Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th February 2025. Vol. 4. No. 2; e-ISSN: 2808-4519

Grade 9 Student Grade Data Analysis Based on Grades Knowledge to Determine Outstanding Students with K-Means Clustering (Case Study of SMPN 2 Sindang)

Putri Andriani[1*], Nana Suarna[2], Irfan Ali[3], Dodi Solihudin[4]

[1,2,3,4]*STMIK IKMI Cirebon*

putrekpipol@gmail.com[1*] , st_nana@yahoo.com[2] , irfanaali0.0@gmail.com[3] , dodisolihudin@gmail.com[4]

Abstract

Increasing student academic achievement is one of the main challenges in the world of education. Students have diverse abilities, so a directed learning approach is needed that suits each individual's needs. At SMPN 2 Sindang, this challenge is of particular concern in improving the achievement of grade 9 students.This research aims to overcome this problem by applying the K-Means Clustering algorithm to group students based on their academic score patterns. The data used includes midterm exam scores and attendance. The K-Means algorithm was used to identify three main groups: Cluster 1 (high value), Cluster 0 (medium value), and Cluster 2 (low value).
The research results show that the K-Means algorithm is effective in identifying groups of students with similar learning needs. This grouping allows schools to design learning programs that are more effective and suit the needs of each group of students. Thus, this research makes an important contribution to the development of data-based learning methods, especially at the secondary education level.

Keywords: *K-Means Clustering; Education; Academic Achievement.*

1. Introduction

Improving student achievement is one of the important focuses in the world of education. Academic achievement is an indicator of student success in understanding the material given. However, many schools face challenges in identifying students' specific needs to improve their performance. At SMPN 2 Sindang, grouping students based on academic scores is a strategic step to develop a more effective and targeted learning program.

One approach that can be used to overcome this problem is the application of the K-Means Clustering algorithm. This algorithm is a data mining method that aims to group data into several clusters based on certain similarities. In clustering, data that has a high level of similarity is grouped in one cluster, while data that is significantly different is in another cluster.

K-Means Clustering uses an iterative approach to determine the centroid as the center of the cluster. In the context of this study, the K-Means algorithm was applied to group 9th grade students at SMPN 2 Sindang based on their academic scores. Which can form three main clusters: students with high, medium, and low scores.

Based on the study of the application of the K-Means Clustering algorithm, it is able to group students into low, sufficient, and high categories, thus helping educators in designing learning strategies that are appropriate to the needs of each group. The use of this method allows schools to obtain more structured information regarding the distribution of student academic achievement [13]. Previous studies have emphasized the importance of utilizing the K-Means algorithm for mapping learning groups, which can help schools understand variations in student abilities and provide more targeted interventions [3]. This algorithm is able to group student data with similar characteristics, making it easier for schools to determine effective learning strategies. This study aims to provide technology-based solutions in the world of education. The results are expected to help schools develop relevant learning strategies, support the needs of each group of students, and encourage significant improvements in academic achievement. In addition, this study also shows how the K-Means algorithm can provide accurate information for decision making in the field of education.

1.1. Data Mining

This is the core stage where data mining techniques are applied to find patterns or hidden information in the data. In this study, the K-Means algorithm was used to cluster student data based on their academic performance, identifying groups that have certain similarities in their grades.

1.2. Clustering

Clustering is a data analysis method that aims to group data into several groups (clusters) based on certain similarities or characteristics. In clustering, data that has similar characteristics will be grouped into one cluster, while data that has significant differences will be put into different clusters.

1.3. K-Means

K-Means is one of the most popular clustering algorithms in machine learning and data analysis. This algorithm aims to divide data into K clusters (groups) based on the similarity between the data, so that each data in a cluster has a higher similarity to data in the same cluster compared to data in other clusters.

2. Research Method

2.1. Research Methods

Research methodology is a systematic framework or approach used in research to collect, analyze, and interpret data to answer questions or achieve research objectives. This methodology includes strategies, methods, and techniques used to ensure that research is conducted in a valid, reliable, and accountable manner.

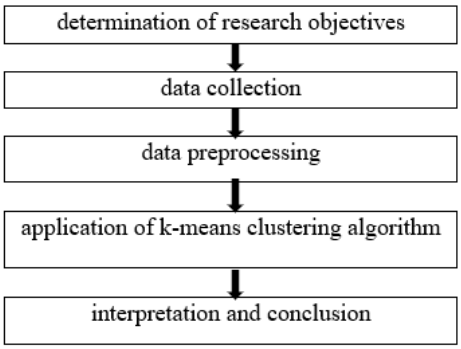

Fig. 1: Research Workflow

1. Determination of research objectives; Identification of objectives and research to analyze the grouping of academic achievement of grade 9 students at SMPN 2 Sindang using the k-means clustering algorithm.
2. Data collection; Collects student academic grade data and school achievement records, including mid-term exam data.
3. Data preprocessing; Perform data cleansing, handle missing or incomplete data, and data normalization to ensure quality and consistency in analysis.
4. Application of k-means clustering algorithm; Applying the k-means clustering algorithm to group students based on their academic scores into several categories, such as high, medium, and low.
5. Interpretation and conclusion; Analyze the grouping results to identify student achievement patterns and draw conclusions that can be used to support the creation of learning programs that are appropriate to student achievement categories.

3. Result and Discussion

3.1. Result

This study uses the K-Means Clustering algorithm to analyze the achievement data of 9th grade students at SMPN 2 Sindang. The purpose of this study is to group students based on their scores in various subjects so that the school can understand the pattern of academic achievement and design appropriate learning programs for each achievement group.

The dataset used in this study consists of 200 student records that include information such as student name and grade. The data were collected from exam results in science subjects.

Table 1: Dataset

No	Nama Siswa	Nilai Uts IPA	Nilai Uts IPS
1	Ahmad Bagus	90	89
2	Ahmad Rifai	90	94
3	Andi Gunawan	90	95
4	Alfi Nuzulil	95	90
5	Ayu Khoirunnisa	85	92
6	Bagas Saputra	90	94
7	Bagus Utomo	90	93
8	Cantika Ayu	95	87
..	……	….	

200	Zaiq Al-Farizky	60	65

Data Selection at this stage, relevant data for analysis is selected, namely only data that includes grades from subjects. Irrelevant attributes, such as student names, are removed so that the analysis focuses more on the attributes "student attendance and student grades".

Table 2: Attributes used for clustering

No.	Atribut	Type Data	Keterangan
1.	Kelas	Integer	Kategori kelas yang diikuti siswa, direpresentasikan dengan angka.
2.	Kehadiran Siswa	real	Persentase kehadiran siswa dalam semester
3.	Nilai UTS	Integer	Nilai siswa dalam angka

Data Pre-Processing this stage involves the data cleaning process, including filling in missing values and removing duplicate data. Because this research has no missing data or duplicate data, the following is the pre-processing data.

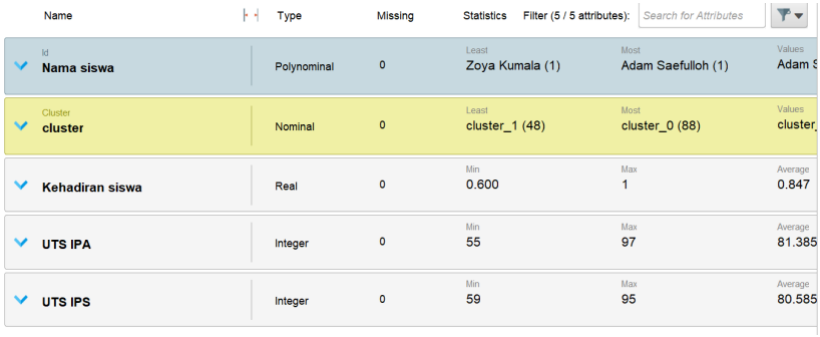

Fig. 2: Data Pre-processing

The author uses the K-Means Clustering algorithm to group data according to certain patterns based on predetermined characteristics. This process is carried out using RapidMiner, which provides a visual interface and automated modules to facilitate the implementation of the K-Means algorithm. Starting the Read Excel operator process which is the most basic used in the RapidMiner application functions as an Excel file reader.

Fig. 3: Operator Read Excel

The use of the K-Means Algorithm operator to group the mid-term exam score data of grade 9 students at SMPN 2 Sindang. In this process, the Cluster Distance Performance operator is used to evaluate the quality of the cluster and determine the best number of clusters (K). In this case, the K value is set to 3 to see the variation of clustering that might provide optimal results.

Fig. 4: Operator K-Means Clustering

The Cluster Distance Performance operator in the K-Means algorithm is used to evaluate the distance between data in each cluster, as well as how well the cluster represents the existing data. The smaller the distance between data points and centroids, the better the quality of the clusters formed.

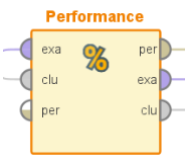

Fig. 5: Operator Performance

At the data mining stage, researchers will use the K-Means Clustering algorithm in RapidMiner to build learning groupings at SMPN 2 Sindang.

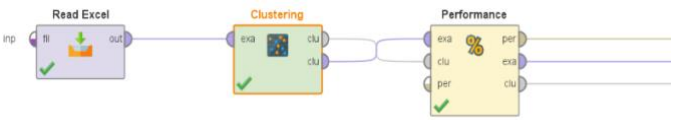

Fig. 6: Data Mining

After processing data using the K-Means Clustering algorithm in the Rapidminer application, the next step is to evaluate the results that have been made. The evaluation stage uses the Performance operator (Cluster Distance Performance) with the application of Davies Bouldin's main criteria on the parameters.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -23.576
Avg. within centroid distance_cluster_0: -18.526
Avg. within centroid distance_cluster_1: -40.842
Avg. within centroid distance_cluster_2: -17.570
Davies Bouldin: -0.479
```

Fig. 7: Performance Model

Based on the evaluation results above, it is known that the optimal DBI value is possessed by an experiment with a total of 3 (three) clusters with a DBI value of 0.479.

4.2. Discussion

Based on the evaluation results, it provides an overview of the scatter plot visualization grouping on the UTS IPS. Each point on the graph represents individual data grouped based on the clustering results, with different colors indicating cluster membership. Where cluster 0 has the highest and best achievement scores than the others, cluster 2 has moderate achievement scores and the last cluster 1 has low achievement scores.

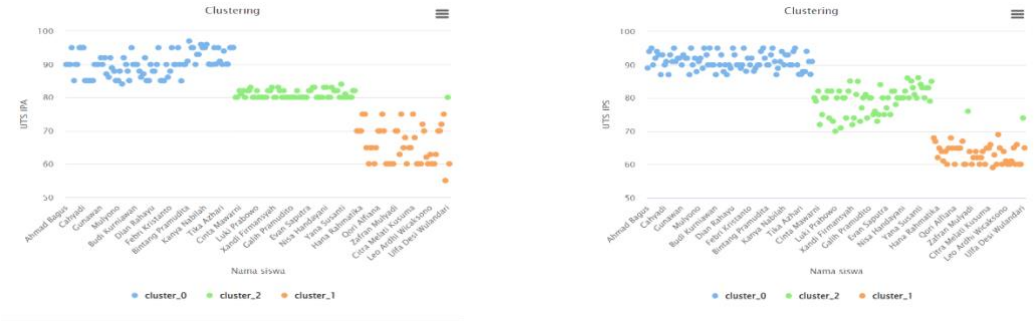

Fig. 8: Scatter plot visualization on UTS IPA & UTS IPS

Explanation scatter plot visualization:

1. Cluster 0 consists of high-achieving students, who show consistency in academic achievement. The strategy for this cluster is to provide additional challenges to encourage higher potential.
2. Cluster 2 consists of students with average achievement, who need more motivation in learning. Interventions in the form of learning assistance can help improve their results.
3. Cluster 1 consists of students with low achievement. The focus for this cluster is to provide additional support, such as remedial programs and intensive tutoring.

4. Conclusion and Suggestions

4.1. Conclusion

This study uses the K-Means Clustering algorithm to group 9th grade students at SMPN 2 Sindang based on their academic achievement. Based on the results of the analysis, students were successfully grouped into three main categories: high, medium, and low achieving students. Thus, the K-Means algorithm has proven to be effective as a tool for understanding student achievement patterns, so that it can

be used by schools as a basis for decision making in designing more targeted and efficient learning strategies. Strategies that are tailored to the needs of each group of students can contribute to improving overall academic achievement.

4.2. Suggestions

1. Utilization of Clustering Results, Schools are expected to use the results of this clustering to design more focused learning programs, such as additional classes for low-achieving students or enrichment activities for high-achieving students.
2. Development of Data-Based Systems, It is recommended that schools integrate the K-Means algorithm into the school information system, so that the clustering process can be carried out automatically and periodically to obtain more up-to-date data.
3. Improvement of Research Variables, This research is limited to academic grades as parameters. Further research can consider other variables, such as attendance, extracurricular participation, and psychological aspects of students, to obtain a more comprehensive picture.
4. Teacher Training, Teachers need to be given training to understand the results of clustering and how to apply them in learning that suits the needs of each group of students.

References

[1] Ishak R. Volume 5 Nomor 2 Juli 2023 Penerapan Algoritma XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah. Jambura Journal of Electrical and Electronics Engineering 192.
[2] Ishak R. Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means.
[3] Fabrianto L, Riza F, Faizah NM. KORELASI ANTARA PROFIL DAN NILAI AKADEMIS SISWA DENGAN MENGGUNAKAN ALGORITMA K-MEANS.
[4] Yuliana DT, Fathoni MIA, Kurniawati N. Penentuan Penerima Kartu Indonesia Pintar KIP Kuliah Dengan Menggunakan Metode K-Means Clustering. Journal Focus Action of Research Mathematic (Factor M) 2022;5(1):127–141.
[5] Febriansyah F, Muntari S. Penerapan Algoritma K-Means untuk Klasterisasi Penduduk Miskin pada Kota Pagar Alam. 2023;
[6] Febrinita F, Dwi Puspitasari W, Zaman WI. Klasterisasi Hasil Belajar Matematika dengan Algoritma K-Means Clustering.
[7] Novi ., Mubarok A. Penerapan Algoritma K-Means untuk Menentukan Kelas Unggulan di SMP Pelita Bandung. Infomatek 2021;32(2):97–106.
[8] Puspita Sari Y, Primajaya A, Susilo Yuda Irawan A, et al. Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang. 5(2):2020.
[9] Ishak R. Volume 5 Nomor 2 Juli 2023 Penerapan Algoritma XGBoost untuk Seleksi Atribut pada K-Means dalam Clustering Penerima KIP Kuliah. Jambura Journal of Electrical and Electronics Engineering 192.
[10] Saputra DA, Pakereng MAI. Analisis Data Nilai Siswa Kelas 8 Berbasis Nilai Pengetahuan Untuk Menentukan Siswa Berprestasi dengan K-Means Clustering (Kasus SMP Negeri 4 Salatiga). 2023;
[11] Isnanto S, Widodo S. PENERAPAN DATA MINING PADA PENERIMAAN MAHASISWA BARU DENGAN ALGORITMA K-MEANS CLUSTERING. Jurnal Teknik Informasi dan Komputer (Tekinkom) 2021;4(2):158.
[12] Abdul Majid MB, Cani YM, Enri U. Penerapan Algoritma K-Means dan Decision Tree Dalam Analisis Prestasi Siswa Sekolah Menengah Kejuruan. Jurnal Sistem Komputer dan Informatika (JSON) 2022;4(2):355.
[13] Sholeh M, Andayati D. Penerapan Metode Clustering dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa. SKANIKA: Sistem Komputer dan Teknik Informatika 2023;6(1):51–60.
[14] Saputra EA, Nataliani Y. Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. Journal of Information Systems and Informatics [homepage on the Internet] 2021;3(3). Available from: http://journal-isi.org/index.php/isi
[15] Dewi S, Defit S, Yuhandri Y. Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means. Jurnal Sistim Informasi dan Teknologi 2021;28–33.