



## Prediksi Diabetes Menggunakan Algoritma Naive Bayes Menggunakan Rapidminer

Viyani Qomarudin Noor<sup>1</sup>, Amali<sup>2</sup>, Raihan Herfiansyah<sup>3</sup>, Arya Syah Ramadhan<sup>4</sup>

<sup>1,2,3,4</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Pelita Bangsa  
Jl. Inspeksi Kalimantan No. 9, Cibatu, Cikarang Selatan, Kab. Bekasi, Jawa Barat,  
Indonesia

Korespondensi email: [amali@pelitabangsa.ac.id](mailto:amali@pelitabangsa.ac.id)

---

### Abstrak

*Diabetes, the fastest-growing life-threatening disease, presents significant global health challenges, with 3.8 million annual deaths and affecting 422 million individuals worldwide. This study employs RapidMiner, a popular data mining tool, to analyze a public diabetes dataset. Addressing common dataset challenges such as extraneous features, efficient feature selection, and algorithm implementation, the research introduces a novel approach utilizing the Naive Bayes algorithm. Despite inherent complexities, the Naive Bayes model achieves a commendable accuracy rate of 75.51%, showcasing its applicability in diabetic data analysis within the RapidMiner environment. This research sheds light on the potential of integrating RapidMiner and data mining techniques to advance our understanding and management of diabetes-related data, thereby contributing to improved healthcare outcomes and patient well-being.*

---

### Informasi Artikel

Diterima: 8 Desember 2023

Direvisi: 6 Februari 2024

Dipublikasikan: 21 Maret 2024

---

### Keywords

*Diabetes prediction, Naive Bayes, Feature selection, Rapidminer, Data Mining*

## I. Pendahuluan

Diabetes adalah penyakit kronis yang ditandai oleh tingkat gula darah yang tinggi (hiperglikemia). Identifikasi dan prediksi dini penyakit ini sangat penting untuk mengurangi risiko komplikasi jangka panjang dan meningkatkan kualitas hidup pasien. Dalam dunia kesehatan, pengambilan keputusan berbasis data telah menjadi hal yang umum, di mana teknik-teknik pembelajaran mesin (machine learning) digunakan untuk memprediksi kondisi kesehatan seperti diabetes. Salah satu algoritma pembelajaran mesin yang banyak digunakan untuk klasifikasi dan prediksi adalah Naive Bayes. Algoritma ini didasarkan pada teorema Bayes, yang mengasumsikan bahwa fitur-fitur yang digunakan untuk prediksi adalah independen satu sama lain. Meskipun asumsi ini mungkin tidak selalu benar dalam praktik, Naive Bayes seringkali memberikan hasil yang baik, terutama dalam kasus di mana fitur-fitur tersebut memang tidak terlalu berkorelasi.

RapidMiner adalah platform analisis data yang user-friendly dan tidak memerlukan pengetahuan mendalam tentang pemrograman, sehingga sangat cocok digunakan oleh praktisi dan peneliti yang ingin menerapkan algoritma pembelajaran mesin seperti Naive Bayes untuk memprediksi diabetes. Dengan menggunakan RapidMiner, proses ekstraksi, transformasi, dan analisis data dapat dilakukan dengan mudah melalui antarmuka yang intuitif. Dalam penelitian atau studi kasus ini, dataset yang berisi berbagai fitur seperti usia, indeks massa tubuh (BMI), tekanan darah, dan faktor-faktor lainnya akan digunakan sebagai input untuk membangun model prediksi diabetes menggunakan algoritma Naive Bayes di RapidMiner. Hasil dari model ini kemudian dapat digunakan untuk mengevaluasi seberapa baik algoritma tersebut dalam

mengklasifikasikan pasien sebagai penderita diabetes atau tidak, berdasarkan fitur-fitur yang tersedia.

Tujuan utama dari studi ini adalah untuk memahami bagaimana Naive Bayes dapat diterapkan dalam konteks prediksi kesehatan, khususnya untuk penyakit diabetes, serta untuk mengevaluasi kinerja model yang dibangun menggunakan RapidMiner. Dengan demikian, penelitian ini tidak hanya berfokus pada aspek teknis dari algoritma dan alat yang digunakan, tetapi juga pada implikasi praktis dalam konteks klinis.

## II. Metodologi

Berisi mengenai tahapan-tahapan penelitian dimulai dari penelitian awal berupa pengumpulan data sampai kepada penerapan algoritma kepada aplikasi rapid miner sehingga ditemukan kecocokan antara hasil akhir terhadap masing-masing perhitungan. Output dari penelitian ini adalah analisa terhadap hasil prediksi faktor seseorang bisa terkena penyakit diabetes. Pada penelitian ini, peneliti menemukan beberapa keadaan yang membuat peneliti tertarik untuk melakukan suatu penganalisaan terhadap kemungkinan – kemungkinan yang akan membuat seseorang terkena penyakit diabetes. Ada 3 metodologi yang digunakan, Dataset, Naive bayes dan Seleksi fitur.

### 1. Dataset

Dataset yang digunakan disini adalah prediksi diabetes. Dataset ini ada 768 record dan 8 atribut. Deskripsi atribut dataset prediksi diabetes bisa dilihat pada tabel.

Tabel 1. Deskripsi Atribut Diabetes

ATRIBUT	DESKRIPSI
Age	Usia
Diabetes Pedigree Fuction	Metode digunakan dalam evaluasi resiko
BMI	Body Mass Index (Masa Tubuh)
Insulin	Hormon yang diproduksi
SkinThickness	Ketebalan pada kulit
BloodPressure	Tekanan darah
glucose	Kadar gula pada tubuh
Pregnancies	Masa kehamilan

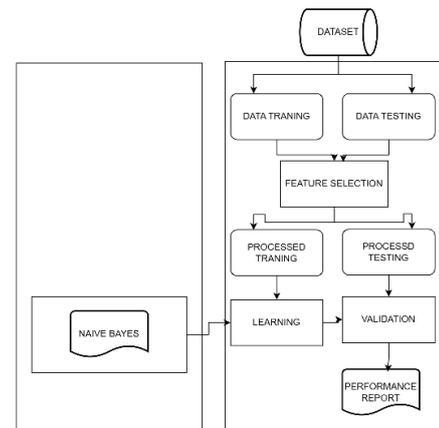
## 2. Naive bayes

Naive bayes merupakan algoritman klasifikasi yang digunakan untuk membuat prediksi. Pada data yang berupa kategori, dengan cara menganalisis sebarapa sering nilai-nilai pada kategori muncul. Naive bayes mampu menghitung probabilitas instance. Algoritma ini efektif dalam menangani dataset dengan fitur-fitur kategori.

## 3. Seleksi fitur

Seleksi fitur merupakan proses dalam data mining dan machine learning. Pentingnya seleksi fitur, dapat meningkatkan kinerja model dengan cara membuang fitur-fitur yang tidak relevan atau pengulangan (redundancy). Hal ini sangat membantu untuk mencegah overfitting dan membuat model lebih efisien. Dengan memilih subset tertentu, interpretasi model menjadi lebih mudah, dan masalah multikolineraritasi dapat diatasi. Mendapatkan model yang lebih sederhana serta lebih mudah di mengerti, jadi langkah seleksi fitur sangat penting dalam membangun model yang efektif dan efisien.

## 4. Kerangka penelitian

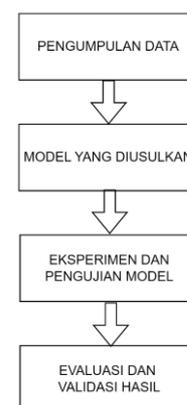


Gambar 1. Kerangka Penelitian

Pada penelitian ini dataset diabetes akan dibagi menjadi 10 bagian menggunakan k-fold cross validation, dimana bagian ini terdiri dari data tranning dan data testing. Seleksi fitur ini untuk menemukan fitur yang paling relevan, kemudian hasil dari proses seleksi ini selanjutnya akan diolah menggunakan alqoritma machine learning Naïve Bayes. Setelah diproses klasifikasi dengan Naïve Bayes, maka akan keluar hasil dari model ini. Gambar 1 merupakan kerangka penelitian yang digunakan.

## 5. Tahapan penelitian

Tahapan penelitian yang digunakan dalam penelitian ini dapat dilihat pada Gambar 2.



Gambar 2. Tahapan Penelitian

Pada gambar 2, tahapan yang dilakukan dalam penelitian ini sebagai berikut:

a. Pengumpulan Data

Pengumpulan data dilakukan dengan mengunduh dataset diabetes di Kaggle Dataset, dataset ini dapat diunduh di [Diabetes Dataset \(kaggle.com\)](https://www.kaggle.com/datasets)

b. Metode yang diusulkan

Model yang digunakan dalam penelitian ini menggunakan algoritma klasifikasi Naïve Bayes, seleksi fitur ini digunakan untuk menseleksi fitur yang relevan sehingga dapat meningkatkan kinerja model yang digunakan.

c. Eksperimen dan Pengujian Model

Eksperimen dan Pengujian Model dilakukan dengan tahapan sebagai berikut:

- 1) Menyapkan dataset.
- 2) Mendisain arsitektur Naïve Bayes.
- 3) Melakukan traning dan testing terhadap model Naïve Bayes.

Spesifikasi komputer yang digunakan dalam eksperimmen ini dapat ditunjukkan pada Tabel 2.

Tabel 2. Spesifikasi Komputer

<b>Processor</b>	<b>Intel Core i3-1005G1</b>
<b>RAM</b>	8 GB DDR4-3200 MHz
<b>Sistem Operasi</b>	Windows 10 Home
<b>Aplikasi</b>	Rapidminer

d. Evaluasi dan Validasi Hasil

Pada saat eksperimen dan pengujian selesai akan keluar hasil dari kinerja model yang disusun yaitu nilai akurasi sebagai tolak ukur seberapa bagus model yang digunakan.

e. Rapidminer

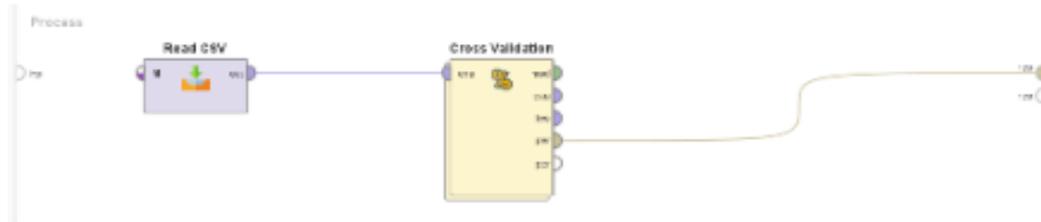
*Rapidminer* merupakan alat atau aplikasi yang digunakan untuk mengolah data, yang bertujuan untuk analisis data. Mulai dari eksperimen pada data hingga pembuatan model prediksi dan evaluasi, dengan desain antarmuka yang sederhana. *Rapidminer* tidak perlu pengetahuan pemrograman mendalam.

f. Data mining

Data mining merupakan proses pengolahan data untuk menentukan pola, prediksi dll. Data mining menggunakan teknik analisis data dan algoritma untuk memproses data yang berukuran besar.

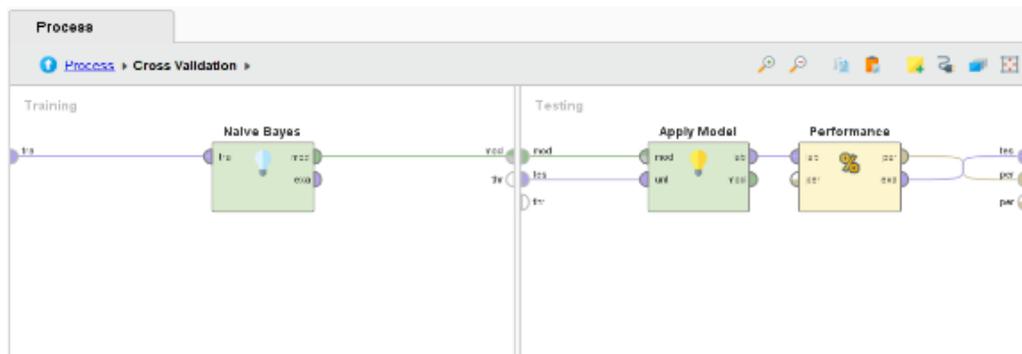
### III. Pembahasan

Dapat dilihat pada gambar 3, dimana operator Read CSV digunakan untuk memanggil dataset diabetes, dan cross Validation digunakan untuk membagi dataset menjadi 10 bagian, yang terdiri dari 2 jenis data yaitu data training dan data testing. Data training berfungsi sebagai melatih data model sedangkan data testing digunakan untuk menguji model yang digunakan.



Gambar 3. Tampilan *Rapidminer* Memanggil Dataset

Pada gambar 4, menunjukan operator Naïve Bayes merupakan algoritma Naïve Bayes, sedangkan operator Apply Model digunakan untuk mengeluarkan rule dari Naïve Bayes dan operator Performance digunakan untuk mengeluarkan hasil kinerja atau akurasi dari model yang digunakan.



Gambar 4. Tampilan *Rapidminer* Mengolah Dataset

Hasil kinerja atau hasil akurasi dapat dilihat pada gambar 5. Dimana hasil akurasi dari eksperimen pada *rapidminer* menggunakan algoritma Naïve Bayes adalah 75.51%.



Gambar 5. Tampilan Kinerja Naive Bayes

pada tabel 3. Dapat dilihat hasil dari confusion matrix algoritma Naïve Bayes, hasil ini berdasarkan eksperimen dan pengujian menggunakan bantuan dari aplikasi *Rapidminer*.

Tabel 3. Confusion Matrix Naive Bayes.

	True positive(1)	True negative(0)	Clas precision
Pred. Positive	162	82	66.39%
Pred. Negative	106	418	79.77%
Class recall	60.45%	83.60%	

#### IV. Kesimpulan

Berdasarkan penelitian di atas, kesimpulan yang di dapat bahwa pada data yang di analisis ada faktor-faktor penyebab kenapa seseorang bisa terkena diabetes, yaitu antar lain kadar gula yang berlebihan (glucose). Dapat memprediksi penyebab terjadinya diabetes dengan data yang digunakan. Dapat memprediksi penyebab terjadinya diabetes dengan cukup akurat, dengan tingkat ke akuratan 75,51%.

#### Daftar Pustaka

- [1] B. A. C. Permana and I. K. Dewi, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naïve Bayes Untuk Prediksi Penyakit Diabetes," *Jurnal Informatika dan Teknologi*, vol. 4, pp. 63-69, 2021.
- [2] N. Nurdiana and A. Algifari, "Studi Komparasi Algoritma Id3 Dan Algoritma Naive Bayes Untuk Klasifikasi Penyakitdiabetes Mellitus," *INFOTECH journal*, vol. 6, pp. 18-23, 2020.
- [3] F. N. Ikhromr, I. Sugiyarto, U. Faddillah and B. Sudarsono, "Implementasi Data Mining Untuk Memprediksi Penyakit Diabetes Menggunakan Algoritma Naive Bayes dan K-Nearest Neighbor," *Journal of Information Technology and Computer Science(INTECOMS)*, vol. 6, pp. 416-428, 2023.
- [4] C. A. Rahayu, "Prediksi Penderita Diabetes Menggunakan Metode Naive Bayes," *Jurnal Informatika Dan Teknik Elektro Terapan*, vol. 11, pp. 1-3, 2023.
- [5] Q. A. Puteri, T. Sagirani and J. Lemantara, "Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor (KNN) untuk Mengetahui Keakuratan Diagnosa Penyakit Diabetes," *Jurnal Nasional Teknologi Dan Sistem Informasi*, vol. 9, pp. 1-3, 2023.
- [6] A. Ardiyansyah, P. A. Rahayuningsih and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner," *Jurnal Khatulistiwa Informatika*, vol. 6, pp. 1-4, 2018.
- [7] S. Rokhanah, A. Hermawan and D. Avianto, "Pengaruh Principal Component Analysis Pada Naïve Bayes dan K-Nearest Neighbor Untuk Prediksi Dini Diabetes Melitus Menggunakan Rapidminer," *Jurnal Sains Dan Manajemen*, vol. 11, pp. 1-12, 2023.
- [8] Fitriyani, "Prediksi Diabetes Menggunakan Algoritma Naïve Bayes dan Greedy Forward Selection," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, pp. 1-9, 2021.
- [9] M. Ardiansyah, A. Sunyoto and E. T. Luthfi, "Analisis Perbandingan Akurasi Algoritma Naïve Bayes dan C4.5 untuk Klasifikasi Diabetes," *EDUMATIC Jurnal Pendidikan Informatika*, vol. 5, pp. 1-10, 2021.
- [10] Hairani, G. S. Nugraha, M. N. Abdillah and M. Innuddin, "Komparasi Akurasi Metode Correlated Naive Bayes Classifier Dan Naive Bayes Classifier Untuk Diagnosis Penyakit Diabetes," *Jurnal*

*Nasional Informatika dan Teknologi  
Jaringan*, vol. 3, pp. 1-5, 2018.