DINAMIKA REKAYASA
http://dinarek.unsoed.ac.id

Volume 22 Nomor 1 (2026) Hal. 126-135
e-ISSN 2527-6131
http://jurnaldinarek.id

# KLASIFIKASI TINGKAT KERUSAKAN JALAN PERKOTAAN MENGGUNAKAN PARAMETER INDEKS KERUSAKAN PERMUKAAN DAN ALGORITMA K-MEANS

## *CLASSIFICATION OF URBAN ROAD DAMAGE LEVELS USING SURFACE DISTRESS INDEX PARAMETERS AND K-MEANS ALGORITHM*

**Yogi Oktopianto*[1], Antonius[2], Abdul Rochim[2]**

*Email: yogi.oktopianto@pktj.ac.id

[1]Politeknik Keselamatan Transportasi Jalan, Tegal, 52125 Indonesia
[2]Department of Civil Engineering, Faculty of Engineering, Universitas Islam Sultan Agung, Semarang, 50112, Indonesiag

***Abstrak***

Kerusakan jalan merupakan masalah signifikan di banyak negara, seperti di Indonesia yang dapat mempengaruhi keselamatan, efisiensi transportasi, dan kualitas hidup Masyarakat terutama jalan perkotaan. Penelitian ini bertujuan mengembangkan model berbasis data menggunakan algoritma K-Means untuk mendeteksi dan mengelompokkan tingkat kerusakan jalan perkotaan berdasarkan parameter Surface Distress Index (SDI). Data yang digunakan mencakup 2467 segmen jalan dengan informasi mengenai jenis dan tingkat kerusakan selama lima tahun terakhir. Model klaster dirancang dengan dua dan empat klaster, dengan hasil menunjukkan bahwa model empat klaster memberikan pemisahan yang lebih jelas dan representatif terhadap kondisi jalan. Nilai Silhouette Coefficient model empat klaster 0.513, menunjukkan pemisahan yang detail dan lebih jelas dibandingkan model dua klaster dengan nilai Silhouette value of 0.423. Perbandingan antara model dua dan empat klaster menunjukkan bahwa model empat klaster lebih baik dalam membedakan struktur data yang kompleks. Hasil penelitian ini berkontribusi dalam pengembangan sistem pemantauan kondisi jalan berbasis data prioritas pemeliharaan berdasarkan kategori kondisi jalan, yang dapat diadaptasi untuk perbaikan kebijakan infrastruktur di Indonesia dan negara berkembang lainnya.

***Kata kunci***: Algoritma K-Means, Kerusakan Jalan, Klaster, Kondisi Jalan, Pemeliharaan Jalan, SDI

***Abstract***

Road damage is a significant problem in many countries, including Indonesia, and it can affect safety, transportation efficiency, and the quality of life of communities, especially on urban roads. This study aims to develop a data-based model using the K-Means algorithm to detect and classify the levels of urban road damage based on the Surface Distress Index (SDI) parameter. The data used comprised 2,467 road segments containing information on the types and levels of damage over the past five years. The clustering model was designed with two and four clusters, and the results indicated that the four-cluster model provided a clearer and more representative separation of road conditions. The Silhouette Coefficient value of the four-cluster model is 0.513, indicating a more detailed and clearer separation compared to the two-cluster model with a Silhouette value of 0.423. The four cluster model is better at telling apart complex data structures than the two cluster model. The results of this study contribute to the development of a road condition monitoring system based on maintenance priority data categorized by road condition, which can be adapted to improve infrastructure policies in Indonesia and other developing countries.

***Keywords:*** K-Means Algorithm, Road Condition, Clustering, Road Damage, Road Maintenance, SDI

# I. INTRODUCTION

One of the important aspects of transportation is road maintenance. Better road conditions significantly improve travel quality and user safety [1], [2], [3]. Current developments require the creation of more efficient and effective methods for assessing road conditions [4], [5]. In assessing the condition of road surfaces, the Surface Distress Index (SDI) value is a key indicator for urban roads [6], [7], [8]. The Surface Distress Index (SDI) is often used to classify and group various categories of road damage [9].

The Surface Distress Index (SDI) value explains the cumulative effect of road damage and directly affects user comfort and traffic safety [9], [10]. In addition, this method consists of several components that provide more detailed and reliable information about road conditions [11]. Therefore, to ensure and maintain a good transportation infrastructure, it is essential to carry out continuous maintenance [12].

Information extraction from datasets generated by the Surface Distress Index (SDI) regarding road conditions can be enhanced using Machine Learning (ML)-based techniques such as K-Means clustering. The K-Means algorithm can classify road areas with similar levels of damage, thereby facilitating the prioritization of road maintenance [13]. This method has advantageous characteristics for managing large and complex road condition assessment data. In addition, K-Means can help identify specific types of road damage learned by the algorithm from patterns that were previously difficult to detect using conventional methods [14], [15].

Previous studies have combined conventional evaluation methods, such as the Surface Distress Index (SDI) and Pavement Condition Index (PCI), with innovative intelligent imaging and Machine Learning (ML) technologies, such as K-Means, resulting in improved accuracy and efficiency in road maintenance management [16], [17], [18]. These technologies work synergistically to accelerate damage detection and analysis, enabling fast and contextual maintenance decisions in accordance with actual conditions [19]. This serves as the foundation for road infrastructure maintenance management strategies based on data and technology.

The Surface Distress Index (SDI) has been applied in several previous studies to assess road damage. However, most of these studies still rely on methods that are less efficient in classifying road damage information. Some studies have developed broader classification methods based on the Surface Distress Index (SDI) and the International Roughness Index (IRI). Nevertheless, there is a very limited number of studies that adopt a big data–based approach to optimize the clustering process, such as K-Means, which is a Machine Learning (ML)–based clustering algorithm. This limitation encompasses aspects of urban roads and the local characteristics of road surface damage in Indonesia. K-Means addresses the problem by achieving faster convergence and producing better clusters compared to conventional methods, which often face convergence issues or non-optimal classification results [20], [21].

This study aims to develop an innovative road damage classification method by integrating road damage parameters from the Surface Distress Index (SDI) method using the K-Means Algorithm, to enhance the efficiency and reliability of overall road maintenance management. This approach not only improves the road damage classification process but also enables the development of a more proactive predictive maintenance system. The results of this research contribute to the development of a road condition monitoring system based on maintenance priority data categorized by road condition, which can be adapted for improving urban road infrastructure policies. Such a system can preserve road lifespan, improve safety, and more broadly support sustainable transportation infrastructure management.

# II. LITERATUREREVIEW

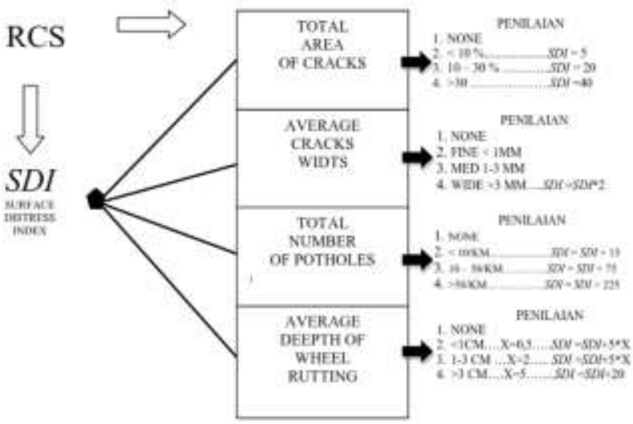


**Figure 1.** SDI Analysis Framework

The Surface Distress Index (SDI) is a method for assessing the condition of road pavement surfaces through visual observation and is used as a reference in the road maintenance program used in Indonesia. The four variables needed to calculate the SDI value using the Road Condition Survey (RCS) are the percentage of cracked area, the average crack gap width, the number of potholes, and the average rut depth [22]. Figure 1 illustrates the details of the Surface Distress Index (SDI) method formula.

The cluster approach used in the assessment of pavement conditions based on the SDI method is in accordance with the guidelines of the Direktorat Jenderal Bina Marga (2011), as outlined in the Indonesian Integrated Road Management Systems (IIRMS) Guide No. SMD-03/RCS on Road Condition Surveys. The SDI rating will indicate the state of the road pavement as good, Moderate, slightly damaged, or badly damaged. Table 1 presents the cluster grouping standards used.

**Table 1**. SDI Cluster Model

| SDI Value | Road Condition | SDI Stability Level |
|---|---|---|
| <50 | Good | Stable Road |
| 50-100 | Moderate | |
| 100-150 | Slightly Damaged | Unstable Road |
| >150 | Severely Damaged | |

Source: Bina Marga, 2011

The Silhouette Index and the Calinski–Harabasz Index are internal evaluation metrics that can be used to check the quality of clustering results without needing reference labels. The Silhouette Index shows how well a data point fits in with its cluster compared to the nearest other cluster. It does this by looking at how close the point is to other points in the cluster and how far away the clusters are from each other. A value closer to 1 means that the clusters are better separated.

The Calinski–Harabasz Index, on the other hand, looks at the difference between clusters and the difference within clusters. The bigger the number, the clearer and more compact the cluster structure is. These two metrics are used because they can objectively give a numerical measure of how well the clusters are separated and how close they are to each other.

T-Distributed Stochastic Neighbor Embedding (t-SNE) is used for visualizing clustering results by reducing the dimensions of high-dimensional data to two or three-dimensional space. T-Distributed Stochastic Neighbor Embedding (t-SNE) is used for visualizing clustering results by reducing the dimensions of high-dimensional data to two or three-dimensional space.

## III. METHOD

### *A. Data Collection*

The road condition data in this study were collected through direct surveys on 42 urban road segments. Historical road damage data from the past five years were obtained from the Department of Public Works (Dinas Pekerjaan Umum; DPU) of Tegal City, Road Maintenance Division, which is responsible for monitoring and maintaining road infrastructure. Out of a total of 42 road sections, the observation was conducted in segments. This means that each road section is divided into several segments, each 50 meters long. The total database used includes 2467 records containing information about the type and level of damage. The collected damage data included crack area, crack width, number of potholes, and rutting or wheel track marks as the basis for cluster calculation using the Surface Distress Index (SDI) method, which represents roads in good, slightly damaged, moderately damaged, and severely damaged conditions.

### B. Data Preprocessing

After the data were collected, a preprocessing stage was carried out to prepare the data for cluster analysis. The technique used was Min-Max Normalization, which aims to normalize the data values to a range between 0 and 1. This normalization is essential to ensure that each data parameter contributes equally to the clustering calculation and to prevent parameters with larger values from dominating the computation. The normalization process was performed using the following formula:

$$x' = \frac{x - \min(X)}{\max(X) - \min(X)} \qquad (1)$$

X represents the original value, min (X) is the minimum value of the dataset, max (X) is the maximum value of the dataset, and $x'$ is the normalized value. After the normalization stage, the data became more homogeneous and ready to be used in the clustering algorithm without bias from differing data scales.

Normalization is performed because the variables used have different scales and units, such as area of cracks (%), cracks widths (mm), potholes (pieces), and rutting (cm). These scale differences can cause variables with larger value ranges to dominate distance calculations in distance-based algorithms like K-Means, so Min-Max Scaling is used to transform the data to the same range. Min-Max is chosen because it better preserves the proportional interpretation of the original data and does not assume a normal distribution, making it more suitable for road condition data, which often does not follow a normal distribution.

### C. Clustering Algorithm

This study used the K-Means algorithm, modeled using machine learning in R-Studio, to group road conditions based on the level of damage. The steps in the K-Means algorithm are as follows:

1. Initialization of Cluster Centers (Centroids)

The first step was to determine the desired number of clusters ($k$), which in this study involved creating two clustering models. The first model consisted of two clusters that represented stable roads and non-stable roads. The second model consisted of four clusters that represented the levels of road damage based on the condition categories of good, moderate, slightly damaged, and severely damaged (Bina Marga, 2011).

2. Distance Calculation Between Data Points

Each data point is measured for its distance to each cluster center using the Euclidean distance in the K-Means algorithm.

3. Data Clustering

Based on the shortest distance, each data point is assigned to the cluster whose center is closest to it.

4. Updating Cluster Centers

After the data are grouped, the cluster centers are recalculated based on the average position of the data points within each cluster. These new cluster centers then serve as reference points for the next iteration.

5. Iteration Until Convergence

This process was repeated until the positions of the cluster centers no longer changed significantly (converged). At this point, the algorithm stopped, and the resulting clusters were analyzed.

Clustering was performed using the stats::kmeans function in R with 25 random initializations (nstart = 25) to reduce the risk of local minima. The algorithm used the default maximum iteration setting (iter.max = 10). Cluster validation was conducted using the Silhouette coefficient from the cluster package, and visualization was performed using factoextra:: fviz_cluster.

### D. Algorithm Cluster Quality Evaluation

To evaluate the quality of the clustering results, the Silhouette Coefficient was used. The Silhouette Coefficient value provides an overview of the quality of separation between clusters in the clustering analysis. In general, a value between 0.5 and 1 indicates good clustering quality, with clear separation between clusters. Values ranging from 0.25 to 0.5 indicate fairly good clustering quality, although there is still room for improvement in cluster separation. Conversely, values below 0.25 indicate poor clustering quality, where there is a high probability of overlapping clusters or unclear separation. By using the Silhouette Coefficient, the quality of the resulting clusters was evaluated to ensure that the clustering based on road damage levels was valid and effective.

## IV. RESULTS AND DISCUSSION

### *A. Analysis Results of the Two-Cluster Model*

This two-cluster model was developed based on the Surface Distress Index (SDI) values to measure the level of road stability. In this model, the two clusters formed represent two main categories of road conditions, namely stable roads and non-stable roads. This division aimed to provide a general overview of road quality in the analyzed area, focusing on whether the roads were in proper condition for use or required further maintenance. The determination of the number of clusters ($k$) is carried out systematically using internal evaluation methods, namely the Silhouette index and Calinski–Harabasz approach. The optimal number of clusters is chosen based on the data structure that is most representative of the characteristics of the Surface Distress Index (SDI) values. Table 2 presents the cluster distribution and centroid values obtained from the K-Means model with k = 2.

**Table 2**. Cluster Distribution and Centroid Values (k = 2)

| Cluster | Cluster Size | Centroid | | | |
|---|---|---|---|---|---|
| | | Crack Area | Crack Width | Number of Potholes | Rutting |
| 1 | 1036 | 0.5796 | 0.4963 | 0.4369 | 0.2353 |
| 2 | 1431 | 0.3704 | 0.1327 | 0.0201 | 0.0048 |

The K-Means results with k = 2 indicate that the pavement conditions can be clearly distinguished into two main clusters with different distributions of pavement conditions in terms of surface pavement damage. Cluster 1 has 1,036 cluster members with high centroid values on indicators such as crack area, crack width, potholes, and rutting, indicating a moderate to severe level of damage and a decrease in structural stability. Meanwhile, Cluster 2 has 1,431 cluster members with much lower centroid values on all indicators, indicating stable pavement conditions with light damage levels. The performance and analysis results of the two-cluster road damage model can be seen in Table 3.

**Table 3.** Summary of the Two-Cluster K-Means Model

| N | k | AIC | BIC | Silhouette Coefficient | Calinski-Harabasz Index |
|---|---|---|---|---|---|
| 2467 | 2 | -4811.3 | -4799.68 | 0.423 | 1703.358 |

*Number of Data (N), Number of Clusters (k)

Overall, the two-cluster model showed a fairly clear separation. However, the relatively low Silhouette value of 0.423 indicates that, although the separation between clusters was acceptable, the quality of this separation could still be improved to achieve more optimal clustering results. A Silhouette Coefficient value within the range of 0.25 to 0.5 reflects fairly good clustering quality, but there remains room for improvement. Based on the values presented in Table 2, although the clustering model provided a reasonably good separation between clusters (as indicated by the relatively high Calinski-Harabasz Index), there is evidence that the separation of data within clusters was not perfect. This indicates that while the clustering performance was fairly good, there is still potential to enhance the separation between clusters for more optimal results. To visualize the separation between the two resulting clusters, the dimensionality reduction results using the t-SNE technique are presented. Figure 1 illustrates how the data were grouped into two distinct clusters with a clear separation.

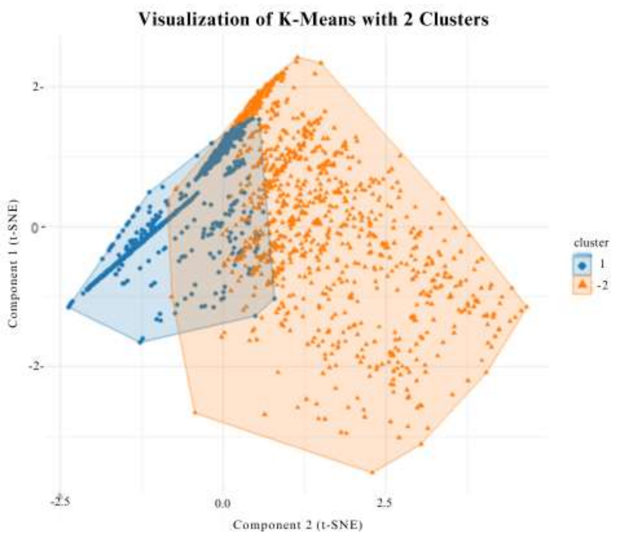


**Figure 2.** Visualization of Model 1 Clustering

The clustering visualization in Figure 2 shows a clear separation between the two identified clusters, each represented by blue (Cluster 1) and orange (Cluster 2). The use of t-SNE as a dimensionality reduction technique successfully separated these clusters along t-SNE Component 1 and Component 2, demonstrating that the two clusters can be easily distinguished in a low-dimensional space. Cluster 1, which appears more concentrated and homogeneous, indicates a high degree of similarity among data points, while Cluster 2 has a broader spread, suggesting greater variation in data characteristics. Although a few points are located farther from the cluster center, the separation between clusters remains noticeably clear.

These results indicate that the clustering method successfully identified two distinct groups of data, with t-SNE effectively reducing the data dimensions for easier analysis. The satisfactory Silhouette Coefficient confirms an adequate quality of cluster separation. Overall, this t-SNE visualization reinforces the finding that the two clusters possess significantly different characteristics. Although the two-cluster grouping provided a fairly clear separation between stable and non-stable road conditions, the model was not yet fully optimal. The high variation within Cluster 1 suggests that the model was not sufficiently effective in handling the diversity of data present in stable road conditions. Therefore, developing a model with more clusters is necessary to achieve a more detailed and representative separation between various levels of road damage.

### *B. Analysis Results of the Four-Cluster Model*

This four-cluster model was developed to classify road conditions based on more detailed levels of damage using the Surface Distress Index (SDI) values. In this model, four clusters were formed, each representing the categories of good, moderate, slightly damaged, and severely damaged road conditions. This classification aimed to provide a deeper analysis of road conditions, focusing on identifying the existing levels of damage. Such an approach can assist in determining the priorities for road repair and maintenance in the analyzed area, thereby ensuring more efficient and well-targeted handling. Table 4 presents the cluster distribution and centroid values obtained from the K-Means model with k = 4.

**Table 4**. Cluster Distribution and Centroid Values (k = 4)

| Cluster | Cluster Size | Centroid | | | |
|---|---|---|---|---|---|
| | | Crack Area | Crack Width | Number of Potholes | Rutting |
| 1 | 295 | 0.5945 | 0.4913 | 0.6356 | 0.6364 |
| 2 | 675 | 0.5665 | 0.5587 | 0.0633 | 0.046 |
| 3 | 1080 | 0.3147 | 0.0478 | 0.0119 | 0.0025 |
| 4 | 417 | 0.5586 | 0.3126 | 0.5713 | 0.07 |

The K-Means algorithm indicates that when k = 4, the segmentation of pavement conditions is based on several surface damage indicators. In cluster 3, where the number of members is 1,080, all the indicators have the lowest centroid values. This indicates that the condition of the pavement is stable and only suffers from minor damage, and this is where preventive measures should be applied. In cluster 2, where the number of members is 675, all the indicators have high centroid values except for potholes and rutting, which have relatively low values. This indicates that the damage to the pavement is caused by cracks, and this is where the damage starts to affect the structural capacity of the pavement.

Cluster 4, which consists of 417 members, has a high centroid value for crack area indication and number of potholes. The centroid value shows that the deterioration in function is quite advanced. More intensive maintenance handling is required. Cluster 1, which consists of 295 members, has the highest value in all indicators that fall into this cluster. These indicators are the number of potholes and rutting. The overall configuration of these four clusters can better explain the gradation of road conditions from good to worse. Thus, it can better support the determination of handling strategies in road asset management.

To provide an overview of the performance of the applied clustering model, Table 3 presents a summary of the results from implementing the K-Means algorithm with four clusters. This table includes evaluation metrics such as AIC, BIC, Silhouette Coefficient, and Calinski-Harabasz Index, which were used to assess the quality of separation between clusters and the model's fit to the data. These metrics offer important insights into how effectively the model can cluster the data and how well it corresponds to the existing data structure.

The four-cluster model showed a significant improvement compared to the two-cluster model, as reflected by the higher Silhouette Coefficient value, which indicates better separation between clusters. With a Silhouette Coefficient value of 0.513, the model reflects good clustering quality and indicates that the more detailed data separation in the four-cluster model resulted in a more accurate and representative analysis of the analyzed data conditions. In general, a value between 0.5 and 1 indicates good clustering quality, with clear separation between clusters.

**Table 5**. Summary of the Four-Cluster K-Means Model

| N | k | AIC | BIC | Silhouette Coefficient | Calinski-Harabasz Index |
|---|---|---|---|---|---|
| 2467 | 4 | -6640.2 | -6616.96 | 0.513 | 2097.482 |

*Number of Data (N), Number of Clusters (k)

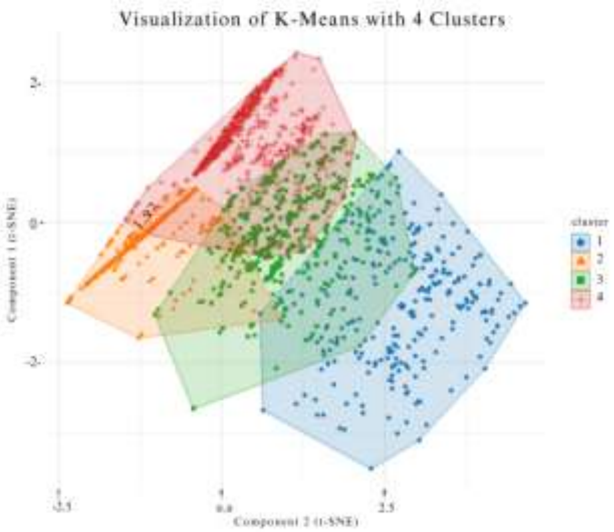


**Figure 3**. Visualization of the Four-Cluster Model

Based on the metrics presented in Table 5, the model showed a high Calinski-Harabasz Index value, which indicates good separation between clusters. The model with four clusters (k=4) provided more detailed results by dividing road conditions into four categories: good, moderate, slightly damaged, and severely damaged. This division provided a clearer and more representative separation of road conditions.

The visualization results shown in Figure 3 display a clear separation among the four identified clusters using the t-SNE technique. These four clusters represent road conditions based on the Surface Distress Index (SDI) parameter, which categorizes them into four levels: good, moderate, slightly damaged, and severely damaged roads. Consistent with similar studies, the appropriate increase in the number of clusters can improve the accuracy and quality of data separation, allowing the clustering model to capture more complex patterns [23], [24], [25].

Overall, this visualization shows that the t-SNE technique successfully reduced the data dimensions and effectively separated the clusters. However, there remains room for improvement in achieving a finer separation, particularly in clusters with greater data variation.

Advanced clustering techniques, such as Hierarchical Clustering, can be used to identify more detailed structural patterns. To provide a clearer illustration of data separation based on road damage parameters, Figure 4 presents a visualization of data distribution using four main parameters: crack area, crack width, number of potholes, and wheel tracks. This visualization illustrates how the four clusters are distributed across each parameter, facilitating the identification of differences in road conditions within each group.

**Figure 4.** Cluster Matrix Plot Based on Road Damage Parameters

The visualization in Figure 4 provides a clear illustration of the distribution of each cluster based on the analyzed road damage parameters. Each cluster exhibits distinct characteristics, reflecting variations in the degree of road damage conditions. The clear separation between these clusters indicates that the data can be effectively grouped based on differences in parameters such as crack area, crack width, number of potholes, and wheel tracks. These differences enable the identification of varying road conditions, which can be used for further analysis, such as determining repair priorities and addressing damage issues based on the severity level detected in each cluster.

In Figure 3, a scatterplot matrix (pairplot) is provided for the four parameters for road damage, where observations are color-coded according to the K-Means clustering algorithm results. The diagonal elements show the distribution for each variable, while the off-diagonal elements show relationships between pairs of variables. From the figure, it is evident that crack width and wheel tracks show the strongest separation for clusters; therefore, these parameters have a strong clustering effect. On the contrary, crack areas show a high degree of overlap for clusters; therefore, it has a relatively weaker clustering effect.

From the figure 3, it is also evident that road condition classification is a multi-dimensional feature; therefore, no single feature is capable of clearly separating clusters from one another. Instead, a non-linear combination of crack characteristics, pothole frequency, and rutting intensity is responsible for clustering behavior. This is a strong indication that machine learning algorithms for clustering are appropriate for road condition classification as opposed to traditional techniques

Overall, these results align with clusters found in other studies, indicating that selecting the appropriate number of clusters is crucial for improving clustering quality [26], [27], [28]. The model with four clusters (k=4) showed better separation and a higher capability in explaining data variation. The value of 0.513 in four clusters falls within the range of $0.50 <$ Silhouette $\leq 0.70$, indicating that the clusters have a moderate structural criterion [29]. Compared to a similar model with different variables, which produced a Silhouette range of 0.407543 to 0.57432 [30], [31], [32].

This value becomes a limitation in the clustering results of road conditions based on the Surface Distress Index (SDI) parameter on urban roads, indicating that further exploration of the cluster (k) test is needed to achieve the most optimal outcome. These results are highly beneficial in practical applications, such as road infrastructure planning and maintenance, where better separation between road damage conditions can assist in determining more targeted and efficient repair priorities.

## V. CONCLUSION

This study successfully developed a clustering model to detect and classify urban road conditions based on their levels of damage. By using the K-Means algorithm, the developed model was able to categorize road conditions into more detailed classifications, allowing for a deeper analysis of road infrastructure conditions. The results of the study showed that the four-cluster model provided a clearer and more representative separation of road conditions compared to the two-cluster model, which tended to produce a coarser distinction between good and poor road conditions.

Based on 2467 observations on 42 urban road, the four-cluster model has a silhouette coefficient value of 0.513, while the two-cluster model has a silhouette value of 0.423. The four-cluster model showed a higher Silhouette Coefficient, with more homogeneous clusters and better clustering quality. This indicates that the model was more effective in explaining variations in road conditions and produced a more accurate separation. In addition, the comparative analysis between the two-cluster and four-cluster models showed that the four-cluster model performed better in distinguishing complex data structures, as evidenced by the higher Calinski-Harabasz Index value in the four-cluster model.

These findings can contribute to improving road maintenance and repair policies, especially in addressing complex variations in road conditions. The results of this study also open the potential for developing a more efficient and effective road condition monitoring system, which is expected to be applicable in other regions of Indonesia as well as in other developing countries facing similar challenges.

While the four-cluster model demonstrates promising outcomes, additional research may be undertaken to evaluate and enhance this model across a wider and more diverse setting. Incorporating other algorithmic types can enhance this model and augment the precision of cluster delineation. The limitations of the road condition clustering results based on the Surface Distress Index (SDI) parameter in urban roads indicate that further exploration of the cluster (k) test is necessary to achieve the most optimal results.

## ACKNOWLEDGEMENTS

This research is supported by the Department of Civil Engineering at Universitas Islam Sultan Agung and Politeknik Keselamatan Transportasi Jalan. The research partner providing the dataset is the Public Works Department of Tegal City.

## REFERENCE

[1] J. A. Padilla and O. G. dela Cruz, “Methods on Calculating the International Roughness Index: A Literature Review,” in *Lecture Notes in Civil Engineering*, 2022. doi: 10.1007/978-3-030-99979-7_2.

[2] M. Montoya-Alcaraz, A. Mungaray-Moctezuma, and L. García, “Sustainable road maintenance planning in developing countries based on pavement management systems: Case study in Baja California, México,” *Sustainability (Switzerland)*, vol. 12, no. 1, 2020, doi: 10.3390/su12010036.

[3] G. Bahati and E. Masabo, “Optimizing Ambulance Location Based on Road Accident Data in Rwanda Using Machine Learning Algorithms ,” 2024, doi: 10.21203/rs.3.rs-5319700/v1.

[4] M. Jan, K. S. Khattak, Z. H. Khan, T. A. Gulliver, and A. B. Altamimi, “Crowdsensing for Road Pavement Condition Monitoring: Trends, Limitations, and Opportunities,” *IEEE Access*, vol. 11, 2023, doi: 10.1109/ACCESS.2023.3332667.

[5] Z. Ahmed Alkaissi, “Traffic congestion evaluation of urban streets based on fuzzy inference system and GIS application,” *Ain Shams Engineering Journal*, vol. 15, no. 6, 2024, doi: 10.1016/j.asej.2024.102725.

[6] M. Isradi, A. D. Hedianto, A. I. Rifai, A. Mufhidin, and J. Prasetijo, “Comparison of PCI (Pavement Condition Index) and SDI (Surface Disstres Index) in Identification of Urban Road Damage,” *ADRI International Journal of Engineering and Natural Science*, vol. 6, no. 01, 2021, doi: 10.29138/aijens.v6i01.34.

[7] E. R. Labaso, M. S. Ishak, and M. Kasan, “Evaluasi Kerusakan Jalan Menggunakan Metode Pavement Condition Index (PCI) dan Surface Distress Iindex (SDI) Studi Kasus Jalan Pue Bongo – Kota Palu,” *REKONSTRUKSI TADULAKO: Civil Engineering Journal on Research and Development*, 2022, doi: 10.22487/renstra.v3i2.428.

[8] M. Isradi, J. Prasetijo, N. Hartatik, Z. Abidin, and Z. Arifin, “Analysis of Urban Road Damage Assessment Using Surface Distress Index (SDI), Pavement Condition Index (PCI), and International Roughness Index (IRI) Methods. Review of International Geographical Education (RIGEO),” *RIGEO* •

*Review of International Geographical Education*, vol. 11, no. 2, 2021.

[9] B. H. Setiadji, D. Purwanto, and Y. I. Wicaksono, "Improvement of Potholes and Rutting Assessment in Surface Distress Index*," 2020, doi: 10.2991/aer.k.200220.034.

[10] Y. Oktopianto, A. Antonius, and A. Rochim, "Evaluation of Urban Road Stability Through the Integration of the Surface Distress Index and International Roughness Index," *Fondasi: Jurnal Teknik Sipil*, vol. 14, no. 1, pp. 1–12, Apr. 2025, doi: 10.62870/fondasi.

[11] R. Utami, "Pavement Maintenance Strategy for Provincial Roads in Tarakan City, North Kalimantan," *E3s Web of Conferences*, vol. 479, p. 07002, 2024, doi: 10.1051/e3sconf/202447907002.

[12] A. M. Hidayat and G. A. Prasetyaningtiyas, "Evaluation of Pavement Condition Based on the Value of Surface Distress Index and Pavement Condition Index on the Road Section of Gedong Tataan - Kedondong, Pesawaran," *Dinamika Teknik Sipil Majalah Ilmiah Teknik Sipil*, vol. 16, no. 2, pp. 68–78, 2023, doi: 10.23917/dts.v16i2.23271.

[13] A. Lubis, I. Iskandar, and M. M. L. W. Panjaitan, "Implementation of KNN Methods and GLCM Extraction for Classification of Road Damage Level," *Iaic Transactions on Sustainable Digital Innovation (Itsdi)*, vol. 4, no. 1, pp. 1–7, 2022, doi: 10.34306/itsdi.v4i1.564.

[14] Y. Khosravi, F. Hosseinali, and M. Adresi, "Assessing Road Accidents In Spatial Context Via Statistical And Nonstatistical Approaches To Detect Road Accident Hotspot Using GIS," *Geodetski Vestnik*, vol. 66, no. 3, 2022, doi: 10.15292/geodetski-vestnik.2022.03.412-431.

[15] Z. Diao, X. Huang, H. Liu, and Z. Liu, "LE-YOLOv5: A Lightweight and Efficient Road Damage Detection Algorithm Based on Improved YOLOv5," *International Journal of Intelligent Systems*, vol. 2023, 2023, doi: 10.1155/2023/8879622.

[16] C. Makendran, M. Karthik, S. M. J. Hasan, M. Harivignesh, and G. V Raahul, "Designing an Intelligent Pavement Maintenance and Management System Using Drone Imagery and Artificial Intelligence," *Matec Web of Conferences*, vol. 393, p. 02005, 2024, doi: 10.1051/matecconf/202439302005.

[17] T. Deng and Y. Tan, "Efficient Pavement Distress Detection and Visual Management in Lean Construction Based on BIM and Deep Learnin," 2023, doi: 10.24928/2023/0230.

[18] Y. Oktopianto, Antonius, and A. Rochim, "An Artificial Neural Network Approach for Predicting Pavement Distress: A Case Study Toward Sustainable Road Maintenance," *Advance Sustainable Science, Engineering and Technology*, vol. 7, no. 3, May 2025, doi: 10.26877/asset.v7i3.2133.

[19] S. Jha *et al.*, "Data-Driven Web-Based Patching Management Tool Using Multi-Sensor Pavement Structure Measurements," *Transportation Research Record Journal of the Transportation Research Board*, vol. 2677, no. 12, pp. 83–98, 2023, doi: 10.1177/03611981231167161.

[20] S. M. H. Hasheminejad and M. Khorrami, "Clustering of bank customers based on lifetime value using data mining methods," *Intelligent Decision Technologies*, vol. 14, no. 4, 2020, doi: 10.3233/IDT-190176.

[21] H. G. Thi Hoang, T. A. Nguyen, H. L. Nguyen, and H. B. Ly, "Neural network approach for GO-modified asphalt properties estimation," *Case Studies in Construction Materials*, vol. 17, 2022, doi: 10.1016/j.cscm.2022.e01617.

[22] Bina Marga, *Bina Marga*. Indonesia, 2011.

[23] M. Gholami, M. Mouhoub, and S. Sadaoui, "Biogeography-Based Optimization for Feature Selection," *The International Flairs Conference Proceedings*, vol. 36, 2023, doi: 10.32473/flairs.36.133230.

[24] K. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," pp. 747–748, 2020, doi: 10.1109/dsaa49011.2020.00096.

[25] L. K. Giang, H. H. T. Lan, D. Van Manh, and T. Q. Hoc, "Applying a Two-Step Cluster Algorithm in Traffic Accident Data Analysis," *The Transport and Communications Science Journal*, vol. 75, no. 4, pp. 1673–1687, 2024, doi: 10.47869/tcsj.75.4.16.

[26] E. Sanjari, F. M. Dehkordi, and H. R. Shahraki, "Clustering Undergraduate Students Based on Academic Burnout and Satisfaction From the Field Using Partitioning Around Medoid," *Comput. Math. Methods Med.*, vol. 2023, no. 1, 2023, doi: 10.1155/2023/8898939.

[27] Y. Chen, Y. Xue, R. Wang, and G. Zhang, "Underwater Source Counting With Local-Confidence-Level-Enhanced Density Clustering," *Sensors*, vol. 23, no. 20, p. 8491, 2023, doi: 10.3390/s23208491.

[28] Q. Li, J. Noel-MacDonnell, D. C. Koestler, E. L. Goode, and B. L. Fridley, "Subject Level Clustering Using a Negative Binomial Model for Small Transcriptomic Studies," *BMC Bioinformatics*, vol. 19, no. 1, 2018, doi: 10.1186/s12859-018-2556-9.

[29] N. Istiqomah, W. Dianita Utami, and D. Yuliati, "Penerapan K-Means Clustering untuk Analisis Kondisi Lalu Lintas di Jalan Ir. H. Soekarno Surabaya," *Jurnal Keselamatan Transportasi Jalan (Indonesian Journal of Road Safety)*, vol. 12, no. 2, pp. 78–88, Dec. 2025, doi: 10.46447/ktj.v12i2.725.

[30] M. S. Firdaus, "Segmentasi Mutu Jalan Nasional Menggunakan Metode Clustering K-Medoids di Kabupaten Bandung Barat dan Kota Bandung," *Prosiding Sains Nasional dan Teknologi*, vol. 12, no. 1, 2022, doi: 10.36499/psnst.v12i1.7173.

[31] A. Wulandari, Irmayansyah, and L. T. Ningrum, "Penerapan Metode K-Means Clustering Untuk Pemetaan Perbaikan Jalan Di Kota Bogor," *Jurnal Informatika Teknologi dan Sains (Jinteks)*, vol. 7, no. 1, 2025, doi: 10.51401/jinteks.v7i1.5323.

[32] S. Asmiatun, "Penerapan Metode K-Medoids Untuk Pengelompokkan Kondisi Jalan Di Kota Semarang," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 6, no. 2, 2019, doi: 10.35957/jatisi.v6i2.193.