

Analisis Komparasi Algoritma KNN dan Naive Bayes untuk Klasifikasi Pasien Rehabilitasi Narkoba di XYZ

Putri Khairunnisa Nabilah*, Triase

Fakultas Sains dan Teknologi, Sistem Informasi, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ¹nabilahputrikhairunnisa@gmail.com, ²triase@uinsu.ac.id

Email Penulis Korespondensi: nabilahputrikhairunnisa@gmail.com

Submitted 29-01-2025; Accepted 19-02-2026; Published 25-02-2026

Abstrak

Perkembangan teknologi informasi dalam bidang data mining membuka peluang signifikan untuk mengoptimalkan klasifikasi data di sektor kesehatan dan rehabilitasi. LRPPN BI Medan saat ini menghadapi kendala dalam menentukan program rehabilitasi karena proses penentuan yang masih bersifat subjektif dan belum memanfaatkan data historis secara sistematis. Ketidaktepatan dalam penentuan program ini berisiko menurunkan efektivitas pemulihan dan meningkatkan potensi kekambuhan pasien. Penelitian ini bertujuan untuk mengaplikasikan dan menguji efektivitas algoritma KNN dan Naive Bayes dalam mengklasifikasikan program rehabilitasi berdasarkan kriteria pasien, seperti lama penggunaan narkoba, tes URICA, riwayat penyakit, dan tingkat kecanduan. Penelitian ini menggunakan pendekatan kuantitatif dengan metode pengembangan Waterfall. Solusi ini diusulkan untuk mengatasi subjektivitas melalui sistem klasifikasi berbasis data yang saling melengkapi dalam akurasi dan kecepatan proses. Hasil penelitian ini ada algoritma Naive Bayes memiliki tingkat akurasi yang lebih tinggi sebesar 96.43%, sedangkan KNN memiliki akurasi sebesar 94.29%.

Kata Kunci: Data Mining; K-Nearest Neighbors; Naive Bayes; Rehabilitasi Narkoba; Klasifikasi

Abstract

The development of information technology in the field of data mining opens up significant opportunities to optimize data classification in the healthcare and rehabilitation sectors. LRPPN BI Medan currently faces challenges in determining rehabilitation programs because the decision-making process is still subjective and has not yet systematically utilized historical data. Inaccuracies in determining these programs can reduce recovery effectiveness and increase the potential for patient relapse. This study aims to apply and test the effectiveness of the KNN and Naive Bayes algorithms in classifying rehabilitation programs based on patient criteria, such as duration of drug use, URICA test results, medical history, and addiction level. This study uses a quantitative approach with the Waterfall development method. This solution is proposed to overcome subjectivity through a data-based classification system that complements each other in accuracy and processing speed. The results of this study show that the Naive Bayes algorithm has a higher accuracy rate of 96.43%, while KNN has an accuracy of 94.29%.

Keywords: Data Mining; K-Nearest Neighbors; Naive Bayes; Drug Rehabilitation; Classification

1. PENDAHULUAN

Di era yang serba canggih ini, teknologi dan informasi berkembang sangat pesat. Teknologi dapat mempermudah aktivitas manusia dan mampu menyediakan informasi yang akurat serta membantu mengorganisasi data instansi. Perkembangan teknologi informasi khususnya dalam bidang data mining telah membuka peluang besar untuk mengolah data yang besar secara efektif dan efisien[1]. Melalui penerapan data mining[2], instansi dapat menganalisis data lebih mendalam dan menghasilkan informasi yang akurat untuk mendukung pengambilan keputusan dalam berbagai bidang, termasuk dalam bidang kesehatan dan rehabilitasi[3][4]. Di bidang rehabilitasi, khususnya rehabilitasi narkoba, data mining berperan penting dalam membantu mengklasifikasikan pasien berdasarkan kriteria-kriteria yang ada di masing-masing lembaga, seperti tingkat kecanduan, riwayat penggunaan zat, kondisi psikologis, serta respons terhadap terapi guna mengembalikan kondisi pasien[5].

Namun, dalam implementasinya, beberapa pasien yang menjalani tahap rehabilitasi menunjukkan hasil kesembuhan yang tidak sesuai dengan prediksi awal yang disebabkan oleh berbagai faktor, salah satunya adalah ketidaktepatan dalam penentuan program rehabilitasi yang diberikan[6]. Penentuan program rehabilitasi pasien masih sering dilakukan secara subjektif tanpa sistem klasifikasi berbasis data[7]. Akibatnya, program yang dijalankan berpotensi menghasilkan output rehabilitasi yang kurang maksimal dan meningkatkan risiko kekambuhan. LRPPN BI Medan sebagai lembaga rehabilitasi masih menghadapi kendala dalam menentukan program rawat inap, rehabilitasi non-medis, dan rawat jalan secara optimal. Masalah ini muncul karena tidak adanya metode klasifikasi sistematis yang memanfaatkan data historis pasien secara komprehensif. Oleh karena itu, dibutuhkan pendekatan ilmiah untuk meningkatkan akurasi klasifikasi program rehabilitasi[8][9][10].

Penelitian ini mengusulkan penerapan algoritma K-Nearest Neighbors (KNN) dan Naive Bayes sebagai metode klasifikasi data pasien rehabilitasi narkoba. KNN merupakan algoritma machine learning berbasis supervised learning yang dapat melakukan klasifikasi dengan membandingkan data baru dengan data historis pasien[11]. Sementara itu, Naive Bayes adalah salah satu algoritma yang menggunakan teori probabilitas[12], yaitu memprediksi probabilitas di masa yang akan datang berdasarkan hasil pengamatan dari pengalaman-pengalaman di masa lalu dengan menggunakan keahlian dan pengetahuan para pakar atau ahli[13]. KNN dan Naive Bayes digunakan bersama dalam penelitian ini karena keduanya saling melengkapi[14][15]. KNN menilai kedekatan data tanpa banyak asumsi, sedangkan Naive Bayes menggunakan perhitungan peluang yang cepat dan sederhana[16][17].

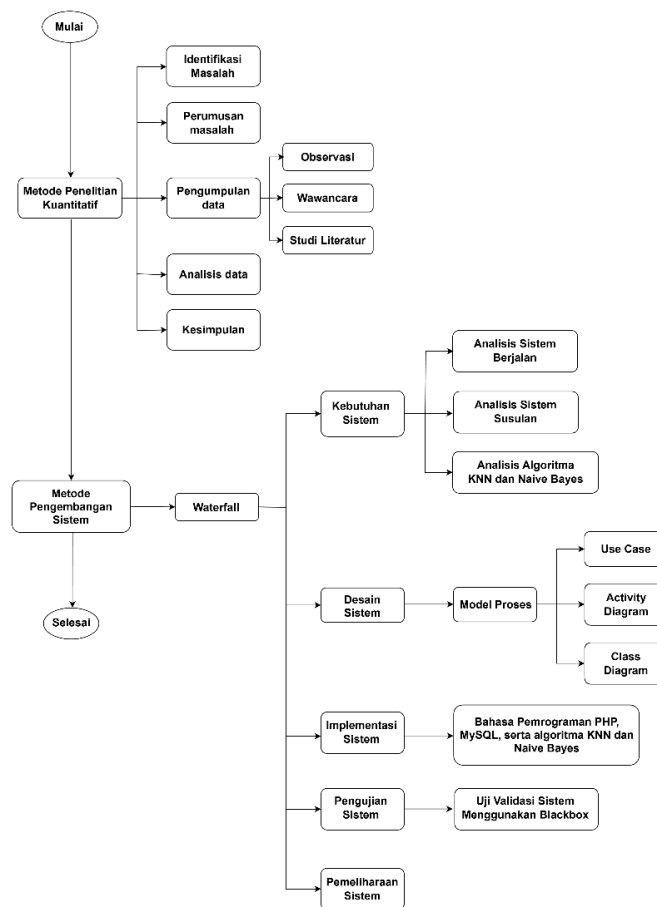
Penelitian terdahulu dilakukan oleh Tursina, Hafiz Muhardi, Dian Aulia Sari (2020) berjudul *Diagnosis Tahapan Pengguna Narkoba Menggunakan Metode K-Nearest Neighbor*. Hasil dari penelitian tersebut metode KNN memiliki tingkat akurasi untuk cross validation sebesar 98,333%, confusion matrix sebesar 100% dan termasuk excellent classification karena memiliki nilai AUC 1,000[18]. Penelitian selanjutnya dilakukan oleh Hari Marfalino, Mutiana Pratiwi, Irzal Arief Wisky, Dinul Akhiyar (2023) berjudul *Penerapan Sistem Pakar dalam Diagnosa Pengguna Narkoba Menggunakan Metode Naïve Bayes*. Hasil yang diperoleh dari penelitian ini adalah terdapat salah satu pengguna narkoba yang terdiagnosa penyalahgunaan narkoba jenis Sabu dengan nilai akurasi 0.4468[19]. Berbeda dengan kedua penelitian sebelumnya, penelitian ini berfokus pada klasifikasi pasien rehabilitasi dan tahapan pemulihan yang berbeda menggunakan algoritma KNN dan Naive Bayes. Keunikan penelitian ini terletak pada konteks penerapannya, yaitu pada lembaga rehabilitasi narkoba, serta penggunaan dua algoritma untuk melihat perbandingan performa dalam menentukan program rehabilitasi yang tepat.

Penelitian ini bertujuan untuk mengaplikasikan dan menguji efektivitas algoritma KNN dan Naive Bayes dalam mengklasifikasikan program rehabilitasi berdasarkan data pasien rehabilitasi narkoba di LRPPN BI Medan. Dengan menerapkan KNN dan Naive Bayes[20] diharapkan dapat diperoleh pola klasifikasi yang lebih akurat dalam mengelompokkan pasien rehabilitasi narkoba berdasarkan faktor-faktor seperti tingkat kecanduan, riwayat penggunaan, dan respons terhadap terapi[21][22]

2. METODOLOGI PENELITIAN

2.1 Metode Penelitian

Metode penelitian mencakup langkah-langkah yang dipilih sesuai tujuan dan jenis penelitian yang dilakukan. Penelitian ini menggunakan metode penelitian kuantitatif. Penelitian kuantitatif adalah suatu pendekatan ilmiah yang dilakukan secara sistematis untuk menganalisis berbagai aspek suatu fenomena serta hubungan kausal di antara variabel-variabel yang terlibat[23]s. Langkah-langkah dalam proses penelitian ini divisualisasikan pada gambar 1.



Gambar 1. Tahapan Penelitian

Diagram pada gambar 1 menggambarkan tahapan metode penelitian kuantitatif dan pengembangan sistem secara terstruktur. Proses dimulai dengan observasi ke LRPPN BI Medan sebagai tempat penelitian yang mencakup wawancara dan studi literatur. Selanjutnya menggunakan metode pengembangan sistem Waterfall[24]. Dimulai dengan analisis kebutuhan, meliputi analisis sistem berjalan yang ada di LRPPN BI Medan, sistem susulan yang akan dibangun, serta

penerapan algoritma KNN dan Naive Bayes untuk pemrosesan data. Setelah analisis, dilakukan desain sistem yang mencakup model proses, use case, activity diagram, dan class diagram.

2.2 Model Kerja Data Sains

2.2.1 Memahami Bisnis

Tahap business understanding dilakukan dengan mengidentifikasi permasalahan pada proses penentuan tingkat keparahan pasien rehabilitasi narkoba di LRPPN BI Medan yang masih dilakukan secara manual, sehingga memerlukan waktu yang lama dan berpotensi menimbulkan ketidakkonsistenan hasil. Berdasarkan permasalahan tersebut, tujuan penelitian ini adalah menganalisis dan membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dan Naive Bayes dalam mengklasifikasikan tingkat keparahan pasien rehabilitasi narkoba[25].

2.2.2 Memahami Data

Tahap data understanding dilakukan dengan mengumpulkan dan mempelajari data pasien rehabilitasi narkoba yang diperoleh dari LRPPN BI Medan. Data yang digunakan mencakup data identitas pasien, riwayat penggunaan narkoba, serta hasil angket tingkat keparahan sebelum menjalani rehabilitasi. Pada tahap ini dilakukan analisis awal untuk mengetahui jumlah data, jenis atribut, serta kondisi data yang digunakan dalam penelitian[26].

2.2.3 Preprocessing

Pada tahap data *preprocessing*, dilakukan proses pembersihan dan pengolahan data agar siap digunakan dalam pemodelan. Proses ini meliputi penghapusan data duplikat, penanganan *missing value*, serta transformasi data ke dalam bentuk numerik. Tahap ini bertujuan untuk meningkatkan kualitas data sehingga tidak memengaruhi hasil klasifikasi[27].

2.2.4 Penerapan Model

Tahap modeling dilakukan dengan menerapkan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes untuk mengklasifikasikan tingkat keparahan pasien rehabilitasi narkoba ke dalam kategori tertentu. Data yang telah dipersiapkan dibagi menjadi data latih dan data uji untuk membangun dan menguji model klasifikasi pada masing-masing algoritma[28].

Pada proses klasifikasi, KNN menghitung jarak antara data uji dengan seluruh data latih menggunakan rumus jarak tertentu, seperti Euclidean Distance. Kemudian, algoritma akan mengambil sejumlah k data terdekat (tetangga terdekat), dan menentukan kelas mayoritas dari tetangga tersebut sebagai hasil prediksi. Berikut urutan proses kerja KNN:

- Menentukan parameter k (jumlah tetangga paling dekat).
- Menghitung kuadrat jarak euclidean (euclidean distance) masing-masing objek terhadap data sampel yang diberikan.

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

Keterangan:

- X_1 = Sampel data
 - X_2 = Data Uji/Testing
 - i = Variabel Data
 - d = Jarak
 - P = Dimensi Data
- Ambil nilai K tetangga terdekat (nilai jarak terkecil)
 - Voting mayoritas kelas tetangga (kelas yang paling sering muncul)
 - Output kelas untuk data uji

Dalam konteks data mining, algoritma Naive Bayes digunakan untuk memprediksi kelas dari suatu data dengan memperhitungkan peluang setiap fitur terhadap masing-masing kelas. Cara kerja algoritma Naive Bayes berikut:

- Hitung probabilitas prior

$$P(C) = \frac{\text{jumlah data pada kelas } C}{\text{total data}} \quad (1)$$

- Hitung likelihood

$$P(X|C) = \frac{\text{jumlah data dengan fitur } X \text{ pada kelas } C}{\text{jumlah data pada kelas } C} \quad (2)$$

- Hitung posterior probabilitas dengan Teorema Bayes

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

Keterangan:

- $P(C|X)$: Probabilitas bebas bersyarat (conditional probability) suatu hipotesis H jika diberikan bukti (Evidence) E terjadi
- $P(X|C)$: Probabilitas sebuah bukti E terjadi akan memengaruhi hipotesis H .

3. $P(C)$: Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun
4. $P(X)$: Probabilitas awal (priori) bukti E terjadi tanpa memandang hipotesis/bukti yang lain
- d. Pilih kelas dengan nilai posterior tertinggi
- e. Output kelas untuk data uji

2.2.5 Evaluasi

Tahap terakhir dalam penelitian ini adalah evaluasi, yaitu mengevaluasi kinerja model klasifikasi yang telah dibangun[29]. Evaluasi dilakukan menggunakan confusion matrix untuk mengetahui nilai akurasi, presisi, recall, dan f-measure dari algoritma KNN dan Naive Bayes. Hasil evaluasi digunakan untuk menentukan algoritma yang memiliki kinerja terbaik dalam mengklasifikasikan tingkat keparahan pasien rehabilitasi narkoba di LRPPN BI Medan.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Sistem Berjalan

Pada sistem yang berjalan saat ini di LRPPN BI Medan, proses pendaftaran dan penanganan awal pasien rehabilitasi narkoba masih dilakukan secara manual dan semi-terkomputerisasi. Seluruh tahapan, mulai dari pendaftaran, pencatatan rekam medis, hingga penentuan tingkat keparahan pasien, belum didukung oleh sistem informasi yang terintegrasi, sehingga berpotensi menimbulkan ketidakefisienan dan kesalahan pengelolaan data.

- a. Pasien melakukan pendaftaran sebagai peserta rehabilitasi narkoba serta pendataan rekam medis oleh dokter dan petugas yang berwenang secara manual dalam dokumen tertulis
- b. Pasien mengisi formulir atau angket berbentuk kertas yang berisi gejala dan riwayat penggunaan narkoba untuk menentukan tingkat keparahan pasien secara manual.
- c. Proses klasifikasi program rehabilitasi pasien belum berbasis sistem informasi terintegrasi.
- d. Pengelolaan data pasien masih dilakukan secara manual dan semi-terkomputerisasi menggunakan aplikasi Microsoft Excel oleh divisi administrasi.

3.2 Analisis Sistem Usulan

Pada sistem usulan yang dirancang, proses pengelolaan data pasien rehabilitasi narkoba dilakukan secara terintegrasi melalui sistem berbasis website guna meningkatkan efisiensi, akurasi, dan konsistensi data.

1. Penginputan data pendaftaran pasien, termasuk identitas, riwayat awal, dan riwayat selama rehabilitasi dilakukan langsung oleh divisi administrasi melalui sistem berbasis website.
2. Formulir tingkat keparahan pasien telah didigitalisasi dan diinput melalui sistem berbasis website yang kemudian diproses secara otomatis untuk penentuan klasifikasi program rehabilitasi

3.3 Preprocessing

Berdasarkan hasil penelitian pada Agustus 2025, didapatkan dataset pasien rehabilitasi narkoba yang bersumber dari LRPPN BI. Dataset pasien yang digunakan pada penelitian ini berjumlah 1399 data dari Januari 2021 hingga Agustus 2025. Dataset terdiri atas beberapa kolom, yaitu jenis kelamin, jenis napza yang digunakan, lama penggunaan, riwayat penyakit, URICA tes, tingkat keparahan, dan program rehabilitasi. Dataset pasien dalam dilihat pada tabel 1.

Tabel 1. Dataset pasien

No	Inisial	Jenis Kelamin	Jenis NAPZA yang digunakan	Lama Penggunaan	Riwayat Penyakit	Urica Tes	Tingkat Keparahannya	Program Rehabilitasi
1	TPW	Laki-Laki	Shabu	8 Tahun	Demam	14	Sedang	Rehabilitasi non-medis
2	YEC	Perempuan	Ganja	2 Tahun	Darah tinggi	19	Ringan	Rawat Jalan
3	RQ	Laki-Laki	Morfin	7 Tahun	Vertigo	14	Sedang	Rehabilitasi Non-Medis
4	RR	Perempuan	Opium, Shabu	3 Tahun	Anemia	20	Ringan	Rawat Jalan
5	RIY	Perempuan	Ekstasi	10 Tahun	Cemas	3	Berat	Rawat Inap
....								
1399	RP	Perempuan	Lem	3 Tahun	Batuk pilek	17	Ringan	Rawat Jalan

Dataset terenkripsi pada tabel 2 menyajikan profil demografis dan klinis subjek rehabilitasi NAPZA, di mana variabel Lama Penggunaan serta Urica Tes menjadi indikator utama dalam menentukan Tingkat Keparahannya. Terdapat korelasi linear yang sistematis antara tingkat keparahan dengan jenis Program Rehabilitasi yang ditetapkan: skor rendah (0) menerima program kategori 0, sedangkan skor yang lebih tinggi (1 dan 2) menerima intervensi kategori 1 dan 2. Hal

ini menunjukkan adanya mekanisme klasifikasi terukur dalam menentukan skema pemulihan pasien berdasarkan parameter input yang telah dikuantifikasi.

Tabel 2. Dataset setelah Encoding

No	Inisial	Jenis Kelamin	Jenis NAPZA yang digunakan	Lama Penggunaan	Riwayat Penyakit	Urica Tes	Tingkat Keparahan	Program Rehabilitasi
1	TPW	0	1	8	1	14	1	1
2	YEC	1	1	2	1	19	0	0
3	RQ	0	1	7	1	14	1	1
4	RR	1	1	3	1	20	0	0
5	RIY	1	1	10	1	3	2	2

3.4 Pemodelan dengan KNN dan Naive Bayes

Tahap modeling dilakukan dengan menerapkan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes untuk mengklasifikasikan tingkat keparahan pasien rehabilitasi narkoba ke dalam kategori tertentu. Dalam konteks data mining, algoritma Naive Bayes digunakan untuk memprediksi kelas dari suatu data dengan memperhitungkan peluang setiap fitur terhadap masing-masing kelas.

Tabel 3 menampilkan contoh sampel data pasien yang digunakan sebagai input dalam proses klasifikasi. Data pasien terdiri dari variabel Gender, Lama Pakai, skor URICA, jenis NAPZA, dan gejala penyakit yang dimasukkan dalam bentuk teks, kemudian dikonversi ke dalam bentuk angka (numerik) agar dapat diproses oleh sistem atau model

Tabel 3. Sampel pasien

Variabel	Input Teks	Input Angka (Mesin)
Gender	Perempuan	1.0
Lama Pakai	8 Tahun	8
URICA	9	0.51
NAPZA	Ekstasi	0.49
Penyakit	Perut kembung	0.0

a. Analisis menggunakan KNN

$$\text{Gender: } (0 - 1)^2 = (-1)^2 = 1$$

$$\text{Lama: } (8 - 5)^2 = (3)^2 = 9$$

$$\text{URICA: } (0.5182 - 0.4)^2 = (0.1182)^2 = 0.0139$$

$$\text{Penyakit: } (0.4909 - 0.3)^2 = (0.1909)^2 = 0.0364$$

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

$$d = \sqrt{1 + 9 + 0.0139 + 0.0364} + \sqrt{10.0503} = 3.17$$

$$d = \sqrt{0.0401}$$

Tabel 4 menyajikan hasil perhitungan metode K-Nearest Neighbor (KNN) untuk menentukan klasifikasi data berdasarkan nilai jarak terdekat:

Tabel 4. Tetangga Terdekat (Rank K)

Rank	Jarak	Kelas Tetangga
1	0.0401	2
2	0.0622	2
3	0.0938	2
4	0.1192	2
5	0.1384	2

Tabel 4 menampilkan hasil perhitungan Tetangga Terdekat (Rank K) yang menunjukkan lima data dengan jarak terdekat terhadap data uji. Nilai jarak terkecil terdapat pada peringkat pertama sebesar 0,0401 dan meningkat secara bertahap hingga 0,1384 pada peringkat kelima.

Hasil prediksi KNN: Sedang

Berdasarkan hasil perhitungan metode K-Nearest Neighbor (KNN), data diklasifikasikan ke dalam kategori Sedang karena mayoritas tetangga terdekat (lima nilai jarak terkecil) berasal dari kelas 2 yang merepresentasikan kategori Sedang.

b. Analisis menggunakan Naive Bayes

1. Likelihood (Peluang Fitur):

$$P(X|C) = \frac{\text{jumlah data dengan fitur X pada kelas C}}{\text{jumlah data pada kelas C}}$$

$$P(\text{Gender} = 1.0 | C): \text{Probabilitas gender perempuan di kelas C.}$$

$P(\text{Lama Pakai} = 8 | C)$: Probabilitas durasi penggunaan tertentu di kelas C.

$P(\text{URICA} = 0.51 | C)$: Probabilitas skor URICA tersebut di kelas C.

$P(\text{NAPZA} = 0.49 | C)$: Probabilitas penggunaan Ekstasi di kelas C.

$P(\text{Penyakit} = 0.0 | C)$: Probabilitas gejala perut kembung di kelas C.

2. Teorema Bayes

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

Hasil Akhir Probabilitas:

- Peluang Berat: 0.16%
- Peluang Ringan: 0%
- Peluang Sedang: 99.84%

Hasil prediksi Naive Bayes: Sedang

Berdasarkan perhitungan metode Naive Bayes, data diklasifikasikan ke dalam kategori Sedang karena memiliki probabilitas tertinggi sebesar 99,84% dibandingkan dengan kategori lainnya.

Setelah melewati tahap pemodelan, diperoleh hasil klasifikasi menggunakan metode Naive Bayes dan KNN dalam menentukan program rehabilitasi bagi pasien penyalahgunaan narkoba pada tabel 5 berikut:

Tabel 5. Klasifikasi KNN dan Naive Bayes

No	Pasien	Hasil KNN	Hasil NB	Status Akhir
1	TPW	Sedang 100%	Sedang 99.8%	Rehabilitasi Non-Medis (Sosial)
2	YEC	Ringan 100%	Ringan 99.4%	Rawat Jalan
3	RQ	Sedang 100%	Sedang 95.7%	Rehabilitasi Non-Medis (Sosial)
4	RR	Ringan 100%	Ringan 99.9%	Rawat Jalan
5	RIY	Berat 80%	Berat 59.1%	Rawat Jalan
6	GOR	Sedang 100%	Sedang 95.3%	Rehabilitasi Non-Medis (Sosial)
7	CXE	Ringan 100%	Ringan 99.4%	Rawat Jalan
8	SDV	Berat 100%	Berat 99.4%	Rawat Inap
9	SDN	Sedang 100%	Sedang 99.6%	Rehabilitasi Non-Medis (Sosial)
10	QQQ	Ringan 100%	Ringan 99.9%	Rawat Jalan
...	...		...	
1399	JYV	Sedang 100%	Sedang 99.6%	Rehabilitasi Non-Medis (Sosial)

Berdasarkan tabel 5, klasifikasi menunjukkan bahwa sebagian besar pasien berada pada kategori Sedang dan Ringan, dengan tingkat akurasi yang sangat tinggi, yaitu mendekati 100% pada kedua metode. Pasien yang diklasifikasikan dalam kategori Sedang oleh metode KNN dan NB selanjutnya ditetapkan menjalani Rehabilitasi Non-Medis (Sosial). Sementara itu, pasien dengan kategori Ringan diarahkan ke Rawat Jalan, karena kondisi pasien masih dapat ditangani tanpa perlu rehabilitasi non-medis secara intensif. Kemudian pasien dalam kategori Berat menjalani program Rawat Inap.

3.5 Evaluasi

Evaluasi adalah tahap untuk mengukur sejauh mana model yang dihasilkan memenuhi tujuan bisnis dan memiliki tingkat akurasi yang valid sebelum diterapkan. Pada tahap ini, performa model dapat dinilai menggunakan *Confusion Matrix* untuk memastikan pola yang ditemukan bukan sekadar kebetulan.

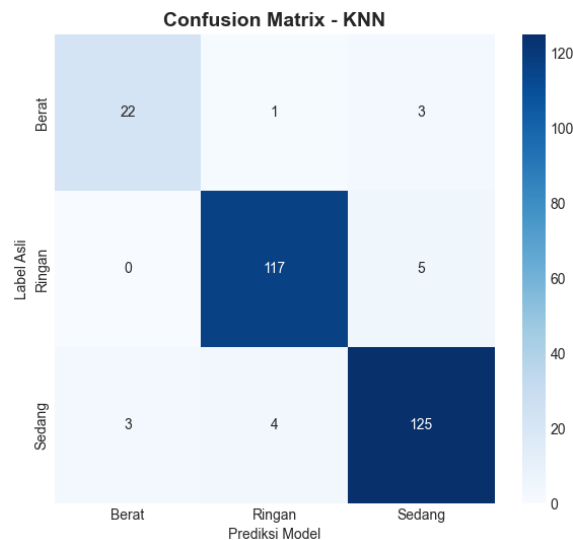
Berikut ini hasil evaluasi model KNN menggunakan *confusion matrix* untuk melihat tingkat akurasi, precision, dan recall pada masing-masing kelas dalam mengklasifikasikan data ke dalam kategori Berat, Ringan, dan Sedang yang terdapat pada tabel 6.

Tabel 6. *Confussion Matrix* KNN

	true Berat	true Ringan	true Sedang	class precision
pred. Berat	22	1	3	88%

pred. Ringan	0	117	5	96%
pred. Sedang	3	4	125	94%
class recall	85%	96%	95%	

Confusion matrix KNN pada tabel 6 tersebut menunjukkan kinerja klasifikasi KNN dalam menentukan tingkat keparahan pasien, yaitu berat, ringan, dan sedang. Nilai precision tertinggi diperoleh pada kelas ringan sebesar 96%, diikuti kelas sedang 94% dan kelas berat 88%. Sementara itu, nilai recall masing-masing kelas menunjukkan hasil yang baik, yaitu 85% untuk berat, 96% untuk ringan, dan 95% untuk sedang. Hasil ini mengindikasikan bahwa model KNN memiliki performa yang cukup akurat dan seimbang dalam mengklasifikasikan tingkat keparahan pasien. Visualisasi *Confussion Matrix* KNN dapat dilihat pada gambar 2.



Gambar 2. Visualisasi *Confussion Matrix* KNN

Visualisasi *confusion matrix* KNN pada gambar 2, berwarna biru mengonfirmasi keberhasilan model dalam memetakan mayoritas data pada garis diagonal, terutama pada kategori “Sedang” yang mencatatkan tingkat identifikasi benar paling tinggi. Berikut adalah hasil evaluasi model KNN berdasarkan metrik precision, recall, dan F1-score pada setiap kelas untuk mengukur tingkat kinerja dan ketepatan model dalam mengklasifikasikan data pasien, tercantum pada tabel 7.

Tabel 7. Evaluasi Model KNN.

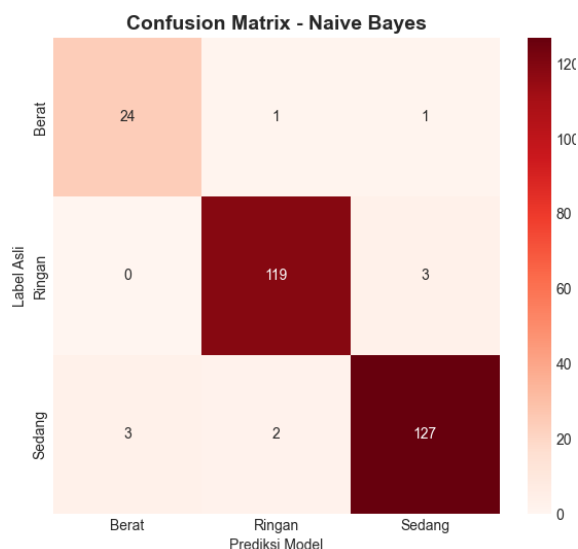
Kelas	Precision	Recall	F1-Score	Support
Berat	0.88	0.85	0.86	26
Ringan	0.96	0.96	0.96	122
Sedang	0.94	0.95	0.94	132
Accuracy			0.94	280

Berikut ini hasil evaluasi model Naive Bayes menggunakan *confusion matrix* untuk melihat tingkat akurasi, precision, dan recall pada masing-masing kelas dalam mengklasifikasikan data ke dalam kategori Berat, Ringan, dan Sedang yang terdapat pada tabel 8.

Tabel 8. *Confussion Matrix* Naive Bayes

	true Berat	true Ringan	true Sedang	class precision
pred. Berat	24	1	1	88%
pred. Ringan	0	119	3	97%
pred. Sedang	3	2	127	96%
class recall	92%	97%	96%	

Confusion matrix Naive Bayes pada tabel 8 menunjukkan kinerja klasifikasi yang baik pada tiga kelas tingkat keparahan, yaitu berat, ringan, dan sedang. Nilai precision tertinggi diperoleh pada kelas ringan sebesar 97%, diikuti kelas sedang 96% dan kelas berat 88%. Sementara itu, nilai recall pada setiap kelas juga tinggi, yaitu 92% untuk berat, 97% untuk ringan, dan 96% untuk sedang. Hasil ini menunjukkan bahwa metode Naive Bayes mampu mengklasifikasikan data dengan tingkat akurasi yang tinggi dan stabil. Visualisasi *Confussion Matrix* Naive Bayes dapat dilihat pada gambar 3.



Gambar 3. Visualisasi *Confusion Matrix* Naive Bayes

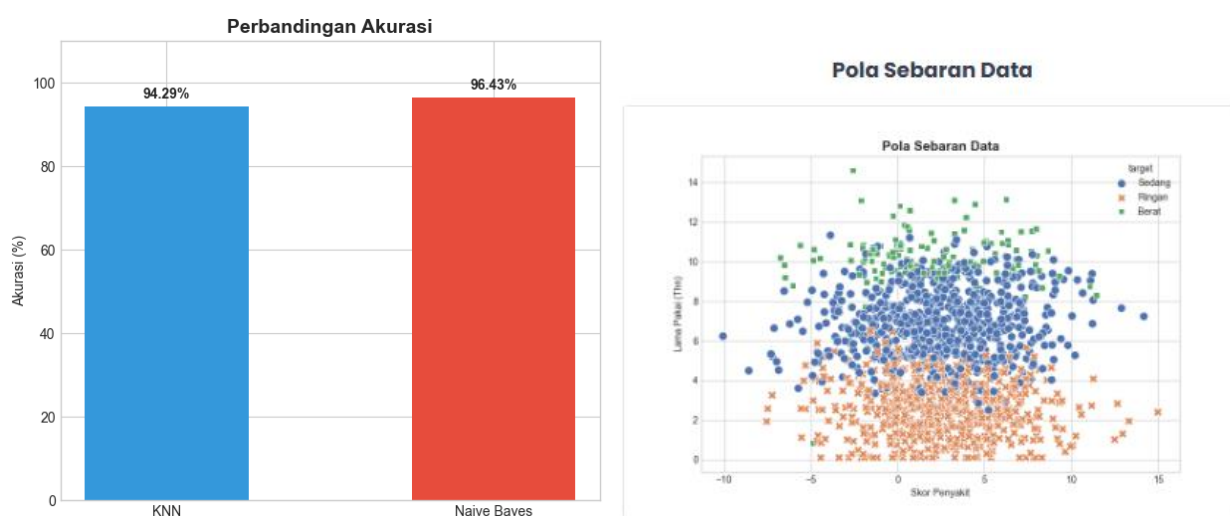
Visualisasi *confusion matrix* Naive Bayes pada gambar 3, berwarna merah menunjukkan konsentrasi prediksi yang sangat akurat di seluruh kelas, di mana model ini terbukti lebih responsif dalam mengidentifikasi kategori “Berat” dengan nilai *recall* mencapai 92%. Berikut adalah hasil evaluasi model Naive Bayes berdasarkan metrik *precision*, *recall*, dan *F1-score* pada setiap kelas untuk mengukur tingkat kinerja dan ketepatan model dalam mengklasifikasikan data pasien, tercantum pada tabel 9.

Tabel 9. Evaluasi Model Naive Bayes

Kelas	Precision	Recall	F1-Score	Support
Berat	0.88	0.92	0.90	26
Ringan	0.97	0.97	0.97	122
Sedang	0.96	0.96	0.96	132
Accuracy			0.96	280

3.6 Perbandingan Model

Perbandingan akurasi antara model KNN dan Naive Bayes serta visualisasi pola sebaran data berdasarkan hasil klasifikasi terdapat pada gambar 4 dan 5. Kedua gambar ini memberikan gambaran mengenai tingkat performa model sekaligus distribusi data pada masing-masing kelas.



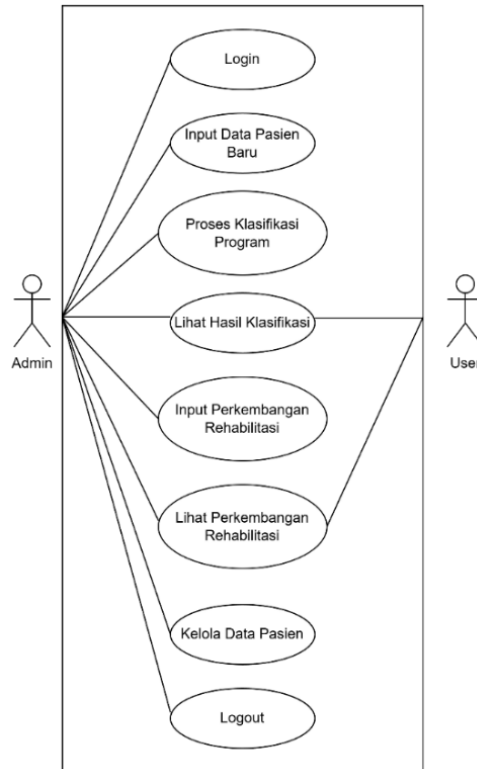
Gambar 4 dan 5. Perbandingan Akurasi Model dan Pola Sebaran Data

Berdasarkan gambar 4, hasil pengujian algoritma Naive Bayes menunjukkan performa yang lebih unggul dengan tingkat akurasi mencapai 96,43%, sementara algoritma KNN menghasilkan akurasi sebesar 94,29%. Secara spesifik, Naive Bayes memiliki kemampuan identifikasi yang lebih konsisten di seluruh kategori. Keseluruhan data ini menegaskan bahwa kedua algoritma sangat reliabel dalam mengklasifikasikan program rehabilitasi pasien berdasarkan parameter klinis yang diberikan.

Gambar 5 menunjukkan pola sebaran data berdasarkan skor penyakit dan lama pakai yang terbagi ke dalam tiga kelas, yaitu Ringan, Sedang, dan Berat. Terlihat bahwa kelas Ringan cenderung berada pada nilai lama pakai dan skor yang lebih rendah, Sedang berada di tengah, sedangkan Berat dominan pada nilai yang lebih tinggi.

3.7 Use case diagram

Diagram use case menggambarkan alur interaksi pengguna dengan sistem klasifikasi program rehabilitasi. Diagram ini menunjukkan fungsi-fungsi utama yang dapat diakses oleh pengguna dalam sistem. Diagram use case dapat dilihat pada gambar 6.



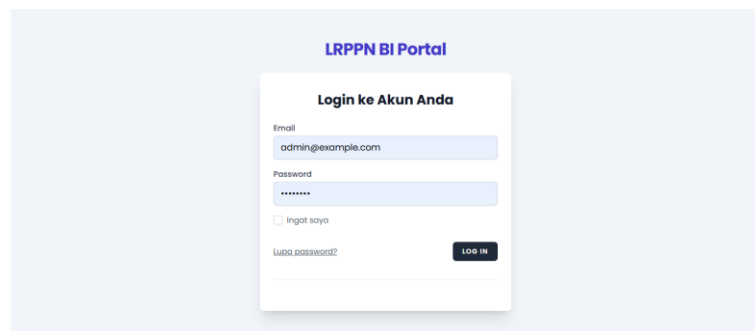
Gambar 6. Use Case Diagram Klasifikasi Program Rehabilitasi

Pada diagram use case yang ditunjukkan pada gambar 6, pengguna sistem (admin) dapat melakukan login terlebih dahulu, kemudian menginput data pasien baru untuk diproses dalam klasifikasi program. Sistem menyediakan fitur untuk melihat hasil klasifikasi, menginput serta memantau perkembangan rehabilitasi, mengelola data pasien, hingga melakukan logout setelah selesai menggunakan sistem.

3.8 Implementasi Sistem

a. Halaman Login

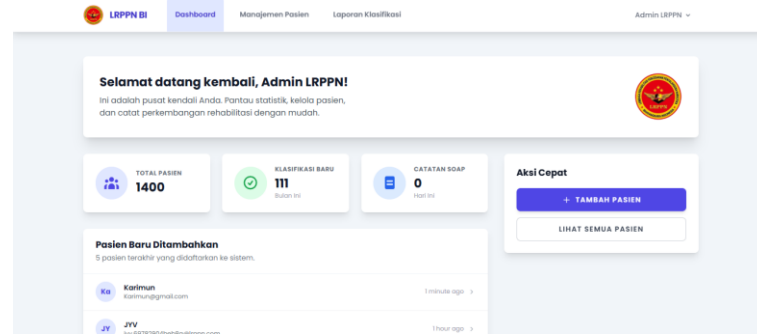
Halaman ini berisi form login dengan alamat email dan kata sandi agar dapat mengakses fitur selanjutnya, terdapat pada gambar 7.



Gambar 7. Tampilan Login

b. Halaman Dashboard

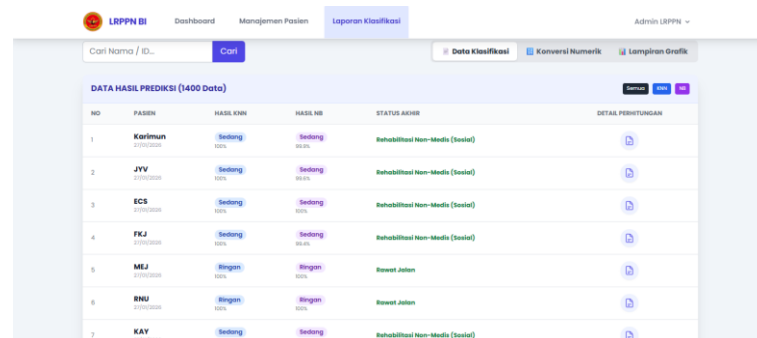
Setelah login, dapat masuk ke halaman dashboard yang berisi terkait jumlah total pasien, penambahan pasien, serta klasifikasi baru, terdapat pada gambar 8.



Gambar 8. Tampilan Dashboard

c. Halaman Klasifikasi KNN dan Naive Bayes

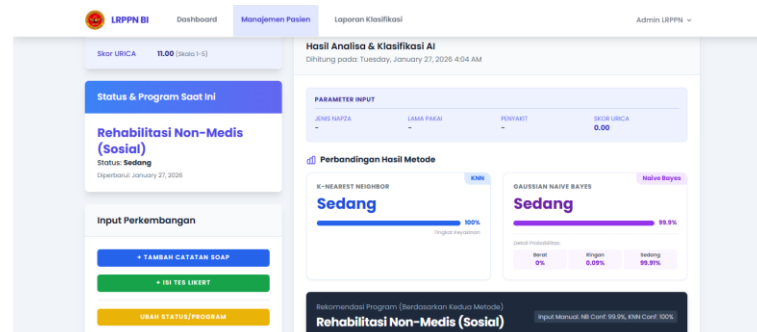
Pada halaman ini, terdapat laporan yang menyajikan hasil klasifikasi tingkat ketergantungan pasien menggunakan metode KNN dan Naive Bayes beserta status akhir rehabilitasinya, terdapat pada gambar 9.



Gambar 9. Tampilan Klasifikasi KNN dan Naive Bayes

d. Halaman Status Program

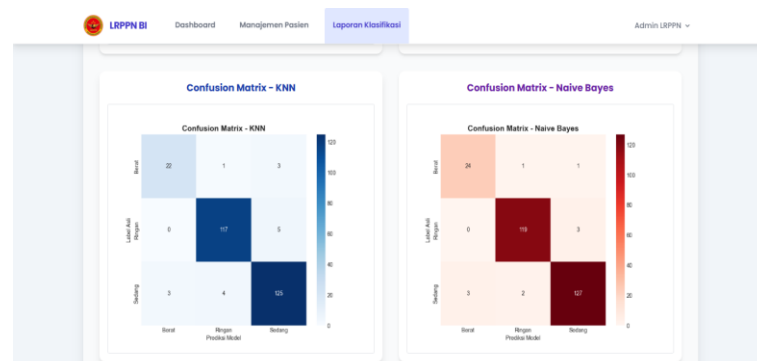
Pada halaman ini, dapat melihat detail perbandingan hasil perhitungan antara dua metode untuk menentukan program rehabilitasi bagi pasien baru, terdapat pada gambar 10.



Gambar 10. Tampilan Status Program Pasien Baru

e. Halaman Evaluasi KNN dan Naive Bayes

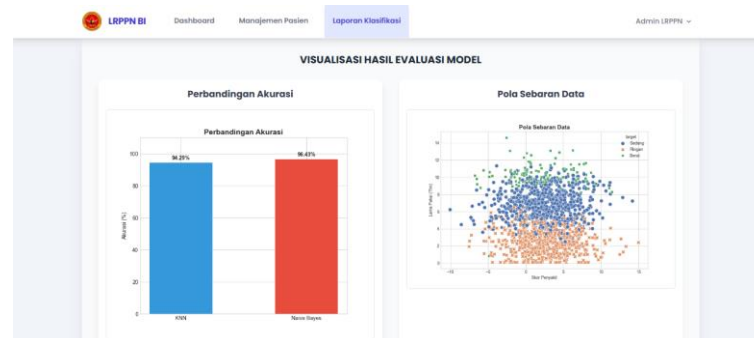
Pada halaman ini, dapat melihat evaluasi model KNN dan Naive Bayes, terdapat pada gambar 11.



Gambar 11. Tampilan Confusion Matrix

f. Halaman Perbandingan Akurasi Model

Pada halaman ini, dapat melihat grafik perbandingan akurasi antara metode KNN dan Naive Bayes dan pola sebaran data pasien, terdapat pada gambar 12.



Gambar 12. Tampilan Perbandingan Akurasi Model

4. KESIMPULAN

Berdasarkan hasil pengujian kedua algoritma, dapat disimpulkan bahwa Naive Bayes menunjukkan performa yang jauh lebih unggul dengan tingkat akurasi mencapai 96,43%, sementara algoritma KNN menghasilkan akurasi sebesar 94,29%. Pada kategori 'Berat', Naive Bayes mencapai tingkat akurasi sebesar 9,6%, sedangkan KNN mencapai akurasi 8,9%. Pada kategori 'Sedang', Naive Bayes dan KNN mencapai tingkat akurasi yang sama, yaitu 43,6%. Pada kategori 'Rendah'. Hal ini menunjukkan bahwa untuk dataset rehabilitasi ini, algoritma Naive Bayes lebih efektif dan konsisten dalam memetakan profil pasien ke target yang sesuai dibandingkan algoritma KNN.

REFERENCES

- [1] Nur Azis, *Analisis Perancangan Sistem Informasi*, vol. 3, no. 1. Bandung: WIDINA BHAKTI PERSADA BANDUNG, 2022.
- [2] Y. Lyu, Q. Xu, Z. Yang, and J. Liu, "Prediction of patient choice tendency in medical decision-making based on machine learning algorithm".
- [3] R. S. Wahono, *Data Mining Data mining*, vol. 2, no. January 2013. 2023. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- [4] T. Triase and S. Samsudin, "Implementasi Data Mining dalam Mengklasifikasikan UKT (Uang Kuliah Tunggal) pada UIN Sumatera Utara Medan," *J. Teknol. Inf.*, vol. 4, no. 2, pp. 370–376, 2020, doi: 10.36294/jurti.v4i2.1711.
- [5] A. Ikhwana *et al.*, "Klasifikasi Kanker Paru-Paru menggunakan Metode Naïve Bayes dengan SMOTE Lung Cancer Classification using the Naïve Bayes Method with SMOTE," vol. 14, pp. 2954–2967, 2025.
- [6] R. Hidayat, F. Sains, U. Islam, and N. Sumatera, "PERANCANGAN APLIKASI PENGARSIPAN SURAT BERBASIS WEB PADA PT. PERKEBUNAN NUSANTARA IV MEDAN," vol. 8, no. 2, 2024.
- [7] K. Huang, C. Yu, X. Chen, Y. Hao, and Y. Ding, "A Quasi-Experimental Study on the Effectiveness of Compulsory and Voluntary Treatment Settings for 1 , 299 Drug Abusers in Hunan , China," vol. 12, no. August, 2021, doi: 10.3389/fpsy.2021.613665.
- [8] H. Nu, N. Lwin, P. Punnakitakashem, and T. Thananusak, "E-Health Research in Southeast Asia : A Bibliometric Review," 2023.
- [9] T. J. Dalimunthe, "Pengembangan Sitem Monitoring Penjualan Dan Prediksi Stok Barang Menggunakan Exponential Smoothing Berbasis Website," pp. 170–183.
- [10] S. Informasi, M. Kesejahteraan, M. Berbasis, and R. J. Sinaga, "Web-Based Welfare Monitoring Information System with Predictive Analytics for Social Assistance and Scholarship Management," vol. 4, 2024.
- [11] B. Hutagalung, "Analysis Of The K-Nearest Neighbor (KNN) Algorithm For Gender Classification Based On Voice Characteristics," vol. 12, no. 5, pp. 705–720, 2025, doi: 10.30865/jurikom.v12i5.9117.
- [12] S. S. Prasetyowati and Y. Sibaroni, "Unlocking the potential of Naive Bayes for spatio temporal classification : a novel approach to feature expansion," *J. Big Data*, 2024, doi: 10.1186/s40537-024-00958-x.
- [13] A. M. Harahap, S. Informasi, U. Islam, and N. Sumatera, "IMPLEMENTASI ALGORITMA NAIVE BAYES UNTUK PREDIKSI KEBUTUHAN," vol. 10, no. 2, pp. 507–513, 2025.
- [14] E. B. Susanto, A. N. Anzila, and B. Ismanto, "Comparison Of The Effectiveness Of K-Nearest Neighbor (KNN) And Naive Bayes Algorithms In Identifying Diabetes Patients," *J. Artif. Intell. Softw. Eng.*, vol. 5, no. 1, p. 22, 2025, doi: 10.30811/jaise.v5i1.6275.
- [15] A. Tangkelayuk and E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN , Naïve Bayes Dan Decision Tree," vol. 9, no. 2, pp. 1109–1119, 2022.
- [16] I. Artikel, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classifier pada Data Set Penyakit Jantung," vol. 1, no. 3, pp. 79–86, 2020.
- [17] B. Delvika, S. Nurhidayarnis, and P. D. Rinada, "Comparison of Classification Between Naive Bayes and K-Nearest Neighbor on Diabetes Risk in Pregnant Women Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil," vol. 2, no. October, pp. 68–75, 2022.
- [18] T. Tursina, H. Muhardi, and D. A. Sari, "Diagnosis Tahapan Pengguna Narkoba Menggunakan Metode K-Nearest Neighbor," *J. Edukasi dan Penelit. Inform.*, vol. 6, no. 1, p. 101, 2020, doi: 10.26418/jp.v6i1.36133.
- [19] H. Marfalino, M. Pratiwi, I. Arief Wisky, and D. Akhiyar, "Penerapan Sistem Pakar Dalam Diagnosa Pengguna Narkoba

- Menggunakan Metode Naïve Bayes,” *J. Sains Inform. Terap.*, vol. 2, no. 2, pp. 30–34, 2023, doi: 10.62357/jsit.v2i2.159.
- [20] U. M. Butt *et al.*, “Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications,” vol. 2021, 2021, doi: 10.1155/2021/9930985.
 - [21] I. G. I. Sudipa, I. G. M. Darmawiguna, I. M. Dendi, and ..., *Buku Ajar Data Mining. PT*, no. April. 2024.
 - [22] C. Lin, Z. Liang, J. Liu, and W. Sun, “A machine learning-based prediction model pre-operatively for functional recovery after 1-year of hip fracture surgery in older people,” no. June, pp. 1–12, 2023, doi: 10.3389/fsurg.2023.1160085.
 - [23] A. Muin, *METODE PENELITIAN KUANTITATIF*. Malang: CV. Literasi Nusantara Abadi, 2023.
 - [24] Adnan Buyung Nasution, B. F. E. Lubis, Nurul Amanda Khairani Lubis, and Friska Andriani, “Perancangan Sistem Pelaporan Keluhan Pelanggan Berbasis Website Menggunakan Metode Waterfall,” *Bull. Comput. Sci. Res.*, vol. 4, no. 1, pp. 40–49, 2023, doi: 10.47065/bulletincsr.v4i1.318.
 - [25] F. D. Pratama and I. Zufria, “Implementasi data mining menggunakan algoritma naïve bayes untuk klasifikasi penerima program indonesia pintar 1) 1,2,3),” vol. 7, no. 1, pp. 77–84, 2022.
 - [26] M. S. Hasibuan, D. Wahyu, and H. Hutabarat, “SISTEM PAKAR DIAGNOSIS AWAL PENYAKIT FARINGITIS DAN LARINGITIS MENGGUNAKAN METODE FORWARD,” vol. 4307, no. August, pp. 1137–1146, 2024.
 - [27] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, “Educational Data Mining Techniques for Student Performance Prediction : Method Review and Comparison Analysis,” vol. 12, no. December, pp. 1–19, 2021, doi: 10.3389/fpsyg.2021.698490.
 - [28] W. Budiharto, “Data science approach to stock prices forecasting in Indonesia during Covid - 19 using Long Short - Term Memory (LSTM),” *J. Big Data*, 2021, doi: 10.1186/s40537-021-00430-0.
 - [29] F. Xie *et al.*, “Deep Learning for Temporal Data Representation in Electronic Health Records : A Systematic Review of Challenges and Methodologies Abstract :,” pp. 1–37.