

# **PENERAPAN ALGORITMA *K-MEANS* UNTUK MENENTUKAN STRATEGI PROMOSI (Studi Kasus: Universitas Katolik Widya Mandala Kampus Kota Madiun )**

**Lorensus Anang Setiyo<sup>\*1)</sup>, Ign. F. Bayu Andoro<sup>2)</sup>**

<sup>1)</sup>Rekayasa Industri, Teknik, Universitas Katolik Widya Mandala Surabaya kampus kota Madiun,  
Jl. Manggis No 15-17, Madiun, 63133, Indonesia

<sup>2)</sup>Sistem Informasi, STMIK Widya Pratama, Jl. Patriot No.25, Pekalongan, 51146, Indonesia  
Email: anangsetiyo@ukwms.ac.id, uyab99@hotmail.com

## **ABSTRAK**

Data penerimaan mahasiswa baru yang diperoleh tiap tahun belum sepenuhnya dimanfaatkan untuk evaluasi dan untuk penentuan bentuk strategi pemasaran yang akan digunakan. Data history sangat bermanfaat untuk menentukan strategi promosi yang tepat. Pengolahan data sangat penting saat ini terutama untuk memperoleh pengetahuan yang baru, dimana pengetahuan baru tersebut nantinya akan digunakan sebagai penunjang penentuan keputusan bagi institusi dalam menentukan strategi pemasaran penerimaan calon mahasiswa baru.

Data mining sangat membantu institusi baik perusahaan maupun perguruan tinggi dalam mendapatkan pola-pola dari data yang tersimpan dalam *data base* yang mereka miliki. Dalam penelitian ini akan digunakan metode algoritma *K-Means* untuk penerapan data mining *clustering*.

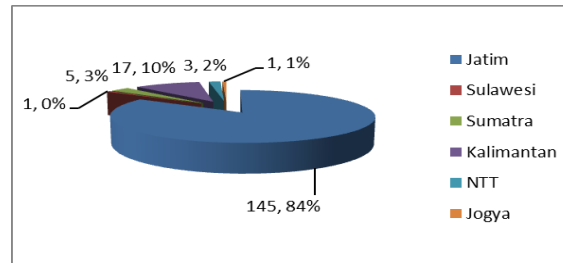
Data yang dipakai dalam penelitian ini merupakan data penerimaan mahasiswa baru (PMB) tahun akademik 2020/2021 dengan jumlah 172 mahasiswa, dengan menggunakan 3 *clustering* maka didapat hasil analisis sebagai berikut : (1) *clustering* 0 terdiri 99 mahasiswa, (2) *clustering* 1 terdiri 49 mahasiswa, dan (3) *clustering* 2 terdiri 24 Mahasiswa. Setelah dilakukan analisis data mining dengan menggunakan metode algoritma *K-Means* maka dapat dilihat untuk *cluster* 1 berdasarkan asal kota terdapat 41% dari Kota Madiun, berdasarkan jenis sekolah SMAN terdapat 20%, berdasar pilihan prodi Akuntansi 21%, dan nilai rata-rata test  $\geq 80$  sebesar 36%. Pada *cluster* 2 terdapat 10% berasal dari kota Madiun, 8% jenis sekolah SMA dan SMA swasta, berdasar pilihan prodi Manajemen 10%, dan 16% nilai rata-rata test  $\geq 80$ . Sedangkan *cluster* 3 berdasarkan asal kota Madiun 4%, berdasarkan jenis sekolah SMAN sebesar 6%, dan 9% nilai rata-rata tes yang  $\geq 80$ .

**Kata kunci:** *data mining, K-Means, clustering, promosi*

## **1. Pendahuluan**

Pada awal tahun ajaran baru perguruan tinggi pasti melakukan proses penerimaan mahasiswa baru, data yang didapatkan tidaklah sedikit, tetapi pemanfaatan data untuk kebutuhan penentuan strategi baik untuk evaluasi promosi maupun strategi pemasaran belum sepenuhnya dilakukan menggunakan data-data yang telah dimiliki. Sebagai contohnya Universitas Katolik Widya Mandala Surabaya (UKWMS) Kampus Kota Madiun data yang diperoleh tiap tahun belum sepenuhnya dimanfaatkan untuk evaluasi dan penentuan bentuk strategi pemasaran yang akan digunakan tahun berikutnya. Data history sangat bermanfaat untuk menentukan strategi promosi yang tepat. Sementara yang sudah dilakukan adalah sekedar analisis sederhana guna mengevaluasi dan memetakan data mahasiswa untuk sekedar mengetahui sebaran mahasiswa baru yang didapat. Pengolahan data sangat penting saat ini terutama untuk memperoleh pengetahuan yang baru, dimana pengetahuan baru tersebut nantinya akan digunakan sebagai penunjang keputusan bagi institusi dalam menentukan strategi pemasaran penerimaan calon mahasiswa baru.

Data mahasiswa UKWMS Kampus Kota Madiun tahun akademik 2020/2021 adalah 172 mahasiswa sebagian besar berasal dari wilayah Jawa Timur dan sampai pulau Jawa (Kalimantan, Sumatera, dan NTT).



Gambar 1. Sebaran mahasiswa berdasarkan asal daerah.

Gambar 1 menunjukkan sebaran berdasarkan asal daerah mahasiswa yang diterima di Universitas Katolik Widya Mandala Surabaya kampus Kota Madiun, 84% dari Jawa Timur, 10% dari Kalimantan dan sisanya dari Sumatra, NTT dan Yogyakarta.

Dengan memanfaatkan proses data mining keberadaan data dan informasi yang cukup besar dapat digunakan untuk menganalisis dalam penentuan kebijakan pimpinan dalam menentukan strategi berikutnya, data mining sangat membantu institusi baik perusahaan maupun perguruan tinggi dalam mendapatkan pola-pola dari data yang dimilikinya, dan tersimpan dalam *data base* yang mereka miliki. Data yang telah diolah akan mendapatkan pengetahuan baru yang akan digunakan dalam menentukan strategi dan sebagai pedoman dalam pengambilan keputusan ditingkat pimpinan. Dengan adanya perangkat lunak (*software*) untuk membantu dalam analisis data mining, tetapi peran manusia masih dibutuhkan dalam setiap langkah penyelesaian masalah yang akan dihadapi. *Software* hanya membantu dalam pengelolaan model statistik dan matematik, sedangkan analisis dan implementasi pengetahuan memerlukan manusia dalam penerapannya. Data mining memiliki kemampuan dalam mencari dan memperoleh informasi-informasi yang sangat penting dari basis data yang sangat besar (Yunita, 2018). Proses yang digunakan dalam data mining menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi serta mengidentifikasi informasi yang berguna dan ada keterkaitannya dengan pengetahuan (Ria et al., 2015). Salah satu algoritma yang sangat populer dalam penelitian data mining adalah algoritma *K-means*, metode ini memiliki prinsip mengelompokkan data sesuai dengan karakteristik yang sama dalam satu *cluster*. Data mining yang menggunakan algoritma *K-means* bertujuan untuk mendapatkan dan menemukan pola-pola data yang tersembunyi. *Clustering* dapat membantu dalam pengelompokan data-data secara otomatis yang memiliki kemiripan/kesamaan, sehingga data-data yang dianggap mirip di gabung bersama karena persamaan atau kedekatannya (Yudiarta et al., 2018)

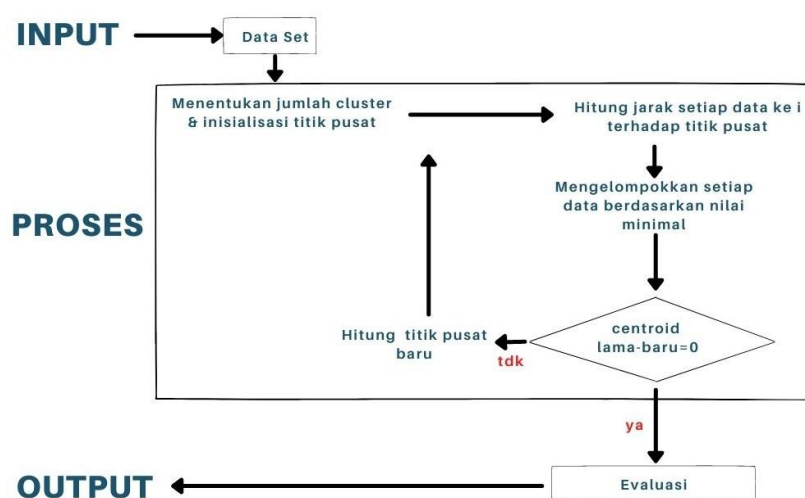
Dalam penelitian sebelumnya, algoritma *K-means* digunakan untuk mendukung strategi promosi dengan memanfaatkan data lulusan dan hasil dari analisis didapat sebaran wilayah berdasarkan potensi akademik yang dimiliki mahasiswa (Asril et al., 2015). Penelitian lainnya yaitu analisa penerapan data mining pada penerimaan mahasiswa Poltek Negri Lhokseumawe menggunakan algoritma *K-means* menghasilkan tampilan grafik pengelompokan data jumlah mahasiswa berdasarkan jalur masuk menurut jurusan yang dipilih (Ria et al., 2015). *Clustering* data mahasiswa dengan menggunakan algoritma *K-means* digunakan untuk menunjang strategi promosi dan di masing-masing program studi sesuai dengan hasil yang paling banyak diminati berdasarkan masing-masing *cluster* (Lestari et al., 2019).

Berdasarkan pertimbangan penelitian sebelumnya dan beberapa referensi yang telah didapat, maka tujuan dari penelitian ini adalah menganalisis data calon mahasiswa baru dengan menggunakan algoritma *K-means* pada UKWMS Kampus Kota Madiun. Atribut yang akan digunakan dalam penelitian ini adalah data mahasiswa baru tahun akademik 2019/2020 dan sumber data berasal dari bagian Penerimaan Mahasiswa Baru (PMB). Manfaat yang diharapkan pada penelitian ini, dapat membantu pimpinan dan tim promosi universitas dalam melaksanakan strategi promosi yang tepat.

## 2. Metode.

### 2.1. Diagram Alir Penelitian

Diagram alir penelitian yang digunakan dalam penelitian ini dapat dilihat pada gambar 1 dibawah ini.



Gambar 2. Diagram Alir Penelitian

### 2.2. Algoritma K-Means.

Dengan algoritma *K-Means* kita dapat pengelompokan data dengan memaksimalkan kemiripan data dalam satu *cluster* dan meminimalkan kemiripan data antar *cluster*. Algoritma ini dinilai cukup efektif dalam pengolahan data mining yang digunakan dalam pengelompokan (*clustering*) data yang akan dianalisa (Helilintar & Farida, 2018). Metode algoritma *K-Means* adalah sebagai berikut (Muhima et al., 2020):

**Input** :  $k = \text{cluster}$ ,  $d = \text{dataset } \{d_1, d_2, d_3, \dots, d_n\}$

**Metode** : a. Tentukan nilai *cluster* yang ada didalam dataset untuk digunakan sebagai *centroid* awal.

b. Menghitung jarak antara setiap titik data  $x_j$  dan *centroid* dengan menggunakan persamaan *Euclidean*.

$$d(x_j, c_j) = \sqrt{\sum_{j=1}^n (x_j - c_j)^2}$$

- c. Mengelompokkan setiap data dengan berdasarkan jarak terdekat antara data dengan *centroid* awal
- d. Menentukan *centroid* baru dengan menghitung nilai rata-rata dari tiap *cluster*
- e. Di uji jika posisi *centroid* baru dengan *centroid* lama sama (*centroid* lama dikurangi *centroid* baru = nol), maka proses selesai, jika tidak sama maka proses kembali ke langkah b untuk iterasi yang ke dua dan seterusnya hingga mendapatkan nilai *centroid* baru dengan *centroid* lama sama.

**Output** : *Cluster* baru untuk dilakukan evaluasi.

### 3. Hasil dan Pembahasan

#### 3.1. Seleksi Data

Dalam tahap ini peneliti menggunakan data profil mahasiswa, dari data yang didapat kemudian dipilih atribut yang bisa dipakai untuk penelitian ini. Dari beberapa atribut yang ada kemudian dipilih 5 atribut yang dipakai sebagai bahan penelitian yaitu Asal Kota, Program Studi, Asal Sekolah, Nilai Matematika, dan Nilai Bahasa Inggris. Tabel 1 menunjukkan data mahasiswa yang dipakai untuk penelitian ini.

Tabel 1. Data penelitian

No	Asal Kota	Program Studi	Sekolah Asal	Mat	Inggris
1	Madiun	Bahasa Inggris	SMAK St. Bonaventura Madiun	74.5	80.25
2	Madiun	Akuntansi	SMAK St. Bonaventura Madiun	79.75	81.5
3	Kab. Ngawi	Manajemen	SMK PGRI 4 Ngawi	81	76.5
4	Madiun	Manajemen	SMAK St. Bonaventura Madiun	77.25	78.5
5	Madiun	Manajemen	SMKN 1 Mejayan	78.5	82.75
6	Madiun	Manajemen	SMKN 1 Mejayan	80.25	81.25
...	.....	.....	.....	.....	.....
172	Kab. Madiun	Farmasi	SMAN 1 Geger	79	82

#### 3.2. Pre-Processing

*Pre-Processing* diperlukan untuk melihat data yang tidak sempurna, dengan cara menghapus data yang hilang, *missing*, dan pada data diatas untuk memudahkan dalam pengelompokan maka atribut nilai Matematika dan nilai Inggris di hitung rata-ratanya. Sehingga atribut yang diperoleh setelah dilakukan *pre-processing* seperti pada tabel 2 dibawah ini.

Tabel 2. Data Setelah dilakukan *Pre-Processing*

No	Asal Kota	Program Studi	Sekolah Asal	Rerata
1	Madiun	Bahasa Inggris	SMAK St. Bonaventura Madiun	77
2	Madiun	Akuntansi	SMAK St. Bonaventura Madiun	81
3	Kab. Ngawi	Manajemen	SMK PGRI 4 Ngawi	79
4	Madiun	Manajemen	SMAK St. Bonaventura Madiun	78
5	Madiun	Manajemen	SMKN 1 Mejayan	81
6	Madiun	Manajemen	SMKN 1 Mejayan	81

...	.....	.....	.....	.....
172	Kab. Madiun	Farmasi	SMAN 1 Geger	81

### 3.3. Transformasi

Tujuan dari transformasi adalah untuk mengelompokkan/mengubah data sehingga dapat dilakukan analisis data dengan menggunakan metode *K-means*. Adapun untuk variabel asal sekolah dikelompokkan menjadi 16 kelompok yaitu untuk Kota Madiun di transformasikan menjadi 1, Kabupaten Madiun di transformasikan menjadi 2, Ngawi di transformasikan menjadi 3, dan seterusnya sampai Kabupaten Pacitan di transformasikan menjadi 16. Variabel prodi dikelompokkan menjadi 10, prodi PBSI di transformasikan menjadi 1, prodi BK di transformasikan menjadi 3, prodi Matematika di transformasikan menjadi 3 dan seterusnya sampai prodi Farmasi di transformasikan menjadi 10, sedangkan variabel asal sekolah SMA Negeri di transformasikan menjadi 1, asal sekolah SMA Swasta di transformasikan menjadi 2, asal sekolah SMK Negeri di transformasikan menjadi 3, asal sekolah SMK Swasta di transformasikan menjadi 4, dan variabel nilai rata-rata  $\Rightarrow >70$  di transformasikan menjadi 1, nilai rata-rata  $69-79$  di transformasikan menjadi 2, dan nilai rata-rata  $80 \leq$  di transformasikan menjadi 3.

Sehingga setelah semua data di transformasi menjadi seperti pada tabel 4 dibawah ini.

Tabel 4. Hasil Transformasi Data

ID	Asal Kota	Prodi	Jenis Sekolah	Rerata
1	1	5	2	2
2	1	8	2	3
3	3	7	4	2
4	1	7	2	2
5	1	7	3	3
6	1	7	3	3
...	.....	.....	.....	.....
172	2	10	1	3

### 3.4. Kalkulasi Menggunakan Algoritma *K-Means*

Setelah dilakukan proses transformasi, langkah berikutnya adalah mengolah data dengan menggunakan algoritma *K-Means*. Pada pengolahan data akan ditentukan cluster baru yang ingin dibuat. Pada penelitian ini cluster yang akan dibuat adalah 3 cluster. Setelah menentukan cluster langkah berikutnya menentukan titik pusat awal dari tiap cluster, dalam penentuan titik pusat ditentukan secara random. Tabel 5 merupakan titik pusat awal dari tiap cluster.

Tabel 5. Titik Pusat dari Awal Cluster

Centroid 1	1	9	4	3
Centroid 2	5	10	4	3
Centroid 3	6	1	2	2

Setelah menentukan titik pusat cluster, langkah berikutnya adalah menghitung jarak objek ke centroid, pada penelitian ini aplikasi yang digunakan untuk membantu perhitungan persamaan dengan menggunakan excel. Tabel 6 hasil perhitungan untuk iterasi pertama.

Tabel 6. Hasil Iterasi pertama

Jarak Data Ke Centroid					
	M1	M2	M3	Cluster	Jarak Terdekat
	4.58	6.78	6.40	1	4.58
	2.24	4.90	8.66	1	2.24
	3.00	3.74	7.00	1	3.00
	3.00	5.48	7.81	1	3.00
	2.24	5.10	7.94	1	2.24
	2.24	5.10	7.94	1	2.24
	.....	.....	.....	.....	.....
	6.86	3.74	9.11	2	3.74
	3.32	4.24	9.95	1	3.32
Total	740.36	861.82	1330.85	269	544.90
Rata-rata	4.30	5.01	7.74	1.56	3.17

Tabel 7. Posisi *cluster* iterasi pertama

C1				C2				C3			
A	B	C	D	A	B	C	D	A	B	C	D
1	5	2	2								
1	8	2	3								
...	....	...	...	....	...	...	...	..	...	...	...
2	10	4	3								
				7	10	1	2				
2	10	1	3								
<b>1.3</b>	<b>7.7</b>	<b>2.6</b>	<b>2.6</b>	<b>5.8</b>	<b>8.4</b>	<b>2.4</b>	<b>2.5</b>	<b>5.5</b>	<b>3.1</b>	<b>2</b>	<b>2.6</b>

Setelah semua data ditempatkan dicluster terdekat seperti pada tabel 7, kemudian dihitung kembali dengan pusat *cluster* yang baru yaitu berdasarkan nilai rata-rata dari *cluster* sebelumnya. Dari tabel 7 didapat nilai rata-rata dari titik pusat *cluster* yang baru.

Tabel 8. Hasil Iterasi Pertama Titik Pusat dari Awal *Cluster*

Centroid 1	1.3	7.7	2.6	2.6
Centroid 2	5.8	8.4	2.4	2.5
Centroid 3	5.5	3.1	2	2.6

Proses selanjutnya menghitung ulang jarak objek ke *centroid* yang baru (nilai rata-rata dari *cluster* iterasi pertama). Setelah iterasi kedua diperoleh dan posisi *cluster* iterasi kedua didapat, langkah selanjutnya adalah dilakukan pengujian jika nilai *centroid* baru dikurangi *centroid* lama sama dengan nol maka iterasi dihentikan, pada penelitian untuk memperoleh nilai nol perlu dilakukan 3 kali iterasi.

Hasil akhir dari penelitian ini adalah pengelompokan berdasarkan kedekatan jarak antara titik pusat dengan semua atribut dari data yang dianalisis. Proses menghasilkan *cluster* model dengan 3 *cluster* dimana *cluster* 0 terdiri dari 99 item, *cluster* 1 terdiri dari 48, dan *cluster* 2 terdiri dari 24 item, sehingga total data sebanyak 172 mahasiswa.

Tabel 9. Hasil Cluster

Cluster 1		Cluster 2		Cluster 3	
Madiun	41%	Magetan	10%	Madiun	4%
Kab. Madiun	12%	Kalimantan	8%	Kalimantan	2%
SMAN	20%	Ponorogo	3%	SMAN	6%
SMK Swasta	21%	SMAN & SMA Swasta	8%	SMA Swasta	6%
Akuntansi	21%	SMK Swasta	6%	SMKN & SMK Swasta	3%
Farmasi	13%	SMKN	6%	Manajemen	3%
Manajemen	12%	Manajemen	10%	Bhs. Inggris & Matematika	3%
Rekayasa Industri	3%	Farmasi	9%	PBSI & BK	3%
Psikologi	5%	Psikologi	4%	Biologi	1%
Nilai $\geq 80$	36%	Nilai $\geq 80$	16%	Nilai $\geq 80$	9%

Berdasarkan hasil analisis didapat hasil *cluster* yang dapat digunakan untuk pengambilan keputusan dalam penetapan strategi promosi di UKWMS Kampus Kota Madiun. Dalam tabel 9 dapat dilihat *cluster* 1 berdasarkan asal kota terdapat 41% dari Kota Madiun, berdasarkan jenis sekolah SMAN terdapat 20%, berdasarkan pilihan prodi Akuntansi 21%, prodi Farmasi 13%, prodi Manajemen 12%, prodi Psikologi 5%, dan nilai rata-rata test  $\geq 80$  sebesar 36%. Pada *cluster* 2 terdapat 10% berasal dari Magetan, 8% jenis sekolah SMA dan SMA swasta, yang memilih prodi Manajemen 10%, prodi Farmasi 9%, prodi Psikologi 4%, dan 16% nilai rata-rata test  $\geq 80$ . Dan *cluster* 3 berdasarkan asal kota Madiun 4%, asal kota Kalimantan 4%, berdasarkan jenis sekolah SMAN sebesar 6%, berdasarkan prodi yang dipilih Bahasa Inggris dan Pendidikan Matematika 3%, prodi PBSI dan BK 3%, prodi Biologi 1%, dan 9% nilai rata-rata tes yang  $\geq 80$ .

#### 4. Simpulan.

Kesimpulan yang dapat diambil dalam penelitian ini, perlu adanya pertimbangan dalam menentukan tempat promosi berdasarkan kota asal calon mahasiswa dan program studi apa yang menjadi pilihan calon mahasiswa selain itu jenis sekolah yang dijadikan target promosi, sehingga pilihan asal kota, program studi dan jenis sekolah tidak salah sasaran dalam pengambilan strategi promosi yang dilakukan tim promosi UKWMS Kampus Kota Madiun.

#### Daftar Pustaka

- Asril, E., Wiza, F., & Yunefri, Y. (2015). Analisis Data Lulusan dengan Data Mining untuk Mendukung Strategi Promosi Universitas Lancang Kuning. In *Jurnal Teknologi Informasi & Komunikasi Digital Zone* (Vol. 6, Issue 2).
- Helilintar, R., & Farida, I. N. (2018). Penerapan Algoritma K-Means Clustering Untuk Prediksi Prestasi Nilai Akademik Mahasiswa. In *Jurnal Sains dan Informatika* (Vol. 4, Issue 2).

- Lestari, W., Bina, S., & Kendari, B. (2019). Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus : STMIK Bina Bangsa Kendari). In *SIMKOM* (Vol. 4, Issue 2). <http://e-jurnal.stmikbinsa.ac.id/index.php/simkom35>
- Muhima, R. R., Kurniawan, M., & Pambudi, O. T. (2020). A LOF K-Means Clustering on Hotspot Data. *International Journal of Artificial Intelligence & Robotics (IJAIR)*, 2(1), 29. <https://doi.org/10.25139/ijair.v2i1.2634>
- Ria, D., Tb, Y., Cerdas Multimedia, J., & Elektro, J. T. (2015). *ANALISA PENERAPAN DATA MINING PADA PENERIMAAN MAHASISWA POLITEKNIK NEGERI LHOKSEUMAWE MENGGUNAKAN ALGORITMA K-MEANS*.
- Yudiarta, N. G., Sudarma, M., & Ariastina, W. G. (2018). Penerapan Metode Clustering Text Mining Untuk Pengelompokan Berita Pada Unstructured Textual Data. *Majalah Ilmiah Teknologi Elektro*, 17(3), 339. <https://doi.org/10.24843/mite.2018.v17i03.p06>
- Yunita, F. (2018). PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA K-MEANS CLUSTRING PADA PENERIMAAN MAHASISWA BARU (STUDI KASUS : UNIVERSITAS ISLAM INDRAGIRI). In *Jurnal SISTEMASI* (Vol. 7).