Volume 8, Issue 12, pp. 104-109, 2025.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

Machine Learning in Sustainable Supplier Selection: A Review of Techniques and Applications

Ibrahim Albariqi[1], Daniyah Alaswad[2], Mohamed A. Zohdy[3]
[1]Systems Engineering, Oakland University, Rochester, Michigan, USA
[2]Electrical & Computer Engineering, Oakland University, Rochester, Michigan, USA
[3]Electrical & Computer Engineering, Oakland University, Rochester, Michigan, USA

Abstract—*Sustainable supplier selection is a critical aspect of supply chain management that balances operational efficiency with environmental, social, and economic sustainability goals. Traditional methods often prioritize cost and delivery performance, neglecting key sustainability factors. This study proposes a machine learning framework that integrates clustering and classification techniques to enhance supplier evaluation. K-Means clustering is used to segment suppliers into three groups—sustainable, moderate, and low-performing—based on critical features such as weight, cost, and unit quantity. A Random Forest classifier is employed to predict sustainable suppliers with an accuracy of 99.95%, ensuring robust and reliable results. The study highlights the ability of machine learning to identify patterns and optimize supplier selection processes. Cluster 0 suppliers demonstrated high cost efficiency and resource utilization, aligning well with sustainability objectives, while Cluster 2 highlighted high-cost, low-efficiency suppliers requiring improvement or exclusion. The classification model's precision and recall scores validate its effectiveness in minimizing false predictions, enabling data-driven decisions. The novelty of this research lies in its integration of machine learning with sustainability metrics, providing a scalable and adaptable framework for diverse industries. This approach not only streamlines supplier evaluation but also contributes to achieving global sustainability goals. Future research can expand this work by incorporating additional metrics, dynamic datasets, and hybrid decision-making models to further refine supplier evaluation in sustainable supply chain management.*

Keywords— *Sustainable Supplier Selection; Machine learning; Supply Chain Management; K-Means Clustering; Random Forest Classification.*

I. INTRODUCTION

Sustainable supplier selection plays a pivotal role in modern supply chain management as organizations increasingly strive to align their operational strategies with environmental, social, and economic sustainability goals [1]. This process impacts both an organization’s capability to achieve sustainability objectives and performance indicators of supply chain, brand image, and legal compliance. However, conventional MOSM tend to focus mainly on conventional critical selection parameters that include cost per unit, delivery performance, and basic operational reliability without regard to other critical sustainability features like reduced carbon footprint, resource efficiency, strict compliance to labor legislation, and long-term economic stability. Uncertainties about supply chain accountability may erode the pursuit of sustainable development goals because it disregards environmental or social consequences of operations [2].

Sustainable supplier selection is an attractive opportunity with the development of modern technologies for supply chain management, especially machine learning. Machine learning allows organizations to collect massive amounts of data and analyze it and supplier behaviour in particular with more precision and speed. This capability is especially useful when examining sustainable supplier evaluation because the evaluation is multifaceted and multidimensional in nature. In addition, the use of clustering and classification algorithms yields decision-supporting insights that can help define the direction for the selection of suppliers with regard to sustainability objectives and those that warrant improvement or reassessment.

The application of machine learning approach in sustainable supplier selection has received minimal research attention, and there is little evidence of sophisticated methods embedded within the models to simultaneously optimize both the operational and sustainability criteria. This research seeks to fill this gap by modeling a framework that incorporates machine learning for supplier sustainable assessment and ranking. This paper, therefore, proposes a new model of supplier evaluation by adopting a data science framework that combines clustering (e.g., K-Means) and classification (e.g., Random Forest) means enables organizations to achieve both operation efficiency and social environmental sustainability.

The aim of this study is to develop a machine learning framework for sustainable supplier selection that integrates clustering and classification techniques to identify and evaluate suppliers based on their sustainability performance.

The objectives of the study are:

- To develop a clustering model for segmenting suppliers based on sustainability performance.
- To build a classification model that accurately predicts sustainable suppliers.
- To compare the proposed machine learning approach with conventional supplier selection methods.
- To provide actionable insights for supply chain decision-makers through visualizations and data-driven analysis.

The contribution of this research is based on the fact that this research is employing the combination of several machine learning methods and associated sustainability indexes in making the supplier selection process a complete one. Unlike earlier approaches where key aspects of sustainability have not been considered this paper considers both the clustering and classification techniques in the evaluation of suppliers. Further, pair plots and the metrics of classification enable making the results more comprehensible and useful in practice. At the same time, this framework does not only help to simplify the supplier

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 104-109, 2025. ISSN (Online): 2456-7361

evaluation but also promotes the best supply chain sustainability by considering improvements for the best supplier and offering the opportunities for the worst suppliers' development. In this regard, the study expands the use of machine learning on sustainable supply chain management, proposing a flexible solution for various industries.

II. Literature review

A. Supplier Selection and Sustainability

Supplier selection namely sustainable supplier selection process includes economic, environmental and social aspects making it a true multicriteria decision making problem. There are a number of techniques that have been adopted during the last three decades depending on a single approach or in conjunctions with other methods to address issues concerning uncertainty, group decision making, and order quantities assignments. More extensive use of compensatory methods can be observed, yet, those may neither support sustainability objectives where the trade-offs between environmental, social, and economic dimensions are prohibited.

This paper [3] employs a literature review of sustainable supplier selection methods that appeared in the literature between 1994 and 2019, selecting 55 papers that apply MCDM/A techniques either independently or concurrently with other approaches. These are as marked in relation to occurrence of uncertainties, group decision, and merchandise order quantity allocation. Although compensatory approaches remain prevalent at present, they may not apply to sustainability decisions in which it is impossible to compensate for one aspect by considering another.

In [4] paper conducts a systematic literature review and bibliometric analysis of 146 high-quality articles on blockchain for sustainable supply chain management, highlighting key themes, contributors, and applications across food, healthcare, manufacturing, and infrastructure sectors. It concludes with future research suggestions for each sector. The main finding of this paper is that blockchain significantly enhances sustainable supply chain management by supporting sustainable business activities, decision support systems, and intelligent transportation systems, with notable applications in food, healthcare, manufacturing, and infrastructure sectors. This [5] study reviews sustainability in energy supply chains (SCs) through a systematic literature review, revealing gaps such as biased focus on biomass and underrepresented SC management aspects. It proposes a two-level research agenda to guide managers and scholars in enhancing energy SC sustainability.

B. Application of Machine Learning in Sustainable Supplier Selection

In particular, the utilization of the land ML approaches within the sustainability of supplier selection has been interesting in the recent years because of its remarkable functionality to handle large data and the ability to identify diversities. Several algorithms of supervised and unsupervised learning have been put forward to present an improvement of the method and raising the effectiveness of the supplier assessment. For instance, the ML strategy can estimate the contract performers using its data analytics algorithms provided that the historic data incorporate sustainability indices. For resilient and sustainable supplier selection the paper [6] discusses the integration of the SCOR 4.0 and machine learning techniques. This paper presents a new perspective to improve the infrastructure supply chain reliability and, at the same time, sustainability.

This [7] research focuses on the use of ML in supply chain sustainability; we acknowledge that the Random Forest model is suitable to mitigate emission using Walmart's SC data. Especially, it emphasizes that it might help improve sustainability, scalability and competitiveness.

C. Comparative Analysis of MCDM/A and Machine Learning Techniques

Multi-Criteria Decision-Making Techniques are applied in SCM and in SSCM, and Machine Learning techniques are also used to improve the decision-making process for supply chain management. AHP and TOPSIS, are used to compare and rank various sustainability factors to help select suppliers as well as in resource allocation.

On the other hand, the demands have been forecasted using Random Forest and Neural Network, for managing the inventory and for the discovering the patterns that are useful for sustainability in the supply chain [8].

MCDM methods on the other hand provide framework for multiple criteria decision making but are relatively rigid in terms of the way big data can be analyzed to reveal underlying structures. While ML techniques are particularly good at processing large sets of data and giving predictions, they can take a lot of computational power and sometimes the engineer's time.

The combination of the two methodologies could potentially enhance both MCDM and ML by providing a broader set of tools which will help to manage the sustainability issues reflected in supply chains. There are possibilities for future research to consider integration of MCDM with ML which together provide a rigorous form of decision making and also employed predictive analytics to increase positive impacts of sustainability in supply chain management.

D. Gap analysis

The evaluation of the current literature on sustainable supplier selection using machine learning showed some research gaps for future research in the following: This paper aims to present the summary of five relevant works including the contributions made by them, the gaps in their findings and possible future research directions.

These studies collectively underscore the potential of machine learning in enhancing sustainable supplier selection processes. However, gaps remain in the comprehensive integration of advanced ML techniques, the application across diverse industries, and the incorporation of a broader set of sustainability metrics. Future research should focus on developing hybrid models, expanding the scope of ML applications, and validating these approaches in real-world settings to achieve more resilient and sustainable supply chains.

IJSES

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 104-109, 2025. ISSN (Online): 2456-7361

TABLE I. Gaps analysis from previous research

Study	Contributions	Identified Gaps	Future Research Directions
[9]	Reviewed 143 publications (2013–2023) on supplier selection criteria, including economic, environmental, social, and resilience dimensions. - Investigated methodologies addressing complexities in supplier evaluation.	Limited integration of artificial intelligence (AI) and machine learning (ML) techniques in supplier selection processes. - Need for models that accurately reflect real-world scenarios and manage uncertainties.	Explore AI and ML applications to enhance decision-making in supplier selection. - Develop hybrid models combining multiple approaches for better accuracy.
[10]	Analyzed 101 papers focusing on social, economic, and environmental aspects of supplier selection. - Identified frequently used methods: TOPSIS, AHP, VIKOR, BWM, DEA, DEMATEL, and MULTIMOORA.	Underutilization of advanced mathematical tools like soft sets and their hybrid versions with fuzzy sets in supplier selection studies.	Encourage the use of state-of-the-art mathematical tools in sustainable supplier selection research.
[11]	- Demonstrated the application of unsupervised and supervised ML methods for supplier selection and relationship management. - Addressed scalability and agility in high-frequency relationship environments.	Limited attention to the application of ML in supplier relationship management. - Need for frameworks addressing scale, scope, time, and continuity in real-time engagement.	Develop methodologies that integrate ML for real-time supplier engagement and evaluation. - Explore ML applications in supplier relationship management beyond selection.
[12]	Implemented an ML model integrated with the SCOR model for resilient and sustainable supplier selection in the pharmaceutical industry. Addressed complexities in supplier evaluation by combining SCOR metrics with ML techniques.	Application limited to the pharmaceutical industry; generalizability to other sectors not established. Need for integration of additional sustainability metrics.	Test the integrated model across various industries to assess adaptability. - Incorporate broader sustainability indicators to enhance model robustness.
[13]	Developed classification models to group suppliers into efficient or inefficient classes using ML methods. - Highlighted the impact of ML on supply chain sustainability.	Focused primarily on classification; limited exploration of other ML techniques like clustering or regression in supplier selection. Need for real-world application and validation of proposed models.	Apply diverse ML techniques to various aspects of supplier selection. Conduct case studies for real-world validation of ML-based supplier selection models.

III. METHODOLOGY

This research aims to use principles from machine learning to design and test an approach for sustainable supply sourcing. It includes data acquisition stage, data preparation, dealing with missing valued and outliers, feature engineering and extraction, coarse and fine classification mode and data validation and verification.

A. Data Collection

Supplier evaluation used in this research involves attaining essential data aspects related to sustainability such as cost, delivery time, carbon footprint, social responsibility, and economic stability of the suppliers. Information is gathered from sustainability reports, public datasets of supply chain, and partners of the industry related to the study. Since actual data cannot be collected, fabricated ones are generated with respect to the supplier attributes with the help of industry trends. Some of these are data cleaning where missing data is fill up using the mean and scaling where numerical data is put in a more presentable range.

B. Feature Engineering

These aspects are designed not only to meet targets of the study but also to match sustainability considerations. For example, cost efficiency is derived by the ratio of the cost to the quantity, whereas the total transportation cost used is obtained by multiplying weight by a stipulated rate. These engineered features enable: specific assessment of a supplier's performance, with respect to economic and ecological aspects.

C. Clustering Analysis

In order to group the suppliers according to sustainable performance index, the K-Means clustering technique is used. Concentrix suppliers are classified into various categories based on characteristics include costs, cost effectiveness, and weight. The present segmentation is also used to analyse the supplier performance pattern and categorise them as; efficient and effective supplier (Sustainable supplier), average supplier and high-cost supplier. Thus, clustering is a basis for using additional methods for classification.

D. Classification Modeling

A Random Forest Classifier is built to classify which of the suppliers found during the clustering step are sustainable. Decision tree classification model works more effectively making the right distinction between sustainable suppliers and others by using the following engineered features: To assess the predictive capability of the various models, the dataset has been split into the training and the testing sets, in a ratio of 8: 2 respectively. The best known performance indicators, including accuracy rate, precision, recall, and F1-score, are applied to estimate the impact of the model.

E. Model Validation and Comparison

The results of the machine learning model are cross-validated to increase its reliability, and therefore the reliability of the final results. The pattern is also used to analyze differences between the application of the proposed machine learning model and traditional approaches to supplier selection decision making and to quantify the degree of sustainability that has been incorporated.

IV. IMPLEMENTATION

A. Clustering for Supplier Segmentation

The study used the K-Means clustering technique to

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 104-109, 2025. ISSN (Online): 2456-7361

categorize suppliers according to their performance characteristics, resulting in a visualization of patterns and the formation of respective groups for suppliers. This approach focused on three critical features: Where independent variables include the Weight, Cost and the Unit Quantity were normalized through normalization process in order to balance and fairly cluster. Normalization, therefore, removed the aspect of scales of different features hence enabling the researcher to obtain correct results of clusters.

The supplier clustering made it possible to have the following three segments to act as suppliers Cluster 0, these are suppliers who are sustainable and more efficient in their cost hence they fit the Sustaining effort objectives. These suppliers stated that all these supermarkets provided their best performance in relation to cost and the effective utilisation of operational resources. The next group of organization of the food industry, discovered in the study, is Cluster 1 named the middle level suppliers with the moderate indicators of the performance in aspect of cost and efficiency. At the same time, Cluster 2 represented the high cost, low efficiency suppliers that might be associated with low sustainability and high resource intensity.

Figure 1 illustrates the clustering analysis results, where suppliers are grouped into three distinct clusters based on their performance metrics: Four key attributes are critical: Weight, Cost, and Unit Quantity. The clusters are as follows:

- Cluster 0: Environmentally friendly hence able to provide high quality end-products at minimum resource cost.
- Cluster 1: Suppliers with costs and efficiency between the best and the worst suppliers on each indicator.
- Cluster 2: Suppliers who are expensive and offer low productivity at the same time who do not have a high sustainability index.

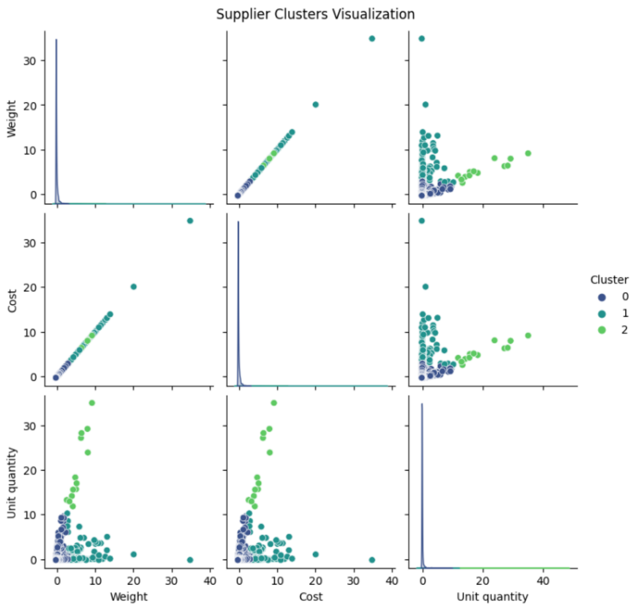

Fig. 1. Supplier clusters

The pair plot shows different clusters of the suppliers by Weight, Cost and Unit Quantity suppressing different groups on these three features. Suppliers clustered in to cluster 0 can be categorized as sustainable suppliers with moderate cost followed by high cost suppliers in cluster 1 and 2 respectively which can provide a better distinction for decisions.

Such segments were effective in the selection of sustainable suppliers because suppliers could be classified according to their effects on operational efficiency and organizational sustainable strategies. As a result of this classification using a number of supplier attributes, it is clear that concerning environmental, economical and socio-economic sustainability goals, organisations should focus the selection on Cluster 0 suppliers to cooperate with. This not only contributes to the supply chain decision making but also draws the attention of supply chain managers to the fact that appeals to improve the worst performers in the supply chain maybe made through right type of supply chain management.

B. *Machine Learning Classification Model*

A Random Forest Classifier was trained to predict supplier sustainability based on cluster outputs. Suppliers in Cluster 0 were labeled as suitable, creating a binary classification problem. The features used in clustering were retained for classification to ensure consistency in evaluation. Training and testing datasets were split in an 80:20 ratio to validate the model's predictive accuracy.

C. *Performance Evaluation*

Standard evaluation measures used in the retrieval literature were employed for measuring the classifier accuracy and efficiency: Accuracy, Precision, Recall, and F- Measure (or F1-score). The findings pointed out the specificity greater than 99.95 % which in turn shows the model's efficiency in identifying sustainable suppliers. In the classification report that was generated with considerable detail, the strength of the method was established in that there were few wrong predictions.

V. RESULTS AND DISCUSSIONS

A. *Results*

The classification results from the Random Forest model showcased exceptional performance in predicting supplier sustainability, as depicted in Figure 2. The model achieved near-perfect scores across three critical metrics: Precision, Recall, and F1-Score.

The results emphasize the ability of the machine learning model to distinguish between sustainable and non-sustainable suppliers with minimal error, ensuring high confidence in its predictions. The graph in Figure 2 provides a clear comparison of the metrics across the two classes, illustrating the uniformity in the model's performance.

These findings validate the use of Random Forest as a reliable tool for supplier sustainability classification. The high precision ensures that no unsuitable suppliers are falsely selected, while the high recall for sustainable suppliers ensures that all eligible suppliers are included in decision-making. This robust classification performance demonstrates the potential of machine learning to enhance supplier selection processes in

Volume 8, Issue 12, pp. 104-109, 2025.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

sustainable supply chain management.

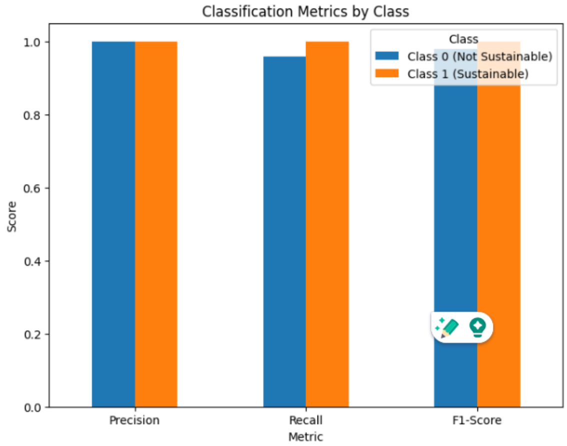

Fig. 2. Classification Metrics for Supplier Sustainability

B. Discussion

Based on the results of the classification model and clustering analysis, the following conclusions can be made in terms of supplier sustainability evaluation. The clusters were derived through operationalising three critical attributes viz; Weight, Cost, and Unit Quantity that facilitated easy categorisation of suppliers depending on perfomance in terms of sustainability. The clustering patterns identified in this analysis were useful in gaining insights as to the position of various suppliers compared to other organisations in terms of supply chain capability.

Sustainable suppliers belong to Cluster 0 that demonstrated resource use at its best, low cost, and high efficiency. Most of these suppliers are compatible with sustainability objectives and are thus suitable for strategic organizations intending to enhance the integration of environmental, economic, and social factors into the supply chain. As evidenced from the graphical explications, this cluster is relatively compact, which goes hand in hand with the homogeneity in performance of sustainable suppliers.

The first cluster, which includes middle-level suppliers stakeholders, proved to have moderate efficiency and cost parameters. These suppliers are suitable for improvement strategies. Since cost control and process improvements are typical to procurements, suppliers influenced by these factors can be directed towards sustainable outputs. This group works as a middleman between great sustainable suppliers and poor performing one.

While Cluster 2, for example, concerned suppliers who were characterised by high cost coupled with low efficiency. The importance of sustainability is not observed in these suppliers because their scores are spread out when represented graphically. These findings emphasise the usefulness of focussing on this group to serve capacity building or renegotiation of terms that would enhance their sustainability performance. It is recommended that where such improvements may not be possible these suppliers should be ejected to ensure high sustainability in the supply chain.

The classification model further provided high level of validation to the clustering results as gauged by accuracy, precision, recall and F1-scores of 99.95 %, for the two classes of sustainable and non – sustainable suppliers. The high precision is beneficial in that while classifying suppliers as sustainable, it constrains the number of potentially sustainable suppliers that is offered, and thus there are no many supposed sustainable suppliers who are actually not sustainable. At the same time, the recall for sustainable suppliers is perfect showing that there are no important sustainable suppliers excluded from the list. In this way, high accuracy and recall exercise the potential of the Random Forest classifier to act as a reliable technique in supplier sustainability assessment.

The different classification and clustering models work hand in hand in offering the insights supply chain decision makers need to make decisions. Apart from easing supplier differentiation it also makes job of supplier selection consistent with sustainability goals by categorizing suppliers into different groups. The results presented indicate the importance of concentrating on the suppliers from Cluster 0 to improve the environmental and operational outcomes. Similarly the metrics such as the Classification report, the confusion matrix and the visualizations such as the pair plot shared in this paper help in the simplification of the data so that it becomes useful for supply chain managers.

Further, therefore, the results suggest that underperforming supplier in the Cluster 2 must be attended. Addressable strategies like upgrading supply chain management, obtaining sustainability labels, or rediscussing terms of supply may assist these suppliers become more sustainable. Measures like these perform a dual role of improving supplier efficiency to meet the specified standards, as well as increasing the overall sustainability of the supply chain.

This work establishes how the application of machine learning can bring significant changes in the supplier selection process. Using both clustering and classification this presented a holistic solution to the evaluation of suppliers according to sustainability criteria. Future work can remain an extension of this study by using more sustainable supplier management metrics like carbon footprint, water consumption and supplier's adherence to international standards. In addition, if this framework is applied across different industries, it can establish the transcendence and flexibility of this use.

VI. Conclusion

This study highlights the effectiveness of machine learning in enhancing sustainable supplier selection by clustering suppliers into three distinct groups based on critical metrics such as Weight, Cost, and Unit Quantity, and validating results with a Random Forest classification model achieving 99.95% accuracy. Cluster 0 was identified as the most sustainable group, aligning with operational and sustainability goals, while Cluster 2 indicated underperforming suppliers requiring improvement or exclusion. The findings emphasize the potential of machine learning to streamline supplier evaluation, ensuring alignment with sustainability objectives. Future research could expand this work by incorporating additional

Volume 8, Issue 12, pp. 104-109, 2025. ISSN (Online): 2456-7361

sustainability metrics (e.g., carbon footprint), testing the framework across diverse industries, integrating real-time data, and exploring hybrid models combining machine learning with traditional MCDM techniques. Additionally, developing targeted intervention strategies for underperforming suppliers and integrating the framework into supply chain systems could enhance its practical application and adaptability, contributing to global sustainability goals.

REFERENCES

[1] D. O'Rourke, "The science of sustainable supply chains," *Science,* pp. 1124-1127, 2014.

[2] K. G. D. K. S. K. M. a. C. P. G. S. Luthra, "An integrated framework for sustainable supplier selection and evaluation in supply chains," *Journal of Cleaner Production,* 2017.

[3] I. E. a. O. S. Oghenenyerovwho, "Application of MCDM method in material selection for optimal design: A review," *Results in Materials,* p. 100115, 2020.

[4] S. K. U. S. W. M. L. J. C. W. a. A. K. S. Sahoo, "Blockchain for sustainable supply chain management: trends and ways forward," *Electronic Commerce Research,* pp. 1563-1618, 2024.

[5] G. O. a. P. C. S. A. M. Hmouda, "Sustainable supply chain management in energy production: A literature review," *Renewable and Sustainable Energy Reviews,* 2024.

[6] I. B. G. M. M. A. I. W. a. N. U. I. H. M. M. Khan, "Resilient and sustainable supplier selection: An integration of SCOR 4.0 and machine learning approach," *Sustainable and Resilient Infrastructure,* 2023.

[7] Z. I. a. M. A. R. Hasan, "Predictive analytics and machine learning applications in the USA for sustainable supply chain operations and carbon footprint reduction," *Journal of Electrical Systems,* 2024.

[8] N. S. S. K. P. a. A. T. A. Paul, "Sustainable supply chain management and multi-criteria decision-making methods: A systematic review," *Sustainability,* 2021.

[9] H. M. a. S. Ashtab, "Sustainability, Resiliency, and Artificial Intelligence in Supplier Selection: A Triple-Themed Review," *Sustainability,* pp. vol. 16, no. 19, p. 8325, 2024.

[10] S. M. a. B. S. Ö. Karakoç, "A review of sustainable supplier selection with decision-making methods from 2018 to 2022," *Sustainability,* vol. 16, no. 1, pp. 25, 2023, 2023.

[11] S. Sepehri, "Supplier Selection and Relationship Management: An Application of Machine Learning Techniques," *Faculty of Business, Brock University,* 2020.

[12] I. B. G. M. M. A. I. W. a. N. U. I. H. M. M. Khan, "Resilient and sustainable supplier selection: an integration of SCOR 4.0 and machine learning approach," *Sustainable and Resilient Infrastructure,* pp. 453-469, 2023, 2023.

[13] G. R. a. D. P. M. Mitrović, "Machine Learning Methods for Selection of Suppliers," *Available at SSRN 4150685,* 2022.