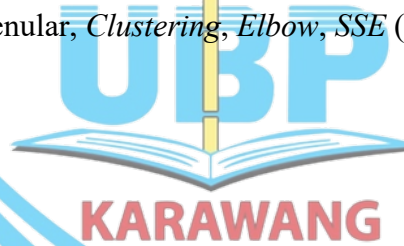


ABSTRAK

Penyakit menular merupakan penyakit yang bisa ditularkan (berpindah dari orang yang satu ke orang yang lain, baik secara langsung ataupun lewat perantara). Jumlah kasus penyakit menular pada tahun 2020 mengalami kenaikan yang cukup signifikan. *Clustering* merupakan salah satu esensi *Data Mining* untuk mengelompokkan data. Terdapat beberapa Algoritma yang dapat digunakan dalam *clustering* diantaranya Algoritma *K-Means*. Penelitian ini bertujuan untuk menerapkan Algoritma *K-Means Clustering* dalam mengelompokkan data penyakit menular pada manusia di Kabupaten Karawang untuk mengetahui puskesmas mana saja yang termasuk penyakit menular tertinggi, sedang, dan rendah. Dari hasil penelitian Algoritma *K-Means Clustering* dalam melakukan pengelompokkan diperoleh 1 data puskesmas pada *cluster* 1, 4 data puskesmas pada *cluster* 2, 45 data puskesmas pada *cluster* 3 dari 50 data puskesmas. Adapun hasil evaluasi menggunakan metode *Elbow* untuk mencari nilai K terbaik, dimana hasil K=3 merupakan jumlah *cluster* terbaik. Kemudian hasil tersebut dievaluasi menggunakan *SSE (Sum of Square Error)*. Hasil evaluasi menggunakan *SSE (Sum of Square Error)* tingkat akurasi yang didapat, yaitu 1.949. Nilai jarak dari 2 *cluster* ke 3 *cluster* tersebut merupakan nilai yang mengalami penurunan paling signifikan atau paling besar sehingga 3 *cluster* merupakan *cluster* terbaik.

Kata Kunci: Penyakit menular, *Clustering*, *Elbow*, *SSE (Sum of Square Error)*.



ABSTRACT

Contagious diseases are diseases transmitted from one person to another, either directly or through intermediaries. The number of infectious disease cases in 2020 experienced a significant increase. Clustering is one of the essences of Data Mining to group data. Several algorithms can be used in clustering including the K-Means Algorithm. This study aimed to apply the K-Means Clustering Algorithm in grouping data on contagious diseases in humans in Karawang Regency to find out which society health centres have the highest, moderate, and low infectious diseases. From the results of the K-Means Clustering Algorithm research in conducting grouping, it was obtained 1 data from society health centres in cluster 1, 4 data from health centres in cluster 2, 45 data from society health centres in cluster 3 of 50 health centres data. The results of the evaluation employed the Elbow method to find the best K value, where the result of $K = 3$ was the best number of clusters. Then the results were evaluated operating SSE (Sum of Square Error). The results of the evaluation utilising SSE (Sum of Square Error) the level of accuracy obtained is 1,949. The value of the distance from 2 clusters to 3 clusters was the value that has the most significant or largest decrease so that 3 clusters was the best cluster.

Keyword: *Infectious diseases, Clustering, Elbow, SSE (Sum of Square Error).*

