

Phase-1 Submission Template

Student Name: Emaya Bharath B

Register Number: 410723104014

Institution: Dhanalakshmi College Of Engineering

Department: Computer Science And Engineering

Date of Submission: 28-04-2025

1.Problem Statement

Cracking the market code with AI-driven stock price prediction using time series analysis

2.Objectives of the Project

- To build a robust AI model that can predict stock prices based on historical market data.
- To evaluate and compare different time series models such as ARIMA, LSTM, and Prophet for forecasting accuracy.
- To identify significant trends, seasonality, and anomalies in stock data using EDA techniques.
- To present findings via interactive visualizations and, if feasible, deploy a web-based dashboard for live or demo predictions.

3.Scope of the Project

Scope of the Project Features to Analyze/Build:

Time series forecasting models (e.g., ARIMA, LSTM, Prophet)

Exploratory Data Analysis (EDA) to uncover hidden patterns

Feature engineering for lag variables, moving averages, etc.

Model comparison and performance evaluation

Limitations/Constraints:

Focus will be on a limited number of stocks (e.g., Apple, Tesla, etc.)

Only historical stock price data (Open, Close, High, Low, Volume) will be used

No deployment on a live trading platform (demo-only application if deployed)

Constraints on real-time predictions unless APIs are used

4.Data Sources

Data set link : [Amazon Stock Market Data \(2015-2024\)](#) 

Source: (Kaggle), and it is public dataset and it is a dynamic dataset.

Type: Public data

Access: Downloadable via API or library functions

Nature: Static for training and experimentation; can be extended to dynamic updates for demo

5.High-Level Methodology

*** Data Collection**

- Gather historical Amazon stock data from Yahoo Finance/Alpha Vantage APIs
- .
- Scrape financial news and social media sentiment (ethical practices ensured).
- Use synthetic data for missing periods (e.g., holidays).

Data Cleaning

- Handle missing values via interpolation or deletion.
- Normalize inconsistent formats (e.g., date formats, currency).
- Detect outliers in trading volumes or price spikes.

Exploratory Data Analysis (EDA)

- Use **time series plots**, candlestick charts, and heatmaps to identify trends.
- Analyze correlations between stock prices and financial indicators.
- Perform clustering to group similar market conditions.

Feature Engineering

- Extract lag features (e.g., 7-day moving average).
- Derive sentiment scores from text data using NLP (e.g., VADER, BERT).
- Create volatility indices (e.g., Bollinger Bands).

Model Building

- Experiment with **ARIMA/SARIMA** for time series forecasting.
- Use **LSTM/GRU** neural networks for sequential data modeling
- Explore **ensemble models** (Random Forest, XGBoost) with technical indicators.

Model Evaluation

- Measure performance using **MAE, RMSE, and MAPE**.
- Validate robustness via **walk-forward validation**.
- Conduct **A/B testing** on trading strategies.

Visualization & Interpretation

- Build **interactive dashboards** (Plotly, Tableau) for trend visualization.
- Use **candlestick charts** and **MACD plots** to display predictions.
- Generate explainable reports using SHAP/LIME.

Deployment

- Deploy as a **web application** using Flask/Django.
- Integrate real-time data feeds via **API connections**.

6.Tools and Technologies

Programming Language: Python

Notebook/IDE: Jupyter Notebook, Google Colab

Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn, TensorFlow/Keras, yfinance

Deployment Tools Streamlit, FastAPI

7.Team Members and Roles

S.NO	NAMES	ROLES	RESPONSIBILITY
1	Mathesh S	Leader	Data Collection
2	Dhanajayan S	Member	Data Cleaning and Feature Engineering
3	Manoj C	Member	Visualization and Interpretation
4	Emaya Bharath	Member	Exploratory Data Analysis
5	Jayanth R	Member	Model Building and Model Evaluation