Phase-2 Submission Template

STUDENT NAME: **DHANUSH S**

REGISTER NUMBER : **622423104008**

INSTITUTION: **Salem College Of Engineering and Technology**

DEPARTMENT: **Computer science and Engineering**

DATE OF SUBMISSION: **15-05-2025**

CRACKING THEMARKET CODE WITH AL-DRIVEN STOCK PRICE PREDICTION USING TIME SERIES ANALYSIS

PROBLEM STATEMENT :

Financial markets are inherently noisy, non-stationary, and influenced by a complex mix of economic, behavioral, and geopolitical factors. Traditional statistical forecasting methods (e.g. ARIMA, exponential smoothing) often struggle to capture nonlinear patterns and regime shifts, leading to sub-optimal prediction accuracy and increased risk for investors.

OBJECTIVE:

To develop an AI-driven model that accurately predicts stock prices by leveraging time series analysis techniques, aiming to uncover hidden patterns and insights within financial market data to enhance prediction reliability and support strategic investment decisions.

KEY AND OUTCOMES:

1. Accurate Stock Price Prediction:

   Develop an AI-based time series model that significantly improves the accuracy of shortterm and medium-term stock price forecasts.

2. Identification of Market Patterns:

   Discover and analyze hidden trends, seasonality, and anomalies in financial time series data using advanced AI techniques.

3. Model Benchmarking:

   Compare the AI-driven models (like LSTM, GRU, Transformer) against traditional methods (like ARIMA) and demonstrate superior performance.

4. Deployment-Ready Prototype:

   Create a functional prediction system or dashboard capable of real-time or near-real-time forecasting with actionable outputs for investors.

5. Investment Decision Support:

   Provide reliable predictive insights that can assist traders and investors in making better timing and strategy decisions.

6. Risk and Uncertainty Estimation:

   Incorporate uncertainty quantification in forecasts (confidence intervals, prediction intervals) to better manage risk in trading strategies.

SCOPE OF THE PROJECT:

1. Data Collection and Preprocessing:

   Gather historical stock price data (open, high, low, close, volume) and relevant external features (e.g., news sentiment, technical indicators). Perform data cleaning, normalization, and feature engineering to prepare the dataset for modeling.

2. Time Series Analysis:

   Analyze trends, seasonality, and anomalies in the data. Apply techniques like stationarity testing, autocorrelation analysis, and decomposition.

3. AI Model Development:

   Build and train deep learning models (e.g., LSTM, GRU, Transformer) tailored for time series forecasting. Optimize hyperparameters and architecture to improve prediction accuracy.

4. Model Evaluation:

   Evaluate models using appropriate metrics like RMSE, MAPE, and directional accuracy. Compare AI models against traditional models like ARIMA and statistical baselines.

5. Forecasting and Visualization:

   Generate future stock price predictions with associated confidence intervals. Visualize results using interactive dashboards or charts for easy interpretation.

6. Deployment (Optional/Extended Scope):

Develop a prototype system for real-time prediction updates using live market feeds. Integrate model predictions into a simple decisionsupport tool for users.

DATA SOURCE:

1. Financial Market Data Providers:

Yahoo Finance (via yfinance Python library) Google Finance Alpha Vantage API Quandl (now part of Nasdaq Data Link) IEX Cloud

2. Historical Stock Prices:

Open, High, Low, Close (OHLC) prices Volume of stocks traded

3. Technical Indicators (Optional Enhancement):

Moving Averages, RSI, MACD, Bollinger Bands, etc. (can be calculated using libraries like ta-lib or pandas-ta)

4. External Datasets (Optional for Enrichment):

Financial news sentiment datasets (e.g., from Reddit, Twitter, or news APIs) Economic indicators (interest rates, CPI, GDP growth) Corporate financial reports (for fundamentalsdriven features)

HIGH LEVEL METHODOLOGY:

1. Data Collection:

Gather historical stock price data (OHLC + volume) using APIs like Yahoo Finance, Alpha Vantage, etc.

2. Data Preprocessing:

   Handle missing data, outliers, and anomalies. Normalize/scale the data (important for many AI models). Create time series features (lags, rolling means, etc.) if necessary.

3. Exploratory Data Analysis (EDA):

   Visualize trends, seasonality, and volatility. Analyze correlations and moving averages.

4. Selection:

   Traditional models: ARIMA, SARIMA, Exponential Smoothing (as baseline). AI models: LSTM (Long Short-Term Memory) networks GRU (Gated Recurrent Unit) networks Transformerbased models (for more advanced setups)

5. Model Training:

   Train models on training datasets. Validate on separate validation sets to tune hyperparameters.

6. Model Evaluation:

   Use metrics like RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percentage Error) to evaluate performance. Compare AI models vs. traditional time series models.

7. Prediction and Visualization:

   Predict future stock prices. Plot predicted vs actual prices to visualize model performance. 8. Deployment (optional): Deploy model into an app or dashboard (e.g., Streamlit, Flask) for live predictions.

TOOLS AND TECHNOLOGIES:

1. Programming Language:

   Python (primary choice for AI and time series analysis)

2. Libraries and Frameworks:

Data Handling: Pandas, NumPy Visualization: Matplotlib, Seaborn, Plotly Time Series Analysis: Statsmodels (for ARIMA, SARIMA) Facebook Prophet (optional for easier forecasting) Machine Learning/Deep Learning: Scikit-learn (for baseline models and preprocessing) TensorFlow/Keras or PyTorch (for LSTM, GRU, Transformer models)

3. Data Sources/APIs:

Yahoo Finance (yfinance library) Alpha Vantage API Quandl API

4. Model Deployment (optional):

Streamlit (for web app) Flask/Django (for backend deployment)

5. Other Tools:

Jupyter Notebook / Google Colab (for development) Git/GitHub (for version control) VS Code or PyCharm

TEAM AND ROLE:

1. [S.SRI YAMINI].---Project Coordinator & Documentation Lead -Responsible for organizims meetings, setting project milestones, and ensuring timely progress. -Maintains comprehensive project documentation, including meeting notes, model design, and final report formattin. Oversees integration of work from all members into a unified and cohesive final submission.

2. [N.SUJITHA]-----Data Preparation & Preprocessing Specialist-Manages acquisition and handling of the MNIST dataset using tools like Keras or PyTorch. -Performs preprocessing tasks such as normalization, reshaping, and one-hot encoding. -Ensures the dataset is in optimal condition for deep learning model training and testing.

3. [P.R.SRI SACHIN]---Model Development & Evaluation Designs and implements the deep learning model using CNN architectures in TensorFlow or PyTorch Trains and validates the model, tunes hyperparameters, and ensures high accuracy. -Evaluates performance using confusion matrix, accuracy, precision, and recall metrics. -Assists with preparing a demo interface or basic deployment setup if applicable.