Journal of Science and Applied Engineering (JSAE)
Vol. 7, No. 2, October 2024
ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/jsae.v7i2.6416
(Page.8-17)

RECOMMENDATION SYSTEM FOR DETERMINING MARKETING LOCATION FOR NEW STUDENTS BASED ON K-MEAN CLUSTERING

D. M. Febrianto[1,*], F. Marisa[2], S. W. Irianandа[3]

[1, 2, 3]Department of Informatics Engineering, Universitas Widya Gama Malang, Malang, Indonesia
**Email:dfebrianto37@gmail.com*

Submitted : 21 August 2024; *Revision :* 21 October 2024; *Accepted :* 22 October 2024

ABSTRACT

Data on new student admissions is often not utilized, especially at STIKES Widyagama Husada Malang. In fact, this data could be very useful for developing marketing strategies, especially in determining marketing locations for the potential to increase the number of prospective new students in a more targeted and efficient manner. In this research, the implementation of the K-Means Clustering algorithm was chosen to solve the problem of the data mining process for new student admissions. The data in this study uses data on new student admissions for 2020–2023, totaling 375 registrant data with attributes such as gender, age, intended study program, school of origin, major of origin, wave of registration, total registration history, and selected registration information. 3 clusters were produced with dominance in the 2nd cluster; the highest number of new student admissions was 103 registrants, with the highest regional percentage of 57%, namely Kab. Malang (57%), followed by Kota Malang (18%). Meanwhile, the most popular study program is S1 Ilmu Keperawatan (48%). Based on this research, a pattern was produced that can be used as a source of new information for higher education institutions, which can be used to support location decisions in an efficient and targeted marketing strategy for new students, namely in Kab. Malang, so universities can promote the most popular study program, namely the Bachelor of Science in Nursing study program, as an effort to increase the number of prospective new student applicants in the coming year.

Keywords K-Means Clustering; New Student Marketing; Marketing Location; Marketing Recommendation System;
Paper type Research paper

INTRODUCTION

In the context of higher education marketing, marketing strategy in determining location is one of the key factors influencing new student admissions [1]. Determining the right target location can be a competitive advantage, the rapidly changing dynamics of the higher education market require them to continuously monitor industry trends, evaluate competitors, and adjust their marketing strategies dynamically. This requires strategies that are more effective and responsive to changes in trends, technology, and market needs [2].

Data mining is the process of finding interesting patterns or information in data sets using certain methods, techniques or algorithms [3]. In addition, new knowledge gathered from old data is used for future decisions and decisions [4]. Clustering is grouping a number of objects or data into groups, so that each group contains data that may be somewhat similar to and different from other groups [5].

In previous research, the K-means Clustering Algorithm was used in selecting promotional locations for the nutrive benecol cholesterol-lowering drink, successfully classifying data with an accuracy percentage of 95% [6]. Other research, the K-means Clustering Algorithm is used in Determining Promotion Locations for New Student Admissions with the results of the promotion location determination system consisting of two groups: feasible (C1) and not feasible (C2). In this K-Means Clustering Method, there are 30 school data that must be initialized starting from the highest data with value 1, then the next data 2, 3 and so on, obtained from several criteria, then making the data into a Cluster [7].

The current problem is that institutions experience difficulties in making decisions about determining the best areas for marketing new students, due to the lack of reference information -

information regarding the calculation of criteria and priorities which have an important influence. Application and results of measuring criteria using the highest priority K-Mean Clustering algorithm on new student admissions data so that it can be used to support decisions on determining marketing locations that are right on target [8], [9].

By using the K means Clustering algorithm in this research, it is hoped that it can make a significant contribution to the marketing field in improving accuracy, efficiency and faster and more accurate decision making in marketing strategies, as well as helping to provide information in shortening the time for determining marketing locations, thereby providing plus points for competing with other competitors and increasing the number of prospective new students [10].

METHOD

This research includes several steps that will be carried out systematically to achieve the goal, The Knowledge Discovery in Databases (KDD) research method uses the process of analyzing data from databases to find new patterns, useful and easy to understand [11]. The processes in data analysis are, selection, pre-processing, transformation, data exploration, and evaluation [12]. In Figure 2.1, the following is a picture of the research method flow:

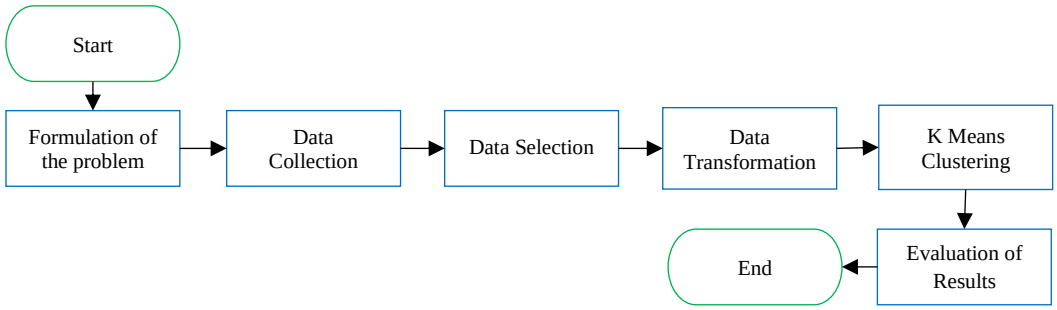

Figure 1. Research methods

Formulation of the problem

Data on new student admissions which continues to increase is the main problem in this research. This allows higher education institutions, especially STIKES Widyagama Husada Malang, to use and obtain new information to support decisions on determining marketing locations in building effective marketing strategies.

Data Collection

This stage is an important stage in this research because it involves the process of collecting data needed to implement a decision support system [13]. Data collection was carried out by requesting previous data from the Marketing STIKES Widygama Husada Malang based on applicable procedures. The data used is data from new students from the 2020/2021, 2021/2022 and 2022/2023 classes.

Data Selection

The data used is 375, At this stage we will only select attributes that are deemed necessary [14].

Pre-Processing

Correcting or deleting data with inconsistent or incomplete formats to ensure data quality that meets requirements [15].

ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/jsae.v7i2.6416
(Page.8-17)

TABLE I. PRE PROCESSING DATA USED

Attributes	Kota Surabaya	Kab. Malang	…	Kab. Nias
Jenis Kelamin	Laki – laki	Perempuan	…	Perempuan
Usia	20 Tahun	19 Tahun	…	20 Tahun
Program studi yang dituju	S1 Keperawatan	S1 Kesling	…	S1 Keperawatan
Asal Sekolah	SMA	SMK	…	SMA
Jurusan Asal	IPA	Kesehatan	…	IPA
Gelombang masuk pendaftaran	3	1	…	2
Jumlah Riwayat pendaftar dari daerah asal	6	57	…	1
Informasi pendaftaran	Sosial Media	Mahasiswa (Kating)	…	Website

In Table I is the data used, with 8 attributes, jenis kelamin, usia, program studi yang dituju, asal sekolah, jurusan asal, masuk gelombang pendafatran, jumlah riwayat pendaftaran dan riwayat informasi pendaftaran. For data processing and analysis needs, not all attributes obtained in data collection will be used.

Data Transformation

In this process, data objects that do not yet have clear entities will be transformed and converted, according to needs, for example into numeric types so that data mining processing becomes easier [16].

TABLE II. ATTRIBUTE DATA TRANSFORMATION JENIS KELAMIN

Jenis Kelamin	Frekuensi	Transformasi
Perempuan	299	1
Laki - laki	76	2

In Table II is the result of the transformation process on the jenis kelamin type attribute with changes to each attribute that has been initialized based on the frequency of the highest value from numbers 1 to 2. Number 1 is female, Number 2 is male.

TABLE III. ATTRIBUTE DATA TRANSFORMATION PROGRAM STUDI YANG DITUJU

Program Studi yang Dituju	Frekuensi	Transformasi
S1 Keperawatan	205	1
S1 Kesehatan Lingkungan	90	2
D3 Kebidanan	80	3

In Table III the results of the program studi attribute transformation process with changes to the initialized attributes from number 1 to number 3. For the study program aimed at S1 Keperawatan, it was changed to number 1. For the study program aimed at S1 Kesehatan Lingkungan, it was changed to number 2. Meanwhile for the study program aimed at D3 Kebidanan, it has been changed to number 3.

TABLE IV. ATTRIBUTE DATA TRANSFORMATION ASAL SEKOLAH

Asal Sekolah	Frekuensi	Transformasi
SMA	206	1
SMK	133	2
MA	30	3
ALJ	6	4

In Table IV above, after the data transformation process, the results of changes to the asal sekolah attribute are obtained. There are various levels of secondary school and equivalent, namely SMA, SMK, MA and ALJ with each attribute initialized from number 1 to number 4.

TABLE V. ATTRIBUTE DATA TRANSFORMATION JURUSAN ASAL

Jurusan Asal	Frekuensi	Transformasi
IPA	155	1
Kesehatan	97	2
IPS	73	3
Bahasa	10	4
Akutansi	8	5
TKJ	7	6
Farmasi	7	6
Agana	5	7
Bisnis	4	8
Lainya	2	9
Teknik	2	10

Table V shows the results of the transformation process on the jurusan asal attribute with changes to each attribute initialized from number 1 to number 10.

TABLE VI. ATTRIBUTE DATA TRANSFORMATION GELOMBANG PENDAFTARAN

Gelombang Pendaftaran	Frekuensi	Transformasi
Gelombang II	149	1
Gelombang III	115	2
Gelombang I	104	3
Gelombang IV	4	4

In Table VI above are the results of the gelombang pendaftaran attribute transformation with each attribute initialized with numbers 1 to 4.

TABLE VII. ATTRIBUTE DATA TRANSFORMATION GELOMBANG PENDAFTARAN

Informasi Pendafataran	Frekuensi	Transformasi
Kampus/Offline	59	1
Sosial Media	58	2
Website	55	3
Mahasiswa (Kakak Tingkat)	54	4
Alumni	36	5
Guru BK	35	6
Brosur	22	7
Dosen	18	8
Spanduk	13	9
Keluarga/Saudara	12	10
Lainnya	8	11
Pegawai Kampus	4	12

In Table VII above, after the data transformation process, the results of changes to the informasi pendaftaran attribute are obtained. There are various reasons obtained from the registration data history. With each attribute that has been initialized from number 1 to number 12.

Clustering K-Means Calculation

K-Means Clustering is an unsupervised data mining technique or data analyzer that carries out a modeling process without supervision, namely an algorithm that works repeatedly (iteratively) from each data point to K groups that have been created to determine the groups in the data set [17]. in figure 1.2 below are the stages of the k means algorithm.

ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/jsae.v7i2.6416
(Page.8-17)

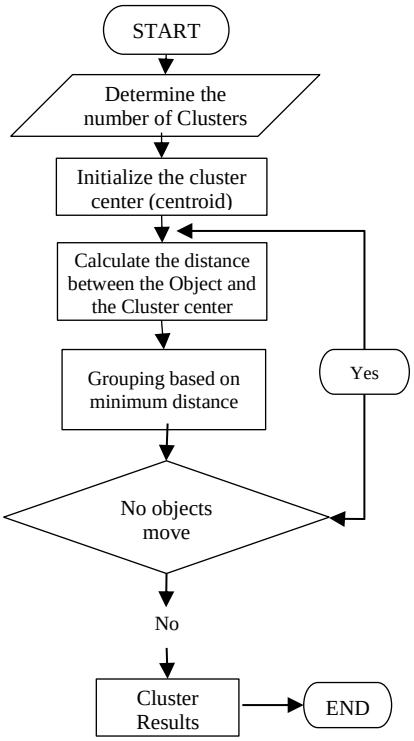

Figure 2. Stages of the K-Means Algorithm [18]

DISCUSSION

K-Means Clustering

The following are the stages in carrying out analysis using the k-means algorithm [19].

Determining the Number of Clusters

In this research, the K-Means clustering algorithm will first determine the number of clusters used, namely 3 clusters.

Determining the Cluster Center (Centroid)

At this stage, we will determine the cluster center value from the data used. The centroid criteria and the code for each attribute used are as follows:

1. Jenis Kelamin (JK)
2. Usia (U)
3. Program Studi Tujuan (PS)
4. Asal Sekolah (AS)
5. Jurusan Asal (JA)
6. Gelombang Masuk Pendaftaran (GP)
7. Jumlah Riwayat Pendaftar dari Daerah Asal (JR)
8. Informasi Pendaftaran

In Table VIII, the following is the centroid obtained randomly from the Student Admissions data that will be used:

TABLE VIII. CENTROID RESULTS OBTAINED RANDOMLY

Baris Data ke-	JK	U	PS	AS	JA	GP	JR	IP
1	1	16	1	2	1	1	1	3
50	1	19	2	3	7	2	57	7
180	1	19	3	3	3	3	9	7

Calculation of Data Distance with Centroid

in table IX of the following Pre-Processing Results, the distance from each data to each existing cluster center will be calculated using the Euclidean distance formula. So that the closest distance from each data to the centroid is found.

TABLE IX. DATA PRE PROCESSING RESULTS

Name of Region of Origin	JK	U	PS	AS	JA	GP	JR	IP
Kab. Bireuen	1	16,42	1	2	1	1	1	3
Kab. Buleleng	1	17,50	2	3	3	2	1	3
Kota Denpasar	1	17,67	1	2	2	3	1	2
Kota Pangkalpinang	1	23,75	3	4	2	2	1	10
Up to data 375	…	…	…	…	…	…	…	…
Kab. Nias	1	20,25	1	1	1	1	1	3

The following is a calculation of the distance between data and the centroid using the Euclidean distance formula [20].

$$dij = \sqrt{(x_{1i} - x_{1j})^2 + (x_{2i} - x_{2j})^2 + \cdots + (x_{ki} - x_{kj})^2} \quad (1)$$

Information:

dij = The distance between 1 cluster j data centers
x_{ki} = Data 1 in data attribute k
x_{kj} = Center point j to attribute k

Distance calculation for area 1 data with centroids 1, 2, and 3:

$C1ij = \sqrt{(1\text{-}1)^2+(16{,}42\text{-}16)^2+(1\text{-}1)^2+(2\text{-}2)^2+(1\text{-}1)^2+(1\text{-}1)+(1\text{-}1)^2+(3\text{-}3)^2} = 0.417$

$C2ij = \sqrt{(1\text{-}1)^2+(16{,}42\text{-}19)^2+(1\text{-}2)^2+(2\text{-}3)^2+(1\text{-}7)^2+(1\text{-}2)+(1\text{-}57)^2+(3\text{-}7)^2} = 56.548$

$C3ij = \sqrt{(1\text{-}1)^2+(16{,}42\text{-}19)^2+(1\text{-}3)^2+(2\text{-}3)^2 +(1\text{-}3)^2+(1\text{-}3)+(1\text{-}9)^2+(3\text{-}7)^2} = 9.984$

The calculation continues until the 375th data. The results of calculating the distance between registrant data in the 1st iteration of the calculation can be seen in Table X below:

TABLE X. CLUSTER RESULTS IN THE 1ST ITERATION

Name of Region of Origin	C1	C2	C3	MINIMUM	CLUSTER
Kab. Bireuen	0,417	56,548	9,984	0,417	1
Kab. Buleleng	3,041	56,305	9,179	3,041	1
Kota Denpasar	2,963	56,487	9,838	2,963	1
Kota Pangkalpinang	10,912	56,520	9,928	9,928	3
Up to data 375	…	…	…	…	…
Kab. Nias	4,366	56,529	9,877	4,366	1

Centroid calculation in the 2nd iteration. The data occupying each cluster will be added up and the average value calculated which will become the new centroid in table XI below:

ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/jsae.v7i2.6416
(Page.8-17)

TABLE XI. NEW CENTROID RESULTS IN THE 1ST ITERATION

Jumlah Anggota Iterasi ke-1	JK	U	PS	AS	JA	GP	JR	IP
C1 = 130 Anggota	1,234	18,632	1,817	1,330	2,193	1,797	7,279	3,680
C2 = 103 Anggota	1,184	18,722	1,417	1,777	2,583	2,019	52,087	4,631
C3 = 142 Anggota	1,268	18,802	1,761	1,556	2,373	1,965	13,845	4,908

The new centroid in table XI above will calculate the distance from each data to each existing cluster center using the Euclidean distance formula in the 2nd iteration and subsequent iterations, then the iteration will be stopped if the cluster position does not change position. In this study, the cluster position did not change position in the 5th iteration so the iteration was stopped. In Figure 3 below are the results of cluster members at every iteration.

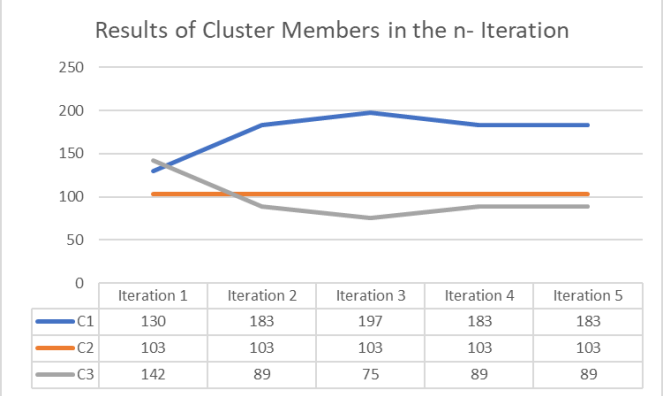

Figure 3. Results of Cluster Members in Each Iteration

Figure 3 shows the results of the analysis of the number of cluster members, namely the dominance of cluster C1 with 183 members, followed by C2 with 103 members and C3 with 89 members. Then an attribute analysis of each cluster was carried out. Table 1.12 shows the results of the overall cluster analysis.

TABLE XII. OVERALL CLUSTER RESULTS

Attribute	Parameter	Cluster 1 has 183 members	Cluster 2 has 103 members	Cluster 3 has 89 members
Nama Daerah	Kota/Kab.	Kab. Sumba Timur = 9 Kota Batu = 9 Kab. Pamekasan = 9 Kota Surabaya = 6 Kab. Blitar = 5 Kab. Blitar = 5 Kab. Sumenep = 5 Kota Probolinggo = 5 Kota Tual = 5 Kota dengan Nilai <5 = 83	Kab. Malang = 57 Kota Malang = 46	Kab. Sumba Barat Daya = 21 Kab. Probolinggo = 19 Kab. Sumba Barat = 18 Kab. Sumba Tengah = 17 Kab. Pasuruan = 14
Jenis Kelamin	Perempuan =	157	84	58
	Laki – laki =	26	19	31
Usia	17<=19	98	58	65
	19>21	65	40	19
	21>=25	20	5	5
Program Studi Tujuan	S1 Ilmu Keperawatan	90	72	0
	S1 Kesehatan Lingkungan	40	19	0
	D3 Kebidanan	53	12	0
Asal Sekolah	SMA	115	32	59
	SMK	44	63	26
	MA	19	7	4
	ALJ	5	1	0
Jurusan Asal	IPA	88	26	41
	Kesehatan	32	49	16
	IPS	48	10	15

ISSN Print : 2621-3745
ISSN Online : 2621-3753
(Page.8-17)
Journal of Science and Applied Engineering (JSAE)
Vol. 7, No. 2, October 2024
DOI : 10.31328/jsae.v7i2.6416

TABLE XII. OVERALL CLUSTER RESULTS

Attribute	Parameter	Cluster 1 has 183 members	Cluster 2 has 103 members	Cluster 3 has 89 members
	Bahasa	3	2	5
	Akutansi	2	4	2
	TKJ	1	5	1
	Farmasi	2	2	3
	Agama	2	1	2
	Bisnis	2	1	1
	Teknik	1	2	1
	Lainya	2	1	2
Gelombang masuk pendaftaran	Gel. I =	48	32	24
	Gel. II=	74	37	39
	Gel. III =	58	31	25
	Gel. IV =	3	3	1
Jumlah Riwayat pendaftar	1>=10	183	0	0
	11>=20	0	0	68
	21>=30	0	0	21
	31>=60	0	103	0
Informasi pendaftaran	Kampus	31	16	12
	Sosial Media	26	15	17
	Website	37	11	7
	Mahasiswa (Kating)	24	10	20
	Alumni	15	13	8
	Guru BK	8	16	11
	Brosur	13	5	4
	Dosen	10	6	2
	Spanduk	6	4	3
	Keluarga/Saudara	6	3	3
	Lainnya	6	0	2
	Pegawai Kampus	1	4	0

Evaluate Results

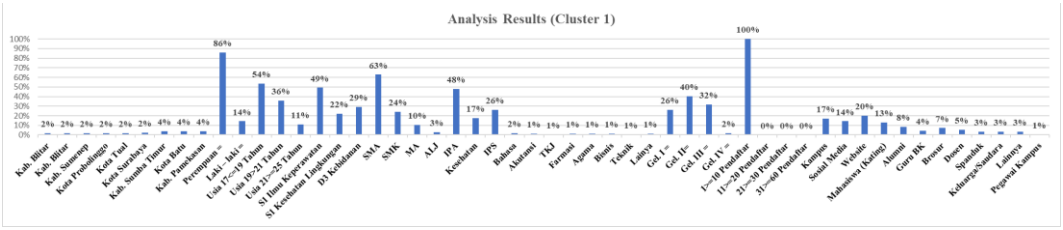

Figure 4. Analysis Visualization Results (Cluster 1)

Figure 4 shows the results of the analysis of the first cluster, with 183 members dominated by registrants from the district area. East Sumba regency, Batu City and Pamekasan regency 3.6%, Number of historical registrants between 1 and 10 registrants. With 85.8% female gender, 53.6% aged between 17 and 19 years, 62.8% from high school (SMA), 48% from science majors (IPA), 40% entered the 2nd wave and 31.7% from the 3rd wave, via website registration information 20%. Meanwhile, the study program in demand is S1 Ilmu Keperawatan at 49%.

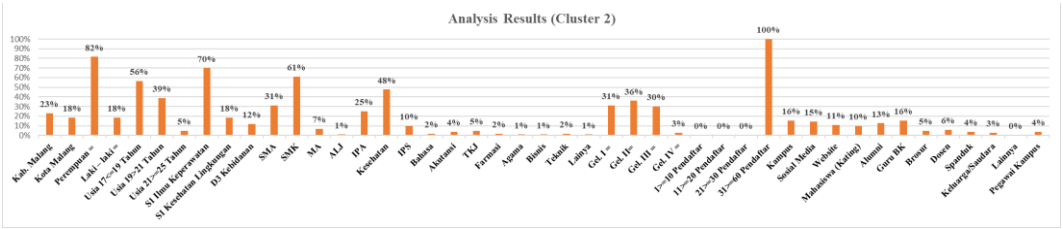

Figure 5. Analysis Visualization Results (Cluster 2)

Figure 5 shows the results of the analysis from the second cluster, with 103 members dominated by registrants from the district area. Malang regency 57% and Malang City 18%. The number of historical registrants is between 10 and 20 registrants. With 65% female gender, 73% aged between 17 and 19 years, 66% from high school (SMA), 46% from science majors (IPA), 44% entered in the

ISSN Print : 2621-3745
ISSN Online : 2621-3753
DOI : 10.31328/jsae.v7i2.6416
(Page.8-17)

2nd wave, 22% through senior student registration information and 19% from social media. %. Meanwhile, the study program of interest is S1 Ilmu Keperawatan 70%.

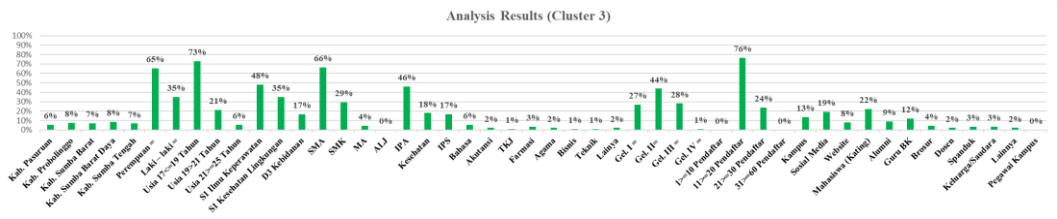

Figure 6. Analysis Visualization Results (Cluster 3)

Figure 6 shows the results of the analysis of the third cluster, with 89 members dominated by registrants from the district area. Probolinggo regency and Southwest Sumba regency 8%, West Sumba regency and Central Sumba regency 7%, and Pasuruan regency 6%. The number of historical registrants is between 31 and 60 registrants. With 82% female gender, 58% aged between 17 and 19 years, 62.8% from vocational school (SMK), 48% majoring in Health (Kesehatan), 36% entered in the 2nd wave and 31.7% in the 1st wave, via registration information came directly to campus 20% and guidance and counseling teachers 16%. Meanwhile, the study programs in demand are S1 Ilmu Keperawatan 48% and S1 Kesehatan Lingkungan 35%.

By using the results of data processing produced by combining clustering techniques with the k-means algorithm mentioned above, patterns are formed that can be studied regarding strategies for determining marketing locations for new students. These patterns will help in future decision making, for example, how to promote each study program in each school that is most in demand from each Cluster.

CONCLUSION

Based on the results of research that has been carried out, 3 clusters were formed with the dominance of the 2nd cluster with the highest number of new student admissions at 103 registrants, with the highest regional percentage of 57%, namely Kab. Malang 57% followed by Kota Malang 18%. Meanwhile the most popular study program is S1 Ilmu Keperawatan 48%. The decision to determine location in the marketing strategy for new students that can be done is by following the clusters that have been formed based on the school's area of origin, namely in the Kab. Malang and the study program that is most in demand is the S1 Ilmu Keperawatan Study Program. Promotions can also be carried out to optimize study programs that have few enthusiasts. So this step provides a great opportunity to increase the number of new student registrants at STIKES Widyagama Husada Malang.

REFERENCES

[1] H. Alzoubi, M. Alshurideh, B. Al Kurdi, I. Akour, and R. Aziz, 'Does BLE technology contribute towards improving marketing strategies, customers' satisfaction and loyalty? The role of open innovation', International Journal of Data and Network Science, vol. 6, no. 2, pp. 449–460, Mar. 2022, doi: 10.5267/j.ijdns.2021.12.009.

[2] J. Z. Zhang and G. F. Watson IV, 'Marketing ecosystem: An outside-in view for sustainable advantage', Industrial Marketing Management, vol. 88, pp. 287–304, Jul. 2020, doi: 10.1016/j.indmarman.2020.04.023.

[3] M. H. Santoso, 'Application of Association Rule Method Using Apriori Algorithm to Find Sales Patterns Case Study of Indomaret Tanjung Anom', Brilliance: Research of Artificial Intelligence, vol. 1, no. 2, pp. 54–66, Dec. 2021, doi: 10.47709/brilliance.v1i2.1228.

[4] I. H. Sarker, 'Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective', Sep. 01, 2021, Springer. doi: 10.1007/s42979-021-00765-8.

[5] M. Piernik and T. Morzy, 'A study on using data clustering for feature extraction to improve the quality of classification', Knowl Inf Syst, vol. 63, no. 7, pp. 1771–1805, Jul. 2021, doi: 10.1007/s10115-021-01572-6.

[6] J. T. Hutabarat, 'Penerapan Algoritma K-Means Untuk Menentukan Lokasi Promosi Produk Minuman Penurun Kolestrol Nutrive Benecol Pada PT. Perkasa/Kalbe Nutrritionals', Medan, Oct. 2018.

[7] N. F. Azmi and N. Sunaryo, 'Perancangan Sistem Pendukung Keputusan Pemilihan Mahasiswa Berprestasi Dengan Metode AHP Berbasis Web Pada Politeknik LP3I Kampus Padang', Jurnal Sains dan Teknologi Informatika, vol. 1, no. 2, pp. 80–89, 2023.
[8] B. Sutara, F. Vulture, and R. Novianti, 'Application of K-Means algorithm with CRISP-DM method in student data analysis as a support for promotion strategy SIDE: Scientifict Development Journal', 2024. [Online]. Available: https://ojs.arbain.co.id/index.php/side/index
[9] T. R. Kurniawan, E. Chumaidiyah, and L. Andrawina, 'Application of the k-means clustering method and simple linear regression to new student admissions as a promotion method', JURNAL INFOTEL, vol. 15, no. 1, pp. 25–33, Feb. 2023, doi: 10.20895/infotel.v15i1.860.
[10] P. Anitha and M. M. Patil, 'RFM model for customer purchase behavior using K-Means algorithm', Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 5, pp. 1785–1792, May 2022, doi: 10.1016/j.jksuci.2019.12.011.
[11] S. Gadal, R. Mokhtar, M. Abdelhaq, R. Alsaqour, E. S. Ali, and R. Saeed, 'Machine Learning-Based Anomaly Detection Using K-Mean Array and Sequential Minimal Optimization', Electronics (Switzerland), vol. 11, no. 14, Jul. 2022, doi: 10.3390/electronics11142158.
[12] P. Bala, H. Tyagi, R. Vaishnav, S. Tiwari, and S. Kumar, 'Citation: Dr Preeti Bala (2024), Data Pre-Processing And Its Implications In Data Mining Educational Administration: Theory and Practice', vol. 2024, no. 5, pp. 9861–981, 2024, doi: 10.53555/kuey.v30i5.4669.
[13] B. Vasey et al., 'Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI', The BMJ, 2022, doi: 10.1136/bmj-2022-070904.
[14] K. P. Sinaga and M. S. Yang, 'Unsupervised K-means clustering algorithm', IEEE Access, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.
[15] K. Maharana, S. Mondal, and B. Nemade, 'A review: Data pre-processing and data augmentation techniques', Global Transitions Proceedings, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.
[16] K. Tabianan, S. Velu, and V. Ravi, 'K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data', Sustainability (Switzerland), vol. 14, no. 12, Jun. 2022, doi: 10.3390/su14127243.
[17] R. Vankayalapati, K. B. Ghutugade, R. Vannapuram, and B. P. S. Prasanna, 'K-means algorithm for clustering of learners performance levels using machine learning techniques', Revue d'Intelligence Artificielle, vol. 35, no. 1, pp. 99–104, Feb. 2021, doi: 10.18280/ria.350112.
[18] E. Umargono, J. E. Suseno, and V. Gunawan, 'K-Means Clustering Optimization Using the Elbow Method and Early Centroid Determination Based on Mean and Median Formula', 2020.
[19] G. Nalli, D. Amendola, A. Perali, and L. Mostarda, 'Comparative analysis of clustering algorithms and moodle plugin for creation of student heterogeneous groups in online university courses', Applied Sciences (Switzerland), vol. 11, no. 13, Jul. 2021, doi: 10.3390/app11135800.
[20] M. Faisal, E. M. Zamzami, and Sutarman, 'Comparative Analysis of Inter-Centroid K-Means Performance using Euclidean Distance, Canberra Distance and Manhattan Distance', in Journal of Physics: Conference Series, Institute of Physics Publishing, Jul. 2020. doi: 10.1088/1742-6596/1566/1/012112.