

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang Masalah

Diabetes adalah penyakit kronis yang terjadi baik ketika pankreas tidak menghasilkan cukup insulin atau ketika tubuh tidak dapat secara efektif menggunakan insulin yang dihasilkannya. Insulin adalah hormon yang mengatur gula darah. Hiperglikemia, atau peningkatan gula darah, adalah efek umum dari diabetes yang tidak terkontrol dan seiring waktu menyebabkan kerusakan serius pada banyak sistem tubuh, terutama saraf dan pembuluh darah [1].

Menurut data WHO di dunia pada tahun 2014 8,5% orang dewasa berusia 18 tahun dan di atasnya memiliki penyakit diabetes. Pada tahun 2019 diabetes menjadi penyebab langsung 1,5 juta kematian. Untuk menyajikan gambaran yang lebih akurat tentang kematian yang disebabkan oleh diabetes, bagaimanapun, kematian karena glukosa darah yang lebih tinggi dari optimal melalui penyakit kardiovaskular, penyakit ginjal kronis dan tuberkulosis harus ditambahkan. Pada tahun 2012 (tahun data terakhir yang tersedia), ada 2,2 juta kematian lagi karena glukosa darah tinggi [1].

Di Indonesia sendiri berdasarkan data International Diabetes Federation (IDF), Indonesia berstatus waspada diabetes karena menempati urutan ke-7 dari 10 negara dengan jumlah pasien diabetes tertinggi. Pasien pengidap diabetes di Indonesia mencapai 6,2 persen, yang artinya ada lebih dari 10,8 juta orang menderita diabetes per tahun 2020 dan Ketua Umum Perkumpulan Endokrinologi Indonesia (Perkeni), Prof Dr dr Ketut Suastika SpPD-KEMD mengatakan bahwa angka ini diperkirakan meningkat menjadi 16,7 juta pasien per tahun 2045. Oleh karena itu dengan melihat data statistik ini analisis penyakit menggunakan teknologi machine learning memainkan peran vital dalam menolong hidup pasien [2].

Salah satu teknik machine learning yang bisa digunakan adalah klasifikasi. Klasifikasi merupakan teknik supervised learning yang memprediksi class target dari setiap data dimana class-nya sudah ditentukan sebelumnya. Tujuan klasifikasi adalah untuk memprediksi secara akurat class target untuk setiap status dalam data. Pengaplikasian klasifikasi termasuk juga pada diagnosis medis pengenalan gambar dan pola, mendeteksi kesalahan dalam industri, dan sebagainya. Interpretabilitas dan akurasi adalah dua parameter yang digunakan untuk menemukan efisiensi model klasifikasi [3].

Salah satu metode klasifikasi yang bisa digunakan adalah Naïve Bayes yang merupakan salah satu metode machine learning yang biasa digunakan untuk mengklasifikasikan dataset. Metode naïve bayes yang digunakan adalah Gaussian Naïve Bayes.

Pemilihan metode ini juga didasari oleh kemampuan Gaussian Naïve Bayes untuk mengolah fitur-fitur yang bernilai kontinu. Gaussian Naïve Bayes mengasumsikan data kontinu ke dalam distribusi tertentu dan memperkirakan parameter distribusi dengan data latih [4].

Penelitian sebelumnya oleh Orabi, Kamal dan Rabah [5] merancang sistem untuk diagnosis diabetes menggunakan Decision Tree dan mendapatkan hasil akurasi di angka 84%. Penelitian yang lain oleh Kamel, Abdulah dan Al-Tuwajari [6] adalah mengklasifikasikan Cancer menggunakan algoritma Gaussian Naïve Bayes dan mendapatkan akurasi yang tinggi di angka 90%. Adapun Penelitian oleh Sarwar, Nasir, Wajeaha dan Shah menggunakan dataset PIMA India namun menggunakan banyak metode dan membandingkan setiap metode yang ada dan Naïve Bayes disini mendapatkan nilai akurasi sebesar 74%.

U N I V E R S I T A S  
M U L T I M E D I A  
N U S A N T A R A

## 1.2 Rumusan Masalah

Berdasarkan kebutuhan yang dijelaskan sebelumnya, beberapa masalah yang dapat dirumuskan sebagai berikut.

- Bagaimana mengimplementasikan Gaussian Naive Bayes pada aplikasi klasifikasi penyakit diabetes ?
- Berapa hasil akurasi, recall, precision, dan F1-score yang didapatkan dari metode Gaussian Naive Bayes ?

## 1.3 Batasan Masalah

Dalam pengklasifikasian dataset diabetes ini diberikan beberapa batasan masalah sebagai berikut.

- Atribut yang digunakan untuk dijadikan bahan untuk perhitungan adalah berapa kali pasien hamil, nilai glukosa pasien, tekanan darah, ketebalan kulit, insulin, Body Mass Index (BMI), diabetes pedigree function, dan umur.
- Dataset yang digunakan didapat dari kaggle dan bersumber dari *National Institute of Diabetes and Digestive and Kidney Diseases* dan Rumah Sakit Frankfurt Jerman

## 1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang dijelaskan, tujuan dari penelitian adalah sebagai berikut.

- Mengimplementasikan Gaussian Naive Bayes pada aplikasi klasifikasi penyakit diabetes.
- Mengetahui hasil akurasi, recall, precision, dan F1-score yang didapatkan dari aplikasi klasifikasi penyakit diabetes dengan mengimplementasikan metode Gaussian Naive Bayes..

## 1.5 Manfaat Penelitian

Manfaat yang diharapkan dari pengimplementasian algoritma Gaussian Naïve Bayes untuk mengklasifikasikan dataset diabetes adalah untuk membantu

masyarakat dalam melakukan diagnosa awal berdasarkan nilai – nilai dari atribut yang telah ditentukan

### **1.6 Sistematika Penulisan**

Sistematika penulisan disusun dan dibagi atas 5(lima) bab sebagai berikut.

- **BAB 1 PENDAHULUAN**

Bab pertama ini menjelaskan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

- **BAB 2 LANDASAN TEORI**

Bab kedua membahas teori-teori dan konsep dasar yang mendukung dalam penelitian ini, seperti Diabetes, Naïve Bayes, Gaussian Naïve Bayes, Confusion Matrix, Performance Evaluation.

- **BAB 3 METODOLOGI DAN PERANCANGAN APLIKASI**

Bab ketiga menjelaskan metode penelitian yang digunakan, pengumpulan data, dan perancangan aplikasi menggunakan *flowchart*.

- **BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM**

Bab keempat berisi tentang implementasi sistem, diikuti dengan data hasil penelitian yang digunakan.

- **BAB 5 SIMPULAN DAN SARAN**

Bab kelima merupakan bab terakhir yang berisi simpulan dari hasil pengujian aplikasi dan juga saran untuk pengembangan di masa mendatang .