

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes Mellitus merupakan salah satu masalah kesehatan yang cenderung meningkat dari tahun ke tahun. Organisasi Kesehatan Dunia (*World Health Organization/WHO*) memperkirakan, bahwa 177 juta penduduk dunia mengidap diabetes. Jumlah ini akan meningkat hingga melebihi 300 juta pada tahun 2025 (Rifqi, 2016). Dr. Paul Zimmet, Direktur International Diabetes Institute (IDI) di Victoria, Australia meramalkan bahwa diabetes akan menjadi epidemi yang paling dahsyat dalam sejarah manusia (Harliani, 2018). Tetapi data epidemiologi di negara berkembang masih belum banyak. Oleh karena itu angka prevalensi yang dapat ditelusuri mayoritas berasal dari negara maju.

Dengan bertambahnya angka harapan hidup di negara Indonesia, perhatian masalah kesehatan beralih dari penyakit infeksi ke penyakit degeneratif. Selain penyakit jantung koroner dan hipertensi, diabetes merupakan salah satu penyakit degeneratif yang saat ini makin bertambah jumlahnya di Indonesia (Rifqi, 2016). Di Indonesia diperkirakan tahun 2030 prevalensi DM mencapai 21,3 juta penduduk. Hasil riset kesehatan dasar (riskesda) diperoleh proporsi penyebab kematian Akibat DM pada Kelompok usia 45-55 tahun didaerah perkotaan menduduki peringkat ke 2 dan daerah pedesaan menduduki peringkat ke 6. Secara umum sekitar 80% prevalensi DM adalah DM Tipe II. Ini berarti gaya hidup tidak sehat menjadi pemicu (Dirjen Pengendalian Penyakit dan Penyehatan Lingkungan. 2009).

Beberapa peneliti dari WHO membagi diabetes menjadi diabetes tipe 1, tipe 2, dan diabetes gestasional (World Health Organization, 2016). Diabetes gestasional adalah jenis diabetes yang hanya terjadi pada kehamilan karena perubahan hormonal. Gejala umum diabetes adalah poliuria, polidipsia, polifagia, penurunan berat badan mendadak (biasanya tipe 1), kelemahan, obesitas (biasanya tipe 2), penyembuhan tertunda, penglihatan kabur, gatal, iritabilitas, sariawan genital, paresis parsial, otot kaku, alopecia, dll. (*American Diabetes Association*, 2016).

Seiring perkembangan ilmu pengetahuan dan teknologi, terdapat berbagai teknologi yang dapat membantu mendeteksi penyakit secara akurat (Ridwan, 2020). Salah satunya yaitu penggunaan data mining, dengan acuan pada analisis data untuk mendiagnosis awal diabetes dengan kecanggihan teknologi berbasis komputer secara akurat yang efisien dan efektifitas waktu dan biaya. Data mining adalah sebuah ilmu yang mempelajari alur kerja pengalihan data atributnya yang saling berkaitan untuk menemukan pengetahuan atau menemukan sebuah pola dari suatu data yang besar (Aris Faizal & Benyamin, 2019). Proses data mining bertujuan untuk menggali, menganalisis, mengekstrak informasi dari kumpulan data yang besar dan mengubahnya menjadi struktur yang dimengerti untuk digunakan sebagai sebuah informasi atau pengetahuan yang berguna (Marzuki et al., n.d.).

Dalam data mining terdapat beberapa metode, salah satunya yaitu klasifikasi. Klasifikasi digunakan dalam menentukan hasil dari dataset yang didapat ke dalam suatu *class*. Tujuan dari klasifikasi yaitu untuk memprediksi

class yang dicari secara akurat pada setiap kasus dalam data. Pemilihan metode klasifikasi pada penelitian ini sangat cocok dan efektif dibandingkan menggunakan metode klusterisasi, karena jenis data yang digunakan pada penelitian ini termasuk ke dalam jenis *supervised learning* (jenis data yang sudah memiliki label atau kelas). Sedangkan jenis data yang digunakan pada metode klusterisasi termasuk ke dalam jenis *unsupervised learning* (jenis data yang belum memiliki label atau kelas tertentu). Oleh karena itu, penggunaan metode klasifikasi ini menjadi pilihan yang sangat tepat untuk digunakan karena memiliki kecocokan dengan jenis *dataset* yang akan digunakan pada penelitian ini.

Guna mendapatkan klasifikasi tentang diabetes mellitus dan penelitian tentang klasifikasi dengan kasus ini, maka pada penelitian ini mencoba mengaitkan kasus dengan menggunakan algoritma Naïve Bayes. Dengan algoritma Naïve Bayes ini mampu untuk menjawab permasalahan-permasalahan di bidang probabilitas dan statistik yang akan diterapkan untuk menjawab persoalan keterjangkitan diabetes mellitus. Oleh karena itu, penelitian ini dilakukan dengan menggunakan algoritma Naïve Bayes untuk pengklasifikasian penyakit diabetes mellitus.

Penelitian (Ridwan, 2020) menggunakan algoritma *Naïve Bayes* untuk proses klasifikasi penyakit diabetes mellitus yang menghasilkan akurasi sebesar 90,20%. Kemudian, penelitian oleh (Nurdiana & Algifari, 2020) membandingkan algoritma *Naïve Bayes* dengan algoritma *ID3*. Hasil yang didapatkan yaitu akurasi algoritma Naïve Bayes lebih unggul dibandingkan dengan algoritma *ID3*. Nilai akurasi dari algoritma *Naïve Bayes* yaitu sebesar 76%%, sedangkan nilai akurasi

dari algoritma *ID3* yaitu sebesar 74%%. Kemudian, penelitian (Putry & Sari, 2022) membandingkan kinerja algoritma K-Nearest Neighbor (KNN) dengan Naïve Bayes. Hasil yang didapat setelah melakukan pengujian kinerja kedua algoritma tersebut bahwa algoritma Naïve Bayes mampu mendapatkan tingkat akurasi yang paling tinggi dibandingkan algoritma KNN. Nilai akurasi dari Naïve Bayes tersebut sebesar 80%.

Berdasarkan latar belakang permasalahan yang terjadi, maka penelitian Tugas Akhir ini berjudul “PENERAPAN DATA MINING PADA KLASIFIKASI PENYAKIT DIABETES MELLITUS MENGGUNAKAN ALGORITMA *NAÏVE BAYES*”

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dipaparkan sebelumnya, maka dapat dibuat rumusan masalah penelitian sebagai berikut.

1. Apakah algoritma *Naïve Bayes* mampu menghasilkan klasifikasi yang sangat baik pada klasifikasi penyakit diabetes mellitus?
2. Bagaimana kinerja algoritma *Naïve Bayes* pada proses klasifikasi penyakit diabetes mellitus?

1.3 Batasan Masalah

Adapun beberapa batasan masalah pada penelitian ini yaitu:

1. Klasifikasi hanya menggunakan algoritma *Naïve Bayes*;
2. Data yang digunakan dalam penelitian ini adalah data pasien poliklinik penyakit dalam tahun 2019 dan 2020 RSUD Bayu Asih Purwakarta;

3. Pengklasifikasian dibagi pada kategori positif dan negatif;
4. *Tools* yang digunakan pada penelitian ini yaitu *RapidMiner*.

1.4 Tujuan Penelitian

Secara umum, penelitian ini memiliki tujuan umum dan khusus. Tujuan umum penelitian ini yaitu untuk mengklasifikasikan data penderita diabetes mellitus. Adapun tujuan khusus penelitian, yaitu:

1. Untuk mengetahui apakah algoritma *Naïve Bayes* mampu menghasilkan klasifikasi yang sangat baik pada klasifikasi penyakit diabetes mellitus.
2. Mengukur kinerja algoritma *Naïve Bayes* untuk penerapan data mining pada hasil klasifikasi penyakit diabetes mellitus.

1.5 Manfaat Penelitian

Penelitian ini mempunyai manfaat sebagai berikut:

1. Membantu mengolah data secara otomatis pada data pasien penyakit diabetes mellitus RSUD Bayu Asih Purwakarta;
2. Menampilkan hasil kinerja dari algoritma *Naive Bayes* pada proses pengklasifikasian penyakit diabetes mellitus.

1.6 Metodologi Penelitian

Langkah – langkah metodologi penelitian yang dilakukan dalam penelitian ini adalah sebagai berikut :

1. Studi Pustaka

Studi pustaka dilakukan untuk mencari referensi dan jurnal, buku dan media lain dalam mendapatkan teori-teori tentang permasalahan yang sudah ditetapkan.

2. Pengumpulan data

Pada penelitian ini data yang digunakan adalah data pasien poliklinik penyakit dalam tahun 2019 dan 2020 RSUD Bayu Asih Purwakarta

3. *Data Preprocessing*

a. *Data Cleaning*

Proses ini untuk dapat mengatasi nilai yang hilang, noise dan data yang tidak konsisten.

b. *Data Intergration*

Proses ini menggabungkan data dari banyak sumber data. Proses ini juga dapat membantu mengurangi redundansi data dan data yang tidak konsisten yang disebabkan pengambilan data dari banyak sumber.

c. *Data Selection*

Proses ini dilakukan untuk menyeleksi jumlah data yang akan digunakan pada proses mining dengan tetap merepresentasikan data aslinya

4. Pengujian data menggunakan *Naïve Bayes*

Setelah dilakukannya tahapan *data preprocessing*, selanjutnya data akan di uji dengan menggunakan algoritma *Niave Bayes*. Langkah-langkah pengujian data pada algoritma *Naïve Bayes* :

a. Dataset

Merupakan data mentah sebelum dilakukannya tahapan *preprocessing data*

b. *Data preprocessing*

Pada tahapan ini proses yang dilakukan yaitu *data cleaning, data integration, data selection*

c. Hasil (akurasi)

Akurasi yang didapat dari kinerja algoritma *Naïve Bayes*

5. Hasil dan Evaluasi pengujian data

Data yg telah di uji kemudian akan di analisis menggunakan metode *confusion matrix* dan kurva ROC

1.7 Sistematika Penulisan

Sistematika dalam penulisan ini disusun untuk mempermudah pembacaan dan gambaran umum tentang penelitian yang akan dilakukan dengan sistematika penulisan penelitian ini sebagai berikut :

BAB I PENDAHULUAN

BAB I membahas tentang latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian yang ingin dicapai, manfaat penelitian, metodologi penelitian yang merupakan langkah untuk menyelesaikan masalah, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab yang membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan penelitian yaitu teori mengenai data *mining*, diabetes mellitus, metode klasifikasi, algoritma *Naïve Bayes*, dan *tools RapidMiner*

BAB III METODOLOGI

Bab ini menjelaskan kebutuhan perangkat lunak yang akan dibangun, dengan berbagai analisis yang akan dijabarkan. Hasil dari analisis tersebut digunakan untuk melakukan perancangan perangkat lunak.

BAB IV HASIL DAN PEMBAHASAN

Merupakan hasil dari proses yang telah dilakukan dari penelitian ini. Lebih tepatnya adalah membahas proses penelitian tersebut disertai pengujian untuk mengetahui apakah penelitian tersebut berhasil atau gagal.

BAB V KESIMPULAN DAN SARAN

Bab yang membahas tentang penarikan kesimpulan dari hasil penelitian yang telah dilakukan, lebih membahas tentang kesimpulan keterkaitan masalah dan tujuan penelitian yang dilakukan. Saran yang bersifat membangun untuk penelitian selanjutnya tentang penelitian ini agar dapat menutupi kekurangan-kekurangan yang ada.