

“Empowerment through quality technical education”

**AJEENKYA DY PATIL SCHOOL OF ENGINEERING**

Dr. D. Y. Patil Knowledge City, Charholi Bk., Via. Lohegaon, Pune – 412 105.

Department of Artificial Intelligence & Data Science Engineering

Case Study Report

**Academic Year: 2024-25 Name: Nikhil Malvi Div:B-50**

**Topic:- The XGBoost Journey – Predicting Stock Price Direction Using Technical Indicators.**

# Introduction

In today’s fast-paced financial markets, making informed and timely trading decisions is a critical challenge faced by investors and analysts. While numerous strategies exist, technical analysis remains a popular method that relies on historical price and volume patterns to forecast future movements. However, the traditional application of technical indicators is often manual, heuristic-based, and prone to subjective interpretation.

With the advancement of machine learning and access to structured financial data, it has become possible to automate and optimize the use of technical indicators through data-driven models. Among various algorithms available, **XGBoost** has emerged as a powerful tool for classification and regression tasks due to its robustness, speed, and performance.This case study explores the application of XGBoost in predicting the **short-term direction of stock prices** using engineered technical indicators. Specifically, we focus on the stock of Apple Inc. (AAPL), using historical OHLCV (Open, High, Low, Close, Volume) data and a set of well-established technical indicators including Moving Averages, RSI, MACD, Bollinger Bands, and On-Balance Volume.

# Case Description

This case study focuses on the development of a machine learning model to predict short-term stock price direction using technical analysis features. The asset chosen for this study is **Apple Inc. (AAPL)**, one of the most actively traded and widely analyzed stocks in the global financial markets.

The core problem is framed as a **binary classification task**: determining whether the **closing price of the stock will go up or down the following trading day**. The motivation behind this approach is to simulate a real-world use case where traders and algorithmic systems rely on directional signals to inform buy or sell decisions.

To support this objective, a dataset containing **daily historical OHLCV (Open, High, Low, Close, Volume)** data was collected over multiple years. From this base data, a range of **technical indicators** was computed, including:

Simple Moving Averages (SMA)

Exponential Moving Averages (EMA)

Relative Strength Index (RSI)

Moving Average Convergence Divergence (MACD and Signal Line)

Bollinger Bands (Upper, Middle, Lower)

On-Balance Volume (OBV)

Daily Return Percentage

Each of these indicators is designed to capture specific aspects of stock behavior such as momentum, volatility, and trend direction. A new column, **Target**, was created to represent the outcome: 1 if the next day’s closing price is higher than the current day’s, and 0 otherwise.

The dataset was preprocessed to handle missing values caused by rolling calculations, and the data was split into **training and testing sets** using a time-ordered 80/20 split. The model chosen for classification was **XGBoost**, due to its efficiency and strong performance in handling structured tabular data.The overall goal was to test the hypothesis that traditional technical indicators, when used in combination and evaluated with a machine learning model, can yield actionable predictions about short-term stock price movement.

# Analysis

The analysis phase of the project involved transforming raw historical stock data into a structured format suitable for machine learning and evaluating the effectiveness of technical indicators in predicting stock price direction. The dataset consisted of daily OHLCV (Open, High, Low, Close, Volume) values for Apple Inc. (AAPL), which served as the foundation for feature extraction and modeling.

*3.1 Feature Engineering*

A total of **10+ technical indicators** were engineered from the original dataset to represent price trends, momentum, and volatility. These included:

**SMA and EMA**: Captured short-term and long-term trend directions.

**RSI (14-day)**: Measured overbought or oversold conditions.

**MACD & Signal Line**: Highlighted momentum shifts and trend reversals.

**Bollinger Bands**: Indicated price volatility and potential breakout levels.

**OBV (On-Balance Volume)**: Incorporated volume as a measure of buying/selling pressure.

**Daily Return**: Reflected the immediate percentage change in closing price.

These indicators were selected due to their widespread use in technical analysis and their ability to capture diverse market behaviors.

*3.2 Target Variable Definition*

A binary target column named **Target** was introduced to represent the next day's price movement:

1 if the next day’s **Close** > current day’s **Close**\

This conversion of a regression-style problem into a **binary classification** task allowed for simplified prediction and evaluation of directional accuracy.

*3.3 Data Cleaning and Preparation*

Rolling indicators introduced **NaN values**, which were removed.

Data was split chronologically into **80% training** and **20% testing** sets.

Only past data was used for predictions to **prevent look-ahead bias**.

*3.4 Model Selection and Training*

The **XGBoost classifier** was chosen due to its superior performance in classification tasks involving tabular data. It is particularly effective for handling non-linear relationships and avoiding overfitting through built-in regularization.

The model was trained on the engineered features and validated using standard classification metrics. Hyperparameter tuning using RandomizedSearchCV was optionally applied to improve performance.

*3.5 Evaluation*

The model's accuracy on the test set reached approximately **58.25%**, outperforming random guessing (50%) and demonstrating that technical indicators do contain informative signals when evaluated collectively. Precision and recall values showed that the model was particularly capable of identifying upward movements with reasonable consistency.

# Findings and Discussion

The objective of this study was to evaluate whether a machine learning model—specifically **XGBoost**—could effectively predict the **direction of stock price movement** using technical indicators derived from historical data. The following key findings were observed as a result of the analysis and model evaluation.

*4.1 Predictive Performance*

The XGBoost classifier achieved an overall **accuracy of approximately 58.25%** on the unseen test data. While this value may appear modest in isolation, it is significant in the context of financial forecasting, where random guessing would yield a baseline accuracy of 50%. This result suggests that the model was able to extract useful patterns from the technical indicators, particularly in identifying days with strong directional movement.

Additionally, the classification report revealed that:

**Precision** for predicting upward movement was slightly higher than that for downward movement.

**Recall** showed a balanced ability to correctly classify both classes, though slightly favoring bullish (upward) predictions.

**F1-Score** reflected consistent performance without significant bias toward either class.

*4.2 Feature Importance*

XGBoost’s built-in feature importance analysis showed that indicators such as **MACD**, **EMA values**, and **OBV** were among the top contributors to model performance. These features are known to reflect both momentum and volume-related trends, aligning with established trading strategies.

This finding supports the idea that machine learning models can validate and quantify the practical utility of technical indicators, beyond their traditional heuristic use.

*4.3 Model Behavior*

The model was observed to perform better in conditions where the market displayed clear trends—either upward or downward. It showed relatively lower accuracy in periods of low volatility or sideways price movement, which is expected due to the lack of strong directional signals in such conditions.

Filtering out flat-return days improved the model’s ability to generalize and removed noise that could have reduced performance.

*4.4 Practical Implications*

These findings suggest that even simple technical indicators, when used collectively and processed through a capable machine learning algorithm, can provide a reasonable degree of directional forecasting. This approach, while not a standalone trading strategy, can be integrated into larger decision-making frameworks or used as a signal enhancement method in algorithmic trading systems.

# Conclusion

This case study demonstrated the application of machine learning—specifically the **XGBoost classifier**—to predict short-term stock price direction using engineered technical indicators. By combining classical financial tools such as Moving Averages, RSI, MACD, Bollinger Bands, and OBV with a data-driven modeling approach, the project aimed to bring objectivity and automation to technical analysis.The model achieved a prediction accuracy of approximately **58.25%**, which, while moderate, is a meaningful improvement over random guessing in the context of financial markets. Feature importance analysis confirmed that indicators grounded in momentum and volume trends contributed most significantly to model performance.

The approach proved effective in capturing price direction during strong market trends but had limitations in low-volatility or sideways market conditions. Nonetheless, the results validate that traditional indicators, when processed collectively and systematically, hold predictive value and can enhance decision-making in short-term trading scenarios.In conclusion, this work lays a solid foundation for further exploration into **machine learning-based financial forecasting** and highlights the potential of integrating technical indicators with modern predictive models in the field of quantitative finance.

# References

[1] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, “Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques,” *Expert Systems with Applications*, vol. 42, no. 1, pp. 259–268, Jan. 2015. doi: 10.1016/j.eswa.2014.07.040

[2] M. Ballings, W. Van den Poel, N. Hespeels, and R. Gryp, “Evaluating multiple classifiers for stock price direction prediction,” *Expert Systems with Applications*, vol. 42, no. 20, pp. 7046–7056, Nov. 2015. doi: 10.1016/j.eswa.2015.05.013

[3] R. F. Engle, “Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation,” *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[4] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in *Proc. 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pp. 785–794, 2016. doi: 10.1145/2939672.2939785

[5] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[6] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” *International Journal of Forecasting*, vol. 14, no. 1, pp. 35–62, 1998.

[7] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.