

1. Pendahuluan

1.1 Latar Belakang

Penyakit diabetes atau dalam Bahasa Indonesia disebut kencing manis adalah gangguan metabolisme yang ditandai dengan gula dalam darah yang tinggi. Penyakit diabetes melitus dapat menyebabkan komplikasi dan meningkatkan resiko penyakit lain diantaranya adalah penyakit ginjal, serangan jantung, kebutaan pada usia dibawah 65 tahun, kerusakan saraf, dan bahkan kecacatan atau amputasi. Pada ibu hamil, diabetes melitus dapat meningkatkan resiko kematian janin [1].

Menurut *International Diabetes Federation*, pada 2019 sebanyak 463 juta orang di seluruh dunia mengidap penyakit ini. Jumlah ini akan meningkat menjadi 700 juta orang pada 2045 [2]. Di Indonesia sendiri, penyakit ini merupakan penyebab utama kematian tertinggi [3]. Dengan perkiraan meningkatnya penderita penyakit diabetes melitus, maka diperlukan teknologi yang dapat membantu menanggulangi penyakit ini dengan menggunakan data yang banyak untuk mendiagnosis penyakit diabetes melitus.

Data mining merupakan kumpulan dari beberapa disiplin ilmu komputer yang didefinisikan sebagai proses *iterative* untuk menemukan suatu informasi yang bermanfaat melalui metode *automatic* ataupun manual [4]. Dalam bidang kesehatan, *data mining* dapat digunakan untuk mengklasifikasi dan memprediksi sebuah penyakit, salah satunya adalah penyakit diabetes melitus. Selain itu, melalui penerapan *data mining*, kita dapat membangun sistem klasifikasi yang dapat menghasilkan informasi dari banyaknya data yang digunakan.

Dalam penelitian sebelumnya, dilakukan klasifikasi data untuk penyakit diabetes melitus menggunakan beberapa metode dalam *data mining* seperti *Binary Logistic Regression*, *Multilayer Perceptron*, dan *k-Nearest Neighbor*. Dalam penelitian ini, *dataset* memiliki 100 baris data dengan atribut sebanyak 7. Hasil akurasi terbaik penelitian ini diperoleh dengan metode *k-Nearest Neighbor*, dengan akurasi, *sensitivity*, dan *specificity* masing-masing adalah 80%, 82,5% dan 78,3% [5].

Dengan ini, penulis tertarik untuk membangun penelitian klasifikasi data menggunakan metode *k-Nearest Neighbor* melalui *dataset* berbeda. Metode *k-Nearest Neighbor* dipilih karena banyak digunakan pada proses klasifikasi data serta algoritmanya yang sederhana untuk diterapkan tetapi efektif ketika menghadapi *noisy data* [4]. Selain itu, dilakukan analisis terhadap performansi pada data yang diterapkan *feature selection* melalui pendekatan *embedded* dengan performansi pada data yang tidak diterapkan *feature selection*. Proses *feature selection* digunakan untuk melihat fitur atau atribut yang relevan untuk digunakan dalam mendiagnosis penyakit diabetes melitus. Alasan dari *feature selection* melalui pendekatan *embedded* dipilih adalah karena sangat tepat untuk *dataset* yang tidak terlalu besar. Pendekatan *embedded* juga akan menangani setiap iterasi dari *data training* dan mengekstraksi fitur-fitur yang berkontribusi paling besar untuk diterapkan pada proses klasifikasi. Penulis berharap untuk menemukan informasi dan performansi terbaik melalui penerapan yang berbeda dari penelitian sebelumnya.

1.2 Topik dan Batasannya

Topik yang dibahas dalam penelitian ini diantaranya melihat pengaruh penggunaan *feature selection* melalui pendekatan *embedded* terhadap nilai evaluasi performansi serta menentukan performansi terbaik dari sistem klasifikasi yang sudah dibangun menggunakan *k-Nearest Neighbor* pada data yang diterapkan *feature selection* dan yang tidak diterapkan *feature selection*.

Adapun *dataset* yang digunakan bersumber dari “*Dataset Gula Karya Medika*” yang mempunyai 470 baris data dan jumlah atribut sebanyak 6. Atribut tersebut diantaranya adalah *Glucose* (Kadar Glukosa), *Gender* (Jenis Kelamin), *Blood Pressure* (Tekanan Darah), BMI (Indeks Masa Tubuh), Usia, dan *Class* (Positif Diabetes atau Negatif Diabetes). Nilai *k* yang digunakan berkisar dari 1, 3, 5, 7, dan 9.

1.3 Tujuan

Tujuan yang akan dicapai diantaranya adalah untuk menganalisis dan membandingkan hasil performansi terhadap data yang diterapkan *feature selection* dengan data yang tidak diterapkan *feature selection*, serta mendapatkan nilai performansi yang baik dan diharapkan dapat menemukan fitur atau atribut terbaik untuk diterapkan pada proses klasifikasi data.

1.4 Organisasi Tulisan

Bab 2 menjelaskan terkait teori atau studi literatur yang mendukung terhadap penelitian yang dikerjakan. Bab 3 menjelaskan desain serta sistem/produk yang dihasilkan. Bab 4 membahas hasil pengujian beserta dengan analisisnya. Dan bab 5 menjelaskan kesimpulan untuk seluruh proses penelitian beserta dengan saran.