

**ANEMIA CLASSIFICATION WITH CLINICAL FEATURE
ENGINEERING AND SHAP INTERPRETATION**

LAPORAN NONREGULER

Diajukan untuk memenuhi salah satu syarat mencapai derajat
Sarjana Program Studi Informatika



Disusun oleh :
IKHLASUL AMALIA
22.11.5304

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

ANEMIA CLASSIFICATION WITH CLINICAL FEATURE ENGINEERING AND SHAP INTERPRETATION

LAPORAN NONREGULER

Diajukan untuk memenuhi salah satu syarat mencapai derajat
Sarjana Program Studi Informatika



Disusun oleh :
IKHLASUL AMALIA
22.11.5304

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**ANEMIA CLASSIFICATION WITH CLINICAL FEATURE
ENGINEERING AND SHAP INTERPRETATION**

yang disusun dan diajukan oleh

Ikhlasul Amalia

22.11.5304

telah disetujui oleh Dosen Pembimbing
pada tanggal 19 November 2025

Dosen Pembimbing,



Rumini, S.Kom., M.Kom.
NIK. 190302246

HALAMAN PENGESAHAN

JALUR NON-REGULER

**ANEMIA CLASSIFICATION WITH CLINICAL
FEATURE ENGINEERING AND SHAP
INTERPRETATION**

yang disusun dan diajukan oleh
Ikhlasul Amalia
22.11.5304

Telah dipertahankan di depan Dewan Penguji
pada tanggal 19 November 2025

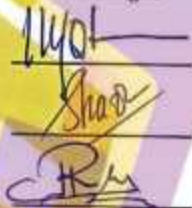
Nama Penguji **Susunan Dewan Penguji**

Tanda Tangan

Uvoek Anggoro Saputra, S.Kom., M.Kom
NIK. 190302419

Sharazita Dyah Angeita, S.Kom., M.Kom
NIK. 190302285

Rumini, S.Kom., M.Kom
NIK. 190302246



Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 19 November 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom
NIK. 1903021

Anonymous Classification with Clinical Feature Engineering and SHAP Interpretation

Dosen Pembimbing : Ramini, S.Kom., M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 November 2025
Yang Menyatakan,

Ikhlasul Amalia

NIM : 22.11.5304

Menyatakan bahwa Laporan dengan judul berikut:

Anemia Classification with Clinical Feature Engineering and SHAP Interpretation

Dosen Pembimbing : Rumini, S.Kom., M.Kom.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali seperti tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 November 2025

Yang Menyatakan,



Ikhtisaul Armatia

HALAMAN PERSEMBAHAN

Dengan menyebut nama Allah Yang Maha Pengasih lagi Maha Penyayang, segala puji syukur penulis panjatkan ke hadirat Allah SWT. Atas limpahan rahmat, kasih sayang, dan kemudahan-Nya, penulis akhirnya dapat menuntaskan Tugas Akhir ini, sebaga wujud tanggung jawab akademik sekaligus langkah akhir dalam menempuh pendidikan sarjana. Kehadiran karya ini menjadi bukti perjalanan panjang penuh doa, usaha, dan cinta dari orang-orang terkasih. Dengan penuh rasa syukur, penulis mempersembahkan karya ini kepada:

1. Dengan segenap rasa syukur, penulis haturkan segala puji ke hadirat Allah SWT yang atas rahmat dan ridha-Nya telah memberikan kekuatan, kesehatan, dan kemudahan, sehingga penulis dapat menyelesaikan penelitian ini.
2. Kedua orang tua tercinta, yang menjadi sumber kekuatan penulis. Atas segala limpahan kasih sayang yang tak hingga, doa yang tiada henti, dukungan moril maupun material yang tak pernah putus, serta pengorbanan dalam mengusahakan segala hal demi menunjang kehidupan dan studi penulis. Rasa terima kasih yang tidak cukup penulis sampaikan di sini, segala cinta kasih dan dukungan yang Mama dan Bapak berikan. Semoga doa Mama dan Bapak selalu menyertai di setiap proses menuju terbaik penulis.
3. Ibu Rumini, S.Kom, M.Kom, selaku Dosen Pembimbing, atas segala bimbingan, arahan, semangat yang diberikan, serta kesabarannya dalam membimbing penulis hingga terselesaikannya penelitian ini.
4. Bapak Abd. Mizwar A. Rahim, S.Kom, M.Kom., selaku Dosen Pengampu mata kuliah Data Mining, atas perhatian, semangat, dukungan, dan arahan kepada penulis.
5. Keluarga besar Bani Suwondo MS., yang selalu mendukung dan mendoakan penulis, terima kasih atas segala perhatian yang selalu menguatkan.

6. Sila Cahya Dewi, sahabat penulis. Terima kasih telah menjadi teman dekat nan erat yang selalu kebersamai dan memberikan kebaikan tiada henti. Lama sebetulnya pertemuan ini telah memberikan arti pertemanan yang tulus bagi penulis.
7. Teman-teman terdekat penulis yang namanya tidak dapat disebutkan satu per satu, terima kasih atas afeksi, tawa, dan kebersamaan yang senantiasa menemani langkah penulis sampai dengan titik ini.
8. Rekan seperjuangan kelas Informatika 12, terima kasih atas kerja sama, dukungan, dan kebersamaan yang terbangun di setiap semester.
9. Ikhlusul Amalia, terima kasih telah berproses dengan sungguh-sungguh versi diri sendiri. Konsistensi dan keberanian untuk belajar membawamu sampai pada pencapaian hari ini. Perjalananmu masih panjang; tetaplah menyediakan ruang kosong agar terus tumbuh, belajar, dan mengisi gelas kehidupanmu dengan lebih baik.

KATA PENGANTAR

Segala puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas rahmat dan karunia-Nya sehingga laporan kelulusan ini dapat terselesaikan dengan baik. Keberhasilan ini tidak terlepas dari bimbingan, dukungan, dan bantuan tulus berbagai pihak. Untuk itu, dengan penuh hormat penulis menyampaikan ucapan terima kasih yang mendalam kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Prof. Dr. Kusri, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi SI Informatika Universitas AMIKOM Yogyakarta.
4. Rumini, S.Kom., M.Kom., selaku Dosen Pembimbing, atas bimbingan, arahan, serta motivasi yang senantiasa diberikan hingga terselesaikannya penelitian ini.
5. Orang tua penulis, atas cinta, kasih sayang, doa, dukungan, dan pengorbanan yang tidak ternilai harganya.

Akhir kata, penulis menyadari bahwa laporan ini masih jauh dari kata sempurna dan memiliki banyak keterbatasan. Oleh karena itu, segala bentuk kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa yang akan datang. Semoga karya ini dapat memberikan manfaat bagi pembaca dan pengembangan ilmu pengetahuan.

Yogyakarta, 17 November 2025

Penulis

DAFTAR ISI

Halaman Judul.....	i
Halaman Persetujuan	ii
Halaman Pengesahan.....	iii
Halaman Pernyataan Keaslian Karya	iv
Halaman Persembahan.....	v
Kata Pengantar	vii
Daftar Isi.....	viii
Daftar Tabel	x
Daftar Gambar.....	xi
Daftar Lambang dan Singkatan	xii
Daftar Istilah	xiii
Intisari.....	xiv
Abstract.....	xv
Bab I Pendahuluan	1
1.1. Gambaran Umum.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah	3
1.4. Tujuan	4
Bab II Tinjauan Pustaka	5
2.1. Studi Literatur	5
2.2. Landasan Teori.....	8
2.2.1 Anemia.....	8
2.2.2 Parameter Hematologi	8
2.2.3 <i>Machine Learning</i>	9
2.2.4 Algoritma Klasifikasi	10
2.2.5 <i>Feature Engineering</i>	12
2.2.6 Evaluasi Model	13
2.2.7 SHAP	15
2.2.8 <i>Preprocessing Data</i>	17

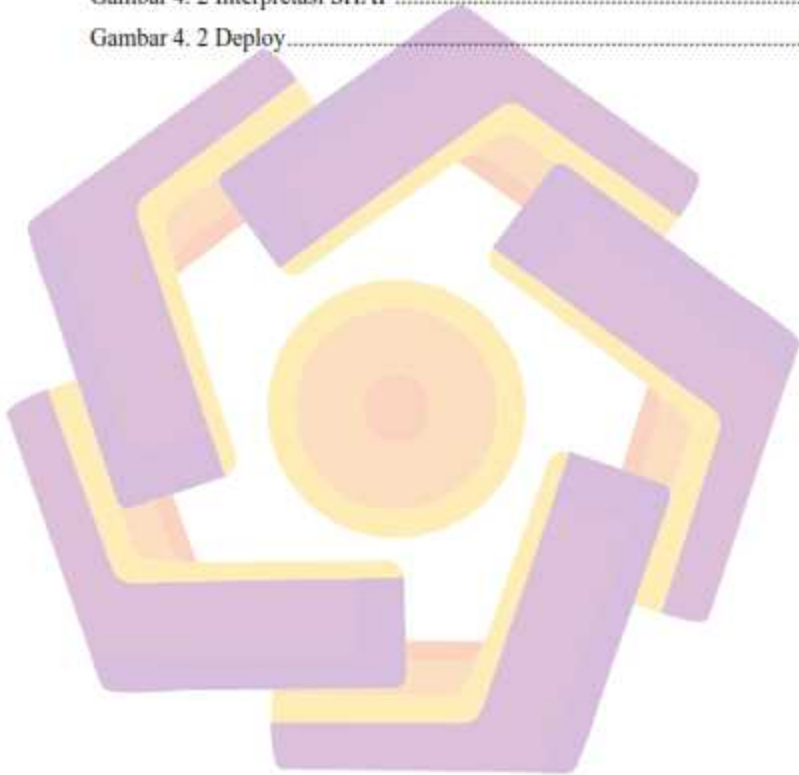
BAB III Metode Penelitian.....	18
3.1. Metode	18
3.2. Alur Penelitian.....	18
3.3. Dataset	20
3.4. Metode <i>Preprocessing</i> Data	21
3.5. <i>Featur Engineering</i>	22
3.6. Splitting Data	23
3.7. Pemodelan	23
3.8. Evaluasi Model.....	23
3.9. Metode Interpretasi SHAP.....	24
BAB IV Pembahasan	26
4.1. Gambaran Umum	26
4.2. Preprocessing Data	26
4.3. Pembagian Data	27
4.4. <i>Feature Engineering</i>	27
4.5. Hasil Evaluasi Model.....	30
4.6. Analisis Confusion Matrix.....	35
4.7. Validasi dengan K-Fold Cross Validation.....	36
4.8. Interpretasi SHAP	37
4.9. Pembahasan.....	37
BAB V Kesimpulan	40
1.1. Kesimpulan	40
1.2. Saran	41
Referensi	42
Lampiran dan Bukti Pendukung	42
a. Letter of Acceptance (LOA)	45
b. Lembar Review	46
c. Bukti Terbit/Terindex.....	49
d. Bukti pembayaran	50

DAFTAR TABEL

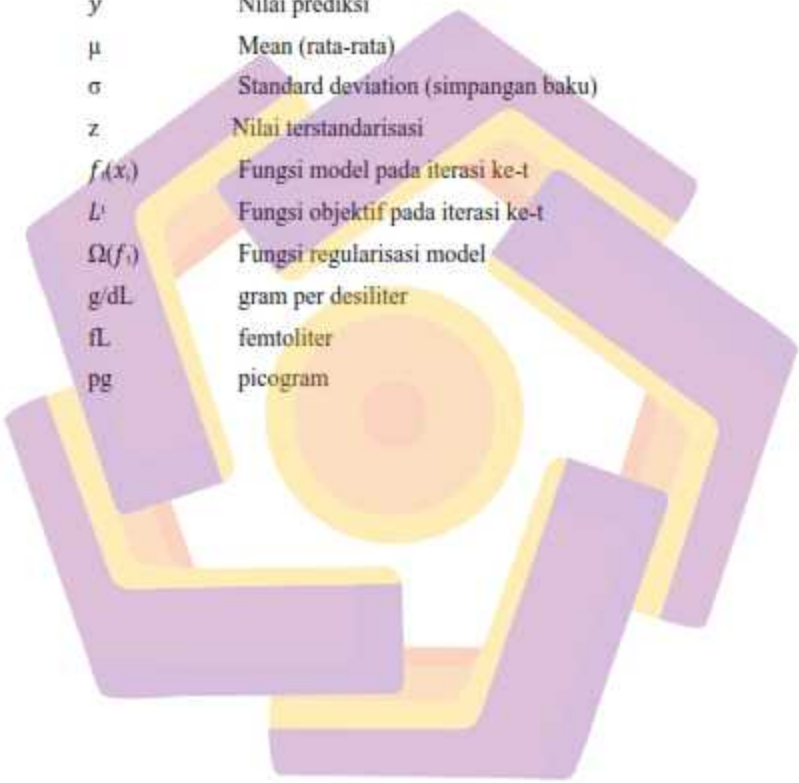
Tabel 2. 1 Confusion Matrix	14
Tabel 3. 1 Deskripsi Atribut Dataset Anemia	19
Tabel 4. 1 Splitting Data	20
Tabel 4. 2 Perbandingan Kinerja Model	30
Tabel 4. 3 Sampel Perhitungan Manual	31
Tabel 4. 4 Perhitungan Linear LogReg	31
Tabel 4. 5 Perhitungan Sigmoidl	31
Tabel 4. 6 Konversi	32
Tabel 4. 7 Prediksi	32
Tabel 4. 8 Majority Voting	33
Tabel 4. 9 Pemisahan Node	34

DAFTAR GAMBAR

Gambar 3. 1 Diagram Alur Penelitian	22
Gambar 4. 1 Confusion Matrix XGBoost.....	36
Gambar 4. 2 Interpretasi SHAP	37
Gambar 4. 2 Deploy.....	38



DAFTAR LAMBANG DAN SINGKATAN



y_i	Nilai label aktual
$\hat{y}$	Nilai prediksi
μ	Mean (rata-rata)
σ	Standard deviation (simpangan baku)
z	Nilai terstandarisasi
$f_t(x_i)$	Fungsi model pada iterasi ke- t
L_t	Fungsi objektif pada iterasi ke- t
$\Omega(f_t)$	Fungsi regularisasi model
g/dL	gram per desiliter
fL	femtoliter
pg	picogram

DAFTAR ISTILAH

<i>Hemoglobin (Hb)</i>	Protein dalam sel darah merah yang membawa oksigen.
<i>Mean Corpuscular Volume (MCV)</i>	Ukuran rata-rata sel darah merah.
<i>Mean Corpuscular Hemoglobin (MCH)</i>	Jumlah rata-rata hemoglobin per sel darah merah.
<i>Mean Corpuscular Hemoglobin Concentration (MCHC)</i>	Konsentrasi rata-rata hemoglobin dalam sel darah merah.
<i>Machine Learning</i>	Pembelajaran pola secara otomatis oleh komputer berdasarkan data.
<i>Overfitting</i>	Kondisi ketika model terlalu sesuai dengan data latih sehingga buruk pada data baru.
<i>Underfitting</i>	Kondisi ketika model terlalu sederhana sehingga tidak menangkap pola penting.
<i>Cross Validation</i>	Teknik evaluasi model dengan membagi data menjadi beberapa lipatan (fold).
<i>Feature Engineering</i>	Proses modifikasi atau pembuatan fitur baru untuk meningkatkan kinerja model.
<i>Confusion Matrix</i>	Matriks yang menunjukkan prediksi benar dan salah dari model.
<i>Evaluation Metrics</i>	Ukuran untuk menilai performa model (misalnya akurasi, presisi, recall).
<i>Precision</i>	Proporsi prediksi positif yang benar ($TP / (TP + FP)$).
<i>Recall</i>	Proporsi kasus positif yang berhasil terdeteksi ($TP / (TP + FN)$).
<i>Oversampling</i>	Menambah jumlah sampel pada kelas minoritas untuk mengatasi ketidakseimbangan data.
<i>RandomizedSearchCV</i>	Metode pencarian hyperparameter secara acak untuk menemukan kombinasi terbaik.
<i>Dataset</i>	Kumpulan data yang digunakan untuk analisis dan pelatihan model.
<i>Label Encoding</i>	Proses mengubah variabel kategorikal menjadi numerik.
<i>Standardization</i>	Teknik normalisasi data agar mengikuti distribusi standar (mean 0, std 1).
<i>Train-Test Split</i>	Pembagian dataset menjadi data pelatihan dan data pengujian.

INTISARI

Anemia merupakan masalah kesehatan global yang memiliki dampak signifikan terhadap kualitas hidup dan produktivitas. Deteksi dini dan akurat sangat penting untuk mencegah komplikasi yang lebih serius. Studi ini bertujuan untuk mengembangkan model klasifikasi anemia berbasis teknologi machine learning menggunakan algoritma XGBoost, serta membandingkan kinerjanya dengan metode Regresi Logistik dan Hutan Acak. Data yang digunakan dalam studi ini diperoleh dari platform Kaggle, terdiri dari 1.421 sampel dan enam atribut klinis, yaitu Jenis Kelamin, Hemoglobin (HGB), Rata-rata Hemoglobin Korpuskular (MCH), Konsentrasi Hemoglobin Korpuskular Rata-rata (MCHC), Volume Korpuskular Rata-rata (MCV), dan Hasil. Selama proses engineering fitur, fitur turunan rasio hemoglobin terhadap MCV (Hb/MCV) ditambahkan, yang secara medis relevan dalam membedakan jenis anemia. Hasil evaluasi menunjukkan bahwa XGBoost dan Random Forest mencapai tingkat akurasi dan F1-Score sebesar 100%, sementara Regresi Logistik mencapai tingkat 98,9%. XGBoost dipilih sebagai model utama karena kemampuannya yang efisien dalam komputasi dan dukungan untuk interpretasi menggunakan SHAP (SHapley Additive exPlanations). Visualisasi SHAP menunjukkan bahwa rasio Hb/MCV dan hemoglobin adalah fitur yang paling berpengaruh dalam klasifikasi. Model ini berpotensi digunakan sebagai sistem pendukung keputusan untuk skrining anemia otomatis dan dapat diintegrasikan lebih lanjut ke dalam sistem klinis.

Kata kunci: Anemia, Rasio Hb/MCV , Machine learning, SHAP, XGBoost

ABSTRACT

Anemia is a global health issue that has a significant impact on quality of life and productivity. Early and accurate detection is essential to prevent more serious complications. This study aims to develop an anemia classification model based on machine learning technology using the XGBoost algorithm, as well as compare its performance with Logistic Regression and Random Forest methods. The dataset used in this study was obtained from the Kaggle platform, consisting of 1,421 samples and six clinical attributes, namely Gender, Hemoglobin (HGB), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Hemoglobin Concentration (MCHC), Mean Corpuscular Volume (MCV), Result. During the feature engineering process, the derived feature of the hemoglobin-to-MCV ratio (Hb/MCV) was added, which is medically relevant in distinguishing types of anemia. Evaluation results showed that XGBoost and Random Forest achieved an accuracy rate and F1-Score of 100%, while Logistic Regression achieved a rate of 98.9%. XGBoost was selected as the primary model due to its efficient computational capabilities and support for Interpretation using SHAP (SHapley Additive exPlanations). SHAP visualization revealed that the Hb/MCV ratio and hemoglobin were the most influential features in classification. This model has the potential to be used as a decision support system for automated anemia screening and can be further integrated into clinical systems.

Keyword: Anemia, Rasio Hb/MCV , Machine learning, SHAP, XGBoost