

BAB I

PENDAHULUAN

1.1. Latar Belakang

Tuberkulosis (TB) adalah penyakit infeksi yang disebabkan oleh bakteri *Mycobacterium tuberculosis*, yang dapat menyerang paru-paru dan organ tubuh lainnya. Penyakit ini menular melalui inhalasi droplet dari penderita, dan infeksi biasanya terjadi pada individu dengan daya tahan tubuh yang lemah. Meskipun TB lebih umum menyerang paru-paru, infeksi ini dapat menyebar ke berbagai organ seperti saluran pencernaan, tulang, otak, dan ginjal. Penularan TB mirip dengan flu, tetapi tidak semudah itu; biasanya terjadi di antara anggota keluarga yang tinggal serumah atau saat berada di ruang publik seperti transportasi umum. Bakteri ini memiliki karakteristik khusus, seperti bentuk batang dengan dinding lemak tebal, yang membuatnya tahan terhadap asam dan alkohol. Tuberkulosis (TB) dapat diklasifikasikan berdasarkan lokasi anatomis menjadi TB paru dan TB ekstra paru. TB paru mencakup infeksi di parenkim paru atau trakeobronkial, termasuk TB milier, sementara TB ekstra paru menyerang organ lain seperti pleura, kelenjar getah bening, tulang, atau selaput otak [1].

Tuberkulosis (TB) adalah masalah kesehatan global serius, dengan Asia Tenggara mencatat 44% kasus baru pada 2019, diikuti Afrika (25%) dan Pasifik Barat (18%). Delapan negara, termasuk Indonesia, menyumbang dua per tiga dari total kasus baru. Tahun tersebut, TB menyebabkan 1,4 juta kematian, termasuk 208.000 pada pasien HIV, dan menjadi penyebab utama kematian akibat infeksi. Sebanyak 10 juta orang terinfeksi TB, meliputi 5,6 juta pria, 3,2 juta wanita, dan 1,2 juta anak. Di Indonesia, setiap tahun ada 250.000 kasus baru TB dan 100.000 kematian. TB adalah penyebab utama kematian akibat infeksi dan peringkat ketiga penyebab kematian semua usia, dengan mayoritas pasien berusia produktif 15–54 tahun [2].

Pada kasus tuberkulosis sensitif obat (TB SO), bakteri *Mycobacterium tuberculosis* masih merespons secara efektif terhadap Obat Anti Tuberkulosis (OAT) lini pertama, seperti *rifampisin*, *isoniazid*, *pirazinamid*, dan *ethambutol*. Berbeda dengan TB resisten obat (TB RO) yang memerlukan pendekatan terapi lebih kompleks karena adanya resistansi, TB SO memungkinkan penggunaan regimen standar dengan tingkat keberhasilan yang tinggi jika diberikan secara tepat dan konsisten. Terapi

umumnya terdiri dari dua fase: fase awal selama dua bulan menggunakan kombinasi HRZE untuk menekan populasi bakteri aktif, diikuti oleh fase lanjutan selama empat bulan dengan HR untuk memastikan eliminasi total terhadap bakteri dorman dan mencegah kekambuhan. Dalam tatalaksana klinis, regimen kategori 1 seperti skema 2HRZE/4HR menjadi protokol yang paling banyak diterapkan pada pasien baru. Dengan pengobatan yang teratur dan diawasi, proses penyembuhan dapat tercapai dalam waktu enam bulan, sekaligus menekan risiko transmisi dan kejadian resistansi lanjutan di tingkat populasi [3].

Penyesuaian regimen Obat Anti Tuberkulosis (OAT) pada pasien TB sensitif-obat (TB-SO) sangat bergantung pada evaluasi menyeluruh terhadap kondisi klinis individu. Dosis obat biasanya dikalibrasi berdasarkan kategori berat badan, yang umumnya berkaitan erat dengan usia, sedangkan durasi terapi ditentukan oleh tingkat keparahan penyakit dan respons terhadap pengobatan. Adanya komorbiditas, seperti diabetes melitus, dapat memperumit perjalanan klinis TB serta mempengaruhi efektivitas terapi, sehingga memerlukan pendekatan yang lebih hati-hati dalam pemilihan kombinasi obat. Penggunaan Tes Cepat Molekular (TCM) seperti Xpert MTB/RIF berperan penting dalam konfirmasi infeksi *Mycobacterium tuberculosis* dan memastikan tidak adanya resistansi terhadap rifampisin, yang merupakan kriteria utama dalam klasifikasi TB-SO. Selain itu, asesmen status HIV tetap diperlukan karena infeksi koinfeksi dapat mengubah pola terapi, baik dari sisi interaksi farmakologis maupun durasi pengobatan [4].

Meskipun terapi tuberkulosis sensitif obat (TB SO) telah tersedia secara luas, tantangan implementasi tetap signifikan, terutama dalam hal konsistensi antara diagnosis klinis dan penentuan regimen Obat Anti Tuberkulosis (OAT) yang sesuai standar. Pada tahun 2024, tercatat 879.354 kasus TB SO, namun hanya 90% pasien yang berhasil memulai pengobatan, dan tingkat keberhasilan terapi baru mencapai 84%, belum memenuhi target 90%. Ketidaktepatan dalam pemilihan regimen yang tidak disesuaikan dengan kondisi klinis pasien berisiko menimbulkan berbagai konsekuensi, termasuk kegagalan terapi, kekambuhan, dan perkembangan kasus *Multi-Drug Resistant* (MDR) yang jauh lebih sulit ditangani. Selain itu, penggunaan OAT yang tidak tepat dapat memperbesar potensi terjadinya resistensi obat, menimbulkan efek samping berat, dan menurunkan motivasi pasien untuk menyelesaikan pengobatan [5]. Oleh sebab itu, diperlukan optimalisasi sistem

penapisan dan penentuan terapi yang berbasis bukti untuk mencegah terjadinya resistansi dan meningkatkan keberhasilan pengendalian TB secara nasional.

Penerapan *machine learning* dalam pemilihan regimen Obat Anti Tuberkulosis (OAT) memberikan pendekatan yang lebih terstruktur dan berbasis data dalam menangani kasus tuberkulosis sensitif obat (TB SO). Model klasifikasi yang dikembangkan dalam penelitian ini dirancang untuk mengolah data klinis dan karakteristik pasien guna memberikan rekomendasi terapeutik yang sesuai dengan kondisi individu. Sistem ini dapat diintegrasikan ke dalam praktik klinis sebagai *decision support tool* yang mempercepat dan menstandarkan proses pengambilan keputusan, terutama dalam situasi kompleks seperti adanya komorbiditas atau riwayat pengobatan sebelumnya. Dengan pendekatan ini, terapi dapat lebih dipersonalisasi, sejalan dengan paradigma *precision medicine*, sehingga meningkatkan efektivitas pengobatan. Selain mempercepat penyesuaian regimen, sistem ini juga membantu meminimalkan kesalahan yang berpotensi memicu resistensi obat. Upaya ini secara strategis mendukung pencegahan munculnya *multi-drug resistant* (MDR) yang menjadi hambatan utama dalam eradikasi TB secara global.

Maka dari itu, untuk mendukung upaya ketepatan dalam pemilihan obat anti tuberkulosis, model klasifikasi *machine learning* dapat diimplementasikan dengan memanfaatkan data rekam medis, termasuk usia, jenis kelamin, riwayat kontak fisik, klasifikasi fisiologi, riwayat diabetes melitus, riwayat HIV, dan fitur relevan lainnya [6]. Namun, beberapa algoritma klasifikasi memiliki keterbatasan, terutama dalam menangani data yang besar dengan variabel diskrit dan numerik, data yang tidak seimbang, efisiensi waktu dan penyimpanan, akurasi uji, serta menghadapi tantangan seperti *overfitting* [7].

Penelitian yang dilakukan oleh Anastasia et al [8], membahas penerapan algoritma CatBoost dalam menganalisis *biomarker matrix metalloproteinase* (MMP) dan inhibitor jaringan (TIMP) pada pasien dengan status resistensi obat *Mycobacterium tuberculosis* (MTB). Dalam penelitian ini, pengembangan prosedur *post-processing* berbasis pohon keputusan digunakan untuk meningkatkan akurasi dalam proses klasifikasi, dengan temuan yang menunjukkan peningkatan signifikan dalam tingkat *True positive Rate* (TPR), *True Negative Rate* (TNR), dan *Positive Predictive Value* (PPV). Hasil ini mengindikasikan efektivitas pendekatan ini dalam diagnosis cepat dan akurat terhadap resistensi obat, yang dapat diterapkan di lingkungan klinis.

Selaras dengan itu, penelitian yang dilakukan oleh Li-Hsing et al [9], mengintegrasikan lima model klasifikasi, yaitu *support vector machine* (SVM), *decision tree* (DT), *random forest* (RF), *extreme gradient boosting* (XGBoost), dan *light gradient boosting machine* (LightGBM) yang dioptimalkan menggunakan Optuna untuk meningkatkan kinerja dalam prediksi penyakit. Penelitian ini menemukan bahwa model *LightGBM* yang dioptimalkan memberikan performa prediksi terbaik, dengan akurasi lebih dari 90%, menyoroti pentingnya optimasi *hyperparameter* untuk meningkatkan ketepatan prediksi penyakit, terutama pada wilayah dengan akses layanan kesehatan.

Di sisi lain, penelitian yang dilakukan oleh Safar et al [10], memanfaatkan *CatBoost* yang dioptimalkan menggunakan Optuna dalam mengklasifikasikan diabetes. Hasil penelitian ini menunjukkan peningkatan akurasi dari 65% menjadi 72% setelah pengoptimalan, meskipun peneliti menyarankan eksperimen lebih lanjut dengan metode lain seperti XGBoost untuk meningkatkan akurasi dan menghasilkan model yang lebih andal dalam aplikasi prediksi penyakit. Kombinasi penelitian ini menunjukkan betapa pentingnya pemilihan algoritma dan pengoptimalan *hyperparameter* dalam meningkatkan akurasi dan efisiensi dalam prediksi dan diagnosis penyakit, baik itu untuk resistensi obat MTB maupun penyakit lainnya.

Berdasarkan penelitian sebelumnya, teknik *ensemble learning* dengan *tuning hyperparameter* menunjukkan hasil yang cukup baik. Algoritma CatBoost digunakan dalam penelitian ini untuk mengatasi permasalahan klasifikasi, terutama pada data yang memiliki kombinasi fitur kategorikal dan numerik. Salah satu kekuatan utama dari CatBoost adalah kemampuannya dalam mengelola fitur kategorikal secara langsung tanpa memerlukan tahapan *encoding* yang kompleks, sehingga alur pemrosesan data menjadi lebih efisien dan risiko kesalahan dapat diminimalkan. Tidak seperti algoritma *gradient boosting* tradisional yang mengharuskan transformasi eksplisit terhadap data kategorikal, CatBoost mengintegrasikan pendekatan berbasis statistik yang memungkinkan pemrosesan kategori secara *native*. Selain itu, algoritma ini membangun model secara bertahap melalui teknik *boosting* berulang, yang secara sistematis mengurangi kesalahan prediksi dari model sebelumnya. Perbedaan signifikan antara CatBoost dan metode *gradient boosting* konvensional terletak pada strategi penanganan *categorical features*, di mana CatBoost menggunakan urutan acak dan *target statistics* yang dikendalikan untuk menghindari *target leakage* dan

overfitting, dua permasalahan yang sering terjadi dalam implementasi *gradient boosting* biasa [41], [43].

Meskipun CatBoost menunjukkan kinerja yang tinggi, optimasi *hyperparameter* tetap menjadi tantangan karena kompleksitas dalam pengaturannya. Untuk mengatasinya, metode optimasi Optuna diterapkan, dengan menggunakan algoritma *Tree-structured Parzen Estimator* (TPE) yang mampu menemukan kombinasi parameter optimal seperti *learning rate*, kedalaman pohon, dan jumlah iterasi [49], [55]. Pendekatan ini diharapkan dapat memaksimalkan performa model dan menghasilkan hasil yang lebih efisien.

Maka dari itu, dilakukan klasifikasi penentuan jenis obat pada penderita tuberkulosis menggunakan algoritma CatBoost dan optimasi optuna. Data yang digunakan dalam penelitian ini mencakup data rekam medis pasien pengobatan yang direkap di Sistem Informasi Tuberkulosis (SITB) milik Kementerian Kesehatan. Data diperoleh melalui RSUD Dr. Iskak Tulungagung dengan variabel numerik seperti usia dan lama pengobatan, serta variabel kategorik seperti riwayat diabetes melitus, hasil tes cepat molekular, tes HIV, dan paduan Obat Anti Tuberkulosis (OAT). Hasil dari penelitian ini adalah perbandingan performa algoritma CatBoost dalam klasifikasi jenis obat yang harus dikonsumsi penderita tuberkulosis, seperti *isoniazid*, *rifampisin*, dan *pirazinamid*. Performanya diukur menggunakan *Confusion matrix* serta analisis *Receiver Operating Characteristic* (ROC) dan *Area Under Coverage* (AUC). Model yang menunjukkan performa terbaik akan diimplementasikan ke dalam sistem berbasis *website* untuk membantu tenaga medis menentukan jenis OAT yang sesuai secara efisien dan akurat.

1.2. Rumusan Masalah

Berdasarkan latar belakang, masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bagaimana pengembangan model prediktif untuk melakukan klasifikasi obat anti tuberkulosis bagi pasien di RSUD Dr. Iskak Tulungagung menggunakan algoritma *Categorical Boosting* (CatBoost) dengan optimasi *hyperparameter* optuna?
2. Bagaimana model prediktif dapat membantu tenaga medis dalam mengidentifikasi penentuan obat anti tuberkulosis dalam sistem berbasis *website* yang akan dikembangkan?

3. Bagaimana performa model prediktif yang dikembangkan menggunakan algoritma *Categorical Boosting* (CatBoost) dengan optimasi *hyperparameter* optuna?

1.3. Batasan Masalah

Batasan masalah merupakan komponen penting dalam suatu penelitian yang berfungsi untuk mencegah asumsi yang terlalu luas. Penjabaran batasan masalah akan disajikan dalam format poin-poin sebagai berikut:

1. Penelitian ini akan berfokus pada pengembangan model klasifikasi *machine learning* untuk mengklasifikasikan jenis obat anti tuberkulosis pada pasien di RSUD Dr. Iskak Tulungagung.
2. Model klasifikasi *machine learning* dirancang menggunakan *Categorical Boosting* (CatBoost) dengan optimasi *hyperparameter* optuna.
3. Data berasal dari hasil rekam medis pasien tuberkulosis tahun 2020-2024 pada platform Sistem Informasi Tuberkulosis (SITB) Kementerian Kesehatan.
4. Model klasifikasi *machine learning* yang dikembangkan akan diterapkan ke dalam sistem berbasis *website* menggunakan *framework* Streamlit python.
5. Model yang dikembangkan hanya akan diuji cobakan sebagai rekomendasi kepada tenaga medis, dan belum diterapkan langsung kepada pasien.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dijabarkan, tujuan dari penelitian ini terdiri dari beberapa poin yang akan dijelaskan berikut ini:

1. Mengembangkan model prediktif untuk melakukan klasifikasi obat anti tuberkulosis bagi pasien menggunakan algoritma *Categorical Boosting* (CatBoost) dengan optimasi *hyperparameter* optuna.
2. Merancang dan mengembangkan sistem berbasis website untuk membantu tenaga medis dalam mengidentifikasi jenis obat anti tuberkulosis berdasarkan hasil prediksi model.
3. Mengevaluasi performa model algoritma CatBoost dengan optimasi Optuna.

1.5. Manfaat Penelitian

Manfaat dari penelitian ini mencakup dua aspek, yaitu aspek teoritis dan praktis, yang akan dijabarkan sebagai berikut:

1. Manfaat teoritis

- Penelitian ini diharapkan dapat meningkatkan pemahaman teoritis untuk mengetahui jenis-jenis obat anti tuberkulosis bagi pasien.
- Penelitian ini diharapkan dapat memberikan pengetahuan mengenai penerapan *Categorical Boosting* (CatBoost) dengan optimasi *hyperparameter* optuna dalam mengklasifikasikan jenis-jenis obat anti tuberkulosis.

2. Manfaat praktis

a. Bagi pasien

- Penelitian ini dapat membantu melakukan identifikasi jenis obat anti tuberkulosis yang dibutuhkan pasien sehingga pasien dapat menerima penanganan lebih baik dengan perawatan yang lebih cepat dan tepat.
- Penelitian ini dapat membantu mempercepat pengobatan tuberkulosis, sehingga meningkatkan kualitas hidup pasien dengan mengurangi gejala yang tidak nyaman dan meningkatkan hasil perawatan yang secara signifikan dapat mengurangi risiko kematian akibat penyakit tuberkulosis.

b. Bagi fasilitas kesehatan

- Penelitian ini dapat membantu menyediakan sistem klasifikasi berbasis *website* yang dapat memantau dan mengidentifikasi jenis obat anti tuberkulosis untuk pasien. Dengan adanya sistem identifikasi ini, rumah sakit dapat mempercepat pengobatan yang dilakukan pada banyaknya pasien.
- Penelitian ini dapat meningkatkan efisiensi perawatan dan mendukung pengambilan keputusan medis terkait dengan jenis obat anti tuberkulosis bagi pasien.

c. Bagi Dunia Medis di Indonesia

- Penelitian ini dapat berkontribusi pada peningkatan standar perawatan medis, terutama bagi pemberian obat anti tuberkulosis di Indonesia.
- Penelitian ini dapat mendorong penyebaran pengetahuan dan pentingnya penggunaan teknologi terutama untuk mendeteksi jenis obat anti tuberkulosis yang dapat diterapkan di berbagai fasilitas kesehatan di Indonesia.

d. Bagi Penulis

- Memberikan peluang bagi penulis dalam menerapkan pengetahuan yang telah dipelajari dan berkontribusi secara nyata khususnya dalam dunia kesehatan. Melalui penelitian ini, penulis dapat menambah pengalaman praktis yang dapat bermanfaat dalam mengembangkan karir terutama di bidang sains data dan kesehatan.