Jurnal Mandiri IT ISSN 2301-8984 (Print)
Vol. 11 No. 4, April (2023), pp. 152-158

Educational data mining for mapping the comprehension of personality subject using K-Means algorithm (case study of SD Insan Mulya)

Wiwiet Herulambang[1], M. Mahaputra Hidayat[2], Nabila Amanda Putri[3]
[123]Department of Informatics Engineering, Universitas Bhayangkara Surabaya, Jawa Timur, Indonesia

A R T I C L E I N F O

Article history:

Received April 02, 2023
Revised April 16, 2023
Accepted April 30, 2023

Keywords:

Clusterizations
Educational Data Mining
K-Means Algorithm
Mapping
Subject Personality

ABSTRACT

The government has emphasized the importance of character education since the elementary school (SD) level by providing various subjects as an effort to build personality/character for each student. It is important for teachers and schools to know the level of understanding in their students. This research is for. This study uses the K-Means Clustering Algorithm method to help map students to the level of understanding of the Personality subject. The results of this study indicate that clustering using the K-Means Algorithm method contains 4 effective clusters in mapping the understanding of personality subjects. The result of the percentage of clusterization results with interviews is 89.39% while the percentage of software is 91.70%. The conclusions obtained in this research are the subjects of Moral Education, Religious Education, Citizenship Education are included in the Subjects that can affect the student's personality.

Corresponding Author:

M. Mahaputra Hidayat,
Department of Informatics Engineering,
Universitas Bhayangkara Surabaya,
Jl. Ahmad Yani No. 144, Surabaya, Jawa Timur, 60231, Indonesia.
Email: mahaputra@ubhara.ac.id

1. INTRODUCTION

In the world of education, character is very much needed by students to form a good, wise, honest, responsible person and can respect others. The government has emphasized the importance of character education since the elementary school (SD) level by providing various subjects as an effort to build personality/character for each student.

As the number of student data continues to increase every year, the number of student data increases so that there is a buildup of data that has not been processed optimally to deepen new information and knowledge through the patterns formed from the data collection. Increasingly increasing data can be processed with several techniques and methods to process it into information that can be applied as consideration for students in the policy and decision-making process as well as an early warning for students based on the results of low personality groupings that have the potential to not be accepted at school. community environment (Hidayat et al., 2013).

Student performance modelling is one of the challenging and popular research topics in educational data mining (Khan & Ghosh, 2021). Educational data mining and learning analytics, these two terms are often used interchangeably. One avenue to provide clarity, uniformity, and consistency around the two fields, is to identify similarities and differences in topics between the two evolving fields (Lemay et al., 2021). By predicting students' performance, we can identify students

risk of academic failure and help teacher to take some actions such as guidance or interventions to help learners as early as possible, or carry out continual evaluation of students as to optimize learning path or personalized learning resources recommendation (Xiao et al., 2022).

Based on this, it becomes an interest to research about understanding mapping in subjects about students' personality/characteristics. With this research, to map understanding on subjects using Educational Data Mining combined with the K-Means Algorithm method.

Based on the above background, the authors are interested in conducting research with the title "Educational Data Mining for Mapping the Understanding of Personality Subjects with the K-Means Algorithm (Case Study of SD Insan Mulya)". This research will give the results of personality class that can be used as material for teacher/school consideration to made policies and decisions as well as an early warning for students.

2. RESEARCH METHOD

Data mining is the process of analyzing data from different perspectives and transform it into important information that can be used to increase profits, reduce expenses, or even both. Technically, data mining can be referred to as a process for finding correlations or patterns from hundreds or thousands of fields from a large relational database. Data mining is a large-scale data processing method therefore data mining has an important role in the fields of industry, finance, weather, science, and technology. In general, data mining studies discuss methods such as clustering, classification, regression, variable selection, and market analysis (Janusz et al., 2022).

Educational Data Mining is the use of Data Mining methods on educational data such as student information, educational records, exam results, student participation in class, and the frequency of students' asking questions. In recent years, EDM has become an effective tool used to identify hidden patterns in educational data, predict academic achievement, and improve the learning/teaching environment (Yağcı, 2022).

One of the techniques known in data mining is clustering. Clustering is a grouping of several data or objects into clusters so that each in that cluster will contain data that is as similar as possible and different from objects in other clusters. There are two clustering methods that is well known, namely hierarchical clustering and partitioning. The hierarchical clustering method itself consists of complete linkage clustering, single linkage clustering, average linkage clustering and centroid linkage clustering (Hussain et al., 2022).

K-Means Clustering is a method of data analysis or data mining method that performs an unsupervised modeling process and is one of the methods that performs data grouping with the partition system. The K-Means method attempts to group existing data into groups, where data in one group has the same characteristics as each other while different characteristics will be grouped into the other group (Ulfah & Irtawty, 2023). The K-means algorithm is an algorithm that needs the input parameters by as much as k and divides a set of n objects into k clusters so that the level of similarity between members in a cluster is high while the level of resemblance to members in other clusters is very low. The similarity of cluster towards members is measured by the object's proximity to the mean value on the cluster or can be referred to as a centroid cluster or mass center (Aldino et al., 2021). The steps of clustering with the K-Means method are as follows:

a. Select the number of clusters k.
b. Initialization of the center of this cluster can be done in a variety of ways, but the most common thing is to do so in the way the cluster centers are initially rated with random numbers.
c. Allocate all data/objects to the nearest cluster. The proximity of the two objects is determined based on the distance of two objects. Likewise, the proximity of a data to a particular cluster is determined by the distance between the data and the center of the cluster. At this stage it is necessary to calculate the distance of each data to each cluster center. The distance betwwen one data and one specific cluster determines which data is entered in which cluster. To distance all data to each cluster center point can use the Euclidean distance theory. (Dewi et al., 2020)

ISSN 2301-8984 (Print)

d. Recalculate the center of the cluster with the current cluster member. The center of the cluster is the average of all data/objects in each cluster. You can also use the median of the cluster. So mean is not the only size you can use.
e. Task each object again using the new cluster center. If the center of the cluster is not changed then the clustering process is complete. Or, go back to step 3 until the center of the cluster is unchanged.
   Calculation of the closest distance Euclidean Distance to calculate the distance

$$d = | x - y | = \sqrt{\sum_{i}^{n} = 1 \ (Xi - Yi)^2}$$

   X=Data
   Y=Centeroid Cluster

f. Renewal of a centroid point can be done with the following formula:

$$New\ Centeroid = \frac{S1 + S2 + \cdots Sn}{\sum s}$$

   S1= Data Value record -1
   Sn= Data value record –n
   $\sum s$ = Total data record

Flowchart can be defined briefly that serves to graphically depict the steps and sequences of procedures of a program (Tomaskova & Tirkolaee, 2021). Monologic flowchart of analysts and programmers to break down problems into smaller segments and help in analyzing other alternatives in the operation of each function. Flowcharts are usually to make it easier if there is a problem, especially a problem that can be studied and can be evaluated further, so the role in making the system is really needed for the flowchart so that you know the problems that can be solved properly and accurately.

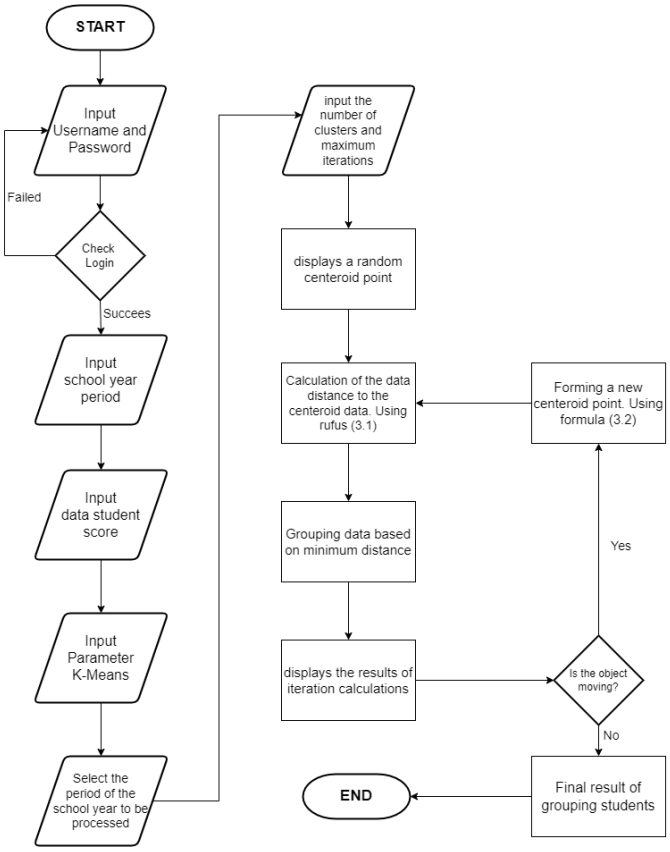

Figure 1. Flowchart system of this research

3. RESULTS AND DISCUSSIONS

The data used in this research is data on values for subjects such as Budi Pekerti, Citizenship Education, and Religion for grades 3, 4 and 5. The data obtained is 3906 in SD Insan Mulya. The following are the results where testing occurs at the time of testing rather than developing an information system for the informatics engineering laboratory at Bhayangkara University Surabaya which will be explained in detail for the functions that happen.

Login page is the initial page of the website that is used for limit user access to the main website.

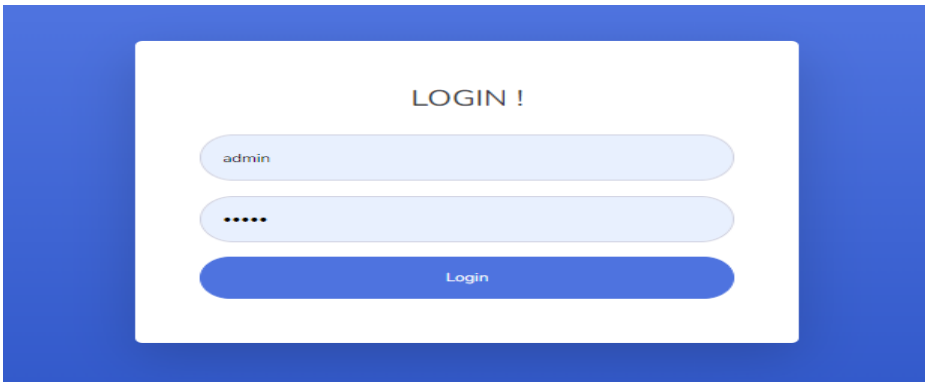

Figure 2. Login page

ISSN 2301-8984 (Print)

Periode page is used to add the period or school year that will be processed by the system.

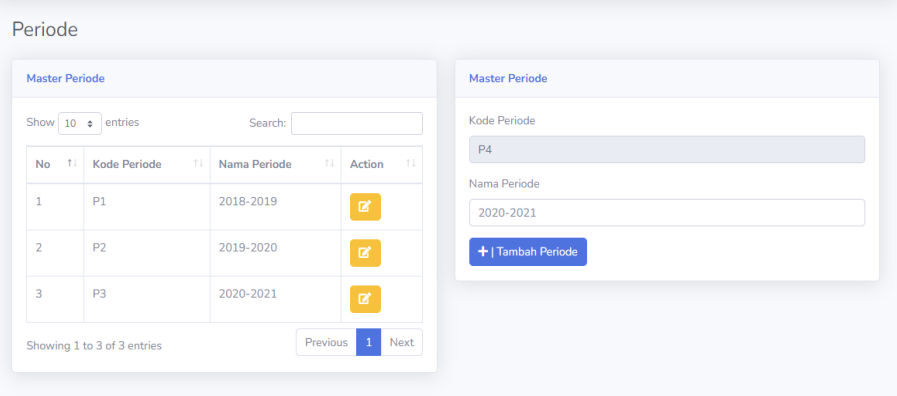

Figure 3. Periode page

This page is used to add the required parameters on the calculation process.

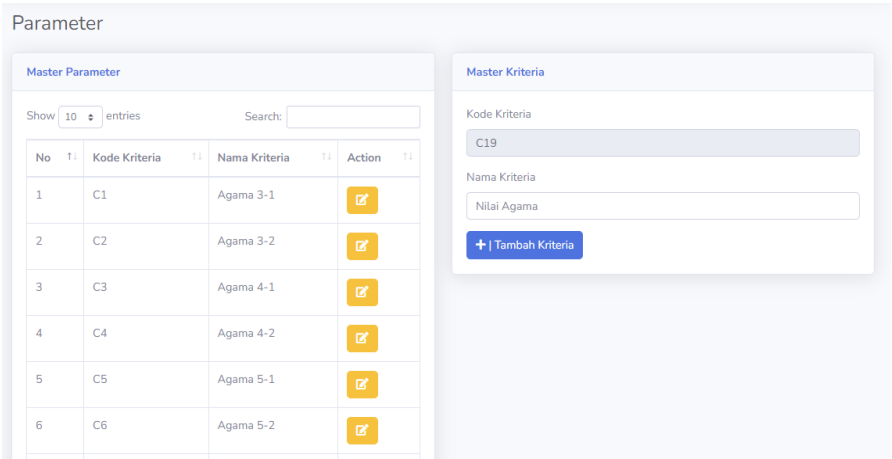

Figure 4. Parameter page

This page is used to display data that has been successful inputted into the application. The data displayed is data that will be processed by system.

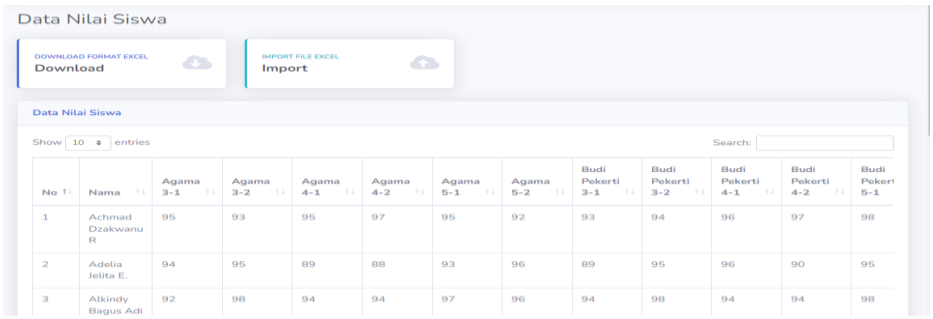

Figure 5. Student dataset page

This page is used to select data to be processed by K-Means.

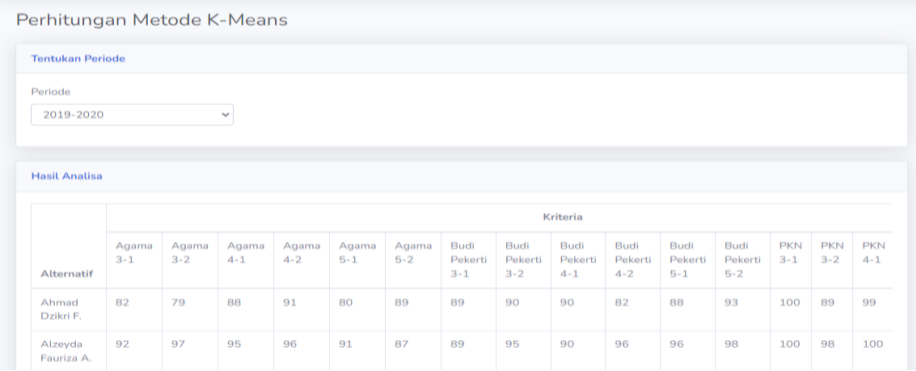

Figure 6. Choose period page

After choose the data or period, it will display the initial center point centeroid chosen randomly by the system, followed by the process calculating the distance of the data to the center point of the centeroid.

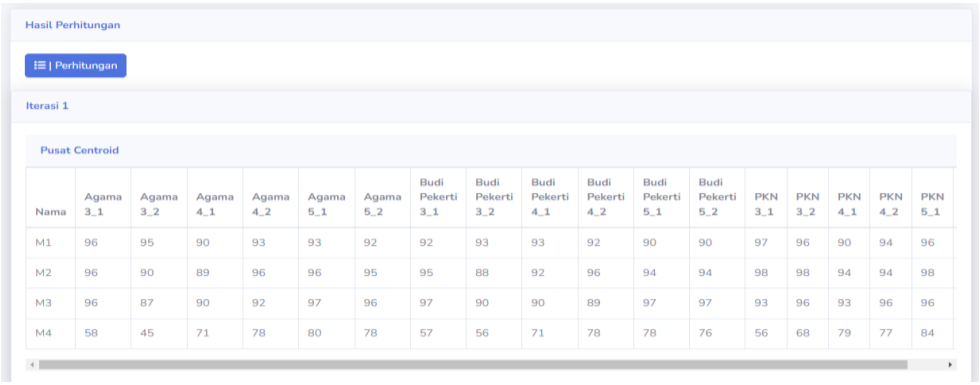

Figure 7. Data calculation page

This page is used to display data that has been successfulprocessed using the system.

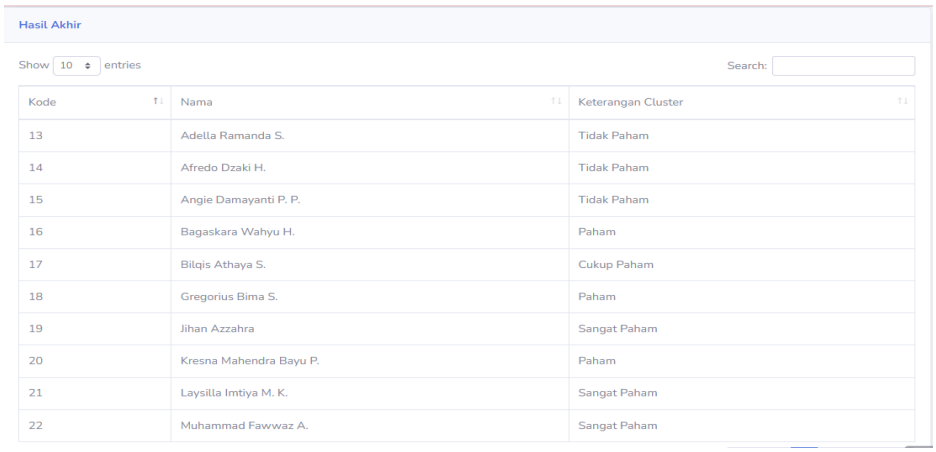

Figure 8. Final result page

4. CONCLUSION

Educational Data Mining for Mapping Understanding of Personality Subjects with the K-Means Algorithm (Case Study: SD Insan Mulya) can be used as a solution to problems to map students' understanding of personality subjects. The conclusions obtained in this research are the subjects of Moral Education, Religious Education, Citizenship Education are included in the Subjects that can affect the student's personality. The final grades obtained by students are a benchmark for the level of understanding of personality subjects. Religious subjects are subjects that have a high influence compared to the other two subjects. With this system, it is hoped that it can help teachers to monitor the level of success in learning, especially religious subjects. The percentage of clusterization results with interviews is 89.39% while the percentage of software is 91.70%. Where the percentage of Software is higher than the percentage of interviews. The K-Means Algorithm method was successfully implemented into the Educational Data Mining program for mapping the Understanding of Personality Subjects as evidenced by a high level of accuracy. The results of personality class that can be used as material for teacher/school consideration to made policies and decisions as well as an early warning for students.

From the results of application testing, it can be suggested that this application still needs development, and is still very simple, it needs a lot of improvement, especially on systems that still use the web, which is expected to be better developed in the future. The author suggests that for future research, problems like this can be developed by combining educational data mining with machine learning.

REFERENCES

Aldino, A. A., Darwis, D., Prastowo, A. T., & Sujana, C. (2021). Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency. *Journal of Physics: Conference Series*, *1751*(1), 012038. https://doi.org/10.1088/1742-6596/1751/1/012038

Dewi, S., Defit, S., & Yunus, Y. (2020). Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means (Studi Kasus SMP Pembangunan Laboratorium UNP). *Jurnal Sistim Informasi Dan Teknologi*, *3*, 7–10. https://doi.org/10.37034/jsisfotek.v3i1.98

Hidayat, M. M., Purwitasari, D., & Ginardi, H. (2013). Analisis Prediksi Drop Out Berdasarkan Perilaku Sosial Mahasiswa Dalam Educational Data Mining Menggunakan Jaringan Syaraf Tiruan. *J. IPTEK*, *17*(2), 109–119.

Hussain, M., Liu, S., Ashraf, U., Ali, M., Hussain, W., Ali, N., & Anees, A. (2022). Application of Machine Learning for Lithofacies Prediction and Cluster Analysis Approach to Identify Rock Type. *Energies*, *15*(12), Article 12. https://doi.org/10.3390/en15124501

Janusz, A., Jamiołkowski, A., & Okulewicz, M. (2022). Predicting the Costs of Forwarding Contracts: Analysis of Data Mining Competition Results. *2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS)*, 399–402. https://doi.org/10.15439/2022F303

Khan, A., & Ghosh, S. K. (2021). Student performance analysis and prediction in classroom learning: A review of educational data mining studies. *Education and Information Technologies*, *26*(1), 205–240. https://doi.org/10.1007/s10639-020-10230-3

Lemay, D. J., Baek, C., & Doleck, T. (2021). Comparison of learning analytics and educational data mining: A topic modeling approach. *Computers and Education: Artificial Intelligence*, *2*, 100016. https://doi.org/10.1016/j.caeai.2021.100016

Tomaskova, H., & Tirkolaee, E. B. (2021). Using a Process Approach to Pandemic Planning: A Case Study. *Applied Sciences*, *11*(9), Article 9. https://doi.org/10.3390/app11094121

Ulfah, M., & Irtawty, A. S. (2023). Implementation of Data Mining Clustering Using the K-Means Method in Grouping Library Books. *International ABEC*, 188–194.

Xiao, W., Ji, P., & Hu, J. (2022). A survey on educational data mining methods used for predicting students' performance. *Engineering Reports*, *4*(5), e12482. https://doi.org/10.1002/eng2.12482

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, *9*(1), 11. https://doi.org/10.1186/s40561-022-00192-z