

KLASTERISASI DATA DISABILITAS MENGGUNAKAN ALGORITMA K-MEANS

M. Atiqul Mutaqin¹, Gunawan², Wresti Andriyani³

¹Program Studi Teknik Informatika, STMIK YMI TEGAL,

atiqulmm1105@gmail.com, gunawan.gayo@gmail.com, wresty.andriyani@gmail.com

Jl. Pendidikan No.1, Pesurungan Lor, Kec. Margadana, Kota Tegal Jawa Tengah 52122

Keywords:

disability, students, clustering, data mining, k-means algorithm

Abstract

Persons with disabilities are people who have physical or mental limitations that can affect them and interfere with normal daily activities. People with special needs (disabled) live with special characteristics and are different from the general public. So far, the placement of classes for students with disabilities is based on the child's entry age when registering at SLB N Tegal City, no Intelligence Quotient (IQ) test has been tried in grouping. This study aims to determine the results of class clustering of students with disabilities. In this study using data mining techniques with the K-Means clustering method. The K-Means method is a clustering method that is useful for breaking datasets into several groups. The source of this research data was obtained from SLB N Tegal City. The data entered is a sample of disability class data for 2021-2022 which is divided into 3 clusters, namely light, medium and severe clusters with an Intelligence Quotient (IQ) assessment. From the results of the k-means calculation, 11 students were found as a light cluster, 9 students as a medium cluster and 10 students as a heavy cluster. The clustering process uses excel and the RapidMiner application is used to help find accurate values.

Kata Kunci:

disabilitas, siswa, klasterisasi, data mining, algoritma k-mean.

Abstrak

Penyandang disabilitas adalah seseorang yang memiliki keterbatasan fisik atau mental yang dapat mempengaruhinya dan mengganggu aktivitas normal sehari-hari. Orang berkebutuhan khusus (disabilitas) hidup dengan ciri khusus dan berbeda dengan masyarakat umum. Selama ini penempatan kelas siswa disabilitas bersumber pada umur masuk anak dikala mendaftar di SLB N Kota Tegal, tidak dicoba uji Intelligence Quotient (IQ) dalam mengelompokkan. Penelitian ini bertujuan untuk mengetahui hasil pengklasteran kelas siswa disabilitas. Dalam penelitian ini menggunakan teknik data mining dengan metode K-Means clustering. Metode K-Means merupakan metode clustering yang berguna untuk memecah dataset menjadi beberapa kelompok. Sumber data penelitian ini diperoleh dari SLB N Kota Tegal. Data yang dimasukkan adalah data sample kelas disabilitas tahun 2021-2022 yang terbagi menjadi 3 klaster yaitu klaster ringan, sedang dan berat dengan penilaian Intelligence Quotient (IQ). Dari hasil perhitungan k-means didapatkan 11 siswa sebagai klaster ringan, 9 siswa sebagai klaster sedang dan 10 siswa sebagai klaster berat. Proses klasterisasi menggunakan excel dan aplikasi RapidMiner digunakan untuk membantu menemukan nilai yang akurat.

Pendahuluan

Anak berkebutuhan khusus atau biasa disebut disabilitas merupakan anak yang memiliki keterbatasan fisik atau mental yang dapat mempengaruhi mereka dan mengganggu aktivitas normal sehari-hari. Tidak semua anak diciptakan dalam kondisi fisik atau mental yang sempurna. Hal ini membuat sebagian pada diri mereka merasa minder atau rendah diri dalam pergaulan. Masalah penyandang disabilitas merupakan masalah yang sangat kompleks dengan keterbatasan. Tentu saja menimbulkan masalah sosial karena keterbatasan fungsi fisik yang tidak sempurna. Kurangnya kemampuan ini dapat menghambat penyandang disabilitas untuk melakukan aktivitas sehari-hari. Orang berkebutuhan khusus (disabilitas) hidup dengan ciri khusus dan berbeda dengan masyarakat umum. Karena karakteristik yang berbeda tersebut, diperlukan pelayanan khusus untuk mendapatkan hak atas kehidupan yang lebih baik. Oleh karena itu, dalam rangka pengelompokan penyandang disabilitas perlu dilakukan di SLB N Kota Tegal [1].

Kota Tegal merupakan sebuah kota dengan 4 kecamatan 27 kelurahan Populasinya adalah 280.940 Jiwa laki-laki Dan wanita. Berdasarkan survei di Kota Tegal terdapat penyandang disabilitas, Jenis kecacatannya berbeda-beda. Penelitian di SLB N Kota Tegal yang merupakan salah satu sekolah luar biasa yang dijadikan sebagai jalur pendidikan bagi anak berkebutuhan khusus di Kota Tegal. Kriteria penentuan tingkat siswa membuat sulit untuk membagi kelas sesuai dengan hasil observasi yang dilakukan oleh SLB N Kota Tegal. Permasalah tersebut sering terjadi di dalam SLB N Kota Tegal, dimana pembagian kelas tidak pada jenis disabilitas yang sama. Dengan begitu siswa – siswi kesulitan saat mengikuti kegiatan belajar mengajar. Hal ini di akibatkan karena wali kelas sering kali lalai saat menentukan pembagian kelas yang tidak sesuai tersebut, tentu hal tersebut mengakibatkan kegiatan belajar mengajar kurang efektif.

Masalah – masalah tersebut bisa di antisipasi dengan cara wali kelas membagi kelas siswa dengan jenis disabilitas yang sama dalam satu kelas, agar guru mata pelajaran yang memberikan materi tidak memakan waktu untuk menjelaskan materi ke siswa – siswi menggunakan bahasa isyarat yang berbeda - beda. Dengan begitu siswa – siswi akan lebih mudah dalam melakukan kegiatan belajar mengajar. Namun, hal tersebut seringkali merepotkan para wali kelas karena harus mengelompokkan ulang siswa - siswi dengan cara manual yaitu dengan cara melihat ulang data penyakit bawaan siswa apakah siswa tersebut masuk ke kelas jenis mana. Oleh karena itu, dilakukan penelitian untuk mengklasifikasikan data siswa penyandang disabilitas dan menentukan kelas mana yang mereka tempati sehingga pihak sekolah dapat mempersiapkannya. Maka dari itu diperlukan adanya sebuah sistem klasterisasi anak disabilitas di SLB N Kota Tegal untuk memudahkan para wali kelas mengelompokkan siswa - siswi pada kelas yang sama lalu dengan demikian memudahkan wali kelas untuk memberikan *treatment* berdasarkan jenis disabilitas. Dengan begitu wali kelas tidak melakukan cara manual yang dinilai kurang efektif dan terkesan ribet.

Data mining merupakan tahap dalam pengklasifikasian terhadap data dengan menghubungkan masing-masing pola pada setiap data set yang berukuran besar dengan jumlah data yang besar pula. Data mining dapat juga didefinisikan menggali data dari banyaknya informasi yang akan dicari sehingga data yang perlu diketahui akan lebih mudah dicari dengan adanya sistem pola yang dibuat berdasarkan titik terdekat dengan informasi yang sering di perlukan. Pemilihan metode atau algoritma dari data mining juga perlu rasional, karena sangat bergantung pada keseluruhan tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan [2]. Secara teknis, data mining dapat disebut sebagai proses untuk menemukan aturan atau pola dari ratusan atau ribuan data dari sebuah relasi basis data yang sangat besar [3].

Algoritma K-Means adalah metode yang berguna untuk mengelompokkan data berdasarkan pencarian iterative dari pusat *cluster*, mendeteksi bahwa lokasi setiap cluster data adalah jarak minimum semua data dari pusat *cluster*. Algoritma K-Means dari teknik clustering membantu mengelompokkan siswa dengan nilai yang termasuk dalam kelas ringan, sedang, atau berat. Metode yang digunakan dalam penelitian ini adalah kuantitatif dan penelitian ini bersifat sistematis dengan memperhatikan beberapa bagian dan hubungannya.

Algoritma K-Means terkenal karena kesederhanaannya dan kemampuannya untuk mengelompokkan data yang besar dan outlier dengan sangat cepat. Algoritma klasterisasi K-Means dapat diterapkan kelas penyandang disabilitas , sehingga dapat diidentifikasi kelompok berdasarkan data tersebut. Karena ciri-cirinya dapat dilihat dari pengelompokan data, maka dapat diketahui bahwa jumlah

penyandang disabilitas di setiap kelas yang mempunyai kluster berat, kluster sedang, dan kluster ringan [4].

Berdasarkan permasalahan diatas, maka penulis mengambil judul “Klasterisasi Data Disabilitas Menggunakan Algoritma K-Means”. Diharapkan penelitian ini dapat dijadikan masukkan kepada Sekolah Luar Biasa Negeri Kota Tegal agar mempermudah para wali kelas mengelompokkan siswa - siswi pada kelas yang sama lalu dengan demikian memudahkan wali kelas untuk memberikan pengajaran berdasarkan jenis disabilitasnya.

Landasan Teori

Berikut beberapa penelitian yang digunakan penulis sebagai bahan referensi dalam melakukan penelitian, penelitian tentang klasterisasi, penelitian tentang disabilitas, penelitian tentang algoritma k-Means. Penelitian oleh Amir Ali tentang *Clustering* Data Antropometri Balita untuk Menentukan Status Gizi Balita di Kelurahan Jumput Rejo Sukodono Sidoarjo menghasilkan Menarik kesimpulan yang dapat dijadikan acuan oleh bidan Puskesmas dan tenaga kesehatan terlatih dalam sosialisasi dan pencegahan gizi buruk di bawah usia 5 tahun berdasarkan tinggi dan berat badan bayi [5].

Penelitian tentang klasterisasi data pertanian Kabupaten Lamongan menggunakan algoritma K-Means dan mean fuzzy C oleh Arif Rohmatullah, Dinita Rahmalia, dan Mohammad Syaiful Pradana. Berdasarkan kesimpulan, beberapa kecamatan dapat dikelompokkan ke dalam Kabupaten Lamongan. Hasil klasterisasi beberapa kecamatan di kawasan Lamongan dapat memberikan informasi potensi pertanian kecamatan tersebut dan memungkinkan untuk dikembangkan penelitian lebih lanjut dengan menggunakan data yang lebih dimensional [6].

Dalam penelitian Fitriana Harahap membandingkan algoritma K-Means dan K-Medoids untuk pengelompokan kelas siswa tunagrahita, dua algoritma pengelompokan, K-Means dan K-Medoids, diuji pada dataset kecil. K-Means lebih efektif dalam memproses data kecil. Dataset keterbelakangan mental menggunakan K-Means menunjukkan nilai DBI sebesar 0,161 dan skor menggunakan K-Medoids menunjukkan skor sebesar 0,281. Ini karena K-Means memiliki daya komputasi yang lebih sedikit daripada K-Medoids [7].

Achmad Solichin and Khansa Khairunisa (2020), tentang klasterisasi penyebaran virus corona (Covid-19) di DKI Jakarta menggunakan metode K-Means. Hal ini bertujuan untuk menjumlahkan penyebaran virus corona di Provinsi DKI Jakarta berdasarkan parameter berupa ODP, PDP, kasus positif, pasien sembuh dan pasien meninggal. Clustering ini dilakukan dengan menggunakan metode K-Means, yaitu mengelompokkan data menjadi beberapa cluster berdasarkan kesamaan data. Pemanfaatan data penyebaran virus corona secara clustered menggunakan algoritma clustering K-Means untuk menekan penyebaran virus corona dan menentukan jumlah kasus secara tepat untuk meminimalisir positif COVID-19 Diharapkan dapat membantu 19 pasien. Metode yang digunakan dalam penelitian ini adalah K-means clustering menggunakan metode pengumpulan dataset berupa ODP, PDP dan data positif Covid-19. Saat menentukan titik awal untuk sebuah cluster atau centroid, ditentukan secara acak berdasarkan jumlah cluster dan jumlah data yang akan diproses. Penelitian ini menghasilkan kesimpulan bahwa metode k-means dapat digunakan dengan benar. Selain itu, prototipe yang dikembangkan dapat menyajikan hasil klasterisasi data kasus Covid-19 di DKI Jakarta yang ditampilkan per kecamatan [8].

Ade Irma Purnamasari, Kajian Radita Dinar Dana tentang analisis penggemar kopi berbasis rasa kopi menggunakan teknik clustering algoritma k-means di Cirebon menemukan bahwa dengan menggunakan teknik clustering algoritma k-means rata-rata usianya adalah 23-30 tahun. Teknik clustering algoritma K-Means untuk menganalisis pecinta kopi berdasarkan preferensi kopi memungkinkan Anda untuk terlebih dahulu mengelompokkan data spasial melalui proses normalisasi. Berdasarkan ini, jarak terdekat dengan objek tidak didasarkan. Mengenai sifat benda, dan hasil pengujian, dapat disimpulkan bahwa rasa kopi lebih disukai oleh masyarakat Cirebon [9].

Miftahul Hasanah, dkk (2019), dalam penelitian tentang *Implementasi Algoritma K-Means untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA*, penelitian ini dilakukan untuk mencari kelompok nilai akademik siswa dengan cara menghitung jarak terpendek antara data dengan titik tengah (*centroid*) pada suatu *cluster*. Sehingga pada pengelompokan tersebut didapatkan beberapa *cluster* siswa yaitu sangat berkompeten, berkompeten dan kurang berkompeten untuk mengikuti Olimpiade Sains Nasional (OSN). Melalui penelitiannya dengan menggunakan algoritma *K-Means*

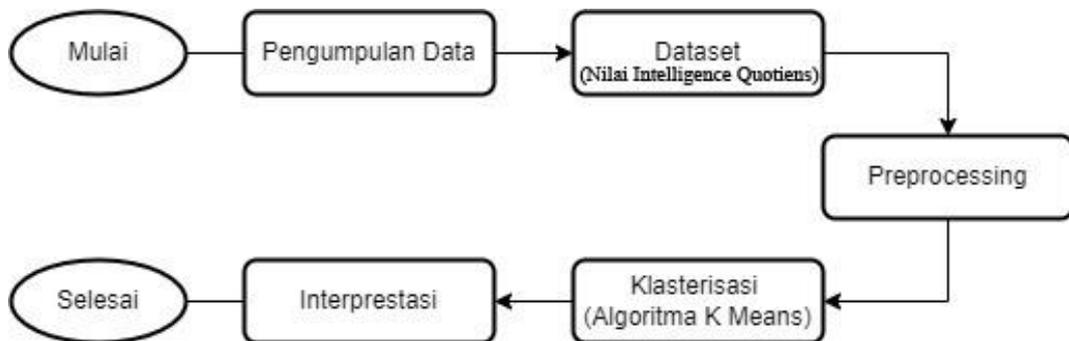
Clustering diharapkan mampu membantu pihak sekolah dan guru dalam pengelompokan siswa berdasarkan nilai akademik dengan hasil yang efektif dan lebih efisien. Metode yang digunakan dalam penelitian tersebut menggunakan kerangka penelitian berupa, identifikasi masalah, tujuan penelitian, mengumpulkan data, mengolah data dengan algoritma *K-Means Clustering* menguji hasil pengolahan data dan menyimpulkan hasil pengujian [10]

Penelitian Arora & Badal, (2013) tentang “Evaluating Student’s Performance Using K-Means Clustering”, menggunakan algoritma k-means karena dinilai dapat dengan cepat dan mudah membantu memantau perkembangan kinerja mahasiswa di suatu instansi pendidikan. Jumlah data yang dianalisis adalah 118 data siswa untuk mendapatkan rata-rata siswa setiap semester. Metode ini dapat berperan penting dalam analisis akademik untuk mengidentifikasi penyebab rendahnya kinerja akademik pada semester tertentu guna mengambil tindakan untuk meningkatkan kinerja akademik pada semester berikutnya [11].

Metode

Penelitian ini membahas tentang metode dan alur penelitian yang diterapkan, berikut tahapan yang dilakukan dalam penelitian yaitu:

Gambar 1. Tahapan Penelitian



1. Pengumpulan Data

Pengumpulan Data Pada tahap ini peneliti melakukan pengumpulan data yang dibutuhkan dalam penelitian yaitu data Disabilitas SLB N Kota Tegal. Untuk mengumpulkan data dalam kegiatan penelitian ini penulis menggunakan beberapa metode pengumpulan data, yaitu :

- a. Pengamatan (Observasi)
Pengamatan (observasi) yang dimaksud dalam penelitian ini adalah melakukan penelitian secara langsung dengan tujuan untuk lebih memahami dan mengetahui langkah-langkah apa saja yang harus diambil dalam menyelesaikan permasalahan yang ditemukan.
- b. Wawancara
Merupakan pengumpulan data dengan cara datang ke lokasi penelitian untuk melakukan tanya jawab langsung dengan pihak SLB N 1 Kota Tegal Kota Tegal, untuk memperoleh informasi yang dibutuhkan.
- c. Penelitian Kepustakaan
Untuk menunjang penelitian yang akan dilakukan, penulis melakukan Library research yakni dengan mencari data-data dari buku maupun jurnal penelitian sejenis yang berhubungan dengan metode yang penulis gunakan dalam penelitian ini yaitu metode Clustering K-Means.

2. Dataset

Tahap pertama adalah melakukan penginputan data, pada Penelitian ini Klasterisasi Disabilitas dilakukan menggunakan dataset Siswa SLB N Kota Tegal. Jenis dan sumber data ini yaitu Data Primer yang diperoleh secara langsung dari narasumber atau responden melalui wawancara.

3. Preprocessing

Data preprocessing adalah proses yang mengubah data mentah ke dalam bentuk yang lebih mudah dipahami. Proses ini penting dilakukan karena data mentah sering kali tidak memiliki

format yang teratur. Selain itu, data mining juga tidak dapat memproses data mentah, sehingga proses ini sangat penting dilakukan untuk mempermudah proses berikutnya, yakni analisis data. Ada beberapa tahapan yang perlu dilakukan ketika akan melakukan data preprocessing. Berikut ini beberapa tahapannya:

- a) Data cleaning
Tahap pertama yang perlu dilakukan ketika akan preprocessing data adalah data cleaning atau membersihkan data. Artinya, data mentah yang telah peroleh perlu diseleksi kembali. Kemudian, hapus atau hilangkan data-data yang tidak lengkap, tidak relevan, dan tidak akurat. Dengan melakukan tahap ini, akan menghindari kesalahpahaman ketika menganalisis data tersebut.
- b) Data integration
Karena data preprocessing akan menggabungkan beberapa data dalam suatu dataset, maka harus mengecek data-data yang datang dari berbagai sumber tersebut supaya memiliki format yang sama. Proses ini menjadi salah satu step penting.
- c) Transformasi data
Proses berikutnya yang harus dilakukan adalah transformasi data. Seperti yang telah dijelaskan di atas, data akan diambil dari berbagai sumber yang kemungkinan memiliki perbedaan format. Anda harus menyamakan seluruh data yang terkumpul supaya dapat mempermudah proses analisis data. Transformasi data untuk mengubah data kedalam bentuk yang sesuai untuk suatu analisis, dalam hal ini data yang menunjukkan jenis disabilitasnya diubah kedalam bentuk numerik. Disini untuk kelas berat (C1), sedang (C2), ringan (C3) menggunakan nilai *Intelligence Quotient* (IQ).
- d) Mengurangi data
Tahap terakhir yang perlu dilakukan adalah mengurangi jumlah data (data reduction). Maksudnya adalah Anda harus mengurangi sampel data yang diambil, tetapi dengan catatan, tidak akan mengubah hasil analisis data. Pada tahap ini untuk mendapatkan suatu data yang baik maka dilakukan tahap seleksi data. Seleksi data dilakukan yaitu merubah beberapa data yang bertujuan memudahkan pemahaman dengan mengacu dalam kesetabilan data, data yang tidak konsisten atau bertumpukan. Apabila sudah selesai maka didapatkan data yang sudah diproses atau data matang yang siap untuk ke tahap selanjutnya.

4. Klasterisasi

Pada tahap K-Means *Clustering* yaitu mengelompokkan data yang sudah ada ke dalam tiga kelompok yaitu berat, sedang dan ringan.

Pengertian Klasterisasi atau pengelompokan data adalah suatu tahapan untuk menggolongkan himpunan data yang atribut kelasnya belum di deskripsikan, secara konsep clustering adalah untuk memaksimalkan dan meminimalkan kesamaan antar kelas. Misalnya, dengan ada suatu himpunan obyek, proses pertama dapat di klasterisasi menjadi beberapa himpunan kelas selanjutnya menjadi sebuah himpunan beraturan sehingga dapat diturunkan berdasarkan kelompok klasifikasi tertentu. Cluster juga dapat diartikan sebagai kelompok. Maka analisa clustering pada dasarnya menghasilkan beberapa cluster (kelompok). Sebelum kita melakukan analisa perlu diterapkan pemahaman bahwa suatu himpunan dari data tertentu sebenarnya sudah memiliki kemiripan di antara anggotanya tersebut. Oleh karena itu, setiap anggota dengan karakteristik yang sama dikelompokkan ke dalam satu atau lebih dari suatu kelompok. Tujuan dari pada clustering data adalah untuk meminimalisasikan fungsi tujuan yang ditetapkan didalam proses clustering, dan umumnya selalu meminimalisasi variasi suatu cluster dan memaksimalisasi kan variasi antar cluster [12].

Algoritma K-Means *Clustering* adalah metode penganalisaan Data Mining dimana proses pemodelan tanpa pengawasan, dan cara mengelompokkan data berdasarkan partisi. Data yang dikelompokkan metode K-Means menjadi beberapa kelompok dan setiap kelompok memiliki karakteristik yang mirip atau sama dengan lainnya tetapi dengan kelompok lainnya memiliki karakteristik berbeda. Dengan tujuan meminimalisasi perbedaan setiap data didalam satu cluster serta memaksimalkan perbedaan dengan cluster yang lain [13].

Langkah – langkah yang dilakukan untuk membuat *clustering* pada metode *K-Means* adalah : [14]

- a) Menentukan jumlah *cluster* (k) dari data set yang digunakan.
- b) Menentukan nilai k sebagai *Centroid*, biasanya dilakukan secara acak.
- c) Menghitung jarak antara data dengan *centroid* menggunakan rumus *Euclidean* (persamaan 1)

$$\text{Distance: } (x_i, \mu_j) = \sqrt{\sum (x_i, \mu_j)^2} \dots\dots\dots(1)$$

Dimana :

- d = titik dokumen
- x_i = data kriteria
- μ_j = centroid pada *cluster* ke-j

- d) Kelompokkan data berdasarkan nilai yang paling dekat dengan *centroid*, kemudian berbarui nilai *centroid* baru dengan lokasi dari pusat cluster menggunakan persamaan kedua. *Euclidean* (persamaan 2)

$$\mu_j (t + 1) = \frac{1}{N_{sj}} \sum_{j \in s_j} x_j \dots\dots\dots(2)$$

Dimana :

- $\mu_j (t+1)$ = centroid baru pada iterasi ke (t+1)
- N_{sj} = banyaknya data pada *cluster* sj

- e) Lakukan langkah ke 2 sampai ke 4 sampai setiap cluster tiap anggota tidak ada yang berubah.

5. Interpretasi

Interprestasi yaitu tahapan terakhir yang membahas hasil analisis *Clustering* data disabilitas dari *cluster* berat, sedang dan ringan dengan nilai *Intelligence Quotient* (IQ).

Hasil dan Pembahasan

Dataset penyandang disabilitas yang diperoleh dari Sekolah Luar Biasa Negeri Kota Tegal tahun ajaran 2021-2022 menggunakan nilai rata-rata, sudah dalam bentuk file spreadsheet, sudah bersih dan tidak perlu melakukan transformasi lagi. Sehingga tahapan selanjutnya yaitu mining klasterisasi algoritma K-Means, untuk mengetahui siswa mana yang masuk ke dalam Klaster berat, Klaster sedang dan Klaster ringan dengan nilai *Intelligence Quotient* (IQ).

Berikut adalah langkah penyelesaian dalam mengelompokkan penyandang disabilitas menggunakan algoritma k-means:

- 1. Menentukan dataset yang akan di klaster

Tabel 1. Data Sample Siswa SLB.

NO	NIS	NAMA SISWA	NILAI RATA-RATA
1	2019C020746	Tria Novitasari	85,1
2	2019C020747	Mohamad Fadhil Imani	71,55
3	2019C020748	Fitri Indriani	75,84
4	2019C020749	Khalifah Nur Artha	68,91
5	2019C020750	Mohamad Fikri Al Rasid	80,74
6	2019C020751	Hud Ibnu Neno	75,62
7	2019C020752	Dian Ayu Safitri	74,86
8	2020C010793	Rizki Akbar Ramadhani	81,42
9	2020C010794	Citra Nur Maulidina	70,61
10	2021C010818	Choerul Umam Muslih	85,57
11	2021C010808	Sultan Aufan Widodo	72,45
12	2021C010809	Fauzio Al Fariz	82,54

13	2021C010810	Syahla Nur Alisyia Erliantoro	86,71
14	2021C010811	Fachri Maulana Firdaus	75,14
15	2021C010812	Safi Andhika Pratama	85,23
16	2021C010813	Wildan Kinnan Afrizal	69,51
17	2021C010814	M. Aenun Najib	74,83
18	2021C010815	Bayu Kurniawan	72,55
19	2020C010779	Artanta Putra S	78,45
20	2020C010780	M Faiz Afandi	80,15
21	2020C010781	Ikhwan Nadhif	87,7
22	2020C010782	Syauqi	82,22
23	2020C010783	Mutiara Al Farah	67,55
24	2020C010792	M. Yusuf Maulana	67.19
25	2021C010816	M Amirul Hisyam	81,15
26	2021C010817	Raihan Risqullah L	76,65
27	2020C010785	Azzam Al Fikri	78,31
28	2020C010786	M Alwi Al Farizy	84.61
29	2020C010787	M Sanubari	65,79
30	2020C010788	Meilani	71,53

2. Menentukan jumlah kluster
 Pada proses ini menentukan nilai k jumlah kluster siswa disabilitas sebanyak 3 kluster (k-3), Kluster yang dibentuk yaitu Kluster berat, Kluster sedang dan Kluster ringan.
3. Menentukan nilai *Centroid* awal
 Menentukan nilai *Centroid* awal yang telah ditentukan secara acak berdasarkan nilai variabel data yang di *Cluster* sebanyak yang ditentukan sebelumnya. Adapun *Cluster* berat (C3) diperoleh dari nilai terkecil , *Cluster* sedang (C2) diperoleh dari nilai rata-rata, dan *Cluster* rendah (C1) diperoleh dari nilai tertinggi. Berikut adalah nilai *centroid* data awal untuk iterasi 1 :

Tabel 2 *Centroid* data awal (iterasi 1)

C1	Maksimal	87,7
C2	Rata-rata	77.14
C3	Minimal	65,79

4. Menghitung jarak *Centroid*
 Mengitung jarak setiap data ke pusat *cluster*. Setelah data nilai pusat *cluster* ditentukan, langkah selanjutnya adalah menghitung jarak masing-masing data terhadap pusat *cluster* dengan menggunakan *Euclidean Distance* sebagai berikut:
 Hitung jarak ke pusat (centroid) di C1, C2, dan C3.
 $D(1.1) = \sqrt{(87,7 - 85,1)^2} = 2,6$
 $D(2.1) = \sqrt{(77,14 - 85,1)^2} = -7.96$
 $D(3.1) = \sqrt{(65,79 - 85,1)^2} = -19.31$
 Dan selanjutnya sampai dengan D (1.31), D (2,31), dan D (3.31) sehingga menghasilkan jarak terpendek dari *centroid*. Berikut data hasil dari perhitungan *centroid* 1 (C1), centroid 2 (C2) dan *centroid* 3 (C3) serta jarak terpendeknya :

Tabel 3 Jarak *centroid* Iterasi 1

NO	NIS	NAMA SISWA	C1	C2	C3	JARAK TERPENDEK
1	2019C020746	Tria Novitasari	2,6	7,96	19,31	2,6
2	2019C020747	Mohamad Fadhil Imani	16,15	5,59	5,76	5,59
3	2019C020748	Fitri Indriani	11,86	1,3	10,05	1,3
4	2019C020749	Khalifah Nur Artha	18,79	8,23	3,12	3,12
5	2019C020750	Mohamad Fikri Al Rasid	6,96	3,6	14,95	3,6
6	2019C020751	Hud Ibnu Neno	12,08	1,52	9,83	1,52
7	2019C020752	Dian Ayu Safitri	12,84	2,28	9,07	2,28
8	2020C010793	Rizki Akbar Ramadhani	6,28	4,28	15,63	4,28
9	2020C010794	Citra Nur Maulidina	17,09	6,53	4,82	4,82
10	2021C010818	Choerul Umam Muslih	2,13	8,43	19,78	2,13
11	2021C010808	Sultan Aufan Widodo	15,25	4,69	6,66	4,69
12	2021C010809	Fauzio Al Fariz	5,16	5,4	16,75	5,16
13	2021C010810	Syahla Nur Alisya Erliantoro	0,99	9,57	20,92	0,99
14	2021C010811	Fachri Maulana Firdaus	12,56	2	9,35	2
15	2021C010812	Safi Andhika Pratama	2,47	8,09	19,44	2,47
16	2021C010813	Wildan Kinnan Afrizal	18,19	7,63	3,72	3,72
17	2021C010814	M. Aenun Najib	12,87	2,31	9,04	2,31
18	2021C010815	Bayu Kurniawan	15,15	4,59	6,76	4,59
19	2020C010779	Artanta Putra S	9,25	1,31	12,66	1,31
20	2020C010780	M Faiz Afandi	7,55	3,01	14,36	3,01
21	2020C010781	Ikhwan Nadhif	0	10,56	21,91	0
22	2020C010782	Syauqi	5,48	5,08	16,43	5,08
23	2020C010783	Mutiara Al Farah	20,15	9,59	1,76	1,76
24	2020C010792	M. Yusuf Maulana	20,51	9,95	1,4	1,4
25	2021C010816	M Amirul Hisyam	6,55	4,01	15,36	4,01
26	2021C010817	Raihan Risqullah L	11,05	0,49	10,86	0,49
27	2020C010785	Azzam Al Fikri	9,39	1,17	12,52	1,17
28	2020C010786	M Alwi Al Farizy	3,09	7,47	18,82	3,09
29	2020C010787	M Sanubari	21,91	11,35	0	0
30	2020C010788	Meilani	16,17	5,61	5,74	5,61

5. Pengelompokan posisi Kluster

Setelah melakukan perhitungan pada iterasi pertama maka didapatkan pengelompokan dari anggota kluster 1, 2 dan 3 seperti di bawah ini:

- a. Kluster 1 terdiri dari 8 siswa dengan nomor urut: (1, 10,12,13,15,21,22,28)
- b. Kluster 2 terdiri dari 14 siswa dengan nomor urut: (5,6,7,8,11,14,17,18,19,20,25,26, 27,30)
- c. Kluster 3 terdiri dari 8 siswa dengan nomor urut: (2,3,4,9,16,23,24,29)

Setelah mendapatkan hasil perhitungan iterasi pertama, maka langkah Selanjutnya menghitung iterasi kedua dengan melakukan kembali langkah 3-5. Jika nilai *centroid* hasil iterasi sama dengan nilai *centroid* sebelumnya, atau jika nilai sudah optimal dan posisi *cluster* tidak berubah, maka proses iteratif berhenti. Namun, jika nilai *centroid* tidak sama,

atau jika nilainya belum optimal, proses iterasi dilanjutkan dengan iterasi berikutnya. Berikut ini menghitung pusat baru menggunakan hasil setiap anggota dari setiap *cluster* :

$$C1_{1x} = \frac{85,1+85,57+82,54+86,71+85,23+87,7+82,22+84,61}{8}$$

$$= 84,96$$

$$C2_{2x} = \frac{80,74+75,62+74,86+8,42+72,45+75,14+74,83+72,55+78,45+80,15+81,15+76,65+78,31+71,53}{14}$$

$$= 76,703$$

$$C3_{3x} = \frac{71,55+75,84+68,91+70,61+69,51+67,55+67,19+65,79}{8}$$

$$= 69,61$$

Setelah melakukan perhitungan pada iterasi kedua didapatkan pengelompokan dari anggota Klaster 1, 2 dan 3 seperti di bawah ini:

- a. Klaster 1 terdiri dari 10 siswa dengan nomor urut: (1, 8,10,12,13,15,21,22,25,28)
- b. Klaster 2 terdiri dari 10 siswa dengan nomor urut: (3,5,6,7,14,17,19,20,26,27)
- c. Klaster 3 terdiri dari 10 siswa dengan nomor urut: (2,4,9,11,16,18,23,24,29,30)

Setelah mendapatkan hasil perhitungan iterasi 2, langkah selanjutnya Ulangi langkah 3-5 untuk menghitung iterasi ketiga. Jika nilai centroid hasil iteratif sama dengan nilai centroid sebelumnya, atau jika nilai sudah optimal dan posisi cluster tidak berubah, maka proses iterasi berhenti. Namun, Namun jika nilai *centroid* tidak sama atau nilainya belum optimal maka proses iterasi berlanjut pada iterasi selanjutnya. Berikut hasil dari Klaster iterasi 3 :

- a. Klaster 1 terdiri dari 11 siswa dengan nomor urut: (1,5, 8,10,12,13,15,21,22,25,28)
- b. Klaster 2 terdiri dari 9 siswa dengan nomor urut: (3,6,7,14,17,19,20,26,27)
- c. Klaster 3 terdiri dari 10 siswa dengan nomor urut: (2,4,9,11,16,18,23,24,29,30)

Dilihat dari posisi Klaster 1, Klaster 2 dan Klaster 3 memiliki nilai belum sama dan ada perubahan maka lanjut ke iterasi ke empat, berikut hasil Klaster iterasi ke 4:

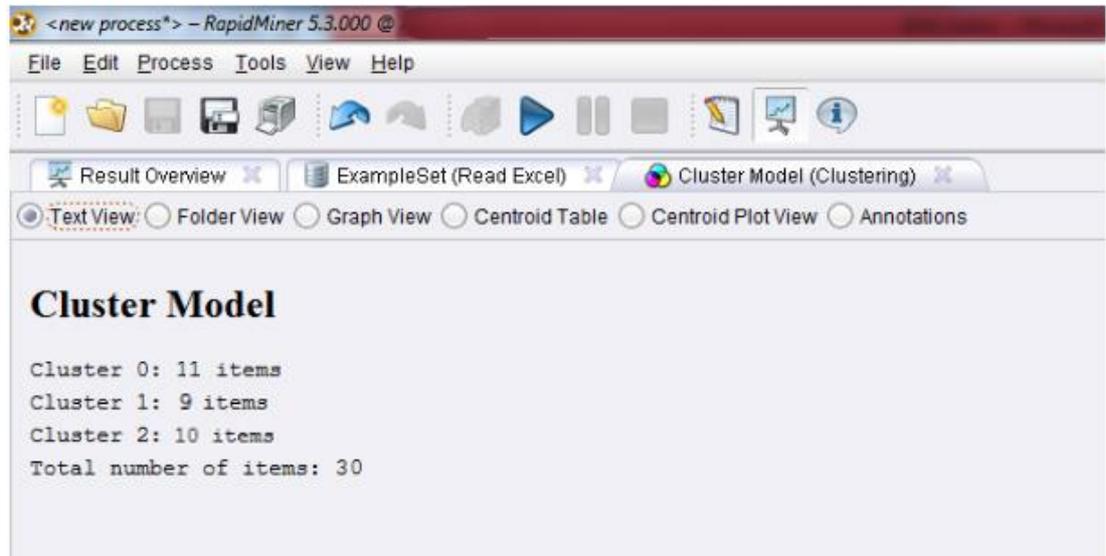
- a. Klaster 1 terdiri dari 11 siswa dengan nomor urut: (1,5, 8,10,12,13,15,21,22,25,28)
- b. Klaster 2 terdiri dari 9 siswa dengan nomor urut: (3,6,7,14,17,19,20,26,27)
- c. Klaster 3 terdiri dari 10 siswa dengan nomor urut: (2,4,9,11,16,18,23,24,29,30)

Dilihat dari daftar posisi Klaster 1, Klaster 2 dan Klaster 3 memiliki nilai yang sama dan tidak ada perubahan. Pada penelitian ini, proses perhitungan K-Means berhenti pada iterasi ke 4, karena iterasi ke 4 sama hasilnya dengan iterasi sebelumnya. Berdasarkan hasil posisi klaster, diperoleh Klaster 1 berjumlah 11 data siswa, Klaster 2 berjumlah 9 data siswa dan Klaster 3 berjumlah 10 data siswa, sehingga diperoleh kesimpulan sebagai berikut :

1. Klaster Ringan dengan jumlah siswa kelas yang masuk disabilitas ringan sebanyak 11, yaitu atas nama:
 - Tria Novitasari, Mohamad Fikri Al Rasid, Rizki Akbar Ramadhani, Chaerul Umam Muslih, Fauzio al Fariz, Syahla Nur Alisya Erliantoro, Safi Andhika Pratama, Ikhwan Nadhif, Syauqi, M Amirul Hisyam, M Alwi Al Farizy.
2. Klaster Sedang dengan jumlah siswa kelas yang masuk disabilitas sedang sebanyak 9, yaitu atas nama:
 - Fitri Indriani, Hud Ibnu Neno, Dian Ayu Safitri, Fachri Maulana Firdaus, M. Aenun Najib, Artanta Putra S, M Faiz Afandi, Raihan Risqullah L, Azzam Al Fikri.
3. Klaster Berat dengan jumlah siswa kelas yang masuk disabilitas berat sebanyak 10, yaitu atas nama:
 - Mohamad Fadhil Imani, Khalifah Nur Artha, Citra Nur Maulidina, Sultan Aufan Widodo, Wildan Kinnan Afrizal, Bayu Kurniawan, Mutiara Al Farah, M. Yusuf Maulana, M. Sanubari, Meilani.

Berdasarkan penjelasan diatas yang telah dianalisa bahwa data tersebut valid. Dibuktikan dengan hasil akhir perhitungan manual dan hasil akhir dari aplikasi *RapidMiner* dapat menampilkan hasil yang sama. Berikut adalah tampilan cluster model berupa teks yang ada didalam aplikasi *RapidMiner* dari *cluster* yang terbentuk :

Gambar 2 *Cluster Model* Siswa Menggunakan aplikasi *RapidMiner*



Keterangan :

1. Jumlah *Cluster* 0 (Ringan) berjumlah 11 Items
 2. Jumlah *Cluster* 1 (Sedang) berjumlah 9 Items
 3. Jumlah *Cluster* 2 (Berat) berjumlah 10 Items
- Dengan jumlah total keseluruhan ada 30 Items/Siswa

Kesimpulan dan Saran

Berdasarkan penelitian yang telah dilakukan tentang pengelompokan kelas disabilitas dengan Algoritma K-Means dapat disimpulkan bahwa Klaster Ringan dengan jumlah data 11 siswa, Klaster Sedang 9 siswa, dan Klaster Berat jumlah data sebanyak 10 siswa dari total 30 data sample kelas disabilitas di SLB N Kota Tegal. Berdasarkan pegujian data yang dilakukan dalam penelitian ini, Klaster berhenti pada data yang terjadi di iterasi 4 dengan jumlah 11, 9, dan 10 siswa. Nilai hasil akurasi dari perhitungan manual *Excel* dan *RapidMiner* bernilai sama.

Referensi

- [1] I. S. S. Ade Indah Sari, Heru Satria Tambunan, Widodo Saputra, Irfan Sudahri Damanik, "Implementasi Algoritma K-Means Dalam Pengelompokan Penyandang Disabilitas Menurut Kecamatan Kabupaten Simalungun," *Pros. Semin. Nas. Ris. Dan Inf. Sci.*, vol. 2, no. 0, pp. 54–61, 2020, [Online]. Available: <http://tunasbangsa.ac.id/seminar/index.php/senaris/article/view/144>
- [2] R. Rosmiati, "PEMEROLEHAN BAHASA INDONESIA PADA ANAK TUNAGRAHITA PADA TAHAP PERKEMBANGAN KOGNITIF," *J. Penelitian, Pendidikan, dan Pembelajaran*, vol. 13, no. 4, 2019.
- [3] M. Sadikin, R. Rosnelly, R. Roslina, T. S. Gunawan, and W. Wanayumini, "Penerapan Data Mining Pada Penerimaan Dosen Tetap Menggunakan Metode Naive Bayes Classifier dan C4.5," *J. MEDIA Inform. BUDIDARMA*, vol. 4, no. 4, pp. 1100–1109, 2020.
- [4] Y. D. Darmi and A. Setiawan, "Penerapan metode clustering k-means dalam pengelompokan penjualan produk," *J. Media Infotama*, vol. 12, no. 2, 2016.
- [5] A. Ali, "Clustering Data Antropometri Balita Untuk Menentukan Status Gizi Balita Di

- Kelurahan Jumput Rejo Sukodono Sidoarjo,” *JATISI (Jurnal Tek. Inform. Dan Sist. Informasi)*, vol. 7, no. 3, pp. 395–407, 2020.
- [6] A. Rohmatullah, D. Rahmalia, and M. S. Pradana, “Klasterisasi Data Pertanian Di Kabupaten,” *J. Ilm. Teknosains*, vol. V, no. 2, pp. 686–93, 2019.
- [7] F. Harahap, “Perbandingan Algoritma K Means dan K Medoids Untuk Clustering Kelas Siswa Tunagrahita,” *TIN Terap. Inform. Nusant.*, vol. 2, no. 4, pp. 191–197, 2021.
- [8] A. Solichin and K. Khairunnisa, “Klasterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means,” *Fountain Informatics J.*, vol. 5, no. 2, p. 52, 2020, doi: 10.21111/fij.v5i2.4905.
- [9] A. Irma Purnamasari and R. Danar Dana, “Analisis Penikmat Kopi Berdasarkan Cita Rasa Kopi Menggunakan Teknik Clustering Algoritma K-Means di Cirebon The Analyzing of Coffee Lover Based on Coffee Flavor Using K-Means Algorithms Clustering Technique in Cirebon,” 2017.
- [10] W. Sirait, S. Defit, and G. W. Nurcahyo, “Algoritma K-Means Untuk Klasterisasi Tugas Akhir Mahasiswa Berdasarkan Keahlian,” *J. Sistim Inf. dan Teknol.*, vol. 1, no. 3, pp. 25–30, 2019, doi: 10.35134/jsisfotek.v1i3.6.
- [11] R. K. Arora and D. Badal, “Evaluating student’s performance using k-means clustering,” *Int. J. Comput. Sci. Technol.*, vol. 4, no. 2, 2013.
- [12] E. Muningsih and S. Kiswati, “Penerapan Metode K-Means Untuk Clustering Produk Online Shop Dalam Penentuan Stok Barang,” *J. Bianglala Inform.*, vol. 3, no. 1, pp. 10–17, 2015.
- [13] P. Sari, B. Pramono, and L. ode H. S. Sagala, “Improve K-Means Terhadap Status Nilai Gizi Pada Balita,” *semanTIK*, vol. 3, no. 1, pp. 143–148, 2017, doi: 10.1063/1.2957900.
- [14] F. Indriyani and E. Irfiani, “Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means,” *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.