



Klasifikasi kecanduan smartphone mahasiswa universitas esa unggul menggunakan machine learning dan sas-sv

Anggoro Verrel^{*1}, Irfan Zidny Maulana², Vico Andrian Liu³, Ary Prabowo⁴

Email: 1anggoroverrel@student.esaunggul.ac.id, 2irfanzidny2611@student.esaunggul.ac.id,
3vicoaja45@student.esaunggul.ac.id, ary.prabowo@esaunggul.ac.id

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul

⁴Fakultas Ilmu Komputer, Universitas Esa Unggul

Diterima: 26 Juli 2025 | Direvisi: 04 Agustus 2025 | Disetujui: 24 Desember 2025

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Era digital telah menjadikan *smartphone* sebagai bagian tak terpisahkan dari kehidupan mahasiswa, namun juga memunculkan risiko kecanduan yang berdampak negatif pada prestasi akademik dan kesehatan mental. Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi model *machine learning* yang mampu mengklasifikasikan tingkat kecanduan *smartphone* pada mahasiswa Universitas Esa Unggul. Data dikumpulkan dari 32 responden mahasiswa melalui kuesioner daring yang mengadopsi instrumen psikometrik tervalidasi, *Smartphone Addiction Scale-Short Version* (SAS-SV). Tingkat kecanduan dikategorikan menjadi dua kelas: 'Tinggi', yang mengacu pada ambang batas risiko kecanduan spesifik gender dari Kwon et al dan 'Sedang', yang mencakup skor di bawah ambang batas tersebut. Empat algoritma klasifikasi—Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, dan Random Forest—diimplementasikan untuk membandingkan kinerjanya. Untuk mengatasi ketidakseimbangan kelas pada data, teknik *oversampling* SMOTE diterapkan pada data latih. Evaluasi model didasarkan pada metrik akurasi, presisi, *recall*, dan F1-score. Hasil penelitian menunjukkan bahwa model Logistic Regression mencapai kinerja terbaik dengan akurasi 1.0000 dan F1-score rata-rata 1.00 pada data uji. Meskipun demikian, perlu dicatat bahwa performa sempurna ini diperoleh dari ukuran data uji yang sangat terbatas (8 sampel), sehingga generalisasi memerlukan validasi lebih lanjut. Analisis tingkat kepentingan fitur dari model Random Forest mengidentifikasi bahwa pertanyaan terkait Pekerjaan/tugas yang saya rencanakan sering terganggu karena penggunaan *smartphone* (Q0) merupakan prediktor paling dominan. Hasil ini mengindikasikan bahwa model *machine learning* berbasis skala psikometrik memiliki potensi awal sebagai alat bantu skrining dan eksplorasi untuk mengidentifikasi mahasiswa yang berisiko mengalami kecanduan *smartphone*, namun memerlukan pengembangan dan validasi ekstensif pada data yang lebih besar sebelum implementasi praktis.

Kata kunci: Kecanduan Smartphone, Klasifikasi, Machine Learning, Mahasiswa, Random Forest, SAS-SV.

Classification of smartphone addiction of esa unggul university students using machine learning and sas-sv

Abstract

The digital era has made smartphones an inseparable part of students' lives, but it also raises the risk of addiction that negatively impacts academic achievement and mental health. This research aims to develop and evaluate machine learning models capable of classifying the level of smartphone addiction among Esa Unggul University students. Data were collected from 32 student respondents through an online questionnaire adopting the validated psychometric instrument, the *Smartphone Addiction Scale-Short Version* (SAS-SV). Addiction levels were categorized into two classes: 'High', which refers to the gender-specific addiction risk threshold from Kwon et al. (2013), and 'Moderate', which includes scores below that threshold. Four classification algorithms—Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, and Random Forest—were implemented to compare their performance. To address class imbalance in the data, the SMOTE oversampling technique was applied to the training data. Model evaluation was based on accuracy, precision, recall, and F1-score. The research results show that the Logistic Regression model achieved the best performance with an accuracy of 1.0000 and an average F1-score of 1.00 on the test data. Nevertheless, it should be noted that this perfect performance was obtained from a very limited test data size (8

samples), so generalization requires further validation. Feature importance analysis from the Random Forest model identified that the question related to Planned tasks/work often interrupted by smartphone use (Q0) was the most dominant predictor. These results indicate that machine learning models based on psychometric scales have initial potential as a screening and exploratory tool to identify students at risk of smartphone addiction, but require extensive development and validation on larger data before practical implementation.

Keywords: *Classification, Machine Learning, Random Forest, SAS-SV, Smartphone Addiction, University Students.*

1. PENDAHULUAN

Perkembangan pesat teknologi informasi dan komunikasi telah membawa transformasi fundamental dalam berbagai aspek kehidupan, salah satunya ditandai dengan meluasnya pemanfaatan smartphone. Bagi kalangan mahasiswa, smartphone tidak lagi sekadar alat komunikasi dasar, melainkan telah berevolusi menjadi perangkat multifungsi esensial untuk mengakses informasi, materi perkuliahan, platform media sosial, serta hiburan.

Namun, intensitas penggunaan yang tinggi ini turut memunculkan fenomena baru yang memerlukan perhatian serius, yakni kecanduan smartphone. Kecanduan smartphone dapat diartikan sebagai pola penggunaan perangkat yang berlebihan dan bersifat kompulsif, yang pada akhirnya menimbulkan dampak merugikan pada berbagai aspek kehidupan, termasuk penurunan kinerja akademik, gangguan pola tidur, masalah kesehatan fisik seperti nyeri leher, dan penurunan interaksi sosial tatap muka [1]. Mahasiswa merupakan salah satu kelompok populasi yang sangat rentan terhadap kecanduan smartphone, didorong oleh tuntutan akademik, tekanan sosial, dan kebutuhan akan konektivitas yang berkelanjutan [2]. Penelitian-penelitian sebelumnya juga telah mengonfirmasi hubungan signifikan antara kecanduan smartphone dan perilaku prokrastinasi akademik pada siswa/mahasiswa [2], [3].

Selama ini, deteksi tingkat kecanduan umumnya masih bergantung pada observasi klinis atau asesmen psikologis konvensional yang cenderung manual dan memakan waktu. Di sisi lain, pemanfaatan teknologi seperti *machine learning* (ML) telah banyak dieksplorasi dalam memprediksi berbagai fenomena kompleks [4], termasuk dalam deteksi serangan jaringan [5] dan prediksi penyakit [6], [7], bahkan pada aplikasi biomedis dan kesehatan [6]. Potensi ML dalam menyediakan sistem yang lebih objektif, scalable, dan efisien dalam mendeteksi pola perilaku kompleks, seperti kecanduan, menjadi sangat relevan. Oleh karena itu, terdapat celah penelitian yang signifikan untuk mengembangkan sistem deteksi kecanduan smartphone berbasis ML yang terukur.

Berangkat dari latar belakang tersebut, penelitian ini berfokus pada pembangunan model klasifikasi menggunakan algoritma *machine learning* untuk memprediksi tingkat kecanduan smartphone pada mahasiswa Universitas Esa Unggul. Studi ini memanfaatkan data yang dikumpulkan melalui instrumen psikometrik yang telah teruji validitasnya, yaitu *Smartphone Addiction Scale-Short Version* (SAS-SV) [1]. Kinerja dari empat algoritma populer—Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, dan Random Forest [8] akan dievaluasi dan dibandingkan guna menentukan model yang menunjukkan performa terbaik. Kontribusi utama dari penelitian ini adalah menyediakan model dasar yang berpotensi dikembangkan lebih lanjut menjadi aplikasi sistem deteksi dini bagi institusi pendidikan, sekaligus menjadi studi eksplorasi terhadap penerapan ML dalam konteks spesifik mahasiswa Esa Unggul, dengan mengakui keterbatasan data yang ada.

2. TINJAUAN PUSTAKA

2.1. Kecanduan Smartphone

Kecanduan smartphone, atau *Smartphone Addiction* (SA), diartikan sebagai pola perilaku penggunaan smartphone yang berlebihan dan tidak terkontrol, yang pada akhirnya membawa dampak negatif signifikan pada kehidupan individu [1]. Konsep ini sering kali dihubungkan dengan kriteria adiksi perilaku lainnya, seperti *internet gaming disorder*, mencakup gejala seperti *preoccupation*, toleransi, *withdrawal*, *relapse*, serta efek negatif terhadap fungsi sosial, pekerjaan, atau akademik [1]. Beberapa studi menunjukkan bahwa mahasiswa cenderung lebih rentan terhadap kecanduan smartphone akibat tekanan akademik, kebutuhan sosial, dan kemudahan akses perangkat tersebut [2]. Dampak negatif ini meliputi penurunan kinerja akademik, gangguan pola tidur, masalah kesehatan fisik seperti nyeri leher, dan penurunan interaksi sosial tatap muka [1], [2]. Penelitian-penelitian sebelumnya juga telah mengonfirmasi hubungan signifikan antara kecanduan smartphone dan perilaku prokrastinasi akademik pada siswa/mahasiswa [2], [3].

2.2. *Smartphone Addiction Scale - Short Version* (SAS-SV)

Smartphone Addiction Scale - Short Version (SAS-SV) merupakan instrumen pengukuran mandiri yang dikembangkan oleh Kwon et al. [1] untuk mengidentifikasi potensi risiko kecanduan smartphone. Skala ini merupakan versi singkat dari *Smartphone*

Addiction Scale (SAS) yang asli dan terdiri dari 10 item (pertanyaan). Pertanyaan-pertanyaan tersebut mencakup enam dimensi kunci yang diidentifikasi dalam penelitian utamanya: (1) gangguan dalam kehidupan sehari-hari (*daily-life disturbance*), (2) gejala putus (*withdrawal*), (3) toleransi, (4) hubungan berorientasi dunia maya (*cyberspace-oriented relationship*), (5) penggunaan berlebihan (*overuse*), dan (6) antisipasi positif (*positive anticipation*) [1]. Setiap item dinilai menggunakan skala Likert 6 poin, mulai dari 1 (sangat tidak setuju) hingga 6 (sangat setuju) [2]. Dengan demikian, skor total SAS-SV berkisar antara 10 hingga 60. Dalam penelitian ini, kategorisasi tingkat kecanduan smartphone didasarkan pada skor total SAS-SV. Kategori 'Tinggi' (berisiko kecanduan) ditentukan berdasarkan *cut-off point* spesifik gender yang divalidasi oleh Kwon et al., yaitu skor total ≥ 31 untuk laki-laki dan ≥ 33 untuk perempuan [1], [2]. Sementara itu, responden dengan skor di bawah ambang batas tersebut dikelompokkan ke dalam kategori 'Sedang'.

2.3. Machine Learning untuk Klasifikasi

Machine learning (ML) merupakan cabang dari kecerdasan buatan (AI) yang memungkinkan sistem untuk belajar dari data dan membuat keputusan atau prediksi tanpa diprogram secara eksplisit [4]. Dalam konteks klasifikasi, algoritma ML dilatih menggunakan *dataset* yang telah diberi label untuk mengenali pola-pola tertentu, kemudian memprediksi kategori untuk data baru yang belum pernah diamati. Penerapan ML untuk tugas klasifikasi telah ditunjukkan dalam berbagai konteks, seperti klasifikasi serangan siber [5], prediksi kondisi medis [6], hingga analisis citra medis untuk deteksi penyakit seperti COVID-19 [7]. Berbagai algoritma klasifikasi ML menawarkan pendekatan yang berbeda dalam memecahkan masalah ini. Dalam studi ini, empat algoritma populer diimplementasikan dan dibandingkan: Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, dan Random Forest.

2.3.1. Logistic Regression

Logistic Regression adalah algoritma klasifikasi linier yang memodelkan probabilitas hasil diskrit (seperti kategori kecanduan) menggunakan fungsi logistik. Meskipun namanya mengandung kata "regresi", algoritma ini banyak digunakan untuk tugas klasifikasi, khususnya dalam memprediksi kemungkinan keanggotaan suatu kelas. Keunggulannya meliputi interpretasi yang relatif mudah dan efisiensi komputasi untuk data yang terpisah secara linier.

2.3.2. K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) adalah algoritma non-parametrik yang diterapkan baik untuk klasifikasi maupun regresi. Dalam tugas klasifikasi, hasilnya adalah penentuan keanggotaan kelas di mana sebuah objek diklasifikasikan berdasarkan suara mayoritas dari tetangga-tetangga terdekatnya [9]. Algoritma ini menetapkan objek ke kelas yang paling umum di antara K tetangga terdekatnya. KNN dikenal karena kesederhanaan dan efektivitasnya, namun sensitif terhadap skala fitur dan "kutukan dimensi" (*curse of dimensionality*) pada *dataset* dengan banyak fitur.

2.3.3. Decision Tree

Decision Tree adalah model prediktif yang menggunakan struktur menyerupai pohon keputusan. Setiap *node* internal mewakili pengujian pada suatu fitur, setiap cabang menunjukkan hasil dari pengujian tersebut, dan setiap *node* daun (*leaf node*) merepresentasikan label kelas atau nilai keputusan. Model ini mudah diinterpretasikan, namun memiliki kecenderungan untuk mengalami *overfitting* pada data yang kompleks jika tidak diatur kedalamannya.

2.3.4. Random Forest

Random Forest merupakan algoritma *ensemble* yang membangun banyak pohon keputusan selama fase pelatihan [8]. Hasil klasifikasinya ditentukan oleh modus kelas (untuk tugas klasifikasi) atau rata-rata prediksi (untuk regresi) dari masing-masing pohon individual. Random Forest dikenal luas karena akurasinya yang tinggi dan kemampuannya dalam menangani banyak fitur [8]. Algoritma ini juga dapat menyediakan informasi penting mengenai signifikansi fitur (*feature importance*) dalam memprediksi target, yang berguna untuk memahami faktor-faktor pendorong prediksi [8]. Pemilihan algoritma ini didasarkan pada popularitas dan kinerjanya yang telah terbukti di berbagai domain klasifikasi.

3. METODE PENELITIAN

3.1. Desain Penelitian

Studi ini menerapkan pendekatan kuantitatif yang berpusat pada perancangan model *machine learning* untuk tujuan klasifikasi. Sasaran utamanya adalah menciptakan suatu sistem yang memiliki kemampuan untuk mengategorikan tingkat kecanduan smartphone pada populasi mahasiswa Universitas Esa Unggul secara otomatis. Diharapkan, model yang dihasilkan ini dapat menjadi fondasi awal bagi upaya deteksi dini dan intervensi yang lebih efektif, meskipun pada tahap eksplorasi awal.

3.2. Pengumpulan Data

Proses pengumpulan data dilaksanakan melalui survei daring yang didistribusikan kepada mahasiswa. Sebanyak 32 partisipan turut serta dalam penelitian ini, setelah sebelumnya memberikan persetujuan (*informed consent*) dan mendapatkan jaminan penuh atas kerahasiaan serta anonimitas data mereka. Semua responden adalah mahasiswa dari Universitas Esa Unggul, yang menjadi fokus utama penelitian ini. Instrumen utama yang dipakai adalah *Smartphone Addiction Scale-Short Version* (SAS-SV), sebuah alat ukur psikometrik yang dikembangkan dari kerangka kerja oleh Kwon et al. [1] yang terdiri dari 10 pertanyaan. Setiap pertanyaan direspons menggunakan skala Likert dengan pilihan jawaban dari 1 (Sangat Tidak Setuju) hingga 6 (Sangat Setuju). Penting untuk diakui bahwa data ini bersifat laporan mandiri (*self-report*), yang mungkin rentan terhadap bias respons.

3.3. Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk menyiapkan data mentah menjadi format yang sesuai untuk pemodelan *machine learning*. Pertama, kolom-kolom yang tidak relevan dengan analisis, seperti *Timestamp*, *Email Address*, dan Nama, dihapus. Kolom-kolom pertanyaan SAS-SV kemudian dipastikan dalam tipe data numerik integer untuk memudahkan pengolahan.

Selanjutnya, skor total kecanduan dihitung dengan menjumlahkan nilai respons dari kesepuluh pertanyaan SAS-SV untuk setiap responden. Proses kategorisasi tingkat kecanduan menjadi variabel target merupakan langkah krusial. Berdasarkan instrumen SAS-SV yang merujuk pada penelitian Kwon et al. [1], klasifikasi tingkat kecanduan dilakukan dengan mengidentifikasi individu yang berisiko tinggi. Kategori 'Tinggi' ditetapkan untuk responden laki-laki dengan skor total SAS-SV ≥ 31 dan untuk responden perempuan dengan skor total ≥ 33 [1], [2]. Pada analisis awal, ditemukan adanya ketidakseimbangan kelas dan bahkan sebuah kategori (Rendah) dengan jumlah sampel yang sangat minim (hanya 1 data), yang dapat menghambat proses pelatihan model *machine learning* yang stabil dan penggunaan teknik *stratified splitting*. Oleh karena itu, untuk tujuan pemodelan, kategori 'Rendah' digabungkan dengan kategori 'Sedang', sehingga menghasilkan dua kelas target akhir: 'Tinggi' dan 'Sedang'. Kategori 'Sedang' ini kini mencakup semua responden yang skornya di bawah ambang batas 'Tinggi' berdasarkan kriteria Kwon et al. [1].

Analisis lebih lanjut menunjukkan adanya ketidakseimbangan antara jumlah sampel di antara kedua kelas yang telah ditentukan (Tinggi dan Sedang). Untuk mengatasi isu ini dan memungkinkan model belajar secara efektif dari semua kelas, teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diterapkan secara eksklusif pada data latih. SMOTE bekerja dengan mensintesis sampel baru untuk kelas minoritas, sehingga jumlah sampel di setiap kategori menjadi lebih seimbang sebelum pelatihan model, yang krusial untuk mencegah bias model terhadap kelas mayoritas.

Terakhir, fitur-fitur numerik yang akan digunakan untuk pelatihan model perlu diseragamkan skalanya. Proses penskalaan fitur ini dilakukan menggunakan *StandardScaler*, yang mengubah distribusi fitur menjadi memiliki rata-rata nol dan standar deviasi satu. Penskalaan ini penting karena beberapa algoritma *machine learning* (seperti Logistic Regression dan K-Nearest Neighbors) sangat sensitif terhadap skala fitur, dan penskalaan yang tepat dapat meningkatkan kinerja model secara signifikan. Penskalaan diterapkan setelah pembagian data latih dan uji untuk menghindari *data leakage*.

3.4. Pemodelan Machine Learning

Dataset yang telah dipra-proses dibagi menjadi data latih (*training set*) dan data uji (*testing set*). Pembagian ini dilakukan dengan rasio 75% untuk data latih dan 25% untuk data uji. Pembagian menggunakan fungsi `train_test_split` dari pustaka Scikit-learn dengan `random_state=42` untuk memastikan reproduktifitas hasil, serta `stratify=y` untuk menjaga proporsi kelas yang sama di kedua set data, yang penting mengingat ketidakseimbangan kelas awal.

Empat algoritma klasifikasi *machine learning* kemudian diimplementasikan dan dibandingkan: Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree Classifier, dan Random Forest Classifier. Untuk setiap model, proses *hyperparameter tuning* dilakukan menggunakan GridSearchCV. Metode ini secara sistematis menjelajahi berbagai kombinasi *hyperparameter* yang telah ditentukan sebelumnya untuk menemukan konfigurasi optimal yang menghasilkan kinerja terbaik pada data latih. Metrik `f1_weighted` dipilih sebagai kriteria penilaian (*scoring*) untuk GridSearchCV karena mampu mengevaluasi kinerja model secara komprehensif pada dataset dengan kelas yang tidak seimbang. Setelah proses *tuning*, parameter terbaik yang ditemukan untuk setiap model adalah sebagai berikut:

- **Logistic Regression:** `C=1`, `multi_class='ovr'`, `solver='liblinear'`, dengan F1-score (Cross-Validation) terbaik 0.9495.
- **K-Nearest Neighbors:** `metric='euclidean'`, `n_neighbors=3`, `weights='uniform'`, dengan F1-score (Cross-Validation) terbaik 0.8708.
- **Decision Tree:** `max_depth=None`, `min_samples_leaf=1`, `min_samples_split=2`, dengan F1-score (Cross-Validation) terbaik 0.8485.
- **Random Forest:** `max_depth=None`, `min_samples_leaf=1`, `n_estimators=50`, dengan F1-score (Cross-Validation) terbaik 0.9749.

Semua model dilatih menggunakan data latih yang telah diseimbangkan oleh SMOTE dan diskalakan oleh StandardScaler untuk memastikan kinerja yang optimal, terutama pada kelas minoritas.

3.5. Evaluasi Model

Kinerja setiap model yang dilatih dievaluasi secara komprehensif menggunakan beberapa metrik standar yang umum dalam tugas klasifikasi. **Akurasi** digunakan sebagai indikator utama persentase prediksi yang benar secara keseluruhan. Untuk evaluasi yang lebih mendalam pada setiap kelas, **Presisi**, **Recall**, dan **F1-Score** juga dihitung. F1-Score adalah rata-rata harmonik dari presisi dan recall, memberikan ukuran keseimbangan antara keduanya, yang sangat berguna pada dataset yang tidak seimbang. Visualisasi *Confusion Matrix* juga disediakan untuk setiap model guna memahami pola kesalahan klasifikasi secara visual.

Lebih lanjut, analisis **tingkat kepentingan fitur** (*feature importance*) dilakukan pada model Random Forest. Meskipun Logistic Regression menunjukkan kinerja terbaik pada data uji, Random Forest dipilih untuk analisis ini karena kemampuannya dalam menyediakan nilai kepentingan fitur secara intrinsik dan mudah diinterpretasikan [8], yang membantu mengidentifikasi pertanyaan-pertanyaan SAS-SV mana yang paling berpengaruh dalam memprediksi tingkat kecanduan.

3.6. Perangkat Lunak dan Lingkungan

Seluruh proses penelitian ini, mulai dari pra-pemrosesan data hingga pemodelan *machine learning* dan evaluasi, diimplementasikan menggunakan bahasa pemrograman Python dalam lingkungan Google Colaboratory. Pustaka (*libraries*) utama yang dimanfaatkan meliputi Pandas untuk manipulasi data, NumPy untuk komputasi numerik, Matplotlib dan Seaborn untuk visualisasi data, Scikit-learn untuk implementasi algoritma *machine learning* dan metrik evaluasi, serta imbalanced-learn untuk teknik *oversampling* SMOTE.

4. HASIL DAN PEMBAHASAN

4.1. Deskripsi Data Responden

Penelitian ini melibatkan total 32 mahasiswa sebagai responden, yang semuanya berasal dari Universitas Esa Unggul. Fokus pada satu institusi ini bertujuan untuk memberikan wawasan mendalam pada konteks spesifik tersebut. Dari jumlah tersebut, mayoritas responden berjenis kelamin Pria dan Wanita, dengan sebaran usia terkonsentrasi antara 19-20 tahun. Responden berasal dari berbagai fakultas/program studi dan semester, dengan proporsi terbesar dari semester 4.

4.2. Distribusi Tingkat Kecanduan Smartphone

Setelah melalui tahapan pra-pemrosesan data dan pengategorian skor SAS-SV, sebaran tingkat kecanduan *smartphone* di antara responden menunjukkan pola sebagai berikut: kategori 'Tinggi' mencakup 26 responden, sementara kategori 'Sedang' mencakup 6 responden. Kategori 'Tinggi' secara spesifik mengacu pada ambang batas risiko kecanduan yang divalidasi oleh Kwon et al. [1], sementara kategori 'Sedang' ditetapkan untuk semua responden di bawah ambang batas risiko tersebut. Ketidakseimbangan awal pada jumlah sampel di kategori minoritas (Sedang) menjadi perhatian utama. Guna mengatasi isu tersebut, teknik *oversampling* SMOTE diaplikasikan pada data latih, memastikan sebaran kelas pada data latih menjadi lebih merata sebelum pelatihan model.

4.3. Kinerja Model Klasifikasi

Keempat model *Machine Learning* yang diuji—Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, dan Random Forest—menunjukkan variasi kinerja dalam mengklasifikasikan tingkat kecanduan *smartphone* pada data uji. Tabel 1 merangkum akurasi, presisi, *recall*, dan F1-score dari masing-masing model pada data uji.

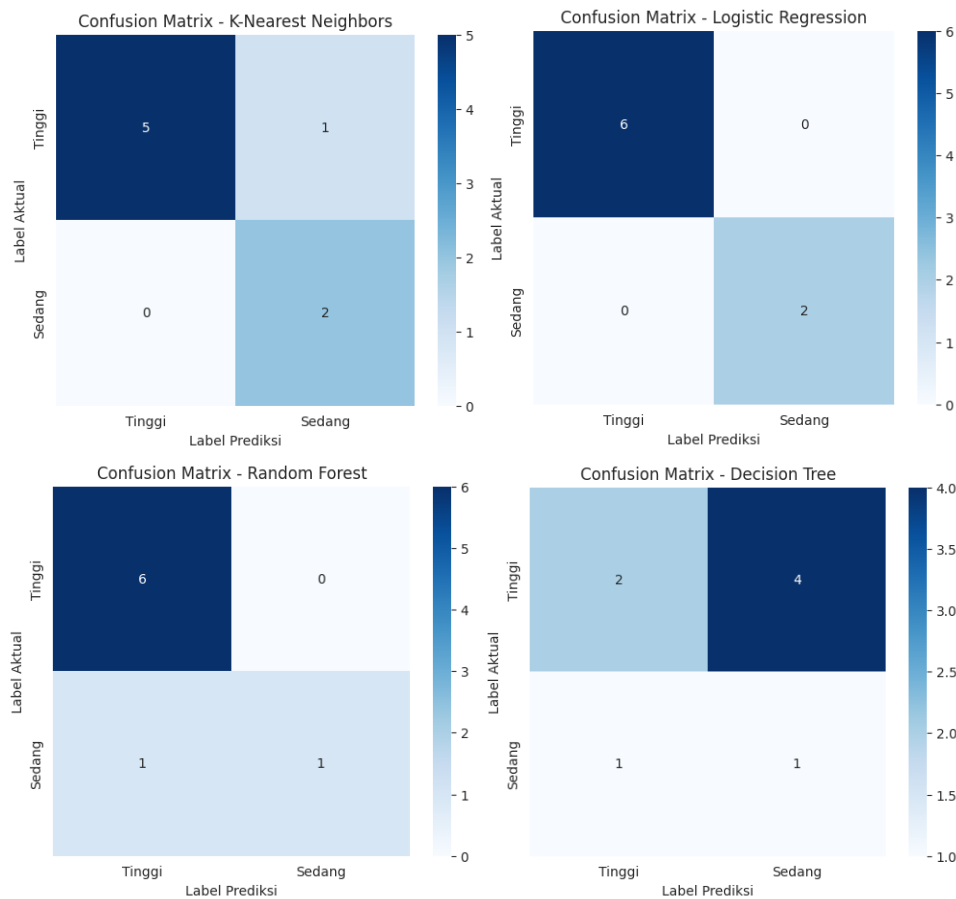
Tabel 1. Perbandingan Kinerja Empat Model Klasifikasi

Model	Akurasi (Accuracy)	Presisi (Weighted Avg)	Recall (Weighted Avg)	F1-Score (Weighted Avg)
-------	--------------------	------------------------	-----------------------	-------------------------

Logistic Regression	1.0000	1.0000	1.0000	1.0000
K-Nearest Neighbors	0.8750	0.9167	0.8750	0.8818
Random Forest	0.8750	0.8929	0.8750	0.8589
Decision Tree	0.3750	0.5500	0.3750	0.4048

Berdasarkan Tabel 1, model **Logistic Regression menunjukkan kinerja superior** dibandingkan dengan algoritma lainnya, mencapai akurasi tertinggi sebesar 1.0000 (100%). Model ini juga menunjukkan nilai Presisi, Recall, dan F1-Score (*Weighted Avg*) yang sempurna. Hasil ini mengindikasikan kemampuan model Logistic Regression yang sangat baik dalam mengklasifikasikan tingkat kecanduan *smartphone* pada *dataset* uji yang terbatas. Sebaliknya, model K-Nearest Neighbors dan Random Forest menunjukkan kinerja yang sama baiknya dengan akurasi 0.8750, sementara model Decision Tree menunjukkan kinerja paling rendah dengan akurasi 0.3750. Gambar 1 menunjukkan confusion matrix pada setiap algoritma.

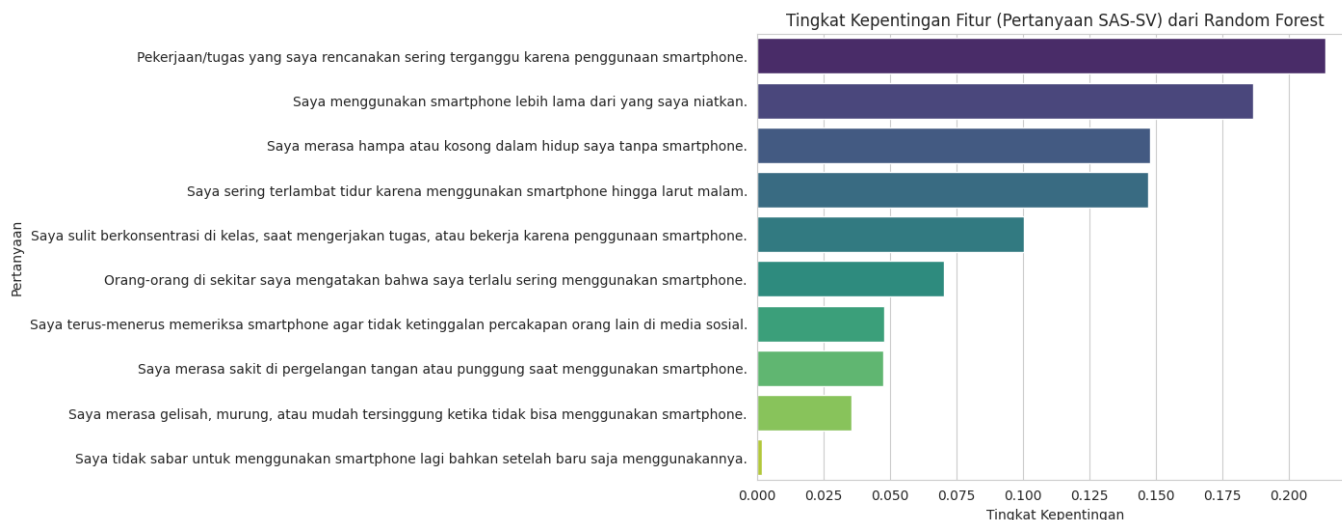
Gambar 1. Perbandingan confusion matrix setiap algoritma



4.4. Analisis Tingkat Kepentingan Fitur (*Feature Importance*)

Analisis tingkat kepentingan fitur dari model Random Forest memberikan wawasan berharga mengenai pertanyaan-pertanyaan SAS-SV mana yang paling berkontribusi dalam memprediksi tingkat kecanduan *smartphone*. Gambar 2 menampilkan urutan kepentingan fitur dari yang paling dominan hingga yang paling rendah.

Gambar 2. Grafik Feature Importance



Berdasarkan analisis tersebut, pertanyaan **Q0 (Pekerjaan/tugas yang saya rencanakan sering terganggu karena penggunaan smartphone.)** teridentifikasi sebagai fitur paling dominan. Hal ini mengindikasikan bahwa **dampak negatif pada produktivitas dan fungsi harian** merupakan prediktor yang sangat kuat untuk tingkat kecanduan. Temuan ini sejalan dengan literatur yang menunjukkan bahwa gangguan dalam kehidupan sehari-hari (*daily-life disturbance*) dan hilangnya kontrol adalah gejala inti dari perilaku adiktif [1].

Fitur penting lainnya termasuk **Q6 (Saya menggunakan smartphone lebih lama dari yang saya niatkan.)** dan **Q9 (Saya merasa hampa atau kosong dalam hidup saya tanpa smartphone.)**. Q6 merefleksikan aspek hilangnya kontrol penggunaan, di mana individu menggunakan *smartphone* melebihi niat awalnya. Sementara Q9 menunjukkan ketergantungan emosional. Hubungan antara penggunaan *smartphone* yang bermasalah dengan prokrastinasi akademik dan kecemasan juga telah didukung dalam penelitian sebelumnya [2]. Sebaliknya, pertanyaan **Q3 (Saya tidak sabar untuk menggunakan smartphone lagi bahkan setelah baru saja menggunakannya.)** menunjukkan tingkat kepentingan yang paling rendah.

4.5. Diskusi Umum dan Implikasi

Hasil penelitian ini menegaskan potensi *Machine Learning* dalam mendeteksi tingkat kecanduan *smartphone*, sebagaimana ditunjukkan oleh kinerja model Logistic Regression yang mencapai akurasi sempurna pada data uji. Identifikasi fitur-fitur penting, khususnya pertanyaan Q0, memberikan wawasan klinis yang dapat membantu psikolog atau konselor. Pertanyaan ini mengindikasikan bahwa dampak pada produktivitas dan fungsi akademik adalah prediktor yang sangat kuat. Temuan ini selaras dengan literatur yang menunjukkan bahwa penggunaan *smartphone* yang bermasalah dapat secara signifikan mengganggu kinerja akademik dan menyebabkan perilaku prokrastinasi [2], [3]. Pentingnya Q0 sebagai prediktor juga diperkuat oleh fitur-fitur lain seperti Q6 (hilangnya kontrol) dan Q9 (ketergantungan emosional), yang keduanya merupakan gejala inti kecanduan perilaku [1].

Meskipun model Logistic Regression menunjukkan performa yang sempurna, penting untuk menempatkan hasil ini dalam perspektif yang kritis. Akurasi 100% pada data uji yang sangat terbatas (8 sampel) merupakan indikasi kuat adanya *overfitting* atau kemungkinan besar terjadi secara kebetulan akibat ukuran *dataset* yang sangat kecil. Dalam praktik *machine learning* pada data dunia nyata, sangat jarang ditemui model yang mencapai akurasi sempurna. Penelitian lain di berbagai domain klasifikasi, termasuk deteksi serangan siber [5] dan prediksi penyakit COVID-19 [7], umumnya melaporkan akurasi tinggi namun tidak mutlak 100%. Hal ini menggarisbawahi bahwa meskipun model menunjukkan kemampuan memisahkan kelas dengan sangat baik pada sampel yang diamati, kemampuan generalisasinya untuk populasi yang lebih luas masih diragukan.

4.6. Keterbatasan Penelitian dan Saran untuk Penelitian Selanjutnya

Studi ini memiliki beberapa keterbatasan. Pertama, **ukuran sampel yang relatif kecil** (total 32 responden) membatasi generalisasi temuan. Akurasi 1.0000 yang dicapai harus diinterpretasikan dengan hati-hati karena risiko *overfitting* yang tinggi pada sampel terbatas. Kedua, data bergantung pada **laporan mandiri (self-report)**, yang rentan terhadap bias. Ketiga, meskipun SMOTE efektif, ketersediaan lebih banyak data asli akan lebih baik. Keempat, penggabungan kategori 'Sedang' dan 'Rendah' dilakukan secara deskriptif dan **belum memiliki validasi eksternal** seperti kategori 'Tinggi' yang mengacu pada Kwon et al. [1].

Untuk penelitian di masa mendatang, disarankan untuk:

- **Mengumpulkan sampel yang lebih besar dan lebih bervariasi** untuk meningkatkan generalisasi model.
- **Mengintegrasikan data kuesioner dengan data objektif**, seperti data penggunaan aplikasi dari fitur *digital wellbeing*.
- **Mengeksplorasi algoritma machine learning yang lebih kompleks** seperti model *deep learning*, yang telah menunjukkan performa superior pada *dataset* besar dalam berbagai tugas prediksi [7], [10].
- **Melakukan validasi model secara eksternal** pada populasi yang berbeda untuk menguji generalisasi.
- **Mengembangkan model untuk memprediksi tingkat keparahan kecanduan** atau dampaknya secara spesifik.

5. KESIMPULAN

Penelitian ini telah berhasil mengembangkan dan mengevaluasi model machine learning untuk mengklasifikasikan tingkat kecanduan smartphone pada mahasiswa Universitas Esa Unggul, menggunakan data dari 32 responden yang dikumpulkan melalui instrumen Smartphone Addiction Scale-Short Version (SAS-SV). Untuk tujuan pemodelan, tingkat kecanduan dikelompokkan menjadi dua kategori: 'Tinggi' (berisiko kecanduan, berdasarkan kriteria spesifik gender dari Kwon et al.) dan 'Sedang' (mencakup responden dengan skor di bawah ambang batas tersebut).

Dari empat algoritma klasifikasi yang diuji (Logistic Regression, K-Nearest Neighbors, Decision Tree, dan Random Forest), model Logistic Regression menunjukkan kinerja superior pada data uji dengan akurasi 1.0000, presisi 1.00, recall 1.00, dan F1-score 1.00. Meskipun performa sempurna ini mengindikasikan kemampuan model yang sangat baik pada dataset terbatas ini, perlu dicatat bahwa hasil tersebut diperoleh dari ukuran data uji yang sangat minim (8 sampel), sehingga memerlukan interpretasi hati-hati terhadap potensi overfitting atau kebetulan.

Analisis tingkat kepentingan fitur dari model Random Forest mengidentifikasi bahwa pertanyaan Pekerjaan/tugas yang saya rencanakan sering terganggu karena penggunaan smartphone (Q0) merupakan prediktor paling signifikan, menyoroti dampak pada produktivitas sebagai indikator kuat kecanduan, yang sejalan dengan dimensi gangguan kehidupan sehari-hari (daily-life disturbance) dalam studi adiksi. Secara keseluruhan, studi ini menunjukkan potensi awal machine learning sebagai alat bantu skrining untuk mengidentifikasi mahasiswa yang berisiko kecanduan smartphone. Namun, karena keterbatasan ukuran sampel dan sifat data self-report, temuan ini bersifat eksploratif dan memerlukan validasi ekstensif pada data yang lebih besar sebelum implementasi praktis.

Ucapan Terimakasih

Penulis menyampaikan apresiasi dan rasa terima kasih yang tulus kepada Bapak ARY PRABOWO, S.Kom., M.Kom., sebagai dosen pembimbing atas bimbingan dan masukan berharga yang telah diberikan. Apresiasi juga disampaikan kepada seluruh responden yang telah berpartisipasi dalam pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] L. J.-Y. Won, W.-Y. Park, and J.-W. Min, "Development and Validation of a Smartphone Addiction Scale (SAS)," *PLoS One*, vol. 8, no. 2, p. 56936, 2013, doi: 10.1371/journal.pone.0056936.
- [2] Z. Yang, K. Asbury, and M. D. Griffiths, "An Exploration of Problematic Smartphone Use among Chinese University Students: Associations with Academic Anxiety, Academic Procrastination, Self-Regulation and Subjective Wellbeing," *Int J Ment Health Addiction*, vol. 17, pp. 596–614, 2019, doi: 10.1007/s11469-018-9961-1.
- [3] F. Indah Putri, I. Sitorus, H. Khair, N. Yanti, and I. Agama Islam Sumatera Barat Pariaman, "Hubungan Kecanduan Smartphone Dengan Prokrastinasi Akademik Siswa," *Ikhtisar: Jurnal Pengetahuan Islam*, vol. 3, no. 2, pp. 237–248, Dec. 2023, doi: 10.55062/IJPI.V3I2.388.
- [4] Z. Hao *et al.*, "Physics-Informed Machine Learning: A Survey on Problems, Methods and Applications."
- [5] R. G. Gunawan, E. S. Handika, and E. Ismanto, "Pendekatan Machine Learning Dengan Menggunakan Algoritma Xgboost (Extreme Gradient Boosting) Untuk Peningkatan Kinerja Klasifikasi Serangan Syn," *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 3, no. 3, pp. 453–463, Dec. 2022, doi: 10.37859/COSCITECH.V3I3.4356.
- [6] R. Gunawan and M. I. Pratama, "Analisa kinerja algoritma machine learning untuk prediksi virus hepatitis C," vol. 4, no. 3, pp. 772–777, 2023, doi: 10.37859/coscitech.v4i3.6513.
- [7] C. Reyes-Aldasoro, B. Alsaaidah, R. Al-Hadidi, H. Al-Nsour, R. Masadeh, and N. AlZubi, "Imaging Comprehensive Survey of Machine Learning Systems for COVID-19 Detection," *J. Imaging*, vol. 2022, p. 267, 2022, doi: 10.3390/jimaging8100267.
- [8] L. Breiman, "Random forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324/METRICS.
- [9] T. M. Cover and P. E. Hart, "Nearest Neighbor Pattern Classification," *IEEE Trans Inf Theory*, vol. 13, no. 1, pp. 21–27, 1967, doi: 10.1109/TIT.1967.1053964.
- [10] T. Hall and K. Rasheed, "A Survey of Machine Learning Methods for Time Series Prediction," *Applied Sciences 2025, Vol. 15, Page 5957*, vol. 15, no. 11, p. 5957, May 2025, doi: 10.3390/APP15115957.