Abstract

K-means clustering is a method of grouping data by looking for similarities between attributes possessed by data points and can overcome high data dimensions because of the simplicity of the algorithms it has. The disadvantage of the k-means method is that the initial centroid initialization will affect the end result of clustering and is very susceptible to outlier data because it will affect computational time. This study combines the huffman tree initialization and k-means to overcome the weaknesses of data grouping in k-means. This study uses 120 students data results taken from the results of try out activities conducted at one of the vocational high schools in Semarang City. The experiment aims to classify data based on the similarity of attributes possessed by the same data. Testing is done by measuring the level of accuracy of the expected results with the results of clustering. The results of this study indicate the highest accuracy value in cluster 1 with a value of 92% with an average value of 67% accuracy in all clusters.

References

Index Terms

Computer Science  Artificial Intelligence

Keywords

Information system, clustering, huffman tree, k-Means clustering, Educational Data Mining