Machine Learning in Momentum Strategies

Hsuan-Ling Chang, Hung-Wen Cheng, Yi-Min Lan*, Jou-Ping Yu

Abstract

The study applies machine learning models to construct momentum strategies and utilizes the information coefficient as an indicator to select stocks with strong and weak momentum characteristics. Through this approach, we have built investment portfolios capable of generating superior returns and conduct a thorough analysis. Compared to existing research on momentum strategies, we incorporate machine learning to capture non-linear interactions. This approach improves the traditional stock selection process, which is frequently hindered by challenges related to timeliness, accuracy, and efficiency due to market risk factors. We find that implementing bidirectional momentum strategies outperforms unidirectional ones, and momentum factors with longer observation periods exhibit stronger correlations with returns. Optimizing the number of stocks in the portfolio, while staying within a certain threshold, leads to the highest excess returns. We present a novel framework for momentum strategies that enhances and improves the operational aspects of asset management. By introducing innovative financial technology applications to traditional investment strategies, we demonstrate significant effectiveness.

Key words： Information Coefficient, Machine Learning, Momentum, Portfolio, Return Prediction

1 Introduction

The efficient market hypothesis has been a fundamental concept in finance. It states that stock prices fully reflect all available information and that consistently outperforming the market is impossible. However, in recent decades, a growing body of research has challenged this traditional theory. One notable contribution to this paradigm shift is the groundbreaking work of Chan et al. (1996) in their seminal paper titled "Momentum Strategies." In their research, the authors explored a novel investment strategy based on momentum. This strategy involves buying stocks with strong historical performance (past winners) and selling stocks with weak historical performance (past losers). Surprisingly, their findings revealed that this momentum strategy could generate abnormal returns, suggesting that stock market efficiency may not be as robust as previously believed. The implications of their discovery are significant, as it challenges the prevailing notion of market efficiency and suggests that active trading strategies, such as momentum, could potentially outperform traditional passive investment approaches. Their research sheds new light on understanding market behavior and raises essential questions about the underlying factors that drive stock prices. Moreover, the paper goes beyond merely identifying profitable trading strategies. It also provides valuable insights into market efficiency by highlighting the presence of price reversals and performance disparities among various types of stocks. Building upon the momentum concept, subsequent research by Fama and French (1992) emphasizes the cross-section of expected stock returns. Their work provides complementary insights into the momentum phenomenon, considering it as a special case of size and value strategies. The relationship between momentum and other well-known factors enhances our comprehension of asset pricing and market anomalies. Another influential study by Shleifer and Vishny (1997) explores the "limits of arbitrage," uncovering the effects of constraints on arbitrage opportunities that can impact momentum strategies. This highlights the importance of considering market dynamics and transaction costs when designing investment strategies.

As the literature on momentum strategies continued to evolve, Barroso and Santa-Clara (2015) introduced a risk-managed version of the momentum strategy. This version incorporates risk management techniques to enhance returns while reducing volatility and risk. Their innovative approach showcases the potential benefits of dynamic momentum strategies and provides investors with a new tool to navigate market complexities. Similarly, Daniel and Moskowitz (2016) examined momentum crashes, providing insights into the reversal of momentum profits over time. Their findings challenge traditional risk-based explanations of momentum, suggesting that other factors, such as behavioral patterns or market inefficiencies, may contribute to the profitability of momentum investing. Constructing momentum strategies is not only theoretically well-supported but also practically feasible. The extensive body of research discussed in the previous sections has consistently shown that momentum strategies have the potential to generate abnormal returns in the stock market. However, there have been challenges in implementing and refining these strategies in real-world settings. One of the key challenges in constructing a traditional momentum strategy is the requirement for precise and timely stock selection. Identifying past winners and losers, as well as determining the optimal holding period and rebalancing frequency, can be a complex task. Additionally, maintaining a dynamic and adaptive approach to changing market conditions poses a significant challenge.

Machine learning offers an innovative solution to these challenges. By leveraging vast amounts of historical market data and employing sophisticated algorithms, machine learning can effectively identify stocks with strong momentum characteristics and optimize trading parameters. This technology can not only enhance the accuracy of stock selection but also automate the entire process, enabling quicker and more efficient decision-making. Furthermore, machine learning algorithms have the ability to identify and exploit patterns and signals that human analysts may overlook. This can lead to a deeper understanding of market dynamics and the discovery of additional factors that may contribute to momentum profits, as demonstrated in the research conducted by Barroso and Santa-Clara (2015). Machine learning can uncover new insights and potentially enhance the performance of momentum strategies. Additionally, the ability of machine learning models to adapt to changing market conditions can help address the issue of momentum profit reversals highlighted by Griffin et al. (2003). These models can continuously analyze market data and adjust trading strategies in response to evolving market trends. This capability potentially helps to mitigate the impact of reversals and enhance the resilience of momentum strategies. The feasibility of constructing momentum strategies is further enhanced by the application of machine learning techniques. By harnessing the power of data analysis and automation, machine learning can overcome past challenges in stock selection, optimize trading parameters, and adjust to evolving market conditions.

In modern finance and machine learning domains, the Information Coefficient (IC) is widely used to measure the correlation between forecasted outcomes and actual results. This metric, originally introduced by Goodwin (1998) in the context of financial analysis, has become an essential tool for assessing the accuracy and skill of various predictions. The IC serves as an indicator of how well a manager's forecasts align with the actual returns observed in financial markets. When the IC is high, it signifies that the manager's predictions are more accurate and capable of identifying mispricing in the market. On the other hand, a low IC suggests that the forecasts do not significantly contribute to generating value. The utility of the IC extends beyond finance and investment management. Researchers, such as Kinney and Atwal (2014), have applied it in various statistical analyses, including the analysis of financial data to identify co-moving stocks, as well as in machine learning applications for feature selection and clustering. Its versatility lies in its ability to capture complex relationships between variables, making it sensitive to both linear and non-linear dependencies. This sensitivity helps improve the performance and accuracy of prediction algorithms.

In this paper, we build upon the research conducted by Gu et al. (2020), who integrated various machine learning techniques with contemporary empirical asset pricing research to analyze the fluctuations of market risk premia for stock returns. Their findings indicated that machine learning improves the description of expected returns. When applied to portfolio construction, it leads to significant performance improvements, especially with more advanced models that consider non-linear predictor interactions that simpler methods overlook. Given these characteristics, machine learning is particularly appealing for investors engaged in short-term speculative behavior, making it well-suited for constructing momentum and reversal strategies. We first verify the ability of momentum and reversal strategies, constructed using the Information Coefficient (IC), to generate significant excess returns. To achieve this, we are motivated to accurately predict the next month's IC by the end of the previous month. This will enable us to obtain a list of investment portfolios that can generate superior returns. Next, we train models using momentum factors and IC to explore the predictive capabilities of different machine learning models when combined with IC attributes. This is done to evaluate the effectiveness of momentum and reversal strategies.

The evaluation approach, based on the confusion matrix (Hernández-Orallo et al., 2012), is unconventional for a non-classification model. However, the objective here is not to directly apply the results of a regression model, but rather to evaluate the consistency in direction between the actual and predicted returns, considering whether they are positive or negative. By utilizing the confusion matrix, we aim to analyze the performance of the model in terms of accurately predicting positive and negative outcomes. The matrix allows for the calculation of key performance metrics, such as true positives (correctly predicted positive returns), false positives (incorrectly predicted positive returns), true negatives (correctly predicted negative returns), and false negatives (incorrectly predicted negative returns). This enables us to calculate four performance metrics that assess the accuracy of predicting positive or negative actual and projected returns. While the model used in this research may not adhere to the conventional classification framework, the utilization of the confusion matrix enables the assessment of the model's accuracy in predicting positive and negative returns. This approach enables a unique evaluation of the model's performance that aligns with the specific research objectives. Our results demonstrate that the applied machine learning models achieve excellent out-of-sample performance.

Observations from our empirical analysis include the following: While momentum and reversal strategies constructed using real IC show the ability to generate absolute returns, those built using machine learning models are only successful for momentum strategies, in line with the findings of Gu et al. (2020). Additionally, we observe that portfolios formed by shorting the weakest momentum stocks outperform those formed by longing the strongest momentum stocks. This provides clear evidence of the "winners' loser" effect, which suggests that price reversal does not occur over the one-month holding period. Furthermore, we find that as the number of companies with long or short positions decreases, the returns increase significantly. However, once the number of stocks falls below a certain threshold, the returns start to decline. Lastly, we compare momentum factors with different formation periods. Our findings reveal that longer formation periods are associated with higher adoption rates. This suggests that momentum factors with longer formation periods exhibit stronger correlations with future returns. This contrasts with the findings of Gu et al. (2020), as we observe a higher importance for momentum factors with shorter formation periods. Previous research has demonstrated that momentum strategies can achieve excess returns. However, these strategies are often influenced by economic conditions, and their effectiveness tends to diminish during economic recessions. Additionally, precise and timely stock selection remains a significant challenge. Our results show that constructing momentum strategies based on machine learning predictions of IC values yields significant and economically meaningful performance, even during periods of market turbulence such as the 2015 stock market crash and the post-2020 COVID-19 pandemic era. While this approach confirms that IC can serve as an evaluation metric for constructing portfolios with excess returns, it is essential to consider the impact of transaction costs and other market factors.

The rest of the paper is structured as follows. In Section 2, we introduce the variables utilized in this study, offering a comprehensive overview of their importance and relevance to our research objectives. In Section 3, we elaborate on the research design adopted and the methodology employed to construct investment portfolios. Additionally, Section 4 offers an explanation of the basic operating principles of the various machine learning models utilized in our analysis. We present our empirical analysis in Section 5. We evaluate the out-of-sample predictability of the constructed models and extensively discuss the relative importance of various predictors. We observe and compare the real-world outcomes with the predictions generated by the machine learning models. Our goal is to derive meaningful insights into the effectiveness of momentum and reversal strategies in the context of market risk premia dynamics. We conclude in Section 6.

2 Variables

2.1 Logarithmic Return

The rate of return utilized in this paper is the logarithmic return, rather than the simple return. Logarithmic return possesses an additive nature, remaining unaffected by the base period and time; thus, the rate of increase or decrease remains constant. This concept can be expressed as follows:

$$R_{t+h} = \ln\left(\frac{S_{t+h}}{S_t}\right), \tag{1}$$

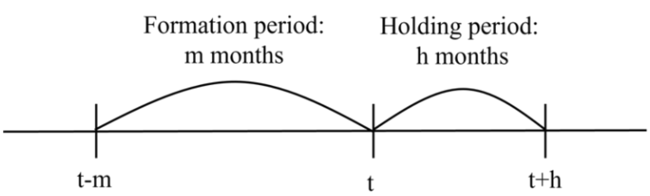
where R_{t+h} is the return for a holding period of h months, S_{t+h} is the closing price of the stock at time $t+h$, and S_t is the closing price of the stock at time t. In this paper, h is set to 1, indicating a holding period of 1 month.

2.2 Momentum

The momentum employed is referred to as return momentum, which utilizes a formation period of m months. The concept involves dividing the current closing price by the closing price m months ago and taking the logarithm of the resulting quotient. This calculation method remains unaffected by the base period and time.

$$M_{t,m} = \ln\left(\frac{S_t}{S_{t-m}}\right). \tag{2}$$

where the momentum factors calculate using five different time intervals: m = 1, 6, 12, 36, and 60 months.



2.3 Information Coefficient (IC)

The Information Coefficient is a measure used to assess the predictive ability of investment strategies or models. It calculates the correlation between observed data and actual returns, where the data can represent different characteristics or macroeconomic variables. The IC values were computed using momentum factors and logarithmic returns as predictive indicators. Pearson Correlation is used as a statistical method to calculate the IC values, which measures the degree of linear correlation between these two sets of data. The IC value is calculated as the covariance of the two variables divided by the product of their respective standard deviations. In this context, X represents the momentum data, and Y represents the returns. It can be expressed as follows:

$$IC = \rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y}. \tag{3}$$

The correlation between the target factor and the return can be determined through the Information Coefficient (IC). The IC value ranges from 1 to -1. When the IC value is greater than 0, it indicates a positive relationship between the current factor and return. Conversely, when the IC value is less than 0, it indicates a negative relationship between the current factor and the return. The larger the absolute value of the IC, the greater the influence of the factor on the return of investment. There are two common types of IC values:

1 Normal IC：The correlation coefficient in the cross section between the target factor and the return of holding for h months at time t.

$$Normal\ IC_t = corr\left(M_{t,m}, R_{t+h}\right). \tag{4}$$

2 Rank IC：The correlation coefficient in the cross section between the ranking of the target factor and the ranking of the return of holding for h months at time t.

$$Rank\ IC_t = corr\left(r(M_{t,m}), r(R_{t+h})\right). \tag{5}$$

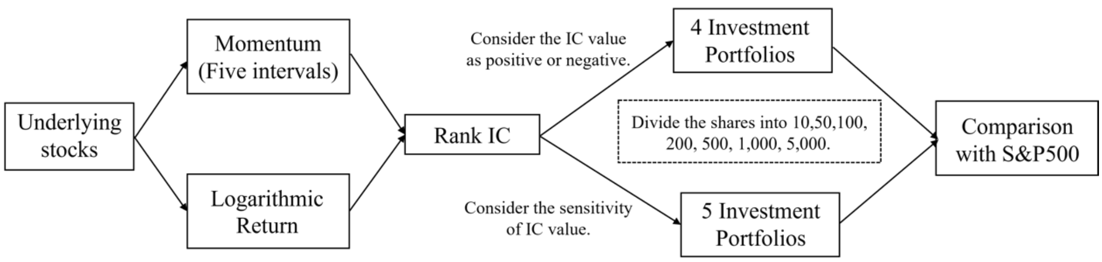
where $r(M_{t,m})$ and $r(R_{t+h})$ indicate that the momentum data and logarithmic return have been ranked before calculating the related coefficients. We use Rank IC because Normal IC requires data to follow a normal distribution, which is often not the case with financial data.

3 Research Design

3.1 Research Steps

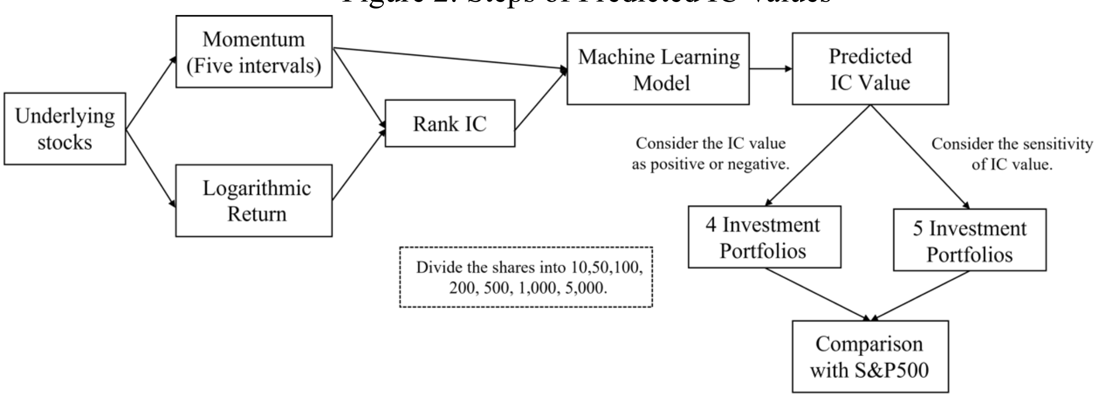
We can be broadly divided into two main parts. The first part focuses on calculating the actual Information Coefficient (IC) as a screening indicator for investment. Initially, the emphasis is on the real IC values rather than incorporating predictions. Stocks with strong and weak performance of momentum are selected based on the actual IC values, and different investment portfolios are constructed accordingly. The objective is to examine whether these portfolios, utilizing the actual IC values, can generate excess returns. This part aims to evaluate whether the IC values calculated using momentum can truly imply excess returns and serve as indicators for distinguishing between strong and weak momentum of stocks. The research steps for the first part, which focuses on actual IC values, are illustrated in Figure 1.

Figure 1: Steps of Actual IC Values



Based on the initial conclusions drawn from the first part, it provides motivation for the second part of the paper. If the investment portfolios using the actual IC values can effectively deliver robust excess returns, we employ machine learning models to predict IC values and obtain a list of companies for the portfolios in advance. This approach aims to achieve similar robust excess returns. The research steps for the second part, involving the application of machine learning, are illustrated in Figure 2.

Figure 2: Steps of Predicted IC Values



3.2 Grouping Shares

After obtaining the predicted IC values for the five different time intervals using machine learning, the absolute values of these predictions are compared, and the maximum absolute value is selected as the final predicted IC value. Stocks with prices below 1 are then removed from consideration. The remaining stocks are sorted based on their momentum strength within this time interval. They are divided into groups of 10, 50, 100, 200, 500, 1000, and 5000, with each group containing 500, 100, 50, 25, 10, 5, and 1 stock, respectively. The group with the strongest momentum is labeled as “Top”, while the group with the weakest momentum is labeled as “Bottom”.

3.3 Investment Portfolio

Two different observation methods are utilized to construct investment portfolios, both employing equal weighting to determine the stock proportions within the portfolios. The first method involves observing the original values of the selected predicted IC values before taking the absolute value. It distinguishes between positive and negative values and conducts different operations accordingly. A positive IC value indicates a positive relationship between the momentum factor and future returns, leading to a momentum strategy. Conversely, a negative IC value indicates a negative relationship, resulting in a reverse momentum strategy. This leads to four construction methods:

1. TB/BT：Buy Top and Sell Bottom when the IC value is positive; Buy Bottom and Sell Top when the IC value is negative. The return is calculated as the difference in returns between the two groups of stocks.
2. buyT/buyB：Buy Top when the IC value is positive; Buy Bottom when the IC value is negative.
3. buyT/sellB：Buy Top when the IC value is positive; Sell Bottom when the IC value is negative.
4. sellB/sellT：Sell Bottom when the IC value is positive; Sell Top when the IC value is negative.

The second method disregards whether the obtained predicted IC value is positive or negative and applies the same operation in all cases. This results in five construction methods:

1. buyT：Buy Top.
2. buyB：Buy Bottom.
3. sellT：Sell Top.
4. sellB：Sell Bottom.
5. buyTsellB：Buy Top and Sell Bottom simultaneously. The return is calculated as the difference in returns between the two groups of stocks.

In total, there are nine different investment portfolio construction methods using these two approaches.

4 Methodology

We describe the relationship between the predicted IC values and the corresponding predictor variables as an additive prediction error model. It can be represented by the following equation：

$$IC_{m,i+1,t+1} = E_t\left(IC_{m,i+1,t+1}\right) + \varepsilon_{m,i+1,t+1}. \tag{6}$$

where:

$$E_t\left(IC_{m,i+1,t+1}\right) = g\left(z_{m,i,t}\right). \tag{7}$$

The IC values are calculated at time $t = 2, \dots, T$, while the momentum factors are calculated at time $i = t - 1 = 1, \dots, I$. The observation period for the momentum factor is as m = 1, 6, 12, 36, 60 respectively. ε represents the random error. It is assumed that all stock data is complete, and the issue of missing values is discussed in section5.1.

Our objective is to isolate a representation of $E_t\left(IC_{m,i+1,t+1}\right)$ as a function of predictor variables that maximizes the out-of-sample explanatory power for realized $IC_{m,i,+1t+1}$. The aim is to improve the accuracy of predicting IC values by appropriately modeling $E_t\left(IC_{m,i+1,t+1}\right)$, ensuring that this predictive ability is maximized not only within the sample (observed data) but also out-of-sample (unobserved data). Therefore, the functional form of $g(\cdot)$is left unspecified. Our target is to search for the prediction model from a set of candidates that gives the best prediction performance. The vector of predictors, $z_{m,i,t}$, consists of the momentum of all stocks at time point t-1 and the IC value at time t. The predictor variables $z_{m,i,t}$ include the momentum of all stocks sorted at time point t-1 and the IC value at time t. It can be represented as:

$$z_{m,i,t} = \begin{pmatrix} r(M_{t-1,m}) \\ IC_t \end{pmatrix}. \tag{8}$$

where $r(M_{t-1,m})$ is a 4983×1 vector of the momentum factors at time t-1, IC_t is a 1×1 vector of the IC value at time t-1. Each result is independent of the other. Also, $g(\cdot)$ depends on z only through $z_{m,i,t}$. This means our prediction does not use information from the history prior to t and the momentum factors prior to t-1. In total, we consider 7 machine learning methods, namely linear regression, random forest, and neural networks (NN1-NN5).

4.1 Simple Linear Regression Model

We begin with the simple linear regression model estimated via ordinary least squares (OLS). The simple linear model imposes those conditional expectations $g(\cdot)$ can be approximated by a linear function of the raw predictor variables and the parameter vector, θ,

$$g(z_{m,i,t}; \theta) = z'_{m,i,t}\theta. \tag{9}$$

This model imposes a simple regression specification and does not allow for nonlinear effects or interactions between predictors. The model combines the original predictor variables and the parameter vector by linearly. As a result, the predicted values of the model are weighted linear combinations of the original predictor variables, with the weights determined by the parameter vector θ.

Our baseline estimation of the simple linear model uses a standard least square, or "ℓ_2", objective function:

$$\mathcal{L}(\theta) = \frac{1}{IT}\sum_{i=1}^{I}\sum_{t=2}^{T}(IC_{m,i+1,t+1} - g(z_{m,i,t}))^2. \tag{10}$$

Minimizing $\boldsymbol{\mathcal{L}}(\theta)$ yields the pooled OLS estimator. The convenience of the baseline l2 objective function is that it offers analytical estimates and thus avoids sophisticated optimization and computation.

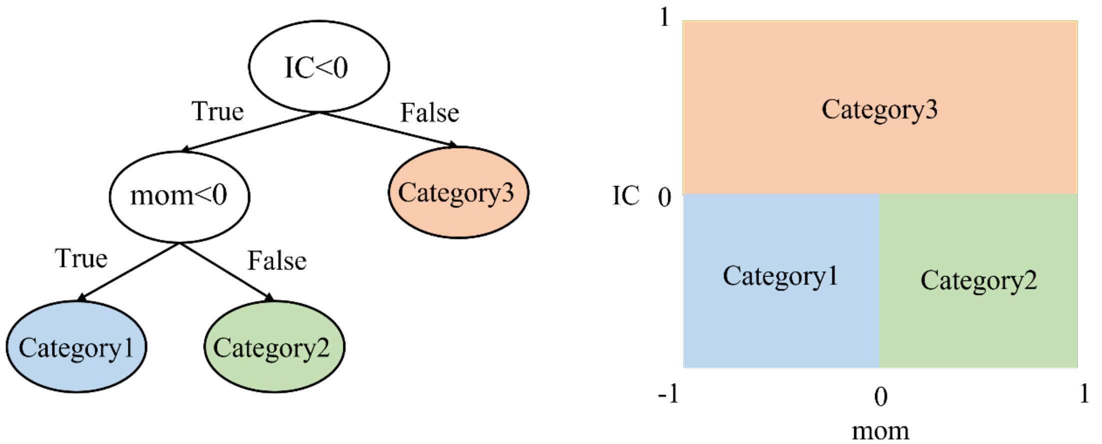
4.2 Random Forest

Simple linear regression captures individual predictors' nonlinear impact on expected returns but does not account for interactions among predictors. One way to add interactions is to expand the generalized model to include multivariate functions of predictors. However, without a priori assumptions for which interactions to include, the generalized linear model becomes computationally infeasible.

Random Forests offer an alternative approach to address this issue. Before discussing Random Forests, it is important to understand Regression Trees. regression trees have become a popular machine learning approach for incorporating multi-way predictor interactions. Unlike linear models, trees are fully nonparametric and possess a logic that departs markedly from traditional regressions. At a basic level, a tree aims to identify groups of samples with similar behaviors. A tree "grows" in a sequence of steps. At each step, a new "branch" sorts the data leftover from the preceding step into bins based on one of the predictor variables. This sequential branching slices the space of predictors into rectangular partitions and approximates the unknown function $g(\cdot)$

with the average value of the outcome variable within each partition.Figure 3 shows an example with two predictors, “IC” and “mom.” The left panel describes how the tree assigns each observation to a partition based on its predictor values. First, observations are sorted on the IC value. Those above the breakpoint of 0 are assigned to Category 3. Those with negative IC value are then further sorted by mom. Observations with negative IC value and mom below 0 are assigned to Category 1, while those with mom above 0 go into Category 2. Finally, forecasts for observations in each partition are defined as the simple average of the outcome variable’s value among observations in that partition.

Figure 3: Regression Tree Example



More formally, the prediction of a tree, T , with K “leaves” (terminal nodes), and depth L, can be written as:

$$g\left(z_{m,i,t};\theta,K,L\right)=\sum_{k=1}^{K}\theta_k 1_{\{z_{m,i,t}\epsilon C_k(L)\}}, \qquad (11)$$

where $C_k(L)$ is one of the K partitions of the data. Each partition is a product of up to L indicator functions of the predictors. The constant associated with partition k (denoted θ_k) is defined to be the sample average of outcomes within the partition.16 In the example of Figure 1, the prediction equation is:

$$g\left(z_{m,i,t};\theta,3,2\right)=\theta_1 1_{\{IC_{m,i,t}<0\}}1_{\{mom_{m,i}<0\}}+\theta_2 1_{\{IC_{m,i,t}<0\}}1_{\{mom_{m,i}\geq 0\}}+\theta_3 1_{\{IC_{m,i,t}\geq 0\}},$$

Random Forest is a model that consists of multiple regression trees created using bootstrapping, with the goal of reducing the correlation between different trees. The steps involved in constructing a Random Forest model are as follows：

Step 1：Randomly select M samples from the original dataset of size N using bootstrapping, allowing for the possibility of selecting the same sample multiple times.

Step 2：Construct a regression tree using the selected M samples. This involves recursively partitioning the data based on different predictor variables to create a tree structure.

Step 3：Repeat Steps 1 and 2 iteratively to generate a total of T regression trees. Each tree is built using a different bootstrap sample.

Step 4：For a new input, obtain predictions from each individual tree in the Random Forest.

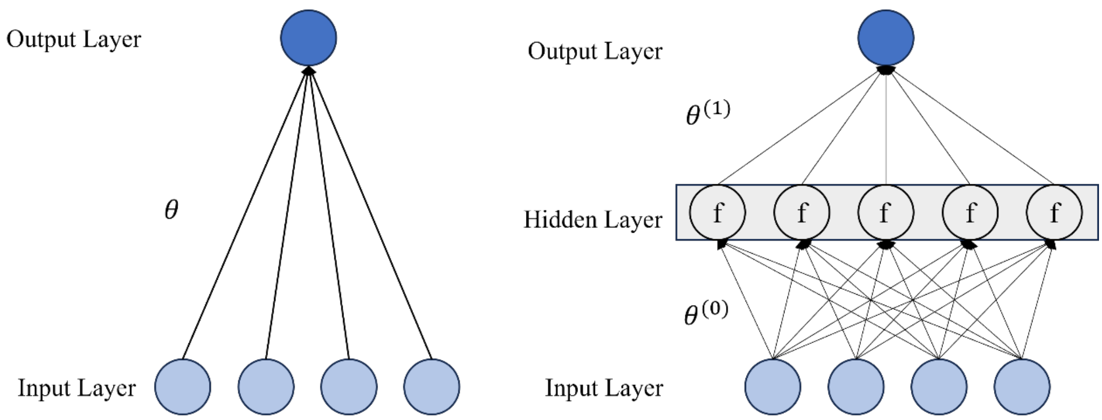
Step 5：Combine the predictions of the T trees, often using majority voting (for classification problems) or averaging (for regression problems), to obtain the final prediction.

In Random Forest, each individual regression tree is trained on a different subset of the data, ensuring diversity in the predictions. By combining the predictions of multiple trees through majority voting or averaging, Random Forest can provide more robust and accurate predictions compared to a single regression tree.

4.3 Neural Networks

Neural Networks consist of an "input layer" of raw predictors, one or more "hidden layers" that interact and nonlinearly transform the predictors, and an "output layer" that aggregates hidden layers into an ultimate outcome prediction. Analogous to axons in a biological brain, layers of the networks represent groups of "neurons" with each layer connected by "synapses" that transmit signals among neurons of different layers. Figure4 shows two illustrative examples.

Figure 4: Neural Networks



The number of units in the input layer is equal to the dimension of the predictors, which we set to four in this example (denoted z_1, z_2, z_3, z_4). The left panel shows the simplest possible network that has no hidden layers. Each of the predictor signals is amplified or attenuated according to a five-dimensional parameter vector, θ, that includes an intercept and one weight parameter per predictor. The output layer aggregates the weighted signals into the forecast$\theta_0 + \sum_{k=1}^{4} z_k \theta_k$; that is, the simplest neural network is a linear regression model.

The model incorporates more flexible predictive associations by adding hidden layers between the inputs and output. The right panel of Figure 4 shows an example with one hidden layer that contains five neurons. Each neuron draws information linearly from all of the input units, just as in the simple network on the left. Then, each neuron applies a nonlinear “activation function” f to its aggregated signal before sending its output to the next layer. For example, the second neuron in the hidden layer transforms inputs into an output as $x_2^{(1)} = f(\theta_{2,0}^{(0)} + \sum_{j=1}^{4} z_j \theta_{2,j}^{(0)})$. Lastly, the results from each neuron are linearly aggregated into an ultimate output forecast:

$$g(z;\theta) = \theta_0^{(1)} + \sum_{j=1}^{5} x_j^{(1)} \theta_j^{(1)}, \tag{12}$$

Thus, in this example, there are a total of $31 = (4+1) \times 5 + 6$ parameters (five parameters to reach each neuron and six weights to aggregate the neurons into a single output). We consider architectures with up to five hidden layers. Our shallowest neural network has a single hidden layer of 32 neurons, which we denoted NN1. Next, NN2 has two hidden layers with 32 and 16 neurons, respectively; NN3 has three hidden layers with 32, 16, and 8 neurons, respectively; NN4 has four hidden layers with 32, 16, 8, 4 neurons, respectively; and NN5 has five hidden layers with 32, 16, 8, 4, and 2 neurons, respectively. We choose the number of neurons in each layer according to the geometric pyramid rule. All architectures are fully connected so each unit receives an input from all units in the layer below. By comparing the performance of NN1 through NN5, we can infer the trade-offs of network depth in the return forecasting problem.

5 An Empirical Study of US Equities

A comprehensive analysis will be conducted by dividing all stocks into seven equally sized groups based on momentum. This results in 63 different portfolio performance outcomes when combined with the nine portfolio construction methods. Additionally, further classification can be done based on different prediction approaches using seven machine learning methods.

However, it is important to note that not all portfolio construction methods may have clear motivation or research significance. To ensure a focused and meaningful analysis, we will only present the performance of investment portfolios that are deemed meaningful and have sufficient justification based on the empirical analysis process. These portfolios will demonstrate robust absolute returns and will be further explained in terms of their research significance and motivation once the results are confirmed.

By focusing on the portfolios that show strong performance and have a clear rationale, we aim to provide insightful and meaningful findings that contribute to the understanding of investment strategies and their implications.

5.1 Data and Handling Missing Data

We obtain monthly individual closing price from CRSP. Our sample begins in January 1995 and ends in October 2022, totaling 334 months. We exclude all financial firms and utility firms (SIC codes between 6,000 and 6,999, and between 4,900 and 4,999, respectively). The number of stocks per month is 4,983. We also obtain the Treasury-bill rate to proxy for the risk-free rate from which we calculate individual excess returns.

Table 1: Statistics of Rank IC in different observation periods

Observation Periods	N	Minimum	25%	50%	75%	Maximum	Mean	SD
1 momth	334	-0.514	-0.118	0.008	0.128	0.421	-0.003	0.189
6 momth	329	-0.658	-0.131	-0.010	0.128	0.446	-0.007	0.196
12 momth	323	-0.604	-0.134	0.004	0.132	0.444	-0.010	0.200
36 momth	299	-0.567	-0.183	-0.034	0.099	0.515	-0.043	0.203
60 momth	275	-0.517	-0.191	-0.035	0.104	0.468	-0.046	0.205

Since not all companies are on the market at the earliest point in the data used, missing values may arise. Information Coefficients are employed as predictive indicators, and adding the mean value does not significantly affect the IC value for that month. To address the issue of missing values and enhance the dataset's representativeness of real data, thus improving the accuracy of the analysis, the mean imputation method is utilized. This method, a type of single imputation, involves calculating the average using the known stock prices for the remaining months of that specific month and filling in the missing data. It avoids the utilization of direct deletion methods, which can result in substantial data loss.

5.2 Portfolio Performance：Real IC Values

In this section, investment portfolios will be constructed based on the real IC values for the current month. The momentum of the five intervals (formation periods) will be calculated using logarithmic returns to obtain the IC values. The maximum absolute IC value will be selected, and stocks will be ranked based on their momentum within that time interval. The stocks will then be divided into different percentiles, including the top and bottom 1%, 2%, and 10%. If any stocks in the list have prices below $1, they will be excluded, and the next eligible stocks will be included to maintain the desired number of stocks (50, 100, or 500) in each group. Portfolios will be constructed based on these company lists. Positions will be entered at the end of the previous month and held for a fixed period of one month, exiting at the end of that month. Profits and losses during this period will be calculated as the returns from holding the portfolios for that specific month.

The performance of these portfolios will be observed and analyzed over a period of 100 months, starting from July 2014 to October 2022. Since the prices one month ahead cannot be known in advance, the actual composition of the portfolios at the time of entry cannot be obtained at the beginning of the month. The existence of these portfolios is based on the assumption of early knowledge of the IC values and their accuracy. The performance of the portfolios will be compared to the S&P 500 (Standard & Poor's 500), which has been tracking the average performance of the U.S. stock market since 1957. The S&P 500 includes 500 common stocks, representing approximately 80% of the total market value. It covers a wide range of sectors and is widely regarded as an index that closely reflects the overall market performance. The monthly returns of the portfolios will be compared to the returns of the S&P 500. If a portfolio's return exceeds that of the S&P 500, it implies two research implications: (1) the Information Coefficient (IC) can serve as a predictive indicator, and (2) provides strong motivation for early prediction of IC values and entering the market at the end of the previous month.

We include some assumptions. In the construction of the investment portfolio, an assumption is made that equal-weighted buying and selling of each stock within the portfolio is considered at the end of the month. However, it is important to note that practical difficulties may arise in achieving equal weighting due to stock prices and fractional shares. Additionally, the assumption is made that the circuit breaker mechanism in the US stock market is ignored. This means that buying and selling transactions can be executed smoothly, even if there are liquidity issues with certain stocks. Furthermore, it is assumed that there are no constraints, such as account closure or zero funds, when the investment portfolio incurs losses. Even if the return rate falls below -100%, the calculations will still be performed. Lastly, no transaction costs are considered in the analysis. This assumption implies that buying and selling stocks within the portfolio do not incur any fees or charges typically associated with transactions in the real market.

5.2.1 Portfolio Performance

Table 2 presents the performance of an investment portfolio constructed using real IC values with an equal-weighted approach, employing a buyTsellB strategy. This strategy involves purchasing the top-performing stocks and selling the bottom-performing stocks. The performance statistics include the average monthly return, monthly standard deviation, and Sharpe ratio, which is calculated using the risk-free rate based on the yield of the US 10-year Treasury bond. The data covers the period from July 2014 to

October 2022, encompassing 100 months. The results demonstrate that portfolios constructed using the top 1% and 2% momentum stocks exhibit significantly better average monthly returns and Sharpe ratio compared to the S&P 500 index. This indicates the potential benefits of early prediction of IC values and the construction of momentum strategies.Table 2: Real IC Stock-level buyTsellB Performance

buyTsellB	1%			2%			10%		
	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio
Portfolio	7.78%	9.30%	0.835	5.24%	7.51%	0.696	0.70%	4.43%	0.153
S&P500	0.75%	4.49%	0.163	0.75%	4.49%	0.163	0.75%	4.49%	0.163

Table 3: Real IC Stock-level buyT/buyB Performance

buyT/buyB	1%			2%			10%		
	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio
Portfolio	-0.93%	9.30%	-0.101	-1.01%	8.99%	-0.115	-1.08%	7.98%	-0.138
S&P500	0.75%	4.49%	0.163	0.75%	4.49%	0.163	0.75%	4.49%	0.163

Table 4: Real IC Stock-level sellB/sellT Performance

sellB/sellT	1%			2%			10%		
	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio
Portfolio	2.37%	10.80%	0.218	2.24%	9.33%	0.238	1.42%	6.75%	0.208
S&P500	0.75%	4.49%	0.163	0.75%	4.49%	0.163	0.75%	4.49%	0.163

Table 5: Real IC Stock-level TB.BT Performance

TB/BT	1%			2%			10%		
	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio
Portfolio	1.45%	12.07%	0.118	1.23%	9.09%	0.133	0.34%	4.47%	0.072
S&P500	0.75%	4.49%	0.163	0.75%	4.49%	0.163	0.75%	4.49%	0.163

A positive IC value indicates a positive relationship between the momentum factor and returns, while a negative IC value indicates a negative relationship. This characteristic is employed to examine whether the contrarian strategy yields abnormal returns. Table 3, Table 4, and Table 5 present the performance of the buyT/buyB, sellB/sellT, and TB/BT strategies, respectively. These strategies involve constructing portfolios using real IC values with an equal-weighted approach. In these tables, a momentum strategy is adopted if the real IC value for a particular month is positive, whereas a contrarian strategy is employed if the IC value is negative. Among these strategies, the buyT/buyB strategy underperforms the S&P 500 index. However, the sellB/sellT strategy exhibits better average monthly returns and Sharpe ratio compared to the S&P 500, particularly for portfolios constructed using the top 1%, 2%, and 10% momentum stocks. The TB/BT strategy also outperforms the S&P 500 in terms of average monthly returns for portfolios constructed using the top 1% and 2% momentum stocks. These findings support the motivation for early prediction of IC values and constructing both momentum and contrarian strategies based on positive and negative IC values.

5.2.2 Summary

Based on the analysis of the empirical results, several preliminary conclusions can be drawn：(1) The positive average monthly returns indicate that the investment portfolios constructed using IC values calculated from momentum factors have implied excess returns. (2) The portfolios constructed using real IC values assume perfect prediction of IC values at the end of the previous month. This assumption is essential for the previous conclusion. If it becomes possible to predict IC values accurately, and the investment portfolio constructed using the predicted values approaches the performance of the portfolio constructed using real IC values, there is a motivation to accurately predict IC values at the end of the previous month.

These conclusions suggest the potential to achieve similar performance to portfolios constructed using real IC values by accurately predicting IC values at the end of the previous month. Therefore, there is a motivation to develop methods that can accurately forecast IC values. Based on these findings, it is recommended to explore and develop techniques or models that can effectively predict IC values, as this has the potential to enhance investment portfolio performance.

5.3 Portfolio Performance：Predictive IC Values

In continuation of section 5.2.2, this section aims to explore the prediction of IC values using three machine learning methods: linear regression, random forest, and neural networks (specifically, NN1, NN2, NN3, NN4, and NN5). The objective is to categorize companies into top-performing and bottom-performing stocks based on the momentum factor, with the goal of generating investment portfolios that can potentially achieve abnormal returns in advance.

The accuracy of the machine learning models will be examined initially, and investment portfolios will be constructed using the models with high accuracy. Similar to the previous approach, the portfolios will be entered into the market at the end of the previous month, and their performance will be observed over a one-month holding period. A comparison will be made against the S&P 500 index to determine whether it is possible to obtain IC values earlier through the predictive power of machine learning. Additionally, different strategies will be examined to evaluate the implications of the momentum factor in generating implicit excess returns. By conducting this analysis, we aim to assess the viability of using machine learning techniques to predict IC values and determine their potential in improving investment strategies. The findings will contribute to the understanding of early prediction of IC values and the development of effective entry strategies for investment portfolios.

5.3.1 Machine Learning Model Setup

In this section, we outline the setup of the machine learning models used in the paper. The input variables for the models are the Ranked Momentum features, while the output variable is the IC value, which serves as the target data. Since there are five observation periods for momentum factors, each machine learning model will be trained five times using the corresponding IC values.

We divide the 334 months of data into training sample, validation sample, and the remaining 1 month for out-of-sample testing. The training sample consists of input variables from 1 to t-101 and output variables from 2 to t-100. The validation sample consists of input variables from t-101 to t-2 and output variables from t-100 to t-1. The length of the validation set is fixed at 100 months. The testing sample includes input variables at time point t-1. The predicted IC value, obtained from the model, corresponds to the output variable at time point t. Because machine learning algorithms are computationally intensive, we avoid recursively refitting models each month. Instead, we refit once every year as most of our signals are updated once per year. Each time we refit, we increase the training sample by one year.

We perform 100 predictions by rolling the time back 100 months. The interval between predictions is one month. It is important to note that the calculated win rate, average excess monthly return, cumulative return, and annualized return do not consider transaction costs, fees, and additional costs associated with shorting. By following this model setup, we aim to evaluate the performance of the machine learning models in predicting IC values and constructing investment portfolios.

5.3.2 Comparison of Machine Learning Model Predictions

We utilize the elements of the Confusion Matrix to calculate four evaluation metrics: Accuracy, Precision, Recall, and F1-Score. The elements of the Confusion Matrix are presented in Table 6 as follows:

Table 6: Confusion Matrix

Confusion Matrix		**TRUE**	
		Positive	**Negative**
Predict	**Positive**	**TP**	**FP**
	Negative	**FN**	**TN**

In the Confusion Matrix, the investment portfolios with positive returns are classified as “Positive,” and those with negative returns are classified as “Negative.”

“Predicted” represents the investment portfolios constructed based on predicted IC values, while “True” represents the investment portfolios constructed based on true IC values. The four elements of the Confusion Matrix are as follows:

- True Positive (TP): Predicted as positive and true value is positive.
- True Negative (TN): Predicted as negative and true value is negative.
- False Positive (FP): Predicted as positive but true value is negative.
- False Negative (FN): Predicted as negative but true value is positive.

These four elements can be used to calculate the four-evaluation metrics: Accuracy, Precision, Recall, and F1-Score.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}}, \tag{13}$$

where Accuracy represents the consistency of the predicted returns with the true returns in terms of their positive or negative direction among all outcomes.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{14}$$

where Precision is a metric that measures the proportion of accurately predicted positive returns out of all instances that were predicted as positive.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{15}$$

where Recall is a metric that measures the proportion of correctly identified positive returns out of all instances that were positive.

$$\text{F1} - \text{score} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}, \tag{16}$$

where F1-score is a metric that combines both precision and recall providing a balanced measure of the two in predicting positive returns. It considers both false positives (FP) and false negatives (FN) and is calculated as the harmonic mean of precision and recall. The F1-score ranges from 0 to 1, with a higher value indicating a better balance between precision and recall and a more accurate and reliable prediction model.

The statistics in Table 7, Table 8, and Table 9 demonstrate the consistency between predicted and actual returns for the buyTsellB strategy at the 1%, 2%, and 10% levels, based on the predictions of various machine learning models over 100 prediction runs. There is a higher probability of correct directional predictions, indicating the effectiveness of the models in capturing the momentum factors. Furthermore, it is observed that the accuracy tends to increase as the number of stocks included in the portfolio decreases. This suggests that constructing portfolios with a smaller number of stocks can potentially yield higher accuracy. In terms of average monthly returns shown in Table 7, Table 8, and Table 9, positive returns are achieved by constructing the buyTsellB strategy based on the predicted IC values. This highlights the explanatory power of using momentum factors to construct momentum strategies. Additionally, portfolios with a smaller number of stocks tend to generate higher returns. It is important to note that there is no significant difference in accuracy among the seven machine learning models used in the paper. This indicates that the choice of the specific machine learning algorithm does not greatly impact the accuracy of the predictions for the buyTsellB strategy.

Table 7: Statistics for the buyTsellB Strategy at 1% Level

1%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Accuracy	90.0%	90.0%	89.0%	87.0%	88.0%	87.0%	82.0%
Precision	93.0%	89.4%	91.0%	89.0%	90.9%	90.8%	88.4%
Recall	95.2%	100.0%	96.4%	96.4%	95.2%	94.0%	90.5%
F1-Score	94.1%	94.4%	93.6%	92.6%	93.0%	92.4%	89.4%

1%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Average Return	9.008%	9.760%	9.505%	9.793%	9.170%	8.638%	9.339%

Table 8: Statistics for the buyTsellB Strategy at 2% Level

2%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Accuracy	83.0%	87.0%	84.0%	84.0%	82.0%	84.0%	82.0%
Precision	89.0%	87.8%	86.5%	86.5%	87.1%	89.2%	87.1%
Recall	90.1%	97.5%	95.1%	95.1%	91.4%	91.4%	91.4%
F1-Score	89.6%	92.4%	90.6%	90.6%	89.2%	90.2%	89.2%

2%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Average Return	5.703%	6.258%	5.989%	6.624%	5.641%	5.204%	5.926%

Table 9: Statistics for the buyTsellB Strategy at 10% Level

10%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Accuracy	84.0%	88.0%	86.0%	84.0%	85.0%	83.0%	73.0%
Precision	88.7%	88.1%	87.7%	86.0%	90.4%	90.0%	75.0%
Recall	82.5%	91.2%	87.7%	86.0%	82.5%	78.9%	78.9%
F1-Score	85.5%	89.7%	87.7%	86.0%	86.2%	84.1%	76.9%

10%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Average Return	0.539%	0.880%	0.954%	0.837%	0.552%	0.397%	1.007%

The statistics in Table10, Table11 and Table12 reflect the accuracy of the TB/BT strategy at the 1%, 2%, and 10% levels based on predictions from various machine learning models, based on the predictions of various machine learning models over 100 prediction runs. It is observed that the accuracy of all four-evaluation metrics significantly decreases compared to the buyTsellB strategy. This suggests that incorporating the contrarian strategy using predicted IC values leads to lower accuracy in general. Additionally, there is no significant difference in accuracy among the seven machine learning models used in the paper, indicating that the choice of specific machine learning algorithms does not greatly impact the accuracy of predictions for the TB/BT strategy. Furthermore, the average monthly returns presented in Table10, Table11, and Tabl12 indicate that almost all average monthly returns are negative for the TB/BT strategy at the 1%, 2%, and 10% levels. This suggests that constructing the TB/BT strategy based on predicted IC values does not improve overall returns and may not be effective in generating positive returns. Incorporating the contrarian strategy using predicted IC values does not appear to yield favorable outcomes in terms of average monthly returns. These findings highlight the limitations and challenges associated with incorporating the contrarian strategy based on predicted IC values and emphasize the need for further research and exploration to improve the performance of such strategies.

Table 10: Accuracy Statistics for the TB/BT Strategy at 1% Level

1%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Accuracy	65.0%	51.0%	48.0%	58.0%	50.0%	54.0%	51.0%
Precision	70.7%	71.4%	50.0%	60.4%	53.1%	56.5%	53.2%
Recall	55.8%	9.6%	42.3%	55.8%	32.7%	50.0%	48.1%
F1-Score	62.4%	16.9%	45.8%	58.0%	40.5%	53.1%	50.5%

1%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Average Return	-0.881%	-9.723%	-0.212%	-0.874%	-3.877%	-0.787%	-2.099%

Table 11: Accuracy Statistics for the TB/BT Strategy at 2% Level

2%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Accuracy	60.0%	52.0%	46.0%	53.0%	48.0%	55.0%	55.0%
Precision	60.5%	54.5%	43.9%	52.3%	45.2%	54.8%	54.5%
Recall	53.1%	12.2%	36.7%	46.9%	28.6%	46.9%	49.0%
F1-Score	56.5%	20.0%	40.0%	49.5%	35.0%	50.5%	51.6%

2%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Average Return	-0.574%	-6.226%	-0.230%	-0.861%	-2.386%	-0.557%	-1.003%

Table 12: Accuracy Statistics for the TB/BT Strategy at 10% Level

10%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Accuracy	63.0%	55.0%	49.0%	51.0%	55.0%	56.0%	50.0%
Precision	59.3%	52.5%	46.0%	48.1%	52.0%	53.1%	47.2%
Recall	68.1%	44.7%	48.9%	55.3%	55.3%	55.3%	53.2%
F1-Score	63.4%	48.3%	47.4%	51.5%	53.6%	54.2%	50.0%

10%	OLS	RF	NN1	NN2	NN3	NN4	NN5
Average Return	0.299%	-0.900%	-0.046%	-0.009%	-0.191%	-0.020%	0.372%

5.3.3 Portfolio Performance

In Table13, the returns and win rates of 9 investment portfolios are compared, with each portfolio constructed based on the top and bottom 1% momentum stocks. The portfolios buyT/sellB, buyT, sellB, and buyTsellB consistently achieve win rates exceeding 50% across different machine learning models, indicating their effectiveness in constructing momentum strategies using predicted IC values. On the other hand, portfolios TB/BT, buyT/buyB, sellB/sellT, buyB, and sellT exhibit win rates below 50%, suggesting limited effectiveness when operations are based on positive and negative IC values or when implementing contrarian strategies. Table13 shows the average excess monthly returns of the 100 constructed portfolios over the period from July 2014 to October 2022. The buyTsellB strategy stands out with the highest average excess monthly return, indicating its significant abnormal returns compared to the S&P 500. These results in Table13 and Table14 complement each other, highlighting the consistent profitability of momentum strategies constructed using predicted IC values in the US stock market.

Overall, the findings emphasize the effectiveness of constructing momentum strategies based on predicted IC values, as demonstrated by the consistently positive returns and win rates of portfolios such as buyT/sellB, buyT, sellB, and buyTsellB. However, strategies involving contrarian operations or solely buying bottom-performing or selling top-performing stocks exhibit limited effectiveness in generating abnormal returns.

Table 13: Comparison Win Rates for 9 Investment Portfolios at 1% Level

	OLS	**RF**	**NN1**	**NN2**	**NN3**	**NN4**	**NN5**
TB/BT	38%	7%	39%	48%	31%	37%	36%
buyT/buyB	31%	16%	40%	33%	28%	34%	34%
buyT/sellB	56%	69%	72%	71%	66%	59%	64%
sellB/sellT	50%	25%	51%	46%	46%	48%	48%
buyT	55%	58%	55%	56%	52%	52%	55%
buyB	18%	16%	19%	18%	17%	19%	19%
sellT	35%	25%	34%	33%	35%	35%	29%
sellB	67%	70%	73%	71%	73%	69%	75%
buyTsellB	77%	86%	82%	87%	80%	79%	80%

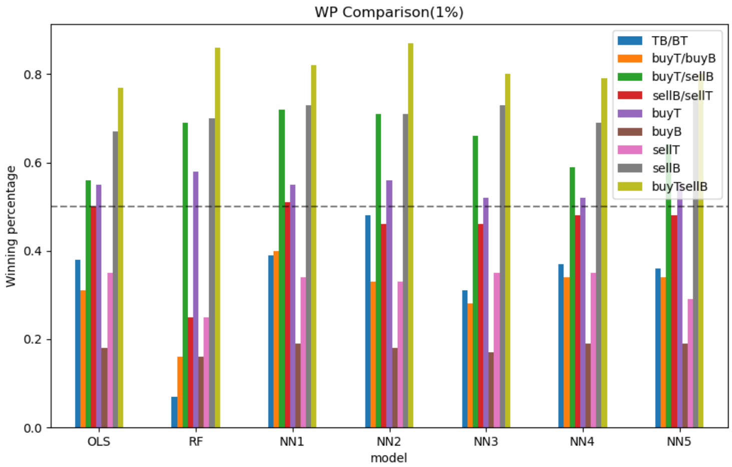
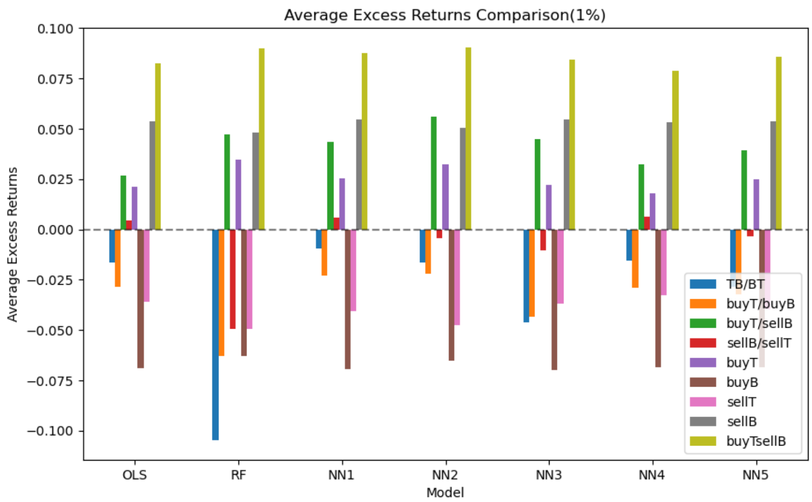


Table 14: Average Excess Monthly Returns at 1% Level

	OLS	RF	NN1	NN2	NN3	NN4	NN5
TB/BT	-1.32%	-6.97%	-0.98%	-1.61%	-3.13%	-1.31%	-1.75%
buyT/buyB	-2.57%	-4.78%	-2.43%	-2.21%	-3.66%	-2.68%	-2.94%
buyT/sellB	0.84%	3.23%	2.66%	3.89%	3.08%	1.65%	2.10%
sellB/sellT	0.49%	-2.94%	0.70%	-0.15%	-0.23%	0.62%	0.44%
buyT	0.57%	1.46%	0.68%	1.53%	0.36%	0.20%	0.52%
buyB	-5.13%	-4.80%	-5.31%	-5.09%	-5.28%	-5.00%	-5.40%
sellT	-2.07%	-2.96%	-2.18%	-3.03%	-1.85%	-1.70%	-2.02%
sellB	3.63%	3.30%	3.81%	3.59%	3.79%	3.50%	3.91%
buyTsellB	4.95%	5.51%	5.24%	5.88%	4.89%	4.46%	5.18%



Next, we will examine the four portfolios (buyT/sellB, buyT, sellB, and buyTsellB) that have significant positive returns.

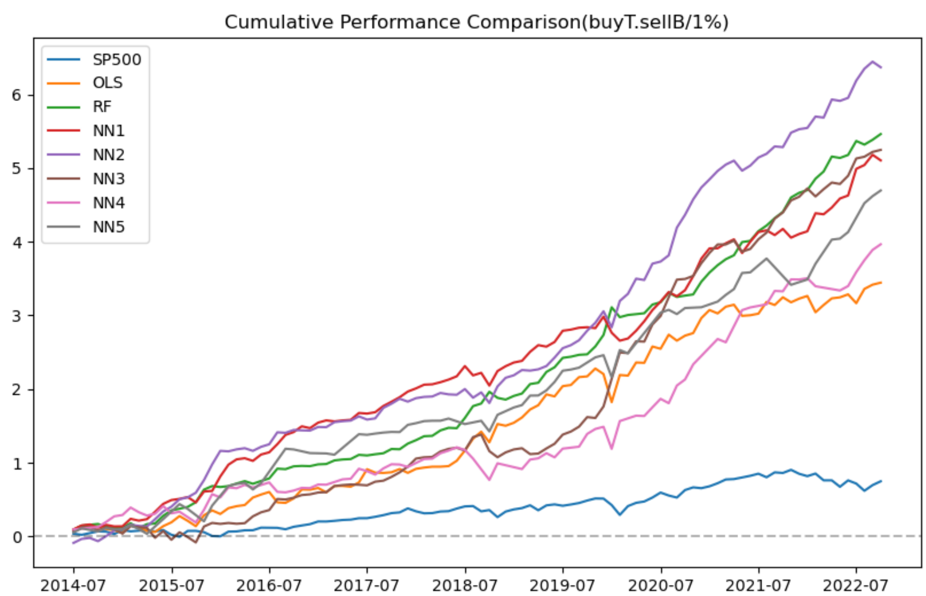
1. buyT/sellB

Table15 presents the performance of the buyT/sellB strategy constructed using machine learning models for the top and bottom 1%, 2%, and 10% portfolios. Across all machine learning models, the buyT/sellB strategy outperforms both the S&P 500 and the buyT/sellB strategy constructed using real IC values for the 1% and 2% portfolios. This suggests that incorporating predicted IC values from the machine learning models leads to improved performance. Furthermore, the portfolios with a smaller number of stocks exhibit higher excess returns. This implies that constructing portfolios with a concentrated selection of top-performing and bottom-performing stocks can potentially enhance returns. Among the machine learning models, the buyT/sellB strategy constructed using the NN2 model yields the most significant positive returns, indicating the effectiveness of this particular model for predicting IC values and generating excess returns. The buyT/sellB strategy operates based on the predicted IC values, where a positive IC value triggers a buyT operation, and a negative IC value triggers a sellB operation. By incorporating the observed positive and negative IC values from the machine learning models into the buyT/sellB strategy, significant absolute returns can be achieved. These results demonstrate the value of utilizing machine learning models for predicting IC values and constructing investment strategies that outperform the benchmark. The findings suggest that accurately predicting IC values can provide an advantage in achieving positive returns and improving the overall performance of investment portfolios.

Table 15: Performance of buyT/sellB Strategy Using Machine Learning Models

buyT/sellB	**1%**			**2%**			**10%**		
	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio
OLS	3.45%	10.00%	0.343	1.59%	9.08%	0.173	-0.96%	7.92%	-0.123
RF	5.46%	7.20%	0.756	3.98%	6.70%	0.591	1.27%	5.97%	0.209
NN1	5.11%	8.93%	0.570	3.41%	8.82%	0.385	0.60%	7.62%	0.076
NN2	**6.37%**	9.51%	0.668	**4.63%**	8.96%	0.515	**1.54%**	7.99%	0.191
NN3	5.25%	9.57%	0.546	3.83%	8.98%	0.425	0.92%	7.26%	0.124
NN4	3.97%	9.76%	0.405	2.39%	9.11%	0.261	-0.08%	8.31%	-0.012
NN5	4.70%	9.33%	0.502	2.84%	8.54%	0.331	0.59%	7.31%	0.079
real IC	2.83%	8.91%	0.315	1.48%	8.92%	0.164	-0.86%	8.00%	-0.110
S&P500	0.75%	4.49%	0.163	0.75%	4.49%	0.163	0.75%	4.49%	0.163

Figure 5: Cumulative Returns of buyT/sellB at 1% Level



2. buyT

Table16 presents the performance of the buyT strategy constructed using machine learning models for the top 1%, 2%, and 10% portfolios. Across all machine learning models, the buyT strategy outperforms both the S&P 500 and the buyT strategy constructed using real IC values for the 1% and 2% portfolios. This suggests that incorporating predicted IC values from the machine learning models leads to improved performance. Furthermore, portfolios with a smaller number of stocks exhibit higher excess returns, indicating the potential benefits of constructing concentrated portfolios of top-performing stocks. Among the machine learning models, the buyT strategy constructed using the Random Forest model yields more significant positive returns, suggesting its effectiveness in predicting IC values and generating excess returns.

Table 16: Performance of buyT Strategy Using Machine Learning Models

buyT	**1%**			**2%**			**10%**		
	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio
OLS	2.86%	10.22%	0.278	1.32%	9.83%	0.133	-0.97%	8.38%	-0.118
RF	**4.21%**	9.92%	0.423	2.21%	9.56%	0.229	-0.45%	8.39%	-0.056
NN1	3.31%	10.02%	0.329	1.43%	9.78%	0.144	-0.65%	8.41%	-0.079
NN2	4.00%	10.05%	0.397	2.28%	9.67%	0.234	-0.63%	8.47%	-0.076
NN3	2.94%	10.10%	0.290	1.11%	9.53%	0.114	-0.95%	8.20%	-0.117
NN4	2.54%	10.73%	0.235	0.95%	10.06%	0.093	-0.95%	8.51%	-0.114
NN5	3.23%	9.90%	0.325	1.27%	9.65%	0.130	-0.84%	8.04%	-0.106
real IC	2.24%	10.42%	0.214	1.00%	9.99%	0.098	-0.91%	8.38%	-0.110
S&P500	0.75%	4.49%	0.163	0.75%	4.49%	0.163	0.75%	4.49%	0.163

Figure6: Cumulative Returns of buyT and buyB Strategies at 1% Level

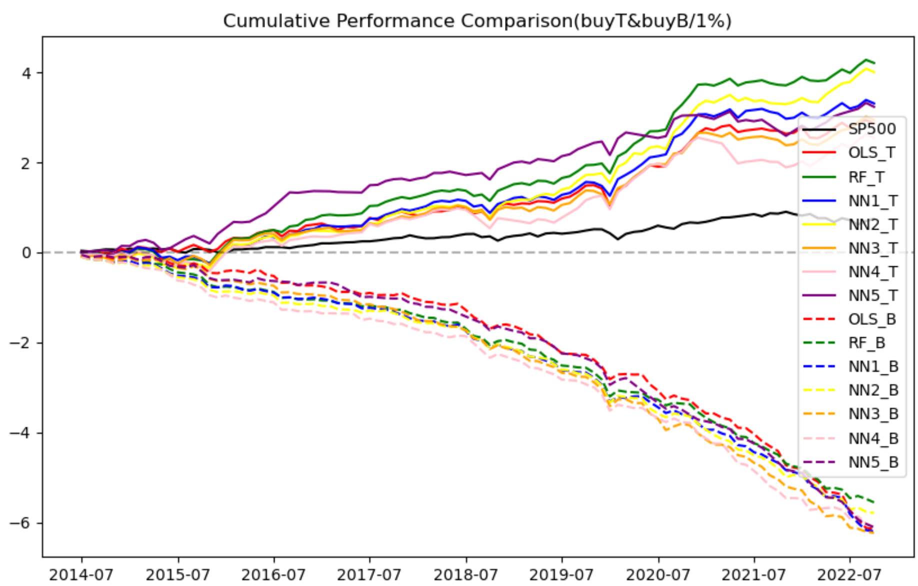


Figure6 illustrates the cumulative returns of the buyT and buyB strategies. It is evident that constructing momentum strategies based on predicted IC values obtained from machine learning models can generate absolute returns. The buyT strategy consistently outperforms the buyB strategy, indicating the effectiveness of the momentum strategy in capturing positive returns. However, it is noted that constructing contrarian strategies based on IC values, as represented by the buyB strategy, does not result in positive returns. This highlights the challenge and limitations associated with implementing contrarian strategies based on predicted IC values. Overall, the findings emphasize the effectiveness of constructing momentum strategies based on predicted IC values, as demonstrated by the positive excess returns of the buyT strategy. The performance of machine learning models plays a crucial role in accurately predicting IC values and generating profitable investment strategies.

3. sellB

Table 17 presents the performance of the sellB strategy constructed using machine learning models for the top 1%, 2%, and 10% portfolios. Across all machine learning models, the sellB strategy outperforms both the S&P 500 and the sellB strategy constructed using real IC values for the 1%, 2%, and 10% portfolios. This suggests that incorporating predicted IC values from the machine learning models leads to improved performance. Moreover, portfolios with a smaller number of stocks exhibit higher excess returns, indicating the potential benefits of constructing concentrated portfolios of bottom-performing stocks. Among the machine learning models, the sellB strategy constructed using the NN3 model generates slightly more significant positive returns, suggesting its effectiveness in predicting IC values and generating excess returns.

Table 17: Performance of sellB Strategy Using Machine Learning Models

sellB	**1%**			**2%**			**10%**		
	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio
OLS	6.15%	8.63%	0.711	4.38%	7.73%	0.564	1.51%	6.69%	0.223
RF	5.55%	7.15%	0.775	4.05%	6.66%	0.605	1.33%	5.95%	0.221
NN1	6.19%	7.82%	0.790	4.56%	7.37%	0.616	1.61%	6.47%	0.245
NN2	5.79%	7.61%	0.758	4.34%	6.99%	0.618	1.46%	6.15%	0.235
NN3	**6.23%**	8.20%	0.757	4.53%	7.54%	0.599	1.50%	6.44%	0.230
NN4	6.10%	7.79%	0.780	4.25%	7.13%	0.594	1.35%	6.70%	0.199
NN5	6.10%	7.54%	0.807	4.66%	7.33%	0.633	1.84%	6.44%	0.283
real IC	4.96%	8.80%	0.562	3.51%	7.94%	0.440	1.01%	6.47%	0.154
S&P500	0.75%	4.49%	0.163	0.75%	4.49%	0.163	0.75%	4.49%	0.163

Figure 7: Cumulative Returns of sellB and sellT Strategies at 1% Level

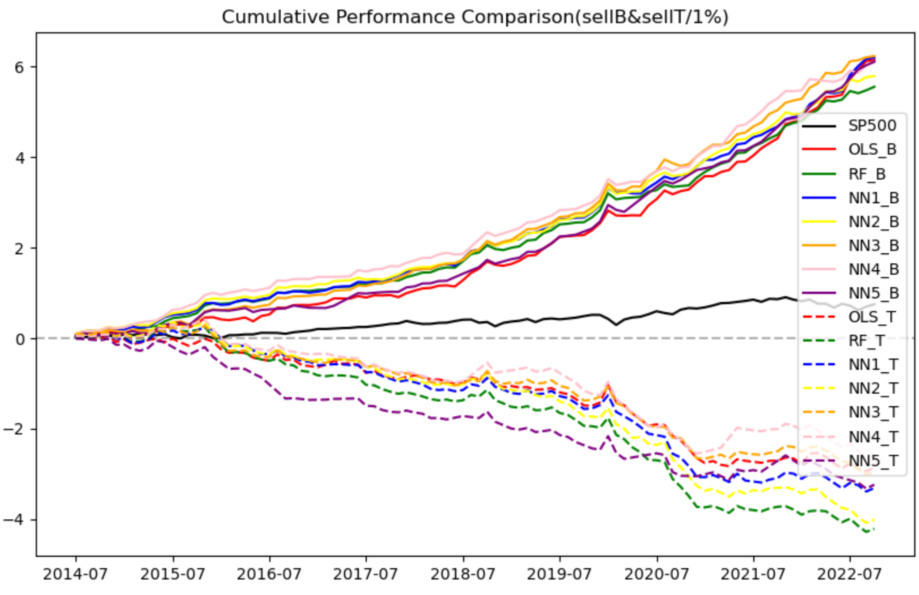


Figure 7 illustrates the cumulative returns of the sellB and sellT strategies. It is evident that constructing momentum strategies based on predicted IC values obtained from machine learning models can generate absolute returns. The sellB strategy consistently outperforms the sellT strategy, indicating the effectiveness of the momentum strategy in capturing positive returns. However, it is noted that constructing contrarian strategies based on IC values, as represented by the sellT strategy, does not result in positive returns. Overall, the findings emphasize the effectiveness of constructing momentum strategies based on predicted IC values, as demonstrated by the positive returns of the sellB strategy.

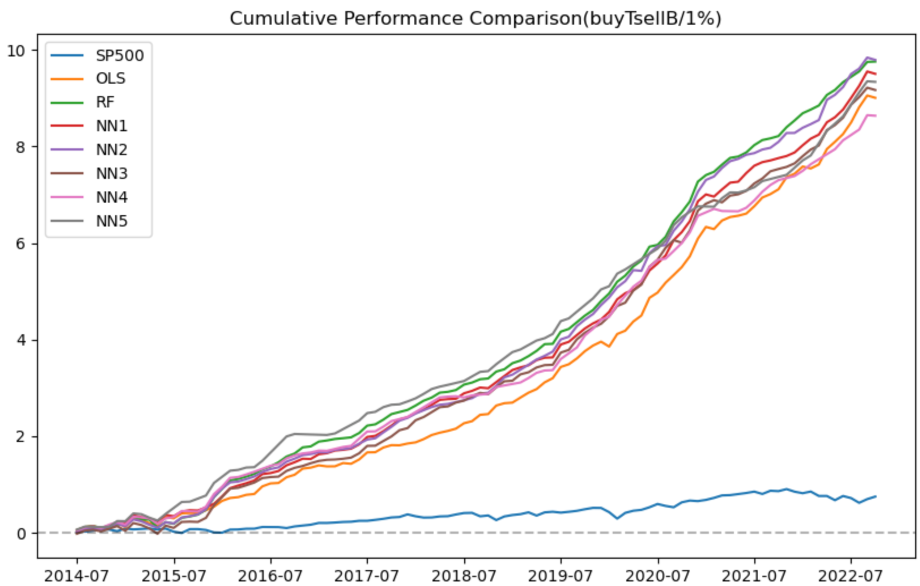
4. buyTsellB

Both implementing the buyT and sellB strategies individually can generate excess returns. We examine the performance when both strategies are implemented simultaneously. Table 18 presents the performance of the buyTsellB strategy constructed using machine learning models for the top 1%, 2%, and 10% portfolios. Across all machine learning models, the buyTsellB strategy outperforms both the S&P 500 and the buyTsellB strategy constructed using real IC values for the 1%, 2%, and 10% portfolios. This suggests that implementing the buyTsellB strategy based on observed IC values from machine learning models leads to improved performance compared to implementing buyT and sellB strategies separately. Among the machine learning models, the buyTsellB strategy constructed using the NN2 model generates slightly more significant positive returns, highlighting the effectiveness of this model in predicting IC values and generating excess returns. These findings indicate that implementing the buyTsellB strategy based on the observed IC values from machine learning models can generate absolute returns, and the returns are notably superior to implementing buyT and sellB strategies separately. This highlights the potential benefits of combining two momentum strategies in a single approach, leveraging the predictive power of machine learning models to enhance portfolio performance.

Table18: Performance of buyTsellB Strategy Using Machine Learning Models

buyTsellB	**1%**			**2%**			**10%**		
	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio	Average return	Standard deviation	Sharpe ratio
OLS	9.01%	9.52%	0.944	5.70%	7.11%	0.800	0.54%	3.98%	0.131
RF	9.76%	8.23%	1.184	6.26%	6.60%	0.946	0.88%	4.30%	0.201
NN1	9.50%	8.55%	1.110	5.99%	6.82%	0.875	0.95%	4.17%	0.225
NN2	**9.79%**	9.23%	1.059	**6.62%**	7.19%	0.918	0.84%	4.44%	0.185
NN3	9.17%	9.44%	0.969	5.64%	7.06%	0.796	0.55%	4.01%	0.134
NN4	8.64%	8.19%	1.053	5.20%	6.36%	0.815	0.40%	3.81%	0.100
NN5	9.34%	7.78%	1.199	5.93%	6.32%	0.935	1.01%	4.30%	0.230
real IC	7.78%	9.30%	0.835	5.24%	7.51%	0.696	0.70%	4.43%	0.153
S&P500	0.75%	4.49%	0.163	0.75%	4.49%	0.163	0.75%	4.49%	0.163

Figure 8: Cumulative Returns of buyTsellB Strategies at 1% Level



5.3.4 Summary

Based on the empirical results discussed above, the result confirms that constructing investment portfolios using machine learning models can generate significant positive returns. The buyT/sellB, buyT, sellB, and buyTsellB portfolios demonstrate strong performance under different machine learning models, with each portfolio achieving the most significant returns in specific scenarios. The buyT/sellB and buyT strategies exhibit notable returns when the portfolio includes stocks in the top and bottom 2% of all stocks, respectively. On the other hand, the sellB and buyTsellB strategies show significant returns when the portfolio contains stocks in the top and bottom 10% of all stocks. This suggests that momentum strategies can be effective, depending on the specific portfolio composition. Furthermore, the empirical results indicate that portfolios with a smaller number of stocks tend to generate better returns. This emphasizes the potential benefits of constructing concentrated portfolios, focusing on a select group of top-performing or bottom-performing stocks.

In contrast to the observation that the contrarian strategy using real IC values can yield abnormal returns, the result finds that the contrarian strategy implemented using machine learning models does not generate positive returns. This suggests that the predictive power of the machine learning models may not be strong enough to capture profitable opportunities with the contrarian approach. Moreover, the results indicate that the sellB strategy outperforms the buyT strategy in terms of returns, suggesting that shorting stocks with weak momentum yields more significant abnormal returns compared to longing stocks with strong momentum. The most significant returns are achieved when both shorting stocks with weak momentum and longing stocks with strong momentum are implemented simultaneously, as demonstrated by the buyTsellB strategy. Overall, the paper examines nine investment portfolios, and among them, four portfolios constructed using machine learning models consistently outperform the S&P 500 and the portfolios constructed using real IC values. This confirms the conclusion that constructing investment portfolios based on predicted IC values obtained from machine learning models can generate absolute positive returns. The findings highlight the potential of leveraging machine learning techniques for improved portfolio performance and the importance of considering both momentum and contrarian strategies in the investment decision-making process.

5.4 Comparison of Momentum Factors' Importance

To evaluate the impact of different momentum factors on the performance of various machine learning models, as described in Section 3.2, we selected the momentum factor with the highest absolute value of predicted IC values for five observation periods per month. We then calculated the frequency of usage for each momentum factor in 100 predictions. A higher frequency indicates a relatively greater importance of the momentum factor for the corresponding machine learning model.

Figure 9: Ranking of Momentum Factors' Importance for Machine Learning Models

Figure 9 presents the ranking of the importance of various momentum factors for different machine learning models, based on different observation periods. The vertical axis represents the momentum factors, with the factor that has the highest frequency of usage positioned at the top of the graph, and the factor with the lowest frequency at the bottom. The color gradient within each column indicates the ranking of the importance of a specific momentum factor for the corresponding machine learning model. The colors range from the least important (lightest color) to the most important (darkest color). From Figure 9, we can observe that the rankings of all six machine learning models, except for the linear regression model, are highly consistent. This consistency suggests a relative importance of different momentum factors that remains consistent across various machine learning models and different observation periods. The results indicate that longer observation periods of momentum factors show a stronger correlation with returns. Using longer observation periods to predict IC values for constructing investment portfolios leads to more significant results. This highlights the importance of considering longer-term momentum factors when predicting future performance and constructing effective investment strategies. This finding emphasizes the consistent and influential role of various momentum factors in the performance of different machine learning models. The results suggest that longer observation periods of momentum factors are more closely linked to returns. This underscores the significance of incorporating these factors into the prediction process for constructing successful investment portfolios.

6 Conclusion

We investigate the predictive ability of several machine learning models for constructing investment portfolios and analyze their performance using different variables and portfolio compositions. At a high level, our research findings indicate that the implemented process can significantly enhance excess returns. Machine learning helps improve momentum strategies by capturing strong correlations between stocks and their returns using information coefficients (IC). Shallow machine learning enables us to trace the sources of advantage and adapt to non-linear correlations that traditional methods might overlook. We confirm that a unidirectional momentum strategy can yield significant returns. Furthermore, we observe that implementing a bidirectional approach leads to even more pronounced gains. We find that shorting weak stocks outperforms longing strong stocks, indicating the persistence of the "winners keep winning, losers keep losing" phenomenon. Additionally, we observe that price reversals do not occur within a month. Optimizing the number of stocks in the portfolio leads to the highest excess returns within a certain threshold range. However, implementing inverse trades using machine learning models does not yield positive returns, indicating that the predictive ability of the models may be inadequate to identify profitable opportunities in these situations. Finally, we discover that momentum factors with longer holding periods are more powerful predictors, accurately capturing the correlation between stocks and their returns. We assume that momentum factors with shorter observation periods are less effective, possibly due to the effect of mean reversion. Our research underscores the importance of accurately predicting IC values and the effectiveness of momentum strategies in generating positive returns. Machine learning models can effectively apply momentum strategies, addressing challenges related to real-time responsiveness and stability. This empowers investors to enhance their decision-making and improve portfolio performance.

References

Barroso, P., & Santa-Clara, P. (2015). Momentum has its moments. *Journal of Financial Economics*, 116(1), 111-120.

Chan, L. K., Jegadeesh, N., & Lakonishok, J. (1996). Momentum strategies. *The Journal of Finance*, 51(5), 1681-1713.

Carhart, M. M. (1997). On persistence in mutual fund performance. *The Journal of Finance*, 52(1), 57-82.

Daniel, K., & Moskowitz, T. J. (2016). Momentum crashes. *Journal of Financial*

Economics, 122(2), 221-247.

Fama, E. F., & French, K. R. (1992). The cross-section of expected stock returns. *The Journal of Finance*, 47(2), 427-465.

Goodwin, T. H. (1998). The information ratio. *Financial Analysts Journal*, 54(4), 34-43.

Griffin, J. M., Ji, X., & Martin, J. S. (2003). Momentum investing and business cycle risk: Evidence from pole to pole. *The Journal of Finance*, 58(6), 2515-2547.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5), 2223-2273.

Hernández-Orallo, J., Flach, P., & Ferri Ramírez, C. (2012). A unified view of performance metrics: Translating threshold choice into expected classification loss. *Journal of Machine Learning Research*, 13, 2813-2869.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *The Journal of finance*, 45(3), 881-898.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of finance*, 48(1), 65-91.

Kinney, J. B., & Atwal, G. S. (2014). Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences*, 111(9), 3354-3359.

Levy, R. A. (1967). Relative strength as a criterion for investment selection. *The Journal of finance,* 22(4), 595-610.

Leippold, M., Wang, Q., & Zhou, W. (2022). Machine learning in the Chinese stock market. *Journal of Financial Economics*, 145(2), 64-82.

Shleifer, A., & Vishny, R. W. (1997). The limits of arbitrage. *The Journal of finance*, 52(1), 35-55.