

Clustering Untuk Menentukan Strategi Promosi Universitas Muhammadiyah Jember Dengan Algoritma K-Medoids

Clustering To Determine The University Of Muhammadiyah Jember Promotion Strategy With K-Medoids Algorithm

Radiyanto Dekaprasetya¹, Ulya Anisatur R.^{2*}, Lutfi Ali Muharom³

¹ Mahasiswa Program studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember
email: raditex@gmail.com

² Fakultas Teknik, Universitas Muhammadiyah Jember* Koresponden Author
email: ulyaanisatur@unmuhjember.ac.id

³ Fakultas Teknik, Universitas Muhammadiyah Jember
email: lutfi.muharom@unmuhjember.ac.id

Abstrak

Dengan semakin banyaknya jumlah lembaga pendidikan setingkat perguruan tinggi baik negeri maupun swasta berdasarkan dokumen Statistik Pendidikan Tinggi tahun 2019 (Kemenristekdikti, 2019) menjadi pemicu persaingan antar lembaga pendidikan tersebut untuk mencari calon mahasiswa. Universitas Muhammadiyah Jember sebagai salah satu perguruan tinggi unggulan di Jawa Timur khususnya di Kabupaten Jember dalam hal menjaring calon mahasiswa membutuhkan strategi promosi yang baik untuk mendapatkan calon mahasiswa sesuai pagu. Penelitian ini dilakukan untuk mendapatkan informasi profil mahasiswa dengan cara mengolah data mahasiswa baru tahun 2020/2021. Pengolahan data dengan cara clustering menggunakan algoritma k-medoids terhadap data mahasiswa seperti Prodi, Kota, Propinsi, Pekerjaan Orang Tua dan Jenis Sekolah. Hasil dari penelitian adalah dengan mendapatkan nilai DBI yang rendah pada cluster $k = 3$ sehingga ditentukan sebagai cluster yang optimal. Mahasiswa terbanyak dari kota Jember, Banyuwangi dan Bondowoso, propinsi Jawa Timur, program studi yang banyak dipilih adalah Teknik Informatika serta pekerjaan orang tua dari Mahasiswa paling banyak sebagai wiraswasta dan pegawai negeri sipil.

Keywords: *clustering; Davies Bouldin Index; k-medoids; strategi promosi*

Abstract

The increasing number of higher education institutions, both public and private, based on the 2019 Higher Education Statistics document (Kemenristekdikti, 2019) has triggered competition between these educational institutions to find prospective students. Muhammadiyah University of Jember as one of the leading universities in East Java, especially in Jember Regency, in terms of attracting prospective students, requires a good promotion strategy to get prospective students according to the ceiling. This research was conducted to obtain student profile information by processing new student data for 2020/2021. Data processing by means of clustering using the k-medoids algorithm on student data such as Study Program, City, Province, Parents' Occupation and Type of School. The result of the research is to get a low DBI value in cluster $k = 3$ so that it is determined as the optimal cluster. Most students are from the cities of Jember, Banyuwangi and Bondowoso, East Java province, the most chosen study program is Informatics Engineering and the work of parents of students is mostly as entrepreneurs and civil servants.

Keywords: *clustering; Davies Bouldin Index; k-medoids; promotion strategy*

1. PENDAHULUAN

Berdasarkan dokumen Statistik Pendidikan Tinggi tahun 2019 (Kemenristekdikti, 2019) yang diterbitkan oleh Setjen Pusdatin Iptek Dikti, Kemenristekdikti bahwa saat ini ada sekitar 615 Universitas, 114 Institut, 1.424 Sekolah Tinggi, 841 Akademi, 34 Akademi Komunitas dan 213 Politeknik baik Negeri maupun Swasta, yang berdampak terhadap persaingan antar Lembaga Pendidikan Tinggi untuk mendapatkan calon mahasiswa sesuai dengan pagu masing-masing Lembaga Pendidikan Tinggi tersebut sehingga dibutuhkan strategi promosi yang baik dan cerdas.

Universitas Muhammadiyah Jember sebagai salah satu perguruan tinggi unggulan di Jawa Timur khususnya di Kabupaten Jember dalam hal menjangkau calon mahasiswa juga melakukan strategi promosi melalui media seperti brosur, banner, billboard, website, youtube, instagram, tiktok, menyelenggarakan program beasiswa baik beasiswa dari pemerintah dan non-pemerintah.

Dengan begitu banyaknya strategi promosi yang telah dilakukan oleh Universitas Muhammadiyah Jember, tetapi masih kurang diminati oleh siswa lulusan sekolah menengah atas / sederajat berdasarkan informasi pendaftaran mahasiswa baru yang dibuka hingga beberapa gelombang setiap tahunnya.

Berdasarkan informasi tersebut peneliti ingin melakukan penelitian menggunakan teknik clustering untuk mengolah data mahasiswa dengan menggunakan metode *k-medoids*, dimana pada proses *clustering* (Nahdliyah et al., 2019) metode *k-medoids* menggunakan *medoid* sebagai pusat *cluster* serta cukup baik dalam mengcluster data yang mengandung pencilan.

Pada penelitian ini hasil yang ingin didapatkan adalah sebuah informasi berupa profil mahasiswa yang nantinya bisa dijadikan sebagai pendukung keputusan bagi tim penerimaan mahasiswa baru yang akan melaksanakan strategi promosi berikutnya agar lebih tepat sasaran.

2. TINJAUAN PUSTAKA

a. Penelitian Sebelumnya

Khoiriya Latifah (Latifah, 2018) dalam penelitiannya analisis dan penerapan algoritma C4.5 dalam data mining untuk menunjang strategi promosi prodi informatika UPGRIS menunjukkan bahwa variabel yang paling tinggi pengaruhnya terhadap hasil registrasi mahasiswa adalah Asal Sekolah dan Jenis Kelamin. Rata-rata berasal dari Semarang dengan jurusan SMU dari IPA dan yang berasal dari luar kota rata-rata berasal dari Batang dan Pati. Dari SMU jurusan IPS dan berjenis kelamin Laki-laki berasal dari Batang dan yang berjenis kelamin Perempuan berasal dari Pati. Akurasi dari pembentukan model ini adalah sebesar 89.33 % (Good Classification).

Erna Hudianti Pujiarini (Pujiarini, 2019) dalam penelitiannya analisa asosiasi untuk menentukan strategi promosi perguruan tinggi dengan algoritma apriori mengamati beberapa variabel yaitu daerah asal mahasiswa, jurusan mahasiswa dan jenis sekolah menengah atas. Hasil dari penelitian tersebut bahwa prodi Teknik Informatika banyak dipilih oleh calon mahasiswa yang berasal dari Sleman, Bantul, Yogya, Klaten dengan nilai support 6% sampai 13% dan diperoleh 16 aturan asosiatif untuk menunjukkan asosiasi pemilihan calon mahasiswa terhadap jurusan dengan daerah asal dan jenis SMA dengan nilai confidence tertinggi 24%.

Rony Setiawan (Rony, 2016) dalam penelitiannya penerapan data mining menggunakan algoritma K-Means *clustering* untuk menentukan strategi promosi mahasiswa baru dengan menggunakan banyak atribut sehingga menghasilkan jumlah kluster 4 ($k=4$) dengan *cluster* pertama 17007 calon mahasiswa, *cluster* kedua sebanyak 83 calon mahasiswa, *cluster* ketiga sebanyak 12919 calon mahasiswa dan *cluster* keempat sebanyak 356 calon mahasiswa.

Surya Darma (Darma & Nurcahyo, 2021) dalam penelitiannya klusterisasi teknik promosi dalam meningkatkan mutu kampus menggunakan algoritma *k-medoids* terhadap data para calon mahasiswa AMIK dan STIKOM Tunas Bangsa Pematangiantar dan hasil dari

penelitian adalah untuk mengetahui dari daerah mana para pelamar yang mendaftar di AMIK dan STIKOM Tunas Bangsa Pematangiantar dengan mengelompokkannya dalam dua klaster yaitu tertinggi dan terendah.

b. Data Mining

Disiplin ilmu (Wahono, 2020) yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar. Nama lain (Han et al., 2011) data mining adalah : *Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence*. Secara umum (Suyanto, 2019) kegunaan data mining dapat dibagi menjadi dua : deskriptif dan prediktif. Deskriptif berarti data mining digunakan untuk mencari pola-pola yang dapat dipahami manusia yang menjelaskan karakteristik data, sedangkan prediktif berarti data mining digunakan untuk membentuk sebuah model pengetahuan yang akan digunakan untuk melakukan prediksi.

c. Clustering

Clustering adalah proses pengelompokan himpunan data ke dalam beberapa grup atau cluster sehingga objek – objek dalam suatu cluster memiliki kemiripan yang tinggi (Kusuma & Agani, 2015), namun berbeda dengan objek – objek di cluster – cluster lainnya (J Han et al. 2012) dalam (Suyanto, 2019). Kemiripan (*similarities*) dan ketidakmiripan (*dissimilarities*) dihitung berdasarkan nilai – nilai atribut yang menggambarkan objek – objek tersebut dan seringkali melibatkan ukuran jarak.

d. Metode Berbasis Partisi

Sesuai dengan Namanya, metode ini bekerja dengan cara membagi atau mempartisi data ke dalam sejumlah kelompok. Metode ini dikenal juga dengan metode berbasis pusat atau metode berbasis representatif (Defiyanti et al., 2017) karena bekerja dengan menentukan pusat – pusat cluster yang bisa berupa rata – rata, modus atau sebuah objek representatif dari semua objek dalam suatu cluster berdasarkan ukuran tertentu.

e. Algoritma K-Medoids

Algoritma *K-medoids* (Suyanto, 2019) menggunakan teknik berbasis objek *representatif* (perwakilan) yang disebut *medoids*. *Medoids* adalah obyek yang dianggap mewakili cluster sekaligus sebagai pusat cluster. Algoritma *k-medoids* membentuk suatu cluster dengan cara menghitung jarak kemiripan yang dimiliki antara *medoids* dengan obyek *non-medoids*.

Langkah – langkah Algoritma *K-medoids* seperti dibawah ini :

1. Menentukan *k* sebagai banyaknya cluster yang akan dibentuk.
2. Memilih *k* obyek menjadi pusat cluster (*medoid*) secara acak.
3. Menghitung kemiripan antara obyek *medoid* dengan obyek *non-medoid* menggunakan jarak *euclidean* dengan persamaan 1 serta menempatkan obyek *non-medoid* ke *medoid* terdekat, kemudian hitung total jaraknya.

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2} \quad (1)$$

Keterangan:

d = jarak obyek

p = data

q = centroid / medoid

4. Memilih secara acak obyek non-medoid pada masing-masing cluster sebagai kandidat medoid baru.
5. Menghitung kemiripan antara obyek *medoid* dengan obyek *non-medoid* menggunakan jarak *euclidean* dengan persamaan 1 serta menempatkan obyek *non-medoid* ke *medoid* terdekat, kemudian hitung total jaraknya.
6. Menghitung selisih total jarak ($S_{\text{total jarak}}$), dengan $S_{\text{total jarak}} = \text{total jarak pada kandidat medoid baru} - \text{total jarak pada medoid lama}$.
7. Jika diperoleh nilai $S_{\text{total jarak}} < 0$, maka kandidat medoid baru tersebut menjadi medoid baru dan jika diperoleh $S_{\text{total jarak}} > 0$ iterasi berhenti.
8. Mengulangi langkah 4 sampai 7 hingga tidak terjadi perubahan medoid atau $S_{\text{total jarak}} > 0$.

f. Metode Davies Bouldin Index

Davies Bouldin Index (DBI) dalam (Dewi & Pramita, 2019) adalah metode validasi pada hasil *clustering*. Kohesi dan Separasi adalah indikator dari DBI yang merupakan pendekatan pengujian nilai. Kohesi menurut (Jumadi & USU, 2018) adalah jumlah kedekatan data terhadap titik pusat *cluster* dari *cluster* yang diikuti. Sedangkan separasi merupakan jarak antar titik pusat *cluster* terhadap *clusternya*. Sebuah *cluster* dikatakan optimal jika nilai kohesi rendah dan nilai separasi tinggi. Nilai DBI yang mendekati 0 menunjukkan bahwa *cluster* tersebut yang optimal.

Sum of square within cluster (SSW) merupakan persamaan untuk mengetahui matrik kohesi dalam sebuah *cluster* ke-*i*, persamaannya sebagai berikut :

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j - c_i) \quad (2)$$

Keterangan :

m_i = jumlah data dalam *cluster* ke-*i*

c_i = *centroid cluster* ke-*i*

$d(x_j - c_i)$ = jarak *euclidean* setiap data ke *centroid*

Sum of square between cluster (SSB) merupakan persamaan untuk mengetahui nilai separasi antar *cluster*, persamaannya sebagai berikut :

$$SSB_{i,j} = d(c_i, c_j) \quad (3)$$

Keterangan :

$d(c_i, c_j)$ = jarak antar *centroid*

Setelah memperoleh nilai kohesi dan separasi, langkah berikutnya adalah menghitung rasio ($R_{i,j}$) untuk mengetahui nilai perbandingan antara *cluster* ke-*i* dan *cluster* ke-*j*, persamaannya sebagai berikut :

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

Selanjutnya untuk menghitung nilai *Davies Bouldin Index* (DBI) menggunakan persamaan seperti berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (5)$$

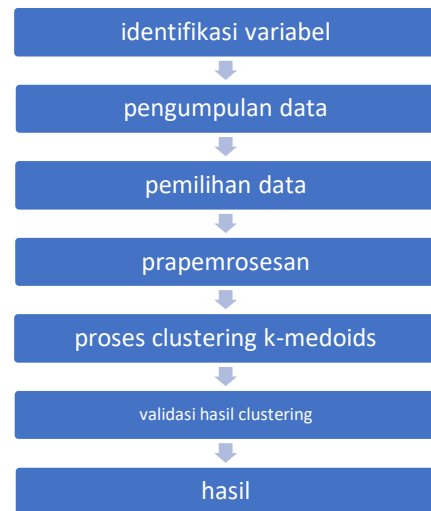
Keterangan :

k = jumlah *cluster* yang digunakan

Dengan memperoleh nilai DBI yang semakin rendah maka akan didapatkan kualitas *cluster* yang semakin baik dari hasil suatu *clustering* data.

3. METODE PENELITIAN

Rancangan tahapan penelitian dilaksanakan seperti tampilan pada gambar di bawah :



Gambar 1. Unjuk Kerja Penelitian
 Sumber : (Sofyan Vivi Dwiyanu, 2019)

a. Identifikasi variabel

Proses identifikasi variabel adalah mengidentifikasi atribut-atribut yang diperlukan

b. Pengumpulan data

Data didapatkan dari UPT. PDI Universitas Muhammadiyah Jember dengan format excel dengan jumlah data sebanyak 13280 baris.

c. Pemilihan data

Memilih atribut – atribut yang relevan pada data mahasiswa, atribut – atribut tersebut adalah : No., Semester, Nama Prodi, NIM, Nama Mahasiswa, Kota, Propinsi, pekerjaan ortu, Nama Sekolah

d. Prapemrosesan data

1. Pembersihan data

Pembersihan data (*Cleaning Data*) dilakukan dengan cara menghapus baris data yang memiliki nilai kosong dan nilai yang tidak sesuai serta data yang sama dikarenakan data 2 semester, dimana setelah dilakukan pembersihan data didapatkan data sejumlah 885 baris. Selanjutnya dilakukan penggantian atribut Nama Sekolah menjadi Jenis Sekolah dengan tujuan memudahkan pengelompokan.

2. Transformasi data

Transformasi data dilakukan untuk merubah data nominal menjadi data numerik dikarenakan pada proses *clustering* untuk pengukuran jarak data menggunakan metode *euclidean distance* yang membutuhkan data numerik.

e. Proses clustering k-medoids

Proses clustering dilakukan dengan jumlah cluster sebanyak $k = 2$, $k=3$, $k = 4$ dan $k = 5$, proses clustering dilakukan dengan menggunakan tool excel dan rapidminer, seperti contoh berikut :

1. Menggunakan excel

Pada proses *clustering* dengan algoritma *K-medoids* ditentukan cluster $k = 2$ kemudian menentukan *medoid* awal secara acak, seperti tabel berikut :

Tabel 1. Medoid Awal

Iterasi 1 dengan C medoid awal						
Medoid	Nama Prodi	Kota	Propinsi	Pekerja an Ortu	Jenis Sekolah	
O885	1	29	0	0	2	6
O612	2	5	0	0	2	2

Sumber : Hasil Perhitungan

Selanjutnya menghitung jarak masing-masing obyek *non-medoid* ke *medoid* awal serta menempatkan obyek *non-medoid* kedalam *cluster* yang paling dekat terhadap *medoid* kemudian menghitung total jarak, seperti perhitungan berikut :

Menggunakan rumus euclidean dengan persamaan 1

Cluster 1

$$\text{Data 1} = d(p_1, q_1) =$$

$$\sqrt{(0-29)^2 + (0-0)^2 + (0-0)^2 + (0-2)^2 + (0-6)^2} = 29,6816441593117$$

$$\text{Data 2} = d(p_2, q_1) =$$

$$\sqrt{(0-29)^2 + (1-0)^2 + (0-0)^2 + (1-2)^2 + (0-6)^2} = 29,6479341607472$$

$$\text{Data 3} = d(p_3, q_1) =$$

$$\sqrt{(0-29)^2 + (2-0)^2 + (0-0)^2 + (2-2)^2 + (1-6)^2} = 29,4957624075053$$

$$\text{Data 4} = d(p_4, q_1) =$$

$$\sqrt{(0-29)^2 + (0-0)^2 + (0-0)^2 + (0-2)^2 + (0-6)^2} = 29,6816441593117$$

Cluster 2

$$\text{Data 1} = d(p_1, q_2) =$$

$$\sqrt{(0-5)^2 + (0-0)^2 + (0-0)^2 + (0-2)^2 + (0-2)^2} = 5,74456264653803$$

$$\text{Data 2} = d(p_2, q_2) =$$

$$\sqrt{(0-5)^2 + (1-0)^2 + (0-0)^2 + (1-2)^2 + (0-2)^2} = 5,56776436283002$$

$$\text{Data 3} = d(p_3, q_2) =$$

$$\sqrt{(0-5)^2 + (2-0)^2 + (0-0)^2 + (2-2)^2 + (1-2)^2} = 5,47722557505166$$

$$\text{Data 4} = d(p_4, q_2) =$$

$$\sqrt{(0-5)^2 + (0-0)^2 + (0-0)^2 + (0-2)^2 + (0-2)^2} = 5,74456264653803$$

Untuk keseluruhan perhitungan dapat dilihat seperti pada tabel berikut :

Tabel 2. Menghitung Jarak Obyek ke Medoid Awal

No.	C1	C2	Kedekatan	Cluster
1	29,68164	5,744563	5,744562647	2
2	29,64793	5,567764	5,567764363	2
3	29,49576	5,477226	5,477225575	2
4	29,68164	5,744563	5,744562647	2
5	29,27456	5	5	2
.....
885	0	24,33105	0	1
jumlah kedekatan			318,1734891	
jumlah cluster 1				361
jumlah cluster 2				524
Total				885

Sumber : Hasil Perhitungan

Selanjutnya melakukan iterasi 2 dengan memilih secara acak obyek *non-medoid* pada masing-masing *cluster* sebagai kandidat *medoid* baru, seperti tabel berikut :

Tabel 3. Iterasi 2 dengan Medoid Baru

Iterasi 2 dengan C medoid awal					
Medoid	Nama Prodi	Kota	Propinsi	Pekerja an Ortu	Jenis Sekolah
O1	C1	0	0	0	0
O2	C2	0	1	0	1

Sumber : Hasil Perhitungan

Selanjutnya menghitung jarak masing-masing obyek *non-medoid* ke *medoid* awal serta menempatkan obyek *non-medoid* kedalam *cluster* yang paling dekat terhadap *medoid* kemudian menghitung total jarak, seperti perhitungan berikut :

Menggunakan rumus euclidean dengan persamaan 1

Cluster 1

$$\text{Data 1} = d(p_1, q_1) =$$

$$\sqrt{(0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2} = 0$$

$$\text{Data 2} = d(p_2, q_1) = \sqrt{(0-0)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2} = 1,4142135623731$$

$$\text{Data 3} = d(p_3, q_1) = \sqrt{(0-0)^2 + (2-0)^2 + (0-0)^2 + (2-0)^2 + (1-0)^2} = 3$$

$$\text{Data 4} = d(p_4, q_1) = \sqrt{(0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2} = 0$$

Cluster 2

$$\text{Data 1} = d(p_1, q_2) = \sqrt{(0-0)^2 + (0-1)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2} = 1,4142135623731$$

$$\text{Data 2} = d(p_2, q_2) = \sqrt{(0-0)^2 + (1-1)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2} = 0$$

$$\text{Data 3} = d(p_3, q_2) = \sqrt{(0-0)^2 + (2-1)^2 + (0-0)^2 + (2-1)^2 + (1-0)^2} = 1,73205080756888$$

$$\text{Data 4} = d(p_4, q_2) = \sqrt{(0-0)^2 + (0-1)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2} = 1,4142135623731$$

Untuk keseluruhan perhitungan dapat dilihat seperti pada tabel berikut :

Tabel 4. Menghitung Jarak Obyek ke Medoid Baru

No.	C1	C2	Kedekatan	Cluster
1	0	1,414214	0	1
2	1,414214	0	0	2
3	3	1,732051	1,732051	2
4	0	1,414214	0	1
5	2,828427	2,44949	2,44949	2
...
885	29,68164	29,64793	29,64793	2
jumlah kedekatan			1705,702	
			jumlah cluster 1	246
			jumlah cluster 2	639
			Total	885

Sumber : Hasil Perhitungan

Selanjutnya menghitung selisih kedekatan atau selisih total jarak dari 2 iterasi yang telah dilakukan, seperti berikut :

Tabel 5. Menghitung Selisih Kedekatan

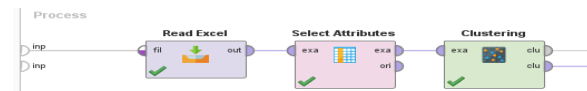
selisih kedekatan	
rumus = jumlah kedekatan iterasi ke-2 - jumlah kedekatan iterasi ke-1	
selisih kedekatan	1387,528786
jika selisih iterasi > 0, maka iterasi berhenti dan cluster ada pada iterasi sebelumnya	

Sumber : Hasil Perhitungan

Dikaraneakan $S > 0$ maka iterasi berhenti pada iterasi ke 2 dan hasil *cluster* yang digunakan adalah hasil perhitungan pada iterasi pertama.

2. Menggunakan rapidminer

Hasil *clustering* dengan menggunakan tool Rapidminer didapatkan *medoid* sebagai pusat *cluster* sebagai berikut :



Gambar 2. Operator yang Digunakan
 Sumber : Hasil Perhitungan

Selanjutnya menentukan parameter pada operator Clustering seperti berikut :

Gambar 3. Parameter pada Operator Clustering
 Sumber : Hasil Perhitungan

Setelah diklik tombol eksekusi maka didapatkan hasil clustering sebagai berikut :

Attribute	cluster_0	cluster_1
Nama Prodi	29	5
Kota	0	0
Propinsi	0	0
Pekerjaan Ortu	2	2
Jenis Sekolah	6	2

Gambar 4. Medoid Sebagai Pusat Cluster k = 2
 Sumber : Hasil Perhitungan

Sedangkan cluster model yang dihasilkan sebagai berikut :

Cluster Model

Cluster 0: 361 items
 Cluster 1: 524 items
 Total number of items: 885

Gambar 5. Cluster Model k = 2
 Sumber : Hasil Perhitungan

Selanjutnya merupakan tabel hasil clustering k-medoid dengan k = 2

Row No.	id	cluster	Nama Prodi	Kota	Propinsi	Pekerjaan O...	Jenis Sekolah
1	1	cluster_1	0	0	0	0	0
2	2	cluster_1	0	1	0	1	0
3	3	cluster_1	0	2	0	2	1
4	4	cluster_1	0	0	0	0	0
5	5	cluster_1	0	0	0	2	2
6	6	cluster_1	0	2	0	0	3
7	7	cluster_1	0	0	0	3	1
8	8	cluster_1	0	2	0	0	3
880	880	cluster_0	28	3	0	0	1
881	881	cluster_0	28	0	0	4	1
882	882	cluster_0	28	0	0	4	4
883	883	cluster_0	28	5	0	0	1
884	884	cluster_0	28	27	4	0	4
885	885	cluster_0	29	0	0	2	6

Gambar 5. Hasil clustering dengan rapidminer cluster k = 2
 Sumber : Hasil Perhitungan

f. Validasi Hasil Clustering

1. Menghitung kohesi / Sum of square within cluster (SSW)

Menggunakan medoid akhir pada proses clustering dengan cluster k = 2 seperti pada tabel berikut :

Tabel 6. Medoid Akhir

medoid	Nama Prodi	Kota	Propinsi	Pekerjaan Ortu	Jenis Sekolah	
						V
O885	C1	29	0	0	2	6
O612	C2	5	0	0	2	2

Sumber : Hasil Perhitungan

a. Menghitung SSW₁:

- Menghitung jarak data ke medoid dari setiap obyek menggunakan persamaan 1, sebagai berikut :

$$\text{Data 1} = d_1(x_1, c_1) = \sqrt{(17-29)^2 + (2-0)^2 + (0-0)^2 + (0-2)^2 + (4-6)^2} = 12,489996$$

$$\text{Data 2} = d_2(x_2, c_1) = \sqrt{(18-29)^2 + (0-0)^2 + (0-0)^2 + (3-2)^2 + (3-6)^2} = 11,44552314$$

$$\text{Data 3} = d_3(x_3, c_1) = \sqrt{(18-29)^2 + (4-0)^2 + (0-0)^2 + (0-2)^2 + (1-6)^2} = 12,88409873$$

$$\text{Data 4} = d_4(x_4, c_1) = \sqrt{(18-29)^2 + (0-0)^2 + (0-0)^2 + (5-2)^2 + (0-6)^2} = 12,88409873$$

- Menghitung seluruh jarak data ke medoid dengan cara sebagai berikut :

$$d = d_1 + d_2 + d_3 + \dots + d_n$$

$$d = 12,489996 + 11,44552314 + 12,88409873 + 12,88409873 + \dots + 0$$

$$\sum d = 3128,667613$$

- Menghitung Sum of square within cluster (SSW) dengan cara sebagai berikut :

$$SSW_1 = \frac{1}{361} \times 3128,667613 = 8,666669288$$

2. Menghitung Separasi / Sum of square between cluster (SSB)

Selanjutnya menghitung SSB menggunakan persamaan 3 dengan contoh perhitungan sebagai berikut :

$$SSB_{1,2} = d_1(c_1, c_2) = \sqrt{(29-5)^2 + (0-0)^2 + (0-0)^2 + (2-2)^2 + (6-2)^2} = 24,33105012$$

Hasil perhitungan SSB setiap cluster seperti tabel berikut :

Tabel 7. Hasil Perhitungan SSB Setiap Cluster

SSB	Medoid	
	1	2
1	0	24,33105012
2	24,33105012	0

Sumber : Hasil Perhitungan

3. Menghitung Rasio

Selanjutnya menghitung Rasio menggunakan persamaan 4 dengan contoh perhitungan sebagai berikut :

$$R_{1,2} = \frac{8,666669288 + 7,316998342}{24,33105012} = 0,656924693$$

Hasil perhitungan Rasio setiap cluster seperti tabel berikut :

Tabel 8. Hasil Perhitungan Rasio Setiap Cluster

R	data ke-i	
	1	2
1	0	0,656924693
2	0,656924693	0

Sumber : Hasil Perhitungan

Selanjutnya mencari rasio terbesar dari setiap cluster seperti ditunjukkan pada tabel berikut :

Tabel 9. Menghitung rasio terbesar dari setiap cluster

R	data ke-i		R Max
	1	2	
1	0	0,656924693	0,656924693
2	0,656924693	0	0,656924693

Sumber : Hasil Perhitungan

Berikutnya adalah menjumlah rasio terbesar dari setiap cluster dengan hasil sebagai berikut : $R \text{ Max} = 0,656924693 + 0,656924693 = 1,313849386$

4. Menghitung DBI

Selanjutnya menghitung DBI menggunakan persamaan 5 dengan menghitung rata-rata dari jumlah rasio terbesar, seperti perhitungan berikut :

$$DBI = \frac{1}{2} \times 1,313849386 = 0,656924693$$

Sehingga nilai DBI yang diperoleh dari cluster $k = 2$ adalah 0,656924693

4. HASIL DAN PEMBAHASAN

Dari keseluruhan perhitungan yang telah dilakukan untuk mencari nilai DBI pada setiap cluster diperoleh hasil sebagai berikut :

Tabel 10. Nilai DBI Tiap Cluster

cluster k =	DBI
2	0,656924693
3	0,634946676
4	1,278481638
5	1,514504045

Sumber : Hasil Perhitungan

Sehingga nilai DBI terkecil didapatkan dari cluster $k = 3 = 0,634946676$

Berdasarkan dari nilai DBI yang telah didapatkan bahwa cluster $k = 3$ adalah cluster yang optimal, sehingga informasi profil mahasiswa dari cluster $k = 3$ adalah sebagai berikut :

Tabel 11. Informasi Profil Mahasiswa

cluster 1	
1.	Prodi yang banyak dipilih adalah Teknik Informatika 33% dan Teknik Sipil 28%
2.	Mahasiswa terbanyak dari kota Pamekasan 11,11 % dan Denpasar 11,11%
3.	Mahasiswa terbanyak dari Propinsi Jawa Timur 44 % dan Bali 22,22%
4.	Pekerjaan Orang Tua Mahasiswa terbanyak adalah Wiraswasta 55,56% dan PNS 16,67%
5.	Asal sekolah Mahasiswa yang terbanyak adalah SMA Negeri 50 % dan MA Swasta 22,22 %
cluster 2	
1.	Prodi yang banyak dipilih adalah Manajemen 28,57% dan Hukum 11,81%
2.	Mahasiswa terbanyak dari kota Jember 49,52%, Banyuwangi 15,24% dan Bondowoso 12,76%
3.	Mahasiswa terbanyak dari Propinsi Jawa Timur 99,05%
4.	Pekerjaan Orang Tua Mahasiswa terbanyak adalah Wiraswasta 43,62% dan PNS 18,1%
5.	Asal sekolah Mahasiswa yang terbanyak adalah SMA Negeri 57,71% dan SMA Swasta 17,71%
cluster 3	
1.	Prodi yang banyak dipilih adalah Teknik Informatika 29,53% dan S1 - Ilmu Keperawatan 23,98%
2.	Mahasiswa terbanyak dari kota Jember 59,06%, Banyuwangi 11,99% dan Bondowoso 10,82%
3.	Mahasiswa terbanyak dari Propinsi Jawa Timur 99,71%
4.	Pekerjaan Orang Tua Mahasiswa terbanyak adalah Wiraswasta 43,57% dan PNS 19,30%
5.	Asal sekolah Mahasiswa yang terbanyak adalah SMA Negeri 44,44% dan SMA Swasta 18,13%

Sumber : Hasil Perhitungan

5. KESIMPULAN

1. Dari hasil validasi DBI dipilih cluster $k = 3$ karena yang paling optimal.
2. Informasi Profil Mahasiswa yang diperoleh dari cluster $k = 3$ adalah Mahasiswa banyak yang memilih Prodi Teknik Informatika, kota asal Mahasiswa yang paling banyak dari Jember, Banyuwangi dan Bondowoso, propinsi Jawa Timur, untuk Pekerjaan Orang tua

yang paling banyak adalah Wiraswasta dan PNS sedangkan jenis sekolah dari Mahasiswa yang paling banyak dari SMA Negeri dan SMA Swasta.

6. SARAN

1. Untuk penelitian selanjutnya bisa menggunakan metode validasi selain DBI, seperti Silhouette Index atau C-Index atau metode validasi yang penulis belum ketahui.

7. REFERENSI

- Darma, S., & Nurcahyo, G. W. (2021). Klasterisasi Teknik Promosi dalam Meningkatkan Mutu Kampus Menggunakan Algoritma K-Medoids. *Jurnal Informatika Ekonomi Bisnis*, 3, 89–94. <https://doi.org/10.37034/infeb.v3i3.87>
- Defiyanti, S., Jajuli, M., & Rohmawati, N. (2017). Optimalisasi K-MEDOID dalam Pengklasteran Mahasiswa Pelamar Beasiswa dengan CUBIC CLUSTERING CRITERION. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 3(1), 211–218. <https://doi.org/10.25077/teknosi.v3i1.2017.211-218>
- Dewi, D. A. I. C., & Pramita, D. A. K. (2019). Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali. *Matrix: Jurnal Manajemen Teknologi Dan Informatika*, 9(3), 102–109. <https://doi.org/10.31940/matrix.v9i3.1662>
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*, 3rd ed. (3rd ed.). Morgan Kaufmann.
- Jumadi, B., & USU. (2018). *Tesis: peningkatan hasil evaluasi clustering davies bouldin dengan penentuan titik pusat cluster awal K means*.
- Kemenristekdikti. (2019). Statistik Pendidikan Tinggi (Higher Education Statistics) 2019. In *Pusdatin Kemenristekdikti*. http://www.mohe.gov.my/web_statistik/
- Kusuma, D. T., & Agani, N. (2015). Prototipe Komparasi Model Clustering Menggunakan Metode K-Means Dan FCM untuk Menentukan Strategi Promosi : Study Kasus Sekolah Tinggi Teknik-PLN Jakarta. *TICOM (Technology of Information and Communication)*, 3(3), 1–10.
- Latifah, K. (2018). *ANALISIS DAN PENERAPAN ALGORITMA C4.5 DALAM DATA MINING UNTUK MENUNJANG STRATEGI PROMOSI PRODI INFORMATIKA UPGRIS*. 11(61).
- Nahdliyah, M. A., Widiharih, T., & Prahutama, A. (2019). Metode K-Medoids Clustering dengan Validasi Silhouette Index dan C-Index. *Jurnal Gaussian*, 8(2), 161–170.
- Pujiarini, E. H. (2019). *Analisis asosiasi untuk menentukan strategi promosi perguruan tinggi dengan algoritma apriori*. 4, 45–51.
- Rony, S. (2016). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru (Studi Kasus : Politeknik Lp3i Jakarta). *Jurnal Lentera Ict*, 3(1), 76–92.
- Sofyan Vivi Dwiyanu. (2019). *KLASIFIKASI KELULUSAN MAHASISWA PROGRAM STUDI TEKNIK INFORMATIKA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR*.
- Suyanto. (2019). *Data Mining Untuk Klasifikasi dan Klasterisasi Data* (Suyanto (ed.); Pertama). informatika Bandung.
- Wahono, R. S. (2020). *Data Mining*. 721. romi@romisatriawahono.net