

# A Cluster Centres Initialization Method for Clustering Categorical Data Using Genetic Algorithm

Kusha Bhatt  
[kusha.bhatt@gmail.com](mailto:kusha.bhatt@gmail.com)

Prof. Pankaj Dalal  
[pkjdalal@gmail.com](mailto:pkjdalal@gmail.com)

Prof Avinash Panwar  
[avinashpanwar@gmail.com](mailto:avinashpanwar@gmail.com)

*Abstract:* The leading partitioned clustering technique, k-modes, is among the most computationally efficient clustering methods for categorical data. However, the k-modes clustering algorithm performance, which converges to numerous local minima strongly depends on initial cluster centres. Currently, most methods of initialization cluster centres are mainly for numerical data. Due to lack of geometry for the categorical data, these methods used in cluster centres initialization for numerical data are not applicable to categorical data. This research proposes a novel initialization method for categorical data which is implemented to the k-modes algorithm using genetic algorithm. The method integrates the distance and the density together to select initial cluster centres and overcomes shortcomings of the existing initialization methods for categorical data. Genetic algorithm is used here to optimize cluster centre initialization in traditional K-mode algorithm in order to find best results.

*Keywords:* Clustering, k-modes, Genetic Algorithm