

**PERBANDINGAN *FUZZY C-MEANS* DAN *K-MEANS* UNTUK
MENGELOMPOKKAN TINGKAT BUTA HURUF BERDASARKAN
PROVINSI DI INDONESIA**

Murtadlo Anugerah Pamungkas¹, Hardian Oktavianto², Reni Umilasari³

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

E-mail: pamungkasmurtadlo@gmail.com¹

ABSTRAK

Bekal dasar dalam menempuh pendidikan adalah memiliki kemampuan menulis, membaca, berkomunikasi dan berhitung secara efisien. Buta huruf merupakan salah satu permasalahan dalam menempuh pendidikan dan menghambat proses pencapaian tujuan peningkatan kualitas pendidikan, sedangkan yang dimaksud buta huruf adalah ketidakmampuan seseorang dalam membaca atau menulis kalimat sederhana. Berdasarkan penjelasan di atas terdapat metode yang dapat menyelesaikan permasalahan tersebut secara kompleks melalui salah satu cabang ilmu komputer seperti *data mining* dengan menggunakan metode *clustering*. Beberapa algoritma yang menggunakan metode *clustering* adalah *Fuzzy C-Means* dan *K-Means*, kedua algoritma tersebut mengelompokkan data ke dalam bentuk satu atau lebih *cluster*. Sehingga data yang memiliki karakteristik sama akan dikelompokkan ke dalam *cluster* yang sama dan data yang mempunyai karakteristik berbeda akan dikelompokkan ke dalam *cluster* yang lain. Untuk *cluster* optimum dalam setiap algoritma menggunakan metode *Dunn Index* dengan ditunjukkan nilai *Dunn Index* yang semakin besar. Data yang digunakan yaitu tingkat buta huruf di 33 provinsi di Indonesia periode tahun 2015-2019. Dari serangkaian pengujian dengan skenario 3 *cluster* sampai 10 *cluster*, menghasilkan *cluster* optimum pada algoritma *Fuzzy C-Means* yang berada pada 4 *cluster* dan untuk algoritma *K-Means* berada pada 6 *cluster*. Setelah menentukan *cluster* optimum dilakukan perbandingan menggunakan metode *Partition Coefficient Index* dan *Silhouette Index* dengan nilai validitas 0,7684 untuk algoritma *Fuzzy C-Means* dan nilai validitas 0,4966 untuk algoritma *K-Means*. Hasil penelitian menunjukkan bahwa algoritma *Fuzzy C-Means* lebih baik dalam melakukan *clustering* dibandingkan algoritma *K-Means*, karena nilai validitas algoritma *Fuzzy C-Means* lebih besar dan lebih mendekati nilai 1.

Kata Kunci: Buta Huruf, *Clustering*, Perbandingan, *Fuzzy C-Means*, *K-Means*, *Dunn Index*, *Partition Coefficient Index*, *Silhouette Index*.

**PERBANDINGAN *FUZZY C-MEANS* DAN *K-MEANS* UNTUK
MENGELOMPOKKAN TINGKAT BUTA HURUF BERDASARKAN
PROVINSI DI INDONESIA**

Murtadlo Anugerah Pamungkas¹, Hardian Oktavianto², Reni Umilasari³

Program Studi Teknik Informatika, Fakultas Teknik

Universitas Muhammadiyah Jember

E-mail: pamungkasmurtadlo@gmail.com¹

ABSTRACT

Basic provisions in education are to have the ability to write, read, communicate and count efficiently. Illiteracy is one of the problems in taking education and hinders the process of achieving the goal of improving the quality of education, while what is meant by illiteracy is the inability of a person to read or write simple sentences. Based on the explanation above, there are methods that can solve these problems in a complex manner through a branch of computer science such as data mining using the clustering method. Several algorithms that use the clustering method are Fuzzy C-Means and K-Means, both algorithms classify data into one or more clusters. So that data that has the same characteristics will be grouped into the same cluster and data that has different characteristics will be grouped into other clusters. For the optimum cluster in each algorithm, the Dunn Index method is used, with the value of the Dunn Index getting bigger. The data used is the illiteracy rate in 33 provinces in Indonesia for the period 2015-2019. From a series of tests with a scenario of 3 clusters to 10 clusters, the optimum cluster results in the Fuzzy C-Means algorithm which are in 4 clusters and for the K-Means algorithm are in 6 clusters. After determining the optimum cluster, comparisons are made using the Partition Coefficient Index and Silhouette Index methods with a validity value of 0.7684 for the Fuzzy C-Means algorithm and a validity value of 0.4966 for the K-Means algorithm. The results showed that the Fuzzy C-Means algorithm is better at clustering than the K-Means algorithm, because the validity value of the Fuzzy C-Means algorithm is greater and closer to the value of 1.

Keywords: Buta Huruf, *Clustering*, Perbandingan, *Fuzzy C-Means*, *K-Means*, *Dunn Index*, *Partition Coefficient Index*, *Silhouette Index*.