

KLASIFIKASI DAMPAK KECANDUAN MEDIA SOSIAL MAHASISWA DENGAN SVM DAN K-MEANS

Syarip Saputra¹⁾, Muhamad Fakhri Hidayat²⁾, Wahyu Gunawan³⁾, Indra Riyana Rahadjeng⁴⁾

^{1,2,3,4)} Informatika Universitas Bina Sarana Informatika

email : 15230747@bsi.ac.id ¹⁾, 15230749@bsi.ac.id ²⁾, 15230776@bsi.ac.id ³⁾,
riyana.irr@bsi.ac.id ⁴⁾

Abstraksi

Penggunaan media sosial yang meningkat di kalangan mahasiswa, disertai pola tidur tidak teratur, menjadi permasalahan serius karena berpotensi menurunkan prestasi akademik dan sering terlambat terdeteksi oleh institusi pendidikan. Penelitian ini bertujuan mengembangkan model prediksi hybrid dua tahap untuk mengklasifikasikan dampak kecanduan media sosial terhadap mahasiswa sebagai sistem peringatan dini. Metode yang digunakan mengintegrasikan algoritma K-Means untuk pelabelan tingkat risiko secara otomatis dan Support Vector Machine (SVM) sebagai tahap akhir klasifikasi. Penelitian menggunakan 705 data responden mahasiswa yang diproses melalui tahap preprocessing. Jumlah cluster optimal ditentukan menggunakan metode Elbow dan Silhouette Score, sedangkan kinerja model dievaluasi menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa K-Means membentuk tiga kluster risiko (rendah, sedang, dan tinggi) dengan nilai Silhouette Score sebesar 0,4188. Model SVM menghasilkan akurasi 93,62%, presisi 94,62%, dan recall 90%, sehingga efektif mendukung pengambilan keputusan preventif di perguruan tinggi.

Kata Kunci :

Kecanduan Media Sosial, Prestasi Akademik, K-Means, Support Vector Machine, Mahasiswa.

Abstract

The increasing use of social media among university students, accompanied by irregular sleep patterns, has become a serious issue as it can reduce academic performance and is often detected too late by educational institutions. This study aims to develop a two-stage hybrid prediction model to classify the impact of social media addiction on students as an early warning system. The proposed method integrates the K-Means algorithm for automatic risk-level labeling and Support Vector Machine (SVM) as the final classification stage. The study used data from 705 student respondents processed through data preprocessing. The optimal number of clusters was determined using the Elbow method and Silhouette Score, while model performance was evaluated using a confusion matrix. The results show that K-Means formed three risk clusters (low, medium, and high) with a Silhouette Score of 0.4188. The SVM model achieved an accuracy of 93.62%, precision of 94.62%, and recall of 90%, indicating its effectiveness in supporting preventive decision-making in higher education.

Keywords :

Social Media Addiction, Academic Performance, K-Means, Support Vector Machine, Students

Pendahuluan

Penggunaan media sosial di kalangan mahasiswa terus meningkat dan berpotensi mengganggu prestasi akademik. Najwa Safira (2025) menemukan positif signifikan antara kecanduan TikTok dengan prokrastinasi akademik pada mahasiswa Gen Z ($r = 0,449$; $p < 0,001$)[1]. Penggunaan TikTok yang berlebihan juga terbukti menurunkan motivasi belajar mahasiswa secara signifikan[2]. Meilisad kk. (2025) juga membuktikan bahwa kecanduan Instagram berhubungan negatif dengan indeks prestasi kumulatif mahasiswa D3 Keperawatan ($p\text{-value} = 0,001$)[3]. Selain itu, pola tidur yang tidak teratur (durasi kurang dari 6 jam atau mulai tidur setelah pukul 00:00) akibat penggunaan gadget pada malam hari terbukti menurunkan konsentrasi dan performa akademik mahasiswa[4]. Saat ini, prediksi risiko penurunan prestasi masih bergantung pada survei manual yang memakan waktu, tidak efisien, dan rentan terhadap subjektivitas[1][3]. Kondisi ini

menjadi permasalahan serius bagi perguruan tinggi karena penurunan prestasi akademik sering kali baru terdeteksi setelah mahasiswa mengalami kegagalan akademik yang signifikan. Ketergantungan pada metode evaluasi manual, seperti kuesioner dan penilaian periodik, menyebabkan keterlambatan dalam mengidentifikasi mahasiswa berisiko serta menyulitkan institusi untuk melakukan intervensi secara tepat waktu. Tanpa adanya sistem prediksi berbasis data yang terstruktur dan objektif, potensi penurunan prestasi akademik akibat kecanduan media sosial dan gangguan pola tidur sulit dicegah sejak dini. Oleh karena itu, dibutuhkan pendekatan analitik berbasis data mining yang mampu mendeteksi risiko akademik secara otomatis, akurat, dan berkelanjutan sebagai sistem peringatan dini di lingkungan perguruan tinggi. Beberapa penelitian telah berhasil mengelompokkan mahasiswa menjadi 3 cluster prestasi menggunakan algoritma K-Means[5][6] serta membuktikan bahwa Support

Vector Machine (SVM) memberikan akurasi tertinggi (94%) di bandingkan Naïve Bayes dan Random Forest. Namun, integrasi faktor non-akademik seperti kecanduan TikTok, Instagram, dan gangguan pola tidur dalam satu model prediksi dua tahap belum pernah dilaporkan sebelumnya.

Penelitian ini menawarkan metode baru berbasis dua tahap. Pertama, mahasiswa diklasifikasikan menggunakan algoritma K-Means berdasarkan

Penelitian sebelumnya [1],[3],[4],[5],[7],[8],[6],[9],[10] belum menunjukkan kombinasi faktor-faktor tersebut dalam satu pipeline hybrid K-Means + SVM. Penelitian ini tidak bertujuan membandingkan performa algoritma, melainkan memanfaatkan keunggulan K-Means sebagai pembentuk kluster risiko dan Support Vector Machine (SVM) sebagai model klasifikasi akhir dalam satu pendekatan hybrid.

Hasil penelitian di harapkan menghasilkan model prediksi dengan akurasi tinggi yang dapat diimplementasikan sebagai sistem peringatan dini (early warning sistem) di perguruan tinggi, sehingga institusi dapat melakukan intervensi preventif sebelum terjadi penurunan prestasi signifikan.

Tinjauan Pustaka

Nazwa Safira (2025) dalam penelitiannya terhadap 107 mahasiswa Gen Z menemukan adanya hubungan positif signifikan antara tingkat kecanduan TikTok dengan prokrastinasi akademik ($r = 0,449$; $p < 0,001$). Semakin tinggi skor kecanduan TikTok, semakin besar kecenderungan mahasiswa menunda penyelesaian tugas akademik. Hal ini memperkuat dugaan bahwa penggunaan TikTok yang berlebihan menjadi salah satu faktor risiko penurunan prestasi akademik[1]. Penelitian berikutnya yang dilakukan oleh Meilisa dkk. (2025) pada 78 mahasiswa D3 Keperawatan Universitas Abulyatama dan memperoleh hasil uji Chi-Square dengan $p\text{-value} = 0,001$ ($\alpha < 0,05$). Mahasiswa dengan tingkat kecanduan Instagram kategori tinggi mayoritas memiliki IPK pada kategori sedang hingga rendah. Temuan ini menegaskan bahwa kecanduan Instagram dapat berdampak negatif terhadap pencapaian akademik mahasiswa[3]. Intan Marhenisaputri (2025) mengungkapkan bahwa pola tidur yang tidak teratur (durasi kurang dari 6 jam atau mulai tidur setelah pukul 00.00) yang sering di picu penggunaan gadget berhubungan negatif signifikan dengan konsentrasi dan performa akademik mahasiswa ($r = -0,412$; $p < 0,001$)[4]. Penelitian Wulandari dan Pranata (2024) terhadap 52 mahasiswa juga menemukan bahwa 27 % responden memiliki kualitas tidur buruk dan 56 % mengalami disfungsi siang hari yang sangat buruk, sehingga pola tidur yang tidak teratur berpotensi besar menurunkan konsentrasi serta prestasi akademik mahasiswa[11]. Dampak kecanduan media sosial terhadap kesehatan

intensitas penggunaan media sosial dan pola tidur mereka, dan kemudian dibagi ke dalam tiga tingkat risiko prestasi akademik: rendah, sedang, dan tinggi. Tahap kedua adalah klasifikasi selesai menggunakan Support Vector Machine (SVM). Penggunaan data non-akademik terutama tingkat kecanduan TikTok, kecanduan Instagram, dan gangguan pola tidur sebagai variabel penentu utama dalam alur prediksi penelitian ini membedakannya dari penelitian sebelumnya.

mental mahasiswa juga menjadi perhatian serius, di mana penggunaan berlebihan terbukti dapat menurunkan kesejahteraan psikologis secara signifikan. Sebanyak 72,4 responden yang tidur <6 jam memiliki IPK <3,00[4]. Moh. Aqi Mukhtar Alfarerad dkk. (2024) menerapkan algoritma K-Means pada data prestasi mahasiswa Universitas Malang tahun 2018-2022. Hasil evaluasi menggunakan Davies-Bouldin menunjukan nilai terkecil 0,079 pada $K = 3$, sehingga terbentuk tiga kelompok prestasi yang jelas: rendah, sedang, dan tinggi. Penelitian ini menjadi bukti bahwa K-Means sangat efektif untuk mengelompokan data prestasi mahasiswa menjadi tiga kategori[5]. Penggunaan TikTok yang berlebihan juga terbukti menurunkan motivasi belajar mahasiswa secara signifikan[12]. Satrio Junaidi dkk. (2024) membandingkan empat algoritma klasifikasi untuk memprediksi kelulusan tepat waktu mahasiswa, yaitu Naive Bayes (87%), Random Forest (84%), ANN (65%), dan Support Vector Machine (94%). SVM menunjukan akurasi tertinggi (94%) dengan pembagian data 70% training dan 30% testing. Oleh karena itu, algoritma SVM di pilih sebagai metode klasifikasi dalam penelitian ini[7]. Pendekatan machine learning telah banyak digunakan dalam memprediksi performa dan kelulusan mahasiswa sebagai bagian dari sistem peringatan dini akademik, dengan Support Vector Machine (SVM) menjadi salah satu algoritma yang paling sering digunakan karena performanya yang stabil dan akurat [13]. Hal ini diperkuat oleh Albreiki et al. (2021) melalui systematic literature review yang mengkaji penelitian Educational Data Mining periode 2009–2021, yang menyimpulkan bahwa algoritma machine learning seperti Support Vector Machine (SVM), Random Forest, dan Neural Network sangat efektif digunakan untuk memprediksi mahasiswa berisiko dan mendukung sistem peringatan dini akademik. Studi tersebut juga menegaskan bahwa pendekatan klasifikasi berbasis machine learning mampu membantu institusi pendidikan dalam melakukan intervensi dini terhadap mahasiswa yang berpotensi mengalami penurunan performa akademik[14]. Selanjutnya, Lu et al. (2025) menegaskan bahwa pendekatan hybrid yang mengombinasikan teknik klasterisasi dengan model prediktif mampu meningkatkan akurasi identifikasi mahasiswa berisiko dengan menangkap perbedaan karakteristik akademik dan perilaku antar kelompok[15]. Selain itu, pengembangan sistem

peringatan dini akademik juga banyak dilakukan dengan pendekatan klasterisasi. Liu et al. (2025) mengusulkan model academic early warning berbasis improved global K-Means yang mampu meningkatkan akurasi dan efisiensi komputasi dalam mendeteksi mahasiswa berisiko. Hasil penelitian tersebut menunjukkan bahwa metode klasterisasi sangat efektif digunakan sebagai dasar pembentukan label risiko sebelum dilakukan analisis lanjutan, sehingga mendukung penerapan K-Means sebagai komponen penting dalam sistem peringatan dini akademik[16]. Penelitian yang dilakukan oleh Sandeepa dan Mohottala menunjukkan bahwa pendekatan *machine learning*, khususnya algoritma klasifikasi seperti Support Vector Machine (SVM) dan neural network, mampu memberikan performa yang kompetitif dalam memprediksi prestasi akademik mahasiswa berbasis data perilaku dan akademik[17]. Penelitian yang dilakukan oleh Alamri et al. (2020) menunjukkan bahwa algoritma Support Vector Machine (SVM) mampu menghasilkan tingkat akurasi hingga 93% dalam memprediksi performa akademik mahasiswa, sehingga efektif digunakan sebagai bagian dari sistem peringatan dini akademik[18]. Temuan tersebut memperkuat pemilihan SVM dalam penelitian ini, karena algoritma tersebut terbukti memiliki performa yang stabil dan akurat dalam menangani data akademik mahasiswa dengan berbagai faktor penentu. Selain itu, penelitian tersebut menegaskan bahwa pemilihan fitur yang tepat memiliki peran penting dalam meningkatkan kinerja model, sehingga integrasi data perilaku dan akademik menjadi faktor krusial dalam sistem prediksi performa mahasiswa.

Berdasarkan kajian pustaka di atas di atas[1],[7], penggunaan TikTok dan Instagram yang berlebihan serta pola tidur yang buruk merupakan faktor-faktor yang berpotensi menurunkan prestasi akademik mahasiswa. Untuk itu, data prestasi mahasiswa akan di kelompokkan terlebih dahulu menggunakan K-means menjadi tiga cluster (rendah, sedang, tinggi), kemudian di lakukan klasifikasi risiko menggunakan SVM guna memberikan peringatan dini kepada mahasiswa yang berpotensi mengalami penurunan prestasi akademik.

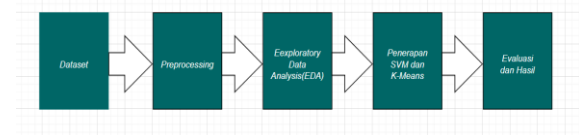
Penelitian sebelumnya umumnya hanya menganalisis satu atau dua faktor risiko secara terpisah dan belum mengintegrasikan ketiga faktor (TikTok, Instagram, dan pola tidur) sekaligus. Selain itu, kombinasi dua tahap K-Means Clustering dan SVM Clasification pada data mahasiswa aktif di suatu institusi yang sama belum pernah di lakukan. Inilah letak kebaruan penelitian ini.

Penelitian ini mengisi celah dengan menggabungkan analisis faktor risiko sekaligus, menerapkan K-Means yang telah terbukti optimal menghasilkan tiga cluster, serta menggunakan SVM yang memiliki akurasi tertinggi (94%) untuk klasifikasi. Hasilnya di

harapkan menjadi sistem peringatan dini yang lebih akurat bagi institusi pendidikan.

Metode Penelitian

Gambar 1 menunjukkan alur tahapan penelitian yang digunakan dalam penelitian ini, mulai dari pengumpulan dataset, preprocessing data, analisis eksploratori (EDA), penerapan algoritma K-Means dan Support Vector Machine (SVM), hingga tahap evaluasi hasil. Metode penelitian ini disusun secara sistematis melalui lima tahapan utama yang berurutan, yaitu pengumpulan data, pra-pemrosesan, analisa eksploratori, pemodelan K-Means dan SVM, serta evaluasi hasil. Alur tersebut mengacu pada kerangka CRISP-DM yang umum di gunakan dalam proyek data mining [7]. Penerapan metode CRISP-DM pada analisis data siswa telah terbukti membantu peneliti dalam mengelola proses data mining secara terstruktur, khususnya pada tahap preprocessing, pemodelan, dan evaluasi hasil [19].



Gambar 1. Tahap penelitian

1. Dataset

Dalam penelitian ini, dataset yang di gunakan berasal dari dataset publik “Students Social Media Addiction” yang di peroleh dari website Kaggle. Dataset ini terdiri dari 705 responden mahasiswa dari berbagai negara dengan 13 atribut, mencakup rata-rata jam penggunaan media sosial per hari, skor kecanduan, jam tidur malam, serta skor Kesehatan mental.

2. Data Preprocessing

Setelah dataset Students Social Media Addiction diperoleh dari platform Kaggle, tahap berikutnya adalah melakukan data preprocessing. Langkah ini bertujuan untuk memastikan bahwa data berada dalam kondisi bersih, konsisten, serta siap digunakan pada proses analisis SVM dan K-Means[7]. Prosedur yang dilakukan pada tahap ini meliputi beberapa langkah, seperti Data cleaning (Memeriksa dan menangani nilai hilang atau missing value), Encoding Variabel Kategorik (variabel kategorik dikonversi menggunakan metode Label Encoding atau One-Hot Encoding), standarisasi (Untuk menghindari dominasi satu variabel terhadap variabel lainnya), Pembuatan Standarisasi (Scaling), (Proses ini bertujuan untuk Menyesuaikan skala semua fitur numerik agar berada dalam rentang yang seragam. Ini sangat penting untuk SVM dan K-Means).

3. Exploratory Data Analysis (EDA)

Proses analisis data dimulai dengan Exploratory Data Analysis (EDA). Tujuannya adalah untuk memperoleh pemahaman tentang struktur, pola, tren, dan ciri-ciri penting dari set data sebelum melakukan

pemodelan statistik atau analisis inferensial lebih lanjut.

4. Klasifikasi SVM & K-MEANS

Pada penelitian ini, algoritma K-Means dan Support Vector Machine (SVM) tidak digunakan untuk tujuan perbandingan performa, melainkan dikombinasikan dalam satu alur kerja (hybrid) secara berurutan. Tahap pertama dilakukan dengan menerapkan algoritma K-Means untuk membentuk kluster risiko prestasi akademik berdasarkan variabel non-akademik, sehingga dihasilkan label risiko rendah, sedang, dan tinggi. Label hasil klusterisasi ini selanjutnya digunakan sebagai data berlabel pada tahap klasifikasi menggunakan algoritma SVM.

Dataset yang telah melalui tahap preprocessing kemudian dibagi menjadi dua subset, yaitu data latih (training data) sebesar 80% dan data uji (testing data) sebesar 20%. Data latih digunakan untuk melatih model SVM agar mampu mempelajari pola risiko prestasi akademik, sedangkan data uji dimanfaatkan untuk menguji performa model yang telah dibangun[5],[7].

5. Evaluasi

Melakukan evaluasi kinerja model adalah langkah terakhir dalam penelitian ini. Pada titik ini, kemampuan kedua model yang digunakan, Support Vector Machine (SVM) untuk klasifikasi dan K-Means untuk klusterisasi, dinilai berdasarkan kapasitas mereka untuk memproses data yang telah melewati tahap preprocessing sebelumnya.

Untuk model SVM, evaluasi dilakukan dengan membandingkan hasil prediksi model dengan label data uji asli. Accuracy, precision, recall, dan F1-score adalah beberapa metrik evaluasi standar yang digunakan untuk mengevaluasi kinerja model[7]. Digunakannya metrik ini untuk menilai kemampuan model untuk mengklasifikasikan tingkat kecanduan secara akurat dan konsisten.

Untuk saat ini, algoritma K-Means dievaluasi menggunakan skor silhouette, metrik klusterisasi khusus. Metrik ini digunakan untuk mengevaluasi kohesi dalam kluster yang terbentuk dan kualitas pemisahan antar kluster[5]. Nilai silhouette yang lebih tinggi menunjukkan bahwa struktur kluster yang dibuat lebih baik dan lebih representatif terhadap pola data daripada yang ditunjukkan oleh nilai silhouette yang lebih rendah.

Hasil dan Pembahasan

1. Dataset

Pada penelitian ini, data yang digunakan diperoleh dari sumber terbuka yang tersedia pada platform Kaggle. Dataset yang digunakan adalah *Students Social Media Addiction*, yaitu kumpulan data survei yang mencatat kebiasaan mahasiswa dalam

menggunakan media sosial serta faktor-faktor yang memengaruhi tingkat kecanduan. Dataset ini terdiri dari 705 responden dengan 13 atribut, yang mencakup informasi demografis, intensitas penggunaan media sosial, pola tidur, serta kondisi kesehatan mental. Deskripsi lengkap setiap atribut yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Deskripsi Atribut Dataset Penelitian

No	Atribut	Keterangan
1	<i>Student_Id</i>	Pengidentifikasi unik responden
2	<i>Age</i>	Usia
3	<i>Gender</i>	"Pria" atau "Wanita"
4	<i>Academic_Level_Categorical</i>	SMA/Magister/Sarjana
5	<i>Country</i>	Negara tempat tinggal
6	<i>Avg_Daily_Usage_Hours</i>	Rata-rata jam per hari di media sosial
7	<i>Most_Used_Platform</i>	Instagram, Facebook, TikTok, dll.
8	<i>Affects_Academic_Performance</i>	Dampak akademi mahasiswa (Ya/Tidak)
9	<i>Sleep_Hours_Per_Night</i>	Jam tidur per malam
10	<i>Mental_Health_Score</i>	Skor kesehatan mental
11	<i>Relationship_Status</i>	Lajang/Berpacaran/Rumit
12	<i>Conflicts_Over_Social_Media</i>	Jumlah masalah di sosial media
13	<i>Addicted_Score</i>	Skor kecanduan medsos (1=rendah hingga 10=Tinggi)

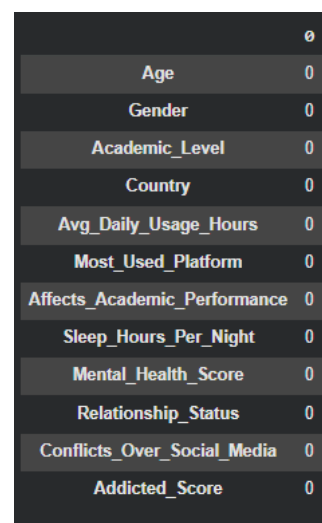
Sumber: (Kaggle, 2025)

2. Preprocessing

Langkah-langkah pra-pemrosesan data dilakukan guna menyiapkan data untuk pengolahan, antara lain:

A. Cleaning

Pembersihan data dilakukan menggunakan dataset *Students_Social_Media_Addiction.csv* untuk menghilangkan *missing value* atau data yang tidak lengkap. Seperti ditunjukkan pada Gambar 2, hasil pemeriksaan menunjukkan bahwa seluruh atribut memiliki nilai *missing* sebesar 0, yang menandakan bahwa dataset berada dalam kondisi lengkap dan siap digunakan untuk tahap pemrosesan selanjutnya tanpa memerlukan imputasi data.



Attribute	Missing Value Count
Age	0
Gender	0
Academic_Level	0
Country	0
Avg_Daily_Usage_Hours	0
Most_Used_Platform	0
Affects_Academic_Performance	0
Sleep_Hours_Per_Night	0
Mental_Health_Score	0
Relationship_Status	0
Conflicts_Over_Social_Media	0
Addicted_Score	0

Gambar 2. Hasil pemeriksaan missing value pada dataset

B. Categorical Data Encoding (One-Hot Encoding)

Pada tahap ini, variabel kategorikal yang tidak dapat diproses secara langsung oleh algoritma machine learning, seperti atribut *gender*, dikonversi ke dalam

bentuk numerik menggunakan metode One-Hot Encoding. Seperti ditunjukkan pada Gambar 3, setiap kategori direpresentasikan sebagai vektor biner untuk menghindari adanya hubungan ordinal semu antar kategori. Proses ini bertujuan agar algoritma K-Means dan Support Vector Machine (SVM) dapat memproses data kategorikal secara optimal tanpa menimbulkan bias pada proses pembelajaran model.

```

--- 5 BARIS PERTAMA KOLOM YANG SUDAH DI ENCODE ---
      Age Academic_Level Gender_Male
Student_ID
1      19 Undergraduate      False
2      22 Graduate          True
3      20 Undergraduate      False
4      18 High School        True
5      21 Graduate          True

--- Keterangan Kolom Gender_Male ---
Nilai 1 = Male (Laki-laki)
Nilai 0 = Female (Perempuan)

```

Gambar 3. Hasil one-hot encoding pada variabel kategorikal

C. Standarisasi (Scaling)

Tahap standarisasi (scaling) dilakukan untuk menyesuaikan skala seluruh fitur numerik agar berada dalam rentang yang seragam. Seperti ditunjukkan pada Gambar 4, proses ini bertujuan untuk mencegah dominasi atribut dengan nilai numerik yang lebih besar terhadap atribut lainnya. Standarisasi sangat penting terutama pada algoritma K-Means yang berbasis jarak dan Support Vector Machine (SVM) yang sensitif terhadap perbedaan skala fitur, sehingga model dapat mempelajari pola data secara lebih adil dan optimal.

```

--- 5 BARIS PERTAMA DATA NUMERIK SETELAH STANDARDISASI ---
      Age Avg_Daily_Usage_Hours Sleep_Hours_Per_Night
0 -1.186915      0.223857      -0.327638
1  0.958662     -2.243309      0.560423
2 -0.471723      0.860545     -1.659730
3 -1.902108     -1.527035      0.116393
4  0.243470     -0.333245     -0.771668

      Mental_Health_Score Conflicts_Over_Social_Media
0      -0.205520      0.157063
1      1.605628     -2.976790
2     -1.111095      1.201681
3      0.700054     -1.932172
4     -0.205520     -0.887554

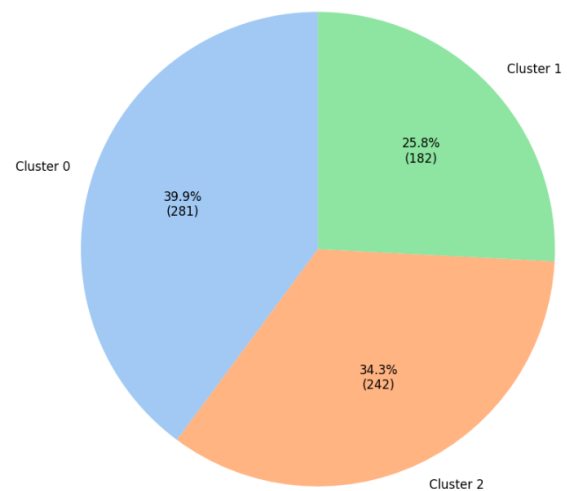
```

Gambar 4. Hasil proses standarisasi (scaling) data

3. Exploratory Data Analysis (EDA)

Proses Exploratory Data Analysis (EDA) dilakukan menggunakan platform Google Colab untuk memahami pola awal dan karakteristik data sebelum proses pemodelan. Analisis ini bertujuan untuk mengidentifikasi distribusi data, hubungan antarvariabel, serta kecenderungan perilaku penggunaan media sosial pada siswa. Berdasarkan hasil klasterisasi yang diterapkan pada 705 data siswa, terbentuk tiga kelompok (cluster) dengan karakteristik yang berbeda terkait pola penggunaan media sosial, durasi tidur, dan kondisi kesehatan mental. Distribusi persentase anggota pada masing-

masing klaster ditunjukkan pada Gambar 5.



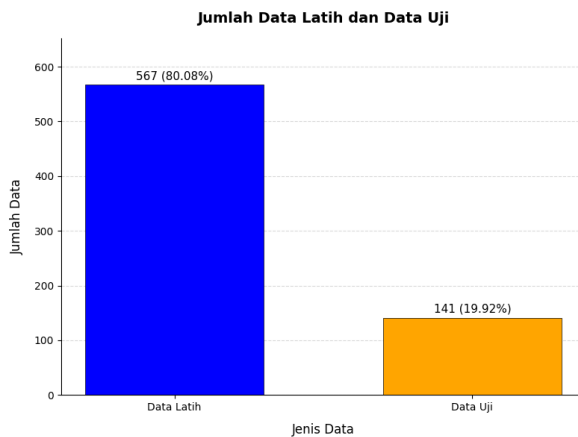
Gambar 5. Distribusi klastering

Secara keseluruhan, hasil segmentasi menunjukkan bahwa mayoritas siswa (gabungan Cluster 0 dan Cluster 2, sebesar 74,2%) berada dalam kategori yang relatif aman dengan kondisi kesehatan yang terjaga, mulai dari penggunaan sedang hingga rendah. Keberadaan Cluster 1 adalah temuan yang harus diperhatikan. Sangat penting untuk melakukan intervensi karena 182 siswa (atau 25,8%) mengalami risiko prestasi tinggi yang dikombinasikan dengan kesehatan mental yang buruk dan pola tidur yang buruk. Menurut data, untuk mencegah penurunan kualitas kesehatan mental lebih lanjut, strategi penanganan atau konseling harus diterapkan secara merata kepada seluruh siswa. Sebaliknya, strategi ini dapat difokuskan pada siswa di Cluster 1 secara khusus.

4. Klasifikasi Algoritma

Pada tahap ini, data siswa diklasifikasikan ke dalam dua kategori risiko prestasi akademik, yaitu risiko rendah dan risiko tinggi. Penentuan label kelas didasarkan pada hasil klasterisasi K-Means yang telah diperoleh sebelumnya. Cluster 1 yang menunjukkan tingkat kecanduan media sosial tinggi, kondisi kesehatan mental buruk, serta pola tidur tidak teratur dikategorikan sebagai kelas risiko tinggi. Sementara itu, Cluster 0 (penggunaan sedang dengan kesejahteraan seimbang) dan Cluster 2 (penggunaan rendah dengan kondisi kesejahteraan baik) digabungkan ke dalam kelas risiko rendah, karena masih berada dalam batas kondisi akademik yang wajar dan sehat. Dataset selanjutnya dibagi menjadi data latih dan data uji dengan proporsi 80% untuk pelatihan dan 20% untuk pengujian. Dari total 705 data siswa, sebanyak 567 data digunakan sebagai data latih untuk membangun model klasifikasi, sedangkan 141 data digunakan sebagai data uji untuk mengevaluasi kinerja dan kemampuan generalisasi

model. Distribusi pembagian data latih dan data uji tersebut ditunjukkan pada Gambar 6.



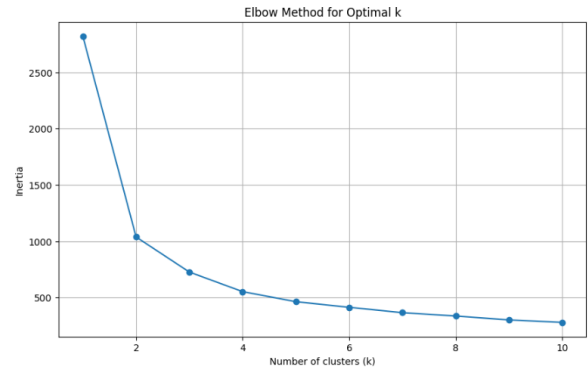
Gambar 6. train & test

Pada tahap selanjutnya, algoritma Support Vector Machine (SVM) digunakan untuk melakukan proses pelatihan model klasifikasi berdasarkan label risiko yang telah dibentuk pada tahap klusterisasi sebelumnya. Evaluasi kinerja model dilakukan menggunakan data uji dengan mengacu pada metrik accuracy, precision, recall, F1-score, dan ROC AUC.

Laporan Klasifikasi SVM:				
	precision	recall	f1-score	support
High	1.00	1.00	1.00	76
Low	1.00	1.00	1.00	55
Medium	1.00	1.00	1.00	9
accuracy			1.00	140
macro avg	1.00	1.00	1.00	140
weighted avg	1.00	1.00	1.00	140

Gambar 7. Hasil klasifikasi support vector machine (SVM)

Berdasarkan hasil evaluasi yang ditunjukkan pada Gambar 7, model SVM memperoleh akurasi sebesar 93,62% dengan nilai ROC AUC sebesar 0,975, yang menunjukkan kemampuan diskriminasi model yang sangat baik. Untuk kelas risiko prestasi akademik rendah, diperoleh nilai precision sebesar 96% dan recall sebesar 95%, sedangkan kelas risiko prestasi akademik tinggi menunjukkan precision sebesar 88% dan recall sebesar 90%. Hasil ini menghasilkan F1-score sebesar 96% pada kelas risiko rendah dan 88% pada kelas risiko tinggi, dengan distribusi data uji masing-masing 101 data risiko rendah dan 40 data risiko tinggi. Tingginya nilai recall pada kelas risiko prestasi akademik tinggi menunjukkan bahwa model mampu mengidentifikasi sebagian besar mahasiswa yang berpotensi mengalami penurunan prestasi akademik dengan tingkat kesalahan *false negative* yang relatif rendah. Hal ini menegaskan bahwa pendekatan hybrid K-Means dan SVM efektif digunakan sebagai sistem peringatan dini untuk mendukung pengambilan keputusan preventif di lingkungan perguruan tinggi.



Gambar 8. Grafik elbow method untuk menentukan jumlah kluster optimal (k)

Evaluasi kualitas klusterisasi pada algoritma K-Means dilakukan menggunakan Elbow Method, sebagaimana ditunjukkan pada Gambar 8. Grafik memperlihatkan penurunan nilai inersia yang signifikan dari $k=1$ ke $k=2$ dan dari $k=2$ ke $k=3$. Setelah $k=3$, penurunan nilai inersia cenderung melandai, yang mengindikasikan terbentuknya titik siku (elbow point) pada $k=3$. Berdasarkan analisis tersebut, ditetapkan bahwa jumlah kluster optimal dalam penelitian ini adalah tiga kluster ($k=3$). Pemilihan ini sejalan dengan tujuan penelitian untuk mengelompokkan mahasiswa ke dalam kategori risiko prestasi akademik yang mudah diinterpretasikan.

The Silhouette Score for 3 clusters is: 0.4188

Gambar 9. Nilai silhouette score untuk tiga kluster ($k=3$)

Evaluasi kualitas klusterisasi dilakukan menggunakan Silhouette Score, sebagaimana ditunjukkan pada Gambar 9. Metrik ini digunakan untuk mengukur sejauh mana setiap objek lebih mirip dengan klusternya sendiri dibandingkan dengan kluster lain. Hasil pengujian menunjukkan bahwa nilai Silhouette Score rata-rata untuk jumlah kluster $k=3$ adalah sebesar 0,4188. Nilai ini berada pada kategori cukup (fair), yang mengindikasikan bahwa struktur kluster yang terbentuk memiliki tingkat kohesi dan pemisahan antar kluster yang memadai. Dengan demikian, kluster yang dihasilkan dapat digunakan secara andal sebagai dasar pelabelan risiko pada tahap klasifikasi selanjutnya.

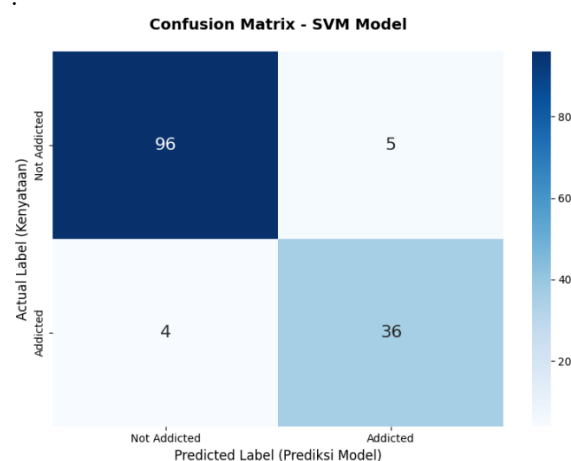
5. Evaluasi Performa Model

Evaluasi kinerja model dilakukan dengan menganalisis hasil dari performa model untuk menentukan model klasifikasi dan klusterisasi dari Support Vector Machine (SVM) dan K-Means.

A. Evaluasi Klasifikasi SVM (Confusion Matrix)

Gambar 9 menunjukkan confusion matrix hasil klasifikasi Support Vector Machine (SVM) terhadap data uji. Berdasarkan gambar tersebut, model berhasil mengklasifikasikan 96 data mahasiswa dengan risiko prestasi akademik rendah secara benar (true negative)

dan 36 data mahasiswa dengan risiko prestasi akademik tinggi secara benar (true positive). Sementara itu, terdapat 5 data yang salah diklasifikasikan sebagai risiko tinggi (false positive) dan 4 data risiko tinggi yang tidak terdeteksi oleh model (false negative). Berdasarkan pelabelan tersebut, evaluasi model klasifikasi Support Vector Machine (SVM) yang dilatih dengan proporsi data 80:20 menunjukkan performa yang sangat memuaskan dengan tingkat akurasi mencapai 93,62% dan nilai ROC AUC sebesar 0,975. Nilai recall sebesar 90% menunjukkan bahwa sistem mampu mengidentifikasi sebagian besar mahasiswa yang berada pada kategori risiko prestasi akademik tinggi dengan tingkat kesalahan false negative yang relatif rendah. Secara keseluruhan, kombinasi metode K-Means dan SVM terbukti efektif sebagai instrumen deteksi dini untuk mengklasifikasikan risiko prestasi akademik mahasiswa.

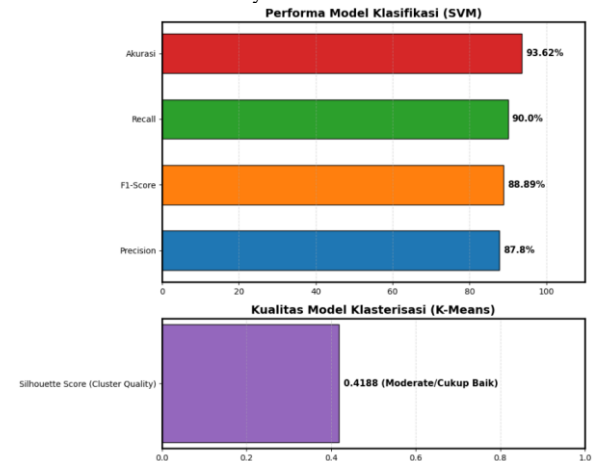


Gambar 10. Confusion matrix model SVM

B. Evaluasi Klasterisasi K-Means

Gambar 10 menunjukkan hasil evaluasi klasterisasi menggunakan algoritma K-Means terhadap data mahasiswa. Analisis dimulai dengan pemetaan karakteristik mahasiswa menggunakan algoritma K-Means, di mana metode Elbow digunakan untuk menentukan jumlah klaster optimal dan menghasilkan tiga segmen utama. Hasil segmentasi menunjukkan bahwa Cluster 1 merupakan kelompok mahasiswa dengan risiko prestasi akademik tinggi, yang ditandai oleh tingkat penggunaan media sosial yang tinggi, kondisi kesehatan mental yang rendah, serta pola tidur yang buruk. Sebaliknya, Cluster 0 dan Cluster 2 merepresentasikan kelompok mahasiswa dengan risiko prestasi akademik rendah, dengan karakteristik penggunaan media sosial yang lebih wajar dan kondisi kesehatan mental yang lebih baik. Dengan nilai Silhouette Score sebesar 0,4188, kualitas pengelompokan ini berada pada kategori moderat, yang menunjukkan bahwa struktur pemisahan antar klaster cukup jelas dan dapat diandalkan sebagai dasar pelabelan risiko pada tahap klasifikasi selanjutnya.

Validasi Performa Model Hybrid



Gambar 11. Validasi Performa Model Hybrid

Gambar 11 memperlihatkan validasi performa model hybrid yang menggabungkan algoritma K-Means dan Support Vector Machine (SVM). Visualisasi hasil menunjukkan bahwa algoritma SVM memberikan performa klasifikasi yang tinggi dengan tingkat akurasi sebesar 93,62% dan nilai Recall sebesar 90%, yang mengindikasikan kemampuan model dalam mendeteksi mahasiswa dengan risiko prestasi akademik tinggi secara efektif. Di sisi lain, algoritma K-Means menghasilkan nilai Silhouette Score sebesar 0,4188 yang termasuk dalam kategori moderat, menandakan bahwa struktur klaster yang terbentuk memiliki kualitas pemisahan yang cukup baik. Klasterisasi ini digunakan sebagai dasar pembentukan label risiko pada tahap klasifikasi. Dengan demikian, kedua algoritma dalam penelitian ini tidak digunakan untuk tujuan perbandingan performa, melainkan saling melengkapi dalam satu alur kerja hybrid, di mana K-Means berperan sebagai pembentuk label risiko dan SVM sebagai model klasifikasi akhir.

6. Pembahasan

Pembicaraan dimulai dengan menganalisis pengelompokan profil siswa menggunakan algoritma K-Means. Metode Elbow menunjukkan bahwa pembagian data menjadi tiga cluster adalah konfigurasi terbaik. Hasil segmentasi menunjukkan bahwa kelompok siswa rentan pada Cluster 1 yang menunjukkan indikasi kecanduan yang tinggi dipisahkan dari kelompok pengguna wajar pada Cluster 0 dan 2, yang divalidasi dengan nilai Silhouette Score sebesar 0.4188. Nilai-nilai tersebut menunjukkan bahwa, meskipun data perilaku siswa sangat beragam, struktur pemisahan kelompok cukup jelas (kategori moderat) dan dapat digunakan sebagai dasar pelabelan otomatis (benar nyata) untuk proses klasifikasi berikutnya.

Dengan tingkat akurasi 93,62% dan nilai ROC AUC 0,975, evaluasi model Support Vector Machine (SVM) menunjukkan kinerja yang lebih baik dari proses klasterisasi. Temuan penting, terutama

tingginya nilai Recall sebesar 90%, menunjukkan bahwa model ini mendeteksi kasus positif (siswa kecanduan) dengan tingkat kesalahan yang sangat rendah (False Negative). Ini menunjukkan bahwa sistem yang menggunakan metode K-Means sebagai pembentuk label dan SVM sebagai prediktor tidak hanya akurat secara statistik, tetapi juga dapat diandalkan sebagai alat untuk mendeteksi kecanduan media sosial secara dini. Hasil 93,62% pada penelitian ini juga lebih tinggi dibandingkan penelitian Riski Qishtiano (2022) yang menggunakan SVM dengan kernel RBF untuk memprediksi kelulusan mahasiswa tepat waktu dan hanya memperoleh akurasi 85,06% [20]

Kesimpulan dan Saran

Berdasarkan tujuan penelitian untuk mengklasifikasikan risiko prestasi akademik mahasiswa akibat kecanduan media sosial, dapat disimpulkan bahwa pendekatan hybrid K-Means dan Support Vector Machine (SVM) mampu memberikan hasil yang efektif dan terstruktur. Algoritma K-Means berhasil mengelompokkan mahasiswa ke dalam tiga kluster dengan karakteristik yang berbeda, sehingga memungkinkan identifikasi kelompok mahasiswa berisiko tinggi dan kelompok pengguna media sosial yang masih berada dalam batas wajar. Nilai Silhouette Score sebesar 0,4188 menunjukkan bahwa kualitas pemisahan kluster berada pada kategori cukup baik dan layak digunakan sebagai dasar pelabelan risiko.

Selanjutnya, model klasifikasi SVM yang dibangun berdasarkan hasil klusterisasi tersebut menunjukkan performa yang tinggi dengan tingkat akurasi sebesar 93,62% dan nilai ROC AUC sebesar 0,975. Nilai recall sebesar 90% pada kelas risiko tinggi menegaskan bahwa model mampu mendeteksi sebagian besar mahasiswa yang berpotensi mengalami penurunan prestasi akademik akibat kecanduan media sosial. Hasil ini menunjukkan bahwa integrasi metode unsupervised dan supervised learning dapat dimanfaatkan sebagai sistem peringatan dini dalam konteks akademik.

Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama pada kualitas pemisahan kluster yang berada pada tingkat moderat serta keterbatasan variabel yang digunakan. Oleh karena itu, penelitian selanjutnya disarankan untuk menambahkan variabel yang lebih spesifik, seperti jenis konten yang dikonsumsi, durasi penggunaan berdasarkan waktu (siang/malam), serta faktor lingkungan akademik. Selain itu, eksplorasi algoritma lain atau optimasi parameter model berpotensi meningkatkan kualitas klusterisasi dan akurasi klasifikasi.

Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh perguruan tinggi sebagai alat skrining awal dalam program bimbingan dan konseling mahasiswa.

Dengan adanya sistem klasifikasi risiko berbasis data, institusi pendidikan dapat melakukan intervensi preventif secara lebih dini dan tepat sasaran untuk meminimalkan dampak negatif kecanduan media sosial terhadap prestasi akademik mahasiswa.

Daftar Pustaka

- [1] N. Safira, "Hubungan Antara Adiksi Tiktok Dengan Prokrastinasi," *J. Couns. Support*, vol. 1, pp. 77–88, 2025.
- [2] N. A.-K. N. F. U. R. R. J. Z. P. D. N. R. M. H. Muzaki, "PERSEPSI MAHASISWA UNIVERSITAS NEGERI MALANG TENTANG PENGGUNAAN MEDIA SOSIAL TIKTOK TERHADAP MOTIVASI BELAJAR," *Ptk Dan Pendidik.*, vol. 9, no. 1, pp. 59–73, 2023, doi: 10.18592/ptk.v9i1.12800.
- [3] Meilisa Meilisa, Syukriadi Syukriadi, and Ellyza Fazlylawati, "Hubungan Social Media Addiction Instagram terhadap Indeks Prestasi Kumulatif pada Mahasiswa Program Studi Diploma Tiga Keperawatan Abulyatama," *Inov. Kesehat. Glob.*, vol. 2, no. 4, pp. 70–81, 2025, doi: 10.62383/ikg.v2i4.2340.
- [4] I. Marhenisaputri *et al.*, "Pengaruh Pola Tidur Terhadap Konsentrasi Dan Performa Akademik Mahasiswa Fakultas Ilmu Tarbiyah Uin Raden Mas Said Surakarta," *J. Ilmu Ekon. Pendidik. dan Tek.*, vol. 2, pp. 72–77, 2025.
- [5] M. A. M. A. Aqil and Z. Fatah, "Analisis Pengelompokan Prestasi Mahasiswa Universitas Malang Menggunakan Metode K-Means Clustering," *JITU J. Inform. Technol. Commun.*, vol. 8, no. 2, pp. 29–40, 2024, doi: 10.36596/jitu.v8i2.1704.
- [6] W. Chang *et al.*, "Analysis of university students' behavior based on a fusion K-means clustering algorithm," *Appl. Sci.*, vol. 10, no. 18, 2020, doi: 10.3390/APP10186566.
- [7] Satrio Junaidi, R. Valicia Anggela, and D. Kariman, "Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Network (ANN)," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 109–119, 2024, doi: 10.52158/jacost.v5i1.489.
- [8] R. S. D. Puspita, "Classification Of Outstanding Students Using Support Vector Machine (SVM) Based on Data Mining," *Classif. Outst. Students Using Support Vector Mach. Based Data Min.*, vol. 7, no. 1, pp. 102–111, 2023, doi: 10.31289/jite.v9i1.13191.
- [9] F. O. W. E. R. Megawati, "Jurnal Kesehatan Masyarakat Indonesia Jurnal Kesehatan

- Masyarakat Indonesia,” vol. 1, no. 18, pp. 46–53, 2025.
- [10] M. Haryanto, A. C. Sidauruk, Y. B. Hendy, J. A. Sabailaket, D. R. Purba, and E. R. Handoyo, “Pengaruh Penggunaan Aplikasi TikTok terhadap Produktivitas Belajar Mahasiswa di Yogyakarta,” *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 330–341, 2023, doi: 10.24002/konstelasi.v3i2.7213.
- [11] S. Wulandari and R. Pranata, “Deskripsi Kualitas Tidur dan Pengaruhnya terhadap Konsentrasi Belajar Mahasiswa,” *J. Pendidik. Kesehat. Rekreasi*, vol. 10, no. 1, pp. 101–108, 2024, doi: 10.59672/jpkr.v10i1.3414.
- [12] S. Suraya, M. Sholeh, and D. Andayati, “Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa,” *Skanika*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skanika.v6i1.2982.
- [13] L. R. Pelima, Y. Sukmana, and Y. Rosmansyah, “Predicting University Student Graduation Using Academic Performance and Machine Learning: A Systematic Literature Review,” *IEEE Access*, vol. 12, no. January, pp. 23451–23465, 2024, doi: 10.1109/ACCESS.2024.3361479.
- [14] B. Albreiki, N. Zaki, and H. Alashwal, “A systematic literature review of student’ performance prediction using machine learning techniques,” *Educ. Sci.*, vol. 11, no. 9, 2021, doi: 10.3390/educsci11090552.
- [15] Y. Lu, S. Yeom, J. Maktoubian, M. M. Rahman, and S. H. Kim, “Improve Student Risk Prediction with Clustering Techniques: A Systematic Review in Education Data Mining,” *Educ. Sci.*, vol. 15, no. 12, pp. 1–38, 2025, doi: 10.3390/educsci15121695.
- [16] J.-J. Liu, Y. Yang, and F.-H. Qu, “Academic Early Warning Model Based on Improved Global K-means Algorithm,” *J. Comput.*, vol. 36, no. 2, pp. 105–118, 2025, doi: 10.63367/199115992025043602008.
- [17] A. G. R. Sandeepa and S. Mohottala, “Evaluation of Machine Learning Models in Student Academic Performance Prediction,” *2025 5th Int. Conf. Adv. Res. Comput. Converging Horizons Uniting Discip. Comput. Res. through AI Innov. ICARC 2025 - Proc.*, 2025, doi: 10.1109/ICARC64760.2025.10963104.
- [18] L. H. Alamri, R. S. Almuslim, M. S. Alotibi, D. K. Alkadi, I. Ullah Khan, and N. Aslam, “Predicting Student Academic Performance using Support Vector Machine and Random Forest,” *ACM Int. Conf. Proceeding Ser.*, vol. PartF16898, no. January, pp. 100–107, 2020, doi: 10.1145/3446590.3446607.
- [19] B. Sutara, F. Vulture, and R. Novianti, “Application of K-Means algorithm with CRISP-DM method in student data analysis as a support for promotion strategy SIDE: Scientifict Development Journal,” *SIDE Sci. Dev. J.*, vol. 1, no. 1, p. 7, 2024, [Online]. Available: <https://ojs.arbain.co.id/index.php/side/index>
- [20] M Riski Qisthiano, “Klasifikasi Terhadap Prediksi Kelulusan Mahasiswa Dengan Menggunakan Metode Support Vector Machine (Svm),” *Semin. Nas. Teknol. dan Multidisiplin Ilmu*, vol. 2, no. 2, pp. 203–207, 2022, doi: 10.51903/semnastekmu.v2i1.170.