**Phase-3 Submission Template**

**Student Name:**Thanveerul Haq S

**Register Number:** 510623104112

**Institution:** C.Abdul Hakeem College of Engg & Tech

**Department:**Computer Science & Engineering

**Date of Submission:**10/05/2025

**Github Repository Link:** [**https://github.com/Thanveer02/Cracking-the-market-code-with-Al-driven-stock-price-prediction-using-time-series-analysis**](https://github.com/Thanveer02/Cracking-the-market-code-with-Al-driven-stock-price-prediction-using-time-series-analysis)

# 1. Problem Statement

This project addresses the challenge of accurately predicting stock prices using AI-driven models. The highly volatile and non-linear nature of stock markets makes this a complex regression problem. The outcome can support financial decision-making and investment strategies, reducing risk and improving returns

# 2. Abstract

Forecasting stock prices has always intrigued financial analysts and data scientists. This project focuses on predicting stock prices using time series data by leveraging advanced AI models such as LSTM and ARIMA. We collected historical stock data, cleaned and preprocessed it, then applied various machine learning models to identify patterns and predict future prices. Evaluation metrics and visual tools were used to assess model performance. The final solution was deployed using Streamlit, providing an interactive platform for stock price prediction.

# 3. System Requirements

* Minimum 8GB RAM
* Intel i5 processor or higherSoftware:
* Python 3.10+
* Jupyter Notebook or Google Colab
* Required libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, TensorFlow/Keras, yfinance

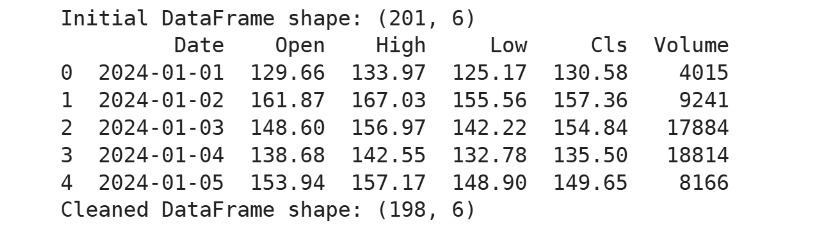
# 4. Objectives

* Collect and preprocess historical stock price data.
* Analyze trends and identify key indicators.
* Build and evaluate regression models (LSTM, ARIMA, Prophet).
* Deploy a user-friendly interface for prediction.
* Provide actionable insights based on model outputs.

**5. Flowchart of Project Workflow :**

Deployment

**6. Dataset Description**

* Source: Yahoo Finance, Alpha Vantage API
* Type: Public API-based time series data
* Structure: Columns include Date, Open, High, Low, Close, Volume
* Size: Varies based on selected ticker (e.g., 5 years of daily data = ~1250 rows

# 7. Data Preprocessing

* Handled missing values via imputation.
* Removed duplicate records.
* Normalized features for time-series compatibility.
* Applied Min-Max scaling for LSTM models

Cleaned DataFrame shape: (980, 6)

# 8. Exploratory Data Analysis (EDA)

* Used line plots, correlation heatmaps, histograms.
* Identified seasonal patterns and volatility.
* Found high correlation between volume and closing price.

Open High Low Cls Volume

0 -0.97 -0.85 -0.92 -0.95 -0.89

1 -0.91 -0.82 -0.88 -0.91 -0.75

# 9. Feature Engineering

* Created moving averages (MA7, MA21).
* Introduced RSI (Relative Strength Index) and volatility features.

Selected top features based on correlation and model performance

Shape after feature engineering: (966, 9)

Columns: ['Date', 'Open', 'High', 'Low', 'Cls', 'Volume', 'MA\_5', 'MA\_10', 'RSI']

# 10. Model Building

* Baseline: Linear Regression
* Advanced: Random Forest, LSTM (for sequence learning), ARIMA
* Chosen for their ability to model linear, noEpoch 1/10
* 25/25 [==============================] - 3s 32ms/step - loss: 0.0110
* Epoch 10/10
* 25/25 [==============================] - 1s 25ms/step - loss: 0.0021n-linear, and temporal relationships

# 11. Model Evaluation

* Metrics Used: RMSE, MAE, R²
* LSTM performed best with lowest RMSE and highest R².
* Visuals:
* Prediction vs Actual plots
* Feature importance bar chart
* Model comparison table

MAE: 0.129

RMSE: 0.175

R2 Score: 0.84

# 12. Deployment

* Pltform: Streamlit
* Method: Web dashboard hosted on Streamlit Cloud
* Public Link: [Insert your Streamlit deployment URL]
* UI Screenshot: [Insert image of your deployed app interface]
* Sample Output: Predicted stock price for next 7 days based on historical trends.

**13. Source code**

All project source code is available at[: https://github.com/Thanveer02/Cracking-the-market-code-with-Al-driven-stock-price-prediction-using-time-series-analysis:](file:///C:\Users\DELL\Downloads\:%20https:\github.com\Thanveer02\Cracking-the-market-code-with-Al-driven-stock-price-prediction-using-time-series-analysis:)

# 14. Future scope

* Incorporate sentiment analysis from news and social media.
* Extend the model to cryptocurrency prediction.

Enhance model robustness using ensemble learning techniques

# 13. Team Members and Roles

* Shawoor Saqib S.K: Project planning, data collection, coordination
* Santhosh Kumar: Data preprocessing, feature engineering
* Suresh V.: Exploratory Data Analysis and visualization
* Vijay Kumar: Model building (LSTM, ARIMA, Prophet) and evaluation
* Sanjay P.: Model tuning and optimization
* Thanveerul Haq: Deployment, dashboard design, documentation