Phase-II

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

Student Name : Dharshini K

Register Number : 620123106018

Institution : AVS Engineering College

Department : ECE

Date of Submission : 10/05/2025

Github Repository Link :
https://github.com/DharshiniK90/Dharshini-K-Naan-Mudhalvan-Project-.git

1. Problem Statement

The stock market is inherently volatile, influenced by numerous dynamic factors such as market trends, economic indicators, political events, and investor behavior. Predicting stock prices with precision is a complex challenge that has drawn the attention of analysts and researchers for decades. Traditional models often fall short in capturing non-linear dependencies and sequential patterns in time series data.

2. Project Objective

Develop an AI-driven model to predict stock prices using historical time series data.

Improve the predictive power over traditional statistical methods by incorporating deep

learning techniques.

Enable investors and traders to make more informed decisions based on predicted trends.

Visualize results for better interpretability and actionable insights.

3. Project Workflow (Flowchart)

Data Collection → Data Preprocessing → Exploratory Data Analysis (EDA) → Feature Engineering → Train-Test Split → Model Building (ARIMA, LSTM, etc.) → Model Evaluation → Result Visualization & Insights

4. Data Description

Source: Yahoo Finance, Alpha Vantage, or Kaggle datasets.

Fields: Date, Open, High, Low, Close, Volume, and optional technical indicators.

Format: Time-series data (daily, weekly, or monthly granularity).

5. Data Preprocessing

Handling missing values (forward fill, interpolation).

Parsing date fields and setting the correct time index.

Normalizing or standardizing numerical columns.

Converting categorical variables (if any).

Smoothing or resampling data to remove noise.

6. Exploratory Data Analysis (EDA)

Time plot of stock prices over time.

Moving average plots to observe trends.Volume vs. price correlation.

Volatility and return distribution.

Heatmaps and correlation matrices.

Stationarity checks (Augmented Dickey-Fuller test).

7. Feature Engineering

Lag features (e.g., previous day's closing price).

Rolling statistics (mean, std dev).

Technical indicators (e.g., Bollinger Bands, RSI, EMA).

Date features (day of week, month, quarter).

Fourier transforms for seasonal patterns.

8. Model Building

Models considered:

Statistical Models: ARIMA, SARIMA

Machine Learning Models: Random Forest, XGBoost

Deep Learning Models: LSTM, GRU, Transformer-based models

Evaluation Metrics:

Mean Absolute Error (MAE)

Root Mean Squared Error (RMSE)

Mean Absolute Percentage Error (MAPE)

R^2 Score

9. Visualization of Results & Model Insights

Predicted vs. Actual price plots

Residual error plots

Confidence intervals for predictions

Feature importance (for ML models)

Hidden state visualization (for LSTM models)

Cumulative returns based on strategy simulation

10. Tools and Technologies Used

Languages: Python

Libraries: pandas, numpy, matplotlib, seaborn, plotly, scikit-learn, xgboost, statsmodels, keras, tensorflow, pytorch, yfinance, alpha_vantage

Platforms: Jupyter Notebook / Google Colab / VS Code

Version Control: Git/GitHub

11. Team Members and Contributions

S.No	Name	Role	Contribution
1	Dharshini.K	Data Engineer	Data Collection, Preprocessing and Feature Engineering
2	Gobika.K	Data Analyst	Performed EDA, Statistical tests and Generated Insights
3	Dharshini.S	ML Engineer	Built and tuned machine learning and deep learning models
4	Gowthami.E	Project Lead & Developer	Workflow design, integration,report writing and final visualisation setup