

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Stroke adalah gangguan fungsional yang terjadi secara mendadak disebabkan karena kurangnya atau terputusnya aliran darah yang mengalir ke otak akibat adanya gumpalan darah, endapan plak, atau karena pecahnya pembuluh darah akibat tekanan darah yang tinggi secara tiba-tiba ke otak [1]. Sel otak akan rusak ketika tidak menerima oksigen dan nutrisi dari darah. Gejala *stroke* terjadi secara tiba-tiba dengan kelemahan pada sisi tubuh [2]. Seperti gangguan fungsional, gangguan *motoric*, psikologis atau perilaku dimana gejala yang paling khas adalah *hemiparesis*, kelemahan *ekstremitas* sisi, hilangnya sensasi wajah, kesulitan bicara dan kehilangan penglihatan sisi [3]. Penyakit *stroke* berdampak pada aktivitas seseorang karena kejadian seperti kelumpuhan, kecacatan, gangguan komunikasi, gangguan emosi, nyeri, gangguan tidur, depresi, disfasia dan sebagainya [4]. Namun keberhasilan penanganan *Stroke* dapat terjadi jika penanganan dilakukan secara tepat, cepat, dan cermat pada penanganan awal pasien [5]. *Stroke* juga merupakan salah satu penyebab kematian dan kecacatan yang cukup besar di seluruh dunia [6].

Menurut data WHO *stroke* merupakan salah satu dari tiga besar penyakit mematikan di dunia, diantara penyakit berbahaya lainnya seperti kanker dan jantung [7]. Pada tahun 2022, terdapat 12,2 juta kasus baru di dunia, 62% diantaranya terjadi pada orang di bawah umur 70 tahun. Jumlah absolut orang yang

memiliki *stroke*, meninggal, atau menjadi disabilitas karena *stroke*, telah meningkat [8]. Diperkirakan ada 4,5 juta kematian pertahun akibat *stroke* di dunia dan lebih dari 9 juta penderita *stroke*. Resiko kekambuhan selama 5 tahun adalah 15-40%. Diperkirakan pada tahun 2023 akan ada absolut peningkatan jumlah pasien yang mengalami pertama kali *stroke* meningkat sekitar 30% dibandingkan dengan 1983 [9]. Salah satu penyebab kematian pada pasien *stroke* adalah *stroke* berulang. Serangan *stroke* berulang masih sangat mungkin terjadi dalam kurung waktu 6 bulan pasca *stroke* pertama [10]. Masalah yang sering ditimbulkan serangan *stroke* berulang adalah kecacatan permanen bahkan dapat mengancam nyawa [11].

Banyaknya kasus *stroke*, membuat prediksi penyakit *stroke* dibutuhkan sebagai pencegahan dan pendeteksian resiko penyakit *stroke* sejak dini. Dengan adanya pendeteksian penyakit *stroke* dapat dilakukan dengan cepat dan tepat untuk memperbesar peluang penyembuhan [12]. Salah satu teknik yang digunakan yaitu data mining. Data mining adalah serangkaian proses untuk mengenali nilai tambahan terutama informasi yang selama ini tidak diketahui secara manual dari suatu basis [13]. Data mining memiliki salah satu metode yang dapat digunakan untuk mengidentifikasi *stroke* yaitu klasifikasi. Klasifikasi adalah proses pengelompokan objek yang memiliki karakteristik atau ciri yang sama ke dalam beberapa kelas [14]. Klasifikasi merupakan salah satu metode untuk mengidentifikasi permasalahan yang di peroleh berdasarkan kelas dilakukan untuk deteksi penyakit *stroke*. Sebagai referensi dalam melakukan penelitian ini, peneliti menggunakan tiga acuan dataset *stroke* sebelumnya.

Menurut penelitian [15], *Stroke* adalah kondisi berbahaya yang perlu ditangani secepatnya, karena sel otak dapat mati dalam hitungan menit. Penanganan yang cepat dapat mengurangi tingkat kerusakan pada otak maka perlu dilakukan prediksi penyakit *stroke*, salah satu cara untuk memprediksi penyakit *stroke* yaitu klasifikasi menggunakan algoritma *K-Nearest Neighbor* dan *Naïve Bayes* dengan hasil akurasi *K-Nearest Neighbor* 68,30% presensi 67,20% dan *recall* 73,34% sedangkan hasil akurasi *naïve bayes* sebesar 74,45% presensi sebesar 74,01% dan *recall* sebesar 75,71%. Dapat disimpulkan bahwa algoritma *naïve bayes* lebih baik dibandingkan *K-Nearest Neighbor*.

Pada penelitian [16], dilakukan klasifikasi *stroke* menggunakan algoritma *Support Vector Regression* dan *Decision Tree C4.5* mendapatkan hasil algoritma *Decision Tree C4.5* sangat bagus dalam memprediksi penyakit *stroke* dengan nilai *root mean square error* yang dihasilkan sebesar 0.235 lebih kecil dibandingkan *Support Vector Regression* yang menghasilkan *root mean square error* sebesar 0.3999.

Berdasarkan pada penelitian [17] teknologi sangat berkembang pesat, sampai manusia sudah hampir terikat sepenuhnya oleh teknologi. Bahkan dibidang kesehatan juga menggunakan teknologi yang canggih untuk membantu menyelesaikan pekerjaan. Penelitian ini menggunakan algoritma *naïve bayes* untuk mempermudah prediksi penyakit *stroke* dengan hasil akurasi *stroke* 0,293 dan non *stroke* 0.706.

Berdasarkan beberapa ulasan diatas *stroke* adalah gangguan fungsional mendadak akibat kurangnya aliran darah ke otak, *stroke* juga salah satu penyebab kematian kecacatan di seluruh dunia. Data dari WHO menunjukkan bahwa *stroke* adalah salah satu dari tiga penyakit mematikan di dunia. Oleh karena itu prediksi penyakit *stroke* menjadi penting untuk pencegahan dan deteksi dini. Salah satu teknik untuk memprediksi penyakit *stroke* yaitu data mining. Untuk mengklasifikasi penyakit *stroke* menggunakan beberapa algoritma seperti *naïve bayes* dan *decision tree C4.5*. Banyak algoritma klasifikasi data mining yang telah digunakan, namun tidak semua parameter kinerja juga dibahas. Oleh karena itu, penelitian ini akan melakukan pengujian kinerja algoritma *C4.5* dan *naïve bayes* dengan beberapa parameter kinerja seperti *accuracy*, *TPR*, *FPR*, *Precision*, *ROC*, dan *RMSE*. Pengujian parameter ini di perlukan untuk mendapatkan gambaran kinerja algoritma secara menyeluruh khususnya dalam mengklasifikasi penyakit *stroke*.

1.2 RUMUSAN MASALAH

Berdasarkan latar belakang masalah yang telah diuraikan di atas, maka Rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana penerapan *Naïve Bayes* dan *Decision Tree C4.5* untuk klasifikasi resiko penyakit *stroke*.
2. Bagaimana perbandingan kinerja algoritma *Naïve Bayes* dan *Decision Tree C4.5* dalam mengklasifikasi penyakit *stroke*, berdasarkan parameter *accuracy*, *TPR*, *FPR*, *Precision*, *ROC* dan *RMSE*.

1.3 BATASAN MASALAH

Adapun Batasan masalah dalam penelitian ini, yaitu:

1. Data yang diperlukan untuk penelitian ini adalah *stroke prediction dataset* yang di dapat secara *online* dari *website* penyediaan dataset kaggle <https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-brain-stroke-dataset>.
2. *Feature* yang digunakan dalam penelitian ini adalah jenis kelamin, usia, hipertensi, *heart_disease*, memiliki penyakit jantung, *work_type*, *residence_type*, *avg_glucose_level*, *BMI*, *Smoking*, *stroke* dengan jumlah data.
3. Tool yang digunakan untuk penelitian ini WEKA.
4. Pengujian menggunakan metode data split dari *10 fold cross validation*.

1.4 TUJUAN DAN MANFAAT

1.4.1 Tujuan Penelitian

Adapun tujuan penelitian yang dilakukan oleh penulis, yaitu:

1. Untuk mengetahui hasil klasifikasian penyakit *stroke* pada dataset yang didapat.
2. Untuk mengetahui perbandingan kinerja algoritma *Naïve Bayes* dan *Decision Tree C4.5* dalam klasifikasi penyakit *stroke*. Berdasarkan parameter *accuracy*, TPR, FPR, *Precision*, ROC dan RMSE.

1.4.2 Manfaat Penelitian

Manfaat dari penulisan ini adalah:

1. Dapat membantu meningkatkan akurasi diagnosis, memungkinkan tindakan medis lebih cepat, merencanakan pengobatan yang tepat, mengelola resiko, dan mendukung penelitian kesehatan sehingga tindakan perawatan pasien dan peluang pencegahan lebih baik.
2. Perbandingan akurasi algoritma *naïve bayes* dan *decision tree C4.5* dalam klasifikasi penyakit *stroke* membantu perbaikan dignositik, dukungan keputusan klinis, dan memperkuat kepercayaan pada prediksi, serta pengoptimal kinerja sistem.

1.5 SISTEMATIKA PENULISAN

Agar dapat membantu dalam pemaparan penulisan laporan penelitian ini, maka penulis menyusun sebuah sistematika penulisan penelitian ini sebagai berikut:

BAB I : PENDAHULUAN

Pada tahapan awal bab ini berisi tentang latar belakang masalah penelitian, berikutnya dilanjutkan oleh perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan yang terakhir sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab ini berisi sebuah teori-teori atau konsep-konsep yang *relevan* sebagai landasan atau kerangka untuk mendapatkan

jawaban masalah dalam penelitian seperti teori tentang metode *Naïve Bayes*, *Decision Tree C4.5*, WEKA, *stroke*, apa saja yang di gunakan dalam mendorong penelitian ini dan lainnya sebagainya. Pada bab ini permasalahan dipusatkan pada literatur-literatur yang membahas konsep teori yang saling berkaitan dengan tujuan dan rumusan masalah penelitian.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan tentang karangka kerja penelitian, metode pengumpulan data, metode *naïve bayes*, *Decision Tree C4.5*, dan yang terakhir alat pendukung yang digunakan pada penelitian ini.

BAB IV : ANALISIS DAN IMPLEMENTASI

Pada tahap ini dilakukan analisa menggunakan metode *naïve bayes* dan *Decision Tree C4.5* terhadap data *stroke* dan merupakan tampilan hasil analisis dari WEKA yang digunakan. Hasil analisa selanjutnya dilakukan perbandingan antara dua algoritma tersebut

BAB V : PENUTUP

Pada bab penutup ini terdiri atas kesimpulan dan saran yang berkaitan dengan hasil penelitian yang penulis lakukan.