

Abstraksi

Klasterisasi adalah proses mengelompokkan data ke dalam suatu kelas atau klaster, sehingga objek pada suatu klaster memiliki kemiripan yang sangat besar dengan objek lain pada klaster yang sama, tetapi sangat tidak mirip dengan objek pada klaster lain.

Salah satu algoritma yang sering digunakan untuk melakukan proses *clustering* data adalah algoritma k-means. K-means sangat populer dalam proses klasterisasi data karena efisiensinya dalam mengklaster data. Namun, algoritma ini hanya terbatas untuk pengelompokan pada data numerik, sedangkan pada kenyataannya di dunia nyata banyak juga data yang atributnya bernilai kategorik.

Untuk menangani masalah data kategorik, dalam Tugas Akhir ini akan dibahas sebuah algoritma bernama k-modes yang merupakan varian dari algoritma k-means. Sama halnya seperti algoritma k-means, algoritma k-modes ini menghasilkan solusi *local optimum*. Hal tersebut berkaitan dengan proses inisialisasi pada penentuan *centroid* awal klaster. Dalam tugas akhir ini dibahas mengenai metode penentuan inisialisasi awal pada algoritma k-modes yaitu, secara random, dan menggunakan metode *frequency based*.

Ditunjukkan dalam tugas akhir ini bahwa metode pemilihan k inisialisasi awal menggunakan metode *frequency based* memiliki tingkat akurasi yang lebih baik dalam mengelompokkan data dibandingkan dengan inisialisasi secara random.

Kata kunci: *Clustering, k-means, k-modes, frequency based*