CRACKING THE MARKET CODE WITH AI-DRIVEN STOCK PRICE PREDICTION USING TIME SERIES ANALYSIS

PHASE-2

Student Name: E. Hemasree

Register Number: 510923205025

Institution: Global Institute of Engineering and Technology

Department: B.Tech IT

Date of Submission: 09-05-2025

Github Repository Link: https://github.com/hemasree2510/phase-2

1.Problem Statement

The volatility and complexity of the stock market make it difficult for investors to make informed decisions. This project addresses the operational challenge of forecasting short- to medium-term stock prices using AI-driven time series models. By analyzing historical stock price data, volume, and possibly sentiment indicators, the goal is to create predictive models that can anticipate future price movements with improved accuracy.

2. Project Objectives

1. Project Goals:

The primary goal of this project is to develop an AI-driven time series model to accurately predict stock prices and uncover underlying market trends. This will support data-informed investment strategies and decision-making in volatile financial markets.

2. Key Questions to Answer Through Data Analysis:

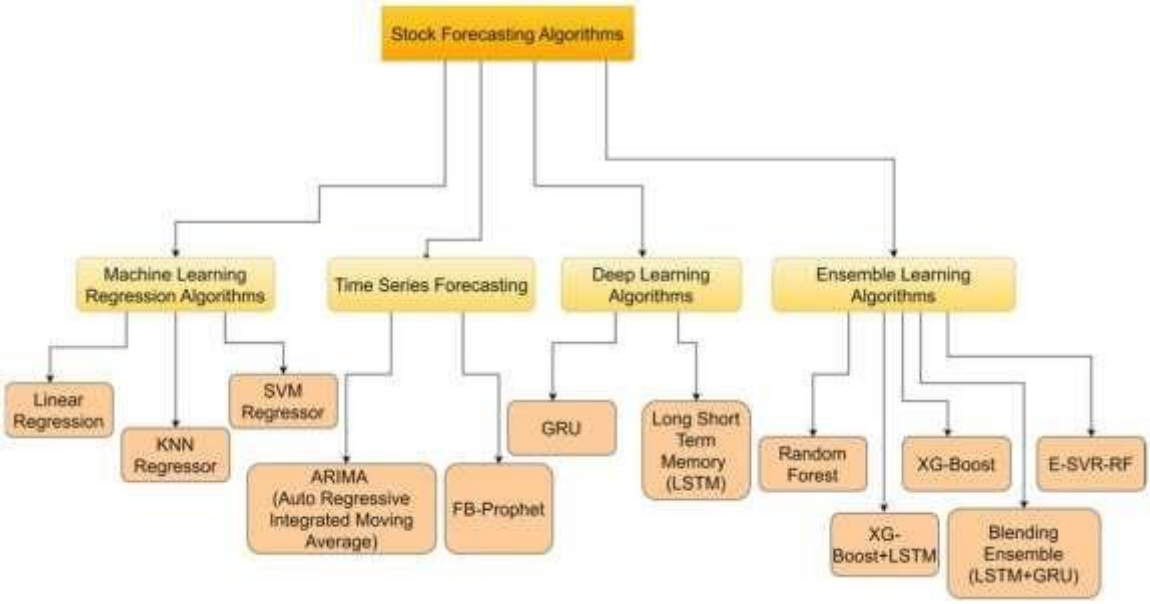
-What historical patterns and trends can be detected in the selected stock's price movement?

-How can these predictions be translated into actionable trading strategies?

3. Expected Deliverables:
Insights & Trends: Key historical patterns, seasonality, volatility, and anomaly detection.
Recommendations: Suggestions on the most reliable model and how to integrate forecasts into trading strategies.

4. Objective Revision (If Any):
After initial exploration of the data, the project scope evolved to include model comparison and the impact of external indicators . This deeper understanding of the data allowed for more refined modeling and enhanced forecasting accuracy

3. Flowchart of the Project Workflow



4. Data Description

- Dataset Name and Source:Name: Historical Stock Price Data,Source: Yahoo Finance / Kaggle

Example Dataset: "NSE Stock Data" or "S&P 500 Historical Prices"

● Data Type:Structured

The data is organized in tabular form with clearly defined columns representing different attributes of stock performance over time.

● Number of Rows and Columns:

Rows: Approximately 5,000–10,000 (depending on stock and time period)

Columns: Typically 6–10

● Static or Dynamic Dataset:

Dynamic: The dataset is continuously updated with new daily trading data, which is essential for time series analysis and model retraining.

● Key Fields or Attributes Relevant to the Problem:

Date: Timestamp for each trading entry.Open / Close Prices: Useful for trend analysis and returns calculation.High / Low Prices: Indicate volatility and intraday range

5. Data Preprocessing

1. Handling Missing Values:

- Checked for missing values in all columns.
- Forward filled (method='ffill') for missing price or volume entries to maintain time series continuity.

2. Removing Duplicates:

- Verified and removed duplicate rows based on the Date field to ensure uniqueness per trading day.

3. Formatting and Parsing Data:

- Converted the 'Date' column to datetime format for accurate time indexing.

- Set 'Date' as the DataFrame index to enable time series operations.

4. Encoding Categorical Variables:

- Not required—dataset contains only numerical and date data.

5. Identifying and Treating Outliers:

- Detected outliers using Z-score and IQR methods in price/volume columns.
- Kept outliers since they reflect real market spikes/dips (important for financial modeling)

6. Documented Transformations:

- Applied log transformation to 'Close' prices to stabilize variance.
- Normalized features (Min-Max scaling) for models like LSTM.

6. Exploratory Data Analysis (EDA)

- Univariate Analysis:Line plots for 'Close' price showed overall trend and seasonality.Histograms and boxplots revealed price distribution and occasional outliers.Volume showed high variance with occasional spikes.

- Bivariate / Multivariate Analysis:

-Heatmap of correlations:Strong positive correlation between 'Open', 'High', 'Low', and 'Close' prices.Moderate inverse relation between price and volume on certain days.

-Pairplots:Visualized linear relationships between price-based fields.

-Moving averages overlay:50-day vs. 200-day moving average crossover points identified as trend indicators.

- Key Metric Analysis (KPIs):Calculated daily returns, volatility, and rolling averages.Identified periods of high volatility and trends across different market cycles.

- Summary of Insights:Stock prices show seasonal patterns and market-driven volatility.‘Volume’ spikes often align with significant price changes.Lag-based features (e.g., previous day’s close) are potentially strong predictors.

7. Tools and Technologies Used

- Programming Language:Python – chosen for its powerful data analysis and machine learning capabilities.

- Notebook/IDE:Google Colab and Jupyter Notebook – used for writing, testing, and visualizing code interactively.

- Libraries:Pandas and numpy – for data manipulation and numerical operations.Matplotlib, seaborn, and plotly – for static and interactive visualizations.

- Pandas-profiling (optional) – for generating automated exploratory data analysis reports.These tools provided a flexible and efficient environment for handling, visualizing, and analyzing time series stock data.

8. Team Members and Contributions

NAME		CONTRIBUTIONS
Hemasree. E	-	Data Collection, cleaning and preprocessing
Raja shri.S	-	Exploratory data Analysis and visualizations
Mageshwars.S.	-	Model Building, evaluation and Report writing