

## ABSTRAK

Hadits merupakan sumber hukum agama Islam kedua setelah Al-Qur'an, di dalam hadits banyak bab yang membahas beberapa kasus dan akan menarik untuk dipadukan dengan teknik data mining khususnya klasterisasi untuk mengelompokkan hadits ke dalam beberapa kelompok berdasarkan *Matan* (isi hadits) secara otomatis. *Clustering* merupakan teknik pengelompokan data berdasarkan kriteria data, dalam *clustering* memiliki beberapa metode diantaranya *K-Means*. Penelitian ini akan mencoba mengelompokkan teks Hadits terjemahan bahasa Indonesia menggunakan algoritma *K-Means* dengan beberapa parameter dan eksperimen yang ditentukan. Penelitian ini digunakan untuk mengetahui bagaimana dan seberapa akurat kinerja algoritma *K-means* dalam pengelompokan Hadits sehingga dapat dikatakan layak untuk digunakan di kehidupan sehari-hari atau tidak. Hasil penelitian ini menunjukkan bahwa beberapa parameter yang digunakan mempengaruhi hasil evaluasi cluster, terutama pada *centroid* 10 dan 15, selain itu pemilihan data dan kata dalam hadits juga sangat mempengaruhi hasil yang didapat. Pada perhitungan *Confusion Matrix*, *K-means* memiliki akurasi sebesar 87% pada 124 data latih dan menggunakan data uji sebanyak 31 data. Dengan hasil di atas menunjukkan bahwa kinerja dan tingkat akurasi metode *K-means* baik dalam pengelompokan teks hadits bahasa Indonesia.

**Kata kunci – Algoritma K-means, Clustering, Data Mining, Hadits.**



## ABSTRACT

Hadith is the second source of Islamic religious law after the Qur'an, in hadith there are many chapters that discuss several cases and it would be interesting to combine it with data mining techniques, especially clustering to group hadiths into groups based on Matan (hadith content) automatically. Clustering is a data grouping technique based on data criteria, in clustering there are several methods including K-Means. This study will try to classify Indonesian translation of Hadith texts using the K-Means algorithm with several parameters and experiments determined. This research is used to find out how and how accurately the K-means algorithm performs in grouping Hadith so that it can be said that it is feasible to use in everyday life or not. The results of this study indicate that several parameters used affect the results of cluster evaluation, especially at centroids 10 and 15, besides that the selection of data and words in the hadith also greatly influences the results obtained. In the Confusion Matrix calculation, K -means has an accuracy of 0.87 at 90% of the training data of 124 training data and uses 10% of the test data of 31 data. The results above show that the performance and accuracy of the K-means method are good in grouping Indonesian hadith texts.

**Keywords – Clustering, Data Mining, Hadith, K-Means Algorithm.**

