Ilkogretim Online - Elementary Education Online, 2021; 20 (1): pp. 699-707
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.01.60

Unsupervised Data Mining Technique for Clustering Library in Indonesia

***Joseph Teguh Santoso,** *Universitas STEKOM, Semarang, Indonesia, joseph@stekom.ac.id*
Sri Jumini, *Universitas Sains Al-Qur'an, Wonosobo, Indonesia*
Gita Widi Bhawika, *Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia*
Daniel Susilo, *School of Communication Science, Universitas Multimedia Nusantara, Jakarta, Indonesia*
Danny Wibowo, *Sekolah Tinggi Ilmu Ekonomi Indonesia Surabaya, Surabaya, Indonesia*
*Corresponding Author

Abstract. Organizing school libraries not only keeps library materials, but helps students and teachers in completing tasks in the teaching process so that national development goals are in order to improve community welfare by producing quality and competitive human resources. The purpose of this study is to analyze the Unsupervised Learning technique in conducting cluster mapping of the number of libraries at education levels in Indonesia. The data source was obtained from the Ministry of Education and Culture which was processed by the Central Statistics Agency (abbreviated as BPS) with url: bps.go.id/. The data consisted of 34 records where the attribute used was the number of libraries at each level of education starting from Elementary School, Junior High School, Senior High School and Vocational High School. The Unsupervised Learning Technique used is the k-medoids method which is part of data mining. The mapping label used is the high cluster (K1) for the number of libraries and the low cluster (K2) for the number of libraries at each level of education. The analysis process uses the help of Rapid Miner software. The results of the study indicated that 3 provinces in the high cluster (K1) and 31 provinces in the low cluster (K2) were for elementary schools; 4 provinces in the high cluster (K1) and 30 provinces in the low cluster (K2) for Junior High Schools; 13 provinces in the high cluster (K1) and 21 provinces in the low cluster (K2) for Senior High Schools; 8 provinces in the high cluster (K1) and 26 provinces in the low cluster (K2) for Vocational High Schools. The cluster formed is the optimal school cluster tested with the parameters of the Davies Bouldin Index (DBI) with results of 0.176 (Elementary School), 0.284 (Junior High School), 0.780 (Senior High School) and 0.662 (Vocational High School). The results of the research are expected to provide information in increasing the number of libraries at each level of education so that students and teachers can take advantage of existing library learning resources at school.
Keywords: Data Mining, K-Medoids Method, Library, School, Unsupervised Learning, Indonesia.

Received: 09.11.2020 Accepted: 03.12.2020 Published: 05.01.2021

INTRODUCTION

In the aspect of the world of education, improving the quality of people's welfare and competitiveness is one of the efforts to achieve development goals [1]–[3]. This is because education is a forum for developing human resources and is expected to be able to educate and bring society, the nation and the state towards the formation of personality, skills, intelligence, self-control and religious spirituality [4], [5]. Therefore, one of the efforts to improve the quality of education is to utilize learning resources such as libraries in schools [6], [7]. According to the Education and Culture Regulation Number 23 of 2015 concerning Character Development which views that every school should be a comfortable place and provide inspiration for students, teachers, and or other educational personnel; In addition, character education must be a joint movement involving the local government, the community, and or parents. According to [8] the school library stated has functions such as an educational function, an informative function, a function of administrative responsibility, a research function, a recreational function that can lead to students' love of reading. The use of school libraries must also be supported by the quality of library managers, especially manners [9]. The purpose of the study was to map the cluster by utilizing Unsupervised Learning techniques in Indonesia based on educational levels. This needs to be done so that the government can prioritize the construction of school libraries at the levels of Elementary Schools, Junior High Schools, Senior High Schools and Vocational High Schools. This needs to be done considering the school library can lead to student love for reading.

Based on these problems, data mining can be a solution in mapping which is an Unsupervised Learning technique. The results displayed depend on the weight values collected at the beginning of the construction of a system of similar value in a certain space or area [10]–[12]. In other words, data mining is a learning method that is suitable for finding or mapping patterns in the form of clusters of many similar objects that are not entirely the same. A popular data mining method for clustering is k-medoids which are widely used in business, academia, and industry [13]. The following is an illustration of unsupervised learning techniques with clustering techniques as shown in the following figure:

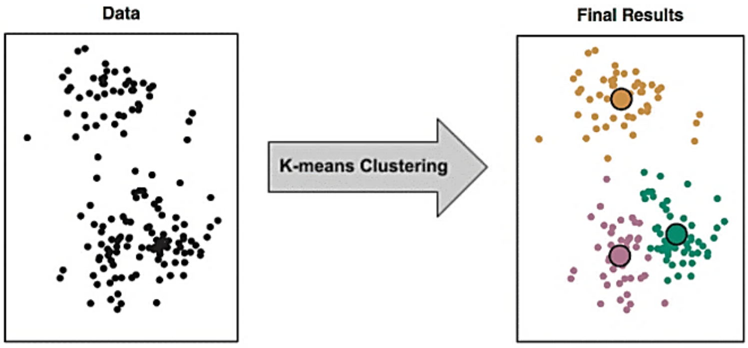


Figure 1. *Unsupervised Learning techniques in data mining clustering*

Several studies related to the k-medoids method in solving problems in terms of clustering such as [14]. This paper proposes the k-medoids method to measure the performance of the method by calculating the purity value of each resulting cluster. The results show that the k-medoids method is more suitable for use in the dataset with the encoded attribute format with the largest data modification value being 91.67%. Next research [15]. This paper proposes a comparison of the two most popular clustering methods k-means and k-medoids which are evaluated on the transaction 10k dataset from KEEL. The comparison results show that the k-medoids method is much better than the k-means method both in terms of execution time, is insensitive to outliers and reduces noise. It is hoped that the results of the research can provide information to the government considering that the role of the government as a facilitator for the establishment of libraries in each region must really be able to provide support for the current existence of libraries.

METHODOLOGY

Data Mining
Data determines a decision in the future, the results of data processed using data mining techniques can be done by generating new knowledge that comes from old data [16]. Examples of data mining processing can be in the form of classification, clustering, association and estimation [17]. Processing with clustering techniques is often done as the first step in the data mining process. There are many clustering methods that have been used by previous researchers such as K-Means, Improved K-Means, K-Medoids (PAM), Fuzzy C-Means [18].
K-Medoids Method
The k-means and k-medoids methods have differences. The k-medoids method uses the object as the representative cluster center (medoid) for each cluster, while the k-means method requires the mean value as the center of the cluster. group. In addition, the k-medoids method is more suitable for grouping data than the k-means method [10].

Data

The data source was obtained from the Ministry of Education and Culture which was processed by the Central Statistics Agency (abbreviated as BPS) with url: bps.go.id/. The data consisted of 34 records where the attribute used was the number of libraries at each level of education starting from Elementary School, Junior High School, Senior High School and Vocational High School. The data used is the number of libraries by province and education level for the 2018/2019 academic year. The following is the research data used in mapping the number of libraries by province in Indonesia.

Table 1. *Research data*

Province	Primary School	Junior High	High School	Vocational High School
Aceh	2765	957	450	157
North Sumatra	5898	2072	941	770
West Sumatra	2886	712	302	183
Riau	2162	861	384	190
Jambi	1705	541	203	141
South Sumatra	318	1047	530	218
Bengkulu	1057	357	127	81
Lampung	3000	1091	402	351
Kep Bangka Belitung	798	197	63	50
Kep Riau	688	270	124	76
DKI Jakarta	2003	1060	475	542
West Java	10958	3999	1296	1985
Central Java	13649	3107	836	135
DI Yogyakarta	1647	449	165	194
East Java	12642	3663	1229	1469
Banten	2612	1122	435	486
Bali	2033	392	146	143
West Nusa Tenggara	2347	663	249	165
East Nusa Tenggara	3476	1185	440	192
West Kalimantan	2951	1022	358	164
Central Kalimantan	1588	595	191	94
South Borneo	2049	572	177	95
East Kalimantan	1238	516	187	165
North Kalimantan	279	128	45	18
North Sulawesi	1500	634	203	140
Central Sulawesi	1864	588	189	121
South Sulawesi	5066	1349	523	337
Southeast Sulawesi	1683	630	260	110
Gorontalo	795	250	56	53
West Sulawesi	801	268	73	73
Maluku	1058	423	226	87
North Maluku	762	284	139	69
West Papua	415	220	95	38
Papua	780	425	163	91

Source: Rahim, 2020 [19]

RESULTS AND DISCUSSION

By utilizing Rapid Miner software in analyzing the k-medoids method, the model design was made using 2 cluster labels. The process of determining cluster labels is carried out by looking at the optimal value of the Davies Bouldin Index (DBI) parameter. The following is the comparison of the DBI values at each level of education for k = 2, 3 and 4 as shown in the following table and figure:

Table 2. *The results of the comparison of the DBI value at each level of education*

	K=2	K=3	K=4
Elementary School	0.176	0.425	2.172
Junior High School	0.284	0.417	0.647
Senior High School	0.780	1.691	0.361
Vocational High School	0.662	2.069	2.289

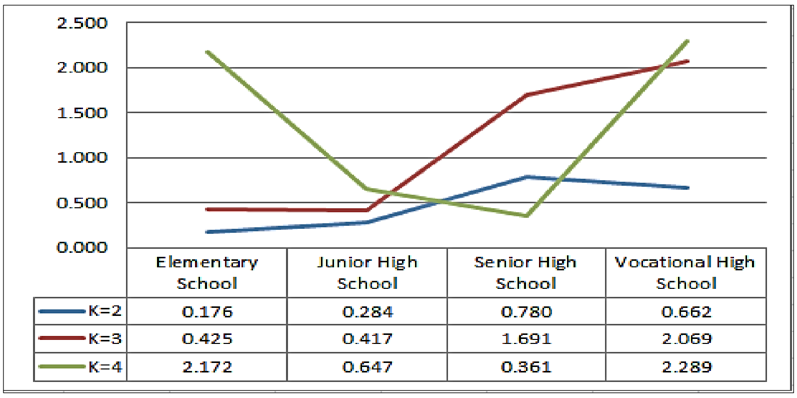


Figure 2. *Graph of the comparison of DBI scores at each level of education*

In Figure 2 it can be explained that the optimal DBI value for clustering is k = 2. Where each level has a small value compared to k = 3 and k = 4. This is the basis for the number of clusters created, namely the high cluster (K1) for the number of libraries and the low cluster (K2) for the number of libraries at each level of education. The following is a clustering model design using Rapid Miner.

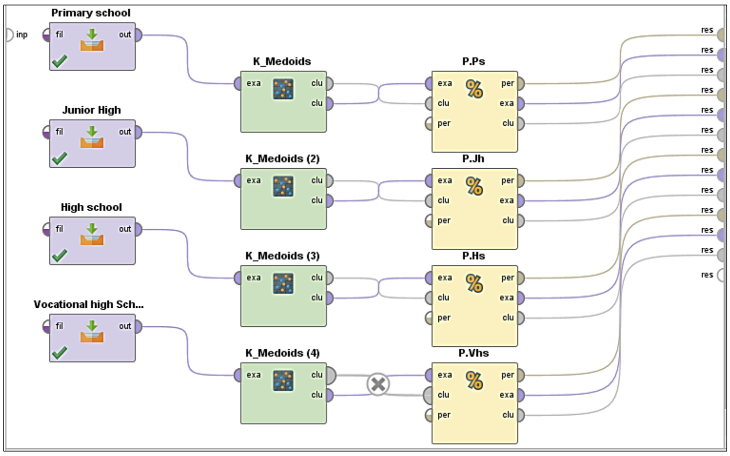
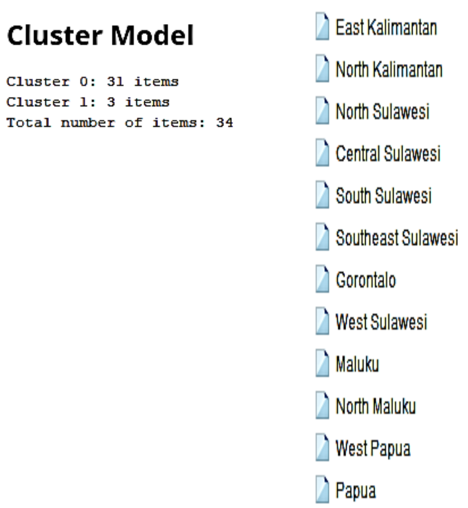


Figure 3. *The k-medoid model in the RapidMiner design*

In Figure 3, it is explained that the data input process is carried out using excel files for each level of education. The method of analysis uses the k-medoids method where the output uses a validity test with parameters from the Davies Bouldin Index (DBI). Following are the results of the mapping in the form of clusters from each level of education as shown in the following figure:



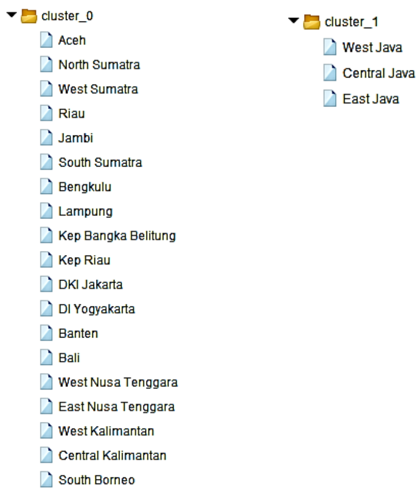


Figure 4. *Results of the cluster at the Primary School level*

In Figure 4 it can be explained that 3 provinces in the high cluster (K1 = cluster_1) and 31 provinces in the low cluster (K2 = cluster_0) for Elementary Schools with the scores from the final centroid for K1 = 12642 and K2 = 780.

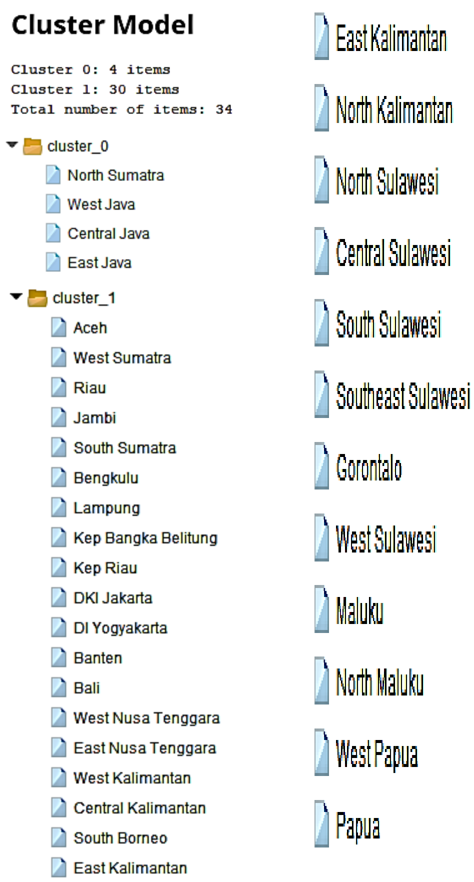


Figure 5. *Results of the cluster at the Junior High School level*

In Figure 5 it can be explained that 4 provinces in the high cluster (K1 = cluster_0) and 30 provinces in the low cluster (K2 = cluster_1) are for Junior High School with the scores from the final centroid for K1 = 3663 and K2 = 425.

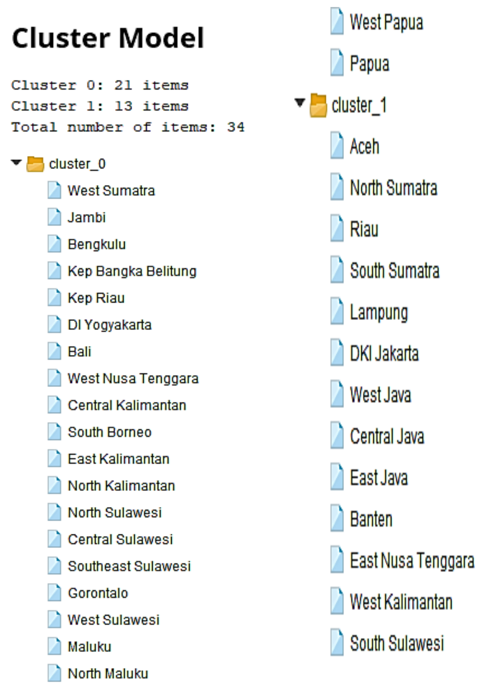


Figure 6. *Results of the cluster at the Senior High School level*

In Figure 6 it can be explained that 13 provinces in the high cluster (K1 = cluster_1) and 21 provinces in the low cluster (K2 = cluster_0) for Senior High School with the scores from the final centroid for K1 = 523 and K2 = 163.

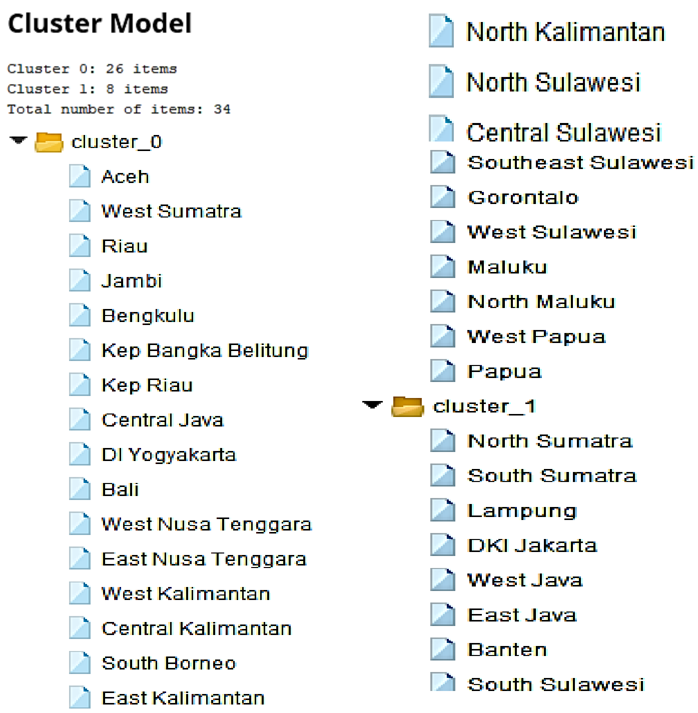


Figure 7. *Results of the cluster at the Vocational High School level*

In Figure 7 it can be explained that 8 provinces in the high cluster (K1 = cluster_1) and 26 provinces in the low cluster (K2 = cluster_0) are for Vocational High Schools with scores from the final centroid for K1 = 337 and K2 = 91.

The following is the recapitulation of the mapping of the number of libraries in Indonesia based on education levels which are clustered by region (province) as shown in the following graph:

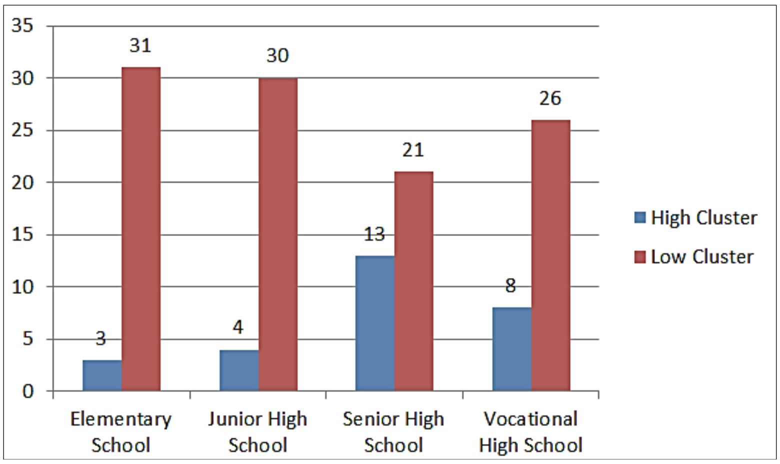
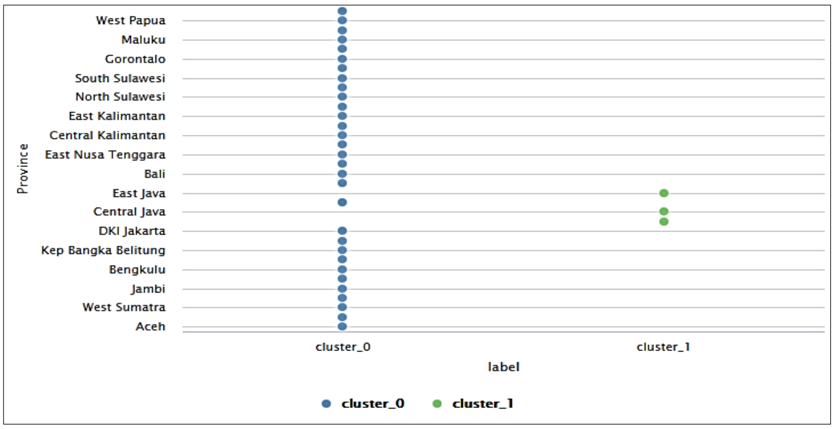
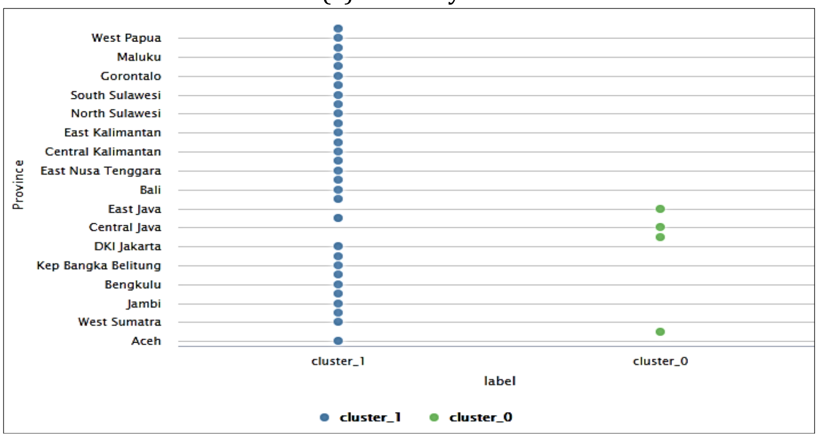


Figure 8. *Results of cluster recapitulation based on education levels*

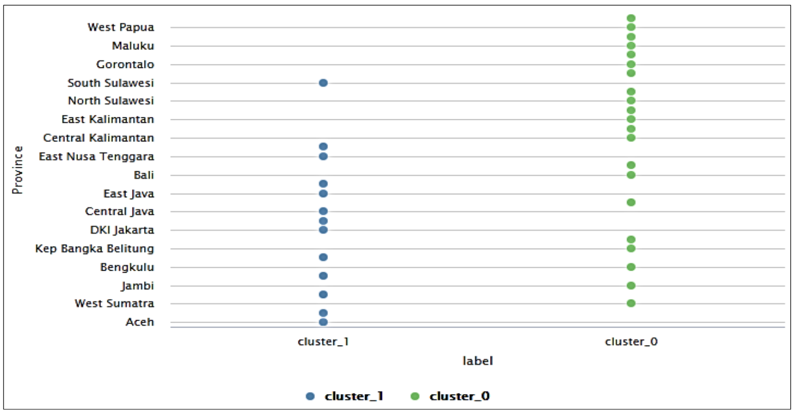
Based on the graph described in Figure 8, the use of school libraries based on education level is still low. Nearly 60% more areas in Indonesia still have a minimal school library at all levels of education. After all, the school library becomes a forum for students and teachers to take advantage of learning resources which are an effort to improve the quality of education.



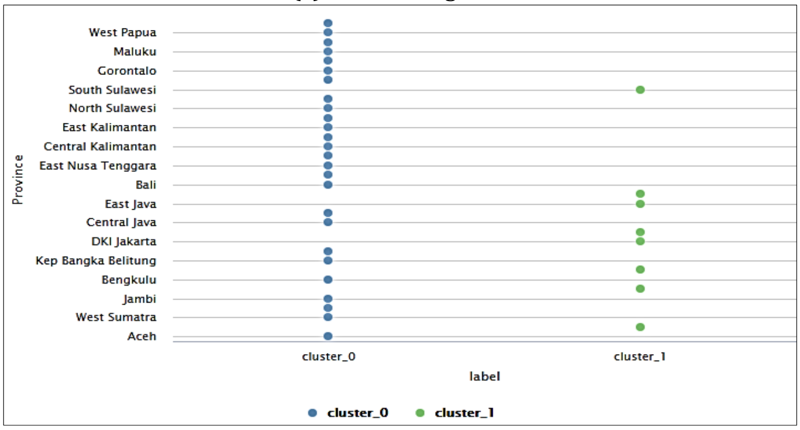
(a) Primary School



(b) Junior High School



(c) Senior High School



(d) Vocational High School

Figure 8. *The mapping graph is in the form of clusters at each level (a)(b)(c)(d)*

CONCLUSION

Based on the research results, it can be explained that the k-medoids method can be applied to mapping the number of libraries based on education levels in Indonesia. The results showed that 3 provinces in the high cluster (K1) and 31 provinces in the low cluster (K2) were for elementary schools; 4 provinces in the high cluster (K1) and 30 provinces in the low cluster (K2) for Junior High Schools; 13 provinces in the high cluster (K1) and 21 provinces in the low cluster (K2) for Senior High Schools; 8 provinces in the high cluster (K1) and 26 provinces in the low cluster (K2) for Vocational High Schools. The cluster formed is the optimal school cluster tested with the parameters of the Davies Bouldin Index (DBI) with the results of 0.176 (Elementary School), 0.284 (Junior High School), 0.780 (Senior High School) and 2.069 (Vocational High School).

REFERENCES

Ilham, A. H. Amirullah, and M. Darwis2, "Pemanfaatan perpustakaan sekolah dalam menunjang kegiatan belajar siswa jurusan administrasi perkantoran di smk negeri 1 sinjai kabupaten sinjai," *J. Off.*, vol. 2, no. 1, pp. 41–44, 2016.

R. Mahande and J. Jasruddin, "The Conceptual model of user satisfaction for e-Learning Edmodo on Undergraduate students: A Preliminary study," in *Proceedings of the The 1st International Conference on Computer Science and Engineering Technology Universitas Muria Kudus*, 2018.

D. N. Marpaung and M. F. Azzajjad, "The Effectiveness of Student Centre Learning in Experiment Method on Acid and Base Solution to Increase Student Achievement," *J. Appl. Sci. Eng. Technol. Educ.*, vol. 2, no. 1, pp. 32–36, Jun. 2020.

R. Novriliam and Yunaldi, "Pemanfaatan Perpustakaan Sekolah Sebagai Pusat Sumber Belajar Di Sekolah Dasar Negeri 23 Painan Utara," *J. Ilmu Inf. Perpust. dan Kearsipan*, vol. 1, no. September, pp. 141–150, 2012.

M. Relawanto, E. Widyarto, and B. Harnadi, "The Math Virtual Reality Game as learning media for 1st-Grade School Students," in *Proceedings of the The 1st International Conference on Computer Science* *and Engineering Technology Universitas Muria Kudus*, 2018.

A. Pakistyaningsih, M. Bahak, U. By, and H. E. Rudyanto, “School Library Utilization Technology Model to Improve Reading Interest and Reading Ability in Elementary Education,” *Univers. J. Educ. Res.*, vol. 7, no. 9, pp. 1945–1955, 2019.

F. S. Hilyana, “Implementation of Schoology-based E-Learning to Improve the ANEKA-based Character,” in *Proceedings of the The 1st International Conference on Computer Science and Engineering Technology Universitas Muria Kudus*, 2018.

A. N. N. M. C. M. A, “Utilization of library outreach services by dental school alumni, 1988-1998,” *J. Am. Dent. Assoc.*, vol. 132, no. 1, pp. 76–82, 2001.

I. Bafadal, *Pengelolaan Perpustakaan Sekolah*. 2011.

F. R. Senduk and F. Nhita, “Clustering of Earthquake Prone Areas in Indonesia Using K-Medoids Algorithm,” *Ind. J. Comput.*, vol. 4, no. 2016, pp. 65–76, 2019.

E. Wijaya and A. Sulaksono, “Implementation Of Neural Network And Canny Edge Detection To Recognize the Crime Through Surveillance Cameras,” in *Proceedings of the Proceedings of The 2nd International Conference On Advance And Scientific Innovation, ICASI 2019, 18 July, Banda Aceh, Indonesia*, 2019.

A. Kadafi and W. Gata, “Particle Swarm Optimization Based C4.5 for Teacher Performance Classification,” in *Proceedings of the Proceedings of The 2nd International Conference On Advance And Scientific Innovation, ICASI 2019, 18 July, Banda Aceh, Indonesia*, 2019.

S. Harikumar and P. V. Surya, “K-Medoid Clustering for Heterogeneous DataSets,” *Procedia Comput. Sci.*, vol. 70, pp. 226–237, 2015.

S. Defiyanti, M. Jajuli, and N. Rohmawati, “K-Medoid Algorithm in Clustering Student Scholarship Applicants,” *Sci. J. Informatics*, vol. 4, no. 1, pp. 27–33, 2017.

P. Arora, Deepali, and S. Varshney, “Analysis of K-Means and K-Medoids Algorithm for Big Data,” *Phys. Procedia*, vol. 78, no. December 2015, pp. 507–512, 2016.

H. Fei, N. Meskens, and C. H. Moreau, “Clustering of patient trajectories with an auto-stopped bisecting K-medoids algorithm,” *IFAC Proc. Vol.*, vol. 13, no. PART 1, pp. 355–360, 2009.

B. P, “A survey of clustering data mining techniques. In: Kogan J, Nicholas C, Teboulle M (eds) Grouping multi- dimensional data: recent advances in clustering,” *Springer, Berlin*, pp. 25–71, 2006.

K. Singh, D. Malik, and N. Sharma, “Evolving limitations in K-means algorithm in data mining and their removal,” *IJCEM Int. J. Comput. Eng. Manag. ISSN*, vol. 12, no. April, pp. 2230–7893, 2011.

R. Rahim, “Educational Data Mining (EDM) on the use of the internet in the world of Indonesian education,” *TEM J.*, 2020.