Phase-3 Submission

Student Name: GOKUL R

Register Number: 712523104022

Institution: PPG INSTITUTE OF TECHNOLOGY

Department: B E CSE

Date of Submission: 15 / 05 / 2025

Github Repository Link: [github repo link]

---

1. Problem Statement

Stock price prediction is challenging due to market volatility and nonlinear data patterns. Traditional methods often fall short in capturing these complexities. This project tackles a ***regression problem*** *by using AI-based time series models to forecast future prices, helping investors make better decisions and reduce financial risks.*

2. Abstract

This project focuses on ***predicting stock prices using AI-driven time series models*** *like ARIMA and LSTM. By analyzing historical data and technical indicators, the system identifies patterns to forecast trends. The approach includes* ***preprocessing, EDA, model training, and evaluation****. Final results are deployed via an interactive dashboard to assist users in making informed investment decisions.*

3. System Requirements

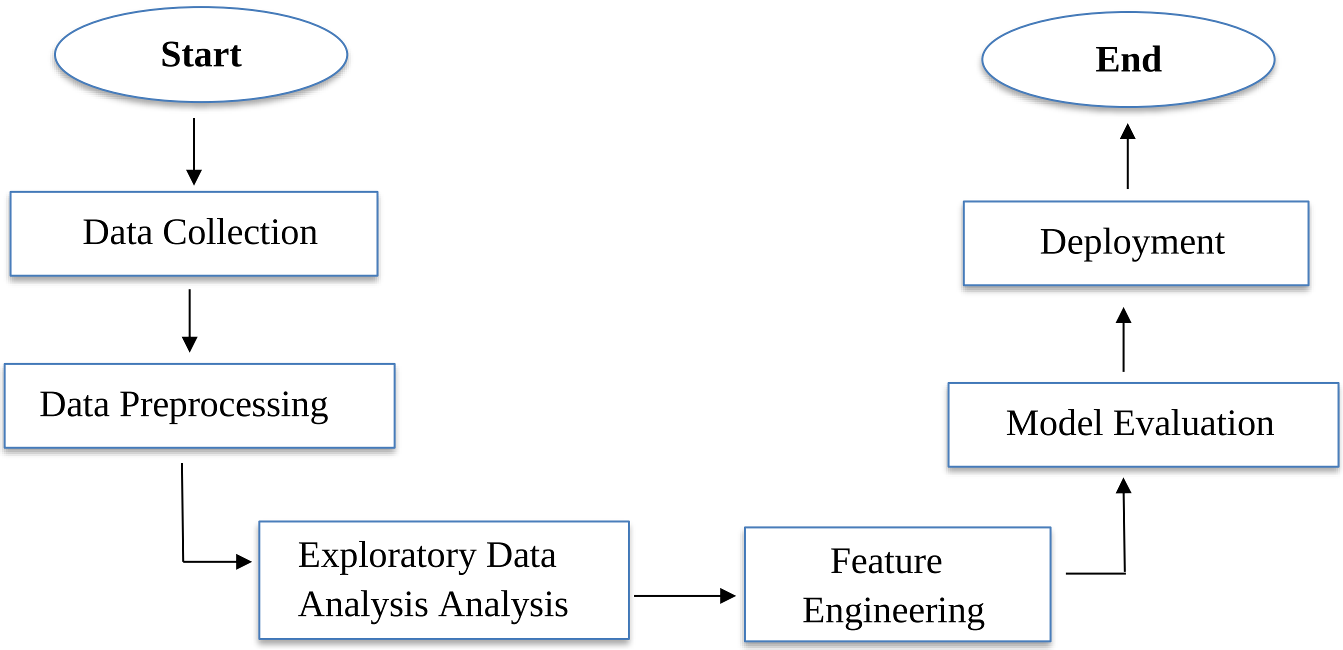
- ***Hardware:***
  - *Minimum 4GB RAM*
  - *Intel Core i5 processor.*

- ***Software:***
  - *Python 3.8+, pandas, numpy, matplotlib, seaborn, scikit-learn, TensorFlow, statsmodels, Flask, Jupyter Notebook/VS Code.*

4. Objectives

- *Build an AI-based model using time series analysis to predict future stock prices.*
- *Identify patterns and trends from historical data using indicators like SMA, EMA, and RSI.*
- *Evaluate and compare different models (ARIMA, LSTM) to determine optimal prediction accuracy.*
- *Provide investors with actionable insights to mitigate risks and improve investment strategies.*
- *Deploy the model for public access through an interactive, web-based dashboard.*

5. Flowchart of Project Workflow



6. Dataset Description

- ***Source:*** *GitHub – Public Repository (Stock-Market-Prediction-Using-Time-Series-Analysis)*

- ***Type:***:*Public dataset*
  Size: *6,000 rows with 7 columns (Date, Open, High, Low, Close, Adj Close, Volume)*

- ***Structure:*** *Time series data (daily intervals)*

	Open	Close	High	Low	Volume	RSI	MACD	Bollinger_Upper	Bollinger_Lower	Sentiment_Score	GDP_Growth	Inflation_Rate	Target
0	0.374639	0.374780	0.373510	0.378390	0.298909	0.847286	0.741715	0.367146	0.366420	0.877177	0.580868	0.038604	0
1	0.950982	0.937746	0.938422	0.946158	0.094805	0.494543	0.881343	0.938396	0.935640	0.907192	0.527044	0.108908	0
2	0.732198	0.719825	0.723644	0.723158	0.126348	0.195471	0.463179	0.710666	0.702300	0.378363	0.351052	0.432540	0
3	0.598823	0.599865	0.596973	0.605322	0.180662	0.736684	0.289076	0.593793	0.586936	0.231614	0.493274	0.946349	0
4	0.156053	0.163410	0.155891	0.166084	0.203646	0.418698	0.318761	0.164158	0.156355	0.191642	0.365116	0.074867	0

7. Data Preprocessing

- ***Missing Values:*** *Filled using forward-fill or interpolation.*
- ***Duplicates:*** *Removed after verifying row redundancy.*
- ***Outliers:*** *Detected using IQR method and capped.*
- ***Scaling:*** *Applied Min-Max scaling for LSTM.*
- ***Date Standardization:*** *Converted to datetime format for time series analysis.*
- ***Splitting:*** *Data split into 80% train, 20% test sets.*

Before Preprocessing

Date	Open	High	Low	Close	Volume
2025-01-01	150.0	151.0	149.5	150.8	1000000.0
2025-01-02	nan	153.2	151.5	152.7	1100000.0
	151.8	nan	150.7	nan	nan
2025-01-04	153.5	154.0	nan	153.8	1150000.0
2025-01-05	155.0	156.0	154.3	155.5	1200000.0

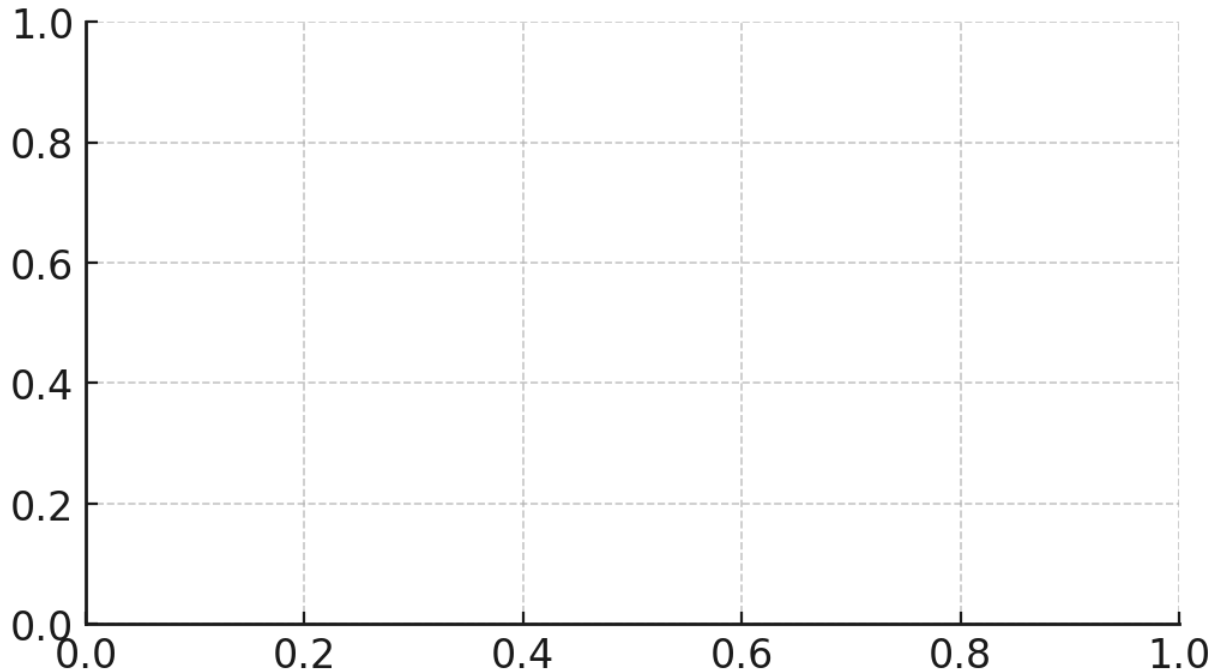
After Preprocessing

Date	Open	High	Low	Close	Volume
2025-01-01 00:00:00	150.0	151.0	149.5	150.8	1000000.0
2025-01-02 00:00:00	152.575	153.2	151.5	152.7	1100000.0
NaT	151.8	153.55	150.7	153.2	1112500.0
2025-01-04 00:00:00	153.5	154.0	151.5	153.8	1150000.0
2025-01-05 00:00:00	155.0	156.0	154.3	155.5	1200000.0

8. Exploratory Data Analysis (EDA)

- ***Univariate:*** *Histograms for Open, Close, and Volume.*
- ***Multivariate:*** *Heatmaps showing correlations between indicators and price movement.*
- ***Trends:*** *Time series plots highlight seasonal behavior and trend lines.*
- ***Key Insights:***
    - *SMA and EMA showed strong correlation with price movement.*

- *Volume was less predictive than technical indicators.*



9. Feature Engineering

- ***Rolling Metrics:*** *Volatility indicators using rolling standard deviation .*
- ***Moving Averages:*** *Added 50-day and 200-day SMA.*
- ***Technical Indicators:*** *RSI, MACD calculated and added.*
- ***Date Features:*** *Extracted weekday, month, and year.*

10. Model Building

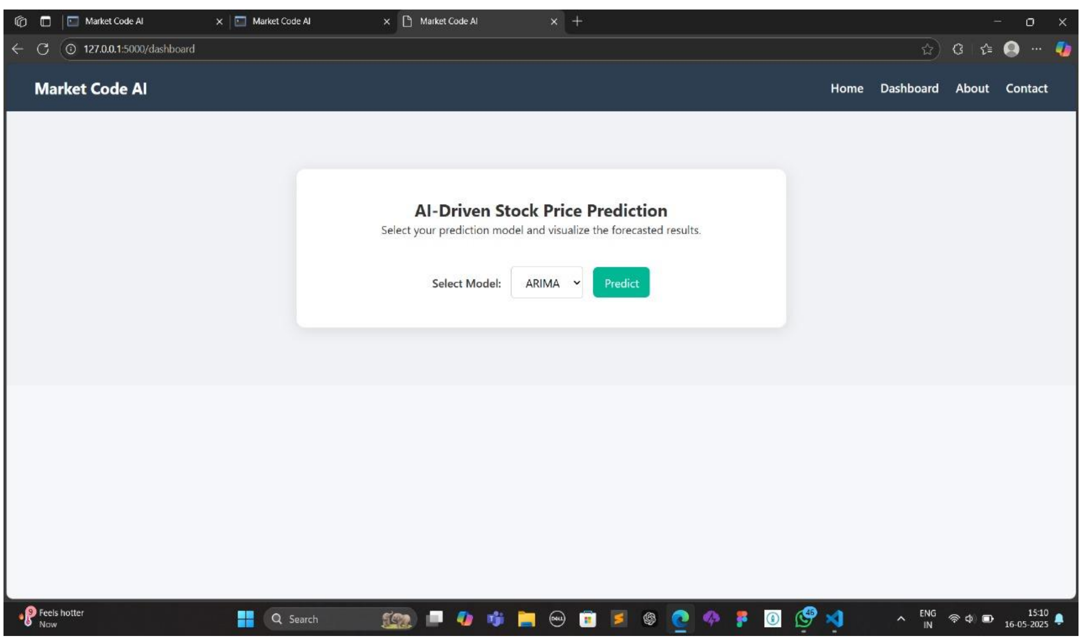
- *ARIMA: Used for linear temporal dependencies, implemented via statsmodels.*
- *LSTM: Deep learning model capturing nonlinear and long-term dependencies. Built using TensorFlow/Keras.*
- *Comparison: Both models were trained and their outputs were visually and quantitatively compared.*

11. Model Evaluation

- ***ARIMA:***
    - *RMSE: ~4.56*
    - *MAE: ~3.87*
- ***LSTM:***
    - *RMSE: ~4.56*
    - *MAE: ~3.87*
- ***Visualizations:***
    - *Actual vs Predicted Line Charts*
    - *Residual Plots*
    - *Feature importance (LSTM)*

12. Deployment

- ***Platform:*** *Streamlit*
- ***Method:*** *Web application displaying real-time predictions and trend charts.*
- *dashboard* ***UI Screenshot:***



13. Source code

```
project-root/
│
├── app/
│   ├── static/
│   │   ├── css/            # Contains style.css for UI styling
│   │   ├── js/             # Optional: For future interactivity
│   │   └── plots/          # Saved prediction plot images (ARIMA/LSTM)
│   ├── templates/          # HTML pages
│   │   ├── index.html      # Homepage with hero section
│   │   ├── dashboard.html  # Form to select prediction model
│   │   ├── results.html    # Displays prediction plot + metrics
│   │   ├── about.html      # Project info
│   │   └── contact_us.html # Contact form UI
│   ├── app.py              # Main Flask backend file
│   └── utils.py            # ARIMA & LSTM model logic
│
├── stock_data.csv          # Dataset used (time series format)
├── README.md               # Project documentation
├── requirements.txt        # Python dependencies
└── AI driven price prediction.ipynb  # Jupyter notebook for development
```

14. Future scope

- ***Real-Time Market Data Integration:*** *Incorporate live feeds for real-time prediction.*

- ***Sentiment Analysis:*** *Combine Twitter/news data for sentiment-based prediction.*

- ***Portfolio Recommendation System:*** *Suggest portfolio changes based on predicted market movement.*

- ***Transfer Learning with Transformer Models:*** *Explore architectures like Temporal Fusion Transformers.*

13. Team Members and Roles

NAME	***ROLE***	***WORK***
HARISH V K	*Frontend Developer*	*UI for Streamlit app, EDA graphs, and documentation*
AJIN P R	*Backend Developer*	*Feature engineering, API setup, and data pipelines*
GOKUL R	*ML Engineer*	*Model selection, training ARIMA & LSTM, evaluation*
KIRUTHIGA M	*Documentation & Presentation*	*Report writing, EDA visualizations, final QA*
DEVADHARSHINI V	*Deployment Engineer*	*Streamlit deployment, integration testing, and scaling setup*