GROUPING OF BOOKS TYPE BASED BY TIME BORROWING (MONTHS) USING FUZZY C-MEANS ALGORITHM. (Case Study: Surabaya Public Library)

[1]BACHTIAR AZHARI, [2]RIFKI FAHRIAL ZAENAL, [3]RANI PURBANINGTYAS

[123]Informatics Engineering of Bhayangkara University Surabaya

Jl. A. Yani 114 Surabaya 60231

Email: nindoujutsu@gmail.com

ABSTRACT

Book lending transactions are the main activities that take place in the Library. With the method of grouping, it will get what kind of books are most often borrowed. The method used is Fuzzy C-Means. This thesis discusses the application of Fuzzy C-Means method to classify book data based on lending per month and measure the effectiveness of the use of methods in the process. This application has been tested by producing 3 clusters and classify the time (month) where the frequency of borrowing done by visitors in the Library.
Keywords: *Cluster, Fuzzy C-means, Library*

1. INTRODUCTION

Surabaya City Library is known as one of the largest libraries in East Java. The cataloging system in the Surabaya City Library currently used to place the type / genre of books to be presented to readers still uses the Dewey Decimal Classification (DDC) classification. Which is where this classification system is used in many libraries around the world.

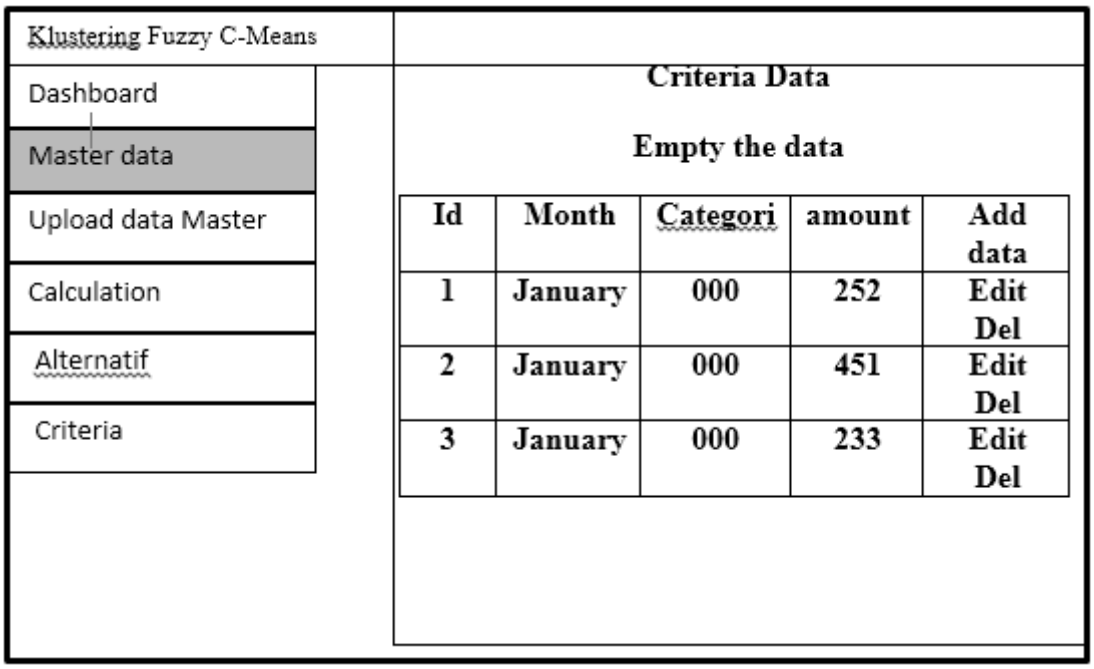
The Surabaya City Library located in Rungkut Asri Tengah still use manual calculations to determine the most borrowing time which takes monthly data in those years through the borrower's books, and begins to calculate the borrower specifically noted by the officers residing in the Library City of Surabaya. One of the clustering algorithms that uses the partition approach is the C-Means algorithm. This algorithm has advantages that are easy to implement and run, relatively fast, easy to adapt, and most widely practiced in data mining tasks. This algorithm is one of the most important algorithms in data mining. C-Means divides the data and then groups it into several clusters that have similarities and separate each cluster based on the differences between each cluster. This algorithm has been put forward by several researchers from different disciplines.

2. RESEARCH METHODS

This research uses C-Means Algorithm. C-Means is one of the algorithms found in the classification technique. This algorithm has advantages that are easy to implement and run, relatively fast, easy to adapt, and most widely practiced in data mining tasks. This algorithm is one of the most important algorithms in data mining. C-Means divides the data and then groups it into several clusters that have similarities and separate each cluster based on the differences between each cluster. This algorithm has been put forward by several researchers from different disciplines.

3. ANALYSIS AND DESIGN SYSTEM

On the main page there are several main menus visible in the sidebar including Dashboard, Master data, Clustering lending per year, and monthly clusters and charts. The dashboard contains the main / main features of this app. The master data menu contains the entire loan data from 2011 to 2014, the yearly clustering menu contains yearly calculations from 2011 to 2014, the master data upload menu, the calculation menu, the alternate menu, and the criteria menu.



Fuzzy clustering is the process of determining the degree of membership, and then using it by inserting it into the data elements into one cluster or more clusters. This will give information of the similarity of each object.

One of the many fuzzy clustering algorithms used is the fuzzy clustering c means algorithm. The vector of fuzzy clustering, V = {v1, v2, v3, ..., vc}, is an objective function that is defined by the degree of membership of data Xj and center cluster Vj. The fuzzy clustering algorithm c means divides the available data from each finite data element and then puts it into the part of the cluster collection that is affected by some given criteria. Please provide a finite set of data. X = {x1, ..., xn} and data centers.

$$J_m(X,U,V) = \sum_{j=l}^{n}\sum_{i=l}^{c}(\mu ij)^m\, d^2\left(X_J, V_i\right) \ldots\ldots\ldots\ldots\ldots..3.1$$

Fuzzy C-Means algorithm is as follows:

1. Input data to be clustered X, a matrix of size n x m (n = number of sample data, m = attribute of each data). Xij = i-th sample data (i = 1,2, ..., n), j-attribute (j = 1,2, ..., m).
2. Specify:
    1. Number of clusters = c
    2. Rank = w
    3. Maximum iteration = MaxIter
    4. The smallest expected error = ξ
    5. Initial objective function = Po = 0
    6. Initial iteration = t = 1
3. Generate random values μik, i = 1,2, ..., n; K = 1,2, ..., c as elements of the initial partition matrix. Mik is the degree of membership that refers to how likely a data can be a member into a cluster. The position and value of the matrix are constructed randomly. Where the value of membership lies in the interval 0 to 1. In the initial position of the partition matrix U is still not accurate as well as its cluster center. So the tendency of data to enter a cluster is also not accurate.

$$Q_i = \sum_{k=1}^{n} \mu_{ik} \dots\dots\dots\dots\dots\dots 3.2$$

$$V_{kj} = \frac{\sum_{i=l}^{n}((\mu_{ik})^w X_{ij})}{\sum_{i=l}^{n}((\mu_{ik})^w)} \dots\dots\dots\dots\dots 3.3$$

4. Calculate the center of the k-cluster: Vkj, with k = 1,2, ... c and j = 1,2, ... m. Where Xij is the fuzzy variable used and w is
5. weights. The objective function is used as a recurrence requirement to get the right cluster center. So that obtained the tendency of data to enter the cluster where in the final step.
6. Calculate the objective function on the t iteration, Pt

$$P_t = \sum_{i=l}^{n}\sum_{i=0}^{n}\left(\left[\sum_{j=l}^{m}\left(X_{ij} - V_{kj}\right)^2\right](\mu_{ik})^w\right) \dots\dots\dots\dots\dots 3.4$$

7. The calculation of the objective function Pt where the Xu fuzzy variable value is less with the Vkj cluster center then the result of the reduction in quadrading then each quadrade result is summed to multiply by the quadrade of the degree of membership for each cluster. Then add all values in all clusters to get the objective function of Pt.
8. Calculate the partition matrix change:

$$\mu_{ik} = \frac{\left[\sum_{j=l}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}}{\sum_{k=l}^{c}\left[\sum_{j=l}^{m}(X_{ij} - V_{kj})^2\right]^{\frac{-1}{w-1}}} \dots\dots\dots\dots\dots 3.5$$

With: i = 1,2, ... n and k = 1,2, .. c.

To find the change in the ik partition matrix, the subtraction of the Xuzhu fuzzy variable value to return to the Vkjlalu cluster center is squared. Then summed and then lifted with -1 / (w-1) with weight, w = 2 result each data is raised by -1. After the calculation process is done, normalize all new membership degree data by adding the new membership degree k = 1, ... c, the result then divided by the new membership degree. This process is done so that the new membership degree has a range between 0 and not more than 1

9. Check the stop condition: a) if :(| Pt - Pt-1 | <ξ) or (t> maxIter) then stop.b) otherwise t = t + 1, repeat step 4.

4. RESULT AND DISCUSSION

At the time of the process of FCM data library library borrowing with the parameters that have been determined that is, lending transactions for the classification 000 books from January 2011 to December 2014 the number of clusters = 3, rank (w) = 2, maximum iteration = 100 and the smallest error = 0.001. The FCM process stops at the 61th iteration with the clustering of 3 clusters:

Table 6.1 The last iteration result and cluster of classification books 000

Month	Criteria	Last membership:	Cluster 1 (Borrowing Rarely)	Cluster 2 (Most Frequent Loans)	Cluster 2 (Most Frequent Loans)
	000				

January 2011	252	0.03417 3499	0.0698849 19	0.8959415 82			*
January 2012	451	0.03202 0534	0.8837548 92	0.0842245 75		*	
January 2013	233	0.00997 0669	0.0123520 63	0.9776772 68			*
January 2014	325	0.03456 0756	0.7754189 83	0.1900202 6		*	
February 2011	455	0.03459 6646	0.8755699 61	0.0898333 94		*	
February 2012	170	0.17411 4647	0.0392864 81	0.7865988 72			*
February 2013	22	0.92430 6544	0.0173694 45	0.0583240 1	*		
February 2014	16	0.91094 4106	0.0208697 03	0.0681861 91	*		
March 2011	85	0.98638 8867	0.0022773 41	0.0113337 92	*		
March 2012	54	0.98668 6301	0.0026707 63	0.0106429 36	*		
Maret 2013	166	0.21146 0122	0.0423018 47	0.7462380 31			*
Maret 2014	110	0.86525 9571	0.0183379 18	0.1164025 11	*		
Apr-11	321	0.03911 3856	0.7362744 46	0.2246116 99		*	
Apr-12	251	0.03284 0354	0.0653824 53	0.9017771 94			*
Apr-13	233	0.00997 0669	0.0123520 63	0.9776772 68			*
Apr-14	325	0.03456 0756	0.7754189 83	0.1900202 6		*	
May 2011	355	0.00668 5913	0.9651274 74	0.0281866 13		*	

May 2012	140	0.52708 2767	0.0442022 43	0.4287149 9	*		
May 2013	321	0.03911 3856	0.7362744 46	0.2246116 99		*	
May 2014	114	0.83126 9692	0.0220575 26	0.1466727 83	*		
June 2011	285	0.06236 6504	0.3260146 3	0.6116188 66			*
June 2012	154	0.34508 5513	0.0472311 52	0.6076833 35			*
June 2013	166	0.21146 0122	0.0423018 47	0.7462380 31			*
June 2014	210	0.00226 2234	0.0015392 01	0.9961985 65			*
July 2011	325	0.03456 0756	0.7754189 83	0.1900202 6		*	
July 2012	345	0.01390 289	0.9227071 75	0.0633899 35		*	
July 2013	233	0.00997 0669	0.0123520 63	0.9776772 68			*
July 2014	325	0.03456 0756	0.7754189 83	0.1900202 6		*	
August 2011	455	0.03459 6646	0.8755699 61	0.0898333 94		*	
August 2012	170	0.17411 4647	0.0392864 81	0.7865988 72			*
August 2013	452	0.03266 3208	0.8817014 48	0.0856353 44		*	
August 2014	315	0.04572 8821	0.6726752 99	0.2815958 8		*	
Sep-11	258	0.04191 0519	0.1008052 27	0.8572842 54			*
Sep-12	351	0.00925 9546	0.9505025 71	0.0402378 83		*	
Sep-13	166	0.21146 0122	0.0423018 47	0.7462380 31			*

Sep-14	110	0.865259571	0.018337918	0.116402511	*		
October 2011	252	0.034173499	0.069884919	0.895941582			*
October 2012	451	0.032020534	0.883754892	0.084224575		*	
October 2013	233	0.009970669	0.012352063	0.977677268			*
October 2014	325	0.034560756	0.775418983	0.19002026		*	
Nov-11	455	0.034596646	0.875569961	0.089833394		*	
Nov-12	170	0.174114647	0.039286481	0.786598872			*
Nov-13	22	0.924306544	0.017369445	0.05832401	*		
Nov-14	16	0.910944106	0.020869703	0.068186191	*		
December 2011	85	0.986388867	0.002277341	0.011333792	*		
December 2012	54	0.986686301	0.002670763	0.010642936	*		
December 2013	166	0.211460122	0.042301847	0.746238031			*
December 2014	110	0.865259571	0.018337918	0.116402511	*		

From the table above can be concluded:

1. Cluster 1 group with lending labels is rare in months "February 2012", "February 2014", "March 2011", "March 2012", "March 2014", "May 2012", "May 2014", "Sep-14", "Nov-13", "Nov- 14 "," December 2011 "," December 2012 "," December 2014 "
2. Cluster 2 groups with the most frequent lending labels in the months "January 2012", "January 2014", "February 2011", "Apr-11", "Apr-14", "May 2011", "May 2013", "July 2011 "," July 2012 "," October 2014 "," August 2012 "," July 2014 "," August 2012 "," August 2014 "," August 2011 "

3. Group Cluster 3 with frequent lending labels in the months of January 2011, January 2013, February 2012, March 2013, Apr-12, Apr-13, June June, , "June 2013", "June 2014", "July 2013", "August 2012", "Sep-11", "Sep-13", "October 2011", "October 2013", "Nov-12", " December 2013 "

The calculation process is done by using the formula:

Precision : $\frac{True\ Positive}{True\ Positive+False\ Positive}$

Recall : $\frac{True\ Positive}{True\ Positive+False\ Negstive}$

5. CONCLUSION

1. Based on the discussion, it can be concluded that the loan data of each classification in 2011 to 2014 can be classified into several clusters by applying the Fuzzy C-Means method.
2. From the lending data of 2011 to 2014, obtained 3 clusters group which is divided based on the book lending.
3. Based on the calculation of precision and recall 3 (three) clusters it can be concluded:

   A. For the classification of 000 books as many as 13 data (27.08%) in cluster 1, 17 data (35.42%) in cluster 2, and 18 data (37.50%) in cluster 3
   B. For classification of 100 books 23 data (47.92%) in cluster 1, 19 data (39.58%) in cluster 2, and 6 data (12.50%) in cluster 3
   C. For classification of 200 books as much as 20 data (41.67%) in cluster 1, 14 data (29.17%) in cluster 2, and 14 data (29.17%) in cluster 3
   D. For classification of 300 books, 4 data (8.33%) in cluster 1, 37 data (77.08%) in cluster 2 and 7 data (14.58%) in cluster 3
   E. For the classification of 400 books as many as 26 data (54.17%) in cluster 1, 17 data (35.42%) in cluster 2, and 5 data (10.42%) in cluster 3
   F. For the classification of 500 books 28 data (58%) in cluster 1, 12 data (25%) in cluster 2 and 8 data (17%) in cluster 3
   G. For classification of 600 books, 12 data (25%) in cluster 1, 14 data (29%) in cluster 2, and 22 data (46%) in cluster 3
   H. For the classification of 700 books 19 data (40%) in cluster 1, 14 data (29%) in cluster 2 and 15 data (31%) in cluster 3
   I. For the classification of 800 books 14 data (29%) in cluster 1, 18 data (28%) in cluster 2, and 16 data (33%) in cluster 3
   J. For the 900 book classification of 18 data (38%) in cluster 1, 15 data (31%) in cluster 2, and 15 data (31%) in cluster 3
4. Clustering process in this research is conducted to determine the number of clusters formed from the beginning of the process in accordance with the number of groups (total borrowers) desired. Thus, it can not be determined exactly how many ideal clusters are formed from existing borrowing data, so the accuracy of grouping can not be measured
5. Clustering is the process of grouping unlabeled objects or data into a class or cluster with objects that have similarities. Clustering using Fuzzy C-Means method to lending data on each book classification that can bring up some data clusters that can be further analyzed equations and differences

REFERENCES

[1] Kusrini, 2006, Data Mining Algorithm, ANDI Publisher, Yogyakarta.
[2] Khoiruddin, (2007), Determining the Final Value of Lecture With Fuzzy C-Means, Proceeding on National Seminar on System and Informatics in Bali, Major Informatics Engineering, Islamic University of Indonesia, Yogyakarta.
[3] Kusumadewi, S., Purnomo, H., 2010, Fuzzy Application For Supporters Decision, Graha Ilmu, Jakarta.

[4] Yulia Yudihartanti, (2010), Interests of Final Task Topics Using Fuzzy C-Means Method, STMIK Banjarbaru, Kalimantan.
[5] Bahar (2011), Determination of Senior High School with Algorithm Fuzzy C-Means, Master of Informatics Engineering, Dian Nuswantoro University, Semarang.
[6] Yunita Permata, 2012, Classification of Pregnancy Risk with Fuzzy C-means method, Department of Informatics, Sebelas Maret University, Surakarta.
[7] Catur Sugeng Pribadi, 2012, Procurement Support System of Library Book Procurement STIKOM Surabaya Using Fuzzy C-means Clustering Method, Department of Information System, STIKOM, Surabaya.
[8] Aditya Nur Santoso, (2012), Decision support system procurement library books STIKOM surabaya using fuzzy c-means clustering method.
[9] Windy Prabowo Wicaksono, (2012), Identification of hand gestures as household appliance controller using fuzzy c-means method, University of Muhammadiyah Ponorogo, Ponorogo.
[10] Nurul Faujiyah, (2012), fuzzy c-mean implementation for the estimated stock of raw materials of tofu production, National Development University "Veteran", Surabaya.
[11] Bastiah, (2013), Decision Support System for determining home purchase using fuzzy c-means (fcm) clustering method, STMIK Budi Darma, Medan.
[12] Ari Irawan, (2014), Application of fuzzy c-means algorithm in determining scholarship, Muria University, Kudus.
[13] Aniq Noviciatie Ulfah, (2014), performance analysis of fuzzy c-means and k-means algorithm on poverty data, State Islamic University of Sunan Kalijaga, Yogyakarta.
[14] A.Hanifa Setianingrum, (2014), mapping model of graduate qualification evaluation based on fuzzy c_means clustering method, Syarif Hidayatullah State Islamic University, Jakarta.
[15] Maria Riana Megawati, (2015), Implementation of fuzzy c-means algorithm for high school grouping in diy based on national examination test and school examination, Sanata Dharma University, Yogyakarta.