

ABSTRAK

Nama : R. Nisfiani Aghisna
NIM : 1157010054
Judul : Analisis Penentuan Jumlah *Cluster* Terbaik pada Algoritma *K-Means* dan *K-Medoids*

Perkembangan teknologi telah menciptakan era dimana setiap sektor kehidupan menyimpan data dalam bentuk digital. Sebagian besar data tersebut disimpan dalam bentuk teks yang tidak terstruktur. Algoritma *clustering K-Means* dan *K-Medoids* merupakan salah satu metode yang dapat digunakan untuk mengatasi permasalahan ini. Pada penelitian ini digunakan algoritma *clustering K-Means* dan *K-Medoids* yang dipadukan dengan dua *proximity measure* yaitu *cosine distance* dan *euclidean distance*. Salah satu masalah utama yang harus diperhatikan dalam proses *clustering* adalah dimensi ruang fitur yang tinggi. Reduksi dimensi dapat dijadikan sebagai salah satu langkah optimasi pada algoritma *clustering* untuk mengurangi jumlah fitur (dimensi). Pada penelitian ini akan digunakan metode *Principal Component Analysis* (PCA) untuk mereduksi fitur-fitur yang kurang berpengaruh dan *redundant* tanpa mengurangi karakteristik data tersebut secara signifikan. Selanjutnya dilakukan perbandingan hasil *clustering* pada data terjemahan hadis yang berjumlah 2175 hadis menggunakan algoritma *K-Means* dan *K-Medoids* dengan masing-masing kombinasi *proximity measure* tanpa reduksi dan menggunakan reduksi PCA untuk mengetahui hasil *clustering* yang terbaik pada algoritma tersebut. Hasil *clustering* diperoleh dengan melakukan evaluasi menggunakan nilai *Silhouette Coefficient* (SC).

Kata Kunci: *K-Means, K-Medoids, Proximity Measure, Principal Component Analysis, Elbow Method, Silhouette Coefficient*

ABSTRACT

Name : R. Nisfiani Aghisna

NIM : 1157010054

Title : *Analysis of Determining the Best Number of Clusters in the K-Means and K-Medoids Algorithm*

Technological developments have created an era where every sector of life stores data in digital form. Most of the data is stored in the form of unstructured text. The K-Means and K-Medoids clustering algorithm is one method that can be used to overcome this problem. In this study, the K-Means and K-Medoids clustering algorithms are used combined with two proximity measures, namely cosine distance and euclidean distance. One of the main problems that must be considered in the clustering process is the high dimension of the feature space. Dimensional reduction can be used as an optimization step in the clustering algorithm to reduce the number of features (dimensions). In this study, the Principal Component Analysis (PCA) method will be used to reduce the features that are less influential and redundant without significantly reducing the characteristics of the data. Furthermore, a comparison of the results of clustering on the translation data of 2175 hadiths was carried out using the K-Means and K-Medoids algorithms with each combination of proximity measures without reduction and using PCA reduction to find out the best clustering results in the algorithm. The results of clustering are obtained by evaluating using the Silhouette Coefficient (SC) values.

Keywords: *K-Means, K-Medoids, Proximity Measure, Principal Component Analysis, Silhouette Coefficient*