

A Work Project, presented as part of the requirements for the Award of a Master's degree in  
Business Analytics from the Nova School of Business and Economics.

Applying Machine Learning Techniques to Identify Companies at Higher Risk of ESG  
Controversy

Sophie Schmitz

Work project carried out under the supervision of:

Qiwei Han

16-12-2021

## Abstract:

ESG controversies may have enormous consequences for an individual company, its customers, investors, and other stakeholders. The objective of this work is to identify companies at high risk of ESG controversy based on public ESG data. By using machine learning solutions, early indicators in ESG data can be identified that provide insight into how likely a company is to face an ESG controversy. By using Random Forest models, the proportion of companies with a controversy among the flagged companies can be increased by 93, 5.6 and 4.3 times for the Environmental, Social and Governance pillar, respectively.

## Keywords

ESG, ESG controversy, CSR, Sustainability, Data Science, Machine Learning, Random Forest

This work used infrastructure and resources funded by Fundação para a Ciência e a Tecnologia (UID/ECO/00124/2013, UID/ECO/00124/2019 and Social Sciences DataLab, Project 22209), POR Lisboa (LISBOA-01-0145-FEDER-007722 and Social Sciences DataLab, Project 22209) and POR Norte (Social Sciences DataLab, Project 22209).

# **1 Environmental, Social & Governance**

In recent decades, investors started considering non-financial data, such as a company's long-term sustainability and ethical practices, when evaluating a company to gain a fuller understanding of the investment (Chen et al. 2021). The ever-rising pressure from climate change and increased social awareness opens the discussion on the impact of Environmental, Social, and Governance (ESG) practices on a firm's financial performance (Allen and Craig 2016; Alsayegh, Abdul Rahman, and Homayoun 2020).

ESG is defined as “a generic term used by investors to evaluate corporate behavior and to determine the future financial performance of companies” (Financial Times Lexicon). More specifically, the Environmental factors evaluate the company's impact on the environment and the climate change, such as emission and waste it produces. Social factors measure a company's reputation and relationship with employees, customers, and local communities. Finally, Governance is the super category for a company's internal structure, controls, and procedures to govern itself and foster effective decision-making. ESG is often used interchangeably with the related term Corporate Social Responsibility (CSR) (Gillan, Koch, and Starks 2021). However, while CSR focuses on engaging stakeholders, ESG is a scoring system that investors use to objectively measure a firm's Social and Environmental impact to improve investment decisions (Cini and Ricci 2018).

## **1.1 The Emergence of ESG**

Although ESG is still a relatively new term, the world's leading newspapers report on ESG investing in their financial pages. In recent years, many companies have created positions such as ESG analysts (McKenzie et al. 2020). Today, more than 80% of the N100, the largest 100 companies of 52 countries, publish sustainability reports (McKenzie et al. 2020), which is the primer source for ESG ratings (Refinitiv). This remarkable growth in ESG-related disclosure

can be attributed to two main factors - a higher intrinsic motivation to voluntarily disclose this type of information and an extrinsic motivation as political and financial institutions publish guidelines and reporting requirements.

Worldwide, awareness of environmental and social crises such as global warming and modern slavery is increasing due to the international exchange of information and the growing use of social media (Dwivedi and Pandey 2013). As a result, investors, consumers, and communities prefer and even actively demand disclosure of companies' sustainability and ESG strategies (Branch, Goldberg, and Hand 2019). Millennials in particular are incorporating more information about companies' ESG practices into their investment and purchasing decisions (Ruggie and Middleton 2019). They prefer purposeful companies that care about the needs of their employees, consumers, and communities (Poole and Sullivan 2021).

Companies with a solid and transparent ESG proposition can differentiate and outperform companies that do not disclose their ESG practices (Friede, Busch, and Bassen 2015). One reason is its attraction of new customers and increased customer loyalty (Kim et al. 2021). Today, nearly two-thirds of consumers are willing to pay a premium for sustainable products, and this trend is expected to continue to grow (Ragbir et al. 2021). In addition, a strong ESG offering improves access to resources through more substantive relationships with the community and government, which ultimately increases revenue and reduces costs. An impactful business often increases employee motivation and helps attract new talent (de Neve and et al. 2018). According to a McKinsey study, this top-line growth and increased employee productivity are two of the key drivers of improved financial performance (Henisz, Koller, and Nuttal 2019).

Several studies have analyzed and confirmed the positive impact of a strong ESG approach on a company's financial performance (Velte 2017). For instance, a study by Bank of America Merrill Lynch revealed that the S&P 500 companies with ESG ratings in the top quantile

outperform those companies in the bottom quantile by more than 25 percentage points (Subramanian et al. 2019). A strong ESG proposition correlates with higher equity returns (Ionescu et al. 2019) and reduces the downside risk, for example, due to higher credit ratings (Henisz and McGlinch 2019). The Covid 19 pandemic contributed to ESG growth by showing that companies that thought about managing environmental or social risks responded better to the situation and were less affected by systemic risks (Adams and Abhayawansa 2021; Wellalage and Kumar 2020). Increased attention to the scientific literature in 2013 and 2014 has clearly accelerated the growth of voluntary reporting (Khan, Serafeim, and Yoon 2015).

In addition to the intrinsic motivation for companies to disclose more information on sustainability, the number of regulatory initiatives requiring non-financial disclosure is also growing rapidly, further enhancing corporate transparency and accountability on environmental and social issues. Regulations can be divided into two types: legally binding regulations and non-binding regulations that must be followed, for example, to be included in sustainable foundations. The first environmental laws, such as the Clean Water Act, were passed in the early 1970s and aimed to monitor and limit the environmental impact of businesses. (United States 1972). In an effort to express their commitment to sustainable business practices through climate stabilization, sustainable resource management and building an inclusive economy, several companies joined together to form an alliance called CERES (Laird 2002).

As the number of non-financial disclosures increases, so does the problem of comparability, as they report different key figures and are written in different styles. For this reason, the Global Reporting Initiative (GRI) was founded in 1997 to create a uniform global language for non-financial reporting (Hedberg and von Malmborg 2003). As of 2020, 73% of the world's 250 largest companies publish CSR/ ESG/ SDG/ sustainability reports according to the GRI Standards, a uniform way of reporting CSR / ESG issues (McKenzie et al. 2020).

Consensus among investors and political institutions about the growing importance of ESG factors and longer-term investments led to a conference in 2005 with 50 CEOs of major financial institutions, supported by the UN Global Compact, the International Finance Corporation (IFC), and the Swiss government (Kell 2018). The goal was to have an open discussion on the role of ESG indicators and create a mutual understanding. In the resulting report, “Who Cares Wins”, the term ESG was coined for the first time (Compact, UN Global 2004). Following, the United Nations launched the United Nations Principles for Responsible Investment (PRI) in 2006 to provide guidance on ESG for investors and companies. The resulting “Blueprint for Responsible Investment” has since been signed by more than 3,900 investors.

In response to increasing calls from academics and society for urgent action on environmental and social issues, CSR oversight has gained prominence on most policy agendas. In line with the EU's CSR agenda, the Non-Financial Reporting Directives (NFRD) have required large companies to publish regular reports on the social and environmental impacts of their activities starting in 2014 (La Torre et al. 2020). The Corporate Sustainability Reporting Directive (CSRD), proposed by the European Commission on April 21, 2021, will expand these responsibilities. When the CSRD comes into force, companies will not only have to report in greater detail, but will also must have their reports audited (European Commission 2021).

## **1.2 Machine Learning approach on ESG data**

Changing regulations and growing awareness for CSR caused reporting on non-financial factors to increase significantly in the last years and, thus, the availability of ESG scoring data (Tschopp and Huefner 2015). With the increasing amount of available data, traditional methods get outperformed on prediction tasks by advanced machine learning (ML) models in several studies (Amin et al. 2021; Singal et al. 2013).

ML is a branch of artificial intelligence and computer science that focuses on using large datasets and algorithms to improve themselves without being explicitly programmed to mimic the way humans learn and incrementally improve accuracy (IBM Cloud Education, 2020). Hence, ML and other data science techniques can be used to find underlying patterns and correlations in the increasing amount of worldwide ESG data that could not be identified before. As a result, new technology based on ML and big data unlocks valuable insights on ESG data to complement conventional financial information for investment decisions.

In the current literature, machine learning is mainly applied to analyzing ESG data regarding financial performance. In fact, most of the existing literature on machine learning indicates a direct link between ESG factors and the company's financial performance (Margot et al. 2021). Additionally, Margot et al. (2020) indicates that the link between ESG profiles and financial performances can only be accessed with non-linear machine learning techniques. Furthermore, Lanza et al. shows that ML techniques can be used to identify those indicators that better contribute to constructing efficient financial portfolios (Lanza, Bernardini, and Faiella 2020).

Nevertheless, ESG data is not only associated with the company's financial performance but also indicates that strong ESG propositions attract B2B and B2C customers with more sustainable products and boost employee motivation (Henisz, Koller, and Nuttal 2019).

Thus, this research aims to identify how machine learning models can be implemented to analyze factors such as ESG controversies, employer attractiveness, ESG score performance, and sustainability reports for new business applications.

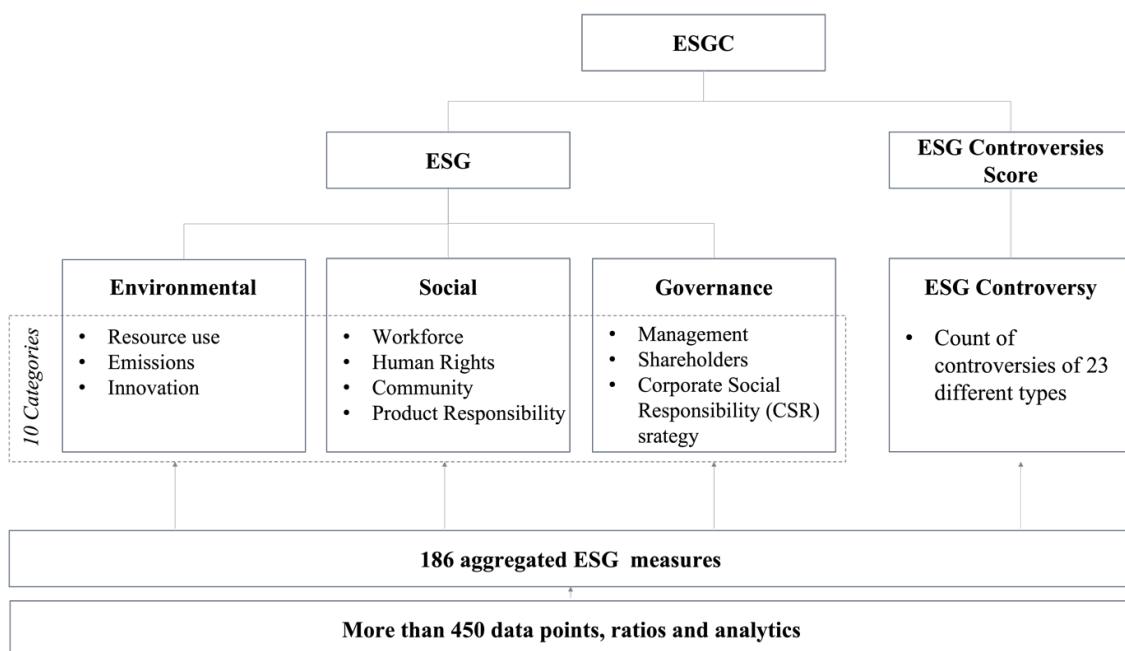
### **1.3 The construction of ESG scores**

While existing policies and guidelines facilitate the reporting of ESG indicators, further manual revision is undertaken for a transparent and standardized comparison. ESG rating agencies serve as independent third parties who make company-level information available and

comprehensive. The information gathered can stem from direct engagement with companies, such as through surveys, as in the case for the Dow Jones Sustainability Index and the Carbon Disclosure Project. A second approach is through the application of ML for automated data collection of publicly available information, such as company websites, reports, and news sites. Studies show that while scores can differ due to different methodologies used to create the ratings, ESG scores are a reliable and robust indicator of CSR performance and are commonly used in financial analyses (Billio et al. 2021; Del Giudice and Rigamonti 2020).

Within the scope of this study, the Refinitiv (former Thomson Reuters - Asset4) database is used, providing one of the largest ESG content collections. Refinitiv gathers data using a combination of both algorithmic and human processes from Company websites, company reports, NGO websites, media and news, and stock exchange filings. Each data entry is error-checked and then run through a consistency quality check to ensure the database's reliability. In total, Refinitiv includes over 450 different ESG metrics for over 10,000 companies, covering over 80% of the global market cap and the history of the observations goes back to 2002.

As shown in Figure 1, Refinitiv evaluates a company with the ESG combined score (ESGC).



**Figure 1: ESG Framework**

The ESGC score represents the discounted ESG score, which accounts for a companies' actions against commitments, based on 23 ESG controversy topics, which are picked up by global media, and includes a severity weight to adjust for a company's market cap.

The non-discounted ESG Score is broken up into the three respective pillars, each of which includes different main categories: Environmental (E) includes emissions, environmental innovation, and resource use; Social (S) includes community, human rights, product responsibility, and workforce; and Governance (G) includes CSR strategy, management, and shareholders. According to their significance for the specific industry group, different industry groups can have different pillar weights factoring the corresponding pillar. For governance, the pillar weights remain the same across all industries.

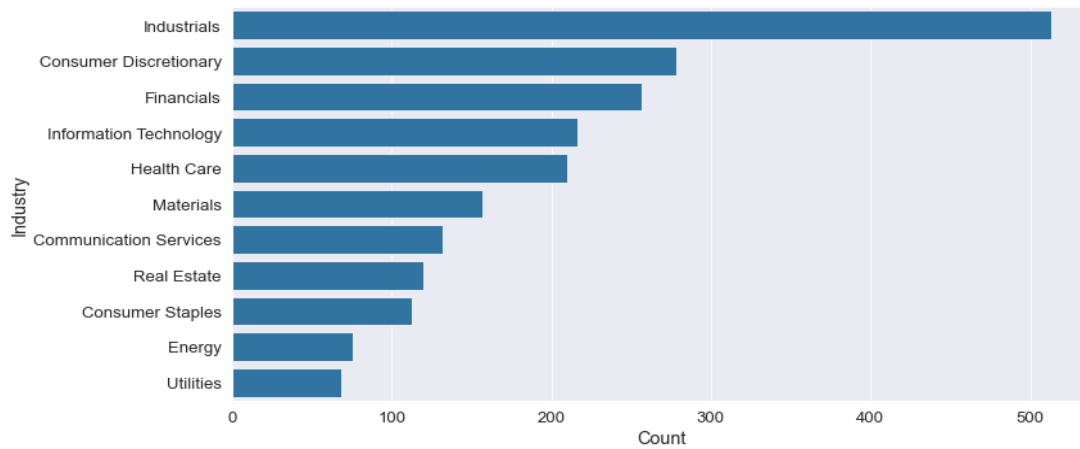
Like the pillar weights, the category scores have different weights depending on the industry group. The weights are used to better benchmark various industries, as some categories are more relevant and material to companies within the same industries. There are ten different categories spread across the three pillars. Each category is made up of multiple measures, which total to 186 included in the ESG score. In total, Environment contains 3, Social 4, and Governance 3 categories with 68, 62, and 56 measures, respectively.

Every Company's performance is ranked within the industry group for environmental and social category scores and the Controversies Score. Refinitiv benchmarks the Governance categories against the country of incorporation for the rank score. The final ESG and ESGC scores are given in both percentile rank scores and letter grades from A+ to D.

## **2 Data Exploration**

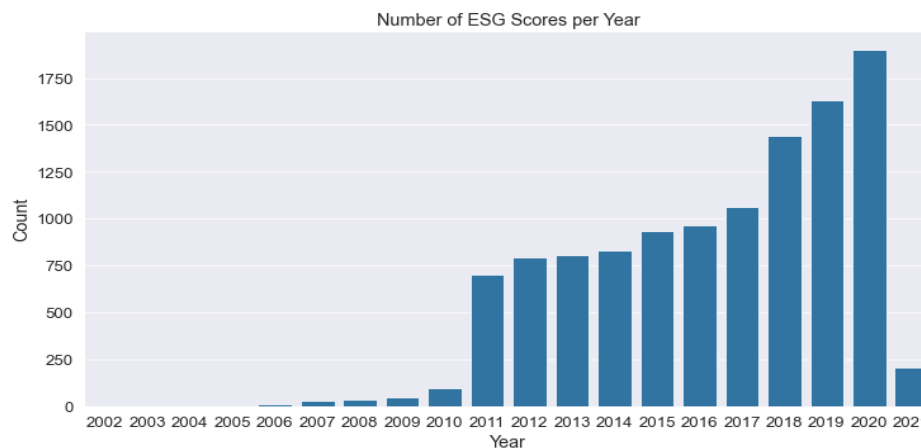
While the wide-ranging Refinitiv database covers over 9,000 global companies, we have focused our work particularly on European companies. In total, we have collected 11,484 ESG ratings for 2,140 European companies. These companies are spread over 11 industries. While

industrials and consumer discretionary are most prevalent industries in our data, the energy and utilities sectors are the least occurring, as shown in Figure 2.



**Figure 2.** Count of Companies per Industry

Looking at the number of companies over the entire period for which Refinitiv collects ESG data, a positive trend is clearly visible. In our dataset, there were less than 100 companies per year from 2002 until 2010, but the number has almost tripled from 2011 (701 companies) to 2020 (1,901 companies). In Figure 3 we can see that the data again underscores the rise and development of ESG, especially in the last decade.



**Figure 3.** Count of ESG Scores per Year

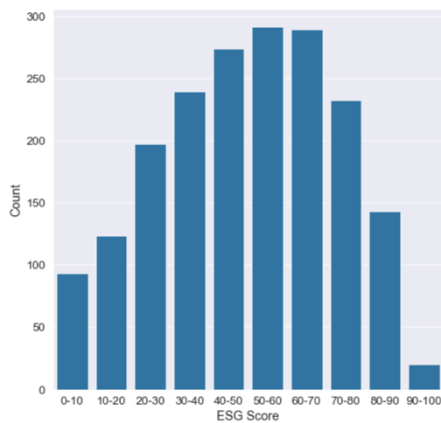
The summary statistics of the separate ESG Scores, presented in Table 1, indicate that both the ESG Score and the ESGC score are similar distributed around 50.

**Table 1.** Summary statistics of ESG scores and controversy scores over the complete data set

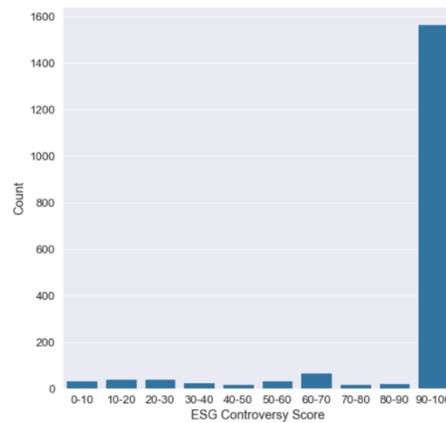
	ESG Score	ESGC Score	Controversy Score	Environmental Pillar Score	Governance Pillar Score	Social Pillar Score
mean	51.58	49.42	90.67	48.15	50.03	55.14
std	20.76	19.50	23.11	28.09	23.15	23.29
min	0.10	0.10	0.51	0.00	0.38	0.12
median	52.69	64.23	100.00	72.75	68.85	74.23
max	94.52	93.66	100.00	99.16	98.59	98.63

The summary statistics also shows that European companies have a slightly higher median for environmental and social scores than for governance scores. In addition, Figure 4 highlights the bell shape of the ESG score distribution that typically occurs with a normal distribution. Most companies have ESG Scores around 50, and 90% have a score between 18 and 86.

While, the ESG score distribution is slightly skewed to the left, the Controversy scores, plotted in Figure 2, are highly skewed to the left.



**Figure 4.** Distribution of ESG Scores for 2020



**Figure 5.** Distribution of Controversy Scores for 2020

The figure shows that most companies do not face a controversy and therefore have a score of 100. Since the consequences of an ESG controversy can nevertheless be very serious for the companies and several stakeholders, the following section focuses on ESG controversies.

### 3 ESG controversy

While the idea of ESG scores is to make sustainability more transparent, accurate, and comparable (Refinitiv 2021), there is an increasing number of companies being accused of having a gap between their external communication and actual ESG practices (Walker and Wan 2012). According to Kanter, for many companies, the motivation of ESG activities is a profit-oriented marketing campaign (Kanter 2009). Many customers, investors, and academics distrust the veracity of claims about the corporate social responsibility benefits of many companies (Mattis 2008).

One reason for that is the rising number of ESG scandals reported in the news. According to the Refinitiv database, the number has almost quadrupled from 2009 and 2019<sup>1</sup>. Some of these scandals can result in tremendous fines. For instance, Google alone is currently facing €8.2 billion in EU antitrust fines (Euronews 2021). Further, ESG related scandals can result in a significant drop in share price. The Bank of America revealed that 24 ESG controversies of Fortune 500 companies resulted together in peak-to-trough market capitalization losses of \$534 billion (Subramanian et al. 2019). Consequently, investors are less likely to include companies at high risk of facing ESG-controversy in sustainable funds reinforcing the falling share price (Flood 2019). In the case of Volkswagen, the loss in customer trust resulted in a decline in revenue of \$3.7 billion (Bachmann, Ehrlich, and Ruzic 2017). Notably, an ESG controversy does not necessarily have to be directly the fault of the company as it can also occur in one of their supplier companies and still have a significant negative impact (Castillo et al. 2018).

Due to these serious consequences for multiple stakeholders, it would be of great interest to know in advance which companies may have an ESG-related scandal. By flagging companies at higher risk of ESG controversies, audit companies can screen those companies in more depth

---

<sup>1</sup> See Figure 10 in Appendix 1

before they face a public ESG-related scandal. Further, rating agencies can create a new risk rating independently of news and social media. Finally, companies can use it as a risk mitigation monitoring method to prevent financial and non-financial impacts. While numerous papers use machine learning (ML) algorithms to detect financial regulatory violations, it is not yet clear whether they are applicable to CSR issues (Kotzian 2021).

Hence, the question to be answered with this paper is: Can companies that are more likely to face a controversy be flagged using ML Learning methods, and if so, how?

### **3.1 Leveraging Machine Learning for ESG controversy prediction**

As data exploration has shown, manual interpretation of the data would be extremely complex due to its volume, variability, and intercorrelation of multiple variables. Several papers show that ML has already been successfully applied in various fields and has led to more evidence-based decision making (Jordan and Mitchell 2015). Recently, ML has also gained attention in the field of sustainability (Lee and Shin 2020). Its potential is underscored by a report from the World Economic Forum, which found that about 70% of goals to address environmental and social problems can be achieved through the use of technology (Herweijer et al. 2020).

The goal of this work is to identify companies with a higher probability of ESG controversy based on their reported CSR practices. Using a supervised machine learning algorithm, we can determine a function that maps the input (ESG input data) to the output (ESG controversy). Supervised algorithms require marked input training data. (Han, Pei, and Kamber 2011). More specifically, as we want to either flag or not flag a company, hence, the output variable is a Boolean variable, measured as 0 or 1 (Han, Pei, and Kamber 2011). Therefore, it is necessary to have a clear definition of the label, ESG controversy.

### **3.2 Definition of ESG controversy**

ESG controversy refers to ongoing activities, events, or products of a company that harms the environment, society, or corporate governance. This includes the violation of laws and practices or events that violate accepted international norm (MSCI ESG Research 2020). Controversies can be the result of individual bad actors or unfortunate events and structural problems of the company, such as poor governance (MSCI ESG Research 2020).

A related concept is green-washing, known as unethical behavior regarding environmental practices (Furlow 2010). However, some definitions not only cover environmental but also social practices (Mattis 2008). Lately, new terms like blue-washing and pink-washing were derived, which stands for cover-ups in the social and LGBT domain, respectively (Smith and Keating 2017; Berliner and Prakash 2015).

In many cases, companies communicate a sustainable performance based on a few selected attributes without attention to other important CSR issues (TerraChoice 2010) or (un-)purposely falsely state social or environmental claims (Lynes 2015). These types of green / ESG laundering practices are referred to as the "sin of hidden compromise" and the "sin of fibbing" (Freitas Netto et al. 2020). Often, scandals become public years after the unethical practices began. For example, in the case of Volkswagen, they manipulated emission levels for years before they were caught (Schiermeier 2015).

In this paper, we only consider ESG controversies that became a public scandal. We use the ESG controversies reported by the Refinitiv database, which automatically captures controversies from media sources on various ESG controversy topics (Refinitiv 2021)

## 4 Methodology

To develop a tool that flags a company at higher risk, the pipeline presented in Figure 6 is executed. This methodology was applied three times for each pillar - Environmental, Social, and Governance.

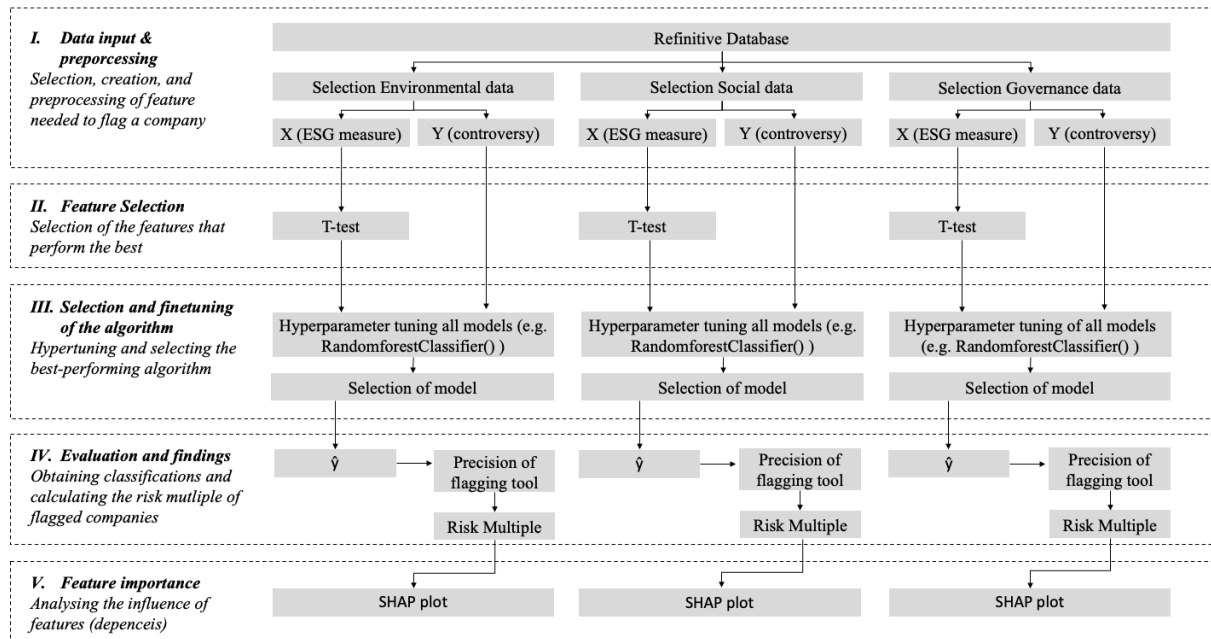


Figure 4. Pipeline Flagging Tool

### 4.1 Data Input and Preprocessing

#### 4.1.1 Label

The target variable is a Boolean variable and indicates whether a company has a controversy in the respective year. It is calculated based on 23 different controversy types from the Refinitiv database, which describe different types of controversy<sup>2</sup>. From these variables, we created three Boolean variables for each of the categories - Environment, Social, and

<sup>2</sup> See overview of all controversy types in [Appendix 2](#)

Governance. We chose to create a Boolean variable instead of a continuous variable indicating the number of controversies in a corporate year for two reasons. First, many controversies are highly correlated and are not necessarily new scandals, but rather consequences of the first scandal. Second, the goal of this work is to flag companies at higher risk for ESG controversy, and therefore no distinction is made between type, impact, and number of controversies.

#### **4.1.2 Features**

The dataset contains six fixed variables that describe the instance: Index, Company Name, Country of Headquarters, Sector, Industry, and Date; these are not included in the classification algorithm. For most companies, scores are available for several years. However, we treated them as independent instances because the number of controversies a company has faced does not give a meaningful prediction of how likely it is to face another controversy in the future. (Schroders 2017). In this paper, a company in a specific year will be referred to as “Company Year”.

The remaining dynamic variables consisted of Boolean variables, categorical variables, and continuous variables. The Boolean variables indicate whether a company has certain standards and policies. Due to the high correlation of the Boolean variables, we replaced them with two new variables, “Number of policy and standards” and “Number of unethical practices”, which represents the number of “True” values for the boolean variables with a positive and those with a negative polarity respectively. Afterwards, we have 73, 65, and 32 input features for Environment, Social, and Governance pillars, respectively. Finally, we normalized all features and replaced the missing data with zeros. In total, the used dataset consisted of 11,484 Company years from 2009 to 2021.

We created a second dataset that captured the percentage change from year t-1 to year t for each company to test whether the evolution of characteristics affected controversies rather than absolute numbers. However, these datasets did not contain any further information on this.

## 4.2 Feature Selection

Feature selection describes the process of selecting unique variables or attributes of the data when building models for machine learning and data science to eliminate irrelevant or less important features (Russell, Norvig, and Davis 2010). Thereby, the computational cost of modeling is reduced, and the quality of the model can be improved (Sarker 2021). While some models such as Support Vector Machine naturally perform feature selection, pre-selecting the most influential features enhances the comparability of different models (Soto et al. 2009). We used an independent t-test for Company Years with controversies and Company Years without controversy to test whether the features are informative. If the p-value for the feature is below 0.05, we can reject the Null-Hypothesis that the distribution is the same (Fadem 2008). 59, 50, 19 variables, for E, S, and G, respectively, had a p-value below 0.05, and therefore were used as input features for our model.

## 4.3 Selection and Finetuning of the Algorithms

Before training the model, we split the dataset randomly into training and testing data in a rate 70:30, as shown in Table 2.

**Table 2.** Train-test split

	Train set		Test set	
	Size	Ratio <sup>1</sup>	Size	Ratio
Environment	8,038	0.7%	3,446	0.6 %
Social	8,038	15.5%	3,446	15.6%
Governance	8,038	2.7%	3,446	2.8 %

<sup>1</sup>Proportion of Company Years that face a controversy.

The training data was used to build and fit the models, while the test data was used to evaluate the models (Vrgazova 2021). Looking at the ratio of company years with a controversy to those without, the data set is extremely unbalanced. Imbalance is a common problem in ML and can negatively affect the prediction probability of the model (Yang and Wu 2006). Therefore, it is necessary to consider it when training the algorithm.

#### 4.3.1 Important Metrics

Not flagging a company facing a controversy (Type I error) can have severe consequences for the company, investors, and other stakeholders, as explained above. The Type I error is minimized when the Recall Score (RS) is maximized, which is calculated as follows:

**Recall** =  $\frac{TP}{TP+FP}$  (TP = True Positive, FP=False Positive). On the other hand, flagging a

company without controversy (Type II error) leads to high investigation costs for validating ESG performance metrics for companies that actually comply with their ESG performance.

Type II error is minimized when the Precision Score (PS) is maximized and is calculated as

follows: **Precision** =  $\frac{TP}{TP+FN}$  (FN=False Negative). In the case of our flagging tool, the

consequences of a high Type I error are worse than the consequences of a high Type II error.

Nevertheless, in minimizing the type I error, we must take into account the budget and the

capacity of the audit firm. Thus, the goal of this model is to reduce the Type I error while

minimizing the Type II error.

To compare the performance of the different models we applied the receiver operating

characteristic (ROC) curve, from which we calculated the area under the ROC curve (AUC)

(Rosset 2004). The ROC curve, plots at different thresholds the true positive rate (TPR) against

the false positive rate (FPR). An AUC value of 0.5 means a random estimate while an AUC

value of 1 would be a perfect prediction. Therefore, the closer the AUC score is to 1, the higher

the predictive power (Barboza, Kimura, and Altman 2017).

### 4.3.2 Selection of the best performing algorithm

There are a variety of ML algorithms, and it is critical to find the algorithm that best suits the target (Lee & Shin 2020). To find the best model, we compared the performance of Logistic Regression, Random Forest Classifier, Gradient Boosting Classifier, AdaBoost Classifier, and Support Vector Machine. For machine learning algorithms to deliver optimal results, their hyperparameters must be set correctly (Syarif, Prugel-Bennett, and Wills 2016). Hence, we performed hyperparameter tuning for all the selected models using a 5-fold cross-validation grid search (Bergstra and Bengio 2012; Nievergelt 2000). We selected the optimal set of hyperparameters based on the best average value of the cross-validation estimator. Using the cross-validation technique can reduce the risk of overfitting (Lin et al. 2008). Since the data set is unbalanced, misclassification in the minority class must be penalized more than in the majority class. This can be achieved by balancing the weights of the two classes in the training phase of each model (Sinha, Purkayastha, and Gichoya 2019).

Next, we calculated the RS and the AUC curve for the optimized models, the results can be found in the following table.

**Table 3.** Algorithms performance

	Environment		Social		Governance	
	Recall	AUC	Recall	AUC	Recall	AUC
Log Regression	0.15	0.91	0.29	0.80	0.00	0.67
<b>Random Forest</b>	<b>0.80</b>	<b>0.95</b>	<b>0.65</b>	<b>0.80</b>	<b>0.69</b>	<b>0.77</b>
Gradient Boosting	0.45	0.73	0.39	0.68	0.09	0.56
AdaBoost	0.60	0.90	0.38	0.79	0.79	0.53
Support Vector Machine	0.15	0.95	0.67	0.79	0.70	0.63

**Note:** All scores are measured on the test set; The highlighted model performed the best

Random Forest Classifier (RF) performed best for all three categories when both metrics are considered. The RF is one of the most popular and powerful techniques in pattern recognition in ML (Shrivastava et al. 2020). It applies bootstrap aggregating to decision trees (Breiman, 1996). This means multiple decision trees are developed in parallel and make various

predictions. The class with the most votes becomes the model’s prediction (Liaw and Wiener 2002). Further, RF selects a random subset of features at each candidate split, a method known as random subspace projection (Tin Kam Ho 1998). Thereby, overfitting on features that are very strong predictors for the target class is minimized (Breiman 1999).

To have a useful labeling system, we need to capture a substantial number of companies facing ESG controversies. Therefore, we adjusted the classification threshold, which indicates the probability that an instance is assigned to one class or another (Fan and Lin 2007). Thereby, we can ensure to flag a sufficient proportion of companies with controversies.

## 5 Findings and implications

Once we selected the best model per domain, we measured its performance on the test set, as presented in Table 4.

**Table 4.** Model Performance

	Recall Score	Precision Score		Risk multiplier
		Flagged	Not Flagged	
Environment	65%	19%	0.2 %	93x
Social	66%	40%	7.1%	5.6x
Governance	67%	6%	1.4%	4.3 x

The model is evaluated with the RS and the PS as well as the risk multiplier. Since we optimized the threshold for the training set, the RS of the test set is not exactly 70%. When evaluating the results, it is important to keep in mind that the number of controversies in the three areas is very different. Therefore, the PS of the marked companies cannot be compared across the three columns. A better indicator is how much higher the proportion of companies with a controversy is among the flagged Company Years compared to the non-flagged Company Years, here given as a risk multiplier. In the environmental category, a flagged company is 93 times more likely to face a controversy a non-flagged company.

To understand which features determine the output, we used the SHAP (SHapley Additive exPlanations) algorithm from Lundberg & Lee (Lundberg and Lee 2017). SHAP uses a cooperative game theory approach to explain the output of complex models. It calculates the mean SHAP values to determine the features' marginal contribution.

## 5.1 Environment

The model for the Environment category has a RS of 65 %, the exact distribution is shown in the Table 5.

**Table 5.** Flagging Tool Performance - Environmental

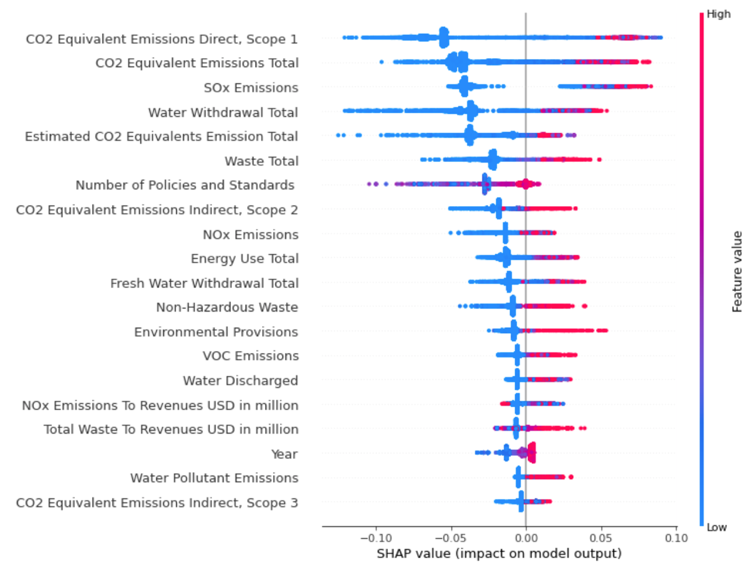
	Not Flagged		Flagged		Total	Share captured correctly
	Absolut	Share of Class <sup>1</sup>	Absolut	Share of Class		
No ESG-related scandal	<b>3,369</b>	99,8%	57	81,4 %	3,426	98,3%
ESG related scandal	7	0,2%	<b>13</b>	19,6%	20	65%
Total	3,376		70		3,446	

**Note:** numbers in bold are correct predictions

<sup>1</sup>Class means whether a company is flagged or not

From the 3,446 instances in the test set, the algorithm flagged 70 (2%). This includes 13 from 20 (65%) companies facing an ESG controversy. Among the flagged Company Years, the ratio of companies with vs without ESG controversy is 19,6 %, meaning that on average every fifth flagged company had a controversy. In comparison to the un-flagged companies, this ratio is 93 times higher.

When looking at the Summary SHAP plot (Figure 7), we can see that there is a positive relationship between ESG-controversies and the variables that measure emissions<sup>3</sup>, resources consumed<sup>4</sup> and waste produced<sup>5</sup>.



**Figure 5.** SHAP Summary Plot - Environmental

On the other side, the recycling ratio<sup>6</sup> negatively correlates with ESG controversies. Notably, the three features with the largest impact are emissions of the company. Practically, that means that Company Years with higher emissions scores are more likely to have an ESG controversy.

## 5.2 Social

In the Social category, the model flagged 349 from the 528 companies who suffered from a controversy representing a RS of 66%, as shown in Table 6. The model flagged 881 Company Years, which is a PS of 40%. In contrast, the probability of a company facing an ESG controversy among the unflagged Company Years is almost times lower.

<sup>3</sup> Measured in “CO2 Equivalent Emissions Total”, “CO2 Equivalent Emissions Direct Scope 1”, “Estimated CO2 Equivalents Emission”, “Total VOC Emissions”, “SOX Emissions”, “CO2 Equivalent Emissions Scope 3”

<sup>4</sup> Measured in: “Water Withdrawal Total”, “Fresh Water Withdrawal Total”, and “Energy Use Total”

<sup>5</sup> Measured in: “Hazardous Waste” and “Non-Hazardous Waste”

<sup>6</sup> Feature name : “Waste Recycled To Total Waste”

**Table 6.** Flagging Tool Performance - Social

	Not Flagged		Flagged		Total	Share captured correctly
	Absolut	Share of Class <sup>1</sup>	Absolut	Share of Class		
No ESG-related scandal	<b>2,382</b>	93%	532	60 %	2,914	82%
ESG related scandal	179	7%	<b>349</b>	40%	528	65%
Total	2,561		881		3,446	

In Figure 8, we can see a positive correlation with features that depend on the firm size<sup>7</sup> and the probability of being flagged.



**Figure 6:** SHAP Summary Plot - Social

While this suggests a positive correlation between company size and ESG controversies, it must be taken into account that larger companies have a greater news presence and scandals are more likely to be uncovered (Drempetic, Klein, and Zwergel 2020). When analyzing the marginal effect that characteristics other than company size have<sup>8</sup>, there is no clear trend for "Total training costs" or "Announced layoffs." In fact, there is a negative relationship between employee salaries and the probability of being labeled<sup>9</sup>.

<sup>7</sup> Measured in "Number of Employees from CSR reporting", "Salaries and Wages from CSR reporting", "Training Costs Total"

<sup>8</sup> Measured in "Number of Employees from CSR reporting"

<sup>9</sup> See Figure 11 and Figure 12 in [Appendix 3](#)

### 5.3 Governance

In the Governance category the prediction led to a RS of 67% (91 of 135 Company Years) and a PS of 6% (64 of 1076 Company Years), as represented in Table 7.

**Table 7.** Performance Flagging Tool - Governance

	Not Flagged		Flagged		Total	Share captured correctly
	Absolut	Share of Class <sup>1</sup>	Absolut	Share of Class		
No ESG-related scandal	<b>2,273</b>	98.6%	1,076	81.4 %	3,349	67,9%
ESG related scandal	32	1.4%	<b>64</b>	5.6%	96	66,7%
Total	2,305		1,140		3,446	

To capture two-third of Company Years with controversies, the model flagged 1,140 (33%) of all Company Years, resulting in a large Type II Error. However, the proportion of Company Years with a controversy is four times higher among the flagged vs. not-flagged Company Years.

As shown in Figure 9, the most impact on a company's classification has “Board Member Affiliation”, “Compensation Committee Independence” and the relative senior compensation “Total Senior Executives Compensation To Revenues”.



**Figure 7.** SHAP Summary Plot - Governance

More board member affiliations lead to a higher likelihood of being flagged; it could mean that the board is less focused and thus hears about fewer internal controversies. Accordingly, a lower average meeting attendance<sup>10</sup> and a lower “Board Member Turnover” also increase the probability of being flagged. Counterintuitively, when interpreting the board members’ independence<sup>11</sup> individually it reveals a positive correlation with ESG controversy, however, when considering the board size, there is no meaningful interpretation<sup>12</sup>.

## **6 Contribution and use cases of the flagging tool**

The introduced flagging system that can be used in at three different ways:

1. Audit firms can make a pre-selection and examine the flagged companies in more detail. This saves audit firms costs by reducing the number of companies they need to examine thoroughly. In addition, close monitoring of flagged companies reduces the risk of overlooking a controversial company, leading to reputational damage and legal consequences for the auditing companies.
2. Companies can use it as a risk screening tool to monitor different departments and partner companies along the supply chain. As mentioned, this can protect the company from reputational damage, financial losses, and legal consequences.
3. Investors/ fund managers can exclude flagged companies from funds to reduce the risk of a stock decline after a public scandal.

---

<sup>10</sup> Feature name: Committee Meetings Attendance Average

<sup>11</sup> Measured in: “Compensation Committee Independence”, “Nomination Committee Independence”

<sup>12</sup> See Figure 13 in [Appendix 4](#)

Overall, this paper shows that it is possible to identify companies with higher risk in all three areas. For the environmental pillar in particular, the RF model found a strong relationship between input data from ESG reports and an ESG controversy. However, this tool can only serve as a warning or recommendation as to where more investigative resources should be applied. A labeled company cannot be accused of being involved in illegal or unethical affairs.

## **7 Limitations**

The proposed model has three main limitations. First, the main limitation of the analysis is that the dependent variable most likely understates the actual number of violations. After a scandal occurs in one industry, similar scandals are uncovered in other industries as other companies are investigated more closely. From this, we can conclude that the number of unreported cases is larger than the actual number of controversies, meaning that we train and evaluate our model with an incorrect label.

Second, ESG is still a young concept, without clear reporting guidelines (Johnson 2021), resulting in many missing values. Therefore, it is recommended to conduct further research as the industry matures and data quality improves.

Third, the data set is biased with respect to company size. Large companies tend to report more, which reduces the amount of missing data. Also, as mentioned earlier, controversies are more likely to be uncovered at large companies, implying that their practices are less ethical than those of smaller companies. Since the new laws in force in the EU from 2022 also require smaller companies to provide ESG reporting, it is also important here is recommended to conduct further research once the data density has increased for smaller companies.

## 8 Bibliography

- Adams, Carol A., and Subhash Abhayawansa. 2021. "Connecting the COVID-19 Pandemic, Environmental, Social and Governance (ESG) Investing and Calls for 'Harmonisation' of Sustainability Reporting." *Critical Perspectives on Accounting*, March, 102309. <https://doi.org/10.1016/j.cpa.2021.102309>.
- Allen, Myria W., and Christopher A. Craig. 2016. "Rethinking Corporate Social Responsibility in the Age of Climate Change: A Communication Perspective." *International Journal of Corporate Social Responsibility* 1 (1): 1. <https://doi.org/10.1186/s40991-016-0002-8>.
- Alsayegh, Maha Faisal, Rashidah Abdul Rahman, and Saeid Homayoun. 2020. "Corporate Economic, Environmental, and Social Sustainability Performance Transformation through ESG Disclosure." *Sustainability* 12 (9): 3910. <https://doi.org/10.3390/su12093910>.
- Amin, Muhammad Nasir, Ammar Iqtidar, Kaffayatullah Khan, Muhammad Faisal Javed, Faisal I. Shalabi, and Muhammad Ghulam Qadir. 2021. "Comparison of Machine Learning Approaches with Traditional Methods for Predicting the Compressive Strength of Rice Husk Ash Concrete." *Crystals* 11 (7): 779. <https://doi.org/10.3390/cryst11070779>.
- Bachmann, Rüdiger, Gabriel Ehrlich, and Dimitrije Ruzic. 2017. "Firms and Collective Reputation: The Volkswagen Emission Scandal as a Case Study," 39.
- Barboza, Flavio, Herbert Kimura, and Edward Altman. 2017. "Machine Learning Models and Bankruptcy Prediction." *Expert Systems with Applications* 83 (October): 405–17. <https://doi.org/10.1016/j.eswa.2017.04.006>.
- Berliner, Daniel, and Aseem Prakash. 2015. "'Bluewashing' the Firm? Voluntary Regulations, Program Design, and Member Compliance with the United Nations Global Compact: 'Bluewashing' the Firm?" *Policy Studies Journal* 43 (1): 115–38. <https://doi.org/10.1111/psj.12085>.
- Billio, Monica, Michele Costola, Iva Hristova, Carmelo Latino, and Lorian Pelizzon. 2021. "Inside the ESG Ratings: (Dis)Agreement and Performance." *Corporate Social Responsibility and Environmental Management* 28 (5): 1426–45. <https://doi.org/10.1002/csr.2177>.
- Branch, Michael, Lisa R. Goldberg, and Pete Hand. 2019. "A Guide to ESG Portfolio Construction." *The Journal of Portfolio Management* 45 (4): 61–66. <https://doi.org/10.3905/jpm.2019.45.4.061>.
- Breiman, Leo. 1999. "RANDOM FORESTS--RANDOM FEATURES," 29.
- Castillo, Vincent E., Diane A. Mollenkopf, John E. Bell, and Hamparsum Bozdogan. 2018. "Supply Chain Integrity: A Key to Sustainable Supply Chain Management." *Journal of Business Logistics* 39 (1): 38–56. <https://doi.org/10.1111/jbl.12176>.
- Chen, Li, Lipei Zhang, Jun Huang, Helu Xiao, and Zhongbao Zhou. 2021. "Social Responsibility Portfolio Optimization Incorporating ESG Criteria." *Journal of Management Science and Engineering* 6 (1): 75–85. <https://doi.org/10.1016/j.jmse.2021.02.005>.

- Cini, Andrea Cincinnati, and Chiara Ricci. 2018. "CSR as a Driver Where ESG Performance Will Ultimately Matter." *Symphonya. Emerging Issues in Management*, no. 1 (July): 68. <https://doi.org/10.4468/2018.1.05cini.ricci>.
- Compact, UN Global. 2004. *Who Cares Wins: Connecting Financial Markets to a Changing World*. New York: Compact, UN Global.
- Del Giudice, Alfonso, and Silvia Rigamonti. 2020. "Does Audit Improve the Quality of ESG Scores? Evidence from Corporate Misconduct." *Sustainability* 12 (14): 5670. <https://doi.org/10.3390/su12145670>.
- Drempetic, Samuel, Christian Klein, and Bernhard Zwergel. 2020. "The Influence of Firm Size on the ESG Score: Corporate Sustainability Ratings Under Review." *Journal of Business Ethics* 167 (2): 333–60. <https://doi.org/10.1007/s10551-019-04164-1>.
- Dwivedi, Pradeep Kumar, and Ingita Pandey. 2013. "ROLE OF MEDIA IN SOCIAL AWARENESS." *Social Sciences* 1: 4.
- Euronews. 2021. "Google and EU Head to Court to Decide the Fate of €4.3 Billion Fine." *Euronews*, September 2021.
- European Commission. 2021. "Proposal for a Directive of the European Parliament and the Council."
- Fadem, Barbara. 2008. *High-Yield Behavioral Science*. Lippincott Williams & Wilkins.
- Fan, Rong-En, and Chih-Jen Lin. 2007. "A Study on Threshold Selection for Multi-Label Classification," 23.
- Flood, Chris. 2019. "ESG Controversies Wipe \$500bn off Value of US Companies." *Financial Times*, December 2019.
- Freitas Netto, Sebastiao, Marcos Sobral, Ana Re Be Ribeiro, and Gleibson Soares. 2020. "Concepts and Forms of Greenwashing: A Systematic Review." *Environmental Sciences Europe* 32 (19).
- Friede, Gunnar, Timo Busch, and Alexander Bassen. 2015. "ESG and Financial Performance: Aggregated Evidence from More than 2000 Empirical Studies." *Journal of Sustainable Finance & Investment* 5 (4): 210–33. <https://doi.org/10.1080/20430795.2015.1118917>.
- Furlow, Nancy E. 2010. "Greenwashing in the New Millennium." *Journal of Applied Business and Economics*, 4.
- Gillan, Stuart L., Andrew Koch, and Laura T. Starks. 2021. "Firms and Social Responsibility: A Review of ESG and CSR Research in Corporate Finance." *Journal of Corporate Finance* 66 (February): 101889. <https://doi.org/10.1016/j.jcorpfin.2021.101889>.
- Han, Jiawei, Jian Pei, and Micheline Kamber. 2011. *Data Mining: Concepts and Techniques*. Elsevier.
- Hedberg, Carl-Johan, and Fredrik von Malmborg. 2003. "The Global Reporting Initiative and Corporate Sustainability Reporting in Swedish Companies." *Corporate Social Responsibility and Environmental Management* 10 (3): 153–64. <https://doi.org/10.1002/csr.38>.
- Henisz, Witold J., Tim Koller, and Robin Nuttal. 2019. "Five Ways That ESG Creates Value." *McKinsey Quarterly*, November.
- Henisz, Witold J., and James McGlinch. 2019. "ESG, Material Credit Events, and Credit Risk." *Journal of Applied Corporate Finance* 31 (2): 105–17. <https://doi.org/10.1111/jacf.12352>.

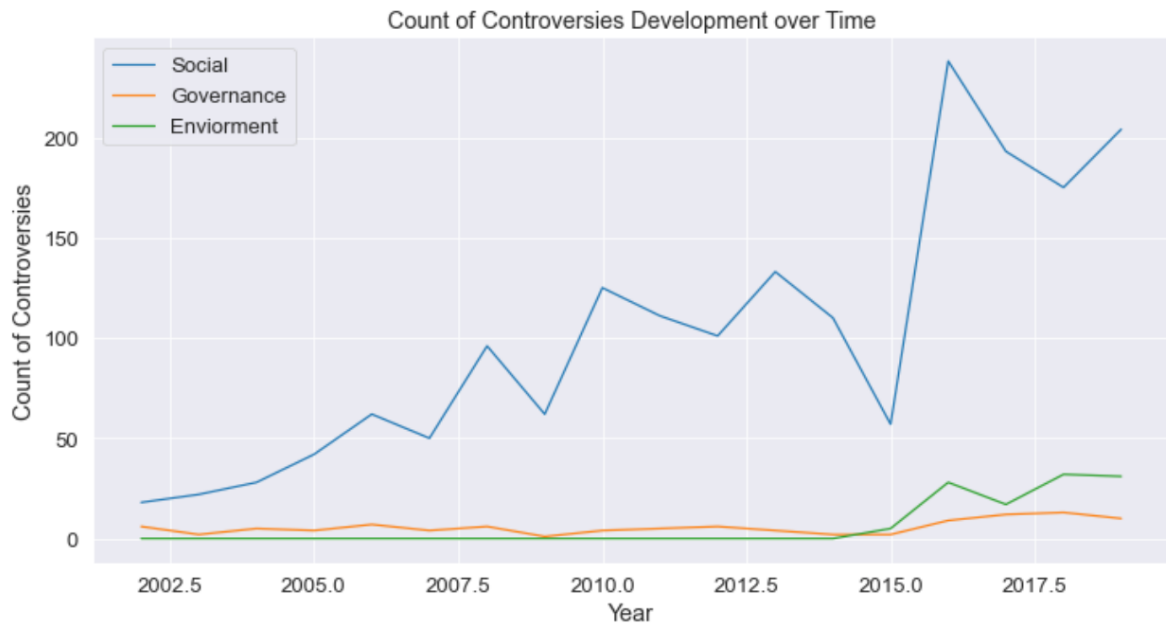
- Herweijer, C, B Combes, A Gawel, and A.E. Larsen. 2020. "Unlocking Technology for the Global Goals." *World Economic Forum, PwC*.
- Ionescu, George H., Daniela Firoiu, Ramona Pirvu, and Ruxandra Dana Vilag. 2019. "THE IMPACT OF ESG FACTORS ON MARKET VALUE OF COMPANIES FROM TRAVEL AND TOURISM INDUSTRY." *Technological and Economic Development of Economy* 25 (5): 820–49. <https://doi.org/10.3846/tede.2019.10294>.
- Johnson, Daniel. 2021. "The Rise of ESG and Why It Matters to Everyone." *SS&C*, 2021. <https://www.ssctech.com/blog/the-rise-of-esg-and-why-it-matters-to-everyone>.
- Jordan, M. I., and T. M. Mitchell. 2015. "Machine Learning: Trends, Perspectives, and Prospects." *Science* 349 (6245): 255–60. <https://doi.org/10.1126/science.aaa8415>.
- Kanter, James. 2009. "Study: For Consumers, Green Is Greenwash." *The New York Times*, April 2009.
- Kell, Georg. 2018. "The Remarkable Rise of ESG." *Forbes*, 2018. <https://www.forbes.com/sites/georgkell/2018/07/11/the-remarkable-rise-of-esg/?sh=25e186f41695>.
- Khan, Mozaffar, George Serafeim, and Aaron Yoon. 2015. "Corporate Sustainability: First Evidence on Materiality." *Harvard Business School Working Paper* 15–073: 37.
- Kim, Jinsung, Minseok Kim, Sehyeuk Im, and Donghyun Choi. 2021. "Competitiveness of E Commerce Firms through ESG Logistics." *Sustainability* 13 (20): 11548. <https://doi.org/10.3390/su132011548>.
- Kotzian, Peter. 2021. "Applying Machine Learning and Artificial Intelligence to CSR-Compliance. A Conceptual Framework with Illustrations." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3788977>.
- La Torre, Matteo, Svetlana Sabelfeld, Marita Blomkvist, and John Dumay. 2020. "Rebuilding Trust: Sustainability and Non-Financial Reporting and the European Union Regulation." *Meditari Accountancy Research* 28 (5).
- Laird, Sarah A. 2002. "Coalition for Environmentally Responsible Economies (CERES)." In *Biodiversity and Traditional Knowledge: Equitable Partnerships in Practice*, 1st ed. <https://www.taylorfrancis.com/books/e/9781849776080>.
- Lanza, Ariel, Enrico Bernardini, and Ivan Faiella. 2020. "Mind the Gap! Machine Learning, ESG Metrics and Sustainable Investment." *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3659584>.
- Lee, In, and Yong Jae Shin. 2020. "Machine Learning for Enterprises: Applications, Algorithm Selection, and Challenges." *Business Horizons* 63 (2): 157–70. <https://doi.org/10.1016/j.bushor.2019.10.005>.
- Liaw, Andy, and Matthew Wiener. 2002. "Classification and Regression by RandomForest" 2: 5.
- Lin, Shih-Wei, Kuo-Ching Ying, Shih-Chieh Chen, and Zne-Jung Lee. 2008. "Particle Swarm Optimization for Parameter Determination and Feature Selection of Support Vector Machines." *Expert Systems with Applications* 35 (4): 1817–24. <https://doi.org/10.1016/j.eswa.2007.08.088>.
- Lundberg, Scott M, and Su-In Lee. 2017. "A Unified Approach to Interpreting Model Predictions," 10.

- Margot, Vincent, Christophe Geissler, Carmine De Franco, and Bruno Monnier. 2021. “ESG Investments: Filtering versus Machine Learning Approaches.” *Applied Economics and Finance* 8 (2): 1. <https://doi.org/10.11114/aef.v8i2.5097>.
- Mattis, Michael. 2008. “CSR-Washing Is the New Greenwashing.” *Moneywatch*, 2008. <https://www.cbsnews.com/news/csr-washing-is-the-new-greenwashing/>.
- McKenzie, Mark, Nancy Barrett, Neil Camm, Melissa Davis, and Rosa Esnard. 2020. “The Time Has Come: The KPMG Survey of Sustainability Reporting 2020.” *KPMG IMPACT*, 63.
- MSCI ESG Research. 2020. “MSCI ESG Research Controversies and Global Norms.” *MSCI ESG Research*, 18.
- Neve, Jan-Emmanuel de, and et al. 2018. “Work and Well-Being: A Global Perspective.” In *Global Happiness Policy Report*. New York, NY: Sustainable Development Solutions Network.
- Poole, Veronica, and Kirsten Sullivan. 2021. “Tectonic Shifts: How ESG Is Changing Business, Moving Markets, and Driving Regulation.” *Deloitte*, 28.
- Ragbir, Nadine K., Stephen Rice, Scott R. Winter, and Elaine C. Choy. 2021. “Emotions and Caring Mediate the Relationship between Knowledge of Sustainability and Willingness to Pay for Greener Aviation.” *Technology in Society* 64 (February): 101491. <https://doi.org/10.1016/j.techsoc.2020.101491>.
- Refinitiv. 2021. “ENVIRONMENTAL, SOCIAL AND GOVERNANCE SCORES FROM REFINITIV.”
- Rosset, Saharon. 2004. “Model Selection via the AUC.” In *Twenty-First International Conference on Machine Learning - ICML '04*, 89. Banff, Alberta, Canada: ACM Press. <https://doi.org/10.1145/1015330.1015400>.
- Ruggie, John Gerard, and Emily K. Middleton. 2019. “Money, Millennials and Human Rights: Sustaining ‘Sustainable Investing.’” *Global Policy* 10 (1): 144–50. <https://doi.org/10.1111/1758-5899.12645>.
- Russell, Stuart J., Peter Norvig, and Ernest Davis. 2010. *Artificial Intelligence: A Modern Approach*. 3rd ed. Prentice Hall Series in Artificial Intelligence. Upper Saddle River: Prentice Hall.
- Sarker, Iqbal H. 2021. “Machine Learning: Algorithms, Real-World Applications and Research Directions.” *SN Computer Science* 2 (3): 160. <https://doi.org/10.1007/s42979-021-00592-x>.
- Schiermeier, Quirin. 2015. “The Science behind the Volkswagen Emissions Scandal.” *Nature*, September, nature.2015.18426. <https://doi.org/10.1038/nature.2015.18426>.
- Schroders. 2017. “Managing Corporate Controversies: The Role of Environmental, Social and Governance (ESG) Ratings.” Schroders.
- Shrivastava, Deepshikha, Sugata Sanyal, Arnab Kumar Maji, and Debdatta Kandar. 2020. “Bone Cancer Detection Using Machine Learning Techniques.” In *Smart Healthcare for Disease Diagnosis and Prevention*, 175–83. Elsevier. <https://doi.org/10.1016/B978-0-12-817913-0.00017-1>.
- Singal, Amit G, Ashin Mukherjee, Joseph B Elmunzer, Peter D R Higgins, Anna S Lok, Ji Zhu, Jorge A Marrero, and Akbar K Waljee. 2013. “Machine Learning Algorithms Outperform Conventional Regression Models in Predicting Development of Hepatocellular Carcinoma.” *American Journal of Gastroenterology* 108 (11): 1723–30. <https://doi.org/10.1038/ajg.2013.332>.

- Sinha, Priyanshu, Saptarshi Purkayastha, and Judy Gichoya. 2019. "Full Training versus Fine Tuning for Radiology Images Concept Detection Task for the ImageCLEF 2019 Challenge," 8.
- Smith, Miriam, and Christine Keating. 2017. "24. Homonationalism and the Comparative Politics of LGBTQ Rights." In *LGBTQ Politics: A Critical Reader*. New York, USA: New York University Press.
- Soto, Axel J., Rocío L. Cecchini, Gustavo E. Vazquez, and Ignacio Ponzoni. 2009. "Multi-Objective Feature Selection in QSAR Using a Machine Learning Approach." *QSAR & Combinatorial Science* 28 (11–12): 1509–23. <https://doi.org/10.1002/qsar.200960053>.
- Subramanian, Savita, James, Jill Carey Hall, Toby Wade, and Alex Makedon. 2019. "ESG Matters - US." Bank of America Merrill Lynch.
- Syarif, Iwan, Adam Prugel-Bennett, and Gary Wills. 2016. "SVM Parameter Optimization Using Grid Search and Genetic Algorithm to Improve Classification Performance." *TELKOMNIKA (Telecommunication Computing Electronics and Control)* 14 (4): 1502. <https://doi.org/10.12928/telkomnika.v14i4.3956>.
- Tin Kam Ho. 1998. "The Random Subspace Method for Constructing Decision Forests." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (8): 832–44. <https://doi.org/10.1109/34.709601>.
- Tschopp, Daniel, and Ronald J. Huefner. 2015. "Comparing the Evolution of CSR Reporting to That of Financial Reporting." *Journal of Business Ethics* 127 (3): 565–77. <https://doi.org/10.1007/s10551-014-2054-6>.
- United States. 1972. "Federal Water Pollution Control Act Amendments of 1972." Pub.L. 92-500.
- Velte, Patrick. 2017. "Does ESG Performance Have an Impact on Financial Performance? Evidence from Germany." *Journal of Global Responsibility* 8 (2).
- Vrgazova, Bois. 2021. "The Proportion for Splitting Data into Training and Test Set for the Bootstrap in Classification Problems." *Business Systems Research* 12 (1).
- Walker, Kent, and Fang Wan. 2012. "The Harm of Symbolic Actions and Green-Washing: Corporate Actions and Communications on Environmental Performance and Their Financial Implications." *Journal of Business Ethics* 109 (2): 227–42. <https://doi.org/10.1007/s10551-011-1122-4>.
- Wellalage, Nirosha Hewa, and Vijay Kumar. 2020. "Does It Pay to Be Green? Environmental Performance and Firm Financing during COVID-19 Outbreaks." Preprint. In Review. <https://doi.org/10.21203/rs.3.rs-83297/v1>.
- Yang, Qiang, and Xindong Wu. 2006. "10 CHALLENGING PROBLEMS IN DATA MINING RESEARCH." *International Journal of Information Technology & Decision Making* 05 (04): 597–604. <https://doi.org/10.1142/S0219622006002258>.

## 9 Appendix

### Appendix 1



**Figure 8.** Development of Controversy Count

## Appendix 2

**Table 8.** Overview Controversy Types

Pillar	Controversy Type
Environmental	'Environmental Controversies' 'Bribery, Corruption and Fraud Controversies', 'Intellectual Property Controversies', 'Public Health Controversies', 'Business Ethics Controversies', 'Tax Fraud Controversies', 'Anti-Competition Controversies Count', 'Critical Countries Controversies', 'Human Rights Controversies', 'Freedom of Association Controversies', 'Child Labor Controversies',
Social	'Responsible Marketing Controversies', 'Product Quality Controversies', 'Consumer Complaints Controversies Count', 'Controversies Customer Health & Safety', 'Controversies Responsible R&D', 'Controversies Privacy', 'Controversies Responsible Marketing', 'Controversies Product Access', 'Diversity and Opportunity Controversies', 'Wages Working Condition Controversies Count', 'Employees Health & Safety Controversies'
Governance	'Executive Compensation Controversies', 'Mgt Compensation Controversies Count', 'Shareholder Rights Controversies Count', 'Insider Dealings Controversies Count', 'Accounting Controversies Count', 'Insider Dealings Controversies', 'Accounting Controversies'

Appendix 3



Figure 9. SHAP value: Salaries and Wages with Respect to Company Size



Figure 10. SHAP value: Training Hours with Respect to Company Size

Appendix 4

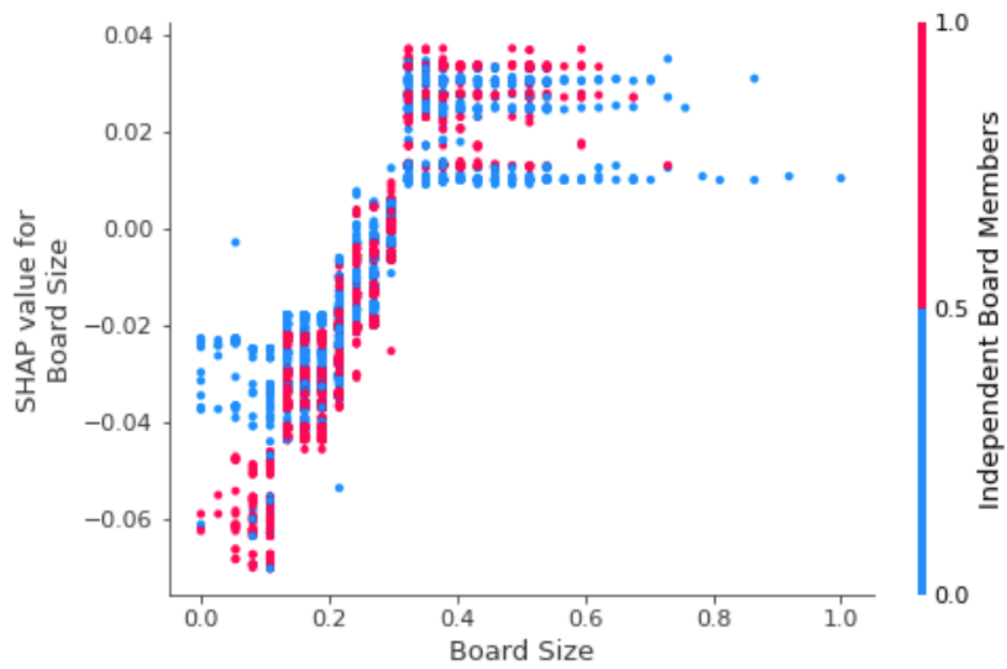


Figure 11. SHAP value: Board Independence with Respect to Board Size