



## Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang

M Reza Alhapizi<sup>1</sup>, Muhammad Nasir<sup>2</sup>, Irman Effendy<sup>3</sup>

<sup>1,2,3</sup>Information System Departement, Universitas Bina Darma, Palembang, Indonesia

Email: <sup>1</sup>reza.alhapizi1996@gmail.com, <sup>2</sup>nasir@binadarma.ac.id,

<sup>3</sup>irman.effendy@binadarma.ac.id

### Abstrak

Kemajuan luar biasa yang terus berlanjut dalam bidang data mining didorong oleh beberapa faktor salah satunya adalah pertumbuhan yang cepat dalam kumpulan data dan perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan. Data calon mahasiswa baru Universitas Bina Darma tak luput untuk dapat perhatian agar dapat dicari nilai emasnya. Banyak metode yang dapat digunakan untuk menggali nilai emas dari kumpulan data yang besar salah satunya adalah clustering. Teknik clustering dapat dijalankan dengan menggunakan salah satu Algoritma yang mudah diimplementasikan yaitu K-Means. Algoritma ini dapat dijalankan dengan menggunakan toolsRapid Miner dan hasilnya dapat dievaluasi sehingga dapat memperoleh metode promosi yang tepat.

**Kata Kunci:** clustering, k-Means, tools, Rapid Miner

### 1. PENDAHULUAN

Kemajuan teknologi informasi sekarang sudah semakin berkembang pesat dan hampir mencangkup di segala bidang kehidupan. Kemajuan tersebut menghasilkan tersedianya data yang sangat besar dan banyak mulai dari bidang industri, ekonomi, ilmu dan teknologi serta berbagai bidang kehidupan lainnya. Kemajuan luar biasa yang terus berlanjut dalam bidang data mining didorong oleh beberapa faktor (Larose, 2005) salah satunya adalah pertumbuhan yang cepat dalam kumpulan data dan perkembangan yang hebat dalam kemampuan komputasi dan pengembangan kapasitas media penyimpanan.



Seperti halnya pada proses penerimaan mahasiswa baru dalam sebuah perguruan tinggi akan menghasilkan data-data baru berupa profil dari mahasiswa baru tersebut. Tahap selanjutnya adalah mahasiswa akan melakukan kegiatan pembelajaran di setiap semester, sehingga dapat diketahui data indeks prestasi mahasiswa pada akhir semester. Hal ini akan terjadi secara berulang pada sebuah perguruan tinggi. Apabila dilakukan pengolahan data pada kedua sumber data tersebut maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu menentukan strategi promosi penerimaan mahasiswa baru tahun berikutnya. Pengolahan tersebut dapat dilakukan menggunakan salah satu metode *Data Mining* dengan algoritma *K-Means Clustering*. Menurut Pramudiono (2006), *Data mining* adalah serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual. Data yang jumlahnya besar tersebut perlu diolah oleh algoritma tersebut untuk diperoleh pengetahuan yang menunjang promosi. Hasil dari pengolahan tersebut akan sangat membantu pihak prodi atau fakultas dengan kondisi prodi atau fakultas tersebut masih belum lama berdiri, karena membutuhkan calon mahasiswa yang mempunyai potensi dalam bidang akademik dan pengaruhnya terhadap universitas adalah dapat meningkatkan jumlah mahasiswa dengan nilai akademik yang cukup tinggi.

Kondisi tersebut sama dengan kondisi yang dialami Prodi Teknik Informatika di Universitas Bina Darma Palembang. Data mahasiswa pada prodi tersebut perlu dilakukan pengolahan data dengan hasil yang menjelaskan tentang berbagai informasi-informasi mengenai persebaran wilayah mahasiswa sehingga informasi dari hasil pengolahan tersebut dapat membantu pihak Prodi Teknik Informatika Universitas Bina Darma Palembang untuk menentukan strategi promosi yang cocok dan tepat sasaran. Selain untuk membantu pihak Prodi Teknik Informatika Universitas Bina Darma Palembang, pengolahan data tersebut dilakukan agar dapat menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun berikutnya.

Atribut yang tersedia pada data calon mahasiswa adalah kota asal mahasiswa, jurusan asal sekolah mahasiswa, jenis sekolah mahasiswa, jenis kelamin mahasiswa, agama mahasiswa dan jurusan sekolah mahasiswa. Berdasarkan atribut tersebut selanjutnya data akan dikelompokkan menjadi asal kota mahasiswa, jurusan asal sekolah mahasiswa, jenis kelamin mahasiswa, agama mahasiswa dan jurusan sekolah mahasiswa menggunakan

algoritma *K-Means Cluster*. Dari penjelasan diatas, maka penulis ingin melakukan penelitian “Penerapan *DataMining* Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang”.

## 2. METODOLOGI PENELITIAN

### 2.2 Metode Pengumpulan

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 Penerapan *Data Mining* menggunakan Algoritma *K-Means Clustering* untuk menentukan strategi promosi mahasiswa baru hingga tahap akhir penelitian.

Pengamatan (Observasi), yaitu metode pengumpulan data dengan cara mengadakan tinjauan secara langsung ke objek yang diteliti. Observasi dilakukan langsung pada Universitas Bina Darma Palembang.

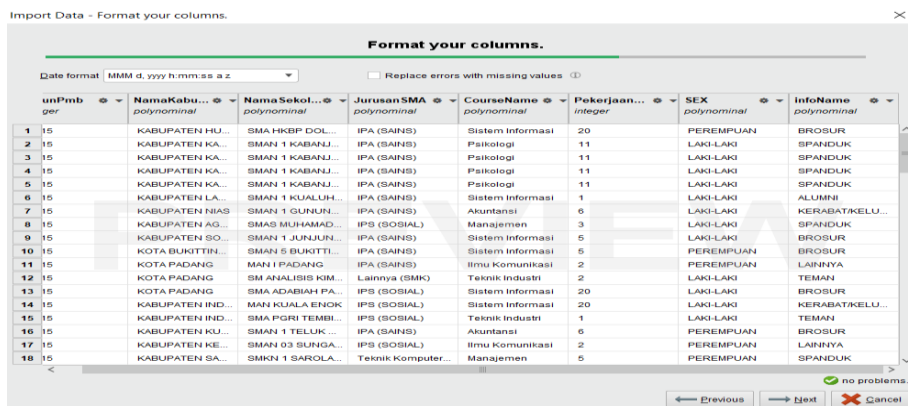
Studi Pustaka, Untuk mendapatkan data-data yang bersifat teoritis maka penulis melakukan pengumpulan data dengan cara membaca dan mempelajari buku-buku, jurnal penelitian, makalah ataupun referensi lain yang berhubungan dengan masalah yang dibahas.

### 2.2 PreProcessing

Tahap Pra pengolahan meliputi tahapan integrasi atau penggabungan data seluruh nya dan juga pembersihan data yang mana hasilnya akan lebih baik ketika dilakukan *clustering* nantinya.

- 1) **Integrasi Data**, data yang diperoleh setelah digabungkan menjadi satu tabel data yang besar sebagai satu kesatuan. Data penerimaan mahasiswa baru dari tahun 2014 sampai dengan 2017 digabung menjadi satu data besar , tidak terpisah serta baik untuk *di clustering*. Dan data yang sudah penulis terima sudah dalam bentuk sudah digabung atau terintegrasi.
- 2) **Cleaning Data**, data yang telah digabung akan dilakukan pembersihan, membuang data yang kosong dan memastikan data tersebut relevan atau, terkait satu sama lain.

Data sebelum dilakukan *cleaning* ,jika ada tampak data yang kosong untuk beberapa kolom. *Record* yang terdapat hal demikian yang layak dihapus, agar data yang diolah nanti mempunyai informasi yang baik. Data yang kami peroleh , memang memiliki sejumlah data yang kosong atau yang diisi dengan tulisan *NULL* sehingga *cleaning* sangat perlu dilakukan. Adanya kemungkinan nilai *NULL* atau kosong dikarenakan kurang lengkapnya data yang terima dari perorangan dari calon mahasiswa baru Universitas Bina Darma.:



Gambar 1. ReFormatDataset

### Penerapan Metode K-Means menggunakan Aplikasi Rapid Miner

1. Statistik Data Penerimaan Mahasiswa Baru Jurusan Teknik Informatika. Setelah data berhasil dilakukan *import*, rapid miner secara otomatis telah melakukan analisa statistika. Hasil statistik data penerimaan mahasiswa baru yang berhasil diperoleh adalah sebagai berikut:

a. Atribut InfoName

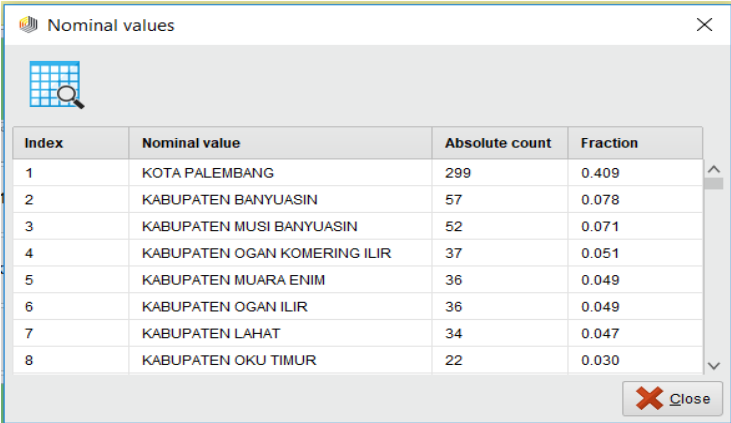
Berdasarkan data statistik diperoleh yang tertinggi jatuh pada Brosur. Mayoritas calon mahasiswa baru mengetahui Universitas Bina Darma dari Brosur. Sedangkan yang terendah yaitu informasi dari twitter.

Index	Nominal value	Absolute count	Fraction
1	BROSUR	186	0.254
2	KERABATKELUARGA	173	0.237
3	TEMAN	126	0.172
4	LAINNYA	93	0.127
5	ALUMNI	80	0.109
6	DOSEN / KARYAWAN B...	24	0.033
7	SPANDUK	23	0.031
8	KORAN / MAJALAH	12	0.016

Gambar 2. Data dan Ringkasan Statistikan Atribut InfoName

b. Atribut NamaKabupaten

Data statistik berkata bahwa calon mahasiswa/i paling banyak diperoleh dari orang-orang yang berasal dari Palembang. Dan yang paling sedikit berasal dari kota pematang siantar. Berikut ringkasan statistik jumlah mahasiswa/i berdasarkan NamaKabupaten:

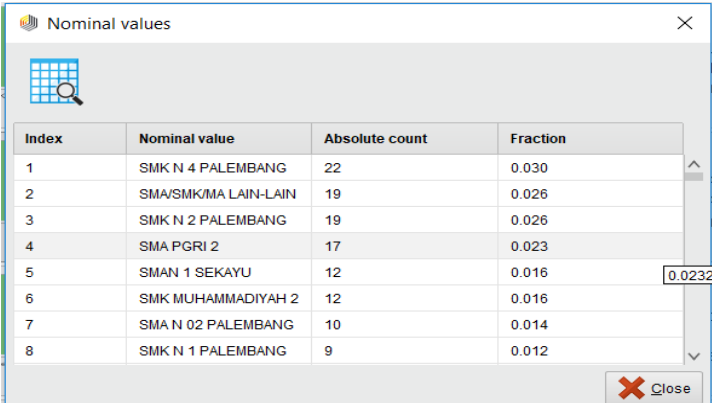


Index	Nominal value	Absolute count	Fraction
1	KOTA PALEMBANG	299	0.409
2	KABUPATEN BANYUASIN	57	0.078
3	KABUPATEN MUSI BANYUASIN	52	0.071
4	KABUPATEN OGAN KOMERING ILIR	37	0.051
5	KABUPATEN MUARA ENIM	36	0.049
6	KABUPATEN OGAN ILIR	36	0.049
7	KABUPATEN LAHAT	34	0.047
8	KABUPATEN OKU TIMUR	22	0.030

Gambar 3. Statistik Jumlah Mahasiswa Baru Berdasarkan kabupaten mereka berasal

c. Atribut Asal Sekolah

Data statistik kali ini menunjukkan sesuatu yang sedikit berbeda, dari data yang diperoleh, dari mayoritas data, kebanyakan asal sekolah dari kota Palembang. Dan yang paling sedikit adalah berasal dari smk batam dan jambi. Berikut data statistik secara ringkas dan menyeluruh:



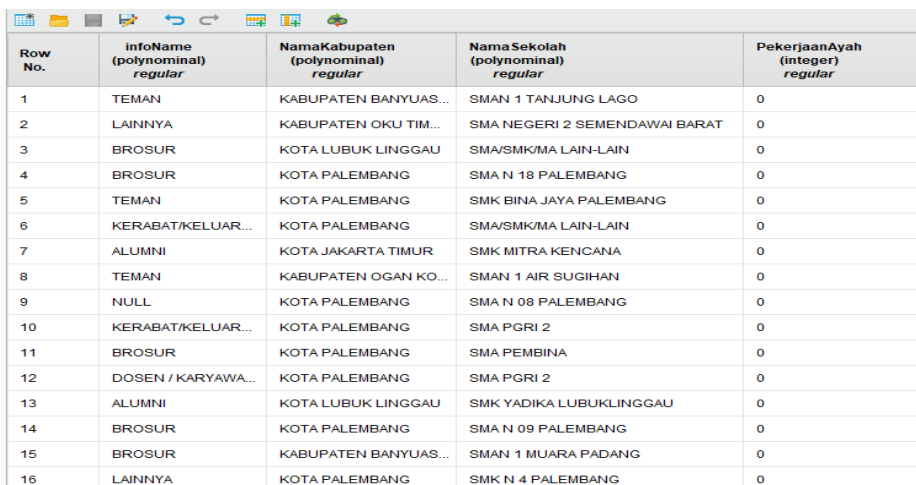
Index	Nominal value	Absolute count	Fraction
1	SMK N 4 PALEMBANG	22	0.030
2	SMA/SMK/MA LAIN-LAIN	19	0.026
3	SMK N 2 PALEMBANG	19	0.026
4	SMA PGRI 2	17	0.023
5	SMAN 1 SEKAYU	12	0.016
6	SMK MUHAMMADIYAH 2	12	0.016
7	SMA N 02 PALEMBANG	10	0.014
8	SMK N 1 PALEMBANG	9	0.012

Gambar 4. Ringkasan jumlah mahasiswa baru berdasarkan asal sekolah

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Centroid K-Means

Atribut fitur yang digunakan pada pengujian ini adalah infoName, NamaKabupaten, NamaSekolah, PekerjaanAyah. Atribut ini atribut yang merupakan bagian referensi penulis untuk menentukan strategi promosi yang tepat untuk ke depannya setelah melihat hasil cluster nantinya. Berikut tampilan data terakhir yang digunakan:



Row No.	InfoName (polynominal) regular	NamaKabupaten (polynominal) regular	NamaSekolah (polynominal) regular	PekerjaanAyah (integer) regular
1	TEMAN	KABUPATEN BANYUAS...	SMAN 1 TANJUNG LAGO	0
2	LAINNYA	KABUPATEN OKU TIM...	SMA NEGERI 2 SEMENDAWAI BARAT	0
3	BROSUR	KOTA LUBUK LINGGAU	SMA/SMK/MA LAIN-LAIN	0
4	BROSUR	KOTA PALEMBANG	SMA N 18 PALEMBANG	0
5	TEMAN	KOTA PALEMBANG	SMK BINA JAYA PALEMBANG	0
6	KERABAT/KELUAR...	KOTA PALEMBANG	SMA/SMK/MA LAIN-LAIN	0
7	ALUMNI	KOTA JAKARTA TIMUR	SMK MITRA KENCANA	0
8	TEMAN	KABUPATEN OGAN KO...	SMAN 1 AIR SUGIHAN	0
9	NULL	KOTA PALEMBANG	SMA N 08 PALEMBANG	0
10	KERABAT/KELUAR...	KOTA PALEMBANG	SMA PGRI 2	0
11	BROSUR	KOTA PALEMBANG	SMA PEMBINA	0
12	DOSEN / KARYAWA...	KOTA PALEMBANG	SMA PGRI 2	0
13	ALUMNI	KOTA LUBUK LINGGAU	SMK YADIKA LUBUKLINGGAU	0
14	BROSUR	KOTA PALEMBANG	SMA N 09 PALEMBANG	0
15	BROSUR	KABUPATEN BANYUAS...	SMAN 1 MUARA PADANG	0
16	LAINNYA	KOTA PALEMBANG	SMK N 4 PALEMBANG	0

Gambar 5. Atribut terpilih untuk dijadikan atribut fitur bagi metode K-Means

#### 3.2 Analisa Algoritma K-Means

Menurut Santosa (2007), langkah-langkah melakukan Clustering dengan metode K-Means adalah sebagai berikut:

- 1) Pilih jumlah *Cluster*  $K$
- 2) Inisialisasi  $K$  pusat *Cluster* ini bisa di lakukan dengan berbagai cara. Namun yang palig sering di lakukan adalah dengan cara random. Pusat-pusat *Cluster* di beri nilai awal dengan angka-angka random
- 3) Alokasikan semua data atau objek ke *Cluster* terdekat. Kedekatan dua objek di tentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *Cluster* tertentu di tentukan jarak antara data dengan pusat *Cluster*. Jarak paling antara satu data dengan satu *Cluster* tertentu akan menentukan suatu data masuk dalam *Cluster* mana.

Untuk menghitung jarak semua data ke setiap titik pusat *Cluster* dapat menggunakan teori jarak *Euclidean* yang di rumuskan sebagai berikut:

$$D(i,j) = \sqrt{(X_{xi} - X_{1j}) + (X_{2i} - X_{2j}) + \dots + (X_{ki} - X_{kj})} \dots\dots\dots(2.1)$$

Dimana:

$D(i,j)$  = jarak data ke  $i$  ke pusat *cluster*  $j$

$X_{ki}$  = Data ke  $i$  atribut data ke  $k$

$X_{kj}$  = Titik pusat ke  $j$  pada atribut  $k$

- 1) Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data atau objek dalam *cluster* tertentu. Jika di kehendaki bisa juga menggunakan medan dari *cluster* tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa di pakai.
- 2) Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah no 3 sampai pusat *cluster* tidak berubah lagi.

No	NamaKabupaten	NamaSekolah	PelajaranAyah	infoName
1	KOTA PALEMBANG	SMA PGRI 2	.0	KERABAT/KELUARGA
2	KABUPATEN BANYUASIN	SMA SATHIA NUSANTARA BETUNG	1.0	TEMAN
3	KABUPATEN LAHAT	SMAN 2 LAHAT	1.0	BROSUR
4	KABUPATEN MUARA ENIM	SMA PGRI GELUMBANG	1.0	LAINNYA
5	KABUPATEN MUARA ENIM	SMKN 1 MUARA ENIM	1.0	KORAN /MAJALAH
6	KABUPATEN MUARA ENIM	SMKN 1 MUARA ENIM	1.0	TEMAN
7	KABUPATEN MUARA ENIM	SMKN 2 MUARA ENIM	1.0	TEMAN
8	KABUPATEN MUSI BANYUASIN	SMAN 1 SEKAYU	1.0	TEMAN
9	KABUPATEN MUSI BANYUASIN	SMAN 1 SEKAYU	1.0	TEMAN
10	KABUPATEN MUSI BANYUASIN	SMAN 1 SEKAYU	1.0	KERABAT/KELUARGA
11	KABUPATEN MUSI BANYUASIN	SMAN 1 SUNGAILILIN	1.0	BROSUR
12	KABUPATEN MUSI BANYUASIN	SMKN 2 SEKAYU	1.0	TEMAN
13	KABUPATEN OGAN ILIR	SMKN INDRALAYA SELATAN	1.0	KERABAT/KELUARGA
14	KABUPATEN OGAN KOMERING ILIR	SMAN 1 PAMPANGAN	1.0	KERABAT/KELUARGA
15	KABUPATEN OGAN KOMERING ILIR	SMAN 2 KAYUAGUNG	1.0	ALUMNI
...	...	...	..	...
730	KOTA PALEMBANG	SMA/SMK/MA LAIN-LAIN	20.0	BROSUR
731	KOTA PALEMBANG	SMA SYAKYAKIRTI	20.0	LAINNYA

Gambar 6. Data Calon Mahasiswa Universitas Bina Darma

Untuk melakukan perhitungan K-Means, data yang dapat diolah harus dalam bentuk angka (numerik). Jika data masih dalam bentuk polynomial atau nominal seperti data pada atribut NamaKabupaten, NamaSekolah dan InfoName, maka perlu dibuat nilai permisalan atau inisialisasi nilai baru. Maka dari itu langkah selanjutnya adalah melakukan inisialisasi data. Inisialisasi data dilakukan berdasarkan frekuensi data yang terdapat dalam setiap instansi dari atribut. Inisialisasi data dari atribut NamaKabupaten adalah sebagai berikut:

No	NamaKabupaten	Frekuensi	Inisial Nilai
1	KOTA PALEMBANG	299.0	1
2	KABUPATEN BANYUASIN	57.0	2
3	KABUPATEN MUSI BANYUASIN	52.0	3
4	KABUPATEN OGAN KOMERING ILIR	37.0	4
5	KABUPATEN MUARA ENIM	36.0	5
6	KABUPATEN OGAN ILIR	36.0	6
7	KABUPATEN LAHAT	34.0	7
8	KABUPATEN OKU TIMUR	22.0	8
9	KABUPATEN PALI	21.0	9
10	KOTA PRABUMULIH	21.0	10
11	KOTA PAGAR ALAM	18.0	11
12	KABUPATEN OGAN KOMERING ULU	17.0	12
13	KABUPATEN OKU SELATAN	16.0	13
14	KOTA LUBUK LINGGAU	13.0	14
15	KABUPATEN EMPATLAWANG	8.0	15

Gambar 7. Data Nama Kabupaten dan Frekuensi Kejadian

Kota Palembang memiliki frekuensi tertinggi dalam data yaitu memiliki 299 kejadian, maka dari itu diberi nilai 1. Kabupaten Banyuasin memiliki peringkat kedua terbanyak kemudian diberi nilai inisial 2 dan berikut seterusnya. Setelah NamaKabupaten selanjutnya yang perlu diinisialisasi adalah Nama sekolah. Inisialisasi data dari atribut NamaSekolah adalah sebagai berikut:

No	Nama Sekolah	Frekuensi	Inisial Data
1	SMKN 4 PALEMBANG	22.0	1
2	SMA/SMK/MA LAIN-LAIN	19.0	2
3	SMKN 2 PALEMBANG	19.0	3
4	SMA PGRI 2	17.0	4
5	SMAN 1 SEKAYU	12.0	5
6	SMK MUHAMMADIYAH 2	12.0	6
7	SMA N 02 PALEMBANG	10.0	7
8	SMKN 1 PALEMBANG	9.0	8
9	SMKN 1 MUARA ENIM	9.0	9
10	SMA N 04 PALEMBANG	8.0	10

Gambar 7. Data Nama Sekolah dan Frekuensi Kejadian



Dan yang terakhir adalah inialisasi atribut InfoName. Atribut penghasil sudah terlebih dahulu nilainya berada dalam nilai numerik sehingga tidak dibutuhkan proses inialisasi. Inisialisasi data untuk atribut InfoName adalah sebagai berikut:

No	infoName	Frekuensi	Inisial Data
1	BROSUR	186.0	1
2	KERABAT/KELUARGA	173.0	2
3	TEMAN	126.0	3
4	LAINNYA	93.0	4
5	ALUMNI	80.0	5
6	DOSEN / KARYAWAN BINA DARMA	24.0	6
7	SPANDUK	23.0	7
8	KORAN / MAJALAH	12.0	8
9	INSTAGRAM	5.0	9
10	TELEVISI	5.0	10
11	FACEBOOK	3.0	11
12	RADIO	1.0	12

**Gambar 8.** Data InfoName dan Frekuensi Kejadian

Berdasarkan inisialisasi data dari ketiga atribut yaitu NamaKabupaten, NamaSekolah, InfoName sebagaimana di atas maka data yang akan diproses oleh metode K-Means menjadi sebagai berikut:

No	Nama Kabupaten	Nama Sekolah	Pekerjaan Ayah	InfoName
1	1	4	0	2
2	2	212	1	3
3	7	11	1	1
4	5	206	1	4
5	5	9	1	8
6	5	9	1	3
7	5	130	1	3
8	3	5	1	3
9	3	5	1	3
10	3	5	1	2
11	3	261	1	1
12	3	46	1	3
13	6	32	1	2
14	4	247	1	2
15	4	110	1	5
16	4	274	1	5
17	4	115	1	3
18	13	100	1	1
19	14	59	1	1
20	11	301	1	2

**Gambar 9.** Data Calon Mahasiswa Setelah diinisialisasi

Setelah data berhasil diubah menjadi tipe numerik maka langkah-langkah atau tahapan-tahapan K-Means dapat diterapkan sebagai berikut:

- 1) Menentukan Jumlah *Cluster*, peneliti memilih 2 *cluster* ( $k=2$ ) yang akan dibuat. Banyaknya *cluster* harus lebih kecil dari pada banyaknya data ( $k < n$ ).
- 2) Menentukan Centroid Tiap *Cluster*, Untuk menentukan *centroid* awal (*initial centroid*) banyak metode yang dapat digunakan. Disini metode

yang digunakan adalah mengambil data dari data sumber, secara acak atau random. Berikut data *centroid* tersebut.

	NamaKabupaten	NamaSekolah	PenghasilanAyah	InfoName
Centroid 1 (c1)	1	1	1	4
Centroid 2 (c2)	29	332	1	8

**Gambar 10.** Data *Inisial Centroid*

- a. Menghitung jarak data dengan *centroid*, Untuk menghitung jarak data dengan *centroid* menggunakan rumus euclidean *distance*. Rumus Euclidean *Distance* adalah sebagai berikut:

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

$d$  = jarak  
 $j$  = banyaknya data  
 $c$  = *centroid*  
 $x$  = data  
 $c$  = *centroid*

Jarak data dengan *centroid* 1 adalah:

$$\begin{aligned} d(1, 1) &= \sqrt{(x1 - p1)^2 + (x2 - p2)^2 + (x3 - p3)^2 + (x4 - p4)^2} \\ &= \sqrt{(1 - 1)^2 + (4 - 1)^2 + (0 - 1)^2 + (2 - 4)^2} \\ &= 3.7416 \end{aligned}$$

$$\begin{aligned} d(2, 1) &= \sqrt{(x1 - p1)^2 + (x2 - p2)^2 + (x3 - p3)^2 + (x4 - p4)^2} \\ &= \sqrt{(2 - 1)^2 + (212 - 1)^2 + (1 - 1)^2 + (3 - 4)^2} \\ &= 211.004 \end{aligned}$$

Dan seterusnya sampai data ke 731, Selanjutnya dihitung jarak data dengan *centroid* 2 adalah:

$$\begin{aligned} d(1, 2) &= \sqrt{(x1 - p1)^2 + (x2 - p2)^2 + (x3 - p3)^2 + (x4 - p4)^2} \\ &= \sqrt{(1 - 29)^2 + (4 - 332)^2 + (0 - 1)^2 + (2 - 8)^2} \\ &= 329.2491 \end{aligned}$$

$$\begin{aligned} d(2, 2) &= \sqrt{(x1 - p1)^2 + (x2 - p2)^2 + (x3 - p3)^2 + (x4 - p4)^2} \\ &= \sqrt{(4 - 29)^2 + (3 - 332)^2 + (1 - 1)^2 + (6 - 8)^2} \\ &= 123.101 \end{aligned}$$

Dan seterusnya sampai data ke 731

b. Hasil Perhitungan Jarak

No	Jarak C1	Jarak C2
1	3.741657387	329.2491458
2	211.0047393	123.1015841
3	12.04159458	321.8291472
4	205.0390207	128.3277055
5	9.797958971	323.8904136
6	9	323.9290046
7	129.0658747	203.482186
8	4.582575695	328.0701145
9	4.582575695	328.0701145
10	4.898979486	328.0868787
11	260.0249988	75.93418203
12	45.0555213	287.2229099
13	31.46426545	300.9401934
14	246.0264213	88.8031531
15	109.0458619	223.4233649
16	273.0183144	63.22973984
17	114.0438512	218.4925628
18	99.76973489	232.656399
19	59.51470407	273.5013711
20	300.1732833	36.34556369

Gambar 11. Data jarak *centroid* terhadap data

Hasil jarak tersebut digunakan untuk menentukan pada cluster mana suatu data terkelompok. Data dikelompokkan menurut nilai jarak terkecil. Misal data1 jarak terkecil jatuh pada C1 maka data 1 termasuk anggota cluster 1. Maka Data Sekarang setelah berhasil dipetakan ke dalam cluster menurut jarak yang sudah diperoleh.

Tabel 1. Cluster menurut jarak

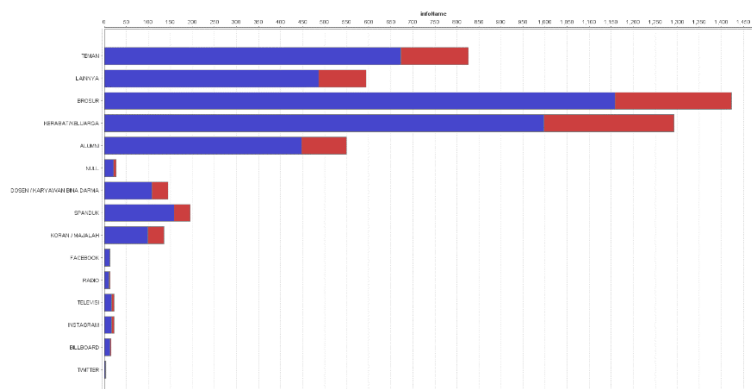
No	Nama Kabupaten	Nama Sekolah	Pekerjaan Ayah	InfoName	Cluster 1	Cluster 2
1	1	4	0	2	*	
2	2	212	1	3		*
3	7	11	1	1	*	
4	5	206	1	4		*
5	5	9	1	8	*	
6	5	9	1	3	*	
7	5	130	1	3	*	
8	3	5	1	3	*	
9	3	5	1	3	*	

10	3	5	1	2	*	
11	3	261	1	1		*
12	3	46	1	3	*	
13	6	32	1	2	*	
14	4	247	1	2		*
15	4	110	1	5	*	
16	4	274	1	5		*
17	4	115	1	3	*	
18	13	100	1	1	*	
19	14	59	1	1	*	
20	11	301	1	2		*

Cluster 0 memiliki lebih dari 600 anggota, dan cluster 1 memiliki lebih dari 100 anggota. Dengan melihat hasil pada tabel tersebut diperoleh total anggota Cluster 0 lebih banyak dari total anggota cluster 1, dan hasil ini tidaklah berbeda dengan hasil yang diolah dari tools rapid miner.

c. Hasil evaluasi (*Evaluation Result*)

Setelah melihat hasil dari cluster model diperoleh pada cluster pertama yaitu cluster 0 memiliki 616 anggota yang terkelompok. Sedangkan cluster kedua yaitu cluster 1 memiliki 115 anggota yang terkelompok. Jika dilihat cluster pertama (cluster 0) memiliki jumlah anggota yang sangat banyak menandakan orang-orang yang mendapatkan informasi dari brosur,teman, kerabat dan lain sebagainya. Sedangkan cluster kedua (cluster 1) memiliki paling sedikit anggota yang terkelompok, jika dilihat dari kelompok datanya cluster 0 cluster yang mayoritas mahasiswa/l diperoleh dari strategi brosur, teman dan kerabat.



Gambar 12. Visualisasi Pembagian Anggota Cluster

Sudah terlihat sekali bahwa ketiga strategi ini adalah strategi yang paling baik dalam mempromosikan lembaga agar memperoleh lebih banyak lagi mahasiswa/i yang meniti ilmu di Universitas Bina Darma. Sehingga yang dapat disimpulkan dari keanggotaan *cluster* tersebut didapati strategi yang bagus untuk diterapkan sekarang dan seterusnya adalah

- 1) Menyebar brosur, membuat desain yang menarik dan menyebarkannya ke berbagai macam tempat, tentu memiliki dampak yang besar bagi banyaknya jumlah penerimaan mahasiswa baru yang akan datang.
- 2) Promosi ke berbagai daerah, dengan menyebarkan informasi ke daerah-daerah, memungkinkan para calon mahasiswa/i mengetahui informasi tentang Universitas Bina Darma dengan lebih baik.

#### 4. KESIMPULAN

Akhirnya sudah sampai kita pada akhir dari laporan. Semua langkah-langkah yang telah diarahkan pada metode CRISM-DM dan hasil dari menggunakan metode clustering *K-Means* maka jatuhlah kepada suatu kesimpulan sebagai berikut:

- 1) Setelah dilakukan pengelompokan data mahasiswa berdasarkan persebaran kabupaten, asal sekolah, penghasilan ayah dan informasi maka diperoleh suatu gagasan bahwa cluster yang jumlahnya terbanyak adalah strategi promosinya dengan menggunakan brosur dan promosi dari teman atau kerabat dekat.
- 2) Fungsi media sosial belum begitu efektif dilihat dari frekuensi statistik calon mahasiswa baru dari instansi atribut ini.
- 3) Strategi promosi yang selanjutnya perlu dipertahankan adalah brosur, berikutnya informasi dari teman atau kerabat dan keikutsertaan karyawan atau dosen dalam mempromosikannya. Dan Media sosial seharusnya menjadi sumber informasi terbaru dan terluas bagi calon mahasiswa baru nantinya.

#### DAFTAR PUSTAKA

- Fauziah Nur, Dkk. 2017. "*Penerapan Algoritma K-MEANS Pada Siswa Baru Sekolah Menengah Kejuruan Untuk Clustering Jurusan*". <https://jurnal.uisu.ac.id/index.php/infotekjar/article/view/70> (4 April 2018)
- Y. Agusta, "*K-Means Penerapan, Permasalahan dan Metode Terkait*," Jurnal Sistem dan Informatika, vol. 3, pp. 47-60, Pebruari 2007.

- Jiawei Han, Micheline Kamber. 2011. *Data Mining: Concepts and Techniques*, 3<sup>rd</sup> ed. *The Morgan Kaufmann Series in Data Management Systems*. Morgan Kaufmann
- Connolly, Thomas dan Begg, Carolyn. (2010). *Database Systems A Practical Approach to Design, Implementation, and Management*. Fifth Edition. Boston: Pearson Education.
- Kurt Thearling, dkk. 1999. *Building Data Mining Applications for CRM*. McGraw-Hill, Inc., New
- Dunham, Margaret H. 2003. *"Data Mining: Introductory and Advanced Topics"*. Pearson Education, Inc.
- Mehmed Kantardzic. 2012. *"Data Mining: Concepts, Models, Methods, and Algorithms"*. Wiley-IEEE Press
- Vercellis, Bernadth. (2009), *Sistem Informasi*, Lokomedia, Yogyakarta
- Larose D, T., 2005, *Discovering knowledge in data : an introduction to data mining*, Jhon Wiley & Sons Inc
- Charles Lamb, W.et.al. 2001. *Pemasaran. Edisi Pertama*, Salemba Empat: Jakarta.
- Kotler, Philip and Gary Armstrong. 2012. *Prinsip-prinsip Pemasaran. Edisi 13 Jilid I*. Jakarta: Erlangga.