Paper Title:

The Application of ESG Factor and Machine Learning in Alpha Strategy for Stock Selection

Authors:

Liu, Chen-Han, Kuang, Xian-Ji, Lai, Xi,-En, & Lin, Shih-Kuei

Status (11/26/23):

Working paper for submitting to The Journal of Portfolio Management

THE APPLICATION OF ESG FACTOR AND MACHINE LEARNING IN ALPHA STRATEGY FOR STOCK SELECTION

ABSTRACT

In an era where sustainable investment aligns with the pursuit of optimal returns, our research integrates traditional finance methodologies with modern technology. We utilize ESG ratings for stock selection within the alpha strategy and evaluate the predictive power of various machine learning models using corporate financial data. We discover that the added return from ESG ratings wanes after their public disclosure. Furthermore, the integration of the alpha strategy with a value-weighted approach consistently outperforms a sole long-only strategy, particularly during heightened systematic risk periods. Notably, models such as Random Forest and XGBoost effectively forecast ESG ratings in advance, contributing to the creation of portfolios that yield stable returns. Conclusively, machine learning models identify correlations between financial data and ESG ratings, thus facilitating the construction of portfolios that consistently realize absolute returns.

Keywords: Machine Learning, ESG, Factor investing, Alpha strategy, Short selling.

INTRODUCTION

Investment strategies have perpetually evolved in response to the demands of investors aiming for higher returns while safeguarding against risks. Despite the myriad of methodologies available, systemic risk remains a looming concern, often undermining absolute returns. In this complex milieu, Hedge funds, especially those leveraging the Alpha strategy, have emerged as vital players. By adeptly navigating investments and harnessing the potential of short-selling, the Alpha strategy seeks to neutralize systemic risks, optimizing returns in the process.

This research endeavors to merge sophisticated machine learning models with the Alpha strategy. Recognizing the mounting significance of ESG (Environmental, Social, and Governance) investing, we utilize relative ESG scores of firms to identify both outperforming and underperforming Alphas. A detailed examination of the short-selling potential of underperforming stocks further enables us to offer insights into systemic risk mitigation. The overarching aim is to unearth entities that promise excess returns suitable for portfolio incorporation.

Tracing the origins of the Alpha strategy, one encounters the foundational Capital Asset Pricing Model (CAPM). Conceived in the 1960s by luminaries like Treynor, Sharpe, Linter, and Mossin, CAPM is predicated on the idealistic belief that in a harmonized market, diversified securities yield returns directly linked to their systematic risk (Sharpe, 1964). Real-world complexities, however, such as taxation, transaction costs, and differing market expectations often introduce inefficiencies, creating opportunities for the discerning investor.

Our research contributes to the discourse by probing the nuances of portfolio construction in ESG investing. We find that value-weighted portfolios often outperform their equal-weighted counterparts. This observation can be attributed to the inherent advantage larger firms possess in adhering to ESG standards, given their extensive resources. Another notable contribution pertains to the monthly performance trends of ESG portfolios. Relying on insights from Siew et al. (2016), we notice that December returns for ESG portfolios typically surpass those in January. This anomaly suggests potential information asymmetry, where specific institutional investors might access pivotal ESG data before its widespread release.

The concept of Alpha further evolved with Jensen's 1967 work, highlighting its capacity to measure an investor's skill in outperforming expectations set by systemic risk (Jensen, 1967). The pursuit of absolute returns, a hallmark of hedge funds, aligns seamlessly with the Alpha strategy's ethos. Subsequent research, such as that by Jacobs & Levy in 1993, refined our understanding of this strategy, distinguishing it from mutual funds' objective of surpassing market indices (Jacobs & Levy, 1993).

The global investment scene, while diverse, is increasingly aligning with the principles of ESG investing. This shift is substantiated by a 2021 UNCTAD report, which recorded a remarkable growth in ESG-focused ETFs (UNCTAD, 2021). Numerous studies, including Khan's 2019 research, support this trend by highlighting superior returns for stocks with elevated ESG ratings (Khan, 2019). Moreover, during the challenging COVID-19 era, Engelhardt et al. (2021) found that firms with commendable ESG scores not only offered higher returns but also demonstrated reduced volatility.

A parallel transformative force in modern finance is Machine Learning. Advanced algorithms, continually evolving through data processing and iterative learning, hold immense promise. The vastness and dynamism of financial data make it a fertile ground for these predictive models. Several studies, from Roko & Gilli's 2008 research (Roko & Gilli, 2008) to more recent explorations by Tan et al. (2019) and Vijh et al. (2020), underscore the potential of machine learning in enhancing returns through accurate predictions.

In summation, our study occupies a unique intersection, combining traditional finance paradigms, the burgeoning realm of ESG, and the transformative capabilities of machine learning. Through this integrative approach, we aim to provide a nuanced perspective on current investment practices while illuminating pathways for future endeavors. By positioning our research at the confluence of these influential domains, we also seek to address gaps in current literature. One such gap pertains to the differential performance of ESG portfolios across distinct timeframes. Drawing from insights derived from Siew et al., 2016, our analysis ventures into the monthly discrepancies in returns (Siew et al., 2016). The curious observation that December returns for ESG portfolios generally surpass January's underscores a potential information advantage enjoyed by certain institutional players.

Moreover, in the intricate tapestry of modern investment strategies, the role of machine learning cannot be overstated. While the initial explorations in this domain, such as the 2008 study by Roko & Gilli, demonstrated the potential of basic algorithms, the evolution of this field has been rapid. Advanced models, from Random Forests to the potent XGBoost, have broadened the horizons of predictive financial analytics. Our research harnesses some of these advanced models, endeavoring to merge the power of machine learning with the robustness of financial data. Such a synthesis aims to provide a more holistic understanding of stock performances, especially in the context of ESG ratings.

Furthermore, as we integrate machine learning into our analysis, we place emphasis on the criticality of data selection. Echoing the findings of Walczak in 2001 (Walczak, 2001) and Adebiyi et al. in 2012 (Adebiyi et al., 2012), we posit that the sheer volume of data isn't always advantageous. Instead, it's the judicious selection of pertinent data, combined with the strategic employment of machine learning algorithms, that promises the most insightful outcomes.

Lastly, our investigation into the Alpha strategy's unique alignment with ESG principles underscores the evolving dynamics of the investment world. While ESG principles introduce an added layer of complexity, they also present new avenues for informed decision-making. Our study, by delving into the interplay between ESG scores, portfolio construction, and machine learning predictions, hopes to offer actionable insights to both seasoned investors and those new to the realm.

To conclude, through a comprehensive and integrative approach, our research seeks to further the discourse on ESG investing, machine learning's role in finance, and the enduring relevance of the Alpha strategy. By addressing both contemporary challenges and emerging opportunities, we aim to offer a forward-looking perspective, setting the stage for subsequent studies in this multifaceted domain.

LITERATURE REVIEW

CAPM and Alpha

Treynor (1962), Sharpe (1964), Linter (1965a, 1965b), and Mossin (1966) established the foundation of the Capital Asset Pricing Model (CAPM) in the 1960s, grounded in the perfect market hypothesis. This model posits that, upon reaching equilibrium and with the successful diversification of unsystematic risk via asset allocation, securities' expected returns linearly correlate with their systematic risk, embodying the reward-risk dynamic. However, real-world deviations from this idealized hypothesis—arising from taxes, transaction costs, and varied investor expectations—introduce market inefficiencies. This divergence underscores the drive to devise investment strategies aimed at maximizing returns while mitigating risks.

Jensen (1967) introduced the concept of "alpha" to gauge a fund's performance. Essentially, alpha signifies the deviation from the efficient expected return, serving as an indicator of an investor's stock selection prowess. A higher alpha denotes superior excess return, reflecting the investor's competency.

In a novel approach, our research integrates Environmental, Social, and Governance (ESG) considerations with the Alpha strategy, employing a machine learning framework. This confluence of previously separate areas provides a comprehensive view. Unlike earlier studies that assessed ESG or Alpha strategies in isolation, our methodology delves into their combined impact on excess returns. Leveraging machine learning, we adeptly incorporate ESG metrics into the Alpha strategy, refining predictive accuracy and presenting a cutting-edge framework for portfolio enhancement.

Alpha Strategy

Alpha strategy's objectives diverge considerably from those of mutual funds. While mutual funds seek returns surpassing market benchmarks, the alpha strategy emphasizes the eradication of systematic risk. This ethos aligns with hedge funds' pursuit of absolute returns, giving rise to the Long/Short equity strategy. This strategy is adept at insulating investment returns from market volatilities, thus neutralizing systematic risk.

Jacobs & Levy (1993) demarcate the Long/Short equity strategy into three distinct categories: market-neutral, equitized, and hedge strategies. Notably, the market-neutral approach is least susceptible to market fluctuations, exhibiting the lowest market correlation.

Grinold & Kahn (2000) illuminate the optimal demographic for alpha strategy deployment and the conditions for its successful application. They posit that proficient fund managers, skilled in discerning high-potential buy and sell stock opportunities, stand to benefit most. The prowess of these managers directly augments alpha, a pivotal metric of portfolio performance. Contrasted with the Long strategy, where short selling is circumscribed, the alpha strategy capitalizes more effectively on the insights fund managers garner. Its merits become pronounced when managing sizable assets, amidst low volatility and heightened active risk.

For a comparative analysis of alpha strategy funds and a gamut of hedge fund categories—including event-driven, global macro, sector-specific, and funds of funds—Edwards & Caglayan (2001) provides an enlightening reference. Their examination spans the 1990-1998 period, juxtaposing the bear and bull market performances of commodity and assorted hedge funds.

During this epoch, the market-neutral strategy, one among various hedge fund methodologies, consistently delivered superior results in both Sharpe ratio and M2 metrics as a standalone asset. When integrated into a broader portfolio, it elevated the Sharpe ratio the most, overshadowing other hedge fund types. Evaluating performances across economic cycles, market-neutral funds secured the pinnacle Sharpe ratio position as a solitary asset in bull markets, trailing only short-sell funds in bearish terrains. Given the latter's subpar bull market performance, one can deduce that market-neutral funds offer the most consistent and commendable returns throughout economic cycles.

ESG Factor Investment

Research into the interplay between ESG (Environmental, Social, and Governance) and corporate financial performance is burgeoning. Given ESG's relative novelty, the temporal scope and company samples in existing literature are somewhat constrained. Nonetheless, there's a growing consensus: adopting ESG metrics can enhance portfolio returns.

Khan (2019) introduced a novel methodology for quantifying the Governance aspect within ESG for global enterprises. Using this refined score as a stock selection criterion and adopting a Long/Short equity strategy, portfolios constructed around high-scoring entities yielded notable excess returns. When the Governance score was updated, the strategy continued to produce superior returns.

Deng & Cheng (2019) centered their study on China's A-share listed companies. Their findings indicated a positive relationship between a firm's ESG scores and stock performance, with this correlation being particularly pronounced for non-state-owned enterprises. Interestingly, ESG's influence was more potent in the secondary sector compared to the tertiary.

Engelhardt et al. (2021) assessed 1,452 European firms across 16 nations. Their results illuminated that entities with elevated ESG ratings experienced both heightened returns and dampened volatility during the COVID-19 upheaval. Upon deconstructing the ESG components, the social dimension emerged as paramount, substantiating that in crises, corporate social commitments are not just ethical gestures—they're financially rewarding.

Our investigation diverges by synergizing ESG tenets with the Alpha strategy, offering a fresh lens to the ESG discourse. Whereas earlier studies leveraged ESG to augment portfolio outcomes, we seamlessly embed these metrics into the Alpha strategy, forging a new frontier in ethically-informed, risk-adjusted returns. Employing machine learning, we probe the intricacies of ESG-enriched Alpha paradigms, spotlighting their potential for both portfolio resilience and superior performance.

Machine Learning and Finance

Machine learning, a subset of artificial intelligence, harnesses computational algorithms that iteratively enhance through data assimilation and experiential learning. Models within this domain span Decision Trees (DTs), Random Forest (RF), Artificial Neural Network (ANN), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost). This segment delineates prominent studies linking these models to financial applications.

Walczak (2001) employed Neural Networks to predict foreign exchange rates, discovering that an influx of data doesn't invariably enhance predictive precision. Superfluous data merely escalates costs and experimental duration. A two-year data span emerged as optimal.

While early Artificial Neural Network models predominantly utilized technical indicators as inputs, Adebiyi et al. (2012) incorporated fundamental metrics to a feedforward neural network model, aiming to forecast stock prices. This fusion notably bolstered predictive capabilities, affirming the pertinence of fundamental data to stock projections.

Tan et al. (2019) crafted two Random Forest models predicated on fundamental/technical and stock momentum factors from the Chinese stock market, targeting both short and long-term stock price predictions. Backtest outcomes showcased Sharpe ratios of 2.75 for the fundamental/technical model and 5 for the momentum model, both yielding remarkable excess returns.

Endri et al. (2020) utilized four Support Vector Machine models to forecast the delisting of Islamic stocks. Post-training on financial metrics, these models exhibited prediction accuracies spanning 71% to 100%, underscoring the efficacy of the Support Vector Machine in stock prognostications.

Vijh et al. (2020) harnessed both Neural Networks and Random Forest to anticipate the subsequent day's closing prices for renowned firms like Nike and JP Morgan, based on technical indicators. The Neural Network outperformed in metrics like Root Mean Square Error, Mean Absolute Percentage Error, and Mean Bias Error, registering values of 0.42, 0.77, and 0.013 respectively.

In our own research, we harmoniously weave the forecasting strength of machine learning with ESG perspectives within the Alpha framework. While earlier studies have delved into the potential of machine learning in financial predictions, we are at the forefront of integrating ESG elements into Alpha strategies. This pioneering synthesis illuminates the intricate interplay between ESG dynamics and Alpha-centric portfolio optimization, offering an innovative pathway for ethical portfolio management that transcends traditional paradigms.

In essence, our research artfully fuses CAPM, Alpha, ESG, and machine learning, presenting a promising avenue for investments that not only target robust returns but also exhibit resilient performance.

METHODOLOGY

Research Process

We divide this study into two sections. The first section examines ESG ratings as the filter for stock selection. Instead of using ESG score predictions, we employ real ESG ratings to separate high ESG rating companies from low ESG rating companies. The second section forms portfolios based on these two groups to see if the actual ESG rating is an ideal factor for investment, applying the performance of the portfolio as our measure. The purpose of this section is to test whether ESG ratings contain an implied excess return that can help differentiate strong buy stocks from strong sell stocks and thus find the best portfolio combination. Figure 1 illustrates the research processes of the first section.

The motivation for the second section comes from the preliminary conclusion in the first section. If portfolios, built from using real ESG scores, in the first section exhibit outstanding performances, then if we indeed can predict real ESG ratings with machine learning models, then we will be able to identify high ESG rating companies and low ESG rating companies in advance and build portfolios with performances approximate to what are built in the first section. Figure 2 shows the research processes of the second section.

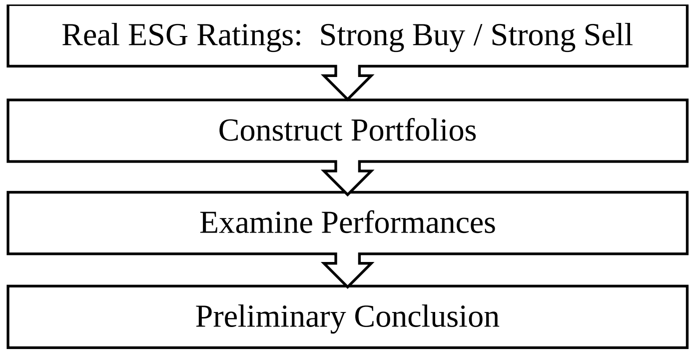

Figure 1: Real ESG rating research process

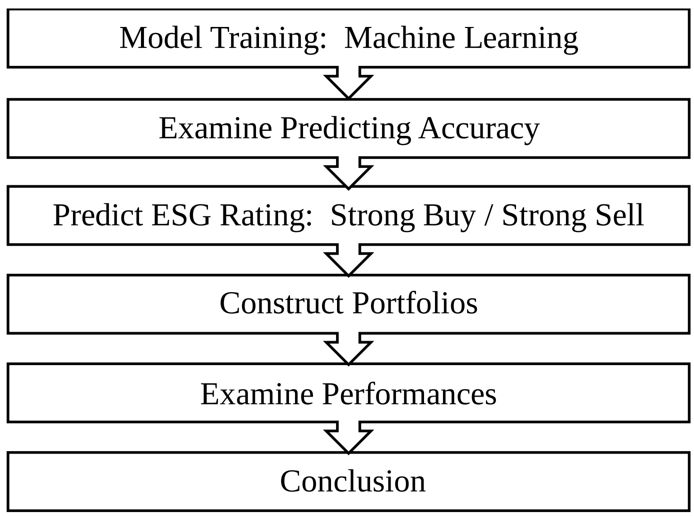

Figure 2: ESG rating prediction research process

Decision Trees (DTs)

The Decision Tree is a fundamental supervised machine learning algorithm with a tree-based model. It has two types: the classification tree used in this study and the regression tree. The basic structure of a Decision Tree starts from a root node, where all the data are unclassified. The root node connects to internal nodes, where data are diverted based on its feature, and to leaf nodes, which are the classification results.

The algorithm proceeds by establishing, growing, and pruning the Decision Tree. The Decision Tree set-up begins with the root node and splits and connects to two internal nodes. The algorithm sends the data to the correct internal node by interpreting the data’s attributes and classifies data based on the chosen attribute. We repeat this process until the conditions set up are met. The goal of minimizing the Gini coefficient after splitting the data determines the decision of classification based on which attribute at what value. The formula is presented below.

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

Dataset W can be divided into n categories, where p_i stands for the distribution probability of the i^{th} category. Let us assume that dataset D has attributes A, B, and C and is split into D_1 and D_2 based on attribute A. The Gini Coefficient for the split is as follow.

$$Gini(D, A) = \frac{D_1}{D} Gini(D_1) + \frac{D_2}{D} Gini(D_2)$$

The Decision Tree chooses the attribute with the smallest value among $Gini(D, A)$, $Gini(D, B)$, and $Gini(D, C)$ as the criteria for dataset D classification.

The purpose of pruning is to prevent overfitting, and we can separate pruning into pre-pruning and post-pruning. The former observes the internal nodes' contribution of splitting for the Decision Tree and decides whether to stop the split and replace the internal node with the leaf node when the Decision Tree is growing. The latter checks whether the performance can be improved by replacing an internal node with a leaf node. Postpruning starts from the leaf node to the internal node after the decision tree is completely established, and replaces internal nodes with leaf nodes if the replacement can improve classification ability.

Random Forest (RF)

The Random Forest methodology employs multiple decision trees to classify a subset of data randomly sampled from the entire training dataset. Two methods are adopted for randomly constructing datasets: sample randomization and feature randomization. Sample randomization involves randomly drawing *n* samples with replacements from the *N* training data samples to construct the training dataset for the decision tree. Feature randomization entails selecting *m* features from the *M* total features to calculate the optimal branching benefit during the generation of the tree branches, where $n \leq N$ and $m \ll M$. Ultimately, the individual decision trees, once trained, vote to determine the final classification result. This approach's advantage lies in its ability to mitigate the impact of a single erroneous decision tree on the overall outcome and reduce the chances of overfitting, leveraging the strengths of multiple trees to address collective shortcomings. The basic outline of the algorithm is as follows:

1. N stands for the number of training data samples; M stands for the number of features.
2. Sampling with replacement n times from N training data samples to form n training sets. Those not picked are used to examine the predicting result.
3. m features are randomly selected in every split of the node. The best result among the m features is calculated.
4. Pruning is not needed.

Artificial Neural Network (ANN)

Artificial Neural Network is a model that imitates a biological neural structure and can automatically modify and train itself through a comparison of input feature and output result. The structure appears in Figure 3 and consists of three main layers.

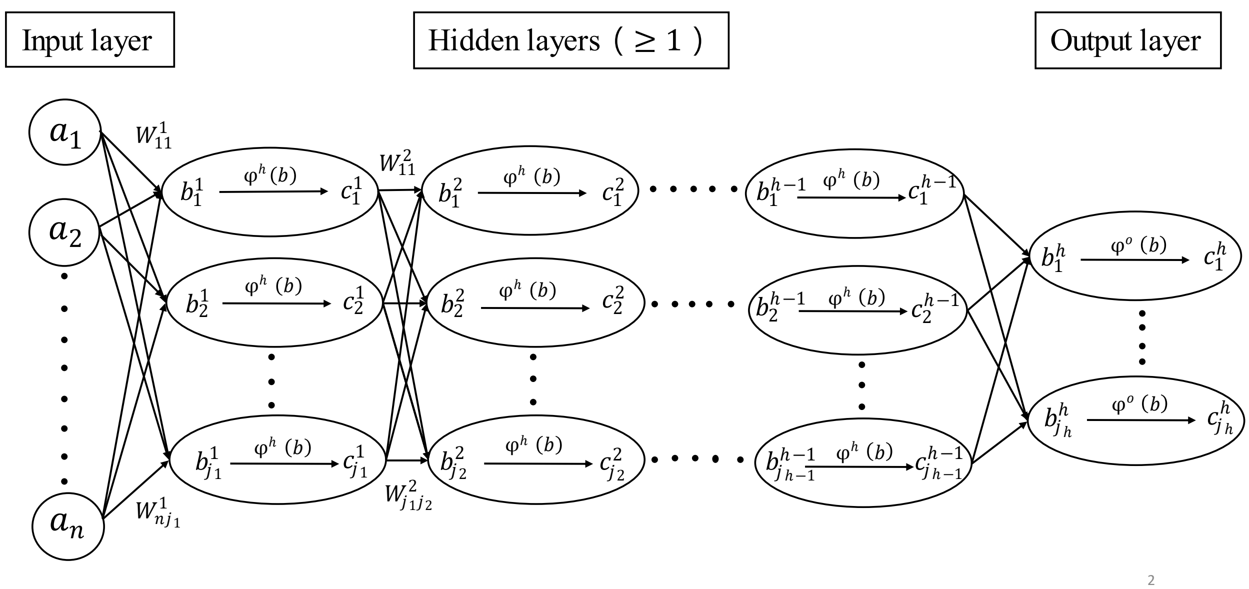

Figure 3 : Artificial Neural Network

$$b_j^1 = \sum_{i=1}^{n} a_i \times W_{ij}^1 + B_j^1 \, , j = 1,2, \ldots j_1$$

$$b_j^H = \sum_{i=1}^{j_{H-1}} c_i^{H-1} \times W_{ij}^H + B_j^H \, , j = 1,2, \ldots j_H \, , \; H = 2,3,4 \ldots , h$$

$$c_j^H = \varphi^h\left(b_j^H\right)\ j = 1,2, \ldots j_H\ ,\ H = 1,2, \ldots , h-1$$

$$c_j^h = \varphi^o\left(b_j^h\right)\ j = 1,2, \ldots j_h$$

Here, a_i stands for the i^{th} feature of the input value, W_{ij}^1 and W_{ij}^H stand for weights, c_j^h is the prediction result, and $\Phi^h\left(b_j^H\right)$ is the activation function for the hidden layers in which we use Rectified Linear Unit (ReLU) in this study.

$$\varphi^h(b) = \max(0,\ b)$$

$\varphi^o\left(b_j^h\right)$ is the activation function for the output layer in which we use Softmax in this study.

$$\varphi^o(b) = \frac{\exp(b)}{\sum_{j=1}^{j_h} \exp\left(b_j^h\right)}$$

During the iteration, the predicted results are put into the Cross Entropy loss function for model optimization.

$$Cross\text{ Entropy Loss} = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{j_h} y_{i,k}\log\left(P_{i,k}\right)$$

If the i^{th} input data are supposed to be classified to the k^{th} category, then $y_{i,k} = 1$; else $y_{i,k} = 0$. $P_{i,k}$ stands for the probability of i^{th} input data being classified into the k^{th} category.

The whole training process is based on the comparison of the predicted result and the target result through a loss function. The steps involve the following: 1) Set up the learning rate. 2) Modify the weights after calculating the contribution of each weight to the error. The training is done alone with the convergence of the loss function.

Support Vector Machine (SVM)

The concept of Support Vector Machine is putting data into hyperspace to find a hyperplane that can divide the data into two groups. To deal with the problem that not all data are linearly separable, kernel functions can help map the data to a high-dimensional space before classification in order to find a hyperplane with the most distance to the critical value datapoints. We display this mathematically as follows.

Assume there are data $(x_1,\ y_1), (x_2,\ y_2) \ldots\ (x_n,\ y_n)$, where x_i is the i^{th} data, or a d-dimensional vector. $y_i \in \{-1,\ 1\}$ is the category that x_i should be classified into. The goal is to find two parallel hyperplanes that could separate data as follows.

$$w \cdot x_i + \mathrm{b} \geq 1 \quad if\ y_i = 1$$

$$w \cdot x_i + \mathrm{b} \leq -1 \quad if\ y_i = -1$$

We can also write them as:

$$y_i(w \cdot x_i + \mathrm{b}) \geq 1$$

The hyperplane in the middle of the two parallel hyperplanes is:

$$w \cdot x_i + \mathrm{b} = 0$$

Two parallel hyperplanes have a distance of $\frac{2}{||w||}$. Minimizing $||w||$ gets the maximum distance. Thus, the objective function of finding the hyperplane that separates the data farthest apart is:

$$\min ||w||,\ s.t.\ \ y_i(w \cdot x_i + \mathrm{b}) \geq 1 \quad \mathrm{i} = 1,\ 2, .. \mathrm{n}$$

Since not all data are linearly separable, mapping function $\Phi(\mathrm{x})$ is used to transform the data into hyperspace. Slack variables ξ_i can also be added to tackle the problem of perfect separation. A positive real number C is used to control the error size. The objective function becomes:

$$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i$$

$$s.t. \begin{cases} y_i(w \cdot \Phi(x_i) + \mathrm{b}) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases} \quad \forall\ i \in N$$

It can then be turned into a Lagrange multiplier problem for optimal solution:

$$min\ L(w,\ \alpha,\ \beta, \xi,\ b) = \frac{1}{2}||w||^2 + C\sum_{i=1}^{n} \xi_i - \sum_{i=1}^{n} \alpha_i[\ y_i(w \cdot \Phi(x_i) + \mathrm{b}) + \xi_i - 1] - \sum_{i=1}^{n} \beta_i \xi_i$$

$$s.t.\ \ \alpha_i\ \beta_i \geq 0\ \ \forall\ i \in N$$

It has the following Karush-Kuhn-Tucker (KKT) Conditions:

$$\frac{\partial L(w,\ \alpha,\ \beta, \xi,\ b)}{\partial w} = \mathrm{w} - \sum_{i=1}^{n} \alpha_i[\ y_i\ \Phi(x_i)] = 0$$

$$\frac{\partial L(w,\ \alpha,\ \beta, \xi,\ b)}{\partial b} = -\sum_{i=1}^{n} \alpha_i(y_i) = 0$$

$$\frac{\partial L(w,\ \alpha,\ \beta, \xi,\ b)}{\partial \xi} = C - \alpha_i - \beta_i = 0$$

Term After considering all the conditions mentioned above and finding the best solution, the hyperplane can be converted to:

$$w \cdot \Phi(x_j) + \mathrm{b} = \sum_{i=1}^{n} \alpha_i \, y_i \, \Phi(x_i)^T \, \Phi(x_j) + \frac{1}{N_s} \sum_{0<\alpha_j<C} \left[y_j - \sum_{i=1}^{n} \alpha_i \, y_i \, \Phi(x_i)^T \, \Phi(x_i) \right] = 0$$

$\Phi(x_i)^T \, \Phi(x_j)$ is called the Kernel function. We use the Radial Basis function in this study.

$$\Phi(x_i)^T \, \Phi(x_j) = \exp\left(-\gamma \left\| x_i - x_j \right\|^2\right) \quad \gamma > 0$$

Extreme Gradient Boosting (XGBoost)

XGBoost is a machine learning model that assembles multiple Decision Trees. Every newly-added Decision Tree modifies the errors for existing Decision Trees and improves the predicting ability of the model. The application of adding a new Decision Tree to rectify any error is called Gradient Boosting. We shall explain the theory through a mathematic model.

$$\hat{y}_i = \sum\nolimits_{k=1}^{K} f_k \, (x_i)$$

Here, x_i stands for the i^{th} datapoint, f_k stands for the prediction function of the k^{th} Decision Tree, and $\hat{y}_i$ is the value predicted by the model made of K Decision Trees.

The objective function is:

$$Obj = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k)$$

Where:

$$\sum_i l(\hat{y}_i, y_i) \;\; = \frac{wrong\ cases}{all\ cases}$$

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum\nolimits_{j=1}^{T} w_j^2$$

$l(\hat{y}_i, y_i)$ is called the loss function and represents the error rate. $\Omega(f_k)$ is called regularization with the purpose of avoiding overfitting. T stands for the number of leaves. w_j is the j^{th} leaf value for the k^{th} Decision Tree. γ is a positive constant that controls the number of leaves. Positive constant λ is L2 regularization, or also known as weight decay that decreases the impact of noises for the model.

All new predicted values form the sum of the new (t^{th} in here) Decision Tree's value and the other Decision Trees' predicted value.

$$\hat{y}_i^t = \sum_{k=1}^{K} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)$$

We put this into the objective function:

$$Obj^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^{t} \Omega(f_i) = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) + constant$$

According to Taylor's expansion, we get:

$$Obj^{(t)} \simeq \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)}\right) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) + constant$$

$$g_i = \partial_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right), \ h_i = \partial^2_{\hat{y}^{(t-1)}} l\left(y_i, \hat{y}_i^{(t-1)}\right)$$

The objective function without the constant is then:

$$Obj^{(t)} \simeq \sum_{i=1}^{n} \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t)$$

Assume that w_j is the predicted value when we put the i^{th} datapoint into the t^{th} Decision Tree prediction function.

$$I_j = \{i | q(x_i) = j\}$$

$$Obj^{(t)} \simeq \sum_{i=1}^{n} \left[g_i w_{q(x_i)} + \frac{1}{2} h_i w_{q(x_i)}^2 \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^{T} w_j^2 = \sum_{j=1}^{T} \left[G_i w_j + (H_i + \lambda) w_j^2 \right] + \gamma T$$

Here, $G_i = \Sigma_{i \in I_j} g_i$, $H_i = \Sigma_{i \in I_j} h_i$. We can consider it as a quadratic equation and solve it.

$$w_j^* = -\frac{G_j}{H_j + \lambda}$$

$$Obj^{(t)} \simeq -\frac{1}{2} \sum_{j=1}^{T} \frac{G_j^2}{H_j + \lambda} + \gamma T$$

The above procedures obtain the leaf value after the Decision Tree structure is built. The establishment of the Decision Tree is based on a greedy algorithm, where we put the split result of each attribute into the following formula.

$$Max: \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma$$

$\frac{G_L^2}{H_L+\lambda}$ is the sum of the leaf values on the left side after the split. $\frac{G_R^2}{H_R+\lambda}$ is the sum of leaf values on the right side after the split. $\frac{(G_L+G_R)^2}{H_L+H_R+\lambda}$ is the sum of the leaf values before the split.

Methods of Evaluation

The four indicators, Accuracy, Precision, Recall, and F1- score, used in this study are from the Confusion Matrix, which is commonly used for machine learning model evaluation. Table 1 displays the elements in the Confusion Matrix.

Table 1: Confusion Matrix

Confusion Matrix		True	
		Positive	Negative
Predict	Positive	TP (True Positive)	FP (False Positive)
	Negative	FN (False Negative)	TN (True Negative)

The term "Predict" in Table 1 delegates the classification outcome. True stands for the actual result. TP (True Positive) is the prediction result aligning with an actual positive outcome, while TN (True Negative) is the prediction result aligning with an actual negative outcome. FP (False Positive) is the positive prediction result that runs in contrast to the actual negative result, or also known as Type II Error. FN (False Negative) is the negative prediction result that runs in contrast to the actual positive result, or also known as Type I Error. The four elements in the Confusion Matrix can be used to calculate the four indicators.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 - score = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}}$$

Portfolio Construction Methods

The portfolio construction method used in this study can be divided into two types: alpha strategy and long strategy. The former buys high ESG-rated strong buy stocks and short sells the same dollar value of strong sell stocks. The latter only buys strong buy stocks without selling any stocks. The two types of portfolio construction method can be combined with an equal-weighted strategy or value-weighted strategy to determine the weight of each stock in the portfolio. This eventually lead to four kinds of portfolio construction methods: (1) Long / Equal-weighted. (2) Long / Value-weighted. (3) Alpha / Equal-weighted. (4) Alpha / Value-weighted.

We also adopt the top 10%, bottom 10%, 20%, and 30% ESG ratings among all the companies as the filter criteria for strong buy stocks and strong sell stocks in this study. These criteria, with four portfolio construction methods create 12 combinations. The portfolios can be classified into more detailed categories that depend on the way the list of strong buy and strong sell companies is obtained, through real ESG ratings or predicted by the five machine learning models. However, parts of the portfolio construction methods lack a clear research motive and research significance. Consequently, this study only lists the performances of portfolios with research significance and supported by sufficient evidence that make it possible to achieve a steady absolute return, based on the empirical result. The research significance and motives of the listed portfolios will be further discussed in the empirical result section.

DATA DESCRIPTION

The data employed in this study spans from January 2003 to January 2019, encompassing financial reports, stock prices, and ESG scores of U.S. listed companies. Financial data is sourced from Compustat monthly data within Wharton Research Data Services (WRDS), stock price data originates from the Center for Research in Security Prices (CRSP), while ESG score data is drawn from annual data in Eikon with Datastream - Thomson Reuters ESG Scores. The financial dataset includes 52 financial ratios, and the average annual count of individual stocks in the dataset stands at 361.68.

The input variables of the machine learning model encompass the financial report information of individual stocks from January to November of each year. These variables consist of the values of 52 financial ratios across the 11 months, yielding a total of 572 features. The output variable signifies the categorical characteristics of ESG scores for the year, classifying into three groups: (1) Low ESG score group: bottom n% of the overall companies; (2) Mid ESG score group: middle 100 – 2n% of the companies; (3) High ESG score group: top n% of the companies. In this study, n is tested with values of 10, 20, and 30.

The machine learning model is divided into a training set and a test set. The training set includes a 30% validation subset and employs a rolling training approach, using data from years t-6 to t-1 for training and data from year t for testing. The ratio of training to testing data years is

6:1, slightly higher than the common 4:1 or 5:1 division. This is attributed to the growing prominence of ESG issues in recent years, leading to an increasing volume of ESG score data. Thus, a 6:1 annual data division is chosen, resulting in t-6 to t-1 years' actual data being approximately four to five times that of year t's data. The computed annualized returns in this study do not account for trading costs, fees, and additional costs associated with shorting.

EMPIRICAL RESULTS

Data and Preprocessing

The data employed in this study spans from January 2003 to January 2019, encompassing financial reports, stock prices, and ESG scores of U.S. listed companies. Financial data is sourced from Compustat monthly data within Wharton Research Data Services (WRDS), stock price data originates from the Center for Research in Security Prices (CRSP), while ESG score data is drawn from annual data in Eikon with Datastream - Thomson Reuters ESG Scores. The financial dataset includes 52 financial ratios, and the average annual count of individual stocks in the dataset stands at 361.68.

The input variables of the machine learning model encompass the financial report information of individual stocks from January to November of each year. These variables consist of the values of 52 financial ratios across the 11 months, yielding a total of 572 features. The output variable signifies the categorical characteristics of ESG scores for the year, classifying into three groups: (1) Low ESG score group: bottom n% of the overall companies; (2) Mid ESG score group: middle 100 – 2n% of the companies; (3) High ESG score group: top n% of the companies. In this study, n is tested with values of 10, 20, and 30.

The machine learning model is divided into a training set and a test set. The training set includes a 30% validation subset and employs a rolling training approach, using data from years t-6 to t-1 for training and data from year t for testing. The ratio of training to testing data years is 6:1, slightly higher than the common 4:1 or 5:1 division. This is attributed to the growing prominence of ESG issues in recent years, leading to an increasing volume of ESG score data. Thus, a 6:1 annual data division is chosen, resulting in t-6 to t-1 years' actual data being approximately four to five times that of year t's data. The computed annualized returns in this study do not account for trading costs, fees, and additional costs associated with shorting.

Examination Procedure and Research Significance

In the preliminary phase of our empirical analysis, we designed portfolios based on a 'strong buy' (or 'sell') distinction, identified by the top (or bottom) n% of genuine ESG ratings for the respective year, where n is set at 10, 20, or 30%. These portfolios tactically enter the market both at the beginning and the end of December. Our evaluation focuses on monthly returns, which includes the return for December and the following January. It is essential to recognize that the formal ESG ratings are disclosed only at the close of December. Thus, portfolios formulated at the start of December do not have access to these ratings in real-time. Their formation hinges on the assumption of accurately anticipating ESG ratings a month in advance.

Performance Explanation (Real ESG Ratings)

Table 2 presents the performance of portfolios crafted using a long strategy combined with an equal-weighted approach. Conversely, Table 3 portrays the outcomes from portfolios that merge a long strategy with a value-weighted technique. All "strong sell" stock classifications are informed by actual ESG ratings. The evaluation period spans each December from 2009 through 2018 and includes the ensuing January (2010-2019). Key performance indicators under scrutiny are the annual returns over ten 1-months, annual standard deviation, and the Sharpe ratio.

A cursory analysis suggests that portfolio performances in January markedly surpass those in December across all metrics. Initial impressions might intimate that the prerogative to forecast ESG ratings for early market entry lacks merit. Strategizing stock classifications based on actual ESG ratings disclosed year-end, and subsequently constructing portfolios for market entry, appears to be the more judicious approach.

However, an intriguing anomaly arises upon a minor alteration in our performance evaluation timeline. Tables 4 and 5, paralleling Tables 2 and 3 in their portfolio construction methodology, truncate the evaluation timeframe, excluding portfolio outcomes from December 2018 and January 2019. The revelations from this altered analysis paint a starkly different picture: December's performances significantly eclipse those of January across annual return, annual standard deviation, and Sharpe ratio metrics. This stark reversal, prompted solely by omitting a single month's data, underscores the peculiarity of the period straddling the close of 2018 and the dawn of 2019. This anomaly necessitates a deeper probe into the financial market dynamics of that interval, as its understanding holds profound implications for the overarching conclusions of this study.

Table 2 : Real ESG/Long/Equal weighted (2009-2018)

	10%			20%			30%		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
December	0.55%	15.49%	0.036	-0.50%	16.05%	-0.031	-2.46%	16.09%	-0.153
January	11.68%	15.06%	0.775	18.77%	16.78%	1.119	20.62%	18.15%	1.136

Table 3 : Real ESG/Long/Value weighted (2009-2018)

	10%			20%			30%		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
December	-3.76%	13.69%	-0.275	-1.95%	13.70%	-0.142	-3.10%	13.57%	-0.228
January	9.74%	17.79%	0.547	8.36%	14.88%	0.562	10.23%	14.81%	0.691

Table 4 : Real ESG/Long/Equal weighted (2009-2017)

	10%			20%			30%		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
December	15.75%	10.01%	1.573	16.75%	7.86%	2.13	14.37%	8.88%	1.619
January	4.28%	13.28%	0.322	7.79%	12.66%	0.615	9.17%	13.96%	0.657

Table 5 : Real ESG/Long/Value weighted (2009-2017)

	10%			20%			30%		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
December	10.35%	7.58%	1.365	12.57%	6.76%	1.859	11.21%	6.86%	1.633
January	2.23%	16.20%	0.138	0.31%	12.61%	0.025	2.16%	12.53%	0.172

The financial terrain in late 2018 was marked by heightened tensions in US-China trade relations, Federal Reserve's interest rate elevations, and the emergence of an inverted yield curve, culminating in notable market fluctuations. In this context, December 2018 saw the S&P 500 diminish by 12.26% and the Nasdaq contract by 13.69%. However, January 2019 marked a phase of market resurgence, largely credited to easing trade frictions and signals from the Federal Reserve hinting at a possible halt in rate increases. During this period, the S&P 500 and the Nasdaq appreciated by 7.61% and 9.41% respectively.

By integrating the portfolio data of December 2018, we are effectively capturing a scenario of entering the market amidst a downturn and exiting just before a subsequent upswing. This dynamic underpins the disparate outcomes noted between Tables 2, 3 and Tables 4, 5. The consistently enhanced outcomes of value-weighted portfolios can be, in part, attributed to the seamless manner in which established corporations engage in and disclose ESG undertakings. Conversely, smaller firms grapple with financial challenges in ESG endeavors, potentially influencing their positioning in equal-weighted portfolios. Notably, portfolios aimed at delivering steady absolute returns encountered impediments in epochs of pronounced systemic risk, as was the case towards 2018's end. Excluding data predicated purely on unfavorable outcomes during portfolio analysis can inadvertently mask the authentic investment landscape.

In our endeavor to devise a portfolio resilient enough to accrue positive gains amid heightened systemic risks, we adopted an alpha-focused construction approach, with the results delineated in Tables 6 and 7. Following this strategic alignment, December's results, even accounting for 2018's turbulent figures, consistently eclipsed January's absolute returns. A deeper inspection emphasizes the dominance of value-weighted portfolios over their equal-weighted counterparts. It becomes evident that sizeable corporations boasting elevated ESG scores have a discernible edge. This could stem from their probable early access to ESG intel and the inherent value they derive from ESG commitments. Conversely, for smaller companies, accentuating ESG metrics might inflate operational expenditures, thereby influencing their bottom line. In our pursuit to ascertain the consistency of our conclusions, especially to offset any potential distortion from a singular year's data, Table 8 chronicles the trajectory of value-weighted portfolios, deliberately excluding 2018's figures. The coherence between its outcomes and those in Table 7 underscores the robustness of portfolios which synergize both the alpha-centric approach and the value-weighted paradigm.

Table 6 : Real ESG/Alpha/Equal weighted (2009-2018)

	10%			**20%**			**30%**		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
December	3.34%	3.71%	0.902	3.03%	3.36%	0.9	1.31%	2.49%	0.526
January	-5.61%	8.49%	-0.661	-2.04%	5.07%	-0.403	0.35%	4.10%	0.084

Table 7 : Real ESG/Alpha/Value weighted (2009-2018)

	10%			**20%**			**30%**		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
December	4.17%	4.04%	1.033	7.55%	6.46%	1.17	5.75%	5.41%	1.062
January	-6.82%	8.12%	-0.839	-11.63%	6.04%	-1.927	-8.45%	5.20%	-1.624

Table 8 : Real ESG/Alpha/Value weighted (2009-2017)

	10%			**20%**			**30%**		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
December	3.27%	4.16%	0.785	6.73%	6.75%	0.996	5.39%	5.71%	0.944
January	-3.28%	8.20%	-0.399	-8.04%	5.61%	-1.432	-5.36%	4.81%	-1.115

Our analysis highlights the resource advantage that larger corporations possess, enabling them to effectively engage in, and broadcast their ESG commitments. Such advantages are clearly visible in value-weighted portfolios. In contrast, smaller firms often face financial hurdles when initiating ESG efforts, a factor that might influence their presence in equal-weighted portfolios. Notably, we found that some institutional entities may obtain early insights into vital ESG data due to their close ties with ESG information agencies or the evaluated companies. This early information can provide them with a competitive edge. A focused review of portfolio returns, comparing December's performance with January's, offers three key insights: firstly, there's a noticeable positive alpha linked to the ESG factor exactly one month before its public release; secondly, the heightened returns of December diminish following the formal ESG ratings disclosure; and thirdly, this points to the critical importance of precise ESG predictions before engaging in the December market.

Preliminary Conclusion

Our empirical analysis yields key insights into the dynamics of ESG factors and portfolio performance. One clear observation is the noticeable positive returns in December, suggesting that the ESG factor inherently contributes to excess returns. However, this advantage lessens once the ESG ratings enter the public domain, as evidenced by the decline into negative returns in January post-disclosure.

The alpha strategy stands out for its efficacy in mitigating systemic risk and consistently delivering stable returns. Within the realm of portfolio weighting, the value-weighted approach, which emphasizes larger firms, shows superior performance over the equal-weighted method. This isn't merely a result of the natural market benefits larger corporations enjoy. It’s also influenced by their efficiency in adopting and communicating ESG initiatives. In contrast, smaller entities face financial challenges when embarking on ESG initiatives, impacting their performance.

Central to these findings is the crucial ability to accurately forecast ESG ratings at the beginning of December, well before their official release at month-end. The precision of this forecast is of paramount importance. Enhanced accuracy ensures that portfolio strategies based on predicted ESG ratings align closely with actual rating outcomes. This fact highlights the importance of honing the accuracy of ESG rating predictions early in December.

In sum, this analysis provides a clearer understanding of the interplay between ESG factors, portfolio weighting approaches, and performance outcomes. These insights hold significant implications for future academic investigations and real-world investment strategies.

Machine Learning Portfolio

Machine Learning Examination and Research Significance

Our subsequent analysis builds upon the foundational conclusions we've established: notably, that the alpha strategy empowers portfolios grounded in authentic ESG ratings to realize absolute returns, even in the face of intensified systemic risks. However, this is predicated upon a pivotal precondition: the acumen to anticipate genuine ESG ratings.

In this investigative pursuit, our methodology taps into the prowess of five divergent machine learning models to simulate the genuine ESG ratings of company listings. While harnessing the capabilities of these predictive instruments, it's salient to acknowledge that larger corporations might innately command superior access to ESG intelligence, affording them an advantageous stance even in the preamble to official disclosures. Our objective is twofold: to precisely forecast both the peak and trough of ESG ratings, and in the aftermath, to classify equities into unambiguous 'strong buy' and 'strong sell' brackets. Prior to the infusion of these machine learning models into our portfolio architecture, a rigorous evaluation of their predictive precision is indispensable. Models demonstrating superlative accuracy will be earmarked for subsequent portfolio synthesis.

By crafting portfolios that resonate with precise ESG ratings and instituting a market entry predicated on the alpha strategy at the dawn of December, we position ourselves to incisively scrutinize performance indicators. This endeavor illuminates whether the integration of machine learning's predictive prowess can truly amplify the potential for excess return harvest.

Machine Learning Model Accuracy

In the scope of this study, we employ financial report data covering a span of eleven months, from January through November, as a foundation to forecast the ESG rating as of the close of December. The efficacy of these predictions is detailed across Tables 9, 10, and 11. Among the quintet of machine learning models enlisted, the Random Forest and XGBoost models—both being advanced tree-based algorithms—stand out with commendable predictive prowess. As a corollary, Table 12 delves into a side-by-side scrutiny of these two models.

An intriguing trend crystallizes: a surge in Accuracy and Precision tends to be paralleled by a contraction in Recall and F1-score, especially when the task at hand is pinpointing the uppermost and lowermost deciles of ESG ratings. Rooted in the underpinnings of these metrics, it appears the models incline towards a circumspect approach while discerning these extremities, predominantly classifying entities within the central 80% ESG rating band. A meager cohort of firms find themselves earmarked for the apex or nadir 10% ESG rating categories. Moreover, it's critical to appreciate that the data propelling these forecasts might be more accessible to behemoth corporations well in advance, which could subtly skew the fidelity and relevance of our machine learning models across a spectrum of firm dimensions. Consequently, while both Accuracy and

Precision are buoyed by the ballooning contingent of the intermediate rating cadre, a proliferation in False Negatives (FN) —emanating from entities rightfully belonging to the extreme tiers but misclassified to the median—acts as a drag on the Recall and F1-score metrics.

Table 9: Top/Bottomt 10% predicting matrix

	DTs	RF	ANN	SVM	XGBoost
Accuracy	0.7656	0.8194	0.7818	0.7976	0.819
Precision	0.4959	0.7965	0.3236	0.2659	0.7954
Recall	0.4336	0.4304	0.3414	0.3333	0.4176
F1-Score	0.4494	0.4647	0.3179	0.2958	0.442

Table 10: Top/Bottom 20% predicting matrix

	DTs	RF	ANN	SVM	XGBoost
Accuracy	0.5767	0.6833	0.5717	0.598	0.6853
Precision	0.499	0.7293	0.2948	0.1993	0.7247
Recall	0.4655	0.5163	0.3458	0.3333	0.5197
F1-Score	0.4739	0.5481	0.2952	0.2495	0.5514

Table 11: Top/Bottom 30% predicting matrix

	DTs	RF	ANN	SVM	XGBoost
Accuracy	0.4847	0.6265	0.4093	0.3988	0.6176
Precision	0.4938	0.6556	0.3147	0.2268	0.6389
Recall	0.48	0.6226	0.3719	0.3337	0.6157
F1-Score	0.4819	0.6303	0.308	0.1955	0.6214

Table 12: Random Forest, XGBoost predicting matrix

		10%	20%	30%
Accuracy	RF	0.8194	0.6833	0.6265
	XGBoost	0.819	0.6853	0.6176

Precision	RF	0.7965	0.7293	0.6556
	XGBoost	0.7954	0.7247	0.6389
Recall	RF	0.4304	0.5163	0.6226
	XGBoost	0.4176	0.5197	0.6157
F1-Score	RF	0.4647	0.5481	0.6303
	XGBoost	0.442	0.5514	0.6214

Performance Explanation (Machine Learning)

In our study, two machine learning models, namely Random Forest and XGBoost, are chosen for their pronounced proficiency in forecasting ESG ratings. Leveraging the defined classifications of 'strong buy' and 'strong sell' for companies, portfolios are structured in accordance with the alpha strategy. Performance metrics and the cumulative return for the equal-weighted portfolios are tabulated in Table 13 and graphically portrayed in Figure 4. In parallel, results pertaining to value-weighted portfolios find their place in Table 14 and are elucidated visually in Figure 5.

A salient observation is the sustained superior performance of value-weighted portfolios compared to their equal-weighted peers, mirroring the patterns discerned in portfolios grounded on actual ESG ratings. This differential can be potentially attributed to the fact that sizable corporations, which are more pronounced in value-weighted portfolios, might enjoy a two-fold advantage. Firstly, they might gain a vantage point from preliminary access to ESG ratings and secondly, reap the rewards inherent to active engagement in ESG undertakings. Another dimension to consider is the innate capability of these large entities, characteristic of value-weighted portfolios, to seamlessly integrate and transparently communicate their ESG initiatives. In contrast, their smaller brethren populating the equal-weighted portfolios might confront the financial challenges tethered to ESG activities. On the continuum of model performance, XGBoost typically edges out Random Forest. However, an anomaly emerges in XGBoost's forecast for the topmost/bottommost decile of ESG ratings within equal-weighted portfolios, manifesting a negative yield. A preponderance of the portfolios, sculpted through the insights of these machine learning models, flaunt positive returns. This underscores the inference that portfolios sculpted around 'strong buy' and 'strong sell' delineations—foretold by Random Forest and XGBoost—offer a trustworthy conduit to realizing absolute returns.

Table 13: Machine Learning/Alpha/Equal weighted (2009-2018)

	10%			20%			30%		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
RF	5.28%	9.72%	0.543	2.65%	4.75%	0.558	2.81%	3.6%	0.781
XGBoost	-2.42%	8.45%	-0.287	4.13%	3.81%	1.085	3.7%	3.73%	0.991
Real ESG	3.34%	3.71%	0.902	3.03%	3.36%	0.9	1.31%	2.49%	0.526

Table 14: Machine Learning/Alpha/Value weighted (2009-2018)

	10%			20%			30%		
	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio	Return	Standard Deviation	Sharpe ratio
RF	8.95%	12.01%	0.745	2.28%	7.72%	0.295	5.28%	4.73%	1.115
XGBoost	10.36%	9.91%	1.045	5.04%	5.79%	0.871	5.45%	4.17%	1.307
Real ESG	4.17%	4.04%	1.033	7.55%	6.46%	1.17	5.75%	5.41%	1.062

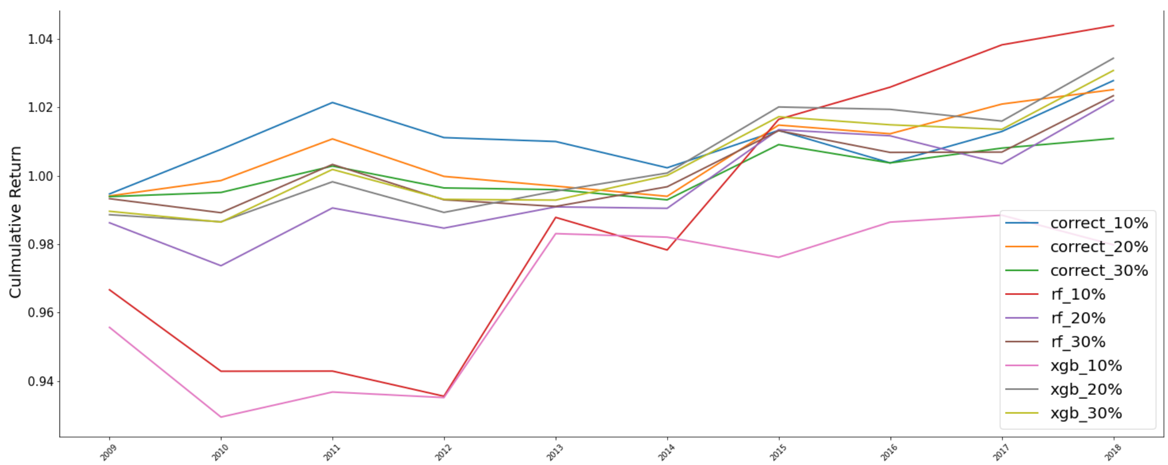

Figure 4: Machine learning/Alpha/Equal weighted accumulated return (2009-2018)

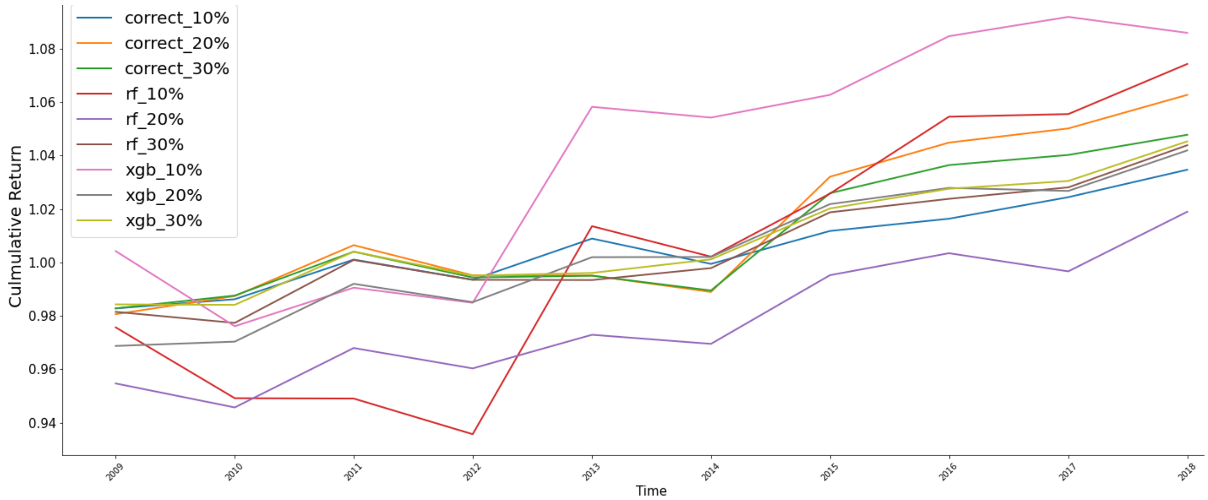

Figure 5: Machine learning/Alpha/Value weighted accumulated return (2009-2018)

CONCLUSION

Our empirical findings detail the positive returns in December for real ESG portfolios, indicating a tangible excess return inherent to the ESG factor. Analyzing the performance differential between real ESG rating portfolios in December and those in January of the subsequent year reveals a tendency for the excess return to diminish following the public disclosure of real ESG ratings at December's end.

A significant contribution of our research lies in the intricate exploration of portfolio construction methodologies. Specifically, we demonstrate that the implementation of an alpha strategy—when compared with market systemic risks—can offer stability against marked volatilities. More critically, combining the alpha strategy with a value-weighted approach proves beneficial, reducing the financial challenges of obtaining ESG ratings for smaller firms. This suggests that larger corporations, with their expansive resources and infrastructural advantages, are better positioned to meet stringent ESG standards.

In our attempt to preemptively access real ESG ratings and thereby optimize portfolios, we utilized five machine learning models, using financial metrics from January to November. Our contribution in this domain emphasizes the effectiveness of models such as Random Forest and XGBoost. Their superior predictive accuracy, when used to shape portfolios, indicated that a fusion of the alpha strategy with value-weighting consistently outperformed the integration of the alpha strategy with equal weighting, ensuring a consistent stream of positive returns. In addition, the observed elevation in December returns for ESG portfolios, relative to their January counterparts, may suggest that a subset of institutional investors have access to preliminary or exclusive data related to ESG ratings before widespread release.

Future enhancements to this research are multifaceted. Exploring further into the hyperparameter tuning of the machine learning models or incorporating diverse data sources can potentially enhance their predictive capabilities. Improving precision metrics such as Accuracy, Precision, Recall, and the F1-score could ensure predictions align more closely with actual ESG ratings. Introducing traditional analytical techniques, like linear regression, can offer a contrasting analytical viewpoint alongside machine learning approaches. While our study assessed the pros and cons between equal and value-weighted portfolio mechanisms, subsequent studies could refine portfolio weights based on specific criteria. Nevertheless, we acknowledged the omission of real-world complexities like transaction costs, additional short-selling expenses, and the timing of financial report releases. Addressing these elements in future research will be crucial in determining the portfolio's performance in practical scenarios.

REFERENCES

1. Adebiyi, A. A., Ayo, C. K., Adebiyi, M., & Otokiti, S. O. (2012). Stock price prediction using neural network with hybridized market indicators. *Journal of Emerging Trends in Computing and Information Sciences*, *3*(1).
2. Brush, J. S. (1997). Comparisons and combinations of long and long/short strategies. *Financial Analysts Journal*, *53*(3), 81-89.
3. Deng, X., & Cheng, X. (2019). Can ESG Indices Improve the Enterprises' Stock Market Performance?—An Empirical Study from China. *Sustainability*, *11*(17), 4765.
4. Edwards, F. R., & Caglayan, M. O. (2001). Hedge fund and commodity fund investments in bull and bear markets. *The Journal of Portfolio Management*, *27*(4), 97-108.
5. Endri, E., Kasmir, K., & Syarif, A. (2020). Delisting sharia stock prediction model based on financial information: Support Vector Machine. *Decision Science Letters*, *9*(2), 207-214.
6. Engelhardt, N., Ekkenga, J., & Posch, P. (2021). ESG ratings and stock performance during the COVID-19 crisis. *Sustainability*, *13*(13), 7133.
7. Grinold, R. C., & Kahn, R. N. (2000). The Efficiency Gains of Long–Short Investing. *Financial Analysts Journal*, *56*(6), 40-53.

8. Jacobs, B. I., & Levy, K. N. (1993). Long/short equity investing. *Journal of Portfolio Management*, *20*(1), 52.
9. Jensen, M. C. (1967). The performance of mutual funds in the period 1945-1964. *The Journal of Finance*, *23*(2), 389-416.
10. Khan, M. (2019). Corporate governance, ESG, and stock returns around the world. *Financial Analysts Journal*, *75*(4), 103-123.
11. Lintner, J. (1965a). The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, 47, 13 – 37.
12. Lintner, J. (1965b). Securities Prices, Risk, and Maximal Gains from Diversification. *Journal of Finance*, 20(4), 587 – 615.
13. Mossin, J. (1966) Equilibrium in a Capital Asset Market. The Econometric Society, 34, 768-783.
14. Roko, I., & Gilli, M. (2008). Using economic and financial information for stock selection. *Computational Management Science*, *5*(4), 317-335.
15. Sharpe, W. F. (1964) Capital Asset Prices A Theory of Market Equilibrium under Conditions of Risk. Journal of Finance, 19, 425-442.
16. Tan, Z., Yan, Z., & Zhu, G. (2019). Stock selection with random forest: An exploitation of excess return in the Chinese stock market. *Heliyon*, *5*(8), e02310.
17. Treynor, J. L. (1962). Toward a Theory of Market Value of Risky Assets. Unpublished manuscript. "Rough Draft" dated by Treynor to the fall of 1962. A final version was published in 1999, in *Asset Pricing and Portfolio Performance.*R. A. Korajczyk (editor), London: Risk Books, 15 – 22.
18. UNCTAD (2021). ESG ETFs 2021 - State of the market and the potential for sustainable development. *Working Paper*.
19. Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia Computer Science*, *167*, 599-606.
20. Walczak, S. (2001). An empirical analysis of data requirements for financial forecasting with neural networks. *Journal of Management Information Systems*, *17*(4), 203-222.