Link:

<https://github.com/ROHITH-0211/Phase-2/blob/main/README.md>

1. Problem Statement

The project aims to predict stock prices using AI techniques, addressing the need for accurate forecasting in financial markets.

Type of Problem:

Regression (predicting continuous values).

Relevance: Accurate stock price predictions can help investors make informed decisions, enhancing financial strategies and minimizing risks.

2. Project Objectives

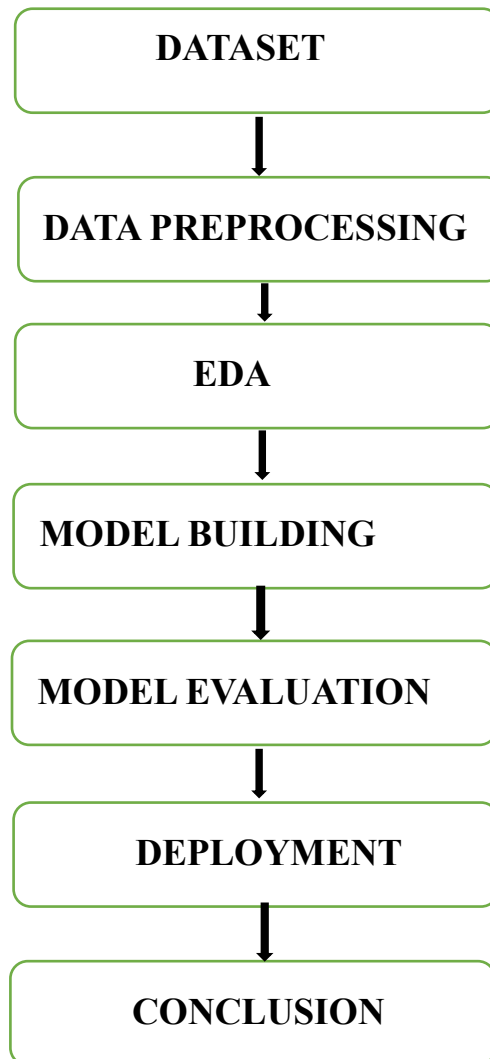
Objectives:

Collect historical stock price data and relevant financial indicators.

Clean and preprocess the data for analysis. Build predictive models using machine learning algorithms.

Evaluate model performance based on accuracy and other metrics.

2. Flowchart of the Project Workflow



3. Data Description.

Dataset Name: Historical Stock Price

Dataset.Source: Financial data APIs (e.g., Yahoo Finance, Alpha Vantage).

Type of Data: Time-series data (stock prices, volume, etc.).

Records and Features: Include features like date, open, high, low, close, and volume.

Target Variable: Future stock price

5. Data Preprocessing

Handle missing values through imputation or removal.

Remove duplicates and outliers.

Normalize or standardize features where necessary.

6. Exploratory Data Analysis (EDA)

Perform univariate and bivariate analysis to understand trends and relationships in the data.

Visualize distributions and correlations.

7. Feature Engineering

Create new features like moving averages, RSI, or other financial indicators.

Select relevant features based on their impact on stock price.

8. Model Building

Models Selected:

Linear Regression (baseline model).

Random Forest or LSTM (for capturing time-series patterns).

Justification: These models are suitable for regression tasks and can handle complex relationships

9. Visualization of Results & Model Insights

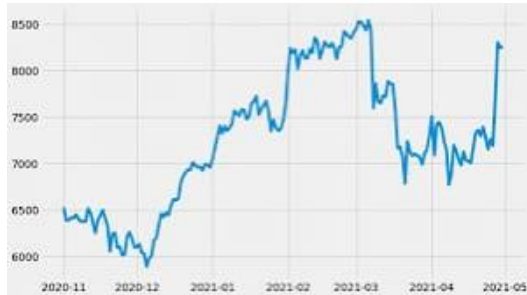
Include visualizations like:

1. Prediction vs. Actual plots:

This plot compares the predicted stock prices against the actual stock prices over a specific time period.

It helps in visually assessing how well the model performs.

Description: The x-axis represents the time (e.g., dates), while the y-axis shows the stock prices. The plot includes two lines: one for actual prices and one for predicted prices.



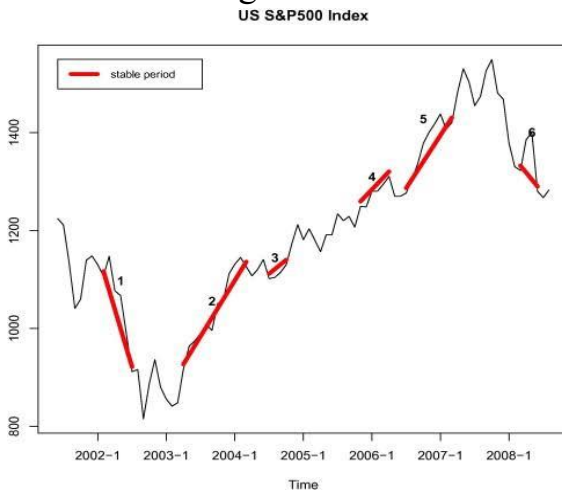
2, Feature importance plots:

This bar chart displays the importance of each feature used in the model.

Understanding feature importance helps in identifying which factors most significantly influence stock price predictions.

Description:

The x-axis represents the importance score, and the y-axis lists the features. Higher scores indicate greater influence on the model's predictions.



10. Tools and Technologies Used

Programming Language: Python.IDE/

Notebook: Jupyter Notebook.

Libraries: pandas, numpy, scikit-learn, matplotlib, TensorFlow/Keras (for LSTM).

11. Team Members and Contributions

1. Rohit S (422223243048)

Role: Data Collection, Exploratory Data Analysis (EDA)

Contributions:

Data Collection: Rohit was responsible for gathering historical stock price data from various financial APIs, ensuring the dataset was comprehensive and relevant for the analysis. He focused on acquiring data that included essential features such as opening prices, closing prices, volume, and other market indicators.

Exploratory Data Analysis (EDA): After collecting the data, Rohit conducted thorough exploratory data analysis. This involved visualizing data distributions, identifying trends, and uncovering patterns. He used tools like matplotlib and seaborn to create informative plots, which helped the team understand the data's behavior and prepare for further analysis.

2. Vadivelan R (422223243057)

Role: Model Building, Feature Engineering

Contributions:

Model Building: Vadivelan took the lead in developing the predictive models. He experimented with various algorithms, including Linear Regression, Random Forest, and LSTM networks, to determine which model provided the best performance for stock price prediction.

Feature Engineering:

He also focused on enhancing the dataset through feature engineering. This included creating new features such as moving averages, Relative Strength Index (RSI), and other technical indicators that could improve the model's predictive capabilities. His efforts ensured that the models had access to relevant and useful data.

3. Thison Bero (422223243056)

Role: Model Evaluation, Report Documentation

Contributions:

Model Evaluation:

Thison was tasked with evaluating the performance of the predictive models. He used metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared to assess how well the models were performing. He also conducted cross-validation to ensure the models were robust and reliable.

Report Documentation:

In addition to model evaluation, Thison compiled the project documentation. He documented the methodologies used, results obtained, and insights drawn from the analysis. His thorough documentation ensured that the project could be understood and replicated by others.

4. Raguman R (422223243046)

Role: Deployment, Dashboard Creation, Presentation.

Contributions:

Deployment: Raguman was responsible for deploying the final model into a production environment. He set up the necessary infrastructure to ensure the model could be accessed and used by stakeholders.

Dashboard Creation: He also created an interactive dashboard that visualized the model's predictions and key metrics. This dashboard allowed users to easily view stock price forecasts and understand the underlying data.

Presentation: Lastly, Raguman prepared and delivered the project presentation. He summarized the project objectives, methodologies, and findings, effectively communicating the team's work to an audience. His presentation skills helped convey the significance of the project and its potential impact.