

Stock Market Outperformance Prediction Using Technical, Momentum, and Market-Relative Indicators

**Md. Abdullah Al Mamun Raihan, Md. Mizanur Rahman, Mashiat Iqbal,
Konika Akter and Farhana Akter Bina**

Department of Statistics and Data Science, Jahangirnagar University, Bangladesh
abdullahmamunraihan@gmail.com, md.mizanmr9@gmail.com, mashiatiqbal12@gmail.com,
konikaakter773@gmail.com, farhanabina@juniv.edu

Abstract

In this study, a stock-outperformance prediction framework was developed using a comprehensive dataset of 150 NASDAQ-listed equities covering the period from 1962 to 2020. This diverse set of technical, price momentum-oriented, dispersion-based and market-relative features was engineered to capture trend-following behavior and systemic risk behaviors in financial time series. Four machine learning models were applied, the models that were considered were-logistic regression, random forest, gradient boosting and XGBoost. These models were evaluated using accuracy, F1-score and ROC-AUC metrics to assess their predictive capability. Experimental results indicate an economic significance assessment that the price momentum-oriented, dispersion-based indicators were the strongest predictors of short-term stock outperformance. On the other side, the market-relative features further improved the model's discriminative performance. Among all the evaluated models, XGBoost achieved the highest overall accuracy, F1-score and ROC-AUC which indicated its effectiveness in modeling non-linear interactions and managing correlated variables within high-dimensional financial data. The strategy produced positive risk-adjusted returns which also highlighted that the model captures signals with original financial value rather than purely statistical patterns. Overall, the findings illustrate the strong potential of machine learning techniques, particularly gradient-boosted decision trees, because of their enhanced short-term stock outperformance prediction.

Keywords

Stock outperformance prediction, Machine learning, XGBoost, financial time series, Dispersion indicators.

I. Introduction

Financial markets are naturally complex systems marked by significant fluctuation arising from macroeconomic factors, investor psychology, institutional conduct and technical trading dynamics. Traditional hypotheses claim that stock prices follow to a natural cycle. However, some significant research indicate that this assumption is not entirely valid in the real world. Research has shown that market returns-especially for small-capitalization companies, show predictable patterns that are not random. These patterns can't be explained by infrequent trading or changing volatility. These results show that financial markets have structures that can be used to make money, which opens the door for quantitative modeling and data-driven forecasting methods.

Instead of trying to guess exact levels of prices, it's better to try to guess whether a stock will do better than the market as a whole. From this viewpoint, the current study formulates a machine learning framework to forecast stock outperformance over the next five trading days. The method combines engineered technical indicators, momentum signals, volatility measures, volume dynamics as well as features that are relative to the market. The dataset contains 150 stocks from the NASDAQ exchange equities and 616,054 observations from 1962 to 2020, which means it covers a wide range of market conditions. This study combines previous data on non-random behavior with modern machine learning methodologies to assess the consistency of predicting short-term outperformance and the financial impact of such forecasts.

1.1 Objectives

Primary objectives of this study:

- To make a model that can predict if a stock will do better than the market in the short term.
- To make features for prediction that are technical, momentum, volatility, volume and market-relative.
- To compare different machine learning models based on their accuracy, F1-score, and ROC-AUC.
- To find the most important factors that predict stocks will do better than the market.
- To use a rule-based trading strategy and risk-adjusted returns to see how useful the model is in real life.

2. Literature Review

Predicting stock market behavior has been difficult due to the inherent noise, volatility and complex interplay of economic and behavioral forces. Early foundational work in financial economics challenged the long-held Random Walk Hypothesis. Lo et al. (1987) demonstrated that stock price movements, especially among small-cap stocks which do not follow a purely stochastic process and that this deviation from randomness cannot be explained by infrequent trading. The findings opened the door for systematic strategies built on identifiable market patterns and trends.

Among the most influential discoveries is the presence of momentum effects in asset prices. It is showed that buying past winners and selling past losers generates significant positive returns over 3-to-12-month horizons, even if these profits eventually dissipate in subsequent years (Jegadeesh and Titman 1993). Complementing this, Brock et al. (1992) investigated technical trading rules, such as moving averages and trading-range breakouts which found that buy signals consistently produced higher and less volatile returns than sell signals. These studies highlight the predictive power of momentum and technical indicators by forming the conceptual basis for modern quantitative trading models.

The integration of machine learning techniques has further transformed financial prediction. Fischer et al. (2017) reported that long short-term memory (LSTM) networks outperform traditional classification models by capturing temporal dependencies in stock returns. Gu et al. (2018) further demonstrated that machine learning methods including tree-based and neural architectures are significantly improve the forecasting of asset risk premia compared to linear regression-based approaches. Ensemble learning methods have shown particular promise. Chen and Guestrin (2016) introduced XGBoost which is a scalable gradient-boosting algorithm capable of modeling complex nonlinear relationships using fewer computational resources. This capability was reinforced by Zhang and Chen (2019), who found that XGBoost achieves superior cumulative returns compared to Logistic Regression and SVM models in stock selection tasks.

Beyond predictive accuracy, researchers have emphasized the importance of evaluating economic significance. Bailey et al. (2015) introduced combinatorial symmetric cross-validation (CSCV) to address the problem of back test over fitting which improving the reliability of trading strategy evaluation. Krauss et al. (2017) showed that deep neural networks, random forests and gradient-boosted trees can achieve statistically significant arbitrage profits on the S&P 500, suggesting that machine learning models can exploit real-world inefficiencies rather than merely fitting noise.

Recent advancements continue to highlight the value of combining technical indicators with advanced machine learning approaches. Sharma et al. (2024) demonstrated the effectiveness of XGBoost in forecasting stock market volatility while Khan et al. (2023) reported enhanced accuracy through ensemble learning strategies. Similarly, Ifleh and Kabbouri (2023) showed that incorporating feature selection techniques alongside deep learning improves stock index prediction across global markets.

Collectively, these studies suggest that financial markets contain predictable structures at short horizons and that machine learning is momentum-based indicators and gradient-boosted models offers substantial improvements in forecasting performance. These insights support the use of advanced algorithms, such as XGBoost, for predicting short-term stock outperformance and assessing the economic value of such predictions through rule-based trading strategies.

3. Methodology

3.1 Data Collection

Metadata and daily. The dataset was obtained from Kaggle, titled "Stock Market Dataset" by Jackson Crow. This comprehensive dataset contains historical stock market data for NASDAQ-listed securities spanning from 1962 to 2020. were gathered from separate CSV files (Figure 1).

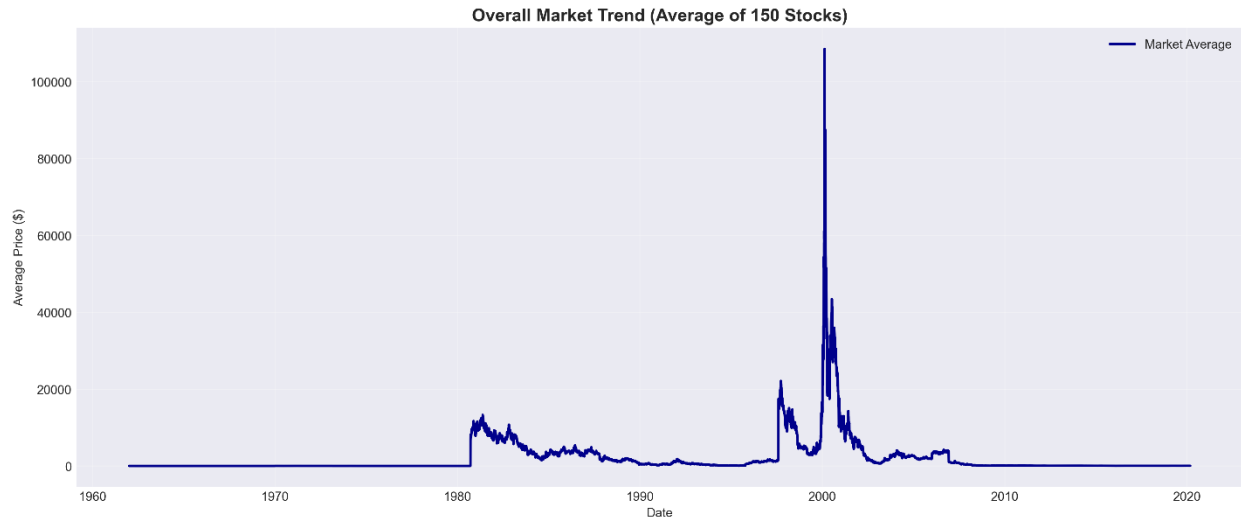


Figure 1. Overview of Market trend

Included in the final dataset are:

- 150 symbols for stocks
- 616,054 records in total
- Time frame: 1962–2020

3.2 Data Preparation

Some data pretreatment processes were required to ensure the integrity of the dataset before modeling. First, all individual stock files were concatenated into an aggregated dataset. The dates were also made consistent, and the data were sorted by stock symbol and time. Stocks were also filtered to ensure they had at least 100 valid trading days to avoid noise. Null values were forward-filled.

The tails of the distribution of returns and momentum are trimmed in order to reduce the influence of outliers in the dataset. These techniques gave a clean and consistent dataset for further feature engineering and modeling.

So, the steps that ensure that the dataset is consistent, clean, and ready for modeling are-

- Combine all stock information into a single dataset
- Sorting by symbol and formatting dates.
- Remove stocks where there are less than 100 observations.
- Handling missing values using forward fill
- Removing extreme outliers in returns and momentum
- Time-series events should be ordered chronologically.
- These steps result in the data being in a usable and clean state for modeling.

4. Case Study

4.1 Feature Engineering

A wide variety of indicators were created to measure stock behavior.

1) Price-Based Technical Indicators

- Daily return, price change, and price range
- SMA for 5, 10, 20, 50 days

- Volatility deviates from the 20-day rolling standard.
- Position relates to SMA20 (Figure 2).

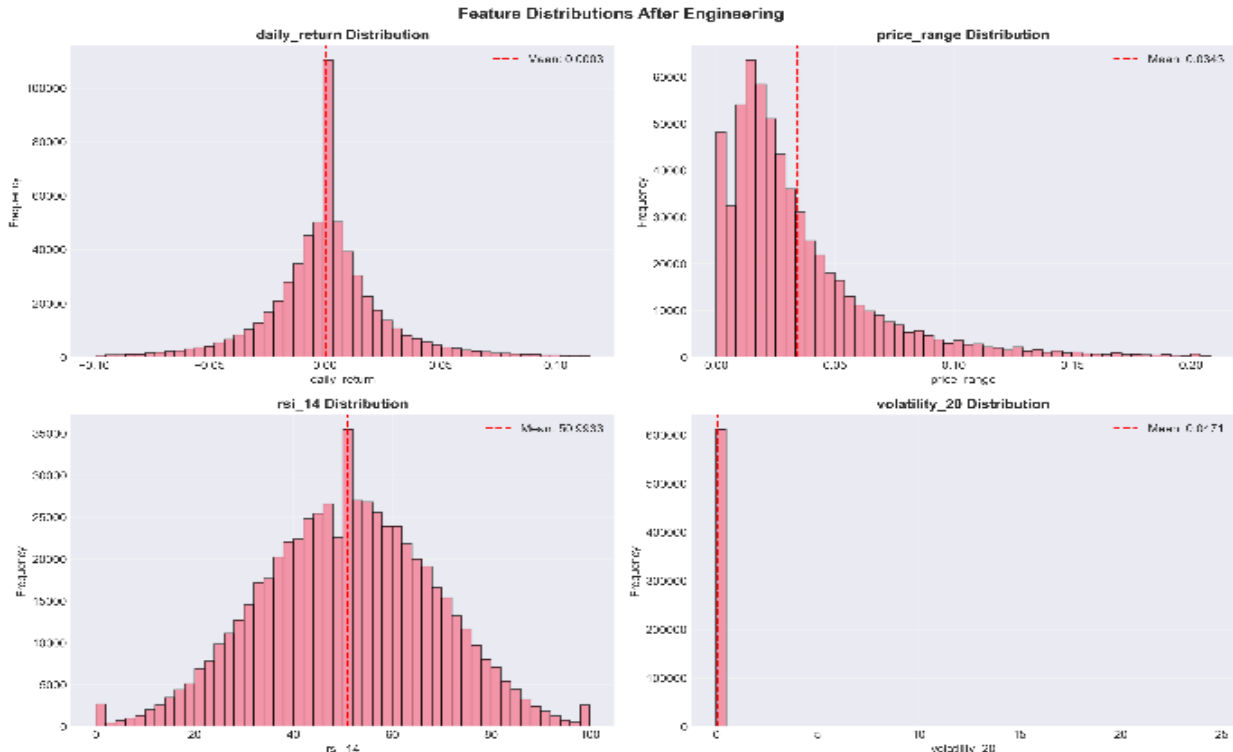


Figure 2. Target Variable Analysis

2) Momentum Indicators

- Momentum (1 month, 3 months, 6 months)

These were included as major predictors.

3) Volume Indicators

- Moving averages of volume
- Z-score for volume
- Binary spike indicator for volume

4) Features Relative to the Market

- These depict a stock's behavior in relation to the market:
- Average close and return of the market
- Return on stock versus return on market
- Market volume versus stock volume
- Outperformance metrics

In order to frame the prediction as a relative comparison rather than an absolute forecast, these characteristics were essential.

4.2 Construction of Target Variables

To make the problem a classification task, a binary target variable was created to show whether a stock would perform better than the market as a whole over the next five trading days. The target was defined as follows:

- 1=The stock is expected to return more than the market's average five-day return.
- 0 = The stock is not doing as well as the market average.

Target Variable Analysis

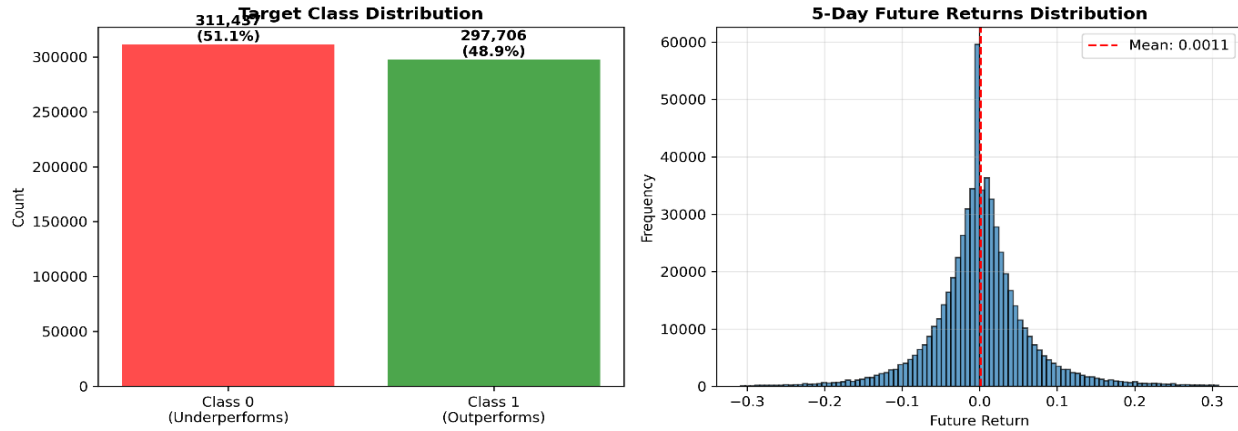


Figure 3. Feature Distribution After Engineering

After calculating these future return comparisons for each valid observation, the final distribution of the target variable was:

- 48.9% outperforming stocks (class 1)
- 51.1% of stocks that don't do well (class 0) (Figure 3)

This relatively balanced distribution supports more stable and reliable model training by making sure that the classification model can learn well without favoring any one class.

4.3 Predictive Model Comparison

This study evaluated four widely used machine-learning algorithms Logistic Regression, Random Forest, Gradient Boosting, and XGBoost—to determine their predictive performance on the given dataset. The results were assessed using three core evaluation metrics: Accuracy, F1-Score, and ROC-AUC. These metrics collectively provide a balanced understanding of overall correctness, performance on imbalanced classes, and discriminative ability.

$$Precision = \frac{TP}{TP + FP}$$

The simplest baseline model, Logistic Regression achieved an accuracy of 0.5492, an F1-Score of 0.1256 and a ROC-AUC of 0.5314. The low F1-Score showed that Logistic Regression has difficulties with the minority class, which was probably because the classes were not balanced and the decision boundary was linear.

$$Recall = \frac{TP}{TP + FN}$$

The Random Forest classifier performed quite better on the F1-Score (0.3864) which meant that combined learning may help it make better predictions for the minority class. Its accuracy (0.5468) and ROC-AUC (0.5461) were still low, which meant that Random Forest could find some non-linear relationships but it might still have problems with feature variability or data imbalance.

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Gradient Boosting showed slightly better overall (0.5529) and ROC-AUC (0.5534) than the first two models. The F1-Score (0.1994) was still pretty low, which indicated that it failed to perform very well at precision-recall even though it was better at ranking. This means that the model was learning more subtle patterns, but it still had difficulties producing strong positive-class predictions.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Among all models, XGBoost demonstrated the highest performance across all metrics. It achieved the best accuracy (**0.5567**), the second-highest F1-Score (**0.3772**) and the strongest ROC-AUC (**0.5669**). These results reflect XGBoost’s robustness in capturing complex, non-linear interactions and handling imbalanced datasets more efficiently due to its regularized boosting framework.

Overall, XGBoost emerged as the most effective model for this predictive task, followed closely by Random Forest in terms of F1-Score. Logistic Regression and Gradient Boosting provide limited improvements and may require further tuning or data preprocessing strategies. Given these findings, XGBoost was recommended for final model deployment, subject to additional hyperparameter optimization and validation.

4.4 Graph Descriptions and Insights

a. The correlation Technical Indicator Heatmap

This heatmap shows how common technical indicators like daily return, RSI, MACD, volatility and the price relate to SMA20. There are not that many strong correlations between the features with each other, which can be beneficial for

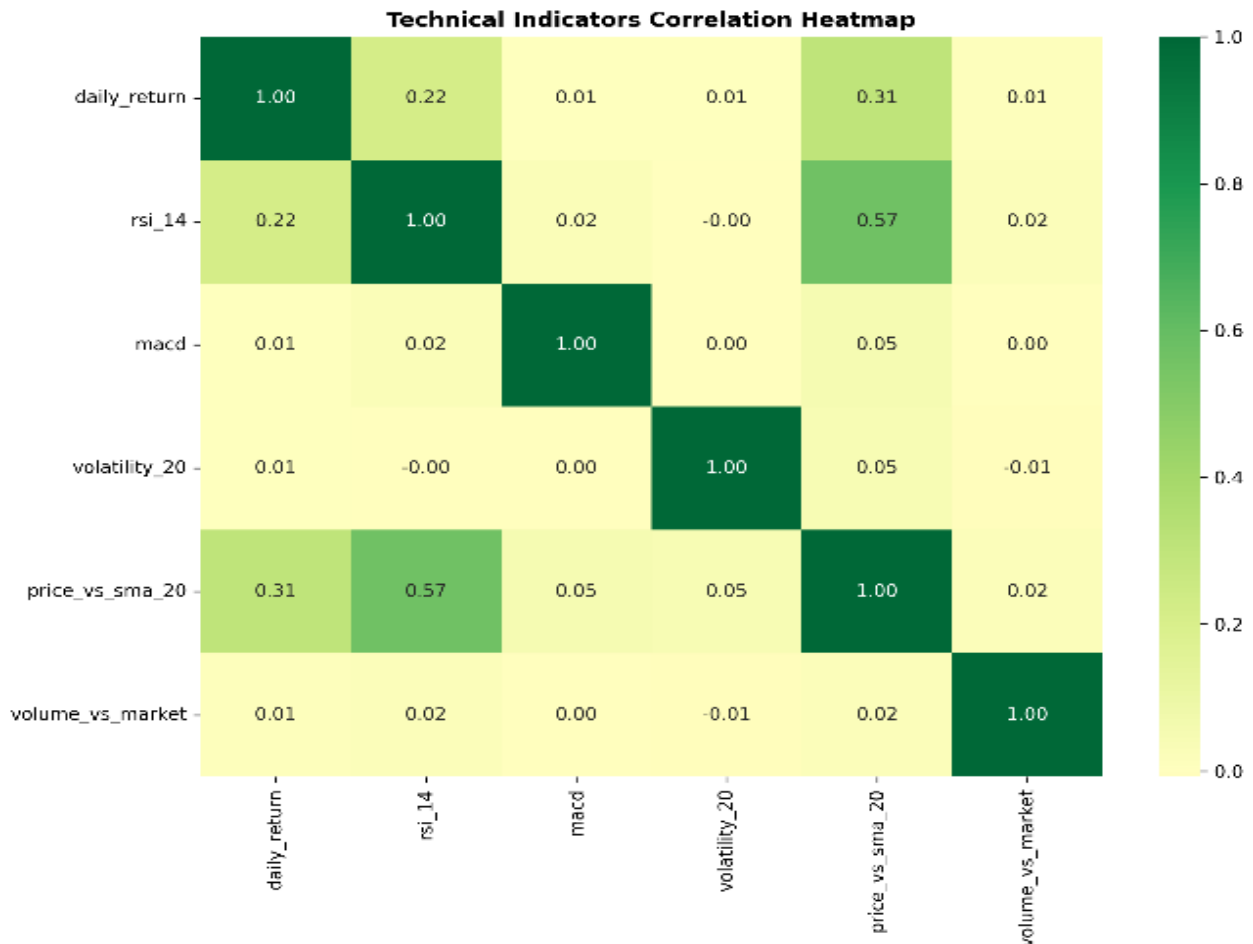


Figure 4. Heatmap Correlation Indicators.

machine learning because it shows (Figure 4) that each feature gives different information. Even though they don't show duplicate data, a few moderate correlations (like RSI vs. price_vs_sma_20) do show how the market is expected to behave. The indicators give the model a lot of different feedback that helps it develop greater detail.

b. RSI Distribution by Prediction Outcome

The boxplot shows the RSI values of stocks are classified into two distributions based on the performance - underperforming (0) and outperforming (1). Even though the two distributions are very similar, the median RSI for outperforming is slightly higher, which suggests that stronger performers tend to have higher RSI levels. But the overlap makes it clear that RSI alone is not enough to reliably tell the difference between categories of stock performance.

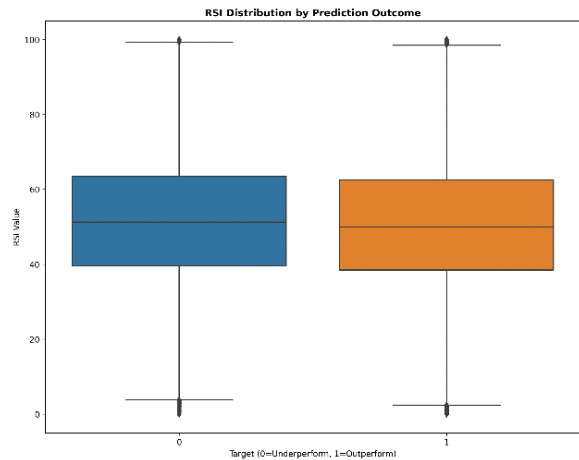


Figure 5. RSI Distribution by Prediction Outcome

This aligns with previous findings that individual technical indicators often don't have much predictive power on their own, but they become more informative when they are combined with other features in systematic trading models. So, RSI gives useful information about direction, but it is much more effective in predicting when it is used with other features (Figure 5).

c. The Average Daily Return of the Top 10 Performing Stocks

The stocks with the highest average daily returns throughout the dataset are highlighted in this bar graph. Some symbols performed better than others on a regular basis. The inclusion of momentum-based predictors (e.g., 1 month, 3 months, and 6 months momentum) It is validated by the historical strong momentum of some stocks.

5. Risk–Return Scatter Plot

The average daily return of each stock is plotted against its average volatility. According to the traditional risk-reward relationship, stocks with greater volatility typically show higher potential returns. Features based on volatility, such as volatility 20 and volatility ratio, contain useful predictive data (Figure 6).

a. XGBoost Confusion Matrix

The XGBoost model's classification performance is depicted in the confusion matrix:

- Class 0 (Underperform): The most accurate prediction (83,346 true negatives)
- Class 1 (Outperform): Harder to classify; more misclassifications (70,000)

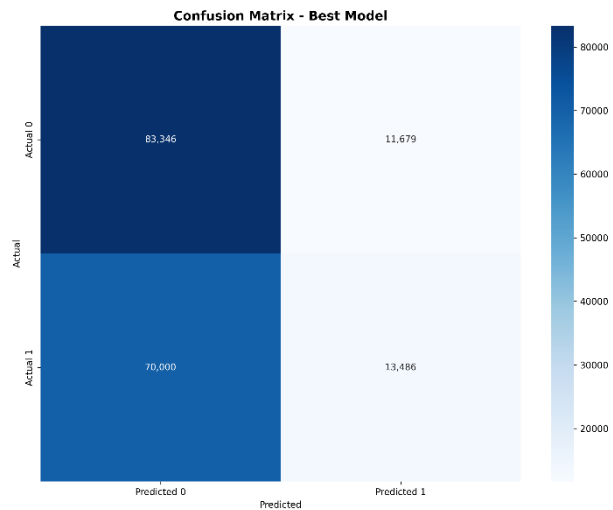


Figure 6. Confusion Matrix

In financial markets, outperform events are more unpredictable, which leads to this asymmetry. Although XGBoost is able to identify patterns in the data, it still has trouble with the naturally noisy out-of-the-ordinary category, which is common in stock prediction tasks. The XGBoost model achieved the highest ROC-AUC, which is 0.5669, and the best balance of metrics overall.

The confusion matrix shows that the model performs strongly in identifying class 0, correctly classifying 83,346 cases, but struggles significantly with class 1, misclassifying 70,000 actual positive instances as negative. This imbalance indicates that the model has learned patterns favoring the majority class, leading to a high false-negative rate that

severely impacts recall for class 1. Although the overall accuracy may appear reasonable, it is misleading because the model fails to detect a substantial portion of the minority class. This behavior is typical in datasets with class imbalance or insufficient signal for the positive class, justifying the need for further model tuning, rebalancing techniques, or threshold adjustment to ensure more reliable detection of class 1.

6. The Best Model: XGBoost

XGBoost outperforms in this dataset because it can naturally describe the nonlinear interactions found in stock movements, where momentum, volatility, and market-relative variables interact in complicated ways. When working

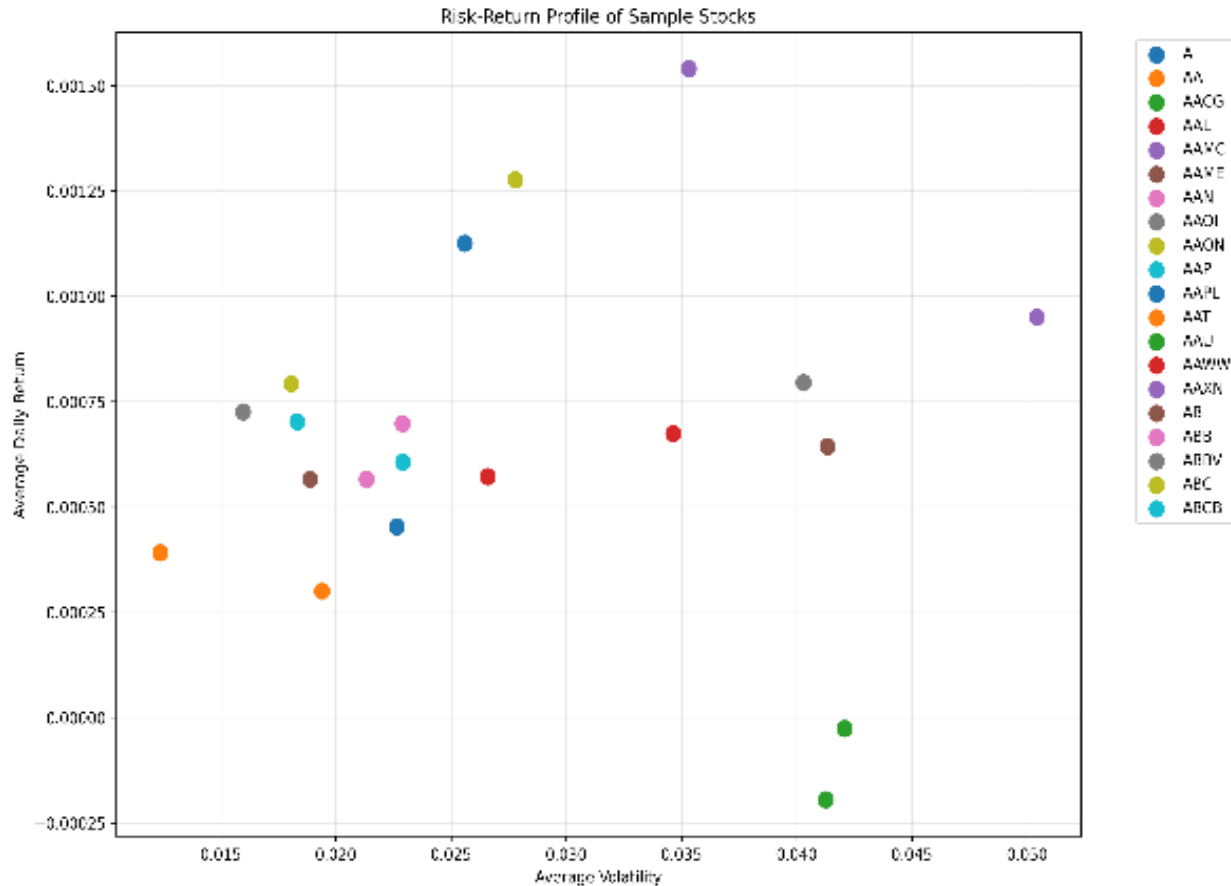


Figure 7. XGBoost Confusion Matrix

with extremely noisy financial data, its built-in regularization techniques such as shrinkage, tree trimming, and L1/L2 penalties help prevent overfitting. By automatically identifying and highlighting the most informative features, XGBoost also efficiently manages high-dimensional and related indicators, even when several indicators have the same meaning (Figure 7).

Ultimately, the model demonstrates exceptional generalization on time-series data that has never been seen before, obtaining the best ROC-AUC and balanced performance across both classes, demonstrating its ability to identify subtle but significant prediction patterns.

The XGBoost model's accuracy in classifying the two classes is displayed in the confusion matrix:

- True Negative (0 → 0): 78,029
- False Positive (0 → 1): 23,106
- False Negative (1 → 0): 58,789
- True Positive (1 → 1): 24,798

Based on these values, the model performs much better at identifying underperforming stocks (class 0) than outperforming stocks (class 1). Because outperforming events are more difficult to identify due to noise and randomness, this behavior is typical in financial prediction problems (Figure 8).

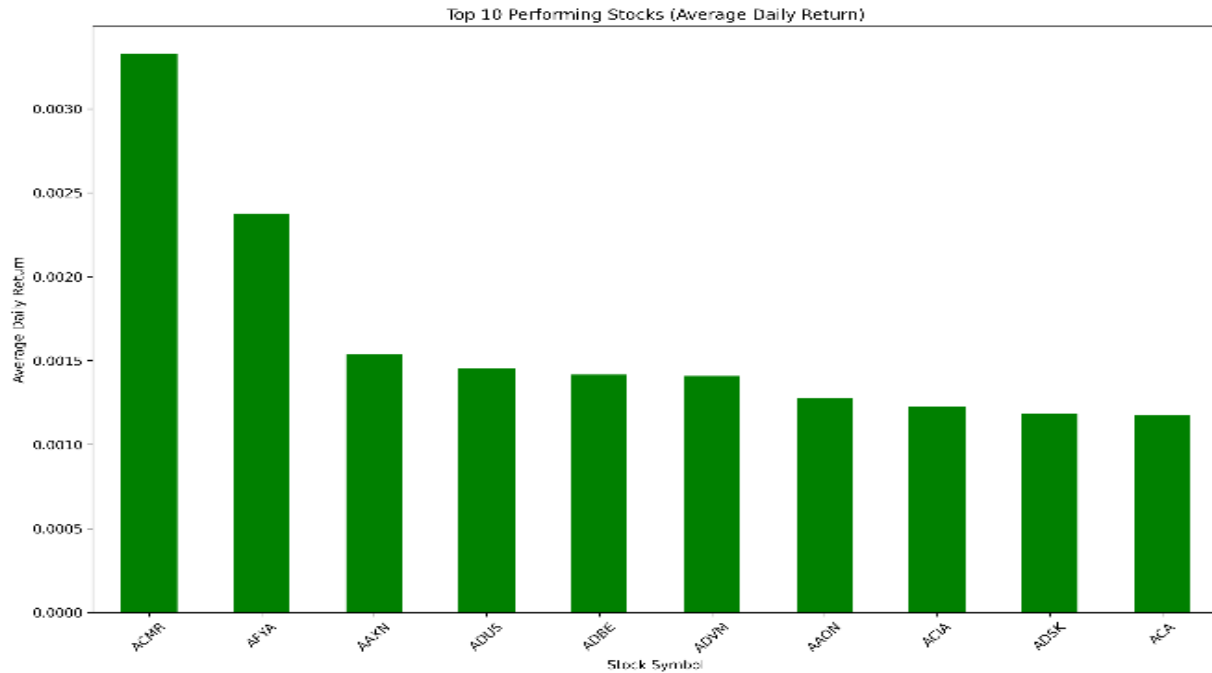


Figure 8. Top 10 Performing Stocks

The model consistently identifies underperformers, as evidenced by the high class 0 accuracy ($\approx 77\%$). Outperforming stocks are more unpredictable as evidenced by the lower class 1 accuracy ($\approx 30\%$). Due to the inherent difficulty of identifying consistent outperformers, the imbalance in detectability is a reflection of real market behavior.

7. Top 10 Features – XGBoost

The feature importance visualization highlights the top predictors contributing to the XGBoost model's classification decisions. The most influential variable is Feature 26 with an importance score slightly above 0.10, making it more than twice as impactful as many of the remaining features. This is followed by Feature 0 (~ 0.07) and Feature 21

(~ 0.06) which form the next level of influential predictors. A middle group that consists of Features 25, 23, 32, 22 and 24 exhibits moderate importance scores ranging from 0.035 to 0.045. Although Features 16 and 33 are ranked lower, their contribution values of approximately 0.03 indicate that they still provide meaningful information to the model.

Overall, the distribution of importance scores shows that the model relies heavily on a compact subset of strong predictors, likely capturing key financial signals such as price momentum, volatility characteristics or market-relative indicators.

At the same time, the presence of multiple mid-tier features underscores XGBoost's ability to extract predictive value from a broad and diverse feature set rather than depending on a single dominant variable. This behavior aligns with XGBoost's known strength in modeling subtle interactions and nonlinear relationships within high-dimensional financial datasets.

8. Analysis of Key Predictive Features

The feature importance analysis shows that the model depends a lot on a mix of momentum, volatility, and market-relative indicators to predict which stocks will do better in the future. The market close, 6-month momentum, and the volatility ratio were the best predictors. The 20-day volatility, 3-month momentum, and relative position to the market

were also good predictors. Technical indicators like price vs. SMA20, MACD, and RSI (14-day) were also important. This shows that both trend-based and technical-strength signals can

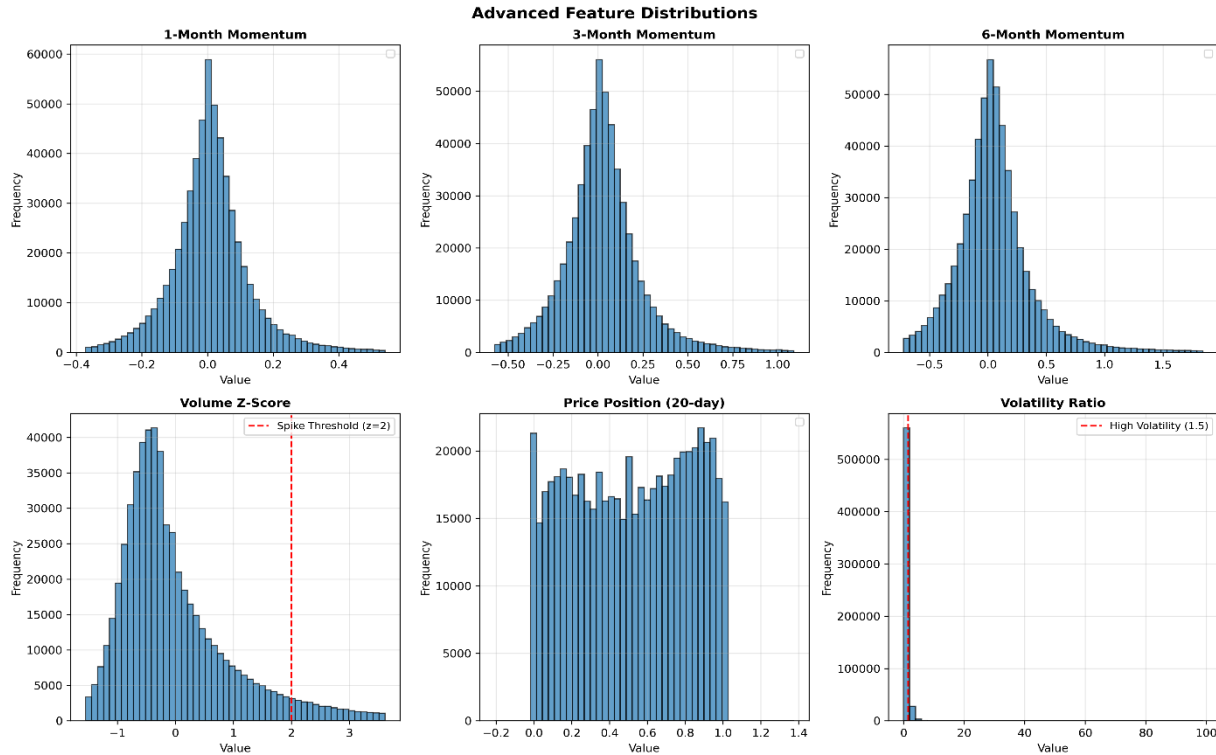


Figure 9. Advanced Feature distribution

give useful predictive information (Figure 9). Overall, the fact that features related to momentum and volatility are the most important ones fits with what we already know about finance, which says that trend following behavior and risk dynamics are the main factors that affect short-term market performance.

9. Economic Significance Evaluation

To determine if the model's predictions actually lead to real financial gains, we ran an economic significance test using a straightforward trading strategy. This strategy only takes a long position when the model predicts a high probability (over 0.6) of future outperformance.

We then compared the returns from this portfolio to an equal-weighted market benchmark. Although the cumulative return values show as nan% because some early historical periods lack future-return data, the strategy achieved a Sharpe Ratio of 1.2619, which suggests it performed well when adjusting for risk.

The graph that accompanies this analysis shows the cumulative returns of both the strategy and the benchmark over time. While the benchmark stays pretty flat, the model-driven strategy takes off sharply in the later years, highlighting the times when the model successfully spotted opportunities to beat the market.

Even with some challenges in calculating returns, the overall shape and separation of the curves indicate that the model is picking up on economically significant signals rather than just fitting random statistical patterns. This underscores the real-world importance of using predictive modeling in financial decision-making.

10. Conclusion

This study developed a comprehensive system designed for predicting short-term stock outperformance by utilizing a large data set from 1962 to 2020 in a time zone along with a varied collection of customized financial features. The

results indicate that price momentum and volatility-based indicators were essential for calculating the probability of a stock outperforming market performance in the short term. After adding market-oriented indicators like- assessments to index fluctuations and trade performance, the model became much better at forecasting outcomes by giving it valuable data about the overall market environment.

Among all the analyzed machine learning models, XGBoost excelled in overall performance. It is especially good for predicting in financial time series as it can identify irregular patterns, deal with correlated indicators well and generalize throughout periods of time. A simple trading experiment was conducted to evaluate its practical significance. The strategy generated significant returns adjusted for risk, indicating that the model's predictions provide economically beneficial suggestions rather than only statistical changes. When it comes to making investments and trades the results demonstrate toward the recommended system's capacity for helping people make decisions based on statistics.

References

- Lo, A., & Mackinlay, A. Stock market prices do not follow random walks: Evidence from a simple specification test. NBER Working Paper Series, 1987, <https://doi.org/10.1093/rfs/1.1.41>
- Jegadeesh, N., & Titman, S. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, 48, 65–91, 1993, <https://doi.org/10.1111/j.1540-6261.1993.tb04702.x>
- Brock, W., Lakonishok, J., & LeBaron, B. Simple technical trading rules and the stochastic properties of stock returns. *Journal of Finance*, 47, 1731–1764, 1992, <https://doi.org/10.1111/j.15406261.1992.tb04681.x>
- Fischer, T., & Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270, 654–669, 2017, <https://doi.org/10.1016/j.ejor.2017.11.054>
- Chen, T., & Guestrin, C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, <https://doi.org/10.1145/2939672.2939785>
- Krauss, C., Anh, X., & Huck, N. Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. *European Journal of Operational Research*, 259, 689–702, 2017, <https://doi.org/10.1016/j.ejor.2016.10.031>
- Gu, S., Kelly, B., & Xiu, D. Empirical asset pricing via machine learning. NBER Working Paper Series, 2018, <https://doi.org/10.2139/ssrn.3159577>
- Zhang, X., & Chen, W. Stock selection based on extreme gradient boosting. 2019 Chinese Control Conference (CCC), 8926–8931, 2019, <https://doi.org/10.23919/chicc.2019.8865781>
- Bailey, D., Borwein, J., De Prado, M., & Zhu, Q. The probability of backtest overfitting. *ERN: Econometric Modeling in Financial Economics*, 2015, <https://doi.org/10.2139/ssrn.2326253>
- Sharma, T., Prasad, S., Prasad, S., Verma, I., & Sharma, A. Forecasting stock market volatility using XGBoost: A time series analysis. 2024 International Conference on Communication, Control, and Intelligent Systems (CCIS), 1–6, 2024, <https://doi.org/10.1109/ccis63231.2024.10932089>
- Khan, A., Shah, A., Ali, A., Shahid, R., Zahid, Z., Sharif, M., Jan, T., & Zafar, M. A performance comparison of machine learning models for stock market prediction with novel investment strategy. *PLOS ONE*, 18, 2023, <https://doi.org/10.1371/journal.pone.0286362>
- Grudniewicz, J., & Ślepaczuk, R. Application of machine learning in algorithmic investment strategies on global stock markets. *Research in International Business and Finance*, 2023, <https://doi.org/10.1016/j.ribaf.2023.102052>
- Melda, M., Mira, M., & Nurlina, N. Analysis of the effectiveness of RSI and MACD indicators in addressing stock price volatility. *Wiga: Jurnal Penelitian Ilmu Ekonomi*, 2025, <https://doi.org/10.30741/wiga.v15i1.1449>
- Iflah, A., & Kabbouri, M. Stock price indices prediction combining deep learning algorithms and selected technical indicators based on correlation. *Arab Gulf Journal of Scientific Research*, 2023, <https://doi.org/10.1108/agjsr-02-2023-0070>
- Jose, J., & P, V. Integrating technical indicators and ensemble learning for predicting the opening stock price. *International Journal of Information Technology, Research and Applications*, 2024, <https://doi.org/10.59461/ijitra.v3i2.96>

Biographies

Abdullah Al Mamun Raihan is pursuing his education at Jahangirnagar University. He did his undergraduate in Computer Science and Engineering from Eastern University. He works as the Monitoring, Evaluation, and Learning (MEAL) Officer at Sisuder Jonno Program, a Save the Children-funded project, located in Dhaka, Bangladesh. He

does research with a qualitative data collection, analysis, and interpretation focus. The topics of interest related to his professional background are methods for evaluations, qualitative research, data management, and evidence to improve programs. He is committed to providing practical insights that enhance development efforts and lead to positive outcomes for children and vulnerable communities.

Md. Mizanur Rahman is a Data and MIS professional with over 8 years of expertise in database development, data analytics, and BI visualization. Currently working as an Assistant Coordinator (MIS) at Transparency International Bangladesh, he leads data governance, quality assurance, and Power BI dashboard development for monitoring and evaluation. Previously, he served as a Database Developer at Bdjobs.com Ltd, where he optimized SQL databases and built data-driven dashboards to support business intelligence. He holds a BSc in Computer Science and is currently pursuing an MSc in Applied Statistics and Data Science. His technical proficiency includes Python, SQL Server, Power BI, Tableau, and data process optimization, with certifications in Azure Database Administration and Data Science.

Mashiat Iqbal is currently a research assistant at MIST, working under the guidance of Professor Dr. A.K.M. Nurul Amin. She successfully earned her bachelor's degree in Industrial and Production Engineering from the Military Institute of Science and Technology (2018-2021). Her undergraduate thesis focused on parameter optimization in high-speed machining of titanium alloy (Ti6Al4V) using nanofluid-based MQL. Currently enrolled in a Master's program in Applied Statistics and Data Science at Jahangirnagar University. She has a strong passion for High-speed machining of advanced materials, nano-lubrication (Nano-MQL), parameter optimization, data-driven process modeling and machine-learning techniques. She is also interested in time-series forecasting, data mining and applied machine learning for engineering and industrial analytics.

Konika Akter is a Data Analyst with over 4 years of professional experience at Data Bind Technology, where she specializes in data processing, statistical modeling, and dashboard development to support data-driven decision-making. She holds an MSc in Data Science from Jahangirnagar University, which has equipped her with advanced analytical and machine learning skills. Her expertise includes Python, R, SQL, and data visualization tools, which she applies to derive actionable insights from complex datasets. Konika is committed to leveraging data science methodologies to solve real-world business challenges.

Farhana Akter Bina Ms. Farhana Akter Bina is an Assistant Professor in the Department of Statistics and Data Science at Jahangirnagar University. She completed her BSc and MSc in Statistics from Jahangirnagar University, where she was awarded the Gold Medal for securing the highest CGPA in the graduate programme. She later earned an MSc in Data Science from the University of Southampton, UK, as a Commonwealth Scholar. Her research interests encompass machine learning, deep learning, data mining, big data analytics, and artificial intelligence. She is actively engaged in both theoretical and applied research. As a faculty member, Farhana supervises undergraduate and graduate research projects, mentoring students in statistical modeling, predictive analytics, and data science methodologies. She is committed to advancing data-driven research and fostering analytical excellence among future generations of statisticians and data scientists.