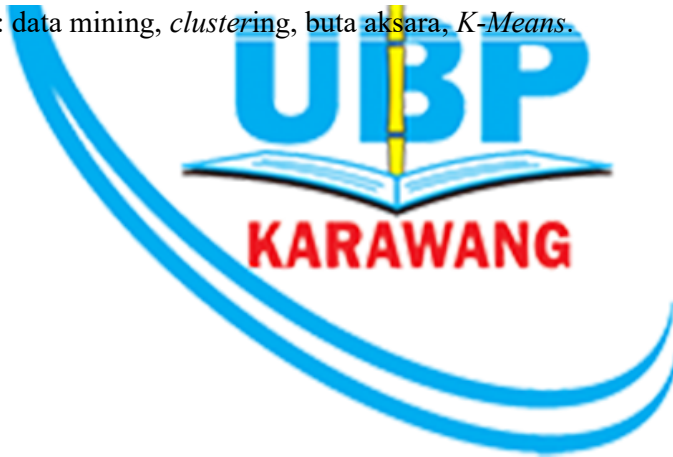


ABSTRAK

Buta Aksara merupakan ketidakmampuan seseorang dalam membaca dan menulis kalimat sederhana dalam bahasa apapun. Hampir setiap provinsi di Indonesia terdapat penduduk yang menderita buta aksara, terbukti dengan data terakhir dari BPS tahun 2017 sebanyak 3.387.035 jiwa penduduk umur 15-59 tahun. Untuk itu perlunya analisis mengenai penyebaran Buta aksara di Indonesia. Tujuan dari penelitian ini untuk melakukan pengelompokan provinsi penderita buta aksara dengan memanfaatkan Teknik *clustering* dalam proses data mining. Metode yang digunakan adalah algoritma K-Means. Algoritma K-Means merupakan salah satu algoritma yang mengelompokkan data yang sama pada kelompok tertentu dan data yang berbeda pada kelompok yang lain. Dalam hal ini mengelompokkan provinsi kedalam beberapa *cluster* berdasarkan kemiripan data, sehingga data provinsi dengan kesamaan karakteristik dikelompokkan menjadi satu *cluster* dan yang memiliki karakteristik yang berbeda akan dikelompokkan dalam *cluster* yang lain. Atribut yang dipakai adalah Provinsi, Umur 15+, umur 15-45, umur 45+. Hasil dari penelitian ini adalah pengelompokan provinsi dengan buta aksara Tertinggi, Sedang dan Terrendah.

Kata Kunci: data mining, *clustering*, buta aksara, *K-Means*.



ABSTRACT

Illiteracy is a person's inability to read and write simple sentences in any language. Almost every province in Indonesia has a population suffering from illiteracy, as evidenced by the latest data from BPS in 2017 as many as 3,387,035 people aged 15-59 years. For this reason, it is necessary to analyze the spread of illiteracy in Indonesia. The purpose of this study is to do provincial grouping of illiterate patients by utilizing clustering techniques in the data mining process. The method used is the K-Means algorithm. K-Means algorithm is one algorithm that groups the same data in certain groups and different data in other groups. In this case grouping provinces into several clusters based on similarity of data, so that provincial data with similarity characteristics are grouped into one cluster and those that have different characteristics will be grouped in another cluster. The attributes used are Province, Age 15+, Ages 15-45, Ages 45+. The results of this study is grouping the provinces with the highest, moderate and lowest literacy.

Keywords: *data mining, clustering, illiteracy, K-Means.*



