

Implementasi Algoritma *K-Medoids* untuk Pengelompokkan Sebaran Mahasiswa Baru

Eka Irawan¹, Sandy Putra Siregar², Irfan Sudahri Damanik³, Ilham Syaputra Saragih⁴

^{1,2,3,4} Program Studi Sistem Informasi, STIKOM Tunas Bangsa

Jl. Jend. Sudirman Blok A no 1,2 dan 3 Pematangsiantar

eka.irawan@amiktunasbangsa.ac.id

Abstract

The existence of new students at a tertiary institution is a routine activity every year in a tertiary institution and can also see the sustainability of the tertiary institution. The variety of regional origin of new students makes the party from the university want to see the distribution of new students based on the origin of the school and its place of residence. STIKOM Tunas Bangsa is one of the tertiary institutions in Pematangsiantar. It aims to promote the university. *K-Medoids* is able to group data on the distribution of new students in STIKOM Tunas Bangsa Pematangsiantar. The clusters produced in this study are of three clusters. The validity used in this study is the validity of Silhouette Coefficient. The validity value generated in the *K-Medoids* algorithm produces a validity value of -116.47 by assuming that if the non-medoids value produced $S < 0$ then the cluster process is stopped.

Keywords: Distribution, Clustering, Data Mining, *KMedoids*, Validation

Abstrak

Keberadaan mahasiswa baru pada suatu perguruan tinggi merupakan kegiatan rutin setiap tahunnya dalam sebuah perguruan tinggi dan dapat juga melihat keberlangsungan suatu perguruan tinggi tersebut. Beragamnya asal wilayah dari mahasiswa baru membuat pihak dari perguruan tinggi ingin melihat persebaran mahasiswa baru berdasarkan asal sekolah dan tempat tinggalnya. STIKOM Tunas Bangsa merupakan salah satu perguruan tinggi yang ada di pematangsiantar bertujuan untuk melakukan promosi dari perguruan tinggi tersebut. *K-Medoids* mampu melakukan pengelompokan pada data sebaran mahasiswa baru yang ada pada STIKOM Tunas Bangsa Pematangsiantar. Kluster yang dihasilkan pada penelitian ini adalah berjumlah tiga kluster. Validitas yang digunakan pada penelitian ini adalah validitas Silhouette Coefficient. Adapun nilai validitas yang dihasilkan pada algoritma *K-Medoids* menghasilkan nilai validitas sebesar -116,47 dengan mengasumsikan bahwa jika nilai non medoids yang dihasilkan $S < 0$ maka proses cluster dihentikan.

Kata Kunci: Sebaran, Clustering, Data Mining, *KMedoids*, Validasi

1. PENDAHULUAN

Data mining dapat diterapkan untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. Terdapat beberapa teknik yang digunakan dalam *data mining*, salah satu teknik *data mining* adalah *clustering* [1]. Terdapat dua jenis metode *clustering* yang digunakan dalam pengelompokan data, yaitu *hierarchical clustering* dan *non-hierarchical clustering*.

Algoritma *clustering* salah satunya adalah *K-Medoids* yang termasuk salah satu metode pengelompokan di dalam *data mining*. Algoritma ini menggunakan objek pada kumpulan objek untuk mewakili sebuah *cluster*. Metode *K-Medoids* cukup efisien untuk dataset yang kecil. Kelebihan dari metode *K-Medoids* mampu mengatasi kelemahan dari metode *K-Means* yang *sensitive* terhadap outlier dan hasil proses *clustering* tidak bergantung pada urutan masuk dataset [2] Dalam hal ini penulis menggunakan algoritma *K-Medoids* dengan *tools* Rapidminer dalam

mengelompokkan kemiskinan menurut provinsi. Diharapkan dengan adanya penelitian ini dapat membantu pemerintah dalam meningkatkan sumber daya ekonomi penduduk, sehingga dapat membantu penduduk yang kurang mampu.

Berdasarkan berlimpahnya data mahasiswa, informasi yang tersembunyi dapat diketahui dengan cara melakukan pengolahan terhadap data tersebut sehingga berguna bagi pihak universitas [3]. Pengolahan data mahasiswa perlu dilakukan untuk mengetahui informasi penting berupa pengetahuan baru (knowledge discovery), misalnya informasi mengenai pengelompokan data mahasiswa berpotensi berdasarkan kota asal mahasiswa [3]. Suatu perguruan tinggi menyimpan data akademik, administrasi, biodata mahasiswa dan lain-lain. Data tersebut akan sangat bermakna jika digali dengan tepat sehingga dapat diketahui pola atau pengetahuan untuk mengambil keputusan [4]. Selain pengambilan data, studi literatur satu cara untuk menemukan atau melengkapi penelitian yang sedang dilakukan sebagai referensi untuk melakukan penelitian menjadi lebih lengkap [5]. Pengetahuan baru tersebut dapat membantu pihak sekolah tinggi dalam menentukan strategi promosi terhadap calon mahasiswa baru dengan tepat sasaran. Berdasarkan latar belakang masalah di atas, penulis mengambil judul "Implementasi Algoritma *K-Medoids* untuk Pengelompokan Asal Sekolah Sebaran Mahasiswa Baru". Diharapkan penelitian ini dapat mengetahui dimana asal sekolah terbanyak dan tersedikit berdasarkan *cluster* tinggi dan rendah untuk melakukan promosi.

2. METODOLOGI PENELITIAN

Dalam penelitian ini menggunakan metode datamining dengan *K-medoid* antara lain: a) Tahap pengumpulan data, b) Analisis kebutuhan, c) Pengolahan Data, d) Tahap *clustering* dan e) Tahap Analisis.

2.1. Tahap Pengumpulan Data

Data yang dibutuhkan dalam penelitian ini adalah data asal sekolah mahasiswa yang terdapat di STIKOM Tunas Bangsa yabf di peroleh dari BAK STIKOM Tunas Bangsa. Data yang digunakan berisi NIM, Nama Mahasiswa, Alamat Dan Asal Sekolah. Data yang digunakan dari tahun 2017 sampai dengan 2019.

2.2. Analisa kebutuhan

Analisa kebutuhan dilakukan dengan menentukan kebutuhan apa saja yang dibutuhkan dalam pengimplementasian metode *K-Medoids clustering* untuk pengelompokan data persebaran mahasiswa STIKOM Tunas Bangsa berdasarkan asal Sekolah.

2.3. Tahap Pengolahan Data

Data diolah menjadi 2 *cluster* yaitu *cluster* tingkat fasilitas tertinggi (C1) dan *cluster* tingkat fasilitas terendah (C2) sehingga dapat diperoleh perhitungan nilai yg akan di proses pada tahap selanjutnya, data setiap wilayah yang memiliki fasilitas perguruan tinggi akan dijumlah setiap aspeknya sehingga pada langkah ini sudah diperoleh data perhitungan nilai yang akan diproses pada prosedur *clustering*. Proses pengolahan menggunakan bantuan aplikasi RapidMiner.

2.4. Tahap Clustering

Clustering merupakan klasifikasi tanpa pengawasan dan merupakan proses partisi sekumpulan objek data dari satu set menjadi beberapa kelas. Hal ini dapat dilakukan dengan menerapkan berbagai persamaan dan langkah-langkah mengenai jarak algoritma.

2.5. Tahap Analisis

Pada tahapan ini dilakukan analisis persebaran mahasiswa berdasarkan asal sekolah kemudian akan dilakukan analisis berdasarkan sekolah asal, mana sekolah yang banyak menyumbangkan siswanya dengan sekolah yang sedikit menyumbangkan siswanya. Tujuannya agar dapat melakukan marketing ke beberapa sekolah yang memiliki peluang masuk STIKOM Tunas Bangsa.

3. HASIL DAN PEMBAHASAN

Pada tahap ini pengolahan data kemiskinan menurut provinsi di Indonesia menggunakan algoritma *k-medoids* dengan tool *Rapidminer*. *Rapidminer* adalah *software* pilihan untuk melakukan ekstraksi data dengan metode-metode data mining. Berikut adalah data kemiskinan di Indonesia diambil dari Badan Pusat Statistika (BPS) pada gambar 1.

Berikut Tahapan dalam proses perhitungan manual menggunakan algoritma *k-medoids*.

- a. Memilih secara acak *centroid* awal, Berikut penentuannya terdapat pada tabel 1.

Tabel 1. *Centroid Awal Medoids*

Nama	Keterangan	X	Y	Z
C1	Data ke-3 dengan sebagai pusat <i>cluster</i> ke-1	2	2	2
C2	Data ke-4 dengan sebagai pusat <i>cluster</i> ke-1	1	1	1
C3	Data ke-5 dengan sebagai pusat <i>cluster</i> ke-1	2	3	1

- b. Penempatan Objek-objek *non medoids* ke dalam *cluster* paling dekat dengan *medoids* berdasarkan jarak *Euclidean*. Contoh perhitungan jarak terdekat pada data ke-1 :

$$Al - Wasliyah_{(C1)} = \sqrt{(1 - 2)^2 + (1 - 2)^2 + (2 - 2)^2} = 1,41$$

$$Al - Washliyah_{(C2)} = \sqrt{(1 - 1)^2 + (1 - 1)^2 + (2 - 1)^2} = 1,00$$

$$Al - Washliyah_{(C3)} = \sqrt{(1 - 2)^2 + (1 - 3)^2 + (2 - 1)^2} = 2,45$$

Perhitungan jarak data ke-2 sampai data ke-73 ke setiap *medoids*, sehingga hasil dari data keseluruhannya dapat diperoleh terdapat pada table 2 berikut.

Tabel 2. Hasil Perhitungan Jarak Ke Setiap Medoids

Data Ke - i	C1	C2	C3	Terdekat	Klaster
AL WASHLIYAH KM.6 MEDAN	1,41	1,00	2,45	1,00	2
AMIK TUNAS BANGSA PEMATANGSIANTAR	23,69	25,42	23,77	23,69	1
MA AL WASHLIYAH SERBELAWAN	0,00	1,73	1,41	0,00	1
MA IBNU SINA	1,73	0,00	2,24	0,00	2

Data Ke - i	C1	C2	C3	Terdekat	Klaster
MA SWASTA IKHLASİYAH GUPPI	1,41	2,24	0,00	0,00	3
.....				
SMK SWASTA TAMAN SISWA PEMATANG SIANTAR	8,12	9,85	8,12	8,12	3
SMK SWASTA TELADAN PEMATANGSIANTAR	22,91	24,54	20,81	20,81	3
SMK SWASTA TELADAN TANAH JAWA	3,32	4,00	3,61	3,32	1
SMK SWASTA TRISAKTI KOTA PEMATANGSIANTAR	2,24	2,45	3,32	2,24	1
SMK SWASTA TUNAS JAYA TANAH JAWA	2,00	1,73	2,45	1,73	2
SMK SWASTA YP HKBP KOTA PEMATANGSIANTAR	2,45	2,24	3,16	2,24	2
Jumlah	420,88	458,58	432,58		
Total Cost	383,97				

c. Mengasumsikan objek telah ditentukan sebagai *non medoids* berikut :

Tabel 3. Non Medoids

Nama	Keterangan	X	Y	Z
C1	Data ke-68 dengan sebagai pusat <i>cluster</i> ke-0	7	7	6
C2	Data ke-73 dengan sebagai pusat <i>cluster</i> ke-1	3	0	1
C3	Data ke-69 dengan sebagai pusat <i>cluster</i> ke-2	12	21	10

d. Penempatan Objek-objek *non medoids* ke dalam *cluster* paling dekat dengan *medoids* berdasarkan jarak *Euclidean*. Contoh perhitungan jarak terdekat pada data ke-1 :

$$Al - Washliyah_{(C1)} = \sqrt{(1-7)^2 + (1-7)^2 + (2-6)^2} = 9,38$$

$$Al - Washliyah_{(C2)} = \sqrt{(1-3)^2 + (1-0)^2 + (2-1)^2} = 2,45$$

$$Al - Washliyah_{(C3)} = \sqrt{(1-12)^2 + (1-21)^2 + (2-10)^2} = 24,19$$

e. Sehingga hasil yang diperoleh seperti tabel 4 berikut.

Tabel 4. Hasil Perhitungan Jarak Ke Setiap Medoids

Asal Sekolah	C1	C2	C3	Terdekat	Klaster
AL WASHLIYAH KM.6 MEDAN	9,38	2,45	24,19	2,45	2
AMIK Tunas Bangsa Pematangsiantar	15,65	24,88	9,38	9,38	3
MA AL WASHLIYAH SERBELAWAN	8,12	2,45	22,91	2,45	2
MA IBNU SINA	9,85	2,24	24,54	2,24	2
MA SWASTA IKHLASİYAH GUPPI	8,12	3,16	22,47	3,16	2
.....				
SMK SWASTA TAMAN SISWA PEMATANG SIANTAR	0,00	9,49	15,39	0,00	1
SMK Swasta Teladan Pematangsiantar	15,39	24,56	0,00	0,00	3

Asal Sekolah	C1	C2	C3	Terdekat	Klaster
SMK SWASTA TELADAN TANAH JAWA	8,06	2,24	23,02	2,24	2
SMK SWASTA TRISAKTI KOTA PEMATANGSIANTAR	9,11	2,24	24,29	2,24	2
SMK SWASTA TUNAS JAYA TANAH JAWA	9,49	3,74	23,85	3,74	2
SMK Swasta YP HKBP Kota Pematangsiantar	9,49	0,00	24,56	0,00	2
Jumlah	645,21	483,22	1488,38		
Total Cost	271,50				

- f. Hitung total simpangan (S) dengan menghitung total *distance* baru – total *distance* lama. Jika $S < 0$ maka tukar objek dengan data *cluster* untuk membentuk sekumpulan k objek baru sebagai *medoids*.

$$S = \text{total distance baru} - \text{total distance lama} \quad (1)$$

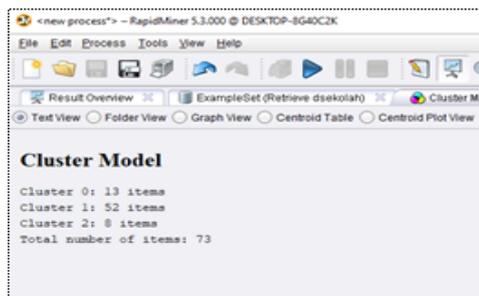
$$S = 271,50 - 387,97 = -116,47$$

- g. Jika nilai $S < 0$ maka proses pengklasteran dihentikan. Sehingga menghasilkan *Cluster* Dapat dilihat pada tabel berikut.

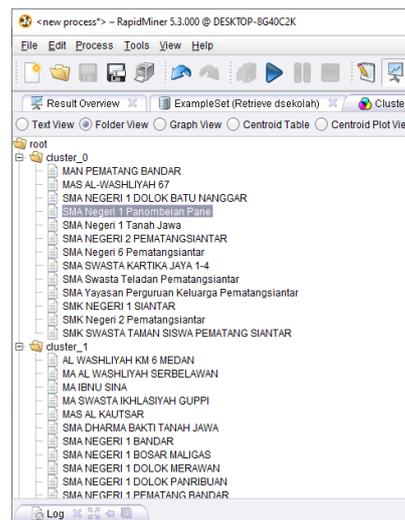
Tabel 5. Hasil Pengklasteran dengan K-Medoids

Asal Sekolah	C1	C2	C3	Terdekat	Klaster
AL WASHLIYAH KM.6 MEDAN	9,38	2,45	24,19	2,45	2
AMIK Tunas Bangsa Pematangsiantar	15,65	24,88	9,38	9,38	3
MA AL WASHLIYAH SERBELAWAN	8,12	2,45	22,91	2,45	2
MA IBNU SINA	9,85	2,24	24,54	2,24	2
MA SWASTA IKHLASIAH GUPPI	8,12	3,16	22,47	3,16	2
.....				
.....				
SMK SWASTA TAMAN SISWA PEMATANG SIANTAR	0,00	9,49	15,39	0,00	1
SMK Swasta Teladan Pematangsiantar	15,39	24,56	0,00	0,00	3
SMK SWASTA TELADAN TANAH JAWA	8,06	2,24	23,02	2,24	2
SMK SWASTA TRISAKTI KOTA PEMATANGSIANTAR	9,11	2,24	24,29	2,24	2
SMK SWASTA TUNAS JAYA TANAH JAWA	9,49	3,74	23,85	3,74	2
SMK Swasta YP HKBP Kota Pematangsiantar	9,49	0,00	24,56	0,00	2

Pada tabel diatas dapat dianalisa kesimpulan yang dimiliki adalah bahwa data yang digunakan *valid*, dengan hasil akhir perhitungan di Rapidminer 5.3 dengan perhitungan excel menampilkan hasil akhir yang sama. Berikut tampilan *cluster* model yang berada di Rapidminer.

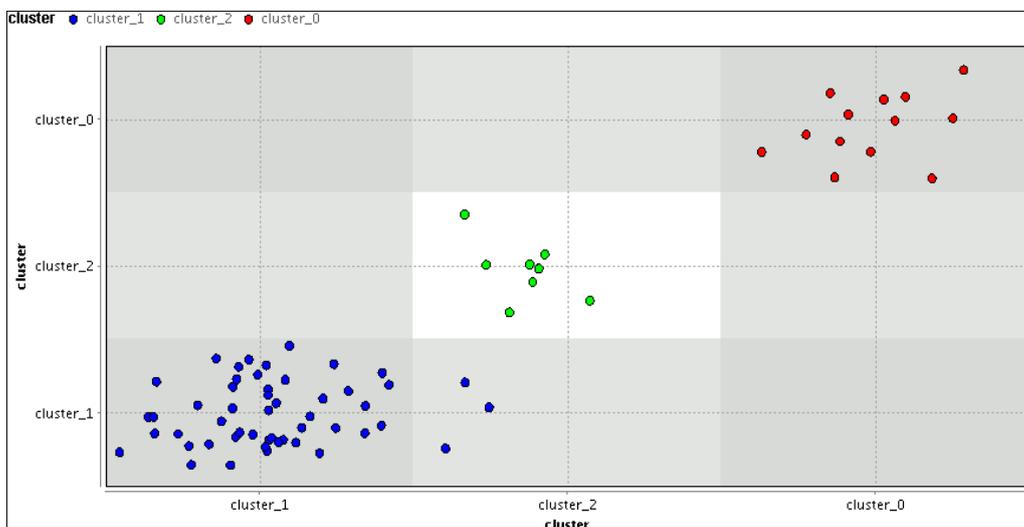


Gambar 1. Cluster Model Pada Rapidminer



Gambar 2. Cluster Folder View Rapidminer 5.3

Persebaran *cluster* pada *software* rapidminer 5.3 dapat dilihat pada gambar 3. Sebagai berikut:



Gambar 3. Persebaran *cluster* Mahasiswa Baru

4. SIMPULAN

Pada hasil analisis dan pembahasan persebaran mahasiswa baru diatas menggunakan 3 *cluster* yakni klaster tinggi (0), klaster sedang (2), dan klaster rendah (1). Berdasarkan perhirungan *k-medoids* didapatkan *cluster* tinggi sebanyak 13 *items*, *Cluster* Sedang sebanyak 8 *items* dan *cluster* rendah sebanyak 52 *items*. Pada Proses *clustering* menggunakan *K-Medoids* dengan metode *silhouette coefficient* menghasilkan nilai validitas sebesar -116,47 dengan mengasumsikan bahwa jika nilai *non medoids* yang dihasilkan $S < 0$ maka proses *cluster* dihentikan. Pengelompokan menggunakan algoritma *K-Medoids* pada kasus sebaran mahasiswa baru diperoleh jumlah *cluster* model yang sama di aplikasi Rapidminer 5.3 dengan perhitungan manual sehingga dihasilkan data yang valid.

DAFTAR PUSTAKA

- [1] Sudirman, A. P. Windarto, and A. Wanto, "Data mining tools | rapidminer: K-means methodon *clustering* of rice crops by province as efforts to stabilize food crops in Indonesia," IOP Conf. Ser. Mater. Sci. Eng., vol. 420, no. 1, 2018, doi: 10.1088/1757899X/420/1/012089.
- [2] Abdillah, G., Putra, F. A., & Renaldi, F. (2016). Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan Klasifikasi Potensi Pemakaian Air Pelanggan Baru Di Pdam Tirta Raharja Menggunakan Algoritma K-Means. *Seminar Nasional Teknologi Informasi Dan Komunikasi, 2016*(Data Mining), 18–19.
- [3] Johan Oscar Ong, "Implementasi Algoritma K-Means *Clustering* Untuk Menentukan Strategi Marketing President University," *Jurnal Ilmiah Teknik Industri*, vol. 12, no. 1, pp. 10-13, Juni 2013
- [4] Ogor. (2007). Student Academic Performance Monitoring and Evaluation Using Data Mining Techniques.
- [5] Sandy Putra Siregar, and A. Wanto. "Analysis Accuracy of Artificial Neural Network Using Backpropagation Algorithm In Predicting Process (Forecasting)" *International Journal Of Information System & Technology* Vol. 1, No. 1, (2017) , pp. 34-42