

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Penyakit asma adalah kondisi kronis yang memengaruhi saluran pernapasan, menyebabkan peradangan dan penyempitan saluran pernapasan, yang menimbulkan gejala seperti sesak napas, batuk, dan mengi. Namun, asma dapat memengaruhi orang-orang dari segala usia. Asma memengaruhi lebih dari 262 juta orang di seluruh dunia, menyebabkan sekitar 460.000 kematian setiap tahunnya. (*World Health Organization, 2024*).

Asma masih merupakan masalah kesehatan global dan cukup umum di Indonesia. Berdasarkan data 2,4% dari populasi Kemenkes (2023), Faktor-faktor seperti polusi udara, kebiasaan merokok, kurangnya olahraga, dan paparan alergen semuanya diperkirakan berkontribusi pada peningkatan prevalensi ini. Kondisi ini menunjukkan bahwa asma merupakan penyakit yang membutuhkan perhatian serius karena dapat menurunkan kualitas hidup dan produktivitas penderitanya jika tidak ditangani dan diobati segera.

Asma bersifat genetik dan tidak dapat dihilangkan sepenuhnya, jadi hingga saat ini belum ada obat yang sepenuhnya menyembuhkan. Namun, gejalanya dapat diatasi dengan baik dengan pengobatan dan menjalani gaya hidup yang sehat. Penanganan dini dan diagnosis telah terbukti sangat membantu mencegah kambuhan dan meningkatkan kualitas hidup pasien. Metode diagnosis konvensional, seperti konsultasi klinis, pemeriksaan tubuh, dan uji spirometri, masih memiliki keterbatasan, seperti biaya yang tinggi, waktu yang lama, dan kesulitan mencapai lokasi tertentu. Akibatnya, pilihan diagnosis alternatif yang tepat sasaran, hemat biaya, cepat dan memiliki akurasi tinggi (*Halimatusyakdiah dkk., 2025*).

Banyak peluang baru telah muncul dalam bidang kesehatan berkat kemajuan teknologi pembelajaran mesin (*machine learning*), terutama untuk klasifikasi penyakit, prediksi risiko, serta analisis data medis. *Machine learning* memungkinkan komputer tanpa pemrograman eksplisit untuk belajar dari data dan membuat keputusan, yang memungkinkan mereka untuk menemukan pola yang

tidak mudah dikenali manusia. Metode pembelajaran *supervised* berbasis kelompok (*ensemble*), yang menggunakan algoritma *random forest* untuk metode klasifikasi, menggunakan banyak pohon keputusan, juga dikenal sebagai (*decision trees*) untuk meningkatkan stabilitas dan akurasi hasil klasifikasi. *Random forest* dikenal mampu memberikan hasil prediksi yang stabil meskipun data tidak sepenuhnya seimbang, dan sangat tahan terhadap data *outlier* dan *noise* (Rahmah dkk., 2023).

Pada kasus data medis seperti penyakit asma sering dijumpai ketidakseimbangan kelas (*class imbalance*), dimana jumlah data pasien yang tidak sama. Ketidakseimbangan ini dapat menyebabkan model bias terhadap kelas mayoritas, yang mengurangi kemampuan model untuk mengidentifikasi kasus asma secara akurat. Untuk mengatasi hal tersebut, digunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE), dengan teknik *oversampling* yang menghasilkan data sintesis baru pada kelas minoritas guna menyeimbangkan distribusi data, sehingga dapat meningkatkan performa model klasifikasi (Hairani dkk., 2024).

Penelitian sebelumnya oleh Al Muqarrabin & Mirzal Safari (2025), telah berhasil mengklasifikasikan asma menggunakan algoritma *KNN*. Namun, belum banyak penelitian yang secara khusus mempelajari penggunaan algoritma *random forest*, yang sangat dikenal kuat dalam menangani data medis yang kompleks, dan mengintegrasikannya dengan metode SMOTE untuk mengatasi kemungkinan masalah ketidakseimbangan kelas pada data keparahan asma. Oleh karena itu, penelitian ini bertujuan untuk mengisi celah tersebut dengan menggunakan algoritma *random forest* untuk mengklasifikasikan asma. Sementara itu, Sari et al. (2024) dalam *PREDATECS Journal*, menunjukkan bahwa algoritma *random forest* memperoleh akurasi 98,58% dalam klasifikasi pasien stroke. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma *random forest* memiliki kemampuan yang baik untuk mengklasifikasikan berbagai kasus penyakit. Namun, kualitas data dan parameter yang digunakan sangat memengaruhi tingkat akurasi algoritma.

Hasil penelusuran literatur menunjukkan bahwa hanya beberapa penelitian yang secara khusus menggunakan algoritma *random forest* untuk klasifikasi penyakit asma dengan menggunakan pendekatan penyeimbangan data SMOTE. Penelitian ini juga menganalisis dampak dari variasi *hyperparameter* utama seperti

n_estimators (jumlah pohon) dan *max_depth* (kedalaman pohon) terhadap metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score* sebagaimana umum dilakukan dalam evaluasi model *random forest* (Ketut dkk., 2023). Oleh karena itu, fokus penelitian ini adalah "Klasifikasi Tingkat Keparahan Penyakit Asma Menggunakan Algoritma *Random Forest*". Diharapkan hasil penelitian ini dapat memberikan kontribusi manfaat yang signifikan untuk kemajuan ilmu di bidang *machine learning* dan kesehatan, terutama dalam upaya untuk menemukan diagnosa penyakit asma dengan cara yang lebih cepat, akurat, dan efektif. Hasil penelitian ini juga akan berfungsi sebagai referensi untuk pengembangan sistem deteksi dini berbasis data medis di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan topik latar belakang di atas, masalah penelitian ini dapat diuraikan sebagai berikut:

1. Berapa nilai metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score* pada algoritma *random forest* setelah penerapan metode SMOTE dalam klasifikasi penyakit asma?
2. Kombinasi *hyperparameter n_estimators* dan *max_depth* mana yang memiliki hasil terbaik berdasarkan evaluasi *accuracy*, *precision*, *recall*, dan *F1-score* dengan penerapan metode SMOTE?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan utama dari penelitian tugas akhir ini adalah sebagai berikut:

1. Menguji kinerja algoritma *random forest* dalam mengklasifikasikan data terkait penyakit asma, berdasarkan evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*, sehingga dapat diketahui sejauh mana efektivitas algoritma *random forest* dalam mengklasifikasikan penyakit asma berdasarkan *dataset* dari Kaggle.
2. Mengimplementasikan algoritma *random forest* pada klasifikasi penyakit asma pada *dataset* yang diperoleh dari Kaggle berdasarkan kombinasi nilai *n_estimators* dan *max_depth* terbaik.

1.4 Batasan Penelitian

Agar terdapat batasan masalah yang ditetapkan untuk membuat penelitian ini lebih fokus, sebagai berikut:

1. *Dataset* yang digunakan merupakan *dataset* terbuka mengenai penyakit asma dari *platform kaggle.com*, yaitu *dataset* milik Deepayan Thakur dengan judul *Asthma Disease Prediction*. Seluruh data dalam *dataset* digunakan sebagai dasar analisis untuk membangun model klasifikasi tingkat keparahan penyakit asma. *Dataset* diakses pada 30 September 2025 melalui tautan <https://www.kaggle.com/datasets/deepayanthakur/asthma-disease-prediction>
2. Penelitian ini menggunakan algoritma *random forest*, tanpa dilakukan perbandingan secara langsung dengan algoritma lain. Parameter utama yang digunakan dan diuji penelitian ini merupakan *n_estimators* dan *max_depth*.
3. Untuk mengatasi masalah ketidakseimbangan kelas (*class imbalance*) pada *dataset* penyakit asma digunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE).
4. Metrik evaluasi yang digunakan untuk mengevaluasi kinerja model, yaitu; *accuracy*, *precision*, *recall*, dan *F1-score*.

1.5 Manfaat Penelitian

Hasil yang diberikan untuk penelitian ini diharapkan dapat memberikan manfaat dalam mengetahui efektivitas algoritma *random forest* serta optimasi *hyperparameter* dengan penerapan metode SMOTE terhadap peningkatan performa klasifikasi penyakit asma.