Combining Artificial Intelligence and Multi Criteria Decision Making Approaches for Supplier Selection

Ahmad Abdulla

Supervisor: Dr George Bargiannis

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

University of Huddersfield

United Kingdom

June 2022

Copyright

Dedication

To my parents, sisters and brothers, my wife and my kids who provided me with endless support.

Acknowledgements

First, I would like to express my sincere gratitude to my advisor, Dr George Bargiannis, who provided encouragement, guidance, and technical support throughout this dissertation. None of this work could have been done without his experience and insight. I would also like to thank my examiners' committee, Dr Ilias Tachmazidis and Dr Kamran Mahroof, for their insightful feedback, constructive criticism, and rigorous examination of my work. Their expertise and scholarly input have greatly enriched the quality and depth of this thesis.

Secondly, I would like to pass on my thanks and appreciation to all the colleagues in the Computing and Engineering department and the staff who, over the years, have made my time a wonderful experience both professionally and personally, as well as my friends who have always been so helpful and encouraging. I want to extend a special thank you to my wife and my kids. Their unwavering support, understanding, and patience during the demanding process of completing this thesis have been invaluable. Their love and encouragement have been a constant source of motivation, and I am grateful for their presence in my life. Also, thanks and gratitude to my brothers and sisters for their high hopes and encouragement. And the souls of my father and mother. Also, I would like to thank my friends Dr Mahmud Ahmed, Dr Aiman Elragig and Dr Salem Othman for their warm feelings and encouragement.

Abstract

Supplier selection is one of the most important activities of a purchasing manager and involves the identification of important criteria to select appropriate suppliers from a pool of available ones, prioritise them and evaluate their performance. Traditionally, multi-criteria decision-making approaches (MCDM) have been utilised for this process. Supplier selection has become increasingly complex both in terms of selection criteria, as a result of expanded data collection processes, and in terms of supplier numbers, due to the effects of globalisation. This complexity has led to considering Artificial Intelligence (AI) techniques, primarily machine learning (ML), to enhance and further improve the supplier selection process. While AI and ML provide the advantage of increased overall performance and efficiency, they are not always understandable and easily adopted by humans. Novel hybrid solutions that integrate MCDM and ML have been introduced to combine their advantages and mitigate their drawbacks. The main hybrid approach proposed involves using interpretable ML algorithms combined with MCDM to solve supplier selection problems in two ways: The first hybrid model uses a decision tree (DT), an interpretable machine learning algorithm, to reduce the complexity of supplier selection in terms of the number of criteria or suppliers, followed by the Analytic Hierarchy Process (AHP), one of the most common MCDM techniques, to rank and select the best suppliers. This maximises familiarity and adaptability while using ML to boost performance. The second hybrid model begins with two MCDM methods, FUCOM to weigh criteria and TOPSIS to rank suppliers. These result in a labelled dataset that is used to train a DT model in the second stage of this hybrid. This classifier is used as the main supplier selection mechanism. The two hybrid approaches are formalised in a generalisable manner, which allows different MCDM and ML approaches to be employed in place of the ones explored in this thesis. The applicability of the proposed approaches is demonstrated through a case study of companies in the oil and gas sector in Libya. The case study involved two datasets, one containing over 20,000 submitted offers of suppliers to requests from different departments, and one recording the evaluation process and outcome for 2,300 requisitions, explaining the selection criteria and the reasons for selecting the chosen supplier.

Experiments show that the first hybrid approach (DT+AHP) achieves considerable gains in precision, up to 90%, outperforming individual ML models in accuracy, precision, recall, and F1 score by over 14%, 18%, 11% and 16%, respectively. Moreover, an F1 score of 87% is achieved for both classes, showing that the approach performs well across both selected and non-selected suppliers. In the second hybrid approach (FUCOM+TOPSIS+DT), experiment results confirm superiority over approaches using only ML, with increases by more than 5% across all metrics and achieving an F1 score of 72.57%. Additionally, the use of DT, an interpretable ML approach, in combination with MCDM methods familiar to procurement stakeholders, allows for the results of both proposed approaches to be more explainable and understandable than pure ML-based approaches.

This research provides practical implications for supply chain stakeholders in supplier selection by facilitating intelligent decision-making by providing two different hybrid approaches; supplier selection solutions can be tailored to individual needs, depending on whether the ability to explain outcomes is prioritised or whether it is vital to a maximising performance by limiting inaccurate selection suggestions. Such hybrid approaches can also lead to increased adoption of intelligent technologies in supplier selection and supply chain processes in general and provide fruitful ground for further interdisciplinary research in AI and supply chains.

List of Publications

Journal Publications:

1. (Under Review) Abdulla, A., Baryannis, G. (2023). Explainable Supplier Selection: A Hybrid Approach Combining Multi-Criteria Decision Making and Machine Learning. *Intelligent Systems with Applications*.

2. (Under Review) Abdulla, A., Baryannis, G. & Badi, I. (2023). Effective Supplier Selection Combining Machine Learning with the MARCOS Method. *Expert Systems*.

Conference Publications:

1. Abdulla, A., Baryannis, G., & Badi, I. (2019). Weighting the Key Features Affecting Supplier Selection using Machine Learning Techniques. *7th International Conference on Transport and Logistics, Niš, Serbia*, December 2019.

Contents

List of Figures

List of Tables

Nomenclature

MAUT	Multi-Attribute Utility Theory
AHP	Analytic Hierarchy Process
α	Cronbach's alpha
ANN	Artificial Neural Network
ANP	Analytic Network Process
BN	Bayesian network
CA	Cluster Analysis
CBR	Case-Based Reasoning
CI	Consistency Index
CV	Cross Validation
DEA	Data Envelopment Analysis
DFC	Deviation in Full Consistency
DM	Decision Maker

ELECTRE Elimination Et Choix Traduisant La Realite

FST Fuzzy Set Theory

GA Genetic Algorithm

GDM Group Decision-Making method

GP Goal Programming

IFS Intuitionistic Fuzzy Set

KNN K-Nearest Neighbors

LLSM Logarithmic Least Square Method

MADM Multi-Attribute Decision Making

MCDA Multi-Criteria Decision Analysis

MCDM Multi-Criteria Decision Making

MIP Mixed-Integer Programming

MLP Multi-Layer Perceptron

MODM Multi-Objective Decision Making

MULTIMOORA Multi-Objective Optimization on the basis of a Ratio Analysis

NIS Negative Ideal Solution

PCA Principal Component Analysis

PIS Positive Ideal Solution

SCM Supply Chain Management

SPSS Statistical Package for the Social Sciences

SVM Support Vector Machine

Chapter 1

Introduction

This chapter begins with some basic information on the motivation for, and problems addressed in this thesis and then discusses the study's main aim and contributions it has made to knowledge.

1.1 Overview

Supplier selection is a critical step in the purchasing process, with the purchasing managers expected to select the suppliers that best match their business requirements. Companies must obtain goods and raw materials at reasonable prices, on time, in sufficient quantity, and of high quality from appropriate suppliers. Companies normally search for the best suppliers to increase their performance and maintain existing long-term relationships, decrease purchasing risk, developing positive connections between buyers and suppliers whilst maximising consumer satisfaction (C. Li et al., 1997). The selection process involves identifying and selecting suitable suppliers from a pool of suppliers while considering certain criteria. Incorrectly selecting suppliers may result in financial problems, which could negatively affect the company's success (Humphreys et al., 2007). The supplier selection process includes identifying a goal, defining the criteria for the purpose of achieving that goal, pre-evaluating qualified suppliers on the basis of established criteria and, lastly, making the final selection (Kıran et al., 2020). The supplier selection process is ongoing, with the goal of following, developing, and, if necessary, replacing existing suppliers with new ones using the criteria established for the process (Kıran et al., 2020).

The most crucial aspect of supply chain management (SCM) is purchasing, with multi-criteria analysis being very important for deciding which suppliers to use, with suppliers playing a determining role in ensuring that raw materials are available for a company's manufacturing operations (Celebi & Bayraktar, 2008).

Supplier selection when purchasing is a strategic management activity that takes place at an important stage of stages of the supply chain (Choi & Hartley, 1996; Cole & Aitken, 2019).

The supplier selection process is divided into three steps:

- Understanding business priorities and strategies.
- Defining and weighing selection criteria.
- Making a final selection on which supplier to choose.

Business priorities and strategies are responsible for selecting which of many possible criteria are relevant and important and determining how these criteria should be measured. Evaluating suppliers and making complex decisions on selecting suitable suppliers is traditionally the responsibility of purchasing department managers, who often give different rankings to the assessment criteria based on personal experience or word-of-mouth, so selecting suitably qualified suppliers is, for many purchasing managers, a difficult and complex task (Liu & Hai, 2005).

The selection of criteria and weighing them to decide on the chosen suppliers is the most significant part of the selection process. The procedure utilised to extract the weights of the criteria is an important issue that should be given more consideration in these investigations, in particular when compensating strategies of MCDM are applied (Schramm et al., 2020).

Once the possible suppliers and selection criteria have been determined, the next step is to choose the best suppliers. This is a complex problem because selecting a supplier is a collective decision, which includes multiple and often conflicting criteria involving different decision-makers, and has the potential for inaccuracy (Boran et al., 2009).

1.2 Motivation

Generally, researchers have been battling to provide answers to supplier selection-related questions (Badi & Pamucar, 2020; Yazdani et al., 2020), such as:

- How can purchasing managers select the best suppliers and make the right decisions to help their companies reach their maximum potential?

- How can the purchasing department evaluate and analyse the decision-making process of selecting suppliers, and which technique should be used to achieve these objectives?

Existing purchasing managers and experts tend to rely on traditional methods to select the best suppliers, and sometimes the expectations on the part of the supplier may not be met (Cavalcante et al., 2019).

The process becomes more complicated as the number of suppliers and criteria increases. Companies may have integrated systems used over many years to evaluate suppliers. Still, much effort is required by such manual work, and the routines can be tedious, time-consuming, and subjective (Abdulla et al., 2019).

However, some businesses have tried to find ways to select suitable suppliers from many different companies so they can be more productive and produce a wider range of products. One such solution is to use integrated methods. Due to the practical complexity of this type of a decision, combining different methods is expected. However, to avoid inconsistent results, it is vital to grasp the underlying distinctions between the methods before integrating them (Schramm et al., 2020).

The following factors contribute to the complexity of the decision-making process and need to be considered when selecting suppliers (De Boer et al., 1998), (Sonmez, 2006):

- The variety of criteria.
- Certain criteria conflict with others.

- The purchasing process is subjected to both internal and external constraints.
- Involvement of several different alternatives.

1.3 Related Research

Researchers have introduced frameworks for solving decision-making problems in SCM; Lee, for example, developed a concept of SCM which enables participating companies to effectively combine products and services into a long-term relationship (Lee et al., 2001). Subsequently, Wu and Barnes noted that the intelligent supply chain had a greater chance of achieving an economic advantage when transforming business situations (C. Wu & Barnes, 2011). It has been found that supplier selection and evaluation processes are critical in influencing whether a company's supply chain succeeds or fails (Sarkis & Talluri, 2002). Therefore, choosing the most suitable approach and criteria for selecting a supplier are among the most important decisions in the procurement process (Durmić, 2019). Traditionally, suppliers are selected based on price, quality, and other widely acknowledged criteria (Cengiz et al., 2017; Rezaei et al., 2016; Weber et al., 1991; Kıran et al., 2020). However, in the modern commercial environment, these criteria are insufficient (Zouggari & Benyoucef, 2012). With increased globalisation providing access to far more suppliers, this issue, in recent years, has become more common, especially with the proliferation of data collection now possible across supply chains, allowing additional criteria to be included in the selection process. Therefore, many practitioners and academics have suggested proposals for multi-criteria systems (dos Santos Amorim et al., 2020). The use of strategies with multiple criteria has resulted in multiple-criteria decision-making (MCDM) methods for evaluating and selecting suitable suppliers. MCDM techniques primarily assist companies in making optimally consistent decisions with predetermined criteria created and assessed by decision-makers (Azadfallah, 2017).

Given the importance and complexity of the supplier selection process, significant research effort has been devoted to facilitating it, particularly in MCDM approaches. For example, the Analytic Hierarchy Process (AHP) (Saaty, 1980), Analytic Network Process (ANP) (Saaty, 1996), Data Envelopment Analysis (DEA) (Charnes et al., 1978), the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) (Huang & Yoon, 1981) and Full COnsistency Method (FUCOM) (Pamučar et al., 2018) are all well-known examples.

One of the important benefits of such approaches is their transparency and explainability: it is straightforward to explain why a particular supplier is selected based on the weights of individual criteria included in, say, the AHP or the objective functions for efficiency of DEA or the decision matrix of TOPSIS. However, this advantage is limited when correlated with scalability in the number of criteria and alternatives. According to the literature, these approaches can perform well when dealing with MCDM situations involving small datasets. While the hierarchical structure of AHP can theoretically be adjusted in size (Velasquez & Hester, 2013), it is difficult to increase their capacity to include the many business requirements for implementing practical systems due to the rapidity with which the size of the problem increases (M. Tang & Liao, 2021).

In practice, their limits and pairwise comparisons and explanations become too complex when there are numerous criteria and alternatives involved (R. J. Kuo et al., 2010; Junior et al., 2014). Therefore applying these methods to complex projects is difficult and demands a significant amount of time and money, and might result in delayed decision-making and, ultimately, inefficiency (Mahmoudi et al., 2020). The rapid changes taking place in the business and market environment imply that traditional multi-criteria decision-making techniques may enhance our understanding of supplier selection rules, but cannot guide us to find the best suppliers when many suppliers and many criteria are involved (Kıran et al., 2020).

Although the problems associated with the selection of suppliers have been identified in many business environments, the emergence of the big data era and artificial intelligence (AI) techniques, especially over the last decade, has made the solving of these problems an integral aspect of computer science. Consequently, researchers increasingly prefer to apply AI techniques to such issues as they are considered more effective than traditional approaches (Martínez-López & Casillas, 2013).

One such type of AI that might further enhance supplier evaluation methods is machine learning (ML), and, nowadays, many researchers rely on ML methods to evaluate the vast number of global suppliers and resolve the associated complexity of selecting a most suitable supplier (Guo et al., 2009; Abdulla et al., 2019; Cavalcante et al., 2019; Ho et al., 2010).

Most AI approaches, such as neural networks (NN) and Support Vector Machines (SVM), work satisfactorily and accurately but their outputs are not interpretable (Baryannis, Dani, & Antoniou, 2019), while they also need large amounts of data for training and testing the model which also means it takes much longer to train the model than most ML algorithms (Zhu et al., 2013).

The difficulty and/or inability to provide explanations is a major hindrance to the adoption of intelligent approaches, as it hinders trust and acceptance of solutions and decisions produced (Molnar, 2019).

Interpretability is one of the important factors in evaluating models, so having a supplier selection methodology that is inherently interpretable raises the level of trust in the decisions made, increasing the possibility that it will be adopted to solve such problems (Baryannis, Dani, Validi, & Antoniou, 2019; Ni et al., 2020). Good interpretability is a criterion that should not be underestimated, and inadequate interpretability can be a major cause of reluctance to utilise a particular technique (Zhu et al., 2013). Integrated approaches may be the answer by combining the benefits of traditional MCDM approaches and more modern, intelligent ones, while mitigating their drawbacks.

1.4 Thesis Contribution

Several researchers have explored the synergies between MCDM and ML (as detailed in Chapter 3). However, the main focus has been on solutions that use any form of ML and where ML always plays the main role in the decision making process, without taking into account issues related to interpretability and explainability that may affect the adoption of such solutions. However, decision-makers in the business sector need to trust the models, and transparency is a major factor in influencing trust. The research hypothesis investigated in this thesis is that a hybrid supplier selection framework that combines MCDM with ML can achieve high performance while retaining strong explainability and trustworthiness, thereby removing barriers to adopting the model for decision-makers to make their decisions. In investigating this hypothesis, two main research questions are considered:

- How can ML techniques be leveraged to reduce the complexity of the supplier selection problem?
- What is the most suitable combination of MCDM and ML given different priorities that range from performance to explainability?

To illustrate the hybrid method's efficacy, this model is applied to a real data set from a Libyan oil and gas company for supplier evaluation and compares the result of the proposed model with the results of the Libyan company's decision-makers. It was found that the comparison showed a high degree of matching of the two sets of results.

To address these research questions, this thesis aims to achieve the following objectives:

- Examine different MCDM approaches to determining and weighing selection criteria and ranking and selecting suppliers.
- Compare different ML algorithms to determine the most appropriate ones to be considered for supplier selection problems.

- Explore how the choice of different MCDM and ML combinations influences the decision-making aspects of a hybrid approach.
- Apply different MCDM and ML combinations in a case study involving oil and gas companies in Libya.
- Validate and evaluate the results of hybrid approaches to determine suitability in different contexts.

This study intends to make a contribution toward giving decision-makers in companies a prediction model that is data-driven and analytical. In addition, the insights provided by a data-driven strategy can offer businesses accurate predictions, which can help them choose the most reliable suppliers.

The main contributions of this thesis are as follows:

- A novel hybrid supplier selection approach that uses interpretable ML to reduce complexity and MCDM to carry out supplier selection decisions. This approach can be easily adopted by purchasing managers who are familiar with MCDM approaches, such as AHP, while also allowing them to harness historical supplier selection data through machine learning algorithms.
- A second novel hybrid supplier selection approach utilises MCDM techniques to label datasets and classify suppliers, followed by training an ML classifier that can be used to select suppliers.
- A demonstration of the efficacy of the proposed approaches through implementation using a real dataset, facilitating supplier selection for oil and gas companies based in Libya.

- An exploration of the processes involved in developing intelligent solutions for supplier selection.

1.5 Thesis Outline

The remainder of this thesis is organised as follows. Chapter 2 presents an overview of background knowledge on machine learning and its various kinds, multi-criteria decision-making strategies and types, and supplier selection and evaluation. Chapter 3 discusses the literature review related to the supplier selection problem and provides a critical analysis of different hybrid methods of solving such problems. Chapter 4 introduces the two proposed hybrid supplier selection frameworks and their main components; ML and MCDM. Chapters 5 and 6 illustrate the applicability of the proposed approaches in a real-world case study, explaining the case study and datasets in detail before providing a detailed account and discussion of experiments and results of applying the approaches. Finally, chapter 7 summarises the main contributions of this thesis and points out future research directions.

Chapter 2

Background

2.1 Introduction

This chapter provides an overview of the research background. This entails various topics, starting with the definition of supplier selection, Artificial Intelligence (AI) and Machine Learning (ML) and some of the related machine learning techniques and their types. Moreover, multi-criteria decision-making and their most popular methods are explained in this chapter. Finally, a brief overview of how the supplier selection problem has been approached in literature is provided, leading to the more focused analysis that is provided in the next chapter.

2.2 Supplier Selection

The supplier selection study focuses on selecting the most appropriate supplier for effectiveness and efficiency, which are related to implementing methodologies and frameworks. Various studies solved the supplier selection problem using many criteria, such as multi-criteria decision-making methodologies and artificial intelligence technologies. Supplier selection is a strategic procedure since it can partially, if not fully, alleviate risk in the supply chain. Thus, supplier selection is a critical component of any trading company (Mukherjee, 2017).

Supplier selection can be conducted in two ways: through a single source or multiple sourcing. A single supplier in a single source manages the whole supply chain. Conversely, the entire supply chain comprises a group of suppliers from multiple sources.

Researchers agree that there are two stages to selecting a good supplier. First, decision criteria should be identified and weighted, and second, the techniques used by the decision-makers to analyse and define the preferred suppliers must be identified (Chan et al., 2008).

Supplier selection and evaluation have been the subject of much earlier research, which has identified a variety number of evaluation criteria and selection models for supplier selection. Dickson's pathfinder study from 1966 has been one of the most widely recognised studies when it comes to determining the criteria to be used in supplier evaluation; 23 criteria were proposed in his research work for supplier selection (Dickson, 1966).

Many researchers have confirmed that quality, cost, and delivery time are selected as essential criteria, which can be quantified (Chan et al., 2008). When choosing a supplier, qualitative variables such as economic stability, top management compliance, a firm's history, and the variety of products available are also crucial, highlighting that the need for qualitative criteria can be as important as quantitative criteria for supplier selection (Ellram, 1990; Weber & Current, 1993; Rezaei et al., 2016).

2.3 Artificial Intelligence

Artificial Intelligence refers to the process of developing systems that exhibit intelligence in their operation. In the field of computer science, artificial intelligence refers to the study of intelligent agents, which can be defined as any technology that can sense its surroundings and take activities that increase the likelihood of successfully attaining its objectives. When a machine is capable of performing functions that people associate with other human minds, such as learning and problem-solving, then it can be considered artificially intelligent (Shinde & Shah, 2018).

There have been varying definitions of AI in literature. In the context of supply chain research, Baryannis, Validi, et al. (2019) consider an approach artificially intelligent if it satisfies the following two characteristics: it is capable of deciding on a course of action autonomously in order to achieve one or more objectives; it can do so under a supply chain environment that is partially unknown.

AI encompasses a wide range of approaches and algorithms that range from knowledge-based to data-driven. In the latter case, it has been mostly associated with the ability to learn from data in the form of ML, which is discussed in the next section.

2.4 Machine Learning

In the current “data age”, computer processing power and storage capacity have improved. These produce a huge volume of data that can only be analysed using intelligent systems based on modern techniques, such as machine learning techniques. Using data to make better decisions has been the essential challenge organisations and businesses have undertaken in the last decade. Huge technological advances have produced a successful framework for AI and ML. These frameworks enable researchers to build intelligent systems and models capable of automating tasks and analysing data to predict outcomes (Subasi, 2020; Sarkar et al., 2018). The most fundamental aspect of machine learning is the study of how to train computers to learn from data (Géron, 2019). ML is a computational approach that employs prior experience to increase performance or create exact predictions (Subasi, 2020). The past data used to train the machine learning model for prediction is referred to as experience.

2.4.1 Incremental and Batch Learning Algorithms

Machine learning algorithms can be classified as ”batch” or ”incremental” depending on whether training data is available fully or in phases. In batch learning approaches, the entire training dataset is available and accessible through the learning model. After the training phase, no additional training datasets can be added. If the amount of training data is excessive, the training process will be prolonged and time-consuming. There may be no space to keep the complete training dataset in some cases (Dulhare et al., 2020), and this kind of learning is called offline learning (Géron, 2019).

On the other hand, the training of the dataset for incremental learning algorithms can be incomplete, and new instances can be added over time, meaning that these algorithms can be used with new data following training (Dulhare et al., 2020). In addition, this kind of algorithm can use incremental learning methods to train systems on huge data sets that exceed the capacity of a single machine's main memory. The algorithm loads a subset of data, performs a training step on that subset, and continues the process until all the data has been loaded (Géron, 2019). This kind of learning is called online learning (Géron, 2019). Both training and predictions are faster with online learning than batch learning (Burlutskiy et al., 2016). However, with online learning, if the system is fed with incorrect data, the system performance will deteriorate steadily (Géron, 2019).

2.4.2 Instance-Based and Model-Based Learning

Another method to categorise machine learning systems is by their generalisability. The majority of machine learning tasks involve prediction; this means that the system must be capable of generalising new instances in the given training cases. An accurate performance metric based on training data is beneficial, but it is not enough; the ultimate goal is effective performance in new cases. A generalisation can be achieved in two ways: Instance-based learning and model-based learning are the two types of learning. The instance-based learning method is based on storing data in the past and predicting a future instance utilising this data. Since all training data must be saved in memory before prediction, it can be computationally expensive. Because it uses memorised instances rather than learned models, instance-based learning may not generalise well to new data sets (Sun et al., 2009).

When classifying new instances using instance-based methods, the nearest existing instance is used to compare each new instance to previous ones using a distance measure. Examples of instance-based classification techniques include K nearest neighbour (KNN) and case-based reasoning (CBR) classifiers, which keep all training data in pattern space and defer generalisation until a test set is provided (Han et al., 2022).

In the model-based learning method, the model uses less storage space (set of observations) and makes predictions depending on these observations. This indicates that the observed data is used to fit the model (Babaei & Bamdad, 2021; Géron, 2019). This is also called eager learning because the models are built during training and are used to classify new examples. Decision trees, neural networks, and Support Vector Machines (SVM) are examples of model-based learning.

2.4.3 Supervised and Unsupervised Learning

This section provides an overview of ML categories based on the types of supervision they receive during training. Generally, there are three major categories (Swamynathan, 2019): supervised learning, unsupervised learning, and reinforcement learning, with this section focusing on the first two as most relevant to supplier selection.

2.4.3.1 Supervised Learning

Supervised learning is a type of learning in which both the input and output variables are present, and an algorithm is used to figure out how they are related. When there is new input data, the aim is that the machine can predict the output variables for that data. This method is named because the algorithm learns from a data set and gives the answers (Géron, 2019). In this type of learning, the learner is provided with labelled data sets, and the supervisor provides information to train algorithms and learn a function from input to output (Dulhare et al., 2020). There are two types of supervised learning tasks: regression and classification.

Regression is a type of supervised learning, and it is capable of making predictions and modelling continuous variables. Linear regression is the most fundamental sort of regression. A straight line is attempted to fit the dataset, which is possible if the linearity of the correlations between the features in the datasets is maintained (Ray, 2019).

In classification tasks, the samples in the dataset are assigned to one of several categories. The data collection may contain only two categories (binary classification) or may contain three or more (multiple classifications). A category is sometimes known as a class or a label. Suppose the model is being used to solve a classification problem. It is assumed that the algorithm would learn the categories in the dataset and then link each new observation with the appropriate category (Kıran et al., 2020). Supervised machine learning algorithms were utilised by (Harikrishnakumar et al., 2019) for efficient and accurate supplier classification. Four machine learning models, namely support vector machines, logistic regression, k-nearest neighbours, and naive Bayes, were employed in this study. The performance of these algorithms was evaluated using a test dataset. The proposed approach offers a comprehensive and robust supplier assessment process by leveraging machine learning. It enhances accuracy and saves time compared to traditional multi-criteria decision-making methods, enabling the identification of the best supplier based on performance measures derived from classification algorithms.

2.4.3.2 Unsupervised Learning

Unsupervised learning is a case where there are no output variables that correlate to the input variables. Therefore, it is referred to as unsupervised learning because there are no labels for output, and the algorithm has no set of data to learn from (Géron, 2019). It is mostly employed in the clustering and feature reduction processes. The K-means algorithm, as in Figure 2.1, is one of the simplest unsupervised learning methods; it is used to solve the well-known clustering problem.

Classifying a given data set into a predetermined number of clusters follows

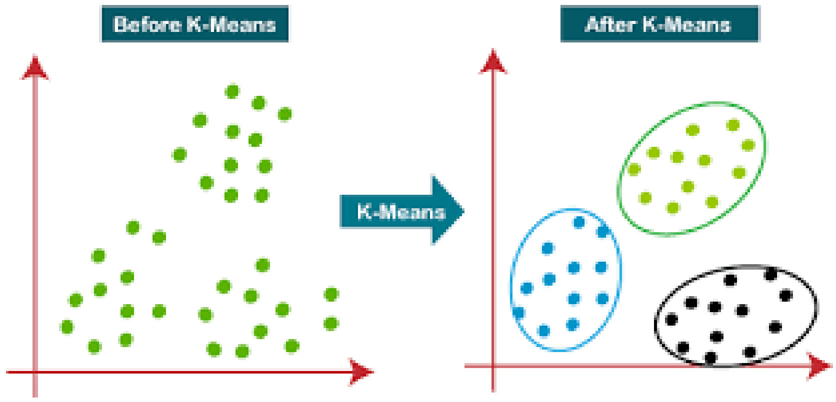

Figure 2.1: K means clustering (Mahesh, 2020)

a clear and uncomplicated procedure. The central concept is to designate k centres, one for each cluster in the data set. These centres should be established strategically because different locations yield varied outcomes. Consequently, it is preferable to situate them as far apart as is reasonably practicable (Mahesh, 2020). Unsupervised learning is relevant to the supplier selection problem as it provides the ability to group different suppliers based on common characteristics. For instance, Akman (2015) employed c-means clustering to group suppliers based on delivery time, price and quality. Then, top suppliers were evaluated using environmental/green criteria resulting in three clusters of good, medium, and bad suppliers.

2.4.4 Supervised ML Algorithms

This section discusses several types of supervised machine learning methods that have been used in relation to supplier selection.

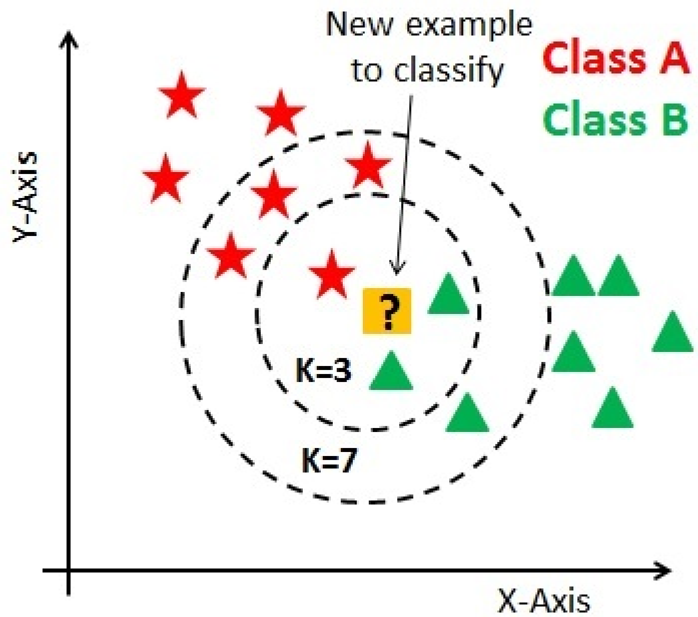

Figure 2.2: K-nearest Neighbour (Taunk et al., 2019)

2.4.4.1 K-Nearest Neighbour (KNN)

The KNN classification model is one of the earliest and most straightforward classification models. The decision rule of the nearest neighbour is to assign an unclassified sample to the closest group of previously classified samples (Cover & Hart, 1967).

As in Figure 2.2, the KNN classification algorithm should be the starting point for any classification research because it is one of the most fundamental and straightforward classification methods available. In situations in which there is scant or no previous knowledge about the data distribution, it should be one of the first approaches explored (Peterson, 2009). The letter K refers to the number of closest neighbours used in the calculation.

In the context of supplier selection, KNN classification can provide the ability to select suppliers based on how closely they match desired characteristics. For instance, Pouraghabagher and Sarfaraz (2018) used a combination of KNN with Collaborative Filtering to propose suppliers to manufacturers and buyers based on how these manufacturers have rated the suppliers.

2.4.4.2 Support Vector Machine (SVM)

SVM is a powerful ML algorithm widely used and can perform linear or nonlinear classification and regression. Therefore, SVM is well adapted for classifying complex problems with small or medium-sized data (Géron, 2019).
Simply, the SVM, as in Figure 2.3, a training dataset of n points of the form is given as: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$
A hyperplane is a subset x that satisfy the Equation 2.1 :

$$w^T x - b = 0 \tag{2.1}$$

Where w is the hyperplane's normal vector. It's like Hesse's normal form, except that w isn't a unit vector-like in Hesse's normal form.
The parameter $\frac{b}{w}$ calculates the hyperplane's normal vector w offset from the origin.

SVM is relevant to supply chain research in general, primarily due to its robust classification performance when the number of features increases. For instance, Baryannis, Dani, and Antoniou (2019) provides a comparative analysis of the capabilities of different ML algorithms for supply chain risk prediction, with SVM achieving the best performance. Similar results for supplier selection have been shown by Liou et al. (2021), where SVM was used to reduce the

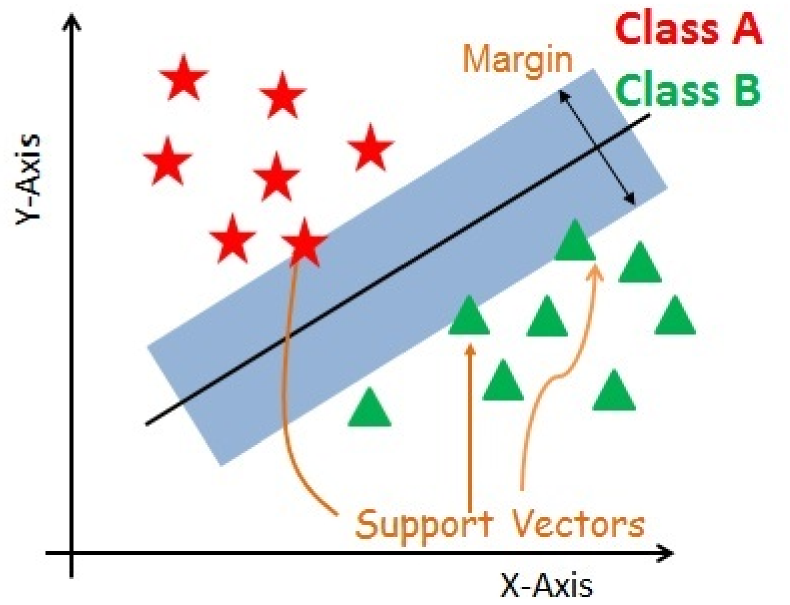

Figure 2.3: Support Vector Machine (Mahesh, 2020)

number of selection criteria based on historical data.

2.4.4.3 Decision Tree (DT)

A DT is a tree structure that is drawn out in the form of a flowchart and is composed of internal nodes that represent a feature test, branches that indicate a test result, and leaf nodes that provide a class label (Sharma & Kumar, 2016). It is possible to illustrate it using Figure 2.4.

DT, like SVMs, are multipurpose ML algorithms that can perform classification, regression, and multi-output problems. In addition, they are extremely powerful algorithms that can fit large datasets and key components of Random Forests (discussed in the following section), among the most sophisticated ML algorithms available today (Géron, 2019). The important aspect of decision tree models is that their decisions are simple to understand. Such models are

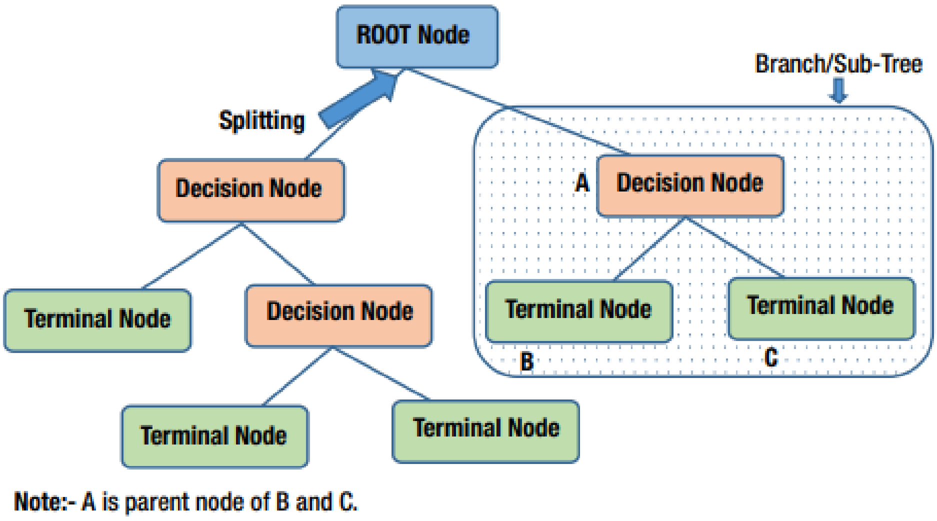

Figure 2.4: Decision Tree (Ayyadevara, 2018)

frequently referred to as white-box models. On the other hand, other models, such as Neural Networks, are commonly regarded as black-box models. They produce excellent predictions, yet, it is often difficult to explain why the predictions were made in simple terms. Decision Trees provide simple and good classification rules that can be applied straightforwardly (Géron, 2019).

In supplier selection, decision trees have been quite attractive due to their inherent interpretability, and have been utilised to derive supplier selection related decision rules that are simple and straightforward to comprehend, hence offering a transparent and rational model that procurement staff can consider adopting more easily than less interpretable methods (D. Wu, 2009).

2.4.4.4 Artificial Neural Networks (ANN)

ANNs are intelligent systems inspired by biology that simulate how the human brain processes information. ANNs are trained by identifying data patterns and correlations. They are not "trained" to learn by programming (Paul et al., 2021).

An ANN, as in Figure 2.5, consists of hundreds of artificial neurons or processing elements (PE) linked together by coefficients (weights) and organised in layers. Networked neurons determine brain computation strength. Each PE has a transfer function and one output. Transfer functions, learning rules, and design affect a neural network's behaviour. Weights are variable parameters. Neural networks are parameterised. The inputs' weights determine neuron activity. A transfer function transfers the activation signal to one neuron; the transfer function is nonlinear. During training, inter-unit connections are fine-tuned to reduce prediction error and increase network accuracy. After the network has been trained and validated, more data can be used to forecast the output (Agatonovic-Kustrin & Beresford, 2000; Abiodun et al., 2018). ANNs that use either supervised or unsupervised learning procedures include convolutional neural networks (CNNs), recurrent neural networks (RNNs), and multilayer perceptron (MLP) (Paul et al., 2021).

ANNs are now widely used in various fields to meet human needs. Because of its wide range of applications, many firms are investing in neural networks to solve problems in numerous sectors and the economic sector that normally fall under the purview of operations research; they have been widely used to help address improvement problems in a range of industries, including industrial production, petroleum exploration, and corporate management, in recent years (Abiodun et al., 2018).

In the context of supplier selection, ANNs have been used either separately

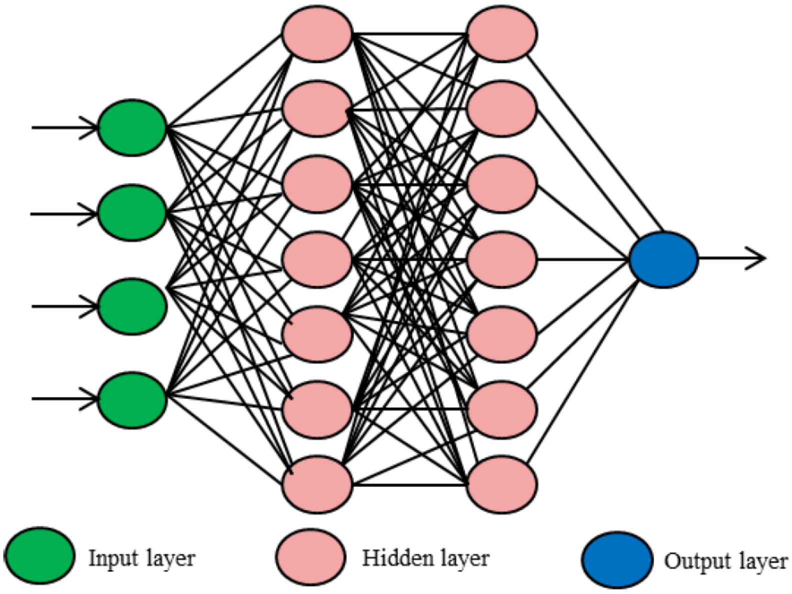

Figure 2.5: Neural Network architecture (Ayyadevara, 2018; Kukreja et al., 2016)

or in combination with multi-criteria decision making. An example of the latter is the work of Kar (2015), which employed ANN to process multiple evaluation criteria, into a one-dimensional output of suppliers that exceed or do not exceed a performance threshold.

2.4.5 The ML Pipeline

The ML process is heavily reliant on data and involves several steps, as shown in Figure 2.6:

- Data preprocessing: This involves assessing the raw data and preparing it for training purposes. If necessary, feature transformations are applied to the data to improve its quality. Then, data is divided into two sets: a larger chunk for training and a smaller portion for testing.

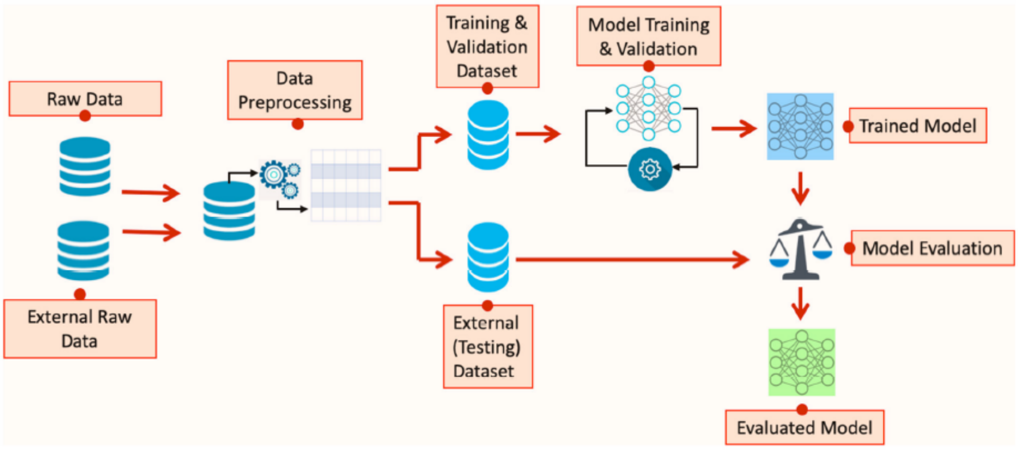

Figure 2.6: ML framework
(Abdulkareem & Petersen, 2021)

- Training: This involves creating a model or "function" and using an available dataset to train it in this stage. Internally, the function would utilise any algorithm and this dataset to train and understand patterns between input and output.
- Validation and Testing: This requires feeding the model with new data and producing an output after the model has been trained to see its accuracy.
- Evaluation: Lastly, performance metrics are used to evaluate and improve the model.

Data preprocessing is a crucial phase in the ML process since the quality of the data that can be derived from it affects the model's learning ability (García et al., 2015). For a small dataset, it is feasible to preprocess and clean the data manually. However, a fully manual process may not be possible with a large dataset, and automated preprocessing approaches may be required.

The data preprocessing stage includes data cleaning (which involves removing noise and inconsistent data), data merging (which combines several data sources into one), data transformation (in which data is transformed and consolidated into appropriate forms), and reducing the amount of data (which includes dimensionality reduction and feature engineering as well as sample selection from datasets). Some of these are explained in more detail next.

2.4.6 Data Cleaning

The most important aspect of the preprocessing task is data cleaning, which tries to improve data quality and consistency. It can be said that data cleaning is where a great deal of total effort is devoted to a data-intensive project (Géron, 2019). Several data cleaning processes have been proposed in the literature, with the seminal books by Dasu and Johnson (2003) and García et al. (2015) offering a detailed analysis of data preprocessing and cleaning. The basic objective of the data cleaning process is identifying and ratifying errors and copied data to create a reliable dataset. With such a process, the quality of the training data improves for analytics, and the decision-making process becomes accurate. The inclusion of an ML component makes the data cleaning part of data preparation more significant since the performance of ML approaches is directly dependent on the quality of the input data (Géron, 2019). For handling missing values, many techniques can be considered depending on the dataset, as listed below (Witten & Frank, 2002):

- Completion of all the missing data if only some are missing by estimating values and replacing the missing value with the mean or mode.

- Removal of the rows with no values or most data is missing.
- Filter with low variance.

After handling missing values, feature scaling should follow. Normalisation and standardisation are the two methods that see the most used when it comes to scaling numerical data in preparation for modelling. When normalisation is performed, each input variable is scaled on its own to the range 0–1. The process of standardisation scales each input variable independently by first subtracting the mean (also known as centring) and then dividing by the standard deviation. This shifts the distribution such that it has a mean of zero and a standard deviation of one.

For instance, if the value ranges of features are different, it is required to normalise these. This can be done using Equation 2.2 to keep the range of values of each feature between [0, 1].

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{2.2}$$

2.4.7 Missing Values

The steps to handle missing values from the data set include the following:

1. Ratio of Missing Values: It is unlikely that columns of data containing many missing values will have a great deal of relevant information. So data columns with a percentage of missing values that exceeds a certain threshold can be removed from a table or database.

2. Filter with a low variance: Columns of data with little variation in the data contain little information, just as they did with the previous technique. As a result, all data columns with a variance less than the specified value can be removed from consideration. Normalisation must be performed before implementing this strategy in practice because the variance varies depending on the column range.

3. The median of that feature across all samples is used to fill in the missing value of a supplier dataset feature.

4. Imputation: An imputation is a group of procedures in which an estimate of the missing value or distribution is utilised to produce predictions from a specified model. An estimate of the value also substitutes a missing value, or the probability distribution of possible missing values is calculated. For example, one frequent practice is to replace a missing value with the attribute's mean or mode (Saar-Tsechansky & Provost, 2007).

2.4.8 Feature Engineering

Following data cleaning and dealing with missing values, several optional steps are collectively referred to as feature engineering. They include dealing with redundant features, scaling features, and reducing dimensionality through feature selection.

Redundancy should be avoided where possible. This usually increases the data set size, increasing algorithm modelling time and model processing time. A feature is redundant if inherited from another, names of dimensions or features can be inconsistent, and correlation analysis can detect feature redundancy (García et al., 2015).

Feature scale transformation is one of the most significant transformations performed on a dataset. Machine learning methods do not work well with a few notable exceptions when numerical features in the input data have widely varying scales (Géron, 2019). Scaling the dataset can be accomplished in various ways, but Min-Max scaling and standardisation are the most popular. The idea of Min-Max scaling, also called normalisation, is very easy to understand: values are changed from 0 to 1, subtraction of minimum and division of maximum minus minimum. Standardisation differs from the Min-Max method in that it first subtracts the mean value before dividing by the variance to produce a distribution with unit variance (Géron, 2019).

2.4.8.1 Dimensionality Reduction

This technique aims to prepare the data for analysis that can be used with ML algorithms. As a result, this section discusses the essential preprocessing methods used for data classification.

Dimensionality reduction is a method of lowering the dimension of a feature space while keeping the most significant data and discarding the rest to minimise the processing time of a classifier and improve visualisation (Tharwat, 2016). Many issues arise while dealing with redundant and irrelevant criteria in large datasets, indicatively:

- The complexity of the model increases, and the difficulty of understanding it.

- Increase the amount of time it takes to train a model.
- The model will be insufficiently robust or confident.

Eliminating such redundant and irrelevant features speeds up the classification process and improves classifiers in making accurate decisions, particularly when working with vast amounts of heterogeneous data (Salo et al., 2019).

Feature Selection

Feature selection is used to select a subset of features from the input that may adequately represent the data while minimising the impact of noise or irrelevant features and still producing accurate predictions (Guyon & Elisseeff, 2003; Chandrashekar & Sahin, 2014). As a result, models are made simpler and easier to understand, and ML performance increases. Feature selection is illustrated in Figure 2.7.

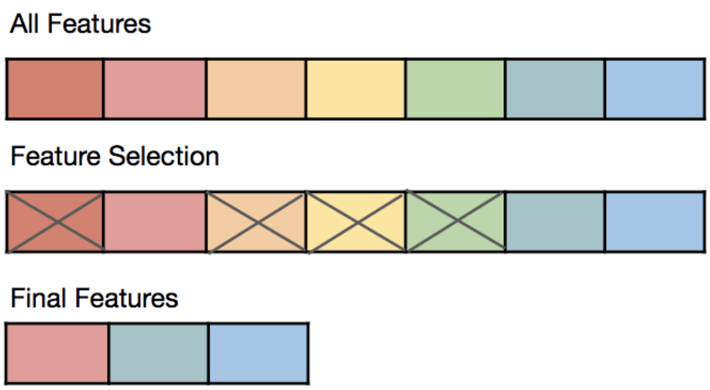

Figure 2.7: Feature selection (J. Li et al., 2017)

2.4.9 Feature Importance

Feature importance in the context of machine learning approaches refers to scoring input features depending on how valuable they are at predicting a target feature. The feature importance values indicate the degree to which a feature influences the class (Al Iqbal et al., 2012). Feature importance assessments are quite crucial when working on model prediction. There are different ways to obtain feature importance scores and from many different sources. Popular examples include decision trees and other algorithms that use the decision tree to calculate the feature importance scores. People commonly look to measures of feature importance computed from tree ensembles to understand how these models get at their predictions (X. Li et al., 2019).

Depending on the algorithm employed, feature importance calculation methods differ, as briefly analysed in the remainder of this section. Note that feature importance is not provided in the case of KNN, due to the nature of the approach to focus on proximity across features rather than focusing on individual features.

The weight of each feature in decision trees is implicitly defined by its value inside the tree structure. The ratio of the reduction in node impurity to the probability of reaching that node is used to calculate a feature's importance. The probability of a node is calculated by dividing the number of samples that reach the node by the total number of samples. Important features are commonly found closer to the root of the tree, while those that are not important will be closer to the leaves or they might not be part of the tree at all. In the case of Random Forests, determining feature importance includes averaging the depth across all the trees in the forest (Géron, 2019).

In the case of SVM, the amount of classifications that each feature contributes to the final model determines the weight assigned to each feature in the SVM. Generally, more weight is given to properties with larger coefficients (Lin et al., 2017).

In ANN, the weights of each feature are determined by the coefficients of each node in the hidden layers. Backpropagation is used to update weight value parameters during model construction internally and may be computationally expensive due to its scaling sensitivity and ability to tune hyperparameters such as hidden layers, neurons, and iterations (Sarker, 2021). The weights of each feature can be found for ANNs with many hidden layers by analysing the weights of the connections between the input layer and the hidden layers, as well as the weights of the connections between the hidden levels and the output layer.

2.4.10 Evaluation Metrics

Precision, recall, and the F-measure are the three metrics that are typically utilised in evaluating a classifier's effectiveness.

2.4.10.1 Precision

Precision (Specificity) is defined as the classifier's ability to unlabelled a negative into an instance positive. It is expressed as the ratio of true positives to the sum of true and false positives for each class of problems. It can be asserted in another way, "for all cases that were classified as positive, what was the correct percentage?" When predicting a positive value, precision is defined as the frequency with which the prediction is correct. The true positive rate refers to the rate wherein the model always predicts a positive value. A high rate is an appropriate measure for the model. False Positive rate can be defined as the rate wherein the wrong positive value is always predicted by the model as in Eq. 2.3. The lower rate is an adequate measure for the model.

$$Precision = \frac{TP}{(TP + FP)} \tag{2.3}$$

2.4.10.2 Recall

Recall (Sensitivity) refers to the classifier's ability to find any and all positive cases. It is expressed as the ratio of true positives to the sum of true positives and false negatives for each class of problems as in Eq. 2.4. It can be referred to in another way, "for all actually positive values, what was the percentage that is classified correctly?"

$$Recall = \frac{TP}{(TP + FN)} \tag{2.4}$$

2.4.10.3 F1 Score

The $F1$ score is calculated by taking the harmonic mean of a classifier's precision and recall and combining them into a single number. It is mostly used to compare the classification results of two distinct classifiers. The $F1$ score is defined as a means of the precision of weighted harmonic and recall, such that 1 is the best score, whereas the worst score is 0. $F1$ scores are often low when compared to accuracy metrics since precision is built into them and then recall into their computation as in Eq. 2.5. As a general rule, when comparing classifier models, a weighted average of $F1$ should be employed instead of global accuracy.

$$F1 - score = \frac{2 * (recall * precision)}{(recall + precision)} \tag{2.5}$$

2.5 Multi-Criteria Decision Making (MCDM)

Decision-making is rather simple when dealing with single-criterion problems because we only need to select the option with the highest preference rating. However, while evaluating alternatives using various criteria, concerns such as criterion weights, preference reliance, and criteria conflicts appear to exacerbate the problems, necessitating more advanced approaches. The stages of evaluating the alternatives are as follows (Tzeng & Huang, 2011):

- Determining the issues: The first step in dealing with multiple criteria decision-making (MCDM) difficulties is figuring out how many criteria are involved and deciphering the problem's path.
- Establishing the preferences: The second step is to collect the necessary data or information to represent and consider DM's choices properly.
- Evaluate the alternatives: To ensure the goal is realised, more work will be done to develop a range of possible alternatives.

- Identifying and selecting the appropriate alternatives: The final stage is to choose an appropriate strategy to assist in evaluating or improving the potential alternatives.

In essence, the goal of MCDM methods is to find the optimum alternative according to many criteria. Each criterion is assigned a numerical value concerning each alternative. One of the drawbacks of MCDM methods is the dimensionality issue. The issue of dimensionality occurs when the number of dimensions surpasses the capacity, which means that the computing cost skyrockets. Many evolution algorithms have recently been proposed to solve this problem, including genetic algorithms, genetic programming, and evolution strategy (Tzeng & Huang, 2011).

2.5.1 Analytic Hierarchy Process (AHP)

The AHP methodology was presented by Saaty (Saaty, 1980) and is a commonly utilised decision-making method owing to its simplicity of use and attractive mathematical features. For example, AHP visualises problem-solving in the hierarchy, which helps decision-makers organise the problem in a way that's more efficient in terms of how it displays and effective in how it functions. AHP involves a hierarchy as in Figure 2.8, which shows how an AHP analyses a problem using performance evaluation of alternatives.

Essentially, AHP is based on three fundamental concepts: organising a complex decision as a hierarchical structure of goals, criteria, and alternatives; comparing the criteria at each level of the hierarchy with each criterion at the previous level; last but not least, vertically synthesising the judgements across

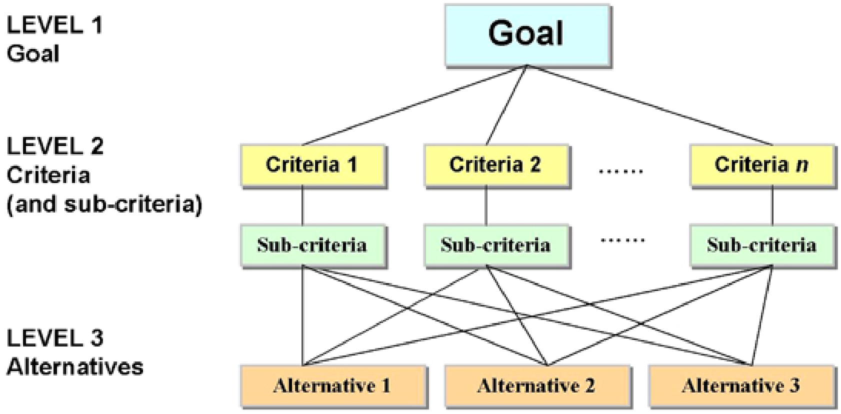

Figure 2.8: AHP hierarchy (Saaty, 1980)

the hierarchy's numerous tiers (also known as vertical synthesising).

For supplier selection problems, AHP has been employed by numerous researchers, indicatively Çalık (2021) and Agrawal and Kant (2020). This is primarily due to its simplicity and adaptability in assisting decision-makers to make precise judgements. However, when the amount of criteria increases, the number of pair comparisons increases, making the process of applying AHP more time-consuming (Oguztimur, 2011; Sakhardande & Gaonkar, 2022). Realistically, efficient decision making with AHP requires the number of criteria to be limited to 7 on average, with 5-9 being the most common range (Shih et al., 2007). This is quite restrictive in more complex situations, making AHP alone an unsuitable solution.

2.5.2 Full Consistency Method (FUCOM)

FUCOM is a novel model that employs the pairwise comparison concept, validates outcomes via divergence from maximal consistency, and overcomes the shortcomings of AHP. Compared to the AHP, FUCOM requires fewer pairwise comparisons (only n-1 comparisons) (Pamučar et al., 2018). It is adaptable in utilising a variety of scales according to the expert's preference. The FUCOM model, like any other subjective approach to establishing criterion weights, is susceptible to the final values of the criterion weights. This could be explained by decision-makers who utilise the FUCOM model to order the criteria based on their personal preferences before performing pairwise comparisons. FUCOM, on the other hand, produces the fewest deviations from ideal values for the calculated weights of the criterion compared to other approaches (Pamucar et al., 2018).

In comparison to other models that use pairwise comparisons, one of the advantages of the FUCOM technique is that it avoids the problem of redundancy in paired criteria comparisons. The FUCOM method has been applied to a range of problems, including airline ranking (Badi & Abdulshahed, 2019) and rail transportation's service quality (Prentkovskis et al., 2018), as well as to evaluate the sustainability criteria in supplier selection (Durmić, 2019; Durmić et al., 2020).

Although FUCOM minimises the number of pairwise comparisons compared to related and popular MCDM methods such as AHP, it is only useful for calculating the weight coefficients of the criteria, as it has not been designed to focus on ranking alternatives (Pamučar et al., 2018).

The FUCOM algorithm can be summarised as follows:

Input: Expert criterion comparison.

Output: The criteria weights

Step 1: Ranking of criteria by the experts.

Step 2: Identifying the relative importance of evaluation criteria.

Step 3: Defining non-linear optimisation model constraints.

Restriction 1: The ratio of the criteria weights is equal to the relative importance of the criterion observed.

Restriction 2: The weight coefficients' values should meet the mathematical transitivity criteria.

Step 4: Develop a model for evaluation criterion weight coefficients.

Step 5: Determining the final evaluation criteria and subcriteria values.

2.5.3 The Technique for Order Preferences by Similarity to an Ideal Solution (TOPSIS)

TOPSIS is an MCDM method originally developed by Hwang and Yoon (1981). TOPSIS is based on the compromise solution concept, which involves selecting the greatest alternative closest to the positive ideal solution (optimal solution) and farthest from the negative ideal solution (inferior solution). An overall index that is based on the separations from the ideal solutions is used to rank the alternatives (Mukherjee, 2017).

The TOPSIS method is broken down into several steps, which are briefly described below (Mukherjee, 2017):

- Make a decision matrix that is normalised.
- Prepare a weighted normalised decision matrix using formula 2.6.
- Differentiate between a positive and a negative ideal solution.

- Calculate the separation distance to be used.
- Calculate the coefficient of relative closeness.
- Sort the alternatives in descending order.

One of TOPSIS's most significant advantages is its ability to find the optimal alternative rapidly (Parkan & Wu, 1997). In addition, as with other MCDM methods, the normalisation approach for TOPSIS can be reduced to a linear transformation (Saghafian & Hejazi, 2005).

TOPSIS formulas can be summarised as follows:

A list of n criteria, $C = \{C_j \mid j = 1, \ldots, n\}$ and a list of m alternatives, $A = \{A_i \mid i = 1, \ldots, m\}$, are given in matrices, and $X = \{x_{ij} \mid i = 1, \ldots, m; j = 1, \ldots, n\}$ indicates a group of performance evaluations whereas the weights of the criteria are denoted as $W = \{w_1, w_2, \ldots, w_n\}$.

The decision matrix incorporates evaluation elements and is the beginning point for TOPSIS analysis. For example, the decision matrix (X_{ij}) is written as follows, where n is the number of assessment criteria, and m denotes the number of decision points:

$$\mathbf{X_{ij}} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \ldots & x_{1n-1} & x_{1n} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \\ \ldots & \ldots & \ldots & \ldots & \ldots & \\ x_{m1} & x_{m2} & x_{m3} & \ldots & x_{mn-1} & x_{mn} \\ & & & & & \end{bmatrix}$$

The following are the elements of a normalised decision matrix 2.6:

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \qquad i = 1, \ldots, m; j = 1, \ldots, n. \tag{2.6}$$

$$\mathbf{R_{ij}} = \begin{bmatrix} r_{11} & r_{12} & \ldots & r_{1n} \\ r_{21} & r_{22} & \ldots & r_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ r_{m1} & r_{m2} & \ldots & r_{mn} \\ \end{bmatrix}$$

The next step is to multiply the normalised decision matrix by each of the criteria weights to create a weighted normalized decision matrix as follows:

$$\mathbf{V_{ij}} = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \ldots & w_m r_{1n} \\ w_1 r_{21} & w_2 r_{22} & \ldots & w_m r_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ w_1 r_{m1} & w_2 r_{m2} & \ldots & w_n r_{mn} \end{bmatrix}$$

where $v_{ij} = w_i * r_{ij}$, $i = 1, \ldots, m,\ j = 1, \ldots, n.$

In the ideal solution set (E+), the best values of the evaluation criteria are used to determine the solution set, which is represented as:

$$E^{+} = \{v_1^*, v_2^*, ...v_n^*\} = (\{\max_j v_{ij} | i \in I \quad , (\min_j v_{ij} | i \in I'\} \quad i = 1, \ldots, m; j = 1, \ldots, n. \tag{2.7}$$

In the negative solution set (E-), the worst values of the evaluation criteria are used to determine the solution set, which is represented as:

$$E^{-} = \{v_1^-, v_2^-, ...v_n^-\} = \{(\min_j v_{ij} | i \in I \quad , (\max_j v_{ij} \quad | i \in I'\} \, i = 1, \ldots, m; j = 1, \ldots, n. \tag{2.8}$$

I represents the positive criteria, whereas I' represents the negative criteria.

The next step is to calculate the distances as follows:

$$S_i^+ = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^*)^2}, \quad i = 1, \ldots, m; j = 1, \ldots, n. \tag{2.9}$$

$$S_i^- = \sqrt{\sum_{j=1}^{n} (v_{ij} - v_j^-)^2}, i = 1, \ldots, m; j = 1, \ldots, n. \tag{2.10}$$

Following that, the distances between each value and the best value were calculated using Eq.2.11.

$$TOPSIS = \frac{S_i^-}{S_i^- + S_i^+} \qquad i = 1, \ldots, m \tag{2.11}$$

The alternatives are then ordered from best to worst; the greater the score, the higher the ranking. The ideal choice and the solution to the problem are at the top of the list.

In the context of supplier selection, TOPSIS has been applied to solve supplier selection by several researchers, such as Shukla et al. (2014), Kannan et al. (2014), Chatterjee and Stević (2019) and Ghamari et al. (2021). Despite TOPSIS being a practical and useful technique for ranking and selecting several alternatives based on distance measurements, the key drawbacks of TOPSIS are the absence of weight elicitation for criteria and the lack of consistency checking for assessments (Shih et al., 2007).

2.5.4 Summary of MCDM Methods

In summary, MCDM methods benefit from the following advantages:

- Explainability: MCDM approaches provide a clear and transparent explanation of the decision-making process, allowing stakeholders to comprehend the rationale behind picking a specific supplier. This makes MCDM approaches more suitable for situations where transparency and accountability are crucial, such as public procurement when these qualities are crucial.

- Expertise of Humans: MCDM approaches rely on the expertise of human decision-makers, who can use their knowledge and experience to make informed decisions. This can be especially useful when the data is intricate or difficult to analyse.

- Simplicity: MCDM procedures can be uncomplicated and simple, making them simple for firms with limited resources or expertise to execute and comprehend.

- Adaptability: MCDM approaches can be adapted to many sorts of challenges, and the evaluation criteria for suppliers can be modified to the company's particular requirements.

A summary of the MCDM methods employed in this thesis and their advantages and disadvantages follows.

- AHP is a well-established and extensively used method in MCDM, although it can be subjective and time-consuming because it relies on expert judgement to assign weights to criteria.
- TOPSIS compares alternatives with an ideal solution and a negative perfect solution using Euclidean distances. TOPSIS is straightforward to comprehend. However, it may not be appropriate for complex decision-making situations involving multiple criteria.
- FUCOM is a novel technique for evaluating criteria following a semi-objective / objective evaluation method that lowers the number of comparisons between criteria. Although FUCOM is faster than other MCDM methods like AHP, it remains a subjective method and is time-consuming when there are large or complex data sets.

ML can be employed to address some of these shortcomings by reducing subjectivity and time required and potentially producing more accurate results. In addition, employing ML models allows for the creation of updatable and adjustable supplier selection approaches as new data become available, further increasing performance.

2.6 Explainability vs Interpretability

ML, as part of AI, excels in efficiency, scalability, and precision, but this comes at the expense of explainability (Papadakis et al., 2022). There is not always a clear reasoning path from data to conclusions, but it is readily available in the traditional analysis of supplier selection utilising MCDM approaches.

In terms of supplier selection and evaluation, today, it is indisputable that AI has the ability to drive digital supply chain transformation. However, its ability to solve more complicated supply chain management (SCM) issues is still ambiguous, in part because of the black box problem that AI faces in literature and practice (Mugurusi & Oluka, 2021).

Although explainability and interpretability are frequently used interchangeably, there may be a difference between the two at this point. Interpretability is more often considered as a narrower term, focusing specifically on the interpretation of machine learning model outputs, or "the degree to which an observer can understand the cause of a decision" of a model as Biran and Cotton (2017) define it. Explainability, however, has much broader coverage, focusing on the ability to derive explanations that may not only be based on machine learning model interpretation but also include expert knowledge and end-to-end reasoning pathways.

Supplier selection approaches based purely on ML can, at best, afford a level of interpretability, such as the approaches used in the conducted experiments. For instance, following the path of a decision tree from root to leaf can provide an interpretation of which criteria and which threshold values contributed to the result of selecting or not selecting a particular supplier. The proposed framework goes one step further than this, as it combines interpretation in relation to feature importance and selection with an explanation of selecting a particular supplier based on the rank they achieve as a result of applying an MCDM approach.

Based on this discussion and in accordance with relevant literature (Antoniou et al., 2022), this thesis interpretability refers exclusively to ML models, while explainability is reserved for hybrid models that go beyond ML interpretability by including a MCDM component.

2.7 Supplier Selection using ML and MCDM

There are various methods of supplier selection accessible in the literature. Supplier selection methods are divided into five categories by Z. Zhang et al. (2003): Linear weighting methods, mathematical programming methods, statistical techniques, artificial intelligence algorithms, and cost-based models.

Researchers facing the selection of a supplier in this kind of MCDM problem where there are conflicts between quantitative and qualitative criteria, there will be trade-offs between these criteria must be considered, with no single best way of selecting and evaluating suppliers (Abdolshah, 2013). Thus, purchasing managers should employ appropriate decision-making strategies based on their circumstances. On the other hand, the hybrid multi-method is a new trend in MCDM applications which combines the advantages of conventional and newer methods (Ortiz-Barrios et al., 2020).

Various scholars have examined the literature on the use of the MCDM approach published between 2008 and 2012 and discovered that the AHP was the most popular method to solve supplier selection problems (Kannan et al., 2014; Gegovska et al., 2020). de Brito and Evers (2016) discovered that the most widely used MCDM was the AHP approach after reviewing 128 papers using the method. However, they recommend that using other MCDM methods should be investigated.

According to Nodeh et al. (2019) and Mukherjee (2017), there are two types of supplier selection approaches, single model and hybrid model methods, as shown in Figure 2.9. Single methods exclusively use either MCDM or AI techniques while hybrids use one or more MCDM methods combined with an AI component. The next chapter provides an analysis of all these hybrid supplier selection approaches.

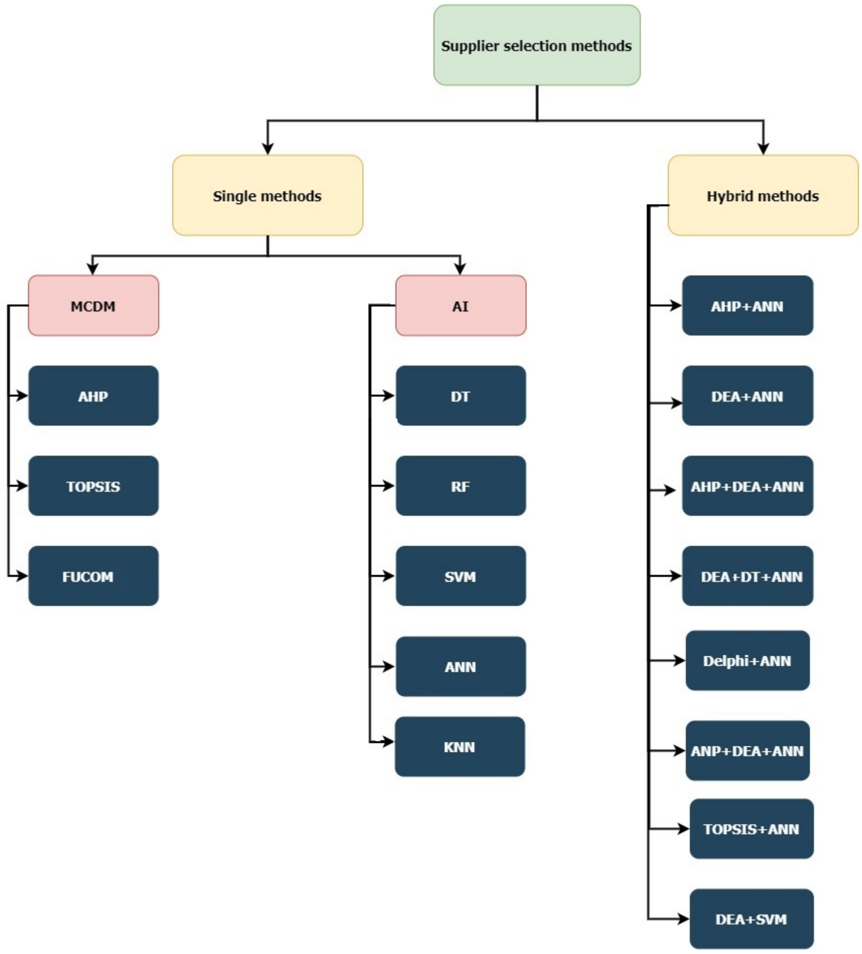

Figure 2.9: Supplier Selection Methods

2.8 Summary

This chapter introduced the concept of machine learning and its various forms used in recent studies. Additionally, it examined multi-criteria decision-making and the methodologies associated with it. Finally, this chapter discussed the process and strategies for selecting suppliers.

Chapter 3

Literature Review

3.1 Introduction

This chapter presents work related to the supplier selection and evaluation problem that MCDM methods have solved and related machine learning algorithms and methods for their integration. Existing state-of-the-art solutions for supplier selection issues are evaluated. Finally, an overview and critical evaluation of recently published work on supplier selection are provided.

3.2 MCDM Approaches

Researchers use various methods to evaluate suppliers, including such MCDM methods as AHP, TOPSIS, mathematical programming, and integrated methods (Ho et al., 2010). Many of these have been discussed as a possible means of developing analytical approaches to evaluate selection problems (De Boer et al., 2001; Partovi et al., 1990; Weber et al., 1991). In this section, some of these approaches are indicatively summarised.

The most popular MCDM methods between 2008 and 2012 for supplier selection, in terms of publications, were AHP, ANP, linear programming, and DEA (Kannan et al., 2014). In addition, today, the AHP is the most widely utilised approach method (Gegovska et al., 2020).

Unfortunately, in many cases, purchasing managers assess suppliers using personally derived criteria, which makes the decision-making process more complicated by introducing subjective factors (Akarte et al., 2001).

For evaluating suppliers in Indian automobile parts manufacturing companies, Shukla and others (Shukla et al., 2014) proposed a method that integrated fuzzy AHP with fuzzy TOPSIS. The criteria weights were calculated using the fuzzy AHP, and viable suppliers were ranked using fuzzy TOPSIS.

The difficulty of obtaining the weights for each criterion through objective or subjective techniques was considered using MCDM but this integrated approach was applied to a small dataset with only three suppliers and four criteria.

For green supplier evaluation, (Kannan et al., 2014) utilised the TOPSIS approach with fuzzy logic for Brazilian electronics companies (Kannan et al., 2014). Wang Chen and his colleagues also evaluated suppliers according to environmental criteria using fuzzy logic with both AHP and TOPSIS (Wang Chen et al., 2016).

A framework was provided to optimise supplier selection by considering sustainability aspects in another integrated method that combined fuzzy AHP and VIKOR techniques. Once again the criteria weights were created using fuzzy AHP, but the alternatives were ranked using fuzzy VIKOR (Awasthi et al., 2018).

The same integrated approach was used again for a manufacturing system (Chatterjee & Stević, 2019), combining fuzzy AHP and fuzzy TOPSIS. As previously, the criteria weights were calculated using the fuzzy AHP approach, and viable suppliers were ranked using the fuzzy TOPSIS method. To calculate the criteria weights, the FUCOM approach was used to assess the relative importance of the criteria used to select sustainable suppliers for the lime industry (Durmić, 2019).

Another integrated method was proposed by Ghamari et al. (2021), combining BWM with TOPSIS. The BWM was used to calculate criteria weights, while TOPSIS was used to rank the suppliers. This study's objective was to address the issue of selecting a sustainable and resilient supplier for a steel manufacturing company.

Recently, Shang et al. (2022) presented a method that combines the Best Worst Method (BWM) which is a subjective method with the fuzzy Shannon Entropy Method (FEM) (objective method) to obtain criteria weights for suppliers and the suppliers are ranked using the fuzzy (Multi-Objective Optimization on the basis of a Ratio Analysis (MULTIMOORA) approach.

Table 3.1 provides a concise summary of the aforementioned approaches, pointing out their strengths and weaknesses. The majority of these techniques are

Reference	MCDM	Case Study	# suppliers	# criteria	Strengths	Weaknesses
(Shukla et al., 2014)	Fuzzy AHP and TOPSIS	Automobile company	3	4	Fuzzy numbers handling and interpretability	Small dataset
(Kannan et al., 2014)	Fuzzy TOPSIS	Brazilian electronics companies	12	17	Dealing with green suppliers	Small dataset
(Wang Chen et al., 2016)	Fuzzy AHP and TOPSIS	Taiwanese manufacturing	4	5	Simple and effective computation	Small dataset
(Awasthi et al., 2018)	Fuzzy AHP and VIKOR	Automotive	3	25	All encompassing criteria	No quantitative verification
(Chatterjee & Stević, 2019)	fuzzy AHP and fuzzy TOPSIS	manufacturing system	5	9	Handles both qualitative and quantitative criteria	Small dataset
(Durmić, 2019)	FUCOM	lime industry	Non	7	Improved performance with FUCOM	Requires additional MCDM for ranking
(Ghamari et al., 2021)	BWM with TOPSIS	Steel manufacturing company	4	14	Includes sustainability criteria	Small dataset
(Shang et al., 2022)	BWM, Fuzzy FEM+ MULTIMOORA	International forklift truck manufacturer	4	17	Includes sustainability criteria	Small dataset

Table 3.1: MCDM approaches to supplier selection.

expert-based or pairwise comparison-based. The main weakness of these technologies is their inability to handle large datasets, which explains why most of these have been validated using only small datasets. As a result, researchers have highlighted the application of ML techniques to the management and evaluation of suppliers (Tirkolaee et al., 2021).

3.3 ML Approaches

The primary advantage of utilising ML prediction methods over traditional prediction methods is that ML approaches may create more accurate predictions and can also deal with complicated and inaccurate data (Islam et al., 2021). In order to select the best suppliers, machine learning algorithms can be used to analyse data on supplier performance and find patterns and trends. For instance, suppliers can be grouped based on their similarities using clustering techniques, and decision trees can be used to determine the most important features when choosing suppliers. Several researchers have considered ML approaches, some of which are presented next.

Golmohammadi et al. (2009), presented different methods combining a genetic algorithm with an artificial neural network, GA-ANN, to solve supplier selection problems. GA was used to obtain the first weights and network architecture to provide an improved search method for training. The ANN used data history to choose the best supplier.

In the same year, (Guo et al., 2009), presented what they termed a hierarchical potential SVM using a hierarchical scheme of features and paired it with a decision tree to address the supplier selection problem, including feature selection and multi-class classification. The study showed that the proposed method had better generalisation performance and used less computational time than a regular SVM.

Güneri et al. (2011), suggested the Adaptive Neuro-Fuzzy Inference System (ANFIS) method to select important criteria to evaluate suppliers and manufacturers of textiles. The results obtained using ANFIS were compared with results from multiple regression analysis, and the ANFIS method achieved more accuracy (Güneri et al., 2011).

More recently, Harikrishnakumar et al. (2019), proposed an efficient model based on supervised ML, models such as SVM, Naïve Bayes, and other classifiers for evaluating suppliers to global petrochemical companies. Generally, the model's accuracy was at least 15% higher than existing methods such as MCDM, with SVM providing the highest accuracy of 87%. But the model lacked dataset size and was applied only to a small dataset (Harikrishnakumar et al., 2019).

In 2021, Sasaki and Sakata (Sasaki & Sakata, 2021), proposed a new method to interpret the ML models. To understand the model, this study uses conditional probability based on Bayes' theorem, the Bayesian Network (BN) model, as a prediction model for suppliers' selection in the Northeast region of Japan. This study compares three ML models, RF, SVM and Logistic Regression, to select the best model. RF had an F1 score of 80% and was the top ML model in terms of prediction performance.

Table 3.2 presents a summary of the works reviewed in this section.

Reference	ML	Case Study	# suppliers	# criteria	Strengths	Weaknesses
(Golmohammadi et al., 2009)	GA+ANN	automotive industry	57	5	Performance ranking	Black box
(Guo et al., 2009)	DT, SVM	Chinese logistics examples	64	30	High accuracy	Verification requires a different dataset
(Güneri et al., 2011)	ANFIS	manufacturer of textiles	6	5	Higher accuracy with ANFIS	Small dataset
(Harikrishnakumar et al., 2019)	SVM	petrochemical companies	350	7	15% more accuracy improvement over other models	Black box
(Sasaki & Sakata, 2021)	BN, RF,SVM and LR	Northeast region of Japan	3	6	RF scored 80% F1	Small dataset

Table 3.2: ML approaches to supplier selection.

3.4 Hybrid Approaches Combining ML and MCDM

MCDM methods such as AHP, TOPSIS, DEA, and FUCOM have been utilised in supplier selection studies due to the difficulties of these approaches when the increasing size of data; however, the application of ML techniques for obtaining accurate findings and enhancing performance is on the rise. MCDM and ML integration provides new possibilities for making difficult judgments in dynamic contexts(Ganesh & Kalpana, 2022). In the context of this thesis, we refer to such approaches combining ML and MCDM as hybrid approaches. According to Gegovska et al. (2020), a hybrid method should be more accurate than a single method, producing a superior result and; therefore, researchers should focus on solving such problems by integrated methods. A critical analysis of such hybrid methods follows in chronological order.

Celebi and Bayraktar (2008), described one such technique in a hybrid model and presented supplier selection to an integrated system comprising ANNs and DEA. This integrated method's main advantage is dealing with incomplete sets of assessment criteria.

Another study combined the analytical method with ANN (Ha & Krishnan, 2008). AHP was integrated with DEA and combined with the ANN approach to select suppliers for a company manufacturing car spare parts. As previously, the evaluation of qualitative criteria was carried out by the AHP. The DEA examined quantitative criteria, and the ANN assessed efficient suppliers.

In the same year, Bottani and Rizzi (2008) improved the AHP and cluster analysis integration in a project for the beverage industry based in Italy to reduce the number of suppliers and select the best supplier possible. To overcome the problem of inconsistencies that invariably arise when there are a large number of alternatives, an approach was developed that combined cluster analysis to reduce the dimensions of the problem with MCDM methods.

Also, D. Wu (2009) published a study in which he combined DEA, DT, and NN in selecting a supplier. First, the DEA model divided suppliers into groups based on their efficiency scores; then, the Decision Tree and NN method trained the data. Finally, the model estimated and evaluated the criteria related to the suppliers. Additionally, it gave classification and regression values of an accurate level. Although this project possessed important features, it was the first hybrid to incorporate DEA, DT and NN, but the results need to be compared to real and large data sets.

A year later, Kumar and Roy (2010), demonstrated a hybrid methodology that combined AHP with NN to evaluate supplier performance. AHP compared the criteria in pairs for all suppliers. The AHP results were used as inputs to the NN model for selecting suitable suppliers; the results give the best supplier and a score suitable for comparing each supplier's relative performance. This hybrid model was able to handle the complexity and criteria of the supplier selection problem. But for validation, the results obtained need to be compared with other methods and real test cases.

Regarding environmental criteria, R. J. Kuo et al. (2010), developed an integrated model for selecting environmentally friendly suppliers, integrating DEA and ANP with ANN. This model, named the ANN-MADA hybrid method, considered both environmental and traditional criteria when evaluating and selecting a supplier. Furthermore, it overcame the constraints associated with the DEA model, with the missing value found by the ANN-MADA method. Also in 2010, the fuzzy Delphi hybrid method was proposed to identify criteria and particle swarm optimisation (PSO) was integrated with a fuzzy NN to deal with qualitative data (R. Kuo et al., 2010).

In 2013, two projects building on previous work focused on integrating AHP with ANN to select and evaluate suppliers. As previously, AHP was used to evaluate criteria, and ANN to determine suitable suppliers. The first model was applied to four suppliers in a shoe manufacturing enterprise (S. Tang et al., 2013). The second model was applied to seven suppliers in a manufacturing company specialising in transformers (Lakshmanpriya et al., 2013). The disadvantage of these hybrids includes, apart from employing black box ANN models that are difficult to interpret, the fact that experiments only included small datasets for evaluation and validation.

Similar AHP/ANN hybrids were further developed by Kar (2015) for a steel production company. Fuzzy AHP was used to evaluate and weigh the criteria according to the opinions of expert decision-makers, and fuzzy NN was used to select suitable suppliers. The main disadvantage of this hybrid is the inability of the ANN component of the hybrid to determine the efficiency score based on input/output variables. The relationship between the independent and dependent criteria is ambiguous, which makes it difficult for decision-makers to quantify the impact of each criterion (independent variable) on the outcome.

Fallahpour et al. (2016) proposed a different method for selecting suitable suppliers, which integrated DEA with Genetic Programming (GP) methods for a firm of clothing manufacturers that specialised in women's underwear and claimed to have overcome the limitations of DEA models. The GP technique compensated for the DEA model's time-consuming elements and complicated computations. Introducing GP offered a promising alternative to compensate for the black box issues associated with the use of powerful AI techniques, such as ANN, for supplier evaluation. This model could compute the suppliers' environmental efficiency, which could then be used as a regression function. Experts were employed to select criteria which included cost, delivery time, environmental protection, material quality, resource usage, and service, which were applied to the data set by this model. The main issue with goal programming is that decision makers must declare goals and priorities a priori (Wey & Wu, 2007); however, the proposed method is capable of integrating numerous objectives in order to find the best solution.

Fallahpour et al. (2018) developed a new model which integrated DEA and SVM to select suitable suppliers for spinning and weaving factories using data collected from an Iranian textile factory. The company produced yarn mixed with cotton, cotton and polyester. The efficiency scores were calculated using DEA, while SVM was used to decide the best suppliers. Afterwards, the data of performance suppliers were used to train by the SVM-DEA. The advantage of this model is that it may be utilised as a regression or classification method. Also, DEA-SVM was demonstrated to be a very powerful and versatile tool for selecting suppliers. But, SVM is considered a black box model.

A novel approach to improving suppliers' selection in pharmaceutical companies was achieved by integrating PCA with TOPSIS (Forghani et al., 2018). Fewer criteria were applied to the suppliers by PCA. The importance of suppliers with each product was obtained via the TOPSIS method. The outputs of PCA and TOPSIS were then used as inputs into a Mixed Integer Linear Programming model to identify the suppliers and the number of items they deliver. This method outperformed the standard TOPSIS method in terms of ranking but was used on a small dataset, hence requiring further validation on larger datasets.

Maghsoodi et al. (2018) employed k-means clustering to group the suppliers and MULTIMOORA as the MCDM method at the same time. Experiments showed that their system was adaptable and useful and could handle large volumes of data. However, the criteria weights need to be recalculated using methods such as the analytic network process, the analytic hierarchy process, and the Best-Worst-Method, as mentioned by the authors.

Liou et al. (2019) demonstrated how to execute data-driven MCDM on green supplier selection problems using massive historical data. They investigated the link between qualities using a random forest technique. They then handled attribute weights using a combination of DEMATEL and an analytical network process (ANP). The evaluation of suppliers by determining the gap between ideal and present suppliers was then determined using multi-objective optimization and ratio analysis. Finally, they reviewed a case study of a green supplier selection process utilised by a Taiwanese electronics firm to demonstrate the efficiency of the proposed approach. But, this hybrid was applied only to three suppliers and used the random forest as a ML model, considered one of the black-box models.

Gegovska et al. (2020) compared five suppliers based on seven criteria using fuzzy MCDM (AHP, TOPSIS, and ELECTRE) along with an ANN. The comparison results showed that the ANN performance was superior to that of the fuzzy MCDM. Apart from depending on a black box model, an additional limitation is the problem of assigning orders after ranking the suppliers.

Cheng et al. (2020) presented an integrated method in which two strategies for evaluating suppliers were combined in the model's design. Instead of using manual labelling, one of the MCDM methods was used to get each supplier's label, and support vector regression (SVR) was used to train the integrated model. Compared to previous models, simulation results suggested that the proposed intelligent model has greater accuracy and robustness. The drawback of this model was that it was validated using a small dataset, and the results were not interpretable by the decision-makers in the company. Finally, Alavi et al. (2021) introduced a hybrid supplier selection approach combining the fuzzy best-worst method (BWM) to calculate the weights of the criteria and evaluate and select the best suppliers utilising a fuzzy inference system (FIS). Table 3.3 concisely summarises the aforementioned approaches, pointing out their strengths and weaknesses.

Reference	MCDM	AI	Case Study	# suppliers	# criteria	Strengths	Weaknesses
(Celebi & Bayraktar, 2008)	DEA	ANN	Automotive	20	4	Supports incomplete data	Small dataset - Black box
(Ha & Krishnan, 2008)	AHP, DEA	ANN	Automotive	27	12	High accuracy	Small dataset - Black box
(D. Wu, 2009)	DEA	Decision Tree, ANN	Telecomm.	11	6	High classification performance	Small dataset
(Golmohammadi et al., 2009)	Decision Matrix	ANN, GA	Automotive	57	6	Decision matrix drawn from managers' insights	Black box
(R. Kuo et al., 2010)	Delphi (Fuzzy)	ANN (Fuzzy), PSO	Electronics	17	10	Low error values	Supports only quantitative criteria
(R. J. Kuo et al., 2010)	ANP, DEA	ANN	Electronics	12	6	Supports incomplete data	No validation
(Kumar & Roy, 2010)	AHP	ANN	Electronics	7	4	Focus on green suppliers	Small dataset - Black box
(Lakshmanpriya et al., 2013)	AHP	ANN	Manufacturing	7	10	Improvement over AHP	Small dataset - Black box
(S. Tang et al., 2013)	AHP	ANN	Manufacturing	4	6	High prediction accuracy	Small dataset - Black box
(Kar, 2015)	AHP (Fuzzy)	ANN (Fuzzy)	Iron & Steel	45	7	Low error on average	Black box, no validation
(Fallahpour et al., 2016)	DEA	GP	Textile	100	6	Focuses on environmental performance	Goals and priorities need to be defined a priori
(Fallahpour et al., 2018)	DEA	SVM	Textile	48	7	Both classification and regression perform well	Black box
(Forghani et al., 2018)	TOPSIS	MILP	Pharma	20	24	Outperforms TOPSIS	No validation
(Cheng et al., 2020)	DEA, TOPSIS	GP, SVR	Manufacturing	26	6	Low error values	Small dataset
(Alavi et al., 2021)	FBWM	FIS	Petrochemical	10	14	Handles fuzzy data	Black box

Table 3.3: Hybrid MCDM/AI approaches to supplier selection.

3.5 Evaluation of Related Work

Table 3.3 offers a comparative overview of most of the works discussed in the previous section. From the related work, it can be noted that the majority of researchers focused more on accuracy than on interpretable methods, which are also important in the context of problems where the methods used and results obtained need to be explained to decision-makers. Also, many researchers used only small datasets for their case studies, and their methods need to be validated on much larger datasets.

In many instances, ML classifiers are used to overcome the limitations of current MCDM methods such as AHP. A common characteristic of supplier selection hybrids is that they begin with an MCDM technique to determine criteria and their importance, feeding this information to an ML algorithm to develop a selection model. This may not be appropriate in cases where there are large numbers of criteria or when there is a requirement to explain supplier selection decisions to stakeholders who may not be knowledgeable of or conversant with ML.

Only two works in Table 3.3, use hybrids combined with interpretable AI techniques that can be readily explained: decision trees (D. Wu, 2009), and fuzzy inference systems (Alavi et al., 2021). In the former case, the proposed hybrid was tested only on a relatively small case study (eleven suppliers and six selection criteria), so further investigation is necessary to determine applicability and scalability. In the latter case, while fuzzy inference systems can be transparent and readily explainable, they are powered by handcrafted rules based on expert knowledge (Baryannis et al., 2016). Moreover, as explained by the authors, the number of inputs and input membership functions needs to be kept quite low (here, the values were 5 and 3, respectively); otherwise, the complexity and processing time are unmanageable. As noted in literature (Ni et al., 2020), it would be incredibly difficult for a corporation to roll out a new machine learning system that functions as a "black box" producing accurate results without explanation. This makes accepting ML in SCM more difficult for a firm.

3.5.1 Comparison to Research in this Thesis

In this thesis, we focus on addressing the aforementioned issues that are common in hybrid supplier selection literature. First, we prioritise interpretability over prediction performance by ensuring that the ML component included in the developed hybrid is inherently interpretable. Second, we go beyond the common trend in literature of using MCDM for determining criteria and weights and ML for selecting suppliers by also including a model that follows the reverse path, using ML for weighing criteria and retaining MCDM at the heart of the actual supplier selection task. In this way, there is an increased likelihood of adopting such a model by stakeholders who are more familiar with MCDM rather than ML.

Finally, we place great emphasis on the validation of the proposed models. This is achieved through a real-world case study of high complexity, both in terms of selection criteria and in terms of the number of alternative suppliers. The case study also relies on a considerably larger dataset compared to what is usually explored in literature.

3.6 Summary

This chapter aimed to introduce and situate the present research by assessing related work that researchers have done in previous research to solve supplier selection problems. This chapter discussed the connection between these studies and the supplier selection criteria utilised to substantiate their findings. This chapter has reviewed pertinent literature of MCDM, ML and integrated MCDM with ML, and identified knowledge gaps in selecting and evaluating suppliers. It has also demonstrated the importance of developing an up-to-date supplier selection and evaluation framework and establishing a decision support system to assist purchasing managers in evaluating and improving their company's performance. An investigation revealed that there is no approach in the literature that sequences ML first to weigh the criteria, followed by MCDM to rank the suppliers, and also, while the reverse sequence is quite common, there is no instance of a hybrid that includes a combination of MCDM techniques such as FUCOM and TOPSIS as a means to prepare training data for a subsequent ML phase. The next chapter details the two hybrid approaches to supplier selection that are the main contribution of this thesis.

Chapter 4

Proposed Hybrid Frameworks for Supplier Selection

4.1 Introduction

The main purpose of this chapter is to identify the research approaches used in this thesis and the reasons for choosing them. The frameworks and models proposed are presented in detail. In addition, the data collection procedures and data analysis methodologies used in this research are described. The proposed hybrid supplier selection approaches presented in this part aim to address some of the issues listed at the end of Chapter 3.

The experimental work and case study are presented in the next chapters, emphasising the need for a hybrid and well-defined context-based approach to evaluating the model in the problem of supplier selection and evaluation, which the integrated model solves.

The proposed hybrid approaches that combine MCDM with ML to solve the problem of supplier selection are as follows:

- Framework One: Using ML to reduce complexity in terms of the number of criteria or /and the number of suppliers and MCDM to rank the alternative suppliers and select the best supplier.
- Framework Two: Applying MCDM approaches to determine criteria weights and rank alternative suppliers in order to create a training dataset for a ML-based binary classifier that is used to decide whether a supplier is the best (the one to be selected) or not.

These are first described in the following two sections, before then instantiating them for the purposes of this thesis, showing particular examples of ML and MCDM combinations that were explored.

4.2 Framework One: ML+MCDM

4.2.1 Overview

Figure 4.1 illustrates the framework in the form of a flow chart. The main constituents include a data-driven AI strand that aims to manage the complexity of large numbers of either criteria or suppliers (or both), followed by the application of an MCDM technique to inform supplier selection decisions, taking advantage of their explainability. While the framework does not prescribe the use of particular ML algorithms or MCDM techniques, the in-depth analysis includes desirable characteristics that should maximise the framework's applicability. A specific instantiation of this framework that employs DT and AHP is discussed in Section 4.4.1.

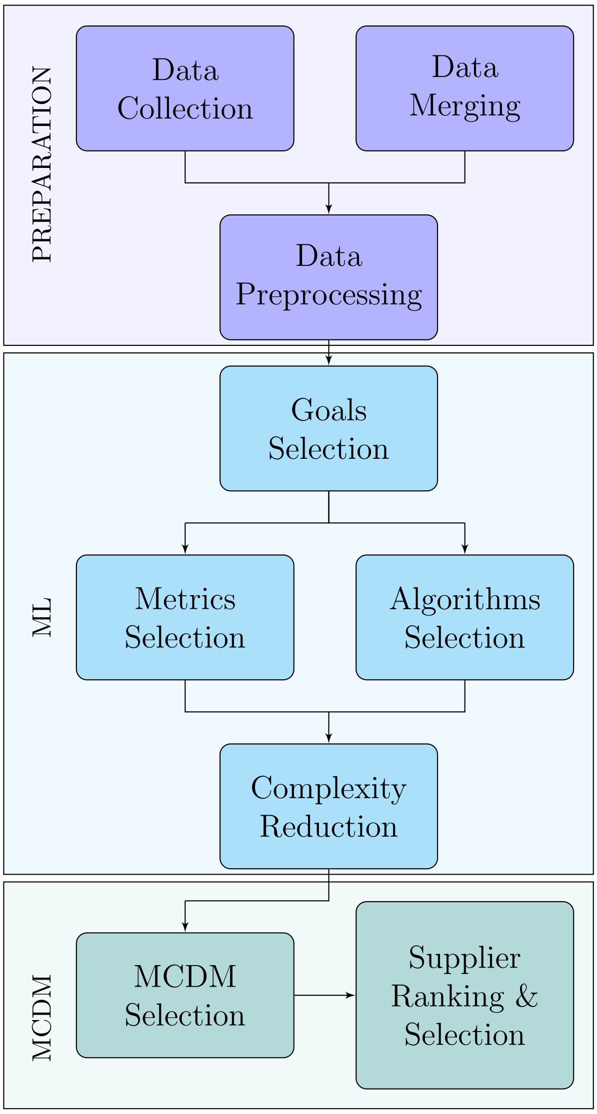

Figure 4.1: Proposed Hybrid ML+MCDM Supplier Selection Framework.

4.2.2 Data Pre-processing

The first step in most supplier selection approach that relies on data is preparing the related data, including gathering relevant data about suppliers and their prior performance because the efficacy of data-driven AI systems directly depends on the input data quality. Supplier databases, assessment spreadsheets, historical procurement data, and other data types may be useful in the data collection process; when multiple data sources are employed, data merging may be necessary to create a single dataset that can be used as the foundation for the rest of the process.

A general overview of data pre-processing has been provided in chapter 2; however, when it comes to supplier selection in particular, data preprocessing tasks that are relevant and should be considered include the following:

- Removing incorrect, duplicate or irrelevant data
- Dealing with missing data through various imputation techniques
- Dealing with data outliers through filtering techniques
- Scaling data through standardisation or normalisation
- Encoding categorical features
- Dealing with data imbalance through sampling techniques
- Using feature engineering approaches to generate new features

Deciding which of these pre-processing tasks will be undertaken and their impact is directly dependent on the available datasets. Considering data imbalance as one of the data pre-processing tasks, this is most problematic in classification problems and can be encountered in supplier selection datasets that result from incomplete data capturing processes where data for discarded suppliers is not recorded adequately. In such cases, dealing with data imbalance is crucial to developing classification models that achieve high performance(Fernández et al., 2018).

The inclusion of a ML component makes the data cleaning part of data pre-processing more significant since the performance of ML approaches is directly dependent on the quality of the input data(Géron, 2019). For handling missing values, any of the techniques discussed in Chapter 2 can be considered depending on the dataset, as listed below (Witten & Frank, 2002):

- Completion of all the missing data if only some are missing by estimating values and replacing the missing value with the mean or mode.
- Removal of the rows which do not have any values or most data are missing.
- Filter with low variance.

4.2.3 ML Phase

After pre-processing data to ensure the highest possible quality level, the framework moves on to its ML phase, using the pre-processed data to calculate selection criteria weights. This process involves choosing appropriate machine learning algorithms as well as metrics to evaluate their results. These choices are directly dependent on which dimension (criteria or suppliers) is targeted to reduce complexity. If the goal is to reduce the amount of selection criteria, then machine learning algorithms need to be targeted at determining which criteria hold the highest importance in terms of determining the most appropriate supplier. This would mean that algorithms and techniques for feature importance and feature selection need to be considered. On the other hand, if the goal is to reduce the amount of suppliers that are considered, then one approach would be to learn supplier selection scores (e.g. through regression) and retain only those that pass a certain threshold.

An additional decisive factor in terms of selecting algorithms is interpretability. Since the main rationale behind proposing this hybrid framework is to suggest a supplier selection approach that is more explainable and understandable from the point of view of supply chain stakeholders, choosing a non-interpretable machine learning algorithm would be counter-intuitive. To ensure that the complete framework can support end-to-end explanations of its operation, the process of reducing the complexity of the supplier's dataset needs to be transparent and this requires machine learning algorithms that are inherently explainable (to an extent), such as decision trees or generalised linear models. Alternatively, post-hoc interpretability methodologies such as SHAP (Lundberg & Lee, 2017) and LIME (Ribeiro, Singh, & Guestrin, 2016) can be leveraged if this additional layer of complexity can be afforded.

In terms of metrics, these again need to be considered in relation to both the chosen goal and the selected algorithms. In the aforementioned examples, Gini importance (Chen et al., 2004) may be considered as a metric to determine the top criteria for supplier selection, while mean squared error may be more appropriate for regression-based scoring of suppliers. If the goal involves classifying suppliers (e.g. depending on their potential) and there are imbalance issues in the dataset, then metrics such as receiver operating characteristic (ROC) or precision-recall curves should be considered.

4.2.4 MCDM Phase

Applying the algorithms considered in the previous step should lead to a less complex problem in terms of the number of selection criteria and/or the number of suppliers considered. The resulting reduced set of criteria and suppliers can then be fed into an MCDM approach in order to make an informed decision about the best supplier according to the criteria or rank suppliers based on their overall (or weighted) score against the criteria. If a feature selection process was followed to reduce the number of selection criteria, then the metrics used in that process (e.g. feature importance metrics) can form the basis for the weighting scheme of the MCDM approach. Our framework does not prescribe a particular MCDM approach, as the choice depends on several factors such as prior knowledge and expertise, the familiarity of stakeholders and organisational culture.

The outcome of the MCDM approach and this hybrid framework as a whole is a conclusion about which supplier to select supported by a full explanation that can be used by purchasing managers and other stakeholders to make an informed decision as to whether they will adopt or discard the outcome. Explanations are end-to-end, including a rationale behind discarding criteria and/or suppliers in the machine learning phase, as well as the output of the MCDM approach.

4.3 Framework Two: MCDM+ML

4.3.1 Overview

The second proposed hybrid approach is illustrated in figure 4.2. The initial preparatory phase is similar to the first hybrid. However, there is an additional need for collecting information that will be used to determine selection criteria, such as through a questionnaire, as described in the next section. The main constituents that follow are now reversed, with MCDM techniques used to weight the criteria and rank the suppliers, resulting in a labelled dataset that can be used to build a ML binary classifier in order to choose (or not choose) a supplier. As before, this framework does not prescribe the use of particular MCDM or ML approaches. However, in this case, MCDM approaches are used for two separate steps, which allows different MCDM approaches to be combined, as explained later. Prioritising interpretable ML algorithms in the final step is again desired for the same reasons of transparency and trustworthiness that were explained in relation to the first hybrid. However, in this case, they are significantly more important because the decision is made by the ML algorithm and not the MCDM technique. A black box algorithm would lead to a supplier selection that cannot be explained. Also, this hybrid differs from methods used in the literature in which this model utilised the MCDM methods as a classifier to label the dataset based on the decision makers' opinions because the dataset collected is not correctly labelled. A specific instantiation of this framework that employs FUCOM and TOPSIS with DT is discussed in Section 4.4.2

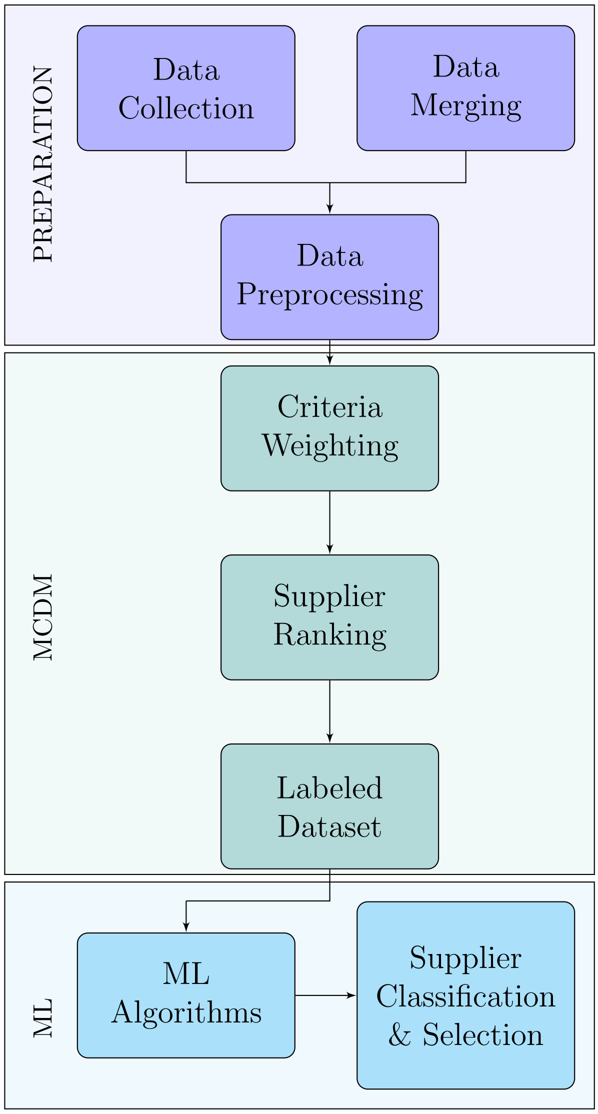

Figure 4.2: Proposed Hybrid MCDM+ML Supplier Selection Framework.

4.3.2 Data Pre-processing

All data collection and preprocessing considerations are similar to those described in Section 4.2.2.

As mentioned above, the preparation phase also needs to include a data collection mechanism to allow criteria weighting at the start of the MCDM phase. Questionnaire surveys are the preferred data collection technique here, primarily because they minimise the potential of bias, which would create issues with the generality of the chosen criteria weights. Each question in the questionnaire survey should denote an element or group of sub-variables that should be quantified in the supplier selection criteria according to the expertise and experience of each participant.

The questionnaire is intended to provide a specific response to how the supplier selection problem is approached by individuals involved in the process and establish the precedence of certain criteria over others. In completing the questionnaire, purchasing managers and decision-makers should be given several options for responding and sufficient time to supply the precise information required. The goal of creating a questionnaire is to narrow the study area in terms of the number of criteria considered. In this sense, this step serves a similar purpose to what ML did in the first hybrid.

Examining the reliability and validity of data gathered through questionnaires is critical for assisting study findings. In this step, statistical metrics such as Cronbach's alpha (*alpha*) can be used to measure data consistency and dependability of the collected data (Vaske et al., 2017). The value of *alpha* can be defined as a formula 12 indicates, which takes into consideration both the number of data items and the average inter-correlation between those items.

$$\alpha = \frac{N.\bar{c}}{\bar{v} + (N-1).\bar{c}} \tag{12}$$

- The number of data elements is N.
- The average inter-correlation between the features is c-bar.
- The v-bar represents the average variance is represented by the v-bar.

4.3.3 MCDM Phase

The training data for most supervised learning algorithms is a set of labelled examples. There is a lot of unlabelled data in many applications, yet manually obtaining labels for such data is time and money-consuming. Despite the quantity of data in many supplier selection datasets, most transactions are poorly classified or unclassified. This creates a major issue in terms of the applicability of ML approaches. This is the main purpose of the MCDM phase in this hybrid framework: having collected relevant data previously, these now need to be used to determine selected suppliers (as well as a relative ranking between them) for every transaction in the available datasets. This way, datasets are correctly and fully labelled (in that every supplier is assigned to one of two classes, selected or not selected), and they can then be fed to the next and final phase (ML).

Based on the questionnaire results, an MCDM method is applied first to calculate the weights for each criterion chosen by experts. Then, the same or another MCDM method is used to rank alternative suppliers. There are different benefits to choosing the same or different MCDM methods. Choosing the same method may be preferable in terms of familiarity in case relevant stakeholders strongly prefer this particular method, e.g. because it is used or has been used by the company in the past. Choosing different methods affords more flexibility and can lead to more optimal results, as the choice of each method will be driven by the individual characteristics of each of the two tasks, weighting criteria and ranking suppliers. This is also a differentiating characteristic of the proposed framework compared to similar MCDM+ML approaches in the literature, where only a single MCDM approach is used.

The output of the MCDM methods is the assignment of ranks to every supplier associated with every transaction based on scores calculated using criteria weights and values. Given that the next and final phase involves binary classification, a conversion step is necessary, where each supplier ranked first is allocated to the class of "good" suppliers (suppliers that are selected) and the rest are allocated to the class of "bad" suppliers (suppliers that are not selected). This ensures that all dataset entries are assigned to a class, and the end result is a labelled dataset that can be fed to the next and final phase.

4.3.4 ML Phase

This step begins similarly to the ML phase in the first hybrid, where suitable metrics and algorithms need to be selected, as analysed in Section 4.2.3. In contrast to the first hybrid, where ML's primary goal is complexity reduction, the ML phase is a standard binary classification process. Hence, algorithms that have been successful in such problems are recommended. Additionally, standard classification metrics, such as precision, recall and F1 score, are suitable, and other metrics may also be considered, especially if there is a significant imbalance in the dataset.

As explained at the start of this section, interpretability is a stronger requirement in this hybrid compared to the first one. This is because decision-making is now performed using ML rather than MCDM. This means that the decision to be provided to the purchasing manager is provided by an algorithm that is probably unfamiliar to them and less likely to be trusted. The more information that can be provided to them to go along with the classifier output (which is simply a binary result, selected or not selected), the more likely it is for them to understand the reasoning behind the decision and adopt it rather than challenge it.

4.4 Instantiating the Frameworks

This thesis explores specific instances of the frameworks presented in Sections 4.2 and 4.3. These are detailed in the remainder of this chapter.

4.4.1 Supplier Selection Combining DT and AHP

This section presents the specific methods and algorithms used as an instantiation of the framework presented in 4.2 focusing on the rationale behind choosing DT and AHP. Figure 4.3 illustrates the framework in a flowchart format. In brief, the components consist of a decision tree as a ML algorithm designed to manage the complexity of a large number of either criteria or suppliers (or both), followed by AHP as an MCDM to rank suppliers.

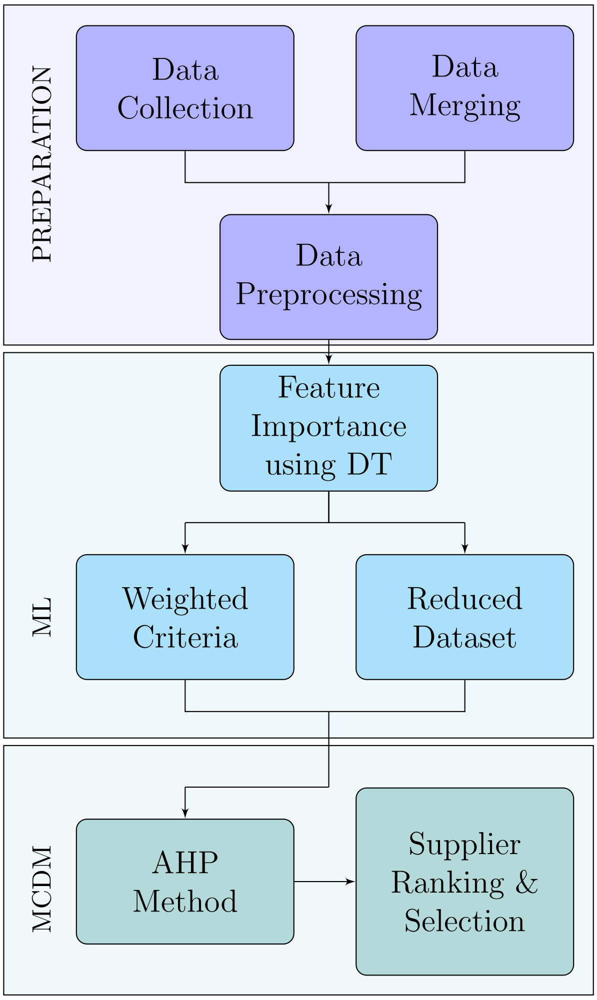

Figure 4.3: Proposed Hybrid DT+AHP Supplier Selection Model

The main goal of the proposed model that follows a ML+MCDM structure is to reduce the complexity of having a wide range of selection criteria. To this end, feature selection approaches are appropriate to determine the relative importance of features and identify which are most important. While there is a wide variety of feature selection approaches to choose from (J. Li et al., 2018), we opted to use decision trees which calculate feature importance as part of the model-building process.

The main reason behind this choice is its explainability: in order to be able to provide as much explanation as possible to purchasing managers and staff, there needs to be a rationale in place for only using some of the features as selection criteria. As shown in figure 4.4, decision trees exhibit the highest notional interpretability among common ML algorithms.

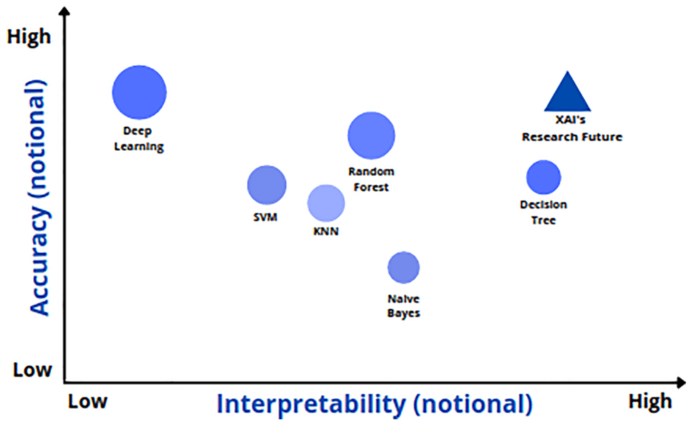

Figure 4.4: Interpretability vs Performance in different ML algorithms (Angelov et al., 2021)

An additional reason in favour of choosing decision trees is their ability to handle both categorical and continuous data. This is useful for supplier selection and evaluation as companies may need to examine both quantitative and qualitative criteria, including quality, delivery time, and cost. Finally, the ability of DT models to handle missing values is also quite desirable, given that the amount of information provided differs from one supplier to the next.

In terms of the MCDM phase of the model, the choice of AHP is primarily due to its widespread popularity. We confirmed this through interviews with purchasing managers and staff as part of the case study described in the next section, which showed that AHP was the only MCDM approach that every-one was familiar with, either having directly applied it in the past or being indirectly aware of it.

4.4.2 Supplier Selection Combining FUCOM and TOPSIS with DT

This section presents the specific methods and algorithms used as an instantiation of the framework presented in 4.3, focusing on the rationale behind choosing FUCOM and TOPSIS. Note that the rationale behind choosing DT for the ML phase of the framework remains the same as discussed in the previous section. Figure 4.5 illustrates the framework in a flowchart format. In brief, the components consist of FUCOM as a criteria weight calculation method and TOPSIS as a ranking method, followed by DT to develop a classification model based on the ranked supplier dataset resulting from the combination of FUCOM and TOPSIS.

In contrast to the previous framework, the choice of MCDM techniques was not influenced by the stakeholders' familiarity with them. In this case, actual supplier selection is not decided using MCDM in the final phase. But it is performed based on machine learning. FUCOM was chosen for weight calculation as it is a relatively more recent technique designed to improve issues compared to AHP, which was used in the previous model in Section 4.4.1. TOPSIS was chosen as a commonly used technique in literature to rank alternatives in a MCDM problem instance.

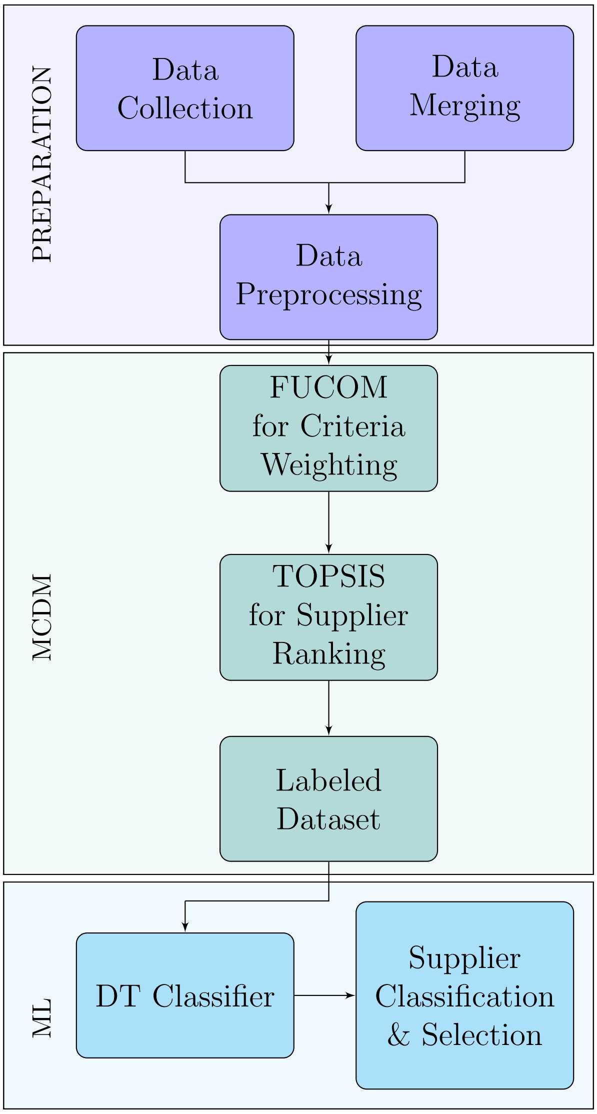

Figure 4.5: Proposed Hybrid FUCOM+TOPSIS+DT Supplier Selection Model

4.5 Summary

This chapter detailed the research methodology used to achieve the thesis aims and objectives and the individual steps within. The two hybrid frameworks proposed in this thesis were defined and explained in depth, also highlighting the differences between them and with relevant literature. Additionally, the two frameworks were instantiated into two specific hybrid models, employing DT and AHP in the first case and FUCOM, TOPSIS and DT in the second case. The case study and related datasets that were used to explore the applicability of the proposed approaches are described in the following chapter.

Chapter 5

Case Study

5.1 Overview

A case study is a research method used to develop a comprehensive, multifaceted understanding of a complicated subject in its real-world context. In this chapter, this established research strategy is utilised to determine the applicability of the hybrid frameworks proposed in the previous chapter in the case of a Libya-based oil and gas company.

The company operates under the Libyan National Oil Corporation (NOC), an oil and gas company based in the middle north of Libya. The activities of NOC include the exploration and development of oil and gas. This resource for oil and gas operations focused on world oil production, offshore oil production, projects for oil management, and gas production. Industrial activities such as refining, manufacturing, marketing, and distribution are all part of the production strategy for Libya's oil sector to operate properly because NOC is the national oil company. This significantly enhances local oil company activity, which benefits the economy, particularly in the oil and gas sector. There have been enhancements to the structure and method of reserve development and several key oil projects and strengthened government authority over the oil and gas sector management. Locals also run the power generation, with the proceeds funding ambitious development plans, most notably to earn foreign exchange to create a balanced economy with different sources of income (Mohamed et al., 2008).

The information provided in the subsequent sections was collected during interviews with one of the purchasing managers, B. Almabrok and other relevant staff that were conducted from 2018 onwards.

5.2 Purchasing Department

The purchasing department is linked with the other departments, such as production, marketing, finance, warehouse, and so on, because it conducts procurement and supply operations to assist and support the company and its other departments. The success of any of these departments is dependent on the others' ability to accomplish their tasks well (Ross et al., 2010). The purchase order is initiated in the Libyan oil and gas company when a department requests it from the purchasing department. For instance, if the department's inventory of raw materials is low, it will request the items from the purchasing department by filling out a detailed and accurate form on the different types of items to avoid any errors in the purchase order. The purchasing department sends a supplier list that matches this request. After that, the suppliers give offers to the purchasing department to evaluate, and the purchasing managers evaluate the offers and select the best supplier.

5.2.1 Supplier Evaluation Process

Interviews with the company's procurement managers and staff involved in the day-to-day procurement processes established an understanding of the company's suppliers' evaluation and selection process. A purchasing department in the company is responsible for all procurement activities, and all other departments process requisition requests through it. The process begins with a request from one of the departments, submitted by filling in details in a form that follows a centrally approved template designed to standardise the process across all departments. The document contains information on the items ordered, their costs, the requested delivery date, and other requirements.

The purchasing department has developed a software system that manages data for all suppliers known to the company (including preferred and frequently used ones) and is updated daily. When a new order request is received, this system identifies suitable suppliers and requests for offers are sent to them. The number of suppliers varies and depends on the complexity of the items ordered. One or more suppliers provide their offers to the purchasing department, and purchasing staff are tasked to evaluate the offers and select the most appropriate supplier. Evaluation is based on various criteria, including technical quality, price and delivery time, payment conditions, offer validity and offered quantity, and any other information available for suppliers. A short questionnaire was provided to staff involved in evaluating supplier offers to further understand the relative importance of selection criteria. The main outcomes of this questionnaire-based interview were:

- All criteria and information available for suppliers are considered relevant and are taken into consideration.
- Technical quality, price and delivery time are considered the most important criteria.

A form-based software system is again used to record supplier selection decisions, with purchasing department staff completing the details of the selected supplier, along with a rationale provided against each relevant criterion. This information is submitted and approved by purchasing managers before going ahead with the remaining steps in the procurement process (e.g. agreeing on a contract with the supplier and delivering items).

5.3 Datasets

The company provided two separate datasets, each exported from one of the two software systems described earlier: (1) a "bidders" dataset with more than 20,000 offers of suppliers made for requisitions in different departments of the company; (2) an "evaluation" dataset, providing information on the selected supplier for 2,300 requisitions, including the criteria behind the decision made. These datasets were collected from this company's purchasing manager as an Excel file sheet. For the purpose of data collection, heads of procurement, as well as procurement staff, were also interviewed to select the best suppliers from the total number of suppliers. This is done on the basis of the best appropriate criteria among the available standards.

The "bidders" dataset contains 30 features with information relevant to a supplier offer, including the following:

- Unique request number
- Supplier information (e.g. unique number, address)
- Number of items requested, and number of items offered
- Total amount of offer, including the cost of items and extra charges (such as transportation costs). Prices might be in different currencies but are converted to USD for uniformity.
- Offer date and offer validity (after which the supplier withdraws their offer)
- Payment method (e.g. cash in advance, paid in full or in instalments)
- Payment terms (e.g. 30 days from delivery, after inspection/testing)
- Delivery terms (e.g. shipping is the supplier's responsibility or the company's responsibility)
- Delivery time (number of weeks that will be required for items to be ready for shipment)

The "evaluation" dataset contains 23 features, and some are common with the "bidders" datasets, such as unique request number, unique supplier number and the total amount of offer. However, most of the other features provide information related to one of the selection criteria and indicatively include the following Boolean criteria (true if the criterion is met, false otherwise):

- Lowest price (the offer has the lowest total amount, including extra charges)
- Delivery date (the offer has the earliest delivery date)
- Previous supplier (there has been a contract with the supplier named in the offer in the past)
- Manufacturer (the supplier named in the offer is also the manufacturer of the items included)

5.3.1 Data Pre-processing

The first large phase for both proposed frameworks involves data pre-processing steps. As two datasets were provided, the initial stage was to try to integrate these into a single dataset. The unique request number is used in both datasets to relate to separate procurement requests, and hence this feature was employed to ensure that data from two different requests would not get mixed up. This is necessary in order to determine which of the first dataset's offers was finally chosen as the best and recorded in the second dataset. Also, when data for a request was not accessible in both databases, the request was eliminated (for instance, offers without an identified chosen supplier). A target feature was also created by merging, assigning a value of 0 to suppliers who were not selected for a specific request and 1 to those who were.

Following the merge, redundant features were removed, as several features, including supplier number, were shared between the two datasets. After any features that were redundant were eliminated, the importance of each remaining feature was assessed by considering how it related to the issue at hand. A number of features were seen as being completely unimportant to the supplier selection process. These include the price in the original currency and currency code (the price in US dollars is preferred for the sake of uniformity); the bid price and extra charges (the total amount is preferred as the sum of both); the delivery place (due to having the same value across all entries); the discount (value of 0 across all entries); the full company name of the supplier; the name of purchasing staff and manager taking the decision; the name of department placing the request; the order description, as well as some binary selection criteria which had a false value across all entries.

As a result of the feature removal steps, the size of the feature set was reduced to 21. And then turned our attention to samples and removed all duplicate samples and samples that were missing most features since they do not contribute to a model's learning ability, as well as requests that only attracted an offer from a single supplier (hence removing the need for a supplier selection process). Then, we converted categorical and alphanumeric values to numerical form. For instance, dates were split into three numbers (year, month and day), while payment terms were converted from a [1-9, A-D] range to a [1-13] range. And also, a new feature has been added for the duration of an offer as the number of days between the offer date and the delivery date.

The final two data pre-processing steps included dealing with missing values and normalisation. For the former task, multivariate feature imputation (van Buuren & Oudshoorn, 2015) was used. For normalisation, the supplier features were normalised using the formula 2.2 in Chapter 4 to keep the range of values of each supplier feature between [0, 1]. At the end of the data preparation, the total amount of samples included in the dataset was 10,580.

5.4 Applying the DT+AHP Hybrid

In this section, a rationale for the appropriateness of the DT+AHP hybrid in this case study is provided. The purchasing department is responsible for procurement across all other departments and has to handle a great variety of orders involving thousands of suppliers. Indicatively, the provided datasets, which cover only a four-year period, contain 3147 distinct suppliers. The information used to select a supplier involves a large number of criteria and associated supplier information. The provided datasets include up to 26 distinct features involved in supplier selection. Additionally, the majority of requests include five or more offers from different suppliers. This significant complexity in terms of both criteria and suppliers makes the work of purchasing staff harder. As a result, they may be prone to make mistakes or risk compromising the integrity and quality of the supplier selection process.

Literature, e.g. (R. Zhang et al., 2016), recommends the use of machine learning to address this issue by partially or fully automating the supplier selection process. However, discussions with one of the purchasing department's managers have indicated a lack of knowledge and understanding of machine learning technologies and how they could be of benefit to them. This leads to a lack of trust in the appropriateness of machine learning for selecting suppliers and reluctance to adopt its outcomes. In contrast, several purchasing department employees are familiar with MCDM approaches such as AHP and understand how they operate. Taking these factors into account, the proposed framework can be considered an appropriate solution given that it would reduce the complexity of the supplier selection process while retaining some level of familiarity (through the use of MCDM approaches) to increase trust and likelihood of adoption.

As a result, a combination of DT and AHP is fully appropriate for this particular case study. In the next chapter, a comparative performance analysis of DT and other ML algorithms is also provided to confirm their positioning in terms of accuracy and other evaluation metrics and justify the selection of DT quantitatively. By interpreting the produced decision tree and assigning important values to features, the proposed model makes sure that end-to-end explanations of the framework are possible and not just available for the MCDM component that follows.

5.4.1 Feature Importance Metrics and Thresholds

The feature importance metric that was chosen is Gini importance, which is calculated as the normalised total reduction of Gini impurity (Breiman et al., 1984) brought by that feature. The impurity reduction of a feature is calculated by subtracting the weighted sum of the impurities of the child nodes from the impurity of the parent node. The weight of each child node is proportional to the number of samples it contains. Features with higher impurity reductions are considered more important, as they contribute more to the overall predictive power of the model.

To determine where to set the threshold of importance (to decide which features to exclude), the literature has taken into account such as (Chan et al., 2008), which refers to 7-9 features as the limit for human understanding and ability to make informed decisions. The results for all features have also been considered, and it was found that Gini importance was negligible for all but 8 of the features. Based on this, the top 8 most important features were selected. More details on selected features are available in the presentation of results in the next chapter. Feature selection led to a reduced dataset in terms of selection criteria, focusing only on the aforementioned eight most important ones.

5.4.2 Ranking and Selecting Suppliers using AHP

The next and final phase of this hybrid is to use this reduced dataset, as well as the calculated feature importance values as input to AHP. Feature importance values were used as weights for each of the 8 criteria, and these weights were multiplied by the criteria values with the products summed up to yield a score for each supplier. Note that positive and negative features are distinguished by attaching a plus or minus sign to the product of weight and value. For instance, offer duration is a positive feature since the longer an offer is valid the better it is, while the price is a negative feature since lower prices are better. The supplier with the highest score among those that have made offers for a particular requisition is the chosen supplier according to the proposed framework. In the next chapter, experiments show how AHP evaluates the suppliers based on the feature importance values calculated by the DT model.

5.5 Applying the FUCOM+TOPSIS+DT Hybrid

As explained in Section 4.4.2 FUCOM is used as a weight calculation method and as an example of a more recent technique that improves on AHP. TOPSIS, on the other hand, uses criteria weights as calculated by FUCOM and calculates a score for each alternative supplier, as well as a ranking based on these scores. This process is, in its essence, a data labelling process that ensures each alternative supplier is associated with an evaluation score.

The final step in this hybrid is to use the results of the FUCOM+TOPSIS phase to train a DT model that can classify a (previously unseen) supplier as selected ("good") or not selected ("bad"). As in the previous hybrid, to confirm the suitability of DT, other classification algorithms were also explored. The results of this exploration are detailed in the next chapter.

5.6 Summary

This chapter detailed the case study that was used to confirm the applicability of the proposed hybrid frameworks. The case study involved real-world data collected from an oil and gas company in Libya. The next chapter provides further details of the implementation of the framework instances summarised in this chapter, detailing and discussing a complete set of results.

Chapter 6

Experiments and Results

6.1 Introduction

In this chapter, results of applying the different frameworks proposed in chapter 4 are presented and evaluated in order to determine the feasibility of each method for supplier selection. For most ML tasks described in this chapter and the previous one, scikit-learn v.1.1.1 (Pedregosa et al., 2011) was used.

This chapter also presents the results of a supplier selection performance evaluation, which demonstrates how the proposed model outperforms other models when it comes to ranking and selecting appropriate suppliers. This allows various tools and techniques to be compared to determine their appropriateness in supplier selection. Performance is evaluated against several criteria, including accuracy, execution time and interpretability of results.

6.2 DT+AHP results

As detailed in Chapter 4, the first proposed hybrid approach utilises DT to reduce the number of selection criteria and apply weights to them before then ranking suppliers in the second stage by using AHP. This section details experiments and results of applying this instance of the general ML+MCDM framework.

6.2.1 Implementing and Evaluating ML Algorithms

In the case of supplier selection and evaluation, ML algorithms can be applied to help the company to make the right decisions in selecting suppliers. Different algorithms can be applied to different problems and to achieve different goals. If the goal is to predict the performance of suppliers, a regression model can be used to solve this kind of problem, such as a linear regression model. If the goal is to identify similar supplies in groups, the cluster analysis model, such as k-means clustering, is used to group suppliers.

Classification models such as decision trees are used to classify suppliers into classes and then use these models to predict the best suppliers in unseen cases. This is the problem at hand in this particular case study. As explained in Section 4.4.1, DT was chosen for this task. However, we will be presenting results of using other ML algorithms to further justify the choice of DT. These algorithms are: SVM, KNN, RF and ANN. These algorithms have been chosen because they have been successfully used in supplier selection studies, as reported in the literature. In Addition, these algorithms have not been comprehensively compared in the context of evaluating suppliers.

The algorithms were used for a standard binary classification problem where suppliers need to be classified as either "good", in terms of being the one that is actually selected as a supplier, or "bad", in all other cases. In evaluating their results in order to make the most appropriate decision, several metrics were taken into account: accuracy, training time, $F1$ score, recall and precision. The ability to interpret each algorithm's results was also considered. 10-fold cross-validation was used in all cases, with a 70%-30% train-test split.

The results of five ML models regarding accuracy and training time are presented in table 6.1.

Table 6.1: Accuracy result of ML algorithms

ML Algorithm	**Accuracy(%)**	**Time (Second)**	**Rank**
DT	76.4	0.01	1
RF	75.7	0.09	2
SVM	74.6	0.05	3
ANN	73.7	0.15	4
KNN	73.0	0.20	5

As shown in Table 6.1, all the ML models in this case study performed better than KNN in accurately evaluating suppliers, but DT outperformed them all. DT has the highest rate of accuracy with an average of 76.4%; this means that 76.4% of the times of classifying suppliers were correctly predicted. RF has the second-highest rate of accuracy and is slightly lower than DT, followed by SVM. The KNN classifier had the lowest accuracy, with an average of 73.0% accuracy on the performance evaluation scale. The ANN algorithm, which had an overall accuracy of 73.7 % on average, was the second worst performance, ranking at number four.

Table 6.1 also includes the training time results of all five ML techniques. The KNN takes the longest time because it needs to calculate the distance between new instances and training data each time, making it a slower algorithm. DT, on the other hand, requires the least time to train the model.

In terms of precision, recall and $F1$ score, as presented in the classification report for DT in figure 6.1, the DT algorithm is better in terms of recall and $F1$ score to classify "bad" suppliers, compared to precision.

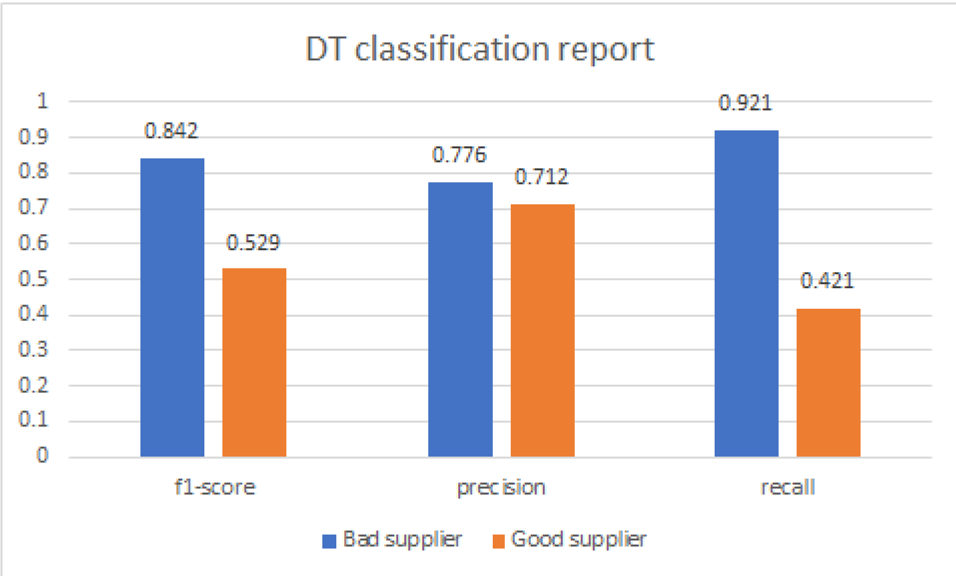

Figure 6.1: DT classification report score

In contrast, to classify "good" suppliers, the precision score was higher than the recall and the $F1$ score. In conclusion, the DT model classified the "bad" supplier class better than the "good" supplier class.

Classification reports for RF, ANN, SVM and KNN are shown in figures 6.2, 6.3, 6.4, and 6.5, similarly showing a better performance at the "bad" supplier class compared to the "good" one.

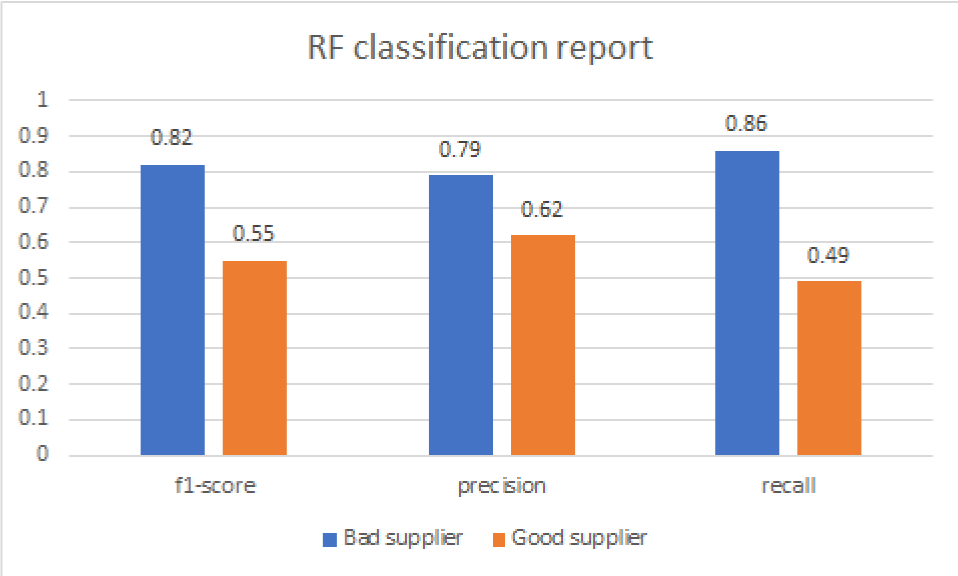

Figure 6.2: RF classification report

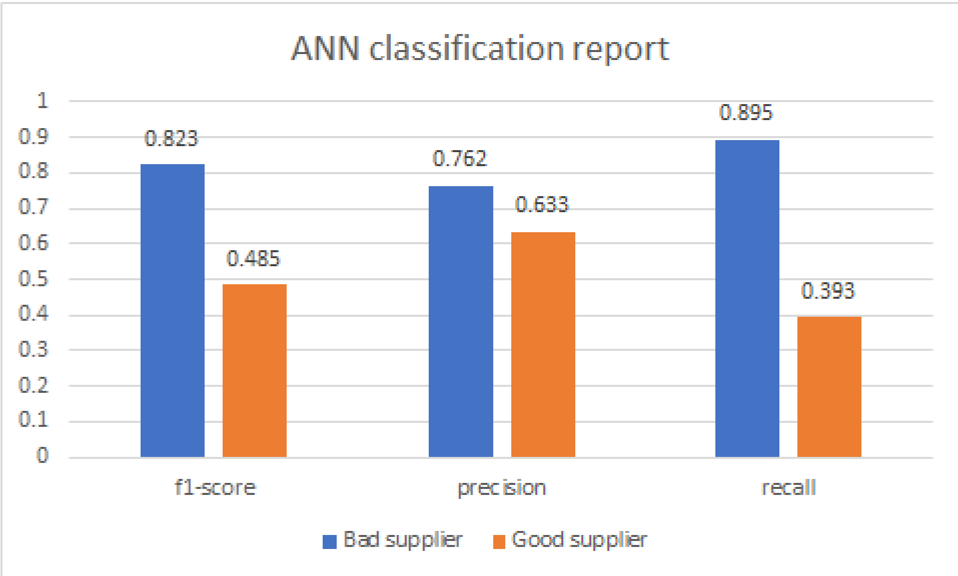

Figure 6.3: ANN classification report

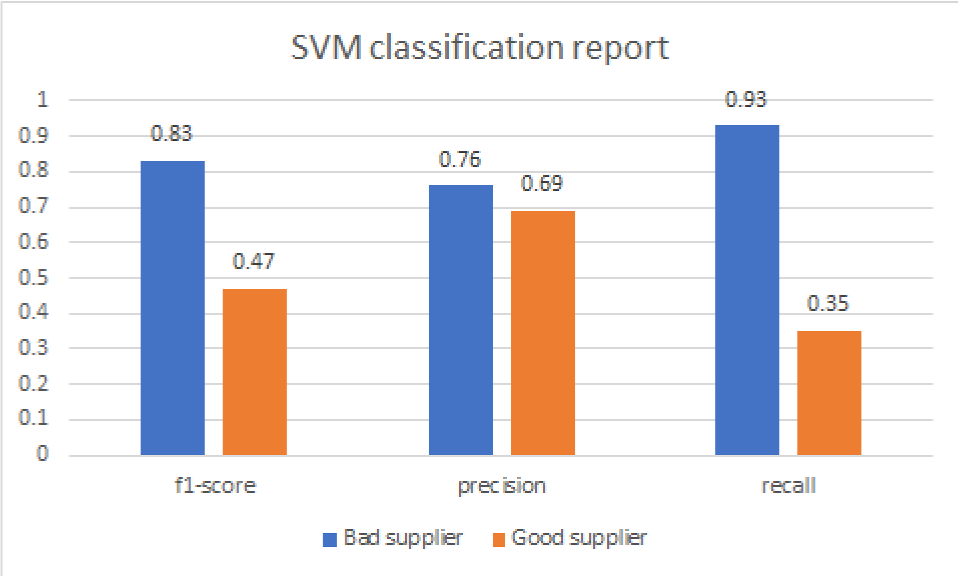

Figure 6.4: SVM classification report

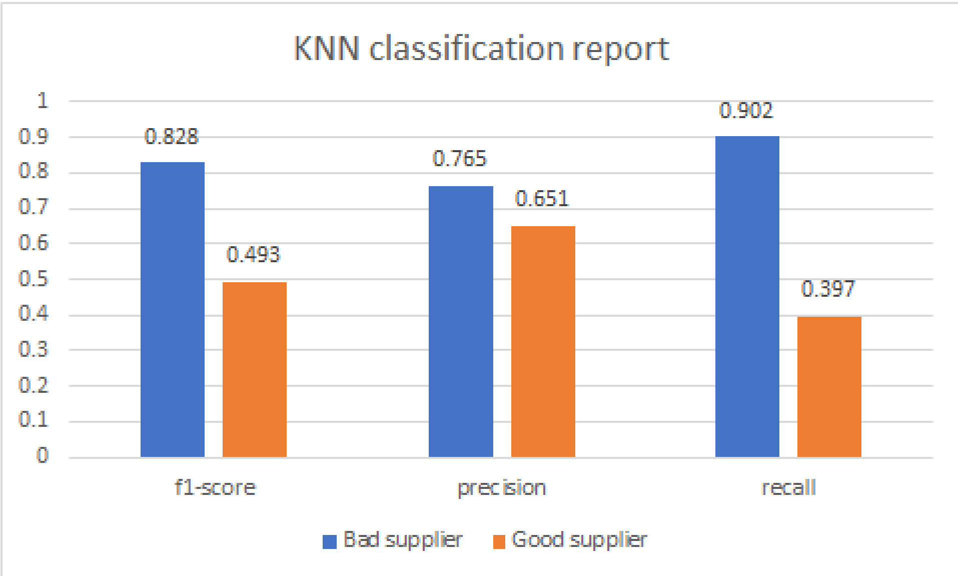

Figure 6.5: KNN classification report

The next three figures provide a head to head comparison of all five ML approaches for each individual metric, starting with precision in figure 6.6.

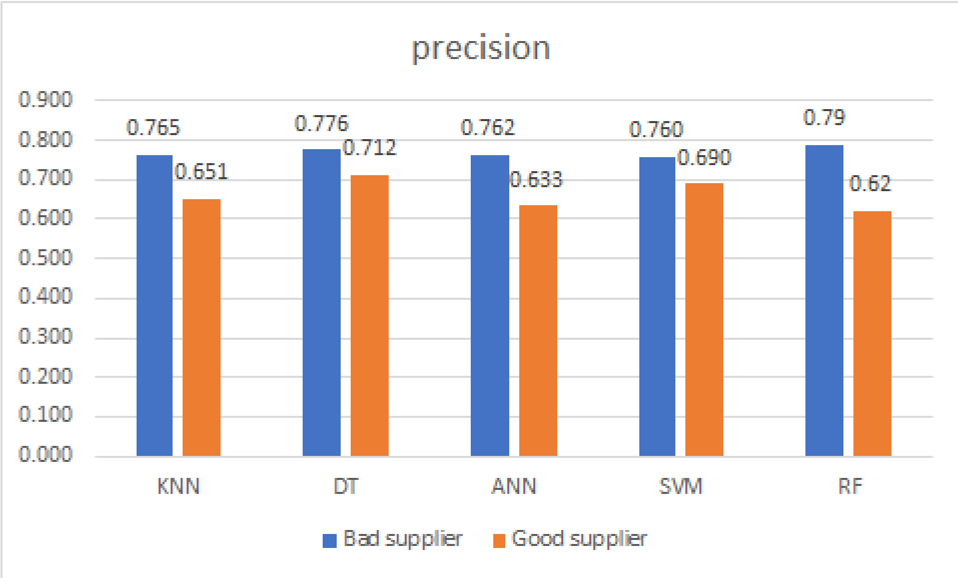

Figure 6.6: Precision score for five ML models of the first hybrid

In terms of precision, the RF classifier was 79% for the "bad" supplier class, ranked highest, followed by the DT classifier with 77.6%. The other ML models, KNN, ANN and SVM, have almost similar results in classifying "bad" suppliers with 76.5%, 76.2% and 76%, respectively. In comparison, to classify good suppliers, the highest precision of 71.2% was achieved by the DT classifier. SVM is the second best one, with 69%. The other classifiers, RF, ANN and KNN, have almost similar results in classifying the "good" supplier class with 62%, 63.3% and 65.1%, respectively.

In terms of recall, as shown in figure 6.7, the SVM classifier outperformed the other classifiers in the "bad" supplier class, which was 93%, but it was the worst classifier in the "good" suppliers class, achieving only 35%. The second

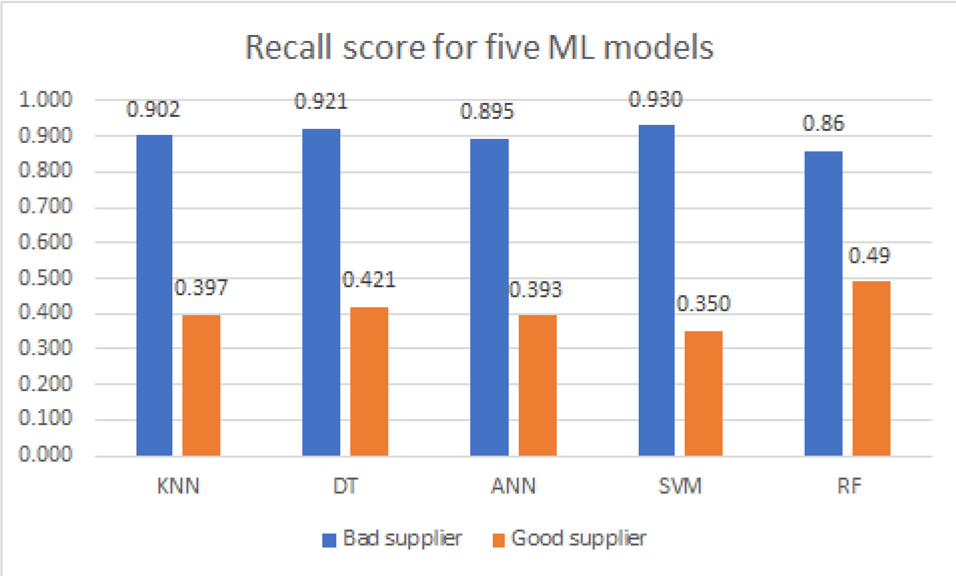

Figure 6.7: Recall for ML models of the first hybrid

highest classifier was the DT classifier which was 92.1% with a difference of less than 1% with the highest model, SVM. For the "good" supplier class, the RF classifier achieved the highest recall score, with 49%, followed by the DT classifier, with 42.1%.

As mentioned in Chapter 2, precision and recall represent two ends of a trade-off. Thus, the $F1$ score is used for an overall, balanced assessment. The $F1$ score of five ML models are presented in figure 6.8, as shown in this figure, the DT classifier achieves the best prediction performance for the "bad" supplier class with a score of 84.2%, while RF achieves the best prediction performance for the "good" suppliers class with a score 55%. ANN and RF achieve similar scores in classifying "bad" suppliers with 82.3% and 82%, respectively. Finally, KNN and SVM result in quite similar scores in classifying "bad" suppliers with 82.8% and 83%, respectively.

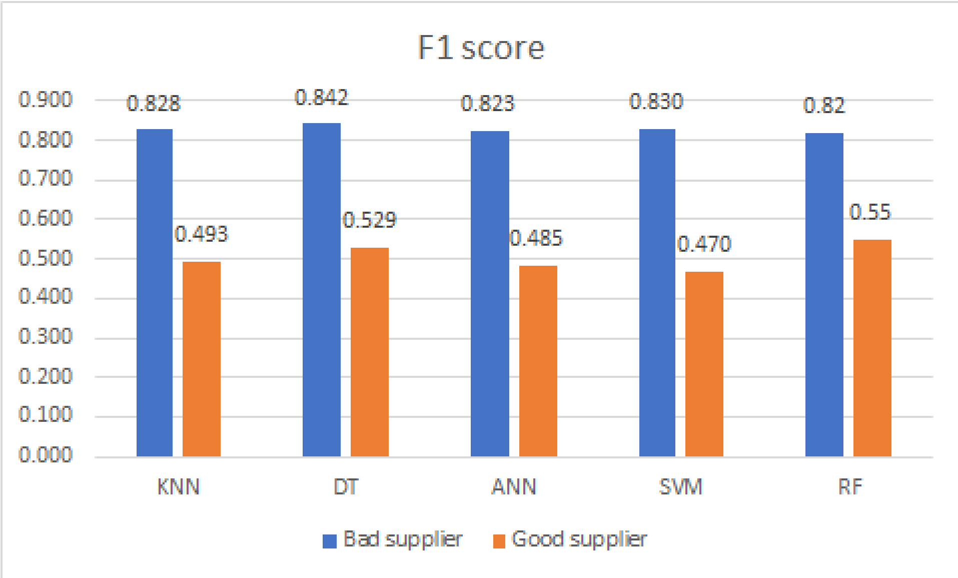

Figure 6.8: $F1$ score for ML models of the first hybrid

In conclusion, the DT classifier was shown to be not only the most accurate but also the most complete in this investigation, with high values for all measures, and was either the first or the second-highest score for both classes, "bad" and "good" suppliers. In addition to this, as explained in the introduction, exploring approaches that have a high likelihood of adoption is of significant importance, and non-interpretable approaches would be problematic in terms of adaptability and trustworthiness, even if their performance might be better. So, a priority is to choose machine learning approaches that are interpretable to an extent, such as DT. Given that in this particular case, DT exhibits high performance, apart from being interpretable, the DT classifier was chosen as the ML model to integrate with AHP in the case study, as described next.

By focusing on interpretability as the main criterion, the applicability of the model in the real world is increased. For instance, to integrate this model in day-to-day procurement practice, the DT can be plotted in a recognisable flowchart as a tree, with the internal node of a decision tree representing a condition on criteria (e.g., high quality or not), each branch reflects the result of the condition and each leaf node as a class label (supplier chosen or not). Procurement staff can then use the tree in its inherent classification rules in the form of paths from the root to the leaf to make supplier selection decisions. Additionally, feature importance plots can help procurement staff understand how an ML model produces predictions.

6.2.2 Weighting Criteria Using DT

Having decided that DT is the most appropriate choice in the case study at hand, the next step is to use DT as a way to both limits the number of selection criteria and assign weights to them that will be considered in the final phase of the approach (the MCDM one). To achieve this, the DT model was used to determine feature importance values for all features and select the subset of these features that is ascribed the highest importance.

Figure 6.9 plots the Gini importance for these features. There is a clear clustering of these features in three groups: (1) the highest importance criteria are price and technical quality, which is in agreement with the results of the questionnaire-based survey conducted with staff involved in purchasing decisions; (2) the second-highest importance (with average importance values less than half of the first group) includes criteria related to quantity, delivery time and offer validity duration; (3) the third group includes delivery mode, delivery terms and payment terms, with important values that are relatively low

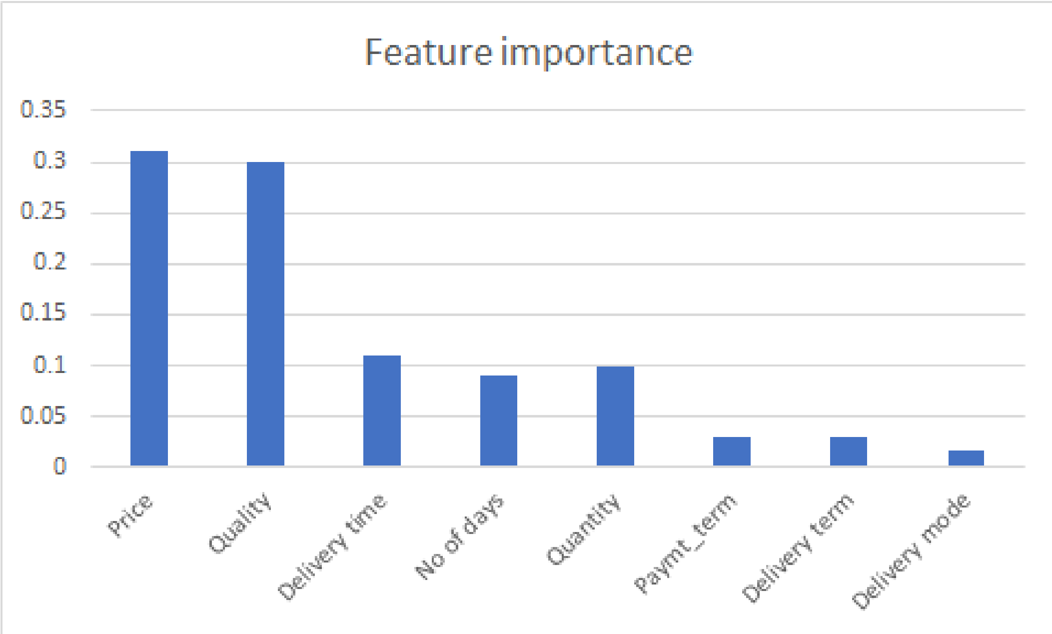

Figure 6.9: Feature importance values using DT

compared to the rest.

The end result of this step is a reduced feature set that only includes these top 8 features and the numerical weights attached to each of them. These two characteristics (limited number of features and calculated weights) are a prerequisite to the final phase. It would be next to impossible to run a meaningful MCDM process relying on weights for the full original set of 21 features. At this point, the ML component of the hybrid has completed its purpose of reducing complexity and facilitating MCDM approaches. The decision as to which supplier to select is left to the MCDM component.

6.2.3 Applying AHP to Rank and Select Suppliers

For the final phase of the approach, out of the wide variety of different MCDM approaches, we opted for AHP. This decision was primarily based on the requirements of the stakeholders involved in the case study, who had prior experience using AHP. Feature importance values were used as weights for each of the eight criteria of the reduced dataset, and these weights were multiplied by the criteria values with the products summed up to yield a score for each supplier. An important thing to highlight at this point is the difference between positive and negative features. Positive features are those that increase as the quality of a supplier increases, whereas negative features represent the opposite. An example of a positive feature is offer duration, as offers with higher duration are preferable and prioritised. Price is an example of a negative feature, as the supplier with the lowest price would be selected, all other things being equal. This meant that in calculating supplier scores in applying AHP, an appropriate positive or negative sign was used for each weight. After calculating scores for each supplier that makes an offer for each requisition, the selected supplier (and the outcome of the ML+MCDM approach) is the supplier with the highest score.

The following example shows how to calculate the best supplier.

Order number 20821 has four suppliers to offer bids as in table 6.2

OrderNo	SupNo	price	delivtime	quality	#days	qty	delivterm	payterm	delivmode
201821	2839	6605.78	1	1	62	1	1	1	1
201821	3905	3757.5	6	1	91	1	1	1	1
201821	3365	3522.45	6	1	88	1	1	1	1
201821	1494	6839.99	27	1	60	1	1	1	1

Table 6.2: Decision matrix of transaction no 201821

Then, the best and worst values were calculated for each criterion, as shown in table 6.3. Again, positive and negative criteria are treated differently, with the lowest value being the best for negative features and the highest value being the best for positive ones.

OrderNo	SupNo	price	delivtime	quality	#days	qty	delivterm	payterm	delivmode
201821	2839	6605.78	1	1	62	1	1	1	1
201821	3905	3757.5	6	1	91	1	1	1	1
201821	3365	3522.45	6	1	88	1	1	1	1
201821	1494	6839.99	27	1	60	1	1	1	1
Best values		3522.45	1	1	91	1	1	1	1
Worst values		6839.99	27	1	60	1	1	1	1

Table 6.3: The best and worst of transaction no 201821

The next step, illustrated in table 6.4, is to normalise the preferred rankings for each alternative so that they all fall on the same scale of [0,1].

Normalisation for positive criteria follows equation 6.1, while equation 6.2 is for negative criteria, with i referring to the criterion and j to the supplier.

$$x_{ij_{norm}} = \frac{x_{ij} - min(x_{ij})}{max(x_{ij}) - min(x_{ij})} \tag{6.1}$$

$$x_{ij_{norm}} = \frac{max(x_{ij}) - x_{ij}}{max(x_{ij}) - min(x_{ij})} \tag{6.2}$$

The data after normalisation is shown in table 6.4 :

SupNo	price	delivtime	quality	#days	qty	delivterm	payterm	delivmode
2839	0.071	1	1	0.065	1	1	1	1
3905	0.929	0.808	1	1	1	1	1	1
3365	1	0.808	1	0.903	1	1	1	1
1494	0	0	1	0	1	1	1	1

Table 6.4: Normalised values for transaction no 201821

The next step is to multiply normalised data with the weights of each criterion. The weights used are the feature importance in figure 6.9, shown in table 6.5. By multiplying each criterion's weights by the preferred scores for that criterion, the weighted normalised ratings were obtained as shown in the table 6.6.

Feature	**Importance**
Price	0.31
Quality	0.3
Deliv time	0.11
No of days	0.09
Quantity	0.1
Paymt_term	0.03
Delivery_term	0.03
Delivery_mode	0.016

Table 6.5: Feature importance values used as weights

SupNo	price	delivtime	quality	#days	qty	delivterm	payterm	delivmode
2839	0.0219	0.11	0.3	0.0058	0.1	0.03	0.03	0.016
3905	0.2880	0.0888	0.3	0.09	0.1	0.03	0.03	0.016
3365	0.31	0.0888	0.3	0.0813	0.1	0.03	0.03	0.016
1494	0	0	0.3	0	0.1	0.03	0.03	0.016

Table 6.6: Weighted normalised transaction no 201821

The last step is to compute the total score for each supplier and rank them based on these scores, the highest score is ranked first, as shown in table 6.7.

order no	**supplier no**	**total**	**Rank**
201821	2839	0.624	3
201821	3905	0.953	2
201821	3365	0.966	1
201821	1494	0.486	4

Table 6.7: Ranked suppliers for transaction no 201821

Results are also shown in figure 6.10, with supplier #3365 achieving the highest score of 0.966 and supplier #3905 following closely with a score of 0.953.

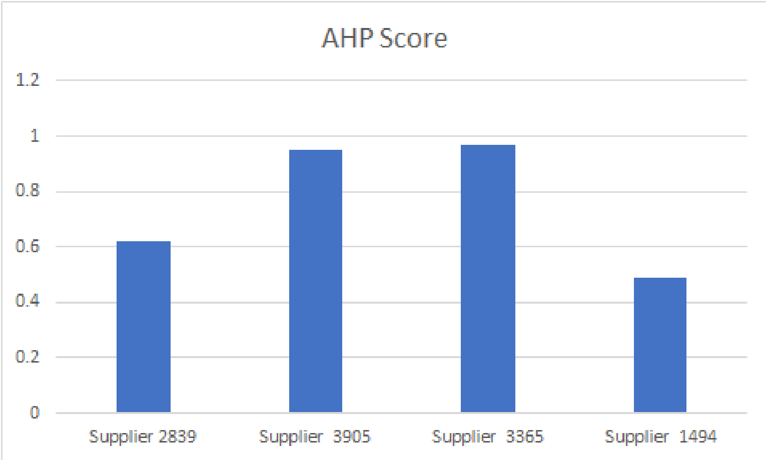

Figure 6.10: Ranking suppliers using AHP

6.2.4 Validating DT+AHP Results

To validate the results of implementing the DT+AHP model, its outputs were compared with the actual results of supplier selection made by the company's purchasing department. The performance of approaches that include only machine learning was also explored to determine how the proposed framework fares in comparison, applying standard binary classification to determine whether a supplier is selected or not and again comparing the classifier output to the actual decision made by the purchasing department. This allows us to quantify the trade-off between aiming for familiarity (through the use of MCDM approaches) and performance.

In the results shown below, a true positive (TP) is when the highest-scored supplier is the one that was actually selected, while a false positive (FP) is

when the highest-scored supplier is not the one actually selected. A true negative (TN) then is when a supplier is not given the highest score and is indeed one of those not selected, while a false negative (FN) is when the supplier that was actually selected is given a score that is not the highest. Metrics follow the standard definitions, which are listed here for convenience:
$Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$;
$Precision = \frac{TP}{TP+FP}$;
$Recall = \frac{TP}{TP+FN}$;
$F_1 = 2 * \frac{Precision * Recall}{Precision + Recall}$.
The number of instances matched between the framework and the actual decisions made by the company in the case study is as follows:

- TP: The same supplier was selected as the top-ranked in 1881 samples out of 2274 (82.7%)
- TN: From 3891 out of 4285 instances where the Libyan company did not select the supplier as the best supplier; also, the proposed model resulted in this supplier being ranked second or lower (90.8%)
- FP: 394 cases out of 4285 (9.2%), the supplier was not selected by the Libyan company but ranked first by the proposed model
- FN: 393 cases out of 2274 (17.3%), the Libyan company selected the best supplier but did not rank first by the proposed model

Results in Table 6.8 show that performance is comparable across both the hybrid approach and standard machine learning approaches, with the hybrid taking the lead in all metrics. In terms of accuracy, the hybrid achieved the

Approach	Accuracy	Precision	Recall	F_1
Hybrid DT+AHP	**0.868**	**0.90**	**0.827**	**0.862**
DT	0.764	0.76	0.76	0.74
RF	0.757	0.744	0.754	0.739
SVM	0.746	0.755	0.759	0.741
ANN	0.737	0.72	0.728	0.718
KNN	0.730	0.757	0.758	0.735

Table 6.8: DT+AHP vs. pure ML

highest result, followed by DT and RF, while KNN was the last. Precision results, the hybrid still has the highest result with 90%, followed by DT, SVM and KNN with similar results, while ANN has the lowest results with 72%. The recall results are very similar across all methods, with the hybrid achieving lower values than the precision results. The hybrid still has the highest result with 82.7%, followed by DT, SVM and KNN with similar results, while ANN has the lowest results.

Finally, in terms of the F1 score, the hybrid model demonstrated superior performance, achieving an impressive 86.2%. All other machine learning (ML) models showcased similar performance, with F1 scores hovering around 74%, except for ANN, which yielded a slightly lower score of 72%.

It is relatively simple to score the model using binary classification using metrics like precision, recall, and $F1$ score. But, when all classes must be treated equally, we use macro average scores to compare the classifier's overall performance to the most prevalent. Table 6.9 provides more information on how different approaches behave in the two different classes (suppliers selected and not selected) in terms of precision, recall and F_1 score. Differences in results per class are due to moderate class imbalance, with roughly two-thirds of sam-

Classifier	Class/Average	Precision	Recall	F_1
Hybrid DT+AHP	Class 0	**0.88**	0.91	**0.87**
	Class 1	**0.91**	**0.83**	**0.87**
	Macro Avg	0.87	0.87	0.87
	Weighted Avg	0.89	0.88	0.87
DT	Class 0	0.78	0.92	0.84
	Class 1	0.71	0.42	0.53
	Macro Avg	0.74	0.67	0.69
	Weighted Avg	0.76	0.76	0.74
SVM	Class 0	0.76	**0.93**	0.83
	Class 1	0.69	0.35	0.47
	Macro Avg	0.74	0.67	0.69
	Weighted Avg	0.76	0.76	0.74
KNN	Class 0	0.77	0.90	0.82
	Class 1	0.65	0.40	0.49
	Macro Avg	0.74	0.67	0.69
	Weighted Avg	0.76	0.76	0.74
RF	Class 0	0.79	0.86	0.82
	Class 1	0.62	0.49	0.55
	Macro Avg	0.71	0.68	0.69
	Weighted Avg	0.74	0.75	0.74
ANN	Class 0	0.76	0.90	0.82
	Class 1	0.63	0.39	0.49
	Macro Avg	0.69	0.66	0.67
	Weighted Avg	0.72	0.73	0.72

Table 6.9: Five ML models vs. hybrid model for the two classes

ples belonging to class 0 (suppliers not selected) and one-third belonging to class 1. Comparing the hybrid model with five ML models, it seems that most ML approaches are not as strong as the hybrid approach. It is important to note, however, that precision results for the hybrid are somewhat more constant between classes, with a slightly lower class 1 recall and the same $F1$ score for both classes.

6.3 FUCOM+TOPSIS+DT Results

In this second hybrid approach, FUCOM and TOPSIS are applied first to calculate criteria weights and supplier scores before using a DT classifier to make the final decision of choosing a supplier.

6.3.1 Identifying Decision Criteria

In addition to the two main datasets, a questionnaire was created and distributed to 9 decision-makers and experts in the case study company to identify which criteria they employ in the supplier selection process. Reliability analysis for questionnaire results was performed using Cronbach's alpha in the SPSS software package.

Different opinions of the questionnaire participants were taken into account by calculating a total score that quantifies the importance placed collectively by them for each criterion and setting a threshold for the lowest total score. If a criterion's total score did not manage to meet the threshold, it was left out. Consequently, the criteria for evaluating suppliers were established. These criteria are as follows: price (noted as C1 in the calculations that follow next), delivery time (C2), quantity (C3), quality (C4), delivery term (C5), delivery mode (C6), number of days (C7), payment term(C8), and action (C9).

6.3.2 Weighting Criteria Using FUCOM

Applying the FUCOM method relied on the experience of the company's purchasing managers, who were used as experts to identify and prioritise selection

Criteria	C1	C4	C2	C5	C3	C9	C7	C8	C6
Weights	1	2.5	3	4	4.5	6	6.5	8	9

Table 6.10: Weights of criteria by FUCOM

criteria and make pairwise comparisons to evaluate each supplier regarding each criterion. A step-by-step explanation of applying FUCOM is provided next.

Based on the criteria scores calculated through processing questionnaire results, selection criteria were ranked as follows:
C1> C4 > C2 > C5 >C3>C9>C7>C8>C6. The experts were asked to perform a pairwise comparison of criteria. On the basis of a scale [1,9], a comparison was conducted with respect to the first-ranked criterion. Table 6.10 shows the weights of the criteria for all of the ranked criteria.

In the next step, comparative priorities are calculated as below:

$\varphi_{C_1/C_4} = 2.5/1.0 = 2.5$

$\varphi_{C_4/C_2} = 3.0/2.5 = 1.20$

$\varphi_{C_2/C_5} = 4.0/3.0 = 1.33$

$\varphi_{C_5/C_3} = 4.5/4.0 = 1.13$

$\varphi_{C_3/C_9} = 6.0/4.5 = 1.33$

$\varphi_{C_9/C_7} = 6.5/6.0 = 1.08$

$\varphi_{C_7/C_8} = 8.0/6.5 = 1.23$

$\varphi_{C_8/C_6} = 9.0/8.0 = 1.13$

The final model for determining the coefficients can be defined as:

$\min \chi$

$$\text{s.t} \begin{cases} \left|\frac{\omega_1}{\omega_4} - 2,5\right| \leq \chi_1 \left|\frac{\omega_4}{\omega_2} - 1,20\right| \leq \chi_1 \left|\frac{\omega_2}{\omega_5} - 1,33\right| \leq \chi \\ ,\left|\frac{\omega_5}{\omega_3} - 1,13\right| \leq \chi_1 \left|\frac{\omega_3}{\omega_9} - 1.33\right| \leq \chi_1 \left|\frac{\omega_9}{\omega_7} - 1,08\right| \leq \chi \\ \left|\frac{\omega_7}{\omega_8} - 1,23\right| \leq \chi_1 \left|\frac{\omega_8}{\omega_6} - 1,13\right| \leq \chi, \left|\frac{\omega_1}{\omega_2} - 3,0\right| \leq \chi \\ \left|\frac{\omega_3}{\omega_7} - 1,60\right| \leq \chi_1 \left|\frac{\omega_2}{\omega_3} - 1,5\right| \leq \chi_1 \left|\frac{\omega_5}{\omega_9} - 1,5\right| \leq \chi_1 \\ \sum_{j=1}^{9} \left|\frac{\omega_9}{\omega_8} - 1,33\right| \leq \chi_1 \left|\frac{\omega_7}{\omega_6} - 1,8\right| \leq \chi \end{cases}$$

Solving this model yields the final values of the weight coefficients as well as the deviation from full consistency (DFC) of the findings. Table 6.11 and figure 6.11 show the obtained weights of the criteria by solving the model using the Microsoft Excel solver. Based on the findings, it can be determined that price is the most essential criterion, followed by quality.

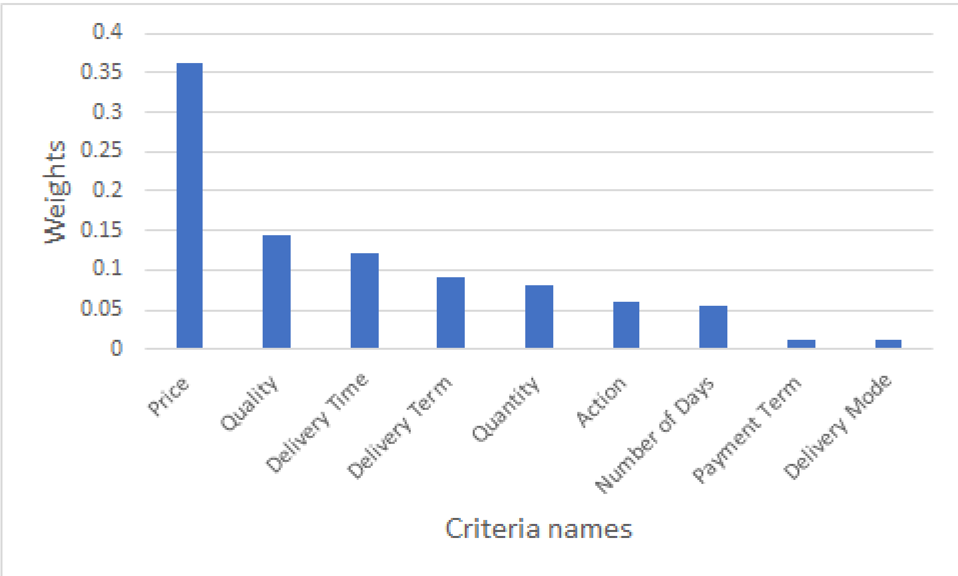

Figure 6.11: Criteria's weights by FUCOM method

Criterion	Weight
Price	0.362
Quality	0.145
Delivery Time	0.121
Delivery Term	0.091
Quantity	0.08
Action	0.06
Number of Days	0.056
Payment Term	0.011
Delivery Mode	0.0109

Table 6.11: Weights of criteria by FUCOM method

6.3.3 Ranking Suppliers Using TOPSIS

Relative standard weights were calculated in the previous step using the FUCOM method. Results from FUCOM are then used as inputs to the TOPSIS method to evaluate and rank suppliers. The rankings were generated by sorting suppliers in the order they were found. The following steps are used to apply the TOPSIS method:

- Determine the type of the criteria: cost or benefits criteria (or, as mentioned earlier, negative and positive ones)
- Normalisation method is applied to ensure all data values are between 0 and 1.
- Multiply the normalised data by the weights calculated by the FUCOM method in the section
- Calculate each supplier's score by an aggregation function
- Rank the suppliers based on their score.

An example of applying TOPSIS is shown below for order number 20821, which has four supplier offers as shown in table 6.12

OrderNo	SupNo	price	delivtime	quality	#days	qty	delivterm	payterm	delivmode	action
201821	2839	6605.78	1	1	62	1	1	1	1	1
201821	3905	3757.5	6	1	91	1	1	1	1	1
201821	3365	3522.45	6	1	88	1	1	1	1	1
201821	1494	6839.99	27	1	60	1	1	1	1	1

Table 6.12: Decision matrix of transaction no 201821

Then, the best and worst values are calculated for each criterion, as shown in table 6.13. As with AHP, positive and negative criteria are treated differently, with the lowest value being the best for positive features and the highest for negative.

OrderNo	SupNo	price	delivtime	quality	#days	qty	delivterm	payterm	delivmode	action
201821	2839	6605.78	1	1	62	1	1	1	1	1
201821	3905	3757.5	6	1	91	1	1	1	1	1
201821	3365	3522.45	6	1	88	1	1	1	1	1
201821	1494	6839.99	27	1	60	1	1	1	1	1
V_j+		3522.45	1	1	91	1	1	1	1	1
V_j-		6839.99	27	1	60	1	1	1	1	1

Table 6.13: The best and worst values of transaction no 201821

Then, normalised values are calculated for each performance (supplier/offer combination) x_{ij} using Equation 6.3.

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} \qquad i = 1, \ldots, m; j = 1, \ldots, n. \tag{6.3}$$

The data after normalisation is shown in table 6.14.

In the next step, the normalised decision matrix is multiplied with the weights calculated by the FUCOM method, associated with each of the criteria in the

SupNo	price	delivtime	quality	#days	qty	delivterm	payterm	delivmode	action
2839	0.6108	0.0353	0.5	0.4047	0.5	0.5	0.5	0.5	0.5
3905	0.3475	0.2119	0.5	0.5940	0.5	0.5	0.5	0.5	0.5
3365	0.3257	0.2119	0.5	0.5744	0.5	0.5	0.5	0.5	0.5
1494	0.6325	0.9534	0.5	0.3917	0.5	0.5	0.5	0.5	0.5

Table 6.14: Normalised decision matrix

second phase. For convenience, the weights are repeated in table 6.15.

price	delivtime	quality	#days	qty	delivterm	payterm	delivmode	action
0.362	0.121	0.145	0.056	0.08	0.091	0.011	0.0109	0.06

Table 6.15: Feature weights

By multiplying each criterion's weight by the data for that criterion, the weighted normalised ratings were obtained as shown in table 6.16.

SupNo	price	delivtime	quality	#days	qty	delivterm	payterm	delivmode	action
2839	0.2211	0.0043	0.0725	0.0227	0.04	0.0455	0.0055	0.0055	0.03
3905	0.1258	0.0256	0.0725	0.0333	0.04	0.0455	0.0055	0.0055	0.03
3365	0.1179	0.0256	0.0725	0.0322	0.04	0.0455	0.0055	0.0055	0.03
1494	0.2290	0.1154	0.0725	0.0219	0.04	0.0455	0.0055	0.0055	0.03

Table 6.16: Weighted normalised transaction no 201821

The following step deviates from other similar techniques in that TOPSIS requires the ideal and anti-ideal alternatives to be determined. This is calculated using the maximum and minimum values for each criterion, shown in tables 6.17 and 6.18.

The positive ideal alternative (E+) is calculated by using Equation 6.4.

$$E^{+} = \{v_1^*, v_2^*, ...v_n^*\} = (\{\max_j v_{ij} | i \in I \quad , (\min_j v_{ij} | i \in I'\} \quad i = 1, \ldots, m; j = 1, \ldots, n. \tag{6.4}$$

price	delivtime	quality	#days	qty	delivterm	payterm	delivmode	action
0.1179	0.0043	0.0725	0.0333	0.04	0.0455	0.0055	0.0055	0.00545

Table 6.17: Ideal TOPSIS values

price	delivtime	quality	#days	qty	delivterm	payterm	delivmode	action
0.2290	0.1154	0.0725	0.0219	0.04	0.0455	0.0055	0.0055	0.00545

Table 6.18: Worst TOPSIS values

The negative ideal alternative (E-) is calculated by using the equation 6.5.

$$E^- = \left\{v_1^-, v_2^-, ..., v_n^-\right\} = \{(\mathrm{min}_j v_{ij} | i \in I \quad , (\mathrm{max}_j v_{ij} \quad | i \in I'\} \, i = 1, \ldots, m; j = 1, \ldots, n. \tag{6.5}$$

In the penultimate step, separation measures are calculated. The euclidean distance to the positive ideal solution (PIS) is calculated using equation 6.6 and to the negative ideal solution(NIS) using equation 6.7 for each alternative, with the results shown in table 6.19 and table 6.20, respectively.

$$S_i^+ = \sqrt{\sum_{i=1}^{n} (v_{ij} - v_j^*)^2}, i = 1, \ldots, m; j = 1, \ldots, n \tag{6.6}$$

$$S_i^- = \sqrt{\sum_{i=1}^{n} (v_{ij} - v_j^-)^2}, i = 1, \ldots, m; j = 1, \ldots, n \tag{6.7}$$

Supplier number	PIS
2839	0.1038
3905	0.0228
3365	0.0214
1494	0.1575

Table 6.19: Positive Ideal Solution (PIS)

Supplier number	NIS
2839	0.1114
3905	0.1372
3365	0.1431
1494	0.0000

Table 6.20: Negative Ideal Solution (NIS)

The final step is to calculate the relative closeness of each alternative to the ideal solution using equation 6.8. Results are shown in table 6.21.

$$TOPSIS = \frac{S_i^-}{S_i^- + S_i^+}, i = 1, \ldots, m \tag{6.8}$$

Supplier number	relative closeness
2839	0.517693
3905	0.857689
3365	0.869983
1494	0

Table 6.21: Relative closeness of each alternative

These values are the output of TOPSIS that is used to rank suppliers, with the supplier with the highest score (highest relative closeness to the ideal) ranked first. Ranks are shown in table 6.22.

order no	**supplier no**	**total**	**Rank**
201821	2839	0.517693	3
201821	3905	0.857689	2
201821	3365	0.869983	1
201821	1494	0	4

Table 6.22: Ranked suppliers for transaction no 201821

Supplier 3365 has the greatest score and ranks first.

Results are also illustrated in figure 6.12, with supplier no 3365 achieving the highest score of 0.869983.

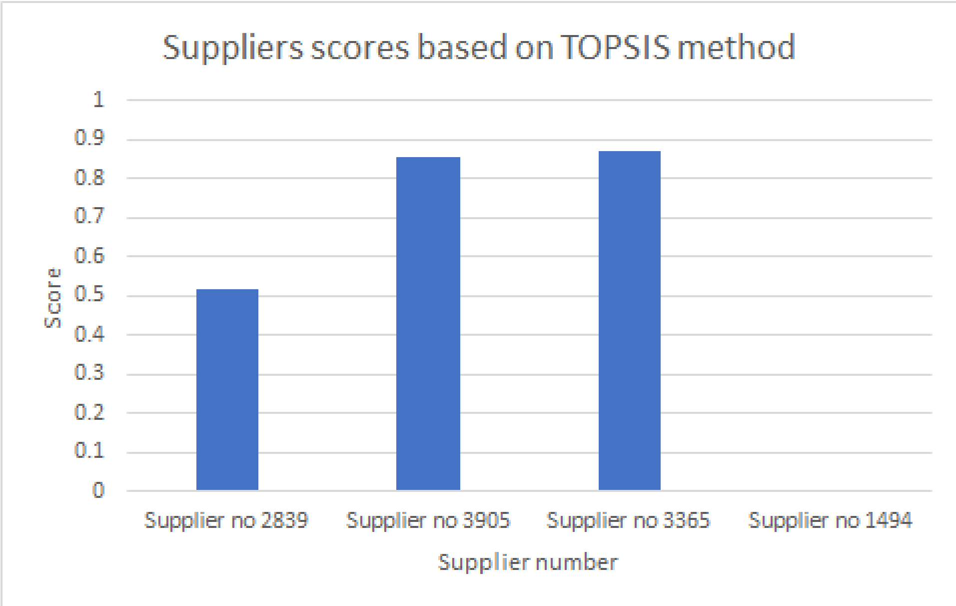

Figure 6.12: Ranking suppliers by TOPSIS

6.3.4 Implementing and Evaluating ML algorithms

In the final phase of applying the FUCOM+TOPSIS+DT hybrid, the same criteria used in FUCOM are used for the binary classification of a supplier as selected or not selected. Also, apart from DT, we also considered the same alternative ML algorithms to confirm DT is the most appropriate choice: SVM, KNN, RF and ANN.

Before applying any ML classifier, the output of TOPSIS needs to be converted from continuous supplier scores (as expected in regression problems) to a binary result (for classification between two classes). A threshold of 0.5 is assumed, with all suppliers scoring 0.5 or above classified in the “good” class, and the rest (less than 0.5) classified in the "bad" class. As a result, a new target value named supplier categorical is included in the dataset, instead of the TOPSIS supplier scores. As before, 10-fold cross-validation was used in all cases, with a 70%-30% train-test split. The same metrics as previously are used to evaluate results: accuracy, training time, $F1$ score, recall and precision.

Figure 6.13 shows accuracy and training time values across all five ML algorithms. In terms of accuracy, RF achieves the highest result, closely followed by DT, with SVM being the least accurate. In terms of training time, DT outperforms the other models with only 0.006 seconds, while the MLP classifier (ANN) has the longest training time with 1.122 seconds.

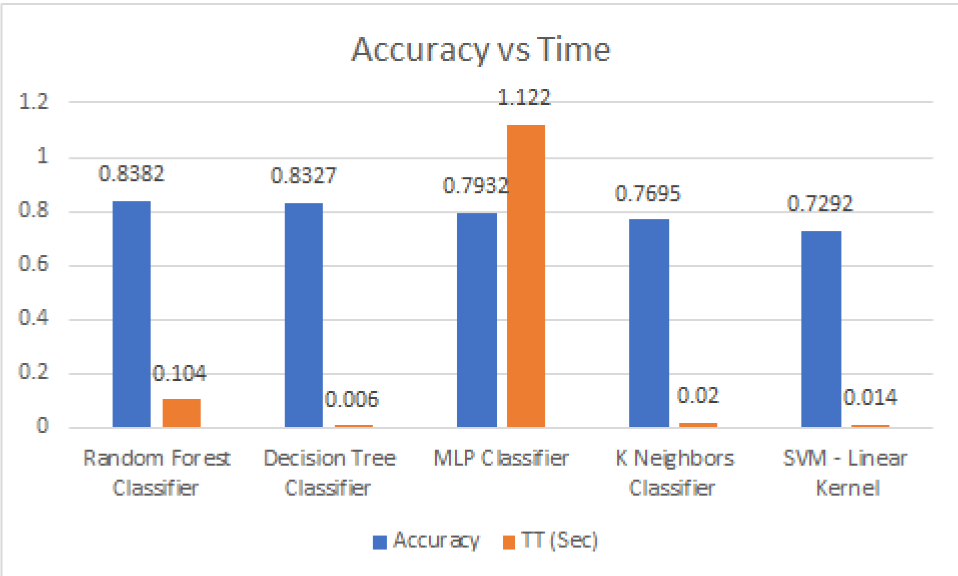

Figure 6.13: TOPSIS-based ML model accuracy vs time

In terms of precision, as shown in Figure 6.14, RF was the highest score at 74%, followed by DT at 72%, while SVM had the least score with only 54%.

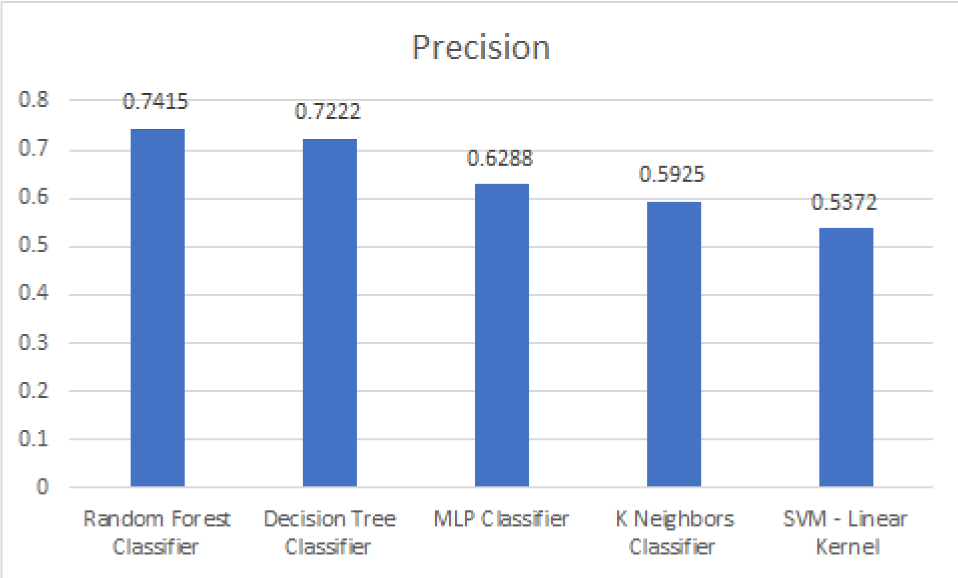

Figure 6.14: Precision for five ML models of the second hybrid

In terms of recall, as shown in figure 6.15, MLP achieves the highest result, followed by KNN and the DT, with SVM and RF lagging slightly.

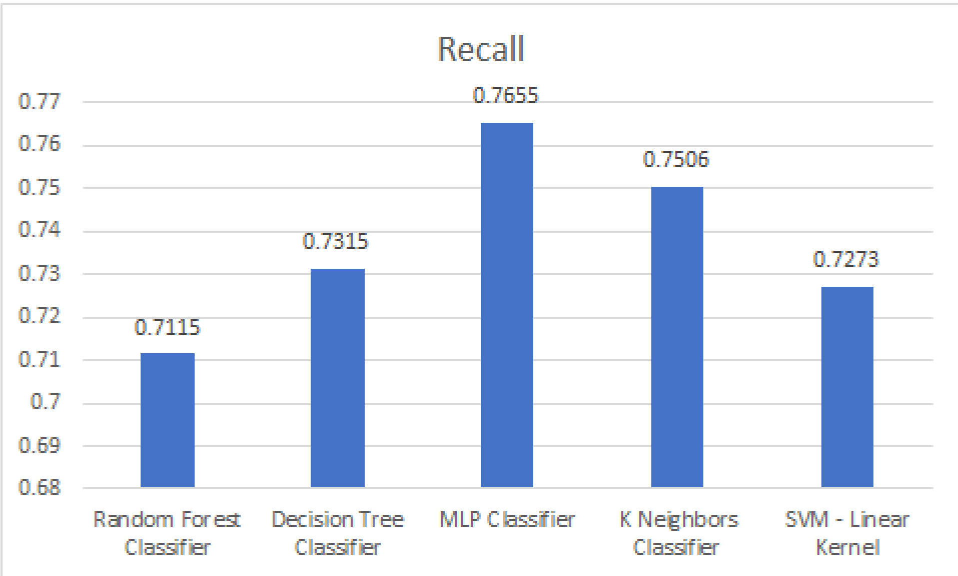

Figure 6.15: Recall for five ML models of the second hybrid

Lastly, $F1$ score results are shown in figure 6.16, with RF and DT achieving similar scores, 72.57% and 72.51%, respectively. In conclusion, DT and RF

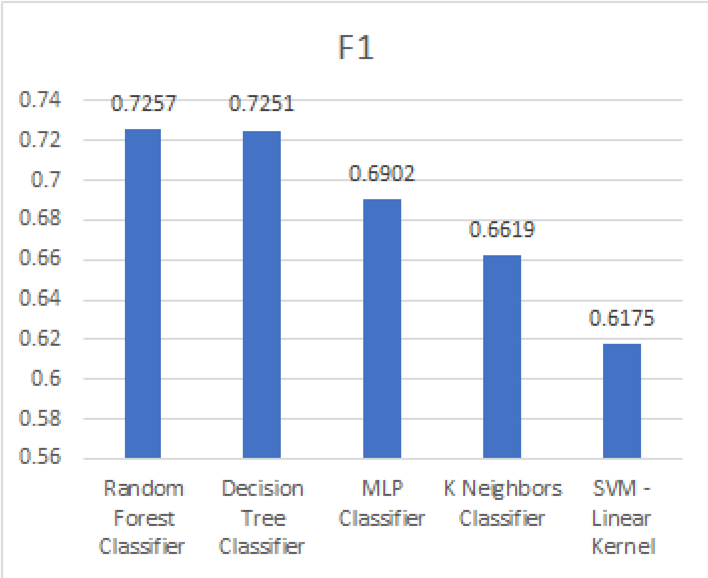

Figure 6.16: $F1$ score for five ML models of the second hybrid

achieved the best results in most evaluation metrics, accuracy, precision and $F1$ score, except recall, where ANN outperforms the rest. In contrast, training time is the highest for ANN and lowest for DT. While RF outperforms DT in terms of precision, it takes slightly more to train models, given the higher complexity of multiple trees. These results confirm a trade-off between algorithm performance and model training time, which should be taken into account when choosing ML techniques for the MCDM+ML hybrid approach.

These metrics are important in evaluating ML models, but it is not the only thing that counts. Interpretability has a higher priority when selecting an ML model in supplier selection and evaluation, which enables purchasing managers and stakeholders to comprehend how the ML model operates and make informed decisions based on its results which may have major effects on the company's financial and operational performance. Hence, while DT is second in terms of accuracy, precision and F1 score, ultimately its inherent interpretability and fast classification time carry more importance, ensuring that not only the FUCOM and TOPSIS parts of the model can be understood by procurement staff, but also the final classification phase that produces the selected supplier.

6.4 Discussion

In this section, we discuss the comparative evaluation results presented in this chapter in terms of performance, explainability and applicability criteria. Section 6.4.1 discusses the ability of the framework to achieve performance in terms of selecting the right supplier that outperforms the standard approaches that rely purely on machine learning, including an assessment of the impact of class imbalance.

Then, Section 6.4.2 summarises the novelty of the FUCOM+TOPSIS combination in the second hybrid. Section 6.4.3 discusses the practical implications of the presented models and results by conducting an applicability assessment. Finally, Section 6.4.4 summarises the discussion by directly comparing the proposed hybrids to other hybrids previously discussed in related work.

6.4.1 Performance

As evidenced by the results of hybrid method one as in Table 6.8, the hybrid framework has achieved better performance than five machine learning approaches, DT, RF, ANN, SVM and KNN. This is partly explained by the fact that a DT algorithm was used for feature selection in the machine-learning component of the hybrid. However, results also support the conclusion that including the MCDM component does not have a negative impact but a positive impact; it has shown improvement in the result in all evaluation metrics, accuracy, precision, recall and $F1$ score, with an average of over 14% of the other ML models, with a significant increase in precision. This is quite important as it shows that, at least in this particular experiment, there is no need to make significant compromises in terms of performance to achieve an increased level of familiarity with the hybrid's components. Note that opting for other, less interpretable machine learning approaches, even if given higher performance results, as explained in chapter 3, this would not be in line with the rationale behind the proposed framework. In the second hybrid method, as seen in Figure6.13, which utilises the MCDM to weigh the criteria and rank the suppliers, the accuracy of all ML models (except SVM) increases compared to the results of using the non-hybrid method, ML only, as in method one as in table 6.1. That is why using MCDM to rank the suppliers improve in terms of the performance of ML models. In terms of the effects of class imbalance, as illustrated in the classification reports in table 6.9, most ML approaches seem to be less robust than the hybrid approach. However, it is worth noting that precision results for the hybrid are relatively more consistent between classes, with a slightly lower class 1 recall and the same $F1$ score for both classes. It can be argued that this robustness against imbalance across most approaches may be due to the fact that the dataset used for the experiments is only moderately imbalanced, with one class containing two times as many samples as the other (imbalance ratio 1:2). This is why the model did not apply any techniques to address imbalance (such as over- or under-sampling), which would be more useful in cases where there is a more severe imbalance (e.g. an imbalance ratio of 1:5 or 1:10).

6.4.2 Novel MCDM combinations

In terms of the second hybrid, it is worth highlighting the novelty of the MCDM combination included. FUCOM is unique in that it has the ability to authenticate the results by defining the deviation from complete consistency (DFC) of comparisons and recognising the transitivity of criteria when comparing them in pairs. FUCOM, in contrast to previous subjective models, has shown only minimal deviations from optimal values in the weights of criteria produced from the obtained values of the weights of criteria.

The TOPSIS technique, on the other hand, is based on ranking the alternatives in regard to ideal values and on a thoroughly rational and reasonable process for applying the procedure. Because the approach is relatively flexible and straightforward, it may also be used to solve a variety of other multi-criteria problems. On the one hand, the TOPSIS approach has been used to evaluate possible suppliers, whilst the FUCOM method was used to determine the relevance of the weight values of the criteria. Experiments confirm that this combination is quite capable of achieving these goals.

6.4.3 Practical Implications

This research utilised a real-world case study involving an oil and gas company to produce more accurate results for the company's users, particularly the decision-makers of comparable companies.

An often neglected aspect of innovative approaches to supplier selection is the extent to which these approaches can be applied in real-world settings and adopted by purchasing staff in their day-to-day operations. In the case of the proposed hybrid framework, applicability was a design requirement from the start in terms of familiarity and explainability of the approaches used. As highlighted in the discussion so far, experiments support the argument of applicability for the hybrid frameworks as it not only achieves satisfactory performance but does so while allowing for end-to-end explainability through the combination of interpretable feature selection and supplier selection using a technique familiar to purchasing staff (AHP) as in the first hybrid, DT+AHP.

If a machine learning model performs well without explaining its conclusion, classification accuracy is an incomplete metric for most real-world activities. Machine learning models can only be debugged and inspected if they are interpretable. An interpretation of an incorrect prediction facilitates comprehension of the error's root cause and guides how to fix it (Molnar, 2020). Using decision trees, which are easy to interpret by decomposing the decision path into feature-specific components, this allows interpreting predictions (Molnar, 2020; Rokach, 2016).

From a practical perspective, by following the DT+AHP hybrid, a supplier selection stakeholder can understand the process of using feature importance values to select particular criteria, setting criteria weights and applying AHP to rank suppliers and select the top-ranked one. This allows for an end-to-end explainable result detailing how particular selection criteria are prioritised and weighted, followed by a transparent calculation of scores for every supplier and the subsequent ranking and selection of the most appropriate one. The same holds for the FUCOM+TOPSIS+DT hybrid, where not only the process of calculating criteria weights and supplier scores is transparent, but also the classification of a supplier into either good or bad can be easily interpreted through the produced DT model.

An additional benefit of retaining an MCDM approach as part of the supplier selection process is that these approaches are capable of ranking suppliers rather than just selecting the most appropriate supplier, which is what the standard machine learning based binary classification approach can provide. This also increases applicability as it is closer to real-world needs, where it may be necessary to look to the second-best supplier if there is an issue (e.g. the best supplier withdrawing their offer). Instead of calculating predictions again, the supplier ranked second based on the MCDM approach results can be selected.

On the other hand, applicability (and the likelihood of adoption) may be limited depending on the choices made for the two main components of the hybrid. The level of interpretability offered by the machine learning component may not be enough for the target users, or there may even be a reluctance to use any machine learning technique. Also, other purchasing staff have different knowledge of MCDM, so the chosen MCDM approach may not be preferable to some, or the fact that weights are automatically generated rather than agreed upon by consensus may be problematic.

Finally, as is common in every approach that involves data-driven AI, results depend highly on the availability of data of sufficient quantity and quality. These limitations can be addressed by designing the hybrid in collaboration with supplier selection stakeholders, ensuring there is agreement at major decision points and that the result is tailored to the needs of each company. As argued by (Baryannis, Dani, & Antoniou, 2019), effective synergies between different experts (AI and supplier selection in this case) is paramount in achieving success in such multidisciplinary settings.

6.4.4 Comparison to Related Work

In this section Table 3.3 in Chapter 3 is revisited, in order to compare against state of the art and further clarify the benefits of the proposed approaches.

A common weakness of state of the art hybrid supplier selection is the use of black box models, like ANN, This is true for the majority of works in Table 3.3 (9 out of 14). As explained in the previous section, both proposed approaches do not suffer from this issue, as they rely on DT models which are interpretable. However, the proposed hybrids go beyond the interpretability of ML models, as they combine MCDM approaches that are inherently explainable, resulting in supplier selection outcomes that can be fully understood.

A further weakness that was identified for several works in Table 3.3 is limited validation that focuses on very small datasets or no validation based on a real world case study. In contrast, this thesis includes a thorough real-world case study using two large datasets containing thousands of suppliers and offers. This allows for a comprehensive validation of the capabilities of the two hybrids, ensuring that they can operate successfully in high complexity environments.

Finally, the hybrid models in this thesis can support both qualitative and quantitative criteria, unlike the work of R. Kuo et al. (2010), while there is no need for a full specification of goals and priorities like in the case of Goal Programming (Fallahpour et al., 2016). All in all, the proposed hybrid supplier selection methods provide a combination of desirable characteristics that are not supported, to the best of our knowledge, by any other hybrid in literature.

Chapter 7

Conclusion and Future Work

7.1 Conclusion

In this thesis, ML and MCDM were studied in relation to the supplier selection problem. Most recent studies have focused on improving supplier selection through the use of MCDM, ML, or their integrated models; nevertheless, as discussed at the end of chapter 3, there are certain drawbacks. Based on this analysis, it is obvious that the combination of MCDM to evaluate the criteria and rank the alternatives, along with ML for calculating the feature importance and as a prediction model, can yield promising approaches for inference in the supplier selection domain. In this thesis, two hybrid approaches are introduced, applied and evaluated in a case study of oil and gas companies based in Libya.

There are two ways to implement these integrated methods. In the first approach, ML models (in particular, DT) are first used to calculate the feature

importance based on the historical transactions of the Libyan company. Then, MCDM techniques (particularly AHP) are used to evaluate and rank the suppliers. In the second approach, MCDM techniques are used first to calculate the weights for each criterion and rank the alternatives, particularly using FUCOM for weighting the evaluation criteria and then TOPSIS for ranking the suppliers. Then, a binary classifier using ML (in particular, DT) is applied to determine whether a supplier is selected or not.
In general, when there are a large number of data points, MCDM theories are less suitable. Integration with machine learning techniques, on the other hand, reduces the number of data points required and extends the advantages of predictive and intelligent decision support to the supplier selection problem. Results show that the proposed models achieve high performance and afford high interpretability.

The developed approaches allow for a very quick and simple evaluation of suppliers that can also be explained to a certain extent, facilitating and enhancing the supplier selection process. This can prove beneficial to companies in any sector, and the fact that two different approaches are provided means that they can cater for a wide range of different requirements (e.g. different levels of performance or interpretability). Results can also prove quite useful to researchers at the confluence of AI and supply chains, providing insights into the successful development of solutions for intelligent supply chains.

In terms of limitations, the following should be noted: (1) ML exploration was limited to five algorithms; (2) experiments involved specific datasets provided by the company; (3) out of a wide variety of MCDM techniques, the thesis has focused on three of them, AHP, FUCOM and TOPSIS.

7.2 Managerial Implications

AI and ML are the primary drivers of data use since they both seek new advantages for companies and greatly influence how companies adopt these technologies, using historical data and the importance of considering them when selecting and evaluating suppliers. By combining ML with MCDM, the resulting methods should make it simpler to select the best suppliers if the important features are considered. In turn, the profitability of a business rises when the best suppliers are taken into account.

Managers should cautiously deploy ML models and assess model compatibility, data quality and the company's capacity to comprehend the findings. In conversations with purchasing managers while conducting the case study reported in this thesis, it was discovered that they favoured traditional MCDM solutions over alternative AI methodologies. Despite its success in other research domains, ML and other methods remain underutilised in purchasing and supply chain research (Baryannis, Validi, et al., 2019). These methodologies are also far less recognisable to industry researchers than conventional MCDM approaches. Due to this discrepancy, industrial decision-makers remain substantially uneducated on AI's potential, notably in supplier selection and evaluation. The proposed hybrids stand as evidence that the combination of MCDM with interpretable ML can lead to supplier selection that is both performant and explainable.

Decision makers are more likely to adopt the proposed hybrids compared to pure black box ML approaches and can work with academics to ensure that they make full use of the hybrid capabilities. It should be noted that this does not imply employing AI as the only decision making mechanism, but rather that decision-makers should leverage AI to uncover new information so that human decision makers can make more informed and accurate supplier evaluations to find the most suitable suppliers. It should be noted, though, that concerns about data safety, security, and openness restrict the accessibility of relevant data. Particularly in a supply chain context, stakeholders may be reluctant to provide precise information, even with their partners(Baryannis, Validi, et al., 2019).

In summary, the benefits of using the proposed hybrids in the supplier selection and evaluation process include the following:

- Using ML models with traditional MCDM techniques helps deal with large volumes of data, with the potential to identify patterns and relationships in data that humans may miss, improving results and related decision making.
- Since processes are automated, this can reduce the time and effort spent to select suppliers, also reducing costs due to errors and delays.
- Reduced prejudice: ML may reduce human bias from supplier evaluation and selection, ensuring suppliers are picked based on their capabilities.
- Insight into criteria: Through feature importance results, procurement staff can gain more insight into different selection criteria, allowing them

to focus attention and resources on the most crucial ones, making better-informed and more effective choices, lowering the risk of supplier failure, and ultimately enhancing the quality and dependability of their supply chains.

It should be noted, however, that the inclusion of ML comes with some particularities that need to be considered by purchasing managers:

- Knowledge and expertise: Applying ML algorithms may require competencies that some companies lack.
- Data quality: For ML algorithms to make appropriate conclusions, companies must guarantee their data is accurate and up-to-date. Reliable data is more important than the size of the data.
- Lack of transparency: we have strongly promoted the use of interpretable ML models; however more sophisticated black box ML models may be considered due to potential advantages in performance, which makes it more difficult for stakeholders to analyse and critique the supplier selection process.

The benefits of employing explainable supplier selection approaches, through the combination of interpretable ML and inherently understandable MCDM approaches, offers several benefits to supply chain stakeholders:

- Trust: Purchasing managers are more likely to trust a model's predictions or choices if they can comprehend how it operates.
- Transparency: Predictions and decisions are understood, and purchasing managers may utilise this knowledge to understand the criteria that affect

the model output and make intelligent decisions.

- Risk management: Evaluating and selecting suppliers entails managing various risks, including financial, delivery and quality risks. Explainable approaches may help purchasing managers make better decisions by providing insight into the many risk factors influencing supplier selection.
- Compliance: Compliance is essential throughout the supply chain management process. A clear and transparent decision-making process can aid in ensuring compliance with the company's rules and processes.
- Communication: Purchasing managers may be more capable of explaining the reasoning behind their supplier selection to interested parties, including senior management or external auditors.

7.3 Future Work

In the future, we intend to explore the following research directions: (a) investigate the applicability and performance of the framework on different case studies of varying complexity and use a range of alternative MCDM approaches for the hybrid one beyond AHP, such as TOPSIS and the Measurement Alternatives and Ranking according to COmpromise Solution (MARCOS) (Stević et al., 2020); (b) consider how the framework can be deployed for use in a multi-site company such as the one in the case study, potentially in the form of a service-based application (Baryannis et al., 2017; Baryannis & Plexousakis, 2014, 2013); and (c) examine alternative approaches to increasing interpretability of the machine learning component of the hybrid, such as neuro symbolic approaches that combine learning and reasoning (Harmelen & Teije, 2019).

References

Abdolshah, M. (2013). A review of quality criteria supporting supplier selection. *Journal of Quality and Reliability Engineering*, *2013*.

Abdulkareem, M., & Petersen, S. E. (2021). The promise of ai in detection, diagnosis, and epidemiology for combating covid-19: Beyond the hype. *Frontiers in Artificial Intelligence*, *4*, 53.

Abdulla, A., Baryannis, G., & Badi, I. (2019). Weighting the key features affecting supplier selection using machine learning techniques. In *7th International Conference on Transport and Logistics, Niš, Serbia, 6 December 2019.* (pp. 15–20). doi: 10.20944/preprints201912.0154.v1

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., & Arshad, H. (2018). State-of-the-art in artificial neural network applications: A survey. *Heliyon*, *4*(11), e00938.

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ann) modeling and its application in pharmaceutical research. *Journal of pharmaceutical and biomedical analysis*, *22*(5), 717–727.

Agrawal, N., & Kant, S. (2020). Supplier selection using fuzzy-ahp: a case study. *Trends in manufacturing processes*, 119–127.

Akarte, M., Surendra, N., Ravi, B., & Rangaraj, N. (2001). Web based casting supplier evaluation using analytical hierarchy process. *Journal of the Operational Research Society*, *52*(5), 511–522.

Akman, G. (2015). Evaluating suppliers to include green supplier development programs via fuzzy c-means and vikor methods. *Computers & industrial engineering*, *86*, 69–82.

Alavi, B., Tavana, M., & Mina, H. (2021). A dynamic decision support system for sustainable supplier selection in circular economy. *Sustainable Production and Consumption*.

Al Iqbal, M. R., Rahman, S., Nabil, S. I., & Chowdhury, I. U. A. (2012). Knowledge based decision tree construction with feature importance domain knowledge. In *2012 7th international conference on electrical and computer engineering* (pp. 659–662).

Angelov, P. P., Soares, E. A., Jiang, R., Arnold, N. I., & Atkinson, P. M. (2021). Explainable artificial intelligence: an analytical review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *11*(5), e1424.

Antoniou, G., Papadakis, E., & Baryannis, G. (2022). Mental health diagnosis: A case for explainable artificial intelligence. *International Journal on Artificial Intelligence Tools*, *31*(03), 2241003. doi: 10.1142/S0218213022410032

Awasthi, A., Govindan, K., & Gold, S. (2018). Multi-tier sustainable global supplier selection using a fuzzy ahp-vikor based approach. *International Journal of Production Economics*, *195*, 106–117.

Ayyadevara, V. K. (2018). Pro machine learning algorithms. *Berkeley: Apress*.

Azadfallah, M. (2017). Evaluation and selection of suppliers in the supply chain using the extended group promethee i procedures. *International Journal of Supply Chain and Operations Resilience*, *3*(1), 56–76.

Babaei, G., & Bamdad, S. (2021). A new hybrid instance-based learning model for decision-making in the p2p lending market. *Computational Economics*, *57*(1), 419–432.

Badi, I., & Abdulshahed, A. (2019). Ranking the libyan airlines by using full consistency method (fucom) and analytical hierarchy process (ahp). *Operational Research in Engineering Sciences: Theory and Applications*, *2*(1), 1–14.

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), 37–48.

Baryannis, G., Dani, S., & Antoniou, G. (2019). Predicting supply chain risks using machine learning: The trade-off between performance and interpretability. *Future Generation Computer Systems*, *101*, 993–1004.

Baryannis, G., Dani, S., Validi, S., & Antoniou, G. (2019). Decision Support Systems and Artificial Intelligence in Supply Chain Risk Management. In G. A. Zsidisin & M. Henke (Eds.), *Revisiting Supply Chain Risk* (p. 53-71). Springer International Publishing. doi: 10.1007/978-3-030-03813-7_4

Baryannis, G., Kritikos, K., & Plexousakis, D. (2017). A specification-based QoS-aware design framework for service-based applications. *Service*

Oriented Computing and Applications, *11*(3), 301–314. doi: 10.1007/s11761-017-0210-4

Baryannis, G., & Plexousakis, D. (2013). WSSL: A Fluent Calculus-Based Language for Web Service Specifications. In C. Salinesi, M. C. Norrie, & Ó. Pastor (Eds.), *25th International Conference on Advanced Information Systems Engineering (CAiSE 2013)* (Vol. 7908, p. 256-271). Springer Berlin Heidelberg. doi: 10.1007/978-3-642-38709-8_17

Baryannis, G., & Plexousakis, D. (2014). Fluent Calculus-based Semantic Web Service Composition and Verification using WSSL. In A. Lomuscio, S. Nepal, F. Patrizi, B. Benatallah, & I. Brandić (Eds.), *9th International Workshop on Semantic Web Enabled Software Engineering (SWESE2013), co-located with ICSOC 2013* (Vol. 8377, pp. 256–270). Springer International Publishing Switzerland. doi: 10.1007/978-3-319-06859-6_23

Baryannis, G., Validi, S., Dani, S., & Antoniou, G. (2019). Supply chain risk management and artificial intelligence: state of the art and future research directions. *International Journal of Production Research*, *57*(7), 2179-2202. doi: 10.1080/00207543.2018.1530476

Baryannis, G., Woznowski, P., & Antoniou, G. (2016). Rule-Based Real-Time ADL Recognition in a Smart Home Environment. In J. J. Alferes, L. Bertossi, G. Governatori, P. Fodor, & D. Roman (Eds.), *Rule technologies. research, tools, and applications: 10th international symposium on rules and rule markup languages for the semantic web (ruleml 2016)* (Vol. 9718, pp. 325–340). Springer International Publishing. doi: 10.1007/978-3-319-42019-6_21

Biran, O., & Cotton, C. V. (2017). Explanation and justification in machine learning: A survey. In *Ijcai-17 workshop on explainable ai (xai)* (p. 1-6). IJCAI.

Boran, F. E., Genç, S., Kurt, M., & Akay, D. (2009). A multi-criteria intuitionistic fuzzy group decision making for supplier selection with topsis method. *Expert Systems with applications*, *36*(8), 11363–11368.

Bottani, E., & Rizzi, A. (2008). An adapted multi-criteria approach to suppliers and products selection—an application oriented to lead-time reduction. *International Journal of Production Economics*, *111*(2), 763–781.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). *Classification and regression trees*. Wadsworth.

Burlutskiy, N., Petridis, M., Fish, A., Chernov, A., & Ali, N. (2016). An investigation on online versus batch learning in predicting user behaviour. In *International conference on innovative techniques and applications of artificial intelligence* (pp. 135–149).

Çalık, A. (2021). A novel pythagorean fuzzy ahp and fuzzy topsis methodology for green supplier selection in the industry 4.0 era. *Soft Computing*, *25*(3), 2253–2265.

Cavalcante, I. M., Frazzon, E. M., Forcellini, F. A., & Ivanov, D. (2019). A supervised machine learning approach to data-driven simulation of resilient supplier selection in digital manufacturing. *International Journal of Information Management*, *49*, 86–97.

Celebi, D., & Bayraktar, D. (2008). An integrated neural network and data envelopment analysis for supplier evaluation under incomplete information. *Expert Systems with Applications*, *35*(4), 1698–1710.

Cengiz, A., Aytekin, O., Ozdemir, I., Kusan, H., & Cabuk, A. (2017). A multi-criteria decision model for construction material supplier selection. *Procedia Engineering*, *196*, 294–301.

Chan, F. T., Kumar, N., Tiwari, M. K., Lau, H. C., & Choy, K. (2008). Global supplier selection: a fuzzy-ahp approach. *International Journal of production research*, *46*(14), 3825–3857.

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering*, *40*(1), 16–28.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, *2*(6), 429 - 444.

Chatterjee, P., & Stević, Ž. (2019). A two-phase fuzzy ahp-fuzzy topsis model for supplier evaluation in manufacturing environment. *Operational Research in Engineering Sciences: Theory and Applications*, *2*(1), 72–90.

Chen, C., Liaw, A., & Breiman, L. (2004). Using random forest to learn imbalanced data. *University of California, Berkeley*, *110*(1-12), 24.

Cheng, Y., Peng, J., Gu, X., Zhang, X., Liu, W., Zhou, Z., ... Huang, Z. (2020). An intelligent supplier evaluation model based on data-driven support vector regression in global supply chain. *Computers & Industrial Engineering*, *139*, 105834.

Choi, T. Y., & Hartley, J. L. (1996). An exploration of supplier selection practices across the supply chain. *Journal of operations management*, *14*(4), 333–343.

Cole, R., & Aitken, J. (2019). Selecting suppliers for socially sustainable supply chain management: post-exchange supplier development activities as pre-selection requirements. *Production Planning & Control*, *30*(14), 1184–1202.

Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21–27.

Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning.* John Wiley & Sons.

De Boer, L., Labro, E., & Morlacchi, P. (2001). A review of methods supporting supplier selection. *European journal of purchasing & supply management*, *7*(2), 75–89.

De Boer, L., van der Wegen, L., & Telgen, J. (1998). Outranking methods in support of supplier selection. *European Journal of Purchasing & Supply Management*, *4*(2-3), 109–118.

de Brito, M. M., & Evers, M. (2016). Multi-criteria decision-making for flood risk management: a survey of the current state of the art. *Natural Hazards and Earth System Sciences*, *16*(4), 1019–1033.

Dickson, G. W. (1966). An analysis of vendor selection systems and decisions. *Journal of purchasing*, *2*(1), 5–17.

dos Santos Amorim, J. M. B., Bezerra, S. d. T. M., Silva, M. M., & de Sousa, L. C. O. (2020). Multicriteria decision support for selection of alternatives directed to integrated urban water management. *Water Resources Management*, *34*(13), 4253–4269.

Dulhare, U. N., Ahmad, K., & Ahmad, K. A. B. (2020). *Machine learning and big data: Concepts, algorithms, tools and applications.* John Wiley & Sons.

Durmić, E. (2019). Evaluation of criteria for sustainable supplier selection using fucom method. *Operational Research in Engineering Sciences: Theory and Applications*, *2*(1), 91–107.

Durmić, E., Stević, Ž., Chatterjee, P., Vasiljević, M., & Tomašević, M. (2020). Sustainable supplier selection using combined fucom–rough saw model. *Reports in mechanical engineering*, *1*(1), 34–43.

Ellram, L. M. (1990). The supplier selection decision in strategic partnerships. *Journal of Purchasing and materials Management*, *26*(4), 8–14.

Fallahpour, A., Kazemi, N., Molani, M., Nayyeri, S., & Ehsani, M. (2018). An intelligence-based model for supplier selection integrating data envelopment analysis and support vector machine. *Iranian Journal of Management Studies*, *11*(2), 209–241.

Fallahpour, A., Olugu, E. U., Musa, S. N., Khezrimotlagh, D., & Wong, K. Y. (2016). An integrated model for green supplier selection under fuzzy environment: application of data envelopment analysis and genetic programming approach. *Neural Computing and Applications*, *27*(3), 707–725.

Fernández, A., García, S., Galar, M., Prati, R. C., Krawczyk, B., & Herrera, F. (2018). *Learning from imbalanced data sets* (Vol. 10). Springer.

Forghani, A., Sadjadi, S. J., & Farhang Moghadam, B. (2018). A supplier selection model in pharmaceutical supply chain using pca, z-topsis and milp: A case study. *PloS one*, *13*(8), e0201604.

Ganesh, A. D., & Kalpana, P. (2022). Future of artificial intelligence and its influence on supply chain risk management–a systematic review. *Computers & Industrial Engineering*, 108206.

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining*

(Vol. 72). Springer.

Gegovska, T., Koker, R., & Cakar, T. (2020). Green supplier selection using fuzzy multiple-criteria decision-making methods and artificial neural networks. *Computational Intelligence and Neuroscience*, *2020*.

Géron, A. (2019). *Hands-on machine learning with scikit-learn, keras, and tensorflow: Concepts, tools, and techniques to build intelligent systems.* O'Reilly Media.

Ghamari, R., Mahdavi-Mazdeh, M., & Ghannadpour, S. F. (2021). Resilient and sustainable supplier selection via a new framework: a case study from the steel industry. *Environment, Development and Sustainability*, 1–39.

Golmohammadi, D., Creese, R. C., Valian, H., & Kolassa, J. (2009). Supplier selection based on a neural network model using genetic algorithm. *IEEE Transactions on Neural Networks*, *20*(9), 1504–1519.

Güneri, A. F., Ertay, T., & Yücel, A. (2011). An approach based on anfis input selection and modeling for supplier selection problem. *Expert Systems with Applications*, *38*(12), 14907–14917.

Guo, X., Yuan, Z., & Tian, B. (2009). Supplier selection based on hierarchical potential support vector machine. *Expert Systems with Applications*, *36*(3), 6978–6985.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157–1182.

Ha, S. H., & Krishnan, R. (2008). A hybrid approach to supplier selection for the maintenance of a competitive supply chain. *Expert systems with applications*, *34*(2), 1303–1311.

Han, J., Pei, J., & Tong, H. (2022). *Data mining: concepts and techniques.* Morgan kaufmann.

Harikrishnakumar, R., Dand, A., Nannapaneni, S., & Krishnan, K. (2019). Supervised machine learning approach for effective supplier classification. In *2019 18th ieee international conference on machine learning and applications (icmla)* (pp. 240–245).

Harmelen, F. v., & Teije, A. t. (2019, January). A Boxology of Design Patterns for Hybrid Learning and Reasoning Systems. *Journal of Web Engineering*, *18*(1), 97–124. Retrieved 2022-02-08, from `https://riverpublishers.com/journal_read_html_article.php?j=JWE/18/1/3` (Publisher: River Publishers) doi: 10.13052/jwe1540-9589.18133

Ho, W., Xu, X., & Dey, P. K. (2010). Multi-criteria decision making approaches for supplier evaluation and selection: A literature review. *European Journal of operational research*, *202*(1), 16–24.

Huang, C., & Yoon, K. (1981). *Multiple attribute decision making: methods and applications.* Springer.

Humphreys, P., Huang, G., Cadden, T., & McIvor, R. (2007). Integrating design metrics within the early supplier selection process. *Journal of Purchasing and Supply Management*, *13*(1), 42–52.

Hwang, C.-L., & Yoon, K. (1981). Methods for multiple attribute decision making. In *Multiple attribute decision making* (pp. 58–191). Springer.

Islam, S., Amin, S. H., & Wardley, L. J. (2021). Machine learning and optimization models for supplier selection and order allocation planning. *International Journal of Production Economics*, *242*, 108315.

Junior, F. R. L., Osiro, L., & Carpinetti, L. C. R. (2014). A comparison between fuzzy ahp and fuzzy topsis methods to supplier selection. *Applied Soft Computing*, *21*, 194–209.

Kannan, D., de Sousa Jabbour, A. B. L., & Jabbour, C. J. C. (2014). Selecting green suppliers based on gscm practices: Using fuzzy topsis applied to a brazilian electronics company. *European Journal of operational research*, *233*(2), 432–447.

Kar, A. K. (2015). A hybrid group decision support system for supplier selection using analytic hierarchy process, fuzzy set theory and neural network. *Journal of Computational Science*, *6*, 23–33.

Kıran, M. S., Eşme, E., Torğul, B., & Paksoy, T. (2020). Machine learning in scm. *Logistics 4.0: Digital Transformation of Supply Chain Management*, 103.

Kukreja, H., Bharath, N., Siddesh, C., & Kuldeep, S. (2016). An introduction to artificial neural network. *Int J Adv Res Innov Ideas Educ*, *1*, 27–30.

Kumar, J., & Roy, N. (2010). A hybrid method for vendor selection using neural network. *International Journal of Computer Applications*, *11*(12), 35–40.

Kuo, R., Hong, S., & Huang, Y. (2010). Integration of particle swarm optimization-based fuzzy neural network and artificial neural network for supplier selection. *Applied Mathematical Modelling*, *34*(12), 3976–3990.

Kuo, R. J., Wang, Y. C., & Tien, F. C. (2010). Integration of artificial neural network and mada methods for green supplier selection. *Journal of cleaner production*, *18*(12), 1161–1170.

Lakshmanpriya, C., Sangeetha, N., & Lavanpriya, C. (2013). Vendor selection in manufacturing industry using ahp and ann. *The SIJ Transactions on Industrial, Financial & Business Management*, *1*(1), 29–34.

Lee, H. Y., Kim, S., & Lee, H. Y. (2001). Expansion of active site area and improvement of kinetic reversibility in electrochemical pseudocapacitor electrode. *Electrochemical and solid state letters*, *4*(3), A19.

Li, C., Fun, Y., & Hung, J. (1997). A new measure for supplier performance evaluation. *IIE Transactions*, *29*(9), 753–758.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2017). Feature selection: A data perspective. *ACM Computing Surveys (CSUR)*, *50*(6), 1–45.

Li, J., Cheng, K., Wang, S., Morstatter, F., Trevino, R. P., Tang, J., & Liu, H. (2018). Feature selection: A data perspective. *ACM Comput. Surv.*, *50*(6), 94:1-94:45.

Li, X., Wang, Y., Basu, S., Kumbier, K., & Yu, B. (2019). A debiased mdi feature importance measure for random forests. *Advances in Neural Information Processing Systems*, *32*.

Lin, X., Li, C., Zhang, Y., Su, B., Fan, M., & Wei, H. (2017). Selecting feature subsets based on svm-rfe and the overlapping ratio with applications in bioinformatics. *Molecules*, *23*(1), 52.

Liou, J. J., Chang, M.-H., Lo, H.-W., & Hsu, M.-H. (2021). Application of an mcdm model with data mining techniques for green supplier evaluation and selection. *Applied Soft Computing*, *109*, 107534.

Liou, J. J., Chuang, Y.-C., Zavadskas, E. K., & Tzeng, G.-H. (2019). Data-driven hybrid multiple attribute decision-making model for green supplier evaluation and performance improvement. *Journal of Cleaner Production*, *241*, 118321.

Liu, F.-H. F., & Hai, H. L. (2005). The voting analytic hierarchy process method for selecting supplier. *International journal of production economics*, *97*(3), 308–317.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Maghsoodi, A. I., Kavian, A., Khalilzadeh, M., & Brauers, W. K. (2018). Clus-mcda: A novel framework based on cluster analysis and multiple criteria decision theory in a supplier selection problem. *Computers & Industrial Engineering*, *118*, 409–422.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*, 381–386.

Mahmoudi, A., Deng, X., Javed, S. A., & Yuan, J. (2020). Large-scale multiple criteria decision-making with missing values: project selection through topsis-opa. *Journal of Ambient Intelligence and Humanized Computing*, 1–22.

Martínez-López, F. J., & Casillas, J. (2013). Artificial intelligence-based systems applied in industrial marketing: An historical overview, current and future insights. *Industrial Marketing Management*, *42*(4), 489–495.

Mohamed, A., Ahmad, M. M., Hamouda, A., & Sapuan, S. (2008). Review of the legal aspect in libyan oil industry and the importance of technology transfer. In *Conference on law and technology.*

Molnar, C. (2019). *Interpretable machine learning.* https://christophm.github.io/interpretable-ml-book/.

Molnar, C. (2020). *Interpretable machine learning.* Lulu. com.

Mugurusi, G., & Oluka, P. N. (2021). Towards explainable artificial intelligence (xai) in supply chain management: A typology and research agenda. In *Advances in production management systems. artificial intelligence for sustainable and resilient production systems: Ifip wg 5.7 international conference, apms 2021, nantes, france, september 5–9, 2021, proceedings, part iv* (pp. 32–38).

Mukherjee, K. (2017). *Supplier selection.* Springer.

Ni, D., Xiao, Z., & Lim, M. K. (2020). A systematic review of the research trends of machine learning in supply chain management. *International Journal of Machine Learning and Cybernetics*, *11*(7), 1463–1482.

Nodeh, M. J., Calp, M. H., & Şahin, İ. (2019). A novel hybrid model for vendor selection in a supply chain by using artificial intelligence techniques case study: Petroleum companies. In *The international conference on artificial intelligence and applied mathematics in engineering* (pp. 226–251).

Oguztimur, S. (2011, September). *Why Fuzzy Analytic Hierarchy Process Approach For Transport Problems?* (ERSA conference papers No. ersa11p438). European Regional Science Association. Retrieved from https://ideas.repec.org/p/wiw/wiwrsa/ersa11p438.html

Ortiz-Barrios, M., Cabarcas-Reyes, J., Ishizaka, A., Barbati, M., Jaramillo-Rueda, N., & de Jesús Carrascal-Zambrano, G. (2020). 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*, 1–39.

Pamucar, D., Badi, I., Sanja, K., & Obradovic, R. (2018). *A novel approach for the selection of power-generation technology using a linguistic neutrosophic codas method: A case study in libya.* Infinite Study.

Pamučar, D., Stević, Ž., & Sremac, S. (2018). A new model for determining weight coefficients of criteria in mcdm models: Full consistency method (fucom). *Symmetry*, *10*(9), 393.

Papadakis, E., Adams, B., Gao, S., Martins, B., Baryannis, G., & Ristea, A. (2022). Explainable artificial intelligence in the spatial domain (x-geoai). *Transactions in GIS*, *26*(6), 2413–2414.

Parkan, C., & Wu, M. (1997). On the equivalence of operational performance measurement and multiple attribute decision making. *International Journal of Production Research*, *35*(11), 2963–2988.

Partovi, F. Y., Burton, J., & Banerjee, A. (1990). Application of analytical hierarchy process in operations management. *International Journal of Operations & Production Management*.

Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. *Drug Discovery Today*, *26*(1), 80.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Peterson, L. E. (2009). K-nearest neighbor. *Scholarpedia*, *4*(2), 1883.

Pouraghabagher, R., & Sarfaraz, O. (2018). Supplier selection using data analytics: a collaborative filtering-based approach. In *Symposium on logistics (isl 2018) big data enabled supply chain innovations.*

Prentkovskis, O., Erceg, Ž., Stević, Ž., Tanackov, I., Vasiljević, M., & Gavranović, M. (2018). A new methodology for improving service quality measurement: Delphi-fucom-servqual model. *Symmetry*, *10*(12), 757.

Ray, S. (2019). A quick review of machine learning algorithms. In *2019 international conference on machine learning, big data, cloud and parallel computing (comitcon)* (pp. 35–39).

Rezaei, J., Nispeling, T., Sarkis, J., & Tavasszy, L. (2016). A supplier selection life cycle approach integrating traditional and environmental criteria using the best worst method. *Journal of Cleaner Production*, *135*, 577–588.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should i trust you? Explaining the predictions of any classifier. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 1135–1144).

Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion*, *27*, 111–125.

Ross, D. F., Weston, F. S., & Stephen, W. (2010). *Introduction to supply chain management technologies.* Crc Press.

Saar-Tsechansky, M., & Provost, F. (2007). Handling missing values when applying classification models. *Journal of Machine Learning Research.*

Saaty, T. (1980). *The analytic hierarchy process.* McGraw-Hill.

Saaty, T. (1996). *Decision making with dependence and feedback: The analytic network process.* RWS Publications.

Saghafian, S., & Hejazi, S. R. (2005). Multi-criteria group decision making using a modified fuzzy topsis procedure. In *International conference* *on computational intelligence for modelling, control and automation and international conference on intelligent agents, web technologies and internet commerce (cimca-iawtic'06)* (Vol. 2, pp. 215–221).

Sakhardande, M. J., & Gaonkar, R. S. P. (2022). On solving large data matrix problems in fuzzy ahp. *Expert Systems with Applications*, *194*, 116488.

Salo, F., Nassif, A. B., & Essex, A. (2019). Dimensionality reduction with ig-pca and ensemble classifier for network intrusion detection. *Computer Networks*, *148*, 164–175.

Sarkar, D., Bali, R., & Sharma, T. (2018). Practical machine learning with python. *A Problem-Solvers Guide To Building Real-World Intelligent Systems. Berkely: Apress*.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, *2*(3), 1–21.

Sarkis, J., & Talluri, S. (2002). A model for strategic supplier selection. *Journal of supply chain management*, *38*(4), 18–28.

Sasaki, H., & Sakata, I. (2021). Business partner selection considering supply-chain centralities and causalities. In *Supply chain forum: An international journal* (Vol. 22, pp. 74–85).

Schramm, V. B., Cabral, L. P. B., & Schramm, F. (2020). Approaches for supporting sustainable supplier selection-a literature review. *Journal of cleaner production*, 123089.

Shang, Z., Yang, X., Barnes, D., & Wu, C. (2022). Supplier selection in sustainable supply chains: Using the integrated bwm, fuzzy shannon entropy, and fuzzy multimoora methods. *Expert Systems with Applications*, *195*, 116567.

Sharma, H., & Kumar, S. (2016). A survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research (IJSR)*, *5*(4), 2094–2097.

Shih, H.-S., Shyur, H.-J., & Lee, E. S. (2007). An extension of topsis for group decision making. *Mathematical and computer modelling*, *45*(7-8), 801–813.

Shinde, P. P., & Shah, S. (2018). A review of machine learning and deep learning applications. In *2018 fourth international conference on computing communication control and automation (iccubea)* (pp. 1–6).

Shukla, R. K., Garg, D., & Agarwal, A. (2014). An integrated approach of fuzzy ahp and fuzzy topsis in modeling supply chain coordination. *Production & Manufacturing Research*, *2*(1), 415–437.

Sonmez, M. (2006). *Review and critique of supplier selection process and practices*. Citeseer.

Stević, Ž., Pamučar, D., Puška, A., & Chatterjee, P. (2020). Sustainable supplier selection in healthcare industries using a new mcdm method: Measurement of alternatives and ranking according to compromise solution (marcos). *Computers & Industrial Engineering*, *140*, 106231.

Subasi, A. (2020). *Practical machine learning for data analysis using python*. Academic Press.

Sun, B., Du, J., & Gao, T. (2009). Study on the improvement of k-nearest-neighbor algorithm. In *2009 international conference on artificial intelligence and computational intelligence* (Vol. 4, pp. 390–393).

Swamynathan, M. (2019). *Mastering machine learning with python in six steps: A practical implementation guide to predictive data analytics using* *python*. Apress.

Tang, M., & Liao, H. (2021). From conventional group decision making to large-scale group decision making: What are the challenges and how to meet them in big data era? a state-of-the-art survey. *Omega*, *100*, 102141.

Tang, S., Hakim, N., Khaksar, W., Ariffin, M., Sulaiman, S., & Pah, P. (2013). A hybrid method using analytic hierarchical process and artificial neural network for supplier selection. *International Journal of Innovation, Management and Technology*, *4*(1), 109–111.

Taunk, K., De, S., Verma, S., & Swetapadma, A. (2019). A brief review of nearest neighbor algorithm for learning and classification. In *2019 international conference on intelligent computing and control systems (iccs)* (pp. 1255–1260).

Tharwat, A. (2016). Principal component analysis-a tutorial. *International Journal of Applied Pattern Recognition*, *3*(3), 197–240.

Tirkolaee, E. B., Sadeghi, S., Mooseloo, F. M., Vandchali, H. R., & Aeini, S. (2021). Application of machine learning in supply chain management: a comprehensive overview of the main areas. *Mathematical problems in engineering*, *2021*.

Tzeng, G.-H., & Huang, J.-J. (2011). *Multiple attribute decision making: methods and applications*. CRC press.

van Buuren, S., & Oudshoorn, K. (2015). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, *45*, 1–67. doi: 10.18637/jss.v045.i03

Vaske, J. J., Beaman, J., & Sponarski, C. C. (2017). Rethinking internal consistency in cronbach's alpha. *Leisure Sciences*, *39*(2), 163–173.

Velasquez, M., & Hester, P. (2013). An analysis of multi-criteria decision making methods. *International Journal of Operations Research*, *10*(2), 56-66.

Wang Chen, H. M., Chou, S.-Y., Luu, Q. D., & Yu, T. H.-K. (2016). A fuzzy mcdm approach for green supplier selection from the economic and environmental aspects. *Mathematical Problems in Engineering*, *2016*.

Weber, C. A., & Current, J. R. (1993). A multiobjective approach to vendor selection. *European journal of operational research*, *68*(2), 173–184.

Weber, C. A., Current, J. R., & Benton, W. (1991). Vendor selection criteria and methods. *European journal of operational research*, *50*(1), 2–18.

Wey, W.-M., & Wu, K.-Y. (2007). Using anp priorities with goal programming in resource allocation in transportation. *Mathematical and computer modelling*, *46*(7-8), 985–1000.

Witten, I. H., & Frank, E. (2002). Data mining: practical machine learning tools and techniques with java implementations. *Acm Sigmod Record*, *31*(1), 76–77.

Wu, C., & Barnes, D. (2011). A literature review of decision-making models and approaches for partner selection in agile supply chains. *Journal of Purchasing and Supply Management*, *17*(4), 256–274.

Wu, D. (2009). Supplier selection: A hybrid model using dea, decision tree and neural network. *Expert systems with Applications*, *36*(5), 9105–9112.

Yazdani, M., Torkayesh, A. E., & Chatterjee, P. (2020). An integrated decision-making model for supplier evaluation in public healthcare system: the case study of a spanish hospital. *Journal of Enterprise Information Man-*

agement.

Zhang, R., Li, J., Wu, S., & Meng, D. (2016). Learning to select supplier portfolios for service supply chain. *PLOS one*, *11*(5), e0155672.

Zhang, Z., Lei, J., Cao, N., To, K., & Ng, K. (2003). Evolution of supplier selection criteria and methods. *European Journal of Operational Research*, *4*(1), 335–342.

Zhu, X., Li, J., Wu, D., Wang, H., & Liang, C. (2013). Balancing accuracy, complexity and interpretability in consumer credit decision making: A c-topsis classification approach. *Knowledge-Based Systems*, *52*, 258–267.

Zouggari, A., & Benyoucef, L. (2012). Simulation based fuzzy topsis approach for group multi-criteria supplier selection problem. *Engineering Applications of Artificial Intelligence*, *25*(3), 507–519.