

Implementasi Klasifikasi Dosis Obat Pil Menggunakan Naive Bayes Pada Penyakit Gagal Jantung

Feiticeira Zulkarnaen
Informatika

Direktorat Kampus Surabaya
Universitas Telkom, Indonesia
feiticeira@student.telkomuniversity.ac.id

Ardian Yusuf Wicaksono
Informatika

Direktorat Kampus Surabaya
Universitas Telkom, Indonesia
ardiany@telkomuniversity.ac.id

Pima Hani Safitri
Informatika

Direktorat Kampus Surabaya
Universitas Telkom, Indonesia
phanisafitri@telkomuniversity.ac.id

Abstrak — Penyakit gagal jantung memerlukan pengelolaan dosis obat yang tepat untuk meningkatkan efektivitas pengobatan. Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi dosis obat menggunakan algoritma Naive Bayes dengan pendekatan Gaussian dan Multinomial, serta meningkatkan performa model dengan teknik oversampling SMOTE dan VAE. Data pasien diklasifikasikan berdasarkan usia, jenis kelamin, berat badan, riwayat penyakit, metabolisme, dan toleransi obat. Implementasi sistem dilakukan melalui GUI berbasis Tkinter, memungkinkan pengguna memasukkan data pasien dan mendapatkan prediksi dosis obat yang sesuai. Hasil evaluasi menunjukkan bahwa kombinasi Naive Bayes dengan SMOTE memberikan performa terbaik dengan akurasi 73,90%, presisi 74,99%, recall 73,36%, dan F1-score 73,90%. Metode VAE memiliki performa lebih rendah dibandingkan SMOTE. Analisis lebih lanjut menunjukkan bahwa Bernoulli Naive Bayes memiliki akurasi tertinggi dalam klasifikasi tanpa oversampling, sementara model Naive Bayes standar kurang efektif dalam menangani data multi-label. Kesimpulannya, pendekatan Naive Bayes dengan oversampling SMOTE dapat meningkatkan akurasi prediksi dosis obat bagi pasien gagal jantung dan dapat digunakan sebagai alat bantu dalam pengambilan keputusan klinis.

Kata kunci— Naive Bayes, Klasifikasi dosis obat, Gagal Jantung, SMOTE, VAE, GUI

I. PENDAHULUAN

Gagal jantung merupakan salah satu penyebab utama angka mortalitas tinggi. Kesalahan dalam resep obat, seperti dosis yang tidak sesuai, dapat berdampak fatal pada pasien [1]. Penelitian ini mengusulkan implementasi metode klasifikasi *Naive Bayes* untuk menentukan dosis obat yang tepat bagi pasien gagal jantung. *Naive Bayes* dikenal sebagai algoritma klasifikasi yang efisien dalam menangani data berukuran besar dan memiliki kecepatan tinggi dalam prediksi.

Dalam penelitian ini, model dikembangkan menggunakan *dataset* rekam medis yang mencakup atribut seperti usia, jenis kelamin, riwayat penyakit, metabolisme, interaksi obat, dan toleransi respon. Evaluasi dilakukan

dengan lima variasi Naive Bayes (Bernoulli, Complement, Gaussian, Multinomial, dan Hybrid Gaussian-Bernoulli) yang diuji sebanyak lima kali untuk mendapatkan hasil yang stabil.

Pengukuran performa model dilakukan dengan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian menunjukkan bahwa metode Naive Bayes mampu memberikan rekomendasi dosis obat dengan akurasi yang tinggi, sehingga dapat mendukung pengambilan keputusan medis secara lebih efektif. Implementasi sistem ini diharapkan dapat meningkatkan kualitas pelayanan medis dan mengurangi risiko kesalahan dosis obat pada pasien gagal jantung.

II. KAJIAN TEORI

A. Penelitian Terdahulu

Beberapa penelitian telah mengimplementasikan algoritma Naive Bayes dalam klasifikasi dosis obat untuk penyakit gagal jantung. [2] Dalam penelitian mereka menunjukkan bahwa Naive Bayes dapat digunakan untuk mengklasifikasikan pasien gagal jantung dengan akurasi mencapai 74,58%, presisi 97,67%, sensitivitas 75%, dan AUC sebesar 0,857, yang menunjukkan performa klasifikasi yang sangat baik.

Penelitian yang dilakukan membandingkan akurasi antara Correlated Naive Bayes Classifier (C-NBC) dan Naive Bayes Classifier (NBC) untuk klasifikasi gagal jantung, di mana C-NBC memiliki akurasi lebih tinggi sebesar 80,6% dibandingkan NBC yang hanya mencapai 67,5%. Studi ini menegaskan bahwa peningkatan metode pada Naive Bayes dapat meningkatkan akurasi klasifikasi pasien gagal jantung [3].

B. Gagal Jantung

Jumlah penderita gagal jantung di seluruh dunia terus meningkat dengan angka kematian dan morbiditas yang tinggi. Pada tahun 2021, tercatat ada 56,5 juta kasus gagal jantung secara global, terdiri dari 27,3 juta perempuan dan 29,2 juta laki-laki. Prevalensi gagal jantung meningkat seiring bertambahnya usia, terutama di negara maju dengan populasi menua seperti Korea yang menuju masyarakat.

Namun, variasi onset, metode pengobatan, dan hasil gagal jantung sangat dipengaruhi oleh demografi, sistem kesehatan, dan sistem asuransi di setiap negara [4].

1. Jantung Koroner

Penyakit jantung koroner yang terdokumentasi pada pasien yang dirawat di rumah sakit karena gagal jantung yang memburuk dengan penurunan HF (Heart Failure) dikaitkan dengan beban penyakit penyerta yang lebih tinggi, penggunaan terapi gagal jantung yang lebih rendah (kecuali beta-blocker), dan peningkatan rawat inap akibat gagal jantung, sementara semua penyebab kematian adalah serupa [5].

2. Infark Miokard

Infark miokard (MI) tetap menjadi penyebab gagal jantung (Heart Failure) paling umum di seluruh dunia. Selama hampir 50 tahun gagal jantung telah diketahui sebagai salah satu faktor penentu prognosis buruk setelah infark miokard, namun upaya untuk mendorong perbaikan miokardium gagal diwujudkan dalam terapi klinis. Infark miokard adalah penyebab utama gagal jantung secara global. Faktor penentu prognosis setelah infark masih menjadi tantangan dalam terapi klinis [6].

3. Penyakit Katup Jantung

Penyakit katup jantung sering terjadi pada lansia dan berhubungan erat dengan gagal jantung. Diagnosis yang akurat diperlukan untuk menentukan terapi yang tepat. Valvular heart disease dan Heart Failure sering terjadi bersamaan, sehingga dapat mempersulit diagnosis akurat mengenai tingkat keparahan stenosis atau regurgitasi katup dan memengaruhi keputusan mengenai pilihan terapi [7].

4. Hipertensi

Hipertensi berkepanjangan dapat menyebabkan gagal jantung, yang ditandai dengan disfungsi diastolik dan penurunan fraksi ejeksi. Tahap akhir penyakit jantung hipertensi diakibatkan oleh kelebihan tekanan dan volume dan terdiri dari kardiomiopati dilatasi dengan disfungsi diastolik dan penurunan fraksi ejeksi [8].

5. Kardiomiopati

Heart Failure merupakan pertanda hasil akhir yang buruk. Kemajuan terkini dalam pemahaman tentang mekanisme yang mendasari perkembangan Heart Failure pada kardiomiopati mempunyai implikasi signifikan terhadap pengambilan keputusan terapeutik [9].

6. Jantung Rematik

Penyakit jantung rematik adalah akibat yang tidak diobati demam rematik akut. Pengetahuan tentang manajemen yang tepat seharusnya dianjurkan pada mereka yang menderita demam rematik untuk menghentikan perkembangan penyakit jantung kerusakan yang dapat menyebabkan gagal jantung [10].

7. Diabetes

Diabetes mellitus dapat memicu atau memperburuk gagal jantung karena akumulasi produk akhir glikasi lanjut, stress oksidatif, penurunan status inflamasi, dan lainnya. Gagal jantung juga mengganggu metabolisme glukosa melalui mekanisme yang kurang diketahui. Perhatian khusus harus diberikan dalam pengobatan karena sering terjadi interaksi merugikan antara kedua penyakit [11].

8. Paru Obstruktif Kronis

Gagal jantung (Heart Failure) dan penyakit paru obstruktif kronik adalah epidemi global yang menimbulkan morbiditas dan mortalitas yang signifikan. Kombinasi ini menghadirkan

banyak tantangan diagnostik. Gejala dan tanda klinis sering kali tumpang tindih. Interpretasi yang cermat diperlukan untuk menghindari kesalahan diagnosis dan pengobatan yang tidak tepat [12].

C. Dosis Obat

Penggunaan 4 jenis obat dalam klasifikasi Naive Bayes seperti Metoprolol, Carvedilol, Cardio Aspirin, Lisinopril dalam bentuk pil disertai berbagai interaksi obat seperti Spironolactone, Warfarin, Salbutamol, Budesonide, Furosemide, Losartan, Enalapril. Spironolactone, Carvedilol, Metoprolol, Lisinopril, Losartan, dan Enalapril membantu mengatur tekanan darah dan mengurangi beban jantung.

1. Carvedilol

Carvedilol dalam menangani gagal jantung, dosis awal adalah 3.125 mg dua kali sehari, dengan peningkatan bertahap hingga 6.25 mg, 12.5 mg, dan 25 mg dua kali sehari dengan interval minimal 2 minggu. Klinisi disarankan untuk mempertahankan dosis yang lebih rendah jika dosis yang lebih tinggi tidak dapat ditoleransi dengan baik oleh pasien [13].

2. Metoprolol

Dosis pemeliharaan disesuaikan sesuai dengan toleransi dan respons klinis pasien. Dosis target biasanya antara 100 dan 200 mg per hari. Metoprolol suksinat harus digunakan selain terapi standar untuk gagal jantung, seperti inhibitor enzim pengubah angiotensin, diuretik, dan digitalis. Dosis dapat berubah seiring bertambahnya usia karena faktor fisiologis [14].

3. Lisinopril

Terdapat dosis awal yang direkomendasikan adalah 2,5 mg per hari, dengan dosis maksimum harian 40 mg. Preferensi terhadap dosis yang lebih tinggi dipengaruhi oleh temuan uji coba SOLVD, di mana orang dalam kelompok dosis tinggi mengalami penurunan kematian sebesar 8% dibandingkan dengan kelompok dosis rendah [15].

4. Cardio Aspirin

Aspirin dikenal sebagai obat antiplatelet yang efektif untuk pencegahan sekunder pada subjek yang berisiko tinggi terkena kejadian kardiovaskular. Namun, sebagian besar pasien yang menerima terapi aspirin jangka panjang masih memiliki risiko besar terjadinya trombotik karena kurangnya penghambatan trombosit, khususnya melalui jalur tromboksan A2 [16].

D. Teori Naive Bayes

Pada teori *Naive Bayes*, klasifikasi *Gaussian Naive Bayes*, yang disebut demikian karena merupakan klasifikasi *Bayesian* yang membuat asumsi sederhana (*naive*) tentang bagaimana fitur Akurasi merupakan metrik yang mengukur proporsi prediksi yang benar terhadap jumlah total prediksi yang dilakukan oleh suatu model fitur berinteraksi [17].

$$c^* = \underset{c \in C}{\operatorname{argmax}} P(c|d) \quad (1)$$

Intuisi dari klasifikasi Bayesian adalah menggunakan aturan Bayes menjadi probabilitas lain yang memiliki beberapa sifat berguna. Aturan Bayes disajikan dalam (1), aturan ini memberi kita cara untuk memecah setiap probabilitas kondisional $P(x|y)$ menjadi tiga probabilitas lain:

$$P(c|d) = \frac{P(d|c)P(c)}{P(d)} \quad (2)$$

Penelitian ini mengklasifikasikan dosis obat berdasarkan empat jenis dosis (Metoprolol, Carvedilol, Cardio Aspirin, Lisinopril) dengan mempertimbangkan delapan riwayat penyakit, faktor individu (usia, jenis kelamin, berat badan, metabolisme), serta 12 interaksi obat. Lima varian Naive Bayes digunakan untuk menangani beragam tipe data dalam dataset: Bernoulli untuk fitur biner, Gaussian untuk data kontinu, Multinomial untuk kategori berbasis frekuensi, Complement untuk ketidakseimbangan kelas, dan Hybrid Gaussian-Bernoulli untuk meningkatkan akurasi prediksi.

E. Oversampling

Salah satu tantangan dalam penerapan model penelitian adalah ketidakseimbangan kelas pada dataset, di mana beberapa kategori dosis obat memiliki jumlah sampel yang jauh lebih sedikit dibandingkan yang lain. Untuk mengatasi masalah ini, dilakukan pendekatan oversampling guna meningkatkan representasi data dari kelas minoritas. Dua metode utama yang digunakan dalam penelitian ini adalah Variational Autoencoder (VAE) dan Synthetic Minority Over-sampling Technique (SMOTE).

1. Synthetic Minority Over-sampling Technique (SMOTE)

Klasifikasi Naive Bayes sering kesulitan dengan data tidak seimbang, lebih akurat untuk kelas mayoritas. Untuk mengatasinya, digunakan kombinasi SMOTE dan Tomek Links pada dataset diagnosis TB dari Rumah Sakit Mayjend HM Ryacudu Kotabumi. Dengan 1.033 data pasien, metode menyeimbangkan data melalui oversampling dan undersampling, meningkatkan kinerja model dan akurasi keseluruhan [18].

2. Variational Autoencoder (VAE)

Masalah klasifikasi tidak seimbang mengakibatkan pengabaian sampel minoritas oleh pengklasifikasi tradisional. Oversampling meningkatkan sampel minoritas tetapi kurang efektif dan memperpanjang waktu pelatihan. Pendekatan dengan auto-encoder variasional lebih optimal karena menyesuaikan distribusi probabilitas tanpa asumsi sebelumnya, sehingga memperluas sampel minoritas secara lebih akurat [19].

F. Akurasi

Confusion Matrix digunakan untuk menganalisis kinerja model machine learning dengan merangkum prediksi benar dan salah. Dengan mengamati nilai diagonal, keakuratan klasifikasi dapat dievaluasi secara efektif [20].

TABEL 1
(Confusion Matrix)

	(Actual Class) Positive	(Actual Class) Negative
(Prediction Class) Positive	True Positive	False Positive
(Prediction Class) Negative	False Negative	True Negative

Dalam menangani ketidakseimbangan kelas dalam dataset rekomendasi dosis obat yang menyebabkan akurasi Naive Bayes rendah sekitar 60%. Untuk mengatasi hal ini, digunakan oversampling dengan SMOTE dan Variational Autoencoder (VAE). SMOTE: Membuat sampel sintesis melalui interpolasi untuk menyeimbangkan kelas. VAE: Menghasilkan data baru berbasis distribusi probabilistik, menciptakan sampel yang lebih beragam dan realistis. Kombinasi kedua metode ini meningkatkan akurasi serta metrik evaluasi lainnya (precision, recall, F1-score),

membuat model lebih adil dan tidak bias terhadap kelas mayoritas.

$$f1 - score = \frac{TP}{TP + TN + FP + FN} \quad (3)$$

Recall mengukur persentase item yang benar-benar ada dalam input yang ada diidentifikasi dengan benar oleh sistem [18].

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

Pada precision berfungsi sebagai memprediksi banyak kasus menjadi positif yang akan menentukan apakah model yang peneliti uji dapat diandalkan atau tidak. Precision sendiri juga merupakan metrik yang berguna jika False Positive memiliki kekhawatiran yang lebih tinggi dibanding False Negative [20].

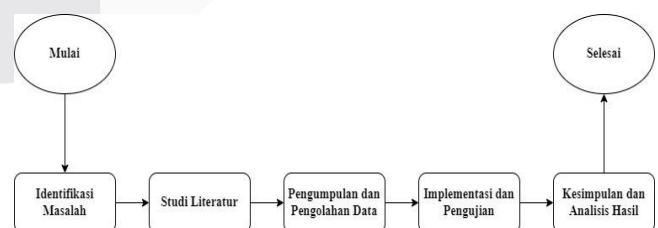
$$Precision = \frac{TP}{TP + FP} \quad (5)$$

Akurasi merupakan metrik yang mengukur proporsi prediksi yang benar terhadap jumlah total prediksi yang dilakukan oleh suatu model seperti terlihat pada (6). Akurasi juga merupakan salah satu metrik evaluasi yang berkaitan erat dengan confusion matrix dalam konteks machine learning dan klasifikasi.

$$Akurasi = \frac{TP + TN}{TP + FN + FP + TN} \times 100\% \quad (6)$$

III. METODE

Terdapat pendekatan kuantitatif dengan menggunakan machine learning. Machine learning menerapkan supervised learning dengan menguji efektivitas algoritma Naive Bayes dalam meningkatkan akurasi dalam pemberian dosis obat. Tujuannya meminimalisir efek samping serta meningkatkan kualitas hidup pasien. Desain penelitian ini melibatkan pengumpulan disertai pengolahan data, perancangan model, pengujian data, dan evaluasi model.



GAMBAR 1
(Desain Penelitian)

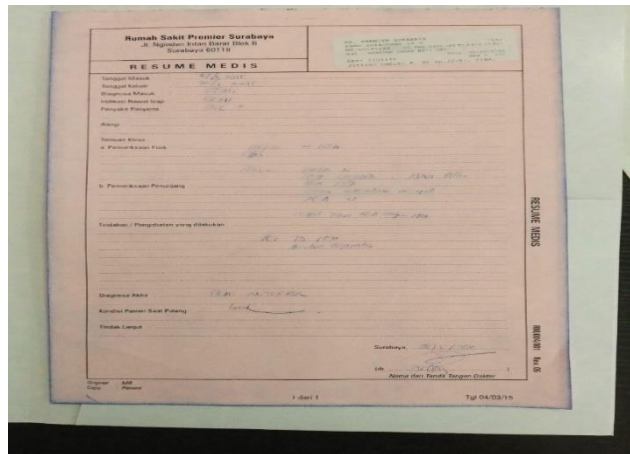
A. Data Penelitian

Setelah studi literatur, penelitian ini melanjutkan pengumpulan data melalui wawancara langsung dengan dr. Jeffrey Daniel Adipranoto, Sp.JP (K), FIHA, FESC, dokter spesialis gagal jantung di Rumah Sakit Premier Surabaya,

pada 7 Januari 2025. Peneliti memperoleh 500 rekam medis beserta resep obat pasien, yang digunakan sebagai data penelitian.

1. Resume Medis Pasien

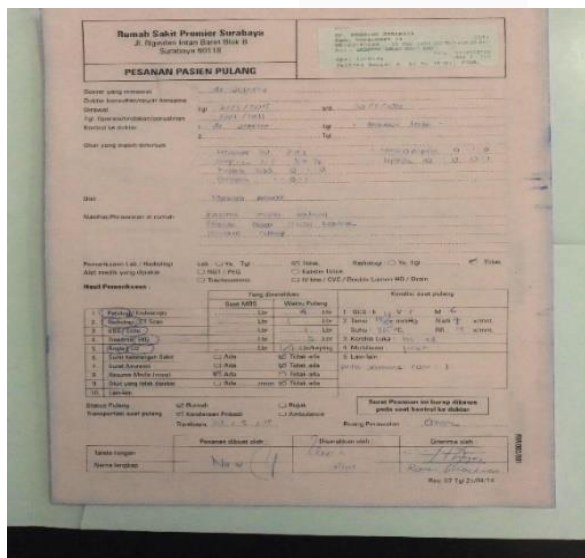
Data rekam medis pasien dari rumah sakit dapat dilihat pada Gambar 2. Pasien diberikan rekam medis beserta resep obat pasien tersebut.



GAMBAR 2
(Rekam Medis Pasien)

2. Pesanan Pasien Pulang

Lembar Pesanan Pasien Pulang berisi rekam medis pasien, termasuk penyakit penyerta dan riwayat konsumsi obat, sebagaimana ditampilkan pada Gambar 3. Data ini digunakan sebagai data pelatihan dalam algoritma machine learning untuk mempelajari pola dari data asli. Fitur yang ditampilkan meliputi Usia, Jenis Kelamin, Berat Badan, Riwayat Penyakit, Metabolisme, Interaksi Obat, Toleransi dan Respon, Formulasi Obat, serta Dosis Awal.

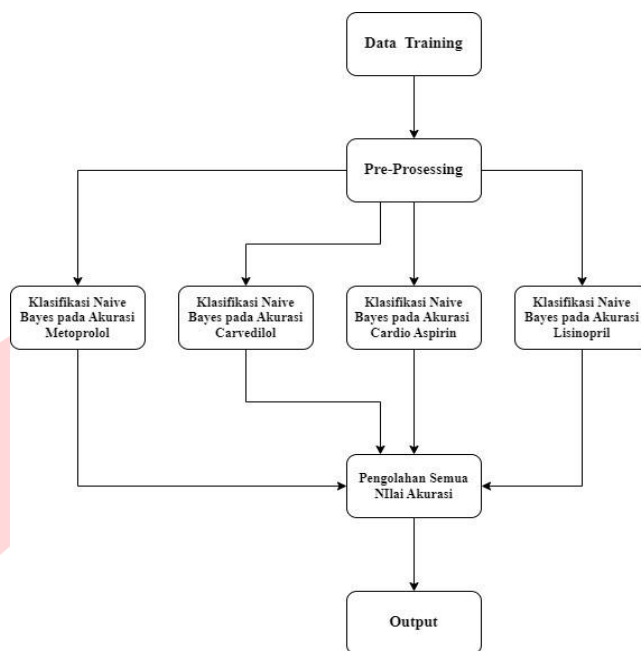


GAMBAR 3
(Resep Obat)

Proses digitalisasi rekam medis mengubah data dari format kertas menjadi numerik untuk dilatihkan pada Naive Bayes. Tujuan utama adalah mengelompokkan data secara rinci berdasarkan 4 fitur kelas dosis obat dan 5 riwayat penyakit pasien. Fitur yang digunakan meliputi Usia, Jenis

Kelamin, Berat Badan, dan Riwayat Penyakit, dengan klasifikasi 2 kelas dosis obat untuk setiap jenis obat, yaitu Dosis Metoprolol, Dosis Carvedilol, Dosis Cardio Aspirin, dan Dosis Lisinopril.

B. Perancangan Model



GAMBAR 4
(Perancangan Model)

Penelitian ini diawali dengan pengumpulan data training melalui wawancara langsung di Rumah Sakit Premier Surabaya. Dalam klasifikasi Naive Bayes, jumlah data yang cukup diperlukan untuk meningkatkan akurasi model. Setelah data diperoleh, dilakukan preprocessing dengan mengonversi data menjadi numerik, di mana nilai 1 menunjukkan pasien memiliki riwayat penyakit, sedangkan nilai 0 menunjukkan sebaliknya.

Klasifikasi Naive Bayes diterapkan pada 4 kelas dosis obat, dengan masing-masing kelas menghitung akurasi secara independen. Total akurasi sistem dihitung dengan merata-ratakan hasil dari keempat kelas. Model diuji sebanyak lima kali dengan pembagian data latih dan uji secara acak untuk memastikan keandalan hasil. Fitur yang digunakan mencakup semua atribut kecuali kolom dosis obat, sedangkan targetnya adalah kolom dosis obat.

Percobaan dilakukan dengan lima varian Naive Bayes seperti Gaussian Naive Bayes untuk data kontinu seperti usia dan berat badan, Bernoulli Naive Bayes untuk atribut biner (0 dan 1), Hybrid Gaussian & Bernoulli Naive Bayes untuk kombinasi Gaussian untuk data kontinu dan Bernoulli untuk data biner, Multinomial Naive Bayes untuk atribut diskret. Complement Naive Bayes untuk menangani ketidakseimbangan data kategori.

Ketidakseimbangan data, terutama pada jenis obat Cardio Aspirin, diatasi dengan oversampling menggunakan dua metode yaitu SMOTE (Synthetic Minority Over-sampling Technique) untuk menghasilkan data sintesis dengan interpolasi antar sampel minoritas menggunakan library imblearn.over_sampling.

dan VAE (Variational Autoencoder) untuk model generatif berbasis neural network yang mempelajari distribusi data asli untuk menghasilkan sampel sintetis yang lebih bervariasi.

Pendekatan ini bertujuan untuk meningkatkan akurasi, precision, recall, dan F1-score guna menghasilkan model rekomendasi dosis obat yang lebih akurat dan adil.

C. Implementasi dan Pengujian

Aplikasi yang dikembangkan ditujukan bagi tenaga medis, termasuk dokter spesialis gagal jantung, apoteker, dan manajemen rumah sakit, sebagai sistem rekomendasi dosis obat untuk mendukung pengambilan keputusan klinis. Model yang telah dilatih diuji menggunakan data uji guna mengevaluasi kinerjanya secara independen pada setiap kelas. Hasil pengujian dikombinasikan menggunakan rata-rata akurasi dari keempat kelas dosis obat, sesuai dengan (7).

$$\text{Total Akurasi Sistem} = \frac{AM + AC + ACA + AL}{4} \quad (7)$$

Pengujian recall dilakukan untuk menilai kemampuan model dalam mengidentifikasi kasus positif yang sebenarnya. Evaluasi dilakukan pada setiap kelas dosis obat, dengan hasil akhir dihitung menggunakan rata-rata recall dari keempat kelas, sesuai dengan (8).

$$\text{Total Recall Sistem} = \frac{RM + RC + RCA + RL}{4} \quad (8)$$

Pengujian precision menilai kemampuan model dalam mengidentifikasi kasus positif yang benar di antara semua prediksi positif. Evaluasi dilakukan pada setiap kelas dosis obat, dengan hasil akhir dihitung menggunakan rata-rata precision dari keempat kelas, sesuai dengan (9).

$$\text{Total Precision Sistem} = \frac{PM + PC + PCA + PL}{4} \quad (9)$$

Pengujian F-measure mengevaluasi keseimbangan antara precision dan recall, memberikan gambaran menyeluruh tentang kinerja model dalam mengidentifikasi kasus positif sekaligus menghindari positif palsu. Hasil akhir dihitung menggunakan rata-rata F-measure dari keempat kelas dosis obat, sesuai dengan (10).

$$\text{Total F Measure Sistem} = \frac{FMM + FMC + FMCA + FML}{4} \quad (10)$$

D. GUI

Dalam penelitian ini untuk menampilkan hasil sistem klasifikasi Gaussian Naive Bayes menggunakan GUI menggunakan library tkinter. User berinteraksi seperti mengisi input berdasarkan data pasien di rumah sakit, lalu terdapat peringatan apabila salah mengisi input pada sistem gui.

IV. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Pengumpulan data dilakukan melalui wawancara dengan dr. Jeffrey Daniel Adipranoto, Sp.JP (K), FIHA, FESC di Rumah Sakit Premier Surabaya pada 7 Januari 2025. Peneliti memperoleh 500 rekam medis pasien beserta resep obat untuk meningkatkan akurasi sistem Naive Bayes. Data dikonversi menjadi data numerik dengan berbagai fitur baru,

yang digunakan dalam klasifikasi 2 kelas dosis obat untuk setiap jenis obat.

B. Pengolahan Data

Sebelumnya, data numerik yang diolah hanya 100 data dengan normalisasi yang belum optimal. Saat ini, peneliti telah menambahkan 500 data dan menerapkan normalisasi untuk meningkatkan akurasi klasifikasi Naive Bayes.

1. Normalisasi Data

Normalisasi data numerik dilakukan untuk memastikan distribusi yang homogen dan meningkatkan kinerja Naive Bayes. Awalnya, hanya 9 fitur yang digunakan, namun setelah penambahan 500 data, jumlah fitur diperbarui menjadi 25, termasuk riwayat penyakit pasien dan interaksi obat.

2. Data Hasil Preprocessing

TABEL 2

(Data Dosis Obat Sebelumnya)

Nama Dosis Obat	Jumlah Data Pada Kelas					
	0	6,25	40	50	100	300
Dosis Metoprolol	45	0	0	55	0	0
Dosis Carvedilol	48	52	0	0	0	0
Dosis Cardio Aspirin	50	0	0	0	30	20
Dosis Lisinopril	37	0	63	0	0	0

Dataset terdiri dari 500 data pasien yang telah melalui pre-processing untuk klasifikasi Naive Bayes dalam rekomendasi dosis obat pil. Dataset mencakup atribut-atribut utama, yaitu: Usia (numerik), Jenis Kelamin (0 = laki-laki, 1 = perempuan), Riwayat Penyakit Gagal Jantung (8 atribut, 0 = tidak, 1 = ya), Metabolisme (0 = lambat, 1 = normal, 2 = cepat), Interaksi Dosis Obat (12 atribut, 0 = tidak, 1 = ya), Toleransi dan Respon (0 = normal, 1 = baik, 2 = sensitif). Klasifikasi Dosis Obat dengan jumlah data seperti Metoprolol yaitu nilai 0 berjumlah 45, nilai 50 mg berjumlah 55, Carvedilol nilai 0 berjumlah 48, nilai 6.25 mg berjumlah 52, Cardio Aspirin nilai 0 berjumlah 50, nilai 100 mg berjumlah 30, nilai 300 mg berjumlah 20, dan Lisinopril nilai 0 berjumlah 37, nilai 40 mg berjumlah 63. Total masing-masing dosis adalah 100 data, dengan distribusi lengkap dapat dilihat pada Tabel 2.

Pada pengumpulan data selanjutnya, data bertambah dari 100 menjadi 500 data. Setiap kelas dosis obat mengalami perubahan jumlah data. Persebaran jumlah data pada setiap dosis obat saat 500 data dapat dilihat pada Tabel 3.

TABEL 3

(Jumlah Kelas Dosis Obat)

Nama Dosis Obat	Jumlah Data Pada Kelas					
	0	6,25	40	50	100	300

Dosis Metoprolol	214	0	0	286	0	0
Dosis Carvedilol	239	261	0	0	0	0
Dosis Cardio Aspirin	223	0	0	0	190	87
Dosis Lisinopril	218	0	282	0	0	0

3. Pembagian Data

Dataset dibagi menjadi dua bagian yaitu Data latih (80%) digunakan untuk melatih model. Data uji (20%) digunakan untuk mengevaluasi kinerja model. Pembagian dilakukan dengan fungsi `train_test_split` dari library `sklearn`.

C. Pengujian Model

Pengujian model melibatkan 4 model pada jenis dosis obat yaitu dosis Carvedilol, dosis Metoprolol, dosis Cardio Aspirin, dosis Lisinopril. Setiap model diuji untuk klasifikasi sistem dalam berbagai varian Naive Bayes beserta menggunakan metode oversampling.

1. Implementasi dan Evaluasi Naive Bayes

Naive Bayes diuji lima kali dengan pembagian data latih dan uji secara acak untuk memastikan konsistensi hasil. Model dilatih menggunakan semua atribut kecuali kolom dosis obat sebagai fitur, sementara kolom dosis obat menjadi target. Evaluasi dilakukan dalam 5 run untuk setiap klasifikasi. Percobaan juga dilakukan untuk menentukan jenis Naive Bayes yang paling optimal, sebagaimana dirinci pada Tabel 4.

TABEL 4
(Metode Klasifikasi Naive Bayes)

No Skema	Metode	Jumlah Data
1	Gaussian Naive Bayes	500
2	Bernoulli Naive Bayes	500
3	Hybrid Gaussian dan Bernoulli Naive Bayes	500
4	Multinomial Naive Bayes	500
5	Complement Naive Bayes	500

Tabel 4 merangkum berbagai skema klasifikasi dengan algoritma Naive Bayes. Penelitian ini menggunakan 500 data asli dan menguji beberapa variasi Naive Bayes Gaussian Naive Bayes untuk data kontinu seperti usia dan berat badan, Bernoulli Naive Bayes untuk atribut biner (0 dan 1), Kombinasi Gaussian dan Bernoulli Naive Bayes, masing-masing untuk data kontinu dan biner, Multinomial Naive Bayes untuk atribut diskret dan Complement Naive Bayes untuk menangani ketidakseimbangan data kategori.

2. Penanganan Ketidakseimbangan Data

Ketidakseimbangan kelas dalam dataset, terutama pada jenis obat Cardio Aspirin, ditangani dengan oversampling menggunakan dua metode yaitu SMOTE (Synthetic Minority Over-sampling Technique) untuk menghasilkan data sintetis dengan menginterpolasi sampel minoritas menggunakan `imblearn.over_sampling` dan VAE (Variational Autoencoder) untuk model generatif berbasis neural network yang mempelajari distribusi data asli untuk menghasilkan data sintetis dari representasi laten. Percobaan oversampling pada variasi Naive Bayes tercantum dalam Tabel 5.

TABEL 5
(Metode Oversampling)

No Skema	Metode	Jumlah Data
1	Gaussian Naive Bayes + SMOTE	827
2	Bernoulli Naive Bayes + SMOTE	827
3	Hybrid Gaussian dan Bernoulli Naive Bayes + SMOTE	827
4	Multinomial Naive Bayes + SMOTE	827
5	Complement Naive Bayes + SMOTE	827
6	Gaussian Naive Bayes + VAE	827
7	Bernoulli Naive Bayes + VAE	827
8	Hybrid Gaussian dan Bernoulli Naive Bayes + VAE	827
9	Multinomial Naive Bayes + VAE	827
10	Complement Naive Bayes + VAE	827

Setelah oversampling dengan SMOTE dan VAE, jumlah data meningkat dari 500 menjadi 827 untuk mengatasi ketidakseimbangan kelas. Berbagai metode Naive Bayes diterapkan kembali pada SMOTE dan VAE. Pada SMOTE dalam Gaussian Naive Bayes dan SMOTE untuk data kontinu, Bernoulli Naive Bayes dan SMOTE untuk atribut biner, Hybrid Gaussian serta Bernoulli Naive Bayes dan SMOTE untuk menggabungkan data kontinu dan biner, Multinomial Naive Bayes dan SMOTE untuk atribut diskret, Complement Naive Bayes dan SMOTE untuk ketidakseimbangan data kategori.

Pada VAE dalam Gaussian Naive Bayes dan VAE untuk data kontinu, Bernoulli Naive Bayes dan VAE untuk atribut biner, Hybrid Gaussian dan Bernoulli Naive Bayes dan VAE untuk menggabungkan data kontinu dan biner, Multinomial Naive Bayes dan VAE untuk atribut diskret, dan Complement

Naive Bayes dan VAE untuk ketidakseimbangan data kategori. Skema ini bertujuan mengevaluasi performa Naive Bayes dengan berbagai pendekatan pengolahan data.

3. Pengujian Data

Hasil evaluasi Naive Bayes dicatat dalam tabel Excel dengan metrik untuk precision untuk akurasi prediksi positif, recall untuk sensitivitas model terhadap kelas target, F1-Score untuk rata-rata harmonik, precision dan recall, accuracy per class untuk akurasi tiap kelas dosis obat. Evaluasi dilakukan pada empat model dosis (Metoprolol, Carvedilol, Cardio Aspirin, Lisinopril), serta oversampling tambahan yaitu SMOTE untuk model dilatih dengan data hasil oversampling untuk menangani ketidakseimbangan kelas dan VAE untuk model dilatih dengan data hasil augmentasi untuk meningkatkan representasi distribusi data. Hasil dibandingkan untuk menentukan metode terbaik berdasarkan metrik evaluasi.

4. GUI

Proses pengolahan data diintegrasikan ke dalam aplikasi GUI berbasis Tkinter dengan varian Gaussian Naive Bayes dan pendekatan One-vs-Rest untuk klasifikasi multi-kelas dan multi-label. Sistem memungkinkan pihak rumah sakit memasukkan data pasien (usia, jenis kelamin, berat badan), memilih riwayat penyakit, metabolisme, dan toleransi obat. Setelah pengguna mengeklik tombol prediksi, sistem menampilkan dosis obat (Metoprolol, Carvedilol, Cardio Aspirin, Lisinopril) yang sesuai dengan kriteria pasien. Tampilan gui ditunjukkan pada Gambar 5.

GAMBAR 5
(GUI)

V. KESIMPULAN

A. Analisis Hasil Sistem

Analisis akhir implementasi klasifikasi dosis obat dengan Naive Bayes pada penyakit gagal jantung mencakup evaluasi dari enam tabel utama: performa keseluruhan Naive Bayes (Tabel 6), sistem dengan SMOTE (Tabel 7), sistem dengan VAE (Tabel 8), keseluruhan sistem gabungan (Tabel 9), akurasi per model dosis obat (Tabel 10), dan Sistem Multinomial SMOTE (Tabel 11). Setiap tabel dianalisis secara mendetail untuk menilai efektivitas masing-masing metode dalam meningkatkan akurasi klasifikasi dosis obat.

1. Analisis Total Sistem Naive Bayes

TABEL 6
(Keseluruhan Sistem Naive Bayes)

Total Keseluruhan Sistem	Presisi Sistem	67,98%
	Recall Sistem	66,62%
	F1 Measure Sistem	65,53%
	Akurasi Sistem	66,89%

Sistem Naive Bayes (Tabel 6) mencapai presisi 67,98%, recall 66,62%, F1-score 65,53%, dan akurasi 66,89%, menunjukkan performa yang kurang optimal dalam memprediksi dosis obat. Akurasi rendah disebabkan oleh keterbatasan jumlah data, yang berdampak pada efektivitas model. Dengan lebih banyak data, hasil prediksi dapat meningkat, seperti terlihat pada oversampling yang meningkatkan akurasi hingga 70%. Selain itu, Naive Bayes kurang cocok untuk klasifikasi multi-label karena mengasumsikan independensi antar label.

2. Analisis Sistem Oversampling

TABEL 7
(Keseluruhan Sistem SMOTE)

Total Keseluruhan Sistem	Presisi Sistem	70,00%
	Recall Sistem	69,24%
	F1 Measure Sistem	68,76%
	Akurasi Sistem	70,20%

TABEL 8
(Keseluruhan Sistem VAE)

Total Keseluruhan Sistem	Presisi Sistem	72,19%
	Recall Sistem	71,15%
	F1 Measure Sistem	70,22%
	Akurasi Sistem	70,70%

Oversampling menggunakan SMOTE (Tabel 7) mencapai presisi 70,00%, recall 69,24%, F1-score 68,76%, dan akurasi 70,20%, menunjukkan efektivitas tinggi dalam menangani ketidakseimbangan data pada Naive Bayes. Sementara itu, oversampling menggunakan VAE (Tabel 8) memperoleh presisi 72,19%, recall 71,15%, F1-score 70,22%, dan akurasi 70,70%, namun performanya lebih rendah dibandingkan SMOTE dalam meningkatkan keseimbangan data dan akurasi model.

3. Analisis Total Sistem Keseluruhan

Terdapat analisis total sistem keseluruhan pada setiap pengujian naive bayes, oversampling vae dan smote. Berdasarkan hasil pada Tabel 9.

TABEL 9
(Keseluruhan Sistem)

Total Keseluruhan Sistem	Presisi Sistem	70,06%
	Recall Sistem	69,00%
	F1 Measure Sistem	68,17%
	Akurasi Sistem	69,26%

Penggabungan sistem Naive Bayes dengan oversampling menggunakan VAE dan SMOTE pada Tabel 9 mencapai akurasi 69,26%, dengan presisi 70,06%, recall 69,00%, dan F1-score 68,17%. Hasil ini menunjukkan bahwa metode probabilistik mampu memprediksi dosis obat dengan cukup baik berdasarkan fitur pasien. Meskipun akurasi tinggi, evaluasi tambahan menggunakan presisi, recall, dan F1-score dilakukan untuk memperoleh gambaran performa model yang lebih komprehensif.

4. Model Dosis Obat Terbaik

TABEL 10
(Akurasi Permodel Dosis Obat)

Total Keseluruhan Sistem	Presisi Sistem	75,92%
	Recall Sistem	75,23%
	F1 Measure Sistem	75,14%
	Akurasi Sistem	75,03%

Semua penilaian sistem dilakukan pada Tabel 10 ditemukan bahwa model dosis obat terbaik yang dapat diklasifikasikan dengan akurasi total keseluruhan tinggi adalah model Carvedilol mencapai 75,92% daripada model lainnya. Hal ini menunjukkan bahwa model mampu menangani data untuk kelas ini dengan baik, bahkan pada dataset yang tidak seimbang. Sistem klasifikasi dosis obat pil menggunakan metode Naive Bayes telah menunjukkan hasil yang baik dalam menangani data pada penyakit gagal jantung. Kombinasi teknik oversampling SMOTE dan VAE dengan 5 variasi Naive Bayes menghasilkan performa terbaik.

B. Analisis Hasil Perbandingan Model Naive Bayes

Pembahasan tentang hasil perbandingan klasifikasi Naive Bayes telah mencapai hasil yang layak, oleh karena itu peneliti memberitahu hasil analisis untuk pencapaian klasifikasi Naive Bayes terbaik pada penelitian ini.

1. Naive Bayes Terbaik

Pengujian pada berbagai varian Naive Bayes seperti Gaussian, Bernoulli, Multinomial, Complement, dan Hybrid Gaussian dan Bernoulli menunjukkan bahwa Bernoulli Naive Bayes memiliki performa terbaik. Berdasarkan Tabel V.3, metode ini mencapai presisi 71,29%, recall 70,31%, F1-score 70,04%, dan akurasi 71,14%.

C. Analisis Hasil Penanganan Ketidakseimbangan Data

Pembahasan tentang hasil perbandingan klasifikasi Naive Bayes pada oversampling mencapai hasil yang memuaskan,

Memberitahu hasil analisis untuk pencapaian oversampling terbaik pada penelitian.

1. Oversampling Terbaik

TABEL 11
(Sistem Multinomial SMOTE)

Total Keseluruhan Sistem	Presisi Sistem	74,99%
	Recall Sistem	73,36%
	F1 Measure Sistem	73,59%
	Akurasi Sistem	73,90%

Teknik oversampling terbaik adalah SMOTE, sebagaimana tercantum dalam Tabel 7, dengan akurasi sistem 70,20%, presisi 70%, recall 69,24%, dan F1-score 68,76%. Hasil pengujian individu pada Tabel 11 menunjukkan bahwa kombinasi SMOTE dan Multinomial Naive Bayes memberikan performa terbaik, meningkatkan akurasi hingga 73,90%, dengan presisi 74,99%, recall 73,36%, dan F1-score 73,90%, serta mengurangi kesalahan pada kelas minoritas.

D. Analisis GUI

GAMBAR 6

(Tampilan Hasil Keseluruhan GUI)

Analisis GUI pada implementasi akhir sistem ditunjukkan pada Gambar 6. Sistem mengonversi setiap input pengguna ke dalam nilai numerik untuk klasifikasi. Usia (55 sampai 90 tahun) dan berat badan (40 sampai 90 kg) dinormalisasi dalam rentang 0 sampai 1. Jenis kelamin dikodekan sebagai 1 untuk "Pria" dan 0 untuk "Wanita". Riwayat penyakit dikategorikan biner (0 yaitu tidak memiliki, 1 yaitu memiliki). Metabolisme diklasifikasikan sebagai Lambat (0), Normal (1), dan Cepat (2), sedangkan toleransi obat dikategorikan sebagai Baik (0), Normal (1), dan Sensitif (2). Hasil klasifikasi bervariasi sesuai data input pasien,

memastikan prediksi dosis obat yang sesuai berdasarkan parameter yang dimasukkan.

REFERENSI

- [1] H. Nursita and A. Pratiwi, "Peningkatan Kualitas Hidup Pada Pasien Gagal Jantung: A Narrative Review Article," *J. Ber. Ilmu Keperawatan*, vol. 13, no. 1, p. 11, 2020, [Online]. Available: <https://doi.org/10.23917/bik.v13i1.11916>
- [2] O. P. Barus, K. Lauwren, J. J. Pangaribuan, and Romindo, "Implementation of the Naive Bayes Algorithm to Predict the Safety of Heart Failure Patients," *Conf. Ser.*, vol. 4, no. 1, pp. 172–177, 2023, doi: 10.34306/conferenceseries.v4i1.651.
- [3] P. Subarkah, W. R. Damayanti, and R. A. Permana, "Comparison of Correlated Algorithm Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Classification of heart failure," *Ilk. J. Ilm.*, vol. 14, no. 2, pp. 120–125, 2022, doi: 10.33096/ilkom.v14i2.1148.120-125.
- [4] C. J. Lee *et al.*, "Heart Failure Statistics 2024 Update: A Report From the Korean Society of Heart Failure," *Int. J. Hear. Fail.*, vol. 6, no. 2, pp. 56–69, 2024, doi: 10.36628/ijhf.2024.0010.
- [5] R. J. Mentz *et al.*, "Influence of documented history of coronary artery disease on outcomes in patients admitted for worsening heart failure with reduced ejection fraction in the EVEREST trial," *Eur. J. Heart Fail.*, vol. 15, no. 1, pp. 61–68, 2013, doi: 10.1093/eurjhf/hfs139.
- [6] B. F. Tetzlaff WF, Meroño T, Menafrá M, Martín M, Botta E, Matoso MD, Sorroche P, De Paula JA, Boero LE, "World Journal of Cardiology," *World J. Cardiol.*, vol. 9, no. 5, 2017, doi: 10.4330/wjc.v9.i5.448.
- [7] S. Ray, "History of Valvular Heart Disease," *Clin. Med. (Northfield. Il.)*, vol. 10, no. 2, pp. 168–171, 2010.
- [8] F. H. Messerli, S. F. Rimoldi, and S. Bangalore, "The Transition From Hypertension to Heart Failure: Contemporary Update," *JACC Hear. Fail.*, vol. 5, no. 8, pp. 543–551, 2017, doi: 10.1016/j.jchf.2017.04.012.
- [9] P. M. Seferović *et al.*, "Heart failure in cardiomyopathies: a position paper from the Heart Failure Association of the European Society of Cardiology," *Eur. J. Heart Fail.*, vol. 21, no. 5, pp. 553–576, 2019, doi: 10.1002/ehf.1461.
- [10] Dian Rozani, "Rheumatic Heart Disease Causes Heart Failure. How is It Treated?," *Cardiovasc. Cardiometaabolic J.*, vol. 4, no. 1, pp. 38–48, 2023, doi: 10.20473/ccj.v4i1.2023.38-48.
- [11] D. Tousoulis, E. Oikonomou, G. Siasos, and C. Stefanadis, "Bidirectional Relationship From Diabetes Mellitus to Heart Failure," *Eur Cardiol*, vol. 9, no. 1, pp. 37–42, 2024.
- [12] N. M. Hawkins, M. C. Petrie, P. S. Jhund, G. W. Chalmers, F. G. Dunn, and J. J. V. McMurray, "Heart failure and chronic obstructive pulmonary disease: Diagnostic pitfalls and epidemiology," *Eur. J. Heart Fail.*, vol. 11, no. 2, pp. 130–139, 2009, doi: 10.1093/eurjhf/hfn013.
- [13] "Carvedilol - StatPearls - NCBI Bookshelf." Accessed: Jun. 28, 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK534868/>
- [14] "Metoprolol - StatPearls - NCBI Bookshelf." Accessed: Jun. 28, 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK532923/>
- [15] "Lisinopril - StatPearls - NCBI Bookshelf." Accessed: Jun. 28, 2024. [Online]. Available: <https://www.ncbi.nlm.nih.gov/books/NBK482230/>
- [16] C. Aspirin, "Cardio spirin ®," vol. 1003661200, pp. 5–10, 2021.
- [17] J. Daniel and J. H. Martin, "Chapter 4 Naive Bayes, Text Classification, and Sentiment | Speech and Language Processing," 2023.
- [18] D. J. Maulana, Siti Saadah, and Prasti Eko Yunanto, "Kmeans-SMOTE Integration for Handling Imbalance Data in Classifying Financial Distress Companies using SVM and Naïve Bayes," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 8, no. 1, pp. 54–61, 2024, doi: 10.29207/resti.v8i1.5140.
- [19] C. Zhang *et al.*, *Over-Sampling Algorithm Based on VAE in Imbalanced Classification*, vol. 10967 LNCS. Springer International Publishing, 2018. doi: 10.1007/978-3-319-94295-7_23.
- [20] K. M. Ting, "Confusion Matrix," *Encycl. Mach. Learn. Data Min.*, no. October, pp. 260–260, 2017, doi: 10.1007/978-1-4899-7687-1_50.