

BAB I

PENDAHULUAN

Pada Bab I terdapat pembahasan tentang penjelasan latar belakang permasalahan, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

1.1 Latar Belakang

Pada era globalisasi saat ini, teknologi informasi semakin berkembang seiring berjalannya waktu. Perkembangan teknologi informasi terjadi di berbagai bidang, baik di bidang pendidikan, ekonomi hingga kesehatan. Penerapan teknologi informasi di bidang kesehatan salah satunya yaitu mengklasifikasi atau menganalisis diagnosa berbagai penyakit, misalnya diabetes melitus. Klasifikasi ini diimplementasikan pada sebuah sistem yang dapat mempermudah pihak medis untuk mengklasifikasi penyakit menggunakan data pasien sebelumnya.

Diabetes melitus atau penyakit kencing manis adalah penyakit yang memiliki ciri tingginya kadar gula darah (*hyperglykemia*) [1]. Peningkatan kadar gula dalam tubuh disebabkan karena ketidakmampuan tubuh dalam mengolah glukosa atau karbohidrat. Hal ini terjadi karena insulin tidak dapat mengerjakan fungsinya dengan baik atau juga bisa disebabkan karena berkurangnya jumlah insulin. Insulin merupakan hormon hasil produksi dari pankreas. Insulin berguna dalam membantu mengendalikan gula darah dalam tubuh. Penyakit diabetes melitus kronis dapat mengakibatkan banyak kerusakan pada organ tubuh dan dapat menjadi penyebab kematian. Terdapat empat jenis penyakit diabetes yaitu diabetes tipe 1, diabetes tipe 2, diabetes tipe 3 dan diabetes gestasional [2].

Pada tahun 2021, *International Diabetes Federation* (IDF) memprediksi bahwa 537 juta orang dengan usia antara 20 dan 79 tahun menderita diabetes di seluruh dunia. Jumlah pengidap penyakit diabetes ini diprediksi akan bertambah hingga 643 juta pada tahun 2030 dan akan bertambah menjadi 784 juta di tahun 2045. Pada tahun 2021 penyakit diabetes menjadi penyebab 6,7 juta kematian di seluruh dunia. Selain itu, IDF menyatakan bahwa dalam daftar negara dengan penderita diabetes terbanyak terbanyak di dunia, Indonesia berada di urutan kelima. Jumlah penderita diabetes di Indonesia sekitar 19,5 juta orang dan

diramalkan akan terus bertambah menjadi 28,6 juta orang di tahun 2045 [3]. Oleh karena itu, untuk mengurangi angka penderita diabetes dibutuhkan suatu sistem yang menerapkan metode klasifikasi agar dapat mengklasifikasi penyakit diabetes. Terdapat beberapa metode *supervised learning* yang dapat digunakan untuk melakukan klasifikasi. Metode tersebut yaitu *K-Nearest Neighbor* (KNN), *Support Vector Machine* (SVM), *Naïve Bayes Classifier*, *Random Forest Classifier* (RFC), *Quadratic Discriminant Analysis* (QDA), *Neural Network* (NN) dan lain-lain [4]. Penelitian klasifikasi penyakit diabetes serta penelitian mengenai metode SVM sebelumnya telah beberapa kali dilakukan. Penelitian tersebut digunakan penulis sebagai referensi dalam melakukan penelitian mengenai klasifikasi penyakit diabetes.

Metode *naïve bayes* dan SVM pernah digunakan dalam mengklasifikasi penyakit diabetes melitus [5]. Pada penelitian tersebut metode *naïve bayes* menghasilkan akurasi sebesar 92.0746%. Sedangkan pada metode SVM menggunakan kernel *polynomial* menghasilkan akurasi sebesar 96.2704%.

Penelitian [6] menggunakan 3 metode linear dan 3 metode non-linear yaitu *logistic regression*, *linear discriminant analysis*, *K-Nearest Neighbors* (KNN), *decision tree*, *gaussian naïve bayes* dan SVM dalam mengklasifikasi penyakit diabetes melitus. Hasil dari penelitian tersebut menunjukkan bahwa dari keenam metode yang diusulkan metode SVM terpilih sebagai metode yang memperoleh nilai akurasi optimal yaitu sebesar 76.3%.

Pada Tahun 2021 terdapat penelitian yang mengklasifikasikan diabetes melitus menggunakan data Pima Indian diabetes dataset [7]. Penelitian ini menggunakan metode KNN, *Naïve Bayes*, *Linear Regression*, *Random Forest* dan SVM. Dari kelima metode tersebut, metode SVM menghasilkan akurasi yang optimal yaitu sebesar 83%.

Pada penelitian lain juga menggunakan metode SVM untuk mengklasifikasi penundaan biaya kuliah mahasiswa [8]. Penelitian tersebut menghasilkan nilai akurasi sebesar 45% pada kernel *linear*. Nilai akurasi yang diperoleh dari kernel *polynomial* yaitu sebesar 65%. Sedangkan nilai akurasi yang diperoleh dari kernel *Radial Basis Function* (RBF) yaitu sebesar 97%.

Berdasarkan hasil penelitian [5], penelitian [6] dan penelitian [7] dapat disimpulkan bahwa metode SVM menghasilkan nilai akurasi yang lebih unggul daripada metode *naïve bayes*, KNN dan beberapa metode lainnya dalam mengklasifikasi penyakit diabetes. Pada metode SVM terdapat beberapa kernel yang umum digunakan diantaranya yaitu kernel *linear*, kernel *polynomial* dan kernel RBF. Kernel RBF juga dapat disebut dengan kernel *Gaussian* [9].

Berdasarkan penjelasan diatas, pada penelitian ini penulis menggunakan metode SVM dalam mengklasifikasi penyakit diabetes melitus. Penulis juga menerapkan 3 fungsi kernel yaitu kernel *linear*, kernel *polynomial* dan kernel RBF [8] untuk mengetahui pengaruh penggunaan kernel pada SVM terhadap akurasi. Pada dataset penyakit diabetes melitus masih terdapat *missing value*. Adanya *missing value* dapat menyebabkan berkurangnya keakuratan suatu data [10]. Oleh karena itu, penulis menerapkan teknik imputasi menggunakan nilai rata-rata untuk mengatasi *missing value* pada dataset tersebut. Penulis berharap dengan adanya penelitian ini dapat mengetahui hasil akurasi dari metode yang diusulkan dalam mengklasifikasi penyakit diabetes melitus.

1.2 Rumusan Masalah

1.2.1 Permasalahan

Berdasarkan uraian dari latar belakang diatas, maka permasalahan pada penelitian ini yaitu kurang optimalnya hasil akurasi dalam mengklasifikasi penyakit diabetes melitus menggunakan beberapa metode klasifikasi seperti *naïve bayes*, *decision tree* dan KNN berdasarkan penelitian terdahulu.

1.2.2 Metode Usulan

Metode yang diusulkan pada penelitian ini yaitu menggunakan metode SVM dengan melakukan pengujian fungsi kernel *linear*, kernel *polynomial* dan kernel RBF.

1.2.3 Pertanyaan Penelitian

Berdasarkan uraian diatas, maka pertanyaan pada penelitian ini yaitu bagaimana hasil perbandingan akurasi menggunakan kernel *linear*, kernel *polynomial* dan kernel RBF pada metode SVM dalam mengklasifikasi penyakit diabetes melitus?

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan masalah, diantaranya yaitu:

1. Penulis menggunakan dataset penyakit diabetes melitus (*diabetes dataset*) sebanyak 768 *record* yang diperoleh dari *website kaggle* <https://www.kaggle.com/datasets/mathchi/diabetes-data-set>
2. Metode usulan yang digunakan yaitu SVM.
3. Fungsi kernel yang digunakan yaitu kernel *linear*, kernel *polynomial* dan kernel RBF.

1.4 Tujuan Penelitian

Berdasarkan pertanyaan pada uraian rumusan masalah di atas, tujuan dari penelitian ini yaitu menghitung dan mengetahui hasil perbandingan akurasi metode SVM menggunakan kernel *linear*, kernel *polynomial* dan kernel RBF dalam mengklasifikasi penyakit diabetes melitus.

1.5 Manfaat Penelitian

1. Manfaat Bagi Pihak Medis

Manfaat penelitian bagi pihak medis yaitu sebagai berikut.

- a. Membantu pihak medis dalam mengklasifikasi penyakit diabetes melitus pada pasien.
- b. Meminimalisir kesalahan dalam diagnosis penyakit diabetes melitus.

2. Manfaat Bagi Pembaca

Berikut merupakan beberapa manfaat bagi pembaca.

- a. Dapat meningkatkan pengetahuan dan wawasan pembaca mengenai metode SVM.
- b. Dapat mengetahui hasil akurasi perbandingan metode SVM menggunakan kernel *linear*, kernel *polynomial* dan kernel RBF.
- c. Sebagai referensi bagi pembaca untuk melakukan penelitian mengenai klasifikasi.

1.6 Sistematika Penulisan

Adapun sistematika penulisan pada penelitian ini yaitu sebagai berikut.

1. Bab I Pendahuluan

Pada bab I membahas tentang latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.

2. Bab II Landasan Teori

Pada bab II membahas tentang penelitian terdahulu yang mengimplementasikan metode SVM serta membahas tentang teori pembahasan yang digunakan untuk menyelesaikan penelitian ini.

3. Bab III Metode Penelitian

Pada bab III membahas tentang kebutuhan sistem, tahapan penelitian dan algoritma metode SVM untuk mengklasifikasi penyakit diabetes melitus.

4. Bab IV Hasil dan Pembahasan

Pada bab IV berisi pembahasan tentang pengelolaan data awal, hasil proses pembagian data, hasil proses tuning parameter menggunakan *gridsearchCV*, hasil skenario uji coba dan hasil analisis skenario uji coba.

5. Bab V Penutup

Pada bab V berisi tentang kesimpulan dan saran untuk pengembangan penelitian selanjutnya.