

ABSTRAK

Rumah Sakit Arun Lhokseumawe memiliki data jenis penyakit pasien rawat inap yang beragam dan terus bertambah, namun belum dimanfaatkan secara optimal untuk analisis lebih lanjut. Selama ini, data hanya menjadi laporan administratif tanpa dilakukan pengolahan guna memperoleh informasi bermakna, seperti pola dominasi penyakit. Riset ini bertujuan untuk mengelompokkan jenis penyakit pasien rawat inap menggunakan dua metode klasterisasi, yaitu *Agglomerative Hierarchical Clustering* dan *K-Medoids*, serta menganalisis perbandingan kinerjanya. Data yang dipergunakan mencakup 84 jenis penyakit yang direkam pada periode Desember 2024 hingga Januari 2025, dengan atribut jumlah pasien laki-laki, pasien perempuan, dan umur pasien. Klasterisasi *Agglomerative Hierarchical Clustering* dilakukan dengan pendekatan *average linkage* dan jarak *Manhattan Distance*, sedangkan *K-Medoids* menggunakan jarak *Euclidean Distance*. Hasil memperlihatkan bahwa metode *Agglomerative Hierarchical Clustering* membentuk 3 cluster, yaitu C1 dengan 4 jenis penyakit, C2 menghasilkan 79 jenis penyakit, dan C3 menghasilkan 1 jenis penyakit. Sedangkan metode *K-Medoids* juga menghasilkan 3 cluster dengan C1 dengan 11 jenis penyakit, C2 menghasilkan 13 jenis penyakit, dan C3 menghasilkan 60 jenis penyakit. Evaluasi hasil dilakukan menggunakan *Silhouette Coefficient*. Berdasarkan pengujian validitas cluster *Agglomerative Hierarchical Clustering* menggunakan *Silhouette Coefficient*, algoritma *Agglomerative Hierarchical Clustering* memperlihatkan kinerja lebih baik dengan rerata nilai 0,5837. Sedangkan pengujian validitas cluster *K-Medoids* menggunakan *Silhouette Coefficient* pada seluruh data, diperoleh nilai rerata sejumlah -0.3558. Nilai ini memperlihatkan bahwa hasil kualitas cluster yang kurang optimal. Perbedaan hasil klasterisasi antara *AHC* dan *K-Medoids* terjadi karena kedua algoritma memiliki cara kerja yang berbeda. *AHC* membentuk cluster secara bertahap dengan menggabungkan data yang paling dekat satu per satu hingga membentuk struktur hierarki menggunakan jarak *Manhattan*, sedangkan *K-Medoids* langsung membagi data ke dalam jumlah cluster di awal dengan menggunakan jarak *Euclidean*. Perbedaan ini memengaruhi jumlah dan susunan anggota pada tiap cluster.

Kata kunci: Klasterisasi, *Agglomerative Hierarchical Clustering*, *K-Medoids*, *Silhouette Coefficient*.

ABSTRACT

Arun Hospital Lhokseumawe has diverse and growing data on the types of diseases of inpatients, but it has not been optimally utilized for further analysis. So far, the data has only been used as administrative reports without being processed to obtain meaningful information, such as disease dominance patterns. This study aims to group the types of diseases of inpatients using two clustering methods, namely Agglomerative Hierarchical Clustering and K-Medoids, and analyze their comparative performance. The data used includes 84 types of diseases recorded from December 2024 to January 2025, with attributes of the number of male patients, female patients, and patient age. Agglomerative Hierarchical Clustering clustering was carried out using the average linkage approach and Manhattan Distance, while K-Medoids used the Euclidean Distance. The results show that the Agglomerative Hierarchical Clustering method formed 3 clusters, namely C1 with 4 types of diseases, C2 produced 79 types of diseases, and C3 produced 1 type of disease. Meanwhile, the K-Medoids method also produced 3 clusters with C1 containing 11 types of diseases, C2 producing 13 types of diseases, and C3 producing 60 types of diseases. Evaluation of the results was carried out using the Silhouette Coefficient. Based on the validity test of the Agglomerative Hierarchical Clustering cluster using the Silhouette Coefficient, the Agglomerative Hierarchical Clustering algorithm showed better performance with an average value of 0.5837. Meanwhile, the validity test of the K-Medoids cluster using the Silhouette Coefficient on all data obtained an average value of -0.3558. This value indicates that the cluster quality results are less than optimal. The difference in clustering results between AHC and K-Medoids occurs because the two algorithms operate differently. AHC forms clusters gradually by combining the closest data points one by one to form a hierarchical structure using the Manhattan distance, while K-Medoids immediately divides the data into clusters at the outset using the Euclidean distance. This difference affects the number and arrangement of members in each cluster.

Keywords: Clustering, Agglomerative Hierarchical Clustering, K-Medoids, Silhouette Coefficient.