

## ABSTRAK

Pada saat ini perkembangan teknologi terjadi sangat pesat dimana teknologi baru semakin banyak bermunculan. Dampak dari perkembangan teknologi tersebut menyebabkan banyaknya kebiasaan, bisnis ataupun proses yang hilang dan digantikan oleh kemudahan yang diberikan oleh teknologi baru. Salah satu dari teknologi itu adalah teknologi informasi yaitu suatu proses yang berhubungan dengan pengolahan data untuk mendapatkan informasi, sehingga informasi ini dapat digunakan untuk mengambil keputusan yang lebih baik. Teknik pengumpulan data yang sering digunakan adalah *data mining*. Pada *data mining* terdapat beberapa metode yang dapat digunakan diantaranya *clustering*, metode ini digunakan untuk mengelompokkan data ke dalam *cluster-cluster* tertentu yang memiliki kemiripan. Salah satu algoritma pada *clustering* yang sering digunakan adalah *k-means* dimana metode ini memiliki tingkat keakuratan yang tinggi dan proses komputasi yang lumayan cepat namun memiliki kekurangan yaitu hasil *clustering* yang didapatkan bergantung dari seberapa baik *centroid* awal dan jumlah *cluster* yang digunakan. Oleh karena itu, diperlukan optimasi pada algoritma *k-means* untuk menghasilkan jumlah *cluster* dan *centroid* yang optimal.

Tahapan pertama dari penelitian ini adalah pengumpulan data yang akan digunakan sebagai bahan dari penelitian. Kemudian dilanjutkan dengan *pre-processing* data agar dapat digunakan dalam proses pengujian. Pada tahapan ini *pre-processing* yang dilakukan diantaranya *cleaning*, *transformation* dan *indexing*. Setelah data di *pre-processing* selesai, kemudian akan dilakukan *clustering* menggunakan metode *elbow* dan *centroid* proses *clustering* sebelumnya pada algoritma *k-means*.

Dalam penelitian ini, optimasi algoritma *k-means* dilakukan dengan cara menentukan jumlah *cluster* optimal menggunakan metode *elbow* dan menentukan *centroid* berdasarkan hasil *clustering* sebelumnya. Penelitian ini diukur dengan menggunakan *silhouette coefficient* untuk mengukur seberapa baik hasil *clustering*. Hasil penelitian ini menunjukkan bahwa hasil *clustering* menggunakan model pada penelitian ini lebih baik yaitu 0.5915 dengan jumlah iterasi 2, sedangkan percobaan *clustering* menggunakan *centroid* acak dan jumlah *cluster* yang berbeda menghasilkan *silhouette coefficient* antara 0.3693 hingga 0.5718 dengan jumlah iterasi 4 hingga 16.

**Kata kunci :** *Data Mining, Clustering, K-Means, Centroid, Silhouette Coefficient*

## ABSTRACT

At this time the development of technology is happening very rapidly where more and more new technologies are emerging. The impact of these technological developments has caused many habits, businesses or processes to disappear and be replaced by the convenience provided by new technology. One of these technologies is information technology, which is a process related to data processing to obtain information, so that this information can be used to make better decisions. The data collection technique that is often used is data mining. In data mining there are several methods that can be used including clustering, this method is used to group data into certain clusters that have similarities. One of the clustering algorithms that is often used is k-means where this method has a high level of accuracy and a fairly fast computational process but has a drawback, namely the clustering results obtained depend on how good the initial centroid is and the number of clusters used. Therefore, it is necessary to optimize the k-means algorithm to produce the optimal number of clusters and centroids.

The first stage of this research is data collection that will be used as research material. Then proceed with pre-processing the data so that it can be used in the testing process. At this stage, pre-processing is carried out including cleaning, transformation and indexing. After the data pre-processing is complete, clustering will then be carried out using the elbow and centroid methods of the previous clustering process in the k-means algorithm.

In this study, the optimization of the k-means algorithm was carried out by determining the optimal number of clusters using the elbow method and determining the centroid based on the previous clustering results. This research is measured using the silhouette coefficient to measure how good the clustering results are. The results of this study indicate that the results of clustering using the model in this study are better, namely 0.5915 with the number of iterations 2, while the clustering experiments using random centroids and different numbers of clusters produce a Silhouette coefficient between 0.3693 to 0.5718 with a number of iterations 4 to 16.

**Keywords:** *Data Mining, Clustering, K-Means, Centroid, Silhouette Coefficient*