International Journal of Computer and Information System (IJCIS)
Peer Reviewed – International Journal
Vol : Vol. 04, Issue 04, December 2023
e-ISSN : 2745-9659
https://ijcis.net/index.php/ijcis/index

Claterization of Primary Schools In The Surakarta Region Using The K-Medoids Method Based on School Costs and Facilities

Siti Rokhmah
Informatika
Institut Teknologi Bisnis AAS Indonesia
Sukoharjo, Indonesia
sitirokhmah.itbaas@gmail.com

Abstract— Basic education has an important role in forming the foundation of a child's character. The city of Surakarta is a city that has many choices of elementary schools, both public and private. The large number of elementary schools requires clustering, so that it can help the government and the community in decision making. The most important factors in choosing an elementary school are school facilities and costs, so the clustering in this research is based on educational costs and the facilities provided by the school. The method used in this research is the K-Medoids clustering method, namely a clustering method that groups data based on groups that have maximum similarity. To evaluate the clustering results, silhoette value calculations are used. It is hoped that this research can help the government, especially the education department, in mapping elementary schools in the Surakarta area and assist parents in determining elementary school choices.

Keywords : Clustering, K-Medoids, Implemetary School, Surakarta.

I. INTRODUCTION

Basic education, especially at the elementary school level, has an important role in forming a strong educational foundation for future generations. The Surakarta region, as one of the important regions in Indonesia, has a number of elementary schools that provide educational services for children in the area. However, in making the decision to choose a school, various factors need to be considered, including the cost and quality of the facilities offered by each school [1], [2], [3]. The large number of elementary schools requires grouping of elementary schools, which is based on certain criteria. This can help the government and society in making decisions. For the government, it can be used as a basis for mapping elementary schools in the Surakarta area, for parents it can be used as a basis for determining their children's school choices.

Differences in operational costs and variations in facilities between elementary schools in the Surakarta area give rise to the need to carry out systematic and structured analysis. Apart from that, an analysis of school costs and facilities can make it easier for the community to determine the choice of elementary school for their children, especially elementary schools in the Surakarta area. This analysis can be carried out by utilizing data mining with clustering techniques. The clustering technique is carried out by grouping elementary schools in the Surakata area based on school costs and facilities [4], [5].

Clustering is a method of grouping data based on groups that have maximum similarity [6], [7]. Many methods are used in data clustering, including the K-Medoids method. The K-Medoids method was chosen as the main approach in this research because it is able to group unlabeled data into groups that have certain internal characteristics in common. By using this method, it is hoped that patterns or groups of schools will be identified that are similar in terms of operational costs and facilities [8], [9].

It is hoped that the results of this clustering will provide a better understanding to local governments, educational institutions and the general public about the differences in characteristics between elementary schools in the Surakarta area. This information can be used as a basis for making better decisions in terms of resource allocation, educational planning, and assisting parents or guardians in choosing a school that suits their needs and preferences.

This research will not only provide benefits in terms of mapping elementary schools in Surakarta, but can also be a basis for further research related to improving the quality of education, school development strategies, as well as formulating more effective education policies at the local and national level. Thus, it is hoped that this research can make a significant contribution in advancing the education sector in Indonesia, especially in the Surakarta area. The main objective of this research is to classify elementary schools in the Surakarta area into groups or clusters based on the level of operational costs and the type of facilities they have.

II. RESEARCH METHODS

There are several stages in this research, namely

1. Literature study In this research, we study references in the form of journals, books and other articles related to research. The journals used as references are journals that discuss data mining, data calsterization and the K-Medoids algorithm. Apart from that, there are also books and articles about calsterization and the K-Medoids method
2. Data collection The data used in this research is elementary school data in the Surakarta area, education cost data and school facility data. Data on elementary schools in the Surakarta area was obtained from data from the Ministry of Education and the results of observations at elementary schools in the Surakarta area. Data on education costs was obtained from observations and interviews with several elementary schools and parties involved.
3. Pre-Processing
   Pre-processing is processing raw data before processing it using the clustering method, there are several steps in pre-processing
   a. Normalisasi
      At this stage, adjustments are made to the scale and value range of each variable (costs and facilities) so that they have a similar value range for further analysis. To calculate data normalization, equation (2) is used.

$$\boldsymbol{Xnorm} = \frac{\boldsymbol{X' - X(min)}}{\boldsymbol{x(max) - x(min)}} \quad \textbf{(2)}$$

   b. Handling Missing Data
      At this stage, it is carried out to handle incomplete or missing data with appropriate data filling techniques.

4. Data analysis
   Data analysis is a step taken to group elementary school data in the Surakarta area. Several stages in data analysis are as follows
   a. K-Medoids Methode
      K-medoids is a clustering algorithm with a partition approach that uses the middle point (median) as a reference in determining clusters [10]. Implementation of the K-Medoids algorithm for clustering data on elementary school costs and facilities in the Surakarta area. Calculations are carried out using matrices and Euclidean distance in calculating similarities between schools. Euclidean distance is used to measure the shortest distance from one point to another. To see the euclidean distance equation, you can look at equation 1.

$$dist(p,q) = \sqrt{\sum_1^m (p-q)^2} \quad (1)$$

      Where p and q are the points whose distance will be found. The simialrity range in Euclidean distance is between 0 and 1. 0 indicates that the objects have no similarities and 1 if the objects are identical.
   b. Cluster Validation
      Cluster data testing is carried out to test how close the relationship between objects in the cluster is, from this test it can be seen how precisely the data is grouped [11]. The silhoette coefficient is a combination of the cohesion method which functions to measure how close the relationship is and the separation method to measure how far apart the cluster is from other clusters. Silhoette coefficient value of equation 6.

$$S_i = \frac{b_i - a_i}{\max(a_i - b_i)}$$

      Si is the silhouette value, ai is the average distance of the ith object to all objects in the cluster and bi is the average distance of the ith object to different clusters. The Silhouette value ranges from 0 to 1, where if the Silhouette value is more than 0.6 it is considered that the resulting cluster structure is good, whereas if the value is below 0.5 then the structure is considered weak.

III. RESULT AND ANALYSIS

1. Pre Processing
   At this stage, pre-processing is carried out on raw elementary school data in the Surakarta area. At this stage, variables are determined to be used to cluster elementary school data in the Surakarta area based on costs and facilities.

   a. Research Dataset
      The research data includes elementary school data in the Surakarta area, while education costs and facilities were obtained from observations and interviews with a number of related parties. To see the research dataset, see table 1.

Tabel 1. Dataset Research

No	Nama SD	Jumlah lab	Jumlah Ruangan	Jam belajar	Jam belajar	Biaya masuk	Biaya SPP
1	Sd N Banyuagung 3	6	1	1	Reguler	0	0
2	Sd N Banyuanyar 2	6	1	1	Reguler	0	0
3	SD N Banyuanyar 3	7	1	1	Reguler	0	0
…	……..	……	….	….	…	…	…
244	SD Muhammadiyah Makam BERGOLA	6	0	1	Reguler	800000	125000

b. Research Variable

1. Jumlah laboratorium
   This variable shows the number of laboratories in each school, laboratories include computer laboratories, language laboratories and science laboratories.

2. Ruang kelas
   The classroom variable is the number of classrooms owned by the school. The classroom is a room used for teaching and learning activities.

3. Library variables
   Shows the number of libraries owned by the school.

4. Study Hours
   Study hours are a facility offered by the school. There are 2 forms of study hours, namely regular which is coded 1, and full day which is coded 2.

5. School fee
   The entrance fee is a fee paid at the start of entering school. Entry fees include construction fees, uniform fees, registration fees and activity fees. SPP costs are presented in 2 digits, namely by dividing by 1,000,000

6. Mountly fee
   These are school fees paid every month by students. SPP costs are presented in 2 digits, namely by dividing by 100,000

2. Data Normalization
   Normalization aims to produce data with the same distance and smaller object values. To calculate data normalization, the formula equation (2) is used and the following results are obtained.

Table 2. data normalization results

No	R. Kelas	R. Lab	R. Perpus	jam belajar	Baiya Masuk	SPP
1	0,240	0,200	0,500	0,000	0,000	0,000
2	0,240	0,200	0,500	0,000	0,000	0,000
3	0,280	0,000	0,500	0,000	0,000	0,000
4	0,240	0,200	0,500	0,000	0,000	0,000
5	0,480	0,200	0,500	0,000	0,000	0,000
…	….	…	….	….	…..	…..
243	0,280	0,000	0,000	0,000	0,054	0,130
244	0,240	0,000	0,500	0,000	0,043	0,125

3. Clusterization using the K-Medoids method

a. Calculation of cluster centers using the K-Medoids method

1. Initiation of the 1st iteration cluster center
   The selection of K-medoids cluster centers for the 1st iteration is determined randomly. The following are the initial cluster values in the 1st test

   Tabel 3. Cluster Center Iterasi 1

K-ke	R. Kelas	R. Lab	R. Perpus	jam belajar	Baiya Masuk	SPP
1	0,240	0,000	0,400	0,000	0,000	0,000
2	0,480	0,400	0,500	0,500	0,216	0,360
3	0,960	1,000	0,500	0,500	1,000	0,800

2. Calculation of the distance between objects.
   To perform clustering, the first step is to calculate the distance between objects and the cluster center in the 1st iteration. Calculations are carried out using the Euclidean distance calculation formula. To calculate the Euclidean distance calculation formula, use equation 1. Following are the results of calculating the distance between objects at the center of the first cluster using the Eucledian distance formula.

3. Grouping data based on the closest distance to the 1st iteration. After calculating between objects, the next step is to group the data according to its clusters. Grouping is done by looking at the closest distance from the cluster center. To see the cluster results in the 1st iteration, you can see the table 4.

Table 4. Clustering result iterasi 1

Data ke	X1	X2	X3	Jarak terdekat	Klaster
1	0,000	0,883	3,420	0,000	1
2	0,000	0,883	3,420	0,000	1
3	0,000	0,803	3,740	0,000	1
…	…	…	…	…	
49	2,930	1,296	0,994	0,994	3
50	1,296	0,345	1,139	0,345	2
…	…	…	…	…	…
238	0,596	0,392	2,826	0,392	2
239	0,021	0,749	3,385	0,021	1
240	0,077	0,596	2,876	0,077	1
241	0,260	0,576	3,360	0,260	1

4. Determine the value of the new medoids
For the calculation of the 2nd iteration, the 2nd medoids are determined, namely by determining the cluster center of the object representation. For the 2nd iteration, the cluster library values are determined which can be seen in the table. 6

Table 6. Cluster Center 2nd Iteration

R. Kelas	**R. Lab**	**R. Perpus**	**jam belajar**	**Baiya Masuk**	**SPP**
0,280	0,000	0,500	0,000	0,000	0,000
0,640	0,600	0,500	0,500	0,385	0,125
0,960	0,600	0,500	0,500	0,769	0,625

5. Distance between objects in the new medoids
Next, distance calculations are carried out in the second iteration using the Euclidean distance formula. The results of calculations between objects towards the 2nd cluster center can be seen in the table 7.

Tabel 7. Result of calculation between objects on 2nd Iteration

Data ke-	X1	X2	X3	
1	0,080	0,974	2,112	0,080
2	0,080	0,974	2,112	0,080
3	0,000	1,134	2,272	0,000
4	0,080	0,974	2,112	0,080
5	0,250	0,883	1,872	0,250
…	…	…	…	..
243	0,270	1,331	2,300	0,270
244	0,057	1,128	2,110	0,057

6. Calculate the total deviation
After calculating between objects, the total deviation is calculated using the equation

$$\boldsymbol{S = b - a}$$

Table 8. Deviation

No	X1	X2	X3
1	0,000	0,883	3,420
2	0,000	0,883	3,420
3	0,000	0,803	3,740
4	0,000	0,883	3,420
…	….	….	….
243	0,290	0,924	3,668
245	0,057	0,726	3,486

Where b is the closest distance in the 1st iteration
a is the closest distance in the 2nd iteration
From these results, the deviation value =

S=61,370 – 61, 030 = 0, 340

Because b>a, the iteration process is stopped, and the 2nd iteration calculation can be used as the calculation result.

4. Cluster Result
Cluster results from calculations using the K-medoids method, obtained the following cluster results

Table 9. Clustering result

	C1	C2	C3
Jumlah anggota	187	35	22

5. Evaluation of the clustering model
Evaluation of the cluster model is carried out using calculations of silhouette values which are calculated using the equation Cluster results can be referred to in the table 10.

Table 9. Equation cluster result

Silhoutte	**interpretation**
0,71-1,00	Klaster kuat
0,51-0,70	Klaster baik
0,26-0,50	Klaster lemah
0,00-o,25	Klaster buruk

The obtained silhouette value = 0.669 which shows that the cluster results are good enough to be used.

VI. CONCLUSION

This research uses data obtained from the Surakarta education office website and from observations at a number of elementary schools in the Surakarta area. Clustering in this research uses grouping based on facilities and educational costs. Clustering was made into three clusters, and data obtained from cluster 1 was 187 schools, cluster 2 consisted of 35 schools and cluster 3 consisted of 32 schools. To evaluate the cluster used, a silhouette value calculation was used which produced a value of 0.669 which indicated a good value, and the cluster was acceptable.

THANK-YOU NOTE

We would like to express our thanks to all parties who supported and assisted in this research process.

REFERENCES

[1] Suhirman, "Pengaruh Biaya Pendidikan terhadap Hasil Belajar melalui Proses Belajar Mengajar di Sma Negeri Se-Kabupaten Rembang Tahun 2011," *J. Econ. Educ.*, vol. 1, no. 2, pp. 117–122, 2012.

[2] N. A. Prastika, H. Zhafirah, A. R. Sari, and ..., "Pengaruh Sarana Prasarana, Biaya, Dan Lokasi Sekolah Dalam Menentukan Pilihan Rasional Orang Tua Memilihkan Sekolah Untuk …," *Pros. Semin. …*, 2022, [Online]. Available: https://prosiding.unimus.ac.id/index.php/semnas/article/view/1239%0Ahttps://prosiding.unimus.ac.id/index.php/semnas/article/viewFile/1239/1240

[3] P. Y. A. Dewi and L. Indrayani, "Persepsi Orang Tua Siswa Terhadap Biaya Pendidikan," *Ekuitas J. Pendidik. Ekon.*, vol. 9, no. 1, p. 69, 2021, doi: 10.23887/ekuitas.v9i1.27034.

[4] M. F. Edy Irwansyah, *Clustering Teori dan Aplikasi*. 2015.

[5] N. Nurahman, A. Purwanto, and S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022, doi: 10.30812/matrik.v21i2.1411.

[6] A. Asroni and R. Adrian, "Penerapan Metode K-Means Untuk Clustering Mahasiswa Berdasarkan Nilai Akademik Dengan Weka Interface Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Semesta Tek.*, vol. 18, no. 1, pp. 76–82, 2016, doi: 10.18196/st.v18i1.708.

[7] M. E. Pratama and A. Finandhita, "Penerapan Metode Clustering untuk Pengelompokan Potensi Wisata di Kabupaten Sumedang," *J. Ilm. Komput. dan Inform.*, no. 112, 2019.

[8] F. Fajriana, "Analisis Algoritma K-Medoids pada Sistem Klasterisasi Produksi Perikanan Tangkap Kabupaten Aceh Utara," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 2, p. 263, 2021, doi: 10.26418/jp.v7i2.47795.

[9] S. Bahri and D. M. Midyanti, "Penerapan Metode K-Medoids untuk Pengelompokan Mahasiswa Berpotensi Drop Out," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 1, pp. 165–172, 2023, doi: 10.25126/jtiik.20231016643.

[10] S. Defiyanti, M. Jajuli, and N. Rohmawati, "Optimalisasi K-MEDOID dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan CUBIC CLUSTERING CRITERION," *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 211–218, 2017, doi: 10.25077/teknosi.v3i1.2017.211-218.

[11] M. A. Nahdliyah, T. Widiharih, and A. Prahutama, "METODE k-MEDOIDS CLUSTERING DENGAN VALIDASI SILHOUETTE INDEX DAN C-INDEX (Studi Kasus Jumlah Kriminalitas Kabupaten/Kota di Jawa Tengah Tahun 2018)," *J. Gaussian*, vol. 8, no. 2, pp. 161–170, 2019, doi: 10.14710/j.gauss.v8i2.26640.