

Klasifikasi Data Penerimaan Mahasiswa Baru Pada Universitas Bina Darma Menggunakan Algoritma *K-Means Clustering*

Devi Udariansyah^{1*}, Deny Rahmat Ibrahim²

^{1,2}Program Studi Teknik informatika, Fakultas Ilmu Komputer,
Universitas Bina Darma Palembang

E-mail: devi.udariansyah@binadarma.ac.id¹, rahmatdeny4@gmail.com²

Abstrak

Tujuan dari penelitian ini adalah mengelompokkan data mahasiswa baru tahun 2019 sampai dengan 2021 pada Universitas Bina Darma dengan memanfaatkan proses data mining dengan menggunakan algoritma *K-means clustering*. Implementasi menggunakan *software Rapidminer* digunakan untuk membantu menemukan nilai yang akurat. Dalam mengerjakan penelitian ini penulis menggunakan 3 atribut dalam mengelompokkan data terdiri dari nama sekolah, daerah sekolah, jenis kelamin. Jenis pengambilan data yang digunakan menggunakan data sekunder. Metode analisis kebutuhan data yang digunakan penulis menggunakan tahap *knowledge discovery-in data base* (KDD). Cluster mahasiswa yang terbentuk adalah tiga cluster, dengan cluster pertama 897 *items*, cluster kedua 2054 *items* dan cluster ketiga berjumlah 389 *items*. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan dalam menentukan strategi mempromosikan Universitas Bina Darma di masing-masing daerah. Berdasarkan hasil cluster algoritma *k-means* dapat dilihat nama sekolah dan daerah sekolah mana saja yang mendaftar di Universitas Bina Darma Palembang.

Kata Kunci: mahasiswa baru, *k-means*, *clustering*

Abstract

The purpose of this study is to classify new student data from 2019 to 2021 at Bina Darma University by utilizing the data mining process using the *K-means clustering algorithm*. Using *Rapidminer software* is used to help find accurate values. In doing this research the writer uses 3 attributes in classifying the data consisting of the name of the school, the school area, and gender. The type of data collection used is secondary data. The data requirements analysis method used by the author uses the *knowledge discovery-in data base* (KDD) stage. The student clusters formed are three clusters, with the first cluster 897 *items*, the second cluster 2054 *items* and the third cluster totaling 389 *items*. The results of this study are used as a basis for decision making in determining strategies to promote Bina Darma University in each region. Based on the results of the *k-means* it can be seen the name of the school and which school area registered at Bina Darma University Palembang.

Keywords: new students, *k-means*, *cluster*

PENDAHULUAN

Perkembangan teknologi informasi sekarang sudah semakin canggih dan modern dalam segala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang ekonomi, industri, pendidikan dan teknologi serta berbagai bidang kehidupan lainnya. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai data mahasiswa dan proses belajar yang dihasilkan (Zendrato et al., 2022).

Didalam Pendidikan tinggi data sangat diperlukan untuk melengkapi setiap proses administrasi mahasiswa mulai dari diterimanya sebagai mahasiswa aktif sampai pada akhirnya menjadi alumni. Tentunya dari data-data tersebut bisa diketahui beberapa Profil penting dari setiap mahasiswa. Tentunya sebuah kesempatan bagi pihak Universitas ketika mendapatkan data-data informasi dari setiap Mahasiswa, selain sebagai bentuk arsip bagi pihak Universitas, dari data-data tersebut bisa digunakan untuk menganalisis kondisi dari setiap mahasiswa sebelum mendaftar pada Universitas tertentu. Sebagai contoh yaitu Nama mahasiswa, nama Sekolah, daerah sekolah, Tahun lahir, jenis kelamin dan Jurusan atau prodi yang di ambil ketika masuk ke perguruan tinggi. Dari data tersebut bisa dianalisis seberapa banyak mahasiswa mengambil Jurusan tertentu di Universitas tersebut. Ketika pihak kampus mendapatkan data ini, maka akan semakin mudah bagi

pihak manajemen universitas untuk mencari sekolah-sekolah yang mana saja yang paling efektif untuk dikunjungi dalam hal mempromosikan Universitas (damanik & sigiro, 2021).

Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Data mining merupakan proses *iterative* dan *interaktif* untuk menemukan pola atau model baru yang sempurna, bermanfaat dan dapat dimengerti dalam suatu database yang besar (C.Pradeepkumar & S.Loganathan, 2015).

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data (Utomo & Mesran, 2020).

K-Means adalah salah satu teknik clustering pada Data Mining proses pemodelan tanpa supervisi dan metode pengelompokan data secara partisi. Data yang dikelompokkan metode K-Means menjadi beberapa kelompok dan setiap kelompok memiliki karakteristik yang mirip atau sama dengan lainnya tetapi dengan kelompok lainnya memiliki karakteristik berbeda. Dengan tujuan meminimalisasi perbedaan setiap data didalam satu cluster serta memaksimalkan perbedaan dengan cluster yang lain.(Muliono & Sembiring, 2019). Clustering adalah mengelompokkan item data ke dalam sejumlah kecil grup sedemikian sehingga masing-masing grup mempunyai sesuatu persamaan yang esensial (PRAKOSO, 2019).

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari Institute of Teknologi Blanchardstown dan Ralf Klinkenberg dari rapid-i.com dengan tampilan GUI (Graphical User Interface) sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. RapidMiner merupakan perangkat lunak yang bersifat terbuka (open source). RapidMiner adalah sebuah solusi untuk melakukan analisis terhadap data mining, text mining dan analisis prediksi. RapidMiner menggunakan berbagai teknik deskriptif dan prediksi dalam memberikan wawasan kepada pengguna sehingga dapat membuat keputusan yang paling baik (Adhinda et al., 2020).

METODE

Metode pengumpulan data yang digunakan dalam penelitian ini adalah:

1. Wawancara (Interview)

Merupakan suatu pengumpulan data yang dilakukan dengan cara tanya jawab secara langsung. Wawancara dilakukan dengan pihak yang terkait, karyawan atau admin di Universitas Bina Darma Palembang mengenai Data Mining menggunakan Algoritma K-Means Clustering untuk menentukan strategi promosi mahasiswa baru hingga tahap akhir penelitian.

2. Pengamatan (Observasi)

Yaitu metode pengumpulan data dengan cara mengadakan tinjauan secara langsung ke objek yang diteliti. Observasi dilakukan langsung pada Universitas Bina Darma Palembang.

3. Studi Pustaka

Untuk mendapatkan data-data yang bersifat teoritis maka penulis melakukan pengumpulan data dengan cara membaca dan mempelajari buku-buku, jurnal penelitian, makalah ataupun referensi lain yang berhubungan dengan masalah yang dibahas (Alhapizi et al., 2020).

Penelitian ini menggunakan metode knowledge discovery in database (KDD) sering kali digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. Dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining (Yunita, 2018).

Proses KDD secara garis besar dapat dijelaskan sebagai berikut:

1. Data Selection

Pemilihan (seleksi) data dari sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam KDD dimulai. Data hasil seleksi yang akan digunakan untuk proses data mining disimpan dalam suatu berkas, terpisah dari basis data operasional.

2. Pre-processing / Cleaning

Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses pembersihan pada data yang menjadi fokus KDD. Proses pembersihan mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi).

3. Transformation

Coding adalah transformasi pada data yang telah dipilih, sehingga data tersebut sesuai untuk proses data mining. Proses coding dalam KDD merupakan proses kreatif dan sangat tergantung pada jenis atau pola informasi yang akan dicari dalam basis data.

Sesudah pengumpulan data yang telah dilakukan, maka langkah selanjutnya adalah tahap pengolahan data yang telah terkumpul. Tempat dilakukannya penelitian di Universitas Bina Darma Palembang Jalan Jenderal Ahmad Yani No.3, 9/10 Ulu, Kecamatan Seberang Ulu I, Kota Palembang, Sumatera Selatan 30111 penelitian ini di lakukan pada bulan febuari sampai dengan bulan maret 2022.

1. Data Selection

Data dalam penelitin ini bersumber dari Universitas Bina Darma dimana data ini merupakan data sekunder yang terdiri atas data mahasiswa baru tahun ajaran 2019 sampai dengan 2021. Adapun jumlah data yang di peroleh sebanyak 3340 data yang derdiri Nama Sekolah, Daerah Sekolah, Jenis Kelamin. Berikut data mahasiswa baru 2019 sampai dengan 2021 yang diperoleh :

Tabel 1. Data mahasiswa baru

NO	NAMA SEKOLAH	DAERAH SEKOLAH	JENIS KELAMIN
1	SMA	KOTA PALEMBANG	PEREMPUAN
2	SMA	KOTA PALEMBANG	LAKI-LAKI
3	SMA	KOTA PALEMBANG	PEREMPUAN
4	SMA	KOTA PALEMBANG	LAKI-LAKI
5	SMA	KABUPATEN BAYUASIN	LAKI-LAKI
6	LAINNYA	KABUPATEN MUARA ENIM	PEREMPUAN
7	LAINNYA	KABUPATEN MUARA ENIM	PEREMPUAN
8	LAINNYA	KABUPATEN OGAN KOMERING ILIR	PEREMPUAN
9	SMK	KABUPATEN MUSI BANYUASIN	LAKI-LAKI
10	SMA	KABUPATEN OGAN ILIR	LAKI-LAKI
11	SMK	KABUPATEN MUSI BANYUASIN	LAKI-LAKI
12	SMK	KABUPATEN MUSI BANYUASIN	PEREMPUAN
13	SMK	KABUPATEN LAHAT	LAKI-LAKI
14	SMA	KABUPATEN BANYUASIN	LAKI-LAKI
15	SMK	KABUPATEN BANYASIN	LAKI-LAKI
16	SMK	KABUPATEN MUSI BANYUASIN	LAKI-LAKI
17	SMA	KABUPATEN OGAN KOMERING ILIR	PEREMPUAN
18	SMA	MUARA ENIM	PEREMPUAN
19	SMA	PALEMBANG	LAKI-LAKI
20	LAINNYA	PALEMBANG	PEREMPUAN

3.2 Tranformasi Data

Tranformasi data dilakukan untuk mengubah data tujuannya agar data dapat diolah dengan menggunakan metode *K-means Clustering*. Adapun variabel yang digunakan pada data mahasiswa baru yaitu data NamaSekolah, DaerahSekolah, JenisKelamin. Untuk data nama sekolah dikelompokan menjadi 3 kelompok. Adapun SMA ditranformasi dengan nilai 1, SMK ditrasformasi dengan nilai 2, dan untuk nilai nama sekolah LAINNYA ditranformasikan dengan nilai 3. Untuk variabel DaerahSekolah dikelompokan menjadi 20 kelompok. Adapun daerah sekolah kota pelambang ditranformasikan dengan nilai 1, daerah sekolah kabupaten banyuasin ditranformasikan dengan nilai 2, daerah sekolah kabupaten muara enim ditranformasikan dengan nilai 3, daerah sekolah kabupaten ogan komering ilir ditranformasikan dengan nilai 4, daerah sekolah kabupaten musu banyuasin ditranformasikan dengan nilai 5, daerah sekolah kabupaten ogan ilir ditranformasikan dengan nilai 6, daerah sekolah kabupaten lahat ditranformasikan dengan nilai 7, daerah sekolah kota prabumulih ditranformasikan dengan nilai 8, daerah sekolah oga komering ulu ditranformasikan dengan nilai 9, daerah sekolah kabupaten oku timur ditranformasikan dengan nilai 10, daerah sekolah kota lubuk linggau ditranformasikan dengan nilai 11, daerah sekolah kabupaten pali ditranformasikan dengan nilai 12, daerah sekolah kota pagar alam ditranformasikan dengan

nilai 13, daerah sekolah kabupaten musi rawas ditransformasikan dengan nilai 14, daerah sekolah kabupaten empat lawang ditransformasikan dengan nilai 15, daerah sekolah kabupaten oku selatan ditransformasikan dengan nilai 16, daerah sekolah kota jambi ditransformasikan dengan nilai 17, daerah sekolah kota pangkal pinang ditransformasikan dengan nilai 18, daerah sekolah kabupaten Bangka barat ditransformasikan dengan nilai 19, daerah sekolah kota bengkulu ditransformasikan dengan nilai 20. Untuk data jenis kelamin dikelompokkan menjadi 2 kelompok. Adapun jenis kelamin laki-laki ditransformasi dengan nilai 1, jenis kelamin perempuan ditransformasi dengan nilai 2. Adapun hasil dari tranformasi data dapat dilihat pada table berikut :

Tabel 2. Data Hasil Transformasi

ID	NAMA SEKOLAH	DAERAH SEKOLAH	JENIS KELAMIN
K1	1	1	2
K2	1	1	1
K3	1	1	2
K4	1	1	1
K5	1	2	1
K6	3	3	2
K7	3	3	2
K8	3	4	2
K9	2	5	1
K10	1	6	1
K11	2	5	1
K12	2	5	2
K13	2	7	1
K14	1	2	1
K15	2	2	1
K16	2	5	1
K17	1	4	2
K18	1	3	2
K19	1	1	1
K20	3	1	2

3.3 Pengolahan Data

Pengolahan data mahasiswa baru dilakukan setelah proses tranformasi sehingga data mahasiswa baru bias diolah menggunakan metode *K-Mean Clustering*. Adapun langkah-langkah proses algoritma *K-MEANS CLUSTERING* adalah sebagai berikut :

1. Dilakukan k dari jumlah *clustering* baru yang ingin dibentuk. Adapun Cluster yang akan dibuat adalah 3 cluster.
2. Tentukan titik puas awal dari setiap cluster. Adapun penentuan titik pusat awal dalam penelitian ini ditentukan secara random dan titik pusat yang didapatkan dapat dilihat pada table berikut :

Tabel 3. titik pusat awal

Titik Pusat			
Centroid 1	3	4	1
Centroid 2	1	3	2
Cendroid 3	2	2	1

3. Hitung jarak setiap data ke pusat cluster antara objek ke centroid terdekat. Centroid terdekat akan menjadi cluster yang diikiti oleh data tersebut. Perhitungan jarak *Eucliden* dapat dilakukan dengan persamaan berikut:

$$d(p,q) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2}$$

Persamaan diatas digunakan karena atribut yang digunakan berjumlah 3. Sebagai contoh, akan dihitung jarak dari data mahasiswa pertama ke pusat *cluster* pertama dengan persamaan

$$\begin{aligned}
 d(1,1) &= \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2} \\
 &= \sqrt{(1 - 3)^2 + (1 - 4)^2 + (2 - 1)^2} \\
 &= 3,7
 \end{aligned}$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data mahasiswa baru pertama ke pusat cluster pertama adalah 3,7. Jarak data mahasiswa baru pertama ke pusat cluster kedua dengan persamaan:

$$\begin{aligned}
 d(1,2) &= \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2} \\
 &= \sqrt{(1 - 1)^2 + (1 - 3)^2 + (2 - 2)^2} \\
 &= 2,0
 \end{aligned}$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data mahasiswa baru pertama ke pusat cluster pertama adalah 2,0. Jarak data mahasiswa baru pertama ke pusat cluster kedua dengan persamaan:

$$\begin{aligned}
 d(1,3) &= \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2} \\
 &= \sqrt{(1 - 2)^2 + (1 - 2)^2 + (2 - 1)^2} \\
 &= 1,7
 \end{aligned}$$

Dari hasil perhitungan diatas didapatkan hasil bahwa jarak data mahasiswa baru pertama ke pusat cluster pertama adalah 1,7.

Berdasarkan hasil ketiga perhitungan diatas dapat disimpulkan bahwa jarak data mahasiswa baru pertama yang paling mendekati adalah cluster 3, sehingga mahasiswa baru pertama dimasukan ke dalam cluster 3 Hasil perhitungan selengkapnya untuk 20 sampel data mahasiswa baru dapat dilihat pada tabel berikut ini :

Table 4. Hasil Perhitungan setiap data ke setiap cluster iterasi 1

DATA	Jarak Ke Cendroid			JARAK TERDEKAT	CLUSTER DIIKUTI
	1	2	3		
K1	3,7	2,0	1,7	1,7	3
K2	3,6	2,2	1,4	1,4	3
K3	3,7	2,0	1,7	1,7	3
K4	3,6	2,2	1,4	1,4	3
K5	2,8	1,4	1,0	1,0	3
K6	1,4	2,0	1,7	1,4	1
K7	1,4	2,0	1,7	1,4	1
K8	1,0	2,2	2,4	1,0	1
K9	1,4	2,4	3,0	1,4	1
K10	2,8	3,2	4,1	2,8	1
K11	1,4	2,4	3,0	1,4	1
K12	1,7	2,2	3,2	1,7	1
K13	3,2	4,2	5,0	3,2	1
K14	2,8	1,4	1,0	1,0	3
K15	2,2	1,7	0,0	0,0	3
K16	1,4	2,4	3,0	1,4	1
K17	2,2	1,0	2,4	1,0	2
K18	2,4	0,0	1,7	0,0	2

K19	3,6	2,2	1,4	1,4	3
K20	3,2	2,8	1,7	1,7	3

1. Setelah semua data ditempatkan kedalam *cluster* yang terdekat, kemudian hitung kembali pusat *cluster* yang baru berdasarkan rata-rata anggota yang ada pada *cluster* tersebut.

2. Jika centroid yang baru *konvergen* dengan *centroid* yang lama maka hentikan iterasi. Jika tidak maka iterasi dilanjutkan ke yang berikutnya.

Selanjutnya mengelompokan hasil *cluster* pada iterasi pertama yang belum *konvergen*. Untuk membangkitkan kembali *centroid* baru dengan rumus berikut :

$$C = \frac{\sum m}{n}$$

Dimana :

C : centroid data

m : anggota data yang termasuk kedalam *centroid* tertentu

n : jumlah data yang menjadi anggota *centroid* tertentu

Contoh perhitungan data sampel 9 orang mahasiswa pada Cluster 1 adalah berikut:

$$\frac{3 + 3 + 3 + 2 + 1 + 2 + 2 + 2 + 2}{9} = 2.2$$

$$\frac{3 + 3 + 4 + 5 + 6 + 5 + 5 + 7 + 5}{9} = 4.8$$

$$\frac{2 + 2 + 2 + 1 + 1 + 1 + 2 + 1 + 1}{9} = 1.4$$

Demikian juga perhitungan pada cluster 2 dan cluster 3.

Pada pengelompokan data dengan menggunakan rumus diatas didapatkan titik pusat cluster dengan nilai sebagai berikut:

table 5. titik pusat iterasi 1 setelah *cluster*

titik pusat			
Centroid 1	2,2	4,8	1,4
Centroid 2	1,0	3,5	2,0
Centroid 3	1,3	1,3	1,3

Karena *centroid* baru yang digunakan belum *konvergen* maka iterasi harus dilanjutkan. Dalam penelitian ini, iterasi terjadi sebanyak 3 kali iterasi karena kondisi *cluster* sudah mencapai konvergen dan proses iterasi pun berhenti. Adapun hasil akhir *clustering* dari 20 data mahasiswa baru dapat dilihat pada tabel berikut :

Table 6. Hasil perhitungan data ke setiap *cluster* pada iterasi-2

DATA	Jarak ke Centroid			JARAK TERDEKAT	CLUSTER DIIKUTI
	1	2	3		
K1	4,0	2,5	0,8	0,8	3
K2	4,0	2,7	0,6	0,6	3
K3	4,0	2,5	0,8	0,8	3
K4	4,0	2,7	0,6	0,6	3
K5	3,1	1,8	0,8	0,8	3
K6	2,0	2,1	2,4	2,0	1
K7	2,0	2,1	2,4	2,0	1
K8	1,2	2,1	3,2	1,2	1
K9	0,5	2,1	3,7	0,5	1
K10	1,8	2,7	4,7	1,8	1
K11	0,5	2,1	3,7	0,5	1

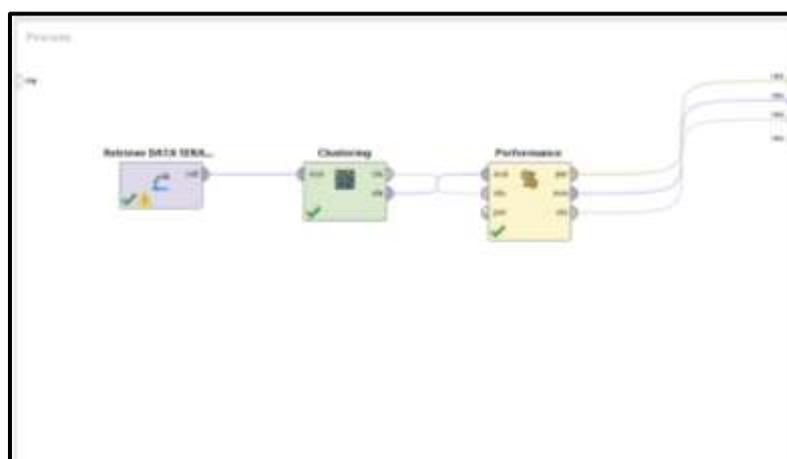
K12	0,6	1,8	3,8	0,6	1
K13	2,3	3,8	5,7	2,3	1
K14	3,1	1,8	0,8	0,8	3
K15	2,8	2,1	1,0	1,0	3
K16	0,5	2,1	3,7	0,5	1
K17	1,6	0,5	2,8	0,5	2
K18	2,2	0,5	1,8	0,5	2
K19	4,0	2,7	0,6	0,6	3
K20	3,9	3,2	1,8	1,8	3

Pada iterasi ketiga, titik pusat dari setiap *cluster* sudah tidak berubah dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain.

HASIL DAN PEMBAHASAN

4.1 Hasil iterasi dengan metode *K-Means*

Adapun pengolahan data mahasiswa baru dengan menggunakan algoritma *k-means clustering* dan *software rapidminer* dapat dilihat pada gambar berikut :



Gambar 1. Pemodelan *K-Means* pada RapidMiner

Dengan menggunakan pemodelan *k-means clustering* seperti gambar dibawah ini, dengan jumlah data inialisasi jumlah *cluster* sebanyak 3, sesuai dengan pendefinisian nilai *k* dengan jumlah *cluster_0* 87 *items*, *cluster_1* 2054 *items*, dan *cluster_2* 389 *items*.

```

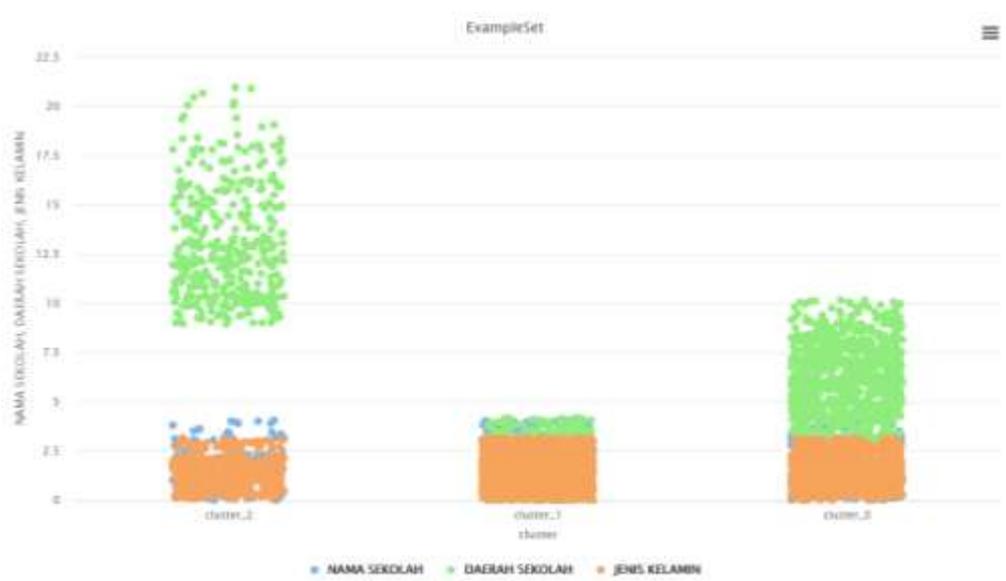
Cluster Model

Cluster 0: 897 items
Cluster 1: 2054 items
Cluster 2: 389 items
Total number of items: 3340

```

Gambar2. Cluster Model

Hasil penyebaran *cluster_0*, *cluster_1* dan *cluster_2* sebanyak 3440 data yang pemodelan *k-means clustering* dengan menggunakan *rapidminer*, untuk 3 kelompok data yang dilihat pada gambar berikut :



Gambar 3. Pemodelan K-means Clusterin pada rapidminer

Untuk kelompok data yang terdapat pada gambar diatas terdiri dari tiga kelompok data. Kelompok pertama terlihat pada titik penyebaran yang ditandai dengan warna biru, kelompok kedua terlihat titik penyebaran yang ditandai dengan warna hijau, dan kelompok ke tiga terlihat pada titik-titik penyebaran yang ditandai dengan warna orange. Adapun hasil analisa cluster pada gambar 4, berisi tentang hasil pengelompokan berdasarkan kedekatan jarak antara titik pusat dengan data mahasiswa pada setiap atribut.

tabel 7. Hasil Cluster 1 (Cluster_0)

HASIL CLUSTER SATU	
Cluster satu terdiri dari mahasiswa yang berasal dari : 897	
NAMA SEKOLAH :	NAMA DAERAH :
SMA = 610	KABUPATEN LAHAT = 153
	KABUPATEN MUSI BANYUASIN = 179
SMK = 203	KABUPATEN OGAN = 163
	KABUPATEN OGAN KOMERING ILIR = 181
LAINNYA = 84	KABUPATEN OGAN KOMERING ULU = 92
	KOTA PRABUMULIH = 131
JENIS KELAMIN	
LAKI-LAKI = 522	PEREMPUAN = 375

tabel 8. Hasil Cluster 2 (Cluster_1)

HASIL CLUSTER DUA	
Cluster satu terdiri dari mahasiswa yang berasal dari : 2054	
NAMA SEKOLAH :	NAMA DAERAH :
SMA = 1260	KABUPATEN BANYUASIN = 207
SMK = 645	KABUPATEM MUARA ENIM = 192
LAINNYA = 149	KOTA PALEMBANG = 1655
JENIS KELAMIN	
LAKI-LAKI = 1194	PEREMPUAN = 860

tabel 9. Hasil Cluster 3 (Cluster_2)

HASIL CLUSTER TIGA	
Cluster satu terdiri dari mahasiswa yang berasal dari : 389	
NAMA SEKOLAH :	NAMA DAERAH :
SMA = 290	KABUPATEN BANGKA BARAT = 7
	KABUPATEN EMPAT LAWANG = 31
	KABUPATEN MUSI RAWAS = 32
	KABUPATEN OKU SELATAN = 29
SMK = 60	KABUPATEN OKU TIMUR = 85
	KABUPATEN PALI = 63
	KOTA BENGKULU = 7
LAINNYA = 39	KOTA JAMBI = 20
	KOTA LUBUK LINGGAU = 66
	KOTA PAGAR ALAM = 37
	KOTA PANGKAL PINANG = 12
JENIS KELAMIN	
LAKI-LAKI = 236	PEREMPUAN = 153

Dari data hasil clustering yang telah dilakukan diatas, maka dapat dilihat data tersusun rapi berdasarkan peminat setiap atribut nama sekolah, daerah sekolah dan jenis kelamin. Dapat disimpulkan bahwa pada Cluster 1 nama sekolah yang diminati adalah SMA, dengan daerah sekolah dominan ke KABUPATEN OKAN KOMERING ILIR, dan jenis kelamin dominan ke LAKI-LAKI. pada Cluster 2 nama sekolah yang diminati adalah SMA, dengan daerah sekolah dominan ke KOTA PALEMBANG, dan jenis kelamin dominan ke LAKI-LAKI. Untuk Cluster 3 nama sekolah yang diminati adalah SMA, dengan daerah sekolah dominan ke KABUPATEN KU TIMUR, dan jenis kelamin dominan ke LAKI-LAKI. Dari hasil clustering diatas yang paling efektif adalah Cluster 2, karena merupakan jumlah yang terbanyak dari hasil clustering. Dari hasil tersebut dapat dilihat bahwa untuk peminat setiap daerah sekolah dari SMA dan SMK walaupun dari setiap Cluster ada beberapa daerah sekolah yang mendominasi, dan untuk jenis kelamin secara umum dari setiap cluster adalah tahun LAKI-LAKI. Dari hasil analisa tersebut dapat dipikirkan strategi promosi kampus ke sekolah-sekolah, yaitu dimaksimalkan pada sekolah SMA dan daerah sekolah yang mendominasi adalah KOTA PALEMBANG.

SIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan pada data mahasiswa baru Universitas Bina Darma Palembang, maka dapat di simpulkan bahwa:

1. Penentuan centroid (titik pusat) pada tahap awal Algoritma K-Means sangat berpengaruh pada hasil cluster seperti pada hasil pengujian yang dilakukan dengan menggunakan 3440 dataset dengan centroid yang berbeda menghasilkan hasil cluster yang berbeda juga.
2. Setelah dilakukan pengelompokan data penerimaan mahasiswa menggunakan metode k-means Clustering terbentuk tiga cluster yaitu cluster satu dengan jumlah 897 items, cluster dua dengan jumlah 2054 items, dan cluster tiga dengan jumlah 389 items.
3. Strategi promosi bagi calon mahasiswa baru nantinya akan mengikuti cluster yang terbentuk berdasarkan Daerah sekolah yang paling banyak diminati

DAFTAR PUSTAKA

- Adhinda, F., Wardani, K., & Kristiana, T. (2020). *Implementasi Data Mining Penjualan Produk Kosmetik Pada PT. Natural Nusantara Menggunakan Algoritma Apriori*. 22(1), 85–90.
- Alhapizi, M. R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. *Journal of Software Engineering Ampera*, 1(1), 1–14. <https://doi.org/10.51519/journalsea.v1i1.10>
- C.Pradeepkumar, & S.Loganathan. (2015). Penerapan Metode Asosiasi Menggunakan Algoritma Apriori Pada Aplikasi Pola Belanja Konsumen (Studi Kasus Toko Buku Gramedia Bintaro). *International Journal of*

<http://journal.uinjkt.ac.id/index.php/ti/article/view/5602/3619>

- damanik, n., & sigiro, m. (2021). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru Sebagai Metode Promosi. *Jutisal Jurnal Teknik Informatika Universal*, 4, 33–43. <https://universal.ac.id/jurnal/index.php/teknikinformatika/article/view/4>
- Genius Zendrato, F. S., Triayudi, A., & E, E. T. (2022). Analisis Clustering Dokumen Tugas Akhir Mahasiswa Sistem Informasi Universitas Nasional menggunakan Metode K-Means Clustering. *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(1), 70–76. <https://doi.org/10.35870/jtik.v6i1.389>
- Muliono, R., & Sembiring, Z. (2019). Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen. *CESS (Journal of Computer Engineering, System and Science)*, 4(2), 2502–2714.
- PRAKOSO, Y. (2019). Sistem Pendukung Keputusan Penerimaan Calon Mahasiswa Baru Menggunakan Metode K-Means Clustering Di Universitas Negeri Surabaya. *Jurnal Manajemen Informatika*, 9(2), 79–86.
- Utomo, D. P., & Mesran, M. (2020). Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung. *Jurnal Media Informatika Budidarma*, 4(2), 437. <https://doi.org/10.30865/mib.v4i2.2080>
- Yunita, F. (2018). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru. *Sistemasi*, 7(3), 238. <https://doi.org/10.32520/stmsi.v7i3.388>