Spare Part Selection for Storage in PLN Suku Cadang Warehouse Using Clustering Analysis

Ahmad Mushawir[a,b], Akhmad Faeda Insani[a,c], Aditya Adiaksa[a,c], Zainuddin Zainuddin[a,c], Sparisoma Viridi[a]

[a]*Institut Teknologi Bandung, Jl. Ganesha No. 10, Bandung, 40132, West Java, Indonesia*
[b]*PLN Nusantara Power, Head Office, Jakarta, 12520, DKI, Indonesia*
[c]*PT PLN (PERSERO), Head Office, Jakarta, 121260, DKI, Indonesia*

Abstract

PLN Suku Cadang (PLNSC) faces challenges in determining which spare parts should be stored in warehouses to support power plant operations. This study employs the k-Means clustering method to analyze historical data, including demand frequency, lead time, quantity, and cost. Using this method, spare parts are grouped into several clusters to identify storage priorities. The clustering results are further validated using specific criteria: high demand frequency, low lead time, high quantity, and low cost. This study demonstrates that the clustering method can help optimize inventory management, reduce downtime risks, and improve cost efficiency in storage.

Keywords: clustering, power plants, spare parts

1. INTRODUCTION

PLN Suku Cadang (PLNSC) is a key supplier of spare parts for power generation units across Indonesia. Its current operational model, which involves directly supplying spare parts from vendors to power plants without centralized storage, presents significant challenges. These include delays in delivery, limited stock availability, and heightened risks of operational downtime due to insufficient inventory planning. In the highly dynamic energy sector, ensuring the availability of critical spare parts is essential to maintaining operational continuity and reliability.

The absence of centralized warehouses necessitates a strategic approach to inventory management. Spare parts vary significantly in their demand frequency, quantity, lead time, and cost to operations. As such, a one-size-fits-all strategy for inventory is not feasible. Traditional inventory management approaches, which often rely on subjective judgment or past experiences, lack the analytical rigor to address these challenges comprehensively. This limitation underscores the need for a systematic and data-driven solution to optimize inventory decisions.

Clustering techniques have been widely adopted in inventory management to address similar challenges in various industries. For instance, clustering retail stores for inventory transshipment has been shown to optimize inventory redistribution and minimize logistical costs through demand correlation analysis [1]. Similarly, hierarchical clustering has been effectively utilized to optimize spare part provision and maintenance scheduling, demonstrating its ability to integrate operational dependencies and cost-efficiency [2, 3]. Selective inventory classification techniques, such as combining K-Means clustering with multi-criteria decision-making methods, also have been used to enhance inventory accuracy and streamline operations by considering multiple attributes like cost, lead time, and demand [4]. Furthermore, modern inventory management solutions often integrate information systems to classify and prioritize warehouse stocks, improving operational efficiency and decision-making processes [5]. Advanced machine learning techniques, such as convolutional neural networks (CNNs), have also been successfully applied to improve inventory accuracy and operational efficiency, offering transformative solutions to traditional inventory challenges [6].

In addition to clustering, predictive maintenance strategies have also been explored to enhance operational reliability. For instance, Bi-LSTM models have been used to predict the remaining useful life (RUL) of warehouse equipment based on stock in-and-out operations, enabling proactive maintenance scheduling and reducing downtime [7]. These advanced methods highlight the growing role of machine learning and deep learning models in warehouse and inventory management systems.

This research proposes the use of clustering analysis, specifically the k-Means clustering method, to address these challenges [8]. By leveraging historical data on spare parts, the study identifies distinct clusters based on key attributes such as demand frequency, lead time, quantity, and cost. These clusters are used to prioritize which spare parts should be stored, enabling a more strategic allocation of resources. The k-Means algorithm offers an effective way to uncover patterns in the data, providing actionable insights into inventory optimization [9, 10].

Through this study, PLNSC gains a robust framework for improving its spare part inventory management. By focusing on critical spare parts that balance cost efficiency and operational readiness, the clustering model reduces downtime risks and storage costs. The findings not only address PLNSC's immediate inventory challenges but also offer a scalable solution applicable to other sectors facing similar issues. This research emphasizes the transformative potential of data-driven approaches in enhancing supply chain efficiency and operational reliability in the energy sector.

2. Objective

The primary objectives of this research are:

- To develop a data-driven clustering model that can effectively classify spare parts based on key attributes such as demand frequency, lead time, cost, and quantity.
- To identify and prioritize spare parts that should be stored in PLN Suku Cadang warehouses, ensuring optimal inventory management.
- To minimize the risks of operational downtime by ensuring the availability of critical spare parts while addressing storage constraints.
- To reduce storage and procurement costs by optimizing the allocation of inventory resources.
- To establish a scalable and adaptable framework for inventory management that can be applied across other units or sectors in the power generation industry.

  These objectives aim to enhance the overall efficiency and reliability of PLN Suku Cadang's inventory management while supporting uninterrupted power plant operations.

2.1. METHODOLOGY

This study employs a systematic and data-driven approach to address the challenges of spare part inventory management in PLN Suku Cadang. The methodology comprises the following steps:

1. Problem Identification
   The research begins with identifying the challenges faced by PLN Suku Cadang in managing spare part inventories, including limited storage facilities, delays in delivery, and risks of operational downtime. A clear understanding of these issues establishes the foundation for designing an effective solution.
2. Literature Review
   A thorough review of existing studies on clustering methods and inventory management is conducted to understand the state-of-the-art techniques and their applications. This step provides insights into the selection of appropriate algorithms and methodologies.
3. Data Collection and Integration
   Historical data on spare parts is collected from PLN Suku Cadang, including:
   - Inventory Data: Item codes, descriptions, specifications, and classifications.
   - Procurement Data: Purchase orders, vendor details, costs, order dates, and delivery dates.
   - Shipment Data: Requested delivery dates, promised dates, and actual receipt dates.

     The collected data is aggregated to create a comprehensive dataset for analysis.
4. Data Preprocessing
   Data cleaning and preprocessing are performed to handle missing values and outliers. Missing values are addressed using mean or mode imputation methods. Features such as cost, demand frequency, and lead time are normalized to ensure compatibility with clustering algorithms.
5. Clustering Analysis
   The k-Means clustering algorithm was selected for this study due to its balance between simplicity, computational efficiency, and interpretability. Spare parts data in this research primarily consists of numeric attributes such as demand frequency, lead time, quantity, and cost, which align well with k-Means' requirements for continuous, well-structured data.
   Despite its sensitivity to outliers and assumptions of spherical cluster shapes, k-Means was deemed appropriate for this context because:

(a) Preprocessing Mitigates Outliers: Data preprocessing steps, including normalization and outlier handling, were implemented to ensure the dataset adhered to k-Means' assumptions.
(b) Operational Applicability: The clustering results needed to be easily interpretable for practical use by PLN Suku Cadang's operational teams. k-Means' use of centroids simplifies the interpretation of cluster characteristics.
(c) Scalability: Given the large volume of historical spare parts data, k-Means offers a computationally efficient solution that can scale to future inventory datasets.

The Elbow Method was used to determine the optimal number of clusters, ensuring that the model balances within-cluster variation and operational feasibility.

6. Cluster Evaluation and Interpretation
   The resulting clusters are evaluated to ensure their validity and relevance to operational needs. Each cluster is analyzed to identify its characteristics, such as high-demand spare parts with low lead time and cost, to prioritize for storage.
7. Model Validation
   To validate the clustering results, the study incorporated a comprehensive approach that involved cross-referencing the clusters with historical operational data from PLN Suku Cadang. Historical downtime records were analyzed to evaluate whether the spare parts in high-priority clusters were linked to operational disruptions caused by spare parts unavailability. Additionally, procurement and inventory patterns were compared to assess the alignment of the clustering results with real-world operational priorities. For example, spare parts grouped into high-demand clusters were expected to match parts frequently procured or consistently maintained in inventory. Feedback from operational experts further substantiated the practical relevance of the clustering results, ensuring that the identified priorities align with their experience in maintaining operational continuity. This multi validation approach ensured that the clustering model is robust and practically applicable for optimizing inventory management.
8. Implementation Framework
   Based on the clustering results, an inventory prioritization framework is developed to guide the storage and procurement decisions of PLN Suku Cadang.

9. Conclusion and Recommendations
   The study concludes with actionable recommendations for optimizing inventory management and future applications of the proposed methodology.
   By following these steps, the research provides a robust and scalable solution for enhancing inventory efficiency and ensuring the operational reliability of PLN Suku Cadang.

3. RESULTS

The initial clustering analysis categorized spare parts into four distinct clusters based on critical operational criteria, including demand frequency, lead time, cost, and quantity. Among these clusters, Cluster 2 was identified as the most promising group for storage, meeting three out of five key criteria: high demand frequency, low lead time, and low cost.

The choice of k-Means clustering was influenced by its ability to effectively partition numeric data into distinct clusters that are easy to interpret. The clustering results demonstrated well-separated centroids, which facilitated the identification of actionable spare part groups. Key reasons for choosing k-Means include:

1. Alignment with Data Characteristics: Spare parts data consists of continuous numeric variables that are well-suited for k-Means' assumptions.
2. Operational Interpretability: Each cluster centroid represents a central tendency of spare part attributes, providing intuitive insights for inventory prioritization.
3. Efficiency and Scalability: The algorithm processed large datasets quickly and reliably, ensuring feasibility for future scaling.

The preprocessing steps, including outlier handling and normalization, mitigated the potential weaknesses of k-Means, such as sensitivity to extreme values. As a result, the algorithm produced clusters that aligned with PLN Suku Cadang's operational needs.

However, Cluster 2 contained a total of 682 types of spare parts, which was deemed too extensive for practical implementation. This large cluster required further refinement to focus on the most critical and manageable spare parts for storage.

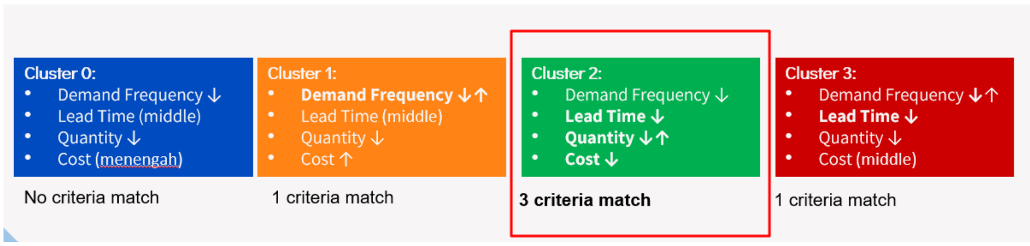

Figure 1: Initial Cluster Result

To address this challenge, a second-level clustering process was implemented using the same k-Means clustering methodology. The aim was to segment Cluster 2 into smaller, more actionable sub-clusters by reapplying the criteria of demand frequency, lead time, and cost. This refined approach ensured that the selected spare parts align closely with operational needs and resource constraints. The secondary clustering provided a more granular view of the data, enabling better prioritization for inventory storage.

From the second-level clustering, As seen in Figure 2, Cluster 6 emerged as the most suitable sub-cluster, meeting three critical criteria: high demand frequency, low lead time, and low cost. Cluster 6 consisted of a smaller, more focused set of spare parts, making it a practical and strategic choice for immediate storage. These spare parts were identified as critical for maintaining operational reliability and minimizing procurement delays, thus addressing key challenges faced by PLN Suku Cadang.

The clustering results were validated against historical operational data, yielding the following insights:

1. Downtime Correlation: Clusters identified as high-priority (Cluster 6) showed a significant correlation with historical downtime events. Parts in this cluster were frequently associated with maintenance delays due to procurement challenges, highlighting the practical relevance of the clustering model.
2. Inventory Patterns: Spare parts in high-priority clusters matched inventory records showing consistent procurement patterns and usage rates. This alignment validated the clustering model's ability to capture operational needs effectively.
3. Expert Feedback: Feedback from operational experts confirmed that the clustering results were consistent with their practical experience, particularly in identifying parts critical to maintaining operational con-

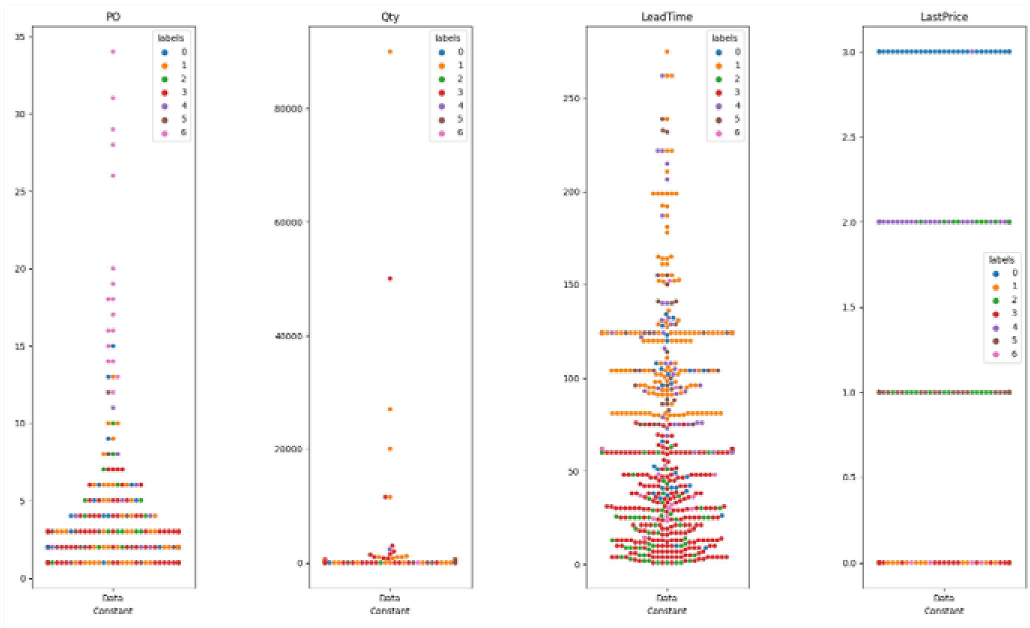

Figure 2: Swarm Plot Second Level Cluster. This visualization illustrates the distribution of spare parts across refined sub-clusters. Each point represents a spare part, color-coded by cluster ID. The distribution highlights critical clusters, such as Cluster 6, characterized by low lead times and high demand frequencies.

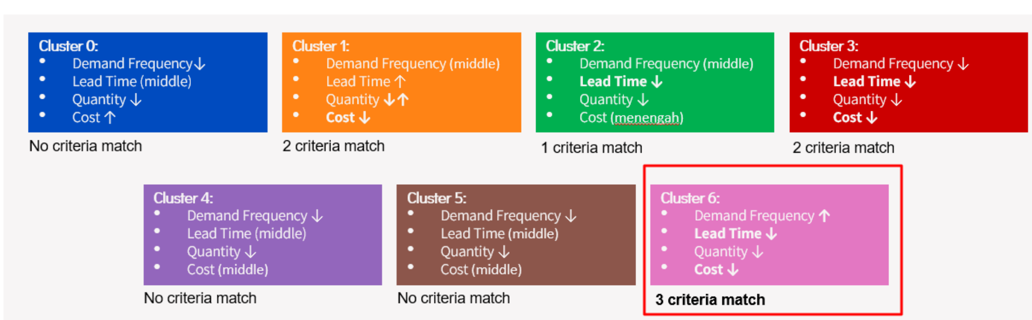

Figure 3: Second Level Cluster Result.

tinuity.

Table 1: Correlation Between Clusters and Downtime

Cluster ID	Types of Spare Parts	Correlation with Downtime Events	Average Procurement Frequency (per year)	Average Lead Time (days)
0	51	Low	7	90
1	267	Moderate	3	139
2	46	High	7	32
3	234	Low	5	43
4	28	Low	3	190
5	40	Moderate	2	102
6	16	High	18	31

The analysis of other sub-clusters also provided valuable insights. For instance, some clusters contained items with lower demand frequency or higher lead times, making them less critical for storage but potentially suitable for just-in-time procurement strategies. These findings highlight the flexibility of the clustering approach in addressing a range of operational scenarios and inventory challenges.

In summary, the transition from initial clustering to a second-level analysis proved highly effective in narrowing down the scope of prioritized spare parts. The multi-level clustering approach not only optimized the selection process but also provided actionable insights for improving inventory management strategies. This iterative analysis enabled PLN Suku Cadang to focus its resources on the most impactful spare parts while maintaining a balance between cost efficiency and operational readiness.

4. IMPLEMENTATION AND CHALLENGES

The practical implementation of the clustering model developed in this study requires integration into PLN Suku Cadang's existing operational framework. This section outlines the key considerations and potential challenges associated with applying the model in a real-world setting.

A significant step in implementation involves integrating the clustering model into PLN Suku Cadang's Enterprise Resource Planning (ERP) or inventory management systems. Such integration would enable automated prioritization of spare parts based on the clusters identified in this study. However, challenges may arise due to limitations in existing legacy systems, which may lack the capability to support advanced analytical models. Modernizing IT infrastructure and ensuring seamless data exchange between the clustering model and operational databases are essential prerequisites for successful implementation.

Another critical aspect is the reliance on high-quality and up-to-date data. The clustering model requires consistent updates to reflect changes in spare part demand, procurement patterns, and operational needs. Data preprocessing and normalization, which were pivotal in this study, should be integrated into routine workflows. However, issues such as incomplete data entries or delays in updating operational records could impact the model's performance. Establishing robust data governance frameworks and improving data collection processes will be necessary to mitigate these risks.

Operational adoption of the model is another area that warrants attention. For the clustering outputs to inform decision-making effectively, operational staff need to be trained to interpret and utilize the insights. Providing clear visualizations, such as dashboards summarizing cluster characteristics and priorities, could facilitate this process. Nevertheless, resistance to change and a lack of technical expertise among personnel may pose barriers. Comprehensive training programs and workshops will be required to address these challenges and ensure smooth adoption.

The scalability and maintenance of the clustering model are also important considerations. As PLN Suku Cadang's inventory grows, the model must adapt to accommodate larger datasets and evolving operational priorities. Regular retraining of the clustering algorithm with new data is essential to maintain its relevance. Ensuring the scalability of the model's computational framework will be critical to supporting future expansions.

Lastly, the insights generated by the clustering model can be leveraged to enhance collaboration with vendors. Sharing prioritized spare part requirements with suppliers could streamline procurement processes and reduce lead times for high-priority items. However, aligning vendor operations with the prioritization framework may require renegotiation of procurement contracts and agreements.

Despite these challenges, the proposed clustering model offers a robust foundation for optimizing inventory management at PLN Suku Cadang. By addressing these operational considerations, the model can contribute significantly to reducing downtime risks, improving cost efficiency, and supporting long-term operational resilience.

5. CONCLUSION

This study demonstrates the effectiveness of a multi-level clustering approach in optimizing spare part inventory management at PLN Suku Cadang. By using the k-Means algorithm, the research successfully identified and prioritized spare parts that are critical to operational reliability. The refinement of Cluster 2 into sub-clusters, with Cluster 6 emerging as the most optimal group, underscores the value of iterative data-driven analysis in addressing complex inventory challenges.

The prioritization of Cluster 6 highlights the importance of focusing on spare parts with high demand frequency, low lead time, and low cost. This targeted approach ensures that critical spare parts are readily available, reducing the risks of operational downtime while maintaining cost efficiency. Moreover, the insights from other clusters provide PLN Suku Cadang with a framework for balancing on-demand procurement and storage strategies based on spare part characteristics.

The clustering results were validated through a comparison with historical operational data and expert reviews. High-priority clusters, such as Cluster 6, demonstrated strong alignment with spare parts historically linked to operational downtime events. Additionally, the clustering outcomes were consistent with procurement and inventory records, reflecting actual operational priorities. Expert feedback further reinforced the practical relevance of the clustering model, highlighting its ability to address critical inventory challenges effectively.

Overall, the findings of this research provide a robust and scalable solution for spare part inventory management. The clustering framework offers PLN Suku Cadang a data-driven tool to improve operational efficiency, optimize storage resources, and support long-term supply chain resilience in the power generation industry.

The use of k-Means clustering in this study aligns with the operational needs and data characteristics of PLN Suku Cadang. Despite its sensitivity to outliers and assumptions of spherical clusters, k-Means was chosen due to its simplicity, computational efficiency, and ease of interpretation. The preprocessing steps effectively addressed potential issues with outliers and data normalization, ensuring that the clusters produced were robust and actionable. While alternative methods like DBSCAN or GMM could offer additional flexibility in handling irregular cluster shapes or outliers, the results demonstrated that k-Means met the study's objectives effectively, balancing accuracy and practicality.

Future research could explore the application of other clustering methods to handle more complex data scenarios, such as datasets with extreme variability or irregular patterns. However, the findings of this study validate the effectiveness of k-Means in optimizing spare parts inventory management for PLN Suku Cadang.

References

[1] E. C. Griffin, B. B. Keskin, A. W. Allaway, Clustering retail stores for inventory transshipment, European Journal of Operational Research 311 (2023) 690–707. doi:10.1016/j.ejor.2023.06.008.

[2] M. Zhu, X. Zhou, Hierarchical-clustering-based joint optimization of spare part provision and maintenance scheduling for serial-parallel multi-station manufacturing systems, International Journal of Production Economics 264 (2023) 108971. doi:10.1016/j.ijpe.2023.108971.

[3] A. Abouelrous, A. F. Gabor, Y. Zhang, Optimizing the inventory and fulfillment of an omnichannel retailer: a stochastic approach with scenario clustering, Computers & Industrial Engineering 173 (2022) 108723. doi:10.1016/j.cie.2022.108723.

[4] Y. Khanorkar, P. Kane, Selective inventory classification using abc classification, multi-criteria decision making techniques, and machine learning techniques, Materials Today: Proceedings 72 (2023) 1270–1274. doi:10.1016/j.matpr.2022.09.298.

[5] S. Yakovlev, V. Sorokin, E. Alexandrova, D. Syromolotov, I. Gorbachenko, Development of an information system for the classification of warehouse stocks, Transportation Research Procedia 68 (2023) 475–482. doi:10.1016/j.trpro.2023.02.064.

[6] W. Villegas-Ch, A. M. Navarro, S. Sanchez-Viteri, Optimization of inventory management through computer vision and machine learning technologies, Intelligent Systems with Applications 24 (2024) 200438. doi:10.1016/j.iswa.2024.200438.

[7] G. Shang, L. Xu, C. Lu, B. Zhang, Warehouse equipment maintenance strategy based on the prediction of total amount of

stock in & out operations, Procedia CIRP 120 (2023) 237–242. doi:10.1016/j.procir.2023.08.042.

[8] H. Abbasimehr, A. Bahrini, An analytical framework based on the recency, frequency, and monetary model and time series clustering techniques for dynamic segmentation, Expert Systems with Applications 192 (2022) 116373.

[9] D. Das, P. Kayal, M. Maiti, K-means clustering model for analyzing the bitcoin extreme value returns, Decision Analytics Journal 6 (2023) 100152. doi:10.1016/j.dajour.2022.100152.
URL `https://doi.org/10.1016/j.dajour.2022.100152`

[10] S. S. Indasari, A. Tjahyanto, Decision support model in compiling owner estimate for fmcgs products from various marketplaces with tf-idf and lsa-based clustering, Procedia Computer Science 234 (2024) 455–462. doi:10.1016/j.procs.2024.03.027.