Phase-1

Student Name: Sasivardhan SK

Register Number: 410723106029

Institution: Dhanalakshmi college of engineering

Department: Electronics and Communication Engineering

Date of Submission: 28/04/2025

1.Problem Statement

Cracking the Market Code with AI-Driven Stock Price Prediction Using Time Series Analysis

The stock market is highly dynamic and influenced by countless factors including economic indicators, political events, and investor sentiment. Accurate stock price prediction has immense value for investors, traders, and financial institutions. However, due to its volatile and non-linear nature, forecasting stock prices remains a challenging task.

This project aims to leverage AI and advanced time series analysis techniques to predict stock price movements, providing valuable insights for making informed trading decisions.

2.Objectives of the Project

1.Develop a predictive model for short-term stock price forecasting.

2.Apply time series analysis to capture market trends and patterns.

3.Compare multiple AI/ML approaches (e.g., ARIMA, LSTM) for best accuracy.

4.Provide visualizations and a dashboard displaying historical vs predicted prices.

5.Deliver actionable insights to help investors manage risk and optimize strategies.

3.Scope of the Project

1.Features to be Analyzed:

1.Stock historical prices (Open, Close, High, Low, Volume).

2.Moving averages, technical indicators (e.g., RSI, MACD).

3.Time features (day of the week, month, quarter).

2.Limitations/Constraints:

1.Focus on predicting stock closing prices only (not full market behavior).

2.Use of publicly available datasets or APIs like Yahoo Finance.

3.Short-term prediction horizon (e.g., next day or next 5 days).

3.Model designing:

1.Model assumes past trends can somewhat predict future prices (limited in cases of sudden news or black-swan events).

4.Data Sources

https://www.kaggle.com/datasets/mrsimple07/stock-price-prediction

1.Source: Yahoo Finance (using yfinance Python library) or Kaggle Datasets related to stock data.

2.Accessibility: Public

3.Type: Dynamic (updated in real-time via API calls)

5.High-Level Methodology

Data Collection

1.Use yfinance API to fetch historical stock price data (e.g., Tesla, Apple).

2.Optionally supplement with Kaggle datasets for back-testing.

- **Data Cleaning**

1.Handle missing timestamps (e.g., weekends, holidays).

2.Fill missing values using forward-fill or interpolation.

3.Ensure consistency in time intervals.

- **Exploratory Data Analysis (EDA)**

1.Plot stock price trends.

2.Analyze autocorrelation and partial autocorrelation (ACF, PACF).

3.Identify seasonality, trends, and outliers.

- **Feature Engineering**

1.Generate technical indicators (e.g., 50-day and 200-day moving averages).

2.Lag features (previous days' prices).

3.Volatility indices.

- **Model Building** –

Traditional time series models:

ARIMA/SARIMA for baseline.

Deep Learning models:

LSTM (Long Short-Term Memory Networks) for sequence prediction.

GRU (Gated Recurrent Units) for comparison.

Justification: LSTM and GRU can capture long-term dependencies and are proven effective for financial time series.

- **Model Evaluation** –

Metrics:

RMSE (Root Mean Squared Error)

MAPE (Mean Absolute Percentage Error)

Validation Strategy:

Walk-forward validation (time series cross-validation).

Training on past data, testing on most recent unseen data.

- **Visualization & Interpretation**

1.Plot historical vs predicted closing prices.

2.Dashboard to allow selection of different stocks for prediction visualization. 3.Feature importance (if applicable).

- **Deployment** –

1. Build an interactive dashboard using Streamlit.

2.Allow users to input a stock ticker and view predicted future prices.

3.Host using Streamlit Cloud or similar free hosting platforms.

6.Tools and Technologies

1. Programming Language: Python
2. Notebook/IDE: Jupyter Notebook, Google Colab 3. Libraries:

4. Data Collection: yfinance, pandas_datareader
5. Data Processing: pandas, numpy
6. Visualization: matplotlib, seaborn, plotly
7. Time Series Analysis: statsmodels, pmdarima
8. Deep Learning: TensorFlow, Keras
9. Deployment: Streamlit
10. Optional Tools for Deployment: Streamlit Cloud

7.Team Members and Roles

S.No	NAME	ROLES	RESPONSIBILITY
1	Sasivardhan S K	Leader	Data Collection, Data Cleaning
2	Joshua Prince S	Member	Visualization & Interpretation
3	Rubesh kumar S	Member	Exploratory Data Analysis (EDA), Feature Engineering
4	Mohit Sai Reddy	Member	Model Building, Model Evaluation