

A Decision-Making System for Medical Transportation Mode Using Machine Learning Methods [†]

Sahar Khodabakhshi, Md Ali Ahsan Habib and Wei Peng ^{*}

Department of Industrial Systems Engineering, University of Regina, Regina, SK S4S 0A2, Canada; sko783@uregina.ca (S.K.); mhs647@uregina.ca (M.A.A.H.)

^{*} Correspondence: wei.peng@uregina.ca

[†] Presented at the 1st International Conference on Industrial, Manufacturing, and Process Engineering (ICIMP-2024), Regina, Canada, 27–29 June 2024.

Abstract: The freight transportation system faces complex operations globally to meet customer demands. Intense competition prompts companies to enhance performance. Transportation modes (road, sea, air) impact service levels, each with distinct features, benefits, costs, environmental effects, and societal risks. Shippers confront challenges in mode selection due to numerous factors, compounded by an increase in low-volume, high-frequency shipments. Rising logistics costs for a few products of exporters affect the socio-economic situation of a country. This research introduces a hybrid approach for a shipment selection model, focusing on pharmaceutical drugs. Utilizing machine learning algorithms (decision tree, Random Forest, logistic regression, XGboost, SVM) and multi-criteria decision-making methods (SAW, MARCOS, TOPSIS, MULTIMOORA, VIKOR), this study predicts the optimal shipping method (land, air, sea) based on dataset features (shipping cost, origin-destination, cargo weight, dimensions). Evaluation metrics include F1 score, Recall, Precision, and Accuracy score. XGboost stands out as the optimal algorithm, demonstrating an accuracy of eighty-four percent, with random forest, decision tree, SVM, and logistic regression following in descending order. This comprehensive approach addresses the complexities of pharmaceutical shipment selection, considering various influential factors.

Keywords: transportation mode; MCDM; machine learning; data analysis



Citation: Khodabakhshi, S.; Habib, M.A.A.; Peng, W. A Decision-Making System for Medical Transportation Mode Using Machine Learning Methods. *Eng. Proc.* **2024**, *76*, 74. <https://doi.org/10.3390/engproc2024076074>

Academic Editors: Golam Kabir, Sharfuddin Khan, Mohammad Khondoker and Hussameldin Ibrahim

Published: 8 November 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The escalating volume of global trade, fueled by economic integration and globalization, has presented numerous market expansion opportunities for companies. However, this surge in trade has concurrently intensified competition, prompting companies to adopt strategic measures such as alliances, mergers, acquisitions, and internal cost reduction to survive. Despite technological advancements reducing production costs, logistics costs, particularly in transportation, remain a significant concern. The proliferation of low-volume, high-frequency shipments due to diverse demand, shorter product life cycles, and agile customer responses has further complicated logistics, cost management, especially for companies exporting small quantities of products abroad [1].

Transportation costs constitute a substantial portion of overall logistics costs, making the selection of an appropriate means of exporting products crucial for cost reduction. Strategic decision making, aimed at minimizing logistic costs and aligning with customer and market needs, revolves around the choice of transportation means, with operational decisions extending to cargo transportation routes [2].

While sea transportation is favored for its cost efficiency in international freight services, shippers also utilize car ferry services for general cargo to close overseas destinations. Shippers weigh factors such as freight rates, transportation speed, and customs clearance quality when choosing sea transportation methods. Similarly, air transportation users

consider shipment costs and service quality, with low-cost carriers gaining popularity due to cost advantages and inherent service benefits over traditional freight carriers.

Logistic management's key decisions, including transportation method and carrier selection, involve considerations of land, sea, and air transportation. Contemporary processes involve the simultaneous selection of transportation type and specific carriers, often facilitated by professional forwarders. Quality of transportation services, highlighted by factors like accurate product transfer billing and high-quality customer service, often takes precedence over transportation cost considerations in land transportation [3,4].

Sea freight carrier selection emphasizes the importance of transportation service quality over reliability, capacity, speed, and flexibility. Factors such as damaged cargo proportion, multimodal transportation capability, on-time delivery reliability, lower price, credit facilities, and environmentally friendly packaging contribute significantly to sea freight carrier choice [5].

Air cargo transportation factors, predominantly focusing on freight-related aspects, consider incentives for transportation cost and product loading quantities. The importance of proper shipping policies and additional factors such as carrier reliability and transportation speed underscores the significance of transportation service quality in air transportation [6,7].

While existing studies have explored land, sea, and air modes separately, a research gap exists in addressing the dynamic transportation approaches faced by companies exporting small quantities of goods. This gap is particularly evident in the context of medical resource transportation, as timely delivery of treatment to patients is critical.

This paper proposes a hybrid approach using machine learning algorithms and multi-criteria decision-making methods to develop a shipment selection model, focusing on pharmaceutical drugs. This research aims to enhance prediction accuracy through decision tree, Random Forest, logistic regression, XGboost, and SVM algorithms, validated by multi-criteria decision-making methods such as SAW, MARCOS, TOPSIS, MULTIMOORA, and VIKOR.

The significance of this research lies in revolutionizing drug distribution networks, optimizing shipping services for companies, and controlling medical center inventories. This model's results are expected to reduce arrival times and transportation costs for medicinal items, fostering competitiveness, savings, and environmentally conscious shipping practices. This research contributes by integrating machine learning and multi-criteria decision making in shipment selection, offering various validation methods and achieving up to eighty-seven percent accuracy in pharmaceutical drug shipment selection. The proposed methodology addresses the pressing challenges in pharmaceutical shipment, providing a systematic and innovative solution for risk mitigation, cost reduction, and timely deliveries in medical supply chains.

2. Methodology

2.1. Data Collection and Analysis Method

In addressing contemporary challenges in global goods transportation necessitated by technological advancements, this section delves into the pivotal role of data science. Defined by its multidisciplinary nature, incorporating statistics, machine learning, and human-computer interaction, data science automates the extraction of valuable insights from extensive data repositories. This process empowers organizations to uncover latent patterns, fostering excellence in operations. The ensuing discussion explores diverse academic fields, encompassing statistics, computer science, mathematics, and domain knowledge, contributing to informed decision making through statistical analysis, machine learning, data mining, and visualization. Data scientists play a pivotal role in tasks such as data cleansing, data modeling, and predictive modeling, enabling the discovery of trends and patterns in complex datasets. The subsequent elucidation on the structured process, as delineated by Han et al. [8], encapsulates sequential steps from data collection and integration to knowledge extraction and illustration. This methodological approach

provides a comprehensive framework for deriving meaningful insights in the context of goods transportation.

2.2. Crisp Method

While the method presented is a step towards research goals, it proves insufficient. Consequently, this study employs CRISP-DM, a comprehensive process developed in 1996 under the ESPRIT initiative by the European Union. Spearheaded by a consortium including SPSS, Teradata, Daimler AG, NCR Corporation, and OHRA, CRISP-DM amalgamates diverse experiences integrated into SPSS. NCR Corporation contributes its data warehouse and data mining software. Originating in 1999, CRISP-DM's six-step approach was unveiled at the 4th CRISP-DM SIG meeting in Brussels in Figure 1.

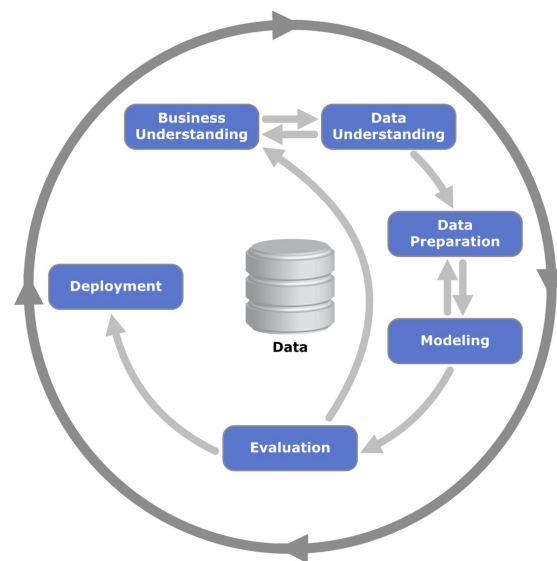


Figure 1. Crisp-DM Process [9].

2.3. Delivery Services in the Supply Chain Management

Rising consumer expectations necessitate supply chains to prioritize efficiency, particularly in delivery services, prompting a demand for agile courier companies. Advanced technologies, notably machine learning-based smart transportation management systems, enhance productivity, optimizing the logistic infrastructure crucial for safe medical specimen transportation. Challenges include uncoordinated processes and technology mismatches, addressed through evaluation and automation in the supply chain. Employing technology ensures smooth supply chain management. On-demand and scheduled on-demand delivery models, each with advantages and drawbacks, handle unforeseen situations. This research focuses on a scheduled on-demand approach for medicine delivery, aiming to identify the most effective, cost-efficient method while ensuring safety and customer satisfaction.

2.4. Proposed Research Method

Data pre-processing and preparation: Data analysis and preparation are pivotal stages in dataset functionality, consuming sixty percent to eighty percent of project time. Steps encompass data integration, combination, transformation, cleaning, and dimensionality reduction, tailored to dataset type. Integrating data from diverse sources ensures comprehensiveness and accuracy. Data combination multiplies separate features, creating new, problem-relevant ones. Data transformation, involving non-numeric control and scaling, addresses categorical values and enhances algorithm reliability. Cleaning manages outliers and missing data, with strategies like deletion or suitable value replacement based on the dataset nature. Following these steps ensures effective dataset preparation for accurate, reliable results in data-driven projects.

Cleaning the data and bringing it to the proper form: In this step, inconsistent, noisy, repetitive, unrecorded, and lost data are identified and addressed. Missing data, not recorded by the logger, can impact results' accuracy, requiring investigation via statistical, descriptive, or expert opinion methods. Methods for inconsistent and outlier data involve establishing relationships to identify incorrect and main outlier data accurately.

Data Transformation and Scaling: In this section, the data are converted into a suitable form for analysis and review. There are different ways to scale data. In this research, the Min-Max scaling method is used. This method uses the lowest and highest values in each feature and places them in the range of zero and one. Finally, all the calculated features will be in the range of zero to one.

Feature Selection: Feature selection is vital in data preparation, employing filter, wrapper, and embedded methods. Filter methods independently rank features based on statistical tests, reducing overfitting but potentially compromising predictive performance. Wrapper methods optimize features based on classifier performance, considering dependencies and yielding higher predictive accuracy but with computational intensity. Embedded methods integrate feature selection into classifier algorithms, adjusting internal parameters for optimal accuracy in a single step. Methods vary in terms of independence, evaluation metrics, computational complexities, and potential for detecting redundancies and interactions between features.

Handle Imbalanced Data: Following target feature analysis, an imbalanced data issue was identified, impacting machine learning algorithm performance. Addressing this issue, Synthetic Minority Over-sampling Technique (SMOTE) was chosen to rectify imbalances by creating synthetic values.

2.5. Proposed Machine Learning Algorithms

The CRISP method guides data science projects, with the fourth step involving modeling to address challenges. Machine learning, categorized as supervised, unsupervised, and reinforcement learning, plays a pivotal role. Supervised learning uses labeled datasets for classification, estimation, and prediction, employing methods like decision tree and support vector machine. Unsupervised learning identifies patterns in unlabeled data through clustering and associative rules. Reinforcement learning focuses on an agent making decisions in dynamic environments, guided by interactions, rewards, and penalties within a Markov decision process framework.

Decision Tree: Decision trees are crucial in data mining, offering efficient classification for vast information. Widely used in machine learning, image processing, and pattern identification, decision trees construct cohesive models with simple rules. Comparing numeric features to thresholds, decision trees excel in categorical classification. Noteworthy advantages include readability, ease of preparation, and minimal data cleaning. However, their instability and limitations in predicting continuous variables are acknowledged disadvantages [10].

Random Forest: Random forests, an ensemble learning method, construct multiple decision trees during training for classification and regression tasks. For classification, the majority-selected class is the output, while regression tasks use the average prediction. The training algorithm employs boot-strap aggregating, reducing model variance without increasing bias. The de-correlating effect of bootstrap sampling ensures improved model performance by averting sensitivity to noise in individual trees' training sets [11].

XGBoost: XGBoost, a recent addition to machine learning, is a highly efficient decision tree gradient boosting algorithm. It supports three main models: Gradient Boosting, Random Gradient Boosting with sub-sampling, and Regularized Gradient Boosting with L1 and L2 regularization. With features like parallelized tree construction, distributed computing, out-of-core computations, and optimization of data structure cache, XGBoost is designed for speed and optimal resource utilization. It excels in handling structured datasets for classification and regression problems, providing fast computation, and supporting user-defined objective functions and evaluation criteria.

Logistic regression: In statistics, the logistic model is used to model the probability of a particular class or event, such as success/failure, victory/failure, alive/dead, or healthy/sick. This approach can be extended to model several classes of events. Logistic regression is a statistical model that, in its original form, uses a logistic function to model a binary dependent variable, although there are many more complex extensions as well. In regression analysis, logistic regression is the estimation of the parameters of a logistic model (a type of binary regression).

Support Vector Classifier: The Support Vector Machine (SVM) is a versatile algorithm for linear and non-linear data classification. In SVM, the algorithm seeks the optimal hyperplane, using a non-linear mapping for higher-dimensional transformation. The hyperplane maximizes the margin between classes. In a linear SVM, the goal is to find the maximum-margin hyperplane that best separates two classes. The SVM's advantages include easy training and good performance in high-dimensional data, but its limitations include reduced efficiency for large datasets, requiring significant training time.

2.6. Proposed Multi-Criteria Decision-Making Method

The multi-criteria decision-making (MCDM) process involves various stages, each crucial for achieving reliable results. Researchers emphasize early stages, including identifying options, determining characteristics, and assigning weights, as vital. Choosing the decision-making method is a pivotal step impacting final outcomes. The eight-step MCDM process includes forming the decision matrix, normalization, and determining criteria weights. Numerous methods exist, such as SAW, TOPSIS, MULTIMOORA, VIKOR, and MARCOS. The Marcos method, a novel approach, innovatively addresses normalization. Each method offers unique advantages and considerations, emphasizing the importance of comprehensive evaluation in decision-making problems. Figure 2 illustrates the entire decision-making process, highlighting its intricate nature and interconnected stages.

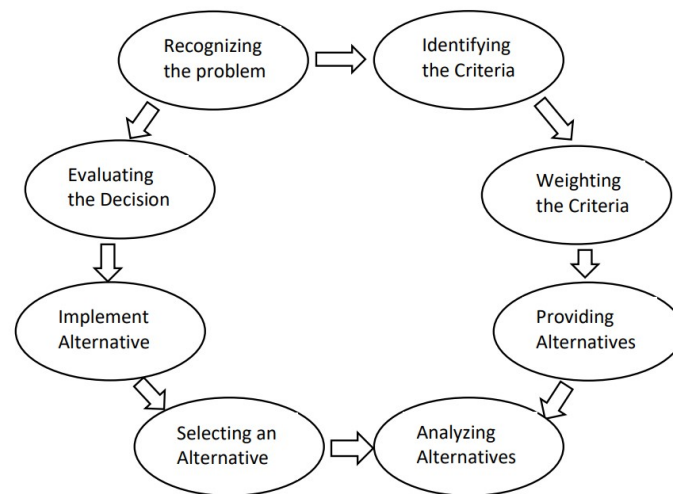


Figure 2. Decision-making process [12].

3. Results

Global advancements in technology have intensified the focus on optimizing shipment methods for raw material producers and distributors. Efficient product movement is crucial for a nation's socioeconomic standing. Active research addresses challenges like rising order volumes, sensitive items, and environmental concerns. Emphasizing intelligence in the shipping business is imperative to address these issues. This section delves into detailed findings, particularly focusing on Exploratory Data Analysis.

The Rank Correlation Coefficient is employed to measure the relationship between variables, ensuring independence and minimizing influences on results in Figure 3. This chart is also beneficial for finding imbalanced data. Feature selection is a fundamental aspect of data preparation, involving various methods like filter, wrapper, and embedded,

depending on the chosen algorithm. Filter methods use independent feature ranking, reducing overfitting. In contrast, wrapper methods enhance performance by considering the classifier algorithm. Embedded methods integrate feature selection into the classifier algorithm, adjusting internal parameters during training.

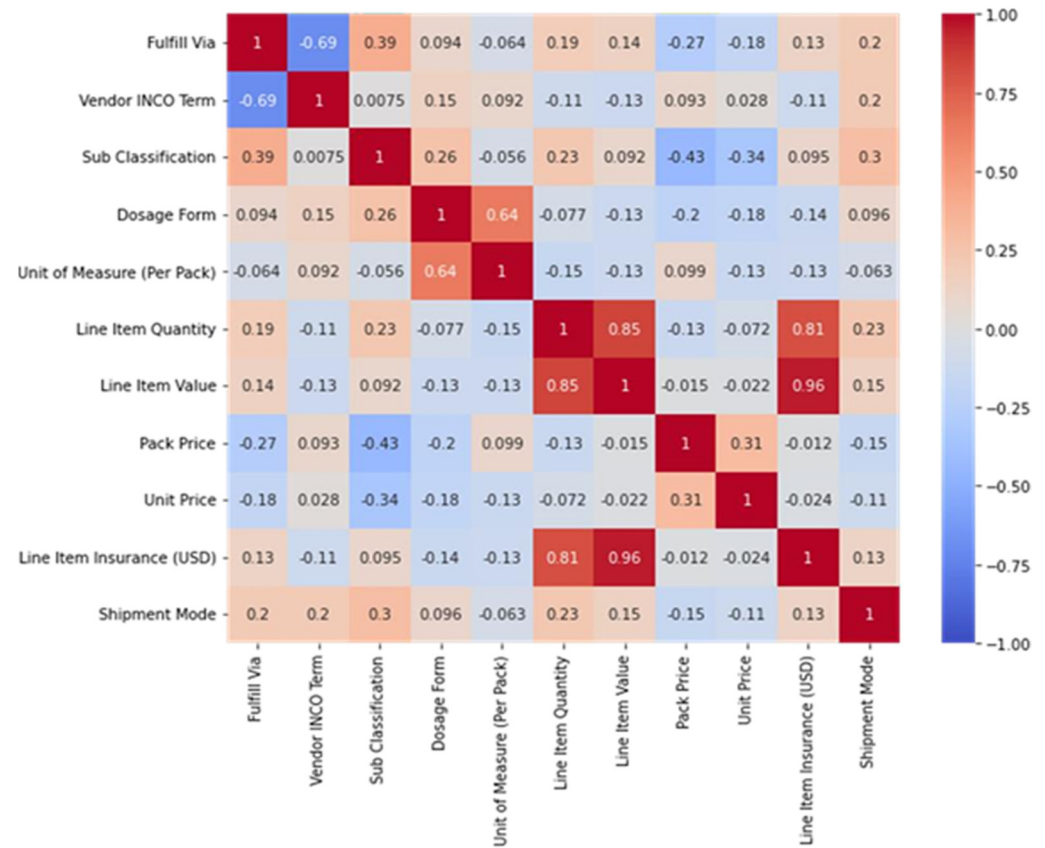


Figure 3. Correlation between fields of dataset.

Hyperparameter tuning was employed in our research. Addressing imbalanced data, we selected Synthetic Minority Over-sampling Technique (SMOTE) to generate synthetic samples, ensuring better machine learning algorithm performance.

In data science, result evaluation ensures the reliability of algorithms. Common techniques include: confusion matrix, which is used to measure the performance of classification models; accuracy, which is used to measure the percentage of correct predictions; precision and recall, which are used to evaluate binary classification models; F1 score, which is the harmonic mean of precision and recall, which is used to balance the two metrics. These methods help identify areas for improvement, ensuring model effectiveness and accuracy in processing datasets for a given problem, ultimately enhancing the decision-making process. Based on reliable results, the performance evaluation of machine learning algorithms is shown in Table 1 as follows:

Table 1. Performance evaluation of machine learning algorithms.

Model	Accuracy	Precision	Recall	F1
DT	0.798	0.8	0.714	0.75
RF	0.854	0.848	0.783	0.812
SVC	0.572	0.41	0.403	0.302
LR	0.49	0.4	0.345	0.3
XGB	0.863	0.894	0.802	0.824

As shown in Table 2, on average, we obtained reasonable results. For choosing the best algorithm, we should use multi-criteria decision making.

Table 2. Results of the MCDM methods.

Index	Marcos	Topsis	SAW	Vikor	Average
1	0.8686	1	0.512	1	0.846
2	0.7991	0.8585	0.477	0.9768	0.779
3	0.4447	0.1539	0.92	0.8367	0.589
4	0.4143	0.0237	1	0.1605	0.399
5	0.4097	0	0.472	0	0.218

4. Conclusions

In the shipping industry, competition is not limited to price, but also includes service quality and the diversity of transportation modes. For drugs that are essential in healthcare, resource constraints require efficient transportation of drugs and equipment. Artificial intelligence, particularly machine learning, proves effective in optimizing supply chains. Machine learning tasks, including supervised, unsupervised, and reinforcement learning, require diverse algorithms for accurate results. Data science, a multidisciplinary field, leverages statistical analysis and machine learning to extract valuable insights. Our research, based on Kaggle data, focuses on drug supply chain transactions, employing CRISP-DM methodology. Addressing imbalanced data, we utilized SMOTE for synthetic sample generation. Evaluation metrics, such as Confusion Matrix, Accuracy, Precision, Recall, and F1 score, guided algorithm selection. Applying multi-criteria decision making (MCDM), XGBoost emerged as the most effective tool. Our approach enhances drug distribution, optimizing services for shipping companies and controlling medical center inventories. This model reduces arrival time and transportation costs, showcasing the impact of combining machine learning and MCDM in medical transportation decision making.

Author Contributions: Conceptualization, S.K. and W.P.; methodology, S.K.; software, S.K.; validation, S.K., M.A.A.H. and W.P.; formal analysis, S.K.; resources, S.K.; writing—original draft preparation, data curation; writing—review and editing, S.K., M.A.A.H. and W.P.; supervision, W.P.; project administration, W.P.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bang, K.T.; Jang, H.H. An empirical study on the transport mode selection factors of Korean exporters -based on the control factor of manufacturing and logistics industry. *J. Shipp. Logist.* **2011**, *69*, 245–263. [\[CrossRef\]](#)
2. Jung, H.; Kim, J.; Shin, K. Importance analysis of decision-making factors for selecting international freight transportation mode. *Asian J. Shipp. Logist.* **2019**, *35*, 55–62. [\[CrossRef\]](#)
3. Lambert, D.M.; Lewis, M.C.; Stock, J.R. How shippers select and evaluate general commodities LTL motor carriers. *J. Bus. Logist.* **1993**, *14*, 131.
4. Danielis, R.; Marcucci, E.; Rotaris, L. Logistics managers' stated preferences for freight service attributes. *Transp. Res. Part E Logist. Transp. Rev.* **2005**, *41*, 201–215. [\[CrossRef\]](#)
5. Shen, L.; Mathiyazhagan, K.; Kannan, D.; Ying, W. Study on analyzing the criteria for selection of shipping carriers in Chinese shipping market using analytical hierarchy process. *Int. J. Shipp. Transp. Logist.* **2015**, *7*, 742–757. [\[CrossRef\]](#)
6. Yoon, S.; Park, J. A study of the competitiveness of airline cargo services departing from Korea: Focusing on the main export routes. *J. Air Transp. Manag.* **2015**, *42*, 232–238. [\[CrossRef\]](#)

7. Park, Y.; Choi, J.K.; Zhang, A. Evaluating competitiveness of air cargo express services. *Transp. Res. Part E Logist. Transp. Rev.* **2009**, *45*, 321–334. [[CrossRef](#)]
8. Han, J.W.; Kamber, M.; Pei, J. *Data Mining Concepts and Techniques*, 3rd ed.; Morgan Kaufmann Publishers: Waltham, MA, USA, 2012.
9. Wikipedia.org. Available online: https://commons.wikimedia.org/wiki/File:CRISP-DM_Process_Diagram.png (accessed on 31 October 2024).
10. Jijo, B.T.; Abdulazeez, A.M. Classification based on decision tree algorithm for machine learning. *Evaluation* **2021**, *6*, 7.
11. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning: With Applications in R*; Springer: Berlin/Heidelberg, Germany, 2013; Available online: www.springer.com/us/book/9781461471370 (accessed on 31 October 2024).
12. Taherdoost, H.; Madanchian, M. Multi-Criteria Decision Making (MCDM) Methods and Concepts. *Encyclopedia* **2023**, *3*, 77–87. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.