Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 9, Nomor: 2, November 2025: 109-116
DOI: 10.30829/algoritma.v9i2.26172
https://jurnal.uinsu.ac.id/index.php/algoritma

Implementasi K-Means Clustering Nilai Ujian Nasional Dalam Peminatan Jurusan Siswa Pada SMAN 1 Gorontalo Utara

Implementation of Kmeans Clustering National Exam Scores In The Specialization Of Student Majors At SMAN 1 Gorontalo Utara

Suhardi Rustam*[1], Zufrianto K Dunggio [2]
[1,]Informatika, Universitas Ichsan Gorontalo Utara
[2]Sistem Informasi, Universitas Ichsan Gorontalo Utara
E-mail: [1]suhardirstm@gmail.com, [2]zufry2dunggio@gmail.com

Abstrak

Permasalahan dalam penentuan peminatan jurusan di SMAN 1 Gorontalo Utara yang masih dilakukan secara manual dan hanya mengandalkan keinginan siswa sering kali menyebabkan hasil penempatan jurusan menjadi kurang akurat dan tidak sepenuhnya mencerminkan kemampuan akademik mereka. Untuk mengatasi hal tersebut, penelitian ini bertujuan menerapkan algoritma clustering K-Means sebagai metode pengelompokan yang lebih objektif, terukur, dan dapat dipertanggungjawabkan. Penelitian menggunakan 90 data nilai Ujian Nasional siswa baru tahun ajaran 2019–2020 dengan empat atribut utama, yaitu nilai Bahasa Indonesia, Matematika, Bahasa Inggris, dan IPA. Seluruh proses analisis dilakukan menggunakan perangkat lunak RapidMiner Studio untuk memastikan hasil perhitungan yang konsisten dan akurat. Berdasarkan hasil pengolahan data, terbentuk empat cluster peminatan yang stabil setelah melalui lima kali iterasi, yakni: Cluster 1 (peminatan IPA 2) dengan 24 siswa, Cluster 2 (peminatan IPS 2) dengan 27 siswa, Cluster 3 (peminatan IPS 1) dengan 27 siswa, dan Cluster 4 (peminatan IPA 1) dengan 12 siswa. Temuan ini menunjukkan bahwa metode K-Means mampu memberikan rekomendasi penentuan jurusan secara cepat, tepat, serta sesuai dengan kemampuan akademik siswa, sehingga dapat dijadikan acuan yang lebih efektif bagi pihak sekolah dalam proses penjurusan.

Kata kunci: Potensi Belajarm Nilai Ujian, Kelompok Jurusan, Algoritma, Kmeans

Abstract

The issue in determining study program specialization at SMAN 1 North Gorontalo, which is still carried out manually and relies solely on students' preferences, often results in inaccurate placement outcomes that do not fully reflect their academic abilities. To address this problem, this study aims to apply the K-Means clustering algorithm as a more objective, measurable, and accountable grouping method. The research uses 90 data points of National Examination scores from new students in the 2019–2020 academic year with four main attributes: Indonesian Language, Mathematics, English, and Science. All analytical processes were conducted using RapidMiner Studio to ensure consistent and accurate computation results. Based on the data processing, four stable specialization clusters were formed after five iterations, namely: Cluster 1 (Science 2 specialization) with 24 students, Cluster 2 (Social Studies 2 specialization) with 27 students, Cluster 3 (Social Studies 1 specialization) with 27 students, and Cluster 4 (Science 1 specialization) with 12 students. These findings indicate that the K-Means method is capable of providing fast, precise, and academically aligned recommendations for study program

Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 9, Nomor: 2, November 2025: 109-116
DOI: 10.30829/algoritma.v9i2.26172
https://jurnal.uinsu.ac.id/index.php/algoritma

placement, making it a more effective reference for schools in the specialization determination process

Keywords: *Learning Potential, Exam Scores, Major Grouping, Algorithm, K-Means*

1. PENDAHULUAN

Penentuan peminatan jurusan di tingkat Sekolah Menengah Atas (SMA) merupakan proses penting untuk mengarahkan siswa pada bidang studi yang sesuai dengan kemampuan dan potensi akademiknya[1]. Namun, praktik penentuan peminatan di SMAN 1 Gorontalo Utara selama ini masih dilakukan secara manual dan hanya mengandalkan pilihan siswa melalui formulir peminatan. Mekanisme tersebut berpotensi menghasilkan ketidaktepatan penempatan jurusan karena tidak mempertimbangkan data akademik secara menyeluruh. Kondisi ini dapat berdampak pada rendahnya kesesuaian minat[2], peningkatan stres akademik, serta ketidakoptimalan siswa dalam mengikuti pembelajaran di jenjang selanjutnya.

Sejumlah penelitian terdahulu menunjukkan bahwa kesalahan pemilihan jurusan dapat menimbulkan gangguan penyesuaian diri, kecemasan, dan penurunan motivasi [3]. Oleh karena itu, diperlukan pendekatan berbasis data untuk membantu proses peminatan secara lebih objektif. Berbagai penelitian sebelumnya telah menerapkan metode data mining, khususnya *clustering*, dalam mendukung pengambilan keputusan pendidikan. Studi[4] menjelaskan bahwa *clustering* mampu mengelompokkan objek berdasarkan kemiripan karakteristik sehingga dapat digunakan untuk analisis akademik, prediksi perilaku[5], maupun segmentasi siswa. Salah satu metode *clustering* yang banyak digunakan adalah K-Means[6], karena mampu memproses data numerik secara efisien dan menghasilkan kelompok dengan tingkat kemiripan tinggi. Penelitian-penelitian sebelumnya telah memanfaatkan K-Means untuk pengelompokkan siswa berdasarkan nilai, performa akademik, ataupun kompetensi. Namun, penelitian terkait implementasi K-Means pada peminatan jurusan di SMAN 1 Gorontalo Utara belum pernah dilakukan.

Berdasarkan kondisi tersebut, penelitian ini berkontribusi dengan menerapkan algoritma K-Means untuk mengelompokkan siswa baru berdasarkan nilai Ujian Nasional[7], sehingga sekolah dapat memperoleh rekomendasi peminatan yang lebih akurat, terukur, dan berbasis data. Dengan demikian, posisi penelitian ini tidak hanya mengisi celah pada mekanisme penjurusan yang masih bersifat manual, tetapi juga memperkuat pemanfaatan data mining sebagai alat bantu pengambilan keputusan di bidang pendidikan.

2. METODE PENELITIAN

Metode penelitian ini disusun untuk menghasilkan model pengelompokan peminatan jurusan yang objektif berdasarkan nilai akademik siswa menggunakan algoritma *clustering* K-Means[8]. Penelitian ini menggunakan pendekatan kuantitatif dengan metode *data mining* berbasis *unsupervised learning*, di mana proses analisis dilakukan dengan cara mengelompokkan data siswa tanpa label awal[9]. Data yang digunakan berjumlah 90 nilai Ujian Nasional siswa baru tahun ajaran 2019–2020 yang mencakup empat atribut utama, yaitu Bahasa Indonesia, Matematika, Bahasa Inggris, dan IPA[10]. Seluruh proses analisis dikerjakan menggunakan perangkat lunak RapidMiner Studio agar pengolahan data berjalan lebih terstruktur dan akurat. Tahapan penelitian dimulai dari proses pengumpulan data melalui dokumentasi resmi nilai UN[11], kemudian dilanjutkan dengan tahap pra-pemrosesan yang meliputi pembersihan data,

Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 9, Nomor: 2, November 2025: 109-116
DOI: 10.30829/algoritma.v9i2.26172
https://jurnal.uinsu.ac.id/index.php/algoritma

normalisasi nilai agar setiap atribut berada pada skala yang seimbang, serta transformasi data ke dalam format yang sesuai untuk analisis lebih lanjut. Setelah data siap digunakan[12], algoritma K-Means diterapkan melalui beberapa langkah inti, yaitu penentuan jumlah cluster sebanyak empat kelompok, inisialisasi centroid awal, perhitungan jarak setiap data terhadap centroid menggunakan Euclidean Distance[13], pengelompokan data berdasarkan jarak terdekat, dan pembaruan centroid hingga proses mencapai kestabilan[14]. Proses ini berjalan selama lima iterasi sampai cluster akhir terbentuk secara konsisten. Implementasi teknik ini dilakukan melalui serangkaian operator dalam RapidMiner seperti *Retrieve*, *Normalize*, *K-Means*, dan *Cluster* Model[15], yang menghasilkan pengelompokan akhir berdasarkan pola persebaran nilai siswa. Evaluasi hasil dilakukan dengan mengamati kestabilan cluster, jumlah anggota pada setiap kelompok, serta pola karakteristik nilai pada masing-masing cluster untuk memastikan kesesuaian dengan peminatan jurusan IPA dan IPS[16]. Hasil akhir penelitian berupa empat cluster peminatan yang dapat digunakan sebagai dasar rekomendasi penjurusan yang lebih cepat, tepat, dan berbasis data dibandingkan metode penjurusan manual.

Tahapan penelitian ini disusun secara sistematis untuk memastikan proses analisis data dan penerapan algoritma K-Means berjalan terstruktur serta dapat direplikasi pada penelitian berikutnya. Tahap pertama adalah

2.1 Pengumpulan Data,

yaitu menghimpun nilai Ujian Nasional siswa baru tahun ajaran 2019–2020 yang terdiri dari empat mata pelajaran inti. Data yang diperoleh kemudian memasuki tahap **pra-pemrosesan (preprocessing)**, yang meliputi pembersihan data dari nilai kosong atau tidak valid, normalisasi nilai agar seluruh atribut berada dalam rentang skala yang sama, serta penyesuaian format dataset agar dapat diproses oleh RapidMiner Studio. Setelah data siap dianalisis, penelitian memasuki tahap

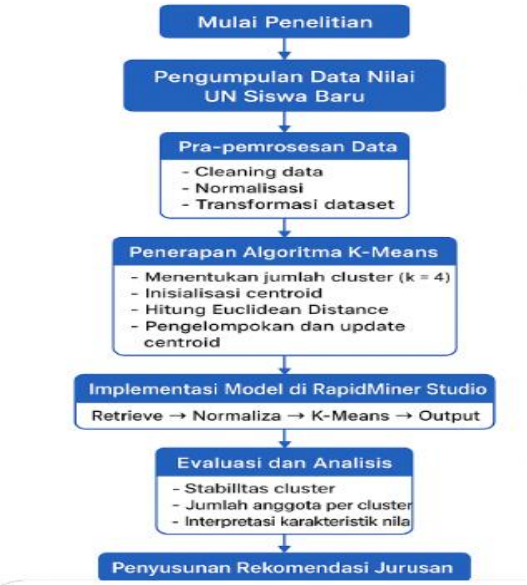
2.2 Penerapan Algoritma K-Means,

yang dimulai dengan penentuan jumlah cluster sebanyak empat kelompok sesuai kebutuhan peminatan jurusan. Proses dilanjutkan dengan inisialisasi centroid, perhitungan jarak menggunakan Euclidean Distance, pengelompokan data berdasarkan kedekatan jarak, dan pembaruan centroid hingga tercapai kestabilan cluster. Algoritma ini dijalankan secara berulang dan berhenti pada iterasi kelima ketika struktur cluster telah konsisten. Tahap berikutnya adalah

2.3 Implementasi Model di RapidMiner Studio, menggunakan operator seperti *Retrieve*, *Normalize*, dan *K-Means* untuk menghasilkan model cluster yang dapat divisualisasikan dan dianalisis. Tahap terakhir adalah **evaluasi hasil**, yaitu mengkaji komposisi masing-masing cluster, interpretasi nilai rata-rata pada tiap kelompok, serta kesesuaian hasil pengelompokan dengan peminatan IPA dan IPS yang ditetapkan pihak sekolah. Secara keseluruhan, tahapan penelitian ini membentuk alur kerja lengkap mulai dari pengumpulan data, pengolahan,

Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 9, Nomor: 2, November 2025: 109-116
DOI: 10.30829/algoritma.v9i2.26172
https://jurnal.uinsu.ac.id/index.php/algoritma

penerapan algoritma, implementasi sistem, hingga interpretasi hasil yang digunakan sebagai dasar rekomendasi penjurusan siswa.



Gambar 2. 1 Tahapan Penelitian dengan K-Means

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pemodelan

Hasil pemodelan *Clustering* data siswa baru berdasarkan nilai UN (Ujian Nasional) dapat dijelaskan sebagai berikut:

1. Pra Pengolahan

Daftar kriteria matapelajaran yang digunakan dapat dilihat pada tabel berikut :

Tabel 3.1 Kriteria Yang Digunakan

No	Nilai UN	Variabel
1	Bahasa Indonesia	X1
2	Matematika	X2
3	Bahasa Inggris	X3
4	IPA	X4

2. Hasil Hitung Algoritma *K-Means*

Perhitungan Algoritma *K-Means Clustering* menggunakan rumus *eucladian distance* atau biasa disebut dengan rumus menghitung jarak. Perhitungan manual Algoritma *K-Means Clustering* dapat dijelaskan pada table sebagai berikut :

1. Dari 90 data diatas telah ditentukan terlebih dahulu jumlah *cluster*yang diinginkan= 4.
2. Menentukan *centroid* awal secara acak. Pada langkah ini secara acak akan dipilih 4 buah data sebagai *centroid*. Misalnya data ke {10, 30, 51, 72}.

Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 9, Nomor: 2, November 2025: 109-116
DOI: 10.30829/algoritma.v9i2.26172
https://jurnal.uinsu.ac.id/index.php/algoritma

3. Pada langkah ini setiap data akan ditentukan *centroid* terdekatnya, dan data tersebut akan ditetapkan sebagai anggota kelompok yang terdekat dengan *centroid*. Perhitungan manual dilakukan menggunakan alat bantu *Microsoft Excel*.Perhitungan jarak dapat dijelaskan sebagai berikut:

$$d1 = \sqrt{(95-90)2 + (82-62)2 + (85-82)2 + (75-75)2}$$
$$= 20{,}78832674$$
$$d2 = \sqrt{(95-85)2 + (82-55)2 + (85-81)2 + (75-67)2}$$
$$= 30{,}22396172$$
$$d3 = \sqrt{(95-81)2 + (82-64)2 + (88-80)2 + (75-76)2}$$
$$= 23{,}26376715$$
$$d4 = \sqrt{(95-90)2 + (82-67)2 + (72-79)2 + (80-66)2}$$
$$= 19{,}28417819$$

Tabel 3.1 Nilai *Centroid* Awal

Cluster	X1	X2	X3	X4
C1 *(dari data ke-10)*	90	62	82	75
C2 *(dari data ke-30)*	85	55	81	67
C3 *(dari data ke-51)*	81	64	80	76
C4*(dari data ke-72)*	90	67	79	66

Dari hasil perhitungan *cluster* awal diatas didapatkan hasil *C1* sebanyak 33 data, *C2* sebanyak 21 data, *C3* sebanyak 24 data dan *C4* sebanyak 12 data. Berikut ini merupakan tabel hasil perhitungan *cluster* awal.

Tabel 3.2 Hasil Perhitungan *Cluster* Awal

Cluster	Jumlah Data
C1	33 Data
C2	21 Data
C3	24 Data
C4	12 Data

Dari hasil perhitungan diatas, langkah selanjutnya adalah untuk mencari *centroid* baru dengan mencari nilai rata-rata dari tiap *Cluster*.

Tabel 3.3 *Centroid* Baru Ke-1

Cluster	X1	X2	X3	X4
C1 *(dari data ke-10)*	90	60,96457115	81	76,14431818
C2 *(dari data ke-30)*	83	56,38367314	79	70,44345238
C3 *(dari data ke-51)*	83	62,73333333	77	74,85
C4*(dari data ke-72)*	89	71,333144	79	71,5375

Setelah menemukan nilai *centroid* baru, dilakukan iterasi dengan menghitung jarak setiap data menggunakan *centroid* baru.

Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 9, Nomor: 2, November 2025: 109-116
DOI: 10.30829/algoritma.v9i2.26172
https://jurnal.uinsu.ac.id/index.php/algoritma

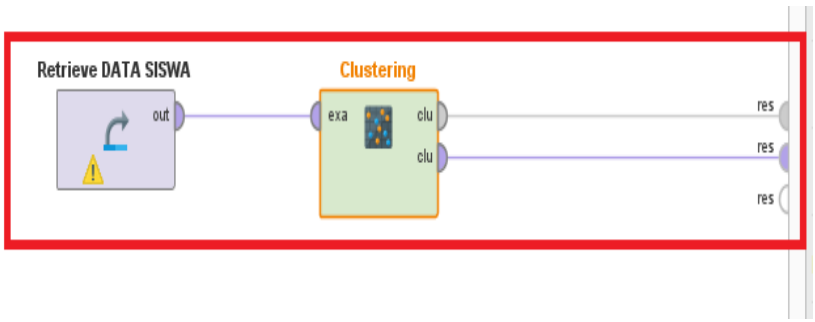
3. Hasil *Clustering*

Berdasarkan hasil perhitungan diatas terlihat bahwa posisi data sudah tidak mengalami perubahan, sehingga proses iterasi sudah dapat dihentikan.Dari hasil data yang diolah diatas, maka dapat disimpulkan bahwa jumlah *cluster* yang terbentuk sebanyak 4 *cluster*. Dimana *cluster* 1 (C1) terdapat 24 siswa yang dikelompokkan ke peminatan jurusan IPA 2, dilihat dari nilai ujian matapelajaran IPA dengan nilai kurang lebih 76,85416667, nilai Matematika sekitar 59,76666633, dan dua nilai penunjang yakni nilai Bahasa Indonesia sekitar 90, dan nilai Bahasa Inggris sekitar 82 dengan nilai rata-rata keseluruhan adalah 77. *Cluster* 2 (C2) terdapat 27 siswa yang dikelompokkan ke peminatan jurusan IPS 2, dengan mempunyai nilai IPA sekitar 71,54768519, nilai Matematika sekitar 56,47619022, dengan dua nilai penunjang Bahasa Indonesia dan Bahasa Inggris masing-masing sebesar 83 dan 79 dengan rata-rata nilai 72. *Cluster* 3 (C3) terdapat 27 siswa yang dikelompokkan ke peminatan jurusan IPS 1, dengan mempunyai nilai IPA sekitar 73,48842593, nilai Matematika sekitar 63,57460326, nilai Bahasa Indonesia sekitar 84 dan nilai Bahasa Inggris sekitar 76 dengan nilai rata-rata 74.*Cluster* 4 (C4) terdapat 12 siswa yang dikelompkkan ke peminatan jurusan IPA 1, dengan mempunyai nilai IPA sekitar 73,87083333, nilai Matematika sekitar 73,47619133, dan nilai Bahasa Indonesia dan nilai Bahasa Inggris masing-masing sekitar 91 dan 79 dengan nilai rata-rata nilai 79.

Berikut ini perbandingan hasil perhitungan iterasi, BCV, WCV dan rasio, sebagai berikut :

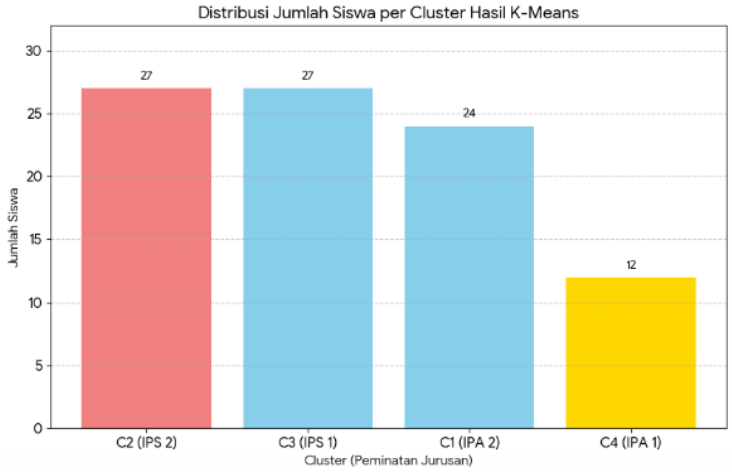
Tabel 3.4 Perbandingan Hasil Iterasi

Proses	BCV	WCV	Rasio
Iterasi 1	65,9543992	361403,2029	0,000182495
Iterasi 2	67,58937617	355233,1267	0,000190268
Iterasi 3	67,7269606	352134,427	0,000192333
Iterasi 4	72,84673355	351003,4857	0,000207538
Iterasi 5	72,84673355	351003,4857	0,000207538



Gambar 3.1 Pemodel K-Means

Algoritma: Jurnal Ilmu Komputer Informatika **e-ISSN 2598-6341**
Volume: 9, Nomor: 2, November 2025: 109-116
DOI: 10.30829/algoritma.v9i2.26172
https://jurnal.uinsu.ac.id/index.php/algoritma



Gambar 3.2 Grafik Distribusi Klustering Siswa

KESIMPULAN DAN SARAN

Penelitian ini menunjukkan bahwa algoritma K-Means mampu mengelompokkan siswa secara objektif berdasarkan nilai Ujian Nasional dan menghasilkan empat cluster peminatan yang stabil. Metode ini lebih akurat dan efisien dibandingkan penjurusan manual karena mengurangi subjektivitas dan memberikan rekomendasi yang lebih terukur. Meskipun demikian, K-Means memiliki keterbatasan, terutama ketergantungan pada kualitas data dan jumlah cluster yang harus ditentukan sejak awal.

Untuk pengembangan selanjutnya, disarankan menggunakan atribut tambahan seperti nilai rapor, tes minat bakat, atau psikotes agar hasil peminatan lebih komprehensif. Penelitian juga dapat membandingkan metode *clustering* lain atau mengembangkan sistem berbasis aplikasi/web agar proses penjurusan lebih otomatis, cepat, dan mudah digunakan oleh pihak sekolah.

DAFTAR PUSTAKA

[1] Kurniawan, W., & Kurniawan, R. (2025). PENERAPAN ALGORITMA K-MEANS CLUSTERING DALAM MENENTUKAN PELUANG MASUK SISWA KE UNIVERSITAS NEGERI. Jurnal Informatika Teknologi dan Sains (Jinteks), 7(1), 386-393.
[2] Fredricka, J. (2023). PENERAPAN METODE K-MEANS CLUSTERING DALAM KLASTERISASI PEMINATAN SISWA TERHADAP MATA PELAJARAN SIMULASI DIGITAL (SIMDIG). Jurnal Media Infotama, 19(1), 20–26.
[3] Palevi, M. R., & Indra, Z. (2024). Implementasi Algoritma K-Means Clustering Dengan Pendekatan Active Learning Pada Siswa SMA Untuk Menentukan Jurusan Ke Perguruan Tinggi. Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer), 23(1), 26-36.
[4] Maulana, H., Idroes, G. M., Kemala, Y., dkk. (2023). Artificial intelligence in education: A systematic literature review of machine learning approaches in student career prediction. Journal of Technology for Education and Learning, 15(1), 1-18.

[5] Susanto, A., & Lestari, D. P. (2024). Komparasi Algoritma Naive Bayes dan C4.5 untuk Prediksi Peminatan Jurusan Siswa di Perguruan Tinggi. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 11(2), 215-224.
[6] Journal-ISI.org. (2025). Clustering of High School Students Academic Scores Using K-Means Algorithm. Jurnal Ilmiah Sistem Informasi, 4(1).
[7] Saputra, I. P. Y., Siswanto, S., & Fredricka, J. (2023). PENERAPAN METODE K-MEANS CLUSTERING DALAM KLASTERISASI PEMINATAN SISWA TERHADAP MATA PELAJARAN SIMULASI DIGITAL (SIMDIG). Jurnal Media Infotama, 19(1), 20-26.
[8] Wijaya, K., & Hartono, R. (2023). Sistem Pendukung Keputusan Pemilihan Jurusan SMA Menggunakan Metode Analytical Hierarchy Process (AHP). Jurnal Sistem Cerdas, 6(1), 45-53.
[9] Hidayat, N., & Sari, Y. A. (2024). Analisis Perbandingan Kinerja Algoritma K-Means dan Fuzzy C-Means dalam Pengelompokan Potensi Akademik Siswa. Jurnal Edukasi dan Penelitian Informatika (JEPIN), 10(1), 112-119.
[10] Rahman, A., & Wulandari, S. (2025). Prediksi Kelulusan Tepat Waktu dan Pemilihan Mata Kuliah Pilihan Menggunakan Algoritma Decision Tree. Indonesian Journal of Computer Science, 14(1), 88-97.
[11] Pratama, R. A., & Nugroho, L. E. (2023). Pengembangan Model Prediktif untuk Rekomendasi Jurusan Kuliah Berdasarkan Nilai Rapor SMA Menggunakan Random Forest. Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI), 12(3), 250-257.
[12] Chen, L., & Zhang, Y. (2024). An Educational Data Mining Approach for Academic Path Recommendation Using Hybrid Filtering. IEEE Access, 12, 14521-14533.
[13] Utami, F. S., & Abdullah, A. (2023). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Klasifikasi Minat dan Bakat Siswa dalam Penentuan Ekstrakurikuler. Jurnal Informatika: Jurnal Pengembangan IT (JPIT), 8(2), 134-140.
[14] Siregar, A. M., & Effendi, S. (2024). Sistem Rekomendasi Peminatan Jurusan di MAN Menggunakan Metode Profile Matching. Jurnal Riset Komputer (JURIKOM), 11(1), 189-196.
[15] Al-Barakati, A., & Al-Dossari, H. (2023). A Machine Learning Framework for Predicting Student Academic Specialization: A Case Study. International Journal of Advanced Computer Science and Applications (IJACSA), 14(5), 201-209.
[16] Santoso, B., & Setiawan, N. A. (2025). Optimasi Parameter pada Algoritma K-Means Menggunakan Algoritma Genetika untuk Meningkatkan Akurasi Klasterisasi Data Siswa. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), 9(1), 78-85.