

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit *stroke* termasuk ke dalam salah satu penyakit tidak menular (PTM) yang kejadiannya menjadi masalah kesehatan yang cukup serius, baik di negara maju ataupun negara berkembang (Khariri & Saraswati, 2021). Penyakit *stroke* menyebabkan kematian kedua dan kecacatan ketiga di dunia, dimana 70% penderita penyakit *stroke* terjadi pada negara berpenghasilan rendah dan menengah. Sementara itu, kematian dan kecacatan yang disebabkan oleh penyakit *stroke* menyumbang 87% (Singh, 2021).

Penyakit *stroke* dapat terjadi begitu saja atau dapat dialami secara mendadak yang dikarenakan adanya gangguan dalam mengantarkan darah ke otak selama 24 jam atau bisa lebih (Khariri & Saraswati, 2021). Menurut Roebers et al (2007) dalam Gofir (2021) menyatakan bahwa penyakit *stroke* dan TIA (*Transient Ischemic Attack*) termasuk ke dalam kasus emergensi. Jika tidak diberikan penanganan sesegera mungkin, maka dapat menimbulkan dampak yang lebih parah. Namun gejala dini penyakit *stroke* sulit untuk diketahui (Gofir, 2021). Penderita juga tidak menyadari atau merasakan adanya gejala dini penyakit *stroke* (Ridwan, 2017). Sehingga ketika penderita telah mengalami gejala yang cukup serius, barulah si penderita dirujuk ke rumah sakit untuk evaluasi dan diberikan penanganan yang lebih baik. Padahal jika gejala dini dari penderita sudah diketahui, dampak buruk yang ditimbulkan oleh penyakit *stroke* dapat diminimalisir.

Data mining dapat dimanfaatkan untuk mendiagnosis penyakit. Dalam penerapannya dengan menggunakan berbagai metode atau algoritma dan membutuhkan data atau informasi yang banyak sehingga berguna untuk mendapatkan informasi yang tersembunyi dan menghasilkan keputusan atas diagnosis medis tersebut (Harahap et al., 2021). Dari metode klasifikasi untuk prediksi, *Random Forest* menjadi metode yang menghasilkan nilai akurasi yang baik (Aprilia et al., 2021).

Penelitian terkait melakukan klasifikasi untuk memprediksi penyakit jantung dengan metode *Random Forest*, yang menghasilkan nilai akurasi sebesar 85.3% (Harahap et al., 2021). Sedangkan penelitian kedua melakukan klasifikasi untuk memprediksi

pasien menderita penyakit ginjal kronik dengan menggunakan metode *Naïve Bayes*, *Naïve Bayes* dengan *Adaboost*, dan *Random Forest*. Dari penelitian tersebut menghasilkan nilai akurasi 95.4% untuk *Naïve Bayes*, *Naïve Bayes* dengan *Adaboost* memperoleh nilai akurasi sebesar 98.6%, dan *Random Forest* menghasilkan nilai akurasi sebesar 99.3% (Tyasnurita & Hapsari, 2020).

Oleh karena itu, penelitian ini akan menggunakan metode *Random Forest* untuk klasifikasi diagnosis penyakit *stroke*. Tujuan penelitian adalah mengetahui performa yang dihasilkan oleh metode *Random Forest* melalui tahapan evaluasi. Luaran akhir yang akan diperoleh berupa model terbaik untuk mendiagnosis penyakit *stroke* dengan menggunakan *Random Forest*.

1.2 Rumusan Masalah

Bagaimana performa yang dihasilkan oleh metode *Random Forest* dalam mendiagnosis penyakit *stroke*?

1.3 Ruang Lingkup

Di bawah ini adalah ruang lingkup untuk penelitian ini, diantaranya yaitu:

1. Dataset yang digunakan adalah dataset penyakit *stroke* yang bersumber dari *website* Kaggle <https://www.kaggle.com/fedesoriano/stroke-prediction-dataset> yang terdiri dari 11 atribut dan 1 *class* serta berjumlah 5110 *records*.
2. Metode yang digunakan adalah metode klasifikasi *Random Forest*.
3. Banyak pohon yang digunakan adalah 10, 20, 30, 40, 50, 60, 70, 80, 90, dan 100 pohon.
4. Evaluasi yang digunakan pada penelitian ini diantaranya adalah nilai *accuracy*, *sensitivity*, *specificity*, *precision*, *F-measure*, dan ROC AUC.

1.4 Tujuan Penelitian

Mengetahui performa yang dihasilkan oleh metode *Random Forest* melalui tahapan evaluasi untuk diagnosis penyakit *stroke*.

1.5 Manfaat Penelitian

Penelitian yang dilakukan akan memberikan manfaat sebagai berikut:

1. Memberikan informasi atau pengetahuan mengenai metode *Random Forest*.
2. Sebagai bahan referensi dalam penelitian yang akan dilakukan selanjutnya.

1.6 Luaran Yang Diharapkan

Luaran yang diperoleh berupa model terbaik dalam mendiagnosis penyakit *stroke* dengan menggunakan *Random Forest*.

1.7 Sistematika Penulisan

Dibawah ini adalah penjelasan dari urutan sistematika penulisan:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, ruang lingkup, tujuan penelitian, manfaat penelitian, luaran yang diharapkan, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisi mengenai teori-teori yang digunakan sebagai acuan dalam melakukan penyusunan penelitian.

BAB III METODOLOGI PENELITIAN

Bab ini berisi mengenai penjelasan atas tahapan-tahapan pelaksanaan penelitian yang akan dilakukan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi mengenai penjelasan yang lebih rinci sesuai dengan tahapan-tahapan pelaksanaan penelitian yang dilakukan dan melakukan proses analisis terhadap hasil penelitian.

BAB V PENUTUP

Bab ini berisi kesimpulan atas hasil penelitian yang sudah diperoleh dan saran untuk dapat mengembangkan atau mengkaji ulang penelitian ini agar menghasilkan hasil yang lebih baik.

DAFTAR PUSTAKA

RIWAYAT HIDUP

LAMPIRAN