

Aplikasi Pendeteksian Dini Hipertensi Pada Lansia Berbasis Klasifikasi Data Menggunakan Multinomial naïve Bayes

Diva Rizky Alfitrah, Ari Kurniawan Saputra

Program studi Informatika, Universitas Bandar Lampung
Program studi Informatika, Universitas Bandar Lampung
Corresponding author: diva.21421036@student.ubl.ac.id

Abstract:

This research aims to develop a data classification-based application capable of early detection of hypertension in the elderly. This system is designed to identify risk factors for hypertension derived from the lifestyle of the elderly by utilizing the Multinomial Naïve Bayes algorithm. The method implemented in this research is the Multinomial Naïve Bayes (MNB) classification algorithm to predict the probability of hypertension risk based on categorical lifestyle data, such as smoking habits, physical activity, stress levels, and diet. This research successfully implemented the Multinomial Naïve Bayes algorithm into a functional web application for early detection of hypertension in the elderly. Based on testing using the K-10 Fold Cross Validation method, the elderly lifestyle data classification model achieved an accuracy rate of 80.55 percent.

Keyword: Hypertension, Elderly, Data Classification, Multinomial Naïve Bayes, K-Fold Cross Validation

Abstrak:

Penelitian ini bertujuan untuk mengembangkan aplikasi berbasis klasifikasi data yang mampu mendeteksi dini hipertensi pada lansia. Sistem ini dirancang untuk mengidentifikasi faktor risiko hipertensi yang berasal dari gaya hidup lansia dengan memanfaatkan algoritma Multinomial Naïve Bayes. Metode yang diterapkan dalam penelitian ini adalah algoritma klasifikasi Multinomial Naïve Bayes (MNB) untuk memprediksi probabilitas risiko hipertensi berdasarkan data gaya hidup kategorikal, seperti kebiasaan merokok, aktivitas fisik, tingkat stres, dan pola makan. Penelitian ini berhasil mengimplementasikan algoritma Multinomial Naïve Bayes ke dalam aplikasi web fungsional untuk deteksi dini hipertensi pada lansia. Berdasarkan pengujian menggunakan metode K-10 Fold Cross Validation, model klasifikasi data gaya hidup lansia mencapai tingkat akurasi sebesar 80,55 persen.

Kata Kunci: Hipertensi, Lansia, Klasifikasi Data, Multinomial Naïve Bayes, K-Fold Cross-Validation

Pendahuluan

mendorong transformasi di berbagai sektor, termasuk pendidikan. Universitas sebagai lembaga pendidikan memerlukan sistem informasi yang efisien untuk mendukung aktivitas administratif, salah satunya adalah pengelolaan data kesehatan seperti deteksi dini penyakit. Hipertensi, atau tekanan darah tinggi, merupakan kondisi di mana tekanan darah melebihi 140/90 mmHg, dan sering disebabkan oleh pola hidup tidak sehat, terutama pada kelompok lanjut usia (lansia). Menurut WHO, lansia adalah individu yang telah mencapai usia lanjut, yang diklasifikasikan lebih lanjut menjadi beberapa kriteria seperti usia pertengahan (45-59 tahun), lanjut usia (60-74 tahun), dan lanjut usia tua (75-90 tahun) [1]. Penelitian oleh Warjiman, dkk. [2]. Menunjukkan bahwa 65,71% lansia penderita hipertensi memiliki kebiasaan buruk seperti konsumsi makanan asin dan kafein. Karena gejalanya yang sering tidak terlihat, hipertensi dijuluki sebagai "silent killer", sehingga pendeteksian dini menjadi sangat penting untuk mencegah komplikasi serius. Pendeteksian dini dapat dilakukan dengan mengklasifikasikan data pola hidup menggunakan teknik data mining. Algoritma Multinomial Naïve Bayes (MNB) merupakan salah satu metode yang efektif untuk kasus ini karena kelebihanannya dalam menangani data kategorikal, memiliki komputasi yang cepat, dan mampu menyelesaikan permasalahan multiple class [3]. Penelitian sebelumnya telah menunjukkan keberhasilan MNB. Zulkipli, dkk [4]. Berhasil memprediksi tingkat kesehatan lingkungan dengan akurasi 93.83%, penelitian oleh Rizki Surya Pratama, dkk [5]. Mencapai akurasi 83%, dan studi oleh Hidayah [6]. Juga menunjukkan hasil yang sangat baik dengan akurasi 85%. Meskipun MNB telah banyak digunakan, penerapan spesifik untuk deteksi hipertensi pada lansia dengan validasi yang kuat masih perlu dieksplorasi. Berdasarkan hal tersebut, penelitian ini bertujuan untuk mengimplementasikan algoritma MNB pada rancang bangun aplikasi pendeteksian dini hipertensi pada lansia. Untuk memastikan validitas dan akurasi model, penelitian ini menggunakan metode K-Fold Cross-Validation. Diharapkan, sistem yang dibangun dapat menjadi alat bantu informasi yang efektif untuk mendukung pengambilan keputusan medis dan meningkatkan kualitas hidup lansia.

Metodologi

A. Data

Data yang dibutuhkan pada penelitian ini berupa dataset rekam medis pasien lansia dari tahun 2023-2024, yang mencakup total 308 data dengan 7 atribut atau variabel pola hidup. Dataset ini menjadi acuan utama dalam proses implementasi dan pelatihan algoritma Multinomial Naïve Bayes. Data-data tersebut didapatkan dari puskesmas tanjung agung.

B. Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan 2 metode, pengumpulan data dilakukan agar data tersebut dapat diolah sebelum proses pembuatan sistem:

1. Observasi

Pada tahap ini, penulis melakukan pengamatan langsung di Puskesmas Tanjung Agung untuk memahami alur kerja dan proses yang berjalan terkait pendataan dan diagnosis dini hipertensi pada pasien lansia.

2. Wawancara

Untuk mendapatkan data dan pemahaman yang lebih mendalam, penulis melakukan wawancara semi-terstruktur dengan dr. Dedi selaku perwakilan tenaga kesehatan di Puskesmas Tanjung Agung. Wawancara ini bertujuan untuk memvalidasi variabel-variabel yang relevan untuk deteksi hipertensi dan untuk memperoleh data historis yang akan digunakan sebagai dataset pelatihan dan pengujian model.

3. Pra-Proses

Sebelum data digunakan untuk pemodelan, dilakukan tahap pra-proses untuk memastikan kualitas data. Tahap ini mencakup penanganan nilai yang hilang (missing value) dan pemisahan data. Untuk nilai yang hilang, sistem dirancang untuk mengabaikan atribut tersebut selama kalkulasi probabilitas tanpa menghentikan proses prediksi. Selanjutnya, data dibagi menjadi dua bagian dengan metode hold-out: 80% sebagai data latih (training set) untuk membangun model, dan 20% sebagai data uji (testing set) untuk mengevaluasi kinerja model secara objektif.

C. Metode

Penelitian ini mengimplementasikan dua metode utama: Multinomial Naïve Bayes untuk klasifikasi dan K-Fold Cross Validation untuk evaluasi model.

1. Multinomial Naïve Bayes

MNB adalah varian dari Naïve Bayes yang dirancang khusus untuk data diskrit atau kategorikal yang merepresentasikan frekuensi. Algoritma ini sangat efisien dan bekerja dengan menghitung probabilitas suatu data masuk ke dalam kelas tertentu berdasarkan probabilitas kemunculan setiap fitur pada kelas tersebut. Keputusan klasifikasi didasarkan pada probabilitas posterior tertinggi yang dihitung menggunakan prinsip Teorema Bayes. Perhitungan skor akhir untuk setiap kelas (C) dilakukan dengan menjumlahkan nilai logaritma dari probabilitas prior dan probabilitas kondisional setiap fitur, seperti pada Persamaan 1.

$$Skor(C) = \log(P(C)) + \sum \log(P(Fitur|C)) \quad (1)$$

2. K-Fold Cross Validation

Ini adalah metode validasi yang digunakan untuk mengevaluasi performa model secara andal. Metode ini membagi dataset menjadi K bagian (lipatan) yang sama besar. Proses evaluasi diulang sebanyak K kali, di mana setiap kali satu bagian digunakan sebagai data uji dan K-1 bagian lainnya digunakan sebagai data latih. Hasil akurasi dari setiap iterasi kemudian dirata-ratakan untuk mendapatkan estimasi performa model yang lebih stabil dan tidak bias. Metode ini membantu memastikan bahwa model tidak overfit pada pembagian data tertentu.

Hasil dan Pembahasan

A. Penerapan Multinomial Naïve Bayes

Untuk mengilustrasikan cara kerja algoritma, dilakukan perhitungan manual pada sebuah kasus baru. Misalkan data baru yang akan diprediksi adalah sebagai berikut: Umur: Lansia Madya, Jenis Kelamin: Laki-laki, Pola Hidup: Tidak sehat, Kebiasaan Merokok: Sering, Aktivitas Fisik: Jarang, Tingkat Stres: Sedang, Pola Makan: Tidak Sehat, Kualitas Tidur: Buruk

1. Perhitungan Probabilitas Prior (P(C))

Berdasarkan data latih, dihitung probabilitas awal untuk setiap kelas (Hipertensi dan Tidak Hipertensi). Jika terdapat 5 data di kelas Hipertensi dan 5 data di kelas Tidak Hipertensi dari total 10 data latih, maka probabilitas priornya adalah:

$$P(\text{Hipertensi}) = (5+1)/(10+2) = 0.5$$

$$P(\text{Tidak_Hipertensi}) = (5+1)/(10+2) = 0.5$$

(Penghitungan menggunakan Laplace Smoothing dengan menambahkan 1 pada pembilang dan jumlah kelas pada penyebut).

2. Perhitungan Probabilitas Kondisional ($P(\text{Fitur}|\text{C})$)

Selanjutnya, dihitung probabilitas kemunculan setiap nilai fitur pada setiap kelas. Contohnya, probabilitas fitur 'Umur = Lansia Madya' pada kelas 'Hipertensi':

$$P(\text{Umur=Lansia Madya} | \text{Hipertensi}) = (\text{Jumlah 'Lansia Madya' di kelas Hipertensi} + 1) / (\text{Total data di kelas Hipertensi} + |V|)$$

Di mana $|V|$ adalah total nilai unik dari semua fitur. Proses ini diulangi untuk semua fitur pada data baru untuk kedua kelas.

3. Kalkulasi Skor Akhir

Skor akhir dihitung dengan menjumlahkan logaritma dari probabilitas prior dan semua probabilitas kondisional yang relevan. Berdasarkan contoh perhitungan dalam skripsi, diperoleh hasil:

$$\text{Skor}(\text{Hipertensi}) = -17.677$$

$$\text{Skor}(\text{Tidak_Hipertensi}) = -24.031$$

1. Keputusan Klasifikasi

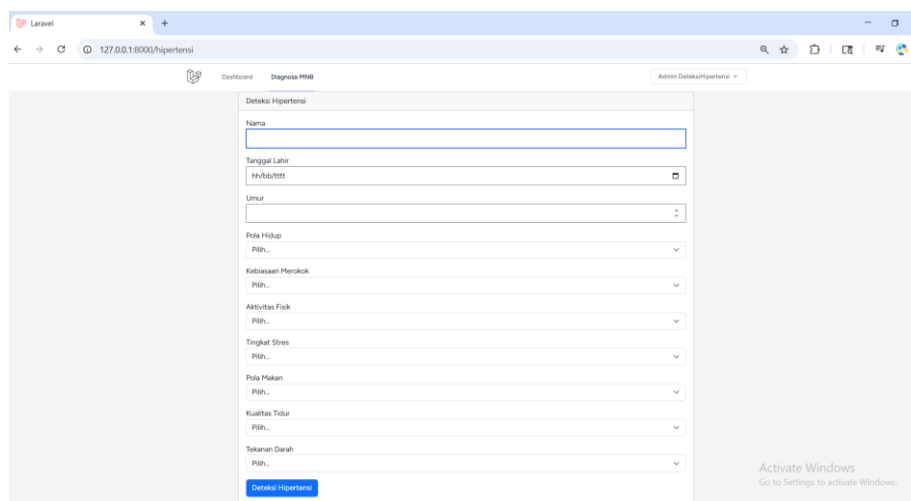
Kelas dengan skor tertinggi (paling mendekati nol) dipilih sebagai hasil prediksi. Karena $-17.677 > -24.031$, model memprediksi bahwa pasien tersebut masuk ke dalam kelas Hipertensi.

2. Tampilan Sistem

Aplikasi yang dibangun memiliki antarmuka berbasis web yang dirancang agar mudah digunakan oleh tenaga kesehatan. Halaman utama dalam fungsionalitas sistem adalah halaman diagnosis dan halaman hasil diagnosis.

3. Halaman Diagnosis

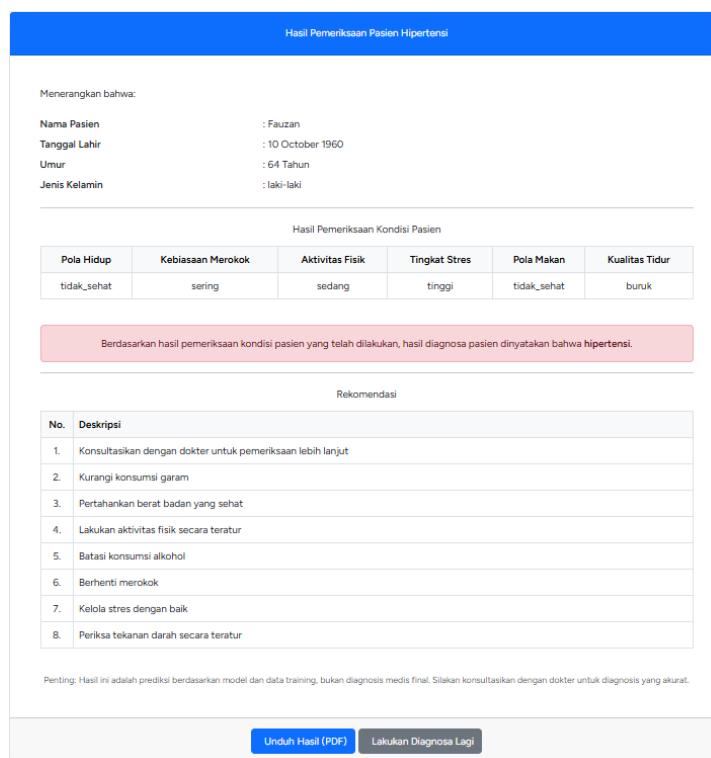
Pada halaman ini, pengguna dapat memasukkan data kesehatan pasien, termasuk nama, tanggal lahir, umur, serta menjawab serangkaian pertanyaan mengenai variabel pola hidup melalui pilihan dropdown. Setelah semua data terisi, pengguna menekan tombol "Deteksi Hipertensi" untuk memulai proses klasifikasi. Tampilan halaman diagnosis dapat dilihat pada Gambar 1.



Gambar 1. Halaman Diagnosis

4. Halaman Hasil Diagnosis

Setelah proses klasifikasi selesai, sistem akan menampilkan halaman hasil. Halaman ini menyajikan kembali data pasien yang diinput, hasil prediksi (misalnya, "dinyatakan bahwa hipertensi"), serta memberikan daftar rekomendasi atau saran kesehatan yang relevan. Terdapat juga fitur untuk mengunduh hasil dalam format PDF untuk keperluan arsip rekam medis. Tampilan halaman Hasil diagnosis dapat dilihat pada Gambar 2.



Hasil Pemeriksaan Pasien Hipertensi

Menerangkan bahwa:

Nama Pasien : Fauzan
 Tanggal Lahir : 10 October 1960
 Umur : 64 Tahun
 Jenis Kelamin : laki-laki

Hasil Pemeriksaan Kondisi Pasien

Pola Hidup	Kebiasaan Merokok	Aktivitas Fisik	Tingkat Stres	Pola Makan	Kualitas Tidur
tidak_sehat	sering	sedang	tinggi	tidak_sehat	buruk

Berdasarkan hasil pemeriksaan kondisi pasien yang telah dilakukan, hasil diagnosa pasien dinyatakan bahwa hipertensi.

Rekomendasi

No.	Deskripsi
1.	Konsultasikan dengan dokter untuk pemeriksaan lebih lanjut
2.	Kurangi konsumsi garam
3.	Pertahankan berat badan yang sehat
4.	Lakukan aktivitas fisik secara teratur
5.	Batasi konsumsi alkohol
6.	Berhenti merokok
7.	Kelola stres dengan baik
8.	Periksa tekanan darah secara teratur

Penting: Hasil ini adalah prediksi berdasarkan model dan data training, bukan diagnosis medis final. Silakan konsultasikan dengan dokter untuk diagnosis yang akurat.

[Unduh Hasil \(PDF\)](#) [Lakukan Diagnosa Lagi](#)

Gambar 2. Halaman Hasil diagnosis

B. Evaluasi Sistem

Untuk memastikan bahwa sistem yang dikembangkan tidak hanya akurat secara algoritmis tetapi juga fungsional dan andal dari sisi pengguna, dilakukan dua tahap pengujian utama. Tahap pertama adalah validasi performa model klasifikasi untuk mengukur akurasi prediksinya. Tahap kedua adalah pengujian fungsionalitas sistem secara keseluruhan untuk memastikan semua fitur berjalan sesuai harapan.

C. Validasi Model Klasifikasi

Validasi model bertujuan untuk menguji kemampuan algoritma Multinomial Naïve Bayes dalam mengklasifikasikan risiko hipertensi secara akurat berdasarkan data pola hidup yang diberikan. Pengujian ini dilakukan menggunakan perangkat lunak RapidMiner dengan metrik evaluasi standar industri: akurasi, presisi, dan recall. Hasil pengujian disajikan dalam bentuk confusion matrix seperti yang terlihat pada Tabel 2.

Tabel 1 Hasil Uji *RapidMiner*

	true hipertensi	true tidak hipertensi
pred. hipertensi	122	2
pred. tidak hipertensi	1	121

Berdasarkan confusion matrix di atas, kinerja model dihitung sebagai berikut:

Akurasi: Kemampuan model untuk memprediksi dengan benar secara keseluruhan.

Akurasi = $(122+121)/(122+121+2+1)=98.78\%$ [cite: 1178, 1179, 1180]

Presisi: Dari semua prediksi hipertensi yang dibuat, berapa persen yang benar-benar hipertensi.

Presisi = $122/(122+2)=98.39\%$ [cite: 1181, 1182, 1183]

Recall: Dari semua kasus hipertensi yang sebenarnya, berapa persen yang berhasil diprediksi oleh model.

Recall = $122/(122+1)=99.19\%$ [cite: 1184, 1185, 1186]

Hasil evaluasi menunjukkan tingkat akurasi sebesar 98.78%, yang tergolong sangat tinggi. Ini menandakan bahwa model klasifikasi yang dibangun memiliki kemampuan yang sangat andal dan efektif dalam membedakan antara lansia yang berisiko hipertensi dan yang tidak, berdasarkan variabel pola hidup yang dianalisis.

Kesimpulan

Berdasarkan hasil penelitian, perancangan, dan pengujian aplikasi pendeteksian dini hipertensi pada lansia menggunakan algoritma Multinomial Naïve Bayes, dapat ditarik beberapa kesimpulan sebagai berikut: Penerapan Algoritma Berhasil: Algoritma Multinomial Naïve Bayes telah berhasil diimplementasikan dalam sebuah sistem aplikasi berbasis web untuk mengklasifikasikan risiko hipertensi pada lansia berdasarkan data pola hidup. Model yang dibangun mampu memproses data kategorikal seperti kebiasaan merokok, aktivitas fisik, dan pola makan untuk menghasilkan prediksi.

Akurasi Model: Hasil pengujian model menggunakan metode K-Fold Cross Validation menunjukkan bahwa algoritma Multinomial Naïve Bayes mampu mencapai tingkat akurasi yang baik. Pada pengujian dengan K-10 Fold, model mencapai akurasi sebesar 98.78%, yang menandakan bahwa model ini cukup andal sebagai alat bantu deteksi dini.

Daftar Pustaka

- [1] A. Nugroho, “Persepsi Anak Muda Terhadap Keberadaan Lansia Di Indonesia,” J. Urban Sociol., vol. 2, no. 2, p. 44, 2020, doi: 10.30742/jus.v2i2.996.
- [2] W. Warjiman, Y. Warni, and A. Rachman, “Gaya Hidup Penderita Hipertensi Di Posyandu Lansia Desa Batu Makap Di Wilayah Kerja Upt Puskesmas Tumbang Kunyi,” J. Keperawatan Suaka Insa., vol. 9, no. 1, pp. 30-34, 2024, doi: 10.51143/jksi.v9i1.588.
- [3] W. Widyawati and S. Sutanto, “Perbandingan Kinerja Variasi Naïve Bayes Multivariate Bernoulli Dan Naïve Bayes Multinomial Dalam Pengklasifikasian Dokumen Teks,” J. Innov. Futur. Technol., vol. 2, no. 1, pp. 108-125, 2020, doi: 10.47080/ifttech.v2i1.859.
- [4] Z. Zulkipli, K. Kusri, and S. Sudarmawan, “Prediksi Tingkat Kesehatan Lingkungan Masyarakat Dalam Program Sustainable Development Goals Menggunakan Algoritma Naive Bayes,” Infotek J. Inform. dan Teknol., vol. 6, no. 2, pp. 431-442, 2023, doi: 10.29408/jit.v6i2.18776.
- [5] D. Rizki Surya Pratama, T. A. Munandar, and K. Fadhilla Ramdhania, “Multinomial Naive Bayes Algorithm for Indonesian language Sentiment Classification Related to Jakarta International Stadium (JIS),” Int. J. Inf. Technol. Comput. Sci. Appl., vol. 2, no. 1, pp. 12-22, 2024, doi: 10.58776/ijitcsa.v2i1.118.
- [6] N. Hidayah, “8 | Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan ... Implementasi Algoritma Multinomial Naïve Bayes, TF-IDF dan Confusion Matrix dalam Pengklasifikasian Saran Monitoring dan Evaluas,” J. Akad. Pendidik. Mat., vol. 10, no. 1, pp. 8-15, 2024, [Online]. Available: <https://doi.org/10.55340/japm.v10i1.1491>