Selangor Science &Technology Review
Vol. 9, No. 2, (2025)

Marketing Intelligence in Determining New Student Admission Promotion Strategy with K-Means Clustering

Widyastuti Andriyani[1)], Setyawan Widyarto[2)]
Digital Technology University of Indonesia, Yogyakarta
widya@utdi.ac.id [1)]
Centre for Graduate Studies (CGS), Universiti Selangor, Malaysia[)]
swidyarto@unisel.edu.my

***Abstract:** Nowadays, the need for information technology to support organizational success is highly significant. Information technology is now utilized in almost all services, including complex activities such as marketing and promotional efforts. Similar to other private universities, Jenderal Achmad Yani University Yogyakarta (Unjaya) remains dependent on student enrollment as a primary source of income, which is often unstable. This study aims to apply Marketing Intelligence by analyzing data from Unjaya's new student admissions between 2014 and 2018 using the K-Means Clustering and Confusion Matrix methods. The findings indicate that these methods can assist New Student Admission (PMB) stakeholders in identifying student clusters. Based on these clusters, several strategic alternatives can be formulated to guide PMB stakeholders in determining effective promotional targets, thereby ensuring the achievement of planned enrollment goals.*

***Keywords:** Business Intelligence, Confusion Matrix, K-Means Clustering, Marketing Intelligence, New Student Admission, Promotion Strategy Universitas Jenderal Achmad Yani Yogyakarta,*

I. INTRODUCTION

In the current era of globalization, the need for information technology facilities to support the success of an organization is highly significant. The use of information technology which was initially only a means to facilitate the administrative needs of the organization is widely applied in almost all services, even in some very complex activities to support the organization, one of them is in supporting marketing and promotion activities. One of the key information technologies to support marketing and promotion activities is Business Intelligence. Business Intelligence is a systematic process to enhance the performance of an organization through the utilization of various data, information, and knowledge possessed by the organization in the decision-making process. The use of Business Intelligence in marketing and promotion strategies is intended to provide evidence-based insights for managerial decision-making by providing various data, information, and knowledge that exist in the organizational environment, where a supporting component in Business Intelligence is Marketing Intelligence (Aaker et al., 2012).

This study focuses on analyzing the use of Marketing Intelligence with the K-Means Clustering method. Clustering is part of the Unsupervised Learning Algorithm that works by inferring patterns from data sets without referring to known results. The K-Means Clustering method is a method that works by grouping data based on similarities and minimizing variations between data in a cluster so that groups that have similar attributes will be produced. The K-Means Clustering method was chosen as the method used in this study because this study aims to present an alternative strategy for promoting PMB Unjaya by grouping data, information, and knowledge from the new student data they have. Thus, it is expected that promotional target clusters can be systematically generated based on data, information, and insights derived from Unjaya's new student records. These clusters may serve as a strategic reference for PMB Unjaya policymakers in evaluating and selecting promotion targets that align with institutional objectives, thereby supporting the achievement of planned enrollment figures..

II. BASIC THEORY

Numerous studies have contributed to the foundational literature of this research, particularly in the domains of Business Intelligence, Marketing Intelligence, K-Means Clustering, and Confusion Matrix methodologies. These studies are summarized as follows:

A. Henderi, Handayani, and Dewi (2012), in their study titled "Business Intelligence Development Model Using Star Schema Methodology," demonstrate that Business Intelligence systems employing the Star Schema methodology are effective in evaluating institutional performance based on Key Performance Indicators (KPIs). Their implementation, focused on new student admissions at Perguruan Tinggi Raharja, illustrates how information technology can enhance organizational output in terms of product and service quality. The admissions process served as a prototype to support managerial evaluation and performance measurement aligned with institutional goals, mission, and vision.

B. Kurniawati, Indrajit, and Fauzi (2017), in "The Role of Business Intelligence in Determining New Student Admission Promotion Strategies," examine the annual admissions process at Institut Sains dan Teknologi Al-Kamal (ISTA). Despite successful candidate selection, a significant proportion of students failed to re-register, resulting in unmet enrollment targets. The study concludes that Business Intelligence can serve as a strategic tool for decision-making in promotional planning to enhance student recruitment outcomes.

C. Alviana and Kurniawan (2019), in their work "Data Analysis of New Student Admissions to Increase University Marketing Potential Using Business Intelligence (XYZ University Case Study)," reveal that Business Intelligence facilitates effective decision-making in university marketing. By analyzing data such as students' city and province of origin, preferred majors, and faculties, institutions can optimize their marketing strategies.

D. Amborowati and Suyanto (2015), through their literature-based study "Marketing Intelligence Support Study on Marketing Strategy," synthesize findings from eight research papers on data mining applications in consumer segmentation. Their analysis underscores the role of Marketing Intelligence in enhancing marketing strategies through technological support, enabling precise market and consumer segmentation for more effective marketing execution.

E. Geria (2018), in "Social Media as Promotion Trend for Increasing Tourist Visit towards the Digital Era in Tegallantang Village, Ubud, Bali in 2017," explores the use of social media as a promotional tool to boost tourism. The study affirms that Marketing Intelligence empowers managers with real-time insights into the marketing environment, thereby facilitating informed planning and strategic development.

F. Purnamaningsih, Saptono, and Aziz (2016), in "Utilization of the K-Means Clustering Method in Determining Majoring of High School Students," propose a clustering-based recommendation system for student major selection. By analyzing academic performance data, the K-Means Clustering method effectively categorizes students into science or social science tracks. The Confusion Matrix is employed to assess the accuracy of the classification, enhancing the reliability of the recommendations.

G. Rarasati (2020), in "A Grouping of Song-Lyric Themes Using K-Means Clustering at Radio Masdha Radio Yogyakarta," develops an automated system for categorizing song lyrics. The system successfully generates four thematic clusters—love, friendship, religion, and struggle—based on 400 data points, achieving an accuracy rate of 93.25% using frequency analysis of unique words.

H. Dinata et al. (2020), in "K-Means Clustering Analysis on Motorcycle Data," address consumer preferences in Dewantara Regency, Aceh. Their clustering model provides actionable recommendations for motorcycle selection, achieving precision and recall rates of 76%, and an overall accuracy of 81%.

I. Suprawoto (2016), in "Student Data Classification Using the K-Means Method to Support the Selection of Marketing Strategies," utilizes academic and administrative data from STMIK AKAKOM Yogyakarta (2014/2015) to inform marketing strategy. The study finds that students with lower national exam scores tend to have weaker academic performance, and that students from the Special Region of Yogyakarta and Central Java exhibit higher average GPAs.

J. SP, BT Sutrisno, and Andriyani (2020), in "Application of MADM with the SAW Method to Determine Promotional Targets Based on the Origin of Majors in Schools," present a decision-support system for promotional targeting. By analyzing data on students' geographic origin, previous majors, and sources of admission information, the system enables policymakers at the Faculty of Health, Universitas Jenderal Achmad Yani Yogyakarta, to optimize promotional strategies and increase enrollment.

III. RESEARCH METHOD

A. *Data*

The data that became the object of this research was obtained from the Unjaya KPP Bureau according to the research permit issued by the Unjaya Research and Community Service Institute (Lembaga Penelitian dan Pengabdian kepada Masyarakat - LPPM) with the reference of B/085/LPPM/UNJAYA/V/2021 a number of 2013 records.

B. *Research methodology*

The methodology used in this study uses the K-Means Clustering and Confusion Matrix methods with the following system design:

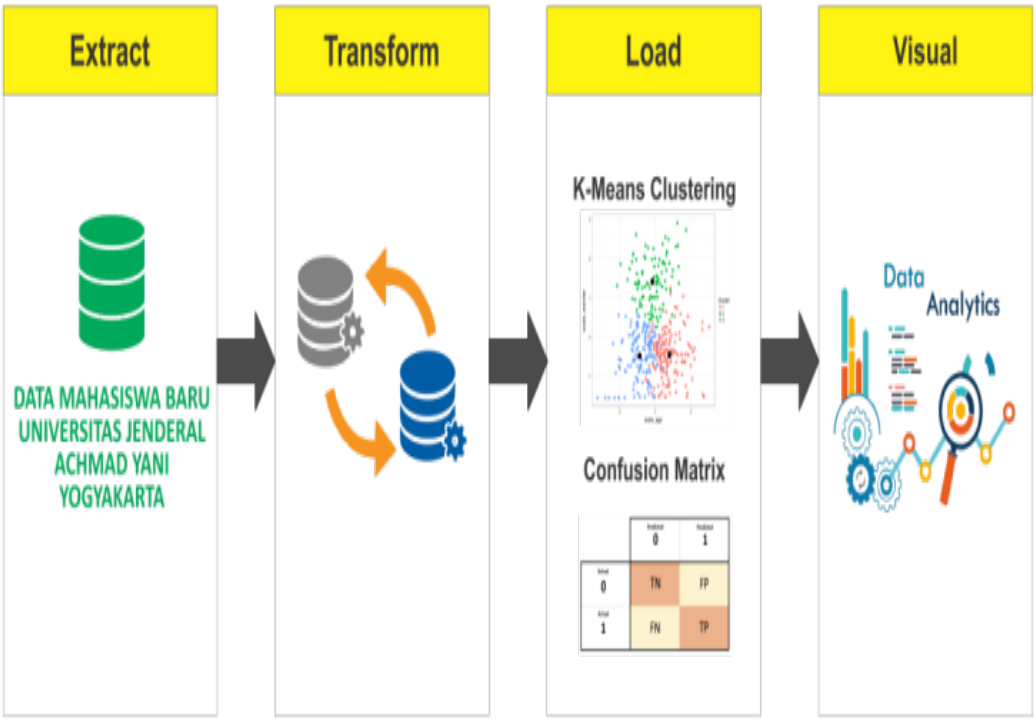

Figure 1. System Design

C. *Extract Process*

It is a process of collecting data, information, and other knowledge-based on data from Unjaya's new students from 2014 to 2018. In this process, the data obtained from the Unjaya KPP Bureau is in the form of a Microsoft Excel file.

D. *Transform Process*

It is a data preparation process by tidying up data according to needs which will later be processed by the system which includes data from the area of origin of prospective students, majors when in SMA/SMK/MA, and media sources that become references for prospective students to find out PMB Unjaya information. At this stage, 2 processes are carried out, namely:

- The process of preparing data by importing Unjaya's new student data files from 2014 to 2018, then determining the data fields that will be used by the system which includes data from the prospective student's area of origin, majors while in SMA/SMK/MA, and media sources that become the reference for prospective students to find out information about PMB Unjaya.
- The process of cleaning data or data cleansing by selecting, correcting and or eliminating data that is not used, after which the data used is combined in the form of a new dataset for later analysis.

E. *Load Process*

The load process is a process that serves to enter data into the methods used to analyze the data so as to obtain the required information. This process is carried out in 3 stages with different methods, namely:

- The process of grouping data using the *K-Means Clustering method* with the aim of grouping based on data similarity and minimizing variations between data in a cluster. Thus *K-Means Clustering* will produce groups that have similar attributes. The K-Means grouping algorithm will produce K groups of records. The *K-Means Clustering* formula used is as follows:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (1)$$

Where:

- De is Euclidean Distance
- i is the number of objects,
- (x,y) are object coordinates
- (s,t) are the coordinates of the centroid

- The process of testing the results of data clusters. In this process using the *Confusion Matrix method* is used to calculate the values of accuracy, precision, and recall. The three evaluation methods are very useful for measuring the performance of the cluster results or the *Machine Learning algorithm* that we use, which will produce accuracy, precision, and recall values. Because there are more than two classified data *(multi-class)*, the Confusion Matrix is used with the following formula:

$$Akurasi = \frac{\sum_{i=1}^{I} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{I} * 100\% \quad (2)$$

$$Presisi = \frac{\sum_{i=1}^{I} TP_i}{\sum_{i=1}^{I} (FP_i + TP_i)} * 100\% \quad (3)$$

$$Recall = \frac{\sum_{i=1}^{I} TP_i}{\sum_{i=1}^{I} (TP_i + FN_i)} * 100\% \quad (4)$$

Where:

- TPi is *True Positive,* ie the number of positive data that is classified correctly by the system for class i.
- TNi is *True Negative* , which is the number of negative data that is classified correctly by the system for class i.
- FNi is *False Negative* , that is, the number of negative data but classified incorrectly by the system for class i.
- FPi is *False Positive* , i.e. the number of positive data but classified incorrectly by the system for class i
- l is the number of classes.

- Visualization process is the final process of displaying the results of data processing to be presented to PMB Unjaya policy makers so that it can be used in making decisions related to promotion strategies. In this process, the Public Tableau platform is used to visualize the data online. Tableau was chosen because it has many advantages in transformation and visualization in *Business Intelligence processes*. With Tableau we can present data dynamically and it is easy for stakeholders to understand the data presented.

IV. RESEARCH RESULTS AND DISCUSSION

A. Data Setup

The data obtained from the Unjaya KPP Bureau is in the form of a Microsoft excel file which is divided into *sheets* per year from 2014 to 2018. For research purposes, the data cleaning process is then carried out by selecting, correcting and/or eliminating unused data. The data *fields* that will be used in this research are the data *field* from the area of prospective students, the *field* majors when in SMA/SMK/MA, and the media source *field* that is the reference for prospective students to find out information about PMB Unjaya with the following data display:

No	Nama Lengkap	Prodi	Asal Daerah	Jurusan	Asal Info
1	DITA NUR INDAHSARI	KEBIDANAN (D-3)	BALI	IPS	DATANG KE KAMPUS
2	DITA DWI ANGGITA	KEBIDANAN (D-3)	BANTEN	IPA	DATANG KE KAMPUS
3	ILHAM CAHYO NUGROHO	KEPERAWATAN (S-1)	BANTEN	IPA	LAINNYA
4	MUTIA TSANY KHAIRUNNISA	KEBIDANAN (D-3)	BANTEN	IPA	DATANG KE KAMPUS
5	BEMI ADRIAN HANAFIAH	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	BENGKULU	IPA	DATANG KE KAMPUS
6	HERIAN ALPAZARA	KEPERAWATAN (S-1)	BENGKULU	IPA	WEBSITE
7	YEKA SAHFITRI	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	BENGKULU	IPA	WEBSITE
8	ERLANGGA MANDALA SAKTI	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	BENGKULU	IPA	WEBSITE
9	M.HASAN BISRI	KEPERAWATAN (S-1)	BENGKULU	IPS	DATANG KE KAMPUS
10	RIZKI DARMAYITASARI	KEBIDANAN (D-3)	BENGKULU	IPA	DATANG KE KAMPUS
11	RENANTHA SILVI JEFRIANY	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	DI YOGYAKARTA	IPA	BROSUR
12	NUR FATHONAH FAJARRINI	KEBIDANAN (D-3)	DI YOGYAKARTA	IPA	BROSUR
13	FATMA ARDIYAN SARI	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	DI YOGYAKARTA	IPS	DATANG KE KAMPUS
14	DEDAH UMILAH MASWATU	KEBIDANAN (D-3)	DI YOGYAKARTA	IPA	DATANG KE KAMPUS
15	RIDWAN DANNY SAPUTRA	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	DI YOGYAKARTA	IPS	DATANG KE KAMPUS
16	NUR ALIFAH MEIRA HARIANTI	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	DI YOGYAKARTA	IPA	DATANG KE KAMPUS
17	IRVAN SAPUTRA	KEPERAWATAN (S-1)	DI YOGYAKARTA	IPA	DATANG KE KAMPUS
18	INKE AYOMI NUR HAPZAH	KEPERAWATAN (S-1)	DI YOGYAKARTA	SMK KES	DATANG KE KAMPUS
19	TYAS YESIANA DWI SAPUTRI	KEBIDANAN (D-3)	DI YOGYAKARTA	SMK KES	DATANG KE KAMPUS
20	SISKA ANIS KHARUNIA	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	DI YOGYAKARTA	SMK KES	DATANG KE KAMPUS
…	…	…	…	…	…
2318	DINDA ZULIA HARIANI	REKAM MEDIS DAN INFORMASI KESEHATAN (D-3)	SUMATERA UTARA	IPA	DATANG KE KAMPUS

Figure 2. Unjaya New Student Data After Cleaning

After cleaning, the amount of data that is ready to be used is 2318 data records, all of which are still in the form of text so that they cannot be used in the calculation of *K-Means Clustering.* Then the transformation process from text data to numeric data is carried out based on the following conditions:

- Province and Zoning Code

The Provincial Code follows the Regulation of the Minister of Home Affairs of the Republic of Indonesia Number 72 of 2019 concerning Amendments to the Regulation of the Minister of Home Affairs Number 137 of 2017 concerning the Code and Data of Government Administration Areas, while the zoning and target options are based on the Decree of the Chancellor of Unjaya number: Skep/088/UNJANI/ XI/2018 concerning Determination of the Amount of Services in Promotional Activities at Jenderal Achmad Yani University Yogyakarta.

- Study Program Code
  Study Program numbering refers to the applicable provisions in the Unjaya Student Prospective Information System (Sicama).
- Code of Origin of Major in SMA/SMK/MA
  The numbering of the code of origin of the majors in SMA/SMK/MA new students refers to the provisions that apply in the Sicama Unjaya application.
- Type of Origin of New Student Information
  Classification of the type of origin of new student information refers to the applicable provisions in the Sicama Unjaya application.
- Unjaya Promotion Target Options
  The classification of promotion target options is determined by the Unjaya KPP Bureau as a reference in implementing operational activities in the field.

After all the transformation processes from text data to numbers are carried out, where the study program data, regional origin, school department origin, and new student reference information have changed from text data to numbers so that they can be processed with *K-Means Clustering calculations*. The final form of the Unjaya new student dataset is shown in table 1 as follows:

Table 1. Unjaya New Student Dataset

No	Nama Lengkap	Prodi	Asal Daerah	Asal Jurusan	Asal Info	Pilihan Opsi
1	Dita Nur Indahsari	3	4	2	1	Opsi 65
2	Dita Dwi Anggita	3	2	1	1	Opsi 21
3	Ilham Cahyo Nugroho	1	2	1	4	Opsi 24
4	Mutia Tsany Khairunnisa	3	2	1	1	Opsi 21
5	Bemi Adrian Hanafiah	4	2	1	1	Opsi 21
6	Herian Alpazara	1	2	1	3	Opsi 23
7	Yeka Sahfitri	4	2	1	3	Opsi 23
8	Erlangga Mandala Sakti	4	2	1	3	Opsi 23
9	M.Hasan Bisri	1	2	2	1	Opsi 25
10	Rizki Darmayitasari	3	2	1	1	Opsi 21
11	Renantha Silvi Jefriany	4	1	1	3	Opsi 3
12	Nur Fathonah Fajarrini	3	1	1	3	Opsi 3
13	Fatma Ardiyan Sari	4	1	2	1	Opsi 5
14	Dedah Umilah Maswatu	3	1	1	1	Opsi 1
15	Ridwan Danny Saputra	4	1	2	1	Opsi 5
16	Nur Alifah Meira Harianti	4	1	1	1	Opsi 1
17	Irvan Saputra	1	1	1	1	Opsi 1
18	Inke Ayomi Nur Hapzah	1	1	3	1	Opsi 9
19	Tyas Yesiana Dwi Saputri	3	1	3	1	Opsi 9
20	Siska Anis Kharunia	4	1	3	1	Opsi 9
...	...	...	...	...	...	...
2318	Dinda Zulia Hariani	4	4	1	1	Opsi 61

B. *K-Means Clustering Flowchart Implementation*

K-Means is one of the methods in data mining that can group data or *Clustering* a data into one or more clusters so that data with the same characteristics are grouped into the same cluster and data with different characteristics are grouped into different groups (Dhuhita, 2015) . To implement the *K-Means Clustering calculation* on Unjaya's new student data, a flow chart is made as follows:

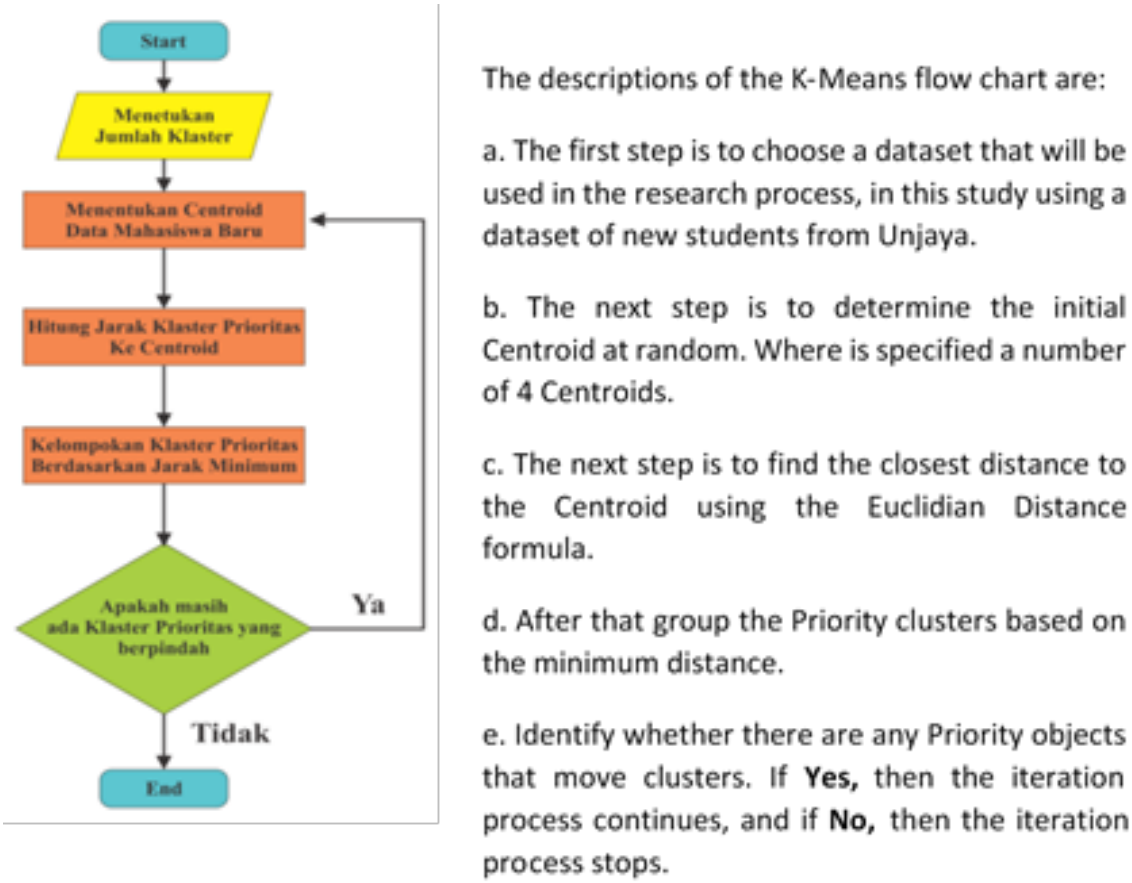

Figure 3. K-Means Process Diagram for New

C. *Implementation of K-Means Clustering Calculation*

To start the calculation of *K-Means Clustering* on the dataset of new Unjaya students, it begins by determining the centroid or the initial arithmetic mean that is assigned randomly. In this calculation, the initial centroid values used are the 10th, 20th, 425th, and 1100th data as follows:

Table 2. Initial Centroid of New Student Dataset

No	Nama Lengkap	Prodi	Asal Daerah	Asal Jurusan	Asal Info	Pilihan Opsi
10	Rizki Darmayitasari	3	2	1	1	Opsi 21
20	Siska Anis Kharunia	4	1	3	1	Opsi 9
425	Nani Fitriyanti Redjeb	3	4	3	1	Opsi 69
1100	Avis Muzdalifah	1	1	1	3	Opsi 3

The calculation in the first test is to find clusters 1 to 4 using the *Euclidean Distance K-Means Clustering formula* as follows:

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (5)$$

Cluster 1 $=\sqrt{(3-3)^2 + (4-2)^2 + (2-1)^2 + (1-1)^2}$
$=\sqrt{(0)^2 + (2)^2 + (1)^2 + (0)^2}$
$=2{,}24$

Cluster 2 $=\sqrt{(3-4)^2 + (4-1)^2 + (2-3)^2 + (1-1)^2}$
$=\sqrt{(-1)^2 + (3)^2 + (-1)^2 + (0)^2}$
$=3{,}32$

Cluster 3 $=\sqrt{(3-3)^2 + (4-4)^2 + (2-3)^2 + (1-1)^2}$
$=\sqrt{(0)^2 + (0)^2 + (-1)^2 + (0)^2}$
$=1$

Cluster 4 $=\sqrt{(3-1)^2 + (4-1)^2 + (2-1)^2 + (1-3)^2}$
$=\sqrt{(2)^2 + (3)^2 + (1)^2 + (-2)^2}$
$=4{,}24$

The equation with the same calculation is applied to 2318 data to get the distance of each data in each cluster so as to produce a new cluster as follows:

Table 3. Distance in Each Cluster First Calculation

No	Full name	K1	K2	K 3	K 4	Jrk	Cluster End
1	Dita Nur Indahsari	2.24	3.32	1.00	4.24	1.00	Cluster 3
2	Dita Dwi Anggita	0.00	2.45	2.83	3.00	0.00	Cluster 1
3	Ilham Cahyo Nugroho	3.61	4.80	4.58	1.41	1.41	Cluster 4
4	Mutia Tsany Khairunnisa	0.00	2.45	2.83	3.00	0.00	Cluster 1
5	Bemi Adrian Hanafiah	1.00	2.24	3.00	3.74	1.00	Cluster 1
6	Herian Alpazara	2.83	4.24	4.00	1.00	1.00	Cluster 4
7	Yeka Sahfitri	2.24	3.00	3.61	3.16	2.24	Cluster 1
8	Erlangga Mandala Sakti	2.24	3.00	3.61	3.16	2.24	Cluster 1
9	M. Hasan Bisri	2.24	3.32	3.00	2.45	2.24	Cluster 1

10	Rizki Darmayitasari	0.00	2.45	2.83	3.00	0.00	Cluster 1
11	Renantha Silvi Jefriany	2.45	2.83	4.24	3.00	2.45	Cluster 1
12	Nur Fathonah Fajarrini	2.24	3.00	4.12	2.00	2.00	Cluster 4
13	Fatma Ardiyan Sari	1.73	1.00	3.32	3.74	1.00	Cluster 2
14	Umilah Maswatu	1.00	2.24	3.61	2.83	1.00	Cluster 1
15	Ridwan Danny Saputra	1.73	1.00	3.32	3.74	1.00	Cluster 2
16	Nur Alifah Meira Harianti	1.41	2.00	3.74	3.61	1.41	Cluster 1
17	Irvan Saputra	2.24	3.61	4.12	2.00	2.00	Cluster 4
18	Inke Ayomi Nur Hapzah	3.00	3.00	3.61	2.83	2.83	Cluster 4
19	Tyas Yesiana Dwi Saputri	2.24	1.00	3.00	3.46	1.00	Cluster 2
20	Siska Anis Kharunia	2.45	0.00	3.16	4.12	0.00	Cluster 2
…	…	…	…	…	…	…	…
2318	Dinda Zulia Hariani	2.24	3.61	2.24	4.69	2.24	Cluster 1

Table 4. Number of First Calculation Clusters

Cluster	Number of Data Members
Cluster 1	873
Cluster 2	487
Cluster 3	168
Cluster 4	790

After getting the value for each data cluster, then looking for the average value by adding up all the members of each cluster which is then divided by the number of members. The new centroids formed are as follows:

Table 5. Second Iteration New Centroid Table

Cluster	Study Program	Origin	Origin of Department	Origin of Info
Cluster 1	3.27	1.92	1.26	1.87
Cluster 2	3.49	1.15	3.19	2.16
Cluster 3	2.44	3.91	2.90	2.30
Cluster 4	1.58	1.37	1.84	2.78

Furthermore, calculations are carried out in the second to ninth iteration tests to find Clusters 1 to 4. In the ninth iteration the results are as follows:

Cluster 1 =

$$\sqrt{(3-3{,}56)^2+(4-1{,}39)^2+(2-1{,}34)^2+(1-2{,}29)^2}$$

$$=\sqrt{(-0{,}56)^2+(2{,}61)^2+(0{,}66)^2+(-1{,}29)^2}$$

$$=3{,}04$$

Cluster 2 =

$$\sqrt{(3-3{,}49)^2+(4-1{,}34)^2+(2-3{,}67)^2+(1-2{,}26)^2}$$
$$=\sqrt{(-0{,}49)^2+(2{,}66)^2+(-1{,}67)^2+(-1{,}26)^2}$$
$$=3{,}42$$

Cluster 3 =

$$\sqrt{(3-2{,}69)^2+(4-3{,}95)^2+(2-1{,}75)^2+(1-2{,}31)^2}$$
$$=\sqrt{(0{,}31)^2+(0{,}05)^2+(0{,}25)^2+(-1{,}31)^2}$$
$$=1{,}37$$

Cluster 4 =

$$\sqrt{(3-1{,}19)^2+(4-1{,}42)^2+(2-2{,}07)^2+(1-2{,}25)^2}$$
$$=\sqrt{(1{,}81)^2+(2{,}58)^2+(-0{,}07)^2+(-1{,}25)^2}$$
$$=3{,}38$$

The equation with the same calculation is applied to 2318 data to get the distance of each data in each cluster resulting in the number of clusters and the new centroid value as follows:

Table 6. Distance in Each Ninth Calculation Cluster

No	Full name	K1	K2	K3	K4	Jrk	Cluster Prediction
1	Dita Nur Indahsari	3.04	3.42	1.37	3.38	1.37	Cluster 3
2	Dita Dwi Anggita	1.57	3.06	2.49	2.51	1.57	Cluster 1
3	Ilham Cahyo Nugroho	3.16	4.09	3.18	2.14	2.14	Cluster 4
4	Mutia Tsany Khairunnisa	1.57	3.06	2.49	2.51	1.57	Cluster 1
5	Bemi Adrian Hanafiah	1.53	3.06	2.79	3.30	1.53	Cluster 1
6	Herian Alpazara	2.75	3.78	2.78	1.44	1.44	Cluster 4
7	Yeka Sahfitri	1.09	2.89	2.56	3.15	1.09	Cluster 1
8	Erlangga Mandala Sakti	1.09	2.89	2.56	3.15	1.09	Cluster 1
9	M. Hasan Bisri	3.00	3.32	2.91	1.39	1.39	Cluster 4
10	Rizki Darmayitasari	1.57	3.06	2.49	2.51	1.57	Cluster 1
11	Renantha Silvi Jefriany	0.98	2.83	3.38	3.12	0.98	Cluster 1
12	Nur Fathonah Fajarrini	1.04	2.83	3.14	2.27	1.04	Cluster 1
13	Fatma Ardiyan Sari	1.56	2.18	3.49	3.10	1.56	Cluster 1
14	Umilah Maswatu	1.50	3.01	3.33	2.48	1.50	Cluster 1
15	Ridwan Danny Saputra	1.56	2.18	3.49	3.10	1.56	Cluster 1
16	Nur Alifah Meira Harianti	1.46	3.01	3.56	3.28	1.46	Cluster 1
17	Irvan Saputra	2.91	3.87	3.72	1.71	1.71	Cluster 4
18	Inke Ayomi Nur Hapzah	3.33	2.89	3.86	1.62	1.62	Cluster 4
19	Tyas Yesiana Dwi Saputri	2.21	1.55	3.48	2.42	1.55	Cluster 2

20	Siska Anis Kharunia	2.18	1.55	3.70	3.24	1.55	Cluster 2
…	…						
2318	Dinda Zulia Hariani	2.96	4.00	2.00	4.15	2.00	Cluster 3

Table 7. Number of Ninth Calculation Clusters

Cluster	Number of Data Members
Cluster 1	932
Cluster 2	353
Cluster 3	291
Cluster 4	742

The iteration stops after there is no change in the number of data cluster members, and produces the data set for each cluster as follows:

Table 8. Table of Cluster Member Association

Group	Data Member
Cluster 1	2,4,5,7,8,10,11,12,13,14,15,16,25,26,27,29,32,33,34,36,37,38,39,44,45, 48,49,54,57,59,62,65,74,75,76,80,83,84,86,87,88,90,92,93,95,96,99,100,101,1 03, 104,105,106,109,116,117,119,120,121,124,125,126,127,129,130,131,132,134, 136,141,142,143,144,149,155, 156,157,159,160,162,163,164,169,172,173,174,177,178,179,180,182,184,186, 189,190,191,192,193,195,196, 197,199,202,203,206,211,213,217,218,219,222,223,224,228,229,230,232,234, 235,236,240,242,243,244,245, 246,249,251,255,256,257,258,259,262,263,269,270,273,276,277,279,280,281, 282,283,287,288,289,290,291,292,293,297,298,299,305,306,307,308,309,312, 314,317,318,319,320,323,324,327,329,330,331,332,333,334,336,338,341,343, 347,348,352,354,355,356,357,358,359,364,365,371,373,381,383,386,387,388, 389,393,394,395,396,397,398,399,403,404,409,412,413,433,437,438,439,442, 443,444,445,446,450,451,452,453,456,459,460,461,462,463,466,468,471,476, 479,481,486,511,513,514,515,516,519,522,523,524,525,531,532,533,534,538, 539,540,543,544,545,549,554,555,563,564,565,566, 567,568,569,578,582,584,586,587,590,593,600,601,605,606,607,608,610,611, 612,613,614,615,616,617,618, 619,620,621,622,623,624,626,627,628,629,630,632,633,634,636,638,639,640, 641,643,644,646,647,653,655, 657,658,659,660,661,665,667,668,674,677,678,679,680,684,687,688,689,696, 699,705,706,707,708,716,717,718,719,720,721,728,729,730,731,732,738,739, 741,742,745,746,747,748,749,750,751,754,755,756,758,759,760,761,763,764, 765,769,770,772,776,778,785,787,788,789,790,792,793,794,795,796,797,798, 800,801,802,803,805,806,811,817,822,824,840,842,843,849,850,851,856,857, 858,859,863,864,866,873,897,898,900,901,903,904,905,907,908,913,916,917, 918,922,955,956,957,958,959,962,963,964,965,966,971,972,974,976,978,979, 980,984,986,987,989,992,995,996,1000,1002,1004,1008,1010,1016,1017,101 8,1019,1021,1023,1024,1031,1033,1035,1038,1042,1045,1046,1047, 1050,1051,1052,1053,1055,1057,1058,1059,1061,1062,1064,1066,1067,1069, 1070,1072,1074,1076,1077,1078,1085,1086,1089,1091,1093, 1098,1102,1103,1104,1105,1107,1109,1111 ,1112,1116,1117,1121,1122,1123,1126,1130,1137,1139,1140,1141,1144,1145 ,1150,1152,1153,1157,1159,1160,1162,1164,1166,1167,1168 ,1169,1170,1171,1172,1173,1178,1179,1180,1183,1188,1194,1198,1202,1205 ,1206,1210,1211,1213,1215,1216,1217,1218,1219,1220,1227

	,1228,1230,1232,1233,1234,1235,1246,1247,1250,1262,1268,1285,1286,1287 ,1288,1289,1290,1291,1292,1294,1295,1297,1304,1305,1307 ,1308,1310,1312,1314,1316,1317,1318,1319,1321,1322,1326,1327,1333,1334 ,1335,1342,1343,1344,1348,1350,1352,1356,1359,1362,1366 ,1367,1369,1370,1371,1372,1373,1381,1383,1389,1390,1391,1392,1393,1394 ,13 95,1398,1399,1401,1402,1403,1405,1408,1417,1420, 1422,1423,1424,1426,1427,1433,1437,1438,1450,1451,1454,1455,1456,1458, 1466,1467,1469,1470,1472,1474,1476,1478,1482,1483,1484, 1486,1492,1494,1495,1496,1502,1503,1505,1506,1507,1508,1509,1511,1523, 1524,1526,1527,1528,1533,1537,1540,1541,1545,1547,1550, 1557,1558,1559,1562,1563,1566,1567,1569,1570,1571,1572,1573,1575,1576, 1579,1580,1582,1584,1589,1592,1593,1595,1596,1597,1598, 159 9,1600,1601,1602,1603,1614,1615,1619,1620,1624,1638,1640,1647,1653,165 6,1659,1660,1662,1666,1667,1668,1672,1673,1676,1684, 1686,1687,1688,1689,1690,1691,1711,1714,1715,1718,1719,1722,1723,1729, 1732,1734,1735,1770,1771,1773,1774,1775,1777,1778,1779, 1782,1783,1785,1786,1787,1794,1797,1799,1800,1803,1804,1805,1813,1814, 1815,1816,1817,1820,1821,1823,1829,1838,1840,1844,1845, 1846,1847,1854,1855,1856,1861,1862,1865,1869,1872,1873,1874,1875,1877, 1882,1883,1886,1887,1888,1889,1892,1893,1904,1906,1910, 1911,1912,1915,1918,1920,1921,1922,1925,1927,1930,1937,1938,1939,1940, 1944,1945,1946,1947,1948,1954,1957,1959,1960,1966,1968, 1971,1979,1980,1990,2000,2001,2002,2006,2007,2011,2016,2019,2020,2022, 2023,2026,2027,2028,2034,2037,2040,2041,2043,2044,2045, 2048,2049,2051,2052,2053,2054,2056,2057,2064,2065,2066,2067,2068,2070, 2075,2076,2077,2080,2081,2083,2085,2086,2089,2091,2097, 2103,2106,2107,2112,2115,2119,2120,2128,2130,2131,2132,2134,2135,2138, 2139,2140,2141,2142,2143,2144,2145,2156,2159,2162,2163, 216 7,2174,2175,2179,2180,2181,2202,2208,2213,2224,2227,2240,2241,2242,224 5,2246,2247,2248,2249,2250,2252,2253,2255,2298,2300, 2301,2305,2306,2307,2308,2311,2312,2314
Cluster 2	19,20,22,23,41,46,52,53,58,60,61,69,70,71,82,85,89,97,111,112,113,114,115, 138,139, 147,148,150,152,153,154,165,166,170,194,205,207,208,231,239,247,248,254, 264,265,271,274,284,285,294,295,302,310,311,316,322,350,374,375,384,401, 406,407,408,410,411,440,441,448,455,472,477,478,487,488,492,510,526,527, 541,547,548,550,552,556,559,560,571,572,576,577,579,580,583,585,588,592, 595,602,603,609,635,637,642,649,650,664,675,676,681,683,685,690,693,697, 701,702,703,704,709,710,711,712,713,714,715,722,724,725,727,734,735,736, 737,743,753,757,768,774,777,780,812,816,819,854,855,865,867,868,914,920, 921,928,929,931,975,985,990,991,993,994,997,998,999,1005,1006,1012,1013 ,1014,1015,1028,1032,1034,1036,1037,1039,1040,1041,1048, 1056,1060,1079,1081,1090,1094,1096,1108,1113,1115,1120,1129,1133,1136, 1143,1148,1151,1154,1163,1165,1174,1176,1185,1190,1200, 1229,1245,1249,1256,1257,1263,1303,1332,1337,1338,1346,1357,1358,1375, 1376,1377,1378,1380,1385,1386,1387,1397,1409,1410,1411, 1412,1413,1416,1419,1425,1435,1436,14 43,1444,1445,1452,1461,1471,1473,1479,1481.1 491,1504,1512,1515,1529,1535,1536,1548,1552,1556,1586,1587,1611,1616,1 621,1639

	,1644,1648,1657,1663,1669,1670,1674,1679,1683,1694,1709,1713,1716,1725 ,1793,1812,1826,1830,1832,1833,1834,1836,1837,1850,1853 ,1859,1868,1870,1876,1890,1891,1898,1905,1909,1917,1924,1931,1932,1933 ,1934,1950,1961,1972,1974,1977,1986,1987,1991,2003,2004 ,2005,2008,2014,2017,2024,2030,2031,2046,2050,2060,2072,2088,2093,2094 ,2101,2109,2111,2113,2125,2126,2147,2160,2166,2173,2193 ,2194,2195,2197,2198,2201,2205,2214,2217,2218,2251,2278,2293,2299,2302 ,2303,2309,2310
Cluster 3	1,363,370,372,379,391,414,415,416,417,418,419,420,421,422,423,424,425,42 6,427,428,429,430,431,432,483,484,485,490,491,493,494,495,496,497,498,49 9,500,501,502,503,504,505,506,507,508,509,528,529,530,813,814,815,820,82 3,825,826,827,828,829,830,831,832,833,834,835,836,837,844,874,875,876,87 7,878,879,880,881,882,883,884,885,886,887,888,889,890,891,892,893,894,89 5,896,924,925,926,927,930,932,933,934,935,936,937,938,939,940,941,942,94 3,944,945,946,947,948,949,950,968,969,970,1191,1192,1195,1196,1197,1203 ,1236,1237,1238,1239,1240,1241,1242,1243,1244,1265,1269,1270,1271,1272 ,1273,1274,1275,1276, 1277,1278,1279,1280,1281,1282,1283,1299,1300,1301,1302,1428,1429,1604, 1605,1606,1608,1610,1612,1613,1617,1618,1622,1623,1629, 1630,1631,1632,1633,1641,1695,1696,1697,1698,1699,1700,1701,1702,1703, 1704,1705,1706,1707,1728,1730,1733,1736,1737,1738,1739, 1740,1741,1742,1743,1744,1745,1746,1747,1748,1749,1750,1751,1752,1753, 1755,1756,1757,1758,1759,1760,1761,1762,1763,1764,1765, 1766,1767,1768,1769,1808,1809,1 810,1964,1965,2148,2149,2150,2152,2153,2154,2155,2157,2168,2169,2170,2 171,2172,2228,2229,2230,2231,2232,2233,2234,2235,2236, 2237,2238,2239,2257,2258,2259,2260,2261,2262,2263,2264,2265,2266,2267, 2268,2269,2270,2271,2272,2273,2274,2275,2276,2279,2280, 2281,2282,2283,2284,2285,2286,2287,2288,2289,2290,2291,2292,2294,2295, 2296,2297,2318
Cluster 4	3,6,9,17,18,21,24,28,30,31,35,40,42,43,47,50,51,55,56,63,64,66,67,68,72, 73,77,78,79,81,91,94,98,102,107,108,110,118,122,123,128,133,135,137,140,1 45,146,151,158,161, 167,168,171,175,176,181,183,185,187,188,198,200,201,204,209,210,212,214, 215,216,220,221,225,226,227, 233,237,238,241,250,252,253,260,261,266,267,268,272,275,278,286,296,300, 301,303,304,313,315,321,325,326,328,335,337,339,340,342,344,345,346,349, 351,353,360,361,362,366,367,368,369,376,377,378,380,382,385,390,392,400, 402,405,434,435,436,447,449,454,457,458,464,465,467,469,470,473,474,475, 480,482,489,512,517,518,520,521,535,536,537,542,546,551,553,557,558,561, 562,570,573,574,575,581,589,591,594,596,597,598,599,604,625,631,645,648, 651,652,654,656,662,663,666,669,670,671,672,673,682,686,691,692,694,695, 698,700,723,726,733,740,744, 752,762,766,767,771,773,775,779,781,782,783,784,786,791,799,804,807,808, 809,810,818,821,838,839,841,845,846,847,848,852,853,860,861,862,869,870, 871,872,899,902,906,909,910,911,912,915,919,923,951,952,953 ,954,960,961,967,973,977,981,982,983,988,1001,1003,1007,1009,1011,1020, 1022,1025,1026,1027,1029,1030,1043,1044,1049,1054,1063,1065,1068,1071, 1073,1075,1080,1082 ,1083,1084,1087,1088,1092,1095,1097,1099,1100,1101,1106,1110,1114,1118

	,1119,1124,1125,1127,1128,1131,1132,1134,1135,1138,1142 ,1146,1147,1149,1155,1156,1158,1161,1175,1177,1181,1182,1184,1186,1187 ,1189,1193,1199,1201,1204,1207,1208,1209,1212,1214,1221 ,1222,1223,1224,1225,1226,1231,1248,1251,1252,1253,1254,1255,1258,1259 ,1260,1261,1264,1266,1267,1284,1293,1296,1298,1306,1309 ,1311,1313,1315,1320,1323,1324,1325,1328,1329,1330,1331,1336,1339,1340 ,1341,1345,1347,1349,1351,1353,1354,1355,1360,1361,1363 ,1364,1365,1368,1374,1379,1382,1384,1388,1396,1400,1404,1406,1407,1414 ,1415,1418,1421,1430,1431,1432,1434,1439,1440,1441,1442 ,1446,1447,1448,1449,1453,1457,1459,1460,1462,1463,1464,1465,1468,1475 ,1477,1480,1485,1487,1488,1489,1490,1493,1497,1498,1499 ,1500,1501,1510,1513,1514,1516,1517,1518,1519,1520,1521,1522,1525,1530 ,1531,1532,1534,1538 ,1539,1542,1543,1544,1546,1549,1551,1553,1554,1555,1560,1561,1564,1565 ,1568,1574,1577,1578,1581,1583,1585,1588,1590,1591,1594 ,1607,1609,1625,1626,1627,1628,1634,1635,1636,1637,1642,1643,1645,1646 ,1649,1650,1651,1652,1654,1655,1658,1661,1664,1665,1671 ,1675,1677,1678,1680,1681,1682,1685,1692,1693,1708,1710,1712,1717,1720 ,1721,1724,1726,1727,1731,1754,1772,1776,1780,1781,1784 ,1788,1789,1790,1791,1792,1795,1796,1798,1801,1802,1806,1807,1811,1818 ,1819,1822,1824,1825,1827,1828,1831,1835,1839,1841,1842 ,1843,1848,1849,1851,1852,1857,1858,1860,1863,1864,1866,1867,1871,1878 ,1879,1880,1881,1884,1885,1894,1895,1896,1897,1899,1900 ,1901,1902,1903,1907,1908,1913,1914,1916,1919,1923,1926,1928,1929,1935 ,1936,1941,1942,1943,1949,1951,1952,1953,1955,1956,1958 ,1962,1963,1967,1969,1970,1973,1975,1976,1978,1981,1982,1983,1984,1985 ,1988,1989,1992,1993,1994,1995,1996,1997,1998,1999,2009 ,2010,2012,2013,2015,2018,2021,2025,2029,2032,2033,2035,2036,2038,2039 ,2042,2047,2055,2058,2059,2061,2062,2063,2069,2071,2073 ,2074,2078,2079,2082,2084,2087,2090,2092,2095,2096,2098,2099,2100,2102 ,2104,2105,2108,2110,2114,2116,2117,2118,2121,2122,2123 ,2124,2127,2129,2133,2136,2137,2146,2151,2158,2161,2164,2165,2176,2177 ,2178,2182,2183,2184,2185,2186,2187,2188,2189,2190,2191 ,2192,2196,2199,2200,2203,2204,2206,2207,2209,2210,2211,2212,2215,2216 ,2219,2220,2221,2222,2223,2225,2226,2243,2244,2254,2256 ,2277,2304,2313,2315,2316,2317

D. Testing Cluster Results with Confusion Matrix

To evaluate the quality of the clustering results generated by the K-Means Clustering algorithm, a validation test was conducted using the Confusion Matrix method. This approach enables the assessment of clustering accuracy by comparing predicted cluster assignments with actual classifications. The actual data were derived from the Unjaya marketing team's historical categorization of promotional targets based on annual patterns and strategic considerations. In contrast, the predicted data represent the cluster assignments produced by the K-Means algorithm. A comparative summary of these two datasets is presented as follows:

Table 9. Table of Actual Clusters and Predicted Clusters of New Students

No	Nama Lengkap	Prodi	Asal Daerah	Asal Jurusan	Asal Info	Klaster Aktual	Klaster Prediksi
1	Dita Nur Indahsari	3	4	2	1	Klaster 3	Klaster 3
2	Dita Dwi Anggita	3	2	1	1	Klaster 1	Klaster 1
3	Ilham Cahyo Nugroho	1	2	1	4	Klaster 4	Klaster 4
4	Mutia Tsany Khairunnisa	3	2	1	1	Klaster 1	Klaster 1
5	Bemi Adrian Hanafiah	4	2	1	1	Klaster 1	Klaster 1
6	Herian Alpazara	1	2	1	3	Klaster 4	Klaster 4
7	Yeka Sahfitri	4	2	1	3	Klaster 1	Klaster 1
8	Erlangga Mandala Sakti	4	2	1	3	Klaster 1	Klaster 1
9	M.Hasan Bisri	1	2	2	1	Klaster 1	Klaster 4
10	Rizki Darmayitasari	3	2	1	1	Klaster 1	Klaster 1
11	Renantha Silvi Jefriany	4	1	1	3	Klaster 1	Klaster 1
12	Nur Fathonah Fajarrini	3	1	1	3	Klaster 4	Klaster 1
13	Fatma Ardiyan Sari	4	1	2	1	Klaster 2	Klaster 1
14	Dedah Umilah Maswatu	3	1	1	1	Klaster 1	Klaster 1
15	Ridwan Danny Saputra	4	1	2	1	Klaster 2	Klaster 1
16	Nur Alifah Meira Harianti	4	1	1	1	Klaster 1	Klaster 1
17	Irvan Saputra	1	1	1	1	Klaster 4	Klaster 4
18	Inke Ayomi Nur Hapzah	1	1	3	1	Klaster 4	Klaster 4
19	Tyas Yesiana Dwi Saputri	3	1	3	1	Klaster 2	Klaster 2
20	Siska Anis Kharunia	4	1	3	1	Klaster 2	Klaster 2
…	…	…	…	…	…	…	…
2318	Dinda Zulia Hariani	4	4	1	1	Klaster 1	Klaster 3

After the actual data and predictive data are known, the *Confusion Matrix calculation is then carried out*. As the number of classes is known to be more than 2, but 4 (clusters 1 to 4), the *Confusion Matrix* used is the ***Confusion Matrix Multi Class*** where the following equation is used to solve it:

$$Akurasi = \frac{\sum_{i=1}^{l} \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} * 100\% \quad (6)$$

$$Presisi = \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l} (FP_i + TP_i)} * 100\% \quad (7)$$

$$Recall = \frac{\sum_{i=1}^{l} TP_i}{\sum_{i=1}^{l} (TP_i + FN_i)} * 100\% \quad (8)$$

Where:

- TPi is *True Positive*, i.e. the number of positive data that is classified correctly by the system for class i.
- TNi is *True Negative*, which is the number of negative data that is classified correctly by the system for class i.
- FNi is *False Negative*, that is, the number of negative data but classified incorrectly by the system for class i.
- FPi is *False Positive*, i.e. the number of positive data but classified incorrectly by the system for class i
- l is the number of classes.

Based on the comparison of the actual data and the predicted data for the Unjaya new student cluster data, the *Confusion Matrix Multi Class* 4x4 table was obtained as follows:

Table 10. *Confusion Matrix Multi Class* 4x4

		PREDICTION			
		Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
AKTUAL	Cluster 1	**636**	0	135	102
	Cluster 2	120	**318**	0	49
	Cluster 3	0	25	**128**	15
	Cluster 4	176	10	28	**576**

Based on the data in the *Confusion Matrix Multi Class* 4x4 table above, we will look for the Accuracy, Precision, and Recall values as follows:

- Value Accuracy

  The accuracy value is the percentage of the total cluster that is predicted to be true and actually true *(True) based on the Confusion Matrix Multi Class* 4 x 4 table with the following calculation:

$$= \frac{(TPi + TNi)}{(TPi + TNi + FPi + FNi) \_\_\_\_\_ l}$$

$$= \frac{(636+318+128+576)}{(636+0+135+102+120+318+0+49+0+25+128+15+176+10+28+576)} \_\_\_\_\_ 4$$

$$= \frac{1658}{2318} \_\_\_\_\_ 4 \qquad = \frac{0,71}{4}$$

From the above calculation, an accuracy value of 0.17 or 17% is obtained which shows the comparison of the number of clusters of new Unjaya students that are classified (predicted) correctly. Seeing the small value of accuracy above, the PMB Unjaya policy makers can evaluate the cluster that has been the target to increase the level of accuracy of the new student promotion target to be implemented.

- Value Precision

The precision value is data that is taken based on information that is lacking or inaccurate. To find the precision value, use the TP/(TP+FP) formula where *the Multi Class Confusion Matrix* is done by finding the FP *(False Positive) value* for each class based on the 4X4 *Multi Class Confusion Matrix table with the following calculations:*

Precision $= \frac{TPi}{TPi+FPi}$

$= \text{FP(Cluster 1)} = \frac{636}{(636+120+0+176)} = \frac{636}{932} = 0.68$

$= \text{FP(Cluster 2)} = \frac{318}{(318+0+25+10)} = \frac{318}{353} = 0.90$

$= \text{FP(Cluster 3)} = \frac{128}{(128+135+0+28)} = \frac{128}{419} = 0.31$

$= \text{FP(Cluster 4)} = \frac{576}{(576+102+49+15)} = \frac{576}{742} = 0.78$

Precision $= \frac{0,68+0,90+0,31+0,78}{4} = \frac{2,67}{4}$

$= 0.67$ or 67%

From the above calculation, the precision value is 0.67 or 67% which shows the comparison of the number of clusters of new Unjaya students who are predicted to be promotion targets and actually produce new students. Thus, PMB Unjaya policymakers can evaluate and or maintain new student clusters that have been proven to be sources of Unjaya's new students.

- Recall Value

A recall is cluster data that cannot be predicted correctly. To find the recall value using the TP/(TP+FN) formula in *the Multi Class Confusion Matrix,* it is done by finding the FN *(False Negative) value* of each class based on the 4X4 *Multi Class Confusion Matrix table with the following calculations:*

Recall $= \frac{TPi}{TPi+FPi}$

$= \text{FN(Cluster 1)} = \frac{636}{(636+0+135+102)} = \frac{636}{873}$

$= 0.73$

$= \text{FN(Cluster 2)} = \frac{318}{(318+120+0+49)} = \frac{318}{805} = 0.40$

$= \text{FN(Cluster 3)} = \frac{128}{(128+0+25+15)} = \frac{128}{168} = 0.76$

$= \text{FN(Cluster 4)} = \frac{576}{(576+176+10+28)} = \frac{576}{790}$

$$= 0.73$$

$$\text{Recall} = \frac{0{,}73+0{,}40+0{,}76+0{,}73}{4} = \frac{2{,}61}{4}$$

$$= 0.65 \text{ or } 65\%$$

From the calculation above, the Recall value is 0.65 or 65% which shows the comparison of the number of new student clusters that are properly classified as the target cluster of promotion targets, so that they can be considered as a promotion target cluster by PMB Unjaya policy makers.

E. Implementation With Python

- Data Cleansing using Python

Upload the data to Google Colabs via the Upload feature, then import the required python library:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.Klaster import KMeans
from sklearn.preprocessing import MinMaxScaler
```

Next, take Unjaya's new student data by using the command:

```
data = pd.read_csv("MABA.csv")
data.head()
```

The initial data is displayed as follows:

Table 11. Display of New Student Initial Data in Python

	NamaLengkap	Prodi	AsalDaerah	Jurusan	AsalInfo	Aktual
0	DITA NUR INDAHSARI	KEBIDANAN	BALI	IPS	DATANG KE KAMPUS	CLUSTER 3
1	DITA DWI ANGGITA	KEBIDANAN	BANTEN	IPA	DATANG KE KAMPUS	CLUSTER 1
2	ILHAM CAHYO NUGROHO	KEPERAWATAN	BANTEN	IPA	LAINNYA	CLUSTER 4
3	MUTIA TSANY KHAIRUNNISA	KEBIDANAN	BANTEN	IPA	DATANG KE KAMPUS	CLUSTER 1
4	BEMI ADRIAN HANAFIAH	REKAM MEDIS DAN INFORMASI KESEHATAN	BENGKULU	IPA	DATANG KE KAMPUS	CLUSTER 1

Then carry out the transformation process from text data to numeric data as follows:

```
# Changing Study Program Data
data[ 'Prodi' ].replace( "MIDOM" , 3 , inplace= True )
data[ 'Prodi' ].replace( "Nursing" , 1 , inplace= True )
data[ 'Prodi' ].replace( "MEDICAL RECORD AND HEALTH INFORMATION" , 4 , inplace= True )
data[ 'Prodi' ].replace( "BLOOD BANK TECHNOLOGY" , 5 , inplace= True )
data[ 'Prodi' ].replace( "PHARMACEUTICAL" , 2 , inplace= True )
# Changing Regional Data
data[ 'OriginDaerah' ].replace([ "IN YOGYAKARTA" , "WEST JAVA" , "CENTRAL JAVA" , "TEST JAVA" ], 1 , inplace= True )
```

```
data[ 'OriginDaerah' ].replace([ "West KALIMANTAN" , "LAMPUNG" , "BANTEN"
, "BENGKULU" , "JAMBI" , "BANGKA BELITUNG ISLANDS" , "NUSA
TENGGARA BARAT" , "SOUTH SUMATERA" ], 2 , inplace= True )
data[ 'OriginDaerah' ].replace([ "DKI JAKARTA" , "SOUTH KALIMANTAN" ,
"CENTRAL KALIMANTAN" , "TIMUR KALIMANTAN" , "RIAU ISLANDS" ,
"NUSA TENGGARA TIMUR" , "SELATAN SULAWESI" ], 3 , inplace = True )
data[ 'OriginDaerah' ].replace([ "BALI" , "GORONTALO" , "NORTH
KALIMANTAN" , "MALUKU" , "NORTH MALUKU" , "RIAU" , "WEST
SULAWESI" , "CENTRAL SULAWESI" , "SOUTHOUTH SULAWESI" , "NORTH
SULAWESI" , "WEST SUMATERA" , "NORTH SUMATRA" ], 4 , inplace= True )
data[ 'Origin' ].replace([ "ACEH" , "PAPUA" , "West Papua" , "TIMOR LESTE" ], 5 ,
inplace= True )
# Changing Department Data
data[ 'Department' ].replace( "IPA" , 1 , inplace= True )
data[ 'Department' ].replace( "IPS" , 2 , inplace= True )
data[ 'Department' ].replace( "SMK KES" , 3 , inplace= True )
data[ 'Department' ].replace( "SMK NON KES" , 4 , inplace= True )
data[ 'Department' ].replace( "OTHERS" , 5 , inplace= True )
# Changing Data Origin Information
data[ 'OriginInfo' ].replace([ "COMING TO CAMPUS" , "LECHERS" ,
"EMPLOYEES" , "FAMILY" , "STUDENTS" ], 1 , inplace= True )
data[ 'OriginInfo' ].replace([ "ALUMNI" , "BK TEACHER" , "CLASS SISTER" ,
"REGIONAL MARKETING" , "EDUCATION EXHIBITION" , "SOCIALIZATION
AT SCHOOL" , "FRIENDS" ], 2 , inplace= True )
data[ 'OriginInfo' ].replace([ "BROCHURE" , "PRINT MEDIA" ,
"BANDUNG/BALIHO" , "WEBSITE" , "SOCIAL MEDIA" ], 3 , inplace= True )
data[ 'OriginInfo' ].replace([ "TNI AD UNIT" , "OTHERS" ], 4 , inplace= True )
# Changing Actual Cluster Data
data[ 'Actual' ].replace([ "CLUSTER 1" ], 1 , inplace= True )
data[ 'Actual' ].replace([ "CLUSTER 2" ], 2 , inplace= True )
data[ 'Actual' ].replace([ "CLUSTER 3" ], 3 , inplace= True )
data[ 'Actual' ].replace([ "CLUSTER 4" ], 4 , inplace= True )
```

The initial data for Unjaya's new students after the *cleansing process* which is then ready to be processed using *K-Means Clustering* is shown as follows:

Table 12. Display of Unjaya's New Student Data

	NamaLengkap	Prodi	AsalDaerah	Jurusan	AsalInfo	Aktual
0	DITA NUR INDAHSARI	3	4	2	1	3
1	DITA DWI ANGGITA	3	2	1	1	1
2	ILHAM CAHYO NUGROHO	1	2	1	4	4
3	MUTIA TSANY KHAIRUNNISA	3	2	1	1	1
4	BEMI ADRIAN HANAFIAH	4	2	1	1	1
...	...	...	...	...	...	...
2313	ELLYSHA MUSTIKA SARI	3	2	2	1	1
2314	NOPI KURNIYATI	1	2	1	3	4
2315	ZERLI RAHMAWATI	2	2	1	3	4
2316	YUNI NURHAYATI	2	2	1	3	4
2317	DINDA ZULIA HARIANI	4	4	1	1	1

2318 rows × 6 columns

To facilitate the process of implementing *K-Means Clustering*, the identity of the Name is removed, and the results are as follows:

```
#Removing Full Name
data = data.drop([ "FullName" ], axis = 1 )
data
```

Table 13. Final Display of Unjaya's New Student Data

	Prodi	AsalDaerah	Jurusan	AsalInfo	Aktual
0	3	4	2	1	3
1	3	2	1	1	1
2	1	2	1	4	4
3	3	2	1	1	1
4	4	2	1	1	1
...	...	...	...	...	...
2313	3	2	2	1	1
2314	1	2	1	3	4
2315	2	2	1	3	4
2316	2	2	1	3	4
2317	4	4	1	1	1

2318 rows × 5 columns

- K-Means Clustering Implementation in Python
  To start implementing with *K-Means Clustering*, the first thing to do is look at the data types to make sure they can be processed in Python.
  data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2318 entries, 0 to 2317
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Prodi       2318 non-null   int64
 1   AsalDaerah  2318 non-null   int64
 2   Jurusan     2318 non-null   int64
 3   AsalInfo    2318 non-null   int64
 4   Aktual      2318 non-null   int64
dtypes: int64(5)
memory usage: 90.7 KB
```

Figure 4. Display of Unjaya's New Student Data Type

Next is displaying Initial Data Visualization
plt.scatter(data.Prodi, data.OriginRegion, data.Department, data.OriginInfo, marker= "o" , alpha= 1)
plt.show()

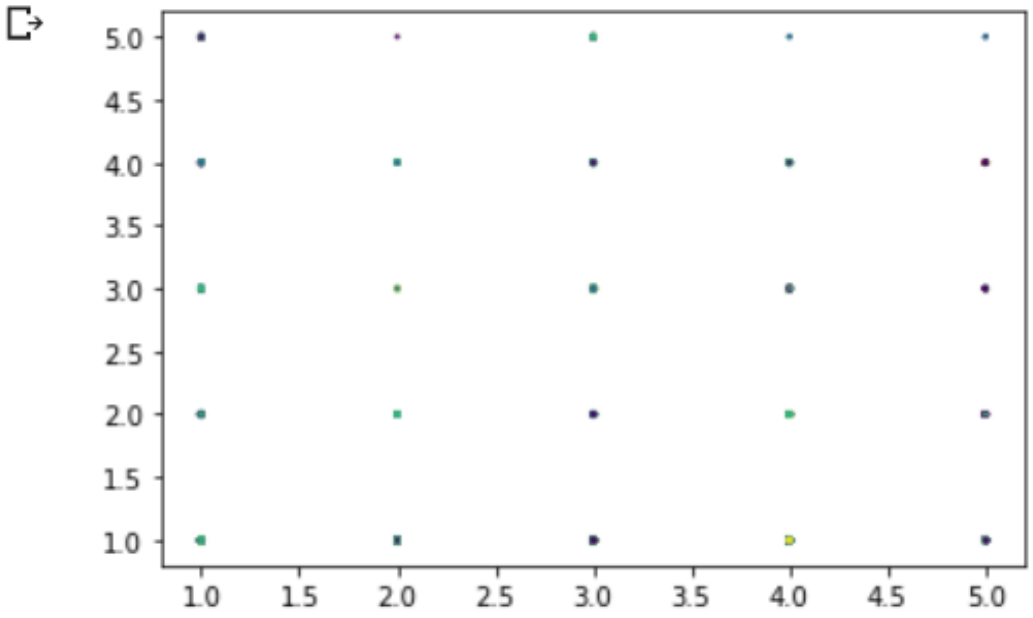

Figure 5. Initial Data Visualization

Next is to determine the variable x for the basis of the calculation as follows:
data_x = data.iloc[:, 0 : 4]

data_x.head()

Table 14. Variable X Data

	Prodi	AsalDaerah	Jurusan	AsalInfo
0	3	4	2	1
1	3	2	1	1
2	1	2	1	4
3	3	2	1	1
4	4	2	1	1

Then change the table to Array form
x_array = np.array(data_x)
print (x_array)

```
[[3 4 2 1]
 [3 2 1 1]
 [1 2 1 4]
 ...
 [2 2 1 3]
 [2 2 1 3]
 [4 4 1 1]]
```

Figure 6. True Array Data

Next perform data visualization using Scaler
scaler = MinMaxScaler()
x_scaled = scaler.fit_transform(x_array)
x_scaled

```
array([[0.5       , 0.75      , 0.25      , 0.        ],
       [0.5       , 0.25      , 0.        , 0.        ],
       [0.        , 0.25      , 0.        , 1.        ],
       ...,
       [0.25      , 0.25      , 0.        , 0.66666667],
       [0.25      , 0.25      , 0.        , 0.66666667],
       [0.75      , 0.75      , 0.        , 0.        ]])
```

Figure 7. Data Visualization with Scaler

The next step is to determine the number of data clusters to be processed with the KMeans function as follows:

```
kmeans = KMeans(n_Klasters = 4 , random_state= 123 )
kmeans.fit(x_scaled)
```

```
KMeans(algorithm='auto', copy_x=True, init='k-means++', max_iter=300,
       n_clusters=4, n_init=10, n_jobs=None, precompute_distances='auto',
       random_state=123, tol=0.0001, verbose=0)
```

Figure 8. Defining Data Clusters

Next displays the new Centroid from the data

```
print (kmeans.Cluster_centers_)
```

```
[[0.63676732 0.09946714 0.23712256 0.76317348]
 [0.34893993 0.73409894 0.19434629 0.4581861 ]
 [0.03871201 0.09479016 0.28762663 0.42932947]
 [0.62932138 0.10883483 0.23111396 0.16303884]]
```

Figure 9. Centroid Data

The next step is to create a new column named Prediction which is the result of the *K-Means Clustering calculation* as follows:

```
data[ "Prediction" ] = kmeans.labels_
output = plt.scatter(x_scaled[:, 0 ], x_scaled[:, 1 ], s = 100 , c = data.Prediction, marker = "o" , alpha = 1 , )
centers = kmeans.Klaster_centers_
plt.scatter(centers[:, 0 ], centers[:, 1 ], c= 'red' , s= 200 , alpha= 1 , marker= "o" );
plt.title( "K-Means Clustering Results" )
plt.colorbar (output)
plt.show()
```

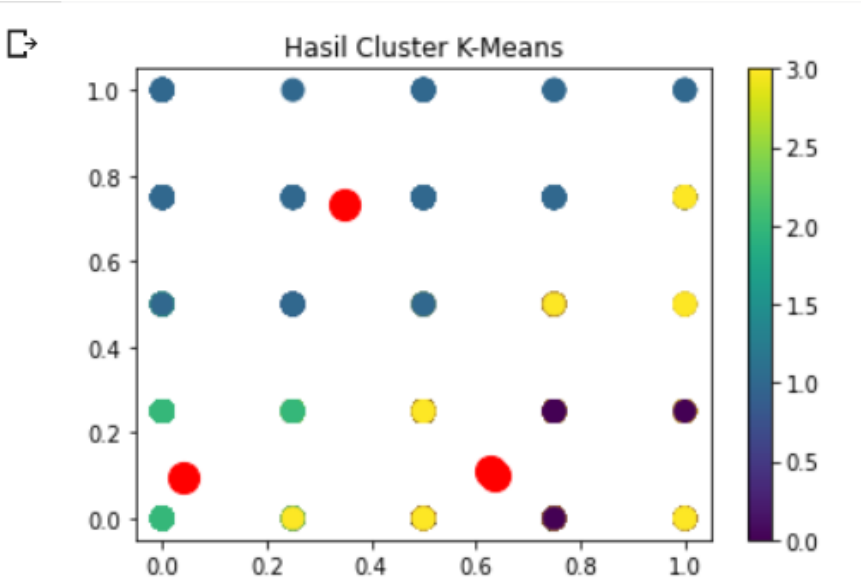

Figure 10. Results of K-Means . Cluster

Then to see the results of the calculation, the Prediction data is displayed as follows:

Data

Table 15. Display of K-Means . Cluster Result Data

	Prodi	AsalDaerah	Jurusan	AsalInfo	Aktual	Prediksi
0	3	4	2	1	3	1
1	3	2	1	1	1	3
2	1	2	1	4	4	2
3	3	2	1	1	1	3
4	4	2	1	1	1	3
...	...	...	...	...	...	...
2313	3	2	2	1	1	3
2314	1	2	1	3	4	2
2315	2	2	1	3	4	2
2316	2	2	1	3	4	2
2317	4	4	1	1	1	1

2318 rows × 6 columns

Because Python always makes the index from the number 0, while the actual Cluster is calculated from the number 1, then the process of transforming numbers from 0 to 1, 1 to 2, 2 to 3, and 3 to 4 is as follows:

```
data[ 'Predict' ].replace( 3 , 4 , inplace= True )
data[ 'Predict' ].replace( 2 , 3 , inplace= True )
data[ 'Predict' ].replace( 1 , 2 , inplace= True )
data[ 'Predict' ].replace( 0 , 1 , inplace= True )
```

Table 16. Display of Predicted Data for *K-Means Clustering Calculation Results*

	Prodi	AsalDaerah	Jurusan	AsalInfo	Aktual	Prediksi
0	3	4	2	1	3	2
1	3	2	1	1	1	4
2	1	2	1	4	4	3
3	3	2	1	1	1	4
4	4	2	1	1	1	4
...	...	...	...	...	...	...
2313	3	2	2	1	1	4
2314	1	2	1	3	4	3
2315	2	2	1	3	4	3
2316	2	2	1	3	4	3
2317	4	4	1	1	1	2

2318 rows × 6 columns

- Implementation of the Confusion Matrix in Python

  Based on the results of the *K-Means Clustering calculation*, then the testing process using the *Confusion Matrix Multi Class is carried out* by first taking the actual and predicted data as follows:

```
data_cm = data.loc[:,[ 'Actual' , 'Predict' ]]
data_cm
```

Table 17. Actual and Predicted Data for *K-Means Clustering Calculation Results*

	Aktual	Prediksi
0	3	1
1	1	3
2	4	2
3	1	3
4	1	3
...	...	...
2313	1	3
2314	4	2
2315	4	2
2316	4	2
2317	1	1

2318 rows × 2 columns

Confusion Matrix calculation is carried out using the Sklern library as follows:

```
from sklearn.metrics import confusion_matrix
confusion_matrix = confusion_matrix(data_cm[ 'Actual' ], data_cm[ 'Predict' ])
print (confusion_matrix)
```

```
[[178 113  74 508]
 [182   0  43 262]
 [ 11 127  19  11]
 [192  43 555   0]]
```

Figure 18. *Confusion Matrix Calculation* of Data

The last step is the process of calculating the value of accuracy, precision, and recall of data using Sklern

```
from sklearn.metrics import classification_report
print (classification_report(data_cm[ 'Actual' ], data_cm[ 'Prediction' ]))
```

Table 18 The Final Result of the *Confusion Matrix Process* on the Data

	precision	recall	f1-score	support
1	0.32	0.20	0.25	873
2	0.00	0.00	0.00	487
3	0.03	0.11	0.04	168
4	0.00	0.00	0.00	790
accuracy			0.08	2318
macro avg	0.09	0.08	0.07	2318
weighted avg	0.12	0.08	0.10	2318

F. Visualization of K-Means Calculation Results

Making a visualization of Unjaya's new student cluster data as a result of the K-Means calculation as follows:

- Visualization of the Number of Cluster Members

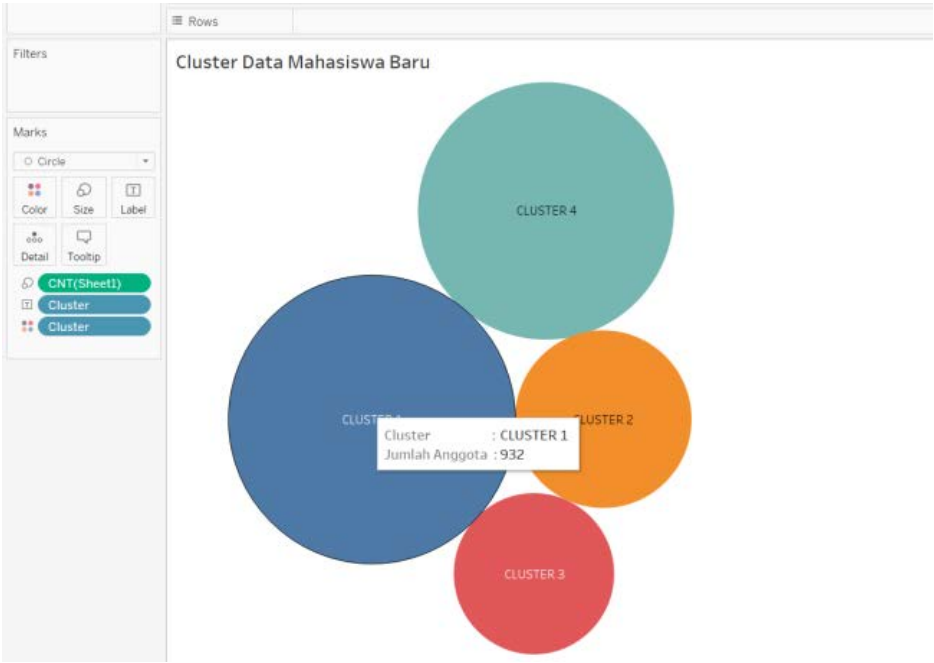

Figure 20. Visualization of K-Means. Cluster Calculation Results

The visualization depicted above illustrates the formation of four distinct data clusters derived from the K-Means Clustering algorithm. Each cluster is represented by a uniquely colored circle, with the size of each circle proportionate to the number of data points contained within the respective cluster. Larger circles indicate a greater volume of data, thereby enabling PMB Unjaya policymakers to readily identify clusters with higher membership. This visual representation serves as a strategic reference in determining appropriate targets for promotional activities, facilitating data-driven decision-making in the context of student recruitment.

- Visualization of Cluster Distribution per Study Program per Year

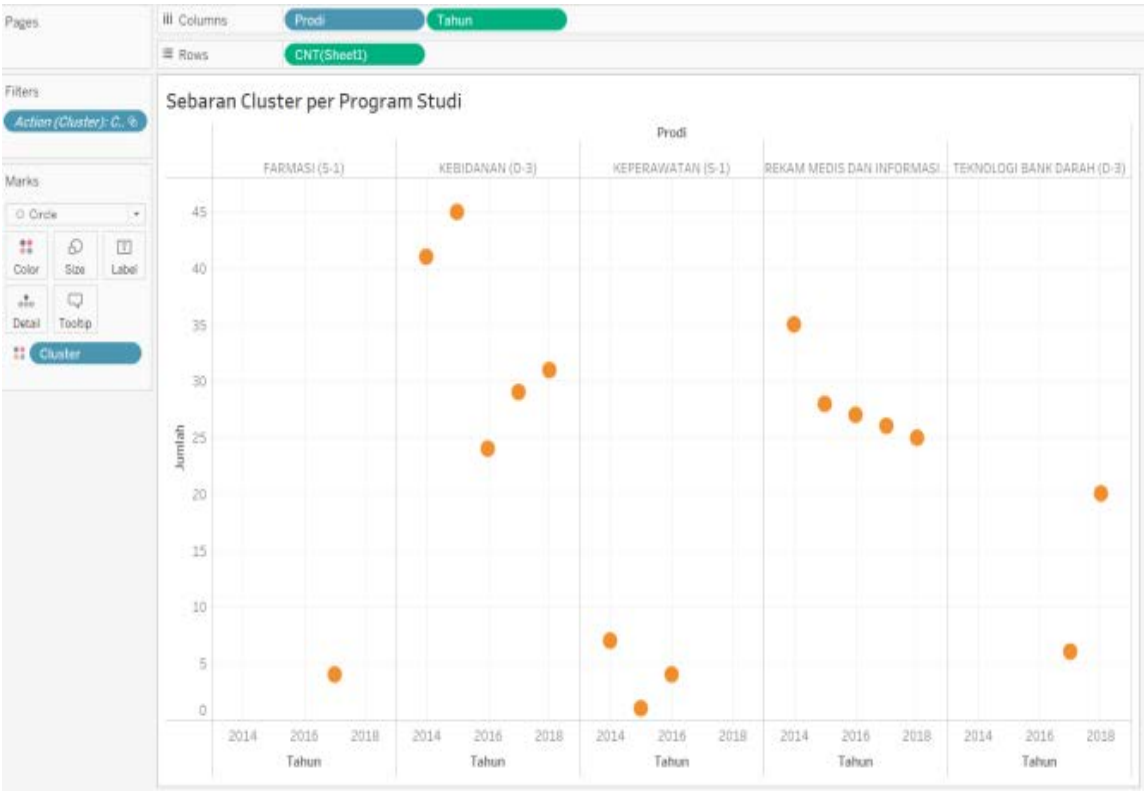

Figure 21. Visualization of Cluster Distribution per Study Program per Year

The visualization presented above depicts the distribution of data across clusters, segmented by study program and academic year. This representation enables the observation of temporal and programmatic trends within each cluster, offering insights into the movement and composition of data over time. Such visual analytics are expected to assist PMB Unjaya policymakers in understanding the dynamics of each cluster in relation to specific study programs and years. Consequently, this information can serve as a strategic foundation for accurately identifying and selecting appropriate targets for promotional initiatives.

- Visualization of the Distribution of New Student Clusters by Region

Figure 22. Visualization of Cluster Distribution from Regions

The picture above is a visualization of the distribution of data from each cluster per region. From this distribution, it will be seen how the data travels for each cluster per region. From this visualization, it is hoped that it can help PMB Unjaya policymakers know the condition of the cluster per region so that they can correctly determine promotion options for certain areas that are the main sources of Unjaya's new students, and evaluate other areas based on data clusters.

After the three cluster visualizations have been developed, the next step is to integrate them into a single information dashboard, which will subsequently be presented to PMB Unjaya stakeholders. In the dashboard design, when a cluster circle is clicked, the corresponding data within that cluster will be displayed. This includes a visualization of the condition of the selected cluster, showing the distribution of its members across study programs by year, as well as the geographical distribution of data points within the cluster. This integrated dashboard visualization is presented as follows:

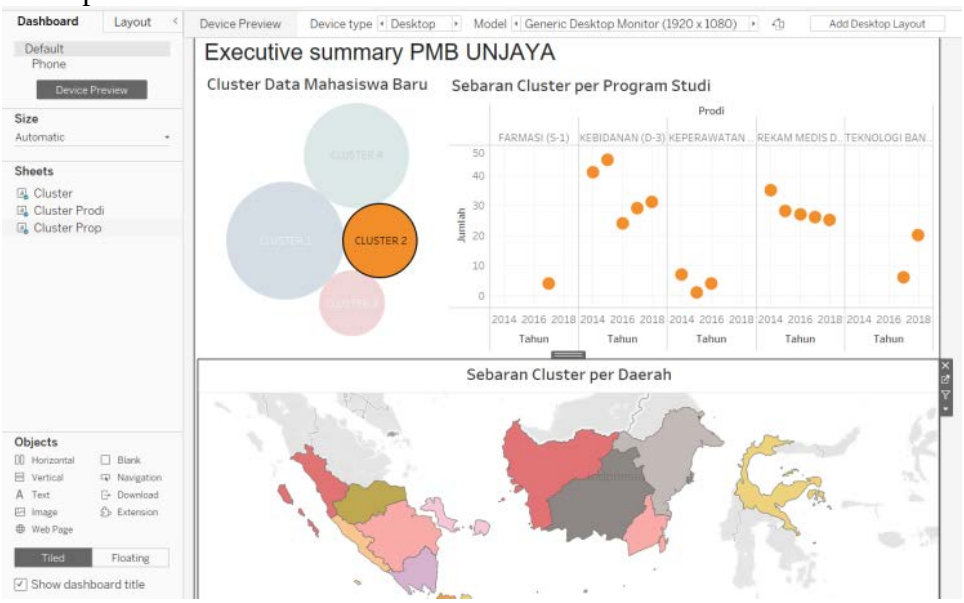

Figure 23. Visualization of PMB Unjaya's *Executive Summary Dashboard*

The final display of the new student cluster dashboard using the K-Means method presented to PMB Unjaya policy makers is as follows:

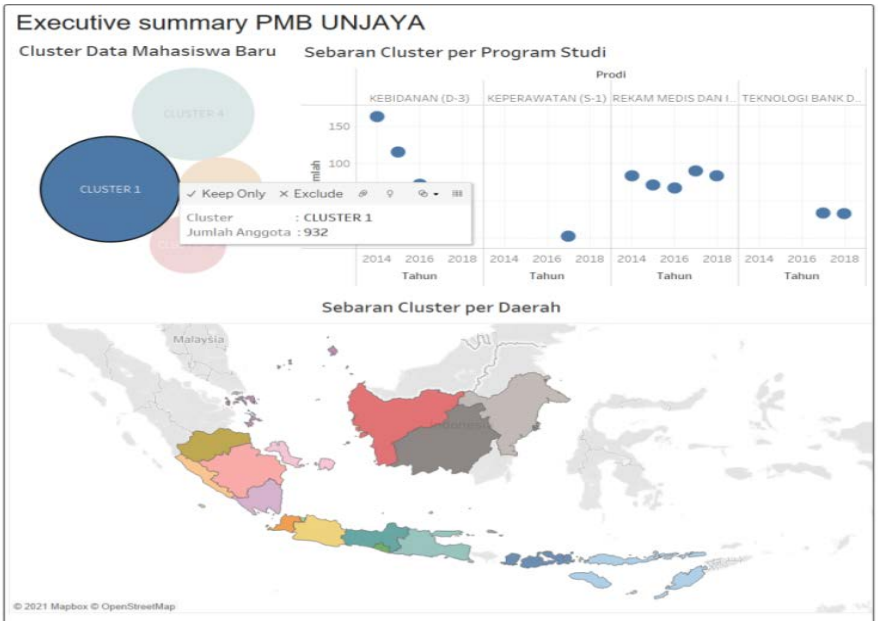

Figure 24. PMB Unjaya *Executive Summary Dashboard* with *the K-Means Clustering Method*

The visualization presented above illustrates that when PMB Unjaya policymakers select a specific cluster, all data associated with that cluster is displayed. For instance, if Cluster 1 is chosen, the system will present the number of members in Cluster 1 over the past five years, accompanied by a visual representation of the distribution of cluster members across study programs by year. Additionally, it will show the geographical distribution of Cluster 1 members. This visualization provides valuable insights for PMB Unjaya policymakers, enabling them to evaluate trends in new student enrollment within each cluster prior to determining which cluster should be targeted for promotional activities. Consequently, PMB Unjaya policymakers can design promotional strategies more efficiently and objectively.

V. CONCLUSIONS AND RECOMMENDATIONS

A. Conclusion

Based on the research conducted using new student enrollment data from Unjaya over the past five years, it can be concluded that the K-Means Clustering algorithm, in conjunction with the Confusion Matrix method, is effective in identifying distinct clusters among incoming students. These clusters provide actionable insights that can be utilized by PMB Unjaya stakeholders in formulating targeted promotional strategies. In alignment with the research objectives, the following conclusions are drawn from the analysis and discussion:

- Cluster 1 represents the largest group among Unjaya's new student population. Therefore, it is essential for PMB Unjaya stakeholders to sustain promotional efforts in the regions associated with this cluster.
- Zone 1, comprising the regions of DI Yogyakarta, West Java, Central Java, and East Java, constitutes the primary geographic origin of Unjaya's new students. These areas should be prioritized as key targets in future promotional campaigns.
- The Science stream is the most prevalent academic background among incoming students from SMA/SMK/MA, indicating a strong preference for science-related programs.
- Option 3, which targets students from Zone 1 with a Science background and utilizes promotional media such as brochures, print media, banners/billboards, websites, and social media, is identified as the most effective strategy for PMB Unjaya policymakers in determining promotional targets.

B. Suggestion

Based on the findings and conclusions of this study, several recommendations are proposed for future research:

- This study was limited to computational analysis and standard visualization outputs. Future research is encouraged to develop an integrated application capable of real-time connectivity with Unjaya's Sicama system. Such integration would enable immediate responsiveness to data changes, thereby supporting policymakers in making timely and informed decisions regarding promotional target selection for PMB Unjaya.
- Subsequent studies are also advised to incorporate additional parameters, such as the income level of prospective students' parents. This enhancement would allow for a more comprehensive mapping of the financial profiles of incoming students, which could be valuable for planning support services and financial aid strategies throughout the study period.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to the Unjaya KPP Bureau for providing the opportunity and support that enabled this research to be conducted successfully.

REFERENCE

[1] Aaker, D. A., Kumar, V., Leone, R., & Day, G. S. (2012). Marketing research. Wiley. Available at: https://books.google.co.id/books?id=lPssygEACAAJ

[2] Alviana, S., & Kurniawan, B. (2019). Analisis data penerimaan mahasiswa baru untuk meningkatkan potensi pemasaran universitas menggunakan Business Intelligence (Studi kasus Universitas XYZ). Infotronik: Jurnal Teknologi Informasi dan Elektronika, 4(1), 10–15. https://doi.org/10.32897/infotronik.2019.4.1.2

[3] Amborowati, A., & Suyanto, M. (2015). Studi dukungan Marketing Intelligence pada strategi pemasaran. Seminar Nasional Informatika 2015.

[4] Dhuhita, W. (2015). Clustering menggunakan metode K-Means untuk menentukan status gizi balita. Jurnal Informatika Darmajaya, 15(2), 160–174.

[5] Dinata, R. K., Safwandi, S., Hasdyna, N., & Azizah, N. (2020). Analisis K-Means Clustering pada data sepeda motor. INFORMAL: Informatics Journal, 5(1), 10. https://doi.org/10.19184/isj.v5i1.17071

[6] Geria, A. A. G. A. (2018). Social media as promotion trend for increasing tourist visit towards digital era. International Journal of Social Sciences and Humanities. https://doi.org/10.29332/ijssh.v2n3.204

[7] Guoxiang, L., & Zhiheng, Q. (2013). Data mining applications in marketing strategy. In Proceedings of the 2013 3rd International Conference on Intelligent System Design and Engineering Applications (ISDEA 2013) (pp. 518–520). https://doi.org/10.1109/ISDEA.2012.126

[8] Han, J., Pei, J., & Kamber, M. (2011). Data mining: Concepts and techniques. Elsevier Science. Available at: https://books.google.co.id/books?id=pQws07tdpjoC

[9] Henderi, H., Handayani, I., & Dewi, M. A. (2012). Business Intelligence development model using Star Schema methodology. CCIT Journal, 5(3), 233–250. https://doi.org/10.33050/ccit.v5i3.154

[10] Kurniawati, I., Indrajit, R. E., & Fauzi, M. (2017). Peran Business Intelligence dalam menentukan strategi promosi penerimaan mahasiswa baru. Ikraith Informatika, 1(2), 70–79.

[11] Laudon, K. C., & Laudon, J. P. (2011). Management information systems: Managing the digital firm. Prentice Hall PTR. Available at: https://books.google.co.id/books?id=4T9EYgEACAAJ

[12] Mishra, R., & Saini, A. K. (2016). Business intelligence and analytics: Paving way for operational excellence in Indian banks. In Proceedings of the International Conference on Industrial Engineering and Operations Management. https://doi.org/10.21013/jmss.v3.n1.p14

[13] Purnamaningsih, C., Saptono, R., & Aziz, A. (2016). Pemanfaatan metode K-Means Clustering dalam penentuan penjurusan siswa SMA. Jurnal Teknologi & Informasi ITSmart, 3(1), 27. https://doi.org/10.20961/its.v3i1.644

[14] Rahman, M. F., Alamsah, D., Darmawidjadja, M. I., & Nurma, I. (2017). Klasifikasi untuk diagnosa diabetes menggunakan metode Bayesian Regularization Neural Network (RBNN). Jurnal Informatika, 11(1), 36. https://doi.org/10.26555/jifo.v11i1.a5452

[15] Rarasati, D. B. (2020). A grouping of song-lyric themes using K-Means Clustering. JISA (Jurnal Informatika dan Sains), 3(2). https://doi.org/10.31326/jisa.v3i2.658

[16] Sutrisno, B. T., & Andriyani, W. (2020). Penerapan MADM dengan metode SAW untuk menentukan target promosi berdasarkan asal jurusan di sekolah. Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer, 11(2). https://doi.org/10.24176/simet.v11i2.4784

[17] Suprawoto, T. (2016). Klasifikasi data mahasiswa menggunakan metode K-Means untuk menunjang pemilihan strategi pemasaran. JIKO (Jurnal Informatika dan Komputer), 1(1). https://doi.org/10.26798/jiko.2016.v1i1.9

[18] Turban, E., Aronson, J. E., & Liang, T. P. (2005). Decision support systems and intelligent systems. Prentice Hall. Available at: https://books.google.co.id/books?id=m0R5QgAACAAJ

[19] Yulianton, H. (2008). Data mining untuk dunia bisnis keputusan informasi. Jurnal Teknologi Informasi DINAMIK. https://doi.org/10.35315/dinamik.v13i1.65