International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-12 Issue-2, July 2023

Driving Performance Improvement of an Organization through Data Object Fusion

Lamia Alhazmi

Abstract: ***To succeed in today's data-driven economy, organizations must find ways to put their massive data stores to work competitively. This research delves into the possibility of using data object fusion techniques and, more significantly, consensus clustering to boost the efficiency of businesses in an area of expertise. A case investigation of the automotive service sector demonstrates potential results and puts theoretical knowledge into practice within an organization. Therefore, this study addresses the prospective benefits of data object fusion in the automotive service sector. Furthermore, by combining the findings of different clustering methods, consensus clustering can provide a more precise and reliable outcome. Moreover, a consistent representation of the data objects is obtained by applying this technique to disparate datasets acquired from different sources inside the organization, which improves decision-making and productivity in operations. The research highlights the significance of data quality and the selection of proper clustering techniques to achieve dependable and accurate data object fusion. The findings add to the expanding knowledge of using data-driven ways to enhance organizational performance in any emerging sector.***

Keywords: Information Fusion, Clustering, Decision-Making, Process Optimization

I. INTRODUCTION

Performance improvement refers to organizations' systematic efforts to enhance their operational effectiveness, productivity, efficiency, and overall outcomes. It involves identifying areas of improvement, setting goals, and implementing strategies to achieve higher performance levels. Performance improvement is crucial for organizations as it directly impacts their competitiveness, profitability, customer satisfaction, and long-term success. However, organizations face various challenges in this pursuit, such as outdated processes, lack of employee engagement, limited resources, and complex business environments. To overcome these challenges, adopting effective strategies is essential. Implementing data-driven approaches, leveraging technology, fostering a culture of continuous improvement, and aligning performance improvement initiatives with organizational goals are some of the critical strategies organizations should embrace. By prioritizing performance improvement and implementing effective strategies, organizations can enhance their capabilities, adapt to change, and achieve sustainable growth in today's dynamic and competitive landscape.

1.1 Data Object Fusion

Data object fusion involves integrating and consolidating diverse data sources, aiming to create a unified and comprehensive view of information within an organization. By integrating formal data objects from multiple sources such as databases, applications, and systems, data object fusion enables organizations to derive meaningful insights and make informed decisions. This approach eliminates data silos, enhances data quality, and improves data consistency across the organization. The benefits of data object fusion include improved reporting accuracy, enhanced data analytics capabilities, better data governance, and increased operational efficiency. Leveraging data object fusion techniques empowers organizations to uncover hidden patterns, identify trends, and gain a holistic understanding of their operations, ultimately driving performance improvement across various functions and processes.

1.2 Significance of Data-Driven Decision Making

Data-driven decision-making has become increasingly vital in today's business environment due to the exponential growth of data and the need for organizations to derive actionable insights. Data object fusion is crucial in enabling organizations to make informed decisions in this context. By integrating and harmonizing data from diverse sources, data object fusion provides a comprehensive and unified view of information, facilitating a holistic understanding of business operations. As a result, this integrated data approach enhances decision-making processes by offering a more accurate and complete picture of the organization's performance, customer behavior, market trends, and operational efficiency. In addition, it allows decision-makers to access timely and reliable data insights, enabling them to identify patterns, detect anomalies, and uncover hidden opportunities. Leveraging data object fusion, organizations can mitigate the risks associated with relying on fragmented and incomplete data, leading to more precise, data-driven decision-making. Ultimately, data-driven decision-making fueled by data object fusion empowers organizations to optimize resources, seize competitive advantages, and drive sustainable growth in today's data-driven business landscape.

1.3 Performance Improvement in Organizations

Performance improvement holds immense relevance for organizations in maintaining their competitiveness, achieving operational excellence, and delivering value to stakeholders.

Manuscript received on 18 May 2023 | Revised Manuscript received on 05 June 2023 | Manuscript Accepted on 15 July 2023 | Manuscript published on 30 July 2023.

*Correspondence Author(s)
Lamia Alhazmi*, Department of Management Information System, College of Business Administration, Taif University, P.O Box 11099, Taif, 21944, Saudi Arabia. Email: Lamia.s@tu.edu.sa, ORCID ID: 0000-0002-7027-5670

In today's dynamic and fast-paced business landscape, organizations must continuously enhance their performance to stay ahead. Performance improvement initiatives enable organizations to optimize processes, increase productivity, and drive innovation, improving efficiency and effectiveness. In addition, by streamlining operations, organizations can reduce costs, minimize waste, and enhance resource utilization, contributing to better financial performance.
Moreover, performance improvement enhances customer satisfaction by consistently delivering high-quality products or services, fostering customer loyalty, and gaining a competitive edge. Additionally, performance improvement initiatives positively impact employee engagement and morale, leading to higher productivity and reduced turnover rates. Furthermore, performance improvement helps organizations align their goals and strategies, enabling them to adapt to market changes and seize opportunities promptly. Ultimately, the holistic impact of performance improvement initiatives translates into improved profitability, customer loyalty, employee satisfaction, and overall business success.

1.4 Importance of Data Object Fusion in Performance Improvement

Data object fusion holds significant significance in performance improvement as it overcomes the limitations of traditional data management approaches. Conventional methods often result in fragmented data stored in separate systems, hindering organizations from obtaining a holistic view of their operations. Data object fusion addresses this challenge by integrating data from multiple sources, enabling organizations to create a unified and comprehensive data environment. This integration facilitates a holistic understanding of various aspects of operations, such as sales, customer behavior, inventory, and production.

Data integration plays a crucial role in data object fusion by combining data from disparate sources into a unified format, eliminating data silos and providing a cohesive view. Moreover, data object fusion emphasizes the importance of data quality, ensuring that integrated data is accurate, complete, and reliable. This ensures that organizations make decisions based on trustworthy information. The primary benefits of data object fusion are illustrated in Figure 1.

Additionally, data analysis is a crucial component of data object fusion. Organizations can extract meaningful insights from integrated data by leveraging advanced analytics techniques. These insights enable organizations to identify trends, patterns, and anomalies, empowering them to make informed decisions and drive performance improvement initiatives.

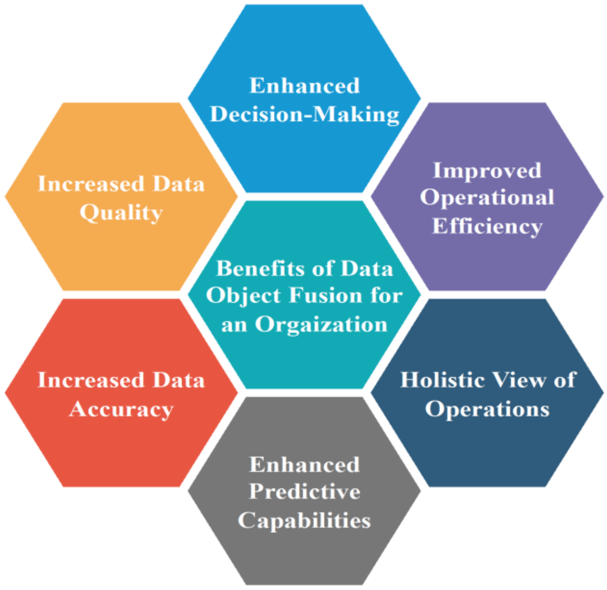

Figure 1. Benefits of Data Object Fusion

Data object fusion facilitates a holistic view of operations by integrating data, ensuring data quality, and enabling practical data analysis. Furthermore, this integration enhances organizational decision-making capabilities and drives performance improvement across various aspects of the business. Thus, in this research, consensus clustering is utilized to improve the performance of an organization.

Consensus clustering combines multiple clustering algorithms or runs on a dataset, resulting in improved accuracy and reliability. In addition, it helps reduce the impact of algorithmic biases or limitations, providing a more robust and accurate clustering solution.

1.5 Motivation and Scope

The study aims to investigate the potential benefits of data object fusion in the automobile service industry, focusing on how consensus clustering can lead to a more robust and reliable clustering solution. Furthermore, it emphasizes the importance of data quality and suitable clustering algorithms in achieving reliable and accurate data object fusion.

The study's outcomes are expected to contribute to the existing knowledge on leveraging data-driven approaches for performance improvement in any developing industry. Furthermore, the abstract implies that the research has broader implications beyond the automobile service industry and can be applied to other sectors seeking to enhance organizational performance through data object fusion methodologies.

The comprehensive study with appropriate investigative outcome discussion is organized as follows: Section 2 deeply reviews the existing work focused on data object fusion processes. Section 3 delineates the core procedures of the proposed cluster process. Section 4 elaborates on the concerned datasets and the experimental utilization to analyze the proposed model. Section 5 briefs the concluded data of the study with actual and observed facts and possible future work.

II. RELATED WORK

The literature survey on data object fusion is significant as it provides insights into practical strategies and methodologies for integrating and leveraging diverse data sources, enabling organizations to gain a holistic view of operations and make informed decisions. Thus, a few relevant works are reviewed in the following manner.

De Vin et al. [1] discuss the concept of information fusion and its application in the defense sector as an established research area. However, it points out that in the manufacturing industry, information fusion is not as well-established, except for sensor/data fusion for automatic decision-making. From a technical standpoint, the article provides a concise overview of information fusion in the defense sector and identifies the need for its application in the manufacturing industry. Furthermore, it introduces the research program's focus on information fusion for human decision-making in manufacturing and highlights the potential relevance of virtual manufacturing and simulation models.

However, the article lacks in-depth technical details and specific examples or case studies to support the claims and findings. It would be beneficial to provide more concrete examples of how information fusion techniques can be applied in manufacturing scenarios and how they can improve decision-making or enhance specific manufacturing processes.

Sanchez-Pi et al. [2] focus on achieving and demonstrating exemplary performance in the oil and gas industry, particularly concerning occupational health and safety (OHS) issues. It highlights the need for petroleum companies to comply with international OHS policies, which require them to register and analyze any anomalies or accidents during operations. The authors proposed a data fusion architecture coupled with a machine-learning layer to provide abstractions and inferences over the data. Their objective is to enable analysts to infer relevant root-cause-and-effect relations within the industry. They have developed a system based on their model and applied it to data from a petroleum company.

From a technical perspective, the article presents a relevant problem statement concerning OHS issues in the oil and gas industry. It highlights the significance of data management and analysis in addressing these concerns. The proposal of data fusion architecture with a machine-learning layer seems promising as it aims to automate the procesures in extracting valuable insights from extensive and diverse data sources.

Meng et al. [3] present a method that combines the Logistic and SVM classification models using evidence theory to address a binary classification problem. It highlights the importance of accuracy in the Logistic and SVM models. They proposes a D-S combination method that assigns uncertainty to the frame of discernment for improved classification effectiveness. Experimental results on three datasets demonstrate that the proposed method outperforms individual classification models, enhancing prediction accuracy.

The researchers utilize the probability output capability of the Logistic and SVM models to perform data fusion at the decision level. It suggests the possibility of exploring multiple classification models suitable for different data sources and types in future research for decision fusion. Additionally, it acknowledges the presence of objects without a definitive classification in the final decision, indicating the need for further consideration of classification and discrimination for such objects. However, the article lacks detailed information about the methodology employed, such as the specific evidence theory techniques used or the process of assigning uncertainty to the frame of discernment. Additionally, the article does not provide particular insights into the datasets used or the experimental setup, limiting the ability to assess the reproducibility and generalizability of the results[4].

Huang & Huang. [5] presents a method for improving the information fusion and early warning system in rail transit signal operation and maintenance using big data from the Internet of Things (IoT) [6-7]. Furthermore, it introduces the concept of an intelligent sensing network within the IoT [8], highlighting the integration of perception, information processing, and communication.

The article explains the critical components of the intelligent sensing network and analyzes their significance. The researchers then discuss a dynamic OD traffic flow estimation model based on path recognition. Next, it proposes using the time a vehicle spends on the main line to calculate its running speed, which is more suitable for free-flowing traffic situations. Finally, the data is normalized using least square estimation, enabling comprehensive analysis of information and data. However, more technical details, empirical evidence, and practical examples would enhance the comprehensibility and credibility of the research.

Dumancas et al. [9] introduce the concept of data fusion and its application in predictive toxicology and chemical risk assessments. It emphasizes the importance of integrating toxicological datasets and endpoints from different sources to estimate chemical risk. The goal is to integrate diverse toxicological datasets and endpoints from various sources to evaluate the potential risk associated with chemicals. From a technical standpoint, the article provides an overview of the methodological approaches used in data fusion architecture. However, it lacks specific details about the techniques and algorithms employed. Therefore, discussing specific data fusion methods commonly used in toxicology and risk assessment studies, such as Bayesian networks, ensemble methods, or machine learning algorithms, would be beneficial.

The review shows that none has provided more technical details, specific methodological approaches, and practical examples that would strengthen the comprehensive understanding of data object fusion strategies.

III. METHODOLOGY

In this research, we employed Consensus Clustering as a data object fusion technique. Consensus Clustering combines multiple clustering algorithms to achieve a more robust and accurate clustering solution.

In this case, we will consider using the k-means and spectral clustering algorithms for consensus clustering in the car manufacturing and service industry. The mathematical computation section is provided in the following sections. Beyond the generalized computation [10], it is essential to illustrate the regulated process through the case study. Thus, the automobile servicing sector is considered to demonstrate the technical procedures of the proposed work.

The work processes of automobile service industries involve multiple stages and activities to ensure vehicle maintenance, repair, and overall servicing. Throughout these processes, various types of data are generated, which are crucial for the efficient functioning of the industry. Figure 2 briefs the work process and the types of data generated across different stages.

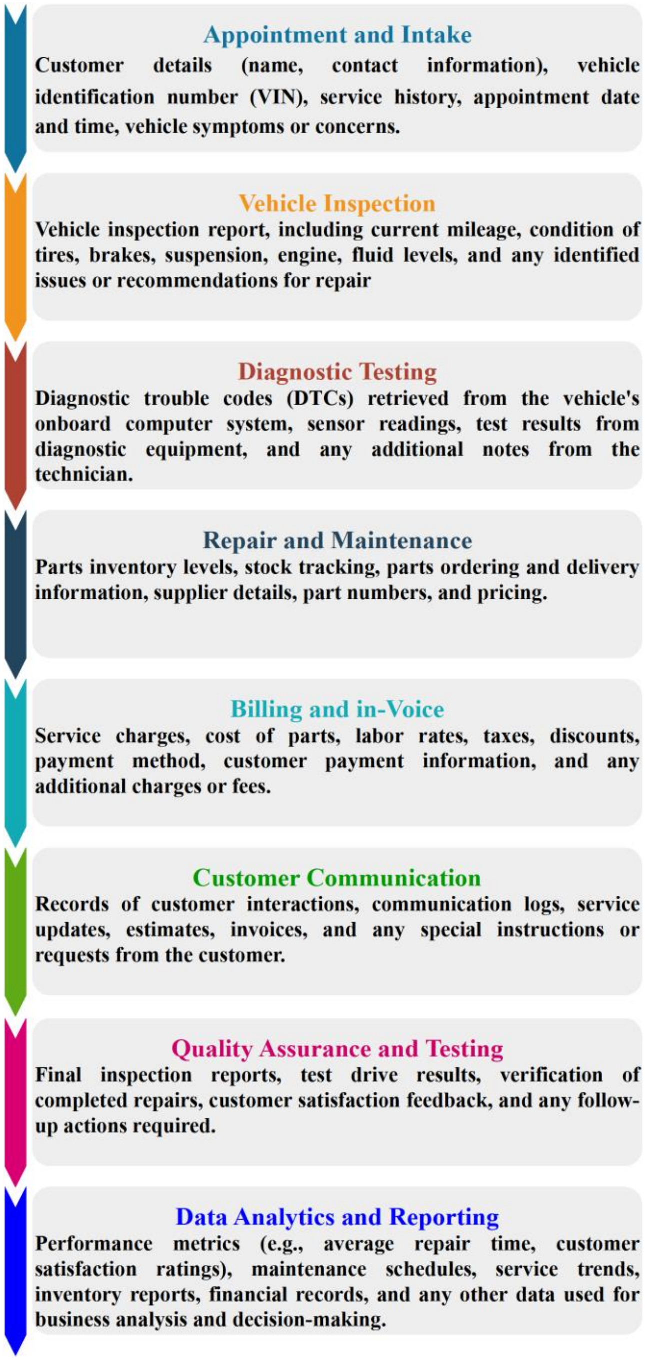

Figure 2. Work Processes of Automobile Service Industries

In the automobile servicing industry, Consensus Clustering can be applied to gain insights into various aspects such as customer segmentation, vehicle maintenance patterns, or component analysis. Let's delve into the detailed explanation of consensus matrix construction using k-means clustering, consensus matrix construction using spectral clustering, Thresholding and Similarity Computation, and data fusion through consensus clustering.

1.1 Consensus Matrix Construction using K-means Clustering

This process comprises two significant steps. They are,

Applying k-means clustering: In this step, the k-means algorithm is used to cluster the data objects into M clusters [11]. Each data cluster is represented as $\mathfrak{x}_k$, where k varies from 1 to M. The k-means algorithm iteratively assigns data objects to clusters based on the proximity to the cluster centroids and updates the centroids until convergence. After convergence, we obtain a set of M clusters as:

$$\mathfrak{x}_k = \left\{\mathfrak{x}_1^k, \mathfrak{x}_2^k, \cdots, \mathfrak{x}_M^k\right\} \tag{1}$$

1.2 Compute the co-occurrence matrix, $m_{(i,j)}^k$

The co-occurrence matrix, $m_{(i,j)}^k$ quantifies the pair-wise similarity between the cluster assignments of data objects. For each pair of clusters, $(\mathfrak{x}_i^k, \mathfrak{x}_j^k)$, we compute their co-occurrence by considering the data objects that belong to both clusters and those that belong to either of the clusters.

$$m_{(i,j)}^k = {\left|\mathfrak{x}_i^k \cap \mathfrak{x}_j^k\right|}\Big/{\left|\mathfrak{x}_i^k \cup \mathfrak{x}_j^k\right|} \tag{2}$$

The numerator, $\left|\mathfrak{x}_i^k \cap \mathfrak{x}_j^k\right|$, represents the number of data objects that are assigned to both cluster $(\mathfrak{x}_i^k, \mathfrak{x}_j^k)$. The denominator $\left|\mathfrak{x}_i^k \cup \mathfrak{x}_j^k\right|$ represents the number of data objects that are assigned to either cluster $(\mathfrak{x}_i^k, \mathfrak{x}_j^k)$ or both.

By dividing the intersection by the union, we obtain a similarity measure that ambit between 0 and 1. A higher value indicates a stronger co-occurrence or similarity between the clusters.

The computation of the co-occurrence matrix, $m_{(i,j)}^k$ enables capturing the pair-wise similarity between the cluster assignments obtained from the K-means clustering. This matrix serves as one component of the consensus matrix used in Consensus Clustering. The co-occurrence matrix quantifies how often pairs of data objects are clustered together, providing insights into the common patterns and associations among the data objects within the automobile service industry context.

1.3 Consensus Matrix Construction using Spectral Clustering

Similar to the K-means clustering step, we compute a co-occurrence matrix, Y spectral, to capture the pair-wise similarity between the cluster assignments obtained from spectral clustering.

Retrieval Number: 100.1/ijrte.B77360712223
DOI: 10.35940/ijrte.B7736.0712223
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

The significant distinction lies in the clustering algorithms themselves. K-means clustering focuses on minimizing the sum of squared distances between data objects and cluster centroids, while spectral clustering considers the spectral properties of a similarity graph. Consequently, the consensus matrices obtained from k-means and spectral clustering capture different aspects of the clustering solutions. Combining these consensus matrices can lead to a more comprehensive and robust clustering solution through the Consensus Clustering approach.

1.4 Thresholding and Similarity Computation

The thresholding and similarity computation step aims to convert the consensus matrix into a binary similarity matrix, where each element represents the presence or absence of consensus between data objects. By setting a threshold, we define the level of agreement required for data objects to be considered similar. This binary similarity matrix serves as input for the subsequent steps in the Consensus Clustering methodology within the automobile service industry, such as the final clustering step and evaluation.

Set a threshold value*:* A threshold value, τ, is determined to establish the strength of consensus between data objects. The threshold selection depends on the desired level of agreement required for data objects to be considered similar[20-21]. The threshold value τ is chosen based on domain knowledge, exploratory analysis, or statistical techniques.

Compute the similarity matrix, $Я_{ij}$: The similarity matrix, $Я$, is computed by applying the threshold to the consensus matrix, $ɰ$, for each element $ɰ_{ij}$ in the consensus matrix. If $ɰ_{ij}$ is greater than the threshold τ (i.e., $ɰ_{ij} > \tau$), then $Я_{ij}$ is set to 1, indicating that there is a consensus or similarity between data objects i and j.

Table 2. Consensus Clustering Process

Input: $\eta = \{d_1, d_2, \cdots, d_n\}$ *//η ⟶ Input dataset; d_n ⟶ 'n' data objects*

Output: $\{d_1 \cup d_2 \cup \cdots \cup d_n\}$ *//Data object fusion*

1: Identify *M*

2: Consensus Matrix Construction

2.1: $ӿ_k = \{ӿ_1^k, ӿ_2^k, \cdots, ӿ_M^k\}$ *//apply K-means*

2.2: Compute Co-occurrence matrix

$$m_{(i,j)}^k = {\left|ӿ_i^k \cap ӿ_j^k\right|}\Big/{\left|ӿ_i^k \cup ӿ_j^k\right|}$$

2.3: $\varpi_s = \{\varpi_1^s, \varpi_2^s, \cdots, \varpi_M^s\}$ *//apply Spectral clustering* [17]

2.4: Compute Co-occurrence matrix

$$m_{(i,j)}^s = {\left|\varpi_i^s \cap \varpi_j^s\right|}\Big/{\left|\varpi_i^s \cup \varpi_j^s\right|}$$

3: Construct Consensus Matrix, $ɰ$

$$ɰ = {\left(m_{(i,j)}^k + m_{(i,j)}^s\right)}\Big/{2}$$

4: Thresholding and Similarity Computation

4.1: set **τ**

4.2: compute similarity matrix, $Я_{ij}$

$$Я_{ij} = \begin{cases} 1, & if\left(ɰ_{ij} > \tau\right) \\ 0, & otherwise \end{cases}$$

4.3: Apply clustering computation to $Я_{ij}$, and obtain consensus clustering

If $ɰ_{ij}$ is less than or equal to the threshold (i.e., $ɰ_{ij} \leq \tau$), then $Я_{ij}$ is set to 0, indicating that there is no consensus or similarity between data objects *i* and *j*. The resulting similarity matrix, $Я$, is a binary matrix where $Я_{ij} = 1$ indicates that data objects i and j are considered similar, while $Я_{ij} = 0$ indicates dissimilarity. Table 2 represents the data fusion process through the consensus clustering process.

IV. PERFORMANCE ANALYSIS

1.1 Datasets

For assessing the proposed model, two open-source datasets are considered: OpenXC and Auto MPG.

OpenXC (Data Set - OpenXC, n.d.)[19] is an open-source platform that collects data from various sensors in a vehicle. It provides access to real-time and logged vehicle data, including engine RPM, throttle position, speed, and more. This dataset can be utilized for diagnostic testing, quality assurance, and maintenance analysis.

The "Auto MPG" dataset [12] from the UCI Machine Learning Repository consists of various attributes related to automobiles. The dataset contains information that can be used for data fusion techniques to improve an organization's performance. The "Auto MPG" dataset, sourced from the UCI Machine Learning Repository, was derived from the StatLib library at Carnegie Mellon University. It is a slightly altered version of the original dataset found in the StatLib library[17]. To align with [12] application of predicting the "mpg" attribute, eight instances with unknown values for "mpg" were eliminated from the dataset. In the context of the "Auto MPG" dataset, data fusion techniques can be applied to leverage the available attributes and improve the organization's understanding of factors affecting fuel efficiency (mpg) in automobiles.

1.2 Experimental Setup

Simulation platforms for data analysis and clustering include Python 3.6 with Scikit-learn version 0.24.1 (Installing Scikit-Learn, n.d.). The hardware configuration for the experimental scenario consists of a XEON 4210R processor 2.4 GHz 13.75 MB. Table 2 represents the essential attributes of each clustering algorithm.

Table 2. Key Attributes of the Clustering Process

Algorithms		Parameters	Values
Algorithms	K-Means	Number of clusters (*M*)	10
		Maximum number of iterations	200
		Distance metric	Euclidean
	Spectral clustering	Number of clusters (*M*)	10
		Affinity matrix	Co-occurrence, Thresholding and similarity computation [18]
		Spectral embedding method	Locally Linear Embedding [13]
		Number of nearest neighbors	10
		Gamma parameter	1.0, or 10.0

Retrieval Number: 100.1/ijrte.B77360712223
DOI: 10.35940/ijrte.B7736.0712223
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Consensus clustering	Number of clusters (M)	10
	Number of clusters	2
	Clustering algorithms	K-Means, or Spectral Clustering [24]
	Similarity measure	Euclidean distance [14]
	Consensus criterion	Cluster assignment agreement [15]
	Re-sampling method	Sub-sampling
	Number of re-sampling iterations	100

1.3 Result Discussions

It is essential to compare the performance of the relevant clustering, like K-means and spectral clustering, to analyze the clustering effectiveness of the proposed model whenever utilized individually. As a result, the clustering accuracy and operational effectiveness are evaluated, which is used to validate the data object fusion approach.

Clustering Accuracy: The clustering accuracy measures the agreement between the consensus clustering results and the ground truth (if available) or an established clustering solution[22-23]. Figure 3 highlights higher clustering accuracy for consensus clusters, indicating a better alignment between the fused data objects and their actual clusters.

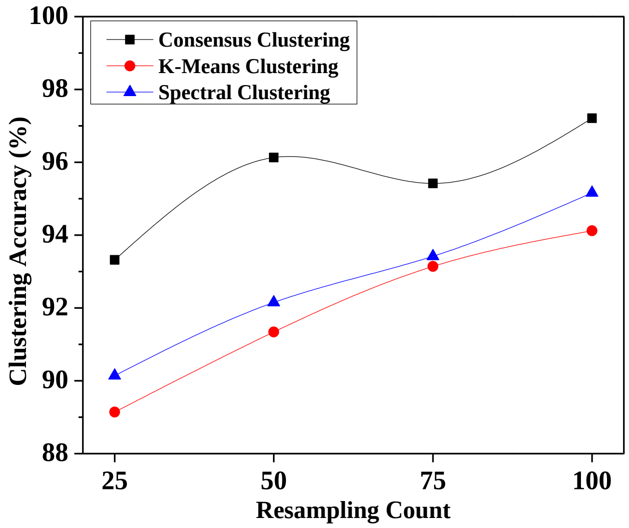

Figure 3. Analysis of Clustering Accuracy

The observed outcome from Figure 3 suggests that the data object fusion process, implemented through consensus clustering, leads to higher clustering accuracy compared to individual clustering algorithms (k-means and spectral clustering). As the number of re-sampling iterations increases to 100, the clustering accuracy improves to 97.21% for consensus clustering, 94.12% for k-means clustering, and 95.16% for spectral clustering. The clustering accuracy consistently enhances with an increase in the number of re-sampling iterations for all three algorithms. However, consensus clustering consistently outperforms k-means and spectral clustering.

Operational Cost Reduction: Figure 4 visually represents how the data object fusion process contributes to optimizing operational expenses over time. Operational cost reduction (ϕ) measures the extent to which the data object fusion process has helped optimize operational expenses within the organization. The computation of estimating the metric is expressed as,

$$\phi = \left[\left(B_\phi - \psi_\phi\right)/B_\phi\right] \times 100 \quad (3)$$

From equation (3), B_ϕ denotes the initial operation cost, whereas ψ_ϕ indicates the current operating cost at a given period T. This metric can consider solid factors such as resource utilization, energy consumption, material waste, or inventory management. A lower operational cost value indicates that the fusion process has enabled organizations to identify areas for cost savings and implement strategies to reduce operational expenses[16]. It allows for the analysis of trends and the assessment of the effectiveness of the fusion process in achieving cost reduction objectives. The X-axis represents the time-period during which the data object fusion process is implemented and monitored. It allows for the analysis of the operational cost reduction trend over time. The Y-axis represents the percentage reduction in operational costs achieved through the data object fusion process. It quantifies the cost-saving impact of the fusion methodology on various factors.

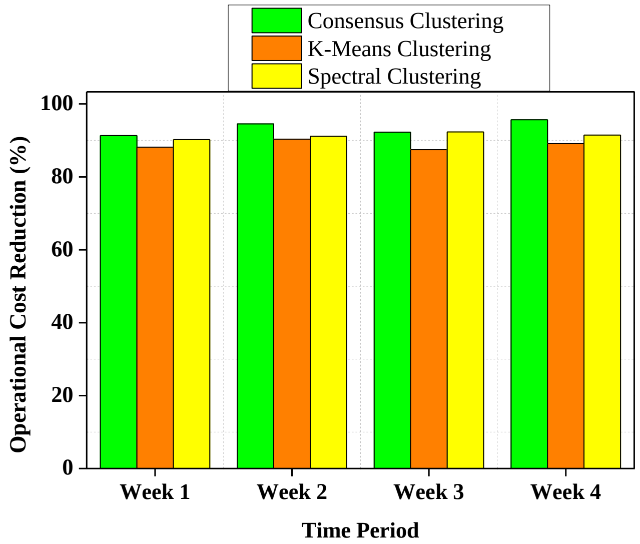

Figure 4. Analysis of Operational Reduction Cost

The operational cost reduction varies across different time periods for all three algorithms, as shown in Figure 4. The observed outcome signifies that implementing service processing through consensus clustering leads to a higher operational cost reduction than individual clustering algorithms. On average, the consensus clustering algorithm achieves the highest operational cost reduction of 93.44%, followed by spectral clustering with an average of 91.28% and k-means clustering with an average of 88.76%.

Quality Improvement: This metric assesses product or service quality improvement resulting from the data object fusion process. It can be measured by tracking the number of defects or errors encountered in the service processes. A lower defect rate indicates that the fusion process has helped identify and address quality issues, improving overall product or service quality. The computation involved in estimating the defect rate (ζ) is expressed as,

$$\zeta = {}^{\mathbb{T}_\zeta}/_{\Upsilon} \quad (4)$$

Retrieval Number: 100.1/ijrte.B77360712223
DOI: 10.35940/ijrte.B7736.0712223
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

where, $\mathbb{T}_{\zeta}$ denotes the number of defects tracked, Υ indicates the total number of services delivered in a given period.

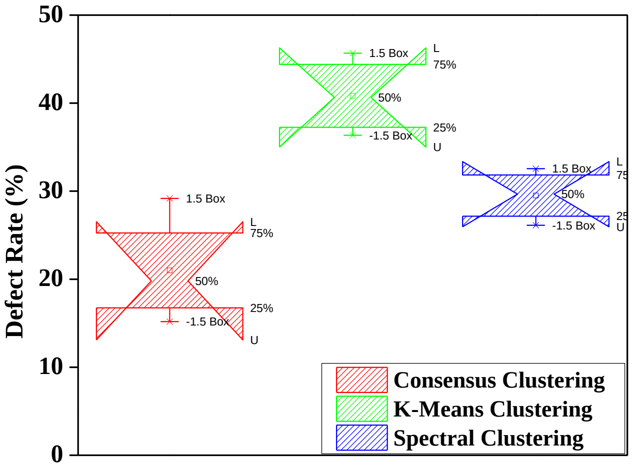

Figure 5. Analysis of Defect Rate(%)

The observed outcome pertains to the defect rate (%) resulting from the data object fusion process. The results are presented for the four-week time-periods, comparing the performance of consensus clustering, k-means clustering, and spectral clustering algorithms. On average, the defect rate achieved through consensus clustering is 21.00%. Through K-means and spectral clustering, it is observed to be 40.83% and 29.50%, respectively. The technical analysis of the observed outcome indicates that the data object fusion process, implemented through consensus clustering, results in the lowest defect rate among the three algorithms. Thus, the proposed model aids in identifying and addressing quality issues, leading to improved overall service quality.

V. CONCLUSION AND FUTURE WORK

The research demonstrates that data object fusion, mainly through consensus clustering, holds significant potential for enhancing organizational efficiency in the automotive service sector. By combining multiple clustering methods, consensus clustering provides a more precise and reliable outcome. It enables the organization to consistently represent data objects by integrating disparate datasets from various sources, leading to improved decision-making and operational productivity. The findings highlight the importance of data quality and the selection of appropriate clustering techniques to ensure dependable and accurate data object fusion.

The study also reveals notable outcomes regarding defect rate reduction and operational cost reduction. Consensus clustering consistently achieves the lowest defect rate, with an average of 21.00%, surpassing both k-means clustering (40.83%) and spectral clustering (29.50%). Furthermore, consensus clustering exhibits the highest operational cost reduction, with an average of 93.44%, followed by spectral clustering (91.28%) and k-means clustering (88.76%).

As a future work, apart from defect rate and operational cost reduction, we planned to consider other relevant performance metrics such as customer satisfaction, response time, and revenue growth can provide a comprehensive assessment of the impact of data object fusion on organizational performance.

DECLARATION

Funding/ Grants/ Financial Support	No, We did not receive.
Conflicts of Interest/ Competing Interests	No conflicts of interest to the best of my knowledge.
Ethical Approval and Consent to Participate	No, the article does not require ethical approval and consent to participate with evidence
Availability of Data and Material/ Data Access Statement	Not relevant.
Author Contributions	I am only the sole author of the article.

REFERENCES

1. De Vin, L. J., Ng, A. H. C., Oscarsson, J., & Andler, S. F. (2006). Information fusion for simulation-based decision support in manufacturing. Robotics and Computer-Integrated Manufacturing, 22(5-6), 429–436. [CrossRef]
2. Sanchez-Pi, N., Martí, L., Molina, J. M., & Bicharra García, A. C. (2016). Information Fusion for Improving Decision-Making in Big Data Applications. Computer Communications and Networks, 171–188. [CrossRef]
3. Meng, F., Li, A., & Liu, Z. (2022). An Evidence theory and data fusion-based classification method for decision making. Procedia Computer Science, 199, 892–899. [CrossRef]
4. Almalawi, A., Khan, A.I., Alsolami, F., Abushark, Y.B. and Alfakeeh, A.S., 2023. Managing Security of Healthcare Data for a Modern Healthcare System. Sensors, 23(7), p.3612.
5. Huang, C., & Huang, Y. (2022). Information fusion early warning of rail transit signal operation and maintenance based on big data of the Internet of things. Sustainable Computing: Informatics and Systems, 35, 100763. [CrossRef]
6. Sarker, I. H., Khan, A. I., Abushark, Y. B., & Alsolami, F. (2022). Internet of Things (IoT) Security Intelligence: A Comprehensive Overview, Machine Learning Solutions and Research Directions. Mobile Networks and Applications. [CrossRef]
7. Karthick Raghunath, K. M., Koti, M. S., Sivakami, R., Vinoth Kumar, V., NagaJyothi, G., & Muthukumaran, V. (2022). Utilization of IoT-assisted computational strategies in wireless sensor networks for smart infrastructure management. International Journal of System Assurance Engineering and Management. [CrossRef]
8. Pandiyan, S., M., A., R., M., K.M., K. R., & G.R., A. R. (2020). Heterogeneous Internet of Things organization Predictive Analysis Platform for Apple Leaf Diseases Recognition. Computer Communications, 154, 99–110. [CrossRef]
9. [Dumancas, G. G., Krichbaum, M., Solivio, B., Lubguban, A. A., & Malaluan, R. M. (2023). Data fusion applications in toxicology. Reference Module in Biomedical Sciences. [CrossRef]
10. H. Almulihi, A., Alassery, F., Irshad Khan, A., Shukla, S., Kumar Gupta, B., & Kumar, R. (2022). Analyzing the Implications of Healthcare Data Breaches through Computational Technique. Intelligent Automation & Soft Computing, 32(3), 1763–1779. [CrossRef]
11. Zheng, D., & Wang, Q. (2013). Selection algorithm for K-means initial clustering center. Journal of Computer Applications, 32(8), 2186–2188. [CrossRef]
12. J. Ross Quinlan. (1993). Combining Instance-Based and Model-Based Learning. Elsevier EBooks, 236–243. [CrossRef]
13. Kanghua Hui, & Chunheng Wang. (2008). Clustering-based locally linear embedding. 2008 19th International Conference on Pattern Recognition. [CrossRef]

Retrieval Number: 100.1/ijrte.B77360712223
DOI: 10.35940/ijrte.B7736.0712223
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

14. Afghani, S. A., & Putra, W. Y. M. (2021). Clustering with Euclidean Distance, Manhattan - Distance, Mahalanobis - Euclidean Distance, and Chebyshev Distance with Their Accuracy. Indonesian Journal of Statistics and Its Applications, 5(2), 369–376. [CrossRef]
15. Jun Wang. (1999). A linear assignment clustering algorithm based on the least similar cluster representatives. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 29(1), 100–104. [CrossRef]
16. Kumar, R., Khan, A.I., Abushark, Y.B., Alam, M.M., Agrawal, A. and Khan, R.A., 2020. An integrated approach of fuzzy logic, AHP and TOPSIS for estimating usable-security of web applications. IEEE Access, 8, pp.50944-50957.
17. sklearn. cluster. Spectral Clustering. (n.d.). Scikit-Learn. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralClustering.html
18. Ling, G., Wang, M., & Feng, J. (2011). Clustering ensemble method based on co-occurrence similarity. Journal of Computer Applications, 31(2), 441–445. [CrossRef]
19. Data Set - OpenXC. (n.d.). Openxcplatform.com. Retrieved May 12, 2023, from http://openxcplatform.com/about/data-set.html
20. von Luxburg, U. (2007). A tutorial on spectral clustering. Statistics and Computing, 17(4), 395–416. [CrossRef]
21. Alam, M.M., Khan, A.I. and Zafar, A., 2016. A comprehensive study of software product line frameworks. International Journal of Computer Applications, 151(3).
22. Barnawi, A., Al-Talhi, A.H., Qureshi, M. and Khan, A.I., 2012. Novel component based development model for sip-based mobile application. arXiv preprint arXiv:1202.2516.
23. Survey Report on K-Means Clustering Algorithm. (2017). International Journal of Modern Trends in Engineering & Research, 4(4), 218–221. [CrossRef]
24. Almalawi, A., Khan, A.I., Alsolami, F., Abushark, Y.B. and Alfakeeh, A.S., 2023. Managing Security of Healthcare Data for a Modern Healthcare System. Sensors, 23(7), p.3612. [CrossRef]

AUTHORS PROFILE

Lamia Alhazmi is an Assistant Professor in the Department of Management Information System at Taif University, Saudi Arabia. Lamia has a Ph.D. in Information System Management from Latrobe University, Australia, a master's degree from Latrobe University, Australia, and a bachelor s degree in Business of Information System Management from King Abdulaziz University, Saudi Arabia. She is a Fellow of the UK Higher Education Academy in recognition of attainment against the UK Professional Standards Framework for teaching and learning support in higher education. Lamia has published papers in esteemed academic journals.

Retrieval Number: 100.1/ijrte.B77360712223
DOI: 10.35940/ijrte.B7736.0712223
Journal Website: www.ijrte.org

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)