

ABSTRACT

PERBANDINGAN ANALISIS ALGORITMA K-MEANS DAN FUZZY C-MEANS UNTUK PENGELOMPOKAN HADITS TERJEMAHAN BAHASA INDONESIA

RIZKY SAM PRATAMA – NIM 1147050146

Jurusan Teknik Informatika

Hadith is the second source of Islamic religious law after the Al-Qur'an scriptures, in the hadith there are many chapters that discuss several cases. The number of chapters in the hadith will be an interesting thing to be combined with data mining techniques, especially text mining in order to group the hadith into several groups based on the contents of the hadith (matan) automatically. Clustering is a grouping technique, there are several methods that can be used for grouping including K-Means and Fuzzy C-Means. From previous research, research is rarely found on hadith clustering based on matan. As for some applications regarding hadith that regulate the grouping of hadiths still manually through database relations, they have not been automated using certain algorithms. This research will try to grouping Indonesian translation of Hadith texts and compare K-Means and Fuzzy C-Means algorithms with some parameters and experiments that are determined. This comparison is used to determine the most accurate method in the Hadith clustering. The results of this study indicate that some of the parameters used affect the results of cluster evaluation, especially in reducing data dimensions. In Silhouette Coefficient and F-Measure calculations, the Fuzzy C-Means method has an accuracy of 0.83079 and 0.97128 while the K-Means method has an accuracy of 0.67828 and 0.95078 with the results above show that the Fuzzy C-Means method is superior in grouping the Indonesian hadith text.

Key Word: *k-means, fuzzy c-means, compare, clustering, hadith, text, indonesian, python, PCA*



ABSTRAK

PERBANDINGAN ANALISIS ALGORITMA K-MEANS DAN FUZZY C-MEANS UNTUK PENGELOMPOKAN HADITS TERJEMAHAN BAHASA INDONESIA

RIZKY SAM PRATAMA – NIM 1147050146

Jurusan Teknik Informatika

Hadits merupakan sumber hukum agama Islam yang kedua setelah kitab suci Al-Qur'an, dalam hadits terdapat banyak bab yang membahas beberapa perkara. Banyaknya bab dalam hadits akan menjadi hal yang menarik untuk dipadukan dengan teknik *data mining* terutama *text mining* guna mengelompokkan hadits kedalam beberapa kelompok berdasarkan isi hadits (matan) secara otomatis. Klusterisasi merupakan sebuah teknik pengelompokan, terdapat beberapa metode yang dapat digunakan untuk pengelompokan diantaranya K-Means dan Fuzzy C-Means. Dari penelitian sebelumnya, masih jarang ditemukan penelitian mengenai klusterisasi hadits berdasarkan matan. Adapun beberapa aplikasi mengenai hadits yang mengatur pengelompokan hadits masih secara manual melalui relasi *database*, belum ter-otomatisasi menggunakan algoritma tertentu. Penelitian sekarang akan mencoba mengelompokkan teks Hadits terjemahan Bahasa Indonesia serta membandingkan algoritma K-Means dan Fuzzy C-Means dengan beberapa parameter dan percobaan yang ditentukan. Perbandingan ini digunakan untuk menentukan metode yang paling akurat dalam klusterisasi hadits. Hasil penelitian ini menunjukkan bahwa beberapa parameter yang digunakan berpengaruh terhadap hasil evaluasi klaster, terutama pada pengurangan dimensi data. Secara perhitungan *Silhouette Coefficient* dan *F-Measure*, metode Fuzzy C-Means mendapat akurasi sebesar 0.83079 dan 0.97128 sedangkan metode K-Means mendapat akurasi sebesar 0.67828 dan 0.95078 dengan hasil diatas menunjukkan metode Fuzzy C-Means lebih unggul dalam pengelompokan teks hadits terjemahan Bahasa Indonesia.

Kata Kunci : k-means, fuzzy c-means, perbandingan, klusterisasi, hadits, teks, indonesia, python, PCA