

Penerapan Algoritma Naive Bayes Untuk Mendiagnosa Penyakit Diabetes

Maulana Fansyuri¹, Muhammad Ikhlas Syahidan Zai², Raka Tri Mustakim^{3*}

^{1,2,3}Ilmu Komputer, Teknik Informatika, Universitas Pamulang, Kota Tangerang Selatan, Indonesia
Email: ¹dosen02359@unpam.ac.id, ²keyraucil28@gmail.com, ^{3*}kumakuga28@gmail.com
(* : coresponding author)

Abstrak – Diabetes atau penyakit gula (gula darah tinggi) merupakan penyakit kronis (jangka panjang) yang perlu diwaspadai. Gejala utama penyakit ini adalah peningkatan kadar gula darah di atas normal. Penderita penyakit diabetes yang semakin meningkat di Indonesia. Sehingga Indonesia menempati peringkat kelima negara dengan jumlah penderita diabetes tertinggi, dengan 19,5 juta orang pada tahun 2021 dan diperkirakan akan mencapai 28,6 juta pada tahun 2045, menurut IDF. Dalam menentukan seseorang dapat memiliki penyakit diabetes bisa dilakukan dengan metode Naive Bayes. Pada penelitian ini digunakan dataset yang didapat dari Kaggle. Hasil yang didapat dari menggunakan algoritma Naive Bayes yaitu 92.60%.

Kata Kunci: Naive Bayes; Diabetes; Data Mining; RapidMiner

Abstract – Diabetes or sugar disease (high blood sugar) is a chronic (long-term) disease that needs to be watched out for. The main symptom of this disease is an increase in blood sugar levels above normal. Diabetes is on the rise in Indonesia. So that Indonesia ranks as the fifth country with the highest number of people with diabetes, with 19.5 million people in 2021 and is expected to reach 28.6 million by 2045, according to IDF. In determining someone can have diabetes, it can be done with the Naive Bayes method. In this study, a dataset obtained from Kaggle was used. The results obtained from using the Naive Bayes algorithm are 92.60%.

Keywords: Naive Bayes; Diabetes; Data Mining; RapidMiner

1. PENDAHULUAN

Diabetes adalah penyakit dimana insulin tidak dapat diproduksi oleh pankreas dan digunakan oleh tubuh atau ketika tubuh manusia tidak dapat menggunakan insulin yang telah dibuat oleh pankreas dengan baik. Insulin merupakan hormon yang diciptakan oleh pankreas, yang berfungsi seperti kunci agar glukosa dari makanan yang manusia makan mengalir ke sel-sel dari darah dalam tubuh yang kemudian menghasilkan energi [1].

Menurut data yang dirilis International Diabetes Federation (IDF) pada tahun 2021, Indonesia menempati peringkat kelima dunia dengan 19,47 orang menderita diabetes atau 10,6% dari total penduduk Indonesia. Jumlah penderita diabetes meningkat pesat selama satu dekade terakhir. Dibandingkan dengan 7,29 juta pasien diabetes pada tahun 2011, jumlah pasien diabetes mengalami peningkatan sebesar 167%. IDF memperkirakan pada tahun 2045, jumlah penderita diabetes di Indonesia akan mencapai 28,57 juta jiwa. Jumlah ini mewakili peningkatan 47% dari tahun 2021 [2].

Data mining adalah pengolahan data untuk menghasilkan informasi baru yang berguna [4]. Proses penggalan informasi dari suatu kumpulan data atau kumpulan data disebut data mining [5].

Identifikasi dini risiko diabetes penting untuk mencegah komplikasi dan meningkatkan kualitas hidup pasien. Dalam konteks ini, teknik berbasis kecerdasan buatan, seperti algoritma Naive Bayes, telah menunjukkan potensi penerapan yang besar dalam analisis data medis. Algoritme ini dikenal karena kesederhanaannya dan kemampuannya memproses data dengan cukup akurat bahkan dalam kumpulan data yang kompleks. Penelitian ini bertujuan untuk mengeksplorasi penerapan algoritma Naive Bayes dalam mengidentifikasi diabetes berdasarkan parameter klinis tertentu, sehingga berkontribusi terhadap pengembangan sistem pendukung keputusan di bidang kesehatan.

2. METODE

Metode utama penelitian ini adalah metode kualitatif dan metode kuantitatif. Metode yang digunakan dalam penelitian ini adalah metode penelitian kuantitatif. Tujuan dari penelitian ini

adalah untuk mengklasifikasikan dan mengevaluasi model algoritma Naive Bayes untuk mengetahui keakuratan algoritma Naive Bayes untuk klasifikasi diabetes.

2.1 Data Preparation

Pada penelitian ini akan digunakan dataset yang tersedia untuk umum pada situs Kaggle. Dataset yang digunakan memiliki judul “Diabetes Dataset” dan memuat 597 data dengan 3 atribut. Dimana 2 atribut berupa input dan 1 atribut berupa target. Target yang dimaksud disini yaitu menyatakan seorang pasien tersebut terdiagnosa penyakit diabetes atau tidak. Target yang bernilai ‘NO’ untuk pasien yang tidak mengalami penyakit diabetes dan ‘YES’ untuk pasien yang mengalami penyakit diabetes. Berikut merupakan deskripsi dari masing-masing atribut pada dataset:

Tabel 1. Deskripsi Tabel Dataset

No.	Atribut	Deskripsi
1	Glucose	Konsentrasi glukosa plasma dalam 2 jam
2	Blood Pressure	Tekanan darah diastol
3	Diabetes	Mengidentifikasi penyakit Diabetes

2.2. Preprocessing

Pemrosesan awal data dilakukan dengan menangani nilai yang hilang dengan mengikuti teknik mengabaikan tupel yang nilainya tidak lengkap. Penjelasan rinci tentang dataset dan propertinya ditunjukkan pada **Tabel 1**. Dua kategori variabel digunakan untuk menentukan apakah seorang pasien berisiko terkena diabetes (Yes) atau tidak (No).

2.3. Model

Menggunakan *Software* RapidMiner pada kumpulan data baru untuk melatih dan menguji algoritma Naive Bayes, dan hasilnya dianalisis, termasuk kesalahan klasifikasi, nilai akurasi probabilitas maksimum untuk setiap kelas, perolehan, dan presisi.

2.4. Evaluasi

Selanjutnya Dataset akan diuji dengan Confusion Matrix serta diukur tingkat akurasinya.

3. ANALISA DAN PEMBAHASAN

Pada bagian ini berisi hasil dari kegiatan penelitian yang sudah dilakukan

3.1 Preprocessing

Dari dataset yang akan di analisis dijadikan sebagai Data Training dan Data Testing yang ada di klasifikasikan oleh algoritma Naïve Bayes Adapun contoh data training dan data testing dapat di lihat pada tabel 2 :

Tabel 2. Deskripsi Tabel Dataset

Atribut	Value 1	Value 2
Glucose	40	45
Blood Pressure	85	63
Diabetes	NO	YES

3.2 Model

Setelah Praprocessing, data Training dan Data Testing akan kita proses klasifikasi menggunakan aplikasi Rapidminer Adapun hasil dari Confusion Matrix.nya :

Confusion Matrix

	true NO	true YES	class precision
pred. NO	133	16	89.26%
pred. YES	5	130	96.30%
class recall	96.38%	89.04%	

Gambar 1. Hasil Confusion Matrix

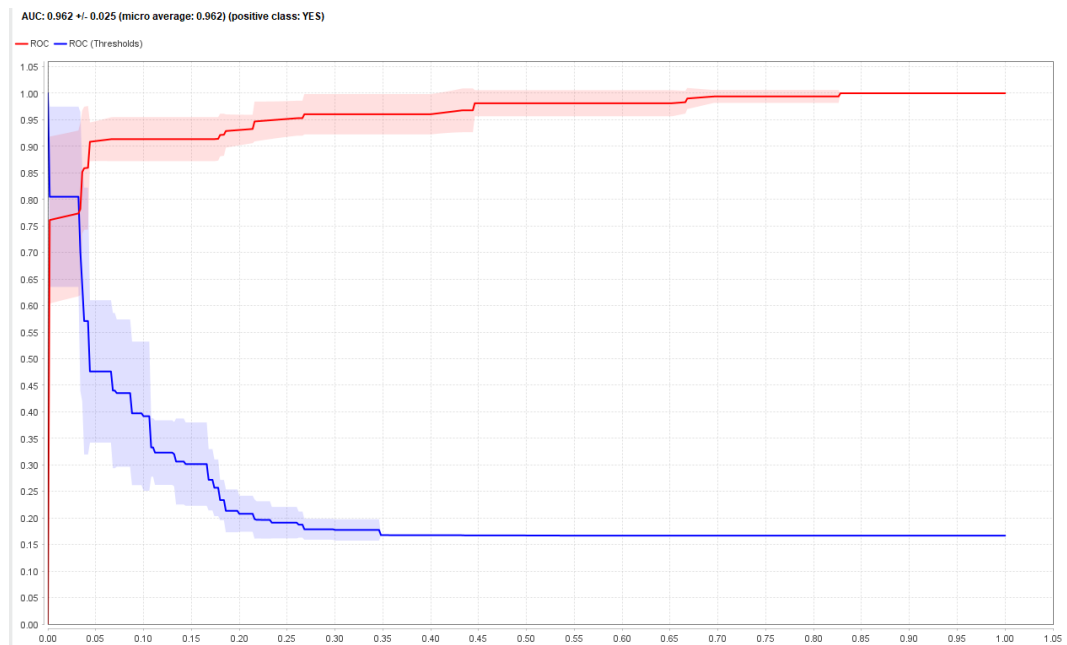
3.3 Evaluasi

Hasil dari evaluasi pengklasifikasian dengan Naive Bayes menghasilkan :

```
PerformanceVector:
accuracy: 92.60% +/- 2.90% (micro average: 92.61%)
ConfusionMatrix:
True:   NO    YES
NO:     133   16
YES:     5    130
```

Gambar 2. Akurasi hasil pengujian Naive Bayes

Dari gambar diatas didapatkan akurasi untuk mengklasifikasikan Dataset Diabetes yaitu 92,60% Adapun hasil Kurv ROC dari pengujian ini Yaitu :



Gambar 3. Kurva ROC hasil pengujian

Kurva ROC (Receiver Operating Characteristic) Algoritma yang ditunjukkan di atas Naive Bayes memiliki nilai AUC sebesar 0,962 Ini berarti klasifikasi sangat baik Hasil penelitian menunjukkan hal itu Naive Bayes memberikan akurasi yang baik untuk Prediksi Diabetes pada UCI dataset Machine Learning Repository.



4. KESIMPULAN

Dari Analisis Algoritma Naive Bayes dapat ditarik kesimpulan bahwa 2 Atribut yang mempengaruhi klasifikasi data set yaitu *Glocose* (Glukosa), dan *Blood Pressure* (Tekanan Darah).

Naive Bayes digunakan untuk pengklasifikasian Dataset Early Stage Diabetes Risk Prediction. Dari 597 data kotor setelah dilakukan pembersihan data tidak lengkap dan melalui proses seleksi sesuai yang dibutuhkan dalam analisa Dataset Stage Diabetes maka didapat data 500 data bersih.

Hasil dari Klasifikasi permodelan menggunakan algoritma naive bayes terhadap Dataset Diabetes menghasilkan Class Precision = 89.26%, dan Class Recall: 89.04%[1].

Hasil evaluasi klasifikasi ini menghasilkan akurasi yang tinggi yaitu 92,60%, yang artinya algoritma Naive Bayes terhadap Dataset Stage Diabetes sudah bagus akurasinya, tetapi perlu peningkatan akurasi dengan Assemble atau dengan cara yang lain[1].

Kurva ROC (Reciver Operating Characteristic) Klasifikasi permodelan algoritma Naïve Bayes terhadap Dataset Stage Diabetes menunjukkan algoritma Naïve Bayes memiliki nilai AUC sebesar 0.962 yang artinya Sangat Bagus[1].

REFERENCES

- Ridwan, A. (2020). Penerapan Algoritma Naïve Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan), 4(1), 15-21.
- Fasnuari, H. A. D., Yuana, H., & Chulkamdi, M. T. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus. ANTIVIRUS: Jurnal Ilmiah Teknik Informatika, 16(2), 133-142.
- Dewi, S., Defit, S., & Yunus, Y. (2020). Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means (Studi Kasus SMP Pembangunan Laboratorium UNP). Jurnal Sistim Informasi Dan Teknologi.
- Mirantika, N.(2021). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat. NUANSA INFORMATIKA, 15(2), 92–98.