

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Diabetes adalah penyakit kronis yang terjadi akibat pankreas tidak dapat bekerja dengan semestinya memproduksi enzim atau hormon dalam jumlah dan dalam waktu yang tepat. Hiperglikemia, atau peningkatan gula darah, adalah efek umum dari diabetes yang tidak terkontrol dan dari waktu ke waktu menyebabkan kerusakan serius pada banyak sistem tubuh, terutama saraf dan pembuluh darah [1]. Menurut data *World Health Organization* (WHO) pada tahun 2016 di Indonesia terjadi sebanyak 48.300 orang yang berusia 30-69 tahun meninggal dunia akibat diabetes [2]. Keterlambatan penanganan serta pendeteksian dari penyakit diabetes ini dapat mengakibatkan bertambahnya kasus kematian sehingga dibutuhkan pendeteksian dini untuk mengurangi tingkat risiko kematian dari penyakit diabetes ini. Banyak penelitian menerapkan metode berbasis *K-Nearest Neighbor* (KNN) namun belum mendapat hasil yang optimal karena metode ini memiliki kekurangan yaitu harus menyimpan dalam jumlah dan karakteristik yang sangat besar pada data latih pada saat proses prediksi [3].

Penelitian sebelumnya yang dilakukan Indriyanti, dkk telah melakukan penelitian untuk memprediksi penyakit diabetes dengan menggunakan KNN pada dataset *Pima Indian Diabetes Dataset*. Pada penelitian tersebut didapatkan tingkat akurasi tertinggi 75.14%, tetapi pada penelitian tersebut untuk KNN masih belum melakukan *preprocessing data* [4]. Pada dataset tersebut terdapat *noise* seperti beberapa data yang hilang. Untuk mengatasi *noisy data* perlu dilakukan *preprocessing* agar hasil prediksi menggunakan KNN lebih optimal [5]. Disamping itu, penggunaan KNN juga membuat penggunaan memori yang tidak sedikit karena harus menghitung jarak ke setiap data latih, terlebih lagi kalau data yang digunakan banyak [6]. Untuk itu perlu dilakukan pemilihan data latih yang lebih optimal agar mendapat hasil yang lebih optimal dengan data yang lebih sedikit. Selain itu, *Preprocessing* perlu dilakukan guna meningkatkan akurasi seperti salah satu hasil penelitian yang dilakukan oleh Darnisa Azzahra Nasution, dkk tentang perbandingan *data preprocessing* dengan normalisasi data [7].

Fokus dari penelitian ini mengembangkan penelitian sebelumnya dilakukan dengan menambahkan *data preprocessing* dan mereduksi data latih menggunakan *K-Support Vector Nearest Neighbor* (KSVNN) untuk mendapatkan performa prediksi yang lebih baik dengan menggunakan data yang lebih sedikit. Dataset yang digunakan di sini sama dengan penelitian sebelumnya. Berdasarkan penelitian yang dilakukan oleh Eko Prasetyo terkait penerapan KSVNN ini untuk klasifikasi berbasis KNN ini merupakan algoritma yang lebih baik daripada KNN karena dapat meningkatkan akurasi dengan data latih yang lebih sedikit hal tersebut sudah dilakukan terhadap beberapa *dataset* [3]. Oleh karena itu, dalam penelitian ini akan dilakukan prediksi penyakit diabetes dengan menerapkan algoritma KSVNN sebagai metode untuk mereduksi data latih.

1.2 Rumusan Masalah

Permasalahan yang dapat dirumuskan pada latar belakang masalah di atas yaitu bagaimana menerapkan *preprocessing* dan mereduksi data latih menggunakan KSVNN.

1.3 Maksud dan Tujuan

Maksud dibuatnya penelitian ini adalah untuk menerapkan *preprocessing data* dan menerapkan algoritma KSVNN untuk mereduksi data latih. Tujuan dilakukannya penelitian ini adalah untuk mengukur hasil dari penerapan *preprocessing data* dan algoritma KSVNN dalam mereduksi data latih.

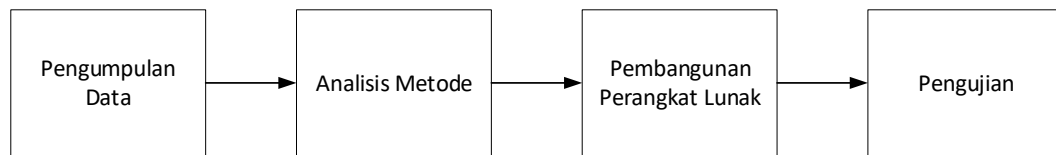
1.4 Batasan Masalah

Batasan masalah pada pembangunan sistem untuk memprediksi penyakit diabetes ini adalah sebagai berikut :

- a. Dataset yang digunakan adalah *Pima Indians Diabetes Database* yang diambil dari *UCI Machine Learning Repository* [8].
- b. Data masukan berupa *file* berformat csv.

1.5 Metodologi Penelitian

Pada penelitian ini digunakan metode *Case Studies Research* yang memfokuskan pada suatu kasus tertentu. Metode ini dilakukan untuk mengumpulkan data serta melakukan penggalan yang lebih pada objek tersebut [9]. Adapun tahapan yang akan dilakukan dapat dilihat pada Gambar 1.1.



Gambar 1.1 Tahapan Penelitian

1.5.1 Metode Pengumpulan Data

Metode yang digunakan untuk mengumpulkan data dalam menyelesaikan penelitian ini yaitu studi literatur. Peneliti mencari referensi dari jurnal, buku, maupun internet yang terkait dengan bidang kesehatan mengenai penyakit diabetes dan metode yang akan digunakan.

Dalam pengujiannya, digunakan dataset dari *Pima Indian Diabetes Database* [8]. Dataset sebanyak 768 *records* tersebut merupakan data dari pasien berjenis kelamin perempuan dan usianya lebih dari atau sama dengan 21 tahun yang diambil dari *National Institute of Diabetes and Digestive and Kidney Diseases*.

1.5.2 Analisis Metode

Analisis metode adalah proses analisa metode-metode yang akan digunakan pada penelitian, adapun metode pada sistem prediksi ini terbagi menjadi dua yaitu:

1. *Preprocessing*

Preprocessing adalah tahapan awal dalam mengolah data *input* sebelum memasuki proses tahapan utama untuk diproses pada tahap *processing*. Adapun tahapan di dalam yang ada pada tahap *preprocessing* yaitu menggunakan *min-max normalization*.

2. *Processing*

Processing adalah tahapan proses utama dalam penelitian, adapun tahapannya yaitu:

a. KSVNN

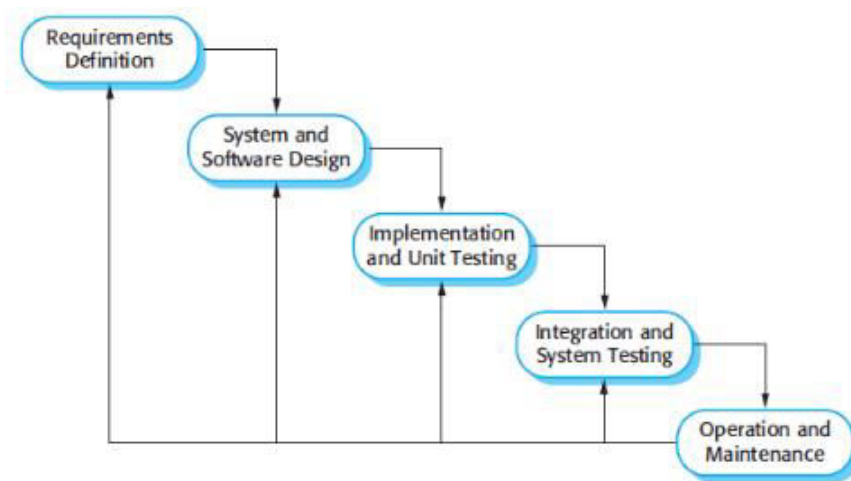
Data latih direduksi menggunakan metode KSVNN dengan mengambil sejumlah *support vector* sebagai data latih yang nantinya digunakan pada tahap berikutnya.

b. KNN

Data latih yang telah direduksi menggunakan KSVNN masuk ke tahap prediksi menggunakan KNN menghitung jarak antara data sampel dengan data lainnya.

1.5.3 Metode Pengembangan Perangkat Lunak

Metode pembangunan perangkat lunak yang digunakan pada penelitian ini menggunakan *waterfall model*. Adapun tahapan diagram pengembangan aplikasi menggunakan waterfall model dapat dilihat pada Gambar 1.2.



Gambar 1.2 Waterfall Model Diagram

- a. *Requirement definition*
Proses menganalisis hal-hal apa saja yang dibutuhkan yang meliputi kebutuhan fungsional dan kebutuhan non-fungsional dalam pembangunan perangkat lunak.
- b. *System and software design*
Proses perancangan aplikasi baik tampilan maupun sistem yang akan dibangun terhadap solusi dari permasalahan yang ada dengan menggunakan perangkat pemodelan seperti *Unified Modeling Language (UML)*.
- c. *Implementation and unit testing*
Proses implementasi perangkat lunak berdasarkan arsitektur sistem dan tampilan antarmuka yang sudah dibuat sebelumnya ke dalam suatu bahasa pemrograman. Pada penelitian ini digunakan bahasa pemrograman Python.

d. *Integration and system testing*

Proses pengujian perangkat lunak apakah sudah sesuai dengan kebutuhan atau belum.

e. *Operation and maintenance*

Proses penerapan perangkat lunak yang sudah dibuat dan sudah melalui tahap pengujian sebelumnya sekaligus melakukan pemeliharaan terhadap perangkat lunak yang sudah diterapkan tersebut. Pemeliharaan dapat berupa perbaikan aplikasi karena suatu *bug* (masalah) ataupun karena adanya penambahan fitur.

1.5.4 Pengujian

. Tahap pengujian, adalah tahap melakukan pengujian terhadap reduksi data dengan KSVNN dan metode KNN, tahap pengujian bertujuan untuk mendapatkan hasil berupa akurasi dalam memprediksi penyakit diabetes.

1.6 Sistematika Penulisan

Penelitian ini disusun dengan sistematika dengan tujuan agar hasil dari laporan lebih terarah dan tidak menyimpang dari permasalahan yang telah direncanakan diawal. Dengan pembahasan sebagai berikut.

BAB 1 PENDAHULUAN

Dalam bab ini menjelaskan tentang latar belakang, identifikasi masalah, maksud dan tujuan, batasan masalah serta metode penelitian yang digunakan dalam melakukan penelitian ini.

BAB 2 TINJAUAN PUSTAKA

Dalam bab ini menjelaskan tentang landasan teori yang dapat mendukung pokok pembahasan dari pembuatan aplikasi untuk memprediksi penyakit diabetes menggunakan metode KSVNN. Seperti pembahasan mengenai penyakit diabetes, *preprocessing*, KNN, dan KSVNN.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Dalam bab ini akan menjelaskan tentang analisis serta perancangan dibutunya aplikasi ini. Analisis yang akan dibahas berupa analisis sistem, analisis algoritma, analisis kebutuhan fungsional, analisis kebutuhan non-fungsional. Bab

ini menjelaskan mengenai analisis reduksi data latih menggunakan KSVNN lalu menggunakan algoritma KNN untuk memprediksi penyakit diabetes.

BAB 4 IMPLEMENTASI DAN PENGUJIAN

Pada bab ini penulis akan melakukan implementasi dan pengujian. Implementasi meliputi implementasi perangkat lunak, implementasi perangkat keras, dan implementasi antarmuka. Kemudian dilakukan pengujian akurasi prediksi penyakit diabetes untuk mengetahui seberapa besar akurasi yang diperoleh dari penerapan reduksi data latih *K-Support Vector Nearest Neighbor* dan metode *K-Nearest Neighbor*.

BAB 5 KESIMPULAN DAN SARAN

Bab ini akan memaparkan kesimpulan hasil analisis dan memberikan masukan atau saran bagi perbaikan sistem guna memperoleh kesempurnaan sistem.