Jurnal TICOM: Technology of Information and Communication
Volume 14, Nomor 2, Januari 2026 Hal: 68-74 E-ISSN: 2962-7982

Prediksi Tingkat Kecanduan Media Sosial Generasi Z Menggunakan Algoritme Pso dan *Decision Tree*

Anisah Masyuuroh[1], Deni Mahdiana[2], Nidya Kusumawardhany[3]

[1,2,3]Fakultas Teknologi Informasi, Sistem Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia
Jl. Raya Ciledug, Petukangan Urara, Jakarta Selatan, 12260
E-mail: [1]2112500372@student.budiluhur.ac.id, [2]deni.mahdiana@budiluhur.ac.id, [3]nidya.kusumawardhany@budiluhur.ac.id
(*: corresponding author)

***Abstrak*— Tingginya penggunaan media sosial di kalangan Generasi Z dapat menyebabkan kecanduan yang sering dianggap wajar, namun berdampak negatif terhadap kesehatan mental dan produktivitas. Dampak tersebut meliputi gangguan tidur, penurunan fokus belajar maupun bekerja, serta meningkatnya kecemasan akibat perbandingan sosial di dunia maya. Beberapa individu juga mengalami ketergantungan emosional terhadap validasi digital seperti jumlah *like*, komentar, atau interaksi lainnya. Oleh karena itu, deteksi dini terhadap tingkat kecanduan media sosial sangat diperlukan untuk mencegah dampak jangka panjang yang merugikan, terutama bagi Generasi Z yang sangat terhubung secara digital. Penelitian ini bertujuan untuk mengestimasi tingkat kecanduan media sosial pada Generasi Z menggunakan metode *Decision tree*, yang dipilih karena mampu menghasilkan klasifikasi yang akurat serta model yang mudah dipahami. Metode ini bekerja dengan membagi data berdasarkan atribut-atribut relevan untuk mengenali pola perilaku digital secara sistematis. Algoritma *Decision tree* dioptimalkan menggunakan *Particle Swarm Optimization* (PSO) untuk memilih fitur paling berpengaruh dan meningkatkan performa model. Hasil penelitian menunjukkan akurasi sebesar 93,75%, *recall* 94,20%, dan precision 95,16%. Setelah optimasi PSO, akurasi meningkat menjadi 96,42%, *recall* 97,03%, dan *precision* 97,00%, yang menunjukkan peningkatan performa prediksi secara signifikan.**

***Kata Kunci*— *Decision Tree*, Generasi Z, Kecanduan Media Sosial, *Particle Swarm Optimization* (PSO)**

Abstract—***The high use of social media among Generation Z can lead to addiction, which is often considered normal but negatively affects mental health and productivity. These effects include sleep disorders, decreased focus on studying and working, and increased anxiety due to social comparisons in the virtual world. Some individuals also experience emotional dependence on digital validation, such as likes, comments, or other interactions. Therefore, early detection of social media addiction levels is necessary to prevent long-term adverse effects, especially for Generation Z who are highly connected digitally. This study aims to estimate the level of social media addiction among Generation Z using the Decision Tree method, which is chosen for its ability to produce accurate classifications and easy-to-understand models. This method works by dividing data based on relevant attributes to systematically identify digital behavior patterns. The Decision Tree algorithm is optimized using Particle Swarm Optimization (PSO) to select influential features and improve model performance. The results show an accuracy of 93.75%, recall of 94.20%, and precision of 95.16%. After PSO optimization, accuracy increased to 96.42%, recall to 97.03%, and precision to 97.00%, indicating a significant improvement in prediction performance.***

Keyword— ***Decision Tree, Generation Z, Social Media Addiction, Particle Swarm Optimization (PSO)***

I. PENDAHULUAN

Generasi Z mencakup individu yang lahir pada periode 1997–2012 dan tumbuh sejalan dengan kemajuan teknologi informasi dan komunikasi, khususnya internet dan media sosial [1]. Tumbuh besar berdampingan dengan teknologi menjadikan Generasi Z memiliki ketergantungan yang tinggi terhadap dunia virtual. Media sosial kini tidak lagi dipandang secara sempit sebagai kanal interaksi sosial, melainkan telah diadopsi sebagai platform pembelajaran dan ekspresi identitas yang krusial. Berdasarkan data statistik dari laporan Digital 2025: Indonesia, terlihat adanya lonjakan signifikan dalam demografi pengguna digital; diperkirakan terdapat sekitar 126 juta penduduk dewasa (18 tahun ke atas) yang aktif di media sosial pada awal tahun 2025. Persentase ini, yang mencakup 62,7% dari populasi dewasa, menegaskan bahwa lanskap informasi di Indonesia saat ini sangat didominasi oleh interaksi di platform digital [2]. Secara demografis, Generasi Z memegang porsi signifikan dalam struktur pengguna internet di tanah air, yakni mencapai 34,40% menurut catatan APJII. Fenomena ini diikuti dengan pola konsumsi konten yang sangat intensif; selaras dengan informasi yang disampaikan oleh Menkominfo pada kegiatan #YukPahamiPemilu 2023, rata-rata konsumsi waktu Generasi Z di platform media sosial adalah 6 jam per hari. Angka tersebut mencerminkan bahwa mayoritas waktu aktif harian generasi ini tercurahkan pada interaksi digital, yang secara langsung berimplikasi pada bagaimana mereka menerima dan mengolah pesan-pesan sosial maupun politik di internet [3]. Penggunaan media sosial yang intensif ini berpotensi menyebabkan kecanduan, yang dapat memengaruhi kesehatan mental dan produktivitas pengguna, terutama pada kalangan remaja [4]. Penggunaan suatu platform yang lebih dari 5 jam sehari dapat dianggap sebuah perilaku problematik [5].

Dalam upaya menganalisis kecanduan media sosial, beragam algoritma data mining telah digunakan, antara lain Multiple Logistic Regression [6], *Decision tree* [7], Support Vector Machine (SVM) [8], Naive Bayes [9], K-Means

Clustering [10], *Random Forest* [11], Extreme Gradient Boosting (XGBoost) [12].

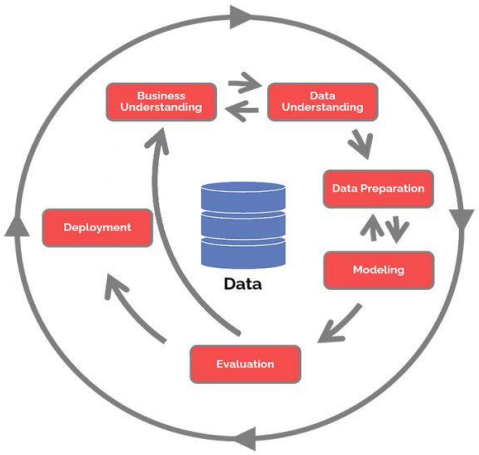
Algoritma yang digunakan pada penelitian-penelitian sebelumnya menunjukkan berbagai kelebihan dan kekurangan. Metode Multiple Logistic Regression dapat mengsevaluasi dampak sejumlah faktor terhadap suatu hasil sekaligus serta memberikan estimasi peluang yang lebih terperinci. Namun, teknik ini memerlukan data yang sepenuhnya memenuhi asumsi statistik tertentu agar mendapatkan hasil yang akurat [6]. Meskipun SVM dapat digunakan untuk mengklasifikasikan tingkat stres berdasarkan aktivitas media sosial, akurasi yang dihasilkan masih tergolong rendah walaupun telah menggunakan metode seleksi fitur Information Gain [8]. Metode Naive Bayes mampu mencapai akurasi rata-rata 93% dalam mengklasifikasikan tingkat kecanduan internet di kalangan remaja. Namun, metode ini memerlukan transformasi data menjadi nilai numerik, yang berisiko menimbulkan kesalahan jika tidak dilakukan dengan hati-hati [9], Meskipun metode K-Means efektif dalam mengelompokkan data, metode ini memerlukan penentuan jumlah cluster sebelumnya, yang dapat mempengaruhi keakuratan hasil jika jumlah yang optimal tidak diketahui [10]. Metode *Random Forest* dapat mengklasifikasikan dengan tingkat akurasi yang baik, namun harus memilih data secara hati-hati karena variabel penting yang tidak sengaja terlewatkan dapat mengganggu keefektifan model [11]. Meskipun nilai F-Measure tidak ideal, menunjukkan bahwa masih ada peluang untuk peningkatan akurasi model, pendekatan XGBoost dapat mengidentifikasi kecanduan internet dan media sosial dengan hasil yang cukup baik [12].

Penelitian ini menggunakan algoritma *Decision tree* karena telah terbukti lebih efisien dalam klasifikasi, seperti dalam penelitian yang dilakukan [7], algoritma *Decision tree* menghasilkan tingkat akurasi lebih tinggi, yaitu 86,67%, dibandingkan dengan algoritma Naive bayes yang hanya mencapai 70,00% dalam mengklasifikasikan data mengenai pengaruh media sosial. Penelitian ini mengimplementasikan algoritma Particle Swarm Optimization (PSO) sebagai instrumen optimasi untuk memperkuat kinerja pemodelan [13]. Pendekatan ini menggabungkan kekuatan *Decision tree* dengan efisiensi pencarian parameter dari PSO untuk menganalisis dan memprediksi kecenderungan kecanduan media sosial di kalangan Generasi Z. Tujuan fundamental dari penggabungan metode ini adalah untuk memperoleh hasil analisis yang lebih akurat dan faktual terkait faktor-faktor yang memengaruhi perilaku digital tersebut, sehingga model yang dihasilkan dapat menjadi acuan yang valid dalam memahami fenomena adiksi digital di masa kini.

II. METODE PENELITIAN

A. CRISP-DM

Dalam mengonstruksi alur penelitian ini, metode Cross Industry Standard for Data Mining (CRISP-DM) diadopsi sebagai kerangka acuan utama yang memandu setiap fase perancangan hingga implementasi [14]. Metodologi ini menyediakan struktur kerja yang komprehensif melalui enam tahapan esensial, yang dimulai dari pemahaman mendalam terhadap konteks masalah (business understanding) dan karakteristik data (data understanding). Proses kemudian dilanjutkan dengan fase persiapan data (data *preparation*), pengembangan arsitektur model (*modeling*), pengujian kualitas melalui evaluasi (*evaluation*), hingga tahap akhir berupa implementasi hasil atau penyebaran (*deployment*).



Gambar 1. CRIPS-DM

Berdasarkan Gambar 1, berikut penjelasan setiap tahap penelitian:

1) *Business Understanding*

Tahap pemahaman data menuntut adanya penguasaan terhadap hakikat kegiatan penggalian data yang bertujuan untuk mengonstruksi model prediksi dengan akurasi tertinggi. Sasaran utama dalam proses ini adalah memetakan tingkat kecanduan media sosial melalui pendekatan klasifikasi yang presisi. Penelitian ini menitikberatkan pada dua persoalan mendasar: pertama, mengidentifikasi variabel-variabel yang memberikan pengaruh dominan terhadap perilaku adiktif Generasi Z di ruang siber; dan kedua, melakukan evaluasi mendalam terhadap kinerja fungsional algoritma *Decision tree*. Melalui pemahaman yang matang pada tahap ini, diharapkan model yang dihasilkan mampu menjawab problematika kecanduan digital secara faktual.

2) *Data Understanding*

Pada tahap *Data Understanding*, data yang digunakan dalam penelitian ini diperoleh dari situs Kaggle, yaitu *Social Media Addiction Dataset* yang bersifat publik dan dapat diakses melalui tautan https://www.kaggle.com/code/necipardakocaba/women-vs-men-social-media-addiction/notebook. Dataset tersebut terdiri dari 1.000 entri data dan 31 atribut, dengan rincian atribut disajikan pada Tabel 1.

TABEL I
ATRIBUT DALAM DATASET

Nama Atribut	Tipe Atribut
*User*ID	*Numeric*
Age	*Numeric*
Gender	*Categorical*
Location	*Categorical*

Nama Atribut	Tipe Atribut
Income	*Numeric*
Debt	*Categorical*
Owns_Property	*Categorical*
Profession	*Categorical*
Demographics	*Categorical*
Platform	*Categorical*
Total_Time_Spent	*Numeric*
Number_of_Session	*Numeric*
Video_ID	*Numeric*
Video_Category	*Categorical*
Video_Length	*Numeric*
Engagement	*Numeric*
Importance_Score	*Numeric*
Time_Spent_on_Video	*Numeric*
Number_of_Videos_Watched	*Numeric*
Scroll_Rate	*Numeric*
Frequency	*Categorical*
Productivity_Loss	*Numeric*
Satisfaction	*Numeric*
Watch_Reason	*Categorical*
Device_Type	*Categorical*
OS	*Categorical*
Watch_Time	*Time*
Self_Control	*Numeric*
Addiction_Level	*Numeric*
Current_Activity	*Categorical*
Connection_Type	*Categorical*

3) *Data Preparation*

Tahap persiapan data dalam penelitian ini menunjukkan bahwa data publik yang terkumpul sudah berada dalam kondisi siap olah tanpa adanya nilai yang hilang (*missing value*). Dengan demikian, tahapan pra-pemrosesan tidak lagi melibatkan manipulasi pengisian data yang kosong. Fokus utama pada fase ini dialihkan ke langkah-langkah optimalisasi data lainnya guna menunjang kinerja algoritma. Beberapa tindakan krusial yang diimplementasikan dalam proses persiapan data ini mencakup antara lain:

a) Menetapkan label klasifikasi dengan menggunakan *operator Set Role* untuk menetapkan atribut *Addiction_Level* sebagai label.
b) Melakukan reduksi data dengan menghapus dua atribut, yaitu *user_id* dan *video_id*, karena keduanya bersifat unik pada setiap baris data dan tidak berkontribusi terhadap proses klasifikasi.
c) Melakukan proses penyaringan data menggunakan *operator Filter Examples* pada *RapidMiner* untuk memilih data dengan rentang usia 13–28 tahun, sesuai dengan kategori Generasi Z. Selain itu, data dengan kombinasi atribut usia dan profesi yang tidak relevan seperti profesi manager pada usia 18–24 tahun juga dihapus karena dianggap tidak sesuai dengan konteks penelitian.
d) Mengelompokkan label *Addiction_Level* menjadi lima kategori, yaitu Sangat Kurang Kecanduan, Kurang Kecanduan, Cukup Kecanduan, Kecanduan, dan Sangat Kecanduan, menggunakan *operator Generate Attributes*.
e) Melakukan transformasi data dengan mengubah tipe data dari numerik menjadi kategorikal pada beberapa atribut, antara lain: *Importance_Score, Self_Control, Satisfaction, Scroll_Rate, Productivity_Loss, Time_Spent_On_Video, Total_Time_Spent, Number_of_Sessions,* dan *Number_of_Videos_Watched*, melalui proses *discretization* menggunakan *operator Generate Attributes*.
f) Proses pembagian data ke dalam data latih dan data uji dilakukan dengan menggunakan metode *10-Fold Cross Validation* untuk memastikan evaluasi model yang komprehensif dan mengurangi bias terhadap data tertentu.

4) *Evaluation*

Pada tahap *evaluation*, dilakukan pengukuran kinerja algoritma model yang digunakan untuk mengklasifikasikan tingkat kecanduan media sosial pada Generasi Z. Tahap *evaluation* dilakukan dengan teknik *confusion matrix* untuk mengukur kinerja model berdasarkan metrik seperti *accuracy*, *precision*, dan *recall*. Tahap *evaluation* turut mencakup penilaian kritis mengenai sejauh mana luaran klasifikasi dapat berkontribusi pada pencapaian tujuan penelitian, khususnya dalam menyediakan data yang aplikatif untuk langkah preventif perilaku digital. Dengan demikian, proses seleksi model mengedepankan aspek fungsionalitas; model terbaik bukan sekadar yang unggul secara performa teknis, namun yang paling mampu mengonstruksi solusi relevan bagi fenomena kecanduan media sosial. Hal ini menjamin bahwa hasil akhir penelitian memiliki nilai guna yang tinggi, melampaui sekadar angka akurasi di atas kertas.

5) *Deployment*

Tahap *deployment* menandai penerapan model klasifikasi dalam skenario nyata guna mengidentifikasi kecenderungan perilaku digital Generasi Z berdasarkan variabel-variabel yang telah ditentukan. Hasil analisis ini diharapkan menjadi referensi krusial bagi tenaga ahli dan pembuat kebijakan untuk merumuskan strategi intervensi yang mampu mendorong budaya digital yang lebih bijak serta produktif. Dengan ketersediaan informasi yang mendalam, berbagai pihak terkait dapat mengembangkan program edukatif yang terukur untuk mencegah eskalasi kecanduan media sosial, sehingga solusi yang ditawarkan menjadi lebih terarah dan berdampak luas.

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Dalam penelitian ini digunakan data sekunder yang diperoleh dari situs *www.kaggle.com*, yaitu dataset dengan judul *Social Media Addiction*. Dataset awal terdiri dari 1.000 *record* yang telah bersih dan tidak mengandung *missing value*. Secara keseluruhan, dataset memiliki 31 atribut. Namun, 2 atribut tidak digunakan. Sebanyak 28 atribut dimanfaatkan sebagai regular atribut untuk menggambarkan secara menyeluruh faktor-faktor yang memengaruhi kecanduan media sosial pada Generasi Z, dan 1 atribut lainnya digunakan sebagai spesial atribut.

Setiap atribut dalam dataset ini memiliki makna yang jelas dalam mengukur kecanduan media sosial. Atribut seperti *Age, Gender, Location, Profession, serta Income, Debt, dan Owns_Property* mencerminkan latar belakang pengguna yang dapat memengaruhi bagaimana mereka menggunakan media

sosial. *Demographics* dan *Platform* menunjukkan demografi pengguna dan media sosial yang digunakan. Atribut waktu seperti *Total_Time_Spent, Number_of_Session, Watch_Time,* dan *Time_Spent_On_Video* adalah atribut yang menunjukkan intensitas penggunaan. Atribut *Video_Category, Video_Length, Engagement, Number_Of_Videos_Watched, Scroll_Rate,* dan *Frequency* menunjukkan jumlah konten yang dikonsumsi. *Importance_Score* dan *Satisfaction* menunjukkan persepsi pengguna. Variabel target seperti *Productivity_Loss, Self_Control, dan Addiction_Level* menunjukkan dampak negatif. Untuk konteks tambahan, atribut seperti *Watch_Reason, Current_Activity, Device_Type,* dan Connection_Type disertakan. UserID dan Video_ID hanya digunakan sebagai identifikasi administratif dan tidak digunakan untuk analisis karena tidak relevan untuk pengukuran kecanduan.

B. Data *Prepocessing*

Tahap data preprocessing dilakukan setelah seluruh data berhasil dikumpulkan. Pada tahap ini, data dianalisis dan dipersiapkan untuk mendukung proses klasifikasi serta mencapai tujuan penelitian. Langkah pertama dalam preprocessing adalah tahap reduksi data, yaitu menghapus atribut-atribut yang tidak relevan serta data yang tidak sesuai dengan kebutuhan penelitian. Karena dataset yang digunakan telah bersih dan tidak mengandung missing value, maka pada tahap data cleaning hanya dilakukan proses filter data berdasarkan kriteria generasi Z. Setelah itu, dilakukan tahap transformasi data agar dataset dapat digunakan secara optimal dalam proses pemodelan menggunakan algoritma *Decision tree*.

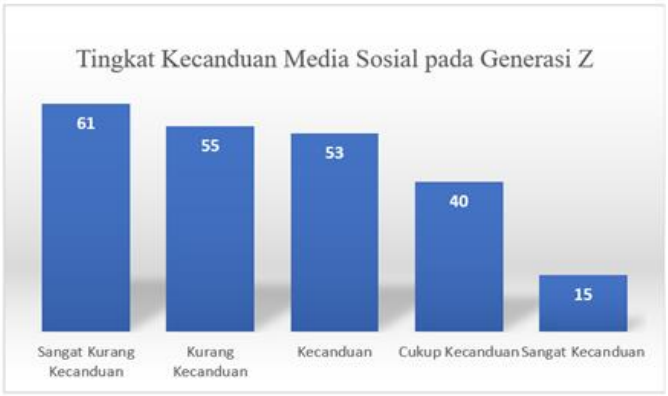
1) Reduksi Data

Dalam proses reduksi data, dua atribut dihapus dari dataset, yaitu *user_id* dan *video_id*. Keduanya tidak digunakan dalam analisis karena memiliki nilai yang unik pada setiap baris data, sehingga bukan variabel yang dapat memengaruhi proses klasifikasi. Hal ini juga berisiko membuat algoritma hanya "menghafal" data, bukan benar-benar mempelajari pola. Awalnya, dataset terdiri dari 31 atribut, dan setelah proses reduksi, jumlahnya menjadi 29 atribut. Selain itu, data dengan profesi manager pada kelompok usia 18–24 tahun juga dihapus, karena dianggap tidak relevan dan berpotensi menurunkan validitas hasil analisis. Langkah ini dilakukan untuk memastikan bahwa data yang dianalisis lebih sederhana dan hanya mencakup informasi yang sesuai dengan konteks karakteristik Generasi Z serta relevan terhadap prediksi tingkat kecanduan media sosial.

2) *Data Cleaning*

Tahapan *data cleaning* dilakukan untuk menyaring data yang sesuai dan menghapus atribut yang tidak berkontribusi terhadap analisis. Dataset awal terdiri dari 1.000 *record* dengan rentang usia yang bervariasi. Oleh karena itu, digunakan *operator Filter Examples* pada *RapidMiner* untuk memilih hanya pengguna berusia 13–28 tahun, sesuai dengan kategori Generasi Z. Setelah dilakukan proses *filter*, data yang tersisa berjumlah 224 *record*. Dengan demikian, dataset menjadi lebih fokus dan sesuai dengan tujuan penelitian. Visualisasi komposisi data berdasarkan label *Addiction_Level* disajikan dalam bentuk diagram pada

Gambar 2 menunjukkan distribusi masing-masing kategori *Addiction_Level* dalam dataset yang telah difilter. Kategori Sangat Kurang Kecanduan memiliki jumlah terbanyak, yaitu 61 data, dilanjutkan oleh Kurang Kecanduan sebanyak 55 data, dan Kecanduan sebanyak 53 data. Sementara itu, Cukup Kecanduan berjumlah 40 data dan Sangat Kecanduan merupakan kategori dengan jumlah paling sedikit, yaitu 15 data.



Gambar 2. Komposisi Data berdasarkan Label Addiction_Level

3) *Data Transformation*

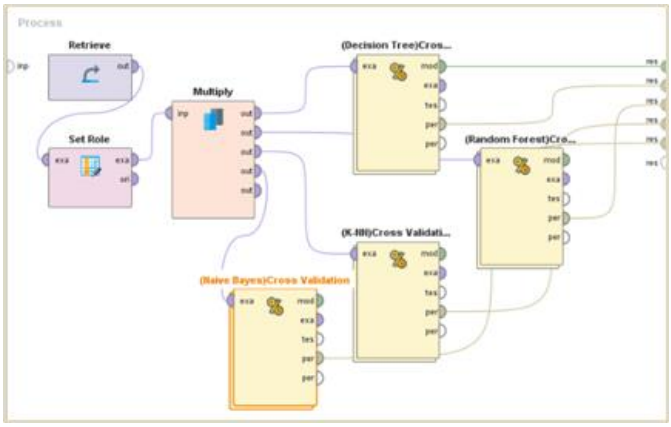
Tahapan *data transformation* dilakukan untuk menyesuaikan format data agar sesuai dengan kebutuhan analisis, khususnya untuk algoritma klasifikasi yang digunakan. Menurut [15], pendekatan ini dapat meningkatkan interpretabilitas model dan sesuai untuk algoritma seperti *Decision tree* yang lebih efektif bekerja dengan data kategorikal. Pertama, dilakukan proses pelabelan data dengan menetapkan atribut *Addiction_Level* sebagai label menggunakan *operator Set Role* pada *RapidMiner*. Selanjutnya, label tersebut dikelompokkan menjadi lima kategori menggunakan *operator Generate Attributes*.

Transformasi lainnya meliputi perubahan tipe data dari *numeric* menjadi *categorical* pada beberapa atribut seperti *Importance_Score, Self_Control, Satisfaction, Scroll_Rate, Productivity_Loss, Time_Spent_On_Video, Total_Time_Spent, Number_of_Sessions,* dan *Number_of_Videos_Watched*. Menurut [15], transformasi numerik ke kategorikal melalui proses discretization dapat meningkatkan interpretabilitas dan mengurangi kompleksitas pada algoritma *Decision tree*.

C. Hasil Data Latih dan Data Uji

Data yang telah melalui proses pembersihan kemudian diolah menggunakan platform *RapidMiner* dengan menerapkan teknik *10-Fold Cross Validation*. Metode ini membagi dataset secara merata menjadi sepuluh sub-set, di mana setiap sub-set berfungsi sebagai data pengujian secara bergantian, sedangkan sisanya digunakan untuk melatih model. Strategi ini diimplementasikan untuk menjamin objektivitas evaluasi dan memastikan model memiliki generalisasi yang baik. Peneliti melakukan pengujian terhadap empat algoritma klasifikasi, yaitu *Decision tree, Naïve Bayes, K-Nearest Neighbor (K-NN),*

dan *Random Forest*. Guna mendapatkan perbandingan performa yang akurat dan *apple-to-apple*, setiap algoritma diproses melalui struktur validasi yang seragam. Alur sistematis dari tahapan pengujian ini secara visual dipaparkan pada Gambar 3.



Gambar 3. Proses K-Fold Cross Validation pada RapidMiner

Pengujian algoritma dilakukan dengan metode *Cross Validation* yang dikombinasikan dengan *operator Apply Model* dan *Performance* untuk evaluasi menggunakan *Confusion Matrix*. Hasil pengujian menunjukkan bahwa algoritma *Decision tree* memperoleh akurasi sebesar 93,75%, *Random Forest* sebesar 91,90%, K-NN sebesar 64,33%, dan *Naïve Bayes* sebesar 90,12%. Hasil pengujian menunjukkan bahwa dalam struktur validasi silang sepuluh lipat, algoritma *Decision tree* berhasil mengungguli kandidat algoritma lainnya dalam aspek akurasi. Temuan ini menjadi landasan bagi peneliti untuk menindaklanjuti model *Decision tree* ke dalam fase optimasi dengan mengimplementasikan teknik meta-heuristik *Particle Swarm Optimization* (PSO). Langkah penggabungan ini dilakukan untuk memaksimalisasi parameter model sehingga diperoleh luaran klasifikasi yang lebih tajam dan akurat. Untuk meninjau perbandingan nilai akurasi secara mendetail dari setiap algoritma yang terlibat, informasi lengkap telah dirangkum dalam Tabel 2.

TABEL II
KOMPILASI ALGORITMA DENGAN 10-FOLD CROSS VALIDATION

Metode Validasi	Algortima	Akurasi (%)
10-Fold Cross Validation	***Decision tree***	***93.75%***
	Random Forest	91.90%
	K-NN	64.33%
	Naïve Bayes	90.12%

1) Pemodelan

Pada tahap ini, dilakukan proses pemodelan untuk menganalisis dan mengukur performa dari beberapa algoritma klasifikasi terhadap dataset yang digunakan. Pemodelan dilakukan berdasarkan hasil pengujian menggunakan metode *10-Fold Cross Validation*. Beberapa algoritma yang digunakan dalam proses ini antara lain *Decision tree, Naïve Bayes, K-Nearest Neighbor (K-NN),* dan *Random Forest*. Melalui proses ini, model dari masing-masing algoritma dibangun berdasarkan data latih, kemudian diuji untuk melihat seberapa baik model tersebut mampu melakukan klasifikasi. Hasil akurasi dari setiap algoritma dibandingkan untuk menentukan model dengan performa terbaik.

2) Hasil Komparasi Model

Evaluasi terhadap efektivitas algoritma *Decision tree, Naïve Bayes, K-NN,* dan *Random Forest* dilakukan melalui pendekatan multidimensional. Selain bersandar pada tingkat akurasi yang dihasilkan dari validasi silang sepuluh lipat, peneliti juga menelaah nilai *precision* dan *recall* guna memperoleh gambaran menyeluruh mengenai reliabilitas model. Melalui peninjauan metrik yang beragam ini, kualitas klasifikasi dapat dianalisis secara lebih presisi, memastikan bahwa model yang terpilih tidak hanya unggul secara kuantitas prediksi benar, tetapi juga memiliki keseimbangan performa pada setiap kategori data.

Dalam mengukur sejauh mana sebuah model mampu meminimalisir luputnya identifikasi pada kelas positif, nilai *recall* menjadi indikator krusial. Temuan penelitian menunjukkan adanya variasi performa yang signifikan: *Decision tree* memimpin dengan persentase 94,20%, disusul oleh *Random Forest* dengan 91,90%, dan *Naïve Bayes* sebesar 89,17%. Di sisi lain, *K-NN* menunjukkan keterbatasan dalam mengenali data positif dengan nilai *recall* sebesar 66,92%. Hasil ini menegaskan bahwa *Decision tree* merupakan algoritma yang paling reliabel untuk memastikan sebagian besar data kelas positif dapat terklasifikasikan dengan tepat. Selanjutnya, nilai *precision* dievaluasi untuk melihat ketepatan model dalam memberikan prediksi positif terhadap data uji. Hasil pengujian menunjukkan bahwa algoritma *Decision tree* memperoleh *precision* sebesar 95,16%, *Random Forest* sebesar 92,85%, K-NN sebesar 64,10%, dan *Naïve Bayes* sebesar 92,30%.

Berdasarkan hasil evaluasi dengan menggunakan akurasi, *precision*, dan recall, diperoleh perbandingan performa dari masing-masing algoritma. Ringkasan hasil komparasi ditampilkan pada Tabel 3, yang menunjukkan bahwa algoritma Decision tree memiliki performa terbaik dengan akurasi 93,75%, recall 94,20%, dan *precision* 95,16%.

TABEL III
KOMPILASI HASIL SETIAP ALGORITMA

Algortima	Akurasi (%)	*Recall* (%)	*Precision* (%)
Decision tree	***93.75%***	**94.20%**	**95.16%**
Random Forest	91.90%	91.90%	92.85%
K-NN	64.33%	66.92%	64.10%
Naïve Bayes	90.12%	89.17%	92.30%

3) Optimasi Model Terbaik dengan PSO

Temuan pada tahap pengujian awal memposisikan *Decision tree* sebagai algoritma dengan performa terbaik, yang kemudian menjadi landasan bagi peneliti untuk melanjutkan ke fase optimasi. Pada tahap ini, metode *Particle Swarm Optimization* (PSO) diintegrasikan untuk memperkuat struktur model. Fokus utama dari penggunaan PSO adalah untuk mengeksplorasi kombinasi parameter yang ideal secara otomatis, sehingga efisiensi dan akurasi model dalam memprediksi tingkat kecanduan media sosial dapat ditingkatkan secara signifikan dibandingkan dengan model *Decision tree* standar.

Setelah proses optimasi, model *Decision tree* menunjukkan peningkatan kinerja dengan akurasi 96,42%, nilai *recall* 97,03%, dan *precision* 97,00%. Perbandingan performa model sebelum dan sesudah optimasi ditampilkan pada Tabel 4, yang memperlihatkan bahwa penerapan PSO mampu meningkatkan kemampuan prediksi model.

TABEL IV
DECISIONS TREE PRA DAN PASCA OPTIMASI

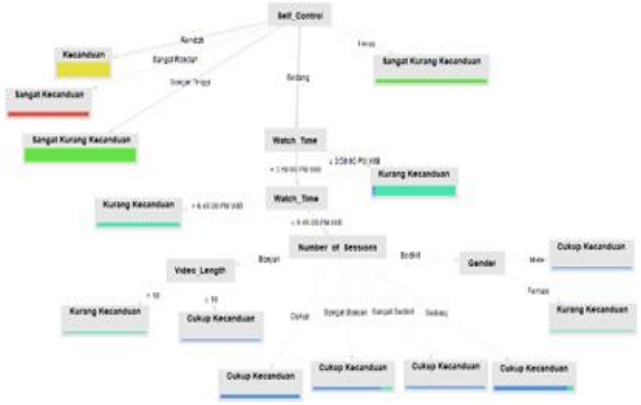
Metode	**Akurasi (%)**	***Recall* (%)**	***Precision* (%)**
10-Fold Cross Validation	*93.75%*	94.20%	95.16%
PSO + *10-Fold Cross Validation*	**96.42%**	**97.03%**	**97.00%**

Selain meningkatkan akurasi, proses PSO juga menghasilkan bobot setiap atribut yang menggambarkan kontribusinya dalam prediksi. Hasil optimasi menunjukkan bahwa atribut dengan bobot tertinggi, seperti *Location, Debt, Demographics, Platform, Scroll_Rate, Frequency, Device_Type, OS, Watch_Time*, dan *Self_Control*, menjadi faktor utama yang memengaruhi tingkat kecanduan media sosial pada Generasi Z. Sementara itu, atribut lain seperti *Age, Video_Length, Connection_Type*, dan *Time_Spent_On_Video* tetap memberikan pengaruh meskipun lebih kecil. Temuan ini menegaskan bahwa optimasi tidak hanya meningkatkan performa model, tetapi juga membantu mengidentifikasi faktor-faktor yang paling relevan dalam prediksi.

4) Penyajian Model Terbaik

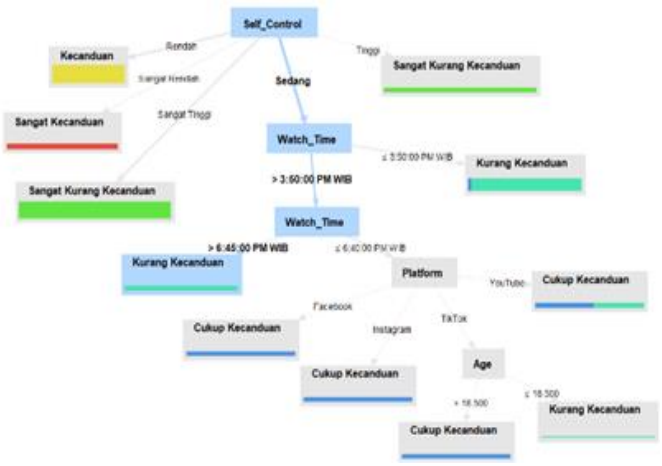
Berdasarkan hasil evaluasi, algoritma *Decision tree* dipilih sebagai model terbaik dan dilanjutkan dengan optimasi menggunakan *Particle Swarm Optimization* (PSO). Penyajian model dilakukan dalam dua tahap, yaitu sebelum dan sesudah optimasi.

Pada Gambar 4 sebagai model awal sebelum optimasi, struktur pohon keputusan dibangun menggunakan algoritma *Decision tree* dengan *10-Fold Cross Validation*. Model ini menjadikan atribut *Self_Control* sebagai akar pohon, kemudian bercabang ke atribut lain seperti *Watch_Time, Number_of_Sessions, Video_Length*, dan *Gender* untuk menghasilkan klasifikasi tingkat kecanduan. Struktur pohon yang dihasilkan masih cukup kompleks dengan banyak percabangan.



Gambar 4. Struktur Pohon *Decision Tree* Sebelum Optimasi

Setelah dilakukan optimasi dengan PSO, jumlah atribut prediktor berkurang dari 28 menjadi 16 atribut yang paling relevan. Pohon keputusan hasil optimasi tetap menjadikan Self_Control sebagai akar pohon, tetapi jalur klasifikasi menjadi lebih sederhana dan terfokus. Atribut Platform dan Age muncul sebagai percabangan dominan setelah Self_Control dan Watch_Time, menggantikan atribut lain yang sebelumnya berperan besar terlihat pada Gambar 5.



Gambar 5. Struktur Pohon *Decision Tree* Hasil Optimasi

Perbandingan ini menunjukkan bahwa proses optimasi berhasil menyederhanakan struktur pohon serta meningkatkan akurasi model. Dengan atribut yang lebih relevan, model tidak hanya lebih efisien, tetapi juga lebih mudah diinterpretasikan untuk memprediksi tingkat kecanduan media sosial pada Generasi Z.

5) Perhitungan dengan Algoritma Terpilih

Guna menguji akurasi model yang telah dioptimasi, dilakukan perhitungan manual pada sepuluh sampel data menggunakan kerangka kerja *Decision tree* hasil optimasi PSO. Pemilihan sampel tersebut dilakukan secara terencana melalui metode *stratified sampling*, yang dioperasikan untuk menjaga integritas distribusi data agar tetap mencerminkan komposisi kelas pada dataset aslinya. Langkah ini diambil guna memastikan bahwa hasil pengujian manual tetap objektif dan mampu memvalidasi kinerja model dalam mengklasifikasikan tingkat kecanduan media sosial sesuai dengan karakteristik distribusi data yang ada.

Langkah pertama adalah menghitung *entropy dataset* untuk mengetahui tingkat ketidakpastian dari distribusi kelas pada data sampel. Perhitungan dilakukan menggunakan Rumus 1:

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (1)$$

Berdasarkan distribusi kelas, yaitu 3 *record* Sangat Kurang Kecanduan, 2 *record* Kurang Kecanduan, 2 *record* Cukup Kecanduan, 2 *record* Kecanduan, dan 1 *record* Sangat Kecanduan, diperoleh nilai *entropy* sebesar 2,246 *bits*.

Tahap berikutnya adalah menghitung *conditional entropy* untuk masing-masing atribut pembentuk pohon keputusan,

yaitu *Self_Control*, *Watch_Time*, *Platform*, dan *Age*. Perhitungan dilakukan dengan menggunakan Rumus 2:

$$Info(D) = -\sum_{i=1}^{m} p_i \log_2(p_i) \quad (2)$$

Hasil perhitungan menunjukkan bahwa nilai *entropy* bersyarat untuk *Self_Control* adalah 0,400, *Watch_Time* sebesar 1,151, *Platform* sebesar 1,351, dan *Age* sebesar 0,875. Nilai-nilai ini selanjutnya digunakan untuk menghitung *information gain* dari setiap atribut. Perhitungan *information gain* dilakukan menggunakan Rumus 3 :

$$Gain(A) = Info(D) - Info_A(D) \quad (3)$$

Berdasarkan hasil perhitungan, diperoleh nilai gain sebesar 1,846 untuk Self_Control, 1,371 untuk Age, 1,095 untuk Watch_Time, dan 0,895 untuk Platform. Dari keempat atribut tersebut, Self_Control memiliki nilai gain tertinggi sehingga dipilih sebagai akar pohon keputusan karena mampu memberikan pemisahan data paling optimal. Ringkasan hasil perhitungan entropy, conditional entropy, dan information gain ditunjukkan pada Tabel 5.

TABEL V
DECISIONS TREE PRA PRA DAN PASCA OPTIMASI

Atribut	**Entropy Awal**	***Conditional Entropy***	***Information Gain***
Self_Control	2,246	0.400	1,846
Watch_Time		1,151	1,095
Platform		1,351	0,895
Age		0,875	1,371

IV. PENUTUP

Berdasarkan hasil penelitian, penerapan algoritma PSO terbukti mampu mengidentifikasi faktor-faktor yang mempengaruhi tingkat kecanduan media sosial pada generasi Z, seperti *Location, Debt, Demographics, Platform, Scroll Rate, Frequency, Device Type, OS, Watch Time, Self Control, Age, Video Length, Connection Type, Time Spent on Video, Income,* dan *Profession*. Selain itu, penerapan algoritma *Decision tree* terbukti dapat memprediksi kecanduan media sosial dengan kinerja yang baik, ditunjukkan oleh akurasi sebesar 93,75%, *recall* 94,20%, dan *precision* 95,16%. Lebih lanjut, penerapan kombinasi algoritma PSO dan *Decision tree* dapat meningkatkan performa model sehingga menghasilkan akurasi yang lebih tinggi, yaitu 96,42%, dibandingkan penggunaan algoritma klasifikasi tunggal.

Dengan hasil ini, disarankan agar penelitian selanjutnya menguji model pada dataset yang lebih luas dan bervariasi untuk memastikan generalisasi model, serta mempertimbangkan metode optimasi lain sebagai pembanding agar diperoleh hasil yang lebih komprehensif.

REFERENSI

[1] L. S. Arum, A. Zahrani, and N. A. Duha, "Karakteristik Generasi Z dan Kesiapannya dalam Menghadapi Bonus Demografi 2030," *Account. Student Res. J.*, vol. 2, no. 1, pp. 59–72, 2023, doi: 10.62108/asrj.v2i1.5812.

[2] S. Kemp, "Digital 2025: Indonesia." [Online]. Available: https://datareportal.com/reports/digital-2025-indonesia

[3] Tempo.co, "Generasi Milenial dan Z Bermedia Sosial 6 Jam Per Hari, Budi Arie: Upayakan Pemilu Damai." [Online] Available: https://www.tempo.co/digital/generasi-milenial-dan-z-bermedia-sosial-6-jam-per-hari-budi-arie-upayakan-pemilu-damai-141312

[4] S. Pappa, H. Pratikto, and A. R. Aristawati, "Leisure Boredom dan Kecenderungan Kecanduan Media Sosial TikTok pada," *Jiwa J. Psikol. Indones.*, vol. 3, no. 02, pp. 88–94, 2024.

[5] N. Unair, "Kecanduan Media Sosial Termasuk Gangguan Mental? Simak Faktanya." [Online]. Available: https://unair.ac.id/kecanduan-media-sosial-termasuk-gangguan-mental-simak-faktanya/

[6] E. Ernawati, "Dampak Kecanduan Media Sosial Terhadap Kesehatan Mental Remaja: Studi Cross Sectional," *Intan Husada J. Ilm. Keperawatan*, vol. 12, no. 01, pp. 78–92, 2024, doi: 10.52236/ih.v12i1.507.

[7] A.- Husaini, I. Hariyanti, and A. R. Raharja, "Perbandingan Algoritma Decision Tree dan Naive Bayes dalam Klasifikasi Data Pengaruh Media Sosial dan Jam Tidur Terhadap Prestasi Akademik Siswa," *Technol. J. Ilm.*, vol. 15, no. 2, pp. 332-340, 2024, doi: 10.31602/tji.v15i2.14381.

[8] J. R. Wiyani, I. Indriati, and S. Sutrisno, "Klasifikasi Stres berdasarkan Unggahan pada Media Sosial Twitter menggunakan Metode Support Vector Machine dan Seleksi Fitur Information Gain," 2022. [Online] Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/12060

[9] M. I. Fikri, E. Budianita, I. Iskandar, and E. P. Cynthia, "Klasifikasi Tingkat Kecanduan Internet Terhadap Remaja Pekanbaru Melalui Pendekatan Algoritma Naïve Bayes," *Zo. J. Sist. Inf.*, vol. 6, no. 2, pp. 424–436, 2024, doi: 10.31849/zn.v6i2.20191.

[10] M. I. Ciptandini and R. Prathivi, "Klasterisasi Tingkat Kecanduan Penggunaan Tiktok Terhadap Minat Belajar Menggunakan Algoritma K-Means Clustering," *KESATRIA J. Penerapan Sist. Inf. (Komputer Manajemen)*, vol. 6, no. 1, pp. 77–84, 2025.

[11] M. C. B. Rahman, Martanto, and U. Hayati, "Analisis Tingkat Kecenderungan Fear of Missing Out Menggunakan Algoritma Random Forest Pada Media Sosial," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 1, pp. 296–302, 2024, doi: 10.36040/jati.v8i1.8356.

[12] R. Rismala, *et al*., "Kajian Ilmiah dan Deteksi Adiksi Internet dan Media Sosial di Indonesia Menggunakan XGBoost," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 1, pp. 1-11, 2021, doi: 10.26418/jp.v7i1.43606.

[13] C. N. Syahputri dan M. S. Hasibuan, "Optimasi Klasifikasi Decision Tree Dengan Teknik Pruning Untuk Mengurangi Overfitting," *JSiI (Jurnal Sist. Informasi)*, vol. 11, no. 2, pp. 87–96, 2024, doi: 10.30656/jsii.v11i2.9161.

[14] R. Fitriana, A. N. Habyba, dan E. Febriani, *Data Mining dan Aplikasinya: Contoh Kasus di Industri Manufaktur dan Jasa*, 1 ed. Purwokerto: Wawasan Ilmu, 2022.

[15] S. Jun, "Evolutionary algorithm for improving decision tree with global discretization in manufacturing," *Sensors*, vol. 21, no. 8, 2021, doi: 10.3390/s21082849.