

StockForecastGenie Pro: A Volatility-Aware Machine Learning Decision-Support System for Short-Horizon Equity Index Forecasting

Chibuike Chiedozie Ibebuchi¹

¹Department of Mathematics, Morgan State University, Baltimore, Maryland, USA

Corresponding author: Chibuike Chiedozie Ibebuchi, (Chibuike.ibebuchi@morgan.edu)

Corresponding author orcid: <https://orcid.org/0000-0001-6010-2330>

<https://doi.org/10.68414/NKSA6726>

Abstract

Forecasting short-term movements in financial markets remains challenging because market prices are influenced by rapidly changing economic conditions, investor sentiment, and volatility. While machine-learning methods have demonstrated promise for financial forecasting, many forecasting tools remain difficult for non-technical users to interpret and often provide predictions without contextual explanations. This paper presents StockForecastGenie Pro, a volatility-aware machine-learning decision-support system designed to forecast short-horizon movements in major United States equity indices while providing transparent and interpretable signals. The system forecasts future index levels, point gains, and percentage returns for the S&P 500 and Dow Jones Industrial Average using a stacked ensemble of machine-learning models. Forecast outputs are combined with market-volatility information derived from the CBOE Volatility Index (VIX) to generate plain-language signals ranging from Strong Negative to Strong Positive. The platform also incorporates confidence estimates, simple benchmark comparisons, and broader market context to support user interpretation. A rolling historical evaluation was conducted, and results indicate that forecasting performance improves substantially as forecast horizon increases. At the five-day horizon, the stacked ensemble achieved directional accuracy of 73.2% for the Dow Jones Industrial Average and 70.9% for the S&P 500. Compared with a no-change benchmark, the system reduced forecast error by approximately 24.3% for the Dow Jones Industrial Average and 22.4% for the S&P 500. The results demonstrate that machine-learning ensembles can provide meaningful improvements over simple forecasting approaches while maintaining interpretability for research, educational, and decision-support applications.

1. Introduction

Financial markets generate enormous volumes of data every trading day, creating opportunities for machine-learning systems to support forecasting and decision making. However, accurately forecasting short-term market movements remains a difficult challenge because market prices are influenced by economic conditions, investor sentiment, monetary policy, volatility, corporate information, and unexpected events (Chiu et al., 2018; Khan et al., 2024). These factors interact in complex ways and contribute to the uncertainty that characterizes financial markets, particularly at short forecasting horizons.

The growing availability of financial and macroeconomic data has accelerated the adoption of machine-learning methods in forecasting applications. Advances in computational power and data availability have enabled the development of forecasting systems capable of identifying nonlinear relationships and interactions that may be difficult to capture using traditional approaches (Ghani et al., 2019; Ibebuchi, 2025). As a result, machine learning has become an increasingly important component of modern financial analytics and decision-support systems (Ghani et al., 2019). Despite these advances, many forecasting platforms remain difficult for non-technical users to interpret. Forecasting systems frequently present numerical predictions without providing sufficient information about forecast confidence, market context, or the conditions under which forecasts should be trusted. Consequently, there is increasing interest in decision-support frameworks that combine predictive performance with transparency and interpretability.

To address this challenge, this study presents StockForecastGenie Pro, a volatility-aware machine-learning decision-support system for short-horizon equity-index forecasting. The platform focuses on two widely followed United States equity indices, the S&P 500 and the Dow Jones Industrial Average and integrates market and macroeconomic information from the Federal Reserve Economic Data platform with a stacked ensemble forecasting framework. The system produces forecasts of future index levels, point gains, and percentage returns while simultaneously translating model outputs into user-friendly signals that incorporate forecast confidence, market volatility, and broader market context.

A distinguishing characteristic of StockForecastGenie Pro is that forecasts are not presented as direct trading instructions. Instead, forecast outputs are evaluated within the context of market volatility measured by the CBOE Volatility Index and converted into interpretable

signals ranging from Strong Negative to Strong Positive. This framework is intended to help users understand not only the forecast itself but also the level of confidence and market conditions associated with that forecast. The objective of this study is to evaluate whether a volatility-aware machine-learning decision-support framework can provide meaningful forecasting value beyond simple benchmark methods while maintaining interpretability for users. Specifically, the study investigates forecasting performance across one-, three-, and five-trading-day horizons using a rolling historical evaluation.

The contributions of this work are threefold. First, it presents an end-to-end machine-learning forecasting and decision-support platform for major United States equity indices. Second, it integrates stacked ensemble forecasting with volatility-aware signal generation and user-oriented explanations. Third, it provides a comprehensive rolling evaluation demonstrating how forecasting performance changes across forecast horizons, benchmark comparisons, and volatility regimes.

2. Literature Review

Financial forecasting has been an active area of research within economics, finance, statistics, and machine learning for several decades (Dingli & Fournier, 2017). Early forecasting studies primarily relied on statistical time-series methods, including autoregressive and moving-average frameworks, to model market behavior (Dingli & Fournier, 2017). While these approaches can capture temporal structure in financial data, they often struggle to represent nonlinear relationships and complex interactions among market and macroeconomic variables.

The emergence of machine-learning techniques introduced more flexible approaches for financial forecasting (Ghani et al., 2019). Tree-based algorithms such as Random Forests, Extra Trees, and gradient-boosting methods have been widely adopted because of their ability to model nonlinear relationships, accommodate large numbers of predictors, and automatically capture interactions among variables (Benghiat & Lahmiri, 2025). These characteristics have made machine-learning methods attractive alternatives to traditional statistical forecasting models in financial applications. More recently, ensemble-learning approaches have gained attention because individual forecasting models often capture different aspects of market behavior (Apu et al., 2022). Rather than relying on a single model, ensemble methods combine multiple forecasts to improve robustness and reduce sensitivity to model-specific weaknesses.

Although substantial progress has been made in improving forecasting accuracy, interpretability remains a persistent challenge. Many forecasting studies focus primarily on predictive performance while providing limited insight into how forecasts should be interpreted or how forecast reliability may change under different market conditions (Poon & Granger, 2003). In addition, some studies evaluate models using static train-test splits rather than operational rolling forecasts that mimic how a model would be used to forecast future market closes in real time. This limitation is particularly important in financial applications, where users often require not only numerical predictions but also context regarding uncertainty, market regime, and practical decision support.

Another area that has received comparatively less attention is the integration of market-volatility information into the decision-support layer of forecasting systems. Volatility measures such as the CBOE Volatility Index provide valuable information regarding prevailing market uncertainty and may influence how forecasting signals should be interpreted (Fernandes et al., 2014). A forecast that appears meaningful during a low-volatility environment may carry a different implication during periods of elevated market stress.

The present study contributes to this literature by combining stacked ensemble forecasting with a volatility-aware decision-support framework. Rather than focusing solely on predictive accuracy, StockForecastGenie Pro integrates forecast magnitude, confidence estimates, benchmark comparisons, volatility conditions, and broader market context into a unified signal-generation process. This approach seeks to improve both operational forecasting performance and interpretability, making the resulting forecasts more useful for research, educational, and financial decision-support applications.

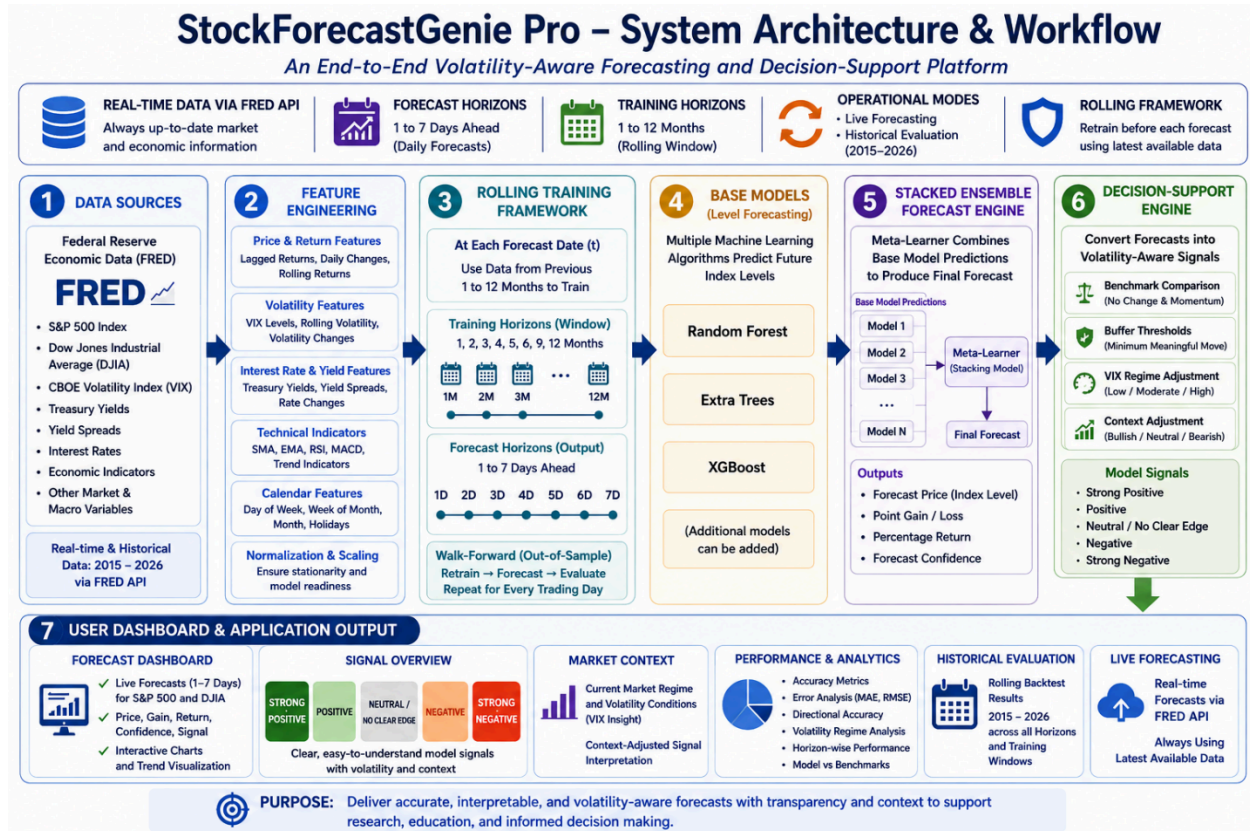
3. Data and Methods

StockForecastGenie Pro was developed as an end-to-end forecasting and decision-support platform for short-horizon equity-index forecasting. As shown in Figure 1, the system integrates data acquisition, feature engineering, machine-learning forecasting, volatility-aware signal generation, and user-facing visualization within a single application workflow.

The platform uses publicly available market and macroeconomic data obtained from the Federal Reserve Economic Data platform (Federal Reserve Bank of St. Louis, n.d.). The

forecasting targets are the S&P 500 Index and the Dow Jones Industrial Average. Input features include market-index behavior, lagged returns, rolling changes, volatility indicators, Treasury yields, yield-spread information, interest-rate variables, calendar features, and technical trend indicators (Figure 1). These variables were included to capture market momentum, economic conditions, and prevailing financial-market uncertainty.

Figure 1. System architecture of StockForecastGenie Pro.



For the experimental evaluation, point gain was used as the primary supervised learning target. The predicted point gain was then converted into an implied forecast price and percentage return, allowing the dashboard to display forecast price, forecast point gain, and forecast percentage return from the same underlying prediction.

The forecasting engine combines three base machine-learning models: Extra Trees (Geurts et al., 2006), Random Forest (Breiman, 2001), and XGBoost (Chen & Guestrin, 2016). Each model is trained using the same engineered feature set and produces an independent forecast. The stacked ensemble combines base-model forecasts using a Huber-regression

meta-learner, with a logistic-regression layer used for directional probability estimation (Ibebuchi, 2025). The stacked ensemble serves as the primary forecasting engine, while the individual models are retained for comparison and diagnostic analysis. In this study, to mimic operational deployment of the StockForecastGenie Pro, the evaluation used a rolling-window design. For each forecast date, the model was trained using only the previous six months of data and then evaluated at one-, three-, and five-trading-day horizons. This prevents look-ahead bias and reflects how the application would operate in real time. The evaluation was conducted over the 2025 calendar year, producing 261 rolling forecasts for each target market and forecast horizon.

Model performance was compared against two simple reference forecasts. The first assumes no market change over the forecast horizon. The second assumes that the most recent comparable market movement continues into the future. These benchmarks were included because a machine-learning model should demonstrate value beyond simple assumptions available to any user.

StockForecastGenie Pro does not present raw forecasts as direct trading instructions. Instead, the forecast is converted into a research-oriented model signal. Signal generation considers forecast direction, forecast magnitude, model confidence, benchmark comparison, and the current volatility environment. The CBOE Volatility Index is used to classify market conditions into volatility regimes. Under higher-volatility conditions, the system requires a larger expected movement before issuing a directional signal. The decision framework uses volatility-dependent minimum-move thresholds of approximately 0.5%, 0.7%, and 1.0% under low-, normal-, and high-volatility regimes, respectively. Forecasts that do not exceed these thresholds are classified as Neutral / No Clear Edge. This buffer logic reduces the likelihood that small forecast differences are overinterpreted as meaningful market opportunities during periods of elevated uncertainty. The application reports five signal categories: Strong Positive Model Signal, Positive Model Signal, Neutral / No Clear Edge, Negative Model Signal, and Strong Negative Model Signal. Neutral / No Clear Edge indicates that the model does not identify a sufficiently strong directional opportunity after applying the benchmark comparison, confidence requirement, volatility adjustment, and buffer logic.

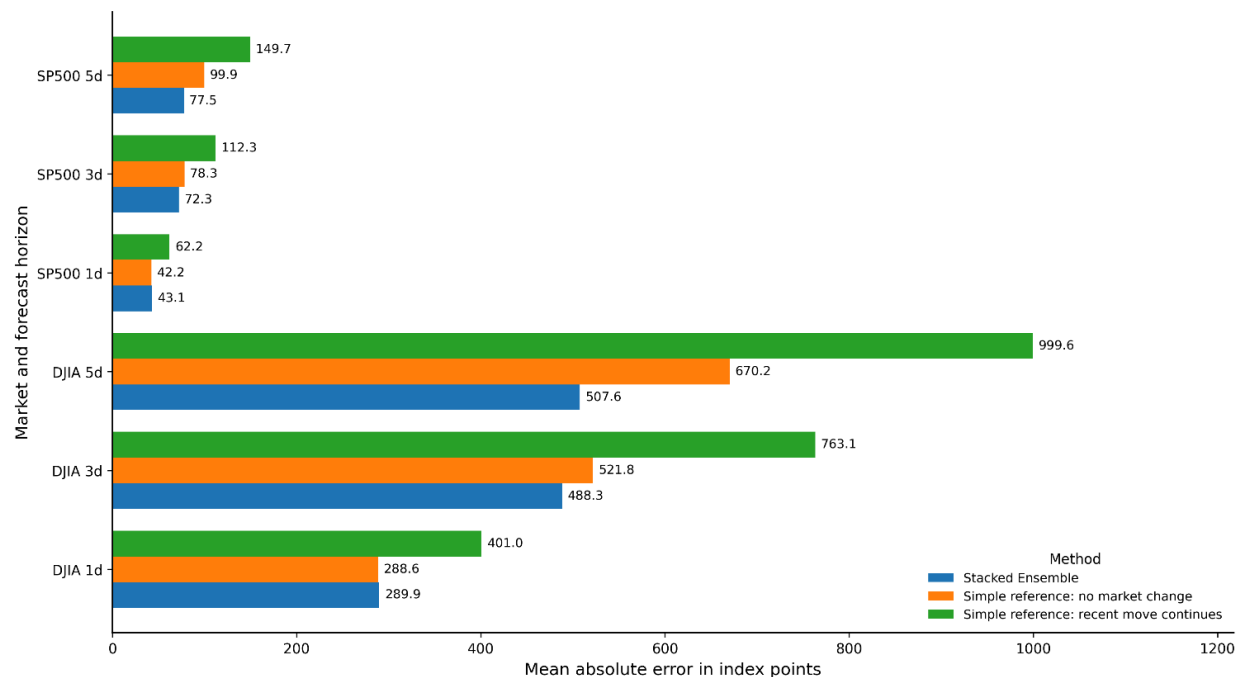
The system also provides a longer-horizon context adjustment. Recent market behavior within the rolling training window is used to classify the broader market environment as bullish, bearish, or neutral. This context does not replace the short-term forecast but helps users interpret whether the short-term signal agrees with or conflicts with the broader market trend. Evaluation metrics included mean absolute error, root mean squared error, directional accuracy, and percentage improvement relative to the no-change reference forecast. Additional analyses examined performance across volatility regimes and the distribution of generated model signals. Together, these metrics evaluate both the forecasting accuracy of the model and the usefulness of the decision-support layer.

4. Results

4.1 Forecast Accuracy Across Horizons

Figure 2 compares forecast error across forecast horizons for the stacked ensemble and the reference forecasts. A clear pattern emerges. Forecasting performance improves as the forecast horizon increases from one to five trading days.

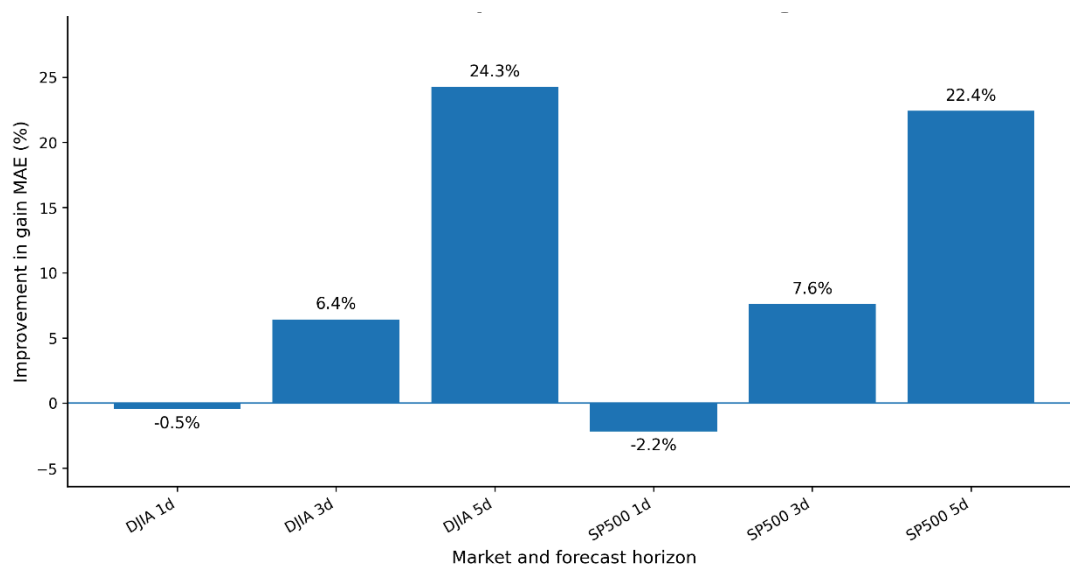
Figure 2. Forecast error (mean absolute error) of the stacked ensemble and simple reference forecasts across the S&P 500 and Dow Jones Industrial Average at one-, three-, and five-day forecast horizons.



At the one-day horizon, the machine-learning framework provides little improvement over the no-change benchmark. This result is not unexpected. Daily movements of broad equity indices are highly noisy and are often difficult to distinguish from random variation. Furthermore, because the S&P 500 and Dow Jones Industrial Average exhibit a long-run upward tendency, simple reference forecasts can remain surprisingly competitive over very short horizons. Performance improves substantially at the three-day horizon and improves further at the five-day horizon (Figure 2). This suggests that the forecasting framework is more effective at identifying short-term market structure than predicting next-day fluctuations. These findings support the use of multi-day horizons for machine-learning-based equity-index forecasting and indicate that the greatest value of the framework lies beyond next-day prediction.

Figure 3 presents the percentage improvement of the stacked ensemble relative to the no-change benchmark. The one-day horizon shows little or no improvement. However, substantial gains emerge at longer horizons. For the five-day horizon, the stacked ensemble reduced forecast error by approximately 22.4% for the S&P 500 and 24.3% for the Dow Jones Industrial Average. These results demonstrate that the machine-learning framework provides useful information beyond what can be obtained from simple benchmark assumptions. Importantly, the improvements are observed consistently across both target indices.

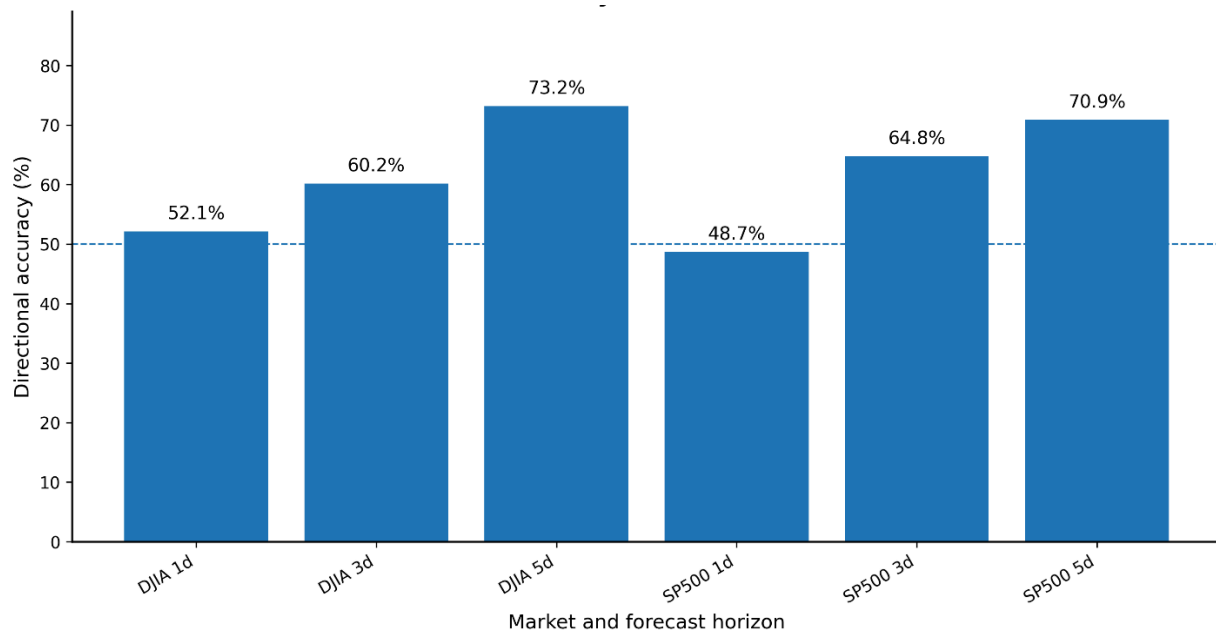
Figure 3. Percentage improvement of the stacked ensemble relative to the no-change reference forecast. Positive values indicate lower forecast error than the benchmark.



4.2 Directional Accuracy

Figure 4 summarizes directional forecasting performance. Again, directional accuracy increased systematically as forecast horizon increased. For the S&P 500, directional accuracy increased from approximately 48.7% on the one-day horizon to 70.9% on the five-day horizon. For the Dow Jones Industrial Average, directional accuracy increased from approximately 52.1% to 73.2%, respectively.

Figure 4. Directional accuracy of the stacked ensemble across forecast horizons for the S&P 500 and Dow Jones Industrial Average.



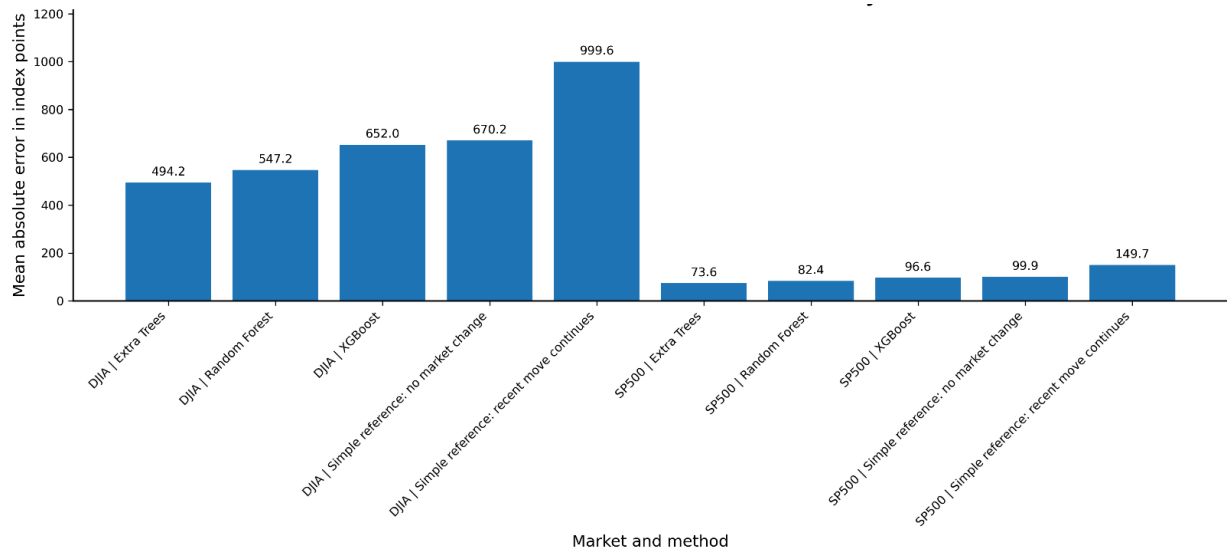
The directional results reinforce the conclusions obtained from the error metrics (Figure 2). The framework is considerably more effective at identifying short-term market direction over multi-day horizons than at predicting next-day movements. From a decision-support perspective, directional accuracy is particularly important because users are often more interested in the expected direction of market movement than in the precise numerical forecast.

4.3 Comparison with Individual Models

Figure 5 compares the stacked ensemble with individual forecasting models and simple reference methods. The results indicate that no single model consistently dominates all markets and forecast horizons. Extra Trees frequently produced the strongest individual-model performance, while Random Forest and XGBoost remained competitive (Figure 5). The stacked ensemble generally produced more stable performance across scenarios by combining

information from multiple forecasting models. This supports the rationale for adopting an ensemble framework rather than relying on a single algorithm. Moreover, the objective of the ensemble is not necessarily to outperform every individual model in every situation but rather to provide a robust and reliable forecasting framework across changing market conditions.

Figure 5. Comparison of individual machine-learning models, the stacked ensemble, and simple reference forecasts at the five-day forecast horizon.



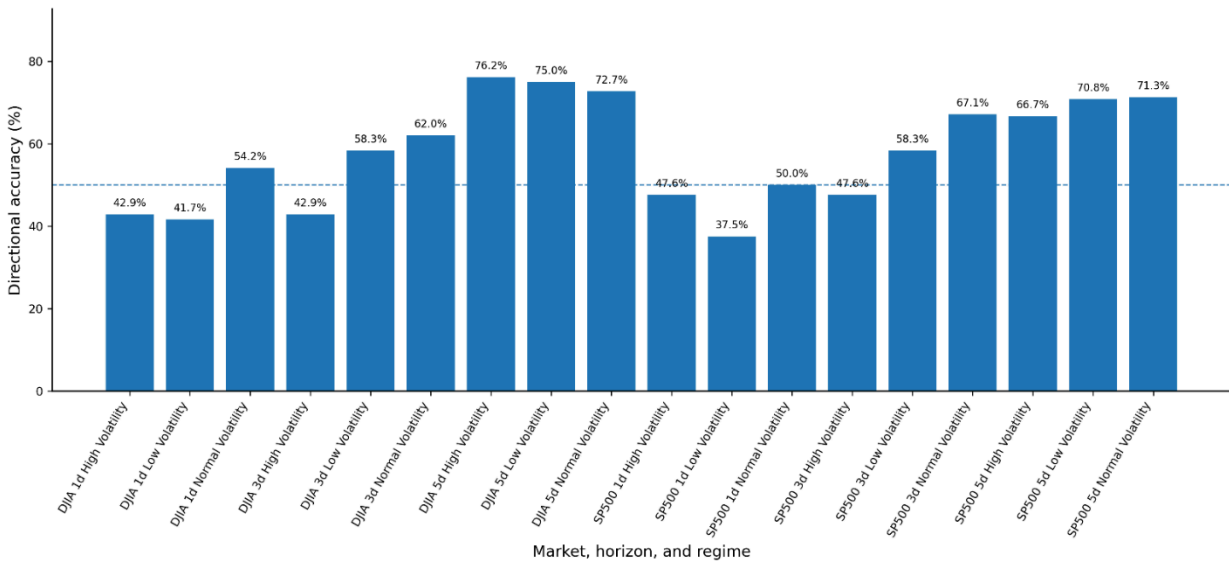
4.4 Volatility Regime Analysis

Figure 6 evaluates forecasting performance across VIX-based volatility regimes. Forecast performance varied across volatility regimes (Figure 6). For the Dow Jones Industrial Average, average directional accuracy increased from 54.0% during high-volatility periods to 58.3% during low-volatility periods and 63.0% during normal-volatility periods. A similar pattern was observed for the S&P 500, where average directional accuracy was 54.0%, 55.5%, and 62.8% under high-, low-, and normal-volatility conditions, respectively. These results suggest that forecasting performance is generally strongest during normal market conditions and tends to deteriorate during periods of elevated uncertainty. Nevertheless, the model maintained directional accuracy near or above 50% across most volatility regimes (Figure 6), indicating that the volatility-aware framework remained reasonably robust even during more challenging market environments.

These findings justify the volatility-aware design of StockForecastGenie Pro. By incorporating VIX into the signal-generation framework, the system adjusts its decision

thresholds according to prevailing market conditions rather than treating all market environments equally.

Figure 6. Directional accuracy of the stacked ensemble across VIX-based volatility regimes. Performance generally varies with market uncertainty.

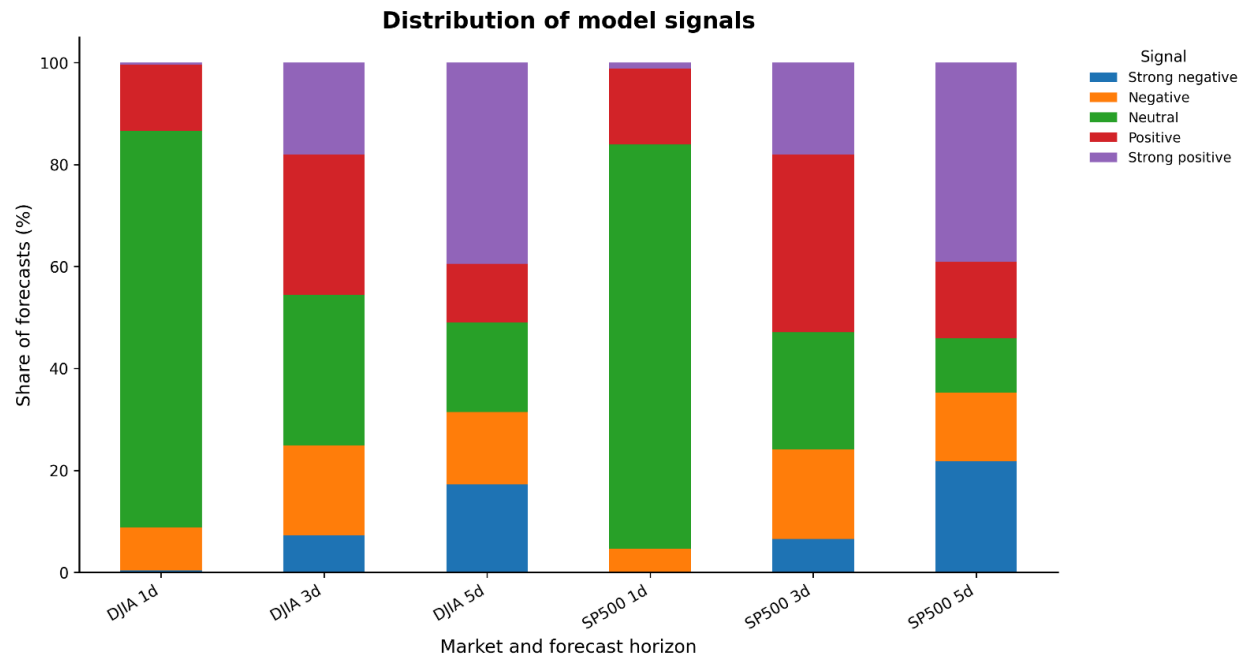


4.5 Signal Distribution and Decision Support

Figure 7 summarizes the distribution of generated signals. A notable characteristic of the framework is that it does not continuously issue strong positive or strong negative signals. Instead, a notable proportion of forecasts are classified as Neutral / No Clear Edge.

This behavior is intentional and reflects the threshold-based decision logic built into the platform. Small forecast advantages are filtered out before a directional signal is generated. This conservative behavior is particularly important for educational and decision-support applications where excessive signal generation can encourage overinterpretation of uncertain forecasts.

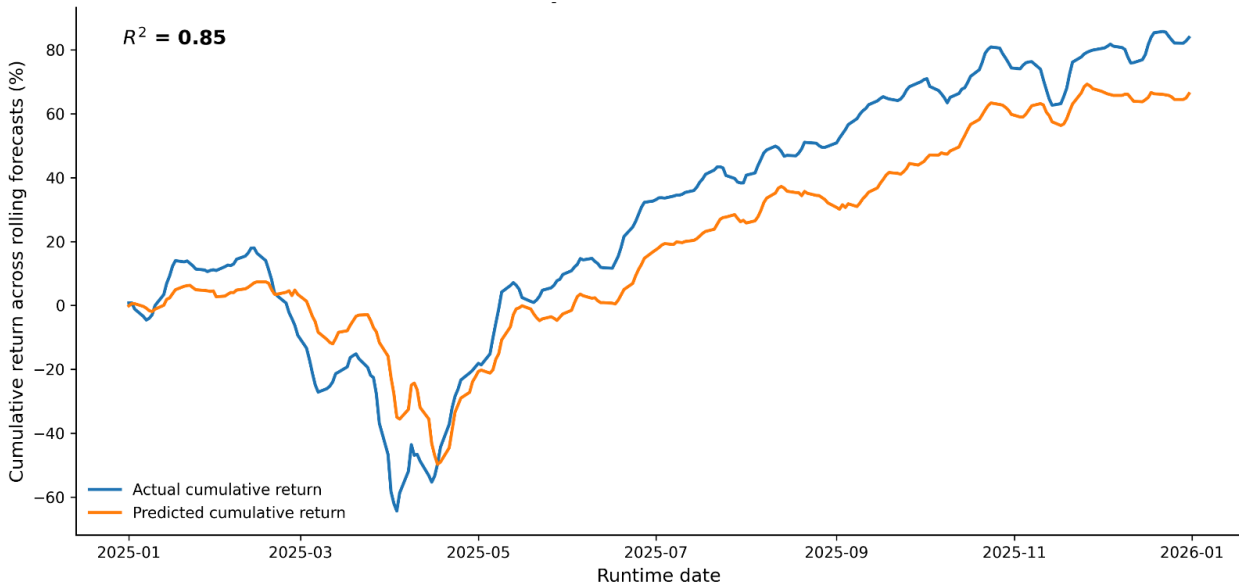
Figure 7. Distribution of generated model signals across markets and forecast horizons, showing the relative frequency of Strong Positive, Positive, Neutral, Negative, and Strong Negative signals.



4.6 Long-Horizon Forecast Tracking

Finally, Figure 8 compares cumulative predicted returns with cumulative realized returns. Although individual forecast errors remain unavoidable, the cumulative trajectories demonstrate that the forecasting framework captures broader market tendencies over time. The predicted cumulative return closely follows the realized cumulative return throughout much of the evaluation period. This result provides additional evidence that the framework is identifying meaningful information beyond simple short-term fluctuations and supports its use as a research-oriented forecasting platform.

Figure 8. Cumulative predicted and realized returns for rolling five-day forecasts. The close correspondence between the two series indicates that the forecasting framework captures broader market tendencies over time.



5. Discussion

This study presented StockForecastGenie Pro, a volatility-aware machine-learning decision-support system for short-horizon forecasting of the S&P 500 and Dow Jones Industrial Average. The platform combines stacked ensemble forecasting with benchmark comparison, volatility-aware thresholding, confidence assessment, and user-oriented explanations.

The 2025 rolling evaluation shows that forecast skill is strongly horizon dependent. One-day forecasts provided limited improvement over simple reference methods, consistent with the noisy and often near-random behavior of short-term equity-index movements (Fama, 1970; Malkiel, 2003; Gu et al., 2020). Performance improved at longer horizons, with the five-day stacked ensemble reducing forecast error by approximately 22.4% for the S&P 500 and 24.3% for the Dow Jones Industrial Average relative to the no-change reference. Directional accuracy also exceeded 70% for both indices. Volatility-regime results further support the platform's decision-support design: accuracy was strongest under normal-volatility conditions and weaker during high-volatility periods, reflecting the greater uncertainty of turbulent markets (Christoffersen, 1998; Ang & Timmermann, 2012). These findings justify the use of VIX-based thresholds and buffer logic to avoid overinterpreting small forecast differences during unstable periods.

The signal-distribution results also highlight the practical value of the platform. StockForecastGenie Pro does not force every forecast into a positive or negative decision. Instead, the Neutral / No Clear Edge category acts as a conservative filter when the model does not identify a sufficiently strong directional opportunity. This feature is important for educational and research settings because it discourages overconfidence and emphasizes uncertainty-aware interpretation. Overall, the findings suggest that the primary value of StockForecastGenie Pro lies in identifying short-term market structure over multi-day horizons rather than predicting next-day fluctuations.

6. Conclusion

StockForecastGenie Pro demonstrates that a volatility-aware machine-learning framework can provide useful short-horizon forecasting support for major United States equity indices while maintaining interpretability for users. For the evaluation in this study, the strongest results occurred at the five-trading-day horizon, where the stacked ensemble outperformed the no-change benchmark and achieved directional accuracy above 70% for both evaluated indices.

Future versions of StockForecastGenie Pro will extend coverage to additional equity indices and exchange-traded funds. FRED-supported additions may include the NASDAQ Composite and NASDAQ-100, while Yahoo Finance can support broader asset coverage, including Russell 2000, index-tracking ETFs, sector ETFs, and selected large-cap equities. Subsequent versions will also evaluate more advanced machine-learning architectures, including neural networks, while maintaining the platform's emphasis on transparency and responsible decision support.

References

- Ang, A., & Timmermann, A. (2012). Regime changes and financial markets. *Annual Review of Financial Economics*, 4(1), 313–337.
- Apu, K. U., Rahman, M. M., Hoque, A. B., & Bhuiyan, M. (2022). Forecasting future investment value with machine learning, neural networks, and ensemble learning: A meta-analytic study. *Review of Applied Science and Technology*, 1(02), 01–25.
- Benghiat, S., & Lahmiri, S. (2025). Evaluating predictive models for three green finance markets: Insights from statistical vs. machine learning approaches. *Computation*, 13(3), 76.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794).
- Chiu, C. W. J., Harris, R. D., Stoja, E., & Chin, M. (2018). Financial market volatility, macroeconomic fundamentals and investor sentiment. *Journal of Banking & Finance*, 92, 130–145.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.
- Dingli, A., & Fournier, K. S. (2017). Financial time series forecasting: A machine learning approach. *Machine Learning and Applications: An International Journal*, 4(1/2/3), 11–27.
- Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2), 383–417.
- Federal Reserve Bank of St. Louis. (n.d.). *Federal Reserve Economic Data (FRED)* [Data repository]. Retrieved June 12, 2026, from <https://fred.stlouisfed.org>
- Fernandes, M., Medeiros, M. C., & Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, 40, 1–10.
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, 63(1), 3–42.
- Ghani, M. U., Awais, M., & Muzammul, M. (2019). Stock market prediction using machine learning algorithms. *Advances in Distributed Computing and Artificial Intelligence Journal*, 8(4), 97–116.
- Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223–2273.
- Ibebuchi, C. C. (2025). Day-ahead energy price forecasting with machine learning: Role of endogenous predictors. *Forecasting*, 7(2), 18.
- Khan, M. A., Ali, H., Shabbir, H., Noor, F., & Majid, M. D. (2024). Impact of macroeconomic indicators on stock market predictions: A cross-country analysis. *Journal of Computing & Biomedical Informatics*, 8(01).
- Malkiel, B. G. (2003). The efficient market hypothesis and its critics. *Journal of Economic Perspectives*, 17(1), 59–82.
- Poon, S. H., & Granger, C. W. J. (2003). Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2), 478–539.