Jurnal Ilmiah Informatika (JIF)
| ISSN (Print) 2337-8379 | ISSN (Online) 2615-1049 |
Jurnal online di akses di http://ejournal.upbatam.ac.id/index.php/jif

Klasifikasi Status Tekanan Darah Berdasarkan Kadar Kolesterol Pada Lansia Di Puskesmas Setiabudi

Vina Nur Amalia Putri[1], Fuad Nur Hasan[2], Lestari Yusuf[3]

[1,2,3]*Universitas Bina Sarana Informatika, Jakarta, 10450, Indonesia*

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 18-07-2026
Revisi Akhir: 28-07-2026
Diterbitkan *Online*: 25-09-2026

KATA KUNCI

Naïve Bayes, Hypertension, Cholesterol, Blood Pressure, Older Adults

KORESPONDENSI

No HP: +62 857 1844 5096
E-mail: vinaptr19@gmail.com

A B S T R A C T

Hypertension is a noncommunicable disease commonly experienced by older adults and is closely associated with elevated cholesterol levels; therefore, a classification method capable of supporting rapid and accurate early detection is needed. This study aims to classify blood pressure status based on cholesterol levels among older adults at the Setiabudi Community Health Center in 2025 using the Naive Bayes algorithm and to evaluate the model's performance through 10-fold cross-validation. The study employed a quantitative approach using a classification method applied to 1,099 secondary medical records of older adults, which included variables such as gender, age, weight, height, body mass index, upper arm circumference, waist circumference, cholesterol levels, systolic blood pressure, and diastolic blood pressure. The research stages included preprocessing, data transformation, feature selection, Naive Bayes model development, and evaluation using a confusion matrix and classification report. The results showed that the model was able to classify blood pressure status into nine categories with an accuracy of 80.44%, precision of 84%, recall of 80%, and an F1-score of 81%. These findings indicate that the Naive Bayes algorithm is effective in classifying blood pressure status based on cholesterol levels in the elderly and has the potential to support decision-making for screening and early detection of hypertension risk in healthcare facilities.

1 PENDAHULUAN

Peningkatan usia harapan hidup di Indonesia menyebabkan bertambahnya jumlah penduduk lanjut usia (lansia) dari tahun ke tahun. Berdasarkan data Badan Pusat Statistik (BPS) tahun 2025, populasi lansia di Provinsi DKI Jakarta mencapai sekitar 1,4 juta jiwa atau sekitar 12,9% dari keseluruhan penduduk. Kondisi tersebut menunjukkan terjadinya perubahan struktur demografi yang berimplikasi pada meningkatnya kebutuhan pelayanan kesehatan bagi kelompok usia lanjut. Seiring bertambahnya usia, kemampuan fisiologis tubuh mengalami penurunan sehingga fungsi berbagai organ tidak lagi bekerja secara optimal. Akibatnya, lansia menjadi kelompok yang memiliki risiko lebih tinggi mengalami penyakit degeneratif, khususnya Penyakit Tidak Menular (PTM), salah satunya adalah hipertensi. Penelitian yang dilakukan oleh Lasriado menyebutkan bahwa hasil Survei Kesehatan Dasar (Riskesdas) menunjukkan prevalensi hipertensi pada kelompok usia di atas 65 tahun berada pada angka yang tinggi sehingga menjadi salah satu permasalahan kesehatan masyarakat yang memerlukan perhatian serius [1].

Hipertensi merupakan salah satu penyakit tidak menular yang menjadi penyebab utama morbiditas dan mortalitas di dunia, terutama pada kelompok lanjut usia (lansia) [2]. Peningkatan tekanan darah sering kali dipengaruhi oleh berbagai faktor, salah satunya adalah kadar kolesterol yang tinggi. Penumpukan kolesterol pada dinding pembuluh darah menyebabkan penyempitan arteri sehingga jantung harus bekerja lebih keras untuk memompa darah, yang pada akhirnya meningkatkan risiko hipertensi [3]. Di Indonesia, tingginya jumlah penderita hipertensi serta meningkatnya populasi lansia menuntut adanya pendekatan yang lebih efektif dalam mendukung deteksi dini dan pengambilan keputusan di fasilitas pelayanan kesehatan. Pemanfaatan teknik data mining dan machine learning menjadi salah satu alternatif yang mampu mengolah data kesehatan secara cepat dan objektif untuk menghasilkan informasi yang bermanfaat.

Algoritma Naive Bayes merupakan salah satu metode klasifikasi yang banyak digunakan karena memiliki proses komputasi yang sederhana, efisien, serta mampu memberikan hasil klasifikasi yang baik pada berbagai permasalahan kesehatan [4]. Berbagai penelitian sebelumnya telah membuktikan bahwa Naive Bayes mampu mengklasifikasikan risiko hipertensi dengan tingkat akurasi yang cukup tinggi [5]. Namun, sebagian besar penelitian tersebut hanya berfokus pada identifikasi hipertensi atau tingkat risikonya secara umum, sedangkan klasifikasi status tekanan darah yang dikombinasikan dengan kategori kadar kolesterol pada populasi lansia masih relatif terbatas. Selain itu,

evaluasi model menggunakan pendekatan 10-Fold Cross Validation untuk memperoleh estimasi performa yang lebih stabil juga belum banyak diterapkan pada penelitian sejenis [6].

Berdasarkan kondisi tersebut, penelitian ini mengajukan pertanyaan penelitian, yaitu: (1) bagaimana klasifikasi status tekanan darah berdasarkan kadar kolesterol pada lansia menggunakan algoritma Naive Bayes, (2) seberapa efektif algoritma Naive Bayes dalam mengklasifikasikan status tekanan darah berdasarkan kadar kolesterol, dan (3) bagaimana penerapan 10-Fold Cross Validation dalam **mengevaluasi kinerja model klasifikasi tersebut.** Penelitian ini bertujuan membangun model klasifikasi status tekanan darah berdasarkan kadar kolesterol pada data lansia, mengukur efektivitas algoritma Naive Bayes menggunakan metrik accuracy, precision, recall, dan F1-score, serta mengevaluasi konsistensi performa model melalui teknik 10-Fold Cross Validation.

Penelitian ini memberikan kontribusi pada penerapan metode klasifikasi berbasis machine learning di bidang kesehatan, khususnya dalam pengelolaan data rekam medis lansia. Berbeda dengan penelitian terdahulu, penelitian ini mengombinasikan variabel kadar kolesterol dengan klasifikasi status tekanan darah menggunakan algoritma Naive Bayes yang divalidasi melalui 10-Fold Cross Validation sehingga menghasilkan evaluasi model yang lebih objektif dan representatif.

Hasil penelitian diharapkan dapat menjadi referensi bagi pengembangan sistem pendukung keputusan dalam deteksi dini hipertensi serta menjadi dasar pengembangan penelitian klasifikasi penyakit berbasis data mining pada bidang kesehatan.

2 TINJAUAN PUSTAKA

2.1 Hipertensi

Hipertensi atau tekanan darah tinggi merupakan penyakit kronis yang ditandai dengan meningkatnya tekanan darah pada dinding arteri secara terus-menerus Menurut Indah Sari. Kondisi ini menyebabkan jantung harus bekerja lebih keras untuk memompa darah agar dapat memenuhi kebutuhan oksigen dan nutrisi ke seluruh jaringan tubuh melalui sistem peredaran darah. Tingkat tekanan darah seseorang dapat berubah sesuai dengan kondisi fisiologis, seperti aktivitas fisik, tingkat stres, maupun keadaan emosional.

Dalam menjalankan fungsinya, jantung memompa darah secara berkesinambungan sehingga setiap organ memperoleh suplai oksigen dan zat gizi yang dibutuhkan. Apabila aliran darah berlangsung tanpa hambatan, tekanan yang dihasilkan akan tetap berada dalam batas normal sesuai kebutuhan tubuh. Sebaliknya, adanya penyempitan atau gangguan pada pembuluh darah dapat meningkatkan tekanan yang diperlukan untuk mengalirkan darah. Oleh karena itu, seseorang perlu mewaspadai hasil pemeriksaan tekanan darah yang menunjukkan nilai melebihi 120/80 mmHg saat berada dalam kondisi istirahat. Untuk memastikan hasil tersebut, pengukuran tekanan darah sebaiknya dilakukan sebanyak dua kali dengan selang waktu sekitar lima menit. Dalam konteks ini, angka diatas 120 menunjukkan tekanan sistolik, sedangkan angka 80 atau dibawahnya menunjukkan tekanan diastolik. Hasil yang menunjukkan angka diatas ambang tersebut sudah termasuk dalam kategori prehipertensi.Tekanan sistolik adalah tekanan darah saat jantung memompa darah, sementara tekanan diastolik adalah tekanan darah saat jantung dalam keadaan relaksasi [2].

Klasifikasi Tekanan Darah	Tekanan Darah Sistol (mmHg)	Tekanan Darah Diastol (mmHg)
Normal	<120	<80
Prehipertensi	120—139	80—89
Hipertensi Tahap 1	140—159	90—99
Hipertensi Tahap 2	≥160	≥100

Tabel 1. Klasifikasi Tekanan Darah

2.2 IMT (Indeks Massa Tubuh)

IMT merupakan perbandingan antra berat badan dengan tinggi badan dalam meter kuadrat [2]. Biasanya pengukuran IMT dilakukan pada orang dewasa dengan usia 18 tahun. IMT bisa ditentukan dengan memanfaatkan rumus yang ada dibawah ini.

$$\text{Indeks Massa Tubuh (IMT)} = \frac{\text{Berat Badan (kg)}}{\text{Tinggi Badan (m}^2\text{)}}$$

Rumus 1. Perhitungan IMT

Seseorang dikatakan melebihi IMT jika hasil tersebut berda diatas 25 kg/m^2. Hal tersebut berdasarkan klasifikasi IMT menurut Depkes RI tahun 1994 sesuai dengan yang tercantum pada gambar berikut[2].

2.3 Kolesterol

Kolesterol adalah elemen krusial dalam struktur membran sel di seluruh tubuh dan menjadi salah satu bagian penting di sel saraf serta otak. Senyawa ini banyak terdapat di jaringan hati dan kelenjar, dimana hati berfungsi sebagai organ utama dalam proses pembuatan dan penyimpanan kolesterol. Selain itu, kolesterol memiliki peranan penting sebagai bahan yang diperlukan untuk membentuk berbagai senyawa steroid yang vital, seperti asam empedu, hormon adrenal, estrogen, progesteron, dan vitamin D.

Tingginya tingkat kolesterol bisa menjadi salah satu penyebab risiko terjadinya hipertensi, terutama pada orang tua. Kenaikan kadar kolesterol, terutama yang berupa LDL, dapat mengakibatkan penumpukkan lemak di dinding pembuluh darah. Hal ini akan membuat pembuluh darah menjadi lebih sempit dan keras, sehingga aliran arah ke jantung terhambat. Akibatnya, pasokan oksigen dan nutrisi untuk jantung berkurang, yang membuat jantung bekerja lebih keras dan dapaat mengurangi fungsi otot jantung dalam jangka panjang.

Seiring bertambahnya usia, risiko kolesterol tinggi juga semakin meningkat. Ini disebabkan oleh perubahan dalam metabolisme tubuh, berkurangnya aktivitas fisik, dan pola makan yang terdiri dari lemak jenuh serta kolesterol yang tinggi. Gaya hidup yang tidak sehat menyebabbkan lemak yang masuk ke tubuh tidak dapat digunakan secara efektif, sehingga terakumulasi di jaringan tubuh dan pembuluh darah. Penumpukkan ini dapat meningakatkan kadar kolesterol dalam darah dan meningkatnya kemungkinan

terjadinya hipertensi serta penyakit kardivaskular lainnya [3].

Distribusi kadar Kolesterol lansia ;

Keterangan	Jumlah Kolesterol
Normal	<200 mg/dL
Borderline	200-239 mg/dL
Tinggi	>240 mg/dL

Tabel 2. Keterangan Kolesterol

2.4 Machine Learning

Machine Learning merupakan salah satu bidang dalam kecerdasan buatan *(Artificial Intelligence)* yang mempelajari cara mengembangkan model atau algoritma agar mampu memperoleh pengetahuan dari data. Melalui proses pembelajaran tersebut, sistem dapat mengidentifikasi pola, melakukan analisis terhadap data, serta menghasilkan prediksi atau keputusan secara otomatis dengan keterlibatan manusia yang sangat terbatas. Sistem-sistem ini dirancang untuk belajar mandiri dan meningkatkan akurasinya seiring bertambahnya volume data yang di proses. *Machine Learning* diterapkan secara luas dalam kehidupan sehari-hari, seperti rekomendasi produk di situs belanja online, sistem rekomendasi film di Netflix, dan asisten digital di ponsel pintar. Dengan kemampuan untuk memproses data dalam jumlah besar, pembelajaran mesin dapat menghasilkan prediksi dan keputusan yang semakin akurat seiring berjalannya waktu.

Dalam proses pembelajaran, terdapat dua pendekatan utama: *Supervised Learning* dan *Unsupervised Learning*. Dalam *Supervised Learning*, algoritma dilatih menggunakan data berlabel sehingga model dapat memprediksi keluaran berdasarkan masukan yang diberikan. Proses ini melibatkan perbandingan hasil prediksi dengan standar emas untuk meminimalkan kesalahan, dan kinerjanya biasanya diukur dengan akurasi atau metrik kesalahan seperti MSE dan MAPE. Pendekatan ini umumnya digunakan untuk memprediksi peristiwa massa depan berdasarkan data historis. Maka dari itu, machine leraning telah menjadi metode penting di berbagai bidang, termasuk ilmu soaial dan kodekteran, karena memberikan analisis prediktif yang lebih cepat dan efisien[7].

2.5 Data Mining

Data mining merupakan rangkaian tahapan untuk mengekstrak nilai tambah secara manual dari sekumpulan data, fokus utamanya adalah mengidentifikasi pola tersembunyi, hubungan signifikan, dan informasi berguna dari data yang terkumpul. Tujuannya adalah mengoptimalkan proses bisnis masyarakat dengan memanfaatkan pengetahuan yang diperoleh melalui data mining untuuk meningkatkan efisiensi serta produktivitas operasional. Secara keseluruhan, data mining bukan sekedar alat analisis data, melainkan landasan bagi perbaikan dan inovasi berkelanjutan dalam pengelolaan bisnis serta pengambilan keputusan di era digital [8].

2.6 Naive Bayes Classifier

Naïve Bayes merupakan teknik probabilistik yang sederhana, menghitung peluang suatu kelas dengan memperhatikan frekuensi serta kombinasi nilai atribut dalam kumpulan data. Algoritma ini mengasumsikan setiap atribut bersifat independen secara kondisional terhadap kelas yang diberikan. Keunggulan utama *Naïve Bayes* terletak pada kebutuhan data pelatihan yang relatif sedikit untuk memperkirakan parameter, serta kemampuannya yang sering lebih baik dari ekspektasi dalam kasus nyata yang kompleks, adapun perhitungan *Naïve Bayes* dengan persamaan berikut [4];

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

Keterangan :
X = Data dengan class yang belum diketahui (bukti)
H = Hipotesis data X merupakan satu class spesifikasi
P(H|X) = Probabilitas hipotesis H (prior prob.)
PX = Probabilitas prior bukti X

2.7 Python

Python digambarkan sebagai bahasa pemrograman tingkat tinggi yang berorientasi objek dengan sintaks yang sederhana, sehingga mudah dipelajari baik oleh pemula maupun pengembang berpengalaman. Kesederhanaan sintaks ini memungkinkan pengguna lebih memusatkan perhatian pada pemecahan masalah ketimbang detail teknis bahasa. Selain itu, Python bersifat open-source dan dilengkapi dengan pustaka standar yang luas, menjadikannya sangat fleksibel untuk berbagai kebutuhan pengembangan aplikasi.

Python juga dikenal sebagai bahasa paling populer dalam riset dan pengembangan pembelajaran mesin, mengungguli bahasa lain seperti R, Java, Scala, dan Julia. Popularitas ini didukung oleh ekosistem pustaka dan kerangka kerja yang kaya, seperti NumPy dan SciPy, yang memudahkan pengolahan array, matriks multidimensi, dan fungsi matematika tingkat tinggi. Dengan kombinasi kemudahan penggunaan dan dukungan pustaka yang kuat, Python menjadi pilihan utama untuk membangun aplikasi berbasis teks, grafis, maupun web dalam bidang kecerdasan buatan dan pembelajaran mesin [9].

2.8 Google Colaborator

Google Interactive Notebook, yang lebih dikenal dengan Google Colaboratoryy atau Google Colab, adalah perangkat lunak yang pada dasarnya mirip dengan jupyter Notebook. Ini dapat digunakan secara gratis berbasis cloud dan dioperasikan melalui browser seperti Mozilla Firefox dan Google Chrome. Perbedaannya, Google Colab dirancang khusus untuk programmer yang mengalami kesulitan dalam mengakses spesifikasi

perangkat keras yang tinggi. Google Colab adalah lingkungan pemrograman untuk bahasa Python dengan format "notebook," mirip dengan Jupyter Notebook. Selain itu, Google Colab berguna untuk menyimpan, menulis, dan membagikan program yang telah dibuat melalui Google Drive [10].

Karena layanan ini gratis, Google tentu menerapkan beberapa aturan atau batasan dalam penggunaan Colab. Mesin virtual ini bisa berhenti berfungsi jika tidak ada aktivitas dalam waktu tertentu, dan notebook dapat berjalan hingga maksimum 12 jam. Walaupun begitu, lama waktu penggunaan mesin virtual ini bisa bervariasi. Bagi mereka yang ingin mendapatkan fitur lebih dari Colab, mereka perlu menggunakan versi berbayar, yaitu Colab Pro. Kode yang dijalankan di Colab operate di mesin virtual terpisah berdasarkan akun masing-masing. Meskipun demikian, Google Colab bisa menjadi tempat belajar untuk membuat machine learning tanpa harus memiliki komputer dengan spesifikasi tinggi [10].

Manfaat Google Colab

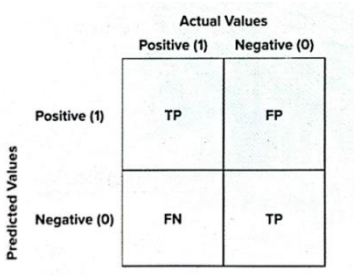
1. Akses gratis ke mesin berkecepatan tinggi (GPU dan TPU);
2. Tidak perlu menginstal apapun di laptop atau komputer pribadi;
3. Google Colab mendukung kolaborasi dengan pengguna lain melalui berbagi kode secara online;
4. Memudahkan integrasi antara Google Colab dan jupyter notebook di komputer lokal;
5. Mudah menghubungkan dengan Google Drive atau GitHub;
6. Menyediakan koleksi built-in library machine learning yang paling populer, seperti Keras, TensorFlow, dan PyTorch;
7. Berbasis cloud sehingga tidak membutuhkan ruang penyimpanan di komputer;
8. Data di Google Colaboratory dapat diakses dan dimodifikasi.

2.9 K-Fold Cross Validation

Pada tahapan ini, seluruh data dibagi menjadi sepuluh kelompok (*fold*) dengan jumlah data yang hampir seimbang menggunakan teknik ***10-Fold Cross Validation***. Dalam setiap putaran pengujian, sembilan *fold* dimanfaatkan sebagai data pelatihan (*training data*) untuk membentuk model ***Naïve Bayes***, sementara satu *fold* sisanya digunakan sebagai data pengujian (*testing data*) untuk menilai kemampuan model dalam melakukan klasifikasi. Proses validasi dilaksanakan sebanyak sepuluh kali, dengan setiap *fold* secara bergantian berfungsi sebagai data pengujian hingga seluruh data memperoleh kesempatan yang sama untuk diuji. Pendekatan ini bertujuan menghasilkan evaluasi model yang lebih objektif dan mampu mengurangi potensi bias akibat pembagian data yang hanya dilakukan satu kali. Selanjutnya, hasil evaluasi dari seluruh iterasi dihitung nilai rata-ratanya sehingga diperoleh estimasi performa model yang lebih stabil, objektif, dan mampu memberikan hasil klasifikasi status tekanan darah berdasarkan kadar kolesterol yang lebih jelas [11].

2.10 Confusion Matrix

Confusion Matrix ialah alat yang digunakan untuk memvisulisasikan kinerja model kecerdasan buatan sesuia dengan data sampel yang hasil sebenarnya sudah diketahui. *Confusion Matrix* membantu dalam menganalisis prediksi prediksi yang dibuat oleh model, memungkinakan identifikasi kategori-kategori yang diklasifikasikan dengan benar atau salah. Dengan demikian, *confusion matrix* memungkinkan evaluasi yang komprehensif terhadap keunggulan dan kelemahan model dalam mengklasifikan data. Dalam penggunaannya, *confusion matrix* sering digunakan dengan konsep *True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN).* Konsep-konsep ini memudahkan pemahaman kinerja model dalam membedakan kelas positif dan negatif yang sebenarnya [12].

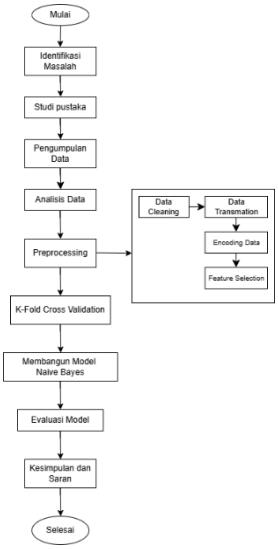


Gambar 1. Distribusi Confusion Matrix

3 METODOLOGI

3.1 Proses dan Langkah Penelitian

Penelitian ini memakai pendekatan kuantitatif dengan metode analisis klasifikasi. Tujuan penelitian ialah menggolongkan kondisi kesehatan lansia berdasarkan kategori kolesterol dan tekanan darah, serta menilai ketepatan model klasifikasi menggunkana algoritma *Naïve Bayes*. Penelitian ini dilakukan dengan cara terstruktur dari tahap pengumpulan informasi sampai penilaian model pengklasifikasian.



Gambar 2. Kerangka Penelitian

Penjelasan alur Kerangka Penelitian :

1. Mulai

Melakukan observasi secara langsung kepada pihak Puskesmas Setiabudi mengenai klasifikasi status tekanan darah berdasarkan kadar kolesterol [13].

2. Identifikasi Masalah

Menganalisis kasus status tekanan darah berdasarkan kadar Kolesterol serta kebutuhan klasifikasi status tekanan darah selanjutnya ditetapkan tujuan penelitian, yaitu membangun model klasifikasi status tekanan darah Berdasarkan kadar Kolesterol [14]

3. Studi Pustaka

Mengumpulkan landasan teori dan referensi terkait Kolesterol, Hipertensi, serta metode *Naïve Bayes*, Klasifikasi, *Python*, Data Mining, *Machine Learning*, *Google Colaboratory, K-Fold Cross Validation, Confusion Matrix*. Diperoleh dari jurnal ilmiah, dan buku [15].

4. Pengumpulan Data

Penelitian ini menggunakan data sekunder yang diperoleh dari rekam medis Puskesmas Setiabudi melalui proses perizinan yang meliputi pengajuan surat izin riset dari Universitas Bina Sarana Informatika, permohonan izin pengambilan data kepada Sudinkes Jakarta Selatan dengan tembusan ke Puskesmas Setiabudi, serta pengajuan permohonan data penelitian yang disertai proposal penelitian. Setelah memperoleh persetujuan, peneliti menerima data pasien lansia tahun 2025.

5. Analisis Data

Teknik pengambilan data pasien lansia yang diperoleh dari Puseksmas Setiabudi Tahun 2025 menggunakan *purposive sampling*, yaitu pemilihan data berdasarkan kriteria yang sesuai dengan tujuan penelitian. Data yang tidak lengkap, duplikat, atau tidak valid dikeluarkan dari analisis . Dengan demikian, data yang digunakan hanya mencakup data pasien lansia dengan variable [16], seperti no, usia, jenis kelamin, Kolesterol, berat badan, tinggi badan, IMT, lingkar lengan atas , lingkar perut, tekanan darah *sistole*, dan tekanan darah *diastole* [17].

6. Preprocessing

Tahap *preporocessing* bertujuan untuk mengubah data pasien lansia menjadi data bersih dan siap digunakan dalam pengolahan data menggunakan pemodelan Naive Bayes sehingga dapat meingkatkan akurasi klasifikasi [18]:

a) *Data Cleaning* : Mengecek tipe data, data kosong, dan menampilkan tipe data, dan nilai kosong pada data
b) *Data Transformation* : Mengubah tipe data kategorikal menjadi numerik.
c) *Encoding Data* : Mengubah label klasifikasi menjadi bentuk numerik.
d) *Feature Selection* : Menentukan variabel input dan output penelitian.

7. K-Fold Cross Validation

Pada tahap ini, data dibagi menjadi 10 *fold* (bagian) dengan jumlah data pada setiap *fold* yang relatif sama menggunakan metode 10-*Fold Cross Validation*. Pada setiap tahap iterasi, 9 *fold* dimanfaatkan sebagai data pelatihan (*training data*) untuk membangun model *Naïve Bayes*, sedangkan 1 *fold* lainnya digunakan sebagai data pengujian (*testing data*) untuk mengevaluasi kinerja model [11].

8. Membangun Model *Naïve Bayes*

Data yang sudah bersih digunakan untuk membangun model klasifikasi dengan algoritma *Naïve Bayes* untuk menentukan Kategori Tekanan Darah [19] .

9. Evaluasi Model

Evaluasi terhadap model dilakukan menggunakan *Confusion Matrix* dengan membandingkan hasil prediksi yang dihasilkan oleh model dengan nilai aktual pada data pengujian. Berdasarkan hasil perbandingan tersebut, performa model kemudian diukur menggunakan beberapa metrik evaluasi, seperti *accuracy, precision, recall, dan F1-score*, untuk menilai tingkat akurasi serta keandalan model dalam melakukan klasifikasi.:

a) *Accuracy* : Mengukur ketepatan klassifikassi secara keseluruhan.
b) *Presion* : Mengukur ketepatan prediksi suatu kelas.
c) *Recall* : Mengukur kemampuan model menemukan data dari suatu kelas.
d) *F1-Score*: Kombinasi precision dan recall dalam satu nilai.
e) *Confusion Matrix*: Tabel yang menunjukkan perbandingan antara hasil prediksi dan data aktual.

10. Kesimpulan dan Saran

Peneliti membuat kesimpulan dari hasil penelitian dan memberikan saran untuk pengembangan penelitian selanjutnya.

11. Selesai

Semua tahapan proses penelitian telah selesai dilakukan.

4 HASIL DAN PEMBAHASAN

4.1 Pengumpulan Data

Data Diperoleh dari Puskesmas Setiabudi dengan persetujuan Dinas Kesehatan Jakarta Selatan. Proses ini menggunakan Bahasa pemrograman *Python* melalui platform *Google Colab.* Data yang dikumpulkan berupa data kuantitatif.

Setelah data diberikan sebanyak 1099 data, seperti yang diperlihatkan pada table III.1 dibawah ini:

No	Usia	JK	Kol	BB	TB	IMT	LLA	LP	TDD	TDS
1	72	P	144	58	165	21.6	10	98	115	85
2	68	L	220	53	155	22.4	27	88	128	61
3	60	L	142	74	161	28.5	39	100	144	87
4	67	L	210	56	160	22.1	29	80	143	61
5	60	L	226	52	160	20.3	28	89	120	89
6	78	P	210	67	156	27.5	30	98	124	84
7	66	P	299	51	143	24.9	29	97	118	80
8	62	L	210	54	155	22.5	23	52	127	88
9	91	L	210	58	165	21.3	22	88	143	70
...	...	...	...	...	...	...	...	...	...	...
1099	70	L	200	70	160	27.34	29	89	159	80

Sumber: (Penelitian,2026)

Tabel 3. Sample Data

Selanjutnya menampilkan Deskripsi dari Data tersebut

Variabel	Definisi Singkat	Kegunaan
No	Nomor urut setiap data atau responden.	Sebagai identitas data untuk memudahkan pengelolaan dan pencarian data.
Usia	Umur responden yang dinyatakan dalam tahun.	Digunakan untuk mengetahui pengaruh pertambahan usia terhadap kadar kolesterol dan tekanan darah.
Jenis Kelamin	Identitas biologis responden (laki-laki atau perempuan).	Digunakan untuk menganalisis perbedaan risiko kesehatan berdasarkan jenis kelamin.
Kolesterol	Kadar kolesterol total dalam darah, biasanya diukur dalam mg/dL.	Digunakan untuk menentukan status kadar kolesterol dan sebagai indikator risiko penyakit kardiovaskular.
Berat Badan	Massa tubuh responden yang diukur dalam kilogram (kg).	untuk menghitung Indeks Massa Tubuh (IMT) dan menilai status gizi.
Tinggi Badan	Tinggi tubuh responden yang diukur dalam sentimeter (cm) atau meter (m).	Digunakan bersama berat badan untuk menghitung IMT.
IMT (Indeks Massa Tubuh)	Indikator status gizi yang dihitung dari berat badan (kg) dibagi kuadrat tinggi badan (m^2).	Dipakai untuk mengklasifikasikan status gizi, seperti kurus, normal, kelebihan berat badan, atau obesitas.
Lingkar Lengan Atas	Ukuran Panjang Lengan atas dari atas punggung atas hingga ujung jari	sebagai variabel prediktor yang menggambarkan status gizi seseorang.
Lingkar Perut	Ukuran keliling perut yang diukur dalam sentimeter (cm).	untuk mengidentifikasi obesitas sentral yang berkaitan dengan peningkatan risiko penyakit metabolik dan kardiovaskular
Tekanan Darah Sistole	Tekanan darah pada saat jantung berkontraksi memompa darah ke seluruh tubuh, diukur dalam mmHg.	untuk menentukan kondisi tekanan darah dan mendeteksi hipertensi.
Tekanan Darah Diastole	Tekanan darah saat jantung berelaksasi di antara dua denyut, diukur dalam mmHg.	Digunakan bersama tekanan darah sistole untuk menentukan klasifikasi tekanan darah dan risiko hipertensi.

Tabel 4. Deskripsi Data

4.2 *Preprocessing*

Mengidentifikasi *data cleaning* dengan :

1. Cek *missing value* dengan *data.isnull().sum().* Hasil pemeriksaan menunjukkan bahwa seluruh variabel tidak memiliki nilai kosong.
2. Cek data duplikat dengan *jumlah_duplikat = data.duplicated().sum().* Hasil pengecekan menjelaskan bahwa jumlah data duplikat pada dataset adalah 0, yang berarti tidak terdapat baris data yang memiliki nilai identik.
3. Cek tipe data dengan print (data.dtypes)

```
NO                        int64
Usia                    float64
Jenis Kelamin            object
Kolestrol                 int64
Berat Badan               int64
Tinggi Badan            float64
IMT                     float64
Lingkar Lengan Atas       int64
Lingkar Perut             int64
Tekanan Darah Sistole     int64
Tekanan Darah Diastole    int64
dtype: object
```

Gambar 1. Tampilan dari Pengecekkan Tipe Data

4. *Data Transformation*

untuk melakukan pengkodean data kategorikal ke bentuk numerik sehingga sesuai dengan kebutuhan algoritma *Naïve Bayes*. Proses diawali dengan mengimpor kelas *LabelEncoder* dari *library scikit-learn*, kemudian membuat objek *LabelEncoder* yang digunakan untuk melakukan transformasi data. Selanjutnya, program melakukan perulangan pada seluruh kolom yang bertipe *object* (teks) dan mengubah setiap nilai kategorikal menjadi nilai numerik. Kolom Keterangan tidak disertakan pada tahap ini karena merupakan variabel label yang akan diproses secara terpisah pada tahap *encoding data*. Setelah proses transformasi selesai, fungsi *print(data.dtypes)* digunakan untuk memastikan bahwa tipe data setiap kolom telah berubah sesuai kebutuhan, sedangkan *data.head()* digunakan untuk melihat lima record pertama dataset sebagai verifikasi hasil transformasi [20].

```
Jenis Kelamin                int64
Berat Badan                  int64
Tinggi Badan               float64
Usia                         int64
Kolestrol                    int64
IMT                        float64
Lingkar Lengan Atas          int64
Lingkar Perut                int64
Tekanan Darah Sistole        int64
Tekanan Darah Diastole       int64
dtype: object
```

Gambar 2. Tampilan Hasil Pengubahan Tipe data

5. Pembentukkan Label Status Tekanan Darah

Pada tahap ini dilakukan pembentukan label kelas berdasarkan kombinasi kategori kadar kolesterol serta nilai tekanan darah sistolik dan diastolik. Proses tersebut menghasilkan sembilan kategori status tekanan darah yang selanjutnya digunakan sebagai variabel target dalam proses klasifikasi.

NO	Usia	Jenis Kelamin	Kolestrol	Berat Badan	Tinggi Badan	IMT	Lingkar Lengan Atas	Linghar Perut	Tekanan Darah Sistole	Tekanan Darah Diastole	Status_Tekanan_Darah_Kolestrol
1	117	1	144	58	165.0	21.634527	10	98	115	85	Normal - Stadium 2 (Sedang)
2	91	0	220	53	155.0	22.434964	27	88	128	61	Batas Tinggi - Stadium 2 (Sedang)
3	42	0	142	74	161.0	28.548281	39	100	144	87	Normal - Stadium 3 (Berat)
4	85	0	210	56	160.0	22.109375	29	80	143	61	Batas Tinggi - Stadium 3 (Berat)
5	42	0	226	52	160.0	20.312500	28	89	120	89	Batas Tinggi - Stadium 2 (Sedang)

Gambar 3. Hasil Pembentukkan label

Hasil pembentukan label status tekanan darah pada data. Berdasarkan hasil tersebut, setiap data telah memiliki kategori status tekanan darah, seperti **Normal-Stadium 1**, **Batas Tinggi-Stadium 1**,**Tinggi-Stadium 1**, **Normal-Stadium 2**, **Batas Tinggi-Stadium 2**, **Tinggi-Stadium 2**, **Normal-Stadium 3**, **Batas Tinggi-Stadium 3**, dan **Tinggi-Stadium 3**.

1. *Encoding* Label

Label status tekanan darah yang masih berbentuk kategorikal diubah menjadi nilai numerik menggunakan proses encoding. Transformasi ini bertujuan agar data dapat diproses oleh algoritma Naive Bayes selama proses pelatihan dan pengujian model.

	NO	Usia	Jenis Kelamin	Kolestrol	Berat Badan	Tinggi Badan	IMT	Lingkar Lengan Atas	Lingkar Perut	Tekanan Darah Sistole	Tekanan Darah Diastole	Status_Tekanan_Darah_Kolestrol
0	1	117	1	144	58	165.0	21.634527	10	98	115	85	4
1	2	91	0	220	53	155.0	22.434964	27	88	128	61	1
2	3	42	0	142	74	161.0	28.548281	39	100	144	87	5
3	4	85	0	210	56	160.0	22.109375	29	80	143	61	2
4	5	42	0	226	52	160.0	20.312500	28	89	120	89	1

Gambar 4. Hasil *Encoding* label

2. Memisahkan Fitur dan Label

Pada tahap penentuan variabel, data dibagi menjadi variabel independen (fitur) dan variabel dependen (label) [8]. Variabel independen meliputi usia, jenis kelamin, kadar kolesterol, berat badan, tinggi badan, IMT, lingkar lengan atas, lingkar perut, tekanan darah sistole, dan tekanan darah diastole, sedangkan variabel dependen adalah status tekanan darah. Selanjutnya dilakukan pemeriksaan jumlah data dan distribusi kelas pada variabel target untuk mengetahui proporsi setiap kelas serta memastikan kesiapan data sebelum proses pelatihan dan evaluasi model klasifikasi.

```
Jumlah data: 1099

Distribusi kelas:
Status_Tekanan_Darah_Kolestrol
5    331
4    297
3    144
2    118
1     65
0     61
8     40
7     22
6     21
Name: count, dtype: int64
```

Gambar 5. Hasil Pemisahan Sesuai Label

Dataset penelitian berjumlah 1.099 data yang terbagi ke dalam sembilan kelas. Hasil distribusi menunjukkan adanya ketidakseimbangan kelas, di mana kelas 5 memiliki jumlah data terbanyak (331 data), sedangkan kelas 6 hanya terdiri dari 21 data.

4.3 Penerapan *K-Fold Cross validation*

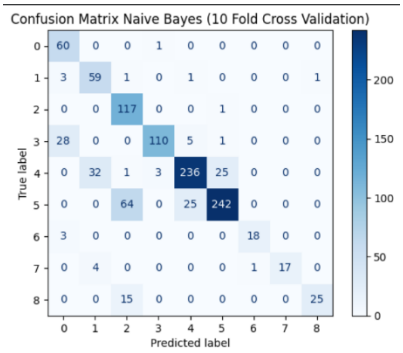
Berdasarkan hasil analisis distribusi kelas, dataset penelitian termasuk *imbalanced dataset*, sehingga digunakan *Stratified K-Fold Cross Validation* untuk mempertahankan ratio setiap kelas pada setiap fold agar hasil evaluasi lebih representatif. Pada penelitian ini digunakan *10 fold (n_splits=10)*, sehingga seluruh data memperoleh kesempatan menjadi data latih dan data uji secara bergantian. Parameter *shuffle=True* digunakan untuk menyusun ulang urutan data sebelum proses pembentukan *fold* sehingga penyebaran data pada setiap *fold* menjadi lebih merata. sedangkan *random_state=42* digunakan untuk menjaga konsistensi pembagian data sehingga hasilnya dapat direplikasi. Tahap ini hanya menginisialisasi konfigurasi validasi silang dan belum melakukan tahapan penyusunan model *Naïve Bayes* dan pengujian kinerjanya.

4.4 Membangun model *Naive Bayes*

Model Naive Bayes dibangun menggunakan data pelatihan untuk mempelajari hubungan antaratribut terhadap status tekanan darah. Model yang telah terbentuk kemudian digunakan untuk melakukan prediksi pada data pengujian sebagai dasar evaluasi performa klasifikasi.

4.5 Evaluasi Model

1. *Confusion Matrix*



Gambar 6. Hasil *Confusion Matrix*

Confusion Matrix digunakan untuk mengevaluasi kinerja algoritma Naive Bayes dengan membandingkan hasil prediksi terhadap data aktual setelah proses 10-Fold Cross Validation. Berdasarkan hasil evaluasi, model berhasil mengklasifikasikan 884 dari total 1.099 data dengan benar sehingga memperoleh tingkat akurasi sebesar **80,44%**. Nilai diagonal utama yang mendominasi menunjukkan bahwa sebagian besar data berhasil diprediksi sesuai dengan kelas sebenarnya. Meskipun demikian, masih terdapat beberapa kesalahan klasifikasi pada beberapa kelas, terutama antara kelas yang memiliki karakteristik serupa. Secara keseluruhan, hasil Confusion Matrix menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam melakukan klasifikasi, meskipun masih terdapat peluang untuk meningkatkan performa pada kelas-kelas yang sering mengalami kesalahan prediksi.

2. Hasil *Classification Report*

	precision	recall	f1-score	support
0	0.64	0.98	0.77	61
1	0.62	0.91	0.74	65
2	0.59	0.99	0.74	118
3	0.96	0.76	0.85	144
4	0.88	0.79	0.84	297
5	0.90	0.73	0.81	331
6	0.95	0.86	0.90	21
7	1.00	0.77	0.87	22
8	0.96	0.62	0.76	40
accuracy			0.80	1099
macro avg	0.83	0.83	0.81	1099
weighted avg	0.84	0.80	0.81	1099

Accuracy : 80%
Precision : 84%
Recall : 80%
F1-Score : 81%

Gambar 7. Hasil *Classification Report*

classification report menghasilkan ***accuracy 80%***, ***weighted precision 84%***, ***weighted recall 80%***, *dan* ***weighted F1-score 81%***. Hasil tersebut menunjukkan bahwa algoritma Naive Bayes memiliki kinerja yang cukup baik dalam mengklasifikasikan status tekanan darah berdasarkan kadar kolesterol pada lansia. Namun, masih terdapat kesalahan klasifikasi pada beberapa kelas akibat kemiripan karakteristik data dan distribusi kelas yang tidak seimbang.

5 KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil menerapkan algoritma Naive Bayes untuk mengklasifikasikan status tekanan darah berdasarkan kadar kolesterol pada lansia di Puskesmas Setiabudi Tahun 2025 menggunakan variabel jenis kelamin, usia, berat badan, tinggi badan, IMT, lingkar lengan atas, lingkar perut, tekanan darah sistolik, dan tekanan darah diastolik. Model menghasilkan sembilan kategori status tekanan darah yang merepresentasikan kombinasi kondisi tekanan darah dan kadar kolesterol.

Hasil evaluasi menggunakan metode 10-Fold Cross Validation menunjukkan bahwa model memperoleh nilai accuracy sebesar 80,44%, weighted precision sebesar 84%, weighted recall sebesar 80%, dan weighted F1-score sebesar 81%. Hasil tersebut menunjukkan bahwa algoritma Naive Bayes memiliki kinerja yang cukup baik dalam mengklasifikasikan status tekanan darah berdasarkan kadar kolesterol pada lansia. Selain itu, penerapan 10-Fold Cross Validation memberikan evaluasi yang lebih objektif dan stabil karena seluruh data memperoleh kesempatan yang sama sebagai data pelatihan maupun data pengujian.

5.2 Saran

Penelitian selanjutnya disarankan menggunakan jumlah data yang lebih besar dan lebih beragam, menambahkan variabel yang berpotensi memengaruhi tekanan darah, menerapkan teknik penyeimbangan data untuk mengatasi ketidakseimbangan kelas, serta membandingkan performa Naive Bayes dengan algoritma klasifikasi lainnya. Selain itu, hasil penelitian dapat dikembangkan menjadi sistem pendukung keputusan berbasis web atau aplikasi guna membantu tenaga kesehatan dalam melakukan skrining awal risiko hipertensi pada lansia.

UCAPAN TERIMA KASIH

Dalam penyusunan penelitian ini, penulis memperoleh dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan apresiasi dan rasa terima kasih kepada:

a. Seluruh sivitas akademika Universitas Bina Sarana Informatika selama proses perkuliahan.
b. Dosen Pembimbing yang telah meluangkan waktu memberikan arahan dengan baik.

DAFTAR PUSTAKA

[1] A. P. Lasriado Girsang, N. Riana Sari, F. W. Rosmala Dewi, and K. Tri Yulianto, "Statistik Penduduk Lanjut Usia 2025," vol. 22, p. 250, 2025, [Online]. Available: https://www.bps.go.id/id/publication/2025/12/12/868d335b088dcddc3ddee052/statistik-penduduk-lanjut-usia-2025.html

[2] Y. N. Indah Sari, *Berdamai dengan Hipertensi*. Jakarta: Bumi Medika, 2022. [Online]. Available: https://ipusnas2.perpusnas.go.id/book/2de504a7-2ad1-44db-81c7-12d4f4156a26/789493d9-4f7c-48d1-ad32-e2c120461f68

[3] I. F. Lesar, D. Modjo, and A. A. Sudirman, "Hubungan

Antara Kadar Kolesterol Dalam Darah dengan Kejadian Hipertensi Pada Lansia di Pkm Tabongo Kabupaten Gorontalo," vol. 1, no. 2, 2023, doi: https://doi.org/10.59680/medika.v1i2.267.

[4] R. Swastika, S. Mukodimah, F. Susanto, M. Muslihudin, and S. Ipnuwati, *Naive Bayes .pdf*. Jawa Barat: CV Adanu Abimata, 2023.

[5] A. A. Hatina, R. Rismayati, B. Fitria, R. Fatimatuzzahra, and R. Amrullah, "JTIM : Jurnal Teknologi Informasi dan Multimedia Klasifikasi Gizi Lansia menggunakan Metode Naïve Bayes," vol. 6, no. 2, pp. 84–96, 2024, doi: https://doi.org/10.35746/jtim.v6i2.502.

[6] M. M. Laia *et al.*, "Analisis Sentimen Program Makan Gratis Pada Platform X Algoritma Naïve Bayes Menggunakan," 2025, doi: https://doi.org/10.33884/jif.v13i02.10427.

[7] M. C. Dr. Yahya Heryadi;Teguh Wahyono, *Machine Learning*, 1st ed. Yogyakarta: PENERBIT GAVA MEDIA, 2023.

[8] D. R. Fitriana, ST., MM., IPM, A. N. Habyba, STP., M.Si, and E. F. STP., M.Si, *Data Mining .pdf*. Banyumas: Wawasan Ilmu, 2022.

[9] E. Suprihadi, *Python.pdf*. Sleman: Deepublish Publisher, 2022.

[10] Wardana, *Dasar-Dasar Data Science dan Aplikasinya Dengan Python*, 1st ed. Jawa Tengah: Wawasan Ilmu, 2024.

[11] K. C. V. Approaches, D. K. Nisa, F. M. Hana, and S. Ulya, "Klasifikasi Kinerja Penjualan Produk Nike Menggunakan Algoritma Random Forest dengan Pendekatan Hold-Out dan K-Fold Cross Validation," vol. 23, no. 1, pp. 79–92, 2026, doi: 10.30595/sainteks.v23i1.29930.

[12] S. Riyadi, *Machine Learning Metode dan Evaluasi*. Yogyakarta: UMY Press, 2024.

[13] M. D. Jannah, A. Fauzi, C. E. Sukmawati, and H. Hikmayanti, "Klasifikasi Penyakit Hipertensi Menggunakan Support Vector Machine dan Naive Bayes," pp. 905–913, 2026, doi: 10.33364/algoritma/v.22-1.2316.

[14] M. N. Zanis, U. Islam, S. Agung, and A. Pendahuluan, "IMPLEMENTASI METODE NAIVE BAYES UNTUK KLASIFIKASI RISIKO HIPERTENSI," vol. 8, no. 2, 2025, doi: https://ojs.unpkediri.ac.id/index.php/noe.

[15] J. Lemantara, "Rancang bangun aplikasi hipertensi . edu sebagai media edukasi dan diagnosis penyakit hipertensi menggunakan metode naïve bayes dengan laplace correction Design and implementation of hipertensi . edu application as media for education and diagnosis of hyp," vol. 5, pp. 146–160, 2024, doi: 10.37373/infotech.v5i1.1197.

[16] C. F. A. Mohtar, O. A. Widyayanti, and M. I. N. Pratiwi, "JURNAL NURSE Hubungan Kadar Kolesterol Total dengan Tekanan Darah pada Pasien Hipertensi di Rumah Sakit Tingkat III 04.06.01 Wijaya Kusuma Purwokerto," *J. Nurse*, vol. 5, no. 2, pp. 1–10, 2022, [Online]. Available: https://doi.org/10.57213/nurse.v5i2.235

[17] M. M. Said, R. Soekarta, and I. Amri, "Sistem Pakar Mendiagnosa Penyakit Hipertensi Dengan Metode Naïve Bayes Berbasis Web (Studi Kasus RSUD Sele Be Solu Kota Sorong)," vol. 01, no. 02, pp. 67–74, 2023, doi: https://doi.org/10.33506/jiki.v1i02.2727.

[18] G. D. Silolongan, F. A. Masse, and D. Kusumawati, "JURNAL LOCUS : Penelitian & Pengabdian Klasifikasi Penyakit Hipertensi Menggunakan Algoritma NAÏVE Di Rumah Sakit Undata Palu," vol. 4, no. 8, pp. 7467–7476, 2025, doi: https://doi.org/10.58344/locus.v4i8.4536.

[19] M. A. Aprihartha and I. Idham, "Optimization of Classification Algorithms Performance with k-Fold Cross Validation," *Eig. Math. J.*, vol. 7, no. 2, pp. 61–66, 2024, doi: 10.29303/emj.v7i2.212.

[20] O. Rainio, "Evaluation metrics and statistical tests for machine learning," *Sci. Rep.*, pp. 1–14, 2024, doi: 10.1038/s41598-024-56706-x.

BIODATA PENULIS

Vina Nur Amalia Putri
Merupakan Mahasiswa Program Studi Informatika Fakultas Teknik dan Informatika di Universitas Bina Sarana Informatika.

Fuad Nur Hasan, S. Kom, M. Kom
Merupakan Dosen Fakultas Teknik dan Informatika di Universitas Bina Sarana Informatika

Lestari Yusuf , S. Kom, M. Kom
Merupakan Dosen Fakultas Teknik dan Informatika di Universitas Bina Sarana Informatika