Board gender diversity and workplace diversity: a machine learning approach

Mikko Ranta and Mika Ylinen

Mikko Ranta and Mika Ylinen are both based at the School of Accounting and Finance, University of Vaasa, Vaasa, Finland.

Abstract

Purpose – *This study aims to examine the association between board gender diversity (BGD) and workplace diversity and the relative importance of various board and firm characteristics in predicting diversity.*

Design/methodology/approach – *With a novel machine learning (ML) approach, this study models the association between three workplace diversity variables and BGD using a social media data set of approximately 250,000 employee reviews. Using the tools of explainable artificial intelligence, the authors interpret the results of the ML model.*

Findings – *The results show that BGD has a strong positive association with the gender equality and inclusiveness dimensions of corporate diversity culture. However, BGD is found to have a weak negative association with age diversity in a company. Furthermore, the authors find that workplace diversity is an important predictor of firm value, indicating a possible channel on how BGD affects firm performance.*

Originality/value – *The effects of BGD on workplace diversity below management levels are mainly omitted in the current corporate governance literature. Furthermore, existing research has not considered different dimensions of this diversity and has mainly focused on its gender aspects. In this study, the authors address this research problem and examine how BGD affects different dimensions of diversity at the overall company level. This study reveals important associations and identifies key variables that should be included as a part of theoretical causal models in future research.*

Keywords *Corporate governance, Board composition, Machine learning, Workplace diversity, Explainable AI*

Paper type *Research paper*

1. Introduction

The gender composition of corporate boards and its influence on other aspects of business is a widely studied research topic in the corporate governance literature. Previous research has studied, for example, how board gender diversity (BGD) affects executive appointments (Cook and Glass, 2014, 2015), board effectiveness (Adams and Ferreira, 2009), corporate value (Carter *et al.*, 2003), operational performance (Campbell and Mínguez-Vera, 2008) and risk-taking (Palvia *et al.*, 2015). For an extensive literature review of BGD research, see Baker *et al.* (2020) or Kirsch (2018). However, although the effects of BGD have been extensively studied, many questions remain on what effects can be expected from a more gender-balanced board composition (Kirsch, 2018).

For example, existing research has mainly ignored the effects of BGD on the diversity culture of workplaces. Few studies have explored the effect of BGD on diversity at the top management level. For instance, Cook and Glass (2014, 2015) studied how BGD is associated with the appointment and success of women chief executive officer (CEOs). Moreover, several studies have found that high diversity at the top level of organizations is positively associated with diversity at the lower managerial levels (Bilimoria, 2006; Matsa and Miller, 2011; Skaggs *et al.*, 2012). However, the effects of BGD on the diversity below the management levels are mainly omitted in the current corporate governance literature.

Received 26 January 2022
Revised 17 May 2022
25 August 2022
16 October 2022
Accepted 30 December 2022

The authors would like to thank the editor and two anonymous reviewers for their very insightful and helpful comments.

Furthermore, existing research has not taken into account different dimensions of workplace diversity and has mainly focused on its gender aspects. Although one could assume that BGD improves the gender equality culture of a company, the same is not necessarily true for other diversity dimensions. This lack of research is surprising, as board diversity is often justified by its positive spillover effects on the rest of the organization (Kirsch, 2018). In general, the existing literature has not provided good answers about how BGD affects diversity culture below the management level and what kinds of effects can be expected from a more balanced board composition.

In this study, we address this research problem and examine how BGD affects different dimensions of workplace diversity at the overall company level. For that purpose, we use a unique data set of approximately 250,000 employee reviews from a social media platform aggregated at the yearly level and measure diversity using three employee assessments from the reviews: *gender equality*, *inclusive/diverse* and *attitude toward older colleagues*. Furthermore, we use a novel machine learning (ML) approach to measure the relative importance of board and firm characteristics to company diversity and analyze the nature of these features' associations (nonlinearities). This will give us tools to analyze the reasons behind the mixed results of the association between BGD and, for example, financial performance (Kirsch, 2018). Moreover, our ML approach allows us to estimate the relative importance of BGD compared to other board and firm characteristics, i.e. how much practical relevance BGD has in affecting workplace diversity.

Our research contributes to the existing literature by examining how BGD is associated with company diversity. The results have important implications. They bring new evidence of how BGD is connected to different dimensions of company diversity below the management level and reveal how the benefits of BGD might be channeled through the company for positive outcomes. Our data offer the possibility of examining diversity in more detail using three dimensions (gender diversity, inclusiveness and age diversity). Thus, we get new information about the connection documented as being causal in some studies (Kirsch, 2018). These measures capture different aspects of workplace diversity and reveal new information about the drivers of company diversity. Furthermore, our method allows nonlinear analysis, revealing new information about the nature of the associations between BGD and company diversity.

Our research also contributes to the growing literature in business disciplines that exploits social media information as sources of company-specific information (Hales *et al.*, 2018; Miller and Skinner, 2015). Previous research has used information from sources like Amazon, Twitter, Glassdoor, different internet bulletin boards and crowdsourced economics estimates. Our work adds to a growing literature that has studied social media information aggregation and the wisdom of the crowd. This literature suggests that "the aggregation of information provided by many individuals often results in predictions that are better than those made by any single member of the group, or even experts" (Bartov *et al.*, 2018, p. 28), showing evidence that social media and specifically employee reviews reveal fundamental information about the firm.

Moreover, our research contributes to the existing literature on ML applications in business disciplines. This approach offers many benefits compared to traditional econometric approaches, where data-driven model selection, efficient predictions, the true nature of dependencies and the ability to deal with multicollinearity are the most important features from the point of view of this study (see, for example, Bertomeu, 2020; Jones, 2017; Karolyi and Van Nieuwerburgh, 2020). In our research setting, traditional econometric models, e.g. linear statistical models, are limited by the possible multicollinearity of features and potential nonlinearities in the data.

More specifically, in this research, we use gradient boosting, a technique that combines the predictions of many weak estimators into one strong prediction (Friedman, 2001). Gradient boosting overcomes many limitations of more traditional research models and provides a potentially useful method for including the board and firm characteristics in a single statistical framework without any cost to model stability or performance (see, for example, Bertomeu, 2020; Jones, 2017, for a more detailed discussion). In addition, the high-dimensional gradient-boosting model uses the full set of input variables and automatically detects all the important interaction effects. This is one clear advantage compared to linear regression. Usually, a gradient-boosting model consists of decision trees, where each tree will sequentially pick the variables that best explain an outcome variable. The initial splits to different branches are typically based on the most critical variables, and the following splits will build interactions with other variables. Thus, the approach also implements a variable selection procedure in each tree, with benefits similar to different regularization techniques, because the number of splits controls the set of variables used in the model (Bertomeu, 2020; Friedman, 2001, 2002).

Furthermore, our paper makes methodological contributions to the growing studies of ML in corporate governance by using the tools of explainable AI to interpret the results of an ML model. We apply the Shapley additive explanations (SHAP) method (Lundberg *et al.*, 2020) to interpret the results of the boosting model. SHAP is a game-theoretic method used to explain the local importance (individual predictions) of features. SHAP indicates how much each feature contributes to an individual prediction. We use SHAP values to explore the relations between individual predictors and company diversity (shape of relations and their statistical significance) and estimate the global importance of the features using several aggregates calculated from the SHAP values.

Our empirical findings demonstrate that BGD is the most important board characteristic and a vital feature overall in predicting company gender equality and inclusiveness. Furthermore, from the local importance perspective, there is a strong positive and statistically significant association between BGD and these company diversity measures. However, for age diversity, the role of BGD is much weaker. Furthermore, this weak association is observed to be negative. With the exception of BGD, firm characteristics are slightly more important predictors of company diversity than board characteristics. Our findings also indicate how the positive aspects of BGD might be channeled as positive financial outcomes for companies through the positive effects of company diversity. Additional analysis reveals that the company diversity variables are important predictors of firm value and significantly more important than BGD. This indicates a possible mechanism of how the positive effects of BGD are channeled through the company via company diversity as positive business outcomes, in this case, firm value.

2. Literature review

2.1 Board composition

The current consensus in the literature is that BGD is beneficial for general company performance. The upper echelons theory (Hambrick and Mason, 1984) predicts that managerial characteristics strongly influence organizational outcomes and this has ignited substantial research on BGD's role in this theory. Diversity is considered an element of social capital that offers diverse perspectives and thus improves companies' governance function (Adams and Ferreira, 2009; Booth-Bell, 2018; Calabrese and Manello, 2021). Overall, the list of positive outcomes from diverse boards is extensive. Adams and Ferreira (2009) found that diversity positively affects governance. Moreover, Gul *et al.* (2011) suggested that diverse boards improve share price informativeness through improved company disclosure, while Loukil *et al.* (2019) found evidence that BGD improves stock market liquidity. Similarly, the findings of and Jurkus *et al.* (2011) indicated that BGD has a positive effect on risk-adjusted stock returns. The research of Dezsö and Ross (2012), Kor (2006) and Miller and Triana (2009) suggests that board diversity has a positive effect on innovation, while Ali *et al.* (2021) found evidence of improved firm efficiency. Furthermore, Bear *et al.* (2010) and Post *et al.* (2011) found positive linkages between BGD and company sustainability practices.

Neutral findings include those from Krishnan and Parsons (2008) and Barua *et al.* (2010), who found that a diverse executive level causes financial reporting to be more cautious and conservative. Furthermore, Menicucci and Paolucci (2022), Hurley and Choudhary (2020), Faccio *et al.* (2016) and Huang and Kisgen (2013) documented that companies with a diverse executive level (especially board) are significantly more risk-averse. Related research by Ben Saad and Belkacem (2022) found that BGD affects capital structure decisions through this risk-taking channel. Chatjuthamard *et al.* (2021) investigated the relationship between BGD and hostile takeover vulnerability and found a negative association between them.

One of the most intensively studied effects of BGD is financial performance, and the findings are mixed. Most studies have found a positive association with financial performance (Anderson *et al.*, 2011; Campbell and Mínguez-Vera, 2008; Carter *et al.*, 2003; Erhardt *et al.*, 2003; Joecks *et al.*, 2013; Leyva-Townsend *et al.*, 2021; Mazzotta and Ferraro, 2020; Saleh *et al.*, 2021). However, several studies have also found a negative or insignificant association between BGD and financial performance (Adams and Ferreira, 2009; Ahern and Dittmar, 2012; Chapple and Humphrey, 2014; Darmadi, 2013; Rose, 2007).

Despite this extensive study base, the relationship between BGD and many other business practices is still unclear (Baker *et al.*, 2020). For example, the effect of BGD on company diversity is largely missing from the existing literature. Few studies have explored the effect of BGD on diversity at the top management level. Cook and Glass (2014, 2015) studied how BGD is associated with the appointment and success of women CEOs. Moreover, several studies have found that high diversity at the top level of organizations is positively associated with diversity at the lower managerial levels (Bilimoria, 2006; Matsa and Miller, 2011; Skaggs *et al.*, 2012). However, the effects of BGD on workplace diversity below the management levels are mainly omitted in the current corporate governance literature. Furthermore, existing research has not considered different dimensions of diversity and has mainly been focusing on the gender aspects of it. Although improved gender equality could be expected from better BGD, all the dimensions of diversity do not necessarily improve with better BGD. This is an evident shortcoming as company diversity is shown to influence company performance (Filbeck *et al.*, 2017; Kirsch, 2018). Therefore, it is interesting to shed light on whether BGD is affecting company diversity and whether this could be the channel for how it influences, for example, financial performance.

What are the mechanisms that drive these effects? Previous research has documented several possible theoretical explanations of how BGD can impact the rest of the organization. Drawing insight from signaling theory (Connelly *et al.*, 2011), high BGD improves company reputation, which could positively affect aspects like market performance (Bear *et al.*, 2010). A similar mechanism could affect companies' recruitment process and send a signal to job seekers about the companies' diversity policies and, therefore, also affect the diversity below the management level. Another line of research examines whether BGD initiates a policy change for companies (Triana *et al.*, 2014). This policy change could be expected (but not necessarily) to include choices like higher social responsibility and more diverse culture (Harjoto *et al.*, 2015; however, see also Rao and Tilt, 2020). From this background, we aim to find more information on the following research questions:

RQ1: Is BGD associated with the diversity below the management level?

RQ2: Are there differences between the dimensions of diversity (gender diversity, inclusiveness and age diversity)?

RQ3: How strongly are different dimensions of workplace diversity associated with market performance?

2.2 Gradient boosting

Ensemble methods are algorithms that combine the prediction of many weak predictors into a single strong prediction. The most popular ensemble methods are random forests and boosting. Overall, they have been applied to research problems in business disciplines at an increasing pace. For example, random forests have been used to forecast stock index returns (Kumar and Thenmozhi, 2014), financial fraud (Liu *et al.*, 2015) and stock market price movements (Khaidem *et al.*, 2016). Barboza *et al.* (2017) evaluated the suitability of many different ML algorithms for bankruptcy prediction and concluded that the ensemble methods are usually the most efficient option. Gu *et al.* (2020) made a similar comparison with regard to empirical asset pricing and found that alongside neural networks, ensemble methods based on decision trees are the best-performing methods. They traced this improved performance to the ability to model nonlinear interactions between predictors that neural networks and ensemble methods possess.

Gradient boosting is a modification of the original boosting model that uses the gradient descent algorithm to train them (Friedman, 2001). One of the latest iterations, the extreme gradient-boosting algorithm (Chen and Guestrin, 2016), adds highly optimized code, scalability, regularization and parallel computing to the algorithm. It has been very popular in recent years for practical applications. Several papers have also applied it in business discipline research. Climent *et al.* (2019) used an extreme gradient-boosting approach to predict bank distress in the eurozone, while Carmona *et al.* (2019) performed a similar analysis of the US banking sector. Moreover, Zie¸ba *et al.* (2016) combined the extreme gradient-boosting algorithm with synthetic features to improve bankruptcy prediction accuracy and evaluated their model using Polish companies. Extreme gradient boosting has also been used to construct credit risk assessment models for financial institutions (Chang *et al.*, 2018), to build a multifactor stock selection model (Zhang and Chen, 2019), to predict customer churn (Gregory, 2018) and to predict the likelihood of loan default on online peer-to-peer lending platforms (Zheng, 2019).

Papers implementing other boosting models have included Cortés *et al.* (2008), who used AdaBoost to predict financial failure; Bao *et al.* (2020), who used boosting to predict accounting fraud in publicly traded US firms; Jones (2017) and Jiang and Jones (2018), who used boosting for bankruptcy prediction; and Pierdzioch *et al.* (2015), who used boosting to predict gold and silver prices successfully. Closely related to our research, a recent study by Yousaf *et al.* (2021) used gradient boosting and several other ML methods to analyze the connection between board diversity and financial distress.

2.3 Shapley additive explanations

Explainable AI is used relatively little in business research. However, the enthusiasm of scholars for these methods is high, and they are implemented at an increasing pace. For example, Doornenbal *et al.* (2021) used tools of explainable AI to analyze personality traits that define a leader. Furthermore, their paper includes good instructions on how researchers can use these tools to interpret ML algorithms, assess model complexity and rank feature importances. They argue that with the advancement of ML, researchers are now better equipped to model complex nonlinear associations between the variables of interest and interpret them using the tools of explainable artificial intelligence (AI). Spisak *et al.* (2019) draw similar conclusions and review the possibilities that explainable AI opens for empirical business research.

Although SHAP is a relatively new methodological innovation for explainable AI, several studies in business disciplines have already adopted it. Jabeur *et al.* (2021) use ML and

SHAP values to predict oil prices during the COVID-19 pandemic and demonstrate how SHAP values can be used to infer the core features that predict these prices. Similarly, Futagami *et al.* (2021) show how SHAP values can be used to identify the most important features that predict mergers and acquisitions. Furthermore, they use SHAP values to interpret how these features contributed to the prediction of acquisition occurrence. Ylinen and Ranta (2021) combine SHAP values with bootstrap methods (Efron, 1979) to include uncertainty estimations to results. More specifically, they use SHAP values to analyze how different characteristics of employee-friendly corporate culture affect company performance. As the last example, we mention Lin and Bai (2021), who use SHAP values to infer the most important determinants of debt financing in heavily polluting enterprises.

We consider the choice of gradient boosting and SHAP to fit well for answering the proposed research questions. Our model includes many variables whose connection a priori is unknown. The ML approach allows data-driven model selection that automatically considers all the interactions and nonlinear associations between variables (Bertomeu, 2020). For example, we can easily analyze the implications of the critical mass theory for BGD with our nonlinear approach (Joecks *et al.*, 2013. Furthermore, the high number of covariates raises multicollinearity concerns that our decision tree-based approach alleviates. Moreover, ML models have proven to be very efficient predictors that help us estimate our findings' practical relevance (Jones, 2017). One major benefit of decision tree-based models is their ability to use observations with missing values for some variables. As decision trees are built using the available information for each observation, models with many covariates (like ours) are more efficiently analyzed using decision tree-based ML algorithms.

3. Data and methodology

3.1 Board and firm characteristics

When collecting the variables for our model, the leading idea has been to add a substantial number of control variables to minimize the risk of endogeneity issues. As discussed earlier, our approach is resistant to multicollinearity, which allows variables to be added to our model with a just-in-case mindset. Our analysis covers the board characteristics that have been extensively used in prior literature to study how board structure affects different aspects of the company (Baker *et al.*, 2020 for an extensive survey). The specific characteristics are *BGD*, measured as the percentage of female board members; *board meeting attendance*, defined as the average overall attendance percentage of board meetings, as reported by the company; *board member affiliation*, defined as the average number of other board memberships of the board members; *board member compensation*, measured as the total compensation of the nonexecutive board members; *independent board member*, defined as the percentage of independent board members, as reported by the company; *number of board meetings*, measured as the number of board meetings during a fiscal year; *average board tenure,* measured as the average number of years each board member has served on the board; *nonexecutive board members*, measured as the percentage of nonexecutive board members; *executive members gender diversity*, measured as the percentage of female board members within the executive members of the board; *specific skills*, measured as the percentage of board members who have either an industry-specific background or a strong financial background; and *board size*, measured as the number of board members at the end of the fiscal year.

Furthermore, we measure the CEO-level participation on corporate boards using two dummy variables. *CEO board member* gets a value of one if the CEO of the company is also a member of the corporate board. *Chairman is ex-CEO* is assigned a value of one if the chairperson of the board has previously served as the CEO of the same company. Finally, we include a variable *strictly independent board members* that measures the percentage of those board members whom the company does not employ; not representing or employed by a majority shareholder; not having served on the board for more than ten years; not a reference shareholder with more than 5% of holdings; having no cross-board membership; having no recent, immediate family ties to the corporation; and having not accepted any compensation other than compensation for board service. The descriptive statistics of the board characteristics are provided in Table 1.

As firm characteristics, we select variables that have been shown to play a significant role in the previous corporate governance literature (Frijns *et al.*, 2016). As a market measure of financial performance, we use Tobin's Q and, as an operational measure of performance, return on assets. Other firm characteristics in the model are the total assets of a company; leverage, calculated as a fraction of long-term debt and total assets; R&D intensity, calculated as a fraction of R&D expenditure and total assets; implied volatility; and firm age. We also control industry and year fixed effects using dummy variables for standard industrial classification (SIC) codes (one-digit) and fiscal years. The descriptive statistics of the firm characteristics are provided in Table 2.

3.2 Employer review sample

Kununu, founded in Austria in 2007, is a social media-based recruiting website similar to Glassdoor.com, from which data have been used recently in accounting and finance research (Green *et al.*, 2019; Hales *et al.*, 2018). The website includes information about job postings, quantitative and qualitative employee reviews on a variety of firm characteristics, salary data, the interview process (before, during and after the job interview), company culture insights and an overall score, which combines scores from

Table 1 Descriptive statistics for the board characteristics, rounded to three significant figures

Characteristic	*Count*	*Mean*	*SD*	*Min*	*25%*	*50%*	*75%*	*Max*
No. of meetings	588	7.83	3.30	4.00	6.00	7.00	9.00	21.0
Meeting attendance	588	79.0	8.39	75.0	75.0	75.0	75.0	100
Board size	588	10.9	1.94	7.00	9.00	11.0	12.0	16.0
CEO board member	588	0.986	0.116	0.00	1.00	1.00	1.00	1.00
Chairman is ex-CEO	588	0.743	0.437	0.00	0.00	1.00	1.00	1.00
Independent members	588	83.8	9.37	50.0	80.0	86.7	90.9	93.8
Strictly independent members	588	48.7	16.6	11.1	37.5	50.0	60.0	88.9
Member affiliations	588	1.14	0.553	0.054	0.750	1.10	1.46	3.22
Nonexecutive members	588	85.2	7.34	62.5	81.8	86.7	91.1	94.1
BGD	588	21.9	9.55	0.00	15.4	21.4	27.3	50.0
Executive members gender diversity	588	17.0	11.9	0.00	9.09	15.4	25.0	50.0
Specific skills	588	53.5	16.0	12.5	42.9	53.9	63.6	91.9
Average tenure	588	9.76	3.70	1.75	7.60	9.30	11.2	20.9
Member compensation (1,000$)	588	2,650,000	1,030,000	457,000	2,000,000	2,590,000	3,150,000	9,310,000

Table 2 Descriptive statistics for the firm characteristics, rounded to three significant figures

Characteristic	*Count*	*Mean*	*SD*	*Min*	*25%*	*50%*	*75%*	*Max*
Total assets (1,000$)	588	44700	101000	415	5570	14400	36700	877000
Leverage	588	0.993	4.49	−15.0	0.346	0.708	1.25	32.1
R&D intensity (%)	588	0.0112	0.0223	0.00	0.00	0.00	0.0135	0.142
ROA (%)	588	0.0631	0.0629	−0.168	0.0262	0.0564	0.0971	0.217
Implied vol.	588	27.7	8.05	15.0	22.0	26.5	31.7	57
Firm age	588	447	225	12.0	264	419	636	804
Tobin's Q	588	2.27	1.32	0.898	1.39	1.84	2.66	7.31

company, application and internship reviews. This score is also compared to the industry average and the average score from all reviews. Moreover, each review includes a recommendation rate, which shows the percentage of people who recommended this company over the past two years. Employees can also enter free-text responses related to suggestions for improvement and issues that the respondent likes and dislikes about the company. Finally, these data also include compensation information concerning benefits and perks (26 items) that are offered to employees. The website allows current and former employees to rate their employers on 18 attributes using a five-point scale. Most of these variables are related to aspects like company image, career opportunities and work–life balance. However, three of these attributes are directly related to company diversity, which allows us to measure company diversity using three dimensions: *gender equality*, *inclusive/diverse* and *attitude toward older colleagues*. The actual survey questions presented to employees for these diversity variables are provided in Table A1 in the Appendix.

By 2019, Kununu had reported hosting over 3.96 million individual employer reviews across over 933,000 companies. We focus our analysis on Standard and Poor's 1,500 firms, of which 591 exist in the raw Kununu data. This decreases the sample size to 250,000. To reduce noise in the data, we include in the sample only those firms that had at least 20 reviews per year in the 2014–2017 period. Furthermore, due to significant differences in the amount of data available for different variables, we keep only observations for which we have data for all the diversity variables to improve reliability. After these actions, the final sample has 588 observations. Table 3 provides descriptive statistics for the modeled diversity variables. The statistics are as expected, indicating a stable close-to-normal distribution for the observations.

3.3 Gradient boosting

We use gradient boosting to model the relationship between the firm/board characteristics and the company diversity measures. A nonlinear ML model is chosen so that we can model the nonlinearities and take into account the interaction effects between covariates. Furthermore, boosting models can naturally handle multicollinearity, which is important as our model includes many covariates that correlate strongly with each other. Using a single model with all the firm and board characteristics included allows us to estimate the pure effects of covariates to the prediction more efficiently using SHAP values. With this approach, we get a model that recognizes nonlinearities efficiently and achieves a very high predictive power.

Boosting models are usually constructed using decision trees or regression trees as simple predictors. The prediction of the model is improved iteratively so that the next decision tree is fed with reweighted data, where the weight of the misclassified data points is increased. Friedman (2001) presented an efficient way to train tree ensemble models using second-order derivatives that is also the basis of the extreme gradient-boosting algorithm. The approach used in this research is based on tree ensemble models:

Table 3 Descriptive statistics for the diversity variables, rounded to three significant figures

Variable	*Count*	*Mean*	*SD*	*Min*	*25%*	*50%*	*75%*	*Max*
Gender equality	588	3.33	0.590	1.17	2.97	3.34	3.73	5.00
Inclusive/diverse	588	3.12	0.729	1.00	2.70	3.14	3.59	5.00
Attitude toward older colleagues	588	3.28	0.586	1.00	2.93	3.32	3.65	5.00

$$\hat{y} = \sum_{j=1}^{J} f_j(x_i),\ \ f_j \in \mathcal{F},$$

where $\mathscr{F}$ is the space of classification and regression trees (CART). The regression trees contain a continuous score on each of the leaves, contrary to the ordinary decision trees that have class labels on the leaves. The functions of the additive model, f_j, correspond to different tree structures and leaf weights. Training the model means minimizing the objective function:

$$\mathcal{L} = \sum_i l(\hat{y}_i, y_i) + \sum_j \ \Omega(f_j),$$

where l is a differentiable convex loss function and Ω is a regularization term that penalizes the model's complexity. We use a squared-log-loss function:

$$l(\hat{y}, y) = \frac{\left(\log(\hat{y}+1) - \log(y+1)\right)^2}{2},$$

because it can be shown to be more robust to outliers than the commonly used squared-loss function (Chen and Guestrin, 2016).

Traditional optimization methods cannot be used because the model uses functions as parameters. Instead, the model is trained in an additive manner, greedily adding a new function (a regression tree) to the model that most improves the predictions. After taking a second-order approximation and removing constant terms, the objective function can be written in a simplified form:

$$\tilde{\mathcal{L}} = \sum_{i=1}^{n} \left[g_i f_t(\boldsymbol{x}_i) + \frac{1}{2} h_i f_t^2(\boldsymbol{x}_i) \right] + \ \Omega(f_t),$$

where g_i and h_i are the first- and second-order derivatives of the model from the previous round, respectively. From this equation, the optimal leaf weights and the optimal split points for tree branches can be calculated. The exact details are omitted and can be found in Friedman (2001) and Chen and Guestrin (2016). When data values and the number of features increase, it is impossible to search and test all the added function's possible tree structures. Therefore, the extreme gradient-boosting algorithm uses an approach where a tree is built starting from a single leaf using a greedy algorithm. The exact greedy algorithm can be used for single-machine solutions and an approximate greedy algorithm for parallel computing.

The extreme gradient boosting allows fine-tuning of numerous parameters related to learning rate and regularization. The parameters fine-tuned in this research are *shrinkage*, *column subsampling*, *row subsampling*, *gamma*, *tree depth*, *minimum leaf weight* and *number of trees*. Because finding optimal parameter values is computationally intensive, we use the default values for the other available parameters in the model. The shrinkage parameter reduces newly added weights by a predefined factor to leave room for future trees to improve the model. The column subsampling method is previously used in random forest algorithms and means selecting a subsample of features for the newly added tree. Similarly, the row subsampling method means selecting a subsample of observations for the tree. The minimum leaf weight parameter controls the minimum number of observations that must be in every leaf, and the gamma parameter controls the minimum loss reduction required to make a further partition on a leaf node of the tree (Chen and Guestrin, 2016). Finally, the tree depth controls the number of divisions allowed in the trees. Overall, these methods counter overfitting, and the column subsampling method also speeds up computation when a parallel computation approach is used. We do the optimization in two steps using fivefold cross-validation. First, a thorough grid-search algorithm is applied for the tree structure and regularization parameters of the model. After that, a separate cross-validation run is applied to the number of decision trees. Table 4 provides the optimal parameters for the models of each diversity dimension.

Alongside efficient parameters for regularization, the extreme gradient-boosting method's key improvements are highly efficient algorithms for optimal split point search, options for parallel learning and sparsity-aware algorithms. Together these innovations make the algorithm an order of magnitude faster than previous implementations and make it possible to use the gradient-boosting algorithm for much larger data sets than before. The details of these benefits can be found in Chen and Guestrin (2016).

3.4 Shapley additive explanations values

The challenge of nonlinear ML models is the difficulty of interpretation. For example, it is impossible to describe a relationship using a single parameter as in linear regression. Thus, we cannot use metrics like *t*-values for a single parameter describing a single feature. Nonlinear models need new ways to analyze the importance of features. Standard metrics used to analyze the importance of the features in ML ensemble models are the weight, gain and cover metrics. For example, in decision tree ensemble models, the weight metric is the relative number of times a feature is used to split data, the gain metric is the improvement in accuracy brought by a feature to the branches it is on, and the cover is the number of observations that go through splits by this feature and which divide that value by the number of times the feature is used in the model. These are very problematic metrics and can give inconsistent and contradictory results (Lundberg and Lee, 2017). For example, the weight metric undervalues the importance of binary features because it divides data into only two categories, and therefore, their appearance in trees is more scarce than features with numerical values. Furthermore, the weight and the cover metric do not consider the importance of the division. Divisions with a small improvement in accuracy are equally important as divisions with a substantial improvement in accuracy.

Because of the shortcomings of these metrics, we interpret the results of the extreme gradient-boosting model with the SHAP method (Lundberg *et al.*, 2020), which is based on game-theoretically optimal Shapley values. This method has many desirable performance metrics properties, such as local accuracy, consistency, efficiency, symmetry and additivity (Lundberg and Lee, 2017). TreeSHAP is a variant of the SHAP for tree-based ML models. It is fast, computes exact Shapley values and can handle feature dependencies. Furthermore, because the method measures individual predictions, it offers versatile ways to measure the model performance.

Thus, SHAP is considered to be the most reliable metric for tree-based ML methods at the moment (Molnar, 2021). Furthermore, the SHAP values have a very intuitive explanation. In the regression setting, they calculate the effect of a single feature on a single prediction. We

Table 4 Optimal parameters for the boosting models

Parameter	*Gender equality model*	*Inclusive/diverse model*	*Attitude toward older colleagues model*
Shrinkage	0.03	0.05	0.05
Column subsampling	0.8	0.8	0.8
Row subsampling	0.8	0.8	0.8
Tree depth	4	4	4
Minimum leaf weight	1	1	1
Gamma	0.0	0.0	0.0
No. of trees	125	80	90

get explanations for every observation and for how each feature affects the given prediction. The sum of the effects (SHAP values) equals the final prediction for a given observation – in this case, the average value of a diversity grade. For a classifier setting, using a logistic regression-type prediction, the interpretation is somewhat more complicated. The SHAP values then estimate the effect of a feature on the (log) probability of a class. For the mathematical details of SHAP, see Lundberg *et al.* (2020).

4. Results

4.1 Model validation

To validate our ML approach, we compare the out-of-sample predictive power of the boosting model to an ordinary least squares (OLS) regression model and a least absolute shrinkage and selection operator (LASSO) regression model. The data are split into training and testing parts (80%/20%), and fivefold cross-validation with the training data is used to optimize the models' hyperparameters. After optimization, the testing part is used to calculate the out-of-sample mean square error and the coefficient of determination. The results are provided in Table 5.

The boosting models have a much improved out-of-sample performance according to both metrics, and in particular, the coefficient of determination is significantly better for the boosting models. The results indicate that significant nonlinearities and interactions are present in the data, as the boosting model is the only model capable of modeling these and has much higher explanative power.

4.2 Gender equality

We proceed by building a gradient-boosting model for *gender equality*. Figure 1 and Table 6 provide the results from the model. We use SHAP values to evaluate the importance of each variable and the nature of the association between the variables. The domain of each variable is divided into four intervals, and the mean effect with statistical significance for the intervals is estimated separately for every variable using bootstrapping with 1,000 samples. We use 10% limits to indicate statistical significance because the condition of an interval being different from zero is much stronger than the mere significance of a linear trend coefficient. The results reveal that almost all firm characteristic controls have a statistically significant effect on at least one of the intervals, the only exceptions being *R&D intensity* and *ROA*. Three board characteristics have a statistically significant effect on at least one of the intervals. These characteristics, in order of importance, are *BGD*, *number of board meetings* and *nonexecutive members*.

Table 5 Model validation results

Model	*Coefficient of determination*	*Mean square error*
Gender equality model		
OLS	0.044	0.422
LASSO	0.002	0.442
Gradient boosting	0.135	0.382
Inclusive/diverse model		
OLS	0.057	0.559
LASSO	0.065	0.554
Gradient boosting	0.195	0.477
Attitude toward older colleagues model		
OLS	0.003	0.598
LASSO	0.008	0.592
Gradient boosting	0.162	0.493

Figure 1 Mean effect for the prediction of gender equality, divided into four intervals. The dashed lines indicate 90% confidence intervals

There is a strong positive association between *BGD* and company gender equality. The effect on the prediction for the companies in the first interval (approximately 0%–12%) is negative and statistically significant. Furthermore, the effect in the third interval (approximately 25%–37%) is positive and also statistically significant. The positive effect for the last interval is as strong as for the third interval. However, the effect is no longer statistically significant due to wider confidence intervals. According to the results, companies with *BGD* of approximately 12%–25% do not have a negative or positive effect on company gender equality.

Table 6 also includes the maximum effect difference for the intervals. *BGD* is the most important board characteristic and the fifth most important variable overall to predict company gender equality. The maximum positive effect of the variable on the average grade of company gender equality is 0.08. The effect is significant, as the most important predictor, *leverage*, is only slightly more important with a maximum effect of 0.157. Overall, the firm characteristics are slightly more important predictors of company gender equality than the board characteristics.

4.3 Inclusive/diverse

Figure 2 and Table 7 provide significance analysis for the *inclusive/diverse* prediction model. Four firm control variables, *total assets*, *leverage*, *R&D intensity* and *Tobin's Q*, have

Table 6 Mean effect of each variable on the prediction of gender equality, divided into four intervals					
Variable	*First interval*	*Second interval*	*Third interval*	*Fourth interval*	*Maximum effect difference*
Leverage	−0.012	0.005	−0.130*	−0.151*	0.157
Implied vol.	−0.066*	0.008	0.034	0.014	0.100
Tobin's Q	−0.020*	0.068*	0.050	0.049	0.087
Total assets (1,000$)	−0.005*	0.077*	0.075*	0.075*	0.082
BGD	−0.050*	0.006	0.030*	0.030	0.080
ROA	−0.062	−0.010	0.010	−0.008	0.072
Number of meetings	0.009*	−0.001	−0.036*	−0.058*	0.067
Specific skills	−0.046	0.001	0.011	−0.008	0.058
Nonexecutive members	−0.050*	−0.004	0.000	0.007	0.057
Firm age	0.027	0.003	−0.010	−0.025	0.052
R&D intensity	−0.004	0.037	0.038	0.013	0.042
Member affiliations	−0.012	0.009	0.000	−0.032	0.041
Executive members gender diversity	−0.012	0.008	0.010	0.019	0.031
Independent members	0.019	−0.008	0.002	0.000	0.027
Strictly independent members	0.014	0.004	0.006	−0.002	0.017
Meeting attendance	0.003	0.000	−0.003	−0.014	0.017
Average tenure	−0.006	−0.002	0.009	0.006	0.015
Board size	−0.002	−0.001	0.005	−0.007	0.012
Member compensation	−0.007	0.003	−0.003	−0.008	0.010
Chairman is ex-CEO	−0.004			0.001	0.005

Notes: The last column has the maximum effect difference for the intervals. The asterisks indicate significance at least at the 10% level

a statistically significant effect on at least one of the intervals. Furthermore, four board characteristics, *number of meetings*, *BGD*, *executive members gender diversity* and *specific skills*, have significant effects.

BGD has a strong positive association with the inclusive/diverse culture of a company. The effect is negative and statistically significant for the first interval (0%–12%), while for companies with high BGD, the last two intervals (25%–50%), the effect is positive and statistically significant. Again, for those companies with mediocre BGD, the effect is neither negative nor positive.

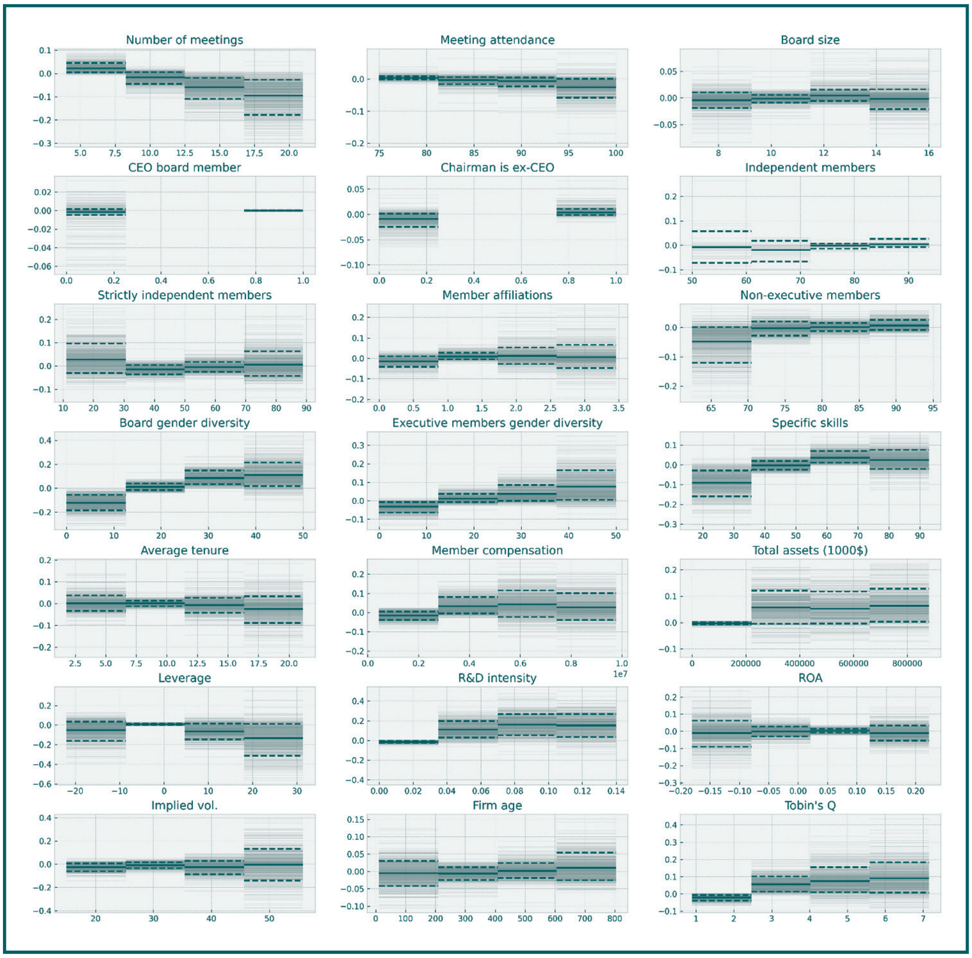
The last column of Table 8 provides the importance analysis of the *inclusive/diverse* model. The strong role of BGD for a company's inclusive/diverse culture is apparent in the results. It is the most important predictor with a maximum effect difference of 0.231, followed by *R&D intensity* with a maximum effect of 0.175. According to the results, the board characteristics are more significant predictors of the *inclusive/diverse* variable than the *gender equality* variable, compared to the firm characteristic controls. The curious negative association of the variable *number of meetings* can be explained by the fact that, usually, companies in crisis have more meetings (Vafeas, 1999). As crisis companies tend to decrease investments for the future and firm diversity can be considered as such, it is expected that crisis companies will lag behind in firm diversity.

4.4 Attitude toward older colleagues

We proceed with the model for the variable *attitude toward older colleagues*. Figure 3 and Table 8 provide the significance analysis for the model. The role of firm characteristics as predictors is more pronounced when compared to the two previous models. All seven variables have a statistically significant effect on at least one of the intervals. Four board characteristics have significant effects: *number of meetings*, *member affiliations*, *nonexecutive members* and *specific skills*.

The effect of *BGD* on the prediction of age diversity in companies is much less pronounced than for the two previous models. There are no statistically significant effects at any of the

Figure 2 Mean effect for the prediction of inclusive/diverse, divided into four intervals. The dashed lines indicate 90% confidence intervals



intervals, and if anything, the association with age diversity is negative. Thus, although BGD is an important predictor of many forms of diversity in companies, it appears that age diversity is not one of them.

The last column of Table 8 has the importance estimates for the *attitude toward older colleagues* model. As the previous results have already indicated, *BGD* is a much less important predictor of age diversity in companies than the other two forms of diversity. It is now the 12th most important variable, with a maximum effect difference of 0.046. Of the ten most important predictors, six are firm characteristics, with *implied volatility*, *leverage* and *ROA* being the three most important. Thus, the firm characteristics are more significant predictors of age diversity in companies than the board characteristics.

4.5 Additional test

The existing research has tried to find the mechanisms and channels for how BGD affects organizations (see, for example, Bear *et al.*, 2010; Dezsö and Ross, 2012; Kor, 2006; Miller and Triana, 2009; Post *et al.*, 2011). Of special interest is the channel through which BGD affects firm value. Thus, we proceed by examining the importance of the company diversity variables in predicting Tobin's Q of a company by constructing a gradient-boosting model

Table 7 Mean effect of each variable on the prediction of inclusive/diverse, divided into four intervals

Variable	*First interval*	*Second interval*	*Third interval*	*Fourth interval*	*Maximum effect difference*
BGD	−0.121*	0.012	0.085*	0.110*	0.231
R&D intensity	−0.015*	0.109*	0.161*	0.152*	0.175
Leverage	−0.052*	0.007*	−0.066	−0.133	0.14
Specific skills	−0.090*	−0.002	0.036*	0.025	0.127
Number of meetings	0.023*	−0.016	−0.059*	−0.095*	0.118
Tobin’s Q	−0.021*	0.056*	0.075*	0.090*	0.111
Executive members gender diversity	−0.032*	0.012	0.037*	0.076*	0.108
Total assets (1,000$)	−0.003	0.058	0.052	0.063*	0.066
Member compensation	−0.015	0.033	0.043	0.027	0.058
Nonexecutive members	−0.048	−0.003	0.002	0.007	0.055
Strictly independent members	0.028	−0.015	−0.005	0.006	0.043
Meeting attendance	0.004	−0.003	−0.007	−0.025	0.029
Member affiliations	−0.015	0.01	0.012	0.007	0.027
Average tenure	0.001	0.001	−0.006	−0.024	0.025
Independent members	−0.008	−0.02	−0.001	0.005	0.025
Implied vol.	−0.026	−0.009	−0.026	−0.006	0.02
Firm age	−0.005	−0.006	0.002	0.011	0.016
ROA	−0.01	−0.001	0.003	−0.01	0.014
Chairman is ex-CEO	−0.009			0.004	0.013
Board size	−0.004	−0.001	0.004	−0.002	0.009

Notes: The last column has the maximum effect difference for the intervals. The asterisks indicate significance at least at the 10% level

Table 8 Mean effect of each variable on the prediction of attitude toward older colleagues, divided into four intervals

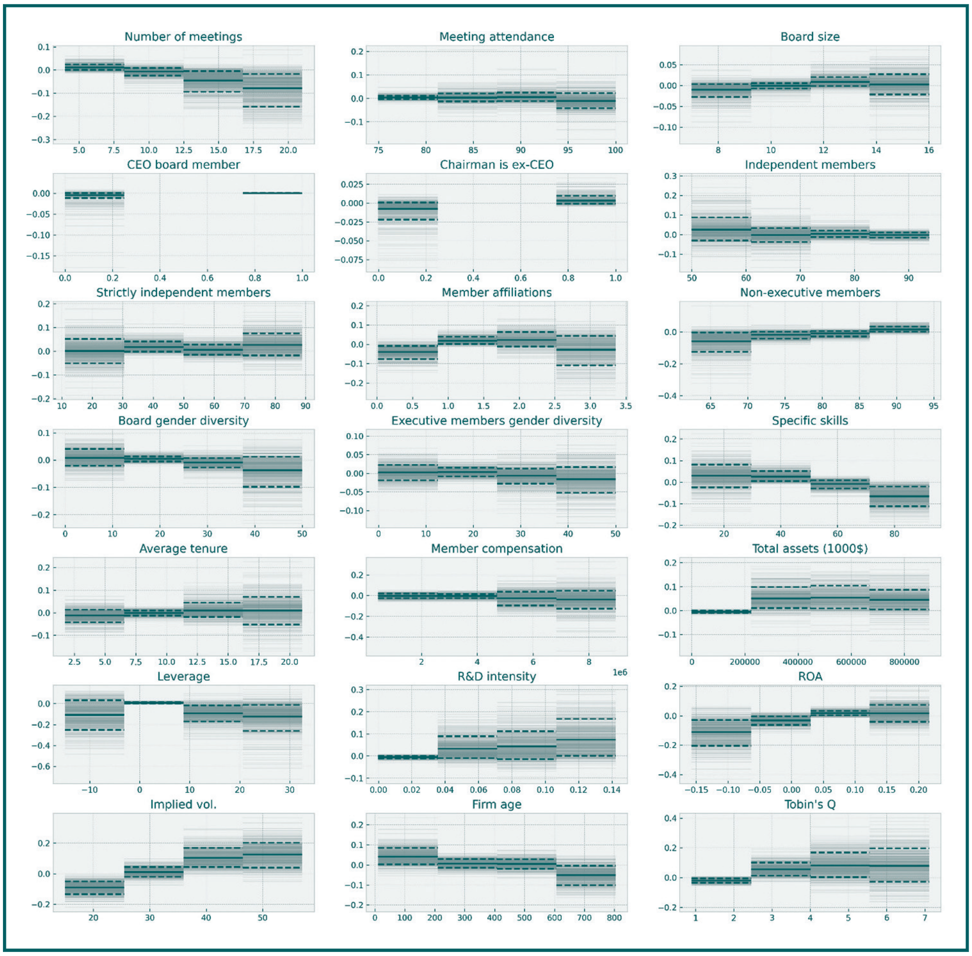
Variable	*First interval*	*Second interval*	*Third interval*	*Fourth interval*	*Maximum effect difference*
Implied vol.	−0.091*	0.011	0.103*	0.125*	0.215
Leverage	−0.108	0.008*	−0.093*	−0.124*	0.132
ROA	−0.111*	−0.032*	0.018*	0.016	0.129
Tobin’s Q	−0.020*	0.056*	0.080*	0.078	0.100
Specific skills	0.030	0.026*	−0.009	−0.066*	0.096
Firm age	0.041*	0.007	0.004	−0.051*	0.092
Number of meetings	0.011*	−0.007	−0.045*	−0.079*	0.09
R&D intensity	−0.005	0.033	0.043	0.073*	0.078
Nonexecutive members	−0.059*	−0.017	−0.009	0.016*	0.075
Member affiliations	−0.040*	0.019*	0.023	−0.027	0.063
Total assets (1,000$)	−0.005*	0.051*	0.054*	0.045*	0.059
BGD	0.008	0.004	−0.009	−0.038	0.046
Member compensation	−0.003	−0.004	−0.025	−0.038	0.035
Independent members	0.025	−0.002	0.004	−0.003	0.028
Strictly independent members	0.000	0.018	0.006	0.027	0.026
Average tenure	−0.012	−0.001	0.010	0.010	0.022
Executive members gender diversity	0.002	0.003	−0.007	−0.015	0.019
Board size	−0.010	−0.001	0.008	0.003	0.018
Meeting attendance	0.004	0.003	0.005	−0.011	0.016
Chairman is ex-CEO	−0.008			0.003	0.011
CEO board member	−0.005			0.000	0.005

Notes: The last column has the maximum effect difference for the intervals. The asterisks indicate significance at least at the 10% level

that uses the firm characteristics, the board characteristics and the three company diversity measures as predictors.

Table 9 provides the importance of the variables for the Tobin’s Q model, where we use the maximum effect difference metric introduced in the previous section to evaluate the variables’ importance. The results verify the importance of company diversity for the market

Figure 3 Mean effect for the prediction of attitude toward older colleagues, divided into four intervals. The dashed lines indicate 90% confidence intervals



performance of a company. For example, *inclusive/diverse* is the eighth most important predictor, right after the firm characteristics and *average tenure.* Moreover, the three company diversity measures are significantly more important predictors than *BGD*. The relative importance of *inclusive/diverse* is almost at the same level as *firm age*, which previous research has shown to have a significant impact on a company's market performance (see, for example, Frijns *et al.*, 2016; Shan *et al.*, 2017). The maximum effect on the prediction for these three variables is between 0.037 and 0.089.

5. Conclusion

In this study, we address how BGD affects different dimensions of diversity culture at the overall company level. For that purpose, we use a unique data set of approximately 250,000 employee reviews from a social media platform aggregated at the yearly level and measure diversity using three employee assessments from the reviews: *gender equality*, *inclusive/diverse values* and *age diversity*. Furthermore, we use tools of explainable AI to measure the relative importance of board and firm characteristics to company diversity and the nature of associations (nonlinearities) these features have.

Our empirical findings demonstrate that BGD is the most crucial board characteristic associated with company gender equality and inclusiveness. Furthermore, from the local

Table 9 Relative importance of the variables for the Tobin's Q model. The importance is measured as the maximum effect difference similarly to the analysis of subsection 4.4

Variable	*Maximum effect difference*
ROA	2.157
R&D intensity	1.147
Leverage	0.333
Implied vol.	0.325
Total assets (1000$)	0.145
Firm age	0.125
Average tenure	0.093
Inclusive/diverse	0.089
Nonexecutive members	0.084
Number of meetings	0.058
Gender equality	0.053
Member compensation	0.050
Member affiliations	0.039
Attitude toward older colleagues	0.037
Specific skills	0.026
Board size	0.024
Independent members	0.023
Meeting attendance	0.021
BGD	0.021
Executive members gender diversity	0.021
SIC 1-digit	0.020
Chairman is ex-CEO	0.014
Strictly independent members	0.013
CEO board member	0.004

importance perspective, the association between BGD and these company diversity measures is strongly positive and statistically significant. However, BGD is not positively associated with all types of diversity. For the age diversity variable, the role of BGD is much weaker, and this weak association is observed to be negative. Thus, our research contributes to the earlier research by showing that BGD is not positively associated with all types of diversity, which supports the view of signaling theory (Bear *et al.*, 2010; Connelly *et al.*, 2011). The improved company reputation affects companies' recruitment process and creates a positive causal link from BGD to some aspects of company diversity, such as gender diversity, but not necessarily to some others, like age diversity. Therefore, our results indicate that focus on BGD might come at the expense of some types of diversity as companies with high BGD lag behind in age diversity principles. This is an interesting finding, and more research is needed to clarify which dimensions of diversity benefit from high BGD and which do not.

Another possible mechanism connecting BGD and different dimensions of firm diversity could be new policies initiated by a new board with stronger BGD (Triana *et al.*, 2014). This is supported by the finding that BGD is more strongly associated with firm diversity than executive diversity, indicating that it is new policies that are working as a driving factor for improved company diversity and top management works as an implementor of these policies, regardless of their background (Kirsch, 2018). However, one could assume that a board implementing new diversity policies would include all dimensions of diversity in these policies. Therefore, our results tend to favor the view of the signaling theory, but more research is needed to clarify this issue. With the exception of BGD, firm characteristics are generally more important predictors of company diversity than board characteristics.

Our ML approach also reveals nonlinearities between BGD and the different dimensions of workplace diversity. The results are interesting in light of the critical mass theory. Previous research has documented a U-shaped association between BGD and firm performance, where a negative association turns positive after a critical mass of about 30% of women on board (Joecks *et al.*, 2013. However, our results indicate different associations between BGD and firm diversity. There is an inverted U-shaped association between BGD and the variable *gender equality*. The association is positive until about 30% of board members are women. After this point, there is no improvement, and the association is slightly negative. Although for the variable *inclusive/diverse*, the association is positive for the whole interval, a similar structure can be seen where the positive effect of BGD starts to decrease after around the 30% point. As discussed, the association is negative and somewhat linear for the age diversity variable.

We also document a possible mechanism of how BGD effects are channeled through the company. Specifically, we analyze the importance of our three diversity measures as predictors of firm value. Previous research has conflicting results on the role of diversity in company performance (Jayne and Dipboye, 2004), and our research reveals new information about the mechanisms that might work in the background. The diversity measures play a significant role in predicting Tobin's Q, and the results demonstrate one viable path to how BGD could affect market performance. They show a possible mediating role of company diversity for the association between BGD and financial performance. Further studies could investigate more of these possible mechanisms on how the effects of BGD are channeled to positive business outcomes through its positive influence on company diversity.

Although our ML approach has many benefits compared to traditional econometric models, it still shares many limitations. For example, although we include an extensive list of firm and board characteristics in our model, we cannot rule out endogeneity caused by omitted variable bias. Furthermore, similarly to traditional econometric models, our analysis cannot give answers about causality. This research question needs field experiments for good answers (Guiso *et al.*, 2015). A mere instrumental variables model or a reverse prediction model cannot reliably identify issues like reverse causality, and neither can our ML model. Therefore, our findings do not confirm that board diversity causes company diversity. For that, we rely on the findings of previous research (Baker *et al.*, 2020; Kirsch, 2018). However, the first crucial step is to understand the potential sources of this link and show that it appears to be present in the data. Our research focuses on finding connections between the examined variables and estimating their relative importance. In this study, we use ML primarily as a descriptive tool that can expose patterns in the data without hypothesizing a particular theory (Delen and Zolbanin, 2018). Thus, our research reveals important findings that can guide researchers in identifying key variables that should be included in theoretical models in future research (Bertomeu, 2020).

References

Adams, R.B. and Ferreira, D. (2009), "Women in the boardroom and their impact on governance and performance", *Journal of Financial Economics*, Vol. 94 No. 2, pp. 291-309, doi: 10.1016/j.jfineco.2008.10.007.

Ahern, K.R. and Dittmar, A.K. (2012), "The changing of the boards: the impact on firm valuation of mandated female board representation", *The Quarterly Journal of Economics*, Vol. 127 No. 1, pp. 137-197.

Ali, F., Wang, M., Jebran, K. and Ali, S.T. (2021), "Board diversity and firm efficiency: evidence from China", *Corporate Governance: The International Journal of Business in Society*, Vol. 21 No. 4, pp. 587-607, doi: 10.1108/CG-10-2019-0312.

Anderson, R.C., Reeb, D.M., Upadhyay, A. and Zhao, W. (2011), "The economics of director heterogeneity", *Financial Management*, Vol. 40 No. 1, pp. 5-38, doi: 10.1111/j.1755-053X.2010.01133.x.

Baker, H.K., Pandey, N., Kumar, S. and Haldar, A. (2020), "A bibliometric analysis of board diversity: current status, development, and future research directions", *Journal of Business Research*, Vol. 108, pp. 232-246, doi: 10.1016/j.jbusres.2019.11.025.

Bao, Y., Ke, B., Li, B., Yu, Y.J. and Zhang, J. (2020), “Detecting accounting fraud in publicly traded US firms using a machine learning approach”, *Journal of Accounting Research*, Vol. 58 No. 1, pp. 199-235, doi: 10.1111/1475-679X.12292.

Barboza, F., Kimura, H. and Altman, E. (2017), “Machine learning models and bankruptcy prediction”, *Expert Systems with Applications*, Vol. 83, pp. 405-417, doi: 10.1016/j.eswa.2017.04.006.

Bartov, E., Faurel, L. and Mohanram, P.S. (2018), “Can Twitter help predict firm-level earnings and stock returns?”, *The Accounting Review*, Vol. 93 No. 3, pp. 25-57, doi: 10.2308/accr-51865.

Barua, A., Davidson, L.F., Rama, D.V. and Thiruvadi, S. (2010), “CFO gender and accruals quality”, *Accounting Horizons*, Vol. 24 No. 1, pp. 25-39, doi: 10.2308/acch.2010.24.1.25.

Bear, S., Rahman, N. and Post, C. (2010), “The impact of board diversity and gender composition on corporate social responsibility and firm reputation”, *Journal of Business Ethics*, Vol. 97 No. 2, pp. 207-221, doi: 10.1007/s10551-010-0505-2.

Ben Saad, S. and Belkacem, L. (2022), “Does board gender diversity affect capital structure decisions?”, *Corporate Governance*, Vol. 22 No. 5, pp. 922-946.

Bertomeu, J. (2020), “Machine learning improves accounting: discussion, implementation and research opportunities”, *Review of Accounting Studies*, Vol. 25 No. 3, pp. 1135-1155, doi: 10.1007/s11142-020-09554-9.

Bilimoria, D. (2006), “The relationship between women corporate directors and women corporate officers”, *Journal of Managerial Issues*, Vol. 18 No. 1, pp. 47-61.

Booth-Bell, D. (2018), “Social capital as a new board diversity rationale for enhanced corporate governance”, *Corporate Governance: The International Journal of Business in Society*, Vol. 18 No. 3, pp. 425-439, doi: 10.1108/CG-02-2017-0035.

Calabrese, G.G. and Manello, A. (2021), “Board diversity and performance in a masculine, aged and global supply chain: new empirical evidence”, *Corporate Governance: The International Journal of Business in Society*, Vol. 21 No. 7, pp. 1440-1459, doi: 10.1108/CG-09-2020-0417.

Campbell, K. and Mínguez-Vera, A. (2008), “Gender diversity in the boardroom and firm financial performance”, *Journal of Business Ethics*, Vol. 83 No. 3, pp. 435-451, doi: 10.1007/s10551-007-9630-y.

Carmona, P., Climent, F. and Momparler, A. (2019), “Predicting failure in the US banking sector: an extreme gradient boosting approach”, *International Review of Economics & Finance*, Vol. 61, pp. 304-323, doi: 10.1016/j.iref.2018.03.008.

Carter, D.A., Simkins, B.J. and Simpson, W.G. (2003), “Corporate governance, board diversity, and firm value”, *The Financial Review*, Vol. 38 No. 1, pp. 33-53, doi: 10.1111/1540-6288.00034.

Chang, Y.-C., Chang, K.-H. and Wu, G.-J. (2018), “Application of extreme gradient boosting trees in the construction of credit risk assessment models for financial institutions”, *Applied Soft Computing*, Vol. 73, pp. 914-920, doi: 10.1016/j.asoc.2018.09.029.

Chapple, L. and Humphrey, J.E. (2014), “Does board gender diversity have a financial impact? Evidence using stock portfolio performance”, *Journal of Business Ethics*, Vol. 122 No. 4, pp. 709-723, doi: 10.1007/s10551-013-1785-0.

Chatjuthamard, P., Jiraporn, P., Lee, S.M., Uyar, A. and Kilic, M. (2021), “Does board gender diversity matter? Evidence from hostile takeover vulnerability”, *Corporate Governance: The International Journal of Business in Society*, Vol. 21 No. 5, pp. 845-864, doi: 10.1108/CG-08-2020-0353.

Chen, T. and Guestrin, C. (2016), “XGBoost: a scalable tree boosting system”, ArXiv:1603.02754 [Cs], doi: 10.1145/2939672.2939785

Climent, F., Momparler, A. and Carmona, P. (2019), “Anticipating bank distress in the Eurozone: an extreme gradient boosting approach”, *Journal of Business Research*, Vol. 101, pp. 885-896, doi: 10.1016/j.jbusres.2018.11.015.

Connelly, B.L., Certo, S.T., Ireland, R.D. and Reutzel, C.R. (2011), “Signaling theory: a review and assessment”, *Journal of Management*, Vol. 37 No. 1, pp. 39-67, doi: 10.1177/0149206310388419.

Cook, A. and Glass, C. (2014), “Women and top leadership positions: towards an institutional analysis”, *Gender, Work & Organization*, Vol. 21 No. 1, pp. 91-103, doi: 10.1111/gwao.12018.

Cook, A. and Glass, C. (2015), “Diversity begets diversity? The effects of board composition on the appointment and success of women CEOs”, *Social Science Research*, Vol. 53, pp. 137-147, doi: 10.1016/j.ssresearch.2015.05.009.

Cortés, E.A., Martínez, M.G. and Rubio, N.G. (2008), "FIAMM return persistence analysis and the determinants of the fees charged", *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad*, Vol. 37 No. 137, pp. 13-32, doi: 10.1080/02102412.2008.10779637.

Darmadi, S. (2013), "Do women in top management affect firm performance? Evidence from Indonesia", *Corporate Governance: The International Journal of Business in Society*, Vol. 13 No. 3, pp. 288-304, doi: 10.1108/CG-12-2010-0096.

Delen, D. and Zolbanin, H.M. (2018), "The analytics paradigm in business research", *Journal of Business Research*, Vol. 90, pp. 186-195, doi: 10.1016/j.jbusres.2018.05.013.

Dezsö, C.L. and Ross, D.G. (2012), "Does female representation in top management improve firm performance? A panel data investigation", *Strategic Management Journal*, Vol. 33 No. 9, pp. 1072-1089, doi: 10.1002/smj.1955.

Doornenbal, B.M., Spisak, B.R. and van der Laken, P.A. (2021), "Opening the black box: uncovering the leader trait paradigm through machine learning", *The Leadership Quarterly*, Vol. 101515, doi: 10.1016/j.leaqua.2021.101515.

Efron, B. (1979), "Bootstrap methods: another look at the jackknife", *The Annals of Statistics*, Vol. 7 No. 1, pp. 1-26, doi: 10.1214/aos/1176344552.

Erhardt, N.L., Werbel, J.D. and Shrader, C.B. (2003), "Board of director diversity and firm financial performance", *Corporate Governance*, Vol. 11 No. 2, pp. 102-111, doi: 10.1111/1467-8683.00011.

Faccio, M., Marchica, M.-T. and Mura, R. (2016), "CEO gender, corporate risk-taking, and the efficiency of capital allocation", *Journal of Corporate Finance*, Vol. 39, pp. 193-209, doi: 10.1016/j.jcorpfin.2016.02.008.

Filbeck, G., Foster, B., Preece, D. and Zhao, X. (2017), "Does diversity improve profits and shareholder returns? Evidence from top rated companies for diversity by DiversityInc", *Advances in Accounting*, Vol. 37, pp. 94-102, doi: 10.1016/j.adiac.2017.02.001.

Friedman, J.H. (2001), "Greedy function approximation: a gradient boosting machine", *The Annals of Statistics*, Vol. 29 No. 5, pp. 1189-1232.

Friedman, J.H. (2002), "Stochastic gradient boosting", *Computational Statistics & Data Analysis*, Vol. 38 No. 4, pp. 367-378, doi: 10.1016/S0167-9473(01)00065-2.

Frijns, B., Dodd, O. and Cimerova, H. (2016), "The impact of cultural diversity in corporate boards on firm performance", *Journal of Corporate Finance*, Vol. 41, pp. 521-541, doi: 10.1016/j.jcorpfin.2016.07.014.

Futagami, K., Fukazawa, Y., Kapoor, N. and Kito, T. (2021), "Pairwise acquisition prediction with SHAP value interpretation", *The Journal of Finance and Data Science*, Vol. 7, pp. 22-44, doi: 10.1016/j.jfds.2021.02.001.

Green, T.C., Huang, R., Wen, Q. and Zhou, D. (2019), "Crowdsourced employer reviews and stock returns", *Journal of Financial Economics*, Vol. 134 No. 1, pp. 236-251, doi: 10.1016/j.jfineco.2019.03.012.

Gregory, B. (2018), "Predicting customer churn: extreme gradient boosting with temporal data", ArXiv:1802.03396 [Cs, Stat], available at: http://arxiv.org/abs/1802.03396

Gu, S., Kelly, B. and Xiu, D. (2020), "Empirical asset pricing via machine learning", *The Review of Financial Studies*, Vol. 33 No. 5, pp. 2223-2273, doi: 10.1093/rfs/hhaa009.

Guiso, L., Sapienza, P. and Zingales, L. (2015), "The value of corporate culture", *Journal of Financial Economics*, Vol. 117 No. 1, pp. 60-76, doi: 10.1016/j.jfineco.2014.05.010.

Gul, F.A., Srinidhi, B. and Ng, A.C. (2011), "Does board gender diversity improve the informativeness of stock prices?", *Journal of Accounting and Economics*, Vol. 51 No. 3, pp. 314-338, doi: 10.1016/j.jacceco.2011.01.005.

Hales, J., Moon, J.R. and Swenson, L.A. (2018), "A new era of voluntary disclosure? Empirical evidence on how employee postings on social media relate to future corporate disclosures", *Accounting, Organizations and Society, Vol. 68–69*, Vol. 68-69, pp. 88-108, doi: 10.1016/j.aos.2018.04.004.

Hambrick, D.C. and Mason, P.A. (1984), "Upper echelons: the organization as a reflection of its top managers", *The Academy of Management Review*, Vol. 9 No. 2, pp. 193-206, doi: 10.2307/258434.

Harjoto, M., Laksmana, I. and Lee, R. (2015), "Board diversity and corporate social responsibility", *Journal of Business Ethics*, Vol. 132 No. 4, pp. 641-660, doi: 10.1007/s10551-014-2343-0.

Huang, J. and Kisgen, D.J. (2013), “Gender and corporate finance: are male executives overconfident relative to female executives?”, *Journal of Financial Economics*, Vol. 108 No. 3, pp. 822-839, doi: 10.1016/j.jfineco.2012.12.005.

Hurley, D. and Choudhary, A. (2020), “Role of gender and corporate risk taking”, *Corporate Governance: The International Journal of Business in Society*, Vol. 20 No. 3, pp. 383-399, doi: 10.1108/CG-10-2018-0313.

Jabeur, B.S., Khalfaoui, R. and Ben Arfi, W. (2021), “The effect of green energy, global environmental indexes, and stock markets in predicting oil price crashes: evidence from explainable machine learning”, *Journal of Environmental Management*, Vol. 298, p. 113511, doi: 10.1016/j.jenvman.2021.113511.

Jayne, M.E.A. and Dipboye, R.L. (2004), “Leveraging diversity to improve business performance: research findings and recommendations for organizations”, *Human Resource Management*, Vol. 43 No. 4, pp. 409-424, doi: 10.1002/hrm.20033.

Jiang, Y. and Jones, S. (2018), “Corporate distress prediction in China: a machine learning approach”, *Accounting & Finance*, Vol. 58 No. 4, pp. 1063-1109, doi: 10.1111/acfi.12432.

Joecks, J., Pull, K. and Vetter, K. (2013), “Gender diversity in the boardroom and firm performance: what exactly constitutes a ‘critical mass’?”, *Journal of Business Ethics*, Vol. 118 No. 1, pp. 61-72, doi: 10.1007/s10551-012-1553-6.

Jones, S. (2017), “Corporate bankruptcy prediction: a high dimensional analysis”, *Review of Accounting Studies*, Vol. 22 No. 3, pp. 1366-1422, doi: 10.1007/s11142-017-9407-1.

Jurkus, A.F., Park, J.C. and Woodard, L.S. (2011), “Women in top management and agency costs”, *Journal of Business Research*, Vol. 64 No. 2, pp. 180-186, doi: 10.1016/j.jbusres.2009.12.010.

Karolyi, G.A. and Van Nieuwerburgh, S. (2020), “New methods for the cross-section of returns”, *The Review of Financial Studies*, Vol. 33 No. 5, pp. 1879-1890, doi: 10.1093/rfs/hhaa019.

Khaidem, L., Saha, S. and Dey, S.R. (2016), “Predicting the direction of stock market prices using random Forest”, ArXiv:1605.00003 [Cs], available at: http://arxiv.org/abs/1605.00003

Kirsch, A. (2018), “The gender composition of corporate boards: a review and research agenda”, *The Leadership Quarterly*, Vol. 29 No. 2, pp. 346-364, doi: 10.1016/j.leaqua.2017.06.001.

Kor, Y.Y. (2006), “Direct and interaction effects of top management team and board compositions on R&D investment strategy”, *Strategic Management Journal*, Vol. 27 No. 11, pp. 1081-1099, doi: 10.1002/smj.554.

Krishnan, G.V. and Parsons, L.M. (2008), “Getting to the bottom line: an exploration of gender and earnings quality”, *Journal of Business Ethics*, Vol. 78 Nos 1/2, pp. 65-76, doi: 10.1007/s10551-006-9314-z.

Kumar, M. and Thenmozhi, M. (2014), “Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random Forest hybrid models”, *International Journal of Banking, Accounting and Finance*, Vol. 5 No. 3, pp. 284-308, doi: 10.1504/IJBAAF.2014.064307.

Leyva-Townsend, P., Rodriguez, W., Idrovo, S. and Pulga, F. (2021), “Female board participation and firm’s financial performance: a panel study from a Latin American economy”, *Corporate Governance: The International Journal of Business in Society*, Vol. 21 No. 5, pp. 920-938, doi: 10.1108/CG-07-2019-0235.

Lin, B. and Bai, R. (2021), “Machine learning approaches for explaining determinants of the debt financing in heavy-polluting enterprises”, *Finance Research Letters*, Vol. 44, doi: 10.1016/j.frl.2021.102094.

Liu, C., Chan, Y., Kazmi, S.H.A. and Fu, H. (2015), “Financial fraud detection model: based on random Forest”, *International Journal of Economics and Finance*, Vol. 7 No. 7, p. 178, doi: 10.5539/ijef.v7n7p178.

Loukil, N., Yousfi, O. and Yerbanga, R. (2019), “Does gender diversity on boards influence stock market liquidity? Empirical evidence from the French market”, *Corporate Governance: The International Journal of Business in Society*, Vol. 19 No. 4, pp. 669-703, doi: 10.1108/CG-09-2018-0291.

Lundberg, S.M. and Lee, S.-I. (2017), “A unified approach to interpreting model predictions”, in Guyon, I. Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S. and Garnett, R. (Eds), *Advances in Neural Information Processing Systems*, Vol. 30, pp. 4765-4774, available at: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N. and Lee, S.-I. (2020), "From local explanations to global understanding with explainable AI for trees", *Nature Machine Intelligence*, Vol. 2 No. 1, pp. 56-67, doi: 10.1038/s42256-019-0138-9.

Matsa, D.A. and Miller, A.R. (2011), "Chipping away at the glass ceiling: gender spillovers in corporate leadership", *American Economic Review*, Vol. 101 No. 3, pp. 635-639, doi: 10.1257/aer.101.3.635.

Mazzotta, R. and Ferraro, O. (2020), "Does the gender quota law affect bank performances? Evidence from Italy", *Corporate Governance: The International Journal of Business in Society*, Vol. 20 No. 6, pp. 1135-1158, doi: 10.1108/CG-08-2019-0252.

Menicucci, E. and Paolucci, G. (2022), "Gender diversity and bank risk-taking: an empirical investigation in Italy", *Corporate Governance*, Vol. 22 No. 2, pp. 317-339.

Miller, G.S. and Skinner, D.J. (2015), "The evolving disclosure landscape: how changes in technology, the media, and capital markets are affecting disclosure", *Journal of Accounting Research*, Vol. 53 No. 2, pp. 221-239, doi: 10.1111/1475-679X.12075.

Miller, T. and Triana, M.D.C. (2009), "Demographic diversity in the boardroom: mediators of the board diversity–firm performance relationship", *Journal of Management Studies*, Vol. 46 No. 5, pp. 755-786, doi: 10.1111/j.1467-6486.2009.00839.x.

Molnar, C. (2021), "Interpretable machine learning", available at: https://christophm.github.io/interpretable-ml-book

Palvia, A., Vähämaa, E. and Vähämaa, S. (2015), "Are female CEOs and chairwomen more conservative and risk averse? Evidence from the banking industry during the financial crisis", *Journal of Business Ethics*, Vol. 131 No. 3, pp. 577-594, doi: 10.1007/s10551-014-2288-3.

Pierdzioch, C., Risse, M. and Rohloff, S. (2015), "Forecasting gold-price fluctuations: a real-time boosting approach", *Applied Economics Letters*, Vol. 22 No. 1, pp. 46-50, doi: 10.1080/13504851.2014.925040.

Post, C., Rahman, N. and Rubow, E. (2011), "Green governance: boards of directors' composition and environmental corporate social responsibility", *Business & Society*, Vol. 50 No. 1, pp. 189-223, doi: 10.1177/0007650310394642.

Rao, K.K. and Tilt, C. (2020), "Gender and CSR decisions: perspectives from Australian boards", *Meditari Accountancy Research*, Vol. 29 No. 1, pp. 60-85, doi: 10.1108/MEDAR-11-2019-0609.

Rose, C. (2007), "Does female board representation influence firm performance? The Danish evidence", *Corporate Governance: An International Review*, Vol. 15 No. 2, pp. 404-413, doi: 10.1111/j.1467-8683.2007.00570.x.

Saleh, M.W.A., Zaid, M.A.A., Shurafa, R., Maigoshi, Z.S., Mansour, M. and Zaid, A. (2021), "Does board gender enhance Palestinian firm performance? The moderating role of corporate social responsibility", *Corporate Governance: The International Journal of Business in Society*, Vol. 1No No. 4, pp. 685-701, doi: 10.1108/CG-08-2020-0325.

Shan, L., Fu, S. and Zheng, L. (2017), "Corporate sexual equality and firm performance", *Strategic Management Journal*, Vol. 38 No. 9, pp. 1812-1826, doi: 10.1002/smj.2624.

Skaggs, S., Stainback, K. and Duncan, P. (2012), "Shaking things up or business as usual? The influence of female corporate executives and board of directors on women's managerial representation", *Social Science Research*, Vol. 41 No. 4, pp. 936-948, doi: 10.1016/j.ssresearch.2012.01.006.

Spisak, B.R., van der Laken, P.A. and Doornenbal, B.M. (2019), "Finding the right fuel for the analytical engine: expanding the leader trait paradigm through machine learning?", *The Leadership Quarterly*, Vol. 30 No. 4, pp. 417-426, doi: 10.1016/j.leaqua.2019.05.005.

Triana, M., del, C., Miller, T.L. and Trzebiatowski, T. (2014), "The double-edged nature of board gender diversity: diversity, firm performance, and the power of women directors as predictors of strategic change", *Organization Science*, Vol. 25 No. 2, pp. 609-632, doi: 10.1287/orsc.2013.0842.

Vafeas, N. (1999), "Board meeting frequency and firm performance", *Journal of Financial Economics*, Vol. 53 No. 1, pp. 113-142, doi: 10.1016/S0304-405X(99)00018-5.

Ylinen, M. and Ranta, M. (2021), "Employee-friendly corporate culture and firm performance: evidence from a machine learning approach (March 26, 2021)", available at SSRN: https://ssrn.com/abstract=3813075 or http://dx.doi.org/10.2139/ssrn.3813075

Yousaf, U.B., Jebran, K. and Wang, M. (2021), “Can board diversity predict the risk of financial distress?”, *Corporate Governance: The International Journal of Business in Society*, Vol. 21 No. 4, pp. 663-684, doi: 10.1108/CG-06-2020-0252.

Zhang, X. and Chen, W. (2019), “Stock selection based on extreme gradient boosting”, *2019 Chinese Control Conference (CCC)*, pp. 8926-8931, doi: 10.23919/ChiCC.2019.8865781.

Zheng, B. (2019), “Financial default payment predictions using a hybrid of simulated annealing heuristics and extreme gradient boosting machines”, *International Journal of Internet Technology and Secured Transactions*, Vol. 9 No. 4, pp. 404-425, doi: 10.1504/IJITST.2019.102796.

Zie¸ba, M., Tomczak, S.K. and Tomczak, J.M. (2016), “Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction”, *Expert Systems with Applications*, Vol. 58, pp. 93-101, doi: 10.1016/j.eswa.2016.04.001.

Further reading

Francoeur, C., Labelle, R. and Sinclair-Desgagné, B. (2008), “Gender diversity in corporate governance and top management”, *Journal of Business Ethics*, Vol. 81 No. 1, pp. 83-95, doi: 10.1007/s10551-007-9482-5.

Zięba, M., Tomczak, S.K. and Tomczak, J.M. (2016), “Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction”, *Expert Systems with Applications*, Vol. 58, pp. 93-101, doi: 10.1016/j.eswa.2016.04.001.

Appendix

Table A1 Survey questions

Variable	*Question*
Gender equality	Are women treated equally and given the same career opportunities?
Attitude toward older colleagues	Does the company hire older workers? Are senior colleagues appreciated, supported and given equal opportunities?
Inclusive/diverse	To what extent does the company value diversity in the workplace? Are diverse ideas and opinions supported?

Corresponding author

Mikko Ranta can be contacted at: mran@uwasa.fi