**Savunma Bilimleri Dergisi, 20(2): 279-302 (2024)**

SAVUNMA BİLİMLERİ DERGİSİ
The Journal of Defense Sciences

https://dergipark.org.tr/tr/pub/khosbd

# Forecasting Sustainability Reports with Financial Performance Indicators using Random Forest for Feature Selection and Gradient Boosting for Learning

## *Özellik Seçimi için Rastgele Orman ve Öğrenme için Gradyan Artırma Kullanılarak Finansal Performans Göstergeleriyle Sürdürülebilirlik Raporlarının Tahmin Edilmesi*

Hakan Ayhan DAĞISTANLI [1*], Figen ÖZEN [2], İlkay SARAÇOĞLU [3]

[1] *Department of Industrial and Systems Engineering, Turkish Military Academy, National Defence University, Ankara, Turkey*
[2] *Department of Electrical and Electronics Engineering, Haliç University, İstanbul, Turkey*
[3] *Department of Industrial Engineering, Haliç University, İstanbul, Turkey*

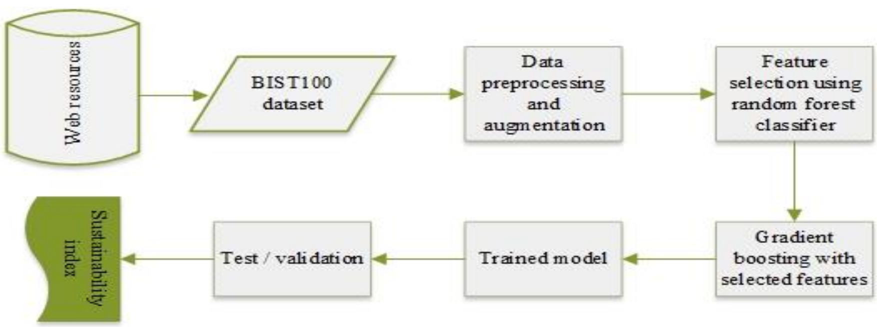
**Makale Bilgisi**

*Araştırma makalesi*
*Başvuru: 31.05.2024*
*Düzeltme: 13.06.2024*
*Kabul: 27.06.2024*

***Highlights***

*•Review the circular economy and sustainability reporting with financial performance data.*
*•Use non-manipulated inputs that are not dependent on market forces for analysis.*
*•Adapt the column-wise random shuffling method for overcome the data problem.*
*•Apply machine learning methods for forecasting in BIST Sustainability Index.*

***Graphical Abstract***




**Keywords**

*Machine Learning BIST Sustainability index*
*Financial Performance*
*Circular Economy*
*Random Forest*
*Gradient Boosting*

**Anahtar Kelimeler**

*Makine Öğrenmesi BIST Sürdürülebilirlik Endeksi*
*Finansal Performans*
*Döngüsel Ekonomi*
*Rastgele Orman*
*Gradyan Artırma*

***Abstract***

*Firms want to be included in the sustainability indices in order to attract the attention of the potential investor. In this study, time series data of financial performance of companies in XUSRD are used. On the other hand, contrary to the statistical analyses in the literature, to predict whether companies will take part in XUSRD, a combination of two machine learning methods, namely random forest for feature selection and gradient boosting for learning, is applied. In addition, to overcome the problem of data scarcity, the column-wise random shuffling method, which is a proven data augmentation technique in predicting stock market indices, has been preferred. The results show that the combination of random forest and gradient boosting reaches a test accuracy of 94.74% and outperforms state-of-the art models, namely, k-nearest neighbor, random forest, decision tree, support vector, naive bayes classifiers.*

***Özet***

*Firmalar potansiyel yatırımcının dikkatini çekebilmek adına sürdürülebilirlik endekslerine dahil olmak istiyor. Bu çalışmada XUSRD'deki şirketlerin finansal performansına ilişkin zaman serisi verileri kullanılmıştır. Öte yandan, literatürdeki istatistiksel analizlerin aksine, şirketlerin XUSRD'de yer alıp almayacağını tahmin etmek için, özellik seçimi için rastgele orman ve öğrenme için gradyan artırma olmak üzere iki makine öğrenmesi yönteminin bir kombinasyonu uygulanmaktadır. Ayrıca veri kıtlığı sorununun aşılması amacıyla borsa endekslerinin tahmininde kanıtlanmış bir veri büyütme tekniği olan sütun bazında rastgele karıştırma yöntemi tercih edilmiştir. Sonuçlar, rastgele orman ve gradyan artırma kombinasyonunun %94,74'lük bir test doğruluğuna ulaştığını ve k-en yakın komşu, rastgele orman, karar ağacı, destek vektörü, saf bayes sınıflandırıcıları gibi en gelişmiş modellerden daha iyi performans gösterdiğini göstermektedir.*

*Corresponding author, e-mail: rdagistanli@kho.msu.edu.tr

DOI: 10.17134/khosbd.1492365

Elektronik/Online ISSN:1303-6831
Basılı/Printed ISSN:2148-1776

## 1. INTRODUCTION

The financial and technological innovations of the last two centuries, which are still ongoing, have caused several environmental, social, and economic problems [1, 2]. In today's world, cities and businesses consume more than two-thirds of the world's energy, while producing 80% of greenhouse gas emissions and 50% of the world's waste. This has highlighted the importance of addressing environmental issues such as uncontrolled consumption of natural resources, pollution, global warming and climate change. In terms of sustainability, it is critical for businesses to reduce the environmental damage they cause in their operations or to operate without impacting the environment. Various strategies, in particular the circular economy, have been developed and used by companies to achieve this goal. The circular economy is a business concept that aims to reuse and recycle products and reduce resource consumption. It promotes economic growth, employment and environmental quality [3]. Before discussing the notion of circular economy (CE), it is necessary to discuss linear economy (LE), which played a significant role in the development of this concept. The linear economy is a type of economy that emerged with the industrial revolution, emphasizing the processing of resources, their conversion into goods, and their disposal at the end of their useful life. With this approach, known as buy-use-dispose, it is understood that resources are employed indefinitely to manufacture items, and products are removed from usage with the intention of "buying better" before they expire. This economic approach has resulted in numerous environmental issues. Among these environmental issues include resource depletion, waste accumulation, and an increase in emissions [4, 5].

The circular economy, on the other hand, is a new economic model that seeks to reduce raw material consumption, waste, and value chain hazards by keeping the resources used in production in the production cycle for as long as possible. As a result, the circular economy appears to be both a sustainable economy and a new economic opportunity [6]. According to the European Action Plan, the circular economy is an economy in which products and materials have value. In this economy, resources are protected for as long as possible and waste generation is limited [7]. This new model, which is widely seen as an alternative to the linear economy, incorporates concepts that are critical for all businesses working within the framework of sustainable development. It focuses on reducing a company's use of resources through reduce, reuse, and recycle (3R principle), ensuring the long-term use of products, and recycling and reusing waste products [8]. Recycling, rethinking, redesign, product renewal and repair are now included in these principles, and it is emphasized that companies carry out their activities in a cycle by developing extended Rs [9].

The circular economy represents a novel approach to more sustainable social and sectoral growth. The circular economy also aims to increase regional competitiveness and achieve a more equitable distribution of economic growth and wealth. The circular economy and the sustainability approach provide the basis for delivering long-term benefits by ensuring an economic system capable of generating long-term growth and income and functioning with high financial performance (FP) [10-12]. As a result, a deeper exploration of the idea of sustainability is needed. Furthermore, as of 2019 [13], Turkey has started to adopt the circular economy by linking it to sustainability efforts.

Sustainability has recently been considered in many different sectors [2-13]. The concept of sustainability has evolved through several stages to reach its current prominence. Globally, there have been rapid technological breakthroughs since the 1950s. Some challenges, such as environmental pollution and energy depletion, have become socially remarkable as a result of these advances in the manufacturing sector. From this perspective, the declaration adopted at the 1972 UN Conference on the Human Environment in Switzerland could be used to discuss an international awareness of sustainability. With the study [14], the concept was recognized by a wide range of circles, especially in the field of economics. According to the report, the most common definition of sustainable development is the use of resources without compromising the needs of the present while considering future generations. Following the publication of the study, the Coalition for Environmentally Responsible Economies (CERES) principles were signed in 1989, the Earth Summit was held in 1992, and the Kyoto Protocol was signed in 1997 as part of the United Nations Framework Convention on Climate Change. The Organization for Economic Cooperation and Development (OECD) Declaration on International Investment and Multinational Enterprises in 2000, the World Business Council on Sustainable Development in 2001, the OECD Parliamentary Commission on the Environment in 2002, and the United Nations Conference on Sustainable Development in 2012 have all focused on global sustainability.

The incorporation of the concept into business management practices is a result of the global importance given to sustainability in the national context. The idea of corporate sustainability has evolved from this perspective. The goal of the business system is to keep the economy viable. However, by considering difficulties related to environmental and social phenomena, rather than just economic objectives, it will be possible to ensure the managerial integration of sustainability [15]. Once integrated, corporate sustainability will no longer be a management strategy, but an integral part of the company's culture [16]. Stakeholder theory, corporate social responsibility, corporate accountability theory, and sustainable development concepts have all been examined in studies of the management concept that internalizes corporate sustainability [17]. Stakeholder theory and legitimacy theory serve as the basis for corporate sustainability. The owners of companies, employee unions, customers, suppliers, competitors, and society are the stakeholders that give the first hypothesis its name. According to stakeholder theory, in addition to meeting the expectations of owners and shareholders, companies should also treat other stakeholders with sufficient transparency [18]. On the other hand, according to legitimacy theory, a firm's actions must be in the best interest of society as a whole in order for its legitimacy to be recognized [19].

The challenge of measuring sustainability in terms of performance has emerged with the integration of corporate sustainability into management processes and the implementation of information exchange with stakeholders without manipulation. Elkington [20] introduced the idea of the Triple Bottom Line (TBL), also referred to as the "Three Pillars of Sustainability", into the literature to be used in this research. According to [21], TBL refers to an approach that considers performance in all three areas of economic, social and environmental aspects. Companies will have achieved their sustainability goal if they approach and apply this three-dimensional performance in a comprehensive and interactive manner.

Sustainability reporting (SR) is used to collect and analyze performance data in TBL dimensions and to communicate with stakeholders. Companies use SR to maintain effective communication with stakeholders by demonstrating the links between their actions and sustainability for their information [22-24]. The ever-increasing number of stakeholders and their growing expectations put pressure on companies to maintain SR, and the fact that management openness is provided also promotes the willingness to do so [25]. Companies that prepare SR work with the goal of being able to achieve success in terms of FP while at the same time being transparent to stakeholders.

Companies' research on sustainability has also shown its impact on the country's stock markets over time. The sustainability index, the first examples of which were observed in developed countries in the 1990s, is one of the mechanisms through which stakeholders evaluate the sustainability of organizations [26]. On November 4, 2014, the sustainability index was included in the indices of Borsa Istanbul (BIST) in Turkey. In the first BIST Sustainability Index

(XUSRD) in Turkey, there were 15 companies; by 2021, this number had increased to 65. Companies can use the index to assess their national and international sustainability performance. Previous research has shown a correlation between the XUSRD and the BIST100, with both indices showing a high degree of similarity in their behavior [27]. Furthermore, it has been found that the sustainability efforts undertaken by the companies have a positive impact on the mission, vision, strategy reports and financial reports. According to Mumcu and Ufacık [28] and Yılmaz et al. [29], the inclusion of companies in the XUSRD index has been found to lead to an acceleration of their sustainability initiatives and an increased focus on reporting.

Both anecdotal and empirical evidence support the notion that increased social and environmental reporting can improve a company's financial performance and market value [30]. This study estimates the effect of FP on SR and the publication status of the SR report by examining the relationship between SR and FP. The sample group of the study consists of XUSRD companies. Due to their structure, financial companies (holding companies, banks, insurance companies, and REITs) included in XUSRD were excluded from the sample group [31]. Firms in XUSRD are classified into two groups based on their SR preparation or non-preparation status. Profit, return on assets (ROA), return on equity (ROE), return on sales (ROS), size, and financial leverage (LVG) are considered as FP indicators. SR was chosen as the decision variable of the model.

The objective of this study is to analyze the importance that companies attribute to SR and its impact on publication status, as evidenced by FP indicators. Numerous studies have been conducted to investigate the relationship between the variables under consideration, as evidenced by the existing literature. This study makes several distinct contributions to the existing body of knowledge. The study makes several significant contributions, which can be summarized as follows: The FP indicators discussed are not dependent on market forces and are immune to manipulation. The studies conducted in the literature used a limited dataset due to their association with the early years of the XUSRD. In contrast, this study used a more extensive and recent dataset. In this context, a comprehensive study was conducted on a sample group of 44 companies, covering the period from 2014 to 2021. The analysis does not include data for 2022, as it is currently unpublished. Rather than using classical statistical analysis based on a single method, as is common in the existing literature, this study used comparative analysis through the application of artificial intelligence techniques. To date, no other empirical investigation has been conducted that examines the correlation between SR and FP using machine learning algorithms. The existing literature on the relationship between circularity and sustainability has been enhanced by the inclusion of the FP relationship.

The remainder of the paper is organized as follows. An extensive literature review is presented in the second part of the study. Data and methodology are presented in the third section. The methods used and the results of the analysis are presented in the fourth section, and the results and comments are presented in the fifth section.

## 2. LITERATURE REVIEW AND RELATED WORK

A review of the literature indicates that the terms social responsibility report, corporate social responsibility report, and SR are used interchangeably in most studies. The relationship between SR and FP is actively debated in a variety of countries, industries, and years [15, 32, 33]. There are studies that examine the relationship between these two variables as positive, negative, neutral, or no relationship based on effect. There are also studies that evaluate the FP indicators as statistically significant or insignificant. Comincioli, Poddi, and Vergalli [34] conducted a study to examine the relationship between corporate social responsibility and financial performance. The sample consisted of 317 companies that were included in at least two of three sustainability indices and a control group of 100 companies that were not included in these indices. The analysis showed that the companies included in the XUSRD index had higher market value added than those not included in the index.

Studies in the literature show that SR analysis can be included in the analysis as 1 if a report has been published by the company for the relevant year and 0 otherwise. For example, Kuzey and Uyar [35] investigated the determinants of sustainability reporting practices of companies listed on BIST. In addition to the Leverage, Size and ROA variables, the SR variable takes the value of 1 if the company publishes a sustainability report, and 0 otherwise. Kim and Kim [36] evaluated the benefits of corporate sustainability management activities, including sustainability reporting, to stakeholders with a similar approach. Conducting their research, they concluded that profitability does not significantly affect sustainability activities, but company size does.

As a second step, studies focusing on sustainability and circular economy were discussed. In Güngör's study [37], a detailed literature review was conducted. According to this study, there are descriptive-explanatory and practical studies on circular economy in the literature, as well as studies that evaluate the relationship between circular economy and sustainability on a theoretical level, examine sustainable business models, and focus on the analysis of sustainability reporting. In addition, the study examined the SR status of 27 companies operating in the food, beverages, chemicals, petroleum, plastics, metals, textiles, leather, mining and forestry, and paper and printing sectors of the BIST100 in 2021. The results obtained are in line with the theory that the circular economy can be analyzed with SR, which was put forward in the research carried out in the textile sector by Stewart and Niero [38].

Despite the growth of multi-variable assessment approaches, the application of supervised and unsupervised learning techniques to sustainability assessment remains largely unexplored [39]. However, many machine learning methods have been used to make predictions on various topics in the stock market [40-44]. Nayak et al. [45] integrated K-Nearest Neighbor (kNN) and Support Vector Machine (SVM) methods to predict the Indian stock market index. The proposed model was found to have partially successful prediction ability based on high-dimensional data. Chen and Hao [46] used a hybrid of feature weighted SVM and feature weighted kNN methods to predict the Chinese stock market index. Ayala et al. [47] tested the performance of Linear Model (LM), Artificial Neural Network (ANN), Random Forests (RF) and Support Vector Regression (SVR) techniques to predict trading signals in three different stock market indices (IBEX, DAX and DJI). Park et al. [48] proposed the LSTM-Forest model by integrating Long Short Term Memory (LSTM) and RF techniques on three different stock market indices (S&P500, SSE, and KOSPI200). The daily forecasting success of their proposed model was more successful than the RF algorithm. Similarly, hybrid learning methods combined with statistical models or other techniques have also been used in stock market research [49-53].

As a result of a two-step literature review; (i) the increasing trend in the studies is on the effect-based analysis of the relationship between various control variables and SR-FP; (ii) the studies on circular economy and sustainability are focused on the descriptive-explanatory or theoretical level, and the analytical application is extremely rare; (iii) numerical analyses are performed using statistical methods. (iv) It has been observed that there is no study on machine learning methods, which is one of the most recent and superior methods in the academic literature.

## 3. DATA AND RESEARCH METHODOLOGY

### 3.1. Data source and description

Companies include detailed information on all their sustainability activities in their SR reports. In this way, both the circular economy and various issues can be examined through these reports. With the global importance of sustainability activities, awareness of this issue has increased in Turkey over the years and XUSRD has been established by BIST. In this study, the data of 44 companies, excluding companies operating in the financial sector (holding, banking, insurance, real estate), included in XUSRD between 2014 and 2021 were collected through the websites of the companies and the Public Disclosure Platform.

The study used FP indicators as independent variables; they can be classified into three types: market-based, accounting-based, and perceptual measures [54]. The perceptual measures of these indicators are subjectively determined by the company's survey ratings. A review of the literature on FP indicators shows that accounting-based indicators have been widely used between 2002 and 2011 because they are objective and verifiable. These studies are followed by studies using partially objective market-based indicators, and the rate of perceptual studies for direct subjective evaluations is very low [55]. In addition to the objectivity of accounting-based measures, the reason for using market-based measures is that they incorporate stakeholders' future expectations for a company's stock [56-58]. ROA, ROE, ROS, and size and financial leverage, which are accounting-based measures with literature consensus on profitability objectivity, were used as variables in this study and are shown in Table 1.

**Table 1:** Explanations of the Variables.

| Symbol | Description/Calculation |
|---|---|
| SR | 1 if the firm publishes a sustainability report, 0 otherwise |
| Profit | Profit |
| ROA | Net Profit/Total Assets |
| ROE | Net Profit/Equity |
| ROS | Net Profit/Total Sales |
| Firm Size | Natural Logarithm of Total Assets |
| lvg | Financial Leverage = Total Liabilities/Equity |

In the literature, Vitezic, Vuko and Mörec [59] examined the effect of FP-SR with ROA, ROE and size variables. Data from Croatian companies from 2002 to 2010 were analyzed using logistic regression. According to the data of Croatian companies, both profitability variables (ROA, ROE) and size variable for FP were found to be effective in publishing SR. Dağıstanlı and Çelik [60] used logistic regression analysis in their study with data between 2014-2021 for companies in XUSRD in Turkey. It was found that the size variable was significantly effective in SR publication, but the profitability variables were not. Firms in the BIST 100 sustainability index were classified by cluster analysis according to their environmental, social and governance scores (ESG) [61].

### 3.2. Data Augmentation Technique: Column-Wise Random Shuffling

In this study, machine learning techniques have been used rather than a single statistical method, as is often the case in the literature on sustainability and financial performance. In many cases where a machine learning method is used, lack of data may not be an issue. However, these models are not suitable for learning patterns from high-dimensional data sets [62]. A large amount of data does not always lead to accurate results. As a result, various feature selection and feature extraction methods are applied to the inputs to reduce the data size. However, since features are learned directly from the data, reducing the input size may result in data loss and poor performance. In addition, the small data set causes the training data to be recalled, and there may be a lack of fit. Using various regularization methods in the data, it is possible to avoid unnecessary repetitions and over fitting [63].

As in this study, data augmentation is one of the best ways to make sense of and avoid fabrication when training datasets are significantly small in industries such as finance. Data augmentation methods are used in many different areas in the literature [64, 65]. For financial time series input datasets, an approach called column-wise random shuffling is proposed, which does not distort the original input dataset [66]. An example of column-wise random shuffling using variables (V), firms (C), and years (T) is visualized in Fig. 1. First, data for 6 variables for 44 companies included in XUSRD were collected over an 8-year period, as shown in Fig. 1. For example, all 8-year values of the first variable of company 5 are reserved. New data sets (represented by cubes in Fig. 1) were obtained by randomly mixing these values on a column-by-column basis. This data increment is then iterated for all V values of each company. All new layers are designed to focus on carrying the common features of each company. It is assumed that a random mixing of all relevant V values will not reflect the characteristics of the company. Therefore, the data augmentation is limited to all company-specific V values recorded for the T period. The number of different data sets that can be generated by random shuffling on a column-by-column basis is nearly infinite. The advantage of this situation is that many similar data sets can be generated, while the disadvantage is that data augmentation with such variation can confuse the model and make learning very difficult.

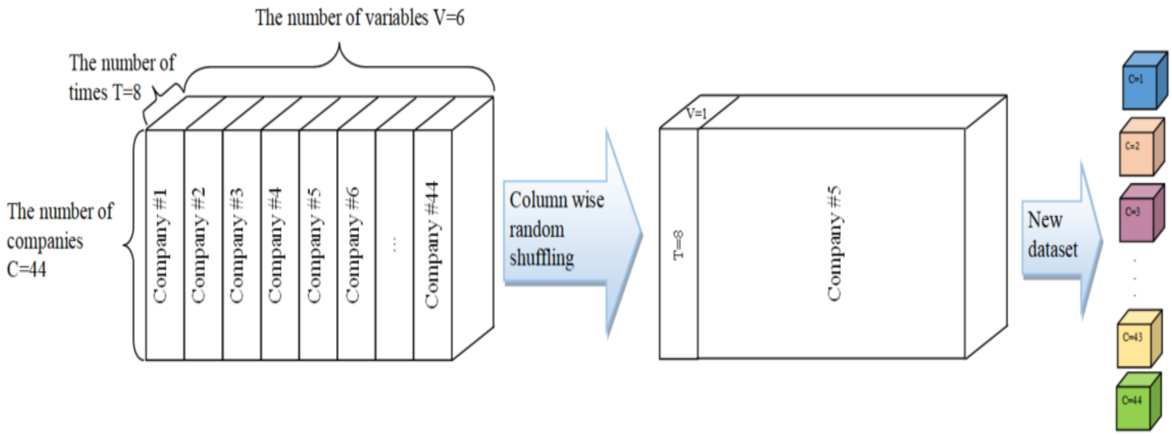



**Figure 1**:An Example of Column-Wise Random Shuffling In XUSRD.

## 4. MACHINE LEARNING METHODS USED FOR THE PREDICTION OF SUSTAINABILITY INDEX

In this study, six different algorithms are used to develop six models for classifying and predicting the sustainability indices of the companies in the BIST100. These methods are naive Bayes, k-nearest neighbor, support vector, decision tree, random forest and gradient boosting classifiers.

### 4.1. Naive Bayes Classifier

This classifier assumes the independence of all the features, hence it is naive. For the features 1 to N $(f1, f2, \dots fN)$, and the $class\ label_i$, where i =1 or 2 for the binary (two-class) problem, the probability of a particular class label given the features can be calculated as in Eq. 1.

### 4.2. K-Nearest Neighbor Classifier

This method measures the distance between the unknown sample and its neighbors. An evaluation is made according to the number of preselected neighbors, and hence k is predefined, and the new sample is categorized as a member of the class to which the majority of the nearest neighbors belong. Different numbers of neighbors can be used and this number is a hyper parameter for the algorithm. It is possible to make a grid search and determine the best k. Fig. 2 shows how the k-nearest neighbor classifier works; in the figure, if k=5, then the classification for the blue triangle which represents the data whose class is unknown, will be determined as the class of the flag sign, since there are four flag signs and one face sign in the five-neighborhood of the blue triangle, whereas if k=11, then the class will be face sign.

$$P(class\ label_i|f1, f2, \dots, fN) = \frac{P(f1, f2, \dots, fN|class\ label_i)P(class\ label_i)}{P(f1, f2, \dots, fN)} \quad (1)$$

If all the features are conditionally independent given the $class\ label_i$:

$$P(f1, f2, \dots, fN|class\ label_i) = P(f1|class\ label_i) \dots P(fN|class\ label_i) \quad (2)$$

Combining Eq. 1 and Eq. 2 gives

$$P(class\ label_i|f1, f2, \dots, fN) = \frac{P(class\ label_i)\prod_{j=1}^{N} P\left(f_j|class\ label_i\right)}{P(f1, f2, \dots, fN)} \quad (3)$$

The classification done by the naïve Bayes canbe summarized by Eq. 4:

$$class\ label = argmax_{i\in\{1,\dots,k\}} P(class\ label_i | f1, f2, \dots, fN) \quad (4)$$

The class labels are assigned using the above algorithm [67]. Naive Bayes model can be used to cluster the data, and hence for classification.

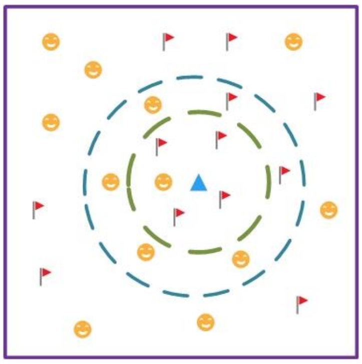


**Figure 2:** Practical Use of kNN.

### 4.3. Support Vector Classifier

Support vector machines, when used as classifiers, employ hyper plane approach for achieving their purpose [68]. The hyper planes are selected to maximize the margin for the given data. The maximization is needed to separate the given classes as apart as possible. A simple support vector classifier for a two-class problem is shown in Fig. 3. The vectors of the closest data points to the hyper plane are called the support vectors [69].

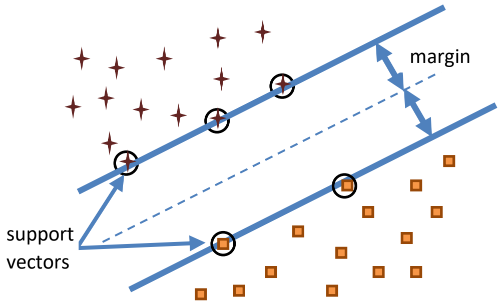



**Figure 3:** Simple Support Vector Classifier For A Two-Class Problem.

### 4.4. Decision Tree Classifier

The idea in a decision tree is to ask successive questions and form branches and sub branches to answer them. When no more questions can be asked, sub branching stops. The point at which no more sub branches are possible is called a leaf node (Fig. 4). The first question at which tree formation starts is called the root [70]. A decision tree classifier can easily over fit because it only learns through a single path and the accuracy decreases on a new sample.

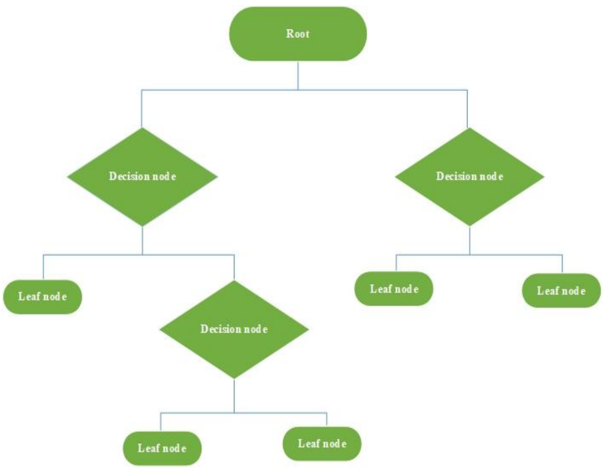



**Figure 4:** Simple Decision Tree Classifier.

### 4.5. Random forest classifier

Leo Breiman described a random forest classifier as a collection of tree-structured classifiers. Each of these classifiers is treated as an independent and identically distributed random vector. Although variations are possible, the most popular class is where each tree has a single vote, as shown in Fig. 5 [71].

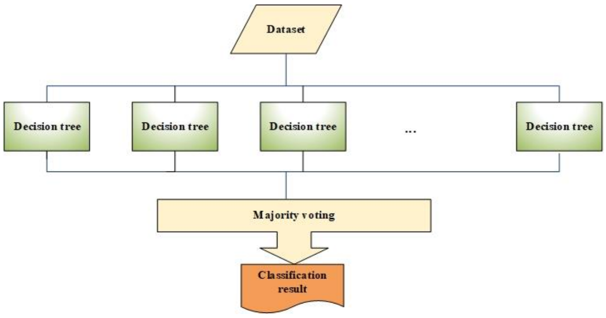



**Figure 5:** Random Forest Classifier.

### 4.6. Gradient boosting classifier

A gradient boosting classifier starts from the steepest descent algorithm but forms tree structures to create better classifications, using the concept of boosting [72]. The basic idea is illustrated in Fig. 6.

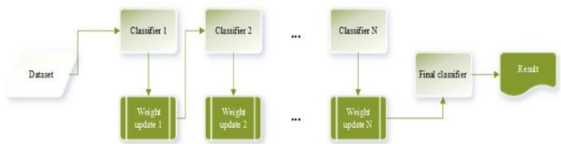



**Figure 6:** Gradient Boosting Classifier.

### 4.7. The Proposed Method

The proposed method uses a random forest classifier to select the features to be used by the machine learning algorithm to be applied later. Once the features are determined, a gradient boosting algorithm is used for classification. The classification algorithm is used to predict the sustainability indices of the companies whose data were not introduced to the algorithm during the training process. The pipeline of the proposed method is shown in Fig. 7.

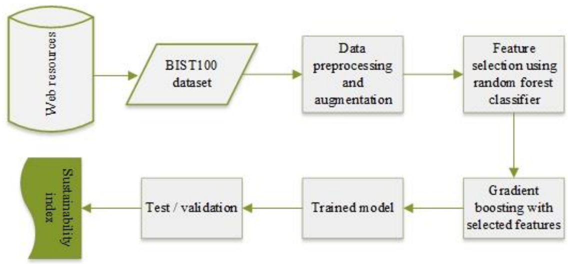



**Figure 7:** Pipeline For The Proposed Method.

#### 4.7.1. Performance criteria

A number of classical and modern machine learning algorithms are used for prediction. Their performance is compared on the basis of training and test accuracy, precision, recall and F1 scores.

For binary classification, as in this work, the performance metrics are calculated using the confusion matrix given in Table 2.

**Table 2:** Confusion Matrix Of A Binary Classification Problem.

| | | Predicted | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| **Actual** | **Positive** | True Positive: ***TP*** | False Negative: ***FN*** |
| | **Negative** | False Positive: ***FP*** | True Negative: ***TN*** |

Using the confusion matrix, a series of performance metrics can be defined [68]. For example, accuracy (acc) is a common metric and is defined by Eq. 5:

$$acc = \frac{TP+TN}{TP+FN+FP+TN} \tag{5}$$

Another metric that can be defined using the confusion matrix is the ***precision*** (Eq. 6):

$$precision = \frac{TP}{TP+FP} \tag{6}$$

Also, recall can be derived from the confusion matrix and is defined by Eq. 7:

$$Recall = \frac{TP}{TP+FN} \tag{7}$$

F1-score can be obtained from precision and recall using their harmonic mean. The definition is given by Eq. 8:

$$F1 - score = \frac{2TP}{2TP+FP+FN} \tag{8}$$

In this work, accuracy, precision, recall and F1-score values for training and test are calculated and used for comparing the models obtained.

## 5. EXPERIMENTAL RESULTS

The simulations in this study are made using 12th Gen Intel Core i7-12650H Processor at 2.3 GHz with 16 GB of RAM and NVIDIA GeForce RTX 3060. The programming language is Python.

Tha data is cleaned and augmented using column-wise random shuffling method. The data file has 911 rows of data. The columns contain profit, roa, roe, ros, size, lvg and sustainability data. To give an idea about the content of data, the first five rows of the file are shown in Table 3. The statistical properties of the variables in the file are as calculated and shown in Table 4.

The dataset is almost balanced as can be seen from Fig. 8. There are 452 unsustainable and 459 sustainable companies.

**Table 3:** The First Five Rows Of The Data File Used In Experiments.

| nr | profit | roa | roe | ros | size | lvg | sustainability |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0.1562 | 0.2165 | 0.1774 | 21.1948 | 0.3864 | 0 |
| 1 | 1 | 0.1660 | 0.2357 | 0.1927 | 21.2569 | 0.4202 | 0 |
| 2 | 1 | 0.1557 | 0.2167 | 0.1774 | 21.1948 | 0.9507 | 0 |
| 3 | 1 | 0.0832 | 0.0951 | 0.0411 | 21.9769 | 1.0147 | 0 |
| 4 | 1 | 0.1557 | 0.2338 | 0.0980 | 21.6119 | 0.4202 | 0 |

**Table 4:** The Statistical Properties Of The Variables In The Dataset.

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **profit** | 911 | 0.8342 | 0.3721 | 0.0 | 1.0 | 1.0 | 1.0 | 1.0 |
| **roa** | 911 | 0.0547 | 0.0775 | -0.2728 | 0.0181 | 0.0580 | 0.0998 | 0.3071 |
| **roe** | 911 | 0.0688 | 1.4762 | -18.9341 | 0.0561 | 0.1521 | 0.2559 | 11.1163 |
| **ros** | 911 | 0.0570 | 0.1247 | -0.7024 | 0.0115 | 0.0572 | 0.1093 | 0.5251 |
| **size** | 911 | 22.6309 | 1.3082 | 18.9284 | 21.7194 | 22.5133 | 23.7155 | 26.5917 |
| **lvg** | 911 | 2.9062 | 32.2149 | -213.4637 | 0.9011 | 1.7677 | 3.0011 | 458.5816 |
| **sustainability** | 911 | 0.5038 | 0.5003 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |

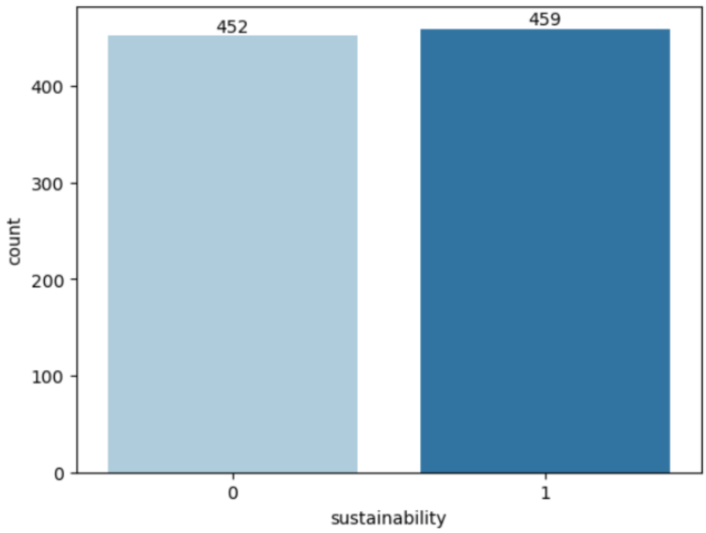



**Figure 8:** Distribution Of Classes In The Dataset.

### 5.1. The selection of features using machine learning

The comparison of attribute importance using random forest classifier is shown in Fig. 9. The maximum depth of the random forest classifier used in the attribute importance comparison is 100. It is clear from Fig. 9 that the attribute "profit" should not be used as a feature because it has the lowest importance of all. The remaining attributes have very close importance levels and they are used as features to determine the sustainability index.

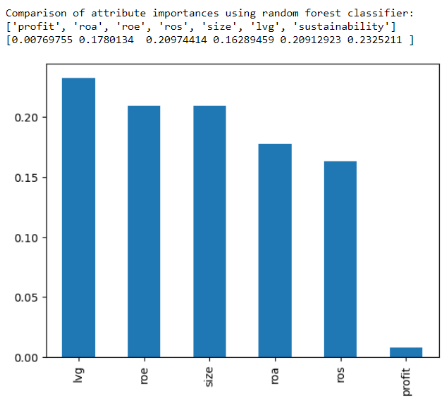



**Figure 9:** Attribute Importance Rating Of The Dataset Using Random Forest Classifier.

To predict the sustainability indices of the companies in the BIST100, data is manually collected from the companies' websites. Since this data is only available for the last few years, the data set is very small. The modern algorithms, especially the deep algorithms, need a large amount of data for prediction. Therefore, in this study, a data augmentation technique, namely column-wise random shuffling, has been used to increase the size of the data. The total size is 911 rows, where the sustainability index is zero for 452 rows of data and one for 459 rows of data.

To predict the sustainability index for the dataset, 98% of the data is used for training and 2% for testing. Most of the data is used for training simply because deep models have many parameters to determine and these parameters can be defined using a lot of data. The results are shown in Table 5 in ascending order as far as the test accuracy, precision, recall and F1 score values are concerned. The same order is observed for training, except for the fourth and fifth rows. It is clear from Table 4 that the best results are obtained using the gradient boosting classifier (GBC). For example, the test accuracy is 12.50% better than its closest competitor, the k nearest neighbor classifier (kNN). On the other hand, the worst accuracy is obtained by a classical algorithm, namely naive Bayes classifier (NBC). A comparison of GBC and NBC shows that GBC has a 62.66% higher test accuracy.

Although the test accuracies of the decision tree classifier (DTC) and the random forest classifier (RFC) are the same, the training accuracy of the random forest classifier (RFC) is better. The same discrepancy is observed for precision, recall and F1 score in training, while the opposite is true for test values. The test scores of k nearest neighbor (kNN) classifier are better than both DTC and RFC, although the training scores of kNN may be a little worse than RFC. Test performance is given a higher priority in this work.

All models obtained in this work are optimized by grid search in simulation. The support vector classifier uses a radial basis function. The nearest neighbor classifier uses 3 neighbors for classification. The decision tree and random forest classifiers use a maximum depth of 5. The gradient boosting classifier uses 100 estimators with a maximum depth of 5.

**Table 5:** Simulation Results For Various Classifiers For The Given Dataset.

| Method | train acc | test acc | prec_train | prec_test | recall_train | recall_test | f1_train | f1_test |
|---|---|---|---|---|---|---|---|---|
| NBC | 57.06 | 57.89 | 0.6464 | 0.6146 | 0.5682 | 0.5611 | 0.5030 | 0.5128 |
| SVC | 63.12 | 73.68 | 0.6497 | 0.7386 | 0.6300 | 0.7333 | 0.6181 | 0.7339 |
| DTC | 78.92 | 78.95 | 0.7962 | 0.8036 | 0.7887 | 0.7833 | 0.7878 | 0.7841 |
| RFC | 84.87 | 78.95 | 0.8492 | 0.8036 | 0.8485 | 0.7833 | 0.8486 | 0.7841 |
| kNN | 83.97 | 84.21 | 0.8416 | 0.8846 | 0.8394 | 0.8333 | 0.8394 | 0.8348 |
| GBC | 99.89 | 94.74 | 0.9989 | 0.9545 | 0.9989 | 0.9444 | 0.9989 | 0.9468 |

To examine the performance of the GBC in more detail, the confusion matrices in training and test are drawn. The results are shown as percentages. In Fig. 10, it is seen that in training phase none of the sustainability “1” data is classified incorrectly, whereas 0.11% of all the data are classified incorrectly as sustainability “1”.

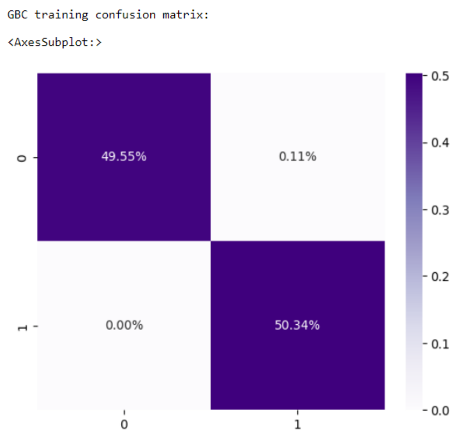



**Figure 10:** GBC Training Confusion Matrix. Values Are Shown In Per Cent.

On the other hand, the test confusion matrix shows more error. Unsustainable data are classified correctly except for 5.26% of all data. Sustainable data is correctly classified (Fig. 11).

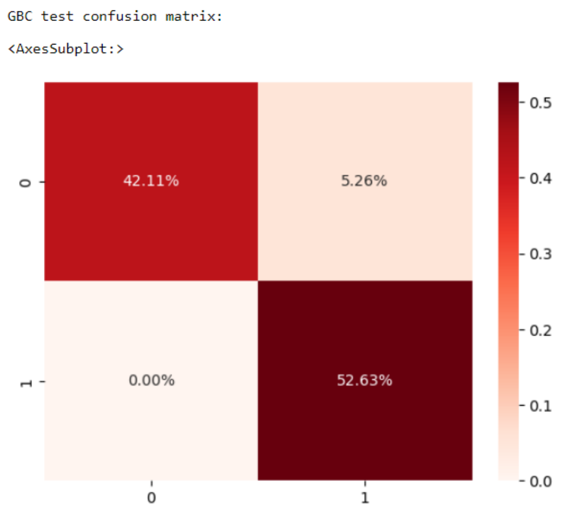



**Figure 11:** 10 GBC Test Confusion Matrix. Values Are Shown In Per Cent.

The number of (in)correctly classified test data is shown in Fig. 12. Eight of the nine 'unsustainable' data are correctly classified, while all ten 'sustainable' data are correctly classified.

| | |
|---|---|
| 8 | 1 |
| 0 | 10 |

**Figure 12:** Confusion Matrix Of Test Data For GBC.

Test accuracy, macro and weighted averages of precision, recall and F1-score values are shown in Fig. 13. It can be seen that the precision value of the unsustainable data is better than the sustainable data, but the recall and F1-score values are better for the sustainable data.

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| **0** | 1.00 | 0.89 | 0.94 | 9 |
| **1** | 0.91 | 1.00 | 0.95 | 10 |
| **accuracy** | | | 0.95 | 19 |
| **macro ave** | 0.95 | 0.94 | 0.95 | 19 |
| **weighted ave** | 0.95 | 0.95 | 0.95 | 19 |

**Figure 13:** Classification Of Test Data For GBC.

Since the errors are within acceptable limits, the proposed combination of random forest classifier for feature selection and gradient boosting classifier for prediction of sustainability indices for BIST100 companies gives successful results.

## 6. CONCLUSION

Companies publish reports containing primarily economic data to facilitate the flow of information to their stakeholders. World developments have shown that economic reporting alone cannot ensure the continuity of this flow. Companies have begun to reflect their sensitivity to participate in sustainability activities, as well as their accountability and transparency to their stakeholders, by preparing reports on their positive social and environmental policies. According to the literature, the basic elements of circular economy are at an extremely high level in these reports [37]. Despite all the advantages, the preparation of a SR is not required by any legal regulation in Turkey. As a result, the factors that influence whether companies publish SRs or not have become a focus of research. In recent years, attention has been drawn to the relationship between SR and FP.

Over time, stock market indices have been established, first in the capital markets of developed countries, where companies are classified into a specific sustainability category. Since 2014, XUSRD has been conducting assessments in Turkey for this specific purpose. Companies' participation in the circular economy encompasses various benefits, such as enhancing reputation, improving corporate image, increasing revenues, increasing profit margins, reducing costs, and promoting employee motivation, among others. There are potentially numerous factors that could be considered. For the purposes of this framework, the sample group of the study has been identified as XUSRD.

This relationship in the XUSRD sample has been discussed using statistical methods from the literature. The study by Vitezic, Vuko, and Mörec [59], which examined the relationship between SR and FP using a logistic regression analysis, is noteworthy. This study was conducted in Croatia using data from 2002 to 2010. ROA, ROE and firm size were used as FP indicators. According to the data from Croatian companies, all three variables are effective in the publication of SR. In their study using 2014-2021 data, Dağıstanlı and Çelik [60] found that the profitability variables had no significant effect, while the size variable did. The studies had classification rates of %80 and %84, respectively.

In this study, contrary to the Turkish Stock Exchange literature, machine learning techniques were used for the first time to predict firms' participation in XUSRD using FP data. Company/year data were collected from the reports of 44 companies in XUSRD, which were studied between 2014 and 2021. The data shortage problem, which is a limitation of machine learning-based predictive modeling in the financial market, was overcome using column-by-column randomization, which is a simple and powerful data augmentation method. The proposed method in this paper is a combination of random forest and gradient boosting, and is compared with five other machine learning methods, namely, k-nearest neighbor, random forest, decision tree, support vector, naive Bayes classifier. The proposed method achieved a test accuracy of 94.74%, while the test accuracy of its closest rival, namely k-nearest neighbor, reached 84.21%. As a result, the SR-FP relationship has been discussed for the first time with machine learning techniques, unlike the methods used so far. The obtained results show that the method applied in the data collection is well suited.

Considering the capital market regulations of developed countries, the SR issuance policy in Turkey is expected to be subject to various regulations. In this way, published reports and cyclical economic data, TBL records and financial data will be made available to investors in a transparent manner. Forecasts made and to be made in the literature with statistical and machine learning-based methods will become much more meaningful in the future.

The analysis in this study has some limitations, and future research may be conducted on other aspects of the topic. The study used accounting-based measures to analyze the profitability data of the companies in the XUSRD. Market-based ratios, which are partially objectively evaluated, as seen in some studies in the literature, can be added to the independent variables. In addition, variables related to the corporate governance structure (number of supervisory and board members, male/female ratio of board members, etc.), as observed in some studies in the literature, can be included in the scope of research on the attitude towards the publication of SR. To understand the extreme changes of profitability variables in Turkey and some technical details, profitability variables of different countries can be compared with XUSRD companies.

## ACKNOWLEDGMENTS

This research received no external funding.

## AUTHOR CONTRIBUTIONS

**Hakan Ayhan DAĞISTANLI:** Conceptual Design, Data Organisation, Research, Drafting and Writing, Revising, Approval

**Figen Özen:** Analysis, Methodology, Software, Audit, Visualisation, Review, Validation

**İlkay Saraçoğlu:** Research, Project Management, Methodology, Writing, Experimental studies, Reviewing and Editing

## CONFLICTS OF INTEREST

The authors declare that they have no conflict of interest.

## REFERENCES

[1] Huang, K., Sim, N., & Zhao, H. (2020). Corporate social responsibility, corporate financial performance and the confounding effects of economic fluctuations: A meta-analysis. International Review of Financial Analysis, 70, 101504. https://doi.org/10.1016/j.irfa.2020.101504

[2] Alatawi, I. A., Ntim, C. G., Zras, A., & Elmagrhi, M. H. (2023). CSR, financial and non-financial performance in the tourism sector: A systematic literature review and future research agenda. International Review of Financial Analysis, 102734. https://doi.org/10.1016/j.irfa.2023.102734

[3] OECD(2021).https://www.oecd.org/cfe/regionaldevelopment/Circular_Economy_Flyer.pdf

[4] Ellen Macarthur Foundation. (2010). Toward the Circular Economy.

[5] Remo-Diez, N., Mendaña-Cuervo, C., & Arenas-Parra, M. (2023). Exploring the asymmetric impact of sustainability reporting on financial performance in the utilities sector: A longitudinal comparative analysis. Utilities Policy, 84, 101650. https://doi.org/10.1016/j.jup.2023.101650

[6] Geisendorf, S., & Pietrulla, F. (2018). The circular economy and circular economic concepts-a literature analysis and redefinition. Thunderbird International

Business Review, 60(5), 771-782. https://doi.org/10.1002/tie.21924

[7] European Commission. (2015). Closing the Loop- An EU Action Plan for the Circular Economy. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Brussels.

[8] Kirchherr, J., Reike, D., & Hekkert, M. (2017). Conceptualizing the circular economy: An analysis of 114 definitions. Resources, conservation and recycling, 127, 221-232. https://doi.org/10.1016/j.resconrec.2017.09.005

[9] Reike, D., Vermeulen, W. J., & Witjes, S. (2018). The circular economy: new or refurbished as CE 3.0?-exploring controversies in the conceptualization of the circular economy through a focus on history and resource value retention options. Resources, conservation and recycling, 135, 246-264. https://doi.org/10.1016/j.resconrec.2017.08.027

[10] Uhrenholt, J. N., Kristensen, J. H., Rincón, M. C., Jensen, S. F., & Waehrens, B. V. (2022). Circular economy: Factors affecting the financial performance of product take-back systems. Journal of Cleaner Production, 335, 130319. https://doi.org/10.1016/j.jclepro.2021.130319

[11] Rodríguez-González, R. M., Maldonado-Guzman, G., Madrid-Guijarro, A., & Garza-Reyes, J. A. (2022). Does circular economy affect financial performance? The mediating role of sustainable supply chain management in the automotive industry. Journal of Cleaner Production, 379, 134670. https://doi.org/10.1016/j.jclepro.2022.134670

[12] Orsini, L. P., Leardini, C., Danesi, L., Guerrini, A., & Frison, N. (2023). Circular economy in the water and wastewater sector: Tariff impact and financial performance of SMARTechs. Utilities Policy, 83, 101593. https://doi.org/10.1016/j.jup.2023.101593

[13] Amankwah-Amoah, J. (2020). Stepping up and stepping out of COVID-19: New challenges for environmental sustainability policies in the global airline industry. Journal of Cleaner Production, 271, 123000. https://doi.org/10.1016/j.jclepro.2020.123000

[14] Bruntland, G.H. (1987). Our Common Future, Report of The World Commission On Environment And Development. http://www.un-documents.net/ourcommon-future.pdf

[15] Shin, J., Moon, J. J., & Kang, J. (2023). Where does ESG pay? The role of national culture in moderating the relationship between ESG performance and financial performance. International Business Review, 32(3), 102071. https://doi.org/10.1016/j.ibusrev.2022.102071

[16] Signitzer, B., & Prexl, A. (2007). Corporate sustainability communications: Aspects of theory and professionalization. Journal of Public Relations Research, 20(1), 1-19. https://doi.org/10.1080/10627260701726996

[17] Lee, M. T., & Raschke, R. L. (2023). Stakeholder legitimacy in firm greening and financial performance: What about

greenwashing temptations? Journal of Business Research, 155, 113393. https://doi.org/10.1016/j.jbusres.2022.113393

[18] Chiong, P.T. (2010). An examination of corporate sustainability disclosure level and its impact on financial performance. Doctor of Philosophy Multimedia University.

[19] Welter, K. A. (2011). A study of publicly-held US corporations on the effects of sustainability measures on financial performance, utilizing a modified regression discontinuity model. Lawrence Technological University.

[20] Elkington, J. (1997). Cannibals with forks: The triple bottom line of 21st century business, Oxford: Capstone Publishing.https://doi.org/10.1002/tqem.3310080106

[21] Elkington, J. (2004). Enter the triple bottom line. http://www.johnelkington.com/archive/TBL-elkington-chapter.pdf

[22] Herzig, C., & Schaltegger, S. (2011). Corporate sustainability reporting. Sustainability communication: Interdisciplinary perspectives and theoretical foundation, 151-169. https://doi.org/10.1007/978-94-007-1697-1_14

[23] Albertini, E. (2013). Does environmental management improve financial performance? A meta-analytical review. Organization & Environment, 26(4), 431-457. https://doi.org/10.1177/1086026613510301

[24] Haffar, M., & Searcy, C. (2017). Classification of trade-offs encountered in the practice of corporate sustainability. Journal of business ethics, 140, 495-522. https://doi.org/10.1007/s10551-015-2678-1

[25] Gao, Y. (2011). CSR in an emerging country: a content analysis of CSR reports of listed companies. Baltic Journal of management, 6(2), 263-291. https://doi.org/10.1108/17465261111131848

[26] Orsato, R. J., Garcia, A., Mendes-Da-Silva, W., Simonetti, R., & Monzoni, M. (2015). Sustainability indexes: why join in? A study of the 'Corporate Sustainability Index (ISE)'in Brazil. Journal of Cleaner Production, 96, 161-170. https://doi.org/10.1016/j.jclepro.2014.10.071

[27] Vardari, D. S. L., & Gashi, R. (2020). The impact of corporate sustainability index on BIST sustainability index. European Journal of Sustainable Development, 9(2), 375-390. https://doi.org/10.14207/ejsd.2020.v9n2p375

[28] Mumcu, A. Y., & Ufacık, O. E. (2016). A research on sustainability indices: BIST Sustainability Index. Social and economic perspectives on sustainability, 264-269.

[29] Yilmaz, M. K., Aksoy, M., & Tatoglu, E. (2020). Does the stock market value inclusion in a sustainability index? Evidence from Borsa Istanbul. Sustainability, 12(2), 483. https://doi.org/10.3390/su12020483

[30] Reddy, K & Gordon, L.W. (2010). The effect of sustainability reporting on financial performance: An empirical study using listed companies. Journal of Asia Entrepreneurship and Sustainability, 6(2), 19-42.

[31] Burhan, A. H. N., & Rahmanti, W. (2012). The impact of sustainability reporting on company performance. Journal of

Economics, Business, & Accountancy Ventura, 15(2), 257-272. https://doi.org/10.14414/jebav.v15i2.79

[32] DasGupta, R. (2022). Financial performance shortfall, ESG controversies, and ESG performance: Evidence from firms around the world. Finance Research Letters, 46, 102487. https://doi.org/10.1016/j.frl.2021.102487

[33] Saini, N., Antil, A., Gunasekaran, A., Malik, K., & Balakumar, S. (2022). Environment-social-governance disclosures nexus between financial performance: A sustainable value chain approach. Resources, Conservation and Recycling, 186, 106571. https://doi.org/10.1016/j.resconrec.2022.106571

[34] Comincioli, N., Poddi, L., & Vergalli, S. (2012). Corporate social responsibility and firms' performance: A stratigraphical analysis. Available at SSRN 2132202. https://doi.org/10.2139/ssrn.2175513

[35] Kuzey, C., & Uyar, A. (2017). Determinants of sustainability reporting and its impact on firm value: Evidence from the emerging market of Turkey. Journal of cleaner production, 143, 27-39. https://doi.org/10.1016/j.jclepro.2016.12.153

[36] Kim, J., & Kim, J. (2018). Corporate sustainability management and its market benefits. Sustainability, 10(5), 1455. https://doi.org/10.3390/su10051455

[37] Güngör, N. (2023). Sürdürülebilirlik Raporlarında Döngüsel Ekonomi: Borsa İstanbul'da Bir Araştırma. Denetim ve Güvence Hizmetleri Dergisi, 3(1), 36-47.

[38] Stewart, R., & Niero, M. (2018). Circular economy in corporate sustainability strategies: A review of corporate sustainability reports in the fast-moving consumer goods sector. Business Strategy and the Environment, 27(7), 1005-1022. https://doi.org/10.1002/bse.2048

[39] Nilashi, M., Rupani, P. F., Rupani, M. M., Kamyab, H., Shao, W., Ahmadi, H., ... & Aljojo, N. (2019). Measuring sustainability through ecological sustainability and human sustainability: A machine learning approach. Journal of Cleaner Production, 240, 118162. https://doi.org/10.1016/j.jclepro.2019.118162

[40] Gorenc Novak, M., & Velušček, D. (2016). Prediction of stock price movement based on daily high prices. Quantitative Finance, 16(5), 793-826. https://doi.org/10.1080/14697688.2015.1070960

[41] Barak, S., Arjmand, A., & Ortobelli, S. (2017). Fusion of multiple diverse predictors in stock market. Information Fusion, 36, 90-102. https://doi.org/10.1016/j.inffus.2016.11.006

[42] Hajek, P., & Henriques, R. (2017). Mining corporate annual reports for intelligent detection of financial statement fraud-A comparative study of machine learning methods. Knowledge-Based Systems, 128, 139-152. https://doi.org/10.1016/j.knosys.2017.05.001

[43] Chen, Y. C., & Huang, W. C. (2021). Constructing a stock-price forecast CNN model with gold and crude oil indicators. Applied Soft Computing, 112, 107760. https://doi.org/10.1016/j.asoc.2021.107760

[44] Kocaarslan, B., & Soytas, U. (2023). The role of major markets in predicting the U.S. municipal green bond market performance:

New evidence from machine learning models. Technological Forecasting and Social Change, 196, 122820. https://doi.org/10.1016/j.techfore.2023.122820

[45] Nayak, R. K., Mishra, D., & Rath, A. K. (2015). A Naïve SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. Applied Soft Computing, 35, 670-680. https://doi.org/10.1016/j.asoc.2015.06.040

[46] Chen, Y., & Hao, Y. (2017). A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction. Expert Systems with Applications, 80, 340-355. https://doi.org/10.1016/j.eswa.2017.02.044

[47] Ayala, J., García-Torres, M., Noguera, J. L. V., Gómez-Vela, F., & Divina, F. (2021). Technical analysis strategy optimization using a machine learning approach in stock market indices. Knowledge-Based Systems, 225, 107119.https://doi.org/10.1016/j.knosys.2021.107119

[48] Park, H. J., Kim, Y., & Kim, H. Y. (2022). Stock market forecasting using a multi-task approach integrating long short-term memory and the random forest framework. Applied Soft Computing, 114, 108106. https://doi.org/10.1016/j.asoc.2021.108106

[49] Kim, H. Y., & Won, C. H. (2018). Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Expert Systems with Applications, 103, 25-37. https://doi.org/10.1016/j.eswa.2018.03.002

[50] Ince, H., & Trafalis, T. B. (2017). A hybrid forecasting model for stock market prediction. Economic Computation & Economic Cybernetics Studies & Research, 51(3).

[51] Li, Z., & Tam, V. (2017, November). A comparative study of a recurrent neural network and support vector machine for predicting price movements of stocks of different volatilites. In 2017 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-8). IEEE. https://doi.org/10.1109/SSCI.2017.8285319

[52] Eachempati, P., Srivastava, P. R., Kumar, A., Tan, K. H., & Gupta, S. (2021). Validating the impact of accounting disclosures on stock market: A deep neural network approach. Technological Forecasting and Social Change, 170, 120903. https://doi.org/10.1016/j.techfore.2021.120903

[53] Çelik, T. B., İcan, Ö., & Bulut, E. (2023). Extending machine learning prediction capabilities by explainable AI in financial time series prediction. Applied Soft Computing, 132, 109876. https://doi.org/10.1016/j.asoc.2022.109876

[54] Orlitzky, M., Schmidt, F. L., & Rynes, S. L. (2003). Corporate social and financial performance: A meta-analysis. Organization studies, 24(3), 403-441. https://doi.org/10.1177/0170840603024003910

[55] Lu, W., Chau, K. W., Wang, H., & Pan, W. (2014). A decade's debate on the nexus between corporate social and corporate financial performance: a critical review of empirical studies 2002-2011. Journal of cleaner production, 79, 195-206. https://doi.org/10.1016/j.jclepro.2014.04.072

[56] Kang, C., Germann, F., & Grewal, R. (2016). Washing away your sins? Corporate social responsibility, corporate social irresponsibility, and firm performance. Journal of Marketing, 80(2), 59-79. https://doi.org/10.1509/jm.15.0324

[57] Buallay, A., El Khoury, R., & Hamdan, A. (2021). Sustainability reporting in smart cities: A multidimensional performance measures. Cities, 119, 103397. https://doi.org/10.1016/j.cities.2021.103397

[58] Safari, K., Njoka, C., & Munkwa, M. G. (2021). Financial literacy and personal retirement planning: a socioeconomic approach. Journal of Business and Socio-Economic Development, 1(2), 121-134. https://doi.org/10.1108/JBSED-04-2021-0052

[59] Vitezić, N., Vuko, T., & Mörec, B. (2012). Does financial performance have an impact on corporate sustainability and CSR disclosure? A case of Croatian companies. Journal of Business Management, 5(Special Edition).

[60] Dağıstanlı, H. A., & Çelik, İ. (2023). Sürdürülebilirlik Raporlaması ve Firma Performansı: BIST Sürdürülebilirlik Endeksi Üzerine Bir Uygulama. Dumlupınar Üniversitesi Sosyal Bilimler Dergisi, (76), 1-16. https://doi.org/10.51290/dpusbe.1153330

[61] Sariyer, G., & Taşkın, D. (2022). Clustering of Firms Based on Environmental, Social, and Governance Ratings: Evidence from BIST Sustainability Index. Borsa Istanbul Review. https://doi.org/10.1016/j.bir.2022.10.009

[62] Duda, R. O., & Hart, P. E. (2012). Pattern classification. John Wiley & Sons.

[63] Chollet, F. (2018). Deep learning mit python und keras: das praxis-handbuch vom entwickler der keras-bibliothek. MITP-Verlags GmbH & Co. KG.

[64] Zhang, J., Rong, W., Liang, Q., Sun, H., & Xiong, Z. (2017). Data augmentation based stock trend prediction using self-organising map. In Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14-18, 2017, Proceedings, Part II 24 (pp. 903-912). Springer International Publishing. https://doi.org/10.1007/978-3-319-70096-0_92

[65] Teng, X., Wang, T., Zhang, X., Lan, L., & Luo, Z. (2020). Enhancing stock price trend prediction via a time-sensitive data augmentation method. Complexity, 2020, 1-8. https://doi.org/10.1155/2020/6737951

[66] Lee, S. W., & Kim, H. Y. (2020). Stock market forecasting with super-high dimensional time-series data using ConvLSTM, trend sampling, and specialized data augmentation. expert systems with applications, 161, 113704. https://doi.org/10.1016/j.eswa.2020.113704

[67] Ertel, W. (2018). Introduction to artificial intelligence. Springer. https://doi.org/10.1007/978-3-319-58487-4

[68] Dinov, I. D. (2018). Data science and predictive analytics. Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-72347-1

[69] Koutroumbas, K., & Theodoridis, S. (2008). Pattern recognition. Academic Press.

[70] Kubat, M., & Kubat, J. A. (2017). An introduction to machine learning (Vol. 2, pp. 321-329). Cham, Switzerland: Springer

International Publishing. https://doi.org/10.1007/978-3-319-63913-0

[71] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32. https://doi.org/10.1023/A:1010933404324

[72] Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. Annals of statistics, 1189-1232. https://doi.org/10.1214/aos/1013203451