

# BAB I PENDAHULUAN

## 1.1. Latar Belakang

Indonesia menempati peringkat kelima sebagai negara dengan jumlah penderita diabetes terbanyak di dunia. Menurut laporan International Diabetes Federation (IDF), terdapat 19,5 juta penduduk Indonesia yang berusia 20-79 tahun yang menderita penyakit diabetes pada tahun 2021 (Mahdi, 2022). Sementara itu, Indonesia juga menduduki peringkat teratas sebagai negara dengan jumlah penderita diabetes tipe satu terbanyak di Asia Tenggara.

*Diabetes mellitus* adalah kondisi di mana tubuh manusia mengalami gangguan dalam produksi dan penggunaan insulin. Insulin, hormon yang diproduksi oleh pankreas memiliki peran penting dalam mengatur kadar gula darah dari makanan yang dikonsumsi agar dapat digunakan sebagai sumber energi oleh sel-sel tubuh (Afif, 2020). *Diabetes mellitus* adalah penyakit serius yang dapat menyebabkan komplikasi jangka panjang yang berkembang secara perlahan. Seiring berjalannya waktu, pengendalian kadar gula darah pada penderita *diabetes mellitus* menjadi semakin sulit, meningkatkan risiko terjadinya komplikasi yang lebih serius. Dalam beberapa kasus, *diabetes mellitus* dapat mengakibatkan kecacatan dan bahkan mengancam jiwa (Anisa & Jumanto, 2022).

*Diabetes Mellitus* dibagi menjadi dua jenis, yaitu (1) Diabetes Mellitus tipe I, yang disebabkan oleh produksi insulin yang sangat sedikit atau bahkan tidak ada oleh pankreas. Diabetes tipe I biasanya terjadi pada anak-anak dan remaja yang umumnya tidak mengalami obesitas. (2) Diabetes Mellitus tipe II, terjadi ketika produksi insulin oleh pankreas tidak mencukupi atau ketika sel lemak dan otot tubuh menjadi kebal terhadap insulin. Akibatnya, terjadi gangguan dalam pengiriman gula ke seluruh tubuh. Diabetes tipe II ini umumnya terkait dengan faktor-faktor seperti obesitas, tingkat aktivitas fisik, pola makan, dan faktor-faktor lainnya (Wahyuni & Alkaff, 2013).

Berdasarkan laporan International Diabetes Federation (IDF), pada tahun 2022 terdapat 41.813 orang yang menderita diabetes tipe satu di dalam negeri. Dari jumlah tersebut, sebanyak 26.781 orang yang menderita diabetes tipe satu berusia

20-59 tahun, dan sebanyak 13.311 orang berusia di bawah 20 tahun. Sementara itu, 1.721 orang penderita diabetes tipe satu berusia di atas 60 tahun (Widi, 2023). Menurut Media Indonesia jumlah penderita *diabetes mellitus* tipe 2 di Indonesia pada tahun terus meningkat. Diperkirakan jumlah penderita DM tipe 2 di Indonesia telah mencapai 19,5 juta orang. Penyakit *diabetes mellitus* tipe 2 pada umumnya diketahui bisa diderita oleh orang-orang yang berusia diatas 40 tahun. Namun saat ini terjadi peningkatan secara global 56% kasus DM tipe 2 pada usia kurang dari 40 tahun (Zuhdi, 2023).

Menurut Susilawati & Rahmawati (2021) terdapat korelasi antara jenis kelamin dan *diabetes mellitus* tipe 2, di mana perempuan memiliki risiko yang lebih tinggi terkena penyakit tersebut. Beberapa faktor yang menyebabkan perbedaan ini mencakup tingginya kadar kolesterol pada perempuan dibandingkan dengan laki-laki, serta perbedaan dalam pola aktivitas dan gaya hidup sehari-hari yang berkontribusi pada perkembangan potensi penyakit *diabetes mellitus* tipe 2. Laki-laki memiliki persentase lemak tubuh sekitar 15-20% dari berat badannya, sedangkan perempuan memiliki persentase lemak tubuh sekitar 20-25% dari berat badannya. Dengan peningkatan kadar lemak yang lebih tinggi pada perempuan, faktor risiko terkena *diabetes mellitus* pada perempuan menjadi 3-7 kali lebih tinggi dibandingkan dengan laki-laki. Menurut Damayanti, wanita memiliki risiko lebih tinggi untuk mengidap diabetes karena secara fisik wanita memiliki peluang peningkatan indeks massa tubuh yang lebih besar (Wahyuni & Alkaff, 2013).

*Klasifikasi* potensi penyakit diabetes *mellitus* tipe II pada pasien menjadi suatu kebutuhan yang mendesak karena penyakit ini memiliki dampak serius dan dapat berakibat fatal bagi kesehatan manusia. Oleh karena itu, untuk meningkatkan upaya penanggulangan *diabetes mellitus* tipe II, perlu dilakukan penelitian yang fokus pada pengembangan sistem yang mampu mendeteksi awal mula penyakit *diabetes mellitus* tipe II. Hal ini akan memungkinkan untuk mengambil langkah-langkah pencegahan sejak dini. Salah satu pendekatan yang dapat mendukung tujuan ini adalah menggunakan teknik *data mining*.

*Data mining* atau penambangan data, merupakan pendekatan yang efektif untuk mengatasi tantangan dalam penanganan penyakit *diabetes mellitus* tipe II.

Pemanfaatan teknik *data mining* bertujuan untuk menggali pola dan informasi berharga dari data pasien, termasuk faktor risiko, gejala, dan rekam medis. Hal ini memungkinkan untuk mengklasifikasi kemungkinan seseorang mengalami *diabetes mellitus* tipe II, sehingga intervensi dini dan perawatan yang tepat dapat dilakukan. Dalam *data mining*, salah satu teknik yang diterapkan adalah menelusuri data yang ada untuk membangun model, yang nantinya dapat digunakan untuk mengenali pola data lain yang tidak terdapat dalam basis data yang tersimpan.

*Klasifikasi* sebagai bagian dari teknik *data mining*, melibatkan penilaian objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia. Dalam *klasifikasi*, dua pekerjaan utama dilakukan, yaitu pembangunan model sebagai prototipe untuk disimpan sebagai memori dan penggunaan model untuk melakukan pengenalan, klasifikasi, atau prediksi pada objek data lain untuk menentukan kelasnya dalam model yang telah disimpan (Prasetyo, 2012).

*Naïve Bayes* sebagai salah satu algoritma dalam *klasifikasi* terbukti dapat menghasilkan performa yang baik (Afif, 2020). Berikut adalah penelitian yang menggunakan algoritma *Naïve Bayes* dan menghasilkan performa baik oleh (Supandi, et al., 2022) dengan judul “Prediksi Penyakit Diabetes Menggunakan *Machine Learning* Dengan Algoritma *Naïve Bayes*” yang menghasilkan akurasi sebesar 95,94%. *K-Nearest Neighbor* merupakan algoritma yang dapat menghasilkan nilai *klasifikasi* yang baik jika menggunakan data dalam jumlah yang besar (Mustafa & Simpen, 2019). Berikut adalah penelitian yang menggunakan algoritma KNN dan menghasilkan nilai *klasifikasi* yang baik oleh (Raharja, et al., 2021) dengan judul “Perbandingan Kinerja Algoritma *Gaussian Naïve Bayes* dan *K-Nearest Neighbor (KNN)* Untuk Mengklasifikasi Penyakit *Hepatitis C Virus (HCV)*” yang menghasilkan nilai akurasi sebesar 91,80%. Terdapat penelitian terdahulu terkait perbandingan metode antara *Naïve Bayes* dan KNN yang disusun oleh (Amien, et al., 2023) dengan judul “Perbandingan Metode *Naïve Bayes* dan *KNN (K-Nearest Neighbor)* dalam *Klasifikasi* Penyakit Diabetes” menghasilkan bahwa pada metode *klasifikasi* KNN mendapatkan nilai akurasi sebesar 90%, sedangkan pada metode *klasifikasi Naïve Bayes* mendapatkan nilai akurasi sebesar 78%. Namun pada penelitian ini hasil akurasi menggunakan Algoritme *Naïve*



*Bayes* lebih besar dari pada KNN. Dari perbandingan diatas algoritme KNN adalah algoritme terbaik, akan tetapi pada *klasifikasi Diabetes Mellitus* Tipe II belum ada yang melakukan perbandingan. Sehingga pada penelitian akan dilakukan perbandingan tersebut.

*Data mining* digunakan untuk mengklasifikasikan penyakit *Diabetes Mellitus* tipe II. Metode *Naïve Bayes* memiliki keunggulan dalam kecepatan pembuatan model, kemampuan prediksi yang baik, dan menyediakan metode baru untuk eksplorasi dan pemahaman data. Kaitan antara *Naïve Bayes* dengan *klasifikasi* terletak pada fakta bahwa hipotesis dalam teorema Bayes menjadi label kelas yang menjadi tujuan pemetaan dalam *klasifikasi* (Prasetyo, 2012).

Terdapat penelitian yang pernah dilakukan sebelumnya oleh (Supandi, et al., 2022) dengan judul “Prediksi Penyakit Diabetes Menggunakan *Machine Learning* Dengan Algoritma *Naïve Bayes*” yang menghasilkan akurasi sebesar 95,94%. Kemudian penelitian oleh (Febrian, et al., 2023) dengan judul “Diabetes prediction using *supervised machine learning*” yang membandingkan algoritma KNN dan *Naïve Bayes*, hasil yang didapat yaitu algoritma *Naïve Bayes* lebih unggul dibandingkan dengan algoritma KNN dengan nilai akurasi dari algoritma *Naïve Bayes* sebesar 78,57% sedangkan nilai akurasi dari algoritma KNN sebesar 77,92%. Kemudian penelitian oleh (Anisa & Jumanto, 2022) dengan judul “*Klasifikasi* Penyakit Diabetes menggunakan Algoritma *Naïve Bayes*” menghasilkan bahwa perempuan yang menderita penyakit diabetes lebih banyak daripada laki-laki, dengan nilai tingkat akurasi sebesar 91,6%. Kemudian penelitian oleh (Afif, 2020) dengan judul “Penerapan Algoritma *Naïve Bayes* Untuk *Klasifikasi* Penyakit *Diabetes Mellitus* di Rumah Sakit Aisyiah” menghasilkan nilai akurasi 90,20%. Kemudian penelitian oleh (Wahyuni & Alkaff, 2013) dengan judul “*Diabetes Mellitus* Pada Perempuan Usia Produksi di Indonesia Tahun 2007” menunjukkan bahwa persentase DM pada perempuan usia reproduksi (15-49 tahun) adalah 3,6%. Kemudian penelitian oleh (Hasanah, et al., 2022) dengan judul “Analisis Algoritma *Gaussian Naïve Bayes* Terhadap *Klasifikasi* Data Pasien Penderita Gagal Jantung” menghasilkan nilai akurasi sebesar 69%, nilai presisi sebesar 65,73% dan nilai recall sebesar 95,91%. Kemudian penelitian oleh (Putra & Akbar, 2016) dengan

judul “*Klasifikasi Pengidap Diabetes Pada Perempuan Menggunakan Penggabungan Metode Support Vector Machine dan K-Nearest Neighbour*” menghasilkan nilai akurasi sebesar 92%. Kemudian penelitian oleh (Apriyani & Kusniati, 2019) dengan judul “*Perbandingan Metode Naïve Bayes Dan Support Vector Machine Dalam Klasifikasi Penyakit Diabetes Mellitus (Studi Kasus: RS. Siti Khadijah Palembang)*” menghasilkan nilai akurasi dari *Confusion Matrix Naïve Bayes* didapat sebesar 92,0746% dan eror sebesar 7,9254%, sedangkan hasil nilai akurasi dari *Confusion Matrix SVM* dengan kernel *Polynomial* didapat sebesar 96,2704% dan eror sebesar 3,7296%. Kemudian penelitian oleh (Ningsih, et al., 2023) dengan judul “*Pendekatan Data Science untuk Deteksi Dini Diabetes Menggunakan Naïve Bayes Classifier*” menghasilkan nilai akurasi sebesar 70% yang berasal dari pembagian dataset sebesar 40% untuk data testing dan 60% untuk data training. Kemudian penelitian oleh (Putra & Akbar, 2016) dengan judul “*Penerapan Metode Naïve Bayes Dalam Mengklasifikasi Tingkat Obesitas Pada Pria*” menghasilkan nilai akurasi sebesar 83,15% dengan menggunakan split data 60% untuk data training dan 40% untuk data testing.

Penelitian ini bertujuan untuk mendeteksi secara dini keberadaan penyakit *diabetes mellitus* dengan menerapkan algoritme *klasifikasi Naïve Bayes Gaussian* dalam data mining. Pengujian dilakukan menggunakan data set yang diperoleh dari Kaggle, sehingga performa algoritme *Naïve Bayes Gaussian* terhadap dataset ini akan diketahui. Pengujian dilakukan menggunakan *Confusion Matrix* dengan menggunakan *Python* sebagai bahasa pemrograman, dan hasil implementasinya akan diterapkan pada suatu sistem yang akan dibangun menggunakan bahasa *Python* dan *framework Streamlit*. Setelah itu, sistem tersebut akan di hosting menggunakan *Streamlit Share*.

## 1.2. Rumusan Masalah

1. Bagaimana implementasi *Naïve Bayes Gaussian* dan KNN untuk *klasifikasi* potensi penyakit *diabetes mellitus* tipe II berdasarkan data pasien?

2. Bagaimana hasil perbandingan *Naïve Bayes Gaussian* dan KNN untuk *klasifikasi* potensi penyakit *diabetes mellitus* tipe II berdasarkan data pasien?
3. Bagaimana rancang bangun sistem *klasifikasi Diabetes Mellitus* tipe II dengan *Naïve Bayes Gaussian* menggunakan *waterfall*?

### 1.3. Batasan Masalah

Batasan masalah yang digunakan dalam penelitian ini adalah :

1. Data diabetes ini hanya tipe II.
2. Parameter yang digunakan adalah Age, Gender, BMI (*Body Mass Index*), SBP (*Systolic Blood Pressure*), DBP (*Diastolic Blood Pressure*), FPG (*Fasting Plasma Glucose*), FFPG (*Final Fasting Plasma Glucose*) dan Cholesterol.
3. Dataset yang digunakan berasal dari studi penelitian Tiongkok pada tahun 2016.

### 1.4. Tujuan Penelitian

Untuk mengetahui apakah pasien terkena penyakit diabetes atau tidak dengan menerapkan algoritme *Naïve Bayes* pada model *machine learning* yang telah dibuat untuk mengklasifikasi penyakit diabetes.

### 1.5. Manfaat Penelitian

Manfaat dari penelitian ini adalah peneliti ingin mengembangkan model *klasifikasi* penyakit diabetes yang dapat membantu dalam mendiagnosis penyakit diabetes dengan baik dan benar. Dengan menggunakan algoritme *Naïve Bayes*, diharapkan dapat memperoleh kemungkinan probabilitas seseorang menderita penyakit diabetes dan menentukan langkah-langkah pencegahan atau pengobatan yang tepat.

## 1.6. Metodologi Penelitian

Metode yang digunakan untuk menyelesaikan penelitian ini adalah :

### a) *Dataset*

Pada penelitian ini, data didapatkan dari situs web Kaggle Diabetes\_Dataset\_With\_18\_Feature. Data set ini memiliki 17 variabel dan 1 label kelas dengan total 4.303 data yang terbagi menjadi dua yaitu 3000 data pasien tidak terkena diabetes dan 1303 data pasien terkena diabetes.

### b) *Preprocessing*

Dalam penelitian ini, tahapan preprocessing data menggunakan bahasa pemrograman *Python* sebagai alat analisis data. Pada tahap ini, variabel akan disortir menjadi 8 dan dilakukan pemisahan data antara atribut dan label. Selanjutnya, data akan dibagi menjadi data train dan data test, dengan total data train sebanyak 4.283 dan data test sebanyak 20.

### c) *Klasifikasi*

Dalam tahap ini, akan dilakukan pembangunan model *machine learning* menggunakan algoritme *klasifikasi Naïve Bayes*.

### d) *Evaluation*

Pada tahap ini, dataset akan diuji dan dievaluasi menggunakan *Confusion Matrix* serta akan dilakukan pengukuran tingkat akurasi.

### e) *Prediction*

Pada tahap ini akan dibuat model prediksi yang berfungsi untuk melakukan pengenalan/klasifikasi/prediksi pada suatu objek data lain agar diketahui di kelas mana objek data tersebut dalam model yang sudah disimpan.

### f) *Simpan model*

Pada tahap ini akan dilakukan penyimpanan model yang berguna sebagai mesin dalam website.



## 1.7 Sistematika Penulisan

Adapun sistematika penulisan skripsi ini dirancang sedemikian rupa agar mempermudah pembaca dalam memahami maksud dan tujuan penelitian ini. Setiap bab dan bagian disusun secara logis dan sistematis untuk memberikan alur yang jelas dan terstruktur. Sistematika penulisan skripsi ini terdiri dari beberapa bagian utama, yang diuraikan sebagai berikut:

**BAB I : PENDAHULUAN**

Dalam bab ini dijelaskan mengenai Latar Belakang, Rumusan Masalah, Batasan Masalah, Tujuan Penelitian, Manfaat Penelitian, Metodologi Penelitian, dan Sistematika Penulisan.

**BAB II : LANDASAN TEORI**

Bab ini membahas teori-teori pendukung yang berkaitan dengan proses perancangan, pembuatan, implementasi, dan pengujian sistem.

**BAB III : ANALISIS DAN PERANCANGAN SISTEM**

Dalam bab ini membahas tentang Spesifikasi Kebutuhan Sistem, Analisis Sistem, Perancangan Metode, Perancangan Sistem, Perancangan Antarmuka Sistem dan Perancangan Pengujian Sistem yang akan digunakan dalam tahap implementasi sistem.

**BAB IV : IMPLEMENTASI DAN PENGUJIAN SISTEM**

Dalam bab ini, akan dibahas mengenai implementasi sistem yang merujuk pada tahap penerapan sistem yang telah dirancang dan dikembangkan, serta pengujian sistem yang melibatkan proses evaluasi terhadap kinerja sistem yang telah diimplementasikan.

**BAB V : PENUTUP**

Dalam bab ini berisi tentang kesimpulan dan saran.