



## **PENGELOMPOKAN PERFORMA SISWA DALAM PEMBELAJARAN BAHASA INDONESIA MENGGUNAKAN ALGORITMA *K-MEANS* *CLUSTERING* DI SMPN SATAP LAMBAKARA**

*Grouping Student Performance In Indonesian Language Learning Using The K-Means  
Algorithm Clustering at SMPN Satap Lambakara*

**Arcelinda Ana Mila<sup>1</sup>, Reynaldi Thimotius Abineno<sup>2</sup>, Arini Aha Pekuwali<sup>3</sup>**

<sup>1,2,3</sup>Program Studi Teknik Informatika, Fakultas Sains Dan Teknologi, Universitas Kristen Wira  
Wacana Sumba

Jl.R.Suprato No.35,Prailiu,Kec.Kota Waingapu,Kabupaten Sumba Timur,Nusa Tenggara Timur

*Corresponding author:* [arcemila69@gmail.com](mailto:arcemila69@gmail.com)<sup>1</sup>, [reynaldi@unkriswina.ac.id](mailto:reynaldi@unkriswina.ac.id)<sup>2</sup>  
[arini.pekuwali@unkriswina.ac.id](mailto:arini.pekuwali@unkriswina.ac.id)<sup>3</sup>

### **ABSTRACT**

*Language literacy, which includes the ability to read and write, is an important component learning Indonesian. subject not only serves as a communication tool but also as a means to add cultural values, enrich insights, and improve students' critical thinking skills. This research makes an important contribution field education by showing how data analysis and data mining techniques can be used to improve the quality of learning. The implementation of K-Means clustering is expected to be a reference for other schools in an effort to improve student learning performance through a more targeted and personalized approach. However, student performance in these subjects often shows significant variation, influenced by various factors such as individual ability, learning interests, and learning environment. The results show that K-Means Clustering can effectively group students into multiple clusters based on the Davias Bouldin Index value test using RapidMiner, produced optimal performance, namely K=2 with a value of 0.882 if this value is close to 0, concluded evaluation show that students a good cluster, cluster 0 consists of 34 students with low language literacy factors, and cluster 1 is the lowest of 16 students with high language proficiency.*

**Keywords:** *Education, Indonesian, literacy, knowledge, skills and K-Means Clustering.*

### **ABSTRAK**

Literasi bahasa yang mencakup kemampuan membaca dan menulis, adalah komponen penting dalam pembelajaran Bahasa Indonesia. Mata pelajaran ini tidak hanya berfungsi sebagai alat komunikasi tetapi juga sebagai sarana untuk menambahkan nilai-nilai budaya, memperkaya wawasan, dan meningkatkan kemampuan berpikir kritis siswa. Penelitian ini memberikan kontribusi penting dalam bidang pendidikan dengan menunjukkan bagaimana analisis data dan teknik data mining dapat digunakan untuk meningkatkan kualitas pembelajaran. Implementasi *K-Means clustering* diharapkan dapat menjadi acuan bagi sekolah lain dalam upaya meningkatkan performa belajar siswa melalui pendekatan yang lebih terarah dan personal. Namun, performa siswa dalam mata pelajaran ini sering kali menunjukkan variasi yang signifikan, dipengaruhi oleh berbagai faktor seperti kemampuan individu, minat belajar, dan lingkungan belajar. Hasil penelitian menunjukkan bahwa *K-Means Clustering* dapat secara efektif mengelompokkan siswa ke dalam beberapa cluster berdasarkan pengujian nilai David Bouldin Index menggunakan RapidMiner, menghasilkan performa yang optimal yaitu K=2 dengan nilai 0.882 apabila nilai ini mendekati dari 0 dapat disimpulkan bahwa hasil evaluasi menunjukkan bahwa siswa berada dalam klaster yang baik, cluster 0 terdiri dari 34 siswa dengan faktor literasi bahasa yang rendah, dan cluster 1 terdiri dari 16 siswa dengan kemahiran bahasa yang tinggi.

**Kata kunci:** *Pendidikan, Bahasa Indonesia, literasi, pengetahuan, keterampilan dan K-Means Clustering.*

## PENDAHULUAN

Literasi bahasa membantu siswa memahami dan mengembangkan strategi yang efektif untuk strategi membaca dan menulis yang membaca dan menulis, sehingga pembelajaran bahasa dalam bahasa Indonesia tidak hanya berfungsi sebagai sarana komunikasi tetapi juga sebagai sarana untuk meningkatkan nilai-nilai budaya, meningkatkan kesadaran, dan meningkatkan kemampuan berpikir siswa. Jadi pembelajaran bahasa Indonesia tidak hanya berfungsi sebagai alat komunikasi saja, tetapi juga sebagai alat untuk meningkatkan nilai-nilai budaya, meningkatkan kesadaran, dan meningkatkan kemampuan peserta didik untuk berpikir kritis. Namun, kinerja siswa pertunjukan dalam bidang studi ini secara konsisten menunjukkan variasi signifikan yang dipengaruhi oleh beberapa faktor seperti kemampuan individu, kemauan belajar, lingkungan belajar, dan beberapa kendala khusus dalam pembelajaran bahasa seperti lupa, gagap, dan tersendat. Dalam konteks pengajaran bahasa Indonesia guru harus mampu merancang proses pembelajaran yang ketat dan fleksibel.

Salah satu masalah utama di SMPN Satap Lambakara adalah belum optimalnya pengelolaan dan pemanfaatan data nilai siswa, variasi dalam performa siswa dalam pembelajaran Bahasa Indonesia menjadi tantangan karena literasi bahasa Indonesia rendah mengacu pada kemampuan yang kurang dalam membaca, menulis, menyimak, dan berbicara dalam bahasa Indonesia. Pada tahun 2023 kemampuan literasi bahasa siswa berkurang, kemampuan literasi bahasa berkurang 13,33 peserta didik yang tercapai kompetensi minimum untuk literasi bahasa perlu untuk berupaya mendorong peserta didik dalam mencapai dalam kompetensi minimum. Pada tahun 2024 kemampuan literasi siswa meningkat, 51,11 peserta didik telah mencapai kompetensi minimum untuk literasi bahasa, namun perlu upaya mendorong lebih banyak peserta didik dalam mencapai kompetensi minimum.

Untuk mengatasi masalah ini, seperti data teknik pengelompokan *K-Means clustering* sangatlah penting. metode memungkinkan ini sekolah untuk memberi peringkat siswa memberi sesuai dengan kinerja mereka yang diukur berdasarkan kemahiran berbahasa, pemahaman, dan ketelitian. Hal ini lebih mudah untuk mengidentifikasi kebutuhan spesifik siswa setiap cluster dan mengembangkan strategi pengajaran yang lebih efektif.

Tujuan dari penelitian ini adalah menerapkan *K-Means clustering* pada data siswa SMPN Satap Lambakara guna membantu pihak sekolah. Diharapkan hasil penelitian ini dapat digunakan sebagai pedoman untuk meningkatkan kinerja siswa melalui pengajaran yang lebih fokus dan individual, sehingga setiap siswa dapat mencapai potensi maksimalnya.

## MATERI DAN METODE

### Data mining

Data mining adalah data proses penggunaan menggunakan kumpulan data besar untuk mengidentifikasi pola yang berpotensi berguna. kumpulan data besar untuk mengidentifikasi pola yang berpotensi berguna. Proses proses pengumpulan data pengumpulan kegunaan menggunakan statistik, matematika, dan penilaian manusia untuk mengumpulkan informasi tentang kemungkinan kejadian di masa depan ( Naibaho & Zahra , 2023 ).

Data mining disebut juga dengan penemuan pengetahuan (penambangan) berdasarkan data, perolehan pengetahuan dari data, analisis pola, data arkeologi, pengerukan data, pengumpulan data, intelijen bisnis, dan bidang terkait lainnya. Ada metode dan teknik lain yang dapat digunakan untuk data mining; lainnya namun, metode ini memerlukan langkah-langkah tertentu agar data yang diinginkan dapat dikumpulkan. Ada dua dua jenis fungsi yang datang disertakan dengan data mining dengan prediktif dan deskriptif. Yang pertama bertujuan untuk menemukan informasi yang lebih rinci tentang data yang sedang dianalisis, sedangkan yang kedua digunakan untuk menemukan model dari data yang sedang dianalisis (Kaligis & Yulianto, 2022).

## **Clustering Data**

*Clustering* adalah suatu teknik untuk membagi kumpulan data menjadi kumpulan beberapa kelompok menjadi beberapa kelompok berdasarkan hasil yang diinginkan. Dalam penambangan data pertambangan, *clustering* adalah proses pengelompokan titik data, atau objek, ke dalam *cluster* ( grup ) dan memastikan bahwa setiap cluster memiliki data yang hampir identik dengan data aslinya dan *clustering* adalah proses pengelompokan titik data, atau objek ke dalam *cluster* ( kelompok ) dan memastikan bahwa setiap *cluster* memiliki data yang hampir identik dengan data aslinya dan dapat dibandingkan dengan objek di *cluster* lainnya (Sibuea & Safta, 2017).

*Clustering* sering sering digunakan sebagai langkah pertama dalam proses penambangan data. Sebagai langkah pertama dalam proses penambangan data. Banyak algoritma klasifikasi algoritma, seperti *K-Means*, *Improved K-Means*, *K-Medoids* (PAM), *Fuzzy C-Means*, DBSCAN, CLARANS, dan *Fuzzy Subtractive*, telah digunakan oleh para peneliti sebelumnya (Harahap, 2021).

### **Metode K-Means**

Pengelompokan adalah metode analisis kluster terarah yang dirancang untuk mengelompokkan objek ke dalam *cluster* yang disebut K. Setiap objek dalam *cluster* yang dihasilkan kemudian dianalisis menggunakan pendekatan rata-rata bersarang. Pendekatan ini paling terkenal. Pendekatan ini juga merupakan alat pembelajaran yang paling mudah dipahami dan langsung untuk memecahkan masalah yang melibatkan ekstraksi data dari kumpulan data tertentu. Alat pembelajaran yang paling mudah digunakan adalah yang memerlukan sedikit usaha untuk dipelajari dan dapat diterapkan pada berbagai skenario masalah mengelompokkan objek tertentu ke dalam cluster k. Setelah itu, setiap objek yang ada dalam cluster dianalisis menggunakan metode rata-rata penempatannya (Harahap, 2021).

### **Davies-Bouldin Index**

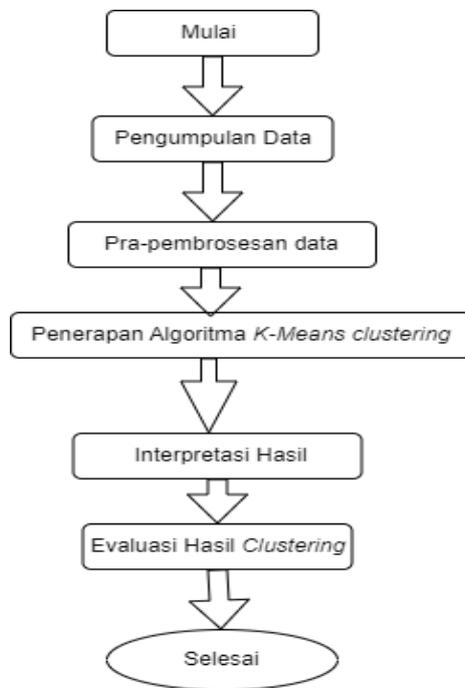
Indeks Davies-Bouldin ( DBI) adalah alat alat statistik yang digunakan untuk menentukan jumlah klasifikasi ideal setelahnya klasifikasi akhir proses. Tujuan sasaran dari pendekatan DBI ini adalah untuk meminimalkan besarnya tumpang tindih antara satu kelas dengan kelas lainnya, Semakin kecil nilai DBI itu yang diperoleh (non-negatif $\geq 0$ ), maka semakin diperoleh klasifikasi yang diperoleh dari K-Means clustering Setelah menerapkan DBI, suatu cluster akhirnya akan memiliki yang optimal skema clustering yang optimal dengan DBI yang minimal ( Harahap , 2021).

### **Rapidminer**

*RapidMiner* adalah mesin perangkat lunak pembelajaran mesin untuk ilmu data. Untuk ilmu data dan pembelajaran mesin platform menawarkan ini banyak alat untuk menganalisis data, membuat model, mengevaluasi, dan mempraktikkannya. *RapidMiner* dirancang agar mudah digunakan, memungkinkan pengguna dengan mudah membuat dan memodifikasi berbagai model. Fungsionalitas *drag-and-drop* kegunaan memungkinkan pengguna membuat kelompok kerja untuk memudahkan ekstraksi dan analisis data (Rafi Nahjan dkk., 2023).

Data khusus data itulah yang bisa dilihat yang bisa dilihat Data yang telah dikumpulkan meliputi data siswa SMPN Satap Lambakara, termasuk nilai yang diambil tahun ajaran 2022–2023 yaitu data semester 1 dan 2, pengetahuan, keterampilan, dan ujian sekolah, serta informasi yang mungkin berdampak pada keterampilan menulis siswa yang pas-pasan.

Berikut adalah beberapa ilustrasi jenis yang akan dibuat.



Gambar 1. Tahapan Penelitian

Pengumpulan data telah dilakukan di SMPN Satap Lambakara. Penelitian menggunakan data pada tahun ajaran 2022/2023, terkait data nilai yang dilakukan dalam pembahasan yang akan Pengelompokan performa siswa dalam pembelajaran Bahasa Indonesia bertujuan untuk mengidentifikasi pola dan trend dalam data hasil belajar siswa dalam menguraikan proses pengklarifikasian atau *clustering* dataset dengan menggunakan algoritma *K-Means* pengelompokkan ini dengan dilakukan dengan proses pengujian menggunakan *software RapidMiner Studio*.

pra-pemrosesan adalah awal data sangat penting sejak itu karena dilakukan setelah rekonsiliasi data ke data dan inisialisasi data selesai. Dataset himpunan data yang digunakan dalam penelitian ini terdiri dari sekitar 50 record dan dibagi menjadi 8 atribut, yaitu sebagai berikut : no, id siswa, jk, pengetahuan semester 1, keterampilan, pengetahuan semester 2, keterampilan, dan ujian sekolah.

Pada fase ini, dilakukan pengelompokan dengan menggunakan metode pengelompokan yang telah ditentukan sebelumnya yaitu *K-Means* pengelompokan dilakukan dengan menggunakan metode pengelompokan yang telah ditentukan sebelumnya, yaitu *K -Means*. Memanfaatkan aplikasi RapidMiner untuk menganalisis data terhambat dari kumpulan data.

Hasil dari proses *clustering* kemudian diinterpretasikan dengan menganalisis karakteristik atau pola yang dimiliki setiap kumpulan titik data. Hasilnya disajikan dalam bentuk tabel dan grafik.

Metode evaluasi yang digunakan untuk menurunkan kualitas setiap cluster dalam proses pengumpulan data adalah *Indeks Davies-Bouldin*. Metode ini digunakan untuk menilai kinerja dan mengatur akurasi saat menggunakan metode *K-Means*

## HASIL DAN PEMBAHASAN

### Pengumpulan Data

Pengumpulan data telah dilakukan di SMPN Satap Lambakara. Penelitian menggunakan data pada tahun ajaran 2022/2023, terkait data nilai yang dilakukan dalam pembahasan yang akan Pengelompokan performa siswa dalam pembelajaran Bahasa Indonesia bertujuan untuk mengidentifikasi pola dan trend dalam data hasil belajar siswa dalam menguraikan proses pengklarifikasian atau *clustering* dataset dengan menggunakan algoritma *K-Means* pengelompokkan ini dengan dilakukan dengan proses pengujian menggunakan *software RapidMiner Studio*.

## Pra-Pemrosesan Data

pra-pemrosesan adalah awal data sangat penting sejak itu karena dilakukan setelah rekonsiliasi data ke data dan inisialisasi data selesai.

Tabel 1. Dataset

No	Nama Siswa	JK	P1	K1	P2	K2	US
1	A*****	0	72	68	72	70	74
2	A*****	1	78	80	82	85	84
3	A*****	0	78	70	79	75	78
4	A*****	1	86	80	86	85	82
5	A*****	1	71	68	74	75	75
6	C*****	0	68	68	90	90	89
.....	.....	.....	.....	.....	.....	.....	.....
48	M*****	0	80	80	77	75	80
49	M*****	1	74	75	77	75	79
50	N*****	1	75	70	74	75	78

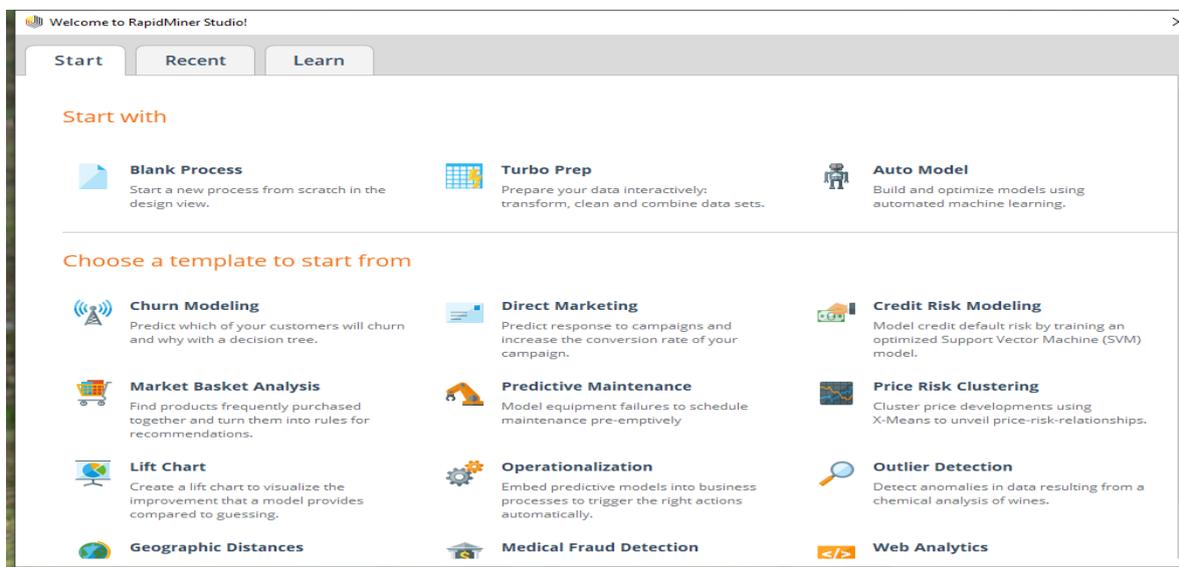
Berdasarkan dataset yang didapatkan dari hasil pengumpulan data, himpunan data yang digunakan dalam penelitian ini terdiri dari sekitar 50 record dan dibagi menjadi 8 atribut, yaitu sebagai berikut : no, nama siswa, jk, pengetahuan semester 1, keterampilan, pengetahuan semester 2, keterampilan, dan ujian sekolah.

## Mengolah Data Clustering

Data yang diperoleh digunakan sebagai input untuk membangun model algoritma *K-Means Clustering* menggunakan perangkat lunak *Rapidminer*.

### Tampilan Menu Utama *RapidMiner*

Menu utama *RapidMiner* memiliki tampilan awal yang mencakup berbagai komponen dan fitur. Salah satunya adalah “*New Process*”, yang merupakan langkah pertama dalam menggunakan *RapidMiner*.



Gambar 2. Tampilan Menu Utama RapidMiner Versi 9.10

### Tampilan *Select The Cells To Import*

Sistem memberikan instruksi tentang bagaimana cara mengimpor data baru yang akan diolah selanjutnya, dengan format data saat ini dalam bentuk Excel. Tahapan ini terlihat dalam gambar berikut ini.

**Select the cells to import.**

Sheet: Sheet1 Cell range: B4:G14 Select All  Define header row: 1

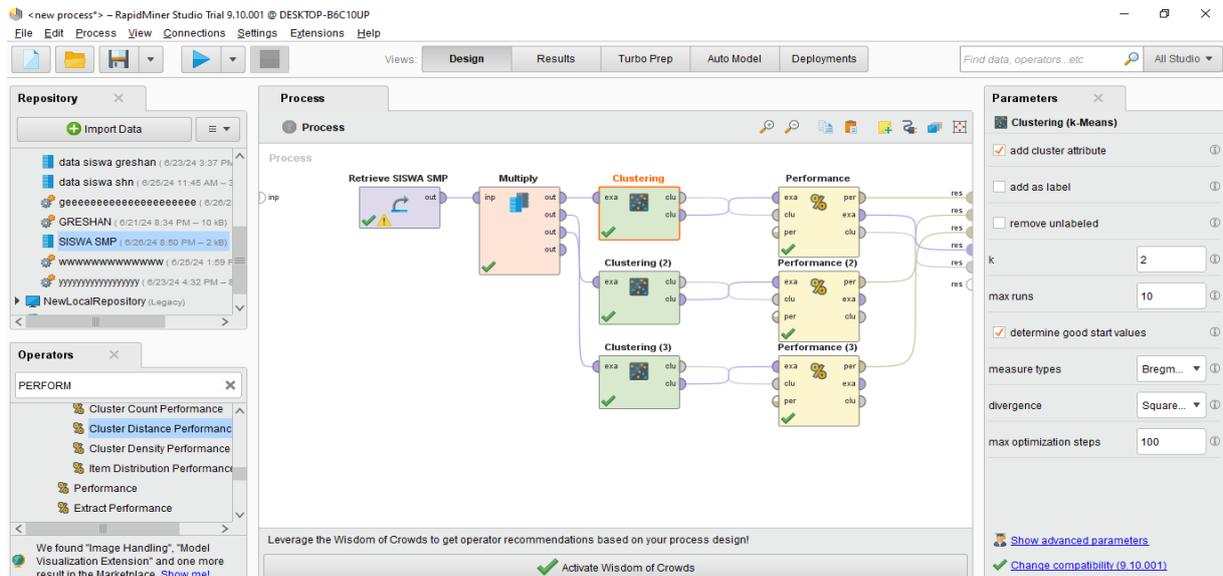
	A	B	C	D	E	F	G
1	No	Id siswa	Semester 1		Semester 2		US
2							
3			P	K	P	K	
4	1.000	Abraham Hina...	72.000	68.000	72.000	70.000	74.000
5	2.000	Ananda Uru A...	78.000	80.000	82.000	85.000	84.000
6	3.000	Ardhan Hia Wi...	78.000	70.000	79.000	75.000	78.000
7	4.000	Aristi Nyanga ...	86.000	80.000	86.000	85.000	82.000
8	5.000	Aurelia Tri Put...	71.000	68.000	74.000	75.000	75.000
9	6.000	Chelsea Olivi...	74.000	70.000	75.000	75.000	76.000
10	7.000	Dirli Syaputra ...	68.000	68.000	90.000	90.000	89.000
11	8.000	Ela Gracia Ra...	89.000	85.000	72.000	75.000	74.000
12	9.000	Ferdinandus ...	71.000	70.000	70.000	75.000	75.000
13	10.000	Indra May Wal...	69.000	70.000	73.000	75.000	75.000
14	11.000	Jeni Wori Hana	71.000	70.000	75.000	75.000	78.000

← Previous Next → ✖ Cancel

Gambar 3. Tampilan select the cells to import

### Tampilan Clustering RapidMiner

Algoritma *K-Means* akan diterapkan dalam proses yang telah dibentuk untuk melakukan klusterisasi atau pengelompokan. Operator klusterisasi yang digunakan adalah *K-Means*. Setelah pembuatan model selesai, proses akan dijalankan untuk mendapatkan hasil dari klusterisasi. Dan beberapa operator seperti *Read Excel* dimana dataset akan dimasukkan dalam *Rad Excel*, *Multiplay* yang berfungsi untuk menjalankan satu atau lebih operator, operator *cluster* yaitu algoritma *K-Means*, kemudian *performance* untuk evaluasi kinerja dari *K-Means*.

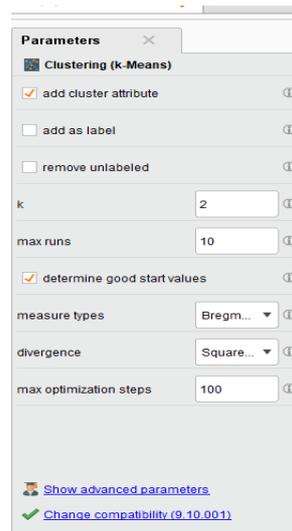


Gambar 4. Tampilan Clustering RapidMiner

### Penentu Jumlah Cluster

Langkah melakukan pengaturan algoritma *K-Means* melalui menu *Parameter Clustering K-Means* perlu memasukkan nilai K yang diinginkan dalam literasi yang dilakukan sebanyak 10 kali, nilai K yang dimasukkan adalah K=2,3,4. Tujuan literasi yang paling kurang, maka dianggap sebagai *cluster* terbaik. Setelah iterasi selesai, proses clusterisasi telah dilakukan untuk mendapatkan hasil dari algoritma *K-Means* yang diimplementasikan melalui aplikasi *RapidMiner*. Setelah melalui beberapa literasi, ditemukan bahwa nilai optimal untuk jumlah

*cluster* K adalah 2. Hasil dari *clusterisasi* akhir diperoleh menggunakan algoritma *K-Means* berdasarkan nilai K yang telah menentukan tersebut.



Gambar 5. Parameter Jumlah *cluster*

### Interprestasi Hasil

Setelah klasterisasi selesai, hasil *cluster* pada setiap data nilai akan ditampilkan dalam bentuk tabel.

Tabel 2. Anggota Hasil *Cluster*

Cluster	Berdasarkan Id Siswa	Jumlah
0	9,10,11,12,13,14,15,20,21 22,23,24,25,26,27,28,29,30 31,32,34,38,39,41,43,44,47 49,50	34
1	2,4,7,8,16,17,18,19,33,40,42,46,48	16
TOTAL		50

Tabel 2 menunjukkan hasil akhir dari aplikasi *K-Means* yang digunakan untuk melakukan pengelompokan data *cluster*. Berdasarkan hasil pengujian, dapat disimpulkan bahwa terdapat 34 siswa dalam klaster 0, sedangkan cluster 1 terdapat 16 siswa.

### Aggregate Value Clustering

Selama proses *clustering*, data diekstraksi dan di kelompokkan menjadi beberapa *cluster* berdasarkan kemiripan tertentu.

Tabel 3. Aggregate Value *Cluster* 0 dan 1

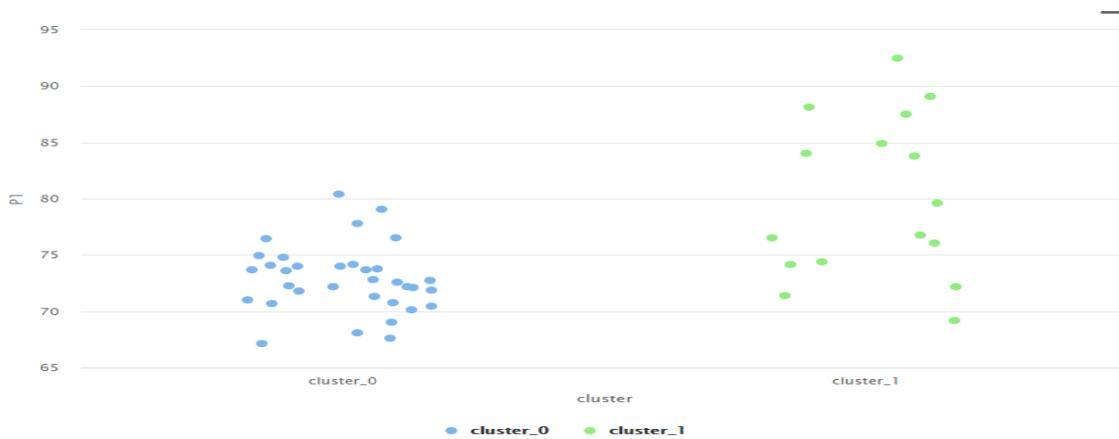
Cluster	Aggregat Value	P1	K1	P2	K2	US
0	Min	67	67	70	70	72
	Max	80	80	81	82	83
	Rata-Rata	72	70	75	74	76
	Standar Deviasi	2,72	2,86	2,70	2,98	2,89
1	Min	68	68	72	75	74
	Max	93	90	92	95	89

Rata-Rata	80	77	81	83	81
Standar Deviasi	7,22	6,95	5,32	6,34	4,14

Tabel 3. ini menampilkan nilai-nilai agregat untuk *cluster* 0, untuk *cluster* nilai semester 1 minimum pengetahuan sebesar 67, sedangkan maksimum 80, nilai minimum keterampilan sebesar 67, Sedangkan maksimum 80, nilai semester 2 minimum pengetahuan sebesar 70, sedangkan maksimum 81, nilai minimum keterampilan sebesar 70, Sedangkan maksimum 82 dan minimum ujian sekolah sebesar 72, sedangkan maksimum 83, dan rata-rata keseluruhan adalah 76 data, kita dapat melihat bahwa anggota *cluster* 0 memiliki nilai semester 1 dan 2 dengan tingkat pengetahuan, keterampilan dan ujian sekolah yang relatif dan sebanding dengan kisaran 76. Sedangkan nilai-nilai agregat untuk *cluster* 1, untuk *cluster* nilai semester 1 minimum pengetahuan sebesar 68, sedangkan maksimum 93, nilai minimum keterampilan sebesar 68, Sedangkan maksimum 90, nilai semester 2 minimum pengetahuan sebesar 72, sedangkan maksimum 92, nilai minimum keterampilan sebesar 75, Sedangkan maksimum 95 dan minimum ujian sekolah sebesar 74, sedangkan maksimum 89, dan rata-rata keseluruhan adalah 81 data, kita dapat melihat bahwa anggota *cluster* 1 memiliki nilai semester 1 dan 2 dengan tingkat pengetahuan, keterampilan dan ujian sekolah yang relatif dan sebanding dengan kisaran 81. Data ini menunjukkan nilai rata-rata standar deviasi yang kecil yang menunjukkan *cluster* 0 yang cenderung dekat dengan nilai rata-rata dibandingkan nilai yang paling besar menunjukkan *cluster* 1 bahwa data tersebut semakin lebar variasi datanya.

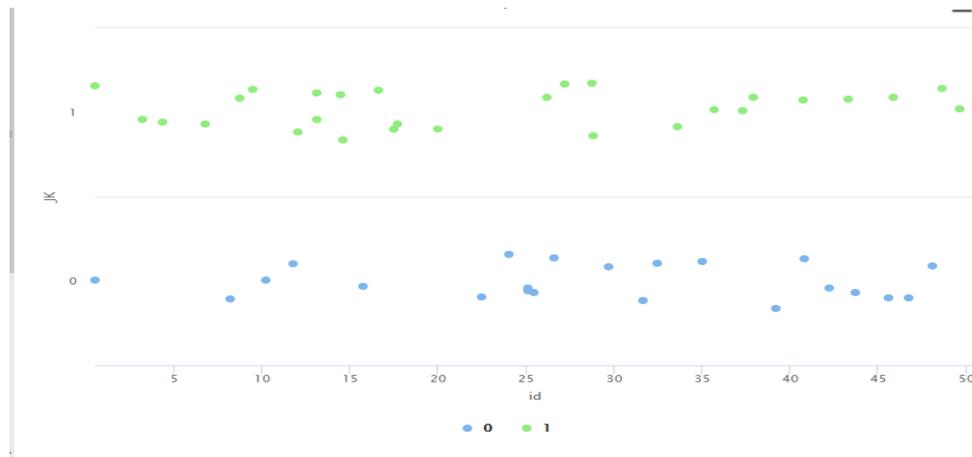
### Visualisasi Hasil Clustering

Hasil analisis data dengan perangkat lunak dengan *RapidMiner* menghasilkan beberapa *output*. Pada gambar visualisasi, terlihat penyebaran data yang menunjukkan jumlah *cluster* data nilai siswa mata pelajaran Bahasa Indonesia di SMPN Satap Lambakara. Visualisasi ini menampilkan sumbu x, sumbu y, dan warna khusus untuk *cluster*.



Gambar 6. Scatter Plot cluster (x), pengetahuan(y)

Gambar 6. menunjukkan tampilan gambar *scatter plot* dari *cluster* yang sebagai variabel x dan pengetahuan sebagai variabel y. sehingga dapat dilihat bahwa nilai yang tertinggi berada pada cluster 1 yaitu berdasarkan *cluster*.

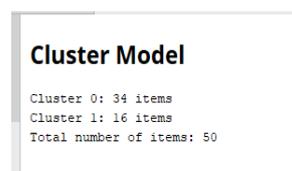


Gambar 7. Scatter plot berdasarkan jenis kelamin

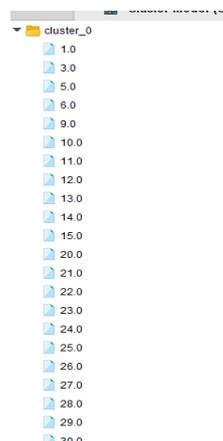
Gambar 7. menunjukkan tampilan gambar scatter plot dari id siswa yang sebagai variabel x dan jenis kelamin siswa sebagai variabel y, sehingga dapat dilihat. Dimana warna biru diidentifikasi laki-laki dan warna hijau diidentifikasi sebagai perempuan. Dapat disimpulkan bahwa nilai tertinggi berada di *cluster* 1 yang diraih oleh perempuan yang ditandai dengan warna hijau.

### Cluster Model

Dalam model *cluster*, terdapat beberapa tampilan yang menampilkan hasil *cluster*, salah satunya adalah tampilan teks yang menunjukkan hasil pengelompokan berdasarkan cluster dan jumlah anggota dalam setiap *cluster*. *Cluster* 0 memiliki 34 item, cluster 16.



Gambar 8. Cluster Model



Gambar 9. Folder View

Dari gambar 9, *Folder view* menunjukkan data secara *komprehensif* dari bagian-bagian *cluster*, dengan setiap anggota dari dua *cluster* menampilkan id variabel. Setelah menerapkan pengelompokan data nilai siswa menggunakan *rapidminer*.

### Evaluasi Hasil Clustering

Metode K-Means clustering digunakan bersama dengan perhitungan David-Bouldin Index (DBI) untuk menentukan *cluster* yang paling optimal. Proses ini melibatkan tiga kali percobaan iterasi untuk mendapatkan nilai DBI terbaik.

Tabel 4. Perhitungan *cluster* DBI yang terpilih

K	Avg. within centroid distance	DBI	Terpilih
2	55.890	0.882	K = 2
3	45.178	0.983	
4	82.522	1.134	



Gambar 10. Davies Bouldin

Gambar 4.10 menampilkan informasi mengenai nilai DBI dan performa klasterisasi. Data ini didapatkan melalui pengujian dengan perhitungan *Davies-Bouldin Index* (DBI) menggunakan *RapidMiner*. Nilai DBI terbaik ditemukan ketika menggunakan  $K=2$ , dengan nilai DBI sebesar 0.882 yang mendekati 0. Ini menunjukkan bahwa semakin kecil nilai DBI yang diperoleh (non-negatif  $\geq 0$ ), semakin baik klasterisasi yang terbentuk. Hasil ini diperoleh dengan metode *K-Means* dalam klasterisasi yang dilakukan menggunakan *RapidMiner*.

Pengujian hasil klasterisasi dilakukan menggunakan perhitungan *Davies-Bouldin Index* dengan jumlah *cluster* sebanyak 2. Pada dataset, nilai *Davies-Bouldin Index* yang diperoleh adalah 0.882. Berdasarkan hasil pengujian ini, dapat disimpulkan bahwa pada data nilai semester 1 dan semester 2 yaitu pengetahuan, keterampilan dan Ujian Sekolah tahun ajaran 2022/2023 terdapat klaster yang sesuai relatif. Nilai pengujian sebesar 0.882 menunjukkan kekuatan cluster yang baik.

## KESIMPULAN

Berdasarkan penelitian yang dilakukan di SMPN 1 Satap Lambakara dengan menggunakan metode *K-Means clustering*, dapat disimpulkan bahwa penerapan data mining ini efektif dalam mengelompokkan berdasarkan performa siswa dalam pembelajaran Bahasa Indonesia dengan nilai pengetahuan, keterampilan, ujian sekolah. Hasil *clustering* menunjukkan bahwa siswa dapat dibagi ke dalam 2 kelompok yang memiliki karakteristik nilai yang serupa artinya karakteristik antar cluster berbeda sehingga perlu dilakukan clustering, sehingga mempermudah pihak sekolah dalam mengidentifikasi kebutuhan dan potensi masing-masing siswa. Selain itu, dari hasil analisis dengan *Davies-Bouldin Index* (DBI), nilai DBI terbaik adalah 0.882 dengan jumlah *cluster* sebanyak dua, yang menunjukkan kekuatan *cluster* yang baik. Penggunaan *K-Means clustering* dalam penelitian ini berhasil memberikan informasi yang mendalam mengenai pola-pola literasi bahasa, yang dapat digunakan oleh sekolah untuk lebih memahami kinerja akademik siswa dan merancang strategi pembelajaran yang lebih personal dan efektif. Dari tabel-tabel *aggregate value* yang dihasilkan, dapat dilihat bahwa *cluster* 0 terdiri dari siswa dengan nilai pengetahuan, keterampilan dan ujian sekolah yang relatif seimbang di sekitar angka 76, *cluster* 1 memiliki nilai pengetahuan, keterampilan dan ujian sekolah dengan rata-rata sekitar 81. Dengan demikian, penelitian ini diharapkan dapat memberikan wawasan tambahan dalam pengelompokan data nilai siswa untuk meningkatkan kualitas pembelajaran di SMP Satap Lambakara.

## SARAN

Saran dari penelitian ini, pengembangan lebih rinci tugas proyek ini pada akhirnya mencakup beberapa aspek. Untuk perkembangan lebih detail dari tugas proyek ini pada akhirnya mencakup beberapa aspek. Pertama, efisiensi klasifikasi dapat ditingkatkan algoritma *k - means* dengan algoritma lain. Dengan memodifikasi atau menggabungkan algoritma *k-means* dengan algoritma lainnya. Selain itu, disarankan untuk menambah jumlah data yang digunakan karena algoritma *k-means* secara konsisten memberikan hasil yang lebih baik dengan lebih banyak data. dilakukan secara lebih akurat dengan memperhatikan data masing-masing sekolah di wilayah Sumba Timur.

## DAFTAR PUSTAKA

- acwanda, D. O., & Nataliani, Y. (2021). Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *Aiti*, 18(2), 125–138. <https://doi.org/10.24246/aiti.v18i2.125-138>
- Harahap, F. (2021). Perbandingan Algoritma K-Means dan K-Medoids untuk Clustering Kelas Siswa Tunagrahita. *TIN: Terapan Informatika Nusantara*, 2(4), 191–197. <https://ejurnal.seminar-id.com/index.php/tin/article/download/873/599>
- Kaligis, G. B., & Yulianto, S. (2022). Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokan Kinerja Pegawai. *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, 1(3), 179–193. <https://doi.org/10.24246/itexplore.v1i3.2022.pp179-193>
- Mustafa, P. S., & Masgumelar, N. K. (2022). Pengembangan Instrumen Penilaian Sikap, Pengetahuan, dan Keterampilan dalam Pendidikan Jasmani. *Biormatika : Jurnal Ilmiah Fakultas Keguruan Dan Ilmu Pendidikan*, 8(1), 31–49. <https://doi.org/10.35569/biormatika.v8i1.1093>
- Nurul Alifia, A., Fahrudi Setiawan, A., & Rudhistiar, D. (2024). Penerapan Algoritma K-Means Clustering Dalam Peringatan Dini Resiko Kegagalan Siswa Pada Mata Pelajaran Bahasa Indonesia. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(2), 1174–1181. <https://doi.org/10.36040/jati.v8i2.9075>
- Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, 3(3), 424–439. <https://doi.org/10.51519/journalisi.v3i3.164>
- Sibuea, M. L., & Safta, A. (2017). Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering. *Jurteksi*, 4(1), 85–92. <https://doi.org/10.33330/jurteksi.v4i1.28>
- Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25. <https://doi.org/10.33365/jtk.v15i2.1162>