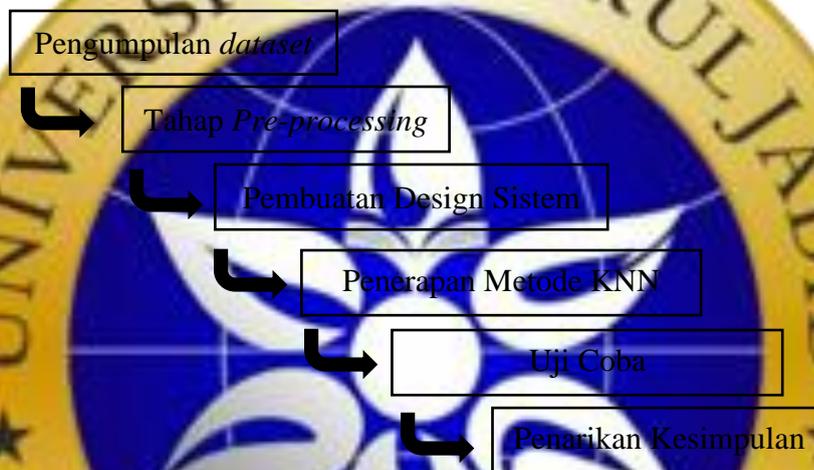


BAB III

METODE PENELITIAN

3.1. Rancangan Penelitian

Model penelitian yang dirancang pada penelitian ini bertujuan untuk memudahkan proses identifikasi penyakit *diabetes melitus*. Perancangan model penelitian terdiri dari beberapa tahapan-tahapan proses yang akan dilakukan sehingga pengerjaan tugas akhir berjalan secara sistematis dan teratur. Tahapan model penelitian yang akan dipaparkan akan dirangkum pada **Gambar 3.1**.



Gambar 3. 1 Rancangan Penelitian

3.2. Tahapan Penelitian

Langkah-langkah yang akan dilakukan selama proses penelitian yang telah dirangkum dalam sub-bab desain penelitian sebelumnya, akan dijelaskan lebih rinci pada sub-bab ini. Berdasarkan **Gambar 3.1**, terdapat 6 tahapan penelitian diantaranya adalah pengumpulan *dataset*, tahap *preprocessing*, pembuatan design sistem untuk deteksi penyakit *diabetes melitus*, penerapan metode KNN untuk deteksi *diabetes melitus*, uji coba, dan penarikan kesimpulan.

3.2.1. Pengumpulan Dataset

Dataset seputar penyakit *diabetes melitus* berasal dari data publik Pima Indians Diabetes Database. *Dataset* tersebut terdiri dari 8 variabel yaitu jumlah

kehamilan, kadar gula darah (*glukosa*), tekanan darah, ketebalan lipatan kulit *trisept*, kadar *insulin*, *index masa tubuh (IMT)*, riwayat *diabetes melitus* dalam keluarga, serta usia. Jumlah keseluruhan data yang terdapat dalam dataset tersebut sebanyak 768 data, dengan jumlah pasien positif sebanyak 268 dan jumlah negatif sebanyak 500 data.

Selain menggunakan *dataset* dari Pima Indians Diabetes Database, juga dilakukan wawancara dengan seorang dokter ahli, wawancara tersebut dilakukan untuk menggali informasi lebih dalam terkait penyakit *diabetes mellitus* serta untuk mengkonfirmasi kebenaran dari 8 variabel penyebab penyakit *diabetes melitus* yang telah didapatkan sebelumnya.

Tabel 3. 1 Wawancara Dokter

No	Pertanyaan	Jawaban
1	Apakah penyakit <i>diabetes melitus</i> merupakan salah satu penyakit dengan jumlah pasien tertinggi?	
2	Bagaimana proses identifikasi penyakit <i>diabetes melitus</i> saat ini?	
3	Apakah saja faktor faktor yang menyebabkan seseorang menderita <i>diabetes melitus</i> ?	

3.2.2. Tahap *Pre-processing*

Tahap *pre-processing* adalah sebuah tahapan penting yang perlu dilakukan sebelum proses pengolahan data dilakukan, tahap ini merupakan tahap pembersihan suatu data, yang meliputi menghilangkan duplikasi data, memeriksa data inkonsisten serta memperbaiki kesalahan. Analisis *dataset* mengungkapkan bahwa tidak semua variabel memiliki data atribut yang lengkap, dan bahwa kelengkapan data memiliki dampak yang signifikan terhadap hasil klasifikasi. Jumlah atribut data tidak lengkap pada masing – masing variable yaitu : variable jumlah kehamilan sebanyak 111, variable kadar gula (*glukosa*) sebanyak 5, variabel tekanan darah sebanyak 35, variable ketebalan lipatan kulit sebanyak 227, variabel *insulin* sebanyak 374, variabel *index masa tubuh* sebanyak 11, sedangkan variabel riwayat *diabetes melitus* dalam keluarga, usia dan hasil prediksi pasien memiliki atribut data yang lengkap.

Tabel 3. 2 Data Atribut yang Hilang (*missing values*)

No	Variabel	Jumlah Data Hilang
1	Jumlah kehamilan	110
2	Kadar gula darah	5
3	Tekanan darah	35
4	Ketebalan lipatan kulit trisep	227
5	Insulin	374
6	Index Massa Tubuh (IMT)	11
7	Riwayat diabetes dalam keluarga	0
8	Usia	0
9	Hasil prediksi pasien	0

Untuk mengatasi ketidak lengkapan atribut data dari masing-masing variabel data dapat dilakukan dengan beberapa cara, antara lain :

1. Mengganti semua nilai yang hilang dengan nilai nol.
2. Mengganti nilai yang hilang dengan mean atau rata rata.
3. Menghapus data yang tidak memiliki nilai.

Penelitian ini menggunakan kombinasi dari ketiga cara diatas untuk mengatasi nilai yang tidak lengkap dari masing-masing variabel, aturannya adalah sebagai berikut :

- a. Nilai nol dalam variable jumlah kehamilan diasumsikan bahwa pasien belum pernah melahirkan,
- b. Nilai nol dalam variabel kadar gula (*glukosa*), tekanan darah dan Index Massa Tubuh (IMT) diganti dengan mean atau rata rata.
- c. Nilai nol dalam variabel ketebalan lipatan kulit trisep dan insulin dihilangkan.

3.2.3. Pembuatan Design Sistem

Pembuatan *design* sistem dimaksudkan untuk mempermudah pengguna sistem dalam pengoperasikan sistem pendukung keputusan yang dibuat. Design sistem pendukung keputusan ini dibuat dengan menggunakan *QtDesigner*. Berikut adalah design dari sistem pendukung keputusan yang akan dibuat :

Aplikasi Deteksi Diabetes

Jumlah Kehamilan

Kadar Glukosa

Tekanan Darah

Ketebalan Lipatan Kulit

Kadar Insulin

Index Masa Tubuh

Riwayat Diabetes Keluarga

Usia

Button Prediksi

Hasil Prediksi

Gambar 3. 2 Design Sistem

3.2.4. Penerapan Metode KNN

Penerapan serta penggunaan metode KNN dilakukan pada dua kali percobaan, pertama pada sistem yang dibuat (dengan menggunakan tampilan sistem pada **Gambar 3.2**) sistem dibuat dengan menggunakan bahasa pemrograman *python* dengan menggunakan aplikasi *jupyter notebook* dan yang kedua pada perhitungan dengan menggunakan *excel*, perhitungan dengan *excel* dilakukan untuk melihat hasil perhitungan jarak *eulidean* yang dihasilkan.

Langkah-langkah yang dilakukan untuk menggunakan metode KNN pada sistem :

1. Mencari Nilai k Terbaik

Nilai k optimal dalam algoritma ini tergantung pada jumlah data. Nilai k yang besar belum tentu menghasilkan tingkat akurasi yang tinggi, begitupun pula sebaliknya. Nilai k yang digunakan biasanya merupakan angka ganjil hal itu untuk menghindari jumlah label class imbang, sehingga akan menyulitkan proses penarikan kesimpulan. Pencarian nilai k optimal merupakan kunci dari keberhasilan sistem yang akan dibuat, untuk itu tahap pertama yang harus dilakukan adalah mencari nilai k terkecil dengan tingkat akurasi tertinggi

2. Pembuatan Sistem

Pada tahap ini hasil pembuatan design serta model akurasi paling optimal yang telah dibuat sebelumnya akan disatukan sehingga nantinya akan menghasilkan suatu sistem pendukung keputusan yang siap digunakan kapan saja

Langkah-langkah yang dilakukan dalam penerapan perhitungan algoritma KNN yaitu:

1. Menentukan nilai k.

Nilai k yang akan digunakan dalam proses perhitungan adalah nilai k terbaik yang telah didapatkan dalam proses sebelumnya.

2. Menghitung kuadrat jarak masing-masing objek terhadap data baru. Penelitian ini menggunakan rumus *euclidian distance* untuk menghitung jarak antar objek. Rumus *euclidean distance* dapat dihat pada persamaan 2.1.

3. Mengurutkan objek dari mulai jarak terdekat sampai jarak terjauh

Hasil perhitungan jarak *euclidean* tersebut kemudian akan diurutkan mulai dari jarak terdekat sampai jarak terjauh. Semakin besar hasil perhitungan jarak *euclidean*, maka jarak antara kedua objek semakin jauh, dan semakin kecil hasil perhitungan jarak *euclidean*, maka jarak antara kedua objek semakin dekat

4. Mengumpulkan label class Y (klasifikasi *nearest neighbor*).

5. Melakukan prediksi berdasarkan label class Y yang paling mayoritas.

3.2.5. Uji Coba

Proses pengujian dalam penelitian ini menggunakan model *confusion matrix*. Tahapan uji coba dilakukan untuk mengetahui keberhasilan dari sistem pendukung keputusan yang dibuat. Tingkat keberhasilan suatu penerapan metode dapat dilihat dari hasil akurasi yang didapatkan. Dalam penelitian ini proses uji coba dilakukan menggunakan nilai k paling optimal. Data uji yang digunakan berjumlah 0.126% dari jumlah dataset. Untuk menghitung hasil akurasi dapat menggunakan persamaan 2.5.

3.2.6. Penarikan Kesimpulan

Setelah melakukan seluruh rangkaian proses pada penelitian yang dilakukan, pada tahapan ini akan dilakukan analisis terhadap hasil uji coba dari sistem yang digunakan. Sehingga dapat disimpulkan keunggulan dan kelemahan dari sistem yang dibuat. Tahap akhir ini akan menentukan tindakan yang akan dilakukan peneliti selanjutnya untuk mengembangkan penelitian yang telah dilakukan.

