



# Penggunaan Algoritma *K-Means* Untuk Menganalisis Performa Siswa Dalam Pembelajaran Matematika di SMP Negeri Satap Lambakara

(Using the *K-Means* Algorithm to Analyze Student Achievement in Mathematics Learning at Satap Lambakara State Middle School)

Florensia Konda Ngguna<sup>1</sup>, Arini Aha Pekuwali<sup>2</sup>, dan Reynaldi Thimotius Abineno<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

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

Corresponding author: [florensiakondangguna@gmail.com](mailto:florensiakondangguna@gmail.com)

## ABSTRACT

Education is one of the main pillars of national development, playing a crucial role in shaping young generations into individuals of quality who can contribute positively to society and the nation. SMP Negeri Satap Lambakara, a school in the East Sumba region, faces challenges in improving its students' literacy and numeracy skills. The data on student's knowledge and skills scores are often not utilized optimally to understand each student's level of literacy and numeracy. This study aims to perform clustering to identify literacy and numeracy levels using the *k-means* method, which utilizes data from students' knowledge and skills scores as well as school exams at SMP Negeri Satap Lambakara. Based on the test results of the Davies Bouldin Index calculated using RapidMiner software, the optimal performance is achieved with  $K = 2$  and the smallest value of 1.052. The evaluation results show that cluster 0, comprising 38 students, represents a group with good literacy and numeracy performance, while cluster 1, comprising 23 students, represents a group with low literacy and numeracy performance.

**Keywords:** Education, Numeracy Literacy, *K-Means* Method, Davies-Bouldin Index.

## ABSTRAK

Pendidikan adalah salah satu pilar utama dalam pembangunan negara, karena berperan penting dalam membentuk generasi muda menjadi individu yang berkualitas dan mampu berkontribusi positif bagi masyarakat dan bangsa. SMP Negeri Satap Lambakara, sebagai salah satu sekolah di daerah Sumba Timur yang menghadapi kendala dalam meningkatkan kemampuan literasi numerasi siswanya. Data nilai pengetahuan dan keterampilan siswa seringkali tidak dimanfaatkan secara optimal untuk memahami tingkat literasi numerasi yang dimiliki oleh masing-masing siswa. Penelitian ini bertujuan Melakukan *clustering* untuk mengidentifikasi tingkat literasi numerasi dengan metode *k-means* yang menggunakan data nilai pengetahuan dan keterampilan serta ujian sekolah siswa di SMP Negeri Satap Lambakara. Berdasarkan dari hasil pengujian nilai *Davies Bouldin Index* dengan perhitungan menggunakan perangkat lunak RapidMiner, menghasilkan performa yaitu  $K = 2$  dengan nilai terkecil 1.052 dimana hasil evaluasi menghasilkan kluster 0 merupakan kluster dengan performa literasi numerasi siswa yang baik yaitu 38 siswa, dan kluster 1 adalah kluster performa literasi numerasi siswa yang rendah terdiri dari 23 siswa.

**Kata kunci:** Pendidikan, Literasi Numerasi, Metode *K-Means*, *Davies-Bouldin Index*.

## PENDAHULUAN

Pendidikan adalah penopang utama dalam mendirikan negara karena berperan penting dalam menciptakan generasi milenial yang unggul dan memberi peran positif bagi masyarakat dan bangsa. Di Indonesia, struktur pendidikan dasar memegang fungsi krusial dalam mengembangkan potensi peserta didik yang memberikan pengetahuan dan keterampilan dasar (Yanto et al., 2023). Salah satu mata pelajaran utama yang diajarkan di sekolah menengah adalah matematika.

Kemampuan untuk menemukan, memahami, menerapkan, dan berkomunikasi konsep matematika (angka dan simbol) untuk menyelesaikan masalah praktis dalam kehidupan sehari-hari disebut sebagai literasi numerasi. Dalam setiap aspek kehidupan, literasi numerasi sangat penting, termasuk di bidang pendidikan, pekerjaan, dan masyarakat. Karena literasi numerasi terus berkembang, manusia harus terus beradaptasi dalam berbagai aspek pendidikan. Literasi numerasi dianggap sebagai kemajuan dan dasar kemajuan pendidikan di Indonesia. Mengembangkan literasi numerasi sangat penting karena ini adalah keterampilan penting yang harus dimiliki untuk berhasil di masa depan. Memiliki kemampuan literasi numerasi yang baik memungkinkan Anda mengikuti perkembangan zaman. Literasi numerik sangat penting untuk menentukan kebijakan dan tujuan yang tepat dalam kehidupan bernegara, ekonomi, pendidikan, dan bidang lainnya. Ini karena banyak data disajikan dalam bentuk numerik atau grafik (Rahmayanti & Utama, 2022).

SMP Negeri Satap Lambakara, sebagai salah satu sekolah di daerah Sumba Timur yang menghadapi kendala dalam meningkatkan kemampuan literasi numerasi siswanya. Data nilai pengetahuan dan keterampilan siswa seringkali tidak dimanfaatkan secara optimal untuk memahami tingkat literasi numerasi yang dimiliki oleh masing-masing siswa. Pada tahun 2023, kemampuan numerasi siswa menurun, hanya 40% dari siswa mencapai kompetensi minimum. Pada tahun 2024, kemampuan mereka meningkat, 40% hingga 70% siswa mencapai kompetensi minimum, tetapi masih membutuhkan untuk mendorong lebih banyak peserta didik untuk memenuhi standar kompetensi dasar.

Untuk mengatasi masalah ini, *Education data mining* merupakan teknik yang dimanfaatkan untuk mengekstrak informasi penting dari kumpulan data yang sangat besar. Dalam konteks pendidikan, teknik ini dapat digunakan untuk menganalisis data terkait prestasi akademik siswa, mengidentifikasi pola-pola, serta menemukan hubungan-hubungan yang mungkin tidak terlihat secara langsung. Kelebihan *data mining* seperti :

1. Kemampuan mengelola dataset yang besar secara cepat dan efisien.
2. Wawasan yang dihasilkan *data mining* dapat digunakan oleh pihak sekolah untuk mengoptimalkan alokasi sumber daya, meningkatkan proses pengambilan keputusan, dan meningkatkan kualitas pendidikan secara keseluruhan.
3. Membantu merancang program pembelajaran yang dapat disesuaikan dengan kebutuhan individual.

Diperlukan strategi yang efektif untuk menganalisis dan mengelompokkan siswa berdasarkan tingkat literasi numerasi mereka. *Clustering* dengan algoritma *k-means* adalah salah satu cara yang dapat digunakan. Alasan memilih teknik pengelompokan ini karena:

1. Metode ini memungkinkan pengelompokan siswa berdasarkan kesamaan nilai pengetahuan dan keterampilan, sehingga memudahkan dalam menentukan tingkat literasi numerasi.
2. Dapat merancang program pembelajaran yang lebih sesuai dengan kebutuhan siswa, seperti memberikan intervensi tambahan bagi siswa dengan tingkat literasi numerasi rendah dan memberikan tantangan tambahan bagi siswa dengan tingkat literasi numerasi

tinggi. Melalui pendekatan ini, diharapkan dapat meningkatkan kualitas pendidikan dan kemampuan literasi numerasi siswa secara keseluruhan.

## MATERI DAN METODE

### *Data mining*

*Data mining* Data mining adalah proses pengumpulan pengetahuan tersembunyi dari berbagai data berskala besar dengan memanfaatkan matematika, statistik, kecerdasan buatan, dan pembelajaran mesin. (Br Sembiring et al., 2022). Secara sederhana, pengembangan data atau *data mining* adalah tahapan pemilihan, analisis mendalam, dan perancangan model sejumlah besar data untuk menemukan pola atau kecenderungan yang biasanya tidak disadari dalam data. Data mining melibatkan penyimpanan dan pemrosesan data secara otomatis oleh komputer menggunakan metode dan perhitungan tertentu (Setiawan, 2020).

*Data mining* mengolah dan menganalisis sejumlah besar data yang disimpan dalam repositori untuk menemukan pola, hubungan, dan tren baru. Proses ini melibatkan penggunaan teknologi pengenalan pola serta berbagai teknik statistik dan matematika untuk mengekstrak informasi penting dari data (Ardiansyah, 2019).

### *Machine learning*

Pembelajaran mesin adalah aplikasi komputer yang menggunakan data untuk membuat prediksi dan algoritma matematika. Proses pembelajaran dimaksudkan untuk memperoleh pengetahuan melalui dua tahap, yaitu latihan dan pengujian (Roihan et al., 2020).

Dengan menggunakan algoritma pembelajaran untuk memadukan dan menangani sejumlah besar data, *machine learning* adalah kumpulan teknik yang dapat membantu dalam menangani dan memprediksi sejumlah besar data (Yuniasari & Maspiyanti, 2021).

### **Penilaian pengetahuan**

Dilakukan secara kuantitatif, penilaian pengetahuan menilai pencapaian kompetensi dasar dengan rentang nilai 0 hingga 100. Kemudian, nilai-nilai ini dianalisis untuk menentukan predikat dan deskripsi pencapaian hasil belajar peserta didik. Kurikulum 2013 mengubah pendekatan di mana predikat digunakan sebagai alat untuk mengukur ketuntasan belajar siswa. Siswa dengan predikat D dinyatakan belum tuntas dan diharuskan mengikuti Kriteria ketuntasan minimal (KKM) yang telah ditetapkan. Ini berbeda dari kurikulum sebelumnya, yang secara konsisten mencantumkan predikat (Khusnuddin, 2018).

Guru melakukan penilaian aspek pengetahuan dengan berbagai metode selama satu semester. Hasilnya direkap dan dicatat dalam penilaian berdasarkan kompetensi dasar untuk tema yang dinilai (Khusnuddin, 2018).

### **Penilaian keterampilan**

Evaluasi kinerja, asesmen proyek, dan teknik portofolio digunakan untuk menilai keterampilan. Pengelolaan penilaian ini diolah secara kuantitatif dengan menggunakan skala 0–100, dan deskripsi level kemampuan peserta didik disertakan. Kalimat positif digunakan untuk menjelaskan capaian kemampuan siswa dalam setiap materi pembelajaran, yang mengacu pada kompetensi dasar masing-masing pelajaran (Khusnuddin, 2018).

Nilai optimum, rata-rata, atau keduanya dapat digunakan untuk menunjukkan hasil penilaian keterampilan. Nilai akhir penilaian keterampilan diperoleh dengan menggabungkan nilai rata-rata dari semua nilai keterampilan (KD). Nilai optimum diberlakukan hanya jika KD berdasarkan sumber pembelajaran dan strategi yang mirip dan jika KD dievaluasi lebih dari satu kali (Khusnuddin, 2018).

## **Clustering**

*Clustering* adalah proses pemisahan sejumlah besar objek data menjadi kelompok terpisah yang dikenal sebagai kluster. Sebelum koefisien siluet dapat dihitung, setiap objek dalam *cluster* memiliki karakteristik tertentu (Ningrum et al., 2023).

Metode *clustering* membagi data menjadi beberapa kelompok yang memiliki karakteristik yang mirip atau serupa. Jumlah pengelompokan yang dapat diidentifikasi bergantung pada variasi dan jumlah objek data yang digunakan. Tujuan dari pengelompokan sekumpulan objek data menjadi beberapa kelompok yang memiliki karakteristik yang mirip satu sama lain ditujukan untuk memungkinkan analisis dan interpretasi lebih lanjut sesuai dengan tujuan penelitian (Arofah & Marisa, 2018).

### **Algoritma K-Means**

*K-Means* adalah sebuah algoritma yang termasuk dalam kategori pembelajaran tanpa pengawasan dan mengklasifikasi data ke dalam beberapa *cluster*. Salah satu kelebihan dari algoritma ini adalah kemampuan untuk mengelola data yang tidak memiliki label kategori sebelumnya, dengan metode ini, data dapat dikelompokkan berdasarkan kemiripan fitur atau atribut tertentu tanpa memerlukan label atau pengawasan. (Ranjawali et al., 2023).

Algoritma *K-Means* adalah teknik analisis *cluster* di mana N observasi dibagi menjadi beberapa kelompok berdasarkan seberapa dekat mereka dengan rata-rata (*mean*) terdekat. Algoritma ini mudah digunakan, kinerjanya relatif cepat, dapat disesuaikan, dan sering digunakan. Algoritma *K-Means* mengklasifikasikan data berdasarkan jarak data tersebut dengan pusat massa kelompok (Ningrum et al., 2023).

Salah satu metode klasterisasi non-hirarkis, algoritma *K-Means* membagi data berdasarkan titik pusat *cluster* (*centroid*) terdekat. Metode ini membagi data menjadi beberapa kelompok berdasarkan kesamaan atribut, sementara data dengan karakteristik yang berbeda ditempatkan dalam *cluster* yang berbeda (Arofah & Marisa, 2018).

### **Davies-Bouldin Index (DBI)**

Mengukur kinerja algoritma *clustering*, *Davies-Bouldin Index* (DBI) pertama kali diperkenalkan pada tahun 1979 oleh David L. Davies dan Donald W. Bouldin. DBI menetapkan bahwa skema *clustering* optimal memiliki nilai DBI minimal (Butsianto & Saepudin, 2020).

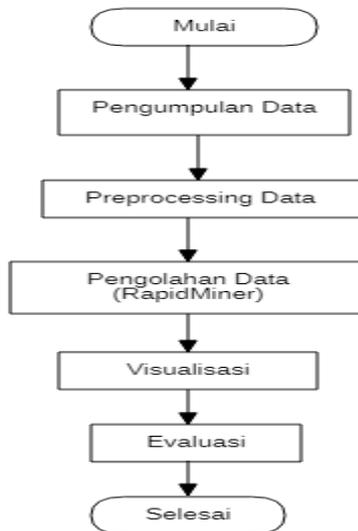
Nilai rata-rata setiap elemen dalam kumpulan data didapat dari *indeks Davies-Bouldin* (DBI), yang dihitung dengan membagi jumlah nilai kekompakan (*compactness*) yang dibagi dengan jarak antara dua pusat *cluster* (*separation*). Persamaan DBI digunakan untuk meminimalkan jarak antara *cluster* dengan persamaan (Maulana & Rosalina, 2020).

### **RapidMiner**

*RapidMiner* adalah perangkat lunak sumber terbuka. *RapidMiner* adalah perangkat lunak untuk pengkajian *data mining*, *text mining*, dan analisis prediktif. Berbagai teknik pemaparan dan peramalan memungkinkan pemakai membuat solusi terbaik. Dengan kurang lebih 500 operator *data mining*, *RapidMiner* adalah alat visualisasi data mandiri dan dapat digunakan sebagai mesin data mining yang diintegrasikan ke dalam produk lain. *RapidMiner*, yang dibuat dalam bahasa *Java*, dapat dioperasikan ke semua sistem operasi. (Ardiansyah, 2019).

Analisis teks, ekstraksi pola dari dataset besar, dan penggabungan pola tersebut dengan metode statistik, kecerdasan buatan, dan basis data adalah beberapa fungsi perangkat lunak *text mining RapidMiner* (Setiawan, 2020).

Berikut gambaran tahapan-tahapan kerangka berpikir yang direncanakan.



Gambar 1. Alur Penelitian

Langkah pertama, data yang akan digunakan dalam penelitian dikumpulkan melalui metode penelitian kualitatif. Hasil observasi dan wawancara, jurnal, dan penelitian sebelumnya adalah fokus penelitian ini. Data dikumpulkan melalui kunjungan atau pengamatan langsung di lokasi penelitian, yaitu SMP Negeri Satap Lambakara. dengan format data yaitu data Excel. Penelitian ini menggunakan data dari 61 siswa kelas IX pada tahun akademik 2022/2023.

Pada tahapan ini melibatkan *data mining* untuk persiapan dan pembersihan data mentah sebelum dilakukan analisis selanjutnya. Proses ini mencakup penghapusan data yang tidak relevan atau duplikat, pengelolaan nilai yang hilang, transformasi data ke format yang konsisten, serta proses normalisasi data dan pemilihan atribut yang akan digunakan. Dalam praprocessing data adapun data yang menggunakan tipe data binominal sebagai penanda dalam variabel.

Pada tahap ini, hasil normalisasi *dataset* akan digunakan untuk mengelompokkan data dengan menggunakan *rapidminer* dengan menggunakan operator *K-means*, yang hasil pengelompokkan datanya akan dianalisis.

Pada langkah ini, akan dilakukan analisis dari hasil pengelompokan dengan tujuan untuk mencari pola, data, hubungan antara variabel, karakteristik data pada setiap *cluster*, serta wawasan yang akan digunakan untuk mengukur tingkat literasi matematika pada penelitian ini. Hasil pengelompokan akan dianalisis menggunakan tabel perbandingan atau *aggreate value*, untuk mengukur nilai rata-rata dari setiap atribut pada masing-masing *cluster*. Setelah dianalisis, hasil analisis akan divisualisasikan menggunakan *bar plot* sehingga mempermudah proses interpretasi hasil.

Pada tahap akhir, peneliti akan mengevaluasi metode clustering menggunakan metode Davies-Bouldin Index (DBI). Ini akan dilakukan untuk mengetahui seberapa baik metode *K-means* yang digunakan.

## HASIL DAN PEMBAHASAN

### Pengumpulan Data

Penelitian ini dilaksanakan di SMP Negeri Satap Lambakara yang berada di Kecamatan

Pahunga Lodu, Kabupaten Sumba Timur, dengan menjadikan siswa kelas IX sebanyak 61 peserta didik pada tahun ajaran 2022/2023 sebagai sampel. Variabel *dataset* yang digunakan yaitu merupakan nilai hasil ulangan pengetahuan dan ketrampilan pada semester 1 dan semester 2 serta nilai hasil ujian sekolah. Data yang telah dikumpulkan kemudian dianalisis untuk mengamati kemampuan literasi numerasi yang dimiliki peserta didik, dari hasil analisis tersebut akan dikelompokkan sehingga siswa yang memiliki tingkat literasi numerasi tinggi akan di biarkan belajar secara mandiri, sedangkan siswa dengan tingkat literasi numerasi rendah akan mendapatkan penanganan khusus.

### Preprocessing Data

Data yang dikumpulkan akan dipraproses sebelum menerapkan algoritma *k-means* untuk memastikan bahwa data yang digunakan sesuai untuk dilakukan penganalisisan. Dari 70 data siswa setelah dilakukan pembersihan data sehingga mendapatkan 61 data siswa yang sesuai. Berikut *dataset* yang siap untuk diolah menggunakan aplikasi *RapidMiner*.

Tabel 1. *Dataset*

No.	Nama siswa	jenis kelamin	nilai matematika 2022-2023				Nilai Ujian Sekolah
			semester 1		semester 2		
			p	k	p	k	
1	Siswa 1	0	75	75	71	70	73
2	Siswa 2	1	82	82	83	81	72
3	Siswa 3	0	73	73	77	77	78
4	Siswa 4	1	80	80	73	74	70
5	Siswa 5	1	76	76	83	78	78
...	.....	....	..	..	..	..	.....
61	Siswa 61	1	73	75	79	78	78

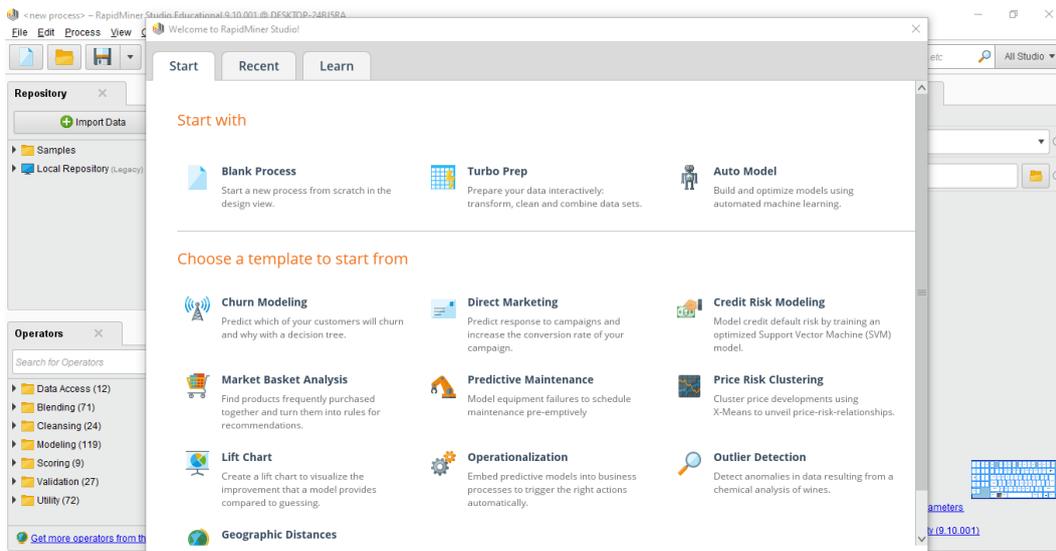
Berdasarkan dataset yang telah dikumpulkan, maka data yang digunakan adalah data dengan atribut nama siswa, jenis kelamin, nilai ulangan pengetahuan dan keterampilan pada semester 1 dan semester 2, serta nilai ujian sekolah.

### Pengolahan Data Menggunakan *Tools RapidMiner*

Data yang telah diperoleh dari proses pengecekan akan digunakan sebagai sebagai inputan untuk membangun model algoritma *K-Means Clustering* menggunakan perangkat lunak *RapidMiner*.

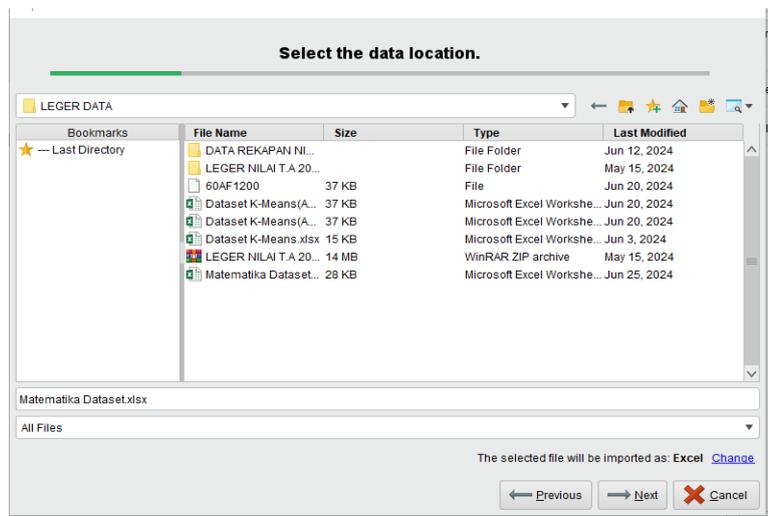
### Tampilan Menu Utama *RapidMiner*

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



Gambar 2. Tampilan Menu awal *RapidMiner* Versi 9.10

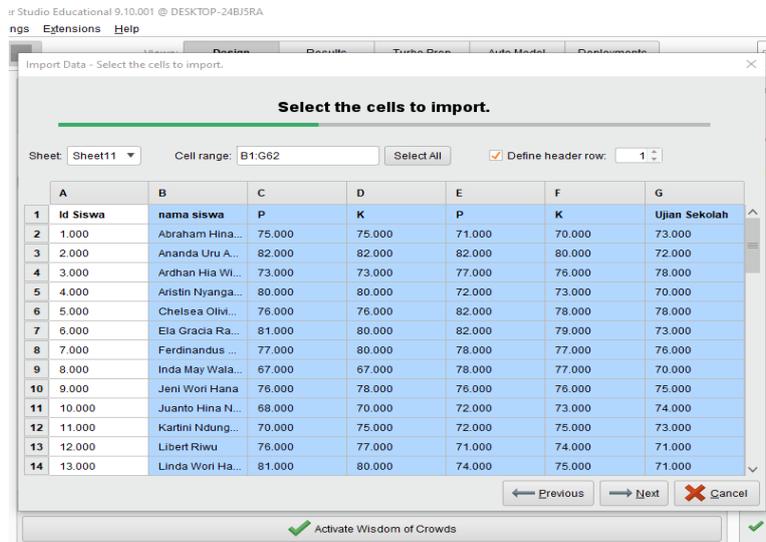
Data yang telah dipersiapkan harus diimport kedalam *RapidMiner*. Data yang di import dalam bentuk file *Excel*. Dapat dilihat pada gambar 3.



Gambar 3. Tampilan menu *import* data

### Tampilan Select The Cells To Import

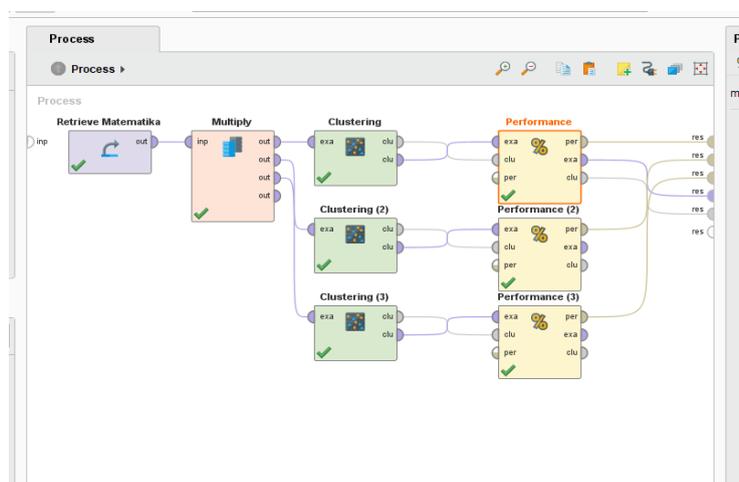
Sistem memberikan instruksi tentang bagaimana cara mengimpor data baru yang akan diolah selanjutnya, dengan format data dalam bentuk *excel*. Tahapan ini dapat diamati dalam ilustrasi berikut.



Gambar 4. Select The Cells To Import

### Tampilan Clustering RapidMiner

Dalam tahapan ini Algoritma *K-Means* akan diterapkan untuk melakukan proses klusterisasi. Operator yang digunakan adalah *K-Means*. Setelah pembuatan model selesai, proses akan dilanjutkan untuk mendapatkan hasil pengelompokan. Langkah ini dapat dilihat dalam ilustrasi berikut.

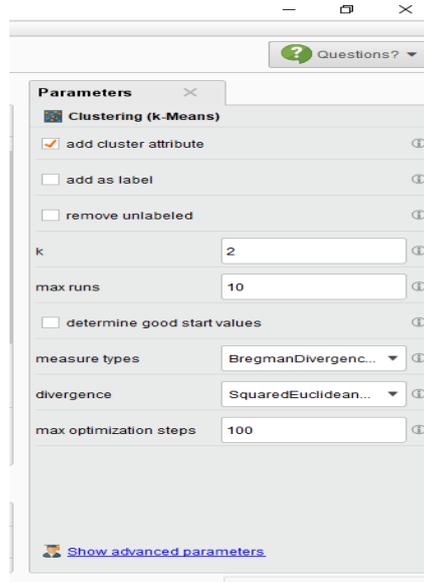


Gambar 5. Clustering RapidMiner

Dari gambar diatas, menunjukkan beberapa operator seperti *Read Excel* dimana *dataset* akan dimasukkan dalam *Read Excel*, *Multiplay* yang berfungsi untuk menjalankan 1 atau lebih operator, operator *cluster* yaitu *k-means*, kemudian performa untuk evaluasi kinerja dari *k-means* yaitu *cluster distance performance*.

### Penentuan Jumlah Klaster

Pengaturan algoritma *K-Means* selanjutnya akan dilakukan melalui menu "Parameter *Clustering K-Means*". Tampilan berikut menunjukkan penentuan jumlah parameter cluster..



Gambar 6. Paramete Jumlah *cluster*

Berdasarkan gambar 6, perlu memasukkan nilai K yang diinginkan. Dalam 10 kali iterasi yang dilakukan, nilai K yang dimasukkan adalah K = 2, 3, dan 4. Tujuannya adalah untuk menemukan *cluster* dengan nilai DBI yang mendekati 0 atau yang paling rendah, maka dianggap sebagai *cluster* terbaik. Setelah proses iterasi selesai, proses *clustering* akan dilakukan agar mendapatkan hasil dari algoritma *K-Means* yang diterapkan. Dari hasil iterasi tersebut, maka diperoleh nilai optimal untuk jumlah *cluster* terbaik adalah K = 2.

### Interpretasi Hasil

Analisis *cluster K-Means* digunakan untuk mengelompokkan siswa berdasarkan metrik literasi numerasi, seperti jenis kelamin, nilai ulangan pengetahuan dan ketrampilan semester pertama dan kedua, serta nilai ujian sekolah. Data telah diubah menjadi dua *cluster*, *cluster* 0 dan *cluster* 1 setelah proses *clustering* selesai. Hasilnya akan ditampilkan dalam tabel dengan operasi variabel berikut :

JK = Jenis Kelamin

P1 = Nilai Pengatahuan semester 1

P2 = nilai pengetahuan semester 2

K1 = nilai keterampilan semester

K2 = nilai keterampilan semester 2

Tabel 2. Anggota Hasil pengelompokkan

No	Nama Siswa	Cluster	JK	P1	K1	P2	K2	Ujian Sekolah
1	Siswa 1	<i>cluster</i> _1	0	75	75	71	70	73
2	Siswa 2	<i>cluster</i> _0	1	82	82	82	80	72
3	Siswa 3	<i>cluster</i> _1	0	73	73	77	76	78
4	Siswa 4	<i>cluster</i> _1	1	80	80	72	73	70
5	Siswa 5	<i>cluster</i> _0	1	76	76	82	78	78
...	.....	.....	....	....	....	....	....	.....
60	Siswa 60	<i>cluster</i> _0	1	76	78	75	75	80
61	Siswa 61	<i>cluster</i> _0	1	73	75	79	78	78

### Aggregate Value Clustering

Setelah pengelompokan data selesai, langkah selanjutnya adalah menentukan nilai min, nilai max, nilai *mean*, dan standar deviasi untuk setiap *cluster* dan atribut. Tujuannya adalah untuk menemukan fitur yang membedakan masing-masing *cluster* dari yang lain. Tabel ini menunjukkan hasil perhitungannya:

Tabel 3. *Aggregate Value*

Cluster 0					
Aggregate Value	P1	K1	P2	K2	Ujian Sekolah
Min	74	75	73	74	70
Max	85	82	85	82	88
Mean	79,39	79,39	79,42	78,42	76,42
Standar Deviasi	2,814436	1,619801	3,243551	2,176429	4,547833
Cluster 1					
Aggregate Value	P1	K1	P2	K2	Ujian Sekolah
Min	67	67	70	70	70
Max	77	78	79	78	80
Mean	72,21	73,34	73,60	75,52	74,34
Standar Deviasi	2,999341	3,269816	2,996045	2,086092	2,740416

Tabel 3 menunjukkan nilai perbandingan dari cluster 0 dan cluster 1, di mana masing-masing variabel memiliki nilai minimum, maksimum, rata-rata, dan standar deviasi yang berbeda. Nilai ujian sekolah di cluster 0 lebih tinggi dibandingkan dengan nilai ujian sekolah di cluster 1. Dapat dilihat pada nilai standar deviasi ujian sekolah yaitu 4,547833 yang menjauhi nol, karena nilainya lebih bervariasi.

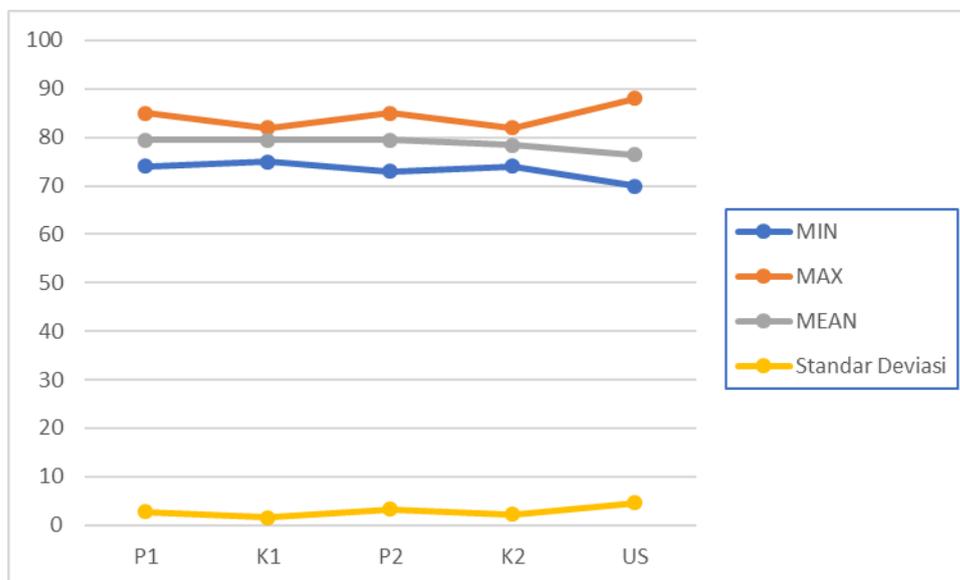
### Visualisasi

Pengujian data menggunakan *RapidMiner 9.10* menghasilkan beberapa *output*. Dalam gambar visualisasi, terdapat *line* atau garis yang menunjukkan jumlah pengelompokan untuk kasus performa peserta didik dalam mata pelajaran matematika di SMP Negeri Satap Lambakara. Visualisasi tersebut menampilkan sumbu x, sumbu y, dan warna khusus untuk *cluster*.

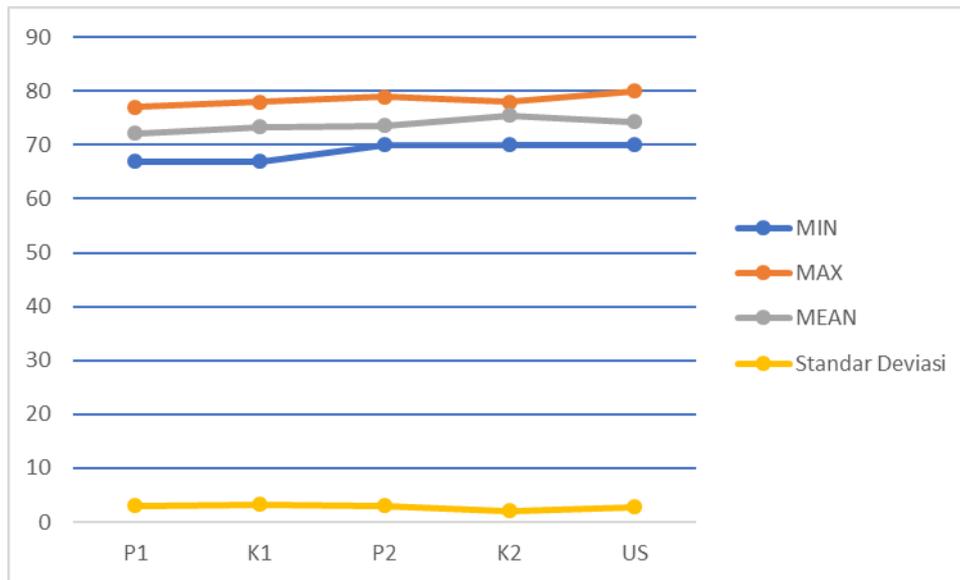


Gambar 7. Scatter/Bubble Cluster (x) dan Jenis kelamin (y)

Pada gambar 7, menggambarkan analisis pengelompokan berdasarkan *cluster* dan jenis kelamin dengan menggunakan *scatter/bubble plot* menunjukkan bahwa *cluster 0* lebih didominasi oleh siswi perempuan, sebanyak 31 orang yang ditandai dengan warna hijau, sementara *cluster 1* didominasi oleh 31 siswa laki-laki yang ditandai dengan warna biru.



Gambar 8. Grafik Line plot Aggregate Value cluster 0



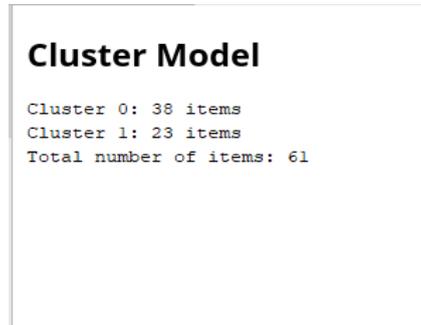
Gambar 9. Grafik Line plot Aggregate Value cluster 1

Pada gambar 8 dan 9, menampilkan line plot hasil *cluster* yang diperoleh dari tabel perbandingan. Yang berwarna biru ditandai sebagai nilai minimum, warna *orange* ditandai sebagai nilai maksimum, warna abu-abu ditandai sebagai nilai rata-rata dan warna kuning ditandai sebagai nilai standar deviasi. dengan operasi variabel sebagai berikut : P1 = Nilai Pengetahuan semester 1, P2 = nilai pengetahuan semester 2, K1 = nilai keterampilan semester 1, K2 = nilai keterampilan semester 2, dan US = nilai ujian sekolah. Nilai setiap atribut pada *cluster* 0 lebih tinggi dibandingkan dengan *cluster* 1, dilihat dari nilai minimum pengetahuan semester 1 yaitu 74, nilai maksimumnya adalah 85, nilai rata-rata yaitu 79,39, dan nilai standar deviasi adalah 2,814436468, sedangkan nilai minimum keterampilan yaitu 75, nilai maksimum adalah 82, nilai rata-rata yaitu 79,39, dan nilai standar deviasinya adalah 1,619801017 nilai standar deviasinya rendah karena nilai dari variabel tidak bervariasi. Nilai minimum pengetahuan semester 2 yaitu 73, maksimum adalah 85, nilai rata-rata adalah 79,49, dan nilai standar deviasi adalah 3,243551432. Sedangkan nilai minimum keterampilan yaitu 74, maksimum adalah 82, nilai rata-rata adalah 78,42, dan nilai standar deviasi yaitu 2,17642875. Nilai minimum ujian sekolah yaitu 70, maksimum yaitu 88, nilai *mean* adalah 76,42, dan nilai standar deviasi adalah 4,547833336, yang menjauhi nol karena nilai pada variabel ini bervariasi. Sedangkan pada *cluster* 1, nilai minimum pengetahuan semester 1 yaitu 67, nilai maksimumnya adalah 77, nilai rata-rata yaitu 72,22, dan nilai standar deviasi adalah 2,999341166, sedangkan nilai minimum keterampilan yaitu 67, nilai maksimum adalah 78, nilai rata-rata yaitu 73,35 dan nilai standar deviasinya adalah 3,269816448 nilai standar deviasinya tinggi karena nilai dari variabel bervariasi. Nilai minimum pengetahuan semester 2 yaitu 70, maksimum adalah 79, nilai rata-rata adalah 73,61, dan nilai standar deviasi adalah 2,996044824. Sedangkan nilai minimum keterampilan yaitu 70, maksimum adalah 78, nilai rata-rata adalah 75,45, dan nilai standar deviasi yaitu 2,086091718. Nilai minimum ujian sekolah yaitu 70, maksimum adalah 80, nilai rata-rata adalah 74,35, serta nilai standar deviasi adalah 2,740416286. Dapat disimpulkan bahwa nilai standar deviasi tertinggi didominasi oleh *cluster* 0, menunjukkan variasi nilai yang lebih tinggi dalam *cluster* ini. Secara keseluruhan, *cluster* 0 memiliki nilai yang lebih baik dibandingkan *cluster* 1.

### Cluster Model

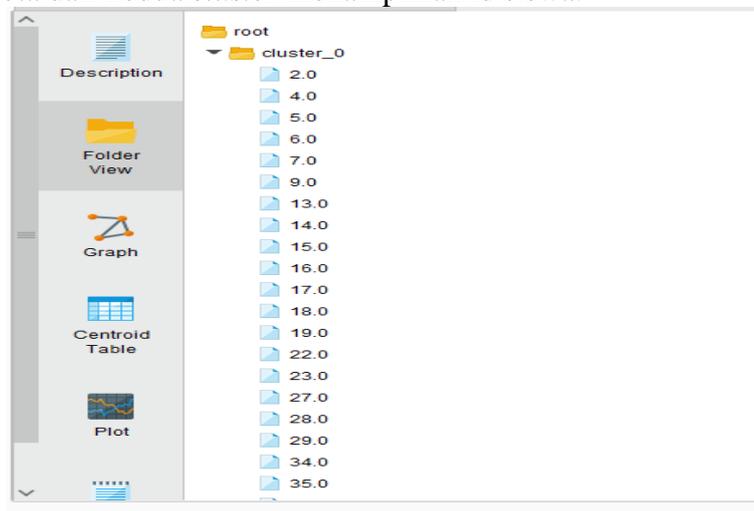
Dalam model *cluster*, terdapat beberapa tampilan yang menampilkan hasil *cluster*, salah satunya yaitu tampilan teks (*view text*) yang menunjukkan hasil pengelompokan berdasarkan klaster dan jumlah anggota dalam setiap klaster. *cluster* 0 memiliki 38 item anggota, sedangkan

kluster 1 memiliki 23 item anggota. Tampilan teksnya dapat dilihat pada gambar.



Gambar 10. Cluster Model

Pada gambar, *folder view* menunjukkan data secara komprehensif dari bagian-bagian *cluster*, dengan setiap anggota dari kedua *cluster* menampilkan id siswa.



Gambar 11. Folder View

### Evaluasi

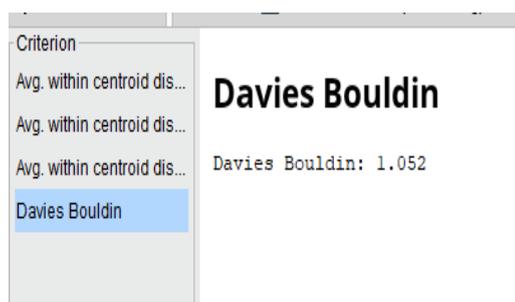
Berdasarkan tabel 4, dengan menggunakan metode *k-means clustering* dan didukung oleh perhitungan *Davies-Bouldin Index* (DBI) untuk menentukan jumlah *cluster* optimum, dilakukan 3 iterasi percobaan untuk menentukan *cluster* yang paling efektif berdasarkan nilai DBI terbaik.

Tabel 4. Nilai *Davies-Bouldin Index* (DBI)

K	Avg. Within centroid distance	DBI	Terpilih
2	42.812	1.052	K = 2
3	32.831	1.136	
4	26.467	1.118	

Penentuan banyaknya *cluster* optimum menggunakan *Davies-Bouldin Index* (DBI), didasarkan pada besar kecilnya nilai yang diperoleh. Berdasarkan tabel 4 diketahui hasil terbaik dari DBI ditemukan ketika *cluster* yang terbentuk sebanyak dua, dengan nilai DBI terkecil 1.052. Ini menunjukkan Pengelompokan meningkat dengan nilai DBI yang dihasilkan. DBI digunakan

untuk mengukur kualitas dari pengelompokan data, dan nilai yang lebih kecil menandakan bahwa pembagian data ke dalam kelompok-kelompok tersebut lebih baik dan lebih terpisah satu sama lain. Nilai DBI yang dihasilkan sebesar 1.052, dikarenakan banyaknya data sampel yang digunakan tidak bervariasi atau cenderung sama pada beberapa atribut sehingga mengakibatkan terjadinya *nois*. Dengan demikian, DBI adalah indikator penting dalam mengevaluasi keberhasilan dalam proses pengelompokan data. Terlihat pada ilustrasi ke-12.



Gambar 12. Nilai DBI K 2

## KESIMPULAN

Dengan melihat hasil klusterisasi hasil belajar matematika siswa-siswi kelas IX di SMP Negeri Satap Lambakara dengan algoritma K-Means Clustering, dapat disimpulkan bahwa penggunaan data mining berdasarkan nilai pengetahuan dan keterampilan siswa pada semester pertama dan kedua, serta nilai ujian sekolah, terbukti efektif. Hasil pengelompokan menunjukkan bahwa siswa dapat dibagi menjadi dua kelompok dengan nilai karakter yang sebanding, sehingga lebih mudah bagi sekolah untuk menganalisis kebutuhan dan potensi masing-masing peserta didik. Hasil pengujian juga menunjukkan bahwa nilai *mean* setiap variabel pada kelompok 0 lebih unggul dibandingkan dengan kelompok 1, dengan nilai rata-rata sekitar 79 untuk kelompok 0 dan sekitar 72 untuk kelompok 1. Selain itu, dengan jumlah cluster 2 dan nilai optimal terkecil 1.052, hasil analisis menggunakan Davies-Bouldin Index (DBI) menunjukkan kekuatan cluster terbaik. Dalam penelitian ini, pengelompokan K-Means berhasil memberikan gambaran yang lebih mendalam tentang bagaimana siswa berhasil memahami literasi numerasi. Hal ini dapat membantu lembaga pendidikan memperoleh pemahaman yang lebih baik tentang bagaimana setiap siswa berprestasi dan membantu mereka membuat pendekatan pembelajaran yang lebih adaptif dan efektif.

## SARAN

Berdasarkan hasil penelitian, untuk pengembangan lebih lanjut mencakup beberapa hal. Pertama mempertimbangkan penggunaan data dengan nilai yang bervariasi. Kedua memodifikasi atau menggabungkan algoritma *K-Means* dengan algoritma lain guna meningkatkan kinerja klasifikasi *cluster*. Selain itu menambah jumlah data yang digunakan, metode *K-Means* biasanya menghasilkan pengelompokan yang lebih optimal jika diterapkan pada kumpulan data yang lebih besar. Oleh karena itu, disarankan agar memperluas kumpulan data dalam penelitian.

## DAFTAR PUSTAKA

- Ardiansyah, D. (2019). Algoritma C4.5 Untuk Klasifikasi Calon Peserta Lomba Cerdas Cermat Siswa Smp Dengan Menggunakan Aplikasi Rapid Miner. *Jurnal Inkofar*, 1(2), 5–12. <https://doi.org/10.46846/jurnalinkofar.v1i2.29>
- Arofah, S. N., & Marisa, F. (2018). Penerapan Data Mining untuk Mengetahui Minat Siswa pada Pelajaran Matematika menggunakan Metode K-Means Clustering. *JOINTECS (Journal of Information Technology and Computer Science)*, 3(2), 85–90. <https://doi.org/10.31328/jointecs.v3i2.787>
- Br Sembiring, S. N., Winata, H., & Kusnasari, S. (2022). Pengelompokan Prestasi Siswa Menggunakan Algoritma K-Means. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 1(1), 31. <https://doi.org/10.53513/jursi.v1i1.4784>
- Butsianto, S., & Saepudin, N. (2020). Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 3(1), 51–59. <https://doi.org/10.32672/jnkti.v3i1.2008>
- Khusnuddin, K. (2018). Model Spreadsheet Excel Application sebagai Pengolahan Hasil Belajar Peserta Didik dalam Penilaian Kurikulum 2013. *Jurnal Kependidikan*, 6(1), 33–52. <https://doi.org/10.24090/jk.v6i1.1695>
- Maulana, I., & Rosalina, U. (2020). Clustering Data Nilai Ujian Akhir Semester Menggunakan Algoritma Data Mining K-Means. *PERISKOP : Jurnal Sains Dan Ilmu Pendidikan*, 1(2), 76–85. <https://doi.org/10.58660/periskop.v1i2.10>
- Ningrum, K. K., Maulindar, J., & Farida, A. (2023). Penerapan Algoritma K-Means Untuk Pengelompokan Penilaian Akhir Semester Di Sdn Kadokan 01 Sukoharjo. *INFOTECH Journal*, 9(1), 190–197. <https://doi.org/10.31949/infotech.v9i1.5343>
- Rahmayanti, D. Z., & Utama. (2022). Pembudayaan Literasi Numerasi dalam Kegiatan Inti Pembelajaran Matematika di Sekolah Menengah Pertama. *Jurnal Riset Pembelajaran Matematika Sekolah*, 6(2), 19–24. <https://doi.org/10.21009/jrpms.062.03>
- Ranjawali, R., Talakua, A. C., & Abineno, R. T. (2023). Clustering Stunting Pada Balita Dengan Metode K- Means Di Puskesmas Katanang. *SATI: Sustainable Agricultural Technology Innovation*, 80–92. <https://ojs.unkriswina.ac.id/index.php/semnas-FST/article/view/587/324>
- Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, 5(1), 75–82. <https://doi.org/10.31294/ijcit.v5i1.7951>
- Setiawan, A. (2020). *Penerapan algoritma c.45 untuk klasifikasi tingkat kenikan kelas di sdn citamiang 2 kota sukabumi*. [https://repository.bsi.ac.id/index.php/unduh/item/302730/19162343\\_Aris-Setiawan-new.pdf](https://repository.bsi.ac.id/index.php/unduh/item/302730/19162343_Aris-Setiawan-new.pdf)
- Yanto, D., Susanto, H., Zulkifli, K., & Gupron, F. R. (2023). Penerapan Algoritma K-Means Dalam Menentukan Kualitas Satuan Pendidikan Berdasarkan Nilai Internal Dan Eksternal. *JEEDCOM Journal of Electrical Engineering and Computer*, 5(2), 319–328. <https://doi.org/10.33650/jeecom.v5i2.6946>
- Yuniasari, P., & Maspiyanti, F. (2021). Analisis Sentimen Data Tweet Menggunakan Metode Support Vector Machine (Studi Kasus: Pindahan Ibukota Baru Republik Indonesia). *Journal of Informatics and Advanced Computing*, 2(1), 1–9. <http://journal.univpancasila.ac.id/index.php/jiac/article/view/2684>