

PREDIKSI DIABETES MELLITUS BERDASARKAN DATA PASIEN SYLHET DIABETES HOSPITAL DENGAN METODE NAIVE BAYES

Pricisilla Cahya Pradhani¹, Aida Shafa Indrayani², Nabilah Azzahra³, Syafikqa Emalinda Aflika⁴, Fatimatuz Zahra⁵, Anna Dina Kalifia⁶

Program Studi Informatika, Universitas Teknologi Yogyakarta, Yogyakarta, Indonesia

Email: ¹pradhanicahya@gmail.com, ²aidashafaiasi@gmail.com,

³nabilah.if16@gmail.com, ⁴emalindasyafiqka@gmail.com, ⁵fatimaif2006@gmail.com,

⁶aruliaavula@gmail.com

Abstrak

Diabetes melitus adalah penyakit metabolisme kronis yang ditandai dengan kadar gula darah yang tinggi dalam jangka waktu yang lama. Gula darah ini berasal dari sel darah yang kemudian akan dipindahkan ke sel tubuh oleh hormone insulin. Seseorang yang terkena diabetes melitus tidak dapat menghasilkan hormon insulin yang cukup, bahkan hormon insulin ini tidak dapat melakukan tugasnya secara optimal. Diabetes melitus dapat menyebabkan kerusakan pada jantung, pembuluh darah, mata, ginjal, saraf, dan yang terparah bisa menyebabkan kematian. Tidak dapat dipungkiri bahwa penyakit ini merupakan salah satu penyumbang kematian terbesar didunia. Dari paparan fakta tersebut, penyakit ini harus dideteksi dini untuk meminimalisir kemungkinan terburuknya. Algoritma Naive Bayes adalah salah satu algoritma yang sederhana tetapi cukup efektif untuk metode klasifikasi. Berdasarkan dataset pasien sylhet diabetes hospital di Bangladesh yang didapat dari platform Kaggle , terdapat 16 atribut yang dapat menjadi data independen untuk mendeteksi dini penyakit diabetes. Dengan dataset tersebut sebagai bahan acuan dan algoritma Naive Bayes sebagai landasan metode, dapat dirancang sebuah program python untuk analisis prediksi dini diabetes melitus. Metode Naive Bayes dinilai cukup akurat, hal ini dibuktikan dengan tingginya nilai akurasi pada percobaan program python yang telah dilakukan, yaitu kurang lebih sebesar 91%. Dengan program dan algoritma yang sederhana, Metode Naive Bayes dapat menangani kompleksitas dataset dan memberikan prediksi yang kredibel.

Sejarah Artikel

Submitted: 4 Januari 2025

Accepted: 10 Januari 2025

Published: 11 Januari 2025

Kata Kunci

Diabetes melitus, dataset, prediksi, dan Naive Bayes

PENDAHULUAN

Di era modern ini, fenomena konsumsi gula berlebih (*excessive sugar consumption*) telah menjadi masalah kesehatan masyarakat yang signifikan. Berbagai penelitian menunjukkan bahwa konsumsi gula yang tidak terkontrol dapat memicu berbagai penyakit serius, termasuk obesitas, penyakit jantung, kanker, dan yang paling mengkhawatirkan adalah diabetes melitus (Sinaga et al., 2024). Konsumsi gula berlebih dapat menyebabkan penumpukan lemak ektopik di dalam otot yang mengakibatkan resistensi insulin (*insulin resistance*), kondisi ini merupakan pemicu utama terjadinya Diabetes Melitus Tipe 2. Diabetes melitus (DM) merupakan salah satu masalah kesehatan global yang prevalensinya terus meningkat secara dramatis dalam tiga dekade terakhir di negara-negara dengan berbagai tingkat pendapatan. *World Health Organization* (WHO) melaporkan bahwa sekitar 830 juta orang di seluruh dunia menderita diabetes, dengan mayoritas berada di negara-negara berpenghasilan rendah dan menengah. Yang lebih mengkhawatirkan, lebih dari setengah penderita diabetes tidak menerima pengobatan yang adekuat. Peningkatan jumlah penderita dan kasus diabetes yang tidak tertangani telah menunjukkan tren yang konsisten selama beberapa dekade terakhir.

Diabetes melitus didefinisikan sebagai penyakit metabolik kronis (*chronic metabolic disease*) yang ditandai dengan peningkatan kadar glukosa darah atau hiperglikemia. Kondisi ini terjadi akibat gangguan dalam produksi atau fungsi insulin, yang dapat menyebabkan serangkaian komplikasi serius seperti penyakit kardiovaskular,

gangguan ginjal, dan kerusakan saraf (Widiastuti et al., 2024). Stanhope (2016) menjelaskan bahwa terdapat mekanisme langsung dan tidak langsung dalam perkembangan diabetes melitus tipe 2. Mekanisme langsung melibatkan uptake dan metabolisme fruktosa yang tidak terregulasi di hati, yang mengakibatkan akumulasi lemak hati, dislipidemia, penurunan sensitivitas insulin, dan peningkatan kadar asam urat.

Dalam konteks diagnosis, diabetes melitus ditandai dengan kadar gula darah yang melebihi batas normal, yaitu lebih dari atau sama dengan 200 mg/dl dan kadar gula darah puasa yang bernilai sama dengan atau lebih dari 126 mg/dl. Insufisiensi fungsi insulin dapat disebabkan oleh dua faktor utama: gangguan atau defisiensi produksi insulin oleh sel-sel beta Langerhans kelenjar pankreas (*pancreatic beta cells*), atau kurang responsifnya sel-sel tubuh terhadap insulin yang dikenal sebagai resistensi insulin (*insulin resistance*). (Muthoharun & Yulianto, 2023) menegaskan bahwa diabetes melitus dapat memicu berbagai komplikasi akibat kadar gula darah yang tidak terkontrol, seperti neuropati, hipertensi, penyakit jantung koroner, retinopati, nefropati, dan gangren. Pengendalian kadar gula darah dapat dilakukan melalui tiga pendekatan utama: diet, olahraga, dan pengobatan. Penelitian mereka menunjukkan bahwa program pengelolaan penyakit kronis (PROLANIS) melalui aktivitas fisik secara rutin efektif dalam mengendalikan kadar gula darah penderita diabetes melitus.

(Resti & Cahyati, 2022) melaporkan bahwa prevalensi diabetes melitus di Indonesia mengalami peningkatan sebesar 0,5% dari tahun 2013 hingga 2018. DKI Jakarta tercatat sebagai provinsi dengan kasus diabetes melitus tertinggi. Penelitian mereka mengungkapkan beberapa faktor yang memiliki hubungan signifikan dengan kejadian diabetes melitus, termasuk tingkat pendidikan, tingkat stres, status merokok, status obesitas, dan riwayat keluarga. Hal ini menunjukkan bahwa diabetes melitus adalah penyakit multifaktorial yang memerlukan pendekatan komprehensif dalam pencegahan dan penanganannya. Diabetes melitus dikenal sebagai "silent killer" karena seringkali penderitanya terlambat mengetahui gejala-gejala awal, yang dapat berakibat pada komplikasi serius (Relica & Mariyati, 2024). Gejala-gejala klasik diabetes melitus meliputi *polyuria* (buang air kecil berlebihan), *polydipsia* (rasa haus berlebihan), *sudden weight loss*, penurunan berat badan yang tidak direncanakan, penglihatan kabur, dan luka yang sulit sembuh. Jika tidak ditangani dengan tepat, diabetes dapat menyebabkan komplikasi akut seperti ketoasidosis diabetik (*diabetic ketoacidosis*), hiperglikemia hiperosmolar (*hyperosmolar hyperglycemia*), dan bahkan kematian.

Mengingat kompleksitas dan dampak serius dari diabetes melitus, diperlukan sistem prediksi dini yang efektif untuk mengidentifikasi risiko penyakit ini sedini mungkin. Dalam konteks ini, penelitian ini mengusulkan penggunaan algoritma Gaussian Naive Bayes untuk mengklasifikasikan risiko dini penyakit diabetes melitus. Pemilihan algoritma ini didasarkan pada kesederhanaan dan efektivitasnya dalam mengimplementasikan dataset kecil. Data yang digunakan berasal dari Sylhet Diabetes Hospital di Bangladesh, yang akan digunakan untuk mengembangkan model prediksi yang dapat menghasilkan diagnosis positif atau negatif berdasarkan parameter-parameter tertentu. Pendekatan machine learning untuk prediksi diabetes melitus ini merupakan langkah strategis dalam upaya pencegahan dan penanganan dini penyakit ini. Dengan memanfaatkan teknologi dan data pasien yang ada, diharapkan dapat membantu tenaga medis dalam mengidentifikasi risiko diabetes melitus secara lebih akurat dan efisien. Hal ini sejalan dengan target global untuk menghentikan peningkatan kasus diabetes dan obesitas pada tahun 2025 (WHO, 2024).

Penelitian ini menjadi sangat relevan mengingat bahwa penanganan diabetes melitus memerlukan pendekatan yang komprehensif dan terintegrasi. (Stanhope, 2016) menekankan pentingnya pemahaman mendalam tentang mekanisme patofisiologi diabetes untuk mengembangkan strategi pencegahan dan penanganan yang efektif. Dengan mengintegrasikan machine learning dalam proses diagnosis dini, diharapkan dapat memberikan kontribusi signifikan dalam upaya menurunkan prevalensi dan komplikasi diabetes melitus di masa mendatang. Berdasarkan latar belakang yang telah diuraikan, masalah utama yang perlu diteliti adalah bagaimana mengembangkan model prediksi yang akurat untuk mendeteksi risiko diabetes melitus menggunakan algoritma Gaussian Naive Bayes dengan memanfaatkan data pasien dari Sylhet Diabetes Hospital. Permasalahan spesifik yang akan dijawab meliputi: (1) Bagaimana tingkat akurasi algoritma Gaussian Naive Bayes dalam memprediksi risiko diabetes melitus? (2) Apa saja faktor-faktor yang paling berpengaruh dalam prediksi diabetes melitus menggunakan model ini? (3) Bagaimana perbandingan performa model ini dengan metode klasifikasi lainnya dalam konteks prediksi diabetes melitus?

Penelitian ini bertujuan untuk mengembangkan sistem prediksi dini diabetes melitus menggunakan algoritma Gaussian Naive Bayes berbasis data pasien Sylhet Diabetes Hospital. Secara khusus, penelitian ini bertujuan untuk: (1) Mengimplementasikan algoritma Gaussian Naive Bayes untuk klasifikasi risiko diabetes melitus, (2) Menganalisis tingkat akurasi, presisi, dan skor-F1 dari model yang dikembangkan, (3) Mengidentifikasi parameter-parameter kunci yang mempengaruhi prediksi diabetes melitus, dan (4) Memberikan rekomendasi untuk pengembangan sistem prediksi diabetes melitus yang lebih efektif.

Hasil penelitian ini diharapkan dapat memberikan kontribusi signifikan baik secara teoretis maupun praktis dalam bidang kesehatan dan teknologi. Secara teoretis, penelitian ini memperkaya literatur tentang penerapan machine learning dalam diagnosis medis, khususnya untuk penyakit diabetes melitus. Secara praktis, model prediksi yang dikembangkan dapat membantu tenaga medis dalam melakukan screening awal dan diagnosis dini diabetes melitus, sehingga dapat meningkatkan efektivitas penanganan dan mencegah komplikasi lebih lanjut. Selain itu, penelitian ini juga dapat menjadi dasar untuk pengembangan sistem prediksi penyakit kronis lainnya menggunakan pendekatan machine learning.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi menggunakan algoritma Gaussian Naive Bayes. Pemilihan algoritma ini didasarkan pada kemampuannya dalam menangani dataset kecil hingga menengah dengan efisiensi komputasi yang tinggi. Dataset yang digunakan berasal dari Sylhet Diabetes Hospital di Bangladesh, yang mencakup berbagai parameter klinis seperti *polyuria* (buang air kecil berlebihan), *polydipsia* (rasa haus berlebihan), *sudden weight loss*, dan faktor risiko lainnya. Dataset ini dipilih karena memiliki kelengkapan data dan representasi yang baik untuk kasus diabetes melitus. Proses pengolahan data dilakukan melalui beberapa tahap. Pertama, data mentah melalui tahap preprocessing yang meliputi penanganan missing values, normalisasi data, dan feature selection. Tahap kedua adalah pembagian dataset menjadi data training. Data training digunakan untuk melatih model Gaussian Naive Bayes. Implementasi algoritma menggunakan Python dengan memanfaatkan library scikit-learn untuk proses klasifikasi dan pandas untuk manipulasi data. Algoritma

Gaussian Naive Bayes yang digunakan dalam penelitian ini didasarkan pada teorema probabilitas Bayes yang diformulasikan sebagai berikut:

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

Rumus teorema Naïve bayes :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

P(H | X) = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob)

P(H) = Probabilitas hipotesis (prior prob)

P(X | H) = Probabilitas X berdasarkan kondisi tersebut

P(X) = Probabilitas dari X

Evaluasi performa model dilakukan menggunakan beberapa metrik standar dalam machine learning, meliputi accuracy, precision, recall, dan F1-score. Visualisasi hasil analisis menggunakan matplotlib dan seaborn untuk menghasilkan confusion matrix dan ROC curve. Selain itu, dilakukan juga analisis feature importance untuk mengidentifikasi parameter-parameter yang memiliki kontribusi signifikan dalam prediksi diabetes melitus. Hasil evaluasi ini kemudian dianalisis secara statistik untuk menilai reliabilitas dan validitas model yang dikembangkan. Untuk memastikan robustness model, dilakukan k-fold cross-validation dengan k=10, yang membagi dataset menjadi 10 bagian secara acak dan melakukan training-testing secara berulang. Pendekatan ini membantu mengurangi overfitting dan memberikan estimasi yang lebih akurat tentang performa model pada data baru. Seluruh proses komputasi dilakukan menggunakan Python 3.8 dengan dukungan pustaka NumPy untuk operasi numerik dan Pandas untuk manipulasi dataset. Hasil prediksi model diklasifikasikan menjadi dua kategori: positif diabetes dan negatif diabetes, berdasarkan threshold probabilitas yang ditentukan melalui analisis ROC curve. Untuk meningkatkan interpretabilitas hasil, dikembangkan juga visualisasi interaktif menggunakan pustaka Plotly yang memungkinkan eksplorasi lebih mendalam terhadap pola-pola dalam data dan hasil prediksi model.

HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Dataset yang digunakan dalam pembuatan program ini merupakan data survey pasien-pasien di Sylhet Diabetes Hospital di Bangladesh. Pasien-pasien tersebut memiliki banyak gejala-gejala yang berbeda, yang merujuk pada penyakit diabetes melitus berdasarkan hasil diagnosis dokter di rumah sakit tersebut. Dataset ini diambil dari salah satu platform dataset repository Kaggle yang nantinya dijadikan sebagai sample hingga 520 data yang dijadikan acuan data dependen dan independent. Berikut ini adalah penggalan gambar dataset yang akan digunakan :

age	gender	polyuria	polydipsia	sudden_weight_loss	weakness	polyphagia	genital_thrush	visual_blurring	itching	irritability	delayed_healing	partial_paresis	muscle_stiffness	alopecia	obesity	(less)
40	Male	0	1	0	1	0	0	0	1	0	1	0	1	1	1	1
53	Male	0	0	0	1	0	0	1	0	0	0	1	0	1	0	1
41	Male	1	0	0	1	1	0	0	1	0	1	0	1	1	1	1
43	Male	0	0	1	1	1	1	0	1	0	1	0	0	0	0	1
60	Male	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
55	Male	1	1	0	1	1	0	1	0	1	0	1	0	1	1	1
37	Male	1	1	0	1	1	1	0	0	1	1	0	0	0	0	1
66	Male	1	1	1	1	0	0	1	1	1	0	1	1	0	0	1
67	Male	1	1	0	1	1	1	0	1	1	1	0	1	1	1	1
70	Male	0	1	1	1	1	0	1	1	1	0	0	0	1	0	1
44	Male	1	1	0	1	0	1	0	0	1	1	0	1	1	0	1
38	Male	1	1	0	0	1	1	0	1	0	1	0	1	0	0	1
55	Male	1	0	0	0	1	1	0	0	1	1	0	0	0	1	1
42	Male	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1
60	Male	1	1	0	1	1	0	1	1	0	1	1	0	0	1	1
59	Male	1	1	0	1	1	0	0	0	0	1	1	1	0	0	1
54	Male	1	1	1	1	0	1	0	0	0	1	0	1	0	0	1
47	Male	0	1	0	1	1	0	1	0	1	1	1	1	1	1	1
66	Male	1	1	0	1	1	0	1	0	0	0	1	1	0	0	1
41	Male	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1

Langkah berikutnya yaitu melakukan pemilihan kriteria data. Pemilihan data adalah Langkah penentuan data yang diperlukan sebagai data training. Data training ini memuat kriteria gejala diabetes yang diperlukan. Adapun data kriteria yang diambil dapat dilihat mulai dari tabel 1 berikut:

Tabel 1 Data Kriteria

Kriteria	Keterangan
Age	$\leq 25 = 1, 26-45 = 2, >45 = 3$
Gender	Male = 1, Female = 2
Polyuria	Iya = 1, Tidak = 0
Polydipsia	Iya = 1, Tidak = 0
Sudden_Weight_Loss	Iya = 1, Tidak = 0
Weakness	Iya = 1, Tidak = 0
Polyphagia	Iya = 1, Tidak = 0
Genital_thrush	Iya = 1, Tidak = 0
Visual_blurring	Iya = 1, Tidak = 0
Itching	Iya = 1, Tidak = 0
Irritability	Iya = 1, Tidak = 0
Delayed_healing	Iya = 1, Tidak = 0
Partial_paresis	Iya = 1, Tidak = 0
Muscle_stiffness	Iya = 1, Tidak = 0
Alopecia	Iya = 1, Tidak = 0
Obesity	Iya = 1, Tidak = 0

Setelah tahap pemilihan data selesai dilakukan, langkah berikutnya adalah melaksanakan proses pembersihan data sebagai bagian dari pra-pemrosesan. Pembersihan data mencakup berbagai kegiatan yang bertujuan untuk mengenali serta memperbaiki ketidaksesuaian atau kekurangan dalam data yang akan dianalisis. Proses ini sangat penting karena data mentah biasanya tidak siap langsung digunakan untuk mining, sehingga perlu dilakukan penyesuaian dan perbaikan agar analisis yang dihasilkan menjadi lebih tepat.

B. Perhitungan Klasifikasi Naïve Bayes

Setelah seluruh data yang dibutuhkan berhasil diperoleh, tahap selanjutnya adalah melakukan perhitungan menggunakan Algoritma Naïve Bayes. Perhitungan ini dilakukan berdasarkan data yang ditampilkan pada Gambar dataset, yang terdiri dari 520 data, dengan kriteria yang telah diuraikan sebelumnya pada Tabel 1.

1. Langkah pertama : Data training dengan jumlah 520 dan 16 faktor kriteria.
2. Langkah kedua : Menghitung probabilitas untuk setiap kelas, yaitu Positive dan Negative.

Probabilitas masing-masing kelas beserta kriterianya dapat dilihat pada Tabel 2 sampai Tabel 18 berikut ini.

Tabel 2 Probabilitas Umur

Age		Positive	Negative
<=25	Anak-anak dan Remaja	3/320	0/200
26-45	Dewasa	125/320	103/200
>45	Lansia	192/320	97/200

Tabel 3 Probabilitas Gender

Gender	Positive	Negative
Famale	173/320	19/200
Male	147/320	181/200

Tabel 4 Probabilitas Polyuria

Polyuria	Positive	Negative
Iya	243/320	15/200
Tidak	77/320	185/200

Tabel 5 Probabilitas Polydipsia

Polydipsia	Positive	Negative
Iya	225/320	8/200
Tidak	95/320	192/200

Tabel 6 P Probabilitas Sudden Weight Loss

Sudden_Weight_Loss	Positive	Negative
Iya	188/320	29/200
Tidak	132/320	171/200

Tabel 7 Probabilitas Weakness

Weakness	Positive	Negative
Iya	218/320	87/200
Tidak	102/320	113/200

Tabel 8 Probabilitas Polyphagia

Polyphagia	Positive	Negative
Iya	189/320	48/200
Tidak	131/320	152/200

Tabel 9 Probabilitas Genital thrush

Genital_thrush	Positive	Negative
Iya	83/320	33/200
Tidak	237/320	167/200

Tabel 10 Probabilitas Visual blurring

Visual_blurring	Positive	Negative
Iya	175/320	58/200
Tidak	145/320	142/200

Tabel 11 Probabilitas Itching

Itching	Positive	Negative
Iya	154/320	99/200
Tidak	166/320	101/200

Tabel 12 Probabilitas Irrability

Irrability	Positive	Negative
Iya	110/320	16/200
Tidak	210/320	184/200

Tabel 13 Probabilitas Delayed healing

Delayed_healing	Positive	Negative
Iya	153/320	86/200
Tidak	167/320	114/200

Tabel 14 Probabilitas Partial paresis

Partial_paresis	Positive	Negative
Iya	192/320	32/200
Tidak	128/320	168/200

Tabel 15 Probabilitas Muscle stiffness

Muscle_stiffness	Positive	Negative
Iya	135/320	60/200
Tidak	185/320	140/200

Tabel 16 Probabilitas Alopecia

Alopecia	Positive	Negative
Iya	78/320	101/200
Tidak	242/320	99/200

Tabel 17 Probabilitas Obesity

Obesity	Positive	Negative
Iya	61/320	27/200
Tidak	259/320	173/200

- Langkah ketiga : Menguji data testing atau data baru. Data yang diuji dapat ditemukan pada Tabel 18 berikut.

Tabel 18 Data Testing

Age	20
Gender	Famale
Polyuria	Tidak
Polydipsia	Iya
Sudden_Weight_Loss	Iya
Weakness	Tidak
Polyphagia	Tidak
Genital_thrush	Iya
Visual_blurring	Iya
Itching	Tidak
Irrability	Tidak
Delayed_healing	Iya
Partial_paresis	Iya
Muscle_stiffness	Tidak
Alopecia	Iya
Obesity	Iya

- Langkah keempat : Memilih hasil dengan peluang probabilitas tertinggi antara positif atau negatif diabetes. Berikut ini adalah perhitungan manual menggunakan metode Naïve Bayes:

P(Positive | Target) =

$$P(\text{Age } 20 \mid \text{Positive}) * P(\text{Gender Female} \mid \text{Positive}) * P(\text{Polyuria Tidak} \mid \text{Positive}) * P(\text{Polydipsia Iya} \mid \text{Positive}) * P(\text{SWL Iya} \mid \text{Positive}) * P(\text{Weakness Tidak} \mid \text{Positive}) * P(\text{Polyphagia Tidak} \mid \text{Positive}) * P(\text{GT Iya} \mid \text{Positive}) * P(\text{VB Iya} \mid \text{Positive}) * P(\text{Ithcing Tidak} \mid \text{Positive}) * P(\text{Irrability Tidak} \mid \text{Positive}) * P(\text{DH Iya} \mid \text{Positive}) * P(\text{PP Iya} \mid \text{Positive}) * P(\text{MS Tidak} \mid \text{Positive}) * P(\text{Alopecia Iya} \mid \text{Positive}) * P(\text{Obesity Iya} \mid \text{Positive}) = 3/320 * 173/320 * 77/320 * 225/320 * 188/320 * 102/320 * 131/320 * 83/320 * 175/320 * 166/320 * 210/320 * 153/320 * 192/320 * 185/320 * 78/320 * 61/320 = 0.00000002446$$

P(Negative | Target) =

$$P(\text{Age } 20 \mid \text{Negative}) * P(\text{Gender Female} \mid \text{Negative}) * P(\text{Polyuria Tidak} \mid \text{Negative}) * P(\text{Polydipsia Iya} \mid \text{Negative}) * P(\text{SWL Iya} \mid \text{Negative}) * P(\text{Weakness Tidak} \mid \text{Negative}) * P(\text{Polyphagia Tidak} \mid \text{Negative}) * P(\text{GT Iya} \mid \text{Negative}) * P(\text{VB Iya} \mid \text{Negative}) * P(\text{Ithcing Tidak} \mid \text{Negative}) * P(\text{Irrability Tidak} \mid \text{Negative}) * P(\text{DH Iya} \mid \text{Negative}) * P(\text{PP Iya} \mid \text{Negative}) * P(\text{MS Tidak} \mid \text{Negative}) * P(\text{Alopecia Iya} \mid \text{Negative}) * P(\text{Obesity Iya} \mid \text{Negative}) = 0/200 * 19/200 * 185/200 * 185/200 * 8/200 * 29/200 * 113/200 * 152/200 * 33/200 * 58/200 * 101/200 * 184/200 * 86/200 * 140/200 * 101/200 * 27/200 = 0$$

- Langkah kelima : Menentukan kategori klasifikasi data yang diuji apakah termasuk dalam kategori diabetes positive atau negative. Jika $P(\text{Positive}|\text{Target})$

Untuk menampilkan hasil nilai probabilitas setiap kriteria terhadap target, yaitu untuk nilai 1 = positif dan nilai 0 = negatif, dapat dilihat pada gambar output

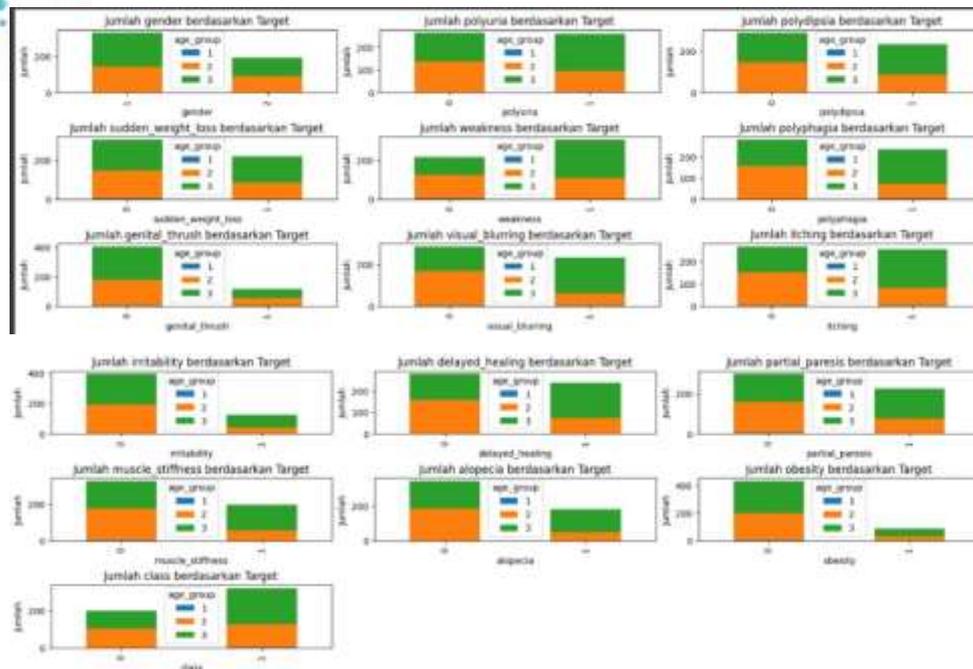


diagram berikut :

Hasil Kode Data Training

KESIMPULAN

Berdasarkan hasil pembahasan, maka dapat disimpulkan sebagai berikut:

1. Klasifikasi risiko diabetes melitus ini menggunakan Algoritma Gaussian Naïve Bayes yang menunjukkan kemampuan tinggi dalam menangani dataset kecil hingga menengah dengan tingkat efisiensi komputasi yang baik.
2. Pengembangan model prediksi diabetes menggunakan dataset pasien dari Sylhet Diabetes Hospital di Bangladesh, yang mencakup 16 atribut klinis. Dataset ini memiliki kelengkapan data yang mendukung validitas model.
3. Algoritma ini efektif dalam menghasilkan prediksi yang relevan karena tingkat akurasi model yang dihasilkan mencapai 91%.
4. Membangun model prediksi menggunakan program berbasis Python memanfaatkan pustaka sklearn, matplotlib, dan pandas yang mampu menghasilkan analisis yang dapat digunakan oleh tenaga medis untuk diagnosis awal diabetes melitus.
5. Hasil prediksi model yang memberikan kontribusi penting dalam membantu tenaga medis dalam mengidentifikasi risiko diabetes secara lebih awal, sehingga dapat digunakan sebagai panduan dalam pengambilan keputusan klinis dan upaya pencegahan komplikasi lebih lanjut

DAFTAR REFERENSI

Al'Ariq, Q., Permadi, G. S., Mashuri, C., & Andriani, Implementasi Naive Bayes dalam Memprediksi Penyakit Diabetes Mellitus. Fakultas Teknologi Informasi, Universitas Hasyim Asy'ari Tebuireng Jombang.
<https://ejournal.unhasy.ac.id/index.php/inovate/article/view/7267>

- Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal Sistem Komputer dan Kecerdasan Buatan*.
<https://jurnal.tau.ac.id/index.php/siskom-kb/article/view/169/146>
- Nofitri, R., & Irawati, N. (2019). Analisis Data Hasil Keuntungan Menggunakan Software RapidMiner. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 5(2), 199–204. <https://doi.org/10.33330/jurteks.v5i2.365>
- Simorangkir, L., & Putra, A. (2024). Klasifikasi Penyakit Diabetes Mellitus Menggunakan Algoritma Naïve Bayes pada Klinik Pratama dan Apotek U.K Jambi. *AKADEMIKA*, 1907-3984, 1–9.
<https://ojs.unh.ac.id/files/journals/1/articles/1296/submission/1296-4-2569-1-2-20241002.docx>
- Citra, A. R., Rudi, H., & Aso, S. (2023). Prediksi Penderita Diabetes Menggunakan Metode Naive Bayes. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 11(3).
<https://pdfs.semanticscholar.org/0fbd/90968454c42d7047d0fdd80b73eb2ea1693b.pdf>
- Fatmawati, S. A. W., & Satrio, H. W. (2023). Implementasi Algoritma Naïve Bayes untuk Mendeteksi Risiko Tinggi Diabetes Melitus pada Ibu Hamil (Studi Kasus: Puskesmas Kabupaten Malang). *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 10(4), 851–856.
<https://jtiik.ub.ac.id/index.php/jtiik/article/view/6422>