

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stroke merupakan salah satu penyebab utama kematian dan kecacatan di seluruh dunia. Berdasarkan data dari Organisasi Kesehatan Dunia (WHO), stroke menyumbang sekitar 11% dari total kematian global. Penyakit stroke adalah penyebab kecacatan nomor satu dan penyebab kematian nomor tiga di dunia setelah penyakit jantung dan kanker. Adapun beberapa faktor penyebab penyakit stroke adalah: tekanan darah, riwayat fibrilasi atrium, kolesterol, diabetes, dan lain sebagainya. [1] Penyakit stroke terjadi karena terputusnya suplai darah menuju otak, dapat terjadi karena adanya sumbatan berupa darah yang menggumpal atau terdapat semburan pada pembuluh darah.

Di Indonesia, prevalensi stroke juga menunjukkan angka yang signifikan, dengan dampak yang besar terhadap kualitas hidup dan ekonomi masyarakat. Pengenalan dini dan intervensi yang tepat sangat penting dalam mengurangi risiko dan dampak stroke[2]. Menurut data Riskesdes pada tahun 2018 dinyatakan bahwa prevalensi stroke (permil) berdasarkan diagnosis dokter pada penduduk umur ≥ 15 tahun provinsi dengan pasien stroke tertinggi terjadi di Provinsi Kalimantan Timur sebesar 14,7% dan terendah ada di Provinsi Papua sebesar 4,1%. Prevalensi pasien stroke berdasarkan diagnosis dokter meningkat seiring dengan bertambahnya usia, tertinggi ada pada usia ≥ 75 tahun yaitu sebesar 50,2% [3] Kematian akibat stroke sebesar 51% di seluruh dunia disebabkan oleh tekanan

darah tinggi. Selain itu, diperkirakan sebesar 16% kematian stroke disebabkan tingginya kadar glukosa darah dalam tubuh. Tingginya kadar gula darah dalam tubuh secara patologis berperan dalam peningkatan konsentrasi glikoprotein, yang merupakan pencetus beberapa penyakit vaskuler. Kadar glukosa darah yang tinggi pada saat stroke akan memperbesar kemungkinan meluasnya area infark karena terbentuknya asam laktat akibat metabolisme glukosa secara anaerobik yang merusak jaringan otak.

Untuk mendukung deteksi dini dan manajemen penyakit stroke, penggunaan teknik prediksi berbasis data memiliki peran yang sangat penting. Dengan memanfaatkan data yang tersedia, kita dapat mengidentifikasi individu yang berisiko tinggi mengalami stroke sebelum gejala klinis muncul. Ini memungkinkan intervensi yang lebih awal dan strategi pencegahan yang lebih efektif. Dalam konteks ini, algoritma pembelajaran mesin, yang merupakan metode canggih dalam analisis data, menawarkan potensi besar. Algoritma ini dapat mengolah dan mengevaluasi sejumlah besar data medis untuk mengungkap pola-pola yang mungkin tidak terlihat dengan metode tradisional. Dengan menggunakan algoritma pembelajaran mesin, kita dapat mengembangkan model yang mampu memprediksi risiko stroke dengan tingkat akurasi yang tinggi, berdasarkan variabel-variabel seperti riwayat kesehatan, faktor risiko, dan hasil pemeriksaan medis[4].

Di antara berbagai algoritma pembelajaran mesin yang tersedia, dua teknik yang sering diterapkan dalam prediksi medis adalah K-Nearest Neighbors (KNN) dan teknik Bagging. K-Nearest Neighbors (KNN) adalah salah satu algoritma

pembelajaran mesin yang sederhana namun efektif. KNN bekerja dengan cara mengklasifikasikan data baru berdasarkan kedekatannya dengan data yang telah ada, yang dinilai dengan metrik jarak tertentu. Meskipun KNN dikenal karena kemudahannya dalam implementasi, ia memiliki keterbatasan dalam hal sensitivitas terhadap data yang tidak seimbang dan variabilitas data yang tinggi[5]. Bagging (Bootstrap Aggregating) adalah teknik ensemble yang dirancang untuk meningkatkan akurasi model dengan mengurangi variabilitas dan overfitting. Bagging bekerja dengan melatih beberapa model pada subset acak dari data dan menggabungkan hasilnya untuk membuat prediksi yang lebih stabil dan akurat. Dalam konteks KNN, teknik Bagging dapat digunakan untuk meningkatkan performa prediksi dengan mengurangi sensitivitas terhadap fluktuasi data dan meningkatkan generalisasi model[6].

Menurut Penelitian Nur Diana Saputri tentang “Comparison of Bagging and Adaboost Methods on C4.5 Algorithm for Stroke Prediction” Metode Bagging dan Adaboost sangat berpengaruh dalam meningkatkan nilai akurasi, presisi, dan spesifitas pada algoritma C4.5 hasil akurasi dari algoritma C4.5 dengan metode bagging yaitu 95.02%. Penelitian Zuriati Z1 "Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN)" Penerapan Algoritma KNN dan Evaluasi kinerja KNN dengan confusion matrix dan penghitungan akurasi. Performa algoritma KNN terbaik didapatkan dengan nilai $k=5$ dan akurasi 93.54%.

Berdasarkan hasil penelitian yang dilakukan peneliti sebelumnya maka masih perlu dikembangkan lagi agar klasifikasi prediksi penyakit stroke mendapatkan tingkat akurasi yang lebih tinggi.

1.2 Batasan Masalah

Agar pembahasan penelitian ini tidak menyimpang pada rumusan masalah, maka penelitian ini akan dibatasi oleh ruang lingkup sebagai berikut :

- a. Prediksi penyakit stroke dengan menggunakan Algoritma *K-Nearest Neighbor* dan *Optimasi Teknik Bagging*
- b. Menggunakan dataset publik yang diambil dari <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>
- c. *Tools* yang digunakan dalam penelitian ini adalah *Rapid Miner*.

1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, bagaimana cara menghasilkan model algoritma yang memiliki akurasi yang sangat tinggi menggunakan model Algoritma *K-Nearest Neighbor* dan *Optimasi Teknik Bagging* pada data prediksi stroke yang di dapat dari situs kaggle.com.

1.4 Tujuan Penelitian

Berdasarkan permasalahan yang ada tujuan dari penelitian ini adalah untuk

1. Mengukur nilai akurasi menggunakan metode Algoritma *K-Nearest Neighbor* dan *Optimasi Teknik Bagging* untuk prediksi penyakit stroke.
2. Mengidentifikasi risiko stroke dengan lebih cepat dan akurat, sehingga dapat membantu intervensi medis lebih awal.

1.5 Manfaat Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah :

1. Mempermudah dalam melakukan prediksi penyakit stroke menggunakan Algoritma *K-Nearest Neighbor* dan *teknik Bagging*
2. Mendapatkan nilai akurasi yang tinggi dengan menggunakan data mining dan menguji beberapa variabel yang diperoleh dari dataset publik
3. Meningkatkan kesembuhan penyakit stroke dengan lebih cepat
4. Memberikan masukan dalam ilmu pengetahuan untuk penelitian selanjutnya dan referensi ilmiah dalam penelitian penerapan data mining.

1.6 Sistematika Penulisan

Sistematika Penulisan Tesis ini menggunakan sistematika pembahasan yang akan diuraikan menjadi susunan bab yang akan di bahas,yaitu sebagai berikut :

a. BAB I PENDAHULUAN

Pada bab ini berisi tentang latar belakang, ruang lingkup, rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan

b. BAB II TINJAUAN PUSTAKA

Pada bab ini berisi menguraikan tentang tinjauan pustaka yang digunakan dalam melakukan evaluasi dan penulisan tesis.

c. BAB III METODE PENELITIAN

Pada bab ini berisi penjelasan tentang metodologi yang digunakan penulis dalam melakukan penelitian.

d. **BAB IV ANALISIS DAN PEMBAHASAN**

Pada bab ini berisi tentang hasil pengujian dan pembahasan mengenai hasil dari algoritma yang digunakan.

e. **BAB V KESIMPULAN DAN SARAN**

Pada bab ini berisi tentang kesimpulan dan saran dari hasil bab – bab yang telah diuraikan.