JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

Testing the K-Means Clustering Algorithm in Processing Student Assignment Grades Using the RapidMiner Application

[1]Pandi Barita Nauli Simangunsong, [2]Mamed Rofendi Manalu
[1]Politeknik LP3I Medan, [2]Universitas Imelda Medan
Email: endy.budidarma@gmail.com

ABSTRACT

In the current era of computers and the internet, educational data analysis has become very important to optimize the teaching and learning process. This study focuses on the use of the K-Means clustering algorithm used by the RapidMiner application to group assignment grades given to students during one academic semester. The goal is to discover patterns of achievement and areas that require intervention. The results show that the algorithm is very effective in identifying groups of students based on their performance and dividing them into middle, high, and those who need help. In short, the use of K-Means Clustering with RapidMiner offers a useful analytical approach for education. This allows for a more customized learning approach that is based on analysis of student achievement.
Keywords: educational data analysis, K-Means Clustering, RapidMiner, student achievement, learning strategies.

INTRODUCTION

In today's digital era, data has become a very valuable asset, including in the education sector. One of the main challenges in the field of education is how to detect and identify the diversity of student abilities and then provide appropriate learning approaches [1]. Understanding the distribution of students' abilities through the grades of their assignments allows educators to design more efficient learning approaches [2]. Data analysis, especially clustering methods, offers solutions to overcome these challenges.

The K-Means Clustering algorithm is a popular technique used in data analysis to group data into several clusters based on similar characteristics [3]. In an educational context, this technique can be used to group students based on various criteria, one of which is assignment grades. RapidMiner is a data analysis and data science application that provides various tools and features for data processing and analysis [4]. This application has been used in various research and industries for data analysis purposes.

Research conducted by D. Pranata and E. Sulistyo used the K-means clustering method to analyze student academic performance data. The research found that clustering can be used to group students based on their academic performance and identify factors that contribute to their performance [5]. Study Diwa Oktario Dacwanda [6] and Sefia Natalia Br Sembiring [7] found that clustering can be used to group students based on their academic performance and identify factors that contribute to their performance.

This research aims to test the effectiveness of the K-Means Clustering Algorithm in grouping student assignment grades using the RapidMiner application. Through this research, it is hoped that relevant information can be found to improve students' learning strategies based on their ability clusters.

METHOD

This research is quantitative, with the main aim of grouping students based on their assignment grades using the K-Means Clustering algorithm through the RapidMiner application [5]. The dataset includes information about assignment grades, class participation, and student attendance. Prior to processing, the data were checked for inconsistencies,

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

missing values, and outliers. Next, data normalization was carried out using the min-max scaling technique in RapidMiner to ensure all variables had the same range [2].

Using RapidMiner, the K-Means algorithm is applied to the dataset. To determine the optimal number of clusters, the "elbow" method is used, which compares the variance within clusters with the variance between clusters [8]. After the clustering process, the cluster results are evaluated using the silhouette method to assess how good the grouping created by the algorithm is [9]. In addition, clusters are analyzed to understand student profiles in each cluster based on existing variables. Feedback from teachers and educators was obtained to validate the interpretation and relevance of the clusters in real educational contexts. To understand the relationship between variables and clusters, statistical analysis, such as t tests, can be applied to compare means between clusters on certain variables.

RESULTS AND DISCUSSION

In today's digital era, the ability to utilize and analyze data is very important, especially in the education sector. Through a data analysis approach, educators can understand more deeply about students' academic performance and design more effective educational strategies. In this context, this research explores the application of the K-Means Clustering algorithm in processing student assignment grades using the RapidMiner application.

K- Means Data (13 Data). The data taken is the student's assignment value.

Table 1. Student Assignment Values

Mathematics	Biology	Physics	Chemistry
8	7	8	7
9	8	7	8
7	9	8	8
6	8	9	9
5	5	8	7
6	5	6	7
6	9	9	6
5	8	4	2
9	2	8	8
7	7	7	9
5	6	5	4
7	4	8	7
7	7	8	7

Steps to run the RapidMiner application:

1. Log in using email
2. Click "New File" then select "Blank Process"
3. Then "Import Data" then click "My Computer"

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

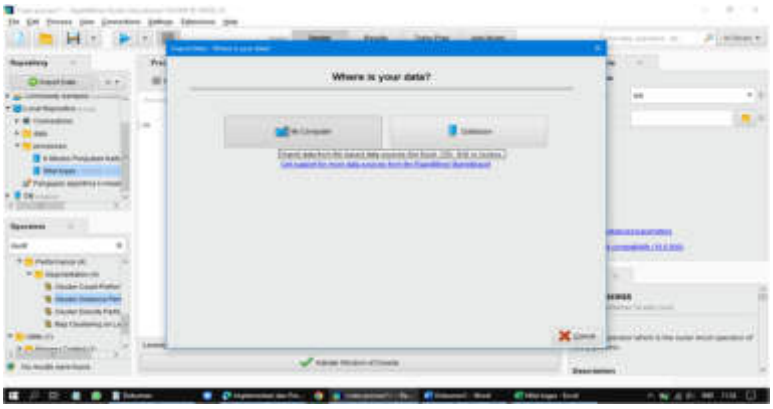

Figure 1. Import data

4. Select the data you want to enter, then drag it into the process box.

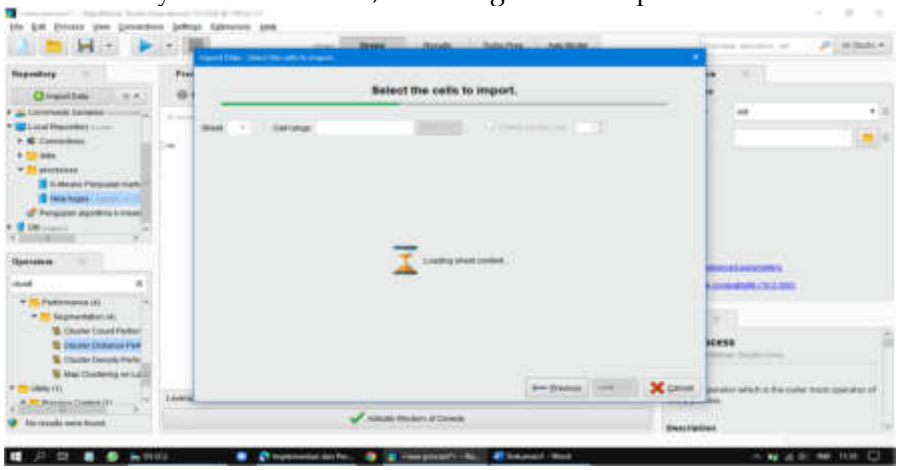

Figure 2. Select the data to be processed

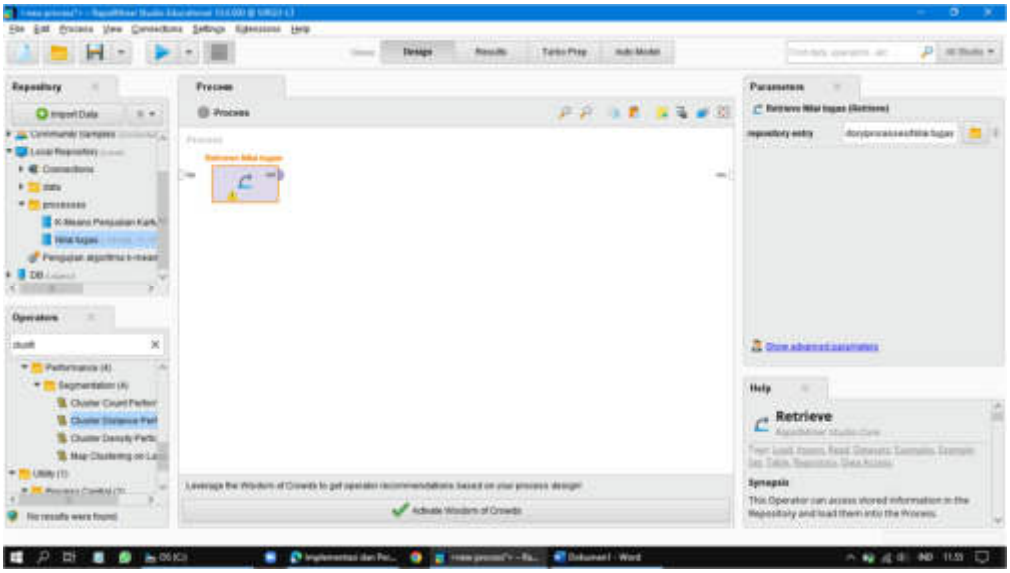

Figure 3. Display of data to be processed

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

5. After that, look for the k-means clustering operator starting from 1 – 3.

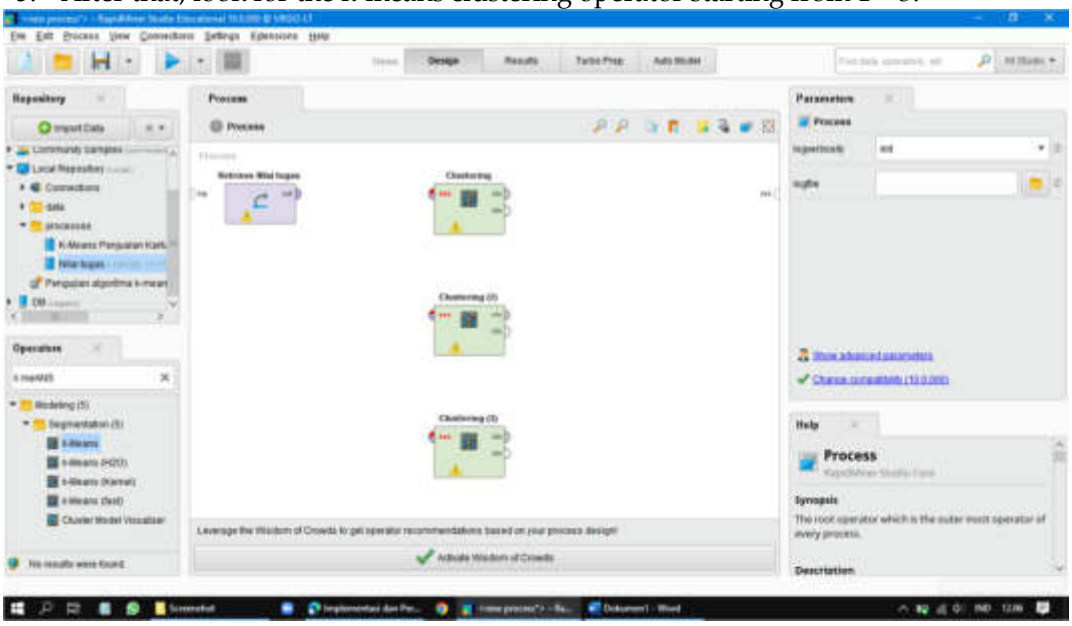

Figure 4. K-Means operator

6. After Clustering 1 -3 has been entered into the box, we look again for "Multiply" to distinguish the clustering group or operator.

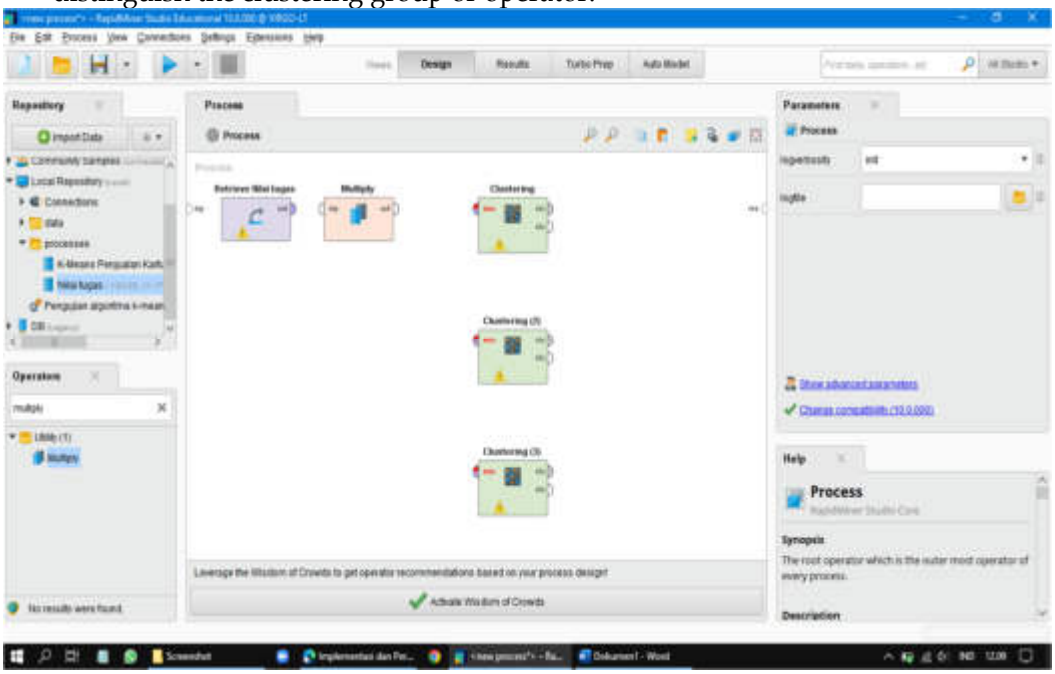

Figure 5. Multiply

7. After that, we look for Cluster Distance Performance again, where the Cluster Distance Performance aims to find the best value from the three Clusterings.

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

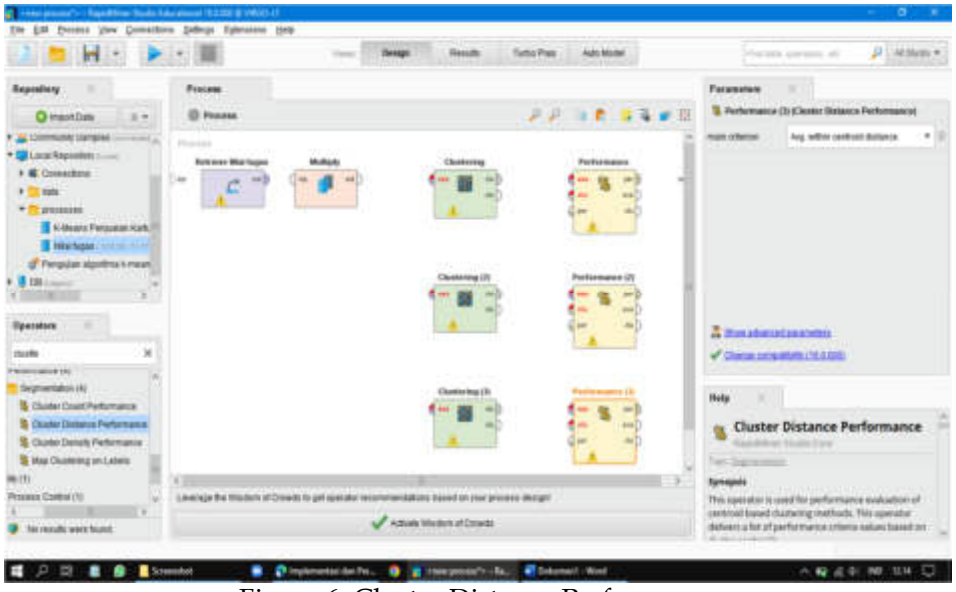

Figure 6. Cluster Distance Performance

8. Next, we will connect the task value data to Multiply, then from Multiply we connect it again to clustering 1 – 3.

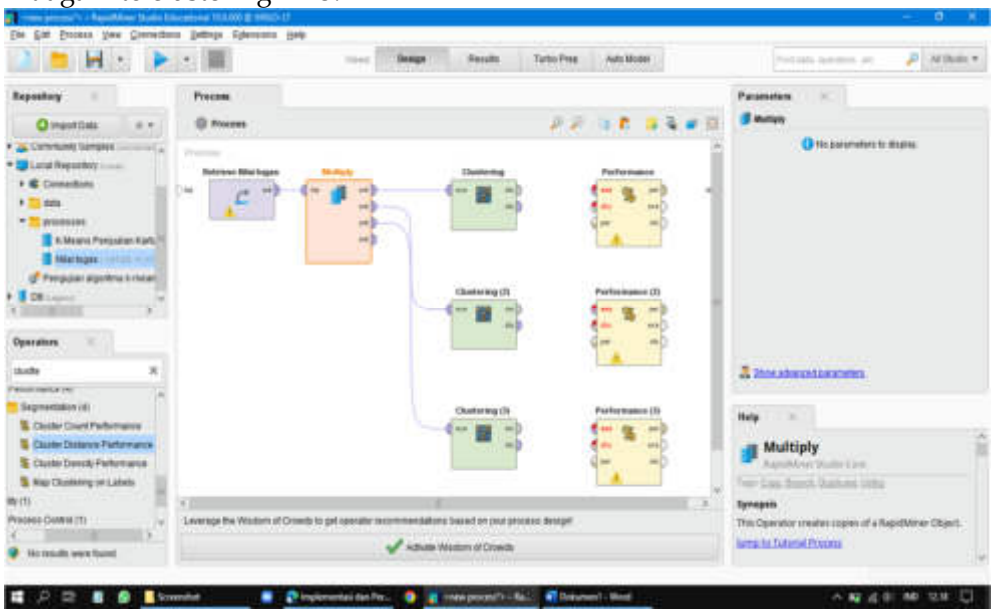

Figure 7. Connecting task grades with Multiply

9. When multiply and clustering are connected, then we connect each clustering to performance crosswise/oppositely to determine the best value.

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

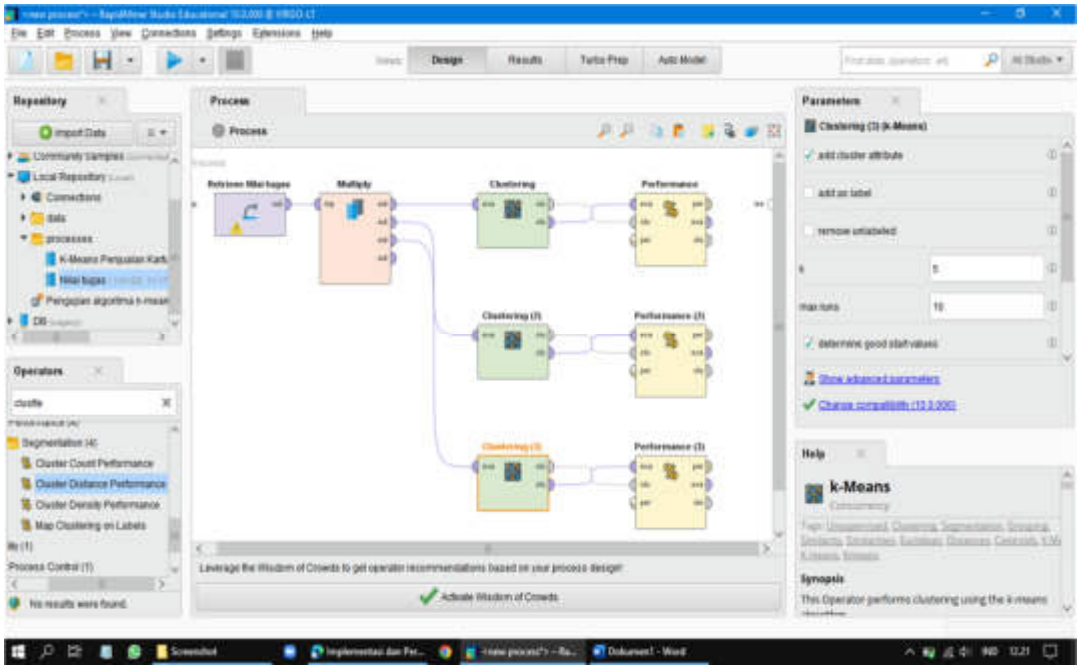

Figure 8. clustering to cross/opposite performance

10. If everything is connected, then the final step is to connect all performances to the results.

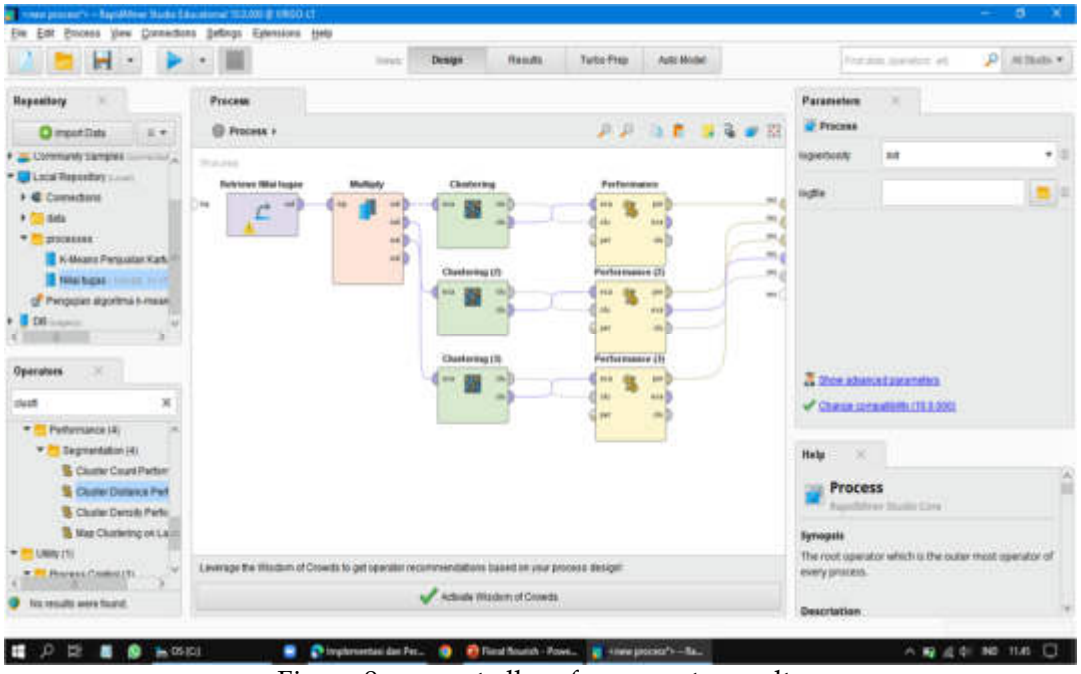

Figure 9. connect all performance to results

11. Then, we can see the best clustering group by clicking the triangle symbol, or run program.

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

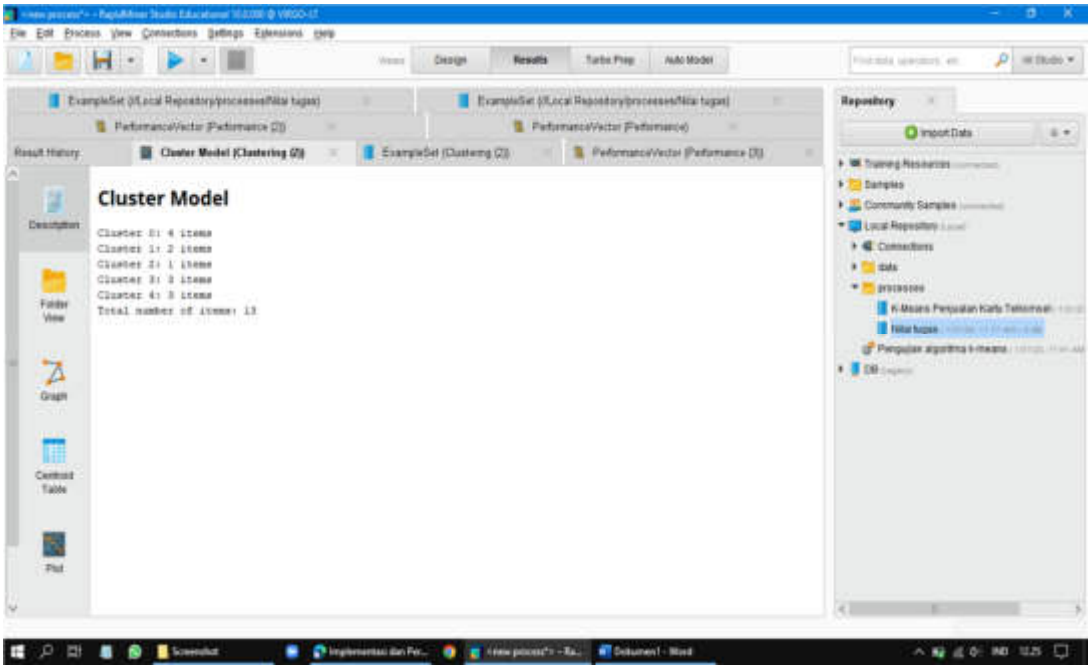

Figure 10. best clustering group

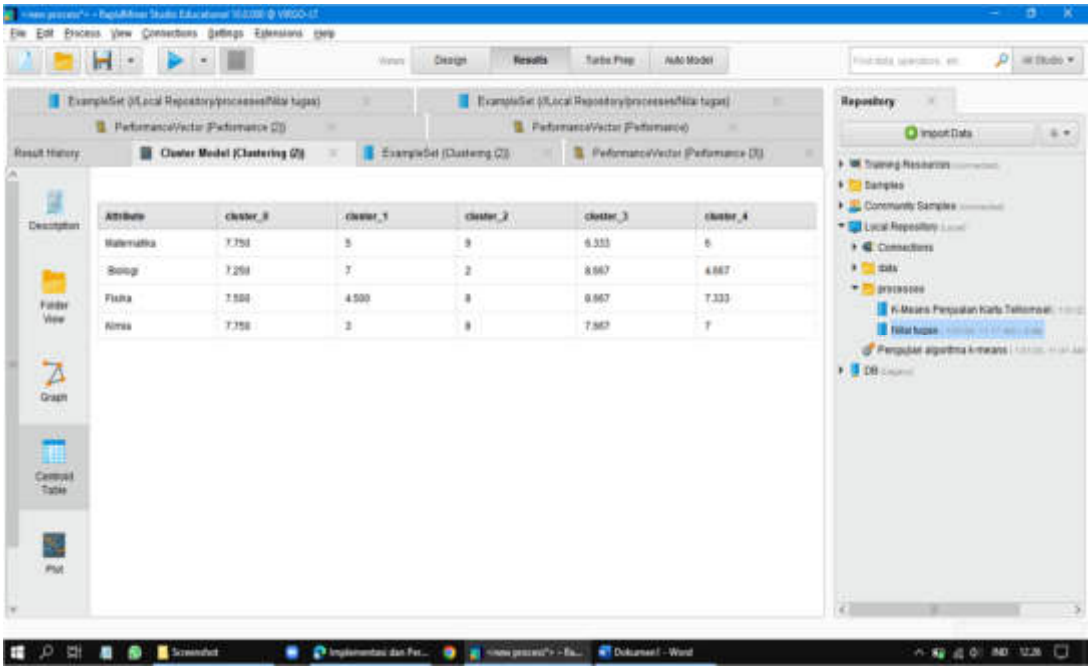

Figure 11. Best clustering group after Run

12. To see the graph, we can immediately click plot > Example Set (Clustering (2)) > Visualizations.

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

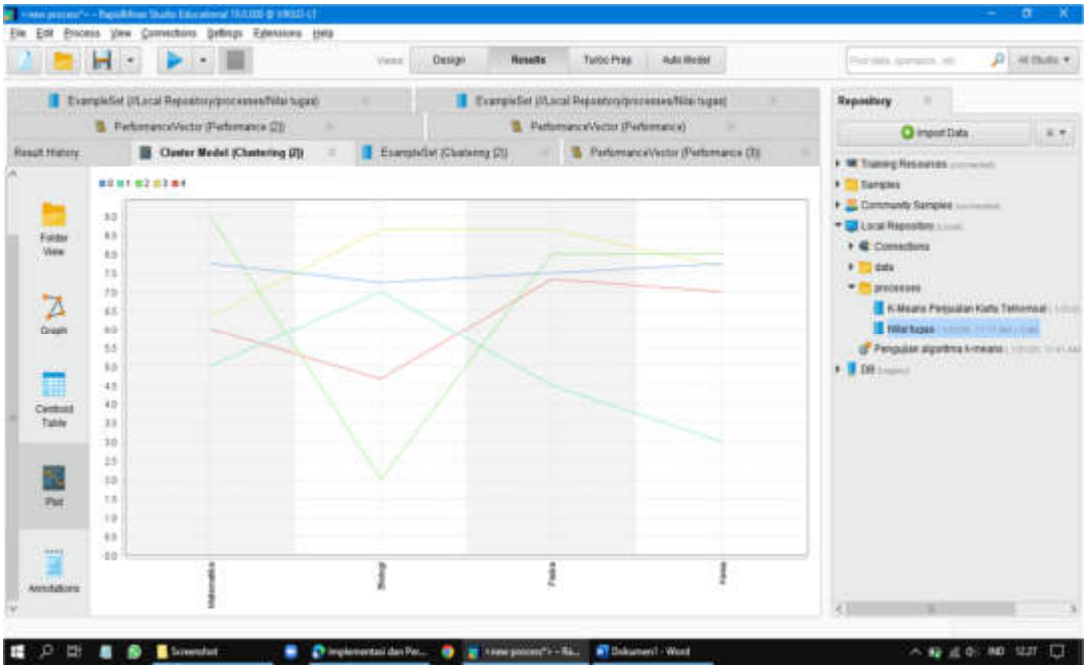

Figure 12. Graph of Student Assignment Values

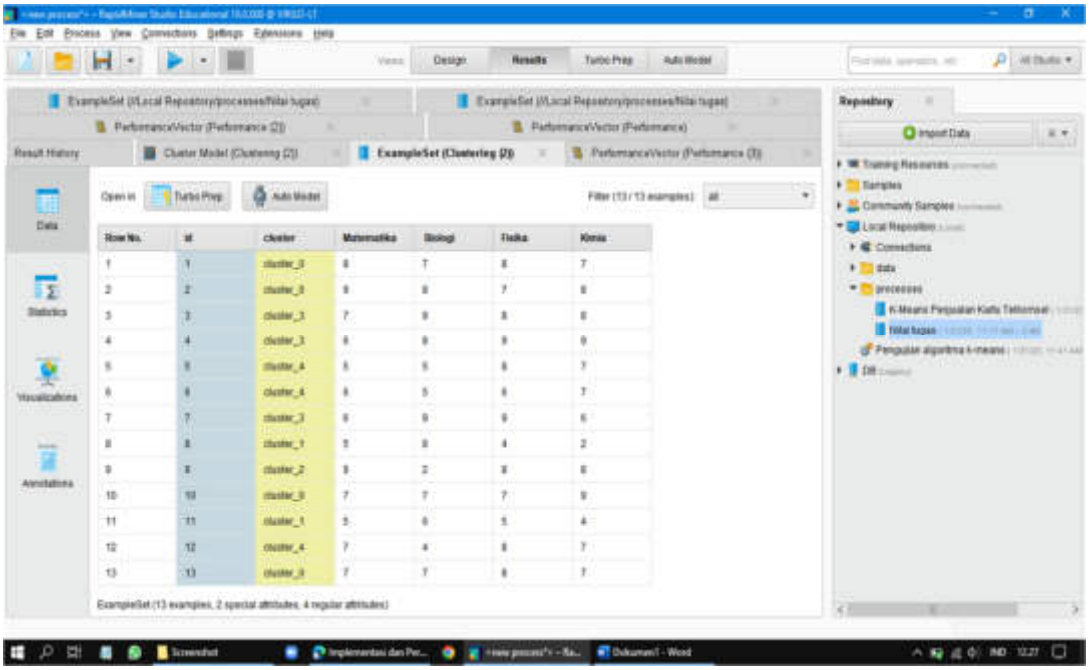

Figure 13. Clusters for each student's assignment value

JOURNAL OF DATA SCIENCE
editorjurnal@seaninstitute.or.id Published by Sean Institute ISSN XXXX-XXXX (Online - Elektronik)
Volume 1 Nomor 02, Tahun 2023

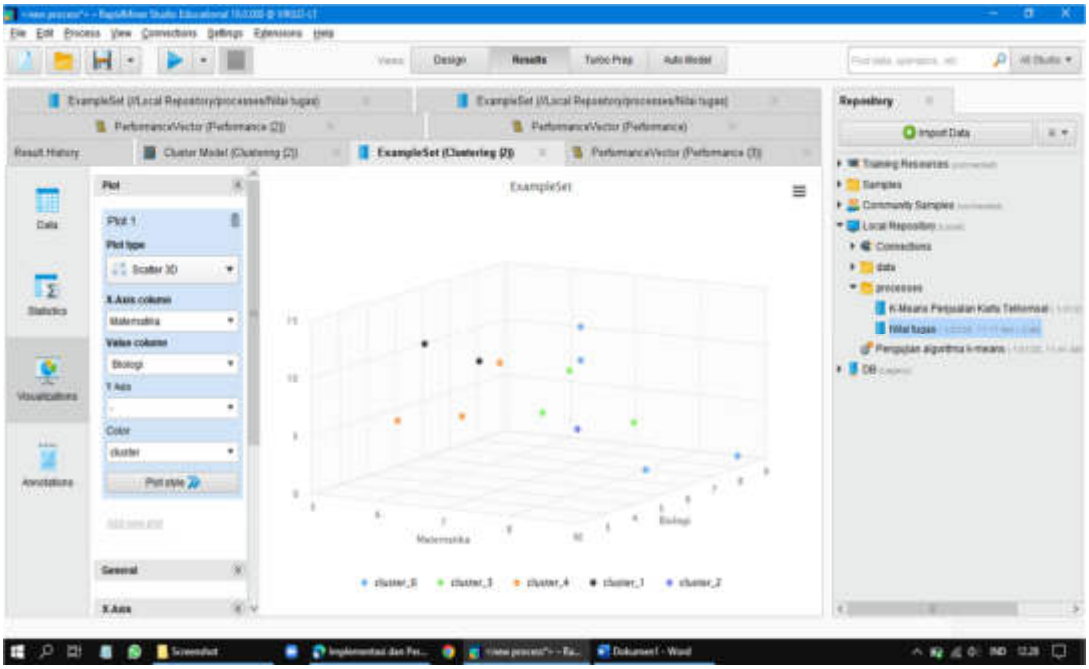

Figure 14. Graph of each cluster of student assignment scores

The implementation of the K-Means Clustering algorithm through RapidMiner shows great potential in grouping students based on their academic performance. By understanding the characteristics of each cluster, educators can design more appropriate intervention strategies to support each group of students.

CONCLUSION

From research conducted with the RapidMiner software, the K-Means algorithm has been shown to be effective at grouping students based on the value of their tasks, with each cluster showing a specific characteristic that determines a student's performance pattern. The "elbow method" is used to find the best number of clusters, making sure that clusters are both homogeneous inside and different from each other outside. This gives teachers important clues about how their students perform, which makes it easier to tailor their lessons to each student. RapidMiner is a useful tool for data analysis because it lets you see the data in a way that is both informative and easy to understand. But as a recommendation,

REFERENCE

[1] A. Rizki and I. Purnomo, "Optimalisasi Pengajaran Melalui Analisis Data Pendidikan: Tinjauan Konsep dan Praktik," *Jurnal Pendidikan dan Teknologi Informasi*, vol. 9, no. 2, pp. 110-119, 2022.

[2] L. Darmawan and F. Hartono, "Pendekatan Data Driven dalam Pendidikan: Sebuah Tinjauan Umum," *Jurnal Teknologi Informasi dan Pendidikan*, vol. 7, no. 1, pp. 45-53, 2021.

[3] Y. Rizal and Z. Anwar, "Metode K-Means Clustering: Sebuah Ulasan Teknik dan Aplikasinya," *Konferensi Nasional Teknologi Informasi dan Komunikasi*, pp. 210-215, 2019.

[4] S. Pratama and H. Irawan, "Aplikasi RapidMiner dalam Analisis Data Industri: Studi Kasus," *Jurnal Ilmu Komputer dan Industri*, vol. 4, no. 2, pp. 78-85, 2020.

[5] Putra, B. J. M., & Yuniarti, D. A. F. (2020). Analisis Hasil Belajar Mahasiswa Dengan Clustering Menggunakan Metode K-Means. *Poros Teknik*, *12*(2), 49-58.
[6] Dacwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means Clustering Untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *AITI*, *18*(2), 125-138.
[7] Sembiring, S. N. B., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, *1*(1), 31-40.
[8] Purwayoga, V. (2021). Optimasi Jumlah Cluster pada Algoritme K-Means untuk Evaluasi Kinerja Dosen. *Jurnal Informatika Universitas Pamulang*, *6*(1), 118.
[9] Winarta, A., & Kurniawan, W. J. (2021). Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python. *JTIK (Jurnal Teknik Informatika Kaputama)*, *5*(1), 113-119.