Vaasan yliopisto
UNIVERSITY OF VAASA

Vijini Witharana

Development of an Integrated Framework using Machine Learning and TOPSIS for Sustainable Supplier Selection in Circular Supply Chain Management

Master’s Thesis

School of Technology and Innovations
Master’s thesis in Industrial Systems Analytics
Master's Programme in Industrial Systems Analytics (ISA)

Vaasa 2025

UNIVERSITY OF VAASA			
School of Technology and Innovations			
Author:	Vijini Witharana		
Title of the thesis:	Development of an Integrated Framework by using Machine Learning and TOPSIS methods for Sustainable Supplier Selection in Circular Supply Chain Management		
Degree:	Master of Science in Technology		
Discipline:	Industrial Systems Analytics (ISA)		
Supervisor:	Ahm Shamsuzzoha		
Year:	2025	**Pages:**	113

ABSTRACT:

Sustainable supplier selection is an imperative part in the process where traditional supply chains are transforming into circular supply chains (CSCs) with emphasis on sustainability, resource efficiency, and minimal waste. The supplier selection methods which are traditional are often based on expert opinions and static evaluation frameworks which might be inappropriate to handle dynamic and intricate supply chain environments such as CSCs. In this research, this challenge is addressed by integrating Machine Learning (ML) with the TOPSIS to select sustainable suppliers in the circular supply chain management (CSCM).

The ML technique, Random Forest, is utilized in the proposed method to perform feature importance analysis. The various selection criteria are given weights based on the feature importance to minimize bias and enhance accuracy. Suppliers are then evaluated and ranked based on the weights given. The evaluation is based on adherence to the circular economy principles and criteria such as cost, delivery performance and quality which are traditional.

To validate the proposed integrated framework, a case study on a food and beverage manufacturing industry was conducted. Data was collected, including primary data obtained through an online structured questionnaire and secondary data from the company's historical supplier records. The primary data questionnaires were completed by 22 industry experts of the case company. The secondary data was followed by preprocessing, where data was merged and cleaned, the data was then input into the ML Model to derive values of feature importance. These values were next applied in the ML-TOPSIS model to rank suppliers.

An analysis between the ML-integrated TOPSIS and traditional TOPSIS methods was conducted to identify the effectiveness of the proposed framework. Findings indicate the objective, adaptive, data-driven decision-making nature of the ML-integrated TOPSIS model. This model dynamically adjusts supplier rankings based on historical data, improving accuracy and efficiency, unlike in traditional TOPSIS, which is heavily based on static, expert-defined weights and longer processing times.

This study addresses the literature gap between the ML techniques and Multi-criteria Decision Making (MCDM) methods in selecting sustainable suppliers for circular supply chains. The study provides a systematic tool for industry participants to improve the supplier selection process and promote global sustainable supply chain management. However, the study is limited by the size and scope of the dataset, which may affect generalizability across industries and regions. Future research can address these limitations by expanding datasets, incorporating diverse industry contexts, and exploring advanced ML techniques to improve decision accuracy and applicability.

KEYWORDS: Supplier Selection, Machine Learning, TOPSIS, Circular Supply Chain Management, Multi-Criteria Decision Making, Sustainability

Contents

Figures

Tables

Abbreviations

AHP – Analytic Hierarchy Process

AI – Artificial Intelligence

ANP – Analytic Network Process

CE – Circular Economy

CSC – Circular Supply Chain

CSCM – Circular Supply Chain Management

DEA – Data Envelopment Analysis

DM – Decision-Making

DT – Decision Tree

ESG – Environmental, Social, and Governance

GA – Genetic Algorithm

GHG – Greenhouse Gas

KNN – K-Nearest Neighbors

ML – Machine Learning

MCDM – Multi-Criteria Decision Making

NN – Neural Network

RF – Random Forest

RO – Research Objective

RQ – Research Question

SC – Supply Chain

SCM – Supply Chain Management

SS – Supplier Selection

SVM – Support Vector Machine

TOPSIS – Technique for Order of Preference by Similarity to Ideal Solution

1 Introduction

1.1 Background of the study

The world is undergoing transformation towards sustainability where the various industries strive to become sustainable and environmentally responsible while growing their business. As a result, supply chains are playing a key role in business operations to minimize the negative global resource consumption and environmental impact. However, for a successful sustainable development not only supply chain is required but also an integrated approach requires which includes advancement in technology, regulatory policies and corporate strategies beyond supply chain alone. The circular economy concept operates as a powerful framework that aids industries in becoming more responsible for sustainability and environmental aspects by minimizing waste, maximizing resource efficiency, and fostering long-term sustainability (Ghisellini et al., 2016). Thus, circular supply chains have become a blooming and important focal point for both researchers and industry practitioners to consider circular economy principles adaption in supply chain operations (Govindan & Hasanagic, 2018).

Supplier selection (SS) plays a key role in SC processes which contributes to adapting circular economy principles in the supply chain processes as choosing suppliers that prioritize the sustainability practices enables a transition towards circular supply chains where materials are reused, remanufactured or recycled. The sustainability, and efficiency of the whole SC are relying heavily on the supplier election process (Rezaei et al., 2016). In most of the cases, cost, quality, and delivery reliability are often prioritized as the selection criteria in the traditional SS process (Ho et al., 2010). However, since industries are now shifting towards circular supply chains, the existing traditional supplier selection approaches needs a refinement. This refinement should incorporate additional supplier selection criteria in the supplier selection processes, such as environmental performance, industry 4.0, the ability to recover and reuse materials, and product life cycle management (Jabbour et al., 2019). This shift necessitates innovative supplier selection and decision-making frameworks that can address the CE principles in modern complex SC processes (Govindan et al., 2020).

In the process of SS, assessing the suppliers based on several muti-dimensional criteria introduces a complex nature into the process of selecting suppliers. To address this issue, Multi-criteria decision-making (MCDM) techniques are being applied in the process of selecting supplier and aids the decision-making process (Velasquez & Hester, 2013). There are several MCDM techniques that are available in practice including the Best-Worst Method (BWM), Analytic Hierarchy Process (AHP), and Technique for Order Preference by Similarity to the Ideal Solution (TOPSIS) (Govindan & Sivakumar, 2016). Yet, the most prominent among these is the TOPSIS method because of its simplicity in computation, ease, and ranking of other selections based on their distance to a solution which is ideal (Hwang & Yoon, 1981). Despite the popularity of the TOPSIS framework, there are some limitations of this framework when dealing with high-dimensional, dynamic, uncertain, and large data environments (Behzadian et al., 2012). In circular supply chains, when selecting feasible suppliers, these supplier selection frameworks often must deal with high-dimensional and dynamic data environments. Therefore, to overcome this challenge, integrating emerging technologies like Machine Learning (ML) with TOPSIS enhances the process of selecting suppliers and its decision-making (DM) process to surmount these limitations and enable the decision-making process to be more effective (Li et al., 2021).

Machine Learning is a powerful technology with tremendous potential in handling large and intricate sets of data, identifying patterns, and draw predictive conclusions (Choi et al., 2018). Machine learning functionalities are extremely useful in decision-making process of supply chains. During the utilization of TOPSIS method, machine learning may be employed to assist in criteria weighting, data normalization, and sensitivity analysis which are the most important components of the TOPSIS method (Kamble et al., 2020). Further, ML-based supplier selection models possess the ability to adjust to dynamic supplier performance changes along with changing environmental regulations that allow the supplier selection model to be dynamic and pertinent over time (Papadopoulos et al., 2017). Thus, the decision-makers can achieve more accurate, effective, and sustainable outcomes through the integration of ML with adaptive and predictive capabilities, and the structured approach of TOPSIS (Govindan et al., 2020). Most importantly, the integration of ML into MCDM models is particularly significant in circular supply chains since the decision-making process of these supply chains is often complex in nature (Kusi-Sarpong et al., 2019). Selecting the suppliers for circular supply chains requires a comprehensive assessment of suppliers' capabilities to evaluate the recycling, remanufacturing, and material recovery aspects (Ghisellini et al., 2016). However, the traditional supplier selection approaches often fail to address multidimensional requirements (Govindan & Hasanagic, 2018). Thus, the ML-enhanced MCDM frameworks can address these limitations and fill this gap by analyzing diverse datasets including supplier innovation capacity, environmental impact metrics, and compliance with circular economy principles (Jabbour et al., 2019). Thus, the incorporation of ML within the MCDM models improves the DM process and also streamlines the activities of supply chains aiding the adaption of CE principles.

Additionally, the external pressures due to regulation necessities, competitiveness, and customers' demand for sustainable products propel the industries to adopt CE principles in their supply chains. However, these pressures compel organizations to use more sophisticated techniques and tools to evaluate suppliers effectively in the context of CSCs (Kamble et al., 2020). Hence, the integrated ML-TOPSIS framework addresses this need by providing a sophisticated, systematic, scalable, and data-driven approach to select suppliers, that is tailored to adapt circular economy principles in the process of selecting suppliers (Govindan et al., 2020).

This research intends to analyse the incorporation of ML into the TOPSIS model for circular supply chain supplier selection with a view to theoretical as well as practical contributions in the area of supply chain management. The research is an extension of the current literature on MCDM, ML, and circular supply chains while addressing the said gaps concerning the utilization of such technologies in the case of supplier selection in real-case applications. Hence, the study aims to provide a robust, scalable, data-driven supplier selection framework for the decision-makers to enhance sustainability and efficiency in circular supply chains by aligning theoretical insight with practical applications contributing to the transformation of supply chains which are focusing on circular economy and sustainability.

1.2 Statement of the Problem

The transformation towards circular supply chains (CSCs) is in full swing and gained pace in recent years with increased environmental concerns, compliance requirements, and lack of availability of resources (Mangla et al., 2018). A key difference between traditional and circular supply chains are in terms of its purpose, with CSCs striving towards closing down loops of materials through reuse, remanufacturing, and recycling to contribute towards sustainability (Jakhar et al., 2019). In CSCs, the process of selecting suppliers, however, is a multi-dimensional issue with complex factors and complex information (Schroeder et al., 2018). The process of selecting suppliers in CSCs, not only considers conventional criteria but also considers the environmental as well as social dimensions (Govindan et al., 2019).

In the DM process of selecting suppliers, many traditional MCDM methods such as Best Worst Method, TOPSIS and AHP are used (Senthil et al., 2014). Among them, TOPSIS is popular among decision-makers due to its simple nature (Azadeh et al., 2016). Nonetheless, the traditional TOPSIS model is beset with processing large, complex, and multi-dimensional data and processing decision factors' dynamic and non-linear behavior in CSCs (Ahmadi & Amin, 2019). Yet, when TOPSIS is compared with other methods such as AHP and Best Worst Method, it is less biased as it relies on objective computation whereas both AHP and BWM involves subjective comparison and subjective identification respectively (Govindan et al., 2019).

Through technological advancement, ML techniques have attained a high capacity for handling big data, recognizing patterns, and forecasting them (Wei & Zhou, 2023). Therefore, when ML techniques are incorporated into MCDM tools such, models enhance the decision function by handling historical data, eliminating subjectivity, and improving accuracy (Ghassemi et al., 2018).

Nonetheless, when considering academic studies, integration between ML and TOPSIS for CSCs' sustainable suppliers is not yet a commonly researched topic (Kusi-Sarpong et al., 2023). Therefore, a vacuum arises for researching a significant issue: How can Machine Learning (ML) and TOPSIS effectively integrate with circular supply chain practice for efficient sustainable supplier selection and decision-making? Therefore, such an issue can be addressed through an emerging framework in an effort to address CSCs' specific needs such as multi-dimensional and dynamic processing, complex assessment, and compatibility with sustainability goals.

1.3 Research Objectives

The study has three main research objectives such as:

RO1: Using Machine Learning Techniques to optimize supplier evaluation and enhancement of decision-making in CSCs.

RO2: Investing the impact on accuracy, and efficiency of sustainable ranking of suppliers with the impact of the ML-Integrated TOPSIS Framework.

RO3: Developing a ML-integrated TOPSIS Framework with the intention of aligning supplier selection with circular economy principles.

By achieving the above three objectives, this study provides a comprehensive framework to select suppliers sustainably and reliably.

1.4 Research Questions

This research work aims to integrate ML and TOPSIS in circular supply chains for effective and efficient selection decisions of suppliers in a sustainable direction. Thus, the problem under investigation is tackled through the three following research questions.

RQ1: How can improvement in evaluation and decision in circular supply chains for a supplier selection be achieved through Machine Learning techniques?

RQ2: How can the accuracy and efficiency in a sustainable supplier selection be affected by an ML-Integrated TOPSIS Model?

RQ3: How can ML-Integrated TOPSIS Model combine circular economy values with selection?

Therefore, addressing these research questions offers an insightful understanding of practical applications of ML and TOPSIS approaches in CSCs from both theoretical and industry practice viewpoints.

1.5 Significance of the Study

The contribution of this research is found in the fact that it adds on the current knowledge of sustainable supply chain management in today's world with the major emphasis on circular economy supply chains. That is why it occurs that SS is a key challange professionals face in the context of CE paradigm shift. There appears to be a lack of specific assessment criteria to apply when identifying and selecting suppliers in conventional, Mass, and hybrid supply chains; this often means that the selection process depends on the expert's feelings and intuitive decisions. It brings novelty concept through application of ML with TOPSIS technique to make a rational decision making approach for supplier selection.

As a result of incorporation of ML and TOPSIS metrics for the assessment of suppliers, it becomes more accurate and efficient for suppliers' evaluation with regards to the sustainability factors which includes the environmental, economical and social factors. This approach eliminates drawbacks inherent in the expert-based models and thus presents a more flexible and efficient way of applying them to SS in CSCs. Furthermore, the study' s results will be helpful especially to industries with an interest in moving to a more sustainable way of operation; as the integrated ML-TOPSIS model can be used across different fields and may help to improve the SC decision making.

Moreover, this research adds to the existing body of knowledge as it addresses a significant research question concerning the use of MCDM techniques, sustainability, and ML in SS. It is beneficial to academia by adding to the stock of knowledge and utility by offering the specifics about the tools and techniques that can facilitate SCMs. From this research, practitioners and industry players will be able to adopt a further comprehensive model that will enhance the application of enhanced supplier choices to help in the optimization of resource usage and the general aim of CE.

1.6 Structure of the Thesis

This paper is divided into several chapters, where the stages of the research process, from the formulation of the topic to the choice of methods and the usage of case study, and the final conclusion, are presented. Here is a general idea of the structure of the text:

In this chapter, the reader will find information regarding background of the research, purpose and objectives of the study as well as its importance. It also presents the nature of the problem under study, the questions that the study seeks to answer and the area of study. They all lead to a postscript where the tentative structure of the thesis is described.

This chapter discusses existing literature on SS in SSC and CSC. It also reviews different MCDM methods, sustainability indicators, and the incorporation of ML in SCM. In this chapter, the author also points out the research gaps in the existing literature that the study intends to fill.

The objectives of the research study followed by the research setting and main variables used in the research context are also outlined in this chapter. The paper describes the integrated ML-TOPSIS model that has been adopted to make the supplier selection, with a justified selection on the method of data collection, selecting the case study, and the criteria adopted in the evaluation of the vendors. This analysis also presents the issues of data preprocessing and feature importance used in the current study.

In this chapter, the authors describe the details of the case study which focuses on food and beverage manufacturing sector. Describes how the research data would be obtained through online questionnaires and secondary records in addition to data cleaning. The outcome of the developed methodology, namely the ML-TOPSIS is presented and compared with the conventional TOPSIS model.

This chapter explicates the research finding that were as a result of using the ML-TOPSIS model in the study. To assess the effectiveness and accuracy of the proposed model, it also compares this model with the conventional model of selecting the suppliers for the organization and their outcomes and its finding for the SSCM.

The last chapter of the study is the conclusion of the findings made and implications derived from the study highlighted in this chapter. It also provides suggestions on the research directions and real-life application involving supplier selection for circular supply chains. The final section illustrates the limitations of the study and the possibilities of the development of the integrated framework.

This is done chronologically where each chapter progresses from the other to present a coherent understanding of the research process and the role of the given study in the development of SSCM.

2 Literature Review

2.1 Multi-Criteria Decision Making

2.1.1 Introduction to Multi-Criteria Decision Making

The researchers and industry practitioners are using different techniques to rank, prioritize, and select between several preferences which are known as MCDM approaches. These techniques provide the decision-makers with the ability to select or rank the alternatives in the most suitable way (Abdulla et al., 2023). When making decisions, MCDM techniques often end up considering criteria which are both quantitative and qualitative. However, this makes the selection process of the alternatives more complex and complicated to reach an agreed decision. In literature, several proposed MCDM techniques can be found. Even though each of these approaches has the same objective of enabling the decision-making capabilities, each of the MCDM techniques brings its unique privileges (Badi & Kridish, 2020).

2.1.2 Studies with MCDM Techniques

There are many previous attempts to integrate a number of technologies and techniques with several MCDM methods practically in various sectors of industry. Over the years, integrating MCDM with other strategies like Fuzzy Logic and Machine learning and algorithms has been of much interest in the supplier evaluation and selection field. Prior studies have mainly dealt with integrating various MCDM techniques to build the models that act like frameworks in the selection of the supplier. For instance, Orji and Wei (2014) proposed an approach TOPSIS and FEDEMATEL for selecting sustainable supplier for a Chinese gear manufacturing company. This was a good strategy that fitted the criterion of sustainability into the purchasing process.

According to the studies conducted so far, several studies were presented by Kaushik et al. (2020) that applies hybrid of VIKOR and BWM for supplier selection. In their case study done in online retail of fashion apparels, they found 38 initial selection criteria from which some important factors like flexibility in order management, return policy, Just in time delivery, technical compliance, flexibility in billing, etc. were select as final criteria for SS.

In a study conducted for a steel producing company, Badi and Pamucar (2020) developed the mixed hybrid of Grey theory with the MARCOS approach for supplier selection. In the study, a comparative analysis was made to show the comparison on the effectiveness of the integrated approach. This is also evidenced by Chattopadhyay et al (2020) work whereby the authors have used the MARCOS method integrated with the D numbers to assess raw material suppliers in the manufacturing of steel. This made it possible for the evaluation of the suppliers in relation to D numbers with the general objective of ranking the suppliers properly.

In recent times, combining machine learning with traditional MCDM methods has become increasingly popular for addressing the SS problem. In the work of Harikrishnakumar et al. (2019), the authors decided to use the Naive Bayes and SVM techniques of supervised ML to assess the suppliers within the petrochemical industry. Enterprises applying this paper's case study were able to experience a 15% improvement compared to traditional MCDM methods. In another study, Sasaki and Sakata (2021) used Bayesian networks to forecast the supplier, and performed causality analysis of the supply chain network. Among ML methods, Random Forests and Logistic Regression as well as SVM were used for business partner selection with Random Forests achieving the best assessment.

Cheng et al., (2020) combined the various MCDM techniques with ML algorithms for supplier assessment. In their approach, they employed TOPSIS and DEA to give labels to the suppliers; secondly, SVM model was used to categorise unseen new suppliers. In an electronics and automation manufacturing company, use of this framework allowed the professionals to uncover unknown suppliers to improve supplier choice.

2.2 Machine Learning and Classifications

Several ML algorithms exist which can be utilized for classification in different tasks and executions. Among those ML algorithms, decision trees, extremely randomized trees, random forests, and gradient boosting through CatBoost are heavily used for classifications. Decision trees can be described as one of the most extremely versatile ML algorithms which is heavily used in classification, multioutput tasks, and regression problems. This algorithm can incorporate complex datasets and also inherently possesses the understandability of a tree structure (Charbuty & Abdulazeez, 2021a). A typical tree structure in a decision tree consists with decision nodes. These nodes are in a recursive structure in as shown in Figure 1.

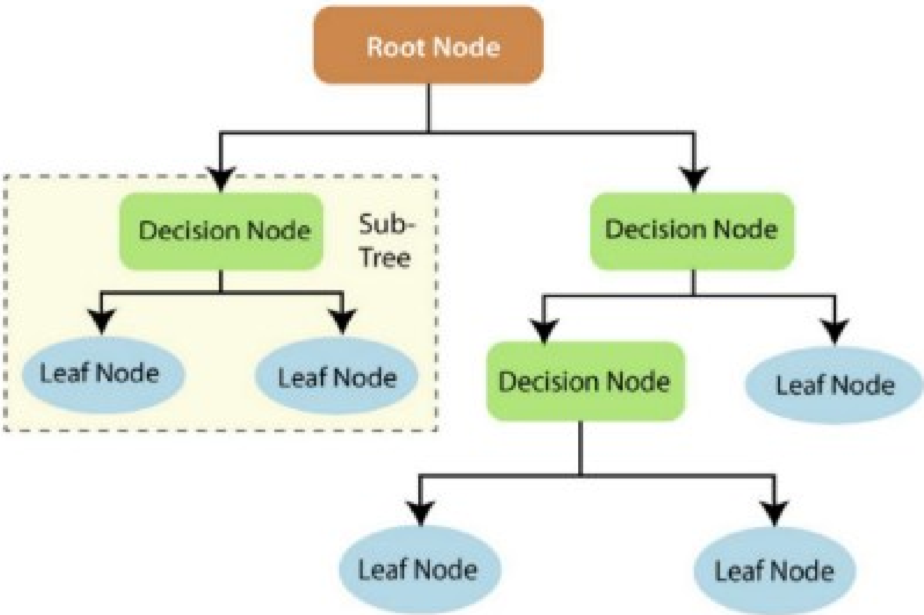


Figure 1.Typical decision tree structure (Charbuty & Abdulazeez, 2021b).

At each decision node, based on the pre-specified criteria and the nature of variables (whether the variables are quantitative or quantitative), the decision to split is determined. The aim of this splitting is to improve the homogeneity using Information Gain or Gini Impurity metrics. The final selection is done when the leaf nodes are reached which is the class which has the most of the samples remaining.

Random forest is combination of multiple decision trees and this method is effective and simple ML algorithm. In the RF, the foundation is the training of several decision tree algorithms which is specifically done based on a specific random subset. Finally, these outcomes are combined, which is often done through a majority voting process.

Additionally, the extremely randomized trees are an expansion of the core concept of random forests, and these are commonly known as Extra Trees which provide a further expanded and more computationally efficient ensemble learning method (Geurts et al., 2006). Extra Trees use a conventional top-down process, in order to create regression trees. Moreover, same as random forests, each base estimator is trained based on a specific random subset. However, in Extra Trees, a random selection is done in selecting the best feature and matching value to split the node (Ahmad et al., 2018). However, while splitting logic is simple, it also weakens the resistance of the algorithm to the noises. However, even though many categorization techniques exist, those differ in their accuracy and efficacy and many studies can be found that have been conducted to optimize them (Saha et al., 2022).

Another ensemble learning technique that is frequently connected to trees is gradient-boosting (Friedman, 2001). However, this technique is also similar to random forests and extra trees as an ensemble technique in the aspect where it generates several models and then integrates their outputs. However, the uniqueness of this technique is that those models are created iteratively to minimize the loss function at each stage. When the categorical features are presented in the dataset, the CatBoost framework can be used (Prokhorenkova et al., 2018). In this framework, a random permutation is used in arranging the training data where the ordered boosting is implemented.

2.3 Supply Chain Management & Supplier Selection

2.3.1 Supply Chain Management (SCM)

SC acts as a bridge between process, flows and stakeholders and it is considered as one of key activities in the organizations (Granlund & Wiktorsson, 2014). Christopher (2016) defines SCM as the strategic coordination of activities or tasks among the company or the business having the focus on improving long term performance both in supply chain and in the business as a whole. Effective SCM is critical to those organizations that would like to be competitive, reduce costs, enhance the quality of their outputs (Lambert et al., 1998). SCM uses activities like sourcing and procurement, conversion, and logistics management to facilitate the effective flow of materials, information, and funds. Integration assists organizations to ensure customers' orders are fulfilled within time and make them competitive in the marketplace. In short, efficient SCM is critical for businesses to operate effectively, fulfil customers' needs, and have sustained success.

2.3.2 Supplier Selection

A key activity of the SC is the supplier selection. This decision-making process enables an organization to lower costs, mitigate risks, and enhance the quality of its outputs (Feng & Zhang, 2017). SS is a responsibility of the procurement function of any company (Giunipero et al., 2012) and this aids in determining the overall supplier performance of suppliers and acts as the driving force behind it (S. A. Khan et al., 2018). However, with the external pressure to drive business towards sustainability, the organizations try to consider environmental, social, and economic, criteria in the process of selecting right suppliers (S. A. Khan et al., 2018). Similarly, since the circular economy concept is emerging to transform waste into resources, this objective also can be achieved by focusing on environmental and socio-economic criteria in supplier selection (Witjes & Lozano, 2016).

2.3.3 Supplier Selection Process and Techniques

The managers and industry leaders consider supplier selection as a decision-making process to sustain a competitive edge in the market (Bai & Sarkis, 2010). The supplier selection has four main stages including defining the problem, defining the selection criteria, determining the competency of the suppliers and identifying their potential, and the final selection. Thus, it is important ensure the accuracy of the first three stages as the quality of the final selection heavily reilies on outputs of first three stages (De Boer et al., 2001).

As noted by J. Chai et al. (2013), the various techniques for decision making can be categorized in three main categories, namely the MCDM methods, mathematical programming techniques as well as artificial intelligence. Some methods of conducting MCDM include AJP, ELECTRE (;Elim-illation et Choice Translating Reality,), ANP, TOPSIS, VIKOR, PROMETHEE, SMART, DEMATEL among them. In the mathematical programming techniques, it is involved linear programming, stochastic programming, multi-objective programming, non-liner programming, and also DEA. The last method of supplier selection procedure is artificial intelligence, which comes under four types such as genetic algorithm, machine learning algorithm, and neural network.

2.4 Sustainable supplier selection (SSC)

The SSC is a supplier selection and evaluation regarding aspects that encompass economic, social, and environmental dynamics (Govindan et al., 2015). Selected suppliers need to fulfil the organization's sustainability goals and promote good practices (Kannan, 2018). The objective is to connect sustainability with the supply chain, enhancing resilience and enabling sustainable development (Seuring & Müller, 2008).

2.4.1 Sustainable supplier selection criteria with their descriptions

Table 1 illustrates the explanations of the most commonly used sustainable SS criteria.

Table 1. Explanations of sustainable SC criteria.

Criteria Number	Dimension	Evaluation Criteria	Description
1	Environmental	Green management	The potential of organizations to improve environmental performance and management
2		Green Manufacturing	Minimizing the usage of energy in the production
3		Green design	Using eco-friendly designs
4		Environment management system	The planning and implementing structure of environmental policies by suppliers to protect the environment
5		Green packing and labeling	Taking environmental considerations into account in the packing and labeling activities
6		Green R&D	Supplier`s ability in investing in clean and green technologies, practices, and methods in their research and development efforts
7		Waste management and pollution prevention	Minimizing the wastage in the production reducing the environmental pollution
8		Environmental competencies	The potential of suppliers to use eco-friendly materials, clean energy, and contribute to less pollution effects
9		Environmental cost	The production is done in a manner where costs and damages to the environment are minimized

10	Economic	Profit on product	Earning a fair and justified profit from the product
11		Price of product	Supplying the product at a reasonable price
12		Flexibility	The capability of suppliers to handle the variations in the market
13		Lead time	Ensuring the product is delivered at the lead time
14		Cost of Transportation	Shipping the products at a minimum transportation cost
15		Quality of product	Supplying the products with good quality level
16		Delivery and Service of product	Ensuring the delivery and service
17		Technological and Financial Capability	The capability of suppliers in the technological and financial aspects
18		Production facilities and capacity	Having the optimal production capacities and facilities to produce the product
19	Social	The rights of stakeholders	Ensuring the moral of sense of protecting the rights of society in the business
20		The rights of employees	Protecting the employee well-being and rights by addressing the employees’ requirements to ensure long-term sustain
21		Occupational safety systems	Ensuring the safety, welfare, and health of the staff
22		Information disclosure	Ensuring information transparency through making details such as

			used materials, carbon emission levels, toxins produced, etc. available to the stakeholders and customers

2.4.2 Sustainable Supplier Selection and MCDM

In particular, MCDM methods can be used for selecting suppliers efficiently with regard to sustainable criteria. For instance, Büyüközkan & Çifçi (2011) applied a fuzzy-based ANP technique to respond to the incomplete information gathered from Turkey's white goods industry for developing an efficient sustainable supply chain management by selecting the best suppliers. am Respective to this policymaking, this framework took into account several factors that were economic, environmental, and social. Likewise, Büyüközkan (2012) recommended a fuzzy AHP-based model for SS that sorts, quality, cost, service performance, and environmental performance, as evaluation criteria.

Freeman & Chen (2015) developed also a green supply chain supplier selection proposal that used AHP, TOPSIS, and entropy techniques. In the context of the identified criteria, AHP was applied in defining the weights of the criteria that consisted of green competency and environmental performance ratings and the entropy technique was incorporated to assess the importance of the criteria as well Supplier ranking was executed with the help of the TOPSIS technique. In the same year, Darabi & Heydari decided to use interval-valued fuzzy entropy to assess the suppliers with green supply chain factors including cost, quality, environmental competency, delivery and technological capability. Moreover, Fallahpour et al. (2017) provided a detailed classification of the criteria related to sustainable supplier evaluation with regard to 46 criteria based on the economic, social, and environmental aspects. Identification of criteria for the purpose was achieved through the Analytical Hierarchical Process (AHP) while fuzzy TOPSIS was employed to rank the top ten suppliers.

Gören (2018) developed a sustainable supplier selection model where the DE-MATEL approach was used to assess the importance of criteria and the Taguchi simulation for evaluating the suppliers. This framework included ER Scope, Environment, Employee,and Ecoproduct initiatives and addressed Environmental, Social and Economic aspects based on the type of consumption of an online retailer business in Canada. Guarnieri & Trojan (2019) proposed the approach for selecting the ethical and sustainable suppliers for the textile industry by categorizing the criteria through ELECTRE-TRI combined with AHP considering the weights. Finally, Li et al. (2019) developed a novel rough cloud TOPSIS model for sustainable SS under uncertain environment and then tested it in a state-owned energy company of China.

2.4.3 Recent studies about sustainable SS in different industries

According to the comprehensive literature review, it can be identified that the green/sustainable supplier selection criteria and methods depend on the industry and the studied case. Thus, Table 2 illustrates the summary of the recent studies about sustainable supplier selection in different industries including used methods, and criteria for the selection.

Table 2. Summary of the recent studies about sustainable SS (source: author).

Authors	Methods	Economic	Environmental	Social	Industry
(Awasthi et al., 2018)	Fuzzy VIKOR, AHP	Quality, flexibility, speed, cost, innovativeness, dependability	Energy water, biodiversity, emissions, materials	Labor practices, society, human rights, product responsibility	Electronic goods manufacturing industry

(Azimifard et al., 2018)	TOPSIS and AHP	Distance	CO2 emissions, water consumption	Number of employees	Steel industry
(Bai et al., 2019)	TODIM and Grey BWM	-	-	Employee training education, safety management system, rights of employees, employment practices, information disclosure, contractual stakeholders' influence	Automobile manufacturing industry
(Guarnieri & Trojan, 2019)	ELECTRE-TRI and AHP	Quality, cost, on time delivery, financial stability, service efficiency, flexibility, delivery delays, R&D	Emission control, eco-friendly packaging, reverse logistics, waste management, green image,	Public disclosure, human rights, management skills, certifications, philanthropy,	Textile industry

		level, production capacity	environmental management	legal compliance	
(Lo et al., 2018)	Fuzzy TOPSIS, BWM	Innovation capability, flexibility, product quality, financial stability	Green logistics, green manufacturing, environmental performance	-	Electronics company
(Abdel-Baset et al., 2019)	VIKOR, ANP	Transportation cost, cost , revenue	Green manufacturing, trash management, green packing	Ethical issues and legal compliance, vocational health and safety system	Importing industry
(Li et al., 2019)	Rough cloud TOPSIS	On time delivery, flexibility, cost, quality	Green logistics, green purchasing. Green development, green production, green design	Philanthropy, employee rights, health and safety, fight for fair trading and corruption, human rights	State-owned energy company

(S. A. Khan et al., 2018)	Fuzzy inference system, Fuzzy Shannon entropy	Service reliability, financial capability, cost, quality	Resource consumption, environmental management system, cleaner technology availability, eco-friendly material, recycled material	Health and safety, employee rights, social commitment, employment practice	Automobile manufacturing company
(Pishchulov et al., 2019)	Voting AHP	Financial capability, technical capacity, delivery, quality, responsiveness, logistics and production, management	Waste management, water usage, environmental management, energy consumption, material consumption, environmental commitment, environmental product performance	Occupational health and safety, social management, diversity, training, working hours and wages, employment relationship, society social commitment	Wood industry

2.5 Circular economy & Circular Supply Chain Management

Circulate economy is an economic strategy of using all the resources without disposal in a manner that enables their repetitive usage. The conventional model is the linear economy where the use of resources is “take-make-dispose.” On the other hand, CE uses “make-use-use” approach. In this regard, the input or waste stream is returned back into the supply chain operations by recycling, reusing, refurbishing or remanufacturing processes to form a closed loop with the help of value restoration and managing the adverse impacts (Yang et al., 2018).

In CSCM, various methods of the CE are utilized. It also establishes the closed-loop cycles whereby the entity should be useful for reuse, remanufacturing or recycling. It is done by focusing on reduction of waste, conservation of resources and sustainability (Batista et al., 2018). SCM in circular economy requires cooperation from the suppliers, manufacturer all the way to the customers.

2.5.1 SS Problem in CSC

To tackle the supplier selection challenge in CSC, a number of scholars have developed standards and models. A strong hybrid MCDM framework that was deployed by Senthil et al. (2014) for selecting contractors. Accordingly, the rank of the contractors was established by applying fuzzy TOPSIS technique, while the weight of the criteria was evaluated by applying the AHP technique. Another model which combines between MCDM was proposed by Azadeh et al. (2016) in an effort to solve the supplier selection problem in closed-loop supply chain with the help of Taguchi and DEA. In this approach the authors have also used five criteria namely cost, product capability, service, delivery time and technology for selection criteria. Ref [34] discussed the network design and the supplier selection in the closed-loop CSC. The three attributes that the authors of the study used in the SC and supplier selection include lean, environmental, and agile. Based on AHP, a rating was then made of the various vendors to be considered for partnership with the outsourcing firm.

Ahmadi and Amin (2019) also proposed a fuzzy theory based MCDM to evaluate different suppliers in mobile phone industry's closed loop CSC. In the supplier selec-tion problem, the authors included the following: Durability, Recyclability, Market Credibility, design requirement, Performance capability, energy consumption, process safety, Lead time, Reject rate, and the Cost of materials as the Supplier selection criteria. Yet, a comprehensive concept of selecting the suppliers in the reverse SC was put forward by developing a fuzzy TOPSIS, fuzzy AHP, and mathematical model as mentioned in the paper of Govindan et al. (2019). In the SS evaluation, the authors integrated all the three economic, social, and environmental aspects in the supplier selection criteria and then finally developed the proposed frame-work among the electronic industry.

2.5.2 Sustainable SS Problem in CSC

Many studies have investigated the problem of SS within the context of circular, green and SSC by exploring the criteria and approaches appropriate for their selection. Supplier selection as a domain has shifted to CSC, and this has taken several studies with it. These changes are conscious and can be attributed to the global concerns on resource optimization, economies of utilization and line with the circular economy concepts. In the area of supplier knowledge, it is possible to distinguish three types of references: green and sustainable supplier selection, general supplier selection of the SCM and SSCM.

The latter of research dealing with sustainable SS in the realm of CSC has not, thus, been much explored; therefore, suggesting a research gap. It will therefore provide a research opportunity in determining the most suitable sustainable supply chain strategies for CSCs in term of supplier selection. Therefore, this research will measure supplier selection in as a respect of SSCM as a means of catering to the above identified need.

To this end, Kusi-Sarpong et al. (2023) made a historical contribution to the course of the CSC theoretical advancement by presenting a framework for the evaluation of vendor sustainability through the two methodologies of BWM and VIKOR. The study also determined the important weights with the help of Best Worst Method (BWM) followed by the ranking of the providers by VIKOR. To achieve the above mentioned research objectives, the following research methodology was used to evaluate the sustainability of suppliers in the context of textile manufacturing sector. It was found that the completeness of the framework was around 75%, and the investigating team observed that differences in the details provided by the matched pair of comparisons for the experts in different topics were very common. Additionally, this framework did not include factors concerning the selection in economic, social, sustainability aspects that are significant to the processes of CSC procedures.

Based on the above-mentioned restriction, Kusi-Sarpong et al. (2023) purposively developed an integrated methodology using interval-valued VIKOR and fuzzy BWM for SS in SCM with the consideration of SDG goals. This was illustrated by the wire and cable business with demonstration on how the methodology of this solution is put into practice. The new idea aims at providing better solution toward integrating sustainable decisions in specific economic, social and environmental contexts to the buying firm as well as improving SS decisions within CSC.

However, the author of the report agreed that the current state has poor understanding regarding the true concept of sustainable SS for the management of CSC which caused the need for further research to develop managerial frameworks to address the other issues that were identified. This study intends to fulfil this gap by assessing and improving the SS management in the CSC context and by incorporating sustainability aspects.

2.5.3 SS Criteria for CSCs

SS is indeed a critical aspect that encompasses multiple factors that are sometimes complementary and sometimes non-mutually exclusive. Thus, this process highly depends on the choice criteria and the analysed business environment. Thus, in most cases the criteria for selecting suppliers are typically distinguished as the economic and non-economic. Moreover, according to the type of the problem deals with, there are many categories of the SS criteria addressed in the literature. In the current circumstances of the development of CSC, there are many studies and industry specialists distinguish three primary categories of supplier selection criteria: social, economic, and circular (Tavana et al., 2023).

2.5.3.1 Economic Criteria for SS

Product costs are heavily influenced by purchasing costs, and the purchasing function significantly raises overall production costs. The purchasing team and organizations consider selecting suppliers who provide at a lower price so that the operational cost can be reduced (Wu et al., 2021). Therefore, one of the most crucial economic factors taken into account throughout the supplier selection process is cost, which can take many various forms and fall into several categories including price, operational costs, product cost, cost of logistics services, and transportation costs. Moreover, in the context of economic criteria, the cost is not the only supplier selection criteria, but also there are several other economic aspects such as on-time delivery, risk and quality (N. Chai et al., 2023). Table 3 illustrates the criteria for supplier selection considering the economic dimensions in different industries.

Table 3. SS criteria for economic dimensions.

Industry	Economic Criteria	Authors
Manufacturing	Financial Capability, Quality, Cost	(Orji & Ojadi, 2021)
Food	Delivery Flexibility, Cost, Quality, Financial Stability	(Fallahpour et al., 2021)
Banking	Quality, Financial Capability	(Ojadi et al., 2023)
Petrochemical	Cost, On-time Delivery, Financial Capability, R&D	(Alavi et al., 2021a)

Aerospace & Defense Industry	Cost, Delivery, Quality, Financial Stability	(Rasmussen et al., 2023)
Mining	Quality, Innovation, Financial Capability, Service Level	(Ortiz-Barrios et al., 2021)
Electric Vehicle	Cost, Quality, Price, Service	(Wei & Zhou, 2023)
Oil Refinery	Quality, Delivery Reliability, Supply Capacity, Price, Service	(Hailiang et al., 2023)

2.5.3.2 Criteria for Circular SS

The government and outside parties are putting a lot of pressure on supply chains these days to transition to circular supply networks that promote sustainability. Thus, the first step of shifting towards circular supply chains requires selecting and collaborating with green and sustainable suppliers. Sustainable suppliers are different from traditional suppliers as they significantly consider the environmental factors in their supply chains including manufacturing, packaging, designing, and distributing products. As the circular economy concept emerged, the industries moved their environmental activities focusing on zero waste concepts. According to the literature review, some of the commonly utilized circular criteria in the supplier selection process are pollution controls, environmental management systems, green technology, and sustainable packaging. According to the literature, the first research to consider the circular economy concept and circular criteria in supplier selection problems was (Kannan et al., 2020). Table 4 illustrates the criteria for supplier selection considering the circular dimensions in different industries.

Table 4. SS criteria for circular dimensions.

Industry	Circular Criteria	Authors
Petrochemical	Using recyclable raw materials, reducing air pollution through production, waste management, utilizing	(Alavi et al., 2021b)

	clean and green technology, an environmental management system, eco-friendly packaging, and compliance with environmental regulations and standards	
Automotive	Using green and clean technology, recycling, reuse, green R&D, green packaging, reducing carbon emissions, having environmental certification, eco-friendly parts	(febGovindan et al., 2020a)
Textile	CE favorable policies, top management dedication towards CE, sustainable product life cycle management, organizational culture towards CE, using recyclable raw materials, eco-friendly packaging, reusing products	(Nasr et al., 2021)
Manufacturing	Green products and technology, pollution control, environmental management system, using recyclable and eco-friendly materials	(Liu et al., 2022)
Oil Refinery	Pollution control, environmental management system, green technology, energy conversation	(Hailiang et al., 2023)
Construction	Environmental management system, eco-friendly materials, green technology, and pollution control	(Tushar et al., 2022)
E-bike Sharing	Environmental management system, pollution control, product recyclability, eco-friendly design	(N. Chai et al., 2023)

Several articles focusing on MCDM techniques in selection of suppliers have been documented in the available literature. Moreover, these MCDM techniques have been improved in the situations where the integration of ML and Fuzzy logics are done with the MCDM techniques. The use of ML and the TOPSIS technique in supplier selection in CSCs is still a relatively unexplored area in literature even though there has been a discussion on MCDM Techniques like AHP and TOPSIS for SS. Although some other researchers have tried to integrate Fuzzy Logic and ML into such methods (Azadeh et al., 2016), to the best of our knowledge, no prior works have focused on the exact application of ML and TOPSIS in the context of CSC.

The purpose of this current study is to bridge this gap by integrating capability of ML to tackle a large and complex datasets with the best traditional TOPSIS method to offer a framework that will complete the CSCs supplier selection system. Also, this research counts for the growing trend of publications on sustainability and CSCs with an interdisciplinary approach on how the latest technologies can be utilized for addressing the issues of supplier selection in CSCs.

3 Methodology

3.1 Research Design

There are four primary stages to this study's research design. These four main stages aids in understanding the supplier selection criteria through literature, collecting data, developing the ML-integrated TOPSIS model and then evaluate the collected data. Figure 2 illustrates the overview of the research design of this study.

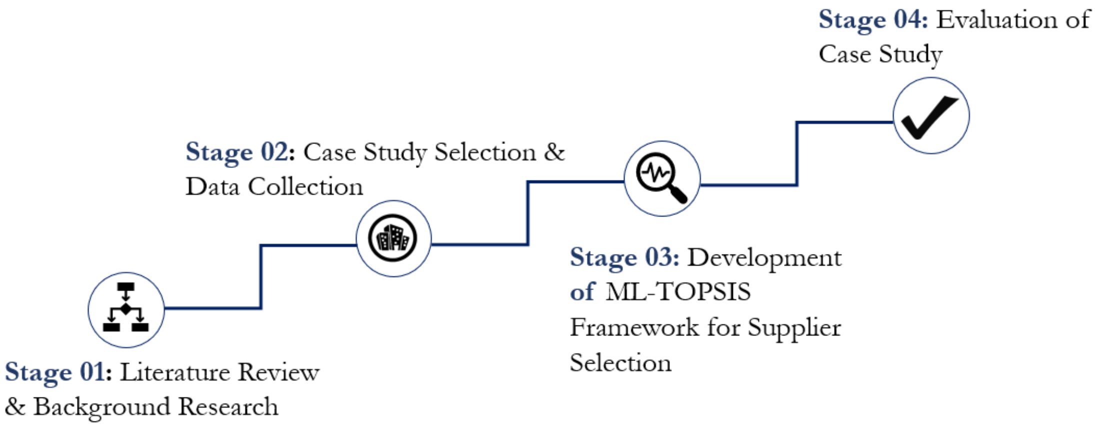


Figure 2. Overview of research design.

- Stage 01: Literature Review and Background Research This step's objective is to determine the SS criteria for CSCs and use of TOPSIS framework in the supplier selection.

- Stage 02: Case Study Selection and Data Collection - This step's objectives are to choose the case company and then gather information from the appropriate sources.

- Stage 03: Development of ML-TOPSIS Framework for Supplier Selection- The following is the objective of this stage, to develop ML-TOPSIS framework in order to analyze the data collected from the research.

- Stage 04: Evaluation of Case Study - This step's objective is to assess the gathered data using the TOPSIS framework's integrated machine learning for supplier selection.

Let’s discuss each stage to understand the research design and the methodology of this research study in depth.

3.2 Stage 01 – Literature Review and Background Research

This phase of the research was discussed under Chapter Two – Literature Review – of the current study. The detailed discussion in this chapter presents the factors for selecting the supplier for circular supply chain. It also provides an insight into a particular approach used in supplier selection, which is TOPSIS, in the domain of CSCM.

3.3 Stage 02 – Case Study Selection and Data Collection

3.3.1 Case Selection

One of the most important steps in this research is choosing an appropriate case study to use. This will ensure that the findings are applicable and relevant to the situation at hand. This research focuses on the development of a framework for selecting suppliers in CSCs that incorporates Machine Learning (ML) and the TOPSIS method. A real-world setting is provided by the case study that was selected for the purpose of evaluating the practicability and applicability of this framework.

3.3.2 Description of the case company

This study focuses on circular supply chains with an emphasis on sustainability and resource efficiency. The selected case company is in the food and beverage industry, which is relevant to the study because it discusses circular supply chains. The practices of CSCM have been implemented by this multinational organization, which also has a vast international network. Due to the fact that the company employs environmentally friendly packaging, recyclability, waste reduction, and sustainable sourcing practices, it is an excellent candidate for the investigation of advanced supplier selection strategies. The study's focus is to investigate the incorporation of ML and TOPSIS in CSCs, which are in line with the principles of a CE.

The case company operates in a market that is highly competitive, and the effectiveness of the suppliers that it chooses to work with has a significant impact on the market positioning that it achieves. As a result, SS is of the utmost importance in order to guarantee cost-effective suppliers who are able to fulfill the quality standards and sustainability requirements of the company. The findings of this study can be generalized to other organizations that have operating environments that are comparable to those of the company, despite the fact that the identity of the company is kept confidential in order to protect sensitive information. The case study helps to shed light on real-world challenges associated with supplier selection in CSCs and demonstrates how ML and TOPSIS can improve DM in this area. It is anticipated that the findings will close the gaps that have been identified in both the existing literature and the practices of the industry.

3.3.3 Data Collection Methods

This is an important part of this investigation because it is the means by which the recently developed TOPSIS-ML model for supplier selection in circular supply chains is evaluated in terms of its quality and effectiveness. The data that will be used in this study will be carefully selected to ensure that they are of the highest possible quality and appropriateness, which will ultimately result in a reliable analysis. This will help to reduce the likelihood of confounding variables occurring in the study. Increasing the generality of the findings and, as a result, more accurately evaluating the supplier selection will be accomplished through the utilization of both primary and secondary data gathering techniques. This will allow for the maximum amount of information to be obtained.

3.3.3.1 Primary Data Collection Method

This data is gathered by means of an online questionnaire that is directed towards procurement and sourcing professionals working for the case company. The purpose of the questionnaire (Appendix 1) is to collect information related to the manner in which these professionals prioritize different supplier selection criteria when selecting the suppliers who are the most suitable for CSCs. The data that was collected is analyzed using regression analyses, which are carried out in Google Colab using Python (Appendices 3 to 6). These analyses are used to identify patterns and correlations in the data. The specifics of the data that was gathered are elaborated upon in greater detail in Section 4.2.

3.3.3.2 Secondary Data Collection Method

This data is collected from the system records of the company that is being investigated. The information that is included here pertains to evaluations, proposals from suppliers, and parameters for making decisions. Within Section 4.1, the specifics of these secondary data are broken down and explained.

Cross-referencing of multiple data sources is performed, and ethical considerations, such as confidentiality agreements, are adhered to in a stringent manner, in order to guarantee the dependability and safety of the data.

The collection of data is carried out in a methodical and multifaceted manner, which guarantees that all pertinent aspects of the supplier selection process are taken into consideration. It is through the incorporation of a wide variety of data sources that the findings are strengthened, which ultimately results in valuable insights and recommendations for the management of sustainable supply chains. Gaining these insights will assist in the development of a framework for DM that is driven by data and has the potential to optimize supplier selection strategies across a variety of industries.

3.4 Stage 03 – Development of ML-integrated TOPSIS Framework for Supplier Selection

In this stage, the ML-integrated TOPSIS framework for supplier selection of circular supply chains is developed. Figure 4, illustrates the framework of the ML-integrated TOPSIS model for supplier selection.

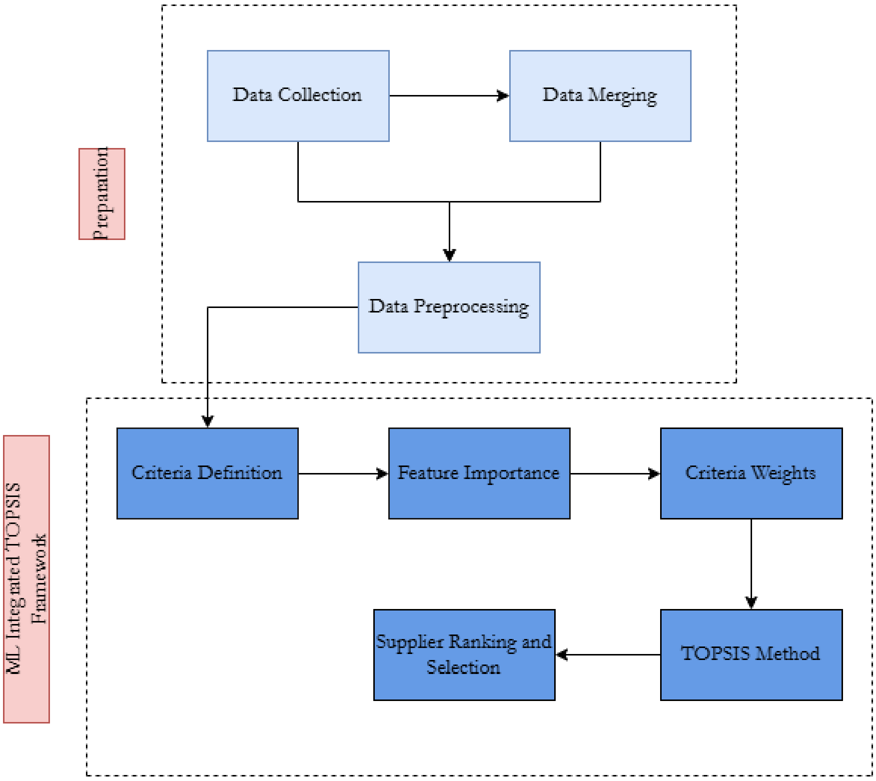


Figure 3. Framework of the ML-integrated TOPSIS model.

In this stage, the framework for selecting suppliers in CSCs is developed by incorporating the techniques of machine learning and the TOPSIS method. With the help of this framework, the process of making informed decisions regarding the selection of suppliers is improved substantially. In order to provide rational and data-driven decision-making processes for the intention of evaluating and selecting suppliers, machine learning techniques are utilized for the purpose of data imputation and feature importance analysis. On the other hand, the methodology that is being implemented here is applicable to the majority of the operations that take place within the context of the SC, thereby enhancing the adaptability, scalability, and robustness of the applications that are used in the industry.

3.4.1 Preparation Phase

The first step in evaluating the suppliers is the “Preparation Phase”. In this stage, the collected data needs to be collected, cleaned and pre-processed before feeding into the ML models and use in decision-making process. An efficient and effective preparation process aids in a reliable dataset and analysis.

3.4.1.1 Data Collection

This is an important step in the analysis process since it ensures a comprehensive and diverse dataset is created to evaluate the suppliers accurately. Information emanates from sources such as the history of supplier performance, historical data on the reliability of suppliers, reports with regard to compliance issues in the environment, and customer feedback. The data sources are categorized into three main dimensions namely; economic, social, and environmental. These play a huge role in assessing the performance of suppliers in a CSC context.

Economic factors may include cost-efficiency, price stability, and financial stability. Supplier’s sustainability practices, such as waste reduction, energy consumption, and compliance with environmental standards are a few factors included under Environmental factors. Social factors take into account aspects such as labor conditions, ethical sourcing, and community engagement. Both quantitative and qualitative aspects of supplier performance are captured to ensure that all relevant evaluation criteria are being considered in the analysis.

3.4.1.2 Data Merging

Data merging is important to ensure consistency in the data presentation as well as its completeness. After 3.4.1.1 step, the data from various sources needs to be integrated into a single data center. This involves merging datasets that come from different formats, scales, and units. This step helps to establish a single, complete representation of each performance across the analyzed evaluation criteria.

The different sources' attributes are matched with due care. For instance, cost data, if from one source, must be converted to units that match those of environmental or social data. Other merging includes the resolution of conflicts and redundancies from the dataset so that the integrity of the latter is maintained.

3.4.1.3 Data Pre-processing

Data preprocessing is required to ensure the data is consistent, clean, and ready for analysis. This is carried out before implementing analysis techniques. Handling missing data, normalizing numerical attributes, and ensuring that data adheres to the required formats are addressed in this stage.

Handling Missing Data - Real-world data often has the problem of missing data. Missing values need to be addressed since it can lead to biased results. Therefore, effective imputation techniques are applied to fill in the missing values. The following are some techniques which will be used in the data preprocessing process:

- K-Nearest Neighbors (KNN) Imputation: KNN imputation finds the k nearest data points using the similarity of the other features and imputing the missing values by the average of the values of those identified neighbors. It works well where there is a correlation between features.

$$X_j^{imputed} = \frac{1}{k} \sum_{i=1}^{k} X_{ij} \tag{1}$$

(Frossard et al., 2016)

- Regression Imputation: Regression imputation is training a model using the features that have complete data to predict the missing values. It uses the relationships between the attributes to provide an estimate for the missing values.

$$X_j^{imputed} = \beta_0 + \sum_{i=1}^{n} X_i\beta_i + \varepsilon \qquad (2)'$$

(Little & Rubin, 2019)

- Mean/Median Imputation: Missing values are imputed using either the mean or median of the available data when there are no such relationships as mentioned above are present. This method is simple but robust, specifically in cases where the missing data is insignificant.

- Normalization - Normalization is done to bring all the different units to a single scale. Different features could be in different measurement units such as monetary units, weights, etc. Normalizing the numerical data is required to ensure that all criteria can be compared effectively. Each feature is normalized into a scale between 0 and 1. This ensures comparability of all the features proportionately without biasing the results.

The normalization formula used is:

$$X_{ij}^{normalized} = \frac{X_{ij} - \min(X_j)}{\max(X_j) - \min(X_j)} \qquad (3)$$

(Ali, 2022)

X_{ij} represents the original data value for criterion Cj for supplier Ai

$\min(X_j)$ and $\max(X_j)$ are the minimum and maximum values for criterion Cj

3.4.2 ML-Integrated TOPSIS Framework

The next step in the methodology is the integration of ML techniques with the TOPSIS framework for supplier evaluation. Defining evaluation criteria is the starting point of this phase followed by feature importance analysis, weight assignment, and finally the application of the TOPSIS method for ranking suppliers.

3.4.2.1 Criteria Definition

The criteria are selected so that they are aligned with the specific goals of a circular supply chain. Sustainability, cost-efficiency, and social responsibility are critical factors. The criteria will encompass a mix of factors that measure various dimensions of performance by a supplier.

The categories may include economic ones, such as total cost, delivery time, or price stability; environmental ones, like carbon emissions, waste management practices, or compliance with eco-friendly regulations; or social ones: working conditions, ethical business practices, and community engagement.

3.4.3 Feature Importance Analysis

The next step, after feature transformation, is to determine the importance of each criterion in the process of supplier selection. This is done by using Random Forest, a machine learning algorithm particularly suitable for the assessment of the importance of features in both classification and regression problems.

Random Forest measures feature importance as a decrease in impurity, for example, the Gini impurity, when the specific feature is used for a split in every decision tree. Feature importance for each criterion is computed as:

$$\mathrm{FI}(\mathrm{C_i}) = \sum_{\{t=1\}}^{T} \frac{\Delta \mathrm{I}_{t(\mathrm{C_i})}}{|\mathrm{N_t}|} \tag{4}$$

(Li et al., 2019)

Where:

$\Delta I_{t(C_i)}$ represents the reduction in impurity for criterion in decision tree t.

$|N_t|$ is the number of nodes in tree t.

T is the total number of trees in the forest.

A weight is given to each criterion after important scores are normalized. This ensures that the weights represent the relative significance of each criterion:

$$W_i = \frac{FI(C_i)}{\sum_{\{j=1\}}^{n} FI(C_j)} \tag{5}$$

(Agarwal et al., 2023)

Where:

W_i represents the weight assigned to criterion C_i.

n is the total number of criteria.

3.4.4 Weight Assignment

The weights derived from the Random Forest analysis are directly used in the TOPSIS method after normalizing. These weights provide a data-driven approach for assigning importance to the different criteria, ensuring that the DM process reflects the relative importance of each aspect of supplier performance. The data-driven nature of this process makes it more accurate and adaptive to changing conditions and priorities in the supply chain.

3.4.5 Application of TOPSIS Method

Next, the suppliers are ranked by applying TOPSIS based on the weighted criteria, following a series of systematic steps.

3.4.5.1 Construction of the Decision Matrix

The decision matrix D is created, with each alternative A1, A2,… representing a supplier, and each criterion C1,C2,… representing a specific aspect of supplier performance. The decision matrix captures the scores for each supplier on each criterion.

3.4.5.2 Normalization of Decision Matrix

Next, to compare the different criteria, normalization is applied to the decision matrix. The normalized matrix R_{ij} is computed using the following formula:

$$R_{ij} = \frac{D_{ij}}{\sqrt{\sum_{k=1}^{m} D_{kj}^2}} \tag{6}$$

(Behzadian et al., 2012)

Where:

R_{ij} is the normalized score.

D_{ij} is the original score of alternative i on criteria j.

3.4.5.3 Weighted Normalized Decision Matrix

Next, the normalized decision matrix is adjusted based on the criteria importance. The weighted normalized matrix V_{ij} is calculated as:

$$V_{ij} = W_j \cdot R_{ij} \tag{7}$$

(Behzadian et al., 2012)

Where:

W_j is the weight criterion.

V_{ij} is the weighted normalized score.

3.4.5.4 Determination of Solutions

The next step of the TOPSIS method is to identify the Positive Ideal Solution (PIS) and Negative Ideal Solution (NIS). These elements illustrate the best and worst possible performance of each criterion.

$$V_j^+ = max_i V_{ij}, \quad V_j^- = min_i V_{ij} \tag{8}$$

(Behzadian et al., 2012)

Where:

V_j^+ is the best possible value for criterion j.

V_j^- is the worst possible value for criterion j.

3.4.5.5 Calculation of Euclidean Distances

This is determined to quantify the closeness to the ideal solutions.

$$S_i^+ = \sqrt{\left\{\sum_{j=1}^{n}\left(V_{ij} - V_j^+\right)^2\right\}}, \quad S_i^- = \sqrt{\left\{\sum_{j=1}^{n}\left(V_{ij} - V_j^-\right)^2\right\}} \tag{9}$$

(Pavić & Novoselac, 2013)

Where:

S_i^+ is the distance from the Positive Ideal Solution.

S_i^- is the distance from the Negative Ideal Solution.

3.4.5.6 Computation of Performance Score

The relative closeness coefficient CC_i can then be calculated for each supplier as:

$$CC_i = \frac{S_i^-}{S_i^+ + S_i^-} \tag{10}$$

(Pavić & Novoselac, 2013)

3.4.6 Implementation of the Framework

The above described ML-integrated TOPSIS frameworks is implemented using Python as the programming language and these scripts are supported by the libraries such as NumPy, Pandas and Scikit-learn improving the capability of handling a large dataset. The implementation stage incorporates the following steps:

- Data Preprocessing: This step includes cleaning the duplicates, and missing values, merging data, and normalization.

- Feature Importance Analysis: In this step, the Random Forest algorithm is trained. Thereafter, the feature importance values are obtained and finally, these obtained values are normalized to obtain the criteria weights.

- TOPSIS Method Analysis: The TOPSIS method is used to rank suppliers by evaluating the weighted decision matrix.

In summary, this methodology presents a comprehensive data-driven approach toward the selection of suppliers in circular supply chains. Integration of machine learning techniques, such as Random Forest for feature importance analysis, along with advanced decision-making frameworks like TOPSIS, makes the proposed methodology robust, scalable, and reliable for evaluating suppliers. This methodology solves not only the problems of missing data management and multi-criteria evaluation but also follows all the rules of the CE since in the decisional process, economic, environmental, and social issues are considered.

3.5 Stage 04 - Evaluation of the Case Study

Stage 04 is one of the most crucial stages in assessing the efficiency and feasibility of the suggested supplier selection process. During this stage, any emerging problems are identified, the effectiveness of the proposed approach is tested, and the overall performance of the suggested framework is evaluated.

In the case study, the supply chain scenario or industry under focus is first specified, along with the criteria for identifying the number of suppliers, such as cost and their impact on the environment and society. The performance of each supplier is then evaluated using the aforementioned methodology, relying on actual data for all three dimensions of evaluation. This process includes checking whether the pre-processing methods, such as filling in missing values and normalizing data, produce accurate, consistent, and comparable data for analysis. Ensuring the validity of the data is critical, and any issues are addressed prior to the evaluation process through appropriate machine learning-based imputation techniques.

Additionally, feature importance analysis using Random Forest or other classification and regression algorithms plays a key role in assigning appropriate weight values to each evaluation criterion. This step is vital for creating the final decision matrix. The correct allocation of weights to individual factors ensures a fair and logical distribution of relative importance when ranking the suppliers. Based on these weighted criteria, suppliers can be compared with the objective ranking position, which will form the foundation for deciding on the best supplier to support the organization's circular supply chain.

The case study also demonstrates how the implementation of the TOPSIS technique facilitates the ranking of suppliers based on the distance of each solution from the ideal solution. By considering both positive and negative ideal solutions, a comprehensive evaluation of all possible supplier options is made possible. The performance scores derived from this process will provide an accurate ranking of the suppliers, offering decision-makers valuable insights for selecting the most suitable supplier for CSC implementation.

Furthermore, in the analysis section, the scalability of the methodology is assessed. Since the framework utilizes machine learning techniques, it is anticipated that it can be applied to various industries and SCs. The study should also address potential issues that may arise during evaluation, such as the ability to scale the methodology for handling large datasets or the effectiveness of the imputation techniques used.

Finally, Stage 04 examines how the integration of ML and the TOPSIS method enhances the supplier selection framework, making it adaptable and applicable in different scenarios within circular supply chain management. This stage is essential for ensuring that the methodology can be effectively used across diverse circumstances, optimizing supplier selection in a variety of supply chain contexts.

4 Results Analysis

4.1 Data Collection and Processing Strategy

4.1.1 Introduction

The methodology employed in this study combined both primary and secondary data collection techniques to enhance the credibility of the research. The primary data was collected from procurement and sourcing professionals within the case company through an online survey. The survey gathered information on how these professionals ranked the importance of various SS criteria to identify the suppliers for CSCs. The questionnaire used for this data collection is provided in Appendix 1.

Along with the primary data, secondary data was gathered from the case company's records, such as procurement documents and supplier evaluation reports. These records were sourced from the company's procurement department and are detailed in Appendix 2.

Given that this study specifically focuses on circular supply chains, the supplier appraisal process went beyond traditional factors. It also incorporated sustainability considerations, such as carbon footprint, ethical sourcing, and recyclability, which are essential components of circular supply chain principles. This approach ensured that the SS process aligned with both operational and environmental sustainability goals.

4.1.2 Data Sources and Integration

When collecting the secondary data from the case company, the data was originally extracted from two distinct systems:

- System 01 – This system records the details of the proposals submitted by the candidate suppliers. Thus, from this system records of supplier offer were extracted.
- System 02 - This system records the details about the supplier evaluations, decision-making criteria, and sustainability assessments. Thus, past supplier performance data and sustainability assessments were extracted.

Thereafter, the two main datasets were integrated using the procurement request numbers as the unique key or the primary key which aids in integrating past supplier performance data with supplier offer data. In the initial dataset, there were around 2184 rows which were then followed by the pre-processing steps to maintain the data integrity and generate a final dataset which statistically robust for ML applications.

4.2 Analysis of Questionnaire Data

A structured online questionnaire was distributed to procurement and sourcing professionals of the case company. The main aim of the questionnaire is to validate and recognize supplier evaluation criteria, identify procurement priorities, and understand sustainable priorities. In the survey study, 22 responses from the case company (specifically from procurement and sourcing departments) were collected through online distribution. The questionnaire responses were used to identify the most important SS criteria from the procurement records from the procurement professionals’ point of view and also identify any additional sustainability factors that might influence the supplier selection for circular supply chains.

4.2.1 Demographics of the survey responses

In this section, an overview of the demographics (age, years of experience, department of working) of the 22 survey respondents are provided.

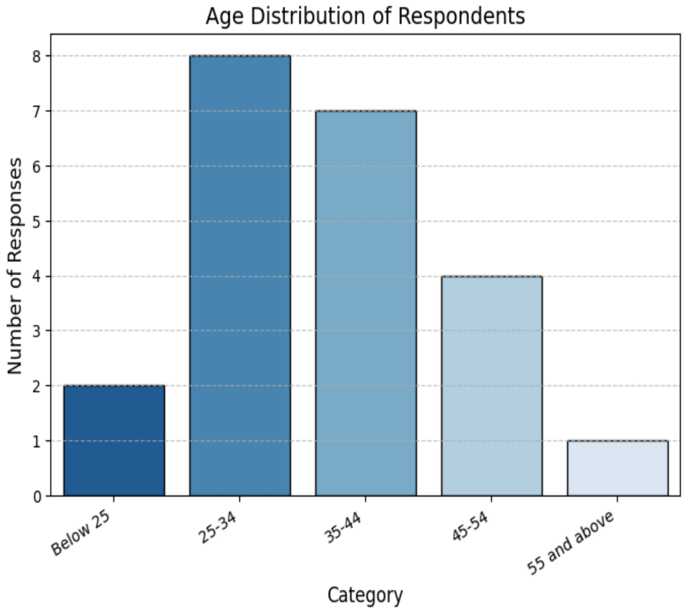


Figure 4. Age Distribution of the Respondents.

According to the Figure 4, the highest number of respondents was from age group between 25 and 34 and the followed by age group 35 and 44, and age group 45 and 34 respectively. The least number of responses was recorded from the age group between 55 and above.

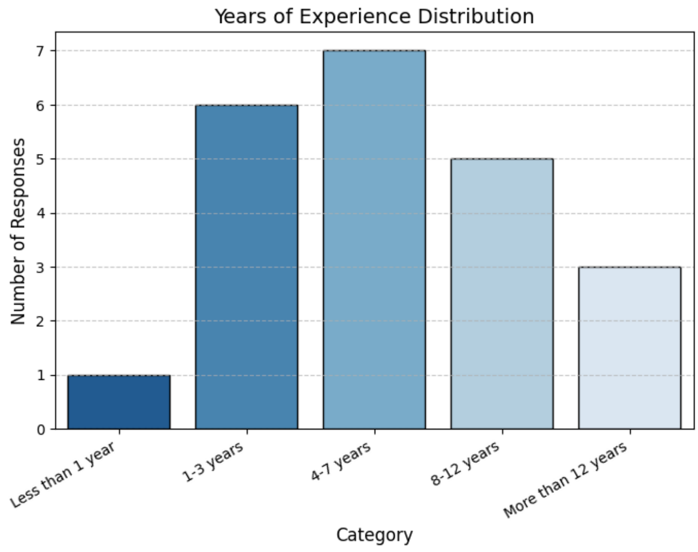


Figure 5. Years of Experience Distribution of the Respondents.

According to Figure 5, it can be seen that the majority of the respondents have experiences between 8 to 12 years whereas the least number of years of experience recorded from less than 1 year experience group.

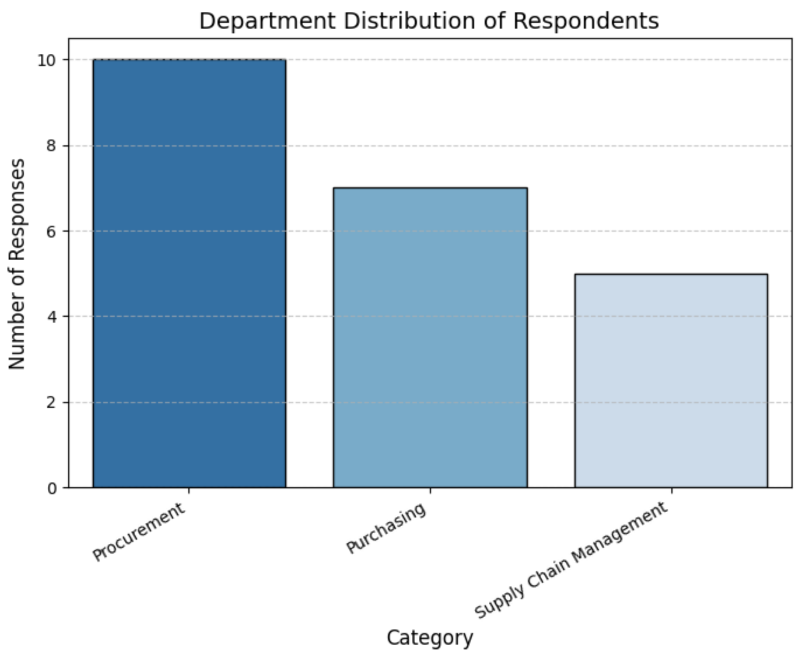


Figure 6. Department Distribution of Respondents.

According to Figure 6, the majority of the respondents recorded from the procurement department and then followed by purchasing and supply chain management department respectively. In the case company, there are two separates department for procurement and purchasing who focuses on long-term and day to day operations respectively.

4.2.2 Identification of Most Important SS Criteria for CSCs

In this study, the supplier selection criteria for CSCs were identified through a review of 43 procurement processes and reports. Through the analysis, 25 available criteria were identified. However, a refined subset of 14 critical criteria was selected to consider in the ML-integrated TOPSIS framework evaluation for supplier selection of circular supply chains ensuring that only the most relevant and often discussed in the case company were considered in the DM process. This aided in reducing complexity and missing values cases of the datasets. This refined subset was determined through a systematic evaluation of the insights from the procurement and sourcing professionals of the case company which were initially collected through the questionnaire.

The refined subset of critical supplier selection criteria was identified through a quantitative threshold approach where participants of the questionnaire survey rated the importance of each criterion on a five-point Likert scale (1= Least important, 5 = Most important). Thereafter, the average rating for each criterion was calculated where a threshold of 3.8 was recognized to identify the most critical SS criteria. The distribution of scores aids in identifying the threshold to distinguish the most critical criteria (which were rated as highly important by the procurement and sourcing professionals).

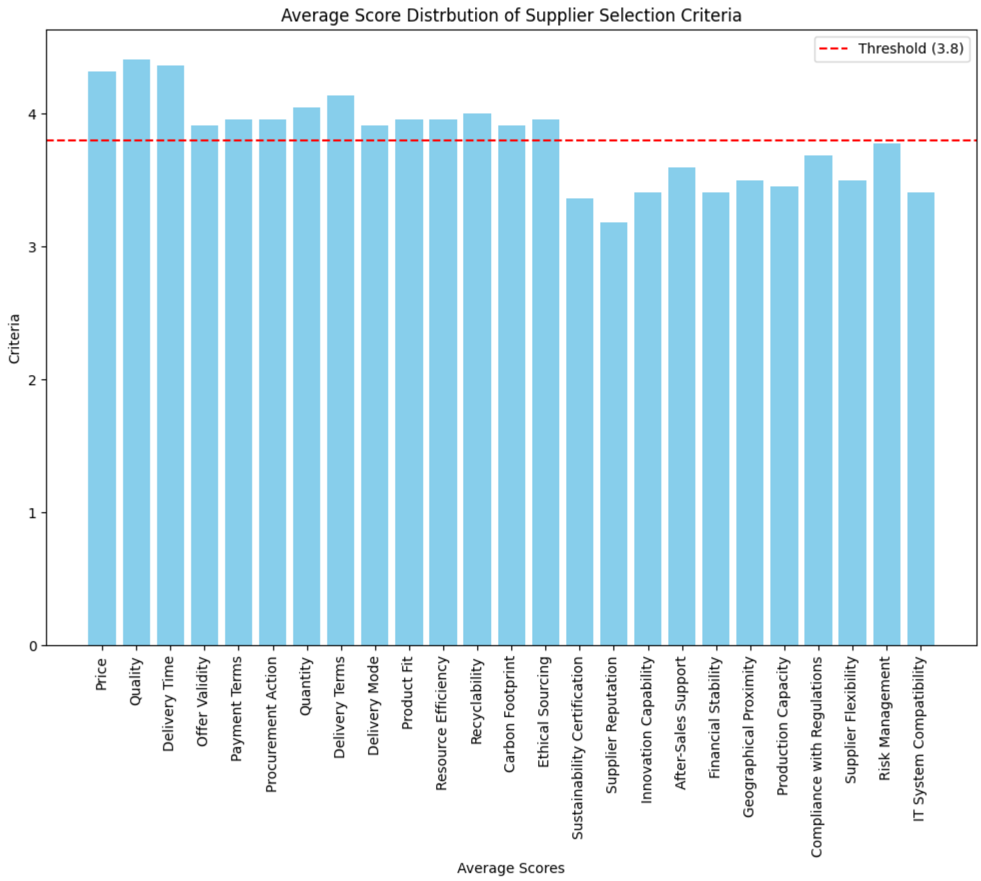


Figure 7. **Average score distribution of SS criteria.**

According to Figure 7, it can be seen that the threshold at 3.8 out of 5 ensures the supplier selection criteria reflects the priorities of survey responders which ultimately filters out the less influential criteria balancing the rigor and inclusivity, capturing the most influential criteria. This approach makes the decision-making framework robust without making it overly complicated. Furthermore, this approach aligns with a practice in MCDM, where an appropriate threshold helps to refine the most influential SS criteria which enhances the decision-making accuracy.

4.2.3 Overview of the most important SS criteria

The final refined subset which includes 14 supplier selection criteria includes both traditional and circular supply chain-specific supplier selection criteria ensuring a selection of a comprehensive supplier evaluation process that aligns with the circular economy principles and sustainability goals. For instance, in the refined subset, the traditional criteria remain essential in the SS as these contribute greatly to the operational efficiency whereas the circular economy-specific criteria such as recyclability, carbon footprint, and resource efficiency are important in ensuring sustainability contribution and practices in the SS process. Thus, this methodology which is driven by both traditional and circular economy-specific supplier selection criteria ensures that the ML-integrated TOPSIS framework is tailored to the CSC needs of the company where informed and sustainability-conscious supplier selection decision making is enabled.

Table 5. Criteria Name and Criteria Definition.

Criteria Name	Criteria Definition	Traditional/Circular Criteria
Price	cost of a unit/ cost per unit	Traditional
Quality	alignment of the predefined quality standards of the supplier products or services	Traditional
Delivery Time	time taken by the supplier to fulfill the order	Traditional
Offer Validity Days	validity period of the supplier offers	Traditional
Payment Terms	conditions of the payments	Traditional
Procurement Action	type of procurement activity	Traditional
Quantity	volume that the supplier is capable of offering	Traditional
Delivery Terms	terms and conditions of the delivery and transportation of the items	Traditional
Delivery Mode	transport method (e.g., air, sea, road)	Traditional

Product Fit	alignment between the product request by the buyer and the products offer by the supplier	Traditional
Resource Efficiency	minimization of wastage, usage of sustainable materials, and efficient use of resources in production	Circular
Recyclability	recyclable capabilities of the products or the materials offered by the suppliers	Circular
Carbon Footprint	supplier's impact on emissions in the production and transportation processes	Circular
Ethical Sourcing	practice of aligning with fair labor laws, sustainability principles, and human rights	Circular

Before proceeding further, it is important to understand if there are any interdependencies among the supplier selection criteria in the refined subset based on the professionals' perspectives. Thus, through a correlation heatmap of the supplier selection criteria, it can be identified whether any notable relationships influence the supplier evaluation and its decision-making processes. However, the findings of this correlation map are only based on the survey which indicates that further validations are required from past performance data of suppliers. In the map, the correlation values varies from –1 to 1, where the positive numbers indicate a positive relationship whereas negative numbers indicate negative relationships. The correlated map was generated using a python code which was executed in Google Colab (see Appendix 3).

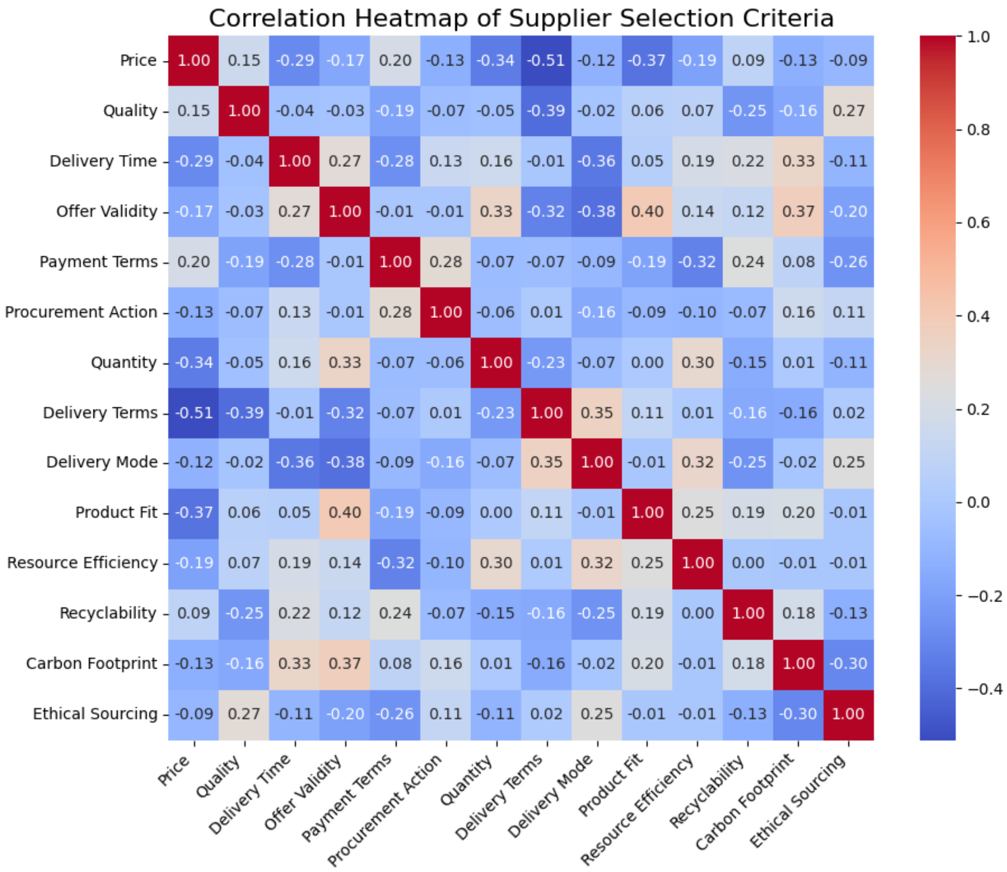


Figure 8. Correlation heatmap of SS criteria.

According to Figure 8, a strong correlation of –0.51 between price and delivery terms can be observed which implies that the cost-reducing procurement decision comes along with the compromise at favorable or flexible delivery options. Furthermore, a negative correlation can be identified between price, carbon footprint, and ethical sourcing which indicates that there are possibilities of compromising these sustainability concerns when focusing on cost efficiency. This showcases some common trade-offs between sustainability considerations and cost reduction in procurement strategies. Moreover, price is negatively correlated with product fit, delivery terms and delivery mode which indicates that the product compatibility, favorable and fast delivery modes are often compromised when focusing on price and cost efficiency.

Additionally, a slightly positive correlation between quantity and ethical sourcing can be observed which might imply that the high-quality suppliers tend to align with ethical sourcing practices. Furthermore, there is only a slight negative correlation between the quality and delivery time which may imply that according to professionals' perspective, the delivery time is sometimes compromised when procuring high-quality products.

The delivery time indicates a negative correlation with price which implies that the supplier who provides a shorter lead time may charge higher prices from the buyer. Moreover, delivery time has a positive relationship with offer validity and carbon footprint which implies that the supplier with timely deliveries tends to offer longer validity periods and minimize the negative environmental impact.

As per professionals' insights, procurement action has a moderate positive correlation with payment terms which implies that payment conditions influence procurement decisions. This aligns with practical supplier selection strategies as organizations tend to favor suppliers who offer flexible and favorable payment structures.

Overall, the professionals' perspective of the case company tends to support common procurement decisions. However, these outcomes need validation with past supplier performance and procurement data to enhance reliability.

4.2.4 Top 10 Supplier Selection Criteria Based on Survey Data

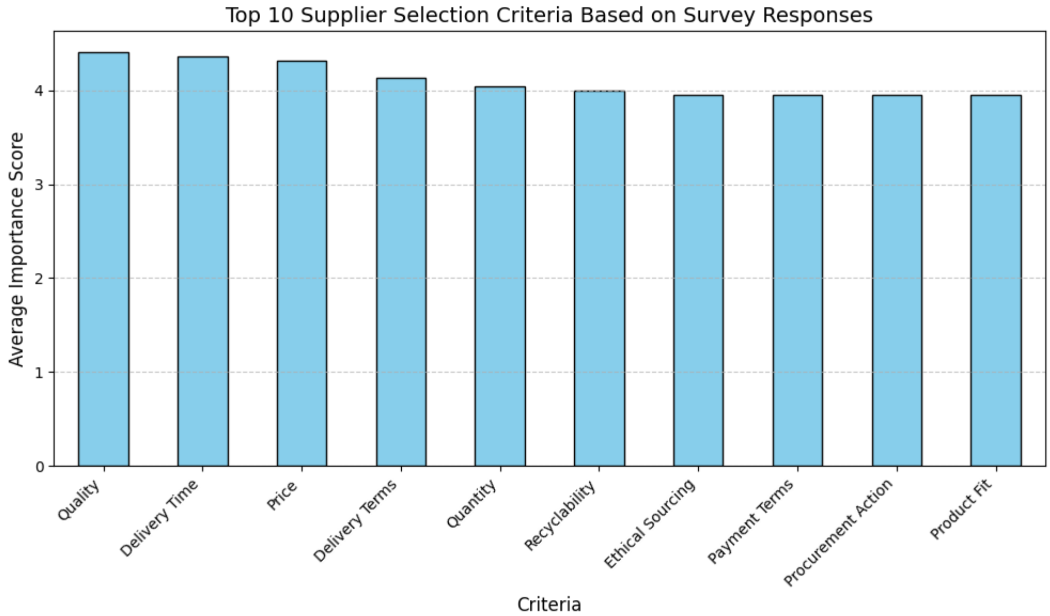


Figure 9. Top 10 Supplier Selection Criteria Based on Survey Data.

Figure 9 was created using a Python code which was executed in Google Colab (see Appendix 4). This Figure demonstrates insights of procurement professionals regarding the prioritization of SS criteria. According to the analysis, quality, delivery time, and price remain the most critical criteria that influence the SS decision-making process. Thus, this suggests that procurement professionals focus on a reliable and cost-effective supply chain first when selecting the suppliers.

Furthermore, delivery terms, quantity, recyclability, and ethical sourcing also score highly which reflects that there is a secondary importance in selecting suppliers based on these criteria. This implies that the professionals of the case company have a mindset of shifting towards a holistic evaluation approach where both operational efficiency and ethical consideration have a significant impact on the DM process.

Moreover, according to the professionals' insights, recyclability, procurement action, and product fit are considered in the top 10 criteria list implying that the case company has a focus on sustainability and alignment with specific procurement needs such as the agreement between the item requested and the item offered.

Overall, this analysis indicates that procurement and sourcing professionals of the case company tend to adopt a balanced approach where both traditional and emerging circular supply chain supplier selection criteria are considered in the supplier evaluation process.

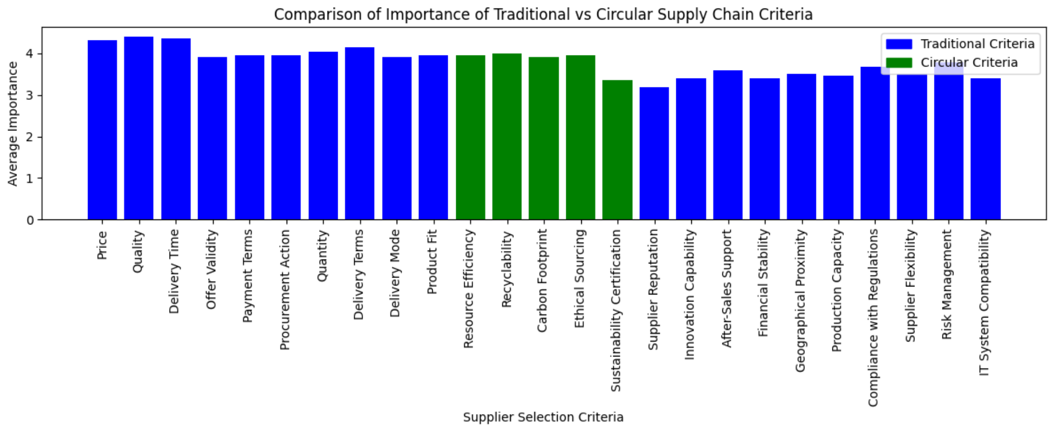


Figure 10. Traditional vs Circular-supply chain specific supplier selection critera.

Figure 10 provides a comparison between the importance of traditional and circular-supply chain supplier selection criteria according to the professionals' insight gathered through the questionnaire. The graph was created using the python code which was executed in the Google Colab (see Appendix 5).

According to the analysis, the conventional SS criteria such as price, quality, delivery time, payment terms, and procurement actions are ranked significantly with an average exceeding 4.0 out of 5 which indicates that according to the professional, these traditional criteria have the most influence on the DM process and plays a central role in supplier evaluation. The cost-effectiveness, operational efficiency, and timely product availability in their supply chain are ensured by incorporating the traditional supplier criteria strategically in the supplier selection.

Furthermore, criteria associated with the CSC like recyclability, efficiency of resource, ethical source, and carbon footprint of the supplier are comparatively higher than the traditional criteria like supplier's solvency and its geographical location. This means that as per the perceived understanding of the experts, the case company is not only catering to the traditional supplier evaluation criteria but is also stressing over some of the CCS supplier selection criteria as well. This might have been caused by the increase in the customer awareness and perception towards green products as well as the external force to provide an environmentally sustainable supply chain.

4.3 Pre-processing of Supplier Performance Data

4.3.1 Data Integration

As described in section 4.1.2, secondary data was collected from two distinct systems of the case company. One system provided the records related to supplier proposals which had the details related to prices, delivery terms, payment conditions product specifications, etc. The other system provided the data related to past supplier evaluation performance records and sustainability assessments.

The initial complete dataset was created by taking the procurement request number as the unique identifier to integrate two different datasets into one. Through this approach, it was able to merge supplier offer records with corresponding evaluation outcomes of the suppliers creating a structured dataset.

After creating the complete dataset, there were 25 evaluation criteria which have different aspects such as quality, operation performance, and sustainability aspects of the suppliers. However, to ensure practical applicability and analytical relevance, a questionnaire was distributed among the procurement and sourcing professionals to determine the most influential supplier selection criteria. Hence, in the final dataset which was ready for the pre-processing, only the 14 most influential supplier selection criteria were retained in the dataset to ensure the balance between practicality and efficiency.

4.3.2 Data Cleaning

The final dataset was then moved to the rigorous preprocessing phase to improve the data quality, suitability, and consistency making it more ready and appropriate for the ML applications. The preprocessing phase included three major steps.

1. Removal of Duplicates

Duplicate records were identified in the dataset, which were related to supplier offers and evaluation records. This step ensured that the ML model would not incorporate duplicate or repeated data points to learn from which might lead to overfitting later.

2. Resolution of Inconsistencies

The categorical criteria such as procurement action types and delivery terms were standardized through feature encoding techniques to make the dataset ready for ML models.

3. Handling of Missing Data

Handling the missing data is a key tasks of the preprocessing phase. The missing data were mainly observed in circular-supply chain-related supply chain criteria and evaluation records as the case company started assessing these criteria recently. This issue was addressed using two ML-oriented approaches.

- Dropping Incomplete Records: If most of the criteria of a row are missing, those records were dropped from the dataset in order to reduce the bias and prevent the ML model being trained on unreliable data.
- Feature Engineering for ML: In some cases, where only one or two missing values are observed in a record, those values were imputed using techniques such as media, mean, or KNN imputation.

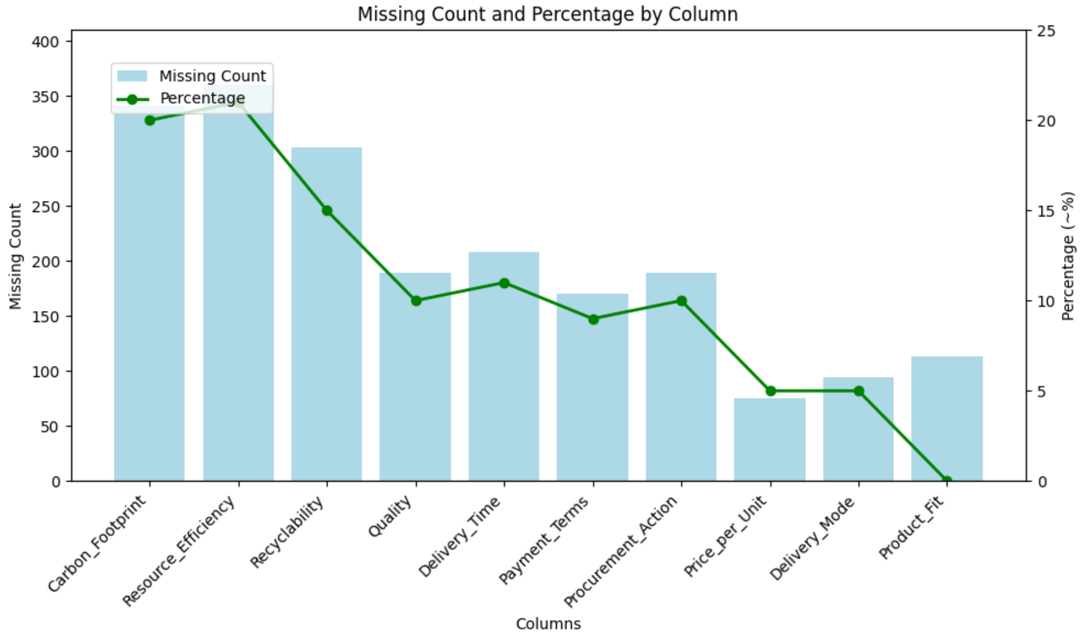


Figure 11. Distribution of missing values in supplier selection criteria.

Figure 11 was created using a Python code which was executed in Google Colab (see Appendix 6). This figure shows the distribution of missing values in each criterion. According to the analysis, resource efficiency has the highest missing data which contributes to 21% of the total dataset followed by the carbon footprint and recyclability as the second and third criteria with the highest missing data rate in the dataset. Conversely, the criteria such as price per unit, delivery mode, and product fit showed the least number of missing data in the final dataset. However, these missing data were addressed through using feature engineering techniques to enhance the reliability and accuracy of the model.

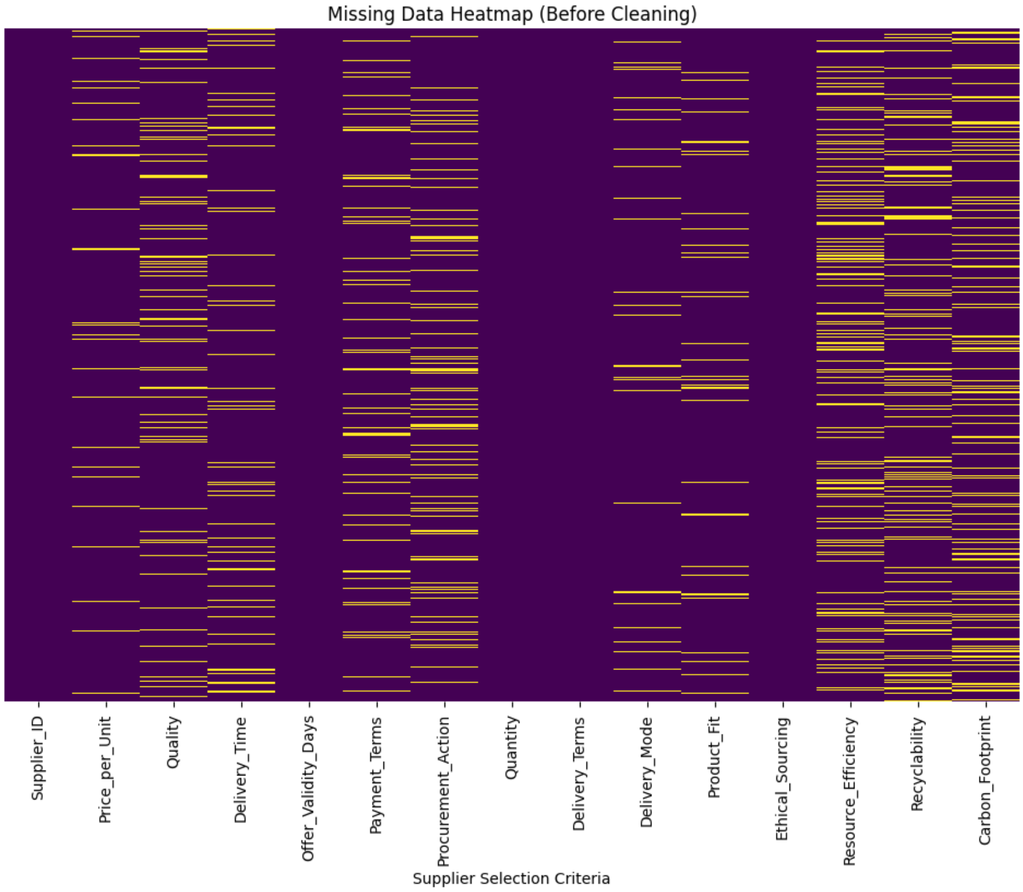


Figure 12. Missing data heatmap (Before cleaning).

Figure 12 which is heatmap illustrates how the missing values are distributed among the dataset. The x axis showcases each criteria or the variable of the dataset. Y axis showcases the place or position of the missing values in the dataset. This figure was generated using a Python code which was executed in the Google Colab (see Appendix 7).

According to the Figure 12, it can be seen that resource efficiency, carbon footprint and recyclability have a significant number of missing values as some past records in the case company's system did not have values for these criteria.

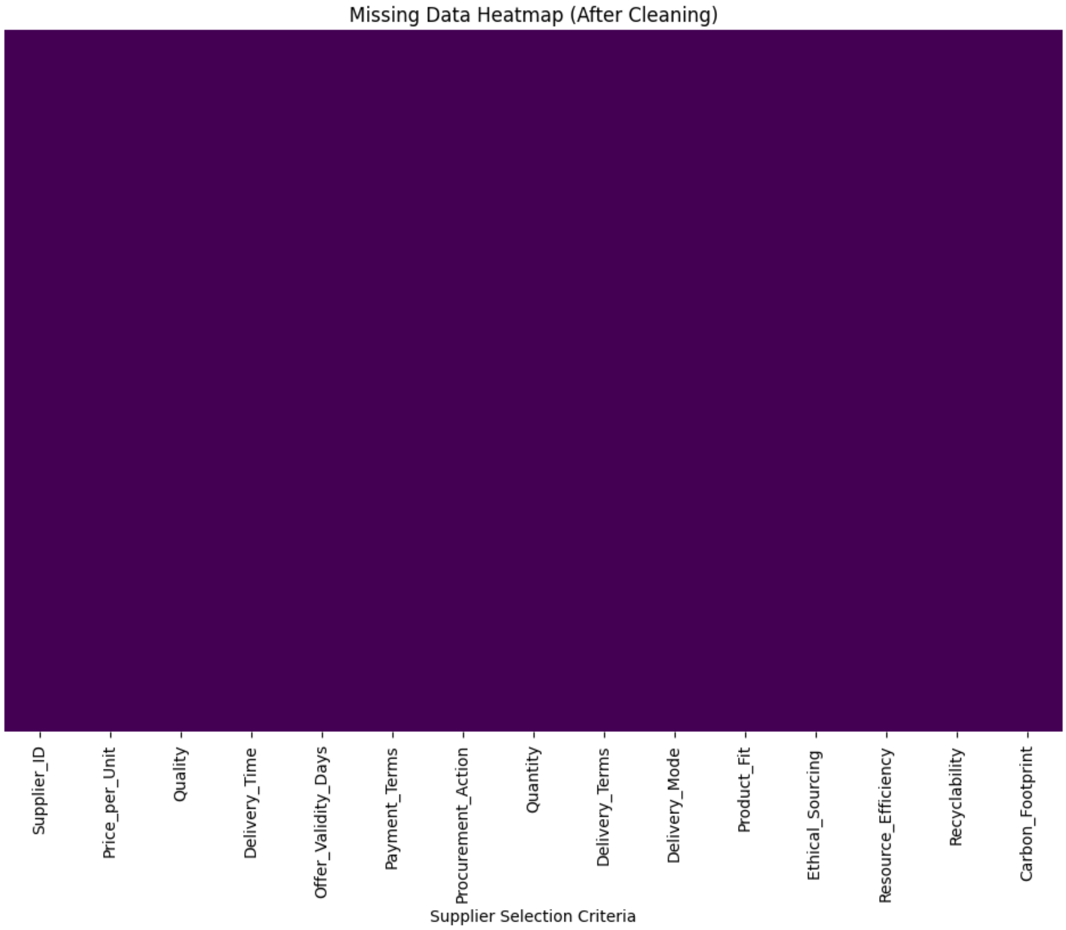


Figure 13. Missing data heatmap (After cleaning).

Figure 13 visualizes the cleaned dataset and any missing values in the dataset. From this figure, it can be seen that the dataset has been fully cleaned addressing all the missing values through dropping records and missing values imputation.

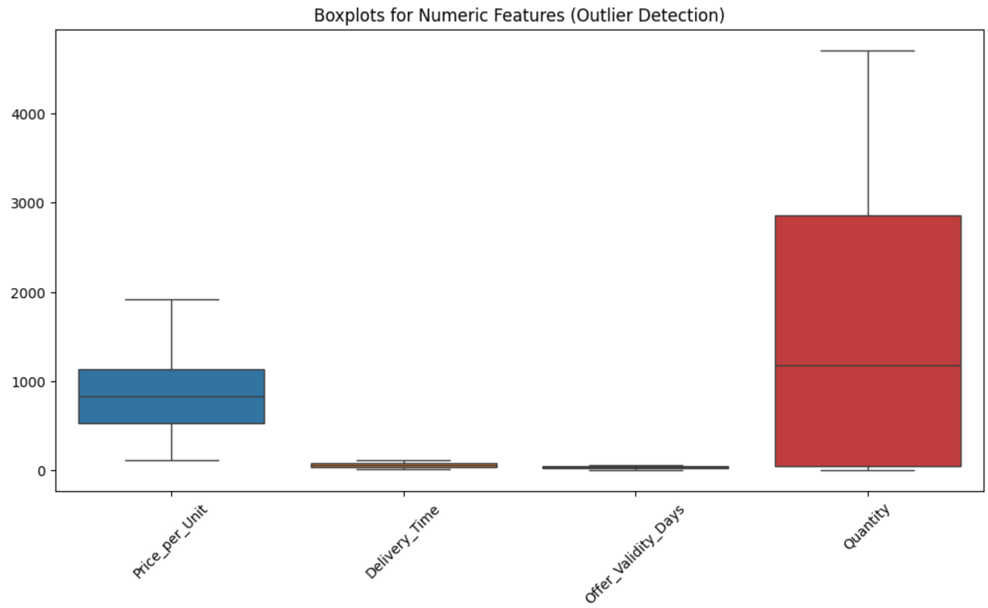


Figure 14. Distribution of the numerical criteria.

Figure 14 which is a boxplot representation showcases the distribution of the numerical attributes or the criteria of the dataset before applying any transformations such as scaling. The figure was created using the python code in Google Colab (see Appendix 8).

According to the Figure 14, the numerical criteria in the dataset include delivery time, offer validity days, quantity, and price per unit. According to the box plot, it can be seen that the distribution of price per unit has a significant variability with a large interquartile range and a few outliers compared to delivery time and offer validity. This is due to the variances in pricing which often occurs due to bulk discounts, and supplier contracts. Delivery time is compactly distributed indicating that many suppliers provide delivery schedules with minimum deviations. Similarly, offer validity days are compactly distributed as well. Conversely, the quantity has a significantly skewed distribution, as the criteria have a large range and multiple large values. This is an expected behavior in the data as the quantity varies and hugely depends on the business needs. However, according to this analysis, the identified outliers should be addressed before feeding the data into the ML model.

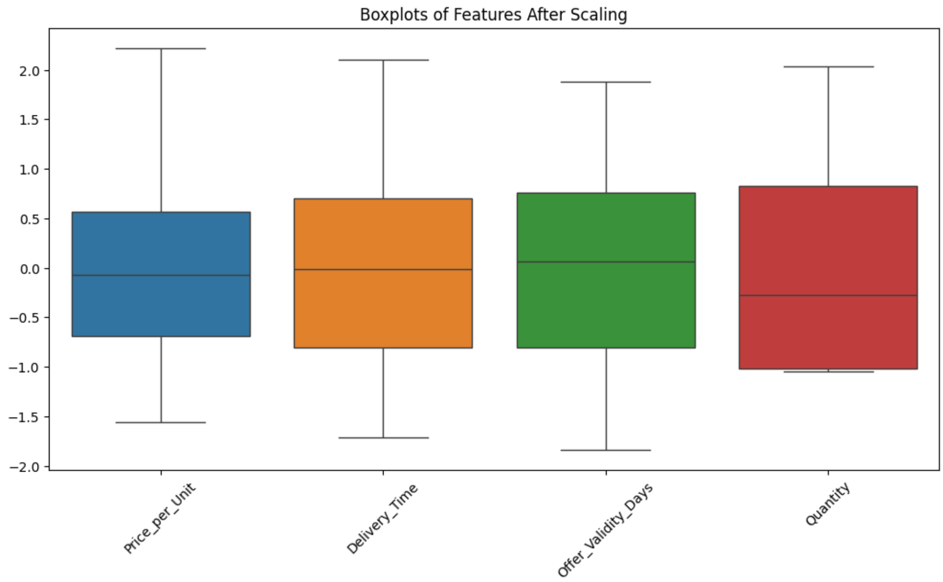


Figure 15. Distribution of the numerical criteria after normalization.

Figure 15 represents the same criteria as in Figure 14, but those attributes are now shown after the z-score normalization. Due to the normalization, all the numerical values are now centered around zero which makes it easier to compare despite their original scales and units. Furthermore, the large magnitudes of price per unit and quantity are now in similar ranges reducing the disproportionate influence on features. After the normalization, the outliers are more balanced without dominating the data, especially the extremely large values. This is an important transformation before feeding the dataset into ML models, as this approach ensures that features are not affecting the results disproportionately due to their original scales.

4.3.3 Final Dataset Overview & Criteria Definition

After the data preprocessing, filtering, and scaling, the dataset had a total of 2478 rows (records) which are from 456 unique suppliers. Thereafter, the past procurement data rows with the same supplier were aggregated into one row, where each row indicates a unique supplier. Hence, in the final data set, there are 456 rows representing 456 suppliers. However, this dataset includes only a subset of suppliers of the case company as not all the suppliers had been evaluated for circular-supply chain-related supplier selection criteria during the previous evaluations.

The final dataset considered for the ML integrated TOPSIS framework deployment consists of 14 key supplier selection criteria, categorized as follows:

Table 6. Criteria Name and Definitions of the Final Dataset.

Criteria Name	Criteria Definition	Numerical/Categorical Values
Price	cost of a unit/ cost per unit	Numerical Values
Quality	alignment of the predefined quality standards of the supplier products or services	Categorical values
Delivery Time	time taken by the supplier to fulfill the order	Numerical Values
Offer Validity Days	validity period of the supplier offers	Numerical Values
Payment Terms	conditions of the payments	Categorical values
Procurement Action	type of procurement activity	Categorical values
Quantity	volume that the supplier is capable of offering	Numerical Values
Delivery Terms	terms and conditions of the delivery and transportation of the items	Categorical values
Delivery Mode	transport method (e.g., air, sea, road)	Categorical values
Product Fit	alignment between the product requested by the buyer and the products offered by the supplier	Categorical values

Resource Efficiency	minimization of wastage, usage of sustainable materials, and efficient use of resources in production	Categorical values
Recyclability	recyclable capabilities of the products or the materials offered by the suppliers	Categorical values
Carbon Footprint	supplier's impact on emissions in the production and transportation processes	Categorical values
Ethical Sourcing	practice of aligning with fair labor laws, sustainability principles, and human rights	Categorical values

4.4 Feature Importance and Criteria Weights

After confirming the final dataset, the next step is to employ the ML-integrated TOPSIS framework for supplier selection. The data-driven approach here is deriving the criteria weights using the feature importance scores obtained from the RF model where the criteria weights are driven by data. Thus, the feature model generates the feature importance values eliminating the subjectivity that often occurs in the traditional TOPSIS method.

4.4.1 Feature Importance Estimation using Random Forest

The RF is a widely used ML technique, and this algorithm creates the decision trees and aggregates the output of these trees to enhance prediction accuracy and interpretability. The main advantage of deploying this algorithm is being able to estimate the importance of each supplier selection criteria based on how those contribute to the model performance. Thus, in this study, RF is used to compute the feature importance criteria by training the provided dataset which has supplier performance data.

In the process, the following steps were followed:

1. Data preparation – This step was already completed in the preprocessing stage where missing values were handled. In this stage, a Python code was added to encode the categorical variables and also to standardize the unit if necessary.

2. Training the RF model – In the model training, the dataset was separated into two sets: training and testing sets. Thereafter, the RF model was trained to predict the supplier performance score.

3. Feature Importance Calculation: After the model training, the importance of each criterion was calculated using the mean decrease in impurity indicating the features with higher importance contribute heavily to the supplier selection decision-making process.

Above mentioned steps from 1 to 3 were followed in the python code and executed in the Google Colab. Figure 16 showcases the python code which was executed following step 1 to step 3.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import LabelEncoder, MinMaxScaler
from sklearn.feature_selection import mutual_info_regression

# Load cleaned dataset
file_path = "/content/final_procurement_data.csv"
df = pd.read_csv(file_path)

# Keep Supplier_ID for ranking but don't use it in ML
supplier_ids = df["Supplier_ID"]

# Separate numerical and categorical columns
numeric_cols = df.select_dtypes(include=[np.number]).columns.tolist()
categorical_cols = df.select_dtypes(include=["object"]).columns.tolist()

# Encode categorical variables using LabelEncoder
label_encoders = {}
for col in categorical_cols:
    le = LabelEncoder()
    df[col] = le.fit_transform(df[col])
    label_encoders[col] = le

# Define X (features) and y (target)
X = df.drop(columns=["Supplier_Performance_Score", "Supplier_ID"])  # Exclude target variable
y = df["Supplier_Performance_Score"]

# Train Random Forest for feature importance analysis
rf = RandomForestRegressor(n_estimators=500, max_depth=10, random_state=42)
rf.fit(X, y)
```

Figure 16. Python Code for Data Pre-processing and Feature Importance Analysis

Figure 17 showcases the Google Colab output of the feature importance of each criterion according to RF model training.

	Feature	Importance
0	Price_per_Unit	0.220421
1	Quality	0.199384
2	Delivery_Time	0.179943
3	Resource_Efficiency	0.101827
4	Quantity	0.084639
5	Ethical_Sourcing	0.070492
6	Delivery_Mode	0.062763
8	Recyclability	0.048276
7	Carbon_Footprint	0.039537
9	Payment_Terms	0.030135
10	Offer_Validity_Days	0.020684
11	Procurement_Action	0.011673
12	Delivery_Terms	0.010951
13	Product_Fit	0.010221

Figure 17. Feature importance output.

According to the figure, it can be identified that quality, delivery time, and price per unit contribute significantly to the supplier performance score or the supplier selection whereas the circular supply chain-related criteria, resource efficiency, ethical sourcing and recyclability contribute to the supplier performance score moderately. The other procurement administrative-related criteria such as procurement action, delivery terms, and product fit have a lower influence on the supplier performance score when compared to other criteria.

4.4.2 Deriving Criteria Weights for TOPSIS

Once the feature importances are calculated from the RF model, these should be normalized to derive the criteria weights that would be used in the TOPSIS calculations. The normalization was done using the python code and after executing in the Google Colab. Figure 18 illustrates the python code executed to derive criteria weights using feature importance analysis.

```
# Train Random Forest for feature importance analysis
rf = RandomForestRegressor(n_estimators=500, max_depth=10, random_state=42)
rf.fit(X, y)

# Get feature importance
feature_importance = pd.DataFrame({"Feature": X.columns, "Importance": rf.feature_importances_})
feature_importance = feature_importance.sort_values(by="Importance", ascending=False)

# Normalize Criteria Weights
feature_importance["Criteria_Weight"] = feature_importance["Importance"] / feature_importance["Importance"].sum()

# Display Feature Importance and Criteria Weights
print("Feature Importance Table:")
print(feature_importance)
print("\nCriteria Weights Table:")
print(feature_importance[["Feature", "Criteria_Weight"]])

# Plot Feature Importance
plt.figure(figsize=(12, 6))
sns.barplot(x="Importance", y="Feature", data=feature_importance, palette="viridis")
plt.title("Feature Importance (Random Forest)")
plt.xlabel("Importance Score")
plt.ylabel("Feature")
plt.show()

# Additional Feature Selection using Mutual Information
mi_scores = mutual_info_regression(X, y)
mi_feature_importance = pd.DataFrame({"Feature": X.columns, "Mutual_Info_Score": mi_scores})
mi_feature_importance = mi_feature_importance.sort_values(by="Mutual_Info_Score", ascending=False)
print("\nMutual Information Scores:")
print(mi_feature_importance)
```

Figure 18. Feature Importance and Criteria Weight Generation.

Figure 19 illustrates the final outputs of the criteria weights which would be subsequently used in the TOPSIS analysis to rank suppliers.

	Feature	Criteria_Weight
0	Price_per_Unit	0.202046
1	Quality	0.182762
2	Delivery_Time	0.164942
3	Resource_Efficiency	0.093338
4	Quantity	0.077583
5	Ethical_Sourcing	0.064615
6	Delivery_Mode	0.057531
7	Carbon_Footprint	0.036241
8	Recyclability	0.044252
9	Payment_Terms	0.027623
10	Offer_Validity_Days	0.018960
11	Procurement_Action	0.010700
12	Delivery_Terms	0.010038
13	Product_Fit	0.009369

Figure 19. Critera weights output.

This approach provides data-driven critical weights that are based on RF models instead of subjective opinions which leads to biases or reduction in data accuracy. However, from this approach, the criteria weights are purely based on data insights.

4.5 Final TOPSIS scores and supplier rankings

Once the criteria weights were generated, the steps of the TOPSIS method to prioritize suppliers were executed. The following steps of TOPSIS analysis were comprehensively described in chapter 3.

- Step 1 – Normalization of the Decision Matrix: The criteria are normalized and scaled to make all the criteria comparable.
- Step 2 – Multiply by Criteria Weights: In this step, the normalized decision matrix formed in Step 1 is multiplied by the criteria weights obtained from the feature importance of the RF model.
- Step 3 – Determination of Ideal Best and Ideal Worst

- Step 4 – Compute Distance from Ideal Best and Ideal Worst
- Step 5 – TOPSIS Score Calculation
- Step 6 – Supplier Ranking

After following all the steps above, the model generates the TOPSIS score and supplier ranking as the output. The above steps were followed in the python code and executed in the Google Colab. Figure 20 illustrates the python code executed which followed step 1 to step 6.

```
def topsis(decision_matrix, weights, benefit_criteria):
    # Step 1: Normalize the decision matrix
    norm_matrix = decision_matrix / np.sqrt((decision_matrix ** 2).sum(axis=0))

    # Step 2: Multiply by Criteria Weights
    weighted_matrix = norm_matrix * weights

    # Step 3: Determine Ideal Best (A+) and Ideal Worst (A-)
    # Apply benefit_criteria element-wise
    ideal_best = np.where(benefit_criteria, np.max(weighted_matrix, axis=0), np.min(weighted_matrix, axis=0))
    ideal_worst = np.where(benefit_criteria, np.min(weighted_matrix, axis=0), np.max(weighted_matrix, axis=0))

    # Step 4: Compute distances from Ideal Best and Ideal Worst
    dist_best = np.sqrt(((weighted_matrix - ideal_best) ** 2).sum(axis=1))
    dist_worst = np.sqrt(((weighted_matrix - ideal_worst) ** 2).sum(axis=1))

    # Step 5: Compute TOPSIS Score
    topsis_scores = dist_worst / (dist_best + dist_worst)

    return topsis_scores

# Prepare Decision Matrix (without Supplier_ID)
decision_matrix = df[X.columns]

# Normalize using Min-Max Scaling
scaler = MinMaxScaler()
decision_matrix = scaler.fit_transform(decision_matrix)

# Get weights from Random Forest results
weights = feature_importance["Criteria_Weight"].values

# Define Benefit Criteria (1 for max benefit, 0 for cost-based)
benefit_criteria = np.ones(len(weights))  # Assuming all criteria are beneficial
```

Figure 20. TOPSIS Calculation and Supplier Rankings.

Figure 21 illustrates the Google Colab output of supplier rankings with corresponding Supplier_ID and TOPSIS scores.

```
TOPSIS Supplier Rankings:
     Supplier_ID  TOPSIS_Score  Ranking
318         1714      0.831715      1.0
0           1889      0.768365      2.0
227         1383      0.743887      3.0
265         1272      0.740061      4.0
331         1419      0.739018      5.0
..           ...           ...      ...
125         1312      0.163543    452.0
313         1307      0.154308    453.0
115         1369      0.153160    454.0
304         1643      0.132105    455.0
349         1109      0.106912    456.0
```

Figure 21. ML-integrated TOPSIS supplier rankings.

This output indicates which suppliers are performing well or have performed well based on the feature importance criteria obtained from the RF model. According to the analysis, it can be seen that Supplier_ID 1714 ranks as the highest-performing supplier.

4.6 Comparison between Traditional TOPSIS Method vs ML integrated TOPSIS Method

In this section, the analysis and results obtained in the previous parts of the chapter 4 will be compared with the traditional TOPSIS method results to identify the difference between the models. Before going forward with the comparison the suppliers ranking through the traditional TOPSIS method should be calculated. Table 7 illustrates the traditional rankings of the suppliers.

Table 7. Traditional TOPSIS method's supplier rankings.

Supplier_ID	Traditional_Rank
1965	1
1027	2

1931	3
1714	4
1971	5
1889	6
1412	7
1453	8
1055	9
1757	10

The comparison between the supplier rankings of ML integrated TOPSIS and the conventional TOPSIS can be shown as in Figure 22.

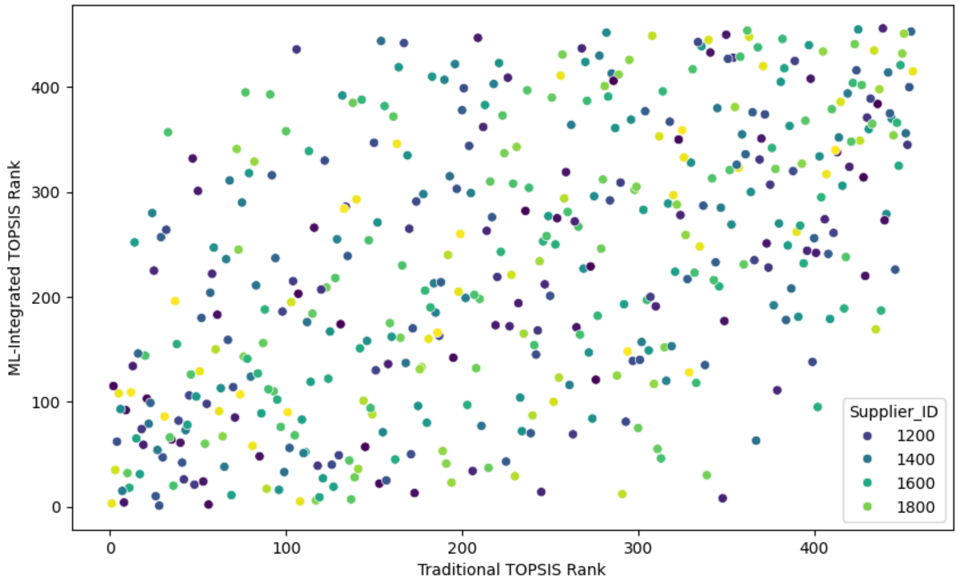


Figure 22. Supplier rankings of ML-integrated TOPSIS and the traditional TOPSIS.

In above scatter plot, the Supplier_IDs are seperated into four ranging categories and these categories are represented in four different colours. The x and y axis showcase the supplier rankings of traditional TOPSIS and ML-integrated TOPSIS respectively. The graph was created using a Python code which was executed in Google Colab (see Appendix 9).

According to Figure 22, it can be seen that the points are scattered indicating the differences in the ranking outcome of the two approaches. If the traditional ranking and ML ranking were identical, all the points would align along a diagonal line. From the analysis, it can be identified that some suppliers ranked high in the traditional TOPSIS approach but lower in ML-integrated TOPSIS and vice versa. However, the higher variability and widespread of scatter plot might indicate a new ranking pattern identified by ML-TOPSIS as it is driven by data-insight-centered adjustments. Moreover, the ML model is capable of assigning feature importance through advanced data analytics reducing human judgments in the decision making which might be the potential reason for having comparatively different rankings in both approaches. Overall, there is not any perfect linear relationship in the supplier ranking which indicates that ML mode has introduced a less biased and more balanced supplier evaluation framework consisting of both traditional and circular supply chain criteria considerations.

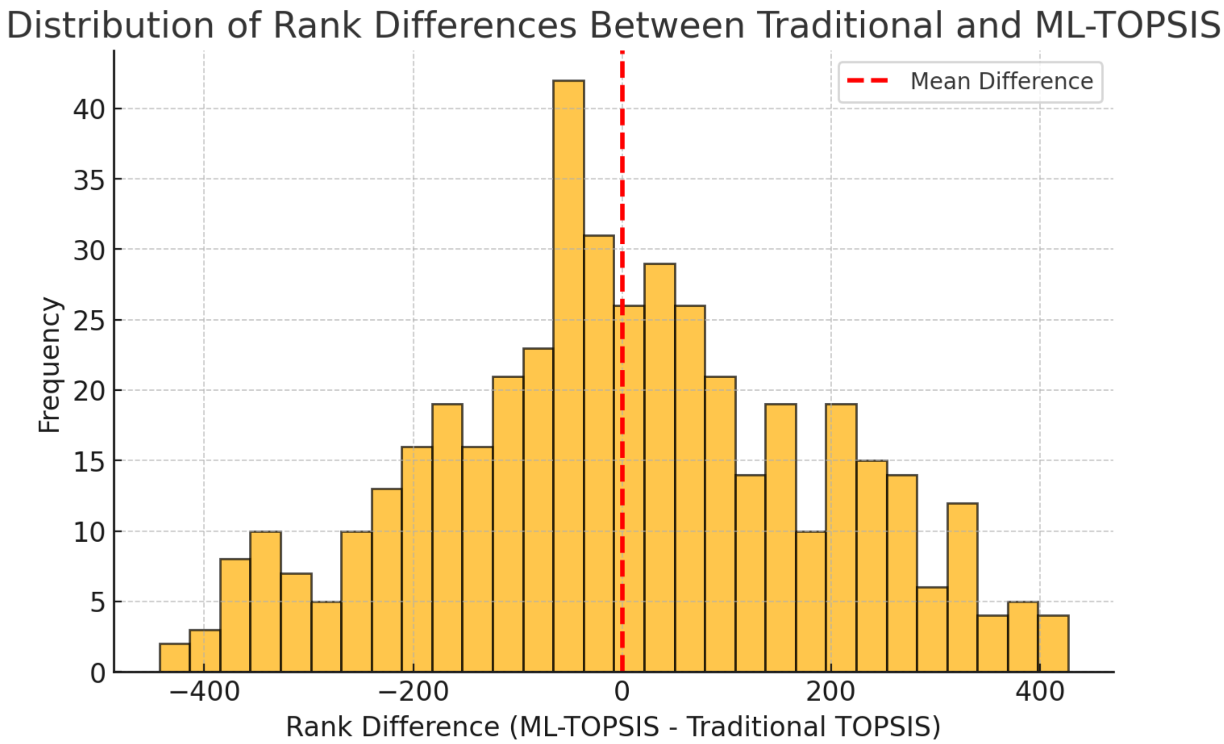


Figure 23. Distribution of rank differences between traditional and ML-TOPSIS.

Figure 23 illustrates how the supplier rankings have shifted between traditional and ML-TOPSIS. The histogram was created by executing a Python code in Google Colab (see

Appendix 10). From the histogram, it can be seen that some suppliers have been significantly moved up and down, while some have remained in the same rankings. Even though the rank differences are centered around zero it can be seen that there is a wide spread of rankings which indicates that ML-integrated TOPSIS has rearranged the suppliers quite differently than the traditional TOPSIS. This dispersed distribution indicates that the ML-integrated TOPSIS has weighted criteria differently than the traditional TOPSIS as the ML model learned the data and assigned the weightage criteria accordingly.

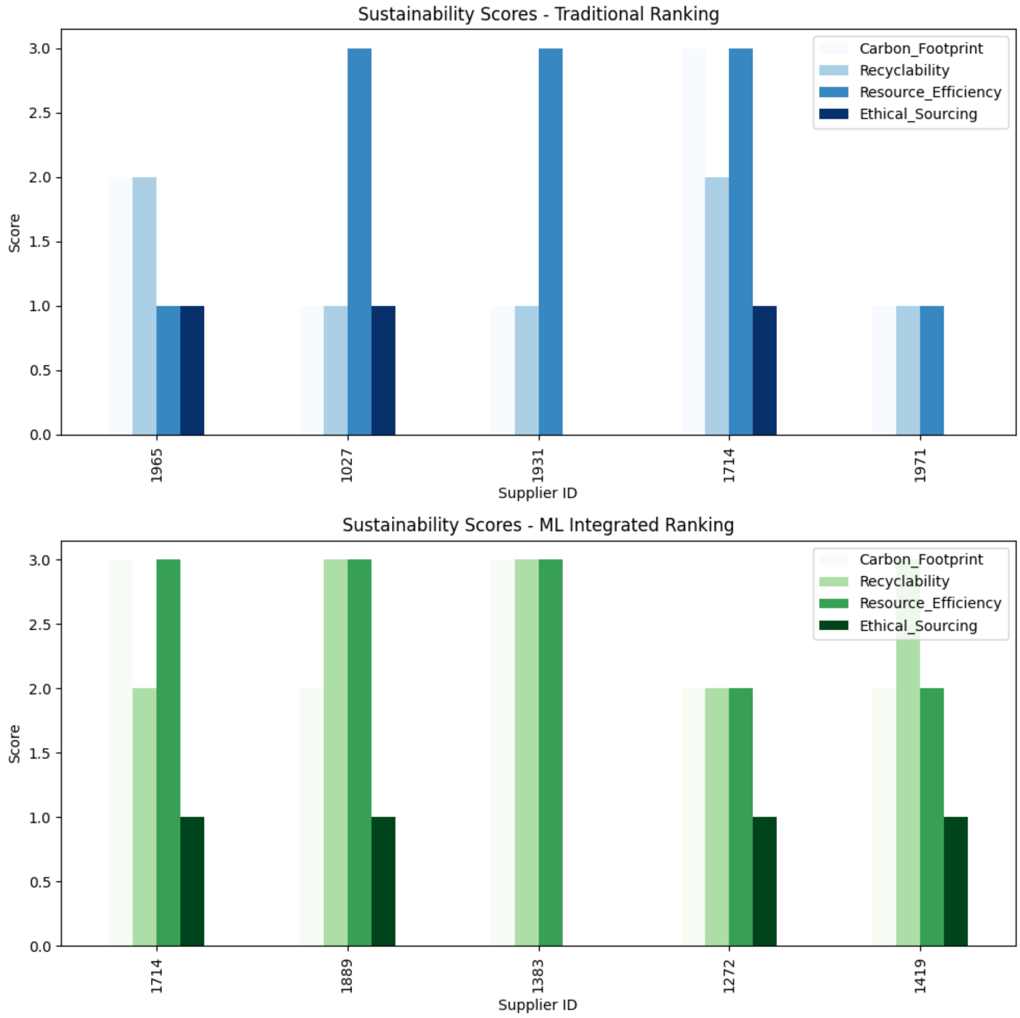


Figure 24. Circular-supply chain-specific criteria performance of the top 5 supplier rankings of traditional and ML-TOPSIS approaches.

Figure 24 illustrates the circular-supply chain-specific criteria performance of the top 5 supplier rankings of traditional and ML-TOPSIS approaches. From the comparison of graphs, it can be seen that the ML-TOPSIS approach is more likely to have given a balance feature importance distribution among circular-supply chain-specific criteria by analyzing the data whereas the traditional TOPSIS approach is more likely to have prioritized resource efficiency among other circular-specific criteria when ranking the suppliers. Overall, it can be identified that the ML-TOPSIS approach has provided a more holistic approach to ranking suppliers in circular supply chains by prioritizing circular-specific criteria covering both environmental and ethical aspects along with the other important operational efficiency criteria.

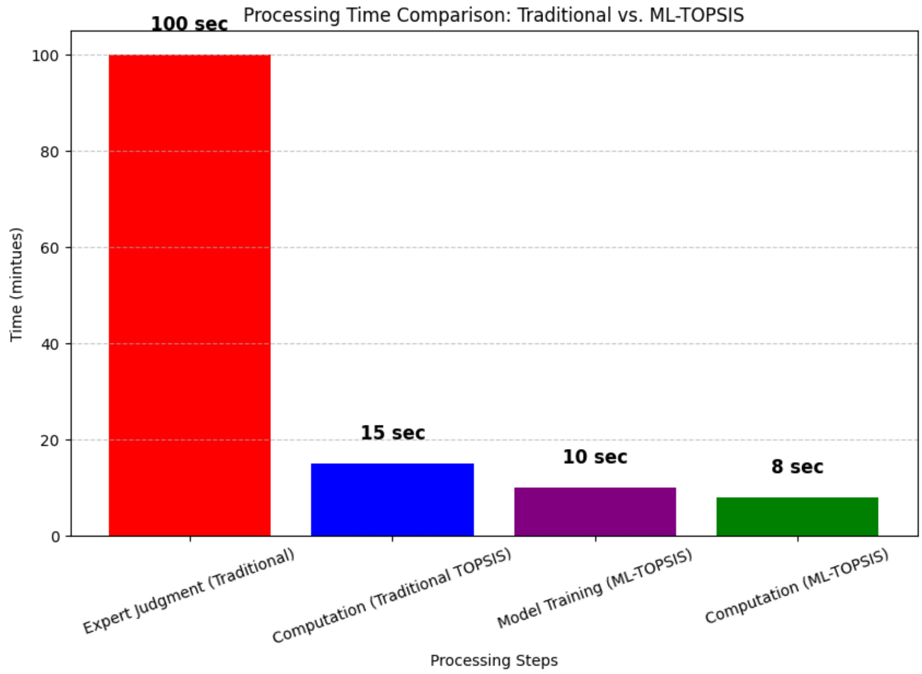


Figure 25. Processing time of ML-integrated TOPSIS and traditional TOPSIS method.

Figure 25 illustrates the processing time of both ML-integrated TOPSIS and traditional TOPSIS. The time taken by traditional TOPSIS for expert judgment is significantly higher than the other processing times. The time taken by expert judgment or manual human input proportionately increases with the number of criteria involved in the supplier evaluation. In this study, the expert judgment was collected through the survey which took a considerable amount of time assuming that all information on the expert judgments was received at the same time. In the ML-TOPSIS model, the time taken for both model training and computation is significantly low when compared with the traditional TOPSIS method. However, the model training time would proportionately increase with the size of the dataset. Yet, since expert judgement takes considerable time for deciding criteria weights, it is clear that processing time of ML integrated TOPSIS model is comparatively low when compared to traditional TOPSIS.

4.7 Summary of the Results Analysis

According to the analysis of the questionnaire data, it was identified that the professionals of the case company have a mindset to shift towards circular supply chain management. According to survey data, quality, delivery time, and price are considered the most critical factors influencing procurement decisions, whereas ethical sourcing, recyclability, and quantity are considered secondary priorities in the supplier selection decision-making process. These prioritizations indicate that the case company is gradually integrating sustainability considerations into its procurement strategy.

In the data preprocessing phase of the ML-integrated TOPSIS framework evaluation, many missing values were observed, especially in the circular supply chain-related criteria. However, these missing values were handled through data imputation, and duplicates of the dataset were removed. Thereafter, the numerical features were standardized using z-score normal distribution to ensure scalability.

The feature importance analysis was performed through training the dataset incorporating RF model. The output of the feature importance analysis showcased that quality, delivery time, and price per unit have the highest influence on the supplier performance score, whereas procurement-related administrative criteria have the lowest impact on the supplier performance score. The ML-integrated TOPSIS framework ranked suppliers after deriving criteria weights based on RF-derived feature importance values.

The comparison between the traditional TOPSIS and the ML-integrated TOPSIS revealed significant differences in the supplier rankings, ranking patterns, and processing efficiencies. According to the comparison of supplier rankings, it was observed that the rankings of the two methods do not align perfectly. This demonstrated that ML-TOPSIS has adjusted the rankings based on the historical data and data-driven insights instead of predefined criteria weights from experts.

Furthermore, the distribution of the ranking differences indicated that some rankings had notable shifts, whereas some rankings maintained the same positions. This indicated that the ML-integrated TOPSIS assigns rankings differently, reducing biases in the decision-making process. Moreover, it was identified that ML-TOPSIS has balanced feature importance across multiple circular supply chain criteria compared to the traditional TOPSIS method.

The processing time analysis of both ML-integrated TOPSIS and traditional TOPSIS highlighted that the traditional TOPSIS consumed a significant amount of time due to the expert judgment process, whereas the ML-TOPSIS method demonstrated higher efficiency in computation and decision-making. Even though the ML model training time increases with the size of the dataset, it remains significantly low compared to the time required for expert judgment in traditional TOPSIS.

Overall, from the outcomes of the analysis, it was identified that the ML-integrated TOPSIS approach provides a less biased, computationally efficient, and more dynamic supplier evaluation framework for the sustainable SS in the CSCM.

5 Discussion

5.1 Introduction

The emerging technologies have been transforming different industries to improve operational efficiency and transparency. Similarly, the supply chains of the organization have been adopting different technologies to improve their processes. Among the advanced technologies, AI and ML are important technologies that adopt advanced data analysis techniques to improve SC operations (Ivanov et al., 2019). Thus, the integration of ML into SS improved efficiency and resilience in the SCs (Govindan et al., 2020). Moreover, the world is now transforming various industries to become sustainable and environmentally responsible while growing their business. Hence, circular supply chains have emerged as a growing and relevant research stream and business strategy for organizations to implement CE concepts in SCM.

Nevertheless, it deserves mention that the supplier selection criterion also significantly contributes to integrating CE principles into the SC. General, actual model that concerns the supplier selection and decision making has a positive influence on the SC effectiveness, use and sustainability. This apparently creates the need of developing new Supplier Selection and Decision-Making Frameworks that will accommodate the circular economy concepts in today’s complex supply chains. In selecting suppliers of the circular supply chains, evaluation of supplier’s competencies is crucial for determining the recycling, remanufacturing, and material recovery tasks. However, the traditional supplier selection approaches often fail to address multidimensional requirements. Hence, this study developed and explored the integration of ML into the TOPSIS framework for the supplier selection of circular supply chains by providing a robust, scalable, data-driven supplier selection framework for the decision makers of the case company to enhance sustainability and efficiency in circular supply chains by making a meaningful contribution to the ongoing transformation of supply chains in the era of sustainability.

5.2 Interpretation of Key Findings

The results of this research study provide meaningful insights derived from the ML integrated TOPSIS framework deployment for select suppliers for circular supply chains. Furthermore, the results and the comprehensive analysis have been able to address three main research questions of this study through comparing the ML-integrated TOPSIS model results with Traditional TOPSIS model results.

A key finding of this study is that the ML techniques optimize supplier evaluation by automating the generation of feature importance and criteria weights for TOPSIS. Furthermore, the ML specifically the RF technique identifies the patterns by learning the data which might have been left overlooked by the traditional approaches. From further analysis, it was identified that the traditional TOPSIS and ML-integrated TOPSIS assigned criteria weights differently where in the traditional approach the weights were assigned based on expert judgments derived from the survey whereas the ML-integrated TOPSIS assigned the weights by training and learning from the dataset. Moreover, it was identified that ML integrated TOPSIS method reduces the human bias of the dataset and criteria weights which ensures more accurate and objective supplier ranking. This aligns with the previous studies from many researchers such as Li et al. (2021) where they found that the ML-based decision models aid in reducing human bias in the criteria weighting process. Hence, these findings address research question 1 of this research study demonstrating how ML techniques optimize supplier evaluation and improve the decision-making circular supply chains.

Furthermore, previous research studies demonstrated that the accuracy and efficiency of the supplier ranking processes improved when ML is integrated into MCDM techniques (Kumar et al., 2022). This research further strengthens the findings of previous research studies, demonstrating that these ML-integrated MCDM techniques not only improve the supplier selection of traditional supply chains but also the supplier selection of circular supply chains. Furthermore, the analysis between the traditional TOPSIS and

ML-integrated TOPSIS showcased there are some significant position shifts in the supplier rankings. This is due to the ability of ML integrated TOPSIS to identify the correlations and latent patterns which often get neglected in the traditional TOPSIS. The findings of the study conducted by Zhang et al. (2020), which demonstrated the ability of ML-integrated MCDM methods to identify hidden correlations and capture latent patterns align with the findings of this study.

Another key finding of the study is the difference between the processing time of traditional TOPSIS and ML-integrated TOPSIS. The traditional TOPSIS required expert judgement which took time to gather the relevant data. This resulted in subjectivity and prolonged DM. Conversely, the ML-TOPSIS automated the criteria weighting, generating TOPSIS score and ranking suppliers by learning from the historical where continuous human input was not required. As a result, this approach took considerably less time to process the data and generate supplier ranking lists. However, this processing time might be proportionately increased with the size of the data. Since the dataset of this study was medium-sized, the processing time was less than the time it would have been if it was a large dataset. Yet, the whole processing time of the traditional TOPSIS would require significantly more time as this method requires human input. This finding addresses research question 2 of this research study which demonstrated that the ML-TOPSIS method enhances the efficiency of the supplier evaluation process reducing expert manual efforts. Similar findings were found in a study conducted by Gunasekaran et al. (2021), which demonstrated the ability of ML-integrated supplier selection processes to eliminate manual data processing by streamlining the decision-making process further.

Another key finding of this study is that the criteria weights have been differently assigned in ML integrated TOPSIS model when compared with expert-defined traditional TOPSIS model criteria weights. From further analysis, it was able to identify how differently the circular-supply chain-related criteria have contributed to the top 5 supplier rankings selection of both TOPSIS methods. In the ML-integrated TOPSIS method, the model prioritized many circular-supply chain-related supplier selection criteria which went underweighted in the traditional TOPSIS method. This addresses the third research question of the research study indicating that the ML-integrated TOPSIS framework improves supplier selection in circular supply chains by integrating data-driven techniques into the decision-making framework and ensuring the supplier selection aligns with circular economy principles which emphasizes sustainability and resource efficiency. The study conducted by Awasthi et al. (2018) which demonstrated that the sustainability-related criteria often remain undervalued in traditional MCDM due to expert biases aligns with the finding of this study.

Overall, this study illustrated the capabilities of ML integrated TOPSIS model in effectively ranking suppliers in the circular supply chains. These findings of this research study addressed the results of the study conducted by Kusi-Sarpong et al. (2019) which illustrated that importance of incorporating real-time adaptability in supplier evaluation frameworks in order to remain effective in sustainable and circular supply chains.

5.3 Theoretical Implications

The results of this study make a valuable contribution to the theory and practice of supply chain management. Practically, this study addresses some research gaps highlighted in the literature. In the literature, many studies have extensively explored MCDM techniques such as AHP and TOPSIS for SS. Furthermore, some have explored how Fuzzy Logic and ML can be integrated into the MCDM techniques for supplier selection (Chai & Ngai, 2020). Yet, the integration of ML into TOPSIS for supplier selection remained underexplored in the literature. Thus, the research study filled this research gap by combining the capabilities of ML in analyzing large and multi-dimensional datasets with the traditional TOPSIS technique.

Furthermore, this study demonstrated further that the traditional MCDM methods can be enhanced significantly with AI and ML integration improving the efficiency. Moreover, this research contributed to the literature on sustainability and CSCs which are growing academic interests in recent years. Therefore, this interdisciplinary approach provided a refined framework to improve the holistic understanding of assimilating advanced and emerging technologies for addressing the complexities involved in SS under the context of CSCs.

5.4 Practical Implications

In terms of practical implications, this study demonstrated how companies can adopt the circular economy principles leveraging ML-integrated TOPSIS framework in their supply chain processes. Furthermore, this study showcased the automation of weight assignment reducing the manual expert inputs reliance on the supplier selection process. Moreover, another key contribution of this study on practical implication is showcasing the ML-based TOPSIS approach allows the incorporation of new supplier performance over time ensuring the supplier evaluations remain relevant. This is also following the findings of the research work conducted by Mangla et al. (2022) that showed how crucial it is to employ a dynamic and scalable supplier selection process to enhance the sustainability issues of SCM in real-life applications. Therefore, it can be verified that the research work makes a considerable practical and theoretical contribution to the SCM field under the domain of SS.

5.5 Limitation of the Study

Despite the study's contribution in several dimensions, the study has several limitations which should be acknowledged. One of the main limitations of this study is the data limitations. This study relied on a dataset of 456 suppliers with 14 supplier selection criteria. While this provides a robust foundation for the ML-integrated TOPSIS framework deployment, the accuracy of the model would have been improved greatly with a larger dataset. In the study, the efforts were given to expand the dataset size further but due to the complexity of data, missing values and duplicate records the number of rows in the dataset had to be removed. Furthermore, in this study, the dataset was not filtered based on the supplier category and supplier location (region) which might have resulted in variability in supplier performance scores. In this study, the dataset was not filtered as the filteration could have reduced the size of the dataset furthermore. Thus, the model interpretability could have been greatly improved with such filtration which might results in more simplified and explainable insights for the decision-making process. In a study conducted by Sarkis et al. (2021), it was stated that the model interpretability can improve the decision-making process significantly by filtering data based on supplier characteristics.

Another limitation of the study is data quality and bias which reduced the accuracy of the ML-integrated TOPSIS results to some extent. The reason behind this was in the initial dataset there were a considerable amount of missing values of circular-supply chain-related criteria as the case company started evaluating supplier performance based on these criteria recently. Even though this issue was addressed in the pre-processing stage, there could have been still biases in the dataset to some extent. However, the reliability of the outcomes of the ML-integrated TOPSIS method could have been greatly with a higher-quality dataset.

These limitations can be overcome by incorporating multiple case studies and also conducting the case studies for more mature environment where shift towards circular supply chain management has been achieved for a considerable extent. This allows to have reduced number of missing values for certain criteria, a larger dataset with filtrations based on dynamics such as product categories and also less complexities in merging data. Hence, addressing the above limitations would improve the ML-integrated TOPSIS framework greatly increasing the accuracy and reliability significantly.

5.6 Future Research Direction

In the future, this study can be further improved by addressing the limitations of this study and integrating the framework with other advanced technologies. Future studies could improve the reliability of the framework and the accuracy of the decision-making process by incorporating real-time supplier performance data from which can be collected and gathered from the Internet of Things and Blockchain technologies so that decisions can be updated in timely manner improving the reliability of the framework. Moreover, this framework could be further validated in future research by incorporating a significantly large dataset to identify the effectiveness of the framework in a significantly high-dimensional environment. Furthermore, the researchers can expand this framework by combining the TOPSIS framework with other highly advanced AI techniques such as deep learning or reinforcement learning which could further enhance the supplier selection framework. Another future research direction is to explore this framework in other industries and several case studies to validate its adaptability and effectiveness.

Overall, the research findings address the three main research questions of this research study. This finding showcases the potential of ML integrated TOPSIS framework in shifting the supplier selection process to align with circular economic principles by relying on data-driven insights. Even though the framework offers great advantages efficiency and accuracy, the reliability of the framework can be significantly improved by addressing the limitations in dataset, real-time data integration and studying several cases in the industry. Moreover, future advancements in AI and ML and integrating them into the framework through several case studies could further advance this approach improving the decision-making process of selecting suppliers in CSCs.

6 Conclusion

As the industry is moving towards CSCs, it is important to align the supply chain processes with economic principles. SS is one of the crucial part of the SC process that hugely contributes to shifting the traditional supply chain towards circular supply chains. Traditional supplier selection incorporates MCDM methods in its decision-making process to select suppliers and evaluate them. However, as technology advances the MCDM methods are being integrated with the ML and AI technologies. Thus, contributing to these trends in industries and literature, this study explored the integration of ML with TOPSIS to select suppliers in CSCM which remained underexplored in the literature. The main research questions of this study were to identify the efficiency and potential of the ML-integrated TOPSIS framework and to understand how it aids in aligning the supplier selection process with circular economy principles. The study was able to successfully address these questions through the findings.

The findings of the study illustrated that the ML-integrated TOPSIS framework enhanced the DM process of SS by automating the criteria weighting and supplier ranking using Random Forest ML technique. Furthermore, the ML-integrated TOPSIS improved the precision and effectiveness of the SS process within the circular supply chain framework. The ML-integrated TOPSIS model dynamically adjusts supplier rankings based on historical data whereas the traditional TOPSIS model requires static expert-defined weights.

The study further highlighted practical implications in optimizing supplier process to align with circular economy principles by incorporating ML- TOPSIS framework. Still, the examination of study limitations on data accuracy and bias, data availability, and model interpretability would additionally enhance the ML-integrated TOPSIS model to be more precise and effective. Furthermore, the assimilation of cutting-edge machine learning and AI methodologies with the TOPSIS approach, along with the verification of the ML-integrated TOPSIS model in various industry environments, are avenues of promising future research opportunities.

With continuous technological advancement specifically with expanding AI technologies the role of multi-criteria decision-making will continue to expand offering new frameworks to optimize the supplier selection process while aligning with circular economy principles. Therefore, this study which developed and validated the ML-integrated TOPSIS model to select suppliers in the CSCM remains a significant step and study toward a more scalable, efficient, accurate, and sustainable decision-making process in the SC and works as a framework to integrate advanced AI technologies into the ML-TOPSIS method to selection suppliers for CSCM in the future research.

Overall, the industry practitioners and organizations who are working on shifting their traditional supply chains to CSCs may find this research important. The output of this research which is the ML-integrated TOPSIS framework contributes to the organizations to optimize supply chain operation while reducing carbon footprints, achieving economic efficiency, and ensuring long-term sustain in a rapidly evolving global market.

References

Ali, P. J. M. (2022). Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements. ARO-The Scientific Journal of Koya University, 10(1), 85–93. https://doi.org/10.14500/aro.10955

Agarwal, A., Kenney, A. M., Tan, Y. S., Tang, T. M., & Yu, B. (2023). MDI+: A Flexible Random Forest-Based Feature Importance Framework. arXiv. https://arxiv.org/abs/2307.01932

Azadeh, A., Zarrin, M., & Salehi, N. (2016). Supplier selection in closed loop supply chain by an integrated simulation-Taguchi-DEA approach. Journal of Enterprise Information Management, 29(3), 302–326. https://doi.org/10.1108/JEIM-09-2014-0089

Azimifard, A., Moosavirad, S. H., & Ariafar, S. (2018). Selecting sustainable supplier countries for Iran's steel industry at three levels by using AHP and TOPSIS methods. Resources Policy, 57, 30–44. https://doi.org/10.1016/j.resourpol.2018.01.002

Badi, I.,& Pamucar, D. (2020). Supplier selection for steelmaking company by using combined grey-marcos methods. Decision Making: Applications in Management and Engineering, 3(2), 37–47. https://doi.org/10.31181/dmame2003037b

Bai, C., & Sarkis, J. (2010). Integrating sustainability into supplier selection with grey system and rough set methodologies. International Journal of Production Economics, 124(1), 252–264. https://doi.org/10.1016/j.ijpe.2009.11.023

Batista, L., Bourlakis, M., Liu, Y., Smart, P., & Sohal, A. (2018). Supply chain operations for a circular economy. In Production Planning and Control (Vol. 29, Issue 6, pp. 419–424). Taylor and Francis Ltd. https://doi.org/10.1080/09537287.2018.1449267

Behzadian, M., Otaghsara, S. K., Yazdani, M., & Ignatius, J. (2012). A comprehensive guide to the TOPSIS method for multi-criteria decision making. Expert Systems with Applications, 39(17), 13051-13069. https://doi.org/10.1016/j.eswa.2012.05.056

Büyüközkan, G., & Çifçi, G. (2011). A novel fuzzy multi-criteria decision framework for sustainable supplier selection with incomplete information. Computers in Industry, 62(2), 164–174. https://doi.org/10.1016/j.compind.2010.10.009

Chai, N., Zhou, W., & Jiang, Z. (2023). Sustainable supplier selection using an intuitionistic and interval-valued fuzzy MCDM approach based on cumulative prospect theory. Information Sciences, 626, 710–737. https://doi.org/10.1016/j.ins.2023.01.070

Charbuty, B., & Abdulazeez, A. (2021a). Classification Based on Decision Tree Algorithm for Machine Learning. Journal of Applied Science and Technology Trends, 2(01), 20–28. https://doi.org/10.38094/jastt20165

Chattopadhyay, R., Chakraborty, S., & Chakraborty, S. (2020). An integrated d-marcos method for supplier selection in an iron and steel industry. Decision Making: Applications in Management and Engineering, 3(2), 49–69. https://doi.org/10.31181/dmame2003049c

Darabi, S., & Heydari, J. (2016). An Interval- Valued Hesitant Fuzzy Ranking Method based on Group Decision Analysis for Green Supplier Selection. IFAC-PapersOnLine, 49(2), 12–17. https://doi.org/10.1016/j.ifacol.2016.03.003

Fallahpour, A., Udoncy Olugu, E., Nurmaya Musa, S., Yew Wong, K., & Noori, S. (2017). A decision support model for sustainable supplier selection in sustainable supply chain management. Computers and Industrial Engineering, 105, 391–410. https://doi.org/10.1016/j.cie.2017.01.005

Fallahpour, A., Yazdani, M., Mohammed, A., & Wong, K. Y. (2021). Green sourcing in the era of industry 4.0: towards green and digitalized competitive advantages. Industrial Management and Data Systems, 121(9), 1997–2025. https://doi.org/10.1108/IMDS-06-2020-0343

Feng, J., & Zhang, M. (2017). Dynamic quotation of leadtime and price for a Make-To-Order system with multiple customer classes and perfect information on customer preferences. European Journal of Operational Research, 258(1), 334–342. https://doi.org/10.1016/j.ejor.2016.08.050

Freeman, J., & Chen, T. (2015). Green supplier selection using an AHP-Entropy-TOPSIS framework. Supply Chain Management, 20(3), 327–340. https://doi.org/10.1108/SCM-04-2014-0142

Frossard, D. E. N., Pimentel, M. D., & de A. Nepomuceno, T. (2016). An approach to dealing with missing values in heterogeneous data using k-nearest neighbors. arXiv. https://arxiv.org/abs/1608.04037

Giunipero, L. C., Hooker, R. E., & Denslow, D. (2012). Purchasing and supply management sustainability: Drivers and barriers. Journal of Purchasing and Supply Management, 18(4), 258–269. https://doi.org/10.1016/j.pursup.2012.06.003

Gören, H. G. (2018). A decision framework for sustainable supplier selection and order allocation with lost sales. Journal of Cleaner Production, 183, 1156–1169. https://doi.org/10.1016/j.jclepro.2018.02.211

Govindan, K., Mina, H., Esmaeili, A., & Gholami-Zanjani, S. M. (2020a). An Integrated Hybrid Approach for Circular supplier selection and Closed loop Supply Chain Network Design under Uncertainty. Journal of Cleaner Production, 242. https://doi.org/10.1016/j.jclepro.2019.118317

Guarnieri, P., & Trojan, F. (2019). Decision making on supplier selection based on social, ethical, and environmental criteria: A study in the textile industry. Resources, Conservation and Recycling, 141, 347–361. https://doi.org/10.1016/j.resconrec.2018.10.023

Hailiang, Z., Khokhar, M., Islam, T., & Sharma, A. (2023). A model for green-resilient supplier selection: fuzzy best–worst multi-criteria decision-making method and its applications. Environmental Science and Pollution Research, 30(18), 54035–54058. https://doi.org/10.1007/s11356-023-25749-4

Harikrishnakumar, R., Dand, A., Nannapaneni, S., & Krishnan, K. (2019). Supervised machine learning approach for effective supplier classification. Proceedings - 18th IEEE International Conference on Machine Learning and Applications, ICMLA 2019, 240–245. https://doi.org/10.1109/ICMLA.2019.00045

Jakhar, S. K., Mangla, S. K., Luthra, S., & Kusi-Sarpong, S. (2019). When stakeholder pressure drives the circular economy: Measuring the mediating role of innovation capabilities. Management Decision, 57(4), 904–920. https://doi.org/10.1108/MD-09-2018-0990

Kannan, D., Mina, H., Nosrati-Abarghooee, S., & Khosrojerdi, G. (2020). Sustainable circular supplier selection: A novel hybrid approach. Science of the Total Environment, 722. https://doi.org/10.1016/j.scitotenv.2020.137936

Kaushik, V., Kumar, A., Gupta, H., & Dixit, G. (2020). A hybrid decision model for supplier selection in Online Fashion Retail (OFR). International Journal of Logistics Research and Applications, 1–25. https://doi.org/10.1080/13675567.2020.1791810

Khan, S. A., Kusi-Sarpong, S., Arhin, F. K., & Kusi-Sarpong, H. (2018). Supplier sustainability performance evaluation and selection: A framework and methodology. Journal of Cleaner Production, 205, 964–979. https://doi.org/10.1016/j.jclepro.2018.09.144

Kusi-Sarpong, S., Gupta, H., Khan, S. A., Chiappetta Jabbour, C. J., Rehman, S. T., & Kusi-Sarpong, H. (2023). Sustainable supplier selection based on industry 4.0 initiatives within the context of circular economy implementation in supply chain operations. Production Planning and Control, 34(10), 999–1019. https://doi.org/10.1080/09537287.2021.1980906

Liu, C., Rani, P., & Pachori, K. (2022). Sustainable circular supplier selection and evaluation in the manufacturing sector using Pythagorean fuzzy EDAS approach. Journal of Enterprise Information Management, 35(4–5), 1040–1066. https://doi.org/10.1108/JEIM-04-2021-0187

Little, R. J. A., & Rubin, D. B. (2019). Statistical analysis with missing data (3rd ed.). Wiley.

Ojadi, F., Kusi-Sarpong, S., Orji, I. J., Bai, C., Gupta, H., & Okwara, U. K. (2023). A decision support framework for socially responsible supplier selection in the Nigerian banking industry. Journal of Business and Industrial Marketing, 38(10), 2220–2239. https://doi.org/10.1108/JBIM-03-2022-0139

Orji, I. J., & Ojadi, F. (2021). Investigating the COVID-19 pandemic's impact on sustainable supplier selection in the Nigerian manufacturing sector. Computers and Industrial Engineering, 160. https://doi.org/10.1016/j.cie.2021.107588

Orji, I. J., & Wei, S. (2014). A decision support tool for sustainable supplier selection in manufacturing firms. Journal of Industrial Engineering and Management, 7(5), 1293–1315. https://doi.org/10.3926/jiem.1203

Ortiz-Barrios, M., Cabarcas-Reyes, J., Ishizaka, A., Barbati, M., Jaramillo-Rueda, N., & de Jesús Carrascal-Zambrano, G. (2021). A hybrid fuzzy multi-criteria decision making model for selecting a sustainable supplier of forklift filters: a case study from the mining industry. Annals of Operations Research, 307(1–2), 443–481. https://doi.org/10.1007/s10479-020-03737-y

Pavić, Z., & Novoselac, V. (2013). Application of TOPSIS method for decision making. ResearchGate. https://www.researchgate.net/publication/342347772

Pishchulov, G., Trautrims, A., Chesney, T., Gold, S., & Schwab, L. (2019). The Voting Analytic Hierarchy Process revisited: A revised method with application to sustainable supplier selection. International Journal of Production Economics, 211, 166–179. https://doi.org/10.1016/j.ijpe.2019.01.025

Prosman, E. J., & Sacchi, R. (2016). New environmental supplier selection criteria for circular supply chains: Lessons from a consequential LCA study on waste recovery. Journal of Cleaner Production, 172, 2782–2792. https://doi.org/10.1016/j.jclepro.2017.11.134

Rasmussen, A., Sabic, H., Saha, S., & Nielsen, I. E. (2023). Supplier selection for aerospace & defense industry through MCDM methods. Cleaner Engineering and Technology, 12. https://doi.org/10.1016/j.clet.2022.100590

Saha, A., Reddy, J., & Kumar, R. (2022). A Fuzzy Similarity Based Classification with Archimedean-Dombi Aggregation Operator. Journal of Intelligent Management Decision, 1(2), 118–127. https://doi.org/10.56578/jimd010205

Sasaki, H., & Sakata, I. (2021). Business partner selection considering supply-chain centralities and causalities. Supply Chain Forum, 22(1), 74–85. https://doi.org/10.1080/16258312.2020.1824531

Schroeder, P., Dewick, P., Kusi-Sarpong, S., & Hofstetter, J. S. (2018). Circular economy and power relations in global value chains: Tensions and trade-offs for lower income countries. In Resources, Conservation and Recycling (Vol. 136, pp. 77–78). Elsevier B.V. https://doi.org/10.1016/j.resconrec.2018.04.003

Senthil, S., Srirangacharyulu, B., & Ramesh, A. (2014). A robust hybrid multi-criteria decision making methodology for contractor evaluation and selection in third-party reverse logistics. Expert Systems with Applications, 41(1), 50–58. https://doi.org/10.1016/j.eswa.2013.07.010

Tavana, M., Sorooshian, S., & Mina, H. (2023). An integrated group fuzzy inference and best–worst method for supplier selection in intelligent circular supply chains. Annals of Operations Research. https://doi.org/10.1007/s10479-023-05680-0

Tushar, Z. N., Bari, A. B. M. M., & Khan, M. A. (2022). Circular supplier selection in the construction industry: A sustainability perspective for the emerging economies. Sustainable Manufacturing and Service Economics, 1, 100005. https://doi.org/10.1016/j.smse.2022.100005

Wei, Q., & Zhou, C. (2023). A multi-criteria decision-making framework for electric vehicle supplier selection of government agencies and public bodies in China. Environmental Science and Pollution Research, 30(4), 10540–10559. https://doi.org/10.1007/s11356-022-22783-6

Witjes, S., & Lozano, R. (2016). Towards a more Circular Economy: Proposing a framework linking sustainable public procurement and sustainable business models. Resources, Conservation and Recycling, 112, 37–44. https://doi.org/10.1016/j.resconrec.2016.04.015

Wu, C., Lin, Y., & Barnes, D. (2021). An integrated decision-making approach for sustainable supplier selection in the chemical industry. Expert Systems with Applications, 184. https://doi.org/10.1016/j.eswa.2021.115553

Yan, L. (2022). Evaluation Method and Application of Power Grid Material Suppliers Based on Entropy Weight TOPSIS Model. EI2 2022 - 6th IEEE Conference on Energy Internet and Energy System Integration, 1924–1927. https://doi.org/10.1109/EI256261.2022.10116353

Yang, M., Smart, P., Kumar, M., Jolly, M., & Evans, S. (2018). Product-service systems business models for circular supply chains. Production Planning and Control, 29(6), 498–508. https://doi.org/10.1080/09537287.2018.1449247

Appendices

Appendix 1. Overview of the Questionnaire

Section 1 of 3

Identifying the importance of different supplier selection criteria in the supplier selection decision making process of the company

Dear Participant,

My name is **Vijini Witharana**, and I am a **second-year Master's student** at the **University of Vaasa, Finland**, specializing in **Industrial Systems Analytics**. As part of my thesis research, I am developing an **integrated framework using Machine Learning and TOPSIS for sustainable supplier selection in circular supply chains**.

This questionnaire aims to gather expert insights on the **importance of different supplier selection criteria** within your company. Your valuable responses will contribute to identifying key factors in sustainable supplier evaluation and improving decision-making processes.

The survey will take approximately **5-10 minutes**, and all responses will be treated with **strict confidentiality**. Thank you for your time and support!

Best regards,
Vijini Witharana

Age

- ○ Below 25
- ○ 25 - 34
- ○ 35 - 44
- ○ 45 - 54
- ○ 55 and Above

Years of Experience in the Industry

- ○ Less than 1 year
- ○ 1 - 3 years
- ○ 4 - 7 years
- ○ 8 - 12 years
- ○ More than 12 years

Years of Experience in the Industry

- ○ Less than 1 year
- ○ 1 - 3 years
- ○ 4 - 7 years
- ○ 8 - 12 years
- ○ More than 12 years

Department

- ○ Procurement
- ○ Purchasing
- ○ Supply Chain Management
- ○ Other

Section 2 of 3

Supplier Selection Criteria Evaluation

Please rate the importance of each **supplier selection criterion** on a scale of **1 to 5**, where:
1 = Not Important, 2 = Slightly Important, 3 = Neutral, 4 = Important, 5 = Very Important

Price

1	2	3	4	5
○	○	○	○	○

Quality

1	2	3	4	5
○	○	○	○	○

Delivery Time

1	2	3	4	5
○	○	○	○	○

Offer Validity Days

1	2	3	4	5
○	○	○	○	○

Payment Terms

1	2	3	4	5
○	○	○	○	○

Procurement Actions

1	2	3	4	5
○	○	○	○	○

Quantity

1	2	3	4	5
○	○	○	○	○

Delivery Terms

1	2	3	4	5
○	○	○	○	○

Delivery Mode

1	2	3	4	5
○	○	○	○	○

Product Fit

1	2	3	4	5
○	○	○	○	○

Resource Efficiency

1	2	3	4	5
○	○	○	○	○

Recyclability

1	2	3	4	5
○	○	○	○	○

Carbon Footprint

1	2	3	4	5
○	○	○	○	○

Ethical Sourcing

1	2	3	4	5
○	○	○	○	○

Sustainability Certification

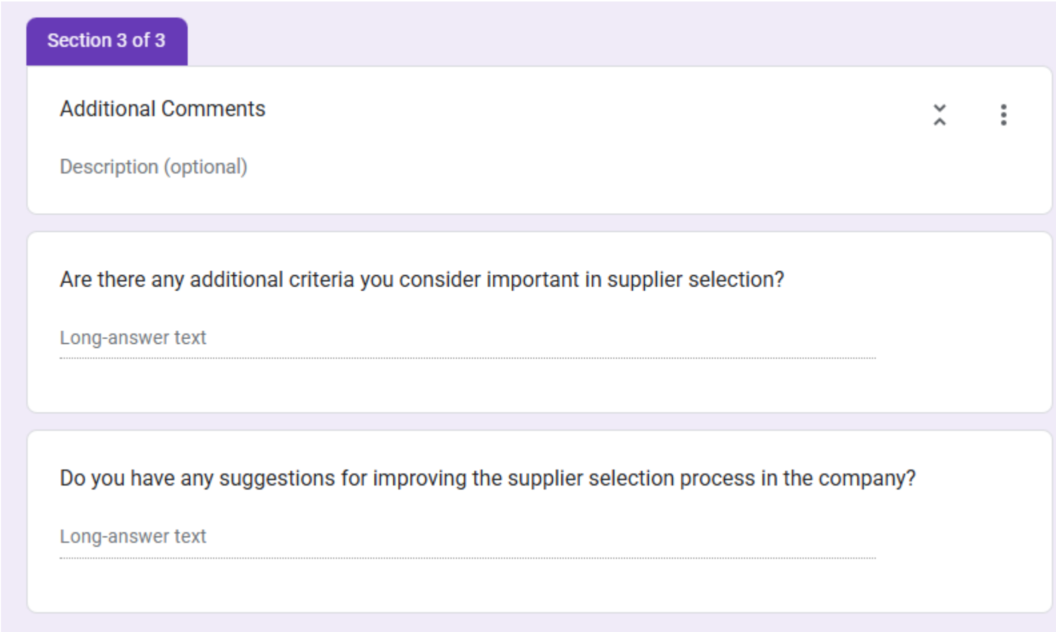
1	2	3	4	5
○	○	○	○	○

Supplier Reputation

1	2	3	4	5
○	○	○	○	○

Innovation Capacity

1	2	3	4	5
○	○	○	○	○



Appendix 2. Overview of Final Dataset

1	Supplier_ID	Price_per_Unit	Quality	ivery_Ti	Validity	Payment_Terms	rement_A	Quantity	livery_Ter	livery_Mo	Product_Fi	ical_Sourc	urce_Effic	ecyclabilit	on_Footp	Performanc
17	1492	263,84		25		Cash on delivery	Purchase	3018	FOB	Air	No	Yes	Low	Medium	Low	85,7
18	1126	975,32	Medium	57	58	Net 60	Purchase	25	CIF	Road	No	Yes	Medium	Low	Medium	83,8
19	1776		Medium	36	32	50% upfront			FOB	Air	No	Yes	Medium	Low	Low	76,4
20	1165	1280,5	Low		16	50% upfront	Purchase	37	CIF	Air	Yes	Yes	High	Medium		87,3
21	1272	1614,12	Low	60	50	Net 60	Purchase	3302		Road	Yes	No	High	High	Medium	81
22	1666	1739,86	Low	24	39	Cash on delivery	Purchase	27	Ex-works	Air	No	No	Medium	High	High	90,7
23	1573	1577,83	Medium	83	35	Cash on delivery	Purchase	2147	Ex-works	Road	No	Yes	Low	High	Low	89,7
24	1534	47,46	High	51	21	Net 60	Purchase	48	CIF	Air	No		Low	High	Low	83,8
25	1787		Medium	39	34	50% upfront	Lease		Ex-works	Air	Yes	Yes	Low	Medium	High	88,4
26	1732	1857,62	High	83	35	50% upfront	Purchase	678	Ex-works	Sea	No	No	Medium	Medium	High	90,2
27	1118	333,49	Low	10	16	50% upfront	Lease	4069	FOB		No	Yes	High	Low	High	87,7
28	1233	101,39	Low	20	15	Net 60	Purchase	3	CIF	Road	Yes	No	Medium		Medium	81,4
29	1507	301,21	High		21	Net 60	Lease	4852	Ex-works	Road	No	No	Medium	High	High	81,6
30	1627	1405,89	Medium	76		Cash on delivery	Lease		Ex-works	Air	No		Medium	High	High	88,8
31	1731	1495,58	Medium	8	8	Net 30	Purchase	1082		Road	Yes	Yes	Medium	Medium	High	84,8
36	1600	1432,8	Low		22		Lease	2937		Sea	Yes	Yes	Low	Low	Low	94,4
37	1732	1893,67	High		24	50% upfront	Lease	28	CIF	Air	No	Yes	Low	Medium	Medium	82,7
38	1902	338,61	Medium	20	10	50% upfront		2765	CIF	Road	No	No		Low		90,6
39	1030	1419,02	Medium	19	50	Cash on delivery	Purchase	2898	CIF	Road	No	Yes	Medium	Medium	Medium	90,2

Appendix 3. Python Code for Correlation Heatmap

```
# Perform correlation analysis
correlation_matrix = filtered_data.corr()

# Visualize the correlation matrix with a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt='.2f', square=True, cbar=True)
plt.title("Correlation Heatmap of Supplier Selection Criteria", fontsize=16)
plt.xticks(rotation=45, ha='right', fontsize=10)
plt.yticks(fontsize=10)
plt.tight_layout()
plt.show()
```

Appendix 4. Python Code for Top 10 Criteria.

```
# Create a bar chart to represent the top 10 criteria
plt.figure(figsize=(10, 6))
top_10_criteria.plot(kind='bar', color='skyblue', edgecolor='black')
plt.title("Top 10 Supplier Selection Criteria Based on Survey Data", fontsize=14)
plt.xlabel("Criteria", fontsize=12)
plt.ylabel("Average Importance Score", fontsize=12)
plt.xticks(rotation=45, ha='right', fontsize=10)
plt.grid(axis='y', linestyle='--', alpha=0.7)
plt.tight_layout()

# Show the chart
plt.show()
```

Appendix 5. Python Code for Traditional vs Circular Supply Chain Criteria

```
# Categorizing traditional vs circular supply chain criteria
traditional_criteria = criteria[:10] + criteria[15:]
circular_criteria = criteria[10:15]

# Colors: blue for traditional, green for circular
colors = ["blue" if c in traditional_criteria else "green" for c in criteria]

# Plotting
plt.figure(figsize=(12, 5))
bars = plt.bar(criteria, values, color=colors)

# Labels and title
plt.xticks(rotation=90)
plt.xlabel("Supplier Selection Criteria")
plt.ylabel("Average Importance")
plt.title("Comparison of Importance of Traditional vs Circular Supply Chain Criteria")

# Adding legend
import matplotlib.patches as mpatches
trad_patch = mpatches.Patch(color='blue', label='Traditional Criteria')
circ_patch = mpatches.Patch(color='green', label='Circular Criteria')
plt.legend(handles=[trad_patch, circ_patch])

plt.tight_layout()
plt.show()
```

,

Appendix 6. Python Code Missing Values Analysis using Bar Charts and Averages

```
# Create figure and axis objects
fig, ax1 = plt.subplots(figsize=(10, 6))

# Bar chart for Missing Count
ax1.bar(columns, missing_count, color='lightblue', label='Missing Count')
ax1.set_xlabel('Columns')
ax1.set_ylabel('Missing Count')
ax1.set_xticklabels(columns, rotation=45, ha='right')

# Create a second y-axis for percentage
ax2 = ax1.twinx()
ax2.plot(columns, percentage, color='O', marker='o', label='Percentage', linestyle='-', linewidth=2)
ax2.set_ylabel('Percentage (~%)')

# Add labels and a title
plt.title('Missing Count and Percentage by Column')
ax1.set_ylim(0, max(missing_count) + 50)
ax2.set_ylim(0, 25)

# Show the legend
fig.legend(loc='upper left', bbox_to_anchor=(0.1, 0.9))
```

Appendix 7. Python Code for Missing Values Heatmap

```
# Create a missing data heatmap
plt.figure(figsize=(12, 8))
sns.heatmap(df.isnull(), cbar=False, cmap='viridis', annot=False, xticklabels=df.columns, yticklabels=False)
plt.title('Missing Data Heatmap (Before Cleaning)')
plt.xlabel('Supplier Selection Criteria')
plt.show()
```

Appendix 8. Python Code for Boxplots of Before and After Scaling

```
# ================================
# 3 Scaling Check (Standardization)
# ================================
scaler = StandardScaler()
scaled_data = scaler.fit_transform(df.select_dtypes(include=[np.number]))  # Only numerical columns
df_scaled = pd.DataFrame(scaled_data, columns=df.select_dtypes(include=[np.number]).columns)

# Boxplot after scaling
plt.figure(figsize=(12, 6))
sns.boxplot(data=df_scaled)
plt.xticks(rotation=45)
plt.title("Boxplots of Features After Scaling")
plt.show()
```

Appendix 9. Python Code for ML-integrated TOPSIS Rankings vs Traditional TOPSIS Rankings

```
# Visualization
plt.figure(figsize=(10, 6))
sns.scatterplot(x=df["Traditional_Rank"], y=df["ML_Integrated_Rank"], hue=df["Supplier_ID"], palette="viridis
plt.xlabel("Traditional TOPSIS Rank")
plt.ylabel("ML-Integrated TOPSIS Rank")
```

Appendix 10. Python Code for Histogram of Rank Differences

```
# Histogram of Rank Differences
df['Rank_Difference'] = df['ML_Integrated_TOPSIS_Rank'] - df['Traditional_TOPSIS_Rank']
plt.figure(figsize=(8, 6))
sns.histplot(df['Rank_Difference'], bins=30, kde=True)
plt.axvline(df['Rank_Difference'].mean(), color='red', linestyle='dashed', linewidth=2)
plt.xlabel("Rank Difference (ML - Traditional)")
plt.ylabel("Frequency")
plt.title("Distribution of Ranking Differences")
plt.show()
```

Appendix 11. Code for ML modelling

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestRegressor
from sklearn.preprocessing import LabelEncoder, MinMaxScaler
from sklearn.feature_selection import mutual_info_regression

# Load cleaned dataset
file_path = "/content/final_procurement_data.csv"
df = pd.read_csv(file_path)

# Keep Supplier_ID for ranking but don't use it in ML
supplier_ids = df["Supplier_ID"]

# Separate numerical and categorical columns
numeric_cols = df.select_dtypes(include=[np.number]).columns.tolist()
categorical_cols = df.select_dtypes(include=["object"]).columns.tolist()

# Encode categorical variables using LabelEncoder
label_encoders = {}
for col in categorical_cols:
    le = LabelEncoder()
    df[col] = le.fit_transform(df[col])
    label_encoders[col] = le

# Define X (features) and y (target)
X = df.drop(columns=["Supplier_Performance_Score", "Supplier_ID"])  # Exclude target variable
y = df["Supplier_Performance_Score"]

# Train Random Forest for feature importance analysis
rf = RandomForestRegressor(n_estimators=500, max_depth=10, random_state=42)
rf.fit(X, y)
```

Appendix 12. Code for Feature Importance and Criteria Weights

```
# Train Random Forest for feature importance analysis
rf = RandomForestRegressor(n_estimators=500, max_depth=10, random_state=42)
rf.fit(X, y)

# Get feature importance
feature_importance = pd.DataFrame({"Feature": X.columns, "Importance": rf.feature_importances_})
feature_importance = feature_importance.sort_values(by="Importance", ascending=False)

# Normalize Criteria Weights
feature_importance["Criteria_Weight"] = feature_importance["Importance"] / feature_importance["Importance"].sum()

# Display Feature Importance and Criteria Weights
print("Feature Importance Table:")
print(feature_importance)
print("\nCriteria Weights Table:")
print(feature_importance[["Feature", "Criteria_Weight"]])

# Plot Feature Importance
plt.figure(figsize=(12, 6))
sns.barplot(x="Importance", y="Feature", data=feature_importance, palette="viridis")
plt.title("Feature Importance (Random Forest)")
plt.xlabel("Importance Score")
plt.ylabel("Feature")
plt.show()

# Additional Feature Selection using Mutual Information
mi_scores = mutual_info_regression(X, y)
mi_feature_importance = pd.DataFrame({"Feature": X.columns, "Mutual_Info_Score": mi_scores})
mi_feature_importance = mi_feature_importance.sort_values(by="Mutual_Info_Score", ascending=False)
print("\nMutual Information Scores:")
print(mi_feature_importance)
```

Appendix 13. Code for TOPSIS Calculations

```
def topsis(decision_matrix, weights, benefit_criteria):
    # Step 1: Normalize the decision matrix
    norm_matrix = decision_matrix / np.sqrt((decision_matrix ** 2).sum(axis=0))

    # Step 2: Multiply by Criteria Weights
    weighted_matrix = norm_matrix * weights

    # Step 3: Determine Ideal Best (A+) and Ideal Worst (A-)
    # Apply benefit_criteria element-wise
    ideal_best = np.where(benefit_criteria, np.max(weighted_matrix, axis=0), np.min(weighted_matrix, axis=0))
    ideal_worst = np.where(benefit_criteria, np.min(weighted_matrix, axis=0), np.max(weighted_matrix, axis=0))

    # Step 4: Compute distances from Ideal Best and Ideal Worst
    dist_best = np.sqrt(((weighted_matrix - ideal_best) ** 2).sum(axis=1))
    dist_worst = np.sqrt(((weighted_matrix - ideal_worst) ** 2).sum(axis=1))

    # Step 5: Compute TOPSIS Score
    topsis_scores = dist_worst / (dist_best + dist_worst)

    return topsis_scores

# Prepare Decision Matrix (without Supplier_ID)
decision_matrix = df[X.columns]

# Normalize using Min-Max Scaling
scaler = MinMaxScaler()
decision_matrix = scaler.fit_transform(decision_matrix)

# Get weights from Random Forest results
weights = feature_importance["Criteria_Weight"].values

# Define Benefit Criteria (1 for max benefit, 0 for cost-based)
benefit_criteria = np.ones(len(weights))  # Assuming all criteria are beneficial
```