



# Catalyzing Sustainable Investment: Revealing ESG Power in Predicting Fund Performance with Machine Learning

Alexandre Momparler<sup>1</sup> · Pedro Carmona<sup>2</sup> · Francisco Climent<sup>3</sup>

Accepted: 24 April 2024 / Published online: 4 May 2024  
© The Author(s) 2024

## Abstract

In today's dynamic financial landscape, the integration of environmental, social, and governance (ESG) principles into investment strategies has gained great significance. Investors and financial advisors are increasingly confronted with the crucial question of whether their dedication to ESG values enhances or hampers their pursuit of financial performance. Addressing this crucial issue, our research delves into the impact of ESG ratings on financial performance, exploring a cutting-edge machine learning approach powered by the Extreme Gradient algorithm. Our study centers on US-registered equity funds with a global investment scope, and performs a cross-sectional data analysis for annualized fund returns for a five-year period (2017–2021). To fortify our analysis, we synergistically amalgamate data from three prominent mutual fund databases, thereby bolstering data completeness, accuracy, and consistency. Through thorough examination, our findings substantiate the positive correlation between ESG ratings and fund performance. In fact, our investigation identifies ESG score as one of the dominant variables, ranking among the top five with the highest predictive capacity for mutual fund performance. As sustainable investing continues to ascend as a central force within financial markets, our study underscores the pivotal role that ESG factors play in shaping investment outcomes. Our research provides socially responsible investors and financial advisors with valuable insights, empowering them to make informed decisions that align their financial objectives with their commitment to ESG values.

**Keywords** Mutual fund · ESG · Sustainability · Investment fund · Machine learning · Classification · Fund selection

## 1 Introduction

The global investment environment is experiencing an unprecedented surge in the allocation of savings into a diverse array of mutual funds, accompanied by a steady rise in the number of investors participating in these funds. This phenomenon carries profound implications for the world economy on two distinct fronts. Firstly, mutual funds play an indispensable role as a vital source of financing for corporations across various sectors, facilitating capital flow to fuel business operations, expansion, and innovation. Secondly, these funds serve as the primary conduit through which millions of investors worldwide channel their savings towards achieving long-term financial objectives, be it retirement planning, wealth accumulation, or the pursuit of specific financial goals.

In this context, socially responsible investors are confronted with a unique confluence of factors that shape their investment decisions. The quest for both financial prosperity and the advancement of personal values calls for a discerning approach to mutual fund selection. Beyond the conventional considerations of investment strategies and historical performance, profit-seeking socially responsible investors recognize the essential role that Environmental, Social, and Governance (ESG) ratings play in the contemporary investment landscape. The integration of ESG criteria into their investment assessment represents a conscientious endeavor to harmonize profit motives with ethical principles, aligning their investments with a broader commitment to sustainability and responsible corporate behavior.

Acknowledging the growing significance of ESG factors, mutual fund databases have taken substantial strides in catering to the evolving needs of investors. Leading financial database providers such as Refinitiv Lipper and Morningstar have incorporated ESG ratings into their comprehensive datasets. This proactive measure empowers both financial advisors and individual investors to infuse the ESG dimension into their fund selection process. The availability of accurate, standardized, and comparable environmental, social, and governance data equips investors with a powerful tool to evaluate the social and ethical impact of their investments, transcending mere financial considerations. This marks a transformative shift, ushering in an era where investments are not only expected to yield financial returns but also to contribute positively to a sustainable and equitable future.

In sum, the burgeoning prominence of mutual funds in the global investment landscape necessitates a holistic approach to investment decision-making. For socially responsible investors, the integration of ESG ratings into their investment strategy is emblematic of a commitment to driving positive change while pursuing financial objectives. The accessibility of ESG data within mutual fund databases empowers investors to make informed choices, thereby reinforcing the symbiotic relationship between finance and social responsibility. This confluence of forces underscores the transformative potential of responsible investing and its capacity to shape a more sustainable and ethical future for global financial markets. Morningstar started rating mutual funds according to ESG criteria in 2016 as a response to investors' growing concern for sustainability values. Morningstar Sustainability Ratings are determined using bottom-up assessments of the underlying holdings within a portfolio, supported by Sustainalytics. The Morningstar sustainability rating is in the form of five differ-

ent categories (1 to 5 globes): low equals 1 globe and high equals 5 globes. Refinitiv Lipper fund ESG numerical scores provide investors and financial advisors with overall fund-level scores that describe the sustainability of a particular mutual fund. All Refinitiv Lipper fund ESG scores are rolled up from the security level to give a fund-level score.

According to the Global Sustainable Fund Flows Report (Morningstar, 2021, 2022) in the global market of sustainable funds, comprising both open-end and exchange-traded funds with a clear focus on sustainability, Europe is leading the sustainable funds market with 83% of global sustainable fund assets, followed by the US with 11% of global sustainable fund assets as of December 2022 (see Table 1). Regarding the number of sustainable funds, Europe also leads the sustainable fund market segment with 5,316 funds (76%), followed by the US with 598 funds (9%) and Asia (excluding Japan) with 459 funds (6%). Although the European fund market plays a prominent role in the sustainable fund segment, US households hold a comparatively larger portion of their assets (about 23%) in regulated mutual funds, whereas European households just hold about 9% of their assets in regulated mutual funds (International Investment Funds Association, 2021). The outstanding portion of US household financial wealth invested in funds and the fact that the US has the largest share of total net assets of worldwide regulated open-end mutual funds are two compelling reasons to focus our research on the US mutual funds market.

Investors' motivations for opting for mutual funds with robust ESG scores are inherently diverse, encompassing a spectrum of orientations ranging from socially driven to profit-seeking objectives. Recognizing this divergence in investor intent is fundamental, as it underscores the significance of aligning individual motivations with mutual fund choices. Socially responsible investors, characterized by a profound commitment to ESG criteria, often exhibit a willingness to forego a portion of their financial returns in pursuit of the noble cause of promoting sustainable corporate practices. While this apparent trade-off between profit and sustainability might be perceived as a potential drawback, a closer examination might reveal a more nuanced dynamic at play.

Companies that operate under a mandate of social responsibility tend to be characterized by a heightened level of transparency in their operations. This heightened transparency has cascading effects, potentially enhancing their financial performance. Transparency serves as a conduit through which companies can communicate their

**Table 1** Global sustainable funds (2021–2022)

| Region                | Assets (USD Billion) |        |              |         | Number of Funds |        |              |         |
|-----------------------|----------------------|--------|--------------|---------|-----------------|--------|--------------|---------|
|                       | 2021                 | %Total | 2022         | % Total | 2022            | %Total | 2022         | % Total |
| Europe                | 2,231                | 81%    | 2,078        | 83%     | 4,461           | 75%    | 5,316        | 76%     |
| United States         | 357                  | 13%    | 286          | 11%     | 534             | 9%     | 598          | 9%      |
| Asia ex-Japan         | 63                   | 2%     | 51           | 2%      | 342             | 6%     | 459          | 7%      |
| Australia/New Zealand | 31                   | 1%     | 29           | 1%      | 152             | 3%     | 198          | 3%      |
| Japan                 | 35                   | 1%     | 25           | 1%      | 212             | 4%     | 238          | 3%      |
| Canada                | 27                   | 1%     | 28           | 1%      | 231             | 4%     | 203          | 3%      |
| <b>Total</b>          | <b>2,744</b>         |        | <b>2,497</b> |         | <b>5,932</b>    |        | <b>7,012</b> |         |

Source: Morningstar Direct, Manager Research. Data as of December 2021 & 2022

commitment to ethical conduct and sustainability practices to stakeholders. In doing so, these companies foster trust and confidence among investors, customers, and the broader market, consequently bolstering their overall performance. This suggests that, contrary to the notion of profit sacrifice, investments in companies with strong ESG values may yield not only ethical satisfaction but also competitive financial returns.

Nevertheless, it is imperative to acknowledge the intricate and multifaceted nature of the relationship between social metrics and financial performance. This relationship is contingent upon an array of factors, including prevailing market conditions, the specific investment strategies employed, and the precise ESG criteria employed to evaluate mutual funds. The interplay of these variables results in a complex, context-dependent landscape where the impact of ESG metrics on fund performance exhibits variability. Consequently, it becomes clear that there is a pressing need for comprehensive studies, such as the one presented in this paper, aimed at unraveling the intricate dynamics between ESG metrics and fund performance.

This study seeks to contribute to the ongoing debate by shedding light on the interrelationships between ESG considerations and mutual fund performance. We aim to provide investors, financial advisors, and market participants with valuable insights into the implications of ESG integration within the investment landscape. In so doing, we intend to foster a deeper understanding of how ethical and financial objectives can coexist harmoniously, ultimately driving sustainable and responsible investment practices in the evolving financial markets.

In our study, we focus on the US mutual fund market because it held about 52% of the world's regulated mutual fund assets in 2021. Within the US mutual fund market, we analyze the performance of US-marketed equity funds with a global geographic scope according to the Refinitiv Eikon database classification. Our analytical framework is a five-year period (2016–2021) and we combine three different databases to obtain comprehensive qualitative and quantitative data, while improving the accuracy and consistency of the data.

By using the Extreme Gradient Boosting algorithm, a robust machine learning technique, we build a reliable prediction model capable of identifying the most relevant fund features that anticipate fund performance. In this model, Environmental, Social and Governance (ESG) rating is established as a relevant determinant of fund performance. In addition, the positive association found between ESG rating and mutual fund performance definitely challenges the common preconception of a prevailing trade-off between sustainability and financial performance, and it has important implications for both investors and financial advisors.

The remainder of this paper is organized as follows. In the “Literature Review” section, the existing literature on mutual fund selection and sustainability is reviewed and the research goal is stated. The “Data Description” section then describes the sources of data and the characteristics of the sample. In the “Methodology” section, we discuss the data characteristics, variables, and model specification. The empirical results are then explained in the “Results” section. Finally, the main findings, implications, and limitations of our study are put forward in the “Conclusions”.

## 2 Literature Review

Researchers and academics have explored the relationship between ESG and financial performance for decades and most studies found positive correlations between ESG scores and operational efficiencies, stock performance, and lower cost of capital. High ESG ratings are now considered leading indicators of the kind of corporate transparency and high-quality management that pave the way for higher returns. Recent studies suggest that ESG investing can, under certain conditions, help improve risk management and lead to returns that are not inferior to returns from traditional investment strategies.

A number of studies have found that ESG investing may provide effective protection during economic downturns. In this sense, a recent study on the role of ESG performance during the worldwide financial crisis caused by the COVID-19 global pandemic found that companies with higher ESG scores outperformed those with lower ESG scores (Broadstock et al., 2021).

The use of machine learning methods in the finance industry has gained significant momentum in recent years. Li and Rossi (2021) highlight that fund performance exhibits a non-linear relationship with fund characteristics, emphasizing substantial interactions among various fund attributes. Their predictive modeling underscores the significance of these non-linearities and interactions. Employing advanced machine learning techniques, particularly Boosted Regression Trees, proves highly effective, surpassing the performance of standard linear frameworks. Notably, the forecasts generated by Boosted Regression Trees encompass and outperform existing predictors of mutual fund performance proposed in the current literature. Kanade et al. (2022) emphasized the use of machine learning algorithms for predicting the most suitable mutual fund scheme for investors. Their research specifically concentrated on chosen Indian companies and their respective mutual fund schemes. The experiments conducted revealed that the Artificial Neural Network demonstrated the highest accuracy in prediction.

Kaniel et al. (2023) employ neural networks for predicting mutual fund alpha, incorporating a broad range of predictors such as stock characteristics, fund characteristics, and macroeconomic variables. Their study reveals that the inclusion of fund characteristics as predictors renders stock characteristics ineffective in predicting alpha. The study conducted by DeMiguel et al. (2023) establishes that employing machine learning techniques enables the creation of portfolios for mutual funds. These portfolios exhibit positive and substantial out-of-sample risk-adjusted returns after accounting for fees. DeMiguel et al. (2023) focus on predicting abnormal returns, exclusively relying on fund-specific characteristics without incorporating macroeconomic information. They highlight the necessity of including macroeconomic variables as predictors in their model to capture the dynamics of the model under varying macroeconomic conditions. This approach aims to enhance the model's ability to adapt and provide more robust predictions by considering a comprehensive set of predictors, including both fund-specific characteristics and macroeconomic variables.

Environmental, Social, and Governance (ESG) investing refers to an investment strategy that considers environmental, social, and governance factors in investment decision-making. According to a Deloitte report, globally, the percentage of investors

who apply ESG principles to at least a quarter of their portfolios jumped from 48% in 2017 to 75% in 2019 (Collins, 2020). Some papers have analyzed the relationship between the performance of investment funds and ESG rating. Abdelsalam et al. (2020) evaluated the relationship between environmental and social governance factors and sustainable investment drivers. An examination of over 1,000 studies showed a positive correlation between ESG rating and financial performance (Whelan et al., 2021). About 53% of papers focused on performance-based ESG ratings found a positive association with fund performance. However, only 26% of papers that focused on a general statement of commitment to ESG values found a positive correlation with fund financial performance. These studies suggest that there is a relationship between mutual fund performance and ESG factors, but the exact nature of that relationship may depend on the specific ESG measures used.

Ruf et al. (2019) compare the risk-adjusted returns of socially responsible mutual funds (SMRF) with funds rated by Morningstar Portfolio ESG Score<sup>TM</sup> grouped into low, medium, and high ratings, during 2005 to 2016 and domiciled in the United States. They find that during the period of pre-recession boom, the funds with high-ESG ratings had a significantly higher risk-adjusted performance than did funds with a low-ESG or mid-ESG rating. This trend continued during the early parts of the Great Recession. During periods of economic recovery and growth, SRMF rated low on ESG performed significantly better than highly-rated SRMFs. For medium-rated SRMFs, other than prior to the Great Recession, fund performance was not consistently significantly different from highly-rated SRMFs.

Dolvin et al. (2019) find that funds with high sustainability scores have about the same risk-adjusted returns (i.e., alphas) as other funds. Thus, Socially Responsible Investing (SRI) investors can apparently follow a social mandate without sacrificing financial performance, but also without garnering any incremental financial benefit. They find, however, that most high-sustainability funds are concentrated in the large-cap space, which implies that strict adherence to social criteria could inadvertently result in less diversified investor portfolios. They also find that funds with high Morningstar sustainability scores generally mimic those of self-proclaimed SRI funds, suggesting that the new metric opens a larger pool of potential funds for investors focused on SRI. Steen et al. (2020) analyze the relationship between Morningstar's ESG ratings and the performance of 146 mutual funds domiciled in Norway. Dividing the sample into ESG quintiles, they find no evidence of rating level effects, nor do they find any abnormal risk-adjusted returns (alphas). However, analyzing the European categorized funds separately, they find significantly higher returns and positive alphas for the top ESG quintiles. Abate et al. (2021), using data envelopment analysis, provide evidence of the superior efficiency of funds investing in high ESG-rated securities. Indeed, investment policies based on ESG criteria seem to enjoy a competitive advantage because of their inclusion of non-financial data. Moreover, they find that funds with high ESG ratings are subject to lower ongoing charges despite their more complex screening activities and associated operating costs. Raghunandan and Rajgopal (2022) verify whether ESG mutual funds invest in firms that have stakeholder-friendly track records, and they find that socially responsible funds do not appear to follow through on proclamations of concerns for stakeholders. Xidonas and Essner (2024) introduce a portfolio optimization model based on multi-objective

minimax principles, aiming to maximize risk performance across the three standard ESG investment objectives. Their research reveals that the ESG portfolios optimized through this model consistently outperform their respective market benchmarks, yielding higher risk-adjusted returns.

Finally, Doukas et al. (2022) explore the advantages of employing both traditional classification methods, including ordinal logit, ordinal probit, and linear discriminant analysis (LDA), and machine learning techniques like k-Nearest Neighbors and Support Vector Machines. Their investigation focuses on developing models to predict the performance of energy efficiency investments. The findings reveal that machine learning methods exhibit a moderate improvement in predictive accuracy compared to traditional methods.

Overall, the literature suggests that ESG integration can lead to positive financial performance under certain conditions, but the relationship between ESG and financial performance is complex and it depends on various factors. This study endeavors to contribute to the existing body of knowledge by providing a nuanced and context-sensitive examination of the relationship between ESG and financial performance. By delving into the intricacies of this relationship, this research aims to offer a deeper understanding of how ESG considerations can be harnessed to enhance mutual fund returns. Such insights are valuable for investors, asset managers, and policymakers as they navigate the evolving landscape of responsible investing and strive to integrate ethical and financial imperatives for a more sustainable and equitable future.<sup>1</sup>

The goal of our research is to provide support to the socially responsible investor in the fund picking process. To that end, we perform an empirical analysis with machine learning techniques and Extreme Gradient Boosting is the algorithm we use to determine those fund features that best anticipate fund performance.

### 3 Data Description

The data was collected in 2022 from different sources. First, from Refinitiv Eikon Database (<https://www.refinitiv.com>) we selected US-registered Mutual equity funds with a global geographic scope (including US stocks), an asset type involving active management in shares of any geographic scope, in dollars, with uninterrupted five-year performance records, and a minimum investment of up to €10,000, in order to focus on retail funds and exclude institutional funds. To ensure the consistency of performance ratios, the sample comprises only capitalization funds; therefore, distribution funds are excluded.

Subsequently, the Morningstar database (<https://morningstar.com>) was used to locate each fund in the sample, linking it to a series of relevant variables that the Morningstar company associates with investment funds. First, in order to show recommendation of these funds, they are assigned stars<sup>2</sup>. Second, the number of

<sup>1</sup> For a more detailed overview of the contributions of the papers that have been published in the relevant field, please refer to Annex 1.

<sup>2</sup> Review Graham et al. (2019) for further detail.

years that the fund manager has been managing each fund is considered. Finally, a variable is included that quantifies the investment style of each fund.

Finally, the Citywire database (<https://citywireselector.com>) is used to specify whether the fund manager is mentioned in this database and to include the fund manager rating when available.

Consequently, there is a sample of 262 funds of which the following variables are available: ESGSCORE, TER, FUNDTNA, MSSTARS, STYLEMATRIX, YEARS-MANAGER, CITYWIRE, CWRATING and a series of variables for a time horizon of five years: ANNUAL RETURN, ANNUAL STANDARD DEVIATION, ALPHA, BETA (Annex II).

## 4 Methodology

Using financial variables, we intend to build a good regression prediction model capable of identifying the most relevant features or key factors that anticipate fund performance. The objective of this regression model is to determine which features are involved in a fund that earns strong returns, while preventing overfitting and making generalizable predictions. We use the algorithm Extreme Gradient Boosting (XGBoost) to achieve this purpose, which belongs to the family of gradient boosting algorithms. This technique is particularly useful in the context of regression situations and yields high predictive models. Boosting models combine large numbers of relatively simple tree models adaptively, to optimize predictive performance; it is particularly useful and frequently used in the context of regression problems, it can improve predictions for many regression methods, and it reduces the variance of a statistical learning method (James et al., 2017). Boosting is a sequential technique that works on the principle of an ensemble, averaging many tree models, which makes it possible to achieve high accurate predictions.

Chen and Guestrin (2016) indicate XGBoost is an effective and scalable implementation of gradient boosting machine framework by Friedman (2001) and Friedman et al. (2000). Boosting consists of creating multiple trees which are then brought together to produce a single combined prediction which substantially improves regression predictions. Hastie et al. (2009) point out the progress of the trees is sequential, that is, each tree grows using information from previously grown trees attempting to correct the mistakes of the previous ones, and each tree is fitted on a modified version of the original data set; therefore, boosting the construction of each tree depends strongly on the trees that have already been grown. Accordingly, this algorithm, in each iteration, fits a new tree using the current residuals, rather than the response variable; then the algorithm adds this new decision tree into the fitted function to update the residuals, sequentially improving the model (James et al., 2017). XGBoost adds new trees until no further improvement can be accomplished, where each subsequent model attempts to correct the errors of the previous model, and these succeeding models are dependent on the previous model.

The general idea behind gradient boosting is to generate and combine multiple weak learners—which are essentially decision trees—to produce a strong rule; therefore, gradient boosting combines multiple decision trees to lower the model error.



After multiple iterations, the weak learners are combined to form a strong learner that will predict a more accurate outcome. XGBoost is a tree ensemble model which contains a set of classification or regression trees. Accumulating multiple tree predictions, the tree ensemble model overcomes the limitations or flaws of single trees regarding low predictive results. Climent et al. (2019) and Carmona et al. (2019) applied XGBoost to predict bank failure in the Eurozone and the US using classification trees, however this paper applies XGBoost based on regression trees.

One of the key elements of XGBoost models is the use of regularization, which reduces overfitting compared to other boosting techniques and allows for model generalization with new data. Regularization or feature penalization controls variable weights or penalizes complexity, performing variable selection, and decreasing high-dimensional situations. XGBoost includes regularization as part of the learning objective, as contrasted with simple gradient boosting and random forests. Moreover, XGBoost works by sequentially adding predictors to an ensemble, each one correcting its predecessor; however, instead of tweaking the instance weights at every iteration, as AdaBoost does, this method tries to fit the new predictor to the residual errors made by the previous predictor (Paul, 2021).

Training an XGBoost algorithm requires the specification of some important hyper-parameters (Kuhn & Vaughan, 2021; and <https://xgboost.readthedocs.io/en/stable/> (visited on December 2022):

- *mtry*. The number of predictors (variables) or proportion that will be randomly sampled at each split when creating the tree models.
- *trees*. The number of trees contained in the ensemble or number of iterations or rounds.
- *min\_n*. The minimum number of observations in a node that is required for the node to be split further. The larger the number, the more conservative the algorithm.
- *tree\_depth*. The maximum depth of the tree or number of splits. Increasing this value will make the model more complex and more likely to overfit.
- *learn\_rate*. The learning rate, or shrinkage parameter, at which the boosting algorithm adapts from iteration to iteration. It is a small positive number that determines the contribution of each tree to the growing model. Typical values are 0.01 or 0.001. To achieve a good performance, a very small value of this parameter can require a very large number of trees.
- *loss\_reduction*. The minimum reduction in the loss function required to split or make a further partition on a leaf node of the tree. The larger it is, the more conservative the algorithm and the higher the regularization.
- *sample\_size*. The number of observations or proportion exposed to the fitting routine. Subsampling occurs once in every boosting iteration.

It is not possible to estimate the best values of these hyperparameters from data; they have to be estimated by the practitioner according to his experience or through tune techniques, such as cross-validation. Hyper-parameter tuning consists of detecting the set of the best hyper-parameters that produce a model with the highest performance on unseen data. To discover the best composition, it is quite usual to create a

*Grid*, that is, a list or range of the most suitable values for each hyper-parameter, and then train all possible models resulting from all parameter combinations. Sometimes, when there are too many parameters' configurations, the training process is very time-consuming and the practitioners abort the process after reaching an optimal result.

*loss\_reduction* (or *gamma*) is the regularization parameter that characterizes the XGBoost algorithm from other gradient boosting techniques—XGBoost is a regularized version of gradient boosting which prevents overfitting by shrinking the leaves weights to make the boosting process more conservative. On the other hand, the decision on the number of trees or iterations (*trees*) is very important; in our study, we will take a value of 1,000, because it is very unlikely to get an improvement after this number of iterations as adding more trees beyond a limit does not enhance the model performance.

As mentioned above, we employ cross-validation methods to determine an optimal combination of the XGBoost hyperparameters; in particular, we choose 10-fold cross-validation ( $k=10$ ). Training data is randomly partitioned into ten subsets or folds of equal size, obtaining an overall estimation of out-of-sample error for each parameter configuration. Using the best parameter selection yielding the highest performance, a new and final model containing all training data is then fitted.

As usual in these type of Machine Learning studies, we divide the available data of investment funds into two different sets: 80% of the observations to train the XGBoost model to find the best model's parameters configuration and the remaining 20%, as a hold-out sample, to the test data or to measure the performance of the best fitted model on independent data not used to train the model.

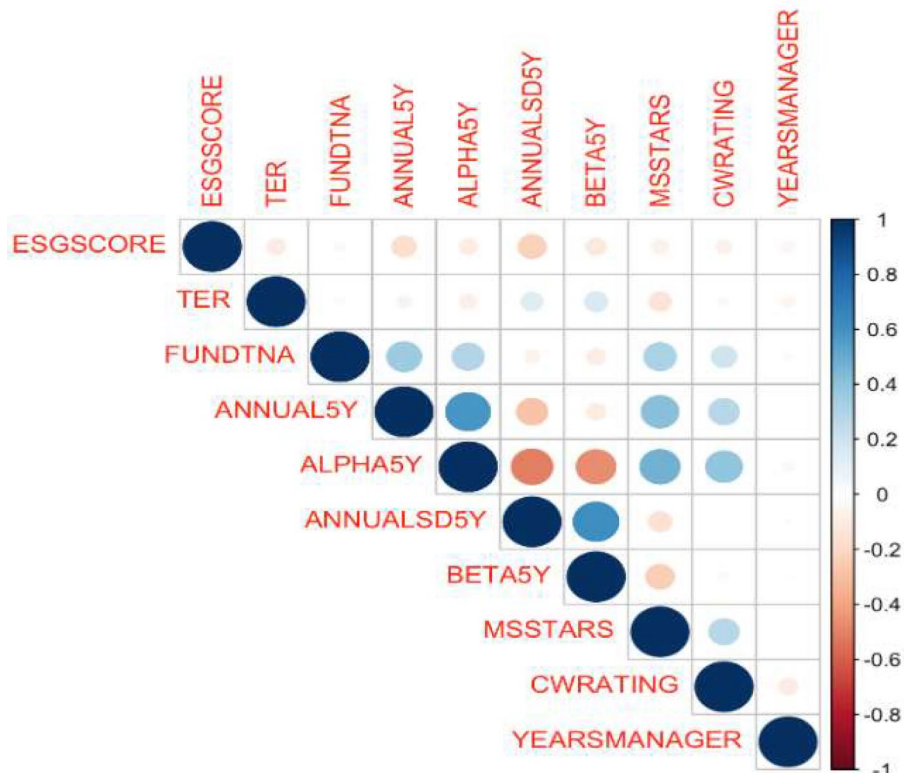
All models were fitted in *R* (R Core Team, 2022) version 4.1.3, and for XGBoost using *h2o* package version 3.36.0.1 (LeDell et al., 2022) and *tidyverse* package version 1.3.2 (Wickham, 2022). Additionally, for model interpretability, we used DALEX package version 2.4.0 (Biecek et al., 2022).

## 5 Results

As a precautionary measure, we ensure there is no correlational problem among the different predictors and the response variable. The size of the dots and the absence of an intense color indicates the nonexistence of this potential problem (see correlogram plot in Fig. 1). A correlogram is a graph that illustrates the correlation matrix, representing each correlation coefficient with a dot color and size according to its value.

Annex III offers a brief summary of the fundamental descriptive statistics, providing a condensed overview that captures essential features of the dataset. These statistical measures offer valuable insights into the dataset's characteristics, aiding in a more complete comprehension of the variables and their variations.

We tune and fit a model considering all variables described in Annex II. The objective is to distinguish the most relevant variables, with the highest predictive content in investment funds. This process requires the ideal values of the hyper-parameters mentioned above to be identified to produce a good XGBoost model. To measure model performance, we take different measures related to model residuals, which are evaluated on an independent test set or hold-out sample, such as root mean squared



**Fig. 1** Correlogram plot of all predictors

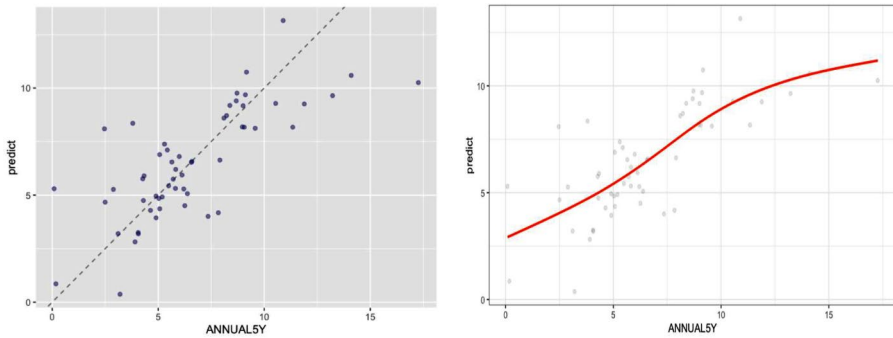
**Table 2** XGBoost cross-validation hyper-parameter tuning,  $k=10$

| Parameter                                                        | Value |
|------------------------------------------------------------------|-------|
| Number of predictors randomly sampled ( <i>mtry</i> )            | 50%   |
| Number of trees contained in the ensemble ( <i>trees</i> )       | 510   |
| Minimum number of observations in a node ( <i>min_n</i> )        | 1     |
| Maximum depth of the tree ( <i>tree_depth</i> )                  | 3     |
| Learning rate, or shrinkage parameter ( <i>learn_rate</i> )      | 0.091 |
| Minimum reduction in the loss function ( <i>loss_reduction</i> ) | 0.001 |
| Amount of observations or proportion ( <i>sample_size</i> )      | 80%   |

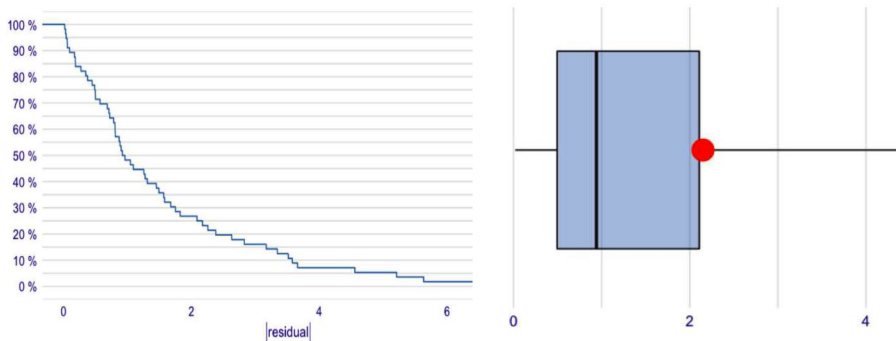
error or coefficient of determination ( $R^2$ ). Therefore, fitting a model—while avoiding over-fitting—that could be generalized on independent data not used during the training process is crucial to tune and choose the best XGBoost model parameters' values.

Using the R *h2o* package (LeDell et al., 2022) and using 10-fold cross-validation ( $k=10$ ), we explore the optimal model's hyper-parameter configuration on the training data (80% of all data). Table 2 shows the parameters' values for the best model after an iteration search process of 100 combinations.

The best XGBoost model—fitted taking the hyper-parameters from Table 1—returns the following resample results using the aforementioned 10-cross-validation:



**Fig. 2** Predicted vs. actual values on test data. XGBoost model



**Fig. 3** Reverse cumulative distribution of residuals' absolute value (left panel) and boxplot of residuals' absolute value (right panel). XGBoost model on test data. *Note:* Red dot represents the mean value of the absolute value of the residuals

a quite reasonable  $R^2$  of 0.618 and a low root mean squared error of 2.039, which prove that the model is quite good. Relating to the hold-out sample or testing data (remaining 20% of all data), we even obtained a high  $R^2$  of 0.694 and also a low root mean squared error of 2.150. That the latter results are not really worse than the former clearly indicates that over-fitting is not present on the final fitted XGBoost model.

Additionally, Fig. 2, left panel, displays predicted and actual funds' values for the test data; as most of the dots are close to the diagonal line, model residuals are quite low, confirming the goodness of fit. The right panel of Fig. 2 shows a smooth regression curve fitted for all residuals; as the result is a relatively straight line, it implies a very good model performance.

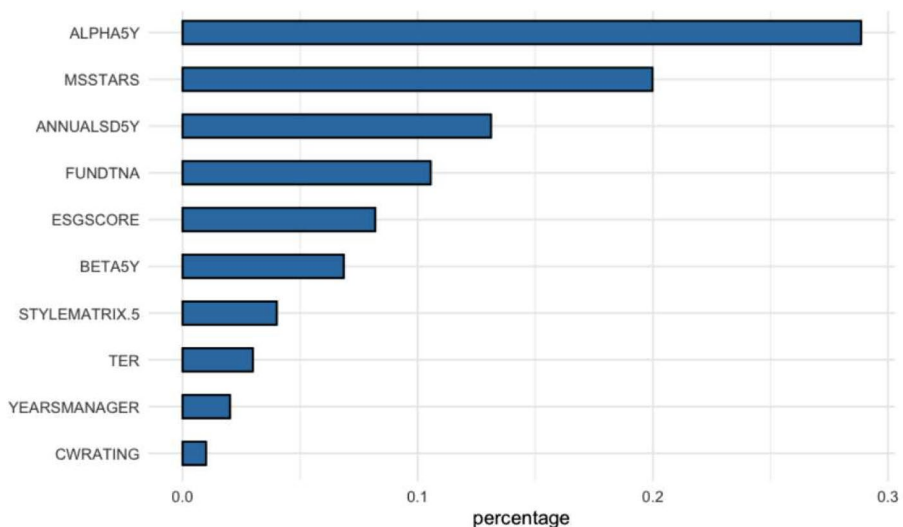
In fact, most of the residuals have very small values, as can be observed in the reverse cumulative distribution of residuals' absolute value (Fig. 3, left panel) and in the boxplot of residuals' absolute value (Fig. 3, right panel).

### 5.1 Model Interpretability

To help to understand the results of the model, that is the relationship between the predictors and the prediction target, the global interpretation of the model usually provides interesting insights. In this way, Fig. 4 shows the features that have a higher contribution to the XGBoost fitted model for fund financial performance. The importance of these features is calculated through Breiman's permutation method (2001). The values of an important feature in the training data are randomly permuted, producing a degradation of the training performance, and destroying any relationship between that feature and the response variable, which facilitates the identification of the most important features.

As can be noted, our XGBoost model has revealed that the variable ESG is one of the most important variables impacting mutual funds' performance. Our findings suggest that ESG is one of the key factors to be prioritized to improve the financial performance and long-term sustainability of mutual funds.

Besides the feature importance, another popular tool for a model's global interpretability is the partial dependence plot. This is a model visualization technique that provides graphical insights into the prediction behavior of a machine learning algorithm, such as XGBoost. These kinds of graphs show how the expected model response behaves as a function of a selected feature, which is an average profile for all observations (Biecek & Burzycowski, 2020). Although partial dependence plots are not a perfect representation of the captured effect and while this may not provide a comprehensive description, they can show overall trends and provide a valuable basis for interpretation, explaining the marginal effect of each one of the predictors on the predicted variable, after accounting for the average joint effect of the other predictors (Friedman, 2002). Accordingly, and following Hall and Gill (2019), partial dependence plots offer a global interpretation of how a model's predictions vary based on

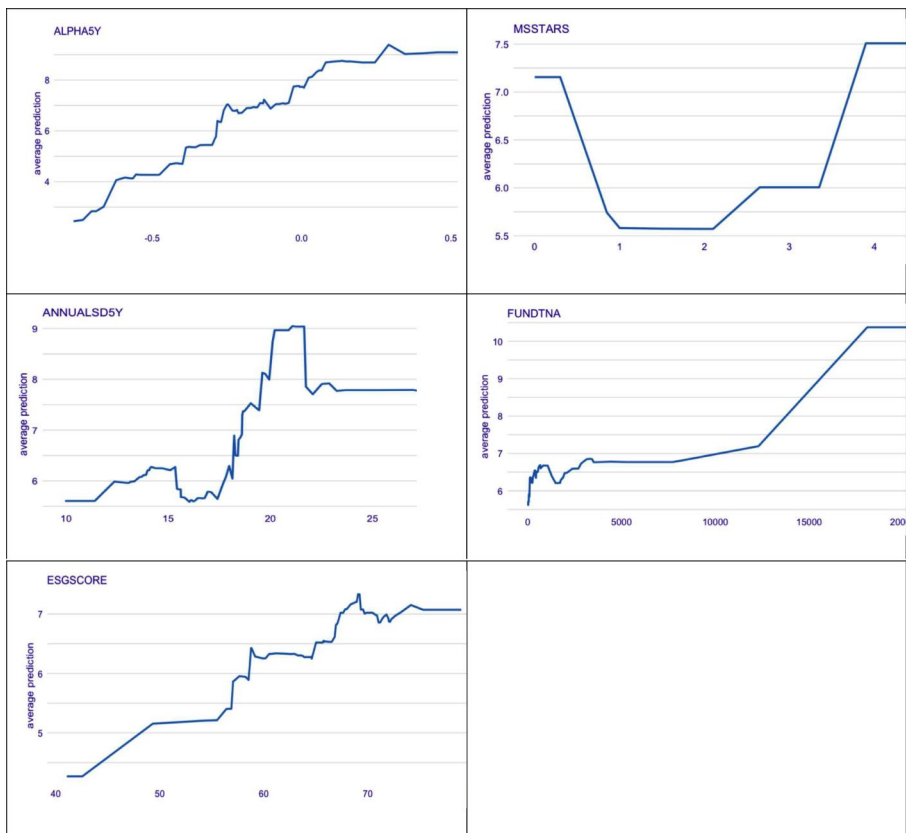


**Fig. 4** Variable importance. XGBoost model

certain features, showing the average manner in which machine-learned response functions change based on the values of one or two input variables of interest, while averaging out the effects of all other input variables. These types of illustrations provide an easy and intuitive explanation that has been noted as practical for different purposes (Natekin & Knoll, 2013).

Figure 5 shows the partial dependence plots for the five most important variables in the final XGBoost model. The partial dependence plots indicate a positive relationship between the dependent variable (fund performance) and the following independent variables: Alpha, Morningstar fund rating (number of stars), Annual Standard Deviation (as long as volatility is under 20%), fund size (measured in total net assets), and ESG score. In addition, the plots suggest a negative relationship between fund performance and Annual Standard Deviation when volatility exceeds 20%. Consequently, the results are rather consistent with the economic interpretation of the variables.

Within the framework of our study, the prominence of the ESG factor emerges as a fundamental determinant of mutual fund performance. The empirical evidence derived from our model supports the notion that mutual funds boasting robust ESG



**Fig. 5** Partial dependence plots

ratings are predisposed to exhibit superior financial performance. This substantiates our hypothesis, aligning with the anticipated positive relationship between elevated ESG scores and the financial prowess of mutual funds.

The association between ESG ratings and financial performance underscores the multifaceted impact of environmental, social, and governance considerations on investment outcomes. By promoting strong ESG practices, mutual funds not only exhibit a commitment to sustainable and responsible investing but also position themselves favorably within the competitive financial landscape.

In addition to global interpretability, there is also a need to understand a model or predictions for a single row of data or a group of similar rows, which is known as a model's local interpretability. Instance-level or local approaches help understand how a model produces a prediction for an individual observation. Local explanations enhance understanding by creating accurate explanations for each observation in a dataset. The principal idea is to estimate the contribution of an explanatory variable to the model's prediction as a shift in the expected model response after conditioning on other variables. To this end, break-down plots provide a detailed summary of the effects of each particular feature on the expected model response that is easy to understand (Biecek & Burzycowski, 2020). For a particular observation of the considered funds, the XGBoost predicted value is broken down into the impact of each individual feature.

Note that the MSSTARS partial dependence plot includes six types of funds, unrated funds which are assigned zero stars and funds with 1 to 5 stars. The fact that we assigned zero stars to unrated funds explains that their performance can be higher than that of funds rated with 1 to 3 stars.

In Annex IV, we show the use of break-down plots, regarding two observations from the test set, in particular, a prediction with a high value of the response variable, and a prediction with a low value of the response variable.

## 6 Conclusions

In pursuit of the objectives outlined in this study, we carried out a comprehensive empirical analysis utilizing the robust machine learning technique known as Extreme Gradient Boosting (XGBoost). The primary aim was to unravel the multifaceted relationship between mutual fund characteristics and performance, with a particular focus on the essential role played by Environmental, Social, and Governance (ESG) ratings as a key determinant of fund performance.

Our research has yielded significant findings that underscore the profound influence of ESG ratings within the complex landscape of mutual fund performance prediction. ESG score emerged as one of the top five most influential factors in predicting mutual fund performance, thus affirming its status as a crucial dimension of the modern investment landscape. This finding highlights the potential for investors to align their financial aspirations with their commitment to sustainability and ethical investing without any compromise on financial returns.

The implications of our findings extend far beyond the boundaries of this study. For socially responsible investors, this revelation offers a clear path to harmonizing

their values with their investment choices. It empowers them to make investment decisions that not only support their ethical principles but also align with their financial objectives. Additionally, asset managers, financial advisors, and market participants should take heed of this discovery, recognizing the increasingly prominent role of ESG considerations in shaping the investment environment. The integration of ESG criteria into investment strategies is no longer merely a matter of ethical alignment but an important factor for optimizing financial performance.

Therefore, our research underscores the significance of ESG ratings as one of the key factors in predicting mutual fund performance. This discovery dispels the notion of a trade-off between sustainability and financial returns, signaling a transformative shift in the investment paradigm. As we navigate an era where ethical and financial imperatives converge, this study serves as a catalyst for responsible investing practices, driving us toward a future where sustainable investment choices are synonymous with financial prudence, ultimately advancing both the goals of investors and the broader global community towards a more equitable and sustainable future.

In summary, this finding carries substantial implications for investors, asset managers, and financial advisors seeking to optimize their investment strategies. It emphasizes the potential for ESG integration to serve as a facilitator for financial success, offering a compelling case for the alignment of ethical and financial objectives. Moreover, our research underscores the need for further exploration into the complex mechanisms through which ESG factors exert their influence on mutual fund performance, thus providing a fertile ground for future investigations in the realm of sustainable finance.

A remarkable finding is that Alpha is the variable that best predicts fund performance by showing the impact of manager skill on fund performance. On the contrary, no relationship is found between fund total expenses and fund return. Therefore, better fund management does not necessarily involve higher management costs.

An interesting outcome is the positive relationship observed between agency ratings and fund performance. Morningstar's rating system seems to have a relevant explanatory capacity for fund performance. Conversely, the fact that a fund manager is rated or even listed in the Citywire database does not have a significant explanatory power for fund performance.

Fund volatility, as measured by the annual standard deviation of fund returns, explains fund performance and investors willing to take some fund volatility will benefit from higher fund returns. However, as return volatility exceeds 20%, fund return declines.

All things considered, the findings of this research will help both investors and advisors focus on those fund characteristics that best anticipate fund performance. In addition, conscientious investors with a strong commitment to sustainability will find it easier to pick performing funds.

The main limitations of this study relate to the type of data and the nature of the sample. First, the empirical analysis is based on cross-sectional data and as longer data series of ESG ratings become available more robust studies may be done with time-series data for periods of market volatility, downturn, and expansion. Second, the funds comprising our sample are US-registered equity funds with a global geographic scope and further research on other fund categories is required to confirm the



generalization of our results. Moreover, by selecting funds with uninterrupted five-year performance records there may be some survival bias in the mutual fund sample. An alternative approach to handle funds with missing data might improve the reliability and robustness of results. Finally, the number of sustainable funds in Europe significantly exceeds the number of sustainable funds in the US. For this reason, we look forward to exploring the performance of sustainable mutual funds in Europe in future research works.

## Annex I: Most representative contributions of the papers that have been published in the relevant field

| Paper                  | Key contribution                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
|------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| DeMiguel et al. (2023) | <ul style="list-style-type: none"> <li>- Application of Machine Learning: Demonstrates the integration of machine learning techniques in the process of selecting mutual funds with positive alpha. By leveraging advanced computational methods, the study aims to enhance the accuracy and efficiency of fund selection processes.</li> <li>- Utilization of Fund Characteristics: Emphasize the importance of incorporating fund characteristics into the selection process. By analyzing various attributes and features of mutual funds, the study aims to identify patterns and relationships that contribute to positive alpha generation.</li> <li>- Enhancing Alpha Generation: The primary goal of the research is to enhance the ability to identify mutual funds that have the potential to generate positive alpha. By leveraging machine learning algorithms and considering fund characteristics, the study seeks to provide investors with improved tools for making informed investment decisions.</li> <li>- Contribution to Financial Economics: Offering insights into the application of machine learning in fund selection. It addresses the challenge of identifying mutual funds with positive alpha, which is of significant interest to both academics and practitioners in the finance industry.</li> </ul> |
| Kaniel et al. (2023)   | <ul style="list-style-type: none"> <li>- Application of Machine Learning: The paper applies machine learning techniques to assess the skill of mutual fund managers.</li> <li>- Skill Assessment: the study provides insights into the skill levels of mutual fund managers. This contributes to the ongoing discussion in finance literature regarding the effectiveness of active management in generating alpha.</li> <li>- Empirical Analysis: The paper likely includes an empirical analysis that demonstrates the effectiveness of the machine learning approach in identifying skillful mutual fund managers. This empirical evidence is crucial for validating the applicability and usefulness of the proposed methodology.</li> <li>- Contribution to Financial Economics: The research contributes to the broader field of financial economics by exploring new methodologies for evaluating fund manager skill, which can have implications for investor decision-making, portfolio management, and the efficiency of financial markets.</li> </ul>                                                                                                                                                                                                                                                                       |

| Paper                | Key contribution                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Doukas et al. (2022) | <ul style="list-style-type: none"> <li>- Comparative Analysis of Energy Efficiency Investments: Examining different investment strategies and approaches, the study aims to assess the success and effectiveness of energy efficiency initiatives in various contexts.</li> <li>- Classification of Investment Strategies: Categorize and classify different investment strategies related to energy efficiency. This classification helps investors and policy-makers understand the diverse approaches available for investing in energy efficiency projects and initiatives.</li> <li>- Performance Prediction: Offers insights into predicting the performance of energy efficiency investments. Employing computational techniques and statistical models, the research seeks to identify factors and variables that influence the success or failure of energy efficiency projects.</li> <li>- Contribution to Financial Economics: Applying advanced computational methods to analyze energy efficiency investments. It enhances our understanding of the economic and financial aspects of energy efficiency initiatives and their impact on sustainability and environmental conservation.</li> </ul>                                                                                                                                                                                                                                                                                                          |
| Kanade et al. (2022) | <ul style="list-style-type: none"> <li>- Application of Machine Learning: The paper focuses on the application of machine learning algorithms to predict suitable mutual funds of Indian companies. By leveraging advanced computational techniques, the study aims to improve the accuracy and efficiency of mutual fund selection processes tailored to the Indian market.</li> <li>- Evaluation of Algorithm Performance: Comprehensive evaluation of various machine learning algorithms to assess their performance in predicting the right mutual funds. This analysis provides valuable insights into the strengths and weaknesses of different algorithms in the context of mutual fund selection.</li> <li>- Customization for Indian Market Dynamics: The research specifically addresses the dynamics of the Indian market, acknowledging the unique characteristics and challenges associated with mutual fund selection in this context. By focusing on Indian companies, the study offers tailored insights and recommendations for investors operating within this market.</li> <li>- Contribution to Financial Economics: The intersection of machine learning and finance by applying advanced computational techniques to the domain of mutual fund selection. By highlighting the efficacy of machine learning algorithms in predicting suitable mutual funds, the study adds to the growing body of research on the application of artificial intelligence in financial decision-making.</li> </ul> |
| Li and Rossi (2020)  | <ul style="list-style-type: none"> <li>- Application of Machine Learning: The paper introduces a machine learning approach for selecting mutual funds based on the stocks they hold. This methodology leverages advanced computational techniques to analyze the composition of mutual fund portfolios.</li> <li>- Portfolio Analysis: By focusing on the stocks held within mutual fund portfolios, offers insights into the underlying investment strategies and preferences of fund managers. This approach provides a more granular understanding of fund performance drivers.</li> <li>- Predictive Modeling: Involves predictive modeling techniques to assess the future performance potential of mutual funds based on their current stock holdings. This aspect is crucial for investors seeking to optimize their investment decisions.</li> <li>- Data-Driven Decision Making: Through the application of machine learning, the paper promotes data-driven decision-making in the context of mutual fund selection. This aligns with the broader trend in finance towards incorporating quantitative methods for investment analysis.</li> <li>- Contribution to Financial Economics: The research likely discusses the practical implications of its findings for investors, fund managers, and other stakeholders in the financial industry. It may provide recommendations or guidelines for utilizing the machine learning approach in real-world investment scenarios.</li> </ul>                       |

## Annex II: Explanatory variables

|                                |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|--------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ESG SCORE                      | Series of relative peer rankings used to evaluate a company's ESG (Environment, Social and Governance) performance at the metric, category, pillar and summary levels. For example, Scores within 0 to 25 (50 to 75) indicates poor (good) relative ESG performance and insufficient (above average) degree of transparency in reporting material ESG data publicly.                                                                                                                                                                                                                   |
| TER                            | The cost of managing a fund that is expressed as a percentage of the assets under management, the TER accounts for all the expenses incurred to run the show.                                                                                                                                                                                                                                                                                                                                                                                                                          |
| FUND TNA                       | The net value of an entity and is calculated as the total value of the entity's assets minus the total value of its liabilities.                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| MS STARS                       | The Morningstar Rating is a measure of a fund's risk-adjusted return, relative to similar funds. Funds are rated from 1 to 5 stars, with the best performers receiving 5 stars and the worst performers receiving a single star.                                                                                                                                                                                                                                                                                                                                                       |
| STYLE MATRIX                   | The Morningstar Style Box is a nine-square grid – with three stock investment styles for each of three size categories: 'small', 'mid' and 'large'. Two of the three style categories are 'value' and 'growth', while the third is 'blend' (funds that own a mixture of growth and value stocks). For example, 1: Large-Cap Value Fund; 4: Large-Cap Blend Fund; 7: Large-Cap Growth Fund.                                                                                                                                                                                             |
| YEARS<br>MANAGER               | The number of years that the current manager has been the portfolio manager of the fund. For funds with more than one manager, the tenure of the manager who has been with the fund the longest is shown.                                                                                                                                                                                                                                                                                                                                                                              |
| CITYWIRE                       | Dichotomous variable that indicates whether the fund manager is listed in the Citywire database.                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |
| CW RATING                      | Citywire Fund Manager Ratings measure performance across all the funds run by a given manager. The ratings provide a clear evaluation of a manager's performance against their direct competitors.                                                                                                                                                                                                                                                                                                                                                                                     |
| ANNUAL RETURN                  | Annual total returns are calculated on a calendar-year and year-to-date basis. Total return includes both capital appreciation and dividends. The year-to-date return is updated daily.                                                                                                                                                                                                                                                                                                                                                                                                |
| ANNUAL STAN-<br>DARD DEVIATION | Annual standard deviation is calculated on a calendar-year and year-to-date basis. Standard deviation measures the dispersion around an average. For a mutual fund, it represents return variability. A higher standard deviation implies a wider predicted performance range and greater volatility. Morningstar calculates total return by taking the change in a fund's NAV (Net Asset Values), assuming the reinvestment of all income and capital gains distributions (on the actual reinvestment date used by the fund) during the period, and then dividing by the initial NAV. |
| ALPHA                          | Alpha gauges how well a manager can pick stocks and measures a mutual fund manager's or strategy's effectiveness. It shows the difference between a fund's actual returns and its expected performance, given its level of risk as measured by beta. A positive alpha indicates the fund has performed better than its beta would predict. In contrast, a negative alpha means the fund performed worse than expected given its beta. Alpha is also after fees, meaning the fund must overcome its management fees as well as its beta to have positive alpha.                         |
| BETA                           | A fund's beta is a measure of its sensitivity to market movements. Morningstar calculates beta by comparing a fund's excess return over Treasury bills to the market's excess return over Treasury bills, so a beta of 1.10 shows that the fund has performed 10% better than its benchmark index in up markets and 10% worse in down markets, assuming all other factors remain constant.                                                                                                                                                                                             |

## Annex III: Descriptive statistics overview

This annex offers a comprehensive examination of key descriptive statistics, providing essential insights into the dataset's variables and their characteristics. These statistics offer a glimpse into the central tendencies, variability, and the overall distribution of the data.

|              | <i>n</i> | mean    | sd      | median | min   | max      | skew  |
|--------------|----------|---------|---------|--------|-------|----------|-------|
| ESGSCORE     | 262      | 66.79   | 7.26    | 68.14  | 41.06 | 81.86    | -1.04 |
| TER          | 262      | 1.75    | 0.46    | 1.83   | 0.04  | 3.19     | -0.95 |
| FUNDTNA      | 262      | 1596.09 | 2948.59 | 558.06 | 5.32  | 21813.27 | 3.73  |
| ANNUAL5Y     | 262      | 6.57    | 3.33    | 6.24   | -2.15 | 17.27    | 0.55  |
| ALPHA5Y      | 262      | -0.16   | 0.29    | -0.14  | -1.06 | 0.54     | -0.23 |
| ANNUALSD5Y   | 262      | 17.74   | 4.71    | 16.50  | 9.95  | 46.25    | 2.52  |
| BETA5Y       | 262      | 0.94    | 0.15    | 0.95   | 0.44  | 1.67     | 0.18  |
| MSSTARS      | 262      | 3.16    | 1.24    | 3.00   | 0.00  | 5.00     | -0.73 |
| CITYWIRE     | 262      | 1.08    | 0.27    | 1.00   | 1.00  | 2.00     | 3.07  |
| CWRATING     | 262      | 0.74    | 1.03    | 0.00   | 0.00  | 4.00     | 1.30  |
| STYLEMATRIX  | 262      | 3.98    | 2.10    | 4.00   | 1.00  | 9.00     | 0.34  |
| YEARSMANAGER | 262      | 7.37    | 5.20    | 6.00   | 0.00  | 27.00    | 1.17  |

### 1. ESGSCORE (Environmental, Social, and Governance Score):

- Mean: 66.79, Median: 68.14.
- The scores are relatively high, suggesting good ESG performance overall.
- Negative skewness (-1.04) indicates more scores on the higher side.
- Standard deviation (7.26) suggests moderate variability in ESG performance among the entities.

### 2. TER (Total Expense Ratio):

- Mean: 1.75%, Median: 1.83%.
- The expenses related to managing funds vary moderately (sd=0.46).
- The negative skewness (-0.95) suggests a concentration of funds towards lower expense ratios.

### 3. FUNDTNA (Fund's Total Net Assets):

- Mean: 1596.09, Median: 558.06.
- Huge difference between mean and median, indicating a few funds with very high assets skewing the average.
- Very high standard deviation (2948.59), indicating significant disparity in fund sizes.

### 4. ANNUAL5Y (5-Year Annual Return):

- Mean: 6.57%, Median: 6.24%.

- Relatively consistent performance across funds ( $sd=3.33$ ).
- Slight positive skewness (0.55), indicating a few funds with exceptionally high returns.

#### **5. ALPHA5Y (5-Year Alpha):**

- Mean: -0.16, Median: -0.14.
- Most funds have slightly poor manager's or strategy's effectiveness.
- Low standard deviation (0.29) suggests most funds have similar fund manager's or strategy's effectiveness.

#### **6. ANNUALSD5Y (5-Year Annual Standard Deviation):**

- Mean: 17.74%, Median: 16.50%.
- Indicates moderate to high variability in fund returns.
- High standard deviation (4.71) points towards a few funds with very high return variability.

#### **7. BETA5Y (5-Year Beta):**

- Mean: 0.94, Median: 0.95.
- Most funds have returns moving slightly less than the market.
- Low standard deviation (0.15) suggests similar market sensitivities across the funds.

#### **8. MSSTARS (Morningstar Rating):**

- Mean: 3.16, Median: 3.00.
- Ratings are generally above average.
- Negative skewness (-0.73) indicates a concentration of higher ratings.

#### **9. CITYWIRE:**

- Mean: 1.08, Median: 1.00.
- Most funds' managers are listed in the Citywire database, indicating a prevalence of recognized management.

#### **10. CWRATING (Citywire Rating):**

- Mean: 0.74, Median: 0.00.
- Indicates that many funds have a low Citywire rating.

#### **11. STYLEMATRIX (Morningstar Style Box):**

- Mean: 3.98, Median: 4.00.
- Suggests a balance between different investment styles among the funds.

## 12. YEARSMANAGER (Years with Current Manager):

- Mean: 7.37 years, Median: 6.00 years.
- Indicates a relatively long tenure for fund managers, suggesting experienced management across funds.

## Annex IV: Break-down plots

Employing break-down plots, the XGBoost fund's prediction model is explained in Fig. A1 for two different cases, taken from the test set. The first case shows a prediction with a high value of the response variable, and the second case shows a low predicted value. The intercept or mean value of all funds is 6.523. This is the average of the fund value on the test data; and it is not the mean of all fund's observations, but the average model response, so a different model would produce a different average.

It is important to note that the lower plot was built by randomly selecting an observation where the fitted XGBoost model predicts a low value of fund performance. Conversely, the upper plot was built by randomly selecting an observation where the fitted XGBoost model predicts a high value of fund performance.

If we focus on the first case—a high predicted value—the model prediction of the fund value is 9.281, which is higher than the average prediction. The prediction is broken down into the influence of each individual feature; precisely, it breaks down the predicted fund value for this given observation. Step-by-step, the fund value is calculated in the following way:

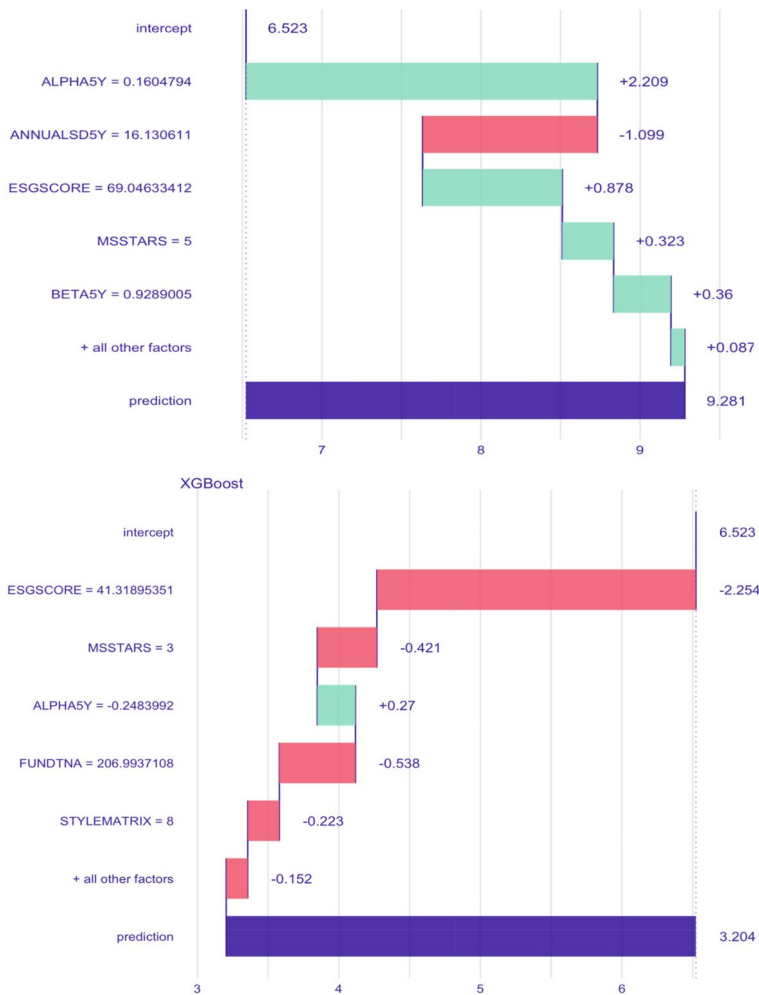
+6.523: Intercept (Average).  
 +2.209: ALPHA5Y=0.160 [prediction is now 8.732].  
 −1.099: ANNUALSD5Y=16.131 [prediction is now 7.633].  
 +0.878: ESGSCORE=69.046 [prediction is now 8.511].  
 +0.323: MSSTARS=5 [prediction is now 8.834].  
 +0.360: BETA5Y=0.929 [prediction is now 9.194].  
 +0.087: all other factors [final prediction is 9.281].

If we consider the second observation—an instance with a low predicted value—the step-by-step representation of the fund value changes as follows:

+6.523: Intercept (Average).  
 −2.254: ESGSCORE=41.319 [prediction is now 4.269].  
 −0.421: MSSTARS=3 [prediction is now 3.848].  
 +0.27: ALPHA5Y=−0.248 [prediction is now 4.118].  
 −0.538: FUNDTNA=206.994 [prediction is now 3.58].  
 −0.223: STYLEMATRIX=7 [prediction is now 3.357].  
 −0.152: all other factors [final prediction is 3.204].

If we consider the test data, this breakdown process can be applied to every single observation to explain all individual XGBoost predictions. It is important to remark that this breakdown implies computing the contributions or impact of each feature

for every tree in the ensemble model, in a similar way to a single decision tree. These impacts are not static coefficients as in a logistic regression. To sum up, each prediction is expressed as the sum of feature impacts. The impact of a feature is dependent on the specific path that the observation took through the ensemble of trees (Foster, 2017).



**Fig. A1** Break-down plots for two observations from the test set. *Notes:* The upper observation has a high predicted value, which is higher than the average prediction. This suggests that ESG has a positive effect on the predicted fund return, which is included in all other factors. On the other hand, the lower observation has a predicted value that is lower than the average, indicating that ESG has a negative effect on the predicted fund return. Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License,

which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abate, G., Basile, I., & Ferrari, P. (2021). The level of sustainability and mutual fund performance in Europe: An empirical analysis using ESG ratings. *Corporate Social Responsibility and Environmental Management*, 28(5), 1446–1455. <https://doi.org/10.1002/csr.2175>.
- Abdelsalam, A., Barake, S., & Elcheikh, A. (2020). Sustainable Investment and ESG Performance. Retrieved from [https://www.researchgate.net/publication/342703817\\_Sustainable\\_Investment\\_and\\_ESG\\_Performance](https://www.researchgate.net/publication/342703817_Sustainable_Investment_and_ESG_Performance).
- Biecek, P., & Burzycowski, T. (2020). Explanatory Model Analysis. Retrieved from <https://pbiecek.github.io/ema/preface.html> [Accessed 21 Sept. 2022].
- Biecek, P., Maksymiuk, S., & Baniecki, B. (2022). DALEX: Descriptive mACHine Learning EXplanations. R package version 2.4.0. Retrieved from <https://cran.r-project.org/web/packages/DALEX/index.html>.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Broadstock, D., Chan, K., Cheng, L., & Wang, X. (2021). The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China. *Finance Research Letters*, 38, 101716. <https://doi.org/10.1016/j.frl.2020.101716>.
- Carmona, P., Climent, F., & Momparler, A. (2019). Predicting failure in the US banking sector: An extreme gradient boosting approach. *International Review of Economics & Finance*, 61, 304–323.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable Tree Boosting System. arXiv:1603.02754v3 [cs.LG].
- Climent, F., Momparler, A., & Carmona, P. (2019). Anticipating bank distress in the Eurozone: An Extreme Gradient boosting approach. *Journal of Business Research*, 101, 885–896.
- Collins, S. (2020). *Advancing environmental, social, and governance investing. A holistic approach for investment management firms*. Deloitte report. Retrieved June 22, from: <https://www2.deloitte.com/us/en/insights/industry/financial-services/esg-investing-performance.html>.
- R Core Team (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Retrieved June 2022, from <http://www.R-project.org/>.
- DeMiguel, V., Gil-Bazo, J., Nogales, F., & Santos, A. (2023). Machine learning and fund characteristics help to select mutual funds with positive alpha. *Journal of Financial Economics*, 150(3), 102737. <https://doi.org/10.1016/j.jfineco.2023.103737>.
- Dolvin, S., Fulkerson, J., & Krukov, A. (2019). Do good guys finish last? The relationship between Morningstar sustainability ratings and mutual fund performance. *The Journal of Investing*, 28(2), 77–91. <https://doi.org/10.3905/joi.2019.28.2.077>.
- Doukas, H., Xidonas, P., & Mastromichalakis, N. (2022). How successful are energy efficiency investments? A comparative analysis for classification and performance prediction. *Computational Economics*, 59, 579–598. <https://doi.org/10.1007/s10614-021-10098-6>.
- Foster, D. (2017). NEW R package that makes XGBoost interpretable. Retrieved from <https://medium.com/applied-data-science/new-r-package-the-xgboost-explainer-51dd7d1aa211> [Accessed December 15, 2021].
- Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5), 1189–1232.
- Friedman, J. H. (2002). Stochastic gradient boosting. *Journal of Computational Statistics & Data Analysis*, 38(4), 367–378.



- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics*, 28(2), 337–407.
- Graham, J. E., Lassala, C., & Ribeiro-Navarrete, B. (2019). A fuzzy-set analysis of conditions influencing mutual fund performance. *International Review of Economics and Finance*, 61, 324–336.
- Hall, P., & Gill, N. (2019). *Introduction to machine learning interpretability* (2nd ed.). O'Reilly Media.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.
- International Investment Funds Association (2021). [https://iifa.ca/page/industry\\_statistics](https://iifa.ca/page/industry_statistics), accessed December 2021.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). *An introduction to statistical learning* (2nd ed.). Springer.
- Kanade, A., Sherdiwala, M., & herekar, S. (2022). Performance of Machine Learning Algorithms to Predict Right Mutual Funds of Indian Companies. In: Gunjan, V.K., Zurada, J.M. (Eds.) *Modern Approaches in Machine Learning & Cognitive Science: A Walkthrough*. Studies in Computational Intelligence, 1027. Springer. [https://doi.org/10.1007/978-3-030-96634-8\\_3](https://doi.org/10.1007/978-3-030-96634-8_3).
- Kaniel, R., Lin, Z., Pelger, M., & Van Nieuwerburgh, S. (2023). Machine-learning the skill of mutual fund managers. *Journal of Financial Economics*, 150(1), 94–138. <https://doi.org/10.1016/j.jfineco.2023.07.004>.
- Kuhn, M., & Vaughan, D. (2021). *parsnip: A Common API to Modeling and Analysis Functions*. R package version 0.1.5. Retrieved June 2022, from <https://CRAN.R-project.org/package=parsnip>.
- LeDell, E., Gill, N., Aiello, S. (2022). *h2o: R Interface for the 'H2O' Scalable Machine Learning Platform*. R package version 3.36.0.1. Retrieved June 2022, from <https://CRAN.R-project.org/package=h2o>.
- Li, B., & Rossi, A. G. (2021). Selecting mutual funds from the stocks they hold: A machine learning approach. Available at SSRN 3737667 papers.[ssrn.com](https://ssrn.com).
- Losada, R. (2022). *Periodic public information on investment funds and how it influences investors' decisions*; CNMV Working Paper N° 76.
- Morningstar (2021). Global Sustainable Fund Flows: Q4 2021 in Review.
- Morningstar (2022). Global Sustainable Fund Flows: Q4 2022 in Review.
- Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, 7, 21. <https://doi.org/10.3389/fnbot.2013.00021>.
- Paul, R. (2021). LightGBM, XGBoost, and CatBoost — Kaggle — Santander Challenge. Retrieve from <https://medium.com/analytics-vidhya/gradient-boosting-lightgbm-xgboost-and-catboost-kaggle-challenge-santander-f3cf0cc56898>. Visited on June 2021.
- Raghunandan, A., & Rajgopal, S. (2022). Do ESG funds make stakeholder-friendly investments? *Review of Accounting Studies*, 27, 822–863. <https://doi.org/10.1007/s11142-022-09693-1>.
- Ruf, B., Das, N., Chatterjee, S., & Sunder, A. (2019). Investments in ESG-Rated mutual funds: Is good better than great? *The Journal of Wealth Management*, 22(1), 56–61. <https://doi.org/10.3905/jwm.2019.1.070>.
- Steen, M., Taghawi, J., & Gjolberg, O. (2020). Is there a relationship between Morningstar's ESG ratings and mutual fund performance? *Journal of Sustainable Finance & Investment*, 10(4), 349–370. <https://doi.org/10.1080/20430795.2019.1700065>.
- Whelan, T., Atz, U., Van Holt, T., & Clark, C. (2021). ESG and financial performance: Uncovering the Relationship by Aggregating Evidence from 1,000 Plus Studies Published between 2015–2020. *Center for Sustainable Business (NYU – Stern)*. Retrieved June 22, from <https://www.stern.nyu.edu/sites/default/files/assets/documents/ESG%20Paper%20Aug%202021.pdf>.
- Wickham, H. (2022). Tidyverse. R package version 1.3.2. Retrieved June 2022, from <https://tidyverse.tidyverse.org>, <https://github.com/tidyverse/tidyverse>.
- Xidonas, P., & Essner, E. (2024). On ESG portfolio construction: A multiobjective optimization approach. *Computational Economics*, 63, 21–45. <https://doi.org/10.1007/s10614-022-10327-6>.

## Authors and Affiliations

Alexandre Momparler<sup>1</sup>  · Pedro Carmona<sup>2</sup>  · Francisco Climent<sup>3</sup> 

---

✉ Alexandre Momparler  
alexandre.momparler@uv.es

Pedro Carmona  
pedro.carmona@uv.es

Francisco Climent  
f.jose.climent@uv.es

<sup>1</sup> Department of Corporate Finance, Universitat de València, València, Spain

<sup>2</sup> Department of Accounting, Universitat de València, València, Spain

<sup>3</sup> Department of Financial Economics, Universitat de València, València, Spain