

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penyakit Alzheimer merupakan penyebab demensia paling umum sekaligus penyakit neurodegeneratif yang merusak koneksi antarsel saraf di otak, sehingga menimbulkan gangguan perilaku dan fungsi kognitif (Alzheimer's Association, 2019). Saat ini Alzheimer telah menjadi isu kesehatan global yang semakin mengkhawatirkan, lebih dari 55 juta penderita di seluruh dunia pada tahun 2024 dan diproyeksikan akan melonjak menjadi 139 juta orang pada tahun 2050 (Alzheimer's Disease International, 2024). Di Indonesia saja, jumlah kasus diperkirakan mencapai ± 518.000 pada tahun 2024 dan akan meningkat seiring penuaan populasi serta gaya hidup modern (kurang aktivitas fisik, pola makan tidak sehat).

Sampai saat ini, belum ditemukan pengobatan yang benar-benar dapat menyembuhkan penyakit Alzheimer. Namun, deteksi dan penanganan sejak dini telah terbukti efektif dalam memperlambat perkembangan penyakit serta meningkatkan kualitas hidup penderita. Sayangnya, metode diagnosis konvensional seperti wawancara klinis mendalam, tes neuropsikologi, dan pencitraan otak masih tergolong mahal, memerlukan waktu lama, dan sulit dijangkau oleh semua kalangan. Oleh karena itu, diperlukan alternatif metode diagnosis yang lebih cepat, akurat, dan mudah diakses.

Machine learning sebagai bagian dari kecerdasan buatan, memungkinkan komputer untuk belajar dan mengambil keputusan berdasarkan data. Teknologi ini telah banyak diterapkan di bidang kesehatan, mulai dari prediksi penyakit, segmentasi kelompok pasien, hingga analisis citra medis. Di antara algoritma *supervised learning*, *Random Forest* menjadi salah satu yang paling banyak digunakan karena akurasi yang tinggi dan kestabilannya terhadap data tidak seimbang dan *outlier* (Breiman, 2001). Namun, masalah ketidakseimbangan kelas dalam data medis terutama ketika jumlah kasus positif jauh lebih sedikit dibanding negatif, sering menyebabkan penurunan akurasi model. Untuk mengatasi hal tersebut, metode *Synthetic Minority Over-sampling Technique* (SMOTE) dapat

digunakan, yaitu dengan membuat data sintetis baru pada kelas minoritas guna menyeimbangkan distribusi data dan meningkatkan kinerja klasifikasi.

Penelitian klasifikasi penyakit alzheimer sebelumnya telah dilakukan oleh (Dewi et al., 2023) dengan judul Algoritma *K-Nearest Neighbor* (K-NN) dan *Single Layer Perceptron* (SLP) Untuk Klasifikasi Penyakit Alzheimer. Pada pengujian dengan variasi parameter K dari 1 sampai 9, hasil terbaik dari penelitian klasifikasi penyakit alzheimer menggunakan algoritma *K-Nearest Neighbor* (K-NN) dan *Single Layer Perceptron* (SLP) ditemukan pada K=5, yang menghasilkan nilai akurasi, berkisar antara 91%.

Adapun penelitian yang menggunakan algoritma *Random Forest* sebelumnya telah dilakukan oleh (Rozy, A. 2024) dengan judul Penerapan Random Forest Untuk Prediksi Virus Hepatitis C. Berdasarkan penelitian yang telah dilakukan menggunakan pemeriksaan darah model *Random Forest* untuk prediksi virus hepatitis C, menghasilkan akurasi yang sangat baik yaitu sebesar 91,86%, *precision* = 85%, *recall* = 95%, *f1 score* = 95% dan *support* = 93%.

Berdasarkan studi literatur yang sudah dilakukan tidak banyak penelitian yang secara khusus menerapkan *Random Forest* untuk klasifikasi Alzheimer dan menangani ketidakseimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE), serta mengevaluasi dampak dari variasi *hyperparameter* kunci *Random Forest* yaitu *n_estimators* (jumlah pohon) dan *max_depth* (kedalaman pohon) terhadap metrik evaluasi (*akurasi*, *precision*, *recall*, dan *F1 score*). Oleh karena itu, penelitian ini mengangkat topik dengan judul “Implementasi Algoritma *Random Forest* Berbasis *Oversampling* SMOTE Untuk Klasifikasi Penyakit Alzheimer” karena ada perbedaan dalam penelitian sebelumnya. Diharapkan bahwa penelitian ini akan memberikan kontribusi terhadap kemajuan ilmu pengetahuan di bidang *machine learning* dalam bidang kesehatan, serta menawarkan solusi praktis dan berguna untuk membantu proses deteksi dini penyakit Alzheimer.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Berapa nilai metrik evaluasi akurasi, *precision*, *recall*, dan *f1 score* algoritma *Random Forest* dalam melakukan klasifikasi penyakit alzheimer menggunakan metode SMOTE (*Synthetic Minority Oversampling Technique*)?
2. Kombinasi nilai *n_estimators* dan *max_depth* manakah yang menghasilkan performa terbaik pada metrik evaluasi akurasi, *precision*, *recall*, dan *F1 score* setelah penerapan SMOTE (*Synthetic Minority Oversampling Technique*)?

1.3 Tujuan

Berdasarkan rumusan masalah diatas maka tujuan dari penelitian tugas akhir ini adalah:

1. Menguji kinerja algoritma *Random Forest* dalam melakukan klasifikasi terhadap data penyakit Alzheimer menggunakan metrik evaluasi yang sesuai seperti nilai akurasi, *precision*, *recall*, dan *f1 score*, sehingga algoritma *Random Forest* ini dapat dikatakan efektif atau tidak, dalam melakukan klasifikasi penyakit Alzheimer berdasarkan *dataset* dari Kaggle.
2. Mengimplementasikan algoritma *Random Forest* untuk melakukan klasifikasi penyakit Alzheimer pada *dataset* yang diperoleh dari Kaggle berdasarkan kombinasi nilai *n_estimators* dan *max_depth* terbaik.

1.4 Batasan Masalah

Agar penelitian ini lebih terfokus dan terarah, maka terdapat beberapa batasan masalah yang ditetapkan, yaitu:

1. *Dataset* yang digunakan merupakan *dataset* terbuka mengenai penyakit Alzheimer yang diunduh dari Kaggle.
2. Algoritma yang digunakan dalam penelitian ini adalah *Random Forest*, tanpa dibandingkan secara langsung dengan algoritma lain yang mana parameter yang digunakan *n_estimator* dan *max_depth*.
3. Menggunakan metode SMOTE (*Synthetic Minority Oversampling Technique*) untuk mengatasi ketidakseimbangan data dengan evaluasi performa model dilakukan dengan menggunakan metrik: akurasi, *precision*, *recall*, dan *f1 score*.

4. Jumlah data yang digunakan sebanyak 2149 data pasien, dataset terdiri dari 35 kolom atau atribut dan memiliki 2 kelas atau target yaitu terindikasi Alzheimer dan tidak terindikasi Alzheimer.

1.5 Manfaat

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menambah wawasan dan referensi ilmiah mengenai penerapan algoritma *machine learning*, khususnya algoritma *Random Forest*, dalam bidang kesehatan.
2. Memberikan gambaran bagaimana *machine learning* dapat digunakan untuk membantu proses deteksi dini penyakit Alzheimer secara lebih sistematis dan berbasis data.
3. Menyediakan aplikasi berbasis *website* yang dapat digunakan sebagai alat bantu dalam proses deteksi dini penyakit Alzheimer dengan model yang telah dievaluasi secara optimal.

