



Implementasi Data Mining Menggunakan Metode Rapidminer Untuk Optimasi Manajemen Akademik Di SMK Secang

Niha Syufa'a¹, Juwari², Muhammad Ikrar Yamin³, Ahmad Soderi⁴, Rinaldo⁵

¹⁻⁵Teknik Informatika, Sekolah Tinggi Manajemen Informatika dan Komputer Mercusuar
Bekasi, Indonesia

Email Penulis Korespondensi: juwari@mercusuar.ac.id

Abstract. Education in vocational high schools (SMKs) requires effective data management to improve students' academic achievement and discipline. At SMK Islam Secang, students' academic scores and attendance data have so far functioned merely as administrative archives, making it difficult to identify patterns of student performance. This study aims to classify students based on academic achievement and discipline by applying the K-Means Clustering algorithm using RapidMiner. The data used in this study consist of scores from six subjects and attendance records of 35 students from the Light Vehicle Engineering (TKR) department over two semesters. The data were obtained from original school records, compiled using Microsoft Excel, and processed in RapidMiner. The clustering process employed four clusters for academic achievement and two clusters for discipline, with Euclidean Distance used as the similarity measure. The results show that in the first semester, students were grouped into four academic achievement clusters: high achievement (6 students), moderate achievement (7 students), potentially problematic (14 students), and problematic (8 students). In the second semester, the distribution changed to high achievement (19 students), moderate achievement (14 students), potentially problematic (4 students), and problematic (1 student). Meanwhile, student discipline was divided into two clusters: disciplined (31 students) and undisciplined (4 students). These results demonstrate that K-Means Clustering is effective in mapping student conditions, revealing patterns in academic performance and attendance, and supporting educational evaluation, learning planning, and early detection of students who require academic or disciplinary intervention.

Keywords: Data Mining; K-Means Clustering; Academic Achievement; Discipline; RapidMiner; Vocational High School (SMK)

Abstrak. Pendidikan di SMK memerlukan pengelolaan data yang efektif untuk meningkatkan prestasi dan kedisiplinan siswa. Di SMK Islam Secang, data nilai dan absensi selama ini hanya berfungsi sebagai arsip sehingga sulit mengidentifikasi pola pencapaian siswa. Penelitian ini bertujuan mengelompokkan siswa berdasarkan prestasi akademik dan kedisiplinan menggunakan algoritma *KMeans Clustering* pada *RapidMiner*. Data yang digunakan mencakup nilai enam mata pelajaran dan absensi 35 siswa jurusan TKR selama dua semester. Data didapatkan dari data sekolah asli lalu direkap dengan *Microsoft Excel*, dan diolah di *RapidMiner*, dengan jumlah kluster empat untuk prestasi akademik dan dua untuk kedisiplinan, serta Euclidean Distance sebagai ukuran kedekatan. Hasil penelitian menunjukkan, pada semester 1 terdapat kluster berprestasi (6 siswa), cukup berprestasi (7), berpotensi bermasalah (14), dan bermasalah (8). Semester 2 berubah menjadi berprestasi (19), cukup berprestasi (14), berpotensi bermasalah (4), dan bermasalah (1). Kedisiplinan terbagi menjadi disiplin (31 siswa) dan tidak disiplin (4). Dengan hasil tersebut membuktikan bahwa *K-Means Clustering* efektif memetakan kondisi siswa, mengungkap pola nilai dan absensi, serta mendukung evaluasi, perencanaan pembelajaran, dan deteksi dini siswa yang memerlukan intervensi.

Kata kunci: Data Mining; K-Means Clustering; Prestasi Akademik; Kedisiplinan; RapidMiner; SMK

1. PENDAHULUAN

Pendidikan memiliki peran yang sangat penting dalam membentuk sumber daya manusia yang berkualitas dan berdaya saing, khususnya pada jenjang Sekolah Menengah Kejuruan (SMK) yang berorientasi pada kesiapan lulusan memasuki dunia kerja. Selain membekali siswa dengan keterampilan teknis sesuai bidang keahlian, SMK juga dituntut untuk memastikan proses pembelajaran berjalan efektif, disiplin siswa terjaga, serta prestasi akademik dapat

dicapai secara optimal. Dalam konteks ini, pengelolaan data akademik yang baik menjadi salah satu faktor kunci dalam mendukung keberhasilan manajemen pendidikan.

Perkembangan teknologi informasi di era digital telah mendorong institusi pendidikan untuk mengadopsi sistem pengelolaan data yang lebih modern dan berbasis analisis. Setiap semester, sekolah menghasilkan data akademik dalam jumlah besar, seperti nilai mata pelajaran dan data absensi siswa. Data absensi mencerminkan tingkat kehadiran dan partisipasi siswa dalam proses pembelajaran, sedangkan nilai mata pelajaran digunakan sebagai indikator capaian akademik. Namun, pada praktiknya, data-data tersebut sering kali hanya disimpan sebagai arsip administratif dan belum dimanfaatkan secara optimal sebagai dasar evaluasi dan pengambilan keputusan akademik.

Kondisi tersebut juga terjadi di SMK Islam Secang. Berdasarkan hasil observasi dan wawancara dengan pihak sekolah, diketahui bahwa data nilai dan absensi siswa selama ini masih dianalisis secara manual dan bersifat umum. Proses evaluasi pembelajaran membutuhkan waktu yang relatif lama dan belum mampu mengungkap pola prestasi maupun kedisiplinan siswa secara menyeluruh. Akibatnya, sekolah mengalami kesulitan dalam mengidentifikasi siswa berprestasi secara objektif maupun mendeteksi siswa yang berpotensi mengalami permasalahan akademik dan kedisiplinan sejak dini. Hal ini menunjukkan perlunya pendekatan analisis data yang lebih sistematis, akurat, dan efisien.

Salah satu pendekatan yang dapat digunakan untuk mengatasi permasalahan tersebut adalah Data Mining. Data Mining merupakan proses penemuan pola, hubungan, dan informasi penting dari kumpulan data dalam jumlah besar dengan memanfaatkan teknik statistik, pembelajaran mesin, dan kecerdasan buatan. Dalam konteks pendidikan, Data Mining dapat digunakan untuk menganalisis data akademik siswa guna mendukung evaluasi pembelajaran, perencanaan pendidikan, serta pengambilan keputusan berbasis data.

Pemanfaatan Data Mining di bidang pendidikan dikenal sebagai Educational Data Mining dan telah banyak diterapkan untuk menganalisis prestasi, perilaku belajar, dan tingkat kehadiran siswa. Algoritma K-Means Clustering merupakan salah satu metode Data Mining yang banyak digunakan karena kesederhanaan dan kemampuannya dalam mengelompokkan data numerik tanpa label (unsupervised learning). Algoritma ini bekerja dengan membagi data ke dalam sejumlah klaster berdasarkan tingkat kemiripan tertentu, yang diukur menggunakan jarak antar data, seperti Euclidean Distance. Melalui proses klusterisasi, data yang memiliki karakteristik serupa akan berada dalam satu kelompok yang sama, sehingga pola-pola tersembunyi dalam data dapat terlihat dengan lebih jelas. Metode ini dinilai sesuai untuk mengelompokkan siswa

berdasarkan prestasi akademik dan kedisiplinan, yang umumnya direpresentasikan dalam bentuk data numerik.

Untuk mendukung proses analisis, penelitian ini memanfaatkan perangkat lunak RapidMiner. RapidMiner merupakan tools analisis data yang menyediakan antarmuka visual sehingga memudahkan pengguna dalam melakukan proses Data Mining tanpa memerlukan pemrograman yang kompleks. Dengan RapidMiner, proses pengolahan data, penerapan algoritma K-Means Clustering, hingga visualisasi hasil klasterisasi dapat dilakukan secara lebih sistematis dan efisien. Penelitian ini menggunakan data akademik berupa nilai enam mata pelajaran, yaitu Teknologi Informasi dan Komunikasi, Matematika, Pendidikan Agama, Bahasa Inggris, mata pelajaran Kejuruan, dan Bahasa Indonesia, serta data absensi siswa yang meliputi kehadiran, sakit, izin, dan alpa. Data tersebut diambil dari 35 siswa jurusan Teknik Kendaraan Ringan (TKR) di SMK Islam Secang selama dua semester. Data awal diperoleh dari arsip sekolah, kemudian direkap menggunakan Microsoft Excel sebelum diolah lebih lanjut dengan RapidMiner.

Melalui penerapan K-Means Clustering, siswa dikelompokkan ke dalam beberapa klaster prestasi akademik dan kedisiplinan. Pengelompokan ini diharapkan mampu memberikan gambaran yang lebih objektif mengenai kondisi akademik dan perilaku kehadiran siswa. Dengan adanya hasil klasterisasi, pihak sekolah dapat dengan mudah mengidentifikasi kelompok siswa berprestasi, siswa dengan prestasi sedang, serta siswa yang berpotensi bermasalah. Selain itu, hasil analisis juga dapat digunakan sebagai dasar evaluasi pembelajaran, perencanaan strategi peningkatan prestasi, serta deteksi dini terhadap siswa yang memerlukan pendampingan atau intervensi khusus.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi nyata dalam pemanfaatan Data Mining untuk optimasi manajemen akademik di SMK, khususnya di SMK Islam Secang. Pendekatan ini tidak hanya membantu sekolah dalam mengelola data akademik secara lebih efektif, tetapi juga mendukung terciptanya pengambilan keputusan yang lebih tepat sasaran dan berbasis data.

2. KAJIAN TEORITIS

Implementasi

Implementasi merupakan proses penerapan suatu metode atau sistem untuk mencapai tujuan tertentu. Dalam penelitian ini, implementasi dilakukan melalui penerapan Data Mining menggunakan aplikasi RapidMiner untuk mengolah data akademik siswa sehingga menghasilkan informasi yang dapat membantu pengambilan keputusan sekolah.

Data Mining

Data Mining adalah proses pengolahan dan analisis data dalam jumlah besar untuk menemukan pola, hubungan, atau informasi tersembunyi yang berguna. Data Mining membantu mengubah data mentah menjadi informasi yang mendukung pengambilan keputusan. Tahapan Data Mining meliputi: pemilihan data, Pra-pemrosesan data, Transformasi data, Proses penambangan data, Interpretasi hasil. Dalam penelitian ini, Data Mining digunakan untuk menganalisis nilai dan absensi siswa guna mengetahui pola prestasi dan kedisiplinan siswa.

Optimasi

Optimasi adalah proses mencari hasil terbaik dengan mempertimbangkan tujuan dan kendala tertentu. Dalam konteks penelitian ini, optimasi digunakan untuk meningkatkan efektivitas manajemen akademik sekolah melalui pengolahan data siswa secara lebih cepat, tepat, dan efisien.

Manajemen

Manajemen merupakan proses perencanaan, pengorganisasian, pengoordinasian, dan pengendalian sumber daya untuk mencapai tujuan tertentu. Dalam penelitian ini, manajemen berkaitan dengan pengelolaan data akademik siswa agar sekolah dapat melakukan evaluasi pembelajaran secara lebih efektif.

Akademik

Akademik adalah segala aktivitas yang berkaitan dengan pendidikan, pembelajaran, dan pengembangan ilmu pengetahuan. Dalam penelitian ini, aspek akademik yang dianalisis meliputi: Nilai mata pelajaran siswa dan Kehadiran /Absensi siswa

RapidMiner

RapidMiner adalah perangkat lunak Data Mining berbasis visual yang menggunakan sistem drag-and-drop sehingga memudahkan pengguna dalam mengolah data tanpa harus menulis kode program secara kompleks. RapidMiner mendukung berbagai metode analisis seperti: Clustering, Classification, Prediction, Association Rule. Dalam penelitian ini, RapidMiner digunakan untuk menjalankan proses K-Means Clustering terhadap data nilai dan absensi siswa.

K-Means Clustering

K-Means Clustering merupakan algoritma unsupervised learning yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik. Algoritma ini bekerja dengan: menentukan jumlah cluster, menentukan centroid awal, menghitung jarak data ke centroid,

mengelompokkan data ke cluster terdekat, mengulang proses hingga centroid stabil. Pada penelitian ini:

$k = 4$: untuk mengelompokkan prestasi akademik dan $k = 2$: untuk mengelompokkan kedisiplinan siswa.

Microsoft Excel

Microsoft Excel digunakan sebagai alat bantu pengolahan awal data sebelum data diimpor ke RapidMiner.

CRISP-DM (Cross Industry Standard Process for Data Mining)

Metode ini digunakan sebagai kerangka kerja penelitian agar proses pengolahan data berjalan sistematis.

Flowchart

Flowchart digunakan untuk menggambarkan alur proses sistem secara visual, mulai dari input data, preprocessing, Clustering, output hasil analisis.

DFD dan Diagram Konteks

Untuk menggambarkan aliran data, hubungan antar proses, dan interaksi sistem dengan pengguna dan sumber data.

Kerangka Pemikiran

Menjelaskan bahwa sekolah belum optimal memanfaatkan data akademik dan dibutuhkan metode otomatis untuk analisis data dengan solusi menggunakan Data Mining dengan K-Means sehingga hasil clustering membantu evaluasi akademik dan kedisiplinan siswa.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode Data Mining untuk menganalisis dan mengelompokkan siswa berdasarkan prestasi akademik dan kedisiplinan. Pendekatan kuantitatif dipilih karena data yang digunakan berupa data numerik, yaitu nilai mata pelajaran dan data absensi siswa, yang dapat diolah secara matematis untuk menemukan pola tertentu. Metode Data Mining digunakan untuk mengekstraksi informasi penting dari kumpulan data akademik yang selama ini belum dimanfaatkan secara optimal oleh pihak sekolah.

Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh langsung dari pihak SMK Islam Secang. Data tersebut meliputi nilai rapor enam mata pelajaran, yaitu Teknologi Informasi dan Komunikasi, Matematika, Pendidikan Agama, Bahasa Inggris, mata pelajaran Kejuruan, dan Bahasa Indonesia, serta data absensi siswa yang terdiri dari kehadiran, sakit, izin, dan alpa. Objek penelitian adalah 35 siswa jurusan Teknik Kendaraan

Ringan (TKR) yang diambil selama dua semester pembelajaran. Data akademik ini dipilih karena mencerminkan dua indikator utama keberhasilan pembelajaran, yaitu prestasi akademik dan kedisiplinan siswa.

Teknik Pengumpulan Data

Pengumpulan data dilakukan melalui beberapa tahap. Pertama, studi literatur dilakukan dengan mengkaji buku, jurnal ilmiah, dan penelitian terdahulu yang berkaitan dengan Data Mining, algoritma K-Means Clustering, serta penerapannya dalam bidang pendidikan. Studi literatur bertujuan untuk memperoleh landasan teori yang kuat dan menentukan metode analisis yang sesuai. Kedua, wawancara dilakukan dengan wakil kepala sekolah bidang kurikulum dan wali kelas jurusan TKR untuk memperoleh informasi mengenai proses pengelolaan data nilai dan absensi siswa yang selama ini diterapkan. Ketiga, observasi dilakukan pada bagian tata usaha dan wali kelas untuk memahami proses pencatatan dan pengolahan data akademik siswa. Hasil wawancara dan observasi menunjukkan bahwa data nilai dan absensi masih dianalisis secara manual dan belum dimanfaatkan untuk analisis pola akademik siswa. Keempat, pengumpulan data akademik dilakukan dengan mengambil arsip nilai rapor dan absensi siswa dari pihak sekolah. Data awal berbentuk dokumen tertulis dan lembar rekap sederhana yang kemudian direkap ulang menggunakan Microsoft Excel.

Tahapan Penelitian

Tahapan penelitian dimulai dari pengumpulan data, praproses data, penerapan algoritma K-Means Clustering, hingga analisis hasil klasterisasi. Pada tahap praproses data, dilakukan pengecekan kelengkapan data, penyesuaian format data, serta normalisasi nilai agar setiap atribut memiliki skala yang seimbang. Proses ini penting untuk menghindari bias dalam perhitungan jarak antar data. Setelah tahap praproses selesai, data diimpor ke dalam perangkat lunak RapidMiner. RapidMiner dipilih karena menyediakan antarmuka visual yang memudahkan proses analisis Data Mining tanpa memerlukan pemrograman yang kompleks. Selanjutnya, algoritma K-Means Clustering diterapkan untuk mengelompokkan siswa berdasarkan kemiripan data nilai dan absensi.

Matode Analisis Data

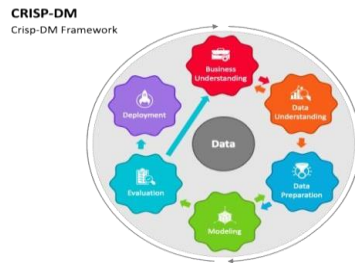
Metode analisis yang digunakan dalam penelitian ini adalah algoritma K-Means Clustering. K-Means merupakan metode pengelompokan data tanpa label (unsupervised learning) yang bertujuan membagi data ke dalam sejumlah klaster berdasarkan tingkat kemiripan tertentu. Dalam penelitian ini, ukuran kemiripan antar data dihitung menggunakan Euclidean Distance. Jumlah klaster ditentukan berdasarkan kebutuhan analisis dan karakteristik

data. Untuk prestasi akademik, digunakan empat klaster yang merepresentasikan kategori siswa berprestasi, cukup berprestasi, berpotensi bermasalah, dan bermasalah. Sementara itu, untuk kedisiplinan digunakan dua klaster, yaitu siswa disiplin dan siswa tidak disiplin. Penentuan jumlah klaster ini bertujuan untuk menghasilkan pengelompokan yang mudah dipahami dan relevan dengan kebutuhan evaluasi akademik di sekolah.

Alat Bantu Perancangan

Dalam merancang sistem yang akan dikembangkan dalam proyek ini, peneliti memerlukan bantuan baik berupa alat bantu ataupun sebuah pendekatan dalam metode perancangan sistem yang berguna dalam memproses hasil observasi yang telah dilakukan sebelumnya, beberapanya antara lain:

CRISP-DM (Cross-Industry Standard Process for Data Mining)



Gambar 1 Proses CRISP-DM



Diagram Alir (Flowchart)


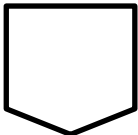


Flowchart adalah alat visual yang efektif untuk merancang dan mendokumentasikan langkah-langkah proses. Akan menggunakan flowchart untuk:

- A. Menggambarkan Alur Proses *Data Mining* secara Keseluruhan
- B. Merinci Langkah-langkah di
- C. Visualisasi Proses Implementasi




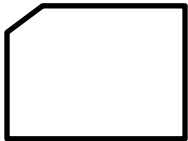
Tabel 1. Simbol Flowchart

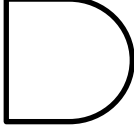
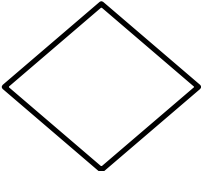

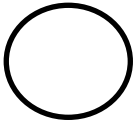
No.	Simbol Flowchart	Nama Simbol	Kegunaan
1.		Terminator	Menunjukkan awal (Start) atau akhir (End) dari sebuah proses.

2.		Input/Output (I/O)	Menunjukkan prose input atau output data (misal: memasukkan data, menampilkan hasil).
3.		Proses / Process	Menunjukkan kegiatan atau proses yang dilakukan (misal: perhitungan, pengolahan data).

4.		Connecting Line / Flow Line	Menunjukkan arah alur proses dari satu simbol ke simbol lainnya dalam <i>flowchart</i> .
5.		Off Page Reference	Simbol untuk keluar masuk atau penyambungan proses dalam lembar kerja yang berbeda
6.		Manual Operation	Simbol yang menyatakan suatu proses yang tidak dilakukan oleh computer
7.		Document	Simbol yang menyatakan bahwa input berasal dari dokumen dalam bentuk fisik, atau

			output yang perlu dicetak
--	--	--	---------------------------

8.		Preparation	Simbol yang menyatakan penyediaan tempat penyimpanan suatu pengolahan untuk memberikan nilai awal
9.		Disk and On-line Storage	Simbol yang menyatakan input yang berasal dari disk atau disimpan ke disk
10.		Magnetik Tape Unit	Simbol yang menyatakan input berasal dari pita magnetic atau output disimpan ke pita magnetic
11.		Punch Card	Simbol yang menyatakan bahwa input berasal dari kartu atau output ditulis ke kartu

12.		Display	Simbol yang menyatakan peralatan output yang digunakan
13.		Decision	Menunjukkan pengambilan keputusan / percabangan (Ya/Tidak atau Benar/Salah).
14.		Predefine Process	Menyatakan suatu proses yang berada dalam sub- bagian atau subprogram
15.		On Page Reference	Simbol untuk keluar masuk atau penyambungan proses dalam lembar kerja yang berbeda

1. Diagram Konteks dan DFD (Data Flow Diagram) Level 0 & 1

Meskipun lebih umum untuk sistem informasi, diagram konteks dan DFD bisa sangat membantu untuk memahami aliran data dalam proyek Anda:

1. Diagram Konteks Level 0 :

Diagram konteks merupakan tingkatan tertinggi dalam Data Flow Diagram (DFD) yang menggambarkan sistem secara umum dalam satu proses tunggal.

2. DFD Level 1:

DFD level 1 merupakan rincian dari diagram konteks. Pada level ini, proses utama dipecah menjadi beberapa subproses yang saling berhubungan melalui aliran data.

1) Alat Bantu Implementasi

Alat bantu implementasi adalah seperangkat perangkat keras, perangkat lunak, metode, dan prosedur yang digunakan untuk memastikan sistem atau aplikasi yang telah dirancang dapat diterapkan di lingkungan operasional secara efektif dan efisien. Alat bantu ini mencakup komputer dan software pendukung, struktur data, dokumentasi, serta prosedur penggunaan yang memudahkan proses penerapan sistem. Didalam penelitian ini kita menggunakan RapidMiner Studi

4. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan metode Data Mining menggunakan algoritma K-Means Clustering dengan bantuan perangkat lunak RapidMiner untuk mengelompokkan siswa SMK Islam Secang berdasarkan prestasi akademik dan kedisiplinan. Data yang digunakan terdiri dari nilai enam mata pelajaran dan data absensi 35 siswa jurusan Teknik Kendaraan Ringan (TKR) selama dua semester. Pada klasterisasi prestasi akademik, jumlah klaster ditentukan sebanyak empat klaster, yang merepresentasikan kategori prestasi siswa. Hasil pengelompokan pada semester 1 menunjukkan bahwa siswa terbagi ke dalam klaster berprestasi (6 siswa), cukup berprestasi (7 siswa), berpotensi bermasalah (14 siswa), dan bermasalah (8 siswa). Distribusi ini menunjukkan bahwa sebagian besar siswa masih berada pada kategori prestasi menengah ke bawah, sehingga diperlukan perhatian khusus dari pihak sekolah.

Sementara itu, pada semester 2 terjadi perubahan signifikan. Hasil klasterisasi menunjukkan peningkatan jumlah siswa pada klaster berprestasi (19 siswa) dan cukup berprestasi (14 siswa), sedangkan jumlah siswa pada klaster berpotensi bermasalah (4 siswa) dan bermasalah (1 siswa) mengalami penurunan. Perubahan ini mengindikasikan adanya peningkatan capaian akademik siswa secara keseluruhan. Untuk klasterisasi kedisiplinan, jumlah klaster ditetapkan sebanyak dua klaster, yaitu siswa disiplin dan siswa tidak disiplin, berdasarkan data absensi. Hasil analisis menunjukkan bahwa 31 siswa termasuk dalam klaster disiplin, sedangkan 4 siswa masuk dalam klaster tidak disiplin, yang ditandai dengan frekuensi ketidakhadiran (izin, sakit, atau alfa) yang lebih tinggi dibandingkan siswa lain.

Perancangan RapidMiner

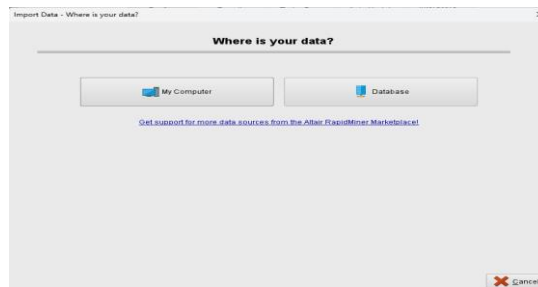
Perancangan teknis dalam *RapidMiner* memerlukan penyusunan operator-operator sebagai berikut:

- A. Sediakan data yang akan diolah berformat file *Excel*.

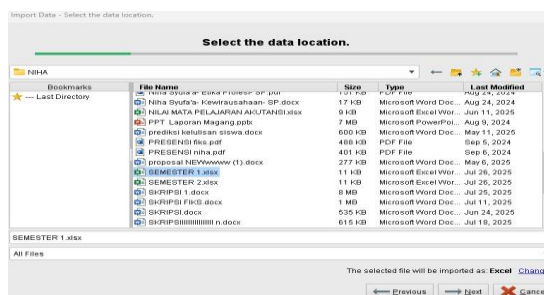
No	NAMA	TK	MTK	ALGAMA	BANJINGSH	KELFELIAN	BANDRO
1	Aditya Firdausy	83	89	85	78	82	88
2	Alvin Firdausy	87	76	81	79	81	85
3	Alvin Indra Awan	80	76	80	74	78	80
4	Alvin Mulya	80	76	80	74	78	80
5	Alvin Prayogo Firdausy	88	88	87	76	79	85
6	Alvin Wicaksono	88	88	87	76	79	85
7	Alvin Wicaksono	88	88	87	76	79	85
8	Alvin Wicaksono	88	88	87	76	79	85
9	Alvin Wicaksono	88	88	87	76	79	85
10	Alvin Wicaksono	88	88	87	76	79	85
11	Alvin Wicaksono	88	88	87	76	79	85
12	Alvin Wicaksono	88	88	87	76	79	85
13	Alvin Wicaksono	88	88	87	76	79	85
14	Alvin Wicaksono	88	88	87	76	79	85
15	Alvin Wicaksono	88	88	87	76	79	85
16	Alvin Wicaksono	88	88	87	76	79	85
17	Alvin Wicaksono	88	88	87	76	79	85
18	Alvin Wicaksono	88	88	87	76	79	85
19	Alvin Wicaksono	88	88	87	76	79	85
20	Alvin Wicaksono	88	88	87	76	79	85

No	NAMA	TK	MTK	ALGAMA	BANJINGSH	KELFELIAN	BANDRO
1	Aditya Firdausy	83	89	85	78	82	88
2	Alvin Firdausy	87	76	81	79	81	85
3	Alvin Indra Awan	80	76	80	74	78	80
4	Alvin Mulya	80	76	80	74	78	80
5	Alvin Prayogo Firdausy	88	88	87	76	79	85
6	Alvin Wicaksono	88	88	87	76	79	85
7	Alvin Wicaksono	88	88	87	76	79	85
8	Alvin Wicaksono	88	88	87	76	79	85
9	Alvin Wicaksono	88	88	87	76	79	85
10	Alvin Wicaksono	88	88	87	76	79	85
11	Alvin Wicaksono	88	88	87	76	79	85
12	Alvin Wicaksono	88	88	87	76	79	85
13	Alvin Wicaksono	88	88	87	76	79	85
14	Alvin Wicaksono	88	88	87	76	79	85
15	Alvin Wicaksono	88	88	87	76	79	85
16	Alvin Wicaksono	88	88	87	76	79	85
17	Alvin Wicaksono	88	88	87	76	79	85
18	Alvin Wicaksono	88	88	87	76	79	85
19	Alvin Wicaksono	88	88	87	76	79	85
20	Alvin Wicaksono	88	88	87	76	79	85

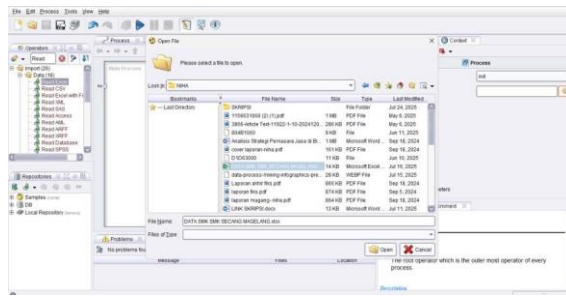
Buka aplikasi *RapidMiner*. Dilanjutkan dengan import data yang akan diolah, yaitu dengan klik button “Add Data” yang terdapat dalam panel Repository.



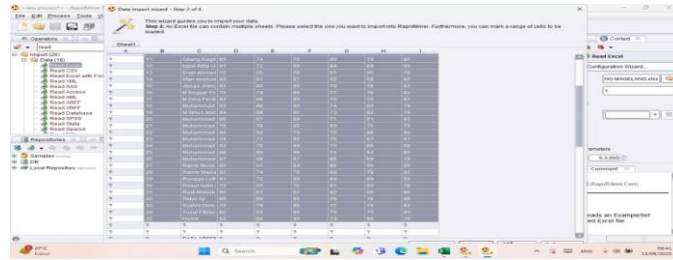
B. Buka aplikasi *RapidMiner*. Dilanjutkan dengan import data yang akan diolah, yaitu dengan klik button “Read Excel” yang terdapat dalam panel Repository. Seperti pada gambar:



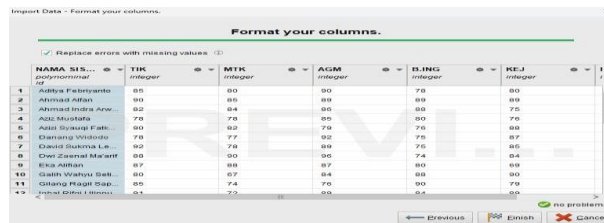
C. Lanjutkan mencari lokasi penyimpanan file data. Apabila sudah ditemukan, maka klik button “Next”. Seperti pada gambar:



D. Pilih cells yang akan di-import-kan, lalu klik button “Next”



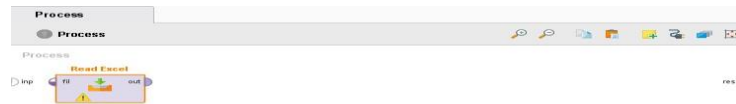
E. Masih dalam proses import data, yaitu pada tahap “format your columns”. Pada bagian ini akan menggunakan tipe data integer, pada nama siswa kita mengubah role id ,dan pada tahap ini merupakan preview/gambaran akhir data yang akan di-import kedalam tools. Kemudian, klik button “Next”



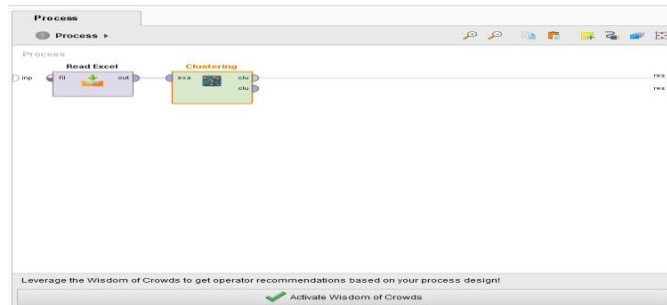
F. Kemudian pada tahap ini, data akan tersimpan di Local Repository dengan nama Data Nilai, dan klik “Finish”.



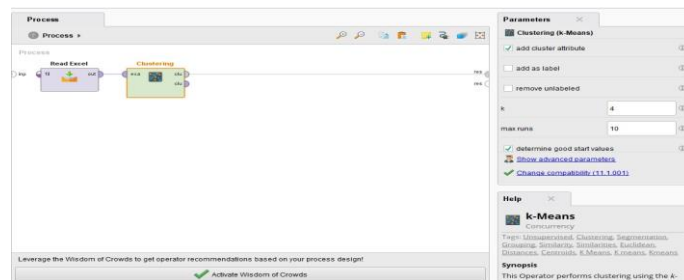
G. Setelah data berhasil di-import, maka klik data tersebut kemudian tarik ke panel process, seperti gambar berikut



- H. Selanjutnya, cari algoritma k-means pada panel Operator, dengan mengetikkan “k-means”. Setelah ditemukannya algoritma k-means, lalu klik algoritma tersebut dan tarik ke panel proses, seperti gambar berikut.



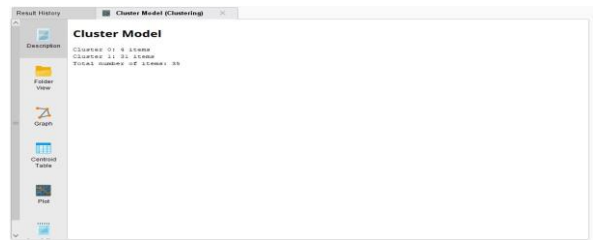
- I. Kemudian, klik *Clustering* yang ada di panel Process, lalu akan muncul Parameters *Clustering* (K-Means). Di dalam parameters tersebut terdapat k, dimana k merupakan jumlah Cluster yang akan dibentuk. Pada tutorial ini menggunakan $k = 4$ (*4Cluster*) untuk nilai siswa dan $k = 2$ untuk absensi siswa, sebagai berikut.



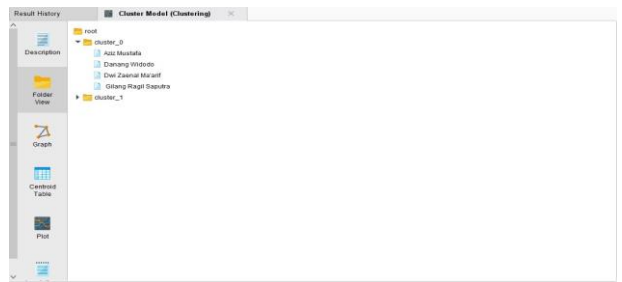
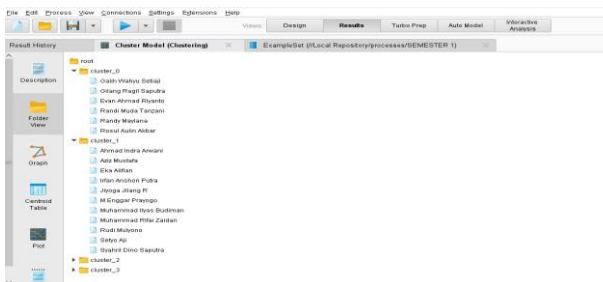
- J. Setelah data selesai dan berhasil diproses, Klik tombol “Run” untuk mengeksekusi proses maka akan menampilkan hasil klusterisasi data seperti gambar dibawah ini.



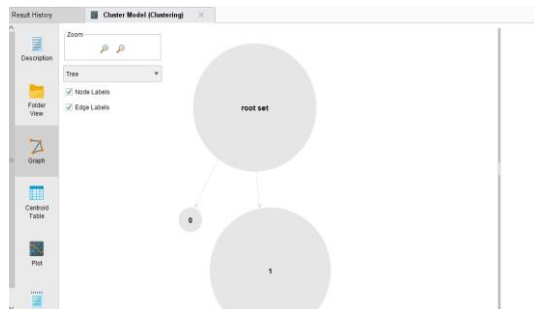
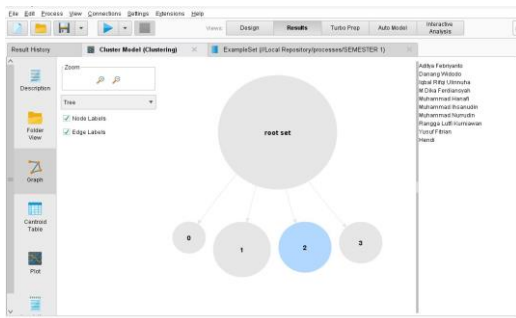
- K. Menampilkan hasil klusterisasi



L. Berikut merupakan hasil Folder View yang disediakan dari *RapidMiner* berupa tampilan folder yang menampilkan output dari model klasterisasi



M. Berikut gambar hasil dari Graph View yang memvisualisasikan hasil klasterisasi dalam bentuk grafik

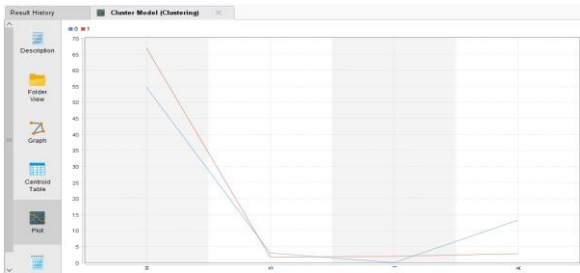
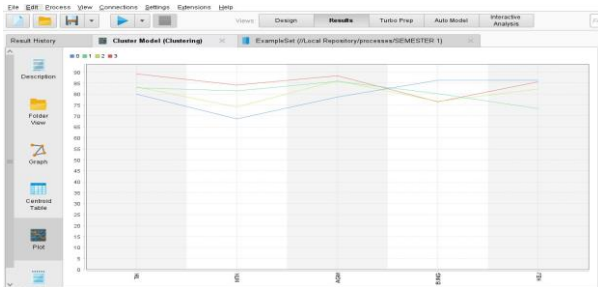


N. Berikut gambar Centroid Table yang menunjukkan nilai rata-rata setiap atribut dalam masing-masing *Cluster*

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
TK	82.829	80.333	80	85.143
MTK	81.143	75.333	83.500	83.714
KEL	74.786	85.333	83.500	85.714
B.RNG	79.143	79.857	83.750	74.714
B.BEND	82.357	78.500	80.125	85.429
AGM	85.286	89	79	89.143

Attribute	cluster_0	cluster_1
H	54.750	66.958
S	3	1.742
I	0	1.958
A	13.250	2.806

O. Berikut gambar Centroid Plot View yang menunjukkan visualisasi sebaran nilai centroid dari tiap *Cluster*



A. Hasil Klaasterisasi

Berikut hasil klasterisasi data nilai siswa dan absensi siswa SMK Secang selama 2 semester dengan metode *K-Means Clustering*:

Hasil Klasterisasi Semester 1

Hasil Klasterisasi Nilai Siswa Semester 1

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
TIK	82.929	90.333	80	85.143
MTK	81.143	75.333	68.500	83.714
AGM	85.286	89	79	89.143
B.ING	79.143	79.667	83.750	74.714
KEJ	74.786	86.333	83.500	85.714
B.INDO	82.357	78.500	80.125	85.429

Dari data hasil centroid table diatas kita mendapatkan nilai rata-rata setiap *Cluster* yang nanti akan digunakan untuk mengetahui peringkat nilai siswa.

Dengan rumus perhitungan:

$$\bar{X} = \frac{\sum x_i}{n}$$

Keterangan:

- \bar{X} (x bar) - notasi untuk Rata-Rata.
- $\sum x_i$ (*Zigma xi*) - jumlah seluruh nilai.
- n - jumlah data.

Tabel 4. 1 Nilai Rata-rata setiap Cluster Nilai Siswa Semester 1

No	Cluster	Rata-rata Nilai
1.	Cluster 0	80,94
2.	Cluster 1	83,86
3.	Cluster 2	79,81
4.	Cluster 3	83,31

Dari Tabel 4.1 diatas kita mendapatkan hasil peringkat nilai siswa semester 1 sebagai berikut:

- Cluster* 1 (Berprestasi, 83,86) : Nilai rata-rata tinggi di semua mata pelajaran, terutama TIK, Pendidikan Agama, dan Kejuruan.
- Cluster* 3 (Cukup Berprestasi, 83,31) : Hampir setara dengan *Cluster* berprestasi, namun terdapat penurunan pada mata pelajaran tertentu seperti Bahasa Inggris.
- Cluster* 0 (Cukup Berprestasi, 80,94) : Nilai akademik cukup baik, tetapi lebih rendah pada beberapa mata pelajaran inti seperti Matematika.
- Cluster* 2 (Bermasalah, 79,81) : Nilai terendah di antara semua klaster, dengan kelemahan jelas pada mata pelajaran Matematika dan Kejuruan.

Interpretasi: Perbedaan antar kluster menunjukkan variasi capaian akademik yang signifikan. Siswa di *Cluster 2* perlu pembinaan akademik intensif, khususnya pada mata pelajaran inti, sedangkan siswa di *Cluster 0* berpotensi meningkat jika kedisiplinan diperbaiki.

a. Hasil Klasterisasi Absensi Siswa Semester 1

Attribute	cluster_0	cluster_1
H	54.750	66.968
S	3	1.742
I	0	1.968
A	13.250	2.806

Gambar 4. 2 Hasil Centroid Tabel Absensi Siswa Semester 1

Dari hasil tersebut kita dapat menyimpulkan hasil *Cluster* absensi siswa semester 1 sebagai berikut :

- a. *Cluster 1* (Disiplin) : Rata-rata hadir tinggi (66,97 hari) dan alpha rendah (2,81 hari), menunjukkan keteraturan hadir ke sekolah.
- b. *Cluster 0* (Kurang Disiplin) : Rata-rata hadir hanya 54,75 hari, dengan alpha tinggi (13,25 hari). Tingkat ketidakhadiran tanpa keterangan tergolong mengkhawatirkan.

Interpretasi: Siswa kurang disiplin kemungkinan besar mengalami hambatan akademik, terutama jika tingkat alpha tinggi berdampak pada keterlambatan materi.

Hasil Klasterisasi Semester 2

a. Hasil klasterisasi nilai siswa semester 2

Attribute	cluster_0	cluster_1	cluster_2	cluster_3
TIK	80	85.636	82.684	72
MTK	67	87.364	79.842	78.750
AGAMA	90	82.364	86.526	84.750
B.INGGRIS	74	73	79.211	78.500
KEJ	73	81.182	81.316	83
B.INDO	89	80.091	79.053	87

Gambar 4. 3 Hasil Centroid Tabel Nilai Siswa Semester 2

Tabel 4. 2 Tabel Rata-rata Nilai Siswa Semester 2

No	Cluster	Rata-rata Nilai
1.	<i>Cluster 0</i>	78,83
2.	<i>Cluster 1</i>	81,61
3.	<i>Cluster 2</i>	83,11
4.	<i>Cluster 3</i>	80,67

Dari Tabel 4.2 kita mendapatkan hasil peringkat nilai siswa semester 2 sebagai berikut:

- a. *Cluster 2* (Berprestasi, 83,11) : Nilai tinggi merata, dengan peningkatan pada mata pelajaran Matematika dan Bahasa Indonesia dibanding semester 1.
- b. *Cluster 1* (Cukup Berprestasi, 81,61) : Nilai baik pada TIK dan Kejuruan, namun menurun pada Pendidikan Agama.
- c. *Cluster 3* (Cukup Berprestasi, 80,67) : Nilai stabil pada mata pelajaran Bahasa Indonesia, tetapi rendah di Matematika dan Pendidikan Agama.
- d. *Cluster 0* (Bermasalah, 78,83) : Nilai terendah, dengan kelemahan konsisten pada Matematika dan Bahasa Inggris.

Interpretasi: Terjadi pergeseran posisi, di mana *Cluster 2* menjadi yang terbaik. Beberapa siswa dari *Cluster* rendah semester 1 berhasil naik peringkat, mengindikasikan adanya keberhasilan intervensi atau perbaikan belajar.

2) Hasil klasterisasi absensi siswa semester 2

Attribute	cluster_0	cluster_1
H	68.419	60.750
S	1.129	6.250
I	0.903	1.750
A	1.548	3.250

Gambar 4. 4 Hasil Centroid Tabel Absensi Siswa Semester 2

Dari hasil tersebut kita dapat menyimpulkan hasil *Cluster* absensi siswa semester 2 sebagai berikut :

- a. *Cluster 0* (Disiplin) : Hadir tertinggi (68,42 hari) dan alpha terendah (1,55 hari), menunjukkan peningkatan kedisiplinan dibanding semester 1.
- b. *Cluster 1* (Kurang Disiplin) : Hadir hanya 60,75 hari, dengan alpha 3,25 hari, serta peningkatan izin dan sakit.

Interpretasi: Perubahan komposisi menunjukkan adanya perbaikan kedisiplinan pada sebagian siswa, namun beberapa siswa tetap berada di kategori kurang disiplin. Hal ini perlu pemantauan untuk mencegah penurunan prestasi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan, implementasi Data Mining menggunakan metode K-Means Clustering pada RapidMiner berhasil membantu proses pengelompokan siswa berdasarkan prestasi akademik dan kedisiplinan di SMK Islam Secang. Pengolahan data nilai enam mata pelajaran dan data absensi 35 siswa selama dua semester mampu menghasilkan beberapa kelompok siswa, yaitu siswa berprestasi, cukup berprestasi,

berpotensi bermasalah, dan bermasalah, serta kelompok siswa disiplin dan tidak disiplin. Hasil penelitian menunjukkan bahwa metode K-Means Clustering efektif dalam menemukan pola tersembunyi dari data akademik sehingga dapat membantu pihak sekolah dalam evaluasi pembelajaran dan pengambilan keputusan akademik secara lebih cepat dan akurat.

Oleh karena itu, disarankan agar pihak sekolah dapat memanfaatkan sistem analisis data secara berkelanjutan untuk memantau perkembangan siswa, melakukan intervensi lebih dini terhadap siswa yang mengalami penurunan prestasi atau kedisiplinan, serta mengembangkan penelitian selanjutnya dengan jumlah data yang lebih besar, variabel yang lebih lengkap, dan penggunaan metode Data Mining lainnya agar hasil analisis menjadi lebih optimal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada STMIK Mercusuar Bekasi, khususnya Program Studi Teknik Informatika, yang telah memberikan dukungan akademik selama proses penelitian dan penyusunan skripsi ini. Ucapan terima kasih juga disampaikan kepada pihak SMK Islam Secang yang telah memberikan izin penelitian serta menyediakan data yang dibutuhkan dalam penelitian ini. Penulis turut menyampaikan penghargaan kepada dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan selama proses penyusunan penelitian hingga selesai. Selain itu, penulis mengucapkan terima kasih kepada keluarga, teman, dan seluruh pihak yang telah memberikan dukungan moral maupun motivasi sehingga penelitian dengan judul “Implementasi Data Mining Menggunakan Metode RapidMiner untuk Optimasi Manajemen Akademik di SMK Islam Secang” ini dapat diselesaikan dengan baik.

DAFTAR REFERENSI

- Abdur Rohman Nurut Toyyibin, & Zaehol Fatah. (2025). Analisis *Data Mining* Menggunakan Metode *Clustering* Terhadap Prestasi Siswa I'Dadiyah Sukorejo. *Jurnal Ilmiah Multidisiplin Ilmu*, 2(1), 96–105. <https://doi.org/10.69714/remqnx91>
- Afiasari, N., Suarna, N., & Rahaningsi, N. (2023). Implementasi *Data Mining* Transaksi Penjualan Menggunakan Algoritma *Clustering* dengan Metode KMeans. *Jurnal SAINTEKOM*, 13(1), 100–110. <https://doi.org/10.33020/saintekom.v13i1.402>
- Agneresa, A., Hananto, A. L., Hilabi, S. S., Hananto, A., & Tukino, T. (2022). Strategi Promosi Penerapan *Data Mining* Mahasiswa Baru Dengan Metode *KMeans Clustering*. *Dirgamaya: Jurnal Manajemen Dan Sistem Informasi*, 2(2), 25–34. <https://doi.org/10.35969/dirgamaya.v2i2.275>
- Antoni, A. R., Sartika, D., & Trianggana, D. A. (2025). Penerapan *Data Mining* Dalam Pengelompokan Data Nilai Siswa SMA Negeri 7 Kota Bengkulu Menggunakan Metode *K-Means Clustering*. *Jl. Meranti Raya No. 32 Kota Bengkulu*, 21(1), 341139.

- Dzalulliyah, A., & Iskandar, A. (2025). *K-Means Clustering Menggunakan RapidMiner Dalam Segmentasi Peminjaman Buku*. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 3937–3945. <https://doi.org/10.36040/jati.v9i3.13422>
- Engineering, I. (2025a). *Klasterisasi Mahasiswa Berdasarkan Performa Akademik Menggunakan Algoritma K-Means pada RapidMiner: Studi Kasus dengan Dataset Student Academic Performance*. 1(1), 46–50.
- Engineering, I. (2025b). *Penerapan Metode Clustering K-Means Menggunakan RapidMiner untuk Klasifikasi Prestasi Siswa di Sekolah Swasta*. 1(1), 20–24.
- Fahrillah Fahrillah, & Zaehol Fatah. (2025). *Pengelompokan Data Nilai Siswa Madrasah Ta’Hiliyah Menggunakan Metode K-Means Clustering*. *JurnalRiset Sistem Informasi*, 2(1), 53–59. <https://doi.org/10.69714/0v1pkz05>
- Ishak, R., & Bengnga, A. (2024). *Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means*. *Jambura Journal of Electrical and Electronics Engineering*, 6(1), 76–81. <https://doi.org/10.37905/jjee.v6i1.2396787>
- Khalish, F., Piranti, N. M., & Martadireja, O. (2025). *Implementasi Menggunakan Teknik Clustering dengan Metode K-Means*. 8, 5392–5397.