

BAB I

PENDAHULUAN

1.1 Latar Belakang

Diabetes Mellitus (DM) merupakan penyakit atau gangguan metabolik kronis yang disebabkan oleh berbagai faktor dan ditandai oleh peningkatan kadar gula, tekanan darah serta gangguan pada metabolisme karbohidrat, lemak, dan protein akibat ketidak cukupan fungsi insulin (Kemkes, 2024). Pada penderita diabetes, kekurangan atau gangguan insulin menimbulkan gejala seperti lesu, haus, sering buang air kecil, dan penglihatan kabur. Diabetes juga merupakan penyakit metabolik yang ditandai dengan hiperglikemia kronis yang dapat menyebabkan kerusakan organ seperti mata, *stroke*, ginjal, saraf, jantung, dan pembuluh darah. Selain itu, diabetes mellitus adalah salah satu masalah kesehatan global yang berkembang paling cepat dan terus meningkat, berkontribusi pada tinggi nya angka kematian (Ghaida, 2024).

Ginting et al (2022), menyatakan bahwa jumlah data penderita diabetes di Indonesia telah melampaui 10,8 juta pada tahun 2020, dan angka ini diprediksi akan terus bertambah setiap tahun. Banyak pasien tidak menyadari kondisi diabetes, sehingga penyakit ini berkembang menjadi komplikasi serius, seperti penyakit kardiovaskular dan kerusakan saraf, yang akhirnya meningkatkan angka morbiditas dan mortalitas.

Penelitian ini bertujuan untuk mengklasifikasikan diabetes mellitus (DM) menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest*. Proses klasifikasi ini dirancang untuk mengidentifikasi pola yang dapat digunakan dalam membangun model prediksi berdasarkan variabel tertentu, guna menentukan apakah seseorang menderita diabetes atau tidak (Abdurrahman, 2022).

Penelitian tentang klasifikasi penyakit DM sangat penting, terutama di Indonesia, di mana prevalensi diabetes terus meningkat. Meningkatnya jumlah penderita diabetes menyoroti pentingnya interensi dini melalui klasifikasi berbasis teknologi, seperti penggunaan algoritma *Machine Learning* yang efektif untuk klasifikasi diabetes karena, mampu mengelola berbagai variabel dan data kompleks secara efisien (Ginting et al., 2022).

Berikutnya Penelitian oleh Siregar [10], membandingkan tiga algoritma klasifikasi dalam prediksi cuaca, yaitu Naïve Bayes, Decision Tree, dan Random Forest. Dari hasil yang diperoleh, akurasi Naïve Bayes sebesar 72,22%, Decision Tree sebesar 54,01%, dan Random Forest mencatat akurasi tertinggi sebesar 82,38%.

Maulana (2024), melakukan optimasi parameter *Support Vector Machine* dengan *Particle Swarm Optimization* untuk peningkatan klasifikasi diabetes, di dapatkan hasil analisis bahwa model SVM yang dioptimasi PSO mengalami peningkatan akurasi dari 74,38% menjadi 74,57%, presisi dari 74,82% menjadi 74,99%, *recall* dari 74,38% menjadi 74,57%, dan *F1-score* dari 74,34% menjadi 74,54%. Peningkatan sekitar 0,2% ini menunjukkan bahwa meskipun kecil, PSO mampu menemukan parameter yang lebih optimal. Penelitian lainnya oleh Desiani et al (2024), membandingkan algoritma SVM dan *Naïve Bayes* dalam klasifikasi penyakit diabetes menggunakan dua teknik evaluasi, yaitu *persentase split* dan *k-fold cross-validation*. Pada pengujian menggunakan *persentase split*, algoritma SVM mencapai akurasi sebesar 77,27%, sementara *Naïve Bayes* memperoleh akurasi sebesar 79%. Sedangkan pada pengujian dengan *k-fold cross-validation*, SVM menghasilkan akurasi sebesar 71%, sementara *Naïve Bayes* memperoleh akurasi sebesar 75%. Selanjutnya penelitian oleh Khairul Huda (2024), membandingkan beberapa algoritma *Naïve Byes*, *Regresi Logistik*, *Random Forest*, SVM, dan KNN untuk prediksi diabetes, *Regresi Logistik* akurasi 78%, presisi 77%, *recall* 73%, dan *F1-score* 74%, lalu untuk RF nilai akurasi nya 83%, presisi 82%, *recall* 80%, dan *F1-score* 81%, sedangkan KNN akurasi 80%, presisi 79%, *recall* 75%, dan *F1-score* 77%, dan SVM nilai akurasi sebesar 81%, presisi 81%, *recall* 76%, dan *F1-score* 77%, *Naïve Bayes* dengan nilai akurasi 77%, presisi 76%, *recall* 74%, dan *F1-score* 74%.

Pahlevi et al (2021), juga melakukan penelitian dengan menerapkan algoritma *Iterative Dichotomiser Tree* dan *Support Vector Machine* pada klasifikasi penyakit diabetes mellitus tipe 2. Algoritma SVM menggunakan *kernel linear* dan RBF. Algoritma SVM dengan *kernel linear* menghasilkan akurasi sebesar 78,5%, SVM dengan *kernel RBF* sebesar 78% dan hasil klasifikasi yang terendah oleh ID3 sebesar akurasi 74%. Selanjutnya penelitian oleh Argina (2020),

melakukan penerapan metode klasifikasi *K-Nearest Neighbor* pada dataset penderita penyakit diabetes, dengan hasil akhir penelitian ini, telah dihitung akurasi tertinggi 39% pada $K=3$, presisi tertinggi 65% pada $K=3$ dan $K=5$, *recall* tertinggi 36% pada $K=3$, dan *F-Measure* tertinggi 46% pada $K=3$. Sementara itu penelitian oleh Rohana et al (2024), membandingkan tiga metode yaitu *Multilayer Perceptron*, *Support Vector Machine*, dan *Decision Tree*, hasil penelitian menunjukkan bahwa *Multilayer Perceptron* memiliki akurasi sebesar 97,1%, *Support Vector Machine* 98,9%, dan *Decision Tree* memiliki akurasi 100%. Berikutnya Diana Dewi et al (2023), melakukan penelitian tentang, perbandingan metode *neural network* dan *support vector machine* dalam klasifikasi diagnosa penyakit diabetes, hasil penelitian menunjukkan bahwa metode ANN memiliki keunggulan dalam mengklasifikasikan diabetes dengan akurasi 77,60% dibandingkan SVM dengan akurasi 65,24%. Selanjutnya penelitian yang dilakukan oleh Siregar et al (2020), membandingkan tiga algoritma klasifikasi dalam prediksi cuaca, yaitu *Naïve Bayes*, *Decision Tree*, dan *Random Forest*. Dari hasil yang diperoleh, akurasi *Naïve Bayes* sebesar 72,22%, *Decision Tree* sebesar 54,01%, dan *Random Forest* mencatat akurasi tertinggi sebesar 82,38%.

Berdasarkan tinjauan penelitian sebelumnya, belum ditemukan penelitian yang secara langsung membandingkan kinerja algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dalam klasifikasi penyakit diabetes mellitus. Oleh karena itu, penelitian ini bertujuan untuk membangun model klasifikasi berdasarkan data pasien dan algoritma mana yang tingkat akurasi nya lebih optimal melalui perbandingan kedua algoritma tersebut. Evaluasi kinerja dilakukan menggunakan empat metrik utama, yaitu akurasi, presisi, *recall*, dan *F1-score*. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan identifikasi pada penyakit diabetes, serta menjadi acuan dalam pengambilan keputusan medis dini guna mengurangi risiko komplikasi dan angka kematian akibat diabetes.

1.2 Rumusan Masalah

Berikut adalah rumusan dalam penelitian ini:

1. Bagaimana membangun model *Support Vector Machine* dan *Random Forest* dalam klasifikasi penyakit diabetes?

2. Bagaimana hasil evaluasi dari kinerja algoritma *Support Vector Machine* dan *Random Forest* dalam Klasifikasi penyakit diabetes mellitus?

1.3 Tujuan Penelitian

Adapun tujuan pada penelitian ini sebagai berikut:

1. Membangun model *Support Vector Machine* dan *Random Forest* untuk klasifikasi penyakit diabetes berdasarkan data pasien.
2. Mengetahui hasil evaluasi dari kinerja algoritma *Support Vector Machine* (SVM) dan *Random Forest* dalam klasifikasi penyakit diabetes mellitus secara menyeluruh.

1.4 Manfaat

Setelah penelitian ini dilakukan, diharapkan dapat memberikan manfaat dan dampak positif bagi berbagai pihak. Berikut adalah manfaat dari penelitian ini:

1. Penelitian ini menjelaskan tentang proses pembangunan model klasifikasi penyakit diabetes menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF), sehingga dapat menjadi referensi dalam penerapan algoritma *machine learning* untuk mengolah data medis secara efektif dan efisien.
2. Penelitian ini bertujuan untuk mengetahui hasil evaluasi kinerja algoritma SVM dan *Random Forest* berdasarkan metrik evaluasi seperti akurasi, presisi, *f1-score*, dan *recall*, sehingga algoritma ini dapat diterapkan pada sistem pendukung keputusan medis.