JOURNAL OF WORLD SCIENCE
https://jws.rivierapublishing.id/index.php/jws
Volume 2 No. 3 March 2023
P-ISSN: 2828-8726
E-ISSN: 2828-9307

ANALYSIS AND DESIGN OF STUDENT POINT SYSTEMS TO IMPROVE STUDENT ACHIEVEMENT USING THE CLUSTERING METHOD

Ade Bani Riyan[1], Mochamad Fikri Rifai[2], Christina Juliane[3]
Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI, Bandung, Indonesia

adebaniriyan@gmail.com[1], fzaki292@gmail.com[2], christina.juliane@likmi.ac.id[3]

ABSTRACT
The student points system is an application for recording students' achievement and offense points. The lack of recording and dissemination of information on achievement results makes students less motivated to improve achievement, and the distribution of scholarships for outstanding students is inappropriate. To improve student achievement, an application program is needed that can record and disseminate student achievement data in real-time, accurate, and effective. So, the purpose in this study is to know and analyze the design of the student point system to improve student achievement using the clustering method. Researchers use the Clustering Method in calculating data to determine the accuracy of scholarship distribution for outstanding students. Clustering with the most achievement points is clustering 2 with 25,254 Achievement Points. The total number in the level 2 cluster is 1,797 which indicates the number is close to 2,000 or 2 which is the result of data transformation from the junior high level. The implication of clustering research on student point data is to provide useful information for the Foundation as an institution that houses schools in allocating scholarships for outstanding students. In this case, clustering 2 with the highest number of Achievement Points indicates that there is a group of students with high achievement points. By using the clustering results, the Foundation can allocate scholarships more effectively and efficiently, because it can identify outstanding students from various school levels more easily.

Keywords: clustering method, student achievement, student point system.

Corresponding Author: Ade Bani Riyan
Email: adebaniriyan@gmail.com

INTRODUCTION

The student point system is an application program for recording student achievement points and violations (Dwijaya, 2020). Not recording and disseminating information on achievement results makes students less motivated to improve. Data on student achievement and violations is needed, especially for scholarship distribution (Rachmawaty, 2016). Having student achievement data will make it easier to distribute accurate scholarships. To improve student achievement, an application program is necessary as a tool for recording and disseminating data on student achievement information in real time, accurately and effectively. Researchers used the Clustering Method in calculating data to determine the accuracy of scholarship distribution for outstanding students.

Education is one indicator of whether a country is progressing or not (Al-Hendawi et al., 2023). If the government does not pay attention to the progress of its education, entry will not move. Education is the main thing in a country because it greatly influences other fields (Muhardi, 2004). Such as government, finance, defense and so on. In the continuity of a country in the future, education is also very important because the education of the next children of a country that is not considered will threaten the sustainability of the country in the future (Ramadhanu et al., 2021).

Data mining is a process of artificial intelligence, machine learning, statistics, and mathematics to extract and identify useful and related information from large databases (Gorunescu, 2011).

The clustering method groups several data or objects into groups (groups) so that each group contains data that is as similar as possible and different from data/objects in other groups (Maukar et al., 2022). While cluster analysis, according to Eka Haryati in his research was used to determine patterns with high characteristics (Haryati et al., 2022). The clustering method has so far been applied in various fields, as written in research and journals (Windarto et al., 2017). Clustering is the process of grouping many data points into two or more groups so that data points belonging to the same group are more like one another than in different groups, based solely on the information available with the data points (Nidheesh et al., 2017).

KDD is a method used to search for knowledge from a database. In his research explains that the results of knowledge can be used as a knowledge base that can be used to make a decision (Adiya & Desnelita, 2019). In more detail, the KKD process is in the following figure adopted from (Gullo, n.d.).

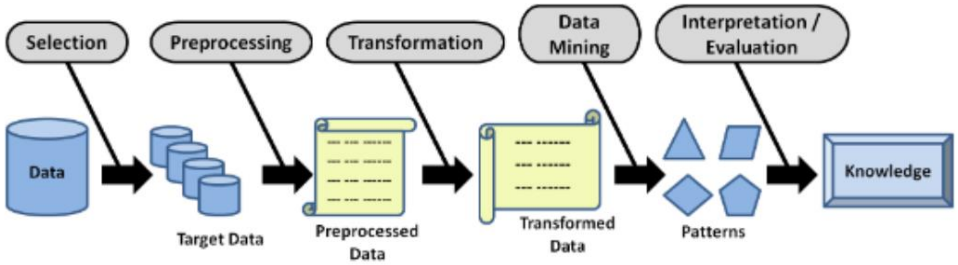

Figure 1. Steps of the KKD Process

The K-Means algorithm is one of the algorithms often applied in grouping because of its efficiency and simplicity (Harding et al., 2006). It is recognized as part of the top 10 data mining algorithms by IEEE (Wu et al., 2008).

Based on the above background, the purpose of this study is to find out and analyze the design of the student point system to improve student achievement using the clustering method. With this research, it is hoped that it can be a solution for recording and disseminating student achievement information, so that students are more motivated to excel and increase the level of accuracy in distributing scholarships for outstanding students.

METHODS

In this study, researchers used data from point system applications that run-in schools. This study has four stages illustrated in Figure 2. Research Stages.

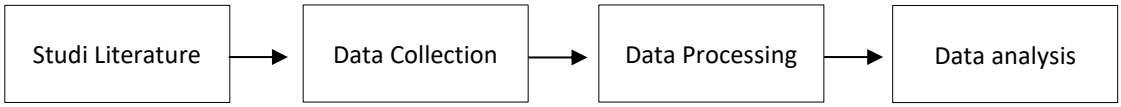

Figure 2. Research Stages

1) Study Literature

The literature study was carried out by collecting various methods and theories related to the problems in the research conducted, namely the use of the K-Means clustering algorithm. Literature studies are obtained from multiple sources, including magazines, articles or scientific papers that are used to strengthen the theoretical basis in research. Several journal references were used as an overview in this study, namely research conducted by " Interactive Web-Based Smart School at Amaliyah Private Elementary School Sunggal with the K-Means Cluster Algorithm Information Systems Study Program" and "Implementation of the K-Means Method in Mapping Student Groups Through Lecture Activity Data" (Fauzi & Samsudin, 2022); (Rosmini et al. , 2018).

2) Data collection

The stages of data collection are the process of collecting data through the student point system application database from July 2021 to December 2021. The total data obtained from this period amounted to 537 rows.

3) Processing Data

The data obtained at the data collection stage is carried out by the next process, namely the pre-processing step. At this stage several activities are carried out, namely:

a) Data selection was carried out to collect data that is suitable for analysis purposes, namely selecting data with the characteristics of Level, Class, Achievement Points, Violation Points;
b) Convert data into a form that is more suitable for analysis, namely converting classes into integer form (numbers) to facilitate research and as a requirement for data to be read by the rapid miner tool;
c) Clearing data is removing some inconsistent data from the collected data. Some of the data characteristics of the data obtained after pre-processing the data can be seen in Table 1 Sample dataset, as follows:

Table 1. Sample Dataset

Name	Gender	Level	Class	Achievement Points	Violation Points
Sister Al Aina	P	3	10	50	25
Ajat Sudrajat	L	3	10	25	30
Aji Soleman	L	3	11	30	75
Alika Trista Aulia	P	3	10	25	30
Amanda	P	3	10	75	20
Amelia Putri	P	3	11	20	20
André Afrijal Maolana	L	3	10	30	25

Description Level
1 = Elementary School
2 = Junior High School
3 = Senior High School

RESULTS AND DISCUSSION

537 data were processed using two methods. First the data is loaded into Rapid Miner and run using the Elbow method to get the right number of clusters before starting the clustering process. Then, the data is processed using the K -means way to get clustering results. An overview of the elbow method process can be seen in Figure 3. The Elbow Method process and an overview of the k-means method process can be seen in Figure 4. The K-Means Method Process.

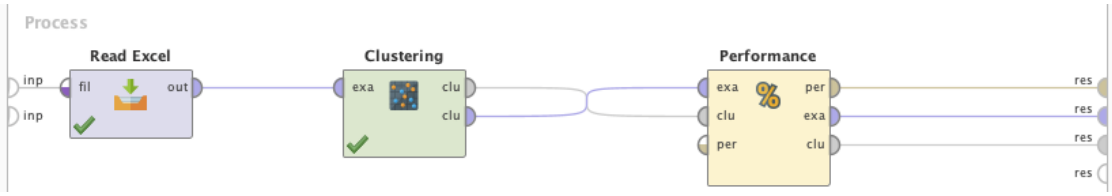

Figure 3. Elbow Method Process

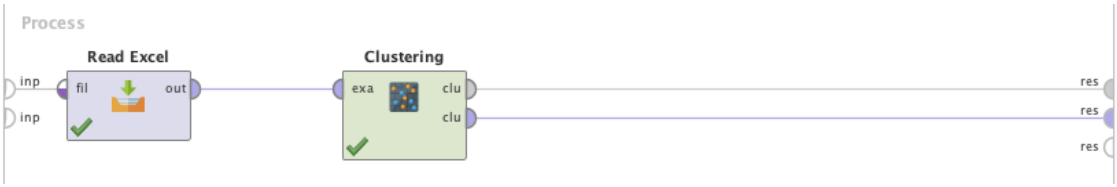

Figure 4. The process of the K-means method

Experimental results k-2 to k-10 and seed value = 10. Seed is a random number in cluster generation with seed value 10 as the default number used as a process reference. This normalization produces an output value between 0 and 1. Then the process of grouping the datasets into their respective groups based on the similarity of characteristics is done by calculating the distance value using the Euclidean Distance in the equation and the K-Means algorithm for processing orders. Then the cluster results were analyzed and evaluated to find the optimal number of K using the Elbow method. The Elbow method calculates the largest SSE depreciation difference and is in the shape of an elbow. Calculation of SSE using the equation. After the clustering trial process on the dataset, the data processing is carried out by calculating the distance value to determine the number of clusters, the results are shown in Table 2—comparison of Average Results.

Table 2. Comparison of Average Results

Clusters	Average
K2	385,788
K3	69,047
K4	36,829
K5	28,161
K6	20,936
K7	16.120
K8	12.172
K9	10,789
K10	9,348

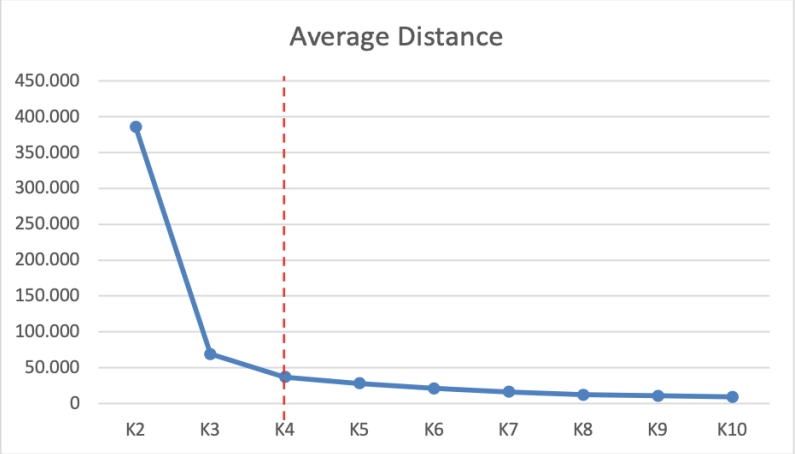

Figure 5. Graph of Clustering Results

We try to do a test with the Elbow method to find out how many clusters are suitable for the analysis process. Based on the test results in Table 2. Comparison of Average Results, we visualize the table in graphical form and look for clusters with the most angular lines. Based on Figure 5 it can be seen that point K4 shows the most angular results compared to other points. So it can be concluded that the most optimal cluster is according to the elbow method using 4 groups. Furthermore, testing was carried out again on the rapid miner using the recommended clustering, the results of the test using the K-Means algorithm with 4 sets can be seen in Figure 6. Results of 4 Cluster Models.

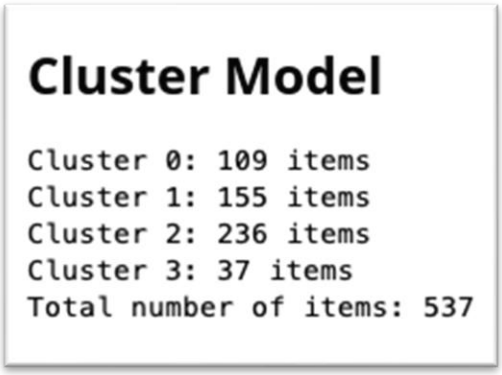

Figure 6. Results of 4 Cluster Models

The results of clustering with 4 clusters show that the lowest achievement point values are in cluster 4 and the highest in cluster 3. Detailed clustering comparisons can be seen in Table 3—results of the Clustering Process.

Table 3. Clustering Process Results

attributes	Cluster_0	Cluster_1	Cluster_2	Cluster_3
Level	1,725	1690	1,797	1676
Class	5514	5,561	6004	5,595
Achievement Points	75	25,032	25,254	50
Violation Points	26.147	75	24,831	25

CONCLUSION

The results of the clustering research on student data points obtained clustering with the most achievement points, namely clustering 2, with a total of 25,254 Achievement Points. The total number at cluster level 2 is 1,797, where these results show the number is close to 2,000 or 2, which is the result of data transformation from the junior high school level raised in Table 1. Sample Dataset. The foundation, as an institution that oversees schools, can see achievement data from the school level, making it easier to allocate scholarships for outstanding students based on their school level.

REFERENCES

Adiya, M. H., & Desnelita, Y. (2019). Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru. *Vol*, *1*, 17–24.

Al-Hendawi, M., Keller, C., & Khair, M. S. (2023). Special Education in the Arab Gulf Countries: An Analysis of Ideals and Realities. *International Journal of Educational Research Open*, *4*, 100217. https://doi.org/10.1016/j.ijedro.2022.100217

Dwijaya, D. A. (2020). Perancangan Aplikasi Untuk Pelanggaran Dan Prestasi Siswa Pada Smp Kartika Ii-2 Bandar Lampung. *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, *1* (2), 127–136. https://doi.org/10.33365/jatika.v1i2.313

Fauzi, M. S., & Samsudin, S. (2022). Smart School Berbasis Web Interaktif di SD Swasta Amaliyah Sunggal dengan Algoritma K-Means Cluster. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *11*(3), 332–341.

Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.

Gullo, F. (n.d.). From patterns in data to knowledge discovery: what data mining can do. Phys. Procedia 62, 18–22 (2015). *3rd International Conference Frontiers in Diagnostic Technologies*.

Harding, J. A., Shahbaz, M., & Kusiak, A. (2006). *Data mining in manufacturing: a review*.

Haryati, A. E., Wijaya, T. T., Wen, G. K., & Thobirin, A. (2022). Fuzzy subtractive clustering (FSC) with exponential membership function for heart failure disease clustering. *International Journal of Artificial Intelligence Research*, *6*(1). https://doi.org/10.29099/ijair.v7i1.306

Maukar, A. L., Marisa, F., & Widodo, A. A. (2022). Analisis Data Penerimaan Mahasiswa Baru Berbasis K-Means. *JIKO (Jurnal Informatika Dan Komputer)*, *6*(2), 142–147.

Muhardi, M. (2004). Kontribusi pendidikan dalam meningkatkan kualitas bangsa Indonesia. *Mimbar: Jurnal Sosial Dan Pembangunan*, *20*(4), 478–492. ttps://doi.org/10.29313/mimbar.v20i4.153

Nidheesh, N., Nazeer, K. A. A., & Ameer, P. M. (2017). An enhanced deterministic K-Means clustering algorithm for cancer subtype prediction from gene expression data. *Computers in Biology and Medicine*, *91*, 213–221. https://doi.org/10.1016/j.compbiomed.2017.10.014

Rachmawaty, D. T. (2016). *Pengaruh beasiswa Bidikmisi terhadap prestasi belajar mahasiswa penerima beasiswa Bidikmisi di UIN Syarif Hidayatullah Jakarta*.

Ramadhanu, A., Defit, S., & Kareem, S. W. (2021). Hybrid Data Mining with the Combination of K-Means Algorithm and C4. 5 to Predict Student Achievement. *International Journal of Artificial Intelligence Research*, *5* (2), 180–189.

Rosmini, R., Fadlil, A., & Sunardi, S. (2018). Implementasi Metode K-Means Dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah. *IT Journal Research and Development*, *3* (1), 22–31.

Windarto, A. P., Komputer, I., & Bangsa, T. (2017). Implementation of Data Mining on Rice Imports by Major Country of Origin Implementation of Data Mining on Rice Imports by Major Country of Origin Using Algorithm Using K-Means Clustering Method. *No. November*.

https://doi.org/10.29099/ijair.v1i2.17

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., & Yu, P. S. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14* (1), 1–37. https://doi.org/10.1007/s10115-007-0114-2