
IMPLEMENTASI ALGORITMA *K-MEANS CLUSTERING* UNTUK MENENTUKAN STRATEGI MARKETING PRESIDENT UNIVERSITY

Johan Oscar Ong¹

Abstract: Information technology advances very rapidly at this time to generate thousands or even millions of data from various aspect of life. However, what can be done with that much data?. In this research, we start from calculation of data set of students who have graduated from President University using k-means clustering algorithm, namely by classifying the data of students into several clusters based on the characteristics of this data in order to discover the information hidden from the data set of student who have graduated from President University. The attribute data that is used in this study is hometown, major and GPA. The purpose of this study is to help the President University's marketing department in predicting promotion strategies undertaken in the cities in Indonesia. Information gained in this study can be used as a references in determining the proper strategy for marketing team in their promotion activities in the cities in Indonesia so that the campaign will be more effective and efficient.

Keywords: student data, major, GPA, k-means clustering, promotional strategies

PENDAHULUAN

Dalam berbagai bidang kehidupan saat ini, banyak sekali data yang dihasilkan oleh teknologi informasi yang semakin canggih. Mulai dari bidang industri, ekonomi, pendidikan, ilmu dan teknologi serta berbagai bidang kehidupan lainnya yang menghasilkan data yang sangat berlimpah. Namun, apa yang dapat dilakukan dari data-data tersebut? Untuk dapat mengetahui informasi yang tersembunyi dari data-data tersebut, maka perlu dilakukan pengolahan terhadap data-data tersebut. Dalam penelitian ini akan dilakukan pengolahan data mahasiswa yang telah lulus dari President University. Atribut data yang akan digunakan adalah nama mahasiswa, jurusan yang diambil, kota asal mahasiswa dan nilai IPK. Hasil dari pengolahan data mahasiswa ini bertujuan untuk membantu pihak marketing President University dalam melakukan pemasaran dan mencari calon mahasiswa baru dari berbagai kota di Indonesia.

President University merupakan salah satu Universitas swasta yang cukup terkenal di Indonesia. President University berlokasi di Kawasan Industri Jababeka yang didalamnya beroperasi lebih dari 1000 perusahaan nasional dan internasional. President University tumbuh sangat pesat setiap tahunnya. Mahasiswa President University sangat banyak yang berasal dari berbagai daerah di Indonesia. Tidak hanya berasal dari Indonesia, mahasiswa President University pun banyak yang berasal dari luar Indonesia, seperti Vietnam, China, Korea dan beberapa negara lainnya.

Oleh karena mahasiswa President University berasal dari berbagai daerah bahkan hingga berbagai negara, maka dibutuhkan strategi khusus oleh bagian

¹ Program Studi Teknik Industri, President University, Jln. Jln. Ki Hajar Dewantara, Kota Jababeka, Cikarang Baru, Bekasi, 17550
E-mail : johanoscarong@gmail.com

marketing dalam melakukan pemasaran untuk mencari calon mahasiswa agar promosi yang dilakukan lebih efektif dan efisien. Untuk dapat melakukan promosi yang lebih efektif dan efisien, maka dalam penelitian ini dilakukan dengan cara mengolah data-data yang telah didapatkan dari mahasiswa yang telah lulus seperti nama mahasiswa, kota asal, jurusan yang diambil dan yang terakhir adalah nilai IPK. Data-data yang telah didapatkan tadi kemudian diolah untuk mengetahui pola dari data-data tersebut sehingga kita dapat mengambil informasi-informasi yang tersembunyi dari data-data tersebut. Metode pengolahan data seperti ini sering disebut sebagai *data mining*. Pada penelitian ini analisa data mining dilakukan dengan menggunakan metode *K-Means clustering*. Dengan menggunakan metode ini, data-data yang telah didapatkan dapat dikelompokkan kedalam beberapa *cluster* berdasarkan kemiripan dari data-data tersebut, sehingga data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster* dan yang memiliki karakteristik yang berbeda dikelompokkan dalam *cluster* yang lain yang memiliki karakteristik yang sama.

Dengan adanya pengelompokan-pengelompokan data seperti ini, diharapkan bagian marketing dapat melakukan pemasaran dengan strategi yang tepat untuk mendapatkan calon mahasiswa baru.

DASAR TEORI

Definisi Data

Dalam *Webster's New World's Dictionary* tertulis bahwa datum: *something known or assumed*. Artinya, datum (bentuk tunggal data) merupakan suatu yang diketahui/dianggap. Dengan demikian, data dapat memberi gambaran tentang suatu keadaan atau persoalan. Sedangkan, data menurut kamus Oxford Dictionary adalah *The Facts*. Jadi, dapat disimpulkan bahwa data adalah sesuatu yang nyata diketahui atau dianggap yang dipakai untuk keperluan suatu analisa, diskusi, presentasi ilmiah atau tes statistik (*Supranto, 2000*).

Jenis Data Berdasar Sifatnya

Jenis-jenis data dapat dibagi menurut sifatnya, menurut sumbernya, menurut cara memperolehnya dan menurut waktu pengumpulannya (*Supranto, 2000*). Menurut sifatnya data dapat terbagi menjadi dua jenis, yaitu data kualitatif (non-metrik) dan data kuantitatif (metrik). Kemudian jenis data kualitatif terbagi lagi menjadi dua jenis, yaitu data nominal dan data ordinal. Begitu pula dengan jenis data kuantitatif terbagi menjadi dua jenis, yaitu data interval dan data rasio (*Hidayat, 2011*).

Data Kualitatif

Data kualitatif secara sederhana dapat disebut data yang bukan berupa angka. Ciri utama data kualitatif didapat dengan cara menghitung, sehingga tidak memiliki nilai desimal. Selain itu data kualitatif memiliki ciri-ciri tidak bisa dilakukan operasi matematika, seperti penambahan, pengurangan, perkalian dan pembagian. Contoh data kualitatif adalah data gender, data golongan darah, data tempat tinggal atau data jenis pekerjaan. Agar dapat dilakukan proses pada data kualitatif atau non metric, data tersebut harus diubah ke dalam bentuk angka, proses ini dinamakan kategorisasi. Data kualitatif dibedakan menjadi dua jenis, yaitu data nominal dan data ordinal (*Santoso, 2010*).

Data Nominal adalah jenis data kualitatif yang digunakan mengidentifikasi, mengklasifikasi, atau membedakan objek. Data nominal merupakan data yang paling

rendah dalam level pengukuran data. Semua data memiliki posisi setara dalam arti tidak ada data yang memiliki tingkat yang lebih atau kurang dibandingkan dengan data yang lain. Jenis data nominal ini tidak memiliki jarak, urutan dan titik origin (*Hidayat, 2011*).

Data Ordinal adalah jenis data kualitatif namun memiliki level lebih tinggi dari data nominal. Data ordinal memiliki karakteristik nominal tapi terdapat perbedaan derajat, urutan, atau peringkat dalam objek tersebut (posisi data tidak setara) (*Hidayat, 2011*).

Data Kuantitatif

Data kuantitatif dapat disebut sebagai data berupa angka dalam arti sebenarnya. Jadi, berbagai jenis operasi matematika dapat dilakukan pada data kuantitatif. Data kuantitatif merupakan data yang didapat dengan jalan mengukur sehingga bisa mempunyai nilai desimal. Contoh data kuantitatif adalah tinggi badan, usia, penjualan barang, dan sebagainya. Sebagai contoh, tinggi badan seseorang bisa bernilai 165 cm atau 165.5 cm. Seperti pada jenis data kualitatif, jenis data kuantitatif juga terbagi menjadi dua, yaitu data interval dan data rasio (*Santoso, 2010*).

Data interval menempati level pengukuran data yang lebih tinggi dari data ordinal karena selain bisa bertingkat urutannya, urutan tersebut juga bisa dikuantitatifkan serta memiliki indikator jarak. Contohnya seperti pengukuran temperatur sebuah ruangan. Interval temperature ruangan tersebut adalah:

- a. Cukup panas jika temperatur antara $50^{\circ}\text{C} - 80^{\circ}\text{C}$
- b. Panas jika temperatur antara $80^{\circ}\text{C} - 110^{\circ}\text{C}$
- c. Sangat panas jika temperatur antara $110^{\circ}\text{C} - 140^{\circ}\text{C}$

Dalam kasus di atas, data temperatur bisa dikatakan data interval karena data mempunyai interval (jarak) tertentu, yaitu 30°C .

Data rasio merupakan data dengan tingkat pengukuran paling tinggi diantara jenis data lainnya. Data rasio adalah data yang bersifat angka dalam arti yang sebenarnya, bukan katagori seperti data nominal dan data ordinal sehingga dapat dilakukan operasi matematika seperti penambahan, pengurangan, perkalian, dan pembagian. Perbedaan dengan data interval adalah data rasio memiliki indikator titik origin yang tidak dapat berubah (absolute). Contoh dari data rasio adalah jumlah suatu produk, jika jumlah produk 0 (nol) berarti memang tidak ada produk atau contoh lainnya adalah berat badan dan tinggi badan, pengukuran-pengukurannya mempunyai nilai 0 (nol) yang sebenarnya. Misalnya jika berat badan 0 berarti memang tanpa berat.

Jenis Data Menurut Sumbernya

Pembagian jenis data menurut sumbernya didasarkan pada sumber perolehan data tersebut, yaitu data internal dan data eksternal (*Supranto, 2007*).

Data internal adalah data yang dikumpulkan oleh suatu organisasi untuk menggambarkan keadaan atau kegiatan organisasi yang bersangkutan serta berguna untuk keperluan kegiatan harian dan pengawasan internal. Misalnya, data penjualan, data produksi suatu perusahaan, data keuangan, data kepegawaian, dan lain sebagainya.

Data eksternal adalah data yang dikumpulkan untuk menggambarkan suatu keadaan atau kegiatan di luar organisasi tersebut. Contoh dari data eksternal seperti data jumlah penduduk dan data pendapatan nasional yang didapat dari kantor pusat statistik setempat. Suatu perusahaan memerlukan data eksternal seperti jumlah

penduduk untuk memprediksi potensi pemerintaan, sedangkan data pendapatan nasional untuk menentukan tingkat daya beli masyarakat yang berguna untuk dasar kebijakan tingkat harga.

Jenis Data Menurut Cara Memperolehnya

Berdasarkan cara memperolehnya, data dapat dibedakan menjadi dua jenis, yaitu data primer dan data sekunder (Supranto, 2000).

Data primer adalah data yang dikumpulkan dan diolah sendiri oleh organisasi atau perorangan langsung dari objeknya. Misalnya suatu perusahaan ingin mengetahui konsumsi rata-rata suatu produk terhadap penduduk disuatu daerah dengan cara melakukan wawancara langsung kepada penduduk setempat.

Data sekunder adalah data yang diperoleh dalam bentuk jadi dan telah diolah oleh pihak lainnya. Biasanya data sekunder ini dalam bentuk publikasi.

Jenis Data Menurut Waktu Pengumpulannya

Berdasarkan waktu pengumpulannya, data dapat dibedakan menjadi dua jenis, yaitu data *cross section* dan data berkala (*time series*) (Supranto, 2000).

Data *cross section* adalah data yang dikumpulkan dalam suatu periode tertentu, biasanya menggambarkan keadaan atau kegiatan dalam periode tersebut. Misalnya, hasil sensus penduduk tahun 2012 menggambarkan keadaan Indonesia pada tahun 2012 menurut, umur, jenis kelamin, agama, tingkat pendidikan, dan sebagainya.

Data berkala (*time series*) adalah data yang dikumpulkan dari waktu ke waktu. Tujuannya adalah untuk menggambarkan perkembangan suatu kegiatan dari waktu ke waktu. Misalnya, perkembangan produksi di suatu perusahaan selama lima tahun terakhir, perkembangan penjualan produk selama lima tahun terakhir, dan lain sebagainya. Jenis data ini juga sering disebut sebagai data historis.

Data Mining

Data mining adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari data tersebut. Hasil dari pengolahan data dengan metode *data mining* ini dapat digunakan untuk mengambil keputusan di masa depan. *Data mining* ini juga dikenal dengan istilah *pattern recognition* (Santosa, 2007).

Data mining merupakan metode pengolahan data berskala besar oleh karena itu data mining ini memiliki peranan penting dalam bidang industri, keuangan, cuaca, ilmu dan teknologi. Secara umum kajian data mining membahas metode-metode seperti, clustering, klasifikasi, regresi, seleksi variable, dan market basket analisis (Santosa, 2007).

Clustering

Pada dasarnya 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*), maksudnya metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*) serta tidak memerlukan target output. Dalam data mining ada dua jenis metode clustering yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering* (Santosa, 2007).

Hierarchical clustering adalah suatu metode pengelompokan data yang dimulai dengan mengelompokkan dua atau lebih objek yang memiliki kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang memiliki kedekatan kedua. Demikian seterusnya sehingga cluster akan membentuk semacam pohon dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut (Santoso, 2010).

Berbeda dengan metode *hierarchical clustering*, metode *non-hierarchical clustering* justru dimulai dengan menentukan terlebih dahulu jumlah cluster yang diinginkan (dua cluster, tiga cluster, atau lain sebagainya). Setelah jumlah cluster diketahui, baru proses cluster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means Clustering* (Santoso, 2010).

K-means Clustering

K-means clustering merupakan salah satu metode data clustering non-hirarki yang mengelompokkan data dalam bentuk satu atau lebih cluster/kelompok. Data-data yang memiliki karakteristik yang sama dikelompokkan dalam satu *cluster*/kelompok dan data yang memiliki karakteristik yang berbeda dikelompokkan dengan *cluster*/kelompok yang lain sehingga data yang berada dalam satu *cluster*/kelompok memiliki tingkat variasi yang kecil (Agusta, 2007).

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

- a. Pilih jumlah *cluster* k .
- b. Inisialisasi k pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberiduberi nilai awal dengan angka-angka random,
- c. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap 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 dirumuskan sebagai berikut:

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

dimana:

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

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

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

- d. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari cluster tersebut. Jadi rata-rata (mean) bukan satu-satunya ukuran yang bisa dipakai.
- e. Tugaskan lagi setiap objek memakai pusat *cluster* yang baru. Jika pusat cluster tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi

METODOLOGI

Pada penelitian ini dimulai dari melakukan pengamatan secara langsung pada Universitas. Pengamatan secara langsung ini dilakukan melalui wawancara terhadap staff-staff dan dosen-dosen yang bekerja di Universitas tersebut. Wawancara ini bertujuan untuk mengetahui gambaran permasalahan yang ada di Universitas tersebut. Berdasarkan hasil wawancara yang dilakukan, lalu dapat ditetapkan rumusan masalah yang ada di Universitas tersebut. Setelah menetapkan rumusan masalah yang ada di dalam Universitas tersebut, kemudian dibuat batasan-batasan masalah agar pembahasan yang akan dijelaskan tidak keluar dari ruang lingkup penelitian yang dilakukan.

Setelah menetapkan rumusan masalah dan batasan-batasan dari masalah tersebut, lalu menentukan tujuan dari penelitian yang akan dilakukan. Tujuan dari penelitian ini akan menjawab semua masalah yang telah dirumuskan. Dari rumusan masalah yang telah didapatkan lalu dilakukan studi literatur untuk mendapatkan teori-teori tentang permasalahan yang telah didapatkan. Teori-teori ini berfungsi sebagai panduan untuk mendapatkan solusi dari permasalahan yang kita temukan.

Data-data yang digunakan pada penelitian ini didapat dari bagian akademik yang berupa data mahasiswa yang telah lulus dari Universitas tersebut pada tahun 2012. Data-data ini berisi data diri dari mahasiswa yang telah lulus tersebut, namun dalam penelitian ini hanya beberapa atribut data saja yang digunakan, seperti nama mahasiswa, kota asal, jurusan dan nilai IPK. Data-data yang telah didapatkan pada tahap pengumpulan data kemudian dilakukan transformasi pada data-data yang berjenis data nominal, yaitu seperti kota asal dan jurusan. Data-data yang berjenis data nominal tersebut diinisialisasikan ke dalam bentuk angka melalui beberapa langkah agar data-data yang berjenis data nominal dapat diolah dengan menggunakan algoritma *K-means Clustering*.

Semua data yang telah didapatkan tadi kemudian diolah untuk dapat dianalisa sehingga menghasilkan informasi yang berguna untuk mengatasi permasalahan tersebut. Pengolahan data ini menggunakan teori-teori dari data mining dengan algoritma *K-Means Clustering* atau yang lebih spesifik lagi algoritma *Hard K-Means Clustering*. Dengan metode *Hard K-Means Clustering* ini, data yang diperoleh dikelompokkan berdasar kemiripan karakteristik dari setiap data, sehingga dapat ditemukan informasi yang tersembunyi dari data-data tersebut. Berdasarkan analisa dan pengolahan data, maka dapat diambil kesimpulan dari penelitian yang telah dilakukan. Kesimpulan tersebut akan menjawab rumusan masalah yang telah ditetapkan diawal.

Data

Data yang diperlukan dalam penelitian ini diperoleh melalui bagian akademik President University yang berupa data mahasiswa yang telah lulus dari President University pada tahun 2012. Data-data tersebut berisi data diri mahasiswa yang telah lulus, namun dalam penelitian ini hanya beberapa atribut data saja yang digunakan, seperti nama mahasiswa, kota asal, jurusan, dan nilai IPK. Berikut ini adalah data mahasiswa yang telah lulus pada tahun 2012.

Tabel 1. Contoh Data Mahasiswa yang Telah Lulus dari President University Tahun 2012

No.	Name	Jurusan	Kota asal	IPK
-----	------	---------	-----------	-----

1	ADE SUPRYAN STEFANUS	IS	Jakarta	3.16
2	ADELINA GANARDI PUTRI HARDI	ACC	Semarang	3.22
3	ADELINE DEWITA	BF	Bekasi	3.29
4	ADIPUTRA	IB	Jakarta	2.83
5	AFRIESKA LAURA TRISYANA	PR	Jakarta	3.15
6	AGAM KHALILULLAH	IB	Banda Aceh	3.25
7	AGUS MULYANA JUNGJUNGAN	IB	Bogor	3.43
8	AGUSMAN	PR	Bekasi	3.06
9	AIDIL FRIADI	BF	Banda Aceh	3.36
10	AJENG PUTRI ARIANDHANI	ACC	Bandung	3.28

Transformasi Data

Agar data di atas dapat diolah dengan menggunakan metode *k-means clustering*, maka data yang berjenis data nominal seperti kota asal dan jurusan harus diinisialisasikan terlebih dahulu dalam bentuk angka.

Untuk melakukan inisialisasi kota asal dilakukan dengan langkah-langkah sebagai berikut:

1. Pada data kota asal terlebih dulu dilakukan pembagian wilayah yang menjadi beberapa bagian wilayah, yaitu:
 - a. Wilayah Sumatera Utara yang terdiri dari kota Aceh Besar, Ambarita Samosir, Banda Aceh, Lhokseumawe, dan Medan
 - b. Wilayah Sumatera Selatan yang terdiri dari kota Padang, Palembang, Bengkulu, Jambi, Bandar Lampung, dan Pangkal Pinang.
 - c. Wilayah DKI Jakarta yang terdiri dari kota Jakarta.
 - d. Wilayah Jawa Barat yang terdiri dari kota, Bandung, Bekasi, Bogor, Ciamis, Cikarang, Depok, Garut, Karawang, Pandeglang, Sukabumi, dan Tangerang.
 - e. Wilayah Jawa Timur yang terdiri dari kota Magelang, Semarang, Sidoarjo, Solo, Surabaya, Tegal Yogyakarta dan Kudus.
 - f. Wilayah Kalimantan yang terdiri dari kota Bontang.
 - g. Wilayah Sulawesi yang terdiri dari kota Makassar, Manado, dan Palu.
 - h. Wilayah Bali yang terdiri dari kota Singaraja dan Denpasar
2. Kemudian wilayah-wilayah tersebut diurutkan dari yang terbesar berdasarkan frekuensi mahasiswa yang berasal dari wilayah tersebut.
3. Setelah itu 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. Hasil dari inisialisasi kota asal dapat dilihat pada tabel 2.

Tabel 2. Inisialisasi Data Wilayah Kota Asal

Wilayah	Frekuensi	Inisial
DKI Jakarta	84	1
Jawa Barat	82	2
Sumatera Utara	28	3
Sulawesi	14	4
Jawa Timur	13	5
Sumatera Selatan	13	6
Bali	8	7
Kalimantan	1	8

Jurusan

Selain kota asal, jurusan juga termasuk ke dalam jenis data nominal sehingga perlu diinisialisasikan ke dalam bentuk angka. Seperti pada kota asal, pada jurusan juga diberikan inisialisasi berdasarkan frekuensi mahasiswa pada jurusan tersebut. Hasil dari inisialisasi jurusan tersebut dapat dilihat pada tabel (3)

Tabel 3. Inisialisasi Data Jurusan

Major	Singkatan	Frekuensi	Inisial
Accounting	ACC	46	1
Management, concentration in International Business	IB	37	2
Public Relation	PR	35	3
Management, concentration in Banking & Finance	BF	28	4
Industrial Engineering	IE	23	5
Information Technology	IT	20	6
Management, concentration in Marketing	MKT	18	7
Visual Communication Design	VCD	12	8
Management, concentration in Hotel & Tourism Management	HTM	9	9
Electrical Engineering	EE	6	10
Business Administration	BA	4	11
International Relations	IR	2	12
Management, concentration in Human Resources Management	HRM	1	13
Information System	IS	1	14
Management	MGT	1	15

PENGOLAHAN DATA

Setelah semua data mahasiswa yang lulus pada tahun 2008 ditransformasi ke dalam bentuk angka, maka data-data tersebut telah dapat dikelompokkan dengan menggunakan algoritma *K-Means Clustering*. Untuk dapat melakukan pengelompokan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu:

1. Tentukan jumlah cluster yang diinginkan. Dalam penelitian ini data-data yang ada akan dikelompokkan mejadi tiga *cluster*.
2. Tentukan titik pusat awal dari setiap cluster. Dalam penelitian ini titik pusat awal ditentukan secara random dan didapat titik pusat dari setiap cluster dapat dilihat pada tabel 4.

Tabel 4. Titik Pusat Awal Setiap Cluster

Titik Pusat awal	Nama	Jurusan	Kota asal	IPK
Cluster 1	DALLY TEGUH SESARIO	9	3	2.94
Cluster 2	HERVINA JULIANA	1	1	3.18
Cluster 3	PASCAL MUHAMMADI	1	2	3.15

3. Tempatkan setiap data pada cluster. Dalam penelitian ini digunakan metode *hard k-means* untuk mengalokasikan setiap data ke dalam suatu *cluster*, sehingga data akan dimasukan dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster*. Sebagai contoh, akan dihitung jarak dari data mahasiswa pertama ke pusat *cluster* pertama:

$$D(1,1) = \sqrt{(14 - 9)^2 + (1 - 3)^2 + (3.16 - 2.94)^2} = 5.390$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data mahasiswa pertama dengan pusat *cluster* pertama adalah 5.390.

Jarak data mahasiswa pertama ke pusat *cluster* kedua:

$$D(1,2) = \sqrt{(14 - 1)^2 + (1 - 1)^2 + (3.16 - 3.18)^2} = 13.000$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data mahasiswa pertama dengan pusat *cluster* kedua adalah 13.

Jarak data mahasiswa pertama ke pusat *cluster* ketiga:

$$D(1,3) = \sqrt{(14 - 1)^2 + (1 - 2)^2 + (3.16 - 3.15)^2} = 13.038$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data mahasiswa pertama dengan pusat *cluster* ketiga adalah 13.038.

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

Tabel 5. Contoh Hasil Perhitungan Setiap Data ke Setiap Cluster

No	Nama	Jurusan	Kota asal	IPK	Jarak Ke			Jarak terdekat ke cluster
					C 1	C 2	C 3	
1	ADE SUPRYAN STEFANUS	14	1	3.16	5.390	13.000	13.038	1
2	ADELINA GANARDI PUTRI HARDI	1	5	3.22	8.251	4.000	3.001	3
3	ADELINE DEWITA	4	2	3.29	5.111	3.164	3.003	3
4	ADIPUTRA	2	1	2.83	7.281	1.059	1.450	2
5	AFRIESKA LAURA TRISYANA	3	1	3.15	6.328	2.000	2.236	2

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

Dalam penelitian ini, iterasi clustering data mahasiswa terjadi sebanyak 7 kali iterasi. Pada iterasi ke-7 ini, 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

Hasil Clustering

Berdasarkan hasil pengelompokan data menggunakan metode k-means clustering, di dapatkan hasil clustering hingga iterasi ke-7, dimana titik pusat tidak lagi berubah dan tidak ada data yang berpindah antar cluster. Hasil dari clustering tersebut seperti pada tabel 6.

Pembahasan

Dari hasil cluster 1, terlihat bahwa karakteristik mahasiswa pada cluster 1 didominasi oleh mahasiswa yang berasal dari jurusan Information Technology dan Marketing. Sedangkan, berdasarkan kota asal didominasi oleh mahasiswa yang berasal dari wilayah kota asal DKI Jakarta dan Jawa Barat, sehingga dapat disimpulkan bahwa rata-rata mahasiswa pada cluster 1 yang berasal dari wilayah kota asal DKI Jakarta dan Jawa Barat mengambil jurusan Information Technology dan Marketing.

Tabel 6. Hasil Analisis Clustering

Hasil Cluster 1	Hasil Cluster 2	Hasil Cluster 3
-----------------	-----------------	-----------------

<p>Cluster 1 terdiri dari 70 orang, yang berasal dari jurusan IT = 19 orang MKT = 15 orang VCD = 12 orang HTM = 9 orang EE = 6 orang BA = 4 orang IR = 2 orang MGT = 1 orang IS = 1 orang HRM = 1 orang Dan berasal dari Wilayah: DKI Jakarta = 30 orang Jawa Barat = 20 orang Sumatera Utara = 12 orang Sulawesi = 2 orang Jawa Timur = 2 orang Sumatera Selatan = 2 orang Bali = 1 orang Kalimantan = 1 orang Dengan rata-rata nilai IPK 3.2</p>	<p>Cluster 2 terdiri dari 132 orang, yang berasal dari aktifis ACC = 39 orang IB = 30 orang BF = 22 orang PR = 21 orang IE = 20 orang Dan berasal dari Wilayah: Jawa Barat = 62 orang DKI Jakarta = 54 orang Sumatera Utara = 16 orang Dengan rata-rata nilai IPK 3.25</p>	<p>Cluster 3 terdiri dari 41 orang, yang berasal dari jurusan: PR = 14 orang ACC = 7 orang IB = 7 orang BF = 6 orang E-3 = 3 orang MKT = 3 orang IT = 1 orang Dan berasal dari Wilayah: Sulawesi = 12 orang. Jawa Timur = 11 orang Sumatera Selatan = 11 orang Bali = 7 orang Dengan rata-rata nilai IPK 3.31</p>
--	--	--

Kemudian, dari hasil cluster 2 di atas dapat dilihat bahwa karakteristik mahasiswa pada cluster 2 didominasi oleh mahasiswa yang berasal dari jurusan Accounting dan International Business. Sedangkan, berdasarkan kota asal didominasi oleh mahasiswa yang berasal dari wilayah kota asal DKI Jakarta dan Jawa Barat, sehingga dapat disimpulkan bahwa rata-rata mahasiswa pada cluster 2 yang berasal dari wilayah kota asal DKI Jakarta dan Jawa Barat mengambil jurusan Information Technology dan Marketing.

Sedangkan, dari hasil cluster 3 di atas dapat dilihat bahwa karakteristik mahasiswa pada cluster 3 didominasi oleh mahasiswa yang berasal dari jurusan Public Relation, Accounting dan International Business. Sedangkan, berdasarkan kota asal didominasi oleh mahasiswa yang berasal dari wilayah kota asal Sulawesi, Jawa Timur dan Sumatera Selatan, sehingga dapat disimpulkan bahwa rata-rata mahasiswa pada cluster 3 yang berasal dari wilayah kota asal Sulawesi, Jawa Timur dan Sumatera Selatan mengambil jurusan Public Relation, Accounting dan International Business.

Strategi Promosi

Dari data hasil clustering yang telah dilakukan di atas, maka dapat dibuat beberapa strategi promosi yang dapat dilakukan oleh pihak marketing President University agar promosi yang dilakukan lebih efektif dan efisien, yaitu:

1. **Promosi Dengan Mengirim Tim Marketing yang Sesuai dengan Jurusan yang Paling Banyak Diminati.** Strategi promosi pertama yang dapat dilakukan pihak marketing President University berdasarkan hasil *clustering* adalah melakukan strategi promosi pada kota-kota di Indonesia berdasarkan jurusan yang paling banyak diminati. Jadi, pihak marketing dapat mengirim tim yang memiliki pengetahuan dan potensi lebih dari suatu jurusan untuk melakukan promosi pada kota yang memiliki minat lebih besar pada jurusan tersebut, sehingga promosi yang dilakukan akan lebih efektif dan efisien karena dengan melakukan strategi ini pihak marketing dapat membagi tim-tim marketing untuk melakukan promosi di kota-kota di Indonesia sehingga sumber daya marketing yang dibutuhkan disetiap kota tidak terlalu banyak.

- 2. Promosi Pada Kota Berdasarkan Tingkat Akademik dari Calon Mahasiswa.** Strategi promosi kedua yang dapat dilakukan pihak marketing President University, yaitu apabila pihak marketing President University ingin mendapat calon mahasiswa yang memiliki tingkat akademik di atas rata-rata, maka dapat dianalisa berdasarkan hasil rata-rata nilai IPK dari setiap mahasiswa yang telah lulus.

SIMPULAN

Dari hasil penelitian yang telah dilakukan dapat disimpulkan bahwa ada tiga strategi promosi yang dapat dilakukan oleh pihak marketing President University, yaitu:

1. Melakukan promosi dengan mengirim tim marketing yang sesuai dengan jurusan yang paling banyak diminati.
2. Melakukan promosi pada kota-kota di Indonesia yang didasarkan pada tingkat kemampuan akademik dari calon mahasiswa.

Daftar Pustaka

- Agusta, Y. 2007. *K-means - Penerapan, Permasalahan dan Metode Terkait*. Jurnal Sistem dan Informatika Vol. 3 (Februari 2007): 47-60.
- Hidayat, T., dan Istiadah, N. 2011 . *Panduan Lengkap Menguasai SPSS 19 untuk Mengolah Data Statistik Penelitian*. Jakarta: Media Kita.
- Santosa, B. 2007. *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Santoso, S. 2010. *Statistik Multivariat*. Jakarta: Elex Media Komputindo.
- Supranto, J. M.A. 2000. *Statistik: Teori dan Aplikasi Edisi Keenam*. Jakarta: Erlangga.