



RESEARCH ARTICLE

Comparational Analysis of K-Means and Fuzzy. Algorithm C-Means in Clasterization of Poor Population Data at Rawa Makmur

Muhammad Ilham ^{a,1*}; Dewi Suranti ^{a,2}; Jhoanne Fredricka ^{a,3}

Published online: 1 February 2022

Abstract

Rawa Makmur Village is one of the villages located in Muara Bangkahulu District, Bengkulu City. So far, the process of processing data on the poor at the Rawa Makmur Village Office is still done manually, namely by conducting a survey of each resident, and providing an assessment of whether the population is categorized as poor or not. However, because the data collection process is still manual, it is difficult for the Rawa Makmur Village to manage the data because it takes a long time. Therefore, a system development was carried out by creating applications that could simplify the process of managing data for the poor in Rawa Makmur Village. The application for clustering data for the poor in Rawa Makmur Village was made using the Visual Basic .Net programming language and SQL Server 2008 database by applying two data mining methods, namely the K-Means Algorithm and the Fuzzy C-Means Algorithm. The grouping is done based on the income and the number of the insured from the data obtained at the Rawa Makmur Village Office. Based on the processing time, the Fuzzy C-Means Algorithm is faster than the K-Means Algorithm in the data clustering process because the K-Means Algorithm has repeated iterations until the final cluster result is found while the Fuzzy C-Means does not have iterative iterations. Based on the results of clusters on the data of the poor as many as 440 people, there are differences in cluster results where the K-Means Algorithm has Cluster 1 (65 people), Cluster 2 (334 people), Cluster 3 (41 people), while the Fuzzy C-Means Algorithm has Clusters. 1 (147 people), Cluster 2 (136 people), Cluster 3 (157 people). From the results of the comparative analysis between the K-Means Algorithm and the Fuzzy C-Means Algorithm, it is found that the K-Means Algorithm is more recommended for classifying poor population data than the Fuzzy C-Means Algorithm because the average value of each cluster does not differ too much, and is still significant according to with data on the poor.

Keyword: Analysis; Comparison; K-Means Algorithm; Fuzzy C-Means Algorithm; Poor Population Data; Kelurahan Rawa Makmur

Introduction

Data mining is a process that uses statistical techniques, mathematics, artificial intelligence, machine learning to extract and identify useful information and related knowledge from large databases. Data mining technology is one of the tools for extracting data in large databases and with specifications of the level of complexity that has been widely used in many application domains such as banking, telecommunications, and government.

Rawa Makmur Village is one of the villages located in Muara Bangkahulu District, Bengkulu City. So far, the process of processing data on the poor at the Rawa Makmur Village Office is still done manually, namely by conducting a survey of each resident, and providing an assessment of whether the population is categorized as poor or not. After collecting data manually, the data is inputted into the office package application, namely Microsoft Word for making reports for the poor in Rawa Makmur Village. This is necessary to provide possible assistance to the poor in the scope of Kelurahan Rawa Makmur. However, because the data collection process is still manual, it is difficult for the Rawa Makmur Village to manage the data because it takes a long time.

Therefore, a system development was carried out by creating applications that could simplify the process of managing data for the poor in Rawa Makmur Village. In addition, the application applies 2 Data Mining Methods, namely the K-Means Algorithm and the Fuzzy C-Means

¹⁻³ Information Study Program, Faculty Of Computer Science, Dehasen University, Bengkulu

**) corresponding author*

Muhammad Ilham

Information Study Program, Faculty Of Computer Science, Dehasen University, Bengkulu

Email: muhammadilham98@gmail.com

Algorithm which are used to classify the data of the poor population into 3 groups/clusters, so that it can make it easier for the Kelurahan to get information on who is in the high, medium group, and low.

The K-Means Algorithm and the Fuzzy C-Means Algorithm have something in common, namely they can group data into several clusters by looking at the closest point in each cluster. However, in practice, these two algorithms have different calculations in the clustering process. Therefore, in this study the authors are interested in conducting an analysis by comparing the K-Means algorithm and the Fuzzy C-Means algorithm in the clustering process by looking at several aspects, namely the results of data grouping and the time of the clustering process that occurs.

Related research on comparative analysis of K-Means and Fuzzy C-Means algorithms, including a research journal by (Agustina & Prihandoko, 2018) entitled "Comparison of K-Means Algorithm with Fuzzy C-Means Algorithm for Clustering Employee Performance Discipline Levels". Based on the research conducted, the two methods classify employee performance data into 3 clusters, namely high-performance levels, medium performance levels and low performance levels. The results of this study indicate that the Fuzzy C-Means method is a better method than the K-Means for clustering data on the performance level of employees at STT Bandung because the validation value is close to 1.

Ease of Use

A. Data Mining

Data mining is an iterative and interactive process to find new patterns or models that are perfect, useful and understandable in a very large database (massive database). Data mining contains the search for desired trends or patterns in large databases to help decision makers in the future, these patterns are recognized by certain tools that can provide a useful and insightful data analysis which can then be studied more carefully, which may be using other decision support tools (Sikumbang, 2018:156).

Data mining as a process to obtain useful information from large database warehouses, which can be interpreted as extracting new information taken from large chunks of data that helps decision making. Data mining can find hidden trends and patterns that do not appear in simple query analysis so that they can have an important part in finding knowledge and making decisions (Wanto, 2020:1).

B. K-Means method

The K-Means algorithm is one of the clustering algorithms included in the unsupervised learning group which is used to divide data into several groups with a partition system. This algorithm accepts input in the form of data without class labels. In the K-Means algorithm, the computer groups its own data into its input without first knowing the target class. The inputs received are data or objects and the desired k clusters. This algorithm will group data or objects into k groups (Wanto, 2020: 3).

Basically the K-Means algorithm only takes part of the number of components obtained to be used as the initial cluster center, in determining the center of this cluster randomly selected from the data population. Then the K-Means algorithm will test each of the components in the data population and mark these components into one of the previously defined cluster centers depending on the minimum distance between components and each cluster center. Then the position of the center of the cluster will be recalculated until all data components are classified into each cluster and finally a new cluster will be formed.

C. Fuzzy C-Means (FCM) Algorithm

Fuzzy logic was first introduced by Prof. Lotfi A. Zadeh in 1965. In many ways, fuzzy logic is used as a way to map problems from input to expected output. In fuzzy logic there is fuzzy clustering which is one method to determine the optimal cluster in a vector space based on the Euclidian normal form for the distance between vectors. The purpose of Fuzzy C-Means is to get the center of the cluster which will later be used to find out the data that enters a cluster. In fuzzy logic there is a method that is often used to cluster data, namely the Fuzzy C-Means method. Fuzzy C-Means is a data clustering method that is determined by the degree of membership (Ramadhan, 2017:221).

Fuzzy C-Means is a data clustering technique in which the existence of each data in a cluster is determined by a certain membership value/degree. This technique was first introduced by Jim Bezdek in 1981. Unlike the classical clustering technique (where an object will only be a member of a certain cluster), in FCM each data can be a member of several clusters. Cluster boundaries in Fuzzy C-Means are soft.

Prepare Your Paper Before Styling

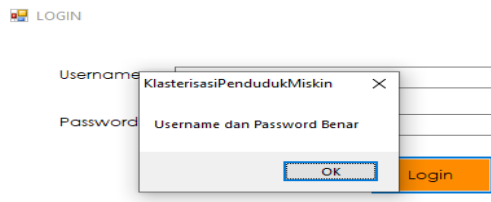
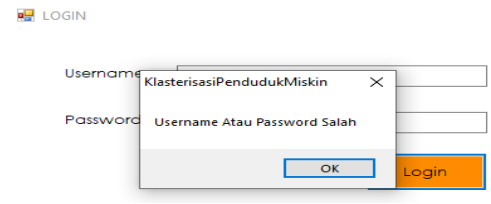
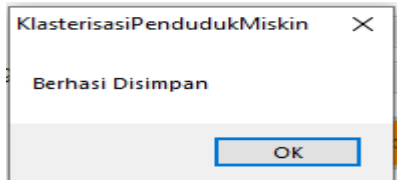
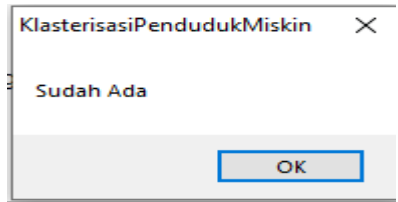
The type of research used by the author is the type of experimental research. The research was conducted by examining the data of the poor population in Rawa Makmur Village using the K-Means and Fuzzy C-Means Algorithms, then validation of the results obtained.

The research method used by the author is the Waterfall method. The waterfall method is often called the classical life cycle, which describes a systematic and sequential approach to software development, starting with the specification of user requirements and then continuing through the stages of planning, modeling, construction, (construction), and delivery of the system to customers/users (deployment), which ends with support for the complete software produced.

Using the Template

Blackbox testing (blackbox testing) is one of the software testing methods that focuses on the functionality side, especially on the input and output of data clustering applications for the poor in Rawa Makmur Village. The results of the black box testing that have been carried out are shown in table 1.

TABLE I. Blackbox Test Results

No.	Tested Form	Testing Scenario	Test result
1	Login Form	fill in the correct username and password	<p>the system has successfully received login access and displays a success message</p> 
		fill in the wrong username and password	<p>the system denies login access by displaying an error message</p> 
2	Population Data Form	enter population data that is not in the database	<p>the system successfully saves population data and displays a success message</p> 
		enter population data that is already in the database	<p>the system failed to save resident data and displays an error message</p> 
3	K-Means Clustering Form	carry out the population clustering process and record the results of clustering	<p>the system successfully displays the results of the data clustering of the poor using the K-Means Algorithm</p>

4	Fuzzy C-Means Clustering Form	carry out the population clustering process and record the results of clustering	<p>the system has succeeded in displaying the results of clustering poor population data using the Fuzzy C-Means Algorithm</p>

Based on the results of the tests that have been carried out, the functional application of clustering data for the poor in Rawa Makmur Village has run as expected and the application is able to provide information on the results of grouping the poor through 3 clusters, namely Cluster 1 (High), Cluster 2 (Medium) and Cluster 3 (High). Low).

Acknowledgment

1. Application of data clustering for the poor in Rawa Makmur Village was made using Visual Basic .Net programming language and SQL Server 2008 database by applying two data mining methods, namely the K-Means Algorithm and the Fuzzy C-Means Algorithm.
2. Based on processing time, the Fuzzy C-Means Algorithm is faster than the K-Means Algorithm in the data clustering process because the K-Means Algorithm has repeated iterations until the final result is found, while the Fuzzy C-Means does not have iterative iterations.
3. Based on the results of clusters of data on the poor as many as 440 people, there are differences in cluster results where the K-Means Algorithm has Cluster 1 (65 people), Cluster 2 (334 people), Cluster 3 (41 people), while the Fuzzy C-Means Algorithm have Cluster 1 (147 people), Cluster 2 (136 people), Cluster 3 (157 people)

Supplementary Material

Supplementary material that may be helpful in the review process should be prepared and provided as a separate electronic file. That file can then be transformed into PDF format and submitted along with the manuscript and graphic files to the appropriate editorial office.

References

- [1] Agustina, N. & Prihandoko, 2018. Perbandingan Algoritma K-Means Dengan Algoritma Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan. *Jurnal Resti (Rekayasa Sistem dan Teknologi Informasi)*, Volume Vol.2 No.3 ISSN:2580-0760.
- [2] Blazing, A., 2018. *Pemrograman Windows Dengan Visual Basic .Net : Praktikum Pemrograman VB.Net*. s.l.:Google Book.
- [3] Chusyairi, A. & Saputra, P. R. N., 2019. Perbandingan Algoritma Fuzzy C-Means dan K-Means Clustering Dalam Pengelompokan Data Puskesmas. *Jurnal Citisee*, Volume ISBN:978-602-60280-1-3.
- [4] Firman, A., 2019. *Analisis dan Perancangan Sistem Informasi*. Surabaya: Penerbit Qiara Media.
- [5] Pamungkas, C. A., 2017. *Pengantar dan Implementasi Basis Data*. Yogyakarta: Penerbit Deepublish.
- [6] Rahakbauw, 2017. Implementasi Fuzzy C-Means Clustering Dalam Penentuan Beasiswa. *Jurnal Ilmu Matematika dan Terapan*, Volume Vol.11 No.1.
- [7] Ramadhan, A., 2017. Perbandingan K-Means dan Fuzzy C-Means Untuk Pengelompokan Data User Knowledge Modeling. Pekanbaru, Seminar Nasional Teknologi Informasi, Komunikasi, dan Industri (SNTIKI) Fakultas Sains dan Teknologi UIN Sultan Syarif Kasim Riau Pekanbaru ISSN 2579-7271.
- [8] Santoso & Nurmalina, R., 2017. Perencanaan dan Pengembangan Aplikasi Absensi Mahasiswa Menggunakan Smart Card Guna Pengembangan Kampus Cerdas (Studi Kasus Politeknik Negeri Tanah Laut). *Jurnal Integrasi*, Volume Vol.9 No.1 . E-ISSN: 2548-9828.
- [9] Sikumbang, E. D., 2018. Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori. *Jurnal Teknik Komputer*, Volume Vol.4 No.1.
- [10] Suprpto, U., 2021. *Pemodelan Perangkat Lunak (C3) Kompetensi Keahlian: Rekayasa Perangkat Lunak Untuk SMK/MAK Kelas XI*. Jakarta: Grasindo.
- [11] Wanto, A. et al., 2020. *Data Mining: Algoritma Dan Implementasi*. Medan: Yayasan Kita Menulis.

