

	<h1 style="color: red;">Jurnal Informatika dan Komputer</h1> <h2 style="color: red;">(JIK)</h2>	
	<p>Vol. 15 No. 2 (2024)</p>	<p>ISSN Media Cetak: 2089 - 4384</p>

PEMANFAATAN ALGORITMA K-MEANS CLUSTERING DALAM PENENTUAN PRIORITAS PENERIMA PROGRAM BANTUAN SOSIAL PENDIDIKAN

M. Andrik Muqorrobin P.¹, Mariatul Qibtiyah²

¹ Teknologi Informasi, Universitas Ibrahimy, Situbondo

² Pendidikan Agama Islam, Universitas Nurul Jadid, Probolinggo

Jl. KHR. Syamsul Arifin No.1-2, Situbondo 68374, Indonesia

Telp: 0123456789; fax: 0735-326169;

e-mail:;, wمامmuqorrobinp@gmail.com¹, mariatulqibtyh@gmail.com²

Abstract - The Education Assistance Program is one of the government's efforts to increase access to education for students from underprivileged families. An efficient and targeted selection mechanism is needed so that aid can be distributed effectively. K-Means Clustering algorithm is used to group students based on certain indicators, such as the number of students per school at a certain level of education (elementary, junior high, high school and college) in a particular region. This method aims to support decision-making in determining the receipt of educational assistance by identifying groups of students who need the most support. The K-Means Clustering algorithm works by dividing student data into several clusters based on similar characteristics.

This process allows the identification of support priorities for each group of students, such as students with high or low levels of need. This approach is expected to improve efficiency, objectivity and equity in the distribution of educational assistance. The results of the analysis with the K-Means algorithm can also reveal hidden patterns in education data, thus supporting policy makers to improve program quality and ensure that aid reaches the neediest students.

The results of this study show that the K-Means Clustering algorithm is able to identify groups of beneficiaries efficiently and objectively.

The resulting clusters provide a clear picture of the level of need of each group based on the analyzed data. By using tools such as Rapid Miner. The clustering process can be done quickly the results are visualized using Matplotlib and Seaborn, which makes data interpretation easier.

The main advantage of this approach is its ability to handle large and complex datasets, while providing results that are easy to interpret. As such, this research makes a significant contribution in providing a more transparent and data-driven decision-making basis for agencies responsible for education social assistance programs.

This research not only offers a technical framework for data clustering, but also provides practical insights that can be applied in various social assistance programs in the future.

Keywords: K-Means Clustering Algorithm, Educational Aid, Recipient Selection, Data Clustering, Distribution Efficiency

Abstrak – Program Bantuan Pendidikan merupakan salah satu upaya pemerintah untuk meningkatkan akses pendidikan bagi siswa yang berasal dari keluarga kurang mampu. Dibutuhkannya mekanisme seleksi yang efisien dan terdistribusikan agar bantuan dapat didistribusikan dengan efektif. Algoritma K-

Means Clustering digunakan untuk mengelompokkan siswa berdasarkan indikator tertentu, seperti jumlah siswa persekolah pada tingkat pendidikan tertentu (SD, SMP, SMA dan Perguruan Tinggi) di wilayah tertentu. Metode ini bertujuan untuk mendukung pengambilan keputusan dalam menentukan penerimaan bantuan pendidikan dengan mengidentifikasi kelompok siswa yang paling membutuhkan dukungan. Algoritma K-Means Clustering bekerja dengan membagi data siswa menjadi beberapa cluster berdasarkan kesamaan karakteristik yang dimiliki. Proses ini memungkinkan identifikasi prioritas dukungan bagi setiap kelompok siswa, seperti siswa dengan tingkat kebutuhan tinggi maupun rendah.

Pendekatan ini diharapkan dapat meningkatkan efisiensi, obyektivitas, dan pemerataan dalam distribusi bantuan pendidikan. Hasil analisis dengan algoritma K-Means juga dapat mengungkap pola tersembunyi dalam data pendidikan, sehingga mendukung pengambil kebijakan untuk memperbaiki kualitas program dan memastikan bahwa bantuan menjangkau siswa yang paling membutuhkan.

Hasil dari penelitian ini menunjukkan bahwa algoritma K-Means Clustering mampu mengidentifikasi kelompok-kelompok penerima bantuan secara efisien dan objektif. Kluster yang dihasilkan memberikan gambaran yang jelas tentang tingkat kebutuhan setiap kelompok berdasarkan data yang dianalisis. Dengan menggunakan alat bantu seperti Rapid Miner. Proses clustering dapat dilakukan dengan cepat hasilnya divisualisasikan menggunakan Matplotlib dan Seaborn, yang mempermudah interpretasi data.

Keunggulan utama dari pendekatan ini adalah kemampuannya untuk menangani dataset yang besar dan kompleks, sekaligus memberikan hasil yang mudah diinterpretasikan. Dengan demikian, penelitian ini memberikan kontribusi signifikan dalam menyediakan dasar pengambilan keputusan yang lebih transparan dan berbasis data bagi instansi yang bertanggung jawab atas program bantuan sosial pendidikan.

Penelitian ini tidak hanya menawarkan kerangka kerja teknis untuk pengelompokan data, tetapi juga memberikan wawasan praktis yang dapat diterapkan dalam berbagai program bantuan sosial di masa depan.

Kata Kunci: *Algoritma K-Means Clustering, Bantuan Pendidikan, Seleksi Penerima, Pengelompokan Data, Afisiensi Distribusi.*

I. PENDAHULUAN

1.1 Latar Belakang Masalah

Algoritma K-Means Clustering merupakan salah satu pendekatan populer dalam bidang analisis data dan data mining. Algoritma ini digunakan untuk mengelompokkan data berdasarkan kemiripan tertentu tanpa acuan yang jelas, sehingga seluruh data dibagi menjadi kelompok-kelompok yang memiliki karakteristik serupa. Metode ini cocok digunakan untuk menyederhanakan analisis data yang kompleks menjadi struktur yang lebih terorganisir dan informatif, terutama dalam konteks data yang besar dan beragam (Rohmah et al., 2021).

Pendidikan memiliki peran fundamental dalam membentuk karakter serta peradaban bangsa, terlebih di Indonesia yang memiliki masyarakat majemuk. Pendidikan bertujuan mencerdaskan kehidupan bangsa dan meningkatkan martabat manusia. Namun, masih terdapat banyak tantangan, terutama dalam hal kesenjangan akses pendidikan akibat keterbatasan finansial. Untuk menjawab tantangan ini, pemerintah meluncurkan Program Indonesia Pintar (PIP), sebuah inisiatif untuk mencegah siswa putus sekolah dan memastikan akses pendidikan yang merata (Dayera, Musa Bundaris Palungan, 2024).

PIP memberikan bantuan finansial kepada siswa yang membutuhkan, namun proses seleksi penerima bantuan ini menghadapi berbagai kendala, seperti data siswa yang besar dan beragam serta kondisi ekonomi yang berbeda-beda antar daerah. Oleh karena itu, diperlukan metode berbasis data yang efektif untuk mendukung proses pengambilan keputusan dalam program ini. Dalam hal ini, Algoritma K-Means Clustering dapat menjadi solusi yang relevan. Dengan mengelompokkan siswa berdasarkan kesamaan karakteristik, algoritma ini dapat membantu pemerintah menentukan penerima bantuan secara lebih cepat, objektif, dan akurat (Rakista, 2021).

Penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering dalam mengelompokkan data siswa dari berbagai tingkat pendidikan (SD, SMP, SMA, dan Perguruan Tinggi). Dengan pendekatan ini, diharapkan distribusi bantuan PIP menjadi lebih efektif dan tepat sasaran. Selain itu, hasil pengelompokan

data diharapkan dapat memberikan wawasan yang lebih baik bagi pengambil kebijakan, sehingga manfaat dari PIP dapat dirasakan secara merata di seluruh Indonesia (Dewi et al., 2022).

1.2 Batasan Masalah Masalah

Ada beberapa batasan masalah terkait penelitian ini, diantaranya:

- Penelitian ini hanya berfokus pada penerapan algoritma K-Means Clustering untuk mengelompokkan calon penerima bantuan sosial pendidikan berdasarkan siswa yang meliputi jenjang pendidikan (SD, SMP, SMA dan Perguruan Tinggi), serta data jumlah siswa per kabupaten/kota.
- Data yang digunakan merupakan dataset sekunder yang berasal dari Badan Pusat Statistik (BPS) dan instansi terkait, dengan fokus pada tahun tertentu.
- Penelitian berikut ini tidak mencakup analisis variabel tambahan seperti tingkat kemiskinan atau aksesibilitas pendidikan, yang dapat memengaruhi hasil clustering.
- Pengolahan data dilakukan menggunakan perangkat lunak Rapid Miner Studio, dengan jumlah kluster (k) ditentukan sebanyak tiga.

1.3 Rumusan Masalah

Ada beberapa rumusan masalah terkait penelitian ini, diantaranya:

- Bagaimana penerapan algoritma K-Means Clustering untuk mengelompokkan siswa berdasarkan karakteristik data yang relevan?
- Bagaimana menentukan jumlah kluster yang optimal untuk memprioritaskan penerima program bantuan pendidikan?
- Apakah hasil pengelompokan dengan algoritma ini dapat meningkatkan efisiensi dan akurasi penyaluran bantuan pendidikan?

1.4 Tujuan Penelitian

Ada beberapa tujuan penelitian yang terkait, diantaranya:

- Mengembangkan metode berbasis algoritma K-Means Clustering untuk mendukung pengambilan keputusan dalam seleksi penerima bantuan pendidikan.
- Mengelompokkan data siswa berdasarkan kesamaan karakteristik untuk mengidentifikasi prioritas penerima bantuan pendidikan.
- Meningkatkan efisiensi, akurasi, dan pemerataan distribusi bantuan pendidikan dengan memanfaatkan hasil clustering.

1.5 Manfaat Penelitian

Ada beberapa manfaat terkait penelitian ini, diantaranya:

- Bagi Pemerintah: Memberikan solusi berbasis data untuk mendukung kebijakan distribusi bantuan pendidikan yang lebih efektif dan tepat sasaran.
- Bagi Peneliti Lain: Menyediakan kerangka teknis yang dapat diterapkan pada berbagai program bantuan sosial lainnya.
- Bagi Pengambil Kebijakan: Menyediakan wawasan tentang pola kebutuhan pendidikan yang tersembunyi dalam data, guna meningkatkan kualitas program bantuan di masa depan.

1.6 Pengembangan Pembahasan Jurnal

1. Studi Literatur yang Relevan

Penggunaan algoritma K-Means Clustering telah diterapkan dalam berbagai bidang, seperti pendidikan, bantuan sosial, dan analisis kesehatan. Studi oleh Dewi et al. (2022) menunjukkan bahwa algoritma ini dapat membantu menentukan prioritas bantuan pendidikan dengan efisiensi yang lebih tinggi dibandingkan metode konvensional. Di sisi lain, penelitian oleh Salam et al. (2020) membandingkan K-Means dengan metode DBSCAN, menunjukkan bahwa K-Means lebih unggul dalam dataset berukuran besar, meskipun memiliki keterbatasan pada data dengan bentuk cluster yang tidak bulat.

2. Detail Proses Penelitian

Dalam penelitian ini, data siswa diolah menggunakan RapidMiner Studio versi 10.3. Langkah pertama adalah melakukan pra-pemrosesan data, seperti normalisasi dan penghapusan data yang tidak relevan. Dataset terdiri dari 35 entri dengan 6 variabel utama, yang kemudian dikonversi menjadi format numerik untuk memudahkan analisis. Setelah itu, algoritma K-Means diterapkan dengan parameter jumlah cluster (k) = 3, dan hasilnya divisualisasikan menggunakan diagram batang dan diagram lingkaran. Visualisasi ini memberikan gambaran yang jelas tentang pembagian data ke dalam cluster.

3. Implikasi Hasil Penelitian

Hasil penelitian ini menunjukkan bahwa cluster dengan prioritas tinggi dapat membantu pemerintah dalam menentukan penerima bantuan pendidikan secara lebih efektif. Dengan implementasi ini, diharapkan distribusi bantuan dapat meningkatkan efisiensi, mengurangi biaya operasional, dan menghindari penyaluran yang

tidak tepat sasaran. Selain itu, analisis ini dapat memberikan wawasan tambahan bagi pembuat kebijakan untuk mengidentifikasi daerah dengan kebutuhan pendidikan yang paling mendesak.

4. Diskusi Kritis

Salah satu tantangan yang dihadapi dalam penelitian ini adalah keterbatasan atribut data, seperti tidak adanya variabel terkait tingkat kemiskinan. Hal ini dapat menyebabkan hasil clustering kurang mencerminkan realitas sosial secara merata. Oleh karena itu, penelitian lanjutan disarankan untuk menambahkan variabel seperti aksesibilitas pendidikan dan tingkat kemiskinan keluarga.

5. Studi Kasus Tambahan

Sebagai contoh aplikasi praktis, hasil penelitian ini dapat diterapkan pada tingkat regional untuk menentukan prioritas bantuan bagi siswa di daerah pedesaan. Simulasi tambahan menunjukkan bahwa dengan menambahkan atribut seperti jumlah guru per siswa dan jarak ke sekolah terdekat, hasil clustering menjadi lebih akurat. Studi ini juga dapat diperluas untuk program bantuan lain, seperti bantuan kesehatan atau sosial ekonomi.

6. Usulan Pengembangan

Untuk pengembangan lebih lanjut, integrasi teknologi seperti AI dan machine learning dapat digunakan untuk memperbaiki akurasi hasil clustering ini. Selain itu, kolaborasi dengan instansi pendidikan dan lembaga sosial akan memperkuat implementasi hasil penelitian. Penggunaan metode clustering alternatif seperti Hierarchical Clustering atau DBSCAN juga dapat dijadikan pembandingan untuk mengukur efektivitas algoritma K-Means.

7. Kesimpulan yang Mendalam

Penelitian ini menunjukkan bahwa algoritma K-Means Clustering adalah alat yang efektif untuk mendukung pengambilan keputusan dalam program bantuan pendidikan. Dengan pendekatan berbasis data, hasil clustering dapat meningkatkan efisiensi, akurasi, dan transparansi dalam distribusi bantuan. Kemudian, studi lanjutan dengan variabel tambahan dan teknologi canggih dapat semakin meningkatkan manfaat dari pendekatan ini.

II. METODE PENELITIAN

2.1. Metode Pengumpulan Data

Dalam pengumpulan data yang diperlukan untuk penelitian ini, ada dua metode, diantaranya:

1. Pendekatan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksploratif. Algoritma K-Means Clustering diterapkan untuk mengelompokkan calon penerima bantuan sosial pendidikan berdasarkan data yang relevan. Pendekatan ini bertujuan untuk menemukan pola dalam data yang mendukung pengambilan keputusan.

2. Referensi

Referensi dilakukan dengan pengumpulan referensi-referensi yang berhubungan dengan permasalahan yang ada, beberapa buku-buku, majalah dan artikel.

2.2 Data Mining

Data Mining adalah Serangkaian proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu basisdata dengan melakukan penggalian polapola dari data dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berharga yang diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data(Jundanuddin et al., 2024).

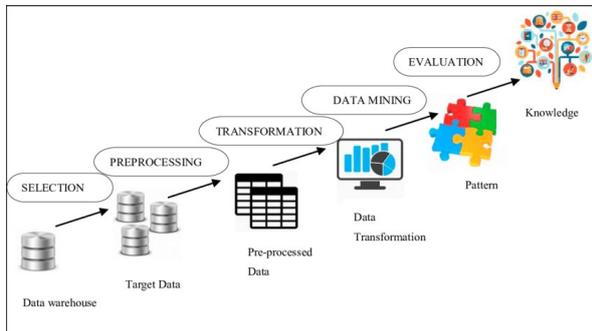
Data mining memiliki banyak metode, metode data mining yang digunakan adalah Knowledge Discovery in Database (KDD) ataupun dapat disebut Patten Recognition.

2.3 Knowledge Discovery in Database

Knowledge discovery in database (KDD) adalah metode teknis yang berguna untuk mencari dan mengidentifikasi pola (patten) dalam data, pola yang sudah ditemukan bersifat sah dan baru sehingga dapat bermanfaat dan dapat dimengerti. Proses dalam KDD terdapat 5 tahapan yaitu seleksi data dari data sumber ke data target, tahap pre-processing, transformasi, data mining dan tahap evaluasi(Fatmawati & Windarto, 2018).

Tahap seleksi dilakukan untuk menargetkan data yang digunakan untuk penelitian, tahap pre-procesing dapat dilakukan integrasi data atau penggabungan data serta dilakukan cleaning data, yaitu dengan menghilangkan noise, data redundan, inkonsistensi data, serta data yang tidak relevan, tahap transformasi adalah penggabungan data dan penyesuaian format data agar dapat diproses pada tahap data mining, tahap data mining dilakukan menggunakan algoritma yang cocok untuk

permasalahan dalam data, serta tahap evaluasi yang digunakan untuk pengujian dalam data (Hadiza et al., 2023).



Gambar 1. Tahapan KDD

Pada gambar 1 digambarkan alur mengenai tahapan KDD, setelah disesuaikan dengan tahapan KDD dihasilkan kerangka kerja penelitian yang dijabarkan pada Gambar 2.



Gambar 2. Kerangka Kerja

2.4 Clustering

Clustering adalah salah satu algoritma unsupervised learning yang paling populer dalam analisis pengelompokan data. Algoritma ini bekerja dengan membagi sejumlah data ke dalam k kelompok (cluster) berdasarkan kemiripan karakteristik data. Proses clustering dilakukan dengan menentukan titik pusat (centroid) untuk setiap cluster, kemudian menghitung jarak setiap data ke masing-masing centroid (Informasi et al., 2024).

Clustering merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan cluster. Objek yang di dalam cluster memiliki kemiripan karakteristik antar satu sama lainnya dan berbeda dengan cluster yang lain. Partisi tidak dilakukan secara manual melainkan dengan algoritma clustering. Oleh karena itu, clustering sangat berguna dan bisa menemukan group atau kelompok yang tidak dikenal dalam data (Irwansyah & Faisal, 2015).

2.5 Algoritma K-Means

Algoritma K-Means merupakan metode yang banyak digunakan untuk mengidentifikasi suatu kelompok dari kasus. K-Means artinya prosedur pemecah pengelompokan secara iteratif yang

melakukan partisi buat mengklasifikasikan atau mengelompokkan sejumlah besar objek. Prosedur pemecahan K-Means adalah metode data mining yang banyak digunakan buat mengidentifikasi suatu kelompok yang alami berasal sebuah kasus yang berdasarkan pada pengelompokan data yang memiliki kemiripan sehingga hasilnya dari pengelompokan dapat dianalisis (Triandini et al., 2021).

Salah satu kelebihan penerapan Algoritma K-Means adalah kemampuannya untuk mengelompokkan data dalam jumlah besar secara efisien, sekaligus meningkatkan kecepatan proses pengelompokan.

III. HASIL DAN PEMBAHASAN

Implementasi penelitian ini menggunakan perhitungan Algoritma K-Means Clustering yang dilakukan dengan tools Rapid Miner Studio. Rapid Miner sebagai tools yang mampu memudahkan user-nya dalam melakukan operator. Operator berfungsi untuk memodifikasi data dan selanjutnya dihubungkan dengan node – node operator (Salam et al., 2020).

Hasil penelitian menunjukkan bahwa dari 35 dataset jumlah siswa di setiap kabupaten atau kota pada tingkat pendidikan (SD, SMP, SMA, dan Perguruan Tinggi) tahun 2021, terdapat 16 data yang tergolong dalam cluster 0 dan 15 data dalam cluster 2. Kedua cluster ini dapat dipertimbangkan untuk menerima bantuan tambahan atau diprioritaskan lebih rendah. Sementara itu, 4 data masuk ke dalam cluster 1, yang menunjukkan daerah dengan kebutuhan mendesak, sehingga siswa di wilayah tersebut dapat difokuskan sebagai penerima utama. Dengan pengelompokan data ke dalam tiga cluster ini, pemerintah dapat lebih mudah menetapkan prioritas penyaluran bantuan PIP secara lebih efektif dan tepat sasaran.

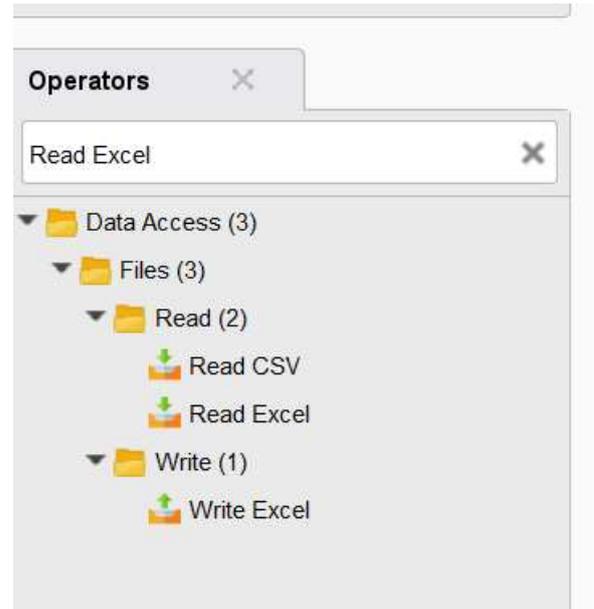
3.1 Dataset

Dataset yang digunakan pada penelitian ini merupakan data jumlah siswa pada setiap kabupaten atau kota pada tingkat (SD, SMP, SMA, Perguruan Tinggi) yang nantinya akan direkomendasikan untuk mendapatkan beasiswa PIP lalu ditransformasikan dari data non-numeric. Selanjutnya dilakukan sebuah proses inisiasi kedalam bentuk numeric agar data diolah dengan mudah (Rady Putra & Angrawan, 2021).

Data penerima beasiswa yang digunakan dalam perhitungan terdiri dari 35 data dan 6 variabel yang diperoleh dari dataset Badan Pusat Statistik (BPS) Indonesia. Data tersebut disajikan pada Gambar 3 berikut ini:

Row No.	Kabupaten	No	SD	SMP	SMA	Perguruan T.	Tahun
1	Kabupaten Cilacap	1	85.980	33.950	16.930	0	2021
2	Kabupaten Banjarnegara	2	72.200	20.990	13.950	0	2021
3	Kabupaten Purwokerto	3	64.670	24.540	16.670	0.800	2021
4	Kabupaten Banjarnegara	4	69.840	26.000	6.770	0	2021
5	Kabupaten Kebumeh	5	66.700	26.930	19.670	2.230	2021
6	Kabupaten Puncung	6	55.130	35.090	24.530	0.960	2021
7	Kabupaten Wonorejo	7	54.230	35.000	16.080	1.130	2021
8	Kabupaten Magelang	8	70.190	22.340	13.430	0	2021
9	Kabupaten Boyali	9	60.300	29.170	26.750	0	2021
10	Kabupaten Klaten	10	63.470	30.120	18.990	2.450	2021
11	Kabupaten Sukoharjo	11	61.070	30.990	36.910	2.250	2021
12	Kabupaten Wonorejo	12	60.380	21.730	16.760	1.290	2021
13	Kabupaten Karanganyar	13	60.250	22.280	16.880	3.390	2021
14	Kabupaten Sragen	14	59.170	36.070	24.430	0	2021
15	Kabupaten Oudya	15	58.030	28.980	24.050	2.190	2021
16	Kabupaten Blora	16	58.020	19.330	20.380	2.580	2021
17	Kabupaten Rembang	17	61.770	27.930	18.020	0	2021
18	Kabupaten Pati	18	64.660	22.010	13.660	1.880	2021
19	Kabupaten Kudus	19	75.190	18.370	13.870	1.350	2021

Gambar 3. Testing data – load current (amperes)



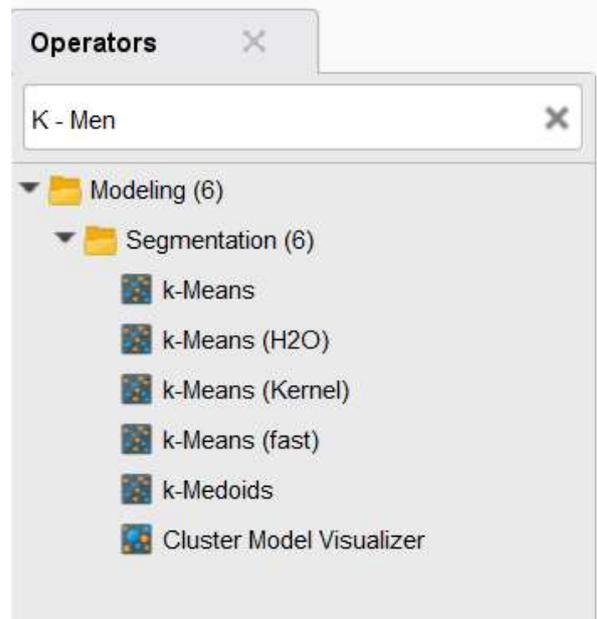
Gambar 4. Read Excel

3.2 Pengolahan Data dengan software Rapidminer

Untuk mengelompokkan data dalam penelitian ini, proses dilakukan menggunakan software RapidMiner Studio versi 10.3 dengan langkah-langkah sebagai berikut:

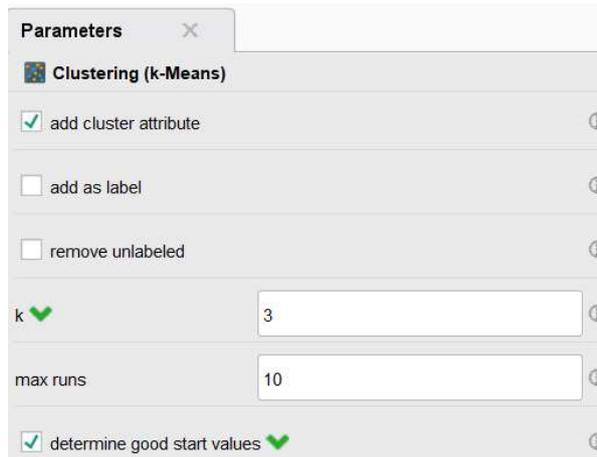
- Buka software Rapid Miner Studio versi 10.3.
- Pilih menu File, kemudian klik New Process.
- J Masuk ke dalam panel operator, ketik "Read Excel", lalu seret operator Read Excel ke dalam panel proses untuk memuat data dari file Excel yang telah melalui tahap data cleansing dan data transformation. Proses ini dapat dilihat pada Gambar 4 berikut:

- Langkah selanjutnya adalah memasukkan model yang akan digunakan untuk proses clustering data mining dengan cara masuk ke menu operator kemudian pilih K-Means seperti pada gambar 5 berikut:



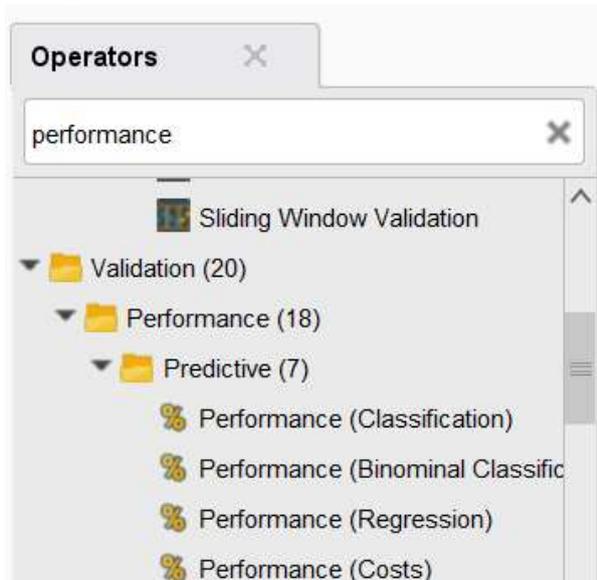
Gambar 5. Memilih Pemodelan Clustering

- Selanjutnya, pada proses clustering tentukan terlebih dahulu jumlah klasternya menjadi K=3 dan mixed measure menjadi mixed Euclidean distance seperti pada gambar 6 berikut:



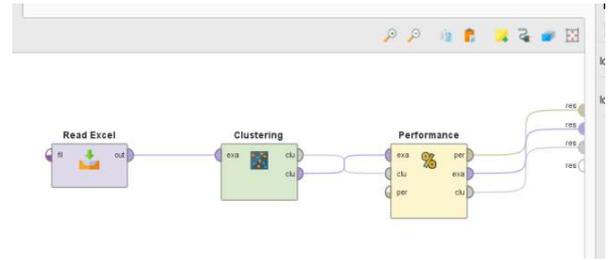
Gambar 6. Menentukan Jumlah K

f. Langkah selanjutnya adalah memasukkan model yang akan mengukur seberapa baik hasil clustering yang dilakukan dengan cara masuk ke menu operator kemudian pilih Performance (classification) seperti pada gambar 7 berikut:



Gambar 7. Memilih Pemodelan Performance (Classification)

g. Selanjutnya adalah menghubungkan konektor masing – masing proses pada main process seperti pada gambar 8 dan selanjutnya klik button run untuk memulai pengklusteran.



h.

Gambar 8. Main Proses Data Mining

3.3 Hasil Klusterisasi dengan Algoritma K-Means

a. Data View

Setelah melewati tahapan proses K-Means Clustering menggunakan aplikasi Rapidminer Studio 10.3 maka proses berikutnya adalah menampilkan hasil dari penerapan tersebut. Pada gambar 9 akan ditampilkan data view hasil clusterisasi yang telah didapatkan sebagai berikut:

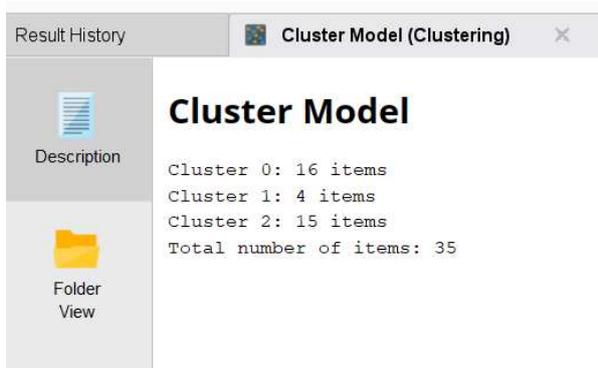
Row No.	Kategori	Kelas	No	SD	RPP	RMA	Program T.	Tahun
1	1	Kabupaten Ciklapa	Cluster_2	1	40.000	13.900	16.900	2017
2	2	Kabupaten Garut	Cluster_2	2	72.200	20.800	13.900	2017
3	3	Kabupaten Pangajene	Cluster_2	3	44.670	14.100	16.900	2017
4	4	Kabupaten Garut	Cluster_2	4	68.800	20.900	17.700	2017
5	5	Kabupaten Sukabumi	Cluster_2	5	60.700	20.400	16.900	2017
6	6	Kabupaten Pangajene	Cluster_2	6	59.100	19.900	24.900	2017
7	7	Kabupaten Sukabumi	Cluster_2	7	54.220	15.500	16.900	2017
8	8	Kabupaten Pangajene	Cluster_2	8	79.100	22.200	13.400	2017
9	9	Kabupaten Sukabumi	Cluster_2	9	40.000	20.100	24.900	2017
10	10	Kabupaten Sukabumi	Cluster_2	10	51.470	10.100	16.900	2017
11	11	Kabupaten Sukabumi	Cluster_2	11	41.070	10.900	16.900	2017
12	12	Kabupaten Pangajene	Cluster_2	12	40.200	21.700	16.900	2017
13	13	Kabupaten Pangajene	Cluster_2	13	40.200	12.200	14.900	2017
14	14	Kabupaten Pangajene	Cluster_2	14	69.170	10.070	24.400	2017
15	15	Kabupaten Pangajene	Cluster_2	15	68.800	20.900	24.400	2017
16	16	Kabupaten Sukabumi	Cluster_2	16	60.000	19.100	20.900	2017
17	17	Kabupaten Pangajene	Cluster_2	17	49.770	17.300	16.900	2017
18	18	Kabupaten Pangajene	Cluster_2	18	44.000	22.010	13.900	2017
19	19	Kabupaten Sukabumi	Cluster_2	19	79.100	16.770	13.900	2017
20	20	Kabupaten Sukabumi	Cluster_2	20	79.100	16.770	13.900	2017

Gambar 9. Tampilan Hasil Cluster Pada Data View

Gambar 9 merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya berjumlah 35 data.

b. Cluster Model

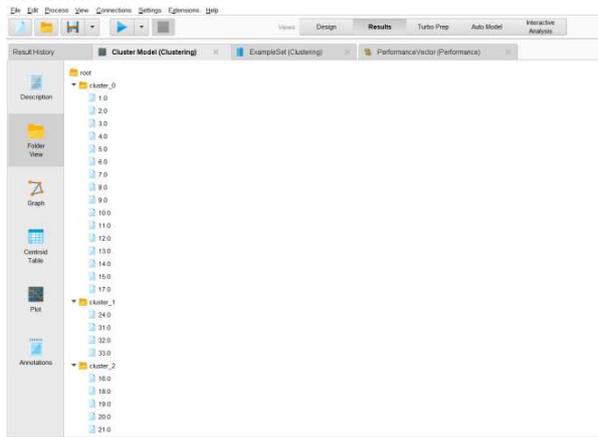
Tampilan hasil seperti gambar 10 berupa hasil pengujian dataset yang berjumlah 35 data menggunakan software rapidminer 10.3, sehingga berbentuk 3 cluster. Pada cluster 0 (cluster pertama) terdapat 16 data, cluster 1 (cluster kedua) terdapat 4 data dan cluster 2 (cluster ketiga) terdapat 15 data.



Gambar 10. Cluster Model Hasil Data Mining

c. Folder View

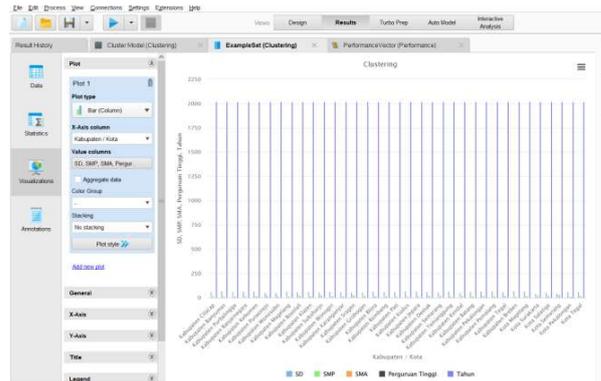
Pada gambar 11 dibawah ini merupakan gambar Folder View jumlah siswa pada setiap kabupaten atau kota pada tingkat (SD, SMP, SMA, Perguruan Tinggi) dengan status layak menerima beasiswa PIP seperti pada gambar 9 berikut:



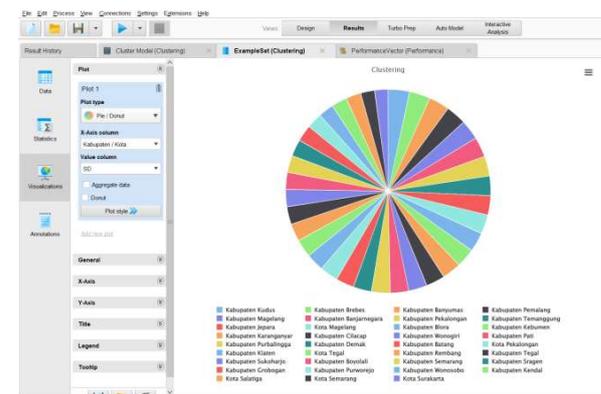
Gambar 11. Folder View

d. Visualization Hasil Clustering

Chart merupakan tampilan grafik hasil pengelompokan atau cluster data siswa pada setiap kabupaten atau kota pada tingkat (SD, SMP, SMA, Perguruan Tinggi) menggunakan 6 variabel penelitian dengan jumlah 3 cluster. Berikut ini tampilan visualization gambar Tampilan Bar (colomn) pada chart dan Tampilan Pie pada gambar 11 dan 12 berikut:



Gambar 12. Tampilan Bar (colomn) pada Chart



Gambar 13. Tampilan Pie pada Chart

IV. KESIMPULAN

Berdasarkan hasil dari penelitian ini dapat disimpulkan penerapan metode Algoritma K-Means Clustering untuk meningkatkan proses seleksi penerima manfaat PIP.

Algoritma K-Means mengelompokkan data siswa berdasarkan kesamaan karakteristik jumlah pelajar per kabupaten atau kota pada jenjang yang berbeda (SD, SMP, SMA, Perguruan Tinggi) dengan membuat tiga cluster yaitu cluster 0 berisi 16 data, cluster 1 terdiri dari 4 data, dan cluster 2 terdapat 15 data.

Maka, untuk cluster yang diutamakan yaitu cluster 1 karena paling rendah dengan 4 data. Cara ini membuktikan peningkatan akurasi dan efisiensi penyaluran dukungan Program Indonesia Pintar (PIP) sehingga meminimalkan risiko kesalahan dalam proses seleksi. Studi ini menunjukkan bahwa Algoritma K-Means Clustering dapat menjawab dan mendukung kebijakan pendidikan yang lebih tepat sasaran,

sehingga menghasilkan distribusi bantuan yang lebih merata dan efektif.

V. SARAN

Penelitian ini dapat dikembangkan lebih lanjut dengan menambah variabel relevan seperti tingkat kemiskinan dan aksesibilitas pendidikan untuk meningkatkan akurasi pengelompokan. Penggunaan metode clustering alternatif, seperti DBSCAN atau Hierarchical Clustering, juga dapat dipertimbangkan untuk membandingkan performa. Selain itu, integrasi hasil penelitian ke dalam sistem penunjang keputusan berbasis teknologi akan mempermudah implementasi dilapangan. Studi lanjutan pada berbagai daerah disarankan untuk memastikan hasil yang lebih relevan dan aplikatif.

DAFTAR PUSTAKA

- Dayera, Musa Bundaris Palungan, F. O. (2024). G-Tech : Jurnal Teknologi Terapan. *G-Tech : Jurnal Teknologi Terapan*, 8(1), 186–195. <https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229>
- Dewi, F. P., Aryni, P. S., & Umaidah, Y. (2022). Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 7(2), 111–121. <https://doi.org/10.14421/jiska.2022.7.2.111-121>
- Fatmawati, K., & Windarto, A. P. (2018). Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi. *Computer Engineering, Science and System Journal*, 3(2), 173. <https://doi.org/10.24114/cess.v3i2.9661>
- Hadiza, N., Sari, W. N., & Afriyadi, H. (2023). Perancangan Sistem Informasi Pengelolaan Surat Masuk dan Surat Keluar pada Kantor Dinas Pengendalian Penduduk dan Keluarga Berencana Kabupaten Sarolangun. 05(04), 11406–11414.
- Informasi, S., Muasaroh, Y. I., & Fatah, Z. (2024). *Jurnal Advance Research Informatika Implementasi RapidMiner dalam Optimasi Pembentukan Kelas Unggulan Menggunakan K-Means Clustering*. 3(1), 66–72.
- Irwansyah, E., & Faisal, M. (2015). *Advanced Clustering: Teori dan Aplikasi*. DeePublish.
- Jundanuddin, M., Fatah, Z., Munazilin, A., Informasi, S., Sains, F., Sukorejo, U. I., Komputer, I., Sains, F., & Sukorejo, U. I. (2024). *Implementasi Data Mining Pada Penilaian Kinerja Guru Menggunakan Metode K-Nearest Neighbor (K-NN) Studi Kasus Pada Mi Alfagiri Silo-Jember Implementation Of Data Mining On Teacher Performance Appraisal Using K-Nearest Neighbor (K-NN) Method Case Study At Mi Alfagiri Silo-Jember*. 13(105), 869–880.
- Rady Putra, L. G., & Anggrawan, A. (2021). Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 21(1), 205–214. <https://doi.org/10.30812/matrik.v21i1.1554>
- Rakista, P. M. (2021). Implementasi Kebijakan Program Indonesia Pintar (PIP). *Sawala : Jurnal Administrasi Negara*, 8(2), 224–232. <https://doi.org/10.30656/sawala.v8i2.2774>
- Rohmah, A., Sembiring, F., & Erfina, A. (2021). Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring. *SISMATIK (Seminar Nasional Sistem Informasi Dan Manajemen Informatika)*, 290–298. <https://www.alfasoleh.com/2019/11/k-means-clustering-contoh>
- Salam, A., Adiatma, D., & Zeniarja, J. (2020). Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS. *JOINS (Journal of Information System)*, 5(1), 62–68. <https://doi.org/10.33633/joins.v5i1.3350>
- Triandini, M., Defit, S., & Nurcahyo, G. W. (2021). Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas

Cendekia. *Jurnal Informasi Dan Teknologi*,
3, 167–173.
<https://doi.org/10.37034/jidt.v3i3.120>