

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Sejak 1991 setiap tahun, IDF (International Diabetes Federation) dan Organisasi Kesehatan Dunia (WHO) telah menetapkan tanggal 14 Nopember sebagai Hari Diabetes. Pada tahun 2007, hari diabetes ini resmi sebagai hari sedunia dalam agenda PBB. Ini menandakan kalau penyakit Diabetes Melitus (disingkat DM) tidak bisa dipandang sebelah mata. WHO memprediksi kenaikan jumlah penyandang DM di Indonesia dari 8,4 juta pada tahun 2000 menjadi sekitar 21,3 juta pada tahun 2030 [PER-11]. Penyakit diabetes ini merupakan penyakit yang ditandai dengan kadar gula darah yang tinggi karena gangguan pada sekresi insulin. Tubuh pasien DM tidak dapat memproduksi atau merespon hormon insulin yang dihasilkan oleh organ pankreas, sehingga kadar gula darah meningkat dan dapat menyebabkan komplikasi jangka pendek maupun jangka panjang pada pasien [REG-12].

Pada perkembangan di dunia kedokteran saat ini, peneliti dan praktisi memusatkan perhatiannya untuk mendeteksi DM dan mencegah atau menghambat berkembangnya komplikasi. Untuk mendeteksi seseorang terkena diabetes, ada beberapa tes lab yang harus dilakukan. US National Institute of Diabetes telah melakukan uji untuk penyakit diabetes sesuai dengan kriteria Organisasi Kesehatan Dunia yang dilakukan pada sejumlah perempuan yang berusia di atas 21 tahun, dari warisan Pima India dan tinggal di dekat Phoenix, Arizona, Amerika Serikat. Lebih dari 50% populasinya menderita DM. *Dataset* Pima meliputi delapan atribut pengukuran dari pasien yang DM positif dan pasien didiagnosis DM negatif. Dari data ini akan diambil sample data sebanyak 290 record (154 record data negatif diabetes dan 136 record data positif diabetes ) data dengan proporsi sampel data training sebanyak 240 objek dan data testing sebanyak 50 objek.

Pendeteksian penyakit diabetes menggunakan *dataset* diabetes “Indian Pima” sudah pernah dilakukan dengan berbagai metode, salah satunya

implementasi metode J48 dan ID3 yang digunakan oleh [LES-12] pada data diabetes Indian prima. Akurasi tertinggi yang didapatkan pada penelitian ini mencapai 74.72% untuk J48 dan 72.64% untuk ID3.

Metode *K-Nearest Neighbor* (K-NN) melakukan klasifikasi pada data baru yang masih belum diketahui masuk ke dalam kelas mana dengan menggunakan beberapa data dengan sejumlah  $k$  yang letaknya terdekat dengan data baru tersebut [JNI-96]. Algoritma *K-Nearest Neighbor* (K-NN) sering digunakan pada penelitian menggunakan teknik klasifikasi karena mempunyai kelebihan, diantaranya sederhana dalam penggunaan dan sangat efektif jika *data training* yang digunakan dalam jumlah besar.

Jika K-NN melakukan prediksi dan klasifikasi secara tegas pada data uji berdasarkan perbandingan  $K$  tetangga terdekat, *fuzzy K-NN* melakukan prediksi menggunakan metode yang sama tetapi tidak secara tegas memprediksi kelas yang harus diikuti data uji. Pemberian label kelas dilakukan berdasarkan teori *fuzzy*, dimana data uji diberikan nilai keanggotaan pada setiap kelas yang terdapat pada data latih. Menurut [KEL-85], *fuzzy K-NN* adalah pengembangan K-NN yang digabungkan dengan teori *fuzzy* dalam memberikan definisi pemberian label kelas pada data uji yang diprediksi.

Penelitian menggunakan algoritma *K-Nearest Neighbor* (K-NN) sebelumnya pernah diterapkan oleh [SHO-12] pada data diagnosis penyakit hati dan didapatkan akurasi mencapai 97.4%. Sedangkan *Fuzzy K-NN* pernah diterapkan oleh [PRA-12] pada klasifikasi bunga iris dan akurasi yang ditemukan mencapai 96%.

Berdasarkan kedua penelitian tersebut, dapat dilihat bahwa kedua algoritma, KNN dan *fuzzy KNN*, mempunyai kinerja yang baik dalam melakukan klasifikasi data. Namun, belum diketahui algoritma mana diantara keduanya yang lebih unggul kinerjanya. Oleh karena itu kedua algoritma ini perlu dibandingkan pada dataset yang sama.

Berdasarkan latar belakang tersebut, maka penulis mengambil judul pada skripsi ini, **"Perbandingan K-Nearest Neighbor dan Fuzzy K-Nearest Neighbor pada Diagnosis Penyakit Diabetes Melitus"**

## 1.2 Perumusan Masalah

Rumusan masalah dalam skripsi ini antara lain:

1. Bagaimana penerapan metode *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor* untuk klasifikasi *dataset* penyakit Diabetes Melitus (DM).
2. Berapa tingkat akurasi yang dipengaruhi sejumlah nilai data k set pada *dataset* Diabetes Melitus menggunakan algoritma *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor*.
3. Bagaimana pengaruh pembobotan (m) pada perhitungan *Fuzzy K-Nearest Neighbor*.

## 1.3 Batasan Masalah

Pada skripsi ini, permasalahan dibatasi oleh hal-hal sebagai berikut :

- a. Data yang digunakan untuk penelitian dalam skripsi ini diperoleh dari database UCI machine learning repository : Indian Pima Diabetes Dataset pada <http://archive.ics.uci.edu>
- b. Parameter DM yang digunakan berupa jumlah hamil, 2 jam PP (OGTT), tekanan diastolik, ketebalan kulit trisep (TSFT), 2 jam serum insulin (INS), indeks massa badan (IMB), riwayat diabetes keluarga (DPF) dan usia.
- c. Output dari diagnosa resiko DM pada sistem yaitu positif DM (1) dan negatif DM (0).
- d. Tidak menangani data yang memiliki *missing value* pada data latihan maupun data uji.

## 1.4 Tujuan

Tujuan pembuatan skripsi ini adalah :

1. Menerapkan metode *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor* untuk klasifikasi *dataset* penyakit Diabetes Melitus (DM).
2. Membandingkan tingkat akurasi sistem yang dipengaruhi data k set pada *dataset* Diabetes Melitus menggunakan algoritma *K-Nearest Neighbor*

dan *Fuzzy K-Nearest Neighbor* berdasarkan input parameter yang diberikan berdasarkan.

3. Mengetahui pengaruh dari pemberian nilai pembobotan pada proses perhitungan *Fuzzy K-Nearest Neighbor* (FK-NN).

### 1.5 Manfaat Penelitian

Manfaat yang dapat diambil dari skripsi ini adalah dapat menemukan kinerja mana yang paling bagus antara *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor* dalam melakukan klasifikasi data serta penggunaan kedua algoritma ini untuk mendiagnosis penyakit Diabetes Melitus.

### 1.6 Sistematika Penulisan

Sistematika penulisan dalam skripsi ini adalah sebagai berikut :

#### 1. BAB I : PENDAHULUAN

Bab ini memuat latar belakang penulisan, perumusan masalah, tujuan, batasan masalah, manfaat penelitian dan sistematika penulisan skripsi.

#### 2. BAB II : KAJIAN PUSTAKA DAN DASAR TEORI

Bab ini memuat jurnal, uraian dasar teori dan referensi yang mendasari pembuatan sistem perbandingan kinerja *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor* dalam melakukan klasifikasi diagnosis Diabetes Melitus (DM).

#### 3. BAB III : METODE PENELITIAN

Bab ini memuat metode yang digunakan dalam penelitian untuk membandingkan kinerja *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor* dalam melakukan klasifikasi diagnosis Diabetes Melitus (DM).

#### 4. BAB IV : PERANCANGAN DAN IMPLEMENTASI

Bab ini memuat perancangan sistem dan implementasi program perbandingan *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor* dalam melakukan klasifikasi diagnosis Diabetes Melitus (DM).

#### 5. BAB V : PENGUJIAN DAN ANALISIS

Bab ini memuat hasil pengujian dan analisa tentang perbandingan *K-Nearest Neighbor* dan *Fuzzy K-Nearest Neighbor* dalam melakukan

klasifikasi diagnosis Diabetes Melitus (DM) dengan *dataset* yang digunakan.

## 6. BAB VI : PENUTUP

Bab ini memuat kesimpulan yang diperoleh dari hasil pengujian dan saran untuk pengembangan lebih lanjut.

