Penerapan Algoritma Random Forest Untuk Klasifikasi Tingkat Kecanduan Smartphone Pada Pengguna Bedasarkan Pola Penggunaan Digital

Caesar Galang Restu Ramadhan[1*], Fatkurrahman Raffie Wibowo[2], Nisha Kalya Salsabilla[3], Ridwan Yoga Pratama[4]

[1]S1 Teknik Informatik/Ilmu Komputer
Universitas Duta Bangsa Surakarta
[1*]230103187@mhs.udb.ac.id

[2]S1 Teknik Informatik/Ilmu Komputer
Universitas Duta Bangsa Surakarta
[2]230103193@mhs.udb.ac.id

[3] S1 Teknik Informatik/Ilmu Komputer
Universitas Duta Bangsa Surakarta
[3]230103203@mhs.udb.ac.id

[4]S1 Teknik Informatik/Ilmu Komputer
Universitas Duta Bangsa Surakarta
[4]230103206@mhs.udb.ac.id

***Abstrak*—Penggunaan smartphone yang semakin meningkat dalam berbagai aktivitas sehari-hari memberikan banyak manfaat, namun juga berpotensi menyebabkan kecanduan yang dapat berdampak pada kesehatan, produktivitas, dan interaksi sosial. Oleh karena itu, diperlukan suatu sistem yang mampu mengidentifikasi tingkat kecanduan smartphone secara cepat dan akurat. Penelitian ini bertujuan mengembangkan sistem klasifikasi tingkat kecanduan smartphone berbasis web menggunakan algoritma Random Forest dengan menerapkan metode Cross Industry Standard Process for Data Mining (CRISP-DM). Tahapan penelitian meliputi *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Dataset yang digunakan merupakan Mobile Addiction Dataset yang terdiri atas 13.589 data dengan 12 atribut, kemudian dilakukan preprocessing berupa penghapusan atribut yang tidak relevan, *encoding*, *standard scaling*, serta pembagian data menjadi data latih dan data uji dengan rasio 80:20. Hasil penelitian menunjukkan bahwa model Random Forest menghasilkan akurasi pengujian sebesar 97,5% dengan nilai *Cross Validation* sebesar 97,6%, sehingga mampu mengklasifikasikan tingkat kecanduan smartphone dengan tingkat kesalahan yang rendah. Implementasi model ke dalam aplikasi berbasis Flask berhasil mengintegrasikan seluruh tahapan klasifikasi dalam satu sistem yang mudah digunakan. Dengan demikian, tujuan penelitian untuk membangun sistem klasifikasi tingkat kecanduan smartphone berbasis web yang akurat dan terintegrasi telah berhasil dicapai.**

***Kata kunci*— Smartphone Addiction, Random Forest, CRISP-DM, Klasifikasi, Flask**

***Abstract*— The increasing use of smartphones in various daily activities provides numerous benefits but also increases the risk of smartphone addiction, which can negatively affect users' health, productivity, and social interactions. Therefore, a system capable of accurately and efficiently identifying smartphone addiction levels is needed. This study aims to develop a web-based smartphone addiction classification system using the Random Forest algorithm by adopting the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology. The research consists of five stages: Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The dataset used is the Mobile Addiction Dataset, which contains 13,589 records with 12 attributes. Data preprocessing included removing irrelevant attributes, encoding categorical data, applying standard scaling, and splitting the dataset into training and testing sets with an 80:20 ratio. The experimental results demonstrate that the Random Forest model achieved a 97.5% testing accuracy and a 97.6% cross-validation score, indicating high classification performance with a low error rate. The trained model was successfully deployed into a Flask-based web application, integrating the entire classification process into a single user-friendly platform. These findings indicate that the proposed system effectively classifies smartphone addiction levels and successfully achieves the objective of developing an accurate, integrated, and web-based classification system.**

***Keywords*— Smartphone Addiction, Random Forest, CRISP-DM, Classification, Flask.**

I. PENDAHULUAN

Perkembangan teknologi digital telah mendorong peningkatan penggunaan smartphone dalam berbagai aktivitas sehari-hari, mulai dari komunikasi, pendidikan, pekerjaan, hiburan, hingga interaksi melalui media sosial. Kemudahan akses yang ditawarkan smartphone menyebabkan durasi penggunaan perangkat ini

terus meningkat dan menjadi bagian yang tidak terpisahkan dari kehidupan masyarakat modern. Namun, penggunaan smartphone yang berlebihan dapat memunculkan berbagai permasalahan, salah satunya adalah kecanduan smartphone (smartphone addiction) yang dapat berdampak pada kesehatan mental, produktivitas, kualitas tidur, serta interaksi sosial pengguna (Verrel et al., 2025)[1].

Identifikasi tingkat kecanduan smartphone menjadi penting karena perilaku adiktif sering kali tidak disadari oleh pengguna. Selama ini, pengukuran tingkat kecanduan umumnya dilakukan melalui survei atau instrumen psikologis yang membutuhkan waktu dan keterlibatan responden secara langsung(Angkasa et al., 2026)[2]. Seiring berkembangnya teknologi data mining dan machine learning, proses identifikasi tingkat kecanduan dapat dilakukan dengan memanfaatkan pola penggunaan digital yang terekam dalam data pengguna. Pendekatan ini memungkinkan proses klasifikasi dilakukan secara lebih cepat, objektif, dan otomatis[3].

Penelitian ini menggunakan Mobile Addiction Dataset yang terdiri dari 13.589 data pengguna smartphone dengan 12 atribut yang merepresentasikan perilaku penggunaan digital. Dataset tersebut memuat berbagai variabel yang berpotensi mempengaruhi tingkat kecanduan smartphone, seperti durasi penggunaan harian (daily_screen_time), jumlah sesi penggunaan aplikasi (app_sessions), durasi penggunaan media sosial (social_media_usage), waktu bermain game (gaming_time), jumlah notifikasi yang diterima (notifications), penggunaan smartphone pada malam hari (night_usage), tingkat stres (stress_level), jumlah aplikasi yang terpasang (apps_installed), serta kategori pekerjaan atau pendidikan pengguna. Selain itu, dataset juga memiliki label target berupa kategori addicted dan not addicted yang dapat digunakan untuk membangun model klasifikasi berbasis machine learning.
Besarnya jumlah data yang tersedia memberikan peluang untuk melakukan analisis terhadap hubungan antara pola penggunaan digital dengan tingkat kecanduan smartphone. Namun, kompleksitas hubungan antar variabel membuat proses identifikasi tidak dapat dilakukan secara efektif menggunakan metode konvensional. Oleh karena itu, diperlukan suatu algoritma klasifikasi yang mampu mempelajari pola dari data historis dan menghasilkan prediksi yang akurat terhadap tingkat kecanduan pengguna smartphone[4], [5], [6].

Salah satu algoritma yang banyak digunakan dalam permasalahan klasifikasi adalah Random Forest. Algoritma ini merupakan metode ensemble learning yang menggabungkan banyak pohon keputusan (decision tree) untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting[7]. Random Forest memiliki kemampuan yang baik dalam menangani dataset berukuran besar, data dengan banyak atribut, serta mampu menghasilkan informasi mengenai tingkat kepentingan fitur (feature importance). Dengan kemampuan tersebut, Random Forest tidak hanya dapat digunakan untuk mengklasifikasikan pengguna ke dalam kategori addicted atau not addicted, tetapi juga dapat mengidentifikasi faktor-faktor penggunaan digital yang paling berpengaruh terhadap tingkat kecanduan smartphone(Setyanto & Wicaksono, 2026)[1], [8].

Untuk memudahkan implementasi model klasifikasi, penelitian ini juga mengembangkan aplikasi berbasis web menggunakan framework Flask. Aplikasi web tersebut dirancang untuk melakukan proses unggah dataset, preprocessing data, pelatihan model Random Forest, pengujian model, serta menampilkan hasil klasifikasi dan visualisasi feature importance secara interaktif. Dengan adanya aplikasi web ini, proses klasifikasi tingkat kecanduan smartphone dapat dilakukan secara lebih mudah dan dapat dimanfaatkan sebagai media analisis bagi peneliti maupun pengguna lainnya.

Berdasarkan uraian tersebut, penelitian ini bertujuan menerapkan algoritma Random Forest untuk mengklasifikasikan tingkat kecanduan

smartphone berdasarkan pola penggunaan digital pengguna menggunakan Mobile Addiction Dataset serta mengimplementasikan model yang dihasilkan ke dalam aplikasi berbasis web untuk mendukung proses identifikasi tingkat kecanduan smartphone secara otomatis dan akurat

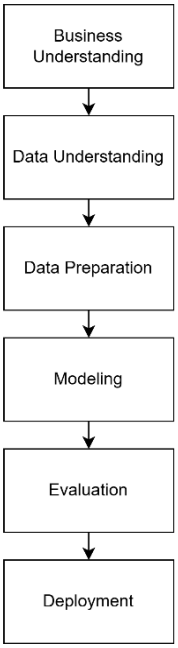
II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM) sebagai kerangka kerja dalam proses pembangunan model klasifikasi tingkat kecanduan smartphone[9]. Metode CRISP-DM dipilih karena menyediakan tahapan yang sistematis mulai dari pemahaman masalah bisnis hingga implementasi model. Algoritma klasifikasi yang digunakan adalah Random Forest, sedangkan evaluasi performa model dilakukan menggunakan Confusion Matrix dengan metrik Accuracy, Precision, Recall, dan F1-Score[10][11].

2.1. Alur Penelitian Menggunakan CRISP-DM

Penelitian ini menggunakan metodologi Cross Industry Standard Process for Data Mining (CRISP-DM) sebagai kerangka kerja dalam proses pembangunan model klasifikasi tingkat kecanduan smartphone. Metode CRISP-DM dipilih karena menyediakan tahapan yang sistematis mulai dari pemahaman masalah bisnis hingga implementasi model. Algoritma klasifikasi yang digunakan adalah Random Forest, sedangkan evaluasi performa model dilakukan menggunakan Confusion Matrix dengan metrik Accuracy, Precision, Recall, dan F1-Score[9].

Tahapan penelitian yang digunakan dalam metodologi CRISP-DM terdiri atas enam fase utama, yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*. Keenam tahapan tersebut saling berkaitan, di mana hasil dari setiap tahapan menjadi dasar untuk melanjutkan proses pada tahapan berikutnya. Kerangka penelitian yang digunakan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Flowchart Alur Penelitian CRISP-DM

2.2. Bussines Understanding

Tahap Business Understanding merupakan tahap awal dalam metodologi CRISP-DM yang bertujuan untuk memahami permasalahan yang akan diselesaikan serta menentukan tujuan penelitian. Pada penelitian ini, permasalahan yang diangkat adalah meningkatnya penggunaan smartphone dalam berbagai aktivitas sehari-hari yang berpotensi menimbulkan kecanduan. Tingginya intensitas penggunaan smartphone dapat memberikan dampak negatif terhadap produktivitas, kesehatan mental, kualitas tidur, maupun interaksi sosial pengguna. Oleh karena itu, diperlukan suatu sistem yang mampu mengidentifikasi tingkat kecanduan smartphone secara otomatis berdasarkan pola penggunaan digital pengguna.

2.3. Data Understanding

Tahap Data Understanding bertujuan untuk memahami karakteristik dataset yang akan digunakan dalam proses pembangunan model klasifikasi. Pada tahap ini dilakukan identifikasi terhadap sumber data, struktur

dataset, jenis atribut, jumlah data, serta distribusi kelas target. Pemahaman terhadap karakteristik data menjadi penting untuk menentukan strategi pengolahan data pada tahap berikutnya.

Dataset yang digunakan dalam penelitian ini adalah Mobile Addiction Dataset yang diperoleh dari Kaggle dan dipublikasikan oleh Prabal Pratap Singh. Dataset tersebut terdiri atas 13.589 data pengguna smartphone dengan 12 atribut, yang merepresentasikan berbagai pola penggunaan smartphone, seperti daily_screen_time, app_sessions, social_media_usage, gaming_time, notifications, night_usage, age, work_study_hours, stress_level, apps_installed, serta atribut target addicted.

Pada tahap ini dilakukan proses eksplorasi data untuk mengetahui jumlah data, tipe data setiap atribut, distribusi kelas target, serta identifikasi awal terhadap kualitas data, seperti keberadaan nilai kosong (missing values) dan inkonsistensi data. Hasil eksplorasi tersebut menjadi dasar dalam menentukan langkah-langkah preprocessing pada tahap Data Preparation.

2.4. Data Preparation

Tahap Data Preparation merupakan proses mempersiapkan dataset agar siap digunakan pada proses pembentukan model klasifikasi. Tahap Data Preparation merupakan lanjutan dari tahap Data Understanding. Berdasarkan hasil eksplorasi data yang telah dilakukan sebelumnya, pada tahap ini dilakukan berbagai proses preprocessing untuk meningkatkan kualitas data sehingga dataset siap digunakan dalam proses pembentukan model klasifikasi menggunakan algoritma Random Forest. Pada penelitian ini dilakukan beberapa tahapan preprocessing sebagai berikut.

a) Penghapusan Data

Tahap pertama dilakukan dengan menghapus atribut yang tidak memberikan kontribusi terhadap proses klasifikasi. Atribut Unnamed: 0 dihapus karena hanya berfungsi sebagai nomor urut data (row identifier) dan tidak memiliki informasi yang berkaitan dengan tingkat kecanduan smartphone. Selain itu, sistem juga melakukan pemeriksaan terhadap atribut yang berpotensi menyebabkan data leakage sehingga model tidak memperoleh informasi yang dapat memengaruhi hasil prediksi secara tidak wajar.

b) Penangan Missing Value

Selanjutnya dilakukan pemeriksaan terhadap keberadaan missing value pada dataset. Apabila ditemukan data yang memiliki nilai kosong, sistem melakukan penanganan sesuai metode yang dipilih pada aplikasi, seperti penghapusan data (drop) atau pengisian nilai menggunakan metode statistik. Tahapan ini bertujuan menjaga kualitas data sehingga proses pelatihan model dapat berlangsung secara optimal.

c) Encoding Data

Selanjutnya dilakukan pemeriksaan terhadap keberadaan missing value pada dataset. Apabila ditemukan data yang memiliki nilai kosong, sistem melakukan penanganan sesuai metode yang dipilih pada aplikasi, seperti penghapusan data (drop) atau pengisian nilai menggunakan metode statistik. Tahapan ini bertujuan menjaga kualitas data sehingga proses pelatihan model dapat berlangsung secara optimal.

d) Feature Selection

Feature selection dilakukan untuk memilih atribut yang relevan terhadap proses klasifikasi. Atribut yang tidak memiliki hubungan dengan target atau hanya berfungsi sebagai identitas dihilangkan sehingga model hanya menggunakan fitur yang berpengaruh terhadap tingkat kecanduan smartphone. Tahapan ini bertujuan meningkatkan efisiensi proses

pelatihan serta mengurangi kompleksitas model.

e) Standart Scaling

Algoritma Random Forest pada dasarnya tidak memerlukan proses feature scaling karena proses pemisahan data pada setiap decision tree didasarkan pada nilai ambang (threshold), bukan pada perhitungan jarak antar data. Oleh karena itu, penerapan Standard Scaling bukan merupakan kebutuhan utama dalam pembentukan model Random Forest.

Meskipun demikian, pada penelitian ini proses Standard Scaling tetap diterapkan pada atribut numerik sebagai bagian dari tahapan preprocessing agar seluruh atribut berada pada skala yang seragam. Langkah ini dilakukan untuk menjaga konsistensi pengolahan data serta mempermudah pengembangan sistem apabila pada penelitian selanjutnya dilakukan perbandingan performa dengan algoritma lain yang sensitif terhadap skala data, seperti Support Vector Machine (SVM) atau K-Nearest Neighbor (KNN). Dengan demikian, proses scaling dilakukan bukan untuk meningkatkan performa Random Forest secara langsung, melainkan sebagai bentuk standardisasi data dalam sistem yang dikembangkan.

f) Split Data

Tahap terakhir pada proses preprocessing adalah membagi dataset menjadi data latih (training) dan data uji (testing). Pembagian data dilakukan menggunakan rasio 80% untuk data training dan 20% untuk data testing dengan nilai random state sebesar 42. Pembagian ini bertujuan agar model dapat dilatih menggunakan sebagian besar data, sedangkan data sisanya digunakan untuk mengukur kemampuan model dalam melakukan prediksi terhadap data yang belum pernah dipelajari sebelumnya.

2.5. Modeling

Tahap Modeling merupakan proses pembangunan model klasifikasi menggunakan algoritma Random Forest. Algoritma ini dipilih karena mampu menangani dataset berukuran besar, memiliki ketahanan terhadap noise dan overfitting, serta dapat memberikan informasi mengenai tingkat kepentingan setiap atribut (feature importance). Random Forest bekerja dengan membangun sejumlah Decision Tree, kemudian hasil prediksi akhir ditentukan berdasarkan mekanisme majority voting dari seluruh pohon keputusan yang terbentuk.

Pada penelitian ini, proses pelatihan model dilakukan menggunakan bahasa pemrograman Python dengan pustaka Scikit-learn. Parameter model yang digunakan pada aplikasi meliputi n_estimators = 120, max_depth = 8, min_samples_split = 20, min_samples_leaf = 8, serta class_weight = balanced. Penggunaan parameter tersebut bertujuan menghasilkan model yang mampu melakukan klasifikasi secara optimal sekaligus mengurangi kemungkinan terjadinya overfitting.

2.6. Evaluasi

Tahap Evaluation dilakukan untuk mengetahui tingkat performa model Random Forest dalam mengklasifikasikan tingkat kecanduan smartphone. Evaluasi dilakukan menggunakan Confusion Matrix yang menghasilkan empat komponen utama, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN). Berdasarkan Confusion Matrix, performa model diukur menggunakan beberapa metrik evaluasi, yaitu Accuracy, Precision, Recall, dan F1-Score. Accuracy digunakan untuk mengetahui persentase prediksi yang benar terhadap seluruh data uji. Precision menunjukkan ketepatan model dalam memprediksi suatu kelas, Recall mengukur kemampuan model menemukan seluruh data

pada kelas yang sebenarnya, sedangkan F1-Score merupakan rata-rata harmonis antara Precision dan Recall sehingga memberikan gambaran performa model secara lebih seimbang.

2.7. Deployment

Tahap terakhir dalam metodologi CRISP-DM adalah Deployment, yaitu mengimplementasikan model yang telah dibangun ke dalam aplikasi berbasis web. Pada penelitian ini aplikasi dikembangkan menggunakan framework Flask dengan bahasa pemrograman Python sehingga pengguna dapat melakukan seluruh proses klasifikasi melalui antarmuka web.

Aplikasi yang dikembangkan terdiri atas enam fitur utama, yaitu Dashboard, Upload Dataset, Preprocessing, Klasifikasi Random Forest, Evaluasi, dan Hasil Klasifikasi. Melalui aplikasi tersebut pengguna dapat mengunggah dataset dalam format CSV, melakukan preprocessing data, melatih model Random Forest, melihat hasil evaluasi berupa Confusion Matrix dan metrik klasifikasi, serta menampilkan hasil prediksi dan visualisasi feature importance. Dengan implementasi ini, proses identifikasi tingkat kecanduan smartphone dapat dilakukan secara otomatis, lebih efisien, dan mudah digunakan tanpa harus menjalankan program Python secara langsung.

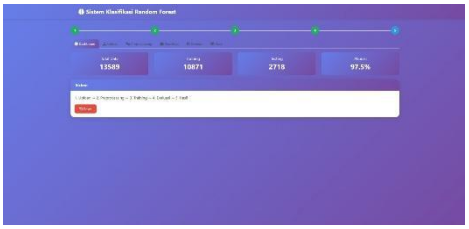
III. HASIL DAN PEMBAHASAN

Pada penelitian ini telah berhasil dikembangkan sebuah sistem klasifikasi tingkat kecanduan smartphone berbasis web menggunakan framework Flask dengan menerapkan algoritma Random Forest. Sistem dibangun berdasarkan tahapan metode CRISP-DM, mulai dari proses pemahaman data (Data Understanding), persiapan data (Data Preparation), pembangunan model (Modeling), evaluasi (Evaluation), hingga implementasi sistem (Deployment). Melalui aplikasi ini, seluruh proses klasifikasi dapat dilakukan secara terintegrasi sehingga pengguna dapat mengunggah dataset, melakukan preprocessing, melatih model, mengevaluasi performa model, dan memperoleh hasil klasifikasi secara otomatis.

3.1.Implementasi Business Understanding

Tahap Business Understanding merupakan tahap awal yang kami lakukan untuk memahami permasalahan serta menentukan tujuan pengembangan sistem. Berdasarkan hasil identifikasi, penggunaan smartphone yang semakin meningkat dapat memicu terjadinya kecanduan sehingga diperlukan suatu sistem yang mampu mengidentifikasi tingkat kecanduan secara otomatis berdasarkan pola penggunaan smartphone.

Sebagai implementasi dari tahap ini, kami mengembangkan aplikasi berbasis web menggunakan framework Flask dengan algoritma Random Forest sebagai metode klasifikasi. Sistem dirancang untuk memudahkan pengguna dalam melakukan seluruh proses data mining, mulai dari pengunggahan dataset, preprocessing, pelatihan model, evaluasi, hingga menampilkan hasil klasifikasi dalam satu aplikasi.



Gambar 3.1 Flowchart Alur Penelitian CRISP-DM

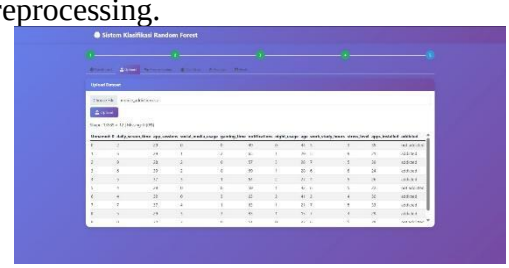
Gambar 3.1 menunjukkan halaman Dashboard sebagai tampilan utama sistem. Halaman ini menyajikan informasi mengenai jumlah dataset, data training, data testing, dan nilai akurasi model. Selain itu, dashboard juga menyediakan menu navigasi menuju halaman Upload Dataset, Preprocessing, Klasifikasi Random Forest, Evaluasi, dan Hasil Klasifikasi. Dengan demikian, dashboard menjadi implementasi dari tahap Business Understanding karena menyediakan antarmuka yang mendukung tujuan penelitian, yaitu membangun sistem

klasifikasi tingkat kecanduan smartphone yang mudah digunakan.

3.2.Implementasi Data Understanding

Tahap Data Understanding dilakukan untuk memahami karakteristik dataset yang akan digunakan dalam proses klasifikasi. Pada penelitian ini, kami menggunakan Mobile Addiction Dataset yang diperoleh dari Kaggle dan dipublikasikan oleh Prabal Pratap Singh. Dataset tersebut dipilih karena memuat informasi mengenai perilaku penggunaan smartphone yang dapat digunakan untuk mengidentifikasi tingkat kecanduan pengguna.

Implementasi tahap ini dilakukan melalui halaman Upload Dataset, dimana pengguna mengunggah dataset berformat CSV ke dalam sistem. Setelah proses unggah berhasil, sistem secara otomatis membaca dataset dan menampilkan informasi seperti jumlah data, jumlah atribut, jumlah missing value, serta contoh isi dataset dalam bentuk tabel. Informasi tersebut digunakan untuk memastikan bahwa dataset telah berhasil dimuat dan siap digunakan pada tahap preprocessing.



Gambar 3.2 Flowchart Alur Penelitian CRISP-DM

Berdasarkan Gambar 3.2, terlihat bahwa dataset berhasil dimuat ke dalam sistem. Halaman ini menampilkan nama file dataset, jumlah data, jumlah atribut, serta beberapa baris pertama dataset yang terdiri dari atribut daily_screen_time, app_sessions, social_media_usage, gaming_time, notifications, night_usage, age, work_study_hours, stress_level, apps_installed, dan addicted sebagai atribut target. Selain itu, sistem juga menampilkan hasil pemeriksaan missing value sebagai langkah awal untuk mengetahui kualitas data sebelum dilakukan proses pengolahan lebih lanjut.

3.1. Implementasi Data Preparation

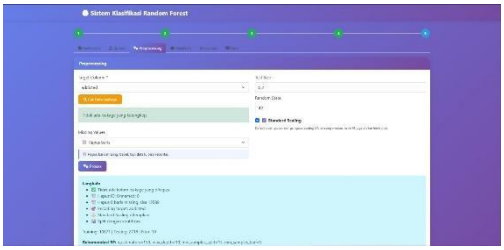
Pada tahap berikutnya adalah Data Preparation yang bertujuan menyiapkan dataset agar siap digunakan dalam proses pelatihan model. Pada penelitian ini seluruh proses preprocessing dilakukan secara otomatis melalui halaman preprocessing pada aplikasi web.

Tahapan pertama yang dilakukan adalah menentukan atribut target yaitu addicted. Selanjutnya sistem melakukan pemeriksaan kemungkinan terjadinya data leakage untuk memastikan tidak terdapat atribut yang dapat menyebabkan model memperoleh informasi target secara langsung. Berdasarkan hasil pemeriksaan, sistem menunjukkan bahwa tidak ditemukan data leakage sehingga seluruh atribut masih layak digunakan.

Tahapan berikutnya adalah menghapus atribut Unnamed:0 karena hanya berfungsi sebagai indeks data dan tidak memberikan kontribusi terhadap proses klasifikasi. Sistem kemudian melakukan proses encoding terhadap atribut target sehingga label addicted dan not addicted dapat diproses oleh algoritma Random Forest.

Selain itu diterapkan proses Standard Scaling agar seluruh atribut numerik memiliki skala yang seragam. Setelah preprocessing selesai dilakukan, dataset dibagi menjadi 80% data training dan 20% data testing menggunakan metode stratified split dengan nilai random state sebesar 42.

Dataset yang digunakan merupakan Mobile Addiction Dataset yang terdiri dari 13.589 data dengan 12 atribut, dimana satu atribut digunakan sebagai target klasifikasi yaitu addicted. Berdasarkan hasil pembacaan sistem diketahui bahwa dataset tidak memiliki missing value, sehingga seluruh data dapat digunakan pada proses selanjutnya tanpa memerlukan proses pembersihan data yang kompleks.



Gambar 3.3 Flowchart Alur Penelitian CRISP-DM

Pada Gambar 3.3 menjunkan hasil preprocessing menunjukkan bahwa diperoleh 10.871 data training, 2.718 data testing, serta 10 atribut yang digunakan dalam proses klasifikasi.

3.2.Implementasi Modeling

Berikutnya adalah tahap modelling, tahap modeling dilakukan menggunakan algoritma Random Forest. Sistem memberikan fasilitas kepada pengguna untuk menentukan parameter model sebelum proses pelatihan dilakukan. Berdasarkan hasil eksperimen, parameter terbaik yang digunakan pada penelitian ini ditunjukkan pada Tabel 3.1.

Tabel 3.1 Tabel Parameter

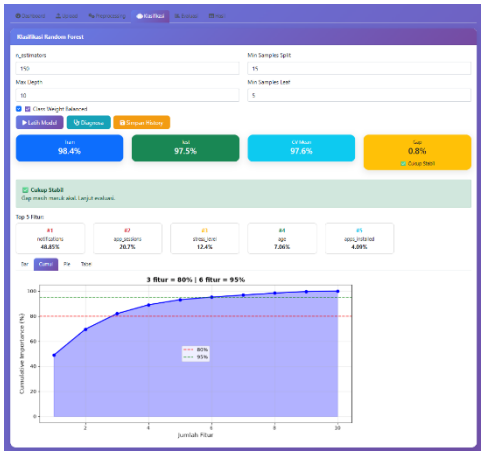
Parameter	**Nilai**
n_estimators	150
max_depth	10
min_samples_split	15
min_samples_leaf	5
class_weight	balanced

Parameter tersebut digunakan untuk menghasilkan model yang stabil dengan kemampuan generalisasi yang baik. Nilai class_weight balanced juga diterapkan untuk menjaga keseimbangan proses klasifikasi apabila distribusi kelas tidak seimbang.Setelah parameter ditentukan, sistem melakukan proses pelatihan menggunakan data training. Hasil pelatihan model ditunjukkan pada Tabel 3.2

Tabel 3.2 Tabel Pengujia

Pengujian	**Nilai**
Training Accuracy	98,4%
Testing Accuracy	97,5%
Cross Validation	97,6%
Gap Accuracy	0,8%

Berdasarkan hasil pelatihan diperoleh akurasi data training sebesar 98,4%, sedangkan akurasi data testing sebesar 97,5%. Nilai rata-rata Cross Validation sebesar 97,6% menunjukkan bahwa model memiliki performa yang konsisten pada beberapa subset data. Selisih akurasi (gap) yang hanya sebesar 0,8% mengindikasikan bahwa model tidak mengalami overfitting dan memiliki kemampuan generalisasi yang baik ketika digunakan pada data baru.



Gambar 3.4 Flowchart Alur Penelitian CRISP-DM

Selain menampilkan hasil pelatihan model, pada Gambar 3.4 sistem juga menghitung nilai Feature Importance untuk mengetahui atribut yang paling berpengaruh terhadap proses klasifikasi. Hasil perhitungan tersebut disajikan pada Gambar.

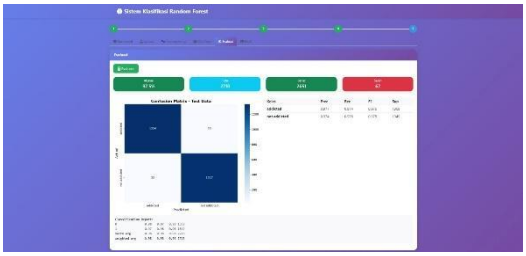
Hasil perhitungan feature importance menunjukkan bahwa atribut notifications memiliki nilai kepentingan sebesar 48,85%, sehingga menjadi atribut yang paling dominan dalam proses klasifikasi. Tingginya nilai tersebut menunjukkan bahwa frekuensi notifikasi yang diterima pengguna merupakan indikator yang paling mampu membedakan antara pengguna yang termasuk kategori addicted dan not addicted. Semakin banyak notifikasi yang diterima, semakin sering pengguna terdorong untuk membuka smartphone, sehingga meningkatkan intensitas interaksi dengan

perangkat. Kondisi tersebut sejalan dengan karakteristik perilaku kecanduan smartphone yang ditandai dengan kecenderungan memeriksa perangkat secara berulang akibat adanya rangsangan dari notifikasi.

Dominannya atribut notifications juga menunjukkan bahwa algoritma Random Forest lebih sering memilih atribut tersebut sebagai pemisah (split attribute) pada banyak decision tree yang dibangun. Hal ini mengindikasikan bahwa atribut tersebut memiliki kemampuan diskriminasi yang lebih baik dibandingkan atribut lainnya dalam memisahkan kedua kelas. Sementara itu, atribut app_sessions, stress_level, age, dan apps_installed tetap memberikan kontribusi terhadap proses klasifikasi, namun tingkat pengaruhnya lebih rendah karena variasi nilainya tidak sekuat atribut notifications dalam membedakan tingkat kecanduan smartphone.

3.3.Implementasi Evaluation

Tahap evaluasi dilakukan menggunakan Confusion Matrix dan Classification Report untuk mengetahui performa model secara lebih rinci. Berdasarkan hasil pengujian terhadap 2.718 data testing, model berhasil mengklasifikasikan 2.651 data dengan benar, sedangkan 67 data mengalami kesalahan klasifikasi sehingga menghasilkan tingkat akurasi sebesar 97,5%.



Gambar 3.5 Flowchart Alur Penelitian CRISP-DM

Berdasarkan Gambar 3.5 Confusion Matrix diketahui bahwa model berhasil mengenali 1.334 data addicted dan 1.317 data not addicted secara benar. Kesalahan klasifikasi hanya terjadi pada 35 data addicted yang diprediksi sebagai not addicted, serta 32 data not addicted yang diprediksi sebagai addicted. Hal tersebut menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik pada kedua kelas.

Berdasarkan hasil Confusion Matrix, terdapat 67 data yang mengalami kesalahan klasifikasi, terdiri atas 35 data addicted yang diprediksi sebagai not addicted (False Negative) dan 32 data not addicted yang diprediksi sebagai addicted (False Positive). Jumlah kesalahan tersebut relatif kecil dibandingkan dengan total 2.718 data pengujian, sehingga model tetap mampu mencapai akurasi sebesar 97,5%.

Kesalahan klasifikasi tersebut kemungkinan disebabkan oleh adanya karakteristik penggunaan smartphone yang memiliki kemiripan antar kelas. Sebagai contoh, beberapa pengguna yang sebenarnya termasuk kategori not addicted memiliki durasi penggunaan, jumlah notifikasi, atau frekuensi penggunaan aplikasi yang cukup tinggi sehingga pola datanya menyerupai pengguna yang mengalami kecanduan. Sebaliknya, terdapat pengguna yang termasuk kategori addicted, namun memiliki nilai pada beberapa atribut yang relatif rendah sehingga model mengelompokkannya ke dalam kelas not addicted. Kondisi ini menunjukkan adanya overlap karakteristik antar kelas yang menyebabkan model sulit menentukan batas klasifikasi secara sempurna.

Selain dipengaruhi oleh kemiripan karakteristik data, kesalahan prediksi juga dapat disebabkan oleh keterbatasan atribut yang tersedia pada dataset. Faktor-faktor lain yang berpotensi memengaruhi tingkat kecanduan smartphone, seperti kondisi psikologis, kebiasaan penggunaan aplikasi tertentu, atau lingkungan sosial pengguna, belum sepenuhnya direpresentasikan dalam dataset. Oleh karena itu, meskipun model Random Forest telah menghasilkan performa yang sangat baik, penambahan atribut yang lebih representatif pada penelitian selanjutnya berpotensi meningkatkan

kemampuan model dalam mengurangi kesalahan klasifikasi.

Nilai Precision, Recall, dan F1-Score pada kedua kelas berada di atas 97%, sehingga menunjukkan bahwa model mampu memberikan performa klasifikasi yang seimbang tanpa menunjukkan kecenderungan terhadap salah satu kelas.

Berdasarkan hasil evaluasi tersebut dapat disimpulkan bahwa algoritma Random Forest mampu menghasilkan performa klasifikasi yang sangat baik dengan tingkat akurasi tinggi dan kesalahan prediksi yang relatif kecil. Selisih akurasi yang rendah antara data training dan testing menunjukkan bahwa model memiliki kemampuan generalisasi yang baik sehingga layak digunakan dalam proses klasifikasi tingkat kecanduan smartphone.

3.4.Implementasi Deployment

Tahap terakhir pada penelitian ini adalah Deployment, yaitu mengimplementasikan model yang telah dibangun ke dalam aplikasi berbasis web menggunakan framework Flask. Melalui implementasi ini, proses klasifikasi tidak lagi dilakukan melalui kode program secara langsung, tetapi dapat dijalankan melalui antarmuka web yang lebih mudah digunakan.

Pada halaman Hasil Klasifikasi, sistem menampilkan hasil prediksi setiap data yang telah diuji. Informasi yang disajikan meliputi kelas aktual, hasil prediksi, status prediksi benar atau salah, serta nilai atribut yang telah melalui proses preprocessing. Selain itu sistem juga menyediakan fitur pencarian data dan ekspor hasil klasifikasi ke dalam format CSV sehingga hasil analisis dapat dimanfaatkan kembali untuk kebutuhan penelitian maupun dokumentasi.

Berdasarkan implementasi yang telah dilakukan, seluruh tahapan metode CRISP-DM berhasil diterapkan ke dalam aplikasi berbasis web. Sistem mampu mengintegrasikan proses pengunggahan dataset, preprocessing, pelatihan model, evaluasi, hingga penyajian hasil klasifikasi dalam satu platform. Dengan tingkat akurasi sebesar 97,5%, sistem yang dikembangkan dapat digunakan sebagai alat bantu untuk mengidentifikasi tingkat kecanduan smartphone secara cepat, konsisten, dan mudah dioperasikan oleh pengguna.

IV. KESIMPULAN.

Penelitian ini berhasil mengembangkan sistem klasifikasi tingkat kecanduan smartphone berbasis web menggunakan algoritma Random Forest dengan menerapkan metode Cross Industry Standard Process for Data Mining (CRISP-DM). Seluruh tahapan CRISP-DM, mulai dari Data Understanding, Data Preparation, Modeling, Evaluation, hingga Deployment, telah berhasil diimplementasikan dalam satu aplikasi berbasis Flask. Sistem yang dikembangkan mampu mengotomatisasi proses klasifikasi sehingga mempermudah pengguna dalam melakukan analisis tingkat kecanduan smartphone secara terintegrasi.

Hasil penelitian menunjukkan bahwa model Random Forest menghasilkan akurasi pengujian sebesar 97,5% dengan nilai Cross Validation sebesar 97,6% dan selisih akurasi sebesar 0,8%, yang menunjukkan kemampuan generalisasi model yang baik. Selain itu, hasil evaluasi memperlihatkan bahwa model mampu mengklasifikasikan data secara konsisten dengan tingkat kesalahan yang rendah, sedangkan analisis feature importance menunjukkan bahwa atribut notifications merupakan faktor yang paling berpengaruh terhadap hasil klasifikasi. Dengan demikian, tujuan penelitian untuk membangun sistem klasifikasi tingkat kecanduan smartphone berbasis web yang akurat dan mudah digunakan telah berhasil tercapai, serta masih berpotensi dikembangkan melalui optimasi model dan penambahan variasi dataset pada penelitian selanjutnya.

REFERENSI

[1] A. Verrel, I. Z. Maulana, V. A. Liu, and A. Prabowo, “Klasifikasi kecanduan smartphone mahasiswa universitas esa unggul menggunakan machine learning dan sas-sv,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 6, no. 3, pp. 585–593, Dec. 2025, doi: 10.37859/coscitech.v6i3.9817.

[2] C. Angkasa, N. P. Dharshinni, and S. Cang Willgert, “PENGARUH SCREEN TIME BERDASARKAN TUJUAN PENGGUNAAN

TERHADAP ADIKSI GADGET MENGGUNAKAN SUPPORT VECTOR MACHINE," 2026. [Online]. Available: https://www.kaggle.com/datasets/sumedh1507/

[3] A. Septiani, A. R. P. Cahyani, H. Damaputra, N. D. Nurfitri, R. Rolastuan, and A. Purnamawati, "Analisis Prediktif Tingkat Kecanduan Penggunaan Smartphone Pada Anak Menggunakan Algoritma C4.5," *Jurnal Komputer Antartika*, vol. 3, no. 4, pp. 165–172, Dec. 2025, doi: 10.70052/jka.v3i4.1246.

[4] F. Salsabilla Nurul Aulia and Z. Fikry, "Hubungan Kesepian dengan Smartphone Addiction Pada Mahasiswa Merantau di Universitas Negeri Padang," *Zulian Fikry INNOVATIVE: Journal Of Social Science Research*, vol. 5, pp. 2967–2989, 2025.

[5] H. Nuramadini *et al.*, "KLASIFIKASI TINGKAT KECANDUAN SMARTPHONE PADA REMAJA MENGGUNAKAN ALGORITMA RANDOM FOREST," 2026. [Online]. Available: https://ikmi.ac.id/page/18/?lang=de

[6] I. Sutoyo, "IMPLEMENTASI ALGORITMA DECISION TREE UNTUK KLASIFIKASI DATA PESERTA DIDIK," vol. 14, no. 2, 2018, [Online]. Available: www.bsi.ac.id

[7] W. Setyanto and D. Y. Wicaksono, "Klasifikasi Tingkat Kecanduan Media Sosial dengan Menggunakan Algoritma Random Forest," 2026.

[8] F. Ayuning Tyas, A. Basir, A. Elistya Ardhin, and P. Korespondensi, "PREDIKSI RESIKO PENGGUNAAN MEDIA SOSIAL TERHADAP KESEHATAN MENTAL MENGGUNAKAN EXPLORATORY DATA ANALYSIS (EDA) DAN CROSS INDUSTRY STANDARD PROCESS FOR DATA MINING (CRISP-DM)," vol. 13, no. 2, pp. 487–498, 2026.

[9] S. Saputra, F. Hidayat, W. Gunawan, I. R. Rahadjeng, and K. Kunci, "KLASIFIKASI DAMPAK KECANDUAN MEDIA SOSIAL MAHASISWA DENGAN SVM DAN K-MEANS," 2026.

[10] D. Maheswari *et al.*, "IMPLEMENTASI ALGORITMA C4.5 UNTUK KLASIFIKASI DAMPAK POLA PENGGUNAAN MEDIA SOSIAL TERHADAP KESEJAHTERAAN EMOSIONAL," 2025.