JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Prediksi Tingkat Kecanduan Media Sosial pada Remaja Menggunakan Berbagai Algoritma Machine Learning untuk Klasifikasi Multikelas

Egi Safitri[1,*], Rini Nurlistiani[2], Nurjoko[1], Rekha Aprilia Andin[1], Novita Triyasri[1]

[1] Fakultas Ilmu Komputer, Sains Data, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, Indonesia
[2] Fakultas Ilmu Komputer, Sistem Informasi, Institut Informatika dan Bisnis Darmajaya, Bandar Lampung, Indonesia
Email : [1]egisafitri@darmajaya.ac.id, [2]rininurlistiani@darmajaya.ac.id, [3]nurjoko@darmajaya.ac.id
Email Penulis korespondensi : egisafitri@darmajaya.ac.id[*]
Submitted **04-06-2026**; Accepted **02-10-2026**; Published **31-10-2026**

Abstrak

Penggunaan media sosial yang berlebihan berkaitan dengan gangguan tidur, kondisi psikologis, aktivitas akademik, dan hubungan sosial pada remaja akhir serta dewasa muda. Penelitian ini bertujuan membandingkan kinerja Decision Tree, Random Forest, dan Support Vector Machine (SVM) linear dalam mengklasifikasikan tingkat kecanduan media sosial menggunakan dataset *Students' Social Media Addiction*. Analisis menggunakan 704 observasi setelah satu dari dua baris dengan nilai prediktor yang sama dikeluarkan. Berdasarkan *Addicted_Score*,data dikelompokkan menjadi kategori *low*(≤5), *medium* (6–8), dan *high* (>8). *Student_ID* dan *Addicted_Score* tidak disertakan sebagai prediktor. Data dibagi secara terstratifikasi menjadi 563 observasi latih dan 141 observasi uji. Pengodean kategorikal dan standardisasi fitur untuk SVM dilakukan melalui *pipeline*, dengan parameter transformasi yang ditentukan hanya dari bagian pelatihan. Pemilihan model didasarkan pada rerata *macro F1-score* validasi silang lima lipatan. Random Forest memperoleh nilai tertinggi sebesar 0,9708 ± 0,0292, diikuti SVM sebesar 0,9701 ± 0,0165 dan Decision Tree sebesar 0,9334 ± 0,0513. Pada data uji, Random Forest mencapai akurasi 97,87% dan *macro F1-score* 0,9836, sedangkan Decision Tree menghasilkan *macro F1-score* tertinggi sebesar 0,9889 dengan akurasi 98,58%. Ketiga model mengenali seluruh 11 observasi kategori *high*. Random Forest dipilih berdasarkan hasil validasi silang, meskipun selisihnya dengan SVM sangat kecil dan keunggulan statistiknya belum diuji. Hasil penelitian menunjukkan kemampuan model dalam mengklasifikasikan kategori skor pada dataset yang digunakan. Pengujian pada data independen dan verifikasi pembentukan skor diperlukan sebelum model dikembangkan sebagai alat skrining.

Kata Kunci: Kecanduan Media Sosial; *Machine Learning*; *Random Forest*; *Support Vector Machine*; *Decision Tree*

Abstract

Excessive social media use is associated with sleep disturbances, psychological conditions, academic activities, and social relationships among late adolescents and young adults. This study compares Decision Tree, Random Forest, and linear Support Vector Machine (SVM) classifiers for classifying social media addiction levels using the *Students' Social Media Addiction* dataset. The analysis included 704 observations after removing one of two rows with identical predictor values. Based on *Addicted_Score*, observations were assigned to low (≤5), medium (6–8), and high (>8) categories. *Student_ID*and *Addicted_Score* were excluded from the predictors. A stratified split yielded 563 training observations and 141 test observations. Categorical encoding and feature standardization for SVM were implemented through pipelines, with transformation parameters fitted only on the training partitions. Model selection was based on the mean macro F1-score from five-fold cross-validation. Random Forest achieved the highest score of 0.9708 ± 0.0292, followed by SVM at 0.9701 ± 0.0165 and Decision Tree at 0.9334 ± 0.0513. On the test set, Random Forest achieved 97.87% accuracy and a macro F1-score of 0.9836, while Decision Tree obtained the highest macro F1-score of 0.9889 with 98.58% accuracy. All three models correctly identified the 11 high-category observations. Random Forest was selected based on cross-validation, although its advantage over SVM was small and statistical significance was not tested. These findings demonstrate the models' ability to classify score-based categories within the dataset. Validation on independent data and verification of score construction are needed before developing the models for screening applications.

Keywords: *Social Media Addiction*; *Machine Learning*; *Random Forest*; *Support Vector Machine*; *Decision Tree*

1. PENDAHULUAN

Media sosial telah menjadi infrastruktur komunikasi sehari-hari bagi generasi muda. Interaksi akademik, hiburan, pencarian informasi, pembentukan identitas, dan relasi pertemanan kini berlangsung secara bersamaan melalui platform seperti Instagram, TikTok, YouTube, X/Twitter, Facebook, dan aplikasi pesan. Kondisi tersebut membuat intensitas penggunaan media sosial tidak lagi dapat dinilai hanya dari lamanya waktu layar, tetapi juga dari pola penggunaan yang kompulsif, kesulitan mengendalikan dorongan untuk terus terhubung, serta munculnya konsekuensi negatif dalam kehidupan sehari-hari. Dalam literatur, pola seperti ini sering dibahas sebagai *problematic social media use* atau *social media addiction*. Meta-analisis pada remaja dan dewasa muda menunjukkan bahwa penggunaan media sosial yang problematik memiliki hubungan yang bermakna dengan gejala depresi, kecemasan, dan stres [1]. Karena itu, upaya identifikasi dini perlu mempertimbangkan aspek perilaku, psikologis, tidur, akademik, dan sosial secara terpadu.

Keterkaitan antara kecanduan media sosial dan kualitas tidur merupakan salah satu pola yang paling konsisten dilaporkan. Zhu et. al. menemukan bahwa *social media addiction* berasosiasi dengan gangguan kualitas tidur pada mahasiswa [2] . Zhuang et. al. juga menunjukkan bahwa kualitas tidur dan kelelahan menjadi jalur penting yang menghubungkan kecanduan media sosial dengan keterlibatan akademik mahasiswa [3]. Temuan-temuan ini relevan bagi analisis berbasis data karena durasi tidur dan intensitas penggunaan media sosial dapat menjadi indikator yang relatif

This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

mudah diukur. Namun, interpretasi hubungan tersebut harus tetap hati-hati karena sebagian besar studi menggunakan desain observasional, sehingga hubungan statistik tidak otomatis menunjukkan hubungan sebab-akibat.

Pada aspek akademik, penggunaan media sosial yang bermasalah dapat menurunkan performa melalui distraksi, *technostress*, kelelahan, penundaan tugas, dan perpindahan waktu belajar. Homaid menunjukkan bahwa *problematic social media use* memengaruhi penurunan performa akademik melalui *technostress* dan *exhaustion* pada mahasiswa [4]. Hamam et. al. menemukan bahwa dimensi kecanduan media sosial berhubungan dengan setidaknya satu indikator performa mahasiswa, termasuk nilai, ketidakhadiran, dan kelulusan mata kuliah [5]. Walaupun demikian, hubungan antara durasi penggunaan dan performa tidak selalu linear. Tafesse menunjukkan adanya pola nonlinier berbentuk *inverted-U* antara penggunaan jejaring sosial dan performa akademik berdasarkan data penggunaan aplikasi seluler [6] , sedangkan Suárez-Perdomo et. al. menunjukkan bahwa profil penggunaan jejaring yang berlebihan lebih konsisten terkait dengan prokrastinasi dibanding perbedaan performa akademik secara langsung [7]. Hal ini menegaskan bahwa analisis kecanduan media sosial sebaiknya tidak bergantung pada satu indikator tunggal.

Dimensi psikologis dan sosial juga penting. Ghanayem et. al. menunjukkan keterkaitan *problematic social media use* dengan stres, kecemasan, depresi, dan kesepian pada mahasiswa [8] . Aslan dan Polat menemukan bahwa kecanduan media sosial, kesepian, depresi, kepuasan hidup, serta kemampuan pemecahan masalah mempunyai hubungan dengan *self-efficacy* dan keberhasilan akademik [9] . Pada tingkat platform, Hendrikse dan Limniou memperlihatkan bahwa penggunaan problematik Instagram dan TikTok berkaitan dengan aspek kesejahteraan, sehingga jenis platform dan pola keterlibatan pengguna dapat memengaruhi konteks risiko [10] . Pada remaja lintas negara, Lahti et. al. menunjukkan bahwa dukungan keluarga, dukungan teman, *health literacy*, gejala depresif, dan kesulitan tidur memiliki hubungan penting dengan *problematic social media use* [11]. Studi longitudinal Salerno et. al. juga memperlihatkan bahwa dukungan sosial dan kecenderungan perbandingan sosial dapat memprediksi lintasan penggunaan media sosial problematik pada remaja [12].

Perkembangan *machine learning* memberi peluang untuk mengintegrasikan banyak indikator tersebut secara simultan. Berbeda dengan analisis bivariat yang menilai satu hubungan pada satu waktu, model klasifikasi dapat belajar dari kombinasi usia, durasi penggunaan, platform, dampak akademik, jam tidur, skor kesehatan mental, status relasi, dan konflik akibat media sosial. Pendekatan prediktif semacam ini mulai digunakan dalam studi adiksi digital. Hong et. al. membangun model *machine learning* untuk memprediksi adiksi telepon seluler pada 3.000 mahasiswa dan memperoleh akurasi 76,68% dengan *Random Forest* sebagai model terbaik [13]. Thity et. al. membandingkan *Decision Tree*, *Random Forest*, SVM, dan Logistic Regression untuk klasifikasi multikelas *internet addiction* pada mahasiswa, serta menekankan pentingnya *confusion matrix* dan validasi lintas lipatan [14]. Studi longitudinal terbaru oleh Zarate et. al. secara khusus menunjukkan bahwa *Random Forest* dapat mengungguli beberapa model lain dalam memprediksi risiko *problematic social media use* dari indikator kesehatan mental, kepribadian, *coping*, dan adiksi lain [15] .

Meskipun penelitian terdahulu telah membahas kecanduan media sosial dan prediksi adiksi digital, perbandingan algoritma tetap perlu memperhatikan pemisahan informasi pelatihan dan pengujian. Identifier individu, transformasi yang dipelajari sebelum pemisahan data, dan dominasi kelas mayoritas dapat mengaburkan interpretasi performa. Penelitian ini membandingkan Decision Tree, Random Forest, dan SVM linear untuk klasifikasi kategori Addicted_Score pada dataset Students' Social Media Addiction. Kontribusi penelitian adalah perbandingan kinerja model dengan penghapusan identifier, preprocessing di dalam pipeline, pemilihan algoritma melalui validasi silang pada data latih, serta pelaporan confusion matrix dan metrik antarkelas. Analisis korelasi data latih dan feature importance Random Forest melengkapi interpretasi pola dataset, tanpa menempatkan hubungan prediktif sebagai bukti kausalitas atau diagnosis klinis.

Sampel penelitian berusia 18–24 tahun, sehingga istilah remaja pada judul dibatasi pada remaja akhir dan dewasa muda dalam dataset. Penelitian ini merupakan klasifikasi potong lintang terhadap kategori skor pada saat observasi, bukan prediksi prospektif timbulnya kecanduan. Hasilnya tidak digeneralisasikan kepada remaja awal, seluruh populasi mahasiswa, maupun prevalensi kecanduan klinis. Batasan tersebut dipertahankan dalam pembahasan hasil dan kesimpulan.

2. METODOLOGI PENELITIAN

2.1 *Dataset*

Penelitian menggunakan data sekunder Students' Social Media Addiction yang dirujuk melalui https://www.kaggle.com/datasets/adilshamim8/social-media-addiction-vs-relationships. Dataset yang dianalisis berisi 705 observasi dan 13 atribut, meliputi Student_ID, usia, gender, jenjang pendidikan, negara, durasi penggunaan, platform utama, dampak akademik, durasi tidur, skor kesehatan mental, status relasi, konflik akibat media sosial, dan Addicted_Score. Rata-rata usia data awal adalah 20,66 tahun dengan rentang 18–24 tahun; terdapat 353 perempuan dan 352 laki-laki. Pemeriksaan berkas tidak menemukan nilai hilang.

Copyright © 2026 The **Egi Safitri**, Page 1376
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

2.2 Pembersihan Data dan Penanganan Pencilan

Pemeriksaan terhadap 705 baris data menunjukkan bahwa setiap baris memiliki *Student_ID* yang berbeda. Meskipun demikian, terdapat dua baris dengan nilai yang sama pada seluruh sebelas variabel prediktor dan *Addicted_Score*. Untuk mencegah kedua baris tersebut masuk ke data latih dan data uji secara terpisah, hanya baris pertama yang dipertahankan sehingga jumlah data menjadi 704. Kesamaan nilai tersebut tidak dapat digunakan untuk memastikan bahwa keduanya berasal dari individu yang sama. Data tidak memerlukan imputasi karena tidak ditemukan nilai hilang. Penghapusan pencilan berdasarkan IQR juga tidak diterapkan. Durasi penggunaan media sosial tertinggi, yaitu 8,5 jam per hari, tetap disertakan karena nilai yang tinggi saja belum cukup menjadi dasar untuk menyatakan adanya kesalahan pencatatan.

Student_ID digunakan sebagai penanda untuk memeriksa pembagian data dan tidak disertakan sebagai prediktor. Sementara itu, *Addicted_Score* digunakan untuk membentuk kelas target, yaitu *Addicted_Category*, kemudian dikeluarkan dari fitur model agar skor pembentuk kelas tidak digunakan untuk memprediksi kelas itu sendiri. Dengan demikian, model menggunakan sebelas prediktor yang terdiri atas lima variabel numerik dan enam variabel kategorikal, sedangkan *Addicted_Category* digunakan sebagai target. Pemeriksaan akhir memastikan bahwa *Student_ID*, *Addicted_Score*, dan *Addicted_Category* tidak termasuk dalam prediktor serta tidak terdapat baris atau kombinasi nilai prediktor yang sama pada data latih dan data uji

2.3 Pembentukan Kategori Tingkat Kecanduan

Tingkat kecanduan dikelompokkan berdasarkan *Addicted_Score* menjadi tiga kategori, yaitu *low* untuk skor ≤5, *medium*untuk skor 6–8, dan *high* untuk skor >8. Batas kategori ditentukan dari eksplorasi kuartil dataset sebelum pemodelan dan tidak disesuaikan berdasarkan hasil pengujian. Karena dataset yang sama telah digunakan pada tahap eksplorasi, penilaian kinerja model terbatas pada validasi internal. Pengelompokan ini digunakan untuk keperluan analisis dan tidak dapat diartikan sebagai diagnosis klinis karena batas skornya belum mengacu pada instrumen diagnostik yang tervalidasi.Setelah satu baris dengan kombinasi nilai prediktor yang sama dikeluarkan, data terdiri atas 236 observasi kategori *low*(33,52%), 413 kategori *medium* (58,66%), dan 55 kategori *high* (7,81%). Jumlah observasi yang tidak seimbang, terutama pada kategori *high*, menjadi dasar penggunaan *macro precision*, *macro recall*, dan *macro F1-score* agar setiap kelas mendapat bobot yang sama dalam penilaian. Nilai *recall* setiap kelas juga dilaporkan untuk melihat kemampuan model mengenali masing-masing kategori. Penelitian ini tidak menerapkan *oversampling*, *undersampling*, maupun pembobotan kelas. Ketiga model dilatih dan diuji menggunakan pembagian data yang sama agar kinerjanya dapat dibandingkan pada sampel yang setara.

2.4 Transformasi Fitur dan Pembagian Data

Data dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan pembagian terstratifikasi dan *random_s tate* 42. Pembagian ini menghasilkan 563 observasi latih dan 141 observasi uji. Data latih mencakup 189 observasi kate gori *low*, 330 *medium*, dan 44 *high*, sedangkan data uji terdiri atas 47 observasi kategori *low*, 83 *medium*, dan 11 *high*. S tratifikasi digunakan untuk mempertahankan proporsi setiap kategori pada kedua kelompok data.Enam variabel kategori kal, yaitu *Gender*, *Academic_Level*, *Country*, *Most_Used_Platform*, *Affects_Academic_Performance*, dan *Relationship_ Status*, diubah menjadi variabel biner melalui *onehot encoding*. Sementara itu, lima variabel numerik meliputi *Age*, *Avg _Daily_Usage_Hours*, *Sleep_Hours_Per_Night*, *Mental_Health_Score*, dan *Conflicts_Over_Social_Media*. Pada Decisi on Tree dan Random Forest, variabel numerik digunakan dalam skala aslinya. Pada SVM, variabel tersebut distandardis asi menggunakan *StandardScaler*.

Transformasi fitur dan algoritma klasifikasi digabungkan dalam satu *pipeline* menggunakan *ColumnTransform er*. Kategori untuk pengodean dan parameter standardisasi ditentukan hanya dari data latih, kemudian diterapkan pada d ata validasi dan data uji. *OneHotEncoder* diatur dengan *handle_unknown="ignore"* dan *sparse_output=False* agar kat egori yang tidak ditemukan saat pelatihan tetap dapat diproses tanpa menyesuaikan kembali pengodean.Validasi silang dilakukan pada data latih menggunakan *StratifiedKFold* sebanyak lima lipatan, dengan pengacakan dan *random_state* 4 2.

2.5 Algoritma Klasifikasi

Penelitian ini membandingkan Decision Tree, Random Forest, dan Support Vector Machine (SVM). Decision Tree membentuk aturan klasifikasi dengan membagi data secara bertahap berdasarkan nilai fitur. Hasilnya berupa struktur pohon yang relatif mudah ditafsirkan [16]. *Random Forest* menggabungkan prediksi sejumlah pohon yang dibangun menggunakan sampel bootstrap dan pemilihan fitur secara acak pada setiap pemisahan. Penggabungan tersebut bertujuan mengurangi variansi dibandingkan penggunaan satu pohon keputusan [17]. Sementara itu, SVM menggunakan kernel linear untuk membentuk batas pemisah antarkelas dengan margin maksimum [18]. Ketiga algoritma dipilih untuk membandingkan pendekatan pohon tunggal, gabungan pohon, dan pemisahan linear. Penggunaan Random Forest dan SVM juga didukung oleh penelitian sebelumnya mengenai klasifikasi adiksi digital dan data perilaku [13], [14], [15] .

This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Analisis dilakukan menggunakan Python 3.12.14 dan scikit-learn 1.8.0. Decision Tree menggunakan kriteria *Gini* dengan *random_state* 42. Random Forest terdiri atas 100 pohon, menggunakan kriteria *Gini*, pengambilan sampel *bootstrap*, dan *random_state* 42. Jumlah fitur yang dipertimbangkan pada setiap pemisahan ditetapkan sebesar akar kuadrat dari jumlah fitur masukan (*max_features="sqrt"*). Kedalaman pohon pada kedua algoritma tidak dibatasi. SVM menggunakan kernel linear dengan parameter C sebesar 1,0. Seluruh model dijalankan tanpa pembobotan kelas dan tanpa penyetelan hiperparameter. Setelah validasi silang selesai, masing-masing model dilatih menggunakan seluruh data latih dan diuji pada data uji yang sama. Konfigurasi model, versi pustaka, dan keluaran analisis dicatat dalam *notebook* untuk mendukung keterulangan penelitian.

2.6 Metrik Evaluasi dan Analisis Interpretabilitas

Kinerja model dinilai menggunakan *accuracy*, *precision*, *recall*, dan *F1-score*. *Accuracy* menunjukkan proporsi prediksi yang benar, *precision* mengukur ketepatan prediksi suatu kelas, sedangkan *recall* menunjukkan kemampuan model mengenali anggota kelas tersebut. *F1-score* merupakan rata-rata harmonik *precision* dan *recall*. Hasilnya dirangkum menggunakan *macro average*, yang memberi bobot sama pada setiap kelas, dan *weighted average*, yang memperhitungkan jumlah observasi setiap kelas. Pola kesalahan klasifikasi diperiksa melalui *confusion matrix*. Algoritma dipilih berdasarkan rerata *macro F1-score* validasi silang lima lipatan pada data latih, dengan *macro recall* sebagai kriteria tambahan jika nilainya sama. Variasi antarlipatan dilaporkan sebagai simpangan baku sampel dengan pembagi n−1, bukan interval kepercayaan. Setelah pemilihan algoritma ditetapkan, seluruh model diuji pada data uji yang sama untuk menilai kinerja akhirnya.

Analisis korelasi Pearson dilakukan pada 563 observasi data latih untuk melihat hubungan antarvariabel numerik dan dampak akademik, yang dikodekan menjadi 0 untuk *No* dan 1 untuk *Yes*. Kontribusi fitur terhadap prediksi Random Forest dianalisis melalui *feature importance* setelah model dilatih menggunakan seluruh data latih. Nilai kepentingan variabel biner hasil *one-hot encoding* dijumlahkan berdasarkan variabel asalnya. Nilai tersebut menunjukkan kontribusi fitur terhadap penurunan *impurity* pada Random Forest, sehingga tidak dapat ditafsirkan sebagai hubungan sebab-akibat maupun digunakan untuk menjelaskan prediksi Decision Tree dan SVM.

Tabel 1. Statistik deskriptif sampel analisis setelah penanganan profil identik (n=704)

Variabel	Mean	SD	Minimum	Median	Maksimum
Age	20,66	1,40	18,00	21,00	24,00
Avg_Daily_Usage_Hours	4,92	1,26	1,50	4,80	8,50
Sleep_Hours_Per_Night	6,87	1,13	3,80	6,90	9,60
Mental_Health_Score	6,23	1,10	4,00	6,00	9,00
Conflicts_Over_Social_Media	2,85	0,96	0,00	3,00	5,00
Addicted_Score	6,44	1,59	2,00	7,00	9,00

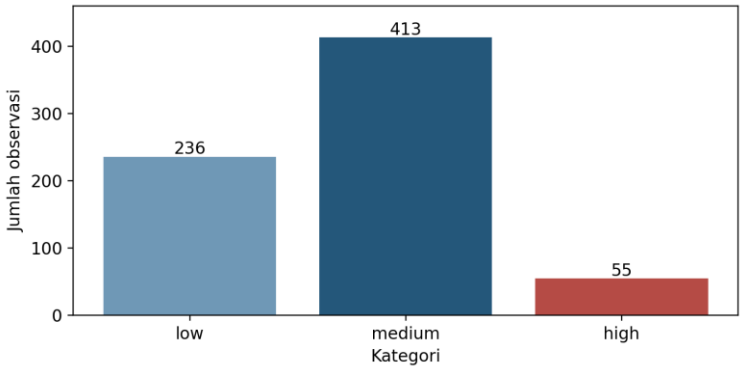
3. HASIL DAN PEMBAHASAN

3.1 Distribusi Tingkat Kecanduan

Distribusi akhir memperlihatkan dominasi kelas medium sebanyak 413 observasi (58,66%), diikuti low sebanyak 236 (33,52%) dan high sebanyak 55 (7,81%). Gambar 1 menampilkan distribusi tersebut dengan nama kelas eksplisit. Distribusi ini diperoleh setelah satu profil identik dikeluarkan, dengan tetap mempertahankan nilai ekstrem yang valid. Distribusi tersebut mendeskripsikan sampel, bukan prevalensi kecanduan pada populasi. Pembagian terstratifikasi mempertahankan representasi ketiga kelas pada data latih dan uji. Namun, kelas high hanya mempunyai 11 observasi dalam data uji, sehingga satu kesalahan tambahan akan mengubah recall kelas tersebut sekitar 9,09 poin persentase. Karena itu, hasil recall harus dibaca bersama dukungan sampel, macro F1-score, dan variasi validasi silang.

Copyright © 2026 The **Egi Safitri**, Page 1378
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



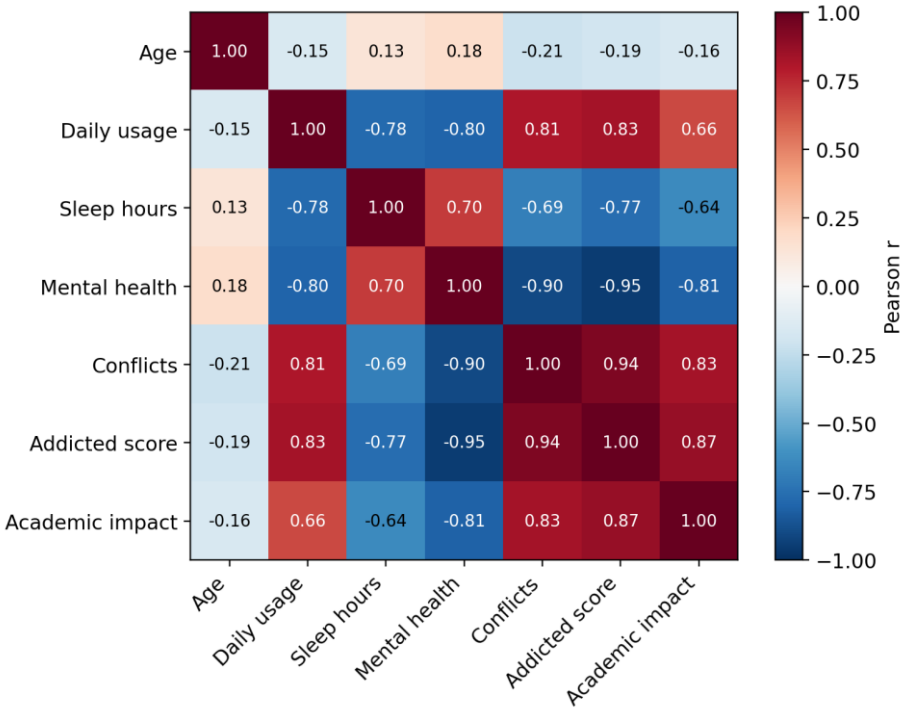
Gambar 1. Distribusi kelas low, medium, dan high pada 704 profil analisis.

3.2 Analisis Korelasi Antarvariabel

Gambar 2 memperlihatkan korelasi Pearson pada data latih, tanpa Student_ID atau kode angka untuk variabel nominal. Korelasi positif Addicted_Score adalah 0,935 dengan Conflicts_Over_Social_Media, 0,871 dengan Affects_Academic_Performance, dan 0,833 dengan Avg_Daily_Usage_Hours. Dengan demikian, pada sampel latih, skor lebih tinggi berkaitan dengan konflik lebih sering, laporan dampak akademik, dan durasi penggunaan lebih panjang. Pola tersebut bersifat asosiasi dan searah dengan pembahasan konsekuensi penggunaan problematik dalam literatur [4], [5], [9].

Korelasi negatif terkuat ditemukan antara Addicted_Score dan Mental_Health_Score (r=−0,947), diikuti Sleep_Hours_Per_Night (r=−0,766). Skor kecanduan yang lebih tinggi cenderung muncul bersama nilai kedua variabel yang lebih rendah. Arah skala Mental_Health_Score tidak boleh diinterpretasikan sebagai kondisi klinis tanpa dokumentasi instrumen. Hubungan durasi tidur juga tidak identik dengan kualitas tidur; penelitian terdahulu membahas aspek tidur dan penggunaan problematik dengan pengukuran yang berbeda [2], [3], [20].

Korelasi tinggi tidak membuktikan sebab-akibat atau menjamin kegunaan model pada populasi lain. Negara, gender, jenjang pendidikan, platform, dan status relasi tidak dimasukkan ke heatmap sebagai kode ordinal. Addicted_Score ditampilkan hanya untuk analisis asosiasi dan tetap tidak masuk matriks prediktor. Analisis ini tidak digunakan untuk memilih atau menghapus fitur berdasarkan informasi data uji.



Gambar 2. Korelasi Pearson variabel numerik dan biner pada data latih (n=563).

Hubungan yang mendekati satu antara skor kecanduan dan beberapa prediktor menuntut kehati-hatian terhadap kemungkinan tumpang tindih konstruk. Pipeline mencegah informasi statistik data uji masuk ke pelatihan, tetapi tidak

Copyright © 2026 The **Egi Safitri**, Page 1379
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

membuktikan bahwa skor kecanduan dibentuk secara independen dari prediktor. Dokumentasi konstruksi Addicted_Score diperlukan sebelum menyatakan seluruh potensi target leakage telah disingkirkan.

3.3 Perbandingan Kinerja Tiga Model

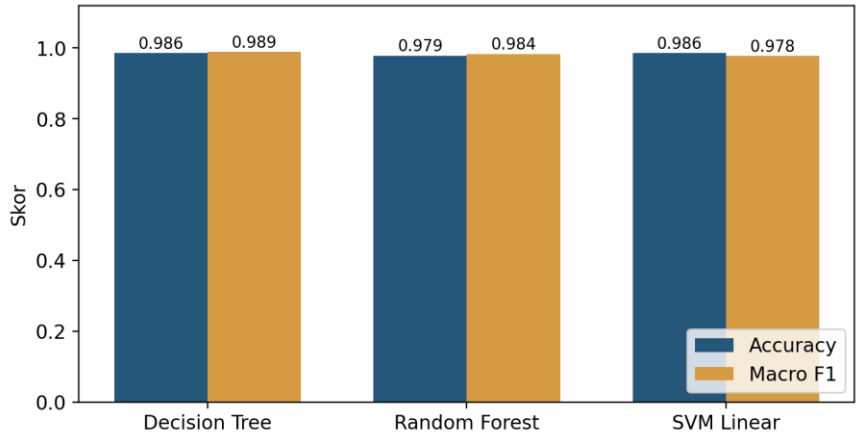
Validasi silang data latih menghasilkan macro F1-score 0,9334 ± 0,0513 untuk Decision Tree, 0,9708 ± 0,0292 untuk Random Forest, dan 0,9701 ± 0,0165 untuk SVM linear. Random Forest dipilih berdasarkan rerata tertinggi sebelum hasil data uji digunakan. Selisih Random Forest terhadap SVM hanya sekitar 0,0007 dan jauh lebih kecil daripada simpangan baku antarlipatan. Karena tidak dilakukan uji perbedaan performa, hasil ini tidak membuktikan superioritas statistik Random Forest; SVM merupakan kandidat dengan rerata yang hampir sama dan variasi antarlipatan lebih kecil.

Tabel 2 menyajikan hasil aktual pada 141 observasi uji setelah semua model dilatih tanpa identifier. Decision Tree dan SVM mencapai accuracy 98,58%, sedangkan Random Forest mencapai 97,87%. Macro F1-score test tertinggi justru diperoleh Decision Tree (0,9889), diikuti Random Forest (0,9836) dan SVM (0,9779). Dengan demikian, model terpilih berdasarkan validasi data latih berbeda dari model dengan macro F1-score tertinggi pada satu holdout. Pilihan algoritma tidak diubah setelah melihat test set agar data uji tetap berfungsi sebagai evaluasi akhir, bukan sarana optimasi.

Tabel 2. Performa ketiga model pada data uji (n=141)

Model	Accuracy	Macro Precision	Macro Recall	Macro F1	Weighted F1	Benar/Salah
Decision Tree	0,9858	0,9889	0,9889	0,9889	0,9858	139/2
Random Forest	0,9787	0,9800	0,9880	0,9836	0,9789	138/3
SVM Linear	0,9858	0,9653	0,9920	0,9779	0,9859	139/2

Gambar 3 membandingkan accuracy dan macro F1-score data uji. Semua model memiliki recall high sebesar 1,0000 pada 11 observasi, sehingga metrik tersebut tidak membedakan kinerja ketiga model. Perbedaan macro F1-score terutama berasal dari pola kesalahan low–medium dan satu false positive high pada SVM.



Gambar 3. Accuracy dan macro F1-score pada data uji.

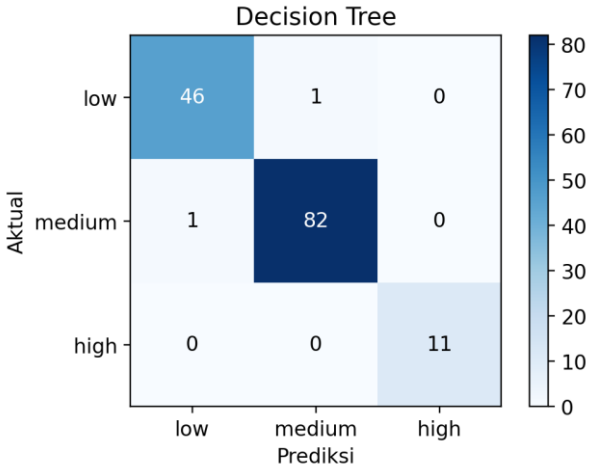
3.4 Analisis *Confusion Matrix Decision Tree*

Gambar 4 menunjukkan 139 prediksi benar dan dua prediksi salah pada Decision Tree. Dari 47 low, sebanyak 46 dikenali benar dan satu diprediksi medium; dari 83 medium, sebanyak 82 benar dan satu diprediksi low. Seluruh 11 high dikenali dengan benar. Precision dan recall high masing-masing 1,0000. Pola kesalahan ini menghasilkan accuracy 0,9858 dan macro F1-score 0,9889, yang merupakan nilai macro F1 tertinggi pada holdout penelitian ini.

Meskipun hasil test Decision Tree tinggi, rerata macro F1-score validasi silangnya lebih rendah dan simpangan bakunya lebih besar daripada dua kandidat lain. Ketidaksesuaian peringkat tersebut memperlihatkan bahwa satu holdout tidak cukup untuk menyatakan keunggulan yang stabil. Model pohon tunggal tetap mempunyai nilai interpretabilitas, tetapi penetapan model akhir mengikuti evaluasi pada data latih yang telah ditetapkan sebelumnya.

Copyright © 2026 The **Egi Safitri**, Page 1380
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

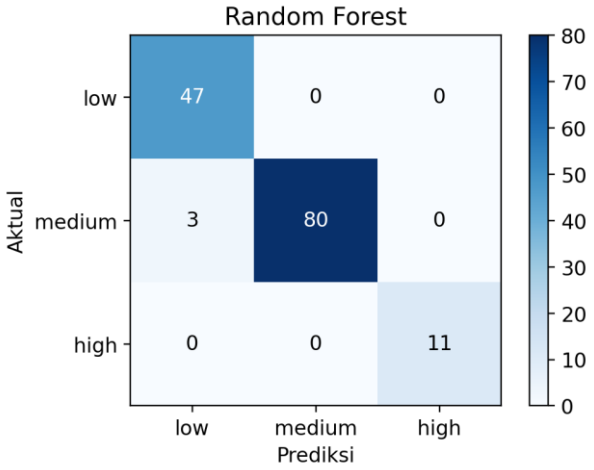


Gambar 4. Confusion matrix Decision Tree

3.5 Analisis *Confusion Matrix Random Forest*

Gambar 5 menunjukkan 138 prediksi benar dan tiga prediksi salah pada Random Forest. Seluruh 47 low dan 11 high dikenali dengan benar. Dari 83 medium, sebanyak 80 dikenali benar dan tiga diprediksi low, sehingga recall medium sebesar 0,9639. Precision low sebesar 0,9400 mencerminkan tiga prediksi low yang sebenarnya medium. Secara keseluruhan, accuracy mencapai 0,9787, macro recall 0,9880, dan macro F1-score 0,9836.

Random Forest dipertahankan sebagai algoritma terpilih karena rerata macro F1-score validasi data latih tertinggi, bukan karena memimpin seluruh metrik test. Literatur juga menunjukkan Random Forest sebagai kandidat yang kompetitif pada adiksi digital [13]–[15], tetapi hasil tersebut tidak dapat dipakai untuk mengabaikan keunggulan Decision Tree pada holdout saat ini. Perbandingan antarpublikasi perlu mempertimbangkan perbedaan target, sampel, instrumen, dan prosedur validasi.



Gambar 5. Confusion matrix Random Forest

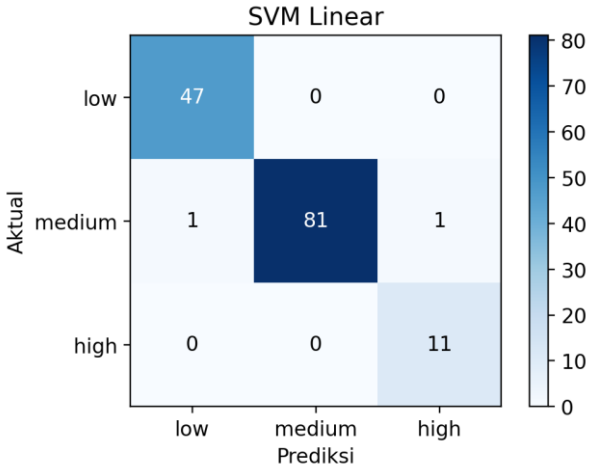
3.6 Analisis *Confusion Matrix* SVM

Gambar 6 menunjukkan 139 prediksi benar dan dua prediksi salah pada SVM linear. Seluruh 47 low dan 11 high dikenali dengan benar. Dari 83 medium, sebanyak 81 dikenali benar, satu diprediksi low, dan satu diprediksi high. Akibat satu false positive high tersebut, precision high sebesar 0,9167 walaupun recall high mencapai 1,0000. Accuracy SVM sebesar 0,9858 sama dengan Decision Tree, tetapi macro F1-score-nya lebih rendah, yaitu 0,9779.

SVM mempunyai macro recall test tertinggi sebesar 0,9920, tetapi accuracy dan macro recall tidak sepenuhnya menggambarkan precision setiap kelas. Kesamaan jumlah prediksi benar dengan Decision Tree menghasilkan profil kesalahan yang berbeda. Dalam penggunaan sebagai skrining, konsekuensi false positive dan false negative perlu dinilai terpisah, dan keputusan klinis tidak dapat ditentukan dari klasifikasi internal dataset ini.

Copyright © 2026 The **Egi Safitri**, Page 1381
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 6. Confusion matrix SVM linear

3.7 *Feature Importance Random Forest*

Tabel 3 menyajikan feature importance Random Forest yang dihitung pada data latih, dengan kontribusi one-hot encoding dijumlahkan menurut fitur asal. Affects_Academic_Performance memiliki importance tertinggi sebesar 0,3236, diikuti Country 0,1460, Conflicts_Over_Social_Media 0,1395, Mental_Health_Score 0,1221, Avg_Daily_Usage_Hours 0,0949, dan Sleep_Hours_Per_Night 0,0706. Student_ID maupun Addicted_Score tidak terdapat dalam model atau daftar importance.

Besarnya kontribusi Country harus ditafsirkan hati-hati. Fitur ini memiliki banyak kategori, dan penjumlahan importance sejumlah dummy dapat memberikan kontribusi agregat besar tanpa menunjukkan mekanisme risiko yang bermakna. Negara juga mungkin menangkap pola komposisi atau konstruksi dataset. Karena itu, hasil tidak digunakan untuk menyimpulkan negara tertentu lebih berisiko mengalami kecanduan. Validasi pada populasi independen diperlukan untuk menilai apakah kontribusi geografis dapat digeneralisasikan.

Impurity-based importance bergantung pada model, distribusi data, dan hubungan antarprediktor. Prediktor yang berkorelasi dapat saling berbagi kontribusi, sehingga peringkat ini bukan ukuran efek kausal atau kepentingan klinis. Interpretasi tersebut hanya berlaku untuk Random Forest, tidak menjelaskan keputusan SVM atau Decision Tree. Analisis permutation importance dan kestabilan peringkat pada sampel eksternal merupakan pengembangan yang relevan.

Tabel 3. Feature importance Random Forest hasil pelatihan, diagregasi per fitur asal

Peringkat	Fitur	Importance
1	*Affects_Academic_Performance*	0,3236
2	*Country*	0,1460
3	*Conflicts_Over_Social_Media*	0,1395
4	*Mental_Health_Score*	0,1221
5	*Avg_Daily_Usage_Hours*	0,0949
6	*Sleep_Hours_Per_Night*	0,0706
7	*Most_Used_Platform*	0,0448
8	*Academic_Level*	0,0171
9	*Age*	0,0157
10	*Relationship_Status*	0,0150
11	*Gender*	0,0107

3.8 Pembahasan terhadap Penelitian Terdahulu

Analisis data latih menunjukkan bahwa skor kecanduan berkaitan dengan aspek tidur, akademik, konflik sosial, dan skor kesehatan mental. Hubungan negatif durasi tidur konsisten secara arah dengan pembahasan Zhu et al. [2] dan Zhuang et al. [3], tetapi kesamaan arah asosiasi tidak membuktikan kesetaraan instrumen atau efek. Dataset hanya mencatat durasi tidur, sedangkan kualitas tidur melibatkan dimensi yang lebih luas. Karena penelitian bersifat potong lintang, urutan sebab-akibat antara penggunaan media sosial dan tidur juga tidak dapat ditetapkan.

Korelasi dampak akademik biner dengan Addicted_Score sebesar 0,871 searah dengan literatur tentang konsekuensi akademik penggunaan problematik [4], [5], [9]. Namun, Affects_Academic_Performance adalah laporan

Copyright © 2026 The **Egi Safitri**, Page 1382
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10139
Hal 1375-1384
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

biner dalam dataset dan bukan nilai akademik objektif seperti GPA, absensi, atau kelulusan. Oleh sebab itu, hasil tidak boleh diubah menjadi klaim bahwa peningkatan skor kecanduan menyebabkan penurunan nilai akademik dengan besaran tertentu.

Korelasi Mental_Health_Score sebesar −0,947 pada data latih menunjukkan keterkaitan yang sangat kuat. Literatur mendiskusikan hubungan penggunaan problematik dengan indikator kesehatan mental [1], [8], [10] , tetapi besaran korelasi dalam dataset ini tidak dapat langsung disetarakan dengan efek pada populasi nyata. Definisi, arah skala, dan konstruksi skor belum dapat diverifikasi dari dataset. Ketergantungan yang tinggi juga dapat mencerminkan variabel yang mengukur aspek serupa, sehingga akurasi klasifikasi perlu dibaca sebagai kemampuan mengenali kategori skor pada dataset.

Dari sisi algoritma, Random Forest mempunyai rerata macro F1 validasi silang tertinggi, sedangkan SVM mempunyai rerata yang hampir sama dengan variasi lebih kecil. Temuan ini mendukung penggunaan keduanya sebagai kandidat pada data tabular perilaku, sejalan dengan kajian adiksi digital [13], [14], [15] . Akan tetapi, Decision Tree memberikan macro F1 test tertinggi pada holdout yang digunakan. Penelitian ini membedakan kriteria seleksi dan pelaporan hasil test sehingga kesimpulan tidak memilih metrik secara selektif untuk mempertahankan pemenang tertentu.

Kinerja model diperoleh melalui rangkaian prosedur yang mencakup penghapusan Student_ID, penanganan profil identik, retensi nilai ekstrem, stratifikasi, encoding kategorikal, preprocessing di dalam pipeline, dan pemilihan algoritma melalui validasi silang. Penelitian ini tidak melakukan studi ablation yang mengubah satu komponen pada satu waktu. Oleh karena itu, kontribusi terpisah dari setiap komponen terhadap kinerja model tidak dapat ditentukan, dan tingginya metrik tidak dapat dikaitkan hanya dengan penghapusan Student_ID.

Dari perspektif penerapan, hasil menunjukkan kemungkinan membangun alat bantu pengelompokan kategori skor, tetapi belum membuktikan manfaat skrining klinis atau kesiapan implementasi. Prediktor harus tersedia pada saat penggunaan, dan konstruksi Addicted_Score harus diverifikasi agar tidak sekadar merekonstruksi komponen yang sama. Penerapan nyata memerlukan instrumen tervalidasi, penilaian profesional, perlindungan data, dan evaluasi pada populasi baru. Model tidak digunakan untuk diagnosis atau keputusan intervensi otomatis.

Untuk kelompok remaja dan dewasa muda, pendekatan preventif sebaiknya tidak berfokus pada larangan penggunaan media sosial secara umum. Studi lintas negara menunjukkan peran dukungan keluarga, dukungan teman, dan literasi kesehatan sebagai faktor protektif [11], sedangkan studi longitudinal menunjukkan pentingnya dukungan sosial dan *social comparison* dalam lintasan penggunaan problematik [12]. Karena itu, keluaran model lebih bermanfaat jika diintegrasikan dengan edukasi literasi digital, pengelolaan waktu, higiene tidur, dukungan psikologis, dan strategi membangun relasi sosial yang sehat.

3.9 Keterbatasan Penelitian dan Validitas Model

Penelitian ini menggunakan satu dataset sekunder dengan informasi pengambilan sampel dan instrumen yang terbatas, sehingga hasilnya belum dapat digeneralisasikan pada populasi tertentu. Batas kategori *low*, *medium*, dan *high* ditentukan melalui eksplorasi dataset dan bukan merupakan batas diagnosis klinis. Kategori *high* juga hanya mencakup 11 observasi pada data uji. Oleh karena itu, keberhasilan mengenali seluruh observasi tersebut belum menjamin kemampuan yang sama pada data lain. Meskipun pemisahan data dan prapemrosesan telah dilakukan melalui *pipeline*, kemungkinan kebocoran target akibat keterkaitan antara komponen pembentuk skor dan prediktor masih memerlukan pemeriksaan berdasarkan dokumentasi instrumen.

Selisih rerata *macro F1-score* Random Forest dan SVM sangat kecil, sementara pengujian signifikansi belum dilakukan. Pemilihan Random Forest karena itu tidak menunjukkan keunggulan statistik atas model lain. Penelitian ini juga belum mencakup penyetelan hiperparameter, penilaian kalibrasi, maupun pengujian manfaat klinis. Analisis *feature importance*terbatas pada Random Forest dan dapat dipengaruhi oleh banyaknya kategori serta korelasi antarfitur. Pengujian pada dataset independen dan verifikasi pembentukan skor diperlukan untuk menilai kemampuan generalisasi model.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa Decision Tree, Random Forest, dan SVM mampu mengklasifikasikan kategori tingkat kecanduan media sosial pada 704 observasi berusia 18–24 tahun. Pemodelan dilakukan tanpa menyertakan *Student_ID*dan *Addicted_Score* sebagai prediktor, dengan pengodean serta standardisasi fitur yang hanya dipelajari dari data latih. Berdasarkan validasi silang lima lipatan, Random Forest dipilih karena memperoleh rerata *macro F1-score* tertinggi sebesar 0,9708 ± 0,0292. Namun, nilainya sangat dekat dengan SVM sebesar 0,9701 ± 0,0165, sehingga belum menunjukkan keunggulan yang signifikan secara statistik. Pada 141 observasi data uji, Random Forest mencapai akurasi 97,87% dan *macro F1-score* 0,9836. Decision Tree memperoleh *macro F1-score* tertinggi pada data uji, yaitu 0,9889 dengan akurasi 98,58%, sedangkan SVM menghasilkan akurasi 98,58% dan *macro F1-score* 0,9779.

Copyright © 2026 The **Egi Safitri**, Page 1383
This Journal is licensed under a Creative Commons Attribution 4.0 International License

Perbedaan peringkat tersebut menunjukkan bahwa kinerja pada satu pembagian data uji tidak selalu sejalan dengan hasil validasi silang yang menjadi dasar pemilihan model.

Ketiga model berhasil mengenali seluruh 11 observasi kategori *high* pada data uji, tetapi jumlah tersebut masih terbatas untuk menilai kemampuan model pada populasi yang lebih luas. Analisis *feature importance* Random Forest menunjukkan bahwa dampak akademik memberikan kontribusi terbesar terhadap prediksi. Kontribusi negara dan variabel lain yang berkaitan erat dengan skor kecanduan tetap memerlukan penelaahan lebih lanjut karena dapat mencerminkan pola khusus dalam dataset. Temuan penelitian ini mendukung penggunaan model untuk mengelompokkan kategori skor pada sampel yang dianalisis, tetapi belum cukup untuk menetapkan diagnosis klinis atau memperkirakan prevalensi kecanduan. Penelitian selanjutnya perlu memverifikasi cara penyusunan skor dan menguji model pada dataset independen sebelum digunakan sebagai alat skrining.

REFERENCES

[1] H. Shannon, K. Bush, P. J. Villeneuve, K. G. C. Hellemans, and S. Guimond, "Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-analysis," 2022. doi: 10.2196/33450.

[2] X. Zhu, T. Zheng, L. Ding, X. Zhang, Z. Li, and H. Jiang, "Exploring associations between social media addiction, social media fatigue, fear of missing out and sleep quality among university students: A cross-section study," *PLoS One*, vol. 18, no. 10 October, 2023, doi: 10.1371/journal.pone.0292429.

[3] J. Zhuang *et al.*, "A serial mediation model of social media addiction and college students' academic engagement: the role of sleep quality and fatigue," *BMC Psychiatry*, vol. 23, no. 1, 2023, doi: 10.1186/s12888-023-04799-5.

[4] A. Ali Homaid, "Problematic social media use and associated consequences on academic performance decrement during Covid-19," *Addictive Behaviors*, vol. 132, 2022, doi: 10.1016/j.addbeh.2022.107370.

[5] B. Hamam, S. Khandaqji, S. Sakr, and A. Ghaddar, "Social media addiction in university students in Lebanon and its effect on student performance," *Journal of American College Health*, vol. 72, no. 8, 2024, doi: 10.1080/07448481.2022.2152690.

[6] W. Tafesse, "Social networking sites use and college students' academic performance: testing for an inverted U-shaped relationship using automated mobile app usage data," *International Journal of Educational Technology in Higher Education*, vol. 19, no. 1, 2022, doi: 10.1186/s41239-022-00322-0.

[7] A. Suárez-Perdomo, Z. Ruiz-Delgado, and Y. Garcés-Delgado, "Profiles of undergraduates' networks addiction: Difference in academic procrastination and performance," *Comput. Educ.*, vol. 181, 2022, doi: 10.1016/j.compedu.2022.104459.

[8] L. K. Ghanayem, H. Shannon, L. Khodr, R. J. McQuaid, and K. G. C. Hellemans, "Lonely and scrolling during the COVID-19 pandemic: understanding the problematic social media use and mental health link among university students," *Front. Psychiatry*, vol. 15, 2024, doi: 10.3389/fpsyt.2024.1247807.

[9] I. Aslan and H. Polat, "Investigating social media addiction and impact of social media addiction, loneliness, depression, life satisfaction and problem-solving skills on academic self-efficacy and academic success among university students," *Front. Public Health*, vol. 12, 2024, doi: 10.3389/fpubh.2024.1359691.

[10] C. Hendrikse and M. Limniou, "The Use of Instagram and TikTok in Relation to Problematic Use and Well-Being," *J. Technol. Behav. Sci.*, vol. 9, no. 4, 2024, doi: 10.1007/s41347-024-00399-6.

[11] H. Lahti *et al.*, "What Counteracts Problematic Social Media Use in Adolescence? A Cross-National Observational Study," *Journal of Adolescent Health*, vol. 74, no. 1, 2024, doi: 10.1016/j.jadohealth.2023.07.026.

[12] L. Salerno, L. Fortunato, V. Ostwald, A. Muscolino, and G. Lo Coco, "Social support and social comparison tendencies predict trajectories of adolescents' problematic social media use: A longitudinal study," *PLoS One*, vol. 20, no. 6 June, 2025, doi: 10.1371/journal.pone.0323320.

[13] Y. Hong, X. Rong, and W. Liu, "Construction of influencing factor segmentation and intelligent prediction model of college students' cell phone addiction model based on machine learning algorithm," *Heliyon*, vol. 10, no. 8, 2024, doi: 10.1016/j.heliyon.2024.e29245.

[14] N. T. Thity, A. Rahman, A. Dulmini, N. Yasmin, and R. Rois, "An illustration of multi-class roc analysis for predicting internet addiction among university students," *PLoS One*, vol. 20, no. 7 July, 2025, doi: 10.1371/journal.pone.0325855.

[15] D. Zarate, J. Jarrad, V. Stavropoulos, C. Conkey-Morrison, and B. Hunt, "Mental health, personality, and cross-addictions as predictors of social media addiction: a machine learning longitudinal study," *Addictive Behaviors Reports*, vol. 24, 2026, doi: 10.1016/j.abrep.2026.100718.

[16] J. R. Quinlan, "Induction of Decision Trees," *Mach. Learn.*, vol. 1, no. 1, 1986, doi: 10.1023/A:1022643204877.

[17] L. Breiman and A. Cutler, "Breiman and Cutler's random forests for classification and regression," *Package "randomForest,"* 2012.

[18] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, vol. 20, no. 3, 1995, doi: 10.1023/A:1022627411411.

Copyright © 2026 The **Egi Safitri**, Page 1384
This Journal is licensed under a Creative Commons Attribution 4.0 International License