

BAB I

PENDAHULUAN

Latar Belakang

Diabetes mellitus atau yang lebih umum dikenal sebagai diabetes merupakan salah satu penyakit yang menjadi masalah kesehatan di seluruh dunia. Menurut Organisasi Kesehatan Dunia (WHO) penyakit diabetes menyebabkan kematian 4.6 juta jiwa pada tahun 2011 (Aamir, Sarfraz, Ramzan, Bilal, Shafi, and Attique, 2021). Pada 2012, ada 2,2 juta kasus diabetes karena gula darah yang tinggi (Hassan, Malaserene, and Leema, 2021). Menurut federasi diabetes internasional pada tahun 2015, kurang lebih ada 415 juta orang mengidap diabetes, dan diperkirakan akan terjadi peningkatan sekitar 227 juta orang pada tahun 2040 (Fadhillah, Rahma, Sepharni, Mufidah, Sari, and Pangestu, 2022). Deteksi dini dan penanganan diabetes menjadi sangat penting untuk mencegah terjadinya komplikasi yang lebih serius. Oleh karena itu, banyak penelitian dilakukan untuk mengembangkan model klasifikasi diabetes yang dapat memberikan prediksi dengan tingkat akurasi yang tinggi dan cepat.

K-Nearest Neighbor (KNN) dan *Naive Bayes* adalah dua metode klasifikasi yang memiliki performa yang baik dalam klasifikasi data (Rahmaulidyah, F., Hayati, M., & Goejantoro, R., 2021). Algoritma *KNN* adalah salah satu metode klasifikasi berbasis jarak yang memilih label kelas dari tetangga terdekat suatu sampel. Sedangkan algoritma *Naive Bayes* adalah metode yang melakukan perhitungan probabilitas terhadap setiap kelas berdasarkan kemunculan setiap variabel pada data training, *Naive Bayes* juga mempunyai kelebihan yaitu: sederhana, cepat, dan presisi yang tinggi (Syarifah, & Muslim, 2015).

Penelitian terdahulu mengenai algoritma *K-Nearest Neighbor (KNN)* dan *Naive Bayes* dalam klasifikasi telah banyak digunakan diantaranya (Gunawan, Zarlis, & Roslina, 2021) meneliti tentang Analisa Komparasi Algoritma *Naive Bayes* dan *K-Nearest Neighbor* dalam melakukan prediksi kelulusan tepat waktu. Penelitian ini bertujuan untuk menerapkan metode *KNN* dan *Naive Bayes* dalam klasifikasi diabetes dan membandingkan performa kedua metode tersebut. Dataset

yang digunakan dalam penelitian ini berasal dari Dinas Kesehatan Kabupaten Malang yang berjumlah 3811 sampel. Hasil dari penelitian ini diharapkan mampu memberikan kontribusi dalam pengembangan model klasifikasi diabetes dengan menggunakan metode *KNN* dan *Naive Bayes*, serta memberikan masukan bagi praktisi kesehatan untuk memperbaiki pelayanan kesehatan kepada pasien diabetes. Di samping itu, diharapkan bahwa penelitian ini dapat menjadi referensi bagi peneliti lainnya yang tertarik untuk mengembangkan model klasifikasi diabetes menggunakan teknik *machine learning* serta dapat memberikan sumbangsih bagi pengembangan teknologi kesehatan di Indonesia dan memperkaya literatur di bidang kesehatan dan teknologi informasi.

1.1 Rumusan Masalah

Berdasarkan latar belakang yang telah disajikan sebelumnya, rumusan masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana perbandingan performa model *Naive Bayes* dan *K-Nearest Neighbor (KNN)* dalam klasifikasi penyakit diabetes melitus?
2. Bagaimana penerapan metode *Naive bayes dan K-Nearest Neighbor (KNN)* dalam klasifikasi penyakit diabetes melitus?

1.2 Tujuan Penelitian

1. Membandingkan performa kedua metode tersebut dalam klasifikasi diabetes, dengan menggunakan beberapa metrics evaluasi.
2. Dapat mengimplementasikan perbandingan model *Naive Bayes* dan *K-Nearest Neighbor (KNN)*.

1.3 Manfaat Penelitian

Berikut merupakan manfaat penelitian ini bagi beberapa pihak, antara lain:

1. Akademik: Menambah literatur dan pengetahuan di bidang kesehatan dan teknologi informasi, khususnya mengenai aplikasi teknik *machine learning* pada klasifikasi penyakit diabetes.

2. Praktisi Kesehatan: Memberikan alternatif metode dalam melakukan deteksi dini diabetes agar pengobatan lebih efektif dan tepat sasaran, sehingga dapat mempercepat penanganan dan mencegah komplikasi yang lebih parah pada pasien.
3. Masyarakat umum: hasil pada penelitian ini diharapkan mampu memberikan kontribusi yang bermanfaat bagi Masyarakat secara luas dalam pemahaman tentang penyakit diabetes melitus dengan mengetahui faktor faktor yang berpengaruh pada penyakit ini dan dapat mengambil Tindakan pencegahan yang tepat.

1.4 Batasan Masalah

Untuk meminimalisir perluasan materi dalam penelitian, berikut adalah beberapa Batasan masalah yang diberlakukan pada penelitian ini:

1. Penelitian ini hanya akan mengevaluasi performa klasifikasi berdasarkan metrics evaluasi sederhana, yaitu akurasi, presisi, recall, dan f1 score.
2. Penelitian ini tidak membahas aspek pengobatan atau pengendalian diabetes, melainkan hanya fokus pada analisis klasifikasi diabetes menggunakan metode *K-Nearest Neighbor* dan *Naïve Bayes*.
3. Perangkat keras yang digunakan untuk menuliskan perintah dalam bentuk Bahasa pemrograman adalah laptop dengan spesifikasi singkat sebagai berikut:

- a. Processor AMD Ryzen 5 5600H
- b. Memory 16 GB RAM
- c. Graphic Card Nvidia GeForce GTX 1650
- d. OS Windows 11