



Penerapan *K-Means Clustering* untuk Menentukan Lokasi Promosi Penerimaan Mahasiswa Baru (Studi Kasus: STMIK Kaputama Kota Binjai)

Ronauli Silaban^{1*}, Achmad Fauzi², Lina Arliana Nur Kadim³

^{1,2,3} STMIK Kaputama Binjai, Indonesia

ronaulisilaban32@gmail.com¹, fauzyrivai88@gmail.com², lina_arliana@yahoo.com³

Alamat: Jl. Veteran No.4A, Tangsi, Kec. Binjai Kota, Kota Binjai, Sumatera Utara 20714

Korespondensi penulis: ronaulisilaban32@gmail.com*

Abstract. *The process of accepting new students generates a lot of data in the form of profiles of students who register. From year to year there is an increase in the number of prospective new students who come from several areas in Binjai City, Langkat Regency and surrounding areas, so the location of the socialization of new student admissions promotions every year is increasing and wider. And from several schools that have been visited and are expected to provide new prospective students, in fact, it is not proportional to the final number of prospective students who register. In this study, applying the K-Means Clustering algorithm using 3 variables namely, region, school origin, major. In determining the location of new student admissions promotions, the promotion team first identifies what factors will influence the determination of promotional locations ranging from region, school origin and majors that are considered to be set as promotional locations. Based on the results of grouping new student admission data of STMIK Kaputama Binjai using the K-means Clustering method from 20 data that has been processed, 3 clusters and 3 iterations are produced where cluster 1 has 9 data, cluster 2 has 2 data and cluster 3 has 9 data.*

Keywords: *Data Mining, Clustering, K-means Algorithm, Promotion*

Abstrak. Proses penerimaan mahasiswa baru menghasilkan banyak data berupa profil mahasiswa yang mendaftar. Dari tahun ke tahun adanya peningkatan jumlah calon mahasiswa baru yang berasal dari beberapa wilayah di Kota Binjai, Kabupaten Langkat dan sekitarnya, maka lokasi sosialisasi promosi penerimaan mahasiswa baru setiap tahun semakin meningkat dan luas. Dan dari beberapa sekolah yang sudah dikunjungi dan diharapkan memberikan calon mahasiswa baru nyatanya tidak sebanding dengan jumlah akhir calon mahasiswa yang mendaftar. Dalam penelitian ini, menerapkan algoritma K-Means Clustering dengan menggunakan 3 variable yakni, wilayah, asal sekolah, jurusan. Dalam menetapkan lokasi promosi penerimaan mahasiswa baru, tim promosi terlebih dahulu mengidentifikasi faktor apa saja yang akan mempengaruhi dalam penentuan lokasi promosi mulai dari wilayah, asal sekolah dan jurusan yang menjadi pertimbangan untuk ditetapkan menjadi lokasi promosi. Berdasarkan hasil pengelompokan data penerimaan mahasiswa baru STMIK Kaputama Binjai menggunakan metode K-means Clustering dari 20 data yang telah diolah maka dihasilkan 3 cluster dan 3 iterasi dimana cluster 1 terdapat 9 data, cluster 2 terdapat 2 data dan cluster 3 terdapat 9 data.

Kata kunci: Data Mining, Clustering, Algoritma K-means, Promosi

1. LATAR BELAKANG

STMIK Kaputama Kota Binjai merupakan salah satu institusi pendidikan di Kota Binjai yang setiap tahun ajaran baru melakukan penerimaan mahasiswa baru. Calon mahasiswa yang mendaftar di STMIK Kaputama Kota Binjai berasal dari Sekolah Menengah Atas (SMA) dan Sekolah Menengah Kejuruan (SMK) yang ada di wilayah Kota Binjai, Kabupaten Langkat dan sekitarnya. Proses penerimaan mahasiswa baru menghasilkan banyak data berupa profil mahasiswa yang mendaftar. Dari tahun ke tahun adanya peningkatan jumlah calon mahasiswa

baru yang berasal dari beberapa wilayah di Kota Binjai, Kabupaten Langkat dan sekitarnya, maka lokasi sosialisasi promosi penerimaan mahasiswa baru setiap tahun semakin meningkat dan luas. Dan dari beberapa sekolah yang sudah dikunjungi dan diharapkan memberikan calon mahasiswa baru nyatanya tidak sebanding dengan jumlah akhir calon mahasiswa yang mendaftar. Hal tersebut mengakibatkan tim promosi penerimaan mahasiswa baru memerlukan pengelompokan lokasi mana yang menjadi prioritas dan efektif untuk mendapatkan calon mahasiswa baru berdasarkan data penerimaan mahasiswa baru di tahun sebelumnya sehingga didapatkan calon mahasiswa baru dengan jumlah yang lebih maksimal dari tahun sebelumnya. Dalam penelitian ini, menerapkan algoritma *K-Means Clustering* dengan menggunakan 3 variable yakni, wilayah, asal sekolah, jurusan.

Adapun penelitian sebelumnya yang dapat mendukung dan diharapkan dapat membantu penelitian ini antara lain: Penelitian yang dilakukan oleh (Widiastuti & Azzat, 2022) menggunakan algoritma *K-means* data penerimaan mahasiswa baru menghasilkan pemetaan wilayah strategis untuk dilakukan promosi. Penelitian yang dilakukan oleh (Azmi et al., 2022) menggunakan metode *K-means* menentukan lokasi promosi peserta didik yang mendapatkan hasil lebih objektif dan efisien.

Adapun tujuan penelitian ini adalah untuk menentukan lokasi promosi prioritas penerimaan mahasiswa baru di STMIK Kaputama Kota Binjai dengan algoritma *K-means clustering*, untuk mengimplementasikan data mining dengan metode *Clustering* dalam pengolahan data penerimaan mahasiswa baru menjadi informasi yang dibutuhkan. Dan manfaat dilakukannya penelitian adalah dapat memberikan informasi tentang lokasi promosi prioritas penerimaan mahasiswa baru di STMIK Kaputama Kota Binjai dengan acuan data tahun 2022 sampai 2023, mengaplikasikan metode *Clustering* dengan algoritma *K-means* dalam pengelompokan lokasi promosi penerimaan mahasiswa baru pada STMIK Kaputama Kota Binjai dengan bantuan program Matlab.

2. KAJIAN TEORITIS

Data mining merupakan suatu proses menggali informasi yang berarti dan tersembunyi pada kumpulan data yang berjumlah sangat banyak. *Data mining* merupakan salah satu bagian atau proses utama dan *Knowledge Discovery in Database* (KDD) yang bentuk kegiatannya yaitu mengumpulkan dan menggunakan data masa lalu untuk menemukan keteraturan, pola atau hubungan dalam suatu set data yang lebih besar (Bulolo, 2020).

Clustering merupakan teknik pengelompokan yang memiliki kemiripan nilai yang sama (homogen). Tujuan dari clustering adalah untuk menemukan struktur yang tersembunyi di dalam data (Reza et al., 2022). *Clustering* merupakan proses pengorganisasian objek-objek kedalam kelompok-kelompok (*cluster*) yang mana kelompok tersebut memiliki karakteristik yang mirip. Ada dua metode clustering yaitu *Hierarchical Clustering* dan *Non-Hierarchical Clustering* (Mirantika et al., 2021).

Algoritma *K-means* merupakan salah satu algoritma unsupervised learning (menganalisis dan memahami data yang tidak berlabel) yang sangat populer. Algoritma ini digunakan untuk mengelompokkan data (*clustering*) dengan cara membagi data kedalam k kelompok (*cluster*) yang berbeda, dimana k ditentukan terlebih dahulu sebelum proses clustering dimulai.

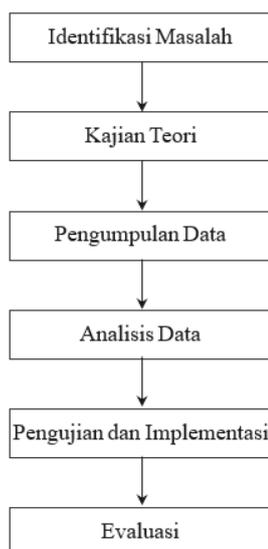
Tahapan penyelesaian algoritma *K-means Clustering* adalah sebagai berikut :

1. Misalkan diberikan matriks $X = \{X_{ij}\}$, data berukuran $n \times p$ dengan $i = 1, 2, \dots, n, j = 1, 2, \dots, p$ dan asumsikan jumlah klaster awal K .
2. Tentukan centroid.
3. Hitung jarak setiap objek ke setiap centroid dengan menggunakan jarak^dEuclidean atau dapat ditulis sebagai berikut

$$J(x_i, c_i) = \sqrt{(x_i - c_i)^2} \dots \dots \dots (1)$$
4. Setiap objek disusun ke centroid terdekat dan kumpulan objek tersebut akan membentuk cluster.
5. Tentukan centroid baru dari cluster yang akan terbentuk, diobjek yang mana centroid baru itu diperoleh dari rata-rata setiap yang terletak pada cluster yang sama.
6. Ulangi langkah 3, jika centroid awal dan baru tidak sama (Reza et al., 2022).

Menurut Kotler (2000:119), Promosi adalah bagian dan proses strategi pemasaran sebagai cara untuk berkomunikasi dengan pasar dengan menggunakan komposisi bauran promosi “promotional mix” (Feriyanto et al., 2019).

3. METODE PENELITIAN



Gambar 1 Kerangka Kerja Penelitian

Berdasarkan gambar diatas dapat disimpulkan bahwa ada beberapa tahapan dalam menyelesaikan penelitian yakni:

1. Identifikasi Masalah, merupakan tahap awal dalam penelitian yang digunakan untuk mengidentifikasi latar belakang masalah, tujuan dan manfaat penelitian. Pada tahap ini dilakukan riset yang bertujuan untuk mendapatkan permasalahan yang dihadapi instansi.
2. Kajian Teori, tahap ini dilakukan kajian teori terhadap masalah yang ada, menentukan konsep yang digunakan dalam penelitian. Kajian teori bertujuan untuk menemukan teori-teori yang terkait dalam penelitian. Teori tersebut dapat berupa jurnal-jurnal pendukung penelitian, buku, e-book dan referensi lainnya.
3. Pengumpulan Data, tahap ini dilakukan pengumpulan data pendukung penelitian yang diperlukan untuk mencapai tujuan penelitian. Data-data diperoleh dari data Penerimaan Mahasiswa Baru STMIK Kaputama tahun 2022-2023.
4. Analisis Data, tahap ini dilakukan analisis data pada hasil pengolahan data pendukung. Analisis data bertujuan untuk menemukan pola, hubungan dan informasi penting yang terdapat dalam data yang diolah.
5. Pengujian dan Implementasi, tahap ini dilakukan pengujian validasi data dan implementasi pengelompokan data mahasiswa baru dengan hasil data yang diolah manual dengan yang diolah menggunakan software pendukung penelitian serta penyusunan program.

6. Evaluasi, merupakan tahap terakhir dalam penelitian. Dalam tahap ini akan dibahas kesimpulan serta saran yang akan diperlukan untuk pengembangan program selanjutnya.

Data input pada sistem berupa data yang didapatkan dari STMIK Kaputama Binjai. Data input sistem telah disimpan pada Ms. Office Excel sebagai penampung data, kemudian data akan ditransformasikan berdasarkan nilai transformasi dari masing-masing variabel yang digunakan. Data pendukung penelitian ini, yang diolah adalah data penerimaan mahasiswa yang berasal dari STMIK Kaputama Binjai.

Tabel 1 Data Penerimaan Mahasiswa Baru

No	Object	Wilayah (X)	Asal Sekolah (Y)	Jurusan (Z)
1	A	Stabat	SMK	TKJ
2	B	Stabat	SMK	TKJ
3	C	Binjai Utara	SMA	IPA
4	D	Kuala	SMK	TKJ
5	E	Stabat	SMA	IPA
6	F	Binjai Barat	SMK	TKRO
7	G	Binjai Kota	SMK	TKJ
8	H	Binjai Kota	SMK	OTKP
9	I	Selesai	SMK	TKJ
10	J	Selesai	SMK	TKJ
11	K	Binjai Utara	SMA	IPA
12	L	Selesai	SMK	TKJ
13	M	Binjai Kota	SMA	IPA
14	N	Binjai Selatan	SMA	IPA
15	O	Binjai Kota	SMA	IPA
16	P	Binjai Timur	SMA	IPS
17	Q	Binjai Timur	SMA	IPA
18	R	Binjai Timur	SMA	IPS
19	S	Kuala	SMA	IPA
20	T	Binjai Selatan	SMA	IPA

Untuk menentukan group dari suatu objek, maka langkah pertama yang dilakukan adalah mengukur jarak Euclidean antara titik objek X, Y, Z yang didefinisikan sebagai berikut:

Tabel 2 Wilayah (X)

Kode	Wilayah (X)
1	Stabat
2	Kuala
3	Selesai
4	Binjai Kota
5	Binjai Barat
6	Binjai Timur
7	Binjai Utara
8	Binjai Selatan

Tabel 3 Asal Sekolah (Y)

Kode	Asal Sekolah (Y)
1	SMA
2	SMK

Tabel 4 Jurusan (Z)

Kode	Jurusan (Z)
1	IPA
2	IPS
3	TKJ
4	TKRO
5	OTKP

Tabel 5 Hasil Transformasi

No.	Object	X	Y	Z
1	A	1	2	3
2	B	2	2	3
3	C	7	1	1
4	D	2	2	3
5	E	1	1	1
6	F	5	2	4
7	G	4	2	3
8	H	4	2	5
9	I	3	2	3
10	J	3	2	3
11	K	7	1	1
12	L	3	2	3
13	M	4	1	1
14	N	8	1	1
15	O	4	1	1
16	P	6	1	2
17	Q	6	1	1

No.	Object	X	Y	Z
18	R	6	1	2
19	S	2	1	1
20	T	8	1	1

Selanjutnya membentuk cluster menjad 3 kelompok ($k=3$) dan menentukan titik pusat centroid secara random. Adapun proses perhitungan clustering adalah sebagai berikut :

$$k = 3$$

Iterasi 1

Centroid 1 = (1, 2, 3), diambil secara acak dari data 1

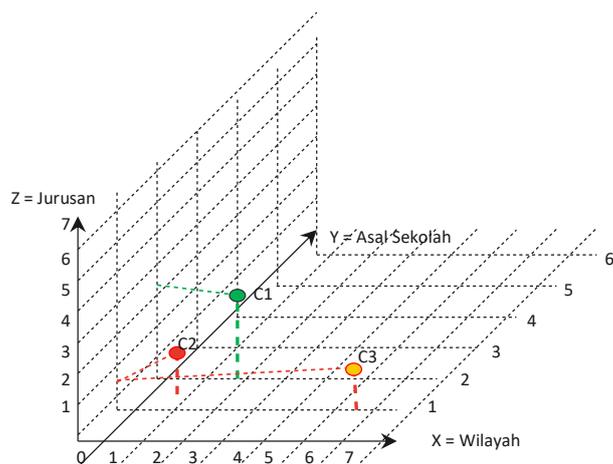
Centroid 2 = (2, 1, 1), diambil secara acak dari data 19

Centroid 3 = (4, 1, 1), diambil secara acak dari data 13

Berdasarkan hasil proses perhitungan dengan metode *Clustering* dengan Algoritma *K-Means*, maka tabel 6 menunjukkan hasil penentuan group, yaitu sebagai berikut:

Tabel 6 Hasil Iterasi 3

No	Object	X	Y	Z	C1	C2	C3	JARAK	GRUP
1	A	1	2	3	2,03	2,29	5,61	2,03	1
2	B	2	2	3	1,05	2,29	4,69	1,05	1
3	C	7	1	1	4,74	5,50	0,81	0,81	3
4	D	2	2	3	1,05	2,29	4,69	1,05	1
5	E	1	1	1	3,23	0,50	5,23	0,50	2
6	F	5	2	4	2,11	4,72	3,20	2,11	1
7	G	4	2	3	1,05	3,35	3,02	1,05	1
8	H	4	2	5	1,94	4,82	4,50	1,94	1
9	I	3	2	3	0,33	2,69	3,81	0,33	1
10	J	3	2	3	0,33	2,69	3,81	0,33	1
11	K	7	1	1	4,74	5,50	0,81	0,81	3
12	L	3	2	3	0,33	2,69	3,81	0,33	1
13	M	4	1	1	2,73	2,50	2,23	2,23	3
14	N	8	1	1	5,61	6,50	1,79	1,79	3
15	O	4	1	1	2,73	2,50	2,23	2,23	3
16	P	6	1	2	3,43	4,61	0,81	0,81	3
17	Q	6	1	1	3,93	4,50	0,31	0,31	3
18	R	6	1	2	3,43	4,61	0,81	0,81	3
19	S	2	1	1	2,73	0,50	4,23	0,50	2
20	T	8	1	1	5,61	6,50	1,79	1,79	3



Gambar 2 Grafik Cluster Berdasarkan perhitungan yang telah dilakukan

4. HASIL DAN PEMBAHASAN

Dari 20 data diperoleh 3 cluster, Cluster 1 terdapat 9 data, Cluster 2 terdapat 2 data dan Cluster 3 terdapat 9 data.

1. Cluster 1 terdapat 9 data

3,00 2,00 3,33

Berdasarkan hasil perhitungan diatas, diketahui bahwasanya Cluster 1 pada kelompok dengan atribut wilayah adalah Selesai, untuk kelompok dengan atribut asal sekolah adalah SMK dan kelompok atribut jurusan adalah TKJ.

2. Cluster 2 terdapat 2 data

1,50 1,00 1,00

Berdasarkan perhitungan diatas, diketahui bahwasanya Cluster 2 pada kelompok dengan atribut wilayah adalah Stabat, untuk kelompok dengan atribut asal sekolah adalah SMA dan kelompok atribut jurusan adalah IPA.

3. Cluster 3 terdapat 9 data

6,22 1,00 1,22

Berdasarkan perhitungan diatas, diketahui bahwasanya Cluster 3 pada kelompok dengan atribut wilayah adalah Binjai Timur, untuk kelompok dengan atribut asal sekolah adalah SMA dan untuk kelompok dengan atribut jurusan adalah IPA.

5. KESIMPULAN DAN SARAN

Dalam menetapkan lokasi promosi penerimaan mahasiswa baru, tim promosi terlebih dahulu mengidentifikasi faktor apa saja yang akan mempengaruhi dalam penentuan lokasi promosi mulai dari wilayah, asal sekolah dan jurusan yang menjadi pertimbangan untuk ditetapkan menjadi lokasi promosi. Berdasarkan hasil pengelompokan data penerimaan mahasiswa baru STMIK Kaputama Binjai menggunakan metode *K-means Clustering* dari 20 data yang telah diolah maka dihasilkan 3 *cluster* dan 3 iterasi dimana *cluster* 1 terdapat 9 data, *cluster* 2 terdapat 2 data dan *cluster* 3 terdapat 9 data. Dengan atribut wilayah, asal sekolah, jurusan pada *cluster* 1 yaitu Selesai- SMK-TKJ, atribut wilayah, asal sekolah, jurusan pada *cluster* 2 yaitu Stabat-SMA-IPA dan atribut wilayah, asal sekolah, jurusan pada *cluster* 3 yaitu Binjai Timur-SMA-IPA

DAFTAR REFERENSI

- Azmi, N., Helmiyah, F., & Sudarmin, S. (2022). Implementasi metode K-Means sebagai upaya penentuan lokasi promosi penerimaan siswa baru. *Building of Informatics, Technology and Science (BITS)*, 3(4), 649–660.
- Buulolo, E. (2020). *Data mining: Untuk perguruan tinggi*. Deepublish Publisher.
- Garaika, G., & Feriyan, W. (2019). Promosi dan pengaruhnya terhadap animo calon mahasiswa baru dalam memilih perguruan tinggi swasta. *Jurnal Aktual*, 16(1), 21–30.
- Lestari, W. (2019). Clustering data mahasiswa menggunakan algoritma K-Means untuk menunjang strategi promosi (Studi kasus: STMIK Bina Bangsa Kendari). *Jurnal Sistem Informasi dan Sistem Komputer*, 4(2), 35–48.
- Mirantika, N., Tsamratul'ain, A., & Diviana Agnia, F. (2021). Penerapan algoritma K-Means clustering untuk pengelompokan penyebaran COVID-19 di Provinsi Jawa Barat. *Jurnal Ilmu Komputer*, 15. <https://journal.uniku.ac.id/index.php/ilkom>
- Reza, A., Serasi Ginting, B., & Fauzi, A. (2022). Penerapan metode clustering untuk pengelompokan data peserta didik baru (PPDB) di SMP Swasta Gotong Royong Kuala. *Jurnal Sistem Informasi Kaputama (JSIK)*, 6(2), 45–55.