

# BAB I PENDAHULUAN

## I.1 Latar Belakang

Diabetes atau sering disebut sebagai penyakit kencing manis merupakan suatu penyakit akibat kelainan metabolik yang diakibatkan oleh tingginya kadar glukosa darah di tubuh dalam waktu yang lama. Gejala umum yang terjadi dari diabetes adalah *polyphagia* (sering lapar), *polydipsia* (sering haus), *polyuria* (sering berkemih), dan menurunnya berat badan. Jika tidak ditangani lebih awal, diabetes bisa menyebabkan terjadinya komplikasi terhadap penyakit lain (Abdillah & Suwarno, 2014).

Berdasarkan rilis Infodatin Diabetes Melitus 2020 Pusat Data dan Informasi Kementerian Kesehatan RI, *International Diabetes Federation* (IDF) memperkirakan sedikitnya terdapat 463 juta jiwa di seluruh dunia pada usia rentang 20-79 tahun menderita penyakit diabetes pada tahun 2019. Jumlah tersebut setara dengan angka prevalensi sebesar 9,3% dari total penduduk pada usia yang sama. Angka prevalensi penderita diabetes diperkirakan akan selalu meningkat seiring dengan penambahan umur penduduk menjadi 19,9% atau 111,2 juta jiwa pada usia 65-79 tahun. Angka prevalensi tersebut diprediksi akan terus mengalami peningkatan hingga mencapai 578 juta jiwa pada tahun 2030 dan 700 juta jiwa pada tahun 2045 (Pangribo, 2020).

Di wilayah Asia Tenggara di mana Indonesia berada, menempati urutan ke-3 dengan angka prevalensi sebesar 11,3%. Negara Indonesia sendiri berada di urutan ke-7 dari total 10 negara dengan jumlah penderita diabetes terbanyak, yaitu sebesar 10,7 juta. Negara Indonesia menjadi satu-satunya negara di wilayah Asia Tenggara yang berada di daftar tersebut (Pangribo, 2020). Jumlah penderita diabetes di Indonesia diprediksi akan terus mengalami kenaikan dengan jumlah 13,7 juta jiwa pada tahun 2030 dan 16,6 juta jiwa pada tahun 2045 (*"IDF Diabetes Atlas Ninth edition 2019,"* 2019).

Banyak orang terdiagnosis penyakit diabetes setelah mengalami komplikasi. Padahal, jika diagnosis dilakukan secara dini, penanganan diabetes dapat

dilakukan lebih cepat dan dapat menghindari komplikasi penyakit lain yang berbahaya. Dalam perkembangan di dunia medis saat ini, para praktisi dan peneliti memusatkan perhatiannya dalam mendeteksi penyakit diabetes agar dapat mencegah berkembangnya komplikasi penyakit lain (Isbandiyo, 2016).

Diagnosa penyakit di dunia medis merupakan hal yang krusial. Kesalahan dalam diagnosa penyakit yang dilakukan oleh seorang dokter dapat mengakibatkan kesalahan dalam praktik medis dan berujung pada kematian seorang pasien. Hasil analisis sebuah data yang diperoleh dari pasien dan kemampuan dokter dalam mengambil keputusan sangatlah penting dan dapat mempengaruhi hasil diagnosa sebuah penyakit. Dengan hal tersebut, dibutuhkanlah suatu metode yang mampu membantu para ahli medis dalam melakukan analisa pola dari data medis pasien (Putri & Kurniadi, 2018).

Seiring dengan perkembangan teknologi dan ilmu pengetahuan, penggunaan teknologi berbasis data di bidang medis dengan menggunakan pengenalan pola baru-baru ini mendapatkan perhatian yang lebih (Abdillah & Suwarno, 2014). Hal tersebut dapat didukung dengan penggunaan teknik *data mining* dalam menggali informasi berharga dari kumpulan data penyakit diabetes (Isbandiyo, 2016).

Salah satu teknik *data mining* adalah tipe pembelajaran *supervised learning*. Teknik ini digunakan untuk data yang memiliki variabel yang ditargetkan sehingga tujuan dari pendekatan ini adalah menemukan pola baru dalam data dan menghubungkan dengan pola data yang sudah ada (Putra & Akbar, 2016). Salah satu jenis dari *supervised learning* adalah klasifikasi. Klasifikasi digunakan untuk menentukan keputusan sesuai dengan pola baru yang didapat dari pola data lama dengan menggunakan perhitungan algoritma (Indrayanti dkk., 2017). Pada penelitian ini, penulis menggunakan algoritma *K-Nearest Neighbor* dan *Logistic Regression*. Dari kedua algoritma tersebut, penulis ingin mengukur performa terbaik untuk klasifikasi *dataset* yang dihasilkan dengan membandingkan akurasi dari algoritma *K-Nearest Neighbor* dan *Logistic Regression*.

*Dataset* yang digunakan pada penelitian ini adalah *dataset* yang berasal dari *National Institute of Diabetes and Digestive and Kidney Diseases* dan dapat diakses di *UCI Machine Learning Repository: Pima Indians Diabetes Database*. *Dataset* ini banyak digunakan oleh peneliti untuk menguji klasifikasi penderita diabetes. *Dataset* ini berisikan 768 pasien wanita dengan 8 atribut diagnosa kondisi medis yang berbeda dan 1 atribut tujuan atau atribut label (Putri & Kurniadi, 2018).

Beberapa penelitian terkait klasifikasi penyakit diabetes dengan menggunakan *dataset Pima Indians Diabetes Database* telah dilakukan sebelumnya. Penelitian serupa berjudul "*Machine Learning Untuk Perbandingan Tingkat Akurasi Prediksi Penyakit Diabetes Metode Supervised learning*" yang ditulis oleh Ali Murtadho (Murtadho, 2020). Penelitian ini membandingkan akurasi dari algoritma *Gradient Boosting*, *K-Nearest Neighbor (KNN)*, *Decision Tree*, *Logistic Regression*, dan *Random Forest*. Dengan hasil akurasi algoritma *K-Nearest Neighbor* mendapatkan nilai 64%. Pada algoritma *Logistic Regression* mendapatkan nilai akurasi sebesar 70%. Penelitian serupa juga sudah pernah dilakukan dengan judul "*Diabetes Prediction Using Machine Learning Classification Algorithms*" yang ditulis oleh Shamriz Nahzat dan Mete Yağanoğlu (Nahzat & Yağanoğlu, 2021). Penelitian ini membandingkan akurasi dari algoritma *K-Nearest Neighbor*, *Random Forest*, *Support Vector Machine*, *Artificial Neural Network*, dan *Decision Tree*. *Dataset* yang digunakan sebelumnya sudah melalui tahap perubahan nilai 0 menjadi nilai *median* dari tiap kolomnya. Hasil akurasi yang diperoleh dari algoritma *K-Nearest Neighbor* sebesar 82%. Penelitian lain yang serupa berjudul "*Klasifikasi Diabetes Menggunakan Model Pembelajaran Ensemble Blending*" ditulis oleh Vinnia Kemala Putri dan Felix Indra Kurniadi menjelaskan tentang kombinasi dari berbagai algoritma dengan metode *ensemble blending* (Putri & Kurniadi, 2018). Dalam pemilihan algoritma, penelitian ini membandingkan beberapa algoritma seperti algoritma *Decision Tree*, *K-Nearest Neighbor*, *Logistic Regression*, dan SVM. Dengan hasil akurasi algoritma *K-Nearest Neighbor* sebesar 73,33% dan algoritma *Logistic Regression* sebesar 75,33%.

Dari beberapa penelitian terkait yang sudah disebutkan, terdapat perbedaan nilai akurasi dari kedua algoritma yang digunakan dari semua penelitian tersebut. Dikarenakan belum ada penelitian yang secara spesifik membandingkan akurasi dari algoritma *K-Nearest Neighbor* dan *Logistic Regression* untuk klasifikasi penyakit diabetes, dalam penelitian ini perlu dilakukan perbandingan algoritma *K-Nearest Neighbor* dan *Logistic Regression* untuk mengetahui tingkat akurasi dalam pengklasifikasian penyakit diabetes.

## **I.2 Perumusan Masalah**

Berdasarkan latar belakang di atas, maka rumusan permasalahan ini diturunkan ke sub-permasalahan yaitu:

- a. Bagaimana hasil akurasi klasifikasi penyakit diabetes pada algoritma *K-Nearest Neighbor*?
- b. Bagaimana hasil akurasi klasifikasi penyakit diabetes pada algoritma *Logistic Regression*?
- c. Bagaimana hasil perbandingan tingkat akurasi klasifikasi penyakit diabetes menggunakan algoritma *K-Nearest Neighbor* dan *Logistic Regression*?

## **I.3 Tujuan Penelitian**

Penelitian ini bertujuan untuk:

- a. Mendapatkan nilai akurasi terbaik klasifikasi penyakit diabetes pada algoritma *K-Nearest Neighbor*.
- b. Mendapatkan nilai akurasi terbaik klasifikasi penyakit diabetes pada algoritma *Logistic Regression*.
- c. Mengetahui perbandingan tingkat akurasi klasifikasi penyakit diabetes menggunakan algoritma *K-Nearest Neighbor* dan *Logistic Regression*.

## **I.4 Batasan Penelitian**

Batasan untuk melakukan penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya dilakukan pada data *Pima Indians Diabetes Database* yang diambil dari situs [atapdata.ai](http://atapdata.ai)
2. Hasil akhir dari penelitian ini berupa perbandingan nilai akurasi dari kedua algoritma.

## **I.5 Manfaat Penelitian**

Melalui penelitian ini diharapkan dapat memberikan manfaat penelitian sebagai berikut:

Manfaat Teoritis:

1. Mengetahui ketepatan hasil akurasi pada algoritma untuk klasifikasi sebuah data.
2. Memahami secara mendalam algoritma yang digunakan dalam klasifikasi sebuah data.

Manfaat Akademis:

1. Sebagai bahan studi dan tambahan ilmu pengetahuan dalam hal klasifikasi penyakit diabetes bagi mahasiswa Fakultas Rekayasa Industri terutama Jurusan Sistem Informasi.
2. Sebagai rujukan mengenai penerapan algoritma yang lebih baik untuk digunakan pada penelitian yang akan datang untuk *dataset* penyakit diabetes.

Manfaat Praktis

Dapat menghasilkan model klasifikasi untuk membuat model prediksi. Model prediksi tersebut dapat digunakan untuk membantu instansi kesehatan dalam mendeteksi dini penyakit diabetes.

## **I.6 Sistematika Penulisan**

Penelitian ini diuraikan dengan sistematika penulisan sebagai berikut:

### **Bab I      Pendahuluan**

Bab ini berisi penjelasan mengenai konteks permasalahan, latar belakang penelitian, perumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, dan sistematika penulisan.

### **Bab II     Tinjauan Pustaka**

Bab ini berisi literatur yang berkaitan dengan permasalahan yang diteliti. Bab ini juga berisi uraian alasan pemilihan kerangka kerja penelitian.

**Bab III Metodologi Penelitian**

Bab ini berisi penjelasan langkah-langkah penelitian secara rinci meliputi model konseptual dan sistematika penyelesaian masalah.

**Analisis dan Perancangan**

**Bab IV**

Bab ini berisi tentang tahapan pengumpulan data dan tahapan *data preprocessing*. Keluaran dari tahap ini berupa data yang digunakan dalam penelitian.

**Bab V Hasil dan Evaluasi**

Bab ini berisi uraian tahap implementasi terhadap penelitian. Pada bab ini akan dilakukan pengujian hasil hingga evaluasi model.

**Bab VI Kesimpulan dan Saran**

Bab ini berisi kesimpulan dari hasil penelitian yang dilakukan. Pada bab ini juga berisi saran yang dapat digunakan untuk pengembangan penelitian selanjutnya.