



Implementasi Metode Algoritma K-Means Clustering Untuk Menentukan Penerima Program Indonesia Pintar (PIP)

Moh. Syahrul Iskandar^{1*}, Zaehol Fatah²

¹ Teknologi Informasi, Universitas Ibrahimy

² Sistem Informasi, Universitas Ibrahimy

^{1*}syahrulganteng975@email.com, ²zaeholfatah@email.com

Abstrak

Program Indonesia Pintar (PIP) merupakan program pemerintah yang bertujuan untuk meningkatkan akses pendidikan bagi siswa dari keluarga kurang mampu. Mekanisme seleksi yang efektif dan akurat diperlukan untuk memastikan bantuan tepat sasaran. Algoritma *k-means clustering* untuk mengelompokkan siswa berdasarkan indikator tertentu, seperti jumlah siswa di sekolah yang berada pada kabupaten atau kota di tingkat (SD, SMP, SMA, dan Perguruan Tinggi). Tujuan dari metode ini untuk memudahkan pengambilan keputusan dalam menentukan penerima PIP dengan mengidentifikasi kelompok siswa yang paling membutuhkan dukungan. *K-means clustering* bekerja dengan membagi data menjadi beberapa *cluster* tergantung pada kesamaan pola fitur yang digunakan. Proses ini memungkinkan untuk mengidentifikasi kelompok siswa dengan prioritas dukungan berbeda, misalnya siswa dengan kebutuhan tinggi dan rendah. Pendekatan ini diharapkan dapat menjadikan hasil seleksi penerima PIP lebih obyektif dan efisien, serta membantu agar dapat diberikan dengan lebih akurat dan merata. Algoritma *K-means* dapat mengungkap pola tersembunyi pada data pendidikan dan lebih mendukung proses distribusi PIP. Hasil ini membantu pengambil kebijakan meningkatkan kualitas program dan memastikan bahwa dukungan menjangkau siswa yang benar – benar membutuhkannya.

Kata Kunci: Algoritma *K-Means Clustering*, Program Indonesia Pintar, Seleksi Penerima, Bantuan Pendidikan, Pengelompokan data.

PENDAHULUAN

Pada dasarnya *Algoritma K-Means Clustering* merupakan bidang penelitian dalam analisis dan *data mining*. Pada algoritma ini teknik pengelompokan berdasarkan kemiripan data yang tidak memiliki acuan apapun, tetapi akan membagi keseluruhan data yang akan menjadi kelompok atau mempunyai kemiripan yang sama[1].

Pendidikan merupakan unsur terpenting dalam membentuk karakter dan budaya bangsa, terlebih bangsa yang memiliki masyarakat yang majemuk seperti halnya Indonesia. Pendidikan berfungsi untuk meningkatkan kemampuan manusia, dalam membentuk karakter serta peradaban bangsa yang bermartabat dalam rangka mencerdaskan kehidupan bangsa[2]. Tetapi, tidak semua siswa di Indonesia, terutama yang berasal dari keluarga keterbatasan finansial, mempunyai kesempatan pendidikan yang sama. Untuk mengatasi kendala tersebut, pemerintah meluncurkan Program Indonesia Pintar (PIP) melalui Kementerian Pendidikan, Kebudayaan dan Teknologi.

Program Indonesia Pintar (PIP) bertujuan untuk mencegah siswa didik dari kemungkinan putus sekolah atau tidak melanjutkan pendidikan akibat kesulitan ekonomi[3]. Dan untuk memberikan dukungan finansial kepada siswa yang membutuhkan untuk memastikan mereka dapat melanjutkan pendidikan dan mengurangi angka putus sekolah. Namun proses seleksi penerima PIP masih menghadapi beberapa kendala. Dengan jumlah siswa dan mahasiswa yang begitu besar dan kondisi perekonomian yang berbeda – beda di setiap daerah, diperlukan pendekatan yang tepat dan baik dalam mengambil keputusan pendanaan.

Tantangan ini memerlukan metode berbasis data yang lebih efektif untuk mendukung pengambilan keputusan. Di sinilah *Algoritma K-Means Clustering* berperan penting. Hal ini memungkinkan untuk pengelompokan data siswa berdasarkan kesamaan karakteristik, sehingga proses seleksi menjadi lebih cepat dan objektif. Penelitian ini bertujuan untuk menerapkan *algoritma k-means clustering* pada siswa kelompok dengan mempertimbangkan berbagai indikator. Jumlah siswa pada setiap kabupaten atau kota pada tingkat (SD, SMP, SMA, dan Perguruan Tinggi). Dengan mengelompokkan data siswa ke dalam beberapa *cluster*, pemerintah dapat lebih mudah menentukan kelompok mana yang paling membutuhkan dukungan.

Hasil pengelompokan ini diharapkan dapat meningkatkan akurasi penyaluran dukungan penerima PIP dan meminimalisir kemungkinan kesalahan dalam proses pemilihan penerima. Melalui kajian ini diharapkan penerapan teknologi *data mining* seperti *k-means clustering* akan memberikan dampak positif terhadap pengaturan kebijakan pendidikan, khususnya dalam hal memastikan sasaran bantuan. Cara ini bertujuan agar PIP lebih efektif dan memberikan manfaat yang lebih merata kepada pelajar di seluruh Indonesia.

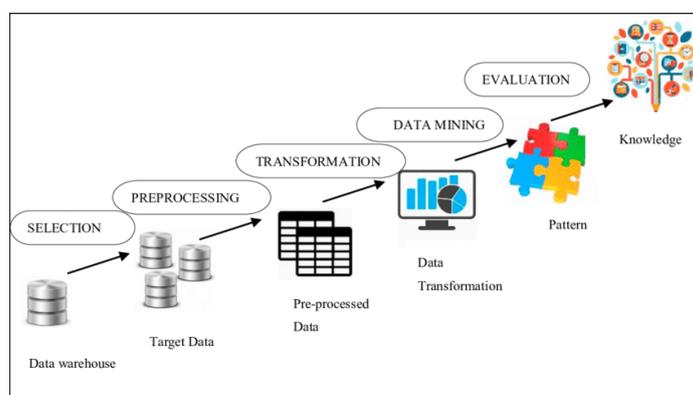
METODE PENELITIAN

Data Mining

Data Mining adalah sebuah proses pencarian secara otomatis informasi yang berguna dalam tempat penyimpanan data berukuran besar. Teknik data mining digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna. Namun tidak semua pekerjaan pencarian informasi dapat dinyatakan sebagai data mining[4]. Data mining memiliki banyak metode, metode data mining yang digunakan adalah *Knowledge Discovery in Database (KDD)* ataupun dapat disebut *Pattem Recognition*.

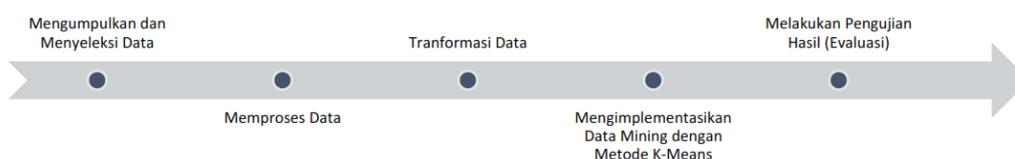
Knowledge Discovery in Database (KDD)

Knowledge discovery in database (KDD) adalah metode teknis yang berguna untuk mencari dan mengidentifikasi pola (*pattem*) 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[5]. 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[6].



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

Clustering

Clustering atau klusterisasi adalah suatu teknik atau metode untuk mengelompokkan data. Menurut Tan, 2006 *clustering* adalah sebuah proses untuk pengelompokan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum.

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[7].

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 dinalisis[8]. Kelebihan dari penerapan *Algoritma K-Means* merupakan mampu mengelompokkan objek besar serta meningkatkan kecepatan proses pengelompokan.

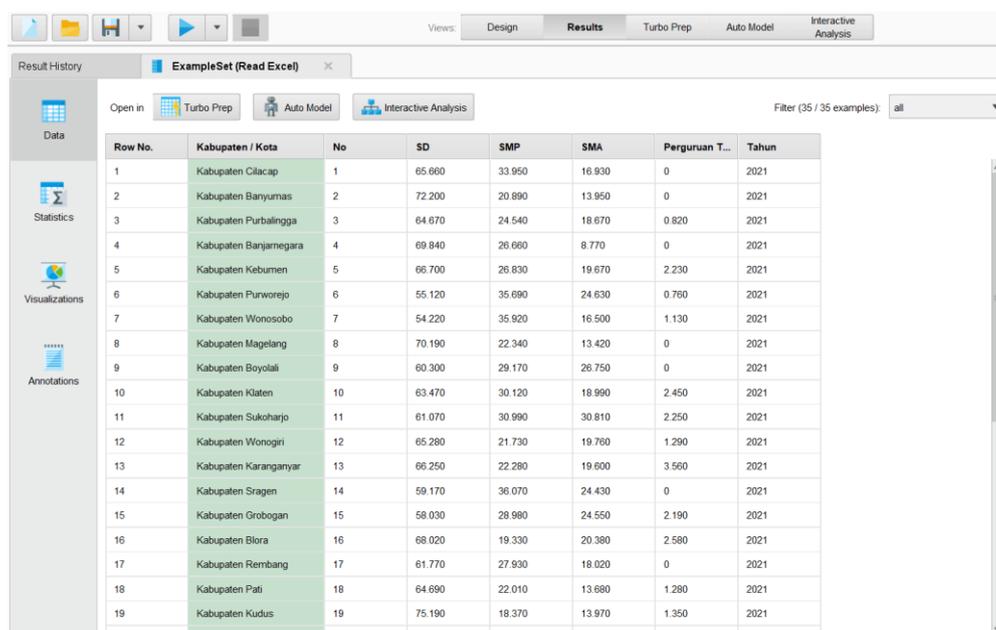
HASIL DAN PEMBAHASAN

Implementasi penelitian ini menggunakan perhitungan *Algoritma K-Means Clustering* yang dilakukan dengan tools Rapid Miner Studio. Rapid Miner sebagai tools yang mampu memudahkan usernya dalam melakukan operator. Operator berfungsi untuk memodifikasi data dan selanjutnya dihubungkan dengan node – node operator[9].

Dari proses penelitian yang dilakukan didapatkan hasil dari 35 dataset dari jumlah siswa pada setiap kabupaten atau kota pada tingkat (SD, SMP, SMA, Perguruan Tinggi) tahun 2021. Terdapat 16 data mengikuti cluster 0 dan terdapat 15 data mengikuti cluster 2 yang artinya dapat dipertimbangkan untuk bantuan tambahan atau dengan prioritas lebih rendah sedangkan terdapat 4 data mengikuti cluster 1 yang artinya siswa dapat difokuskan sebagai penerima utama karena berisi daerah yang membutuhkan bantuan lebih mendesak. Dengan membagi data ke dalam tiga cluster, pemerintah dapat lebih mudah menentukan prioritas penyaluran penerima PIP.

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[10]. Data penerima beasiswa yang digunakan untuk perhitungan sebanyak 35 data dan 6 variabel yang telah didapatkan dari dataset Badan Pusat Statistik (BPS) Indonesia. Seperti ditampilkan pada gambar 1 berikut ini:



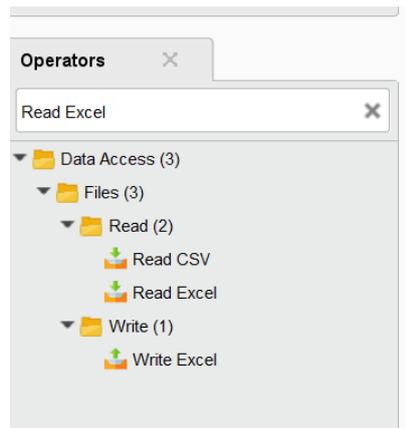
Row No.	Kabupaten / Kota	No	SD	SMP	SMA	Perguruan T...	Tahun
1	Kabupaten Cilacap	1	65.660	33.950	16.930	0	2021
2	Kabupaten Banyumas	2	72.200	20.890	13.950	0	2021
3	Kabupaten Purbalingga	3	64.670	24.540	18.670	0.820	2021
4	Kabupaten Banjarnegara	4	69.840	26.660	8.770	0	2021
5	Kabupaten Kebumen	5	66.700	26.830	19.670	2.230	2021
6	Kabupaten Purworejo	6	55.120	35.690	24.630	0.760	2021
7	Kabupaten Wonosobo	7	54.220	35.920	16.500	1.130	2021
8	Kabupaten Magelang	8	70.190	22.340	13.420	0	2021
9	Kabupaten Boyolali	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	30.810	2.250	2021
12	Kabupaten Wonogiri	12	65.280	21.730	19.760	1.290	2021
13	Kabupaten Karanganyar	13	66.250	22.280	19.600	3.560	2021
14	Kabupaten Sragen	14	59.170	36.070	24.430	0	2021
15	Kabupaten Grobogan	15	58.030	28.980	24.550	2.190	2021
16	Kabupaten Blora	16	68.020	19.330	20.380	2.580	2021
17	Kabupaten Rembang	17	61.770	27.930	18.020	0	2021
18	Kabupaten Pati	18	64.690	22.010	13.680	1.280	2021
19	Kabupaten Kudus	19	75.190	18.370	13.970	1.350	2021

Gambar 1. Testing data- load current (amperes)

Pengolahan Data dengan software Rapidminer

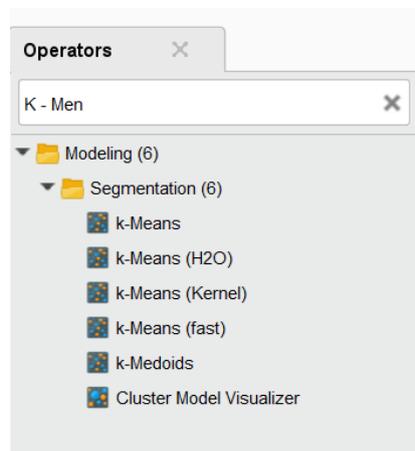
Untuk mengelompokkan data dalam penelitian ini, akan dilakukan dengan menggunakan software rapidminer 10.3 sebagai berikut:

- Buka *Software Rapidminer Studio Versi 10.3*.
- Lalu pilih menu file kemudian pilih *new* proses.
- Selanjutnya masuk kedalam panel operator kemudian ketik *Read Excel*, lalu drag operator *read excel* agar operator tersebut muncul di panel proses dan bisa digunakan untuk memasukkan data *excel* yang telah melalui tahap data *cleansing* dan data *transformation*. Seperti yang terlihat pada gambar 2 berikut:



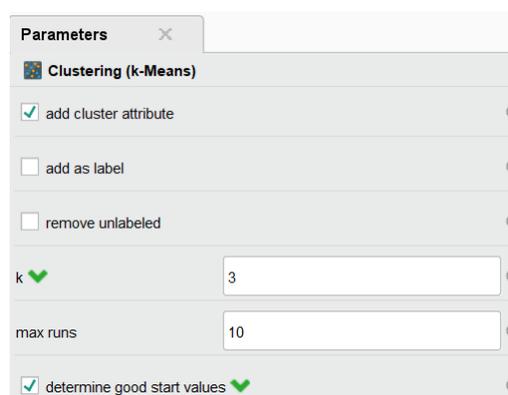
Gambar 2. Read Excel

- d. 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 3 berikut:



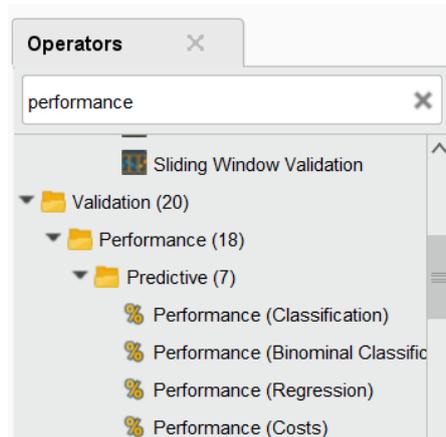
Gambar 3. Memilih Pemodelan Clustering

- e. Selanjutnya, pada proses *clustering* tentukan terlebih dahulu jumlah klasternya menjadi $K=3$ dan *mixed measure* menjadi *mixed Euclidean distance* seperti pada gambar 4 berikut:



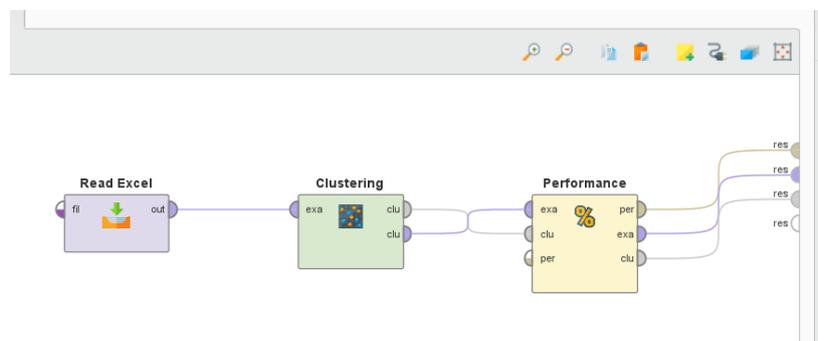
Gambar 4. Menentukan Jumlah K

- 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 5 berikut:



Gambar 5. Memilih Pemodelan *Performance (Classification)*

g. Selanjutnya adalah menghubungkan konektor masing – masing proses pada main process seperti pada gambar 6 dan selanjutnya klik button run untuk memulai pengklusteran.



Gambar 6. Main Proses Data Mining

Hasil Klasterisasi 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 7 akan ditampilkan data view hasil clusterisasi yang telah didapatkan sebagai berikut:

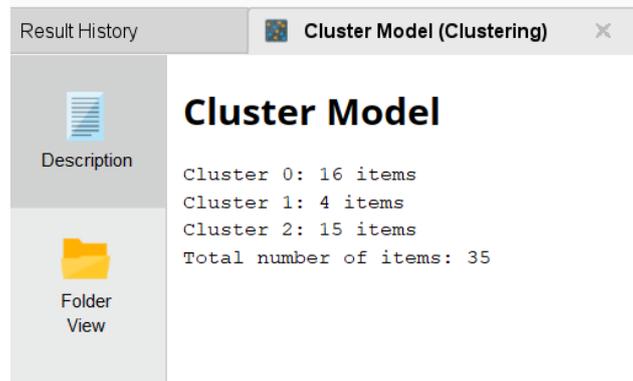
Row No.	Id	Kabupaten / Kota	cluster	No	SD	SMP	SMA	Perguruan T.	Tahun
1	1	Kabupaten Cilacap	cluster_0	1	65.660	33.950	16.930	0	2021
2	2	Kabupaten Banyumas	cluster_0	2	72.200	20.890	13.950	0	2021
3	3	Kabupaten Purbalingga	cluster_0	3	64.670	24.540	18.670	0.820	2021
4	4	Kabupaten Banjarnegara	cluster_0	4	69.840	26.660	8.770	0	2021
5	5	Kabupaten Kebumen	cluster_0	5	66.700	26.930	19.670	2.230	2021
6	6	Kabupaten Purworep	cluster_0	6	55.120	35.690	24.630	0.760	2021
7	7	Kabupaten Wonosobo	cluster_0	7	54.220	35.920	16.500	1.130	2021
8	8	Kabupaten Magelang	cluster_0	8	70.190	22.340	13.420	0	2021
9	9	Kabupaten Boyali	cluster_0	9	60.300	29.170	26.750	0	2021
10	10	Kabupaten Kliten	cluster_0	10	63.470	30.120	18.990	2.450	2021
11	11	Kabupaten Sukoharjo	cluster_0	11	61.070	30.990	30.810	2.250	2021
12	12	Kabupaten Wonogri	cluster_0	12	65.280	21.730	19.760	1.290	2021
13	13	Kabupaten Karanganyar	cluster_0	13	66.250	22.280	19.600	3.560	2021
14	14	Kabupaten Sragen	cluster_0	14	59.170	36.070	24.430	0	2021
15	15	Kabupaten Grobogan	cluster_0	15	58.030	29.980	24.550	2.190	2021
16	16	Kabupaten Blora	cluster_2	16	68.020	19.330	20.380	2.580	2021
17	17	Kabupaten Rembang	cluster_0	17	61.770	27.930	18.020	0	2021
18	18	Kabupaten Pati	cluster_2	18	64.690	22.910	13.680	1.280	2021
19	19	Kabupaten Kudus	cluster_2	19	75.190	18.370	13.970	1.350	2021

Gambar 7. Tampilan Hasil Cluster Pada Data View

Gambar 7 merupakan sheet untuk menampilkan database yang telah diolah secara keseluruhan lengkap dengan clusternya berjumlah 35 data.

b. Cluster Model

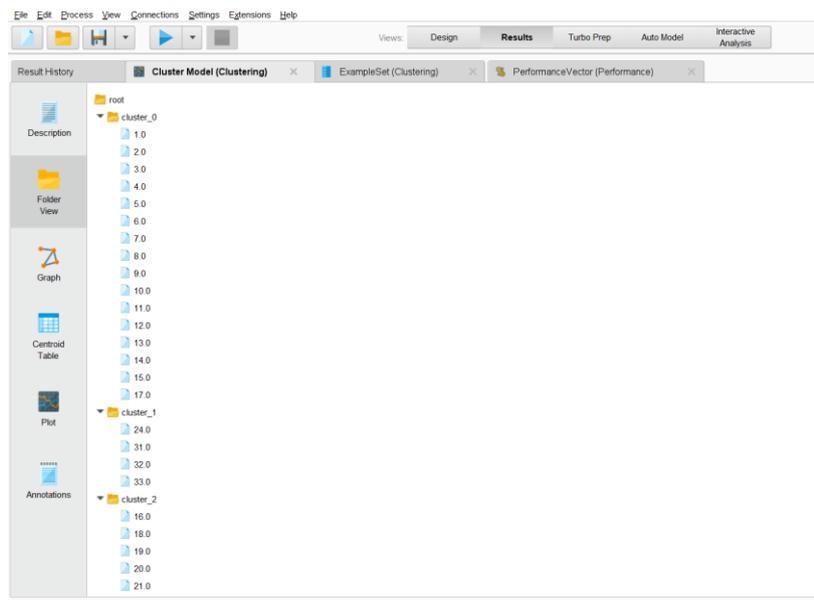
Tampilan hasil seperti gambar 8 berupa hasil pengujian dataset yang berjumlah 35 data menggunakan software rapidminer 10.3, sehingga berbentuk 3 cluster. Pada cluster 0 (cluster pertama) terdapat 16 data, cluster 1 (cluster kedua) terdapat 4 data dan cluster 2 (cluster ketiga) terdapat 15 data.



Gambar 8. Cluster Model Hasil Data Mining

c. Folder View

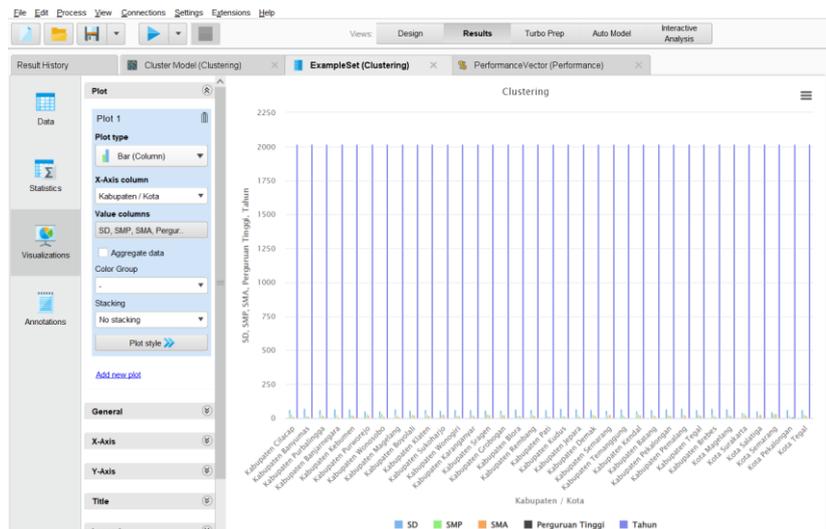
Pada gambar 9 dibawah ini merupakan gambar Folder View jumlah siswa pada setiap kabupaten atau kota pada tingkat (SD, SMP, SMA, Perguruan Tinggi) dengan status layak menerima beasiswa PIP seperti pada gambar 9 berikut:



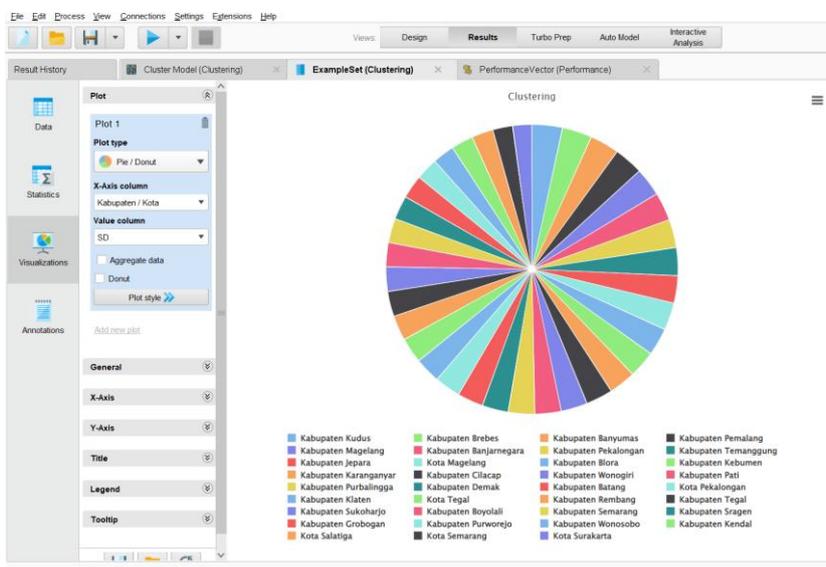
Gambar 9. Folder View

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 (column) pada chart dan Tampilan Pie pada gambar 10 dan 11 berikut:



Gambar 10. Tampilan Bar (column) pada Chart



Gambar 11. Tampilan Pie pada Chart

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.

UCAPAN TERIMA KASIH

Puji dan Syukur saya panjatkan kehadirat Allah SWT. yang telah melimpahkan taufiq dan hidayah-nya kepada kami, sehingga saya dapat menyelesaikan jurnal ini. Terima kasih yang sebesar – besarnya kepada dosen pembimbing yang telah mengarahkan saya dengan penuh kesabaran. Dan juga terima kasih untuk segenap dukungan, khususnya kepada teman – teman yang selalu mengingatkan saya akan terselesaikannya jurnal ini. Tanpa adanya semua dukungan ini, sulit bagi saya untuk menyelesaikan jurnal ini.

DAFTAR PUSTAKA

- [1] A. Rohmah, F. Sembiring, and A. Erfina, "Implementasi Algoritma K-Means Clustering Analysis Untuk Menentukan Hambatan Pembelajaran Daring," *SISMATIK (Seminar Nas. Sist. Inf. dan Manaj. Inform.*, pp. 290–298, 2021, [Online]. Available: <https://www.alfasoleh.com/2019/11/k-means-clustering-contoh>
- [2] F. O. Dayera, Musa Bundaris Palungan, "G-Tech : Jurnal Teknologi Terapan," *G-Tech J. Teknol. Terap.*, vol. 8, no. 1, pp. 186–195, 2024, [Online]. Available: <https://ejournal.uniramalang.ac.id/index.php/g-tech/article/view/1823/1229>
- [3] S. Kasus and K. Banyumas, "IMPLEMENTASI KEBIJAKAN PROGRAM INDONESIA PINTAR (PIP) (Studi Kasus pada Sekolah Dasar di Kabupaten Banyumas)," vol. 8, pp. 224–232, 2020.
- [4] Anggada Maulana, "Konsep Dasar Data Mining," *Konsep Data Min.*, vol. 1, pp. 1–16, 2018.
- [5] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
- [6] N. Hadiza, W. N. Sari, and H. Afriyadi, "Perancangan Sistem Informasi Pengelolaan Surat Masuk dan Surat Keluar pada Kantor Dinas Pengendalian Penduduk dan Keluarga Berencana Kabupaten Sarolangun," vol. 05, no. 04, pp. 11406–11414, 2023.
- [7] E. Irwansyah and M. Faisal, *Advanced Clustering: Teori dan Aplikasi*. DeePublish, 2015. [Online]. Available: <https://books.google.co.id/books?id=8y80BgAAQBAJ>
- [8] M. Triandini, S. Defit, and G. W. Nurcahyo, "Data Mining dalam Mengukur Tingkat Keaktifan Siswa dalam Mengikuti Proses Belajar pada SMP IT Andalas Cendekia," *J. Inf. dan Teknol.*, vol. 3, pp. 167–173, 2021, doi: 10.37034/jjdt.v3i3.120.
- [9] A. Salam, D. Adiatma, and J. Zeniarja, "Implementasi Algoritma K-Means Dalam Pengklasteraan untuk Rekomendasi Penerima Beasiswa PPA di UDINUS," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 62–68, 2020, doi: 10.33633/joins.v5i1.3350.
- [10] L. G. Rady Putra and A. Anggrawan, "Pengelompokan Penerima Bantuan Sosial Masyarakat dengan Metode K-Means," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 1, pp. 205–214, 2021, doi: 10.30812/matrik.v21i1.1554.