

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Teknologi yang berkembang dalam dunia medis dapat berupa sistem penganalisaan atau pengdiagnosisan data dari riwayat penyakit pasien terhadap keluhannya. Teknologi ini memegang hal penting dalam pengembangan sistem sistem pemantauan kesehatan masyarakat yang dapat mendeteksi sebuah penyakit secara lebih cepat, tepat dan akurat (Mutiarani, 2023). Implementasi dari penganalisaan dari data riwayat pasien dengan cara memanfaatkan analisis prediksi, klasifikasi dan *clustering* sehingga memungkinkan para ahli kesehatan untuk memantau pola penyakit pasien, mengidentifikasi titik-titik panas (hotspots) pada tubuh pasien dan meramalkan penyebaran penyakit dengan lebih baik.

Penyakit yang sangat rawan terjadi pada manusia adalah penyakit diabetes. Diabetes adalah kondisi medis yang ditandai oleh tingginya kadar glukosa (gula) dalam darah (Kardika & Herawati, 2018). Hal ini terjadi ketika tubuh tidak dapat memproduksi cukup insulin atau tidak dapat menggunakan insulin dengan efektif. Penyakit diabetes juga sulit dideteksi dikarenakan gejala yang terjadi tidak seperti orang yang terkena penyakit. Terdapat beberapa tanda seseorang mengidap penyakit diabetes yaitu hiperglikemia, Hiperglikemia adalah kadar glukosa di dalam darah meningkat melampaui batas normal (Hardianto, 2021).

Berdasarkan data Organisasi Internasional Diabetes Federation tahun 2019, sekurang-kurangnya terdapat 483 juta orang pada usia 20-79 tahun di dunia menderita diabetes. Angka ini setara dengan angka prevalensi sebesar 9.3% dari total penduduk pada usia 20-79 tahun didunia. Faktor-faktor penyakit diabetes dapat berupa tekanan darah tinggi, kadar gula berlebih, berat badan, riwayat keturunan diabetes, usia, jumlah kehamilan seseorang, ketebalan lipatan kulit, jumlah kadar insulin dalam tubuh, kurangnya aktivitas fisik, pola hidup yang tidak sehat serta diet yang tidak teratur (Hardianto, 2021). Metode umum yang digunakan dalam deteksi diabetes dengan menggunakan tes laboratorium seperti glukosa

darah. Kelemahan metode ini adalah metode yang menggunakan waktu yang cukup lama.

Fenomena diabetes mellitus di dunia terkhususnya di India menunjukkan tren yang mengkhawatirkan. Secara global, diabetes telah menjadi salah satu masalah kesehatan utama, dengan jumlah penderita yang terus meningkat. Menurut Organisasi Kesehatan Dunia (WHO), lebih dari 400 juta orang di seluruh dunia menderita diabetes, dan angka ini diperkirakan akan terus meningkat seiring dengan perubahan gaya hidup, urbanisasi, serta peningkatan angka obesitas. Di India, fenomena ini sangat mencolok, karena negara ini tercatat memiliki salah satu populasi penderita diabetes terbesar di dunia, dengan lebih dari 77 juta orang yang terdiagnosis diabetes. Peningkatan prevalensi diabetes di India sebagian besar disebabkan oleh pola makan yang tidak sehat, kurangnya aktivitas fisik, serta faktor genetik. Selain itu, fenomena ini juga semakin banyak ditemukan pada usia yang lebih muda, mencerminkan dampak negatif dari perubahan gaya hidup modern.

Melalui pemanfaatan teknologi, dalam hal ini dapat memanfaatkan sistem pembelajaran mesin. Dengan memanfaatkan sistem pembelajaran mesin, diagnosis yang dilakukan oleh para ahli dapat dilakukan dan dikalkulasikan dengan baik sehingga peluang kemungkinan kesalahan diagnosis medis dapat dihindari serta data medis yang diperoleh dapat diperiksa dalam kurun waktu yang singkat dan mendetail. Selain itu, pemanfaatan analisis teknik data mining dan machine learning dalam mendiagnosa dan memprediksi penyakit dari gejala serta keluhan yang dirasakan menjadi salah satu solusi untuk menganalisa pola dari sebuah data untuk memprediksi penyakit diabetes (Wardhana et al., 2023). Adapun metode *machine learning* yang umum dilakukan pada penelitian yaitu prediksi (*Regresi Linear, KNN, Regresi Logistic, Decision Tree, Random Forest dll*), *clustering* (*K-Means Clustering, K-Medoids Clustering, DBSCAN dll*), klasifikasi (*NBC dll*).

Terdapat beberapa penelitian terkait teknik pembelajaran mesin dalam memprediksi sebuah penyakit. Penelitian yang dilakukan oleh Muhamad Ichsan Gunawan dkk dengan judul penelitian “Peningkatan Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode *Grid Search* pada Algoritma *Logistic Regression*”. Penelitian ini memberikan kesimpulan bahwa Kinerja Akurasi Prediksi Penyakit Diabetes Mellitus Menggunakan Metode *Grid Search*

pada Algoritma *Logistic Regression* pada *Classification Report* memiliki rata-rata akurasi model sekitar 79% dan akurasi data check sebesar 83,33% (Gunawan et al., 2020).

Regresi logistik (*Logistic Regretion*) adalah salah satu metode analisis statistik yang digunakan untuk memodelkan hubungan antara variabel independen (prediktor) dan variabel dependen biner (*outcome*) dengan menggunakan fungsi logistic (Setyati et al., 2020). Terdapat penelitian yang menggabungkan Regresi logistik dan KNN untuk memprediksi sebuah penyakit. Penelitian yang dilakukan oleh I Wayan Trisna Wahyudi dkk dengan judul penelitian “Implementasi *Logistic Regression* dalam Sistem Diagnosa Penyakit Diabetes dengan *KNN*”. Adapun kesimpulan penelitian yang diberikan adalah Prediksi risiko diabetes menggunakan algoritma regresi logistik menggunakan liblinear dengan normalisasi menghasilkan recall sebesar 58% (Wahyudi & Kadyanan, 2023).

Data penelitian yang digunakan pada implementasi penelitian ini adalah database Diabetes Mellitus yang memiliki beberapa paramater. Data yang digunakan merupakan kumpulan data pasien dan data yang digunakan merupakan data Diabetes National Institute of Diabetes and Digestive and Kidney Diseases yang dihimpun pada tahun 2022. Data ini memiliki 768 data dengan jumlah 8 parameter. Pada penelitian ini, data yang digunakan merupakan data yang berhubungan dan berkaitan dengan algoritma machine learning yang digunakan pada penelitian yakni dalam proses prediksi.

Berdasarkan latar belakang masalah dan data penelitian yang digunakan dan yang telah dipaparkan sebelumnya, maka pada penelitian ini mengangkat topik untuk Analisis Perbandingan Prediksi Penyakit Diabetes Menggunakan Algoritma *Logistic Regression* Dan *K-Nearest Neighbor* (Studi Kasus: Original Data National Institute of Diabetes and Disgestive).

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang telah diuraikan, maka penulis mengambil langkah pengindentifikasian masalah yaitu sebagai berikut,

1. Penyakit diabetes melitus (DM) di Dunia khususnya di India merupakan masalah kesehatan yang cukup signifikan. Diabetes adalah kondisi dimana

kadar gula darah (glukosa) dalam tubuh tidak terkontrol dengan baik, yang dapat disebabkan oleh kurangnya produksi insulin oleh pankreas atau ketidakmampuan tubuh untuk menggunakan insulin dengan efektif.

2. Tingkat prevalensi penyakit diabetes di dunia menunjukkan angka yang signifikan, dengan peningkatan yang cukup besar dalam beberapa tahun terakhir. Di dunia, Berdasarkan laporan International Diabetes Federation (IDF) 2023 tingkat persentase penyakit diabetes yaitu prevalensi diabetes global diperkirakan mencapai 11,3% dari populasi orang dewasa dunia, yang setara dengan sekitar 572 juta orang yang menderita diabetes.
3. Penyakit diabetes di Dunia khususnya di India, Prevalensi diabetes di India pada tahun 2023 diperkirakan sekitar 6,9% dari populasi orang dewasa, yang setara dengan sekitar 12 juta orang.
4. Teknologi untuk mengenali dan memantau penyakit diabetes telah berkembang pesat dalam beberapa tahun terakhir. Berbagai teknologi dan alat canggih digunakan untuk membantu mendeteksi diabetes, memantau kadar gula darah, dan mengelola penyakit ini dengan lebih efektif.
5. Teknologi *machine learning* (ML) memiliki potensi besar dalam mengenali dan mendiagnosis penyakit diabetes dari data pasien. Machine learning dapat menganalisis sejumlah besar data medis dan mendeteksi pola yang mungkin tidak terlihat oleh manusia.

1.3 Pembatasan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka penulis mengambil batasan masalah yaitu sebagai berikut:

1. Algoritma machine learning yang digunakan pada penelitian ini adalah algoritma *Logistic Regresion* dan *K-NN*
2. Dataset gejala penyakit yang digunakan pada penelitian ini bersumber pada situs *website kaggle*
3. Dataset yang digunakan merupakan yang bersumber pada Sampel *Data Diabetes National Institute of Diabetes and Digestive and Kidney Diseases*

4. Dataset penelitian memiliki beberapa parameter data diabetes yaitu *Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age, Outcome*

1.4 Perumusan Masalah

Berdasarkan latar belakang yang telah diuraikan diatas maka, bisa dirumuskan permasalahan yang dapat diambil yaitu Bagaimana hasil analisis perbandingan prediksi penyakit jantung menggunakan metode *Logistic Regression* dan *K-Nearest Neighbor* (KNN)?

1.5 Tujuan dan Manfaat Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Untuk mengetahui penerapan algoritma *Logistic Regresion* dalam mendeteksi penyakit diabetes dari data gejala pasien
2. Untuk mengetahui penerapan algoritma K-NN dalam mendeteksi penyakit diabetes dari data gejala pasien
3. Untuk mengetahui analisis perbandingan algoritma *Logistic Regression* dan K-NN dalam mendeteksi penyakit diabetes dari data gejala pasien

Adapun manfaat pada penelitian ini adalah untuk memberikan wawasan kepada pembaca bagaimana analisis perbandingan algoritma *Logistic Regresion* dan K-NN dalam mendeteksi penyakit diabetes dari data gejala pasien serta mengimplementasikan pembelajaran semasa kuliah.