

## 1. Pendahuluan

### Latar Belakang

Diabetes termasuk ke dalam jenis/kelompok penyakit metabolik (metabolic disorder) yaitu suatu jenis penyakit yang ditandai dengan kenaikan kadar gula dalam darah. Penyakit metabolic seperti diabetes adalah penyakit genetik yang biasanya disebabkan oleh penyakit keturunan, lalu penyebab lainnya bisa dari racun makanan dan infeksi. Penyakit mempengaruhi destruksi sel beta, diabetagonik, dan olahraga yang kurang. Biasanya penyakit ini menyebabkan komplikasi kronis yang terjadi pada mata, ginjal, syaraf, dan pembuluh darah[1].

Menurut laporan The International Diabetes Federation, kenaikan jumlah penduduk yang terjangkit diabetes dari tahun 2003 hingga 2025 akan terus meningkat. Indonesia sendiri terdapat pada peringkat ke-5 dengan jumlah penderita diabetes sebesar 12.9 juta jiwa dan diperkirakan akan terus naik hingga peringkat ke-3 dunia pada tahun 2025[20].

Terdapat banyak metode klasifikasi pada supervised learning pada machine learning, diantaranya adalah K-Nearest Neighbor, Naïve Bayes, Support Vector Machine, Neural Network, Random Forest Classifier, Ada Boost Classifier, serta Quadratic Discriminant Analysis[10].

Dalam penelitian ini penulis menggunakan metode KNN dan Naïve Bayes untuk proses klasifikasinya, metode pengklasifikasian menggunakan KNN dan juga Naïve Bayes merupakan proses metode yang akan dikembangkan kedepannya untuk pengklasifikasian penyakit diabetes. Metode KNN memiliki beberapa kelebihan yaitu, pelatihan sangat cepat, sederhana, efektif pada data pelatihan yang besar. Sedangkan kekurangan KNN sendiri adalah nilai k bias harus ditentukan, komputasi kompleks, keterbatasan memori, dan dapat tertipu oleh atribut yang tidak relevan[3]. Adapun metode Naïve Bayes yaitu metode klasifikasi yang menggunakan teori Bayes Probabilitas untuk memprediksi kelas yang tidak diketahui. Pada Naïve Bayes dataset di anggap paling independent[4]. Evaluasi tingkat keberhasilan akan didasarkan pada nilai akurasi, baik menggunakan algoritma KNN maupun Naïve Bayes menggunakan data pada UCI Machine Learning Pima Indian Diabetes, sehingga hasil yang bisa didapat dari klasifikasi yang akan dibandingkan, dan bisa digunakan dalam kebutuhan medis maupun pengklasifikasian dalam penyakit diabetes.

Ada beberapa permasalahan lainnya dalam penelitian diantaranya missing data, missing data merupakan informasi yang tidak tersedia untuk sebuah obyek oleh karena itu akan dilakukan penanganan missing data oleh nilai mean dan median.

### Topik dan Batasannya

Hal utama dalam penelitian ini yang pertama yaitu membandingkan hasil dari penggunaan auto machine learning dengan menggunakan metode KNN dan Naïve Bayes hasil dengan akurasi tertinggi yang akan direkomendasikan dalam penggunaan klasifikasi dengan jumlah data yang telah ada.

KNN baik diterapkan penggunaannya dalam jumlah data yang besar, metode ini juga cukup sederhana untuk dipahami dan mudah diimplementasikan, selain itu metode KNN baik digunakan untuk masalah klasifikasi. Proses data mining dilakukan untuk mempreprocessing data terlebih dahulu agar data bersih dan layak untuk digunakan.

Batasan penelitian yang dilakukan yaitu :

- Menggunakan metode klasifikasi KNN dan Naïve Bayes
- Melakukan Analisis hasil dari klasifikasi untuk akurasi tertinggi dan terendah
- Menggunakan Teknik Data Mining untuk melakukan Preprocessing pada data Pima Indian Diabetes
- Penangan Missing Value

### Tujuan

Adapun beberapa tujuan dari penelitian ini diantaranya :

- Mendapatkan Hasil Optimum dari Klasifikasi KNN dan Naïve Bayes pada data uji dan data testing yang disajikan.
- Mendapatkan Perbandingan Hasil dari Klasifikasi KNN dan Naïve Bayes.
- Dapat dimanfaatkan untuk pendeteksian penyakit Diabetes.
- Menangani Missing Value