DIJEMSS: Dinasti International Journal of Education Management and Social Science
E-ISSN: 2686-6331
P-ISSN: 2686-6358
https://dinastipub.org/DIJEMSS dinasti.info@gmail.com +62 811 7404 455
DOI: https://doi.org/10.38035/dijemss.v6i6

Clusterization of New Student Promotion Strategy Determination at Prisma University Manado Using K-Means Algorithm and Agglomerative Hierarchical Clustering

Ngurah Daniel Palar[1]
[1]Information Systems Study Program, Faculty of Information Technology, Budi Luhur University, danielpalar1000mmi@gmail.com

Corresponding Author: danielpalar1000mmi@gmail.com[1]

Abstract: Prisma Manado University, a newly established higher education institution in 2017, has faced challenges in increasing its number of new students. Due to the increasingly competitive landscape in higher education, determining effective promotional strategies has become essential. In an effort to achieve this goal, this research analyzes the application of Data Mining with a focus on clustering using the K-Means and Hierarchical Clustering algorithms. This study aims to identify groups of prospective students with similar characteristics based on available data, such as educational background, field of interest, and location preferences, so that more targeted promotional strategies can be developed. The K-Means method is used to group prospective new students into clusters based on their similar characteristics. The data used in this study includes demographic data and preferences of prospective students that have been collected by Prisma Manado University. The results of the cluster analysis will provide valuable insights into the preferences and needs of prospective students, which will then help the university in designing more efficient and suitable promotional strategies for each group of prospective students. This helps the university allocate resources more effectively and maximize promotional efforts to achieve new student enrollment targets. By applying the Data Mining approach using the K-Means and Hierarchical Clustering algorithms, this research is expected to provide a deeper understanding of the prospective students of Prisma Manado University and encourage the university to take concrete steps in improving the effectiveness of their promotional strategies. The contribution of this research is to compare two algorithms, namely K-Means Clustering and Agglomerative Hierarchical Clustering.

Keywords: Data Mining, Clustering, K-Means, Hierarchical Clustering, Promotion Strategies.

INTRODUCTION

Higher education at universities plays a crucial role in shaping the future of individuals and society. Through higher education, students gain in-depth knowledge and skills relevant to their field of study, helping them become experts and ready to contribute professionally. University graduates have better career opportunities and generally earn higher salaries

compared to those with only secondary education, as well as access to jobs requiring specialized qualifications.

In addition, higher education also contributes to students' personal development. They learn to become more independent, responsible, and develop strong critical thinking skills. Universities are centers of innovation and research, and through higher education, students can participate in research projects that bring positive change to society.

For children across Indonesia, higher education is a dream that can be achieved. Through various scholarship programs and support from the government and private institutions, children from diverse economic backgrounds have the opportunity to pursue university education. Scholarships offered by the government and private organizations help academically capable but financially disadvantaged students. The Indonesian government also continues to improve access to and quality of higher education through various policies and programs, including building better educational facilities and enhancing the quality of teaching staff.

The role of communities and private institutions is also vital in supporting higher education, by providing scholarships, training, and other assistance to help children achieve their dreams. The success stories of children who manage to continue their higher education despite coming from underprivileged families serve as inspiration for others to keep striving and never give up. With the right support, children throughout Indonesia can attain higher education and a brighter future.

Higher education at universities is essential for individual development and societal progress. According to data from the Ministry of Home Affairs, only around 6 percent of the population has pursued university education—more precisely, 6.52 percent. As of December 31, 2022, the number of Indonesians enrolled in higher education at the D1 and D2 levels was 1.11 million people or 0.4% of the total population. At the D3 level, there were 3.56 million people or 1.28%, and at the S1 (Bachelor's) level, 12.44 million people or 4.47%. At the Master's level, there were 882,113 (0.31%), but only 63,315 (0.02%) were enrolled in doctoral (S3) programs. (Kemendikbud, 2025)

Currently, there are 66.07 million Indonesians who are not enrolled in any education, representing around 23.78% of the population. Additionally, 30.89 million people (11.12%) did not complete primary school, 64.30 million (23.15%) completed primary school, 40.21 million (14.47%) completed junior high school, and 100 million (23.15%) completed senior high school. Meanwhile, 58.57 million people (21.08%) are unregistered.

Universitas Prisma Manado stands in the community as a provider of quality higher education. Since its establishment in 2017, the university has offered a variety of study programs designed to meet students' needs and interests. Through its 11 comprehensive study programs, Universitas Prisma Manado aims to provide education that not only emphasizes academic excellence but also the development of skills and character.

As a relatively young educational institution, Universitas Prisma Manado continues to innovate and adapt to meet contemporary challenges. With the support of competent lecturers and adequate facilities, the university is committed to producing graduates who are ready to contribute in various sectors, both locally and globally.

Universitas Prisma Manado is currently facing challenges in recruiting new students. These challenges arise from various factors, including increasing competition from other universities and a lack of awareness among prospective students about the university's strengths and program offerings.

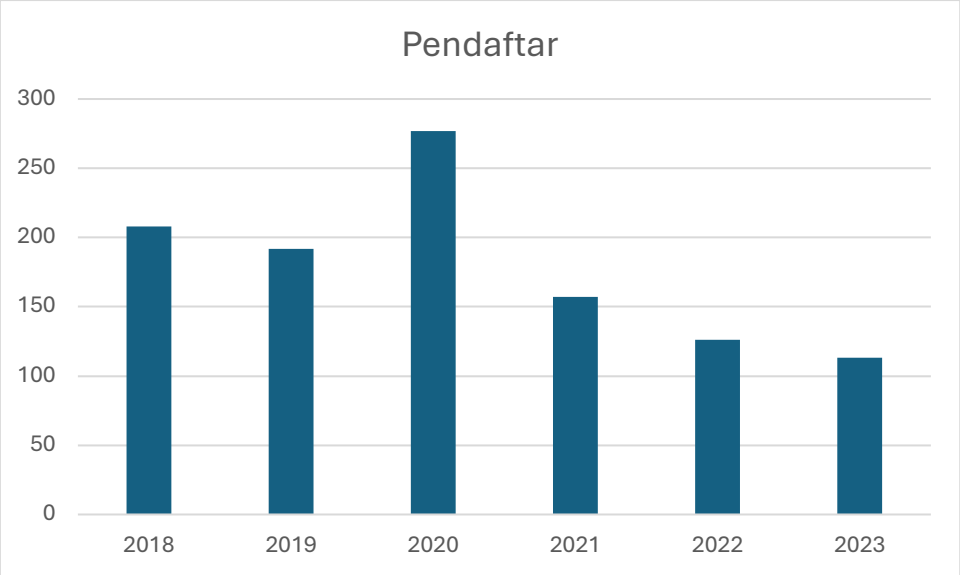


Figure 1. Number of Applicants to Prisma University

Figure 1 illustrates the number of applicants to Prisma University from 2018 to 2023. Although many factors influence the increase or decrease in student enrollment—such as the university's reputation, accreditation, quality of teaching, and others—this study focuses on the marketing strategies that affect the number of applicants at Prisma University.

The use of information technology and data analysis has become a critical component in strategic decision-making across various sectors. In the context of higher education, the application of data mining techniques, particularly clustering using the K-Means algorithm, can help universities identify potential new student segments and design more targeted promotional strategies. Data mining is an approach that enables the extraction of valuable insights from large and complex datasets, helping to identify patterns, trends, and groupings based on similar characteristics. This allows institutions to identify different market segments and tailor promotional strategies that are more specific and relevant.

Moreover, by identifying the most promising segments, universities can allocate promotional budgets and other resources more efficiently. This means focusing on the most effective channels and promotional methods to reach segmented target audiences. After identifying market segments, universities can customize promotional messages to better match the needs and preferences of each group. For example, prospective students interested in technology programs may respond better to messages emphasizing innovation and technological development at the university.

K-Means clustering also helps identify trends and preferences among prospective students. This allows universities to stay up-to-date with changing interests and needs and adjust their promotional strategies accordingly. Institutions can also evaluate the effectiveness of their promotional campaigns based on responses from various market segments, offering deeper insights into what works and what needs improvement.

Another clustering method that can assist universities in decision-making is Agglomerative Hierarchical Clustering. Unlike K-Means clustering, agglomerative hierarchical clustering does not require a predefined number of clusters. This algorithm works by gradually merging data points—starting with each data point as an individual cluster and then progressively combining them into larger clusters until all data is grouped into one large cluster. This process produces a dendrogram, a visual representation of hierarchical clustering that helps in understanding the overall structure of the data.

In the context of promotional strategies to attract new students, agglomerative hierarchical clustering offers advantages in identifying more specific and layered market

segments. This enables universities to design more accurate and relevant promotional strategies for each market segment. Additionally, this algorithm allows universities to identify sub-segments within main groups, so that promotional messages can be tailored even further to meet the specific needs and preferences of each sub-segment.

This algorithm also assists in identifying more complex patterns and trends among prospective students, enabling the university to remain relevant and responsive to shifts in trends and dynamics in the education market.

The application of data mining in the context of higher education—particularly at Prisma University Manado—is critical. In an increasingly connected and information-rich world, universities must move beyond traditional, general promotional strategies and shift towards more data-driven and tailored approaches that align with prospective student preferences. K-Means, as one of the most widely used clustering algorithms in data mining, along with Hierarchical Clustering, can help group prospective students based on characteristics such as educational background, major interests, and location preferences.

The proposed solution in this research is to perform clustering to determine promotional strategies for new student recruitment at Prisma University Manado using both K-Means and Agglomerative Hierarchical Clustering algorithms. This study aims to understand how the application of data mining through these two algorithms can enhance promotional strategies for new student enrollment at Prisma University Manado. By analyzing the preferences and needs of prospective students in a deeper and more segmented way, the university can better direct its resources and optimize promotional efforts.

The contribution of this research lies in comparing the two algorithms—K-Means Clustering and Agglomerative Hierarchical Clustering. The attributes used in the K-Means algorithm include School Origin, Regional Origin, and Promotion Reference. In contrast, the Agglomerative Hierarchical Clustering algorithm considers all fields as attributes for analysis.

METHOD

This research was conducted to obtain new students for a university by determining the most appropriate promotion target using the K-Means Algorithm and Agglomerative Hierarchical clustering (AHC). To support the data collection process, an object to be studied is needed, in this case the research will be conducted through direct observation at Prisma University. Before conducting research, of course, it is necessary to know in advance the steps to be taken, so that the research process can run in a structured manner, based on the research process that refers to data mining, this research will use one of the methods commonly used in the data mining process, namely the CRISP-DM method (Cross-Industry Standard Process for Data Mining). CRISP-DM steps that can be applied to determine new student promotion strategies using the K-Means algorithm and Agglomerative Hierarchical Clustering.

Sampling/Sample Selection Method

In determining the new student promotion strategy, the use of secondary data can provide valuable additional support. Secondary data is data that has been previously collected by various parties for a specific purpose. The components of this secondary data will be integrated with the sample selection method through the K-Means and Agglomerative Hierarchical Clustering algorithms so that it can provide a deeper understanding of student behavior. The secondary data that will be used in this study was taken from data generated from 2017 to 2023. The selection of secondary data was due to the instability of enthusiasts and applicants at the university.

Based on existing primary data, such as student registration history in several years, it can provide context about previous registration trends. This is very helpful in determining

more relevant parameters or variables to be included in the clustering process, such as promotional media preferences or what sources of information are often accessed by previous students. In sample selection, the use of secondary data allows for selecting a more representative and relevant sample, taking into account demographic data or other data that influence promotional preferences.

Data Collection Methods

The methods used in data collection include, the use of secondary data which is an effective method, involving analysis of historical data on student registration or previous promotions, then direct interviews with the marketing team either online or offline.

online surveys that can be used through online forms, this allows for fast data collection and can be accessed by many respondents, the use of social media analysis, by analyzing student interactions and responses to promotional campaigns on platforms such as Instagram and Facebook.

The use of online analytics can also be used to track student online behavior, such as website visit data and ad clicks. The combination of these data collection methods can provide a holistic and in-depth view of student needs, preferences, and behavior, thus helping to design more appropriate and effective promotional strategies. With this approach, Universitas Prisma Manado can be better at attracting prospective students and increasing the number of registrations.

RESULT AND DISCUSSION

K-Means Algorithm

In the K-Means algorithm, the first stage is to determine the number of groups. The groups or columns that will be processed in this system have been set by default, namely City of Origin, School of Origin, and Reference.

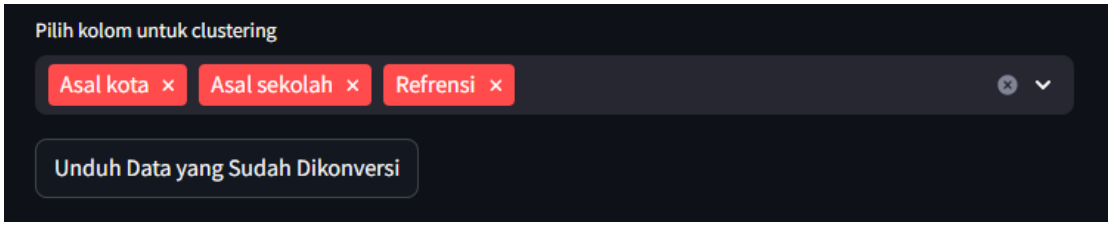


Figure 1. Column Interface to be processed by K-Means

The three columns to be processed, the data will be taken and grouped. After being grouped, the data will be converted into numbers. After the data is converted into numbers, the data will be divided into groups based on the similarities in the group.

Table 1. The result of changing the dataset into numbers

City of Origin	School of Origin	Reference
22	65	3
22	65	0
34	17	0
37	96	0
42	98	1
25	34	1
22	203	1
20	17	3
41	17	0
38	17	1

22	150	2
24	40	1
38	146	3
38	146	0
38	137	0

After the data is converted into numbers, K-Means will initialize the cluster assignment using the initial centroid. Here the author creates 3 clusters, so that the initial centroid will be determined randomly where the initial centroid is obtained as follows

- C1 = [0.0,0.0,0.0]
- C2 = [0.5,0.3333,0.3333]
- C3 = [1.0,1.0,1.0]

Then, the system will calculate the distance to the Centroid using the Euclidean formula.

$$d(x,c) = \sqrt{\sum_{i=1}^{n}(x_i - c_i)^2}$$

Figure 2. Euclidean Formula

The distance obtained by the system for centroid calculation is as follows:

Data	$d(C_1)$	$d(C_2)$	$d(C_3)$
1	0	0.623	1.732
2	0.623	0	1.226
3	0.943	0.471	1.054
4	1.732	1.226	0
5	0	0.623	1.732

Figure 3. Data Distance to Centroid

From the calculated distance results, the system will assign clusters based on the closest distance and the initial clustering results obtained are as follows:

- Data 1 → C_1
- Data 2 → C_2
- Data 3 → C_2
- Data 4 → C_3
- Data 5 → C_1

Figure 4. Data Distance to Centroid

The system will continue to update new centroids as the average data in each cluster and reassign data to the cluster until after several iterations a stable cluster is obtained. The following is the final result of K-Means clustering that is already stable from several sampled data.

	Asal kota	Asal sekolah	Refrensi	KMeans_Cluster
0	Kab. Minahasa	SMAN 3 TONDANO	Sosial Media	0
1	Kab. Minahasa	SMAN 3 TONDANO	Pencarian Goc	0
2	Kota Jakarta Pusat	SMA Prisma	Pencarian Goc	0
3	Kota Makassar	SMAS KATOLIK SANTO DOMINIKUS	Kerabat	1
4	Kota. Kotamobagu	SMAS KATOLIK THEODORUS KOTAMOBAGU	Kerabat	1
5	Kab. Minahasa Utara	SMAN 1 MANADO	Brosur	1
6	Kab. Minahasa	SMKS YADIKA	Brosur	2
7	Kab. Marowali	SMA Prisma	Pencarian Goc	0
8	Kota Tomohon	SMA Prisma	Sosial Media	0
9	Kota Manado	SMA Prisma	Pencarian Goc	0
10	Kab. Minahasa	SMKN 1 SONDER	Brosur	1
11	Kab. Minahasa Tengga	SMAN 1 PUSUMAEN	Brosur	1
12	Kota Manado	SMKN 1 MANADO	Sosial Media	0
13	Kota Manado	SMKN 1 MANADO	Brosur	1
14	Kota Manado	SMK NEGERI 1 SINONSAYANG	Brosur	1
15	Kab. Lanny Jaya	SMAN 1 TIOM	Pencarian Goc	2
16	Kota Manado	SMAS TRIDHARMA	Pencarian Goc	0
17	Kota Manado	SMKN 6 MANADO	Kerabat	1
18	Kab. Bolaang Mongon	SMKS BPD NANASI	Brosur	2
19	Kab. Minahasa Utara	SMAN 8 MANADO	Kerabat	1
20	Kab. Banggai	SMKN 6 MANADO	Kerabat	2
21	Kota Manado	SMAN 7 MANADO	Kerabat	1
22	Kota Manado	SMAN 7 MANADO	Kerabat	1
23	Kota Manado	SMAS ADVENT KLABAT	Sosial Media	0
24	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Sosial Media	0

Figure 5. Final Result of K-Means Clustering

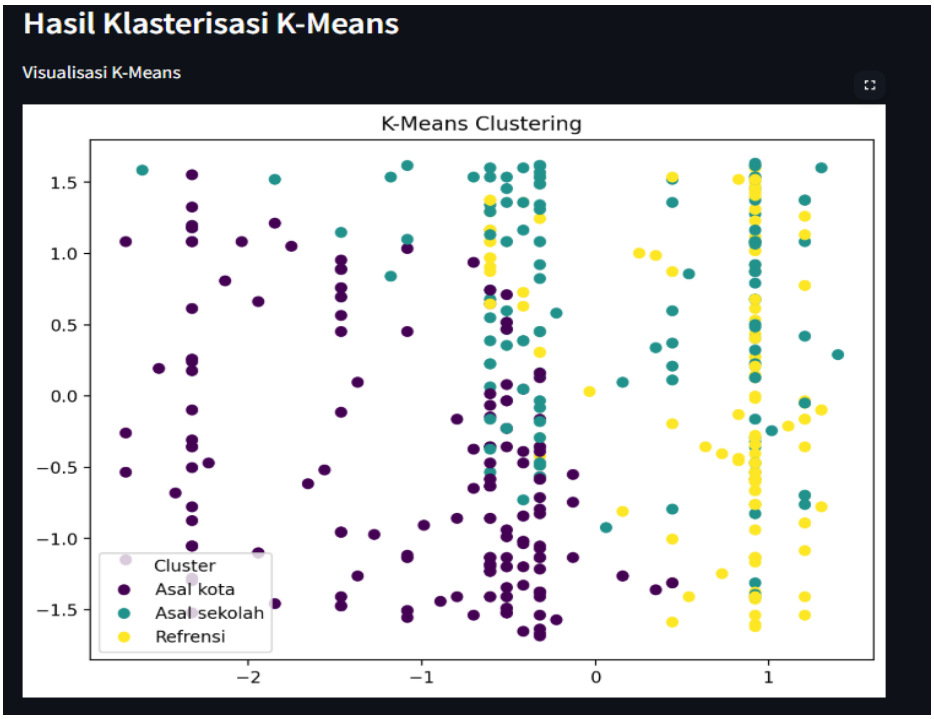
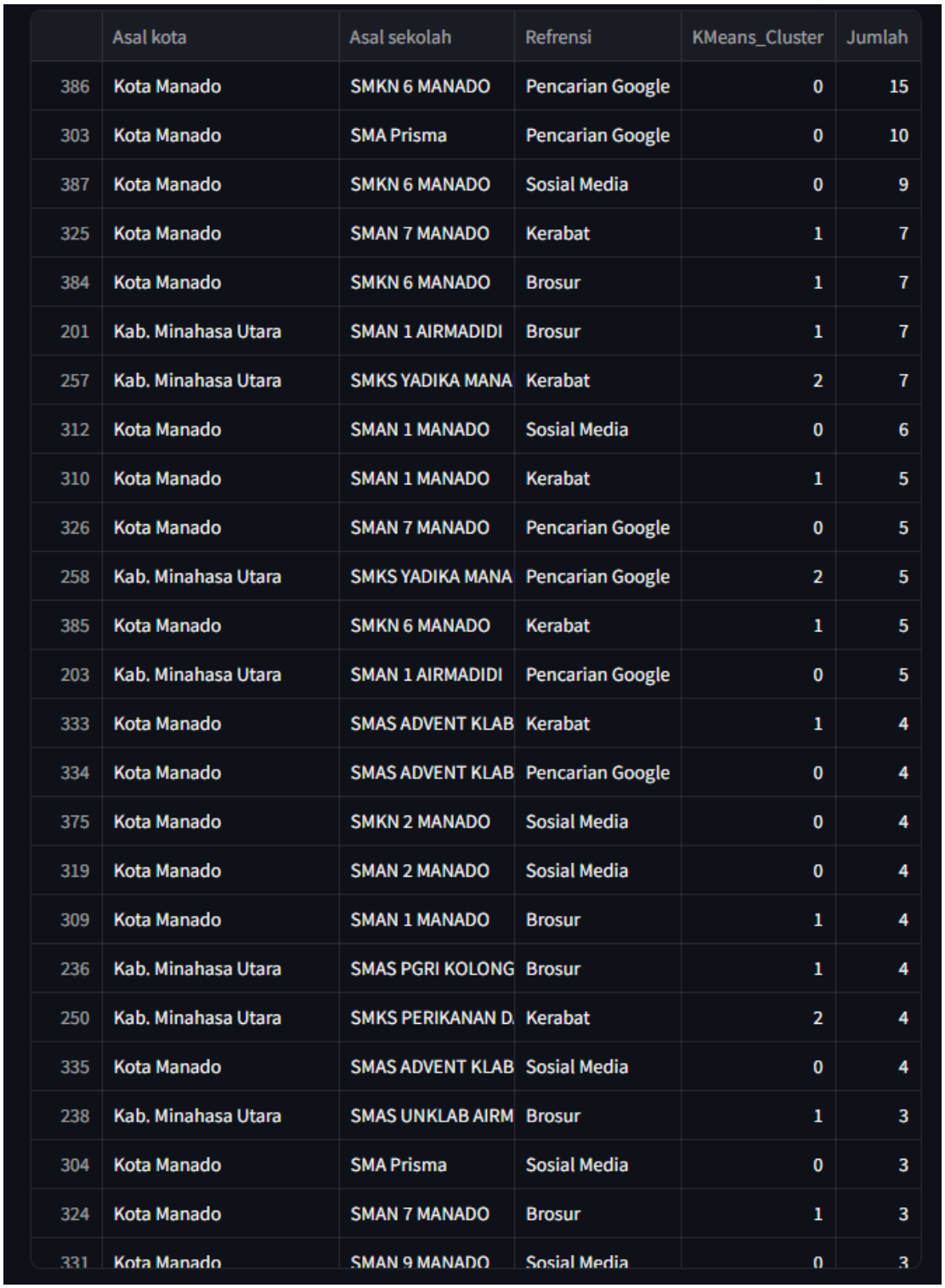


Figure 6. K-Means Visualization

After the clustering results are obtained, the system will perform a promotion analysis by calculating the amount of data in each cluster using the following command:

col_analysis_kmeans = data.groupby([col.strip() for col in columns] + ['KMeans_Cluster']).size().reset_index(name='Jumlah').sort_values(by='Jumlah', ascending=False)

And the results of the analysis obtained are as follows:



	Asal kota	Asal sekolah	Refrensi	KMeans_Cluster	Jumlah
386	Kota Manado	SMKN 6 MANADO	Pencarian Google	0	15
303	Kota Manado	SMA Prisma	Pencarian Google	0	10
387	Kota Manado	SMKN 6 MANADO	Sosial Media	0	9
325	Kota Manado	SMAN 7 MANADO	Kerabat	1	7
384	Kota Manado	SMKN 6 MANADO	Brosur	1	7
201	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Brosur	1	7
257	Kab. Minahasa Utara	SMKS YADIKA MANA	Kerabat	2	7
312	Kota Manado	SMAN 1 MANADO	Sosial Media	0	6
310	Kota Manado	SMAN 1 MANADO	Kerabat	1	5
326	Kota Manado	SMAN 7 MANADO	Pencarian Google	0	5
258	Kab. Minahasa Utara	SMKS YADIKA MANA	Pencarian Google	2	5
385	Kota Manado	SMKN 6 MANADO	Kerabat	1	5
203	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Pencarian Google	0	5
333	Kota Manado	SMAS ADVENT KLAB	Kerabat	1	4
334	Kota Manado	SMAS ADVENT KLAB	Pencarian Google	0	4
375	Kota Manado	SMKN 2 MANADO	Sosial Media	0	4
319	Kota Manado	SMAN 2 MANADO	Sosial Media	0	4
309	Kota Manado	SMAN 1 MANADO	Brosur	1	4
236	Kab. Minahasa Utara	SMAS PGRI KOLONG	Brosur	1	4
250	Kab. Minahasa Utara	SMKS PERIKANAN D	Kerabat	2	4
335	Kota Manado	SMAS ADVENT KLAB	Sosial Media	0	4
238	Kab. Minahasa Utara	SMAS UNKLAB AIRM	Brosur	1	3
304	Kota Manado	SMA Prisma	Sosial Media	0	3
324	Kota Manado	SMAN 7 MANADO	Brosur	1	3
331	Kota Manado	SMAN 9 MANADO	Sosial Media	0	3

Figure 7. Promotion Analysis

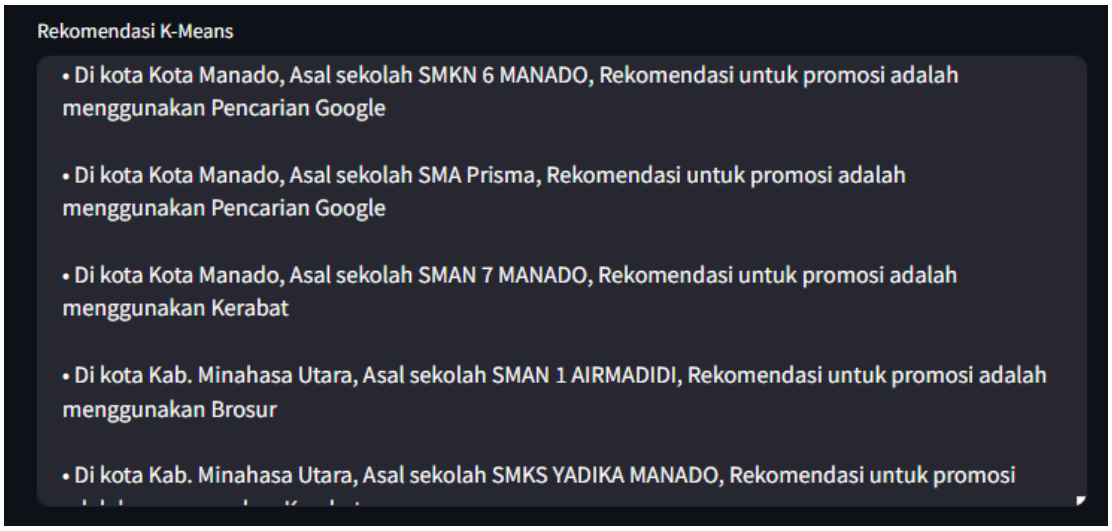


Figure 8. Promotion Recommendations

The K-Mean clustering process takes 0.50 seconds with an average distance to the Centroid of 1.2286. The command for calculating the time and average distance is as follows:
Time Calculation:
st.write(f"Waktu Proses K-Means: {end_time_kmeans - start_time_kmeans:.2f} detik")
Perhitungan Jarak Rata-rata:
*centroids = kmeans.cluster_centers_*
avg_distance = np.mean([cdist(X_scaled[kmeans_labels == i], [centroids[i]], 'euclidean').mean() for i in range(number_cluster)])
st.write(f"Rata-rata Jarak ke Centroid (K-Means): {avg_distance:.4f}")

Hierarchical Clustering Algorithm

In the Hierarchical Clustering algorithm, the first stage is the algorithm will combine data gradually, starting from individual clusters to become a large cluster. The groups or columns that will be processed in this system have been set by default, namely City of Origin, School of Origin, and Reference.

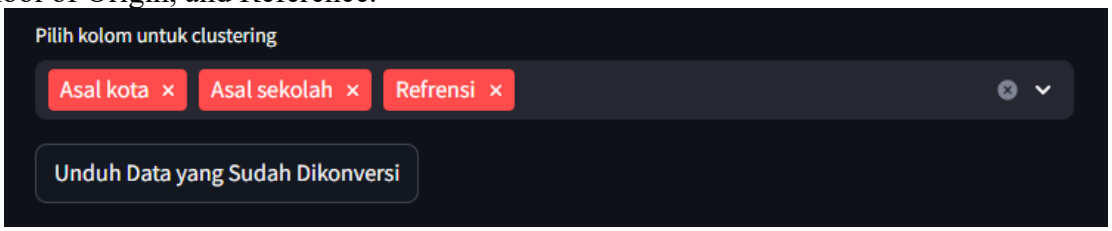


Figure 9. Column Interface to be processed by K-Means

The three columns to be processed, the data will be taken and grouped. After being grouped, the data will be converted into numbers. After the data is converted into numbers, the data will be divided into groups based on the similarities in the group.

Table 2. The result of changing the dataset into numbers

City of Origin	School of Origin	Reference
22	65	3
22	65	0
34	17	0
37	96	0
42	98	1
25	34	1

22	203	1
20	17	3
41	17	0
38	17	1
22	150	2
24	40	1
38	146	3
38	146	0
38	137	0

Based on the existing dataset, each student is a group. After that, the process of searching and grouping the most similar data will be carried out by calculating the distance between data points. For example, data that comes from the same city will be one group. Or data that has similarities from the same school will be grouped into one group. And so on. The formula used to find the distance between data is as follows.

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

Figure 10. Euclidean formula used by Hierarchical

The distance between data obtained by the system is as follows:

Data	1	2	3	4	5
1	0.0000	0.623	0.943	1.732	0.0000
2	0.623	0.0000	0.471	1.226	0.623
3	0.943	0.471	0.0000	1.054	0.943
4	1.732	1.226	1.054	0.0000	1.732
5	0.0000	0.623	0.943	1.732	0.0000

Figure 11. Euclidean results used by Hierarchical

Unlike K-Means, Hierarchical will merge the closest clusters into a large cluster, as follows

Iteration 1:

- Closest distance: Data 1 and 5 (d=0.0d = 0.0d=0.0).
- Merge Data 1 and 5 into Cluster C1C_1C1: [1,5][1, 5][1,5].

Iteration 2:

- New cluster:
  - Data 2 and Data 3 have the next closest distance (d=0.471d = 0.471d=0.471).
- Merge Data 2 and Data 3 into Cluster C2C_2C2: [2,3][2, 3][2,3].

Iterasi 3:

- Merge Cluster C1C_1C1 ([1,5][1, 5][1,5]) with Cluster C2C_2C2 ([2,3][2, 3][2,3]) based on minimum distance.

Iteration 4:

- Merge all clusters into one large cluster.

The Clustering Results obtained are as follows:

	Asal kota	Asal sekolah	Refrensi	Hierarchical_Cluster
0	Kab. Minahasa	SMAN 3 TONDANO	Sosial Media	0
1	Kab. Minahasa	SMAN 3 TONDANO	Pencarian Goc	0
2	Kota Jakarta Pusat	SMA Prisma	Pencarian Goc	0
3	Kota Makassar	SMAS KATOLIK SANTO DOMINIKUS	Kerabat	2
4	Kota. Kotamobagu	SMAS KATOLIK THEODORUS KOTAMOBAGU	Kerabat	2
5	Kab. Minahasa Utara	SMAN 1 MANADO	Brosur	0
6	Kab. Minahasa	SMKS YADIKA	Brosur	1
7	Kab. Marowali	SMA Prisma	Pencarian Goc	0
8	Kota Tomohon	SMA Prisma	Sosial Media	0
9	Kota Manado	SMA Prisma	Pencarian Goc	0
10	Kab. Minahasa	SMKN 1 SONDER	Brosur	1
11	Kab. Minahasa Tengga	SMAN 1 PUSUMAEN	Brosur	0
12	Kota Manado	SMKN 1 MANADO	Sosial Media	1
13	Kota Manado	SMKN 1 MANADO	Brosur	2
14	Kota Manado	SMK NEGERI 1 SINONSAYANG	Brosur	2
15	Kab. Lanny Jaya	SMAN 1 TIOM	Pencarian Goc	0
16	Kota Manado	SMAS TRIDHARMA	Pencarian Goc	0
17	Kota Manado	SMKN 6 MANADO	Kerabat	1
18	Kab. Bolaang Mongon	SMKS BPD NANASI	Brosur	1
19	Kab. Minahasa Utara	SMAN 8 MANADO	Kerabat	0
20	Kab. Banggai	SMKN 6 MANADO	Kerabat	1
21	Kota Manado	SMAN 7 MANADO	Kerabat	2
22	Kota Manado	SMAN 7 MANADO	Kerabat	2
23	Kota Manado	SMAS ADVENT KLABAT	Sosial Media	0
24	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Sosial Media	0

Figure 12. Hierarchical Results

Different from K-Means visualization, for Hierarchical visualization a dendogram is used where the dendogram will visualize how data or clusters are combined and the formula used is as follows:

$$D(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{x \in C_i} \sum_{y \in C_j} d(x, y)$$

Figure 13. Average Linkage Formula

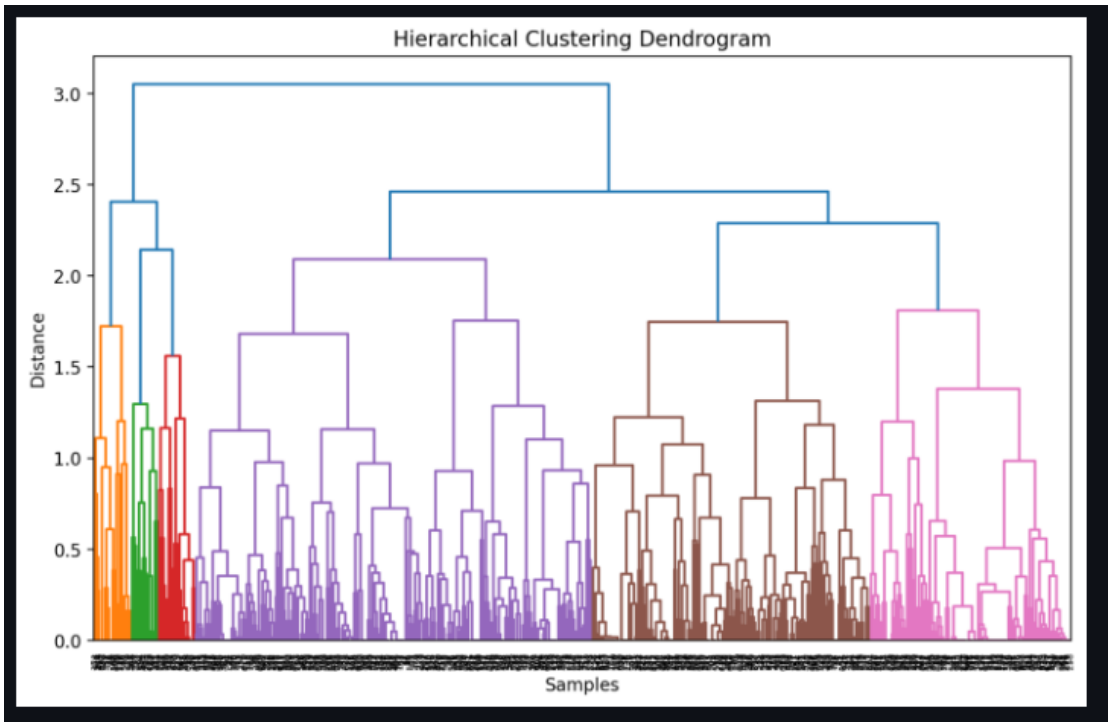


Figure 14. Hierarchical Dendogram Visualization

After the clustering results are obtained, the system will perform a promotion analysis by calculating the amount of data in each cluster using the following command.

col_analysis_hier = data.groupby([col.strip() for col in columns] + ['KMeans_Cluster']).size().reset_index(name='Jumlah').sort_values(by='Jumlah', ascending=False)

From the results of the analysis carried out, the following results were obtained:

	Asal kota	Asal sekolah	Refrensi	Hierarchical_Cluster	Jumlah
386	Kota Manado	SMKN 6 MANADO	Pencarian Google	1	15
303	Kota Manado	SMA Prisma	Pencarian Google	0	10
387	Kota Manado	SMKN 6 MANADO	Sosial Media	1	9
325	Kota Manado	SMAN 7 MANADO	Kerabat	2	7
384	Kota Manado	SMKN 6 MANADO	Brosur	2	7
201	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Brosur	0	7
257	Kab. Minahasa Utara	SMKS YADIKA MANA	Kerabat	1	7
312	Kota Manado	SMAN 1 MANADO	Sosial Media	0	6
310	Kota Manado	SMAN 1 MANADO	Kerabat	2	5
326	Kota Manado	SMAN 7 MANADO	Pencarian Google	0	5
258	Kab. Minahasa Utara	SMKS YADIKA MANA	Pencarian Google	1	5
385	Kota Manado	SMKN 6 MANADO	Kerabat	1	5
203	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Pencarian Google	0	5
333	Kota Manado	SMAS ADVENT KLAB	Kerabat	2	4
334	Kota Manado	SMAS ADVENT KLAB	Pencarian Google	0	4
375	Kota Manado	SMKN 2 MANADO	Sosial Media	1	4
319	Kota Manado	SMAN 2 MANADO	Sosial Media	0	4
309	Kota Manado	SMAN 1 MANADO	Brosur	2	4
236	Kab. Minahasa Utara	SMAS PGRI KOLONG	Brosur	1	4
250	Kab. Minahasa Utara	SMKS PERIKANAN D	Kerabat	1	4
335	Kota Manado	SMAS ADVENT KLAB	Sosial Media	0	4
238	Kab. Minahasa Utara	SMAS UNKLAB AIRM	Brosur	1	3
304	Kota Manado	SMA Prisma	Sosial Media	0	3
324	Kota Manado	SMAN 7 MANADO	Brosur	2	3
331	Kota Manado	SMAN 9 MANADO	Sosial Media	0	3

Figure 15. Promotion Analysis Results

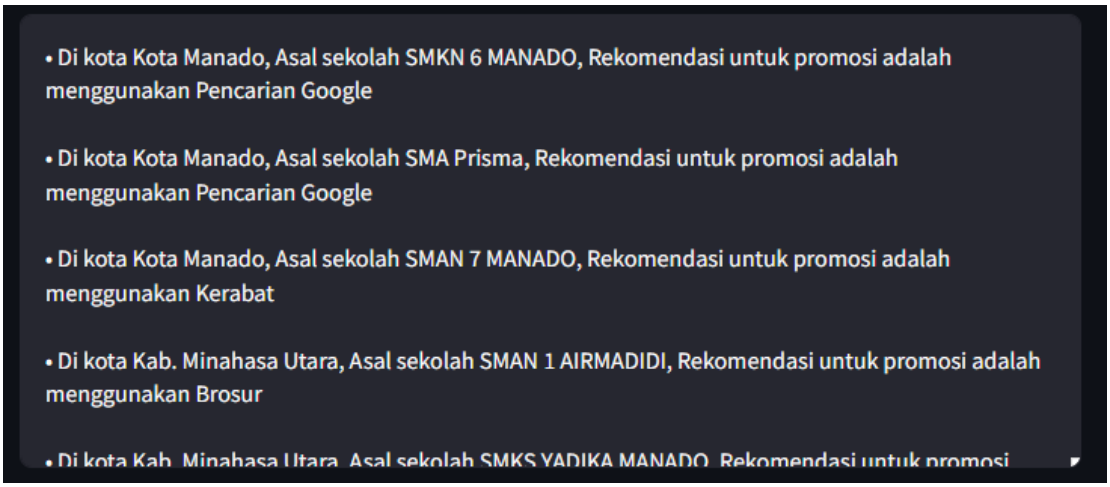


Figure 16. Hierarchical Clustering Recommendations

The Hierarchical clustering process takes 13.30 seconds with an average distance between data of 1.6641. The command for calculating the time and average distance is as follows:

Time Calculation:

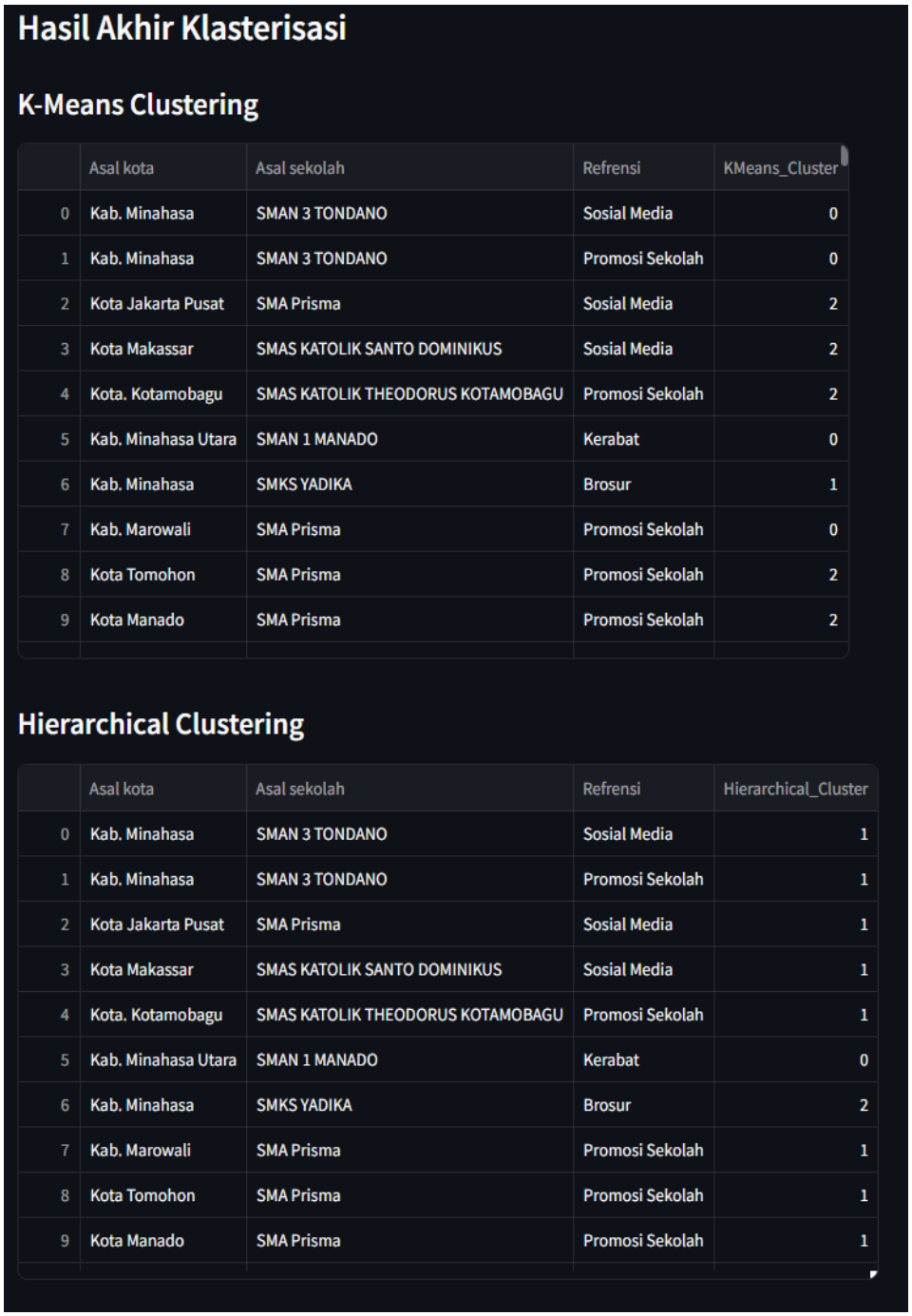
```
st.write(f"Waktu Proses Hierarchical Clustering: {end_time_hier - start_time_hier:.2f} detik")
```

Average Distance Calculation:

```
avg_distance_hierarchical = np.mean([
pdist(X_scaled[hierarchical_labels == i], 'euclidean').mean() if
len(X_scaled[hierarchical_labels == i]) > 1 else 0
for i in range(number_cluster)
])
st.write(f"Rata-rata Jarak Antar Titik (Hierarchical): {avg_distance_hierarchical:.4f}")
```

Final Results of Clustering and Promotion Potential Analysis

After the data is processed, the final results of the clustering will be shown, along with the Promotion Potential analysis, and Promotion Recommendations. The final results will be placed in a table form.



Hasil Akhir Klasterisasi

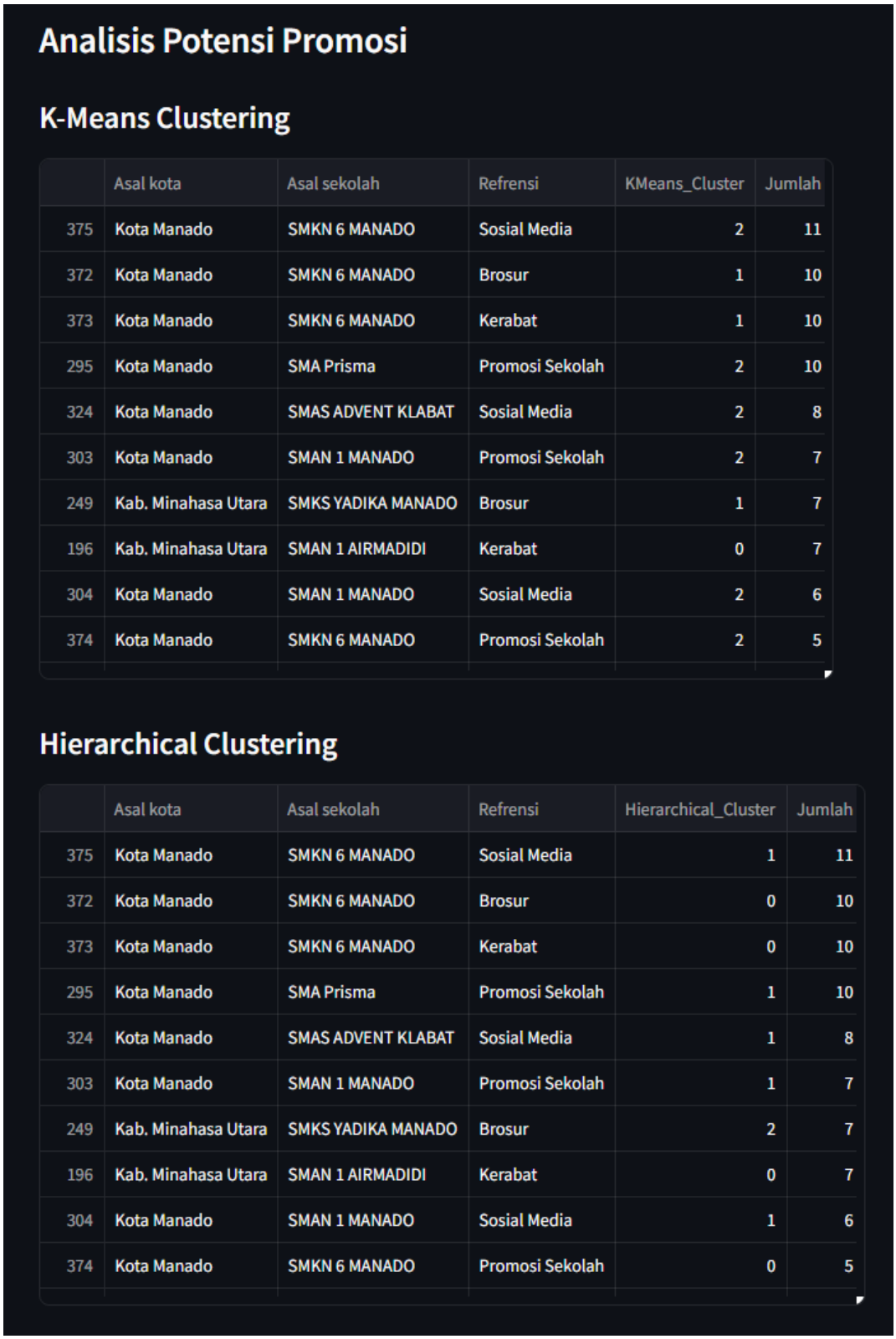
K-Means Clustering

	Asal kota	Asal sekolah	Refrensi	KMeans_Cluster
0	Kab. Minahasa	SMAN 3 TONDANO	Sosial Media	0
1	Kab. Minahasa	SMAN 3 TONDANO	Promosi Sekolah	0
2	Kota Jakarta Pusat	SMA Prisma	Sosial Media	2
3	Kota Makassar	SMAS KATOLIK SANTO DOMINIKUS	Sosial Media	2
4	Kota. Kotamobagu	SMAS KATOLIK THEODORUS KOTAMOBAGU	Promosi Sekolah	2
5	Kab. Minahasa Utara	SMAN 1 MANADO	Kerabat	0
6	Kab. Minahasa	SMKS YADIKA	Brosur	1
7	Kab. Marowali	SMA Prisma	Promosi Sekolah	0
8	Kota Tomohon	SMA Prisma	Promosi Sekolah	2
9	Kota Manado	SMA Prisma	Promosi Sekolah	2

Hierarchical Clustering

	Asal kota	Asal sekolah	Refrensi	Hierarchical_Cluster
0	Kab. Minahasa	SMAN 3 TONDANO	Sosial Media	1
1	Kab. Minahasa	SMAN 3 TONDANO	Promosi Sekolah	1
2	Kota Jakarta Pusat	SMA Prisma	Sosial Media	1
3	Kota Makassar	SMAS KATOLIK SANTO DOMINIKUS	Sosial Media	1
4	Kota. Kotamobagu	SMAS KATOLIK THEODORUS KOTAMOBAGU	Promosi Sekolah	1
5	Kab. Minahasa Utara	SMAN 1 MANADO	Kerabat	0
6	Kab. Minahasa	SMKS YADIKA	Brosur	2
7	Kab. Marowali	SMA Prisma	Promosi Sekolah	1
8	Kota Tomohon	SMA Prisma	Promosi Sekolah	1
9	Kota Manado	SMA Prisma	Promosi Sekolah	1

Figure 17. Visualization of the final clustering results



Analisis Potensi Promosi

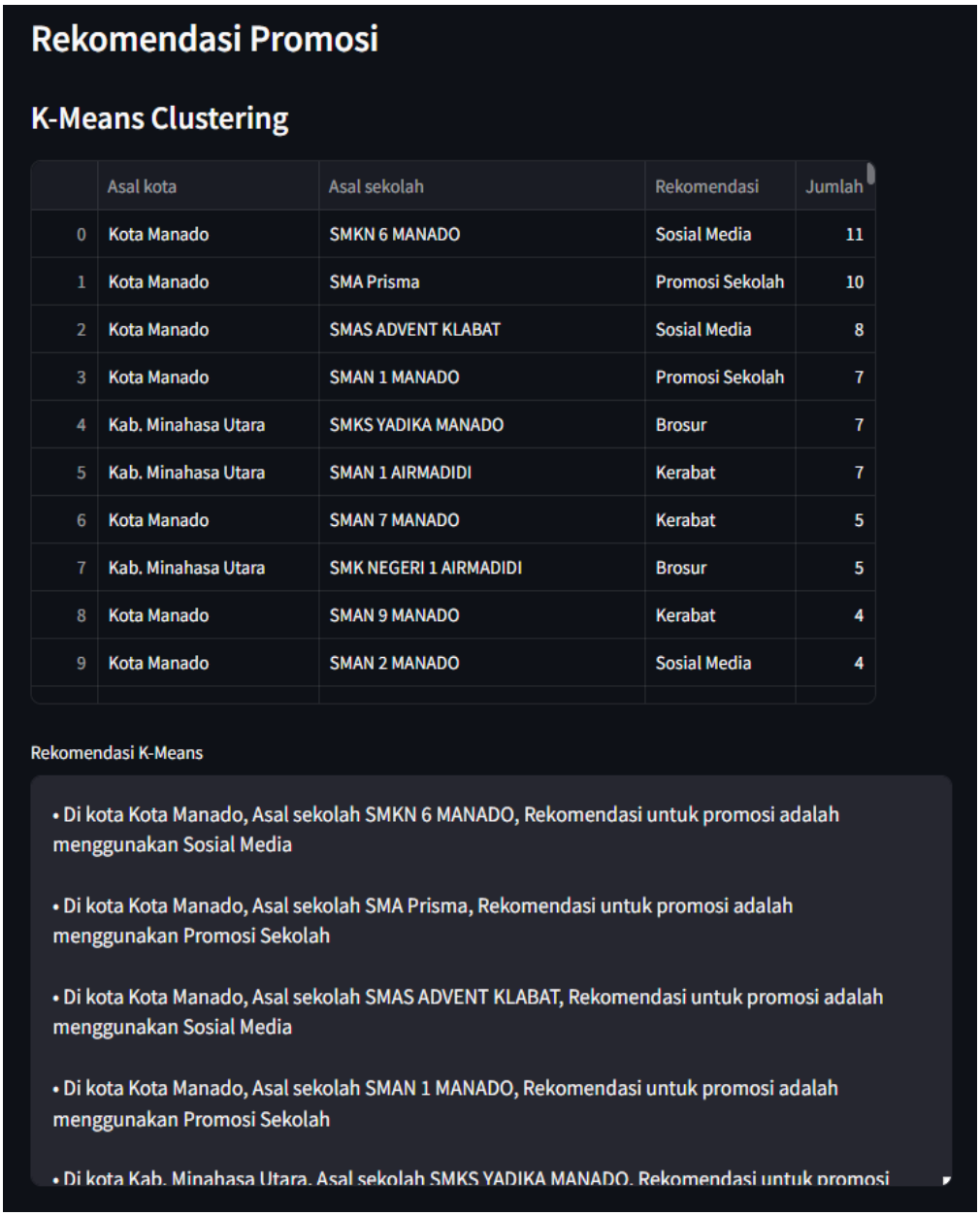
K-Means Clustering

	Asal kota	Asal sekolah	Refrensi	KMeans_Cluster	Jumlah
375	Kota Manado	SMKN 6 MANADO	Sosial Media	2	11
372	Kota Manado	SMKN 6 MANADO	Brosur	1	10
373	Kota Manado	SMKN 6 MANADO	Kerabat	1	10
295	Kota Manado	SMA Prisma	Promosi Sekolah	2	10
324	Kota Manado	SMAS ADVENT KLABAT	Sosial Media	2	8
303	Kota Manado	SMAN 1 MANADO	Promosi Sekolah	2	7
249	Kab. Minahasa Utara	SMKS YADIKA MANADO	Brosur	1	7
196	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Kerabat	0	7
304	Kota Manado	SMAN 1 MANADO	Sosial Media	2	6
374	Kota Manado	SMKN 6 MANADO	Promosi Sekolah	2	5

Hierarchical Clustering

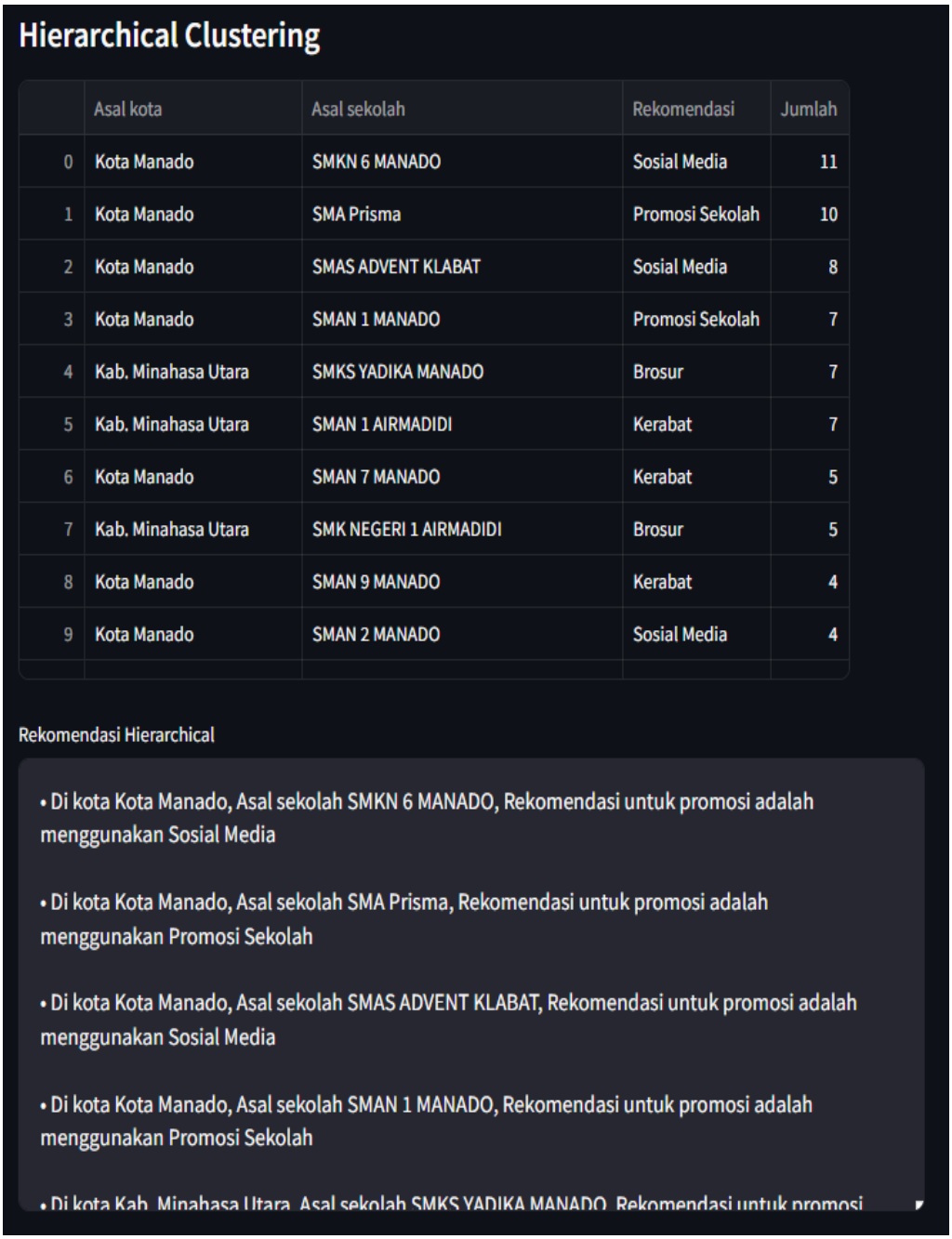
	Asal kota	Asal sekolah	Refrensi	Hierarchical_Cluster	Jumlah
375	Kota Manado	SMKN 6 MANADO	Sosial Media	1	11
372	Kota Manado	SMKN 6 MANADO	Brosur	0	10
373	Kota Manado	SMKN 6 MANADO	Kerabat	0	10
295	Kota Manado	SMA Prisma	Promosi Sekolah	1	10
324	Kota Manado	SMAS ADVENT KLABAT	Sosial Media	1	8
303	Kota Manado	SMAN 1 MANADO	Promosi Sekolah	1	7
249	Kab. Minahasa Utara	SMKS YADIKA MANADO	Brosur	2	7
196	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Kerabat	0	7
304	Kota Manado	SMAN 1 MANADO	Sosial Media	1	6
374	Kota Manado	SMKN 6 MANADO	Promosi Sekolah	0	5

Figure 18. Promotion Potential Analysis



	Asal kota	Asal sekolah	Rekomendasi	Jumlah
0	Kota Manado	SMKN 6 MANADO	Sosial Media	11
1	Kota Manado	SMA Prisma	Promosi Sekolah	10
2	Kota Manado	SMAS ADVENT KLABAT	Sosial Media	8
3	Kota Manado	SMAN 1 MANADO	Promosi Sekolah	7
4	Kab. Minahasa Utara	SMKS YADIKA MANADO	Brosur	7
5	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Kerabat	7
6	Kota Manado	SMAN 7 MANADO	Kerabat	5
7	Kab. Minahasa Utara	SMK NEGERI 1 AIRMADIDI	Brosur	5
8	Kota Manado	SMAN 9 MANADO	Kerabat	4
9	Kota Manado	SMAN 2 MANADO	Sosial Media	4

Figure 19. Recommendations for K-Means Clustering Results



Hierarchical Clustering

	Asal kota	Asal sekolah	Rekomendasi	Jumlah
0	Kota Manado	SMKN 6 MANADO	Sosial Media	11
1	Kota Manado	SMA Prisma	Promosi Sekolah	10
2	Kota Manado	SMAS ADVENT KLABAT	Sosial Media	8
3	Kota Manado	SMAN 1 MANADO	Promosi Sekolah	7
4	Kab. Minahasa Utara	SMKS YADIKA MANADO	Brosur	7
5	Kab. Minahasa Utara	SMAN 1 AIRMADIDI	Kerabat	7
6	Kota Manado	SMAN 7 MANADO	Kerabat	5
7	Kab. Minahasa Utara	SMK NEGERI 1 AIRMADIDI	Brosur	5
8	Kota Manado	SMAN 9 MANADO	Kerabat	4
9	Kota Manado	SMAN 2 MANADO	Sosial Media	4

Figure 20. Recommendations for Hierarchical Clustering Results

The calculated results table can be exported in Excel format.

```
# Export hasil
output_file = 'hasil_klasterisasi.xlsx'
data.to_excel(output_file, index=False)
with open(output_file, 'rb') as f:
    st.download_button("Unduh Hasil Klasterisasi", f, file_name=output_file)
```

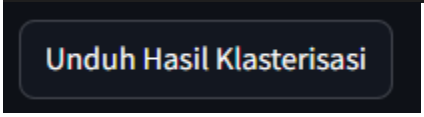


Figure 21. Export Excel data

Results

From each implementation process that has been carried out, the following results were obtained:

1. At the data processing stage using the K-Means Clustering algorithm, with key attributes using School Origin, City Origin and Promotion Strategy, and the processed dataset of 647 data, it was found that the processing time was 0.50 seconds. And the average distance to the centroid in the K-Means algorithm is 1.2408
2. At the data processing stage using the Hierarchical Clustering algorithm, with the processed dataset of 647 data, it was found that the processing time was 2.09 seconds. And the average distance between points was 1.7857
3. The results of the analysis obtained based on the 2 algorithms tested, showed similarities. In the city of Manado, the origin of the school SMKN 6 MANADO, the K-Means recommendation for promotion is to use Social Media. Likewise, the results of Hierarchical Clustering, in the city of Manado, the origin of the school SMKN 6 MANADO, the recommendation of Hierarchical Clustering for promotion is to use Social Media.

CONCLUSION

The following are the temporary conclusions of the research conducted

1. In the K-Means process, with a dataset of 647, the processing time was 0.50 seconds, with an average distance to the Centroid of 1.2408. While in the Hierarchical Clustering process, the time required was 2.09 with the same dataset, and the average distance between points was 1.7857. Based on these results, K-Means Clustering has a faster time. This is because the K-Means calculation only requires 3 attributes, namely school origin, city origin and promotion strategy. While in Hierarchical Clustering, the attributes calculated are each student. So that 647 data are calculated one by one. This causes the processing time on Hierarchical to be longer
2. The analysis obtained on K-Means clustering and Hierarchical is the same. This proves that based on the results, both algorithms are equally good. But based on the processing time, K-Means has a faster processing time.

REFERENCES

Abriyanto, A., & Damastuti, N. (2019). SEGMENTASI MAHASISWA DENGAN "UNSUPERVISED" ALGORITMA GUNA MEMBANGUN STRATEGI MARKETING PENERIMAAN MAHASISWA. *Jurnal Insand Comtech*, *4*(2).

Analisis Dan Penerapan, ., Handayanto, A., Latifa, K., Saputro, N. D., & Waliyansyah, R. R. (2019). *Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi (Analysis and Application of Algorithm Support Vector Machine (SVM) in Data Mining to Support Promotional Strategies)* (Vol. 7, Issue 2).

Budiman, R. (2019). Penerapan Data Mining Untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru Pada Universitas Banten Jaya (Metode K-Means Clustering). *Jurnal ProTekInfo*, *6*(1), 2406–7741.

Dharmayanti, D., Mukharil Bachtiar, A., & Prasetyo, A. C. (2017). PENERAPAN METODE CLUSTERING UNTUK MEMBENTUK KELOMPOK BELAJAR MENGGUNAKAN DI SMPN 19 BANDUNG. *Ilmiah Komputer Dan*, *6*(2).

Kristanto, T., Hadiansyah, W. M., Nasrullah, M., Amalia, A., Anggraini, E. Y., & Firmansyah, A. (2020). Strategi Pemasaran Digital Dalam Peningkatan Penerimaan Mahasiswa Baru Menggunakan Analisis SWOT. *MULTINETICS*, *6*(2), 128–133. https://doi.org/10.32722/multinetics.v6i2.3415

Lestari, W., Bina, S., & Kendari, B. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari). *Jurnal Sistem Informasi Dan Sistem Komputer*, *4*(2). http://e-jurnal.stmikbinsa.ac.id/index.php/simkom35

Mukharil Bachtiar, A., Dharmayanti, D., & Hamzah, R. L. (2017). PENERAPAN METODE HIERARCHICAL AGGLOMERATIVE CLUSTERING UNTUK SEGMENTASI PELANGGAN POTENSIAL DI JEGER JERSEY INDONESIA. *Ilmiah Komputer Dan*, *6*(1).

Schüller, D., & Pekárek, J. (2016). Market Attractiveness Classification of European Union Countries for Establishing Logistics Centres. *Acta Oeconomica Pragensia*, *24*(5), 3–13. https://doi.org/10.18267/j.aop.554

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, *8*, 80716–80727. https://doi.org/10.1109/ACCESS.2020.2988796

Suhanda, Y., Kurniati, I., & Norma, S. (2020). Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik. *Jurnal Teknologi Informatika Dan Komputer*, *6*(2), 12–20. https://doi.org/10.37012/jtik.v6i2.299

Tri, A., Dani, R., Wahyuningsih, S., & Rizki, N. A. (2019). Penerapan Hierarchical Clustering Metode Agglomerative pada Data Runtun Waktu. *Jambura Journal of Mathematics*, *1*. http://ejurnal.ung.ac.id/index.php/jjom,P-

Wang, S., Li, M., Hu, N., Zhu, E., Hu, J., Liu, X., & Yin, J. (2019). K-Means Clustering With Incomplete Data. *IEEE Access*, *7*, 69162–69171. https://doi.org/10.1109/ACCESS.2019.2910287

Wang, X. D., Chen, R. C., Yan, F., Zeng, Z. Q., & Hong, C. Q. (2019). Fast Adaptive K-Means Subspace Clustering for High-Dimensional Data. *IEEE Access*, *7*, 42639–42651. https://doi.org/10.1109/ACCESS.2019.2907043

Yunita, F. (2018). PENERAPAN DATA MINING MENGGUNKAN ALGORITMA K-MEANS CLUSTRING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS ISLAM INDRAGIRI). In *Jurnal SISTEMASI* (Vol. 7).