

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Diabetes merupakan salah satu penyakit yang berbahaya juga penyakit yang kompleks dan rumit. Tingkat diagnosa diabetes memberikan kontribusi yang signifikan terhadap komorbiditas, atau penyakit lain yang dapat menyertai penyakit utama, dan tingkat komplikasi diabetes [1]. Pada diabetes mellitus, komorbiditas tersebut dapat berupa serangan jantung, stroke, gagal ginjal bahkan kebutaan. Pemicu diabetes mellitus ini terdiri dari 2 faktor, yaitu faktor internal seseorang dan faktor external. Faktor internal meliputi keturunan atau genetika, di mana tubuh tidak menghasilkan insulin atau mengalami kelainan atau kerusakan pankreas sejak kecil. Sedangkan faktor external adalah pola hidup yang tidak sehat seperti terlalu banyak mengkonsumsi gula, obesitas, dan jarang berolahraga. Di Indonesia sendiri, jumlah penderita diabetes mellitus naik setiap tahunnya [2].

Oleh karena itu deteksi dan diagnosis diabetes pada tahap awal dibutuhkan. Diagnosis medis merupakan sebuah proses klasifikasi. Seorang dokter harus menganalisis banyak faktor terlebih dahulu sebelum dapat memulai diagnosis diabetes, yang mana membuat pekerjaan dokter sulit [3]. Berdasarkan data histori penderita penyakit diabetes, penyakit diabetes ini dapat diklasifikasikan. Tujuan dari klasifikasi data adalah untuk mengelompokkan suatu objek menjadi sejumlah kategori atau kelas. Sehingga dapat dibuat suatu sistem yang dapat melakukan klasifikasi atau prediksi dalam mendiagnosa penyakit diabetes pada seseorang. Untuk melakukan klasifikasi data, dapat digunakan metode pembelajaran mesin atau *machine learning*. Dengan adanya pembelajaran mesin tersebut, dapat dilakukan deteksi penyakit dengan pengukuran klinis yang lebih cepat. Sehingga, diharapkan dapat membantu masyarakat terutama tenaga medis dalam melakukan diagnosis.

Pada saat ini sudah terdapat beberapa penelitian mengenai prediksi diabetes, dataset yang digunakan pun beragam. Seperti data hasil observasi peneliti terhadap penduduk di Indonesia, mulai dari *random sampling* Rumah Tangga (RT) oleh Wahyuni [4], dan data salah satu Rumah Sakit oleh Anggraeni [5]. Ada pula yang menggunakan dataset internasional “Pima Indian Diabetes Database” dari The John Hopkins University.

Dalam mengolah dataset internasional tersebut, sudah banyak metode *machine learning* yang digunakan. Pada salah satu penelitian sebelumnya, metode *machine learning* yang digunakan adalah KNN, *Naive Bayes*, dan *Random Forest* [2]. Pada penelitian yang lain, digunakan metode *Artificial Neural Network* (ANN) atau jaringan saraf tiruan [6], dan ada pula metode *Support Vector Machine* (SVM) [3][7]. Dari ke-4 penelitian tersebut, metode SVM memiliki tingkat akurasi yang kurang maksimal. Hal ini ditunjukkan dengan nilai akurasi dari SVM yang lebih rendah dibandingkan dengan metode ANN dan *Random Forest*. Metode SVM hanya menghasilkan akurasi sebesar 78%, sedangkan untuk metode ANN dan *Random Forest* masing-masing menghasilkan akurasi sebesar 81,8% dan 86,4%.

Metode SVM adalah salah satu metode pembelajaran mesin terbimbing atau *Supervised Machine Learning*. Metode SVM ini biasa digunakan untuk melakukan klasifikasi dan regresi. SVM diwakili dengan bantuan hyper plane. Sebagai contoh, diberikan suatu set karakteristik yang termasuk ke dalam salah satu kelas dari dua kelas, SVM akan menemukan bidang hiper yang memiliki kemungkinan pecahan karakteristik terbesar dari kelas yang sama pada bidang yang sama. Bidang hiper pemisah ini disebut bidang hiper pemisah optimal yang memaksimalkan jarak antara dua bidang hiper paralel dan dapat meminimalkan risiko kesalahan klasifikasi.[3]

Metode SVM memiliki berbagai macam turunan, salah satunya adalah Metode *Reduced Support Vector Machine* (RSVM). Metode RSVM ini diusulkan sebagai alternatif dari SVM standar, dan dipercaya dapat menyelesaikan kesulitan dalam menangani dataset yang cukup besar [8]. Metode RSVM juga sudah terbukti lebih baik dari metode *Smooth Support Vector Machine* (SSVM) dalam hal performansi terhadap data yang relatif besar (lebih dari 1000) [9].

Berdasarkan apa yang sudah dijelaskan di atas, penelitian ini bertujuan untuk menguji performa dari metode *Reduced Support Vector Machine* (RSVM) yang mana merupakan turunan dari metode SVM tersebut. Metode RSVM ini dipilih selain sebagai antisipasi jika suatu saat digunakan pada data yang relatif lebih besar di masa mendatang, juga karena waktu untuk pembentukan model yang cepat [9].

1.2 Identifikasi Masalah

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya, terdapat identifikasi masalah yang dijadikan bahan penelitian, yaitu apakah metode *Reduced Support Vector Machine* (RSVM) dapat lebih baik dari metode SVM dalam melakukan prediksi diabetes.

1.3 Perumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya, permasalahan dalam penelitian ini adalah bagaimana cara untuk menerapkan metode *Reduced Support Vector Machine* (RSVM) dalam melakukan klasifikasi atau memprediksi positif atau negatifnya penyakit diabetes pada seorang pasien.

1.4 Maksud dan Tujuan

Berdasarkan masalah yang sudah dijelaskan sebelumnya, maksud dari penelitian ini adalah untuk menerapkan metode *Reduced Support Vector Machine* (RSVM) dalam melakukan klasifikasi atau memprediksi positif atau negatifnya penyakit diabetes pada seorang pasien. Sedangkan tujuan dari penelitian ini adalah untuk mengukur hasil dari penerapan metode *Reduced Support Vector Machine* (RSVM) dalam melakukan klasifikasi atau memprediksi positif atau negatifnya penyakit diabetes pada seorang pasien.

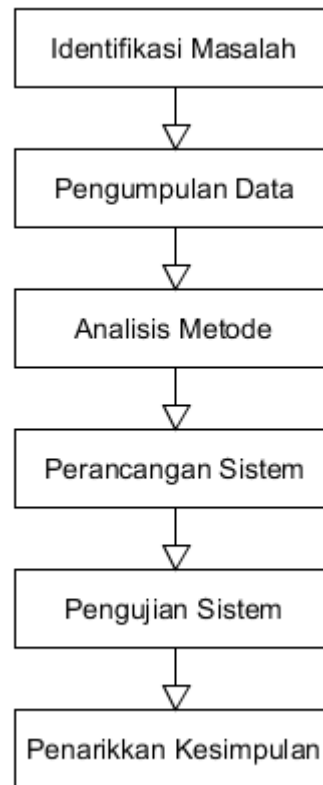
1.5 Batasan Masalah

Pembatasan masalah diperlukan agar penelitian menjadi lebih terarah dan sesuai dengan tujuan yang sudah ditetapkan sebelumnya. Oleh karena itu beberapa batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan adalah dataset “*Pima Indian Diabetes Database*” yang diambil secara online pada situs kaggle.com,
2. Data masukkan berupa *file* dengan format .csv,
3. Penelitian ini berfokus pada performa yang diberikan dari metode *Reduced Support Vector Machine* yang digunakan dalam melakukan prediksi,
4. Pembangunan sistem untuk menguji penelitian ini menggunakan bahasa pemrograman python,
5. Kernel yang digunakan adalah kernel linier.

1.6 Metode Penelitian

Metode penelitian merupakan tahapan yang dilalui oleh peneliti, mulai dari perumusan masalah hingga kesimpulan. Metode penelitian yang digunakan adalah metode penelitian kuantitatif. Metode ini bertujuan untuk menunjukkan atau membuktikan teori atau hipotesa yang sudah ditentukan sebelumnya. Metode penelitian ini digunakan sebagai pedoman untuk peneliti dalam melaksanakan penelitiannya. Hal ini dilakukan agar hasil yang dicapai tidak menyimpang dari tujuan yang telah ditentukan sebelumnya. Alur atau kerangka kerja dari metode penelitian dapat dilihat pada Gambar 1.1 berikut:



Gambar 1.1 Alur Metode Penelitian

Berikut ini adalah penjelasan dari alur metode penelitian pada

1. Identifikasi Masalah

Tahap identifikasi masalah ini merupakan tahap pertama dilakukan. Pada tahap ini dilakukan proses pengamatan terhadap penelitian terkait yang sudah pernah dilakukan sebelumnya. Tujuan dari identifikasi masalah ini adalah untuk mendefinisikan kebutuhan sistem yang akan dibangun dan tujuan yang akan dicapai.

2. Pengumpulan Data

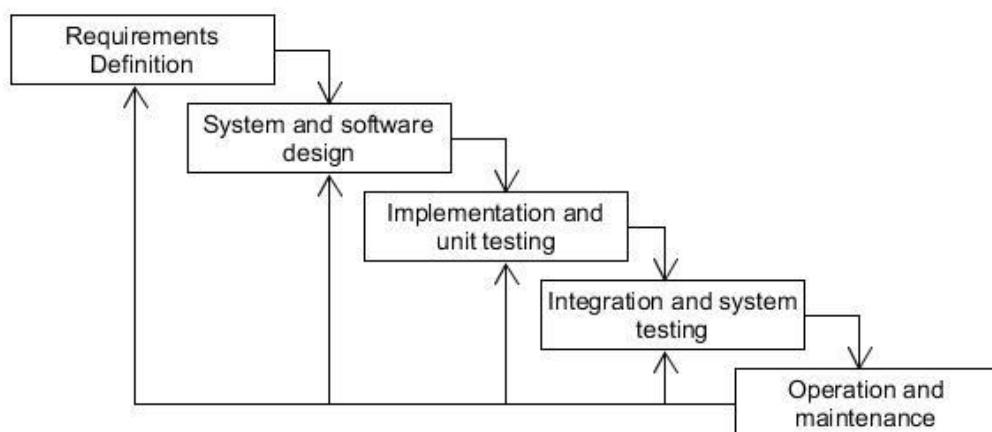
Dalam Penelitian ini, data yang digunakan adalah dataset “*Pima Indian Diabetes Database*” dari The John Hopkins University. Dateset ini diambil langsung secara online melalui situs [kaggle.com](https://www.kaggle.com).

3. Analisis Metode

Analisis adalah suatu kegiatan atau usaha yang dilakukan untuk mengamati suatu objek atau masalah yang bertujuan untuk mendapatkan pemahaman tentang suatu masalah yang dikaji. Pada tahap ini diambil sebuah hipotesa awal mengenai penelitian yang akan dilakukan.

4. Perancangan Sistem

Dalam perancangan sistem atau perangkat lunak ini dilakukan dengan menggunakan Model Air Terjun atau *Waterfall Model*. Model ini dipilih karena kelebihanannya, yaitu sederhana, tidak memerlukan biaya besar, mudah diestimasi, dan mudah diimplementasikan [10]. Untuk alur dari *waterfall* model ini dapat dilihat pada Gambar 1.2.



Gambar 1.2 Waterfall Model

Berikut ini adalah penjelasan dari alur *waterfall* model pada Gambar 1.2

A. *Requirements Analysis and Definition*

Pada tahap ini dilakukan deskripsi perilaku sistem untuk dikembangkan. Ditetapkan beberapa spesifikasi dan fitur perangkat lunak. Beberapa kebutuhan yang diperlukan dikumpulkan, dianalisis, dan kemudian dilakukan dokumentasi, yang dapat membantu lebih lanjut proses pengembangan.

B. *System and Software Design*

Pada tahap ini, informasi yang sudah dikumpulkan dari tahap sebelumnya akan dievaluasi, lalu dirumuskanlah implementasinya. Ini adalah proses perencanaan dan pemecahan masalah untuk solusi perangkat lunak. Pada tahap ini juga dilakukan pemilihan algoritma yang sesuai desain, dan desain arsitektur atau model perangkat lunak.

C. *Implementation and Unit Testing*

Pada tahap ini yang dilakukan adalah menterjemahkan desain program ke dalam kode-kode dengan menggunakan bahasa pemrograman yang sudah ditentukan.

D. *Integration and System Testing*

Tahap ini berkaitan dengan pengujian nyata dan pemeriksaan solusi yang diberikan oleh perangkat lunak yang telah dikembangkan untuk memenuhi kebutuhan. Pada tahap ini juga dilakukan pengecekan *bug* atau gangguan sistem, perbaikan, dan penyempurnaan.

E. *Operation and Maintenance*

Tahap ini adalah tahap dilakukannya modifikasi, perbaikan, koreksi kesalahan, dan penyempurnaan yang sesuai, setelah perangkat lunak sudah dirilis. Karena ada beberapa kemungkinan bahwa perangkat lunak tersebut perlu diubah atau disesuaikan sedemikian rupa. Jadi tahap ini adalah sebuah tahap mengurus masalah tersebut.

5. Pengujian Sistem

Tahapan pengujian adalah tahap melakukan pengujian terhadap sistem yang dibangun, yaitu prediksi penyakit diabetes menggunakan metode *Reduced Support Vector Machine*. Tahap pengujian ini bertujuan untuk mendapatkan hasil berupa akurasi dalam memprediksi penyakit diabetes.

6. Penarikan Kesimpulan

Pada tahap yang terakhir ini ditarik kesimpulan terhadap kinerja sistem yang telah dibangun.

1.7 Sistematika Penulisan

Sistematika penulisan disusun untuk memberikan gambaran secara umum mulai dari permasalahan hingga solusi atau pemecahannya. Sistematika penulisan yang digunakan pada penelitian ini adalah sebagai berikut.

BAB 1 PENDAHULUAN

Bab 1 ini berisi tentang latar belakang masalah, identifikasi masalah, perumusan masalah, maksud dan tujuan, batasan masalah, dan metodologi penelitian, serta sistematika penulisan sebagai gambaran dalam penelitian ini.

BAB 2 LANDASAN TEORI

Bab 2 ini menjelaskan teori-teori dasar terkait penelitian yang digunakan untuk menganalisis masalah, teori pendukung yang dipakai dalam penelitian, dan pengertian atau definisi yang dikutip dari berbagai macam sumber yang berkaitan dengan penelitian. Mulai dari penyakit diabetes, metode yang digunakan, perhitungan evaluasi, dan metode pengujian.

BAB 3 ANALISIS DAN PERANCANGAN SISTEM

Bab 3 ini menjelaskan menjelaskan secara rinci analisa dan perancangan dari sistem yang akan dibangun, meliputi analisis sistem, analisis masalah, analisis data masukan, analisis keluaran, analisis klasifikasi dan prediksi menggunakan metode *Reduced Support Vector Machine*, analisis kebutuhan non fungsional, dan perancangan antarmuka.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Bab 4 ini menjelaskan tentang implementasi, meliputi implementasi perangkat lunak, implemementasi perangkat keras, dan impelementasi antarmuka. Di sini juga dijelaskan analisa dari tahap sebelumnya yang diubah ke dalam bentuk program dengan bahasa pemrograman python. Hasil implementasi ini kemudian

akan diuji performanya untuk mengetahui seberapa besar akurasi yang diperoleh sistem.

BAB 5 KESIMPULAN DAN SARAN

Bab 5 ini menjelaskan hasil dari penelitian yang telah dilakukan berdasarkan dari pengujian. Dari hasil pengujian tersebut lalu akan ditarik kesimpulan serta saran untuk perbaikan sistem pada penelitian-penelitian berikutnya.

