

SENASTIKA Universitas Malikussaleh

DIAGNOSIS PENYAKIT DIABETES DALAM DATA MEDIS MENGGUNAKAN METODE NAÏVE BAYES

M.Novil¹, Nadya Putri Dwinta², Muhammad Noval³, Anya Regina Putri⁴

Teknik Informatika, Universitas Malikussaleh, Lhokseumawe, Aceh

Email: ¹novil.210170203@mhs.unimal.ac.id, ²nadya.210170196@mhs.unimal.ac.id, ³muhammad.210170189@mhs.unimal.ac.id, ⁴anya.210170251@mhs.unimal.ac.id

Abstrak

Diabetes yaitu penyakit kronis yang prevalensinya terus meningkat secara global. Deteksi dini dan prediksi risiko diabetes sangat penting untuk pencegahan dan penanganan yang lebih baik. Penelitian ini bertujuan untuk mengembangkan model prediksi penyakit diabetes menggunakan metode Naïve Bayes berdasarkan data medis pasien. Data yang digunakan mencakup faktor-faktor risiko seperti usia, indeks pada massa tubuh, riwayat keluarga, tekanan darah, juga hasil tes laboratorium. Metode Naïve Bayes dipilih karena kemampuannya dalam menangani data kategorik dan numerik, serta efisiensinya pada dataset besar. Dalam penelitian ini terdapat 9 atribut digunakan untuk mengklasifikasi dataset diabetes. Usulan penelitian ini adalah perbandingan beberapa algoritma klasifikasi lain untuk mencapai nilai akurasi yang lebih tinggi. Dengan menggunakan teknik pengukuran ini, akurasi model dan Area Under Curve (AUC) sebagai indikator kualitas prediksi dari algoritma yang digunakan dapat dihitung. Berbeda dengan penelitian sebelumnya, penelitian ini mempertimbangkan beberapa faktor risiko seperti usia, indeks massa tubuh, kadar glukosa, riwayat keluarga, dan hasil tes laboratorium. Fokus penelitian ini pada diabetes. seperti glukosa, hasil laboratorium, dan faktor risiko genetik. Namun, akurasi dapat meningkat hingga 96,27% dengan menggunakan algoritma lain seperti SVM (Support Vector Machine). Namun, penelitian sebelumnya lebih fokus pada penyakit diare pada balita dengan memprediksi gejala yang relevan, seperti jenis kelamin, usia, dan konsistensi tinja. Selain itu, model ini menunjukkan kinerja yang stabil dan dapat diandalkan, baik untuk pasien yang sudah terdiagnosis maupun yang belum menunjukkan gejala yang signifikan. Dengan demikian, penggunaan algoritma Naive Bayes dalam diagnosis penyakit diabetes terbukti efektif, menawarkan solusi yang signifikan dalam membantu dokter dan tenaga medis untuk melakukan deteksi dini serta intervensi yang lebih tepat waktu.

Kata Kunci: Algoritma Naïve Bayes, Penyakit Jantung

Abstract

Diabetes is a chronic disease with a globally increasing prevalence. Early detection risk prediction of diabetes are crucial for better prevention and management. This study aims to develop a diabetes prediction model using the Naïve Bayes method based on patient medical data. The data used includes risk factors such as age, body mass index, family history, blood pressure, and laboratory test results. The Naïve Bayes method was chosen for its ability to handle categorical and numerical data, as well as its efficiency on large datasets. In this study, there are 9 attributes used to classify diabetes datasets. Using this measurement technique, the accuracy of the model and the Area Under Curve (AUC) as an indicator of the prediction quality of the algorithm used can be calculated. Unlike previous studies, this study considered risk factors such as age, body mass index, glucose levels, family history, and laboratory test results. The focus of this research is on diabetes. such as glucose, laboratory results, and genetic risk factors. Using this measurement technique, the accuracy of the model and the Area Under Curve (AUC) as an indicator of the prediction quality of the algorithm used can be calculated. Unlike previous studies, this study considered risk factors such as age, body mass index, glucose levels, family history, and laboratory test results. The focus of this research is on diabetes. such as glucose, laboratory results, and genetic risk factors.

Keyword: Naïve Bayes Algorithm, Heart Disease

1. PENDAHULUAN

Diabetes adalah kondisi nan memengaruhi metabolisme pada tubuh dan menyebabkan hiperglikemia akibat masalah pada sekresi atau fungsi insulin, atau keduanya. Diabetes, yang juga dikenal sebagai diabetes mellitus (DM), merupakan gangguan metabolisme jangka waktu panjang akan terjadi saat pankreas tidak mampu produksi cukup insulin atau ketika tubuh tidak bisa memakai insulin oleh dihasilkan secara efektif. Ibu hamil sering mengalami komplikasi yang terkait dengan diabetes mellitus. Penyakit ini terbagi menjadi beberapa macam, termasuk diabetes mellitus jenis 1 (bergantung pada insulin), diabetes mellitus jenis 2 (tidak bergantung pada insulin), dan diabetes mellitus gestasional. [1]. Naiknya tingkat glukosa atau gula dalam darah adalah tanda penyakit metabolik yang dikenal sebagai diabetes. Glukosa darah sangat penting untuk Kesehatan sangat bergantung pada glukosa sebagai sumber energi utama bagi sel dan jaringan. Jika tidak dikelola dengan baik, diabetes dapat menyebabkan berbagai komplikasi seperti penyakit jantung koroner, stroke, obesitas, serta masalah pada mata, ginjal, dan saraf [2].

Sistem kekebalan tubuh individu yang menderita diabetes tipe 1 dapat merusakkan sel-sel pankreas yang memproduksi insulin, yang menyebabkan peningkatan kadar glukosa darah, sehingga organ tubuh rusak. Diabetes autoimun adalah istilah lain untuk diabetes tipe 1. Keadaan autoimun ini belum diketahui penyebabnya. Faktor genetik pasien, yang dipengaruhi oleh faktor lingkungan, adalah yang paling umum. Diabetes yang paling umum adalah diabetes tipe 2. Resistensi sel tubuh terhadap insulin menyebabkan diabetes jenis ini karena sel-sel tubuh menjadi kurang sensitif terhadap insulin, sehingga insulin yang dihasilkan tidak dapat digunakan dengan baik. Diabetes jenis ini terjadi pada sekitar 90 hingga 95 persen penderita diabetes di seluruh dunia. Diabetes mellitus (DM) adalah penyakit kronis yang akan diderita sepanjang hidup dan berdampak signifikan terhadap kualitas hidup seseorang. Selain itu, pengobatannya memerlukan biaya yang signifikan. Diabetes mellitus dapat menyebabkan banyak komplikasi berbahaya yang membutuhkan pengobatan jangka panjang. Akan tetapi, banyak masyarakat belum menyadari bahaya penyakit ini, tidak tahu dasar tentangnya, dan jarang melihat dokter karena tidak cukup waktu. Tidak ada website yang dapat membantu orang mendiagnosis diabetes mellitus saat ini seperti dokter.[3].

Menurut International Diabetes Federation (IDF), jumlah penderita diabetes mencapai 537 juta orang pada tahun 2021, atau 10,5% dari seluruh populasi. Bahkan diperkirakan akan mencapai 643 juta orang, atau 11,3%, pada tahun 2030. Jika diabetes militus tidak diobati dengan segera dan efektif, ada kemungkinan jumlah kasus akan meningkat di masa depan. Oleh karena itu, sistem klasifikasi penyakit diabetes yang menggunakan metode pengolahan data Naive Bayes ini sangat efektif untuk mengelompokkan jenis pasien untuk penanganan yang lebih cepat dan efisien. [4]

Untuk menyelesaikan masalah di atas, sistem pakar berbasis web harus dibuat. Sistem ini harus dapat mengidentifikasi penyakit diabetes mellitus mengikuti gejala bahwa ditunjukkan dan memberikan saran informasi dan solusi untuk penyakit tersebut. Untuk menghitung probabilitasnya, metode Naive Bayes dapat digunakan. Pendekatan ini mengukur ketidaktentuan dengan probabilitas dan memprediksi probabilitas di masa depan berdasarkan pengalaman masa lalu. Penelitian terdahulu yang dilakukan oleh Hilda Apriyani dan Kurniati dengan judul "Perbandingan Metode Naive Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus" Selanjutnya, akurasi diukur pada melakukan Penilaian matriks kebingungan pada beberapa percobaan unggulan dengan menggunakan data pelatihan dan pengujian. Algoritma SVM yang menggunakan dua kernel, yaitu kernel polinomial dan RBF, memberikan nilai akurasi yang lebih tinggi. Algoritma SVM dengan kernel polynomial mempunyai nilai akurasi tertinggi, 96.2704%, dibandingkan oleh algoritma naif Bayes, yang memiliki nilai akurasi 92.0746%. Ada kemungkinan untuk menyimpulkan mengapa beberapa komponen menyebabkan pada yang lain hasilnya berbeda, algoritma naive Bayes yang lebih baik. Ini disebabkan oleh fakta bahwa peneliti memasukkan fase penghapusan stopword, yang memungkinkan perubahan isi konteks bahwa dasarnya mempunyai arti yang beda, bahkan berlawanan. Berdasarkan analisis, Dapat disimpulkan pula bahwa penghitungan secara manual tidak selalu menjamin klasifikasi data yang akurat [5].

2. METODE PENELITIAN

2.1 Metode Pengumpulan Data

Data yang dikumpulkan menggunakan Metode Studi Pustaka, yang mencari informasi tentang penelitian dari buku-buku dan dokumen, lalu data terkait akan diperlukan pada pembangunan sistem. Data penelitian yang digunakan adalah dataset yang bersumber dari situs data publik yang dapat kita temukan di platform Kaggle.com[6]. <https://www.kaggle.com/code/mbalvi75/08-knn-diabetes-dataset/input?select=diabetes.csv>

2.2 Pembagian Data

Pembagian data membagi kumpulan data menjadi 2 bagian: data pelatihan serta data uji dalam proporsi tertentu. Tujuan dibaginya data yaitu demi memastikan pada model nan] dirancang mampu diterapkan selama baik

pada kumpulan data yang belum pernah diobservasi sebelumnya [7] yaitu atas membagi data latih sebesar 70% serta data uji sebesar 30%.

2.3 Metode Naïve Bayes

Dalam penelitian ini, Naive Bayes digunakan untuk menyelesaikan masalah oleh kemudahan penggunaannya dan pemrosesan nan cepat, Naive Bayes merupakan salah satu algoritma nan paling banyak dipakai pada keperluan data mining. Ini mudah digunakan, memiliki struktur yang sederhana, dan memiliki tingkat efektifitas yang tinggi. Naive Bayes adalah algoritma yang menggunakan metode probabilitas dan statistik untuk mengklasifikasikan variabel [8].

3. ANALISIS DAN PEMBAHASAN

3.1 Analisis Atribut

Penelitian ini menggunakan algoritma Naive Bayes, tujuannya adalah untuk menilai keakuratan algoritma ini dalam mengklasifikasikan diabetes. Terdapat 9 atribut yang dipilih untuk penelitian ini diantaranya kehamilan, gula darah, tekanan darah, ketebalan kulit, insulin, BMI, diabetes, umur, serta hasil akhir. Pada tabel 1 memberikan penjelasan tentang atribut yang ada dalam kumpulan data yang dipakai untuk penelitian ini. Data ini diambil berdasarkan website Kaggle.com dan jumlahnya ada 768 data.

Tabel 1. Atribut Dataset

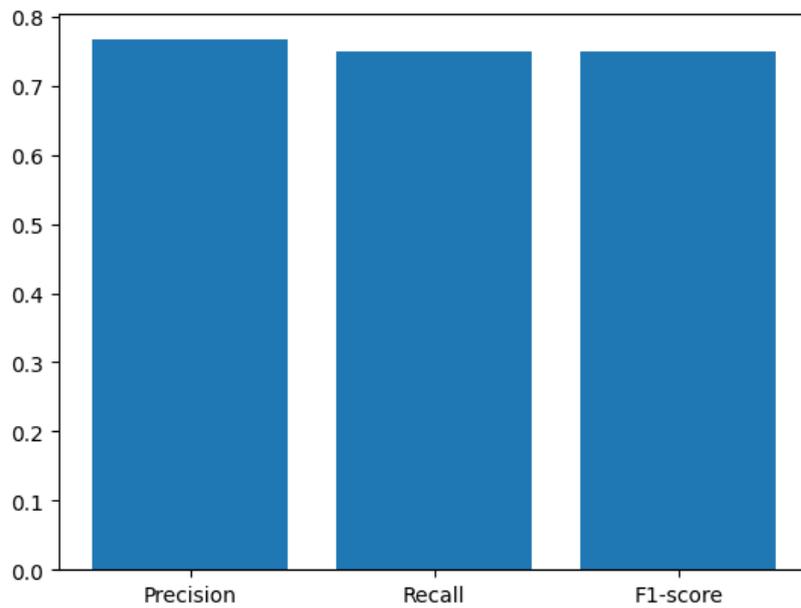
No.	Atribut	Keterangan
1	Pregnancies	Mengungkapkan Jumlah Kehamilan
2	Glucose	Kadar Gula dalam Darah
3	BloodPressure	Pengukuran Tekanan Darah dengan satuan mm Hg
4	SkinThickness	Mengungkapkan Ketebalan Kulit dengan satuan mm
5	Insulin	Menunjukkan Tingkat Kadar Insulin dalam Darah
6	BMI	Mengungkapkan Indeks Massa Tubuh
7	DiabetesPedigreeFunction	Persentase Diabetes
8	Age	Mengungkapkan Umur
9	Outcome	Mengungkapkan Hasil Akhir Positive dan Negative

Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
6	148	72	35	0	33.6	0.627	50	1
1	85	66	29	0	26.6	0.351	31	0
8	183	64	0	0	23.3	0.672	32	1
1	89	66	23	94	28.1	0.167	21	0
0	137	40	35	168	43.1	2.288	33	1
5	116	74	0	0	25.6	0.201	30	0
3	78	50	32	88	31	0.248	26	1
10	115	0	0	0	35.3	0.134	29	0
2	197	70	45	543	30.5	0.158	53	1
8	125	96	0	0	0	0.232	54	1
4	110	92	0	0	37.6	0.191	30	0

10	168	74	0	0	38	0.537	34	1
10	139	80	0	0	27. 1	1.441	57	0
1	189	60	23	846	30. 1	0.398	59	1
...
1	93	70	31	0	30. 4	0.315	23	0

3.2 Preview Model Naïve Bayes

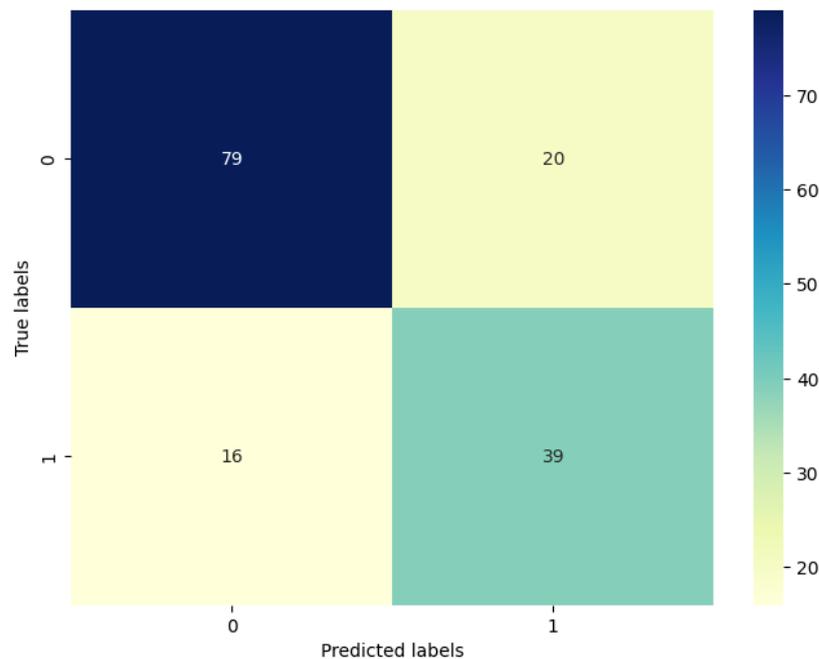
Langkah selanjutnya menampilkan tabel yang sudah diprediksi diagnosa penyakit diabetes untuk melihat fitur-fitur yang memengaruhi hasil prediksi.



Gambar 1

3.3 Hasil Pemodelan Data

Selanjutnya menggunakan confusion matrix agar mengevaluasi kinerja model klasifikasi. Khususnya pada masalah klasifikasi biner atau multi-kelas. Tabel ini menyajikan hasil prediksi model dengan membandingkan antara nilai actual dan nilai yang diprediksi oleh model.



Gambar 2

4. Diskusi

Pada kedua penelitian ini, digunakan algoritma Naive Bayes, yang merupakan metode klasifikasi yang berbasis pada probabilitas. Algoritma ini terbukti efektif dalam mengelola data dengan berbagai atribut, baik yang bersifat kategori maupun numerik, serta mampu menghasilkan model yang cepat dan efisien. Tujuan dari kedua penelitian ini adalah untuk meningkatkan akurasi dalam diagnosis dini penyakit yang diteliti, yaitu diare pada balita dan diabetes pada pasien dewasa. Diagnosis dini ini bertujuan untuk memberikan penanganan yang lebih cepat dan tepat, serta mengurangi risiko terjadinya komplikasi penyakit. Evaluasi kinerja model dalam kedua penelitian dilakukan dengan menggunakan confusion matrix dan kurva ROC (Receiver Operating Characteristic). Metode pengukuran ini digunakan untuk menghitung akurasi model serta Area Under Curve (AUC) sebagai indikator kualitas prediksi dari algoritma yang digunakan. Berbeda dengan penelitian sebelumnya, penelitian ini lebih fokus pada diabetes dengan mempertimbangkan faktor risiko seperti usia, indeks massa tubuh, kadar glukosa, riwayat keluarga, dan hasil tes laboratorium. Sementara itu, penelitian sebelumnya lebih menekankan pada penyakit diare pada balita dengan memprediksi kehadiran gejala-gejala relevan seperti jenis kelamin, usia, konsistensi tinja, dan lain-lain. Penelitian mengenai diabetes menunjukkan bahwa meskipun metode yang digunakan serupa, tingkat akurasi yang diperoleh sedikit lebih rendah, yaitu sekitar 92,07%. Namun, dengan penerapan algoritma lain seperti SVM (Support Vector Machine), akurasi dapat meningkat hingga 96,27%. Dataset yang digunakan dalam penelitian diabetes lebih besar, terdiri dari 768 data yang mencakup atribut yang lebih kompleks, termasuk hasil laboratorium, kadar glukosa, dan faktor risiko genetik. Di sisi lain, penelitian mengenai penyakit diare menunjukkan hasil yang sangat memuaskan dengan akurasi mencapai 98,57% dan AUC sebesar 1.000, yang mencerminkan klasifikasi yang sangat baik. Dataset yang digunakan dalam penelitian diare terdiri dari 68 data pasien dengan atribut seperti usia dan gejala fisik lainnya.

5. KESIMPULAN

Diabetes merupakan kondisi penyakit yang sering kita temui di lingkungan sekitar. Penyakit ini memiliki tanda-tanda berupa tingginya kadar gula darah pada tubuh manusia, yang disebabkan karena tubuh manusia tidak dapat memproduksi insulin secara maksimal. Dalam penelitian ini terdapat 9 atribut digunakan untuk mengklasifikasi dataset diabetes. Usulan penelitian ini adalah perbandingan beberapa algoritma klasifikasi lain untuk mencapai Tingkat akurasi yang lebih unggul.

DAFTAR PUSTAKA

- [1] J. Homepage, B. Delvika, S. Nurhidayarnis, P. D. Rinada, N. Abror, and A. Hidayat, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Comparison of Classification Between Naive Bayes and K-Nearest Neighbor on Diabetes Risk in Pregnant Women Perbandingan Klasifikasi Antara Naive Bayes dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil," vol. 2, pp. 68–75, 2022.
- [2] A. M. Argina, "Indonesian Journal of Data and Science Penerapan Metode Klasifikasi K-Nearest Neighbor pada Dataset Penderita Penyakit Diabetes," vol. 1, no. 2, pp. 29–33, 2020.
- [3] R. Antoni, "OKTAL : Jurnal Ilmu Komputer dan Science SISTEM PAKAR DETEKSI PENYAKIT DIABETES MELLITUS DENGAN MENGGUNAKAN PENDEKATAN NAÏVE BAYESIAN BERBASIS WEB (Studi Kasus : PUSKESMAS Kelurahan Grogol 3)," vol. 2, no. 1, 2023, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal>
- [4] Khoiriah Annisa, Trimaysella Anne, and Adinda Putri Gea, "Penerapan Klasifikasi Data Mining pada Diabetes Menggunakan Metode Naive Bayes," 2024.
- [5] H. Apriyani, "Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Melitus," 2020. [Online]. Available: <https://journal-computing.org/index.php/journal-ita/index>
- [6] A. Desiani, M. Akbar, I. Irmeilyana, and A. Amran, "Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular," *Jurnal Teknik Elektro dan Komputasi (ELKOM)*, vol. 4, no. 2, pp. 207–214, Aug. 2022, doi: 10.32528/elkom.v4i2.7691.
- [7] F. Sholekhah, A. D. Putri, R. Rahmaddeni, and L. Efrizoni, "Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbors untuk Klasifikasi Metabolik Sindrom," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 507–514, Feb. 2024, doi: 10.57152/malcom.v4i2.1249.
- [8] A. C. Khotimah and E. Utami, "COMPARISON NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE IN THE CLASSIFICATION OF INDIVIDUAL ON TWITTER ACCOUNT," *Jurnal Teknik Informatika (JUTIF)*, vol. 3, no. 3, 2022, doi: 10.20884/1.jutif.2022.3.3.254.