

IMPLEMENTASI ALGORITMA NAÏVE BAYES UNTUK PREDIKSI
RISIKO RELAPS PADA SISTEM MONITORING PASIEN REHABILITASI
NARKOBA

¹⁾ Steve Jeremy Rarumangkay, ²⁾ Yonathan Kazu Datumbanua, ³⁾ Levi Elia Pitoy, ⁴⁾ Marike Amelda Silvia Kondo, ⁵⁾ Franky Gerald Clifford Manoppo
^{1,2,3,4,5)} Program Studi Teknik Informatika, Jurusan Teknik Elektro, Politeknik Negeri Manado
¹⁾ jrmmngky@gmail.com, ²⁾ mariekondoj@polimdo.ac.id, ³⁾ franky.cliford@gmail.com

INFO ARTIKEL	ABSTRAK
Riwayat Artikel : Diterima : 15 Mei 2026 Disetujui : 9 Juli 2026	Penelitian ini bertujuan mengimplementasikan algoritma Gaussian Naïve Bayes pada sistem Analisis Risiko Relaps untuk memprediksi risiko relaps pasien rehabilitasi narkoba secara cepat, terstruktur, dan berbasis data. Sistem dikembangkan menggunakan Python, Streamlit, dan MySQL dengan pendekatan Research and Development. Tahapan penelitian meliputi pengumpulan data, prapemrosesan, encoding, standardisasi, SMOTE, pembagian data 80:20, pelatihan model, evaluasi, dan implementasi sistem. Variabel prediktor mencakup usia, riwayat pakai, hasil tes urine, lama pakai, status kerja, dan dukungan keluarga. Dari 87 data sampel pemeriksaan dengan 76 pasien unik, sistem mengklasifikasikan 44 pasien berisiko tinggi dan 32 pasien berisiko rendah. Evaluasi awal pada 18 data uji menghasilkan accuracy, precision, recall, dan F1-score sebesar 1,00. Meskipun hasil ini menunjukkan performa model yang sangat baik pada data uji, ukuran sampel pengujian yang masih terbatas menyebabkan hasil tersebut belum dapat digeneralisasi secara luas. Oleh karena itu, sistem masih memerlukan pengujian lebih lanjut menggunakan dataset yang lebih besar, lebih beragam, serta validasi silang untuk memastikan stabilitas dan generalisasi model sebelum diterapkan secara operasional dalam skala luas.

ARTICLE INFO	ABSTRACT
Article History : Received : May 15, 2026 Accepted : Jul 9, 2026	<i>This study aims to implement the Gaussian Naïve Bayes algorithm in the Relapse Risk Analysis system to predict relapse risk among drug rehabilitation patients in a fast, structured, and data-driven manner. The system was developed using Python, Streamlit, and MySQL through a Research and Development approach. The research stages included data collection, preprocessing, encoding, standardization, SMOTE, an 80:20 data split, model training, evaluation, and system implementation. The predictor variables consisted of age, drug use history, urine test results, duration of use, employment status, and family support. From 87 examination sample records involving 76 unique patients, the system classified 44 patients as high risk and 32 patients as low risk. Initial evaluation on 18 test data records yielded an accuracy, precision, recall, and F1-score of 1.00. Although these results indicate excellent model performance on the test data, the limited size of the testing sample means that the findings cannot yet be generalized broadly. Therefore, the system still requires further evaluation using</i>
Keywords: Gaussian Naïve Bayes, relapse, drug rehabilitation, SMOTE, monitoring system.	

larger and more diverse datasets, as well as cross-validation, to ensure the model's stability and generalizability before being deployed operationally on a wider scale.

1. PENDAHULUAN

Relaps pada pasien rehabilitasi narkoba merupakan persoalan yang membutuhkan pemantauan berkelanjutan berbasis data. Proses rehabilitasi tidak hanya berkaitan dengan penghentian penggunaan zat, tetapi juga pengawasan terhadap perubahan kondisi pasien setelah menjalani program pemulihan. Pada praktik konvensional, penilaian risiko relaps sering bergantung pada observasi petugas, catatan manual, dan pengalaman klinis. Cara tersebut masih memiliki kelemahan karena data pasien yang terus bertambah sulit dianalisis secara cepat tanpa bantuan sistem komputasi. Kebutuhan sistem monitoring yang mampu menyimpan riwayat pasien, membaca pola risiko, dan menghasilkan rekomendasi awal menjadi relevan dalam pengembangan teknologi informasi pada layanan rehabilitasi.

Pemanfaatan data mining dalam kasus penyalahgunaan narkoba telah digunakan untuk membantu proses klasifikasi dan pengambilan keputusan. Data pemeriksaan, riwayat pasien, serta atribut sosial dapat diolah menjadi informasi klasifikatif melalui algoritma pembelajaran mesin. Beberapa studi pada data BNN menunjukkan bahwa algoritma Naïve Bayes dapat diterapkan untuk mengklasifikasikan data pencegahans narkoba dan pasien tes urine, sehingga data tidak hanya berfungsi sebagai arsip administratif, tetapi juga sebagai dasar analisis prediktif. (Nugraha, Okrpna and Sormin, 2022; Situmorang, Sumarno and Hidayati, 2022)

Algoritma Naïve Bayes dapat dikatakan memiliki karakteristik sederhana, cepat, dan sesuai untuk klasifikasi berbasis atribut. Metode tersebut telah digunakan untuk mendukung diagnosis, klasifikasi tingkat ketergantungan, dan identifikasi penyakit adiktif. Sistem pakar diagnosis pengguna narkoba dapat dibangun dengan memanfaatkan atribut seperti usia, jenis kelamin, riwayat penggunaan, dan gejala (Marfalino *et al.*, 2023). Perbandingan Naïve Bayes dan K-Nearest Neighbor pada klasifikasi penyakit adiktif pengguna narkoba juga menunjukkan performa Naïve Bayes yang

kompetitif, dengan akurasi mencapai 94,5% (Setiawan *et al.*, 2023). Klasifikasi tingkat ketergantungan narkoba menggunakan Naïve Bayes Classifier memperoleh accuracy 94%, precision 95%, dan recall 92% (Puspitasari *et al.*, 2024). Hasil tersebut memperkuat alasan pemilihan Naïve Bayes sebagai algoritma inti pada sistem prediksi risiko relaps.

Pengembangan sistem prediksi risiko relaps membutuhkan proses prapemrosesan data yang baik. Data pasien rehabilitasi umumnya terdiri atas atribut numerik dan kategorikal, seperti usia, lama pemakaian, hasil tes urine, status kerja, riwayat pemakaian, dan dukungan keluarga. Data kategorikal perlu dikonversi ke bentuk numerik agar dapat diproses oleh model machine learning. Data numerik perlu distandardisasi agar perbedaan skala antarvariabel tidak mengganggu proses pelatihan model. Kualitas data, konsistensi format, dan kesesuaian fitur berpengaruh langsung terhadap hasil klasifikasi yang diberikan sistem.

Klasifikasi pada data kesehatan sering berhadapan dengan ketidakseimbangan kelas. Pada data rehabilitasi, jumlah pasien berisiko rendah dan berisiko tinggi dapat berbeda jauh. Kondisi ini dapat membuat model terlalu dominan mengenali kelas mayoritas dan kurang sensitif terhadap kelas minoritas. Teknik oversampling seperti SMOTE digunakan untuk mengatasi ketidakseimbangan data dengan membentuk sampel sintesis pada kelas minoritas. Pendekatan serupa telah digunakan dalam klasifikasi indikator kesehatan karena ketidakseimbangan data dapat menurunkan kemampuan model dalam mendeteksi kelompok berisiko (Pamungkas, Vikri and Sa'ida, 2025).

Machine learning juga mulai digunakan secara luas dalam prediksi luaran perawatan pada substance use disorder. Kajian systematic review menunjukkan bahwa algoritma machine learning telah digunakan untuk memprediksi berbagai luaran perawatan, termasuk kepatuhan terapi, tingkat keparahan, relaps, dan readmission (de Mattos *et al.*, 2024). Pendekatan interpretable machine learning juga digunakan

untuk mengidentifikasi faktor risiko perilaku relaps pada individu dengan substance use disorders, sehingga model prediktif tidak hanya menghasilkan klasifikasi, tetapi juga membantu membaca faktor yang berkontribusi terhadap risiko pasien (Xu *et al.*, 2025).

Sistem pendukung keputusan berbasis machine learning semakin banyak diterapkan dalam bidang kesehatan. Clinical decision support system dapat membantu proses pengambilan keputusan, pelayanan pasien, dan evaluasi luaran kesehatan melalui pengolahan data klinis secara terstruktur (Susanto *et al.*, 2023). Pada sisi implementasi antarmuka, Streamlit dapat digunakan untuk membangun aplikasi prediksi berbasis kecerdasan buatan yang interaktif dan mudah digunakan oleh pengguna non-teknis. Penggunaan Streamlit dalam sistem monitoring kesehatan memperlihatkan bahwa model machine learning dapat disajikan dalam bentuk aplikasi yang mudah diakses, menampilkan hasil prediksi, dan mendukung pemantauan pasien secara real-time (Buga *et al.*, 2025).

Meskipun berbagai penelitian telah menerapkan algoritma Naïve Bayes untuk klasifikasi data penyalahgunaan narkoba dan substance use disorder, sebagian besar penelitian masih berfokus pada evaluasi model klasifikasi secara terpisah tanpa integrasi ke dalam sistem monitoring yang dapat digunakan secara langsung oleh petugas rehabilitasi. Selain itu, penelitian terkait prediksi relaps umumnya menitikberatkan pada performa model machine learning, namun belum banyak yang menggabungkan prediksi risiko relaps dengan antarmuka berbasis web yang mendukung pencatatan riwayat pasien, visualisasi risiko, dan pembuatan laporan otomatis secara real-time.

Kesenjangan penelitian (research gap) tersebut menunjukkan perlunya pengembangan sistem monitoring terintegrasi yang tidak hanya mampu melakukan prediksi, tetapi juga mendukung proses pengambilan keputusan operasional di pusat rehabilitasi.

Kebaruan (novelty) penelitian ini terletak pada integrasi algoritma Gaussian Naïve Bayes dengan sistem monitoring berbasis web menggunakan Streamlit dan MySQL untuk prediksi risiko relaps pasien rehabilitasi narkoba. Sistem ini tidak hanya menghasilkan klasifikasi

risiko, tetapi juga menyediakan fitur monitoring longitudinal pasien, evaluasi model secara real-time, serta laporan otomatis yang mendukung petugas rehabilitasi dalam melakukan tindak lanjut.

Berdasarkan kebutuhan tersebut, penelitian ini mengimplementasikan algoritma Gaussian Naïve Bayes pada sistem Analisis Risiko Relaps, yaitu sistem monitoring pasien rehabilitasi narkoba berbasis Python dan Streamlit. Sistem menggunakan enam variabel prediktor, yaitu usia, riwayat pemakaian, hasil tes urine, lama pemakaian, status kerja, dan dukungan keluarga. Variabel tersebut diproses melalui tahapan encoding, standardisasi, penanganan ketidakseimbangan data menggunakan SMOTE, pelatihan model, pengujian model, dan prediksi risiko.

2. METODE

Penelitian ini menggunakan pendekatan Research and Development (R&D) dengan orientasi pengembangan sistem berbasis machine learning. Produk yang dikembangkan adalah aplikasi Analisis Risiko Relaps, yaitu sistem monitoring pasien rehabilitasi narkoba yang mengimplementasikan algoritma Gaussian Naïve Bayes untuk memprediksi risiko relaps pasien. Pendekatan pengembangan dipilih karena penelitian tidak hanya menguji algoritma, tetapi juga menghasilkan sistem aplikasi yang dapat digunakan oleh petugas rehabilitasi dalam proses input data, monitoring pasien, evaluasi model, dan pencetakan laporan.

Pengembangan sistem dilakukan menggunakan bahasa pemrograman Python, framework Streamlit, dan basis data MySQL. Pemilihan Python didasarkan pada ketersediaan pustaka machine learning yang mendukung proses prapemrosesan data, pelatihan model, evaluasi model, dan integrasi ke aplikasi. Streamlit digunakan sebagai antarmuka aplikasi berbasis web karena mendukung penyajian model prediktif secara interaktif. Pemanfaatan sistem pendukung keputusan berbasis web dalam layanan kesehatan dapat membantu proses klasifikasi, diagnosis awal, serta rekomendasi tindakan berbasis data (Yaseliani, Soryani and Setarehdan, 2024).

Data penelitian berasal dari data pasien rehabilitasi narkoba yang tersimpan pada

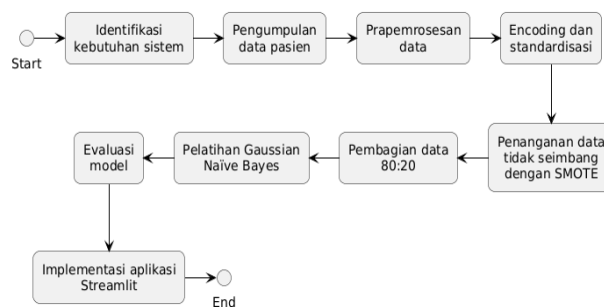
database db_rehabilitasi, khususnya tabel data_pasien. Data bersifat dinamis karena dapat bertambah melalui input manual maupun unggahan file CSV. Variabel prediktor yang digunakan meliputi usia, riwayat pakai, tes urine, lama pakai, status kerja, dan dukungan keluarga. Variabel target adalah hasil prediksi risiko relaps, yang diklasifikasikan menjadi dua kategori, yaitu Risiko Rendah dan Risiko Tinggi.

Dataset penelitian dikumpulkan dari data pemeriksaan pasien rehabilitasi narkoba di Yayasan Sosial Bunda Meifa selama periode [Januari 2024–Mei 2025]. Dataset awal terdiri atas 87 record pemeriksaan dari 76 pasien unik. Setelah dilakukan preprocessing berupa pemeriksaan missing value, validasi format data, dan pembersihan atribut yang tidak relevan, seluruh 87 record dinyatakan layak digunakan untuk proses analisis. Distribusi kelas awal menunjukkan 44 data termasuk kategori Risiko Tinggi dan 32 data termasuk kategori Risiko Rendah, sehingga terdapat ketidakseimbangan kelas yang perlu ditangani pada tahap pelatihan model.

Untuk mencegah data leakage, pipeline eksperimen dilakukan secara berurutan. Dataset terlebih dahulu dibagi menjadi data latih (80%) dan data uji (20%) menggunakan stratified split agar proporsi kelas tetap terjaga. Proses standarisasi menggunakan StandardScaler dilakukan dengan fitting hanya pada data latih, kemudian parameter scaler diterapkan pada data uji. Teknik SMOTE diterapkan hanya pada data latih setelah standarisasi untuk menyeimbangkan distribusi kelas. Data uji tidak mengalami oversampling agar tetap merepresentasikan data nyata. Pendekatan ini dilakukan untuk memastikan evaluasi model bersifat objektif dan dapat direplikasi.

Model klasifikasi yang digunakan adalah Gaussian Naïve Bayes. Algoritma ini dipilih karena memiliki proses komputasi yang ringan, sederhana, dan sesuai untuk klasifikasi berbasis atribut. Penggunaan Naïve Bayes juga banyak diterapkan pada masalah klasifikasi data kesehatan karena mampu menghasilkan prediksi secara cepat dengan struktur data yang relatif terbatas. Kombinasi Naïve Bayes dan teknik penanganan data tidak seimbang dapat meningkatkan kemampuan model dalam

mengenali kelas minoritas pada klasifikasi kesehatan (Wibowo, 2025).



Gambar 1. Tahapan Penelitian

Tahapan penelitian dimulai dari identifikasi kebutuhan sistem dan pengumpulan data pasien rehabilitasi yang mencakup identitas pasien, riwayat penggunaan narkoba, hasil tes urine, lama pemakaian, status pekerjaan, dukungan keluarga, dan status risiko relaps. Data historis digunakan sebagai data latih, sedangkan data pasien baru digunakan untuk implementasi prediksi risiko relaps.

Tahap berikutnya adalah prapemrosesan data. Data diperiksa untuk memastikan tidak terdapat nilai kosong, kesalahan format, atau atribut yang tidak sesuai. Data kategorikal dikonversi menjadi numerik melalui encoding, misalnya kategori Negatif, Kerja, Baik, dan < 30 hari diberi nilai 0, sedangkan Positif, Tidak Kerja, Buruk, dan > 30 hari diberi nilai 1. Data numerik kemudian distandarisasi menggunakan StandardScaler agar seluruh variabel berada pada skala yang seimbang. Rumus standarisasi adalah:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

di mana Z merupakan nilai hasil standarisasi, X merupakan nilai asli data, μ merupakan nilai rata-rata, dan σ merupakan standar deviasi. Standarisasi diperlukan agar variabel numerik seperti usia dan lama pemakaian tidak mendominasi variabel biner.

Setelah itu, dilakukan penanganan ketidakseimbangan kelas menggunakan Synthetic Minority Over-sampling Technique atau SMOTE. Teknik ini digunakan apabila distribusi data risiko rendah dan risiko tinggi tidak seimbang. SMOTE membentuk data sintesis pada kelas minoritas agar model tidak terlalu dominan mengenali kelas mayoritas. Rumus sederhana SMOTE adalah:

$$X_{new} = X_i + \lambda(X_{nn} - X_i) \quad (2)$$

di mana X_{new} merupakan data sintetis yang dibentuk, X_i merupakan data dari kelas minoritas, X_{nn} merupakan data tetangga terdekat dari kelas minoritas, dan λ merupakan nilai acak antara 0 sampai 1. Ketidakseimbangan kelas perlu ditangani karena dapat menurunkan sensitivitas model terhadap kelas minoritas (Li *et al.*, 2025).

Setelah penerapan SMOTE pada data latih, distribusi kelas menjadi lebih seimbang antara kelas Risiko Rendah dan Risiko Tinggi. Penyeimbangan ini bertujuan untuk meningkatkan sensitivitas model terhadap kelas minoritas sehingga model tidak bias terhadap kelas mayoritas selama proses pelatihan.

Data yang telah siap kemudian dibagi menjadi data latih dan data uji dengan rasio 80:20. Data latih digunakan untuk membentuk pola klasifikasi, sedangkan data uji digunakan untuk menilai kemampuan model dalam mengenali data baru. Model yang digunakan adalah Gaussian Naïve Bayes, yaitu algoritma berbasis Teorema Bayes dengan asumsi bahwa setiap atribut bersifat independen. Rumus dasar Teorema Bayes adalah:

$$P(C|X) = \frac{P(X|C)P(C)}{P(X)} \quad (3)$$

di mana $P(C|X)$ merupakan probabilitas kelas C berdasarkan data X, $P(X|C)$ merupakan probabilitas data X pada kelas C, $P(C)$ merupakan probabilitas awal kelas, dan $P(X)$ merupakan probabilitas data X.

Pemilihan rasio 80:20 dilakukan karena proporsi ini umum digunakan dalam supervised learning untuk memberikan jumlah data latih yang cukup besar dalam membentuk pola klasifikasi, sekaligus mempertahankan sebagian data sebagai data uji independen. Untuk menghindari data leakage, pembagian data dilakukan sebelum proses standarisasi dan oversampling. Standarisasi menggunakan StandardScaler dilakukan dengan fitting hanya pada data latih, kemudian parameter yang sama diterapkan pada data uji. Selanjutnya, teknik SMOTE diterapkan hanya pada data latih untuk menyeimbangkan distribusi kelas tanpa memengaruhi karakteristik data uji.

Pada Gaussian Naïve Bayes, atribut numerik diasumsikan mengikuti distribusi normal. Peluang atribut terhadap suatu kelas dihitung menggunakan fungsi Gaussian berikut:

$$P(x_i|C) = \frac{1}{\sqrt{2\pi\sigma_C^2}} e^{-\frac{(x_i-\mu_C)^2}{2\sigma_C^2}} \quad (4)$$

di mana $P(x_i|C)$ merupakan probabilitas atribut x_i pada kelas C, x_i merupakan nilai atribut, μ_C merupakan rata-rata atribut pada kelas C, σ_C^2 merupakan varians atribut pada kelas C, dan e merupakan konstanta eksponensial. Kelas dengan probabilitas terbesar dipilih sebagai hasil prediksi, yang dapat dituliskan sebagai berikut:

$$C_{prediksi} = \arg \max_C P(C) \prod_{i=1}^n P(x_i|C) \quad (5)$$

di mana $C_{prediksi}$ merupakan kelas hasil prediksi, $\arg \max$ merupakan proses memilih kelas dengan nilai probabilitas terbesar, $P(C)$ merupakan probabilitas awal kelas, $P(x_i|C)$ merupakan probabilitas atribut ke- i terhadap kelas C, dan n merupakan jumlah atribut.

Evaluasi model dilakukan menggunakan accuracy, precision, recall, dan confusion matrix. Rumus metrik evaluasi yang digunakan adalah:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

di mana TP merupakan data risiko tinggi yang benar diprediksi sebagai risiko tinggi, TN merupakan data risiko rendah yang benar diprediksi sebagai risiko rendah, FP merupakan data risiko rendah yang salah diprediksi sebagai risiko tinggi, dan FN merupakan data risiko tinggi yang salah diprediksi sebagai risiko rendah. Evaluasi diperlukan untuk memastikan model layak digunakan sebelum diterapkan dalam sistem operasional (Navaei and Tabrizi, 2022).

Selain evaluasi menggunakan train-test split, validasi model juga diperkuat dengan Stratified K-Fold Cross Validation sebanyak 5

fold mengingat ukuran dataset relatif kecil. Metode ini membagi dataset ke dalam lima subset dengan proporsi kelas yang tetap seimbang pada setiap fold. Model dilatih dan diuji sebanyak lima iterasi, di mana setiap subset digunakan sebagai data uji satu kali. Nilai rata-rata accuracy, precision, recall, dan F1-score dari seluruh fold digunakan untuk menilai stabilitas dan generalisasi model secara lebih robust.

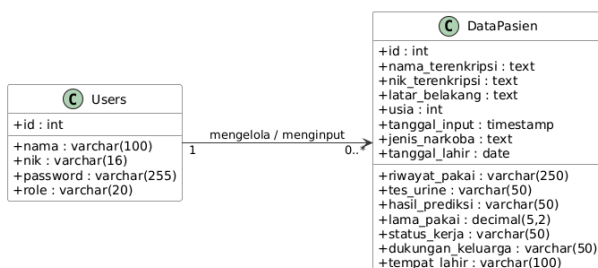
Tahap akhir adalah implementasi model Gaussian Naïve Bayes ke dalam aplikasi Streamlit. Sistem menyediakan fitur login, input prediksi, riwayat pasien, grafik tren pasien, import data latih, evaluasi model, manajemen akun, dan cetak laporan PDF.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Implementasi Sistem

Hasil penelitian ini berupa aplikasi atau sistem Analisis Risiko Relaps, yaitu sistem monitoring pasien rehabilitasi narkoba berbasis machine learning yang dikembangkan menggunakan bahasa pemrograman Python dengan framework Streamlit. Sistem ini dirancang sebagai sistem pakar atau pendukung keputusan yang membantu petugas rehabilitasi dalam memprediksi tingkat risiko relaps pasien berdasarkan data klinis dan sosial. Sistem ini diterapkan pada lingkungan rehabilitasi Yayasan Sosial “Bunda Meifa”.

3.1.1. Perancangan Struktur Sistem



Gambar 2. Class Diagram

Class diagram memiliki dua kelas utama, yaitu Users dan DataPasien. Kelas Users memuat data pengguna sistem, seperti id, nama, nik, password, dan role, yang digunakan untuk autentikasi dan pengaturan hak akses. Kelas DataPasien memuat data pasien rehabilitasi, meliputi identitas terenkripsi, usia, riwayat pemakaian, hasil tes urine, lama pakai, status

kerja, dukungan keluarga, jenis narkoba, hasil prediksi, dan tanggal input. Data ini menjadi dasar proses monitoring dan prediksi risiko relaps. Relasi antara Users dan DataPasien bersifat one to many, artinya satu pengguna dapat mengelola banyak data pasien sesuai kewenangannya.

3.1.2. Hasil Implementasi Sistem

Aplikasi Analisis Risiko Relaps memiliki beberapa fitur utama yang mendukung proses kerja petugas rehabilitasi.

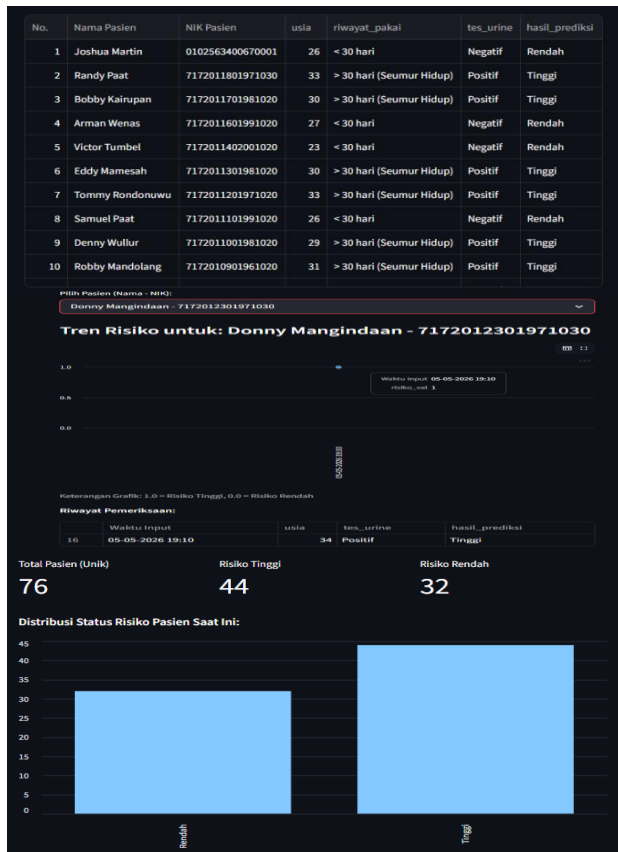


Gambar 3. Sistem autentikasi keamanan

Gambar 3 adalah hasil implementasi sistem autentikasi keamanan yang berupa halaman login yang membatasi akses hanya kepada pengguna terdaftar. Pengguna masuk ke sistem menggunakan NIK dan password. Autentikasi ini berfungsi untuk menjaga keamanan data pasien dan memastikan bahwa sistem hanya digunakan oleh petugas yang memiliki kewenangan.

Gambar 4. Halaman Input prediksi

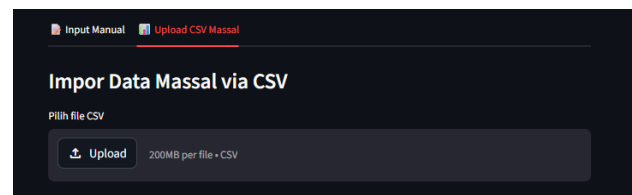
Pada gambar 4. digunakan untuk memasukkan data pasien secara langsung. Data yang dimasukkan meliputi biodata pasien, riwayat pemakaian, hasil tes urine, lama pemakaian, status kerja, dan dukungan keluarga. Setelah data dikirim, sistem melakukan proses prediksi secara otomatis dan menampilkan hasil klasifikasi risiko relaps. Hasil yang ditampilkan terdiri atas dua kategori, yaitu Risiko Rendah dan Risiko Tinggi.



Gambar 5. Halaman riwayat dan perbandingan

Pada gambar 5. menampilkan rekam jejak pemeriksaan pasien yang telah tersimpan di dalam database. Petugas dapat melihat daftar pasien, hasil prediksi, dan perkembangan risiko dari waktu ke waktu. Pada menu ini juga tersedia dasbor rekapitulasi jumlah pasien dan distribusi status risiko. Pada dasbor rekapitulasi, sistem menampilkan data sebagai sampel sebanyak 76 pasien unik. Dari jumlah tersebut, sebanyak 44 pasien berada pada kategori risiko tinggi, sedangkan 32 pasien berada pada kategori risiko rendah. Distribusi ini menunjukkan bahwa jumlah pasien dengan risiko relaps tinggi lebih dominan dibandingkan pasien berisiko rendah. Temuan ini dapat menjadi dasar awal bagi petugas rehabilitasi untuk memprioritaskan

pemantauan, konseling intensif, dan penguatan dukungan keluarga pada kelompok pasien yang masuk kategori risiko tinggi.



Gambar 6. Kelola knowledge base

Gambar 6, merupakan hasil implementasi sisten untuk import data lama atau kelola knowledge base. Menu ini digunakan untuk menambah data pembelajaran bagi model machine learning. Petugas dapat memasukkan data historis secara manual atau mengunggah file CSV secara massal. Fitur ini penting karena sistem membutuhkan data historis untuk membentuk pola klasifikasi. Semakin banyak data historis yang valid, semakin baik basis pembelajaran yang dimiliki sistem.

	precision	recall	f1-score	support	
0		1	1	1	7
1		1	1	1	11
accuracy	1	1	1	1	1
macro avg	1	1	1	1	18
weighted avg	1	1	1	1	18

Gambar 7. Evaluasi Model

Pada gambar 7, menampilkan performa algoritma secara real-time. Halaman evaluasi model, sistem menampilkan hasil pengujian algoritma Gaussian Naïve Bayes dengan akurasi keseluruhan sebesar 100% pada data uji. Confusion matrix menunjukkan bahwa 7 data aktual risiko rendah berhasil diprediksi sebagai risiko rendah, dan 11 data aktual risiko tinggi berhasil diprediksi sebagai risiko tinggi. Tidak terdapat kesalahan prediksi pada data uji yang ditampilkan.

Nilai precision, recall, dan F1-score pada kedua kelas juga mencapai 1,00. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan data uji secara tepat pada sampel pengujian tersebut. Meski demikian, akurasi sempurna perlu dibaca secara hati-hati karena jumlah data uji masih terbatas, yaitu 18 data, sehingga pengujian dengan jumlah data yang lebih besar tetap diperlukan untuk memastikan kestabilan performa model.

3.1.2. Hasil Implementasi Hak Akses

Sistem Sistem memiliki dua role pengguna, yaitu Admin dan Petugas. Admin memiliki akses penuh ke seluruh fitur, termasuk manajemen akun, pengelolaan data pasien, prediksi, riwayat, data latih, evaluasi model, dan cetak laporan. Petugas hanya memiliki akses operasional, seperti input prediksi, melihat riwayat pasien, mengelola data latih, mengevaluasi model, dan mencetak laporan. Pembatasan akses ini diterapkan untuk menjaga keamanan sistem, terutama pada pengelolaan akun dan kredensial pengguna.

Tabel 1. Pembagian Hak akses pengguna

Fitur	Admin	Petugas
Login	Dapat mengakses	Dapat mengakses
Input prediksi pasien	Dapat mengakses	Dapat mengakses
Riwayat dan monitoring	Dapat mengakses	Dapat mengakses
Grafik tren pasien	Dapat mengakses	Dapat mengakses
Import data lama	Dapat mengakses	Dapat mengakses
Evaluasi model	Dapat mengakses	Dapat mengakses
Cetak laporan PDF	Dapat mengakses	Dapat mengakses
Edit dan hapus data	Dapat mengakses	Dapat mengakses sesuai kebutuhan operasional
Manajemen akun	Dapat mengakses	Tidak dapat mengakses

3.1.3. Hasil Prediksi Risiko Relaps

Hasil utama dari sistem adalah klasifikasi risiko relaps pasien rehabilitasi narkoba. Sistem menghasilkan dua kategori prediksi, yaitu Risiko Rendah dan Risiko Tinggi. Prediksi diperoleh dari proses perhitungan probabilitas menggunakan algoritma Gaussian Naïve Bayes terhadap enam variabel prediktor, yaitu usia, riwayat pakai, hasil tes urine, lama pakai, status kerja, dan dukungan keluarga.

Gambar 8. Halaman input pasien untuk Analisis Risiko Relaps

Pasien dikategorikan sebagai Risiko Tinggi apabila kombinasi atribut yang dimiliki menunjukkan kecenderungan kuat terhadap relaps. Kondisi ini dapat muncul ketika pasien memiliki hasil tes urine positif, riwayat pemakaian yang lebih panjang, tidak bekerja, dan dukungan keluarga yang buruk. Pasien dikategorikan sebagai Risiko Rendah apabila atribut yang dimiliki menunjukkan kondisi yang lebih stabil, seperti hasil tes urine negatif, status kerja aktif, dan dukungan keluarga baik.

Gambar 9. Hasil Prediksi

Hasil prediksi disajikan secara langsung setelah petugas memasukkan data pasien. Sistem juga menyimpan hasil tersebut ke dalam database sebagai bagian dari riwayat monitoring. Penyimpanan ini membuat perkembangan pasien dapat dipantau pada pemeriksaan berikutnya.

3.1.4. Hasil Implementasi Laporan Otomatis

Output penting dari sistem adalah Laporan Evaluasi Risiko Relaps dalam format PDF. Laporan ini dihasilkan secara otomatis melalui

fitur cetak laporan. Isi laporan meliputi identitas pasien, data pemeriksaan, hasil prediksi Naïve Bayes, interpretasi risiko, rekomendasi tindakan, nama petugas, dan waktu pencetakan laporan.

YAYASAN SOSIAL 'BUNDA MEIFA'
Jl. Mapalus, Ling. 3, Kel. Watulambot Kec. Tondano Barat, Kab. Minahasa, Prov. Sulawesi Utara.
Email: Meifaawarokka14@gmail.com; Telp. 085240473335

LAPORAN EVALUASI RISIKO RELAPS

I. DATA PASIEN & RIWAYAT KLINIS

Nama Lengkap	: Andre Sangihe	Status Kerja	: Tidak
NIK	: 7172010601991010	Riwayat Pakai	: < 30 hari
Tempat, Tgl Lahir	: Pineleng, 2001-06-06	Dukungan Kel.	: Baik
Usia	: 24 Tahun		
Lama Pemakaian	: 8 Bulan		

II. INTERPRETASI HASIL (NAIVE BAYES)

Tingkat Risiko Relaps	RENDAH
Estimasi Sisa Hari Aman	60 Hari

III. REKOMENDASI TINDAKAN

Hasil analisis menunjukkan risiko relaps yang RENDAH. Kondisi pasien saat ini cenderung stabil dengan estimasi aman sekitar 60 hari. Pertahankan pola hidup sehat dan dukungan sosial yang ada saat ini.

Gambar 10. Laporan evaluasi resiko

Laporan PDF dirancang dengan format kop surat yayasan sehingga dapat digunakan sebagai dokumen resmi internal. Fitur ini membantu petugas membuat laporan secara cepat tanpa harus menyusun ulang data secara manual. Laporan yang tersimpan juga dapat digunakan sebagai arsip pemantauan, bahan rapat evaluasi, atau dokumen pendukung tindak lanjut pasien.

3.2. Pembahasan

Implementasi Analisis Risiko Relaps menunjukkan bahwa sistem monitoring berbasis machine learning dapat membantu petugas rehabilitasi membaca risiko relaps pasien secara lebih cepat dan terstruktur. Sistem mengolah data klinis dan sosial pasien menjadi dua kategori prediksi, yaitu Risiko Rendah dan Risiko Tinggi. Penelitian ini menggunakan 87 data input pasien sebagai sampel, yaitu total seluruh entri pemeriksaan yang masuk ke sistem. Pada dasbor, sistem menampilkan 76 pasien unik karena beberapa pasien memiliki lebih dari satu riwayat pemeriksaan. Dari 76 pasien unik tersebut, sebanyak 44 pasien masuk kategori risiko tinggi dan 32 pasien masuk kategori risiko rendah. Kondisi ini menunjukkan bahwa pasien risiko tinggi lebih dominan sehingga perlu menjadi prioritas pemantauan dan pendampingan.

Penggunaan Gaussian Naïve Bayes sesuai dengan kebutuhan sistem karena algoritma ini ringan, cepat, dan dapat digunakan untuk klasifikasi berbasis atribut. Variabel yang digunakan meliputi usia, riwayat pakai, tes urine, lama pakai, status kerja, dan dukungan keluarga. Atribut tersebut diproses untuk menghitung probabilitas pasien berada pada kelas risiko rendah atau risiko tinggi. Penerapan Naïve Bayes pada sistem pendukung keputusan kesehatan relevan karena mampu membantu klasifikasi risiko secara praktis berbasis data (Kurniawan *et al.*, 2025).

Berdasarkan hasil klasifikasi dan pola data yang diperoleh, beberapa variabel menunjukkan kontribusi yang lebih dominan terhadap prediksi risiko relaps pasien. Analisis variabel menunjukkan bahwa hasil tes urine merupakan indikator paling kuat dalam prediksi risiko relaps. Pasien dengan hasil tes urine positif cenderung diklasifikasikan ke dalam kelompok risiko tinggi karena menunjukkan kemungkinan penggunaan zat kembali setelah rehabilitasi.

Lama pemakaian juga berkontribusi signifikan terhadap peningkatan risiko relaps. Semakin lama riwayat penggunaan narkoba, semakin besar kemungkinan terbentuknya ketergantungan kronis yang mempersulit proses pemulihan.

Status pekerjaan berpengaruh terhadap stabilitas psikososial pasien. Pasien yang tidak memiliki pekerjaan cenderung memiliki risiko relaps lebih tinggi akibat tekanan ekonomi, rendahnya rutinitas harian, dan meningkatnya paparan lingkungan negatif.

Dukungan keluarga menjadi faktor protektif penting dalam rehabilitasi. Pasien dengan dukungan keluarga yang baik cenderung memiliki risiko relaps lebih rendah karena memperoleh pengawasan, motivasi, dan dukungan emosional selama proses pemulihan. Temuan ini sejalan dengan teori rehabilitasi adiksi yang menyatakan bahwa faktor biologis, psikologis, dan sosial secara bersama-sama memengaruhi keberhasilan pemulihan pasien.

Penerapan SMOTE membantu mengatasi kemungkinan ketidakseimbangan data antara kelas risiko rendah dan risiko tinggi. Teknik ini digunakan agar model tidak terlalu condong pada kelas mayoritas dan tetap mampu mengenali kelas minoritas. Hasil evaluasi

menunjukkan nilai accuracy, precision, recall, dan F1-score sebesar 1,00 pada 18 data uji. Confusion matrix menunjukkan 7 data risiko rendah dan 11 data risiko tinggi berhasil diklasifikasikan dengan tepat. Hasil ini menunjukkan performa model yang sangat baik pada sampel uji, tetapi tetap perlu diuji dengan jumlah data lebih besar agar kestabilannya dapat dipastikan (Aryuni *et al.*, 2023; Hairani *et al.*, 2024).

Untuk memperkuat evaluasi model pada dataset yang relatif kecil, validasi tambahan menggunakan Stratified K-Fold Cross Validation dilakukan guna mengurangi bias akibat pembagian data tunggal. Pendekatan ini membantu mengevaluasi konsistensi performa model pada beberapa subset data yang berbeda. Melalui validasi silang, performa model dapat dinilai secara lebih robust sehingga hasil evaluasi tidak hanya bergantung pada satu skenario train-test split. Pendekatan ini penting untuk memastikan bahwa performa tinggi yang diperoleh pada data uji tidak semata-mata disebabkan oleh distribusi data yang kebetulan menguntungkan model.

Dari sisi sistem, pembagian role Admin dan Petugas mendukung keamanan pengelolaan data. Admin memiliki akses penuh, sedangkan petugas hanya mengakses fitur operasional. Pengaturan ini penting karena sistem menyimpan data sensitif pasien, termasuk identitas terenkripsi dan riwayat pemeriksaan. Secara umum, aplikasi ini mampu mengintegrasikan prediksi risiko, monitoring riwayat pasien, visualisasi distribusi risiko, evaluasi model, dan cetak laporan PDF. Sistem tetap diposisikan sebagai alat bantu keputusan, sedangkan keputusan akhir tetap memerlukan asesmen petugas rehabilitasi yang memahami kondisi pasien secara langsung (Cobrado *et al.*, 2024; Revathy, Sindhuja and Jayashree, 2025).

4. PENUTUP

4.1. Kesimpulan

Penelitian ini menghasilkan aplikasi Analisis Risiko Relaps, yaitu sistem monitoring pasien rehabilitasi narkoba berbasis machine learning dengan algoritma Gaussian Naïve Bayes. Sistem dikembangkan menggunakan Python, Streamlit, dan MySQL untuk membantu

petugas rehabilitasi dalam input data pasien, prediksi risiko relaps, monitoring riwayat pemeriksaan, evaluasi model, serta pencetakan laporan otomatis.

Hasil implementasi menunjukkan bahwa sistem mampu mengklasifikasikan pasien ke dalam dua kategori, yaitu Risiko Rendah dan Risiko Tinggi, berdasarkan variabel usia, riwayat pakai, tes urine, lama pakai, status kerja, dan dukungan keluarga. Dari 87 data pemeriksaan dengan 76 pasien unik, terdapat 44 pasien risiko tinggi dan 32 pasien risiko rendah. Evaluasi model pada 18 data uji menghasilkan nilai accuracy, precision, recall, dan F1-score sebesar 1,00. Hasil ini menunjukkan model bekerja sangat baik pada sampel uji, tetapi tetap perlu diuji dengan data yang lebih besar agar performanya lebih stabil dan representatif.

Penelitian ini memiliki beberapa keterbatasan. Pertama, jumlah dataset yang digunakan masih relatif kecil sehingga generalisasi model masih terbatas. Kedua, data hanya berasal dari satu lokasi rehabilitasi sehingga belum merepresentasikan populasi pasien rehabilitasi narkoba yang lebih luas. Ketiga, penelitian ini belum melakukan external validation menggunakan data dari institusi lain. Oleh karena itu, sistem ini harus diposisikan sebagai alat bantu pengambilan keputusan (decision support tool), bukan pengganti asesmen profesional dari petugas rehabilitasi atau tenaga klinis.

4.2. Saran

1. Sistem Analisis Risiko Relaps dapat digunakan sebagai alat bantu awal bagi petugas rehabilitasi dalam mendeteksi risiko relaps secara cepat, terdokumentasi, dan berbasis data. Sistem ini sebaiknya dimanfaatkan untuk mendukung penentuan prioritas pemantauan, konseling, penguatan dukungan keluarga, serta tindak lanjut pasien rehabilitasi narkoba.
2. Penelitian selanjutnya disarankan menggunakan jumlah data yang lebih besar dan variasi kasus yang lebih luas agar hasil evaluasi model lebih akurat dan generalisasi sistem semakin baik. Selain itu, sistem dapat dikembangkan dengan penambahan fitur rekomendasi intervensi yang lebih spesifik, integrasi dengan data rehabilitasi secara

berkala, serta pengujian menggunakan algoritma machine learning lain sebagai pembandingan.

5. DAFTAR PUSTAKA

- Aryuni, M. *et al.* (2023) ‘Imbalanced learning in heart disease categorization: Improving minority class prediction accuracy using the SMOTE algorithm’, *International Journal of Fuzzy Logic and Intelligent Systems*, 23(2), pp. 140–151. Available at: <https://doi.org/10.5391/IJFIS.2023.23.2.140>.
- Buga, R. *et al.* (2025) ‘Streamlit application and deep learning model for brain metastasis monitoring after Gamma Knife treatment’, *Biomedicines*, 13(2), p. 423. Available at: <https://doi.org/10.3390/biomedicines13020423>.
- Cobrado, U.N. *et al.* (2024) ‘Access control solutions in electronic health record systems: A systematic literature review’, *Healthcare Analytics*, 5, p. 100311.
- Hairani, H. *et al.* (2024) ‘Addressing class imbalance of health data: A systematic literature review on SMOTE-based approaches’, *JOIV: International Journal on Informatics Visualization*, 8(2).
- Kurniawan, W.I. *et al.* (2025) ‘Application of Naïve Bayes Classifiers for family risk classification’, *Journal of Applied Informatics and Computing* [Preprint].
- Li, Y. *et al.* (2025) ‘An improved SMOTE algorithm for enhanced imbalanced data classification’, *Scientific Reports*, 15, p. 18495. Available at: <https://doi.org/10.1038/s41598-025-03105-3>.
- Marfalino, H. *et al.* (2023) ‘Penerapan sistem pakar dalam diagnosa pengguna narkoba menggunakan metode Naïve Bayes’, *Jurnal Sains Informatika Terapan*, 2(2), pp. 30–34. Available at: <https://doi.org/10.62357/jsit.v2i2.159>.
- de Mattos, B.P. *et al.* (2024) ‘Craving for a robust methodology: A systematic review of machine learning algorithms on substance-use disorders treatment outcomes’, *International Journal of Mental Health and Addiction* [Preprint]. Available at: <https://doi.org/10.1007/s11469-024-01403-z>.
- Navaei, M. and Tabrizi, N. (2022) ‘Machine learning in software development life cycle: A comprehensive review’, in *Proceedings of the 17th International Conference on Evaluation of Novel Approaches to Software Engineering*, pp. 521–528.
- Nugraha, K.A., Okrpana, H. and Sormin, R.K. (2022) ‘Implementasi algoritma Naïve Bayes dalam mengklasifikasi pasien test urine pada BNN Pematangsiantar’, *ZAHRA: Bulletin Big Data, Data Science, and Artificial Intelligence*, 1(1), pp. 34–43.
- Pamungkas, B.P., Vikri, M.J. and Sa’ida, I.A. (2025) ‘Application of SMOTE-ENN method in data balancing for classification of diabetes health indicators with C4.5 algorithm’, *Jurnal SISFOKOM Sistem Informasi dan Komputer*, 14(2), pp. 183–188. Available at: <https://doi.org/10.32736/sisfokom.v14i2.2350>.
- Puspitasari, N. *et al.* (2024) ‘Classification for determining the level of drugs dependence using the Naïve Bayes Classifier’, *IT Journal Research and Development*, 9(1), pp. 68–79. Available at: <https://doi.org/10.25299/itjrd.2024.16319>.
- Revathy, S.P., Sindhuja, M. and Jayashree, R. (2025) ‘Streamlit-based web application for Parkinson’s detection using machine learning’, *Journal of Artificial Intelligence and Capsule Networks*, 6(4), pp. 466–478. Available at: <https://doi.org/10.36548/jaicn.2024.4.006>.
- Setiawan, Z. *et al.* (2023) *Buku Ajar Data Mining*. PT. Sonpedia Publishing

Indonesia.

- Situmorang, R.D., Sumarno and Hidayati, N. (2022) ‘Penerapan data mining dalam klasifikasi pencegahan narkoba menggunakan algoritma Naïve Bayes di BNN Kota Pematangsiantar’, *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, 1(4), pp. 295–302. Available at: <https://doi.org/10.55123/jomlai.v1i4.1667>.
- Susanto, A.P. *et al.* (2023) ‘Effects of machine learning-based clinical decision support systems on decision-making, care delivery, and patient outcomes: A scoping review’, *Journal of the American Medical Informatics Association*, 30(12), pp. 2050–2063. Available at: <https://doi.org/10.1093/jamia/ocad180>.
- Wibowo, J.J. (2025) ‘A comparative analysis of SMOTE and Naïve Bayes for heart disease classification’, *JOIV: International Journal on Informatics Visualization*, 9(1), pp. 160–167. Available at: <https://doi.org/10.62527/joiv.9.1.3248>.
- Xu, J. *et al.* (2025) ‘The risk factors for relapse behavior in individuals with substance use disorders: An interpretable machine learning study’.
- Yaseliani, M., Soryani, M. and Setarehdan, S.K. (2024) ‘Diagnostic clinical decision support based on deep learning and decision fusion for psoriasis classification and treatment recommendation’, *Computers in Biology and Medicine*, 168, p. 107778. Available at: <https://doi.org/10.1016/j.compbimed.2023.107778>.