

Optimalisasi *K-Medoid* Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan *Cubic Clustering Criterion*

Sofi Defiyanti^{1*}, Mohamad Jajuli², Nurul Rohmawati W³

^{1,2,3}Teknik Informatika Fakultas Ilmu Komputer Universitas Singaperbangsa Karawang
Jl. H.S Ronggowaluyo Telukjambe Timur Karawang
(corresponding author) sofi.defiyanti@unsika.ac.id*

Abstract—The scholarship is one learning assistance given to students. One of the existing scholarships is a scholarship given by the state with the name of Student Learning Aid (BBM). Grouping the data scholarship recipients are useful for determining the eligible student, considered or not entitled. By grouping, these scholarship recipients may facilitate the administration to determine the scholarship recipients, especially fuel. The grouping in was done by using a partition-based clustering algorithm that is by *K-Medoids*. The data obtained to do the grouping consists of attributes credits, GPA, Dependent parents and salaries of the parents. From the data obtained have diverse values and ranges far apart from one another. So do three scenarios, namely 1: all data obtained do grouping with *K-Medoids*, 2: partial data obtained do codification, 3: all existing data do codification. Of the three scenarios that do get the value *Cubic Clustering Criterion* (CCC). A dataset with the codification of the entire data occupies honors in uniformity in the grouping with a value of 2.245, with the value of the overall codification dataset shows the value of CCC is between 2 to 3 shows that the codification of the overall dataset has a good uniformity. This is because all the values of each attribute have a value that is almost the same.

Keywords: *scholarships, clustering, data mining, K-Medoids, Cubic Clustering Criterion (CCC)*

Intisari— Beasiswa merupakan salah satu bantuan belajar yang diberikan kepada mahasiswa. Salah satu beasiswa yang ada adalah beasiswa yang diberikan oleh negara dengan nama Bantuan Belajar Mahasiswa (BBM). Pengelompokan data mahasiswa penerima beasiswa berguna untuk menentukan mahasiswa yang berhak, dipertimbangkan atau tidak berhak. Dengan pengelompokan mahasiswa penerima beasiswa ini dapat memudahkan pihak tata usaha dalam menentukan penerima beasiswa khususnya beasiswa BBM. Pengelompokan tersebut dalam dilakukan dengan menggunakan teknik klustering berbasis partisi yaitu dengan algoritma *K-Medoids*. Data-data yang didapat untuk dilakukan pengelompokan terdiri dari atribut SKS, IPK, Tanggungan orang tua dan jumlah penghasilan orang tua. Dari data-data yang didapat memiliki nilai yang beragam dan memiliki rentang satu dengan yang lainnya berjauhan. Maka dilakukan tiga buah skenario, yaitu 1: semua data yang didapat dilakukan pengelompokan dengan *K-Medoids*, 2 : sebagian data yang didapat dilakukan kodefikasi, 3 : semua data yang ada dilakukan kodefikasi. Dari ketiga skenario yang dilakukan didapat nilai *Cubic Clustering Criterion* (CCC). Dataset dengan kodifikasi keseluruhan data menempati predikat terbaik dalam keseragaman dalam pengelompokan dengan nilai 2,245, dengan nilai tersebut dataset kodifikasi keseluruhan menunjukkan nilai CCC berada diantara 2 sampai 3 ini menunjukkan bahwa dataset kodifikasi keseluruhan mempunyai keseragaman yang baik. Hal ini dikarenakan semua nilai pada setiap atribut memiliki nilai yang hampir sama.

Kata kunci: *beasiswa, clustering, data mining, K-Medoids, Cubic Clustering Criterion (CCC)*

I. PENDAHULUAN

Beasiswa merupakan tunjangan yang diberikan kepada pelajar atau mahasiswa sebagai bantuan biaya belajar. Beasiswa diberikan untuk meningkatkan prestasi mahasiswa penerima baik kurikuler, kokurikuler ataupun ekstrakurikuler serta motivasi berprestasi bagi mahasiswa lain. Pemberian beasiswa juga memiliki tujuan untuk mengurangi jumlah mahasiswa yang putus kuliah, karena tidak mampu membiayai pendidikan serta meningkatkan akses dan pemerataan kesempatan belajar diperguruan tinggi.

Pemerintah melalui peraturan pemerintah dan mengacu kepada Undang-undang, maka pemerintah mengupayakan pemberian beasiswa bagi yang berprestasi dan bantuan biaya pendidikan bagi mahasiswa yang memiliki keterbatasan kemampuan ekonomi.

Beasiswa yang diberikan memiliki syarat dan ketentuan yang berlaku dan penetapan penerima beasiswa juga memiliki aturan-aturan yang sudah ditentukan. Penetapan beasiswa bantuan belajar ditentukan berdasarkan :

- 1) Mahasiswa yang orang tuanya paling tidak mampu
- 2) Mahasiswa yang memiliki prestasi
- 3) Mahasiswa yang memiliki IPK paling tinggi
- 4) Mahasiswa yang mencapai SKS paling banyak dengan jumlah semester paling sedikit.[1]

Banyaknya kriteria penetapan penerima beasiswa bantuan belajar mahasiswa dengan jumlah kuota terbatas dan pendaftar beasiswa yang banyak menyebabkan kesulitan bagi pihak-pihak tertentu untuk menetapkan penerima beasiswa.

Algoritma *k-medoids* dari teknik *clustering* dapat membantu dalam pengelompokan mahasiswa yang berhak menerima beasiswa, mahasiswa yang di pertimbangkan dan mahasiswa yang tidak berhak menerima beasiswa.

Tujuan dari penelitian ini adalah untuk mengelompokkan penerima beasiswa menjadi tiga buah klaster, selain itu akan dilakukan analisis terhadap hasil pengelompokan yang telah dilakukan dengan menggunakan nilai Cubic Clustering Criterion untuk mengetahui *cluster* yang lebih baik.

II. TINJAUAN PUSTAKA

A. Data Mining

Secara sederhana *data mining* adalah penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar[2]. *Data mining*, sering juga disebut sebagai *knowledge discovery in database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar[3].

Data mining adalah kegiatan menemukan pola yang menarik dari data dalam jumlah besar, data dapat disimpan dalam *database*, *data warehouse*, atau penyimpanan informasi lainnya. *Data mining* berkaitan dengan bidang ilmu – ilmu lain, seperti *database system*, *data warehousing*, statistik, *machine learning*, *information retrieval*, dan komputasi tingkat tinggi. Selain itu, *data mining* didukung oleh ilmu lain seperti *neural network*, pengenalan pola, *spatial data analysis*, *image database*, *signal processing*[4]. *Data mining* didefinisikan sebagai proses menemukan pola - pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar[5].

Karakteristik data mining sebagai berikut:

1. Data mining berhubungan dengan penemuan sesuatu yang tersembunyi dan pola data tertentu yang tidak diketahui sebelumnya.
2. Data mining biasa menggunakan data yang sangat besar. Biasanya data yang besar digunakan untuk membuat hasil lebih dipercaya.
3. Data mining berguna untuk membuat keputusan yang kritis, terutama dalam strategi[2].

Data mining dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu:

1. Deskripsi

Terkadang peneliti dan analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Sebagai contoh, petugas pengumpulan suara mungkin tidak dapat menemukan keterangan atau fakta bahwa siapa yang tidak cukup profesional akan sedikit didukung dalam pemilihan presiden. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih kearah numerik daripada kearah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi. Sebagai contoh, akan dilakukan estimasi tekanan darah sistolik pada pasien rumah sakit berdasarkan umur pasien, jenis kelamin, indeks berat badan dan level sodium darah. Hubungan antara tekanan darah sistolik dan nilai variabel prediksi dalam proses pembelajaran akan menghasilkan model estimasi. Model estimasi yang dihasilkan dapat digunakan untuk kasus baru lainnya.

3. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa yang akan datang. Beberapa metode dan teknik yang digunakan dalam klasifikasi dan estimasi dapat pula digunakan (untuk keadaan yang tepat) untuk prediksi.

4. Klasifikasi

Dalam klasifikasi terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam 3 kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah.

5. Pengklasteran

Pengklasteran merupakan pengelompokan *record*, pengamatan atau memperhatikan dan membentuk kelas objek - objek yang memiliki kemiripan. Kluster adalah kumpulan *record* yang memiliki kemiripan satu dengan yanglainnya dan memiliki ketidak miripan dengan *record-record* dalam kluster lain. Pengklasteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengklasteran. Pengklasteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, memprediksi nilai dari variabel target. Akan tetapi, algoritma pengklasteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok -

kelompok yang memiliki kemiripan (homogen), yang mana kemiripan *record* dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan *record* dalam kelompok lain akan bernilai minimal.

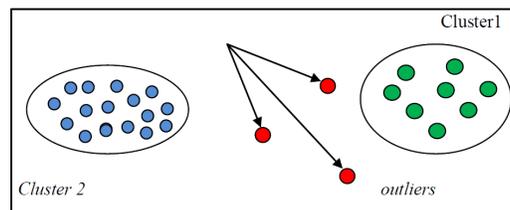
6. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum di sebut analisis keranjang belanja[6].

B. Clustering

Clustering atau klusterisasi adalah salah satu alat bantu pada data mining yang bertujuan mengelompokkan objek-objek ke dalam *cluster - cluster*. *Cluster* adalah sekelompok atau sekumpulan objek - objek data yang similar satu sama lain dalam cluster yang sama dan disimilar terhadap objek-objek yang berbeda cluster. Objek akan dikelompokkan ke dalam satu atau lebih cluster sehingga objek - objek yang berada dalam satu cluster akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Objek - objek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan objek pada cluster yang sama dan memaksimalkan ketidaksamaan pada cluster yang berbeda. Kesamaan objek biasanya diperoleh dari nilai - nilai atribut yang menjelaskan objek data, sedangkan objek - objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi.

Dengan menggunakan klusterisasi, dapat mengidentifikasi daerah yang padat, menemukan pola - pola distribusi secara keseluruhan, dan menemukan keterkaitan yang menarik antara atribut - atribut data seperti pada gambar 1. Dalam *data mining*, usaha difokuskan pada metode - metode penemuan untuk cluster pada basis data berukuran besar secara efektif dan efisien. Beberapa kebutuhan klusterisasi dalam *data mining* meliputi skalabilitas, kemampuan untuk menangani tipe atribut yang berbeda, mampu menangani dimensionalitas yang tinggi, menangani data yang mempunyai *noise*, dan dapat diterjemahkan dengan mudah.



Gambar 1. Contoh Clustering[7]

Adapun tujuan dari data *clustering* ini adalah untuk meminimalisasikan *objective function* yang diset dalam proses *clustering*, yang pada umumnya berusaha meminimalisasikan variasi di dalam suatu *cluster* dan memaksimalkan variasi antar *cluster*[7].

Pada umumnya terdapat dua pendekatan *clustering* yaitu pendekatan partisi dan pendekatan hirarki. *Clustering* dengan pendekatan partisi merupakan pengelompokan data dari satu kelompok besar kemudian dibagi menjadi beberapa kelompok yang lebih kecil. Contoh metode *clustering* dengan pendekatan partisi adalah *K-Means Clustering*. *Clustering* dengan pendekatan hirarki atau sering disebut dengan *Hierarchical Clustering* mengelompokkan data dengan menggabungkan masing-masing *record* atau individu pada data menjadi *cluster-cluster*. Contoh metode *clustering* dengan pendekatan hirarki adalah *Agglomerative Hierarchical Clustering* [8].

1) Algoritma *K-Medoids Clustering*

Algoritma *k-means* sensitif terhadap outlier, karena objek yang sangat besar mungkin secara substansial mengganggu distribusi data[9].

Berbeda dengan *k-means*, *k-medoids* tidak menentukan nilai rata - rata dari objek dalam *cluster* sebagai titik acuan, tapi menggunakan *medoid* (median), yang merupakan objek yang paling terletak dipusat sebuah cluster. Dengan demikian, metode partisi masih dapat dilakukan berdasarkan prinsip meminimalkan jumlah dari ketidaksamaan antara setiap objek dan titik acuan yang sesuai (*medoid*). Hal ini merupakan dasar dari metode *k-medoids*.

Strategi dasar dari algoritma *k-medoids* adalah untuk menemukan *cluster k* pada objek *n* dengan terlebih dahulu menemukan objek awal (*medoid*) secara acak sebagai perwakilan untuk setiap *cluster*. Setiap objek yang tersisa dikelompokkan dengan *medoid* yang paling mirip. Metode *k-medoids* menggunakan objek representatif sebagai titik yang menjadi acuan dan bukan rata - rata objek dari setiap *cluster* adalah kunci dari metode ini. Algoritma akan mengambil parameter masukan *k* jumlah *cluster* yang akan di partisi antara himpunan jumlah *n* objek.

Algoritma *k-medoids* [4] adalah sebagai berikut:

Input

: *k* : jumlah *cluster*

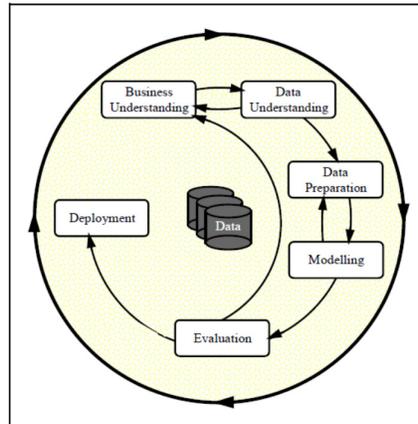
D: dataset dengan *n* buah objek

Output : satu set *k cluster* dengan sejumlah data yang paling dekat dengan *medoidnya*;

1. Secara acak memilih k objek dari D sebagai titik acuan;
Repeat:
2. Tetapkan setiap objek yang tersisa ke dalam *cluster* dengan medoid yang paling dekat;
3. Pilih secara acak objek non-medoid : O_{random} ;
4. Hitung total jarak poin S dari pertukaran objek O_j dengan O_{random} ;
5. Jika, $S < 0$ kemudian tukar O_j dengan O_{random} untuk membentuk set baru k medoid;
6. **Until:** tidak ada perubahan[10].

III. METODE PENELITIAN

Metode yang dipergunakan dalam penelitian ini menggunakan CRISP-DM. Terdapat enam proses *data mining*[11], [12] seperti tergambar dalam Gambar 2 berikut ini :



Gambar 2 Model Crips-DM[13]

a. *Bussiness Understanding*

Bussiness understanding merupakan tahap awal dalam metode CRIPS-DM. Dalam bussiness understanding akan dilakukan penentuan tujuan dari penelitian yang dilaksanakan dan menyiapkan strategi untuk mencapai tujuan yang telah ditetapkan.

b. *Data Understanding*

Setelah ditentukan tujuan yang akan dicapai maka masuk ke data understanding, pada data understanding akan dilakukan pengumpulan data yang didapat dari bagian akademik yang mengurus beasiswa mahasiswa.

c. *Data Preparation*

Setelah didapat data yang akan dipakai, lalu masuk ke dalam data preparation. Pada tahapan ini akan dilakukan persiapan data berupa pembersihan data, integrasi data dan kodefikasi.

d. *Modeling*

Setelah semua data siap maka masuk ketahap modelling. Dipemodelan ini akan diterapkan teknik clustering yaitu algoritma K-medoids.

e. *Evaluasi*

Setelah model yang dihasilkan menggunakan algoritma K-Medoid maka akan dilakukan evaluasi dengan menghitung nilai cubic cluster criterion dari setiap model yang dihasilkan sehingga dapat diketahui mana yang lebih baik.

f. *Deployment*

Setelah tahap evaluasi dilakukan maka akan dilakukan tahap pelaporan yang berhubungan dengan penemuan pengetahuan baru yang selanjutnya akan dikaitkan dengan tujuan-tujuan yang ingin dicapai.

IV. HASIL PENELITIAN

a. *Pemahaman Bisnis*

Tujuan bisnis yang ingin dicapai adalah pengelompokan penerima beasiswa yang terdiri dari tiga kelompok yaitu yang berhak menerima, dipertimbangkan dan yang tidak berhak menerima beasiswa.

Tujuan lain yang ingin dicapai adalah mengetahui jenis data yang terbaik dalam pengelompokkan penerima beasiswa menggunakan k-medoid.

b. Pemahaman Data

Pengumpulan data dilakukan dengan cara meminta data penerima beasiswa BBM yang ada di fakultas ilmu computer dari angkatan tahun 2008 sampai angkatan tahun 2011. Didapat data sebanyak 36 data mahasiswa yang mengajukan beasiswa BBM.

Atribut yang di dapat adalah NPM, SKS, IPK, semester, pendapatan orang tua dan jumlah tanggungan orang tua. Dari semua atribut yang didapat hanya atribut NPM yang tidak akan dipakai.

Data yang didapat memiliki tipe data seperti pada tabel 1. Karena bentuk data yang didapat beragam dan dengan jangkauan yang berbeda disetiap atributnya. Maka akan dilakukan percobaan dengan membagi menjadi tiga buah skenario yaitu menggunakan data asli yang didapat, menerapkan kodefikasi terhadap atribut penghasilan orang tua dan SKS, dan yang terakhir akan dilakukan kodefikasi secara keseluruhan dari data yang didapat.

Tabel 1. Atribut dan jenis data

No	Atribut	Tipe data	Contoh
1	NPM	Karakter	104117004186
2	SKS	Integer	105
3	IPK	Real	3.65
4	Semester	Integer	6
5	Pendapatan orang tua	Integer	Rp. 2.500.000
6	Jumlah tanggungan	integer	4

c. Pengolahan Data

Data yang didapat mengandung missing value. Maka akan dilakukan penanganan khusus untuk missing value yaitu menggunakan mean imputation[5,6] dengan menggunakan rumus 1. Dari rumus 1 didapat nilai rata-rata dari atribut yang mengandung missing value. Yaitu atribut penghasilan orang tua dengan nilai Rp. 1.728.025,- jika terdapat data yang hilang pada atribut penghasilan orang tua maka akan diisi dengan nilai rata-rata tersebut.

$$\bar{X} = \frac{\text{Jumlah total atribut}}{\text{penghasilan orang tua}} \quad (1)$$

$$\bar{X} = \frac{\text{Jumlah data}}{62208900}$$

$$\bar{X} = \frac{36}{1728025}$$

Dilakukan tiga buah skenario terhadap dari data yang telah didapat. Yaitu skenario pertama menggunakan data asli dari data yang telah didapat, skenario dua dilakukan kodefikasi terhadap atribut penghasilan orang tua dan SKS, dan skenario ketiga akan dilakukan keseluruhan kodefikasi.

Kodefikasi keseluruhan atribut dilakukan dengan menggunakan teknik pengelompokan yaitu K-means dengan mengelompokkannya menjadi empat kelompok untuk masing-masing atribut seperti pada tabel 2. Pengelompokan menjadi empat kelompok dilakukan agar mendekati nilai satu atribut dengan atribut yang lain, sehingga dari semua atribut yang dipakai semua dikelompokkan menjadi empat kelompok.

Tabel 2. Kodefikasi atribut

Atribut	Kodefikasi
SKS	84 - 90 = kategori 1 105 - 123 = kategori 2 60 - 63 = kategori 3 22 - 42 = kategori 4
Penghasilan orang tua	Rp. 400.00 - Rp. 1.000.000 = Kategori 4 Rp. 1.728.025 - Rp. 2.500.000 = kategori 1 Rp. 3.300.000 - 3.686.000 = Kategori 3 Rp. 4.000.000 - Rp. 5.000.000 = Kategori 2
IPK	2,38 - 2,47 = kategori 1 3,00 - 3,24 = kategori 3 3,27 - 3,52 = kategori 2 3,64 - 3,97 = kategori 4
Jumlah tanggungan orang tua	1 - 2 = kategori 2 3 = kategori 3 d. = kategori 4 e. = katagori 1
Semester	2 = kategori 4 3 = kategori 3 4 - 5 = kategori 1 6 = kategori 2

Dari ketiga skenario yang akan dilakukan maka akan menghasilkan tiga buah dataset yakni *dataset* kodifikasi sebagian, *dataset* kodifikasi keseluruhan dan *dataset* data asli (atribut yang tidak di kategorikan).

d. Pemodelan

Pemodelan *data mining* dalam penelitian ini dibuat dengan menggunakan perangkat lunak SAS. Pada aplikasi ini telah tersedia algoritma *clustering* berupa algoritma *k-medoid* [10], [16]–[18].

Algoritma K-Medoids

1. Dataset kodifikasi sebagian

Hasil dari pemodelan *k-medoid* dengan skenario kodifikasi sebagian yaitu atribut penghasilan orang tua dan SKS seperti terlihat pada Table 3. Pada hasil *clustering* dengan menggunakan data yang dikodifikasi sebagian didapat hasil cluster 1 berisi 13 data, cluster 2 berisi 8 data sedangkan cluster 3 terdiri dari 5 data. Dari hasil yang telah didapat nilai RMS Std Deviation pada masing-masing cluster memiliki nilai yang kecil sehingga cluster yang terbentuk cukup homogen.

Tabel 3. Hasil skenario kodifikasi sebagian

Cluster Summary		
Cluster	Frequency	RMS Std Deviation
1	13	0.8016
2	8	0.7830
3	5	0.6703

2. Dataset kodifikasi keseluruhan

Hasil dari pemodelan *k-medoid* dengan skenario kodifikasi keseluruhan seperti terlihat pada tabel 4. . Pada hasil *clustering* dengan menggunakan data yang dikodifikasi keseluruhan didapat hasil cluster 1 berisi 8 data, cluster 2 berisi 10 data sedangkan cluster 3 terdiri dari 18 data. Dari hasil yang telah didapat nilai RMS Std Deviation pada masing-masing cluster memiliki nilai yang kecil sehingga cluster yang terbentuk cukup homogen.

Tabel 4. Hasil skenario kodifikasi keseluruhan

Cluster Summary		
Cluster	Frequency	RMS Std Deviation
1	8	0.6492
2	10	0.8433
3	18	0.7379

3. Dataset data asli

Hasil dari pemodelan *k-medoid* dengan skenario menggunakan data asli terlihat pada tabel 5. . Pada hasil *clustering* dengan menggunakan data yang dikodifikasi sebagian didapat hasil cluster 1 berisi 9 data, cluster 2 berisi 8 data sedangkan cluster 3 terdiri dari 16 data. Dari hasil yang telah didapat nilai RMS Std Deviation pada masing-masing cluster memiliki nilai yang besar jika dibandingkan dengan scenario yang lainnya ini menunjukkan bahwa hasil cluster yang terbentuk tidak homogen. Hal ini disebabkan karena dari beberapa atribut yang dipakai memiliki rentang yang berbeda antara satu atribut dengan atribut yang lain.

Tabel 5. Hasil skenario data asli

Cluster Summary		
Cluster	Frequency	RMS Std Deviation
1	9	251934
2	8	173066
3	16	95535.0

e. Evaluasi

Evaluasi menggunakan nilai Cubic Clustering Criterion (CCC), nilai CCC menjelaskan perbandingan koefisien nilai pengamatan R^2 dengan pendekatan nilai harapan dari R^2 [19]-[21] Nilai CCC dapat dihitung dengan rumus 2 dengan bantuan *software SAS*.

$$CCC = \ln \left[\frac{1 - E(R^2)}{1 - R^2} \right] \frac{\left(\frac{np^*}{2} \right)^{\frac{1}{2}}}{(0.001 + E(R^2))^2} \quad (2)$$

Dimana :

R^2 = keragaman yang dapat dijelaskan gerombol

$E(R^2)$ = Nilai harapan dari R^2

n = Jumlah pengamatan

$p^* < p$, p = jumlah peubah

Nilai CCC dapat dilihat dari tabel 6.

Tabel 6. Predikat nilai CCC[22]

Nilai CCC	Predikat
$\geq 2 - 3$	Baik
$0 - 2$	Potensial
< 0	Menunjukkan adanya pencilan

Dan didapat hasil seperti pada tabel 7.

Tabel 7. Hasil nilai CCC

	Dataset asli	Dataset kodifikasi sebagian	Dataset kodifikasi keseluruhan
Nilai CCC	0.773	-0.855	2.245

f. Deployment

Hasil dari ketiga scenario yang telah dilakukan memiliki jumlah anggota yang berbeda antara scenario satu dengan scenario yang lain dapat dilihat pada table 3, table 4 dan table 5. Perbedaan jumlah anggota atau data antara scenario satu dengan scenario yang lain disebabkan karena pada algoritma K-medoids memiliki kelemahan saying sama dengan algoritma K-means yaitu sensitive terhadap pemilihan nilai awal sehingga jika nilai awal yang ditentukan secara acak maka akan terbentuk cluster yang berbeda[23].

Evaluasi yang telah dilakukan dengan menghitung nilai Cubic Clustering Criterion (CCC) didapat cluster yang terbaik adalah cluster dengan dataset yang dikodifikasi keseluruhan. Nilai CCC merupakan nilai yang digunakan untuk melihat keseragaman yang dihasilkan dari setiap cluster. Dataset kodifikasi keseluruhan menunjukkan nilai CCC berada diantara 2 sampai 3 ini menunjukkan bahwa dataset kodifikasi keseluruhan mempunyai keseragaman yang baik hal ini juga didukung dengan nilai RMS Std Deviation yang memiliki nilai kecil yang menandakan bahwa cluster yang didapat adalah homogen. Hal ini dikarenakan semua nilai pada setiap atribut memiliki nilai yang hampir sama. Sedangkan dataset kodifikasi sebagian memiliki nilai kurang dari 0 maka didalam dataset tersebut setelah di kelompokkan ada outlier.

V. KESIMPULAN

Terbentuknya tiga buah pengelompokan untuk data pengujian beasiswa BBM yang terdiri dari kelompok penerima, kelompok dipertibangkan menerima dan kelompok penerima beasiswa. Berdasarkan hasil evaluasi yang telah dilakukan dengan menghitung nilai Cubic Clustering Criterion (CCC) didapat bahwa dataset dengan kodifikasi keseluruhan data menempati predikat terbaik dalam keseragaman dalam pengelompokan dengan nilai 2,245 yang menempati predikat baik hal ini disebabkan oleh keseragaman data dari setiap atribut dan juga didukung dengan nilai RMS Std Deviation yang kecil.

UCAPAN TERIMA KASIH

Terimakasih diberikan kepada Fakultas ilmu Komputer UNSIKA, LPPM UNSIKA dan semua yang terlibat

REFERENSI

- [1] K. P. Nasional, D. Jenderal, P. Tinggi, dan D. Kelembagaan, "Beasiswa peningkatan prestasi akademik (ppa) dan bantuan belajar mahasiswa (bbm)," 2010.
- [2] P. Beynon-Davies, *Database Systems*. Palgrave Macmillan, 2004.
- [3] B. Santoso, *Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis*. 2007.
- [4] J. Han, J. Pei, dan M. Kamber, *Data Mining: Concepts and Techniques*. Elsevier Science, 2011.
- [5] I. H. Witten, E. Frank, dan M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier Science, 2011.
- [6] E. T. Kusri dan Luthfi, *algoritma data mining*. Penerbit Andi.
- [7] D. N. Nango, "Penerapan Algoritma K-Means untuk Clustering Data Anggaran Pendapatan Belanja Daerah Di Kabupaten XYZ," 2012.
- [8] A. Yusuf dan H. Tjandrasa, "Prediksi Nilai Dengan Metode Spectral Clustering Dan Clusterwise Regression," vol. VIII, no. 1, hal. 39-45, 2013.
- [9] H. Park et al., "A K-means-like Algorithm for K-medoids Clustering and Its Performance," *System*, no. C, hal. 2005-2008, 2008.
- [10] D. Velmurugan, "Efficiency of k-Means and K-Medoids Algorithms for Clustering Arbitrary Data Points," *International Journal of Computer Technology & ...*, vol. 3, no. 5, hal. 1758-1764, 2012.
- [11] L. a. Kurgan dan P. Musilek, "A survey of Knowledge Discovery and Data Mining process models," *The Knowledge Engineering Review*, vol. 21, no. 1, hal. 1, 2006.
- [12] R. Wirth, "CRISP-DM : Towards a Standard Process Model for Data Mining," *Proceedings of the Fourth International Conference on the Practical Application of Knowledge Discovery and Data Mining*, no. 24959, hal. 29-39, 2000.
- [13] G. Mariscal, Ó. Marbán, dan C. Fernández, "A survey of data mining and knowledge discovery process models and methodologies," *The Knowledge Engineering Review*, vol. 25, no. 2, hal. 137-166, 2010.
- [14] A. A. Romdhoni, Wiharto, dan E. Suryani, "Peramalan Nilai Tukar Mata Uang Dollar AS terhadap Rupiah Menggunakan Neural Network Ensemble Bagging," *Jurnal ITSMART*, vol. 2, no. 2, hal. 42-47, 2013.
- [15] L. S. Mandani, E. Sumarminingsih, W. H. Nugroho, J. Matematika, F. Mipa, dan U. Brawijaya, "Study Tentang Metode Complete Case , Last Observation Carried Forward , Dan Unconditional Mean Imputation Untuk Mengatasi Data Hilang Dengan Pola Monoton Dan Non Monoton Pada Data Longitudinal," *Jurnal Mahasiswa Statistik*, vol. 1, no. 2, hal. 129, 2013.
- [16] B. Setiyono dan I. Mukhlash, "Kajian Algoritma GDBScan , Clarans dan Cure untuk Spatial Clustering," *limits*, vol. 2, no. 2, hal. 117-128, 2005.
- [17] T. Wuryandari, A. Rusgionyono, dan E. Setyowati, "Pengelompokan Kabupaten/Kota Berdasarkan Komoditas Pertanian Menggunakan Metode K Medoids," *Media Statistika*, vol. 9, no. 1, hal. 41-49, 2016.
- [18] W. A. Triyanto, "Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran," *Simetris*, vol. 6, no. 1, hal. 183-188, 2015.
- [19] C. Bian, C. Yuan, W. Kuang, dan D. Wu, "Evaluation, Classification, and Influential Factors Analysis of Traffic Congestion in Chinese Cities Using the Online Map Data," *Mathematical Problems in Engineering*, vol. 2016, 2016.
- [20] S. K. A. Fahad dan M. Alam, "A Modified K-Means Algorithm for Big Data Clustering," vol. 6, no. 4, hal. 129-132, 2016.
- [21] R. J. Broderick, K. Munoz-Ramos, dan M. J. Reno, "Accuracy of clustering as a method to group distribution feeders by PV hosting capacity," *Proceedings of the IEEE Power Engineering Society Transmission and Distribution Conference*, vol. 2016-July, no. Ccc, 2016.
- [22] Ahmad Nur Rohman, "Pengelompokan wilayah di indonesia dengan analisis gerombol dan analisis input output dalam bidang teknologi komunikasi ahmad nur rohman," 2015.
- [23] L. Aiguo, Z. Shuli, Q. Taorong, dan B. Xiaoming, "Research on K-medoids clustering algorithm based on data density and its parallel processing based on MapReduce," vol. 13, no. 7, hal. 1-6, 2016.