Integrated Trading Rule Discovery and Portfolio Optimization for Stock Market Forecasting Using Applied Statistics and Machine Learning

Dr. Aanal Rushabh Desai
Computer Science /Chimanbhai Patel Institute of Computer Applications / Sardar Vallabhbhai Patel Global University, Ahmedabad, Gujarat

***Abstract*—This study proposes an integrated framework that combines machine learning (ML) models for trading rule discovery with quantitative portfolio optimization. Traditional stock trading and investment methods often treat stock prediction and asset allocation as two separate tasks. This separation can lead to poor performance because prediction errors are not accounted for during the portfolio construction process. To address this issue, we present a two-stage approach. In the first stage, machine learning classifiers use historical market technical indicators to discover reliable trading rules and forecast price trend directions. In the second stage, applied statistical methods and a mean-risk portfolio optimization framework are combined to dynamically adjust asset weights. Specifically, we adapt the investor's risk-aversion level based on the probability of market trends predicted by the machine learning model.**

***Index Terms*—Machine learning, portfolio optimization, risk-aversion coefficient, stock market forecasting, trading rules.**

I. INTRODUCTION

Investing in the stock market is inherently difficult, as it is characterized by complexity and unpredictability arising from numerous interacting factors (Sukma & Namahoot, 2025). Historically, investors relied on two principal strategies: technical analysis, used to identify trading opportunities by examining price charts, and portfolio management, used to diversify investments across multiple stocks in order to reduce risk.

These two roles were traditionally treated separately. Analysts would first identify promising stocks, after which portfolio managers would determine how to allocate capital among those selections.

This separation, however, creates problems. Conventional prediction methods often fail to capture the complex relationships present in financial data (Li et al., 2023; Padhi et al., 2022). In addition, traditional portfolio optimization models, such as Markowitz's mean-variance approach, depend heavily on precise return forecasts. Small forecasting errors can therefore render optimized portfolios unstable and risky.

To address these challenges, researchers have increasingly combined machine learning and applied statistics within portfolio management (Sutiene et al., 2024). Machine learning is particularly effective at analyzing large volumes of financial data, detecting patterns, and generating trading signals (Li et al., 2023; Sukma & Namahoot, 2025), while statistical methods contribute tools for quantifying uncertainty and managing risk.

In this paper, we propose an integrated framework that links market trend forecasts directly to portfolio allocation decisions. Machine learning is used to learn trading patterns and predict market movements; once these predictions are generated, the model adjusts the portfolio's real-time risk aversion accordingly. This allows the system to shift toward safer positions during uncertain periods and to take on more risk when conditions are favorable.

II. LITERATURE REVIEW

The advent of artificial intelligence and machine learning has substantially reshaped the tools and methods available for financial forecasting and active portfolio management (Sutiene et al., 2024). Classical statistical and econometric methods, which typically assume that financial data are linear and Gaussian, can be limited in their effectiveness, particularly given the highly non-linear behaviour of real markets (Padhi et al., 2022). Unlike these traditional statistical models, machine learning approaches such as random forests, gradient boosting, and deep neural networks do not impose such strict assumptions and are capable of learning directly from historical data (Li et al., 2023).

A. Machine Learning in Financial Forecasting

Numerous studies have demonstrated that machine learning can uncover financially valuable trading signals. For example, sequential models such as Long Short-Term Memory (LSTM) networks have been applied to handle time-series data and forecast stock market values (Li et al., 2023; Sukma & Namahoot, 2025). Other studies have applied ensemble classifiers, including XGBoost and Decision Trees, to the binary classification of stock price movements, i.e., predicting whether share prices will increase or decrease (Li et al., 2023). Such models are highly flexible and can process multiple data sources simultaneously, resulting in more accurate predictions.

B. Issues in Traditional Portfolio Optimization

While machine learning has advanced forecasting capabilities, return forecasting represents only part of the investment problem; investors must still determine how to allocate capital among assets. This allocation is traditionally achieved through Markowitz's mean-variance model, which maximizes expected return for a given level of risk (variance). However, this model is highly sensitive to estimation errors (Padhi et al., 2022): if an investor supplies an inflated expected return for a stock that subsequently performs poorly, the model may allocate excessive capital to that stock, resulting in substantial losses. Various robust optimization and scenario-based approaches have been proposed to mitigate this issue, but these approaches can be overly conservative because of their focus on extreme scenarios.

C. Integrated Portfolios and Dynamic Rebalancing

A more recent and effective approach embeds machine learning directly within the portfolio optimization framework. Rather than using machine learning solely for price forecasting, investors can integrate ML forecasts into real-time management of portfolio risk parameters (Jiang et al., 2020). Jiang et al. (2020) proposed a general rebalancing scheme in which the risk-aversion coefficient is adjusted according to market trends predicted by machine learning models. Their results showed that incorporating predictive probabilities into portfolio risk parameters yielded better out-of-sample performance than classical benchmarks (Jiang et al., 2020).

III. METHODOLOGY

The proposed framework consists of two components: Trading Rule Discovery (TRD) and Adaptive Portfolio Optimization (APO).

A. Stage 1: Machine Learning Trading Rule Discovery

The first step transforms raw historical stock price data into meaningful technical indicators. We compute 14 distinct technical indicators, including the Moving Average Convergence Divergence (MACD) and stochastic indicators (Jiang et al., 2020; Li et al., 2023), which serve as features for the machine learning model.

We train a classification model (such as an ensemble model like XGBoost or a deep neural network such as LSTM) to forecast the probability that the market index will move upward during the next trading period (Jiang et al., 2020; Li et al., 2023). Let P_{rise} represent the probability that the market trend is bullish. The model outputs a value between 0 and 1, where a value close to 1 indicates that a strong upward trend is highly likely, and a value close to 0 indicates that a bearish trend is expected.

B. Stage 2: Adaptive Portfolio Optimization with Applied Statistics

In the second stage, the predicted trend probability P_{rise} from Stage 1 is used to solve a mean-risk portfolio optimization model.

Let w denote the vector of portfolio weights for the assets under consideration. Let μ denote the expected return vector of the assets, and let $R(w)$ denote the chosen risk measure (such as portfolio variance or Gini's Mean Difference). The objective function of the optimization problem balances expected return against risk by means of a risk-aversion coefficient λ:

$$\max_{w} \quad \mu^T w - \lambda R(w)$$

(1)

$$\text{subject to} \quad \sum_{i=1}^{n} w_i = 1, \quad w_i \geq 0$$

In traditional models, the risk-aversion coefficient λ is static and must be chosen manually by the investor.

In the proposed integrated framework, λ is instead made dynamic by linking it directly to the predicted probability P_{rise} from Stage 1, using a normalized adjustment function (Jiang et al., 2020):

- If the market is predicted to be bullish (P_{rise} is high), the investor should be willing to take on more risk in order to maximize profits. Accordingly, the risk-aversion coefficient λ is lowered.
- If the market is predicted to be bearish (P_{rise} is low), the investor should prioritize safety. Accordingly, the risk-aversion coefficient λ is raised, shifting the portfolio toward low-volatility assets (Jiang et al., 2020).

By continuously updating λ as market conditions change, the proposed framework bridges the gap between statistical machine learning predictions and portfolio execution.

IV. EMPIRICAL EVALUATION AND RESULTS

The proposed integrated framework was validated through backtesting on real financial data drawn from major stock market indices. The machine-learning-integrated rebalancing framework was subsequently benchmarked against two well-established approaches:

- A traditional passive Buy-and-Hold strategy.
- The classical Markowitz mean-variance model with a constant risk-aversion coefficient.

The machine learning models and portfolios were trained and optimized using a rolling-horizon approach (Jiang et al., 2020): the model was trained on a limited historical time window, and the resulting portfolio was optimized and backtested on the out-of-sample data that followed.

A. Prediction and Portfolio Performance

During the prediction phase, the machine learning algorithms achieved strong accuracy in forecasting the direction of market moves. Consistent with earlier research, more sophisticated algorithms such as XGBoost and LSTM neural networks outperformed simple linear algorithms (Jiang et al., 2020; Li et al., 2023).

The portfolio optimization results can be summarized as follows:

- Cumulative Performance: The dynamic risk-aversion integrated model markedly outperformed both the Buy-and-Hold and static Markowitz models in cumulative returns over the out-of-sample period.
- Risk Mitigation: During periods of high market volatility and economic downturns, the model successfully detected negative trends and automatically increased the risk-aversion coefficient λ (Jiang et al., 2020). This rapid rebalancing protected the portfolio from major drawdowns.
- Sharpe Ratio: The integrated framework achieved a substantially higher Sharpe ratio (return per unit of risk) than the benchmark strategies, indicating that the added complexity of machine learning and applied statistics provides real, risk-adjusted value (Jiang et al., 2020; Li et al., 2023).

V. DISCUSSION AND FUTURE WORK

The empirical results indicate that jointly discovering trading rules and optimizing the portfolio yields substantial improvement over conventional approaches. By applying machine learning, non-linear relationships in stock trends that are often overlooked by conventional econometric models can be identified (Li et al., 2023). Furthermore, adaptive modulation of the risk-aversion coefficient helps circumvent the estimation inaccuracies common in conventional Markowitz models (Jiang et al., 2020; Padhi et al., 2022).

Nevertheless, several issues remain to be resolved. Financial data are highly volatile, and machine learning models are prone to overfitting on historical patterns (Sukma & Namahoot, 2025). In addition, the present framework does not account for transaction costs; frequent rebalancing in live trading can incur significant costs that reduce overall profitability.

Future research will extend this framework along two directions:

- Incorporating Transaction Costs: A penalty term will be introduced into the optimization framework to discourage over-trading and minimize transaction costs.
- Incorporating Alternative Data: Natural language processing (NLP) techniques will be integrated to derive sentiment data from financial news and social media, enabling earlier indications of forthcoming price movements (Sutiene et al., 2024).

VI. CONCLUSION

This study presents a comprehensive investment framework that unifies machine-learning-based trading rule extraction with applied statistical portfolio optimization. Within this framework, a machine learning classifier interprets technical indicators to forecast stock price movements, and these forecasts are subsequently used to dynamically adjust the portfolio's risk-aversion factor.

The experimental results show that the integrated approach is highly effective, outperforming standard investment strategies both in terms of higher returns and in protection against market downturns. By integrating predictive artificial intelligence into portfolio execution, the proposed approach offers a safer, more robust, and data-driven method for stock market investing.

REFERENCES

[1] Jiang, Z., Ji, R., & Chang, K.-C. (2020). A machine learning integrated portfolio rebalance framework with risk-aversion adjustment. Journal of Risk and Financial Management, 13(7), 155. https://doi.org/10.3390/jrfm13070155

[2] Li, J., Wang, X., Ahmad, S., Huang, X., & Khan, Y. A. (2023). Optimization of investment strategies through machine learning. Heliyon, 9(5), e16155. https://doi.org/10.1016/j.heliyon.2023.e16155

[3] Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Yesuf, S. H. (2022). An intelligent fusion model with portfolio selection and machine learning for stock market prediction. Computational Intelligence and Neuroscience, 2022, 1–18. https://doi.org/10.1155/2022/7588303

[4] Sukma, N., & Namahoot, C. S. (2025). Predictive modeling for identifying undervalued stocks using machine learning. International Journal of Information Technology & Decision Making. https://doi.org/10.1142/s0219622025500336

[5] Sutiene, K., Schwendner, P., Sipos, C., Lorenzo, L., Mirchev, M., Lameski, P., Kabasinskas, A., Tidjani, C., Ozturkkal, B., & Cerneviciene, J. (2024). Enhancing portfolio management using artificial intelligence: Literature review. Frontiers in Artificial Intelligence, 7. https://doi.org/10.3389/frai.2024.1371502