

## Diversity-based attribute weighting for k-modes clustering

M. Misbachul Huda, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20448143&lokasi=lokal>

---

### Abstrak

Categorical data is a kind of data that is used for computational in computer science. To obtain the information from categorical data input, it needs a clustering algorithm. There are so many clustering algorithms that are given by the researchers. One of the clustering algorithms for categorical data is k-modes. K-modes uses a simple matching approach. This simple matching approach uses similarity values. In K-modes, the two similar objects have similarity value 1, and 0 if it is otherwise. Actually, in each attribute, there are some kinds of different attribute value and each kind of attribute value has different number. The similarity value 0 and 1 is not enough to represent the real semantic distance between a data object and a cluster. Thus in this paper, we generalize a k-modes algorithm for categorical data by adding the weight and diversity value of each attribute value to optimize categorical data clustering.

<hr>Data Kategorial merupakan suatu jenis data perhitungan di ilmu komputer. Untuk mendapatkan informasi dari input data kategorial diperlukan algoritma klastering. Ada berbagai jenis algoritma klastering yang dikembangkan peneliti terdahulu. Salah satunya adalah K-modes. K-modes menggunakan pendekatan simple matching. Pendekatan simple matching ini menggunakan nilai similarity. Pada K-modes, jika dua objek data mirip, maka akan diberi nilai 1. Jika dua objek data tidak mirip, maka diberi nilai 0. Pada kenyataannya, tiap atribut data terdiri dari beberapa jenis nilai atribut dan tiap jenis nilai atribut terdiri dari jumlah yang berbeda. Nilai similarity 0 dan 1 kurang merepresentasi jarak antara sebuah objek data dan klaster secara nyata. Oleh karena itu, pada paper ini, kami mengembangkan algoritma K-modes untuk data kategorial dengan penambahan bobot dan nilai diversity pada setiap atribut untuk mengoptimalkan klastering data kategorial.