

## ABSTRACT

Fuzzy C-Means (FCM) algorithm is one of the popular clustering methods. In this algorithm, data can be members of more than one cluster, whose existence is determined by the degree of membership. The FCM algorithm has a good performance, especially for handling large amounts of data and overlapping data set. However, the FCM algorithm has a weakness, which is sensitive to the initialization of clustering parameters, such as initialization of the membership degree, the initial cluster center, and number of clusters.

This study attempt to overcome these issues by modifying the initialization of the initial membership degree and fixing the cluster center in the first iteration. The initial membership degree is determined based on the placement of each data with several data points designated as an initial cluster center. Because the degree of the membership is incorrect, then in the first iteration, the cluster center needs to be repaired. The modified FCM algorithm is named as Modified Fuzzy C-Means (MFCM). This study also conducted cluster validation to look for optimal number of clusters so that clustering results are even better. Cluster validation methods used in this thesis are Xie-Beni (XB), Partition Coefficient (PC), and Partition Entropy (PE).

This study uses sales data of 49 products for 25 months. The number of clusters tested are 2, 3, 4, 5, and 6. The results of this study are that the MFCM method has a better performance than FCM. MFCM requires fewer iterations than FCM. In addition, MFCM has a smaller XB and PE value and a PC value bigger than FCM. In the data used, based on the values of XB, PC, and PE, cluster 2 is the optimal number of clusters.

**Keywords** : clustering, fuzzy c-means, cluster validation, xie-beni index, industrial product

## INTISARI

Algoritme Fuzzy C-Means (FCM) merupakan salah satu metode *clustering* yang populer digunakan. Pada algoritme ini, data dapat menjadi anggota lebih dari satu kelompok (*cluster*), yang keberadaannya ditentukan oleh derajat keanggotaan. Algoritme FCM memiliki performa yang cukup baik, terutama untuk menangani data yang banyak dan overlap. Namun algoritme FCM memiliki kelemahan, yaitu sensitif terhadap inisialisasi parameter-parameter *clustering* yang digunakan, seperti penentuan derajat keanggotaan, pusat *cluster* awal, dan cacah *cluster* yang digunakan.

Penelitian ini mencoba mengatasi isu-isu tersebut dengan memodifikasi penentuan derajat keanggotaan awal dan memperbaiki pusat *cluster* pada iterasi pertama. Derajat keanggotaan awal ditentukan berdasarkan kedekatan setiap data dengan beberapa titik data yang ditunjuk sebagai pusat *cluster* awal. Karena derajat keanggotaan tersebut belum tepat, maka pada iterasi pertama pusat *cluster* perlu diperbaiki. Algoritme FCM yang telah dimodifikasi diberi nama *Modified Fuzzy C-Means* (MFCM). Penelitian ini juga melakukan validasi *cluster* untuk mencari cacah *cluster* optimal agar hasil *clustering* lebih baik lagi. Metode validasi *cluster* yang digunakan dalam tesis ini adalah indeks Xie-Beni (XB), *Partition Coefficient* (PC), dan *Partition Entropy* (PE).

Penelitian ini menggunakan data penjualan 49 produk kerajinan selama 25 bulan. Cacah *cluster* yang diuji adalah 2, 3, 4, 5, dan 6. Hasil penelitian ini adalah metode MFCM memiliki kinerja yang lebih baik daripada FCM. MFCM memerlukan iterasi yang lebih sedikit daripada FCM. Selain itu, MFCM memiliki nilai XB dan PE yang lebih kecil dan nilai PC yang lebih besar daripada FCM. Pada data yang digunakan, berdasarkan nilai XB, PC, dan PE, cacah *cluster* 2 merupakan cacah *cluster* yang optimal.

**Kata kunci** – *clustering, fuzzy c-means, validasi cluster, indeks xie-beni, produk industri.*