Project Title:
Predictive Modeling and Forecasting of Stock Prices Using Machine Learning

Submitted by:
Madhu Priyaa S -910622104058
Nivetha R S   -910622104068
Department of Computer Science and Engineering
K.L.N. College of Engineering.

Abstract

Stock market forecasting has always been a challenging task due to the unpredictable, non-linear, and volatile nature of financial data. Traditional models like ARIMA and moving averages often fail to adapt to real-time fluctuations and are limited to single-variable analysis. Even advanced deep learning models like LSTM, while effective in sequence learning, are resource-heavy, slow to train, and often act as black boxes — difficult to interpret or deploy in real-time systems. To overcome these challenges, this project proposes a **new machine learning-based forecasting model using XGBoost (Extreme Gradient Boosting)** — a highly efficient and interpretable algorithm capable of modeling complex relationships in historical stock data. Unlike existing approaches, this model emphasizes speed, clarity, and multi-feature flexibility, making it ideal for real-world financial prediction.

The system introduces **a new structure of feature engineering** where raw stock data (Open, High, Low, Close, Volume) collected using the **yfinance API** is transformed into a rich predictive dataset. Features like **lag values (past prices), percentage returns, moving averages (MA5, MA10), and volatility patterns** are extracted to give the model a deeper understanding of stock behavior. The XGBoost algorithm is then trained on this enriched dataset to forecast the **next-day closing price** with high precision. Performance is measured using **MAE (average error size), RMSE (penalizing larger errors), and R^2 Score (explaining how well the model predicts actual prices)**. What makes this system different is not just the accuracy — but its **plug-and-play architecture**, which allows it to work with **any stock symbol**, handle **multivariate data**, and be **updated in real time** without retraining from scratch.

The true innovation of this project lies in its vision beyond academic prediction — it is a **real-world-ready, intelligent forecasting engine**. This system can be directly connected to live stock feeds, integrated into financial dashboards or advisory platforms, and scaled for multiple users or stock categories without manual intervention. It removes the mystery from machine learning by offering clear feature importance, transparency, and faster processing than LSTM models. It also lays the groundwork for future enhancements such as **sentiment analysis using financial news**, **integration of macroeconomic indicators**, and **alert systems for high-risk market movements**. In short, this project contributes not just a new model, but a **new way of thinking** about how machine learning can be applied to finance — combining accuracy, explainability, and real-time usability in one unified system.