

BAB 1

PENDAHULUAN

1.1 Latar Belakang Masalah

Strok atau *Cerebrovascular Disease* merupakan sebuah penyakit tidak menular yang disebabkan oleh berkurang atau terhambatnya aliran darah dan oksigen ke otak secara tiba-tiba, sehingga terjadi gangguan atau kerusakan pada otak [1]. Menurut Direktur Pencegahan dan Pengendalian Penyakit Tidak Menular Kemenkes, Siti Nadia Tarmizi, pasien strok mengalami tren peningkatan jumlah pasien. Hal ini didukung dengan adanya data Riset Kesehatan Dasar pada 2023, terdapat prevalensi strok sebesar 8,3 persen di Indonesia [2]. Tingginya penderita strok di Indonesia disebabkan oleh gaya hidup yang tidak sehat seperti merokok, mengonsumsi alkohol, hingga tidak melakukan kegiatan olahraga [1].

Berdasarkan pada penelitian terdahulu yang berjudul "Hubungan Usia dan Jenis Kelamin Terhadap Jenis *Stroke* di IGD RSPON Jakarta" pada tahun 2021 yang dilakukan di RS Pusat Otak Nasional, jumlah pasien strok dengan jenis kelamin laki-laki terdapat 3001 orang, sedangkan jumlah pasien strok dengan jenis kelamin perempuan terdapat 1684 orang. Kemudian, berdasar dari sumber tersebut, sebaran kelompok usia yang mengalami strok terbanyak yaitu pada usia 55-64 tahun atau 34,6%, sedangkan sebaran kelompok usia yang mengalami strok tersedikit yaitu pada usia kurang dari 15 tahun atau 7% [3].

Dalam melakukan klasifikasi data, terdapat beberapa algoritma yang dapat digunakan, salah satunya adalah algoritma *Decision Tree*. Algoritma *Decision Tree* merupakan algoritma yang bersifat *supervised* dan memiliki struktur seperti pohon. Algoritma *Decision Tree* mampu menangani data dengan berbagai jenis atribut seperti kategori dan numerik, serta mudah untuk diinterpretasikan karena struktur pohon yang dihasilkan dapat divisualisasikan dengan jelas [4]. Selain itu, algoritma *Decision Tree* juga efektif dalam menangani *dataset* yang besar dan fleksibel dalam menangani masalah klasifikasi seperti pada penelitian yang berjudul "*Heart Disease Prediction Using Decision Tree and SVM*" yang menunjukkan bahwa algoritma *Decision Tree* dapat menghasilkan akurasi sebesar 89,6% dengan menggunakan *dataset* yang terdiri dari 14 atribut dan 1025 baris data [5].

Algoritma *Decision Tree* tidak hanya dapat digunakan untuk penyakit jantung, tetapi juga dapat digunakan untuk penyakit strok. Pada penelitian yang

berjudul "Classification of stroke patients using data mining with AdaBoost, Decision Tree, and Random Forest models", ditemukan bahwa algoritma *Decision Tree* menghasilkan akurasi sebesar 95.3%, presisi sebesar 90.7%, dan *recall* sebesar 95.3% [6]. Penelitian tersebut menggunakan *dataset* yang berisi 5100 sampel dan 11 atribut serta dibagi menjadi 2 bagian, yaitu data *training* sebanyak 80% dan data *testing* sebanyak 20%.

Pada penelitian serupa dengan judul "Klasifikasi Penyakit Stroke dengan Metode *Support Vector Machine (SVM)*" yang menggunakan algoritma *Support Vector Machine* untuk mengklasifikasikan penyakit stroke dengan *kernel linear*, *RBF*, *polynomial*, dan *sigmoid*, didapatkan beberapa hasil, yaitu Untuk *kernel linear*, didapatkan nilai akurasi 75.73%, nilai presisi 74.07%, serta nilai *recall* sebesar 56.16%. Kemudian, untuk *kernel RBF* didapatkan nilai akurasi 73.38%, nilai presisi 71.97%, dan nilai *recall* sebesar 55.85%. Untuk *kernel polynomial*, didapatkan nilai akurasi sebesar 78.86%, nilai presisi 73.98%, serta nilai *recall* 56.75%. Terakhir, untuk *kernel sigmoid* didapatkan nilai akurasi sebesar 46.14%, nilai presisi 38.84% dan nilai *recall* sebesar 48.13% [7].

Kemudian, Pada penelitian lain yang berjudul "Analysis of Stroke Classification using Random Forest Method" menggunakan algoritma *Random Forest* untuk melakukan klasifikasi penyakit stroke dengan atribut umur, kadar gula rata-rata, dan BMI. Hasil nilai akurasi dari penelitian ini sebesar 86.82%, nilai presisi 15.76%, nilai *recall* 38.15%, dan *F1-Score* sebesar 22.30% dengan menggunakan 100 *trees* [8].

Selain itu, terdapat penelitian serupa tentang penyakit stroke dengan menggunakan algoritma *Logistic Regression*. Penelitian dengan judul "Implementasi Algoritma *Logistic Regression* Untuk Klasifikasi Penyakit Stroke" ini menggunakan 3984 data *training* dan 996 data *testing* menghasilkan nilai akurasi sebesar 94% [9].

Berdasarkan latar belakang yang telah dijabarkan, penelitian ini berfokus pada implementasi algoritma *Decision Tree* dalam mendeteksi penyakit stroke pada pria dengan menggunakan *Brain Stroke Dataset* yang diperoleh dari *Kaggle*.

1.2 Rumusan Masalah

Berdasarkan pada latar belakang yang telah dipaparkan, maka tujuan penelitian adalah sebagai berikut:

1. Bagaimana mengimplementasikan algoritma *Decision Tree* dalam

mendeteksi penyakit stroke pada pria?

2. Berapa tingkat akurasi, *precision*, *recall*, dan *f1-score* algoritma *Decision Tree* dalam mendeteksi penyakit stroke pada pria?

1.3 Batasan Permasalahan

1. *Dataset* yang digunakan adalah *Brain Stroke Dataset* yang berasal dari *Kaggle*.
2. Penelitian berfokus pada *gender* pria, sehingga *class* yang digunakan dari atribut *gender* adalah *male*.

1.4 Tujuan Penelitian

Berdasarkan pada latar belakang yang telah dipaparkan, maka tujuan penelitian adalah sebagai berikut:

1. Mengimplementasikan algoritma *Decision Tree* dalam mendeteksi penyakit stroke pada pria.
2. Mengukur tingkat akurasi, *precision*, *recall*, dan *f1-score* algoritma *Decision Tree* dalam mendeteksi penyakit stroke pada pria.

1.5 Manfaat Penelitian

Berdasarkan pada latar belakang yang telah dipaparkan, manfaat dari penelitian ini adalah model yang dihasilkan dapat dijadikan dasar dan dikembangkan menjadi aplikasi untuk membantu tenaga medis dalam mendeteksi penyakit stroke pada pria.

1.6 Sistematika Penulisan

Sistematika yang digunakan dalam penulisan laporan tugas akhir ini adalah sebagai berikut:

- Bab 1 PENDAHULUAN

Bab ini membahas latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan.

- Bab 2 LANDASAN TEORI

Bab ini menjelaskan teori-teori atau literatur yang digunakan sebagai dasar penelitian, termasuk pembahasan mengenai strok, *decision tree*, *Grid Search*, *SMOTEENN*, dan *confusion matrix*.

- Bab 3 METODOLOGI PENELITIAN

Bab ini menjelaskan tahapan penelitian meliputi studi literatur, pengumpulan data, perancangan serta pengujian dan evaluasi.

- Bab 4 HASIL DAN DISKUSI

Bab ini menyajikan spesifikasi sistem, implementasi sistem, hasil uji coba penelitian yang diperoleh serta pembahasannya yang bertujuan untuk menjawab rumusan masalah yang telah ditetapkan.

- Bab 5 KESIMPULAN DAN SARAN

Bab ini berisi simpulan dari penelitian yang telah dilakukan serta saran untuk pengembangan penelitian di masa mendatang.

