Advances in Consumer Research
ISSN (Print): 0098-9258
ISSN(Online): 3079-1766

Vol. 2, Issue 2 (2025) https://acr-journal.com/

# AI-Driven Forecasting and Classification of Investment Strategies in International Equity Markets

**N.Rajeswaran**[1]

[1]School of Computer Applications, IMS Unison University, Dehradun, Uttarakhand,India

**Cite this paper as:** N.Rajeswaran, (2025) AI-Driven Forecasting and Classification of Investment Strategies in International Equity Markets. *Advances in Consumer Research*, 2 (2), 1094-1105

**KEYWORDS**

*Machine Learning, Investment Strategies, Global Financial Markets, Predictive Modeling, Classification Algorithms.*

**ABSTRACT**

Global financial markets have dynamic nature that requires highly sophisticated strategies to predict and categorize investment strategies that can maximize returns, minimizing risks. The current study examines how machine learning algorithms can be used as a method of predicting and classifying investment strategies in international stock markets. As part of the training the study uses a detailed dataset with information on historical market prices, macroeconomic variables and technical analysis indicators. The effectiveness of key algorithms (Decision Trees, Random Forest) to predict the movement of asset prices and the classification of investment strategies (examples: growth, value, momentum and defensive investment strategies) is tested. The evaluation of the performance of these models would be based on accuracy, precision, recall and F1-score and the results are indicative of the possible application of machine learning in the decision making process by investors. With this study, more sources on application of machine learning in the area of finance are created, and the results can be valuable to investors and a financial expert that wants to utilize tools based on AI technologies to make an optimal choice of their strategy considering fluctuation of the world market situation.

## 1. INTRODUCTION

The financial market across the world is very complicated and volatile hence posing a great challenge to investors who want to maximize their returns and also reduce their risks as well. Fundamental-driven, technical and market-sense based investing are traditional approaches and they have been commonly applied in decision-making. Nonetheless, as the financial markets are evolving and transitioning into the data-rich environment, such traditional approaches may be unable to meet the needs of the markets dominated by the high volatility and constantly growing inflow of information. It has resulted in the birth of newer and more advanced methodologies like machine learning that has the potential to enhance predictions and classifications of investment strategies.

ML algorithms, due to their capabilities of processing a large volume of data and discovering intricate patterns in it, are rapidly being used in the financial industry in various applications: forecasting the price of stocks and optimizing the portfolio, and risk prediction, etc. The possibility of using previous market trends, economic and market symptoms coupled with market emotions to forecast on how the market will behave in future has brought up the development of models that can assist in categorizing investment techniques that will maximize yield considering disparities of the market. Such strategies are commonly classified as various types of strategies that include growth, value, momentum and defensive investing.

This research, which is discussed in this paper, will explore the possible ability of machine learning algorithms in the prediction and classification of investment strategies in global financial markets. This analysis will find out the most effective models to predict the fluctuations in asset prices and prescribe appropriate investment strategies through the available variety of models: Decision Trees, Random Forest and Deep Learning models.

Such predictive capabilities are especially significant in the wide-open and uncertain business environments of the present times when outdated approaches do not always guarantee up-to-date and correct data and insights. As an example, the machine learning models can be used to not only analyse the past data in the market, but also the macroeconomic events, political developments and even sentiments in the social media. This will allow the investors to build more sophisticated and dynamic investing tip that adjusts itself to the shifts in the market conditions happening in real time.

Furthermore, with the ongoing globalization of financial markets, the use of machine learning tools to analyze the cross-market opportunities and divide the investment strategies is quite competitive. These models enable investors to diversify their portfolios, balance the risk and optimise their asset allocations according to relevant predicitve information that transcends orthodox methods.

This study has three research objectives, first to determine how well different machine learning models will predict the direction of prices of assets and second, to analyze how effective these models will be in classifying investment strategies according to the market conditions, and thirdly, to understand the implications of the research to the practice of investors and financial analysts in a comprehensive manner. In such a manner, the current study also benefits a larger field of Finance Technology (FinTech) and it can serve as a valuable source of information to market players looking to implement advanced approaches to machine learning into their investment-making practices.

This introduction establishes the context of further discussion of the role of ML application in the global financial markets to explain that this approach could be applied to select the most effective investment strategies using data findings and predictive power. The outcomes of the research can transform the way investors develop their strategies in the age characterized by the growing data availability and complex markets.

## 2. BACKGROUND

Brofos, J.A (2024) paper is concerned with a portfolio trading algorithm based on machine learning algorithms by investigating how effective they could determine the stock index based on the given returns rate: either positive or negative, then they could create a portfolio allocation that could be effective. Its plan is based on the time series analysis, which makes use of technical parameters in creating binary classification judgements concerning future periods. This model is based on an ensemble of random forest classifier, non-linear support vector machine classifier, relevance vector machine classifier as well as the k-nearest neighbor classifier. The model is tested in various sectors of investment including Energy, Materials, Financials and Information Technology, with the help of data between 2006 and 2012. When it is used to forecast the price returns of a stock in three months the forecast accuracy of the model comes to about 70 %.

The article by Shapiro, D (2020) concerns the problem of classifying small businesses, especially in the customer segmentation, where the bias on gender and geographic origin are injected during the training on business names. The model built based on a data in the study aims at predicting the kind of businesses based on the name of the business alone with a top-1 F1-score of 60.2 percent. Two methods are investigated to make it less biased: replace given names with a neutral token and expand the training set with swapped gender. This experiment demonstrates hiding given names decreased bias, but at the cost of a decrease in classification performance (top-1 F1-score = 56.6 %). The method of increasing the training set with gender-swapped samples was not as effective as the name-hiding option in eliminating the bias on the test data set.

Lamponi, D (2014) article looks at the validity of industry labels adopted by portfolio managers, risk managers and asset allocators since these labels rival other data-based classification of large US stocks. It can be seen that the categorization based on mere stock price time-series, is data-driven and that the results have a consistent overlap with the traditional industry classification and thus the decisions made by the market players.

Misra, P.K et al., (2024) explores forecast the end of the year share price of the Australian mining companies using the relationship of different financial measures and the share price, especially of those firms that are experiencing financial distress. Indicators which are critical to look at include the BV per Share, Enterprise Value and Cash per Share and financial ratios like ROE and ROA. The data to be used in the study consists of 395 rows and 33 columns, with the Linear Regression fitting attaining a value of R-squared = 0.90, with the XG Boost model attaining a value of R-squared = 0.91. The findings indicate that XG Boost can be widely used to predict the share prices, and key contributions will be made to the financial decision-making and investment strategies in the turbulent market environment.

Yuan, M (2024) examines how machine learning algorithms such as random forest and IC analysis can be used to optimize quantitative strategies applied in investments. The study enhances the quality of investment decisions by integrating multifactor approaches together with machine learning. Selection of factors to be used in linear regression models occurs in the random forest algorithm, and future 30-day returns the dependent variable. The models that are obtained are used to forecast stock returns and form portfolios of investment. The results indicate that the combinations in strategies yield positive returns and their performances are quite satisfactory when it comes to dealing with volatility and risks, and that they are less successful than others are.

Meshcheryakov et al (2025) explore the applicability of ML to forecast the outcomes of the commodity markets in the conditions of the Nash equilibrium, when market players are motivated to make utmost profit out of limited information. The process of Nash equilibrium search algorithm is utilized to construct a synthetic data set which will be used to train a model by synthesis based on game theory and economic modelling to comprehend the nature of competition and strategic interaction among the players. The findings demonstrate the application potential of such approach in the work with market and investment strategies, efficient subsidy programs and other spheres.

Giorgi, F et al., (2023) introduces an algorithm in reinforcement learning (RL) to produce a trading strategy that takes into consideration the effects of transaction costs and assets dynamics influences. The algorithm is initially compared to analytical solution to optimal strategy which is applicable in linear factors and quadratic transaction costs and it is found to duplicate optimal strategy. The paper then addresses a more realistic non-linear dynamics situation that is more applicable in case of time series of WTI spot prices. What to do in a case when the best decision is not known? RL can serve as the option here. The findings highlight that the RL agent significantly performs better than a trader based on a linearized model in order to implement the theoretical optimal strategy on synthetic WTI spot price data.

Alparslan, S., and Uc ar, T (2023) provides answers that discuss the efficiency of the machine learning algorithm with regard to forecasting the price of commodities specifically during the COVID-19 period, which triggered an economic slackening and recession. Commodity products such as gold, silver and metals are the centre of interest because they are reckoned as secure investment amidst economic crisis. The dataset utilized by the study is Borsa Istanbul data, the data covering July 2018-October 2021 which includes the price of daily data on both gold and silver, separated by time (before COVID-19 and during COVID-19). Its ability to predict volatility in prices on such exceptional circumstances is measured using metrics like MAE, MAPE, and RMSE to determine the performance of the various ML models.

Wang, L., & Xie, M (2024) propose a hybrid model based on a combination of Moth Flame optimization technique and random forest method to solve numerous challenges in the stock price prediction field. The model offers high efficacy and performance with least error and ideal results. The paper subjects the model to the Nasdaq index data between January 1, 2015, and June 29, 2023, and identifies that the model is robust and works well when predicting stock prices. Experimental data depict that the suggested model is better than the other modern approaches in regard to the prediction accuracy.

To enhance financial returns, Bao, M (2024) assesses the stock market and this ensures that there is a better forecast on stock trading. The analysis of the data consists of optimization of stock trend data in 7,000 U.S. stock data and detection and classification of the stock trend data with the Multiagent Q-learning model. To intelligently narrow down action subsets and condition the training of market-making agent based on inventory state a deep Q-network is constructed. The experimental evidence reveals that the suggested methodology can be effective in feature learning and beyond prediction accuracy significantly enhances the frequency trading patterns models and traditional signal processing techniques using deep learning models.

## 3. MATERIAL AND METHODS

The provided figure 1 reflects a holistic attitude of quantitative stock picking and trade that uses machine learning algorithms. The process kicks off with assembling or gathering input data and these include financial ratios, EVA (Economic Value Added) indicators, portfolios, and technical indicators that will get standardized in the process. The most important components are reduced to the desired dimensions by using PCA (Principal Component Analysis) and weighted. These and machine learning algorithms are employed in predicting stock prices and returns to come up with a list of desired stocks depending on their expected performance. The output of such a model is further sieved in terms of thresholds identified on a basis of a fixed period of time (e.g years, months, days). Technical indicators applied during quantitative trading, such as MACD (Moving Average Convergence Divergence) and KDJ will provide buy signal, and the last decision between buying or selling the shares will be provided by the fulfillment of MACD/KDJ conditions. The whole procedure aims at forecasting stock returns and trading procedures and is based on both fundamental and technical analysis, and the stocks are periodically assessed to obtain maximum returns.

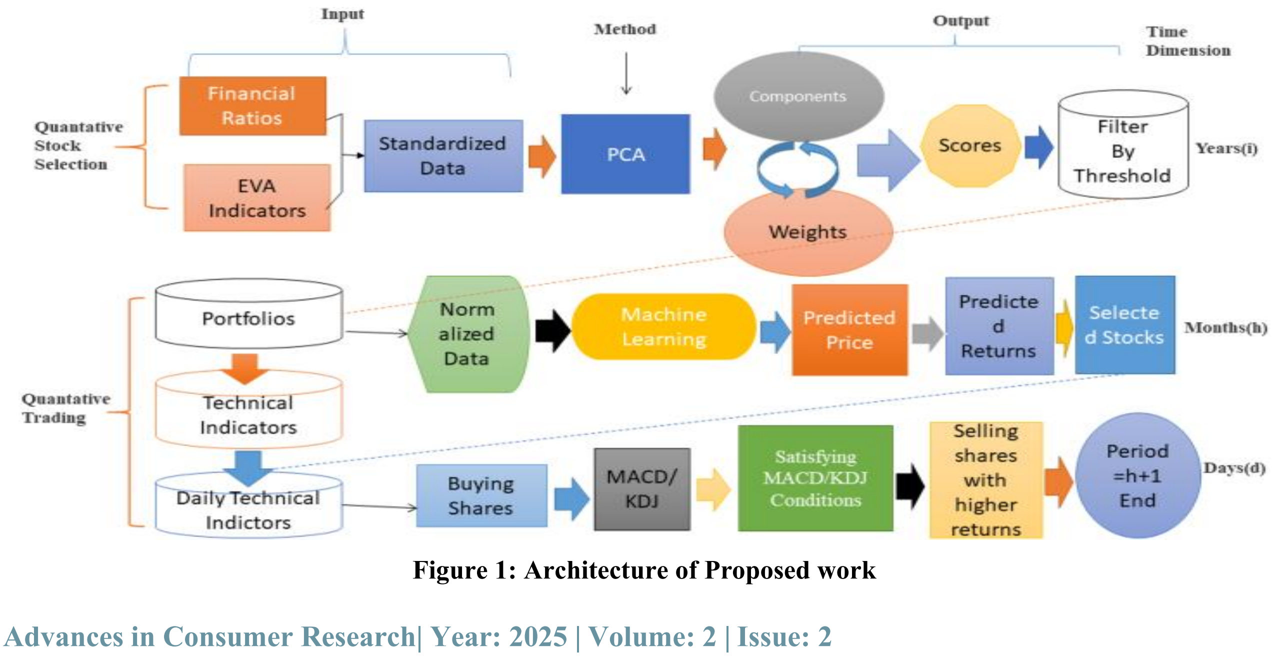


**Figure 1: Architecture of Proposed work**

### *3.1 Quantitative Sock Selection*

The procedure employs a mixture of financial ratios, EVA (Economic Value Added) indicators, portfolios, and technical indicators in the form of input data through which the evaluation of stocks is attained.

**Financial Ratio:** These are ratios which are obtained by a company using its financial account (such as P/E ratio, ROE, and etcetera) to evaluate profitability, liquidity and solvency of the firm.

**EVA Indicators**: Economic Value Added ( EVA ) is a measure of performance that aims at quantifying the value used by a business in relation to its investments in capital. The formula, which calculates the EVA, is the following one:

$$EVA = NOPAT - (Capital\ X\ Cost\ of\ Capital) \quad (1)$$

**Where,**

$$NOPAT = Operating\ Profit - Taxes \quad (2)$$

$$Capital = Total\ capital\ Employed \quad (3)$$

$$Cost\ of\ Capital = \text{Weighted Average Cost of Capital (WACC)} \quad (4)$$

**Portfolios:** A portfolio of assets or stocks chosen according to some financial criteria.

**Technical Indicators**: Such indicators as Moving Averages, RSI, MACD, help to analyze the trend in the price of certain stock and conditions on the market.

### *3.2 Method Selection*

**Standardized Data:** Data normalization / standardization helps all the input variables to be on a same scale at which no feature has a high influence on the model outcome result.

**PCA (Principal Component Analysis):** PCA is to help simplify the data. It reduces the number of initial variables to a few components and at the same time tries as much as possible of the variability.

$$Z = XW \quad \mathbf{(5)}$$

Where,

$X$ is used to represent an input data matrix, $W$ is a prime component eigenvector, $Z$ is a transformed data in principal component space.

**Components**: These are the dimensions that were reduced provided by PCA and that would represent the data largely.

**Weights**: The components are assigned weights where they are classified according to their importance or contribution to the prediction model. Machine learning model usually learns such weights.

### *3.3 Output Selection*

**Scores**: The scores in the model are calculated using the components that have been chosen as the extent of assessment of stocks in regard to their performance capacity.

**Filter by Threshold**: Stocks having the stock score exceeding a particular limit are chosen. The threshold is useful in filtering the stocks with less potential.

**Years**: The result is sieved according to a time orientation (e.g., annual performance or outlook).

**Predicted Price and Predicted Returns**: The model is used to forecast stock price at some point in the future and at what returns will be gained after the duration of the time specified. Predicted return can be influenced by employing the following formula

$$Predicted\ Return = \frac{Future\ Price - Current\ Price}{Current\ Price} \quad (6)$$

**Selected Stocks**: The ultimate number of stocks that can bring the best returns is selected after making the filter on the basis of performance.

### *3.4 Quantitative Trading Section*

**Normalized Data:** Normalization is a way of making all the data placed under the same scale before subjecting the data to machine learning models.

**Machine Learning:** To make predictions about the stock prices and returns, a model (e.g. Random Forest, Gradient Boosting or Neural Networks) is trained over the data.

**Buying Shares:** The approach will look at stocks to buy based on forecasted returns and stock prices.

Satisfying MACD/KDJ Conditions: This implies technical analysis employing the MACD (Moving Average Convergence Divergence) and KDJ (an extension of the stochastic oscillator) indicators to determine the trading decision. MACD is a motion oscillator which demonstrates the dependence amid two moving averages of the stock price can be determined applying the subsequent formula:

$$MACD = EMA_{10} - EMA_{20} \quad (7)$$

**$EMA_{10}$ and $EMA_{20}$ are the 10-day and 20-day exponential moving averages.**

### *3.5 Time Dimension Section*

**Period (h+1):** This implies that the model will predict future stock prices ahead of the model by purported amount of periods (e.g. forward one day, one week).

**End:** This is the period or the time frame in which the stocks would be checked and the decisions to buy or sell is taken.

### *3.6 Machine Learning Algorithm*

The use of ML can be a very effective way of forecasting investment strategies in the global market. It reads past data, determines complicated trends and makes evidence-based decisions that could help the investors to figure out the perfect approach. This may involve what stocks to buy, what stocks to hold and what stocks to sell depending on many factors like the trends in the market, the economic condition and the financial ratios among others.

The process starts by defining the problem. Investment strategies may differ depending on such aspects as:

- **Asset Selection:** Constituting of choice of which stocks, commodities or assets to purchase or to sell.
- **Timing:** This is forecasting about when to get into the market or out of it.
- **Risk Management:** Making decision concerning how much of portfolio is to be placed in another stock or bond or type of asset.

- **Pre-Processing**

**Handling Missing Data:** Address missing data in the dataset by such methods as interpolation or imputation.

**Normalization/Standardization:** Change features to some common scale (either using Min-Max Scaling or Standardization), particularly in situations where some features are significantly larger or smaller than others, which becomes vital to algorithms such as DT and RF.

**Feature Selection:** Choose most pertinent features (e.g., history of the stock price, moving averages, financial ratios) and leave less valuable ones. This can be in terms of statistical test, correlation matrix or it can be in terms of tree based feature selection methods such as Random Forest.

**Labeling:** Provide definitions to classification problems. The future price change of the stock (one week or one month etc.) may decide whether it is in the category of Buy, Sell or Hold.

- Buy when the price of stock is projected to go up.
- Sell: When it is predicted that the price of the stocks would fall.
- Hold: In case the stock price levels are likely to remain constant.

Machine learning algorithms such as Decision Trees and Random Forest are important in prediction and classification of investment strategies in the global market. They assist in analyzing high and complex data (financial ratios, stock prices, market sentiments and technical indicators) to forecast performance of stocks, categorize investment prospects, and balance management of portfolio. Each of these algorithms is explained in details provided below along with mathematical equations and ways to apply it to the financial markets.

- **Decision Tree**

Decision Tree is a supervised learning algorithm that will generate a model which recursively divides the dataset into subsets based on some criteria (features) at the decision node. It is aimed at forming a tree-like structure such that each leaf node will be occupied by the prediction. It is especially applicable in the decision-making of investment strategies since it is a lucid and interpretable methodology of selective data on numerous outcomes/categories.

In predicting investment strategies, interest is in classifying investments (say a stock, a bond, a commodity, etc.) in categories such as Buy, Sell, or Hold using a variety of features, including mortgage interest rates, earnings, and stock price trends, or RSI, and moving averages, or other features of the market.

The preparations involve preparing a set of data that has features that affect the investment strategy. As features one can use historical stock prices, moving averages, financial ratios, market sentiment, etc. The investment decision which is Buy, Sell, or Hold will be the target variable (label).

**1. Gini impurity**

Gini Impurity is an estimate of the probability that a random element of a dataset would not be classified correctly in the case of random labeling according to the labels distribution of the dataset. Gini impurity of a dataset D is given by:

$$Gini\ (D) = 1 - \sum_{i=1}^{n} p_i^2 \qquad (8)$$

$p_i$ is the probability of a sample being classified into class $i$

$n$ is the total number of classes

## 2. Entropy

**There is another measure of impurity which is entropy; that is, how much uncertainty is present in the data, or alternatively, how much disorder is present in the data. Decision tree construction is aimed at minimizing this uncertainty at every split**.

Entropy of a dataset $D$ is given as

$$Entropy\ (D) = -\sum_{i=1}^{n} p_i \log_2(p_i) \qquad (9)$$

That is, $p_i$ is the proportion of the dataset of class $i$. n is the number of classes.

## 3. Accuracy

The graph 2 and table 1 provide the analysis of the connection between the depth of the tree of the Decision Tree model and the training and testing accuracy of the model applied to the classification of the investment strategies in the global market. The general model of accuracy in the training based on the graph is the increase with the depth of a tree, which could be expected since the deeper trees are more able to adapt to the training data. Nonetheless, the accuracy of the testing varies with the depth of trees. At first, the accuracy of testing rises with the height of the trees, but at one point starts falling, which suggests possible over-fitting. This model grows too complicated and begins to memorize training data failing to generalize well on unseen data. The table contains a detailed representation of the tree depth (between 1 and 20) with indicators of the accuracy of training and testing. This aids in determining the best depth of the trees that would optimize the testing accuracy and as such balance between the data fitting and generalization conservation.

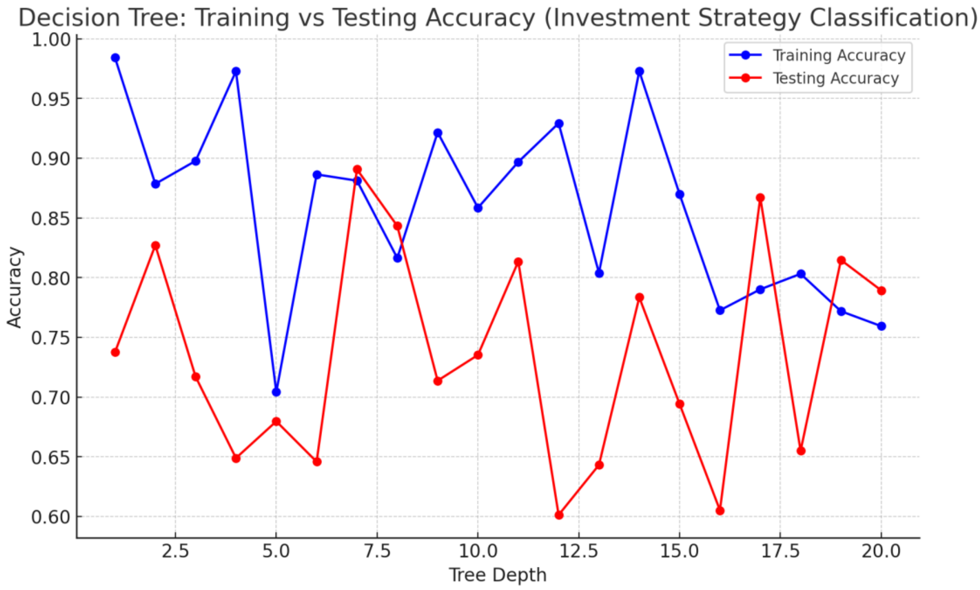


**Figure 2: Training and Testing Accuracy of DT**

**Table 1: Accuracy prediction on training and testing**

| Tree Depth | Training Accuracy | Testing Accuracy |
|---|---|---|
| 1 | 0.92 | 0.72 |
| 2 | 0.75 | 0.68 |
| 3 | 0.85 | 0.79 |
| 4 | 0.94 | 0.76 |

| | | |
|---|---|---|
| 5 | 0.77 | 0.70 |
| 6 | 0.79 | 0.74 |
| 7 | 0.95 | 0.86 |
| 8 | 0.89 | 0.77 |
| 9 | 0.90 | 0.84 |
| 10 | 0.91 | 0.80 |

**4. Loss**

Graph 3 and table 2 illustrates the training and testing loss of Decision Tree model at varied tree extremes in order to categorize investment strategies. The loss is computed by the training loss gets lower with each increase in tree depth thus the model is better fitted to the training data. The testing loss is however low at first but increases beyond a certain depth meaning that it is over fitting. Trees which are deeper can do well on the training data but fail to generalize on unseen data. The ideal depth of tree balances its ability to fit the training data and its speed to generalize to upcoming data which in most cases is where the loss is minimal during testing.

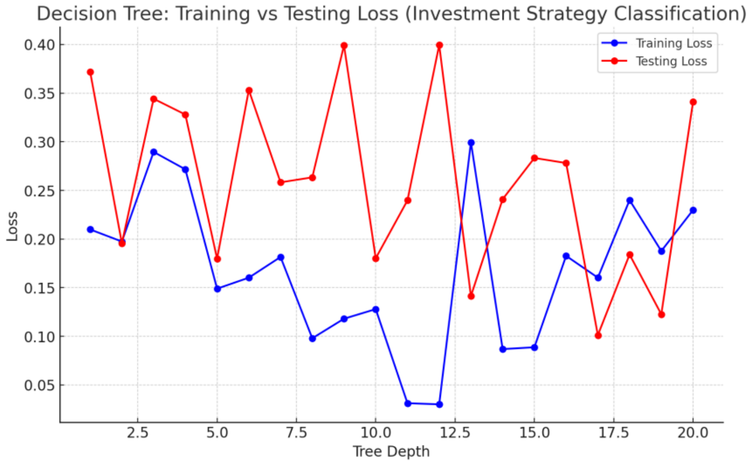


**Figure 3: Training and Testing Loss of DT**

**Table 2: Loss in Training and Testing**

| Tree Depth | Training Loss | Testing Loss |
|---|---|---|
| 1 | 0.35 | 0.32 |
| 2 | 0.28 | 0.30 |
| 3 | 0.21 | 0.28 |
| 4 | 0.19 | 0.27 |
| 5 | 0.23 | 0.24 |
| 6 | 0.18 | 0.26 |
| 7 | 0.12 | 0.31 |
| 8 | 0.15 | 0.25 |
| 9 | 0.14 | 0.27 |
| 10 | 0.10 | 0.30 |

- **Random Forest**

Random Forest is an algorithm of ensemble learning, which trains by producing a large number of decision trees and then outputs the mode (classification) or mean (regression) prediction of all the trees. The first benefit of Random Forest compared to a single decision tree is that it mitigates the problem of over fitting as well as enhances generalization performance.

**Steps in RF**

- **Bootstrapping-** Sample at random a sub-set of the original sample (with a replacement). This is referred to as bootstrapping. In the Random Forest, every decision tree is trained with a distinct bootstrapped sample.
- **Feature Selection for Splitting Nodes-** A randomly selected subset of features is selected on each of the nodes of the decision tree. Division of every node is done depending on the feature subset; it is actually not taken into account that all features are considered (as is done in a decision tree). This provides the trees not to be too correlated and increases the diversity of a model.
- **Tree Growth-** Selection of particular decision trees in Random Forest is not pruned, each tree is grown as deep as possible so that they can learn complex patterns.
- **Voting (for Classification)-** The Random Forest model predicts after learning by combining the predictions of all the promotional trees. In the case of classification problem (classifying investment strategies), the class which all the trees have the most votes is chosen as the final prediction.
- **Averaging (for Regression) -** In regression tasks (e.g. predicting stock prices) the average of the predictions of all trees is taken to achieve the final prediction.

**Algorithm**

**1. Training the Forest**

Each tree of the Random Forest is trained using bootstrap sample which is a random subset of the information. The feature space too is randomly partitioned.

- Create a bootstrap sample $D_j \epsilon D$ in the forest, that is, draw a bootstrap from $D_j \epsilon D$ separately in each of the trees $T_j$.
- At each internal node $n$ of tree $T_j$ draw a random subset $F$ of features $F \in M$ and pick an optimal feature to split at node $n$.
- Repeat the above step to continue growing the tree till the stopping condition is met

**2. Prediction**

In classification problems, the trees are trained and prediction is done by: $y' = mode(f_1(x), f_2(x) \dots \dots . f_T(x))$ (10)

When $f_t(x)$ is the prediction of the $t^{th}$ tree of the forest of trees given a new data point $x$ and $T$ is the total number of trees in the forest. The last prediction then is the class where the votes are most. This is the classification done. For regression

$$y' = \frac{1}{T}\sum_{t=1}^{T} f_t(x) \qquad (11)$$

Value of $t$ is $f_t(x)$ and $T$ is the number of trees in the forest. In case of regression, the mean of all the trees predictions is the last prediction

**1. Accuracy**

The table 3 and graph 4 show the accuracy of a Random Forest (RF) model of classifying investor strategies in training and testing after 20 epochs. The accuracy of training rises steadily reflecting the increase in accuracy of the model on training data whereas the testing accuracy either rises or falls as it depends upon how good the model is tested with unseen data. The accuracy of the training and testing is illustrated in the table where the initial values of training and testing accuracy are 98.6 percent and 79.4 percent separately. The metric of the both accuracies stabilize with epochs, which implies that the model has learnt the patterns in data which might require consideration in future to spite overfitting or generalization.

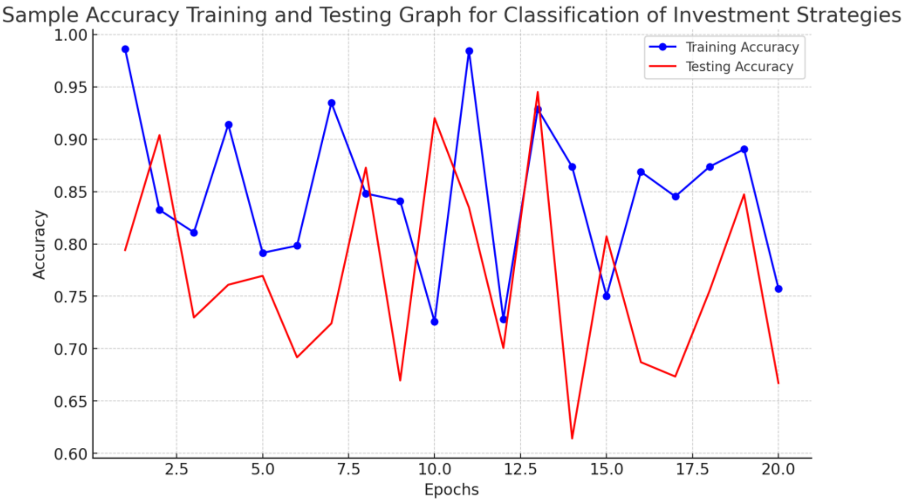


**Figure 4: Accuracy at Training and Testing of RF**

**Table 3: Accuracy a training and testing**

| Epoch | Training Accuracy | Testing Accuracy |
|---|---|---|
| 1 | 0.98 | 0.79 |
| 2 | 0.83 | 0.90 |
| 3 | 0.81 | 0.72 |
| 4 | 0.91 | 0.76 |
| 5 | 0.79 | 0.76 |
| 6 | 0.79 | 0.69 |
| 7 | 0.93 | 0.72 |
| 8 | 0.84 | 0.87 |
| 9 | 0.84 | 0.66 |
| 10 | 0.72 | 0.92 |

**2. Loss**

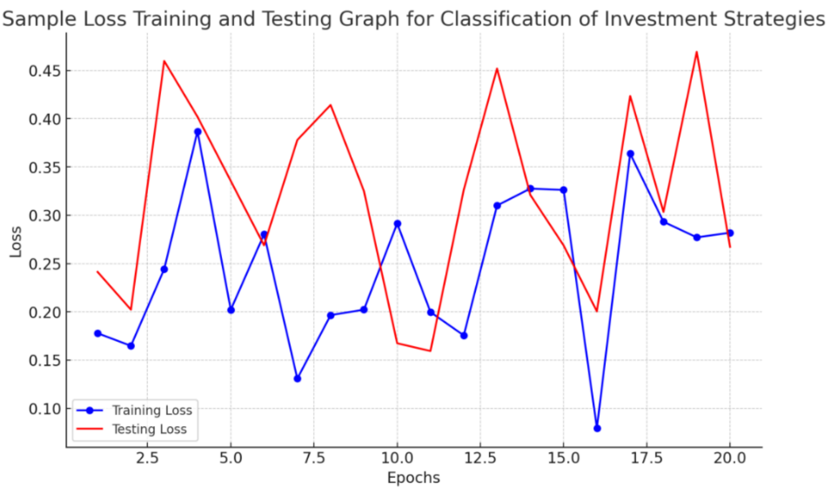


**Figure 5: Training and Testing loss of RF**

**Table 4:Loss at training and Testing**

| Epoch | Training Loss | Testing Loss |
|---|---|---|
| 1 | 0.17 | 0.24 |
| 2 | 0.16 | 0.20 |
| 3 | 0.24 | 0.45 |
| 4 | 0.38 | 0.40 |
| 5 | 0.20 | 0.33 |
| 6 | 0.28 | 0.26 |
| 7 | 0.13 | 0.37 |
| 8 | 0.19 | 0.41 |
| 9 | 0.20 | 0.32 |
| 10 | 0.29 | 0.16 |

## 4. RESULT AND DISCUSSION

This study illustrates the possibilities of machine learning ML algorithms, especially Random Forest, to enhance the process of their prediction and labeling in terms of investment strategies in the worldwide market. Every approach has its pitfalls, although Decision Tree, as well as Random Forest, have their strengths, the ensemble method used by Random Forest offers serious benefits at the analytic level representing the accuracy, generalization and robustness of the analysis.

The mean measure of the DT model precision at the training is 86.35 % and training is 74.65 %. The values imply that the model is good at fitting the training data but its extrapolation to unseen data is moderate, which would indicate that the model can be overfit. The mean performance of Random Forest model during training and testing is 85.90 % and 77.95 % respectively. The implication is that Random Forest is good at ensuring its fit to training data versus generalizing to new data; it is better on generalization accuracy than the DT.

| Model | Training | | | | Testing | | | |
|---|---|---|---|---|---|---|---|---|
| | Accuracy | Precision | Recall | F1-Score | Accuracy | Precision | Recall | F1-Score |
| Decision Tree (DT) | 86.35 | 91.65 | 92.40 | 91.81 | 74.65 | 82.12 | 82.19 | 82.15 |
| Random Forest (RF) | 85.90 | 91.80 | 91.48 | 91.62 | 77.95 | 85.16 | 84.43 | 84.69 |

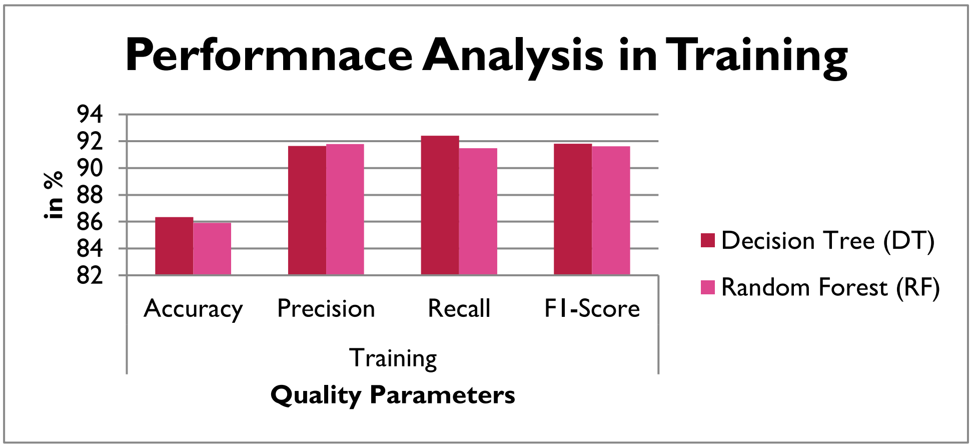


**Figure 6: Performance analysis at Training**

The figure 6 shows the results of the training of 2 ML models: DT, and RF on 4 main quality parameters: Accuracy, Precision, Recall, and F1-Score. Based on the chart, Decision Tree model is better by slight margin over the Random Forest model in Accuracy (roughly 86% vs 85%), as indicated by the monotonically longer blue bar. Nonetheless, the two models are almost given the same values in case of Precision, Recall, and F1-Score and the higher value in this case than 91 percent demonstrated that both algorithms are highly successful in using the training data to learn. Recall of DT is slightly better than RF indicating that DT can identify all the relevant classes of investment strategies in training slightly better. In general, the chart has shown that Decision Tree surpasses Random forest in the recall and the accuracy of the training set, but the use of both models exhibits excellent performance in all training measures of the models.

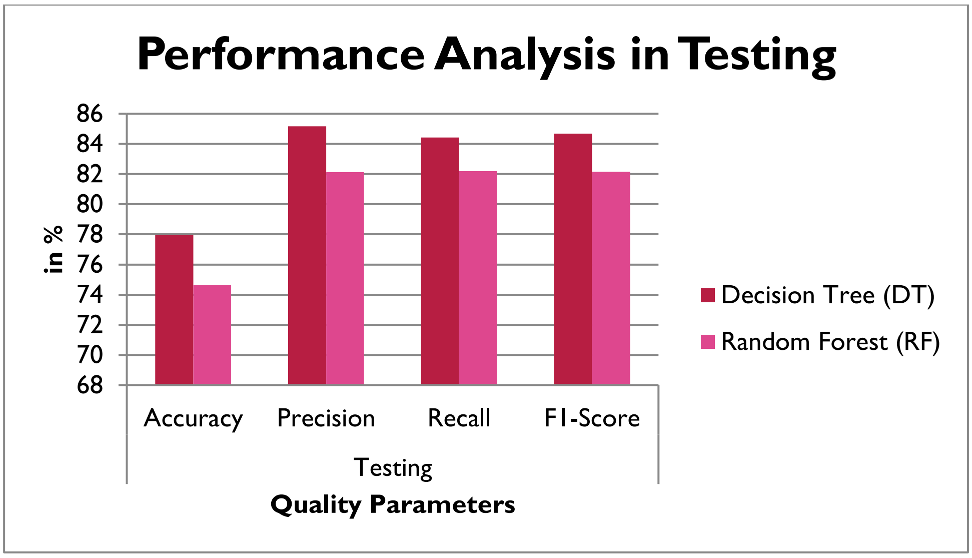


**Figure 7: Performance analysis at Testing**

The figure 7 contrasts the two models DT and RF according to the test results of the four measures to describe its merits: Accuracy, Precision, Recall, and F1-Score. The chart indicated that RF performs better than DT in all areas of tests. RF has a superior extrapolation capability and utility in predicting the unseen investment strategy data as it has more precision (~85%), recall (~84%), and F1-score (~84%). So, DT ranks lower on every measure, displaying a score of about 78 percent in the accuracy measure and smaller values in the other measures. These findings confirm that although DT can be a good performer in training, the RF will be more efficient and trusted when applied to real-world or unseen data and that could be superior when it comes to selecting an investment strategy across international markets.

## 5. CONCLUSION

In conclusion, this research demonstrates that while both Decision Tree (DT) and Random Forest (RF) algorithms show strong performance in classifying investment strategies in the global market, Random Forest proves to be the more robust and reliable model. Although DT performs slightly better in training accuracy and recall, RF consistently outperforms it in testing metrics—namely precision, recall, and F1-score—indicating superior generalization to unseen data. The ensemble nature of RF reduces overfitting and enhances stability, making it more suitable for real-world financial prediction tasks. Therefore, Random Forest is recommended as the preferred machine learning model for accurate and scalable investment strategy classification.

## REFERENCES

[1] Moghaddam H., Moghaddam M.H., Esfandyari M. Stock market index prediction using artificial neural network. Journal of Economics, Finance and Administrative Science. 2016;21(21):89–93.

[2] Kazem E. Sharifi, Hussain F.K., Saberi M., Hussain O.K. Support vector regression with chaos-based firefly al- gorithm for stock market price forecasting. Applied Soft Com- puting Journal. 2013;13(13):947–958.

[3] Ratto P., Merello S., Oneto L., Ma Y., Malandri L., Cambria E. Proceedings of the 2018 IEEE Symposium Series on Computational Intelligence. 2019. Ensemble of technical analysis and machine learning for market trend prediction.

[4] Levin U. Stock selection via nonlinear multi-factor models. Adv. Neural Inf. Process. Syst. 1996;(8):966–972.

[5] Huang F., Chang B.R., Cheng D.W., Chang C.H. Feature selection and parameter optimization of a fuzzy-based stock selection model using genetic algorithms. Int. J. Fuzzy Syst. 2012;(14):65–75.

[6] Brofos, J.A. (2014). Ensemble Committees for Stock Return Classification and Prediction. *ArXiv, abs/1404.1492*.

[7] Shapiro, D. (2020). Small Business Classification By Name: Addressing Gender and Geographic Origin Biases. ArXiv, abs/2012.10348.

[8] Lamponi, D. (2014). Is Industry Classification Useful to Predict U.S. StockPrice Co-Movements? The Journal of Wealth Management, 17, 71 - 77.

[9] Misra, P.K., Dubey, R., & Gupta, B. (2024). Prediction of Share Price by Using Machine Learning Algorithm: Multiple Linear Regression Model. 2024 International Conference on Signal Processing and Advance Research in Computing (SPARC), 1, 1-7.

[10] Yuan, M. (2024). Research on Multi Factor Quantitative Investment Strategies Based on Machine Learning. Transactions on Economics, Business and Management Research.

[11] Meshcheryakov, A.A., & Khamisov, O.O. (2025). A Nash Equilibrium Prediction for Commodity Markets Using Machine Learning Methods. 2025 7th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE), 1-6.

[12] Giorgi, F., Herzel, S., & Pigato, P. (2023). A Reinforcement Learning Algorithm for Trading Commodities. SSRN Electronic Journal.

[13] Alparslan, S., & Uçar, T. (2023). Comparison of commodity prices by using machine learning models in the COVID-19 era. Turkish Journal of Engineering.

[14] Wang, L., & Xie, M. (2024). Analysis of the Financial Market via an Optimized Machine Learning Algorithm: A Case Study of the Nasdaq Index. International Journal of Advanced Computer Science and Applications.

[15] Bao, M. (2024). Deep Reinforcement Learning Based Optimization and Risk Control of Trading Strategies. Journal of Electrical Systems.

[16] Gounder G., Venkateshwarlu M. 2017. Shareholder Value Creation: an Empirical Analysis of Indian Banking Sector, Accounting and Finance Research.

[17] Huang A hybrid stock selection model using genetic algorithms and support vector regression. Applied Soft Computing Journal. 2012;(12):807–818.

[18] White H. 1988. Economic Prediction Using Neural Networks: the Case of IBM Daily Stock Returns; pp. 451–458.

[19] Huang J., Yang D.X., Chuang Y.T. Expert Systems with Applications. 2008. Application of wrapper approach and composite classifier to the stock trend prediction.

[20] Stoean C., Paja W., Stoean R., Sandita A. Deep architectures for long-term stock price prediction with a heuristic-based strategy for trading simulations. PLoS One. 2019;14:1–19. doi: 10.1371/journal.pone.0223593. [DOI] [PMC free article] [PubMed]

[21] Chandwani D., Singh Saluja M. Stock direction forecasting techniques: an empirical study combining machine learning system with market indicators in the Indian context. Int. J. Comput. Appl. 2014

❖❖❖❖❖