

# Analysis of Sosio-Environmental Factors on The Risk of Drug Abuse Using Naïve Bayes and Decision Tree Algorithms

Fitra Galatya Putry<sup>1</sup>, Besse Helmi Mustawinar<sup>2\*</sup>, Fitriani A<sup>3</sup>

*1Program Studi Matematika, Fakultas Sains, Universitas Cokroaminoto Palopo, Indonesia*

*2 Program Studi Matematika, Fakultas Sains, Universitas Cokroaminoto Palopo, Indonesia*

*3Program Studi Matematika, Fakultas Sains, Universitas Cokroaminoto Palopo, Indonesia*

Email: [emmy.emm92@gmail.com](mailto:emmy.emm92@gmail.com)

*\*Penulis Korespondensi*

## ABSTRACT

This study aims to analyse and compare the performance of the Naïve Bayes classification model and Decision Tree C5.0 in predicting environmental risk factors for drug abuse in Palopo City. The research data come from the Badan Narkotika Nasional Kota Palopo, comprising a total of 74 data points with environmental label attributes and five main attributes. The data were processed through feature selection stages using Random Forest and model validation using the Leave-One-Out Cross-Validation (LOOCV) method to obtain unbiased performance estimates. The results show that personality is the most dominant attribute in determining environmental factors for drug abuse. In the performance comparison, the Naïve Bayes algorithm proved superior with an overall accuracy of 50.00% and a Mean Balanced Accuracy of 62.33%, surpassing Decision Tree C5.0 (accuracy of 47.30%). The main implication of this finding is that the Naïve Bayes model exhibits high specificity (79.39%), making it a reliable tool for early screening to validate the safety of the social environment. Practically, these findings suggest that BNNK Palopo prioritise interventions that focus on strengthening psychosocial aspects (personality) and utilise the Naïve Bayes model for more efficient and targeted resource allocation.

*Keywords: Drugs; Data Mining; Naïve Bayes; Decision Tree; LOOCV*

## ABSTRAK

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja model klasifikasi Naïve Bayes dan Decision Tree C5.0 dalam memprediksi faktor lingkungan yang berisiko terhadap penyalahgunaan narkoba di Kota Palopo. Data penelitian bersumber dari Badan Narkotika Nasional Kota Palopo dengan jumlah data sebanyak 74 dengan atribut label lingkungan dan 5 atribut utama. Data diolah melalui tahapan feature selection menggunakan Random Forest dan validasi model menggunakan metode Leave-One-Out Cross-Validation (LOOCV) untuk mendapatkan estimasi kinerja yang tidak bias. Hasil penelitian menunjukkan kepribadian adalah atribut paling dominan dalam menentukan faktor lingkungan penyalahgunaan narkoba. Dalam perbandingan kinerja, algoritma Naïve Bayes terbukti unggul dengan akurasi keseluruhan 50.00% dan Mean Balanced Accuracy 62.33%, melampaui Decision Tree C5.0 (akurasi 47.30%). Implikasi utama dari temuan ini adalah model Naïve Bayes menunjukkan spesifisitas yang tinggi (79.39%), menjadikannya alat yang andal sebagai sistem penyaring awal (early screening system) untuk memvalidasi keamanan lingkungan sosial. Secara praktis, temuan ini menyarankan BNNK Palopo untuk memprioritaskan intervensi yang berfokus pada penguatan aspek psikososial (kepribadian) dan memanfaatkan model Naïve Bayes untuk alokasi sumber daya yang lebih efisien dan tepat sasaran.

*Kata Kunci: Narkoba; Data Mining; Naïve Bayes; Decision Tree; LOOCV*



*This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial 4.0 International License. Copyright © 2025 by the Author(s).*

## I. Pendahuluan

Penyalahgunaan narkoba di Indonesia merupakan isu krusial dan kompleks yang menimbulkan dampak multidimensional terhadap tatanan sosial, ekonomi, dan kesehatan masyarakat. Badan Narkotika Nasional menjelaskan tren kasus penyalahgunaan narkotika terus menunjukkan peningkatan signifikan dari tahun ke tahun, yang ditandai dengan bertambahnya jumlah pengguna dan meluasnya jaringan sindikat peredaran. NAPZA (Narkotika, Psikotropika, dan Zat Adiktif lainnya) sebagai senyawa psikoaktif, dapat memicu ketergantungan fisik dan psikologis yang destruktif bagi individu. Survei nasional oleh BNN (2024) mengonfirmasi urgensi masalah ini dengan melaporkan prevalensi penyalahgunaan narkoba mencapai 1,73% atau setara dengan 3,3 juta penduduk, dengan dominasi pada kelompok usia produktif dan remaja [1].

Secara spesifik di tingkat lokal, BNN Kota Palopo mencatat data yang mengkhawatirkan. Selama periode Januari hingga Juli 2024, Satuan Reserse Narkoba Polres Palopo telah menangani 36 kasus penyalahgunaan narkoba yang melibatkan 53 tersangka, mayoritas berjenis kelamin laki-laki dan termasuk satu perempuan di bawah umur. Data ini mengindikasikan bahwa peredaran dan penyalahgunaan narkoba di Kota Palopo tidak hanya meningkat, tetapi juga telah merambah ke berbagai lapisan masyarakat, termasuk kelompok rentan. Faktor pemicunya bermacam-macam, mulai dari dinamika keluarga, tekanan sosial dari teman sebaya, konflik internal individu, hingga lingkungan yang permisif [2]. Analisis mendalam terhadap faktor-faktor ini menjadi krusial untuk merumuskan strategi pencegahan yang efektif dan berbasis bukti.

Dalam beberapa tahun terakhir, penerapan teknik data mining dan machine learning telah terbukti menjadi pendekatan yang kuat untuk menganalisis dan memprediksi perilaku sosial kompleks, termasuk risiko penyalahgunaan narkoba. Korelasi dari faktor perilaku, lingkungan, sosial ekonomi, geografis terhadap kasus kecanduan narkoba di Indonesia [3]. Sejumlah penelitian di Indonesia telah menunjukkan efektivitasnya dalam data mining. Penerapan data mining dengan algoritma Naïve Bayes digunakan untuk mengklasifikasikan upaya pencegahan narkoba berdasarkan data dari BNN [4], [5]. Naïve Bayes juga diterapkan untuk mengidentifikasi kelompok rentan dengan akurasi yang tinggi [6]. Lebih lanjut, penelitian oleh Moningga dan Kusrini membandingkan beberapa teknik dalam data mining, seperti algoritma apriori, naïve bayes, random forest, dan support vector machine (SVM) untuk menganalisis peredaran narkoba di Provinsi Maluku [7].

Secara paralel, komunitas riset internasional telah lebih jauh mengeksplorasi kapabilitas machine learning dalam domain ini. Sebagai contoh, pengembangan model prediktif untuk mengidentifikasi individu berisiko tinggi terhadap penggunaan opioid problematik dengan menganalisis data rekam medis elektronik (electronic health records) [8]. Berdasarkan data survei dan penggunaan data mining diprediksi jika penggunaan zat psikoaktif dengan performa model yang sangat baik, di mana faktor sosial-demografis memegang peranan penting, misalnya variabel seperti status merokok, konsumsi alkohol, dan kondisi kesehatan mental merupakan prediktor universal yang kuat untuk penggunaan zat di kalangan remaja [9], [10]. Model machine learning, yaitu Random Forest, Decision Tree, Logistic Regression, Support Vector Machine, K-Nearest neighbors, and Gaussian Naive Bayes, efektif dalam memprediksi risiko penyalahgunaan zat adiktif berdasarkan data lingkungan dan sosial [11].

Sebuah tinjauan komprehensif menjelaskan machine learning tidak hanya efektif untuk prediksi tetapi juga untuk mengidentifikasi fitur-fitur paling berpengaruh, mulai dari faktor genetik hingga lingkungan sosial, yang sangat krusial untuk intervensi dini. Pengelompokan

pola pengguna narkoba juga dilakukan dengan menggunakan data mining. Temuan-temuan ini secara kolektif menunjukkan bahwa pendekatan komputasional mampu mengungkap pola tersembunyi dari data yang kompleks untuk tujuan prediksi dan pencegahan [12], [13].

Meskipun metode probabilistik Naïve Bayes digunakan karena kecepatannya, penelitian ini juga mengadopsi algoritma Decision Tree C5.0 sebagai pembanding yang strategis. Pemilihan C5.0 didasarkan pada keunggulannya dalam interpretabilitas: algoritma ini menghasilkan model berbentuk pohon keputusan dan rule set (jika-maka) yang mudah dipahami oleh manusia, berbeda dengan model black-box lainnya. Fitur ini sangat vital bagi pemangku kebijakan seperti BNN untuk memahami alur logika penyebab risiko secara transparan. Keunggulan ini menyoroti bagaimana pohon keputusan memberikan visibilitas langsung terhadap fitur mana yang paling penting dalam membuat keputusan klasifikasi, suatu aspek krusial untuk intervensi berbasis bukti [14]. Selain itu, C5.0 merupakan penyempurnaan dari algoritma C4.5 yang terbukti memiliki akurasi lebih tinggi, efisiensi memori yang lebih baik, serta kemampuan menangani fitur yang tidak relevan melalui teknik boosting dan pruning yang canggih [15]. Algoritma berbasis Decision Tree sangat tangguh dalam menangani data sosial yang seringkali memiliki ketidakseimbangan kelas, menjadikannya kandidat yang tepat untuk memetakan risiko narkoba yang kompleks. Dengan akurasi yang kompetitif dan transparansi pengambilan keputusan, C5.0 ideal untuk merumuskan strategi pencegahan yang terarah dan dapat dipertanggungjawabkan [16].

Meskipun potensi data mining telah terbukti relevan, aplikasinya untuk menganalisis faktor lingkungan secara spesifik dalam konteks pencegahan narkoba di Kota Palopo masih terbatas. Berdasarkan hal tersebut maka penelitian ini bertujuan untuk menerapkan dan membandingkan kinerja dua algoritma klasifikasi populer, yaitu Naïve Bayes dan Decision Tree C5.0, untuk menganalisis pengaruh faktor lingkungan terhadap risiko penyalahgunaan narkoba pada data yang bersumber dari BNN Kota Palopo. Hasil dari penelitian ini diharapkan dapat memberikan landasan ilmiah bagi BNN Kota Palopo untuk merancang strategi pencegahan berbasis data (data-driven prevention strategy) yang lebih akurat, terarah, dan efisien.

## II. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan data sekunder sebagai sumber utama analisis. Data yang digunakan merupakan data rehabilitasi penyalahgunaan narkoba tahun 2024 yang diperoleh dari Badan Narkotika Nasional (BNN) Kota Palopo. Data tersebut mencakup berbagai atribut sosial yang relevan dengan upaya pencegahan serta rehabilitasi penyalahgunaan narkoba. Atribut-atribut yang dipilih berdasarkan relevansinya terhadap faktor risiko penyalahgunaan narkoba menurut pedoman BNN.

Dalam penelitian ini terdapat lima atribut utama yang dijadikan variabel klasifikasi, yaitu: umur, pekerjaan, pendidikan, penggunaan narkoba, kepribadian, dan atribut lingkungan sebagai atribut label, yang dapat diamati pada tabel berikut.

Tabel 1. Data Atribut Pengguna Narkoba di BNNK Palopo

Atribut	Kategori
Umur	Remaja (14 – 18 tahun)
	Dewasa Muda (19 – 25 tahun)
	Dewasa (26 – 59 tahun)
Pekerjaan	Pelajar
	Wiraswasta
	Pemerintahan
	Pengangguran
Pendidikan	Tinggi
	Rendah
Penggunaan Narkoba	Tidak Lama
	Lama
	Sangat Lama
Kepribadian	Ajakan Teman
	Anjuran Keluarga
	Penambah Stamina
	Coba-coba
Lingkungan	Depresi
	Sahabat
	Teman
	Keluarga
	Masyarakat

Sumber: Data BNNK Palopo (diolah Penulis, 2025)

Alat analisis dalam penelitian ini menggunakan *software R*. Metode analisis yang digunakan adalah algoritma *Naïve Bayes* dan *Decision Tree* algoritma C5.0, dua teknik populer dalam *machine learning* yang digunakan untuk proses klasifikasi. Algoritma *Naïve Bayes* memiliki kemampuan dalam memproses dataset berdimensi tinggi dengan asumsi independensi antar atribut [17]. Tujuan utamanya adalah menghitung probabilitas posterior  $P(C_k|X)$  untuk setiap kelas  $C_k$  berdasarkan vektor fitur  $X = (x_1, x_2, \dots, x_n)$  yang dirumuskan sebagai berikut.

$$\text{Kelas}_{\text{prediksi}} = \underset{k}{\operatorname{argmax}} P(C_k) \cdot \prod_{i=1}^n P(x_i|C_k)$$

Dimana  $P(C_k)$  adalah probabilitas prior kelas  $C_k$ ,  $P(x_i|C_k)$  adalah *likelihood* fitur  $x_i$  pada kelas  $C_k$ , dan  $\prod_{i=1}^n P(x_i|C_k)$  adalah produk dari probabilitas independen semua fitur.

Sedangkan, *Decision Tree* digunakan sebagai pembanding karena keunggulannya dalam menghasilkan model yang mudah diinterpretasikan serta mampu menangani atribut kategorikal dan numerik secara bersamaan. Algoritma C5.0 merupakan Algoritma berbasis *Decision Tree* yang merupakan penyempurnaan dari Algoritma ID3 dan C4.5 yang dibentuk oleh Ross Quinlan pada tahun 1987. Langkah untuk pembentukan *decision tree* pada Algoritma C5.0 serupa dengan langkah pembentukan pohon keputusan pada Algoritma C4.5. Hal yang menyerupai terdapat pada perhitungan *entropy* dan *gain*, yang dirumuskan:

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (2)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (3)$$

Namun, perbedaannya terletak pada tahapan setelah perhitungan *gain*, algoritma C5.0 melanjutkan perhitungannya dengan menghitung *gain ratio*, yang dirumuskan sebagai berikut.

$$Gain\ ratio = \frac{Gain(A)}{\sum_{i=1}^m Entropy(A_{ij})} \quad (4)$$

Tahapan penelitian dilaksanakan melalui beberapa proses, yaitu:

1. Pengumpulan data sekunder, diperoleh dari dokumentasi resmi BNN Kota Palopo, dengan target klasifikasi adalah lingkungan pengguna narkoba yang merupakan variabel kategorikal.
2. Pra-pemrosesan data (*data preprocessing*) yang meliputi pembersihan data (*data cleaning*), penanganan data kosong, normalisasi nilai atribut, serta transformasi tipe kategorikal (*factor*), dan penanganan *missing value*.
3. Reduksi Dimensi dan Seleksi Fitur untuk meningkatkan efisiensi komputasi dan mengurangi risiko *overfitting* dikarenakan jumlah dataset yang kecil, dengan Teknik *Feature Selection* berbasis *Random Forest* dengan skor *variable importance* diukur menggunakan metrik *Mean Decrease Gini*.
4. Konfigurasi metode validasi dengan metode *Leave-One-Out Cross-Validation* (LOOCV) untuk evaluasi model, dimana LOOCV memaksimalkan ukuran training set pada setiap iterasi, menghasilkan estimasi kinerja model yang paling tidak bias (*unbiased*) untuk dataset berukuran kecil.
5. Model akan dilatih dengan tiga fitur yang terpilih dengan control validasi LOOCV untuk metode *Naïve Bayes* dan *Decision Tree C5.0*

Kinerja kedua model (*Naïve Bayes* dan *Decision Tree C5.0*) dievaluasi melalui i) akurasi (*accuracy*): Tingkat prediksi benar secara keseluruhan, ii) kappa: Ukuran kesepakatan yang disesuaikan untuk peluang (lebih kuat dari akurasi), iii) Akurasi Seimbang (*Balanced Accuracy*): Rata-rata dari Sensitivitas dan Spesifisitas (penting untuk data tidak seimbang), iv) Sensitivitas (*Sensitivity*): Kemampuan model mendeteksi kasus positif, v) Spesifisitas (*Specificity*): Kemampuan model mengidentifikasi kasus negatif. Metrik ini dihitung berdasarkan nilai dari *Confusion Matrix*.

### III. Hasil dan Pembahasan

Sebagai bagian dari preprossesing data, data sekunder yang diambil sebanyak 74 data pengguna narkoba yang menjalani proses rehabilitasi di Kantor BNN Kota Palopo sebagai dasar analisis. Pemilihan atribut yang dianggap memiliki pengaruh pada potensi pengaruh lingkungan terhadap pencegahan penggunaan narkoba, dengan beberapa contoh dataset ditampilkan melalui gambar berikut.

Umur	Pekerjaan	Pendidikan	Penggunaan Narkoba	Kepribadian	Lingkungan
Dewasa Muda	Wiraswasta	Rendah	Lama	Penambah stamina	Teman
Dewasa	Pengangguran	Rendah	Sangat lama	Coba-coba	Teman
Dewasa Muda	Pengangguran	Rendah	Lama	ajakan teman	Teman
Dewasa	Pengangguran	Rendah	Lama	ajakan teman	Teman
Remaja	Pelajar	Rendah	Lama	depresi	Masyarakat
Dewasa	Wiraswasta	Tinggi	Sangat lama	Penambah stamina	Teman
Dewasa	Pemerintahan	Tinggi	Sangat lama	Anjuran keluarga	Keluarga
Remaja	Pelajar	Rendah	Lama	ajakan teman	Masyarakat
Dewasa	Wiraswasta	Tinggi	Sangat lama	Anjuran keluarga	Keluarga

Gambar 1. Dataset Penelitian

Berdasarkan gambar 1 diketahui jika ukuran sampel yang digunakan terbatas yaitu 74 maka Langkah awal dilakukan adalah mengidentifikasi variabel prediktor yang paling informatif untuk mengurangi risiko *overfitting*. Sebelum melakukan pemodelan klasifikasi, penelitian ini menerapkan seleksi fitur menggunakan algoritma *Random Forest* dengan metrik *mean decrease gini* untuk mengidentifikasi atribut yang memiliki kontribusi paling signifikan terhadap klasifikasi lingkungan risiko narkoba. Metrik ini mengevaluasi kontribusi setiap variabel dalam meningkatkan kemurnian klasifikasi di seluruh pohon keputusan yang dibangun. Hasil dari analisis kepentingan variabel disajikan pada Tabel 2. Berdasarkan metrik *mean decrease gini*, ditemukan bahwa dari lima atribut klasifikasi, terdapat tiga atribut dengan nilai kepentingan tertinggi yang dipilih untuk pemodelan selanjutnya.

Tabel 2. Peringkat Kepentingan Fitur berdasarkan *Mean Decrease Gini*

Peringkat	Fitur	<i>Mean Decrease Gini</i>	<i>Mean Decrease Accuracy</i>
1	Kepribadian	12.967	10.397
2	Pekerjaan	5.497	2.142
3	Penggunaan Narkoba	5.360	2.348
4	Umur	4.110	1.966
5	Pendidikan	3.267	3.926

Sumber : data diolah (2025)

Berdasarkan Tabel 2 yang menjelaskan bahwa atribut kepribadian (alasan penggunaan narkoba) merupakan fitur dengan daya pembeda tertinggi secara signifikan, dengan skor *mean decrease gini* tertinggi sebesar 12.967. Fitur ini jauh lebih berpengaruh dibandingkan fitur lainnya. Atribut Pekerjaan dan Penggunaan Narkoba menempati peringkat kedua dan ketiga dengan skor yang relatif sebanding. Hal ini mengindikasikan bahwa karakteristik internal individu (kepribadian) menjadi determinan utama dalam menentukan jenis lingkungan pergaulan seseorang, dibandingkan faktor demografis seperti umur atau pendidikan.

Berdasarkan temuan ini, tiga fitur teratas yaitu Kepribadian, Pekerjaan, dan Penggunaan Narkoba dipilih untuk digunakan dalam tahap pemodelan klasifikasi selanjutnya, untuk mengurangi *noise* dan menghindari *overfitting* pada dataset yang terbatas ( $N=74$ ).

Selanjutnya adalah melakukan evaluasi untuk dua algoritma klasifikasi, *Naïve Bayes* dan *Decision Tree C5.0*. Evaluasi kinerja model dilakukan menggunakan metode *Leave-One-Out Cross-Validation* (LOOCV) untuk memastikan validitas hasil pada dataset berskala kecil. Dalam metode ini, model dilatih sebanyak  $N$  kali pada  $N-1$  data dan diuji pada satu data yang tersisa, di mana  $N$  adalah jumlah total sampel. Kinerja model dievaluasi berdasarkan rata-rata metrik yang dihitung dari keseluruhan proses validasi. Model C5.0 secara internal dioptimalkan oleh perintah *caret* dengan mencoba beberapa parameter, seperti jumlah *trials* (*boosting*) dan penggunaan *winnowing*. Untuk *Naïve Bayes*, perintah *caret* menguji penggunaan *kernel density* dan parameter *Laplace smoothing*. Model dengan konfigurasi parameter terbaik berdasarkan metrik *accuracy* dapat diamati pada tabel 3.

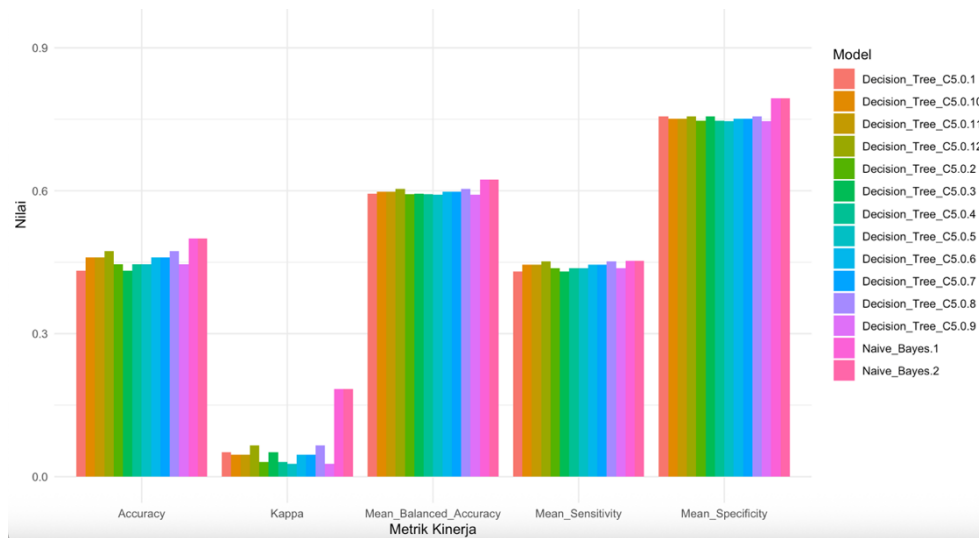
Tabel 3. Ringkasan Perbandingan Kinerja Model LOOCV

<i>Model</i>	<i>Naïve Bayes</i>	<i>Decision Tree C5.0</i>
<i>Accuracy</i>	0.500	0.473
<i>Kappa</i>	0.184	0.066
<i>Mean Balanced Accuracy</i>	0.623	0.604
<i>Mean Sensitivity</i>	0.453	0.451
<i>Mean Specificity</i>	0.794	0.756

Sumber : data diolah (2025)

Hasil evaluasi pada Tabel 3 menunjukkan bahwa model *Naïve Bayes* menunjukkan kinerja yang sedikit lebih unggul secara konsisten di semua metrik utama dibandingkan dengan *Decision Tree C5.0*. Model *Naïve Bayes* mencatat akurasi seluruhnya sebesar 50.0% dengan *balanced accuracy* mencapai 62,33%. Sedangkan untuk *decision tree C5.0* yang dioptimalkan dengan *boosting* 20 *trials*, hanya mencapai tingkat akurasi maksimal 47.30%. Nilai Kappa pada *Naïve Bayes* (0.184) menunjukkan adanya kesepakatan (*agreement*) yang "sedikit" (*slight agreement*) di atas kebetulan. Sedangkan, *decision tree C5.0* memiliki nilai Kappa yang sangat rendah (0.066), mendekati nol. Hal menarik terlihat pada nilai Spesifisitas *Naïve Bayes* yang cukup tinggi (79.39%), yang berarti model ini cukup andal dalam memprediksi kelas negatif, namun masih berjuang dalam mendeteksi kelas positif (Sensitivitas ~45%).

Untuk memvisualisasikan perbandingan metrik kinerja secara intuitif, Gambar 2 menyajikan perbedaan nilai akurasi, kappa, dan *Balanced Accuracy* antara kedua model. Secara visualisasi data, terlihat jelas bahwa *Naïve Bayes* secara konsisten menghasilkan batang (*bar*) yang lebih tinggi daripada *Decision Tree C5.0* di semua metrik yang ditampilkan. Hal ini menunjukkan jika *Naïve Bayes* adalah algoritma terbaik untuk pemodelan data ini. Keunggulan ini dipicu oleh arsitektur *Naïve Bayes* yang sederhana dan probabilistik, yang efektif dalam menangkap pola kunci dari sedikit data tanpa membangun struktur yang terlalu kompleks, seperti yang terjadi pada *Decision Tree C5.0*.



Gambar 2. Perbandingan Kinerja *Naïve Bayes* dan *Decision Tree C5.0* Berdasarkan Metrik LOOCV

Temuan penelitian ini mengonfirmasi bahwa pada dataset dengan jumlah sampel terbatas ( $N=74$ ), pendekatan probabilistik sederhana seperti *Naïve Bayes* cenderung lebih stabil dan akurat dibandingkan pendekatan berbasis aturan yang kompleks seperti *decision tree C5.0*. Algoritma C5.0 membutuhkan jumlah data yang lebih besar untuk membentuk struktur pohon keputusan yang mendalam dan melakukan pruning yang efektif. Pada data kecil, C5.0 rentan mengalami kesulitan dalam generalisasi, yang terlihat dari rendahnya akurasi meskipun telah diterapkan teknik *boosting* (hingga 20 trials). Sebaliknya, *Naïve Bayes* yang bekerja dengan asumsi independensi antar fitur mampu memanfaatkan pola probabilitas global dari fitur dominan (seperti Kepribadian) tanpa terpengaruh kompleksitas struktur data.

Meskipun *balanced accuracy* mencapai 62%, akurasi keseluruhan hanya berkisar 50% menunjukkan kompleksitas masalah klasifikasi lingkungan sosial ini. Rendahnya akurasi ini dapat dikaitkan dengan dua faktor, yaitu 1) ambiguitas data sosial dimana perilaku manusia dan lingkungan sosial memiliki variabilitas yang sangat tinggi dan tidak selalu linier. Faktor "Kepribadian" yang menjadi fitur terkuat mungkin memiliki irisan pola antar kelas yang menyulitkan model untuk menarik garis pemisah yang tegas, 2) keterbatasan jumlah sampel membuat model memiliki kesempatan belajar yang terbatas untuk menangkap seluruh variasi pola risiko.

Tingginya skor *mean decrease gini* pada variabel kepribadian (12.967) menyiratkan bahwa intervensi pencegahan narkoba di Kota Palopo tidak bisa hanya berfokus pada faktor eksternal (seperti merazia tempat hiburan), tetapi harus menyentuh aspek psikososial individu. Strategi penguatan karakter dan kesehatan mental (kepribadian) berpotensi lebih efektif dalam membentengi individu dari pengaruh lingkungan yang buruk, mengingat hasil data menunjukkan bahwa kepribadian adalah prediktor terkuat lingkungan di mana seseorang berada.

#### IV. Kesimpulan

Hasil analisis menunjukkan bahwa dari semua atribut yang diuji, variabel kepribadian teridentifikasi sebagai prediktor yang paling dominan, diikuti oleh pekerjaan dan penggunaan narkoba, sebuah temuan yang menegaskan peran krusial faktor psikososial individu dalam membentuk faktor lingkungan dalam penyalahgunaan

narkoba. Dalam evaluasi kinerja, algoritma Naïve Bayes terbukti lebih unggul dan stabil, mencapai akurasi keseluruhan 50.00% dan mean balanced accuracy 62.33%, mengungguli Decision Tree C5.0 yang hanya mencapai akurasi 47.30%. Kinerja superior Naïve Bayes ini mengindikasikan bahwa pada dataset berukuran kecil, pendekatan probabilistik sederhana lebih efektif dalam menangkap pola klasifikasi. Lebih lanjut, model Naïve Bayes menunjukkan nilai spesifisitas yang tinggi (79.39%), menyiratkan bahwa model ini sangat andal dalam mengonfirmasi bahwa individu berada dalam lingkungan yang aman atau tidak berisiko. Oleh karena itu, Naïve Bayes direkomendasikan sebagai algoritma yang paling sesuai untuk studi kasus ini dan dapat diaplikasikan sebagai alat skrining awal yang andal oleh instansi terkait. Untuk penelitian mendatang, disarankan untuk menguji efektivitas model ini pada dataset yang lebih besar untuk meningkatkan sensitivitas (kemampuan mendeteksi risiko tinggi) dan mengeksplorasi teknik ensemble yang lebih maju untuk perbandingan kinerja yang lebih komprehensif.

### Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Badan Narkotika Nasional (BNN) Kota Palopo atas dukungan, kerja sama, dan bantuan yang telah diberikan selama proses penelitian. Ketersediaan data yang akurat dan kemudahan dalam berkoordinasi dengan pihak BNN Kota Palopo menjadi salah satu faktor utama yang menunjang kelancaran penelitian, khususnya dalam memahami dinamika lingkungan sosial dan upaya pencegahan penyalahgunaan narkoba.

### Daftar Pustaka

- [1] Badan Narkotika Nasional. (2025). PERINGATAN HANI 2025: MEMUTUS RANTAI PEREDARAN GELAP NARKOBA MELALUI PENCEGAHAN, REHABILITASI, DAN PEMBERANTASAN MENUJU INDONESIA EMAS 2045. Diakses pada 9 November 2025 dari <https://bnn.go.id/peringatan-hani-2025-memutus-rantai-peredaran-gelap-narkoba-melalui-pencegahan-rehabilitasi-dan-pemberantasan-menuju-indonesia-emas-2045/>
- [2] Andi Bunayya Nandini. (2024). Dalam 6 Bulan, Polisi Tangani 36 Kasus Penyalahgunaan Narkoba di Palopo Sulsel. (Online). <https://makassar.tribunnews.com/2024/07/05/dalam-6-bulan-polisi-tangani-36-kasus-penyalahgunaan-narkoba-di-palopo-sulsel>
- [3] Dewabhrata, W., Ahsan, A., Bella, A., Amalia, N., Kusuma, D., & Pertiwi, Y. B. A., "Mental Health, Environmental, and Socioeconomic Geographic Factors of Severe Drug Addiction: Analysis of Rehabilitation Center Data in Indonesia," Substance Abuse: Research and Treatment, vol. 17, 2023. <https://doi.org/10.1177/11782218231203687>
- [4] S. R. D. Situmorang and N. Hidayati, "Penerapan Data Mining dalam Klasifikasi Pencegahan Narkoba Menggunakan Algoritma Naïve Bayes di BNN Kota Pematangsiantar Application of Data Mining in Drug Prevention Classification Using the Naïve Bayes Algorithm in BNN Pematangsiantar City," JOMLAI: Journal of Machine Learning and Artificial Intelligence, vol. 1, no. 4, pp. 295–302, Sep. 2022. <http://doi.org/10.55123/jomlai.v1i4.1667>
- [5] D. Setiadi and R. Syahri, "PENERAPAN ALGORITMA NAÏVE BAYES PADA SISTEM PREDIKSI PENGGUNA NARKOBA DI KOTA PAGAR ALAM," JUTIM (Jurnal Teknik Informatika Musirawas), vol. 7, no. 1, pp. 1-10, 2022.
- [6] D. Haryadi, D. M. U. Atmaja, and A. R. Hakim, "Implementation of text mining for classification of drug effectiveness using the naïve bayes algorithm," in AIP Conference Proceedings, Mar. 2024, vol. 2927, no. 1, p. 060030. <https://doi.org/10.1063/5.0205662>
- [7] N. Moningga and K. Kusrini, "Use of Data Mining Technology to Identify Narcotics Distribution Patterns," SITEKIN: Jurnal Sains, Teknologi dan Industri, vol. 22, no. 1, pp. 75-82.
- [8] C. Bharat, M. Hickman, S. Barbieri, and L. Degenhardt, "Big data and predictive modelling for the opioid crisis: existing research and future potential," The Lancet Digital Health, vol. 3, no. 6, pp. e397-e407, 2021. DOI: [10.1016/S2589-7500\(21\)00058-3](https://doi.org/10.1016/S2589-7500(21)00058-3)

- [9] S. Kim, H. Kim, S. Kim, H. Lee, A. Hammoodi, Y. Choi, A. T. H. C. Team, and D. K. Yon, "Machine learning-based prediction of substance use in adolescents in three independent worldwide cohorts: Algorithm development and validation study," *Journal of Medical Internet Research*, vol. 27, p. e62805, 2025. [doi:10.2196/62805](https://doi.org/10.2196/62805)
- [10] A. Unlu and A. Subasi, "Substance use prediction using artificial intelligence techniques," *Journal of Computational Social Science*, vol. 8, no. 1, p. 21, 2025. <https://doi.org/10.1007/s42001-024-00356-6>
- [11] U. I. Islam, E. Haque, D. Alsalman, M. N. Islam, M. A. Moni, and I. H. Sarker, "A machine learning model for predicting individual substance abuse with associated risk-factors," *Annals of Data Science*, vol. 10, no. 6, pp. 1607-1634, 2023. <https://doi.org/10.1007/s40745-022-00381-0>
- [12] K. K. Mak, K. Lee, and C. Park, "Applications of machine learning in addiction studies: A systematic review," *Psychiatry research*, vol. 275, pp. 53-60, 2019. <https://doi.org/10.1016/j.psychres.2019.03.001>
- [13] Mustawinar, B. H., Irenita, I., Yuliani, Y., Wulandari, N., & Hirman, H. (2025). Analysis of Drug User Pattern Grouping Based on Social and Psychological Factors Using the K-Modes Method. ICSIGER- International Conference on Science and Emerging Technology , 1(1). Retrieved from <https://conference.lppm.unila.ac.id/index.php/icsiger/article/view/63>
- [14] R. Hasan, "Comparative analysis of machine learning algorithms for heart disease prediction," in *ITM Web of Conferences*, 2021, vol. 40, p. 03007. <https://doi.org/10.1051/itmconf/20214003007>
- [15] Y. Crismayella, N. Satyahadewi, and H. Perdana, "Comparison of Adaboost Application to C4. 5 and C5. 0 Algorithms in Student Graduation Classification," *Pattimura International Journal of Mathematics (PIJMath)*, vol. 2, no. 1, pp. 07-16, 2023. <https://doi.org/10.30598/pijmathvol2iss1pp07-16>
- [16] Maulani, G., Hasan, F. N., Setiawan, D., Bowo, I. T., Ardhana, V. Y. P., Ramdhani, Y., ... & Safitri, R. (2025). Machine Learning. Mega Press Nusantara.
- [17] Ananta, A., Wulandari, N., Mustawinar, B. H., & Putri, F. G. (2025). Klasifikasi Pembelian Produk Rumah Tangga Melalui Metode Naïve Bayes. *indonesian journal of material and applied physics*, 1(1), 16-21.