

CHAPTER V

CONCLUSION

This chapter presents a summary of the research that has been conducted, including the main conclusions derived from the analysis and discussions in the previous chapters. In addition, this chapter provides several recommendations that may serve as references for future research development, either in terms of data enrichment, methodological improvement, or the practical application of clustering results for decision-making in public transportation operations.

5.1. Conclusion

Based on the research and analysis that have been carried out, several conclusions can be drawn as follows:

1. This study successfully performed clustering on the Suroboyo Bus passenger density data using three different methods: Agglomerative Hierarchical Clustering (AHC), Gaussian Mixture Model (GMM), and K-Medoids. The clustering process was conducted using two main attributes—operational time and number of passengers—after completing several preprocessing stages, including attribute selection, time transformation into numerical format, outlier detection using the IQR method, and normalization using the Z-Score.
2. The clustering results from the three methods produced four clusters, which can generally be categorized as low, medium, high, and very high based on the average passenger count. Based on the analysis of cluster centers, AHC and K-Medoids were able to form groups with relatively clear separation, while GMM tended to produce cluster centers that were closer to each other due to its probabilistic nature, which allows overlap between clusters.
3. In terms of cluster characteristics, the AHC method was able to differentiate between density levels more distinctly, particularly between low- and high-density groups. The K-Medoids method produced stable results, especially in the presence of extreme values (outliers), and was able to maintain a good balance between time and passenger density. Meanwhile, GMM demonstrated flexibility in capturing complex variations within the data, although its cluster boundaries tended to be less distinct, particularly when the data distribution was not clearly separated.

4. The evaluation using the Silhouette Score showed that K-Medoids achieved the highest score of 0.4222, followed by AHC with 0.3657, and GMM with 0.3024. These values indicate that K-Medoids produced the best clustering quality with the clearest cluster separation and highest internal compactness, while GMM demonstrated the least optimal separation performance.
5. The cluster labeling results indicate that there are noticeable differences in passenger density patterns across specific times and locations. This information can be utilized by transportation authorities, particularly the Surabaya City Department of Transportation, as a basis for improving operational efficiency—such as adjusting bus schedules or determining strategic locations for new bus stops.

5.2. Recommendation for Future Research

The following recommendations are provided to support further development of future studies and to enhance the applicability of the results:

1. This study used only two main attributes. Adding more variables, such as operational day, type of bus fleet, or detailed bus stop location data, may provide deeper and more comprehensive clustering results.
2. Utilizing a larger dataset with a longer time span (e.g., covering monthly or annual periods) may increase the reliability of the results and allow the identification of seasonal patterns in passenger density.
3. Future research may consider exploring hybrid methods or combining several clustering algorithms to leverage the strengths of each technique, thereby producing more accurate and stable clustering outcomes.