JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 3, No. 1 Juni 2025
ISSN 3025-0919 (media online)
Hal 29-33

Pemanfaatan Metode Naive Bayes untuk Klasifikasi Diabetes Berdasarkan Data Medis

Muahammad Rafif Rabbani[1*], Iftikhar Rizqullah[2], Muhammad Wifqi Aufal Maulana[3], Diana Laily Fithri[4]

[1,2,3,4]Fakultas Teknik, Program Studi Sistem informasi, Universitas Muria Kudus, Kudus, Indonesia
Email: [1*]rffrafid@email.com, [2]iftikharrizqullah08@email.com, [3]wifqiam13@email.com, [4]diana.laily@umk.ac.id

Abstrak−Penelitian ini mengimplementasikan algoritma *Naive Bayes* untuk klasifikasi risiko penyakit diabetes berdasarkan dataset Pima Indians Diabetes. Dataset tersebut berisi data medis dari perempuan keturunan Pima Indian berusia ≥21 tahun. Proses analisis mencakup *preprocessing* data, seperti penanganan *missing values* dan normalisasi, serta evaluasi model menggunakan *RapidMiner* dan perhitungan manual. Hasil menunjukkan akurasi sebesar 75.25%, dengan *precision* 61.57% dan recall 65.48%. Temuan ini memperlihatkan bahwa *Naive Bayes* mampu memberikan prediksi yang layak dan dapat dijadikan referensi dalam sistem pendukung keputusan medis.

Kata Kunci: Naive Bayes; Data Mining; Klasifikasi; Diabetes; Medis

Abstract−*This study implements the Naive Bayes algorithm for classifying diabetes risk based on the Pima Indians Diabetes dataset. The dataset contains medical data from Pima Indian women aged ≥21 years. The analysis process includes data preprocessing, such as handling missing values and normalization, and model evaluation using RapidMiner and manual calculation. The results showed an accuracy of 75.25%, with a precision of 61.57% and recall of 65.48%. These findings indicate that Naive Bayes provides reliable predictions and can be referenced in decision support systems for healthcare*.

Keywords: *Naive Bayes; Data Mining; Classification; Diabetes; Medical*

1. PENDAHULUAN

Diabetes melitus merupakan salah satu penyakit tidak menular yang telah menjadi masalah kesehatan global. Berdasarkan data dari *World Health Organization* (WHO), jumlah penderita diabetes meningkat secara signifikan dari tahun ke tahun. Penyakit ini menyebabkan komplikasi serius seperti penyakit jantung, gagal ginjal, dan kebutaan, yang pada akhirnya dapat meningkatkan angka kematian (Rahayu, 2023).

Oleh karena itu, dibutuhkan pendekatan yang efektif untuk deteksi dini dan penanganan cepat terhadap penyakit ini. Namun, proses diagnosis dini yang dilakukan di fasilitas pelayanan kesehatan sering kali memerlukan alat dan metode yang cukup kompleks dan mahal, sehingga tidak semua masyarakat bisa mengaksesnya. Dalam kondisi ini, pemanfaatan teknologi informasi, khususnya teknik data mining, menjadi solusi yang menjanjikan. Data mining memungkinkan proses analisis data berskala besar untuk menemukan pola-pola tertentu yang relevan dalam proses diagnosis (Ridwan, 2020).

Salah satu metode data mining yang banyak digunakan dalam klasifikasi data medis adalah algoritma *Naive Bayes*. Algoritma ini didasarkan pada teori probabilitas Bayes dan memiliki keunggulan dalam hal kesederhanaan, efisiensi komputasi, serta kecepatan dalam pengolahan data (Widiastuti et al., n.d.). Meskipun memiliki asumsi independensi antar fitur, metode ini terbukti efektif dalam berbagai penelitian klasifikasi penyakit (Nurul Anisa, 2022; Nurwijayanti, 2023)

Penelitian ini bertujuan untuk menerapkan metode *Naive Bayes* pada dataset medis Pima Indians Diabetes guna mengevaluasi kemampuannya dalam memprediksi risiko diabetes dan membandingkan hasilnya dengan literatur yang relevan (Veronica Agustin et al., 2023).

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 3, No. 1 Juni 2025
ISSN 3025-0919 (media online)
Hal 29-33

2. METODE

2.1 *Dataset dan Preprocessing*

Dataset yang digunakan dalam penelitian ini adalah Pima Indians Diabetes yang bersumber dari platform Kaggle. Dataset ini berasal dari *National Institute of Diabetes and Digestive and Kidney Diseases*. Dataset ini berisi 768 sampel data dari perempuan keturunan Pima Indian yang berusia minimal 21 tahun. Terdapat 8 fitur prediktor dan 1 target variabel, yaitu *Outcome*, yang bernilai 1 jika pasien menderita diabetes dan 0 jika tidak. Fitur-fitur yang digunakan antara lain jumlah kehamilan (*Pregnancies*), konsentrasi glukosa plasma (*Glucose*), tekanan darah diastolik (*BloodPressure*), ketebalan lipatan kulit (*SkinThickness*), kadar insulin (*Insulin*), indeks massa tubuh (BMI), fungsi silsilah diabetes (*DiabetesPedigreeFunction*), dan usia (*Age*).

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
2	6	148	72	35	0	33.6	0.627	50	1
3	1	85	66	29	0	26.6	0.351	31	0
4	8	183	64	0	0	23.3	0.672	32	1
5	1	89	66	23	94	28.1	0.167	21	0
6	0	137	40	35	168	43.1	2.288	33	1
7	5	116	74	0	0	25.6	0.201	30	0
8	3	78	50	32	88	31	0.248	26	1
9	10	115	0	0	0	35.3	0.134	29	0
10	2	197	70	45	543	30.5	0.158	53	1
11	8	125	96	0	0	0	0.232	54	1
12	4	110	92	0	0	37.6	0.191	30	0
13	10	168	74	0	0	38	0.537	34	1
14	10	139	80	0	0	27.1	1.441	57	0
15	1	189	60	23	846	30.1	0.398	59	1
16	5	166	72	19	175	25.8	0.587	51	1
17	7	100	0	0	0	30	0.484	32	1
18	0	118	84	47	230	45.8	0.551	31	1
19	7	107	74	0	0	29.6	0.254	31	1
20	1	103	30	38	83	43.3	0.183	33	0
21	1	115	70	30	96	34.6	0.529	32	1
22	3	126	88	41	235	39.3	0.704	27	0

Gambar 1. Dataset Pima Indians Diabetes

Sebelum dilakukan analisis, data mengalami tahapan *preprocessing*. Tahapan ini meliputi penanganan nilai hilang (*missing values*) dengan mengganti nilai nol atau kosong pada atribut medis dengan nilai rata-rata kolom terkait (Widiastuti et al., n.d.). Kemudian dilakukan normalisasi menggunakan metode *Min-Max Scaling* untuk menyamakan skala antar fitur. Selanjutnya dilakukan encoding data kategorikal agar dapat diproses oleh algoritma. Semua tahap *preprocessing* dilakukan secara manual di Microsoft Excel sebelum data diimpor ke dalam *RapidMiner*.

2.2. Implementasi Naive Bayes

Setelah data siap, implementasi algoritma *Naive Bayes* dilakukan menggunakan perangkat lunak *RapidMiner*. Teknik validasi yang digunakan adalah *Split Validation*, dengan pembagian data sebesar 70% untuk *training* dan 30% untuk *testing*. Model dilatih dengan operator *Naive Bayes* dan hasilnya diterapkan pada data *testing*.

Evaluasi kinerja model dilakukan melalui *Confusion Matrix* yang menghasilkan nilai-nilai metrik seperti akurasi, *precision*, *recall*, dan *F1-score*. Untuk memastikan validitas hasil klasifikasi, dilakukan juga perhitungan manual menggunakan rumus *Teorema Bayes* pada sebagian data sebagai sampel (Kurniawan, 2023).

3. ANALISA DAN PEMBAHASAN

3.1 Pengujian Data Set

Pengujian dilakukan pada dataset Pima Indians Diabetes yang telah melalui tahap *preprocessing*. Teknik *Split Validation* digunakan untuk membagi data menjadi 70% data latih dan 30% data uji. Model *Naive Bayes* dilatih menggunakan *RapidMiner* dan diuji menggunakan data uji untuk menghasilkan prediksi klasifikasi apakah pasien mengidap diabetes atau tidak.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 3, No. 1 Juni 2025
ISSN 3025-0919 (media online)
Hal 29-33

3.2 Rumus Metode Data Mining

Untuk mengukur kinerja klasifikasi, digunakan metrik evaluasi berbasis *Confusion Matrix* dengan rumus sebagai berikut:

- *Accuracy* = (TP + TN) / (TP + TN + FP + FN)
- *Precision* = TP / (TP + FP)
- *Recall* = TP / (TP + FN)
- *F1-Score* = 2 × (*Precision* × *Recall*) / (*Precision* + *Recall*)

Dimana:

- TP (*True Positive*): Prediksi = 1, Aktual = 1
- TN (*True Negative*): Prediksi = 0, Aktual = 0
- FP (*False Positive*): Prediksi = 1, Aktual = 0
- FN (*False Negative*): Prediksi = 0, Aktual = 1

3.3 Uji Coba Perhitungan Data Set

Hasil pengujian berdasarkan *Confusion Matrix* adalah:

- TP = 165
- FN = 87
- FP = 103
- TN = 413

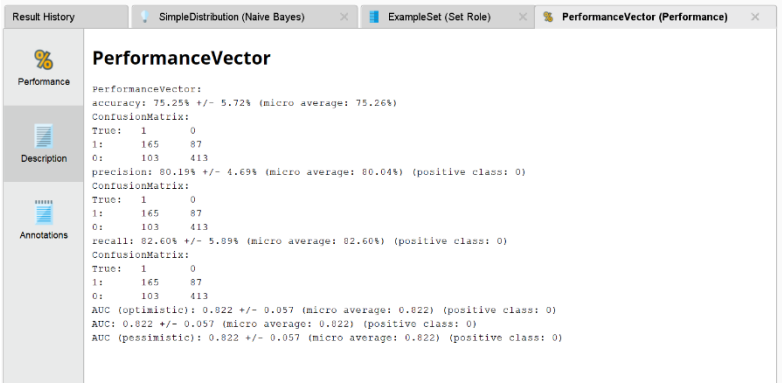
Dengan menggunakan rumus yang telah disebutkan:

- *Accuracy* = (165 + 413) / 768 = 578 / 768 ≈ 75.25%
- *Precision* = 165 / (165 + 103) ≈ 61.57%
- *Recall* = 165 / (165 + 87) ≈ 65.48%
- *F1-Score* = 2 × (0.6157 × 0.6548) / (0.6157 + 0.6548) ≈ 63.49%

3.4 Hasil Data Set yang Sudah Diuji

Hasil tambahan dari pengujian menggunakan operator *Performance* (*Classification*) di *RapidMiner* dengan 10-*fold cross-validation* adalah sebagai berikut:

- *Overall Accuracy*: 75.25% ± 5.72%
- *Class Precision* (kelas 0 – Tidak Diabetes): 80.19% ± 4.69%
- *Class Recall* (kelas 0 – Tidak Diabetes): 82.60% ± 5.89%
- Area *Under ROC Curve* (AUC): 0.822 ± 0.057



Gambar 2. Hasil Visusalisasi Confusion Matrix

Hasil ini menunjukkan bahwa model *Naive Bayes* memiliki performa yang stabil dan cukup akurat dalam mengklasifikasikan risiko diabetes pada pasien berdasarkan data medis.

JRIIN : Jurnal Riset Informatika dan Inovasi
Volume 3, No. 1 Juni 2025
ISSN 3025-0919 (media online)
Hal 29-33

3.5. Perbandingan dengan Literatur

Hasil penelitian ini sejalan dengan hasil studi sebelumnya yang membahas penerapan algoritma *Naive Bayes* dalam klasifikasi penyakit diabetes. Studi oleh (Veronica Agustin et al., 2023) mencatat akurasi sebesar 74.50% dengan dataset serupa, sementara penelitian oleh (Nurwijayanti, 2023) dan (Nurul Anisa, 2022) menunjukkan efektivitas metode ini dalam aplikasi berbasis web dan prediksi medis.

Studi lain oleh (Rahayu, 2023) dan (Ridwan, 2020) juga menyimpulkan bahwa metode *Naive Bayes* unggul dalam efisiensi komputasi dan sederhana dalam implementasi, menjadikannya pilihan populer untuk sistem pendukung keputusan medis. Dengan demikian, hasil dari penelitian ini mendukung kesimpulan literatur bahwa *Naive Bayes* layak digunakan dalam analisis data medis.

Penelitian serupa oleh (Haza Mahendra & Kusumawati, 2025) juga menunjukkan efektivitas algoritma *Naïve Bayes* dalam mendeteksi diabetes mellitus. Dalam penelitian tersebut, digunakan teknik imputasi *mean* pada dataset yang berbeda, menghasilkan akurasi tertinggi sebesar 78.09%. Temuan ini memperkuat kesimpulan bahwa algoritma *Naïve Bayes* merupakan pendekatan yang andal dalam klasifikasi data medis, khususnya untuk prediksi penyakit kronis seperti diabetes.

4. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *Naive Bayes* merupakan metode yang efektif untuk klasifikasi risiko penyakit diabetes berdasarkan data medis. Melalui *preprocessing* yang baik dan implementasi yang tepat, algoritma ini mampu memberikan hasil yang akurat, efisien, dan cukup andal dalam mendeteksi kemungkinan seseorang menderita diabetes.

Dengan akurasi 75.25% serta hasil evaluasi metrik lainnya yang memadai, *Naive Bayes* dapat dijadikan salah satu alternatif metode klasifikasi dalam pengembangan sistem pendukung keputusan di bidang kesehatan. Untuk penelitian selanjutnya, disarankan dilakukan perbandingan dengan algoritma lain seperti *Decision Tree*, *Random Forest*, dan *Support Vector Machine* agar diperoleh analisis yang lebih komprehensif. Selain itu, penerapan teknik *balancing* seperti *SMOTE* dapat membantu mengatasi masalah ketidakseimbangan kelas dalam dataset.

REFERENCES

Haza Mahendra, Y., & Kusumawati, R. (2025). *Analisis Data Mining Untuk Deteksi Diabetes Mellitus Menggunakan Naïve Bayes Data Mining Analysis for Detecting Diabetes Mellitus Using Naïve Bayes*. *6*(1), 39–44. https://doi.org/10.37859/coscitech.v6i1.7954

Kurniawan, A. (2023). SISTEM PENDUKUNG KEPUTUSAN PENYAKIT DIABETES MEGGUNAKAN METODE KLASIFIKASI NAÏVE BAYES. In *Scientia Sacra: Jurnal Sains* (Vol. 3, Issue 2). http://pijarpemikiran.com/index.php/Scientia

Nurul Anisa, D. (2022). KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA NAIVE BAYES. *Dinamika Informatika*, *14*(1).

Nurwijayanti, K. (2023). KLASIFIKASI DIAGNOSA PENYAKIT DIABETES DENGAN METODE NAÏVE BAYES BERBASIS WEB. *Jurnal Kecerdasan Buatan Dan Teknologi Informasi*, *2*(3), 115–121.

Rahayu, C. A. (2023). PREDIKSI PENDERITA DIABETES MENGGUNAKAN METODE NAIVE BAYES. *Jurnal Informatika Dan Teknik Elektro Terapan*, *11*(3). https://doi.org/10.23960/jitet.v11i3.3055

Ridwan, A. (2020). *Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus*.

Veronica Agustin, A., Voutama, A., Singaperbangsa Karawang HS Ronggo Waluyo, U. J., & Barat, J. (2023). IMPLEMENTASI DATA MINING KLASIFIKASI PENYAKIT DIABETES PADA PEREMPUAN MENGGUNAKAN NAÏVE BAYES. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 7, Issue 2).

Widiastuti, W., Abdullah, D., Marwazi, M., Deny, F., & Baiturrahmah, U. (n.d.). EFEKTIVITAS TERAPI INSULIN PADA LATENT AUTOIMMUNE DIABETES IN ADULTS (LADA) DAN KAJIAN KLASIFIKASI DIABETES TIPE 5. In *Journal of Public Health Science (JoPHS)* (Vol. 2).