Phase-1 Submission

Student Name: Gokul R

Register Number: 712523104022

Institution: PPG INSTITUTE OF TECHNOLOGY

Department: Computer Science and Engineering

Date of Submission: 25.04.2025

1.Problem Statement

The problem is predicting stock prices accurately amid market volatility, which challenges investors and businesses alike. AI-driven time series analysis can uncover complex patterns in data, enabling smarter investment decisions, reducing risks, and promoting economic stability.

2.Objectives of the Project

The goal of this project is to build an AI model that predicts stock price trends using time series analysis. It will generate precise forecasts, uncover market patterns, and provide actionable insights to enhance investment strategies and mitigate risks effectively.

3.Scope of the Project

The project will analyze stock prices, technical indicators, and market sentiment using time series models like ARIMA and LSTM. It focuses on identifying patterns for accurate predictions, but relies on specific datasets and is limited to simulation rather than real-world trading applications.

4.Data Sources

The dataset will include historical stock prices, trading volumes, and technical indicators sourced from public APIs like Yahoo Finance or Alpha Vantage. It will be dynamic, with real-time updates, supplemented by sentiment data from public platforms like github and news sources.

Data source link : Stock-Market-Prediction-Using-Time-Series-Analysis/Dataset at master · farhanhira/Stock-Market-Prediction-Using-Time-Series-Analysis · GitHub

5.High-Level Methodology

- **Data Collection** – Obtain historical stock price data using APIs like Yahoo Finance (via `yfinance`) or Alpha Vantage for reliable and structured time series. Optionally, supplement with web scraping or synthetically generated data for additional scenarios or testing.

- **Data Cleaning** – Data cleaning includes handling missing values, removing duplicates, standardizing formats, treating outliers statistically, and refining sentiment data for relevance.

- **Exploratory Data Analysis (EDA)** – EDA will use time-series plots, correlation heatmaps, and box plots to uncover trends, relationships, and outliers.

- **Feature Engineering** – Feature engineering will involve creating new features like volatility metrics and transforming existing ones, such as scaling prices, to enhance model performance.

- **Model Building** – The project will explore ARIMA for statistical time series, LSTM for capturing temporal dependencies, and XGBoost for feature-rich datasets to ensure accurate stock price predictions.

- **Model Evaluation** – Model evaluation will use metrics like RMSE, MAE, and R-squared, alongside cross-validation strategies, to assess accuracy and

reliability.

- **Visualization & Interpretation** – Key findings and predictions will be presented through interactive dashboards, time-series charts, correlation graphs, and feature importance plots for clear and actionable insights.

- **Deployment** – The project will be deployed as an interactive dashboard or notebook using tools like Streamlit or Jupyter for user-friendly access to insights and predictions.

6.Tools and Technologies

- **Programming Language** – The primary programming language for this project will be Python, chosen for its robust data analysis, machine learning, and visualization libraries.

- **Notebook/IDE** – The project will be developed in "Jupyter Notebook" 'for its interactive environment and seamless integration with Python libraries.

- **Libraries** – Key libraries include pandas and numpy for data processing, matplotlib and seaborn for visualization, and scikit-learn and TensorFlow for modeling.

- **Optional Tools for Deployment** – Optional deployment tools include Streamlit, Flask, or FastAPI for creating interactive and user-friendly web applications.

7.Team Members and Roles

NAME	ROLE	WORK
HARISH VK	Frontend Developer	Design and build the user interface (dashboard) to visualize stock trends and predictions.
AJIN.PR	Backend Developer	Develop APIs to connect frontend with ML model, manage data flow and server logic.
GOKUL.R	ML Engineer	Implement and train time series models (e.g., LSTM), handle data preprocessing, evaluation.
KIRUTHIGA.M	Documentation & Presentation	Prepare detailed project documentation, reports, and project presentation slides.
DEVADHARSHINI.V	Deployment Engineer	Handle model deployment, integrate all components, ensure scalability and uptime.