

# Penggunaan Algoritma *Clustering K-means* Untuk Melihat Daerah-Daerah Penyuplai Mahasiswa Di Biro Promosi UKSW

<http://dx.doi.org/10.28932/jutisi.v5i2.1968>

Rivort Pormes <sup>✉#1</sup>, Daniel H. F. Manongga<sup>\*2</sup>

# *Magister Sistem Informasi, Universitas Kristen Satya Wacana*

*Jl. Diponegoro 52-60 Salatiga, 50711-Indonesia*

<sup>1</sup>*rivopormes@gmail.com*

\* *Magister Sistem Informasi, Universitas Kristen Satya Wacana*

*Jl. Diponegoro 52-60 Salatiga, 50711-Indonesia*

<sup>2</sup>*dmanongga@yahoo.co.uk*

**Abstract**— The use of K-means algorithm in students supply process is done to see students supply tendency in each region who study at Satya Wacana Christian University (SWCU). This data processing with K-means is useful for extracting information and knowledge from students data at SWCU. With data mining processing, the stakeholders at SWCU can take strategic steps for the penetration process in regions with potential student supply indication. The use of K-means algorithms facilitates the process of data analysis as well as grouping student data during a 5-year period. The aim of this study is to provide on-target strategic picture for regions that potentially have a significant impact on students supply each year.

**Keywords**— *Clustering; K-means; Data mining.*

## I. PENDAHULUAN

Penerimaan mahasiswa baru adalah kegiatan tahunan yang dilakukan oleh setiap Universitas yang ada di Indonesia. Universitas Kristen Satya Wacana Setiap tahunnya mengirimkan tim promosi untuk mendapatkan peminat dari tiap sekolah yang di daerah atau provinsi yang berada di hampir seluruh wilayah Indonesia.

Tahapan promosi biasanya dilakukan dengan cara kunjungan secara langsung ke sekolah pada provinsi yang telah di tentukan. Proses wawancara pada biro promosi Universitas Kristen Satya Wacana dijelaskan bahwa, promosi masih belum berjalan maksimal menjaring target calon peminat di tiap daerah yang di kunjungi karna terbatasnya waktu dalam penjangkaran di provinsi-provinsi, serta kurangnya kelengkapan dari data calon mahasiswa yang di dapat tim promosi pada saat turun secara langsung ke daerah-daerah yang dituju.

Data yang terkumpul dan berada pada biro promosi selama ini tidak dilakukan proses pengolahan untuk mendapatkan *knowledge* yang tersembunyi dari data-data

tersebut. Biro promosi masih menggunakan acuan data dari mahasiswa yang telah berkuliah di UKSW sebelumnya namun belum menggunakan pengolahan *data mining* dalam proses pengolahan dari data-data tersebut, sehingga *knowledge* yang di dapat pun terbatas hanya kepada data mahasiswa pada umumnya, tanpa dapat dikreasikan untuk mendapatkan informasi berharga yang terdapat pada data-data tersebut.

Melihat pentingnya hal tersebut dan pentingnya dalam keberlangsungan biro promosi yang tugas dan tanggung jawabnya dalam menjaring serta melakukan promosi di tiap daerah maka perlu dilakukan proses pengolahan yang sistematis serta terukur guna mempermudah biro promosi UKSW saat ke daerah-daerah pada provinsi yang dituju.

*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.

Dari masalah tersebut penelitian ini bertujuan untuk melihat daerah-daerah penyuplai mahasiswa serta pola-pola menarik dari hasil pengolahan data di Universitas Kristen Satya Wacana dengan pemanfaatan proses *data mining* menggunakan *algoritma Clustering K-means* guna mempermudah biro promosi Universitas Kristen Satya Wacana dalam menentukan strategi saat menjalankan penjangkaran serta promosi di tiap-tiap daerah yang menjadi prioritas serta daerah berpotensi memiliki kenaikan signifikan dalam menyuplai mahasiswa dan membatasi ruang lingkup penjangkaran pada daerah-daerah yang memberikan suplai mahasiswa terkecil pada Universitas Kristen Satya Wacana.

## II. PENELITIAN TERKAIT

Penelitian Asril, Wiza, Yuneфри, tentang Analisis Data Lulusan dengan Data Mining untuk Mendukung Strategi

Promosi Universitas Lancang Kuning, penelitian ini berfokus mengamati beberapa variabel penelitian yang sering dipertimbangkan oleh perguruan tinggi dalam menentukan sasaran promosi yaitu asal sekolah, daerah, dan jurusan. Hasil penelitian ini berupa pola menarik hasil *data mining* yang merupakan informasi penting untuk mendukung strategi promosi yang tepat dalam mendapatkan calon mahasiswa baru.[1]

Nugroho dengan Analisa Potensi Mahasiswa Di Daerah Bali Menggunakan Pendekatan *K-means clustering*, penelitian ini bertujuan mengetahui potensi mahasiswa di setiap daerah di pulau Bali ini untuk keperluan analisis, baik analisis untuk marketing ataupun pengembangan lainnya. Sehingga dalam penyajian laporan dan informasi dapat diperoleh data distribusi potensi mahasiswa di masing-masing daerah dengan maksud untuk kegiatan marketing di tahun-tahun berikutnya.[2]

Ong membahas Implementasi Algoritma *K-means clustering* Untuk Menentukan Strategi Marketing President University, penelitian ini berfokus pengambilan data mahasiswa yang sudah lulus untuk memprediksikan strategi promosi yang akan dilakukan di kota-kota di Indonesia. Informasi yang didapatkan dari hasil penelitian ini dapat digunakan sebagai referensi dalam menentukan strategi yang tepat bagi tim marketing dalam kegiatan promosi sehingga akan lebih efektif dan efisien.[3]

Prasetyo, Saharuna, Abdul menulis Analisis Data Atribut Mahasiswa untuk Menentukan Strategi Promosi Kampus Menggunakan Metode *Data Mining*. Penelitian ini bertujuan untuk menentukan strategi promosi yang efektif dan efisien sehingga informasi promosi dapat tersebar di setiap daerah. Pada penelitian ini, data yang digunakan adalah data peminat PNUP yang terdiri dari 12.850 data dengan menggunakan tiga atribut, yaitu jenis kelamin, program studi, dan daerah asal. Teknik Analisis data menggunakan metode *data mining* dengan teknik *clustering* menggunakan algoritma *K-means* dan *tools* yang digunakan yaitu *Matlab*. Berdasarkan hasil penelitian ini, ditemukan pola atau informasi baru pada data peminat. Data peminat terbentuk tiga *cluster*, yaitu *cluster* peminat banyak, sedang, dan sedikit. Dari hasil *cluster* tersebut, terbentuk strategi promosi yang dapat menjadi rujukan tim promosi PNUP untuk melakukan promosi yang lebih efektif dan efisien.[4]

Setiawan menulis Penerapan *Data Mining* Menggunakan Algoritma *K-means clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik LP3I Jakarta) Penelitian ini membahas tentang penerapan *data mining*, menggunakan algoritma *K-means clustering* untuk menghasilkan profil yang memiliki kemiripan atribut yang sama. Metode yang digunakan adalah CRISPDM dengan melalui proses *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment*. Implementasi proses *K-Means Clustering* menggunakan *Rapid Miner 7.0*. *data mining* dengan algoritma *K-Means Clustering* dapat memberikan pengetahuan untuk menentukan strategi promosi di Politeknik LP3I Jakarta.

Pengetahuan yang didapat dari hasil *clustering* yaitu calon mahasiswa didominasi dari masyarakat ekonomi rendah dan menengah. [5]

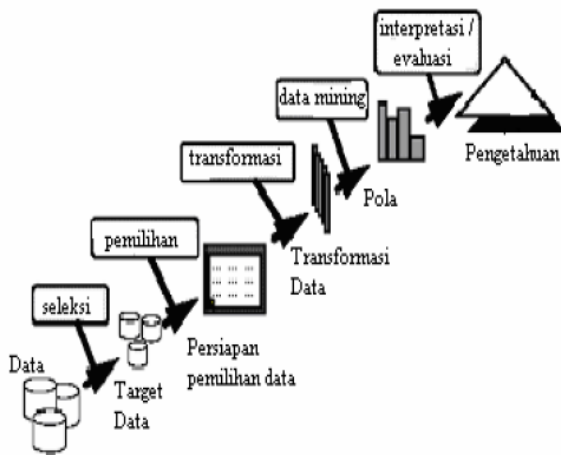
Asroni dan Adrian dengan Penerapan penelitian Metode *K-means* Untuk *clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan *Weka Interface* Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang, Menguji data yang telah ada di data *warehouse* Universitas Muhammadiyah Magelang untuk memudahkan untuk mencari 5 orang mahasiswa pada jurusan Teknik Informatika dalam melakukan penyeleksian untuk mengikuti lomba. Kriteria tersebut diproses dengan menggunakan metode pengelompokan *K-Means*. Diharapkan dengan adanya pengujian data ini dapat memberikan rekomendasi mahasiswa yang layak maju sebagai peserta *event Cyber* jawara.[6]

Ramadhani pada penelitian Data Mining Menggunakan Algoritma *K-Means Clustering* Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro. 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*. Hasil dari penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk menentukan strategi promosi berdasarkan *cluster* yang terbentuk oleh pihak admisi UDINUS.[7]

### III. TINJAUAN PUSTAKA

#### A. DATA MINING

*Data mining* (DM) adalah serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Patut diingat bahwa kata mining sendiri berarti usaha untuk mendapatkan sedikit barang berharga dari sejumlah besar material dasar. Karena itu DM sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (*artificial intelligent*), *machine learning*, statistik dan basis data. Beberapa teknik yang sering disebut-sebut dalam literatur DM antara lain: *clustering*, *classification*, *association rule mining*, *neural network*, dan *genetic algorithm*. [8]



Gambar 1. Tahap-tahap data mining

Tahapan data mining pada Gambar 1 dibagi menjadi enam bagian yaitu :

1. Pembersihan data (*data cleaning*) Sebelum proses data mining dapat dilaksanakan, perlu dilakukan proses cleaning pada data yang menjadi fokus KDD. Proses cleaning mencakup antara lain membuang 4 duplikasi data, memeriksa data yang inkonsisten, dan memperbaiki kesalahan pada data, seperti kesalahan cetak (tipografi). Juga dilakukan proses *enrichment*, yaitu proses “memperkaya” data yang sudah ada dengan data atau informasi lain yang relevan dan diperlukan untuk KDD, seperti data atau informasi eksternal.
2. Integrasi data (*data integration*) Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk *data mining* tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau *file teks*. Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggandan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.
3. Seleksi Data (*Data Selection*) Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang

membeli dalam kasus *market basket analysis*, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.

4. Transformasi data (*Data Transformation*) Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan *clustering* hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagi-bagi menjadi beberapa interval. Proses ini sering disebut transformasi data.
5. Proses *mining* adalah sebuah proses yang paling utama pada saat metode diterapkan untuk mencari pengetahuan tersembunyi dan berharga dari data.
6. Evaluasi pola (*pattern evaluation*), Untuk mengidentifikasi pola-pola menarik kedalam *knowledge based* yang ditemukan. Dalam tahap ini hasil dari teknik *data mining* berupa pola-pola yang khas maupun model prediksi di evaluasi untuk menilai apakah hipotesa yang ada memang tercapai.
7. Presentasi pengetahuan (*knowledge presentation*), Merupakan penyajian dan visualisasi pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses *data mining* adalah bagaimana memformulasikan keputusan atau 5 aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami *data mining*. Karenanya presentasi hasil *data mining* dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses *data mining*. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil *data mining* [8].

#### B. Clustering

*Clustering* adalah proses pengelompokan benda serupa ke dalam kelompok yang berbeda, atau lebih tepatnya partisi dari sebuah data set kedalam subset, sehingga data dalam setiap subset memiliki arti yang bermanfaat. Sebuah *cluster* terdiri dari kumpulan benda-benda yang mirip antara satu dengan yang lainnya dan berbeda dengan benda yang terdapat pada *cluster* lainnya. Algoritma *clustering* terdiri dari dua bagian yaitu secara hirarkis dan secara partitional. Algoritma hirarkis menemukan *cluster* secara berurutan dimana *cluster* ditetapkan sebelumnya, sedangkan algoritma partitional menentukan semua kelompok pada waktu tertentu.[9]

*Clustering* atau klasifikasi adalah metode yang digunakan untuk membagi rangkaian data menjadi beberapa group berdasarkan kesamaan-kesamaan yang telah ditentukan sebelumnya. [11]

### C. K-means

Algoritma *K-means* merupakan salah satu algoritma dengan partitional, karena *K-means* didasarkan pada penentuan jumlah awal kelompok dengan mendefinisikan nilai *centroid* awalnya. Algoritma *K-means* menggunakan proses secara berulang-ulang untuk mendapatkan basis data *cluster*. Dibutuhkan jumlah *cluster* awal yang diinginkan sebagai masukan dan menghasilkan jumlah *cluster* akhir sebagai output.

Algoritma *K-means* merupakan Algoritma pengelompokan iteratif yang melakukan *partisiet* data kedalam sejumlah *K cluster* yang sudah ditetapkan diawal. Algoritma *K-means* sederhana untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, umum penggunaannya dalam praktek. Secara historis, *K-Means* menjadi salah satu algoritma yang paling penting dalam bidang *data mining* [12]

Jika algoritma diperlukan untuk menghasilkan *cluster K* maka akan ada *K* awal dan *K* akhir. Metode *K-means* akan memilih pola *K* sebagai titik awal *centroid* secara acak. Jumlah iterasi untuk mencapai *cluster centroid* akan dipengaruhi oleh calon *cluster centroid* awal secara random dimana jika posisi *centroid* baru tidak berubah. Data yang memiliki jarak pendek atau terdekat dengan *centroid* akan membentuk sebuah *cluster*. [9]

Algoritma *K-means*

1. Tentukan *k* sebagai jumlah *cluster* yang akan dibentuk
2. Tentukan *k Centroid* (titik pusat *cluster*) awal secara random/acak.

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i=1,2,3,...,n \quad (1)$$

Rumus *K-mean*

Dimana;

*v* : *centroid* pada *cluster*

*x<sub>i</sub>* : objek ke-*i*

*n* : banyaknya objek/jumlah objek yang menjadi anggota *cluster*

3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *cluster*. Untuk menghitung jarak antara objek dengan *centroid* dapat menggunakan *Euclidian Distance*.

$$d(x,y) = \|x-y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i=1,2,3,...,n \quad (2)$$

Rumus *Euclidian Distance*

Dimana;

*x<sub>i</sub>* : objek x ke-*i*

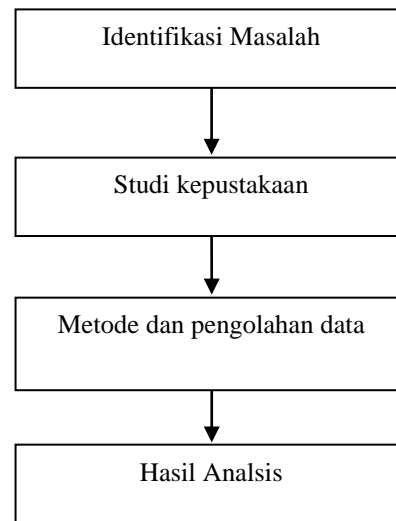
*y<sub>i</sub>*: daya y ke-*i*

*n* : banyaknya objek

4. Alokasikan masing-masing objek ke dalam *centroid* yang paling dekat
5. Lakukan iterasi, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (1)
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama. [10]

## IV. METODE PENELITIAN

Pada metode penelitian ini diuraikan langkah-langkah sistematis dan terarah yang akan dijadikan acuan sebagai kerangka penelitian. Tahapan penelitian ini dibagi menjadi Empat bagian yaitu, (1) Identifikasi masalah, (2) studi kepustakaan, (3) metode, dan (4) hasil analisis.

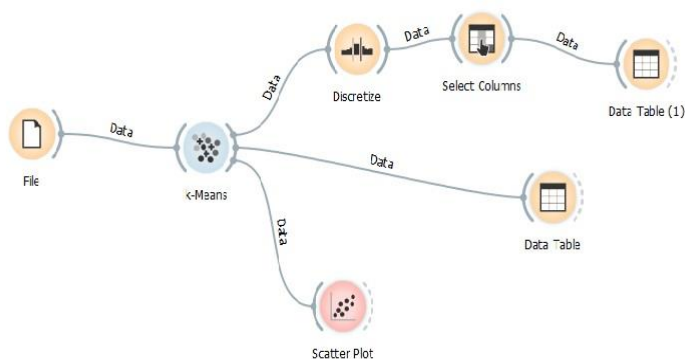


Gambar 2. Tahapan penelitian

Pada Gambar 2 penelitian ini akan mencakup 4 tahapan. Tahap pertama dilakukan identifikasi masalah seperti yang telah di jelaskan pada latar belakang. Tahap kedua yaitu studi kepustakaan, yang mencakup literatur serta penelitian-penelitian terkait, tahap ketiga, penggunaan metode *clustering* dengan algoritma *K-Means* dalam proses pengolahan data selama kurang waktu 5 tahun, dan tahap keempat yaitu hasil analisis dari proses pengolahan data yang telah diterjemahkan dalam bentuk rekomendasi, *knowledge* baru serta gambaran kepada biro promosi Universitas Kristen Satya Wacana dalam menentukan strategi penjangkaran yang lebih sistematis. Diharapkan penelitian ini dapat memberikan dampak yang cukup signifikan bagi Biro Promosi Universitas Kristen Satya Wacana.

## V. HASIL DAN PEMBAHASAN

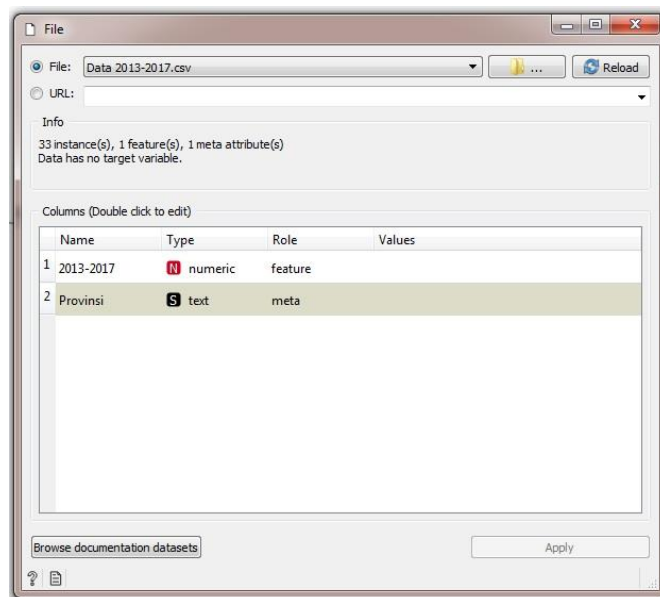
Pada proses pengolahan data digunakan metode *clustering* dengan algoritma *K-means*. Pada proses pengolahan dilakukan dengan menginisialisasi 5 *cluster* untuk mengelompokkan data-data selama kurun waktu 5 tahun. Dengan penggunaan 5 buah *cluster* mempermudah dalam menentukan skala prioritas, potensi suplai, serta batasan atau ruang lingkup pada data 5 tahun dalam kurun waktu 2013-2017.



Gambar 3. Proses pengolahan Program Orange

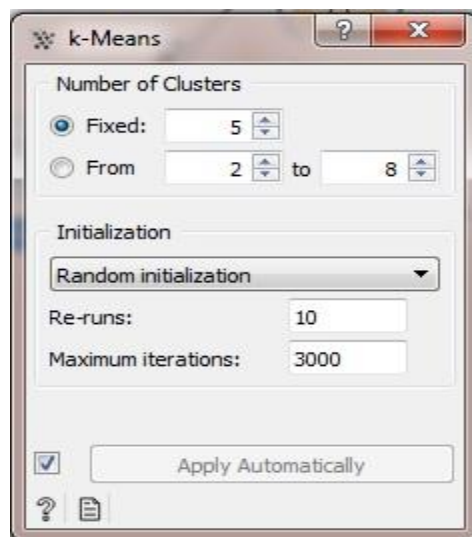
Pada Gambar 3 adalah proses pengolahan data mining menggunakan program Orange. Ada beberapa tools yang digunakan dalam proses pengolahan pada gambar 3, antara lain:

1. *file*: berisi data berkecstensi *.csv* yang telah di normalisasi sebelumnya.
2. *K-mean* : *algoritma* yang digunakan dalam proses pengolahan. Pada proses *k-mean* di tentukan sebelumnya jumlah *cluster* yang akan digunakan.
3. *Scatter plot*: *tools* untuk visualisasi hasil clustering
4. *Data table*: visualisasi data-data hasil inputan serta hasil *cluster* dalam bentuk tabel.
5. *Descrctize* : *tools* untuk memberikan batasan dan nilai rata-rata pada data-data yang diinputkan
6. *Select Columns* : *tool* yang digunakan untuk mengolah hasil *meta attributes* serta data *features* dari *file* data dan selanjutnya ditampilkan ke *data table*



Gambar 4. File data mahasiswa tahun 2014-2017

Pada Gambar 4 merupakan data *.csv* dari kurun waktu 5 tahun yang telah melewati proses normalisasi untuk menghilangkan data-data anomali serta data-data yang tidak memiliki nilai. Terdapat 2 *attribute* yang ditampilkan, pertama adalah data 2013-2017 yang terbaca sebagai data *numeric* dan data provinsi yang terbaca sebagai data *text*.



Gambar 5. Proses menentukan jumlah *cluster*

Pada Gambar 5 merupakan proses penentuan jumlah *cluster*. Pada gambar di atas menggunakan 5 buah *cluster* dalam proses pengolahan data dengan proses maksimum iterasi sebanyak 3000 kali.

D. Hasil *clustering K-means*

Pada data hasil pengolahan selama 5 tahun terdapat atribut yang menjadi acuan pengolahan, yaitu provinsi, jumlah mahasiswa, serta tahun, dengan menginialisasi 5 *cluster*. Pada hasil *cluster* juga menunjukkan provinsi lain memiliki potensi suplai signifikan dalam menyuplai mahasiswa dalam tahun-tahun berikutnya.

TABEL I  
HASIL *CLUSTERING* DATA 2013-2017

Provinsi	Cluster	Silhouette	2013-2017
25 Jawa Tengah	C1	0.500	9666
12 NTT	C2	0.664	762
11 Papua	C2	0.677	686
26 Jawa Barat	C2	0.601	555
16 Maluku	C2	0.599	554
4 Sulawesi Utara	C5	0.690	438
23 Kalimantan Barat	C5	0.711	417
7 Sulawesi Selatan	C5	0.713	391
18 Lampung	C5	0.695	359
1 Sumatera Utara	C5	0.637	327
24 Jawa Timur	C3	0.638	252
15 Maluku Utara	C3	0.679	233
21 Kalimantan Tengah	C3	0.683	230
32 Banten	C3	0.692	175
29 DKI Jakarta	C3	0.689	171
6 Sulawesi Tengah	C3	0.671	159
20 Kalimantan Timur	C3	0.460	110
2 Sumatera Selatan	C4	0.624	86
33 Bali	C4	0.675	70
9 Riau	C4	0.686	65
22 Kalimantan Selatan	C4	0.720	40
5 Sulawesi Tenggara	C4	0.726	33
30 D.I. Yogyakarta	C4	0.727	32
27 Jambi	C4	0.730	25
31 Bengkulu	C4	0.730	23
3 Sumatera Barat	C4	0.731	19
17 Luar Negeri	C4	0.731	14
13 NTB	C4	0.730	12
8 Sulawesi Barat	C4	0.730	12
19 Kepulauan Bangka Belitung	C4	0.730	10
14 NAD	C4	0.729	8
10 Papua Barat	C4	0.728	7
28 Gorontalo	C4	0.723	1

Pada data di tabel I menunjukkan 33 provinsi dengan jumlah suplai berbeda selama kurun 5 tahun terakhir.

Dengan jumlah tersebut, biro promosi memiliki keterbatasan untuk dapat melakukan proses promosi dan penjangring pada semua provinsi tersebut. Dengan model pengolahan *cluster K-Means* memfokuskan daerah mana yang menjadi perhatian dan konsentrasi dalam proses penjangringan. Pada tabel I provinsi-provinsi di bagi atau dipartisi berdasarkan kemiripan nilai dan di bagi menjadi 5 *cluster* dengan label C1, C2, C3, C4, dan C5.

TABEL II  
*CLUSTER* PROVINSI JAWA TENGAH

Provinsi	Cluster	Silhouette	2013-2017
25 Jawa Tengah	C1	0.500	9666

Pada tabel II *cluster* C1 adalah *cluster* dengan nilai terbesar di antara *cluster* lain. Pada *cluster* C1 jawa tengah adalah penyumbang mahasiswa terbanyak dari seluruh suplai tiap provinsi yang terdapat di Universitas Kristen Satya Wacan selama kurun waktu 5 tahun. Dengan nilai 9666. Dari hasil *cluster* C1 tidak memiliki kemiripan dengan *cluster* lain karna perbedaan nilai yang sangat signifikan

TABEL III *CLUSTER* PROVINSI TERBANYAK KEDUA

Provinsi	Cluster	Silhouette	2013-2017
12 NTT	C2	0.664	762
11 Papua	C2	0.677	686
26 Jawa Barat	C2	0.601	555
16 Maluku	C2	0.599	554

Pada tabel III *cluster* C2 adalah *cluster* dengan nilai suplai terbesar kedua setelah jawa tengah. Terdapat 4 provinsi yang memiliki kemipiran nilai antara satu dengan yang lainnya. Antara lain NTT, papua, jawa barat, serta malaku. Dengan nilai suplai antara 554-762.

TABEL IV *CLUSTER* PROVINSI DENGAN NILAI MENENGAH

Provinsi	Cluster	Silhouette	2013-2017
4 Sulawesi Utara	C5	0.690	438
23 Kalimantan Barat	C5	0.711	417
7 Sulawesi Selatan	C5	0.713	391
18 Lampung	C5	0.695	359
1 Sumatera Utara	C5	0.637	327

Pada tabel IV *cluster* C5 adalah *cluster* dengan nilai suplai menengah. Dengan suplai rata-rata di angka 327-438. Pada *cluster* C5 terdapat 5 provinsi antara lain, sulawesi utara, kalimantan barat, sulawesi selatan, lampung, dan sumatera utara.

TABEL V CLUSTER DENGAN NILAI RENDAH

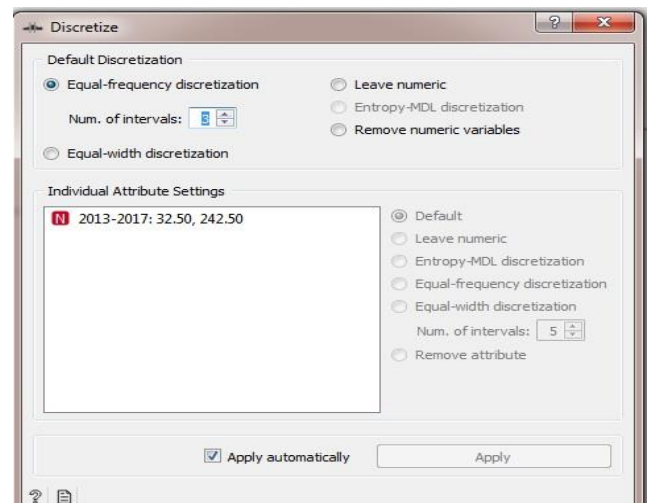
	Provinsi	Cluster	Silhouette	2013 - 2017
24	Jawa Timur	C3	0.638	252
15	Maluku Utara	C3	0.679	233
21	Kalimantan Tengah	C3	0.683	230
32	Banten	C3	0.692	175
29	DKI Jakarta	C3	0.689	171
6	Sulawesi Tengah	C3	0.671	159
20	Kalimantan Timur	C3	0.460	110

Pada tabel V cluster C3 adalah cluster dengan nilai suplai rendah selama kurun waktu 5 tahun. Pada data hasil cluster C3 terdapat 7 provinsi yang masuk dalam kelompok cluster ini antara lain jawa timur, maluku utara, kalimantan tengah, banten, DKI jakarta, sulawesi tengah, dan kalimantan timur.

TABEL VI CLUSTER DENGAN SUPLAI TERKECIL

	Provinsi	Cluster	Silhouette	2013-2017
2	Sumatera Selatan	C4	0.624	86
33	Bali	C4	0.675	70
9	Riau	C4	0.686	65
22	Kalimantan Selatan	C4	0.720	40
5	Sulawesi Tenggara	C4	0.726	33
30	D.I. Yogyakarta	C4	0.727	32
27	Jambi	C4	0.730	25
31	Bengkulu	C4	0.730	23
3	Sumatera Barat	C4	0.731	19
17	Luar Negeri	C4	0.731	14
13	NTB	C4	0.730	12
8	Sulawesi Barat	C4	0.730	12
19	Kepulauan Bangka Belitung	C4	0.730	10
14	NAD	C4	0.729	8
10	Papua Barat	C4	0.728	7
28	Gorontalo	C4	0.723	1

Pada tabel VI cluster C4 adalah cluster dengan jumlah suplai terkecil selama kurun waktu 5 tahun. Dari hasil cluster provinsi-provinsi ini memberikan jumlah suplai paling sedikit dan dibandingkan dengan cluster lainnya pada cluster C4 adalah cluster dengan jumlah terbanyak dengan total 16 provinsi.



Gambar 6. Proses pembagian data menjadi 3 interval

Pada gambar 6 hasil pengolahan pada cluster ini dibagi menjadi 3 interval, antara lain suplai prioritas, potensi suplai, dan suplai terendah. Dengan membagi jumlah data berupa nama tiap provinsi, cluster, serta nilai. Dengan membagi data menjadi 3 interval biro promosi dapat menentukan skala prioritas dari provinsi yang memberikan suplai terbanyak, yang berpotensi memberikan suplai, serta daerah yang memberikan suplai terkecil. Hal ini berguna dalam proses strategi promosi ke provinsi-provinsi tersebut.

E. Provinsi dengan suplai prioritas tahun 2013-2017

TABEL VII PROVINSI PRIORITAS

Provinsi	Cluster	2013-2017
Jawa Tengah	C1	≥ 242
Papua	C2	≥ 242
NTT	C2	≥ 242
Maluku	C2	≥ 242
Jawa Barat	C2	≥ 242
Jawa Timur	C3	≥ 242
Sumatera Utara	C5	≥ 242
Sulawesi Utara	C5	≥ 242
Sulawesi Selatan	C5	≥ 242
Lampung	C5	≥ 242
Kalimantan Barat	C5	≥ 242

Pada tabel VII dari hasil pembagian data menjadi 3 interval, terdapat 11 provinsi dengan suplai prioritas yang memberikan suplai dengan jumlah terbanyak karena memiliki kenaikan signifikan. Dapat dilihat jumlah rata-rata dari tiap daerah tersebut lebih dari atau sama dengan 242.

Cluster yang termasuk dalam provinsi prioritas adalah C1 meliputi provinsi Jawa Tengah, C2 meliputi Papua, NTT, Maluku, dan Jawa Barat, C3 dengan provinsi Jawa timur, dan C5 yang meliputi provinsi Sumatera Utara, Sulawesi Utara, Sulawesi selatan, Lampung, dan Kalimantan barat.

F. Provinsi dengan Potensi Suplai tahun 2013-2017

TABEL VIII PROVINSI POTENSI SUPLAI

Provinsi	Cluster	2013-2017
Sulawesi Tengah	C3	32 - 242
Maluku Utara	C3	32 - 242
Kalimantan Timur	C3	32 - 242
Kalimantan Tengah	C3	32 - 242
DKI Jakarta	C3	32 - 242
Banten	C3	32 - 242
Sumatera Selatan	C4	32 - 242
Sulawesi Tenggara	C4	32 - 242
Riau	C4	32 - 242
Kalimantan Selatan	C4	32 - 242
Bali	C4	32 - 242

Pada tabel VIII menunjukkan provinsi dengan potensi suplai yang terdiri 11 provinsi dengan nilai rata-rata 32-242 yang meliputi cluster C3 dengan cakupan provinsi Sulawesi tengah, Maluku utara, Kalimantan timur, Kalimantan tengah, DKI Jakarta, dan Banten serta C4 dengan cakupan provinsi Sumatera selatan, Sulawesi tenggara, Riau, Kalimantan selatan, dan Bali. Pada data tabel 8 menunjukkan 5 provinsi yang termasuk dalam cluster C4 atau daerah potensi suplai, hal ini karna dari data 5 tahun ada kenaikan signifikan pada

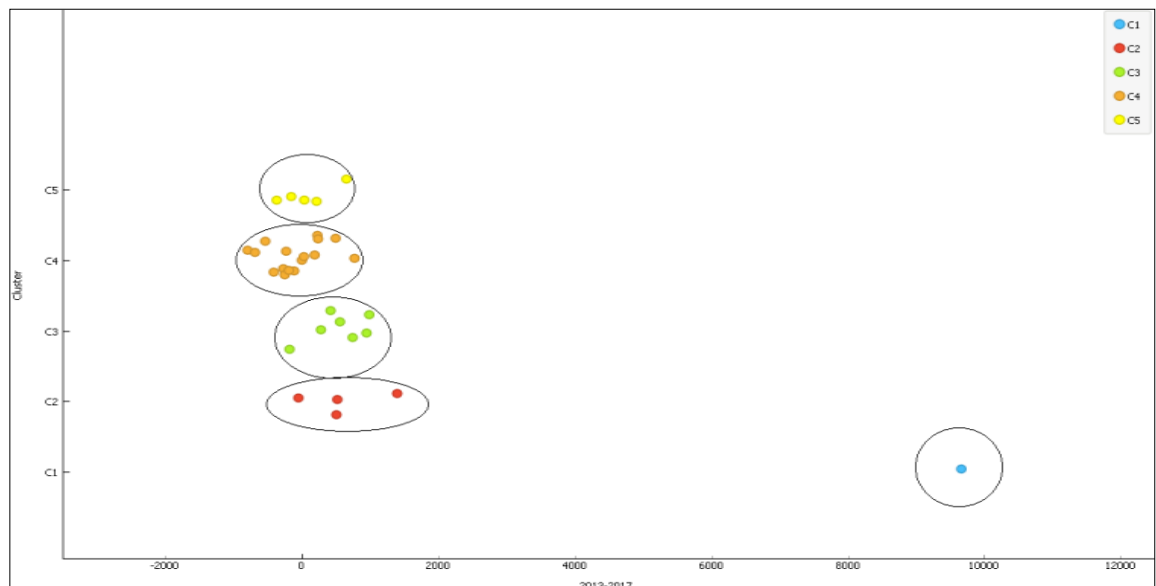
daerah tersebut. Proses *Discretize* membagi data hasil *clustering* menjadi 3 interval karna ada cluster yang suplai provinsinya mendekati daerah potensi suplai tetapi dari hasil *cluster* masuk dalam label C4 atau daerah dengan suplai nilai terendah. Proses *Discretize* ini membantu agar biro promosi tidak melewatkan perhatian pada provinsi tersebut meskipun masuk dalam label *cluster* C4.

D. Provinsi dengan nilai suplai terendah tahun 2013-2017

Tabel IX PROVINSI dengan SUPLAI TERKECIL

Provinsi	Cluster	2013-2017
Sumatera Barat	C4	< 32
Sulawesi Barat	C4	< 32
Papua Barat	C4	< 32
NTB	C4	< 32
NAD	C4	< 32
Luar Negeri	C4	< 32
Kepulauan Bangka Belitung	C4	< 32
Jambi	C4	< 32
Gorontalo	C4	< 32
D.I. Yogyakarta	C4	< 32
Bengkulu	C4	< 32

Pada tabel IX menunjukkan provinsi dengan suplai terkecil, dengan nilai rata-rata di bawah 32. Terdapat 11 provinsi yang meliputi Sumatera barat, Sulawesi barat, Papua barat, NTB, NAD, Luar negeri, Kepulauan bangka belitung, Jambi, Gorontalo, D.I Yogyakarta, dan Bengkulu yang termasuk dalam cluster C4.



Gambar 7. Hasil visualisasi cluster



Pada Gambar 7 adalah hasil visualisasi dari proses penggunaan 5 buah *cluster* terlihat bahwa penyebaran tiap-tiap provinsi membentuk *cluster-cluster* yang saling berdekat berdasarkan dengan kemiripan nilai yang dimiliki. Pada proses tersebut telah melewati iterasi dari 300-3000 iterasi, sehingga hasil yang didapatkan tidak lagi mengalami pergeseran posisi dan jarak dari masing-masing *cluster*.

Kelebihan dari proses pengolahan ini dibandingkan dengan penelitian sebelumnya adalah pada penilitan sebelumnya menggunakan data kurang dari kurang waktu 5 tahun yang pada dasarnya tiap tahunnya terjadi perbedaan suplai pada tiap provinsi ataupun kota dan juga hanya pada beberapa wilayah yang ada di Indonesia. Kedua dari penilitan sebelumnya belum menjangkau hampir setiap provinsi yang mungkin terdapat populasi suplai yang signifikan tiap tahunnya. Penelitian yang dilakukan pada biro promosi UKSW mendapatkan hasil yang memiliki pola menarik dan pengetahuan baru. Penelitian ini juga memakai pembagian interval, sehingga provinsi yang terdapat pada tiap *cluster* yang dibentuk menjadi suplai prioritas, potensi prioritas, dan suplai terkecil mendapat porsi perhatian yang sama, sehingga pengetahuan akan pola tiap provinsi lebih akurat. Pada penilitan ini pun hampir menjangkau semua provinsi yang ada di Indonesia, dan telah ditentukan skala prioritas berdasarkan *cluster* maupun interval data, yang dibagi pada *cluster-cluster* tersebut.

Dari hasil pengolahan pada data-data selama kurun waktu 5 tahun dari periode 2013-2017 terdapat 2 hal yang menjadi acuan dan fokus. Pertama adalah data hasil provinsi prioritas dan data dari provinsi potensi suplai. Dari hasil pengolahan dan penggunaan metode *Clustering* membantu dalam mempartisi atau membagi tiap provinsi ke dalam kelompok-kelompok dengan nilai kemiripan yang sama. Selanjutnya membagi data menjadi 3 interval sehingga dari data hasil *cluster* dapat ditentukan skala prioritas, sehingga hal ini berguna dalam memfokuskan penjangkauan berdasarkan prioritas dan potensi suplai di tiap provinsi.

Pembagian fokus penjangkauan menjadi hal yang penting karena Biro Promosi dapat merencanakan jumlah tim promosi yang akan turun di tiap-tiap provinsi berdasarkan jumlah suplai dari yang terbesar sampai dengan terkecil, selanjutnya, biro promosi dapat memperhitungkan simulasi biaya serta pengeluaran yang dilakukan sebelum proses promosi dan penjangkauan, yang terakhir biro promosi dapat menentukan dan menskenariokan lama waktu yang diperlukan dalam proses penjangkauan di tiap-tiap provinsi yang dikunjungi.

#### E. Strategi promosi

Dari hasil pengolahan data serta pengelompokan berdasarkan *cluster* pada tiap provinsi. Ada beberapa strategi yang dapat dilakukan oleh biro promosi universitas Kristen Satya Wacana antara lain: perhitungan jumlah tim yang akan turun langsung untuk promosi ke provinsi tersebut. Provinsi prioritas dan provinsi dengan potensi suplai memiliki

kecenderungan pertumbuhan suplai mahasiswa yang cukup signifikan tercatat ada 21 provinsi yang menjadi fokus. Dari 21 provinsi tersebut hanya 3 provinsi yang berada di pulau Jawa. Hal ini berpengaruh kepada jumlah tim yang lebih banyak untuk turun secara langsung. Karena dari hasil pengolahan *data mining*, provinsi-provinsi tersebut yang memberikan suplai paling banyak serta signifikan. Provinsi dengan nilai suplai terendah lebih diutamakan dengan cara promosi menggunakan media elektronik berupa *social media*, *video* promosi, *pamflet*, ataupun baliho pada provinsi-provinsi tersebut. Dengan cara demikian dapat meminimalisir pengeluaran untuk pengiriman tim ke provinsi-provinsi tersebut, yang memberikan nilai suplai terkecil. Tercatat ada 11 provinsi dan hanya 1 provinsi yang berada di pulau Jawa, sehingga alternatif ini dapat dipakai untuk menjangkau tiap provinsi tersebut.

#### V. KESIMPULAN

Dari proses pengolahan serta hasil analisis provinsi dengan potensi suplai, tiap tahunnya, terjadi kenaikan dan penurunan suplai dari provinsi yang memiliki potensi di Indonesia. Proses *cluster* membantu dalam memecah dan mempartisi beberapa provinsi yang memiliki potensi suplai dari data 5 tahun dalam kurun waktu 2013-2017. Kedua, hasil dari proses *cluster* membantu dalam mengelompokkan provinsi-provinsi tersebut berdasarkan jumlah kemiripan data menjadi satu label sehingga biro promosi dapat menentukan fokus penjangkauan dari provinsi dengan potensi suplai terlebih dahulu dan selanjutnya ke provinsi dengan suplai prioritas. Ketiga, dari hasil pengolahan dan analisis yang dilakukan provinsi dengan suplai prioritas tidak banyak mengalami perubahan yang signifikan dari kurun waktu 2013-2017. Keempat, dengan proses *cluster K-means* pembagian fokus pada proses penjangkauan mahasiswa di tiap daerah lebih tepat sasaran. Sehingga mengurangi penjangkauan ke segala provinsi yang tidak memberikan suplai jumlah mahasiswa yang signifikan. Hal ini berimbas pada pembekakan pengeluaran besar, serta tidak tercapainya kuota yang telah ditargetkan. Terakhir, dengan melakukan proses data mining menggunakan metode *clustering* biro promosi menerapkan strategi yang efisien serta dapat meminimalkan pengeluaran untuk turun secara langsung ke tiap-tiap provinsi.

#### DAFTAR PUSTAKA

- [1] Asril, Wiza, F., & Yunefri, "Analisis Data Lulusan dengan *Data Mining*," Jurnal Teknologi Informasi & Komunikasi *Digital Zone*, Vol. 6, Nomor 2, pp.24-32, 2015.
- [2] Nugroho, "Analisa Potensi Mahasiswa Di Daerah Bali Menggunakan Pendekatan *K-mean clustering*" Jurnal Sistem Dan Informatika, vol. 11, pp.45-50, 2016.
- [3] Ong, J. O. "Implementasi *Algoritma K-means Clustering* untuk menentukan strategi marketing *President University*," Jurnal Ilmiah Teknik Industri, vol.12, pp.10-20. 2013.
- [4] Prasetyo, Saharuna, Abduh, "Analisis Data Atribut Mahasiswa untuk Menentukan Strategi Promosi Kampus Menggunakan Metode *Data*

- Mining," Seminar Nasional Teknik Elektro dan Informatika (SNTEI) Makassar, Nov. 2017.
- [5] Setiawan, "Penerapan *Data Mining* Menggunakan *Algoritma K-Means Clustering* Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik LP3I Jakarta)," *Jurnal Lentera ICT*, vol.3, No.1, 2016.
- [6] Asroni, Adrian, "Penerapan Metode *K-Means* Untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan *Weka Interface* Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *Jurnal Ilmiah Semesta Teknik* vol. 18, No. 1, pp.76-82, 2015.
- [7] Ramadhani, Dias, "*Data Mining* Menggunakan *Algoritma K-Means Clustering* Untuk Menentukan Strategi Promosi Universitas Dian Nuswantoro Universitas Dian Nuswantoro," 2014.
- [8] Eska, "Penerapan *Data Mining* untuk prediksi penjualan wallpaper menggunakan *Algoritma C4.5*" *JURTEKSI*, vol. 2, No. 2, 2016.
- [9] Merliana, Ernawati, Santoso, "Analisa Penentuan Jumlah Cluster Terbaik Pada Metode *K-Means Clustering*." *Prosiding seminar nasional multi disiplin ilmu & call for papers unisbank*, 2016.
- [10] Ediyanto, Mara, Satyahadewi, "Pengklasifikasian Karakteristik Dengan Metode *K-Means Cluster Analysis*." *Buletin Ilmiah Mat. Stat. dan Terapannya* , vol. 02 , No. 2, pp. 133 – 136, 2013.
- [11] Robani, Widodo, "*Algoritma K-Means Clustering* Untuk Pengelompokan Ayat *Al Quran* Pada Terjemahan Bahasa Indonesia," *JSINBIS*, vol. 6, no. 2, pp. 164-176, 2016.
- [12] Wu, X. and Kumar, "*The Top Ten Algorithms in Data Mining*" *London: CRC Press Taylor & Francis Group*, 2009.