

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kesehatan adalah salah satu faktor penting untuk menunjang aktivitas sehari-hari. Namun, kepedulian manusia terhadap menjaga kesehatan sangat rendah, yang mengakibatkan keterlambatan dalam mendiagnosis penyakit salah satu tersebut adalah anemia [1]. Menurut data Kementerian Kesehatan Republik Indonesia tahun 2023, prevalensi anemia di Indonesia mencapai 48,9% orang dewasa dan 32% pada anak-anak usia di bawah lima tahun[2]. Anemia adalah kondisi di mana menyebabkan daya tahan tubuh menurun akibat kurangnya sel darah merah dengan Gejala seperti merasa mudah lelah, kepala pusing, sering merasa mengantuk, kulit tampak pucat atau kekuningan, detak jantung tidak teratur, nyeri dada, dingin di tangan dan kaki. [3],[4].

Anemia ditunjukkan oleh penurunan kadar hemoglobin, hematokrit, dan jumlah eritrosit (red cell count) dibawah nilai normal[2]. Anemia dapat diagnosis melalui pemeriksaan hematologi (profil darah) yang berguna untuk status kesehatan dan menjadi acuan nilai awal (baseline)[5]. Maka karena itu, untuk mengetahui profil darah seperti kadar hemoglobin, hematokrit, dan jumlah eritrosit pemeriksaan dilakukan menggunakan alat hematologi analyzer di laboratorium pada sebuah rumah sakit.

Rumah Sakit Pelengkap Medical Center Jombang, salah satu pusat layanan kesehatan di wilayah jombang dan pada tahun 2012 telah berhasil mendapatkan akreditasi 5 Pokja Pelayanan Dasar dari KARS (Komite Akreditasi Rumah Sakit) Jakarta. Dengan visi menjadi Rumah Sakit yang terdepan dan terpercaya dalam pelayanan medis, karena itu rumah sakit ini memiliki fasilitas laboratorium hematologi yang digunakan untuk mendiagnosis berbagai penyakit darah seperti anemia.

Hasil pemeriksaan laboratorium hematologi biasanya digunakan dalam kelengkapan dokumen dalam rekam medis. Pada setiap hari data rekam medis akan

selalu bertambah sesuai dengan pasien yang masuk pada rumah sakit pelengkap medical center jombang, secara tidak langsung data rekam medis akan menjadi tumpukan data. Jadi guna proses prediksi penyakit dapat berjalan lebih efektif dan efisien maka pemeriksaan dapat dilakukan secara digital dan dapat dilakukan lebih muda maka digunakan proses otomatisasi. Proses otomatisasi tersebut dimaksudkan adalah dengan menerapkan *Machine Learning* yang telah sering dimanfaatkan dalam bidang kesehatan [6].

Terdapat beberapa penelitian yang membahas mengenai latar belakang tersebut. Pada penelitian sebelumnya [7] implementasi machine learning metode fuzzy Tsukamoto dilakukan dari 40 data hematologi dengan hasil 34 data sesuai dan 6 yang tidak sesuai sehingga di peroleh presentase keberhasilan 85%. Penelitian berikutnya oleh [8] menunjukan metode theorema bayes sebagai metode yang digunakan untuk mendiagnosa penyakit anemia pada manusia. Berdasarkan hasil pengujian diperoleh presentase keberhasilan 90% dengan data yang digunakan berjumlah 50 pasien sesuai 45 dan 5 pasien tidak sesuai. Penelitian sebelumnya [9] menggunakan metode SVM untuk klasifikasi kanker paru-paru, dengan mengatasi masalah ketidakseimbangan kelas menggunakan teknik oversampling Adaptive Synthetic Sampling (ADASYN). Penelitian ini menunjukkan bahwa kombinasi ADSYN dan SVM mampu menghasilkan model klasifikasi yang efektif, dengan akurasi sebesar 98,56%.

Selanjutnya Penelitian [10] membandingkan algoritma machine learning untuk mendeteksi penyakit jantung menggunakan dataset dari UCI Machine Learning Repositor yang berjumlah 303 data. XGBoost menunjukkan performa terbaik dengan akurasi 95,08%, mengungguli MLP (93,44%), Random Forest (91,80%), Logistic Regression (90,16%), SVM (90,16%), Gradient Boosting (88,52%), dan KNN (88,52%). Penelitian lain [11] juga membandingkan xgboost untuk memprediksi penyakit kanker payudara dengan menggunakan dataset total 569 data dari kaggle yang di split menjadi 70:30 data train dan test. Penelitian ini menunjukan bahwa hasil akurasi performa kinerja algoritma XGBoost sebesar 97,03%, sedangkan LightGBM hanya mendapatkan akurasi sebesar 95,59% .

Berdasarkan penelitian sebelumnya, penelitian ini berfokus pada pengembangan sistem klasifikasi untuk diagnosis penyakit anemia menggunakan algoritma *eXtreme Gradient Boosting* (XGBoost), karena menunjukkan akurasi yang cukup baik dan terbukti sangat efektif dalam memodelkan dataset berukuran kecil hingga menengah dibandingkan algoritma lainnya [10][11]. XGBoost juga memiliki kemampuan untuk memastikan generalisasi yang baik pada data baru[12]. Selain itu, penelitian ini juga melibatkan teknik hyperparameter tuning mengingat pemilihan hyperparameter yang tepat merupakan kunci dalam meningkatkan akurasi model Machine Learning[13]. Teknik optimasi GridSearchCV digunakan karena memiliki keunggulan dalam melakukan pencarian kombinasi parameter secara sistematis, meskipun membutuhkan waktu lebih banyak[14]. Penelitian ini juga mengevaluasi teknik oversampling *Adaptive Synthetic Sampling* (ADASYN) [9] sebagai strategi penambahan data sintesis pada kelas minoritas, sehingga model lebih responsif. Untuk itu, dilakukan perbandingan kinerja model XGBoost sebelum dan sesudah menerapkan teknik ADASYN dengan menggunakan GridSearchCV. Penelitian bertujuan untuk mengembangkan model XGBoost yang secara optimal meningkatkan akurasi diagnosis anemia berbasis data hematologi.

1.2 Rumusan Masalah

Berdasarkan penjabaran latar belakang di atas, permasalahan dalam tugas akhir ini dapat dirumuskan sebagai berikut:

1. Bagaimana penerapan *hyperparameter* menggunakan algoritma *eXtreme Gradient Boosting* (XGBoost) secara optimal pada klasifikasi diagnosis penyakit anemia berdasarkan data hematologi?
2. Bagaimana pengaruh teknik oversampling ADASYN dengan hyperparameter menggunakan algoritma *eXtreme Gradient Boosting* (XGBoost) pada klasifikasi penyakit anemia berdasarkan data hematologi?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan penelitian ini adalah sebagai berikut :

1. Mengembangkan algoritma *eXtreme Gradient Boosting* (XGBoost) secara optimal untuk klasifikasi penyakit anemia berdasarkan data hematologi.

2. Mengevaluasi pengaruh teknik oversampling ADASYN untuk mengatasi ketidak seimbangan kelas pada klasifikasi penyakit anemia menggunakan *eXtreme Gradient Boosting* (XGBoost) dengan hyperparameter tuning.

1.4 Batasan Masalah

1. Data yang digunakan dalam penelitian ini berasal dari dokumentasi rekam medis Rumah Sakit Pelengkap Medical Jombang.
2. Menggunakan platform google colab dengan bahasa pemrograman Python.
3. Fokus penelitian adalah pada klasifikasi kategori anemia dan tidak anemia.
4. Penelitian ini hanya menggunakan data hematologi yang terdiri dari HGB, PCV, RBC, dan MCV untuk klasifikasi diagnosis anemia.
5. Penelitian ini menggunakan hyperparameter tuning *GridSearchCV* , untuk mengoptimalkan performa model *eXtreme Gradient Boosting* (XGBoost).
6. Teknik oversampling yang digunakan untuk mengatasi ketidakseimbangan kelas adalah *Adaptive Synthetic Sampling* (ADASYN).