

Sistem Rekomendasi Program Studi berbasis K-Means Clustering menggunakan Asesmen Komprehensif dan Nilai Akademik

Arvelynia Deriska ^{1,*}, Suhirman ²

¹ Program Studi Informatika, Universitas Teknologi Yogyakarta, Indonesia

² Program Studi Teknologi Informasi, Universitas Teknologi Yogyakarta, Indonesia

* Correspondence: arvelyniaderiska@gmail.com

Copyright: © 2025 by the authors

Received: 13 Oktober 2025 | Revised: 31 Oktober 2025 | Accepted: 4 Desember 2025 | Published: 10 Desember 2025

Abstrak

Tingginya ketidaktepatan pemilihan program studi di Indonesia menunjukkan perlunya alat asesmen yang mampu memetakan potensi siswa secara holistik. Penelitian ini bertujuan mengembangkan Sistem Rekomendasi Jurusan Impian (SRJI) berbasis *K-Means Clustering* yang mengintegrasikan nilai akademik enam semester dan empat asesmen non-akademik untuk membantu siswa SMAN 3 Batang Hari memilih program studi sesuai potensi multidimensi mereka. Penelitian ini menggunakan desain kuantitatif terapan dengan Model Prototype melalui tahapan identifikasi kebutuhan, pembangunan prototipe *K-Means*, dan pengujian iteratif. Data 152 siswa diproses melalui *median imputation*, *Isolation Forest*, normalisasi, dan PCA menghasilkan 123 data valid yang dikelompokkan menggunakan jarak *Euclidean*. Hasil temuan berupa sistem rekomendasi program studi dengan tiga klaster optimal Silhouette Score 0,6164; DBI 0,5684. Hasil analisis menunjukkan Cluster 0 (n=42, IPS 64%, rapor 75,6) berpola social-enterprising, Cluster 1 (n=58, IPA 72%, rapor 87,5) bertipe investigative-realistic, dan Cluster 2 (n=35, seimbang, rapor 81,2) berkarakter *realistic-conventional*. Hasil pengujian menunjukkan kesesuaian rekomendasi 80% berdasarkan validasi guru BK. Kontribusi berupa model *Integrated Multidimensional Student Profiling* yang memperluas literatur *educational data mining* dan mendukung keputusan penjurusan objektif berbasis data.

Kata kunci: *educational data mining; k-means clustering; multi-domain assessment; pemilihan program studi, sistem rekomendasi*

Abstract

The high level of inaccuracy in choosing study programmes in Indonesia demonstrates the need for an assessment tool that can map students' potential holistically. This study aims to develop a Dream Major Recommendation System (SRJI) based on K-Means Clustering that integrates six semesters of academic grades and four non-academic assessments to help students at SMAN 3 Batang Hari choose study programmes according to their multidimensional potential. This study uses an applied quantitative design with a Prototype Model through the stages of needs identification, K-Means prototype development, and iterative testing. Data from 152 students were processed using median imputation, Isolation Forest, normalization, and Principal Component Analysis (PCA), resulting in 123 valid datasets clustered using Euclidean distance. The study resulted in a study program recommendation system featuring three optimal clusters with a Silhouette Score of 0.6164 and a Davies–Bouldin Index (DBI) of 0.5684. The analysis identified Cluster 0 (n=42, Social Science 64%, mean score 75.6) with a social-enterprising pattern, Cluster 1 (n=58, Natural Science 72%, mean score 87.5) as an investigative-realistic type, and Cluster 2 (n=35, balanced distribution, mean score 81.2) with realistic-conventional characteristics. Testing results demonstrated an 80% recommendation suitability rate based on validation by Guidance and Counseling teachers. This study contributes an Integrated Multidimensional Student Profiling model that expands the educational data mining literature and supports objective, data-driven decision-making in study program selection.



Keywords: *educational data mining; k-means clustering; multi-domain assessment; study program selection; recommendation system*

PENDAHULUAN

Ketidaktepatan pemilihan program studi merupakan permasalahan global yang berdampak signifikan terhadap keberlanjutan pendidikan, motivasi akademik, serta kesiapan karier mahasiswa. Fenomena *major switching*, yang mencerminkan ketidaksesuaian antara minat, potensi, dan pilihan studi, menjadi indikasi kuat bahwa proses penjurusan masih menghadapi tantangan mendasar. Penelitian di Kamboja menunjukkan lebih dari 54% mahasiswa berganti jurusan saat masuk perguruan tinggi (Kao et al., 2023). Di Indonesia, 87% mahasiswa merasa salah memilih jurusan sehingga berdampak pada motivasi, ketidakpastian karier, dan keterlambatan kelulusan (Diana et al., 2023). Fenomena ini tidak hanya terkait aspek akademik, tetapi juga dinamika emosional seperti ketidakpastian identitas belajar dan tekanan psikologis (Silver, 2023), yang menunjukkan mekanisme pemilihan program studi belum mampu memetakan profil mahasiswa secara komprehensif dan berpotensi menimbulkan *academic misfit*.

Pemilihan program studi dipengaruhi oleh sejumlah faktor, baik internal seperti kapasitas kognitif, minat, gaya belajar, dan disposisi kepribadian, maupun eksternal seperti ekspektasi keluarga dan lingkungan sosial. Namun, praktik penjurusan di sekolah menengah di Indonesia masih didominasi oleh penilaian satu dimensi, yaitu nilai akademik semata (Aisyah et al., 2023; Arifin et al., 2020; Kolne et al., 2025;). Keterbatasan asesmen holistik mengakibatkan potensi siswa tidak tergambar secara multidimensi dan keputusan penjurusan menjadi sangat rentan terhadap bias subjektif. Perkembangan *educational data mining* menawarkan peluang strategis untuk merancang sistem rekomendasi yang mampu menyintesis berbagai dimensi data (Zhao et al., 2024) siswa sehingga pemilihan program studi dapat dilakukan secara lebih objektif dan terukur. Dalam konteks ini, algoritma *K-Means Clustering* menjadi salah satu pendekatan yang banyak digunakan karena kemampuannya mengelompokkan data multivariabel secara efisien (Basalamah & Setyadi, 2023; Karmanita & Hendrik, 2023) dan menghasilkan representasi pola profil siswa yang dapat dijadikan dasar rekomendasi (Eirlangga & Syaputra, 2022; Darsono et al., 2022; Hidayat, 2022).

Sebagai respons terhadap keterbatasan penilaian tradisional, sistem rekomendasi berbasis *K-Means Clustering* menyediakan mekanisme pemetaan siswa yang lebih objektif melalui analisis kesamaan profil multidimensional (Lubis & Sriani, 2025), sehingga rekomendasi program studi tidak bergantung pada intuisi, tetapi pada kedekatan karakteristik siswa dengan kelompok berperforma serupa (Walangare, & Sujatmiko, 2022). Sistem ini mengintegrasikan nilai akademik enam semester dengan empat instrumen non-akademik tes kemampuan, gaya belajar, minat penjurusan, dan kepribadian yang selaras dengan *Multidimensional Assessment Theory*, bahwa pemetaan potensi individu harus mencakup dimensi kognitif, afektif, dan konatif untuk meningkatkan validitas prediktif rekomendasi. Proses clustering kemudian dioptimalkan melalui tahapan preprocessing, normalisasi, deteksi outlier, serta evaluasi kualitas cluster menggunakan Elbow Method, Silhouette Score, dan *Davies Bouldin Index*, sehingga centroid yang terbentuk lebih akurat dan representatif.

Penelitian ini mengacu pada *Person Environment Fit Theory* yang menyatakan bahwa keberhasilan akademik dan kepuasan karier sangat dipengaruhi oleh tingkat keselarasan antara karakteristik personal dan tuntutan lingkungan studi. Model RIASEC Holland memberikan kerangka konseptual yang relevan untuk memetakan kecenderungan minat dan kepribadian siswa, sehingga *cluster* yang terbentuk dari model *K-Means* dapat dikonseptualisasikan sebagai kelompok “kecocokan” yang secara empiris mendukung kompatibilitas pilihan jurusan (Ramadania & Fatah, 2024). Integrasi teori ini ke dalam sistem rekomendasi memastikan

bawa keputusan penjurusan tidak hanya komputasional akurat, tetapi juga valid dari sisi psikopedagogis.

Meskipun sejumlah penelitian sebelumnya telah menerapkan *K-Means* dalam konteks rekomendasi jurusan, mayoritas studi menunjukkan keterbatasan pada variabel input yang sempit, biasanya hanya melibatkan nilai akademik atau minat, sehingga tidak mampu menangkap kompleksitas profil siswa secara multidimensional (Syahril et al., 2023; Khoirunnisa & Arkhiansyah, 2024). Selain itu, akurasi model yang dilaporkan relatif rendah, seperti Silhouette 0,340 pada penelitian Setiawan & Mailoa (2025), serta minimnya validasi praktis oleh guru bimbingan konseling yang semestinya menjadi pemakai utama sistem rekomendasi (Utama et al., 2024). Hal tersebut menunjukkan perlunya model rekomendasi yang lebih komprehensif, akurat, dan teruji secara praktis.

Penelitian ini bertujuan menghasilkan sistem rekomendasi program studi yang objektif, personal, terukur, dan relevan dengan kebutuhan siswa serta praktik konseling pendidikan. Kontribusi penelitian ini meliputi pengembangan model *cluster* multidimensional yang memperkuat literatur *educational data mining*, integrasi teori psikologis dan pendekatan algoritmik untuk meningkatkan validitas rekomendasi, serta implementasi sistem berbasis web yang dapat digunakan secara langsung oleh guru BK dan siswa dalam proses pengambilan keputusan penjurusan yang berbasis data.

METODE

Penelitian ini mengembangkan Sistem Rekomendasi Jurusan Impian (SRJI) dengan pendekatan Model *Prototype* untuk menghasilkan sistem yang adaptif terhadap kebutuhan pengguna. Pengembangan dimulai dari identifikasi kebutuhan melalui konsultasi dengan guru BK, diikuti perancangan prototipe yang mengintegrasikan algoritma *K-Means* sebagai komponen utama. Prototipe tersebut diuji secara iteratif bersama guru BK untuk menilai akurasi, kegunaan, dan kesesuaian fitur, sehingga perbaikannya dapat dilakukan secara berkelanjutan hingga sistem mencapai stabilitas fungsional. Penelitian melibatkan 152 siswa kelas XII SMAN 3 Batang Hari (69 IPA, 83 IPS), yang memberikan validitas kontekstual kuat dalam lingkungan pendidikan menengah atas dan turut menentukan pemilihan variabel akademik dan non-akademik dalam analisis.

Data penelitian terdiri atas dua kategori utama. Data akademik berupa nilai rapor berskala numerik (0–100) yang diverifikasi oleh sekolah untuk memastikan konsistensi dan keakuratannya. Data non-akademik berskala ordinal (Likert 1–5) diperoleh melalui empat instrumen psikologis, yaitu Tes Kemampuan, Tes Gaya Belajar, Tes Penjurusan, dan Tes Kepribadian, yang disusun berdasarkan kajian teori psikologi pendidikan dan divalidasi secara konseptual untuk memastikan kesesuaian konstruk. Setelah proses verifikasi, sebanyak 123 data siswa memenuhi kriteria kelengkapan dan konsistensi sehingga layak dianalisis. Kedua jenis data ini memberikan representasi komprehensif mengenai karakteristik kognitif, preferensi belajar, kecenderungan minat, dan aspek kepribadian yang relevan bagi pengambilan keputusan penjurusan.

Sebelum analisis dilakukan, data diproses melalui tahapan *preprocessing* yang mencakup penanganan data hilang sebesar 1–4% per variabel menggunakan *Median Imputation* karena ketahanannya terhadap outlier. Deteksi outlier dilakukan menggunakan *Isolation Forest* (*contamination* = 0,19) dan divalidasi dengan *Local Outlier Factor* (*n_neighbors* = 20), menghasilkan 29 dari 152 data sebagai outlier sehingga 123 data dinyatakan layak untuk analisis. Seluruh fitur kemudian dinormalisasi menggunakan *StandardScaler* untuk menyeragamkan skala numerik sesuai kebutuhan algoritma *K-Means* dan PCA. Setelah proses *preprocessing*, pengelompokan dilakukan dengan menghitung jarak setiap data terhadap centroid menggunakan *Euclidean Distance*, di mana siswa dengan jarak minimum ditempatkan pada cluster yang sesuai.

$$d(x_i, \mu_j) = \sqrt{\sum_{k=1}^n (x_{ik} - \mu_{jk})^2} \quad (1)$$

Keterangan:

$d(x_i, \mu_j)$: Jarak antara data ke- i dan pusat *cluster* ke- j

x_{ik} : Nilai ke- k dari data ke- i

μ_{jk} : Nilai ke- k dari centroid ke- j

n : Jumlah variabel yang digunakan dalam perhitungan

Jarak minimum $d(x_i, \mu_j)$ menentukan penempatan data ke *cluster* terdekat. Pada tahap validasi, $K = 3$ dipilih karena menjadi titik siku yang paling jelas pada Elbow Method serta menghasilkan kombinasi kualitas *cluster* terbaik, yaitu Silhouette Score 0.6164 dan DBI 0.5684. Hasil *clusterisasi* ini kemudian diintegrasikan ke dalam sistem rekomendasi program studi berbasis web, di mana model *K-Means* digunakan untuk menghitung kedekatan siswa baru terhadap centroid yang telah terbentuk. Sistem ini diuji secara fungsional dan iteratif bersama guru BK untuk memastikan rekomendasi yang dihasilkan tidak hanya akurat secara komputasional, tetapi juga aplikatif dan relevan dalam konteks pendidikan (Saputra & Nataliani, 2021).

HASIL DAN PEMBAHASAN

Hasil

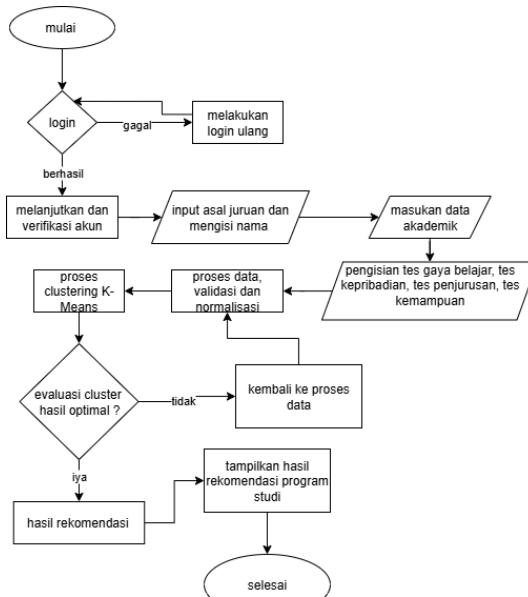
Penelitian ini menghasilkan sistem rekomendasi program studi berbasis algoritma *K-Means Clustering* yang dirancang untuk mendukung pengambilan keputusan akademik siswa kelas XII secara lebih objektif dan terukur. Sistem dikembangkan berdasarkan temuan observasi dan wawancara dengan Guru Bimbingan dan Konseling SMAN 3 Batang Hari, yang menekankan kebutuhan akan alat pemetaan profil siswa yang komprehensif serta layanan asesmen yang mudah diakses oleh siswa. Untuk itu, sistem mengintegrasikan data multivariabel yang mencakup nilai rapor enam semester dan empat instrumen non-akademik—tes kemampuan, gaya belajar, minat penjurusan, dan kepribadian ke dalam arsitektur pemrosesan berbasis *machine learning* guna menghasilkan rekomendasi program studi yang terpersonalisasi.

Alur kerja sistem (gambar 1) dimulai dari proses autentikasi pengguna, pengisian data akademik dan instrumen asesmen, hingga tahap validasi dan normalisasi data sebelum diproses menggunakan algoritma K-Means. Model selanjutnya membentuk cluster siswa dan mengevaluasi kualitas pengelompokan untuk memastikan optimalitas model. Apabila hasil belum memenuhi kriteria kualitas, pemrosesan dilakukan kembali. Setelah cluster optimal diperoleh, sistem menghasilkan rekomendasi berdasarkan kedekatan profil siswa dengan centroid yang terbentuk. Penentuan jumlah cluster terbaik divisualisasikan pada Gambar 2 menggunakan kombinasi Elbow Method, Silhouette Score, dan Davies–Bouldin Index (DBI), yang memastikan pemilihan nilai k yang paling representatif.

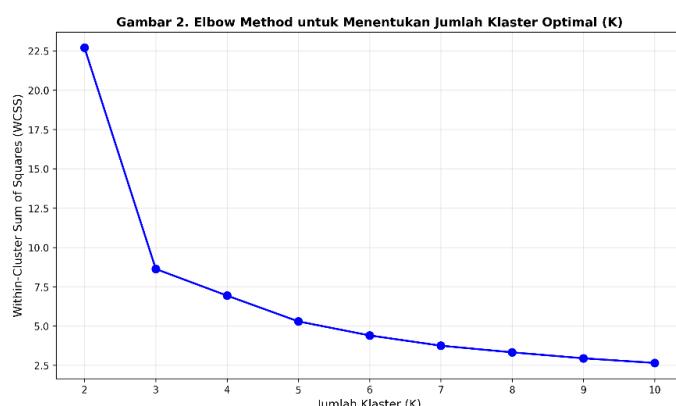
Hasil gambar 2 menunjukkan bahwa nilai *Within-Cluster Sum of Squares* (WCSS) mengalami penurunan signifikan hingga titik $k = 3$ sebelum mencapai fase konvergensi, membentuk pola *elbow* yang mengindikasikan jumlah *cluster* optimal. Validasi kuantitatif menunjukkan nilai *Silhouette Score* sebesar 0,6164 dan *Davies–Bouldin Index* (DBI) sebesar 0,5684, yang menunjukkan bahwa setiap *cluster* memiliki kesamaan internal yang kuat dan perbedaan yang jelas dengan *cluster* lainnya.

Selanjutnya hasil pada gambar 3 menampilkan visualisasi *cluster* menggunakan dua komponen utama PCA, yang digunakan untuk memetakan sebaran data siswa dalam ruang

berdimensi rendah. Hasil pada tabel 1 merangkum karakteristik setiap *cluster* berdasarkan gabungan indikator akademik dan non-akademik. Pola pemisahan antar *cluster* terlihat jelas, konsisten dengan nilai Silhouette Score sebesar 0,6164 dan DBI sebesar 0,5684 yang menunjukkan kualitas *cluster* yang baik. Berdasarkan struktur *cluster* ini, sistem merekomendasikan program studi dengan memetakan siswa baru ke centroid terdekat, sehingga setiap rekomendasi didasarkan pada kedekatan profil siswa dengan karakteristik *cluster* yang telah terbentuk.



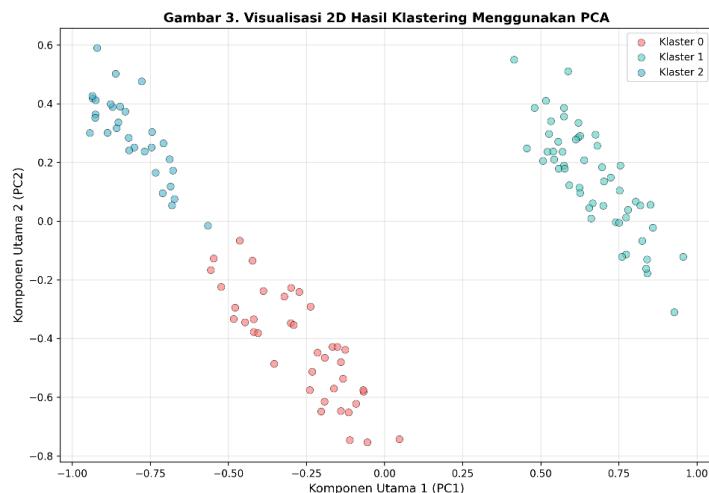
Gambar 1. Flowchart



Gambar 2. Elbow method untuk menentukan jumlah *cluster* optimal (k)

Table 1. Profil dan karakteristik *cluster* hasil analisis k-means

Variabel	Indikator	Cluster 0	Cluster 1	Cluster 2
Jumlah Siswa (%)		42 (31,1%)	58 (43,0%)	35 (25,9%)
Akademik	Weighted Rapor	75,6	87,5	81,2
	Tren Peningkatan	4,8	3,6	4
	Stabilitas (Standar Deviasi)	2,1	1,5	1,8
Non-Akademik	Rata-rata 4 Tes	70	84,1	76,9
Distribusi Jurusan	IPA : IPS	36% : 64%	72% : 28%	46% : 54%



Gambar 3. Visualisasi 2d hasil clustering menggunakan pca

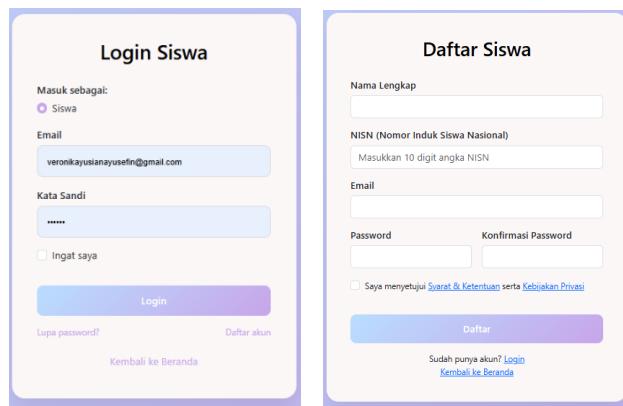
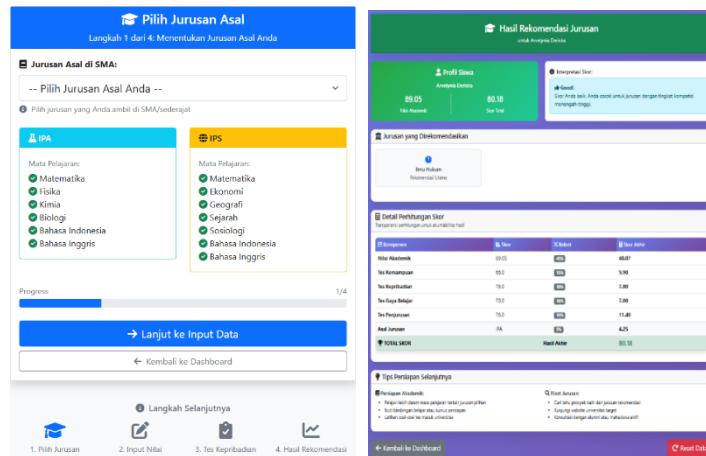
Tabel 2. Hasil rekomendasi berbasis *clustering*

Kode Siswa	Jurusan Asal	Fakultas	Jurusan	Cluster	Rata-rata Rapor	Rata-rata Tes
S01	IPA	Fakultas Keguruan dan Ilmu Pendidikan	Pendidikan Matematika	1	76,7	67,8
S02	IPS	Fakultas Ilmu Komputer	Sistem Informasi	0	87,8	78,6
S03	IPS	Fakultas Pertanian	Kehutanan	2	74,1	69,6

Tabel 2 menyajikan contoh rekomendasi program studi untuk tiga siswa dengan karakteristik berbeda. Siswa S01 (IPA), yang berada pada cluster 1, direkomendasikan ke Pendidikan Matematika berdasarkan performa akademik yang baik dan hasil asesmen non-akademik yang seimbang, selaras dengan tuntutan konseptual dan pedagogis bidang tersebut. Siswa S02 (IPS) pada cluster 0 direkomendasikan ke Sistem Informasi karena memiliki performa akademik dan non-akademik tertinggi, mencerminkan kemampuan analitis yang kuat. Adapun siswa S03 (IPS) pada cluster 2 direkomendasikan ke Kehutanan, sesuai dengan kecenderungan profil praktis dan nilai yang stabil. Perbedaan rekomendasi ini menunjukkan bahwa sistem mampu memetakan karakteristik masing-masing cluster secara tepat ke program studi yang relevan.

Validasi sistem dilakukan melalui tiga tahap. Validasi teknis menunjukkan stabilitas model dengan *Silhouette Score* 0,6164 dan Davies Bouldin Index 0,5684 serta deviasi yang sangat rendah, menegaskan konsistensi hasil clustering. Validasi praktis oleh guru BK menghasilkan tingkat kesesuaian rekomendasi sebesar 80%, mengindikasikan relevansi edukatif yang kuat. Selain itu, evaluasi pengguna oleh siswa menunjukkan respons positif terkait kemudahan penggunaan, kesesuaian rekomendasi, dan efisiensi sistem. Gambar 4 memperlihatkan antarmuka sistem yang menyajikan hasil analisis secara interaktif dan mudah diakses.

Sistem menyediakan mekanisme autentikasi yang memungkinkan siswa mengakses aplikasi secara personal dan aman melalui proses registrasi akun baru atau login menggunakan kredensial terdaftar. Desain halaman autentikasi dirancang sederhana namun fungsional untuk memastikan kemudahan akses bagi seluruh pengguna. Setelah berhasil login, pengguna diarahkan ke halaman dashboard sebagai beranda utama yang menyediakan menu navigasi, termasuk fitur input data akademik dan empat instrumen asesmen non-akademik (gambar 5).

Gambar 4. Menu *login* dan *signup*Gambar 5. Halaman *Input* data dan tes dan halaman hasil rekomendasi

Tabel 3. Hasil pengujian sistem

No.	Fitur yang Diuji	Skenario	Hasil yang Diharapkan	Status
1	Login	Pengguna memasukkan username & password valid	Berhasil masuk ke dashboard	Berhasil
2	Input Data Akademik	Pengguna mengisi nilai rapor lengkap	Data tersimpan & lanjut ke tes	Berhasil
3	Pengisian Tes	Pengguna menyelesaikan 4 tes non-akademik	Nilai tes tersimpan	Berhasil
4	Proses <i>Clustering</i>	Sistem memproses data dengan K-Means	<i>Cluster</i> muncul (0/1/2)	Berhasil
5	Hasil Rekomendasi	Sistem menampilkan rekomendasi program studi	Jurusan, fakultas, dan skor tampil	Berhasil
6	Validasi Input	Form dikirim dengan data tidak lengkap	Sistem menolak dan beri pesan error	Berhasil
7	Logout	Pengguna keluar dari aplikasi	Session terhapus & kembali ke login	Berhasil

Proses pengumpulan data dilakukan melalui halaman input pada gambar 5 yang memfasilitasi siswa untuk memasukkan nilai rapor dan menyelesaikan tes kepribadian serta minat sesuai parameter *clustering*. Halaman ini juga menampilkan hasil rekomendasi program studi setelah data diproses oleh model *K-Means*, di mana sistem menghitung jarak siswa ke

centroid terdekat dan memetakannya ke fakultas serta jurusan yang sesuai. Informasi yang ditampilkan mencakup nama fakultas, program studi, total skor, rata-rata rapor, dan rata-rata tes, memberikan panduan komprehensif bagi siswa dalam memilih program studi. Secara keseluruhan, integrasi antara algoritma *clustering* dan antarmuka sistem telah menghasilkan aplikasi rekomendasi program studi yang akurasi modelnya teruji dan bermanfaat bagi pengguna di lingkungan sekolah.

Hasil pengujian fungsional pada Tabel 3 menunjukkan bahwa seluruh fitur sistem berjalan sesuai skenario. Proses autentikasi, input nilai rapor, dan pengisian empat instrumen non-akademik beroperasi dengan baik dan tersimpan secara konsisten. Data yang terkumpul dapat diproses oleh algoritma K-Means untuk menghasilkan label cluster secara akurat, yang kemudian ditampilkan dalam bentuk rekomendasi fakultas, program studi, serta skor perhitungan. Mekanisme validasi input berfungsi efektif dengan menolak formulir tidak lengkap, dan fitur logout bekerja sesuai harapan. Secara keseluruhan, sistem terbukti stabil, responsif, dan layak digunakan oleh siswa maupun guru BK sebagai alat rekomendasi program studi.

Pembahasan

Hasil *clustering* menunjukkan bahwa model *Integrated Multidimensional Student Profiling* mampu merepresentasikan pola psikopedagogis siswa secara konsisten dengan teori dan riset terdahulu. Tiga cluster yang terbentuk bukan sekadar pemisahan matematis, tetapi mencerminkan kecenderungan belajar dan karakter individual yang selaras dengan tipologi RIASEC dan prinsip *Person Environment Fit Theory*. Cluster pertama berpola *social-enterprising* dengan dominasi siswa IPS dan tren peningkatan nilai yang stabil. Cluster kedua menunjukkan tipe *investigative-realistic* dengan performa akademik tertinggi dan karakter analitis khas rumpun STEM. Cluster ketiga berpola *realistic-conventional* dengan keseimbangan siswa IPA dan IPS dan preferensi aktivitas praktis. Konsistensi ini mengindikasikan bahwa struktur cluster memiliki *construct validity* yang kuat, karena selaras dengan teori kepribadian dan minat akademik.

Secara teknis, kualitas pengelompokan menunjukkan performa yang lebih tinggi dibandingkan sejumlah penelitian sebelumnya. Nilai Silhouette Score sebesar 0,6164 dan DBI 0,5684 menunjukkan pemisahan cluster yang jelas dan kompak. Nilai ini lebih baik daripada model berbasis variabel tunggal seperti Darsono et al. (2022) maupun model campuran terbatas seperti Setiawan & Mailoa (2025). Keunggulan ini terutama berasal dari integrasi enam semester nilai akademik dengan empat instrumen non-akademik, yang memberikan representasi multidimensi lebih kaya dan responsif terhadap variasi karakter siswa. Dengan demikian, model terbukti lebih akurat dalam memetakan profil siswa dibanding pendekatan konvensional yang hanya berbasis nilai atau minat.

Validitas praktis juga terkonfirmasi melalui penilaian guru Bimbingan dan Konseling yang menunjukkan tingkat kesesuaian rekomendasi sebesar 80%. Hal ini menegaskan bahwa model tidak hanya akurat secara komputasional, tetapi juga relevan secara edukatif serta dapat digunakan untuk mendukung keputusan penjurusan berbasis data. Implementasi sistem berbasis web memungkinkan proses rekomendasi menjadi lebih objektif, mengurangi bias subjektif, dan meningkatkan konsistensi layanan konseling akademik. Temuan ini memperkuat argumen bahwa *data-driven decision making* dapat meningkatkan kualitas penjurusan dan meminimalkan risiko *misfit* studi lanjut.

Meski demikian, penelitian ini memiliki keterbatasan, di antaranya cakupan data yang hanya berasal dari satu sekolah dan belum adanya validasi longitudinal terhadap keberhasilan mahasiswa setelah memasuki perguruan tinggi. Selain itu, instrumen non-akademik belum diuji reliabilitas psikometriknya secara luas. Studi lanjutan perlu mencakup sampel multi-

lokasi, uji validitas dan reliabilitas instrumen secara formal, serta pengembangan model adaptif berbasis pembaruan data periodik.

Hasil penelitian ini memberikan kontribusi penting bagi literatur *educational data mining* melalui pengembangan model clustering multidimensi yang memadukan teori psikologis dan pendekatan *machine learning*. Temuan memperlihatkan bahwa integrasi variabel kognitif, afektif, dan konatif menghasilkan rekomendasi program studi yang lebih akurat, personal, dan potensial untuk diterapkan dalam konteks pendidikan menengah.

SIMPULAN

Sistem Rekomendasi Jurusan Impian (SRJI) berbasis *K-Means Clustering* berhasil dikembangkan dengan mengintegrasikan nilai akademik enam semester dan empat asesmen non-akademik yaitu tes kemampuan, gaya belajar, minat, serta kepribadian. Algoritma K-Means mengelompokkan siswa ke dalam tiga cluster berkualitas tinggi (Silhouette Score 0.6164, Davies-Bouldin Index 0.5684), yang merepresentasikan pola kecenderungan social-enterprising, investigative-realistic, dan realistic-conventional. Validasi menunjukkan bahwa sistem memberikan rekomendan program studi yang konsisten dengan profil siswa, dengan tingkat kesesuaian 80% berdasarkan penilaian Guru Bimbingan Konseling. Model Integrated Multidimensional Student Profiling menangkap karakteristik siswa secara lebih komprehensif dibandingkan pendekatan tunggal atau parsial, sehingga menyediakan dasar empiris untuk sistem rekomendasi yang objektif, personal, dan mengurangi bias pengambilan keputusan. Keterbatasan penelitian meliputi cakupan satu sekolah dan belum tervalidasi secara psikometrik, sehingga studi lanjutan disarankan untuk memperluas data secara multi-site, melakukan evaluasi longitudinal terhadap prestasi perguruan tinggi, serta mengembangkan sistem adaptif yang memperbarui model berkala. *Integrated Multidimensional Student Profiling* berpotensi memperkuat praktik penjurusan berbasis data dan membuka arah baru bagi penelitian educational data mining multidimensional.

REFERENSI

Aisyah, J., Risnasari, M., & Ni'mah, A. T. (2023). Sistem rekomendasi program studi menggunakan metode hybrid recommendation (studi kasus: MAN Sumenep). *Journal of Education and Informatics*, 4(1), 1–10. <https://doi.org/10.30864/eksplora.v12i1.992>

Arifin, I., Primayasa, W., & Baharsyah, M. Y. (2020). Pengaruh salah pilih jurusan terhadap rasa putus asa mahasiswa Teknik Informatika. *Nathiqiyah Jurnal Psikologi Islam*, 3(1), 1-5. <https://doi.org/10.46781/nathiqiyah.v3i1.76>

Basalamah, A., & Setyadi, R. (2023). Penerapan algoritma K-Means clustering pada tingkat penyelesaian pendidikan di Provinsi Indonesia. *Jurnal Informatika dan Teknologi Komputer (J-ICOM)*, 4(2), 114-121. <https://doi.org/10.55377/j-icom.v4i2.7893>

Darsono, V., Amroni, & Andrianti, A. (2022). Penerapan data mining algoritma k-means untuk rekomendasi pemilihan bidang studi perguruan tinggi pada siswa SMKN 1 Kota Jambi. *Jurnal Informatika Dan Rekayasa Komputer(JAKAKOM)*, 2(2), 161–171. <https://doi.org/10.33998/jakakom.2022.2.2.80>

Diana, D., Guntur, I., Roebianto, A., & Christy, C. (2023). Choosing the Wrong Major: what is the profile of students who feel they have chosen the wrong major?. *Jurnal Psikologi Pendidikan dan Konseling: Jurnal Kajian Psikologi Pendidikan dan Bimbingan Konseling (JPPK)*, 9(2), 77-85. <https://doi.org/10.26858/jppk.v9i2.45753>

Eirlangga, Y. S., & Syaputra, A. E. (2022). Analisis pengelompokan data nilai siswa untuk menentukan siswa berprestasi menggunakan metode clustering k-means. *Jurnal Informasi Dan Teknologi*, 4(3), 160–165. <https://doi.org/10.37034/jidt.v4i3.235>

Hidayat, R. (2022). Pemanfaatan data mining untuk melihat minat siswa setelah menyelesaikan pendidikan sekolah menengah atas (SMA) dengan algoritma k-means clustering.

Technology and Informatics Insight Journal, 1(2), 136–151.
<https://doi.org/10.32639/tijj.v1i2.220>

Kao, S., Chea, P., & Song, S. (2023). Upper secondary school tracking and major choices in higher education : to switch or not to switch. *Educational Research for Policy and Practice*, 23(0123456789), 89–113. <https://doi.org/10.1007/s10671-023-09356-1>

Karmanita, D., & Hendrik, B. (2023). Penerapan metode clustering dengan algoritma k-means pada pengelompokan peminatan mata kuliah. *Jurnal Ilmiah Dan Karya Mahasiswa*, 1(6), 1–10. <https://doi.org/10.54066/jikma.v1i6.1028>

Khoirunnisa, C. V., & Arkhiansyah, Y. (2024). Implementasi algoritma machine learning untuk rekomendasi program studi bagi siswa SMA (Studi Kasus : Institut Informatika dan Bisnis Darmajaya). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(5), 10102–10108. <https://doi.org/10.36040/jati.v8i5.10866>

Kolne, M. A., Kelen, Y. P. K., & Ullu, H. H. (2025). Analisis kelayakan pemilihan jurusan siswa-siswi SMA/sederajat di era merdeka belajar menggunakan algoritma naïve bayes. *Jurnal Sistem Informasi*, 7(1), 149–161.

Lubis, A. S., & Sriani. (2025). Sistem pendukung keputusan dalam menentukan jurusan siswa di sekolah menengah menggunakan metode fuzzy tahani. *Edumatic : Jurnal Pendidikan Informatika*, 9(2), 543–551. <https://doi.org/10.29408/edumatic.v9i2.30722>

Ramadania, M. S., & Fatah, Z. (2024). Analisis pengelompokan data nilai siswa untuk menentukan siswa berprestasi menggunakan metode clustering k-means. *Jurnal Riset Sistem Informasi*, 1(4), 103–110.

Saputra, E. A., & Nataliani, Y. (2021). Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *Journal of Information Systems and Informatics*, 3(3), 424-439. <https://doi.org/10.51519/journalisi.v3i3.164>

Setiawan, K., & Mailoa, E. (2025). Pengelompokan gaya belajar siswa untuk mendukung efektifitas pembelajaran dengan k-means. *JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 16(2), 26–35.

Silver, B. R. (2023). Major transitions : how college students interpret the process of changing fields of study. *Higher Education*, 87(0123456789), 1–20. <https://doi.org/10.1007/s10734-023-01050-8>

Syahril, M., Kusnasari, S., Sobirin, Muhamzir, A., & Syahputri, A. (2023). Implementasi data mining untuk rekomendasi jurusan menggunakan algoritma k-means clustering. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 6(1), 235–245. <https://doi.org/10.53513/jsk.v6i1.7456>

Utama, A., Utama, S., Sabilla, W. I., Wakhidah, R., A, W. I., & Wakhidah, R. (2024). Sistem rekomendasi tempat wisata di malang raya menggunakan metode k-means clustering. *Jurnal Informatika & Multimedia*, 16(1), 1–15. <https://doi.org/10.33795/jtim.v16i1.5178>

Walangare, R. A. C., & Sujatmiko, B. (2022). Penerapan Algoritma Naive Bayes Dalam Sistem Pendukung Keputusan Pemilihan Peminatan Konsentrasi Berdasarkan Nilai Akademik Berbasis Web Pada Program Studi S1 Pendidikan Teknologi Informasi. *IT-Edu: Jurnal Information Technology and Education*, 7(3), 74-83. <https://doi.org/10.26740/itedu.v7i3.50086>

Zhao, Y., Zhao, C., Wang, Z., & Min, Z. (2024). Cluster analysis and its application in teaching resources of university curriculum: A personalized method. *Computer Applications in Engineering Education*, 32(2), e22696. <https://doi.org/10.1002/cae.22696>