

## ABSTRAK

**Nama** : Nanda Priatna  
**NIM** : 1147010048  
**Judul** : Analisis Perbandingan Algoritma *Clustering K-Means* dan DBSCAN untuk Data Teks Terjemah Hadits Menggunakan Ekstraksi Ciri *Hybrid*

Hasil *clustering* sangat dipengaruhi oleh metode *clustering* yang digunakan. Terdapat banyak sekali metode *clustering* yang sudah diperkenalkan dalam berbagai literatur. Pada penelitian ini, penulis akan membandingkan dua metode *clustering* yaitu metode *k-means* dengan metode DBSCAN untuk mengetahui karakteristik dari metode *k-Means* dan DBSCAN. Selain pemilihan metode *clustering*, dimensi ruang fitur yang tinggi juga merupakan salah satu masalah utama yang harus diperhatikan dalam proses *clustering*. Oleh karena itu, diperlukan adanya proses reduksi dimensi untuk memilih atau menyeleksi subfitur yang informatif. Pada penelitian ini, akan digunakan metode reduksi *hybrid* dengan memadukan metode *feature selection* dan metode *feature extraction* untuk memilih atau menyeleksi subfitur yang informatif. Keefektifan metode yang diajukan diuji pada dataset berupa terjemahan hadits dalam bahasa Inggris yang berjumlah 892 hadits. Kemudian, melakukan evaluasi *clustering* dengan menggunakan metode Davies-Bouldin Index (DBI) dan *silhouette coefficient*. Hasil uji coba menunjukkan bahwa metode *k-means* mempunyai keunggulan untuk mengatasi data yang mempunyai perbedaan kerapatan yang signifikan. Artinya, tidak akan ada data yang menjadi derau/*noise* sehingga semua data masuk ke dalam *cluster*. Pada algoritma DBSCAN, hal ini tidak dapat dilakukan. Metode DBSCAN sangat bagus digunakan untuk mengelompokkan data-data dengan kerapatan yang tinggi. Sedangkan *k-means* hanya menemukan klaster yang terbentuk saja tanpa memperhatikan kerapatan data klaster. DBSCAN juga mempunyai keunggulan untuk mendeteksi derau/*noise*. Kemudian, dengan diterapkannya metode reduksi yang memadukan metode *feature selection* dan metode *feature extraction* pada proses *text clustering* dapat meningkatkan akurasi *clustering*.

**Kata Kunci:** *K-Means*, DBSCAN, *feature selection*, *feature extraction*, dan PCA.

## **ABSTRACT**

**Name** : Nanda Priatna  
**NIM** : 1147010048  
**Title** : *Comparison Analysis Algorithm Clustering K-Means and DBSCAN for Translated Hadith Text Data using Extraction Hybrid Characteristic*

*Result of clustering is very influenced by clusterin method that used. There are a lot clustering method that introduced in various literature. In this research, the writer will compare two clustering method that is k-means method with DBSCAN method to know the characteristic from k-means method and DBSCAN. Besides clustering method choice, feature dimensional space also one of the main problem which must be considered in clustering process. Because of that, there a need of dimension reduction proses to choose or selecting subfeature that informative. In this research, hybrid reduction method used by integrating feature selection method and feature extraction method to choose or selecting subfeature that informative. Effectiveness of method that submitted tested in dataset in the form of hadith translation in english which amount to 892 hadith, after that, do clustering evaluation using davies-bouldin index (DBI) method and silhoutte coefficient. Test result show that k-means method have superiority to overcome data that have significant difference in density. That means, there is no data that become noise so every data enter into cluster. In DBSCAN algorithm, this could not be done, DBSCAN method is good in using for classify data with high density. While k-means used only to find cluster that formed without see density of cluster data. DBSCAN also have superiority to detect noise. Afterward, with applied reduction method that combine feature selection method and feature extraction method in text clustering process can increse clustering accuration.*

**Keywords:** *K-Means, DBSCAN, feature selection, feature extraction, dan PCA.*