



Implementasi Metode K-Means Clustering untuk Meningkatkan Penjaringan Mahasiswa

Implementation of the K-Means Clustering Method to Improve Student Recruitment

Muhammad Zulfadhilah¹, Mambang², Septyan Eka Prastya³

^{1,2,3}Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Sari Mulia

¹zulfadhilah@unism.ac.id, ²mambang@unism.ac.id, ³septyan.e.prastya@unism.ac.id

Abstract

Recruitment of students in private universities has its challenges for the campus. Several things must be seen by the campus in its efforts to recruit new students every year, namely the existence of student data that can be used in developing a strategy for recruiting new students. This study aims to assist the campus in determining strategic policies to conduct the screening of new students every year. This study uses data from the last 3 years from one of the study programs, the student data used are city origin, school origin, and province origin. Grouping students based on school origin, city, and the province is one way to attract new students. the use of data mining in the grouping can be done using the K-Means Clustering algorithm. The implementation of the K-Means Clustering Algorithm is relatively easy and has fast computing, so it can be used to analyze college student profile data to help increase the selection of new students in the following year. The applications used to run the K-Means Clustering Algorithm are Orange Data Mining and RapidMiner, these 2 applications are used to ensure the results obtained regarding data grouping or clustering are the same. In the data that has been grouped with K-Means Clustering, there are 3 clusters, namely cluster_2 / C2 is the cluster with the most cluster members, in this cluster, South Kalimantan Province with Banjarmasin City has the most members so it becomes a separate cluster. Data cluster_0 / C3 is the cluster with the second highest value, this cluster is obtained in Central Kalimantan Province with various cities in it. The data cluster_1 / C1 has the lowest value of 10, in this cluster it is obtained in the Province of South Kalimantan, but with different cities. The grouping of the data will make it easier to formulate strategies for recruiting new students in the following year.

Keywords: clustering, data mining, k-means,

Abstrak

Penjaringan mahasiswa di perguruan tinggi swasta memiliki tantangan tersendiri bagi pihak kampus. Beberapa hal yang harus bisa dilihat oleh pihak kampus dalam upaya penjaringan mahasiswa baru disetiap tahunnya, yaitu adanya data mahasiswa yang dapat digunakan dalam menyusun strategi penjaringan mahasiswa baru. Penelitian ini bertujuan untuk membantu pihak kampus dalam menentukan kebijakan strategi untuk melakukan penjaringan mahasiswa baru disetiap tahunnya. Penelitian ini menggunakan data 3 tahun terakhir dari salah satu program studi, data mahasiswa yang digunakan adalah asal kota, asal sekolah, dan asal provinsi. Pengelompokan mahasiswa berdasarkan asal sekolah, kota dan provinsi merupakan salah satu cara untuk menjanging mahasiswa baru. penggunaan data mining dalam pengelompokan dapat dilakukan dengan menggunakan algoritma K-Means Clustering. Implementasi Algoritma K-Means Clustering relatif mudah dan memiliki komputasi yang cepat, sehingga dapat digunakan untuk menganalisis data profil mahasiswa perguruan tinggi untuk membantu meningkatkan penjaringan mahasiswa baru pada tahun berikutnya. Aplikasi yang digunakan untuk menjalankan Algoritma K-Means Clustering adalah Orange Data Mining dan RapidMiner, 2 aplikasi ini digunakan untuk memastikan hasil yang didapat terkait pengelompokan atau klasterisasi data adalah sama. Pada data yang telah dikelompokkan dengan K-Means Clustering terdapat 3 cluster, yaitu cluster_2 / C2 merupakan klaster dengan anggota klaster terbanyak, pada klaster ini Provinsi Kalimantan Selatan dengan Kota Banjarmasin yang memiliki anggota terbanyak, sehingga menjadi klaster tersendiri. Data cluster_0 / C3 merupakan klaster dengan nilai kedua tertinggi, klaster ini didapatkan pada Provinsi Kalimantan Tengah dengan berbagai Kota di dalamnya. Data cluster_1 / C1 memiliki nilai terendah yaitu 10, pada klaster ini didapatkan pada Provinsi Kalimantan Selatan, tetapi dengan kota yang berbeda-beda. Pengelompokan data tersebut akan mempermudah penyusunan strategi dalam penjaringan mahasiswa baru di tahun berikutnya.

Kata kunci : : klastering, data mining, k-means

1. Pendahuluan

Penerapan teknologi informasi diberbagai bidang sudah menjadi sebuah keharusan saat ini, tidak terkecuali dalam dunia pendidikan yang menghasilkan data yang berlimpah terkait pembelajaran dan data mahasiswa [1]. Penerimaan mahasiswa baru yang dilakukan setiap tahun oleh setiap perguruan tinggi menghasilkan data yang berlimpah, terutama pada data profil mahasiswa. Hal tersebut dapat ditindaklanjuti dengan melakukan pengolahan data yang akan sangat membantu untuk mengetahui informasi penting berupa pengetahuan baru. Informasi yang telah diolah tadi dapat digunakan sebagai penunjang dalam pengambilan keputusan bagi institusi [2] dalam hal menentukan strategi pencarian mahasiswa baru nantinya. Pengelompokan data yang dilakukan berdasarkan kesamaan atau similaritas akan mempermudah untuk dilakukannya analisis[3]. Data yang didapat dan disimpan mempunyai keragaman, dan perubahan datanya cepat, maka dapat disebut dengan big data [4]. Big Data meliputi tiga hal yang biasa disingkat dengan 3V, yaitu *volume*, *velocity*, dan *variety*. *Volume* adalah besaran data yang harus dikelola berukuran super besar. Sedangkan *Velocity* berkenaan dengan kecepatan pemrosesan data yang harus mengimbangi pesatnya pertumbuhan jumlah data yang ada. Kemudian *variety* merujuk pada karakteristik sumber data yang sangat beragam, baik itu yang berasal dari basis data yang terstruktur maupun juga dari data-data yang tidak terstruktur [5].

Big data tersebut jika diolah menjadi informasi [4] yang tepat maka dapat digunakan untuk merencanakan strategi yang tepat dalam penjangkaran mahasiswa. Salah satu cara pengolahan data adalah dengan menggunakan data mining, dengan menggunakan data mining para pengambil keputusan dapat memanfaatkannya untuk menentukan cara yang tepat dan cepat [6]. Pengelompokan data mahasiswa dengan data mining dapat menggunakan teknik Clustering yang merupakan proses membagi data dalam suatu himpunan ke dalam beberapa kelompok yang memiliki kesamaan data [7]. Teknik Clustering ini juga dapat meminimalisir suatu objek data yang mirip dan mengklasterkan data tersebut secara ilmiah [8]. Teknik *clustering* saat ini banyak digunakan dan dipelajari secara ekstensif dalam klasifikasi unsupervised [9].

K-Means Clustering merupakan salah satu algoritma yang mampu melakukan klasterisasi terhadap data yang bersifat heterogen, hal ini dikarenakan algoritma pada dasarnya hanya mampu melakukan pengelompokan dengan nilai atribut homogen saja [10]. *K-Means Clustering* juga disebut dengan algoritma yang menentukan nilai *Cluster* (*k*) dengan sembarang dan nilai yang telah ditentukan tersebut menjadi titik pusat dari *Cluster* atau sebagai *centroid* [11]. *K-Means* merupakan metode data untuk klasterisasi *non-hierarchical* yang dapat mengelompokkan data menjadi

beberapa *cluster* berdasarkan kesamaan data, sehingga data yang memiliki karakteristik yang sama dikelompokkan menjadi satu cluster dan yang memiliki karakteristik berbeda dikelompokkan dalam cluster yang lain [12]. [13] menyebutkan bahwa hasil uji klaster *K-Means* merupakan solusi yang optimal. *K-Means Clustering* memiliki kelebihan sebagai algoritma pengelompokan yang mudah diimplementasikan, relatif cepat dalam proses komputasi [14]. Namun [15] menuliskan bahwa *K-Means Clustering* memiliki beberapa kelemahan pada penentuan nilai *centroid*-nya sehingga bisa mengakibatkan klasterisasi tidak maksimal dilakukan.

Pada penelitian yang dilakukan oleh [2] menyebutkan bahwa algoritma *K-Means Clustering* dapat mengklasterkan data yang nantinya digunakan untuk melakukan strategi promosi berdasarkan program studi yang paling banyak diminati. Hal serupa juga disampaikan pada penelitian yang dilakukan oleh [8] yang menyebutkan bahwa penerapan data mining dengan metode *K-means Clustering* dapat diterapkan pada pengelompokan buku sehingga membantu pihak Perpustakaan untuk mengetahui buku mana yang sering dipinjam. Hal ini sesuai dengan apa yang dituliskan oleh [16] yang menyatakan bahwa dengan adanya pengolahan data berdasarkan kelompok tersebut, maka penentuan tempat promosi perguruan tinggi akan tepat sasaran dan meminimalisir penurunan jumlah mahasiswa pada tahun berikutnya.

Penelitian yang dilakukan oleh [17] menyatakan bahwa hasil *K-Means Clustering* dapat digunakan untuk menyimpulkan dalam menentukan siswa yang berpotensi DO. Pada penelitian yang dilakukan oleh [18] menyebutkan bahwa dari perbandingan metode algoritma klasterisasi, metode *K-Means* menghasilkan pengelompokan terbaik. Pada penelitian yang dilakukan oleh [10] menyebutkan bahwa semakin banyak data yang diproses, maka hasil yang didapatkan juga akan semakin baik. Hal ini serupa dengan penelitian yang dilakukan oleh [19] menyebutkan bahwa hasil klasterisasi dari metode *K-Means Clustering* memiliki akurasi sebesar 100% dengan menggunakan 1000 dataset.

Berdasarkan uraian di atas, maka peneliti akan melakukan pengelompokan data mahasiswa berdasarkan asal sekolah, kota dan provinsi dengan menggunakan algoritma *K-Means Clustering*. Hal ini bertujuan untuk dapat menentukan strategi penjangkaran mahasiswa baru pada tahun-tahun berikutnya dan pengambilan keputusan dalam beberapa strategi penjangkaran mahasiswa.

2. Metode Penelitian

Penelitian ini memiliki beberapa tahapan yang dimulai dari pengumpulan data, pengolahan awal, dan proses klasterisasi yang akan dianalisis dan visualisasikan.

Tahap satu dimulai dengan pengumpulan data yang didapat dari program studi disalah satu perguruan tinggi swasta, data yang diambil adalah 3 tahun terakhir. Tahapan berikutnya adalah melakukan seleksi data, pada tahapan ini data mahasiswa yang didapat akan diseleksi berdasarkan asal sekolah, kota asal dan provinsi asal mahasiswa. Pada tahapan terakhir yaitu implementasi menggunakan metode *K-Means Clustering* dengan menggunakan aplikasi *Orange Data Mining* dan *RapidMiner*. Hasil dari klasterisasi 2 buah aplikasi tersebut akan dianalisis dan visualisasikan untuk mempermudah pembaca memahami hasilnya.

Pada metode *K-Means Clustering* ada langkah-langkah yang dilakukan, berikut adalah langkah-langkah yang dilakukan [20] yaitu pada langkah pertama menentukan jumlah klaster yang akan dipergunakan dalam pembagian data; kemudian langkah kedua dilanjutkan dengan menentukan centroid awal yang diperoleh secara acak serta jumlah centroid sebanyak klaster yang akan dibuat. Pengertian centroid sendiri adalah titik pusat klaster atau awal pusat klaster; setelah itu langkah ketiga dilakukan perhitungan jarak pada setiap inputan data terhadap pusat klaster hingga ditemukan jarak paling dekat dari setiap data terhadap centroid. Perhitungan jarak dilakukan dengan menggunakan persamaan *Euclidean Distance*.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i + t_i)^2} \quad (1)$$

dengan D_e adalah *Euclidean Distance*, i adalah banyaknya objek, (x,y) merupakan koordinat objek dan (s,t) merupakan koordinat *centroid*.

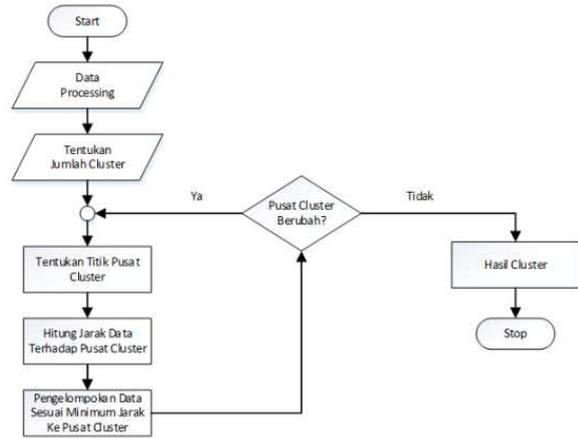
Langkah keempat adalah mengelompokkan setiap data terhadap jarak pada titik pusat centroid terdekat; kemudian langkah kelima adalah mengubah nilai centroid yang diperoleh dari rata-rata klaster yang bersangkutan dengan persamaan:

$$C_k = \frac{1}{n_k} \sum d_i \quad (2)$$

dengan n_k adalah jumlah data dalam klaster dan d_i adalah jumlah dari nilai jarak yang masuk dalam masing-masing klaster.

Jika anggota tiap klaster tidak ada yang berubah, maka iterasi selesai dan nilai rata-rata pusat klaster (μ_j) akan digunakan sebagai parameter dalam penentuan pembagian data; jika tidak, maka iterasi akan dilakukan dengan menggunakan langkah kedua hingga langkah kelima.

Gambar 1 merupakan alur penggunaan metode *K-Means Clustering* yang digunakan dalam penelitian ini.



Gambar 1. Algoritma K-Means Clustering [20]

3. Hasil dan Pembahasan

Data yang diolah pada penelitian ini adalah data mahasiswa pada salah satu perguruan tinggi swasta dengan mengambil satu program studi yang akan dianalisa menggunakan *K-Means Clustering*. Data yang diambil adalah data dari 3 angkatan mahasiswa program studi dengan jumlah 66 orang mahasiswa aktif. Aplikasi yang digunakan dalam pengolahan data dan analisa adalah Aplikasi *RapidMiner* dan *Orange Data Mining*, 2 aplikasi tersebut merupakan aplikasi yang sering digunakan dalam pengolahan dan analisa data mining.

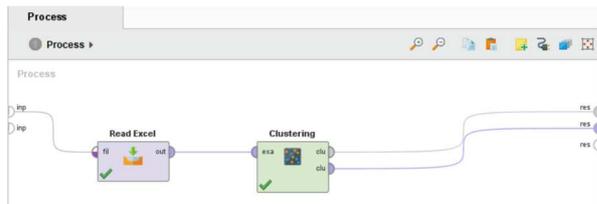
Data mahasiswa akan dikelompokkan menjadi 3 ($K=3$), hal ini bertujuan untuk mengetahui kelompok mahasiswa berdasarkan dari data asal sekolah, kota dan provinsi mahasiswa. Berikut data mahasiswa yang akan dilakukan pengolahan dan analisa dengan menggunakan *K-Means Clustering*.

Tabel 1. Tabel Data Mahasiswa

No	Asal Sekolah	Provinsi	Kota
1	SMK Negeri Banjarmasin	3 Kalimantan Selatan	Banjarmasin
2	SMK Negeri Banjarmasin	1 Kalimantan Selatan	Banjarmasin
3	SMA Negeri Beruntung Baru	1 Kalimantan Selatan	Banjarmasin
4	SMK Negeri Banjarmasin	3 Kalimantan Selatan	Banjarmasin
5	SMA Negeri Muara Teweh	1 Kalimantan Tengah	Muara Teweh
6	SMK Negeri Palangkaraya	2 Kalimantan Tengah	Palangkaraya
7	SMK Negeri Puruk Cahu	1 Kalimantan Tengah	Puruk Cahu
8	SMK Negeri Mihing Raya	1 Kalimantan Tengah	Palangkaraya
9	SMA Negeri Dusun Selatan	2 Kalimantan Tengah	Buntok
....
44	SMK Negeri Banjarmasin	4 Kalimantan Selatan	Banjarmasin
45	Tunas Bangsa	Kalimantan Selatan	Banjarmasin

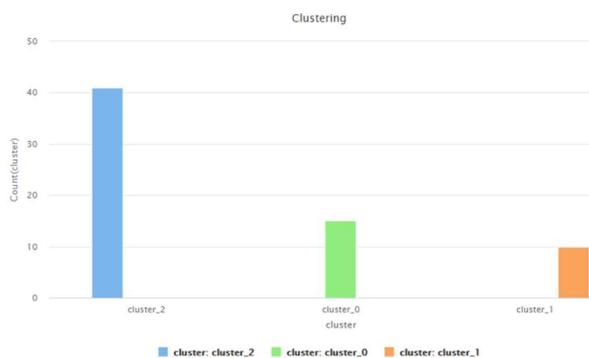
No	Asal Sekolah	Provinsi	Kota
46	SMK Islam Sabibal Muhtadin	Kalimantan Selatan	Banjarmasin
47	SMK ISFI Banjarmasin	Kalimantan selatan	Banjarmasin
48	SMK Negeri 1 Banjarmasin	Kalimantan Selatan	Banjarmasin
...

Tab 1 Data Mahasiswa tersebut kemudian diolah dan dianalisa dengan menggunakan *RapidMiner* dan *Orange Data Mining*. Pada gambar 2 berikut merupakan desain proses yang dilakukan pada aplikasi *RapidMiner*:



Gambar 2. Desain Pemrosesan Pada *RapidMiner*

Setelah proses desain tersebut, maka aplikasi dijalankan dan mendapatkan hasil yang ada pada gambar 3 yang merupakan pengelompokan data berdasarkan provinsi asal mahasiswa sebagai berikut:



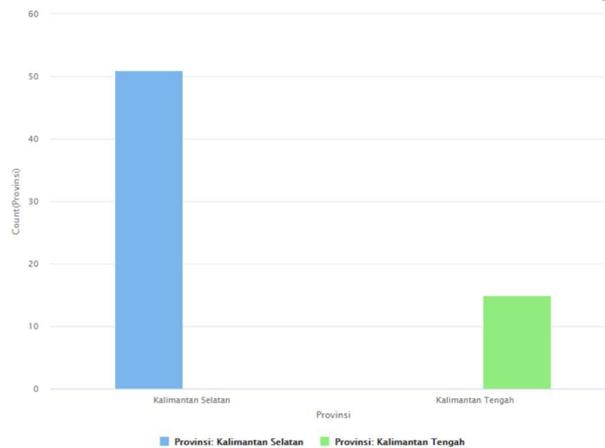
Gambar 3. Hasil dari *K-Means Clustering RapidMiner*

Dari gambar 3 terlihat hasil dari klasterisasi aplikasi *RapidMiner* terdapat 3 klaster, dengan masing-masing data setiap klaster dapat dilihat pada tabel 2 berikut.

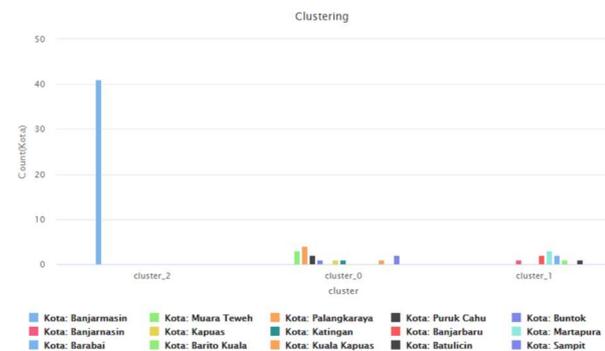
Tabel 2. Model Cluster

Nilai	Total	Persentasi
cluster_0	15 items	22.73 %
cluster_1	10 items	15.15 %
cluster_2	41 items	62.12 %
Total	66	100 %

Pada data yang telah dilakukan klasterisasi juga dapat dilihat hasil berdasarkan asal provinsi mahasiswa, hasil tersebut bisa dilihat pada gambar 4 berikut.



Gambar 4 Data berdasarkan asal provinsi mahasiswa



Gambar 5. Hasil dari *K-Means Clustering* dengan Keterangan Kota

Pada gambar 5 di atas merupakan hasil pengelompokan data berdasarkan Kota asal mahasiswa. Tabel 3 di bawah ini adalah hasil dari klasterisasi.

Tabel 3 Hasil dari *K-Means Clustering*

id	cluster	Asal Sekolah	Provinsi	Kota
1	cluster_2	SMK Negeri 3 Banjarmasin	Kalimantan Selatan	Banjarmasin
2	cluster_2	SMK Negeri 1 Banjarmasin	Kalimantan Selatan	Banjarmasin
3	cluster_2	SMA Negeri 1 Beruntung Baru	Kalimantan Selatan	Banjarmasin
4	cluster_2	SMK Negeri 3 Banjarmasin	Kalimantan Selatan	Banjarmasin
5	cluster_0	SMA Negeri 1 Muara Teweh	Kalimantan Tengah	Muara Teweh
6	cluster_0	SMK Negeri 2 Palangkaraya	Kalimantan Tengah	Palangkaraya
7	cluster_0	SMK Negeri 1 Puruk Cahu	Kalimantan Tengah	Puruk Cahu
8	cluster_0	SMK Negeri 1 Mihing Raya	Kalimantan Tengah	Palangkaraya

id	cluster	Asal Sekolah	Provinsi	Kota	id	cluster	Asal Sekolah	Provinsi	Kota
9	cluster_0	SMA Negeri 2 Dusun Selatan	Kalimantan Tengah	Buntok	32	cluster_1	SMA Negeri 8 Barabai	Kalimantan Selatan	Barabai
10	cluster_2	SMK Negeri 1 Gambut	Kalimantan Selatan	Banjarmasin	33	cluster_2	SMK Maestro Islamic School	Kalimantan Selatan	Banjarmasin
11	cluster_1	SMK Maestro Islamic School Banjarmasin	Kalimantan Selatan	Banjarmasin	34	cluster_2	SMA Negeri 11 BANJARMASIN	Kalimantan Selatan	Banjarmasin
12	cluster_0	SMA Negeri 4 Muara Teweh	Kalimantan Tengah	Muara Teweh	35	cluster_2	SMK Negeri 2 Banjarmasin	Kalimantan Selatan	Banjarmasin
13	cluster_0	SMA Negeri 3 Palangkaraya	Kalimantan Tengah	Palangkaraya	36	cluster_1	MA Raudhatussuyuban	Kalimantan Selatan	Martapura
14	cluster_0	SMK Negeri 5 Kuala Kapuas	Kalimantan Tengah	Kapuas	37	cluster_1	SMK Negeri 1 Banjarbaru	Kalimantan Selatan	Banjarbaru
15	cluster_2	SMA PGRI 1 Banjarmasin	Kalimantan Selatan	Banjarmasin	38	cluster_2	SMA Negeri 11 Banjarmasin	Kalimantan Selatan	Banjarmasin
16	cluster_2	SMA PGRI 1 Banjarmasin	Kalimantan Selatan	Banjarmasin	39	cluster_1	SMA Negeri 1 Alalak	Kalimantan Selatan	Barito Kuala
17	cluster_0	SMA N 1 Puruk Cahu	Kalimantan Tengah	Puruk Cahu	40	cluster_0	SMA Negeri 2 Palangkaraya	Kalimantan Tengah	Palangkaraya
18	cluster_2	SMK Negeri 3 Banjarmasin	Kalimantan Selatan	Banjarmasin	41	cluster_0	SMA Negeri 1 Kapuas Timur	Kalimantan Tengah	Kuala Kapuas
19	cluster_2	SMA Negeri 4 Banjar baru	Kalimantan Selatan	Banjarmasin	42	cluster_2	SMA Negeri 11 Banjarmasin	Kalimantan Selatan	Banjarmasin
20	cluster_0	SMK Negeri 1 Sanaman Mantikei	Kalimantan Tengah	Katingan	43	cluster_2	SMA Negeri 4 Banjarmasin	Kalimantan Selatan	Banjarmasin
21	cluster_2	SMK Negeri 2 Banjarmasin	Kalimantan Selatan	Banjarmasin	44	cluster_2	SMK Negeri 4 Banjarmasin	Kalimantan Selatan	Banjarmasin
22	cluster_2	SMK Negeri 3 Banjarmasin	Kalimantan Selatan	Banjarmasin	45	cluster_2	Tunas Bangsa	Kalimantan Selatan	Banjarmasin
23	cluster_2	SMK ISFI Farmasi Banjarmasin	Kalimantan Selatan	Banjarmasin	46	cluster_2	SMK Islam Sabilal Muhtadin	Kalimantan Selatan	Banjarmasin
24	cluster_2	MAN 1 Banjarmasin	Kalimantan Selatan	Banjarmasin	47	cluster_2	SMK ISFI banjarmasin	Kalimantan Selatan	Banjarmasin
25	cluster_2	SMA Negeri 11 Banjarmasin	Kalimantan Selatan	Banjarmasin	48	cluster_2	SMK Negeri 1 Banjarmasin	Kalimantan Selatan	Banjarmasin
26	cluster_2	SMA Negeri 11 Banjarmasin	Kalimantan Selatan	Banjarmasin	49	cluster_2	MAN 1 Banjarmasin	Kalimantan Selatan	Banjarmasin
27	cluster_1	SMA Plus Citra Madinatul Ilmi	Kalimantan Selatan	Banjarbaru	50	cluster_1	Madrasah Aliyah Pertasi Kencana NU Haruyan	Kalimantan Selatan	Barabai
28	cluster_2	MAN 3 Banjarmasin	Kalimantan Selatan	Banjarmasin	51	cluster_2	SMA Negeri 8 Banjarmasin	Kalimantan Selatan	Banjarmasin
29	cluster_1	MA Raudhatussuyuban	Kalimantan Selatan	Martapura	52	cluster_1	SMA Negeri 1 Martapura	Kalimantan Selatan	Martapura
30	cluster_2	SMA Negeri 5 Banjarmasin	Kalimantan Selatan	Banjarmasin	53	cluster_2	SMK Negeri 3 Banjarmasin	Kalimantan Selatan	Banjarmasin
31	cluster_2	SMA Negeri 11 Banjarmasin	Kalimantan Selatan	Banjarmasin	54	cluster_2	SMK Negeri 3 Banjarmasin	Kalimantan Selatan	Banjarmasin

id	cluster	Asal Sekolah	Provinsi	Kota
55	cluster_1	SMA Negeri 1 Karang Bintang	Kalimantan Selatan	Batulicin
56	cluster_2	MAN 2 Model Banjarmasin	Kalimantan Selatan	Banjarmasin
57	cluster_2	MAN 1 Banjarmasin	Kalimantan Selatan	Banjarmasin
58	cluster_2	SMA Negeri 9 Banjarmasin	Kalimantan Selatan	Banjarmasin
59	cluster_2	SMA Negeri 11 Banjarmasin	Kalimantan Selatan	Banjarmasin
60	cluster_0	SMA N 2 Sampit	Kalimantan Tengah	Sampit
61	cluster_2	Sma Negeri 6 Banjarmasin	Kalimantan Selatan	Banjarmasin
62	cluster_2	MAN 1 Banjarmasin	Kalimantan Selatan	Banjarmasin
63	cluster_2	SMAN 1 Beruntung Baru	Kalimantan Selatan	Banjarmasin
64	cluster_2	SMK Negeri 4 Banjarmasin	Kalimantan Selatan	Banjarmasin
65	cluster_0	SMA Negeri 1 Paranggean	Kalimantan Tengah	Sampit
66	cluster_0	SMK Negeri 1 Bukit Sawit	Kalimantan Tengah	Muara Teweh

Berikut tabel 4 merupakan *centroid* pada pemrosesan *K-Means Clustering*:

Tabel 4. *Centorid*

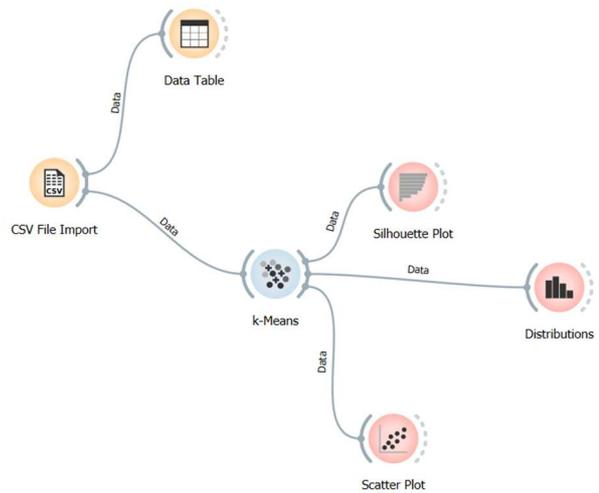
Atribut	cluster_0	cluster_1	cluster_2
Asal Sekolah	20.333	29.300	21.902
Provinsi	2.0	1.0	1.0
Kota	5.933	10.2	1.0

Pada *RapidMiner* juga menghasilkan *Performance Vector* dari metode *K-Means Clustering*, yaitu pada tabel 5 sebagai berikut:

Tabel 5. *Performance Vector*

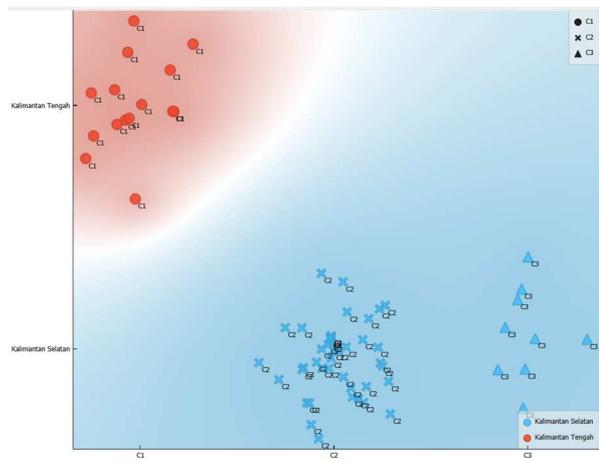
Performance Vector	
Number of clusters	3.000 Cluster
Cluster Number Index	0.955

Selain menggunakan aplikasi *RapidMiner*, peneliti juga melakukan pengolahan dan analisis data menggunakan aplikasi *Orange Data Mining*. Pada gambar 6 berikut adalah proses klusterisasi dengan menggunakan aplikasi *Orange Data Mining*.



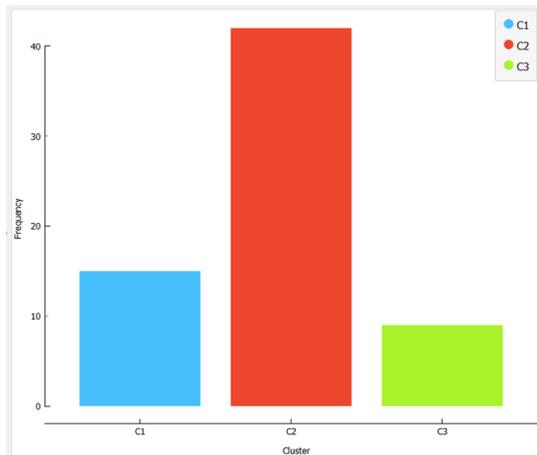
Gambar 6. Desain Pemrosesan pada *Orange Data Mining*

Pada gambar 6 di atas terlihat ada beberapa percabangan dengan keterangan, yaitu *CSV File Import* merupakan data mahasiswa program studi yang akan diproses dalam bentuk format *.csv*; *Data Table* digunakan untuk melihat data yang sudah dimasukkan dari *file .csv* tadi; *K-Means* merupakan metode yang digunakan dan memiliki nilai $K=3$, serta pemrosesan klusterisasi berlangsung pada bagian ini; dan *Scatter Plot*, *Distributions* dan *Silhouette Plot* digunakan untuk melihat hasil visualisasi *K-Means Clustering* yang telah diproses sebelumnya.



Gambar 7. Hasil *K-Means Clustering* menggunakan Scatter Plot

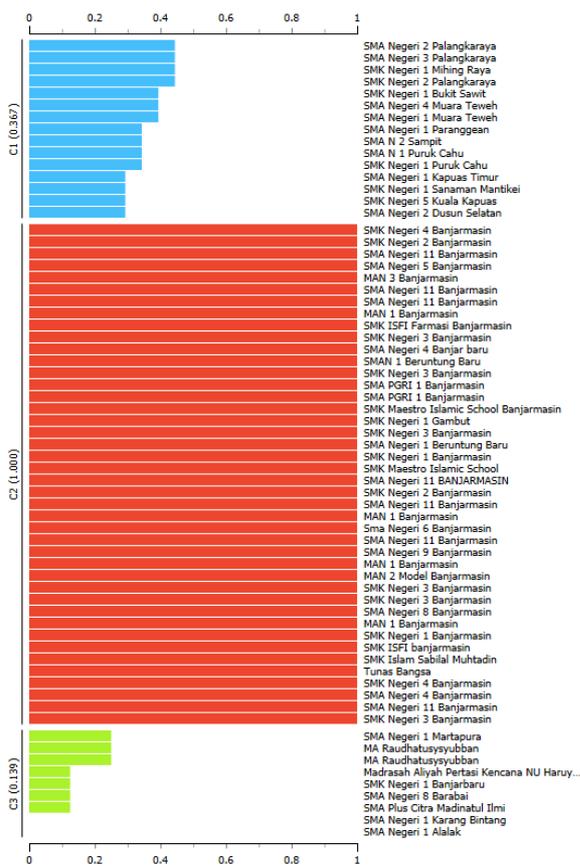
Pada gambar 7 di atas merupakan hasil visualisasi dari klusterisasi menggunakan metode *K-Means Clustering* berdasarkan dari provinsi asal mahasiswa.



Gambar 8. Hasil K-Means Clustering menggunakan visualisasi Distributions

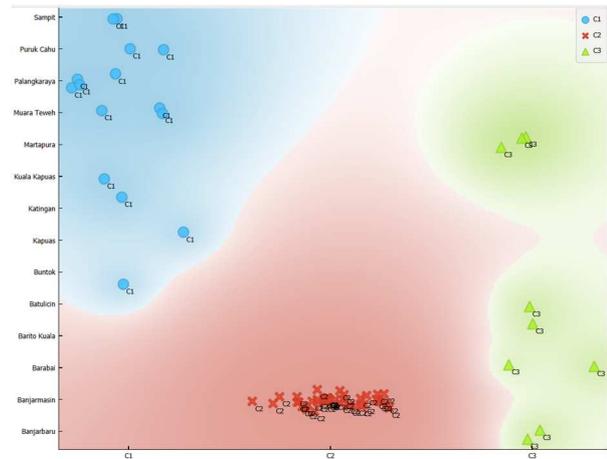
Pada gambar 8 di atas merupakan hasil visualisasi Distributions dari klasterisasi menggunakan metode K-Means Clustering.

Pada gambar 9 di bawah ini merupakan visualisasi berdasarkan perhitungan Euclidean Distance dengan menampilkan keterangan asal sekolah dari mahasiswa. Nilai jarak Euclidean Distance pada C3 adalah 0.139, C2 adalah 1.000 dan C1 0.367.



Gambar 9 Hasil K-Means Clustering menggunakan perhitungan Euclidean Distance

Pada gambar 10 berikut merupakan hasil klasterisasi berdasarkan dari asal kota mahasiswa menggunakan Scatter Plot.



Gambar 10 Hasil K-Means Clustering menggunakan Scatter Plot

Penelitian ini dilakukan dengan tujuan untuk mengetahui kelompok mahasiswa berdasarkan Asal Sekolah, Kota dan Provinsi. Dataset yang digunakan adalah data mahasiswa salah program studi di perguruan tinggi swasta, dataset yang diambil adalah data 3 tahun terakhir yang tersedia. Pengelompokan dilakukan dengan menggunakan Algoritma K-Means Clustering yang merupakan salah satu algoritma dengan pengimplementasian yang relatif mudah [21].

Pada penggunaan aplikasi RapidMiner dan Orange Data Mining mendapatkan hasil yang sama. Hal ini menunjukkan penentuan nilai centroid yang disampaikan oleh [15] yang menjadi kelemahan pada metode K-Means Clustering bisa ditangani dengan baik oleh aplikasi yang dipakai. Berdasarkan hasil klaster, terlihat ada 3 klaster dengan nilai yang berbeda-beda, pada data cluster_2 (pada RapidMiner) / C2 (Pada Orange Data Mining) jumlah anggota klaster yang didapat adalah 41, pada cluster_0 / C3 jumlah anggota klaster yang didapat adalah 15 dan pada cluster_1 / C1 jumlah anggota klaster yang didapat adalah 10.

Data cluster_2 / C2 merupakan klaster dengan anggota klaster terbanyak, pada klaster ini Provinsi Kalimantan Selatan dengan Kota Banjarmasin yang memiliki anggota terbanyak, sehingga menjadi klaster tersendiri. Data cluster_0 / C3 merupakan klaster dengan nilai kedua tertinggi, klaster ini didapatkan pada Provinsi Kalimantan Tengah dengan berbagai Kota di dalamnya. Data cluster_1 / C1 memiliki nilai terendah yaitu 10. Pada klaster ini didapatkan pada Provinsi Kalimantan Selatan, tetapi dengan kota yang berbeda-beda.

Pada perhitungan performa dengan menggunakan RapidMiner terlihat metode K-Means Clustering memiliki nilai 95%. Hal ini sesuai dengan penelitian yang dilakukan oleh [19] yang menyebutkan bahwa

klasterisasi yang dilakukan oleh *K-Means Clustering* memiliki tingkat akurasi 100%.

Berdasarkan hasil diatas, terlihat penggunaan metode *K-Means Clustering* pada 2 aplikasi data mining berjalan dengan baik dengan menunjukkan hasil klasterisasi yang sama. Hasil diatas dapat digunakan oleh tim penjangkaran mahasiswa untuk meningkatkan pola promosi sesuai dengan klaster yang ada. Selain itu pada pihak pengambil keputusan akan lebih mudah melakukan strategi selanjutnya dalam menentukan lokasi promosi.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka kesimpulan yang didapat adalah implementasi Metode *K-Means Clustering* pada pengelompokan mahasiswa berdasarkan asal sekolah, kota dan provinsi dengan menggunakan aplikasi *RapidMiner* dan *Orange Data Mining* menghasilkan 3 klaster, yaitu cluster_2 /C2 yang memiliki nilai 41 dengan beranggotakan data Kota Banjarmasin pada Provinsi Kalimantan Selatan, kemudian cluster_0/C3 memiliki nilai 15 dengan anggota beberapa kota di Provinsi Kalimantan Tengah, serta cluster_1/C1 dengan nilai 10 yang beranggotakan beberapa kota di Provinsi Kalimantan Selatan.

Hasil tersebut akan membantu pengambil kebijakan dan tim promosi untuk menentukan strategi penjangkaran mahasiswa baru pada tahun berikutnya. Hasil dari klasterisasi ini juga akan dapat membantu memaksimalkan penjangkaran pada lokasi yang tepat bagi tim promosi kampus.

Penelitian selanjutnya bisa menambahkan data mahasiswa dari program studi lain yang memungkinkan untuk dilakukannya klasterisasi wilayah berdasarkan peminatan program studi.

Daftar Rujukan

- [1] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Islam Indragiri)," *Jurnal SISTEMASI*, vol. 7, no. 3, pp. 238–249, 2018, doi: <https://doi.org/10.32520/stmsi.v7i3.388>.
- [2] W. Lestari, "Clustering Data Mahasiswa Menggunakan Algoritma K-Means Untuk Menunjang Strategi Promosi (Studi Kasus: STMIK Bina Bangsa Kendari)," *SIMKOM*, vol. 4, no. 2, pp. 35–48, Jul. 2019, [Online]. Available: <http://e-jurnal.stmikbinsa.ac.id/index.php/simkom35>
- [3] L. Qadrini, "Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017," *J Statistika*, vol. 13, no. 2, pp. 5–11, 2020, doi: <https://doi.org/10.36456/jstat.vol13.no2.a2886>.
- [4] A. Octaviani and P. Dewi, "Big Data di Perpustakaan dengan Memanfaatkan Data Mining," *ANUVA*, vol. 4, no. 2, pp. 223–230, 2020, doi: [10.14710/anuva.4.2.223-230](https://doi.org/10.14710/anuva.4.2.223-230).
- [5] B. Maryanto, "Big Data dan Pemanfaatannya Dalam Berbagai Sektor," *Media Informatika*, vol. 16, no. 2, pp. 14–19, 2017, doi: [10.37595/mediainfo.v17i1](https://doi.org/10.37595/mediainfo.v17i1).
- [6] G. Triyandana, L. A. Putri, and Y. Umidah, "Penerapan Data Mining Pengelompokan Menu Makanan dan Minuman

- Berdasarkan Tingkat Penjualan Menggunakan Metode K-Means," *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 1, pp. 40–46, Jul. 2022, doi: [10.30871/jaic.v6i1.3824](https://doi.org/10.30871/jaic.v6i1.3824).
- [7] D. K. Sitinjak, B. A. Pangestu, and B. N. Sari, "Clustering Jumlah Tenaga Kesehatan Berdasarkan Kecamatan di Kabupaten Karawang Menggunakan Algoritma K-Means," *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 1, pp. 46–54, Jul. 2022, doi: [10.30871/jaic.v6i1.3855](https://doi.org/10.30871/jaic.v6i1.3855).
- [8] J. Nasir, "Penerapan Data Mining Clustering Dalam Mengelompokan Buku Dengan Metode K-Means," *Jurnal SIMETRIS*, vol. 11, no. 2, Nov. 2020, doi: [10.24176/simet.v11i2.5482](https://doi.org/10.24176/simet.v11i2.5482).
- [9] B. V and R. G, "Hyperspectral Image Processing in Internet of Things model using Clustering Algorithm," *Journal of ISMAC*, vol. 3, no. 2, pp. 163–175, Jun. 2021, doi: [10.36548/jismac.2021.2.008](https://doi.org/10.36548/jismac.2021.2.008).
- [10] R. Muliono and Z. Sembiring, "Data Mining Clustering Menggunakan Algoritma K-Means Untuk Klasterisasi Tingkat Tridarma Pengajaran Dosen," *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 2, pp. 272–279, Jul. 2019, doi: [10.24114/cess.v4i2.13620](https://doi.org/10.24114/cess.v4i2.13620).
- [11] N. N. Fransiska R, D. S. Anggraeni, and U. Enri, "Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient," *TEMATIK Jurnal Teknologi Informasi Komunikasi*, vol. 9, no. 1, pp. 29–35, Jun. 2022, doi: [10.38204/tematik.v9i1.921](https://doi.org/10.38204/tematik.v9i1.921).
- [12] E. Desi, S. Aliyah, S. Lestari, and W. Dari, "Implementasi Algoritma K-Means Untuk Penerimaan Siswa Baru Di SMANPAS Berdasarkan Nilai Rapot dan Hasil Tes," *IT (INFORMATIC TECHNIQUE) JOURNAL*, vol. 10, no. 1, pp. 01–10, Apr. 2022, doi: [10.22303/it.10.1.2022.01-10](https://doi.org/10.22303/it.10.1.2022.01-10).
- [13] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster," in *IOP Conference Series: Materials Science and Engineering*, Apr. 2018, vol. 336, no. 1. doi: [10.1088/1757-899X/336/1/012017](https://doi.org/10.1088/1757-899X/336/1/012017).
- [14] D. Triyansyah and D. Fitriana, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *IncomiTech, Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 3, pp. 163–182, Jul. 2018, doi: [10.22441/incomtech.v8i2.4174](https://doi.org/10.22441/incomtech.v8i2.4174).
- [15] Priati and A. Fauzi, "Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore," in *Seminar Nasional Informatikadan Aplikasinya (SNIA)*, Sep. 2017, pp. 15–19. [Online]. Available: <http://community.tableau.com>.
- [16] R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *Journal of Software Engineering Ampera*, vol. 1, no. 1, pp. 1–14, Feb. 2020, doi: <https://doi.org/10.51519/journalsea.v1i1.10>.
- [17] W. Purba, S. Tamba, and J. Saragih, "The effect of mining data k-means clustering toward students profile model drop out potential," in *Journal of Physics: Conference Series*, Apr. 2018, vol. 1007, no. 1, pp. 1–6. doi: [10.1088/1742-6596/1007/1/012049](https://doi.org/10.1088/1742-6596/1007/1/012049).
- [18] D. Herawatie, E. Wuryanto, and Purbandini, "Perbandingan Algoritma Pengelompokan Non-Hierarki untuk Dataset Dokumen," in *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) Yogyakarta*, Jun. 2014, pp. 11–16. [Online]. Available: www.kompas.com
- [19] I. Firman Ashari, R. Banjarmasin, D. R. Farida, S. P. Aisyah, A. P. Dewi, and N. Humaya, "Application of Data Mining with the K-Means Clustering Method and Davies Bouldin Index for Grouping IMDB Movies," *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 1, pp. 7–15, Jul. 2022, doi: [10.30871/jaic.v6i1.3485](https://doi.org/10.30871/jaic.v6i1.3485).

- [20] A. Chusyairi and P. Ramadar Noor Saputra, "Pengelompokan Data Puskesmas Banyuwangi Dalam Pemberian Imunisasi Menggunakan Metode K-Means Clustering," *Telematika*, vol. 12, no. 2, pp. 139–148, Aug. 2019, doi: 10.35671/telematika.v12i2.848.
- [21] M. A. P. Sari and U. Chotijah, "Pengelompokan Anggota Divisi Himpunan Mahasiswa Jurusan Pada Universitas 'XYZ' Dengan Metode K-Means Clustering," *ANTIVIRUS: Jurnal Ilmiah Teknik Informatika*, vol. 16, no. 1, pp. 52–62, May 2022, doi: doi.org/10.35457/antivirus.v16i1.2139.