Applied Information System and Management (AISM)
Volume 7, (1) 2024, p. 23-28
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v7i1.33544

The Implementation of K-Means Clustering on E-Learning Feature Development for Outstanding Student Recommendation

Lubsul Jannah[1*], Supriyono[2]

***Abstract*—Schools still need to carry out the process of selecting outstanding students, which has several weaknesses. The data processing process takes a long time and tends to result in human decision-making errors. Although the selection of outstanding students is essential in giving awards and praise to students who excel, the school's current method could be more optimal. The process often takes a long time and requires a lot of human resources to collect and process student data, which can disrupt the school's daily operations. this research aims to group and select students as outstanding students by implementing the k-means clustering method and utilizing E-Learning features. The data used in this study are 30 samples of MIN 2 Malang City student grades, five criteria and grouped into 3 clusters. Experiments conducted are the best criteria weight, the best centroid, the best radius and the best number of clusters to obtain groups (clusters) of students according to the ability and assessment of students. The experimental results show that the best criterion weight is the 4th criterion weight with the percentage of criterion weights: K1 = 25%, K2 = 20%, K3 = 25%, K4 = 15% and K5 = 15%. The best centroid is the 1st test with a Percussion value of 97%, Recall of 98%, Specificity of 98% and Accuracy of 98% obtained in the 1st test. The best radius is obtained in the first and fifth tests with the farthest distance of 10.42. The best number of clusters from the trial results with division into three groups and four obtained is 3 clusters with Precision of 79%, Recall of 78%, Specificity of 89% and Accuracy of 87%. Then the implementation of the k-means method with the system resulted in grouping the highest scores (C1) in as many as 21 students, medium scores (C2) in as many as 5 students and low scores (C3) in as many as 4 students. C1 = 21 students with student data (2, 4, 6, 7, 8, 12, 13, 14, 15, 16, 17, 18, 19, 21, 24, 25, 26, 27, 28, 29, 30), C2 = 5 students with student data (9, 10, 20, 22, 23) and C3 = 4 students with student data (1, 3, 5, 11).**

***Index Terms*—Centroid, cluster, k-means clustering.**

Received: 4 July 2023; Revised: 11 July 2023; Accepted: 16 October 2023
*Corresponding author
[1]L. Jannah, Department of Informatics Engineering Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia (e-mail: 17650030@student.uin-malang.ac.id).
[2]Supriyono, Department of Informatics Engineering Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia (e-mail: priyono@ti.uin-malang.ac.id).

I. Introduction

Every individual student has the potential to develop hard and soft skills that can help them achieve a successful future. However, not all students can fully utilize their potential. This can help them achieve their career and life goals and provide benefits to society and the environment around them [1].

Schools still need to carry out the process of selecting outstanding students, which has several weaknesses. One of them is the data processing process which takes a long time and tends to result in human error in decision-making. Although selecting outstanding students is important in giving awards and praise to outstanding students, the school's current method could be more optimal. The process often takes a long time and requires a lot of human resources to collect and process student data, which can disrupt the school's daily operations.

However, the possibility of human error in decision-making is also a problem that often occurs when selecting outstanding students. This can lead to injustice in giving awards to students who should have achieved them or inappropriate awards to students who did not. One solution to the problem is to use technology that can help process data automatically and accurately, thus speeding up the decision-making process.

Many intelligent algorithms have been developed that can be used to analyze data, including the K-Means algorithm. Implementing the K-Means Clustering method in developing e-learning features for the recommendation of outstanding students can utilize several features contained in the e-learning of MIN 2 Malang City, which also becomes a criterion in calculating K-Means Clustering. Some of these features are 1) Science assessment in the form of daily assignment scores, UTS scores and UAS scores, 2) Homeroom teacher assessment in the form of student social attitude scores, 3) Assessment of extracurricular activities, 4) Achievement History and 5) Percentage of Attendance. This algorithm is often used in data clustering because it is simple to implement, relatively fast, and adaptable to various data types [2]. The number of clusters for this study amounted to 3 clusters, with cluster 1 (C1) for high-scoring students, cluster 2 (C2) for medium-scoring students and cluster 3 (C3) for low-scoring students.

II. Related Work

A. Related Research

One of the biggest problems in Indonesia is malnutrition. In the 2018 Riskesdas data, as many as 17.7% of infants under 60 months still have problems with nutritional intake, while 3.9% are malnourished. As a result, the mortality rate is increasing. This study was conducted to classify the nutritional status of infants under the age of 60 months using a clustering method called K-Means clustering. The conclusions obtained in the clustering of the nutritional state of infants through the division of groups into 4 clusters formed, namely 23 infants with poor nutrition, 17 infants with malnutrition, seven infants with good nutrition and ten infants with excess nutrition. The difference with the research conducted [3] lies in several aspects. This research uses two variables and 4 clusters and calculates accuracy with benchmarks. The goal is to classify the nutritional status of infants under 60 months using the K-Means clustering method. Meanwhile, the current research uses five variables and 3 clusters. It calculates accuracy using a confusion matrix the similarity between the two lies in using the K-Means Clustering method.

Research conducted by [4], South Lampung is a regency with the capital city Kalianda which has an area of 2007.01 km2 and dominates the agricultural area. Based on data on corn crops in the South Lampung Regency Agriculture Office through BPS (Central Bureau of Statistics) shows several areas with varying amounts of corn crops. Therefore, clustering potential corn-producing areas is needed to show which areas obtain large or small amounts of corn. The K-Means clustering method is a data mining method that is non-hierarchical clustering; grouping can use one or more clusters. Based on the results, the region with the highest corn harvest is Penengahan, with a centroid value of 2 is 79448.30257. The region with the lowest corn harvest is Candipuro, with a centroid value of 2 is 1,424,036868. Eight regions (High) with high maize yields are Natar, Tanjung Bintang, Katibung, Sidomulyo, Kalianda, Palas, Penengahan, and Ketapang. In contrast, nine regions (Low) have low maize yields. These areas are Jati Agung, Tanjung Sari, Merbau Mataram, Way Sulan, Candipuro, Way Panji, Rajabasa, Sragi, and Bakauheni. The difference with the research conducted by [4] entitled "Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency" also includes several aspects. This research uses four variables and 2 clusters and does not calculate accuracy. The focus is on grouping areas that have greater or lesser corn income. The current research uses five variables and 3 clusters and calculates accuracy with a confusion matrix. The main similarity is the use of the K-Means Clustering method.

Research conducted by [5]. The transmission of the disease caused by the coronavirus, which is spread under Covid-19, has now been confirmed. The death rate in the Southeast Asia region is increasing and quite alarming. Therefore, the discussion regarding the clustering of COVID-19 Cases and Deaths in the Southeast Asia region in this study chose the K-Means clustering method to process data mining using the K-Means algorithm. With research conducted by [5], there are differences in the number of variables and do not calculate accuracy. This study uses two variables and focuses on clustering COVID-19 cases and deaths in Southeast Asia. The current research uses five variables and calculates accuracy with a confusion matrix. The similarity between the two is using 3 clusters and the K-Means Clustering method.

Research conducted by [2]. The purpose of this research is to introduce TA research areas for students. Based on grades A, B and C in 10 compulsory courses (MKW) received during the six semesters of 2015, 2016, 2017 and 2018 academic years for students of the Informatics Study Program, Faculty of Engineering, Mulawarman University. Clustering analysis was carried out using the K-Means method with cluster accuracy testing using the Sum of Squared Errors (SSE) method obtained a value using 3 clusters of 0.4506% and a value using 4 clusters of 1.1072%. As for cluster evaluation using SC, the three clusters average 0.5852%, meaning that 3 clusters are better at distinguishing MKW, while 4 clusters average 0.4591%, meaning they are relatively weak in placing objects in a cluster. The results show that the K-Means method can be an alternative way to analyze the relationship between MKW and TA in the hope that students, lecturers and academic programs can document their decisions or formulate student regulation policies to verify TA research. The difference with the research conducted by [2] entitled "Implementation of the K-Means Method for Grouping Final Project Recommendations" lies in the number of variables and accuracy calculations. This study uses four variables and calculates accuracy with the sum of squared errors. The goal is to help informatics engineering students choose a final project research area based on ten compulsory courses. The current research uses five variables and calculates accuracy with a confusion matrix. The main similarity is using 3 clusters and the K-Means Clustering method.

Research conducted by [6], Since the emergence of COVID-19 in Indonesia, the government has a new policy in defence efforts. Researchers need to contribute thoughts to determine the level of online learning barriers, which are divided into low and high groups. In this study, researchers analyzed the level of online learning barriers at SMK YASPIM using the K-Means clustering analysis algorithm. The category with low learning barriers is Class 10 TKJ 1, 10 TKJ 2, 10 TKJ 3, 10 RPL, 10 TBSM 1, 10 TBSM 2, 11 RPL, 12 TKJ 1 and 12 TKJ 2 and 6 classes are included in the High-Level Learning Barriers category: 10 RPL, 11 TKJ 1, 11 TKJ 2, 11 TKJ 3, 11 TBSM and 12 RPL. The difference with the research [6] includes the number of variables, clusters, and accuracy calculations. This research uses three variables and 2 clusters and does not calculate accuracy. This research aims to measure barriers to online learning during the COVID-19 pandemic. The current research uses five variables and calculates accuracy with a confusion matrix. The similarity lies in the clustering method, namely K-Means Clustering.

B. *Clustering*

Clustering refers to grouping such as records, reviewing and forming classes of objects with something in common. A cluster is a set of records that are similar and different from records from other clusters. Clustering tries to divide all data sets. Clustering attempts to divide all data sets into relatively similar groups, where the similarity among records within one group is maximized. Meanwhile, the similarity with records in other groups is minimized [7].

There are several methods involved in implementing clustering techniques. In clustering, two main approaches are commonly used: the partition approach and the hierarchical approach. Partition-based clustering is also known as the partitioning method. Clustering is the process of grouping data by sorting analyzed data into available clusters. On the other hand, the hierarchical approach is a data clustering technique that builds a hierarchy based on a dendrogram. A dendrogram is a diagram or graph resembling a tree-like structure that summarizes the hierarchical clustering process, showing evolutionary changes. In this technique, data with high similarity are placed nearby within the hierarchy, while data with low similarity are placed farther away [8].

C. K-Means

The k-means data mining method stands as a pivotal technique within the realm of data clustering. In the intricate domain of data clustering, intricate datasets find cohesion through a meticulous process: they are organized into cohesive clusters, each bearing its unique identity shaped by shared characteristics. This methodological approach hinges on the fundamental principle of segregating data points into one or more clusters based on their inherent traits. When similarities echo among specific data points, they naturally converge within the confines of a singular cluster, forging a collective identity. Conversely, data points embodying distinct characteristics find their belongingness in disparate clusters, creating a systematic and nuanced classification system. Through this nuanced stratagem, the k-Means methodology profoundly shapes the landscape of data clustering, illuminating the underlying patterns within datasets with precision and clarity [9].

The K-means Clustering algorithm stands at the forefront of data analysis, representing a sophisticated data mining methodology deeply rooted in unsupervised learning. This algorithm is a linchpin within the expansive domain of data analysis, exemplifying the fusion of intricate datasets through meticulously crafted partitioning systems. Its prominence lies in its ability to discern patterns and unveil hidden insights within datasets without explicit guidance. Through the intricate dance of unsupervised learning, the K-means Clustering algorithm artfully groups data points into clusters, illuminating the underlying structures within the data tapestry. This technique represents not just a computational tool but a gateway to unraveling the complexities of datasets, offering researchers and analysts a powerful lens to dissect and comprehend the intricate relationships interwoven in the vast expanse of data [10].

The K-means method embarks on a sophisticated endeavor, aiming to amalgamate diverse datasets into distinct categories, each marked by a cohesive set of shared traits. In this intricate process, data points within a specific category echo one another in their inherent characteristics, forging a unified identity within the cluster. Simultaneously, these shared attributes set them apart from their counterparts inhabiting other categories. Through this meticulous partitioning, the K-means method unravels the intricate tapestry of data and magnifies the subtle nuances that delineate one cluster. It is a profound exploration into the essence of data, illuminating the unique qualities and commonalities that define each category, thereby enabling a profound understanding of the underlying patterns that permeate the dataset [11].

The k-means clustering method has several advantages compared to other clustering methods. This is revealed in research [12]: Simple and Fast: K-means is one of the simplest and fastest clustering methods. This makes it suitable for use when you have extensive data and want to divide it into groups quickly. Good Scalability: K-means usually performs well when the number of data points (observations) is relatively large. It can handle data sets with thousands or millions of data points well. Straightforward Interpretation: The results of k-means clustering can be easily interpreted. The centroid of its group identifies each group, and data points are labelled according to their group.

III. Research Method

The clustering process through the K-Means method encompasses a series of intricate steps, each meticulously designed to unveil the underlying patterns within the dataset. Determining the appropriate number of clusters, denoted by the variable 'k,' initially sets the stage. Typically, this number is chosen based on careful consideration and domain knowledge. Subsequently, the pivotal task of establishing the cluster center, often called the centroid, takes center stage. In the inaugural iteration, the centroid is arbitrarily chosen, either from the existing dataset or from the data set to be processed. However, as the process evolves, subsequent iterations involve a refined approach. The average data points within each cluster are meticulously calculated, shaping the evolving centroids.

Central to this method is utilizing the Euclidean Distance formula, a cornerstone in clustering algorithms. This formula is the bedrock for calculating the spatial distance between every input data point and each centroid. This meticulous computation aims to pinpoint the closest proximity between each data point and its corresponding centroid, illuminating the data's intricate relationships within the clusters. Through these intricate calculations, the K-Means method not only refines the clustering process but also offers a profound lens through which the intricate interconnections and spatial dimensions of the dataset are comprehensively explored and understood [13].

$$(i,j) = \sqrt{\sum_{i=1}^{n_k} (X_{ki} - C_{kj})^2} \tag{1}$$

In the formula, $(i,)$ is the distance of the first data to the J centroid. X_ki is the first data in the cluster on the k data attribute, while C_kj is the j centroid on the k data attribute. n_k is the number of data in the K data attribute. With this formula, the closest distance of each data to the centroid can be found to determine the right cluster for the data.

Following the meticulous calculation of distances, the dataset embarks on a transformative journey, where each data point finds its rightful place within a specific cluster, determined by its closest proximity to the cluster's center. This pivotal step hinges on the crucial measurement of the shortest distance between individual data points and the designated cluster center, effectively delineating the cluster members. With the cluster compositions established, the centroid, serving as the heartbeat of each cluster, undergoes a metamorphosis. This transformation is orchestrated through meticulous calculations, where the centroid's value is recalibrated, drawing upon the collective essence of all data points residing within the cluster's newfound membership. This recalibration is executed through a precise formula, a mathematical dance that harmoniously fuses the data points' attributes, resulting in an updated centroid value that embodies the cluster's evolving identity [14]:

$$C_{ki} = \frac{1}{n_i} \sum_{q=1}^{n_{ki}} X_{kq} \quad (2)$$

C_{ki} is the centre point of a cluster at the k attribute, and i cluster. X_{kq} refers to the q data in the i cluster at the k attribute, while n_{ki} is the number of data in the i cluster at the k attribute.

According to the Euclidean Distance formula, the K-Means method goes through a series of steps that are all very complicated. The first step is to update the centers and draw lines between clusters based on the proximity criteria. This intricate dance continues unabated until stability is achieved—a point wherein the data members within each cluster cease their fluctuations or alterations. At this juncture, a critical transformation occurs: the clusters solidify into distinct, coherent entities, representing the culmination of the iterative journey. If, during the iterative process, no discernible changes reverberate through the dataset, the algorithm gracefully halts, signifying the formation of the clusters. The picture shown in Fig. 1 shows this complicated ballet of data analysis, which is like a carefully planned performance. It gives a deep and clear picture of the many steps in this changing process.

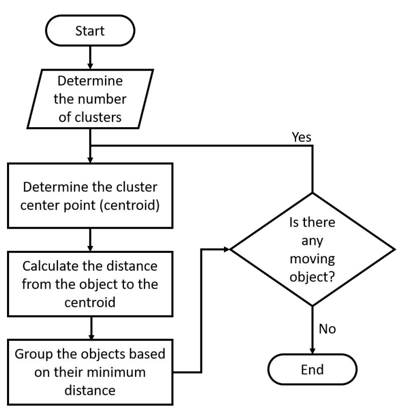

Fig. 1. Flowchat of K-Means Method [15]

IV. RESULT

This research utilized data from 30 samples of 6th-grade students from MIN 2 Malang City, evaluated based on five criteria. These data were then grouped into three clusters: C1 representing high scores, C2 representing medium scores, and C3 representing low scores. Researchers conducted a series of experiments to match students' abilities and grades with these groupings. These experiments involved searching for the best weight values, determining the optimal centroids, identifying the best radii, and finding the best number of clusters. All the criteria to be used in this study are detailed in Table 1:

Table 1.Criteria

No	Criteria code	Name criteria	value
1	K1	science	Task 1, task 2, task 3, task 4, UTS, UAS
2	K2	Social attitude	Honest, disciplined, responsible, polite
3	K3	Extracurricular	activities
4	K4	Achievement History	Highest achievement
5	K5	attendence	Present, permission. alpha

During this process, researchers explored various combinations of weight values to find the best ones. They focused on determining the optimal centroids and the best distances between data points and centroids. Additionally, the study aimed to find the optimal number of clusters to best represent the variation in the data.

By conducting these experiments, the researchers aimed to create an accurate system of grouping that aligns with students' abilities and achievements. The final outcome was dividing students into three groups, carefully considering their abilities and grades: C1 for high-achieving students, C2 for students with intermediate achievements, and C3 for students with low achievements. Therefore, this research involved standard statistical analyses, in-depth experimental steps, and careful criteria selection to ensure the relevance and accuracy of the grouping results. All these processes form a solid foundation for the methodology of this research.

In the process of selecting the most suitable criterion weights, the researchers had the opportunity to explore five different options. Through rigorous experimentation, the researchers aimed to identify the combination of criterion weights that yielded the highest accuracy values. These experiments served as a crucial step in determining the optimal criteria for the research. Ultimately, the researchers chose the set of criterion weights that demonstrated superior accuracy, ensuring the robustness and reliability of the chosen criteria for the study.

Based on the outcomes of the conducted experiments, it was determined that the most optimal criterion weight configuration was the fourth set, where the criteria were distributed as follows: K1=25%, K2=20%, K3=25%, K4=15%, and K5=15%. Furthermore, the best-performing centroid was identified in the initial test, displaying exceptional metrics such as a Precision value of 97%, a Recall rate of 98%, a Specificity rate of 98%, and an Accuracy rate of 98%. Notably, this first test utilized randomized data, wherein C1 represented the 16th student, C2

Applied Information System and Management (AISM)
Volume 7, (1) 2024, p. 23-28
P-ISSN: 2621-2536; E-ISSN: 2621-2544; DOI: 10.15408/aism.v7i1.33544

represented the 29th student, and C3 represented the 27th student. These findings underscore the meticulous evaluation process to determine the most accurate and effective configurations for the study's criteria and centroids.

The optimal radius was identified through meticulous testing, with the most favorable results emerging from the initial and fifth tests. In these tests, the calculated distance, set at 10.42 units, represented the farthest reach of the radius. This critical parameter signifies the extent to which data points were considered within the clusters, underscoring the precision of the clustering methodology. For a comprehensive understanding, the visualization of the first test configuration can be observed in Figure 2, while the fifth test configuration is depicted in Figure 3. These visual representations provide a detailed insight into the spatial distribution of data points within the defined clusters, aiding in a comprehensive analysis of the clustering outcomes.

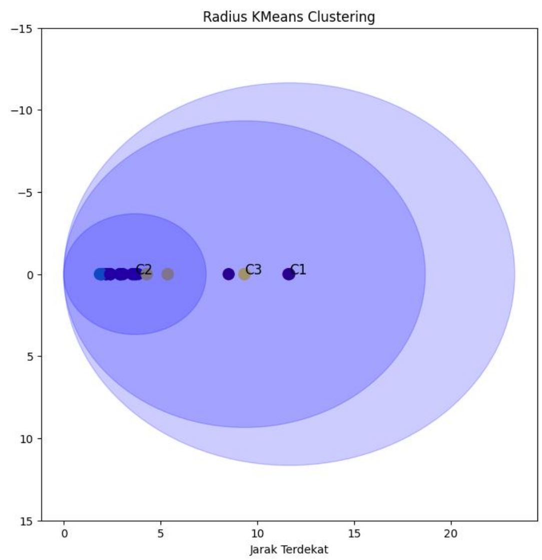

Fig. 2. Radius of K-Means clustering 1st test

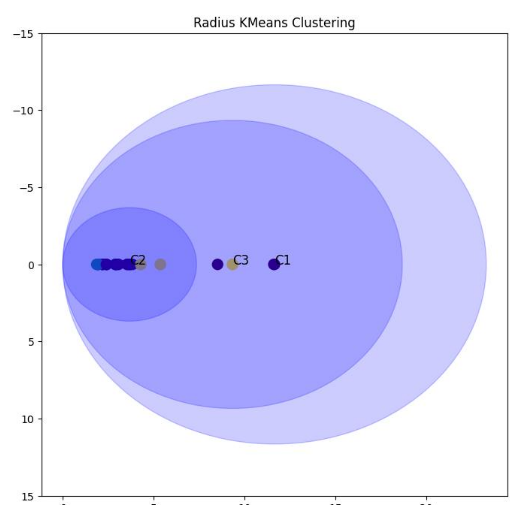

Fig. 3. Radius of K-Means clustering 5th testing

Figures 2 and 3 describe the outcomes derived from a meticulous analysis involving ten extensive trials. Within this experiment, Fig. 2 represents the inaugural trial, while Figure 3 encapsulates the results of the fifth trial. Remarkably, both configurations yielded an identical radius measurement of 10.42 units. These visual representations serve as a profound exploration into the calculated radius of each cluster. Significantly, the radius signifies the extent to which elements within a cluster are encompassed, emphasizing the meticulousness of the clustering methodology applied.

The rigorous experimentation involving varied cluster divisions, specifically into three and four groups, culminated in a revelation. The analysis indicated that the optimal number of clusters was three. This configuration showcased an impressive performance, boasting a precision rate of 79%, a recall rate of 78%, a specificity rate of 89%, and an accuracy rate of 87%. Consequently, the meticulous implementation of the k-means method within this system resulted in the formation of distinct groups. Notably, the highest-scoring group denoted as C1, comprised an impressive 21 students. Additionally, a medium-scoring group (C2) comprising five students and a low-scoring group (C3) encompassing four students. To delve into specifics, C1 included students with the following data points: (2, 4, 6, 7, 8, 12, 13, 14, 15, 16, 17, 18, 19, 21, 24, 25, 26, 27, 28, 29, 30), C2 encompassed students (9, 10, 20, 22, 23), and C3 comprised students (1, 3, 5, 11).

For an in-depth exploration of the K-Means distribution, mainly focusing on the exceptional students' grouping, Figure 3 serves as a comprehensive visual reference. This figure encapsulates the intricacies of the clustering outcomes, offering profound insights into the strategic arrangement of outstanding students within the defined clusters.

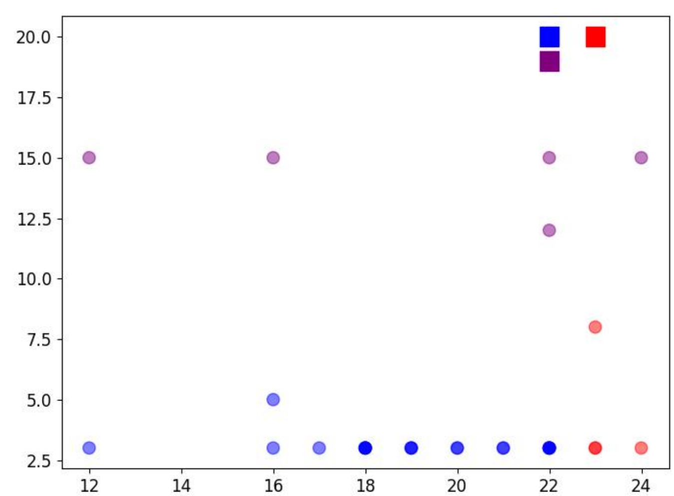

Fig. 4. Visualization of cluster results

Figure 4 is a comprehensive visual representation that delineates the intricate distribution of student data within each cluster, derived from the conclusive outcomes of employing the advanced K-Means clustering method. Within this visualization, Cluster C1, encompassing 21 students, is vividly portrayed in a striking shade of blue, signifying their high achievement levels. Similarly, Cluster C2, housing a cohort of 5 students, is elegantly depicted in regal purple, representing their intermediate scores. Contrastingly, Cluster C3, accommodating a smaller group of 4 students, is vividly highlighted in a vibrant red hue, denoting their comparatively lower academic achievements.

Notably, this visualization process is meticulously crafted through manual coding, utilizing the sophisticated Python programming language on the innovative Jupiter notebook platform. Through this intricate methodology, the intricate patterns and distinctions within each cluster come to life, providing an illuminating visual narrative that aids in a profound understanding of the academic stratification among students, meticulously captured through the lens of the K-Means clustering technique.

V. Conclusion

In this research, my main contribution lies in applying the k-means clustering method to group 6th-grade students from MIN 2 Malang City based on the earlier criteria, such as science scores, social attitudes, extracurricular activity scores, achievement history, and attendance percentage. I also measured the accuracy of this method using a confusion matrix and obtained significant results in terms of precision, recall, specificity, and accuracy, which were 79%, 78%, 89%, and 87%, respectively. These results demonstrate the effectiveness of using the k-means method in student clustering.

However, for future research, several suggestions can be considered to enhance the quality of the study:

- Further Research on Clustering Criteria: Conduct in-depth research to determine more relevant clustering criteria that significantly impact the final outcomes. For instance, consider psychological or environmental factors that might influence student performance.
- Exploration of Other Clustering Algorithms: Apart from k-means, explore other clustering algorithms such as hierarchical clustering, DBSCAN, or more complex machine learning algorithms to compare their results and select the best method based on specific research needs.
- Analysis of Dynamic Data: Expand the research towards dynamic data analysis, where student clustering is performed in real-time based on changing data over time. This requires the development of adaptive strategies and continuous monitoring of data changes.
- Utilization of Additional Validation Methods: Use additional validation methods like cross-validation or bootstrapping to ensure the stability and reliability of the obtained clustering results.
- Consideration of New Technologies: Consider using new technologies such as machine learning, big data analytics, or artificial intelligence to enhance the accuracy and efficiency of the clustering process.

Taking these considerations into account, it is expected that future research will provide a more profound and more significant contribution to the field of student clustering based on relevant criteria.

References

[1] E. R. Siagian, “Implementasi Metode Profile Matching untuk Penentuan Mahasiswa Berprestasi,” *MEANS (Media Inf. Anal. dan Sist.)*, vol. 5, no. 1, pp. 61–66, 2020, doi: 10.54367/means.v5i1.752.

[2] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, “Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir,” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 16, no. 1, pp. 13-18, 2021, doi: 10.30872/jim.v16i1.5182.

[3] S. S. Nagari and L. Inayati, “Implementation of Clustering Using K-Means Method to Determine Nutritional Status,” *J. Biometrika dan Kependud.*, vol. 9, no. 1, pp. 62-68, 2020, doi: 10.20473/jbk.v9i1.2020.62-68.

[4] A. A. Aldino, D. Darwis, A. T. Prastowo, and C. Sujana, “Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung Regency,” *J. Phys. Conf. Ser.*, vol. 1751, no. 1, pp. 1-9, 2021, doi: 10.1088/1742-6596/1751/1/012038.

[5] J. Hutagalung, N. L. W. S. R. Ginantra, G. W. Bhawika, W. G. S. Parwita, A. Wanto, and P. D. Panjaitan, “COVID-19 Cases and Deaths in Southeast Asia Clustering using K-Means Algorithm,” *J. Phys. Conf. Ser.*, vol. 1783, no. 1, pp. 1-6, 2021, doi: 10.1088/1742-6596/1783/1/012027.

[6] A. Rohmah, F. Sembiring, and A. Erfina, “Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring (Studi Kasus: Smk Yaspim Gegerbitung,” *Sistem Informasi dan Manajemen Informatika*, vol. 1, no. 1, pp. 290–298, 2021.

[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, pp. 100-108, 2021.

[8] B. Karlina and A. M. Sanoyo, “Pengaruh Cluster Emiten terhadap Return Saham JSX Berbasis Parameter Rasio Analisa Fundamental,” *J. Akuntansi, Keuangan, dan Manaj.*, vol. 2, no. 4, pp. 279–291, 2021, doi: 10.35912/jakman.v2i4.294.

[9] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, “Analisis K-Means Clustering pada Data Sepeda Motor,” *INFORMAL Informatics J.*, vol. 5, no. 1, pp. 10-17, 2020, doi: 10.19184/isj.v5i1.17071.

[10] S. P. Tamba, F. T. Kesuma, and Feryanto, “Penerapan Data Mining Untuk Menentukan Penjualan Sparepart Toyota Dengan Metode K-Means Clustering,” *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 67–72, 2020, doi: 10.34012/jusikom.v2i2.376.

[11] F. Amin, D. S. Anggraeni, and Q. Aini, “Penerapan Metode K-Means dalam Penjualan Produk Souq.Com,” *Appl. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 7–14, 2022, doi: 10.15408/aism.v5i1.22534.

[12] N. Sirait, “Implementasi K-Means Clustering Pada Pengelompokan Mutu Biji Sawit,” *J. Pelita Inform.*, vol. 16, no. 4, pp. 368–372, 2017.

[13] D. Xia, F. Ning, and W. He, “Research on Parallel Adaptive Canopy-K-Means Clustering Algorithm for Big Data Mining Based on Cloud Platform,” *J. Grid Comput.*, vol. 18, no. 2, pp. 263–273, 2020, doi: 10.1007/s10723-019-09504-z.

[14] N. Nur Afidah, “Penerapan Metode Clustering dengan Algoritma K-means untuk Pengelompokkan Data Migrasi Penduduk Tiap Kecamatan di Kabupaten Rembang,” *Prism. Pros. Semin. Nas. Mat.*, vol. 6, pp. 729–738, 2023, [Online]. Available: https://journal.unnes.ac.id/sju/index.php/prisma/

[15] W. Afifi, D. R. Nastiti, and Q. Aini, “Clustering K-Means Pada Data Ekspor (Studi Kasus: Pt. Gaikindo),” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 1, pp. 45–50, 2020, doi: 10.24176/simet.v11i1.3568.