

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Diabetes merupakan salah satu penyakit kronis yang menjadi permasalahan kesehatan global. Penyakit ini ditandai dengan meningkatnya kadar glukosa dalam darah akibat gangguan metabolisme tubuh dalam memanfaatkan glukosa sebagai sumber energi. Apabila kondisi tersebut tidak dikendalikan dengan baik, diabetes dapat menyebabkan berbagai komplikasi serius pada organ tubuh, seperti gangguan ginjal, jantung, mata, dan sistem saraf (Azizah et al., 2023).

Jumlah penderita diabetes terus mengalami peningkatan, baik secara global maupun nasional. Jumlah penderita diabetes tercatat mencapai 830 juta orang pada tahun 2022, dengan sekitar 1,6 juta kematian akibat diabetes pada tahun 2021 (World Health Organization, 2024). Selain itu, Indonesia menempati peringkat kelima dunia dengan jumlah penderita diabetes sebanyak 19,5 juta orang pada tahun 2021 dan meningkat menjadi 28,6 juta orang pada tahun 2045 (*International Diabetes Federation*, 2024 sebagaimana dikutip dalam Yuntarisa et al., 2025).

Peningkatan jumlah penderita diabetes juga terlihat berdasarkan hasil Survei Kesehatan Indonesia (SKI) tahun 2023. Berdasarkan diagnosis dokter, prevalensi diabetes tercatat sebesar 1,7%, sedangkan berdasarkan pemeriksaan kadar gula darah langsung pada penduduk usia ≥ 15 tahun mencapai 11,7% (Kemenkes RI, 2023). Perbedaan angka tersebut menunjukkan bahwa masih terdapat banyak kasus diabetes yang belum teridentifikasi secara klinis. Oleh karena itu, diperlukan metode deteksi dini yang dapat membantu memberikan informasi awal mengenai risiko diabetes sebelum dilakukan pemeriksaan lebih lanjut.

Salah satu pendekatan yang dapat digunakan untuk mendukung deteksi dini diabetes adalah pemanfaatan teknologi *data mining* dan *machine learning*. *Data mining* merupakan proses pengolahan data untuk menemukan pola atau informasi

tertentu yang dapat digunakan sebagai pendukung pengambilan Keputusan (Yuniar&Rohayani, 2025). Dalam bidang kesehatan, *machine learning* dapat diterapkan untuk menganalisis data pasien dan menemukan pola yang berkaitan dengan kondisi tertentu, salah satunya melalui metode klasifikasi.

Metode klasifikasi digunakan untuk mengelompokkan data berdasarkan pola atau karakteristik tertentu ke dalam kelas yang telah ditentukan. Beberapa algoritma klasifikasi yang sering digunakan dalam prediksi penyakit adalah *Naive Bayes* dan *Decision Tree*. *Naive Bayes* merupakan algoritma berbasis probabilitas yang memiliki proses komputasi sederhana serta mampu bekerja dengan baik pada proses klasifikasi data (Kawani, 2019). Sementara itu, *Decision Tree* merupakan algoritma yang menghasilkan aturan keputusan dalam bentuk struktur pohon sehingga hasil prediksinya lebih mudah dipahami dan diinterpretasikan (Permana & Patwari, 2021). Perbedaan karakteristik kedua algoritma tersebut menjadi dasar dilakukannya perbandingan performa antara *Naive Bayes* dan *Decision Tree* dalam penelitian ini. Perbandingan algoritma diperlukan karena setiap metode memiliki kelebihan dan keterbatasan yang berbeda dalam mengolah suatu dataset. Melalui perbandingan tersebut, dapat diketahui algoritma yang memiliki kinerja lebih baik dalam melakukan prediksi diabetes berdasarkan data yang digunakan.

Beberapa penelitian sebelumnya telah menerapkan *machine learning* untuk prediksi diabetes. (Huwaidi et al., 2022) mengembangkan sistem deteksi diabetes berbasis web menggunakan algoritma *K-Nearest Neighbor* pada dataset *Pima Indians*, sedangkan (Fauzi, 2025) mengembangkan aplikasi sistem pakar diabetes menggunakan algoritma *Naive Bayes*. Kedua penelitian tersebut telah mengimplementasikan model prediksi dalam suatu sistem, tetapi masih menggunakan satu algoritma sehingga belum memberikan informasi mengenai perbandingan performa beberapa algoritma.

Selain itu, penelitian (Nuryamin & Risyda, 2025) telah melakukan analisis perbandingan algoritma *Decision Tree C4.5* dan *Naive Bayes* untuk prediksi penyakit diabetes. Namun, penelitian tersebut masih berfokus pada pengujian performa model

tanpa implementasi sistem berbasis web. Berdasarkan penelitian terdahulu tersebut, terdapat peluang untuk mengembangkan penelitian yang tidak hanya membandingkan performa algoritma, tetapi juga mengimplementasikan model terbaik ke dalam sistem berbasis web agar hasil prediksi dapat digunakan secara lebih mudah.

Pada penelitian ini digunakan *Pima Indians Diabetes Dataset* sebagai sumber data penelitian. Dataset tersebut digunakan untuk membangun model prediksi diabetes dengan menerapkan algoritma *Naive Bayes* dan *Decision Tree*. Tahapan penelitian meliputi *preprocessing* data, pelatihan model, optimasi parameter menggunakan *GridSearchCV*, serta evaluasi performa menggunakan akurasi, sensitivitas, dan spesifisitas berdasarkan *confusion matrix*.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk melakukan analisis perbandingan algoritma *Naive Bayes* dan *Decision Tree* dalam melakukan prediksi risiko diabetes serta mengimplementasikan model terbaik ke dalam sistem berbasis web. Pada penelitian ini, sistem yang dikembangkan tidak bertujuan untuk menggantikan proses diagnosis medis yang dilakukan oleh tenaga kesehatan, melainkan sebagai alat bantu dalam melakukan skrining awal risiko diabetes berdasarkan pola data yang tersedia. Hasil prediksi yang diberikan oleh sistem dapat digunakan sebagai informasi pendukung sehingga pengguna dapat melakukan tindakan lanjutan berupa pemeriksaan medis apabila diperlukan.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan algoritma *Naive Bayes* dan *Decision Tree* untuk memprediksi penyakit diabetes berdasarkan data pasien?
2. Bagaimana hasil perbandingan performa algoritma *Naive Bayes* dan *Decision Tree* untuk memprediksi penyakit diabetes berdasarkan metrik evaluasi model?

3. Bagaimana implementasi sistem berbasis web yang dapat digunakan untuk melakukan prediksi penyakit diabetes serta menampilkan hasil komparasi performa algoritma?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijelaskan, maka tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma *Naive Bayes* dan *Decision Tree* untuk melakukan prediksi penyakit diabetes berdasarkan data pasien.
2. Menganalisis hasil perbandingan performa algoritma *Naive Bayes* dan *Decision Tree* dalam memprediksi penyakit diabetes berdasarkan metrik evaluasi, yaitu akurasi, sensitivitas, dan spesifisitas.
3. Mengimplementasikan sistem berbasis web yang dapat digunakan untuk melakukan prediksi penyakit diabetes serta menampilkan hasil komparasi performa algoritma.

1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Manfaat bagi peneliti

Penelitian ini dapat menambah pemahaman peneliti mengenai penerapan *machine learning*, khususnya algoritma *Naive Bayes* dan *Decision Tree* dalam proses klasifikasi penyakit diabetes. Selain itu, penelitian ini juga memberikan pengalaman dalam melakukan *preprocessing* data, pelatihan model, evaluasi model, perbandingan algoritma, dan implementasi model ke dalam sistem berbasis web.

2. Manfaat bagi pengguna sistem

Penelitian ini memberikan sistem berbasis web yang dapat membantu admin atau tenaga kesehatan dalam melakukan prediksi awal risiko diabetes pasien secara cepat, mudah, dan terdokumentasi.

3. Manfaat bagi pengembangan ilmu pengetahuan dan sistem

Penelitian ini dapat menjadi referensi bagi penelitian selanjutnya yang membahas penerapan algoritma *machine learning* dalam bidang Kesehatan sekaligus menjadi dasar pengembangan sistem prediksi penyakit berbasis web yang lebih lanjut.

1.5 Batasan Masalah

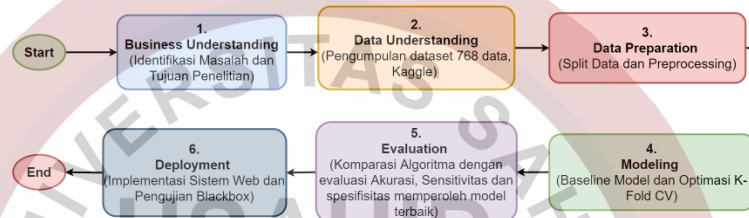
Agar penelitian ini lebih terarah dan tidak menyimpang dari tujuan yang telah ditentukan, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini menggunakan *Pima Indians Diabetes Dataset* yang diperoleh dari *Kaggle* sebagai data sekunder untuk proses pelatihan dan pengujian model prediksi diabetes.
2. Algoritma klasifikasi yang digunakan dalam penelitian ini terbatas pada *Naive Bayes* dan *Decision Tree* untuk memprediksi status diabetes.
3. Variabel input yang digunakan terdiri dari lima fitur, yaitu *Age*, *Glucose*, *BloodPressure*, *Insulin*, dan *BMI*, dengan *Outcome* sebagai label target.
4. Evaluasi performa model dilakukan menggunakan metrik akurasi, sensitivitas, spesifisitas, dan *confusion matrix* untuk membandingkan kinerja kedua algoritma.
5. Sistem yang dikembangkan berbasis web dan digunakan sebagai media pendukung skrining awal risiko diabetes melalui proses input data pasien dan prediksi menggunakan model yang telah dibangun.
6. Penelitian ini berfokus pada pengembangan sistem prediksi dan perbandingan performa algoritma, sehingga tidak membahas diagnosis medis, keamanan data, maupun pengelolaan rekam medis secara mendalam.

1.6 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan jenis non-eksperimental dan metode komparatif. Pendekatan kuantitatif digunakan karena data yang dianalisis berupa angka, yaitu hasil pengukuran performa algoritma. Penelitian bersifat non-eksperimental karena data yang digunakan adalah data

sekunder, yaitu Pima Indians Diabetes *Dataset* yang diperoleh dari *Kaggle*, sehingga peneliti tidak melakukan pengambilan data atau perlakuan langsung terhadap pasien. Adapun metode komparatif digunakan untuk membandingkan performa algoritma *Naive Bayes* dan *Decision Tree* dalam memprediksi risiko penyakit diabetes. Untuk menjalankan penelitian ini secara sistematis, digunakan kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri atas enam tahapan, sebagaimana ditunjukkan pada Gambar 1.1.



Gambar 1. 1 Metode Penelitian

Sumber: Olahan Peneliti (2026)

Berdasarkan gambar tersebut, penelitian ini dilakukan melalui enam tahapan mengikuti kerangka CRISP-DM. Tahap pertama, *Business Understanding*, dilakukan untuk mengidentifikasi masalah penelitian dan menetapkan tujuan penelitian. Tahap kedua, *Data Understanding*, dilakukan dengan mengumpulkan dataset yang terdiri atas 768 data dari repositori publik Kaggle. Tahap ketiga, *Data Preparation*, dilakukan melalui proses *split data* dan *preprocessing* untuk mempersiapkan data sebelum digunakan dalam pemodelan. Tahap keempat, *Modeling*, dilakukan dengan membangun *baseline model* menggunakan algoritma *Naive Bayes* dan *Decision Tree*, yang kemudian dioptimasi menggunakan *K-Fold Cross Validation*. Tahap kelima, *Evaluation*, dilakukan dengan mengomparasikan performa kedua algoritma melalui evaluasi akurasi, sensitivitas, dan spesifisitas untuk memperoleh model terbaik. Tahap terakhir, *Deployment*, dilakukan dengan mengimplementasikan model terbaik ke dalam sistem berbasis web dan melakukan pengujian menggunakan metode *Black Box Testing*.

1.7 Sistematika Penulisan

Sistematika penulisan terdiri dari lima bab dengan uraian sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, metode penelitian, serta sistematika penulisan, termasuk alasan penggunaan algoritma *Naive Bayes* dan *Decision Tree* dalam penelitian.

BAB II LANDASAN TEORI

Bab ini membahas teori pendukung penelitian, meliputi konsep diabetes, *data mining*, *machine learning*, klasifikasi, algoritma *Naive Bayes* dan *Decision Tree*, metode evaluasi model, penelitian terdahulu, serta kerangka pemikiran penelitian.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Bab ini membahas analisis sistem berjalan dan sistem yang diusulkan, tahapan pengumpulan dataset, pembagian data *training* dan *testing*, *preprocessing*, pembangunan dan evaluasi *baseline model*, optimasi menggunakan *K-Fold Cross Validation* dan *GridSearch*, perancangan komparasi algoritma, serta perancangan sistem berbasis *web* meliputi *use case*, basis data, dan desain antarmuka.

BAB IV IMPLEMENTASI DAN ANALISIS HASIL

Bab ini membahas hasil implementasi berdasarkan rancangan pada bab sebelumnya, meliputi hasil pembagian data, *preprocessing*, pengujian *baseline model*, evaluasi model dengan *K-Fold Cross Validation*, analisis perbandingan performa algoritma, implementasi sistem berbasis *web*, serta pengujian *BlackBox Testing*.

BAB V SIMPULAN DAN SARAN

Bab ini berisi simpulan berdasarkan hasil penelitian dan pembahasan, serta saran bagi pengembangan penelitian selanjutnya.