

Penerapan Algoritma *K-Means* pada Pengelompokan Data Pendaftar Bantuan Biaya Pendidikan

Abdul Fadlil¹⁾, Imam Riadi²⁾, Yana Mulyana^{3)*}

¹⁾Teknik Elektro, Universitas Ahmad Dahlan

²⁾Sistem Informasi, Universitas Ahmad Dahlan

³⁾Magister Informatika, Universitas Ahmad Dahlan

Correspondence author : yana2107048012@webmail.uad.ac.id, Indonesia

DOI: <https://doi.org/10.37012/jtik.v8i2.1261>

Abstrak

Universitas Muhammadiyah Tasikmalaya merupakan salah satu perguruan tinggi yang mendapatkan bantuan biaya pendidikan untuk mahasiswa penerima program Kartu Indonesia Pintar Kuliah (KIP-Kuliah) disetiap tahunnya. Program ini diperuntukan bagi lulusan SMA/SMK/ sederajat dari keluarga miskin/rentan miskin/afirmasi yang memiliki keinginan untuk melanjutkan belajar ke jenjang yang lebih tinggi. Hasil dari evaluasi pelaksanaan dalam penetapan data penerimaannya terdapat masalah karena data pendaftar masih banyak yang berasal dari keluarga mampu, disamping itu jumlah kuota yang diberikan oleh pemerintah sebanyak 30 kuota, jauh lebih sedikit daripada jumlah pendaftar yang berjumlah 191, sehingga harus ada metode yang dapat mengoptimalkan pengelompokan data pendaftar terlebih dahulu agar penetapan penerima bantuan biaya pendidikan KIP-Kuliah tepat sasaran. Dalam penelitian ini penulis menggunakan metode *K-means* untuk pengelompokan data pendaftar dengan jumlah klaster sebanyak 3 ($K=3$). Hasil dari penelitian ini yaitu klaster C0 sebanyak 109 data (57,1%), klaster C1 sebanyak 52 data (27,2%), dan klaster C2 sebanyak 30 data (15,7%). Hasil dari sebaran data pada masing-masing kelompok, penulis merekomendasikan klaster C0 sebagai data kelompok yang dipertimbangkan, klaster C1 sebagai kelompok yang tidak layak dan klaster C2 sebagai kelompok yang layak mendapatkan bantuan biaya pendidikan / KIP-Kuliah.

Kata kunci: Bantuan biaya pendidikan; KIP-Kuliah; K-Means; Clustering; Rapidminer

Abstract

Universitas Muhammadiyah Tasikmalaya is one of the universities that receives tuition fee assistance for students receiving Kartu Indonesia Pintar Kuliah (KIP-Kuliah) program every year. This program is intended for high school / vocational / equivalent graduates from poor / vulnerable poor / affirmation families who have the desire to continue their studies to a higher level. The results of the implementation evaluation in determining the admission data are problematic because there are still many registrant data from capable families, besides that the number of quotas given by the government is 30 quotas, far less than the number of registrants totaling 191, so there must be a method that can optimize the grouping of registrant data first so that the determination of KIP-Kuliah education fee assistance recipients is right on target. In this study, the authors used the K-means method to group registrant data with a total of 3 clusters ($K=3$). The results of this study were cluster C0 as many as 109 data (57.1%), cluster C1 as many as 52 data (27.2%), and cluster C2 as many as 30 data (15.7%). As a result of the distribution of data in each group, the authors recommend cluster C0 as the data group considered, cluster C1 as an unfit group and cluster C2 as a group that deserves tuition assistance / KIP-Kuliah.

Keywords: Tuition assistance; KIP-Kuliah; K-Means; Clustering; Rapidminer

PENDAHULUAN

Pendidikan merupakan salah satu unsur penting dalam sebuah negara karena memiliki peran yang sangat signifikan dalam perubahan sosial masyarakat tidak terkecuali di negara maju ataupun negara berkembang. Adanya pendidikan yang merata dan memadai di semua kalangan masyarakat akan berimbas terhadap kemajuan sebuah negara. Di

Indonesia, pendidikan menjadi tujuan utama dalam pembangunan nasional sehingga tercantum dalam amandemen UUD 1945 yang menyatakan bahwa negara berkewajiban mencerdaskan kehidupan bangsa dan memajukan kesejahteraan umum (Rohaeni & Saryono, 2018). Menanggapi hal tersebut, pemerintah mengeluarkan program Kartu Indonesia Pintar Kuliah (KIP-Kuliah). KIP-Kuliah merupakan salah satu bantuan biaya pendidikan yang mencakup biaya akademik dan biaya hidup sehari-hari sesuai dengan jenjang program pendidikan. Program jenjang sarjana dibiayai selama 8 semester dan untuk program jenjang diploma selama 6 semester.

Universitas Muhammadiyah Tasikmalaya (UMTAS) merupakan salah satu universitas swasta dibawah naungan LLDIKTI Wilayah IV, yang memiliki 3 fakultas dan tiga belas prodi. Setiap tahunnya, UMTAS selalu mendapatkan alokasi kuota KIP-Kuliah dan jumlah kuota yang diterimanya variatif sesuai dengan kuota yang diberikan oleh pemerintah. Kuota KIP-Kuliah yang didapat UMTAS di tahun 2022 sebanyak 30 mahasiswa, sementara data pendaftar sebanyak 191 sehingga membingungkan pihak pengelola dalam memutuskan siapa saja yang berhak menerimanya dikarenakan dalam penentuannya masih menggunakan perhitungan secara manual. Oleh karena itu diperlukan sebuah metode untuk pengelompokan data pendaftar agar penentuan hasilnya bisa lebih efektif dan akurat, salah satu metode untuk pengelompokan data yaitu metode algoritma *K-Means clustering*

Penelitian terdahulu mengenai metode k-means yang dilakukan oleh Darlinda dan Utamajaya untuk penentuan penerima beasiswa PIP menggunakan *tools* rapidminer dengan 5 variable penilaiannya, dari total 236 data pendaftar didapatkan 3 klaster, yaitu klaster 1 sebanyak 70 data sebagai penerima tepat sasaran, klaster 2 sebanyak 118 data sebagai penerima tidak tepat sasaran dan klaster 3 sebanyak 48 data sebagai penerima sangat tidak tepat sasaran (Darlinda & Utamajaya, 2022). Penelitian yang dilakukan oleh Sabrina Aulia Rahman dan Jovi Antares yaitu klasterisasi data mahasiswa calon penerima beasiswa yayasan menggunakan metode k-means, dimana jumlah klaster dari hasil perhitungan sebanyak 3 klaster yaitu diterima, dipertimbangkan dan di tolak dengan inisialisasi C1, C2 dan C3. Dari total data pendaftar sebanyak 80 pendaftar dihasilkan sebanyak 16% kategori C1, 61% kategori C2 dan 23% kategori C3 (Rahmah & Antares, 2021).

Dari latar belakang yang sudah dipaparkan diatas, penulis menerapkan konsep data mining dalam penyelesaian masalah pengelompokan data penerima dana bantuan biaya pendidikan atau KIP-Kuliah menggunakan metode algoritma *K-Means*. Data yang akan diolah sebanyak 191 dengan nilai $K=3$ dan variable penilaiannya sebanyak 5 terdiri dari status

DTKS, kepemilikan kartu KIP/KKS, Penghasilan orang tua, Status Orang tua, jumlah tanggungan dan prestasi. Hasil penelitian ini nantinya dapat memberikan informasi kepada pihak Perguruan Tinggi dalam menentukan klaster data mana yang berhak sebagai penerima bantuan biaya pendidikan atau KIP-Kuliah.

METODE

Metode memberikan gambaran-gambaran penting dalam pembuatan langkah awal penelitian. Langkah awal penelitian yang dilakukan penulis sebagai berikut :

1. Studi literatur

Studi literatur merupakan kegiatan ilmiah yang dilakukan untuk menemukan jawaban atas suatu masalah, tujuannya memberikan sumbangan teoretis atau praktis bagi pengembangan disiplin ilmu yang relevan (Wardhani, 2016). Studi literatur pada penelitian ini meliputi pengolahan data pendaftar KIP-Kuliah di UMTAS dan penggunaan algoritma K-Means. Studi literatur merupakan salah satu elemen penting dalam penelitian untuk mengkaji masalah yang akan dibahas secara teoritis melalui kajian berbagai jurnal dan referensi lainnya.

2. Pengumpulan data

Metode pengumpulan data merupakan teknik atau metode yang melakukan pengumpulan data dalam mendapatkan informasi yang dibutuhkan untuk mencapai tujuan penelitian (Kusnasari³, 2014). Pada tahap ini proses pengumpulan data dilakukan dengan cara wawancara dengan kepala biro administrasi akademik kemahasiswaan dan alumni Universitas Muhammadiyah Tasikmalaya. untuk prosedur pengumpulan data sebagai berikut :

- 1) Penulis melakukan wawancara kepada kepala biro administrasi akademik kemahasiswaan dan alumni
- 2) Penulis meminta data pendaftar KIP-Kuliah yang ada di UMTAS
- 3) Penulis mengumpulkan data pendaftar KIPK-Kuliah yang diberikan oleh kepala biro administrasi akademik kemahasiswaan dan alumni
- 4) Penulis melakukan pengelompokan data dari yang sudah dikumpulkan berdasarkan nama pendaftar
- 5) Penulis membuat analisis dari data yang sudah di kelompokkan KIP-Kuliah
- 6) Dari hasil analisis data maka didapatkan sebuah hasil yang diharapkan

Analisis merupakan suatu usaha dalam memahami suatu masalah sehingga dapat memberikan penilaian dengan tujuan mencapai pemahaman yang lebih spesifik tentang

masalah yang sedang diselidiki (Asahar Johar, 2019). Data adalah sekumpulan informasi yang diperoleh dari hasil pencarian atau penyelidikan dari suatu sumber informasi tertentu yang dianggap kompetibel. Dalam penelitian ini penulis menggunakan analisis data kualitatif.

Penelitian ini melibatkan seluruh data pendaftar KIP-Kuliah UMTAS, data tersebut meliputi status DTKS, kepemilikan kartu KIP/KKS, penghasilan orang tua, Status Orang tua, Jumlah Tanggungan Orang tua dan prestasi. Semua data yang disiapkan dalam format .xlsx. dapat dilihat pada Tabel 1.

Tabel 1. Data pendaftar KIP-K

NO	Nama Pendaftar	Status DTKS	No. KIP	No. KKS	Penghasilan Orangtua	Status Ayah	Status Ibu	Tanggungan	Prestasi Siswa
1	Dali Fadilah	Terdata	-	-	-	Hidup	Hidup	2	Sebagai Peserta - Kwartir Gerakan Pramuka 2018 - Tingkat Kabupaten/Kota
2	Fitriyani	Tidak	F43P6H	-	Tidak Berpenghasilan	Wafat	Hidup	2	
3	Farhan	Terdata	-	-	-	Hidup	Hidup	1	
...
189	Azmi Ajkyia	Tidak	-	-	Rp. 2.500.000	Hidup	Hidup	4	
190	Firda Nabila	Tidak	-	-	Rp. 2.000.000	Hidup	Hidup	4	
191	Silvi Silviani	Terdata	P57KC1	-	-	Hidup	Hidup	6	

Pra-pemrosesan data merupakan langkah yang penting dan diperlukan dalam proses penambangan data karena berkaitan erat dengan persiapan dan pembuatan kumpulan data asli. Kualitas data menjadi penentu dalam proses data mining oleh karena itu data tidak boleh kosong atau ada nilai data yang hilang, sehingga harus dipastikan terlebih dahulu bahwa data yang dimasukkan sudah benar dan tidak ada data yang bersifat anomali karena dapat mempengaruhi hasil klasterisasi (Fadhilah et al., 2020)(Setiawan, 2016). Untuk mempermudah dalam klasterisasi data harus dilakukan inialisasi data terlebih dahulu yaitu merubah data nominal menjadi data numerik. Atribut kepemilikan kartu dilakukan perubahan dari nominal menjadi numerik, parameter penilaian pada atribut ini yaitu tidak memiliki kartu, memiliki kartu KKS, memiliki kartu KIP, memiliki kartu KIP dan KKS. Perubahan nominal menjadi numerik pada atribut ini bisa dilihat pada Tabel 2.

Tabel 2. Inisial kepemilikan kartu KIP/KKS

Kepemilikan Kartu KIP/KKS	Inisial
Tidak memiliki kartu	1
Memiliki kartu KKS	2
Memiliki kartu KIP	3
Memiliki kartu KIP dan KKS	4

Untuk atribut status DTKS dilakukan perubahan dari nominal menjadi numerik, parameter pada penilaian ini yaitu terdeteksinya DTKS atau tidak. Untuk lebih jelas dapat dilihat pada Tabel 3.

Tabel 3. Inisial status DTKS

Status DTKS	Inisial
Tidak terdeteksi DTKS	1
Terdeteksi DTKS	2

Atribut penghasilan orangtua dilakukan perubahan dari nominal menjadi numerik, parameter pada penghasilan orangtua dinilai dari besaran penghasilan rata-rata perbulan. Untuk lebih jelas perubahan penghasilan rata-rata perbulan dapat dilihat pada Tabel 4.

Tabel 4. Inisial penghasilan orang tua /bulan

Penghasilan rata-rata perbulan	Inisial
>1.500.001	1
1.25.001 – 1.500.000	2
1.000.001 – 1.250.000	3
750.001 – 1.000.000	4
500.001 – 750.000	5
<500.000	6

Atribut status orangtua dilakukan perubahan dari nominal menjadi numerik, parameter pada status orangtua dapat dilihat pada Tabel 5.

Tabel 5. Inisial status orang tua

Status orangtua	Inisial
Ayah dan ibu hidup	1
Ayah dan ibu bercerai	2
Ibu meninggal	3
Ayah meninggal	4
Ayah dan ibu meninggal	5

Atribut prestasi dilakukan perubahan dari nominal menjadi numerik, parameter pada prestasi yaitu tingkat internasional, nasional, provinsi, kota/kabupaten. Untuk lebih jelas perubahan pada prestasi dari nominal menjadi numerik dapat dilihat pada Tabel 6.

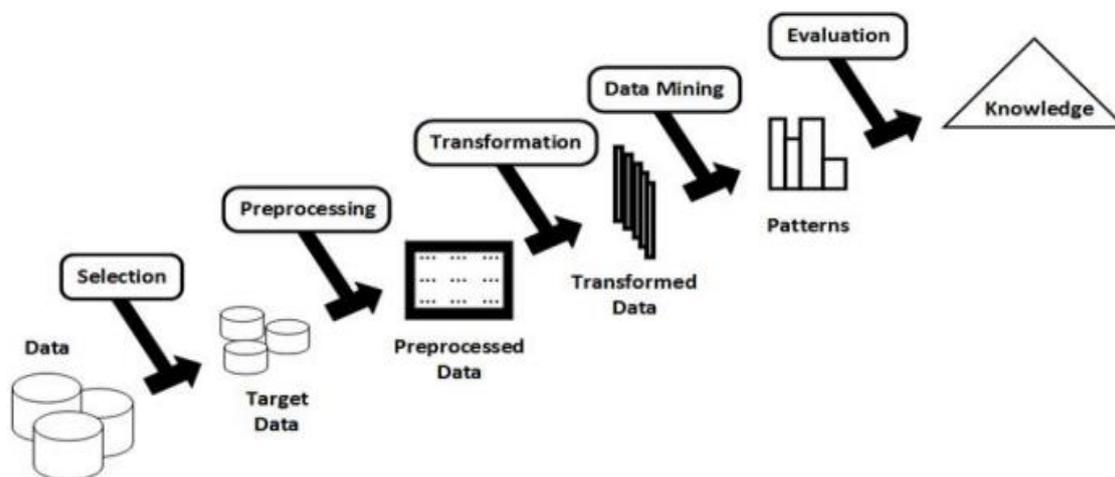
Tabel 6. Inisial prestasi

Prestasi	Inisial
Tidak ada prestasi	1
Prestasi Tk. Kota/kabupaten	2
Prestasi Tk. Provinsi	3
Prestasi Tk. Nasional	4
Prestasi Tk. Internasioanl	5

Data mining merupakan metode yang digunakan untuk mengekstrak informasi prediksi yang tersembunyi dalam *database*, ini merupakan sebuah teknologi yang memiliki

potensi besar untuk meningkatkan gudang data. *Data mining* memiliki arti yang berbeda, yaitu penemuan pengetahuan atau pengenalan pola yang digunakan untuk mengekstrak pengetahuan tersembunyi dari kumpulan data yang sangat besar (Noviyanto, 2020). Tujuan utama dari data mining yaitu untuk menggali, mencari, dan menelaah data yang ada sehingga dapat memberikan informasi sebagai sebuah pengetahuan (Ridwan et al., 2013). Secara garis besar, data mining dapat dibagi menjadi dua kategori utama:

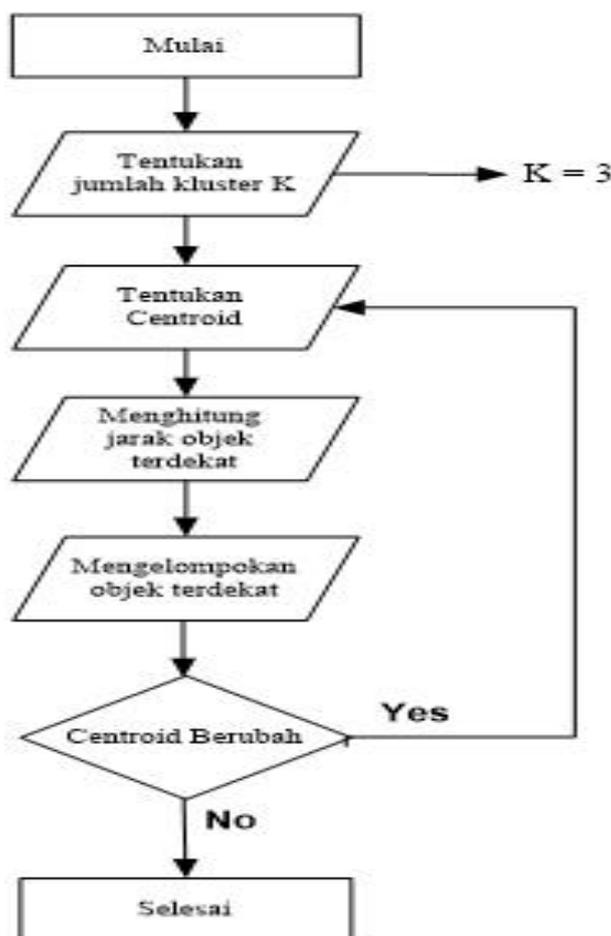
1. *Penambangan deskriptif*, Proses menemukan karakteristik penting dari data dalam database. Teknik penambangan data, termasuk penambangan deskriptif, adalah pengelompokan, asosiasi, dan penambangan sekuensial.
2. *Predictive*, Ini adalah proses menemukan pola dalam data menggunakan beberapa variabel lain di masa mendatang. Salah satu teknik yang terlibat dalam penambangan prediktif adalah klasifikasi. Sederhananya, *data mining* biasanya disebut sebagai proses dimana pengetahuan disaring atau "ditambang" dari sejumlah besar data istilah lain untuk data mining adalah *Knowledge Discovery Database (KDD)* (Darmi & Setiawan, 2017). Untuk Langkah-langkah *KDD* bisa dilihat pada Gambar 1.



Gambar 1. Langkah-langkah *Knowledge Discovery Database*

K-Means kadang disebut juga sebagai teknik analisis data atau teknik data mining yang menjalankan proses pemodelan data tanpa supervisi. Algoritma ini merupakan salah satu teknik untuk mengelompokkan data dalam kelompok atau sistem partisi yang disesuaikan dengan karakteristik masing-masing bagian data. Istilah pengelompokan dalam *k-means* digunakan untuk menggambarkan suatu algoritma yang menetapkan setiap elemen terhadap kluster dengan centroid terdekat (*mean*) dimulai dengan split atau memisahkan objek menjadi *k* kluster awal, lalu ditetapkan objek ke seluruh kluster *centroid* terdekat

biasanya dihitung berdasarkan jarak Euclidean (Annuril Chusna & Tuti Rumiati, 2020). Cara kinerja algoritma *K-Means* dapat dilihat pada Gambar 2.



Gambar 2. Flowchart algoritma metode *K-Means*

Pada gambar 3 memperlihatkan langkah-langkah dalam menentukan klusterisasi menggunakan metode algoritma *K-means*. Tahapan *clustering* menggunakan algoritma *K-means* sebagai berikut (Agustina & Prihandoko, 2018) (Harahap & Zulvia, 2021):

1. Masukkan data set
2. Menentukan jumlah *cluster* $k=3$
3. Ambil data secara acak sebanyak jumlah kluster sebagai pusat kluster (*centroid*).
4. Hitung jarak antara data dengan pusat kluster, dengan menggunakan persamaan :

$$d(x, y) = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1, 2, 3, \dots, n \quad (1)$$

dimana :

$d(x, y)$ adalah jarak antara data dengan pusat *cluster*

x_i adalah titik pusat *cluster* x ke i

y_i adalah data y ke i
 n adalah banyaknya atribut pada data set
 i adalah data atribut ke- i

5. Hitung kembali pusat cluster dengan keanggotaan cluster yang baru, Untuk menentukan pusat cluster (centroid) dari data yang ada pada masing-masing cluster yang didapatkan dari hasil persamaan berikut :

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1, 2, 3, \dots, n \quad (2)$$

dimana :

v adalah *centroid* pada *cluster*

x_i adalah nilai data atribut ke i

n adalah banyaknya data yang menjadi anggota *cluster*

Jika pusat *cluster* tidak berubah maka proses *clustering* telah selesai, jika belum maka ulangi langkah ke d sampai pusat *cluster* tidak berubah lagi.

Rapidminer adalah Perangkat lunak yang dikembangkan oleh Dr. Markus Hofmann dari Institut Teknologi Blanchardstown dan Ralf Klinkenberg dari *rapid-i.com*. Rapidminer menampilkan tampilan antarmuka pengguna yang berbasis grafis sehingga lebih mudah digunakan. Perangkat lunak ini berbasis open *source* ditulis menggunakan program Java di bawah Lisensi Publik GNU, rapidminer dapat berjalan di sistem operasi apa pun. Rapidminer memiliki sekitar 500 operator data mining seperti *input*, *output*, *preprocessing data*, operator visualisasi dan lain-lain (Sibuea & Safta, 2017). Pada penelitian ini penulis akan membandingkan hasil pengolahan data secara manual dan pengolahan data menggunakan aplikasi rapid miner studio versi 9.0.

HASIL DAN PEMBAHASAN

Metode penelitian yang dilakukan oleh penulis menggunakan metode algoritma *k-means*, perhitungan metode *k-means* dilakukan secara manual dengan menghitung kedekatan nilai dari setiap atribut dengan jarak pusat data (*cluster*). Pada proses perhitungan ini penulis menggunakan 3 klaster ($K=3$) yang diambil dari data set yang sudah ada. Perhitungan jarak kedekatan dalam penelitian ini menggunakan persamaan *Euclidean Distance*, dimana setiap nilai pada masing-masing atribut akan dihitung sampai dengan tidak ada nilai yang berpindah di setiap klasternya atau *konvergen*. Selanjutnya data set yang sudah ada akan dicoba dihitung kembali menggunakan aplikasi/*tools* rapid miner studio dengan menggunakan 3 klaster ($K=3$) seperti perhitungan manual untuk memastikan hasilnya.

Hasil penelitian dari pengelompokan data penerima bantuan biaya pendidikan menggunakan algoritma *K-Means* yang dilakukan secara manual dan *tools* rapid miner

didapatkan hasil yang sama yaitu kluster C0 sebanyak 109 data (57,1%), kluster C1 sebanyak 52 data (27,2%), dan kluster C2 sebanyak 30 data (15,7%).

Pengolahan data tentang penerimaan dana bantuan biaya pendidikan dihitung dengan cara manual terlebih dahulu, proses ini menggunakan 3 kluster yang diambil secara acak dari data set yang sudah ada. Untuk atribut data yang digunakan yaitu status DTKS, penerima kartu KIP dan KKS, penghasilan orang tua, status orang tua, jumlah tanggungan orang tua dan prestasi siswa. Untuk data atribut dapat dilihat pada Tabel 7.

Tabel 7. Dataset pendaftar penerima bantuan biaya pendidikan

No	Nama Pendaftar	Status DTKS	KIP & KKS	Penghasilan Orangtua	Status Orangtua	Jumlah Tanggungan	Prestasi Siswa
1	Dali Fadilah	2	1	6	1	2	1
2	Fitriyani	1	3	6	4	2	1
3	Farhan	2	1	6	1	1	1
.....
189	Azmi Ajkya	1	1	1	1	4	1
190	Firda Nabila	1	1	1	1	4	1
191	Silvi Silviani	2	3	6	1	6	1

Selanjutnya pemilihan kluster dari dataset yang ada untuk penentuan *centroid* secara acak/random sebagai awal perhitungan untuk menentukan kedekatan jarak antara data dan pusat *cluster*. Penentuan *centroid* secara acak/random dapat dilihat pada Tabel 8.

Tabel 8. Sample Centroid

Cluster	Nama Pendaftar	DTKS	KIP & KKS	Penghasilan Orangtua	Status Orangtua	Jumlah Tanggungan	Prestasi Siswa
C0	Hesti Sri Wulandari	1	1	5	1	4	1
C1	Muhammad Fahd	1	1	4	1	1	4
C2	Siti Nuraeni	1	1	4	1	5	1

Penghitungan jarak dari setiap data ke pusat *cluster* (C0, C1, C2) menggunakan persamaan *Euclidean Distance*, dimana setiap nilai yang ada dalam atribut dihitung secara keseluruhan dari data pertama sampai akhir. Perhitungan ini dilakukan untuk mendapatkan nilai *centroid* baru, perhitungan mendapatkan nilai centroid baru (C0, C1, C2) sebagai berikut :

1. Matriks mendapatkan nilai *centroid* C0

$$d = \sqrt{(1 - 2)^2 + (1 - 1)^2 + (5 - 6)^2 + (1 - 1)^2 + (4 - 2)^2 + (1 - 1)^2} = 6$$

$$d = \sqrt{(1 - 1)^2 + (1 - 3)^2 + (5 - 6)^2 + (1 - 4)^2 + (4 - 2)^2 + (1 - 1)^2} = 18$$

$$d = \sqrt{(1 - 2)^2 + (1 - 1)^2 + (5 - 6)^2 + (1 - 1)^2 + (4 - 1)^2 + (1 - 1)^2} = 11$$

dilanjut sampai data ke 191..

2. Matriks mendapatkan nilai centroid C1

$$d = \sqrt{(1 - 2)^2 + (1 - 1)^2 + (4 - 6)^2 + (1 - 1)^2 + (1 - 2)^2 + (4 - 1)^2} = 15$$

$$d = \sqrt{(1 - 1)^2 + (1 - 3)^2 + (4 - 6)^2 + (1 - 4)^2 + (1 - 2)^2 + (4 - 1)^2} = 27$$

$$d = \sqrt{(1 - 2)^2 + (1 - 1)^2 + (4 - 6)^2 + (1 - 1)^2 + (1 - 1)^2 + (4 - 1)^2} = 14$$

dilanjut sampai data ke 191..

3. Matriks mendapatkan nilai centroid C2

$$d = \sqrt{(1 - 2)^2 + (1 - 1)^2 + (4 - 6)^2 + (1 - 1)^2 + (5 - 2)^2 + (1 - 1)^2} = 14$$

$$d = \sqrt{(1 - 1)^2 + (1 - 3)^2 + (4 - 6)^2 + (1 - 4)^2 + (5 - 2)^2 + (1 - 1)^2} = 26$$

$$d = \sqrt{(1 - 2)^2 + (1 - 1)^2 + (4 - 6)^2 + (1 - 1)^2 + (5 - 1)^2 + (1 - 1)^2} = 21$$

dilanjut sampai data ke 191..

Hasil perhitungan yang dilakukan pada iterasi pertama menghasilkan data klaster pada C0, C1 dan C2, data yang termasuk pada klaster C0 sebanyak 145 data, C1 sebanyak 19 dan C2 sebanyak 27 data. Untuk data *iterasi* pertama dapat dilihat pada Tabel 9.

Tabel 9. Data *iterasi* pertama

No	Nama Pendaftar	C0	C1	C2	Jarak Terdekat	Kelompok Data
1	Dali Fadilah	6	15	14	6	C0
2	Fitriyani	18	27	26	18	C0
3	Farhan	11	14	21	11	C0
.....
189	Azmi Ajkya	16	27	10	10	C2
190	Firda Nabila	16	27	10	10	C2
191	Silvi Silviani	10	43	10	10	C0

Selanjutnya dilakukan perhitungan untuk mendapatkan *centroid* baru dari hasil pengelompokan data disetiap klasternya. Dalam proses mendapatkan nilai *centroid* baru dilakukan penjumlahan nilai data pada atribut data set dibagi banyaknya data yang menjadi anggota klaster. Dalam perhitungan ini jumlah data yang termasuk anggota klaster C0 yaitu sebanyak 145 data. Berikut perhitungan untuk mendapatkan *centroid* baru :

1. Perhitungan mendapatkan *centroid* di setiap atribut pada klaster C0

Status DTKS	$\frac{2+1+2+2+1+2+2+2+1+ \dots\dots\dots+1+1+2+2+2+1+2}{145} = 1,46$
KIP & KKS	$\frac{1+3+1+3+1+1+3+1+1+ \dots\dots\dots+1+1+3+3+3+1+3}{145} = 1,61$
Penghasilan orangtua	$\frac{6+6+6+6+5+6+6+6+5+ \dots\dots\dots+4+6+6+6+6+3+6}{145} = 5,61$
Status ortu	$\frac{1+4+1+1+1+1+1+2+1+ \dots\dots\dots+1+4+1+4+1+1+1}{145} = 1,55$
Jumlah tanggungan	$\frac{2+2+1+3+4+3+2+3+2+ \dots\dots\dots+2+3+3+2+3+3+6}{145} = 2,80$
Prestasi Siswa	$\frac{1+1+1+1+1+1+1+1+1+ \dots\dots\dots+1+3+1+1+1+1+1}{145} = 1,22$

2. Pusat *centroid* baru pada klaster C1 dan C2 dihitung seperti pada klaster C0, dimana nilai yang ada pada atribut data set dibagi dengan banyaknya data yang menjadi anggota klaster masing-masing. Untuk data yang menjadi anggota klaster C1 dalam perhitungan *iterasi* pertama sebanyak 27 data dan C2 sebanyak 19 data. hasil perhitungan nilai *centroid* baru pada iterasi pertama dapat dilihat pada Tabel 10.

Tabel 10. Nilai *centroid* baru pada *iterasi* pertama

Cluster	status DTKS	KIPK&KKS	Penghasilan orangtua	Status ortu	Jml Tanggungan	Prestasi
C0	1,46	1,61	5,61	1,55	2,80	1,22
C1	1,47	1,63	4,74	1,68	1,68	2,95
C2	1,04	1,33	2,04	1,00	3,78	1,22

Untuk menghitung *iterasi* ke- dua dan selanjutnya, dilakukan perhitungan seperti *iterasi* pertama sampai *konvergen* dimana nilai dari setiap *centroid* tidak ada yang berubah. Dalam penelitian ini nilai *konvergen* didapat pada *iterasi* ke 12, hasil dari perhitungan ini mendapatkan nilai pengelompokan C0 sebanyak 109, C1 sebanyak 52 dan C2 sebanyak 30 data. Data yang termasuk klaster C0 yaitu data ke- :

1, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 22, 23, 26, 27, 28, 29, 34, 36, 37, 42, 45, 46, 48, 51, 52, 53, 54, 56, 58, 59, 60, 61, 64, 65, 66, 67, 68, 75, 76, 79, 85, 92, 93, 94, 95, 98, 99, 100, 101, 102, 104, 106, 108, 110, 111, 112, 115, 116, 117, 118, 120, 122, 123, 124, 125, 127, 128, 130, 131, 132, 133, 134, 136, 138, 139, 142, 143, 144, 147, 148, 149, 150, 152, 156, 157, 158, 159, 163, 165, 168, 169, 170, 172, 173, 174, 175, 176, 178, 183, 184, 186, 191.

Data yang termasuk klaster C1 yaitu data ke- :

6, 18, 21, 25, 32, 35, 39, 40, 41, 50, 55, 62, 69, 70, 71, 73, 77, 78, 80, 84, 86, 87, 88, 89, 90, 91, 96, 97, 103, 105, 107, 121, 126, 135, 137, 140, 146, 154, 160, 161, 162, 164, 166, 167, 177, 179, 180, 182, 187, 188, 189, 190.

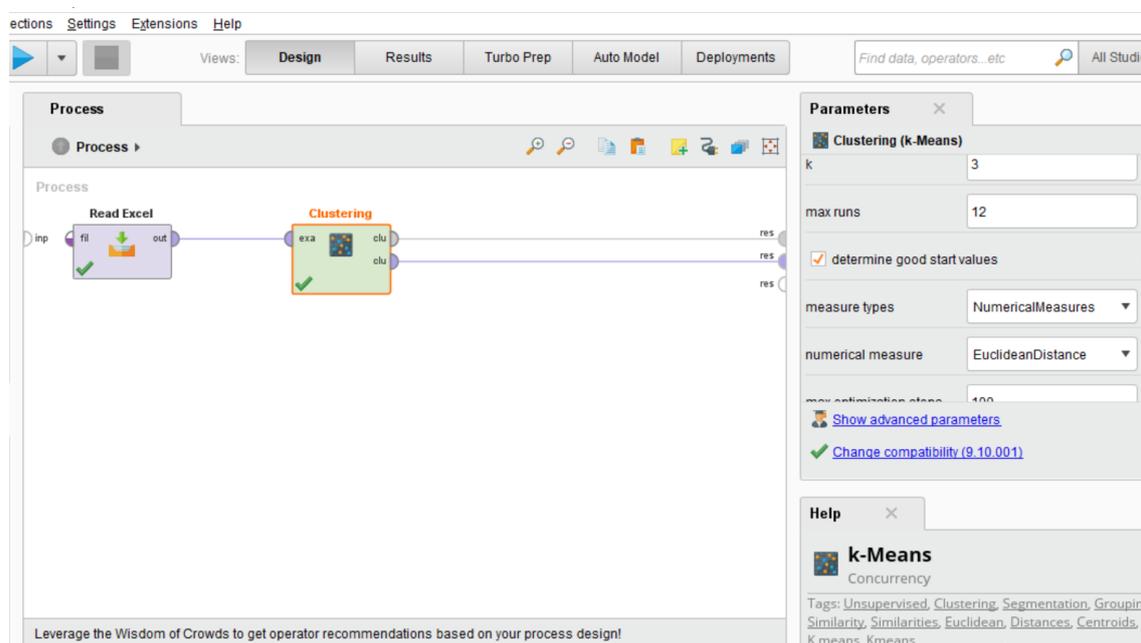
Data yang termasuk klaster C2 yaitu data ke- :

2, 24, 30, 31, 33, 38, 43, 44, 47, 49, 57, 63, 72, 74, 81, 82, 83, 109, 113, 114, 119, 129, 141, 145, 151, 153, 155, 171, 181, 185

Hasil pengelompokan data klaster C0 dari 109 data yang termasuk atribut data penghasilan orang tua terdapat sebanyak 78 data berpenghasilan <500.000 dan 31 data berpenghasilan >500.000. Atribut status orang tua menunjukkan bahwa sebanyak 99 data berstatus hidup, 10 data dengan status meninggal atau bercerai. Pada klaster C1 sebanyak 52 data yang termasuk pada atribut penghasilan orang tua dengan status penghasilan >1.000.000 sebanyak 40 data, penghasilan >2.000.000 sebanyak 12 data. Atribut status orang tua

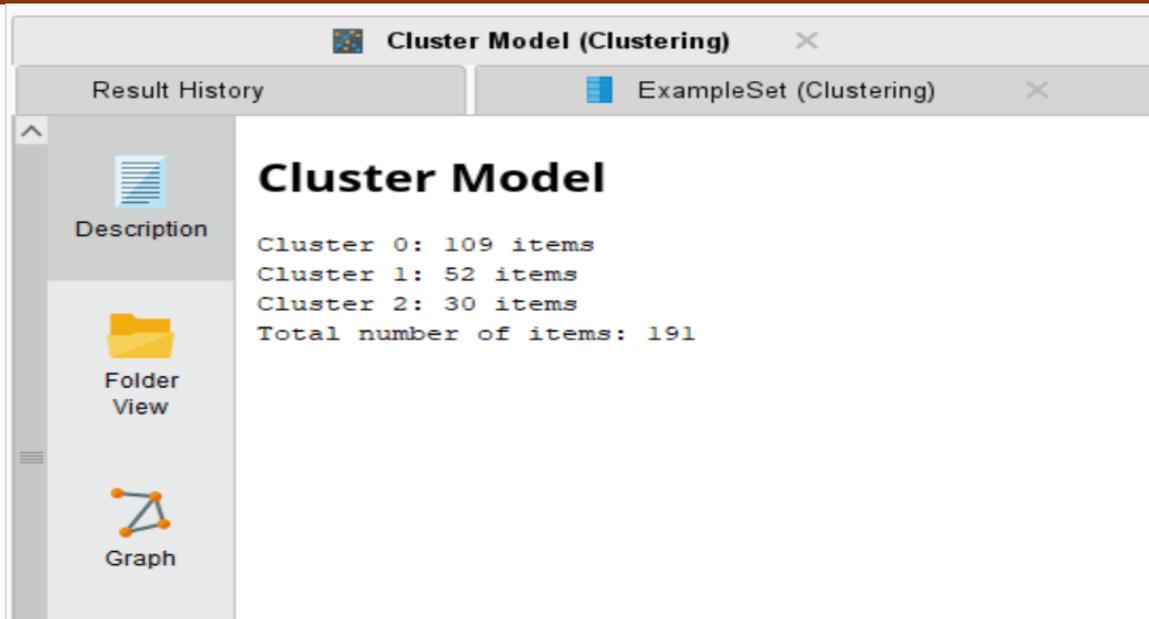
menunjukkan sebanyak 52 dengan status hidup (ayah&ibu). Pada kelompok C2 berjumlah 30 data, yang termasuk pada atribut penghasilan orang tua dengan status tidak berpenghasilan sebanyak 26 data dan 4 data berpenghasilan <750.000. Atribut status orang tua menunjukkan sebanyak 29 data berstatus meninggal (ayah/ibu) atau bercerai dan 1 data berstatus meninggal (ayah&ibu). Melihat dari sebaran data pada masing-masing kelompok, penulis merekomendasikan klaster C0 sebagai data kelompok yang dipertimbangkan, klaster C1 sebagai kelompok yang tidak layak dan klaster C2 sebagai kelompok yang layak sebagai data penerima bantuan biaya pendidikan / KIP-Kuliah.

Setelah mendapatkan hasil analisis pengelompokan data pendaftar penerima bantuan biaya pendidikan / KIP-Kuliah secara manual menggunakan excel, selanjutnya penulis melakukan analisis data menggunakan tools Rapidminer Studio. Data yang akan di olah menggunakan *tools* rapidminer sama dengan data yang di olah secara manual. Dalam pemodelan *k-means* menggunakan rapidminer dengan pengaturan $K = 3$, $max\ run = 12$, $measure\ types = NumricalMeasures$, $Numrical\ Measures = EuclideanDistance$ dan $max\ optimization\ steps = 100$, pengolahan data menggunakan *tools* rapid miner dapat dilihat pada Gambar 3.



Gambar 3. Pengolahan data menggunakan tools rapid miner

Dari hasil pengolahan data didapatkan pengelompokan sebanyak 109 data, 52 data dan 30 data. Hasil dari perhitungan pengelompokan data menggunakan *tools* Rapidminer dan perhitungan manual menggunakan Ms.Excel didapatkan hasil yang sama di setiap klasternya. Hasil pemodelan data dan *example set* dapat dilihat pada Gambar 4 dan Gambar 5.



Gambar 4. Hasil pemodelan data menggunakan rapid miner



Gambar 5. Example set pengelompokan disetiap cluster

KESIMPULAN DAN REKOMENDASI

Dari hasil pembahasan dapat disimpulkan bahwa pengelompokan data penerima bantuan biaya pendidikan sebanyak 191 data dengan menggunakan metode k-means menghasilkan data yang termasuk pada klaster C0 sebanyak 109 data (57,1%), C1 sebanyak 52 data (27,2%), dan C2 sebanyak 30 data (15,7%). Hasil pengelompokan data klaster C0 yang termasuk atribut data penghasilan orang tua terdapat sebanyak 78 data berpenghasilan <500.000 dan 31 data berpenghasilan >500.000. Atribut status orang tua menunjukkan bahwa sebanyak 99 data berstatus hidup, 10 data dengan status meninggal atau bercerai. Pada

klaster C1 yang termasuk pada atribut penghasilan orang tua dengan status penghasilan >1.000.000 sebanyak 40 data, penghasilan >2.000.000 sebanyak 12 data. Atribut status orang tua menunjukkan seluruh data berstatus hidup (ayah&ibu). Pada kelompok C2 yang termasuk pada atribut penghasilan orang tua dengan status tidak berpenghasilan sebanyak 26 data dan 4 data berpenghasilan <750.000. Atribut status orang tua menunjukkan sebanyak 29 data berstatus meninggal (ayah/ibu) atau bercerai dan 1 data berstatus meninggal (ayah&ibu). Melihat dari sebaran data pada masing-masing kelompok, penulis merekomendasikan klaster C2 sebagai data kelompok layak, klaster C0 sebagai kelompok yang dipertimbangkan dan klaster C1 sebagai kelompok yang tidak layak sebagai penerima bantuan biaya pendidikan / KIP-Kuliah.

REFERENSI

- Agustina, N., & Prihandoko, P. (2018). Perbandingan Algoritma K-Means dengan Fuzzy C-Means Untuk Clustering Tingkat Kedisiplinan Kinerja Karyawan. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 2(3), 621–626. <https://doi.org/10.29207/resti.v2i3.492>
- Annuril Chusna, H., & Tuti Rumiati, A. (2020). Penerapan Metode K-Means dan Fuzzy C-Means untuk Pengelompokan Sekolah Menengah Pertama (SMP) di Indonesia Berdasarkan Standar Nasional Pendidikan (SNP). 9(2), 216–223.
- Asahar Johar. (2019). Implementasi Algoritme Hard K- Means Clustering Dalam Penentuan Masa Pensiun (Studi Kasus: Badan Kepegawaian Daerah Provinsi Bengkulu). 7(2), 179–186. <http://ejournal.unib.ac.id/index.php/rekursif/>
- Darlinda, & Utamajaya, N. (2022). Sistem Pendukung Keputusan Penerima Beasiswa Program Indonesia Pintar Menggunakan Metode Algoritma K-Means Clustering. *JURIKOM (Jurnal Riset Komputer)*, 9 (2), 167–175. <https://doi.org/10.30865/jurikom.v9i2.3971>
- Darmi, Y. D., & Setiawan, A. (2017). Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. *Jurnal Media Infotama*, 12(2), 148–157. <https://doi.org/10.37676/jmi.v12i2.418>
- Fadhilah, A. M., Wahyuddin, M. I., & Hidayatullah, D. (2020). Analisis Faktor yang Mempengaruhi Perokok Beralih ke Produk Alternatif Tembakau (VAPE) menggunakan Metode K-Means Clustering. *Jurnal JTik (Jurnal Teknologi Informasi Dan Komunikasi)*, 5(2), 219. <https://doi.org/10.35870/jtik.v5i2.182>
- Harahap, A. S., & Zulvia, P. (2021). Klasterisasi Desa dengan Menggunakan Algoritma K-

- Means pada Data Potensi Desa. *JURIKOM (Jurnal Riset Komputer)*, 8(6), 237–246.
<https://doi.org/10.30865/jurikom.v8i6.3724>
- Kusnasari³, S. N. B. S. H. W. S. (2014). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Paper Knowledge . Toward a Media History of Documents*, 1, 82–86. <https://ojs.trigunadharma.ac.id/index.php/jsi/article/download/4784/320>
- Noviyanto, N. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Paradigma - Jurnal Komputer Dan Informatika*, 22(2), 183–188. <https://doi.org/10.31294/p.v22i2.8808>
- Rahmah, S. A., & Antares, J. (2021). *Klasterisasi Seleksi Mahasiswa Calon Penerima Beasiswa Yayasan Menggunakan K-Means Clustering*. 13(2), 25–30.
- Ridwan, M., Suyono, H., & Sarosa, M. (2013). Penerapan Data Mining Untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Eeccis*, 7(1), 59–64. <https://doi.org/10.1038/hdy.2009.180>
- Rohaeni, N. E., & Saryono, O. (2018). Implementasi Kebijakan Program Indonesia Pintar (PIP) Melalui Kartu Indonesia Pintar (KIP) dalam Upaya Pemerataan Pendidikan. *Journal of Education Management and Administration Review*, 2(1), 193–204.
- Setiawan, R. (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.
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteks*, 4(1), 85–92. <https://doi.org/10.33330/jurteks.v4i1.28>
- Wardhani, A. K. (2016). Implementasi Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Kajen Pekalongan. *Jurnal Transformatika*, 14(1), 30–37.