Jurnal Penelitian dan Aplikasi Sistem dan Teknik Industri (PASTI)
Vol. XVIII, No. 3, Desember 2024, 405-416
p-ISSN 2085-5869/ e-ISSN 2598-4853

Structured Defect Data-Based K-Means Clustering Analysis and Framework for Quality Control (QC) Prioritization in Manufacturing

Jakfat Haekal[1*] and Rizaldi Mu'min[2]
[1)] Departement of Industrial Engineering, Engineering Faculty, Universitas Bhayangkara Jakarta Raya, Jakarta
[2)] Departement of Management, Faculty of Economics, Universitas Negeri Jakarta, Jakarta
Email: jakfat.haekal@dsn.ubharajaya.ac.id

Abstrak

Quality Control (QC) sangat penting dalam manufaktur untuk memastikan kualitas produk dan meminimalisir cacat produk. Namun, meningkatnya kompleksitas produk dan proses manufaktur telah membuat identifikasi dan prioritas cacat untuk QC menjadi lebih menantang, sementara sebagian besar studi hanya berfokus pada inspeksi visual. Oleh karena itu, studi ini mengusulkan analisis berbasis data cacat terstruktur dan kerangka kerja untuk menemukan prioritas dalam proses QC. Kerangka kerja ini menggunakan pengelompokan K-Means untuk mengelompokkan cacat berdasarkan karakteristiknya, seperti jenis, lokasi, dan tingkat keparahan yang memengaruhi tingkat biaya perbaikan. Untuk memvalidasi model, *Davis-Bouldin Index* dan *Silhouette Score* digunakan untuk mengukur kualitas model. Eksplorasi data menunjukkan bahwa setiap fitur memiliki hubungan serta dampak terhadap biaya perbaikan di mana tingkat keparahan yang lebih besar sejalan dengan biaya perbaikan yang lebih tinggi. Penemuan menunjukkan bahwa *cluster 0* adalah yang harus diprioritaskan karena memiliki biaya perbaikan tertinggi di antara yang lain. Hasil penelitian menunjukkan bahwa kerangka kerja tersebut dapat secara efektif mengidentifikasi dan memprioritaskan cacat untuk QC, yang berpotensi mengarah pada peningkatan kualitas produk dan pengurangan biaya manufaktur.

Kata kunci: Cacat produk; Quality Control (QC); K-Means Clustering; Manufaktur; Prioritas

Abstract

Quality control (QC) is crucial in manufacturing to ensure product quality and minimize defects. However, the increasing complexity of products and manufacturing processes has made it more challenging to identify and prioritize defects for QC, while most studies focus only on visual inspection. Therefore, this paper proposes a structured defect data-based analysis and framework for QC prioritization. The framework uses K-Means clustering to group defects based on their characteristics, such as type, location, and severity, which impact the repair cost rate. To validate the model, the Davis-Bouldin Index and Silhouette Score are used to measure model quality. Data exploration shows that each feature has a relationship with and impacts repair costs, where greater severity aligns with higher repair costs. The findings indicate cluster 0 as the main priority due to highest repair cost among others. The results show that the framework can effectively identify and prioritize defects for QC, potentially leading to improved product quality and reduced manufacturing costs.

Keywords: *Defects; Quality Control; K-Means Clustering; Manufacturing; Prioritization*

INTRODUCTION

The manufacturing sector is pivotal in advancing Sustainable Development Goals (SDGs) by producing high-quality, complex products at lower costs while minimizing resource waste (Stershic et al., 2021; Colledani et al., 2014). However, manufacturing defects remain a significant challenge, directly affecting product performance, costs, and customer perception. Despite advancements in design, material selection, and manufacturing processes, defects persist in production systems (Sreedharan et al., 2018). For developed countries competing with lower-cost producers, addressing these challenges is crucial. Manufacturers are increasingly focusing on productivity and quality enhancement through defect analysis and prevention to stay competitive (Elmekkawy et al., 2006).

Defects negatively affect the manufacturing process (Sanaei & Fatemi, 2021) and impact product quality (Psarommatis et al., 2020). Its analysis is essential for enhancing quality and productivity. Various methods, such as simulations, historical and Pareto analysis, cause-effect diagrams, and neural networks, help identify process improvements (Kumar et al., 2016; Casey et al., 1991). A comprehensive approach classifies defects by factors like size and location, then analyzes potential causes related to design, material, and process parameters (Mane et al., 2011). Integrated computer modules support small industries by combining expertise for defect identification and correction (Mehta et al., 2020). Simulation analysis optimizes parameters to reduce defects, especially in foundries with manual processes and unskilled labor (Mane et al., 2011).

K-means clustering has proven effective in detecting manufacturing defects across various industries. In semiconductor wafer production, it can segment defects based on color features, enhancing efficiency compared to manual inspection (Nor Hidayah Saad et al., 2015). In biomass particle production, K-means, combined with prior knowledge, accurately detects issues such as poor roundness and cracks, achieving high accuracy and speed (Wei Wang & Gong, 2020). Although K-means itself is not directly used, a related K-Nearest Neighbor algorithm has been employed to predict defect rates, with the Minkowski distance method reaching 86.41% accuracy under optimal conditions (Muhammad David & Muhammad Azka Firdaus, 2024). These studies underscore the adaptability of K-means and similar algorithms in improving quality control and production efficiency across various industries.

The primary issue addressed in this research pertains to defects within the manufacturing industry, which remain a critical challenge in ensuring product quality. This study highlights that defect analysis should not be limited to visual inspections alone. By leveraging structured data for further analysis, industries can significantly improve their quality control (QC) processes. Through the application of clustering techniques, this research provides a practical framework to assist managers in prioritizing defects, thereby enhancing the overall efficiency of defect management and quality assurance. The novelty of this research lies in its innovative approach to integrating structured defect data with K-Means clustering to enhance quality control (QC) prioritization in the manufacturing industry. Traditional QC methods often rely on visual inspections and reactive strategies, which can be inefficient and fail to uncover underlying defect patterns. Traditional quality control methods often rely on reactive strategies and visual inspections, which can be inefficient and fail to uncover underlying defect patterns (Desai & Mital, 2009; Watts, 2011). These approaches are costly and do not guarantee defect-free products, especially for items with multiple quality characteristics (Hussein & Diab, 2010). To address these limitations, there has been a shift towards proactive quality management strategies that incorporate design techniques to eliminate the need for inspection (Desai & Mital, 2009). This study introduces a data-driven framework that systematically analyzes defect data to identify patterns and clusters based on factors such as severity, frequency, and impact. This structured methodology provides a fresh perspective on addressing manufacturing defects by enabling proactive and informed decision-making.

Research on defect detection identifies various attributes, including defect type (Monje et al., 2019), location (Wang & Cheng, 2019), and severity (Psarommatis et al., 2020). Defects are classified as cosmetic (e.g., scratches) (Zhang et al., 2021), structural (e.g., cracks) (Brennan et al., 2021), or functional (affecting performance) (Fu et al., 2022). Location determines whether defects appear on surfaces or components (Brennan et al., 2021), while severity influences repair costs (Powell et al., 2022). Clustering these attributes helps prioritize defect management based on criticality and cost.

The findings of this research highlight the significant advantages of utilizing structured defect data for clustering analysis. By applying K-Means clustering, the study reveals hidden trends and patterns in defect occurrences, offering deeper insights that are not readily apparent through conventional QC practices. Additionally, the framework prioritizes critical defect clusters, allowing managers to allocate resources more effectively to address the most impactful issues. Data-driven approaches have shown promise in identifying root causes of defects. Chen et al. (2004) as a result, the proposed methodology not only streamlines inspection processes but also minimizes waste and improves overall product quality, showcasing its potential to revolutionize QC practices in the manufacturing sector.

METHODS

1. Data Collection and Preprocessing

The primary objective of this step is to prepare manufacturing defect data for clustering analysis. Data collection involves gathering defect-related information, including categorical features such as defect type, defect location, and severity.

To ensure compatibility with clustering algorithms, categorical data is transformed into numerical formats using label encoding. For example, defect types are encoded as cosmetic (0), functional (1), and structural (2). Similarly, defect locations are labeled as component (0), internal (1), and surface (2), while severity is categorized into critical (0), minor (1), and moderate (2).

After encoding, the dataset undergoes cleaning processes to address missing values, outliers, and duplicates. This ensures that the input data is accurate and reliable for subsequent clustering analysis.

2. Cluster Analysis Using K-Means

The goal of this step is to group defects into clusters based on shared characteristics to support prioritization. The first task involves defining the optimal number of clusters (“k”) using the elbow method. This technique identifies the point where adding more clusters no longer significantly reduces the Within-Cluster Sum of Squares (WCSS). K-Means will define the centroid (the center-point) randomly (Romanuke, 2023), and place based on its cluster to the nearest centroid (Figure 1), then it will count the average point (Ikotun et.al., 2023).

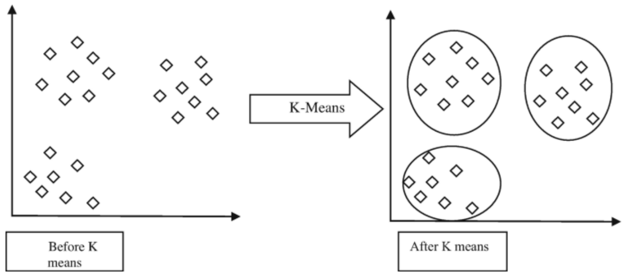


Figure 1. How K-Means Works on Data Clustering

Each data point in the dataset is assigned to the nearest cluster based on a distance metric, typically the Euclidean distance (Ikotun et.al., 2023). It calculates the straight-line distance between two points in a multidimensional space (Faisal & Zamzami, 2020), it also helps the algorithms to update the centroid. The objective function (J(z, A)) in k-Means is used to minimize the distance between data points and their cluster centres, once data points assigned to clusters, the centroid are updated (a_k). *k* greatly influences the clustering results. four methods used for selecting the value of *k:* the elbow method & silhouette score (Yuan & Yang, 2019). This study will measure the silhouette score to validate clustering quality, Once the optimal "k" is determined (e.g., k = 3), the K-Means algorithm is applied. The resulting clusters are analyzed by examining their centroids to identify key patterns such as repair costs, defect types, defect locations, and severity.

3. Dimensionality Reduction and Visualization

Dimensionality reduction techniques are essential for effective visualization of clustering results. Principal Component Analysis (PCA) is employed to reduce the number of dimensions while retaining meaningful data variance.

The PCA plots highlight group differences based on repair costs and defect characteristics, providing a clear visual representation of the clusters. This aids in interpreting the clustering results and identifying trends among the defects.

4. Cluster Interpretation and Prioritization

This step focuses on interpreting the clustering results and determining priorities for quality control.

- **Cluster 0**: This cluster is characterized by high repair costs and functional defects, predominantly located internally. Severity ranges from minor to moderate. These defects require prioritization due to their complexity and expense.
- **Cluster 1**: This group contains low-cost cosmetic defects that are primarily located on the surface. With minor severity, these defects can be managed with minimal resources.
- **Cluster 2**: Comprising moderate-cost structural defects, this cluster shows a balanced distribution of defect locations. Severity levels range from minor to moderate, warranting investigations into assembly processes.

Actionable insights include allocating skilled technicians and larger budgets for Cluster 0, minimal resources for Cluster 1, and process optimization for Cluster 2.

5. Model Validation

Model validation ensures the clustering model's performance and reliability. Internal validation involves calculating silhouette scores and the Davies-Bouldin Index to assess clustering quality and overlap between clusters.

External validation tests the model's performance on unseen datasets, using silhouette scores to confirm its generalization capability. These validation steps ensure that the model is robust and reliable for real-world applications.

6. Implementation

To integrate the clustering model into manufacturing quality control processes, the model is deployed to classify ongoing defect data into clusters. Managers can leverage visualizations and statistical summaries to make informed decisions.
Periodic retraining of the model with updated defect data is crucial to adapt to new patterns and maintain its effectiveness.

RESULT & DISCUSSION

1. Data Preprocessing

	defect_date	defect_id	product_id	defect_type	defect_location	severity	nspection_metho	repair_cost
1	6/6/2024	1	15	Structural	Component	Minor	Visual Inspe...	245.47
2	4/26/2024	2	6	Functional	Component	Minor	Visual Inspe...	26.87
3	2/15/2024	3	84	Structural	Internal	Minor	Automated T...	835.81
4	3/28/2024	4	10	Functional	Internal	Critical	Automated T...	444.47
5	4/26/2024	5	14	Cosmetic	Component	Minor	Manual Testi...	823.64
6	5/11/2024	6	17	Functional	Internal	Moderate	Visual Inspe...	788.11
7	5/23/2024	7	85	Cosmetic	Internal	Critical	Manual Testi...	33.68
8	1/15/2024	8	90	Structural	Internal	Moderate	Manual Testi...	65.56
9	1/26/2024	9	30	Structural	Component	Critical	Manual Testi...	848.61
10	6/5/2024	10	20	Structural	Component	Critical	Visual Inspe...	478.48

Figure 2. Sample Dataset of Manufacturing Defect

The manufacturing defect dataset underwent preprocessing to prepare it for K-Means clustering. This involved transforming categorical features into numerical representations. Defect type has categorical feature was converted into numerical data using label encoding: cosmetic (0), functional (1), structural (2). While defect location's categorical features was also converted using label encoding: component (0), internal (1), surface (2). Severity was converted using label encoding: critical (0), minor (1), moderate (2).

2. Defining Number of Clusters *(k)*

In doing clustering K-Means task, defining centroids is an essential part. The method that will be used for defining centroids is elbow method which is used for illustrating the relationship between the number of clusters *(k)* and Within-Cluster Sum of Squares (WCSS). Figure 3 shows the elbows graph, y-axis represents the WCSS values. This measures how far data points within a cluster are from their cluster center. While x-axis represents the number of clusters being tested. Elbow point on the graph indicates the optimal number of clusters. Beyond this point, the decrease in WCSS becomes insignificant.

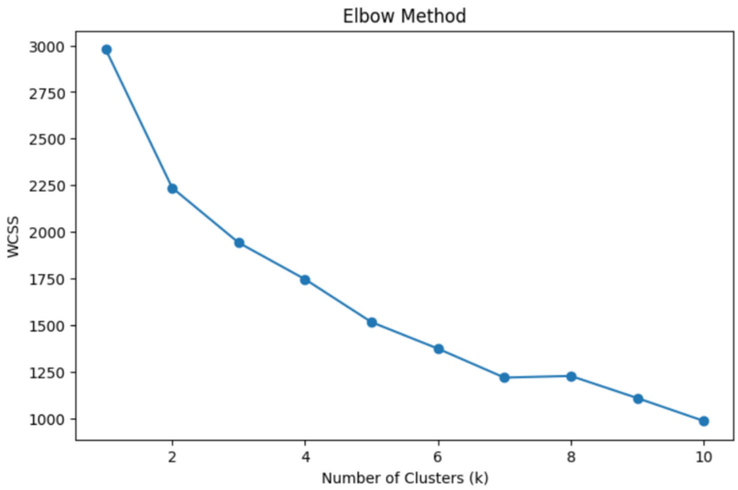


Figure 3. Elbow method to define centroid

Before the elbow, decrease in WCSS is very steep, indicating that adding more clusters significantly improves the quality of clustering. While after the elbow, decrease in WCSS slows down, suggesting that adding more clusters does not provide significant improvement. The optimal number of clusters is at the "elbow" point. In this graph, the "elbow" is observed around *k*=3, thus, the research use three clusters.

2. K-Means Clustering

Result shows that there is a positive correlation is observed between repair cost and severity (figure 4), meaning that defects with higher repair costs also tend to have higher severity levels. This relationship highlights that more severe defects often demand more resources to address. Additionally, there is a notable relationship between defect location and defect type. Despite the general patterns observed within each cluster, there is still significant variation among the defects within the same cluster.

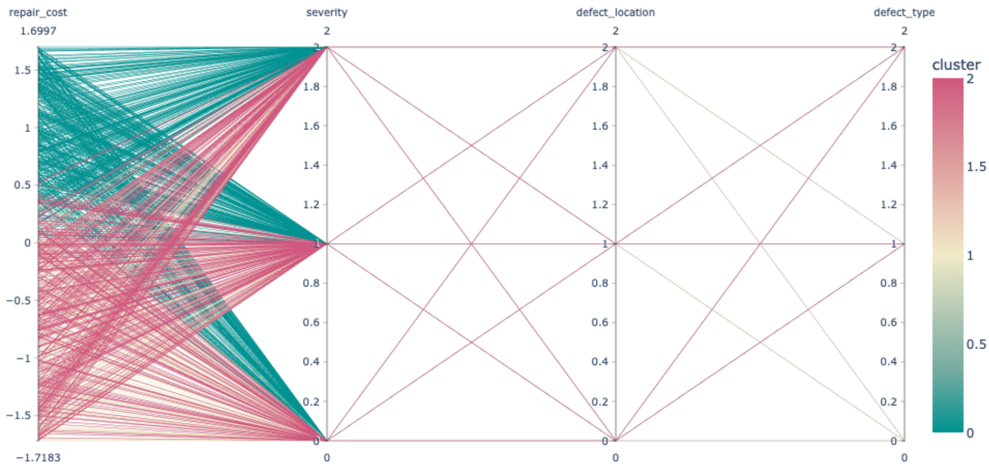
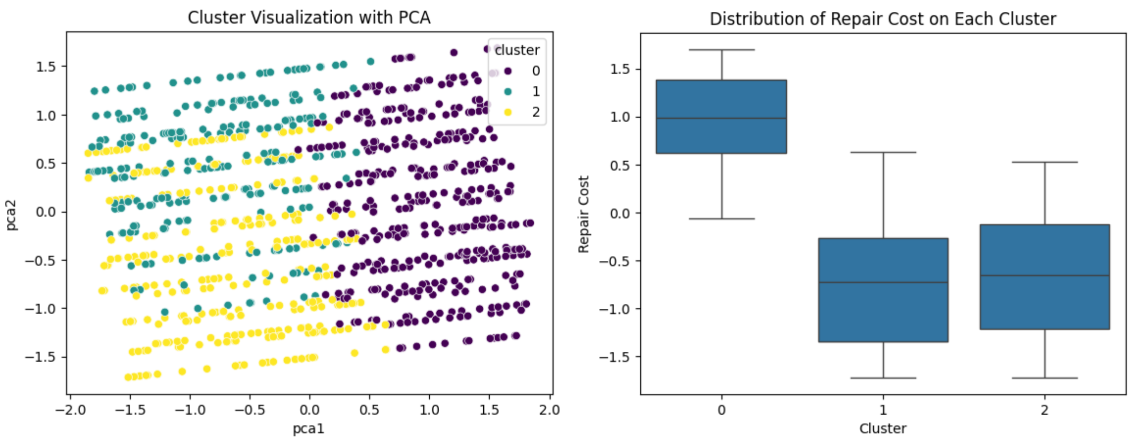


Figure 4. Parallel Coordinates Plot Describing Data Spread and Variability Each Cluster

Cluster 2 (red) tends to have lower repair costs and severity, suggesting that defects within this cluster are generally easier and more affordable to fix. In contrast, cluster 0 (green) is characterized by higher repair costs and severity, indicating that defects in this cluster are likely more complex and require more expensive repairs. Cluster 1 (yellow) has characteristics that fall between cluster 0 and cluster 2, representing a middle ground in terms of repair costs and defect severity. Variability between defects in clustering is common, cluster 2, while most defects are easier and cheaper to fix, some may still present unique challenges. It may contain defects with various levels of complexity and cost. PCA (Principal Component Analysis) is used in clustering visualization as a dimensionality reduction technique where the details of each data point will be identified to maintain the accuracy and objectivity of clustering. Figure 6a shows that cluster 0 is grouped in the upper part of the PCA graph which indicates higher repair costs, as it tends to be separated from other clusters that may have lower repair costs.



(a) Cluster visualization with PCA (b) Distribution of Repair Cost Each Cluster

Figure 5. Principal Component Analysis (PCA) and Boxplot Defines Cluster with Highest Repair Cost

Cluster 1 is positioned in the lower-left, indicated lower repair costs, the same with cluster 2 which is located in the lower-right of the PCA graph (Figure 5a). This is supported by the boxplot (Figure 5b) that shows the difference each cluster where Cluster 0 has the highest median repair cost, supporting our earlier assumption based on the PCA graph that this cluster is dominated by defects with higher repair costs. Clusters 1 and 2 have lower and relatively similar median repair costs. However, Cluster 1 exhibits a wider data distribution, indicating greater variability in repair costs within this cluster. Table 1 shows that cluster 0 is dominated by functional defects (average defect_type 959.620), mostly located internally (average defect_location 942.993), with a severity leaning towards minor but including some moderate cases (average severity 1.023.753). Cluster 1 mostly contains cosmetic defects (average defect_type 338.462), found predominantly on the surface (average defect_location 1.380.769), with minor severity (average severity 807.692). Cluster 2 is dominated by structural defects (average defect_type 1.727.273), occurring across all locations but slightly favoring internal areas (average defect_location 849.530), with a severity leaning towards minor but including some moderate cases (average severity 1.050.157).

Table 1. Average feature each cluster

Cluster	defect_type	defect_location	severity
0	959.620	942.993	1.023.753
1	338.462	1.380.769	807.692
2	1.727.273	849.530	1.050.157

Cluster 0, comprising 421 data points, is characterized by the highest average repair cost (mean 0.97, median 0.98), indicating expensive repairs. The defect type in this cluster is predominantly "functional", while defects are mostly located internally and have a severity level categorized as "minor". Cluster 1, with the smallest count of 260 data points, exhibits the lowest average repair cost (mean -0.77, median -0.72), suggesting inexpensive repairs. Defects in this cluster are primarily "cosmetic" and predominantly located on the surface, with a severity level also classified as "minor." Cluster 2, the largest cluster with 319 data points, displays relatively low repair costs (mean -0.66, median -0.64). The defect type in this cluster is primarily "structural," with defect locations more evenly distributed but slightly skewed toward "component" locations. The severity in this cluster is classified as "minor" but trends closer to "moderate."

Table 2. Descriptive Statistics each Cluster

k	**Fitur**	**count**	**mean**	**std**	**min**	**25%**	**50%**	**75%**	**max**
0	defect_type	421.0	959.620	781.198	0.0	0.0	1.0	2.0	2.0
0	defect_location	421.0	942.993	817.417	0.0	0.0	1.0	2.0	2.0
0	severity	421.0	1.023.753	801.431	0.0	0.0	1.0	2.0	2.0
0	repair_cost	421.0	979.471	453.846	-56.471	616.255	984.33	1.385.1	1.699
1	defect_type	260.0	338.462	474.099	0.0	0.0	0.0	1.0	1.0
1	defect_location	260.0	1.380.7	728.101	0.0	1.0	2.0	2.0	2.0
1	severity	260.0	807.692	801.506	0.0	0.0	1.0	1.0	2.0
1	repair_cost	260.0	-772.40	606.535	-1.718.285	-1.343.1	-721.02	-264.98	628.0
2	defect_type	319.0	1.727.273	446.061	1.0	1.0	2.0	2.0	2.0
2	defect_location	319.0	849.530	821.828	0.0	0.0	1.0	2.0	2.0
2	severity	319.0	1.050.1	783.472	0.0	0.0	1.0	2.0	2.0
2	repair_cost	319.0	-663.11	611.548	-1.717.525	-1.214.1	-646.01	-124.83	526.5

From this observation, Cluster 0 stands out with significantly higher repair costs, while Clusters 1 and 2 have relatively lower repair costs. In terms of defect types, Cluster 0 is dominated by "functional" defects, Cluster 1 by "cosmetic" defects, and Cluster 2 by "structural" defects. Defect locations also vary, with Cluster 0's defects predominantly "internal," Cluster 1's defects more prevalent on the "surface," and Cluster 2 showing a relatively balanced distribution with a slight skew toward "component" defects. Although all clusters are largely associated with "minor" severity, Cluster 2 leans closer to "moderate." Cluster 0 requires prioritization for repairs due to its high costs and the need for skilled technicians and larger budgets. For Cluster 1, less specialized resources can be allocated to manage inexpensive "cosmetic" defects on the surface. Cluster 2 warrants further investigation into assembly processes and materials to address "structural" defects.

3. Model Validation

The validation is to measure if this model can be deployed to future data, we use internal validation and external validation. External validation evaluates how well the model generalizes to new, unseen data, which consist of silhouette score and the davies-bouldin index are used to evaluate clustering K-Means algorithm (Suraya, 2024). The test set silhouette score is 0.232, indicates that the model generalizes fairly well, as it reflects adequate clustering on unseen data. Internal validation measures the quality of clustering on the training dataset. The silhouette score for the entire training data of 0.228, shows a quite good clustering performance. Additionally, the Davies-Bouldin Index of 1.213 reveals that there is noticeable overlap between clusters, as lower values typically indicate better-defined clusters. Overall, K-Means model demonstrates sufficient capability in segmenting manufacturing defect data into clusters. It generalizes well to new data, but the overlap between clusters suggests there is room for further refinement and optimization.

CONCLUSION

This study demonstrates that the defect analysis does not stop only in visual inspection, for further analysis, industries can put that on a structured data to enhance the quality control (QC). Clustering in this study can help managers to determine which defect should be prioritized. This study prove that there is positive correlation between repair costs and severity, highlighting the increased expense associated with fixing more severe defects. A relationship between defect location and defect type, indicating that certain types of defects are more likely to occur in specific locations within the product. The analysis reveals significant variability within each cluster, underscoring the diverse characteristics that defects can exhibit even within the same cluster. PCA visualization effectively distinguishes clusters based on repair costs, with one cluster incurring the highest expenses and the other clusters showing lower costs. This financial disparity is further supported by the boxplot analysis, which confirms the higher median repair cost for the high-cost cluster. This study contribute to a deeper understanding of defect patterns in manufacturing and provide a robust framework for prioritization in QC. The K-Means model successfully identifies distinct clusters based on defect characteristics, enabling targeted QC prioritization in manufacturing.

REFERENCES

Stershic, J., Igogo, T., & Carpenter, A. (2021). *Sustainability of the US Manufacturing Sector: Use of the United Nations Sustainable Development Goals* (No. NREL/TP-6A20-80732). National Renewable Energy Lab.(NREL), Golden, CO (United States).

Colledani, M., Copani, G., & Rosa, P. (2014). Zero Waste PCBs: a new integrated solution for key-metals recovery from PCBs. In *Proceedings of SUM 2014, Second Symposium on Urban Mining* (pp. 1-16).

Sreedharan V, R., Kannan S, S., P, A., & Trehan, R. (2018). Defect reduction in an electrical parts manufacturer: a case study. *The TQM Journal*, *30*(6), 650-678. ElMekkawy, T. Y., Hachkowski, P. A., Strong, D., & Mann, D. D. (2006). Defect analysis for quality and productivity improvements in a manufacturing system. *Canadian Biosystems Engineering*, *48*, 7.

Kumar, M., Singh, B., Yadav, P., Bhatt, V., Kumar, M., Singh, K., ... & Yun, J. H. (2017). Effect of structural defects, surface roughness on sensing properties of Al doped ZnO thin films deposited by chemical spray pyrolysis technique. *Ceramics International*, *43*(4), 3562-3568.

Mane, M. L., Dhage, V. N., Sundar, R., Ranganathan, K., Oak, S. M., Shengule, D. R., & Jadhav, K. M. (2011). Effects of Nd: YAG laser irradiation on structural, morphological, cation distribution and magnetic properties of nanocrystalline CoFe2O4. *Applied surface science*, *257*(20), 8511-8517.

Mehta, N., Gohil, A. V., Dave, K. G., & Patel, V. B. (2021). Development of casting defect analysis module through integrated approach for small and medium scale industries. *Materials Today: Proceedings*, *38*, 2935-2942.

Saad, N. H., Sabri, N. M., Hasan, A. F., Ali, A., & Saleh, H. M. (2015). Defect segmentation of semiconductor wafer image using k-means clustering. *Applied Mechanics and Materials*, *815*, 374-379.

Xiao, Q., Wang, Y., Zhao, Z. J., Pei, C., Chen, S., Gao, L., ... & Gong, J. (2020). Defect-mediated reactivity of Pt/TiO 2 catalysts: the different role of titanium and oxygen vacancies. *Science China Chemistry*, *63*, 1323-1330.

David, M., & Firdaus, M. (2024). Defect Rate Prediction in Manufacturing Process Using K-Nearest Neighbor Algorithm. *International Journal Of Informatics, Economics, Management And Science*, *3*(2), 161-173. doi:10.52362/ijiems.v3i2.1599

Desai, Anoop. (2009). Managing quality: The transition from reactive to proactive strategies. International Journal of Product Development - Int J Prod Dev. 8. 10.1504/IJPD.2009.023749.

Watts, D. H., Huang, S., Culnane, M., Kaiser, K. A., Scheuerle, A., Mofenson, L., Stanley, K., Newell, M. L., Mandelbrot, L., Delfraissy, J. F., & Cunningham, C. K. (2011). Birth defects among a cohort of infants born to HIV-infected women on antiretroviral medication. *Journal of perinatal medicine*, *39*(2), 163–170. https://doi.org/10.1515/jpm.2010.139

Chen, S. T., Wilson, T. G., Jr, & Hämmerle, C. H. (2004). Immediate or early placement of implants following tooth extraction: review of biologic basis, clinical procedures, and outcomes. *The International journal of oral & maxillofacial implants*, *19 Suppl*, 12–25.

Ran, X., Zhou, X., Lei, M., Tepsan, W., & Deng, W. (2021). A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots. *Applied Sciences*, *11*(23), 11202. https://doi.org/10.3390/app112311202

Breneman, J. E., Sahay, C., & Lewis, E. E. (2022). *Introduction to reliability engineering*. John Wiley & Sons.

Sanaei, N., & Fatemi, A. (2021). Defects in additive manufactured metals and their effect on fatigue performance: A state-of-the-art review. *Progress in Materials Science*, *117*, 100724.

Yang, J., Li, S., Wang, Z., Dong, H., Wang, J., & Tang, S. (2020). Using deep learning to detect defects in manufacturing: a comprehensive survey and current challenges. *Materials*, *13*(24), 5755.

Psarommatis, F., May, G., Dreyfus, P. A., & Kiritsis, D. (2020). Zero defect manufacturing: state-of-the-art review, shortcomings and future directions in research. *International journal of production research*, *58*(1), 1-17.

Shrifan, N. H., Jawad, G. N., Isa, N. A. M., & Akbar, M. F. (2020). Microwave nondestructive testing for defect detection in composites based on K-means clustering algorithm. *IEEE Access*, *9*, 4820-4828.

Banerjee, S., Chaudhuri, A., & Chakrabarty, K. (2020). Analysis of the impact of process variations and manufacturing defects on the performance of carbon-nanotube FETs. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, *28*(6), 1513-1526.

Bhatt, P. M., Malhan, R. K., Rajendran, P., Shah, B. C., Thakar, S., Yoon, Y. J., & Gupta, S. K. (2021). Image-based surface defect detection using deep learning: A review. *Journal of Computing and Information Science in Engineering*, *21*(4), 040801.

Le, X., Mei, J., Zhang, H., Zhou, B., & Xi, J. (2020). A learning-based approach for surface defect detection using small image datasets. *Neurocomputing*, *408*, 112-120.

Ullah, I., Khan, R. U., Yang, F., & Wuttisittikulkij, L. (2020). Deep learning image-based defect detection in high voltage electrical equipment. *Energies*, *13*(2), 392.

Zhao, Y., & Chen, J. (2022). A survey on differential privacy for unstructured data content. *ACM Computing Surveys (CSUR)*, *54*(10s), 1-28.

Adnan, K., & Akbar, R. (2019). An analytical study of information extraction from unstructured and multidimensional big data. *Journal of Big Data*, *6*(1), 1-38.

Imoto, K., Nakai, T., Ike, T., Haruki, K., & Sato, Y. (2019). A CNN-based transfer learning method for defect classification in semiconductor manufacturing. *IEEE Transactions on Semiconductor Manufacturing*, *32*(4), 455-459.

Czimmermann, T., Ciuti, G., Milazzo, M., Chiurazzi, M., Roccella, S., Oddo, C. M., & Dario, P. (2020). Visual-based defect detection and classification approaches for industrial applications—a survey. *Sensors*, *20*(5), 1459.

Di, H., Ke, X., Peng, Z., & Dongdong, Z. (2019). Surface defect classification of steels with a new semi-supervised learning method. *Optics and Lasers in Engineering*, *117*, 40-48.

Monje, A., Pons, R., Insua, A., Nart, J., Wang, H. L., & Schwarz, F. (2019). Morphology and severity of peri-implantitis bone defects. *Clinical Implant Dentistry and Related Research*, *21*(4), 635-643.

Wang, M., & Cheng, J. C. (2020). A unified convolutional neural network integrated with conditional random field for pipe defect segmentation. *Computer-Aided Civil and Infrastructure Engineering*, *35*(2), 162-177.

Zhang, H., Jiang, L., & Li, C. (2021). CS-ResNet: Cost-sensitive residual convolutional neural network for PCB cosmetic defect detection. *Expert Systems with Applications*, *185*, 115673.

Brennan, M. C., Keist, J. S., & Palmer, T. A. (2021). Defects in metal additive manufacturing processes.

Fu, Y., Downey, A. R., Yuan, L., Zhang, T., Pratt, A., & Balogun, Y. (2022). Machine learning algorithms for defect detection in metal laser-based additive manufacturing: A review. *Journal of Manufacturing Processes*, *75*, 693-710.

Goyal, A., Agrawal, R., & Saha, C. R. (2019). Quality management for sustainable manufacturing: Moving from number to impact of defects. *Journal of Cleaner Production*, *241*, 118348.

Powell, D., Magnanini, M. C., Colledani, M., & Myklebust, O. (2022). Advancing zero defect manufacturing: A state-of-the-art perspective and future research directions. *Computers in Industry*, *136*, 103596.

Mittal, H., Pandey, A. C., Saraswat, M., Kumar, S., Pal, R., & Modwel, G. (2022). A comprehensive survey of image segmentation: clustering methods, performance parameters, and benchmark datasets. *Multimedia Tools and Applications*, 1-26.

Ruspini, E. H., Bezdek, J. C., & Keller, J. M. (2019). Fuzzy clustering: A historical perspective. *IEEE Computational Intelligence Magazine*, *14*(1), 45-55.

Serban, C., & Vescan, A. (2019, August). Predicting reliability by severity and priority of defects. In *Proceedings of the 2nd ACM SIGSOFT International Workshop on Software Qualities and their Dependencies* (pp. 27-34).

Akhmedov, B. A. (2021). Innovative cluster model for improving the quality of education. *Academic Research in Educational Sciences*, *2*(3), 528-534.

Somu, N., MR, G. R., & Ramamritham, K. (2021). A deep learning framework for building energy consumption forecast. *Renewable and Sustainable Energy Reviews*, *137*, 110591.

Ikotun, A. M., Ezugwu, A. E., Abualigah, L., Abuhaija, B., & Heming, J. (2023). K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data. *Information Sciences*, *622*, 178-210.

Pérez-Ortega, J., Almanza-Ortega, N. N., Vega-Villalobos, A., Pazos-Rangel, R., Zavala-Díaz, C., & Martínez-Rebollar, A. (2019). The K-means algorithm evolution. *Introduction to data science and machine learning*, 69-90.

Romanuke, V. (2023). Random centroid initialization for improving centroid-based clustering. *Decision Making: Applications in Management and Engineering*, *6*(2), 734-746.

Faisal, M., & Zamzami, E. M. (2020, June). Comparative analysis of inter-centroid K-Means performance using euclidean distance, canberra distance and manhattan distance. In *Journal of Physics: Conference Series* (Vol. 1566, No. 1, p. 012112). IOP Publishing.

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE access*, *8*, 80716-80727.

Yuan, C., & Yang, H. (2019). Research on K-value selection method of K-means clustering algorithm. *J*, *2*(2), 226-235.

Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023). An unsupervised machine learning algorithms: Comprehensive review. *International Journal of Computing and Digital Systems*.