

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

*K-Nearest Neighbor* (KNN) merupakan suatu algoritma yang sering dipakai untuk melakukan klasifikasi terhadap objek berdasarkan data yang mempunyai jarak terdekat dengan objek tersebut. Pada konsepnya algoritma KNN terdapat dua data yaitu data latih dan data uji, data latih digunakan untuk melatih algoritma sesuai dengan yang dibentuk, sedangkan data uji digunakan untuk menguji algoritma yang sudah dilatih. Untuk mencari jarak paling terdekat dengan objek berdasarkan data yaitu dengan menentukan parameter K tetangga terdekat. Dari K tetangga terdekat dapat diambil keputusan hasil klasifikasi berdasarkan dari mayoritas atau nilai yang lebih banyak muncul. Namun ada kelemahan pada algoritma ini yaitu bobot dari setiap data pada K jarak terdekat besarnya sama yaitu  $1/k$ , pemberian bobot pada algoritma KNN dapat memperbaiki tingkat kesalahan. Pemberian bobot pada KNN disebut juga *Weighted K-Nearest Neighbor* (WKNN).[1]

*Weighted K-Nearest Neighbor* (WKNN) merupakan pengembangan dari algoritma KNN dan memiliki prinsip yang sama yaitu mencari jarak terdekat antara data yang akan di uji dengan sejumlah K tetangga terdekatnya dalam data latih, dengan menggunakan rumus *Euclidean distance*. Yang membedakan WKNN dengan KNN adalah pada perhitungan setelah mendapatkan hasil K tetangga terdekat, dilanjutkan dengan memberi pembobotan pada tetangga terdekat kemudian menentukan kelompok berdasarkan kelasnya.[2]

Diabetes melitus adalah suatu penyakit pada manusia sebagai dampak atas gangguan pada mengontrol kadar gula darah. Gangguan ini dapat disebabkan oleh sekresi hormon insulin yang tidak memadai atau fungsi insulin terganggu “resistensi insulin”, atau kombinasi dari keduanya. Seringkali, orang tidak mengetahui jika seseorang menderita diabetes dan seringkali mengalami keterlambatan dalam pengobatan sehingga menyebabkan banyak komplikasi. Diabetes juga dikenal sebagai “*Mother of Disease*” karena sebagai induk atau ibu dari penyakit lainnya seperti darah tinggi, penyakit jantung, stroke, gagal ginjal,

hingga kebutaan[3]. Diabetes melitus yang tidak ditangani dengan baik dapat berdampak buruk bagi orang yang terkena karena diabetes melitus adalah penyakit kronis. Kebanyakan penderita diabetes melitus tidak memahami tujuan pengobatan diabetes melitus, sehingga tidak menyadari risiko komplikasi yang dapat ditimbulkan dari penyakit diabetes melitus[4]. Berdasarkan situs [databoks.katadata.co.id](http://databoks.katadata.co.id) menunjukkan penelitian dari *International Diabetes Federation* (IDF) pada tahun 2021, Indonesia berada pada urutan ke kelima dengan jumlah penderita diabetes sebanyak 19,47 juta jiwa. Dengan penduduk sebesar 179,2 juta jiwa, berarti populasi diabetes di Indonesia sebesar 10,6%.

Dalam menyelesaikan suatu proses pengambilan kesimpulan, sistem identifikasi dapat menggunakan algoritma tertentu untuk diterapkan. Seperti pada penelitian sebelumnya yang dilakukan oleh Andi Maulida Argina yang berjudul “Penerapan Metode Klasifikasi *K-Nearest Neighbor* pada Dataset Penderita Diabetes”[5]. Dalam penelitian ini menggunakan metode KNN untuk menghitung akurasi, presisi, *recall*, dan *F-Measure*. Dan mendapatkan hasil akurasi tertinggi 39% pada  $K=3$ , presisi tertinggi 65% pada  $K=3$  dan  $K=5$ , *recall* tertinggi 36% pada  $K=3$ , dan *F-Measure* tertinggi 46% pada  $K=3$ . Namun jumlah data yang digunakan sedikit, sehingga nilai yang diperoleh tidak mencukupi. Penelitian selanjutnya dilakukan oleh Yessivha Imanuela Claudy, Rizal Setya Perdana, M. Ali Fauzi dengan judul “Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritma *K-Nearest Neighbor* (KNN)” [6]. Penelitian ini menggunakan 160 data dengan pembagian 50% data latih dan 50% data uji. Dari pengujian yang dilakukan didapatkan hasil akurasi sebesar 66% dengan hasil klasifikasi benar 53 data dan klasifikasi salah 27 data.

Saat ini sudah cukup banyak penelitian klasifikasi menggunakan algoritma KNN, penulis ingin menunjukkan bahwa klasifikasi bukan hanya menggunakan algoritma KNN saja, melainkan masih banyak algoritma lain yang masih jarang dipakai namun nilai akurasinya tidak kalah bagus dengan algoritma KNN, salah satunya yaitu algoritma WKNN. Seperti pada penelitian sebelumnya yang dilakukan oleh Aisyah Syahidah yang berjudul “Klasifikasi Imbalanced Data Menggunakan Metode *Weighted K-Nearest Neighbor* pada data debitur kartu kredit bank” [7]. Hasil klasifikasi terbaik pada data asli untuk data uji ke-8 yaitu

nilai akurasi terbesar berada pada K-10 dengan 75.26%. Selanjutnya pada penelitian Sigit Bagus Setiawan, Adiwijaya, dan Muhammad Syahrul Mubarak dengan judul “Klasifikasi Topik Berita Berbahasa Indonesia menggunakan Metode *Weighted K-Nearest Neighbor*”[1]. Berdasarkan data yang diuji menggunakan metode *Weighted K-Nearest Neighbor* dengan k=27 dengan performansi sebesar 75,86% dengan data training 80% dan data testing 20%.

Berdasarkan penjelasan latar belakang di atas, penulis berencana membuat sebuah sistem “Implementasi *K-Nearest Neighbor* Pada Penyakit Diabetes Melitus Dengan Pembobotan”. Algoritma berikut dipilih karena pada penelitian sebelumnya yang diproses oleh Cintya Surya Utami dengan judul “Klasifikasi Kinerja Perusahaan di Indonesia dengan menggunakan Metode *Weighted K-Nearest Neighbor*”[2], menunjukkan bahwa hasil terbaik klasifikasi menggunakan metode WKNN dengan studi kasus 436 perusahaan didapatkan nilai akurasi tertinggi 98.8% pada parameter k=3. Dari uraian tersebut penulis tertarik mengembangkan sebuah sistem untuk membandingkan performa algoritma KNN dengan algoritma WKNN. Yang diharapkan dapat mengetahui hasil akurasi terbaik dari kedua algoritma yang di uji.

## **1.2 Perumusan Masalah**

### **1.2.1 Permasalahan**

Pada algoritma *K-Nearest Neighbor* (KNN) bobot setiap data pada K tetangga terdekat besarnya sama yaitu  $\frac{1}{K}$ , sehingga tingkat performansi nya kurang akurat.

### **1.2.2 Usulan Solusi**

Memberikan pembobotan pada K tetangga terdekat dengan menggunakan algoritma *Weighted K-Nearest Neighbor* (WKNN).

### **1.2.3 Pertanyaan Penelitian**

Benarkah algoritma *Weighted K-Nearest Neighbor* (WKNN) lebih akurat daripada algoritma *K-Nearest Neighbor* (KNN) dalam mendiagnosa penyakit diabetes melitus?

### **1.3 Tujuan dan Manfaat**

#### **1.3.1 Tujuan Penelitian**

Tujuan dari penelitian ini adalah untuk membandingkan hasil akurasi tertinggi dengan menggunakan algoritma WKNN dan KNN dalam mendiagnosa penyakit diabetes melitus.

#### **1.3.2 Manfaat Penelitian**

Penelitian ini akan bermanfaat bagi peneliti berikutnya dalam mengetahui tingkat akurasi yang dihasilkan oleh metode *Weighted K-Nearest Neighbor* (WKNN).

### **1.4 Batasan-batasan**

Adapun batasan masalah dalam penelitian ini adalah :

1. Data yang digunakan pada penelitian ini adalah data National Institute of Diabetes and Digestive and Kidney Diseases dari situs kaggle.com  
( <https://www.kaggle.com/kandij/diabetes-dataset/version/1> )
2. Jumlah data sebanyak 768.
3. Parameter yang digunakan ada 8 gejala pada penyakit diabetes mellitus yaitu kehamilan, glukosa, tekanan darah, ketebalan kulit, insulin, BMI, silsilah diabetes, umur.

### **1.5 Sistematika Penulisan**

Sistematika penulisan skripsi ini terdiri dari beberapa bab, yaitu :

#### **BAB I PENDAHULUAN**

Bab ini membahas tentang latar belakang, rumusan masalah, tujuan, manfaat, batasan masalah, dan sistematika penulisan.

#### **BAB II KAJIAN PUSTAKA**

Bab ini membahas tentang landasan teori dan penelitian sebelumnya.

#### **BAB III METODE USULAN**

Bab ini membahas tentang metode yang diusulkan (algoritma), arsitektur (input,proses,ouput), data penelitian, tahapan penelitian, dan rencana pengujian.

#### **BAB IV HASIL DAN PEMBAHASAN**

Bab ini membahas mengenai lingkungan pengembangan, hasil pembuatan sistem serta hasil uji coba yang telah dilakukan.

## BAB V PENUTUP

Bab ini membahas mengenai kesimpulan dari penelitian yang telah dilakukan serta saran pada penelitian selanjutnya.

REFERENSI Berisi daftar referensi atau daftar pustaka yang menjadi acuan dalam penelitian.

