

1. Latar Belakang

Anemia merupakan salah satu masalah kesehatan global dengan prevalensi tinggi, terutama di negara berkembang termasuk Indonesia. Kondisi ini terjadi ketika kadar hemoglobin dalam darah lebih rendah dari normal, sehingga menyebabkan tubuh kekurangan oksigen. Diagnosis anemia biasanya dilakukan melalui pemeriksaan laboratorium terhadap beberapa parameter hematologi, seperti hemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), serta faktor penunjang lain seperti ferritin dan kadar zat besi.

Meskipun metode konvensional masih digunakan, proses diagnosis manual membutuhkan waktu, rentan terhadap kesalahan interpretasi, serta sering kali kurang efektif dalam menangani jumlah pasien yang besar. Oleh karena itu, penerapan teknologi *machine learning* menjadi salah satu solusi potensial untuk meningkatkan akurasi dan efisiensi dalam klasifikasi penyakit anemia.

Beberapa penelitian terdahulu telah menggunakan algoritma seperti SVM, Random Forest, dan XGBoost dalam klasifikasi anemia. Namun, penggunaan CatBoost—algoritma berbasis *gradient boosting* yang optimal untuk data tabular dan mampu menangani data kategorikal tanpa pra-pemrosesan kompleks—masih jarang dieksplorasi dalam konteks ini.

Selain akurasi, masalah lain yang muncul adalah sifat model *machine learning* yang sering dianggap sebagai “black box”. Hal ini menyulitkan praktisi medis untuk memahami alasan di balik prediksi yang dihasilkan. Untuk mengatasi hal tersebut, pendekatan interpretabilitas model dibutuhkan. Salah satunya adalah SHAP (SHapley Additive exPlanations), yang dapat memberikan penjelasan kontribusi setiap fitur terhadap prediksi model.

Dengan demikian, riset ini berfokus pada pembangunan model klasifikasi anemia menggunakan CatBoost dan interpretasi faktor penentu menggunakan SHAP, sehingga hasil penelitian dapat memberikan kontribusi pada bidang *explainable AI* untuk kesehatan.

2. Rumusan Masalah

Berdasarkan uraian di atas, permasalahan yang dapat dirumuskan adalah sebagai berikut:

1. Bagaimana membangun model klasifikasi penyakit anemia menggunakan algoritma CatBoost?
2. Bagaimana menerapkan metode SHAP untuk menginterpretasi faktor-faktor yang berpengaruh terhadap klasifikasi anemia?
3. Faktor medis apa saja yang paling signifikan memengaruhi klasifikasi anemia menurut hasil interpretasi SHAP?

3. Identifikasi Masalah

Beberapa masalah yang dapat diidentifikasi antara lain:

- Penelitian klasifikasi anemia dengan algoritma CatBoost masih terbatas.
- Model machine learning umumnya sulit dipahami (*black box problem*).
- Diperlukan metode interpretabilitas untuk menjelaskan prediksi model agar dapat dipercaya oleh tenaga medis.
- Dataset medis sering kali terbatas, tidak seimbang, dan memiliki variasi yang kompleks.

4. Tujuan

Tujuan dari penelitian ini adalah:

1. Membangun model klasifikasi penyakit anemia menggunakan algoritma CatBoost.
2. Menggunakan metode SHAP untuk menginterpretasi faktor penentu anemia.
3. Membandingkan hasil interpretasi model dengan literatur medis yang relevan.