

DATA MINING MENGGUNAKAN ALGORITMA *K-MEANS* CLUSTERING UNTUK MENENTUKAN STRATEGI PROMOSI UNIVERSITAS DIAN NUSWANTORO

Oleh: Rima Dias Ramadhani
Jurusan Sistem Informasi, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro
Email: rimadias93@gmail.com

ABSTRAK

Proses penerimaan mahasiswa baru Universitas Dian Nuswantoro menghasilkan data mahasiswa yang sangat berlimpah berupa data profil mahasiswa dan data kegiatan belajar mengajar. Hal tersebut terjadi secara berulang dan menimbulkan penumpukan terhadap data mahasiswa, sehingga mempengaruhi pencarian informasi terhadap data tersebut. Penelitian ini bertujuan untuk melakukan pengelompokan terhadap data mahasiswa Universitas Dian Nuswantoro dengan memanfaatkan proses data mining dengan menggunakan teknik *Clustering*. Metode yang digunakan adalah CRISP-DM dengan melalui proses *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment*. Algoritma yang digunakan untuk pembentukan *cluster* adalah algoritma *K-Means*. *K-Means* merupakan salah satu metode data *non-hierarchical clustering* yang dapat mengelompokkan data mahasiswa ke dalam beberapa *cluster* berdasarkan kemiripan dari data tersebut, sehingga data mahasiswa yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain. Implementasi menggunakan *RapidMiner 5.3* digunakan untuk membantu menemukan nilai yang akurat. Atribut yang digunakan adalah kota asal, program studi dan IPK mahasiswa. *Cluster* mahasiswa yang terbentuk adalah tiga *cluster*, dengan *cluster* pertama 804 mahasiswa, *cluster* kedua 2792 mahasiswa dan *cluster* ketiga sejumlah 223 mahasiswa. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan *cluster* yang terbentuk oleh pihak admisi UDINUS.

Kata kunci: kota asal, program studi, IPK, *k-means clustering*, strategi promosi

ABSTRACT

New admissions process Dian Nuswantoro University students generate data that are highly abundant in the form of student profile data and the data of study. This happens over and over and cause a build up of the student data, there by affecting the information retrieval of the data. This study aims to perform the data clustering Dian Nuswantoro University students by utilizing data mining process using clustering techniques. The method used is the CRISP-DM with a through understanding of business processes, understanding the data, the data preparation, modeling, evaluation and deployment. The algorithm used for the formation of clusters is K-Means algorithm. K-Means is one method of non-hierarchical clustering of data that can group student data into several clusters based on the similarity of the data, so the data of students who have similar characteristics are grouped into one cluster and that have different characteristics grouped in another cluster. Implementation using *RapidMiner 5.3* is used to help find an accurate value. The attribute used is the county, course and GPA student. Cluster students are three clusters formed, with the first cluster of 804 students, the second cluster of students in 2792 and the third cluster number of 223 students. The results of this study are used as a basis for decision making to determine promotional strategies based cluster formed by the admission UDINUS.

Keywords: county, course and GPA student, k-means clustering, promotional strategies

1. PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi informasi sudah semakin berkembang pesat disegala bidang kehidupan. Banyak sekali data yang dihasilkan oleh teknologi informasi yang canggih, mulai dari bidang industri, ekonomi, ilmu dan teknologi serta berbagai bidang kehidupan lainnya. Penerapan teknologi informasi dalam dunia pendidikan juga dapat menghasilkan data yang berlimpah mengenai siswa dan proses pembelajaran yang dihasilkan.

Pada institusi pendidikan perguruan tinggi, data dapat diperoleh berdasarkan data historis, sehingga data akan bertambah secara terus menerus, misalnya data mahasiswa. Proses penerimaan mahasiswa baru dalam sebuah perguruan tinggi menghasilkan data yang berlimpah berupa profil dari mahasiswa baru tersebut. Tahap selanjutnya mahasiswa akan melakukan kegiatan belajar mengajar, sehingga dapat diketahui data prestasi mereka setiap akhir semester. Hal ini akan terjadi secara berulang pada sebuah perguruan tinggi. Penumpukan data mahasiswa secara menerus akan memperlambat pencarian informasi terhadap data tersebut.

Berdasarkan berlimpahnya data mahasiswa, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data tersebut sehingga berguna bagi pihak universitas [1]. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (*knowledge discovery*), misalnya informasi mengenai pengelompokan data mahasiswa berpotensi berdasarkan kota asal mahasiswa [1]. Pengetahuan baru tersebut dapat membantu pihak universitas dalam menentukan strategi promosi terhadap calon mahasiswa baru dengan tepat sasaran.

Berdasarkan data yang diperoleh dari bagian UPT Data dan Informasi (PSI) UDINUS, pada tahun 2010 jumlah mahasiswa baru UDINUS berjumlah 2343 mahasiswa, pada tahun 2011 berjumlah 2448 mahasiswa, pada tahun 2012 berjumlah 2707 sedangkan jumlah mahasiswa baru tahun 2013 berjumlah 2277 mahasiswa. Pada tahun 2010 hingga 2011 jumlah mahasiswa baru pada masing-masing program studi mengalami peningkatan sejumlah 105 mahasiswa. Pada tahun 2011 hingga 2012 jumlah mahasiswa baru pada masing-masing program studi juga mengalami

peningkatan sejumlah 259 mahasiswa. Pada tahun 2013 terjadi penurunan jumlah mahasiswa baru pada masing-masing program studi yaitu sejumlah 430 mahasiswa.

Salah satu penyebab terjadinya penurunan jumlah mahasiswa baru pada tahun 2013 adalah kurang dilakukan pengolahan data mahasiswa secara tepat berdasarkan data historis oleh pihak admisi UDINUS. Hal tersebut dapat mempengaruhi pengambilan keputusan dalam menentukan wilayah promosi yang tepat sasaran. Pengolahan data mahasiswa seharusnya dilakukan agar dapat menentukan wilayah promosi yang tepat sasaran sehingga tidak terjadi penurunan jumlah mahasiswa pada tahun berikutnya.

Penelitian ini akan melakukan pengelompokan berdasarkan pada data mahasiswa UDINUS. Atribut yang akan digunakan dalam penentuan pengelompokan wilayah untuk melakukan promosi UDINUS adalah program studi mahasiswa, kota asal mahasiswa dan Indeks Prestasi Mahasiswa (IPK) mahasiswa.

1.2 Tujuan Penelitian

Tujuan dari penelitian yang dilakukan adalah:

- Mengetahui persebaran wilayah berdasarkan data mahasiswa tahun 2012, 2011 dan 2012 berdasarkan potensi akademik mahasiswa.
- Menentukan strategi promosi yang tepat untuk setiap *cluster* yang terbentuk.

2 LANDASAN TEORI

2.1 Promosi

Menurut Kotler, *promotion, the fourth marketing mix tools, stand for various activities, the company undertakes to communicate its products merits and to persuade target customers to buy them*. Definisi tersebut mempunyai pengertian bahwa promosi meliputi semua alat yang terdapat dalam bauran promosi yang peranan utamanya adalah mengadakan komunikasi yang bersifat membujuk [7].

2.1.1 Strategi Promosi

Menurut Kotler & Armstrong variabel-variabel yang ada di dalam *promotional mix* ada lima, yaitu [8]:

- Periklanan (*advertising*)
- Penjualan Personal (*personal selling*)
- Promosi penjualan (*sales promotion*)

- d. Hubungan masyarakat (*public relation*)
- e. Pemasaran langsung (*direct marketing*)

2.2 Penerimaan Mahasiswa Baru

Penerimaan Mahasiswa baru di UDINUS dilakukan melalui 4 tahap, yaitu [9]:

- a. Jalur khusus (vokasi)
- b. Jalur PMDK
- c. Program pindahan
- d. Program transfer
- e. Program reguler

2.3 Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar [10].

2.4 Pengelompokan Data Mining

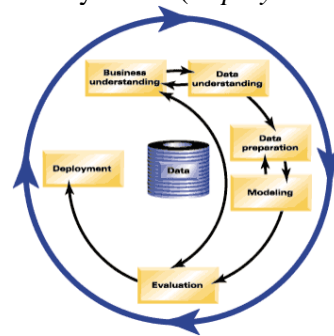
Menurut Larose, *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat di lakukan, yaitu [11]:

- a. Deskripsi
Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
- b. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik dari pada ke arah kategori.
- c. Prediksi
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa mendatang.
- d. Klasifikasi
Dalam klasifikasi, terdapat target variabel kategori.
- e. Pengklusteran
Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan (*unsupervised*).
- f. Asosiasi
Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam suatu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

2.5 Metode Data Mining

Menurut Larose, data mining memiliki enam fase CRISP-DM (*Cross Industry Standard Process for Data Mining*) [11].

- a. Fase Pemahaman Bisnis (*Business Understanding Phase*)
- b. Fase Pemahaman Data (*Data Understanding Phase*)
- c. Fase Pengolahan Data (*Data Preparation Phase*)
- d. Fase Pemodelan (*Modeling Phase*)
- e. Fase Evaluasi (*Evaluation Phase*)
- f. Fase Penyebaran (*Deployment Phase*)



Gambar 2.1 Proses CRISP-DM [11]

2.6 Algoritma K-Means

Menurut Hasn & Kamber algoritma *K-Means* bekerja dengan membagi data ke dalam k buah *cluster* yang telah ditentukan [12].

Beberapa cara penghitungan jarak yang biasa digunakan yaitu:

- *Euclidean distance*
Formula jarak antar dua titik dalam satu, dua dan tiga dimensi secara berurutan ditunjukkan pada formula 1, 2, 3 berikut ini :

$$\sqrt{(x - y)^2} = |x - y| \quad (2-1)$$

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

$$d(p, q) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2} \quad (2-3)$$
- *Manhattan Distance*
Manhattan distance disebut juga *taxicab distance*.

$$d1(p, q) = ||p - q||_1 = \sum_{i=1}^n |pi - q1| \quad (2-4)$$
- *Chebichev Distance*
Di dalam *Chebichev distance* atau *Maximum Metric* jarak antar titik didefinisikan dengan cara mengambil nilai selisih terbesar dari tiap koordinat dimensinya.

$$D_{cheb}(p, q) = \max(|pi - q1|) \quad (2-5)$$

2.7 RapidMiner

RapidMiner adalah sebuah lingkungan *machine learning data mining*, *text mining* dan *predictive analytics* [13].

3 HASIL DAN PEMBAHASAN

3.1 Pemahaman Bisnis (*Business Understanding*)

Pemahaman bisnis (*business understanding*), tahap pertama dalam proses CRISP-DM yang juga dapat disebut sebagai tahap pemahaman bisnis (penelitian).

3.1.1 Menentukan Tujuan Bisnis

Tujuan bisnis berdasarkan pengolahan data mahasiswa tahun 2009, 2010, 2011, 2012 dan 2013 antara lain untuk meningkatkan jumlah mahasiswa pada tahun 2014. Pengolahan data dilakukan karena telah terjadi penurunan sejumlah 430 mahasiswa pada tahun 2013. Tujuan pengolahan data dapat dijadikan sebagai salah satu dasar pengambilan keputusan untuk menentukan wilayah promosi yang tepat sasaran.

3.1.2 Melakukan Penilaian Situasi

Pada tahapan ini dibutuhkan pemahaman terhadap tujuan bisnis dan menerjemahkan ke dalam tujuan *data mining*.

3.1.3 Menentukan strategi awal *data mining*

Strategi awal dalam menerapkan tujuan dilakukannya *data mining* adalah melakukan permintaan data mahasiswa terlebih dahulu ke bagian UPT Data dan Informasi (PSI) UDINUS.

3.2 Pemahaman Data (*Data Understanding*)

Dataset mahasiswa yang didapatkan dari UPT Data dan Informasi (PSI) UDINUS berupa dokumen *excel* sejumlah 11997 *record*.

3.2.1 Pengumpulan data awal

Adapun sumber data utama yang digunakan dalam penelitian ini adalah *dataset* mahasiswa UDINUS yang berjenjang DIII dan S1 pada tahun 2009, 2010, 2011, 2012 dan 2013 dengan format *.xlsx* karena data yang diberikan berupa dokumen *excel*.

3.2.2 Mendeskripsikan data

Dataset mahasiswa terdiri dari atribut NIM, nama, program studi, alamat asal, nama provinsi, kota asal, tempat lahir, jenis kelamin, sekolah asal, alamat sekolah, IP Semester 1 & 2 serta IPK mahasiswa. Jumlah data yang ada pada atribut berjumlah 11997 *record*.

3.2.3 Evaluasi kualitas data

Hasil evaluasi terhadap kualitas data yaitu menemukan banyak nilai kosong / *null*

yang disebut dengan *missing value* pada atribut dalam *dataset* mahasiswa.

3.2.4 Pemilihan Atribut

Atribut yang digunakan adalah program studi mahasiswa, kota asal mahasiswa dan IPK mahasiswa dengan standar $\geq 2,75$.

3.3 Persiapan Data (*Data Preparation*)

Persiapan data mencakup semua kegiatan untuk membangun *dataset* mahasiswa yang akan diterapkan ke dalam alat pemodelan, dari data mentah awal berupa *dataset* mahasiswa dan selanjutnya akan melakukan proses *data mining*.

3.3.1 Seleksi data

Atribut yang digunakan adalah program studi mahasiswa, kota asal mahasiswa dan IPK mahasiswa $\geq 2,75$.

3.3.2 Pengolahan data mentah (*Preprocessing Data*)

Pada tahap ini merupakan tahap untuk memastikan data mahasiswa yang dipilih telah layak untuk dilakukan proses pengolahan.

3.3.3 Transformasi data

Data yang berjenis nominal seperti kota asal dan program studi harus dilakukan proses inialisasi data terlebih dahulu ke dalam bentuk angka/numerikal. Untuk melakukan inialisasi kota asal dapat dilakukan dengan:

a. Pada kota asal mahasiswa terlebih dahulu dilakukan pembagian wilayah-wilayah menjadi beberapa bagian wilayah.

Tabel 3.1. Inialisasi Data Wilayah Kota Asal

Kota Asal	Frekuensi	Inisial
JAWA TENGAH 1	1821	1
JAWA TENGAH 2	909	2
JAWA TENGAH 5	492	3
JAWA TENGAH 4	135	4
JAWA TENGAH 3	121	5
JAWA TIMUR	94	6
JAWA BARAT	46	7
SUMATERA SELATAN	42	8
KALIMANTAN TENGAH	24	9
D.I YOGYAKARTA	23	10
SUMATERA UTARA	23	11
D.K.I JAKARTA	19	12
RIAU	11	13
NUSA TENGGARA BARAT	11	14
KALIMANTAN BARAT	10	15
KALIMANTAN SELATAN	8	16
SULAWESI SELATAN	7	17
KALIMANTAN TIMUR	5	18
MALUKU	5	19
SUMATERA BARAT	4	20
NUSA TENGGARA TIMUR	4	21
BALI	3	22
PAPUA	3	23
SULAWESI UTARA	2	24
SULAWESI TENGAH	2	25

b. Kemudian wilayah-wilayah tersebut dilakukan pengurutan angka dari yang terbesar berdasarkan frekuensi mahasiswa yang berasal dari wilayah tersebut.

Tabel 3.2. Inisialisasi Data Program Studi

Program Studi	Frekuensi	Inisial
Teknik Informatika / S1	1110	1
Akuntansi / S1	501	2
Sistem Informasi / S1	435	3
DKV / S1	348	4
Manajemen / S1	295	5
Kesehatan Masyarakat / S1	269	6
Rekam Medis & Info Kes / D3	231	7
Teknik Informatika / D3	127	8
Sastra Inggris / S1	106	9
Broadcast / D3	104	10
Teknik Industri / S1	83	11
Sastra Jepang / S1	72	12
Manajemen Informatika / S1	72	13
Teknik Elektro / S1	66	14

c. Wilayah yang memiliki frekuensi terbesar diberi inisial dengan angka 1 dan wilayah yang memiliki frekuensi terbesar kedua diberi inisial dengan angka 2, begitu seterusnya hingga wilayah dengan frekuensi paling sedikit.

Selain kota asal, program studi juga termasuk ke dalam jenis data nominal ke dalam bentuk angka/numerikal.

Tabel 3.3. Contoh *dataset* mahasiswa yang telah dilakukan inisialisasi

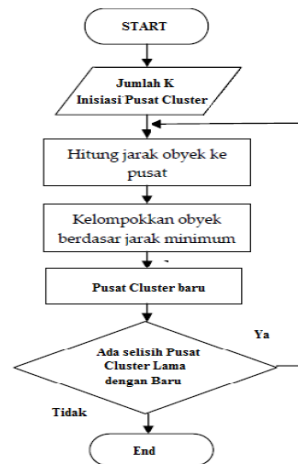
No.	Program Studi	Kota Asal	IPK
1.	3	1	3,85
2.	7	2	3,64
3.	8	5	2,91
4.	1	8	3,20
5.	3	1	3,85
6.	6	1	2,99
7.	12	4	2,93
8.	6	14	3,62
9.	7	2	3,64
10.	14	13	3,02
11.	7	2	3,45
12.	14	4	3,10
13.	11	1	3,29
14.	11	1	3,63
15.	5	7	3,28
16.	8	1	2,75
17.	5	4	3,11
18.	3	2	3,05
19.	3	18	3,14
20.	1	5	3,37
21.	1	7	3,07
22.	10	1	3
23.	2	2	3,11
24.	9	11	3,18
25.	6	1	3,3

3.4 Pemodelan (*Modeling*)

Pemodelan adalah fase yang secara langsung melibatkan teknik *data mining* yaitu dengan melakukan pemilihan teknik *data mining* dan menentukan algoritma yang akan digunakan.

3.4.1 Pemilihan teknik pemodelan

Tool yang digunakan adalah *RapidMiner* versi 5.3:



Gambar 3.1. Flowcart *RapidMiner*

Flowcart algoritma *K-means*:

a. Penetapan jumlah *cluster* (K) pada penelitian ini yaitu berjumlah 3 *cluster*.

Tabel 3.4. Titik Pusat Awal Tiap *Cluster*

Titik Pusat Awal	Jurusan	Kota Asal	IPK
<i>Cluster</i> 1	1	2	3,33
<i>Cluster</i> 2	3	1	3,53
<i>Cluster</i> 3	2	6	2,81

b. Hitung jarak setiap data ke pusat *cluster* antara objek ke *centroid* dengan perhitungan jarak *Euclidean*.

Persamaan yang digunakan adalah :

$$d(p,q) = \sqrt{(p1 - q1)^2 + (p2 - q2)^2 + (p3 - q3)^2} \quad (4-1)$$

Persamaan 2-3 diatas digunakan karena atribut yang digunakan berjumlah 3.

Sebagai contoh, akan dihitung jarak dari data mahasiswa pertama ke pusat *cluster* pertama dengan persamaan:

$$\begin{aligned}
 d(1,1) &= \sqrt{(1 - 1)^2 + (1 - 2)^2 + (3,48 - 3,33)^2} \\
 &= \sqrt{(0)^2 + (-1)^2 + (0,15)^2} \\
 &= \sqrt{0 + (-1) + 0,0225} \\
 &= \sqrt{1,0225}
 \end{aligned}$$

$$d(1,1) = 1,011$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* pertama adalah 1,011.

Jarak dari data mahasiswa pertama ke pusat *cluster* kedua dengan persamaan:

$$\begin{aligned}
 d(1,2) &= \sqrt{(1-3)^2 + (1-1)^2 + (3,48-3,53)^2} \\
 &= \sqrt{(-2)^2 + (0)^2 + (0,15)^2} \\
 &= \sqrt{4 + (0) + 0,0025} \\
 &= \sqrt{4,0025}
 \end{aligned}$$

$$d(1,1) = 2,001$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* pertama adalah 2,001.

Jarak dari data mahasiswa pertama ke pusat *cluster* ketiga dengan perhitungan:

$$\begin{aligned}
 d(1,3) &= \sqrt{(1-2)^2 + (1-6)^2 + (3,48-2,81)^2} \\
 &= \sqrt{(-1)^2 + (-5)^2 + (0,67)^2} \\
 &= \sqrt{1 + (25) + 0,4489} \\
 &= \sqrt{25,4489}
 \end{aligned}$$

$$d(1,1) = 5,044$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan *cluster* pertama adalah 5,044.

Berdasarkan hasil ketiga perhitungan di atas dapat disimpulkan bahwa jarak data mahasiswa pertama yang paling dekat adalah *cluster* 1, sehingga mahasiswa pertama dimasukkan ke dalam *cluster* 1. Hasil perhitungan selengkapnya untuk 5 data mahasiswa pertama dapat dilihat pada **Tabel 3.5**.

c. Kelompokkan data ke dalam *cluster* dengan jarak minimal

Tabel 3.5. Contoh hasil perhitungan setiap data ke setiap *cluster*

No	Program Studi	Kota Asal	IPK	Jarak Ke			Jarak Terdekat
				C1	C2	C3	
1.	3	1	3,85	1,011	2,001	5,044	1
2.	7	2	3,64	2,296	0,32	6,04	2
3.	8	5	2,91	6,141	4,125	6,457	2
4.	1	8	3,20	5,262	6,267	5,402	1
5.	3	1	3,85	5,259	4,722	4,358	3
6.	6	1	2,99	6,412	5,855	4,004	3
7.	12	4	2,93	11,187	9,505	10,198	2
8.	6	14	3,62	13,003	13,341	8,980	3
9.	7	2	3,64	6,008	4,124	6,456	2
10.	14	13	3,02	17,032	16,286	13,894	3
11.	7	2	3,45	6,001	4,123	6,435	2
12.	14	4	3,10	13,154	11,409	12,168	2
13.	11	1	3,29	10,049	8,003	10,306	2
14.	11	1	3,63	10,054	8,000	10,328	2
15.	5	7	3,28	6,403	6,329	5,855	3
16.	8	1	2,75	7,094	5,060	7,810	2
17.	5	4	3,11	4,477	3,629	3,618	3
18.	3	2	3,05	2,019	1,109	4,130	2
19.	3	18	3,14	16,125	17,004	12,046	3
20.	1	5	3,37	3,000	4,474	1,521	3
21.	1	7	3,07	5,006	6,341	1,437	3
22.	10	1	3	9,061	7,020	9,435	2
23.	2	2	3,11	1,023	1,475	4,011	1
24.	9	11	3,18	12,042	11,667	8,610	3

Matriks jarak dengan *centroid* awal seperti tersebut di atas adalah sebagai berikut:

Tabel 3.6. Contoh tabel jarak dengan *centroid* awal

Cluster 1	Cluster 2	Cluster 3
1	0	0
0	1	0
0	1	0
1	0	0
0	0	1
0	0	1
0	1	0
0	0	1
0	1	0
0	0	1
0	1	0
0	1	0
0	1	0
0	0	1
0	1	0
0	0	1
0	1	0
0	0	1
0	0	1
0	0	1
1	0	0
0	0	1
0	1	0

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

Contoh perhitungan pada *cluster* 1 adalah sebagai berikut:

$$\frac{13 + 8 + 10 + 9 + 12 + 7 + \dots + 11}{11}, \frac{1 + 1 + 1 + 3 + 3 + 2 + \dots + 1}{11}$$

Jadi, hasil *centroid* baru *cluster* 1 adalah (9.609 , 1.988 , 3.165).

Contoh perhitungan pada *cluster* 2 adalah sebagai berikut:

$$\frac{1 + 1 + 2 + 3 + 4 + 5 + \dots + 6}{6}, \frac{1 + 2 + 3 + 4 + 6 + 2 + \dots + 4}{6}$$

Jadi, hasil *centroid* baru *cluster* 2 adalah (2.621 , 1.939 , 3.148).

Contoh perhitungan pada *cluster* 3 adalah sebagai berikut:

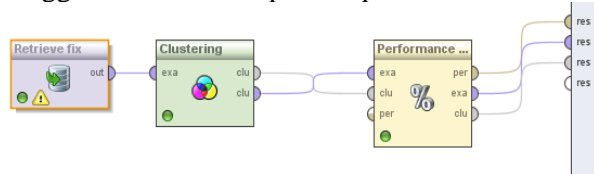
$$\frac{3 + 4 + 9 + 12 + 6 + 7 + \dots + 11}{11}, \frac{8 + 18 + 9 + 10 + 12 + 8 + \dots + 9}{11}$$

Jadi, hasil *centroid* baru *cluster* 3 adalah (5.009 , 11.960 , 3.200).

e. Setelah didapatkan titik pusat yang baru dari setiap *cluster*, lakukan kembali dari langkah ketiga hingga titik pusat dari setiap *cluster* tidak berubah lagi dan tidak ada lagi data yang berpindah dari satu *cluster* ke *cluster* yang lain.

3.4.2 Implementasi dengan RapidMiner

Berikut adalah pengolahan data dengan menggunakan *k-means* pada *RapidMiner* :



Gambar 3.2. Pemodelan *K-Means* pada *RapidMiner*

Dengan menggunakan pemodelan *k-means clustering* seperti **Gambar 3.2** diatas, dengan inialisasi jumlah *cluster* sebanyak 3 buah, maka didapatkan hasil dengan *cluster* yang terbentuk adalah 3, sesuai dengan pendefinisian nilai *k* dengan jumlah *cluster_0* ada 804 item, *cluster_1* ada 2792 item, *cluster_3* ada 223 item sejumlah 3819 item.

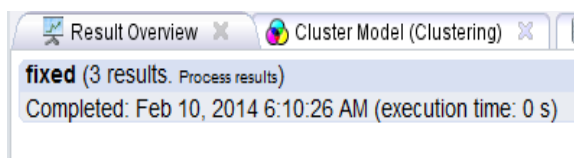
Tabel 3.7. Hasil *cluster K-Means* dalam implementasi *RapidMiner*

Cluster Model	
Cluster 0	804 items
Cluster 1	2792 items
Cluster 2	223 items
Total number of items	3819 items

Hasil perhitungan manual antara jarak *cluster* dengan *centroid* dan perhitungan menggunakan *RapidMiner* menghasilkan perhitungan yang sama serta tidak mengalami perubahan.

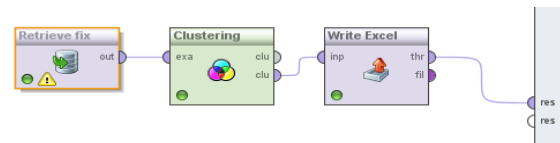
Tabel 3.8. Hasil perhitungan manual antara jarak *cluster* dengan *centroid*

Attribute	cluster_0	cluster_1	cluster_3
Progdi	9,609	2,621	5,009
Kota_asal	1,988	1,939	11,960
Ipk	3,165	3,148	3,200



Gambar 3.3. *Result Overview*

Proses melakukan *export* data ke *excel* mempermudah pada saat melakukan pengolahan data dan evaluasi terhadap hasil dari *cluster* yang terbentuk saat melakukan implementasi pada *RapidMiner*.



Gambar 3.4. *Eksport data dari RapidMiner ke Excel*

3.5 Evaluasi (Evaluation)

Evaluasi adalah fase lanjutan terhadap tujuan *data mining*. Evaluasi dilakukan secara mendalam dengan tujuan agar hasil pada tahap pemodelan sesuai dengan sasaran yang ingin dicapai dalam tahap *business understanding*.

3.5.1 Evaluasi Hasil (Evaluation Results)

Tahap ini menilai sejauh mana hasil pemodelan *data mining* memenuhi tujuan *data mining* yang telah ditentukan pada tahap *business understanding*.

3.5.2 Pengecekan Ulang Proses (Review Process)

Pada tahapan ini penulis memastikan bahwa semua tahapan / faktor penting yang telah dilakukan dalam pengolahan data tidak ada yang terlewatkan.

3.5.3 Menentukan Langkah Selanjutnya (Determine Next Steps)

Pada tahap ini adalah tahapan dalam menentukan langkah selanjutnya yang dilakukan. Terdapat 2 pilihan yaitu kembali pada tahap awal (*business understanding*) atau melanjutkan ke tahap akhir (*deployment*).

3.6 Persebaran (Deployment)

3.6.1 Analisis cluster

Tabel 3.9. Berisi tentang hasil pengelompokan berdasarkan kedekatan jarak antara titik pusat dengan data mahasiswa pada setiap atribut.

Tabel 3.9. Hasil Analisis *Cluster Satu*

Hasil Cluster Satu	
Cluster Satu terdiri dari 804 mahasiswa yang berasal dari:	
Program Studi :	Berasal dari Wilayah :
a. Rekam Medis&Info. Kesehatan = 210	a. Jawa Tengah 1 = 438
b. Teknik Informatika / D3 = 125	b. Jawa Tengah 2 = 167
c. Sastra Inggris = 98	c. Jawa Tengah 5 = 94
d. Broadcasting = 98	d. Jawa Tengah 4 = 38
e. Teknik Industri = 77	e. Jawa Tengah 3 = 29
f. Sastra Jepang = 65	f. Jawa Timur = 20
g. Manajemen Informatika = 67	g. Jawa Barat = 13
h. Teknik Elektro = 64	h. Sumatera Selatan = 4
	i. Kalimantan Tengah = 1
Dengan rata-rata IPK = 3,16	

Tabel 3.10. Hasil Analisis *Cluster* Dua

Hasil <i>Cluster</i> Dua	
<i>Cluster</i> Dua terdiri dari 2792 mahasiswa yang berasal dari:	
Program Studi :	Berasal dari Wilayah :
a. Teknik Informatika = 1070	a. Jawa Tengah 1 = 1383
b. Akuntansi = 478	b. Jawa Tengah 2 = 742
c. Sistem informasi = 412	c. Jawa Tengah 5 = 399
d. DKV = 329	d. Jawa Tengah 4 = 97
e. Manajemen = 277	e. Jawa Tengah 3 = 92
f. Kesehatan Masyarakat = 226	f. Jawa Timur = 54
	g. Jawa Barat = 25
Dengan rata-rata IPK = 3,15	

Tabel 3.11. Hasil Analisis *Cluster* Tiga

Hasil <i>Cluster</i> Tiga	
<i>Cluster</i> 3 terdiri dari 223 mahasiswa yang berasal dari:	
Program Studi :	Berasal dari Wilayah :
a. Teknik Informatika = 41	a. Sumatera Selatan = 39
b. Akuntansi = 23	b. Kalimantan Tengah = 30
c. Sistem informasi = 23	c. D. I Yogyakarta = 21
d. DKV = 18	d. Sumatera Utara = 25
e. Manajemen = 18	e. D.K.I Jakarta = 19
f. Kesehatan Masyarakat = 43	f. Riau = 11
g. Rekam Medis dan Info Kesehatan = 21	g. Kalimantan Barat = 11
h. Teknik Informatika D3 = 2	h. Kalimantan Selatan = 8
i. Sastra Inggris = 8	i. Sulawesi Selatan = 6
j. Broadcasting = 6	j. Kalimantan Timur = 6
k. Teknik Industri = 6	k. Sumatera Barat = 4
l. Sastra Jepang = 7	l. Sulawesi Tengah = 1
m. Manajemen Informatika = 5	m. Bali = 3
n. Teknik Elektro = 2	n. Jawa Barat = 14
	o. NTT = 4
	p. Papua = 3
	q. Maluku = 5
	r. NTB = 11
	s. Sulut = 2
Dengan rata-rata IPK = 3,2	

3.7 Strategi Promosi

Dari data hasil *clustering* yang telah dilakukan di atas, maka dapat ditentukan beberapa strategi promosi yang dapat dilakukan oleh pihak admisi UDINUS.

a. Promosi dengan mengirim tim marketing yang sesuai dengan program studi yang paling banyak diminati.

b. Promosi pada persebaran wilayah berdasarkan tingkat akademik mahasiswa.

Selanjutnya akan dilakukan analisis strategi promosi berdasarkan *promotion mix* pada masing-masing *cluster* yang terbentuk. Diharapkan dengan strategi yang diberikan pada masing-masing *cluster* dapat sesuai dengan tujuan bisnis yang telah dijelaskan.

Tabel 3.12. Strategi Promosi berdasarkan *Promotion Mix*

No.	Strategi Promosi	<i>Cluster</i> Satu	<i>Cluster</i> Dua	<i>Cluster</i> Tiga
1.	Periklanan	✓	✓	✓
2.	Penjualan Personal	✓	✓	✓
3.	Promosi Penjualan	✓	✓	
4.	Hubungan Masyarakat	✓	✓	
5.	Pemasaran Langsung	✓	✓	

4 KESIMPULAN DAN SARAN

4.1 Kesimpulan

a. Setelah dilakukan pengelompokan data mahasiswa melalui persebaran wilayah berdasarkan potensi akademik menggunakan *K-Means clustering* terbentuk tiga *cluster* yaitu, *cluster* satu dengan jumlah 804 mahasiswa dengan rata-rata IPK 3.16, *cluster* dua dengan jumlah 2792 mahasiswa dengan rata-rata IPK 3.15 dan *cluster* tiga dengan jumlah 223 mahasiswa dengan rata-rata IPK 3.2.

b. Strategi promosi bagi calon mahasiswa baru yang tepat sasaran untuk setiap wilayah berdasarkan *cluster* yang terbentuk adalah dengan mengirim tim admisi UDINUS yang sesuai dengan program studi yang paling banyak diminati dan melakukan promosi berdasarkan potensi akademik mahasiswa dengan melakukan penyelarasan menggunakan *promotion mix* dan dengan melihat rata-rata IPK pada setiap *cluster*.

4.2 Saran

a. Pengelompokan terhadap data mahasiswa UDINUS sebaiknya dilakukan secara rutin setiap tahun untuk menghindari penurunan mahasiswa baru pada tahun berikutnya.

b. Penelitian ini dapat dijadikan sebagai salah satu referensi bagi pihak admisi UDINUS untuk mengambil keputusan dalam melakukan strategi promosi bagi calon mahasiswa baru.

5 DAFTAR PUSTAKA

- [1] Johan Oscar Ong, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING UNTUK MENENTUKAN STRATEGI MARKETING PRESIDENT UNIVERSITY," *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, pp. 10-13, Juni 2013.
- [2] Universitas Dian Nuswantoro, *Buku Panduan Akademik Mahasiswa Tahun Ajaran 2012-2013*. Semarang, Jawa Tengah: Universitas Dian Nuswantoro, 2012.
- [3] Irwan Budiman, Toni Prahasto, and Yuli Christyono, "DATA CLUSTERING MENGGUNAKAN METODOLOGI CRISP-DM UNTUK PENGENALAN POLA PROPORSI PELAKSANAAN TRIDHARMA," in *Seminar Nasional Aplikasi Teknologi Informasi 2012 (SNATI 2012)*, Yogyakarta, 2012.
- [4] Yudi Agusta, "K-Means – Penerapan, Permasalahan dan Metode Terkait," *Jurnal Sistem dan Informatika*, vol. 3, pp. 47-60, Pebruari 2007.
- [5] Sri Kusumadewi Tedy Rismawan, "APLIKASI K-MEANS UNTUK PENGELOMPOKAN MAHASISWA BERDASARKAN NILAI BODY MASS INDEX (BMI) & UKURAN KERANGKA," in *Seminar Nasional Aplikasi Teknologi Informasi*, Yogyakarta, 2008, pp. ISSN: 1907-5022.
- [6] Suprihatin, "Klastering K-Means Untuk Penentuan Nilai Ujian," *JUSI*, vol. 1, no. 1, pp. ISSN: 2087-8737, Februari 2011.
- [7] Philip Kotler, *Manajemen Pemasaran: Analisis, Perencanaan, Implementasi dan Kontrol*, Prenhallindo, Ed. Jakarta, 1997.
- [8] Philip dan Gary Armstrong Kotler, *Prinsip-Prinsip Pemasaran*, Alih Bahasa Imam Nurmawan, Ed. Jakarta: Salemba Empat, 2001.
- [9] Universitas Dian Nuswantoro. (2014, Jan) UDINUS. [Online]. HYPERLINK "file:///E:\dari%20data%20D\ta\bismillah\www.dinus.ac.id" www.dinus.ac.id
- [10] Turban, E. dkk, *Decicion Support Systems and Intelligent Systems.*: Andi Offset, 2005.
- [11] Larose, Daniel T, *Data Mining Methods and Models*. Hoboken New Jersey: Jhon Wiley & Sons, Inc, 2006.
- [12] Larose, Daniel T, *Discovering Knowledge in Data: An Introduction to Data Mining.*: John Willey & Sons. Inc, 2005.
- [13] Han,J. and Kamber,M. "Data mining: Concepts and Techniques", 2nd