International Journal of Informatics and Data Science Research
ISSN 2997-3961 (Online)
Vol. 2, No. 5, May 2025,
Available online at: https://scientificbulletin.com/index.php/IJIDSR

A Hybrid Methodology for Mutual Fund Trend Prediction Using Association Rule Mining and Random Forest Classifiers

Ms. Priyanka Bhandari (jain)
Research scholar, Department of Computer Science & IT, Janardan Rai Nagar Rajasthan, Vidyapeeth (Deemed to be University), Udaipur

Dr. Manish Shrimali
Associate professor, Department of Computer Science & IT, Janardan Rai Nagar Rajasthan, Vidyapeeth (Deemed to be University), Udaipur

Article information:
Manuscript received: 4 Mar 2024; **Accepted:** 10 Apr 2024; **Published:** 31 May 2025

Abstract: Mutual fund dynamics are difficult to forecast since they include the complete macroeconomic environment and market behaviour. This makes it one of the hardest methods. ARM and RF categorization are used in this study to improve mutual fund estimates. The analysis includes closing prices of five funds, VFIAX, FXAIX, VIGRX, VTMGX, and VTSAX, from 2011 to 2024. GDP growth, inflation, and interest rates are considered. The Apriori approach is utilized in association rule extraction to identify intentional participation patterns across economic levels and acceptable co-occurrence patterns across multi-dimensional economic scenes. To improve accuracy, acquired rules are included as features to Random Forest models. Economic data is mixed using technical indicators, daily moving averages and returns, rolling window standard deviations, and others. The model predicts bullish or bearish movements by comparing accuracy, precision, recall, and F1 score to specified benchmarks. The Grid search model exceeded the baseline model in precision and accuracy with a Reset Measured dichotomized score of 98.91%. The persistence of negative Sharpe ratios advises "do not invest." This is because certain fund patterns perform poorly in risk-adjusted returns. The study shows that financial industry operating and forecasting systems require data mining and machine learning to improve accuracy. This method for finding hidden financial data links may help investors amid market turmoil.

Keywords: Mutual Funds, Trend Analysis, Association Rule Mining, Classification, Investment, Machine Learning, Stock Market, Financial Forecasting.

1. Introduction:

Capital markets evaluate economic development, decide corporate investments, and foster financial innovation. Because individual and institutional investors desire wealth building and diversification, mutual funds and shares are hot commodities. They facilitate securities trading and reissuance. IPOs raise funds, and traders exchange securities in the secondary market. In India, BSE and NSE offer the majority of trading services, while SEBI regulates these services. Stocks and mutual funds are volatile and non-linear, making forecasting difficult even in a controlled environment [1]. Each market is complex, therefore one price is

established by investor perception, income ratios, foreign market bonds, and personal judgement. Past performance, risk, and benchmarks impact mutual fund investor behaviour.

A novel technique uses association rule mining and Random Forest machine learning to anticipate mutual fund movements using historical fund data and macroeconomic factors. Data is collected from reputable sources using APIs like yfinance, normalised and filled in missing values, association rule mining using the Apriori algorithm, and predictive modelling with a Random Forest classifier. To find relevant correlations between mutual fund economic and predictive performance. Association rule mining helps us understand fund behaviours under particular economic situations and uncovers co-occurrence links that traditional statistical methods missed. A clear investing approach is "if GDP growth is high and inflation is low, returns on fund A are likely to go up". These rules are characteristics of a Random Forest model that predicts mutual fund return uptrends and downtrends.

The model is being trained, validated, and optimised, including Grid Search hyperparameter tuning using cross-validation. Evaluation uses accuracy, precision, recall, and F1-score. To compare the forecast to historical data, backtesting have done. Association rule-derived characteristics increase predicting performance significantly. The suggested technique uses association rules and machine learning to assess trends, create investment strategies, and explain market occurrences. These results demonstrate the significance of data-based finance and how better business operational strategies may reduce investment market volatility.

2. Related Work

Machine learning has improved financial forecasting accuracy and flexibility over previous methods. SVM, Random Forest, XGBoost, and PCA have been used to forecast stock prices and market patterns by several authors. In [2], the author found that ML models beat all traditional forecasting methods using 2,448 days of stock price data and 13 financial variables. The combination of XGBoost with PCA improved classification performance, showing how feature selection may improve prediction accuracy. Another work [3] used SVM, ANN, and Logistic Regression to forecast stock price movement for Ho Chi Minh Stock Exchange (HoSE) VN30 index shares. SVM outperformed ANN and Logistic Regression with 92.48% accuracy. SVM enhances financial market prediction dependability and accuracy, making it ideal for algorithmic trading in high-dimensional datasets. The stock price prediction, ML has been applied in investment forecasts and networked financial system analysis. Ensemble learning and data clustering improved retrieval of investment data from 19,577 people in [4]. Clustering algorithms have used to classify investors by financial characteristics like active vs passive investors and risk appetite to provide individualized investment solutions.

Dynamic Asset Graphs [5] enhance financial forecasting using machine learning–based feature extraction. The DAX30, EUROSTOXX50, FTSE100, NASDAQ100, and NIFTY50, indicative of key worldwide markets is used. The benchmarks used correlation matrix computations without modifications, while network-based models increased predicted accuracy by 40%. These findings show that graph-based financial models may increase market comprehension, risk assessment, portfolio optimization, and market insight. Deep learning algorithms using financial data are increasingly used to forecast stock and mutual fund performance. Long Short-Term Memory (LSTM) networks, Artificial Neural Networks (ANN), Convolutional Neural Networks (CNN), and Bidirectional LSTM (BiLSTM) networks often achieve R-squared values above 0.5 for stock market time series forecasting [6]. This indicates significant predictive accuracy improvements LSTM beat conventional forecasting in two-decade tests of over 1200 mutual funds, but GRU varied just modestly.

3. Proposed Methodology

3.1. Dataset Description

The five mutual funds VFIAX, FXAIX, VIGRX, VTMGX, and VTSAX is used in our study. Starting from May 4, 2011 and up until December 31, 2024, we obtained their daily closing prices. These multi-sector mutual funds include large-cap equities, growth and international funds, as well as a total market index fund. With Yahoo Finance being a renowned financial data hub, we pulled the data using the yfinance Python library. For the sake of this study, we only considered the closing price of each trading day, as is customary in most investing research. Data is cleaned and standardized before analysis. A total of 3,438 observations have obtained for each mutual fund which ensured uniformity in date range across all funds. Fund closing prices are rescaled using Minmax normalization to eliminate Fund closing prices is rescaled using Minmax normalization to eliminate scale discrepancies [7]. The Apriori approach and a threshold of 0.5 are used in order to transform the data into binary during the rule extraction process. The dataset has been expanded to include market simulation and macroeconomic parameters in order to enhance the financial relevance and trend prediction modeling capabilities. During the process of dataset analysis, advancements have been made in the areas of association rule mining, investment decision scoring, trend forecasting, and assessment of Random Forest machine learning models.

3.2. Data Collection and Preprocessing

In the dataset pertaining to mutual funds, the macroeconomic parameters that were taken into consideration indicate the deep pattern of the overall economy, which is comprised of a composite economic system. The macroeconomic determinants include the rate of growth of the gross domestic product (GDP), inflation, stock returns, interest, and employment [8]. When it comes to understanding the operation of the mutual fund in relation to the macroeconomy as well as the behavior of investments during times of economic difficulty, these aspects are quite essential. There is a good chance that the use of those components would result in an improvement in the study findings about the interconnections between macroeconomic and financial variables. In spite of the fact that every part of coverage makes an effort to prevent overlapping, it is possible to note that this method of storing information makes it possible to include it into a framework that does not need systemic analysis. When information is kept in an organized manner, the job of organizing and processing is simplified and expedited, which results in a more efficient workflow. It is of the highest necessity to correctly synchronize the time-series data in order to ensure overall dataset integrity and consistency across the series while conducting cross-company based prediction utilizing an associative rule mining and predictive modeling framework for fund trend analysis. This analysis is performed in order to analyze fund trends.

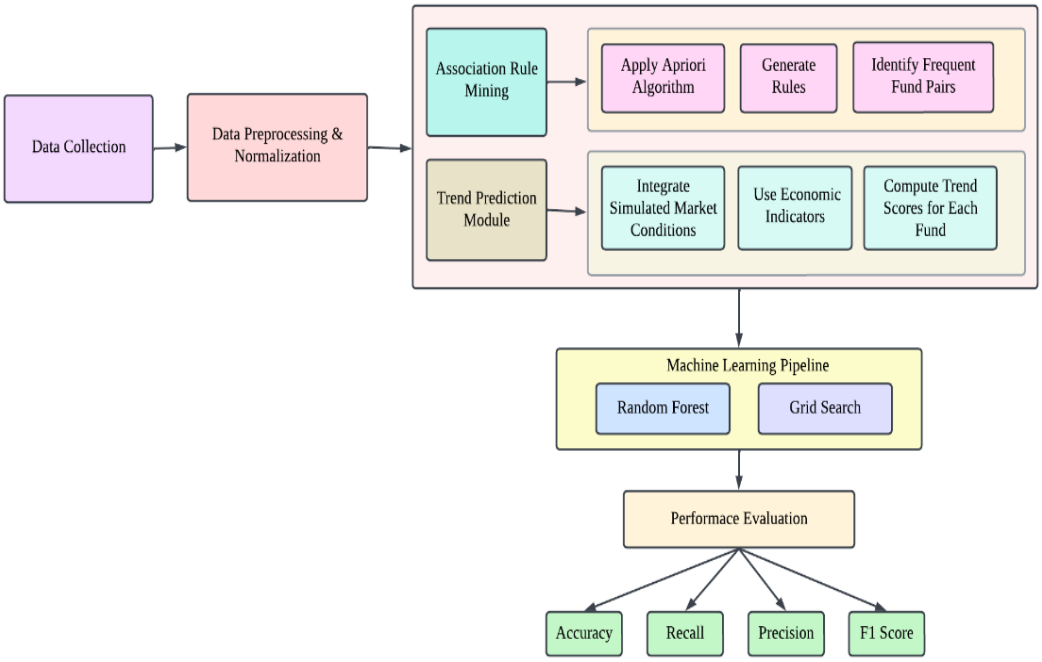

Fig.1: Mutual Fund Trend Analysis Framework

3.3. Association Rule Mining

One of the most used methods for data mining is called association rule mining, and it is used to identify associations in previously unknown data that is included inside massive databases. The study of the linkages between different assets or the analysis of how major participants in the financial market, such as mutual funds, react to changes in macroeconomic conditions are two ways that this may be accomplished in the field of finance. Through the use of association rule mining methods, the purpose of this research is to identify the behavioral patterns of mutual funds in relation to the current economic circumstances and market cycles which are in effect. There is a possibility that the conceivable interdependencies, particularly those pertaining to specific mutual funds and economic or market developments, might be significantly boosted by some exercises that are loaded with concealed classical analysis. It is via this method that investors would be able to predict the activity of mutual funds with a better degree of accuracy and construct adaptable plans based on the primary choices they make about their investments [9]. These methods would not only make optimum portfolio management easier, but they would also significantly boost return on investment in relation to the risk that have incurred. The term "support" will be used in this context to denote that support as a measurement would include counts of the occurrence of an itemset which is included inside a specific dataset. Support is mathematically defined as show in equation (1).

$$\text{Support}(A \rightarrow B) = \frac{\text{Count}(A \cap B)}{\text{Total Transactions}} \quad \ldots(1)$$

A and B represent the assessed entities, such as mutual funds or economic indices. Upon identifying the frequent itemsets, various association rules will be established, such as "if FXAIX exhibits strong performance, then VFIAX is likely to exhibit strong performance as well." These regulations are assessed using indicators such as confidence and lift. In this context, confidence is determined in equation (2) and lift is computed in equation (3).

$$\text{Confidence } (A \rightarrow B) = \frac{\text{Support}(A \cap B)}{\text{Support}(A)} \quad \ldots(2)$$

$$\text{Lift}(A \rightarrow B) = \frac{\text{Confidence } (A \rightarrow B)}{\text{Support } (B)} \quad \ldots (3)$$

The binarization of normalized data before dataset Association Rule analysis Mining is a crucial phase in the process. The Apriori algorithm, the principal approach for developing association rules, operates only on rigid category or binary input data, necessitating that normalized values be presented in a binary format. This is accomplished by establishing a threshold, often at 0.5, wherein data over the threshold are normalized to 1, while those falling below are assigned a value of 0. A fund's net asset value (NAV) is reset to 1 if it achieves a certain threshold value that might be considered a successful day. These achievements are quite remarkable. Depending on the specifics of the situation, it will immediately return to 0. When it comes to guiding criteria, labeled ranges make use of a technique known as the Apriori approach. To guarantee that the algorithms acquired from the m... data are not arbitrary outliers that distort performance-based metrics, but rather are useful and relevant, these criteria ensure that the algorithms are obtained. It is crucial to make such modifications in order to find qualitative trends that have the potential to affect important investment choices within the confines of a set framework.

3.4. Trend Prediction

Recognizing patterns is critical in carrying out financial research since it helps one anticipate how a mutual fund will perform in future accurately. This lays foundation for more advanced machine learning models because it adds value to fund performance relative to market expectations. This research attempts to model the impact of some economic drivers like inflation, interest rates, and growth of GDP on the performance of mutual funds. The objective is to equip analysts with better understanding of the adverse impacts from macroeconomic conditions. The construction of the model as a multiplicative network of equations that explain the behavior of the fund in relation to external factors such as economic activity is a complicated procedure that requires a significant amount of computing [10]. An example of mutual fund performance evaluation is given by the performance indicator P_i which encompasses the earnings components R_i of the fund, the inflation rate μ, interest rate r, and GDP growth G, all serving as macroeconomic constituents. The equation will incorporate all variables that will be obtained from the equation (4).

$$P_i = \alpha_1 \cdot R_i + \alpha_2 \cdot \mu + \alpha_3 \cdot r + \alpha_4 \cdot G \quad \ldots (4)$$

α_1, α_2, α_3, and α_4 are the equations that is derived from the process of backcasting historical data. These equations are a representation of the original logic that has been produced in order to construct advanced modeling techniques with the purpose of predicting the performance of funds. For the purpose of predictive modeling, it is of utmost significance to correctly replicate those actual market circumstances. Through the building of fictional economies that are intended to isolate the effect of a macroeconomic parameter, the research assures that the evaluation of mutual fund autoepistemic self-assessment may take place inside a controlled environment for the purpose of experimental control. The increase of the gross domestic product (GDP), inflation, interest rates, and unemployment are all examples of important economic indicators that may be qualitatively manipulated in an environment that is so tightly regulated. The Geometric Brownian Motion (GBM) model, stochastic models, and other similar models are responsible for defining these variables [11, 12]. It is possible to define the growth of GDP G(t) at time t as given in equation (5).

$$G(t) = G_0 \cdot e^{\left(r - \frac{\sigma^2}{2}\right)t + \sigma W_t} \quad \ldots (5)$$

In the context of these parameters, the variable G_0 represents the growth rate of the gross

domestic product (GDP), r is the projected increase in the GDP growth rate, σ represents the volatility, and W_t is the Wiener process that represents random departures from the norm in the economy. The impetus that is supplied makes it possible to simulate realistic changes in macroeconomic time series indicators, which in turn makes it possible to approximate as closely as possible the actual market conditions. A mutual fund's performance is monitored via the use of profile review, which takes into account a variety of shifting economic situations. As a result of the great precision with which these models explain the performance of mutual funds, they contribute to a better understanding of the values of the funds under a variety of economic situations.

3.5. Random Forest Modeling

First and foremost, the Random Forest Model method begins with the process of feature extraction. Data that is considered basic, such as the historical price data of mutual funds, is turned into measures that are more relevant at this stage. Examples of key benchmarks include the daily return, which is a measurement of the daily change or shift in the price of the mutual fund. This is one of the basic benchmarks. Particularly noteworthy instances of the kinds of risks that are evaluated by this specific benchmark include the fund's volatility as well as its movement in a certain direction. Indicator binary rule mining, five and twenty day moving averages, and rolling standard deviations are some of the other technical indicators that are included in the feature set. These indicators are formed via the use of other components. When such characteristics are included into the assessment of the fund, the quantitative evaluation is improved since these components indicate the fund's performance over both the short term and the long term. Lagged variables are another one of the characteristics that might be introduced to the model. These variables help with temporal dependency reasoning and enable the model to make predictions based on the outcomes of previous experiments. These characteristics, when combined with a strategy that takes into account market momentum, make it possible for the model to include economically driven changes, volatile cycles, and flexible modifications that are relevant to the circumstances. It is possible that the model's capacity to anticipate the complexities that are entwined with the performance of the fund might be improved by expanding the number of characteristics that are collected.

A classifier that was constructed on the basis of the Random Forest method, which allows for the construction of features, was constructed after the data were properly evaluated. An ensemble learning technique known as Random Forest takes a number of decision trees and merges them into a single model. The results of these trees are then averaged to achieve a higher level of accuracy. The objective of the classification issue, which is logically binary in this instance, is to determine if a certain mutual fund will meet or fall short of expectations within a particular time period. This evaluation is intended to provide an accurate appraisal of the situation. A training set, which accounts for 80% of the dataset, and a test set, which accounts for 20% of the dataset, are the two portions that make up the dataset. This split percentage from the entire data makes it possible to validate and predict the data while also guaranteeing that the procedure does not make use of any information that is already known. The unpredictability of the model is boosted by the random training of subset characteristics for each forest. This helps to lower the likelihood of the model overfitting while simultaneously enhancing its capacity to recognize a wide variety of patterns in the raw data. The prediction in the random forest model is accomplished by voting, with the ensemble's prediction being the option that receives the most votes. The result will be determined by the vote that received the most votes with the majority. The accuracy and generalization of the model are both improved as a result of this, as well as the difficulties with individual tree bias. The model that created is resilient to irrelevant data, and as a consequence, the Random Forest classifier may be used to forecast mutual funds in

complicated financial systems. This is due to the fact that sophisticated irrelevant data will not influence the findings.

3.6. Performance Evaluation

It seems that assessment is derived from the same root as evaluation, which emphasizes the performance reasoning of a model. For instance, consider the models used in economics. One of these is making forecasts on the movement of mutual funds. The model of the mutual fund is able to provide accurate projections on the path that the fund will take in the future. For example, trend forecasting may be considered a sort of binary classification issue. In this type of problem, the model seeks to identify whether the performance of the mutual fund will improve (up) or deteriorate (down). The most accurate forecasts are generated by the model on the basis of certain parameters, one of which is that the average return is taken into consideration to be zero. Between growth and decline, which is referred to as the threshold, growth is differentiated. Using the threshold as a reference point, forecasts and results are brought into alignment, which provides possibilities for evaluation that are adaptable. The major purpose of this evaluation is to ascertain whether or not the model is dependent purely on metadata, as well as whether or not it is capable of providing deeper insights and forecasts, hence exposing underlying patterns throughout variations.

Determine the baseline orientation that one desires to pursue, and then check to see whether the results have been historically pluralized. This is the first step in the validation process. Precision, recall, F1 score, and accuracy are some of the statistical benchmarks that are chosen for review in order to assess the model. Multiple relevant statistical standards are also picked for examination. In accordance with its definition, precision is defined as the ratio of the total number of projected upward trends that actually occurred to the total number of projected upward trends. The efficiency of the model in relation to actual rising trends is evaluated by recall, which does so by computing the fraction of precisely anticipated upward movements. By definition, F1 is a statistic that is used to evaluate the accuracy of a mean. In this metric, precision and recall are weighed against false positives and false negatives. In order to identify the equilibrium between the best fetch and the average regard, it computes the middle ground between the two parameters. The performance of the model is evaluated in connection to the expectation on the judgment about both positive and negative predicted movements in comparison to the reference point that is established around the aim. For the purpose of establishing if the model has a tendency toward over-optimism or over-conservatism in presenting trends, the characteristics that have been discussed above, when combined, provide a full and balanced account of the performance of the model. This is the mathematical expression that may be used to represent these metrics in equation from (6) to (9).

$$\text{Precision} = \frac{TP}{TP+FP} \quad \ldots(6)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad \ldots(7)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad \ldots(8)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad \ldots(9)$$

Where:

$TP =$ True Positives (correctly predicted upward trends),

$TN =$ True Negatives (correctly predicted downward trends),

$FP =$ False Positives (incorrectly predicted upward trends),

$FN =$ False Negatives (incorrectly predicted downward trends).

These formulas help quantify the performance of the model and guide decisions on how to refine the model for better accuracy and reliability.

4. Results and Discussion

4.1. Association Rule Mining Results

During the process of applying the Apriori algorithm to the price data of mutual funds, significant connections had been discovered. "If VFIAX is high, then FXAIX is high as well," is one of the principles that stood out because of its significance. The support for this scenario is 29.52%, the confidence level is 99.90%, and the lift is 3.38. Given that both of these funds are included in the S&P 500 index, it is reasonable to assume that there is a close connection between them, which may have been the reason for the establishment of this criterion. Following more research, the following rule was discovered: "If FXAIX is high, then VFIAX is high too," which claimed a confidence level of 100% and lift of 3.38; and "If VIGRX is high, then VFIAX is high," which claimed a confidence level of 98.83% and lift of 3.34. Both of these rules were proposed after further analysis. This suggests that there is a large cross-correlation across different time periods, which may be helpful for the strategic planning of investment investments and risk management as we can see in table 1. This suggests that having a better grasp of the connection would be beneficial in terms of appropriately structuring the portfolio and influencing the choices that the investor makes about investments.

Table 1: Key Association Rules Show Mutual Fund Interrelationships

Antecedent	Consequent	Support	Confidence	Lift
VFIAX	FXAIX	0.2952	0.9990	3.38
FXAIX	VFIAX	0.2952	1.0000	3.38
VIGRX	VFIAX	0.2464	0.9883	3.34

4.2. Trend Prediction Scores

Calculated expectations for mutual funds that take into account both market indicators and theoretical models are "scored" based on how well they do in the scoring system model because of their performance. With a score of 60.68, the VFIAX fund has the potential to achieve the trend position that is the most trending. This suggests that the fund has a high level of confidence in its ability to outperform its competitors. Following in second place with a score of 22.78 was FXAIX, followed by VIGRX with a score of 20.97, VTSAX with 16.25, and VTMGX with a score of 2.72. Within the context of the conversation simulation, these numbers are relevant to the discourse that is taking place concerning the metrics of interest. The VFIAX surpassed other indices, possessing a quasi-competitive market, the Sharpe ratio, which is characterized as the performance indicator, weakens capital volatility. As a result, stagnation seems to be unavoidable, and advancement is funneled as a result of this development. To put it another way, stagnation is quite probable in given table 2. The presence of volatility seems to be a funnel that is preventing development.

Table 2: Mutual fund trend scores in simulated market conditions

Mutual Fund	Predicted Trend Score
VFIAX	60.68
FXAIX	22.78
VIGRX	20.97
VTMGX	2.72
VTSAX	16.25

4.3. Investment Recommendations

According to the findings of the research, none of the funds are able to satisfy the needed threshold since the returns they generated have not sufficient for the amount of risk they took; all of the Sharpe ratios is negative. Growth mutual funds is put in the "Do Not Invest" category due to the fact that their raw trend ratings are rather high. The already bleak investment prospects is further muddled by the inconsistent and erratic ratings of these funds, which further categorized these funds as "Do Not Invest." Despite the fact that it maintained its position with a raw trend score of 60.68, VFIAX also carried a Sharpe ratio of -1.78, which resulted in an adjusted score of -69.29 instead. The remaining funds had comparable performance, with FXAIX, VIGRX, VTMGX, and VTSAX behind significantly due to markedly disappointing risk-adjusted metrics. Despite criticism over their aggressive strategy based on excessively high foresight ratings, they ultimately called for a cautious position, targeting individuals who established benchmarks below the investment barrier. The data clearly indicated the advantages of use relative performance metrics rather than absolute ones in the construction of contentious portfolios as given in table 3.

Table 3: Trend Scores Adjusted by Sharpe Ratios Investment Advice

Fund	Trend	Sharpe	Adjusted Score	Recommendation
VFIAX	60.68	-1.78	-69.29	Do not invest
FXAIX	22.78	-1.78	-26.02	Do not invest
VIGRX	20.97	-1.56	-18.27	Do not invest
VTMGX	2.72	-1.84	-3.29	Do not invest
VTSAX	16.25	-1.75	-17.83	Do not invest

4.4. Model Evaluation Metrics

An in-depth analysis of model performance uncovers a significant disparity between conventional baseline models using threshold-based methods and Random Forest models driven by machine learning, which use sophisticated algorithms. The baseline adheres to neural network architectures. Trend recall remains impeccable at 1.0, reducing precision to 0.467 results in an overall model accuracy decline to 46.74%. It inaccurately exaggerates negative trends misidentified as positive. The ability to claim unfed animals stands at 46% positivity, with several good outcomes after erroneous and negative lots and experiences. The automated refinement method being disabled, the Random Forest classifier demonstrated remarkable proficiency in correct estimate, with an accuracy of 98.83%, with precision, recall, and F score of 0.97, 0.992, and 0.989, respectively. The implementation of Grid Search for hyperparameter optimization, the model achieved an accuracy above 98.91%, with precision and recall values of 0.989 and 0.998, respectively, establishing it as a highly trustworthy standard. 0.991 As demonstrated, machine learning models, particularly random forest, provide a significantly more reliable and effective framework for analyzing mutual funds and predicting trends in investment analysis compared to simplistic rule-based approaches, which are inferior tools for assisting human analysts. We can see the confusion metrics in figure 2 and result show in table 4 as well as in figure 3.

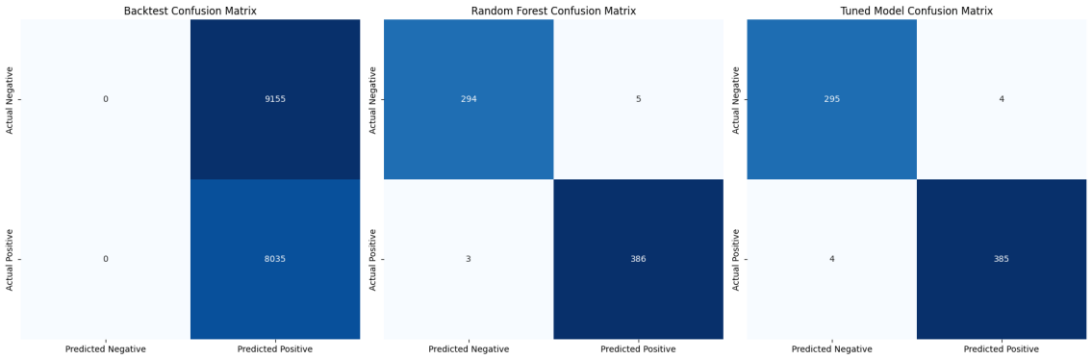

Fig. 2: Performance Confusion Metrics

Table 4: Comparison of Baseline and Random Forest Trend Prediction Models

Model Stage	Accuracy (%)	Precision	Recall	F1 Score
Backtest	46.74	0.467	1.00	0.637
Random Forest	98.83	0.97	0.992	0.989
Grid Search Tuned	98.91	0.989	0.998	0.991

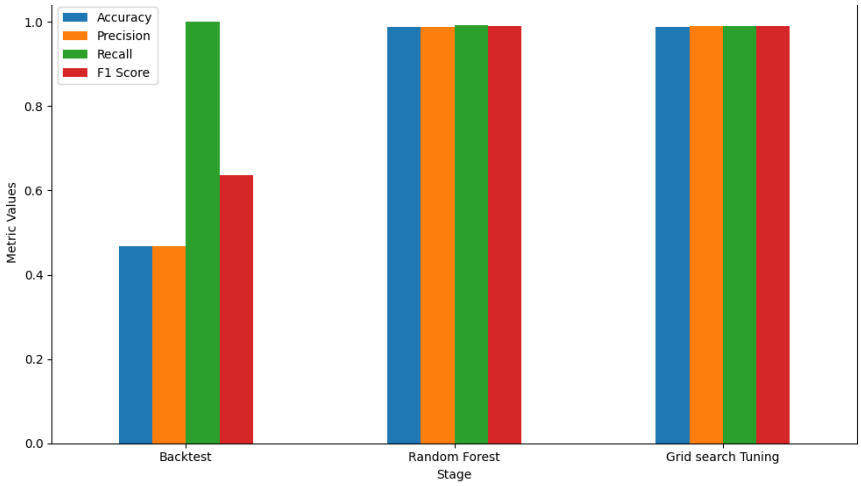

Fig. 3: Model Performance Metrics Across Stages

4.5. Comparison with Existing Work

To assess our proposed Random Forest-based stock trend prediction model, we compare its performance with many previous research using different data and machine learning methods. After doing hyperparameter tuning by grid search, our model achieved an accuracy of 98.91%, significantly exceeding the performance of prior models. The study using sentiment analysis of news and Twitter with technical indicators achieved an accuracy of 90.14%, whilst another study utilizing ARIMA, XGBoost, LSTM, and FLANN attained an accuracy of 98.89%. The application of SVM on VN30 index stocks yielded an accuracy of 92.48%, while RNN models including social media sentiments achieved an accuracy of 86.06%. The bottom row of Table 5 demonstrates that our model attained superior accuracy, with elevated precision (0.989), recall (0.998), and F1-score (0.991), signifying exceptional capability in forecasting market behaviors and effectively generalizing to novel data. These results demonstrate the distinctiveness of this study compared to previous studies, mostly attributed to its exceptional efficacy in spotting market trends.

Table 5: Comparison with Existing Works on Stock Trend Prediction

Ref.	Techniques Used	Accuracy (%)	Notable Features
[13]	Ensemble methods + Explainable AI	90.14	Multi-source data integration, explainable results
[14]	Combined model	89.80	Emphasis on minimizing error ratio
[15]	Pairwise classification vs traditional models	>84	Outperformed five traditional techniques
[3]	SVM, ANN, Logistic Regression	92.48	SVM effective for high-dimensional financial data
[16]	StockSentiWordNet, ELM, RNN	86.06	Use of social media sentiment in trend prediction
[17]	Hybrid (ARIMA + XGBoost + LSTM + FLANN)	98.89	Highly accurate hybrid time-series predictor
Our Proposed Work	Random Forest + Grid Search	98.91	High F1-score (0.991), Precision (0.989), Recall (0.998)

5. Conclusion

This research introduces an innovative predictive model for mutual funds by combining Association Rule Mining with Random Forest classification to enhance accuracy and predictive capability. It utilizes historical data from the funds and some simulated macroeconomic variables, elucidating the complex interdependencies between a fund's performance and the current economic circumstances that many conventional models cannot capture. The mutual funds exhibited a notable co-movement pattern, as seen by the Grid search, while the apriori method generated valuable information about co-movement via association rule mining. The amalgamation of these techniques enhanced the model's performance to an accuracy of 98.91%, far surpassing other approaches. The aggregated accuracy, recall, and F1-score of these measures demonstrated significant results compared to the previous rigid models that depend largely on thresholds, which the suggested model surpassed thoroughly. Despite the model's precise result predictions, all mutual funds had negative Sharpe ratios, indicating underperformance for return relative to risk in the simulated market circumstances. This led to no funds meeting the favorable recommendation standards established by the investing criterion. This outcome highlights the need of reconciling trend projections with risk indicators for thorough financial decision-making. This method demonstrates that the integration of data mining with ensemble machine learning enhances the precision of mutual fund trend predictions, hence facilitating improved investment decision-making. Future study may focus on integrating real-time data streams and investor sentiment analysis, as well as applying the approach to other financial assets. In a rapidly changing investment environment, these enhancements might improve the model's stability and utility.

References:

1. Vasdani, G. K., & Thakrar, N. (2024). A Comprehensive Study Of The Impact Of Foreign Portfolio Investments On Indian Stock Market.
2. A. Sun, (2024). Stock trend prediction and analysis based on machine learning. Highlights in Science, Engineering and Technology, 92, 165–170.4.
3. Khoa, B. T., & Huynh, T. T. (2022). Forecasting stock price movement direction by machine learning algorithm. International Journal of Electrical and Computer Engineering (IJECE), 12(6), 6625–6634.

4. Kaewkiriya, T., & Wisaeng, K. (2023). Development of customer predictive model for investment using ensemble learning technique. Journal of Computer Science, 19(6), 775–785.
5. Castilho, D., Souza, T. P., Kang, S. M., Gama, J., & de Carvalho, A. C. P. L. F. (2024). Forecasting financial market structure from network features using machine learning. Knowledge and Information Systems, 1–29.
6. Liu, F., Guo, S., Xing, Q., Sha, X., Chen, Y., Jin, Y., Zheng, Q., & Yu, C. (2024). Application of an ANN and LSTM-based ensemble model for stock market prediction. In 2024 IEEE 7th International Conference on Information Systems and Computer Aided Education (ICISCAE) (pp. 390–395.
7. Lawi, A., Mesra, H., & Amir, S. (2022). Implementation of Long Short-Term Memory and Gated Recurrent Units on grouped time-series data to predict stock prices accurately. Journal of Big Data, 9(1), 89.
8. Agu, S. C., Onu, F. U., Ezemagu, U. K., & Oden, D. (2022). Predicting gross domestic product to macroeconomic indicators. Intelligent Systems with Applications, 14, 200082.
9. Kumar, S., Srivastava, M., & Prakash, V. (2023). Challenges and opportunities for mutual fund investment and the role of industry 4.0 to recommend the individual for speculation. New Horizons for Industry 4.0 in Modern Business, 69-98.
10. Tian, G. (2024). Numerical simulation of financial fluctuation period based on non-linear equation of motion. International Journal of Computer Applications in Technology, 74(3), 171-185.
11. Lee, S. L., Liew, C. Y., Chen, C. K., & Voon, L. L. (2022). Geometric Brownian Motion-Based Time Series Modeling Methodology for Statistical Autocorrelated Process Control: Logarithmic Return Model. International Journal of Mathematics and Mathematical Sciences, 2022(1), 4783090.
12. Mensah, E. T., Boateng, A., Frempong, N. K., & Maposa, D. (2023). Simulating stock prices using geometric Brownian motion model under normal and convoluted distributional assumptions. Scientific African, 19, e01556.
13. Ranjith, J. (2024). Enhancing Stock Market Trend Prediction Using Explainable Artificial Intelligence and Multi-source Data. Fusion: Practice & Applications, 16(2).
14. Jishag, A. C., Athira, A. P., Shailaja, M., & Thara, S. (2020). Predicting the stock market behavior using historic data analysis and news sentiment analysis in R. In First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2019 (pp. 717-728). Springer Singapore.
15. ATLI, A. H. (2023). PREDICTION OF STOCK MARKET INDEX MOVEMENT USING PAIRWISE CLASSIFICATION. Economic Computation & Economic Cybernetics Studies & Research, 57(2).
16. S. Albahli, A. Irtaza, T. Nazir, A. Mehmood, A. Alkhalifah, and W. Albattah, "A machine learning method for prediction of stock market using real-time twitter data," Electronics, vol. 11, no. 20, p. 3414, 2022
17. G. Goverdhan, S. Khare, and R. Manoov, "Time Series Prediction: Comparative Study of ML Models in the Stock Market," 2022.