

PROPOSAL RISET INFORMATIKA

KLASIFIKASI ANEMIA BERBASIS CATBOOST DAN ANALISIS INTERPRETABILITAS MODEL DENGAN SHAP

Vika Rafi Ana

NPM : 22081010009

Latar Belakang

Anemia masih menjadi masalah kesehatan utama di Indonesia dengan prevalensi tinggi pada berbagai kelompok masyarakat. Deteksi biasanya dilakukan melalui pemeriksaan darah (CBC), namun interpretasinya masih manual dan bergantung pada tenaga medis. Untuk mengatasinya, penelitian ini mengembangkan model klasifikasi anemia menggunakan algoritma CatBoost, yang mampu mengenali pola data medis secara akurat. Selain itu, digunakan metode SHAP untuk menjelaskan pengaruh setiap parameter darah terhadap hasil klasifikasi, sehingga model menjadi lebih transparan dan mudah dipahami oleh tenaga medis.

Rumusan Masalah

01.

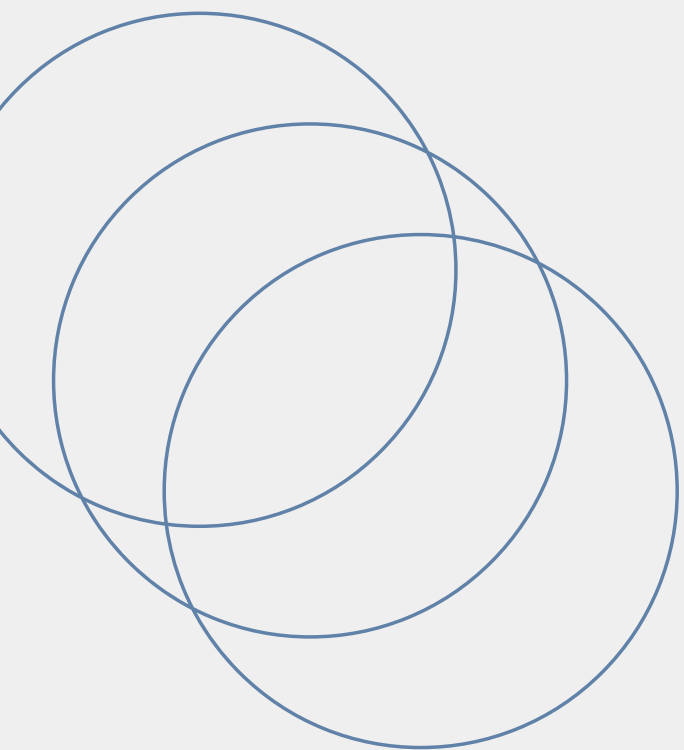
Bagaimana penerapan algoritma CatBoost dalam mengklasifikasikan penyakit anemia berdasarkan data Complete Blood Count (CBC)?

02.

Bagaimana hasil evaluasi performa model klasifikasi CatBoost dibandingkan dengan model pembandingan lainnya?

03.

Bagaimana interpretasi hasil klasifikasi menggunakan metode SHAP untuk mengetahui faktor-faktor dominan dalam prediksi anemia?



Tujuan Penelitian

01.

Menerapkan algoritma CatBoost untuk melakukan klasifikasi penyakit anemia berdasarkan data CBC.

02.

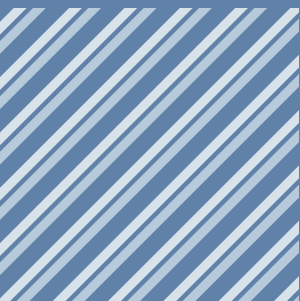
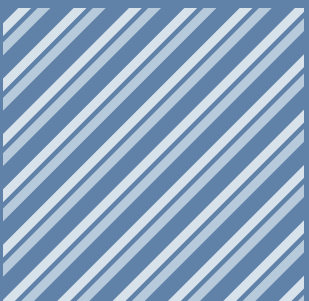
Mengevaluasi performa model menggunakan metrik akurasi, presisi, recall, dan f1-score.

03.

Menganalisis interpretabilitas model menggunakan metode SHAP untuk menjelaskan pengaruh fitur terhadap hasil prediksi.

Penelitian Terdahulu

- Penelitian “Predicting Diabetes Mellitus Using CatBoost Classifier and SHAP Approach” menunjukkan bahwa CatBoost menghasilkan akurasi tertinggi sebesar 95,7%, dan metode SHAP berhasil mengidentifikasi fitur paling berpengaruh seperti kadar glukosa dan tekanan darah.
- Penelitian “SHAP-Based Predictive Modeling for 1-Year Readmission Risk in Elderly Heart Failure Patients” menggunakan CatBoost + SHAP dengan hasil AUC 0.89, menunjukkan model mampu memprediksi risiko pasien secara akurat sekaligus memberikan interpretasi fitur yang jelas.



Gap Riset




Sebagian besar penelitian terkait klasifikasi anemia masih menggunakan algoritma seperti Random Forest, SVM, dan Logistic Regression. Namun, penerapan CatBoost dalam konteks ini masih jarang ditemukan. Selain itu, banyak penelitian yang hanya menekankan pencapaian akurasi tanpa memperhatikan interpretabilitas model.



Dataset

Data yang digunakan dalam penelitian ini berasal dari dataset publik berjudul “Anemia Types Classification”, yang tersedia di platform Kaggle.

Dataset tersebut terdiri atas sekitar 1.282 baris data dengan 15 atribut, yang merepresentasikan berbagai parameter hasil pemeriksaan darah atau Complete Blood Count (CBC).



Tahapan Penelitian



Pengumpulan Data

Prarocessing Data

Pembagian Data (Train-Test Split)

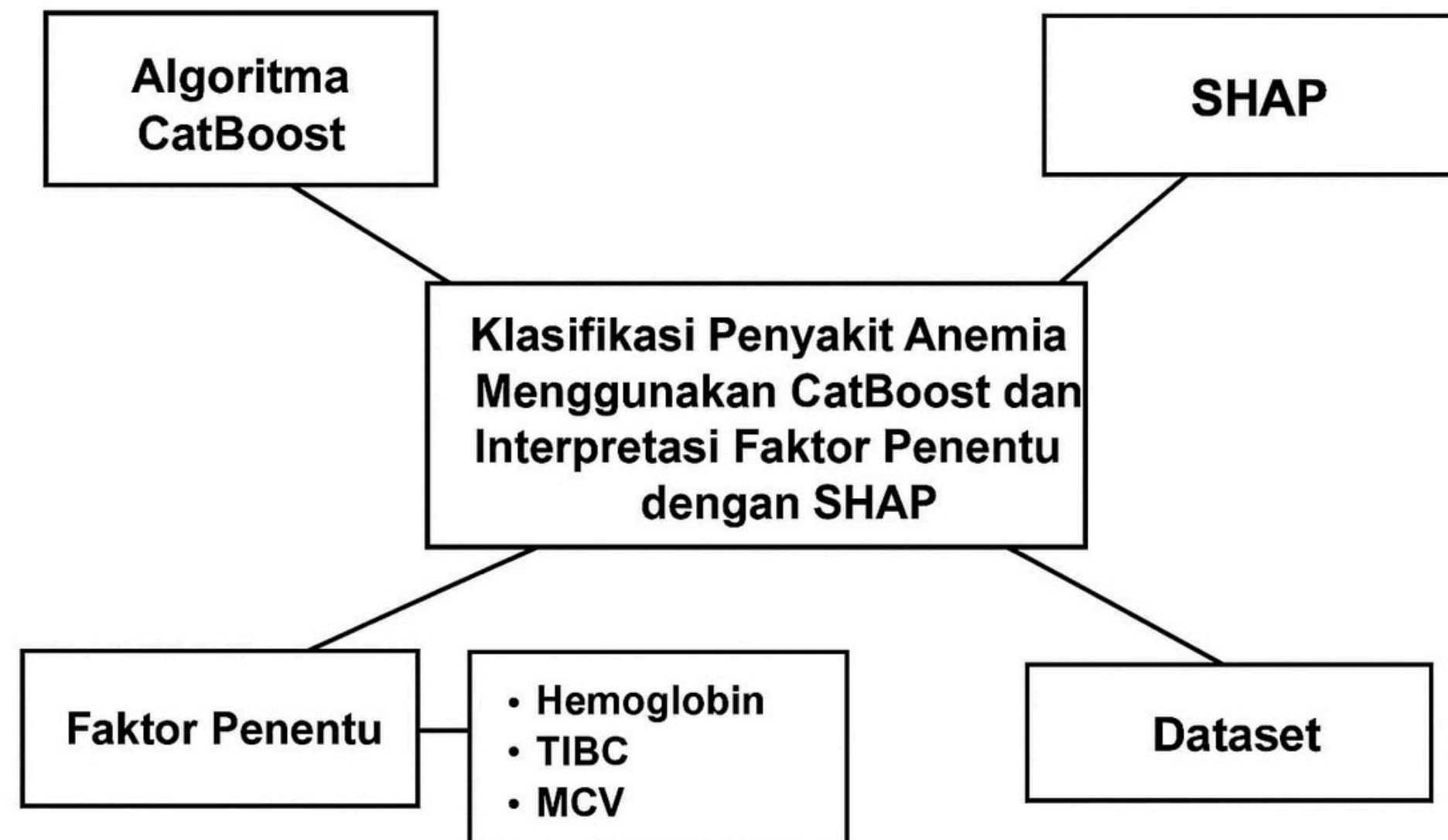
Penerapan algoritma CatBoost

Interpretasi Model dengan SHAP

Evaluasi Hasil

Evaluasi performa model dilakukan dengan menggunakan Confusion Matrix serta lima metrik utama, yaitu accuracy, precision, recall, F1-score, dan AUC (Area Under the Curve).

MIND MAP





Terima Kasih

