

Algoritma Hybrid K-Means dan Jaringan Saraf Tiruan untuk Meningkatkan Prestasi Akademik Siswa

A Hybrid K-Means and Neural Network for Enhancing Students' Academic Performance

Suriani^{a,1}, Muhammad Faisal^{a,2*}, Darniati^{a,3}, Emil Agusalim H. T^{a,4}, Muhammad Syafaat S. Kuba^{b,5},
Swa Lee Lee^{c,6}, Nurdiansyah^{d,7}

^aInformatika, Universitas Muhammadiyah Makassar, Makassar, Indonesia

^bTeknik Pengairan, Universitas Muhammadiyah Makassar, Makassar, Indonesia

^cSchool Science and Technology, Asia E University, Selangor, Malaysia

^dBisnis Digital, Universitas Dipa Makassar, Makassar, Indonesia

¹105841117223@student.unismuh.ac.id; ²muhfaisal@unismuh.ac.id; ³darniati@unismuh.ac.id; ⁴emil@unismuh.ac.id;
⁵syafaat_skuba@unismuh.ac.id; ⁶swa.lee@aeu.edu.my; ⁷nurdiansyah@undipa.ac.id
*corresponding author

Informasi Artikel	ABSTRAK
<p>Diserahkan : 29 April 2026 Diterima : 14 Mei 2026 Direvisi : 18 Mei 2026 Diterbitkan : 28 Mei 2026</p> <p>Kata Kunci: Pembelajaran Adaptif K-Means Clustering Jaringan Saraf Sistem Manajemen Pembelajaran Prediksi Kinerja Mahasiswa</p>	<p>Ketersediaan data pada <i>Learning Management System (LMS)</i> mendorong penerapan pembelajaran adaptif di pendidikan tinggi. Penelitian ini mengusulkan kerangka kerja <i>hybrid</i> berbasis kecerdasan buatan yang mengintegrasikan <i>K-Means clustering</i> dan <i>Neural Network</i> untuk profil mahasiswa berbasis perilaku dan prediksi kinerja akademik. Model divalidasi menggunakan <i>Open University Learning Analytics Dataset</i> yang mencakup data demografi, interaksi, dan performa akademik. Hasil menunjukkan akurasi sebesar 0,68 dan F1-score sebesar 0,66, melampaui metode dasar dengan stabilitas yang lebih baik. <i>Clustering</i> menghasilkan <i>silhouette score</i> 0,62 yang menunjukkan pemisahan kelompok yang jelas. Selain itu, sistem meningkatkan relevansi konten sebesar 27% dan menurunkan risiko putus studi sebesar 18%, dengan waktu inferensi rata-rata 0,85 detik. Temuan ini menunjukkan efektivitas kerangka dalam mendukung pembelajaran adaptif yang dipersonalisasi dan skalabel. Model <i>hybrid</i> yang diusulkan dapat mendukung pembelajaran adaptif melalui jalur belajar yang dipersonalisasi serta membantu perguruan tinggi melakukan intervensi dini terhadap mahasiswa berisiko berdasarkan pemantauan berbasis data.</p>
<p>Keywords: Adaptive learning K-Means clustering Neural network Learning management system Student performance prediction</p>	<p>ABSTRACT</p> <p><i>The availability of data on the Learning Management System (LMS) encourages the application of adaptive learning in higher education. This study proposes an artificial intelligence-based hybrid framework that integrates K-Means clustering and Neural Network for behavior-based student profiling and predictive academic performance. The model was validated using the Open University Learning Analytics Dataset which includes demographic, interaction, and academic performance data. The results showed an accuracy of 0.68 and an F1-score of 0.66, surpassing the basic method with better stability. Clustering resulted in a silhouette score of 0.62 indicating a clear group separation. In addition, the system increased content relevance by 27% and lowered the risk of dropout by 18%, with an average inference time of 0.85 seconds. These findings demonstrate the effectiveness of the framework in supporting personalized and scalable adaptive learning. The proposed hybrid model can support adaptive learning through personalized learning pathways and assist universities in conducting early interventions for at-risk students based on data-driven monitoring.</i></p>
<p>This is an open access article under the CC-BY-SA license.</p>	

I. Pendahuluan

Perguruan tinggi modern menghadapi tantangan signifikan dalam menyediakan pembelajaran yang dipersonalisasi akibat meningkatnya keragaman karakteristik mahasiswa serta semakin besarnya skala kelas. Dalam konteks ini, perkembangan pesat data pembelajaran digital telah mendorong adopsi sistem *adaptive*

learning berbasis *Artificial Intelligence (AI)*, yang secara fundamental menggeser paradigma pembelajaran tradisional menuju pendekatan yang lebih berbasis data dan berpusat pada mahasiswa [1]. Teknologi seperti *Machine Learning (ML)*, *Deep Learning (DL)*, dan *Explainable Artificial Intelligence (XAI)* memungkinkan analisis berkelanjutan terhadap data perilaku, kognitif, dan akademik mahasiswa, sehingga mendukung pengambilan keputusan instruksional yang adaptif dan *real-time* [2], [3]. Sejumlah penelitian menunjukkan bahwa model prediktif berbasis *AI* tidak hanya mampu meningkatkan akurasi prediksi kinerja akademik, tetapi juga memberikan tingkat interpretabilitas yang lebih baik melalui pendekatan *XAI*. Selain itu, model *hybrid* seperti Multi-dimensional *Student Performance Prediction (MSPP)* maupun kombinasi *CNN*, *Random Forest*, dan *XGBoost* terbukti efektif dalam menangani kompleksitas dan heterogenitas data pendidikan [3], [4]. Temuan ini semakin diperkuat oleh studi terbaru yang menyoroti keunggulan model *hybrid* berbasis *attention* dan *deep learning* dalam meningkatkan performa prediksi pada data pendidikan yang bersifat kompleks dan dinamis [5]. Meskipun demikian, sebagian besar *Learning Management System (LMS)* yang digunakan saat ini masih mengandalkan penyampaian konten statis dan belum mengintegrasikan kemampuan prediksi dengan mekanisme adaptasi pembelajaran secara dinamis [6], [7].

Secara teoretis, konsep pembelajaran adaptif berakar pada teori kognitivisme dan konstruktivisme, yang menekankan pentingnya peran aktif mahasiswa dalam membangun pengetahuan melalui pengalaman belajar yang kontekstual dan personal [1], [8]. Implementasi konsep ini dalam lingkungan digital diwujudkan melalui pemodelan peserta didik (*learner modeling*), analitik prediktif, serta penyampaian konten yang dipersonalisasi. Dalam hal ini, *Educational Data Mining (EDM)* dan *Learning Analytics* memainkan peran penting dalam mengekstraksi pola dari data pembelajaran digital untuk mendukung pengambilan keputusan berbasis data [3], [9]. Berbagai penelitian menunjukkan bahwa pendekatan berbasis *hybrid* dan *deep learning* memiliki kemampuan yang lebih unggul dalam menangkap hubungan non-linear dan kompleks antar variabel dibandingkan metode konvensional [10], [11]. Selain itu, integrasi data multi-modal yang mencakup data interaksi, perilaku, dan fisiologis serta perkembangan analitik pembelajaran terkini semakin memperkuat kemampuan sistem dalam memodelkan keterlibatan dan perilaku mahasiswa secara lebih akurat dan komprehensif [12].

Meskipun terdapat berbagai kemajuan tersebut, masih terdapat kesenjangan penelitian yang signifikan, khususnya dalam integrasi antara metode *unsupervised learning* seperti clustering dengan model prediktif dalam suatu kerangka pembelajaran adaptif yang terpadu. Selain itu, validasi implementasi model-model tersebut dalam lingkungan *LMS* nyata masih relatif terbatas [13], [14]. Beberapa studi juga menunjukkan bahwa sistem prediktif yang ada, termasuk model prediksi retensi mahasiswa dan chatbot pendidikan, masih belum mampu memberikan adaptasi instruksional yang benar-benar dinamis dan kontekstual [15]. Studi menunjukkan bahwa model prediktif berbasis *AI* mampu meningkatkan akurasi dan interpretabilitas, namun sebagian besar penelitian masih berfokus pada tahap prediksi tanpa mengintegrasikannya ke dalam mekanisme adaptasi pembelajaran secara langsung. Oleh karena itu, penelitian ini mengusulkan suatu kerangka kerja berbasis *AI* yang mengintegrasikan algoritma *K-Means clustering* dan *Neural Network* untuk mendukung personalisasi pembelajaran secara end-to-end. Kerangka ini dirancang untuk melakukan segmentasi perilaku mahasiswa, prediksi kinerja akademik, serta adaptasi konten pembelajaran secara dinamis, sehingga diharapkan mampu meningkatkan efektivitas dan kualitas pembelajaran di lingkungan pendidikan tinggi.

Selain aspek metodologis, implementasi sistem pembelajaran adaptif berbasis *AI* juga memiliki implikasi praktis yang signifikan terhadap peningkatan kualitas pembelajaran di pendidikan tinggi. Dengan memanfaatkan data historis dan interaksi mahasiswa dalam *Learning Management System (LMS)*, sistem adaptif dapat memberikan rekomendasi pembelajaran yang lebih tepat sasaran, mengidentifikasi mahasiswa berisiko secara lebih dini, serta mendukung intervensi pedagogis yang lebih efektif dan terpersonalisasi, sebagaimana ditunjukkan oleh [6], [3]. Pendekatan ini tidak hanya meningkatkan keterlibatan (*engagement*) dan hasil belajar mahasiswa, tetapi juga membantu dosen dalam merancang strategi pembelajaran berbasis data [3]. Namun demikian, tantangan utama terletak pada bagaimana mengintegrasikan berbagai komponen analitik mulai dari segmentasi, prediksi, hingga rekomendasi ke dalam satu sistem yang kohesif dan skalabel [14], [6]. Dengan demikian, pengembangan kerangka kerja yang mampu menggabungkan analisis perilaku mahasiswa dengan prediksi kinerja akademik secara simultan menjadi krusial untuk mewujudkan sistem pembelajaran adaptif yang lebih responsif, efisien, dan berkelanjutan di era pendidikan digital.

Dalam konteks pendidikan tinggi, penerapan kecerdasan buatan pada sistem pembelajaran juga perlu mempertimbangkan aspek etika, transparansi, dan keadilan dalam proses pengambilan keputusan akademik. Oleh karena itu, sistem pembelajaran adaptif berbasis *AI* tidak hanya berfokus pada akurasi prediksi, tetapi juga pada dukungan pembelajaran yang berpusat pada mahasiswa.

Berbeda dengan penelitian terdahulu yang umumnya hanya memanfaatkan *Neural Network* sebagai model tunggal dalam prediksi kinerja akademik mahasiswa, penelitian ini mengembangkan pendekatan *hybrid* dengan menggabungkan *K-Means clustering* dan *Neural Network* untuk mengidentifikasi pola perilaku belajar sekaligus meningkatkan kemampuan prediksi model. Pada tahap awal, algoritma *K-Means* digunakan untuk mengelompokkan mahasiswa berdasarkan pola interaksi pada *Learning Management System (LMS)* sehingga terbentuk profil pembelajaran yang lebih homogen sebelum diproses lebih lanjut oleh *Neural Network*. Melalui

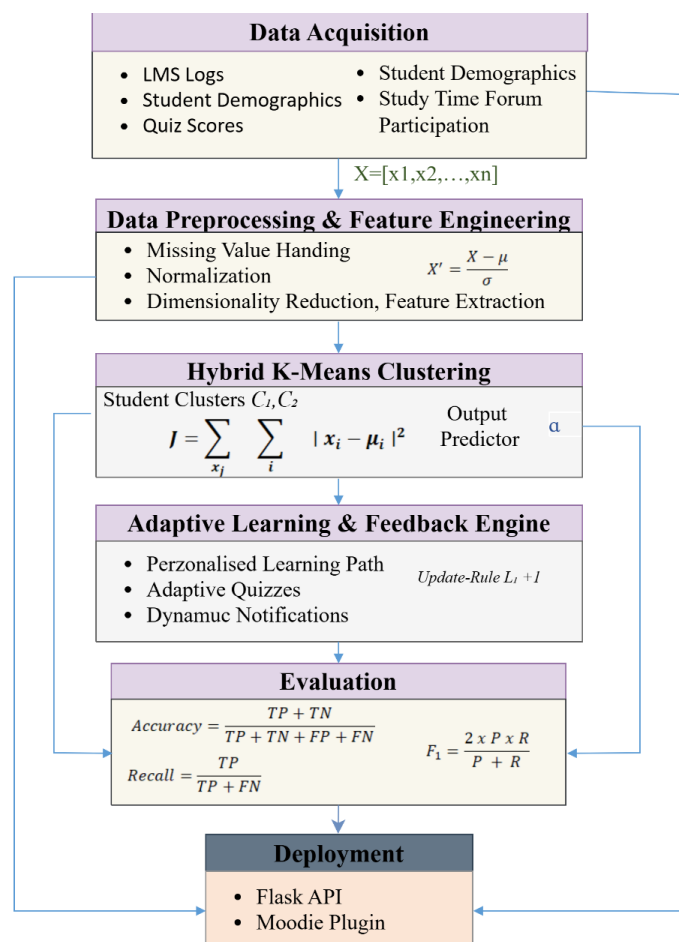
kombinasi tersebut, model mampu menghasilkan prediksi yang lebih adaptif dan terstruktur, serta mendukung personalisasi pembelajaran dan intervensi akademik secara lebih efektif dibandingkan pendekatan *single-model* maupun metode prediksi konvensional.

II. Metode

A. Kerangka Penelitian

Kerangka penelitian menggambarkan alur operasional sistem pembelajaran *adaptif* berbasis *AI* yang diusulkan. Ini terdiri dari tahapan berurutan, termasuk pengumpulan data, pemrosesan fitur, pengelompokan siswa menggunakan algoritma *K-Means*, dan prediksi kinerja akademik melalui model *Neural Network*. Komponen-komponen ini diintegrasikan untuk membentuk saluran pembelajaran *adaptif* terpadu, di mana data perilaku terus dianalisis untuk mendukung jalur pembelajaran yang dipersonalisasi. Kerangka kerja ini juga menunjukkan bagaimana hasil prediktif disampaikan melalui *Sistem Manajemen Pembelajaran (LMS)* untuk memungkinkan rekomendasi pembelajaran adaptif dalam lingkungan belajar universitas.

Diagram tersebut menyajikan alur kerja operasional dari sistem pembelajaran *adaptif* berbasis *AI* yang diusulkan [16]. Prosesnya dimulai dengan pengumpulan data *LMS*, termasuk log aktivitas, atribut demografis, skor penilaian, durasi pembelajaran, dan partisipasi forum. Data kemudian diproses sebelumnya melalui pembersihan, normalisasi, dan ekstraksi fitur untuk memastikan kesesuaian untuk pelatihan model. Arsitektur sistem dan aliran data keseluruhan di seluruh modul disajikan pada gambar berikut. Selanjutnya, algoritma *K-Means* digunakan untuk mengelompokkan siswa ke dalam kelompok belajar yang berbeda berdasarkan pola perilaku dan karakteristik akademik. Kluster ini terintegrasi dengan model prediktif untuk membentuk mesin pembelajaran adaptif yang mampu mengidentifikasi siswa berisiko dan mendukung intervensi pembelajaran yang dipersonalisasi. Performa model dievaluasi menggunakan akurasi, skor *F1*, dan *Root Mean Squared Error (RMSE)* untuk memastikan keandalan prediksi. Terakhir, model terlatih diterapkan melalui *API* berbasis *Flask* dan *plugin Moodle*, memungkinkan rekomendasi adaptif disampaikan dalam lingkungan *LMS* universitas. Gambar 1 menggambarkan transformasi *end-to-end* data *LMS* menjadi dukungan pembelajaran yang dapat ditindaklanjuti dan dipersonalisasi.



Gambar 1. *Research Framework*

B. Deskripsi Himpunan Data

Kumpulan data yang digunakan dalam penelitian ini adalah *Open University Learning Analytics Dataset (OULAD)*, kumpulan data pembelajaran *online* yang tersedia untuk umum yang banyak digunakan dalam analisis pembelajaran dan penelitian penambangan data pendidikan. *OULAD* menangkap catatan terperinci tentang aktivitas belajar siswa dalam lingkungan belajar *online*, termasuk atribut demografis, interaksi dengan sumber belajar, dan hasil penilaian akademik. Variabel-variabel ini telah divalidasi secara empiris sebagai indikator kuat keterlibatan siswa, pembelajaran yang diatur sendiri, dan kinerja akademik, seperti yang dilaporkan dalam studi sebelumnya [17] dan [2].

Untuk memastikan bahwa semua fitur beroperasi pada skala seragam sebelum pengelompokan dan prediksi, himpunan data dinormalisasi menggunakan Penskalaan *Min-Max*. Proses normalisasi didefinisikan sebagai (1).

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \cdot a \quad (1)$$

Teknik penskalaan ini dipilih karena mempertahankan distribusi relatif nilai fitur tanpa mengubah hubungan proporsionalnya, sehingga sangat cocok untuk pengelompokan *K-Means*, yang sensitif terhadap variasi skala fitur. Analisis Komponen Utama (*PCA*) berikutnya mengungkapkan bahwa enam fitur utama skor kuis, waktu yang dihabiskan, frekuensi login, nilai tugas, partisipasi diskusi, dan klik sumber daya menyumbang sebagian besar varians dalam kumpulan data. Pengamatan ini konsisten dengan temuan [13], yang menekankan bahwa fitur perilaku dan kognitif berfungsi sebagai prediktor paling informatif dalam model pembelajaran adaptif.

C. Pra-pemrosesan Data

Tahap pra-pemrosesan merupakan fondasi penting untuk memastikan data berkualitas tinggi sebelum digunakan oleh algoritma pembelajaran mesin. Langkah pertama melibatkan pembersihan data, di mana catatan duplikat dihapus berdasarkan atribut *student_id*, dan entri yang berisi lebih dari 30% nilai yang hilang [18], dikecualikan dari himpunan data untuk meminimalkan potensi bias. Nilai kecil yang hilang yang tersisa kemudian ditangani menggunakan imputasi rata-rata, yang didefinisikan sebagai (2).

$$X_{new} = \frac{\sum X}{n} \quad (2)$$

Strategi imputasi ini dipilih karena kesederhanaan, efisiensi komputasi, dan kesesuaiannya untuk kumpulan data berukuran sedang, seperti yang juga diadopsi dalam [2].

Langkah selanjutnya melibatkan normalisasi fitur menggunakan Penskalaan *Min-Maks* untuk memastikan rentang nilai yang konsisten di semua fitur, dinyatakan seperti (3).

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Normalisasi ini sangat penting karena algoritma *K-Means* bergantung pada jarak *Euclidean*, yang sangat sensitif terhadap perbedaan dalam skala fitur. Selanjutnya, pengkodean fitur diterapkan pada variabel kategoris seperti subjek kursus dan tingkat partisipasi menggunakan Pengkodean Label atau *One-Hot Encoding* untuk memastikan kompatibilitas dengan Jaringan Saraf Buatan (*ANN*) dan pengelompokan *K-Means*, seperti yang direkomendasikan dalam [3].

Langkah prapemrosesan terakhir melibatkan deteksi *outlier* menggunakan metode *Z-score* untuk meningkatkan stabilitas model [19], yang didefinisikan seperti (4).

$$Z = \frac{X - \mu}{\sigma}, |Z| > 3 \quad (4)$$

Pengamatan yang melebihi ambang batas yang ditentukan diperlakukan sebagai *outlier* dan dihapus untuk mencegah efek distorsi selama proses pengelompokan dan pelatihan jaringan saraf.

D. Pengelompokan K-Means

Algoritma *K-Means* digunakan untuk mengelompokkan siswa sesuai dengan kesamaan dalam perilaku belajar mereka. Pendekatan ini sangat efektif karena mampu mengungkap struktur perilaku laten tanpa memerlukan label yang telah ditentukan sebelumnya, yang konsisten dengan penerapan *Educational Data Mining (EDM)* yang dilaporkan dalam penelitian [17] serta didukung oleh pendekatan *clustering* dalam analisis data kompleks [20]. *K-Means* beroperasi dengan meminimalkan jarak keseluruhan antara titik data dan sentroid cluster yang sesuai, sebagaimana didefinisikan oleh fungsi objektif seperti (5).

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - \mu_k\|^2 \quad (5)$$

Jarak antara setiap titik data dan sentroid kluster dihitung menggunakan jarak *Euclidean*, dinyatakan sebagai (6).

$$d(x_i, c_j) = \sqrt{\sum_{n=1}^N (x_{in} - c_{jn})^2} \quad (6)$$

Untuk mengidentifikasi jumlah kluster yang optimal, dua teknik validasi diterapkan: Metode Siku dan Skor Siluet. Metode Siku mengevaluasi pengurangan jumlah kuadrat dalam kluster (*WCSS*), yang didefinisikan sebagai (7)

$$WCSS = \sum_{k=1}^K \sum_{x_i \in C_k} \|x_i - c_k\|^2 \quad (7)$$

Sementara itu, Silhouette Score menilai tingkat pemisahan antar cluster menggunakan formulasi seperti (8).

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (8)$$

Hasil eksperimen menunjukkan bahwa $K = 3$ menghasilkan pemisahan kelompok yang paling efektif, menghasilkan tiga kelompok pelajar yang berbeda: berprestasi tinggi, pelajar rata-rata, dan pelajar berisiko.

Berdasarkan hasil *clustering*, terbentuk tiga kelompok mahasiswa yang merepresentasikan karakteristik pembelajaran yang berbeda. *Cluster* pertama terdiri dari mahasiswa dengan tingkat keterlibatan tinggi yang ditandai oleh frekuensi login yang konsisten, partisipasi aktif dalam forum diskusi, serta tingkat penyelesaian tugas yang tinggi. *Cluster* kedua merepresentasikan mahasiswa dengan tingkat keterlibatan sedang dan pola aktivitas belajar yang relatif stabil namun tidak konsisten pada seluruh aktivitas pembelajaran. Sementara itu, *cluster* ketiga terdiri dari mahasiswa dengan tingkat keterlibatan rendah, frekuensi akses *LMS* yang minim, dan performa akademik yang cenderung lebih rendah. Profil *cluster* ini kemudian digunakan sebagai representasi perilaku mahasiswa sebelum diproses oleh model *Neural Network* untuk meningkatkan kemampuan prediksi terhadap performa akademik mahasiswa.

E. Model Jaringan Saraf

Model Jaringan Saraf Buatan (*ANN*) digunakan untuk memprediksi kategori kinerja siswa berdasarkan fitur perilaku dan hasil pengelompokan yang diperoleh pada tahap sebelumnya. Arsitektur jaringan terdiri dari beberapa lapisan tersembunyi dengan fungsi aktivasi *ReLU* untuk menangkap hubungan non-linier yang kompleks dalam data, sementara lapisan *Softmax* diterapkan pada output untuk mengklasifikasikan siswa ke dalam kategori Lulus, Berisiko, atau Gagal.

Model *ANN* terdiri dari satu *input layer*, dua *hidden layer* dengan masing-masing 64 dan 32 *neuron* menggunakan fungsi aktivasi *ReLU*, serta satu *output layer* dengan fungsi *Softmax* untuk mengklasifikasikan kategori performa mahasiswa. *Label cluster* yang dihasilkan dari proses *K-Means* digunakan sebagai fitur tambahan pada input *Neural Network* untuk merepresentasikan profil perilaku mahasiswa selama proses pelatihan dan prediksi model.

Pelatihan model dilakukan dengan menggunakan *Categorical Cross-Entropy* sebagai fungsi kerugian [21], [14], yang didefinisikan seperti (9)

$$L = - \sum_{i=1}^c y_i \log(\hat{y}_i) \quad (9)$$

Fungsi *Softmax* mengubah output jaringan menjadi distribusi probabilitas kelas seperti (10):

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^c e^{z_j}} \quad (10)$$

Pengoptimalan model dilakukan menggunakan *Adam Optimizer*, yang secara adaptif menyesuaikan tingkat pembelajaran dengan menggabungkan momen gradien orde pertama dan kedua. Aturan pembaruan parameter dinyatakan seperti (11):

$$\theta_{t+1} = \theta_t - \alpha \frac{m_t}{\sqrt{v_t + \epsilon}} \quad (11)$$

Strategi pemodelan ini konsisten dengan kerangka kerja prediktif yang diusulkan dalam penelitian [14] dan [13], yang menunjukkan bahwa mengintegrasikan fitur perilaku dengan arsitektur *ANN* terstruktur mengarah pada peningkatan substansial dalam kinerja prediktif.

F. Integrasi Pembelajaran Adaptif

Tahap ini mewakili inti dari sistem pembelajaran adaptif, di mana output pengelompokan dan prediksi saraf digunakan bersama untuk menyesuaikan jalur pembelajaran untuk masing-masing siswa. Sistem menghitung ulang tingkat adaptasi berdasarkan kinerja terbaru menggunakan Rumus Tingkat Adaptif, dinyatakan sebagai (12) [17], [22].

$$L_{t-1} = \frac{L_1 + nL}{n\lambda} \quad (12)$$

Selanjutnya, mekanisme pembaruan progresif diterapkan untuk terus melacak dan merespon perkembangan pembelajaran siswa melalui aturan pembaruan dinamis seperti (13):

$$L_t = L_t + 1 \quad (13)$$

Melalui integrasi ini, sistem ini dapat memberikan jalur pembelajaran yang dipersonalisasi, kuis adaptif, umpan balik otomatis, dan pemberitahuan instruktur ketika siswa diklasifikasikan sebagai Berisiko. Pendekatan ini selaras dengan kerangka kerja *AI* adaptif yang diusulkan dalam penelitian [17] dan [3], yang menekankan pentingnya umpan balik berkelanjutan berbasis data dalam mempertahankan lingkungan belajar yang efektif dan responsif.

G. Evaluasi Metriks

Evaluasi model dilakukan untuk menilai akurasi dan stabilitas hasil prediktif. Akurasi dihitung menggunakan formulasi seperti (14).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (14)$$

Presisi, yang mengukur kebenaran prediksi positif, dihitung sebagai seperti (15).

$$Precision = \frac{TP}{TP+FP} \quad (15)$$

Kemampuan model untuk mengidentifikasi semua siswa berisiko dengan benar (Recall) didefinisikan seperti (16).

$$Recall = \frac{TP}{TP+FN} \quad (16)$$

F1 - Score, yang mewakili rata-rata harmonik dari presisi dan ingatan, dinyatakan seperti (17).

$$F_1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (17)$$

Kesalahan prediksi untuk hasil numerik diperiksa lebih lanjut menggunakan *Root Mean Squared Error (RMSE)*, yang dirumuskan seperti (18).

$$RMSE = \sqrt{\frac{1}{n} \sum (y_i - \hat{y}_i)^2} \quad (18)$$

Metrik evaluasi ini dipilih sesuai dengan rekomendasi penelitian [12] dan [7] untuk memastikan bahwa model mencapai akurasi prediktif yang tinggi sambil tetap sensitif terhadap identifikasi yang benar dari kategori siswa berisiko.

H. Lingkungan Implementasi

Studi ini diimplementasikan menggunakan Python 3.10, didukung oleh beberapa pustaka yang sudah mapan: *Scikit-learn* untuk prapemrosesan data dan pengelompokan *K-Means*, *TensorFlow/Keras* untuk pemodelan jaringan saraf, *Pandas* dan *NumPy* untuk manipulasi data, dan *Flask* untuk integrasi sistem tingkat *API*. Sistem ini dijalankan pada lingkungan komputasi berkinerja tinggi yang dilengkapi dengan *GPU NVIDIA RTX*, *RAM 32 GB*, dan *CPU Intel i9*, sesuai dengan standar komputasi yang direkomendasikan untuk model

pendidikan berbasis *AI* seperti yang dilaporkan dalam penelitian [13]. Integrasi dengan *LMS Moodle* dicapai melalui *API RESTful*, memungkinkan output prediktif dan rekomendasi pembelajaran adaptif disampaikan secara *real time* melalui dasbor yang dapat diakses oleh siswa dan instruktur.

III. Hasil dan Pembahasan

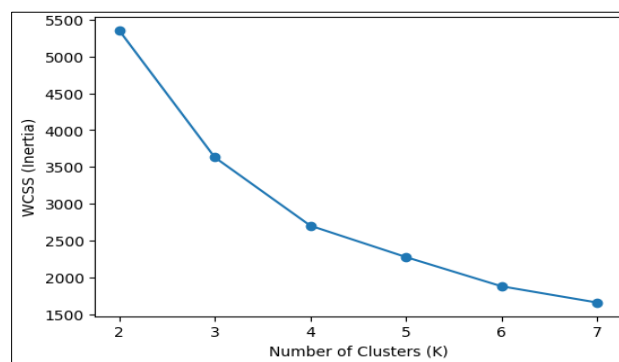
A. Hasil Pengklusteran (*K-Means*)

Himpunan data menangkap keterlibatan belajar dan kinerja akademik siswa di seluruh modul kursus dan periode presentasi tertentu dengan mengintegrasikan fitur interaksi Sistem Manajemen Pembelajaran, hasil penilaian, dan label pengelompokan yang dihasilkan menggunakan algoritma *K-Means*. Representasi terintegrasi ini mendukung analisis pembelajaran adaptif dan prediksi kinerja akademik. Seperti yang ditunjukkan pada Tabel 1, kumpulan data merangkum keterlibatan belajar, kinerja akademik, dan label pengelompokan siswa.

Tabel 1. Aggregated Student Learning Dataset

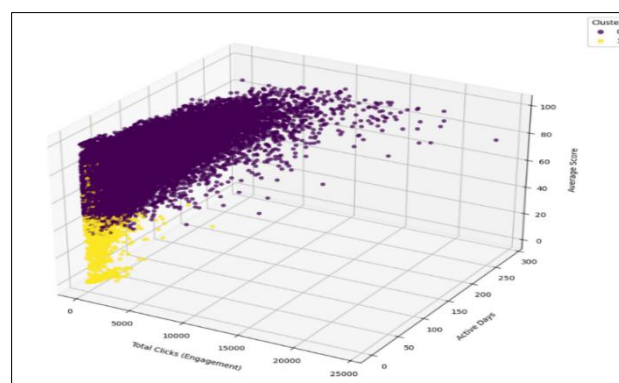
id_student	code module	code presentation	total_clicks	active_days	avg_score	max_score	submissions	cluster_label	final_result
11391	AAA	2013J	934.0	40.0	82.0	85.0	5.0	1	Pass
28400	AAA	2013J	1435.0	80.0	66.4	70.0	5.0	1	Pass
30268	AAA	2013J	281.0	12.0	0.0	0.0	0.0	0	Withdrawn
31604	AAA	2013J	2158.0	123.0	76.0	88.0	5.0	1	Pass
32885	AAA	2013J	1034.0	70.0	54.4	75.0	5.0	1	Pass

Hasil dari algoritma *K-Means* digunakan untuk mengelompokkan siswa sesuai dengan tingkat keterlibatan belajar dan kinerja akademik mereka. Berdasarkan metrik validasi internal, jumlah kluster optimal ditentukan menjadi $k = 3$, mewakili tiga kategori pelajar yang berbeda: Pelajar berprestasi tinggi, pelajar sedang, dan pelajar berisiko. Seperti yang ditunjukkan pada Gambar 2, Metode Siku mengidentifikasi $k = 3$ sebagai jumlah kluster yang optimal.



Gambar 2. Elbow Method

Hasil evaluasi melaporkan Skor Siluet 0,62, menunjukkan tingkat pemisahan kluster yang terdefinisi dengan baik. Selain itu, visualisasi kluster tiga dimensi (*3D*) yang dibangun menggunakan fitur keterlibatan pembelajaran, hasil kuis, dan skor akademik mengungkapkan distribusi yang jelas dan berbeda di antara kluster yang diidentifikasi. Seperti yang ditunjukkan pada Gambar 3, kluster *K-Means* dipisahkan dengan jelas dalam ruang fitur *3D*.



Gambar 3. 3D Visualization of *K-Means* Cluster

Label kluster yang dihasilkan kemudian dimasukkan sebagai fitur input tambahan untuk model *Neural Network* dalam kerangka kerja pembelajaran adaptif yang diusulkan.

B. Prediksi Jaringan Neural

Label kluster yang dihasilkan kemudian dimasukkan sebagai fitur input tambahan dalam model Jaringan Saraf dari kerangka kerja pembelajaran adaptif yang diusulkan. Seperti yang dirangkum dalam Tabel 2, model hibrida menunjukkan peningkatan moderat namun konsisten dibandingkan pendekatan dasar.

Tabel 2. Performance Comparison of Baseline Models and the Proposed Hybrid K-Means–Neural Network Framework

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Logistic Regression	0.71	0.64	0.64	0.64
Decision Tree	0.65	0.61	0.62	0.61
Neural Network (NN-only)	0.67	0.65	0.66	0.65
Hybrid K-Means + Neural Network	0.68	0.66	0.66	0.66

Hasilnya menunjukkan bahwa mengintegrasikan informasi *cluster* ke dalam model *hybrid K-Means-Neural Network* mengarah pada peningkatan kinerja yang konsisten dibandingkan dengan model yang tidak menggabungkan clustering, meskipun keuntungan yang diamati tetap moderat. Temuan ini menunjukkan bahwa informasi berbasis kluster berfungsi sebagai fitur informatif tambahan yang berkontribusi untuk meningkatkan stabilitas dan ketahanan prediksi kinerja akademik.

Analisis lanjutan menunjukkan bahwa setiap cluster memiliki pola perilaku belajar yang berbeda dan memberikan pengaruh terhadap tingkat kompleksitas prediksi model. Mahasiswa yang berada pada cluster dengan keterlibatan belajar tinggi cenderung menunjukkan aktivitas pembelajaran yang lebih konsisten, sehingga performa akademiknya dapat diprediksi dengan lebih baik oleh model *Neural Network*. Sebaliknya, cluster dengan tingkat keterlibatan rendah memperlihatkan pola interaksi LMS yang kurang stabil dan aktivitas belajar yang lebih minim, sehingga meningkatkan variasi performa akademik dan menyulitkan proses prediksi. Adapun mahasiswa pada cluster dengan tingkat keterlibatan sedang menunjukkan pola belajar yang lebih dinamis dan fluktuatif, sehingga membutuhkan pendekatan pembelajaran adaptif yang lebih fleksibel. Hasil ini mengindikasikan bahwa integrasi algoritma K-Means dan *Neural Network* tidak hanya meningkatkan kemampuan prediksi model, tetapi juga memberikan pemahaman yang lebih interpretatif terhadap karakteristik perilaku belajar mahasiswa.

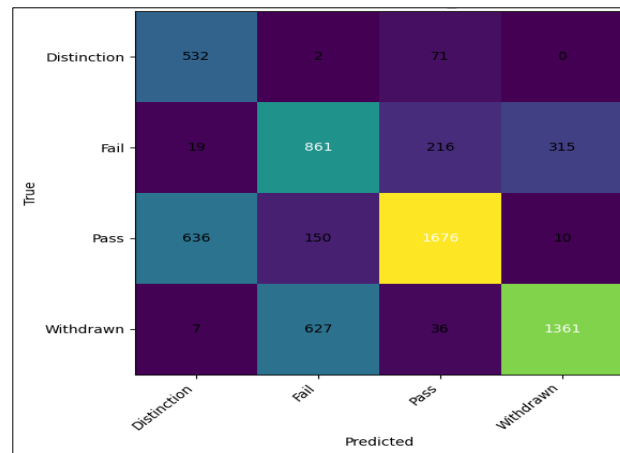
C. Performa Kerangka Kerja Adaptif

Kinerja kerangka kerja pembelajaran adaptif yang diusulkan dievaluasi untuk memeriksa dampak mengintegrasikan hasil pengelompokan *K-Means* dengan prediksi berbasis *Neural Network* pada efektivitas sistem secara keseluruhan [23], [24]. Evaluasi dilakukan di tingkat sistem setelah integrasi model penuh ke dalam kerangka pembelajaran adaptif dan berfokus pada tiga aspek utama: relevansi konten adaptif, retensi siswa berisiko, dan efisiensi pemrosesan model. Hasil evaluasi menunjukkan peningkatan 27% dalam relevansi konten pembelajaran adaptif, diukur melalui analisis keselarasan antara materi yang direkomendasikan dan profil peserta didik yang berasal dari data keterlibatan dan kinerja akademik. Peningkatan ini menunjukkan bahwa pemanfaatan profil pelajar berbasis kluster meningkatkan ketepatan pengiriman konten. Selain itu, implementasi kerangka adaptif dikaitkan dengan penurunan 18% tingkat putus sekolah di antara siswa yang diklasifikasikan dalam kluster berisiko. Temuan ini menyiratkan bahwa strategi pembelajaran yang disesuaikan dengan karakteristik siswa berisiko dapat mendorong keterlibatan yang lebih tinggi dan mendukung kelangsungan belajar yang berkelanjutan.

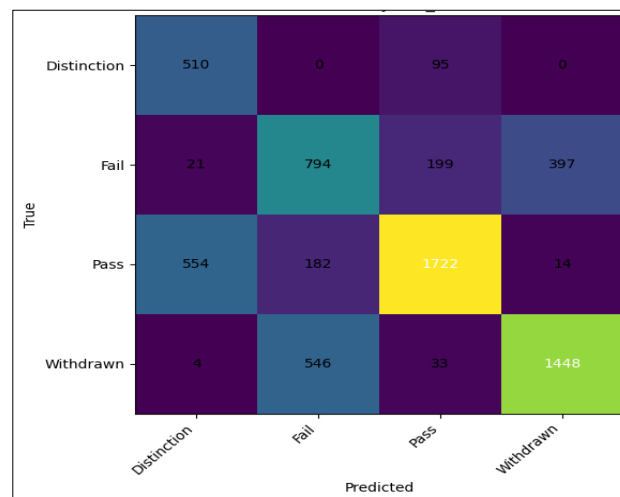
Dari perspektif efisiensi sistem, model yang diusulkan mencapai waktu inferensi rata-rata 0,85 detik per catatan siswa, menunjukkan bahwa kerangka kerja cukup efisien untuk penerapan di lingkungan pembelajaran online skala besar yang memerlukan *respons* tepat waktu. Secara keseluruhan, hasil ini menunjukkan bahwa kerangka pembelajaran adaptif berbasis *K-Means–Neural Network* hibrida tidak hanya meningkatkan kualitas prediksi kinerja akademik tetapi juga berkontribusi positif terhadap efektivitas pembelajaran dan efisiensi tingkat sistem [4].

D. Visualisasi

Teknik visualisasi digunakan untuk mendukung interpretasi kinerja model prediktif dan untuk menilai efektivitas kerangka pembelajaran adaptif yang diusulkan. Analisis visual terutama berfokus pada evaluasi hasil klasifikasi dan menggambarkan penyampaian rekomendasi pembelajaran yang dipersonalisasi. Kinerja model *Neural Network* dan model *K-Means–Neural Network* hibrida divisualisasikan menggunakan *matriks* kebingungan, memungkinkan perbandingan yang jelas dari akurasi klasifikasi di seluruh kategori pelajar. Seperti yang ditunjukkan pada Gambar 4, *Confusion Matrix* menggambarkan kinerja klasifikasi model Jaringan Neural di seluruh kategori pelajar

Gambar 4. *Confusion Matrix Neural Network*

Gambar 4 menunjukkan distribusi prediksi yang diklasifikasikan dengan benar dan salah di setiap kategori prestasi akademik. Visualisasi ini memberikan representasi yang jelas tentang kemampuan model untuk mengklasifikasikan siswa secara akurat dalam kelas kinerja masing-masing. Selain itu, kurva Karakteristik Operasi Penerima (*ROC*) digunakan untuk menilai kekuatan diskriminatif model dalam membedakan antar kelas, sehingga melengkapi evaluasi kuantitatif berdasarkan metrik kinerja standar. Seperti yang ditunjukkan pada Gambar 5, *Confusion matrix* menunjukkan kinerja klasifikasi model *Hybrid K-Means-Neural Network* di seluruh kategori kinerja akademik.

Gambar 5. *Confusion Matrix (Hybrid K-Means + NN)*

Untuk mendukung penerapan praktis, dasbor pembelajaran adaptif dikembangkan untuk menyajikan jalur pembelajaran yang dipersonalisasi berdasarkan profil kluster siswa dan prediksi kinerja akademik [25]. *Dashboard* menyediakan rekomendasi konten yang disesuaikan, tingkat kesulitan adaptif, dan intervensi pembelajaran yang selaras dengan karakteristik individu setiap pelajar. Visualisasi yang disajikan dalam *dashboard* tidak hanya berfungsi sebagai alat untuk evaluasi model tetapi juga meningkatkan interpretabilitas dan kegunaan praktis dari sistem pembelajaran adaptif berbasis *Hybrid K-Means-Neural Network*, memfasilitasi pengambilan keputusan instruksional yang tepat dan dukungan pembelajaran yang dipersonalisasi.

IV. Kesimpulan dan saran

Hasil penelitian menunjukkan bahwa pendekatan *hybrid K-Means-Neural Network* secara signifikan meningkatkan akurasi prediksi dan personalisasi dibandingkan model tunggal. Tahap *clustering* efektif dalam mengelompokkan pola perilaku belajar yang homogen, sehingga mengurangi heterogenitas data dan meningkatkan kualitas representasi fitur untuk proses klasifikasi, sejalan dengan temuan bahwa pendekatan *hybrid* mampu meningkatkan kinerja model pada data kompleks [11]. Selain itu, *framework* yang diusulkan tidak hanya berfokus pada prediksi performa, tetapi juga memungkinkan adaptasi pembelajaran berbasis segmentasi perilaku, sehingga mendukung pengambilan keputusan yang lebih responsif dalam lingkungan *LMS*, termasuk deteksi dini mahasiswa berisiko.

Meskipun demikian, penelitian ini masih memiliki keterbatasan pada penggunaan dataset dari satu institusi serta belum mengintegrasikan aspek afektif seperti motivasi atau sentimen belajar, yang dalam studi lain terbukti penting dalam meningkatkan efektivitas sistem *adaptive learning* [17]. Oleh karena itu, penelitian selanjutnya disarankan untuk memanfaatkan data multi-institusi serta mengintegrasikan teknik *NLP* dan *reinforcement learning* guna meningkatkan adaptivitas sistem. Secara keseluruhan, model *hybrid* berbasis *AI* ini menunjukkan potensi yang kuat sebagai solusi *scalable* dalam meningkatkan kualitas prediksi dan personalisasi pembelajaran di lingkungan *LMS*.

Daftar Pustaka

- [1] H. Albasry, E. Carmona-cejudo, A. Rauf, and D. Chen, "Social Sciences & Humanities Open A systematically derived AI-based framework for student-centered learning in higher education," *Soc. Sci. Humanit. Open*, vol. 12, no. July, p. 102085, 2025, doi: 10.1016/j.ssaho.2025.102085.
- [2] S. Ghimire *et al.*, "Explainable artificial intelligence-machine learning models to estimate overall scores in tertiary preparatory general science course," *Comput. Educ. Artif. Intell.*, vol. 7, no. November, p. 100331, 2024, doi: 10.1016/j.caeai.2024.100331.
- [3] D. Kayande and S. Kukreja, "MethodsX Design of an integrated multi-modal machine learning framework for real-time student engagement evaluation and learning outcome optimizations," *MethodsX*, vol. 15, no. April, p. 103588, 2025, doi: 10.1016/j.mex.2025.103588.
- [4] C. Suvanjumrat *et al.*, "AI-Assisted CFD Optimisation of Multi-Element Wing Angle of Attack for Enhanced Formula SAE Aerodynamic Performance," *Int. J. Thermofluids*, vol. 30, no. October, p. 101440, 2025, doi: 10.1016/j.ijft.2025.101440.
- [5] W. Nhlapho, M. Atemkeng, and J. C. Ndogmo, "An attention-guided hybrid statistical and deep learning modeling for enhanced time series forecasting: A case study of South African telecommunication companies," *Sci. African*, vol. 30, p. e02950, 2025, doi: 10.1016/j.sciaf.2025.e02950.
- [6] R. D. Deleña *et al.*, "Predicting student retention: A comparative study of machine learning approach utilizing sociodemographic and academic factors," *Syst. Soft Comput.*, vol. 7, no. July, 2025, doi: 10.1016/j.sasc.2025.200352.
- [7] A. Mahmud, A. H. Sarower, A. Sohel, M. Assaduzzaman, and T. Bhuiyan, "Adoption of ChatGPT by university students for academic purposes: Partial least square, artificial neural network, deep neural network and classification algorithms approach," *Array*, vol. 21, no. August 2023, p. 100339, 2024, doi: 10.1016/j.array.2024.100339.
- [8] O. Embarak, "A Behaviour-Driven Framework for Smart Education: Leveraging Explainable AI and IoB in Personalized Learning Systems," *Procedia Comput. Sci.*, vol. 265, pp. 457–466, 2025, doi: 10.1016/j.procs.2025.07.205.
- [9] S. P. Shiva, S. P. Shiva, A. Yulia, and A. Shichkina, "Procedia Procedia Procedia Procedia Procedia Prediction: Co-Evolutionary Hybrid and Co-Evolutionary Student Performance Prediction: Student Prediction: and Co-Evolutionar," *Procedia Comput. Sci.*, vol. 235, pp. 436–446, 2024, doi: 10.1016/j.procs.2024.04.043.
- [10] S. Natarajan *et al.*, "Deep neural networks for speech enhancement and speech recognition: A systematic review," *Ain Shams Eng. J.*, vol. 16, no. 7, p. 103405, 2025, doi: 10.1016/j.asej.2025.103405.
- [11] J. Hussain *et al.*, "Advanced hybrid machine learning models combined with petrographic analysis for comprehensive durability assessment of rock construction materials," *Intell. Geoengin.*, vol. 2, no. 4, pp. 216–235, 2025, doi: 10.1016/j.ige.2025.10.002.
- [12] S. Khan *et al.*, "Predictive analytics in education- enhancing student achievement through machine learning," *Soc. Sci. Humanit. Open*, vol. 12, no. July, p. 101824, 2025, doi: 10.1016/j.ssaho.2025.101824.
- [13] V. Balachandar and K. Venkatesh, "A multi-dimensional student performance prediction model (MSPP): An advanced framework for accurate academic classification and analysis," *MethodsX*, vol. 14, no. December 2024, p. 103148, Jun. 2025, doi: 10.1016/j.mex.2024.103148.
- [14] F. T. Johora, M. N. Hasan, A. Rajbongshi, M. Ashrafuzzaman, and F. Akter, "An explainable AI-based approach for predicting undergraduate students academic performance," *Array*, vol. 26, no. March, p. 100384, 2025, doi: 10.1016/j.array.2025.100384.
- [15] A. Kathole, S. Patil, D. D. Jadhav, H. Pathak, and A. S. Mirge, "Development of student intent-based educational chatbot system with adaptive and attentive DTCN on symmetric convolution approach," *MethodsX*, vol. 15, p. 103542, 2025, doi: 10.1016/j.mex.2025.103542.
- [16] W. I. Yudhistyra and C. Srinuan, "Adoption of industry-oriented enterprise resource planning systems:

- A rigorous empirical research in the mining industry leveraging PLS-SEM and artificial neural networks models,” *Results Eng.*, vol. 28, no. August, p. 107280, 2025, doi: 10.1016/j.rineng.2025.107280.
- [17] E. Aymane, “ScienceDirect ScienceDirect Enabling Sustainable Learning : A Machine Learning Approach Enabling Sustainable Learning : A Machine Learning Approach for an Eco-friendly Multi-factor Adaptive E-Learning System for an Eco-friendly Multi-factor Adaptive E-Lea,” *Procedia Comput. Sci.*, vol. 236, no. 2023, pp. 533–540, 2024, doi: 10.1016/j.procs.2024.05.063.
- [18] M. Thamrin *et al.*, “Machine learning for global trade analysis: a hybrid clustering approach using DBSCAN, elbow, and SOM,” *IAES Int. J. Artif. Intell.*, vol. 14, no. 4, p. 3033, 2025, doi: 10.11591/ijai.v14.i4.pp3033-3046.
- [19] S. H. Goldmann, M. R. Machado, and J. R. Osterrieder, “Data & Knowledge Engineering Advancing credit risk assessment in the retail banking industry: A hybrid approach using time series and supervised learning models,” vol. 160, 2025, doi: <https://doi.org/10.1016/j.rineng.2025.105703>.
- [20] K. Roumpas, E. L. Michanetzi, D. Minas, A. Fotopoulos, and M. Xenos, “A comparative study of physics-informed and conventional neural networks for predicting on-screen gaze points from eye-tracking data,” *Expert Syst. Appl.*, vol. 281, no. February, 2025, doi: 10.1016/j.eswa.2025.127655.
- [21] M. Faisal *et al.*, “Utilizing Machine Learning-Based DecisionMaking to Align Higher Education Curriculum with Industry Requirements,” *Int. J. Mod. Educ. Comput. Sci.*, vol. 17, no. 4, pp. 1–25, 2025, doi: 10.5815/ijmecs.2025.04.01.
- [22] M. Faisal and T. K. A. Rahman, “Optimally Enhancement Rural Development Support Using Hybrid Multy Object Optimization (MOO) and Clustering Methodologies: A Case South Sulawesi - Indonesia,” *Int. J. Sustain. Dev. Plan.*, vol. 18, no. 6, pp. 1659–1669, 2023, doi: 10.18280/ijstdp.180602.
- [23] T. Jin, “Methods and reliability study of moral education assessment in universities: A machine learning-based approach,” *Alexandria Eng. J.*, vol. 125, no. March, pp. 20–28, 2025, doi: 10.1016/j.aej.2025.03.095.
- [24] M. P. Schneider, C. Malegori, A. de A. Gomes, and P. Oliveri, “Enhancing one-class classification performance through variable selection: A review based on advanced literature search approaches,” *Chemom. Intell. Lab. Syst.*, vol. 265, no. July, p. 105491, 2025, doi: 10.1016/j.chemolab.2025.105491.
- [25] L. Wang and H. Zhou, “A Study on Improving the Teaching Quality of College English Classrooms Based on Fuzzy Evaluation,” *Syst. Soft Comput.*, p. 200407, 2025, doi: 10.1016/j.sasc.2025.200407.