IMPLEMENTASI METODE RANDOM FOREST UNTUK KLASIFIKASI ANEMIA

SKRIPSI

Oleh :
MARSYA ADELINE DESWINTARANI
NIM. 220605110065

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

IMPLEMENTASI METODE RANDOM FOREST UNTUK KLASIFIKASI ANEMIA

SKRIPSI

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

Oleh:
MARSYA ADELINE DESWINTARANI
NIM. 220605110065

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

HALAMAN PERSETUJUAN

IMPLEMENTASI METODE *RANDOM FOREST* UNTUK KLASIFIKASI ANEMIA

SKRIPSI

Oleh :
MARSYA ADELINE DESWINTARANI
NIM. 220605110065

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 25 Juni 2026

Pembimbing I,	Pembimbing II,
Prof. Dr. Suhartono, S.Si., M.Kom	A'la Syauqi, M.Kom
NIP. 19680519 200312 1 001	NIP. 19771201 200801 1 007

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

Supriyono, M. Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

IMPLEMENTASI METODE *RANDOM FOREST* UNTUK KLASIFIKASI ANEMIA

SKRIPSI

Oleh :
MARSYA ADELINE DESWINTARANI
NIM. 220605110065

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Tanggal: 25 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: Okta Qomaruddin Aziz, M.Kom NIP. 19911019 201903 1 013	( )
Anggota Penguji I	: Khadijah Fahmi Hayati Halle, M.Kom NIP. 19900626 202203 1 002	( )
Anggota Penguji II	: Prof. Dr. Suhartono, S.Si., M.Kom NIP. 19680519 200312 1 001	( )
Anggota Penguji III	: A'la Syauqi, M.Kom NIP. 19771201 200801 1 007	( )

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	: Marsya Adeline Deswintarani
NIM	: 220605110065
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Implementasi Metode Random Forest Uintuk Klasifikasi Anemia

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, 30 Juni 2026
Yang membuat pernyataan,

Marsya Adeline Deswintarani
NIM.220605110065

MOTTO

“Debug your life, optimize your soul”

-me

HALAMAN PERSEMBAHAN

Karya ini penulis persembahkan kepada Sujud syukur penulis persembahkan ke hadirat Allah SWT, Sang Pemilik Ilmu, atas segala ketetapan, kemudahan, dan kasih sayang-Nya yang tak terhingga hingga langkah kaki ini sampai pada akhir perjuangan. Karya sederhana ini penulis dedikasikan dengan penuh cinta kepada:

Orang Tua Penulis,
Yang dukungannya menjadi sandaran paling kokoh dan kehadirannya menjadi penyejuk di tengah hiruk-pikuk perjuangan akademik ini.

Dosen Pembimbing dan Dosen Penguji,
Terimakasih atas limpahan ilmu, bimbingan serta kesabaran yang diberikan selama proses penyusunan skripsi.

Untuk Diri Penulis Sendiri,
Terima kasih karena telah menjadi kuat. Terima kasih telah memaafkan diri saat lelah, namun menolak untuk menyerah.

Semoga keberhasilan ini menjadi pembuka gerbang keberkahan, membawa manfaat bagi sesama, dan menjadi langkah awal menuju masa depan yang diridhai-Nya.

KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Alhamdulillah, segala puji dan syukur penulis panjatkan ke hadirat Allah Subhanahu Wa Ta'ala atas limpahan rahmat, karunia, dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul "Implementasi Metode Random Forest Untuk Klasifikasi Anemia" dengan baik. Sholawat serta salam semoga selalu tercurah kepada Nabi Muhammad Shallallahu Alaihi Wasallam, yang menjadi teladan bagi penulis dalam menuntut ilmu dalam setiap langkah perjuangan ini.

Penulis menyadari sepenuhnya bahwa skripsi ini bukanlah hasil jerih payah sendiri, melainkan sebuah simfoni yang tersusun dari doa, dukungan, dan bimbingan berbagai pihak. Dengan segala kerendahan hati, penulis menyampaikan apresiasi setinggi-tingginya 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. Dr. Ir. Yunifa Miftachul Arif, S.ST., M.T., SMIEEE. selaku Dosen Wali yang telah dengan sabar menuntun perjalanan akademik penulis sejak awal perkuliahan dengan penuh dedikasi.
5. Prof. Dr. Suhartono, M.Kom. selaku Dosen Pembimbing I, atas segala bimbingan, perspektif ilmu yang mencerahkan, dan kesabaran dalam

mengarahkan penulis menyelesaikan penelitian ini.

6. A'la Syauqi, M.Kom. selaku Dosen Pembimbing II, yang telah banyak memberikan motivasi, kritik membangun, serta dukungan teknis yang sangat berarti.
7. Okta Qomaruddin Aziz, M.Kom. selaku Ketua Penguji, terima kasih atas ketelitian, saran yang konstruktif, serta kesediaan waktu dalam membekali penulis melalui berbagai diskusi dan latihan demi kualitas skripsi yang lebih baik.
8. Khadijah Fahmi Hayati Holle, M.Kom. selaku Anggota Penguji I, atas segala masukan berharga yang telah menyempurnakan setiap celah dalam penelitian ini
9. Segenap Dosen dan Civitas Academica Teknik Informatika, yang telah membekali penulis dengan ilmu pengetahuan dan pengalaman berharga selama masa studi.
10. Kedua orang tua, yang menjadi rumah bagi segala keluh kesah penulis. Doa kalian adalah pelita dalam gelapnya proses revisi. Terima kasih telah menjadi alasan penulis untuk tetap berjuang.
11. Rekan-Rekan Teknik Informatika Angkatan 2022, INFINITY '22, terima kasih atas solidaritas dan kebersamaan yang membuat masa-masa sulit terasa lebih ringan.
12. Diri sendiri. Terima kasih karena telah percaya pada kemampuan diri sendiri, sudah bekerja keras tanpa kenal lelah, dan karena telah memilih untuk tidak pernah berhenti berjuang.

Penulis menyadari skripsi ini masih jauh dari kata sempurna dan memiliki banyak keterbatasan. Oleh karena itu, penulis sangat terbuka dalam menerima kritik maupun saran yang membangun demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat nyata bagi pembaca dan pengembangan ilmu pengetahuan, sehingga tidak hanya menjadi syarat kelulusan.

Wassalamu'alaikum Wr, Wb.

Malang, 30 Juni 2026

Penulis

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

ABSTRAK

Deswintarani, Marsya Adeline. 2026. **Implementasi Random Forest Untuk Klasifikasi Anemia.** 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) A'la Syauqi, M.Kom.

Kata kunci: Anemia, Random Forest, *GridSearchCV*, Hemoglobin, Klasifikasi.

Anemia merupakan masalah kesehatan global yang memengaruhi lebih dari 1,6 miliar orang, dengan prevelansi yang masih tiinggi di Indonesia, terutama pada remaja dan ibu hamil. Penyakit ini disebabkan oleh rendahnya kadar hemoglobin, sehingga diperlukan deteksi dini yang akurat guna menghindari risiko komplikasi berat. Penelitian ini mengimplementasikan algoritma Random Forest untuk mengklasifikasikan status anemiadengan memanfatkan teknik *GridSearchCV* guna mencari parameter model yang optimal. Pengujian dilakukan melalui dua skenario: Model A (seluruh fitur klinis) dan Model B (tanpa fitur hemoglobin). Hasil penelitian membuktikan bahwa fitur hemoglobin merupakan prediktor paling krusial; tanpa fitur ini, akurasi model menurun drastis ke rentang 61,75% - 69,23%. Sebaliknya, Model A2 menunjukkan peforma optimal pada rasio data 80:20 dengan akurasi mencapai 94,39%. Analisis confusion matrix menunjukkan bahwa ketiadaan fitur hemoglobin menyebabkan bias pada kelas mayoritas yang meningkatkan angka false negative. Penelitian ini menyimpulkan bahwa penggunaan Random Forest dengan integrasi fitur klinis yang lengkap dapat memberikan diagnosa yang presisi dan stabil untuk deteksi anemia.

ABSTRACT

Deswintarani, Marsya Adeline. 2026. **The Implementation of the Random Forest to Classify Anemia.** Thesis. Informatics Engineering Study Program. Faculty of Sience and Technology Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisor: (I) Prof. Dr. Suhartono, S.Si., M.Kom. (II) A'la Syauqi, M.Kom.

Keywords: Anemia, Random Forest, GridSearchCV, Hemoglobin, Classification.

Anemia is a global health problem affecting more than 1.6 billion people worldwide, with a high prevalence in Indonesia, particularly among adolescents and pregnant women. This condition results from low hemoglobin levels and requires accurate early detection to prevent severe complications. This research employed the Random Forest algorithm to classify anemia status, using GridSearchCV to identify optimal model parameters. The researcher conducted two scenarios: Model A (using all clinical features) and Model B (excluding the hemoglobin feature). The research results show that the hemoglobin feature is the most critical predictor of anemia. Without this feature, the model's accuracy declined substantially, ranging from 61.75% to 69.23%. In contrast, Model A2 showed optimal performance with an 80:20 data ratio and an accuracy of 94.39%. Furthermore, confusion matrix analysis revealed that the absence of the hemoglobin feature caused a bias toward the majority class, resulting in a higher rate of false negatives. The research concludes that the Random Forest algorithm integrated with comprehensive clinical features can provide accurate and robust diagnostic performance for anemia detection.

البحث مستخلص

ديسوينتاراني، مارشا أدلين. 2026. تطبيق غابة عشوائية لتصنيف فقر الدم. البحث الجامعي. قسم الهندسة المعلوماتية، كلية العلوم والتكنولوجيا بجامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف الأول: أ. د. سوهارتونو، الماجستير؛ المشرف الثاني: أعلا شوقي، الماجستير.

الكلمات الرئيسية: فقر دم، غابة عشوائية، *GridSearchCV*، خضاب الدم، تصنيف.

يُعد فقر الدم مشكلة صحية عالمية تؤثر على أكثر من 1.6 مليار شخص، مع معدل انتشار لا يزال مرتفعًا في إندونيسيا، خصوصًا بين المراهقين والنساء الحوامل. يحدث هذا المرض بسبب انخفاض مستوى خضاب الدم أو الهيموغلوبين، لذلك من الضروري الكشف المبكر بدقة لتجنب مخاطر المضاعفات الخطيرة. نفذ هذا البحث خوارزمية الغابة العشوائية لتصنيف حالة فقر الدم باستخدام تقنية *GridSearchCV* للبحث عن أفضل معلمات للنموذج. تم اختبار ذلك من خلال سيناريوهين: النموذج أ (جميع الخصائص السريرية) والنموذج ب (بدون خاصية الهيموغلوبين). أثبتت نتائج البحث أن ميزة الهيموغلوبين هي المعيار الأكثر أهمية؛ بدون هذه الميزة، تنخفض دقة النموذج بشكل كبير إلى النطاق 61.75٪ – 69.23٪. بالمقابل، أظهر النموذج أ2 أداءً مثاليًا عند نسبة بيانات 80:20 مع دقة تصل إلى 94.39٪. تحليل مصفوفة الالتباس أظهر أن غياب ميزة الهيموغلوبين يؤدي إلى انحياز نحو الفئة الكبرى مما يزيد من معدل السلبيات الكاذبة. توصل هذا البحث إلى أن استخدام الغابات العشوائية مع دمج الميزات السريرية الكاملة يمكن أن يوفر تشخيصًا دقيقًا ومستقرًا للكشف عن فقر الدم.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Salah satu masalah kesehatan yang masih terjadi hingga saat ini di berbagai belahan dunia adalah anemia. Penyakit ini terjadi karena kadar hemoglobin dalam darah terlalu rendah, sehingga darah kurang mampu untuk mengangkut oksigen ke seluruh tubuh. Hal tersebut dapat mengakibatkan kelelahan yang terus menerus, sulit fokus, atau bahkan meningkatkan risiko berat jika tidak segera ditangani. Menurut organisasi kesehatan dunia, WHO, lebih dari 1,6 miliar orang diseluruh dunia mengalami anemia, terutama pada anak-anak dan wanita di usia subur (WHO, 2021).

Di Indonesia, anemia masih menjadi masalah kesehatan yang masih sering terjadi, terutama pada remaja perempuan dan ibu hamil. Menurut data Riskesdas tahun 2018, sekitar 48,9% ibu hamil mengalami anemia sedangkan pada tahun 2020 sekitar 48,9% anemia terjadi pada usia remaja (Riskesdas, 2020). Sementara itu, berdasarkan laporan Survei Kesehatan Indonesia (SKI) 2023, prevalensi anemia pada remaja dalam rentang usia 15-24 tahun sekitar 15,5% yakni remaja putri sebesar 18% dan remaja putra sebesar 14,4%. Meskipun, anemia mengalami penurunan dibandingkan 5 tahun yang lalu, namun 15,5% masih tergolong tinggi. Menurut WHO, populasi terjadi anemia sebesar 10% termasuk tinggi (Kemenkes, 2023). Maka, deteksi anemia sejak dini sangat penting dilakukan agar pengobatan dapat dilakukan lebih cepat dan tepat.

Diagnosis anemia biasanya dilakukan dengan pemeriksaan laboratorium, seperti mengukur kadar Hemoglobin (Hb), *Mean Corpuscular Volume* (MCV), *Mean Corpuscular Hemoglobin* (MCH), *Mean Corpuscular Hemoglobin Concentration* (MCHC). Namun, dalam menafsirkan hasil pemeriksaan tersebut, tenaga medis masih harus melakukan analisis secara manual. Proses tersebut memerlukan waktu, biaya, dan berisiko adanya kesalahan karena faktor subyektif. Dengan semakin banyaknya data pasien, diperlukan sistem yang dapat melakukan analisis data secara otomatis dan akurat.

Kemajuan teknologi informasi, terutama dibidang *machine learning*, telah memberikan peluang besar dalam penggunaan data medis. Beberapa penelitian menunjukan bahwa machine learning sangat efektif dalam mengklasifikasikan penyakit anemia. Misalnya, studi yang dilakukan oleh Kusumawati, menunjukkan bahwa algoritma *Random Forest* mampu mengklasifikasikan anemia menggunakan data *Complete Blood Count* (CBC) (Kusumawati et al., 2023). Adapun penelitian yang dilakukan oleh Wijaya, mengungkapkan bahwa metode *Random Forest* dapat mengklasifikasikan penyakit berdasarkan data laboratorium (Wijaya, A.P., 2025).

Random Forest adalah algoritma dari *machine learning* yang menggunakan metode *ensemble learning* dengan menggabungkan banyak pohon keputusan (*decision tree*) untuk mengklasifikasikan serta mengurangi risiko overfitting. Namun, kebanyakan penelitian sebelumnya masih menggunakan parameter bawaan (*default*), sehingga performa model belum optimal (Zhu et al., 2022). Maka, proses optimasi parameter sangat penting untuk menemukan

untuk menemukan konfigurasi terbaik. Salah satu proses tuning parameter yang sering digunakan adalah *GridSearchCV*, sebuah teknik untuk mencari parameter terbaik secara sistematis (Dimgba & Andreas, 2024).

Penelitian yang dilakukan oleh Ramzan et al., memaparkan bahwa dalam klasifikasi anemia dengan metode *Support Vector Machines* (SVM) menunjukkan bahwa akurasi meningkat sebesar 5-10% setelah parameter dioptimalkan menggunakan *GridSearchCV* (Ramzan et al., 2023). Hal tersebut menunjukkan, bahwa optimasi bukan hanya proses tambahan, tetapi juga merupakan bagian penting dalam menciptakan model prediksi yang lebih akurat dan stabil.

Studi yang dilakukan oleh Hossain, menemukan bahwa hemoglobin adalah indikator terpenting, diikuti oleh MCV dan MCH (Tepakhan et al., 2025). Analisis tersebut juga penting agar nantinya hasil penelitian tidak hanya bersifat *blackbox*, tetapi juga memberikan pemahaman klinis yang dapat membantu tenaga medis dalam mengenali faktor risiko utama anemia.

Jika ditinjau dari sudut pandang Islam, upaya penelitian ini sesuai dengan ajaran agama yang mengingatkan kita mengenai pentingnya menjaga kesehatan. Allah SWT berfirman dalam Surah Al Baqarah ayat 195:

وَلَا تُلْقُوا بِأَيْدِيكُمْ إِلَى التَّهْلُكَةِ

“.. dan janganlah kamu menjatuhkan dirimu sendiri ke dalam kebinasaan..” (Q.S. Al-Baqarah/2:195).

Ayat ini memberi tahu kita bahwa menjaga kesehatan, seperti mendeteksi penyakit secara dini, misalkan anemia, adalah kewajiban setiap orang muslim untuk menjaga dirinya sendiri.

Rasulullah SAW juga memperkuat hal tersebut yakni pentingnya menjaga kesehatan dalam sebuah hadist:

إِنَّ لِجَسَدِكَ عَلَيْكَ حَقًّا

"Sesungguhnya badanmu memiliki hak atasmu" (HR. Bukhari dan Muslim).

Oleh karena itu, penggunaan *machine learning* dalam bidang kesehatan dapat dianggap sebagai upaya ilmiah kita untuk memenuhi amanah tersebut, yakni menjaga tubuh yang diberikan oleh Allah SWT. Hal tersebut juga sesuai dengan tujuan maqashid syariah, yaitu melindungi jiwa (hifz an-nafs), yang merupakan salah satu dasar utama dalam syariat islam.

Berdasar beberapa uraian tersebut, penelitian ini penting untuk dilakukan karena dapat mengklasifikasikan Anemia menggunakan metode *Random Forest* dengan metode *GridSearchCV*. Selain mengklasifikasikan anemia, penelitian ini juga memberikan kontribusi ilmu pengetahuan melalui analisis mendalam mengenai peran fitur medis dalam melakukan klasifikasi anemia.

1.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, perumusan masalah dalam penelitian ini, antara lain:

1. Bagaimana performa model terbaik dari metode *Random Forest* dengan optimasi *GridSearch* dalam mengklasifikasikan penyakit anemia?
2. Bagaimana pengaruh seleksi fitur terhadap pengklasifikasian yang dihasilkan oleh model?

1.3 Batasan Masalah

Agar penelitian lebih terarah, batasan masalah yang digunakan dalam penelitian ini, antara lain:

1. Dataset yang digunakan yaitu Anemia Dataset yang bersumber dari Kaggle.
2. Fokus penelitian ini adalah pada proses klasifikasi penyakit anemia yakni terkena Anemia atau tidak dan evaluasi model berdasarkan metrik akurasi, presisi, recall, dan F-1 score.
3. Proses optimasi hyperpararmeter tuning dilakukan menggunakan GridSearch dengan kombinasi parameter (n_estimators, max_depth, max_features dan min_samples_leaf).

1.4 Tujuan Penelitian

Tujuan penelitian yang ingin dicapai dalam penelitian ini, antara lain:

1. Untuk mengetahui peforma model terbaik yang dihasilkan oleh metode *Random Forest* dengan *GridSearch* dalam mengklasifikasikan penyakit

anemia.

2. Untuk mengetahui bagaimana pengaruh seleksi fitur terhadap pengklasifikasian yang dihhasilkan oleh model.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini dapat dirasakan berbagai pihak, antara lain:

1. Bagi tenaga medis, penelitian ini dapat digunakan sebagai alat bantu diagnosis awal anemia berdasarkan data.
2. Bagi pasien, penelitian ini bermanfaat dalam deteksi dini anemia sehingga penanganan dapat dilakukan dengan cepat.
3. Bagi akademisi dan peneliti, penelitian ini memperkaya literature mengenai penerapan machine learning dibidang kesehatan, khususnya dalam klasifikasi penyakit berbasis data.
4. Bagi pengembang teknologi, penelitian ini dapat dijadikan dasar pengembangan sistem pendukung keputusan yang dapat diterapkan kedalam aplikasi.

BAB II

STUDI PUSTAKA

2.1 Penelitian Terkait

Penelitian terkait digunakan untuk meninjau studi-studi sebelumnya yang saling berhubungan, sehingga dapat menjadi dasar dan pembanding bagi penelitian ini.

Airlangga (2024) melakukan penelitian mengenai klasifikasi anemia menggunakan metode *machine learning* dengan memanfaatkan data *Complete Blood Count* (CBC). Penelitian tersebut menggunakan dataset yang berasal dari rekam medis dengan jumlah sampel sebesar 564 data. Model klasifikasi yang digunakan dalam penelitian tersebut meliputi *Random Forest, Decision Tree,* dan *XGBoost,* dengan proses tuning hiperparameter melalui *GridSearch* dan validasi silang k-fold. Hasil akhir yang diperoleh dari penelitian tersebut menunjukkan bahwa *Decision Tree* yang dioptimasi dengan *GridSearch* menghasilkan *balanced accuracy* tertinggi, yakni sebesar 94,17%.

Penelitian terkait juga dilakukan oleh Ahamad et al. (2023) yang membahas mengenai evaluasi kinerja model *machine learning* dengan tuning hiperparameter menggunakan *GridSearch*. Penelitian tersebut menggunakan data medis publik dari UCI I dan II yang mencakup sekitar 1300 data rekam medis. Dengan menguji beberapa model seperti *Logistic Regression, Random Forest,* dan *Gradient Boosting* diperoleh hasil akhir yang menunjukkan adanya peningkatan signifikan setelah optimasi, dimana pada data I yang dituning memperoleh akurasi sebesar 87,91% sedangkan pada data II yang dituning hanya sebesar 99,03%.

Sementara itu, Samtani et al. (2025) meneliti terkait deteksi anemia dengan membandingkan berbagai algoritma *machine learning* termasuk *Random Forest, Logistic Regretion*, dan SVM menggunakan dataset CBC sebesar kurang lebih 1.000 data. Optimasi dilakukan menggunakan *GridSearch* serta validasi silang untuk menilai kestabilan model. Hasil akhir yang diperoleh menunjukkkan bahwa *Random Forest* dengan optimasi hiperparameter mencapai akurasi 99,48%, lebih tinggi dibandingkan dengan *Logistic Regretion* sebesar 57,23% dan SVM sebesar 23,81%.

Tepakhan et al. (2025) juga melakukan penelitian terkait dengan membedakan Anemia defisiensi besi dan talasemia menggunakan *machine learning*. Dataset yang digunakan mencakup 1143 sampe data IDA sebesar 382 dan thalesemia sebesar 635 data dengan 15 parameter CBC. *Random Forest* digunakan sebagai metode utama, dengan hasil evaluasi menunjukkan akurasi sebesar 82,2% dan AUC-ROC sebesar 0,899.

Setiawan et al. (2023) melakukan penelitian mengenai klasifikasi subtipe anemia menggunakan pendekatan data mining. Metode yang digunakan adalah *Random Forest, J48 Decission Tree*, dan *Naïve Bayes* dengan dataset yang berasal dari data CBC dan data demografi pasien. Sehingga didapatkan hasil berupa *Decission Tree* dan J48 sebesar 99,61% memberikan peforma dengan akurasi yang cukup tinggi dalam membedakan subtipe anemia.

Sari et al. (2024) juga melakukan penelitian mengenai klasifikasi penyakit anemia dengan menggunakan pendekatan *Decision Tree* C4.5 dan *K-Nearest Neighbor* (KNN). Penelitian ini menggunakan dataset publik Kaggle yang berisi

1.421 sampel dengan 6 atribut yakni gender, Hb, MCH, MCV, MCHC, MCV, dan result. Hasil penelitian menunjukan algoritma C4.5 memberikan peforma terbaik dengan akurasi sebesar 99,75%.

Tabel 2.1 merangkum penelitian-penelitian terkait yang menjadi referensi untuk pengembangan penelitian ini.

Tabel 2.1 Penelitian Terkait

No.	Nama Peneliti	Judul Penelitian	Metode	Dataset	Hasil Penelitian
1.	(Airlangga, 2024)	Leveranging Machine Learning for Accurate Anemia DAgnosis Using Complete Blood Count Data	Random Forest, Decision Tree + GridSearch	Dataset CBC, 564 sampel data pasien	Decision Tree + GridSearch menghasilkan balanced accuracy sebesar 93,8%
2.	(Ahamad et al., 2023)	Performance of Optimal Hyperparameters on the Peformance of Machine Learning	Logistic Regression, Random Forest, Gradient Boosting + GridSearch	Dataset UCI, 303 pada data I dan 1025 pada data ke II	GridSearch meningkatkan performa dari data I sebesar 87,91% menjadi 99,03% pada data II
3.	(Samtani et al., 2025)	Anemia Detection Using CBC Data: A Comparative Study	Random Forest, Logistic Regretion, SVM + GridSearch	Dataset CBC, kurang lebih 1000 sampel pasien	Random Forest + GridSearch mencapai akurasi tertinggi sebesar 99,48%
4.	(Tepakhan et al., 2025)	Machine Learning Approach For DifferentAting Iron Deficiency Anemia and Thalassemia	Random Forest + gradient boosting	Dataset CBC, 1143 sampel data	Akurasi Random Forest sebesar 82,2% dan AUC-ROC sebesar 0,899
5.	(Setiawan et al., 2023)	Data Mining Techniques for Predictive Classification of Anemia Disease Subtypes	Random Forest, J48, Naïve Bayes, Decision Tree	Dataset CBC + demografi pasien (UMN)	Akurasi Random Forest & J48 sebesar 99,6%
6.	(Sari et al., 2024)	Klasifikasi Penyakit Anemia Menggunakan Machine Learning	Decision Tree C4.5 dan K-NN	Dataset publik Kaggle Anemia sebanyak 1421 sampel	C4.5 dengan akurasi 99,75%

Berdasarkan hasil penelitian pada tabel 2.1, menunjukkan bahwa metode machine learning yang dituning dengan *GridSearch* memiliki kinerja unggul dalam klasifikasi medis, termasuk anemia. Namun, sebagian besar penelitian tersebut menggunakan dataset yang terbatas. Oleh karena itu, penelitian ini memanfaatkan dataset publik dari Kaggle untuk mengimplementasikan metode *Random Forest* menggunakan *GridSearch*, sehingga hasil yang diperoleh lebih akurat dan dapat diandalkan.

2.2 Anemia

Anemia adalah kondisi penyakit yang ditandai oleh rendahnya kadar hemoglobin yang mengakibatkan darah tidak dapat mengangkut oksigen secara optimal ke seluruh tubuh. Anemia disebabkan oleh berbagai hal, mulai dari kekurangan zat besi, kehilangan darah kronis, penyakit kronis, hingga kelainan gen seperti thalasemia. Kurangnya zat besi masih menjadi penyebab utama anemia secara global hingga saat ini (Chaparro et al., 2019; Belo et al., 2021). Kondisi tersebut paling banyak ditemukan pada kelompok rentan seperti wanita hamil, anak-anak, hingga remaja. Anemia tidak hanya berdampak pada gangguan fisik, tetapi juga pada penurunan produktivitas kerja dan gangguan perkembangan kognitif pada anak. Sehingga, kondisi tersebut juga meningkatkan risiko mortalitas pada ibu dan bayi pada masa mengandung.

Diagnosis anemia umumnya menggunakan pemeriksaan *berupa Complete Blood Count* (CBC) yang mencakup parameter seperti Hb, HCT, MCV, MCH, dan RDW. Data CBC relatif mudah diperoleh untuk menjadi standar fasilitas kesehatan. Selain itu, integrasi antara data CBC dengan *machine learning* dapat meningkatkan akurasi diagnosis secara signifikan (Kitaw et al., 2024).

Diferensiasi subtipe anemia seperti *Iron Deficiency Anemia* (IDA) dan thalassemia mampu dibedakan dengan hasil akhir Hb dan MCV menjadi fitur dominan yang dapat membantu prediksi (Tepakhan et al., 2025). Hal tersebut berpeluang besar dalam pemanfaatan data hematologi sederhana untuk klasifikasi yang lebih presisi. Dengan tingginya prevelensi anemia, pemanfaatan teknologi berbasis data medis sangat diperlukan. Pendekatan menggunakan *machine learning* dapat mempercepat diagnosis dan mengurangi kesalahan klinis. Model klasifikasi juga dapat membantu dokter dalam menentukan terapi yang lebih sesuai untuk pasien, sehingga penelitian ini memiliki tingkat urgensi yang tinggi untuk mendukung praktik medis dibidang kesehatan.

2.3 GridSearch

GridSearch adalah metode optimasi hiperparameter untuk melakukan pencarian sistematis terhadap kombinasi parameter yang telah ditentukan dimana setiap kombinasi dievaluasi menggunakan teknik *cross-validation* sehingga diperoleh parameter dengan peforma terbaik. Kesederhanaan dan keterbukaan terhadap interpretasi menjadi faktor keunggulan *GridSearch*. Metode tersebut juga mudah direplikasi sehingga menjadi populer dalam penelitian medis. Selain itu, struktur pencarian yang jelas dan sistematis serta fleksibel dalam menentukan ruang pencarian parameter, membuat peneliti dapat menyesuaikan eksperimen sesuai kebutuhan.

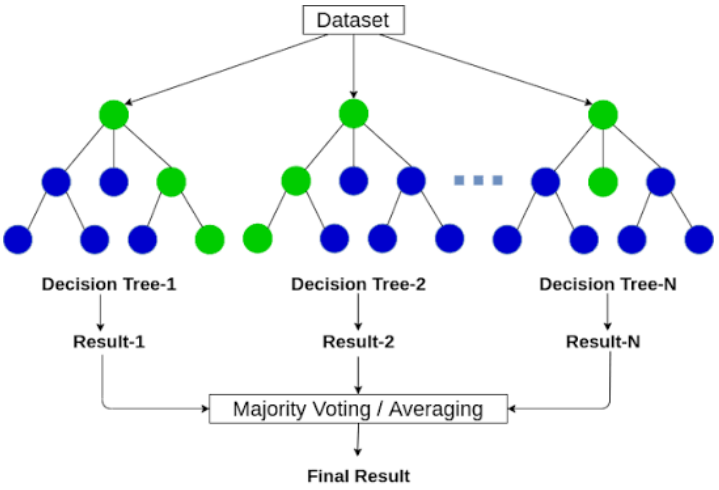
GridSearch digunakan untuk melakukan *hyperparameter tuning* pada beberapa model *machine learning* seperti *Random Forest* (Airlangga, 2024). Dataset yang digunakan berupa data CBC dari fasilitas medis dibeberapa tempat, dengan validasi silang 5-fold yang digunakan untuk eksperimen.

Hasil yang didapat menunjukkan *DesicionTreeClassifier* mencapai *balanced accuracy* 94,17%, walaupun teknik ensemble seperti *Random Forest* setelah tuning mendekati peforma yang bagus. Hal tersebut menunjukkan bahwa walau *GridSearch* dapat memakan waktu, namun hasilnya akan memberikan peningkatan dalam klasifikasi anemia.

Selain itu, *GridSearch* juga digunakan untuk menemukan kombinasi parameter terbaik dari beberapa metode klasifikasi, termasuk *Random Forest* (Ahamad et al., 2023). Dataset medis diuji dengan dan tanpa tuning *hyperparameter*, dimana hasil uji dengan tuning menunjukkan peningkatan *metric* seperti akurasi, presisi, *recall,* dan *f1-score* dibandingkan dengan versi tanpa tuning. Metode ini membuktikan bahwa penggunaan *GridSearch* pada dataset medis tidak hanya pada aspek teoritis saja, tetapi juga praktis untuk memberikan manfaat dalam meningkatkan diagnostik model.

2.4 Random Forest

Random Forest adalah teknik *ensemble learning* berbasis *decision tree* yang menggabungkan banyak pohon keputusan untuk menghasilkan prediksi yang lebih akurat. Setiap pohon dibangun dari smoke data acak dan subset fitur yang berbeda. *Random Forest* memiliki keunggulan utama seperti kemampuan dalam menangani dataset besar dengan berbagai fitur serta ketahanan terhadap *outlier* (Kim et al., 2025).



Gambar 2.1 Alur Metode Random Forest

Gambar 2.1 menerangkan mengenai bagaimana cara kerja dari metode *Random Forest*, yakni teknik *ensemble learning* yang membangun banyak decision tree dari dataset yang sama dengan teknik bootstrap sampling serta pemilihan fitur secara acak. Setiap pohon menghasilkan prediksi masing-masing, kemudian digabungkan melalui mekanisme majority voting (untuk klasifikasi) atau averaging (untuk regresi). Proses penggabungkan ini membuat hasil akhir yang lebih akurat, stabil, dan tahan terhadap *overfitting* dibandingkan hanya dengan menggunakan satu decision tree saja (Samtani et al., 2025).

Metode ini juga dapat menghasilkan ukuran *feature importance* sehingga intepretasi hasil menjadi lebih mudah. Hal tersebut membuktikan efektivitas *Random Forest* dalam mengklasifikasikan tingkat keparahan COVID-19 menggunakan data klinis. Hal tersebut menunjukkan revalansi pada kasus medis yang lebih kompleks (Chen et al., 2020).

Random Forest konsisten dalam menunjukan peforma yang lebih baik dibandingkan logistic regession dan SVM (Yimer et al., 2025). Model tersebut sangat sesuai dengan dataset CBC yang berisi banyak parameter yang saling berelasi. diagnosis medis.

Random Forest yang dioptimasi berhasil mencapai akurasi tinggi, dengan Hb dan MCV sebagai fitur terpenting (Tepakhan et al., 2025). Hasil tersebut menegaskan bahwa *Random Forest* tidak hanya kuat dalam prediksi tetapi juga dapat membantu klinisi dalam memahami variabel mana yang paling berpengaruh. Temuan tersebut berkontribusi pada penerapan model interpretatif dalam bidang kesehatan. Maka, *Random Forest* menjadi metode yang sesuai untuk penelitian ini, terlebih jika dipadukan dengan optimasi *GridSearch*, sehingga peformanya dapat dimaksimalkan. Diharapkan nantinya penelitian ini akan mampu dalam meningkatkan akurasi klasifikasi anemia serta mendukung pengambilan keputusan klinis yang lebih sesuai.

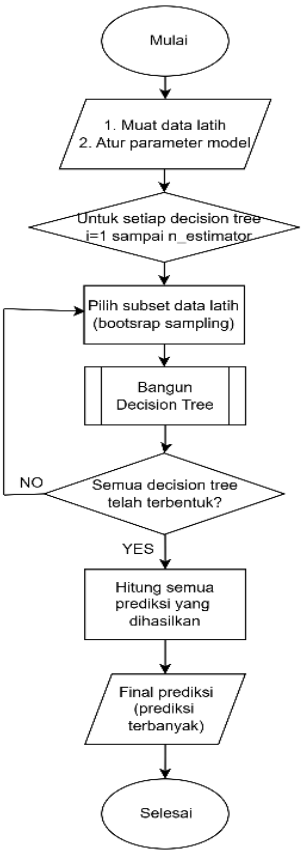
BAB III

METODE PENELITIAN

3.1 Desain Sistem

Desain sistem ini seperti terlihat pada gambar 3.1, disusun secara sistematis untuk memastikan proses pemodelan berjalan sesuai tujuan. Langkah pertama yang harus dilakukan yakni mempersiapkan dataset anemia serta menentukan parameter awal misalnya *n_estimators* dan *max_features* sebagai dasar konfigurasi model. Selanjutnya, data tersebut dibagi (*split*) menjadi data latih dan data uji. Kemudian sistem dirancang melalui *boostrap sampling*, pemilihan fitur acak, hingga pembentukan *decision tree* secara bertahap. Setelah itu, sistem kemudian diimplementasikan melalui pembangunan model *Random Forest.* Setelah model selesai dibangun, dilakukan proses pengujian dengan cara memasukkan data uji untuk memperoleh hasil klasifikasi dari setiap pohon dalam ensembel. Yang terakhir, yakni tahap evaluasi yang dilakukan dengan cara menggabungkan hasil klasifikasi tadi melalui majority vote berdasarkan metrik evaluasi.

Pada proses pelatihan, setiap *decision tree* dibangun menggunakan subset data hasil *bootstrap sampling* sehingga masing-masing pohon mempelajari pola data yang berbeda. Pendekatan ini bertujuan untuk meningkatkan kemampuan generalisasi model serta mengurangi risiko *overfitting*, sehingga hasil klasifikasi yang dihasilkan menjadi lebih stabil, konsisten, dan memiliki tingkat akurasi yang lebih baik ketika diterapkan pada data yang belum pernah dipelajari sebelumnya.



Gambar 3. 1 Alur Penelitian

3.1.1 Persiapkan Dataset dan Hyperparameter

Persiapkan Dataset dan *Hyperparameter* Masukkan dataset anemia dan konfigurasikan parameter seperti *n_estimators* (jumlah pohon) dan *max_features* (jumlah fitur yang harus dipertimbangkan saat melakukan *split*).

3.1.2 Looping Iterasi

Lakukan *looping* yang dimulai dari iterasi ke 1 hingga n_estimators untuk setiap decision tree.

3.1.3 Bootstrap Sampling

Bootstrap sampling yakni suatu proses untuk pengambilan subset data latih secara acak yang berasal dari dataset asli disertai dengan penggantian untuk mengolah tree (Fu et al., 2024).

Kode 3.1 Persiapan Data

```
# 1. Persiapan Data (Simulasi)

# Membuat data dummy: 100 sampel, 4 fitur,

2 kelas X = np.random.rand(100, 4) * 10

y = (X[:, 0] + X[:, 3] >

10).astype(int) #

Memisahkan data latih dan

data uji

X_train, X_test, y_train, y_test = train_test_split( X, y,

            test_size=0.3, random_state=42

)
```

Latih setiap decision tree pada subset data pelatihan yang berbeda. Kemudian pilih N *sampel* 100 atau 10 persen yaitu sama dengan jumlah keseluruhan data training yang dipilih secara acak dengan pengembalian dari dataset asli. Yang berarti beberapa baris data mungkin muncul lebih dari sekali pada subset, atau bahkan tidak (Breiman, 2020).

3.1.4 Tahapan Menerapkan Metode Decision Tree

Sebelum masuk ke tahap tersebut, terlebih dahulu pilih fitur *random* (*random feature selection*), dimana selain data yang diacak, *Random Forest* juga mengacak fitur. Saat setiap *node* pada *tree* yang sedang dipertimbangkan untuk dilakukan *split*, hanya subset acak fitur (misalnya $\sqrt{p}$ atau log2(p) fitur dari total p fitur) yang dibolehkan untuk dipilih sebagai split terbaik.

Kode 3.2 Inisialisasi

```
# 2. Inisialisasi   dan Pelatihan Model
# n_estimators: Jumlah pohon (sesuai n_estimators di
diagram) # random_state: Untuk hasil yang dapat
direproduksi model_rf = RandomForestClassifier(
n_estimators=100,
max_features='sqrt',  # Menggunakan akar kuadrat dari jumlah
                      fitur random_state=42
)
```

Pengacakan fitur tersebut mencegah tree terlalu mirip, terutama jika ada satu atau dua fitur yang sangat dominan. Sehingga hal tersebut mengurangi hubungan antar *tree*, dimana hal tersebut dapat meningkatkan akurasi ensemble secara keseluruhan.

Kode 3.3 Bangun Model

```
# 'Bangun Decision Tree'(Train the model)
model_rf.fit(X_train, y_train)
```

Selanjutnya bangun decision tree yang utuh berdasarkan sampel data hasil *bootstrap* dan *set* fitur tersebatas dari langkah sebelumnya.

3.1.5 Pengecekan kondisi

Langkah selanjutnya, cek terlebih dahulu apakah semua *decision tree* telah terbentuk, jika tidak ulangi langkah ke tiga.

3.1.6 Ensemble Agregation

Pada langkah ini terdiri dari 2 bagian yakni klasifikasi dimana prediksi akhir ditentukan oleh suara mayoritas (*majority vote*) dari semua tree dan regresi yakni prediksi akhir ditentukan oleh rata-rata (*average*) dari semua *output tree*.

Kode 3.4 Prediksi

```
# 3. Prediksi dan Evaluasi

# 'Hitung semua Prediksi yang dihasilkan'

(Predict) y_pred = model_rf.predict(X_test)
```

Setelah semua *tree n_estimator* dibangun, masukkan data uji ke setiap pohon untuk mendapatkan prediksi individual. Kemudian pada tahap ini tentukan hasil akhir berdasar kelas yang paling banyak dipilih (mayoritas).

3.1.7 Split Data

Split data adalah tahap untuk membagi data menjadi beberapa bagian. Pada tahap tersebut, dataset yang telah diproses tadi kemudian dibagi menjadi dua bagian lagi agar dapat digunakan untuk proses pelatihan (*training*) dan pengujian (*testing*) model. Pembagian data tersebut penting agar model tidak hanya menghafal pola data testing, tetapi juga dapat melakukan generalisasi pada data baru yang belum pernah dilihat sebelumnya (Purba et al., 2022). Dengan adanya split data, performa model dapat dievaluasi dengan objektif berdasarkan data testing (Nurhopipah et al., 2020).

Pembagian dataset dilakukan menggunakan fungsi *train_test_split* dari library scikit-learn dengan sintaks pada kode 3.5:

Kode 3.5 Split Data

```
X_train, X_test, y_train, y_test = train_test_split(X[fitur], y, test_size=0.2,
random_state=rs)
```

Biasanya, proses *split* data dapat dinyatakan dengan persamaan 3.1:

$$D = D\ train \cup D\ test, D\ train \cap D\ test = \emptyset \quad (3.1)$$

Dengan proporsi:

$$|D\ train| = (1-\propto) \cdot |D|, \quad |D\ test| = \propto \cdot |D| \quad (3.2)$$

Dalam persamaan (3.1) dijelaskan bahwa D yakni semua dataset dibagi menjadi dua himpunan yang saling lepas, yakni *D train* sebagai data latih(*training*) dan *D test* sebagai data uji (*testing*). Simbol ∪ yang berati gabungan himpunan, yang berarti seluruh data yang terdiri dari data latih dan data uji, sedangkan simbol ∩ = ∅ yang berarti tidak ada data yang menumpuk antara keduanya. Dengan begitu, setiap sampel hanya berada pada salah satu bagian saja, sehingga hasil evaluasi benar menjelaskan kemampuan model terhadap data baru (Jin et al., 2021).

Sedangkan pada persamaan (3.2) menjelaskan proporsi pembagian data latih dan uji. Simbol |*D train*| berarti banyaknya data latih yang diperoleh dari $(1-\propto)$ yang dikalikan dengan jumlah seluruh data, sedangkan |*D test*| yakni banyak data uji yang diperoleh dari $\propto \cdot |D|$. Simbol $\propto$ yakni porsi data uji (sebagai contoh, disini penulis menggunakan 0.2 untuk pembagian sebesar 20%). Dengan persamaan tersebut, proses split data dapat diatur sesuai kebutuhan.

Pada penelitian tersebut, sebagai contoh, penulis membagi data menjadi 80% untuk data training dan 20% untuk data testing. Percobaan tersebut diulang sebanyak tiga kali dengan *random state yang* berbeda (42, 123, 2025) agar kestabilan model bisa dilihat pada beberapa senario pembagian data. Dengan cara tersebut, dapat diperoleh gambaran yang lebih menyeluruh mengenai konsistensi model *Random Forest* dalam klasifikasi anemia.

3.1.8 Pemodelan

Tahap selanjutnya yakni pemodelan yang merupakan proses dalam membangun model klasifikasi menggunakan metode random forest, yang diimplementasikan dalam library *scikit-learn* dengan sintaks 3.6 :

Kode 3.6 Bangun Random Forest

```
RandomForestClassifier(random_state=rs)
```

Model yang digunakan pada penelitian ini yakni *Random Forest* yang merupakan salah satu teknik *ensemble learning* yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi pengklasifikasian dan mengurangi risiko *overfitting* (ZemarAm at al., 2024). Proses pelatihan model dilakukan dengan parameter default dan hasilnya kemudian dioptimasi menggunakan *GridSearchCV* yang merupakan salah satu metode pencarian sistematis dengan kombinasi parameter terbaik melalui teknik *cross-validation* (Jiang et al., 2022). Pemilihan metode *Random Forest* didasarkan pada keefektifan metode tersebut dalam melakukan suatu pengklasifikasian yang menangani dataset dengan fitur numerik dan kategorikal. Untuk *hyperparaneter* yang diuji dalam penelitian ini dapat dilihat pada kode 3.7 :

Kode 3.7 Menentukan Parameter

```
param_grid = {
    "n_estimators": [10, 20],
    "max_depth": [3, 5],
    "max_features": [1]
    "min_samples_leaf": [5]
  }
```

Dimana setiap kombinasi diuji menggunakan *GridSearchCV* sehingga didaparkan parameter terbaik untuk setiap skenario. Parameter *n_estimators* yang merupakan jumlah pohon keputusan yang memiliki nilai 10 dan 20. Kemudian parameter *max_depth* yang merupakan kedalaman tree yang digunakan untuk membatasi seberapa jauh tree dapat melakukan pemisahan (*split*) pada data. Parameter ini memiliki variasi 3 dan 5. Selanjutnya, adalah parameter *max_features* yakni jumlah fitur yang dipilih secara acak pada setiap split yang yang bernilai 1. Dan yang terakhir yakni *min_samples_leaf* yakni jumlah maxium data pada leaf yang bernilai 5 (Thomas & Kaliraj, 2024). Sehingga dari kombinasi tersebut dapat diperoleh total 2 x 2 x 1 x 1 = 4 percobaan untuk setiap skenario.

Kode 3.8 Inisialisasi GridSearch

```
GridSearchCV(..., cv=5)
```

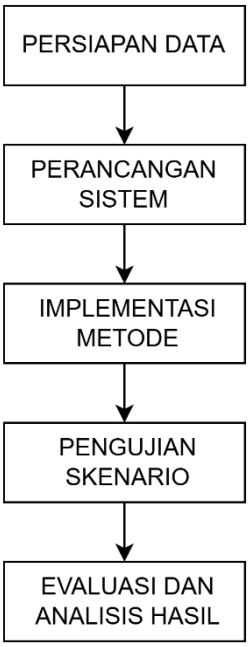
Selanjutnya, setiap kombinasi diuji menggunakan 5-*fold cross-validation*, sehingga total percobaan menjadi 4 x 5 = 20 kali pelatihan untuk setiap skenario.

Dalam 5-*fold* CV, data latih dibagi lagi menjadi lima lipatan (5-*fold* CV) yang fungsinya bergantian sebagai data latih dan data validasi, sehingga proses evaluasi nantinya dapat lebih objektif dalam menentukan parameter terbaik. Pemodelan *Random Forest* pada penelitian ini tidak hanya membangun model klasifikasi dasar, tetapi juga mengintegrasikan optimasi *hyperparameter* untuk mendapatkan pengklasifikasian yang terbaik dalam mengklasifikasikan anemia (Azis et al., 2024).

Penggunaan *random state* yang berbeda (42, 123, 2025) bertujuan untuk menjamin *reproducibility* agar hasil pembagian data dan pembentukan model tetap konsisten saat dijalankan ulang. Dengan mengunci nilai tertentu, hasil penelitian menjadi lebih stabil dan dapat diverifikasi tanpa adanya perubahan nilai. Selain itu, beberapa variasi sekaligus berfungsi sebagai uji *robustness* untuk memastikan peforma model dapat stabil (Ahmed et al., 2022).

3.2 Desain Penelitian

Desain penelitian ini ini seperti terlihat pada gambar 3.2, disusun secara bertahap agar sesuai dengan tujuan yang diharapkan oleh penulis. Langkah pertama yang harus dilakukan yakni menyiapkan data sebagai dasar penelitian. Kemudian sistem dirancang, yakni dengan menentukan alur kerja, model, serta parameter yang akan diuji. Untuk tahap implementasi metode, model dikembangkan serta dijalankan menggunakan data yang telah dikumpulkan tadi. Setelah itu, dilakukan pengujian dengan berbagai skenario untuk melihat adanya perubahan fitur atau parameter yang memengaruhi kinerja model. Yang terakhir, yakni melakukan evaluasi serta analisis hasil, dengan cara membandingkan model berdasarkan metrik evaluasi untuk mendapatkan hasil akhir (Susanto, 2025).



Gambar 3.2 Desain Penelitian

3.2.1 Persiapan Data

Masukkan dataset anemia dan konfigurasikan parameter seperti *n_estimators* (jumlah pohon) dan *max_features* (jumlah fitur yang harus dipertimbangkan saat melakukan *split*). Pada penelitian ini, dalam mengklasifikasikan anemia, penulis menggunakan dataset yang didapatkan melalui website Kaggle. Dalam "Anemia Dataset" yang berisi data sebanyak 1421 baris dan 6 kolom, dengan salah satu kolom target yang bernama *result* yang terdiri dari pengklasifikasian meliputi 0 yang berarti normal dan 1 yang berarti anemia dengan kategori normal sebanyak 801 dan anemia sebesar 620. Distribusi dataset sebesar 56% : 44% dikategorikan sebagai seibang (*well-balanced*) (Kaope et al., 2023), sehingga penggunaan SMOTE tidak diperlukan. Hal ini dilakukan untuk menjaga originalitas pola data asli dan menghindari risiko bias. Namun, SMOTE tetap direkomendasikan untuk penelitian kedepannya jika peneliti memggunakan dataset dengan tingkat ketimpangan yang lebih ekstrem demi menjaga generalisasi model (Kaope et al., 2023).

Selain distribusi kelas, karakteristik awal dataset juga dianalisis untuk memahami struktur data sebelum dilakukan proses pemodelan. Dataset terdiri atas lima variabel prediktor, yaitu gender, hemoglobin, MCH, MCHC, dan MCV, serta satu variabel target yaitu result. Dari keseluruhan atribut tersebut, fitur hemoglobin, MCH, MCHC, dan MCV merupakan data numerik kontinu, sedangkan variabel *gender* dan *result* merupakan data kategorikal biner. Berdasarkan pemeriksaan awal pada dataset, tidak ditemukan *missing value* atau data kosong sehingga seluruh data dapat digunakan secara langsung pada tahap analisis berikutnya tanpa memerlukan proses imputasi data.

Tabel 3.1 Dataset Anemia

Gender	Hemoglobin	MCH	MCHC	MCV	Result
1	14.9	22.7	29.1	83.7	0
0	15.9	25.4	28.3	72	0
0	9	21.5	29.6	71.2	1
0	14.9	16	31.4	87.5	0
1	14.7	22	28.2	99.5	0
..	..	..	..	..	..
1	13.1	17.7	28.1	80.7	1
0	14.3	16.2	29.5	95.2	0
0	11.8	21.2	28.4	98.1	1

Sumber : https://www.kaggle.com/datasets/biswaranjanrao/Anemia-dataset

Tabel 3.1 merupakan potongan data yang berasal dari “Anemia Dataset“ yang digunakan penulis dalam penelitian ini. Dataset ini berisi informasi terkait kondisi hematologi responden, seperti variabel gender, kadar hemoglobin, serta parameter sel darah merah lain seperti MCH, MCHC, dan MCV. Data tersebut berupa tabel sehingga mudah dipahami serta dianalisis dalam konteks klasifikasi medis (Airlangga et al., 2024). Beragamnya nilai dalam setiap fitur dapat memberikan fondasi yang kuat untuk melatih model dalam melakukan klasifikasi secara tepat serta akurat (Kitaw et al., 2024). Dengan adanya struktur dataset yang rinci, proses pengolahan data dapat dilakukan secara sistematis.

Tabel 3.2 Deskripsi Fitur

Fitur	Deskripsi
Gender	Jenis kelamin responden yakni 0 = laki-laki, 1 = perempuan
Hemoglobin	Hemoglobin yakni protein dalam sel darah merah yang membawa oklsigen ke jaringan tubuh dan mengangkut karbondioksida kembali ke paru-paru; dengan rentang nilai 6.6 – 16.9 g/dL
MCH	Mean Corpuscular Hemoglobin yakni rata-rata jumlah hemoglobin dalam setiap sel darah merah; dengan rentang nilai 16 – 30 pg
MCHC	Mean Corpuscular Hemoglobin Concentration yakni mengukur konsentrasi rata-rata hemoglobin dalam satu sel darah merah; dengan rentang nilai 27.8 – 32.5 g/dL
MCV	Mean Corpuscular Volume yakni mengukur ukuran rata-rata sel darah merah; dengan rentang nilai 69.4 – 102 fL
Result	Label target klasifikasi yakni 0 = tidak Anemia, 1 = Anemia

Tabel 3.2 merupakan deskripsi fitur yang menjelaskan pengertian dari setiap kolom yang terdapat dalam dataset “Anemia Dataset”. Deskripsi fitur ini penting dalam penelitian ini agar konteks dari variabel - variabel tersebut menjadi jelas dan mudah dipahami. Fitur numerik seperti Hemoglobin, MCH, MCHC, dan MCV dilengkapi dengan rentang nilai yang menggambarkan variasi biologis pada responden. Sedangan fitur kategorikal seperti gender dan result menjelaskan perbedaan individu serta status anemia yang menjadi target klasifikasi. Penyajian deskripsi fitur ini juga bermanfaat dalam tahap prapemrosesan data, khususnya untuk memastikan agar data yang dianalisis dapat sesuai dengan kriteria serta kaidah medis (Awaad et al., 2025). Sehingga, tabel 3.2 ini berfungsi sebagai awal dalam memahami variabel penelitian sebelum dilakukan pemodelan lebih lanjut.

Sebelum memasuki tahap pemodelan, dilakukan analisis eksplorasi data (*Exploratory Data Analysis*) untuk melihat karakteristik statistik awal dari setiap fitur. Berdasarkan pengamatan terhadap rentang nilai pada dataset, fitur hemoglobin memiliki rentang nilai antara 6.6–16.9 g/dL, fitur MCH berada pada kisaran 16–30 pg, MCHC berada pada rentang 27.8–32.5 g/dL, sedangkan MCV berada pada rentang 69.4–102 fL. Variasi nilai yang cukup lebar pada masing-masing fitur menunjukkan adanya perbedaan kondisi hematologi antar responden sehingga memberikan informasi yang dibutuhkan model untuk mempelajari pola klasifikasi secara lebih efektif.

Selanjutnya dilakukan analisis hubungan antar fitur (*feature correlation*) untuk memahami keterkaitan antar variabel sebelum proses pelatihan model dilakukan. Analisis korelasi penting untuk mengetahui fitur mana yang memiliki hubungan paling besar terhadap target klasifikasi anemia. Secara medis, variabel hemoglobin diketahui memiliki hubungan yang paling erat terhadap kondisi anemia karena merepresentasikan kadar protein pembawa oksigen dalam darah yang menjadi indikator utama diagnosis anemia. Sementara itu, fitur lain seperti MCV, MCH, dan MCHC berfungsi sebagai parameter pendukung yang menjelaskan karakteristik sel darah merah. Analisis ini dilakukan untuk memastikan bahwa fitur-fitur yang digunakan memang memiliki relevansi terhadap proses klasifikasi sehingga model yang dibangun dapat mempelajari pola hubungan data secara lebih optimal (Kitaw et al., 2024).

3.2.2 Perancangan Sistem

Perancangan sistem dilakukan untuk menentukan alur penelitian yang dimulai dari penentuan struktur data, pemilihan fitur, dan mekanisme alur (input, proses, output) yang akan digunakan dalam model *machine learning*. Menurut Shaveta, rancangan sistem yang terstruktur diperlukan untuk memastikan tahapan machine learnig berjalan konsisten dan dapat digunakan dengan baik (Shaveta, 2023).

Kemudian sistem dirancang menggunakan *machine learning* seperti Random Forest karena efektif dalam menangani data *multivariabel*. Parameter yang akan dioptimasi disini yakni *n_estimators, max_depth, max_features* dan *min_samples_leaf*. Studi yang dilakukan oleh Fouodo, menekankan bahwa penting untuk melakukan perancangan sistem sejak awal agar tuning dapat berjalan dengan efisien dan terarah (Fouodo et al., 2023).

3.2.3 Implementasi Metode

Implementasi metode dimulai dari membangun model *Random Forest* menggunakan *library scikit-learn*. Random Forest terbukti efektif dalam melakukan klasifikasi hematologi sebagaimana dibuktikan oleh Yimer et al., yang menunjukkan model ini unggul dibandingkan dengan metode lain dalam mengklasifikasikan anemia (Yimer et al., 2025). Tahap implementasi melibatkan pembagian dataset menjadi data latih dan data uji, yang kemudian model dilatih menggunakan parameter awal. Setelah itu dilakukan hyperparameter tuning menggunakan *GridSearchCV*.

Proses tuning tersebut dilakukan secara terstruktur dan dijelaskan dalam studi yang dilakukan oleh Nurcahyo dan Sasongko, bahwa pencarian parameter yang dilakukan secara sistematis dapat meningkatkan performa model secara signifikan (Nurcahyo, J.A, & Sasongko, T.B. , 2023). Sebagaimana diuraikan oleh Nugroho, penggabungan banyak pohon keputusan dengan parameter optimal memberikan model yang lebih kuat dan tangguh terhadap *outlier* (Nugroho, 2025).

3.2.4 Skenario Pengujian

Sistem juga dirancang untuk mendukung beberapa skenario pengujian agar model dapat dievaluasi secara menyeluruh. Desain *multi-skenario* ini mengikuti prinsip penelitian seperti variasi fitur yang dapat memengaruhi model klasifikasi dan harus diuji secara sistematis sehingga didapatkan hasil yang akurat (Chen et al., 2020).

Tabel 3.3 Skenario Model

Fitur	Model A	Model B
Semua Fitur	**V**	-
Beberapa fitur (Selain Hemoglobin)	-	**V**

Pembagian skenario pengujian pada penelitian ini dilakukan untuk mengevaluasi performa model berdasarkan variasi fitur yang digunakan. Pada Model A, seluruh fitur CBC (Complete Blood Count) seperti Hb, MCV, MCH, dan MCHC dimasukkan sebagai input ke dalam model. Penggunaan semua fitur ini bertujuan untuk melihat performa maksimal model ketika seluruh informasi

yang tersedia dijadikan dasar prediksi. Skenario ini bertujuan untuk melihat performa maksimal model ketika seluruh informasi tersedia. *GridSearchCV* yang digunakan untuk mengoptimasi parameter sehingga diperoleh konfigurasi dengan hasil akurasi yang tinggi. Pendekatan ini selaras dengan pendapat Ahamad et al. (2023) yang menyatakan bahwa penggunaan fitur lengkap dapat meningkatkan sensitivitas model, meskipun berpotensi menyebabkan *overfitting* sehingga memerlukan proses tuning parameter yang tepat.

Pemilihan rentang nilai pada parameter yang telah dijelaskan sebelumnya pada subbab 3.1.8 didasarkan pada karakteristik dataset yang memiliki jumlah fitur ringkas setelah dilakukan reduksi fitur. Adapun rincian justifikasi pemilihan nilai parameter tersebut disajikan pada tabel 3.4 :

Tabel 3.4 Parameter Random Forest

Parameter	**Rentang Nilai (Grid)**	**Justifikasi (Alasan Pemilihan)**
n_estimators	[10, 20]	Menjaga efisiensi komputasi dan mencegah overfitting. Jumlah pohon yang sedikit sudah optimal untuk fitur yang terbatas.
max_depth	[3, 5]	Membatasi kedalaman pohon agar model melakukan generalisasi dengan baik dan tidak terjebak pada noise data.
max_features	[1]	Meningkatkan diversitas antar pohon dan meminimalisir korelasi antar pohon untuk meningkatkan kekuatan prediksi kolektif.
max_samples_leaf	[5]	Memastikan setiap keputusan akhir didasarkan pada sampel yang representatif guna menghindari sensitivitas terhadap outlier.

Sementara itu, Model B dilakukan dengan menghilangkan fitur Hemoglobin (Hb) untuk menguji robustnya model ketika salah satu fitur paling dominan tidak digunakan. Skenario ini penting karena Hb merupakan indikator utama dalam diagnosis anemia, sehingga penghapusannya dapat menggambarkan seberapa kuat model dalam memanfaatkan fitur lain sebagai dasar klasifikasi. Menurut Tepakhan et al. (2025), Hb memiliki pengaruh signifikan dalam membedakan sampel normal dan anemia, sehingga skenario ini menjadi evaluasi krusal terhadap kemampuan generalisasi model.

Tabel 3.5 Uji Model

Model	Sub	Training	Testing
A	A1	90	10
	A2	80	20
	A3	70	30
	A4	60	40
B	B1	90	10
	B2	80	20
	B3	70	30
	B4	60	40

Pembagian data pada Tabel 3.4 dilakukan untuk mengevaluasi performa model berdasarkan variasi rasio training dan testing. Pada model A maupun B menggunakan distribusi 90% data untuk training dan 10% untuk testing, yang umumnya menghasilkan model dengan kemampuan belajar yang kuat karena data latih lebih besar. Selanjutnya, porsi data untuk *testing* meningkat bertahap dari 20% hingga 40%, sementara porsi data *training* berkurang secara proporsional. Pendekatan bertahap ini memberikan gambaran mengenai sensitivitas model terhadap jumlah data yang digunakan untuk proses pelatihan.

Sistem dirancang untuk mendukung desain *multi-skenario* pengujian yang komprehensif, menggabungkan varasi *set* fitur input (Model A dan B) dan varasi rasio pembagian data *training* dan *testing* (90:10 hingga 60:40). Desain ini mengikuti prinsip penelitian untuk menguji model secara sistematis guna mendapatkan hasil akurat mengenai robustitas dan performa maksimal model klasifikasi (Chen et al., 2020).

Guna memperkuat validitas hasil, setiap konfigurasi diuji menggunakan metode 5-*Fold Cross Validation* pada tahap pencarian parameter terbaik (GridSearchCV). Kombinasi skenario ini penting tidak hanya untuk mengevaluasi sensitivitas model terhadap jumlah data latih dan kebergantungan pada fitur dominan seperti Hemoglobin, tetapi juga untuk membuktikan stabilitas model melalui pengujian berulang pada lipatan data yang berbeda. Dengan demikan, hasil akurasi yang diperoleh merupakan representasi performa model yang konsisten dan reliabel. Pembagian variasi tersebut disajikan lebih lanjut pada tabel 3.6 di bawah ini.

Tabel 3.6 Konfigurasi Skenario Pengujian Model

Model	**Sub**	**Fitur**	**Rasio (Train : Test)**	**Validasi Internal**
A	A1	Hemoglobin, MCH, MCV, MCHC	90 : 10	5-Fold CV
	A2		80 : 20	
	A3		70 : 30	
	A4		60 : 40	
B	B1	MCH, MCV, MCHC (Tanpa Hb)	90 : 10	
	B2		80 : 20	
	B3		70 : 30	
	B4		60 : 40	

Secara total, penelitian ini menguji delapan konfigurasi model yang berbeda untuk memastikan bahwa performa yang diperoleh stabil dan tidak hanya baik pada satu konfigurasi tertentu. Kombinasi skenario ini penting untuk mengevaluasi sensitivitas model terhadap jumlah data latih dan kebergantungan model pada fitur-fitur dominan. Pembagian variasi tersebut akan diuraikan lebih lanjut pada sub poin 3.2.4.1 dan 3.2.4.2.

3.2.4.1 Skenario Pengujian Model A (Semua Fitur)

3.2.4.1.1 Model A1 (90% Training : 10% Testing)

Skenario ini mewakili kondisi pengujian dengan set fitur terlengkap dan data latih terbanyak (90%), bertujuan untuk mengidentifikasi potensi akurasi tertinggi dan batas atas performa model *Random Forest*. Dengan menggunakan semua fitur CBC (Hb, MCV, MCH, MCHC) dan porsi data *training* yang melimpah, model memiliki kesempatan maksimal untuk mempelajari pola data secara mendalam. Porsi *testing* yang kecil (10%) memungkinkan evaluasi terhadap kemampuan model yang telah terlatih sangat baik. Hasil dari skenario ini diharapkan dapat menunjukkan performa model yang optimal ketika semua informasi diagnostik tersedia. Konfigurasi ini adalah titik awal untuk melihat kemampuan sensitivitas model, meskipun memerlukan tuning yang tepat untuk menghindari overfitting (Ahamad et al., 2023).

3.2.4.1.2 Model A2 (80% Training : 20% Testing)

Pengujian Model A2 dengan rasio 80% *training* dan 20% *testing* berfungsi sebagai tolak ukur *baseline* performa model menggunakan fitur lengkap. Rasio ini merupakan standar industri yang menawarkan keseimbangan yang baik antara porsi data untuk proses pembelajaran model dan porsi untuk evaluasi independen.

Model diuji untuk mengukur kemampuan generalisasinya ketika semua fitur dipertahankan, memastikan bahwa model tidak hanya menghafal data tetapi mampu membuat prediksi yang akurat pada data baru. Skenario ini akan memberikan gambaran yang stabil mengenai seberapa efektif *Random Forest* dan *GridSearch* bekerja dengan set fitur penuh. Hasilnya menjadi acuan utama untuk membandingkan semua skenario pengujian lainnya.

3.2.4.1.3 Model A3 (70% Training : 30% Testing)

Dalam skenario ini, Model A3 diuji dengan peningkatan porsi data *testing* (30%), yang secara otomatis mengurangi data latih menjadi 70%. Peningkatan ini memberikan evaluasi yang lebih ketat terhadap model yang dilatih dengan fitur lengkap. Tujuannya adalah menantang model untuk mempertahankan tingkat akurasi yang tinggi meskipun mengalami sedikit penurunan dalam jumlah sampel training yang digunakan. Skenario ini akan membantu menilai seberapa stabil performa model terhadap sedikit keterbatasan data latih. Jika performa tetap tinggi, ini mengindikasikan bahwa model memiliki kemampuan generalization yang kuat dan tidak terlalu sensitif terhadap pengurangan moderat pada data training.

3.2.4.1.4 Model A4 (60% Training : 40% Testing)

Skenario ini mewakili pengujian paling menantang bagi Model A4 dalam hal pembagian data, karena porsi *testing* menjadi terbesar (40%), dan data *training* paling terbatas (60%). Desain ini dirancang untuk menguji robustness model ketika data yang digunakan untuk pembelajaran sangat berkurang. Keberhasilan model mempertahankan performa yang baik di sini akan menunjukkan bahwa model mampu mengekstrak pola penting dengan sampel yang terbatas. Skenario ini sangat penting untuk mengidentifikasi potensi underfitting yang mungkin terjadi ketika data training terlalu sedikit.

3.2.4.2 Skenario Pengujian Model B (Fitur Tanpa Hemoglobin)

3.2.4.2.1 Model B1 (90% Training : 10% Testing)

Skenario ini menguji model tanpa fitur Hemoglobin (Hb), namun dengan data latih maksimal (90%). Tujuannya adalah untuk melihat sejauh mana fitur sekunder (MCV, MCH, MCHC) dapat mengompensasi hilangnya fitur dominan (Hb) ketika didukung oleh *volume* data training yang besar. Meskipun fitur utama dihilangkan, data latih yang melimpah memberikan kesempatan terbaik bagi model untuk menemukan pola tersembunyi dari fitur tersisa. Hasil dari skenario ini akan menunjukkan potensi akurasi tertinggi dari subset fitur ini.

3.2.4.2.2 Model B2 (80% Training : 20% Testing)

Pengujian Model B dengan rasio 80% *training* dan 20% *testing* ini sangat penting karena berfungsi sebagai perbandingan kuantitatif langsung terhadap Model A2 (80:20). Skenario ini secara spesifik mengukur dampak kerugian performa yang diakibatkan oleh penghapusan fitur Hb pada rasio pembagian data yang seimbang. Perbedaan metrik performa antara Model A dan B pada rasio ini akan mengukur secara empiris signifikansi fitur Hb pada klasifikasi Anemia, sebagaimana dikemukakan oleh Tepakhan et al. (2025). Hasilnya akan menjadi dasar untuk menilai seberapa kuat model dapat beroperasi hanya dengan fitur tersisa.

3.2.4.2.3 Model B3 (70% Training : 30% Testing)

Dalam skenario ini, Model B3 diuji pada kondisi ganda yang menantang: tanpa fitur dominan (Hb) dan evaluasi yang lebih ketat (30% testing). Peningkatan porsi testing menantang model untuk mempertahankan performa meskipun data latih berkurang. Tujuan utamanya adalah menguji stabilitas dan daya tahan model dalam memanfaatkan fitur sekunder. Skenario ini akan menggambarkan seberapa besar *trade-off* yang terjadi ketika model harus mengandalkan fitur sekunder di bawah tekanan evaluasi yang lebih ketat.

3.2.4.2.4 Model B4 (60% Training : 40% Testing)

Skenario ini merupakan pengujian paling ekstrem bagi Model B, menggabungkan tidak adanya fitur dominan (Hb) dengan data *training* paling sedikit (60%) dan data *testing* paling banyak (40%).

Skenario ini krusial untuk mengidentifikasi batas terendah performa model tanpa Hb. Hasilnya akan menentukan batas kemampuan generalisasi model pada kondisi data latih yang minim, menunjukkan robustness model secara keseluruhan.

Variasi rasio tersebut penting dilakukan karena jumlah data *training* yang lebih kecil dapat memengaruhi kemampuan model dalam mempelajari pola data, sementara jumlah data *testing* yang lebih besar memberikan ruang evaluasi yang lebih ketat. Dengan menguji berbagai rasio, penelitian dapat memastikan bahwa performa model tidak hanya baik pada satu konfigurasi tertentu, tetapi stabil pada kondisi evaluasi yang berbeda. Pendekatan ini juga membantu mengidentifikasi apakah model mengalami overfitting ketika data training terlalu sedikit atau justru tetap konsisten pada berbagai rasio pembagian data.

3.2.5 Evaluasi

Tahap evaluasi dilakukan untuk mengukur kinerja model yang telah dibangun menggunakan metode *Random Forest* dengan optimasi *hyperparameter* melalui *GridSearchCV*. Pada penelitian tersebut, menggunakan empat metriks utama, yakni *Accuracy, Precision, Recall,* dan *F1-Score*. Keempat perhitungan tersebut dilakukan dengan memanggil fungsi dari *library scikit-learn* sebagaimana ditunjukan pada sintaks 3.9:

Kode 3.9 Metrik Evaluasi

```
acc = accuracy_score(y_test, y_pred)
prec = precision_score(y_test, y_pred, average="weighted",
zero_division=0) rec = recall_score(y_test, y_pred, average="weighted",
zero_division=0)
f1 = f1_score(y_test, y_pred, average="weighted", zero_division=0)
```

Pertama, *Accuracy* dihitung untuk memberikan gambaran umum mengenai efektivitas model dalam memprediksi data yang tepat secara keseluruhan. Hal ini digunakan untuk akurasi dengan persamaan 3.3 :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (3.3)$$

Metrik ini merepresentasikan rasio total prediksi benar terhadap seluruh 35 jumlah data uji yang tersedia. Selanjutnya, untuk mengukur ketepatan klasifikasi pada kelas positif, digunakan *Precision* dengan persamaan 3.4:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3.4)$$

Presisi menjadi indikator seberapa akurat model dalam menentukan subjek yang benar-benar menderita Anemia di antara seluruh subjek yang diprediksi anemia oleh sistem. Sementara itu, untuk memastikan tidak ada kasus nemia yang terlewatkan yakni mengukur sejauh mana model mampu menemukan kasus anemia dengan benar, digunakan metrik *Recall* (sensitivitas) yang dihitung melalui persamaan 3.5:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3.5)$$

Recall menjadi sangat krusial dalam penelitian ini karena berfungsi untuk meminimalisir risiko *False Negative*, yaitu kondisi di mana penderita anemia salah terdiagnosa sebagai subjek sehat. Terakhir, untuk mengevaluasi performa model secara seimbang, terutama jika terdapat ketidakseimbangan distribusi data, digunakan *F1-Score* dengan persamaan 3.6:

$$\text{F1-Score} = 2\ x\ \frac{Precision\ x\ Recall}{Precision+\ Recall} \quad (3.6)$$

Penggunaan *F1-Score* memastikan bahwa model yang dibangun tidak hanya unggul pada salah satu metrik saja, melainkan memiliki stabilitas yang baik antara presisi dan *recall* (Cyntha et al., 2021).

Selain metriks kuantitatif, hasil klasifikasi yang diperoleh akan ditampilkan dalam bentuk *confusion matrix* untuk menggambarkan distribusi klasifikasi model dari setiap kelas, baik anemia ataupun tidak (Farida et al., 2025). Confusion matrix memberikan informasi yang lebih detail meliputi *True Positive (TP), True 36 Negative (TN), False Positive (FP),* dan *False Negative (FN)*.

Setelah melakukan uji skenario pada pemilihan fitur (Skenario A dan B) selanjutnya dilakukan pengujian ulang pada setiap *random state* yang berbeda (42, 123, 2025) untuk menilai konsistensi model terhadap variasi data latih dan uji, sehingga performa dapat lebih diandalkan (Hidayat et al., 2024). Tahap evaluasi sangat berperan dalam memastikan model yang dihasilkan benar-benar memiliki performa yang baik, konsisten dan siap untuk diaplikasikan lebih lanjut.

3.2.6 Analisis Hasil

Untuk mempermudah analisis, hasil evaluasi dari seluruh percobaan kemudian direkap dan dibandingkan. Hasil perbandingan rata-rata metriks dari setiap skenario digambarkan dalam bentuk tabel dan grafik, sehingga kontribusi masing-masing fitur terhadap kinerja model dapat diamati secara lebih jelas. Penggunaan beragam metriks ini penting untuk membantu peneliti dalam menilai kekuatan dan kelemahan model, serta menjadi dasar dalam pengambilan keputusan mengenai pemodelan terbaik.

BAB IV

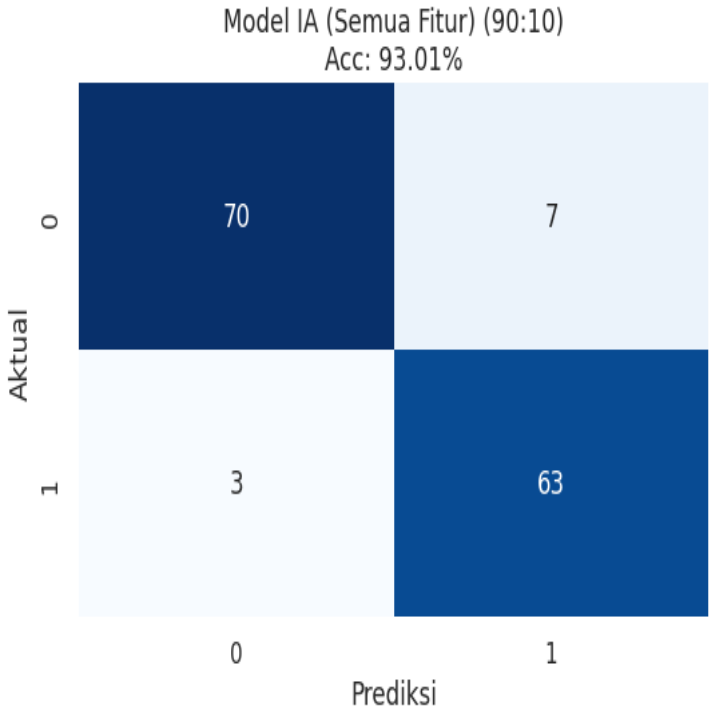
HASIL DAN PEMBAHASAN

Pada bab iv ini, penulis akan membahas mengenai hasil skenario model, metrik evaluasi dan integrasi islami.

4.1 Hasil Uji Coba Model A

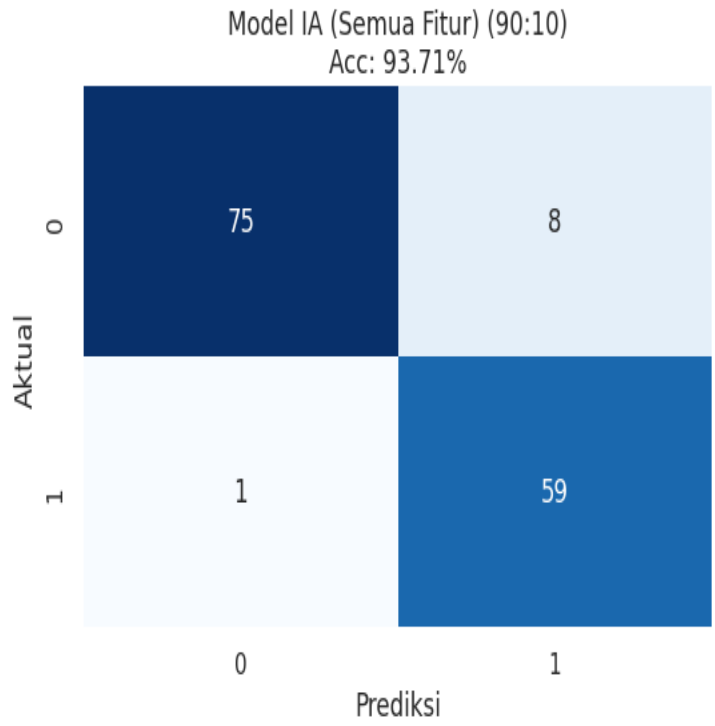
Sebelum dilakukan analisis komparatif antar model, terlebih dahulu dilakukan serangkaian pengujian terhadap Model A, yaitu model yang dibangun menggunakan seluruh fitur hematologi meliputi Hemoglobin, MCH, MCHC, dan MCV pada berbagai rasio pembagian data. Berdasarkan seluruh hasil pengujian pada skenario A1, A2, A3, dan A4 dengan variasi *random state* yang berbeda, performa model menunjukkan tingkat akurasi yang relatif tinggi dan konsisten, yaitu berada pada rentang 89,93% hingga 94,39%. Hasil terbaik diperoleh pada skenario A2 dengan rasio pembagian data 80:20, yang berhasil mencapai tingkat akurasi tertinggi sebesar 94,39%, sedangkan performa terendah terjadi pada skenario A3 sebesar 89,93%. Secara umum, hasil tersebut menunjukkan bahwa penggunaan seluruh fitur memberikan kemampuan yang baik bagi algoritma *Random Forest* dalam mempelajari pola hubungan antar parameter hematologi sehingga proses klasifikasi anemia dapat dilakukan secara optimal. Konsistensi performa yang diperoleh pada seluruh skenario pengujian juga mengindikasikan bahwa Model A memiliki tingkat stabilitas yang cukup baik dalam melakukan prediksi terhadap data yang belum pernah dipelajari sebelumnya.

4.1.1 Hasil Uji Coba Model A1



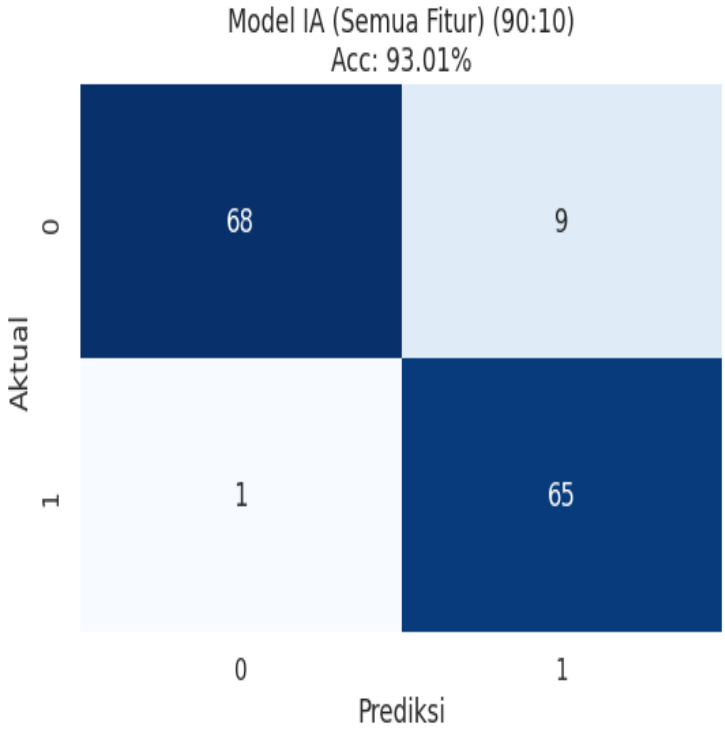
Gambar 4.1 *Confusion Matrix A1 random state 42*

Pada Gambar 4.1, model A1 dengan *random state* 42 menghasilkan akurasi sebesar 93,01% dimana didapatkan hasil *True Positive* (TP) sebesar 63 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 7 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 70 data pasien tidak terkena anemia yang diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebesar 3 data pasien anemia namun diprediksi tidak terkena anemia.



Gambar 4.2 *Confusion Matrix A1 random state 123*

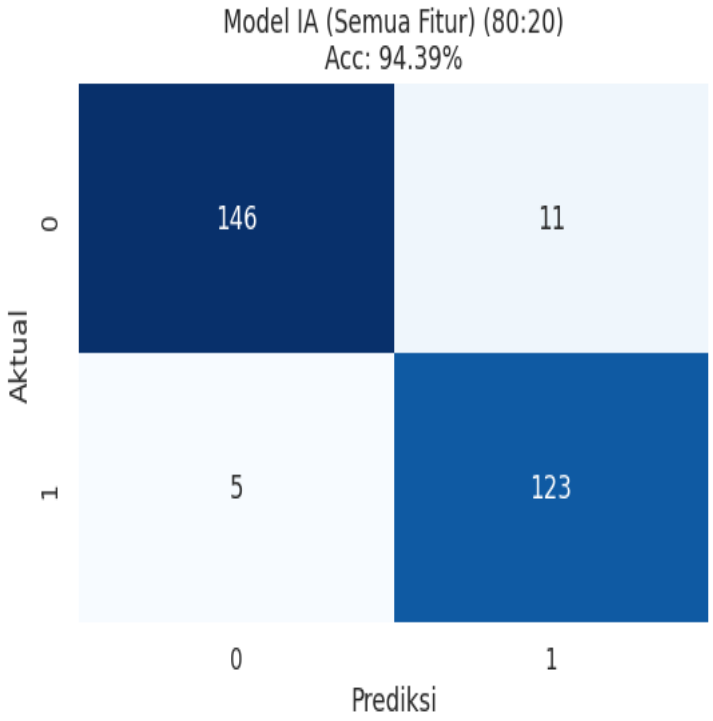
Selanjutnya pada Gambar 4.2, model A1 dengan *random state* 123 menghasilkan akurasi sebesar 93,71% dimana didapatkan hasil *True Positive* (TP) sebesar 59 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 8 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 75 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 1 data pasien namun diprediksi tidak terkena anemia.



Gambar 4.3 *Confusion Matrix A1 random state 2025*

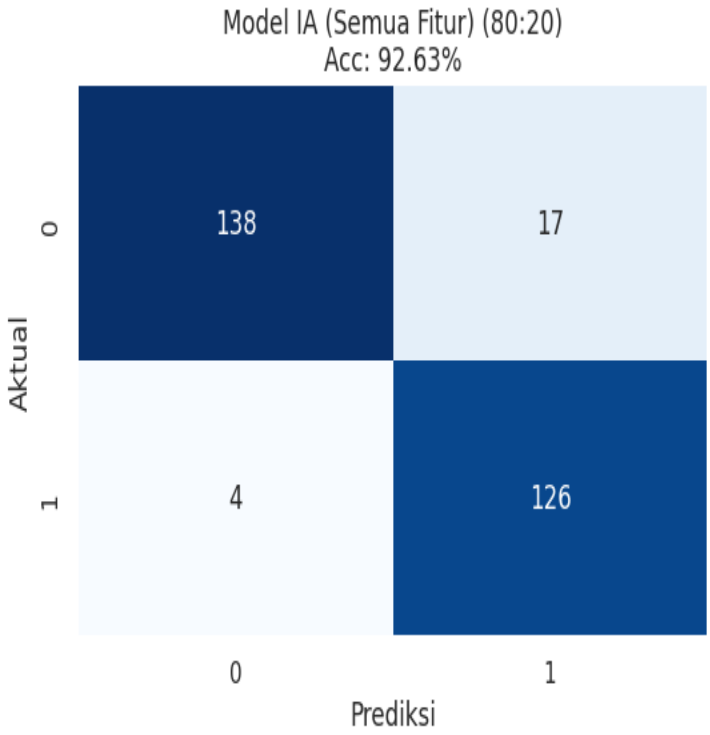
Selajutnya pada Gambar 4.3, model A1 dengan *random state* 2025 menghasilkan akurasi sebesar 93,01% dimana didapatkan hasil *True Positive* (TP) sebesar 65 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 9 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 68 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 1 data pasien namun diprediksi tidak terkena anemia.

4.1.2 Hasil Uji Coba Model A2



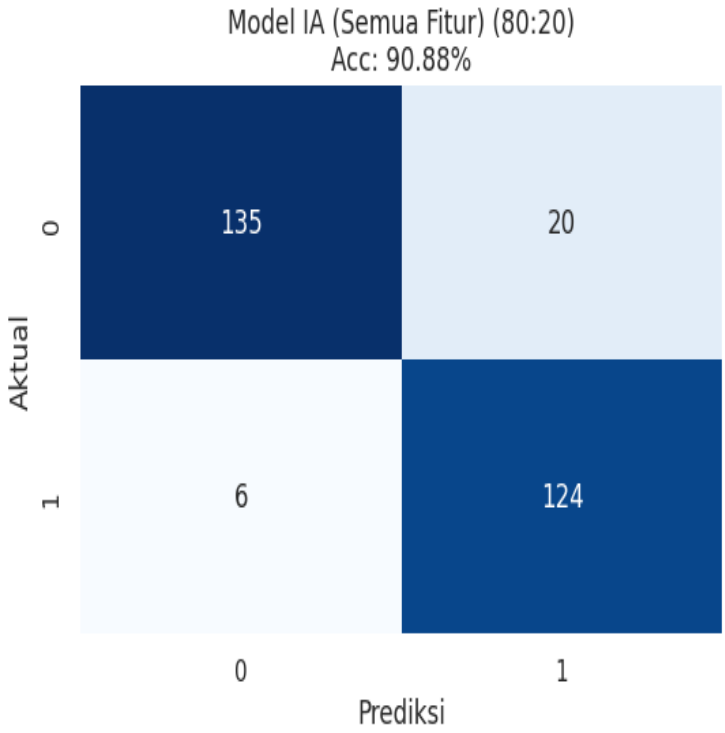
Gambar 4.4 *Confusion Matrix A2 random state 42*

Pada Gambar 4.4, model A2 dengan *random state* 42 menghasilkan akurasi sebesar 94,39% dimana didapatkan hasil *True Positive* (TP) sebesar 123 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 11 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 146 data pasien tidak terkena anemia yang diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebesar 5 data pasien anemia namun diprediksi tidak terkena anemia.



Gambar 4.5 *Confusion Matrix A2 random state 123*

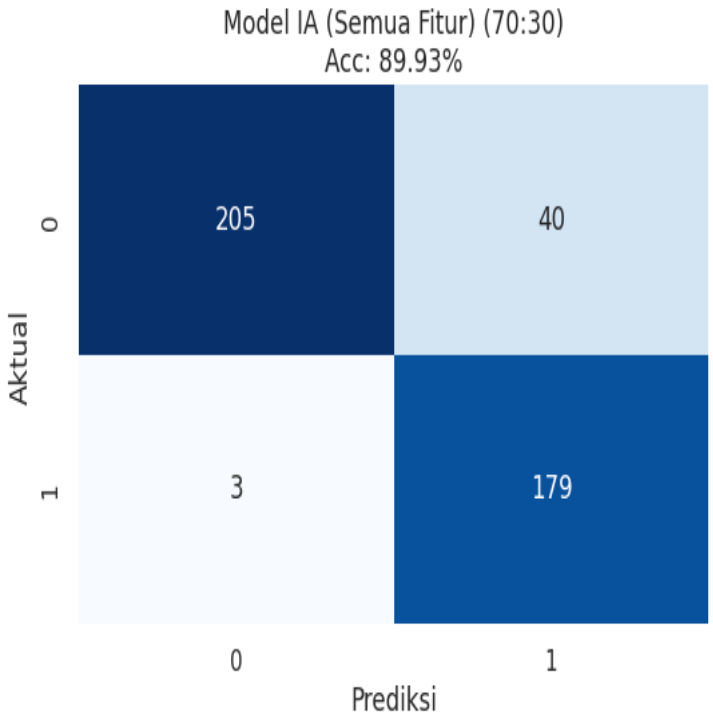
Selanjutnya pada Gambar 4.5 model A2 dengan *random state* 123 menghasilkan akurasi sebesar 92,63% dimana didapatkan hasil *True Positive* (TP) sebesar 126 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 17 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 138 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 4 data pasien namun diprediksi tidak terkena anemia.



Gambar 4.6 *Confusion Matrix A2 random state 2025*

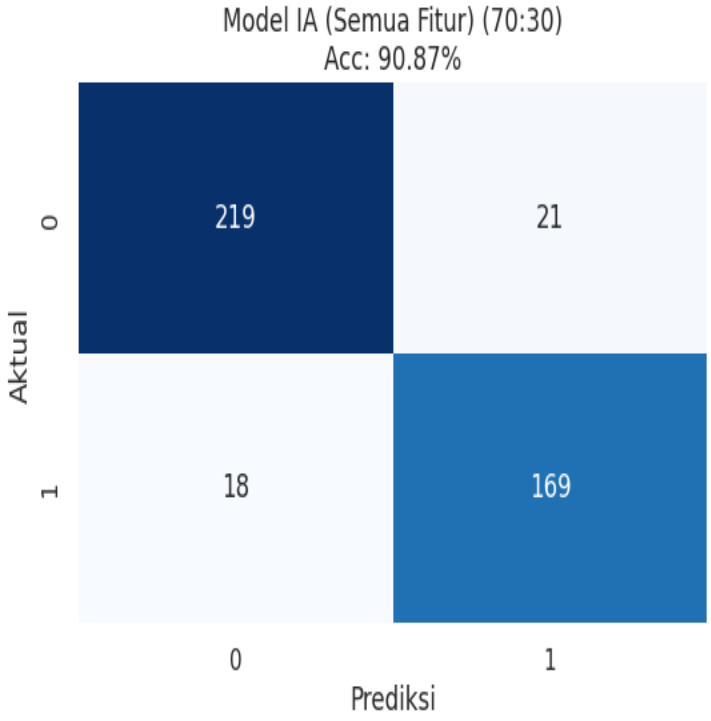
Selanjutnya pada Gambar 4.6, model A2 dengan *random state* 2025 menghasilkan akurasi sebesar 90,88% dimana didapatkan hasil *True Positive* (TP) sebesar 124 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 20 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 135 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 6 data pasien namun diprediksi tidak terkena anemia.

4.1.3 Hasil Uji Coba Model A3



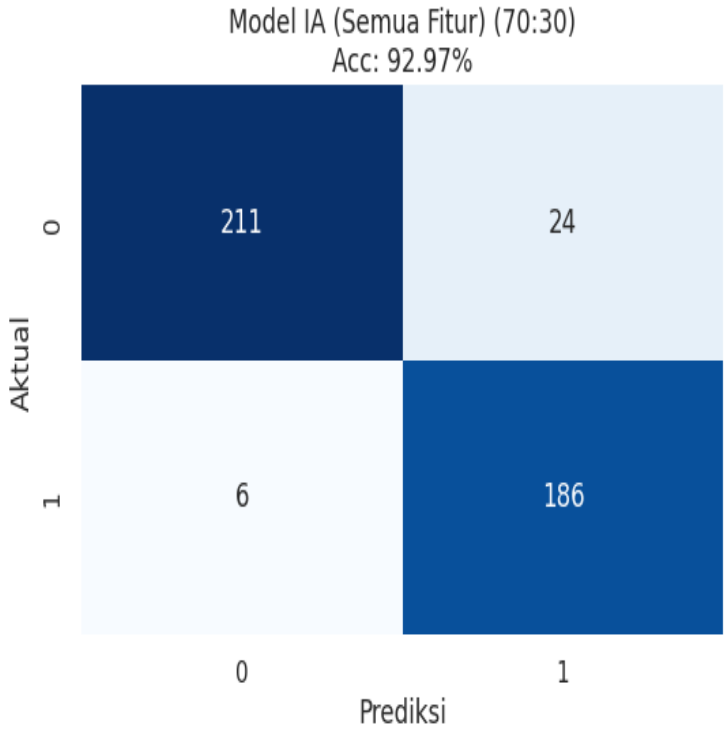
Gambar 4.7 Confusion Matrix A3 random state 42

Pada Gambar 4.7, model A3 dengan *random state* 42 menghasilkan akurasi sebesar 89,93% dimana didapatkan hasil *True Positive* (TP) sebesar 179 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 40 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 205 data pasien tidak terkena anemia yang diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebesar 3 data pasien anemia namun diprediksi tidak terkena anemia.



Gambar 4.8 Confusion Matrix A3 random state 123

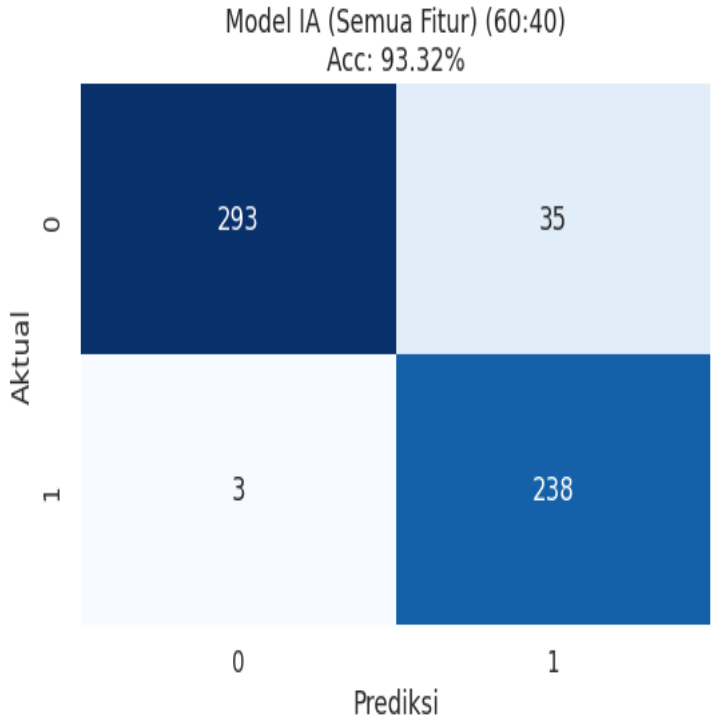
Selanjutnya pada Gambar 4.8, model A3 dengan *random state* 123 menghasilkan akurasi sebesar 90,87% dimana didapatkan hasil *True Positive* (TP) sebesar 169 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 21 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 219 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 18 data pasien namun diprediksi tidak terkena anemia.



Gambar 4.8 Confusion Matrix A3 random state 2025

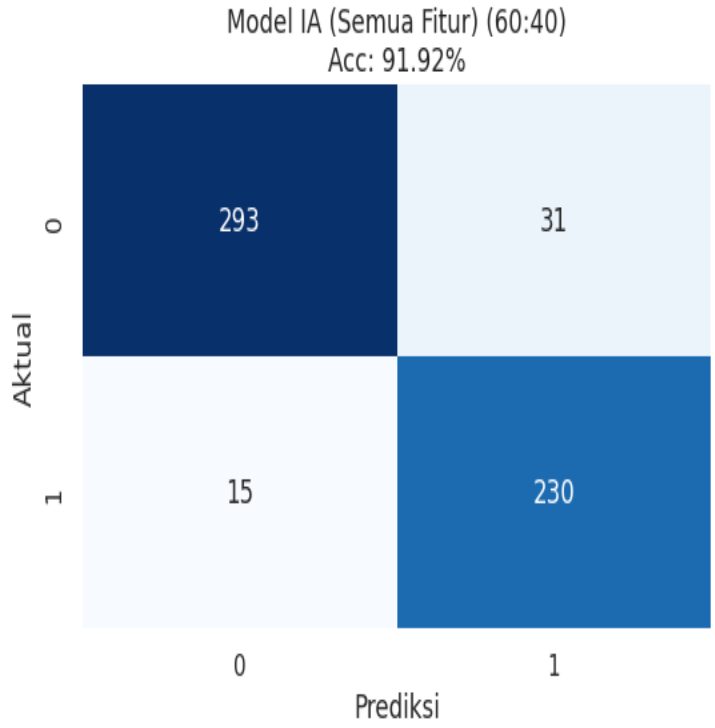
Selanjutnya pada Gambar 4.24, model A3 dengan *random state* 2025 menghasilkan akurasi sebesar 92,97% dimana didapatkan hasil *True Positive* (TP) sebesar 186 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 24 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 211 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 6 data pasien namun diprediksi tidak terkena anemia.

4.1.4 Hasil Uji Coba Model A4



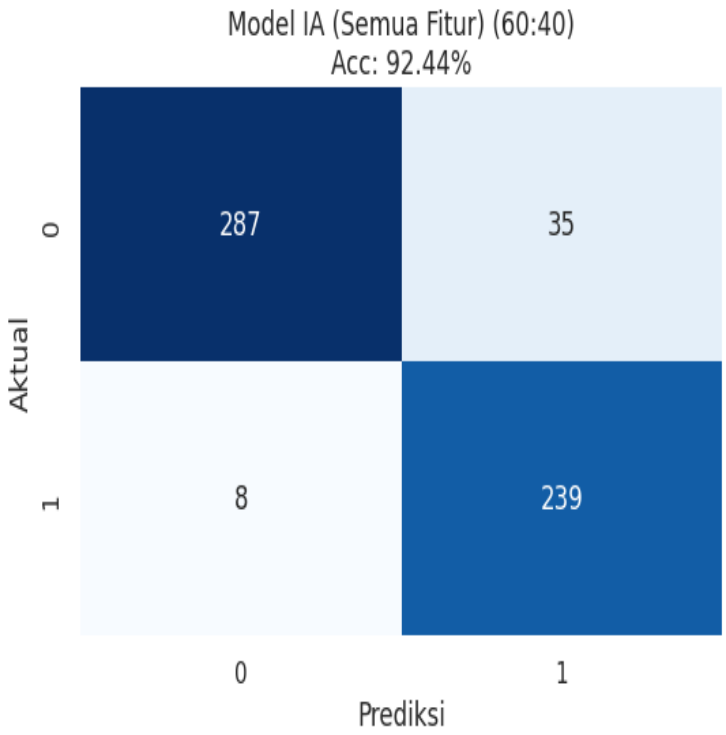
Gambar 4.9 Confusion Matrix A3 random state 42

Pada Gambar 4.9, model A4 dengan *random state* 42 menghasilkan akurasi sebesar 93,32% dimana didapatkan hasil *True Positive* (TP) sebanyak 238 data pasien anemia yang dipresiksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 35 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 293 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 3 data pasien namun diprediksi tidak terkena anemia.



Gambar 4.10 Confusion Matrix A4 random state 123

Selanjutnya pada Gambar 4.10, model A4 dengan *random state* 123 menghasilkan akurasi sebesar 91,92% dimana didapatkan hasil *True Positive* (TP) sebesar 230 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 31 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 293 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 15 data pasien namun diprediksi tidak terkena anemia.



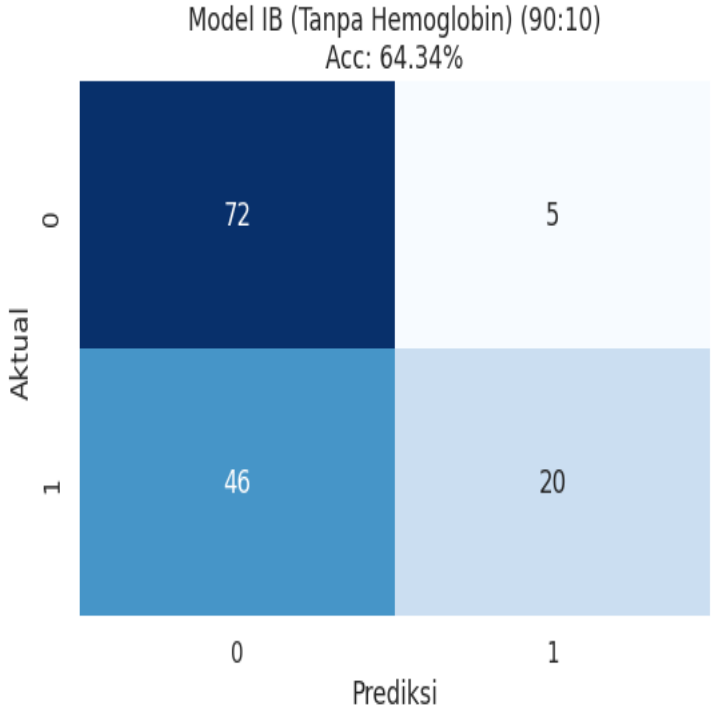
Gambar 4.11 Confusion Matrix A4 random state 2025

Selanjutnya pada Gambar 4.11, model A4 dengan *random state* 2025 menghasilkan akurasi sebesar 92,44% dimana didapatkan hasil *True Positive* (TP) sebesar 239 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 35 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 287 data pasien tidak terkena anemia yang dipresiksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 8 data pasien namun diprediksi tidak terkena anemia.

4.2 Hasil Uji Coba Model B

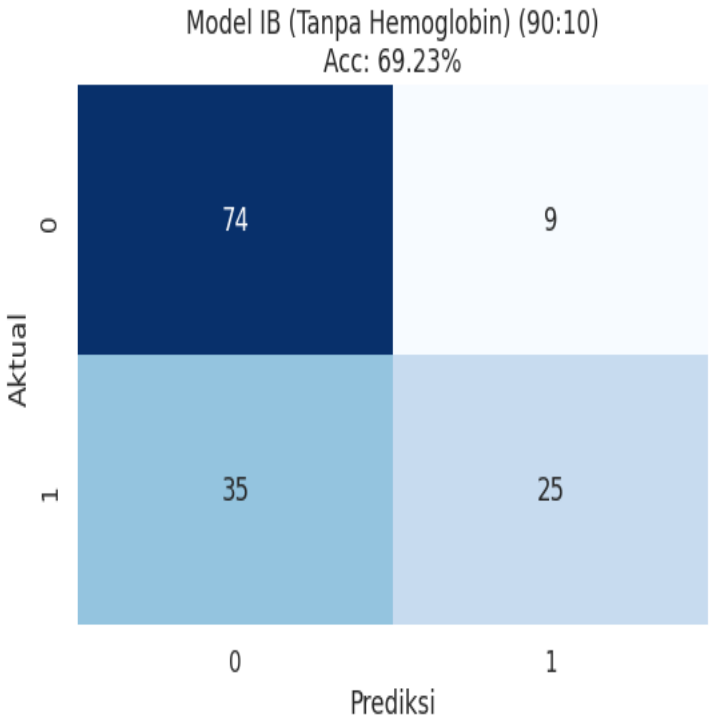
Sama seperti pengujian sebelumnya, serangkaian eksperimen juga dilakukan pada Model B, yaitu model yang dibangun tanpa menggunakan fitur Hemoglobin untuk mengukur pengaruh penghapusan fitur utama terhadap performa klasifikasi. Berdasarkan keseluruhan hasil pengujian pada skenario B1, B2, B3, dan B4 dengan variasi *random state* yang berbeda, performa model mengalami penurunan yang cukup signifikan dibandingkan Model A, dengan tingkat akurasi hanya berada pada rentang 61,75% hingga 69,23%. Performa terbaik diperoleh pada skenario B1 dengan akurasi tertinggi sebesar 69,23%, sedangkan hasil terendah terjadi pada skenario B2 dengan akurasi sebesar 61,75%. Penurunan performa yang terjadi secara konsisten pada seluruh skenario menunjukkan bahwa penghapusan fitur Hemoglobin menyebabkan algoritma *Random Forest* kehilangan informasi penting yang dibutuhkan dalam membedakan kelas anemia dan non-anemia secara akurat. Hasil ini semakin memperkuat dugaan bahwa Hemoglobin merupakan fitur dominan yang memiliki kontribusi paling besar dalam proses klasifikasi pada dataset anemia sehingga keberadaannya sangat berpengaruh terhadap kualitas prediksi model.

4.2.1 Hasil Uji Coba Model B1



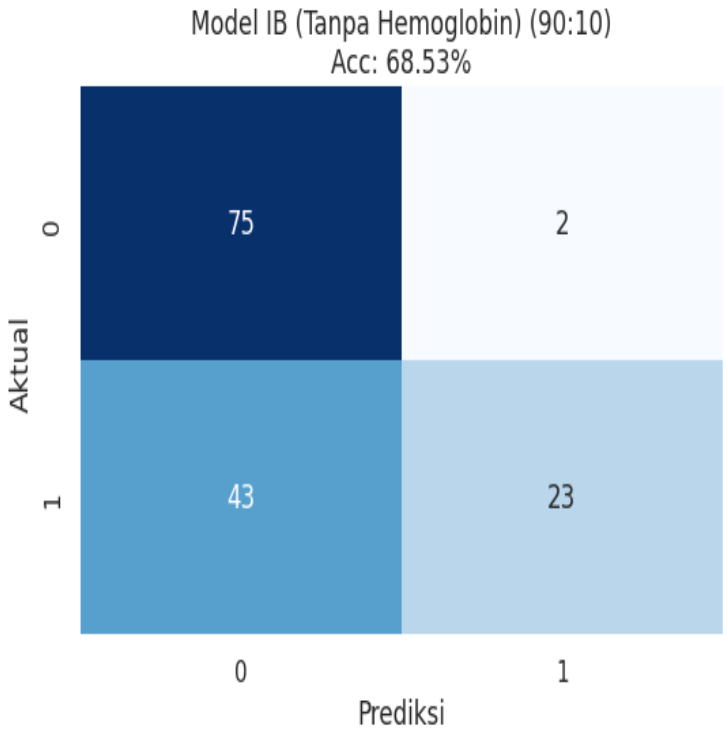
Gambar 4.12 Confusion Matrix B1 random state 42

Pada Gambar 4.12 terlihat bahwa Model B1 (Tanpa Hemoglobin), peforma model mengalami penurunan tajam dengan akurasi hanya sebesar 64,34% dimana didapatkan hasil *True Positive* (TP) sebesar 20 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 5 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 72 data pasien tidak terkena anemia yang diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebesar 46 data pasien anemia namun diprediksi tidak terkena anemia.



Gambar 4.13 Confusion Matrix B1 random state 123

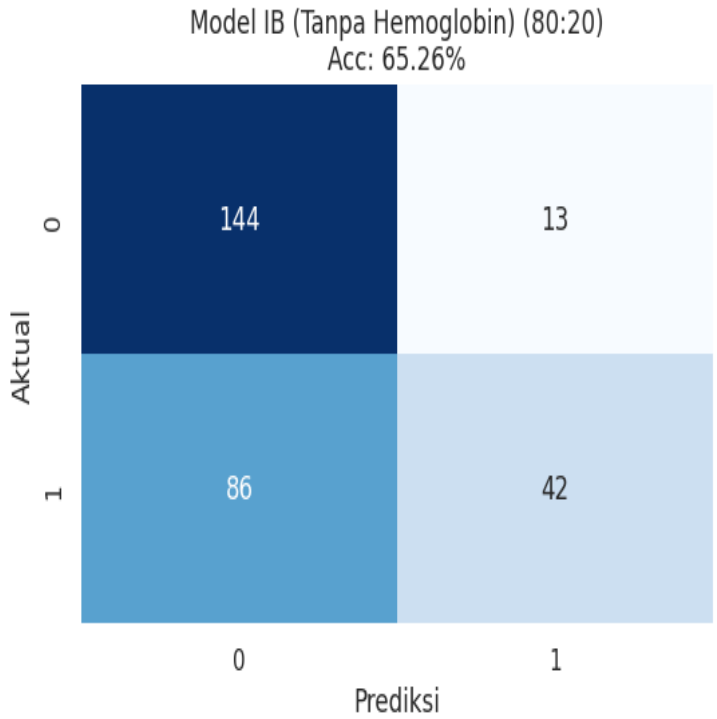
Selanjutnya pada Gambar 4.13 terlihat bahwa Model B1 (Tanpa Hemoglobin), peforma model mengalami penurunan tajam dengan akurasi sebesar 69,23% dimana didapatkan hasil *True Positive* (TP) sebesar 25 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 9 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 74 data pasien tidak terkena anemia diprediksi benar (non anemia). *False Negative* (FN) menghasilkan sebesar 35 data pasien anemia namun diprediksi terkena anemia.



Gambar 4.14 Confusion Matrix B1 random state 2025

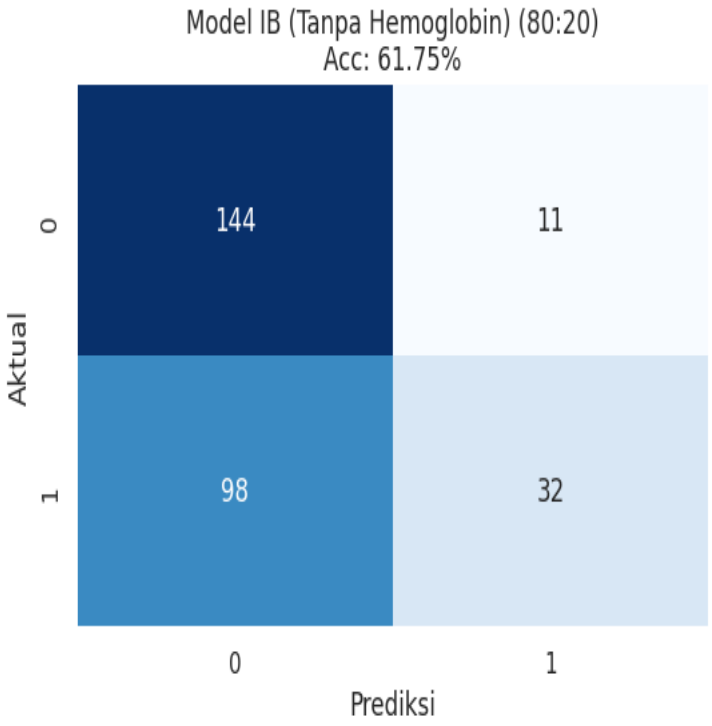
Selanjutnya pada Gambar 4.14 terlihat bahwa Model B1 (Tanpa Hemoglobin), peforma model mengalami penurunan dengan akurasi sebesar 68,53% dimana didapatkan hasil *True Positive* (TP) sebesar 23 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 2 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 75 data pasien tidak terkena anemia diprediksi benar (non anemia). *False Negative* (FN) menghasilkan sebesar 43 data pasien anemia namun diprediksi terkena anemia.

4.2.2 Hasil Uji Coba Model B2



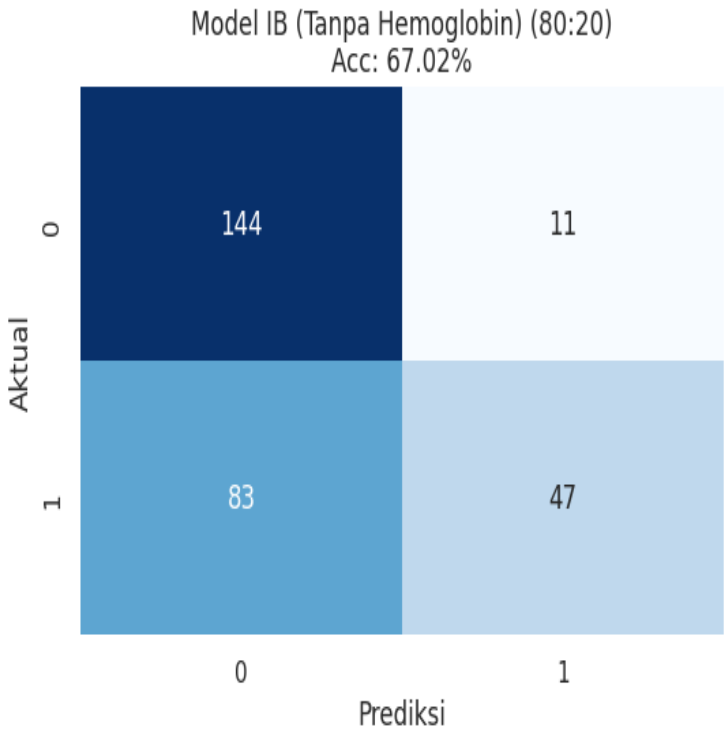
Gambar 4.15 Confusion Matrix B2 random state 42

Pada Gambar 4.15 terlihat bahwa Model B2 (Tanpa Hemoglobin) didapatkan hasil akurasi sebesar 65,26% dimana didapatkan hasil *True Positive* (TP) sebesar 42 data pasien anemia yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 13 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 144 data pasien tidak terkena anemia yang diprediksi benar (non-anemia). *False Negative (FN)* menghasilkan sebesar 86 data pasien anemia namun diprediksi tidak terkena anemia.



Gambar 4.16 Confusion Matrix B2 random state 123

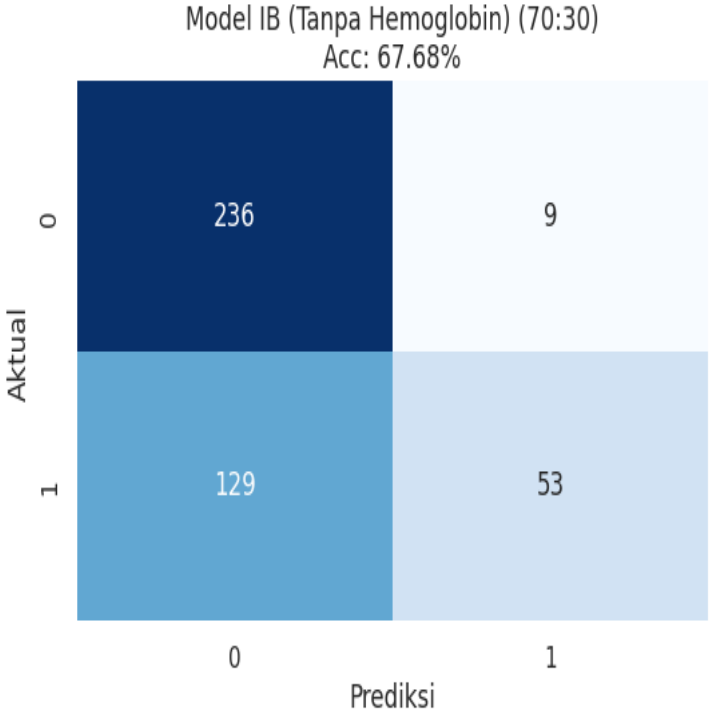
Sebaliknya, pada Gambar 4.16 terlihat bahwa Model B2 (Tanpa Hemoglobin), peforma model mengalami penurunan tajam dengan akurasi sebesar 61,75% dimana didapatkan hasil *True Positive* (TP) sebesar 32 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 11 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 144 data pasien tidak terkena anemia diprediksi benar (non anemia). *False Negative* (FN) menghasilkan sebesar 98 data pasien anemia namun diprediksi terkena anemia.



Gambar 4.17 Confusion Matrix B2 random state 2025

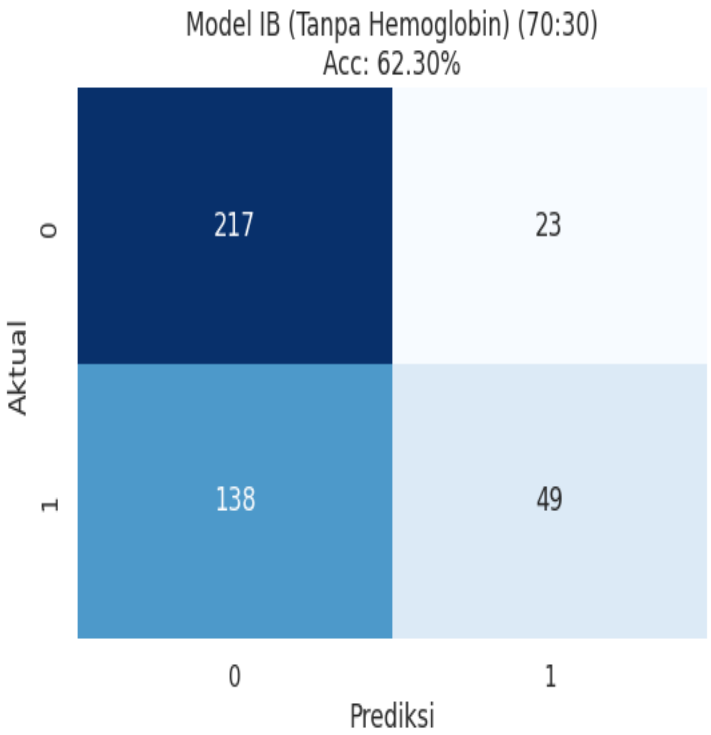
Selanjutnya pada Gambar 4.17 terlihat bahwa Model B (Tanpa Hemoglobin), peforma model mengalami penurunan tajam dengan akurasi sebesar 67,02% dimana didapatkan hasil *True Positive* (TP) sebesar 47 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 11 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 144 data pasien tidak terkena anemia diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebesar 83 data pasien anemia namun diprediksi terkena anemia.

4.2.3 Hasil Uji Coba Model B3



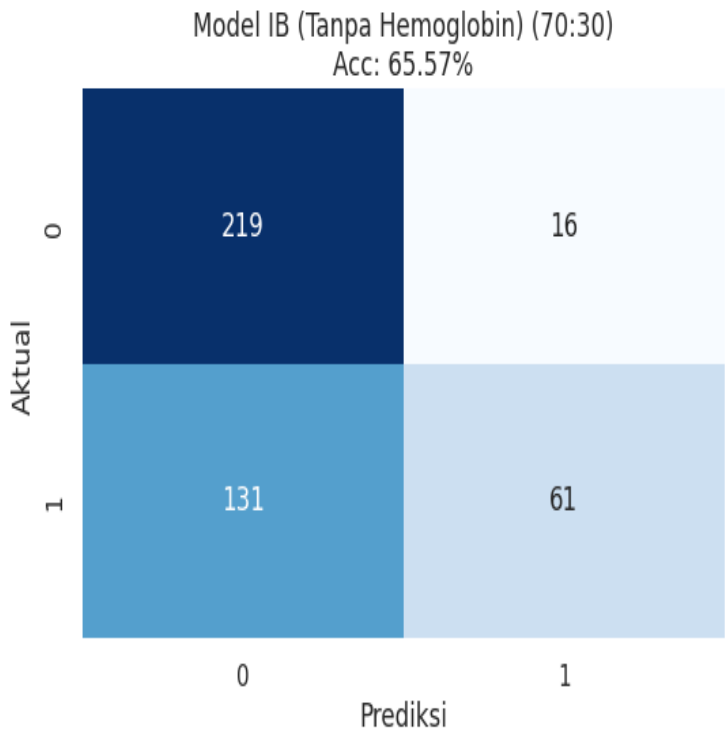
Gambar 4.18 Confusion Matrix B3 random state 42

Pada Gambar 4.18 terlihat bahwa model B3 (Tanpa Hemoglobin) menghasilkan akurasi sebesar 67,68% dimana didapatkan hasil *True Positive* (TP) sebesar 53 data pasien anemia yang diprediksi benar (terkena anemia), *False Positive* (FP) menghasilkan sebesar 9 data pasien tidak terkena anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 236 data pasien tidak terkena anemia yang diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebanyak 129 data pasien anemia namun diprediksi tidak terkena anemia.



Gambar 4.19 Confusion Matrix B3 random state 123

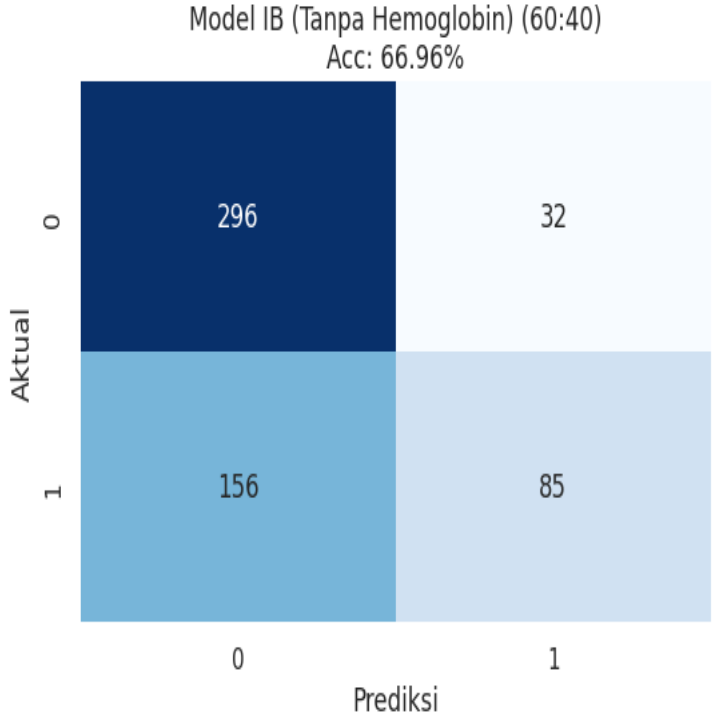
Selanjutnya pada Gambar 4.19 terlihat bahwa model B3 (Tanpa Hemoglobin) dengan akurasi sebesar 62,30% dimana didapatkan hasil *True Positive* (TP) sebesar 49 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 23 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 217 data pasien tidak terkena anemia diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebesar 138 data pasien anemia namun diprediksi terkena anemia.



Gambar 4.20 Confusion Matrix B3 random state 2025

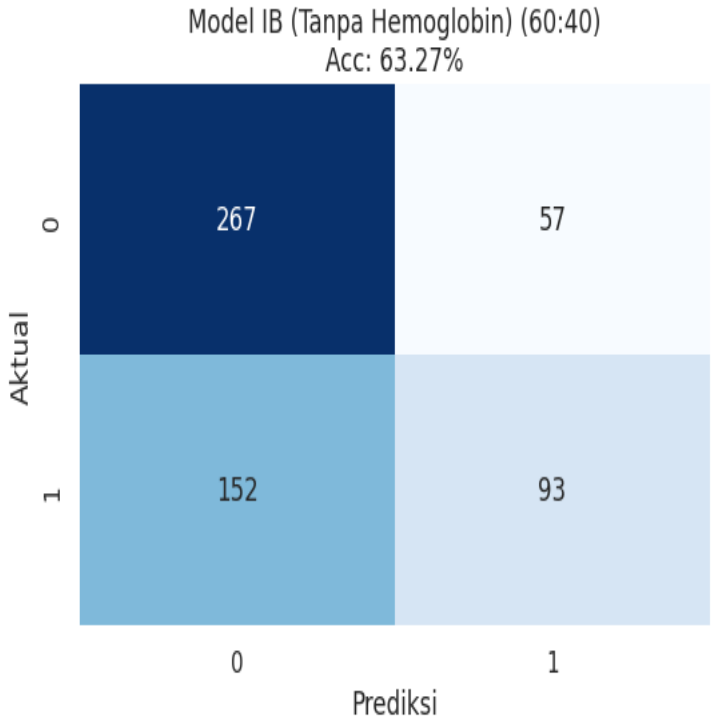
Selanjutnya pada Gambar 4.20 terlihat bahwa Model B3 (Tanpa Hemoglobin), peforma model mengalami penurunan tajam dengan akurasi sebesar 65,57% dimana didapatkan hasil *True Positive* (TP) sebesar 61 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 16 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 219 data pasien tidak terkena anemia diprediksi benar (non anemia). *False Negative* (FN) menghasilkan sebesar 131 data pasien anemia namun diprediksi terkena anemia.

4.2.4 Hasil Uji Coba Model B4



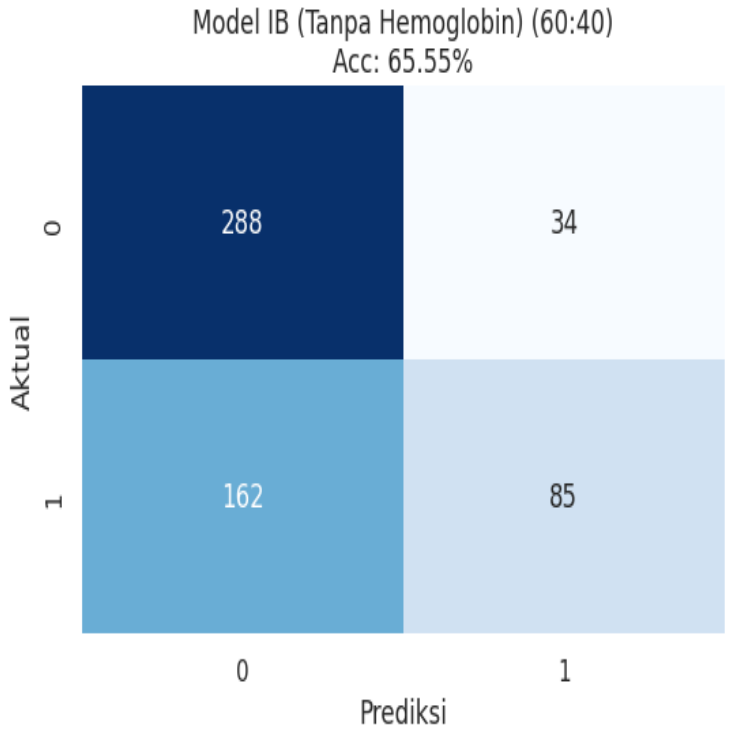
Gambar 4.21 Confusion Matrix B4 random state 42

Pada Gambar 4.21 terlihat bahwa model B4 (Tanpa Hemoglobin) menghasilkan akurasi sebesar 66,96% dimana didapatkan hasil *True Positive* (TP) sebesar 85 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 32 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 296 data pasien tidak terkena anemia diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebesar 156 data pasien anemia namun diprediksi terkena anemia.



Gambar 4.22 Confusion Matrix B4 random state 123

Selanjutnya pada Gambar 4.22 terlihat bahwa model B4 (Tanpa Hemoglobin) menghasilkan akurasi sebesar 63,27% dimana didapatkan hasil *True Positive* (TP) 49 sebesar 93 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 57 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 267 data pasien tidak terkena anemia diprediksi benar (non-anemia*)*. *False Negative* (FN) menghasilkan sebesar 152 data pasien anemia namun diprediksi terkena anemia.

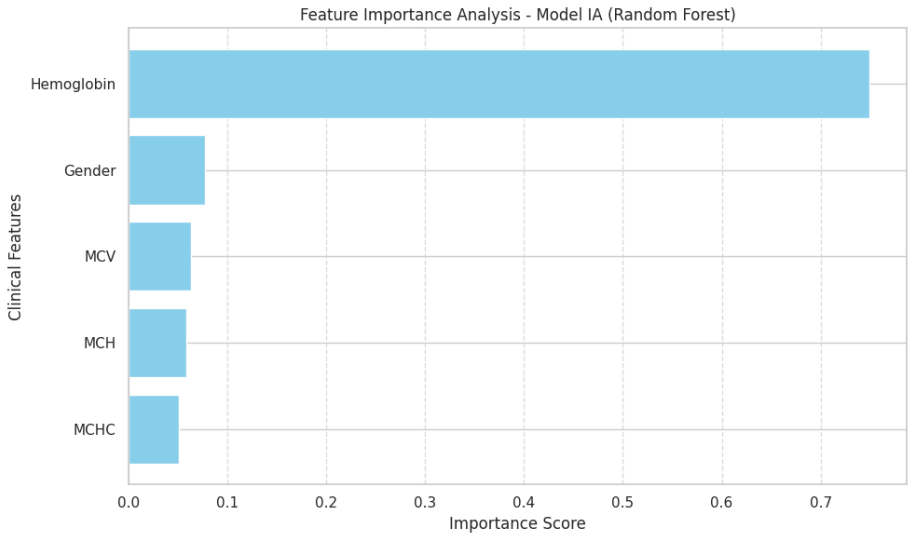


Gambar 4.23 Confusion Matrix B4 random state 2025

Selanjutnya pada Gambar 4.23 terlihat bahwa model B4 (Tanpa Hemoglobin), peforma model mengalami penurunan tajam dengan akurasi sebesar 65,55% dimana didapatkan hasil *True Positive* (TP) sebesar 85 data pasien yang diprediksi benar (terkena anemia). *False Positive* (FP) menghasilkan sebesar 34 data pasien tidak anemia namun diprediksi anemia. *True Negative* (TN) menghasilkan sebesar 288 data pasien tidak terkena anemia diprediksi benar (non-anemia). *False Negative* (FN) menghasilkan sebesar 162 data pasien anemia namun diprediksi terkena anemia.

4.3 Pembahasan Feature Importance

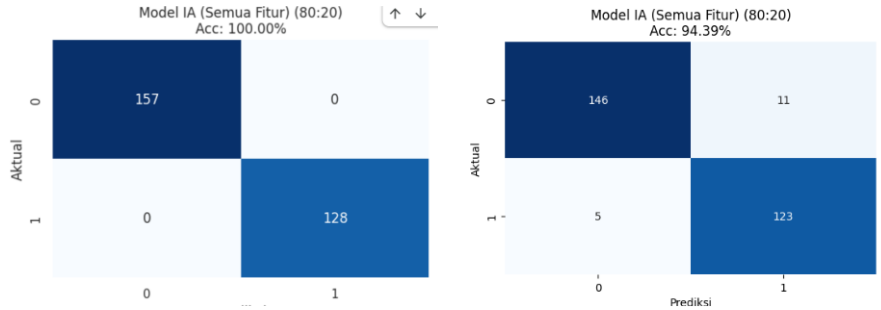
Berdasarkan hasil pengujian yang telah dilakukan pada masing-masing skenario, dilakukan analisis lanjut untuk melihat kontribusi masing-masing fitur terhadap hasil klasifikasi model. *Random Forest* memungkinkan penghitungan nilai *feature importance* yang menunjukkan seberapa besar pengaruh sebuah variabel dalam menentukan keputusan pada pohon-pohon keputusan yang terbentuk. Hal ini dilakukan untuk memvalidasi temuan pada sub-bab sebelumnya mengenai penyebab penurunan akurasi yang signifikan pada Model B.



Gambar 4.28 Feature Importance

Hasil analisis pada Gambar 4.28 menunjukkan bahwa variabel Hemoglobin secara konsisten menempati urutan pertama dengan bobot kepentingan yang jauh melampaui fitur lainnya. Dominasi nilai kepentingan Hemoglobin ini menjelaskan mengapa Model A mampu menjaga akurasi di atas 90% pada seluruh random state dan rasio data. Sebaliknya, variabel lain seperti MCH, MCV, dan MCHC memiliki nilai kepentingan yang jauh lebih rendah.

Hal ini membuktikan secara teknis bahwa fitur-fitur tersebut bersifat sebagai parameter pendukung dan tidak memiliki daya pembeda (*discriminative power*) yang cukup kuat untuk berdiri sendiri tanpa kehadiran fitur Hemoglobin. Ketergantungan model yang sangat tinggi terhadap satu fitur kunci ini menjadi alasan utama mengapa Model B mengalami bias dan peningkatan angka *False Negative* yang sangat tajam di seluruh skenario pengujian.



Gambar 4.31 Pengaruh Seleksi Fitur

Selain analisis fitur tersebut, hasil eksperimen awal, penyertaan fitur Gender menyebabkan model mengalami indikasi overfitting yang ekstrem atau perfect classification (akurasi mencapai 100%) seperti pada gambar 4.31, dimana gambar sebelah kiri adalah sebelum dilakukan seleksi fitur sedangkan gambar sebelah kanan kondisi setelah dilakukan seleksi fitur. Hal ini menunjukkan adanya korelasi yang terlalu tinggi dan tidak representatif antara jenis kelamin dengan target kelas pada dataset yang digunakan. Untuk menjaga objektivitas model dan memastikan bahwa klasifikasi didasarkan pada parameter klinis biokimia yang valid secara medis, maka fitur Gender dieliminasi (Neufang et al., 2025). Keputusan ini diambil agar model memiliki kemampuan generalisasi yang lebih baik saat dihadapkan pada data baru di luar dataset penelitian, serta untuk menghindari bias yang dapat mendistorsi hasil deteksi anemia berdasarkan parameter laboratorium utama (Pagano et al., 2023).

4.4 Pembahasan Confusion Matrix

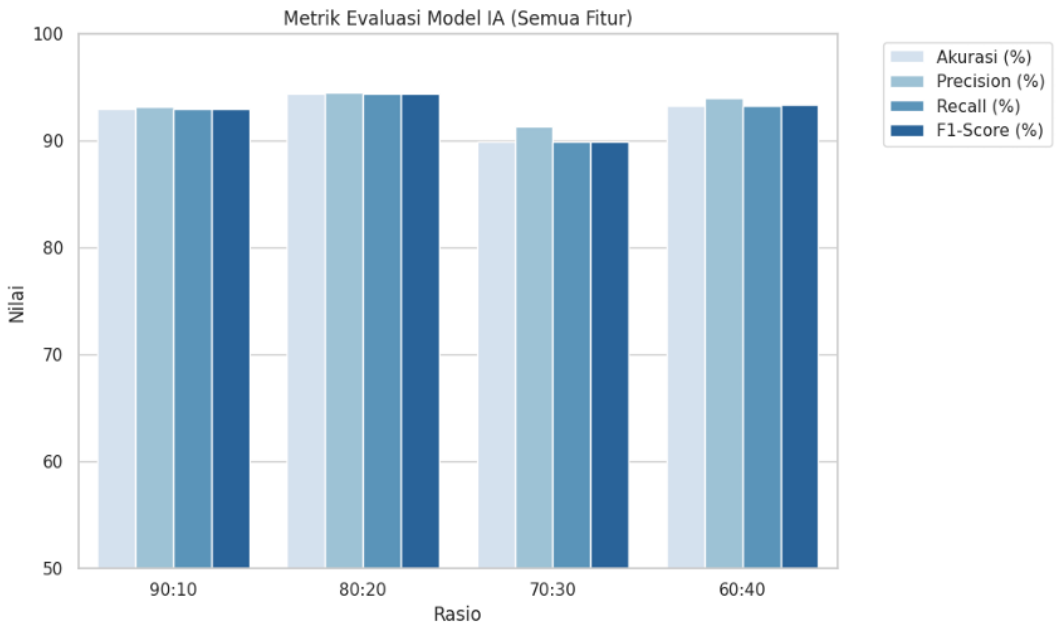
Berdasarkan *confusion matrix* pada Gambar 4.1 pada subbab sebelumnya, model berhasil memprediksi X data dengan benar. Dari angka-angka tersebut, penulis kemudian melakukan perhitungan metrik evaluasi untuk mendapatkan nilai *Accuracy, Precision, Recall*, dan *F1-Score* untuk melihat performa model secara lebih terukur.

Tabel 4.1 Metrik Evaluasi A

Rasio (Train:Test)	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
90:10	93,02	93.17	93.01	93.02
80:20	94,39	94,49	94,39	94,40
70:30	89,93	91,39	89,93	89,98
60:40	93,32	93,99	93,32	93,36

Berdasarkan Tabel 4.1, hasil pengujian oodel A yang menggunakan seluruh fitur (Hemoglobin, MCH, MCV, dan MCHC) menunjukkan performa yang sangat luar biasa dan stabil di berbagai skenario rasio data. Akurasi tertinggi berhasil dicapai pada rasio data 80:20, yakni sebesar 94,39%. Hal tersebut mengindikasikan bahwa pembagian data dengan komposisi 80% untuk pelatihan dan 20% untuk Pengujian memberikan ruang yang cukup bagi *Random Forest* untuk mempelajari pola klasifikasi Anemia secara optimal. Meskipun data belum melalui proses penyeimbangan (*oversampling*), tingginya nilai akurasi pada rasio lainnya yang secara konsisten berada di rentang 89% hingga 93%, membuktikan bahwa fitur-fitur yang dipilih memiliki korelasi yang sangat kuat dengan target klasifikasi.

Selanjutnya, jika ditinjau berdasarkan metrik evaluasi lainnya, model A pada rasio 80:20 juga menghasilkan nilai *Precision* sebesar 94,49%, Recall sebesar 94,39%, dan *F1-Score* sebesar 94,40%. Keselarasan angka antara keempat metrik ini menandakan bahwa model tidak hanya unggul dalam menebak secara keseluruhan (akurasi), tetapi juga memiliki keseimbangan yang sangat baik dalam meminimalisir kesalahan prediksi, baik itu *False Positive* maupun *False Negative*. Nilai *F1-Score* yang mencapai angka 94% menjadi bukti kuat bahwa model tetap mampu mengenali kelas minoritas dengan akurat meskipun terdapat potensi ketidakseimbangan data awal. Hal ini memperkuat kesimpulan bahwa kombinasi fitur hematologi yang digunakan dalam model A merupakan parameter yang valid untuk deteksi dini penyakit anemia menggunakan pendekatan machine learning.



Gambar 4.29 Metrik Evaluasi Model A

Berdasarkan grafik hasil pengujian pada Gambar 4.29, terlihat bahwa skenario A (Model A) yang menggunakan seluruh fitur (Hemoglobin, MCH, MCV, dan MCHC) menunjukkan performa yang sangat impresif dan konsisten di

seluruh rasio data. Akurasi tertinggi berhasil dicapai pada rasio 80:20, di mana model mampu mengenali pola data dengan optimal dibandingkan rasio lainnya. Stabilitas ini membuktikan bahwa kombinasi fitur-fitur hematologi tersebut memiliki relevansi yang tinggi dalam mendeteksi kondisi anemia, sehingga *Random Forest* dapat melakukan klasifikasi dengan tingkat kesalahan yang sangat rendah.

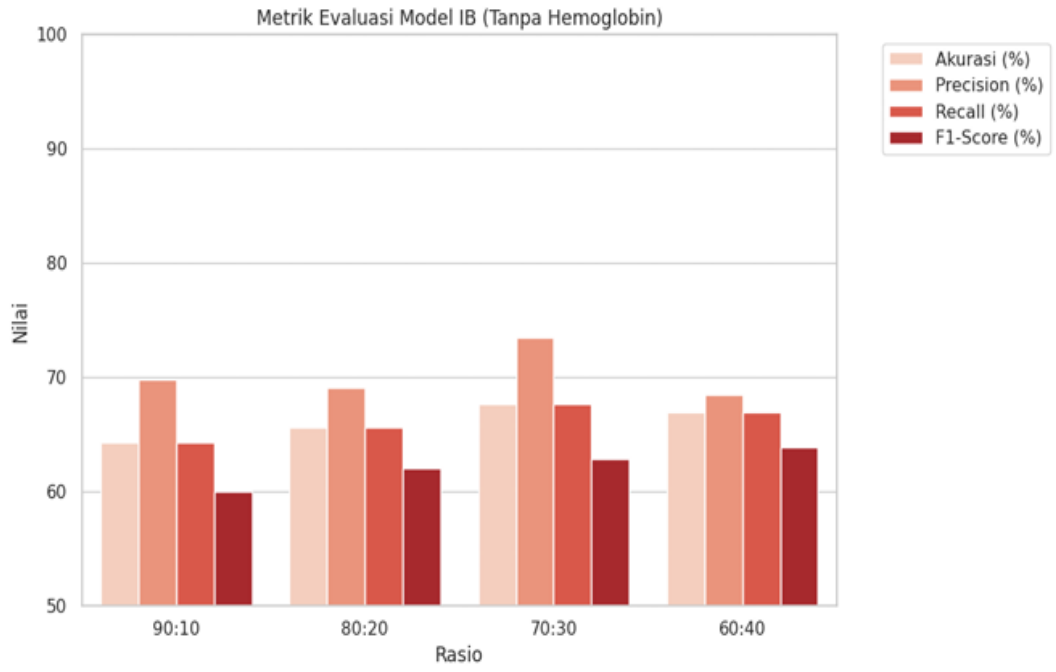
Selanjutnya, grafik tersebut menunjukkan harmoni yang sangat baik antara keempat metrik evaluasi, yaitu Akurasi, *Precision, Recall*, dan *F1-Score*. Nilai keempat metrik yang hampir sejajar pada setiap batang grafik mengindikasikan bahwa model memiliki kemampuan generalisasi yang stabil, tidak hanya unggul dalam akurasi secara keseluruhan tetapi juga handal dalam menyeimbangkan prediksi antara kelas positif dan negatif. Khusus pada rasio 80:20, nilai *F1-Score* yang tinggi menjadi indikator kuat bahwa model tetap mampu bekerja efektif pada distribusi data yang ada, menjadikannya skenario terbaik dalam penelitian ini untuk dijadikan acuan dalam sistem klasifikasi anemia.

Tabel 4.2 Metrik Evaluasi B

Rasio (Train:Test)	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
90:10	64,34	69.78	64.34	60.05
80:20	65,26	68,79	65,26	61,61
70:30	67,68	73,53	67,68	62,91
60:40	66,96	68,52	66,96	63,86

Berdasarkan hasil Pengujian pada Tabel 4.2 model B, terlihat penurunan performa yang sangat signifikan dibandingkan dengan Model A. Nilai akurasi 60 tertinggi pada skenario ini hanya mencapai 67,68% pada rasio data 70:30. Hal tersebut mengindikasikan bahwa penghapusan fitur Hemoglobin (Hb) menghilangkan informasi krusial yang digunakan Random Forest untuk melakukan klasifikasi.

Sementara itu, nilai *F1-Score* yang berada di kisaran 60%—63% menunjukkan bahwa model tidak hanya mengalami penurunan akurasi, tetapi juga mengalami kesulitan dalam menyeimbangkan antara *Precision* dan *Recall*. Tanpa fitur Hemoglobin, model cenderung lebih banyak melakukan kesalahan prediksi (*misclassification*), yang membuktikan bahwa kadar hemoglobin merupakan indikator utama dalam diagnosa anemia secara komputasional.



Gambar 4.30 Metrik Evaluasi Model B

Berdasarkan grafik hasil pengujian pada Gambar 4.30, terlihat bahwa skenario B (Model B) yang tidak menggunakan fitur Hemoglobin menunjukkan performa yang cenderung fluktuatif dan berada jauh di bawah performa model A. Akurasi tertinggi pada skenario tersebut hanya mampu mencapai rentang 60% hingga 67%, dengan hasil terbaik terlihat pada rasio data 70:30. Rendahnya capaian metrik evaluasi secara keseluruhan pada grafik tersebut membuktikan secara visual bahwa penghapusan variabel Hemoglobin menyebabkan model kehilangan informasi paling krusial, sehingga *Random Forest* mengalami kesulitan dalam membedakan antara subjek yang menderita anemia dan yang tidak.

Selain itu, jika dilakukan observasi mendalam pada komposisi batang grafik di setiap rasio, terlihat adanya ketidakseimbangan yang cukup mencolok antara nilai *Precision* dan *F1-Score*. Nilai *F1-Score* yang secara konsisten menjadi metrik terendah (berada di bawah 65%) menandakan bahwa model B memiliki kelemahan dalam menyeimbangkan ketepatan prediksi dengan kemampuan menangkap seluruh data positif. Fenomena ini mempertegas kesimpulan penelitian bahwa fitur hematologi pendukung seperti MCH, MCV, dan MCHC tidak dapat memberikan hasil klasifikasi yang optimal jika tidak dibarengi dengan fitur utama Hemoglobin, sehingga model ini dianggap kurang layak untuk digunakan dalam diagnosa klinis yang memerlukan akurasi tinggi.

Tabel 4.3 Metrik Evaluasi Gabungan (Rata-Rata)

Rasio Data	**Skenario**							
	A1	**B2**	**A2**	**B2**	**A3**	**B3**	**A4**	**B4**
K-Fold								
Fold 1 (%)	91,8	67,9	89	68,4	92,9	64,3	91.8	63,1
Fold 2 (%)	91,4	63,6	91,6	58,5	94,4	67,3	90	64,9
Fold 3 (%)	93,3	67,9	91,1	66,9	90,4	60,8	90	6,23
Fold 4 (%)	90,5	64,7	93,3	62,5	92,9	60,8	90	56,4
Fold 5 (%)	88,6	67,4	88,5	65,6	86,8	67,6	88,2	67,6
Mean (%)	**91**	**66**	**91**	**64**	**92**	**64**	**90**	**63**
Std	**0.0155**	**0.0181**	**0.0177**	**0.0350**	**0.0267**	**0.0300**	**0.0113**	**0.0370**
Acc (%)	93	64,3	94,3	65,6	89,9	67,6	93,3	66,9
Pss (%)	93,1	69,7	94,4	69,1	91,3	73,5	93,9	68,5
Recall (%)	93	64,3	94,3	65,6	89,9	67,6	93,3	66,9
F1 (%)	93	60	94,4	62	89,9	62,9	93,3	63,8

Berdasarkan seluruh rangkaian eksperimen yang telah dipetakan pada Tabel 4.3, perbandingan performa antara Model A (menggunakan seluruh fitur hematologi) dan Model B (menghapus fitur Hemoglobin/Hb) menunjukkan perbedaan performa yang sangat signifikan. Model A secara konsisten menghasilkan performa terbaik pada seluruh skenario pembagian data, dengan hasil tertinggi diperoleh pada rasio 80:20 yang mencapai akurasi testing sebesar 94,39% serta nilai *F1-Score* sebesar 94,40%. Hasil ini menunjukkan bahwa kombinasi fitur Hemoglobin, MCH, MCV, dan MCHC mampu membentuk representasi pola data yang lebih lengkap sehingga algoritma *Random Forest* dapat melakukan proses klasifikasi secara lebih akurat. Kondisi tersebut sejalan dengan penelitian *Machine Learning* yang menunjukkan bahwa performa model klasifikasi meningkat ketika fitur yang digunakan memiliki relevansi tinggi terhadap target prediksi (Sharma et al., 2023).

Sebaliknya, ketika fitur Hemoglobin dihilangkan pada Model B, terjadi penurunan performa yang cukup drastis pada seluruh skenario pengujian. Pada rasio data 90:10 misalnya, akurasi pengujian turun dari 93,01% pada Model A menjadi hanya 64,34% pada Model B. Penurunan serupa terlihat pada rasio 80:20 di mana nilai *F1-Score* turun dari 94,40% menjadi 62,09%. Hasil tersebut menunjukkan bahwa fitur lain seperti MCH, MCV, dan MCHC belum mampu memberikan informasi pembeda kelas yang cukup kuat ketika fitur Hemoglobin tidak tersedia. Dengan kata lain, keberadaan Hb berperan besar dalam membantu model membedakan kondisi anemia dan non-anemia secara lebih presisi.

Hal serupa juga dilaporkan pada penelitian klasifikasi anemia berbasis hematologi oleh Medical *Data Classification* yang menunjukkan bahwa parameter Hb menjadi variabel dengan kontribusi prediktif tertinggi (Sivakumar et al., 2024).

Validasi performa model tidak hanya dilakukan melalui pengujian *testing set*, tetapi juga diperkuat menggunakan metode 5-*Fold Cross Validation* untuk mengevaluasi kestabilan model pada subset data yang berbeda. Berdasarkan hasil validasi, Model A pada rasio data 90:10 mampu mempertahankan rata-rata akurasi (*mean accuracy*) sebesar 91% dengan nilai standar deviasi sebesar 0,0155. Nilai standar deviasi yang kecil menunjukkan bahwa model memiliki tingkat konsistensi yang baik dan menghasilkan performa yang relatif stabil pada setiap fold pengujian. Sebaliknya, pada Model B, rata-rata akurasi turun menjadi 66% disertai peningkatan standar deviasi menjadi 0,0181. Fluktuasi ini menunjukkan bahwa ketika fitur Hb dihilangkan, performa model menjadi lebih tidak stabil dan sensitif terhadap variasi data pelatihan. Menurut Sarker, validasi silang seperti *K-Fold* digunakan untuk mengukur generalisasi model dan mengurangi bias evaluasi akibat ketergantungan pada satu pembagian data tertentu (Sarker, 2022).

Skenario pengujian dengan membandingkan dua model berbeda dirancang berdasarkan pendekatan *feature ablation study*, yaitu metode eksperimen yang dilakukan dengan cara menghapus satu fitur tertentu untuk mengukur tingkat kontribusi fitur tersebut terhadap performa sistem secara keseluruhan. Penggunaan skenario ini penting karena dalam pengembangan machine learning, tidak cukup hanya mengetahui model memiliki akurasi tinggi, tetapi juga perlu dipahami fitur mana yang paling memengaruhi proses pengambilan keputusan model. Dengan melakukan penghapusan fitur secara

terkontrol, peneliti dapat menganalisis sensitivitas algoritma terhadap perubahan struktur input dan mengetahui tingkat ketergantungan model terhadap fitur tertentu. Pendekatan *ablation study* banyak digunakan dalam penelitian modern untuk mengevaluasi *feature importance* dan *robustnes* model klasifikasi (Chen et al., 2020).

Secara medis, Hemoglobin memang telah lama dikenal sebagai indikator utama dalam diagnosis anemia sehingga secara teori sudah dapat diasumsikan sebagai fitur yang sangat penting. Namun, dalam konteks penelitian komputasi dan *machine learning*, asumsi klinis tersebut tidak dapat langsung diterima tanpa dilakukan pembuktian empiris melalui eksperimen sistem. Oleh karena itu, fitur Hb tetap diuji secara khusus bukan untuk membuktikan kembali pengetahuan medis yang sudah ada, melainkan untuk menjawab pertanyaan komputasional: apakah algoritma masih mampu mempertahankan performa klasifikasi yang baik ketika fitur utama tersebut dihilangkan, dan apakah fitur-fitur lain dapat menggantikan informasi yang dibawa oleh Hemoglobin. Hasil eksperimen menunjukkan bahwa ketika Hemoglobin dihapus, performa model turun secara signifikan sehingga membuktikan bahwa fitur lain belum mampu menggantikan peran Hemoglobin sebagai sumber informasi utama dalam proses klasifikasi. Dengan demikian, pengujian ini memberikan bukti kuantitatif bahwa dalam sistem klasifikasi berbasis Random Forest, Hemoglobin memiliki tingkat *feature importance* dominan dan tidak tergantikan. Pendekatan semacam ini direkomendasikan dalam penelitian *explainable artificial intelligence* untuk memahami kontribusi individual setiap fitur terhadap keputusan model (Zhang et al., 2022).

4.5 Integrasi Nilai Islam

Pengembangan model klasifikasi anemia menggunakan *Random Forest* merupakan bentuk implementasi dari perintah Allah SWT dalam QS. Al-Alaq (96): 1-5:

اِقْرَأْ بِاسْمِ رَبِّكَ الَّذِيْ خَلَقَۚ (١) خَلَقَ الْاِنْسَانَ مِنْ عَلَقٍۚ (٢) اِقْرَأْ وَرَبُّكَ الْاَكْرَمُۙ (٣) الَّذِيْ عَلَّمَ بِالْقَلَمِۙ

عَلَّمَ الْاِنْسَانَ مَا لَمْ يَعْلَمْۗ (٤)(٥)

"Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan. Dia telah menciptakan manusia dari segumpal darah. Bacalah, dan Tuhanmulah Yang Mahamulia, Yang mengajar (manusia) dengan perantaraan kalam. Dia mengajarkan manusia apa yang tidak diketahuinya."

Proses analisis pola data medis dalam penelitian tersebut yakni upaya manusia untuk memahami "apa yang tidak diketahuinya" dengan melalui riset ilmiah demi kemaslahatan umat. Semangat pencarian ilmu tersebut bertujuan untuk menjaga amanah berupa kesehatan tubuh, sebagaimana ditegaskan dalam sebuah hadist:

"Sesungguhnya badanmu memiliki hak atasmu" (HR. Bukhari dan Muslim).

Dengan memanfaatkan *Machine Learning*, peneliti berupaya memenuhi hak tubuh tersebut melalui sistem deteksi dini yang objektif dan terukur. Penulis juga membuktikan kebesaran Allah SWT dalam menciptakan keseimbangan komponen tubuh manusia, di mana fitur Hemoglobin ditemukan sebagai variabel paling vital. Hal ini selaras dengan QS. Al-Infithar (82): 7:

الَّذِيْ خَلَقَكَ فَسَوّٰىكَ فَعَدَلَكَۙ

"(Tuhan) yang telah menciptakanmu lalu menyempurnakan kejadianmu dan menjadikan susunan tubuhmu seimbang."

Pentingnya akurasi tinggi pada Model A mencerminkan prinsip Tabayyun (verifikasi kebenaran) sebagaimana diperintahkan dalam QS. Al-Hujurat (49): 6:

ۖتَبَيَّنُوٓا۟ أَن تُصِيبُوا۟ قَوْمًۢا بِجَهَٰلَةٍ

"*...maka telitilah kebenarannya agar kamu tidak mencelakakan suatu kaum karena kebodohan (kecerobohan)...*"

Sebaliknya, penggunaan parameter yang tidak lengkap pada Model B merupakan bentuk kelalaian yang berisiko menjatuhkan manusia ke dalam bahaya, sesuai dengan peringatan dalam QS. Al-Baqarah (2): 195:

وَلَا تُلْقُوْا بِاَيْدِيْكُمْ اِلَى التَّهْلُكَةِ

"..dan janganlah kamu menjatuhkan dirimu sendiri ke dalam kebinasaan..."

Oleh karena itu, ketelitian dalam memilih fitur medis adalah syarat mutlak untuk menghasilkan diagnosa yang aman dan dapat dipertanggungjawabkan dalam menjaga jiwa (Hifz an-Nafs). Integrasi antara teknologi dan nilai syariat ini pada akhirnya bertujuan untuk mencapai kemaslahatan melalui diagnosa yang presisi. Hal ini merupakan langkah awal menuju penyembuhan yang tepat, sebagaimana sabda Rasulullah. SAW dalam hadist yang berbunyi:

“Setiap penyakit ada obatnya. Apabila obat itu tepat sasaran mengenai penyakitnya, maka sembuhlah ia dengan izin Allah SWT” (HR. Muslim).

Melalui akurasi model yang dihasilkan dalam penelitian tersebut, penulis mengharapkan proses penanganan medis menjadi lebih efektif dan efisien. Dengan demikian, penerapan ilmu teknologi dalam nilai islami bukan sekadar aktivitas teknis, melainkan bentuk pengabdian seorang hamba dalam menjaga kehidupan dan memberikan manfaat sebesar-besarnya bagi sesama manusia (Anfa'uhum linnas).

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dan pembahasan mengenai klasifikasi penyakit anemia menggunakan *Random Forest*, maka penulis menarik beberapa kesimpulan, antara lain:

1. Performa Model Terbaik: Model A (semua fitur) menunjukkan performa optimal pada skenario rasio data 80:20 atau lebih tepatnya pada Model A2 dengan tingkat akurasi yang mencapai 94,39% (pada random state 42). Stabilitas model ini teruji pada berbagai random state dan pembagian data, yang membuktikan bahwa integrasi fitur klinis yang lengkap memberikan hasil diagnosa yang paling presisi.

2. Pengaruh Fitur Hemoglobin: Penelitian ini membuktikan secara empiris bahwa fitur Hemoglobin merupakan prediktor paling krusial dalam pendeteksian anemia. Hal ini ditunjukkan oleh penurunan akurasi yang sangat signifikan pada model B (Tanpa Hemoglobin) yang hanya mencapai rentang 61,75% – 69,23%, dibandingkan Model A (Semua Fitur) yang konsisten di atas 90%.

5.2 Saran

Berdasarkan hasil penelitian yang telah penulis lakukan, terdapat beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya:

1. Penambahan Variasi Fitur: Disarankan untuk menambahkan fitur klinis lain yang relevan secara medis, seperti kadar zat besi (ferritin) atau riwayat penyakit penyerta, untuk meningkatkan kemampuan model dalam mengenali jenis-jenis anemia yang lebih spesifik.

2. Penanganan Ketidakseimbangan Data: Untuk penelitian selanjutnya, disarankan menggunakan teknik oversampling (seperti SMOTE) atau undersampling guna meminimalisir angka False Negative pada skenario fitur yang terbatas, sehingga model lebih sensitif terhadap kelas minoritas (positif anemia).

3. Eksperimen Algoritma Lain: Perlu dilakukan komparasi dengan algoritma ensemble learning lain seperti XGBoost atau CatBoost untuk melihat apakah terdapat peningkatan efisiensi komputasi atau akurasi yang lebih tinggi pada dataset yang sama.

4. Implementasi Sistem: Mengembangkan hasil model A yang telah optimal ke dalam bentuk aplikasi berbasis mobile atau web agar dapat digunakan secara praktis oleh tenaga medis atau masyarakat umum sebagai alat deteksi dini anemia.

DAFTAR PUSTAKA

Ahmed, H., Lofstead, J. (2022). Managing Randomness to Enable Reproducible Machine Learning. https://doi.org/10.1145/3526062.3536353

Airlangga, G. (2024). Leveraging Machine Learning for Accurate Anemia Diagnosis Using Complete Blood Count Data. IndonesAn Journal of ArtificAl Intelligence and Data Mining, 7(3), 489–500. http://dx.doi.org/10.24014/ijaidm.v7i2.29869

Ahamad, G. N., Shafiullah, Fatima, H., Imdadullah, Zakariya, S. M., Abbas, M., Alqahtani, M. S., & Usman, M. (2023). Influence of optimal hyperparameters on the performance of machine learning algorithms for predicting heart disease. 11(3), 734. https://doi.org/10.3390/pr11030734

Al-Bukhari & Muslim. (2017). Ringkasan Shahih Bukhari dan Muslim. Jakarta: Pustaka Amani.

Alghifari, F., dkk. (2021). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK).

Al Hazmi, M. I. U. (2024). Klasifikasi risiko diabetes menggunakan Random Forest dengan kombinasi SMOTE-Tomek Links dan Recursive Feature Elimination [Skripsi, Universitas Islam Negeri Maulana Malik Brahim Malang]. UIN Malang Institutional Repository. http://etheses.uin-malang.ac.id/78390/

Al-Qur'an. (2019). Al-Qur'an dan Terjemahannya. Jakarta: Kementerian Agama RI.

Azis, H. ; Rismayanti, N. (2024). Prediksi anemia dari pixel gambar dan level hemoglobin menggunakan random forest classifier. Networking Engineering Research Operation, 9(1), 21–34. https://doi.org/10.21107/nero.v9i1.27916

Awaad, Amira S., Yomna M.Elbarawy, H. Mancy, and Naglaa E. Ghannam. 2025. "Exploring CBC Data for Anemia Diagnosis: A Machine Learning and Ontology Perspective" BioMedInformatics 5, no. 3: 35. https://doi.org/10.3390/biomedinformatics5030035

Belo, J. F., Salim, L. A., IndrAni, D., & Handayani, S. (2022). Tablet zat besi untuk menurunkan Anemia pada wanita tidak hamil: Systematic review & meta-analysis. Jurnal Penelitian Kesehatan Suara Forikes, 13(4), 939–947. https://doi.org/10.33846/sf13410

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

Cabanillas-Carbonell M, Zapata-Paulini J. Evaluation of machine learning models for the prediction of Alzheimer's: In search of the best performance. Brain Behav Immun Health. 2025 Jan 31;44:100957. https://doi.org/10.1016/j.bbih.2025.100957

Chaparro, C. M., & Suchdev, P. S. (2019). Anemia epidemiology, pathophysiology, and etiology in low- and middle-income countries. Annals of the New York Academy of Sciences, 1450(1), 15–31. https://doi.org/10.1111/nyas.14092

Chen, X., et al. (2020). Machine learning-based classification of COVID-19 severity. Journal of Infection, 81(5), 798–816. https://doi.org/10.1016/j.jinf.2020.08.011

Cynthia, E. P., Afif Rizky A., Nazir, A., & SyafrA, F. (2021). Random Forest Algorithm to Investigate the Case of Acute Coronary Syndrome. Jurnal RESTI, 5(2), 369-378. https://doi.org/10.29207/resti.v5i2.3000

Dimgba, M., & Andreas, R. (2024). Optimal hyperparameter search strategies: Benchmarking grid search, random search, and genetic algorithms across regression, classification, and clustering tasks. ResearchGate. https://doi.org/10.13140/RG.2.2.24957.37603

Faradila, Z., Homaidi, A., & Prasetyo, J. D. (2025). Classification of anaemA status using the K-Nearest Neighbor algorithm. G-Tech: Jurnal Teknologi Terapan, 9(1), 436–444. https://doi.org/10.70609/gtech.v9i1.6377

Farida, L. N., Novitasari, D. C. R., & Utami, W. D. (2025). Identifikasi Tipe Penyakit Anemia dengan Menggunakan Metode Random Forest. Journal of Informatics Engineering, 6(1). https://doi.org/10.53682/jointer.v6i01.386

Fu, S., & Avdelidis, N. P. (2024). Novel prognostic methodology of bootstrap forest and hyperbolic tangent boosted neural network for aircraft system. Applied Sciences, 14(12), 5057. https://doi.org/10.3390/app14125057

Fouodo, C.J., Kronziel, L.L., König, I.R., & Szymczak, S. (2023). Effect of hyperparameters on varAble selection in random forests. ArXiv, abs/2309.06943. https://www.semanticscholar.org/reader/78716d527c45cacc2cd73fd1104f675503ce3560

Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing data dan klasifikasi untuk prediksi kinerja akademik siswa. Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), 11(1), 215–224. https://doi.org/10.25126/jtiik.20241118074

Hidayat, M. A., Hidayatullah, N., & Mursyid, A. (2024). Classification Comparison Performance of Supervised Machine Learning Random Forest and Decision Tree Algorithms Using Confusion Matrix. Jurnal SISFOKOM, 13(1), 23-32. https://doi.org/10.32736/sisfokom.v13i1.1551

Hossain, M., Islam, Z., Sultana, S., Rahman, A. S., Hotz, C., Haque, Md. A., Dhillon, C. N., Khondker, R., Neufeld, L. M., & Ahmed, T. (2019). Effectiveness of Workplace Nutrition Programs on Anemia Status among Female Readymade Garment Workers in Bangladesh: A Program Evaluation. Nutrients, 11(6), 1259 . https://doi.org/10.3390/nu11061259

Jiang, X., et al. (2022). Deep Learning and Machine Learning with Grid Search to Optimize Performance: A Review and Practical Guide. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9570670/

Jin, X., et al. (2021). Generalization performance and experiment design: Linked Classification bounds. Pattern Recognition, 120, 108159. https://doi.org/10.1016/j.patcog.2021.108159

Kaope, C., & Pristyanto, Y. (2023). The effect of class imbalance handling on datasets toward classification algorithm performance. Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer, 22(2), 227–238. https://doi.org/10.30812/matrik.v22i2.2515

Kim, H., Lee, J., & Park, S. (2025). Machine learning approach for differentAting iron deficiency Anemia and thalassemA. Scientific Reports, 15,12345. https://pubmed.ncbi.nlm.nih.gov/40374805/

Kitaw, B., et al. (2024). Leveraging machine learning models for classification of Anemia severity/type. BMC Public Health. https://doi.org/10.1186/s12889-024- 20827-1

Kusumawati, A., Pratiwi, A. D., Sahla, H. I., Puspita Sari, N. A., & SetAwan, W. (2023). Penerapan Data Mining untuk Prediksi Penyakit Anemia menggunakan Metode Machine Learning. Diunggah di Scribd oleh Diah Pratiwi. DAkses dari https://id.scrBd.com/document/772435154/Penerapan-Data-Mining-Untuk-Prediksi-Penyakit-Anemia-Menggunakan-Metode-Mechine-Learning

Marpaung, S. H., Sinaga, F. M., Rambe, K. H., Simamora, F. P., & Kelvin. (2025). Random forest optimization using recursive feature elimination for stunting classification. IndonesAn Journal of ArtificAl Intelligence and Data Mining, 8(1), 281–287. https://doi.org/10.24014/ijaidm.v8i1.35295

Mashuri, M. (2025). Transformative fiqh and health technology: An Islamic ethical study of the use of artificAl intelligence in medical practice. KodifikasA: Jurnal Penelitian Islam, 19(1). https://doi.org/10.21154/kodifikasA.v19i1.12159

Neufang, S., Li, F., Akhrif, A. et al. Toward a fair, gender-debAsed classifier for the dAgnosis of attention deficit/hyperactivity disorder- a Machine-Learning based classification study. BMC Med Inform Decis Mak 25, 290 (2025). https://doi.org/10.1186/s12911-025-03126-0

Nurcahyo, J. A., & Sasongko, T. B. (2023). Hyperparameter Tuning Algoritma Supervised Learning untuk Klasifikasi Keluarga Penerima Bantuan Pangan Beras. IndonesAn Journal of Computer Science, 12(3), 1365. DOI:10.33022/ijcs.v12i3.3254

Nugroho, M. W. (2025). Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi. Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi), 9(4), 1562–1571. DOI:10.35870/jtik.v9i4.4236

Nurhopipah, A., & Hasanah, U. (2020). Dataset Splitting Techniques Comparison for Face Classification on CCTV Images. IJCCS (IndonesAn Journal of Computing and Cybernetics Systems), 14(4). https://doi.org/10.22146/ijccs.58092

Pagano, T. P., Loureiro, R. B., Lisboa, F. V. N., Peixoto, R. M., Guimarães, G. A. S., Cruz, G. O. R., Araujo, M. M., Santos, L. L., Cruz, M. A. S., Oliveira, E. L. S., Winkler, I., & Nascimento, E. G. S. (2023). BAs and Unfairness in Machine Learning Models: A Systematic Review on Datasets, Tools, Fairness Metrics, and Identification and Mitigation Methods. Big Data and Cognitive Computing, 7(1), 15. https://doi.org/10.3390/bdcc7010015

Pasricha, S. R., Tye-Din, J., Muckenthaler, M. U., & Swinkels, D. W. (2021). Iron deficiency. The Lancet, 397(10270), 233–248 https://doi.org/10.1016/S0140-6736(20)32594-0

Pullakhandam, S., & McRoy, S. (2024). Classification and Explanation of Iron Deficiency Anemia from Complete Blood Count Data Using Machine Learning. BioMedInformatics, 4(1), 661-672. https://doi.org/10.3390/biomedinformatics4010036

Purba, M., Ermatita, E., & AbdAnsah, A. (2022). Effect of Random Splitting and Cross Validation for IndonesAn Opinion Mining using Machine Learning Approach. International Journal of Advanced Computer Science and Applications, 13(9). https://doi.org/10.14569/IJACSA.2022.0130917

Ramzan M, Sheng J, Saeed MU, Wang B, Duraihem FZ. Revolutionizing Anemia detection: integrative machine learning models and advanced attention mechanisms. Vis Comput Ind Biomed Art. 2024 Jul 17;7(1):18. doi: https://pmc.ncbi.nlm.nih.gov/articles/PMC11255163/10.1186/s42492-024-00169-4. PMID: 39017765; PMCID: PMC11255163.

Samtani, P., Raghuwanshi, V., Raval, D., & Sahu, R. (2025). Anemia Detection Using CBC Data: A Comparative Study. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 13(5), 3184–3191. https://www.ijraset.com/research-paper/Anemia-detection-using-cbc-data-a-comparative-study

Sari, N. A. A., Jannah, U. M., & Nurmalitasari. (2024). Komparasi algoritma C4.5 dan K-Nearest Neighbor (K-NN) pada klasifikasi penyakit Anemia. Jurnal Sistem Informasi dan Teknik Komputer, 9(2), 110–113. https://ejournal.catursakti.ac.id/index.php/simtek/article/view/399

Sarker I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN computer science*, *2*(3), 160. https://doi.org/10.1007/s42979-021-00592-x

Shaveta. (2023). A review on machine learning. International Journal of Science and Research Archive, 9(1), 281–285. https://doi.org/10.30574/ijsra.2023.9.1.0410

Sivakumar, M., Parthasarathy, S., & Padmapriya, T. (2024). "Trade-off between training and testing ratio in machine learning for medical image processing." PeerJ Computer Science, 2024, e2245. DOI:10.7717/peerj-cs.2245

Susanto, E. R., & Pranajaya, A. (2025). Optimasi Random Forest untuk Prediksi Penyakit Jantung Menggunakan SMOTEENN dan Grid Search. Jurnal Pendidikan Dan Teknologi IndonesA, 5(7), 1965-1979. https://doi.org/10.52436/1.jpti.855

Tepakhan, W., et al. (2025). Machine learning approach for differentAting iron deficiency Anemia and thalassemA. Scientific Reports. https://doi.org/10.1038/s41598-025-12345-x

Thomas, N.S., Kaliraj, S. An Improved and Optimized Random Forest Based Approach to Predict the Software Faults. SN COMPUT. SCI. 5, 530 (2024). https://doi.org/10.1007/s42979-024-02764-x

Ulya, M., & NurlAna, L. (2025). Qur'anic perspectives on digital health ethics and artificAl intelligence. Jurnal Studi Islam dan Teknologi Kesehatan, 6(1).

Wijaya, A. P. (2025). Perbandingan algoritma klasifikasi Random Forest dengan Naïve Bayes Classifier pada studi penyakit berdasarkan pola nutrisi. REMik: Riset dan E-Jurnal Manajemen Informatika Komputer, 9(1), 429–440. https://doi.org/10.33395/remik.v9i1.14652

Yimer, A. (2025). Optimizing machine learning models for predicting Anemia: A systematic review. PLOS Digital Health, 4(2), e0000123. https://doi.org/10.1371/journal.pdig.0000123

Yunos, A. S., & Hamdan, M. N. (2024). Kecerdasan buatan dalam penjagaan kesihatan: Analisis dari sudut pandang maqasid al-sharAh. Journal Information and Technology Management (JISTM), 9(35). https://gaexcellence.com/jistm/article/view/461

Zemariam, A. B., et al. (2024). Employing supervised machine learning algorithms for classification and prediction of Anemia among youth girls in EthiopA. Scientific Reports. https://doi.org/10.1371/journal.pdig.0000123

Zhu, N., Zhu, C., Zhou, L., Zhu, Y., & Zhang, X. (2022). Optimization of the Random Forest Hyperparameters for Power IndustrAl Control Systems Intrusion Detection Using an Improved Grid Search Algorithm. Applied Sciences, 12(20), 10456. https://doi.org/10.3390/app122010456