

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Stroke merupakan salah satu masalah kesehatan yang serius. Terjadinya stroke karena ketidaknormalan fungsi sistem saraf pusat dimana pasokan aliran darah yang menuju ke otak terganggu. Penyakit Stroke menjadi penyebab kematian kedua dan menjadi penyebab kecacatan yang keenam yang paling umum terjadi [1].

Di Indonesia penyakit stroke dalam beberapa tahun terakhir mengalami peningkatan, akan tetapi pemerintah masih belum memiliki solusi dalam menyelesaikan permasalahan penyakit tersebut. Padahal proses penanganan penyakit stroke butuh waktu yang cukup dalam pendeteksiannya, apabila tidak ditangani dengan cepat bisa mengakibatkan cacat atau bahkan kematian. Dengan adanya permasalahan tersebut maka dibutuhkan sebuah sistem yang dapat mempercepat dan mempermudah dalam mendeteksi resiko untuk mengurangi angka seseorang terserang penyakit stroke [2].

Peneliti sebelumnya terkait dengan memprediksi penyakit stroke menggunakan metode *Deep Neural Network* mendapatkan hasil akurasi 71,6% dan sensitifitas sebesar 67,4%. Dengan mengatasi *imbalance class* menggunakan metode *SMOTE* dan *ADASYN* dalam meningkatkan performa klasifikasi herregistrasi mahasiswa baru mendapatkan hasil *SMOTE* lebih tinggi dari *ADASYN* dalam menyeimbangkan data [3].

Penelitian selanjutnya melakukan penelitian dengan judul “Klasifikasi Penyakit Kanker Serviks Menggunakan Algoritma *Support Vector Machine* (SVM). Pada penelitian menggunakan perbandingan 80:20. Pada penelitian menggunakan *tools orange* sebagai *data test* dan *python* digunakan sebagai *data train*. Hasil pengujian dengan *tools orange* menghasilkan *akurasi* 92.9% dan dengan *python* menghasilkan nilai *akurasi* 87% [4].

Penelitian selanjutnya telah meningkatkan *akurasi prediksi* dengan berbagai metode namun masih terbatas dalam pemilihan fitur optimal dan efisiensi pemodelan. Oleh karena itu, penelitian ini bertujuan untuk menganalisis kinerja kernel pada algoritma *Support Vector Machine* (SVM) seperti *Dot*, *Radial*, *Polynomial* dan menganalisis efektivitas *Forward Selection* (FS) dalam memilih fitur paling signifikan guna mengoptimalkan prediksi risiko gagal jantung. Hasil penelitian menunjukkan bahwa kernel *Radial* dengan FS memiliki performa terbaik dengan AUC 0.881, *Accuracy* 84,64%, dan *Recall* 92,55% [5].

Penelitian selanjutnya melakukan dua algoritma dengan menerapkan algoritma *Support Vector Machine* (SVM) dan C4.5. Hasilnya menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memiliki tingkat akurasi sebesar 87% dan dianggap sebagai algoritma yang paling tepat dan akurat untuk memprediksi penyakit jantung [6].

Support Vector Machine (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. *Support Vector Machine* (SVM) memiliki prinsip dasar *linier classifier* yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun *Support Vector Machine* (SVM) telah

dikembangkan agar dapat bekerja pada *problem non-linier* dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada ruang berdimensi tinggi, akan dicari *hyperplane* yang dapat memaksimalkan jarak (*margin*) antara kelas data [7].

SVM salah satu set metode pembelajaran yang diawasi terkait yang digunakan dalam diagnosis medis untuk klasifikasi dan regresi. SVM secara bersamaan meminimalkan kesalahan klasifikasi empiris dan memaksimalkan *margin* geometrik. Jadi SVM disebut Pengklasifikasi *Margin* Maksimum. SVM merupakan algoritma umum yang didasarkan pada batasan risiko terjamin dari teori pembelajaran statistik yaitu yang disebut prinsip minimisasi risiko struktural [8].

Machine learning dapat didefinisikan sebagai aplikasi komputer dan algoritma matematika yang diadopsi dengan cara pembelajaran yang berasal dari data dan menghasilkan *prediksi* di masa yang akan datang. Adapun proses pembelajaran yang dimaksud adalah suatu usaha dalam memperoleh kecerdasan yang melalui dua tahap antara lain latihan (*training*) dan pengujian (*testing*) [9].

Dengan berkembangnya teknologi, khususnya di bidang *machine learning*, penggunaan model berbasis algoritma untuk membantu proses diagnosa mulai meningkat. *Support Vector Machine* (SVM) merupakan salah satu algoritma yang banyak digunakan dalam klasifikasi data, terutama pada kasus dengan data yang memiliki dimensi yang tinggi dan memerlukan klasifikasi yang akurat [10]

Berdasarkan data *Global Burden of Disease* yang dirilis oleh *American Heart Association* dan *World Stroke Organization*, jumlah penderita stroke di dunia

masih menunjukkan angka yang sangat tinggi dan cenderung meningkat setiap tahunnya.

Pada tahun 2020, tercatat sekitar 89,1 juta orang hidup dengan kondisi stroke, dengan rincian 68,2 juta mengalami stroke iskemik, 18,9 juta mengalami perdarahan otak (*intracerebral hemorrhage*), dan 8,1 juta mengalami *subarachnoid hemorrhage*. Pada tahun yang sama, stroke juga menjadi penyebab kematian sekitar 7,1 juta jiwa.

Kemudian pada tahun 2021, jumlah penderita stroke meningkat menjadi 93,8 juta orang di seluruh dunia, dengan kasus stroke iskemik sebanyak 69,94 juta, *intracerebral hemorrhage* sebanyak 16,60 juta, dan *subarachnoid hemorrhage* sebanyak 7,85 juta. Jumlah kematian akibat stroke juga meningkat menjadi sekitar 7,25 juta jiwa.

Untuk tahun 2022 hingga 2025, data resmi terperinci masih dalam proses pengumpulan dan publikasi. Namun, para peneliti memperkirakan bahwa tren ini akan terus meningkat seiring bertambahnya populasi lansia di dunia. Bahkan, sebuah laporan memprediksi bahwa kematian akibat stroke dapat mencapai 9,7 juta jiwa pada tahun 2050 jika tidak ada upaya pencegahan yang lebih efektif.

Dengan data tersebut dapat disimpulkan bahwa stroke masih menjadi salah satu masalah kesehatan global utama, sehingga penelitian terkait deteksi dini dan pencegahan, termasuk melalui penerapan machine learning, sangat relevan untuk dilakukan.

1.2 RUMUSAN MASALAH

Berdasarkan dari uraian latar belakan diatas, maka didapatkan rumusan masalah pada penelitian ini sebagai berikut:

1. Bagaimana menerapkan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan risiko penyakit stroke?
2. Bagaimana mengevaluasi kinerja algoritma SVM dalam mengklasifikasikan risiko penyakit stroke ditinjau dari *metrik evaluasi* seperti *akurasi*, *presisi*, *recall*, dan *F1-score*?

1.3 BATASAN MASALAH

Untuk mempersempit fokus penulis terhadap penelitian ini, berikut adalah batasan masalah yang terdapat pada penelitian:

1. Bahasa pemrograman yang digunakan adalah *python* dengan menggunakan *google colab*.
2. Data yang digunakan dalam penelitian ini adalah data penyakit stroke.
3. Dataset diambil dari *kaggle*, yang berjumlah 5.111 data yang terdiri dari *id*, *gender*, *age*, *hypertension*, *heart_disease*, *ever_married*, *work_type*, *residence_type*, *avg_glucose_level*, *bmi*, *smoking_status*, *stroke*.
4. Menggunakan metode *Support Vector Machine*.

1.4 TUJUAN PENELITIAN

Adapun tujuan yang dapat dicapai dalam penelitian ini, yaitu:

1. Menerapkan algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan tingkat risiko penyakit stroke berdasarkan data medis pasien.
2. Mengevaluasi kinerja algoritma SVM dalam klasifikasi risiko penyakit stroke dengan menggunakan metrik evaluasi seperti *akurasi*, *presisi*, *recall*, dan *F1-score*.

1.5 MANFAAT PENELITIAN

Adapun manfaat yang dapat dicapai dalam penelitian ini, yaitu:

1. Membantu tenaga medis dalam menganalisis risiko penyakit jantung dengan cepat.
2. Memberikan referensi bagi penelitian selanjutnya dalam bidang kesehatan berbasis *Machine Learning*.
3. Penelitian ini diharapkan dapat menjadi salah satu cara penyelesaian dalam menentukan prediksi seseorang terkena penyakit stroke dengan menggunakan *machine learning*.
4. Menambah pengetahuan bagi mahasiswa dan pembaca mengenai *machine learning* dan penggunaannya.

1.6 SISTEMATIKA PENULISAN

Untuk memberikan gambaran umum mengenai keseluruhan bab yang akan dibahas dapat dilihat dalam sistematika penulisan berikut:

BAB I PENDAHULUAN

Bab ini menjelaskan mengenai latar Belakang Masalah, Perumusan Masalah, Batasan masalah, Tujuan Penelitian, Manfaat Penelitian dan Sistematika Penulisan.

BAB II LANDASAN TEORI

Pada bab ini berisikan tentang teori-teori terkait yang berhubungan dengan masalah yang diteliti dari pokok permasalahan stroke diangkat, yakni Klasifikasi Resiko Penyakit Stroke Menggunakan *Support Vector Machine* (SVM). Dengan poin-poin yang tertera pada penelitian yaitu stroke, *machine learning*, algoritma *Support Vector Machine*. Bahasa pemrograman yang digunakan. Serta penelitian sejenis yang dapat mendukung tugas akhir ini.

BAB III METODE PENELITIAN

Bab ini berisi langkah penelitian, waktu penelitian dan tempat penelitian, alat yang digunakan hingga prosedur penelitian dari awal sampai akhir penelitian.

BAB IV ANALISIS DAN HASIL

Dalam bab ini penulis memparkan analisis data yang diperoleh selama penelitian dan hasil dari analisis tersebut. Bab ini juga mencakup interpretasi hasil serta diskusi mengenai temuan penelitian.

BAB V PENUTUP

Dalam bab ini penulis memberikan kesimpulan dari hasil penelitian dan menyampaikan saran-saran yang relevan untuk penelitian lebih lanjut atau penerapan praktis dari temuan penelitian.

Dengan sistematika penulisan ini, diharapkan pembaca dapat dengan mudah mengikuti alur penelitian dan memahami setiap bagian dari laporan penelitian yang disusun.