

## Phase-1 Submission

**Student Name:** Mohamed Saif

**Register Number:** 712523205041

**Institution:** PPG Institute of Technology

**Department:** B TECH

**Date of Submission:** 04:05:2025

---

### 1.Problem Statement

Stock market forecasting remains a challenging and complex task due to the volatile and non-linear nature of financial markets. Traditional prediction methods often fail to account for temporal dependencies and dynamic patterns. This project aims to develop an AI-driven time series prediction system that:

- Predicts stock prices with high accuracy
- Captures complex market patterns using deep learning models
- Provides interpretable results to support informed investment decisions

### 2.Objectives of the Project

By the end of the project, we aim to:

- Build a stock price prediction model with high accuracy using time series analysis
- Implement LSTM (Long Short-Term Memory) networks for sequential data forecasting
- Analyze market trends and significant influencing features
- Optionally deploy a Streamlit web app for real-time predictions

### 3.Scope of the Project

#### Features Included:

- ✓ Time series analysis of stock data (OHLC and volume)
- ✓ Comparative model evaluation (ARIMA vs. LSTM)
- ✓ Visual dashboards for trend and prediction analysis

#### Limitations:

- ✗ Focused only on select company stocks
- ✗ Historical data only (no live API feeds)
- ✗ Lightweight local deployment

### 4.Data Sources

This project will utilize publicly available datasets such as:

- Yahoo Finance historical stock data (CSV via yfinance API)
- Optional: NSE/BSE datasets for Indian stocks ([Dataset for Stock Market Prediction](#))

These datasets provide historical price, volume, and technical indicators required for training time series models.

## 5.High-Level Methodology

### Data Collection

Download stock data using **yfinance** or from Kaggle/NSE

### Data Cleaning

Handle missing entries, normalize values, convert time to proper format

### Exploratory Data Analysis (EDA)

Plot moving averages, RSI, MACD, and price trends

### Feature Engineering

Generate lag features, rolling means, and technical indicators

### Model Building

- Baseline: ARIMA/Prophet
- Advanced: LSTM/GRU using TensorFlow or PyTorch

### Model Evaluation

Metrics: RMSE, MAE,  $R^2$  score

Train-test split by time sequence

### Visualization & Interpretation

Plot actual vs. predicted prices

Include model explainability via attention scores (if applicable)

## Deployment

Deploy via Streamlit for local web interface

## 6.Tools and Technologies

**Programming Language:** Python 3.8

**IDE/Notebook:** Google Colab

### Libraries Used:

- Data Processing: **pandas, numpy**
- Time Series: **statsmodels, yfinance, fbprophet**
- Deep Learning: **tensorflow, keras**
- Visualization: **matplotlib, seaborn, plotly**

**Deployment:** Streamlit (for web interface)

## 7.Team Members and Roles

### Team Members and Roles

Name	Role	Responsibilities
Mohamed Saif	Project Lead	Model architecture design and LSTM implementation
Diya Angelin S.P.	Data Engineer	Data acquisition, cleaning, feature engineering
Naveen M	Visualization Specialist	EDA charts, trend visualizations, forecasting graphs

Sanjay M	Deployment Engineer	Streamlit app development and optimization
----------	---------------------	--