Classification of Student Grade Data Using the K-Means Clustering Method

Lanjar Pamungkas[1], Nur Aela Dewi[2]\*, Nessia Alfadila Putri[3]

Information System, Faculty of Informatics[1], [2], [3]

Telkom Purwokerto Institute of Technology

Purwokerto, Banyumas, Indonesia

20103043@ittelkom-pwt.ac.id[3], 20103064@ittelkom-pwt.ac.id[1], [20103061@ittelkom-pwt.ac.id[2](mailto:20103061@ittelkom-pwt.ac.id[2)],

***Abstract*—** **The advent of industrial revolution 4.0 is bringing rapid changes in education thanks to advances in technology and automation. These changes affect lifestyles and learning processes. In the education sector, challenges arise in the automation of student-grade data processing due to large volumes of data. One of the relevant approaches to overcome this problem is to classify the data. In this study, student grade data is structured using the K-Means Clustering technique. The purpose of this research is to understand the pattern of students' academic behavior and provide recommendations to improve the quality of learning. This study encompasses the gathering of student grade data, the application of the K-Means Clustering algorithm, and the use of the Elbow method for testing to ascertain the most optimal quantity of clusters. This study sampled is student grade data from the Information System Study Programme at Telkom Institute of Technology Purwokerto. The findings indicated that students could be categorized into three segments: "Diligent" with high academic performance (GPA: 3.74 - 3.96, SKS Taken: 63 - 150), "Not bad" with fairly good performance (GPA: 2.92 - 3.72, SKS Taken: varied), and "Lazy" with low performance (GPA: 2.35, SKS Taken: 14). The application of the K-Means Clustering technique and the Elbow method testing facilitated the effective recognition of patterns in student academic conduct. The findings provide valuable insights for educational institutions to make informed decisions in improving learning and providing better support for students.**

***Keywords—*** ***Clustering, Data, Education, Elbow, K-Means***

***Abstrak*— Munculnya revolusi industri 4.0 membawa perubahan pesat dalam dunia pendidikan berkat kemajuan teknologi dan otomasi. Perubahan ini mempengaruhi gaya hidup dan proses belajar. Di sektor pendidikan, tantangan muncul dalam otomatisasi pemrosesan data tingkat siswa karena besarnya volume data. Salah satu pendekatan yang relevan untuk mengatasi masalah ini adalah dengan mengklasifikasikan data. Pada penelitian ini data nilai siswa distrukturisasi dengan teknik K-Means Clustering. Tujuan penelitian ini adalah untuk memahami pola perilaku akademik siswa dan memberikan rekomendasi untuk meningkatkan kualitas pembelajaran. Penelitian ini meliputi pengumpulan data nilai siswa, penerapan algoritma K-Means Clustering, dan penggunaan metode Elbow untuk pengujian guna memastikan jumlah cluster yang paling optimal. Sampel penelitian ini adalah data nilai mahasiswa Program Studi Sistem Informasi Institut Teknologi Telkom Purwokerto. Temuan menunjukkan bahwa mahasiswa dapat dikategorikan menjadi tiga segmen: “Rajin” dengan prestasi akademik tinggi (IPK: 3,74 – 3,96, SKS Diambil: 63 – 150), “Lumayan” dengan prestasi cukup baik (IPK: 2,92 – 3,72, SKS Diambil: bervariasi), dan “Malas” dengan prestasi rendah (IPK: 2,35, SKS Diambil: 14). Penerapan teknik K-Means Clustering dan pengujian metode Elbow memfasilitasi pengenalan pola yang efektif dalam perilaku akademik siswa. Temuan ini memberikan wawasan berharga bagi institusi pendidikan untuk mengambil keputusan yang tepat dalam meningkatkan pembelajaran dan memberikan dukungan yang lebih baik bagi siswa.**

***Kata Kunci—*** ***Clustering, Data, Pendidikan, Elbow, K-Means***

# INTRODUCTION

The epoch of the fourth industrial revolution, marked by swift progress in technology and machinery, has induced substantial transformations in diverse aspects of life. One of the sectors most affected is the world of education. In this context, these changes force the world of education to adapt to the major changes currently occurring[1]. With the rapid development of technology in this modern era, automation processes have become increasingly common in various aspects of life. This leads to reduced human involvement in many jobs that previously required manual intervention. One sector that has adopted the application of automation quite significantly is the education sector. Through automation, the education sector has undergone a transformation that has resulted in convenience and efficiency in various aspects[2]. The application of automation in the education sector has brought great benefits, especially in processing student grade data. This data is not only valuable for increasing understanding of student academic performance and achievement, but also becomes an important asset for educational institutions. However, the increasing abundance of data collected has brought new challenges in processing it efficiently and effectively[3].

In the era of advances in information technology and increasingly widespread use of data, it cannot be denied that data classification has become a very relevant and important approach to facing complex challenges. Through the data classification process, the main goal to be achieved is to carefully group data into various relevant categories or classes, based on patterns or special features contained in the dataset. In the context of the education sector, the application of advanced data classification techniques provides a major opportunity to improve understanding, decision making, and efficiency in managing student grades data, opening the door to significant improvements in education delivery[4]. The ability to have a model or function that is able to accurately differentiate between various classes of data has opened a wide door to increasing in-depth understanding of the characteristics of students in certain groups in the world of education. Using these models, it is possible to gain deeper insight into student academic behavior patterns, identify significant trends, and identify challenges that need to be addressed. More than just analysis, these capabilities make it possible to provide appropriate recommendations and appropriate actions to improve the quality of learning and assessment, bringing meaningful improvements to sustainable education systems.

K-Means clustering is a technique employed to categorize data into groups in a classifying. This technique is commonly employed in research and falls under the category of clustering methods in Data Mining[5]. The K-Means Clustering technique is utilized for categorizing student score data. Through experimentation and analysis, it is hoped that it can reveal valuable insights about academic patterns that provide benefits for educational institutions in improving the learning process and student guidance. The decision-making process regarding the results of evaluations and assessments of students is carried out by lecturers during the learning process. The ratings for the "Lazy," "Decent," and "Diligent" categories are determined based on several factors, including the number of credits taken, credits not yet taken, number of D grades, number of E grades, and the student's GPA. The "Lazy" category is given if a student has a lot of credits not taken, a high number of D and E grades, and a low GPA. Meanwhile, the "Excellent" category is given to students who have made fairly good progress, with an adequate number of credits taken, a moderate number of D and E grades, and a satisfactory GPA. Students who are included in the "Diligent" category are students who have completed many credits, have a small number of D and E grades, and have a high GPA. This category assessment helps describe academic performance and the level of student seriousness in facing learning challenges.

Data mining, as an increasingly important concept in the digital information era, is a very significant process in modern data analysis. This process involves searching, mining, and analyzing large amounts of data stored in repositories with a very systematic approach. Through data mining, the main goal is to discover hidden relationships, unexpected patterns, and new trends that may not be discovered by conventional analysis methods. Thus, the data mining process plays a central role in extracting previously unknown, but of high significance, valuable information from large and complex databases. In this context, data mining is not just an analytical tool, but is also the key to uncovering deep knowledge and has the potential to change the way we understand and make decisions in various fields, from business to science[6]. This complex data mining process relies on sophisticated pattern recognition technology, and utilizes a variety of powerful statistical and mathematical techniques. The goal is to uncover hidden patterns and knowledge that has significant value from within this large amount of data. By combining computational intelligence, in-depth mathematical analysis, and specialized algorithms, this process becomes a highly effective tool in unraveling the complexity of modern data. Thus, data mining is not just an analytical tool, but also an important bridge to deeper understanding and more efficient problem solving in a variety of contexts, from business to scientific research[7]. Data Mining is an important concept in data analysis, which clearly describes the process of searching and discovering hidden information or knowledge in a database. This process involves the use of a variety of methods and techniques, including statistics, mathematics, and pattern recognition technology, with the aim of unearthing valuable insights that may not be discovered through conventional analysis. Thus, Data Mining is not just an analytical tool, but also a means that makes it possible to explore the enormous potential of available data, help make better decisions, and understand complex phenomena in various areas of life[8].

Clustering is an important process in data analysis which aims to group records, observations, or classes based on the similarity of the objects contained in them. Using various statistical methods and techniques, clustering helps identify emerging patterns in large and complex data, ultimately aiding in a deeper understanding of the relationships between those objects. This process makes it possible to recognize similar groups, thereby facilitating analysis, decision making, and understanding of phenomena in various fields, including science, business, and technology[9]. Clustering is a method employed to categorize data into clusters or sets according to shared characteristics among the data. Multiple methodologies, including k-means, k-medoid, k-mode, hierarchical clustering, and alternative techniques, can be utilized to organize the data into distinct groups. Clustering methods have unique advantages and disadvantages. The formation of optimal clusters can be impacted by factors such as the clustering method applied, dataset characteristics, data density structure, dataset volume, and the specified number of clusters. Besides various clustering methods, there are also multiple techniques to identify the appropriate number of clusters, one of which includes the use of the Elbow Method. This technique involves graphing the explained variation in relation to the number of clusters, and choosing the "elbow" point on the curve as the most suitable number of clusters to employ[10].

The Elbow Method is employed to ascertain the most suitable value for k when creating a cluster[11]. The Elbow Method is when the point forms a sharp angle. This point indicates a significant decrease between the two cluster points followed by a value that tends to remain constant. The table above shows the SSE (Sum of Squares Error) values for the most important or largest errors[12]. The Elbow Method is used to determine the most suitable number of clusters by contrasting outcomes across varying numbers of clusters, where a notable percentage shift creates an elbow at a certain point[13].

K-Means is a method utilized to categorize data into two or more groups, employing the mean value as the central point for each group[14]. K-Means is a non-hierarchical clustering technique designed to assist in categorizing variables into predefined classes upon completion of the computation. The K-Means algorithm stands out as one of the most straightforward and extensively applied implementations of partitional clustering algorithms. K-Means uses a squared error criterion. This algorithm starts by partitioning the data space randomly while assigning existing samples into clusters based on the similarity between clusters and samples, until a convergent criterion is found[15].

# LITERATURE REVIEW

* 1. Data

Data is a raw source of information that reflects the realities of events and entities. It can take the form of symbols such as letters, numbers, images, sounds, signals, and more. For it to be useful, data needs to undergo processing, and the outcomes can transform into information. Essentially, data is a record of facts, concepts, or instructions that undergoes a process to become understandable to humans[16].

* 1. Clustering

Clustering, a method within the realm of data mining, is utilized for the analysis of data with the primary objective of resolving issues pertaining to data categorization. More precisely, it involves the partitioning of a dataset into distinct subsets. In the realm of clustering techniques, the emphasis is placed on the allocation of cases (be they objects, individuals, events, etc.) into specific groups. This is done to ensure that the level of connectivity among members within the same cluster is robust, while the strength of connections between members belonging to different clusters exhibits variability ranging from strong to weak[17].

* 1. K-Means

K-Means is a non-hierarchical data clustering technique that divides data into one or more clusters. It groups data with similar features into a single cluster, while placing data with distinct characteristics into separate clusters. Functioning as a distance-based clustering method, K-Means algorithmically separates data into multiple clusters, focusing specifically on numeric attributes.[18].

* 1. K-Means Clustering

The K-Means clustering method, first presented by Stuart Lloyd from Bell Labs in 1957, encompasses an Unsupervised Learning strategy marked by a repetitive process. In this method, the dataset is partitioned into a predetermined number, k, of distinct clusters or subgroups. The iterative nature of the algorithm aims to group data points closely together within each cluster while maintaining separation among clusters in different spaces. It designates data points to clusters in such a way that it reduces the total of squared distances between the cluster centers and the data points. Consequently, the centroid of a cluster serves as the average value for the data points within that specific cluster[19]. Generally, the steps of the K-Means algorithm proceed as follows: general, the steps of the K-Means algorithm are as follows:

1. Establish the preferred number of clusters, represented as k.
2. Select k objects randomly to function as the central points or centroids for the clusters.
3. Identify the k centroids, representing the central positions.
4. Categorize objects into the nearest cluster centroid, utilizing Euclidean distance as the determining factor:

(1)

Description:

x\_i denotes the i-th object in the set x,

y\_i signifies the i-th object in the set y, and

n represents the total number of objects.

* 1. Elbow Method

The elbow method is a clustering technique utilized to ascertain the most optimal number of clusters. Its name is derived from the graph it generates, which usually shows a bend, referred to as the optimal cluster count. The graph illustrates various percentage results for each cluster value, providing valuable insights. If the graph shows a noticeable bend with a gradual decrease in cluster values, without a sharp decline afterward, that specific cluster value is considered the most appropriate[20].

# Research Method

In this research, several stages were carried out, including the following:

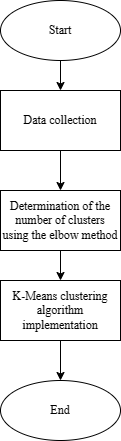


Figure 1. Research Flow Chart

Figure 1 explains the stages of the research carried out:

* 1. Data Collection

The data used in this research was obtained from student assessments taking part in the Information Systems Study Program at the Telkom Purwokerto Institute of Technology. The student score data is secondary data obtained from the score recording system from 2019 to 2022. The dataset used for this research contains the results of student score data from the Information Systems Study Program at the Telkom Purwokerto Institute of Technology as many as 550 students with the attributes of Credits Taken, GPA , Total D Grades, Total E Grades, and Credits Not Yet Taken. The K-Means algorithm, in conjunction with the Elbow method as a cluster optimization strategy, will be used to group the data. This is intended to pinpoint the most suitable number of clusters for efficient categorization.

* 1. Determining The Number of Clusters Using Elbow

The elbow method is utilized to ascertain the most suitable or optimal number of clusters in the k-means algorithm. This technique is employed to identify the elbow point on the curve of clustering results, where an increase in the number of clusters does not yield a considerable enhancement in the quality of clustering. The elbow method allows identifying the most suitable number of clusters for the k-means algorithm. In research involving assessment data from students from the Information Systems Study Program at the Telkom Purwokerto Institute of Technology, the elbow method strategy was utilized as the optimal means to identify the most suitable number of clusters.

* 1. Implementation of the K-Means Clustering Algorithm

The implementation of the K-Means clustering algorithm in this study is intended to categorize the evaluation data of students participating in the Information Systems Study Program at the Telkom Purwokerto Institute of Technology into various groups, according to the existing attributes. The implementation of K-Means aims to help data analysis and further understanding of student group patterns based on value data. Clustering results can provide valuable insight for academic decision making or the implementation of more appropriate educational strategies.

# RESULT AND DISCUSSION

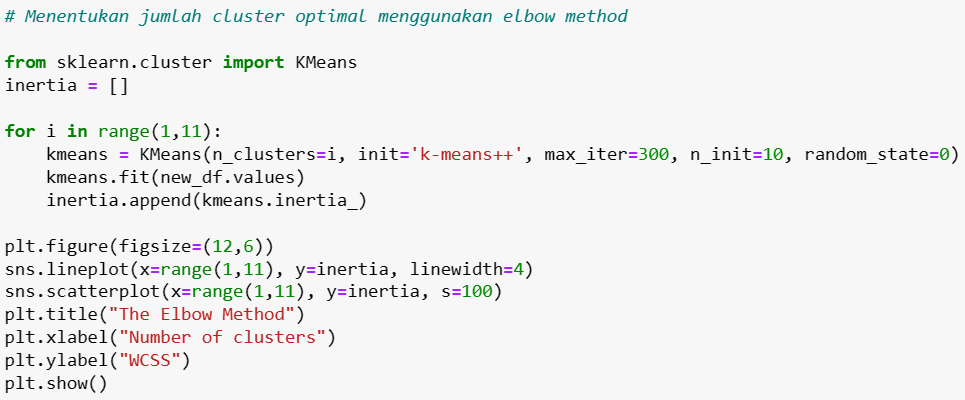
The data used in this research comes from the scores of students from the Information Systems Study Program at the Telkom Purwokerto Institute of Technology. The student score data is secondary data obtained from the score recording system from 2019 to 2022. The dataset used for this research contains the results of student score data from the Information Systems Study Program at the Telkom Purwokerto Institute of Technology as many as 550 students with the attributes of Credits Taken, GPA , Total D Grades, Total E Grades, and Credits Not Yet Taken. The following example of student data used can be seen in table 1.

TABLE 1. STUDENT VALUE DATASET

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **SKS Taken** | **IPK** | **Number of values D** | **Number of values E** | **SKS Not Taken** |
| 151 | 2,46 | 3 | 4 | 21 |
| 136 | 2,92 | 3 | 0 | 8 |
| 138 | 3,77 | 0 | 0 | 6 |
| 140 | 3,56 | 0 | 0 | 4 |
| 132 | 1,69 | 6 | 21 | 67 |
| 140 | 3 | 0 | 0 | 4 |
| 141 | 2,68 | 3 | 6 | 14 |
| 140 | 2,97 | 1 | 0 | 4 |
| 85 | 1,46 | 6 | 17 | 32 |
| 120 | 1,72 | 1 | 17 | 57 |

The results of determining the number of clusters using elbows are displayed in the form of SSE (Sum of Squares Error) value data and also in the form of a graph that has an x-axis and a y-axis. On the x-axis, there are variations in the number of clusters ranging from 2 to 10 clusters, while on the y-axis, there are SSE values derived from each cluster. The test outcomes indicate that two clusters are the optimal number of clusters suitable for segregating student data into the "Lazy", "Not bad" and "Diligent" groups.

Upon examining the test outcomes utilizing the elbow method, a graph is revealed displaying the position of the "elbow", which signifies the optimal number of clusters. The location of the "elbow" will dictate the best number of clusters to be employed in the procedure of categorizing student grade data using the K-Means algorithm.



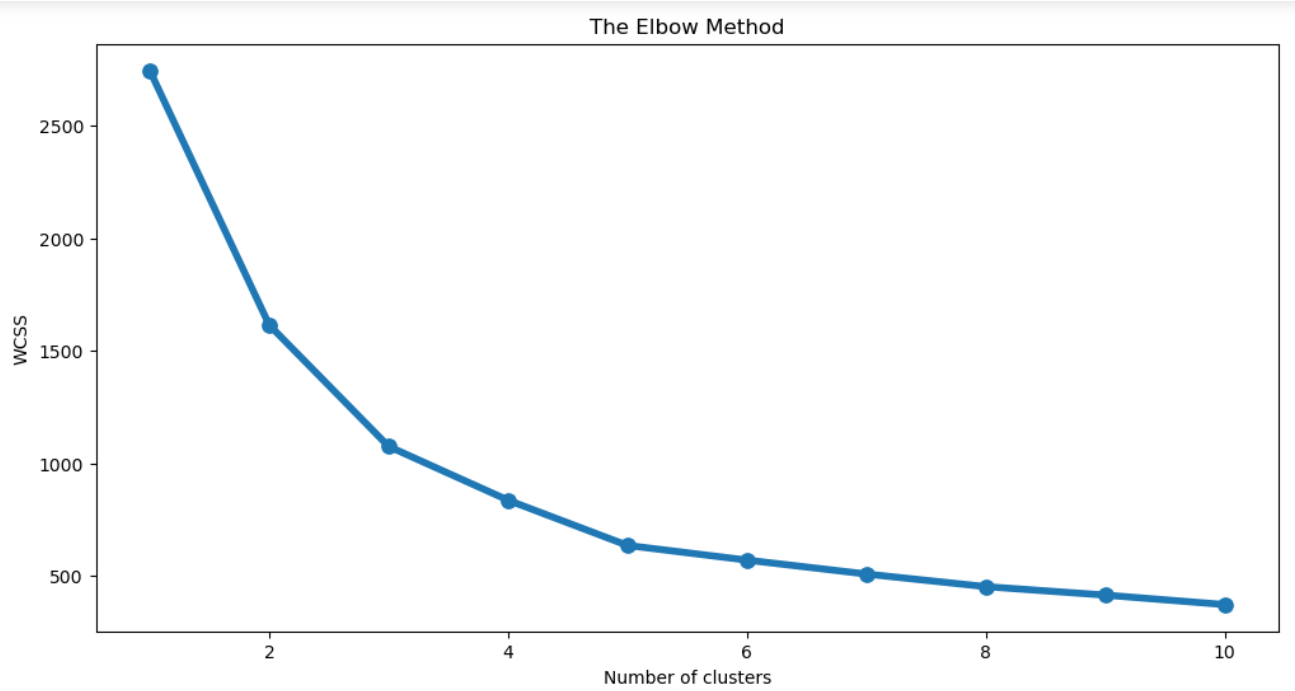


Figure 2. Results of Determining the Number of Clusters

The next step is to standardize the data. Standardization is a process to rescale data so that it has a mean (average) of 0 and a standard deviation of 1. This process is carried out so that all attributes have balanced weights in the clustering process and prevent attributes with a large scale from dominating in distance calculations, to standardize the attributes of SKS Terambil, IPK, Jumlah Nilai D, Jumlah Nilai E, and SKS Belum Terambil in the dataset of student grades. After standardization, the data will be ready to be implemented using the K-Means clustering algorithm to group student grade data into several groups based on existing attributes.

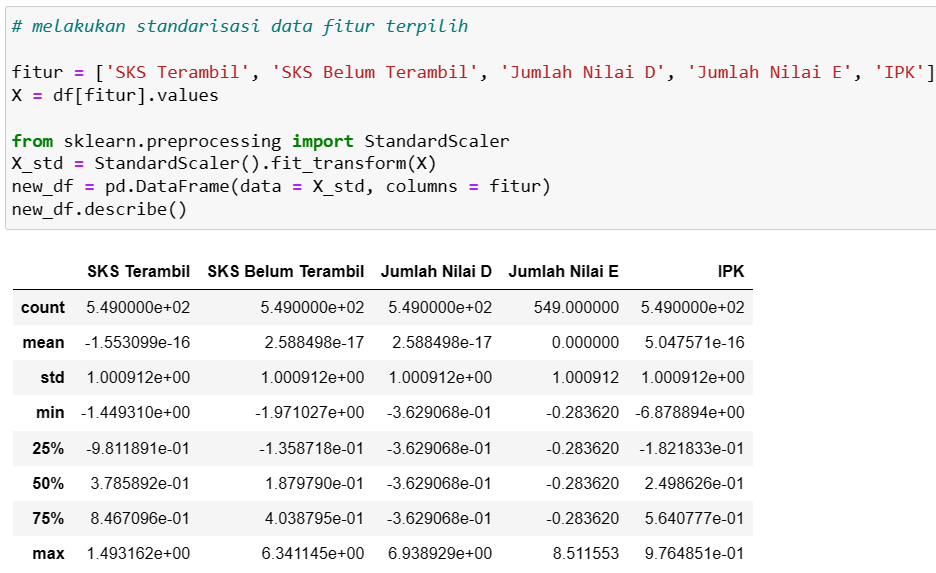


Figure 3. Standardization

The subsequent step involves forming segments based on specific criteria from the student grade data. Creating segments is done by identifying and grouping student data into segments or groups based on certain characteristics or patterns found from the clustering results. For example, segments can be created based on the range of GPA values, the number of credits taken, or the distribution of D and E grades. he outcomes of this segment creation can be observed in the table below:

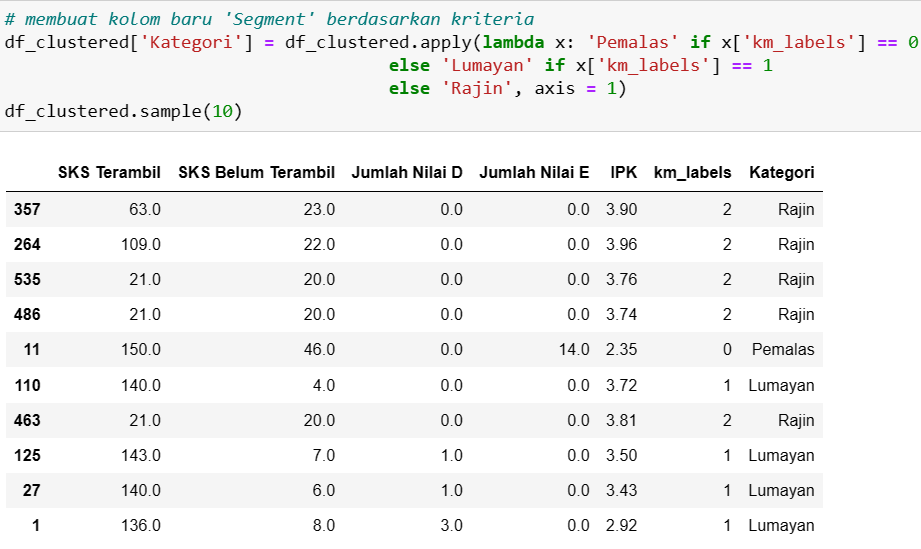


Figure 4. Criteria Segment Results

Based on the table above, student data has been classified into three segments based on certain characteristics resulting from the K-Means clustering method, namely "Lazy", "Not bad", and "Diligent".

As a result of the clustering process using the K-Means method, student data has been divided into three different segments. First, there is the "Diligent" segment consisting of five students. Students in this segment show excellent academic performance with a high number of credits taken, ranging from 63 to 150 credits. They also never got an E, which indicates consistency in completing academic assignments without bad grades. The GPA of students in this segment is also high, ranging from 3.74 to 3.96, indicating extraordinary academic achievement.

Then, there is the "Not bad" segment consisting of four students. Students in this segment have variations in the number of credits taken, but the majority of them have higher credits taken than credits not yet taken. They also show quite good academic performance with GPAs ranging from 2.92 to 3.72. Even though it is not as good as the "Diligent" segment, students in this segment still show dedication in pursuing good academic achievements.

Finally, there is the "Lazy" segment which only consists of one student. Students in this segment show low academic performance with the number of credits taken which is much lower than credits not yet taken. There are also 14 credits with a grade of E, indicating the number of bad grades in courses. This is reflected in the low GPA, namely 2.35. The "Lazy" segment reflects students who have challenges completing academic assignments and need further support to improve their academic performance.

# CONSLUSION

The K-Means Clustering method succeeded in grouping student score data into three segments: "Diligent" with very good academic performance (GPA: 3.74 - 3.96, credits taken: 63 - 150), "Not bad" with fairly good performance (GPA: 2.92 - 3.72, Credits Taken: varies), and "Lazy" with low performance (GPA: 2.35, Credits Taken: 14). Testing using the elbow method confirms that three clusters are the optimum number for dividing student groups. These results provide important insights into students' academic behavior patterns and can help educational institutions improve learning and provide appropriate support for students.

##### REFERENCES

[1] A. R. Taraju, N. Nurdin, and A. Pettalongi, “Tantangan dan Strategi Guru Menghadapi Era Revolusi Industri 4 . 0,” *Pros. Kaji. Islam dan Integr. Ilmu di Era Soc. 5.0 (KIIIES 5.0) Pascasarj. Univ. Islam Negeri Datokarama Palu*, vol. 1, pp. 314–315, 2022.

[2] T. Kurniati and N. A. Wiyani, “Pembelajaran Berbasis Information and Communication Technology pada Era Revolusi Industri 4.0,” *J. Imiah Pendidik. dan Pembelajaran*, vol. 6, no. 1, p. 182, 2022, doi: 10.23887/jipp.v6i1.41411.

[3] A. Yudhistira and R. Andika, “Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering,” *J. Artif. Intell. Technol. Inf.*, vol. 1, no. 1, pp. 20–28, 2023, doi: 10.58602/jaiti.v1i1.22.

[4] N. B. Putri and A. W. Wijayanto, “Analisis Komparasi Algoritma Klasifikasi Data Mining Dalam Klasifikasi Website Phishing,” *Komputika J. Sist. Komput.*, vol. 11, no. 1, pp. 59–66, 2022, doi: 10.34010/komputika.v11i1.4350.

[5] E. Muningsih, I. Maryani, and V. R. Handayani, “Penerapan Metode K-Means dan Optimasi Jumlah Cluster dengan Index Davies Bouldin untuk Clustering Propinsi Berdasarkan Potensi Desa,” *J. Sains dan Manaj.*, vol. 9, no. 1, pp. 95–100, 2021, [Online]. Available: https://ejournal.bsi.ac.id/ejurnal/index.php/evolusi/article/view/10428/4839

[6] M. I. T. B. N. Sumadi, R. Putra, and A. Firmansyah, “Peran Perkembangan Teknologi Pada Profesi Akuntan Dalam Menghadapi Industri 4.0 Dan Society 5.0,” *J. Law, Adm. Soc. Sci.*, vol. 2, no. 1, pp. 56–68, 2022, doi: 10.54957/jolas.v2i1.162.

[7] Z. Nabila, A. Rahman Isnain, and Z. Abidin, “Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means,” *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: http://jim.teknokrat.ac.id/index.php/JTSI

[8] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.

[9] F. Hardiyanti, H. S. Tambunan, and I. S. Saragih, “Penerapan Metode K-Medoids Clustering Pada Penanganan Kasus Diare Di Indonesia,” *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 598–603, 2019, doi: 10.30865/komik.v3i1.1666.

[10] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix J. Manaj. Teknol. dan Inform.*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.

[11] A. W. Fuadah, F. N. Arifin, and O. Juwita, “Optimasi K-Klasterisasi Ketahanan Pangan Kabupaten Jember Menggunakan Metode Elbow,” *INFORMAL Informatics J.*, vol. 6, no. 3, p. 136, 2021, doi: 10.19184/isj.v6i3.28363.

[12] R. Yuliana Sari, H. Oktavianto, and H. Wahyu Sulistyo, “Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia K-Means Algorithm With Elbow Method To Grouping District/City in Central Java Based on Components of Human D,” *J. Smart Teknol.*, vol. 3, no. 2, pp. 2774–1702, 2022, [Online]. Available: http://jurnal.unmuhjember.ac.id/index.php/JST

[13] A. R. Said, D. Arifianto, and H. A. Al Faruq, “Pengelompokan Kecamatan Di Kabupaten Jember Berdasarkan Tanaman Pangan Dengan Algoritma Fuzzy C-Means Dan Metode Elbow,” *J. Smart Teknol.*, vol. 2, no. 1, pp. 1–12, 2020.

[14] D. Widyadhana, R. B. Hastuti, I. Kharisudin, and F. Fauzi, “Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah,” *Prism. Pros. Semin. Nas. Mat.*, vol. 4, pp. 584–594, 2021, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[15] S. Sonang, A. T. Purba, and F. O. I. Pardede, “Pengelompokan Jumlah Penduduk Berdasarkan Kategori Usia Dengan Metode K-Means,” *J. Tek. Inf. dan Komput.*, vol. 2, no. 2, p. 166, 2019, doi: 10.37600/tekinkom.v2i2.115.

[16] I. Ahmad, S. Samsugi, and Y. Irawan, “Implementasi Data Mining Sebagai Pengolahan Data,” *J. Teknoinfo*, vol. 16, no. 1, p. 46, 2022, [Online]. Available: http://portaldata.org/index.php/portaldata/article/view/107

[17] A. Trisnawan, W. Hariyanto, and S. -, “Klasifikasi Beras Menggunakan Metode K-Means Clustering Berbasis Pengolahan Citra Digital,” *RAINSTEK J. Terap. Sains Teknol.*, vol. 1, no. 1, pp. 16–24, 2019, doi: 10.21067/jtst.v1i1.3013.

[18] K. S. H. Kusuma Al Atros, A. R. Padri, O. Nurdiawan, A. Faqih, and S. Anwar, “Model Klasifikasi Analisis Kepuasan Pengguna Perpustakaan Online Menggunakan K-Means dan Decission Tree,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 8, no. 6, p. 323, 2021, doi: 10.30865/jurikom.v8i6.3680.

[19] R. Ishak and A. Bengnga, “Clustering Tingkat Pemahaman Mahasiswa Pada Perkuliahan Probabilitas Statistika Dengan Metode K-Means,” *Jambura J. Electr. Electron. Eng.*, vol. 4, no. 1, pp. 65–69, 2022, doi: 10.37905/jjeee.v4i1.11997.

[20] A. Prasetya, R. Salkiawati, and A. D. Alexander, “Analisis Cluster K-Means dengan Metode Elbow untuk Menentukan Pola Penjualan Produk Traffic Room Summarecon Mal Bekasi,” *J. Students‘ Res. Comput. Sci.*, vol. 4, no. 1, pp. 105–118, 2023, doi: 10.31599/jsrcs.v4i1.2480.