

BAB I PENDAHULUAN

I.1 Latar Belakang

Diabetes merupakan penyakit yang terus meningkat dan semakin tinggi kasus kematian yang memakan korban, Penyakit kronis serius ini disebabkan oleh gangguan metabolik yang terjadi karena pankreas tidak dapat menghasilkan atau memproduksi cukup *Insulin* (hormon yang mengatur glukosa) atau dimana ketika tubuh tidak dapat secara efektif menggunakan *Insulin* yang diproduksi, sehingga menghasilkan kadar gula darah tinggi atau disebut hiperglikemia (Yuniastuti et al., 2018).

Diabetes adalah penyakit tidak menular (PTM) merupakan penyebab utama kematian dan kecacatan di dunia, menjadi penyebab masalah kesehatan masyarakat yang sangat penting dan menjadi salah satu dari empat prioritas penyakit tidak menular untuk ditindaklanjuti oleh pemerintah di seluruh dunia (World Health Organization, 2019). Pada tabel I.1 menunjukkan kasus diabetes yang terjadi pada tahun 2019 dengan rentang usia 20-79 tahun (*“IDF Diabetes Atlas Ninth edition 2019,”* 2019)..

Tabel I.1 Estimasi Prevalensi dan Jumlah Penderita Diabetes

<i>Rank</i>	<i>IDF Region World</i>	<i>Number of people with diabetes, Million</i>	<i>World-Age standardised diabetes prevalence, %</i>
	World	463.0 (368.7–600.6)	8.3 (6.2–11.8)
1	MENA	54.8 (30.7–75.1)	12.2 (8.3–16.1)
2	WP	162.6 (146.6–203.0)	11.4 (8.3–15.6)
3	SEA	87.6 (70.9–110.9)	11.3 (8.0–15.9)
4	NAC	47.6 (37.4–56.4)	11.1 (9.0–14.5)
5	SACA	31.6 (26.3–39.2)	8.5 (6.7–11.3)
6	EUR	59.3 (46.3–80.2)	6.3 (4.9–9.2)
7	AFR	19.4 (10.6–35.8)	4.7 (3.2–8.1)

Pada tabel I.1 dapat dilihat bahwa, secara global diperkirakan 463 juta orang dewasa mengidap diabetes pada tahun 2019, Menurut *World Health Organization* (WHO) jumlah ini meningkat drastis dibandingkan pada tahun 1980 dengan 108 juta orang dewasa mengidap diabetes dan juga pada tahun 2014 sebanyak 422 juta. Prevalensi diabetes di dunia secara global telah meningkat sejak tahun 1980. Hal

ini mencerminkan peningkatan faktor risiko terkait seperti kelebihan berat badan atau obesitas. Selama beberapa dekade terakhir, prevalensi diabetes meningkat lebih cepat di negara berpenghasilan rendah dan menengah daripada di negara berpenghasilan tinggi.

Identifikasi penyakit diabetes dapat dilakukan dengan melakukan pengklasifikasian untuk membantu dalam penanggulangan penyakit diabetes, maka langkah untuk melakukan analisis dan mencari solusi yang nantinya dapat membantu untuk menurunkan angka tingkat diabetes yang semakin tinggi dengan dilakukannya klasifikasi dalam *Machine learning*, untuk melakukan klasifikasi dapat dilakukan dengan teknik *Data Mining* (Isbandiyo, 2013). Data mining merupakan gabungan dari berbagai disiplin ilmu seperti statistik, matematika, pengenalan pola, gudang data (data warehouse), kecerdasan buatan, dan visualisasi data (data visualization) (Han & Kamber, 2006). Saat ini banyak algoritma-algoritma yang dapat digunakan untuk melakukan metode *Machine learning*. *Data Mining* merupakan suatu proses pengumpulan informasi penting dari sebuah data yang besar pada suatu keahlian yang berkaitan dengan informatika. *Data Mining* juga dapat digunakan pada penelitian yang bergerak di aspek lainnya, salah satunya pada bagian kesehatan untuk melakukan prediksi penyakit Diabetes pada suatu kelompok individu dengan teknik klasifikasi. Pada penelitian kali ini peneliti ingin membandingkan algoritma klasifikasi yang memiliki akurasi paling tinggi khususnya dalam pemodelan *Supervised Learning* dengan menggunakan algoritma klasifikasi dalam *Machine learning*.

Pada penelitian sebelumnya dilakukan klasifikasi menggunakan subjek yang sama yaitu pada data diabetes, Ali Murtadho pada tahun 2020 melakukan penerapan model yang berjudul "*Machine learning* Untuk Perbandingan Tingkat Akurasi Prediksi Penyakit Diabetes Metode *Supervised learning*". Penelitian ini melakukan komparasi akurasi pada algoritma *Gradient Boosting*, *K-Nearest Neighbor* (KNN), *Decision Tree*, *Logistic Regression*, dan *Random Forest*. Hasil penelitian menunjukkan klasifikasi menggunakan *Gradient Boosting* mendapatkan nilai terbaik yaitu sebesar 81.3%, untuk algoritma *K-Nearest Neighbor* mendapatkan nilai 64%, *Decision Tree* mendapatkan nilai akurasi sebesar 72%, *Logistic Regression* mendapatkan nilai akurasi sebesar 70%,

sementara *Random Forest* mendapatkan nilai akurasi sebesar 72%. (Murtadho, 2020).

Pada tahun 2017, Deepti Sisodia dan koleganya melakukan penelitian dengan melakukan perbandingan algoritma klasifikasi untuk prediksi diabetes pada *Pima Indians Diabetes Database* (PIDD). Dari ketiga algoritma yang telah dibandingkan terdapat nilai hasil akurasinya dimana Algoritma Naive Bayes sebesar 76.30%, Support Vector Machine sebesar 65.10% dan *Decision Tree* sebesar 73.82% (Sisodia & Sisodia, 2018).

Berdasarkan dengan permasalahan dan penelitian sebelumnya, maka pada penelitian ini akan dibangun model pada klasifikasi penyakit diabetes. Dataset yang digunakan pada penelitian ini adalah *dataset Pima Indian Diabetes Database* yang dimana di dalam *dataset* ini terdapat macam-macam atribut dan atribut tersebut juga menjadi beberapa faktor sebagai penyebab tingginya jumlah penyandang diabetes. Untuk klasifikasi data penyakit diabetes, dan *gender* yang terdapat pada dataset ini adalah wanita. Menurut *International Diabetes Federation* (IDF) pada tahun 2017 wanita lebih berisiko terkena diabetes ketimbang pria. Faktor wanita memiliki resiko terkena penyakit yang lebih tinggi secara internal adalah *Insulin* resistance atau resistensi *Insulin* dimana wanita mempunyai satu komponen resistensi *Insulin* yang akan meningkat ketika hamil. Oleh karena itu diharapkan dengan adanya penelitian ini dapat diketahui tingkat akurasi dalam melakukan klasifikasi permasalahan pada data penyakit diabetes tersebut.

I.2 Perumusan Masalah

Berdasarkan latar belakang yang ada, maka terdapat beberapa permasalahan yang dapat dikaji di antaranya :

- a. Bagaimana hasil akurasi klasifikasi penyakit diabetes pada Algoritma *Naïve Bayes*?
- b. Bagaimana hasil akurasi klasifikasi penyakit diabetes pada Algoritma *XGBoost*?
- c. Bagaimana hasil perbandingan tingkat akurasi klasifikasi penyakit diabetes menggunakan algoritma *Naïve Bayes* dan *XGBoost*?

I.3 Tujuan Penelitian

Penelitian ini bertujuan untuk:

- a. Mendapatkan nilai akurasi terbaik klasifikasi penyakit diabetes pada Algoritma *Naïve Bayes*.
- b. Mendapatkan nilai akurasi terbaik klasifikasi penyakit diabetes pada Algoritma *XGBoost*.
- c. Menghasilkan perbandingan tingkat akurasi klasifikasi penyakit diabetes menggunakan Algoritma Naive Bayes dan *XGBoost*.

I.4 Batasan Penelitian

Dari hasil penjabaran di atas, batasan masalah dari penelitian ini yaitu:

1. Penelitian ini hanya dilakukan pada data *Pima Indians Diabetes Database* yang diambil dari situs atapdata.ai.
2. Hasil akhir dari penelitian ini berupa perbandingan nilai akurasi dari kedua algoritma tersebut.

I.5 Manfaat Penelitian

Adapun manfaat yang didapatkan dari penelitian ini adalah sebagai berikut.

Bagi Peneliti:

1. Mengetahui tingkat akurasi masing-masing algoritma klasifikasi dalam mengklasifikasikan penyakit diabetes.
2. Memahami penerapan masing-masing algoritma dalam mengklasifikasikan penyakit diabetes.
3. Memahami penerapan Algoritma *Supervised Learning*.

Bagi Peneliti lainnya :

1. Sebagai bahan studi dan rujukan dalam hal klasifikasi penyakit diabetes atau sebagai dokumen bagi mahasiswa Fakultas Rekayasa Industri terutama Jurusan Sistem Informasi.
2. Sebagai rujukan mengenai algoritma mana yang lebih efisien digunakan pada penelitian yang akan datang untuk dataset penyakit diabetes.

I.6 Sistematika Penulisan

Penelitian ini diuraikan dengan sistematika penulisan sebagai berikut:

Bab I Pendahuluan

Pada bab ini berisi uraian mengenai konteks permasalahan, latar belakang penelitian, perumusan masalah, tujuan penelitian, batasan penelitian, manfaat penelitian, dan sistematika penelitian.

Bab II Tinjauan Pustaka

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

Bab III Metodologi Penelitian

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

Bab V Implementasi Dan Pengujian

Pada bab ini berisi uraian tahap implementasi terhadap penelitian. Bab ini juga akan dilakukan pengujian hasil hingga evaluasi model. maka penamaan bab ini membahas tentang tahapan implementasi Algoritma *Naïve Bayes* dan Algoritma *XGBoost*

Bab VI Kesimpulan dan Saran

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