

ARTIKEL ILMIAH NABELLA.docx

by 1 1

Submission date: 17-Apr-2024 09:08AM (UTC+0300)

Submission ID: 2352502678

File name: ARTIKEL_ILMIAH_NABELLA.docx (1.33M)

Word count: 2929

Character count: 18055

Implementasi Rapidminer Untuk Menentukan Siswa Unggulan Menggunakan Metode K-Means

Nabella Rosyefa Wahyudi¹, Yunianita Rahmawati², Suprianto³

^{1,2,3}Fakultas Sains dan Teknologi, Teknik Informatika, Universitas Muhammadiyah, Sidoarjo, Indonesia

E-mail: ¹nabellaa2407@gmail.com, ²yunianita@umsida.ac.id, ³suprianto@umsida.ac.id

(* : corresponding author)

Abstrak

Penelitian ini bertujuan untuk menerapkan Metode K-Means dan RapidMiner dalam mengidentifikasi siswa unggulan berdasarkan nilai mata pelajaran Matematika, IPA, dan IPS di SMP Negeri 11 Sampit. Dengan dataset yang terdiri dari 24 atribut dan 31 data siswa, Metode K-Means berhasil mengelompokkan siswa berdasarkan kemiripan nilai yang siswa tersebut peroleh. Hasil evaluasi model menunjukkan bahwa nilai cluster terbaik berjumlah 2 cluster. Penggunaan Metode K-Means dalam penelitian ini merupakan solusi untuk kesulitan dalam menentukan siswa unggulan di SMP Negeri 11 Sampit karena kemampuan siswa cenderung seimbang di setiap semester. Penelitian ini memberikan wawasan bahwa Metode K-Means efektif dalam menentukan siswa unggulan dan dapat menjadi alat yang berguna dalam evaluasi Pendidikan.

Kata kunci: RapidMiner, K-Means, Siswa, Akademik.

Abstract

This research aims to apply the K-Means Method and RapidMiner in identifying superior students based on the scores of Mathematics, Science, and Social Studies subjects at SMP Negeri 11 Sampit. With a dataset consisting of 24 attributes and 31 student data, the K-Means Method successfully clusters students based on the similarity of the grades the students obtained. The model evaluation results show that the best cluster value is 2 clusters. The use of the K-Means Method in this study is a solution to the difficulty in determining superior students at SMP Negeri 11 Sampit because students' abilities tend to be balanced in each semester. This research provides insight that the K-Means Method is effective in determining superior students and can be a useful tool in educational evaluation.

Keywords: RapidMiner, K-Means, Student, Academic.

1. PENDAHULUAN

Di era globalisasi saat ini, cara belajar siswa sangat bergantung pada kehadiran media teknologi informasi. Hal ini mencerminkan karakteristik siswa abad ke-21 yang memegang kendali, lebih menyukai beragam pilihan pembelajaran, dan menikmati kolaborasi teknis. [1] Pentingnya media yang dihasilkan teknologi dapat menjadi sumber peningkatan standar mutu pendidikan dan sekolah. Pendidikan memegang peranan penting dalam kehidupan manusia, menciptakan kondisi bagi individu untuk mengekspresikan bakat, inovasi sebagai salah satu faktor penunjang mutu pendidikan di Indonesia. Setiap sekolah berupaya untuk meningkatkan mutu sekolahnya melalui pengelompokan siswa unggulan yang didasari kemampuan siswa dan kriteria sekolah.

Pengelompokan siswa unggulan, secara umum ditentukan melalui nilai siswa pada setiap mata pelajaran di kelas. Objek penelitian ini menggunakan data nilai siswa kelas 7 SMP Negeri 11 Sampit. Berdasarkan hasil observasi yang dilakukan bahwa SMP Negeri 11 Sampit mengalami kesulitan untuk menentukan perwakilan siswa lomba seperti lomba OSN (Olimpiade Sains Nasional) karena kemampuan siswa SMP Negeri 11 Sampit di setiap generasi cenderung seimbang. Selain itu penempatan siswa unggulan dapat berubah menyesuaikan nilai yang diperoleh siswa setiap semester. Salah satu solusi yang dapat dilakukan pada SMP Negeri 11 Sampit yaitu dengan menerapkan metode data mining *clustering*. Salah satu jenis teknik dalam data mining *clustering* adalah Metode K-Means. Metode ini berfungsi untuk membagi data ke dalam *cluster-cluster* sehingga yang serupa akan dikelompokkan kedalam *cluster* yang sama,

sementara data yang berbeda akan dikelompokkan dalam *cluster* yang serupa dengan data tersebut.[2] 22

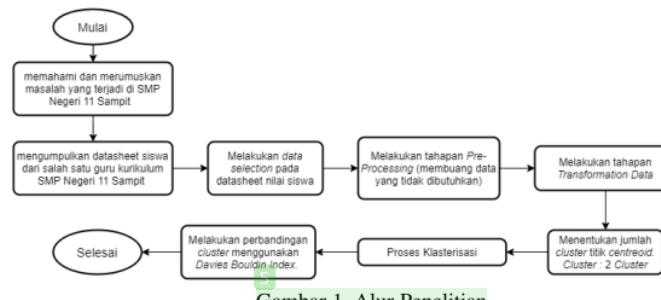
Penelitian ini akan melakukan pengelompokan data siswa dengan menerapkan Metode K-Means yang memiliki tingkat keakuratan tinggi, hal ini dibuktikan dalam paper [3] yang berjudul ‘Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional’. Selain itu, penerapan Algoritma K-Means lebih mudah dan sederhana serta sering digunakan pada proses data mining. [4],[5] Algoritma K-Means merupakan salah satu jenis algoritma *clustering* dalam mengelompokkan data dengan menentukan titik pusat *cluster* terdekat [6]. Penerapan metode ini termasuk metode berulang yang sederhana karena metode ini cenderung *flexible* dan tidak memakan waktu lama. [7] Algoritma K-Means juga dipahami meminimalkan banyak kesalahan. [8]

Penelitian terdahulu juga telah menjelaskan tentang Algoritma K-Means seperti paper [9] yang berjudul ‘Implementasi Algoritma K-Means Clustering dalam Penentuan Siswa Kelas Unggulan’, penelitian ini bertujuan untuk mengelompokkan siswa pada kelas yang lebih tinggi dengan mengimplementasikan Metode K-Means *Clustering*. Kemudian, penelitian yang telah dijelaskan oleh Reno Supardi dalam paper [10], penelitian ini bertujuan untuk membantu pelaporan penjualan di Toko Edelwis. Selanjutnya, penelitian yang telah dijelaskan oleh Putra Primanda dalam paper [11], penelitian ini bertujuan untuk membantu mengklastering kelas unggulan dengan menerapkan Metode K-Means. Selanjutnya paper [12], penelitian ini bertujuan memudahkan proses klasterisasi untuk memantau perkembangan kemakmuran masyarakat di Daerah Aceh. Serta penelitian yang telah di jelaskan oleh Suraya dalam paper [13], penelitian ini bertujuan untuk membantu mengklastering IPK (Indeks Prestasi Mahasiswa) dengan menerapkan Algoritma K-Means. Penelitian sebelumnya yang bertujuan untuk mengelompokkan siswa yang unggul dengan mengimplementasikan algoritma K-Means telah dilakukan di berbagai lembaga pendidikan, namun penelitian saya hanya terfokus pada SMP Negeri 11 Sampit, dimana pengimplementasian Algoritma K-Means menggunakan RapidMiner belum dilakukan sebelumnya disana. Penelitian ini mengisi kesenjangan dalam literatur dengan memberikan lebih banyak informasi tentang penentuan siswa yang unggul di suatu institusi tertentu.

Berdasarkan pemaparan di atas, kajian ini bertujuan yaitu untuk mempermudah SMP 11 SAMPIT dalam mengelompokkan siswa unggulan sebagai bahan evaluasi meningkatkan mutu sekolah serta membuktikan tingkat keakuratan performa *cluster Davies Bouldin Index* (DBI). Penelitian ini menggunakan *tools* RapidMiner untuk mempermudah perhitungan pada data mining. Untuk mencapai tujuan tersebut maka dibuatlah solusi yaitu Implementasi Rapidminner Untuk Menentukan Siswa Unggulan Menggunakan Metode K-Means.

2. METODE PENELITIAN

Peneliti melakukan serangkaian proses dimana setiap proses berpotensi membantu penelitian ini menjadi lebih mudah. Berikut adalah diagram alur dari tahapan proses tersebut :



Gambar 1. Alur Penelitian

Gambar 1, menjelaskan alur dari penelitian yang peneliti susun. Gambar 1 menampilkan langkah mengidentifikasi, merumuskan dan memahami permasalahan yang akan diteliti, salah satunya

yaitu dengan melakukan studi literatur, lalu selanjutnya peneliti mengumpulkan data berupa data nilai siswa kelas 7. Data yang terkumpul sebanyak 31 siswa. Kemudian, peneliti melakukan analisis hasil dengan menerapkan Metode K-Means pada setiap data hingga menemukan hasil akhir.

5

2.1. Metode Pengambilan data

Peneliti menggunakan metode pengumpulan data kuantitatif yang meliputi,
a. Dokumentatif
Peneliti menggunakan metode ini untuk mengumpulkan data nilai keseluruhan siswa kelas VII yang akan diteliti.

2.2. Data Mining

Data mining merupakan proses yang menerapkan matematika, berbagai teknik statistik, kecerdasan buatan yang bertujuan untuk mengidentifikasi dan mengekstraksi data/ informasi berharga dari sekumpulan data yang lebih besar [14]. Data mining memiliki beberapa teknik seperti *clustering*, analisis regresi, *text mining* dan lain lain.

2.3. Davies Bouldin Indeks (DBI)

Evaluasi DBI (*Davies Bouldin Index*) perlu dilakukan untuk mencari *cluster* terbaik dari metode *clustering*. Nilai *performance* proses *clustering* akan semakin baik apabila nilai tersebut semakin kecil [15]. DBI berperan cukup penting dalam penelitian ini karena dengan DBI, peneliti dapat mengetahui kevalidan tingkat performa K-Means dalam menentukan K (Kluster).

10

2.4. Metode K-Means

Metode K-Means merupakan metode *clustering* yang mengelompokkan data yang mirip satu sama lain kedalam cluster yang sama. Data yang memiliki kemiripan atau sama akan dikelompokkan kedalam satu cluster dan yang berbeda akan di kelompokkan ke *cluster* yang lain. [16]. Berikut adalah gambaran umum proses *clustering* Algoritma K-Means :

1. Menentukan jumlah K (*cluster*).
2. Memilih titik pusat cluster (*centreoid*) secara acak. Titik *centreoid* ditentukan acak sesuai dengan jumlah cluster.
3. Menghitung jarak setiap *cluster* dengan menggunakan rumus *Euclidean Distance*. Rumus ini menghitung selisih data dengan titik *centreoid* lalu menjumlahkan tiap data tersebut. Berikut adalah rumus *Euclidean Distance* :

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Gambar 2. Rumus *Euclidean Distance*

Keterangan Gambar 2 :

- $\sum_{i=1}^n, i=1$ bermakna bahwa perhitungan penjumlahan dimulai dari atribut pertama hingga n, n merupakan jumlah total dari atribut data yang akan dihitung.
- $(x_i - y_i)^2, x_i$ merupakan data awal dari atribut pertama dan nilai dari y_i merupakan hasil rata rata dari semua data *cluster* pada iterasi ke-2 dan seterusnya.

4. Memperbarui titik pusat *cluster* (*centreoid*).

Berikut adalah rumus mencari titik pusat *cluster* :

$$C_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i$$

Gambar 3. Rumus memperbarui *centreoid*

Keterangan Gambar 3 :

- n_j , merepresentasikan jumlah data dari j (*cluster*) yang sudah di alokasikan sebelumnya.
 - $\sum_{i=1}^{n_j} x_i$, $i=1$ bermakna bahwa perhitungan penjumlahan dimulai dari atribut pertama hingga n_j , n_j merupakan jumlah total dari atribut data *cluster* yang akan dihitung sesuai dengan posisi atribut ke i .
14. Ulangi langkah-langkah 2 dan 3 hingga tidak ada perubahan yang signifikan dalam proses pengelompokan data.

17

2.5. Datasheet (Data Nilai Siswa)

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	85	78	78	78	79	78	71	71	80	80
2	Akbar Risky B.	cluster_0	87	85	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	Ametsa Putri	cluster_0	92	94	79	79	84	82	76	78	73	85	85
5	APRIL SIEK	cluster_0	88	94	78	78	80	81	79	78	73	85	85
6	Aulta Aini Nur.	cluster_0	90	94	78	78	82	80	80	77	71	85	84
7	Avi Sufah	cluster_0	89	94	78	78	85	80	78	83	75	85	85
8	Beri Abd Pa...	cluster_0	86	85	78	78	78	79	72	71	53	80	80
9	Caesha Ahi...	cluster_0	92	94	80	80	85	85	79	78	77	85	87
10	Dendi Putra	cluster_0	89	94	78	78	81	80	76	73	71	85	83
11	DIDA YULIANI	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 Muhi...	cluster_0	92	94	80	80	81	85	81	78	72	85	85
14	IKKA ARIANES	cluster_0	89	94	78	78	85	82	81	78	73	85	84
15	Iryas Abdulah	cluster_0	88	94	80	80	80	81	75	73	78	85	85
16	M. Jona Apri...	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	85
18	M. RARYA HA...	cluster_0	92	94	80	80	87	81	82	78	87	85	88

Gambar 4. Datasheet (Data Nilai Siswa)

Gambar 4 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.

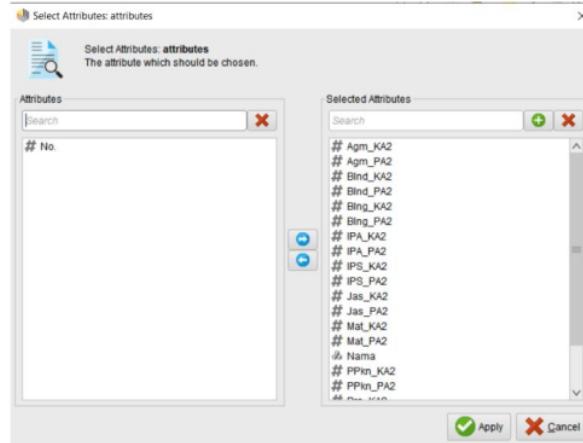
3. HASIL DAN PEMBAHASAN

Proses perhitungan penerapan Metode K-Means untuk menentukan siswa unggulan ini dilakukan menggunakan *tools* RapidMiner. Proses perhitungan ini hanya berfokus pada mata pelajaran yang diujikan OSN (Olimpiade Sains Nasional) saja seperti Matematika, IPA dan IPS. Berikut hasil dan pembahasan yang sudah dilakukan oleh peneliti dengan menggunakan tools yang sudah dijelaskan diatas.

3.1 Data Selection

Setelah semua data terkumpul, tahapan yang perlu dilakukan selanjutnya yaitu menyeleksi *attribute* yang akan di gunakan dalam pengimplementasian Algoritma K-Means. Tahapan seleksi data dalam rapidminer dapat menggunakan operator *select attribute*. Fitur ini berfungsi untuk

memilah keseluruhan atribut data yang dibutuhkan oleh operator perhitungan. Berikut dibawah ini merupakan proses tahapan seleksi atribut dari datasheet nilai siswa kelas 7 SMP 11 Sampit.



Gambar 5. Proses seleksi atribut data kelas 7

Gambar 5 menampilkan proses seleksi data kelas 7 menggunakan operator *select attribute* pada RapidMiner. Data tersebut memiliki 21 *attribute* yaitu 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 Atribut nama diubah menjadi type data id untuk mempermudah proses pengelolahan data. Setelah diubah, atribut tersebut akan terhitung sebagai label objek yang akan di teliti.

3.2 Pre-Processing

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. Berikut dibawah ini merupakan proses tahapan *Pre Processing* dari datasheet nilai siswa kelas 7 SMP 11 Sampit.

Result History							Repository	
Exemplar Set (Choosing)		Cluster Model (Content)					Import Data	
Name	Type	Missing	Statistics	Filter (22 (22 attributes))			Search for attributes	
Name	Parameter	0	Last	YUSTIWI FENDY (1)	ADHYA R. PUTRA (1)	ACTHA RAMADHAN SAPUTRA (1), APRIL SEDIH (1), DIAH	Filter	All Studio
cluster	nominal	3	Last	cluster_1 (1)	cluster_0 (30)	cluster_0 (30), cluster_1 (1)	cluster_0 (30), cluster_1 (1)	cluster_0 (30), cluster_1 (1)
Agn_PA2	integer	9	Min	54	Average	57.032	Min	Average
Agn_KA2	integer	9	Min	54	Average	56.194	Min	Average
PPW_PA2	integer	9	Min	77	Average	79.419	Min	Average
PPW_KA2	integer	9	Min	77	Average	79.419	Min	Average
Btd_PA2	integer	9	Min	78	Average	82.258	Min	Average
Btd_KA2	integer	9	Min	79	Average	81.971	Min	Average
Mat_PA2	integer	9	Min	72	Average	78.258	Min	Average

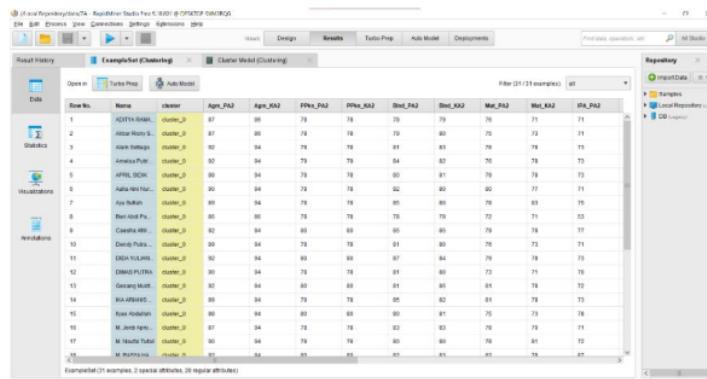
Gambar 6. Proses *Pre-Processing* data kelas 7

Gambar 6 menampilkan stastistika hasil seleksi data kelas 7 setelah dilakukan *running* di *software* Rapid Minner. Gambar 6 diatas menjelaskan bahwa data kelas 7 memiliki 2 atribut spesial dan

31 example dimana atribut Nama masuk kedalam salah satu atribut spesial karena memiliki tipe *polynominal* yang tidak bisa diolah dengan perhitungan Algoritma *K-Means*.

3.3 Transformation Data

Setelah tahapan *pre-processing* selesai, langkah berikutnya adalah melakukan *transformation* data pada hasil dari tahapan sebelumnya. *Transformation* data merupakan proses mengubah / memanipulasi data mentah menjadi lebih sesuai untuk dianalisis. Berikut dibawah ini merupakan proses tahapan *transformation data* dari datasheet nilai siswa kelas 7 SMP 11 Sampit.



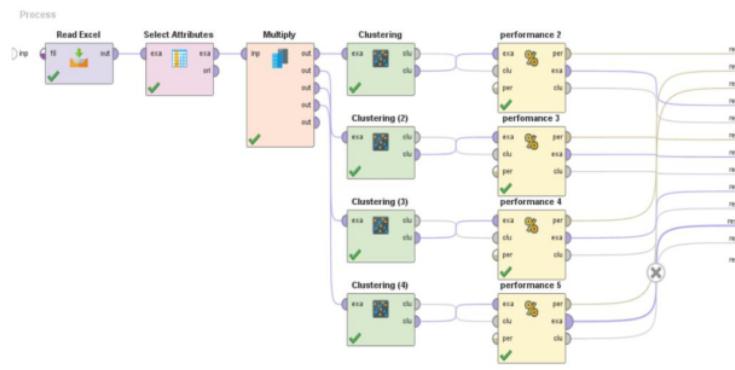
Row No.	Name	cluster	Agm_PA2	Agm_PA2	Prth_PA2	Prth_PA2	Smr_PA2	Smr_PA2	Mat_PA2	Mat_PA2	IPA_PA2
1	ADITYA BRIAN	cluster_0	87	86	78	78	70	70	71	71	71
2	Astar Wury S.	cluster_0	87	86	78	78	70	80	75	73	71
3	Ilman Ridhoq	cluster_0	92	94	78	78	81	83	79	78	73
4	Amelia Putri	cluster_0	92	94	78	78	84	87	76	78	73
5	AFRIK INDRA	cluster_0	88	94	78	78	80	81	79	78	73
6	Adila Rini Huz...	cluster_0	80	94	78	78	82	80	80	77	71
7	Ayu Sulisti	cluster_0	89	94	78	78	85	80	79	83	76
8	Beri Andri Pa...	cluster_0	86	86	78	78	78	79	72	71	53
9	Cendita ARI	cluster_0	82	94	68	68	95	85	79	78	77
10	Denyli Putri...	cluster_0	89	94	78	78	91	80	76	72	71
11	DEWI YULIAH	cluster_0	92	94	68	68	97	84	79	78	73
12	DEWI YULIAH	cluster_0	90	94	78	78	91	80	73	71	78
13	Gerryati Mu...	cluster_0	82	94	68	68	91	81	81	78	72
14	HA. ABRAHAM	cluster_0	89	94	78	78	95	82	81	78	73
15	Irene Andriah...	cluster_0	88	94	68	68	95	81	76	73	78
16	M. Jend. Hery...	cluster_0	87	94	78	78	92	82	70	73	71
17	M. Nizahulul	cluster_0	90	94	78	78	90	80	76	81	72
18	M. Nurul Huda	cluster_0	92	94	68	68	93	81	82	78	87
19	M. Nurul Huda	cluster_0	92	94	68	68	93	81	82	78	87
20	M. Nurul Huda	cluster_0	92	94	68	68	93	81	82	78	87
21	M. Nurul Huda	cluster_0	92	94	68	68	93	81	82	78	87

Gambar 7. Proses *Transformation Data* data kelas 7

Gambar 7 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.

3.4 Data Mining

Setelah melalui tahapan *Transformation data*. Langkah berikutnya yaitu melakukan tahapan *data mining* pada hasil dari tahapan sebelumnya. Berikut dibawah ini merupakan proses tahapan *Data Mining* dari datasheet nilai siswa kelas 7 SMP 11 Sampit :



Gambar 8. Proses *Clustering* pada data kelas 7

Gambar 8 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*.

3.5 Analisis Cluster

Berikut adalah hasil *cluster* dari proses penerapan Metode K-Means untuk menentukan siswa unggulan menggunakan Rapidminner dari *datasheet* nilai siswa kelas 7 yang sudah dilampirkan diatas. Berikut adalah hasil pemodelan jumlah *cluster* yang terbentuk dari perhitungan Metode K-Means :

Cluster Model

```
Cluster 0: 30 items
Cluster 1: 1 items
Total number of items: 31
```

Gambar 9. Hasil *Cluster* Kelas 7

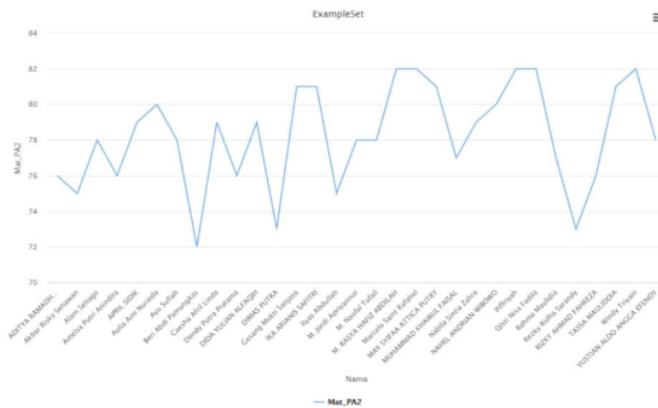
Gambar 9 menampilkan hasil dari jumlah *cluster* yang terbentuk di kelas 7 dimana *cluster* 0 mendapatkan 30 data dan *cluster* 1 mendapatkan 1 data. *Cluster* 0 merupakan representasi dari siswa unggul dan cluster 1 merupakan representasi dari siswa bukan unggulan dari seluruh mata pelajaran. Berikut di bawah ini Gambar 10 yang merupakan tampilan hasil dari keseluruhan data yang sudah melalui tahapan *Clustering* menggunakan Metode *K-Means* :

Row No.	Nama	cluster	Agm_PA2	Agm_KA2	PPkn_PA2	PPkn_KA2	Bind_PA2	Bind_KA2	Mat_PA2	Mat_KA2	IPA_PA2
1	ADITYA RAMA...	cluster_0	87	86	78	78	78	79	76	71	71
2	Albar Risky S...	cluster_0	87	86	78	78	79	80	75	73	71
3	Alam Setiago	cluster_0	92	94	78	78	81	83	78	78	73
4	Amelisa 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	Aulia Aini Nur...	cluster_0	90	94	78	78	82	80	80	77	71
7	Ayu Sulfah	cluster_0	89	94	78	78	85	80	78	83	75
8	Beri Abdi Pa...	cluster_0	86	86	78	78	78	79	72	71	53
9	Caesha Atil...	cluster_0	92	94	80	80	85	85	79	78	77
10	Dendy 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	90	94	78	78	81	80	73	71	70
13	Gesang Muidi...	cluster_0	92	94	80	80	81	85	81	78	72
14	IKA ARIANIS...	cluster_0	89	94	78	78	85	82	81	78	73
15	Iyas Abdullah	cluster_0	88	94	80	80	80	81	75	73	78
16	M. Inerti Antri...	cluster_0	87	94	78	78	83	83	78	78	71

Gambar 10. Data hasil *Cluster* Kelas 7

Data yang ditampilkan pada Gambar 10 dapat digambarkan juga dengan grafik untuk diketahui perankingan data di setiap mata pelajarannya. Berikut tampilan grafik perankingan data kelas 7 dapat dilihat pada gambar 10 dibawah ini :

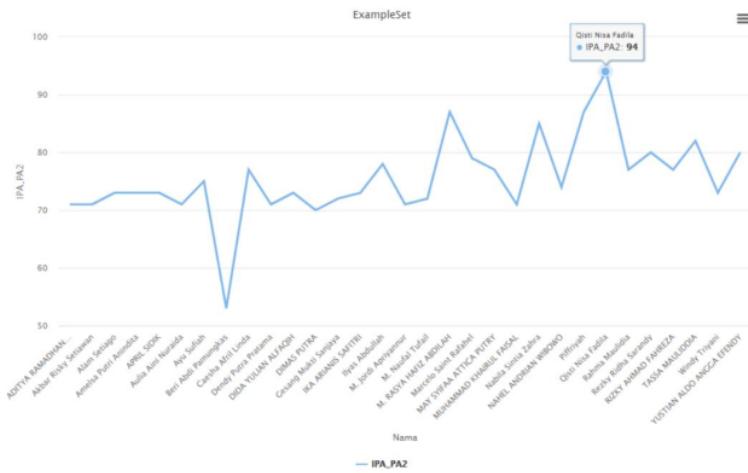
3.5.1 Mata Pelajaran Matematika



Gambar 11. Grafik Ranking Nilai Mata Pelajaran Matematika Kelas 7

Pada Gambar 11 diketahui secara visual dari grafik bahwa siswa dengan nama M. Rasya Hafiz Abdillah, Marcelo Saint Rahafel, Qisti Nisa Fadilla dan Windy Triyani menjadi siswa paling unggul pada mata pelajaran Matematika dibandingkan temannya yang lain.

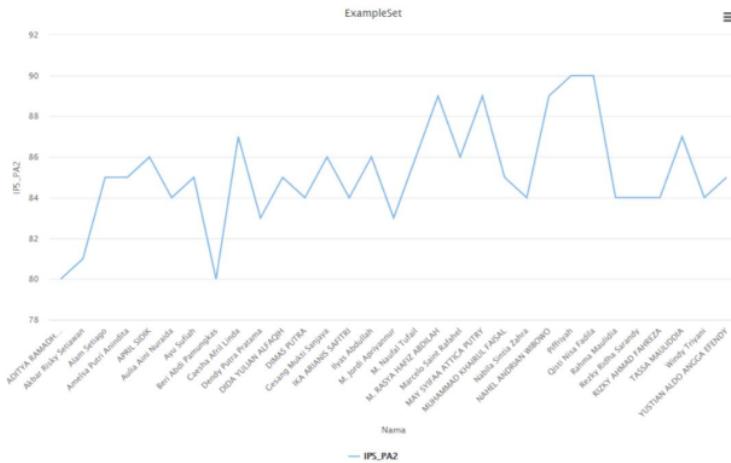
3.5.2 Mata Pelajaran IPA



Gambar 12. Grafik Ranking Nilai Mata Pelajaran IPA Kelas 7

Pada Gambar 12 diketahui secara visual dari grafik bahwa siswi dengan nama Qisti Nisa Fadilla menjadi siswa paling unggul pada mata pelajaran IPS dibandingkan temannya yang lain.

3.5.3 Mata Pelajaran IPS

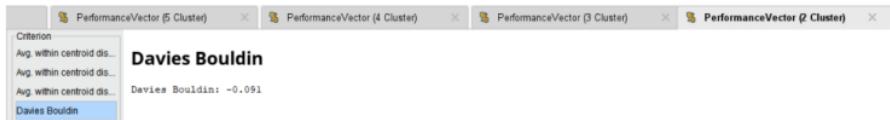


Gambar 13. Grafik Ranking Nilai Mata Pelajaran IPS Kelas 7

Pada Gambar 13 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.

3.6 Evaluasi Model

Setelah proses seleksi data nilai selesai, langkah berikutnya adalah melakukan evaluasi model pada hasil dari tahapan sebelumnya. Setiap pengujian menggunakan algoritma K-Means menghasilkan cluster yang berbeda-beda tergantung pada nilai K yang digunakan. Dalam empat percobaan yang dilakukan, nilai K yang berbeda menghasilkan hasil yang berbeda pula. Berikut dibawah ini hasil nilai pengujian *performance* data kelas 7 yang paling baik yang ditampilkan pada Gambar 14 :



Gambar 14. Hasil Nilai pengujian *Performance* data Kelas 7

Berikut di bawah ini adalah tabel hasil pengujian performa K-2 hingga K-5 (K = Jumlah Cluster):

Tabel 1. Hasil Pengujian *Performance* data kelas 7

Nilai K	Davies Bouldin Index (DBI)
2	-0.091
3	-0.866
4	-0.681
5	-0.906

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.

9
4. 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.

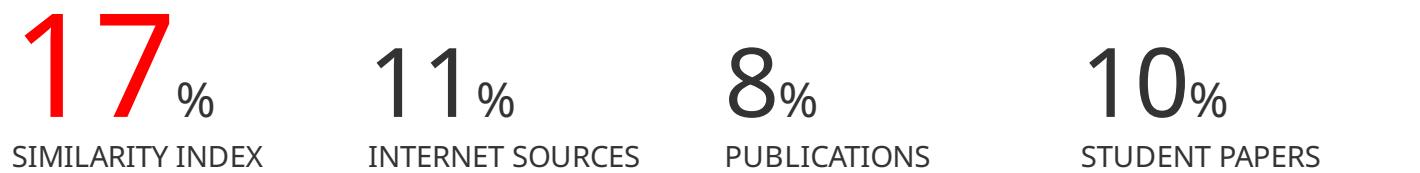
DAFTAR PUSTAKA

- [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 Pengelompokan 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 eksport minyak kelapa sawit menurut negara tujuan,” *Semin. Nas. Teknol. Komput. Sains 2020*, pp. 314–318, 2020.
- [9] A. Sulistiawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penetuan 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. Umatermate, 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.

ARTIKEL ILMIAH NABELLA.docx

ORIGINALITY REPORT



PRIMARY SOURCES

1	Submitted to Universitas Budi Luhur Student Paper	6%
2	jurnal.kopertipindonesia.or.id Internet Source	1%
3	core.ac.uk Internet Source	1%
4	Submitted to Universitas Muhammadiyah Sidoarjo Student Paper	1%
5	csrid.potensi-utama.ac.id Internet Source	1%
6	repository.teknokrat.ac.id Internet Source	1%
7	www.coursehero.com Internet Source	1%
8	Submitted to Universitas Jember Student Paper	<1%
9	repositori.uma.ac.id Internet Source	<1%

10	Submitted to UIN Syarif Hidayatullah Jakarta Student Paper	<1 %
11	jurnal.stmikroyal.ac.id Internet Source	<1 %
12	Submitted to Yonkers High School Student Paper	<1 %
13	Listiyani Wahyuningsih, Ahmad Rifai. "IMPLEMENTASI ALGORITMA K-MEANS TERHADAP PENGELOMPOKAN KEPEMILIKAN AKTA KELAHIRAN DI JAWA BARAT", E-Link: Jurnal Teknik Elektro dan Informatika, 2023 Publication	<1 %
14	ejurnal.ung.ac.id Internet Source	<1 %
15	jurnal.atmaluhur.ac.id Internet Source	<1 %
16	Yogi Yunefri, Eddisyah Putra Pane, Sutejo Sutejo. "Pengembangan Sistem Pengelompokan Belajar Mahasiswa pada Matakuliah Struktur Data dengan Metode K- Means", INTECOMS: Journal of Information Technology and Computer Science, 2019 Publication	<1 %
17	Submitted to STT PLN Student Paper	<1 %
	ejournal.itn.ac.id	

18	Internet Source	<1 %
19	ejurnal.stmik-budidarma.ac.id Internet Source	<1 %
20	jurnal.unimed.ac.id Internet Source	<1 %
21	ojs.trigunadharma.ac.id Internet Source	<1 %
22	sistemasi.ftik.unisi.ac.id Internet Source	<1 %
23	www.scribd.com Internet Source	<1 %
24	Yopi Hendro Syahputra, Juniar Hutagalung. "Superior Class to Improve Student Achievement Using the K-Means Algorithm", SinkrOn, 2022 Publication	<1 %
25	Hairani Hairani, Muhammad Ridho Hansyah, Lalu Zazuli Azhar Mardedi. "Integrasi Metode Naive Bayes dengan K-Means dan K-Means- Smote untuk Klasifikasi Jurusan SMAN 3 Mataram", Jurnal Sistem dan Informatika (JSI), 2020 Publication	<1 %
26	jurnal.polibatam.ac.id Internet Source	<1 %

Exclude quotes On

Exclude bibliography On

Exclude matches Off

ARTIKEL ILMIAH NABELLA.docx

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

Implementasi Rapidminer Untuk Menentukan Siswa Unggulan Menggunakan Metode K-Means

Nabella Rosyefa Wahyudi¹, Yunianita Rahmawati², Suprianto³

^{1,2,3}Fakultas Sains dan Teknologi, Tekhnik Informatika, Universitas Muhammadiyah, Sidoarjo, Indonesia

E-mail: ¹nabellaa2407@gmail.com, ²yunianita@umsida.ac.id, ³suprianto@umsida.ac.id

(* : corresponding author)

Abstrak

Penelitian ini bertujuan untuk menerapkan Metode K-Means dan RapidMiner dalam mengidentifikasi siswa unggulan berdasarkan nilai mata pelajaran Matematika, IPA, dan IPS di SMP Negeri 11 Sampit. Dengan dataset yang terdiri dari 24 atribut dan 31 data siswa, Metode K-Means berhasil mengelompokkan siswa berdasarkan kemiripan nilai yang siswa tersebut peroleh. Hasil evaluasi model menunjukkan bahwa nilai cluster terbaik berjumlah 2 cluster. Penggunaan Metode K-Means dalam penelitian ini merupakan solusi untuk kesulitan dalam menentukan siswa unggulan di SMP Negeri 11 Sampit karena kemampuan siswa cenderung seimbang di setiap semester. Penelitian ini memberikan wawasan bahwa Metode K-Means efektif dalam menentukan siswa unggulan dan dapat menjadi alat yang berguna dalam evaluasi Pendidikan.

Kata kunci: RapidMiner, K-Means, Siswa, Akademik.

Abstract

This research aims to apply the K-Means Method and RapidMiner in identifying superior students based on the scores of Mathematics, Science, and Social Studies subjects at SMP Negeri 11 Sampit. With a dataset consisting of 24 attributes and 31 student data, the K-Means Method successfully clusters students based on the similarity of the grades the students obtained. The model evaluation results show that the best cluster value is 2 clusters. The use of the K-Means Method in this study is a solution to the difficulty in determining superior students at SMP Negeri 11 Sampit because students' abilities tend to be balanced in each semester. This research provides insight that the K-Means Method is effective in determining superior students and can be a useful tool in educational evaluation.

Keywords: RapidMiner, K-Means, Student, Academic.

1. PENDAHULUAN

Di era globalisasi saat ini, cara belajar siswa sangat bergantung pada kehadiran media teknologi informasi. Hal ini mencerminkan karakteristik siswa abad ke-21 yang memegang kendali, lebih menyukai beragam pilihan pembelajaran, dan menikmati kolaborasi teknis. [1] Pentingnya media yang dihasilkan teknologi dapat menjadi sumber peningkatan standar mutu pendidikan dan sekolah. Pendidikan memegang peranan penting dalam kehidupan manusia, menciptakan kondisi bagi individu untuk mengekspresikan bakat, inovasi sebagai salah satu faktor penunjang mutu pendidikan di Indonesia. Setiap sekolah berupaya untuk meningkatkan mutu sekolahnya melalui pengelompokan siswa unggulan yang didasari kemampuan siswa dan kriteria sekolah.

Pengelompokan siswa unggulan, secara umum ditentukan melalui nilai siswa pada setiap mata pelajaran di kelas. Objek penelitian ini menggunakan data nilai siswa kelas 7 SMP Negeri 11 Sampit. Berdasarkan hasil observasi yang dilakukan bahwa SMP Negeri 11 Sampit mengalami kesulitan untuk menentukan perwakilan siswa lomba seperti lomba OSN (Olimpiade Sains Nasional) karena kemampuan siswa SMP Negeri 11 Sampit di setiap generasi cenderung seimbang. Selain itu penempatan siswa unggulan dapat berubah menyesuaikan nilai yang diperoleh siswa setiap semester. Salah satu solusi yang dapat dilakukan pada SMP Negeri 11 Sampit yaitu dengan menerapkan metode data mining *clustering*. Salah satu jenis teknik dalam data mining *clustering* adalah Metode K-Means. Metode ini berfungsi untuk membagi data ke dalam *cluster-cluster* sehingga yang serupa akan dikelompokkan kedalam *cluster* yang sama,

sementara data yang berbeda akan dikelompokkan dalam *cluster* yang serupa dengan data tersebut.[2]

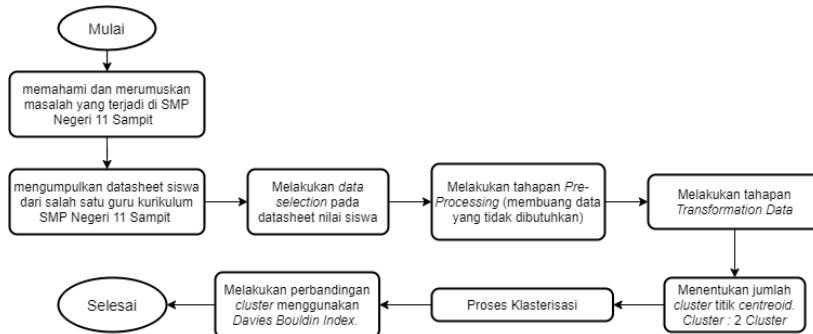
Penelitian ini akan melakukan pengelompokan data siswa dengan menerapkan Metode K-Means yang memiliki tingkat keakuratan tinggi, hal ini dibuktikan dalam paper [3] yang berjudul ‘Analisis Algoritma K-Means dan K-Medoids Untuk Clustering Data Kinerja Karyawan Pada Perusahaan Perumahan Nasional’. Selain itu, penerapan Algoritma K-Means lebih mudah dan sederhana serta sering digunakan pada proses data mining. [4],[5] Algoritma K-Means merupakan salah satu jenis algoritma *clustering* dalam mengelompokkan data dengan menentukan titik pusat *cluster* terdekat [6]. Penerapan metode ini termasuk metode berulang yang sederhana karena metode ini cenderung *flexible* dan tidak memakan waktu lama. [7] Algoritma K-Means juga dipahami meminimalkan banyak kesalahan. [8]

Penelitian terdahulu juga telah menjelaskan tentang Algoritma K-Means seperti paper [9] yang berjudul ‘Implementasi Algoritma K-Means *Clustering* dalam Penentuan Siswa Kelas Unggulan’, penelitian ini bertujuan untuk mengelompokkan siswa pada kelas yang lebih tinggi dengan mengimplementasikan Metode K-Means *Clustering*. Kemudian, penelitian yang telah dijelaskan oleh Reno Supardi dalam paper [10], penelitian ini bertujuan untuk membantu pelaporan penjualan di Toko Edelwis. Selanjutnya, penelitian yang telah dijelaskan oleh Putra Primanda dalam paper [11], penelitian ini bertujuan untuk membantu mengklastering kelas unggulan dengan menerapkan Metode K-Means. Selanjutnya paper [12], penelitian ini bertujuan memudahkan proses klasterisasi untuk memantau perkembangan kemakmuran masyarakat di Daerah Aceh. Serta penelitian yang telah dijelaskan oleh Suraya dalam paper [13], penelitian ini bertujuan untuk membantu mengklastering IPK (Indeks Prestasi Mahasiswa) dengan menerapkan Algoritma K-Means. Penelitian sebelumnya yang bertujuan untuk mengelompokkan siswa yang unggul dengan mengimplementasikan algoritma K-Means telah dilakukan di berbagai lembaga pendidikan, namun penelitian saya hanya terfokus pada SMP Negeri 11 Sampit, dimana pengimplementasian Algoritma K-Means menggunakan RapidMiner belum dilakukan sebelumnya disana. Penelitian ini mengisi kesenjangan dalam literatur dengan memberikan lebih banyak informasi tentang penentuan siswa yang unggul di suatu institusi tertentu.

Berdasarkan pemaparan di atas, kajian ini bertujuan yaitu untuk mempermudah SMP 11 SAMPIT dalam mengelompokkan siswa unggulan sebagai bahan evaluasi meningkatkan mutu sekolah serta membuktikan tingkat keakuratan performa *cluster Davies Bouldin Index* (DBI). Penelitian ini menggunakan *tools RapidMiner* untuk mempermudah perhitungan pada data mining. Untuk mencapai tujuan tersebut maka dibuatlah solusi yaitu Implementasi Rapidminner Untuk Menentukan Siswa Unggulan Menggunakan Metode K-Means.

2. METODE PENELITIAN

Peneliti melakukan serangkaian proses dimana setiap proses berpotensi membantu penelitian ini menjadi lebih mudah. Berikut adalah diagram alur dari tahapan proses tersebut :



Gambar 1. Alur Penelitian

Gambar 1, menjelaskan alur dari penelitian yang peneliti susun. Gambar 1 menampilkan langkah mengidentifikasi, merumuskan dan memahami permasalahan yang akan diteliti, salah satunya

yaitu dengan melakukan studi literatur, lalu selanjutnya peneliti mengumpulkan data berupa data nilai siswa kelas 7. Data yang terkumpul sebanyak 31 siswa. Kemudian, peneliti melakukan analisis hasil dengan menerapkan Metode K-Means pada setiap data hingga menemukan hasil akhir.

2.1. Metode Pengambilan data

Peneliti menggunakan metode pengumpulan data kuantitatif yang meliputi,

a. Dokumentatif

Peneliti menggunakan metode ini untuk mengumpulkan data nilai keseluruhan siswa kelas VII yang akan diteliti.

2.2. Data Mining

Data mining merupakan proses yang menerapkan matematika, berbagai teknik statistik, kecerdasan buatan yang bertujuan untuk mengidentifikasi dan mengekstraksi data/ informasi berharga dari sekumpulan data yang lebih besar [14]. Data mining memiliki beberapa teknik seperti *clustering*, analisis regresi, *text mining* dan lain lain.

2.3. Davies Bouldin Indeks (DBI)

Evaluasi DBI (*Davies Bouldin Index*) perlu dilakukan untuk mencari *cluster* terbaik dari metode *clustering*. Nilai *performance* proses *clustering* akan semakin baik apabila nilai tersebut semakin kecil [15]. DBI berperan cukup penting dalam penelitian ini karena dengan DBI, peneliti dapat mengetahui kevalidan tingkat performa K-Means dalam menentukan K (Kluster).

2.4. Metode K-Means

Metode K-Means merupakan metode *clustering* yang mengelompokkan data yang mirip satu sama lain kedalam cluster yang sama. Data yang memiliki kemiripan atau sama akan dikelompokkan kedalam satu cluster dan yang berbeda akan di kelompokkan ke *cluster* yang lain. [16]. Berikut adalah gambaran umum proses *clustering* Algoritma K-Means :

1. Menentukan jumlah K (*cluster*).
2. Memilih titik pusat cluster (*centreoid*) secara acak. Titik *centreoid* ditentukan acak sesuai dengan jumlah cluster.
3. Menghitung jarak setiap *cluster* dengan menggunakan rumus *Euclidean Distance*. Rumus ini menghitung selisih data dengan titik *centreoid* lalu menjumlahkan tiap data tersebut. Berikut adalah rumus *Euclidean Distance* :

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Gambar 2. Rumus *Euclidean Distance*

Keterangan Gambar 2 :

- $\sum_{i=1}^n$, $i=1$ bermakna bahwa perhitungan penjumlahan dimulai dari atribut pertama hingga n, n merupakan jumlah total dari atribut data yang akan dihitung.
- $(x_i - y_i)^2$, x_i merupakan data awal dari atribut pertama dan nilai dari y_i merupakan hasil rata rata dari semua data *cluster* pada iterasi ke-2 dan seterusnya.

4. Memperbarui titik pusat *cluster* (*centreoid*).

Berikut adalah rumus mencari titik pusat *cluster* :

$$C_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_i$$

Gambar 3. Rumus memperbarui *centreoid*

Keterangan Gambar 3 :

- n_j , merepresentasikan jumlah data dari j (*cluster*) yang sudah di alokasikan sebelumnya.
 - $\sum_{i=1}^{n_j} x_i$, $i=1$ bermakna bahwa perhitungan penjumlahan dimulai dari atribut pertama hingga n_j , n_j merupakan jumlah total dari atribut data *cluster* yang akan dihitung sesuai dengan posisi atribut ke i .
5. Ulangi langkah-langkah 2 dan 3 hingga tidak ada perubahan yang signifikan dalam proses pengelompokan data.

2.5. Datasheet (Data Nilai Siswa)

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	Amelia 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 Alini Nur...	cluster_0	90	94	78	78	82	80	80	77	71	85	84
7	Ayu Suliah	cluster_0	89	94	78	78	85	80	78	83	75	85	85
8	Ben Abdi Pa...	cluster_0	86	86	78	78	78	79	72	71	53	80	80
9	Caesha Afri ...	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 ARIANS ...	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 Aprily...	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	86	89

Gambar 4. Datasheet (Data Nilai Siswa)

Gambar 4 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.

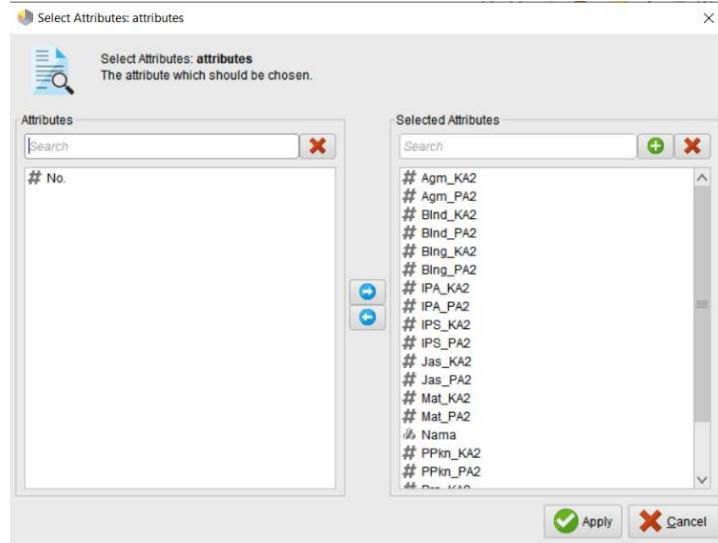
3. HASIL DAN PEMBAHASAN

Proses perhitungan penerapan Metode K-Means untuk menentukan siswa unggulan ini dilakukan menggunakan *tools* RapidMiner. Proses perhitungan ini hanya berfokus pada mata pelajaran yang diujikan OSN (Olimpiade Sains Nasional) saja seperti Matematika, IPA dan IPS. Berikut hasil dan pembahasan yang sudah dilakukan oleh peneliti dengan menggunakan tools yang sudah dijelaskan diatas.

3.1 Data Selection

Setelah semua data terkumpul, tahapan yang perlu dilakukan selanjutnya yaitu menyeleksi *attribute* yang akan di gunakan dalam pengimplementasian Algoritma K-Means. Tahapan seleksi data dalam rapidminer dapat menggunakan operator *select attribute*. Fitur ini berfungsi untuk

memilah keseluruhan atribut data yang dibutuhkan oleh operator perhitungan. Berikut dibawah ini merupakan proses tahapan seleksi atribut dari datasheet nilai siswa kelas 7 SMP 11 Sampit.

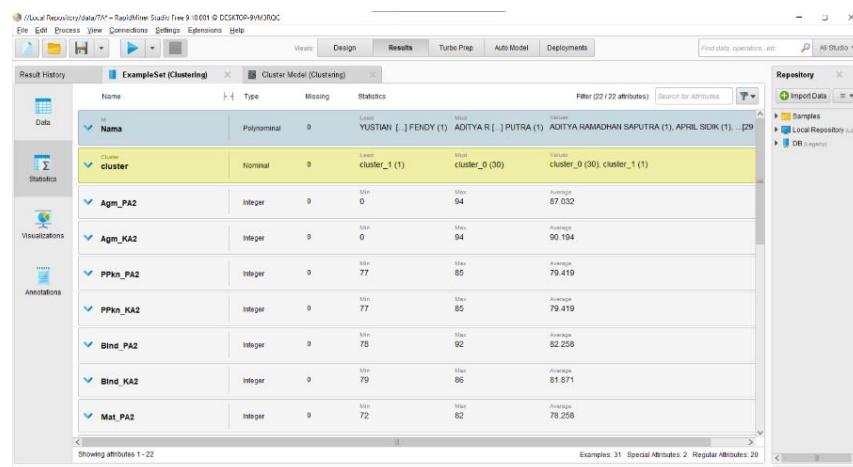


Gambar 5. Proses seleksi atribut data kelas 7

Gambar 5 menampilkan proses seleksi data kelas 7 menggunakan operator *select attribute* pada RapidMiner. Data tersebut memiliki 21 *attribute* yaitu 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. Atribut nama diubah menjadi type data id untuk mempermudah proses pengelolahan data. Setelah diubah, atribut tersebut akan terhitung sebagai label objek yang akan di teliti.

3.2 Pre-Processing

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. Berikut dibawah ini merupakan proses tahapan *Pre Processing* dari datasheet nilai siswa kelas 7 SMP 11 Sampit.



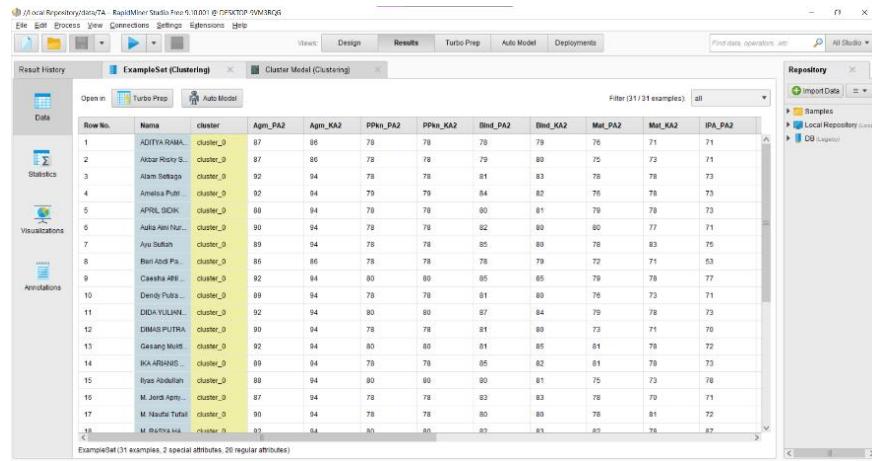
Gambar 6. Proses *Pre-Processing* data kelas 7

Gambar 6 menampilkan stastitika hasil seleksi data kelas 7 setelah dilakukan *running* di *software* Rapid Minner. Gambar 6 diatas menjelaskan bahwa data kelas 7 memiliki 2 atribut spesial dan

31 example dimana atribut Nama masuk kedalam salah satu atribut spesial karena memiliki tipe *polynomial* yang tidak bisa diolah dengan perhitungan Algoritma *K-Means*.

3.3 Transformation Data

Setelah tahapan *pre-processing* selesai, langkah berikutnya adalah melakukan *transformation* data pada hasil dari tahapan sebelumnya. *Transformation* data merupakan proses mengubah / memanipulasi data mentah menjadi lebih sesuai untuk dianalisis. Berikut dibawah ini merupakan proses tahapan *transformation data* dari datasheet nilai siswa kelas 7 SMP 11 Sampit.



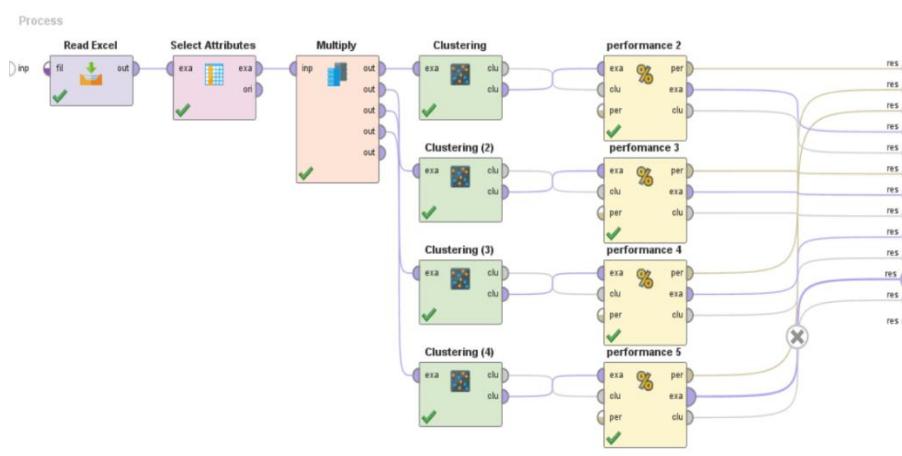
Row No.	Nama	cluster	Agm_PA2	Agm_KA2	PPkm_PA2	PPkm_KA2	Bmt_PA2	Bmt_KA2	Mat_PA2	Mat_KA2	IPA_PA2
1	ADITYA RAMA...	cluster_0	87	86	78	78	79	78	76	71	71
2	Albar Risya S...	cluster_0	87	96	78	78	79	80	75	73	71
3	Alam Setiaga	cluster_0	92	94	78	78	81	83	78	78	73
4	Amerita Putri	cluster_0	92	94	79	79	84	82	76	78	73
5	APRIL SIDIA	cluster_0	80	94	78	78	80	81	79	78	73
6	Asila Aini Nur...	cluster_0	90	94	78	78	82	80	80	77	71
7	Ayu Sulfan	cluster_0	89	94	78	78	85	89	78	83	75
8	Beri Abdillah	cluster_0	86	86	78	78	78	79	72	71	53
9	Cassella Afri...	cluster_0	92	94	80	80	85	85	79	78	77
10	Dendy Putri...	cluster_0	89	94	78	78	81	89	76	73	71
11	DIDA YULIAN...	cluster_0	92	94	80	80	87	84	79	78	73
12	DIMAS PUTRA	cluster_0	90	94	78	78	81	89	73	71	70
13	Gessang Muli...	cluster_0	92	94	80	80	81	85	81	78	72
14	IKA ARABANIS...	cluster_0	89	94	78	78	85	82	81	78	73
15	Iyasa Abdillah	cluster_0	88	94	80	80	80	81	75	73	78
16	M. Jenda Apri...	cluster_0	87	94	78	78	83	83	78	70	71
17	M. Naufit Tufail	cluster_0	90	94	78	78	80	89	78	81	72
18	M. RAYNA HA...	cluster_0	92	94	81	81	82	81	78	87	87

Gambar 7. Proses *Transformation Data* data kelas 7

Gambar 7 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*.

3.4 Data Mining

Setelah melalui tahapan *Transformation data*. Langkah berikutnya yaitu melakukan tahapan *data mining* pada hasil dari tahapan sebelumnya. Berikut dibawah ini merupakan proses tahapan *Data Mining* dari datasheet nilai siswa kelas 7 SMP 11 Sampit :



Gambar 8. Proses *Clustering* pada data kelas 7

Gambar 8 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*.

3.5 Analisis Cluster

Berikut adalah hasil *cluster* dari proses penerapan Metode K-Means untuk menentukan siswa unggulan menggunakan Rapidminner dari *datasheet* nilai siswa kelas 7 yang sudah dilampirkan diatas. Berikut adalah hasil pemodelan jumlah *cluster* yang terbentuk dari perhitungan Metode K-Means :

Cluster Model

Cluster 0: 30 items
Cluster 1: 1 items
Total number of items: 31

Gambar 9. Hasil *Cluster* Kelas 7

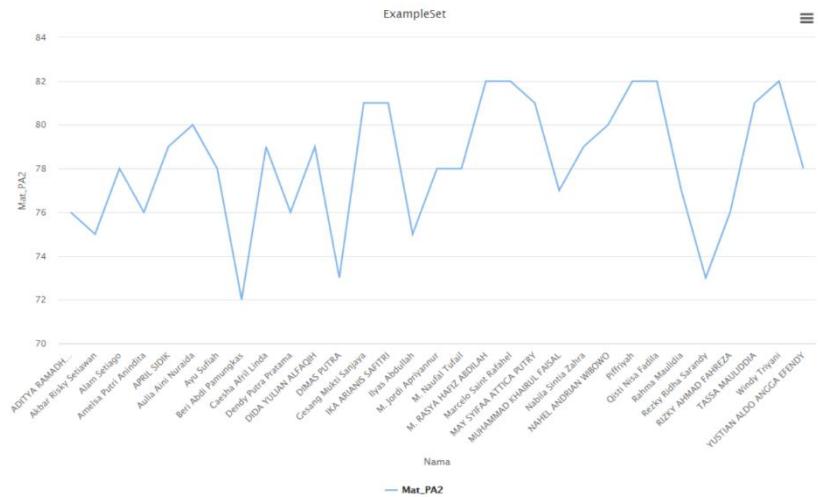
Gambar 9 menampilkan hasil dari jumlah *cluster* yang terbentuk di kelas 7 dimana *cluster* 0 mendapatkan 30 data dan *cluster* 1 mendapatkan 1 data. *Cluster* 0 merupakan representasi dari siswa unggul dan cluster 1 merupakan representasi dari siswa bukan unggulan dari seluruh mata pelajaran. Berikut di bawah ini Gambar 10 yang merupakan tampilan hasil dari keseluruhan data yang sudah melalui tahapan *Clustering* menggunakan Metode *K-Means* :

Row No.	Nama	cluster	Agrm_PA2	Agrm_KA2	PPkn_PA2	PPkn_KA2	Bind_PA2	Bind_KA2	Mat_PA2	Mat_KA2	IPA_PA2	I
1	ADITYA RAMA...	cluster_0	87	86	78	78	78	79	76	71	71	
2	Akbar Risky S...	cluster_0	87	86	78	78	79	80	75	73	71	
3	Alam Setiago	cluster_0	92	94	78	78	81	83	78	78	73	
4	Amelia 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	Aulia Aini Nur...	cluster_0	90	94	78	78	82	80	80	77	71	
7	Ayu Sulfah	cluster_0	89	94	78	78	85	80	78	83	75	
8	Beri Abdi Pa...	cluster_0	86	86	78	78	78	79	72	71	53	
9	Caesha Afri ...	cluster_0	92	94	80	80	85	85	79	78	77	
10	Dendy 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	90	94	78	78	81	80	73	71	70	
13	Gesang Mukti...	cluster_0	92	94	80	80	81	85	81	78	72	
14	IKA ARIANIS ...	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. Inriqi Anriq	cluster_0	87	94	78	78	83	83	78	70	71	

Gambar 10. Data hasil *Cluster* Kelas 7

Data yang ditampilkan pada Gambar 10 dapat digambarkan juga dengan grafik untuk diketahui perankingan data di setiap mata pelajarannya. Berikut tampilan grafik perankingan data kelas 7 dapat dilihat pada gambar 10 dibawah ini :

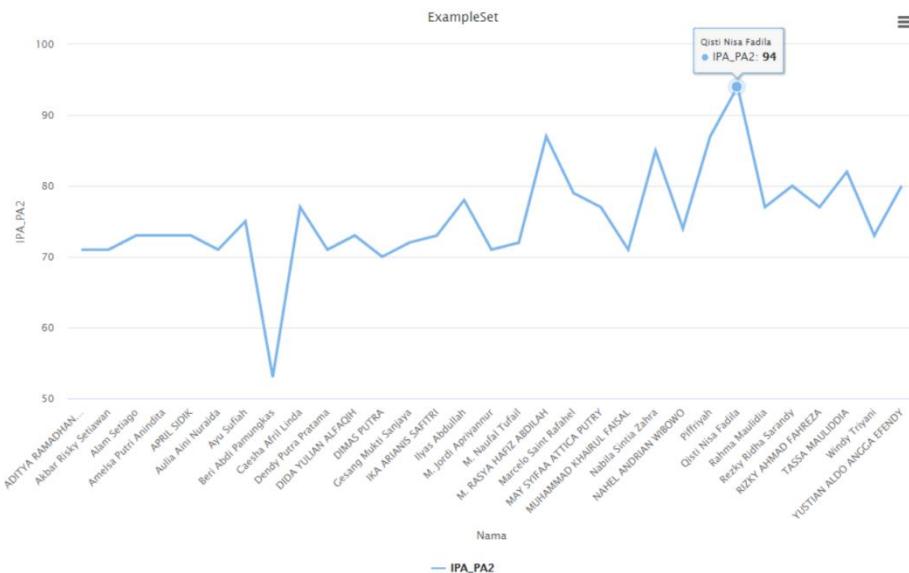
3.5.1 Mata Pelajaran Matematika



Gambar 11. Grafik Ranking Nilai Mata Pelajaran Matematika Kelas 7

Pada Gambar 11 diketahui secara visual dari grafik bahwa siswa dengan nama M. Rasya Hafiz Abdillah, Marcelo Saint Rahafel, Piffriyah, Qisti Nisa Fadilla dan Windy Triyani menjadi siswa paling unggul pada mata pelajaran Matematika dibandingkan temannya yang lain.

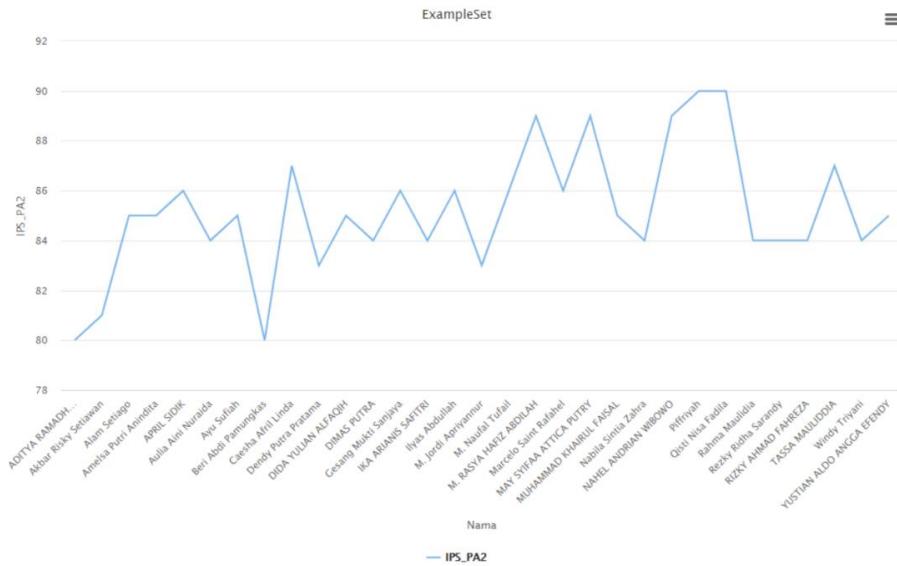
3.5.2 Mata Pelajaran IPA



Gambar 12. Grafik Ranking Nilai Mata Pelajaran IPA Kelas 7

Pada Gambar 12 diketahui secara visual dari grafik bahwa siswi dengan nama Qisti Nisa Fadilla menjadi siswa paling unggul pada mata pelajaran IPS dibandingkan temannya yang lain.

3.5.3 Mata Pelajaran IPS



Gambar 13. Grafik Ranking Nilai Mata Pelajaran IPS Kelas 7

Pada Gambar 13 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.

3.6 Evaluasi Model

Setelah proses seleksi data nilai selesai, langkah berikutnya adalah melakukan evaluasi model pada hasil dari tahapan sebelumnya. Setiap pengujian menggunakan algoritma K-Means menghasilkan cluster yang berbeda-beda tergantung pada nilai K yang digunakan. Dalam empat percobaan yang dilakukan, nilai K yang berbeda menghasilkan hasil yang berbeda pula. Berikut dibawah ini hasil nilai pengujian *performance* data kelas 7 yang paling baik yang ditampilkan pada Gambar 14 :



Gambar 14. Hasil Nilai pengujian *Performance* data Kelas 7

Berikut di bawah ini adalah tabel hasil pengujian performa K-2 hingga K-5 (K = Jumlah Cluster):

Tabel 1. Hasil Pengujian *Performance* data kelas 7

Nilai K	Davies Bouldin Index (DBI)
2	-0.091
3	-0,866
4	-0,681
5	-0,906

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.

4. 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.

DAFTAR PUSTAKA

- [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 Pengelompokan 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.
- [9] A. Sulistiawati and E. Supriyanto, “Implementasi Algoritma K-means Clustering dalam Penetuan 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),” *Komutek*, 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 Menggunkan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru,” *Sistemasi*, vol. 7, no. 3, p. 238, 2018.