

Business Model Innovation for Sustainable Competitive Advantage: A Machine Learning Approach

Kingshuk Srivastava^{1,*}, Nalla Bala Kalyan², Wasid Khan³, Kanchan Natha Shinde⁴, Shakhlo Begmatova⁵, and Manish Kaushik⁶

¹*Department of Data Science and Big Data, School of Computer Science, IILM, Greater Noida, India*

²*Department of Management Studies, S.A. Engineering College, Chennai, India*

³*EFSIM Facilities Management Company, Riyadh, Saudi Arabia*

⁴*Commerce and Management, Savitribai Phule Pune University, Pune, India*

⁵*Department of Corporate Finance and Securities, Tashkent State University of Economics, Tashkent, Uzbekistan*

⁶*Apex School of Basic, Life and Applied Sciences, Apex University, Jaipur, India*

*Corresponding author: kingshuk.india@gmail.com

Abstract. In an increasingly competitive business environment, organizations must innovate continuously and adopt workable practices to succeed. Business model innovation has become an effective means of adding value, improving efficiency, and securing a favorable market position. This paper investigates the link between sustainability indicators and competitive advantage using machine learning methods. A dataset of publicly listed companies containing environmental, social, and governance (ESG) variables was analyzed to predict overall corporate sustainability grades. Three classification models were implemented: logistic regression, support vector machine (SVM), and k-nearest neighbors (KNN). Logistic regression and SVM achieved the highest prediction accuracy, 94.37%, outperforming KNN (88.73%). These results suggest that ESG indicators are reliable predictors of market positioning and sustainability performance. The study also indicates that sustainability-based, innovative business models can strengthen stakeholder trust and sustainable competitive advantage, and that machine learning can help investors and managers make informed strategic decisions.

Keywords. Business model innovation; sustainable competitive advantage; ESG; machine learning; logistic regression; support vector machine; corporate sustainability; predictive analytics

I. INTRODUCTION

In the present business environment, firms must meet rising demands on governance, social, and environmental issues while remaining competitive. Traditional profit-centered approaches are becoming insufficient because of a growing need for responsible and sustainable operations. A substantial number of companies are therefore adopting business models that integrate sustainability with profit [1]. Environmental, social, and governance (ESG) metrics have become common tools for evaluating efficiency and corporate sustainability. A positive ESG score can be associated with an enhanced reputation, greater stakeholder trust, and lower risk [2], [3], and these traits can result in sustainable competitive advantage.

At the same time, machine learning offers effective ways to analyze data on enterprise activities and to predict future outcomes [4]. This paper applies logistic regression, support vector machine (SVM), and k-nearest neighbors (KNN) methods to evaluate ESG grades and the influence of sustainability measures on competitive advantage [5].

II. LITERATURE REVIEW

Recent studies have shown that sustainability plays an increasingly important role in determining the future success of businesses. Research has pointed out that companies with a good track record of sustainable practices have built a favorable reputation and reduced their risks.

Business model innovation is one of the essential components of competitiveness [6]. Organizations that respond to changes in technology, customer needs, and regulatory conditions tend to succeed more often than those that keep to traditional methods. Sustainable operations and planned resource use are now regarded as vital elements of innovation [7].

ESG ratings are widely used to measure the governance, social, and environmental performance of companies and to assess their capability to deal with these issues. Strong ESG scores are associated with sound governance structures and with value creation [8]. Investors are therefore building ESG criteria into their core investment decisions.

Machine learning algorithms support data analysis within business firms. KNN, SVM, and logistic regression are fun-

damental methods for prediction and classification [9], [10]. Few studies, however, have combined machine learning with ESG performance to investigate sustainable competitive advantage. The present research addresses this gap by predicting ESG grades with machine learning models [11].

III. METHODOLOGY

This study trained machine learning classifiers on labeled data to predict the sustainability grades of organizations from ESG factors. The method consisted of five phases: data collection, preprocessing, variable selection, model building, and assessment [12].

A. Dataset Collection

The dataset used in this research consists of the ESG scores of 722 listed firms from various sectors. It contains both categorical and continuous variables, namely industry sector, environmental score, social score, governance score, and total ESG score. The dependent variable is the total grade, which reflects the overall sustainability rating of each firm [13].

B. Data Preprocessing

The following preprocessing steps were applied to the dataset:

- records with missing values were deleted;
- the industry variable was converted to numeric form by label encoding;
- the target variable (total grade) was encoded as numeric classes; and
- all features were scaled.

C. Feature Selection

The environmental score, social score, governance score, and industry category were used as the input variables for model building, because they represent sustainability performance and position [14]. The dependent variable was the overall ESG grade (total grade).

D. Model Development

Three machine learning classifiers were compared:

1. logistic regression, a basic linear classifier;
2. SVM, which classifies ESG grades using optimal decision boundaries [15], [16]; and
3. KNN, a distance-based classifier.

The dataset was split into a training set (80%) and a test set (20%), with the random state set to 42.

E. Performance Evaluation

The models were evaluated with accuracy, precision, recall, the F1-score, and the confusion matrix. These measures were used to assess classification effectiveness and to compare model performance [17]. Accuracy is defined in (1), where TP , TN , FP , and FN denote true positives, true negatives, false positives, and false negatives:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}. \quad (1)$$

F. Research Framework

The overall workflow followed six steps: dataset collection, preprocessing, feature engineering, model training, predic-

tion, and evaluation.

IV. RESULTS AND DISCUSSION

This section presents the results of the three classification models (logistic regression, SVM, and KNN) used to predict the total grade assigned to the listed firms.

A. Comparative Model Performance

The accuracy, precision, recall, and F1-score of the algorithms were used to assess their predictive performance, and the results are given in Table 1. Logistic regression and SVM achieved the highest classification accuracy, 94.37%, whereas KNN reached 88.73%. This implies that the linear and margin-based classifiers performed better than the distance-based algorithm on this ESG dataset.

TABLE 1. PERFORMANCE OF THE CLASSIFICATION MODELS

Model	Accuracy (%)	Precision	Recall	F1-score
Logistic regression	94.37	0.94	0.94	0.94
SVM	94.37	0.94	0.94	0.94
KNN	88.73	0.89	0.89	0.88

Fig. 1 compares the accuracy of the three models. Logistic regression and SVM clearly outperform KNN.

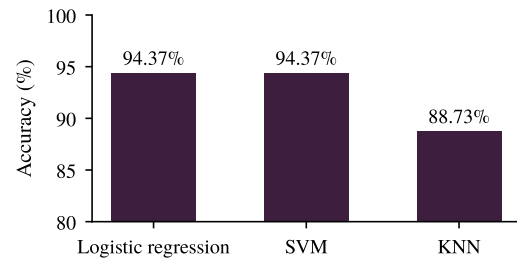


Fig. 1. Accuracy comparison of the machine learning models.

B. Classification Behavior Analysis

The models show good predictive accuracy, especially for the classes with more data. The classes with higher support have good precision and recall because their decision boundaries are stable.

SVM performs consistently because it is well suited to multiclass prediction of ESG grades. Logistic regression is also consistent, because the relationship between the indicators and the final grade is structured.

The confusion matrix of the SVM model in Fig. 2 shows the correct and incorrect classifications of the ESG grades on the test set. Most classes are classified accurately, and the few errors fall between adjacent classes.

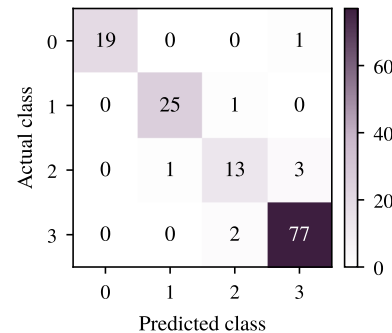


Fig. 2. Confusion matrix of the SVM model on the test set (classes 0 to 3 are the label-encoded ESG grades).

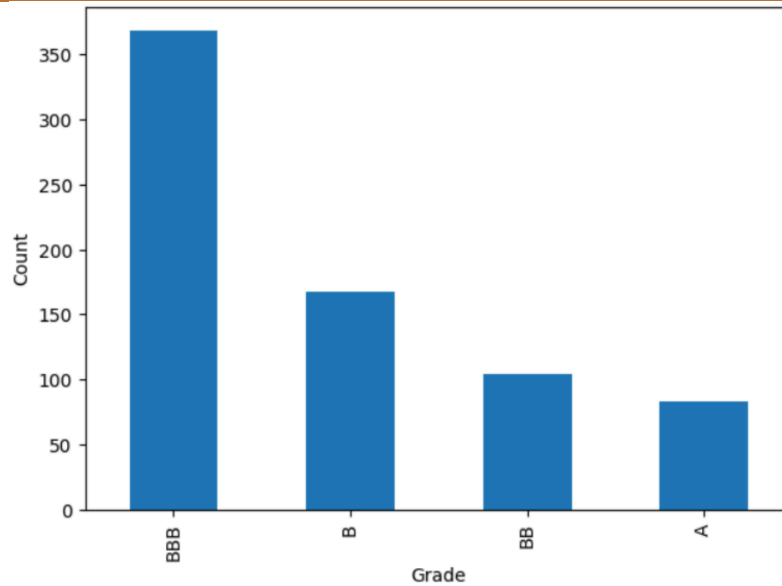


Fig. 3. Distribution of the ESG grades.

C. Target Class Distribution

The ESG grades are not uniformly distributed in the dataset, which may explain the relatively weaker performance of the classifiers on some minority classes. Fig. 3 shows the distribution of the ESG grades and the class imbalance among the sustainability categories. Class imbalance is common in real-world business datasets.

D. ESG Indicator Relationships

A correlation analysis was carried out to identify the relationships between the environmental, social, and governance scores and the overall ESG score. The correlation heatmap in Fig. 4 shows positive correlations among all the ESG categories, which means that companies that performed well in one sustainability area tended to perform well in the others. Each individual ESG score was also highly correlated with the overall ESG score.

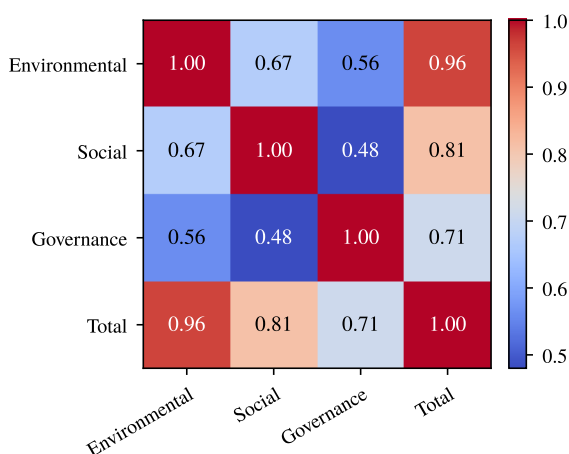


Fig. 4. Correlation heatmap of the ESG indicators.

E. Managerial Implications

The analysis shows that ESG metrics can successfully predict sustainability grades and competitive positioning. Firms with stronger environmental, social, and governance

performance are expected to attain higher sustainability grades.

F. Key Findings

Logistic regression and SVM performed better than KNN in predicting the ESG grades of listed firms. Both achieved a classification accuracy of 94.37%, whereas KNN reached a lower 88.73%. This difference suggests that the ESG dataset has a structure that is better captured by linear and margin-based methods than by distance-based models. The results also show that strategies that address sustainability issues can give firms a competitive advantage.

Although KNN gave satisfactory results, its lower accuracy shows that instance-based learners may struggle when classes overlap or are imbalanced. From a management perspective, this study shows that sustainability ratings can be predicted from ESG indicators. Companies that strengthen their governance processes, improve their environmental sustainability, and fulfill their social responsibilities obtain better ESG ratings.

V. CONCLUSION

This study explored machine learning techniques for predicting ESG grades. Logistic regression and SVM attained the highest accuracy, 94.37%, while KNN had the lowest accuracy, 88.73%. The findings show that ESG performance is important in predicting the overall sustainability grade, and they highlight environmental and social commitment and good governance as critical factors for future competitive advantage.

The analysis also shows that machine learning can support ESG analysis and decision-making. Future research should use more data, more advanced algorithms, and live sustainability measures, and should benchmark the results against earlier ESG prediction studies.

AUTHOR CONTRIBUTIONS

All authors contributed to the conception and preparation of this work and approved the final version of the manuscript.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

DECLARATION OF GENERATIVE AI USE

The authors declare that no generative artificial intelligence tools were used in the preparation of this manuscript. The authors take full responsibility for the content of the publication.

DATA AVAILABILITY

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

This study used publicly available, de-identified data and did not involve direct interaction with human participants or animals; ethical approval was therefore not required.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- [1] Michael E. Porter, *Competitive Advantage: Creating and Sustaining Superior Performance*. New York, NY, USA: Free Press, 1985, ISBN 0-02-925090-0.
- [2] Jay B. Barney, "Firm resources and sustained competitive advantage," *J. Manage.*, vol. 17, no. 1, pp. 99-120, 1991, doi: 10.1177/014920639101700108.
- [3] Christoph Zott and Raphael Amit, "Business model design and the performance of entrepreneurial firms," *Organ. Sci.*, vol. 18, no. 2, pp. 181-199, 2007, doi: 10.1287/orsc.1060.0232.
- [4] David J. Teece, "Business models, business strategy and innovation," *Long Range Plann.*, vol. 43, no. 2-3, pp. 172-194, 2010, doi: 10.1016/j.lrp.2009.07.003.
- [5] Nancy M. P. Bocken, Samuel W. Short, Padmakshi Rana, and Steve Evans, "A literature and practice review to develop sustainable business model archetypes," *J. Clean. Prod.*, vol. 65, pp. 42-56, 2014, doi: 10.1016/j.jclepro.2013.11.039.
- [6] Gunnar Friede, Timo Busch, and Alexander Bassen, "ESG and financial performance: Aggregated evidence from more than 2000 empirical studies," *J. Sustain. Finance Invest.*, vol. 5, no. 4, pp. 210-233, 2015, doi: 10.1080/20430795.2015.1118917.
- [7] Robert G. Eccles, Ioannis Ioannou, and George Serafeim, "The impact of corporate sustainability on organizational processes and performance," *Manage. Sci.*, vol. 60, no. 11, pp. 2835-2857, 2014, doi: 10.1287/mnsc.2014.1984.
- [8] Tensie Whelan, Ulrich Atz, Tracy Van Holt, and Casey Clark, "ESG and financial performance: Uncovering the relationship by aggregating evidence from 1,000 plus studies published between 2015-2020," NYU Stern Center for Sustainable Business and Rockefeller Asset Management, New York, NY, USA, 2021. [Online]. Available: https://www.stern.nyu.edu/sites/default/files/assets/documents/NYU-RAM_ESG-Paper_2021%20Rev_0.pdf
- [9] Trevor Hastie, Robert Tibshirani, and Jerome Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2009, doi: 10.1007/978-0-387-84858-7.
- [10] Christopher M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006, ISBN 978-0-387-31073-2.
- [11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825-2830, 2011. [Online]. Available: <https://jmlr.org/papers/v12/pedregosa11a.html>
- [12] Corinna Cortes and Vladimir Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273-297, 1995, doi: 10.1007/BF00994018.
- [13] D. R. Cox, "The regression analysis of binary sequences," *J. R. Stat. Soc. Ser. B (Methodol.)*, vol. 20, no. 2, pp. 215-232, 1958, doi: 10.1111/j.2517-6161.1958.tb00292.x.
- [14] Sebastian Raschka and Vahid Mirjalili, *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2*, 3rd ed. Birmingham, U.K.: Packt Publishing, 2019, ISBN 978-1-78995-575-0.
- [15] Stefan Schaltegger, Florian Lüdeke-Freund, and Erik G. Hansen, "Business cases for sustainability: The role of business model innovation for corporate sustainability," *Int. J. Innov. Sustain. Dev.*, vol. 6, no. 2, pp. 95-119, 2012, doi: 10.1504/IJISD.2012.046944.
- [16] Ram Nidumolu, C. K. Prahalad, and M. R. Rangaswami, "Why sustainability is now the key driver of innovation," *Harv. Bus. Rev.*, vol. 87, no. 9, pp. 56-64, 2009. [Online]. Available: <https://hbr.org/2009/09/why-sustainability-is-now-the-key-driver-of-innovation>
- [17] World Economic Forum, "Measuring stakeholder capitalism: Towards common metrics and consistent reporting of sustainable value creation," World Economic Forum, Geneva, Switzerland, Sep. 2020. [Online]. Available: <https://www.weforum.org/publications/measuring-stakeholder-capitalism-towards-common-metrics-and-consistent-reporting-of-sustainable-value-creation/>