Indonesian Journal of Computer Science
ISSN 2549-7286 (*online*)
Jln. Khatib Sulaiman Dalam No. 1, Padang, Indonesia
Website: ijcs.stmikindonesia.ac.id | E-mail: ijcs@stmikindonesia.ac.id

Classification Of Potential Patients Using the Fuzzy K-Means Algorithm at Woman's and Children's Hospital

Ahmad Sahab[1], Hidayatus Sibyan[1], Nulngafan[1], Irnawati[2], Pawiliyah[3]
sahab1304@gmail.com, hsibyan@unsiq.ac.id, affan@unsiq.ac.id, sakinah.jogja@ymail.com, pawiliyah@yahoo.com
[1]Informatics Engineering, Faculty of Engineering and Computer Science, University of Sains Al Qur'an Wonosobo
[2] Department of Nursing, Faculty of Health Sciences, University of Muhammadiyah Pekajangan Pekalongan
[3]Department of Nursing, Bengkulu Tri Mandiri Sakti High School of Health Sciences

Article Information

Submitted : 22 Jul 2023
Reviewed: 14 Sep 2023
Accepted : 28 Oct 2023

Keywords

Patient, Promotion, Clustering, Fuzzy K-Means

Abstract

Adina Wonosobo Woman's and Children's Hopsital is one of the hospitals in the Wonosobo Regency. To increase the number of patient visits, promotion is carried out through broadcast messages to patients. In evaluating its implementation, problems were found in determining the patient list of recipients of promotional broadcast messages. This problem arises because the determination of the list of recipients of promotional messages is only done on the contact numbers stored on the marketing team's mobile phones. The Fuzzy K-Means method is a data grouping method that is used as a reference for decision-making. Patient visit data were obtained from the RSIA Adina Wonosobo information system database. Furthermore, the data was analyzed to determine the potential of digital marketing priority patients using the clustering method with the Fuzzy K-Means algorithm.

A. Introduction

At present the Mother and Child Hospital of Adina Wonosobo in an effort to increase the number of children's immunization visits, so it takes a promotion to increase understanding of immunization to patients in line with efforts to increase children's immunization visits in the Mother and Child Hospital of Adina Wonosobo. The use of the WhatsApp Blasting promotional media at the Adina Mother and Child Hospital in marketing activities is currently not running optimally. The promotion target is only directed to the patient's contact number stored on the health promotion telephone. There is no classification of patient contacts and not all patient numbers are stored in health promotion telephones causing the determination of promotion targets to be inaccurate. Adina Wonosobo Mother and Child Hospital has a patient database with more than 70,000 patient data. With this patient database, there is an opportunity to increase digital marketing targets. However, it is less efficient to promote all patient data simultaneously. Therefore, it is necessary to group and classify patients to obtain priority potential patient data so that digital marketing through WhatsApp Blasting can run effectively and efficiently[1].

Adina Wonosobo Mother and Child Hospital is currently in an effort to increase the number of child immunization patient visits, so that in this study an analysis of patient data will be carried out to determine potential priority patients for the promotion of child immunization services. The patient analysis will classify patient data so as to produce priority potential patient data as marketing targets. The process of separating data that has different characteristics and grouping data that has the same characteristics uses the clustering method[2].

The Fuzzy K-Means Algorithm is an algorithm that randomly determines the value of a cluster (k). The value that temporarily becomes the center of the cluster is often referred to as the centroid. Then calculate the distance from all existing data to each centroid using the Euclidean Distance formula until the closest distance to the centroid is found for all data until the centroid value does not change[3].

The Fuzzy K-Means algorithm is designed to make it easier to group data into different groups based on certain variables without having to go through a training process. This is because Fuzzy K-Means is a centroid-based unsupervised learning algorithm, where each cluster is associated with a centroid. The main goal of the Fuzzy K-Means algorithm is to minimize the sum of the distances between data points and their corresponding clusters[4].

The Fuzzy K-Means algorithm makes it possible to group patient data with the same characteristics to obtain priority potential patient data so that digital marketing via WhatsApp Blasting[5].

Cluster analysis is a multivariate analysis technique used to find information about variables and organize them so that they can be grouped into relatively homogeneous groups. Clusters are formed using the proximity method, which must be internally homogeneous (its members are similar to one another) and externally different (its members are not like other cluster members). Cluster analysis can accept different input data, often referred to as similarity measures, but can also be referred to as proximity and similarity. Some experts recommend using standardized data, clusters can be

calculated at different scales, and standardization provides measurements in different units[6].

B. Research Methods

This study uses a type of quantitative research. Quantitative research method is a type of research whose objectives are systematic, planned and clearly structured from the beginning to the creation of a research plan. Data collection is a way to obtain information about the cases that are the problem in this study. This study collected information about the services of the Adina Wonosobo Mother and Child Hospital, patient data, marketing targets and the method used in this case study, namely the Fuzzy K-Means algorithm.The object of research in this study were the characteristics of the target patients in marketing promotions, such as age, gender and history of patient visits collected from patient visit lists at Adina Wonosobo Mother and Child Hospital[7].

The data source used in this study was patient visit data at the Adina Wonosobo Women and Children Hospital for the period November 1 2022 to February 28 2023 which were taken on March 10 2023 with a total of 9262 visits This research data is used to promote child immunization services so that sample data from pediatric clinic visits with doctor code 2731XXX are taken as many as 346 visits[8].

A population of 346 visits obtained a sample of 263 visit data with a valid telephone number format. Furthermore, grouping of visits was carried out based on the patient. From 263 data visits with a valid telephone number format, 236 patient data were obtained. Then, from 236 patient data, patients with a maximum age of 24 months were selected and 189 patient data were obtained[7].

C. Result And Discussion

1. Data Processing

The method used in this study is the clustering method using the Fuzzy K-Means algorithm. The data that has been pre-processed will be processed from the data using the clustering method. This method works by grouping data that has the same characteristics in each data[4].

1. Determine the number of clusters, in this study it is divided into three clusters because the results of this study are expected to get three groups of criteria.
2. Determines the initial centroid value. In this study, the centroid value was determined in Table 1:

Table 1. Initial Centroid Value

Centroid		
C1	2	1
C2	9	2
C3	24	3

3. Determination of the next centroid value. In this study, determining the centroid value then uses the formula:

$$\frac{x_1 + x_2 + x_3 + \cdots x_n}{\Sigma_n}$$

Information :

x : Cluster member values
Σn : Number of cluster members

4. Termining the cluster value as a reference in calculating the distance of the object to the centroid, the distance calculation refers to the Euclidean Distance formula:

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Manual calculation example:
Patient 1 data to cluster center
Cluster 0

$$\sqrt{(24 - 2)^2 + (1 - 1)^2} = 22$$

Cluster 1

$$\sqrt{(24 - 9)^2 + (1 - 2)^2} = 15.033296378373$$

Cluster 2

$$\sqrt{(24 - 24)^2 + (1 - 3)^2} = 2$$

In this study the process stops at the 3rd iteration with the centroid value and the results of the number of each cluster member in each iteration in Tables 2, 3 and 4:

Table 2. Iteration 2 Centroid Value

Centroid		
C1	2.6194690265487	1.1769911504425
C2	10.298245614035	1.1052631578947
C3	19.789473684211	1.0526315789474

Table 3. Iteration 3 Centroid Value

Centroid		
C1	2.8429752066116	1.1652892561983
C2	10.555555555556	1.1333333333333
C3	19.130434782609	1.0434782608696

Table 4. Calculation Results for the Number of Cluster Members

Iteration	**C1**	**C2**	**C3**
1	113	57	19
	Priority 1	Priority 2	Priority 3

2	121	45	23
	Priority 1	Priority 2	Priority 3
3	121	43	25
	Priority 1	Priority 2	Priority 3

The results of the calculations can be concluded that potential patients as digital marketing targets are cluster 1 with a total of 121 patient cluster members[9].

2. System Implementation

Based on calculations using the Fuzzy K-Means Algorithm then implemented in software development using the php and MySQL programming languages[10]. The following are the results of software development in this study:

1. Initial Data Display Page

This page displays initial patient visit data from the database.

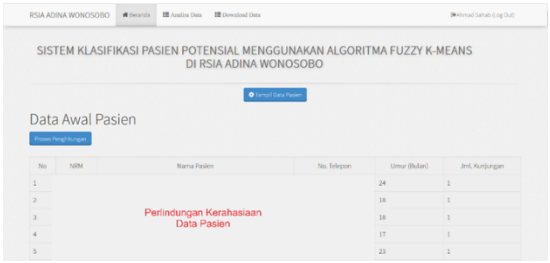

Figure 1. Initial Data Display Page

2. Data Clustering page

This page displays data from patient data clustering calculation results.

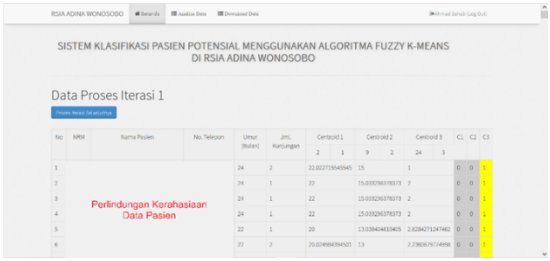

Figure 2. Data Clustering Page

3. Completed Data Clustering Process page

This page displays notifications when the data clustering process has been completed.

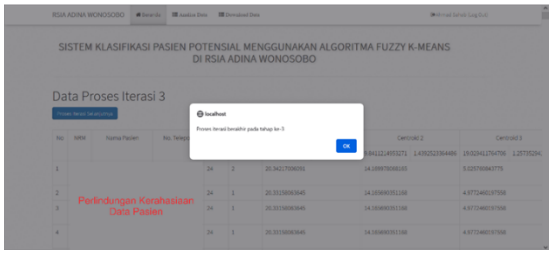

Figure 3. Completed Data Clustering Process Page

4. Clustering Result Data Download Page

This page displays notifications when the data clustering process has been completed.

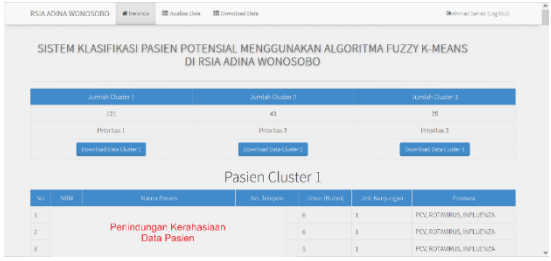

Figure 4. Result Data Download Page

3. Accuracy Level Testing

In testing the level of accuracy of the calculation results is done by comparison with manual calculations using the Euclidean Distance. The comparison results show an accuracy rate of 100% with the results in Table 5:

Table 5. Accuracy Level Testing

Iteration	System Results			Euclidean Distance		
	C1	C2	C3	C1	C2	C3
1	113	57	19	113	57	19
2	121	45	23	121	45	23
3	121	43	25	121	43	25

D. Conclusion

Based on the description of the analysis and discussion that has been carried out in the study of potential patient classification using the Fuzzy K-Means algorithm at the Adina Wonosobo Mother and Child Hospital, it can be concluded that the selection of priority potential patients can be done by clustering method using the Fuzzy K-Means algorithm. Analysis of 9262 patient visits at the Adina Wonosobo Women's and Children's Hospital took a sample of 346 patient data. After pre-processing the sample data, 189 valid patient data were obtained. The sample data is calculated by dividing it into 3 clusters, which then calculates the distance between the data and the centroid value using the Euclidean Distance

formula. when there is no movement of cluster members and results in cluster 1 as a priority potential patient for digital marketing targets with a total of 121 patients.

Suggestions from this study are that the next researcher can carry out a calculation process using other methods to be used as a comparison of the level of effectiveness in determining potential patients at the Adina Mother and Child Hospital Wonosobo. It needs direct testing in the implementation of digital marketing processes to increase patients reusing health services at the Adina Wonosobo Mother and Child Hospital.

E. References

[1] S. Puspitorini, R. W. Astuti, M. Jannah, and V. Putri, "ASSOCIATION RULE MINING UNTUK PROMO WHATSAPP BLAST PADA PRODUK ACE HARDWARE JAMBI PRIMA MALL," pp. 90–96, 1907.

[2] M. Sinan, J. Leng, K. Shah, and T. Abdeljawad, "Advances in numerical simulation with a clustering method based on K–means algorithm and Adams Bashforth scheme for fractional order laser chaotic system," *Alexandria Eng. J.*, vol. 75, pp. 165–179, 2023, doi: 10.1016/j.aej.2023.05.080.

[3] X. Wu, B. Wu, J. Sun, S. Qiu, and X. Li, "A hybrid fuzzy K-harmonic means clustering algorithm," *Appl. Math. Model.*, vol. 39, no. 12, pp. 3398–3409, 2015, doi: 10.1016/j.apm.2014.11.041.

[4] Q. Ren, H. zhang, D. Zhang, X. Zhao, L. Yan, and J. Rui, "A novel hybrid method of lithology identification based on k-means++ algorithm and fuzzy decision tree," *J. Pet. Sci. Eng.*, vol. 208, no. PD, p. 109681, 2022, doi: 10.1016/j.petrol.2021.109681.

[5] S. Toukola, M. Ståhle, and T. Mahlam, "Renaissance of project marketing : Avenues for the utilisation of digital tools," vol. 4, no. April, 2023, doi: 10.1016/j.plas.2023.100091.

[6] Á. López-Oriona, P. D'Urso, J. A. Vilar, and B. Lafuente-Rego, "Quantile-based fuzzy C-means clustering of multivariate time series: Robust techniques," *Int. J. Approx. Reason.*, vol. 150, pp. 55–82, 2022, doi: 10.1016/j.ijar.2022.07.010.

[7] L. Patricia, *Research Design Quantitative, Qualitative, Mixed Methods, Arts-Based, and Community-Based Participatory Research Approaches*. New York, USA, 2017.

[8] J. W. Creswell, *Research Design Qualitative and quantitative research methods*, 4th ed., no. 1. London: SAGE Publications, 2016.

[9] S. Methods *et al.*, "Sampling," 2015.

[10] S. Tummalapalli and V. R. Machavarapu, "Managing Mysql Cluster Data Using Cloudera Impala," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 463–474, 2016, doi: 10.1016/j.procs.2016.05.193.