Ekonomi, Politika & Finans Araştırmaları Dergisi, 2025, 10(3): 949-970
Journal of Research in Economics, Politics & Finance, 2025, 10(3): 949-970
Araştırma Makalesi / Research Article, https://doi.org/10.30784/epfad.1651693

# THE ROLE OF FINANCIAL INDICATORS IN THE PREDICTION OF VOLUNTARY CARBON DISCLOSURE: A COMPARATIVE ANALYSIS WITH MACHINE LEARNING METHODS

## Gönüllü Karbon Açıklaması Tahmininde Finansal Göstergelerin Rolü: Makine Öğrenmesi Yöntemleri ile Karşılaştırmalı Bir Analiz

**Yunus Emre AKDOĞAN***

**Keywords:** Climate Change, Voluntary Carbon Disclosure, Machine Learning, Explainable Artificial Intelligence, Borsa İstanbul.

**JEL Codes:** Q54, C45, C53, C55, G30, G32.

**Abstract**

Since the Industrial Revolution, carbon dioxide emissions and deforestation have been considered the primary causes of climate change. Many countries are developing policies to reduce greenhouse gas emissions and are encouraging firms to disclose and reduce their carbon emissions. This study aims to identify the potential financial determinants of carbon risk awareness, as measured by the willingness to respond to the CDP (Carbon Disclosure Project) survey, among firms listed on the Borsa Istanbul between 2016 and 2023, using machine learning methods. The findings reveal that whether firms will make voluntary carbon disclosures can be predicted with an accuracy rate exceeding 92% using nonlinear, ensemble learning-based Random Forest and XGBoost algorithms in models based on financial indicators. Furthermore, analyses conducted with explainable artificial intelligence tools indicate that specific financial ratios, such as the ratio of equity to total debt, the ratio of fixed assets to equity, and the ratio of long-term debt to total debt, significantly enhance the model's explainability within the XGBoost algorithm. Finally, the study highlights the potential of machine learning algorithms to improve investors' risk analysis in predicting corporate carbon emissions and demonstrates that this finding contributes to both the theoretical and practical development of sustainable investment strategies.

**Anahtar Kelimeler:** İklim Değişikliği, Gönüllü Karbon Saydamlığı, Makine Öğrenmesi, Açıklanabilir Yapay Zekâ, Borsa İstanbul.

**JEL Kodları**: Q54, C45, C53, C55, G30, G32.

**Öz**

Sanayi Devriminden bu yana atmosferdeki karbondioksit emisyonları ve ormansızlaşmanın iklim değişikliğinin başlıca nedenleri olduğu düşünülmektedir. Birçok ülke sera gazı emisyonlarını azaltmak için politikalar geliştirmekte ve firmaları karbon emisyonlarını açıklamaya ve azaltmaya teşvik etmektedir. Bu çalışma, makine öğrenmesi yöntemlerini kullanarak, 2016-2023 yılları arasında Borsa İstanbul'da işlem gören firmaların CDP (Carbon Disclosure Project) anketine yanıt verme istekliliği ile ölçülen karbon riski farkındalığının potansiyel finansal belirleyicilerini ortaya çıkarmayı amaçlamaktadır. Çalışmanın bulguları, firmaların finansal göstergelerine dayalı modeller aracılığıyla, doğrusal olmayan, topluluk öğrenmesi tabanlı Rastgele Orman ve XGBoost algoritmaları kullanılarak, gönüllü karbon açıklaması yapma eğilimlerinin %92'nin üzerinde bir doğruluk oranıyla tahmin edilebildiğini ortaya koymaktadır. Ayrıca açıklanabilir yapay zekâ araçları kullanılarak yapılan analizler, özkaynakların toplam borçlara oranı, duran varlıkların özkaynaklara oranı ve uzun vadeli borçların toplam borçlara oranı gibi belirli finansal oranların, XGBoost algoritmasında modelin açıklayıcılığına önemli düzeyde katkı sağladığını göstermektedir. Son olarak, çalışma, makine öğrenmesi algoritmalarının kurumsal karbon emisyonlarının tahmininde yatırımcıların risk analizini iyileştirme potansiyeline ve bu bulgunun sürdürülebilir yatırım stratejilerinin hem kuramsal hem de uygulamalı olarak geliştirilmesine katkı sunabileceğine dikkat çekmektedir.

* Assist. Prof. Dr., Yozgat Bozok University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Türkiye, emre.akdogan@bozok.edu.tr

*Received Date (Makale Geliş Tarihi): 05.03.2025 Accepted Date (Makale Kabul Tarihi): 04.07.2025*

## 1. Introduction

Climate change is defined as long-term changes in average weather patterns that describe the Earth's local, regional and global climates (López-Pacheco et al., 2021; NASA, 2024). These changes lead to temperature increases in land and oceans, rising water levels in the seas and melting of polar ice caps, and provide evidence of key indicators of climate change, such as changes in the frequency and severity of extreme weather conditions such as hurricanes, wildfires, droughts and floods (Costa de Oliveira et al., 2014; Arendt et al., 2021).

Carbon dioxide emissions and deforestation in the atmosphere since the industrial era are thought to be at the root of climate change (Rogoff, 2014; Pires, 2017). Therefore, changes in climatic conditions require institutions and organizations to make new regulations to combat carbon emissions that cause climate change. In this context, in 1992, the United Nations Framework Convention on Climate Change focused on reducing greenhouse gases in the atmosphere and addressed consequences such as degradation of natural resources and global warming. Subsequently, the Kyoto Protocol, which was adopted in 1997 and entered into force in 2005, was implemented as one of the important institutional steps towards climate change, limiting the amount of carbon emissions for developed economies to a certain limit (History of the convention, n.d.). Türkiye became a party to the Kyoto Protocol in 2009 (Kyoto Protocol, n.d.). The Paris Agreement on Climate Change is an international agreement signed in 2015 that aims to keep global warming below 2°C. The agreement requires countries to submit national contribution declarations that include commitments to reduce carbon emissions and to review these commitments every five years. It also encourages the provision of financial and technical support to developing countries (The Paris Agreement, n.d.). Türkiye signed the Paris Agreement on Climate Change on April 22, 2016 and officially became a party on November 10, 2021. As a party to the agreement, Türkiye has set a net class carbon emission target by 2053 (Paris Anlaşması, n.d.).

These significant changes in climate have the potential to drastically alter humanity's way of life and livelihoods (Füssel and Klein, 2006; Pires, 2017). Therefore, climate change risks deeply affect the economy and business environment (Kolk and Pinkse, 2005). In this framework, many countries have developed various policies on reducing greenhouse gas emissions of firms within the framework of the aforementioned regulations and encourage firms to disclose and reduce their carbon emissions (Jung et al., 2008; Harker et al., 2022). In addition, the fact that the awareness of climate change risks has increased among all institutions and investors related to financial markets points to the potential to be an indirect pressure factor for firms (Rohleder et al., 2022). Considering all these, it becomes even more important for firms and regulatory bodies to disclose their sustainability and corporate carbon performance (Kolk and Pinkse, 2005; Jones and Levy, 2007; Velte et al., 2020). This study aims to identify the most important financial indicators in predicting whether firms traded on Borsa Istanbul between 2016 and 2023 will voluntarily report carbon information under CDP (Carbon Disclosure Project). In other words, the aim of this study is to reveal the potential predictive financial variables of carbon risk awareness, measured by the willingness to respond to the CDP survey, using machine learning methods.

The contributions of the study to the literature can be summarized under the following headings: Firstly, the study offers a more sophisticated and data-driven approach by transcending traditional analytical methods through the application of machine learning algorithms (such as

Random Forest and XGBoost) to predict carbon reporting. This significantly enriches the literature by illustrating that innovative methodologies can be effectively utilized in the analysis of environmental reporting. Furthermore, the study systematically enhances the literature concerning the interplay between corporate finance and environmental transparency by establishing those specific financial ratios (Equity to Total Liabilities, Long-Term Liabilities to Total Liabilities, and Fixed Assets to Equity Ratio) may serve as predictive indicators of firms' carbon reporting behaviors. Another notable contribution of this study is the examination of firms listed on Borsa Istanbul, which yields valuable insights into the dynamics of carbon reporting, particularly within emerging markets. Lastly, in terms of investor risk management, the study suggests that carbon reporting represents a significant indicator that can be instrumental in investor risk assessments. This insight will contribute to formulating sustainable investment strategies in both theoretical and practical contexts.

This study organizes the content into four distinct sections. Section 2 reviews the existing literature. Section 3, titled Empirical Analysis, includes five sub-sections: Section 3.1 specifies the dataset employed, Section 3.2 describes the methodological framework adopted, and Section 3.3 clarifies the performance metrics. Section 3.4 presents the empirical findings of the research, whereas Section 3.5 engages in a theoretical discussion regarding these findings. Section 4 contains the concluding remarks of the paper.

## 2. Literature Review

The literature review on climate change risks generally shows that three main issues come to the fore. One of them examines the connection between corporate governance and environmental accountability regarding the disclosure of greenhouse gas emissions. In these studies, gender diversity on corporate boards and board independence positively affect greenhouse gas disclosure practices (Liao et al., 2015); having women directors in shaping corporate environmental strategies promotes sustainable practices and reduces carbon footprints across firms (Fan et. al, 2023); the inclusion of diverse perspectives in management can lead to more comprehensive risk assessments and innovative solutions that traditional leadership may overlook, thereby strengthening decision-making processes (Hollindale et al., 2019); board characteristics and committee structures significantly affect carbon performance and voluntary disclosure of greenhouse gas emissions (Haque, 2017; Krishnamurti and Velayutham, 2018).

Another important area highlighted in the scientific literature is the correlation between carbon emissions and corporate performance. In this context, empirical studies have primarily focused on corporate valuation, stock market prices, and investment portfolio performance. These studies have found that high corporate social responsibility scores increase the negative impact on firm value when expectations are not met (Cooper et al., 2018); reducing emissions intensity by half in carbon-based portfolios does not negatively affect portfolio returns (Anquetin et al., 2022); decarbonizing mutual fund portfolios creates sustained selling pressure on shares of carbon-intensive firms, leading to a permanent decline in share prices (Rohleder et al., 2022); firms with higher carbon emissions are more likely to acquire foreign firms when operating in countries with lower GDP or weaker environmental, regulatory, or governance standards, while their tendency to acquire domestic targets is lower (Bose et al., 2021); carbon emissions show a negative correlation with asset returns and earnings per share (Güneysu and Atasel, 2022); environmental performance information is associated with analysts' information processing costs

(Griffin et al., 2020); the market tends to punish negative environmental performance more consistently than rewarding positive performance, particularly in the context of R&D investments (Lee et al., 2015); corporate social responsibility disclosures in the FinTech sector have a positive impact on the market value/book value ratio (Merello et al., 2022); green stocks, which represent shares of environmentally friendly and low-carbon firms, outperform black stocks, which represent shares of firms with high carbon emissions and significant environmental risks (Rahat and Nguyen, 2022); firms that tweet about climate issues achieve positive abnormal returns in the short term (Guastella et al., 2022); better carbon performance, reflected in lower carbon emission levels, has a positive impact on the market value of firms, especially those with higher gender diversity and innovation capacity (Benkraiem et al., 2022); carbon emissions have a significant negative impact on accounting measures such as ROA and market-based performance indicators such as Tobin's Q (Desai et al., 2022); carbon disclosure has a significant negative impact on accounting measures such as ROA and market-based performance indicators such as Tobin's Q (Desai et al., 2022); carbon disclosure has a significant negative impact on accounting measures such as ROA and market- have a significant negative impact on accounting metrics such as ROA and market-based performance indicators such as Tobin's Q (Desai et al., 2022); carbon disclosure does not significantly improve the financial performance of high-carbon industries unless implemented alongside actual emission reductions, while increased transparency and strengthened investor confidence provide sustainable financial benefits for low-carbon industries (Lu et al., 2021); institutional investors, exclusionary screening practices that remove firms with high direct emissions from their portfolios as part of their investment strategies to address carbon risk (Bolton and Kacperczyk, 2021).

A literature review reveals that some studies have focused on the relationship between corporate carbon emissions and financing costs. These studies generally aim to determine whether lenders consider climate change risks as a risk factor for loan defaults. These studies suggest that direct carbon emissions have a more negative impact on credit ratings than indirect emissions (Safiullah et al., 2021); firms with high carbon emissions have lower financial leverage ratios, and increased carbon risk contributes to financial distress (Nguyen and Phan, 2020); and that an increase in carbon emissions leads to increased credit spreads (Kleimeier and Viehs, 2021); voluntary carbon disclosure improves financial performance (Alsaifi et al., 2020); carbon emissions negatively affect default risk, and green initiatives reduce this risk (Kabir et al., 2021); firms aware of carbon risk benefit from benefiting from lower borrowing costs (Jung et al., 2018); many European investors actively participating in the green bond market and significant demand for green bonds (Sangiorgi and Schopohl, 2021); climate change scores of environmentally friendly firms being estimable using machine learning models based on scope and credit ratings; voluntary carbon reporting levels can be predicted using machine learning methods by employing financial performance indicators such as leverage, one-year raw return, current ratio, and corporate governance factors such as ownership concentration and the number of board members (Frost et al., 2023).

## 3. Empirical Analysis

### 3.1. Data Description

This section presents the main characteristics of the dataset and provides an overview of the variables included in the analysis.

### 3.1.1. Explained Variable

This study uses data from the Carbon Disclosure Project (CDP), which reports firm-level carbon emissions. CDP surveys large publicly traded firms annually to gather emission data and calculates their climate change scores. These scores can pressure firms to lower emissions and inform investors about firms' environmental performance. The study period was set between 2016 and 2023, as the number of firms responding voluntarily to the CDP climate change questionnaire in Türkiye increased significantly starting from 2016 (see Figure 1: The number of disclosing firms in Türkiye). Prior to 2016, the number of firms disclosing firm-level carbon risk information was quite limited, which could have compromised the representativeness and reliability of the dataset. The 2016–2023 period thus provides a more balanced, comprehensive, and consistent dataset, enhancing the robustness and generalizability of the machine learning analyses conducted in this study.

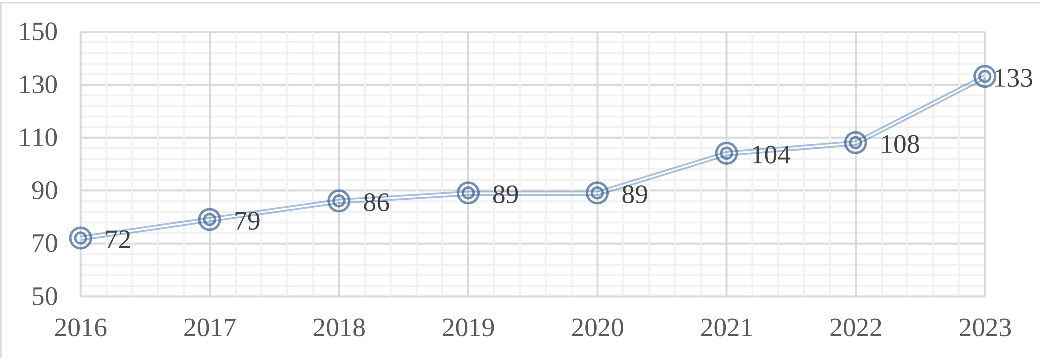


**Figure 1. The Number of Disclosing Firms in Türkiye**

Numerous studies have used the CDP database for carbon emissions research (Matsumura et al., 2014; Liao et al., 2015; Griffin et al., 2017; Jung et al., 2018; Caby et al., 2020; Bose et al., 2021; Kleimeier and Viehs, 2021; Desai et al., 2022; Harker et al., 2022). Following Jung et al. (2018), this study gathered data on the willingness of non-financial firms listed on Borsa Istanbul to participate in carbon emission surveys from 2016 to 2023. This study, following Jung et al. (2018), gathered data on the willingness of non-financial firms listed on Borsa Istanbul to participate in carbon emission surveys from 2016 to 2023. A total of 100 firms participated in the study. Assuming firms that disclose carbon emissions voluntarily have made progress or care about the issue, the dependent variable is defined as their willingness to respond to climate change surveys.

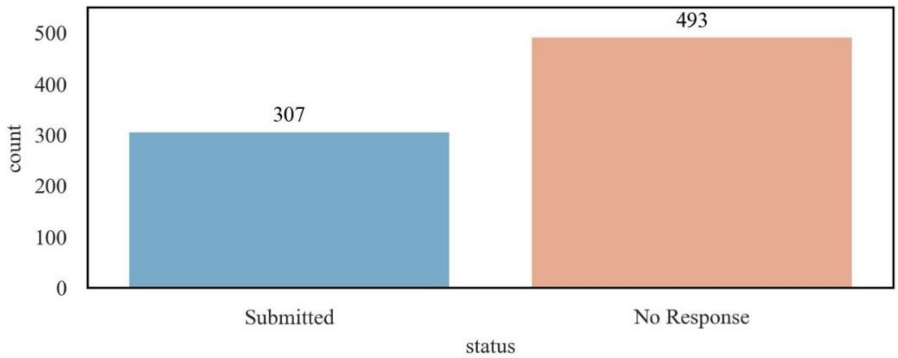


**Figure 2. Distribution of Total Public Disclosure Responses**

As shown in Figure 2, of the total firms considered, 307 submitted disclosure responses, while 493 provided no response. This distribution reflects the varying levels of engagement by firms in carbon disclosure practices.

### 3.1.2. Explanatory Variables

The explanatory variables in the study are financial ratios obtained from the balance sheets and income statements of the firms. 36 different financial ratios and beta score as a risk measure are calculated. The financial ratios calculated within the scope of the analysis are presented in Table 1. Beta score for firms is calculated using three years of data at the daily level.

**Table 1. Categorization of Financial Ratios Used in the Study**

| Variable Name | Category |
|---|---|
| Beta Score | Risk Indicator |
| Current Ratio<br>Acid-Test Ratio<br>Cash Ratio<br>Inventory / Total Assets Ratio<br>Inventory Dependency Ratio<br>Short-Term Receivables / Total Assets Ratio | Liquidity Ratios |
| Total Liabilities / Total Assets Ratio (Leverage Ratio)<br>Equity / Total Assets Ratio<br>Equity / Total Liabilities Ratio<br>Short-Term Liabilities / Total Liabilities Ratio<br>Long-Term Liabilities / Total Liabilities Ratio<br>Tangible Fixed Assets / Long-Term Liabilities Ratio<br>Fixed Assets / Total Liabilities Ratio<br>Fixed Assets / Equity Ratio<br>Short-Term Bank Loans / Short-Term Liabilities Ratio<br>Bank Loans / Total Liabilities Ratio<br>Current Assets / Total Assets Ratio<br>Net Tangible Fixed Assets / Total Assets Ratio | Leverage Ratios |
| Inventory Turnover Ratio<br>Receivables Turnover Ratio<br>Working Capital Turnover Ratio<br>Net Working Capital Turnover Ratio<br>Fixed Assets Turnover Ratio<br>Debt Turnover Ratio<br>Asset Turnover Ratio | Efficiency Ratios |
| EBIT / Total Liabilities Ratio<br>Net Profit / Total Assets Ratio<br>Cumulative Profitability Ratio<br>Operating Profit / Net Sales Ratio<br>Gross Profit / Net Sales Ratio<br>Net Profit / Net Sales Ratio<br>Operating Expenses / Net Sales Ratio<br>Interest Expenses / Net Sales Ratio<br>Profit Before Interest and Tax / Interest Expenses Ratio<br>Net Profit + Interest Expenses / Interest Expenses Ratio | Profitability Ratios |
| Asset Size | Size |

This study has some limitations. First, the analysis is limited to non-financial firms listed on the Borsa Istanbul. This may limit the generalizability of the findings to other countries and markets. Additionally, the model does not include managerial variables such as the number of independent board members or the number of female board members; only financial ratios were considered. Therefore, the findings of this study do not encompass the effects of other managerial and governance variables on carbon disclosure willingness.

### 3.2. Methodology

In this study, several machine learning algorithms with different techniques, including K-Nearest Neighbors, Naive Bayes, Support Vector Machines, Decision Trees, Random Forest, and XGBoost, are used to predict whether firms will disclose their carbon emissions based on their financial ratios. 80% of the dataset is allocated for training and 20% for testing.

#### 3.2.1. K-Nearest Neighbors

K-Nearest Neighbors (KNN) works on the principle that similar data points are likely to be close to each other in feature space. When making predictions for a new data point, KNN identifies the 'k' closest training samples (neighbors), based on a distance metric, typically Manhattan, Minkowski, or Euclidean distance. The algorithm then classifies the new point among its neighbors based on the majority of votes (Müller and Guido, 2016; Huang et al., 2023).

#### 3.2.2. Naive Bayes

Naïve Bayes is a classification technique based on the probability theory in Bayes' Theorem (Sarang, 2023). It uses available class information to determine the probability that a data point belongs to a class. It is based on the assumption that the presence of one feature does not affect the other. In practice, Naïve Bayes calculates the probability of classification for a data sample by examining the probabilities of features within each category and their combined probabilities. Ultimately, it identifies the class with the highest likelihood for a data point (Frank et al., 2000; Zhang, 2004).

#### 3.2.3. Support Vector Machines

Support vector machines are essentially based on the idea of transforming input vectors through training data into a high-dimensional feature space using a nonlinear transformation and then performing a linear discrimination in the feature space (Vapnik, 1998; Dhanalakshmi et al., 2009). In linear separation, the goal is to categorize class members according to the number of classes by selecting the line that provides the largest margin among an unlimited number of lines that can separate classes. The line corresponding to this margin is drawn parallel to the members of the class closest to the selected line. In fact, these lines in the feature space, called hyperplanes, correspond to a decision boundary in the input space (Cristianini and Ricci, 2008). In addition, SVMs map data into a higher-dimensional space using various kernel functions, enabling the classification of data that cannot be linearly separated (Deisenroth et al., 2020).

#### 3.2.4. Decision Trees

Decision tree algorithms, which can be used in classification and regression problems, divide the data into subsets according to the conditions specified at each internal node, and this process continues recursively until a defined stopping criterion is reached. Internal nodes represent tests based on features, and each branch corresponds to the results of these tests. Leaf nodes represent the class label or prediction value (Sarang, 2023). This tree structure guides the decision-making process by following a path from the root to the leaf node. For classification problems, prediction is based on the majority class observed at the leaf node (Loh, 2011).

#### 3.2.5. Random Forest

Based on ensemble learning, the Random Forest algorithm is based on the results of multiple classifiers randomly generated from existing situations rather than one classifier. In other words, the algorithm generates a large number of independently sampled decision trees, and the final prediction is based on majority voting among these trees. Unlike decision trees, the Random Forest algorithm branches each node using the best of randomly selected variables at each node. As the quantity of trees increases, the generalization error tends to decrease. This adds robustness and accuracy to the model. Furthermore, the random forest algorithm has the capacity to identify nonlinear patterns in a way that is robust to noise and overfitting (Breiman, 2001).

#### 3.2.6. XGBoost

The XGBoost algorithm is based on the logic that each tree is built sequentially, optimizing each tree to minimize the errors of the previous trees. Unlike the random forest approach, where the majority of votes between trees determines the final prediction, the XGBoost model aggregates the outputs of all trees to produce the final prediction (Chen and Guestrin, 2016; Sagi and Rokach, 2021). The XGBoost algorithm supports both L1 (Lasso) and L2 (Ridge) regularization, which helps prevent overfitting. It also allows parallel processing in distributed environments (Sarang, 2023).

### 3.3. Performance Metrics

For classification tasks, several commonly used metrics are used to evaluate a model's performance (Hossin and Sulaiman, 2015).

#### 3.3.1. Accuracy

Accuracy constitutes a quantitative measure that evaluates the proportion of instances wherein a machine learning model correctly forecasts the resultant outcome. As illustrated in Equation 1, the accuracy metric is derived by computing the ratio of the count of correct predictions to the aggregate number of predictions made.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

### 3.3.2. Precision

Precision represents the ratio of accurately classified positive instances to the total number of positive instances, as shown in Equation 2. Precision serves as a metric that indicates the extent to which the machine learning model successfully predicts the positive class.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

### 3.3.3. Recall

Recall is defined as the ratio of the quantity of accurately identified positive samples to the aggregate number of positive samples, as illustrated in Equation 3. In other terms, this metric encompasses both true positives and false negatives.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

The metric of recall is particularly advantageous in scenarios where the repercussions of false negatives are significantly elevated.

### 3.3.4. F1 Score

As shown in Equation 4, this metric summarizes the harmonic mean of recall and precision values. It thoroughly evaluates the model's ability to differentiate between positive and negative examples accurately.

$$F1\ Score = 2 \times \frac{Precision\ \times Recall}{Precision + Recall} \tag{4}$$

An F1 score close to 1 indicates strong model performance, while a score near 0 reflects poor performance.

### 3.3.5. ROC Curve

The Receiver Operating Characteristic (ROC) curve represents the relationship between precision and recall for a classification model. It plots the true positive rate against the false positive rate. Specifically, the X-axis of the ROC curve represents the true positive rate, which is calculated using Equation 5. The Y-axis represents the false positive rate, derived from Equation 6.

$$TPR = \frac{TP}{TP + FN} \tag{5}$$

$$FPR = \frac{FP}{FP + TN} \tag{6}$$

Each point on the ROC curve corresponds to a specific pair of precision and recall, thereby indicating the model's effectiveness in discriminating between positive and negative classes. The Area Under the Curve (AUC) is a numerical expression of the results on the ROC curve. In other words, it corresponds to the ratio of correctly classified positive samples to misclassified positive samples, and an AUC value less than 0.5 means that the predictions failed. A higher area under the ROC curve indicates classification performance (Balbal, 2024).

### 3.3.6. Confusion Matrix

The confusion matrix tabulates the performance of the model by comparing the predicted values of the target feature with the actual observed values. As shown in Table 2, the lower right cell of the confusion matrix represents the number of true positive predictions, where the model correctly identifies positive samples. The upper right cell represents false positive predictions, where the model incorrectly classifies negative samples as positive. The upper left cell reflects true negative predictions where the model correctly identified negative examples. The bottom left cell represents false negative predictions where the model failed to recognize positive examples (Harrison, 2024).

**Tablo 2. Confusion Matrix**

| | **Predicted Negative** | **Predicted Positive** |
|---|---|---|
| Actual Negative | TN | FP |
| Actual Positive | FN | TP |

A confusion matrix, which is a matrix detailing the counts of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), can be utilized to calculate various metrics for the evaluation of binary classification classifier (Rainio et al., 2024).

## 3.4. Findings

### 3.4.1. Machine Learning Results

Table 3 presents the model summary for each machine learning algorithm, including accuracy, precision, recall, f1 score, area under the ROC curve (AUC), macro avg and weighted avg scores. Table 3 Panel A shows the classification results of the KNN algorithm. The KNN method has an accuracy of 0.58, ROC of 0.56, and f1 score of 0.50. Table 3 Panel B shows the classification results of the NB algorithm. In the NB method, accuracy is 0.59, ROC is 0.59 and f1 score is 0.59. Table 3 Panel C shows the classification results of the SVM algorithm. The SVM method has an accuracy of 0.61, ROC of 0.54, and f1 score of 0.47. Table 3 Panel D presents the classification results of the DT algorithm. In the DT method, the accuracy is 0.73, the ROC is 0.71, and the f1 score is 0.72. Table 3 Panel E presents the classification results of the RF algorithm. In the RF method, the accuracy is 0.86, the ROC is 0.93, and the f1 score is 0.85. Table 3 Panel F presents the classification results of the XGBoost algorithm. In the XGBoost method, the accuracy is 0.87, the ROC is 0.92, and f1 score is 0.87. According to these results, the machine learning algorithms with the highest classification success are RF and XGBoost, while the algorithm with the lowest classification success is SVM.

**Tablo 3. Classification Report for Machine Learning Models**

| | | | | |
|---|---|---|---|---|
| **Panel A: Classification Report for KNN** | | | | |
| | **precision** | **recall** | **f1-score** | **support** |
| Submitted | 0.60 | 0.88 | 0.71 | 92 |
| No Response | 0.39 | 0.11 | 0.18 | 61 |
| macro avg | 0.49 | 0.50 | 0.45 | 153 |
| weighted avg | 0.52 | 0.58 | 0.50 | 153 |
| | | | accuracy | 0.58 |
| | | | ROC | 0.56 |
| **Panel B: Classification Report for NB** | | | | |
| | **precision** | **recall** | **f1-score** | **support** |
| Submitted | 0.62 | 0.85 | 0.72 | 92 |
| No Response | 0.48 | 0.21 | 0.30 | 61 |
| macro avg | 0.55 | 0.53 | 0.51 | 153 |
| weighted avg | 0.56 | 0.59 | 0.55 | 153 |
| | | | accuracy | 0.59 |
| | | | ROC | 0.59 |
| **Panel C: Classification Report for SVM** | | | | |
| | **precision** | **recall** | **f1-score** | **support** |
| Submitted | 0.61 | 1.00 | 0.75 | 92 |
| No Response | 1.00 | 0.02 | 0.03 | 61 |
| macro avg | 0.80 | 0.51 | 0.39 | 153 |
| weighted avg | 0.76 | 0.61 | 0.47 | 153 |
| | | | accuracy | 0.61 |
| | | | ROC | 0.54 |
| **Panel D: Classification Report for DT** | | | | |
| | **precision** | **recall** | **f1-score** | **support** |
| Submitted | 0.77 | 0.78 | 0.77 | 92 |
| No Response | 0.66 | 0.64 | 0.65 | 61 |
| macro avg | 0.71 | 0.71 | 0.71 | 153 |
| weighted avg | 0.72 | 0.73 | 0.72 | 153 |
| | | | accuracy | 0.73 |
| | | | ROC | 0.71 |
| **Panel E: Classification Report for RF** | | | | |
| | **precision** | **recall** | **f1-score** | **support** |
| Submitted | 0.84 | 0.93 | 0.89 | 92 |
| No Response | 0.88 | 0.74 | 0.80 | 61 |
| macro avg | 0.86 | 0.84 | 0.85 | 153 |
| weighted avg | 0.86 | 0.86 | 0.85 | 153 |
| | | | accuracy | 0.86 |
| | | | ROC | 0.93 |
| **Panel F: Classification Report for XGBoost** | | | | |
| | **precision** | **recall** | **f1-score** | **support** |
| Submitted | 0.89 | 0.89 | 0.89 | 92 |
| No Response | 0.84 | 0.84 | 0.84 | 61 |
| macro avg | 0.86 | 0.86 | 0.86 | 153 |
| weighted avg | 0.87 | 0.87 | 0.87 | 153 |
| | | | accuracy | 0.87 |
| | | | ROC | 0.92 |

Table 4 shows the confusion matrices obtained from each machine learning algorithm. Table 4 Panel A shows the classification results of the KNN algorithm, Table 4 Panel B shows the NB algorithm, Table 4 Panel C shows the SVM algorithm, Table 4 Panel D shows the DT algorithm, Table 4 Panel E shows the RF algorithm, and Table 4 Panel F shows the XGBoost algorithm. When the confusion matrix of each machine learning method is analyzed, it is seen that RF and XGBoost methods based on ensemble learning are more successful in predicting whether firms will voluntarily disclose carbon.

**Tablo 4. Confusion Matrices for Machine Learning Models**

| **Panel A: K-Nearest Neighbors Confusion Matrix** | | |
|---|---|---|
| Actual / Predicted | Submitted | No Response |
| Submitted | 81 (True Submitted) | 11 (False Submitted) |
| No Response | 54 (False No Response) | 7 (True No Response) |
| **Panel B: Naïve Bayes Confusion Matrix** | | |
| Actual / Predicted | Submitted | No Response |
| Submitted | 78 (True Submitted) | 14 (False Submitted) |
| No Response | 48 (False No Response) | 13 (True No Response) |
| **Panel C: Support Vector Machine Confusion Matrix** | | |
| Actual / Predicted | Submitted | No Response |
| Submitted | 92 (True Submitted) | 0 (False Submitted) |
| No Response | 60 (False No Response) | 1 (True No Response) |
| **Panel D: Decision Tree Confusion Matrix** | | |
| Actual / Predicted | Submitted | No Response |
| Submitted | 74 (True Submitted) | 18 (False Submitted) |
| No Response | 23 (False No Response) | 38 (True No Response) |
| **Panel E: Random Forest Confusion Matrix** | | |
| Actual / Predicted | Submitted | No Response |
| Submitted | 86 (True Submitted) | 6 (False Submitted) |
| No Response | 16 (False No Response) | 45 (True No Response) |
| **Panel F: XGBoost Confusion Matrix** | | |
| Actual / Predicted | Submitted | No Response |
| Submitted | 82 (True Submitted) | 10 (False Submitted) |
| No Response | 10 (False No Response) | 51 (True No Response) |

Figure 3 presents plots of the areas under the ROC curve (AUC) for each machine learning algorithm. These plots illustrate the relationship between the true positive rate and the false positive rate across all possible classification thresholds. The AUC quantifies the entire area beneath the ROC curve.

As shown in Figure 3 Panel A, the K-Nearest Neighbors algorithm achieved an AUC of 0.56, indicating a relatively modest classification performance. In Figure 3 Panel B, the Naive Bayes algorithm slightly outperformed KNN with an AUC of 0.59. Conversely, the Support Vector Machine model, as shown in Figure 3 Panel C, yielded a lower AUC of 0.46, suggesting limited discriminatory power. Figure 3 Panel D illustrates the performance of the Decision Tree algorithm, which achieved a more promising AUC of 0.71. Notably, both the Random Forest and XGBoost algorithms, depicted in Figure 3 Panel E and 3 Panel F, respectively, demonstrated the highest predictive performance, each obtaining an AUC of 0.93 and 0.92. These results highlight the superior classification capabilities of ensemble-based methods compared to individual classifiers. Based on these findings, the SVM algorithm exhibits the lowest prediction success, while the RF algorithm achieves the highest ROC score.

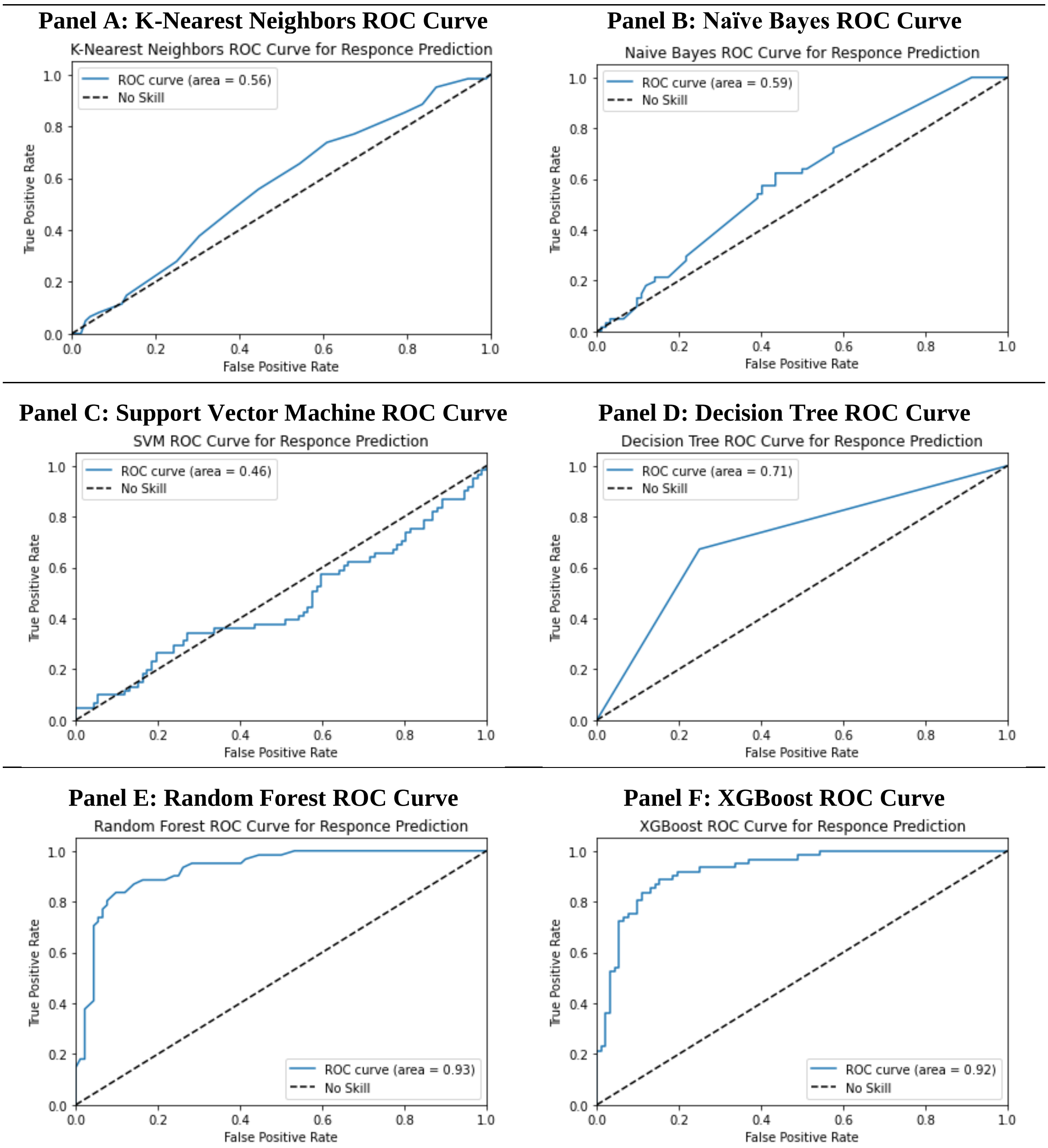


**Figure 3. ROC Curves for Machine Learning Models**

### 3.4.2. Additional Model Validation with Different Parameters and Data Splits

In this context, additional analyses were conducted to test the robustness of the model and evaluate the impact of parameter selections on classification performance. In this direction, the model's sensitivity was analyzed using different training-test ratios and hyperparameter configurations, and the findings were discussed. In this scope, XGBoost classifiers were evaluated using both 80% training – 20% test and 75% training – 25% test data splits, and two different parameter combinations presented in Table 5 were applied.

**Tablo 5. Classification Metrics of XGBoost with Varying Train-Test Splits and Parameter Settings**

| Train Size | Test Size | Parameters | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|---|
| 0.80 | 0.20 | n_estimators: 120,<br>max_depth: 11,<br>learning_rate: 0.1,<br>subsample: 0.8,<br>colsample_bytree: 0.8,<br>gamma: 0.1,<br>reg_alpha: 0.2,<br>reg_lambda: 0.1 | 0.88 | 0.88 | 0.87 | 0.88 |
| 0.80 | 0.20 | n_estimators: 150,<br>max_depth: 8,<br>learning_rate: 0.05,<br>subsample: 0.7,<br>colsample_bytree: 0.7,<br>gamma: 0.05,<br>reg_alpha: 0.1,<br>reg_lambda: 0.2} | 0.82 | 0.84 | 0.81 | 0.81 |
| 0.75 | 0.25 | n_estimators: 120,<br>max_depth: 11,<br>learning_rate: 0.1,<br>subsample: 0.8,<br>colsample_bytree: 0.8,<br>gamma: 0.1,<br>reg_alpha: 0.2,<br>reg_lambda: 0.1 | 0.83 | 0.84 | 0.82 | 0.82 |
| 0.75 | 0.25 | n_estimators: 150,<br>max_depth: 8,<br>learning_rate: 0.05,<br>subsample: 0.7,<br>colsample_bytree: 0.7,<br>gamma: 0.05,<br>reg_alpha: 0.1,<br>reg_lambda: 0.2 | 0.80 | 0.82 | 0.79 | 0.79 |

The findings revealed that the first parameter set performed better than the second in both data splits and achieved an accuracy rate of 88% in the 80-20 data split. This comparative analysis supports the robustness of the findings and once again emphasizes the importance of model selection and validation strategies in machine learning applications.

### 3.4.3. Feature Importance

Feature importance is a metric that determines which features (variables) are important in the learning process of a model, especially in machine learning algorithms such as XGBoost, which play a critical role in modern artificial intelligence systems and the development of explainable artificial intelligence (XAI) techniques. Feature importance improves the understandability of the model and provides guidance on which features should be optimized to improve model performance. It also allows simplifying the model and reducing training time by identifying less important features of the model. Feature importance analysis in XGBoost makes the decision-making process more transparent and explainable by identifying the key features that have an impact on the model's outputs.

Figure 4 shows the importance levels of firms' carbon disclosure prediction with the XGBoost method in a descending manner from top to bottom. Accordingly, the top three most explanatory variables are Equity / Total Liabilities Ratio, Long-Term Liabilities / Total Liabilities Ratio, and Fixed Assets / Equity Ratio.

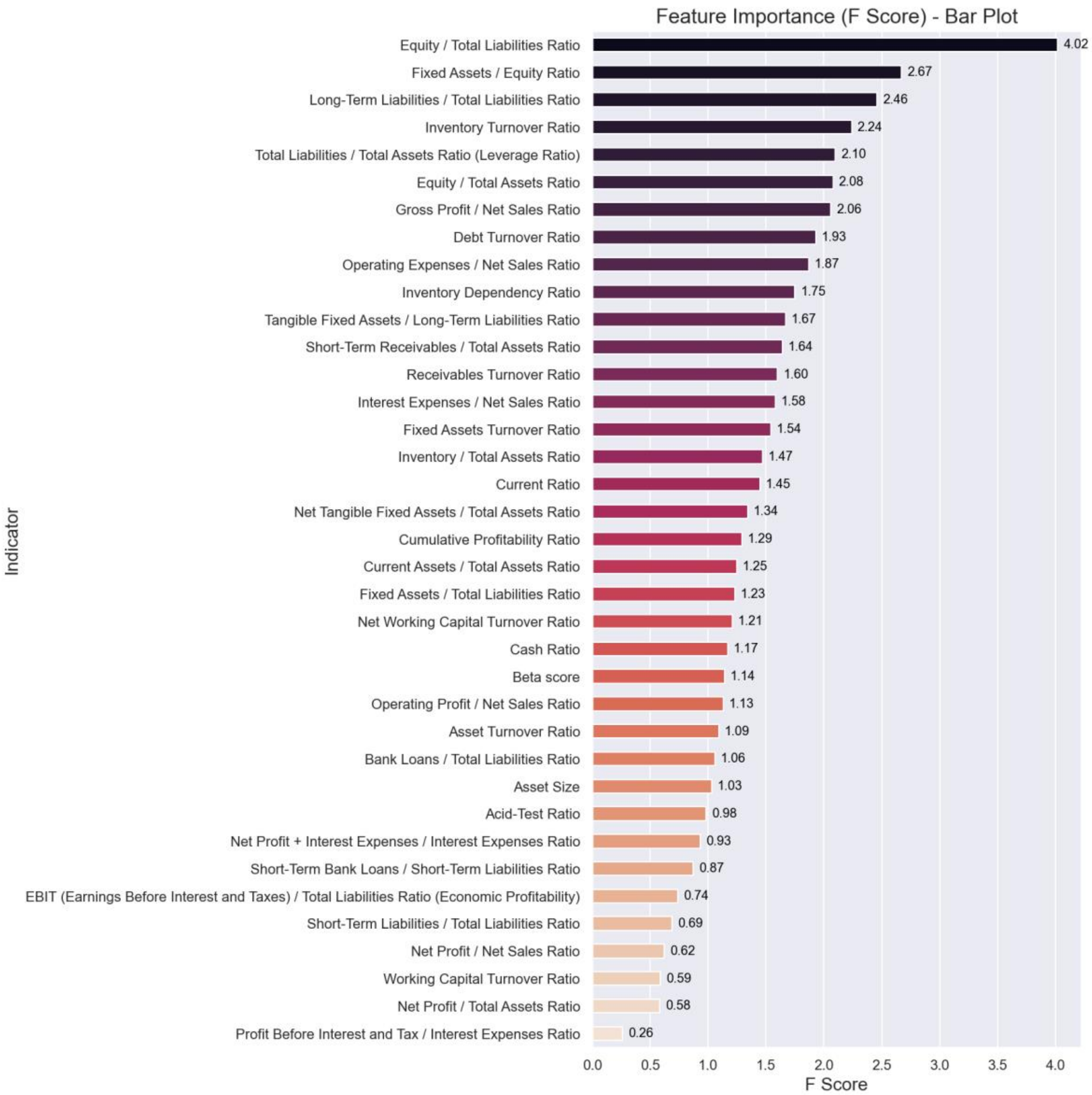


**Figure 4. Feature Importance for Carbon Reporting Prediction with XGBoost**

The ratio of shareholders' equity to total liabilities is one of the most important financial indicators that assesses a firm's capital structure. A high level of this ratio indicates that the firm is financed with equity rather than foreign equity and is less risky. In contrast, a low level suggests that the firm is financed with foreign equity rather than equity and is more dangerous, especially during economic downturns.

In this context, another important indicator is the ratio of long-term liabilities to total liabilities. This ratio reveals the debt structure of firms in terms of short-term and long-term liabilities. A high ratio indicates that a significant portion of liabilities is composed of long-term liabilities, while a low ratio indicates that liabilities are mainly composed of short-term liabilities. While a significant share of long-term liabilities in total liabilities corresponds to a more robust financing structure, a significant share of short-term liabilities indicates an increase in liquidity risk.

Another essential ratio that increases the model's explanatory power is the ratio of fixed assets to equity. This ratio indicates how much a firm's fixed assets are financed by equity. A high ratio indicates that a significant portion of equity is allocated to fixed assets. In contrast, a low ratio indicates that a smaller amount of equity is associated with fixed assets.

The findings suggest that firms' propensity to make voluntary disclosures may be linked to their asset and capital structure. In this regard, the results are consistent with those of Choi et al. (2013), Nguyen and Phan (2020), D'Amato et al. (2021), and Frost et al. (2023).

### 3.5. Discussion

The empirical findings indicate that firms' propensity to respond to the CDP survey can be effectively predicted using financial ratios as indicators. This finding suggests that firms' voluntary reporting of carbon emissions can be successfully predicted based on financial indicators, a more detailed explanation of the relationship between financial indicators and carbon transparency.

In evaluating the findings within this context, it is evident that leverage and efficiency ratios emerge as the most effective financial indicators in predicting firms' propensity to disclose carbon emissions voluntarily. As articulated by Nguyen and Phan (2020), the primary rationale for this phenomenon pertains to the capacity of these ratios to provide crucial insights into a firm's effectiveness in utilizing its resources. Conversely, undercapitalization, high indebtedness, and low productivity levels are likely to negatively affect firms' environmental performance and reduce their propensity to disclose carbon emissions voluntarily.

The propensity to disclose carbon emissions can be associated with financial indicators in the context of corporate governance and transparency (D'Amato et al., 2021). Financial indicators reflect firms' risk management practices, growth strategies, sustainability policies, and overall corporate governance quality. In this context, financially strong firms are likely to engage in more comprehensive and voluntary environmental reporting to enhance their credibility and reputation with investors (Birkey et al., 2016). From this perspective, the success of financial indicators in predicting attitudes towards voluntary disclosure of carbon emissions may also be associated with the risk perception and expectations of the market. This is because a firm's financial structure shapes the level of risk perceived by market actors towards the firm, thus increasing the capacity of firms with a strong financial structure to manage environmental risks more effectively (Cho et al., 2012).

Furthermore, the explanatory power of financial indicators on the willingness to voluntarily disclose carbon emissions can be evaluated within the framework of signaling theory. Voluntary carbon emission reporting signals firms' commitment to environmental sustainability and their performance in this area to external stakeholders (Choi et al., 2013). Consequently, positive financial indicators can be interpreted as signals that firms are willing to invest in environmentally friendly practices and share these practices transparently in the context of long-term risk management and sustainability (Harker et al., 2022). As a result, financial indicators, which provide essential information about a firm's asset and resource management, risk management, and market relationships, stand out as practical tools for predicting the propensity to disclose voluntary carbon emissions.

## 4. Conclusion

This paper investigates whether financial indicators can predict whether firms will voluntarily report their carbon emission information within the scope of CDP, as well as the most important financial indicators for successful predictions. To this end, the present study employs a unique dataset, encompassing financial ratios of 100 firms traded on Borsa Istanbul from 2016 to 2023, and leverages a range of machine learning algorithms, including KNN, NB, SVM, DT, RF, and XGBoost. The findings indicate that the Random Forest and XGBoost algorithms can accurately predict whether firms will make voluntary carbon emission disclosures with over 92% precision. The study further demonstrates that specific financial ratios, including equity/total liabilities, long-term liabilities/total liabilities, and fixed assets/equity, play a pivotal role in enhancing the model's explanatory power within the XGBoost algorithm.

The results suggest that firms' propensity to make voluntary disclosures may be closely related to their asset and capital structures. This finding highlights the importance of incorporating environmental performance criteria into internal risk management processes by considering carbon emissions and other environmental factors, which could help reduce the operational and financial risks that firms may encounter. Additionally, the findings indicate that firms that integrate environmental reporting into their reporting processes could enhance their reputation with stakeholders and investors, potentially positively impacting their market values. Indeed, the findings indicate that firms that adopt environmental reporting may be better positioned to adapt to market conditions over the long term, capitalize on new sustainability-driven business opportunities, and thus gain a competitive advantage.

The findings of this paper also indicate a significant opportunity for investors, which is related to the ability to make more informed investment decisions by effectively estimating firms' tendencies to disclose their carbon emissions. In other words, assessing firms' inclination to disclose carbon emissions presents an opportunity to identify and manage the risks associated with these emissions more clearly. Predicting whether firms will make voluntary carbon disclosures helps investors mitigate financial losses from potential environmental regulations, criminal sanctions, and reputational damage to their portfolios. This foresight enables investors to proactively address risks before they materialize, thereby avoiding potential losses. In summary, these findings underscore the capacity of advanced analytical tools to enhance risk analysis frameworks and ultimately facilitate more informed financial decision-making.

The paper contributes to the literature by demonstrating that machine learning algorithms provide a more sophisticated, data-driven approach to predicting carbon disclosure willingness, surpassing traditional analytical methods, and that innovative methodologies can be effectively employed in environmental reporting analysis. Furthermore, this paper systematically advances the literature on the interaction between corporate finance and environmental transparency by showing that specific financial ratios can act as predictive indicators of firms' carbon-reporting behavior. Finally, the paper suggests that carbon reporting is a critical indicator that can significantly influence investor risk assessments and that this insight contributes to the development of sustainable investment strategies in both theoretical and practical contexts regarding investor risk management.

Finally, future studies could be expanded to include data from different countries and stock exchanges, thereby increasing the external validity of the findings. Including managerial variables such as board structure, gender diversity, and CEO characteristics in the analysis could provide more in-depth and comprehensive results regarding the determinants of carbon disclosure behavior.

**Declaration of Research and Publication Ethics**
This study, which does not require ethics committee approval and/or legal/specific permission, complies with the research and publication ethics.

**Researcher's Contribution Rate Statement**
I am the single author of this paper. My contribution is 100%.

**Declaration of Researcher's Conflict of Interest**
There are no potential conflicts of interest in this study.

## References

Alsaifi, K., Elnahass, M. and Salama, A. (2020). Carbon disclosure and financial performance: UK environmental policy. *Business Strategy and the Environment*, 29(2), 711-726. https://doi.org/10.1002/bse.2426

Anquetin, T., Coqueret, G., Tavin, B. and Welgryn, L. (2022). Scopes of carbon emissions and their impact on green portfolios. *Economic Modelling*, 115, 105951. https://doi.org/10.1016/j.econmod.2022.105951

Arendt, C.A., Hyland, E.G. and Piliouras, A. (2021). The geological consequences of global climate change. In D. Alderton and S.A. Elias (Eds.), *Encyclopedia of geology second edition* (pp. 510–522). New York: Academic Press.

Balbal, K.F. (2024). MIPART: A partial decision tree-based method for multiple-instance classification. *Applied Sciences*, 14(24), 11696. https://doi.org/10.3390/app142411696

Benkraiem, R., Shuwaikh, F., Lakhal, F. and Guizani, A. (2022). Carbon performance and firm value of the World's most sustainable companies. *Economic Modelling*, 116, 106002. https://doi.org/10.1016/j.econmod.2022.106002

Birkey, R.N., Michelon, G., Patten, D.M. and Sankara, J. (2016). Does assurance on CSR reporting enhance environmental reputation? An examination in the U.S. context. *Accounting Forum*, 40(3), 143–152. https://doi.org/10.1016/j.accfor.2016.07.001

Bolton, P. and Kacperczyk, M. (2021). Do investors care about carbon risk? *Journal of Financial Economics*, 142(2), 517-549. https://doi.org/10.1016/j.jfineco.2021.05.008

Bose, S., Minnick, K. and Shams, S. (2021). Does carbon risk matter for corporate acquisition decisions? *Journal of Corporate Finance*, 70, 102058. https://doi.org/10.1016/j.jcorpfin.2021.102058

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Caby, J., Ziane, Y. and Lamarque, E. (2020). The determinants of voluntary climate change disclosure commitment and quality in the banking industry. *Technological Forecasting and Social Change*, 161, 120282. https://doi.org/10.1016/j.techfore.2020.120282

Chen, T. and Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In B. Krishnapuram and M. Shah (Eds.), *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785-794). New York: Association for Computing Machinery.

Cho, C.H., Guidry, R.P., Hageman, A.M. and Patten, D.M. (2012). Do actions speak louder than words? An empirical investigation of corporate environmental reputation. *Accounting, Organizations and Society*, 37(1), 14-25. https://doi.org/10.1016/j.aos.2011.12.001

Choi, B.B., Lee, D. and Psaros, J. (2013). An analysis of Australian company carbon emission disclosures. *Pacific Accounting Review*, 25(1), 58-79. https://doi.org/10.1108/01140581311318968

Cooper, S.A., Raman, K.K. and Yin, J. (2018). Halo effect or fallen angel effect? Firm value consequences of greenhouse gas emissions and reputation for corporate social responsibility. *Journal of Accounting and Public Policy*, 37(3), 226-240. https://doi.org/10.1016/j.jaccpubpol.2018.04.003

Costa de Oliveira, A., Marini, N. and Farias, D.R. (2014). Climate change: New breeding pressures and goals. In N.K. Van Alfen (Ed.), *Encyclopedia of agriculture and food systems* (pp. 284–293). New York: Academic Press.

Cristianini, N. and Ricci, E. (2008). Support vector machines. In M.-Y. Kao (Ed.), *Encyclopedia of algorithms* (pp. 928–932). US: Springer.

D'Amato, V., D'Ecclesia, R. and Levantesi, S. (2021). Fundamental ratios as predictors of ESG scores: A machine learning approach. *Decisions in Economics and Finance*, 44(2), 1087-1110. https://doi.org/10.1007/s10203-021-00364-5

Deisenroth, M.P., Faisal, A.A. and Ong, C.S. (2020). *Mathematics for machine learning*. Cambridge: Cambridge University Press.

Desai, R., Raval, A., Baser, N. and Desai, J. (2022). Impact of carbon emission on financial performance: Empirical evidence from India. *South Asian Journal of Business Studies*, 11(4), 450-470. https://doi.org/10.1108/SAJBS-10-2020-0384

Dhanalakshmi, P., Palanivel, S. and Ramalingam, V. (2009). Classification of audio signals using SVM and RBFNN. *Expert Systems with Applications*, 36(3), 6069-6075. https://doi.org/10.1016/j.eswa.2008.06.126

Fan, P., Qian, X. and Wang, J. (2023). Does gender diversity matter? Female directors and firm carbon emissions in Japan. *Pacific-Basin Finance Journal*, 77, 101931. https://doi.org/10.1016/j.pacfin.2022.101931

Frank, E., Trigg, L., Holmes, G. and Witten, I.H. (2000). Naive Bayes for regression. *Machine Learning*, 41, 5-25. https://doi.org/10.1023/A:1007670802811

Frost, G., Jones, S. and Yu, M. (2023). Voluntary carbon reporting prediction: A machine learning approach. *Abacus*, 59(4), 1116-1166. https://doi.org/10.1111/abac.12298

Füssel, H.M. and Klein, R.J. (2006). Climate change vulnerability assessments: An evolution of conceptual thinking. *Climatic Change*, 75(3), 301-329. https://doi.org/10.1007/s10584-006-0329-3

Griffin, P.A., Lont, D.H. and Sun, E.Y. (2017). The relevance to investors of greenhouse gas emission disclosures. *Contemporary Accounting Research*, 34(2), 1265-1297. https://doi.org/10.1111/1911-3846.12298

Griffin, P.A., Neururer, T. and Sun, E.Y. (2020). Environmental performance and analyst information processing costs. *Journal of Corporate Finance*, 61, 101397. https://doi.org/10.1016/j.jcorpfin.2018.08.008

Guastella, G., Mazzarano, M., Pareglio, S. and Xepapadeas, A. (2022). Climate reputation risk and abnormal returns in the stock markets: A focus on large emitters. *International Review of Financial Analysis*, 84, 102365. https://doi.org/10.1016/j.irfa.2022.102365

Güneysu, Y. ve Atasel, O.Y. (2022). Karbon emisyonları ile finansal performans arasındaki ilişkinin incelenmesi: BIST100 endeksinde bir araştırma. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 32, 3(1183-1193). https://doi.org/10.18069/firatsbed.1125859

Haque, F. (2017). The effects of board characteristics and sustainable compensation policy on carbon performance of UK firms. *The British Accounting Review*, 49(3), 347-364. https://doi.org/10.1016/j.bar.2017.01.001

Harker, C., Hassall, M., Lant, P., Rybak, N. and Dargusch, P. (2022). What can machine learning teach us about Australian climate risk disclosures? *Sustainability*, 14(16), 10000. https://doi.org/10.3390/su141610000

Harrison, M. (2024). *Effective XGBoost: Optimizing, tuning, understanding, and deploying classification models*. Salt Lake City: MetaSnake.

History of the convention. (n.d.). *United Nations framework convention on climate change.* Retrieved from https://unfccc.int/process/the-convention/history-of-the-convention#Climate-Change-in-context

Hollindale, J., Kent, P., Routledge, J. and Chapple, L. (2019). Women on boards and greenhouse gas emission disclosures. *Accounting & Finance*, 59(1), 277-308. https://doi.org/10.1111/acfi.12258

Hossin, M. and Sulaiman, M.N. (2015). A review on evaluation metrics for data classification evaluations. *International Journal of Data Mining & Knowledge Management Process*, 5(2), 1-11. https://doi.org/10.5121/ijdkp.2015.5201

Huang, X., Jin, G. and Ruan, W. (2023). Naive Bayes. In X. Huang, G. Jin and W. Ruan (Eds.), *Machine learning safety. Artificial intelligence: Foundations, theory, and algorithms* (pp. 95–102). Berlin: Springer Nature.

Jung, J., Herbohn, K. and Clarkson, P. (2018). Carbon risk, carbon risk awareness and the cost of debt financing. *Journal of Business Ethics*, 150, 1151-1171. https://doi.org/10.1007/s10551-016-3207-6

Kabir, M.N., Rahman, S., Rahman, M.A. and Anwar, M. (2021). Carbon emissions and default risk: International evidence from firm-level data. *Economic Modelling*, 103, 105617. https://doi.org/10.1016/j.econmod.2021.105617

Kleimeier, S. and Viehs, M. (2021). Pricing carbon risk: Investor preferences or risk mitigation? *Economics Letters*, 205, 109936. https://doi.org/10.1016/j.econlet.2021.109936

Kolk, A. and Pinkse, J. (2005). Business responses to climate change: Identifying emergent strategies. *California Management Review*, 47(3), 6-20. https://doi.org/10.2307/41166304

Krishnamurti, C. and Velayutham, E. (2018). The influence of board committee structures on voluntary disclosure of greenhouse gas emissions: Australian evidence. *Pacific-Basin Finance Journal*, 50, 65-81. https://doi.org/10.1016/j.pacfin.2017.09.003

Kyoto Protocol. (n.d.). *Climate change presidency.* Retrieved from https://iklim.gov.tr/kyoto-protokolu-i-35#

Lee, K.H., Min, B. and Yook, K.H. (2015). The impacts of carbon (CO2) emissions and environmental research and development (R&D) investment on firm performance. *International Journal of Production Economics*, 167, 1-11. https://doi.org/10.1016/j.ijpe.2015.05.018

Liao, L., Luo, L. and Tang, Q. (2015). Gender diversity, board independence, environmental committee and greenhouse gas disclosure. *The British Accounting Review*, 47(4), 409-424. https://doi.org/10.1016/j.bar.2014.01.002

Loh, W.Y. (2011). Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1), 14-23. https://doi.org/10.1002/widm.8

López-Pacheco, I.Y., Rodas-Zuluaga, L.I., Fuentes-Tristan, S., Castillo-Zacarías, C., Sosa-Hernández, J.E., Barceló, D., … Parra-Saldívar, R. (2021). Phycocapture of CO2 as an option to reduce greenhouse gases in cities: Carbon sinks in urban spaces. *Journal of CO2 Utilization*, 53, 101704. https://doi.org/10.1016/j.jcou.2021.101704

Lu, W., Zhu, N. and Zhang, J. (2021). The impact of carbon disclosure on financial performance under low carbon constraints. *Energies*, 14 (14). https://doi.org/10.3390/en14144126

Matsumura, E.M., Prakash, R. and Vera-Munoz, S.C. (2014). Firm-value effects of carbon emissions and carbon disclosures. *The Accounting Review*, 89(2), 695-724. https://doi.org/10.2308/accr-50629

Merello, P., Barberá, A. and De la Poza, E. (2022). Is the sustainability profile of FinTech companies a key driver of their value? *Technological Forecasting and Social Change*, 174, 121290. https://doi.org/10.1016/j.techfore.2021.121290

Müller, A.C. and Guido, S. (2016). *Introduction to machine learning with Python: A guide for data scientists* (1st ed.). California: O'Reilly Media.

NASA. (2024). *What is climate change?* Retrieved from https://science.nasa.gov/climate-change/what-is-climate-change/

Nguyen, J.H. and Phan, H.V. (2020). Carbon risk and corporate capital structure. *Journal of Corporate Finance*, 64, 101713. https://doi.org/10.1016/j.jcorpfin.2020.101713

Paris Anlaşması. (n.d.). *T.C. Dışişleri Bakanlığı.* Retrieved from https://www.mfa.gov.tr/paris-anlasmasi.tr.mfa

Pires, J.C.M. (2017). COP21: The algae opportunity? *Renewable and Sustainable Energy Reviews*, 79, 867–877. https://doi.org/10.1016/J.RSER.2017.05.197

Rahat, B. and Nguyen, P. (2022). Risk-adjusted investment performance of green and black portfolios and impact of toxic divestments in emerging markets. *Energy Economics*, 116, 106423. https://doi.org/10.1016/j.eneco.2022.106423

Rainio, O., Teuho, J. and Klén, R. (2024). Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, 14(1), 6086. https://doi.org/10.1038/s41598-024-56706-x

Rogoff, M.J. (2014). Introduction. In M.J. Rogoff (Ed.), *Solid waste recycling and processing* (pp. 1–9). Amsterdam: William Andrew.

Rohleder, M., Wilkens, M. and Zink, J. (2022). The effects of mutual fund decarbonization on stock prices and carbon emissions. *Journal of Banking & Finance*, 134, 106352. https://doi.org/10.1016/j.jbankfin.2021.106352

Safiullah, M., Kabir, M.N. and Miah, M.D. (2021). Carbon emissions and credit ratings. *Energy Economics*, 100, 105330. https://doi.org/10.1016/j.eneco.2021.105330

Sagi, O. and Rokach, L. (2021). Approximating XGBoost with an interpretable decision tree. *Information Sciences*, 572, 522-542. https://doi.org/10.1016/j.ins.2021.05.055

Sangiorgi, I. and Schopohl, L. (2021). Why do institutional investors buy green bonds: Evidence from a survey of European asset managers. *International Review of Financial Analysis*, 75, 101738. https://doi.org/10.1016/j.irfa.2021.101738

Sarang, P. (2023). Naive Bayes. In O. Marques (Ed.), *Thinking data science: A data science practitioner's guide* (pp. 143–152). Cham: Springer International Publishing.

The Paris Agreement. (n.d.). *UNFCCC*. Retrieved from https://unfccc.int/process-and-meetings/the-paris-agreement

Vapnik, V. (1998). *Statistical learning theory*. New York: Wiley.

Velte, P., Stawinoga, M. and Lueg, R. (2020). Carbon performance and disclosure: A systematic review of governance-related determinants and financial consequences. *Journal of Cleaner Production*, 254, 120063. https://doi.org/10.1016/j.jclepro.2020.120063

Yu, B., Li, C., Mirza, N. and Umar, M. (2021). Forecasting credit ratings of decarbonized firms: Comparative assessment of machine learning models. *Technological Forecasting and Social Change*, 174, 121255. https://doi.org/10.1016/j.techfore.2021.121255

Zhang, H. (2004). *The optimality of Naive Bayes*. Paper presented at the *Seventeenth international Florida artificial intelligence research society conference (FLAIRS 2004)*. Menlo Park, California. Retrieved from https://cdn.aaai.org/FLAIRS/2004/Flairs04-097.pdf