Jurnal Ilmu Komputer dan Sistem Informasi (JIRSI) **e-ISSN 2830-3954**
Volume: 5, Nomor: 2, Mei 2026: 198-208 **p-ISSN 2830-6031**
https://jurnal.unity-academy.sch.id/index.php/jirsi/index

Analisis Kejelasan Tujuan Pendidikan Remaja Berbasis Machine Learning dan Data Digital Psikometrik

Analysis of Educational Goal Clarity among Adolescents Using Machine Learning and Psychometric Digital Data

Dwi Wulandari Sari*[1], Kurnia Gusti Ayu[2], Dana Riksa Buana[3], Hariesa Budi Prabowo[4], Muhammad Irvan[5] (* corespondent author)
[1,2,4,5] Fakultas Ilmu Komputer/Sistem Informasi, [3] Fakultas Psikologi/Psikologi Universitas Mercu Buana
E-mail: [1]dwi.wulandari@mercubuana.ac.id, [2]kurnia.gusti@mercubuana.ac.id, [3]dana.riksa@mercubuana.ac.id, [4]hariesa@mercubuana.ac.id, [5]muhammad.irvan@mercubuana.ac.id,

Abstrak

Penelitian ini bertujuan untuk menganalisis kejelasan tujuan pendidikan remaja berdasarkan integrasi data psikometrik dan aktivitas digital menggunakan pendekatan machine learning. Permasalahan utama dalam penelitian ini adalah belum optimalnya pemanfaatan data digital dan psikologis untuk memahami karakteristik tujuan pendidikan remaja. Metodologi yang digunakan meliputi preprocessing data, analisis clustering menggunakan algoritma K-Means untuk mengidentifikasi pola tersembunyi, serta klasifikasi menggunakan Logistic Regression dan Decision Tree untuk membangun model prediksi. Hasil penelitian menunjukkan bahwa data responden terbagi menjadi tiga kelompok utama, yaitu rendah, sedang, dan tinggi, yang mencerminkan tingkat keterlibatan digital dan kejelasan tujuan pendidikan. Dengan pembagian data 80:20, menunjukkan bahwa pada data test (20%), model Logistic Regression menghasilkan performa terbaik dengan akurasi sebesar 0.95 dan ROC-AUC sebesar 0.996, sedangkan Decision Tree memberikan interpretasi pola yang lebih mudah dipahami dengan akurasi sebesar 0.80. Variabel yang paling berpengaruh meliputi frekuensi pencarian informasi, durasi akses konten edukasi, dan hasil clustering. Kesimpulan penelitian ini menunjukkan bahwa perilaku digital yang produktif berkontribusi signifikan terhadap kejelasan tujuan pendidikan remaja, serta pendekatan hybrid machine learning efektif dalam menggabungkan analisis pola dan prediksi.

***Kata kunci**: machine learning, clustering, klasifikasi, perilaku digital, tujuan pendidikan*

Abstract

This study aims to analyze adolescents' educational goal clarity based on the integration of psychometric and digital activity data using a machine learning approach. The main problem addressed in this study is the limited utilization of digital and psychological data to understand adolescents' educational goal characteristics. The methodology includes data preprocessing, clustering analysis using the K-Means algorithm to identify hidden patterns, and classification using Logistic Regression and Decision Tree to build predictive models. The results show that respondents are grouped into three main clusters: low, medium, and high, reflecting different levels of digital engagement and educational goal clarity. Using an 80:20 data split, the results show that on the test set (20%), Logistic Regression achieved the best performance with an accuracy of 0.95 and ROC-AUC of 0.996, while Decision Tree provided more interpretable patterns with an accuracy of 0.80. The most influential variables include frequency of information searching, duration of accessing educational content, and clustering results. The study concludes that productive digital behavior significantly contributes to adolescents'

educational goal clarity, and the hybrid machine learning approach is effective in combining pattern analysis and prediction.

***Keywords**: machine learning, clustering, classification, digital behavior, educational goals*

1. PENDAHULUAN

Masa remaja merupakan fase krusial dalam pembentukan identitas diri serta pengambilan keputusan strategis terkait masa depan pendidikan dan karier. Pada tahap ini, kejelasan tujuan pendidikan menjadi faktor penting yang mempengaruhi keberhasilan transisi menuju pendidikan tinggi maupun dunia kerja. Kejelasan tujuan pendidikan dapat diartikan sebagai pemahaman individu terhadap arah akademik yang akan ditempuh, termasuk pemilihan program studi dan pengembangan kompetensi yang sesuai dengan minat, bakat, serta aspirasi karier [1], [4]. Remaja yang memiliki tujuan pendidikan yang jelas cenderung memiliki motivasi belajar yang lebih tinggi, resiliensi akademik yang baik, serta kemampuan pengambilan keputusan yang lebih matang [2], [10].

Namun demikian, fenomena yang terjadi menunjukkan bahwa masih banyak remaja yang mengalami kebingungan dalam menentukan arah pendidikan. Kondisi ini dapat menyebabkan berbagai permasalahan, seperti ketidaksesuaian pilihan jurusan (*academic misalignment*), rendahnya motivasi belajar, hingga meningkatnya kecemasan terhadap masa depan (*future anxiety*) [3], [4]. Faktor-faktor yang mempengaruhi kondisi tersebut tidak hanya berasal dari aspek psikologis seperti *self-efficacy* dan motivasi belajar, tetapi juga dipengaruhi oleh lingkungan sosial serta pola aktivitas digital yang semakin dominan di era transformasi digital saat ini [1], [2].

Perkembangan teknologi digital telah mengubah cara remaja dalam memperoleh informasi dan membentuk preferensi pendidikan. Aktivitas digital seperti pencarian informasi, konsumsi konten edukatif, serta penggunaan media sosial dapat menjadi indikator penting dalam memahami karakteristik dan orientasi masa depan remaja [1], [2]. Jejak digital (*digital footprint*) yang dihasilkan dari aktivitas tersebut memberikan peluang baru untuk dianalisis secara lebih mendalam menggunakan pendekatan berbasis data (*data-driven approach*) [3]. Dalam konteks ini, pemanfaatan teknologi *machine learning* menjadi solusi yang relevan untuk mengidentifikasi pola tersembunyi (*hidden patterns*) serta membangun model prediktif dalam bidang pendidikan [5], [8].

Penelitian sebelumnya menunjukkan bahwa algoritma klasifikasi seperti *Decision Tree*, *Random Forest*, dan *Support Vector Machine* mampu memprediksi performa akademik dan pilihan pendidikan dengan tingkat akurasi yang baik [5], [6], [8], [10], [13], [16]. Selain itu, integrasi faktor psikometrik dalam model prediktif juga terbukti meningkatkan performa klasifikasi dalam analisis pendidikan [7], [11], [15]. Berbagai penelitian juga telah mengembangkan sistem rekomendasi berbasis kecerdasan buatan untuk membantu pengambilan keputusan pendidikan [9], [12], [14]. Namun, sebagian besar penelitian masih berfokus pada data akademik formal dan variabel psikometrik statis, sehingga belum sepenuhnya mampu menangkap dinamika perilaku remaja di era digital.

Berdasarkan hal tersebut, terdapat celah penelitian yang signifikan dalam mengintegrasikan data psikometrik dengan data aktivitas digital untuk menganalisis kejelasan tujuan pendidikan remaja secara lebih komprehensif.

Penelitian ini menawarkan pendekatan inovatif melalui integrasi data multimodal yang mencakup faktor internal (psikometrik) dan faktor eksternal (aktivitas digital) dengan memanfaatkan algoritma *machine learning*. Pendekatan ini diharapkan mampu memberikan pemahaman yang lebih akurat terhadap karakteristik remaja serta meningkatkan performa model prediksi.

Secara metodologis, penelitian ini menggunakan kombinasi pendekatan *unsupervised learning* dan *supervised learning*. Algoritma *K-Means clustering* digunakan untuk melakukan segmentasi karakteristik remaja berdasarkan kemiripan pola perilaku, sedangkan algoritma *Logistic Regression* dan *Decision Tree* digunakan untuk membangun model klasifikasi tingkat kejelasan tujuan pendidikan. Pendekatan *hybrid* ini memungkinkan analisis yang tidak hanya bersifat prediktif, tetapi juga eksploratif dan interpretatif.

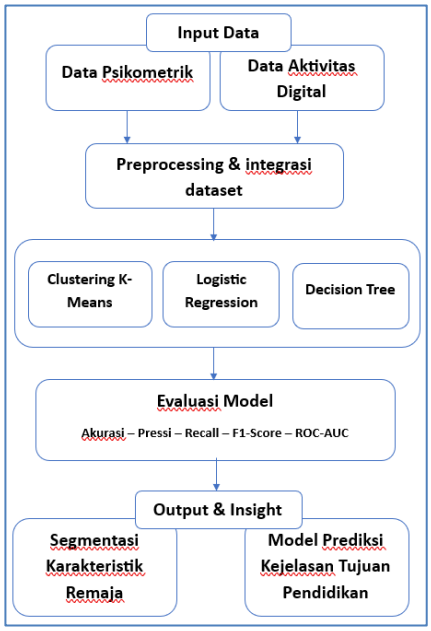
Berdasarkan uraian tersebut, rumusan masalah dalam penelitian ini meliputi: (1) bagaimana karakteristik kejelasan tujuan pendidikan remaja berdasarkan integrasi data psikometrik dan aktivitas digital; (2) variabel apa yang paling berpengaruh terhadap kejelasan tujuan pendidikan; (3) bagaimana performa model *machine learning* dalam mengklasifikasikan tingkat kejelasan tujuan pendidikan; serta (4) bagaimana segmentasi karakteristik remaja berdasarkan hasil clustering dan klasifikasi.

Adapun nilai kebaruan (*novelty*) dari penelitian ini terletak pada integrasi data psikometrik dan aktivitas digital dalam satu kerangka analitik berbasis *machine learning*, yang belum banyak dikaji secara komprehensif dalam penelitian sebelumnya. Penelitian ini tidak hanya menghasilkan model prediksi yang akurat, tetapi juga memberikan insight mengenai pola perilaku remaja yang dapat digunakan sebagai dasar pengambilan keputusan dalam bidang pendidikan.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis *data-driven* dengan metode *machine learning* untuk menganalisis karakteristik kejelasan tujuan pendidikan remaja melalui integrasi data psikometrik dan aktivitas digital. Desain penelitian bersifat *explanatory predictive*, yaitu tidak hanya mengidentifikasi pola tetapi juga membangun model prediksi yang mampu menjelaskan faktor-faktor yang memengaruhi kejelasan tujuan pendidikan.

Tahapan penelitian disusun secara sistematis mulai dari pengumpulan data, *preprocessing*, integrasi data, pembangunan model, hingga evaluasi dan interpretasi hasil. Diagram alir penelitian yang menggambarkan keseluruhan proses dapat dilihat pada Gambar 1.



Gambar 1. Diagram Penelitian.

2.1. Desain dan Pengumpulan Data

Penelitian ini menggunakan data primer yang diperoleh melalui penyebaran kuesioner kepada responden siswa SMA/sederajat dengan jumlah minimal n ≥ 300. Data yang dikumpulkan terdiri dari dua jenis utama, yaitu:

1. Data Psikometrik, meliputi:
    - *Self-efficacy* akademik
    - Motivasi belajar
    - Dukungan sosial
    - Literasi digital
2. Data Aktivitas Digital, meliputi:
    - Durasi penggunaan media digital
    - Frekuensi
    - Rasio konten
    - Partisipasi online
    - Aplikasi produktivitas

Instrumen penelitian disusun menggunakan skala Likert dan telah disesuaikan dengan konstruk variabel penelitian. Data dikumpulkan melalui platform digital (*online survey*) untuk memastikan efisiensi dan jangkauan responden yang lebih luas.

2.2. Prosedur Penelitian dan Pemodelan

Prosedur penelitian dilakukan melalui beberapa tahapan utama sebagai berikut:

1. Preprocessing & Integrasi: Untuk memastikan model dapat mengidentifikasi pola secara akurat, dilakukan tahap Feature Engineering yang bekerja dengan menggabungkan data statis (psikometrik) dan data dinamis (aktivitas digital) yang telah dinormalisasi.

2. Pengembangan Model: Pada tahap ini dilakukan pembangunan model analitik menggunakan algoritma machine learning yaitu K-means, Logistic Regression dan Decision Tree. Dimana :
   - Segmentasi Pola dengan *Clustering* K-Means: Penentuan jumlah cluster optimal dilakukan menggunakan metode *Elbow* dengan menganalisis nilai *Within Cluster Sum of Squares* (WCSS) untuk menemukan titik dimana penurunan mulai melandai. Selanjutnya, algoritma K-Means digunakan untuk mengidentifikasi pola tersembunyi dengan mengelompokkan responden berdasarkan kedekatan jarak Euclidean dari variabel multimodal yang terdiri dari aspek psikometrik (self-efficacy, motivasi belajar, dukungan sosial, dan literasi digital) serta aktivitas digital (durasi akses edukatif, frekuensi pencarian informasi, rasio konsumsi konten, partisipasi online, dan penggunaan aplikasi produktivitas). Seluruh variabel dinormalisasi sebelum proses clustering untuk memastikan kontribusi yang seimbang dalam perhitungan jarak, sehingga menghasilkan segmentasi karakteristik remaja yang menjadi dasar pelabelan pada tahap klasifikasi.
   - Klasifikasi dan Prediksi dengan *Logistic Regression*: Model ini digunakan untuk mengidentifikasi variabel mana yang memiliki pengaruh (koefisien) paling signifikan terhadap kejelasan tujuan pendidikan. Dengan fungsi *sigmoid*, model ini memetakan probabilitas keterhubungan antara variabel digital (misalnya: frekuensi pencarian informasi karir) dan variabel psikometrik terhadap tingkat kesiapan siswa.
   - Pemetaan Keputusan dengan *Decision Tree*: Algoritma ini dipilih karena kemampuannya dalam memodelkan kombinasi variabel ke dalam aturan keputusan (*decision rules*) yang mudah diinterpretasikan. Model ini akan membangun pohon keputusan yang secara otomatis menyeleksi variabel paling informatif (*Information Gain*) untuk memecah data menjadi sub-kelompok yang lebih homogen, sehingga dapat terlihat jelas jalur bagaimana kombinasi aktivitas digital tertentu dapat memprediksi arah pendidikan yang jelas.
3. Evaluasi Model: Metrik performa utama (Akurasi, Presisi, Recall, F1-Score, dan ROC-AUC) digunakan untuk memberikan kejelasan visual mengenai bagaimana model divalidasi.
4. Output & Insight: Penelitian ini menghasilkan dua keluaran utama, yaitu segmentasi karakteristik remaja berbasis data serta model prediksi kejelasan tujuan pendidikan. Segmentasi diperoleh melalui algoritma *K-Means* sedangkan Selanjutnya, model prediksi dibangun menggunakan *Logistic Regression* dan *Decision Tree.* Kontribusi utama penelitian ini terletak pada penerapan pendekatan hybrid machine learning yang mengintegrasikan proses segmentasi dan klasifikasi dalam satu kerangka analitik. Selain itu, penelitian ini memperkenalkan penggunaan data multimodal yang menggabungkan aspek psikometrik dan aktivitas digital untuk menganalisis kejelasan tujuan pendidikan remaja. Pendekatan ini tidak hanya meningkatkan akurasi prediksi,

tetapi juga menghasilkan insight yang relevan untuk mendukung pengambilan keputusan dalam bidang pendidikan.

3. HASIL DAN PEMBAHASAN

3.1. Data Preprocessing dan Integrasi

Tahap awal dilakukan pembersihan data (*data cleaning*), normalisasi, serta transformasi variabel untuk memastikan kualitas data. Selanjutnya dilakukan *feature engineering* dengan mengintegrasikan data psikometrik dan aktivitas digital menjadi dataset terpadu (*multimodal dataset*).

Tipe_Sekolah	Jurusan	A1	A2	A3	M1	M2	M3	S1	S2	S3	L1	L2	L3	D1_Durasi	D2_Frekue	D3_Rasio_	D4_Partisi	D5_App_P	Target_Kej
SMA	IPA	3	3	4	4	4	4	4	3	3	4	4	4	3	3	2	1	2	Tinggi
SMA	Bahasa	3	3	3	3	3	3	4	3	2	4	4	4	2	1	0	1	0	Sedang
SMK	Farmasi	4	4	3	4	4	4	3	4	3	4	4	4	1	1	1	1	2	Sedang
SMA	Bahasa	5	5	4	5	4	5	4	4	3	5	5	4	3	3	2	2	2	Tinggi
SMK	Multimedia	3	3	3	3	4	4	4	3	3	4	4	3	1	1	0	0	0	Sedang
SMA	Bahasa	3	3	4	4	3	3	3	2	4	4	3	2	2	1	2	0	0	Sedang
SMA	IPA	4	4	4	5	4	5	4	4	4	4	4	5	3	3	2	2	2	Tinggi
SMK	Multimedia	4	5	5	4	4	5	4	3	4	5	3	3	1	2	2	2	2	Tinggi
SMK	Farmasi	4	2	2	3	3	3	2	3	2	4	3	3	1	0	1	1	0	Rendah

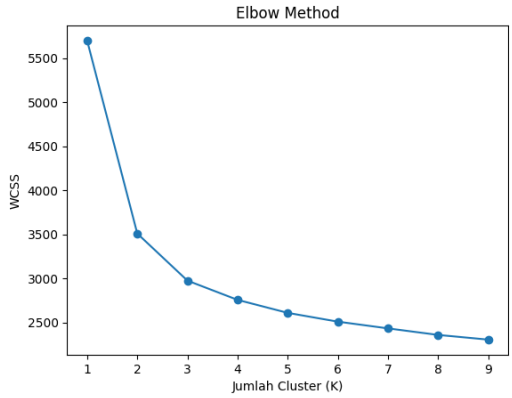
Gambar 2. Data Integrasi antara psikometri dan aktivitas digital

Tipe_Seko	Jurusan_E	A1	A2	A3	M1	M2	M3	S1	S2	S3	L1	L2	L3	D1_Durasi	D2_Frekue	D3_Rasio_	D4_Partisi	D5_App_P	Target_Enc
0	3	-0.25315	-0.14012	0.834234	0.612165	0.845104	0.524671	1.159858	-0.05999	0.057831	0.471763	0.647946	0.838089	1.407952	1.31648	1.245668	0.039086	1.299588	2
0	1	-0.25315	-0.14012	-0.33525	-0.38593	-0.2157	-0.49082	1.159858	-0.05999	-1.18141	0.471763	0.647946	0.838089	0.467227	-0.39323	-1.18096	0.039086	-1.21574	1
1	2	0.81651	0.885146	-0.33525	0.612165	0.845104	0.524671	0.007681	1.064754	0.057831	0.471763	0.647946	0.838089	-0.4735	-0.39323	0.032355	0.039086	1.299588	1
0	1	1.886173	1.910411	0.834234	1.610259	0.845104	1.540162	1.159858	1.064754	0.057831	1.577457	1.69302	0.838089	1.407952	1.31648	1.245668	1.34197	1.299588	2
1	5	-0.25315	-0.14012	-0.33525	-0.38593	0.845104	0.524671	1.159858	-0.05999	0.057831	0.471763	0.647946	-0.27936	-0.4735	-0.39323	-1.18096	-1.2638	-1.21574	1
0	1	-0.25315	-0.14012	0.834234	0.612165	-0.2157	-0.49082	0.007681	-1.18473	1.297078	0.471763	-0.39713	-1.39682	0.467227	-0.39323	1.245668	-1.2638	-1.21574	1
0	3	0.81651	0.885146	0.834234	1.610259	0.845104	1.540162	1.159858	1.064754	1.297078	0.471763	0.647946	1.955541	1.407952	1.31648	1.245668	1.34197	1.299588	2
1	5	0.81651	1.910411	2.00372	0.612165	0.845104	1.540162	1.159858	-0.05999	1.297078	1.577457	-0.39713	-0.27936	-0.4735	0.461623	1.245668	1.34197	1.299588	2
1	2	0.81651	-1.16538	-1.50474	-0.38593	-0.2157	-0.49082	-1.1445	-0.05999	-1.18141	0.471763	-0.39713	-0.27936	-0.4735	-1.24809	0.032355	0.039086	-1.21574	0

Gambar 3. Data Integrasi hasil preprocessing (*multimodal dataset*)

3.2. Clustering dengan K-Means

Algoritma *K-Means clustering* digunakan untuk mengelompokkan responden berdasarkan kemiripan karakteristik. Penentuan jumlah cluster optimal dilakukan menggunakan metode *Elbow* dengan menghitung nilai *Within Cluster Sum of Squares (WCSS)*.



Gambar 4. Grafik *Elbow* hasil dari pengujian data

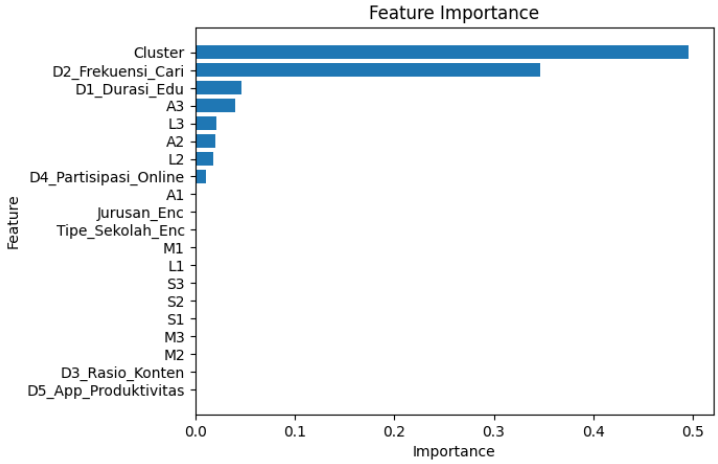
Berdasarkan hasil analisis menggunakan metode *Elbow*, terlihat bahwa nilai *Within Cluster Sum of Squares* (WCSS) mengalami penurunan signifikan dari K=1 hingga K=3, kemudian cenderung melandai setelahnya. Hal ini menunjukkan bahwa jumlah cluster optimal berada pada K=3. Pemilihan K=3 juga konsisten dengan jumlah kategori pada variabel target, yaitu tinggi, sedang, dan rendah. Dengan demikian, hasil clustering menunjukkan bahwa pola alami data sejalan dengan struktur kelas yang ada.

Tipe_Seko	Jurusan_E	A1	A2	A3	M1	M2	M3	S1	S2	S3	L1	L2	L3	D1_Durasi	D2_Frekue	D3_Rasio_	D4_Partisi	D5_App_P	Target_En	Cluster
0	3	-0.25315	-0.14012	0.834234	0.612165	0.845104	0.524671	1.159858	-0.05999	0.057831	0.471763	0.647946	0.838089	1.407952	1.31648	1.245668	0.039086	1.299588	2	2
0	1	-0.25315	-0.14012	-0.33525	-0.38593	-0.2157	-0.49082	1.159858	-0.05999	-1.18141	0.471763	0.647946	0.838089	0.467227	-0.39323	-1.18096	0.039086	-1.21574	1	0
1	2	0.81651	0.885146	-0.33525	0.612165	0.845104	0.524671	0.007681	1.064754	0.057831	0.471763	0.647946	0.838089	-0.4735	-0.39323	0.032355	0.039086	1.299588	1	0
0	1	1.886173	1.910411	0.834234	1.610259	0.845104	1.540162	1.159858	1.064754	0.057831	1.577457	1.69302	0.838089	1.407952	1.31648	1.245668	1.34197	1.299588	2	2
1	5	-0.25315	-0.14012	-0.33525	-0.38593	0.845104	0.524671	1.159858	-0.05999	0.057831	0.471763	0.647946	-0.27936	-0.4735	-0.39323	-1.18096	-1.2638	-1.21574	1	0
0	1	-0.25315	-0.14012	0.834234	0.612165	-0.2157	-0.49082	0.007681	-1.18473	1.297078	0.471763	-0.39713	-1.39682	0.467227	-0.39323	1.245668	-1.2638	-1.21574	1	0
0	3	0.81651	0.885146	0.834234	1.610259	0.845104	1.540162	1.159858	1.064754	1.297078	0.471763	0.647946	1.955541	1.407952	1.31648	1.245668	1.34197	1.299588	2	2
1	5	0.81651	1.910411	2.00372	0.612165	0.845104	1.540162	1.159858	-0.05999	1.297078	1.577457	-0.39713	-0.27936	-0.4735	0.461623	1.245668	1.34197	1.299588	2	2
1	2	0.81651	-1.16538	-1.50474	-0.38593	-0.2157	-0.49082	-1.1445	-0.05999	-1.18141	0.471763	-0.39713	-0.27936	-0.4735	-1.24809	0.032355	0.039086	-1.21574	0	1

Gambar 5. Hasil pemetaan cluster ke *multimodal dataset*

3. 3 Klasifikasi dengan Logistic Regression

Model *Logistic Regression* digunakan untuk memprediksi probabilitas tingkat kejelasan tujuan pendidikan (tinggi, sedang, rendah). Model ini juga digunakan untuk mengidentifikasi variabel yang paling berpengaruh melalui nilai koefisien.



Gambar 6. *Feature Importance* hasil olah data multimodal dataset

Tabel koefisien asli Logistic Regression yang didapatkan dari hasil olah dataset sebagai berikut :

Tabel 1. Koefisien Logistic Regression

Rank	**Variabel**	**Makna**
1	Cluster	Paling dominan
2	D2_Frekuensi_Cari	Perilaku digital
3	D1_Durasi_Edu	Intensitas belajar
4	A3	Faktor psikometrik
5	L3	Variabel tambahan

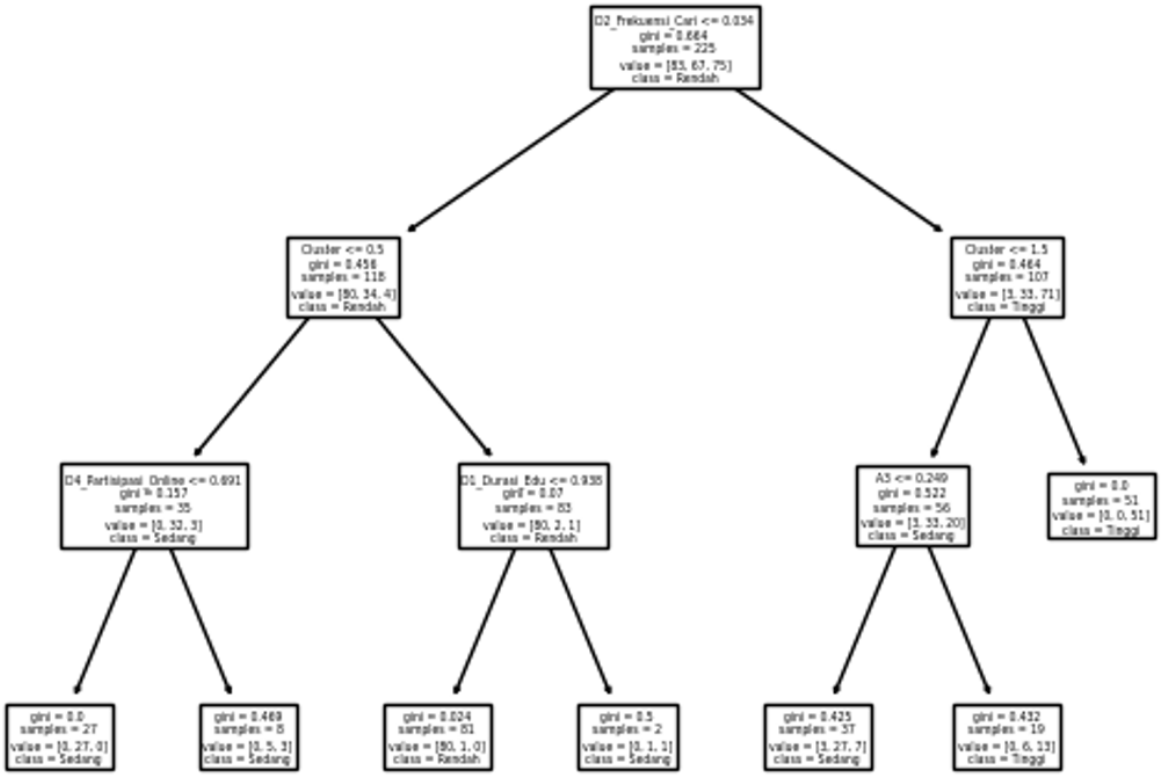
Hasil analisis menggunakan algoritma *Logistic Regression* menunjukkan bahwa beberapa variabel memiliki kontribusi signifikan dalam menentukan tingkat kejelasan tujuan pendidikan remaja. Variabel yang paling dominan adalah *Cluster*, hasil segmentasi menggunakan algoritma K-Means, yang menunjukkan bahwa pola karakteristik responden berperan penting dalam meningkatkan akurasi model. Selanjutnya, variabel *D2_Frekuensi_Cari* menjadi faktor penting yang merepresentasikan perilaku digital dalam mencari informasi, diikuti oleh *D1_Durasi_Edu* yang mencerminkan intensitas akses terhadap konten edukatif. Kedua variabel ini menunjukkan bahwa aktivitas digital yang bersifat eksploratif dan edukatif berkontribusi besar terhadap kejelasan tujuan pendidikan.

Selain faktor digital, variabel psikometrik seperti *A3* juga memberikan kontribusi dalam model, yang menunjukkan pentingnya aspek internal seperti *self-efficacy* dalam menentukan arah pendidikan. Variabel tambahan *L3* turut berperan

meskipun tidak sebesar variabel utama. Secara keseluruhan, hasil ini menunjukkan bahwa kombinasi antara faktor eksternal (aktivitas digital) dan faktor internal (psikometrik) memiliki peran penting dalam membentuk kejelasan tujuan pendidikan remaja, serta menegaskan bahwa pendekatan hybrid yang menggabungkan *clustering* dan *classification* mampu menghasilkan model yang lebih optimal dan komprehensif.

3.4 Klasifikasi dengan Decision Tree

Algoritma *Decision Tree* digunakan untuk membentuk aturan keputusan (*decision rules*) yang mudah diinterpretasikan. Model ini menghasilkan struktur pohon keputusan yang menggambarkan hubungan antar variabel dalam menentukan kejelasan tujuan pendidikan.



Gambar 7. Decision Tree

Visualisasi Decision Tree menunjukkan bahwa variabel frekuensi pencarian informasi (D2_Frekuensi_Cari) menjadi faktor utama dalam proses klasifikasi. Model kemudian membentuk percabangan berdasarkan variabel cluster, durasi akses konten edukasi, serta faktor psikometrik. Struktur pohon ini menunjukkan bahwa keputusan klasifikasi tidak hanya dipengaruhi oleh satu variabel, tetapi oleh kombinasi beberapa faktor yang saling berinteraksi.

Selain itu, pola yang terbentuk menunjukkan bahwa responden dengan frekuensi pencarian tinggi dan tingkat *self-efficacy* yang baik cenderung berada pada kategori tujuan pendidikan yang tinggi, sedangkan responden dengan aktivitas digital rendah cenderung berada pada kategori rendah.

4. Evaluasi Model.

Evaluasi performa model dilakukan menggunakan beberapa metrik berikut: Akurasi (*Accuracy*), Presisi (*Precision*), Recall, F1-Score, ROC-AUC. Proses evaluasi

model dilakukan dengan membagi dataset menggunakan metode *train-test split* dengan rasio 80:20. Selanjutnya, untuk memastikan kestabilan dan generalisasi model, dilakukan *5-fold cross-validation*.

Tabel 2. Matrix Evaluasi Model

Model	***Accuracy***	***Precision***	**Recall**	**F1-Score**	**ROC-AUC**	**CV Accuracy (5-fold) ± std**
Logistic Regression	0.95	0.951	0.95	0.950	0.996	0.967 ± 0.02
Decision Tree	0.80	0.799	0.80	0.799	0.851	0.877 ± 0.05

Hasil menunjukkan bahwa Logistic Regression mencapai akurasi sebesar 0,95 pada data uji, dengan rata-rata akurasi *cross-validation* sebesar 0,967 ± 0,02. Sementara itu, Decision Tree memperoleh akurasi sebesar 0,80 pada data uji dan 0,877 ± 0,05 pada *cross-validation*. Pelaporan dalam bentuk rata-rata dan deviasi standar menunjukkan kestabilan model pada setiap fold.Pendekatan ini memungkinkan evaluasi performa model secara lebih komprehensif dan mengurangi bias akibat pembagian data tunggal. Karena penelitian ini melibatkan tiga kelas (rendah, sedang, dan tinggi), perhitungan ROC-AUC dilakukan menggunakan pendekatan *one-vs-rest (OvR)*, dimana setiap kelas dibandingkan terhadap kelas lainnya. Nilai ROC-AUC kemudian dihitung untuk masing-masing kelas dan dirata-ratakan menggunakan metode *macro-average*. Nilai ROC-AUC sebesar 0,996 menunjukkan rata-rata kemampuan model dalam membedakan setiap kelas secara sangat baik.
Sementara itu, *Decision Tree* meskipun memiliki performa yang lebih rendah, model ini memberikan keunggulan dalam interpretasi pola melalui struktur pohon keputusan.
Berdasarkan confusion matrix, kedua model mampu mengklasifikasikan data dengan baik, namun kesalahan prediksi lebih banyak terjadi pada kelas menengah.
Dengan demikian, *Logistic Regression* dipilih sebagai model utama, sedangkan *Decision Tree* digunakan sebagai model pendukung untuk interpretasi pola.

5.Output dan Insight

Penelitian ini menghasilkan dua output utama, yaitu segmentasi karakteristik remaja menjadi tiga kelompok (rendah, sedang, tinggi) menggunakan K-Means, serta model prediksi kejelasan tujuan pendidikan menggunakan *Logistic Regression* dan *Decision Tree*. Hasil menunjukkan bahwa *Logistic Regression* memiliki performa terbaik, sedangkan *Decision Tree* memberikan interpretasi pola yang jelas. Insight utama menunjukkan bahwa aktivitas digital edukatif dan faktor psikometrik berperan signifikan dalam menentukan kejelasan tujuan pendidikan**.** Penelitian ini juga menunjukkan bahwa perilaku digital yang produktif, khususnya frekuensi pencarian informasi dan akses terhadap konten edukatif, merupakan faktor utama dalam membentuk kejelasan tujuan pendidikan remaja. Selain itu, pendekatan *hybrid machine learning* terbukti efektif dalam mengidentifikasi pola sekaligus menghasilkan model prediksi yang akurat dan *interpretable*.

KESIMPULAN

Berdasarkan tujuan penelitian yang telah dirumuskan pada bagian pendahuluan, yaitu untuk menganalisis karakteristik dan memprediksi kejelasan tujuan pendidikan remaja melalui integrasi data psikometrik dan aktivitas digital, hasil penelitian menunjukkan bahwa tujuan tersebut telah berhasil dicapai. Melalui tahapan *preprocessing*, *clustering* menggunakan K-Means, serta klasifikasi menggunakan Logistic Regression dan Decision Tree, diperoleh segmentasi responden ke dalam tiga kelompok utama (rendah, sedang, dan tinggi) yang mencerminkan pola perilaku yang terstruktur.

Hasil analisis menunjukkan bahwa variabel *cluster*, frekuensi pencarian informasi, dan durasi akses konten edukatif merupakan faktor dominan dalam menentukan kejelasan tujuan pendidikan. Model *Logistic Regression* memberikan performa terbaik dengan tingkat akurasi yang tinggi dan stabil, sedangkan *Decision Tree* memberikan keunggulan dalam interpretasi pola. Selain itu, pendekatan *hybrid machine learning* terbukti efektif dalam menggabungkan analisis pola dan prediksi secara komprehensif.

Secara keseluruhan, hasil penelitian ini konsisten dengan permasalahan yang diangkat, yaitu bahwa perilaku digital yang produktif berkontribusi signifikan terhadap kejelasan tujuan pendidikan remaja. Sebagai prospek pengembangan, penelitian selanjutnya dapat memperluas jumlah dan variasi data, menggunakan algoritma yang lebih kompleks seperti *Random Forest* atau *Neural Network*, serta mengembangkan sistem berbasis aplikasi atau dashboard untuk implementasi praktis dalam bidang pendidikan.

DAFTAR PUSTAKA

[1] N. Sevila, R. A. Ningsih, M. A. M. Huda, and A. Malik, "Tren konsumsi digital di kalangan remaja," *JIIC: Jurnal Intelek Insan Cendikia*, vol. 2, no. 5, May 2025. [Online]. Available: https://jicnusantara.com/index.php/jiic

[2] Z. M. N. Al-Rahbi and N. S. A. Al-Daraai, "The impact of social media on academic performance, health and social interaction of students in University of Technology and Applied Sciences, Nizwa, Sultanate of Oman," *Journal of Education, Society and Behavioural Science*, vol. 36, no. 12, pp. 1–13, 2023.

[3] S. C. Berzin, J. Singer, and C. Chan, "Practice innovation through technology in the digital age: A grand challenge for social work," *American Academy of Social Work & Social Welfare*, vol. 12, pp. 3–21, 2015.

[4] N. Sany *et al.*, "Prediction of student major selection at high school using a machine learning approach," *International Journal of Engineering and Computer Science Applications (IJECSA)*, 2025. [Online]. Available: https://doi.org/10.30812/ijecsa.v4i1.4983

[5] K. Huynh and N. T. Ly, "A deep learning model of major consulting support," *Science & Technology Development Journal*, 2023. [Online]. Available: https://doi.org/10.32508/stdj.v26i2.4087

[6] N. B. Sulikeri, "Making career choices and AI based counselling accessible to every child at secondary level along with aptitude tests and detailed career paths," *International Journal for Research in Applied Science and Engineering Technology*, 2025. [Online]. Available: https://doi.org/10.22214/ijraset.2025.70380

[7] O. B. Famodimu *et al.*, "Enhancing pre-tertiary students decision-making using a web-based admission recommender system," in *2024 International Conference on Science,*

Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), 2024, pp. 1–6. [Online]. Available: https://doi.org/10.1109/SEB4SDG60871.2024.10630106

[8] N. Kamal, F. Sarker, and K. Mamun, "A comparative study of machine learning approaches for recommending university faculty," in *2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI)*, 2020, pp. 1–6. [Online]. Available: https://doi.org/10.1109/STI50764.2020.9350461

[9] B. Ondiek, L. Waruguru, and S. Njenga, "Recommender system for STEM enrolment in universities using machine learning algorithms: Case of Kenyan universities," *International Journal of Science, Technology & Management*, vol. 4, no. 6, 2023. [Online]. Available: https://doi.org/10.46729/ijstm.v4i6.1009

[10] R. Saluja and M. Rai, "A streamlined approach to student stream prediction using an ensemble machine learning model," *Communications on Applied Nonlinear Analysis*, vol. 32, 2024. [Online]. Available: https://doi.org/10.52783/cana.v32.2668

[11] B. Berlikozha *et al.*, "Development of method to predict career choice of IT students in Kazakhstan by applying machine learning methods," *Journal of Robotics and Control (JRC)*, vol. 6, no. 1, 2025. [Online]. Available: https://doi.org/10.18196/jrc.v6i1.25558

[12] R. Phummapooti, J. Thanyaphongphat, and P. Panjaburee, "AI-driven recommendations for digital majors based on learning styles," in *2024 5th Technology Innovation Management and Engineering Science International Conference (TIMES-iCON)*, 2024, pp. 1–5. [Online]. Available: https://doi.org/10.1109/TIMES-iCON61890.2024.10630767

[13] M. Aloud, N. Alkhamees, N. Almezeini, and M. AlYahya, "Employing supervised learning techniques for college major prediction: Empowering decision-making in university admission systems," *Journal of Information Systems Engineering and Management*, vol. 10, 2025. [Online]. Available: https://doi.org/10.52783/jisem.v10i15s.2429

[14] J. P. A. Acang *et al.*, "Generating program recommendations to aid in the decision-making of students in higher education using non-parametric supervised machine learning," in *2023 IEEE 15th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM)*, 2023, pp. 1–5. [Online]. Available: https://doi.org/10.1109/HNICEM60674.2023.10589265

[15] M. F. Abdullah and Y. A. S. Elewaa, "A predictive model for academic major selection using AI and labor market trends," *Formosa Journal of Science and Technology*, vol. 4, no. 4, 2025. [Online]. Available: https://doi.org/10.55927/fjst.v4i4.55

[16] A. Wibowo and W. Gunawan, "Pemanfaatan Algoritma K-Means dalam Klasterisasi Gempa Sulawesi," Faktor Exacta, vol. 17, no. 3, pp. 228–240, Sep. 2024. [Online]. Available: https://journal.lppmunindra.ac.id/index.php/Faktor_Exacta/article/view/23169. DOI: 10.30998/faktorexacta.v17i3.23169