

# Stock Market Prediction

## 1. Problem Statement

The goal of this project is to analyze and forecast stock market movements using data from the National Stock Exchange (NSE) of India. The project focuses on predicting short-term price direction (Up/Down trend) and forecasting future price values. This is achieved by applying machine learning classification (Random Forest) and time-series forecasting (Facebook Prophet) models.

## 2. Dataset

The dataset used in this project is sourced from NSE India, specifically historical stock data (e.g., State Bank of India 'SBIN'). The raw dataset included columns such as Date, Open, High, Low, Close, Volume, and VWAP. Unnecessary or redundant columns such as 'Series' were ignored. A correlation heatmap was also used to identify and drop unrelated or highly collinear columns, ensuring only meaningful features were kept.

## 3. Feature Engineering

Several technical indicators and derived features were added to capture market patterns:

- Simple Moving Averages (SMA20, SMA50): Trend direction and crossovers.
- RSI (14): Strength of momentum (overbought/oversold signals).
- MACD, MACD Signal, MACD Histogram: Detects bullish or bearish momentum changes.
- Returns & Volatility (5-day): Short-term risk measurement.
- ROC (5) and Relative Strength: Momentum comparison.
- Lag Features (Close\_lag1, Close\_lag2): Past dependency.

Target variables were defined as:

- Binary Trend (Up = 1, Down = 0)
- Multiclass Trend (Bullish, Neutral, Bearish).
- Target Price and Target Return for regression tasks.

Charts were drawn for SMA20, SMA50, RSI(14), and MACD indicators to visualize trend direction, overbought/oversold conditions, and bullish/neutral/bearish momentum. These visualizations help traders clearly understand market signals.

## 4. Methodology

Two models were applied for different objectives:

### 1. Random Forest Classifier:

- Input: Technical indicators and derived features.
- Output: Predicts Up (1) or Down (0) trend for the next day.

- Training: 80% historical data, 20% testing using time-based split.
- Evaluation: Accuracy ~75%, Precision, Recall, and F1-score used for validation.

## 2. Prophet (by Facebook):

- Input: Date (ds), Closing Price (y).
- Output: Forecasts stock prices for the next 30 days.
- Accuracy: The model was evaluated using error metrics:
  - Root Mean Squared Error (RMSE): **30.12**
  - Mean Absolute Error (MAE): **23.22**

- Visualization: Charts included trend, upper bound, and lower bound to evaluate accumulation zones.

## 5. Visual Analysis

From the visuals we have seen, the stock shows a clear long-term downtrend with price trading below the SMA-200 and the short-term averages (SMA-20, SMA-50) also sloping lower. RSI mostly stays under 50 with frequent dips near oversold zones, highlighting persistent weakness. The MACD remains below zero with repeated bearish crossovers, confirming downward momentum. Volume patterns show heavy spikes during declines and only sideways activity in 2025, suggesting distribution rather than fresh accumulation.

Overall, the trend has shifted from bullish (early 2024) to bearish (late 2024) and is now in sideways consolidation, keeping the long-term outlook cautious.

## 6. Results & Evaluation

Random Forest achieved around 75% classification accuracy, effectively capturing SMA, RSI, MACD, and volatility patterns to determine likely short-term trends. This supports decisions about bullish, neutral, or bearish momentum.

Prophet produced reliable 30-day forecasts. The forecast plots displayed upper and lower thresholds, which helped in identifying whether a stock was consolidating (accumulation) or showing strong directional momentum. Adding volume to Prophet charts further enhanced the ability to detect accumulation phases.

## 7. Conclusion

This project demonstrates how combining correlation analysis, technical indicators, and machine learning with time-series forecasting can provide actionable insights for traders and investors. Random Forest assists in daily buy/sell decisions using trend signals, while Prophet provides broader forecasts with uncertainty intervals. Together, they offer a comprehensive toolkit for understanding market momentum and managing risk.