International Journal of Information Technology and Computer Science Applications (IJITCSA)
p-ISSN: 2964-3139 e-ISSN: 2985-5330
Vol. 02, No. 03, page 130 - 138
Submitted 09/06/2024; Accepted 20/08/2024; Published 30/09/2024

Descriptive Analysis Of K-Means and Apriori Methods to Find Promotion Strategies For University Bhayangkara.

[1]Sultan Bacharuddin Yusuf Hidayat and [2]Dr. Tb. Ai. Munandar S.Kom., M.T.
[1-2]Informatics, Universitas Bhayangkara Jakarta Raya, Bekasi, INDONESIA
e-mail: [1]sultanbyh@gmail.com, [2] tbaimunandar@gmail.com

Corresponding Autor: Dr. Tb. Ai. Munandar S.Kom., M.T.

Abstract
The increasing number of higher education institutions in Indonesia has intensified competition between universities. Universitas Bhayangkara Jakarta Raya (Ubhara Jaya) must develop an effective marketing plan to stand out. This study used segmentation and associative analysis on 2023 student enrolment data. The Apriori algorithm identified patterns in student preferences for study programmes, while the K-Means method categorized students based on demographics and family income. Three income-based clusters were identified: C0 'Already stable' (IDR 1,000,000 - 2,500,000), C1 'Focus on promotion strategy' (IDR 20,000,000), and C2 'Maximise promotion again' (IDR 5,000,000 - 10,000,000). The Davies-Bouldin Index (DBI) indicated k=5 as the optimal cluster number, but k=3 was adequate with a minimal score difference. The most popular programmes were Communication Science and Management, with high support and confidence values. This data helps Ubhara Jaya manage study programme demand and room availability. Combining K-Means and Apriori algorithms is expected to enhance data segmentation[1] and support effective marketing strategies, aiding strategic decisions in higher education marketing.

Keywords—: Segmentation, Associative Analysis, Data Mining, K-Means, Apriori, Promotion Strategy

1 Introduction
The development of education in Indonesia has increased the competitiveness of universities over time. With 4,593[2] institutions, universities - especially private universities - need to stand out from the competition and attract students, who are their main source of funding. As a private institution, Ubhara Jaya needs to use effective marketing techniques to attract prospective students. Student enrolment data for 2023 was used in this study to create a marketing plan for Ubhara Jaya. Students were categorised based on their family income and [1]demographics using [3]K-Means, while students' preference patterns for study programmes were discovered using Apriori algorithm[4]. For the purpose of the study, student data from Jabodetabek (Jakarta, Bogor, Depok, Tangerang, and Bekasi).

Utilising data mining in particular, information technology is used to transform data into meaningful information. The Apriori method is used to identify frequently occurring itemsets and examine the relationship between the preferences of study programme I and study programme II, while K-Means organises the data according to comparable features. It is expected that the combination of these two algorithms will produce the best segmentation for marketing plans. This research uses K-Means and Apriori algorithms to extract data from 2023 student data to categorise prospective new Ubhara Jaya students in Jabodetabek for efficient marketing efforts.

2 Reacearch Method

2.1 Data mining, Clustering, Association
Data mining[5], also known as knowledge discovery or pattern recognition, aims to extract hidden knowledge from data through automated or semi-automated processes to find meaningful patterns that generate economic benefits. The advancement of data mining is fuelled by the increasing amount of data, storage in data warehouses, internet access, business competition, software technology, and computing capabilities.

One of the main techniques in data mining is clustering[6][7][8][9], which groups data based on similarity in the absence of a target variable, dividing data into homogeneous groups to discover hidden structures. In addition, association techniques, or Market Basket Analysis[1][10][11][12], are used to find attributes that often appear together, looking for association rules between combinations of items. These techniques are used to design effective marketing strategies such as placement of frequently purchased items together, catalogue design, coupon giving, and package sales, helping businesses understand customer buying patterns and optimise product offerings.

3 Reasearch Method

The Cross Industry Standard Process Model for Data Mining (CRISP-DM) approach was used to implement the research. An outline of the CRISP-DM[5] process flow is shown below.

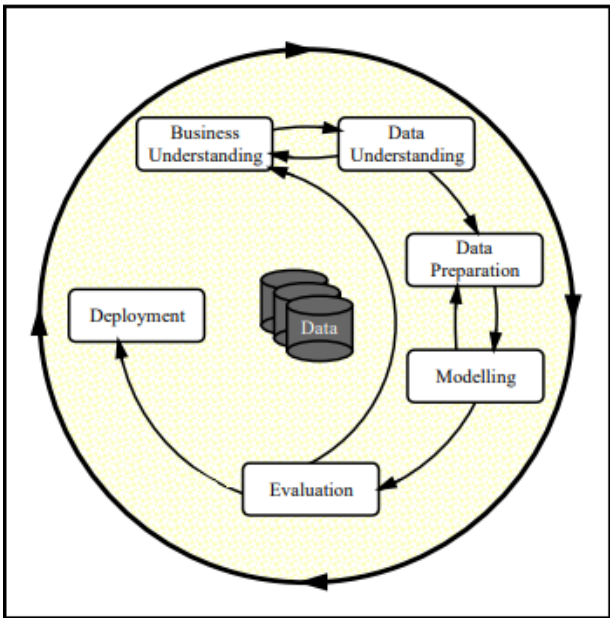


Figure 1. CRIPS-DM Flow

3.1 Busniess Understanding

Universitas Bhayangkara Jakarta Raya (Ubhara Jaya) is facing intense competition in the higher education market and needs to improve its promotional strategy. This project aims to apply segmentation-based descriptive and associative analysis to identify potential student segments as well as study programme interest patterns. By using K-Means algorithm for segmentation based on demographic and geographic attributes, and Apriori algorithm for associative analysis, Ubhara Jaya can understand the preferences of potential students and the relationship between study programmes. The results will help develop more effective promotion strategies, optimise resources, and make better strategic decisions in marketing and student recruitment.

3.2 Data Understanding

After understanding the company's objectives, you should understand the data provided by the Ubhara Marketing Bureau. The K-Means algorithm uses data from 1980, which includes sub-districts, hours travelling time, minutes travelling time, KM, and income of parents of different students. The data for the Apriori algorithm is study programme choice I and study programme choice II.

	ID_KECAMATAN	WAKTU TEMPUH JAM	WAKTU TEMPUH MENIT	KM	Gaji	Pilihan Prodi I	Pilihan Prodi II
0	46.0	1.0	30.0	55.0	2000.0	MANAJEMEN 1	MANAJEMEN 2
1	47.0	1.0	5.0	45.0	2000.0	INFORMATIKA 1	MANAJEMEN 2
2	45.0	1.0	2.0	40.0	2000.0	INFORMATIKA 1	INFORMATIKA 2
3	47.0	1.0	5.0	45.0	20000.0	PSIKOLOGI 1	PSIKOLOGI 2
4	48.0	1.0	30.0	60.0	5000.0	ILMU KOMUNIKASI 1	PENDIDIKAN GURU SEKOLAH DASAR 2
5	49.0	1.0	30.0	55.0	2000.0	MANAJEMEN 1	MANAJEMEN 2
6	31.0	1.0	15.0	70.0	500.0	AKUNTANSI 1	ILMU KOMUNIKASI 2
7	32.0	1.0	20.0	65.0	1000.0	AKUNTANSI 1	MANAJEMEN 2
8	32.0	1.0	20.0	65.0	1000.0	AKUNTANSI 1	MANAJEMEN 2
9	34.0	1.0	30.0	70.0	5000.0	PSIKOLOGI 1	PSIKOLOGI 2
10	33.0	1.0	30.0	55.0	5000.0	PSIKOLOGI 1	PSIKOLOGI 2
11	36.0	1.0	1.0	25.0	1000.0	MANAJEMEN 1	MANAJEMEN 2
12	36.0	1.0	1.0	25.0	2000.0	MANAJEMEN 1	TEKNIK INDUSTRI 2
13	36.0	1.0	1.0	25.0	2000.0	INFORMATIKA 1	TEKNIK INDUSTRI 2

Figure 2. Result Displaying Data

3.3 Data Preparation

Data preparation is a stage to fix problems contained in the data before the data enters the modelling stage so as to produce good modelling.

1. Data Selection
   Student admission data at Ubhara consists of 7 columns of variables, which will be used 4 columns of variable travel time hours, travel time minutes, Kilometers, parents' income for k-means and 2 columns of study programme I, study programme II for a priori from 2023 student data. There are no discarded variables in the student data.

2. Data Preprocessing
   At this step, cleaning invalid data and missing values is carried out. In this data, there is no missing value or invalid data.

```
ID_KECAMATAN          0
WAKTU TEMPUH JAM      0
WAKTU TEMPUH MENIT    0
KM                    0
Gaji                  0
Pilihan Prodi I       0
Pilihan Prodi II      0
dtype: int64
```

Figure 3. Missing Value Check Result

3.4 Modeling

In the modelling stage for Ubhara student data processing with the help of two main techniques used, namely K-Means for segmentation and Apriori for associative analysis.

1. Application of K-Means

At this stage k-Means is used to segment students or students of Ubhara based on certain characteristics, such as demographics and geography. With the help of Elbow method to determine the optimal cluster centre value.

To calculate the ideal number of clusters, first select columns from the organised data. Calculating and graphing the elbow method can be done using the variables travel time hours, travel time minutes, Kilometres, parents' income. The graph will show the optimal number of clusters, with two elbow points at k=2 and k=3. The significant decrease in cluster values after k=3 indicates that the optimal number of clusters is three, which means that the data can be grouped into three clusters.

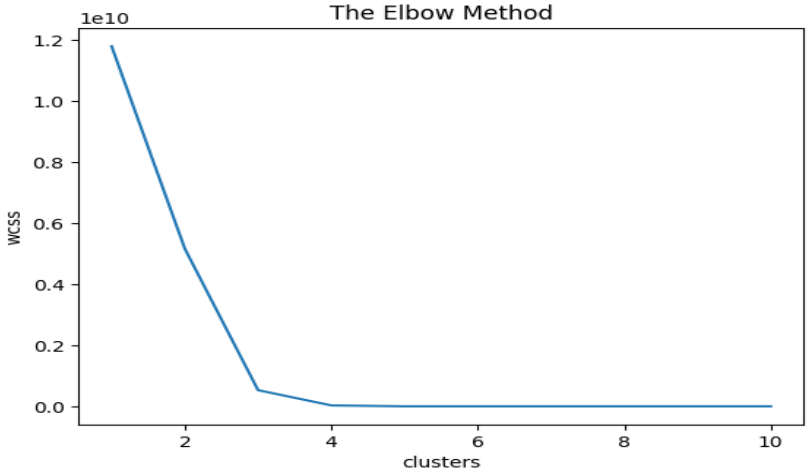


Figure 4. Elbow Method Graphics

The purpose of this k-means algorithm in 2023 is to collect data on student area, parent income, and school type. Since the elbow method produces the best cluster of 3, n_cluster is filled with 3.
Next, a data clustering graph is created with distance points, hours, minutes, kilometres, and salary after clustering. In the graph, the x-axis shows the cluster result, and the y-axis shows the centroid result.

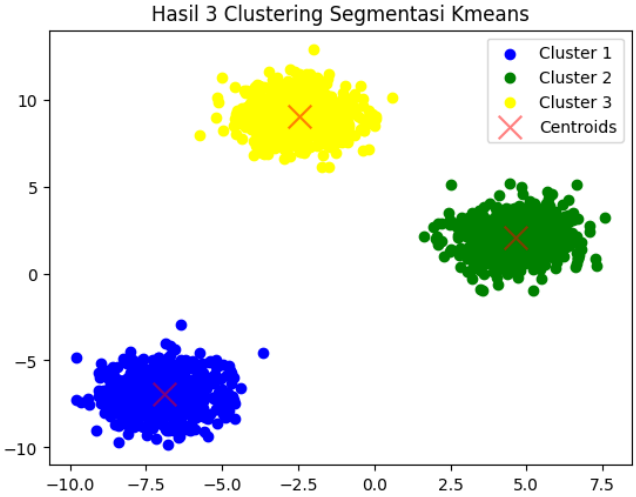


Figure 4. Scatter Plot Clustering

cluster 1 Districts of North Bekasi, South Tambun, North Tambun, Babelan, Mustika jaya, Cibitung, and Cakung the average income of parents in cluster 1 is IDR 1,000,000 to IDR 2,500,000, cluster 2 Districts of Tarumajaya, Rawalumbu, Tapos, and Pamijahan the average income of parents in cluster 2 exceeds IDR 20. 000,000 so it is highly recommended to focus on promotion in these sub-districts, cluster 3 West Bekasi, East Bekasi, South Bekasi, West Cikarang, and Duren Sawit sub-districts, the average income of parents in this cluster is IDR 5,000,000 to IDR 10,000,000 so it is recommended to maximise it again in these sub-districts. The cluster results are shown in Table 1

Table 1: Clustering Results

Cluster Name	Sub-district	Initials Parent Income	Cluster Number	Promotion Rate
0	1, 2 ,3 ,4 ,5 ,6 ,7 ,8 ,9 ,10 , 12 ,13 ,14 ,15 ,16 ,17 ,18 , 19 ,20 ,22 ,23 ,24 ,26 ,27 , 28 ,29 ,30 , 45 , 46, 47, 31 ,32 ,36 ,37 ,38 ,39 ,44 ,50, 51, 52 ,53, 54, 57, 58, 59, 60, 61, 62, 63, 64, 66, 67, 68, 69, 70, 71, 72, 73	IDR. 1.000.000 – IDR. 2.500.000	1410	Stabilised
1	1, 2, 3, 5, 12 ,16 ,19, 22, 24, 30	IDR. 20.000.000	22	Focus on Promotion
2	1, 2, 3, 4, 5, 6, 8, 9, 12, 13, 14, 15, 16 ,17 ,18, 20, 21, 22, 24, 25, 28, 29, 30, 48, 34, 33, 36, 40, 41, 42, 50, 54, 56, 60, 61, 62, 63, 64, 65, 68	IDR 5.000.000 –IDR. 10.000.000	548	Maximal again in promotion

2. Application of Apriori

In this step, Apriori is used to find associative rules in the transaction data. This is used to create better promotion plans and find patterns of student interest in study programmes. The process includes:

1. Selecting a dataset to process.
2. Determining the minimum values 7% of support and confidence.
3. Presenting the created association rules.

The selected association rules are applied through a coding process, which creates patterns of student interest in study programmes.

A. Combination of 1 Itemsets

Tabel 2: Combination Result 1 itemset

No.	Support	Itemset
1	0.072272	Akuntnasi 1
2	0.115564	Ilmu Hukum 1
3	0.103914	Ilmu Hukum 2
4	0.162162	Ilmu Komunikasi 1
5	0.186859	Ilmu Komunikasi 2
6	0.157036	Informatika 1
7	0.105312	Informatika 1
8	0.185461	Manajemen 1
9	0.211090	Manajemen 2
10	0.118826	Psikologi 1
11	0.125815	Psikologi 2
12	0.118360	Teknik Industri 1
13	0.094129	Teknik Industri 2

Based on the process of forming itemsets in Table 2 with a minimum support of 7%, it can be seen that meeting the minimum support standards is in the formation of frequent items AK1, IH1, IH2, IK1, IK2, IF1, IF2, MJ1, MJ2, PSG1, PSG2, TI1, TI2. Then from the results of 1 itemset, a combination of 2 itemsets will be made as in Table 3.

B. Combination of 2 Itemsets

Tabel 3: Combination Result 2 itemset

No.	Support	Itemset
1	0.074091	Ilmu Komunikasi 1, Ilmu Komunikasi 2
2	0.079683	Manajemen 1, Manajemen 2

Based on the combination of 2 itemsets in Table 4.4, it can be seen that the combination of 2 itemsets that meet the minimum support standard of 7% is IK1 / IK2, MJ1 / MJ2 because only 2 features can be used as an itemset combination, so it cannot be calculated for the formation of 3 itemsets, so association rules are formed.

C. Finding Association Rules

After all the high frequency patterns are found. Then look for association rules that meet the minimum requirements for confidence by calculating the confidence of associative rules with a minimum value of confidence = 40%.

Tabel 4: Frequent Item Results

No.	Itemset	Confidence
1	Ilmu Komunikasi 1-> Ilmu Komunikasi 2	0.456897
2	Manajemen 1-> Manajemen 2	0.429648
3	Ilmu Komunikasi 2-> Ilmu Komunikasi 1	0.396509
4	Manajemen 2-> Manajemen 1	0.377483

The output results can be seen that students who choose study programme I Communication Science will choose Communication Science in the selection of study programme II and if they choose study programme I Management then in the selection of study programme II they will choose Management.

3.5 Evaluation DBI

Based on the results of the Davies Bouldin Index calculation, a score for the number of clusters is obtained in Table 5.

Tabel 5. Davies Boulding Index Score

Number of K	Score
2	0.080
3	0.093
4	0.123
5	0.026
6	0.223

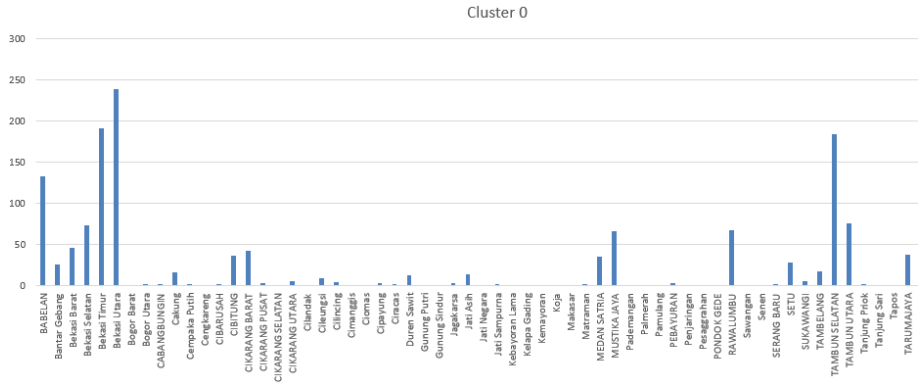
Based on the results from Table 5, the best k scores are 2 and 5, but with the comparison of k scores between 3 and 5 only 0.07, k = 3 is considered quite appropriate as the total number of clusters for clustering on this dataset.

4 Results and Discussion

1. K-means Analysis

Clustering 0 and its Recommendation

Figure 5. District Cluster 0



Based on Figure 5, cluster 0 includes sub-districts with an average income of Rp 1,000,000 - Rp 2,500,000, with 1410 students. Babelan, East Bekasi, North Bekasi, South Bekasi, and South Tambun sub-districts show a good number of students, and it is recommended that the marketing agency maintain promotions in these areas.

Clustering 1 and its Recommendations

Figure 6. District Cluster 1

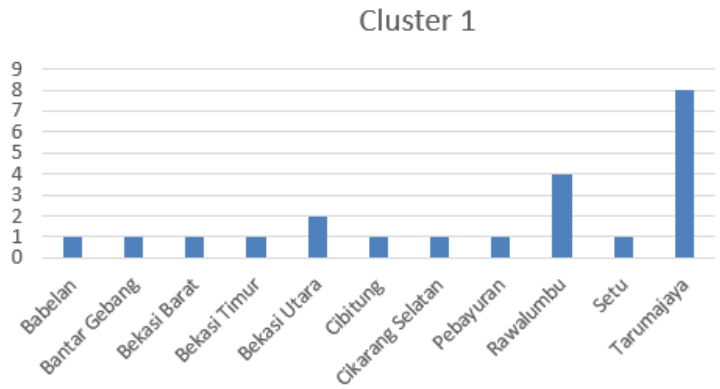
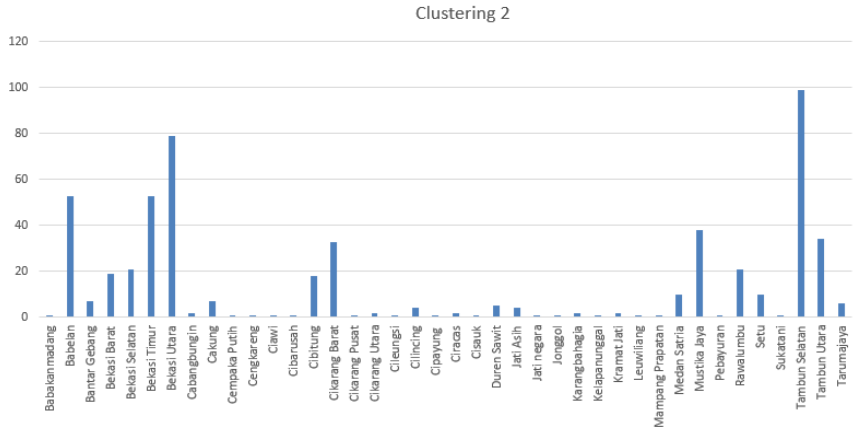


Figure 6. shows cluster 1 with an average income of IDR 20,000,000 and 22 students. Tarumajaya sub-district has a very high number of students compared to other sub-districts. It is recommended to focus on promotion in Tarumajaya sub-district.

Clustering 2 and its Recommendations

Figure 7. District Cluster 2



Based on Figure 7. cluster 2 includes sub-districts with an average income of Rp 5,000,000 - Rp 10,000,000 and 548 students. The sub-districts of South Tambun, North Bekasi, East Bekasi, and Babelan are already quite good. It is recommended that promotion in cluster 2 be maximised.

2. Apriori Analysis

Mahasiswa Universitas Bhayangkara Jakarta Raya menunjukkan minat pada beberapa program studi, dengan persentase support tertinggi untuk program studi I Manajemen (18%) dan program studi II Manajemen (21%). Program studi I dan II Ilmu Komunikasi juga diminati dengan persentase support masing-masing 16% dan 18%. Terdapat juga itemset dengan support sebesar 7,49% untuk kombinasi program studi I dan II Ilmu Komunikasi, serta 7,9% untuk kombinasi program studi I dan II Manajemen. Nilai confidence tertinggi terjadi pada kombinasi program studi I dan II Ilmu Komunikasi (45,6%) dan program studi I dan II Manajemen (42,9%).

3. Correlation and Relationship

Cluster 0:

Babelan sub-district has 140 students, East Bekasi 190, North Bekasi 240, South Bekasi 80, and South Tambun 190, with an average income of Rp 1,000,000 - Rp 2,500,000. Most choose Management I and II study programmes, with strong Apriori rules and strong correlations.

Cluster 1:

Tarumajaya sub-district has 8 students and Rawalumbu 4, with an average income of IDR 20,000,000. Most choose Law I and II study programmes, but the Apriori rule is not strong enough.

Cluster 2:

South Tambun sub-district has 100 students, North Bekasi 80, East Bekasi 55, and Babelan 55, with an average income of IDR 5,000,000 - IDR 10,000,000. Most chose Communication Science I and II study programmes, with a strong Apriori rule and strong correlation

5 Conclusion

The results of the analysis show several significant points. Firstly, the marketing agency can use the sub-district clustering based on parents' income for a more effective promotion strategy. Second, the clustering results produced three groups: C0 ‘Already stable,’ C1 ‘Focus on promotion strategy,’ and C2 ‘Maximise promotion again.’ Third, the DBI evaluation showed that the best number of clusters was three, although there was an option with five clusters. Fourthly, the a priori algorithm was used to generate association rules from student admission data, which showed the pattern of student interest in study programmes. Fifth, Cluster 0 and Cluster 2 show a correlation between the a priori rules and k-means segmentation, indicating a certain preference for choosing study programmes based on the income of each group.

BIBLIOGRAPHY

[1] P.-Y. Dolbec, “Digital marketing strategy,” *Digit. Mark. Strateg.*, pp. 1–245, 2021, doi: 10.4337/9781035311316.

[2] Kemendikbud, “Statistik Pendidikan Tinggi (Higer Education Statistic) 2020,” *PDDikti Kemendikbud*, vol. 5, pp. 81–85, 2020, [Online]. Available: https://pddikti.kemdikbud.go.id/publikasi

[3] A. Abriyanto and N. Damastuti, “Segmentasi Mahasiswa Dengan ‘Unsupervised’ Algoritma Guna Membangun Strategi Marketing Penerimaan Mahasiswa,” *Insa. Comtech Inf. Sci. Comput. Technol. J.*, vol. 4, no. 2, pp. 10–18, 2019, [Online]. Available: http://ejournal.unira.ac.id/index.php/insand_comtech/article/view/677

[4] E. H. Pujiarini, “Analisis Asosiasi Untuk Menentukan Strategi Promosi Perguruan Tinggi Dengan Algoritma Apriori,” *J. Inform. dan Komput.*, vol. 4, no. 1, pp. 45–51, 2019.

[5] O. Maimon and L. Rokach, *Data Mining and Knowledge Discovery Handbook*, Second. 2010.

[6] D. Sugianti, H. Wim Hapsoro, and W. Setianto, “Penerapan Machine Learning Untuk Penentuan Segmentasi Mahasiswa Baru Dengan Metode K Modes,” *IC-Tech*, vol. 17, no. 1, pp. 1–9, 2022, doi: 10.47775/ictech.v17i1.227.

[7] N. A. Rahmalinda and A. Jananto, “Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru,” *J. Tekno Kompak*, vol. 16, no. 2, pp. 163–175, 2022.

[8] H. Hairani, D. Susilowati, I. Puji Lestari, K. Marzuki, and L. Z. A. Mardedi, “Segmentasi Lokasi Promosi Penerimaan Mahasiswa Baru Menggunakan Metode RFM dan K-Means Clustering,” *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 275–282, 2022, doi: 10.30812/matrik.v21i2.1542.

[9] Y. Wicaksono, U. Nendra Pratama, S. Nurhasanah, T. Utari Ramadania, and W. Juslan, “Pemetaan Strategi Promosi Penerimaan Mahasiswa Baru Menggunakan K-Means,” *Syst. Inf. Syst. Informatics J.*, vol. 7, no. 1, pp. 42–48, 2021, doi: 10.29080/systemic.v7i1.1268.

[10] F. N. Nasari, “Algoritma A Priori Dalam Pengelompokkan Data Pendaftaran Mahasiswa Baru,” *J. Sains dan Ilmu Terap.*, vol. 4, no. 1, pp. 40–45, 2021, doi: 10.59061/jsit.v4i1.102.

[11] I. P. Astuti, “Algoritma Apriori Untuk Menemukan Hubungan Antara Jurusan Sekolah Dengan Tingkat Kelulusan Mahasiswa,” *J. Tek. Inform.*, vol. 12, no. 1, pp. 69–78, 2019, doi: 10.15408/jti.v12i1.10525.

[12] H. Kusumo, E. Sediyono, and M. Marwata, “Analisis Algoritma Apriori untuk Mendukung Strategi Promosi Perguruan Tinggi,” *Walisongo J. Inf. Technol.*, vol. 1, no. 1, p. 49, 2019, doi: 10.21580/wjit.2019.1.1.4000.