Journal Of Global Computer Science

Volume 1 ; Number 2 ; August 2025 ; Page 70-74
Doi : https://doi.org/10.59435/jgcs.v1i2.2025.39
Web : https://journal.padangtekno.com/index.php/jgcs
E-ISSN : 3032-372X

Penggunaan Metode Naive Bayes untuk Prediksi Penyakit Diabetes

Linny Fadhilah [1*], Zaehol Fatah[2], Irma Yunita[3]

[1] Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy
[2] Sistem Informasi, Sains dan Teknologi, Universitas Ibrahimy
[3] Teknologi Informasi, Sains dan Teknologi, Universitas Ibrahimy
[1]lenifadhilahns@gmail.com [2] zaeholfatah@gmail.com [3]irmayunitasaid@gmail.com

ABSTRAK

Prediksi penyakit diabetes menjadi salah satu tantangan penting dalam dunia kesehatan mengingat meningkatnya prevalensi penyakit ini secara global. Penelitian ini bertujuan untuk mengimplementasikan metode Naive Bayes dalam memprediksi risiko diabetes berdasarkan dataset medis yang berisi berbagai fitur seperti usia, berat badan, kadar gula darah, tekanan darah, dan riwayat keluarga. Metode ini dipilih karena kemampuannya dalam mengolah data dengan asumsi independensi antar variabel. Pada penelitian menggunakan metode Naive Bayes menunjukkan bahwa memiliki capaian tingkat akurasi sebesar 83.75% dengan precision untuk kelas "Yes" sebesar 86.47% dan recall sebesar 93.50%, yang menandakan performa yang cukup baik dalam mendeteksi pasien diabetes. Namun, performa model dalam mendeteksi pasien non-diabetes perlu ditingkatkan mengingat recall untuk kelas "No" hanya mencapai 51.35%. Hasil ini menunjukkan bahwa Naive Bayes adalah metode yang sederhana namun efektif untuk memprediksi diabetes, dan dapat dikembangkan lebih lanjut untuk membantu pengambilan keputusan medis secara lebih akurat.

Kata kunci: Naive Bayes, prediksi diabetes, metode klasifikasi, akurasi, dataset medis.

PENDAHULUAN

Diabetes merupakan kondisi medis yang ditandai oleh peningkatan kadar glukosa darah secara kronis. Penyakit ini terjadi ketika tubuh tidak mampu memproduksi atau menggunakan insulin dengan efektif, sehingga menyebabkan gangguan pada pengaturan kadar gula darah. Meskipun istilah "diabetes" sering kali digunakan untuk merujuk pada diabetes mellitus, yang merupakan jenis diabetes paling umum, istilah ini secara lebih luas mencakup berbagai jenis gangguan metabolik yang melibatkan ketidakmampuan tubuh dalam mengatur gula darah. Di Indonesia, prevalensi diabetes terus meningkat seiring dengan perubahan gaya hidup yang lebih sedentari dan pola makan yang tidak sehat.[1]

Penyakit diabetes terbagi menjadi beberapa tipe, seperti diabetes tipe 1, tipe 2, dan diabetes gestasional. Diabetes tipe 1 biasanya terjadi pada usia muda dan disebabkan oleh kerusakan pada sel penghasil insulin di pankreas, sementara diabetes tipe 2 lebih sering ditemukan pada orang dewasa dan sering kali dikaitkan dengan obesitas serta gaya hidup yang tidak sehat. Selain itu, diabetes gestasional adalah kondisi sementara yang terjadi selama kehamilan, tetapi dapat meningkatkan risiko terkena diabetes tipe 2 di kemudian hari.[2]

Prediksi dini dan deteksi diabetes sangat penting untuk mencegah komplikasi jangka panjang, seperti penyakit jantung, kerusakan ginjal, kebutaan, dan gangguan saraf. Dalam beberapa tahun terakhir, perkembangan teknologi dalam bidang data science dan machine learning telah memberikan peluang baru

dalam mendiagnosis dan memprediksi penyakit diabetes. Salah satu metode yang banyak digunakan untuk analisis data medis adalah Naive Bayes. Metode ini memiliki keunggulan dalam menangani data dengan banyak fitur dan variabel kategorikal, serta kemampuannya untuk memberikan hasil yang cepat dan akurat dengan sedikit data pelatihan.[3]

Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang didasarkan pada teorema Bayes, yang mengasumsikan bahwa fitur-fitur yang ada bersifat independen satu sama lain. Meskipun asumsi independensi ini sering kali tidak sepenuhnya realistis dalam dunia medis, Naive Bayes tetap terbukti efektif dalam banyak kasus klasifikasi, termasuk prediksi penyakit diabetes. Oleh karena itu, pendekatan ini dapat digunakan untuk memodelkan faktor-faktor risiko yang berkontribusi pada perkembangan diabetes, seperti usia, berat badan, kadar gula darah, dan riwayat keluarga.[4]

Tingkat akurasi yang dihasilkan dengan nilai accuracy sebesar 84.00%, precisionsebesar 75.00%, recallsebesar 75.00%, AUC Optimisticsebesar 0.909%, AUC sebesar 0.907% dan AUC Pessimisticsebesar 0.905%, maka dapat disimpulkanbahwa algoritma ini cocok digunakan untuk menghitung kelayakan penerima bantuan pemerintah yang dimaksud.[5]

METODE

Dataset

Dataset yang digunakan merupakan data yang bersifat simulasi (bukan data medis asli), yang dirancang untuk tujuan eksperimen dan tidak berasal dari data medis nyata. Dataset diabetes yang digunakan adalah data yang diperoleh dari database Kesehatan diabetes yang bersifat public.

Dataset ini berisi 200 record data dengan beberapa variabel atau atribut prediktor yang relevan untuk mendeteksi risiko diabetes. Atribut yang tersedia meliputi Usia, Berat Badan, Kadar Gula Darah, Riwayat Keluarga (Ya/Tidak), Tekanan Darah (sistolik), Aktivitas Fisik (Rendah, Sedang, Tinggi), dan Indeks Massa Tubuh (BMI). Variabel target yang digunakan adalah label Diabetes dengan dua kemungkinan nilai, yaitu Ya (positif diabetes) atau Tidak (negatif diabetes). Dataset tersebut ditampilkan dalam table berikut:

Table 1. Dataset Training

Usia	Berat_Badan (kg)	Glicose (mg/dL)	Riwayat_keluarga	Tekanan_Darah (mmHg)	Aktivitas_Fisik	BMI	Label (yes/no)
48	94.9	152.2	No	108.8	Medium	27.5	Yes
49	61	97.6	Yes	117.6	Low	22.1	Yes
66	101	155.4	No	150.8	Low	36.3	Yes
57	52.6	111.9	No	114.7	Low	13.9	No
54	75.9	70	Yes	139.1	Higt	20.3	Yes
63	78.6	101	No	140.2	Low	25	No
41	113.6	171.2	No	132.4	Low	36.5	Yes
72	117.4	101.4	Yes	110.2	Low	34	Yes
41	111.4	132.1	Yes	126.6	Medium	37.3	Yes
33	96.4	166.2	Yes	125	Low	31.6	Yes

Tabel 2. Dataset Testing

Usia	Berat_Badan (kg)	Glicose (mg/dL)	Riwayat_Keluarga	Tekanan_Darah (mmHg)	Aktivitas_Fisik	BMI
48	94.9	152.2	No	108.8	Medium	27.5
49	61	97.6	Yes	117.6	Low	22.1
66	101	155.4	No	150.8	Low	36.3
57	52.6	111.9	No	114.7	Low	13.9
54	75.9	70	Yes	139.1	Higt	20.3
63	78.6	101	No	140.2	Low	25
41	113.6	171.2	No	132.4	Low	36.5
72	117.4	101.4	Yes	110.2	Low	34
41	111.4	132.1	Yes	126.6	Medium	37.3
33	96.4	166.2	Yes	125	Low	31.6

Naïve Bayes

Metode Naive Bayes adalah algoritma klasifikasi berbasis probabilitas yang didasarkan pada teorema Bayes.[6] Algoritma ini sering digunakan untuk memprediksi kategori target berdasarkan hubungan antara fitur-fitur dalam dataset[7]. Nama "Naive" mencerminkan asumsi bahwa setiap fitur bersifat independen satu sama lain dalam mempengaruhi probabilitas target, meskipun pada kenyataannya, hubungan antar fitur mungkin ada.

Teorema Bayes dijelaskan dengan formula berikut:

$$P(H|E) = \frac{P(E|H).P(H)}{P(E)}$$

Dimana:
P(H|E) : Probabilitas hipotesis H berdasarkan bukti E
P(E|H) : Probabilitas bukti E diberikan hipotesis H
P(H) : Probabilitas prior dari hipotesis H
P(E) : Probabilitas bukti E secara keseluruhan

Metode Naive Bayes memiliki keunggulan dalam menangani dataset dengan jumlah data yang relatif kecil, serta mampu bekerja dengan baik untuk data dengan kombinasi fitur kategorikal dan numerik. Dalam konteks prediksi penyakit diabetes, algoritma ini dapat memanfaatkan berbagai variabel medis seperti usia, kadar gula darah, indeks massa tubuh, dan riwayat keluarga untuk menentukan apakah seseorang memiliki risiko menderita diabetes.

Pada proses klasifikasi, Naive Bayes menghitung probabilitas posterior untuk setiap kelas target berdasarkan fitur-fitur yang ada, kemudian memilih kelas dengan probabilitas tertinggi sebagai hasil prediksi. Probabilitas ini dihitung dengan memanfaatkan distribusi data pada masing-masing fitur, seperti distribusi Gaussian untuk data numerik.

HASIL DAN PEMBAHASAN

Proses Awal

Proses awal dalam penelitian ini melibatkan pengolahan dataset yang telah disiapkan dengan atribut-atribut relevan, seperti usia, berat badan, kadar gula darah, riwayat keluarga, tekanan darah, tingkat aktivitas fisik, indeks massa tubuh (BMI), dan label target (diabetes: Ya/Tidak). Dataset tersebut diproses menggunakan metode data preprocessing untuk memastikan data bersih, bebas dari nilai kosong, dan dalam format yang sesuai untuk algoritma Naive Bayes.

Penentuan Jenis Data

Dataset dibagi menjadi dua subset: data pelatihan (training set) sebesar 80% dari total data dan data pengujian (testing set) sebesar 20%. Data pelatihan digunakan untuk melatih model Naive Bayes, sementara data pengujian digunakan untuk mengevaluasi performa model. Penentuan jenis data yang relevan untuk fitur numerik dan kategorikal dilakukan untuk memastikan algoritma bekerja optimal dalam mempelajari pola dari data tersebut.[8]

Model Data Mining

Model Naive Bayes diterapkan pada dataset menggunakan asumsi distribusi Gaussian untuk fitur numerik, seperti kadar gula darah dan BMI. Probabilitas posterior dihitung untuk setiap kelas target, dan hasil prediksi ditentukan berdasarkan probabilitas tertinggi. Model ini mampu mengklasifikasikan data dengan baik meskipun ada variasi dalam atribut kategorikal, seperti riwayat keluarga dan tingkat aktivitas fisik.[9]

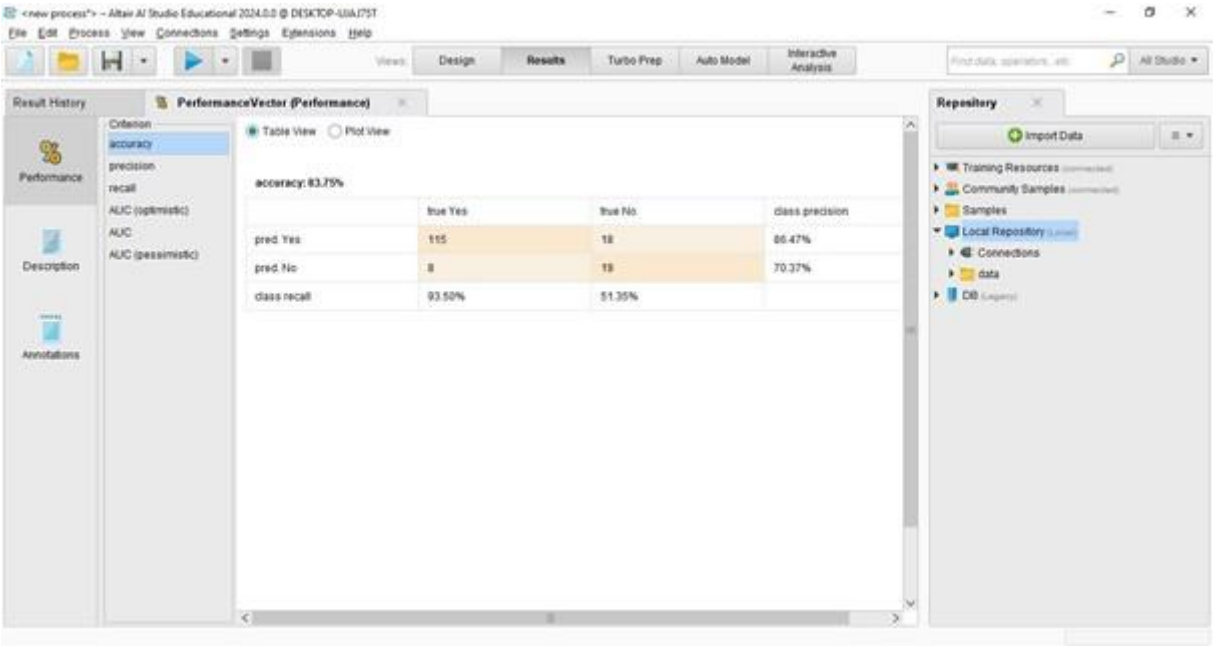
Hasil Accuracy Perhitungan Naïve Bayes

Hasil evaluasi model Naive Bayes menunjukkan bahwa model memiliki tingkat akurasi sebesar 83.75%, yang berarti model mampu memprediksi secara benar 83.75% dari total data pengujian. Tingginya akurasi ini menunjukkan bahwa metode Naive Bayes cukup efektif dalam memprediksi penyakit diabetes.

Selain itu, precision untuk prediksi kelas "Yes" sebesar 86.47% mengindikasikan bahwa sebagian besar prediksi positif (Yes) yang dihasilkan oleh model adalah benar. Namun, precision untuk kelas "No" sedikit lebih rendah, yaitu 70.37%, yang menunjukkan adanya beberapa kasus di mana model salah memprediksi kelas "No" menjadi "Yes".

Recall untuk kelas "Yes" mencapai 93.50%, yang berarti model dapat menangkap hampir semua kasus aktual diabetes dalam dataset. Sebaliknya, recall untuk kelas "No" relatif lebih rendah, yaitu 51.35%, mengindikasikan bahwa model kurang mampu mendeteksi kasus aktual "No" atau pasien yang tidak memiliki diabetes. Hal ini menunjukkan bahwa model cenderung lebih baik dalam mendeteksi pasien dengan diabetes dibandingkan mendeteksi pasien tanpa diabetes.

Secara keseluruhan, hasil evaluasi ini menunjukkan performa yang cukup baik, tetapi terdapat kekurangan pada kemampuan model dalam mendeteksi kelas "No". Untuk meningkatkan performa, disarankan melakukan analisis lebih lanjut, seperti penyeimbangan dataset atau penyesuaian threshold prediksi, guna meminimalkan kesalahan pada kelas tertentu.



Gambar 1. Hasil accuracy

KESIMPULAN

Metode Naive Bayes telah berhasil diterapkan untuk prediksi penyakit diabetes dengan hasil evaluasi yang cukup baik. Model yang dibangun mampu mencapai tingkat akurasi sebesar 83.75%, yang menunjukkan bahwa metode ini efektif dalam memprediksi risiko diabetes. Precision untuk kelas "Yes" sebesar 86.47% menunjukkan kemampuan model untuk memprediksi pasien diabetes dengan tingkat kesalahan rendah. Sementara itu, recall untuk kelas "Yes" mencapai 93.50%, yang berarti model dapat menangkap sebagian besar kasus diabetes aktual dalam dataset.

Namun, terdapat kekurangan pada kemampuan model dalam mendeteksi kelas "No," dengan recall sebesar 51.35%, yang menunjukkan bahwa model lebih cenderung memprioritaskan identifikasi pasien dengan diabetes daripada yang tanpa diabetes. Hal ini dapat menjadi area yang memerlukan perbaikan, seperti penyeimbangan data atau optimasi parameter model.

Secara keseluruhan, Naive *Bayes* terbukti menjadi metode yang sederhana namun efektif untuk memodelkan faktor risiko diabetes, serta dapat digunakan sebagai dasar untuk pengembangan sistem prediksi penyakit yang lebih canggih di masa depan.

DAFTAR PUSTAKA

[1] C. M. Porth, *Essentials of Pathophysiology: Concepts of Altered Health States*. Wolters Kluwer, 2020.

[2] B. Khan, F., & Chaurasia, *Diabetes: Understanding and Management*. 2021.

[3] S. Cohen, J., & Burk, *Medical Informatics: An Executive Primer (3rd ed.)*. 2022.

[4] and M. Ç.-R. Alicia Johnson, Miles Ott, *Bayes Rules! An Introduction to Applied Bayesian Modeling*. 2021.

[5] Z. F. Zainur Rohman, Ahmad Homaidi, "Penerapan Metode Naïve Bayes untuk Menentukan Penerima Kartu Indonesia Pintar (KIP)," *G-Tech : Jurnal Teknologi Terapan*, vol. 8, no. 3, pp. 1806–1815, 2024.

[6] E. Alpaydin, *Introduction to Machine Learning*. MIT Press, 2021.

[7] J. F. Trevor Hastie, Robert Tibshirani, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, 2020.

[8] S. B. Joseph M. Cohen, *Medical Informatics: An Executive Primer*. 2022.

[9] J. P. Jiawei Han, Micheline Kamber, *Data Mining: Concepts and Techniques*. Elsevier, 2022.