Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

Article DOI : 10.36350/jbs.v12i2.139
Received: May; Accepted: June; Published: July

Penerapan Algoritma K-Means Untuk Pemetaan Potensi Calon Mahasiswa Baru

Safi'I Eko Triyono[1*], Irmayansyah[2]
[1]Sistem Informasi/Universitas Binaniga Indonesia
Email: safeeeko@gmail.com
[2]Sistem Informasi/Universitas Binaniga Indonesia
Email: irma@unbin.ac.id

ABSTRACT

The process of mapping the potential of prospective new students is a grouping of prospective students based on various criteria which will be grouped based on their potential, both low and high, in order to assist the PR/promotion bureau in providing reference data and information in order to determine what promotion strategy will be carried out in the next period. This research can provide a mapping of potential new students using the K-Means Clustering Algorithm, namely by analyzing the initial data group, transforming the initial data and performing grouping calculations, after that the results of the grouping calculations can be re-analyzed to see where the school came from, the location of the school and the location of the school. the origin of information for each member in each group. In it applied the variables, namely the origin of the school, the location of the origin of the school and the origin of the information. This is done to map the potential of prospective new students, in order to assist the public relations/promotion bureau in providing reference data and information in order to determine what promotion strategy will be carried out in the next period. A cluster validity test has been carried out using the Silhoutte Coefficiency of the K-Means algorithm which is applied with a value of 0.6113 which means that the cluster created is included in the "Medium Structure" category, it is based on the Sillhoutte Category table according to Kauffman and Roussseeuw.

***Keywords**: Clustering; Mapping; Potential; K-Means Algorithm; Promotion.*

ABSTRAK

Proses pemetaan potensi calon mahasiswa baru merupakan pengelompokan calon mahasiswa berdasarkan berbagai kriteria yang nantinya dikelompokan berdasarkan potensinya baik itu rendah maupun tinggi guna membantu bagian promosi dalam menyediakan acuan data dan informasi agar dapat menentukan strategi promosi apa yang akan dijalankan di periode selanjutnya. Pada penelitian ini dapat memberikan pemetaan potensi calon mahasiswa baru menggunakan Algoritma K-Means Clustering yaitu dengan cara menganalisis kelompok data awal, mentransformasi data awal dan melakukan perhitungan pengelompokan, setelah itu hasil perhitungan pengelompokan dapat dianalisis kembali untuk melihat darimana asal sekolah, lokasi asal sekolah dan asal informasi masing-masing anggota yang ada di masing-masing kelompok. Didalamnya diterapkan variabel-variabel yaitu asal sekolah, lokasi asal sekolah dan asal informasi. Hal ini dilakukan untuk memetakan potensi calon mahasiswa baru, supaya dapat membantu bagian promosi dalam menyediakan acuan data dan informasi agar dapat menentukan strategi promosi apa yang akan dijalankan di periode selanjutnya. Telah dilakukan uji validitas cluster menggunakan *Silhoutte Coeffisient* terhadap algoritma K-Means yang diterapkan dengan nilai yang di dapat sebesar 0.6113 yang bermakna cluster yang dibuat termasuk dalam kategori "*Medium Structure*", hal tersebut didasarkan pada tabel kategori sillhoutte menurut Kauffman dan Rousseuw

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

***Keywords**: Klusterisasi; Pemetaan; Potensi; Algoritma K-Means; Promosi.*

A. PENDAHULUAN

1. Latar Belakang

Mahasiswa diambil dari dua kata yaitu maha dan siswa, dengan kata lain adalah pelajar yang paling tinggi levelnya. Sebagai seorang pelajar tertinggi, tentu mahasiswa seharusnya sudah terpelajar, sebab mereka tinggal menyempurnakan pembelajaranya. Dalam peraturan pemerintah RI No. 30 tahun 1990, mahasiswa adalah peserta didik yang terdaftar dan belajar di perguruan tinggi tertentu. Sedangkan menurut Sarwono (1978) mahasiswa adalah setiap orang yang secara resmi terdaftar untuk mengikuti setiap pelajaran yang ada di suatu perguruan tinggi. Mahasiswa juga merupakan insan-insan calon sarjana yang dalam keterlibatannya dengan perguruan tinggi dididik dan diharapkan menjadi calon-calon intelektual dalam suatu kelompok ataupun dalam masyarakat yang telah memperoleh statusnya karena ikatan dengan suatu lembaga perguruan tinggi.

Promosi adalah kegiatan yang berperan aktif dalam menyuarakan kembali keunggulan dari suatu produk atau jasa guna mendorong konsumen untuk membeli produk atau jasa yang dipromosikan. Untuk menjalankan promosi, perlu menentukan dengan tepat dimana lokasi dan alat promosi apa yang dapat digunakan untuk dapat mencapai kesuksesan dalam penjualan. Menurut Nickels dalam Swastha & Irawan (2008) promosi adalah arus informasi atau persuasi satu arah yang dirancang untuk mengarahkan satu orang atau lebih kepada tindakan yang mampu menghasilkan pertukaran atau transaksi dalam pemasaran. Metode yang digunakan dalam kegiatan promosi meliputi periklanan, promosi penjualan, penjualan perseorangan hingga hubungan masyarakat. Promosi menunjuk pada berbagai aktivitas yang dilakukan untuk mengkomunikasikan kebaikan produknya dan membujuk para sasaran yaitu pelanggan dan konsumen untuk membeli produk tersebut. Sehingga dapat disimpulkan suatu promosi yaitu dasar kegiatan komunikasi dengan para pelanggan ataupun konsumen untuk mendorong terciptanya penjualan.

Kata strategi berasal dari kata Strategos yang dalam bahasa Yunani merupakan gabungan dari Stratos atau tentara dan Ego atau pemimpin. Suatu strategi pasti memiliki dasar atau skema untuk dapat mencapai sasaran yang dituju. Jadi pada dasarnya strategi adalah suatu alat untuk mencapai suatu tujuan. Menurut Marrus (2002) strategi diartikan sebagai suatu proses pendefinisian suatu rencana yang menitik beratkan pada tujuan dengan jangka panjang, disertai dengan penyusunan metode atau cara bagaimana agar dapat mencapai tujuan tersebut. Strategi juga dapat didefinisikan sebagai suatu bentuk rencana yang mengintegrasikan tujuan, kebijakan dan rangkaian tindakan dalam suatu organisasi menjadi satu kesatuan yang utuh. Strategi yang dirancang dengan baik akan membantu mengelola sumber daya yang dimiliki perusahaan menjadi suatu bentuk yang unik dan berkelanjutan.

Penelitian ini akan me-metakan potensi calon mahasiswa baru yang ada pada sekolah-sekolah di wilayah Kota Bogor dengan menerapkan metode Algoritma K-Means Clustering. Algoritma K-Means merupakan algoritma pengelompokan berulang yang mempartisi set data ke dalam jumlah K cluster yang sudah ditentukan sebelumnya. Algoritma K-Means mudah untuk diimplementasikan dan dijalankan, relatif cepat, mudah beradaptasi, banyak digunakan dalam praktik. Secara historis, K-Means telah menjadi salah satu algoritma terpenting dalam topik atau bidang data mining. K-Means termasuk dalam partitioning clustering yaitu setiap data harus dimasukkan ke dalam cluster pada suatu tahapan proses, pada tahapan berikutnya berpindah, diikuti dengan beralih ke cluster yang lain. Algoritma K-Means dikenal luas karena kemudahan dan kemampuannya untuk mengklasifikasikan data (Prasetyo, 2014).

Berdasarkan uraian diatas, metode K-Means Clustering diharapkan dapat membantu menyelesaikan permasalahan Pemetaan Potensi Calon Mahasiswa Baru" berdasarkan asal sekolah, lokasi asal sekolah, asal informasi dan tahun mendaftar.

2. Permasalahan

Permasalahan yang dihadapi yaitu belum diketahui secara akurat peta potensi calon mahasiswa baru dan belum efektifnya proses pemetaan potensi calon mahasiswa baru. Sejauh ini pihak biro humas/promosi dalam penentuannya melakukan promosi dilakukan tanpa melihat terlebih dahulu potensi-potensi yang ada pada wilayah-wilayah tertentu, sehingga potensi calon mahasiswa yang ada di wilayah Kota Bogor tidak bisa diketahui secara pasti dan potensi untuk bisa menampung siswa-siswa di masing-masing wilayah belum bisa dilihat dengan jelas. Berdasarkan hal tersebut maka dilakukanlah penelitian ini yang berjudul ”Penerapan Algoritma K-Means Clustering untuk Pemetaan Potensi Calon Mahasiswa Baru”.

3. Tujuan

Adapun tujuan dari penelitian ini adalah :

a. Mendapatkan pemetaan potensi calon mahasiswa baru
b. Mendapatkan proses pemetaan potensi calon mahasiswa baru yang akurat
c. Mengembangkan prototype aplikasi pemetaan potensi calon mahasiswa baru menggunakan metode Algoritma K-Means;
d. Mengukur tingkat akurasi dan efektivitas metode K-Means untuk pemetaan potensi calon mahasiswa baru.

4. Tinjauan Pustaka

a. Clustering

Menurut Prasetyo (2014:186) menyatakan bahwa *Clustering* adalah suatu teknik untuk menemukan sekelompok data dari pemecahan atau pemisahan sekumpulan data menurut karakteristik atau kriteria yang telah ditentukan. Dalam pengelompokan tersebut nilai label nya belum diketahui sehingga diharapkan setelah melakukan pengelompokan maka label data dapat diketahui dari data tersebut. Metode *clustering* juga sering disebut tahapan awal sebelum melakukan metode lain seperti klasifikasi

Analisis *Clustering* adalah mengelompokkan suatu data objek pada informasi yang mirip atau memiliki kesamaan antara satu dengan yang lainnya, tujuannya agar dapat menemukan kelompok yang berkualitas sama seperti kelompok yang merupakan objek-objek yang mirip atau memiliki hubungan satu dengan yang lain, dan sebaliknya yaitu kelompok yang tidak memiliki hubungan dengan objek dalam kelompok yang lain. *Clustering* cocok digunakan untuk menjelajahi suatu data. Jika ada banyak kasus tetapi tidak ada pengelompokan yang jelas, algoritma *clustering* dapat digunakan untuk melakukan pengelompokan dari data tersebut. *Clustering* juga dapat berguna sebagai data-*preprocessing* yaitu suatu langkah untuk mengidentifikasi kelompok-kelompok yang berhubungan dalam membangun suatu model.

b. Data Mining

Menurut Hermawati (2013:1) data mining adalah suatu proses pengolahan satu atau lebih teknik pembelajaran komputer (machine learning) untuk menganalisa dan mengektraksi pengetahuan (knowledge). Definisi lain di antaranya adalah pembelajaran berbasis induksi (induction-based learning) adalah suatu proses pembentukan definisi-definisi dari konsep umum yang dilakukan dengan cara mengobservasi contoh-contoh detail dan spesifik dari konsep-konsep yang nantinya akan dipelajari.

Data mining berisi pencarian trend atau pola yang dibutuhkan dalam database yang besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dapat dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih rinci, yang dapat menggunakan perangkat pendukung keputusan yang lainnya.

Menurut Turban (2001:3) Knowledge Discovery in Databases (KDD) adalah suatu penerapan metode saintifik pada data mining. Dalam konteks ini data mining merupakan termasuk ke dalam salah satu langkah dari proses KDD.

KDD berhubungan dengan suatu teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola suatu data. Serangkaian proses tersebut memiliki tahap sebagai berikut ini:

1) Pembersihan data yaitu untuk membuang data yang tidak konsisten dan noise.

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

2) Integrasi data yaitu penggabungan data dari beberapa sumber
3) Transformasi data yaitu data diubah menjadi bentuk yang sesuai untuk di mining.
4) Aplikasi teknik data mining yaitu proses ekstraksi pola dari data yang ada.
5) Evaluasi pola yang ditemukan yaitu proses interpretasi pola menjadi pengetahuan yang dapat digunakan untuk menunjang pengambilan keputusan.
6) Presentasi pengetahuan yaitu dengan teknik visualisasi. Langkah ini merupakan bagian dari proses pencarian pengetahuan, yang melibatkan pengecekan apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau asumsi yang sudah ada sebelumnya. Langkah terakhir dalam KDD adalah menyajikan pengetahuan dalam bentuk yang mudah dipahami oleh pengguna.

B. METODE

1. Algoritma K-Means

Parameter yang harus dimasukkan ketika menggunakan algoritma K-Means adalah nilai K. Nilai K yang digunakan biasanya didasarkan pada informasi yang diketahui sebelumnya tentang berapa banyak cluster data yang muncul dalam X, berapa banyak cluster yang dibutuhkan untuk penerapannya, atau jenis cluster dicari dengan mengeksplorasi atau melakukan percobaan dengan beberapa nilai K. Berapa nilai K yang dipilih tidak perlu memahami bagaimana K-Means mempartisi set data X.

Dalam K-Means setiap cluster dari K cluster diwakili oleh titik tunggal dalam R^d. Set representatif cluster dinyatakan C={c_j |j=1,…,K}. Sejumlah K representatif cluster tersebut disebut juga sebagai cluster means atau cluster centroid (atau centroid saja). Untuk set data dalam X dikelompokkan bersama berdasarkan konsep kedekatan atau kemiripan. Meskipun konsep yang dimaksud untuk data – data yang berkumpul dalam satu cluster adalah data – data yang memiliki kemiripan, tetapi besaran yang digunakan untuk mengukurnya adalah ketidakmiripan (dissimilarity). Artinya, data-data dengan ketidakmiripan (jarak) yang kecil atau dekat maka lebih besar kemungkinannya untuk bergabung dalam satu cluster. Metrik yang umum digunakan untuk ketidakmiripannya adalah Euclidean.Saat data sudah dihitung ketidakmiripan terhadap setiap centroid, maka selanjutnya dipilih ketidakmiripan yang paling kecil sebagai cluster yang akan diikuti sebagai pergerakan data dalam cluster pada sebuah iterasi. Pergerakan sebuah data dalam cluster yang diikuti dapat dinyatakan dengan nilai 0 atau 1. Nilai 0 jika tidak menjadi anggota sebuah cluster dan 1 jika menjadi anggota sebuah cluster. Karena K-Means mengelompokkan secara tegas data hanya pada satu cluster, maka dari nilai a sebuah data pada semua cluster, hanya satu yang bemilai1, sedangkan lainnya 0. seperti dinyatakan oleh persamaan berikut:

$$a_{ij} = \begin{cases} 1 & arg\ \min_{j} \{(X_i, C_j)\} \\ 0 & L = lainnya \end{cases}$$

Sementara relokasi centroid untuk mendapatkan titik centroid C didapatkan dengan menghitung rata – rata setiap fitur dari semua data yang tergabung dalam setiap cluster. Rata-rata sebuah fitur dari semua data dalam sebuah cluster dinyatakan oleh persamaan berikut:

$$C_j = \frac{1}{Nk} \sum_{l=1}^{Nk} x_{jl}$$

Nk adalah jumlah data yang tergabung dalam sebuah cluster. Jika diperhatikan dari langkahnya yang selalu memilih cluster terdekat, maka sebenarnya K-Means berusaha untuk meminimalkan fungsi objektif/fungsi biaya non-negatif, seperti dinyatakan oleh persamaan berikut:

$$J = \sum_{i=1}^{N} \sum_{l=1}^{K} a_{ic} d(x_i, C_l)^2$$

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

Dengan kata lain, K-Means berusaha untuk meminimalkan total jarak kuadrat (*squared distance*) di antara setiap titik x_i dan representasi cluster c_j terdekat.

2. Teknik Analisa Data

Metode pengujian yang diterapkan dalam penelitian ini adalah metode *Silhoutte Coeffisient*. Metode ini akan menguji kualitas dari setiap *cluster* yang dihasilkan dengan menggabungkan metode *cohesion* dan *separation*. Ada tiga langkah yang perlu dilakukan untuk menghitung *Silhoutte Coeffisient*, yaitu:

Untuk setiap objek i, hitung rata-rata jarak objek i dengan seluruh objek yang berada dalam satu *cluster*. Maka akan didapatkan nilai rata-rata yang disebut dengan a_i.

a. Untuk setiap objek i, hitung rata-rata jarak objek i dengan seluruh objek yang berada dalam satu *cluster*. Maka akan didapatkan nilai rata-rata yang disebut dengan a_i.
b. Untuk setiap objek i, hitung rata-rata jarak dari objek i dengan objek yang berada di *cluster* lainnya. Dari semua jarak rata-rata tersebut diambil nilai yang paling kecil. Nilai ini disebut dengan b_i
c. Setelah itu maka nilai *Silhoutte Coeffisient* dari objek i adalah:

$S_i = (b_i - a_i)/max(a_i, b_i)$

Keterangan

a_i: Rata-rata jarak objek i terhadap seluruh objek di dalam *cluster* b_i: Rata-rata jarak objek i terhadap seluruh objek di luar *cluster*

Ukuran nilai *Silhoutte Coeffisient* dapat dilihat pada tabel 1

Tabel 1. Kategori Sillhoutte Menurut Kauffman dan Roussseeuw

Nilai Silhoutte Coeffisient	Keterangan
0,7 < SC <= 1	Strong structure
0,5 < SC <= 0,7	Medium structure
0,25 < SC <= 0,5	Weak structure
SC <= 0,25	No structure

C. HASIL DAN PEMBAHASAN

1. Hasil

a. Data Transformation

Data yang diperoleh adalah data kualitatif berjenis nominal maka perlu dilakukan transformasi data pada beberapa atribut, yaitu kecamatan, asal sekolah dan jurusan. Agar data tersebut dapat diolah menggunakan Algoritma K-Means Clustering maka dilakukan transformasi data ke dalam bentuk angka (Simovici, 2012). Berikut proses:

1) Urutkan data berdasarkan frekuensi terbesar dari masing-masing atribut tersebut.
2) Kemudian dimulai dari frekuensi yang terbesar diberi inisial mulai dari angka 1, 2 dan seterusnya secara berurut-turut sampai frekuensi data terkecil.

Tabel 2. Transformasi Data Data Atribut Asal Sekolah

Inisial	Keterangan	Frekuensi	Tahun
1	SMK WIKRAMA	25	2018
2	SMK NEGERI 1	20	2018
3	SMK PEMBANGUNAN	15	2018
4	SMK NEGERI 4	8	2018
5	MA NEGERI 1	8	2018
6	SMK PGRI 3	7	2018
7	SMK BHAKTI INSANI	7	2018
8	SMK NEGERI 3	6	2018
…	…	…	…
56	SMK CITRA PARIWISATA	1	2018

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

Tabel 3. Transformasi Data Data Atribut Lokasi Asal Sekolah

Inisial	Keterangan	Frekuensi	Tahun
1	Bogor Utara	47	2018
2	Tanah Sareal	43	2018
3	Bogor Barat	36	2018
4	Bogor Timur	31	2018
5	Bogor Selatan	26	2018
6	Bogor Tengah	14	2018

Tabel 4. Transformasi Data Data Atribut Asal Informasi

Inisial	Keterangan	Frekuensi	Tahun
1	Teman/Saudara	124	2018
2	Internet/Website	52	2018
3	Brosur/Spanduk	14	2018
4	Staff/Dosen	7	2018

Tabel 5. Data PMB tahun 2018 Hasil Transformasi

No	Nama	Asal Sekolah	Lokasi Asal Sekolah	Asal Informasi	Tahun
1	BTARI ASMORODEWI	4	3	3	2018
2	HERRIS SETIAWATY	4	3	1	2018
3	MOCHAMAD ILHAM	4	3	4	2018
4	MUHAMMAD KAHFI SAHRUDIN	4	3	1	2018
5	ALZIA RAHMA FAUZIAH	4	3	1	2018
6	MUHAMMAD RAAFI RASYIIDIN	4	3	1	2018
7	REVYANADA	4	3	1	2018
8	ILHAM ADIYATNA	4	3	3	2018
9	EVA FITRIANI	34	4	2	2018
10	TRI RACHMI MAULIDI	24	6	2	2018
…	……	……	…….	……	…..
197	REVY MAYTRIANTY	56	3	2	2018

Tabel 5 menunjukkan data hasil proses transformasi yang dilakukan pada variabel Asal Sekolah, Lokasi Asal Sekolah dan Asal Informasi.

b. Penentuan Variabel Penelitian

Variabel yang digunakan dalam pemetaan potensi calon mahasiswa baru ini adalah asal sekolah, lokasi asal sekolah dan asal informasi.

1) Asal Sekolah
2) Lokasi Asal Sekolah
3) Asal Informasi

c. Menyiapkan Dataset

Dataset yang digunakan pada perhitungan ini menggunakan variabel data asal sekolah, lokasi asal sekolah, asal informasi dan tahun dari data PMB yang telah ditransformasi. Dataset tersebut terdiri dari 197 baris dan 6 kolom.

Tabel 6. Dataset Perhitungan K-Means Potensi Calon Mahasiswa Baru

No	Nama	Asal Sekolah	Lokasi Asal Sekolah	Asal Informasi	Tahun
1	BTARI	4	3	3	2018

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

	ASMORODEWI				
2	HERRIS SETIAWATY	4	3	1	2018
3	MOCHAMAD ILHAM	4	3	4	2018
4	MUHAMMAD KAHFI SAHRUDIN	4	3	1	2018
5	ALZIA RAHMA FAUZIAH	4	3	1	2018
6	MUHAMMAD RAAFI RASYIIDIN	4	3	1	2018
7	REVYANADA	4	3	1	2018
…	…	…	…	…	…
197	REVY MAYTRIANTY	56	3	2	2018

d. Menentukan Jumlah Cluster
Dari dataset perhitungan k-means potensi calon mahasiswa baru akan dibagi menjadi 2 kelompok/cluster sehingga nilai K pada perhitungan ini adalah 2. Dimana cluster 1 untuk kluster yang berisi anggota dengan potensi rendah dan cluster 2 untuk kluster yang berisi anggota dengan potensi tinggi.

e. Menentukan Titik Centroid
Pada langkah sebelumnya, dataset perhitungan k-means potensi calon mahasiswa baru akan dibagi menjadi 2 (dua) kelompok/cluster sehingga titik centroid awal yang dipilih juga sejumlah 2 (dua). Titik centroid awal yang digunakan dapat dilihat pada tabel 7.

Tabel 7. Nilai Centroid Awal

Centroid	x	y	z
Centroid 1	56	3	2
Centroid 2	1	4	2

f. Hitung Jarak Data Dengan Centroid
Perhitungan jarak data dengan centroid berfungsi untuk menentukan jarak terpendek untuk menentukan pengelompokan cluster, menghitung jarak data dengan centroid dengan menggunakan rumus *Euclidean Distance.Euclidean Distance* adalah perhitungan jarak dari 2 buah titik dan untuk mempelajari hubungan antara sudut dan jarak, rumusnya sebagai berikut:

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

Perhitungan Jarak dengan cluster 1:
Data ke 1:
$D(X_1,C_1)$
$=\sqrt{(a1 - C1a)^2 + (b1 - C1b)^2 + (c1 - C1c)^2}$
$=\sqrt{(4 - 56)^2 + (3 - 3)^2 + (3 - 2)^2}$
= 52.01
Data ke 2 :
$D(X_2,C_1)$
$=\sqrt{(a2 - C1a)^2 + (b2 - C1b)^2 + (c2 - C1c)^2}$
$=\sqrt{(4 - 56)^2 + (3 - 3)^2 + (1 - 2)^2}$
= 52.01
…
Data ke 197 :
$D(X_{197},C_1)$
$=\sqrt{(a197 - C1a)^2 + (b197 - C1b)^2 + (c197 - C1c)^2}$
$=\sqrt{(56 - 56)^2 + (3 - 3)^2 + (2 - 3)^2}$

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

= 0

Perhitungan Jarak dengan cluster 2:

Data ke 1:

$D(X_1,C_2)$

$=\sqrt{(a1 - C2a)^2 + (b1 - C2b)^2 + (c1 - C2c)^2}$

$=\sqrt{(4 - 1)^2 + (1 - 4)^2 + (2 - 2)^2}$

= 3.3166

Data ke 2:

$D(X_2,C_2)$

$=\sqrt{(a2 - C2a)^2 + (b2 - C2b)^2 + (c2 - C2c)^2}$

$=\sqrt{(4 - 1)^2 + (3 - 4)^2 + (1 - 2)^2}$

= 3.3166

…

Data ke 197:

$D(X_{197},C_2)$

$=\sqrt{(a197 - C2a)^2 + (b197 - C2b)^2 + (c197 - C2c)^2}$

$=\sqrt{(56 - 1)^2 + (3 - 41)^2 + (2 - 2)^2}$

=55.009

Tabel 8 Hasil Pengelompokan Iterasi Pertama

No	Nama	Asal Sekolah	Lokasi Asal Sekolah	Asal Informasi	c1	c2	Jarak Terpendek	Hasil
1	BTARI ASMORODEWI	4	3	3	52.01	3.3166	3.3166	Kluster C2
2	HERRIS SETIAWATY	4	3	1	52.01	3.3166	3.3166	Kluster C2
3	MOCHAMAD ILHAM	4	3	4	52.038	3.7417	3.7417	Kluster C2
4	MUHAMMAD KAHFI SAHRUDIN	4	3	1	52.01	3.3166	3.3166	Kluster C2
5	ALZIA RAHMA FAUZIAH	4	3	1	52.01	3.3166	3.3166	Kluster C2
6	MUHAMMAD RAAFI RASYIIDIN	4	3	1	52.01	3.3166	3.3166	Kluster C2
7	REVYANADA	4	3	1	52.01	3.3166	3.3166	Kluster C2
8	ILHAM ADIYATNA	4	3	3	52.01	3.3166	3.3166	Kluster C2
9	EVA FITRIANI	34	4	2	22.023	33	22.023	Kluster C1
10	TRI RACHMI MAULIDI	24	6	2	32.14	23.087	23.087	Kluster C2
…	…	…	…	…	...	...	...	…
197	REVY MAYTRIANTY	56	3	2	0	55.009	0	Kluster C1

Tabel 8 Merupakan hasil perhitungan dengan menggunakan rumus *Euclidean Distance* antara asal sekolah (variabel 1), lokasi asal sekolah (variabel 2) dan asal informasi (variabel 3) dengan pusat centroid 1 dan centroid 2 sehingga didapat jarak terdekat dari masing-masing atrbut (data N), kemudian untuk ditentukan masuk ke cluster masing-masing.

Setelah proses pengelompokan data lalu dilakukan penentuan titik centroid baru. Nilai centroid baru didapatkan dengan menggunakan rumus

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

$$\mu_k = \frac{1}{N_k}\sum_{t=1}^{N_k} x_i$$

Keterangan :
μ_k : Titik centroid dari *cluster* ke-K
N_k : Banyaknya data pada *cluster* ke-K
x_i : Data ke-*i cluster* ke-K

1) Titik centroid 1 baru untuk iterasi ke 2

$$x = \frac{1345}{33} = 40.75757576 \qquad y = \frac{114}{33} = 3.454545455 \qquad z = \frac{56}{33} = 1.696969697$$

2) Titik centroid 2 baru untuk iterasi ke 2

$$x = \frac{1442}{164} = 8.792682927 \qquad y = \frac{465}{164} = 2.835365854 \qquad z = \frac{242}{164} = 1.475609756$$

Nilai titik centroid baru dapat dilihat pada tabel 9.

Tabel 9 Nilai Centroid Baru

Centroid Baru	x	Y	z
Centroid 1	40.75757576	3.454545455	1.696969697
Centroid 2	8.792682927	2.835365854	1.475609756

Ulangi perhitungan jarak data dengan centroid sampai nilai dari titik centroid tidak lagi berubah.Dari perhitungan yang dilakukan pada penelitian ini terjadi iterasi sebanyak 5 kali titik pusat cluster tidak ada lagi perubahan dan data pertama hingga data yang terakhir tidak ada lagi yang bergeser dari satu cluster ke cluster yang lainnya. Adapun hasil perhitungan pada iterasi terakhir dapat dilihat pada tabel 10 di bawah ini :

Tabel 10. Hasil Kluster 1

Kluster 1 / Potensi Rendah					
Asal Sekolah		Lokasi Asal Sekolah		Asal Informasi	
Nama	Jumlah	Nama	Jumlah	Nama	Jumlah
SMK PERMATA 2	3 / 1.5%	Bogor Barat	12 / 6,1%	Internet/Website	30/ 15,2%
SMK PGRI 1	3 / 1.5%	Tanah Sareal	9 / 4,6%	Teman/Saudara	19/ 9,6%
SMA AL-GHAZALY	2 / 1%	Bogor Utara	9 / 4,6%		
SMA BHAKTI INSANI	2 / 1%	Bogor Tengah	8/ 4,1%		
SMA NEGERI 10	2 / 1%	Bogor Selatan	8/ 4,1%		
SMA NEGERI 2	2 / 1%	Bogor Timur	3/ 1,5%		
SMA NEGERI 7	2 / 1%				
SMK BINA INFORMATIKA	2 / 1%				
SMK BINA SEJAHTERA 3	2 / 1%				
SMK MEKANIKA	2 / 1%				
SMK RANTI MULA	2 / 1%				
SMK TRI DHARMA 4	2 / 1%				
MA NEGERI 2	1 / 0,5%				
SMA KESATUAN	1 / 0,5%				
SMA NEGERI 5	1 / 0,5%				

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

Kluster 1 / Potensi Rendah					
Asal Sekolah		Lokasi Asal Sekolah		Asal Informasi	
Nama	Jumlah	Nama	Jumlah	Nama	Jumlah
SMA NEGERI 8	1 / 0,5%				
SMA NEGERI 9	1 / 0,5%				
SMA YPHB	1 / 0,5%				
SMK AL-AZHAR	1 / 0,5%				
SMK AL-GHAZALY	1 / 0,5%				
SMK BINA SEJAHTERA 1	1 / 0,5%				
SMK BINA SEJAHTERA 2	1 / 0,5%				
SMK BINA WARGA 2	1 / 0,5%				
SMK CITRA PARIWISATA	1 / 0,5%				
SMK GRAFIKA MARDI YUANA	1 / 0,5%				
SMK INSAN MUTIARA AWANI BANGSA	1 / 0,5%				
SMK KOSGORO	1 / 0,5%				
SMK PERMATA 1	1 / 0,5%				
SMK PGRI 2	1 / 0,5%				
SMK PUI	1 / 0,5%				
SMK SIROJUL HUDA	1 / 0,5%				
SMK SURYAKENCANA YAPIS	1 / 0,5%				
SMK TELEKOMEDIKA	1 / 0,5%				
SMK YZA 3	1 / 0,5%				
SMK YZA 4	1 / 0,5%				

Tabel 11. Hasil Kluster 2

Kluster 2 / Potensi Tinggi					
Asal Sekolah		Lokasi Asal Sekolah		Asal Informasi	
Nama	Jumlah	Nama	Jumlah	Nama	Jumlah
SMK WIKRAMA	25 /12,6%	Bogor Utara	38 / 19,2%	Teman/Saudara	105 / 53,3%
SMK NEGERI 1	20 /10,1%	Tanah Sareal	34 / 17,2%	Internet/Website	22 / 11,2%
SMK PEMBANGUNAN	15 /7,6%	Bogor Timur	28 / 14,2%	Brosur/Spanduk	14 / 7,1%
MA NEGERI 1	8 /4,1%	Bogor Barat	24 / 12,2%	Staff/Dosen	7 / 3,5%
SMK NEGERI 4	8 /4,1%	Bogor Selatan	18 / 9,1%		
SMK BHAKTI INSANI	7 /3,5%	Bogor Tengah	6 / 3,4%	-	
SMK PGRI 3	7 /3,5 %				
SMA PGRI 3	6 / 3%				
SMK NEGERI 3	6 / 3%				
SMK TARUNA ANDIGHA	6 / 3%				
SMK YKTB 2	6 / 3%				
SMK TRI DHARMA 2	5 /2,5%				
SMA RIMBA MADYA	4 / 2%				
SMK NEGERI 2	4 / 2%				
SMA KOSGORO	3 /1,5%				
SMA NEGERI 4	3 /1,5%				

Volume 12 Number 2 July 2022 Page. 139-150
Journal Homepage : http://teknois.stikombinaniaga.ac.id/index.php/JBS
DOI Link : http://doi.org/10.36350/jbs.v12i2

SMA NEGERI 6	3 /1,5%		
SMA PGRI 1	3 /1,5%		
SMK BINA WARGA 1	3 /1,5%		
SMK INFOKOM	3 /1,5%		
SMK INFORMATIKAPESAT	3 /1,5%		

2. Pembahasan

Pada tahap ini dilakukan uji hasil dengan menerapkan metode Silhoutte Coeffisient. Metode ini bertujuan untuk menguji kualitas dari setiap cluster yang dihasilkan dengan menggabungkan metode cohesion dan separation. Dengan menggunakan aplikasi matlab, source code pada aplikasi matlab untuk menguji cluster dengan metode Silhoutte Coeffisient adalah sebagai berikut :

```
dataw = readtable('Bahan Rapidminer.xlsx');
X = [dataw.idAsalSekolah,dataw.idLokasi,dataw.idInformasi];jumlah_cluster = 2;
rand('seed',10);
[label,centroid] = kmeans(X,jumlah_cluster);
figure;
[S,H] = silhouette(X, label,'Euclidean'); title('nilai silhouette di tiap
cluster');
info=array2table([S,label],'variablenames',{'nilai','label'});disp(info);
S_cluster= [mean(S(label==1)) mean(S(label==2))];S_Score = mean(S_cluster);
```

Tabel 12 menunjukkan nilai silhouette dari masing-masing data yang berjumlah 197, Semakin nilai silhouette coefficient mendekati nilai 1, maka semakin baik pengelompokan data dalam satu cluster. Sebaliknya jika nilai silhouette mendekati nilai 0, maka semakin buruk pengelompokan data didalamsatu cluster.

Tabel 12. Nilai S

No	Nilai S
1	0.8227
2	0.8316
3	0.8084
4	0.8316
5	0.8316
6	0.8316
7	0.8316
8	0.8227
9	0.6495
10	0.2940
…	…
197	0.5563

Tabel 13. Nilai S_Cluster

S1	*S2*
0.4955	0.7271

Tabel 13 menunjukkan S1 dan S2 yang berisi hasil rata-rata jarak objek yang berada di dalam masing-masing cluster.

Tabel 14. Nilai S_Semua

S_Semua
0.6113

Tabel 14 menunjukkan nilai rata-rata dari S1 dan S2 yang merupakan hasil akhir dari perhitungan Silhouette Coefficient. Hasil perhitungan nilai Silhoutte berada pada nilai *S*1 (cluster 1) dengan point 0.4955, nilai *S*2 (cluster 2) dengan point 0.7271, Semakin sedikit nilai S yang nilainya dibawah 0 berarti obyek *I* sudah berada pada clusternya.

Mengacu pada tabel 1 nilai rata-rata validasi dari setiap cluster berada pada point 0.6113 (Medium Structure)

D. KESIMPULAN

Berdasarkan Hasil Penelitian yang dilakukan, kesimpulan yang bisa diuraikan antara lain :

1. Menerapkan metode Algoritma K-Means dapat memberikan peta potensi calon mahasiswa baru dengan akurat karena telah dilakukan uji akurasi menggunakan *silhouette coefficient*.
2. Hasil uji akurasi sebesar 0.6113 yang berarti termasuk kedalam kategori *Medium Structure*.

E. DAFTAR PUSTAKA

[1] Hermawati, F. A. (2013). *Data Mining*. Yogyakarta: Penerbit Andi.

[2] Kusrini & E.T Luthfi. (2009). *Algoritma Data Mining*. Andi. Yogyakarta.

[3] Moore, Andrew. (2001). K-*means and Hierarchical Clustering*. Pennsylvania.

[4] Santosa, Budi. (2007). *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Yogyakarta.

[5] Simovici, D. A. (2012). *The k -Means Clustering . In Linear Algebra Tools for Data Mining.* United States of America.

[6] Sugiyono. (2018). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta. Bandung.

[7] Tan, P.N., Steinbach, M., Kumar, V. (2006). *Introduction to Data Mining*. Pearson Education. Boston.

[8] Wanto, Anjar. (2020). *Data Mining : Algoritma dan Implementasi*. Medan. Yayasan Kita Menulis.

[9] Xindong Wu & Vipin Kumar. (2009). *The Top Ten Algorithms in Data Mining*. United States of America

[10] Yunita, Fitri. (2018). *Penerapan Data Mining Menggunkan Algoritma K-Means Clustring Pada Penerimaan Mahasiswa Baru*. Jurnal SISTEMASI, Volume 7, Nomor 3 September 2018 : 238 – 249