JIIFKOM (Jurnal Ilmiah Informatika & Komputer) STTR Cepu
E-Mail: jiifkom@sttrcepu.ac.id

Clustering Menggunakan Algoritma *K-Medoids* Untuk Menentukan Strategi Promosi Sekolah Tinggi Teknologi Ronggolawe Cepu

Muhammad Abdimas Khalifuddin [1], Retno Wahyusari [2*]
[1]mabdimaskhalifuddin@gmail.com, *[2]retnowahyusari@gmail.com
Sekolah Tinggi TeknologiRonggolawe Cepu, Indonesia

Intisari

Upaya menciptakan strategi manajemen pemasaran yang efektif dan efisien, perlu pemahaman mendetail dan objektif mengenai pasar tempat mereka beroperasi. Dalam menganalisis permasalahan ini, bidang manajemen pemasaran sering tumpang-tindih bidang perencanaan strategis. Strategi pemasaran terdiri dari pengambilan keputusan tentang biaya pemasaran dari perusahaan, bauran pemasaran, dan lokasi pemasaran. Manajemen pemasaran harus memutuskan biaya yang perlu dikeluarkan untuk pemasaran dan bagaimana mengalokasikan seluruh anggaran biaya pemasaran untuk berbagai alat dalam bauran pemasaran. Data mahasiswa baru di Cluster dengan metode Algoritma K-medoids dengan bantuan Rapid minner, serta mengetahui tingkat keeratan dengan *Davies Bouldin Index* (DBI). Hasil dari penelitian ini adalah dari 118 data, dengan menggukan *2 cluster* menghasilkan *0 cluster* sebanyak 72 dan *1 cluster* sebanyak 46 dan *3 cluster* menghasilkan *0 cluster* mendapat 72 anggota, *1 cluster* dengan 3 anggota dan *2 cluster* mendapatkan 43 anggota. Nilai DBI dengan *2 cluster* sebesar 1.04 dan *3 cluster* sebersar 1.01.

Kata kunci:
Davies Bouldin Index (DBI), K-medoid, Promosi

Abstract

Efforts to create an effective and efficient marketing management strategy require a detailed and objective understanding of the market in which they operate. In analyzing this problem, the field of marketing management often overlaps the field of strategic planning. Marketing strategy consists of making decisions about the company's marketing costs, marketing mix, and marketing location. Marketing management must decide what costs need to be spent on marketing and how to allocate the entire marketing budget to various tools in the marketing mix. New student data in the Cluster using the K-medoids Algorithm method with the help of Rapid Minner, as well as knowing the level of correlation with the Davies Bouldin Index (DBI). The results of this research are from 118 data, using 2 clusters produces 0 clusters of 72 and 1 cluster of 46 and 3 clusters produce 0 clusters with 72 members, 1 cluster with 3 members and 2 clusters with 43 members. The DBI value with 2 clusters is 1.04 and 3 clusters is 1.01.

Tentang naskah:
-di terima, 8 Jan 2024
-diterbitkan, 31 Jan 2024

Keyword:
Davies Bouldin Index (DBI), K-medoid, Promotion

1. Pendahuluan

Menurut Kotler mengatakan bahwa promosi memiliki peran utama yang mengandakan komunikasi yang bersifat membujuk. Strategi yang digunakan Kotler & Armstrong variable yang memiliki 5 variabel promotional mix, yaitu Periklanan (*advertising*), Penjualan Personal (*personal selling*), Promosi penjualan (*sales promotion*), Hubungan masyarakat (*public relation*), Pemasaran langsung (*direct marketing*) (Kotler dan Armstrong, 2001).

Kegiatan promosi merupakan ujung tombak dari setiap kegiatan pemasaran, sebab promosi mempengaruhi target pasar dengan mengkomunikaskan suatu produk kepada target pasar. Kegiatan yang berhubungan dengan promosi salah satunya adalah strategi promosi kampus untuk mendapatkan mahasiswa baru. STT Ronggolawe Cepu merupakan salah satu perguruan tinggi yang melakukan kegiatan promosi tersebut. Disisi lain berdasarkan hasil wawancara dengan ketua PMB STTR tahun 2021 menyatakan pada kegiatan promosi yang dilakukan oleh STT Ronggolawe Cepu menyebabkan pembengkakan biaya, akibat pemasangan spanduk yang tidak tepat sasaran. Pembengkakan terjadi

hingga 50% dari biaya pengajuan. Pembengkakan biaya terjadi adanya penambahan jumlah pemasangan spanduk dilokasi strategis, terjadinya kerusakan pada spanduk dan adanya pajak yang tidak sesuai dengan teknis dilapangan. Pemasangan dilakukan daerah Cepu, Padangan, Ngraho, Blora, Jiken, Tuban dan Bojonegoro. Dengan adanya hal tersebut, perlu adanya metode yang efektif dan efisian dalam pemilihan kegiatan promosi. Salah satu metode yang dapat digunakan berupa klustering. Algoritma K-Medoids merupakan salah satu algoritma yang memiliki kelebihan untuk mengatasi kelemahan pada pada algoritma K-Means yang sensitive terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang pada dari distribusi data. Kelebihan lainnya yaitu hasil proses clustering tidak bergantung pada urutan masuk dataset (Furqon, dkk, 2015).

Beberapa penelitian terdahulu yang berkaitan dengan algoritma K-Medoids diantaranya: memilih teknik promosi berdasarkan data mahasiswa baru dengan 399 dataset ke dalam 2 (dua) klaster dengan atribut kota/kabupaten, nama prodi, sumber informasi, asal sekolah dan status kelas. Menentukan strategi promosi mahasiswa baru tahun 2017/2018 dan data sosialisasi ke dalam 3 (tiga) klaster yang bertujuan untuk menentukan prioritas sekolah-sekolah berpotensi dengan atribut sekolah, jenjang, jumlah siswa, jarak (Km), dan promosi kampus. Menentukan strategi promosi berdasarkan data mahasiswa dengan 885 dataset ke dalam 2 (dua) klaster dengan atribut semester, nama prodi, NIM, nama mahasiswa, kota, proponsi, pekerjaan ortu, dan nama sekolah.

Berdasarkan latar belakang yang telah dipaparkan, maka penulis akan menerapkan Algoritma K-medoids dengan dataset mahasiswa baru dengan di *Clustering* 2 atau 3 untuk memperkecil biaya promosi di suatu daerah. Hasil pengolahan data dengan algoritma K-Medoids digunakan untuk menemukan pola strategi promosi.

2. Kerangka Teori

2.1. Penelitian Sebelumnya

Penelitian yang dilakukan (Minarni, Elsa Indah Sari, Anna Syahrani, Putri Mandarani, 2021), mendapatkan hasil bahwa 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.

Penelitian (Daffa Rafif Agustian, Budi Arif Dermawan., 2022) menyimpulkan bahwa hasil perhitungan clustering pada dataset Penyakit DBD Tahun 2020 menghasilkan 2 *cluster* yaitu *cluster* 1 dan *cluster* 2. *Cluster* 1 memiliki 6 daerah dikategorikan sebagai daerah yang memiliki kasus penyakit DBD dalam tingkatan tinggi diantaranya Karawang Barat, Telukjambe Timur, Klari, Cikampek, Kotabaru, Karawang Timur. Pada *cluster* 2 memiliki 22 daerah, dan dikategorikan sebagai daerah yang memiliki kasus penyakit diare dalam tingkatan rendah di Kabupaten Karawang diantaranya Pangkalan, Ciampel, Rengasdengklok, Kutawaluya, Batujaya, Pedes, Pakisjaya, Jatisari, Cilamaya Wetan, Tirtamulya, Telagasari, Rawamerta, Lemahabang, Tempuran, Majalaya, Jayakerta, Cilamaya Kulon, Banyusari, Telukjambe Barat, Tegalwaru, Purwasari, Cilebar. Hasil clustering dievaluasi menggunakan *Silhouette Coefficient* untuk mengetahui kualitas cluster yang dihasilkan berdasarkan proses pengujian dari dataset Penyakit DBD di Kabupaten karawang Tahun 2020 menghasilkan nilai *Silhouette Coefficient* 0.78793, hal tersebut menunjukan bahwa jumlah cluster 2 merupakan *cluster* yang memiliki kualitas yang baik karena mendekati angka 1. Hal tersebut dalam analisis pengelompokan data daerah yang memiliki kasus penyakit DBD di Kabupaten Karawang lebih baik menggunakan *cluster* 2, se - hingga data yang dihasilkan dari penerapan algoritma K-Medoids akan mendekati ketepatan dalam pengelompokan daerah yang memiliki kasus penyakit DBD di Kabupaten Karawang.

Penelitian selanjutnya yang dilakukan oleh Surya Darma menggunakan algorima K-Medoids dapat mengetahui hasil *cluster* yang diperoleh dari iterasi terakhir. Diperoleh cluster rendah terdapat 376 data dan cluster

tinggi 23 data. Dengan dilakukan analisis terhadap data para mahasiswa baru yang mendaftar di AMIK dan STIKOM Tunas Bangsa Pematangsiantar dengan algoritma K-Medoids dapat diketahui bahwa sumber informasi dari alumi merupakan pendaftar terbanyak pada tahun ajaran 2020/2021dan diikutin berasal dari brosur, website, radio, koran, melihat baliho, dan asal sekolah. Sedangkan bila dilihat dari asal kota maka kabupaten simalungun sebagai pendaftar terbanyak dengan memilih program studi sistem informasi (Surya Darma, 2021).

2.2. Data Mining

Data mining merupakan sebuah inti dari proses KDD, meliputi dugaan algoritma yang mengeksplor data, membangun model dan menemukan pola yang belum diketahui. KDD bersifat otomatis, dapat diartikan sebagai pengorganisasian proses untuk pengidentifikasian yang benar, berguna dan penemuan pola dari kumpulan data yang besar dan komplek (Gambbeta Windy.,2012).

2.3. Clustering

Clustering merupakan suatu proses pengelompokkan *record* suatu, observasi, atau mengelompokkan kelas yang mempunyai kesamaan objek. Perbedaan clustering dengan klasifikasi yaitu tidak adanya variabel target dalam melakukan suatu pengelompokan pada proses clustering. *Clustering* sering dilakukan sebagai langkah awal dalam proses data mining saat melakukkan suatu metode analisis. (Kaur, dkk., 2014).

2.4. Algoritma K-Medoids

Menurut Algoritma *Partition Around Medoids* (PAM) atau dikenal juga dengan K-medoids dikembangkan oleh Leonard Kaufman dan Peter J. Rousseeuw pada tahun 1987. Algoritma PAM termasuk metode partitioning clustering untuk mengelompokkan sekumpulan n objek menjadi sekumpulan k *cluster.* Representasi cluster pada PAM adalah objek dari sekumpulan objek yang mewakili *cluster*, disebut *medoid* (Darma, S.,2021).

Beberapa langkah-langkah dalam perhitungan algoritma K-Medoids sebagai berikut:

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.
   $(p, q) = \sqrt{(p1 - q1)\ 2 + (p2 - q2)\ 2 + (p3 - q3)2}$ .....(2.1)
   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 (Stotal jarak), dengan Stotal jarak = total jarak pada kandidat medoid baru – total jarak pada *medoid* lama .
7. Jika diperoleh nilai Stotal jarak < 0, maka kandidat *medoid* baru tersebut menjadi *medoid* baru dan jika diperoleh Stotal jarak > 0 iterasi berhenti.
8. Mengulangi langkah 4 sampai 7 hingga tidak terjadi perubahan medoid atau Stotal jarak > 0.

2.5. RapidMiner

Rapidminer merupakan sebuah solusi untuk melakukan analisis terhadap data *mining*, *text mining* dan analisis prediksi. Rapidminer memakai banyak sekali teknik naratif serta prediksi dalam menyampaikan wawasan pada pengguna sehingga bisa membuat keputusan yang paling baik. Rapidminer menempati peringkat pertama sebagai software data mining pada polling oleh KDnuggets, sebuah portal data mining pada 2010-2011 (Aprilia C et al., 2013).

2.6. Metode Device Boulding 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 *cluster*nya. Sebuah *cluster* dikatakan optimal jika nilai kohesi rendah dan nilai separasi tinggi. Nilai DBI yang mendekati 0 menunjukkan bahwa *cluster* tersebut yang optimal (Jumadi & USU, 2018).

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

$$SSW_i = \frac{1}{m_i} \sum_{m=1}^{m_i} d(x_j + c_i) \text{.....(2)}$$

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:

$$SSB_{i,j} = \text{d}(x_j - c_i) \text{.....(3)}$$

Keterangan:
d(x_j –c_i) = jarak antar *centroid*

Setelah memperoleh nilai kohesi dan separasi, langkah berikutnya adalah menghitung rasio (Ri,j) untuk mengetahui nilai perbandingan antara cluster ke-i dan cluster ke-j, persamannya sebagai berikut :

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

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

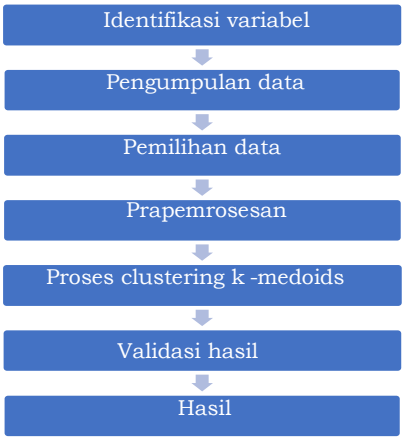
$$DBI = \frac{1}{k} \sum_{i=1}^{k} max_{i \neq j} (R_{i,j}) \text{.....(5)}$$

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. Metodologi

Rancangan tahapan penelitian dilaksanakan seperti tampilan pada gambar di bawah:



Gambar 3.1 Unjuk Kerja Penelitian

a. Identifikasi variabel

Proses identifikasi variabel adalah mengidentifikasi atribut-atribut yang diperlukan

b. Pengumpulan data

Data didapatkan dari Penerimaan mahasiswa baru angkatan 2022-2023 Sekolah Tinggi Ronggolawe Cepu dengan format excel dengan jumlah data sebanyak 118 baris.

c. Pemilihan data

Memilih atribut – atribut yang relevan pada data mahasiswa, atribut – atribut tersebut adalah Kelas, Asal Pendidikan, Kecamatan, Jurusan, Pekerjaan Orang tua, Gaji.

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.

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 dan k=3, proses *clustering* dilakukan dengan menggunakan tool excel dan rapidminer, seperti contoh berikut :

1. Meggunakan excel

Pada proses clustering dengan algorima Kmedoids ditentukan cluster k = 2 kemudian menentukan medoid awal secara acak, seperti tabel berikut :

Tabel 1. Medoid Awal

Iterasi 1 Dengan Medoid awal						
No	X_1	X_2	X_3	X_4	X_5	X_6
1	1	1	1	5	4	1
2	1	2	2	5	4	2

X_1= Kelas
X_2= Asal Pendidikan
X_3= Kecamatan
X_4= Jurusan
X_5= Pekerjaan Orang tua
X_6= Gaji
Sumber : Hasil Perhitungan

Selanjutnya menghitung jarak 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

Data 1 = $(p1, q1)$ = $\sqrt{(1-1)2 + (3-1)2 + (38-1)2 + (1-5)2 + (3-4)2 + (2-1)2}$ = 37.30

Data 2 = $(p2, q1)$ = $\sqrt{(1-2)2 + (3-2)2 + (38-2)2 + (1-5)2 + (3-4)2 + (2-2)2}$ = 36.25

Data 3 = $(p3, q1)$ = $\sqrt{(1-3)2 + (3-3)2 + (38-3)2 + (1-5)2 + (3-7)2 + (2-1)2}$ = 35.47

Cluster 2

Data 1 = $(p1, q1)$ = $\sqrt{(1-1)2 + (41-1)2 + (10-1)2 + (1-5)2 + (8-4)2 + (1-1)2}$ = 41.39

Data 2 = $(p2, q1)$ = $\sqrt{(1-2)2 + (41-2)2 + (10-2)2 + (1-5)2 + (8-4)2 + (1-2)2}$ = 40.22

Data 3 = $(p3, q1)$ = $\sqrt{(1-3)2 + (41-3)2 + (10-3)2 + (1-5)2 + (8-7)2 + (1-1)2}$ = 38.86

Untuk keseluruhan perhitungan dapat dilihat seperti pada tabel berikut:

Tabel 2. Menghitung Jarak Obyek ke Medoid Awal

No	C 1	C2 2	kedekatan	Cluster
1	37.30	41.39	37.30	1
2	36.25	40.22	36.25	1
3	35.47	38.86	35.47	1
4	35.27	38.08	35.27	1
5	34.29	37.07	34.29	1
.....	.....	.....	.....	.....
118	47.48	0.00	0.00	2
Jumlah kedekatan			2791.62	
			Jumlah Cluster 1	46
			Jumlah Cluster 2	72
			Total	118

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 1 Dengan Medoid awal						
No	X_1	X_2	X_3	X_4	X_5	X_6
1	1	1	1	5	4	1
2	1	2	2	5	4	2

X_1= Kelas
X_2= Asal Pendidikan
X_3= Kecamatan
X_4= Jurusan
X_5= Pekerjaan Orang tua
X_6= Gaji
Sumber: Hasil Perhitungan

Selanjutnya menghitung jarak masingmasing 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

Data 1 = $(p1, q1)$ = $\sqrt{(1-1)2 + (40-1)2 + (39-1)2 + (1-5)2 + (8-4)2 + (1-1)2}$ = 54.74

Data 2 = $(p2, q1)$ = $\sqrt{(1-2)2 + (40-2)2 + (39-2)2 + (1-5)2 + (8-4)2 + (1-2)2}$ = 53.35

Data 3 = $(p3, q1)$ = $\sqrt{(1-3)2 + (40-3)2 + (39-3)2 + (1-5)2 + (8-7)2 + (1-1)2}$ = 51.79

Cluster 2

Data 1 = $(p1, q1)$ = $\sqrt{(1-1)2 + (41-1)2 + (10-1)2 + (1-5)2 + (8-4)2 + (1-1)2}$ = 41.39

Data 2 = $(p2, q1)$ = $\sqrt{(1-2)2 + (41-2)2 + (10-2)2 + (1-5)2 + (8-4)2 + (1-2)2}$ = 40.22

Data 3 = $(p3, q1)$ = $\sqrt{(1-3)2 + (41-3)2 + (10-3)2 + (1-5)2 + (8-7)2 + (1-1)2}$ = 38.86

Untuk keseluruhan perhitungan dapat dilihat seperti pada tabel berikut:

Tabel 4. Menghitung Jarak Obyek ke Medoid Baru

No	C 1	C2 2	kedekatan	Cluster
1	54.74	41.39	41.39	2
2	53.35	40.22	40.22	2
3	51.79	38.86	38.86	2
4	51.22	38.08	38.08	2
5	49.92	37.07	37.07	2
.....	.....	.....	.....	.....
118	29.02	0	0	2
Jumlah kedekatan			3026.78	
		Jumlah Cluster 1		17
		Jumlah Cluster 2		101
			Total	118

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

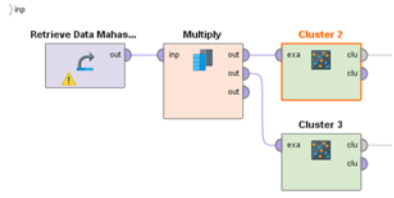
Tabel 5. Menghitung Selisih Kedekatan

selisih kedekatan	
rumus = jumlah kedekatan iterasi ke-2 - jumlah kedekatan iterasi ke-1	
selisih kedekatan	235.16
jika selisih iterasi > 0, maka iterasi berhenti dan cluster ada pada iterasi sebelumnya	

Dikaranekan 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:



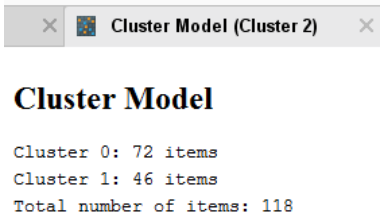
Gambar 3. Parameter pada Operator *Clustering*
Sumber: Hasil Perhitungan

Setelah diklik tombol eksekusi maka didapatkan hasil clustering sebagai berikut:

Attribute	cluster_0	cluster_1
Kelas	1	1
Asal Pendidikan	41	3
Kecamatan	10	38
Jurusan	1	1
pekerjaan	8	3
gaji	1	2

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

Sedangkan *cluster* model yang dihasilkan sebagai berikut:



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

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

Row No.	id	cluster	Kelas	Asal Pendidikan	Kecamatan	Jurusan	pekerjaan	gaji
1	1	cluster_1	1	1	1	5	4	1
2	2	cluster_1	1	2	2	5	4	2
3	3	cluster_1	1	3	3	5	7	1
4	4	cluster_1	1	4	3	5	4	1
5	5	cluster_1	1	5	4	5	3	2
6	6	cluster_1	1	6	5	5	7	1
114	114	cluster_0	1	48	1	1	1	1
115	115	cluster_0	1	40	13	1	8	1
116	116	cluster_0	1	40	13	1	7	1
117	117	cluster_0	1	40	39	1	8	1
118	118	cluster_0	1	41	10	1	8	1

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

No	Kelas	Asal Pendidikan	Kec	Jur	kerja	gaji
113	1	3	38	1	3	2
118	1	41	10	1	8	1

Sumber: Hasil Perhitungan

a. Menghitung SSW_1:

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

Data 1 = $(p1, q1)$ = $\sqrt{(1-1)2 + (3-1)2 + (38-1)2 + (1-5)2 + (3-4)2 + (2-1)2}$ = 37.30

Data 2 = $(p2, q1)$ = $\sqrt{(1-2)2 + (3-2)2 + (38-2)2 + (1-5)2 + (3-4)2 + (2-2)2}$ = 36.25

Data 3 = $(p3, q1$ = $\sqrt{(1-3)2 + (3-3)2 + (38-3)2 + (1-5)2 + (3-7)2 + (2-1)2}$ = 35.47

2. Menghitung seluruh jarak data ke *medoid* dengan cara sebagai berikut:
   d = d1+d2+d3….. dn
   d= 37.30 + 36.25 + 35.47 + 35.27 + ….. + 0
   ∑d = 1333.27

3. Menghitung *Sum of square within cluster* (SSW) dengan cara sebagai berikut:
   $SSW_1 = \frac{1}{46} \times 1333.27 = 28.98$

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

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

$SSB_{1,2} = d1\ (c1, c2\ ) = \sqrt{(1-2)2 + (3-2)2 + (38-2)2 + (1-5)2 + (3-4)2 + (2-2)2}$ = 36.25

Hasil perhitungan SSB setiap *cluster* seperti tabel berikut:

Tabel 7. Hasil Perhitungan SSB Setiap Cluster

SSB	*Medoid*	
	1	2
1	0	47.48
2	47.48	0

Sumber: Hasil Perhitungan

3) Menghitung Rasio

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

$R_{1,2} = R_{1,2} = \frac{28,29 + 20,27}{47,48}$

$R_{1,2}$ = 1,04

Hasil perhitungan Rasio setiap cluster seperti tabel berikut:

Tabel 8. Hasil Perhitungan Rasio Setiap Cluster

R	Data ke-i	
	1	2
1	0	1.04
2	1.04	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
	1	2	MAX
1	0	1.04	1.04
2	1.04	0	1.04

Sumber: Hasil Perhitungan

Berikutnya adalah menjumlah rasio terbesar dari setiap *cluster* dengan hasil sebagai berikut:

R Max = 1.04 + 1.04 = 2.08

4) Menghitung DBI

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

DBI = $\frac{1}{2} \times 2.08 = 1.04$

Sehingga nilai DBI yang diperoleh dari *cluster* k = 2 adalah 1.04

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	1.04

3	1.01

Sumber: Hasil Perhitungan

Sehingga nilai DBI terkecil didapatkan dari *cluster* k = 3 = 1.01

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	Mahasiswa terbanyak dari Kecamatan Cepu 33,33% dan Sambong 12,50%
2	Prodi yang di pilih adalah Sipil 31.94% , Mesin D3 23.26% , Informatika 19.44% , Elektro 13,89% dan Mesin 11.11%
3	Pekerjaan Orang Tua Mahasiswa terbanyak Wiraswata 27.78% dan Petani 23.61%
4	Gaji orang tua terbanyak adalah kurang dari 2 juta 69.44% dan 2-5 juta 29.17%
	Cluster 2
1	Mahasiswa terbanyak dari kecamatan Biting 33.33% dan JatiBiting 33.33%
2	Prodi terbanyak adalah Mesin 100%
3	Pekerjaan Orang Tua Mahasiswa terbanyak Wiraswasta 33.33% dan Lainnya 33.33%
4	Gaji orang tua terbanyak adalah kurang dari 2 juta 66.67% dan 2-5 juta 33.33%
	Cluster 3
1	Mahasiswa terbanyak dari kecamatan Blora 16.28 % dan Kasiman 11.63%
2	Prodi terbanyak adalah Sipil 53.49% dan Mesin 46.51%
3	Pekerjaan Orang Tua Mahasiswa terbanyak Wiraswasta 23.26%, Petani 23.26% dan Lainnya 23.26%
4	Gaji orang tua terbanyak adalah kurang dari 2 juta 86.05% dan 2-5 juta 11.63%

Sumber : Hasil Perhitungan

5) Simpulan

Adapun kesimpulan dalam penelitian ini adalah sebagai berikut:

1. Pengaplikasian perhitungan algoritma K-medoids baik menggunakan perhitungan secara manual maupun aplikasi dari 118 data, dengan menggukan *2 cluster* menghasilkan *0 cluster* sebanyak 72 dan *1 cluster* sebanayak 46 dan *3 cluster* menghasilakan *0 cluster* mendapat 72 anggota, *1 cluster* dengan 3 anggota dan *2 cluster* mendapatkan 43 anggota.
2. Nilai DBI dengan *2 cluster* sebesar 1.04 dan *3 cluster* sebersar 1.01.
3. Hasil yang digunakan untuk mengetahui kecamatan, pekerjaan, gaji dan jurusan adalah hasil *3 cluster*. Berdasarkan Jurusan terbanyak adalah Teknik Mesin dan Teknik Sipil. Kecamatan yang banyak mahasiswa berasal dari Kecamaam Cepu dan *Biting*. Pekerjaan orang tua paling banyak Wiraswata dan Lainnya. Serta Gaji paling banyak kurang dari 2 juta dan antara 2-5 Juta

Daftar Pustaka

Artikel jurnal:

Agustian, D. R., & Darmawan, B. A. (2022). Analisis Clustering Demam Berdarah Dengue Dengan Algoritma K-Medoids (Studi Kasus Kabupaten Karawang). *Jiko (Jurnal Informatika dan Komputer)*, *6*(1), 18-26.

Darma, S. (2021). Klasterisasi Teknik Promosi dalam Meningkatkan Mutu Kampus Menggunakan Algoritma K-Medoids. *Jurnal Informatika Ekonomi Bisnis*, 89-94.

Dekaprasetya, R., & Muharom, L. A. (2022). Clustering Untuk Menentukan Strategi Promosi Universitas Muhammadiyah Jember Dengan Algoritma K-Medoids. *Jurnal Smart Teknologi*, *3*(3), 312-320.

Harist N, A., Munthe, I. R., & Juledi, A. P. (2021). Implementasi Data Mining Algoritma Apriori untuk Meningkatkan Penjualan. *Jurnal Teknik Informatika UNIKA Santo Thomas*, *06*, 188–197. https://doi.org/10.54367/jtiust.v6i1.1276

Iriadi, N., Priatno, & Ishaq, A. (2020). *Penerapan Data Mining Dengan Rapid Miner* (Pertama). Graha Ilmu.

Kaur, Noor K., Kaur, Usvir. & Singh, Dr. Dheerendra. (2014). K-Medoid Clustering Algorithm- A Review, International Journal of Computer Application and Technology.

Minarni, M., Sari, E. I., Syahrani, A., & Mandarani, P. (2021). Klasterisasi Penyakit Menggunakan Algoritma K-Medoids pada Dinas Kesehatan Kabupaten Agam. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, *10*(3), 137.

https://doi.org/10.23887/janapati.v10i3.34904

Y. H. Chrisnanto And G. Abdillah, Penerapan Algoritma Partitioning Around Medoids (PAM) Clustering Untuk Melihat Gambaran Umum Kemampuan, Vol. 2015, No. Sentika, Pp. 444–448, 2015.

Buku:

Aprilia C, D., Aji Baskoro, D., Ambarwati, L., & Wicaksana, I. W. S. (2013). *Belajar Data Mining Dengan Rapid Minner*. 139.

Umar, A. (2021). *Mengenal Menu-Menu RapidMiner*. SoraTemplates.