

Implementasi RapidMiner Untuk Menentukan Siswa Unggulan Menggunakan Metode K-Means

Oleh :

Nabella Rosyefa Wahyudi,

Yunianita Rahmawati

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

April, 2024

Pendahuluan

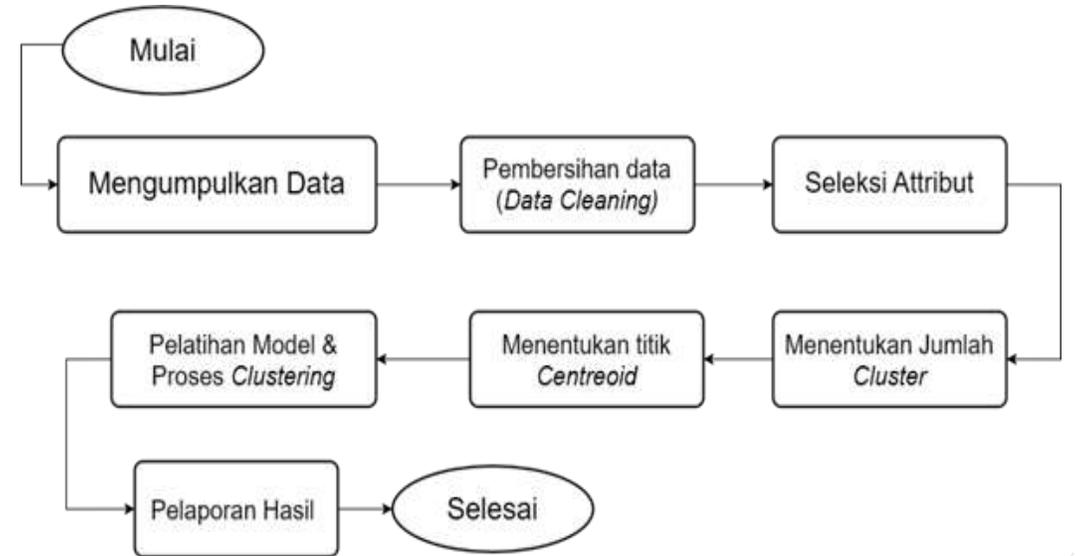
SMP Negeri 11 SAMPIT mengalami kesulitan dalam menentukan siswa lomba OSN (OLIMPIADE SAINS NASIONAL), hal ini disebabkan karena faktor penempatan siswa yang cenderung berubah – rubah setiap semester. Penelitian ini didorong untuk mengatasi tantangan tersebut dengan menerapkan metode data mining clustering, terutama Algoritma K-Means. Metode data mining clustering, terutama Algoritma K-Means. Pemilihan Algoritma K-Means dilatarbelakangi oleh tingkat keakuratan yang tinggi, kesederhanaan, dan kemudahan penerapannya dalam data mining. Dukungan dari penelitian terdahulu yang serupa memperkuat keyakinan bahwa metode ini efektif untuk meningkatkan pengelompokan siswa unggulan. Oleh karena itu, penelitian ini tidak hanya memberikan solusi praktis untuk SMP Negeri 11 Sampit, tetapi juga membangun fondasi pada konsep yang telah terbukti berhasil dalam situasi serupa.

Pertanyaan Penelitian (Rumusan Masalah)

Berdasarkan latar belakang yang telah dibahas di atas, maka dapat disimpulkan bahwa rumusan masalah penelitian ini adalah bagaimana penerapan algoritma K-means dalam pengelompokan siswa berbakat di SMP Negeri 11 Sampit.

Metode

Gambar 1 menampilkan proses data mining dari Metode K-Means yang diterapkan pada penelitian. Penelitian ini menggunakan Metode K-Means. Metode K-Means merupakan metode clustering yang mengelompokkan data yang mirip satu sama lain ke dalam cluster yang sama. Data yang memiliki kemiripan atau sama akan dikelompokkan ke dalam satu cluster dan yang berbeda akan dikelompokkan ke cluster yang lain.



Gambar 1. Proses data mining

Pembahasan

Gambar 2 menampilkan datasheet nilai siswa kelas 7 yang akan di kelolah oleh peneliti. *Dataset* yang diolah memiliki 24 *attribute* sama seperti Nama, Agm_PA, Agm_KA, PPkn_PA, PPkn_KA, Bind_PA, Bind_KA, Mat_PA, Mat_KA, IPA_PA, IPA_KA, IPS_PA, IPS_KA, Bing_PA, Bing_KA, Seni_PA, Seni_KA, Jas_PA, Jas_KA, Pra_PA, Pra_KA. *Dataset* memiliki jumlah data siswa sebanyak 31 data. *Dataset* ini diperoleh dari guru kurikulum kelas 7 SMP Negeri 11 Sampit.

Row No.	Nama	cluster	Agm_PA2	Agm_KA2	PPkn_PA2	PPkn_KA2	Bind_PA2	Bind_KA2	Mat_PA2	Mat_KA2	IPA_PA2	IPA_KA2	IPS_PA2
1	ADITYA RAMA...	cluster_0	87	86	78	78	78	79	76	71	71	80	80
2	Akbar Risky S...	cluster_0	87	86	78	78	79	80	75	73	71	85	81
3	Alam Setiago	cluster_0	92	94	78	78	81	83	78	78	73	85	85
4	Amelsa Putri ...	cluster_0	92	94	79	79	84	82	76	78	73	85	85
5	APRIL SIDIK	cluster_0	88	94	78	78	80	81	79	78	73	85	86
6	Aulia Aini Nur...	cluster_0	90	94	78	78	82	80	80	77	71	85	84
7	Ayu Sufiah	cluster_0	89	94	78	78	85	80	78	83	75	85	85
8	Beri Abdi Pa...	cluster_0	86	86	78	78	78	79	72	71	53	80	80
9	Caesha Afril ...	cluster_0	92	94	80	80	85	85	79	78	77	85	87
10	Dendy Putra ...	cluster_0	89	94	78	78	81	80	76	73	71	85	83
11	DIDA YULIAN...	cluster_0	92	94	80	80	87	84	79	78	73	85	85
12	DIMAS PUTRA	cluster_0	90	94	78	78	81	80	73	71	70	80	84
13	Gesang Mukti...	cluster_0	92	94	80	80	81	85	81	78	72	85	86
14	IKA ARIANIS ...	cluster_0	89	94	78	78	85	82	81	78	73	85	84
15	Ilyas Abdullah	cluster_0	88	94	80	80	80	81	75	73	78	85	86
16	M. Jordi Apry...	cluster_0	87	94	78	78	83	83	78	70	71	85	83
17	M. Naufal Tufail	cluster_0	90	94	78	78	80	80	78	81	72	85	86
18	M. RASYA HA	cluster_0	92	94	80	80	82	83	82	78	87	85	89

ExampleSet (31 examples, 2 special attributes, 20 regular attributes)

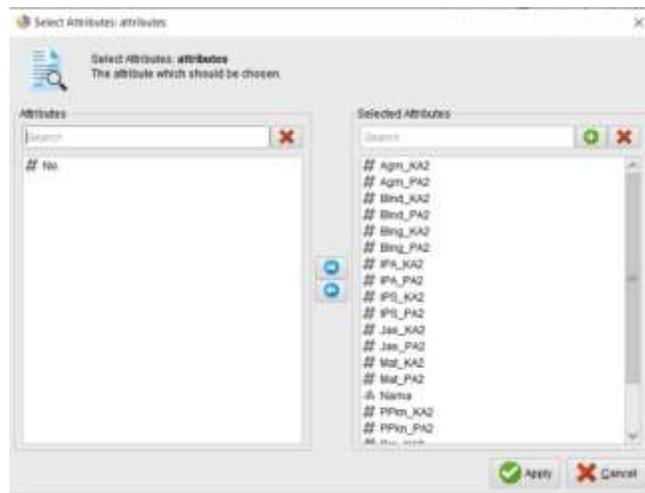
Gambar 2. Data Nilai Siswa Kelas 7

Pembahasan

Berikut hasil dan pembahasan yang sudah dilakukan oleh peneliti dengan menggunakan tools yang sudah dijelaskan diatas.

1. Data Selection

Gambar 3 menampilkan datasheet nilai siswa kelas 7 yang akan di kelolah oleh peneliti. *Dataset* yang diolah memiliki 24 *attribute* sama seperti Nama, Agm_PA, Agm_KA, PPkn_PA, PPkn_KA, Bind_PA, Bind_KA, Mat_PA, Mat_KA, IPA_PA, IPA_KA, IPS_PA, IPS_KA, Bing_PA, Bing_KA, Seni_PA, Seni_KA, Jas_PA, Jas_KA, Pra_PA, Pra_KA. *Dataset* memiliki jumlah data siswa sebanyak 31 data. *Dataset* ini diperoleh dari guru kurikulum kelas 7 SMP Negeri 11 Sampit.



Gambar 3. Proses seleksi data

2. Preprocessing

Setelah proses seleksi data nilai selesai, langkah berikutnya adalah melakukan pre-processing data pada hasil dari tahapan sebelumnya. Pre-Processing bertujuan untuk membersihkan data yang tidak perlu digunakan / data kosong / data ganda dan sejenisnya. Gambar 4 menampilkan stastitika hasil seleksi data kelas 7 setelah dilakukan *running* di *software* Rapid Minner.

Attribute	Type	Missing	Minimum	Maximum	Standard Deviation
Nama	Text	0	YUSMANI JEMEN NORTA PUTRI ARIYA LAMONGAN ANITAN ANIS NOR ...		
Agm_KA2	Nominal	0	1	34	87.032
Agm_PA2	Nominal	0	1	34	80.184
PPkn_KA2	Nominal	0	77	80	70.219
PPkn_PA2	Nominal	0	77	80	70.419
Bind_KA2	Nominal	0	75	80	82.258
Bind_PA2	Nominal	0	79	86	87.877
Mat_KA2	Nominal	0	82	82	78.258

Gambar 4. Hasil proses pre-processing

Pembahasan

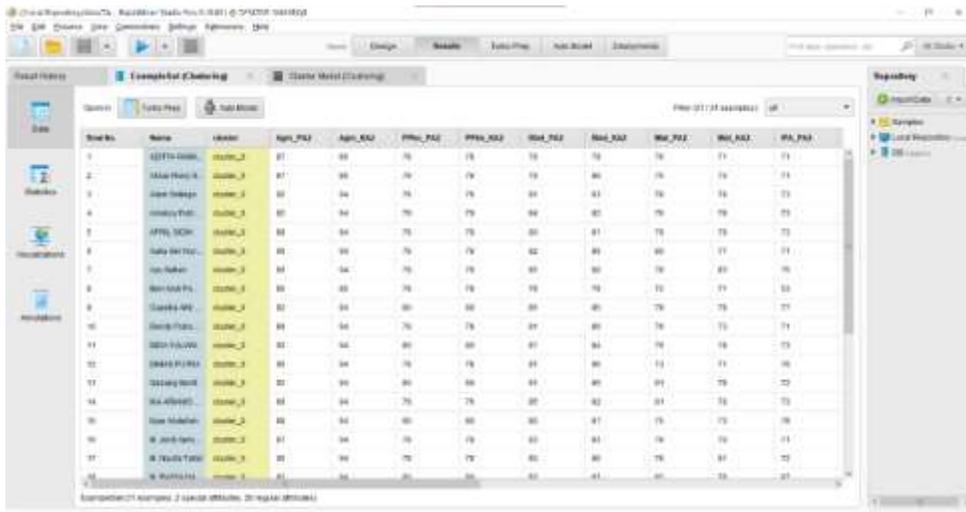
Berikut hasil dan pembahasan yang sudah dilakukan oleh peneliti dengan menggunakan tools yang sudah dijelaskan diatas.

3. Transformation Data

Setelah tahapan *pre-processing* selesai, langkah berikutnya adalah melakukan *tranformation* data pada hasil dari tahapan sebelumnya. Gambar 5 menampilkan hasil dari proses Transformation data yang sudah dilakukan di kelas 7. Pada tahapan Transformation data, Hasil data yang di dapat sudah berupa integer/angka sehingga pada tahapan ini tidak ada perlu diubah karena memenuhi syarat penerapan Algoritma K-Means.

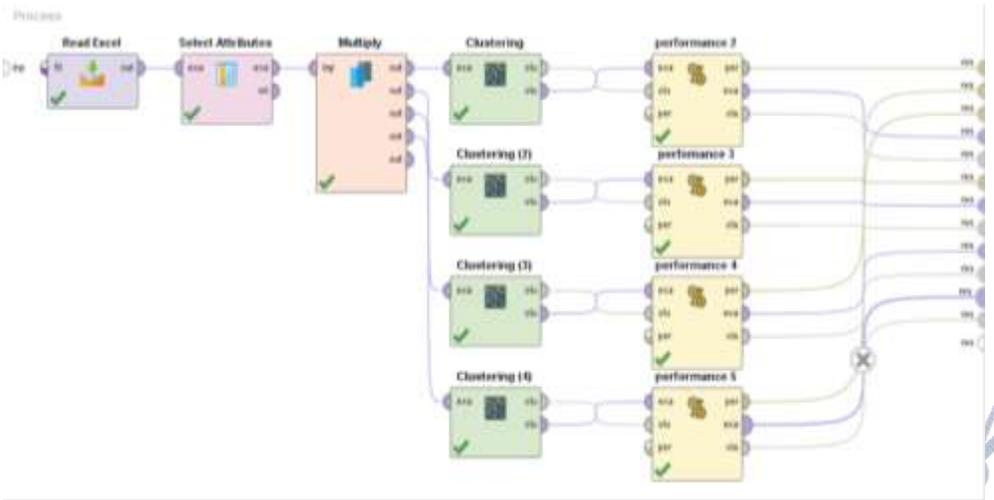
2. Data Mining

Gambar 6 diatas menampilkan proses *clustering* dilakukan menggunakan tools RapidMiner. Proses diatas menggunakan 2 - 4 cluster berbeda yang bermaksud untuk membandingkan nilai cluster lalu mencari nilai yang terbaik dari setiap cluster. Sebelum proses cluster dilakukan, data akan dibaca menggunakan operator *read excel* lalu selanjutnya dilakukan proses seleksi atribut menggunakan operator *select attribute*



No	Nama	Agri_702	Agri_802	PPH_702	PPH_802	IMA_702	IMA_802	IMU_702	IMU_802	PA_P04
1	ESTER GIAN	87	88	78	78	78	78	78	78	78
2	ALIA PRATIKA	87	88	78	78	78	78	78	78	78
3	ANNA DEWI	87	88	78	78	78	78	78	78	78
4	ANINDA PUTRI	87	88	78	78	78	78	78	78	78
5	AFRIYAH NICHOLAS	87	88	78	78	78	78	78	78	78
6	ADITYA ANANDA	87	88	78	78	78	78	78	78	78
7	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
8	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
9	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
10	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
11	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
12	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
13	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
14	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
15	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
16	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
17	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
18	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
19	ANANDA PUTRI	87	88	78	78	78	78	78	78	78
20	ANANDA PUTRI	87	88	78	78	78	78	78	78	78

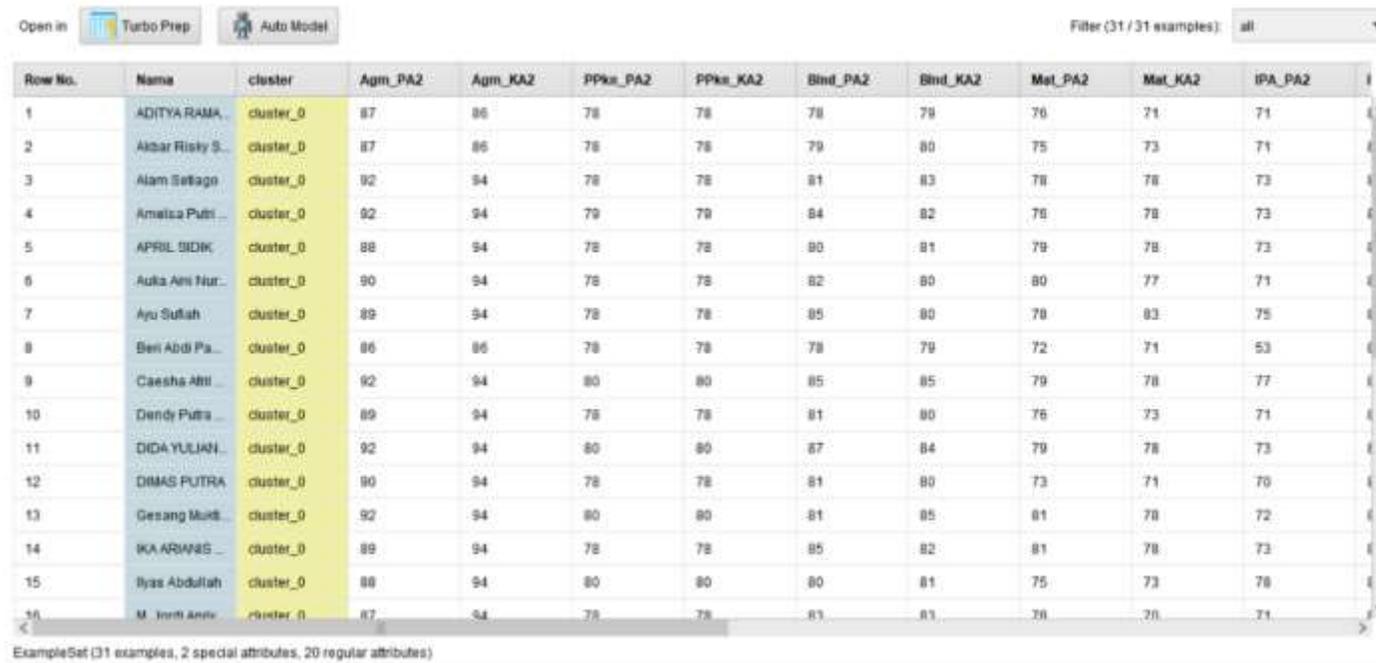
Gambar 5. Hasil Proses transformation data



Gambar 6. Hasil proses pre-processing data

Hasil

Cluster 0 merupakan representasi dari siswa unggul dan cluster 1 merupakan representasi dari siswa bukan unggulan dari seluruh mata pelajaran. Berikut di bawah ini Gambar 7 yang merupakan tampilan hasil dari keseluruhan data yang sudah melalui tahapan *Clustering* menggunakan Metode *K-Means* :



The screenshot shows a data table with 16 rows and 12 columns. The columns are: Row No., Nama, cluster, Agm_PA2, Agm_KA2, PPkn_PA2, PPkn_KA2, Bhd_PA2, Bhd_KA2, Mat_PA2, Mat_KA2, and IPA_PA2. All rows are assigned to 'cluster_0'. The scores for each subject are as follows:

Row No.	Nama	cluster	Agm_PA2	Agm_KA2	PPkn_PA2	PPkn_KA2	Bhd_PA2	Bhd_KA2	Mat_PA2	Mat_KA2	IPA_PA2
1	ADITYA RAMA	cluster_0	87	86	78	78	78	79	76	71	71
2	Aldar Risky S.	cluster_0	87	86	78	78	79	80	75	73	71
3	Alam Setiogo	cluster_0	92	94	78	78	81	83	78	78	73
4	Amelza Putri	cluster_0	92	94	79	79	84	82	76	78	73
5	APRIL SIDIK	cluster_0	88	94	78	78	80	81	79	78	73
6	Aula Ani Nur	cluster_0	90	94	78	78	82	80	80	77	71
7	Ayu Sufah	cluster_0	89	94	78	78	85	80	78	83	75
8	Beri Addi Pa	cluster_0	86	86	78	78	78	79	72	71	53
9	Caesha Atli	cluster_0	92	94	80	80	85	85	79	78	77
10	Dendi Putra	cluster_0	89	94	78	78	81	80	76	73	71
11	DIDA YULIAN	cluster_0	92	94	80	80	87	84	79	78	73
12	DIMAS PUTRA	cluster_0	80	94	78	78	81	80	73	71	70
13	Gasang Muhi	cluster_0	92	94	80	80	81	85	81	78	72
14	KA ARMANIS	cluster_0	89	94	78	78	85	82	81	78	73
15	Ilyas Abdullah	cluster_0	88	94	80	80	80	81	75	73	78
16	M. Irfan Anis	cluster_0	87	94	78	78	81	81	76	76	71

Data yang ditampilkan pada Gambar 7 dapat digambarkan juga dengan grafik untuk diketahui perankingan data di setiap mata pelajarannya

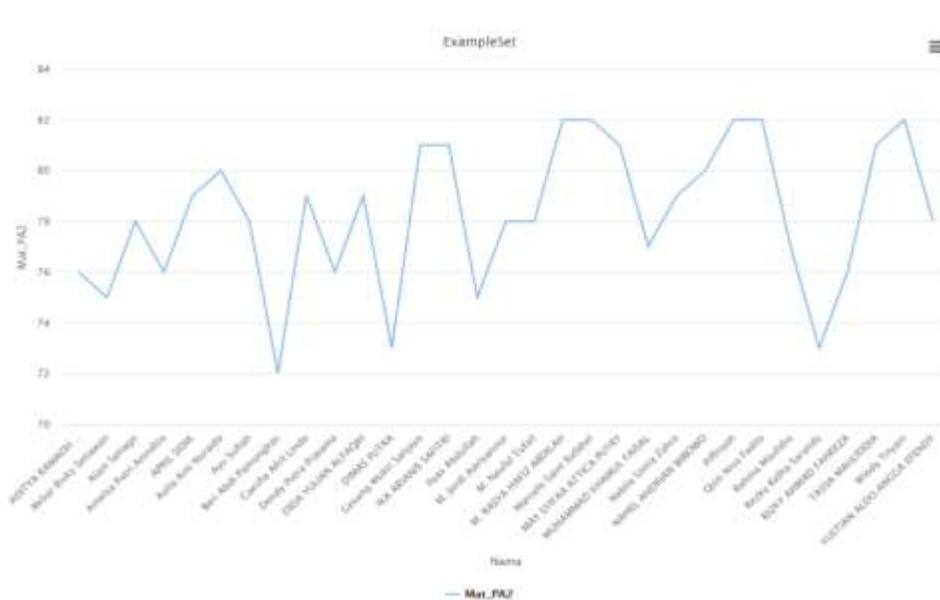
Gambar 7. Hasil Proses transformation data

Hasil

Berikut tampilan grafik perankingan data kelas 7 mata pelajaran matematika, IPA dan IPS :

1. MATEMATIKA

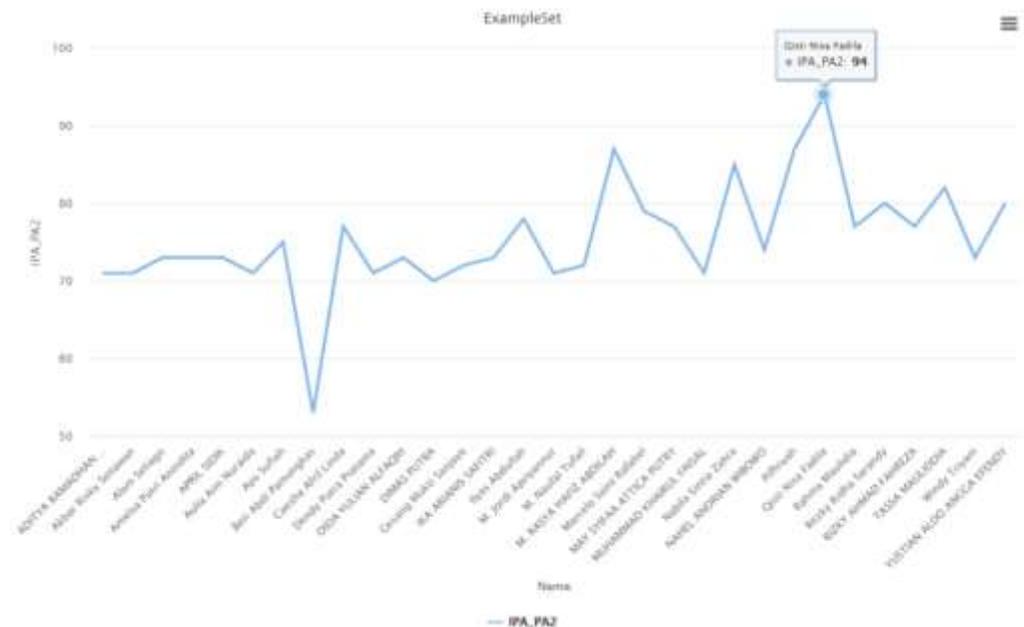
Pada Gambar 8 diketahui secara visual dari grafik bahwa siswa dengan nama M. Rasya Hafiz Abdullah, Marcelo Saint Rahafel, Piffriyah, Qisti Nisa Fadilla dan Windy Triyani menjadi siswa paling unggul pada mata pelajaran Matematika dibandingkan temannya yang lain.



Gambar 8. Grafik Ranking Nilai Mata Pelajaran Matematika Kelas 7

2. IPA

Pada Gambar 9 diketahui secara visual dari grafik bahwa siswi dengan nama Qisti Nisa Fadilla menjadi siswa paling unggul pada mata pelajaran IPA dibandingkan temannya yang lain.



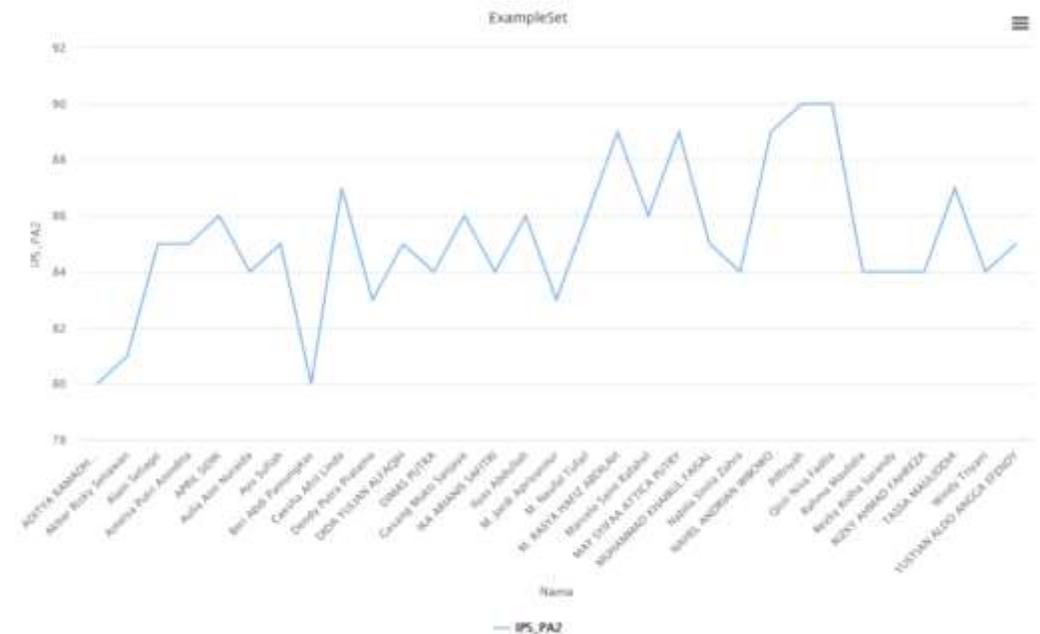
Gambar 9. Grafik Ranking Nilai Mata Pelajaran IPA Kelas 7

Hasil

Berikut tampilan grafik perankingan data kelas 7 mata pelajaran matematika, IPA dan IPS :

3. IPS

Pada Gambar 10 diketahui secara visual dari grafik bahwa siswi dengan nama Qisti Nisa Fadilla dan Piffriyah menjadi siswa paling unggul pada mata pelajaran IPS dibandingkan temannya yang lain.



Gambar 10. Grafik Ranking Nilai Mata Pelajaran IPS Kelas 7

Hasil

Hasil Pengujian *Performance* data kelas 7

Berikut table disamping adalah tabel hasil pengujian performa K-2 hingga K-5 (K = Jumlah *Cluster*). Mengacu ke Tabel 1, menjelaskan bahwa cluster yang cocok untuk di terapkan dalam proses perhitungan sebelumnya yaitu sejumlah 2 kluster dengan perolehan nilai performa sebesar -0,091 karena merupakan nilai terkecil dari nilai performa yang lain.

NILAI K	DAVIES BOULDIN INDEX (DBI)
2	-0.091
3	-0,866
4	-0,681
5	-0,906

Tabel 1. Hasil Pengujian *Performance* data kelas 7

Kesimpulan dan Saran

Berdasarkan hasil penelitian yang telah dilakukan oleh peneliti, dapat ditarik kesimpulan bahwa penerapan Metode K-Means menggunakan *tools* RapidMiner bekerja dengan baik dan mendapatkan nilai *cluster* yang akurat. Dari hasil perhitungan menggunakan *tools* RapidMiner mendapatkan beberapa nama siswa/i kelas 7 yang unggul di mata pelajaran Matematika, IPA dan IPS. Siswa yang Bernama M. Rasya Hafiz Abdillah, Marcelo Saint Rahafel, Piffriyah, Qisti Nisa Fadilla dan Windy Triyani unggul di mata pelajaran Matematika, Siswi yang bernama Qisti Nisa Fadilla unggul di mata pelajaran IPA dan Siswi yang bernama Qisti Nisa Fadilla dan Piffriyah unggul di mata pelajaran IPS. Untuk harapan penelitian selanjutnya, penelitian ini dapat dikembangkan serta direalisasikan dalam bentuk aplikasi/website sehingga mempermudah pengguna untuk menerapkan Metode K-Means selain menggunakan *tools* RapidMiner.

Referensi

- [1] S. D. Laksana, "Pentingnya Pendidikan Karakter Dalam Menghadapi Teknologi Pendidikan Abad 21," *J. Teknol. Pembelajaran*, vol. 1, no. 01, pp. 14–22, 2021.
- [2] F. P. Dewi, P. S. Aryni, and Y. Umaidah, "Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 7, no. 2, pp. 111–121, 2022.
- [3] A. upi Fitriyadi, "Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional," *Kilat*, vol. 10, no. 1, pp. 157–168, 2021.
- [4] Y. P. Sari, A. Primajaya, and A. S. Y. Irawan, "Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang," *INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, p. 229, 2020.
- [5] J. Hutagalung, "Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 1, pp. 606–620, 2022.
- [6] D. D. Darmansah and N. W. Wardani, "Analisis Pesebaran Penularan Virus Corona di Provinsi Jawa Tengah Menggunakan Metode K-Means Clustering," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 8, no. 1, pp. 105–117, 2021.
- [7] H. Haviluddin, S. J. Patandianan, G. M. Putra, N. Puspitasari, and H. S. Pakpahan, "Implementasi Metode K-Means Untuk Pengelompokkan Rekomendasi Tugas Akhir," *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 16, no. 1, p. 13, 2021.
- [8] S. Hajar, A. A. Novany, A. P. Windarto, A. Wanto, and E. Irawan, "Penerapan K-Means Clustering pada ekspor minyak kelapa sawit menurut negara tujuan," *Semin. Nas. Teknol. Komput. Sains 2020*, pp. 314–318, 2020.

Referensi

- [9] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. Tekno Kompak*, vol. 15, no. 2, p. 25, 2021.
- [10] R. Supardi and I. Kanedi, "Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 270–277, 2020.
- [11] R. P. Primanda, A. Alwi, and D. Mustikasari, "DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri)," *Komputek*, vol. 5, no. 1, p. 88, 2021.
- [12] M. I. Zuhendra and R. Hidayat, "Penerapan Data Mining Untuk Klasterisasi Tingkat Kemiskinan Berdasarkan Data Terpadu Kesejahteraan Sosial (DTKS)," vol. 7, pp. 32–42, 2024.
- [13] D. A. Suraya, Muhammad Sholeh, "Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa," *SKANIKA Sist. Komput. dan Tek. Inform.*, vol. 6, no. 6, pp. 51–60, 2023.
- [14] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020.
- [15] I. T. Umagapi, B. Umaternate, H. Hazriani, and ..., "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa," *Pros. ...*, pp. 303–308, 2023.
- [16] F. Yunita, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru," *Sistemasi*, vol. 7, no. 3, p. 238, 2018.

