KLASIFIKASI ANEMIA MENGGUNAKAN ALGORITMA XGBOOST BERDASARKAN DATA HEMATOLOGI

SKRIPSI

Oleh :
ZAUDA SALMA SALSABILA
NIM. 220605110043

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026

KLASIFIKASI ANEMIA MENGGUNAKAN ALGORITMA XGBOOST BERDASARKAN DATA HEMATOLOGI

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
ZAUDA SALMA SALSABILA
NIM. 220605110043

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2026

HALAMAN PERSETUJUAN

KLASIFIKASI ANEMIA MENGGUNAKAN ALGORITMA XGBOOST BERDASARKAN DATA HEMATOLOGI

SKRIPSI

Oleh :
ZAUDA SALMA SALSABILA
NIM. 220605110043

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 30 Mei 2026

Pembimbing I,	Pembimbing II,
Prof. Dr. Suhartono, M.Kom	Roro Inda Melani, MT, M.Sc
NIP. 19680519 200312 1 001	NIP. 19780925 200501 2 008

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Sunardi, M. Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

KLASIFIKASI ANEMIA MENGGUNAKAN ALGORITMA XGBOOST BERDASARKAN DATA HEMATOLOGI

SKRIPSI

Oleh :
ZAUDA SALMA SALSABILA
NIM. 220605110043

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 19 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: Prof. Dr. Muhammad Faisal, M.T NIP. 19740510 200501 1 007	( )
Anggota Penguji I	: A'la Syauqi, M.Kom NIP. 19771201 200801 1 007	( )
Anggota Penguji II	: Prof. Dr. Suhartono, M.Kom NIP. 19680519 200312 1 001	( )
Anggota Penguji III	: Roro Inda Melani, MT, M.Sc NIP. 19780925 200501 2 008	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M. Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

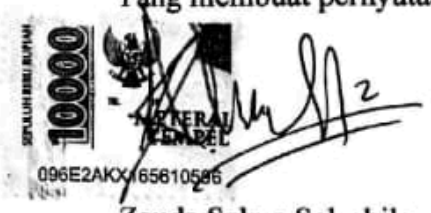
Nama	: Zauda Salma Salsabila
NIM	: 220605110043
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Klasifikasi Anemia Menggunakan Algoritma XGBoost Berdasarkan Data Hematologi

Menyatakan dengan sebenarnya bahwa Skripsi yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsi ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 19 Juni 2026

Yang membuat pernyataan,



Zauda Salma Salsabila

NIM.220605110043

MOTTO

“tetap jadi dirimu sendiri, sampai kau menemukan cinta yang membuatmu tetap hidup tanpa memintamu untuk berubah”

"If I have seen further, it is by standing on the shoulders of giants."

- *Isaac Newton* -

HALAMAN PERSEMBAHAN

Alhamdulillahi rabbil ‘alamin puji syukur kepada Allah Subḥānahu wa Ta‘ālā karena berkat rahmat dan petunjuk-Nya penulis dapat menyelesaikan skripsi ini. Shalawat serta salam selalu tercurahkan kepada Rasulullah Ṣallallāhu ‘Alaihi Wasallam yang telah membawa kita dari zaman jahiliyah menuju addinul Islam. Skripsi ini penulis persembahkan kepada seluruh pihak yang telah berperan dan berjasa dalam perjalanan penelitian ini.

Pertama, kepada diri sendiri. Terima kasih karena telah bertahan, berjuang, dan terus melangkah melewati berbagai tantangan hingga mampu menyelesaikan perjalanan akademik ini dengan sebaik-baiknya. .

Kedua, penulis persembahkan skripsi ini kepada kedua orang tua. Terima kasih atas doa yang tidak pernah putus, kasih sayang yang tulus, dukungan yang selalu menguatkan, serta keyakinan dan kepercayaan yang diberikan kepada penulis dalam setiap langkah untuk menggapai cita-cita. Penulis juga mempersembahkan kepada adik-adik tercinta yang senantiasa menjadi penguat dan motivasi, semangat, dan dukungan selama proses penyusunan skripsi ini.

Terakhir, kepada sahabat-sahabat seperjuangan skripsi dan teman-teman angkatan 2022 yang telah hadir sebagai teman berbagi cerita, memberikan dukungan, semangat, serta menemani setiap proses yang dilalui, baik dalam suka maupun duka, hingga skripsi ini dapat terselesaikan.

Semoga segala kebaikan, bantuan dan doa yang telah diberikan mendapatkan balasan terbaik dari Allah Subḥānahu wa Ta‘ālā. Aamiin.

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh.

Segala puji hanya milik Allah Subḥānahu wa Ta'ālā atas segala nikmat dan kasih sayang-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul "Klasifikasi Anemia Menggunakan Algoritma XGBoost Berdasarkan Data Hematologi". Shalawat dan salam senantiasa terlimpah kepada Nabi Muhammad Ṣallallāhu 'Alaihi Wasallam. Dan semoga kita semua mendapat syafaatnya di hari kiamat nanti, Aamiin.

Dalam penulisan skripsi ini, penulis mengucapkan rasa syukur dan terima kasih yang tak terhingga kepada semua pihak-pihak yang selalu memberikan bantuan dan motivasi kepada penulis untuk dapat menyelesaikan skripsi ini. Ucapan ini penulis sampaikan kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., CAHRM., CRMP., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim Malang.
2. Dr. Agus Mulyono, M.Kes., selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi Teknik Informatika Universitas Islam Negeri Maulana Malik Ibrahim Malang.
4. Prof. Dr. Suhartono, S.Si, M.Kom., selaku Dosen Pembimbing I sekaligus Dosen Penguji II, yang sejak awal pengajuan pra proposal hingga skripsi ini selesai selalu membersamai proses penulis dengan penuh kesabaran. Terima kasih yang sebesar-besarnya atas waktu dan ilmu dari Bapak.

Kehadiran dan bimbingan Bapak menjadi bagian penting yang menguatkan penulis hingga skripsi ini dapat terselesaikan.

5. Roro Inda Melani, M.T, M.Sc., selaku Dosen Pembimbing II sekaligus Dosen Penguji III, terima kasih banyak atas segala saran dan koreksinya. Penulis sangat terbantu dengan ketelitian dalam memeriksa bagian-bagian yang detail, sehingga skripsi ini menjadi lebih baik dan rapi.
6. Prof. Dr. Muhammad Faisal, M.T., selaku Dosen Ketua Penguji dan A'la Syauqi, M.Kom selaku Dosen Penguji I yang telah bersedia menguji serta memberikan masukan sehingga penulis dapat menuntaskan skripsi dengan baik.
7. Segenap Dosen, Admin, Laboran dan Teman-teman Mahasiswa Program Studi Teknik Informatika yang telah memberikan banyak dukungan dan bimbingan selama proses kuliah, dan pengerjaan skripsi ini.
8. Orang tua serta saudara saya yang selalu memberikan dukungan dan motivasi untuk terus berusaha, dan doa yang tak putus-putusnya selalu disampaikan agar dapat menuntaskan skripsi ini dengan lancar dan baik.
9. Terakhir, kepada diri sendiri yang telah berjuang dan berproses hingga sampai pada tahap ini. Terima kasih karena tetap bertahan, terus belajar, dan berusaha memberikan yang terbaik di tengah berbagai tantangan yang dihadapi selama perjalanan penyusunan skripsi ini.

Malang, 19 Juni 2026

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Salsabila, Zauda Salma. 2026. **Klasifikasi Anemia Menggunakan Algoritma XGBoost Berdasarkan Data Hematologi**. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Prof. Dr. Suhartono, S.Si, M.Kom (II) Roro Inda Melani, MT, M.Sc.

Kata Kunci: Anemia, XGBoost, Data Hematologi, Grid Search.

Anemia merupakan masalah kesehatan global yang membutuhkan deteksi dini secara cepat dan akurat. Diagnosis manual berbasis parameter hematologi seringkali membutuhkan tenaga ahli dan rentan terhadap inkonsistensi, terutama di fasilitas layanan kesehatan primer. Penelitian ini bertujuan mengevaluasi kinerja algoritma Extreme Gradient Boosting (XGBoost) dalam mengklasifikasikan anemia menggunakan dataset hematologi minimalis yang erdiri dari parameter hemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), dan mean corpuscular hemoglobin concentration (MCHC). Dataset berjumlah 1.421 data dengan label biner (0 = tidak anemia, 1 = anemia). Optimasi dilakukan menggunakan hyperparameter tuning Grid Search dengan variasi n_estimators (50, 70, 100) dan tiga rasio pembagian data (60:40, 70:30, 80:20), menghasilkan sembilan kombinasi eksperimen. Hasil terbaik diperoleh pada rasio 70:30 dengan n_estimators=100, menghasilkan accuracy 0,9415, precision 0,8889, recall 0,9892, dan F1-score 0,9364. Analisis misklasifikasi menunjukkan seluruh kesalahan terpusat pada kasus borderline dengan hemoglobin 12,0–13,3 g/dL. Penelitian ini membuktikan XGBoost mampu mempertahankan performa klasifikasi tinggi menggunakan parameter hematologi dasar, sehingga berpotensi diterapkan sebagai alat bantu deteksi dini anemia di fasilitas kesehatan primer.

ABSTRACT

Salsabila, Zauda Salma. 2026. **Classification of Anemia Using the XGBoost Algorithm Based on Hematology Data.** Thesis. Department of Informatics Engineering, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University Malang. Advisors: (I) Prof. Dr. Suhartono, S.Si, M.Kom (II) Roro Inda Melani, MT, M.Sc

Keywords: Anemia, XGBoost, Hematological Data, Grid Search .

Anemia is a global health problem with high prevalence requiring fast and accurate early detection. Manual diagnosis based on hematological parameters often requires expert knowledge and is prone to inconsistency, particularly in primary healthcare facilities with limited resources. This study aims to evaluate the performance of the Extreme Gradient Boosting (XGBoost) algorithm in classifying anemia using a minimal hematological dataset consisting of hemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), and mean corpuscular hemoglobin concentration (MCHC). The dataset comprises 1,421 records with binary labels (0 = non-anemia, 1 = anemia). Optimization was performed using Grid Search hyperparameter tuning with n_estimators variations (50, 70, 100) and three data splitting ratios (60:40, 70:30, 80:20), producing nine experimental combinations. The best configuration was achieved at a 70:30 split ratio with n_estimators=100, yielding accuracy of 0.9415, precision of 0.8889, recall of 0.9892, and F1-score of 0.9364. Misclassification analysis revealed all errors were concentrated in borderline cases with hemoglobin values of 12.0–13.3 g/dL. This study demonstrates that XGBoost maintains high classification performance using only basic hematological parameters, making it potentially applicable as an early anemia detection tool in primary healthcare facilities.

مستخلص البحث

سلسبيلا، زودا سلمى. ٢٠٢٦. **تصنيف حالات فقر الدم باستخدام خوارزمية XGBoost استنادًا إلى بيانات أمراض الدم**. أطروحة. قسم هندسة المعلوماتية، كلية العلوم والتكنولوجيا، جامعة مولانا مالك إبراهيم الإسلامية الحكومية في مالانغ. المشرفان: (١) الأستاذ الدكتور سوهارتونو، S.Si، M.Kom (٢) رورو إندا ميلاني، MT، M.Sc

الكلمات المفتاحية: فقر الدم، XGBoost، البيانات الدموية، البحث الشبكي.

يُعد فقر الدم مشكلة صحية عالمية ذات معدل انتشار مرتفع تتطلب الكشف المبكر السريع والدقيق. وغالبًا ما يتطلب التشخيص اليدوي القائم على المعايير الدموية معرفة متخصصة، كما أنه عرضة للتباين، لا سيما في مرافق الرعاية الصحية الأولية ذات الموارد المحدودة. تهدف هذه الدراسة إلى تقييم أداء خوارزمية Extreme Gradient Boosting (XGBoost) في تصنيف فقر الدم باستخدام مجموعة بيانات دموية محدودة تتكون من الهيموجلوبين (Hb)، وحجم الخلية الكرياتي المتوسط (MCV)، والهيموجلوبين الكرياتي المتوسط (MCH)، وتركيز الهيموجلوبين الكرياتي المتوسط (MCHC). تم إجراء التحسين باستخدام ضبط المعلمات الفائقة باستخدام «Grid Search» مع تباينات في n_estimators (٥٠ ، ٧٠ ، ١٠٠) وثلاث نسب لتقسيم البيانات (٦٠:٤٠، ٧٠:٣٠، ٨٠:٢٠)، مما أدى إلى إنتاج تسع تركيبات تجريبية. تم تحقيق أفضل تكوين عند نسبة تقسيم ٧٠:٣٠ مع n_estimators=100، مما أدى إلى دقة قدرها ٠,٩٤١٥، وضبط قدره ٠,٨٨٨٩، واسترجاع قدره ٠,٩٣٦٤. كشف تحليل الأخطاء التصنيفية عن جميع الأخطاء. استخدم تقييم النموذج F1 ٠,٩٨٩٢، ودرجة F1. وكشف تحليل الأخطاء التصنيفية أن جميع الأخطاء تركزت في الحالات مقاييس الدقة، والضبط، والاسترجاع، ومؤشر XGBoost الحدية التي تتراوح فيها قيم الهيموجلوبين بين ١٢,٠ و١٣,٣ غ/ديسيلتر. وتُظهر هذه الدراسة أن نموذج يحافظ على أداء تصنيفي عالٍ باستخدام المعلمات الدموية الأساسية فقط، مما يجعله قابلاً للتطبيق كأداة للكشف المبكر عن فقر الدم في مرافق الرعاية الصحية الأولية.

BAB I
PENDAHULUAN

1.1 Latar Belakang

Anemia merupakan salah satu masalah kesehatan yang umum terjadi di berbagai negara, termasuk Indonesia. Menurut World Health Organization (WHO) lebih dari 1,62 miliar penduduk dunia mengalami anemia, dengan prevalensi tertinggi pada ibu hamil, anak-anak, dan remaja (Safiri et al., 2021). Di Indonesia, hasil Riset Kesehatan Dasar (Riskesdas) Kemenkes 2018 menunjukkan bahwa prevalensi anemia pada ibu hamil mencapai 48,9%, sementara pada remaja putri mencapai 32%, menjadikannya salah satu masalah kesehatan masyarakat yang perlu mendapatkan perhatian khusus (Attaqy et al., 2022). Anemia ditandai oleh kadar hemoglobin yang rendah sehingga mengganggu kemampuan darah dalam membawa oksigen ke jaringan tubuh. Kondisi ini dapat menyebabkan kelelahan, penurunan fungsi fisik, gangguan kognitif, dan meningkatkan risiko infeksi serta komplikasi lain, terutama pada kelompok rentan seperti ibu hamil, anak-anak, dan remaja (An et al., 2021; Sundararajan & Rabe, 2021).

Tingginya angka kejadian anemia tersebut menunjukkan bahwa deteksi dini menjadi kebutuhan yang mendesak. Dalam praktik medis, diagnosis anemia umumnya dilakukan melalui pemeriksaan hematologi meliputi *hemoglobin*, *mean corpuscular volume* (MCV), *mean corpuscular hemoglobin* (MCH), dan *mean corpuscular hemoglobin concentration* (MCHC) merupakan indikator penting dalam diagnosis anemia (Ramzan et al., 2024). Namun, proses penilaian manual berdasarkan parameter hematologi seringkali membutuhkan tenaga ahli, memakan

waktu, dan memiliki potensi inkonsistensi, terutama pada fasilitas layanan kesehatan primer yang memiliki keterbatasan sumber daya manusia dan peralatan. Sehingga banyak kasus anemia tidak terdeteksi sejak dini. Keterlambatan deteksi inilah yang pada akhirnya meningkatkan risiko komplikasi serius, terutama pada kelompok rentan.

Selain pentingnya deteksi dini anemia dari perspektif medis, Islam juga menekankan kewajiban menjaga kesehatan sebagai bentuk amanah dari Allah Subhanahu Wa Ta'ala. Islam menempatkan kesehatan sebagai nikmat yang harus dijaga, karena kesehatan adalah syarat optimal untuk beribadah dalam menjalankan tugas sebagai khalifah di bumi(Sumarno et al., 2022). Dalam Al-Quran, Allah berfirman dalam QS. Asy-Syu'ara ayat 80, sebagai berikut:

وَاِذَا مَرِضْتُ فَهُوَ يَشْفِيْنِ

“Dan apabila aku sakit, Dialah (Allah) yang menyembuhkan aku.” (Q.S. Asy-Syua’ra: 80).

Ayat tersebut menjelaskan bahwa segala bentuk kesembuhan pada hakikatnya berasal dari Allah SWT. Namun demikian, Islam juga mengajarkan umatnya untuk melakukan ikhtiar dalam menjaga kesehatan dan mencari pengobatan ketika mengalami penyakit. Oleh karena itu, pemanfaatan teknologi dalam bidang kesehatan, seperti penerapan algoritma *Extreme Gradient Boosting* (XGBoost) untuk mengklasifikasikan kondisi anemia berdasarkan data hematologi, dapat dipandang sebagai bentuk *ikhtiar ilmiah* dalam melakukan deteksi penyakit. Upaya pendeteksian anemia melalui analisis data hematologi merupakan salah satu bentuk

untuk menjaga kesehatan yang sejalan dengan nilai-nilai Islam dan mencegah resiko gangguan tubuh yang lebih serius.

Selain itu, penggunaan teknologi seperti machine learning dalam penelitian ini sejalan dengan tujuan syariat Islam yang dikenal sebagai Maqashid Al-Syari'ah. Maqashid Al-Syariah merupakan lima tujuan pokok syariat Islam yang meliputi *hifz al-din*, *hifz al-nafs*, *hifz al-aql*, *hifz al-nasl*, dan *hifz al-mal* (Abi Ishaq Al-Syatibi, 1423/2003). Penelitian ini secara khusus berkaitan dengan aspek *hifz an-nafs*, yaitu menjaga keselamatan jiwa melalui upaya pencegahan dan penanganan penyakit secara lebih efektif dengan memanfaatkan teknologi machine learning sebagai alat bantu deteksi dini anemia (Yunos & Hamdan, 2024). Machine learning membantu menjaga keselamatan jiwa melalui deteksi dini, prediksi risiko, dan pengelolaan penyakit yang lebih akurat dan efisien. Sehingga dapat mendukung layanan kesehatan dalam meningkatkan kualitas diagnosis pengobatan. Hal ini sejalan dengan kaidah fiqh sebagai berikut:

الضرر يزال...

"Kemudharatan atau bahaya harus dihilangkan.." (Abu Bakr Al-Ahdali Al-Yamani, 2004. Faraidul Bahiyah hal.25).

Sehingga segala bentuk kemudharatan, termasuk pemanfaatan teknologi kesehatan modern dipandang selaras dengan nilai-nilai syariat (Irfan et al., 2025). Dengan demikian, penerapan machine learning dengan deteksi anemia tidak hanya menjadi inovasi ilmiah. Tetapi juga merupakan bentuk implementasi nilai Islam dalam menjaga kesehatan dan keselamatan secara lebih luas .

Seiring berkembangnya teknologi, pendekatan berbasis *machine learning* menawarkan solusi yang lebih cepat, efisien, dan konsisten untuk membantu proses deteksi anemia. Pada beberapa tahun terakhir, penelitian mengenai penerapan machine learning di bidang hematologi meningkat pesat. Beberapa studi menggunakan algoritma seperti *Logistic Regression*, *Random Forest*, *Support Vector Machine* (SVM) untuk mengklasifikasi anemia ataupun kondisi hematologis lainnya (Abdul-Jabbar et al., 2023; Gómez et al., 2025; Putri, 2025). Hasil penelitian-penelitian tersebut menunjukkan bahwa parameter hematologi dasar seperti hemoglobin, MCH, MCHC, MCV mampu menjadi dasar yang efektif dalam membangun model klasifikasi anemia karena merepresentasikan kondisi fisiologis sel darah merah secara kuantitatif (Amalia, 2025).

Berdasarkan penelitian Airlangga (2024) salah satu algoritma yang menunjukkan performa sangat kompetitif dalam berbagai penelitian tersebut adalah *Extreme Gradient Boosting* (XGBoost). XGBoost adalah salah satu algoritma yang dikenal memiliki kemampuan generalisasi yang tinggi, ketahanan terhadap overfitting, serta kinerja unggul dalam menangani data tabular numerik. Beberapa studi melaporkan bahwa XGBoost mampu mencapai akurasi sangat tinggi (hingga 100% menurut Schweta & Pande (2023) dan pada beberapa studi yang umumnya 94% menurut Darshan et al (2025)) dalam mendeteksi anemia dan subtipenya, baik pada data sederhana (CBC) maupun data kompleks. Namun demikian, terdapat kesenjangan yang belum banyak dieksplorasi dalam literatur yang ada.

Sebagian besar penelitian berfokus pada perbandingkan banyak algoritma sekaligus. Çakmak (2025) membandingkan Random Forest, XGBoost, LightGBM,

dan CatBoost pada data hematologi lengkap. Airlangga (2024) membandingkan enam algoritma sekaligus yaitu Decision Tree, Extra Tree, Random Forest, XGBoost, LightGBM, dan CatBoost. Sementara itu, Putri (2025) membandingkan Random Forest, SVM, Naive Bayes, dan XGBoost, dan Abdul-Jabbar et al. (2023) membandingkan LogitBoost, Random Forest, dan XGBoost. Penelitian-penelitian tersebut menggunakan dataset dengan fitur klinis yang lebih banyak dan parameter biologis lanjutan lainnya. Kondisi tersebut berbeda dengan situasi nyata di fasilitas layanan kesehatan primer yang umumnya hanya memiliki akses terhadap parameter terbatas, yaitu hemoglobin, MCH, MCHC, MCV (Darshan et al., 2025). Belum banyak penelitian yang secara khusus mengevaluasi performa XGBoost secara tunggal pada dataset hematologi minimalis seperti ini, sehingga belum ada gambaran yang jelas mengenai sejauh mana XGBoost mampu mempertahankan fitur-fitur dasar tersebut. Selain itu pengaruh variasi hyperparameter terhadap performa XGBoost pada dataset hematologi minimalis juga belum dieksplorasi secara mendalam.

Berdasarkan kesenjangan tersebut, penelitian ini difokuskan untuk mengevaluasi kinerja algoritma XGBoost dalam klasifikasi kondisi anemia menggunakan dataset hematologi yang hanya terdiri dari parameter hemoglobin, MCV, MCH, dan MCHC dengan label biner (0 = tidak anemia, 1 = anemia). Melalui penerapan hyperparameter tuning menggunakan Grid Search pada parameter n_estimators, penelitian ini diharapkan dapat memberikan gambaran empiris mengenai performa XGBoost pada kondisi daraser sederhana yang merepresentasikan keterbatasan data di fasilitas layanan kesehatan primer.

Sehingga dapat menjadi dasar pengembangan sistem deteksi anemia yang lebih efisien, presisi, dan adaptif.

1.2 Rumusan Masalah

"Bagaimana tingkat performa model klasifikasi anemia menggunakan algoritma XGBoost berdasarkan data hematologi?".

1.3 Batasan Masalah

1. Penelitian ini menggunakan dataset hematologi yang terdiri dari atribut Gender, Hemoglobin, MCV, MCH, dan MCHC. Total data yang digunakan berjumlah 1.421 baris data.
2. Label yang digunakan adalah nilai biner (0 = tidak anemia, 1 = anemia).
3. Algoritma yang digunakan adalah XGBoost dengan algoritma pembanding Logistic Regression dan Random Forest.
4. Penelitian ini menerapkan proses *hyperparameter tuning* pada XGBoost terhadap parameter n_estimators.
5. Evaluasi kinerja hanya mencakup metrik klasifikasi; *accuracy*, *precision*, *recall*, dan *F1-score.*

1.4 Tujuan Penelitian

Tujuan utama penelitian ini adalah mengetahui tingkat performa model klasifikasi anemia menggunakan algoritma XGBoost berdasarkan data hematologi.

1.5 Manfaat Penelitian

1. Memberikan informasi empiris mengenai tingkat akurasi algoritma XGBoost dalam klasifikasi anemia berdasarkan data hematologi.
2. Menjadi referensi bagi penelitian selanjutnya yang ingin mengembangkan atau membandingkan model machine learning untuk diagnosis anemia.
3. Memberikan kontribusi dalam pemanfaatan parameter hematologi dasar (Hemoglobin, MCV, MCH, MCHC) sebagai fitur klasifikasi.
4. Memberikan gambaran performa algoritma XGBoost yang dapat digunakan sebagai dasar untuk pengembangan sistem klasifikasi anemia berbasis data di fasilitas kesehatan.
5. Menjadi acuan bagi praktisi teknologi kesehatan dalam memahami potensi dan keterbatasan penggunaan algoritma XGBoost untuk klasifikasi kondisi anemia.

BAB II
STUDI PUSTAKA

2.1 Penelitian Terdahulu

Penelitian mengenai deteksi dan klasifikasi anemia berbasis data medis telah banyak dilakukan dengan memanfaatkan pendekatan pembelajaran mesin (*machine learning*). Berbagai algoritma klasifikasi, seperti Decision Tree, Random Forest, Support Vector Machine, dan XGBoost, telah diterapkan untuk menganalisis data hematologi guna mengidentifikasi kondisi anemia secara akurat. Sejumlah studi menunjukkan bahwa parameter hematologi, seperti kadar hemoglobin, MCV, MCH, dan MCHC, dapat digunakan sebagai sumber data yang efektif dalam membangun model klasifikasi anemia, karena mampu merepresentasikan kondisi fisiologis sel darah merah secara kuantitatif.

Penelitian yang dilakukan oleh Çakmak (2025) dengan judul *Machine Learning Approaches for Enhanced Diagnosis of Hematological Disorders* bertujuan untuk meningkatkan akurasi diagnosis gangguan hematologi, termasuk anemia, dengan memanfaatkan algoritma machine learning berbasis data darah lengkap (Complete Blood Count). Dataset yang digunakan terdiri dari beberapa parameter hematologi utama seperti hemoglobin, jumlah eritrosit (RBC), MCV, MCH, dan trombosit. Penelitian ini mengevaluasi beberapa algoritma, yaitu Random Forest, XGBoost, LightGBM, dan CatBoost. Hasil pengujian menunjukkan bahwa algoritma berbasis gradient boosting memberikan performa terbaik, di mana LightGBM mencapai akurasi sebesar 98,38%, CatBoost 98,37%, dan XGBoost 97,72%.Hasil ini menunjukkan bahwa algoritma XGBoost memiliki

kemampuan yang sangat baik dalam menangani data tabular medis dan mampu memodelkan hubungan non-linear antar fitur hematologi secara efektif.

Penelitian oleh Amalia dan Rumini (2025) yang berjudul *Anemia Classification with Clinical Feature Engineering and SHAP Interpretation* bertujuan untuk mengklasifikasikan kondisi anemia menggunakan data hematologi berbasis machine learning. Dataset yang digunakan merupakan dataset publik Kaggle yang berisi lebih dari 5.000 data pasien dengan parameter hemoglobin, MCV, MCH, dan MCHC. Penelitian ini membandingkan algoritma XGBoost, Random Forest, dan Logistic Regression. Hasil evaluasi menunjukkan bahwa algoritma XGBoost menghasilkan performa terbaik dengan akurasi sebesar 97,8%, precision 97,5%, recall 98,1%, dan nilai AUC sebesar 0,98. Analisis fitur menggunakan SHAP menunjukkan bahwa kadar hemoglobin dan MCV merupakan fitur paling dominan dalam menentukan klasifikasi anemia.

Penelitian yang dilakukan oleh Putri et al. (2023) dengan judul *A Comparative Analysis of Machine Learning Classifier of Anemia Diagnosis Based on Complete Blood Count (CBC) Data* bertujuan untuk mengembangkan model klasifikasi anemia menggunakan pendekatan machine learning berbasis data hematologi. Dataset yang digunakan berasal dari data pemeriksaan darah yang terdiri dari beberapa parameter utama seperti Hb, MCV, MCH, MCHC yang umum digunakan dalam diagnosis anemia. Penelitian ini membandingkan beberapa algoritma klasifikasi, yaitu Support Vector Machine (SVM), Random Forest, dan XGBoost untuk mengidentifikasi model yang paling optimal dalam mendeteksi kondisi anemia. Hasil pengujian menunjukkan bahwa algoritma XGBoost

memberikan performa klasifikasi terbaik dengan tingkat akurasi sebesar 96,4%, lebih tinggi dibandingkan metode lainnya. Temuan ini menunjukkan bahwa algoritma berbasis gradient boosting mampu memodelkan hubungan kompleks antar parameter hematologi secara efektif serta memiliki potensi yang baik dalam mendukung sistem deteksi anemia berbasis data medis.

Penelitian yang dilakukan oleh Abdul Jabbar et al. (2024) yang berjudul *A Comparative Study of Anemia Classification Algorithms for International and Newly CBC Datasets* bertujuan untuk mengidentifikasi kondisi anemia dengan memanfaatkan metode machine learning berdasarkan parameter hematologi. Data yang digunakan dalam penelitian ini berasal dari pemeriksaan *Complete Blood Count* (CBC) yang meliputi sejumlah indikator penting, seperti kadar hemoglobin, jumlah sel darah merah (*Red Blood Cell* atau RBC), *Mean Corpuscular Volume*(MCV), dan *Mean Corpuscular Hemoglobin* (MCH). Beberapa algoritma klasifikasi diterapkan dan dibandingkan dalam penelitian tersebut, yaitu Decision Tree, Random Forest, dan XGBoost. Berdasarkan hasil pengujian, algoritma XGBoost memberikan kinerja terbaik dengan tingkat akurasi mencapai 97,1% serta menunjukkan kemampuan yang konsisten dalam mengklasifikasikan data ke dalam kategori anemia maupun non-anemia. Temuan ini mengindikasikan bahwa metode *ensemble learning* yang menggunakan pendekatan *gradient boosting* efektif dalam mengolah data medis berbentuk tabular dan memiliki potensi untuk mendukung proses analisis klinis secara lebih optimal dan efisien.

Penelitian oleh Airlangga (2024) mengevaluasi beberapa algoritma machine learning untuk diagnosis anemia menggunakan data *Complete Blood Count* (CBC)

dengan jumlah lebih dari 6.000 data pasien. Algoritma yang diuji meliputi Decision Tree, Random Forest, XGBoost, LightGBM, dan CatBoost. Hasil eksperimen menunjukkan bahwa XGBoost memberikan performa terbaik dengan akurasi sebesar 98,1% dan nilai AUC 0,99, mengungguli Random Forest (96,4%) dan Decision Tree (90,8%). Penelitian ini menegaskan keunggulan XGBoost dalam mengolah data tabular medis.

Penelitian oleh Hidayah et al. (2023) berfokus pada penanganan data tidak seimbang dalam klasifikasi anemia menggunakan SMOTE. Dataset yang digunakan memiliki rasio kelas minoritas hanya sekitar 30%. Setelah penerapan SMOTE, performa model XGBoost meningkat signifikan, dengan nilai recall kelas anemia naik dari 81,4% menjadi 92,6%, serta F1-score meningkat dari 85,2% menjadi 93,1%. Hasil ini menunjukkan bahwa teknik penyeimbangan data sangat berpengaruh terhadap kinerja model klasifikasi anemia.

Penelitian yang dipublikasikan pada jurnal MDPI oleh Gomez et al. (2025) dengan judul *Anemia Classification System Using Machine Learning* berfokus pada pengembangan sistem klasifikasi anemia menggunakan pendekatan machine learning yang diterapkan pada data hematologi dan citra darah. Dataset yang digunakan berasal dari sumber publik dan data klinis dengan parameter utama seperti hemoglobin, hematokrit, MCV, dan MCH. Beberapa algoritma diuji dalam penelitian ini, termasuk Support Vector Machine, Random Forest, dan XGBoost. Hasil evaluasi menunjukkan bahwa algoritma XGBoost memberikan performa terbaik dengan tingkat akurasi mencapai 96,95%, nilai AUC lebih dari 0,97, serta tingkat recall yang tinggi pada kelas anemia. Penelitian ini menegaskan bahwa

XGBoost tidak hanya unggul dari segi akurasi, tetapi juga mampu memberikan interpretasi fitur yang baik, sehingga cocok digunakan dalam sistem pendukung keputusan medis.

Tabel 2.1 Penelitian Terdahulu

No	Nama Peneliti	Metode	Dataset/Fitur	Fokus Penelitian	Keterbatasan
1	Çakmak, 2025	RF, XGBoost, LightGBM, CatBoost	Hematologi dataset	Perbandingan beberapa algoritma ML	Tidak fokus pada analisis performa tunggal XGBoost.
2	Ikhlasul Amalia & Rumini, 2025	XGBoost, RF, Logistic Regression	Hematologi dataset	Interpretabilitas model	Tidak mengevaluasi performa klasifikasi sederhana
3	Abdul-Jabbar et al., 2023	LogitBoost, RF, XGBoost	CBC dataset	Klasifikasi anemia	Perbandingan algoritma
4	Putri, 2025	RF, SVM, NB, XGBoost	CBC dataset	Perbandingan algoritma	Tidak mengevaluasi parameterisasi XGBoost
5	Airlangga, 2024	Decision Tree, ExtraTree, RF, XGBoost, LightGBM, CatBoost	CBC dataset	Perbandingan model	Tidak melakukan eksplorasi fitur hematologi paling berpengaruh.
6	Hidayah et al., 2025	Random Forest, XGBoost + SMOTE	CBC dataset	Data balancing dan aplikasi Streamlit	Tidak mengevaluasi performa klasifikasi sederhana
7	Gómez et al., 2025	XGBoost, Random Forest, SVM	CBC dataset	Kajian khusus mengenai karakteristik data CBC dan perbandingan ML	Analisis sensitivitas fitur hematologi masih terbatas.
8	Penelitian ini	XGBoost + *Hyperparameter Tuning Grid Search*	Anemia dataset (Hb, MCV, MCH, MCHC)	Evaluasi performa XGBoost pada dataset hematologi minimalis	-

Berdasarkan Tabel 2.1, dapat dilihat bahwa sebagian besar penelitian sebelumnya berfokus pada perbandingan beberapa algoritma machine learning atau menggunakan dataset dengan fitur klinis yang lebih kompleks. Beberapa penelitian

juga menitikberatkan pada pengembangan sistem atau interpretabilitas model. Namun, masih terbatas penelitian yang secara khusus mengevaluasi performa algoritma XGBoost secara tunggal pada dataset hematologi yang sederhana dengan jumlah fitur minimal. Oleh karena itu, penelitian ini berfokus pada analisis kinerja algoritma XGBoost dalam mengklasifikasikan kondisi anemia menggunakan parameter hematologi dasar, yaitu hemoglobin, MCV, MCH, dan MCHC, melalui penerapan hyperparameter tuning untuk memperoleh konfigurasi model yang optimal.

2.2 Anemia

Anemia adalah kondisi medis yang ditandai oleh penurunan kadar hemoglobin (Hb), hematokrit, atau jumlah sel darah merah di bawah batas normal, yang berakibat pada berkurangnya kemampuan darah untuk mengantarkanoksigen ke seluruh jaringan tubuh. Kondisi ini umumnya menimbulkan sejumlah keluhan klinis, antara lain rasa lelah berlebihan, pusing, sesak napas, sertamenurunnya kemampuan konsentrasi dan produktivitas penderita. Secara diagnostik, kriteria yang paling lazim digunakan secara internasional, termasuk oleh World Health Organization (WHO), menetapkan bahwa seseorang dikategorikan anemia apabila kadar hemoglobinnya berada di bawah 13 g/dL pada laki-laki dewasa dan di bawah 12 g/dL pada perempuan dewasa(Chopra & Anker, 2020; Hain et al., 2023). Ambang nilai tersebut turut diadopsi dalam berbagai pedoman klinis maupun survei kesehatan masyarakat di tingkat global. Penetapannya didasarkan pada distribusi normal kadar Hb dalam populasi yang sehat. Kendati demikian, sejumlah penelitian

terkini mengindikasikan adanya 14 variasi nilai ambang tersebut yang dipengaruhi oleh faktor usia, jenis kelamin, latar belakang etnis, serta kondisi fisiologis individu.

Secara umum, anemia dapat diklasifikasikan menjadi beberapa tipe berdasarkan penyebabnya, yaitu anemia defisiensi zat besi, anemia akibat kehilangan darah, anemia akibat gangguan produksi eritrosit dan anemia hemolitik. Kekurangan zat besi menyebabkan gangguan sintesis hemoglobin yang berperan penting dalam transportasi oksigen. Diantara tipe tersebut, anemia defisiensi besi merupakan yang paling umum, terutama pada kelompok rentan seperti wanita usia subur dan anak-anak(Safiri et al., 2021). Disebabkan oleh asupan zat besi yang tidak cukup, gangguan penyerapan, atau kehilangan darah kronis. Anemia akibat kehilangan darah umumnya terjadi pada pendarahan akut, misal akibat menstruasi berat atau pendarahan saluran cerna. Anemia akibat gangguan produksi eritrosit termasuk anemia aplastik, diakibatkan adanya penyakit kronis, dan gangguan sumsum tulang. Sedangkan anemia hemolotik disebabkan oleh penghancuran eritrosit yang berlebihan, baik karena faktor imun maupun non-imun (Dawidowski & Pietrzak, 2022).

Diagnosis anemia umumnya dilakukan melalui pemeriksaan hematologi lengkap (Complete Blood Count/CBC) yang meliputi pengukuran hemoglobin, hematokrit, Mean Corpuscular Volume (MCV), Mean Corpuscular Hemoglobin (MCH), dan Mean Corpuscular Hemoglobin Concentration (MCHC). MCV membantu mengklasifikasi anemia menjadi mikro-, normo-, atau makrositik, yang penting untuk menentukan penyebabnya. MCH dan MCHC memberikan informasi tentang kandungan hemoglobin dalam eritrosit, membantu membedakan anemia

defisiensi besi dari tipe lain(Gómez-Pastora et al., 2022). Parameter-parameter tersebut membantu menentukan jenis anemia dan tingkat keparahannya.

2.3 Data Hematologi

Data Hematologi merupakan data medis yang diperoleh dari hasil pemeriksaan laboratorium darah yang bertujuan untuk mengevaluasi kondisi dan fungsi komponen darah dalam tubuh manusia. Pemeriksaan hematologi, terutama Complete Blood Count(CBC) digunakan secara luas dalam dunia medis untuk membantu proses diagnosis, pemantauan penyakit, serta evaluasi kondisi kesehatan seseorang(Celkan, 2020; Seo & Lee, 2022). Data ini bersifat kuantitatif dan terdiri dari berbagai parameter yang mencerminkan karakteristik sel darah merah, sel darah putih, dan trombosit.

Parameter utama dalam data hematologi meliputi banyak hal. Seperti, sel darah merah (RBC), hemoglobin (Hb), dan hematokrit (Hct) menilai kapasitas pengangkutan oksigen dan mendeteksi anemia. Indeks eritrosit (MCV, MCH, MCHC, RDW) membantu klasifikasi anemia dan gangguan morfologi eritrosit. Sel darah putih (WBC) dan diferensial mengidentifikasi infeksi, inflamasi, dan gangguan imun. Trombosit (PLT) dan MPV menilai resiko pendarahan atau trombosis. Serta, rasio seperti NLR (neutrophil-to-lymphocyte), PLR (platelet-to-lymphocyte ratio) yang digunakan sebagai penanda prognosis pada berbagai penyakit, termasuk infeksi, kanker, dan penyakit metabolik(Celkan, 2020).

Dalam konteks penelitian deteksi anemia, data hematologi memiliki peranan yang sangat penting karena anemia secara langsung berkaitan dengan kondisi sel darah merah dan kadar hemoglobin dalam darah. Oleh karena itu

parameter hematologi yang berkaitan dengan eritrosit menjadi fokus utama dalam penelitian ini (Miglio et al., 2023). Parameter-parameter tersebut mencerminkan jumlah, ukuran, serta kandungan hemoglobin dalam sel darah merah yang digunakan sebagai dasar dalam menentukan kondisi anemia atau non-anemia.

2.4 Machine Learning

Machine learning merupakan salah satu cabang kecerdasan buatan (artificial intelligence) yang menitikberatkan pada pengembangan sistem komputasional yang memiliki kemampuan untuk belajar dari data dan meningkatkan performanya secara mandiri tanpa memerlukan pemrograman secara eksplisit. Melalui pendekatan ini, sistem dapat mengidentifikasi pola tersembunyi, menghasilkan prediksi, serta menformulasikan keputusan berdasarkan data historis yang tersedia. Pendekatan ini dinilai sangat efektif dan telah diadopsi secara luas di berbagai bidang karena kapasitasnya dalam mengelola data bervolume besar dengan tingkat kompleksitas tinggi (Aldahiri et al., 2021; Jhaveri et al., 2022).

Ditinjau dari metode pembelajarannya, machine learning secara umum dikategorikan ke dalam beberapa tipe, di antaranya supervised learning (pembelajaran terawasi) dan unsupervised learning (pembelajaran tidak terawasi). Saxena & Chandra (2021) dalam karyanya yang berjudul "Artificial Intelligence and Machine Learning in Healthcare" menjelaskan bahwa supervised learning adalah metode pembelajaran yang bertumpu pada data berlabel, di mana setiap data masukan telah disertai dengan nilai keluaran yang sudah diketahui sebelumnya. Tujuan utama metode ini adalah mempelajari relasi antara data masukan dan label keluarannya guna membangun model yang mampu mengeneralisasi prediksi

terhadap data baru. Di sisi lain, unsupervised learning merupakan metode pembelajaran yang bekerja tanpa label, dengan tujuan menemukan pola atau struktur laten dalam data, seperti melalui pengelompokan (clustering) maupun reduksi dimensi.

Salah satu permasalahan utama yang umum diselesaikan melalui supervised learning adalah klasifikasi. Klasifikasi merupakan proses pengelompokan data ke dalam kelas atau kategori tertentu berdasarkan karakteristik yang terkandung di dalamnya (Saxena & Chandra, 2021). Dalam kerangka machine learning, model klasifikasi dibangun dengan tujuan memetakan data masukan ke dalam kelas-kelas yang telah ditetapkan sebelumnya. Pada penelitian ini, permasalahan klasifikasi yang dihadapi adalah menentukan apakah seorang individu tergolong ke dalam kategori anemia atau non-anemia berdasarkan data hematologi, sehingga permasalahan ini termasuk dalam klasifikasi biner.

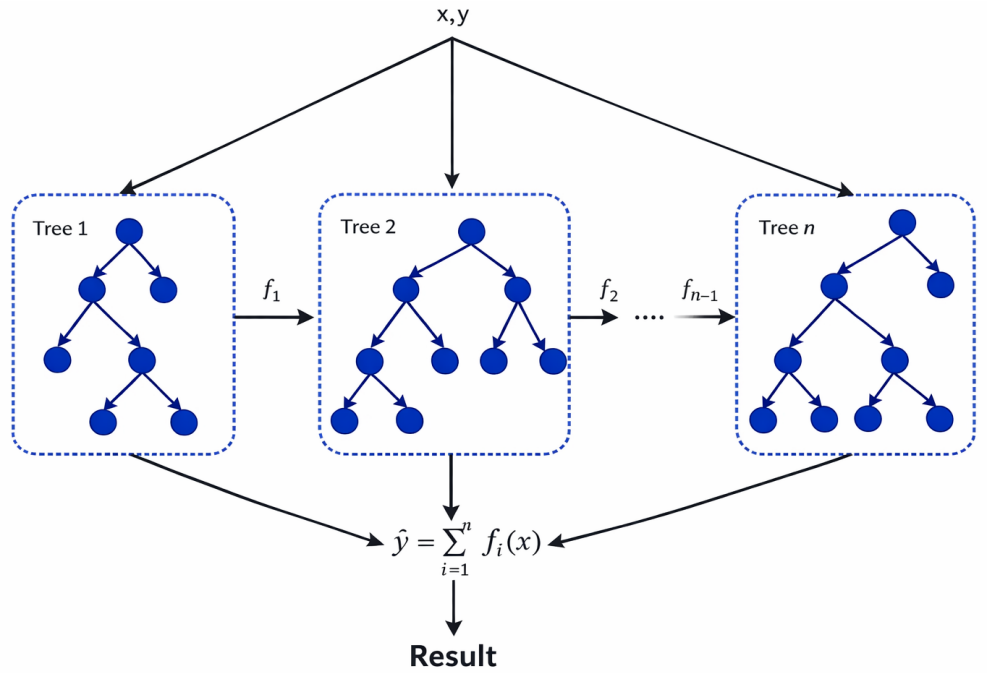
Penerapan machine learning di bidang kesehatan telah mengalami perkembangan yang pesat dan memberikan kontribusi yang substansial, terutama dalam mendukung proses diagnosis medis dan prediksi penyakit. Machine learning dimanfaatkan untuk menganalisis data klinis, data laboratorium, maupun data rekam medis guna mendukung pengambilan keputusan medis yang lebih efisien dan presisi. Beberapa contoh aplikasinya di ranah kesehatan mencakup deteksi penyakit berbasis citra medis, prediksi risiko penyakit kronis, analisis hasil pemeriksaan laboratorium, serta pengembangan sistem pendukung keputusan klinis (Bhatt et al., 2023; Botlagunta et al., 2023). Dalam penelitian ini, machine learning didayagunakan untuk menganalisis data hematologi sebagai landasan dalam

mendeteksi anemia. Dengan memanfaatkan kemampuan machine learning dalam mempelajari pola keterkaitan antar parameter darah, model yang dikembangkan diharapkan mampu menghasilkan klasifikasi yang akurat dan dapat difungsikan sebagai instrumen bantu dalam proses deteksi dini anemia.

2.5 Algoritma XGBoost

XGBoost (*Extreme Gradient Boosting*) adalah algoritma machine learning berbasis *ensemble learning* yang mengimplementasikan teknik gradient boosting dengan decision tree sebagai model dasarnya. Algoritma ini dirancang untuk meningkatkan kecepatan, akurasi, serta efisiensi komputasi dibandingkan dengan metode boosting konvensional (Mienye & Sun, 2022). XGBoost telah banyak digunakan dalam berbagai kompetisi maupun penelitian ilmiah berkat kemampuannya menghasilkan performa tinggi, terutama pada permasalahan klasifikasi dan regresi dengan data bertipe tabular. Konsep inti yang mendasari XGBoost adalah *boosting*, yakni metode ensemble yang mengombinasikan sejumlah model pembelajar lemah (*weak learners*) untuk membentuk satu model yang lebih kuat secara prediktif. Dalam teknik ini, model dibangun secara sekuensial, di mana setiap model baru diarahkan untuk memperbaiki kesalahan yang dihasilkan oleh model sebelumnya. Melalui mekanisme tersebut, kesalahan prediksi dapat direduksi secara bertahap hingga diperoleh model dengan kinerja yang optimal (Sahin, 2020). Sebagaimana Gambar 2.1 menunjukkan arsitektur XGBoost.

.



Gambar 2.1 Arsitektur XGBoost
Sumber : https://www.researchgate.net/figure/Model-architecture-for-XGBoost-algorithm_fig5_377596115

Secara umum, mekanisme kerja XGBoost diawali dengan pembangunan model awal menggunakan decision tree sederhana, yang kemudian secara iteratif dilengkapi dengan penambahan pohon-pohon baru untuk mengoreksi kesalahan prediksi model sebelumnya. Pada setiap iterasi, model mengevaluasi nilai kesalahan prediksi melalui fungsi kerugian (*loss function*). Berdasarkan nilai tersebut, XGBoost membangun decision tree berikutnya dengan tujuan memperbaiki kesalahan sebelumnya menggunakan pendekatan *gradient descent*, yaitu setiap pohon baru diorientasikan ke arah negatif gradien dari fungsi loss (Wiens et al., 2025). Proses ini berlangsung hingga jumlah pohon yang telah ditentukan tercapai atau hingga nilai kesalahan mencapai titik minimum. Di samping itu, XGBoost mengintegrasikan teknik regularisasi guna mengendalikan

kompleksitas model sehingga risiko overfitting dapat ditekan (Noorunnahar et al., 2023).

Library XGBoost yang digunakan dalam penelitian ini merupakan salah satu implementasi populer dari algoritma Gradient Boosting yang dikembangkan secara open-source oleh Tianqi Chen dkk. pada tahun 2016 (Shwartz-Ziv & Armon, 2021). Library ini ditulis dalam bahasa pemrograman C++ dengan antarmuka (binding) ke berbagai bahasa pemrograman, termasuk Python melalui modul xgboost, sehingga mampu melakukan komputasi paralel dan efisien untuk data dalam jumlah besar. Dalam penelitian ini, fungsi utama yang digunakan adalah XGBClassifier() dari modul xgboost.sklearn, yang menyediakan antarmuka serupa dengan library scikit-learn sehingga memudahkan dalam proses pelatihan (fit), prediksi (predict), serta evaluasi dan penyetelan hiperparameter model klasifikasi (Vu et al., 2021).

Dalam penerapannya, XGBoost memiliki sejumlah hiperparameter yang berpengaruh terhadap performa model. Beberapa hiperparameter utama yang lazim digunakan antara lain n_estimators yang merepresentasikan jumlah decision tree yang dibangun, max_depth yang menentukan kedalaman maksimum setiap pohon, serta learning_rate yang mengatur laju pembelajaran model. Selain itu, terdapat hiperparameter tambahan seperti subsample dan colsample_bytree yang menentukan proporsi data dan fitur yang diikutsertakan dalam pembangunan setiap pohon, serta parameter regularisasi lambda dan alpha yang berfungsi mengendalikan kompleksitas model (Bakasa & Viriri, 2023; Sahin, 2020).

XGBoost memiliki sejumlah keunggulan yang menjadikannya sangat relevan untuk diterapkan pada data tabular medis, termasuk data hematologi. Algoritma ini mampu menangkap hubungan non-linear yang kompleks antar fitur serta bekerja secara optimal pada data numerik. Selain itu, XGBoost dilengkapi dengan mekanisme penanganan nilai hilang (missing value) serta fitur regularisasi yang berkontribusi pada peningkatan generalisasi model terhadap data medis yang cenderung memiliki variasi tinggi. Berbagai penelitian terkini menyebutkan bahwa efisiensi komputasi dan kemampuan XGBoost dalam menghasilkan akurasi tinggi menjadikannya pilihan yang tepat dalam penelitian ini untuk membangun model klasifikasi anemia berbasis data hematologi (Yıldız & Kalayci, 2025).

2.6 Hyperparameter Tuning Grid Search

Hyperparameter tuning adalah proses optimasi parameter eksternal pada model machine learning yang tidak diperoleh secara langsung melalui data pelatihan, melainkan ditetapkan sebelum proses training berlangsung. Berbeda dengan parameter internal model yang dihasilkan melalui proses pembelajaran (seperti bobot pada setiap node), hyperparameter berperan dalam menentukan struktur dan mekanisme pembelajaran model secara keseluruhan. Pada algoritma XGBoost, performa model sangat bergantung pada konfigurasi hyperparameter seperti n_estimators, max_depth, learning_rate, dan parameter lainnya. Penetapan nilai hyperparameter yang tidak tepat berpotensi menyebabkan model jatuh ke kondisi underfitting, yakni model yang terlalu sederhana, atau sebaliknya overfitting, yakni model yang terlalu kompleks terhadap data pelatihan, sehingga diperlukan proses penyesuaian yang cermat untuk mencapai performa yang

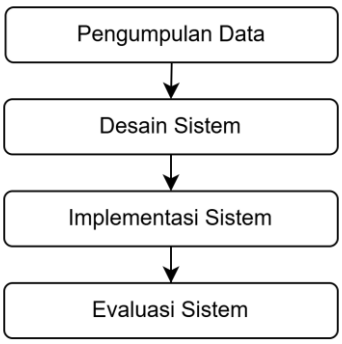
optimal. Oleh karena itu, penerapan hyperparameter tuning pada XGBoost secara konsisten terbukti mampu meningkatkan akurasi sekaligus meminimalkan risiko overfitting maupun underfitting, terlebih apabila dilakukan melalui metode pencarian yang sistematis atau berbasis pendekatan optimasi (Omotehinwa & Oyewola, 2023).

Salah satu metode yang umum digunakan dalam proses tuning adalah Grid Search. Grid Search bekerja dengan menguji seluruh kombinasi nilai hyperparameter yang telah ditentukan sebelumnya dalam sebuah ruang parameter (*parameter grid*). Setiap kombinasi diuji menggunakan metode validasi tertentu, kemudian dibandingkan berdasarkan metrik evaluasi yang telah ditentukan untuk memperoleh konfigurasi terbaik. Proses kerja Grid Search dimulai dengan mendefinisikan rentang atau daftar nilai dari setiap hyperparameter yang akan diuji. Selanjutnya, sistem akan melatih model untuk setiap kombinasi tersebut dan menghitung nilai evaluasi. Kombinasi dengan skor evaluasi terbaik akan dipilih sebagai konfigurasi optimal. Meskipun metode ini bersifat komputasional lebih tinggi dibandingkan metode acak, Grid Search memiliki keunggulan dalam hal konsistensi, kemudahan interpretasi, serta reproduktibilitas hasil penelitian (A Ilemobayo et al., 2024).

BAB III
DESAIN DAN IMPLEMENTASI

3.1 Desain Penelitian

Desain penelitian merupakan rancangan atau kerangka kerja yang digunakan sebagai landasan untuk mencapai tujuan penelitian secara sistematis. Kerangka ini berfungsi sebagai panduan agar seluruh proses penelitian dapat berjalan secara terarah, efisien, dan sesuai dengan sasaran yang telah ditetapkan sebelumnya. Pada penelitian ini, proses pelaksanaannya ditempuh melalui sejumlah tahapan yang saling berkesinambungan, mulai dari tahap awal hingga tahap akhir penelitian. Secara umum, alur penelitian ini mencakup beberapa tahapan utama, yaitu studi literatur, pengumpulan data, desain sistem, preprocessing data, implementasi atau pembangunan model menggunakan algoritma XGBoost, pengujian, evaluasi, serta penarikan kesimpulan. Setiap tahapan memegang peran yang signifikan dalam memastikan hasil penelitian yang diperoleh bersifat valid dan dapat dipertanggungjawabkan secara ilmiah. Berikut adalah desain penelitian yang digunakan dalam penelitian ini.



Gambar 3.1 Desain Penelitian

Tahap pertama adalah pengumpulan data, yaitu proses memperoleh data yang akan digunakan sebagai dasar analisis dan pengujian sistem. Selanjutnya, dilakukan desain sistem, yang mencakup perencanaan arsitektur dan alur proses sistem sesuai dengan kebutuhan penelitian. Tahap berikutnya adalah preprocessing data, yaitu pembersihan dan pengolahan awal data agar siap untuk dianalisis. Setelah itu, dilakukan implementasi dan pengujian sistem, dimana sistem dijalankan menggunakan metode klasifikasi yang telah ditentukan untuk menguji performa dan akurasi hasilnya.

3.2 Pengumpulan Data

Penelitian ini memanfaatkan data sekunder, yaitu data yang tidak dikumpulkan secara langsung oleh peneliti, melainkan diperoleh dari sumber yang telah menjalani proses pengumpulan sebelumnya. Dataset yang digunakan dalam penelitian ini bersumber dari platform Kaggle dengan fokus pada data hematologi yang dimanfaatkan untuk keperluan deteksi anemia. Dataset tersebut memuat sejumlah parameter darah yang lazim digunakan dalam diagnosis klinis, antara lain hemoglobin (Hb), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), mean corpuscular hemoglobin concentration (MCHC), serta informasi jenis kelamin. Selain itu, dataset ini telah dilengkapi dengan label diagnosis dalam format biner, yaitu 0 untuk kategori "tidak anemia" dan 1 untuk kategori "anemia", sehingga memungkinkan penerapan algoritma klasifikasi seperti XGBoost secara langsung tanpa memerlukan proses pelabelan ulang. Dataset ini tersedia dalam format file CSV (Comma Separated Values) yang dapat diakses secara publik dan

diunduh melalui platform Kaggle. Informasi mengenai fitur-fitur dataset beserta penjelasannya dapat dilihat pada Tabel 3.1 Fitur Dataset.

Tabel 3.1 Fitur Dataset

No	Nama Fitur	Deskripsi	Satuan
1	Gender	Jenis Kelamin	-
2	Hemoglobin	Protein dalam sel darah merah yang bertugas mengikat dan mengangkut oksigen, serta indikator utama untuk menentukan anemia.	gram per desiliter (g/dL)
3	Mean Corpuscular Hemoglobin	Rata-rata jumlah hemoglobin dalam satu sel darah merah (erythrocyte).	picogram (pg)
4	Mean Corpuscular Hemoglobin Concentration	Mengukur konsentrasi hemoglobin per volume sel darah merah	gram per desiliter (g/dL)
5	Mean Corpuscular Volume	Rata-rata ukuran atau volume sel darah merah	femtoliter (fL)
6	Result	Label diagnosis	-

Dataset diambil dari sumber publik yang kredibel dan telah melalui proses kurasi sebelumnya, sehingga data siap untuk digunakan pada tahap pra proses dan pemodelan. Seluruh data dihasilkan dari pemeriksaan hematologi dan direpresentasikan dalam format terstruktur yang memudahkan proses analisis. Dalam penelitian ini, data tersebut akan diolah lebih lanjut melalui tahapan preprocessing, serta pemisahan data menjadi data latih dan data uji untuk mengukur performa model secara objektif. Total data yang digunakan berjumlah 1.421 baris data, yang kemudian akan dipetakan menjadi dua kategori besar sesuai tujuan penelitian. Berikut merupakan sample data dari dataset yang pada Tabel 3.2 Sample Dataset.

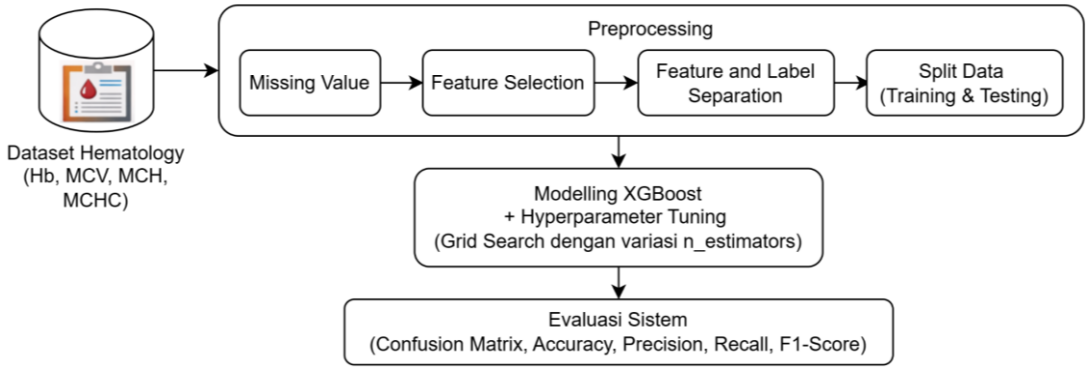
Tabel 3.2 Sample Dataset

Gender	Hb	MCH	MCHC	MCV	Result
1	14.9	22.7	29.1	83.7	0
0	15.9	25.4	28.3	72	0
…	…	…	…	…	…

Gender	Hb	MCH	MCHC	MCV	Result
0	7.5	28.8	30.8	86.4	1
1	16.7	20.1	31.4	73	0
…	…	…	…	…	…
0	14.3	16.2	29.5	95.2	0
0	11.8	21.2	28.4	98.1	1

3.3 Desain Sistem

Dalam penelitian ini peneliti akan menggunakan *Google Colab* dan bahasa pemrograman Python dalam membangun sistem. Berikut Gambar 3.2 yang menunjukkan desain sistem secara umum.



Gambar 3.2 Desain Sistem

Secara umum, sistem deteksi anemia menggunakan algoritma XGBoost pada penelitian ini dibagi menjadi beberapa komponen utama. Tahap pertama adalah input data, yaitu, proses pengambilan data hematologi yang telah tersedia dalam bentuk dataset terstruktur. Data ini mencakup beberapa parameter pemeriksaan darah seperti hemoglobin, MCV, MCH, dan MCHC yang menjadi dasar dalam penentuan anemia. Tahap kedua adalah preprocessing data, yaitu proses pemberian dan penyiapan data agar siap digunakan dalam pemodelan. Proses ini meliputi pemeriksaan nilai kosong (*missing value*), seleksi fitur, dan pemisahan

fitur dan label. Tahap ini penting agar model dapat bekerja lebih optimal dan menghindari bias akibat kualitas data yang buruk.

Setelah data bersih, dilakukan tahap pembagian dataset menjadi data latih dan data uji. Selanjutnya adalah tahap klasifikasi menggunakan XGBoost, dimana data yang telah dipreproses menjadi input bagi model untuk memprediksi apakah seseorang masuk kategori “anemia” atau “tidak anemia” berdasarkan parameter hematologinya. XGBoost dipilih karena kemampuannya menangani data tabular, kestabilan performa, dan pemrosesan fitur non-linear yang baik.

3.4 Exploratory Data Analysis (EDA)

Exploratory Data Analysis(EDA) dilakukan untuk memahami karakteristik data hematologi yang digunakan dalam penelitian serta memastikan kualitas data sebelum diproses lebih lanjut menggunakan algoritma XGBoost. Dataset yang digunakan terdiri dari 1.421 data pasien dengan 6 atribut, yaitu Gender, Hemoglobin, MCH, MCHC, MCV dan Result sebagai label kelas. Seluruh atribut numerik merepresentasikan parameter hematologi yang umum digunakan dalam pemeriksaan darah rutin untuk mendeteksi anemia. Berdasarkan hasil pemeriksaan awal, dataset tidak memiliki nilai kosong (missing value) pada seluruh fitur, sehingga dapat langsung digunakan tanpa proses imputasi. Seluruh fitur numerik memiliki rentang nilai yang wajar dan sesuai dengan standar klinis pemeriksaan darah, yang menunjukkan bahwa dataset memiliki kualitas baik dan layak untuk analisis lebih lanjut menggunakan machine learning.

Distribusi label kelas menunjukkan bahwa data terbagi menjadi dua kategori yaitu tidak anemia (0) dan anemia (1). Analisis statik deskriptif dilakukan untuk

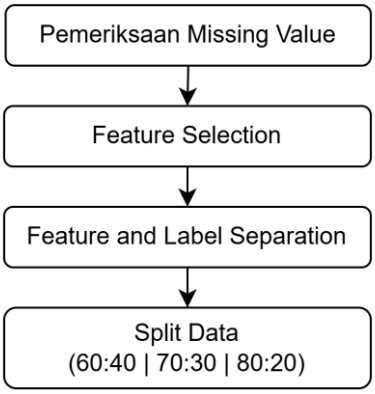
memahami karakteristik masing-masing fitur numerik. Rata-rata kadar hemoglobin pada dataset adalah 13,41 g/dL, dengan nilai minimum 6,6 g/dL dan maksimum 16 g/dL. Rentang ini mencerminkan perbedaan yang jelas antara individu dengan kondisi anemia dan non-anemia. Fitur lain seperti MCH, MCHC, dan MCV juga menunjukkan variasi nilai yang signifikan, yang mengindikasikan bahwa fitur-fitur tersebut memiliki potensi sebagai pembeda antara kelas anemia dan tidak anemia.

Selanjutnya, distribusi nilai hemoglobin dianalisis berdasarkan kelas label. Hasil visualisasi menunjukkan bahwa individu dengan label anemia cenderung memiliki nilai hemoglobin yang lebih rendah dibandingkan dengan individu tanpa anemia. Temuan ini selaras dengan definisi medis anemia yang ditandai oleh rendahnya kadar hemoglobin dalam darah, sehingga memperkuat relevansi fitur hemoglobin sebagai indikator utama dalam proses klasifikasi. Selain hemoglobin, fitur MCV dan MCH juga dianalisis terhadap kelas label. Individu dengan anemia umumnya menunjukkan nilai MCV dan MCH yang lebih rendah, yang berkaitan dengan jenis anemia mikrositik dan hipokromik. Pola ini menunjukkan bahwa fitur-fitur hematologi dalam dataset memiliki hubungan yang logis dan bermakna secara klinis terhadap kondisi anemia.

3.5 Preprocessing Data

Tahap preprocessing data dilakukan untuk menyiapkan dataset hematologi agar dapat digunakan secara optimal dalam proses pemodelan menggunakan algoritma XGBoost. Proses preprocessing pada penelitian ini mencakup beberapa tahapan, yaitu pemeriksaan nilai kosong (*missing value*), seleksi fitur, pemisahan

fitur dan label, serta pembagian dataset menjadi data latih dan data uji. Gambar 3.3 menunjukkan tahapan preprocessing data sebagai berikut :



Gambar 3.3 Tahapan Preprocessing Data

Pada tahap awal, dilakukan pemeriksaan terhadap keberadaan nilai kosong (*missing value*) pada seluruh fitur. Langkah ini bertujuan untuk memastikan bahwa tidak terdapat data yang hilang yang dapat memengaruhi kinerja model. Apabila ditemukan nilai kosong, maka akan dilakukan penanganan yang sesuai agar kualitas data tetap terjaga. Selanjutnya, dilakukan seleksi fitur untuk memastikan bahwa fitur yang digunakan dalam pemodelan memiliki relevansi terhadap tujuan penelitian. Fitur yang digunakan dalam penelitian ini difokuskan pada parameter hematologi utama, yaitu hemoglobin (Hb), MCH, MCHC, dan MCV. Variabel Gender tidak disertakan sebagai fitur dalam model. Keputusan ini didasarkan pada prinsip seleksi fitur dalam machine learning yang menyatakan bahwa fitur yang berpotensi menimbulkan bias atau tidak secara langsung merepresentasikan variabel utama penelitian dapat dieliminasi untuk meningkatkan objektivitas model. Penggunaan variabel demografis seperti gender dalam model klasifikasi berpotensi menyebabkan model memanfaatkan korelasi distribusi kelas, bukan mempelajari

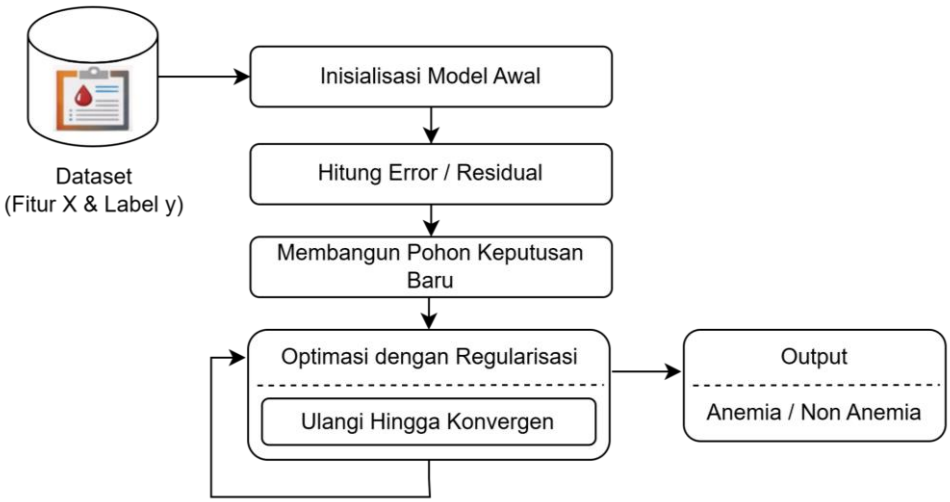
pola intrinsik dari parameter hematologi(Terven et al., 2025; Xing et al., 2021). Oleh karena itu, untuk menjaga fokus penelitian pada indikator klinis anemia, variabel Gender dihapus dari proses pemodelan.

Setelah proses pemeriksaan dan seleksi fitur, dataset dipisahkan menjadi fitur dan label. Fitur terdiri dari variabel hemoglobin, MCH, MCHC, dan MCV, sedangkan label merepresentasikan kondisi anemia dengan dua kelas, yaitu tidak anemia (0) dan anemia (1). Selanjutnya, dataset dibagi menjadi data latih dan data uji dengan beberapa variasi rasio pembagian sesuai dengan *experimental setup* yang telah ditetapkan. Pembagian ini dilakukan untuk mengukur kemampuan model dalam melakukan generalisasi terhadap data yang belum pernah digunakan dalam proses pelatihan. Proses normalisasi atau *scaling* data tidak dilakukan dalam penelitian ini karena algoritma XGBoost merupakan algoritma berbasis pohon keputusan yang tidak sensitif terhadap perbedaan skala antar fitur. Dengan demikian, model dapat langsung mempelajari pola dari data asli tanpa memerlukan transformasi skala tambahan.

3.6 Implementasi Algoritma Extreme Gradient Boosting

Pada tahap ini dilakukan pemodelan menggunakan algoritma Extreme Gradient Boosting (XGBoost). XGBoost merupakan salah satu algoritma *ensemble learning* berbasis *gradient boosting* yang bekerja dengan membangun sekumpulan *decision tree* secara bertahap untuk meningkatkan akurasi prediksi, di mana setiap model baru berfokus untuk memperbaiki kesalahan prediksi yang dihasilkan oleh model sebelumnya. Algoritma Extreme Gradient Boosting (XGBoost) digunakan dalam penelitian ini sebagai metode klasifikasi untuk mendeteksi anemia

berdasarkan data hematologi. Pendekatan ini memungkinkan XGBoost menghasilkan model klasifikasi yang kuat dan akurat, terutama pada data tabular yang bersifat numerik seperti data hematologi. Proses kerja XGBoost dapat dilihat pada Gambar 3.4, yang menggambarkan alur pembentukan model secara sistematis.



Gambar 3.4 Desain Implementasi XGBoost

Pada tahap implementasi, data latih yang telah melalui proses preprocessing digunakan sebagai masukan untuk melatih model XGBoost. Proses pelatihan bertujuan untuk mempelajari pola dan hubungan antara variabel input, yaitu Gender, Hemoglobin, MCH, MCHC, dan MCV, dengan label kelas anemia. Parameter-parameter hematologi tersebut merupakan indikator medis yang umum digunakan dalam pemeriksaan darah dan memiliki peran penting dalam menentukan kondisi anemia seseorang. Melalui proses pelatihan ini, model XGBoost diharapkan mampu mengenali karakteristik data yang membedakan individu dengan anemia dan tanpa anemia, sehingga dapat menghasilkan prediksi kelas yang akurat ketika diberikan data baru yang belum pernah dilihat sebelumnya.

Tahap pertama adalah inisialisasi model awal, di mana sistem membuat prediksi dasar (biasanya menggunakan rata-rata nilai target). Selanjutnya dilakukan perhitungan error atau residual, yaitu selisih antara nilai aktual dengan nilai prediksi sebelumnya. Nilai residual ini kemudian digunakan dalam tahap pembangunan pohon keputusan baru, yang bertujuan untuk memperbaiki kesalahan model sebelumnya. Setelah itu dilakukan pembaruan model (update model) dengan menambahkan hasil prediksi dari pohon baru yang telah dibentuk, dikalikan dengan *learning rate (η)* untuk mengontrol laju pembelajaran. Tahap berikutnya adalah optimasi dengan regularisasi, di mana XGBoost menerapkan regularisasi L1 dan L2 guna mengurangi kompleksitas model dan mencegah *overfitting*. Proses ini kemudian diulang dari tahap perhitungan residual hingga optimasi regularisasi sampai model mencapai kondisi konvergen, yakni ketika peningkatan akurasi sudah tidak signifikan lagi. Hasil akhir dari proses ini adalah model yang terdiri dari kumpulan beberapa *weak learners* (pohon keputusan kecil) yang secara kolektif menghasilkan prediksi akhir terhadap data uji.

Selain menjelaskan tahapan algoritma XGBoost, penelitian ini juga menguraikan implementasi teknis dari library XGBoost yang digunakan dalam pemodelan, termasuk fungsi dan parameter utama yang mendasari proses klasifikasi. Dalam penelitian ini, fungsi utama yang digunakan adalah XGBClassifier() dari modul xgboost.sklearn, yang menyediakan antarmuka serupa dengan library *scikit-learn*, sehingga memudahkan dalam proses pelatihan (fit), prediksi (predict), serta evaluasi model. Fungsi XGBClassifier() digunakan untuk membangun model klasifikasi berbasis *Extreme Gradient Boosting*. Parameter

objective='binary:logistic' digunakan karena permasalahan yang dihadapi bersifat klasifikasi biner, yaitu anemia (1) dan tidak anemia (0). Parameter eval_metric='logloss' digunakan untuk menghitung loss function selama proses pelatihan model. Sementara itu, parameter random_state=7 digunakan untuk memastikan proses pembagian data dan pelatihan model bersifat *reproducible* (hasilnya konsisten setiap kali dijalankan) (Mehdary et al., 2024; Terven et al., 2025). Parameter lainnya seperti n_estimators, max_depth, dan learning_rate dibiarkan pada nilai default karena penelitian ini berfokus pada evaluasi kinerja dari model XGBoost dengan optimasi hyperparameter tuning.

Dalam penelitian ini, model XGBoost tidak sepenuhnya menggunakan parameter bawaan (default), melainkan dilakukan penyesuaian awal pada beberapa parameter untuk mengurangi risiko overfitting. Penyesuaian dilakukan pada parameter *max_depth*, dan *learning_rate* guna mengontrol kompleksitas model dan meningkatkan kemampuan generalisasi. Selanjutnya, dilakukan proses *hyperparameter tuning* menggunakan metode Grid Search yang difokuskan pada parameter *n_estimators*, yaitu jumlah pohon keputusan (*decision trees*) yang dibangun dalam model. Pemilihan parameter ini untuk menganalisis pengaruh jumlah pohon terhadap performa klasifikasi anemia.

Tabel 3.3 Parameter Algoritma XGBoost

Parameter	**Nilai**
Objective	binary:logistic
Eval_metric	logloss
n_estimators	50, 70, 100 (variasi tuning)
max_depth	4
Learning_rate	0.1
Random_state	7

Pada Tabel 3.3 ditampilkan parameter yang digunakan dalam implementasi algoritma XGBoost pada penelitian ini. Parameter objective ditetapkan sebagai *binary:logistic* karena permasalahan yang dikaji merupakan klasifikasi biner antara kondisi anemia dan tidak anemia. Parameter eval_metric menggunakan *logloss* untuk mengevaluasi kesalahan prediksi berbasis probabilitas selama proses pelatihan model. Beberapa parameter model ditentukan secara eksplisit untuk mengontrol kompleksitas dan meningkatkan kemampuan generalisasi model. Parameter max_depth ditetapkan sebesar 4 guna membatasi kedalaman pohon keputusan sehingga risiko overfitting dapat diminimalkan. Selain itu, learning_rate sebesar 0,1 digunakan untuk mengatur laju pembelajaran agar proses boosting berlangsung lebih stabil. Parameter random_state ditetapkan sebesar 7 untuk memastikan reproduktibilitas hasil eksperimen.

Selanjutnya, dilakukan proses hyperparameter tuning menggunakan pendekatan *Grid Search* sederhana pada parameter n_estimators, dengan variasi nilai 50, 70, dan 100. Variasi ini bertujuan untuk menganalisis pengaruh jumlah pohon terhadap performa klasifikasi anemia pada dataset hematologi. Setelah proses pelatihan selesai, model digunakan untuk menghasilkan prediksi pada data uji dalam bentuk probabilitas kelas yang kemudian dikonversi menjadi label biner (0 = tidak anemia, 1 = anemia). Hasil prediksi tersebut selanjutnya digunakan dalam tahap evaluasi untuk mengukur kinerja model menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Dengan demikian, implementasi algoritma XGBoost pada penelitian ini dilakukan secara sistematis, mulai dari penentuan

konfigurasi parameter, proses pelatihan dan tuning, hingga evaluasi performa model pada data uji.

3.7 Experimental Setup

Pada penelitian ini, rancangan eksperimen dirancang untuk mengevaluasi performa algoritma XGBoost dalam mengklasifikasikan kondisi anemia berdasarkan data hematologi. Eksperimen yang dilakukan tidak hanya ditujukan untuk mencapai nilai akurasi yang tinggi, melainkan juga untuk menilai kestabilan model serta menganalisis pengaruh variasi pembagian data latih dan data uji terhadap hasil klasifikasi yang diperoleh. Hal ini dianggap penting mengingat dataset yang dimanfaatkan dalam penelitian memiliki jumlah sampel yang terbatas, sehingga pembagian data yang tidak sesuai berpotensi menghasilkan bias dan fenomena overfitting pada model. Rancangan eksperimen utama dalam penelitian ini dilakukan dengan membandingkan beberapa variasi proporsi pembagian antara data latih dan data uji, yakni 80:20, 70:30, dan 60:40, sebagaimana ditampilkan pada Tabel 3.4.

Penggunaan pembagian menjadi tiga variasi rasio data, yaitu 60:40, 70:30, dan 80:20 merupakan rentang rasio yang umum digunakan dalam penelitian klasifikasi berbasis data tabular numerik klinis(Ashisha et al., 2024; Gómez et al., 2025a). Rasio 90:10 tidak digunakan karena dataset berukuran sedang, peningkatan proporsi data latih hingga 90% justru berpotensi menurunkan performa pengujian model karena data terlalu sedikit menjadi kurang representatif (Nguyen et al., 2021). Sementara itu, rasio 50:50 tidak digunakan karena dapat mengurangi porsi

data latih secara signifikan sehingga model tidak memiliki cukup sampel untuk mempelajari pola secara optimal

Pembagian data dilakukan dengan memanfaatkan teknik *stratified splitting* guna memastikan bahwa distribusi kedua kelas yaitu anemia dan tidak anemia pada data pelatihan tetap seimbang dan representatif. Tujuan dari rancangan eksperimen ini adalah untuk memverifikasi bahwa model dilatih pada data yang mencerminkan karakteristik populasi secara keseluruhan, sekaligus untuk mengobservasi konsistensi dan stabilitas performa model terhadap variasi dalam proporsi data pelatihan dan data pengujian. Kedua, dilakukan pengujian variasi nilai hiperparameter n_estimators dengan memanfaatkan metode Grid Search sederhana, mencakup nilai-nilai 50, 70, dan 100. Variasi pengujian ini bertujuan untuk menganalisis dampak jumlah pohon keputusan yang dibangun terhadap performa model klasifikasi anemia, serta untuk mengidentifikasi konfigurasi hiperparameter yang menghasilkan evaluasi terbaik dalam konteks penelitian ini.

Tabel 3.4 Experimental Setup Model XGBoost

Experimental Setup	**Rasio Data Latih : Data Uji**	**n_estimators**
Eksperimen XGBoost dengan variasi split data dan Hyperparameter tuning.	60:40	50
		70
		100
	70:30	50
		70
		100
	80:20	50
		70
		100

Berdasarkan Tabel 3.4, total eksperimen yang dilakukan berjumlah sembilan kombinasi pengujian yang diperoleh dari variasi tiga rasio pembagian data dan tiga nilai parameter n_estimators. Pada setiap eksperimen pembagian data, model XGBoost dilatih menggunakan data latih dan selanjutnya dievaluasi menggunakan data uji. Evaluasi dilakukan dengan menggunakan metrik evaluasi klasifikasi, yaitu akurasi, precision, recall, dan F1-score, serta analisis confusion matrix untuk melihat kemampuan model dalam mengklasifikasikan masing-masing kelas secara benar. Dengan demikian, *experimental setup* dalam penelitian ini tidak hanya mengevaluasi pengaruh pembagian data, tetapi juga mengkaji dampak variasi hyperparameter terhadap kinerja model XGBoost secara komprehensif.

3.8 Evaluasi Model

Evaluasi model dilakukan untuk mengetahui kinerja algoritma Extreme Gradient Boosting (XGBoost) dalam mendeteksi anemia berdasarkan data hematologi. Proses evaluasi dilakukan dengan membandingkan hasil prediksi model terhadap label sebenarnya pada data uji. Data uji merupakan bagian dari dataset yang tidak digunakan selama proses pelatihan model, sehingga dapat memberikan gambaran objektif mengenai kemampuan generalisasi model dalam mengklasifikasikan kondisi anemia. Metode evaluasi yang digunakan dalam penelitian ini meliputi confusion matrix dan beberapa metrik performa klasifikasi, yaitu akurasi, presisi, recall, dan f1-score.

Confusion matrix digunakan untuk menggambarkan distribusi prediksi model. Melalui confusion matrix, dapat diketahui sejauh mana model mampu mengidentifikasi kasus anemia dan tidak anemia secara tepat. Evaluasi kinerja

model dilakukan menggunakan confusion matrix untuk menggambarkan perbandingan antara hasil prediksi dan label aktual. Confusion matrix terdiri dari empat komponen utama, yaitu true positive (TP), true negative (TN), false positive (FP), dan false negative (FN), yang digunakan sebagai dasar dalam perhitungan metrik evaluasi model. Sebagaimana pada Tabel 3.5.

Tabel 3.5 Confusion Matrix

Kelas Aktual/Prediksi	**Tidak Anemia (0)**	**Anemia (1)**
Tidak Anemia (0)	True Negative (TN)	False Positive (FP)
Anemia (1)	False Negative (FN)	True Positif (TP)

Akurasi digunakan untuk mengukur tingkat ketepatan model secara keseluruhan dalam mengklasifikasikan data. Namun, karena penelitian ini berkaitan dengan data medis, metrik presisi dan recall juga digunakan untuk memberikan evaluasi yang lebih komprehensif. Presisi menunjukkan ketepatan model dalam memprediksi kelas anemia, sedangkan recall menunjukkan kemampuan model dalam mendeteksi seluruh kasus anemia yang ada. Untuk menyeimbangkan nilai presisi dan recall, digunakan metrik f1-score sebagai indikator performa model secara keseluruhan. Metrik evaluasi yang digunakan dalam penelitian ini meliputi *accuracy*, *precision*, *recall*, dan f1-score.

Metrik-metrik tersebut dihitung berdasarkan nilai true positive (TP), true negative (TN), false positive (FP), dan false negative (FN) yang diperoleh dari confusion matrix. Penggunaan beberapa metrik evaluasi bertujuan untuk memberikan gambaran kinerja model secara lebih komprehensif, khususnya pada konteks deteksi anemia yang termasuk dalam domain kesehatan. Dengan rumus sebagai berikut:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.1)$$

Accuracy mengukur ketepatan prediksi model secara keseluruhan.

$$Precision = \frac{TP}{TP+FP} \quad (3.2)$$

Precision mengukur ketepatan model dalam memprediksi kelas anemia.

$$Recall = \frac{TP}{TP+FN} \quad (3.3)$$

Recall (Sensitivity) mengukur kemampuan model dalam mendeteksi seluruh kasus anemia.

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (3.4)$$

F1-Score merupakan rata-rata harmonis antara precision dan recall.

Keterangan:

- TP : True Positive, jumlah data yang sebenarnya termasuk dalam kelas positif dan berhasil diprediksi
- TN : True Negative, jumlah data yang sebenarnya termasuk dalam kelas negatif dan berhasil diprediksi
- FP : False Positive, jumlah data yang sebenarnya termasuk dalam kelas negatif, tetapi diprediksi secara keliru oleh model sebagai kelas positif
- FN : False Negatif, jumlah data yang sebenarnya termasuk dalam kelas positif, namun salah diklasifikasikan oleh model sebagai kelas negatif.

Hasil evaluasi model ini digunakan sebagai dasar untuk menilai performa model klasifikasi menggunakan algoritma XGBoost dalam mendeteksi anemia. Evaluasi ini juga menjadi acuan awal sebelum dilakukan pengembangan lebih lanjut, seperti optimasi parameter atau penggunaan metode evaluasi tambahan pada penelitian selanjutnya.

BAB IV

HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil penelitian yang telah dilakukan serta pembahasan terhadap penerapan algoritma XGBoost dalam klasifikasi anemia berdasarkan data hematologi. Pembahasan meliputi hasil pengujian model dengan berbagai skenario pembagian data latih dan data uji, serta variasi parameter yang digunakan. Untuk mengetahui performa model dalam klasifikasi secara optimal.

4.1 Hasil Preprocessing

Pada penelitian, tahap preprocessing data merupakan proses yang krusial yang dilakukan sebelum pembangunan model klasifikasi. Tahap ini bertujuan untuk memastikan kualitas data yang digunakan berada dalam kondisi optimal, sehingga dapat meminimalkan bias serta meningkatkan akurasi model yang dihasilkan. Proses preprocessing meliputi beberapa tahapan, yaitu identifikasi dan penanganan missing value, pemilihan fitur yang relevan (feature selection), pemisahan antara variabel independen (fitur) dan variabel dependen (label), serta pembagian data ke dalam data latih dan data uji.

4.1.1 Hasil Pemeriksaan Missing Value

Berikut merupakan hasil pemeriksaan missing value pada dataset yang digunakan dalam penelitian ini. Berdasarkan hasil analisis yang telah dilakukan, diketahui bahwa seluruh atribut pada dataset tidak memiliki nilai yang hilang (missing value). Dengan demikian, tidak diperlukan proses penanganan seperti

imputasi atau penghapusan data. Hasil pemeriksaan missing value tersebut ditampilkan pada Tabel 4.1.

Tabel 4.1 Hasil Pemeriksaan Missing Value

Fitur	Missing Value	Presentase (%)
Gender	0	.0
Hemoglobin	0	.0
MCH	0	.0
MCHC	0	.0
MCV	0	.0
Result	0	.0

Hasil pada tabel menunjukkan bahwa seluruh atribut memiliki nilai *missing value* sebesar 0 atau 0%, yang berarti tidak terdapat data yang hilang pada dataset. Hal ini juga diperkuat dengan pengecekan menggunakan data.isnull().values.any() yang menghasilkan kondisi *False*, sehingga dapat disimpulkan bahwa dataset dalam kondisi lengkap dan tidak memerlukan penanganan khusus terkait *missing value*.

4.1.2 Hasil *Feature and Label Separation*

Pada tahap pemisahan fitur dan label, dataset yang digunakan dalam penelitian dipisahkan menjadi variabel fitur (feature) dan variabel target (label). Variabel label yang digunakan dalam atribut *Result* yang merepresentasikan kondisi anemia, sedangkan atribut lainnya digunakan sebagai variabel fitur yang berperan sebagai input dalam proses klasifikasi. Proses pemisahan ini bertujuan untuk mempermudah pengolahan data serta menyesuaikan kebutuhan model dalam membedakan antara data masukan dan keluaran. Hasil dari proses pemisahan fitur dan label tersebut dapat dilihat pada Tabel 4.2 dan Tabel 4.3, yang menunjukkan pembagian atribut ke dalam masing-masing peran secara jelas.

Tabel 4.2 Hasil Pemisahan Fitur

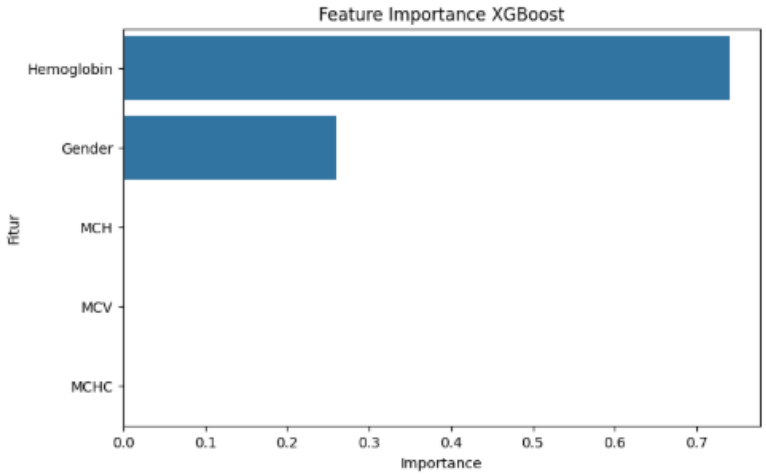
	Gender	Hemoglobin	MCH	MCHC	MCV
0	1	14,9	22,7	29,1	83,7
1	0	15,9	25,4	28,3	72,0
2	0	9,0	21,5	29,6	71,2
3	0	14,9	16,0	31,4	87,5
4	1	14,7	22,0	28,2	99,5
…	…	…	…	…	…

Tabel 4.3 Hasil Pemisahan Label

	Result
0	0
1	0
2	1
3	0
4	0
…	…

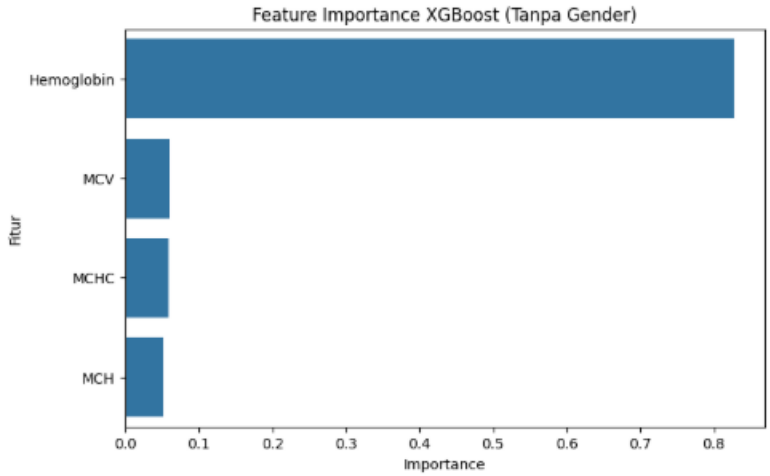
4.1.3 Hasil *Feature Selection*

Pada penelitian ini, proses seleksi fitur dilakukan dengan memanfaatkan nilai *feature importance* yang dihasilkan oleh algoritma XGBoost. Nilai ini menunjukkan kontribusi relatif setiap fitur dalam membangun model klasifikasi. Berdasarkan hasil tersebut, dilakukan analisis terhadap fitur-fitur yang memiliki pengaruh signifikan maupun yang cenderung rendah kontribusinya.



Gambar 4.1 Diagram Hasil *Feature Importance* dengan Gender.

Hasil visualisasi *feature importance* sebelum penghapusan fitur tertentu ditunjukkan pada Gambar 4.1 . Berdasarkan gambar tersebut, terlihat bahwa setiap fitur memiliki tingkat kontribusi yang berbeda dalam proses klasifikasi, termasuk variabel gender yang turut dianalisis pengaruhnya terhadap model. Namun demikian, variabel gender merupakan variabel demografis yang tidak secara langsung merepresentasikan kondisi klinis atau parameter hematologi pasien. Selain itu, penggunaan variabel ini berpotensi menimbulkan bias dalam proses klasifikasi, sehingga dapat memengaruhi objektivitas model. Oleh karena itu, dilakukan pengujian ulang dengan menghapus fitur gender untuk mengevaluasi dampaknya terhadap distribusi tingkat kepentingan fitur lainnya.



Gambar 4.2 Diagram Hasil *Feature Importance* tanpa Gender.

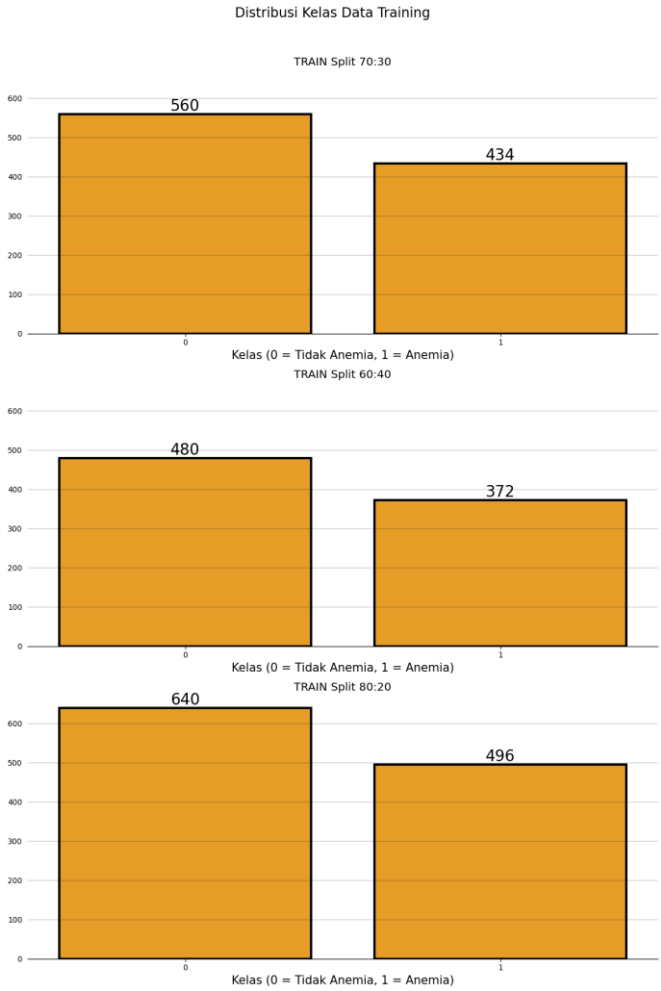
Hasil *feature importance* setelah penghapusan fitur gender ditunjukkan pada Gambar 4.2 . Perbandingan kedua hasil tersebut menunjukkan adanya perubahan nilai kepentingan pada beberapa fitur. Hal tersebut mengindikasi bahwa keberadaan fitur gender dapat memengaruhi struktur pembelajaran model. Berdasarkan hasil tersebut, fitur-fitur yang memiliki kontribusi paling relevan dipertahankan seperti MCV, MCH, dan MCHC, serta digunakan dalam tahap pemodelan. Sedangkan fitur yang kurang berpengaruh dapat dipertimbangkan untuk dihilangkan. Langkah ini juga bertujuan untuk memastikan bahwa model yang dihasilkan lebih fokus pada fitur-fitur klinis yang relevan. Hasil data setelah melakukan seleksi terdapat pada Tabel 4.4 berikut.

Tabel 4.4 Hasil Seleksi Fitur Tanpa Gender

	Hemoglobin	**MCH**	**MCHC**	**MCV**
0	14,9	22,7	29,1	83,7
1	15,9	25,4	28,3	72,0
2	9,0	21,5	29,6	71,2
3	14,9	16,0	31,4	87,5
4	14,7	22,0	28,2	99,5
…	…	…	…	…

4.1.4 Hasil *Split Data*

Hasil pembagian data divisualisasikan untuk melihat distribusi kelas pada data latih. Visualisasi tersebut ditunjukkan pada Gambar 4.3, yang menampilkan distribusi kelas untuk setiap eksperimen pembagian data. Berdasarkan hasil visualisasi terlihat bahwa proporsi antara kelas tidak anemia (0) dan anemia (1) tetap terjaga pada setiap skenario pembagian. Hal ini menunjukkan bahwa penggunaan teknik *stratified sampling* berhasil mempertahankan keseimbangan distribusi kelas pada data latih. Pembagian ini dilakukan menggunakan fungsi *train_test_split* dengan parameter *stratify* untuk menjaga proporsi distribusi kelas antara data latih dan data uji tetap seimbang. Selain itu, digunakan *random_state* bernilai 13 untuk memastikan hasil pembagian data bersifat konsisten dan dapat direproduksi.



Gambar 4.3 Distribusi Kelas Data Training

Berdasarkan Gambar 4.3, ditampilkan distribusi kelas pada data training untuk tiga eksperimen pembagian data, yaitu 80:20, 70:30, dan 60:40. Pada eksperimen rasio 70:30 jumlah data latih untuk kelas tidak anemia(0) sebanyak 560 data, sedangkan kelas anemia (1) sebanyak 434 data. Pada eksperimen rasio 60:40, jumlah data latih untuk kelas tidak anemia sebesar 480 data dan anemia sebesar 372 data. Sementara itu, pada eksperimen rasio 80:20 diperoleh jumlah data latih sebanyak 640 data untuk kelas tidak anemia dan 496 data untuk kelas anemia. Perbedaan rasio pembagian data memberikan variasi dalam jumlah data latih yang

digunakan untuk membangun model. Semakin besar proporsi data latih, seperti pada eksperimen rasio 80:20, maka semakin banyak data yang dapat dipelajari oleh model. Sebaliknya, pada eksperimen dengan proporsi data uji lebih besar, seperti 60:40, jumlah data latih menjadi lebih sedikit, namun evaluasi model dapat dilakukan dengan data uji yang lebih banyak. Oleh karena itu, penggunaan beberapa eksperimen ini bertujuan untuk membandingkan performa model pada kondisi pembagian data yang berbeda.

4.2 Hasil Eksperimen

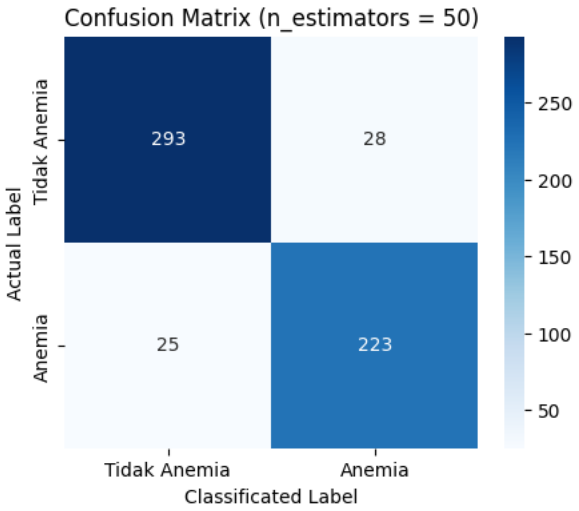
Pada subbab ini akan membahas hasil dari eksperimen dengan fitur yang telah diseleksi dan diklasifikasi menggunakan algoritma XGBoost dengan rasio-rasio, serta parameter yang sudah ditentukan.

4.2.1 Eksperimen Rasio 60:40

Pada pengujian ini, dilakukan pembagian dataset menggunakan metode *hold-out* dengan rasio 60:40, yaitu sebesar 60% data digunakan sebagai data latih (*training data*) dan 40% sebagai data uji (*testing data*).

4.2.1.1 Eksperimen Model A

Pada eksperimen pertama, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 50. Sebagaimana ditunjukkan pada Gambar 4.4 berikut.



Gambar 4.4 Confusion Matrix Model A

Pada Gambar 4.4 dipaparkan hasil confusion matrix model XGBoost dengan nilai n_estimators sebesar 50 pada proporsi pembagian data 60:40. Berdasarkan hasil yang diperoleh, terdapat 293 data yang berhasil diklasifikasikan sebagai True Negative (TN), merepresentasikan data tidak anemia yang diprediksi dengan akurat, dan sebanyak 223 data yang masuk dalam kategori True Positive (TP), merepresentasikan data anemia yang diprediksi dengan benar. Di samping itu, ditemukan 28 data yang tergolong False Positive (FP), yaitu data tidak anemia yang salah diklasifikasikan sebagai anemia, serta 25 data yang masuk kategori False Negative (FN), merepresentasikan data anemia yang salah diprediksi sebagai kondisi tidak anemia. Selain disajikan dalam format confusion matrix visual, performa model juga dievaluasi menggunakan sejumlah metrik evaluasi komprehensif, meliputi *accuracy*, *precision*, *recall*, dan F1-score. Hasil perhitungan dari metrik-metrik evaluasi tersebut untuk model A ditampilkan pada Tabel 4.5 berikut.

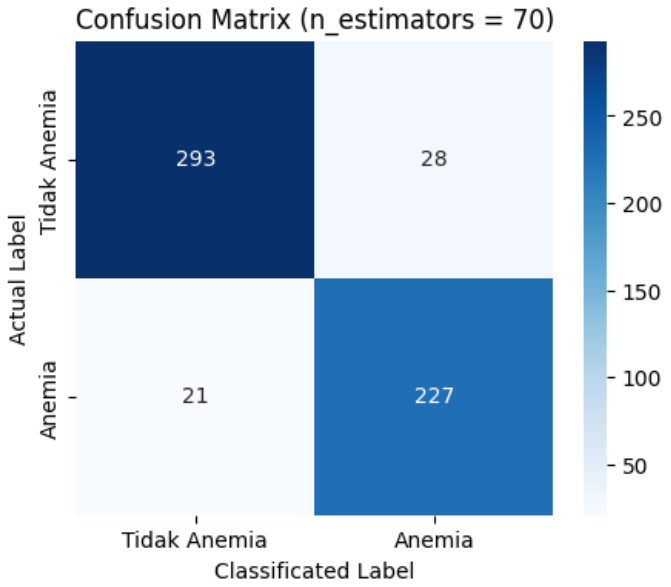
Tabel 4.5 Hasil Metrik Evaluasi Model A

Class	Model XGBoost			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.92	0.91	0.92	321
1	0.89	0.90	0.89	248
Accuracy			0.91	569
Macro avg	0.90	0.91	0.91	569
Weight avg	0.91	0.91	0.91	569

Berdasarkan Tabel 4.5, terlihat bahwa model XGBoost pada pengujian ini menghasilkan nilai akurasi sebesar 0,91. Untuk kelas 0 (tidak anemia), diperoleh nilai presisi sebesar 0,92, recall sebesar 0,91, dan F1-score sebesar 0,92. Sebaliknya, pada kelas 1 (anemia), diperoleh nilai presisi sebesar 0,89, recall sebesar 0,90, dan F1-score sebesar 0,89. Nilai-nilai ini mengindikasikan bahwa model mampu mendeteksi data anemia dengan cukup baik, meskipun kinerjanya sedikit lebih rendah jika dibandingkan dengan performa pada kelas tidak anemia. Selain itu, nilai macro average dan weighted average yang terletak pada rentang 0,90–0,91 menunjukkan bahwa secara komprehensif model memiliki performa yang relatif seimbang dan konsisten dalam mengklasifikasikan kedua kelas tersebut.

4.2.1.2 Eksperimen Model B

Pada eksperimen kedua, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 70. Sebagaimana ditunjukkan pada Gambar 4.5 berikut.



Gambar 4.5 Confusion Matrix Model B

Pada Gambar 4.5 dipaparkan hasil confusion matrix model XGBoost dengan nilai n_estimators sebesar 70 pada proporsi pembagian data 60:40. Berdasarkan hasil yang diperoleh, terdapat 293 data yang berhasil diklasifikasikan sebagai True Negative (TN), merepresentasikan data tidak anemia yang diprediksi dengan akurat, dan sebanyak 227 data yang masuk dalam kategori True Positive (TP), merepresentasikan data anemia yang diprediksi dengan benar. Di samping itu, ditemukan 28 data yang tergolong False Positive (FP), yaitu data tidak anemia yang salah diklasifikasikan sebagai anemia, serta 21 data yang termasuk False Negative (FN), merepresentasikan data anemia yang salah diprediksi sebagai kondisi tidak anemia. Selain disajikan dalam format confusion matrix visual, performa model juga dievaluasi menggunakan sejumlah metrik evaluasi komprehensif, meliputi accuracy, precision, recall, dan F1-score. Hasil perhitungan dari metrik-metrik evaluasi tersebut untuk model B ditampilkan pada Tabel 4.6 berikut.

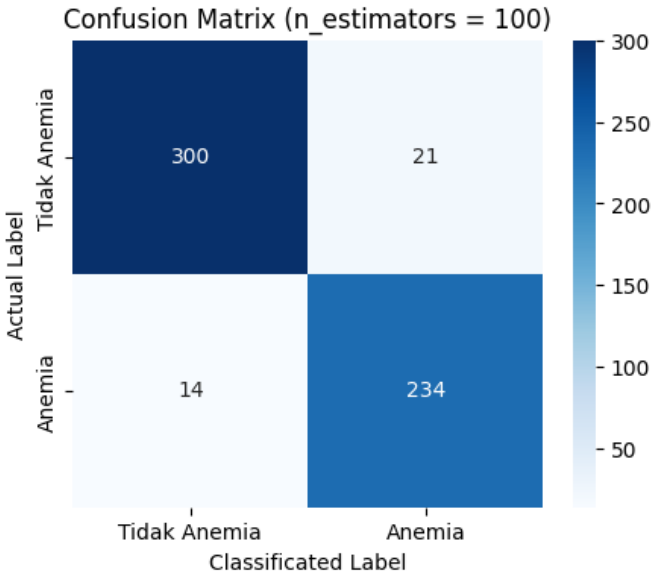
Tabel 4.6 Hasil Metrik Evaluasi Model B

Class	Model XGBoost			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.93	0.91	0.92	321
1	0.89	0.92	0.90	248
Accuracy			0.91	569
Macro avg	0.91	0.91	0.91	569
Weight avg	0.91	0.91	0.91	569

Berdasarkan Tabel 4.6, terlihat bahwa model XGBoost pada pengujian ini menghasilkan nilai akurasi sebesar 0,91. Untuk kelas 0 (tidak anemia), diperoleh nilai presisi sebesar 0,93, recall sebesar 0,89, dan F1-score sebesar 0,92. Sebaliknya, pada kelas 1 (anemia), diperoleh nilai presisi sebesar 0,89, recall sebesar 0,92, dan F1-score sebesar 0,90. Nilai-nilai ini mengindikasikan bahwa model menunjukkan kemampuan yang cukup memuaskan dalam mengidentifikasi kasus anemia, meskipun tingkat kinerjanya terukur sedikit lebih rendah jika dikomparasikan dengan performa pada kategori tidak anemia. Lebih lanjut, nilai macro average dan weighted average yang keduanya mencapai 0,91 mengungkapkan bahwa dalam perspektif menyeluruh, model mendemonstrasikan performa yang relatif selaras dan stabil dalam mengklasifikasikan kedua kelas yang diuji.

4.2.1.3 Eksperimen Model C

Pada eksperimen ketiga, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 100. Sebagaimana ditunjukkan pada Gambar 4.6 berikut.



Gambar 4.6 Confusion Matrix Model C

Pada Gambar 4.6 dipaparkan hasil confusion matrix model XGBoost dengan nilai n_estimators sebesar 100 pada proporsi pembagian data 60:40. Berdasarkan hasil yang diperoleh, terdapat 300 data yang berhasil diklasifikasikan sebagai True Negative (TN), merepresentasikan data tidak anemia yang diprediksi dengan akurat, dan sebanyak 234 data yang masuk dalam kategori True Positive (TP), merepresentasikan data anemia yang diprediksi dengan benar. Di samping itu, ditemukan 21 data yang tergolong False Positive (FP), yaitu data tidak anemia yang salah diklasifikasikan sebagai anemia, serta 14 data yang termasuk False Negative (FN), merepresentasikan data anemia yang salah diprediksi sebagai kondisi tidak anemia. Selain disajikan dalam format confusion matrix visual, performa model juga dievaluasi menggunakan sejumlah metrik evaluasi komprehensif, meliputi accuracy, precision, recall, dan F1-score. Hasil perhitungan dari metrik-metrik evaluasi tersebut untuk model C ditampilkan pada Tabel 4.7 berikut.

Tabel 4.7 Hasil Metrik Evaluasi Model C

Class	Model XGBoost			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.96	0.93	0.94	321
1	0.92	0.94	0.93	248
Accuracy			**0.94**	569
Macro avg	0.94	0.94	0.94	569
Weight avg	0.94	0.94	0.94	569

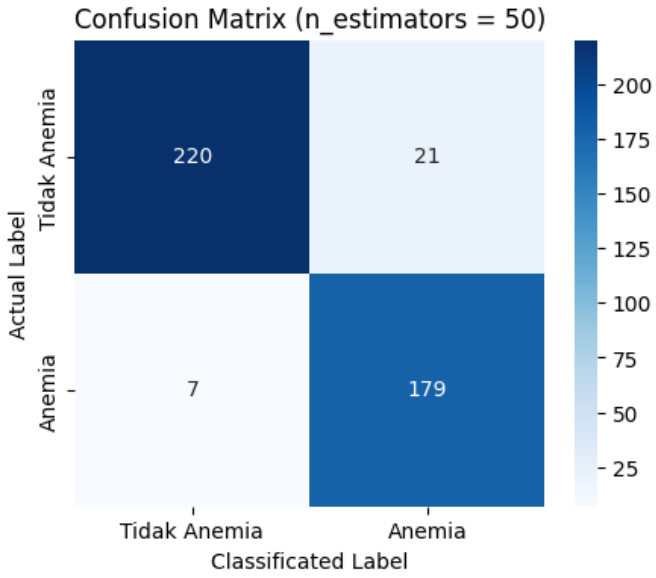
Berdasarkan Tabel 4.7, terlihat bahwa model XGBoost pada pengujian ini menghasilkan nilai akurasi sebesar 0,94. Untuk kelas 0 (tidak anemia), diperoleh nilai presisi sebesar 0,96, recall sebesar 0,93, dan F1-score sebesar 0,94. Sebaliknya, pada kelas 1 (anemia), diperoleh nilai presisi sebesar 0,92, recall sebesar 0,94, dan F1-score sebesar 0,93. Nilai-nilai ini mengindikasikan bahwa model menunjukkan kemampuan yang cukup memuaskan dalam mengidentifikasi kasus anemia, meskipun tingkat kinerjanya terukur sedikit lebih rendah jika dikomparasikan dengan performa pada kategori tidak anemia. Lebih lanjut, nilai *macro average* dan *weighted average* yang keduanya mencapai 0,94 mengungkapkan bahwa dalam perspektif menyeluruh, model mendemonstrasikan performa yang relatif selaras dan konsisten dalam mengklasifikasikan kedua kelas yang diuji.

4.2.2 Eksperimen Rasio 70:30

Pada pengujian ini, dilakukan pembagian dataset menggunakan metode *hold-out* dengan rasio 70:30, yaitu sebesar 70% data digunakan sebagai data latih (*training data*) dan 30% sebagai data uji (*testing data*).

4.2.2.1 Eksperimen Model D

Pada eksperimen keempat, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 50. Sebagaimana ditunjukkan pada Gambar 4.7 berikut.



Gambar 4.7 Confusion Matrix Model D

Pada Gambar 4.7 dipaparkan hasil confusion matrix model XGBoost dengan nilai n_estimators sebesar 50 pada proporsi pembagian data 70:30. Berdasarkan hasil yang diperoleh, terdapat 220 data yang berhasil diklasifikasikan sebagai True Negative (TN), merepresentasikan data tidak anemia yang diprediksi dengan akurat, dan sebanyak 179 data yang masuk dalam kategori True Positive (TP), merepresentasikan data anemia yang diprediksi dengan benar. Di samping itu, ditemukan 21 data yang tergolong False Positive (FP), yaitu data tidak anemia yang salah diklasifikasikan sebagai anemia, serta 7 data yang termasuk False Negative (FN), merepresentasikan data anemia yang salah diprediksi sebagai kondisi tidak anemia. Selain disajikan dalam format confusion matrix visual, performa model

juga dievaluasi menggunakan sejumlah metrik evaluasi komprehensif, meliputi *accuracy, precision, recall,* dan F1-score. Hasil perhitungan dari metrik-metrik evaluasi tersebut untuk model D ditampilkan pada Tabel 4.8 berikut.

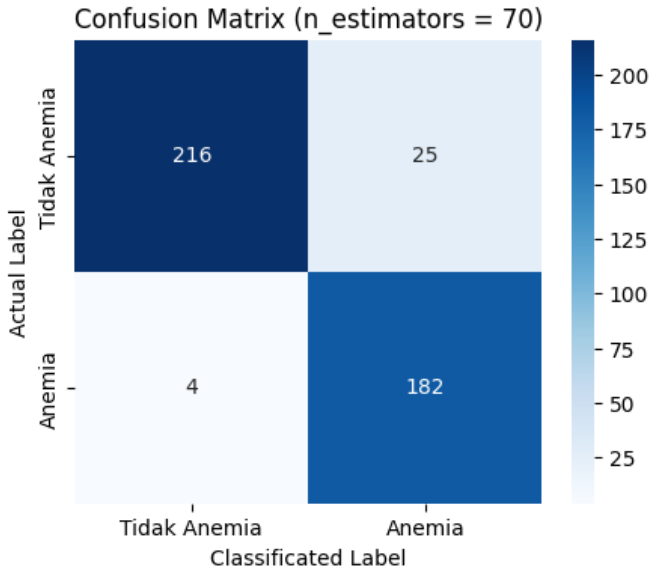
Tabel 4.8 Hasil Metrik Evaluasi Model D

Class	**Model XGBoost**			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.97	0.91	0.94	241
1	0.90	0.96	0.93	186
Accuracy			**0.93**	427
Macro avg	0.93	0.94	0.93	427
Weight avg	0.94	0.93	0.93	427

Berdasarkan Tabel 4.8, terlihat bahwa model XGBoost pada pengujian ini menghasilkan nilai akurasi sebesar 0,93. Untuk kelas 0 (tidak anemia), diperoleh nilai presisi sebesar 0,97, recall sebesar 0,91, dan F1-score sebesar 0,94. Sebaliknya, pada kelas 1 (anemia), diperoleh nilai presisi sebesar 0,90, recall sebesar 0,96, dan F1-score sebesar 0,93. Nilai-nilai ini mengindikasikan bahwa model menunjukkan kemampuan yang cukup memuaskan dalam mengidentifikasi kasus anemia, meskipun tingkat kinerjanya terukur sedikit lebih rendah jika dikomparasikan dengan performa pada kategori tidak anemia. Lebih lanjut, nilai macro average dan weighted average yang berada pada rentang 0,93–0,94 mengungkapkan bahwa dalam perspektif menyeluruh, model mendemonstrasikan performa yang relatif selaras dan konsisten dalam mengklasifikasikan kedua kelas yang diuji.

4.2.2.2 Eksperimen Model E

Pada eksperimen model kelima, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 70. Sebagaimana ditunjukkan pada Gambar 4.8 berikut.



Gambar 4.8 Confusion Matrix Model E

Pada Gambar 4.8 dipaparkan hasil confusion matrix model XGBoost dengan nilai n_estimators sebesar 70 pada proporsi pembagian data 70:30. Berdasarkan hasil yang diperoleh, terdapat 216 data yang berhasil diklasifikasikan sebagai True Negative (TN), merepresentasikan data tidak anemia yang diprediksi dengan akurat, dan sebanyak 182 data yang masuk dalam kategori True Positive (TP), merepresentasikan data anemia yang diprediksi dengan benar. Di samping itu, ditemukan 25 data yang tergolong False Positive (FP), yaitu data tidak anemia yang salah diklasifikasikan sebagai anemia, serta 4 data yang termasuk False Negative (FN), merepresentasikan data anemia yang salah diprediksi sebagai kondisi tidak anemia. Selain disajikan dalam format confusion matrix visual, performa model

juga dievaluasi menggunakan sejumlah metrik evaluasi komprehensif, meliputi accuracy, precision, recall, dan F1-score. Hasil perhitungan dari metrik-metrik evaluasi tersebut untuk model E ditampilkan pada Tabel 4.9 berikut.

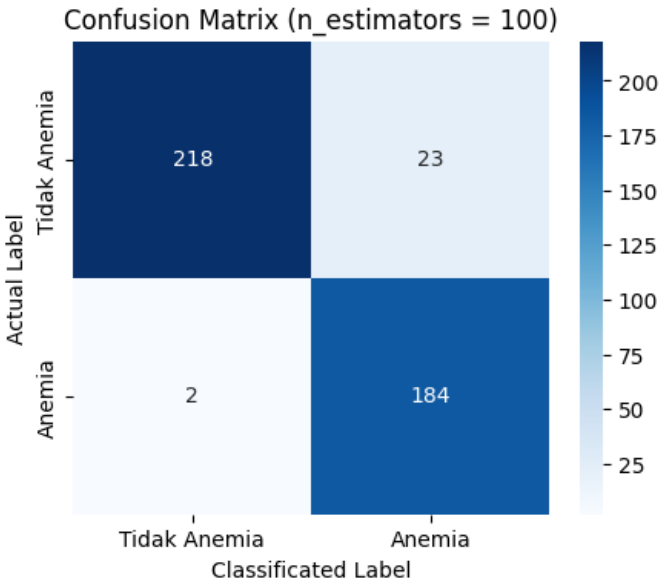
Tabel 4.9 Hasil Metrik Evaluasi Model E

Class	**Model XGBoost**			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.98	0.90	0.94	241
1	0.88	0.98	0.93	186
Accuracy			0.93	427
Macro avg	0.93	0.94	0.93	427
Weight avg	0.94	0.93	0.93	427

Hasil pada Tabel 4.9 menunjukkan bahwa model XGBoost mencapai *accuracy* sebesar 0,93. Untuk kelas tidak anemia (kelas 0), model memiliki *precision*, *recall*, dan *F1-score* masing-masing sebesar 0,98, 0,90, dan 0,94. Sementara itu, pada kelas anemia (kelas 1), nilai ketiga metrik tersebut tercatat sebesar 0,88, 0,98, dan 0,93. Perbedaan nilai tersebut menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengenali kedua kelas, dengan performa yang sedikit lebih unggul pada kelas tidak anemia. Nilai *macro average* dan *weighted average* yang berkisar antara 0,93–0,94 juga mengindikasikan bahwa model mampu mempertahankan kinerja klasifikasi yang relatif seimbang pada seluruh kelas data.

4.2.2.3 Eksperimen Model F

Pada eksperimen model keenam, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 100. Sebagaimana ditunjukkan pada Gambar 4.9 berikut.



Gambar 4.9 Confusion Matrix Model F

Hasil *confusion matrix* model XGBoost dengan nilai *n_estimators* 100 dan rasio pembagian data 70:30 ditampilkan pada Gambar 4.9. Dari hasil pengujian tersebut diketahui bahwa model mampu mengklasifikasikan 218 data tidak anemia secara benar (*True Negative*) dan 184 data anemia secara benar (*True Positive*). Sementara itu, kesalahan klasifikasi terjadi pada 23 data tidak anemia yang diprediksi sebagai anemia (*False Positive*) serta 2 data anemia yang diprediksi sebagai tidak anemia (*False Negative*). Temuan ini menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan antara data anemia dan tidak anemia. Untuk memperoleh gambaran performa yang lebih komprehensif, evaluasi juga dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Adapun hasil pengukuran metrik evaluasi pada model F dapat dilihat pada Tabel 4.10

Tabel 4.10 Hasil Metrik Evaluasi Model F

Class	Model XGBoost			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.99	0.90	0.95	241
1	0.89	0.99	0.94	186
Accuracy			0.94	427
Macro avg	0.94	0.95	0.94	427
Weight avg	0.95	0.94	0.94	427

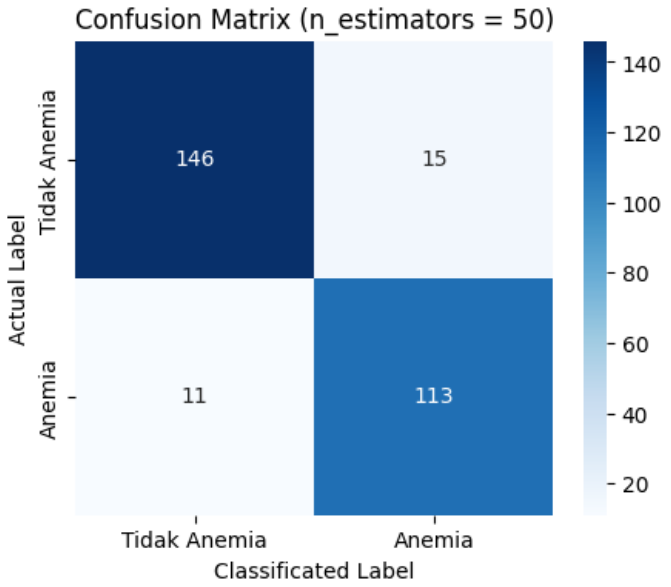
Tabel 4.10 menunjukkan bahwa model XGBoost mencapai tingkat *accuracy* sebesar 0,94 pada pengujian yang dilakukan. Untuk kelas tidak anemia (kelas 0), nilai *precision*, *recall*, dan *F1-score* masing-masing tercatat sebesar 0,99, 0,90, dan 0,95. Adapun pada kelas anemia (kelas 1), model memperoleh nilai *precision* sebesar 0,89, *recall* sebesar 0,99, dan *F1-score* sebesar 0,94. Perolehan nilai tersebut mengindikasikan bahwa model memiliki kemampuan yang sangat baik dalam mengenali kedua kelas, dengan performa yang sedikit lebih unggul pada kelas tidak anemia. Nilai *macro average* dan *weighted average* yang berada pada kisaran 0,94–0,95 juga menunjukkan bahwa model mampu mempertahankan keseimbangan performa klasifikasi secara keseluruhan.

4.2.3 Eksperimen Rasio 80:20

Pada pengujian ini, dilakukan pembagian dataset menggunakan metode *hold-out* dengan rasio 80:20, yaitu sebesar 80% data digunakan sebagai data latih (*training data*) dan 20% sebagai data uji (*testing data*).

4.2.3.1 Eksperimen Model G

Pada eksperimen model ketujuh, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 50. Sebagaimana ditunjukkan pada Gambar 4.10 berikut.



Gambar 4.10 Confusion Matrix Model G

Hasil *confusion matrix* model XGBoost dengan nilai *n_estimators* 50 dan rasio pembagian data 60:40 ditampilkan pada Gambar 4.10. Dari hasil tersebut diketahui bahwa model mampu mengklasifikasikan 146 data tidak anemia secara benar (*True Negative*) dan 113 data anemia secara benar (*True Positive*). Sementara itu, kesalahan klasifikasi terjadi pada 15 data tidak anemia yang diprediksi sebagai anemia (*False Positive*) serta 11 data anemia yang diprediksi sebagai tidak anemia (*False Negative*). Hasil ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan data anemia dan tidak anemia. Untuk memperoleh evaluasi yang lebih menyeluruh, performa model juga diukur menggunakan metrik

accuracy, *precision*, *recall*, dan *F1-score*. Hasil pengukuran metrik evaluasi pada model G dapat dilihat pada Tabel 4.11.

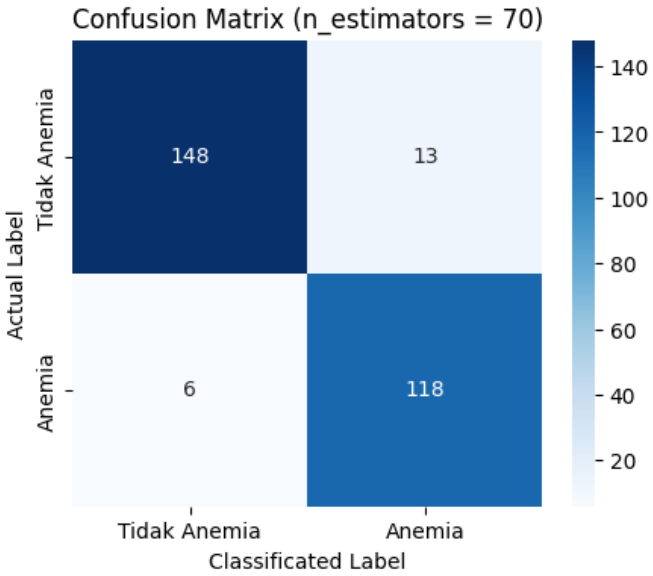
Tabel 4.11 Hasil Metrik Evaluasi Model G

Class	**Model XGBoost**			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.93	0.91	0.92	161
1	0.88	0.91	0.90	124
Accuracy			0.91	285
Macro avg	0.91	0.91	0.91	285
Weight avg	0.91	0.91	0.91	285

Berdasarkan hasil evaluasi pada Tabel 4.11, model XGBoost berhasil mencapai tingkat *accuracy* sebesar 0,91. Untuk kelas tidak anemia (kelas 0), nilai *precision*, *recall*, dan *F1-score* yang diperoleh masing-masing adalah 0,93, 0,88, dan 0,92. Adapun pada kelas anemia (kelas 1), model memperoleh nilai *precision* sebesar 0,88, *recall* sebesar 0,91, dan *F1-score* sebesar 0,90. Perolehan nilai tersebut menunjukkan bahwa model mampu mengenali kedua kelas dengan cukup baik, walaupun performa pada kelas tidak anemia masih sedikit lebih unggul. Nilai *macro average* dan *weighted average* sebesar 0,91 juga mengindikasikan bahwa model memiliki performa klasifikasi yang konsisten dan seimbang secara keseluruhan.

4.2.3.2 Eksperimen Model H

Pada eksperimen model kedelapan, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 70. Sebagaimana ditunjukkan pada Gambar 4.11 berikut.



Gambar 4.11 Confusion Matrix Model H

Gambar 4.11 menyajikan hasil *confusion matrix* dari model XGBoost dengan parameter *n_estimators* sebesar 70 pada skenario pembagian data 60:40. Berdasarkan hasil klasifikasi yang diperoleh, model berhasil mengidentifikasi 148 data tidak anemia secara tepat sebagai *True Negative* (TN) dan 118 data anemia secara tepat sebagai *True Positive* (TP). Sementara itu, terdapat 13 data yang termasuk kategori *False Positive* (FP), yaitu data tidak anemia yang diprediksi sebagai anemia, serta 6 data yang termasuk kategori *False Negative* (FN), yaitu data anemia yang diprediksi sebagai tidak anemia. Hasil tersebut menunjukkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar oleh model. Selain dianalisis melalui *confusion matrix*, performa model juga dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Adapun hasil perhitungan metrik evaluasi untuk model H disajikan pada Tabel 4.12.

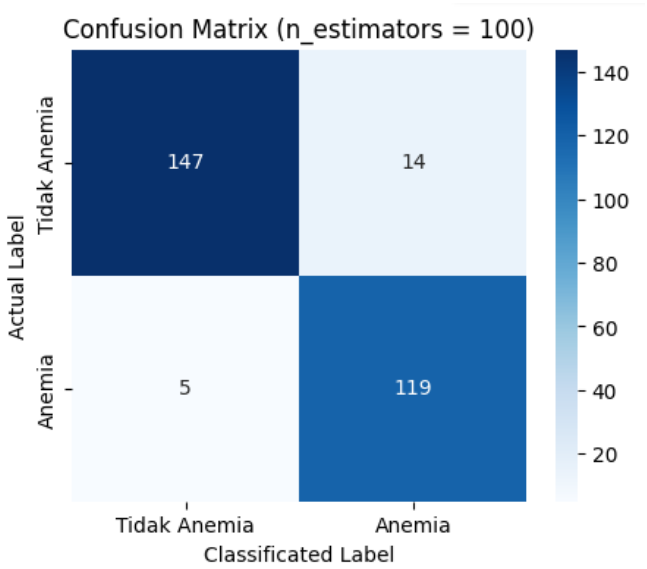
Tabel 4.12 Hasil Metrik Evaluasi Model H

Class	Model XGBoost			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.96	0.92	0.94	161
1	0.90	0.95	0.93	124
Accuracy			0.93	285
Macro avg	0.93	0.94	0.93	285
Weight avg	0.93	0.93	0.93	285

Hasil evaluasi yang ditampilkan pada Tabel 4.12 menunjukkan bahwa model XGBoost mencapai nilai *accuracy* sebesar 0,93. Pada kelas 0 (tidak anemia), model memperoleh nilai *precision* sebesar 0,96, *recall* sebesar 0,92, dan *F1-score* sebesar 0,94. Sementara itu, untuk kelas 1 (anemia), nilai *precision*, *recall*, dan *F1-score* yang diperoleh masing-masing adalah 0,90, 0,95, dan 0,93. Perolehan nilai tersebut mengindikasikan bahwa model mampu mengidentifikasi data anemia dengan baik, meskipun performanya masih sedikit lebih tinggi pada kelas tidak anemia. Selain itu, nilai *macro average* dan *weighted average* yang berada pada rentang 0,93 hingga 0,94 menunjukkan bahwa model memiliki kemampuan klasifikasi yang relatif konsisten dan seimbang pada kedua kelas.

4.2.3.3 Eksperimen Model I

Pada eksperimen model kesembilan, didapat hasil confusion matrix dari eksperimen yang dilakukan menggunakan variasi parameter n_estimators = 100. Sebagaimana ditunjukkan pada Gambar 4.12 berikut.



Gambar 4.12 Confusion Matrix Model I

Gambar 4.12 memperlihatkan hasil *confusion matrix* dari model XGBoost yang menggunakan nilai *n_estimators* sebesar 100 dengan rasio pembagian data 60:40. Berdasarkan hasil klasifikasi tersebut, sebanyak 147 data tidak anemia berhasil dikenali dengan benar sebagai *True Negative* (TN), sedangkan 119 data anemia berhasil diprediksi secara tepat sebagai *True Positive* (TP). Di samping itu, terdapat 14 data yang termasuk *False Positive* (FP), yaitu data tidak anemia yang keliru diklasifikasikan sebagai anemia. Sementara itu, sebanyak 5 data termasuk *False Negative* (FN), yaitu data anemia yang salah diprediksi sebagai tidak anemia. Secara umum, hasil ini menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat ketepatan yang cukup baik. Selain menggunakan *confusion matrix*, evaluasi performa model juga dilakukan melalui perhitungan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi untuk model I selanjutnya disajikan pada Tabel 4.13.

Tabel 4.13 Hasil Metrik Evaluasi Model I

Class	Model XGBoost			
	Precision	**Recall**	**F1-Score**	**Support**
0	0.97	0.91	0.94	161
1	0.89	0.96	0.93	124
Accuracy			**0.93**	285
Macro avg	0.93	0.94	0.93	285
Weight avg	0.94	0.93	0.93	285

Berdasarkan hasil evaluasi pada Tabel 4.13, model XGBoost memperoleh nilai *accuracy* sebesar 0,93. Pada kelas 0 (tidak anemia), nilai *precision*, *recall*, dan *F1-score* yang dihasilkan masing-masing sebesar 0,97, 0,91, dan 0,94. Sementara itu, pada kelas 1 (anemia), model mencapai nilai *precision* sebesar 0,89, *recall* sebesar 0,96, dan *F1-score* sebesar 0,93. Hasil tersebut menunjukkan bahwa model mampu mengklasifikasikan kedua kelas dengan baik. Tingginya nilai *recall* pada kelas anemia mengindikasikan bahwa sebagian besar data anemia berhasil terdeteksi oleh model, sehingga risiko kesalahan dalam mengidentifikasi kasus anemia relatif rendah. Selain itu, nilai *macro average* dan *weighted average* yang berada pada kisaran 0,93–0,94 menunjukkan bahwa performa model cukup konsisten dan seimbang pada seluruh kelas data yang diuji.

4.3 Pembahasan

Pada tahap ini dilakukan analisis terhadap hasil pengujian model klasifikasi menggunakan algoritma XGBoost pada berbagai eksperimen. Yaitu, variasi rasio pembagian data (60:40, 70:30, dan 80:20) serta variasi nilai *n_estimators* (50, 70, dan 100). Analisis ini dilakukan untuk membandingkan kinerja model serta

menentukan konfigurasi parameter yang menghasilkan performa terbaik berdasarkan metrik evaluasi berupa *accuracy, precision, recall,* dan f1-score. Untuk mempermudah analisis secara menyeluruh, hasil evaluasi dari seluruh skenario pengujian dirangkum dalam bentuk tabel perbandingan. Tabel ini menyajikan kinerja model berdasarkan variasi rasio pembagian data dan nilai *n_estimators,* sehingga dapat digunakan untuk mengidentifikasi pola performa serta menentukan konfigurasi model terbaik.

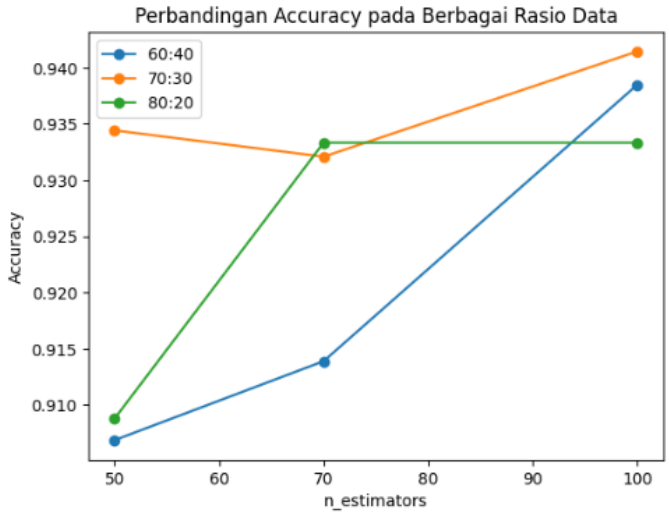
Tabel 4.14 Ringkasan Hasil Metrik Evaluasi

Model	**Rasio Data Latih : Data Uji**	**n_estimators**	**accuracy**	**precision**	**recall**	**f1-score**
A	60:40	50	0.906854	0.888446	0.899194	0.893788
B		70	0.913884	0.890196	0.915323	0.902584
C		100	0.938489	0.917647	0.943548	0.930417
D	70:30	50	0.934426	0.895000	0.962366	0.927461
E		70	0.932084	0.879227	0.978495	0.926209
F		100	0.941452	0.888889	0.989247	0.936387
G	80:20	50	0.908772	0.882812	0.911290	0.896825
H		70	0.933333	0.900763	0.951613	0.925490
I		100	0.933333	0.894737	0.959677	0.926070

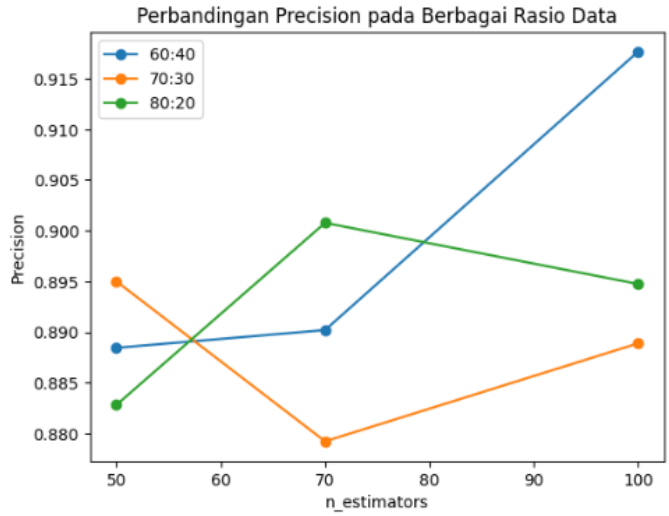
Berdasarkan ringkasan hasil metrik evaluasi pada Tabel 4.14, dapat diketahui bahwa peningkatan nilai *n_estimators* secara umum memberikan peningkatan performa model. Pada rasio 60:40, nilai *accuracy* meningkat dari 0,9069 pada *n_estimators* 50 menjadi 0,9385 pada *n_estimators* 100, dengan peningkatan yang sejalan pada *precision, recall,* dan F1-score. Hal serupa juga terlihat pada rasio 70:30, di mana nilai *accuracy* tertinggi diperoleh pada *n_estimators* 100 sebesar 0,9415 dengan F1-score sebesar 0,9364. Sementara itu,

pada rasio 80:20, peningkatan performa terjadi dari *n_estimators* 50 ke 70, namun cenderung stabil pada *n_estimators* 100 dengan nilai *accuracy* sebesar 0,9333.

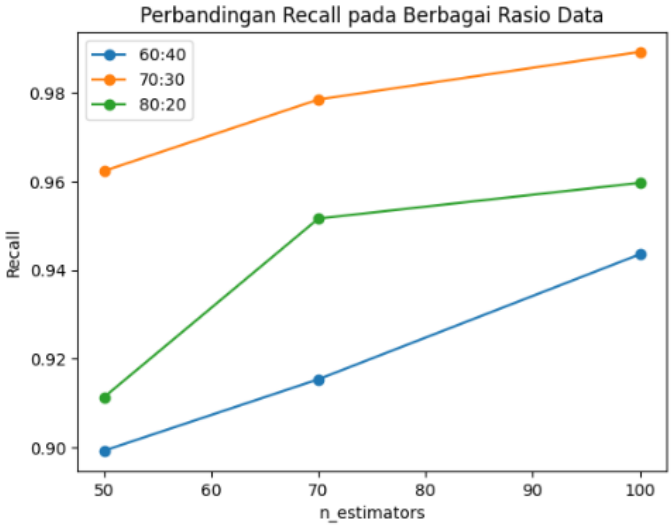
Untuk memperjelas perbandingan kinerja model pada seluruh skenario pengujian, hasil evaluasi juga disajikan dalam bentuk diagram berdasarkan masing-masing metrik. Diagram ini bertujuan untuk memberikan visualisasi yang lebih jelas terhadap perubahan performa mode pada setiap variasi rasio pembagian data dan nilai n_estimators.



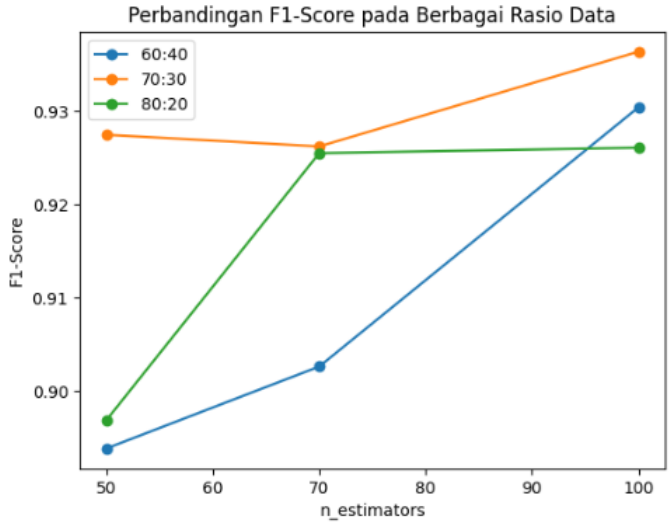
Gambar 4.13 Diagram Perbandingan *Accuracy* pada Berbagai Rasio Data



Gambar 4.14 Diagram Perbandingan Precision pada Berbagai Rasio Data



Gambar 4.15 Diagram Perbandingan Recall pada Berbagai Rasio Data



Gambar 4.16 Diagram Perbandingan F1-Score pada Berbagai Rasio Data

Berdasarkan diagram perbandingan yang ditampilkan, jika dibandingkan antar rasio pembagian data, rasio 70:30 menghasilkan performa terbaik secara keseluruhan, khususnya pada nilai *accuracy* dan *recall*. Hal ini menunjukkan bahwa proporsi data latih yang lebih besar dibandingkan data uji mampu meningkatkan kemampuan model dalam mempelajari pola data. Rasio 60:40 juga menunjukkan performa yang baik, namun sedikit lebih rendah dibandingkan rasio 70:30.

Sementara itu, pada rasio 80:20, meskipun data latih lebih banyak, peningkatan performa tidak terlalu signifikan, yang mengindikasikan bahwa penambahan data latih tidak selalu berbanding lurus dengan peningkatan kinerja model.

Dari sisi metrik evaluasi, terlihat adanya kecenderungan bahwa nilai recall cenderung lebih tinggi dibandingkan precision, terutama pada rasio 70:30 dan 80:20. Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam mendeteksi kelas positif (anemia), namun dengan konsekuensi adanya peningkatan kesalahan dalam bentuk *false positive*. Kondisi ini menunjukkan adanya *trade-off* antara precision dan recall, di mana peningkatan kemampuan deteksi dapat menyebabkan penurunan ketepatan prediksi pada kelas tertentu. Dalam konteks klasifikasi anemia, nilai *recall* menjadi salah satu metrik yang penting karena berkaitan dengan kemampuan model dalam mendeteksi pasien yang benar-benar mengalami anemia. Nilai *recall* yang tinggi menunjukkan bahwa model mampu meminimalkan kesalahan dalam mengklasifikasikan pasien anemia sebagai tidak anemia (*false negative*), yang sangat penting dalam aplikasi medis.

Salah satu pertanyaan mendasar yang melatarbelakangi penelitian ini adalah sejauh mana algoritma XGBoost mampu mempertahankan performa klasifikasi yang baik meskipun hanya mengandalkan parameter hematologi dasar, yaitu Hb, MCV, MCH, MCHC. Berdasarkan hasil eksperimen yang telah dilakukan, dapat disimpulkan bahwa XGBoost terbukti mampu mempertahankan performa klasifikasi yang tinggi pada dataset hematologi minimalis ini. Hal ini dibuktikan oleh konfigurasi terbaik pada Model F yang menghasilkan *accuracy* sebesar 0,9415, F1-Score sebesar 0,9364, dan recall sebesar 0,9892. Yang berarti model berhasil

mendeteksi 98,92% kasus anemia yang sebenarnya hanya dengan empat parameter dasar.

Kemampuan ini dapat dijelaskan oleh tiga faktor utama. Pertama, hemoglobin sebagai fitur dengan kontribusi tertinggi berdasarkan analisis *feature importance* sudah mampu menjadi pembeda kelas yang kuat pada sebagian besar data, karena secara klinis hemoglobin memang merupakan indikator utama diagnosis anemia. Kedua, fitur MCV, MCH, dan MCHC berperan melengkapi keputusan klasifikasi pada kasus-kasus borderline, di mana nilai hemoglobin saja tidak cukup untuk menentukan kelas. Terbukti dari analisis misklasifikasi yang menunjukkan bahwa nilai MCH yang rendah (Tabel 4.16 mean 22,04 pg) pada kasus False Positive turut memengaruhi keputusan model. Ketiga, kemampuan XGBoost dalam menangkap interaksi non-linear antar fitur melalui 100 pohon keputusan memungkinkan model memaksimalkan informasi yang tersedia meskipun jumlah fiturnya terbatas. Dengan demikian, meskipun dataset yang digunakan bersifat minimalis, kombinasi keempat parameter hematologi dasar tersebut sudah mengandung informasi yang cukup untuk mendukung klasifikasi anemia yang akurat ketika dimodelkan menggunakan XGBoost.

Analisis lebih lanjut terhadap FP dan FN menunjukkan pola yang bervariasi antar model dan rasio pembagian data. Perbandingan angka absolut FP dan FN antar rasio tidak dapat dilakukan secara langsung. Hal ini dikarenakan ukuran data uji yang berbeda-beda (569, 427, dan 285 data) pada setiap rasio. Apabila dihitung secara proporsional terhadap jumlah data aktual masing-masing kelas, rasio 70:30 secara konsisten menghasilkan FN rate terendah di seluruh variasi n_estimators.

Bahkan lebih rendah dibandingkan rasio 80:20 yang memiliki data training lebih banyak. Hal ini disebabkan oleh ukuran data uji pada rasio 80:20 yang lebih kecil, sehingga distribusinya kurang representatif terhadap populasi, khususnya pada kasus borderline yang berbeda di ambang batas keputusan model. Sebagaimana pada Tabel 4.15 dibawah.

Terkait pola fluktuasi FP, peningkatan n_estimators tidak selalu menghasilkan penurunan FP secara linear. Pada rasio 70:30, nilai FP justru meningkat dari 21 pada n_estimators=50 menjadi 25 pada n_estimators=70, sebelum kembali turun menjadi 23 pada n_estimators=100. Fenomena ini mencerminkan adanya *trade-off* antara precision dan recall. Dimana peningkatan sensitivitas model dalam mendeteksi kelas anemia (recall naik) menyebabkan lebih banyak kasus tidak anemia yang ikut terprediksi positif (FP naik). Pada n_estimators=100, model telah memiliki cukup pohon keputusan antar kelas, sehingga FP kembali turun sementara FN terus berkurang hingga mencapai nilai terendah sebesar 2. Untuk menganalisis karakteristik data yang mengalami misklasifikasi secara lebih mendalam, dilakukan pemeriksaan terhadap nilai rata-rata seluruh fitur hematologi pada sampel False Positive dan False Negative. Hasil pemeriksaan tersebut ditampilkan pada Tabel 4.15, dan Tabel 4.16 berikut.

Tabel 4.15 Perbandingan Rata-rata Hemoglobin Data Misklasifikasi terhadap Rata-rata Kelas pada Data Uji Model F.

Kategori	Label Aktual	Label Prediksi	Jumlah	Mean Hb (g/dL)	Rentang Hb (g/dL)
True Negative	0	0	218	14,81	12,0-16,9
True Positive	1	1	184	11,63	6,6-13,4
False Positive	0	1	23	12,69	12,0-13,3
False Negative	1	0	2	12,60	12,0-13,2

Tabel 4.16 Rata-rata Seluruh Fitur Hematologi pada Data Misklasifikasi Model F

Kategori	Satuan	False Positive (n=23)	False Negative (n-2)
Hemoglobin	g/dL	12,69	12,60
MCH	pg	22,04	22,00
MCHC	g/dL	29,98	28,15
MCV	fL	88,41	93,60

Untuk memahami penyebab fluktuasi nilai FP dan FN, dilakukan analisis terhadap karakteristik fitur data yang mengalami misklasifikasi pada model terbaik, yaitu Model F (rasio 70:30, n_estimators=100) dalam Tabel 4.14. Hasil analisis menunjukkan bahwa seluruh data yang tergolong False Positive maupun False Negative berada pada zona nilai hemoglobin yang sangat berdekatan. Masing-masing memiliki rata-rata hemoglobin antara 12,69 g/dL dan 12,60 g/dL dengan selisih hanya 0,09 g/dL. Kedua nilai ini jauh berbeda dari rata-rata hemoglobin kelas tidak anemia (14,81 g/dL) maupun kelas anemia (11,63 g/dL). Pada data uji secara keseluruhan, yang menunjukkan bahwa kesalahan klasifikasi tidak terjadi secara acak, melainkan terpusat pada kasus-kasus borderline yang nilai hematologinya berada di zona ambang batas.

Seluruh 23 data False Positive memiliki hemoglobin dalam rentang 12,0-13,3 g/dL, yang secara klinis berada sangat dekat dengan nilai ambang diagnosis anemia WHO. Meskipun label aktualnya adalah tidak anemia, nilai hemoglobin yang rendah dikombinasikan dengan nilai MCH yang rendah (mean 22,04 pg) menyebabkan model menginterpretasikannya sebagai kondisi anemia. Sebaliknya, 2 data False Negative memiliki hemoglobin dalam rentang 12,0-13,2 g/dL yang tergolong tinggi untuk pasien anemia. Sehingga model cenderung mengklasifikasinya sebagai tidak anemia.

Kondisi ini menjelaskan mengapa nilai FP dan FN berfluktuasi antar model dan antar rasio. Ketika n_estimators bertambah, batas keputusan model menjadi lebih halus sehingga sebagian kasus borderline berhasil diprediksi dengan benar. Namun karena karakteristik fitur FP dan FN berada pada zona nilai yang hampir identik, tidak semua kesalahan dapat diperbaiki sekaligus. Perbaikan pada FN belum tentu diikuti perbaikan pada FP secara proporsional. Hal ini juga menjelaskan mengapa Model F meskipun memiliki FP yang lebih banyak dibandingkan beberapa model lain, tetap menjadi konfigurasi terbaik karena berhasil menekan FN hingga hanya 2 kasus, yang dalam konteks klinis merupakan capaian yang sangat signifikan mengingat kasus anemia yang tidak terdeteksi memiliki dampak lebih serius dibandingkan kasus tidak anemia yang salah terklasifikasi sebagai anemia.

Analisis lebih lanjut terhadap nilai precision menunjukkan bebeapa pola menarik. Secara keseluruhan, nilai precision pada seluruh model secara konsisten lebih rendah dibandingkan nilai recall. Kondisi ini bukan merupakan kelemahan model secara acak, melainkan mencerminkan karakteristik struktural dataset yang digunakan. Sampel tidak anemia dengan nilai hemoglobin borderline (12,03-13,3 g/dL) lebih banyak jumlahnya dibandingkan sampel anemia pada rentan yang sama. Sehingga, potensi terjadinya False Positive secara struktural lebih besar dibandingkan False Negative. Hal ini mengakibatkan nilai precision secara konsisten berada di bawah nilai recall pada seluruh variasi eksperimen. Terdapat pola trade-off yang semakin jelas antara precision dan recall seiring meningkatnya n_estimators, khususnya pada rasio 70:30. Selisih antara recall dan precision

melebar dari 0,0674 pada n_estimators=50 menjadi 0,1003 pada n_estimators=100. Hal ini menunjukkan bahwa semakin banyak pohon keputusan yang dibangun, model semakin sensitif dalam mendeteksi kelas anemia namun dengan konsekuensi peningkatan False Positive yang menekan nilai precision.

Hal menarik lainnya adalah bahwa model dengan precision tertinggi justru bukan model terbaik yang dipilih dalam penelitian ini. Model C (rasio 60:40, n_estimators=100) menghasilkan precision tertinggi sebesar 0,9176, lebih tinggi dibandingkan Model F (0,8889). Namun Model C memiliki False Negative sebanyak 14 kasus, jauh lebih banyak dibandingkan Model F yang hanya 2 kasus. Kondisi ini menunjukkan bahwa Model C bersifat lebih konservatif dalam memprediksi anemia, sehingga FP-nya lebih sedikit dan *precision*-nya lebih tinggi, namun dengan konsekuensi banyak kasus anemia yang tidak terdeteksi. Dalam konteks klasifikasi anemia sebagai domain kesehatan, model yang terlalu konservatif justru lebih berbahaya karena pasien anemia yang tidak terdeteksi berisiko tidak mendapatkan penanganan. Oleh karena itu, meskipun *precision* Model F lebih rendah, keunggulannya pada *recall* dan F1-Score menjadikannya konfigurasi yang lebih optimal secara klinis maupun statistik.

Berdasarkan keseluruhan hasil pengujian, dapat disimpulkan bahwa konfigurasi terbaik diperoleh pada rasio 70:30 dengan nilai *n_estimators* sebesar 100, yang menghasilkan nilai *accuracy* sebesar 0,9415 dan F1-score sebesar 0,9364. Pemilihan Model F sebagai konfigurasi terbaik didasarkan pada tiga pertimbangan utama. Pertama, Model F memiliki nilai *accuracy* tertinggi sebesar 0.9415 yang mencerminkan kemampuan klasifikasi terbaik secara keseluruhan.

Kedua, Model F memiliki nilai F1-Score tertinggi sebesar 0,9364 yang menunjukkan keseimbangan terbaik antara *precision* dan *recall*. Meskipun nilai *precision* Model F (0,8889) lebih rendah dibanding Model C (0,9176), hal ini diimbangi oleh nilai *recall* Model F yang secara signifikan lebih tinggi (0,9892 vs 0,9435), sehingga menghasilkan *F1-Score* yang lebih unggul. Ketiga, dalam konteks klasifikasi anemia sebagai domain kesehatan, nilai *recall* yang tinggi lebih diprioritaskan karena meminimalkan resiko kasus anemia yang tidak terdeteksi (false negative), yang dalam praktik klinis dapat berdampak lebih serius dibandingkan *false positive*.

Kombinasi ini menunjukkan keseimbangan yang baik antara precision dan recall, serta mampu memberikan performa klasifikasi yang optimal. Penggunaan rasio 70:30 selaras dengan penelitian Abdul-Jabbar et al., (2023) yang juga menggunakan pembagian 70% data latih dan 30% data uji pada model XGBoost kemudian mengevaluasi dengan precision, recall, dan F1-Score. Konfigurasi n_estimators=100 pada model XGBoost juga sejalan dengan memperoleh akurasi tertinggi 98% serta performa yang stabil. Meskipun demikian, model yang dibangun masih memiliki keterbatasan, ditunjukkan dengan masih adanya kesalahan klasifikasi pada beberapa eksperimen. Hal ini dapat disebabkan oleh keterbatasan jumlah data, distribusi data, maupun karakteristik fitur yang digunakan dalam proses pelatihan model.

4.4 Perbandingan Algoritma

Untuk memberikan konteks terhadap performa algoritma XGBoost yang dievaluasi dalam penelitian ini, dilakukan perbandingan dengan dua algoritma pembanding yaitu Logistic Regression dan Random Forest. Pemilihan kedua algoritma pembanding ini didasarkan pada perbedaan paradigma yang diwakili masing-masing algoritma. Logistic Regression sebagai representasi model linear yang bersifat sederhana. Random Forest sebagai representasi algoritma yang berbasis bagging yang merupakan pendekatan berbeda dari boosting yang digunakan XGBoost.

Perlu diperhatikan bahwa Random Forest merupakan algoritma yang dikembangkan lebih awal dibandingkan XGBoost. Algoritma ini pertama kali diperkenalkan oleh Breiman pada tahun 2001 dan menerapkan metode *bagging* dengan membangun banyak pohon keputusan secara independen dan paralel menggunakan subset data yang berbeda. Hasil prediksi dari seluruh pohon kemudian dikombinasikan melalui mekanisme *majority voting* untuk menghasilkan keputusan akhir (Breiman, 2001). Di sisi lain, XGBoost diperkenalkan oleh Chen dan Guestrin pada tahun 2016 sebagai pengembangan dari metode *gradient boosting*. Berbeda dengan Random Forest, XGBoost membangun pohon keputusan secara bertahap (*sequential*), di mana setiap pohon baru dirancang untuk memperbaiki kesalahan yang dihasilkan oleh pohon sebelumnya. Selain itu, algoritma ini dilengkapi dengan teknik regularisasi yang berfungsi untuk mengurangi risiko *overfitting* (Chen & Guestrin, 2016). Adapun Logistic Regression merupakan salah satu algoritma klasifikasi berbasis model linear yang

telah banyak diterapkan dalam bidang medis karena memiliki struktur yang sederhana serta hasil yang relatif mudah untuk diinterpretasikan.

Hasil perbandingan ketiga algoritma pada seluruh variasi rasio pembagian data disajikan pada Tabel 4.17 berikut.

Tabel 4.17 Perbandingan Hasil Evaluasi Algoritma pada Berbagai Rasio Pembagian Data

Rasio	Algoritma	Accuracy	F1-Score
60:40	XGBoost	0.9385	0.9304
	Logistic Regression	0.8910	0.8750
	Random Forest	0.9754	0.9720
70:30	XGBoost	0.9414	0.9364
	Logistic Regression	0.8969	0.8835
	Random Forest	0.9836	0.9813
80:20	XGBoost	0.9333	0.9260
	Logistic Regression	0.9052	0.8924
	Random Forest	0.9929	0.9918

Perbandingan dilakukan dengan menyamakan nilai n_estimators=100 dan random_state=7 pada seluruh algoritma berbasis pohon keputusan untuk memastikan kondisi eksperimen yang terkontrol. Sehingga perbedaan performa yang dihasilkan mencerminkan karakteristik masing-masing algoritma dan bukan perbedaan konfigurasi. XGBoost dalam penelitian ini sengaja dibatasi pada max_depth=4 dan learning_rate=0,1 untuk mencegah overfitting, sehingga kompleksitas modelnya lebih terkontrol. Pada XGBoost, max_depth=4 merupakan bagian dari konfigurasi eksperimen yang bertujuan mengontrol kompleksitas model sesuai paradigma boosting. Sementara pada Random Forest, max_depth dibiarkan pada nilai default tanpa batasan karena regularisasi pada algoritma bagging ditangani melalui mekanisme bootstrap sampling dan random feature selection,

sehingga pembatasan kedalaman pohon tidak diperlukan dan justru tidak sesuai dengan karakteristik algoritmanya.

Berdasarkan Tabel 4.17, terlihat bahwa Random Forest secara konsisten menghasilkan nilai *accuracy* dan F1-Score tertinggi diseluruh variasi rasio pembagian data. Hal ini dapat dijelaskan oleh perbedaan mekanisme ensemble yang digunakan. Random Forest membangun pohon keputusan secara penuh tanpa batasan kedalaman. Sehingga, pada dataset yang bersih dan terstruktur seperti dataset hematologi yang digunakan dalam penelitian ini, Random Forest mampu menangkap pola klasifikasi dengan sangat presisi. Di sisi lain, Logistic Regression secara konsisten menghasilkan performa terendah di semua rasio. Hal ini disebabkan oleh asumsi linearitas yang menjadi dasar algoritma tersebut. Logistic Regression mengasumsikan hubungan linear antara fitur input dan label output. Sementara hubungan antar parameter hematologi dalam menentukan kondisi anemia bersifat non-linear dan melibatkan interaksi kompleks antar fitur (Hb, MCV, MCH, MCHC). Algoritma berbasis decision tree seperti XGBoost dan Random Forest mampu menangkap hubungan non-linear ini secara alami sehingga menghasilkan performa yang lebih tinggi.

4.5 Integrasi Sains dan Islam

Berdasarkan hasil pengujian yang telah dilakukan, model klasifikasi menggunakan algoritma XGBoost menunjukkan kemampuan yang baik dalam mengidentifikasi kondisi anemia berdasarkan data hematologi. Hasil ini menunjukkan bahwa pemanfaatan teknologi *machine learning* dapat digunakan

sebagai alat bantu dalam proses deteksi dini suatu penyakit, sehingga dapat mendukung pengambilan keputusan dalam bidang kesehatan.

Hasil penelitian ini dapat dimaknai secara mendalam melalui kerangka Maqashid Al-Syariah, yaitu tujuan-tujuan pokok syariat Islam dalam menjaga kemaslahatan manusia. Secara khusus, penelitian ini berkaitan erat dengan aspek *hifz an-nafs* (menjaga keselamatan jiwa) yang merupakan salah satu dari lima penjagaan utama dalam Maqashid Al-Syariah. Model klasifikasi XGBoost yang dikembangkan dalam penelitian ini berkontribusi pada *hifz an-nafs* melalui kemampuannya mendeteksi 98,92% kasus anemia secara akurat hanya dengan parameter hematologi dasar. Sehingga, berpotensi mendukung deteksi dini anemia terutama di fasilitas kesehatan primer yang memiliki keterbatasan sumber daya. Deteksi dini yang akurat memungkinkan penanganan yang lebih cepat dan tepat, sehingga risiko komplikasi serius akibat anemia yang tidak terdeteksi dapat diminimalkan.

Hal ini juga sejalan dengan kaidah fiqh yang telah disebutkan sebelumnya, yaitu *al-dhararu yuzalu* (kemudharatan harus dihilangkan). Anemia yang tidak terdeteksi merupakan bentuk kemudharatan nyata, terutama bagi kelompok rentan seperti ibu hamil dan anak-anak. Dengan demikian, penerapan teknologi machine learning dalam deteksi anemia bukan sekadar inovasi ilmiah. Melainkan merupakan bentuk nyata dari upaya menghilangkan kemudharatan melalui pemanfaatan ilmu pengetahuan yang sejalan dengan nilai-nilai syariat Islam.

Dalam perspektif Islam, pemanfaatan ilmu pengetahuan dan teknologi merupakan bagian dari ikhtiar manusia dalam menjaga amanah berupa kesehatan

yang telah diberikan oleh Allah *Subhanahu Wa Ta'ala*. Hasil penelitian ini mencerminkan bahwa penggunaan model klasifikasi tidak dimaksudkan untuk menggantikan peran tenaga medis, melainkan sebagai sarana pendukung dalam meningkatkan ketepatan analisis terhadap kondisi kesehatan seseorang. Hal ini sejalan dengan konsep bahwa manusia diperintahkan untuk berusaha secara optimal, sementara hasil akhir tetap berada dalam ketentuan Allah *Subhanahu WaTa'ala*. Sebagaimana dalam hadits Rasulullah *Shalallahu Alaihi Wassalam* pada kitab Shahih Bukhari (Musnad Ahmad:17726) :

تَدَاوَوْا فَإِنَّ اللَّهَ لَمْ يَضَعْ دَاءً إِلَّا وَضَعَ لَهُ دَوَاءً

"Berobatlah kalian, karena sesungguhnya Allah tidak menurunkan suatu penyakit melainkan Dia juga menurunkan obatnya." (HR. Abu Dawud)

Hadits tersebut menunjukkan bahwa setiap penyakit memiliki potensi untuk diatasi, sehingga manusia dianjurkan untuk melakukan usaha, termasuk melalui pemanfaatan teknologi dan ilmu pengetahuan. Dalam hal ini, penggunaan algoritma XGBoost untuk mendeteksi anemia dapat dipandang sebagai bagian dari ikhtiar ilmiah dalam membantu proses diagnosis secara lebih cepat dan akurat. Islam juga mendorong umatnya untuk menggunakan akal dalam memahami fenomena yang terjadi. Dalam hal ini, penggunaan algoritma XGBoost dalam menganalisis data kesehatan merupakan bentuk pemanfaatan akal dan ilmu pengetahuan dalam mengolah informasi secara sistematis dan objektif.

Dengan demikian, hasil penelitian ini menunjukkan bahwa penerapan teknologi *machine learning* dalam deteksi anemia tidak hanya memberikan

kontribusi dalam bidang ilmu pengetahuan. Tetapi juga sejalan dengan nilai-nilai Islam dalam menjaga kesehatan, menggunakan akal secara optimal, serta melakukan ikhtiar dalam mencegah kemudharatan. Oleh karena itu, integrasi antara sains dan Islam dalam penelitian ini dapat memperkuat pemanfaatan teknologi sebagai sarana untuk mencapai kemaslahatan manusia.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma XGBoost mampu digunakan dengan baik dalam melakukan klasifikasi anemia berdasarkan data hematologi. Model yang dibangun menunjukkan performa yang cukup tinggi pada berbagai skenario pengujian, yang ditunjukkan melalui nilai *accuracy, precision, recall*, dan F1-score yang relatif baik pada setiap variasi rasio pembagian data dan nilai *n_estimators*. Hasil pengujian menunjukkan bahwa variasi nilai *n_estimators* berpengaruh terhadap kinerja model. Semakin besar nilai *n_estimators*, performa model cenderung mengalami peningkatan, khususnya pada metrik recall dan F1-score. Hal ini menunjukkan bahwa model menjadi lebih mampu dalam mendeteksi data anemia secara lebih akurat, meskipun pada beberapa kondisi terjadi penurunan kecil pada nilai precision yang menunjukkan adanya *trade-off* antar metrik evaluasi.

Analisis lebih lanjut terhadap data yang mengalami misklasifikasi pada model terbaik menunjukkan bahwa seluruh kesalahan klasifikasi terpusat pada kasus-kasus borderline dengan nilai hemoglobin di rentang 12,0–13,3 g/dL, yang berada di zona ambang batas diagnosis anemia secara klinis. Kondisi ini mengindikasikan bahwa kesalahan yang terjadi bukan disebabkan oleh kelemahan model, melainkan oleh karakteristik inheren data yang secara klinis pun sulit dibedakan. Hal ini sekaligus membuktikan bahwa XGBoost tidak hanya bergantung

pada hemoglobin sebagai satu-satunya penentu klasifikasi, melainkan oleh karakteristik inheren data yang secara klinis pun sulit dibedakan. Hal ini sekaligus membuktikan bahwa XGBoost tidak hanya bergantung pada hemoglobin sebagai satu-satunya penentu klasifikasi, melainkan mempertimbangkan interaksi antar fitur hematologi secara bersamaan dalam mengambil keputusan.

Selain itu, variasi rasio pembagian data juga mempengaruhi performa model. Berdasarkan hasil pengujian, rasio 70:30 memberikan hasil terbaik dibandingkan rasio lainnya, dengan nilai *accuracy* sebesar 0,9415 dan F1-score sebesar 0,9364 pada n_estimators sebesar 100. Hal ini menunjukkan bahwa proporsi data latih yang cukup besar mampu meningkatkan kemampuan model dalam mempelajari pola data secara optimal. Secara keseluruhan, model terbaik dalam penelitian ini diperoleh pada kombinasi rasio 70:30 dengan n_estimators sebesar 100, yang memberikan keseimbangan yang baik antara precision dan recall. Dengan demikian, penelitian ini membuktikan bahwa XGBoost mampu mempertahankan performa klasifikasi yang tinggi meskipun hanya menggunakan empat parameter hematologi dasar, dengan *accuracy* terbaik 0,9415 dan recall 0,9892. Sehingga berpotensi diterapkan sebagai alat bantu deteksi dini anemia di fasilitas kesehatan primer yang memiliki keterbatasan akses terhadap pemeriksaan hematologi lanjutan.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya maupun implementasi sistem yang lebih lanjut.

1. Penelitian selanjutnya dapat mengembangkan model dengan menggunakan algoritma lain sebagai pembanding, seperti Random Forest, Support Vector Machine (SVM), atau metode *deep learning*, sehingga dapat diketahui perbandingan kinerja antar algoritma dalam kasus klasifikasi anemia.
2. Proses optimasi parameter (*hyperparameter tuning*) dapat dilakukan secara lebih mendalam menggunakan metode seperti Grid Search atau Random Search untuk memperoleh kombinasi parameter yang lebih optimal dibandingkan pengaturan parameter secara manual.
3. Sistem yang dikembangkan dalam penelitian ini dapat diimplementasikan dalam bentuk aplikasi berbasis web atau mobile, sehingga dapat dimanfaatkan secara lebih luas sebagai alat bantu dalam proses deteksi dini anemia.
4. Penelitian ini diharapkan dapat dikembangkan lebih lanjut dengan mengintegrasikan sistem ke dalam lingkungan layanan kesehatan, sehingga dapat membantu tenaga medis dalam proses pengambilan keputusan secara lebih cepat dan akurat.

DAFTAR PUSTAKA

A Ilemobayo, J., Durodola, O., Alade, O., J Awotunde, O., T Olanrewaju, A., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., Ifeanyi, A., E Odezuligbo, I., & E Edu, O. (2024). Hyperparameter Tuning in Machine Learning: A Comprehensive Review. Journal of Engineering Research and Reports, 26(6), 388–395. https://doi.org/10.9734/jerr/2024/v26i61188

Abdul-Jabbar, S. S., Farhan, A. K., & Luchinin, A. S. (2023). Comparative Study of Anemia Classification Algorithms for International and Newly CBC Datasets. International Journal of Online and Biomedical Engineering (iJOE), 19(06), 141–157. https://doi.org/10.3991/ijoe.v19i06.38157

Abi Ishaq Al-Syatibi. (2003). Al-Muwafaqat (Abdullah Daraz, Ed.; Vol. 2). Wizarat al-Syu'un al-Islamiyyah al Awqaf (Kementerian Urusan Islam Arab Saudi). (Original work published 1423)

Abu Bakr Al-Ahdali Al-Yamani. (2004). Al-Faraidul Bahiyyah. Madrasah Hidayatul Mubtadi-ien (Darul Mubtadi-ien).

Airlangga, G. (2024). Leveraging Machine Learning for Accurate Anemia Diagnosis Using Complete Blood Count Data. Indonesian Journal of Artificial Intelligence and Data Mining, 7(2), 318. https://doi.org/10.24014/ijaidm.v7i2.29869

Aldahiri, A., Alrashed, B., & Hussain, W. (2021). Trends in Using IoT with Machine Learning in Health Prediction System. Forecasting, 3(1), 181–206. https://doi.org/10.3390/forecast3010012

Amalia, I. (n.d.). Anemia Classification with Clinical Feature Engineering and SHAP Interpretation. 2025, 9(5). https://doi.org/https://doi.org/10.30871/jaic.v9i5.10912

An, R., Huang, Y., Man, Y., Valentine, R. W., Kucukal, E., Goreke, U., Sekyonda, Z., Piccone, C., Owusu-Ansah, A., Ahuja, S., Little, J. A., & Gurkan, U. A. (2021). Emerging point-of-care technologies for anemia detection. Lab on a Chip, 21(10), 1843–1865. https://doi.org/10.1039/D0LC01235A

Ashisha, G. R., Mary, X. A., Kanaga, E. G. M., Andrew, J., & Eunice, R. J. (2024). Random Oversampling-Based Diabetes Classification via Machine Learning Algorithms. International Journal of Computational Intelligence Systems, 17(1), 270. https://doi.org/10.1007/s44196-024-00678-3

Attaqy, F. C., Kalsum, U., & Syukri, M. (2022). Determinan Anemia pada Wanita Usia Subur (15-49 Tahun) Pernah Hamil di Indonesia: Analisis Data Riskesdas 2018. 6(1).

Bakasa, W., & Viriri, S. (2023). Stacked ensemble deep learning for pancreas cancer classification using extreme gradient boosting. Frontiers in Artificial Intelligence, 6, 1232640. https://doi.org/10.3389/frai.2023.1232640

Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective Heart Disease Prediction Using Machine Learning Techniques. Algorithms, 16(2), 88. https://doi.org/10.3390/a16020088

Botlagunta, M., Botlagunta, M. D., Myneni, M. B., Lakshmi, D., Nayyar, A., Gullapalli, J. S., & Shah, M. A. (2023). Classification and diagnostic prediction of breast cancer metastasis on clinical data using machine learning algorithms. Scientific Reports, 13(1), 485. https://doi.org/10.1038/s41598-023-27548-w

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

Çakmak, Y. (2025). Machine Learning Approaches for Enhanced Diagnosis of Hematological Disorders. Computational Systems and Artificial Intelligence, 1(1), 8–14. https://doi.org/10.69882/adba.csai.2025072

Celkan, T. (2020). Hemogram bize neler söyler? Türk Pediatri Arşivi. https://doi.org/10.14744/TurkPediatriArs.2019.76301

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794. https://doi.org/10.1145/2939672.2939785

Chopra, V. K., & Anker, S. D. (2020). Anaemia, iron deficiency and heart failure in 2020: Facts and numbers. ESC Heart Failure, 7(5), 2007–2011. https://doi.org/10.1002/ehf2.12797

Darshan, B. S. D., Sampathila, N., Bairy, G. M., Prabhu, S., Belurkar, S., Chadaga, K., & Nandish, S. (2025). Differential diagnosis of iron deficiency anemia from aplastic anemia using machine learning and explainable Artificial Intelligence utilizing blood attributes. Scientific Reports, 15(1), 505. https://doi.org/10.1038/s41598-024-84120-w

Dawidowski, J., & Pietrzak, A. (2022). Rare causes of anemia in liver diseases. Advances in Clinical and Experimental Medicine, 31(5), 567–574. https://doi.org/10.17219/acem/145984

Gómez, J. G., Parra Urueta, C., Álvarez, D. S., Hernández Riaño, V., & Ramirez-Gonzalez, G. (2025a). Anemia Classification System Using Machine Learning. Informatics, 12(1), 19. https://doi.org/10.3390/informatics12010019

Gómez, J. G., Parra Urueta, C., Álvarez, D. S., Hernández Riaño, V., & Ramirez-Gonzalez, G. (2025b). Anemia Classification System Using Machine

Learning. Informatics, 12(1), Article 1. https://doi.org/10.3390/informatics12010019

Gómez-Pastora, J., Weigand, M., Kim, J., Palmer, A. F., Yazer, M., Desai, P. C., Zborowski, M., & Chalmers, J. J. (2022). Potential of cell tracking velocimetry as an economical and portable hematology analyzer. Scientific Reports, 12(1), 1692. https://doi.org/10.1038/s41598-022-05654-5

Hain, D., Bednarski, D., Cahill, M., Dix, A., Foote, B., Haras, M. S., Pace, R., & Gutiérrez, O. M. (2023). Iron-Deficiency Anemia in CKD: A Narrative Review for the Kidney Care Team. Kidney Medicine, 5(8), 100677. https://doi.org/10.1016/j.xkme.2023.100677

Hidayah, Y. N., Daniati, E., & Azis, M. A. (2025). Penyeimbangan Data Untuk Klasifikasi Jenis Anemia Menggunakan Smote Dan Pengembangan Diagnostik Streamlit.

Irfan, B., Khleif, A., Badarneh, J., Abutaqa, J., Allam, A., Kweis, S., Tarab, B., Abu Shammala, A., Nasser, E., Shweiki, S., Wajahath, M., & Padela, A. (2025). Considering Islamic Frameworks to Infectious Disease Prevention. Open Forum Infectious Diseases, 12(10), ofaf011. https://doi.org/10.1093/ofid/ofaf011

Jhaveri, R. H., Revathi, A., Ramana, K., Raut, R., & Dhanaraj, R. K. (2022). A Review on Machine Learning Strategies for Real-World Engineering Applications. Mobile Information Systems, 2022, 1–26. https://doi.org/10.1155/2022/1833507

Mehdary, A., Chehri, A., Jakimi, A., & Saadane, R. (2024). Hyperparameter Optimization with Genetic Algorithms and XGBoost: A Step Forward in Smart Grid Fraud Detection. Sensors, 24(4), 1230. https://doi.org/10.3390/s24041230

Mienye, I. D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access, 10, 99129–99149. https://doi.org/10.1109/ACCESS.2022.3207287

Miglio, A., Valente, C., & Guglielmini, C. (2023). Red Blood Cell Distribution Width as a Novel Parameter in Canine Disorders: Literature Review and Future Prospective. Animals, 13(6), 985. https://doi.org/10.3390/ani13060985

Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Le, H. V., Tran, V. Q., Prakash, I., & Pham, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. Mathematical Problems in Engineering, 2021(1), 4832864. https://doi.org/10.1155/2021/4832864

Noorunnahar, M., Chowdhury, A. H., & Mila, F. A. (2023). A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual

rice production in Bangladesh. PLOS ONE, 18(3), e0283452. https://doi.org/10.1371/journal.pone.0283452

Omotehinwa, T. O., & Oyewola, D. O. (2023). Hyperparameter Optimization of Ensemble Models for Spam Email Detection. Applied Sciences, 13(3), 1971. https://doi.org/10.3390/app13031971

Putri, N. A. (2025). A Comparative Analysis of Machine Learning Classifier of Anemia Diagnosis Based on Complete Blood Count (CBC) Data. IJIIS: International Journal of Informatics and Information Systems, 8(4), 188–200. https://doi.org/10.47738/ijiis.v8i4.286

Ramzan, M., Sheng, J., Saeed, M. U., Wang, B., & Duraihem, F. Z. (2024). Revolutionizing anemia detection: Integrative machine learning models and advanced attention mechanisms. Visual Computing for Industry, Biomedicine, and Art, 7(1), 18. https://doi.org/10.1186/s42492-024-00169-4

Safiri, S., Kolahi, A.-A., Noori, M., Nejadghaderi, S. A., Karamzad, N., Bragazzi, N. L., Sullman, M. J. M., Abdollahi, M., Collins, G. S., Kaufman, J. S., & Grieger, J. A. (2021). Burden of anemia and its underlying causes in 204 countries and territories, 1990–2019: Results from the Global Burden of Disease Study 2019. Journal of Hematology & Oncology, 14(1), 185. https://doi.org/10.1186/s13045-021-01202-2

Sahin, E. K. (2020). Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost, gradient boosting machine, and random forest. SN Applied Sciences, 2(7), 1308. https://doi.org/10.1007/s42452-020-3060-1

Saxena, A., & Chandra, S. (Eds.). (2021). Artificial Intelligence and Machine Learning in Healthcare. Springer. https://doi.org/10.1007/978-981-16-0811-7

Schweta, N., & Pande, S. D. (2023). Prediction of Anemia using various Ensemble Learning and Boosting Techniques. EAI Endorsed Transactions on Pervasive Health and Technology, 9. https://doi.org/10.4108/eetpht.9.4197

Seo, I.-H., & Lee, Y.-J. (2022). Usefulness of Complete Blood Count (CBC) to Assess Cardiovascular and Metabolic Diseases in Clinical Settings: A Comprehensive Literature Review. Biomedicines, 10(11), 2697. https://doi.org/10.3390/biomedicines10112697

Shwartz-Ziv, R., & Armon, A. (2021). Tabular Data: Deep Learning is Not All You Need (arXiv:2106.03253). arXiv. https://doi.org/10.48550/arXiv.2106.03253

Sumarno, Haddade, H., & Damis, R. (2022). Wawasan Al-Qur'an Tentang Kesehatan. Jurnal Pendidikan Islam, 8(2), 293–304. https://doi.org/10.37286/ojs.v8i2.166

Sundararajan, S., & Rabe, H. (2021). Prevention of iron deficiency anemia in infants and toddlers. Pediatric Research, 89(1), 63–73. https://doi.org/10.1038/s41390-020-0907-5

Terven, J., Cordova-Esparza, D.-M., Romero-González, J.-A., Ramírez-Pedraza, A., & Chávez-Urbiola, E. A. (2025). A comprehensive survey of loss functions and metrics in deep learning. Artificial Intelligence Review, 58(7), 195. https://doi.org/10.1007/s10462-025-11198-7

Vu, T., Vo, T., Thai, H. T., & Nguyen, H. (2021). Efficient machine learning models for prediction of concrete strengths. https://doi.org/10.26181/5fb33e9fdeb85

Wiens, M., Verone-Boyle, A., Henscheid, N., Podichetty, J. T., & Burton, J. (2025). A Tutorial and Use Case Example of the eXtreme Gradient Boosting (XGBoost) Artificial Intelligence Algorithm for Drug Development Applications. Clinical and Translational Science, 18(3), e70172. https://doi.org/10.1111/cts.70172

Xing, X., Liu, H., Chen, C., & Li, J. (2021). Fairness-Aware Unsupervised Feature Selection (arXiv:2106.02216). arXiv. https://doi.org/10.48550/arXiv.2106.02216

Yıldız, A. Y., & Kalayci, A. (2025). Gradient Boosting Decision Trees on Medical Diagnosis over Tabular Data. 2025 IEEE International Conference on AI and Data Analytics (ICAD), 1–8. https://doi.org/10.1109/ICAD65464.2025.11114069

Yunos, A. S., & Hamdan, M. N. (2024). Artificial Intelligence In Healthcare: Analysis From The Perspective Of Maqasid Al-Shariah. Journal of Information System and Technology Management, 9(35), 58–74. https://doi.org/10.35631/JISTM.935004