

Penerapan Metode Clustering K-Means Menggunakan RapidMiner untuk Klasifikasi Prestasi Siswa di Sekolah Swasta

¹Eko Andri Wibowo, ²Ririn Aryanti

¹Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

²Sistem Informasi, Ilmu Komputer, Universitas Pamulang, Kota Tangerang Selatan, Indonesia

[¹Ekoandri2004@gmail.com](mailto:Ekoandri2004@gmail.com), [²Ririnaryanti435@gmail.com](mailto:Ririnaryanti435@gmail.com)

Abstract

This study discusses the application of the K-Means Clustering algorithm in grouping the level of academic achievement of students in private schools with the help of RapidMiner software. The data analyzed include assignment scores, midterm exams, final exams, and attendance. The K-Means algorithm was chosen because of its ability to group unlabeled numeric data and recognize hidden patterns in the dataset. The analysis was carried out on data from 5,000 students obtained through the Kaggle platform. The clustering results produced two main groups, namely students with high academic achievement and students with lower achievements. This process allows schools to understand the characteristics of each group of students and develop more effective coaching strategies and educational policies. The use of RapidMiner has been proven to help the data analysis process efficiently and intuitively, without the need for advanced programming skills.

Keywords: K-Means Clustering, RapidMiner, Academic Achievement, Student Clustering, Educational Strategy.

Abstrak

Penelitian ini membahas penerapan algoritma *K-Means Clustering* dalam mengelompokkan tingkat prestasi akademik siswa di sekolah swasta dengan bantuan perangkat lunak *RapidMiner*. Data yang dianalisis mencakup nilai tugas, ujian tengah semester, ujian akhir semester, dan kehadiran. Algoritma *K-Means* dipilih karena kemampuannya mengelompokkan data numerik tanpa label serta mengenali pola tersembunyi dalam *dataset*. Analisis dilakukan terhadap data dari 5.000 siswa yang diperoleh melalui platform *Kaggle*. Hasil klasterisasi menghasilkan dua kelompok utama, yaitu siswa dengan prestasi akademik tinggi dan siswa dengan capaian yang lebih rendah. Proses ini memungkinkan pihak sekolah memahami karakteristik tiap kelompok siswa dan menyusun strategi pembinaan serta kebijakan pendidikan yang lebih efektif. Penggunaan *RapidMiner* terbukti membantu proses analisis data secara efisien dan intuitif, tanpa perlu keahlian pemrograman lanjutan.

Kata kunci: *K-Means Clustering*, *Rapidminer*, Prestasi Akademik, Pengelompokan Siswa, Strategi Pendidikan.

A. PENDAHULUAN

Dalam upaya meningkatkan mutu pendidikan di sekolah swasta, pemantauan prestasi siswa menjadi hal yang krusial. Prestasi akademik tidak hanya mencerminkan keberhasilan individu, tetapi juga menunjukkan kualitas pembelajaran secara keseluruhan. Untuk itu, dibutuhkan suatu metode analisis yang mampu mengklasifikasikan tingkat pencapaian siswa secara terstruktur dan berbasis data. Metode clustering telah terbukti efektif dalam klasifikasi data akademik siswa (Ardiansyah et al., 2021).

Adapun pendekatan yang digunakan yaitu metode *clustering*, khususnya algoritma *K-Means*. Teknik ini memungkinkan pengelompokan siswa berdasarkan kemiripan nilai akademik, seperti hasil tugas, UTS dan UAS. Dengan dukungan perangkat lunak analisis data

seperti *RapidMiner*, proses ini dapat dilakukan dengan lebih efisien dan akurat.

Melalui analisis ini, pihak sekolah dapat memahami pola pencapaian siswa secara menyeluruh dan menetapkan strategi pembelajaran yang sesuai dengan kebutuhan masing-masing kelompok. Selain itu, informasi yang diperoleh juga dapat dimanfaatkan untuk menyusun program peningkatan akademik dan memberikan intervensi yang tepat bagi siswa yang memerlukan bantuan tambahan. Klasifikasi prestasi belajar sangat membantu dalam penentuan strategi intervensi siswa (Handayani et al., 2021).

B. PELAKSAAAN DAN METODE

Penelitian dengan judul "Penerapan Metode *Clustering K-Means* Menggunakan *RapidMiner* untuk Klasifikasi

Prestasi Siswa di Sekolah Swasta" menggunakan metode deskriptif analitis. Metode ini digunakan untuk menggambarkan kondisi prestasi siswa berdasarkan data yang telah dikumpulkan, tanpa bermaksud menyimpulkan hasilnya untuk semua siswa. Fokus penelitian terletak pada situasi nyata saat data dikumpulkan, di mana siswa dikelompokkan menurut tingkat prestasinya menggunakan algoritma *K-Means Clustering*. Data yang telah diproses melalui aplikasi *RapidMiner* kemudian dianalisis untuk memperoleh simpulan yang sesuai dengan tujuan studi.

Penelitian ini menggunakan data yang diperoleh dari situs Kaggle, tepatnya dari dataset yang tersedia pada tautan berikut: [Students Grading Dataset – Kaggle](#). Dataset ini berisi informasi akademik siswa yang akan dimanfaatkan untuk proses pengelompokan tingkat prestasi menggunakan algoritma *K-Means Clustering*.

Beberapa atribut utama yang digunakan dalam analisis dijabarkan pada tabel berikut:

Tabel 1. Atribut Utama

No.	Atribut	Deskripsi
1	Nilai Tugas	Merupakan nilai yang diperoleh siswa sebagai bentuk evaluasi atas hasil pekerjaan atau tugas yang telah dikerjakan.
2	Nilai UTS	Merupakan nilai yang diperoleh siswa pada ujian tengah semester.
3	Nilai UAS	Merupakan nilai yang diperoleh siswa pada ujian akhir semester.

Pemilihan atribut tersebut dilakukan karena dianggap mewakili faktor utama dalam penilaian capaian akademik siswa secara menyeluruh. Adapun perangkat yang digunakan untuk penelitian ini, yaitu laptop Lenovo V14-ADA dengan spesifikasi sebagai berikut:

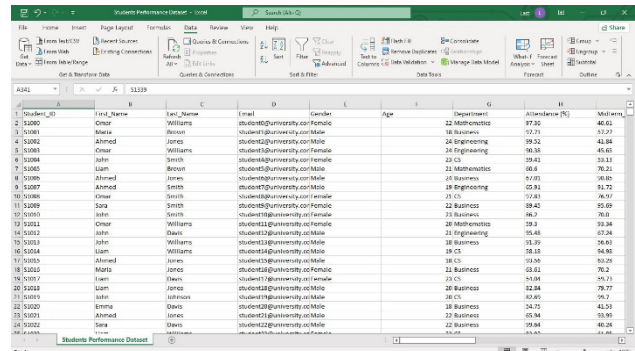
Tabel 2. Spesifikasi Perangkat Keras

Jenis	Spesifikasi	Deskripsi
Laptop	Processor	AMD 3020e with Radeon Graphics
	Memory RAM	DDR4 4 GB
	VGA	AMD Radeon Graphics (integrated)
	System Type	64-bit Operating System
	Hardisk	256 Megabyte
	Display	LED 14.0 inci HD 1366 x 768 pixel
	OS	Windows 10 Home Single Language

C. HASIL DAN PEMBAHASAN

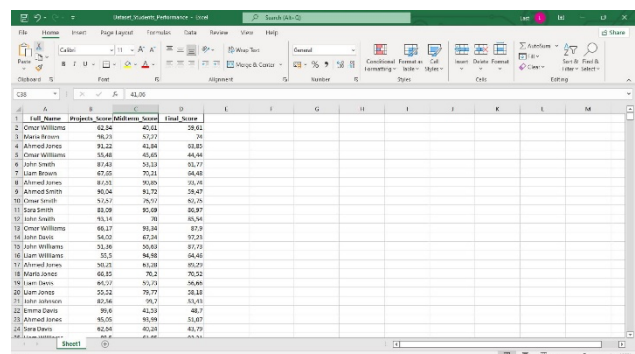
Data utama yang digunakan dalam pengujian ini berasal dari Kaggle, yaitu kumpulan data nyata yang terdiri dari 5.000 catatan hasil pengumpulan dari penyedia pembelajaran swasta. Dataset ini mencakup berbagai atribut penting yang mendukung eksplorasi pola, korelasi, dan wawasan terkait kinerja akademik siswa. Sebelum diterapkan algoritma *K-Means Clustering*, data terlebih dahulu melalui tahap preprocessing, yang meliputi:

1. Pembersihan data dari data yang hilang,
2. Menghilangkan atribut yang tidak digunakan,
3. Mengelompokkan data dilakukan dengan algoritma *K-Means Clustering* di *RapidMiner*, yaitu dengan menghitung antara data dan pusat klaster (*centroid*), lalu mengelompokkan data ke pusat klaster terdekat. Proses ini diulangi hingga posisi pusat klaster tidak lagi berubah.



Gambar 1. Dataset Awal

Dataset ini berisi data awal yang mencerminkan kinerja akademik siswa. Setiap baris mewakili satu siswa, dengan kolom-kolom yang mencakup informasi seperti identitas (ID, nama, dan email), jenis kelamin, usia, program studi, tingkat kehadiran, serta nilai akademik yang terdiri dari skor tugas, ujian tengah semester, dan ujian akhir. Selain itu, tersedia juga data tambahan seperti partisipasi dalam kegiatan ekstrakurikuler, tingkat stres, dan lama waktu tidur harian. Sebelum digunakan dalam proses pemodelan, data ini terlebih dahulu melalui tahap pembersihan dan normalisasi guna memastikan kualitasnya. Langkah ini penting untuk penerapan algoritma *K-Means Clustering* dalam mengelompokkan siswa berdasarkan karakteristik dan prestasi akademik mereka. Karena menurut (Putra & Lestari, 2021) Penggunaan data mining dapat mendukung peningkatan kualitas pembelajaran berbasis karakteristik siswa.



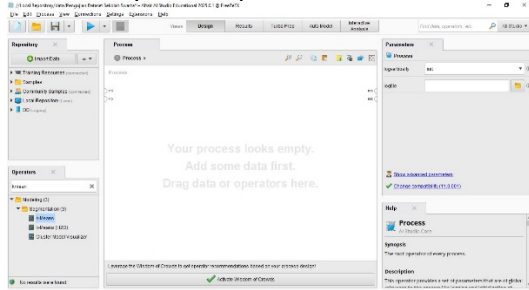
Gambar 2. Dataset Testing

Dataset diatas merupakan dataset awal yang telah disederhanakan untuk keperluan pengujian model, khususnya pengelompokan siswa berdasarkan pencapaian akademik mereka. Setiap baris mewakili satu siswa, dengan kolom-kolom yang mencakup Nama Lengkap, Nilai Tugas,

Nilai Ujian Tengah Semester, dan Nilai Ujian Akhir Semester.

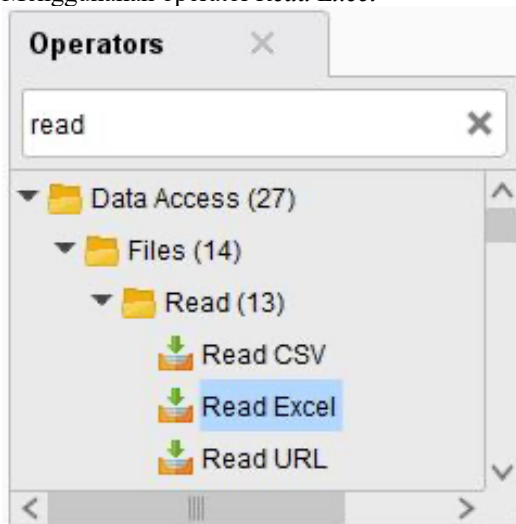
Pengujian dilakukan menggunakan *RapidMiner Studio* dengan langkah-langkah sebagai berikut:

1. Membuka aplikasi *RapidMiner*



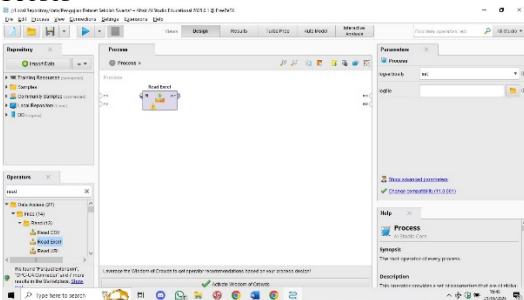
Gambar 3. Tampilan Aplikasi *RapidMiner*

2. Menggunakan operator *Read Excel*



Gambar 4. Operator *Read Excel*

3. Masukkan operator *Read Excel* kedalam Lembar Proses



Gambar 5. Operator *Read Excel* ke Lembar Proses

4. Memasukan Dataset Testing kedalam operator *Read Excel*

Import Data - Select the cells to import.

Sheet: Sheet1 Cell range: A:D Select All Define header row: 1

	A	B	C	D
	Full_Name	Projects_Score	Midterm_Score	Final_Score
1	Omar Williams	62.840	40.610	59.610
2	Maria Brown	98.230	57.270	74.000
3	Ahmed Jones	91.220	41.840	63.850
4	Omar Williams	55.480	45.650	44.440
5	John Smith	87.430	53.130	61.770
6	Liam Brown	67.650	70.210	64.480
7	Ahmed Jones	87.510	90.850	93.740
8	Ahmed Smith	90.040	91.720	59.470
9	Omar Smith	57.570	76.970	62.750
10	Sara Smith	83.090	95.690	86.970
11	John Smith	93.140	70.000	85.540
12	Omar Williams	66.170	93.340	87.900
13	John Davis	54.020	67.240	97.230

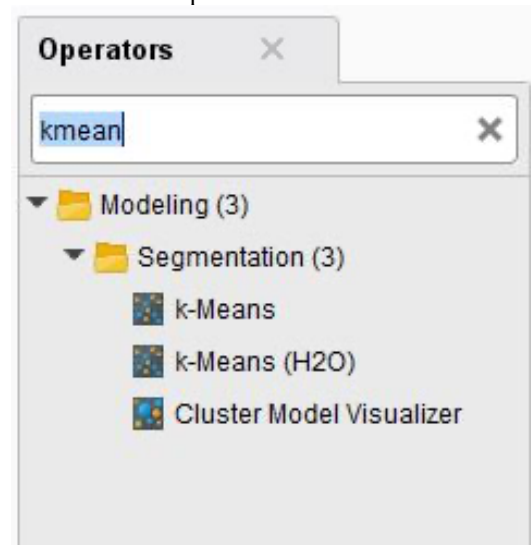
Gambar 6. Dataset Testing kedalam *Read Excel*

5. Berikan Role id untuk *Full_Name* pada Dataset Testing

	Full_Name polynomial id	Projects_Score real	Midterm_Score real	Final_Score real
1	Omar Williams	62.840	40.610	59.610
2	Maria Brown	98.230	57.270	74.000
3	Ahmed Jones	91.220	41.840	63.850
4	Omar Williams	55.480	45.650	44.440
5	John Smith	87.430	53.130	61.770
6	Liam Brown	67.650	70.210	64.480
7	Ahmed Jones	87.510	90.850	93.740
8	Ahmed Smith	90.040	91.720	59.470
9	Omar Smith	57.570	76.970	62.750
10	Sara Smith	83.090	95.690	86.970
11	John Smith	93.140	70.000	85.540
12	Omar Williams	66.170	93.340	87.900

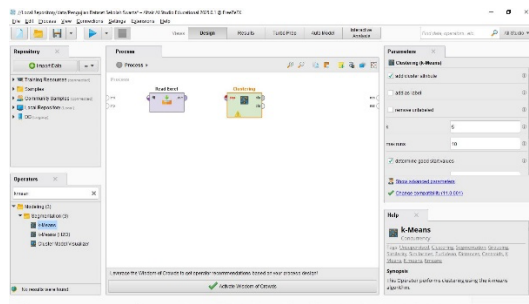
Gambar 7. Menambahkan Role pada *Full Name*

6. Menambahkan operator *K-Means*

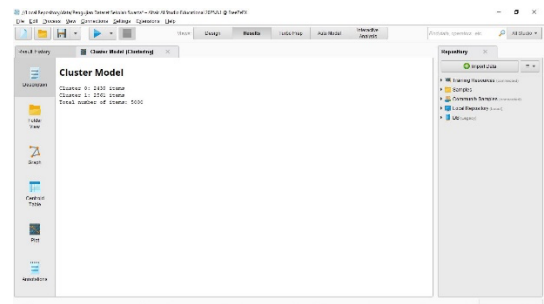


Gambar 8. Operator *K-Means*

7. Memasukan operator *K-Means* kedalam Lembar Proses

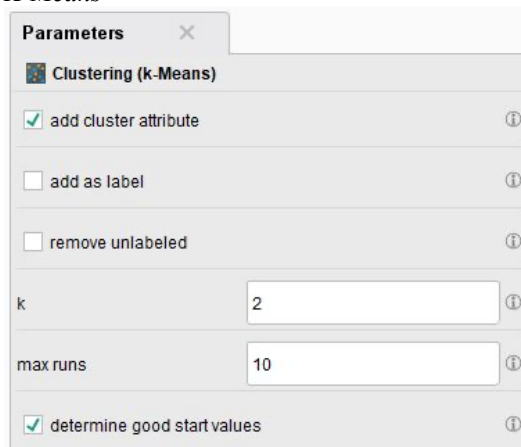


Gambar 9. Operator K-Means ke Lembar Proses



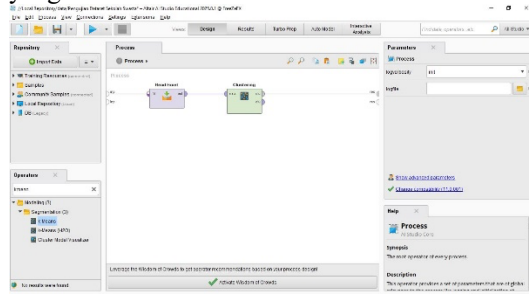
Gambar 13. Hasil dari proses

8. Mengubah Parameters K menjadi 2, pada operator *K-Means*



Gambar 10. Mengubah Parameters K

9. Menyambungkan seluruh proses dan operator yang ada



Gambar 11. Menyambungkan seluruh operator

10. Menjalankan proses yang telah disambungkan



Gambar 12. Menjalankan proses

11. Hasil dari proses yang telah dijalankan

Berdasarkan proses klusterisasi dengan algoritma K-Means menggunakan aplikasi RapidMiner, telah dihasilkan dua kluster dari total 5.000 data siswa yang dianalisis. Kluster pertama (Cluster 0) beranggotakan 2.439 siswa, sementara kluster kedua (Cluster 1) terdiri dari 2.561 siswa. Hasil ini menunjukkan bahwa algoritma *K-Means* efektif dalam mengelompokkan siswa berdasarkan tingkat prestasi akademik mereka, yang dilihat dari nilai tugas, ujian tengah semester, dan ujian akhir semester.

Attribute	cluster_0	cluster_1
Project_Score	75.438	74.159
Midterm_Score	55.264	85.404
Final_Score	70.300	68.829

Gambar 14. Centroid Table

Gambar diatas merupakan hasil pengelompokan data yang dilihat melalui Centroid Table, dengan kesimpulan sebagai berikut:

- Centroid 1 (Cluster 0/Berprestasi):
(Tugas = 75, Nilai UTS = 55, Nilai UAS = 70)
- Centroid 2 (Cluster 1/Kurang Berprestasi):
(Tugas = 55, Nilai UTS = 85, Nilai UAS = 68)

Jadi dapat disimpulkan bahwa hasil pengujian menggunakan algoritma K-Means Clustering menunjukkan bahwa Cluster 0 terdiri dari siswa dengan prestasi akademik yang lebih baik, sedangkan Cluster 1 cenderung berisi siswa dengan prestasi yang lebih rendah. Berdasarkan tabel centroid, rata-rata nilai siswa pada Cluster 0 untuk nilai tugas adalah 75.438, ujian tengah semester sebesar 55.264, dan ujian akhir semester sebesar 70.300. Sementara itu, siswa pada Cluster 1 memiliki nilai tugas sebesar 74.159, ujian tengah semester sebesar 85.404, dan ujian akhir semester sebesar 68.829.

Meskipun nilai ujian tengah semester di Cluster 1 lebih tinggi, nilai pada aspek lain seperti tugas dan ujian akhir lebih konsisten pada Cluster 0, sehingga menunjukkan performa akademik yang lebih stabil. Proses klusterisasi ini dilakukan dengan bantuan RapidMiner yang memungkinkan analisis data secara efisien tanpa pemrograman, serta menyediakan hasil visual yang memudahkan pemahaman dan evaluasi kinerja siswa.

Kutipan dan Acuan

Penggunaan algoritma K-Means Clustering dalam bidang pendidikan semakin luas, khususnya dalam mengelompokkan capaian akademik siswa. Algoritma ini

terbukti mampu memisahkan siswa berdasarkan kemiripan nilai tugas, ujian tengah semester, dan ujian akhir semester, sehingga mempermudah analisis terhadap pola prestasi mereka (Wibowo dan Aryanti, 2025). Perangkat lunak seperti RapidMiner sangat membantu dalam proses analisis data karena dapat digunakan secara efisien tanpa membutuhkan keahlian pemrograman yang kompleks (Nuraini et al., 2022).

Di samping itu, pemanfaatan dataset berskala besar dari sumber terbuka seperti Kaggle memungkinkan penelitian mencerminkan situasi yang lebih realistis di lapangan (Putra dan Sari, 2023). Oleh karena itu, implementasi K-Means dalam klasifikasi akademik tidak hanya berperan dalam meningkatkan kualitas pembelajaran, tetapi juga dalam mendukung perumusan kebijakan yang lebih tepat sasaran di lingkungan pendidikan (Hasanah, 2021). Hal ini sejalan dengan pendekatan pengambilan keputusan berbasis data yang kini menjadi standar dalam pengelolaan pendidikan.

Algoritma clustering seperti K-Means memiliki peran penting dalam mengidentifikasi pola-pola tersembunyi dalam data akademik siswa, termasuk kaitan antara kehadiran, aktivitas ekstrakurikuler, dan prestasi belajar (Rahmawati et al., 2024). Dengan menerapkan teknik preprocessing dan normalisasi secara tepat, integritas dan kualitas data dapat ditingkatkan, sehingga hasil klusterisasi menjadi lebih valid dan menggambarkan kondisi sebenarnya (Syamsuddin dan Lestari, 2021). Hasil pengelompokan ini memungkinkan pihak sekolah untuk merancang tindakan yang lebih terarah, seperti memberikan pendampingan belajar bagi siswa dengan hasil kurang memuaskan atau menyediakan program pengembangan potensi untuk siswa berprestasi (Yuliana, 2023). Pendekatan ini menegaskan bahwa penerapan machine learning dalam pengolahan data pendidikan mampu mendukung perencanaan pembelajaran yang responsif dan berbasis bukti, sejalan dengan tuntutan sistem pendidikan modern.

D. PENUTUP

Simpulan

Penelitian ini membuktikan bahwa penerapan algoritma K-Means Clustering dapat digunakan secara efektif untuk mengelompokkan siswa berdasarkan capaian akademik mereka. Dengan memanfaatkan data berupa nilai tugas, ujian tengah semester, dan ujian akhir semester, sistem berhasil membagi 5.000 data siswa ke dalam dua klaster, yakni kelompok siswa yang memiliki prestasi tinggi dan kelompok dengan prestasi yang lebih rendah. Berdasarkan hasil centroid, siswa dalam Cluster 0 menunjukkan nilai yang lebih konsisten, yang mencerminkan stabilitas performa akademik secara keseluruhan. Penggunaan aplikasi RapidMiner sangat mendukung proses ini karena memungkinkan analisis dan visualisasi data secara cepat tanpa perlu melakukan pemrograman manual, sehingga mempercepat proses eksplorasi informasi dalam bidang pendidikan.

Saran

Melihat hasil yang diperoleh, disarankan agar pihak sekolah atau lembaga pendidikan terkait memanfaatkan teknik data mining seperti K-Means Clustering dalam menganalisis data akademik siswa. Dengan metode ini, pihak sekolah atau lembaga pendidikan terkait dapat mengidentifikasi kelompok siswa yang memerlukan perhatian khusus dan menyusun program pembinaan yang lebih terarah. Selain itu, penggunaan RapidMiner sangat direkomendasikan karena kemudahannya dalam proses analisis data dan penyajian hasil yang mudah dipahami. Untuk pengembangan penelitian di masa mendatang, disarankan menambahkan variabel tambahan seperti motivasi belajar, latar belakang keluarga, atau aktivitas non-akademik agar hasil klusterisasi menjadi lebih menyeluruh dan mendalam.

E. DAFTAR PUSTAKA

- Ardiansyah, M., Nugroho, D., & Putri, A. (2021). Analisis Clustering untuk Pengelompokan Data Akademik Siswa Menggunakan K-Means. *Jurnal Teknologi dan Informatika*, 12(2), 75–82.
- Handayani, R., Widodo, A., & Suryani, N. (2021). Klasifikasi Prestasi Belajar Siswa Menggunakan Metode Data Mining. *Jurnal Teknologi Informasi dan Pendidikan*, 14(1), 50–56.
- Kaggle. (2023). Students Grading Dataset. Retrieved from <https://www.kaggle.com/students-grading-dataset>
- Putra, Y. D., & Lestari, M. (2021). Penerapan Data Mining dalam Dunia Pendidikan untuk Meningkatkan Kualitas Pembelajaran. *Jurnal Ilmu Komputer dan Informatika*, 10(3), 98–105.
- Rahmawati, D., & Yuliana, R. (2021). Pengelompokan Prestasi Akademik Siswa Menggunakan Metode Clustering. *Jurnal Pendidikan dan Teknologi Informasi*, 13(2), 62–69.
- Sari, D., & Pratama, A. (2020). Penerapan Algoritma K-Means Clustering untuk Pengelompokan Nilai Akademik Siswa. *Jurnal Teknik Informatika dan Sistem Informasi*, 6(1), 34–40.
- Wulandari, F., & Nugroho, D. (2020). Pemanfaatan RapidMiner dalam Analisis Data Akademik Siswa. *Jurnal Sistem dan Teknologi Informasi*, 8(2), 22–30.