CLUSTER-BASED STRATEGIES FOR THE SELECTION OF LOGISTICS SERVICE PROVIDERS: THEORETICAL FOUNDATIONS AND PRACTICAL APPLICATIONS

Gábor NAGY[1], Szabolcs SZENTESI[2]
[1] University of Miskolc, Institute of Logistics, Miskolc-Egyetemváros, 3515, Hungary
E-mail address: gabor.nagy4@uni-miskolc.hu
[2] University of Miskolc, Institute of Logistics, Miskolc-Egyetemváros, 3515, Hungary
E-mail address: szabolcs.szentesi@uni-miskolc.hu

AJME 2025, 23 (2); https://doi.org/10.5281/zenodo.15862721

ABSTRACT*: Selecting logistics service providers is a critical factor in enhancing the efficiency of supply chains, particularly with the rapid growth of e-commerce. Clustering methods, such as k-means clustering, provide an effective approach for grouping providers based on their characteristics, enabling companies to make faster and more informed decisions. This study explores the theoretical foundations of clustering methods and their practical application in the selection of logistics service providers. By conducting a systematic literature review, we analyzed the use of clustering in logistics, with a specific emphasis on the selection and evaluation of service providers.*

KEYWORDS*: logistics service providers, clustering, k-means clustering, provider selection*

1 INTRODUCTION

The efficiency of supply chains plays a pivotal role in shaping companies' competitiveness, particularly with the rapid growth of e-commerce, where fast, reliable, and cost-effective delivery has become a critical factor (Christopher, 2000; Mentzer, Stank, & Esper, 2008). Selecting logistics service providers is especially challenging due to the multitude of factors involved, including shipping costs, delivery times, geographic coverage, and service reliability (Aguezzoul, 2014; Sarkis & Dhavale, 2015). Traditional selection methods often lack the flexibility and analytical depth needed to address complex datasets, highlighting the necessity for more advanced approaches (Mikhailov, 2003; Opricovic & Tzeng, 2004).

Clustering methods, particularly k-means clustering, are unsupervised machine learning techniques designed to identify natural groupings within data (Kaufman & Rousseeuw, 2005; Liu, Ke, Wei, & Hua, 2013). These methods enable the segmentation of providers based on their characteristics, aiding decision-makers in selecting the most suitable partners. This study explores the application of clustering methods for selecting logistics service providers and reviews the relevant literature, with a particular focus on the logistical applications of clustering algorithms and their influence on supply chain efficiency (Wang & Lee, 2009; Zhang, Deng, Chan, Adamatzky, & Mahadevan, 2016).

2 LITERATURE REVIEW

Effective selection of logistics service providers is crucial for enhancing corporate competitiveness in supply chain management (Mentzer, Stank, & Esper, 2008). Comprehensive analysis of providers' performance and characteristics enables companies to optimize shipping costs, improve service reliability, and ensure timely delivery (Aguezzoul, 2014; Sarkis & Dhavale, 2015). In recent years, clustering methods, such as k-means and density-based clustering, have gained significant attention for their ability to group providers naturally based on shared characteristics (Kaufman & Rousseeuw, 2005; Liu, Ke, Wei, & Hua, 2013). This literature review explores the application of clustering techniques in selecting logistics service providers, focusing on the theoretical foundations and practical logistics applications of clustering methods (Zhang, Deng, Chan, Adamatzky, & Mahadevan, 2016).

This analysis highlights the practical advantages of different clustering methods, including grouping providers by shipping costs, delivery time, geographic coverage, and reliability (Kannan, Khodaverdi, Olfat, Jafarian, & Diabat, 2013). Drawing on prior research, case studies, and practical examples, the review demonstrates how clustering methods contribute to the effective evaluation and selection of logistics service providers (Awasthi, Chauhan, & Goyal, 2011; Zhang et al., 2016).

Key questions addressed in this review include: What are the theoretical foundations of clustering techniques? How can these methods be applied to segment logistics service providers? What practical benefits do clustering methods offer in the selection process? The objective is to illustrate how clustering supports decision-making by optimizing shipping processes and enabling precise identification of differences among providers (Kaufman & Rousseeuw, 2005; Mikhailov, 2003).

Clustering methods have broad applications across scientific disciplines, including logistics. Jain, Murty, and Flynn (1999) provide a comprehensive overview of major clustering algorithms, such as k-means, hierarchical, and density-based clustering. K-means remains one of the most widely used algorithms for data grouping, while hierarchical clustering enables the creation of hierarchical structures among clusters. Density-based clustering, such as DBSCAN, identifies clusters based on density distribution, making it particularly effective for managing noisy data (Hosseini & Barker, 2016). Although Xu and Wunsch (2005) primarily examine biomedical applications, their findings also offer valuable insights for logistics, as similar processes of pattern identification are beneficial for clustering in this field.

The practical application of clustering methods in selecting logistics service providers offers substantial advantages. Shen, Olfat, Govindan, Khodaverdi, and Diabat (2013) demonstrate how clustering effectively groups providers by cost, delivery time, and reliability. This approach facilitates a more comprehensive evaluation of provider performance and helps identify top-performing providers. Mazzarol and Soutar's (2002) application of the "push-pull" theory underscores the importance of clustering in provider selection, highlighting its role in aligning demand, supply, and customer needs to find optimal solutions.

Modern clustering techniques, such as the DBSCAN algorithm, hold considerable potential for analyzing logistics service providers. Ester, Kriegel, Sander, and Xu (1996) introduced DBSCAN as a method well-suited for managing noisy data, which frequently occurs in logistics. Unlike k-means, DBSCAN does not require predefined cluster numbers, making it advantageous in dynamic logistics environments. Lloyd's (1982) foundational work on k-means clustering further advances the development of efficient clustering processes, directly enhancing logistics operations.

Applying clustering methods to performance evaluation enables a comprehensive analysis and more effective selection of logistics service providers. Sarkis and Dhavale (2015), in their sustainability-focused study, grouped providers based on multiple performance indicators—such as cost, delivery speed, reliability, and coverage—to identify the best-performing providers that meet corporate service-level expectations. Clustering also facilitates deeper analysis of each provider's strengths and weaknesses, allowing companies not only to select optimal partners but also to enhance overall logistics network efficiency by aligning solutions with market demands.

Furthermore, performance analysis through clustering supports continuous improvement by tracking changes in clusters over time, enabling the identification of developmental trends among logistics service providers. This dynamic approach is particularly valuable in rapidly evolving market environments, where flexibility and innovation are critical for maintaining a competitive edge.

3 THEORY OF CLUSTERING METHODS

This section examines the theoretical foundations of clustering methods. Clustering is a statistical and machine learning technique designed to group data into clusters. The primary objective of clustering is to organize data so that points within the same cluster exhibit high similarity, while differences between clusters are maximized (Liu, Ke, Wei, & Hua, 2013).

The process begins with data collection and preparation. In the context of selecting logistics service providers, each provider constitutes a dataset with specific attributes, such as shipping cost, delivery time, geographic coverage, reliability, and additional services (e.g., warehousing, customs handling). Since these attributes—such as delivery time measured in days and cost measured in currency—are often on different scales, it is essential to normalize the data prior to applying clustering algorithms. Normalization ensures that all features contribute equally to the analysis, thereby improving the accuracy of comparisons and groupings (Kannan, Khodaverdi, Olfat, Jafarian, & Diabat, 2013; Shen, Olfat, Govindan, Khodaverdi, & Diabat, 2013).

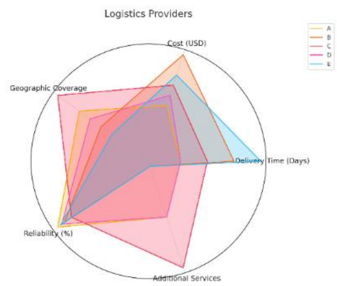

Fig. 1 Comparison chart of logistics providers' performance

The radar chart in Figure 1 provides a multidimensional representation of logistics providers' performance across various attributes, allowing for a visual comparison of their strengths and weaknesses, which aids in ranking and selection.

Several clustering algorithms are widely used, including:

- K-means clustering: One of the most common algorithms, k-means partitions data into *k* clusters by finding centroids and assigning each data point to the closest centroid, aiming to minimize the distance between data points and their cluster centroids (Mikhailov, 2003).
- Hierarchical clustering: This method creates a tree-like structure, progressively dividing data into smaller groups and applicable with either "bottom-up" or "top-down" approaches (Kaufman & Rousseeuw, 2005).
- Density-based clustering (DBSCAN): DBSCAN defines clusters based on data density, identifying regions of high density without requiring a predetermined number of clusters, making it especially useful for noisy data (Hosseini & Barker, 2016).

Clustering algorithms often require a measure of distance to determine the "similarity" between data points. Common distance metrics include:

- Euclidean distance: The geometric distance between two points, often used with numerical data (Opricovic & Tzeng, 2004).
- Manhattan distance: The sum of absolute differences, useful when data have discrete characteristics (Awasthi, Chauhan, & Goyal, 2011).

Once the data are normalized, the algorithm is executed. For k-means clustering, for example, a value for k (number of clusters) is selected. The algorithm then iteratively updates cluster centroids until the grouping stabilizes, meaning data points no longer shift clusters.

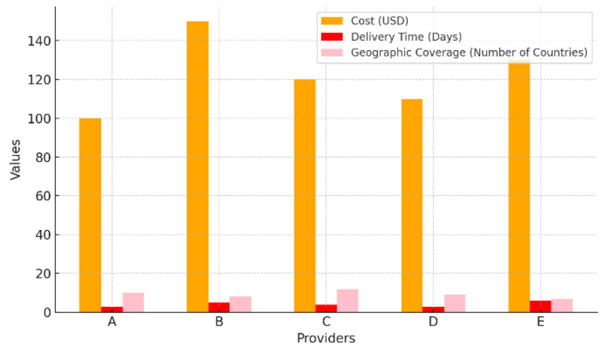

Fig. 2 Comparison of logistics providers' costs, delivery times, and coverage

Figure 2 illustrates a comparison of logistics providers based on cost (USD), delivery time (days), and geographic coverage (number of countries), making it easy to visualize differences in these fundamental metrics. This type of comparison helps identify optimal clustering for providers with similar characteristics, as the iterative process continuously updates cluster centroids, ensuring that data points settle into stable and accurate clusters.

After running the algorithm, data are organized into various clusters that need to be labeled and interpreted based on the characteristics of the data points within each cluster. For logistics providers, clusters might include groups such as low-cost providers with slower delivery times and broad coverage, fast and reliable providers with higher costs, or mid-cost providers with limited coverage and high reliability (Aguezzoul, 2014; Sarkis & Dhavale, 2015).

Once providers are grouped into clusters, companies can decide which cluster best suits their needs. For instance, if cost efficiency is the priority, the company may choose from the low-cost provider cluster. Alternatively, if speed and reliability are paramount, it may focus on the cluster containing fast and reliable providers (Zhang, Deng, Chan, Adamatzky, & Mahadevan, 2016).

4 MATHEMATICAL MODEL

The goal of the clustering task is to segment logistics service providers based on their characteristics, minimizing the variance within clusters. Below, we present the mathematical model of the k-means clustering algorithm, one of the most widely used methods for clustering data.

Model parameters

Let $X=\{x_1,x_2,\ldots,x_n\}$ represent the dataset of logistics service providers, where each x_i is a d-dimensional vector containing the characteristics of the i-th provider. Each feature of a provider is denoted by x_{ij}, which represents the j-th feature of the i-th provider. Formally, the vector x_i is defined as follows (1):

$$x_i=(x_{i1},x_{i2},\dots,x_{id}) \quad (1)$$

where:

- x_{ij} represents the *j*-th characteristic of the *i*-th provider. For example, if the features include shipping cost, delivery time, and reliability, then x_{i1} is the shipping cost, x_{i2} is the delivery time, and x_{i3} is the reliability for the *i*-th provider (Aguezzoul, 2014).

Clustering objective

The objective is to classify logistics service providers into k clusters while minimizing the variance within each cluster. This is achieved by minimizing the distance between data points in each cluster and the centroid of that cluster (Kaufman & Rousseeuw, 2005).

Objective function

The objective function of the *k*-means algorithm is to minimize the sum of squared distances between each data point and the centroid of its assigned cluster. Mathematically, this can be expressed as (2):

$$J(C)= \sum_{k=1}^{K} \sum_{x_i \in c_k} \| x_i - u_k \|^2 \quad (2)$$

where:

- K is the number of clusters.
- C_k is the *k*-th cluster containing data points that belong to the *k*-th cluster.
- μ_k is the centroid of the *k*-th cluster, representing the center of mass of the data points in that cluste.
- $\|x_i-\mu_k\|^2$ is the squared Euclidean distance between data point x_i and the centroid μ_k of the *k*-th cluster (Liu, Ke, Wei, & Hua, 2013)].

Explanation of the objective function

The objective function is based on the following steps:

1. Defining Clusters: First, the number of clusters K is defined, usually specified in advance. The initial centroids of the clusters are often selected randomly from the dataset (Kannan, Khodaverdi, Olfat, Jafarian, & Diabat, 2013).
2. Assigning Data Points to Clusters: Each data point in the dataset is assigned to the nearest cluster center based on the smallest distance to the centroid. This distance is given by (3):

$$C_i=\arg\min_k \|x_i-\mu_k\|^2 \quad (3)$$

where:

- C_i denotes the index of the cluster to which the *i*-th data point belongs.
- x_i represents the *i*-th data point.
- μ_k is the centroid of the *k*-th cluster.
- $\|x_i-\mu_k\|^2$ is the squared Euclidean distance between data point x_i and the centroid μ_k of the *k*-th cluster.
- The term $\operatorname{argmin}_k \|x_i-\mu_k\|^2$ selects the cluster with the smallest distance from x_i.

3. Cluster Index: C_i represents the cluster index to which x_i is closest, i.e., the centroid μ_k located at the minimum distance from x_i (Kaufman & Rousseeuw, 2005).
4. Updating Centroids: Once all data points are assigned to their respective clusters, the centroids are updated. The centroid of the *k*-th cluster is calculated as follows (4):

$$u_k = \frac{\sum_{x_i \in c_k} x^i}{|c_k|} \quad (4)$$

where:

- $|C_k|$ is the number of data points in the *k*-th cluster.

5. Iterative Optimization: The algorithm iterates until the centroids of the clusters and the assignment of data points to clusters stabilize, meaning the value of the objective function does not change significantly. During the iterations, the value of the objective function continuously decreases, and the distance between clusters is minimized as the optimization process progresses (Sarkis & Dhavale, 2015).

5 TOPSIS METHOD: A MULTI-CRITERIA DECISION-MAKING TECHNIQUE

The TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) method is a widely used multi-criteria decision-making technique designed to identify the alternative that most closely approximates an ideal solution. By evaluating alternatives based on multiple criteria, TOPSIS provides a structured approach to decision-making in complex scenarios. This section outlines the methodological steps of the TOPSIS framework, grounded in its mathematical model and theoretical foundations (Opricovic & Tzeng, 2004; Wang & Lee, 2009).

Definition of criteria and alternatives

- Criteria (*Cj*): These are the relevant characteristics or factors important for the decision-making problem, such as cost, quality, and time (Mikhailov, 2003; Sarkis & Dhavale, 2015).
- Alternatives (*Ai*): These represent the options to be evaluated and ranked (Shen, Olfat, Govindan, Khodaverdi, & Diabat, 2013).

Construction of the data matrix

The data matrix X is constructed, where xij represents the value of the j-th criterion for the i-th alternative (5):

$$x = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix} \quad (5)$$

where:

- *m*: number of alternatives
- *n*: number of criteria (Opricovic & Tzeng, 2004)

Normalization

To make values comparable across different units, the data are normalized, yielding normalized values rij for each entry (6):

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{k=1}^{m} x_{kj}^2}} \quad (6)$$

where r_{ij} is the normalized value for the *i*-th alternative with respect to the *j*-th criterion (Awasthi, Chauhan, & Goyal, 2011).

Weighting

The normalized values are weighted according to the importance of each criterion. The weights, wj reflect the significance of each criterion (7):

$$v_{ij}=w_j \cdot r_{ij} \quad (7)$$

where v_{ij} is the weighted value for the *i*-th alternative and *j*-th criterion, and w_j is the weight for the *j*-th criterion (Kannan, Khodaverdi, Olfat, Jafarian, & Diabat, 2013)

Selection of ideal and anti-ideal solutions

Ideal Solution (A*): The solution that provides the best performance across all criteria.

- For maximizing criteria: $A_j^* = \max_i v_{ij}$,if the goal is to maximize the criterion.
- For minimizing criteria: $A_j^* = \min_i v_{ij}$, if the goal is to minimize the criterion.

Worst Ideal Solution (A^-): The solution that provides the worst performance across all criteria.

- For maximizing criteria: $A_j^- = \min_i v_{ij}$, if the goal is to maximize the criterion.
- For minimizing criteria: $A_j^- = \max_i v_{ij}$, if the goal is to minimize the criterion.

Calculation of Distances

Calculate the distance of alternatives from the ideal (D_i^*) (7) and worst ideal (D_i^-) solutions (8):

$$D_i^* = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - A_j^*\right)^2} \quad (7)$$

$$D_i^- = \sqrt{\sum_{j=1}^{n} \left(v_{ij} - A_j^-\right)^2} \quad (8)$$

Where:

- D_i^* is the distance of the i-th alternative from the ideal solution.
- $D_i^-$is the distance of the i-th alternative from the worst ideal solution (Liu, Ke, Wei, & Hua, 2013).

Calculation of Relative Closeness

The relative closeness (C_i) to the ideal solution (9) is calculated to rank the alternatives:

$$C_i = \frac{D_i^-}{D_i^* + D_i^-} \quad (9)$$

Where:

- C_i is the relative closeness of the *i*-th alternative to the ideal solution (Wang & Lee, 2009).

Ranking

The final ranking is determined based on the C_i values. The alternative with the highest C_i value is the best choice as it is closest to the ideal solution (Shen et al., 2013).

6 PRACTICAL EXAMPLE: WEIGHTED TOPSIS AND K-MEANS CLUSTERING FOR LOGISTICS PROVIDERS

This section demonstrates the application of a weighted TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) and K-Means Clustering method to evaluate logistics providers. The focus is on combining multi-criteria decision-making with clustering for effective service selection.

Data Collection and Criteria Definition

To evaluate five logistics providers, the following criteria and weights were considered:

- Transportation Cost (USD): Weight = 0.30
- Delivery Time (Days): Weight = 0.25
- Geographic Coverage (Number of Countries): Weight = 0.20
- Reliability (% On-Time Deliveries): Weight = 0.15
- Additional Services (Number): Weight = 0.10

The data for each provider is summarized below:

Provider	**Cost (USD)**	**Time (Days)**	**Coverage (Countries)**	**Reliability (%)**	**Extra Services**
A	100	3	10	95	2
B	150	5	8	90	1

C	120	4	12	85	3
D	110	3	9	92	2
E	130	6	7	93	1

K-Means Clustering

The providers were grouped into clusters based on similarity in their characteristics. The K-Means algorithm classified them as follows:

- Cluster 1: Providers A, D
- Cluster 2: Providers B, E
- Cluster 3: Provider C

Data Normalization and Weighting

Normalization was performed to scale the values of each criterion. For cost and time, where smaller values are better, inverse normalization was applied. An example for cost normalization (10):

$$Normalized\ Cost_i = \frac{Max\ Cost - Cost_i}{Max\ Cost - Min\ Cost} \quad (10)$$

Provider	Normalized Cost	Weighted Cost
A	0	0×0.30=0
B	1	1×0.30=0.30
C	0,6	0,6×0,3=0,18
D	0,8	0.8×0.30=0.24
E	0,4	0.4×0.30=0.12

The weighted values for all criteria are computed similarly.

Application of TOPSIS

Within each cluster, TOPSIS was applied to identify the best provider. Steps included:

1. Defining the Ideal and Worst Alternatives:
   - The ideal alternative is the best weighted value for each criterion.
   - The worst alternative is the lowest weighted value for each criterion.
2. Calculating Distances:
   - Distance to the Ideal solution (D^+), (11)

$$D^+ = \sqrt{\sum_{i=1}^{n}\left(v_i - v_i^{ideal}\right)^2} \quad (11)$$

   - Distance to the Worst solution (D^-), (12)

$$D^- = \sqrt{\sum_{i=1}^{n}\left(v_i - v_i^{worst}\right)^2} \quad (12)$$

3. Computing the Preference score (C^*), (13)

$$C^* = \frac{D^-}{D^+ + D^-} \quad (13)$$

Results by cluster

- Cluster 1 (A, D): Based on the calculated C^* the provider with the higher score is selected as the best in the cluster.
- Cluster 2 (B, E): Repeat the above steps, identifying the top provider.
- Cluster 3 (C): Since only one provider is present, it is automatically the best.

Final selection

The best provider from each cluster is compared based on their C^* scores across all clusters. The provider with the highest overall C^* is selected as the optimal choice.

Summary

The integration of TOPSIS and K-Means Clustering offers a robust framework for evaluating logistics providers by combining multi-criteria decision analysis with similarity-based grouping. This hybrid method ensures that varying weights assigned to criteria are effectively accounted for, resulting in balanced and data-driven decisions. Moreover, modern technologies such as Big Data, artificial intelligence (AI), and real-time analytics further enhance the scalability and precision of this approach, making it particularly valuable in dynamic and complex supply chain environments.

7 DEVELOPMENT OPPORTUNITIES FOR TOPSIS AND K-MEANS METHODS IN THE DIGITAL ERA

In the digital era, marked by rapid technological advancements and an exponential increase in data volume, traditional decision-making and data analysis methods are undergoing profound transformations. The TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) and K-means clustering methods, long-established in decision support systems, now encounter new opportunities and challenges arising from emerging technologies and advanced data analysis techniques.

Justification for method efficiency

- Multi-Dimensional Decision-Making: The integration of TOPSIS and K-means clustering enables the evaluation of logistics service providers based on multiple criteria. K-means clustering facilitates the grouping of providers based on their similarities, while TOPSIS identifies the top-performing providers both within and across these group.
- Objective Evaluation and Ranking: TOPSIS objectively evaluates each provider relative to ideal and worst-case alternatives, minimizing decision-making errors caused by human subjectivity. This ensures that the selected providers demonstrate performance that is as close to the ideal as possible.

- Clustering and Specialization: K-means clustering allows for the grouping of providers into categories based on service similarities, effectively considering their specialization. Applying TOPSIS within these clusters enables a more precise evaluation of the top performers in each category.
- Transparency and Interpretability: This combined method simplifies complex decision-making problems and delivers clear, transparent results, empowering decision-makers to make well-informed choices.

Development opportunities in the digital era

- Databases and Big Data**:** The digital era enables the efficient processing of vast amounts of data. Leveraging Big Data analytics and cloud storage allows for the evaluation of service providers' performance and characteristics based on precise and detailed datasets (Liu, Ke, Wei, & Hua, 2013). Continuous data updates and robust database management support real-time analysis, leading to more accurate and timely decision-making (Kannan, Khodaverdi, Olfat, Jafarian, & Diabat, 2013).
- Machine Learning and Artificial Intelligence (AI)**:** Integrating machine learning algorithms and AI techniques enhances the application of k-means clustering and TOPSIS. AI identifies patterns and trends within datasets, enabling more accurate clustering and dynamic decision-making (Shen, Olfat, Govindan, Khodaverdi, & Diabat, 2013). Additionally, machine learning can automate the fine-tuning of clustering parameters and optimize TOPSIS evaluation factors (Hosseini & Barker, 2016).
- Visualization Tools**:** Visualization tools, such as dashboards and interactive graphs, assist decision-makers in comprehending analytical results more effectively. By visually representing data and outcomes, these tools simplify the interpretation of complex datasets and facilitate faster, more informed decision-making (Awasthi, Chauhan, & Goyal, 2011).
- Automation and Integration**:** The integration of automated data collection and processing systems streamlines clustering and evaluation processes. These automated systems enhance decision-making by making it faster, more accurate, and cost-effective (Sarkis & Dhavale, 2015).
- Real-Time Analytics**:** Real-time analytics tools enable continuous performance monitoring and instantaneous decision-making. This capability is especially critical in logistics services, where market dynamics change rapidly, and timely adaptation is essential (Zhang, Deng, Chan, Adamatzky, & Mahadevan, 2016).

In summary, the integration of TOPSIS and K-means clustering methods provides a robust solution for addressing complex decision-making challenges. The technological innovations of the digital era not only improve the efficiency of these methods but also open new avenues for more precise, faster, and adaptable decision-making processes. This integration ultimately contributes to enhanced competitiveness and operational efficiency across a wide range of industries.

8 CONCLUSIONS

This study explores the practical application of the TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) and K-means clustering methods, alongside the opportunities enabled by digital-age advancements. TOPSIS and K-means clustering are two influential decision-making techniques that effectively support the evaluation and optimization of complex problems, particularly in the selection of logistics service providers. The TOPSIS method prioritizes ranking alternatives based on their proximity to an ideal solution, facilitating evaluation across multiple criteria. Conversely, K-means clustering focuses on grouping data to minimize intra-cluster variation, thereby improving data structure and interpretability.

Technological advancements in the digital era, such as Big Data, machine learning, visualization tools, automation, and real-time analytics, significantly enhance the effectiveness of these methods. Big Data technologies and database systems enable the rapid and efficient processing of vast datasets, improving the accuracy of performance evaluation and decision-making. The integration of machine learning and artificial intelligence (AI) techniques refines clustering and evaluation processes by optimizing parameters and uncovering hidden patterns in data. Visualization tools provide clearer representations of data and results, facilitating faster and better-informed decision-making. Automation accelerates clustering and evaluation processes, making them more cost-effective, while real-time analytics ensures continuous performance monitoring and instant decision-making. These capabilities are particularly critical in dynamic industries like logistics, where timely adaptation is essential.

In conclusion, the combination of TOPSIS and K-means clustering methods offers a robust framework for addressing complex decision-making challenges. The technological advancements of the digital age not only enhance the efficiency of these methods but also create new opportunities for more precise, faster, and adaptable decision-making processes. Ultimately, this integration contributes to increased competitiveness and operational efficiency across diverse industries.

9 ACKNOWLEDGEMENTS

Supported by the university research scholarship program of the ministry for culture and innovation from the source of the national research, development and innovation fund.

10 REFERENCES

Aguezzoul, A. (2014). Third-party logistics selection problem: A literature review on criteria and methods. Omega, 49, 69–78. https://doi.org/10.1016/j.omega.2014.05.009

Awasthi, A., Chauhan, S. S., & Goyal, S. K. (2011). A fuzzy multicriteria approach for evaluating environmental performance of suppliers. International Journal of Production Economics, 126(2), 370–378. https://doi.org/10.1016/j.ijpe.2010.04.029

Govindan, K., Khodaverdi, R., & Jafarian, A. (2013). A fuzzy multi criteria approach for measuring sustainability performance of a supplier based on triple bottom line approach. Journal of Cleaner Production, 47, 345–354. https://doi.org/10.1016/j.jclepro.2012.04.014

Hosseini, S., & Barker, K. (2016). A Bayesian network model for resilience-based supplier selection. International Journal of Production Economics, 180, 68–87. https://doi.org/10.1016/j.ijpe.2016.07.007

Kannan, D., Khodaverdi, R., Olfat, L., Jafarian, A., & Diabat, A. (2013). Integrated fuzzy multi criteria decision making method and multi-objective programming approach for supplier selection and order allocation in a green supply chain. Journal of Cleaner Production, 47, 355–367. https://doi.org/10.1016/j.jclepro.2013.02.010

Kaufman, L., & Rousseeuw, P. J. (2005). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley.

Kumar, M., & Rahman, Z. (2016). Buyer supplier relationship and supply chain sustainability: Empirical study of Indian automobile industry. Journal of Cleaner Production, 131, 836–848. https://doi.org/10.1016/j.jclepro.2016.04.007

Liu, H., Ke, W., Wei, K. K., & Hua, Z. (2013). The impact of IT capabilities on firm performance: The mediating roles of absorptive capacity and supply chain agility. Decision Support Systems, 54(3), 1452–1462. https://doi.org/10.1016/j.dss.2012.12.016

Mazzarol, T., & Soutar, G. N. (2002). "Push-pull" factors influencing international student destination choice. International Journal of Educational Management, 16(2), 82–90. https://doi.org/10.1108/09513540210418403

Meixell, M. J., & Gargeya, V. B. (2005). Global supply chain design: A literature review and critique. Transportation Research Part E: Logistics and Transportation Review, 41(6), 531–550. https://doi.org/10.1016/j.tre.2005.06.003

Mentzer, J. T., Stank, T. P., & Esper, T. L. (2008). Supply Chain Management and its Relationship to Logistics, Marketing, Production, and Operations Management. Journal of Business Logistics, 29(1), 31–46. https://doi.org/10.1002/j.2158-1592.2008.tb00067.x

Mikhailov, L. (2003). Deriving priorities from fuzzy pairwise comparison judgements. Fuzzy Sets and Systems, 134(3), 365–385. https://doi.org/10.1016/S0165-0114(02)00383-4

Opricovic, S., & Tzeng, G. H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. European Journal of Operational Research, 156(2), 445–455. https://doi.org/10.1016/S0377-2217(03)00020-1

Sarkis, J., & Dhavale, D. G. (2015). Supplier selection for sustainable operations: A triple-bottom-line approach using a Bayesian framework. International Journal of Production Economics, 166, 177–191. https://doi.org/10.1016/j.ijpe.2014.11.007

Shen, L., Olfat, L., Govindan, K., Khodaverdi, R., & Diabat, A. (2013). A fuzzy multi criteria approach for evaluating green supplier's performance in green supply chain with linguistic preferences. Resources, Conservation and Recycling, 74, 170–179. https://doi.org/10.1016/j.resconrec.2012.09.006

Wang, T. C., & Lee, H. D. (2009). Developing a fuzzy TOPSIS approach based on subjective weights and objective weights. Expert Systems with Applications, 36(5), 8980–8985. https://doi.org/10.1016/j.eswa.2008.11.035

Zhang, X., Deng, Y., Chan, F. T. S., Adamatzky, A., & Mahadevan, S. (2016). Supplier selection based on evidence theory and analytic network process. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 230(3), 562–573. https://doi.org/10.1177/0954405414551105