

ANALISIS DATA MAHASISWA MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING SEBAGAI DASAR PELAKSANA PROMOSI

Data Analysis of Students using K-Means Clustering Algorithm as a Basis for The Implementation of Promotion

Sterry Fleanry Mulaki¹⁾, Nina Setiyawati¹⁾ dan Agustinus Fritz Wijaya²⁾

¹⁾ Fakultas Teknologi Informasi/Teknik Informatika, Universitas Kristen Satya Wacana

²⁾ Fakultas Teknologi Informasi/Sistem Informasi, Universitas Kristen Satya Wacana

Diterima 08 Juni 2018 / Disetujui 31 Juli 2018

ABSTRACT

The Promotion Bureau of Satya Wacana Christian University's Information Technology Faculty of Engineering undertakes marketing techniques by sending promotional teams randomly for promotion. The Promotion team sent at random may not be able to achieve the expected target. In order to reach the target, the appropriate promotion team placement will get more efficient results. Data already stored at the time of the student registration at Satya Wacana Christian University can be used to create marketing support strategy that can be very useful for the Faculty of Information Technology. Data used in decision support marketing strategy such as data of student origin, study program, and the origin of school that analyzed using K-Means Clustering so that decision making as basis of promotion implementation can be more efficient. The results obtained in the form of information displayed on desktop applications.

Keywords: *K-Means Clustering, Promotion, Student Data.*

ABSTRAK

Biro Promosi Fakultas Teknologi Informasi Universitas Kristen Satya Wacana melakukan teknik pemasaran dengan mengirimkan tim promosi yang secara acak untuk melakukan promosi. Tim Promosi yang dikirim secara acak belum tentu dapat mencapai target yang diharapkan. Agar mencapai target, penempatan tim promosi yang sesuai akan mendapatkan hasil yang lebih efisien. Data yang sudah tersimpan pada saat mahasiswa melakukan registrasi di Universitas Kristen Satya Wacana, dapat dimanfaatkan untuk membuat strategi pendukung pemasaran yang dapat sangat berguna untuk Fakultas Teknologi Informasi tersebut. Data yang dipakai dalam pendukung keputusan strategi pemasaran antara lain data asal mahasiswa, program studi, serta asal sekolah yang dianalisa menggunakan *K-Means Clustering* sehingga pembuatan keputusan sebagai dasar pelaksanaan promosi bisa lebih efisien. Hasil yang didapat berupa informasi yang ditampilkan pada aplikasi *desktop*.

Kata Kunci: *K-Means Clustering, Promosi, Data Mahasiswa.*

PENDAHULUAN

Berdasarkan wawancara dengan ketua Biro Promosi Fakultas Teknologi Informasi Universitas Kristen Satya Wacana selain ketertarikan dari keunggulan sebuah perusahaan atau instansi, hal yang harus dipikirkan juga adalah strategi dalam hal mempromosikan perusahaan atau instansi tersebut, contohnya strategi untuk

penempatan berapa orang dalam tim yang akan mempromosikan perusahaan atau instansi tersebut.

Tim promosi yang dikirim belum tentu dapat mencapai target peminat yang diharapkan. Saat Biro Promosi FTI melakukan promosi yang berupa ekspo, berdasarkan wawancara jumlah orang yang akan dikirim untuk melakukan promosi adalah 3 orang. Dimana jumlah orang yang

dikirim ditentukan secara random dengan jumlah maksimal 3 orang tanpa melihat data mahasiswa yang sudah menjadi mahasiswa di FTI UKSW sehingga seluruh provinsi akan dikirim jumlah anggota promosi yang sama, padahal setiap provinsi memiliki jumlah peminat yang berbeda. Penempatan tim promosi untuk beberapa daerah masih ditentukan secara manual oleh Biro Promosi FTI UKSW, sama halnya dengan penentuan jumlah anggota tim promosi masih dilakukan secara acak tanpa memperhatikan data mahasiswa FTI yang berasal dari daerah tersebut. Hal tersebut menjelaskan bahwa adanya strategi marketing untuk melakukan promosi FTI UKSW itu penting sehingga Biro Promosi bisa lebih efisien saat melakukan promosi terhadap pengiriman jumlah anggota dalam tim promosi disetiap daerah tujuan promosi.

Clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteristik (*similarity*) antara satu data dengan data yang lain, serta *k-means* merupakan pengelompokan data kedalam suatu *cluster* melalui titik yang sudah ditentukan.

Berdasarkan permasalahan tersebut Implementasi Algoritma *K-Means Clustering* untuk strategi marketing di Fakultas Teknologi Informasi Universitas Kristen Satya Wacana ini dapat memberikan strategi yang efisien sehingga tim promosi FTI UKSW yang dikirim untuk melakukan promosi bisa membawa peminat yang lebih banyak serta meminimalisir pengeluaran biaya terhadap tim promosi yang dikirim melalui pengelompokan data yang diolah dari data mahasiswa yang sudah menjadi mahasiswa di FTI UKSW.

*Korespondensi Penulis:

E-mail: nina.setiyawati@staff.uksw.edu

TINJAUAN PUSKATA

Penelitian Terdahulu

Penelitian yang berjudul Implementasi Algoritma *K-Means Clustering* untuk menentukan Strategi *Marketing* President University membahas

tentang pengimplementasian algoritma *K-Means Clustering* ini pada data mahasiswa yang sudah lulus dari President University untuk dilihat kota/kabupaten mana yang menghasilkan lulusan terbanyak dan terbaik dengan mengambil beberapa atribut antara lain, Kota/Kabupaten asal, Program Studi serta IPK dari mahasiswa yang sudah lulus untuk strategi *marketing* dari President University nantinya, pada artikel ini, digunakan *excel* sebagai aplikasi untuk mengimplementasikan algoritma *K-Means Clustering* tersebut (Ong, 2013).

Penelitian lain yang berjudul Data Mining menggunakan Algoritma *K-Means Clustering* untuk menentukan Strategi Promosi Universitas Dian Nuswantoro membahas tentang pengimplementasian algoritma *K-Means Clustering* untuk strategi *marketing* di Universitas Dian Nuswantoro yang menggunakan acuan penelitian yang dijabarkan sebelumnya, yaitu Implementasi Algoritma *K-Means Clustering* untuk menentukan Strategi *Marketing* President University. Penelitian ini dilakukan analisis menggunakan RapidMiner. Selain dari implementasi analisis tersebut, penelitian ini dan penelitian sebelumnya menggunakan data mahasiswa yang sudah lulus dengan mengambil 3 atribut yang sama yakni, Kota/Kabupaten, Program Studi serta IPK (Ramadhani, 2013).

Penelaahan Teori

Berdasarkan penelitian yang pernah dilakukan tentang penerapan algoritma *K-Means Clustering* untuk strategi marketing suatu universitas maka akan dilakukan penelitian yang membahas tentang Analisa Data Mahasiswa Menggunakan Algoritma *K-Means Clustering* Sebagai Dasar Pelaksanaan Promosi (Studi Kasus: Biro Promosi FTI UKSW), dimana pengelompokan data berdasarkan karakteristik yang sama dalam satu *cluster*, bisa menghasilkan informasi yang berguna untuk strategi marketing promosi FTI UKSW. Berbeda dengan penelitian terdahulu yang menggunakan data mahasiswa yang sudah lulus serta 3 atribut yang sama, maka dalam penelitian ini peneliti menggunakan data mahasiswa yang

sudah mendaftar menjadi mahasiswa Fakultas Teknologi Informasi Universitas Kristen Satya Wacana serta menggunakan 3 atribut yang sedikit berbeda yakni Asal Kota/Kabupaten, Program Studi serta Asal sekolah (SMA, SMK/STM, dan SLTA lainnya) yang diproses menggunakan IBM SPSS Statistics sehingga Fakultas Teknologi Informasi Universitas Kristen Satya Wacana dapat membuat strategi marketing yang lebih efisien sebelum mengirimkan tim promosi untuk melakukan promosi.

Data mining adalah suatu proses pencarian korelasi, pola dan tren baru yang berguna dalam media penyimpanan data berukuran besar menggunakan teknologi pengenalan pola seperti teknik-teknik statistik dan matematis. Istilah lain yang sering digunakan antara lain *knowledge mining from data*, *knowledge extraction*, *data/pattern analysis*, *data archeology*, dan *data dredging* (Larose, 2005).

Data mining merupakan suatu kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menentukan keteraturan, pola atau hubungan dalam set data berukuran besar. Salah satu tugas utama dari *data mining* adalah pengelompokan *clustering* dimana *data* yang dikelompokkan belum mempunyai contoh kelompok (Santosa, 2007).

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).

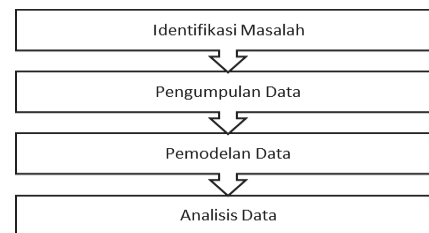
K-means clustering merupakan salah satu metode *data clustering* non-hirarki yang mengelompokkan *data* dalam bentuk satu atau lebih *cluster*/kelompok. *Data*

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).

METODOLOGI PENELITIAN

Langkah-Langkah Penelitian

Tahapan penelitian yang digunakan dalam Implementasi Algoritma *K-Means Clustering* Untuk Strategi *Marketing Promosi* Fakultas Teknologi Informasi Universitas Kristen Satya Wacana, ditunjukkan pada Gambar 1:



Gambar 1. Tahapan Penelitian

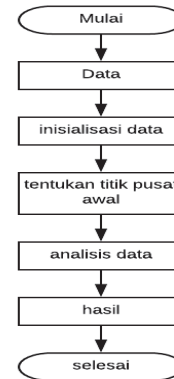
Tahapan penelitian pada Gambar 1 dijelaskan sebagai berikut: tahap 1 adalah identifikasi masalah, berdasarkan hasil wawancara, Biro Promosi FTI UKSW cenderung mengirimkan jumlah anggota tim promosi dengan melihat apakah kegiatan tersebut merupakan kegiatan ekspo atau tidak. Jika kegiatan tersebut ekspo maka dikirim 2-3 orang untuk melakukan promosi, dan ketika kegiatan promosi bukan ekspo maka yang dikirim hanya 1 orang saja. Biro Promosi FTI UKSW mengirimkan tim promosi dengan jumlah anggota secara random tanpa melihat efisiensi dari tindakan tersebut. Pengiriman jumlah tim promosi yang tepat menghasilkan target promosi yang diharapkan, dan membuktikan bahwa kegiatan promosi tersebut efisien. Oleh karena itu penelitian ini dibuat untuk memberikan informasi yang nantinya mendukung strategi *marketing*, sehingga kegiatan promosi dari Biro Promosi FTI menjadi lebih efisien berdasarkan data yang sudah ada. Tahap 2 adalah pengumpulan

data, data yang diperlukan dalam penelitian ini diperoleh melalui bagian akademik Universitas Kristen Satya Wacana yang berupa data mahasiswa yang sudah mendaftar ulang pada saat masuk Universitas Kristen Satya Wacana pada tahun 2017 dan data tersebut telah di *pre-processing* sehingga hanya data mahasiswa yang diperlukan saja yang diambil. Data yang diambil antara lain, Nomor Induk Mahasiswa (NIM), Program Studi, Kota Asal, serta Asal sekolah (SMA, SMK/STM). Berikut ini adalah beberapa data mahasiswa yang sudah mendaftar ulang pada Universitas Kristen Satya Wacana yang diurutkan berdasarkan abjad Kota Asal.

Tabel 1. Contoh Data Mahasiswa yang Telah Mendaftar Ulang pada Universitas Kristen Satya Wacana yang Sudah di *Pre-processing*

NO	NIM	Asal Sekolah	Kota Asal	Program Studi
1	732017002	SMA	Kab. Alor	Ilmu Perpustakaan
2	672017294	SMA	Kab. Banjarnegara	Teknik Informatika
3	682017072	SLTA LAIN	Kab. Banyumas	Sistem Informasi
4	682017088	SMA	Kab. Banyumas	Sistem Informasi
5	682017028	SMK	Kab. Barito Selatan	Sistem Informasi
6	672017701	SMA	Kab. Barito Timur	Teknik Informatika
7	672017235	SMA	Kab. Batang	Teknik Informatika
8	672017288	SMA	Kab. Batang	Teknik Informatika
9	682017106	SMK	Kab. Batang	Sistem Informasi
10	692017067	SMA	Kab. Batang	Desain Komunikasi Visual
...

Tahap 3 adalah pemodelan data, data sebelumnya merupakan data yang belum dapat kita olah karena masih berbentuk karakter, *K-Means Clustering* merupakan sebuah algoritma yang hanya dapat bekerja saat data yang diolah merupakan data yang berupa bilangan atau integer. Maka, data diatas harus diinisialisasi sehingga bisa dianalisa menggunakan algoritma *K-Means Clustering*. Sebelum langkah inialisasi adapun proses analisis *K-Means Clustering* yang dapat kita lihat pada Gambar 2.



Gambar 2. Flowchart Analisis K-Means

Langkah-langkah yang digunakan saat melakukan inialisasi antara lain, nilai yang ada di atribut Kota Asal dibuat pembagian wilayah yang menjadi beberapa bagian wilayah, dibagi berdasarkan provinsi dari kabupaten atau kota. Contohnya Sulawesi Utara yang terdiri dari Kota Manado, Kota Tomohon, Kabupaten Minahasa, Kota Bitung, Kabupaten Kepulauan Talaud, Kabupaten Kepulauan Sangihe dan Kabupaten Bolaang Mongondow, kemudian wilayah-wilayah tersebut diurutkan berdasarkan besar frekuensi asal provinsi ke frekuensi yang paling kecil, setelah itu wilayah yang memiliki frekuensi paling besar diberikan inisial 1, selanjutnya yang kedua besar diberikan inisial 2, dan seterusnya seperti itu sampai ke wilayah yang mempunyai frekuensi terkecil. Berikut adalah data Kota Asal Mahasiswa yang sudah di inialisasikan.

Tabel 2. Data Kota Asal Mahasiswa yang Sudah Diinisialisasi

Provinsi	Frekuensi	Inisial
Jateng	313	1
NTT	33	2
Maluku	26	3
Papua	24	4
Sulut	20	5
...

Setelah menginisialisasi Kota Asal selanjutnya yang harus di inialisasi adalah nilai yang ada di atribut Program Studi, seperti pada Tabel 3.

Tabel 3. Inialisasi Program Studi

Prodi	Frekuensi	Inisial
Teknik Informatika	289	1
Sistem Informasi	115	2
Desain Komunikasi Visual	72	3
Public Relation	29	4

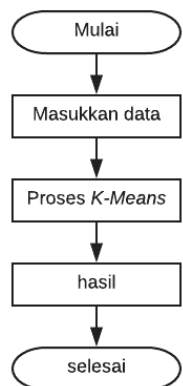
Destinasi Pariwisata	23	5
Pendidikan Teknologi Informasi dan Komunikasi	21	6
Ilmu Perpustakaan	4	7

Sama seperti langkah yang dijelaskan pada penginisialisasian Kota asal, program studi yang paling besar frekuensinya mendapatkan inisial 1, terbesar kedua diinisialisasi 2, dan diinisialkan seterusnya sampai ke frekuensi paling kecil. Selanjutnya inialisasi asal sekolah (SMA, SMK/STM, atau SMTA lain) yang di lakukan sama seperti menginisialisasi Kota Asal dan Program Studi diatas.

Tabel 4. Inialisasi Asal Sekolah

Asal Sekolah	Frekuensi	Inisial
SMA	400	1
SMK/STM	129	2
SLTA Lain	22	3
N/A	2	4

Pada penelitian ini, proses *analisis K-Means* dan hasilnya ditampilkan pada Aplikasi yang dibangun untuk mendukung penelitian ini agar menjadi lebih *valid*. Aplikasi ini dirancang menggunakan *Flowchart* terkait sistem yang dibangun.



Gambar 3. *Flowchart System*

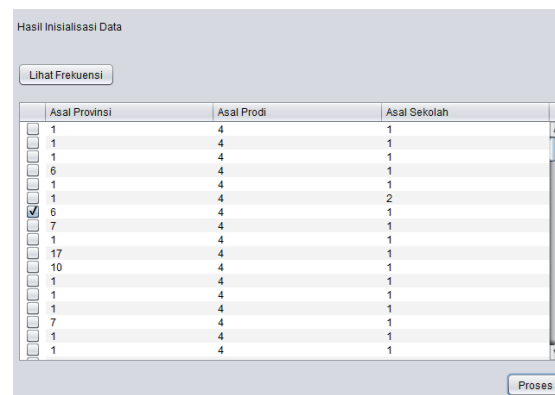
Gambar 3 menunjukkan alur sistem yang berjalan dalam sistem yang dibangun untuk menampilkan hasil dari analisa data menggunakan Algoritma *K-Means Clustering* untuk strategi marketing FTI UKSW. Dimulai dari mengimpor data dalam format xls atau xlsx, setelah itu data yang dimasukkan harus diolah terlebih dahulu dengan menginisialisasi data berdasarkan frekuensi (besar ke kecil). Data

yang sudah diinisialisasi tersebut diolah menggunakan Algoritma *K-Means Clustering* sehingga hasil akhir dari analisis data berbentuk laporan yang menunjukkan pengelompokan data yang terbentuk.

HASIL DAN PEMBAHASAN

Proses analisis *K-Means* pada data mahasiswa dilakukan setelah inialisasi. Pada penelitian ini proses inialisasi sampai interpretasi hasil analisis menggunakan *K-Means* ditampilkan menggunakan aplikasi berbasis *desktop* yang dibangun menggunakan bahasa pemrograman Java. Adapun data diimpor dalam format xls atau xlsx. Proses *K-means* selanjutnya adalah menentukan jumlah *cluster*, menentukan titik awal dari setiap *cluster* sampai jarak terdekat data pada *cluster* yang sudah dibentuk. Berikut langkah-langkah yang dilakukan:

1. Menentukan jumlah cluster yang diinginkan oleh user (Biro Promosi). Dalam penelitian ini data yang ada dikelompokkan menjadi empat *cluster*.
2. Menentukan titik pusat awal (*Centroid* awal) dari setiap *cluster*. Dalam penelitian ini titik pusat awal (*Centroid* awal) ditentukan oleh *user* (Biro Promosi) dengan cara memilih data mahasiswa yang akan dijadikan sebagai titik pusat awal seperti yang terlihat pada Gambar 4:



Gambar 4. Penentuan Titik Pusat Awal pada Aplikasi

Gambar 4 merupakan merupakan penentuan inialisasi titik pusat yang ditentukan random oleh *user*. Selanjutnya,

akan ditunjukkan di tabel 5 empat titik pusat yang digunakan.

Tabel 5. Titik Pusat Awal

Titik Pusat Awal	NIM	Asal Provinsi	Asal Prodi
Cluster 1	602017007	Lampung	Public Relation
Cluster 2	672017008	Jateng	Teknik Informatika
Cluster 3	692017006	Jatim	Desain Komunikasi Visual
Cluster 4	702017019	NTT	Pendidikan Teknologi Informasi dan Komunikasi

3. Menempatkan setiap data pada cluster. Pada penelitian ini digunakan teori Euclidean untuk mengalokasikan setiap data pada cluster terdekat yang ditunjukkan pada Rumus 1.

$$D(i, j) =$$

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

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

Rumus tersebut diimplementasikan pada data yang sudah ada, contohnya adalah sebagai berikut:

$$D(1,1) = \sqrt{(2-6)^2 + (7-4)^2 +}$$

$$D(1,2) = \sqrt{(2-1)^2 + (7-1)^2 +}$$

$$D(1,3) = \sqrt{(2-13)^2 + (7-3)^2 +}$$

$$D(1,4) = \sqrt{(2-2)^2 + (7-6)^2 +}$$

D(1,1) menunjukkan data pertama yang di operasikan dengan titik pusat awal pertama. X_{1i}, X_{2i} dan X_{ki} menjelaskan data setiap parameter yang ada yakni angka 2 data dari parameter Asal Provinsi, angka 7 adalah data dari parameter Asal Prodi dan angka 1 adalah data dari parameter Asal sekolah dimana data tersebut diambil dari data mahasiswa pada field pertama yang diimpor. Sedangkan X_{1j}, X_{2j} dan X_{3j} menjelaskan data setiap parameter dari titik pusat awal pertama, yakni angka 2 dari

parameter Asal Provinsi, angka 6 dari parameter Asal Prodi dan angka 3 dari Asal Sekolah.

Berdasarkan hasil keempat perhitungan tersebut data mahasiswa pertama dimasukkan kedalam cluster 4. Hasil perhitungan selengkapnya untuk beberapa data mahasiswa pertama dapat dilihat pada Tabel 6.

Tabel 6. Hasil Perhitungan Jarak Setiap Data ke Setiap Cluster

No	NIM	Jarak ke-				Jarak terdekat ke cluster
		C1	C2	C3	C4	
1	732017002	5,66	6,16	11,75	2,24	4
2	672017294	5,83	1	12,69	5,48	2
3	682017072	5,74	1,41	12,08	3,16	2
4	682017088	5,39	1,41	12,08	4,58	2
5	682017028	4,58	9,06	5,09	9	3

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

5. Setelah mendapatkan pusat cluster baru, lakukan kembali langkah ketiga

6. Dan, seterusnya sampai titik pusat dari setiap cluster tidak berubah lagi dan tidak ada data yang berpindah cluster.

Dalam penelitian ini, iterasi clustering data terjadi sebanyak 5 kali iterasi. Pada iterasi ke- 5 ini, titik pusat dari setiap cluster sudah tidak berubah dan tidak ada lagi data yang berpindah dari satu cluster ke cluster yang lain sehingga proses iterasi berakhir. Seperti yang ditunjukkan oleh Tabel 7.

Tabel 7. Jumlah Iterasi yang Terjadi

Iterasi	Perubahan pada cluster			
	1	2	3	4
1	1,902	0,884	2,456	2,010
2	0,472	0,000	0,896	0,110
3	0,499	0,033	0,824	0,000
4	0,357	0,000	0,843	0,000
5	0,000	0,000	0,000	0,000

Berdasarkan hasil pengelompokkan data menggunakan metode K-Means Clustering didapatkan hasil clustering hingga iterasi ke- 5. Hasil dari clustering tersebut dapat dilihat pada Tabel 8.

Tabel 8. Hasil Analisis K-Means Clustering

Cluster	Asal Provinsi	Asal Prodi	Asal Sekolah
Cluster 1	Sulut = 20 mahasiswa Lampung = 19 mahasiswa Jabar = 17 mahasiswa Sulsel = 14 mahasiswa Kalbar = 13 mahasiswa Kalteng = 11 mahasiswa Sumut = 10 mahasiswa Papua Barat = 9 mahasiswa Total 113 Mahasiswa	TI = 56 mahasiswa SI = 25 mahasiswa DKV = 14 mahasiswa PR = 9 mahasiswa DesPar = 8 mahasiswa IP = 1 mahasiswa	SMA = 98 mahasiswa SLTA Lain = 8 mahasiswa SMK/STM = 7 mahasiswa
Cluster 2	Jateng = 271 mahasiswa NTT = 23 mahasiswa Papua = 22 mahasiswa Maluku = 19 mahasiswa Total 335 Mahasiswa	TI = 205 mahasiswa SI = 81 mahasiswa DKV = 49 mahasiswa	SMA = 224 mahasiswa SMK/STM = 97 mahasiswa SLTA Lain = 14 mahasiswa
Cluster 3	Jatim = 7 mahasiswa Banten = 7 mahasiswa Sulteng = 6 mahasiswa DKI Jakarta = 5 mahasiswa Sumsel = 5 mahasiswa N/A = 4 mahasiswa Maluku Utara = 4 mahasiswa DIY = 3 mahasiswa Sultenggara = 2 mahasiswa Kaltim = 2 mahasiswa Bali = 1 mahasiswa Kalmut = 1 mahasiswa Kalut = 1 mahasiswa Riau = 1 mahasiswa Bengkulu = 1 mahasiswa Jambi = 1 mahasiswa Bangka belitung = 1 mahasiswa Total 52 Mahasiswa	TI = 28 mahasiswa SI = 9 mahasiswa DKV = 9 mahasiswa PR = 3 mahasiswa DesPar = 2 mahasiswa PTIK = 1 mahasiswa	SMA = 36 mahasiswa SMK/STM = 8 mahasiswa SLTA Lain = 6 Mahasiswa N/A = 2
Cluster 4	Jateng = 40 mahasiswa NTT = 10 mahasiswa Maluku = 2 mahasiswa Papua = 1 mahasiswa Total 53 Mahasiswa	PTIK = 20 mahasiswa PR = 17 mahasiswa DesPar = 13 mahasiswa IP = 3 mahasiswa	SMA = 39 mahasiswa SMK/STM = 10 mahasiswa SLTA Lain = 4 mahasiswa

Berikut adalah *output* hasil perhitungan menggunakan algoritma *K-Means* pada aplikasi.

Asal Provinsi	Asal Prodi	Asal Sekolah
Sulut = 20	TEKNIK INFORMATIKA = 56	SMA = 98
Lampung = 19	SISTEM INFORMASI = 25	SLTA lain = 8
Jabar = 17	DESAIN KOMUNIKASI VISUAL = 14	SMK/STM = 7
Sulsel = 14	PUBLIC RELATION = 9	
Kalbar = 13	DESTINASI PARIWISATA = 8	
Kalteng = 11	ILMU PERPUSTAKAAN = 1	
Sumut = 10		
Papua Barat = 9		

Gambar 5. Hasil Akhir Cluster 1

Dilakukan pengujian proses *clustering* sebanyak empat *cluster*. Pada Gambar 5 dapat dilihat bahwa *cluster* satu didominasi oleh mahasiswa yang berasal provinsi dari Sulawesi Utara dan Lampung yakni 20 mahasiswa dan 19 mahasiswa. Sedangkan berdasarkan minat program studi adalah Teknik Informatika yakni 56 mahasiswa dan Sistem Informasi yang berjumlah 25 mahasiswa. Yang dimana

rata-rata mahasiswa berasal dari SMA. *Cluster* ke-dua didominasi oleh Jawa Tengah yakni 271 mahasiswa dan minat program studi Teknik Informatika yang berjumlah 205 mahasiswa dan rata-rata mahasiswa masih berasal dari SMA. Selanjutnya, *Cluster* ke-tiga didominasi oleh Jawa Timur dan Banten yang memiliki jumlah yang sama yakni 7 mahasiswa, yang minat program studinya masih dari Teknik Informatika dan Sistem Informasi dan lulusan dari SMA. Terakhir, *cluster* keempat didominasi oleh Jawa Tengah dengan jumlah mahasiswa 40 mahasiswa dan Nusa Tenggara Timur yang berjumlah 10 mahasiswa. Yang dimana memiliki minat program studi Pendidikan Teknologi Informasi dan Komunikasi yang berjumlah 20 mahasiswa dan *Public Relation* 17 mahasiswa yang rata-rata mahasiswa lulusan dari SMA.

Proses perhitungan menggunakan *K-Means* juga dilakukan melalui aplikasi lain

yakni IBM SPSS menggunakan data yang sama yaitu data mahasiswa tahun 2017. Hasil dari SPSS digunakan untuk membandingkan hasil perhitungan yang didapatkan dari aplikasi yang dibangun sesuai dengan hasil dari SPSS atau tidak. Hasil yang didapat dari perhitungan melalui SPSS sama dengan hasil yang didapat pada proses perhitungan melalui aplikasi yang dibangun pada penelitian ini. Proses iterasi yang didapat dari SPSS yakni 5 iterasi, dapat dilihat pada Gambar 6.

Iteration History ^a				
Iteration	Change in Cluster Centers			
	1	2	3	4
1	1,902	,884	2,456	2,010
2	,472	,000	,896	,110
3	,499	,033	,824	,000
4	,357	,000	,843	,000
5	,000	,000	,000	,000

a. Convergence achieved due to no or small change in cluster centers. The maximum absolute coordinate change for any center is ,000. The current iteration is 5. The minimum distance between initial centers is 4,899.

Gambar 6. Proses Iterasi Pengujian SPSS

Gambar 6 adalah proses iterasi yang dilakukan melalui SPSS, sedangkan proses iterasi yang terjadi pada aplikasi terlihat pada Tabel 7. Dari kedua gambar tersebut dapat disimpulkan bahwa proses iterasi pada SPSS sama dengan proses iterasi pada aplikasi. Adapun jumlah data mahasiswa yang dianalisa menggunakan aplikasi dan SPSS menghasilkan data yang sesuai seperti yang ditunjukkan pada Tabel 7, dimana cluster 1 memiliki 113 mahasiswa, cluster 2 memiliki 335 mahasiswa, cluster 3 memiliki 52 mahasiswa dan cluster 4 memiliki 53 mahasiswa. Jumlah mahasiswa untuk setiap cluster pada perhitungan *K-Means* yang dilakukan pada SPSS terlihat pada Gambar 6.

Number of Cases in each Cluster

Cluster	1	113,000
	2	335,000
	3	52,000
	4	53,000
Valid		553,000
Missing		,000

Gambar 7. Hasil Jumlah Data Setiap Cluster

Hasil titik pusat akhir yang didapat dari perbandingan menggunakan SPSS sesuai dengan yang dianalisa pada Gambar 8 hanya saja SPSS menggunakan pembulatan, dapat dilihat pada Gambar 8.

Final Cluster Centers

	Cluster			
	1	2	3	4
V2	8	1	18	1
V3	2	2	2	5
V4	1	1	2	1

Gambar 8. Titik Pusat Akhir Menggunakan SPSS

Pengujian hasil perhitungan *K-Means* juga dilakukan dengan membandingkan hasil pengelompokan data mahasiswa secara manual sesuai atau tidak dengan hasil perhitungan menggunakan aplikasi yang dibangun. Pada proses pengelompokan secara manual, data mahasiswa berdasarkan asal provinsi diambil secara acak, kemudian dihitung jumlah mahasiswa untuk setiap prodi dan asal sekolah. Hasil pengelompokan secara manual terlihat pada Tabel 9.

Tabel 9. Hasil Perhitungan Manual

Cluster	Asal Provinsi	Asal Prodi	Asal Sekolah
1	Sulut = 20 mahasiswa Lampung = 19 mahasiswa Jabar = 17 mahasiswa Sulsel = 14 mahasiswa Kalbar = 13 mahasiswa Kalteng = 11 mahasiswa Sumut = 10 mahasiswa Papua Barat = 9 mahasiswa Total 113 Mahasiswa	TI = 56 mahasiswa SI = 25 mahasiswa DKV = 14 mahasiswa PR = 9 mahasiswa DesPar = 8 mahasiswa IP = 1 mahasiswa	SMA = 98 mahasiswa SLTA Lain = 8 mahasiswa SMK/STM = 7 mahasiswa
2	Jateng = 271 mahasiswa NTT = 23 mahasiswa Papua = 22 mahasiswa Maluku = 19 mahasiswa Total 335 Mahasiswa	TI = 205 mahasiswa SI = 81 mahasiswa DKV = 32 mahasiswa PR = 17 mahasiswa	SMA = 229 mahasiswa SMK/STM = 91 mahasiswa SLTA Lain = 15 mahasiswa
3	Jatim = 7 mahasiswa Banten = 7 mahasiswa Sulteng = 6 mahasiswa DKI Jakarta = 5 mahasiswa Sumsel = 5 mahasiswa N/A = 4 mahasiswa Maluku Utara = 4 mahasiswa DIY = 3 mahasiswa Sultenggara = 2 mahasiswa Kaltim = 2 mahasiswa Bali = 1 mahasiswa Kalmut = 1 mahasiswa Kalut = 1 mahasiswa Riau = 1 mahasiswa Bengkulu = 1 mahasiswa Jambi = 1 mahasiswa Bangka belitung = 1 mahasiswa Total 52 Mahasiswa	TI = 28 mahasiswa SI = 9 mahasiswa DKV = 9 mahasiswa PR = 3 mahasiswa DesPar = 2 mahasiswa PTIK = 1 mahasiswa	SMA = 36 mahasiswa SMK/STM = 8 mahasiswa SLTA Lain = 6 Mahasiswa N/A = 2
4	Jateng = 40 Mahasiswa NTT = 10 mahasiswa Maluku = 2 mahasiswa Papua = 1 mahasiswa Total 53 Mahasiswa	TI = 36 mahasiswa PR = 17 Mahasiswa	SMA = 47 Mahasiswa SMK/STM = 5 mahasiswa SLTA Lain = 1 mahasiswa

Dari Tabel 9 dan Tabel 8, terlihat bahwa pada cluster 1 data yang dihasilkan sama dimana Sulut mempunyai 20 mahasiswa, lampung mempunyai 19 mahasiswa dimana prodi yang diminati adalah Teknik Informatika dan Sistem Informasi rata-rata berasal dari SMA. Pada cluster 2 hasil perhitungan *K-Means* pada aplikasi memiliki beberapa perbedaan yang tidak terlalu jauh dengan pengelompokan secara manual yaitu pada pengelompokan manual jumlah mahasiswa DKV adalah 32 orang sedangkan pengelompokan menggunakan aplikasi jumlah mahasiswa DKV adalah 49, serta prodi PR yang muncul pada pengelompokan secara manual dan tidak ada pada pengelompokan pada aplikasi. Pada cluster 3 pengelompokan data dari hasil perhitungan *K-Means* pada aplikasi dan secara manual sesuai. Pada *cluster* 4 pengelompokan data dari hasil perhitungan *K-Means* pada

aplikasi dan secara manual memiliki perbedaan yakni prodi IP memiliki 3 mahasiswa sedangkan pada aplikasi tidak ada prodi IP. Hasil yang didapatkan secara manual terlihat pada Tabel 9.

Pengelompokan data terbentuk empat *cluster* yaitu, *cluster* satu yang berjumlah 113 mahasiswa dengan rata-rata mahasiswa berasal dari SMA memiliki minat program studi Teknik Informatika dan Sistem Informasi, *cluster* dua yang berjumlah 335 mahasiswa dengan rata-rata mahasiswa berasal dari SMA memiliki minat program studi Teknik Informatika dan Sistem Informasi, *cluster* tiga yang berjumlah 52 mahasiswa dimana rata-rata mahasiswa berasal dari SMA dan memiliki minat program studi yang sama dengan *cluster* satu dan dua yakni Teknik Informatika dan Sistem Informasi, dan *cluster* empat berjumlah 53 mahasiswa yang rata-rata mahasiswa berasal dari SMA memiliki

minat program studi Pendidikan Teknologi Informasi dan Komputer serta *Public Relation*.

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan penelitian yang dilakukan maka dapat disimpulkan bahwa analisa data mahasiswa menggunakan Algoritma *K-Means Clustering* menghasilkan pengelompokan data mahasiswa berdasarkan asal provinsi, prodi dan asal sekolah. Pemilihan titik *centroid* awal dipilih oleh user agar dapat melihat titik awal yang dipilih.

Berdasarkan pengujian terhadap hasil perhitungan *K-Means* secara aplikasi didapat tidak jauh berbeda dengan perhitungan secara manual. Dimana perbedaan yang ada hanya pada *cluster 2* yang memiliki jumlah mahasiswa DKV yang berbeda serta prodi PR yang masuk pada *cluster 2* dan prodi IP pada *cluster 4* pada saat pengujian secara manual.

Dari hasil pengelompokan menggunakan *K-Means*, Biro Promosi dapat melihat asal daerah yang memiliki jumlah mahasiswa yang besar pada suatu prodi serta asal sekolah mahasiswa yang dominan pada suatu prodi. Seperti pada *cluster 1*, provinsi Sulawesi Utara dan Lampung mempunyai mahasiswa yang banyak di program studi Teknik Informatika dan Sistem Informasi. Informasi tersebut dapat digunakan Biro Promosi sebagai dasar dari pelaksanaan promosi.

Saran

Berdasarkan hasil penulisan yang telah dilakukan maka diharapkan penulisan selanjutnya untuk menguji dengan algoritma clustering lainnya seperti *Fuzzy C-Means* atau metode algoritma yang lain, serta aplikasi yang dibangun bisa dikembangkan lagi menjadi lebih *user friendly*.

DAFTAR PUSTAKA

- Agusta, Y., 2007. *K-Means – Penerapan, Permasalahan dan Metode Terkait: Jurusan Informatika*, Fakultas Sistem dan Informatika, STMIK STIKOM Bali.
- Larose, D., 2005. *Discovering Knowledge In Data: An Introduction to Data Mining*. Kota: Hoboken, New Jersey, Penerbit: A John Wiley & Sons, inc., Publication.
- Ong, J. O., 2013. *Implementasi Algoritma K-Means Clustering untuk Menentukan Strategi Marketing President University*, Jurnal Ilmiah Teknik Industri: Jurusan Teknik Industri, Fakultas Teknik, President University.
- Ramadhani, R. D., 2013. *Data Mining Menggunakan Algoritma K-Means Clustering untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro: Jurusan Sistem Informasi*, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro.
- Santosa, B., 2007. *Data Mining Terapan dengan MATLAB*. Kota: Yogyakarta, Penerbit: Graha Ilmu.