

**ANALISIS DAN PENERAPAN MODEL MACHINE
LEARNING UNTUK KLASIFIKASI DATASET PENYAKIT
ANEMIA**

SKRIPSI



disusun oleh

ADI PRASETYO

21.12.2248

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

**ANALISIS DAN PENERAPAN MODEL MACHINE
LEARNING UNTUK KLASIFIKASI DATASET PENYAKIT
ANEMIA**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Sistem Informasi



disusun oleh

ADI PRASETYO

21.12.2248

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS DAN PENERAPAN MODEL MACHINE LEARNING UNTUK
KLASIFIKASI DATASET PENYAKIT ANEMIA**

yang disusun dan diajukan oleh

Adi Prasetyo

21.12.2248

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 2 Mei 2025

Dosen Pembimbing,



Yoga Pristyanto, S.Kom, M.Eng

NIK. 190302412

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS DAN PENERAPAN MODEL MACHINE LEARNING UNTUK
KLASIFIKASI DATASET PENYAKIT ANEMIA**

yang disusun dan diajukan oleh

Adi Prasetyo

21.12.2248

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 Mei 2025

Susunan Dewan Penguji

Nama Penguji

Irma Rofni Wulandari, S.Pd., M.Eng
NIK. 190302329

Agung Nugroho, S.Kom., M.Kom
NIK. 190302242

Yoga Pristyanto, S.Kom., M.Eng.
NIK. 190302412

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 Mei 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Adi Prasetyo
NIM : 21.12.2248

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS DAN PENERAPAN MODEL MACHINE LEARNING UNTUK KLASIFIKASI DATASET PENYAKIT ANEMIA

Dosen Pembimbing : Yoga Pristyanto, S.Kom, M.Eng

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

Yogyakarta, 23 Mei 2025

Yang Menyatakan,



Adi Prasetyo

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur ke hadirat Allah Subhanahu wa Ta'ala, saya persembahkan skripsi ini kepada Bapak Sugiyanto dan Ibu Lilis Amanah, orang tua tercinta yang telah menjadi sumber kekuatan, kasih sayang, dan doa yang tiada henti dalam setiap langkah perjalanan hidup dan pendidikan saya; kepada Almarhumah Ibu Sudarsih, yang meskipun telah tiada, tetap menjadi inspirasi dan semangat dalam hati saya; serta kepada kakak-kakak saya, Arief Kurniawan, Dwi Setyaningrum, dan Ratri Wulandari, yang selalu memberikan dukungan moral, semangat, dan motivasi yang tak ternilai harganya. Semoga karya ini menjadi wujud penghargaan atas segala pengorbanan, kasih sayang, dan dukungan yang telah diberikan.

Dengan penuh rasa syukur kepada Allah Subhanahu wa Ta'ala, saya persembahkan skripsi ini kepada Universitas AMIKOM Yogyakarta dan Bapak Yoga Pristyanto, S.Kom., M.Eng., dosen pembimbing yang telah dengan sabar dan penuh dedikasi membimbing saya sejak semester 6 hingga saat ini. Bimbingan, arahan, dan motivasi yang Bapak berikan sangat membantu dalam proses akademik saya. Semoga karya sederhana ini dapat menjadi bentuk penghargaan atas segala pengorbanan dan kasih sayang yang telah diberikan.

Dengan tulus saya persembahkan karya ini kepada rekan-rekan SI 06 dan teman seperjuangan. Kebersamaan dan dukungan kalian telah memperkaya perjalanan akademik saya dan menjadikan proses belajar lebih bermakna. Semoga persahabatan ini terus terjaga.

KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Allah Subhanahu wa Ta'ala atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga saya dapat menyelesaikan skripsi ini dengan judul “Analisis dan Penerapan Model *Machine Learning* untuk Klafikasi Dataset Penyakit Anemia” sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Universitas AMIKOM Yogyakarta.

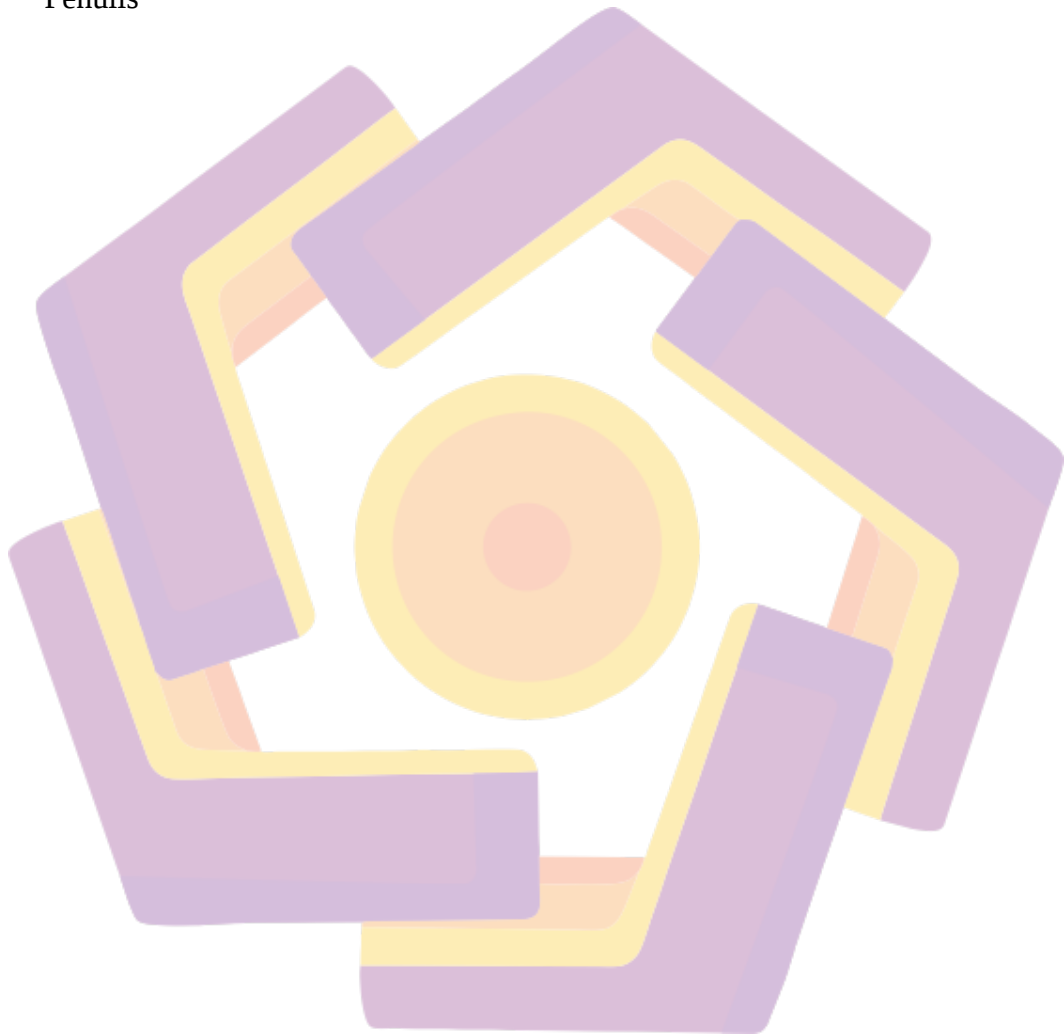
Dalam proses penyusunan skripsi ini, saya menyadari bahwa keberhasilan ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, saya ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. M. Suyanto, MM, selaku Rektor Universitas AMIKOM Yogyakarta, atas dukungan dan fasilitas yang diberikan selama masa studi saya.
2. Prof. Dr. Kusrini, M.Kom., selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta, atas bimbingan, arahan, dan dukungan yang telah diberikan selama proses perkuliahan hingga penyusunan skripsi ini.
3. Bapak Anggit Dwi Hartanto, S.Kom., M.Kom, selaku Ketua Program Studi Sistem Informasi, atas arahan dan dukungan yang diberikan selama masa perkuliahan.
4. Bapak Yoga Pristyanto, S.Kom., M.Eng, selaku dosen pembimbing, yang telah dengan sabar dan penuh dedikasi membimbing saya sejak semester 6 hingga saat ini.
5. Kedua orang tua saya, Bapak Sugiyanto dan Ibu Lilis Amanah, serta Almarhumah Ibu Sudarsih, yang telah menjadi sumber kekuatan dan motivasi melalui doa, dukungan, serta semangat yang tiada henti dalam setiap langkah perjalanan hidup dan pendidikan saya.

6. Seluruh teman-teman sekelas SI06 angkatan 2021, atas kebersamaan, dukungan, dan semangat yang telah kita bagi bersama selama masa perkuliahan.

Yogyakarta, <tanggal bulan tahun>

Penulis



DAFTAR ISI

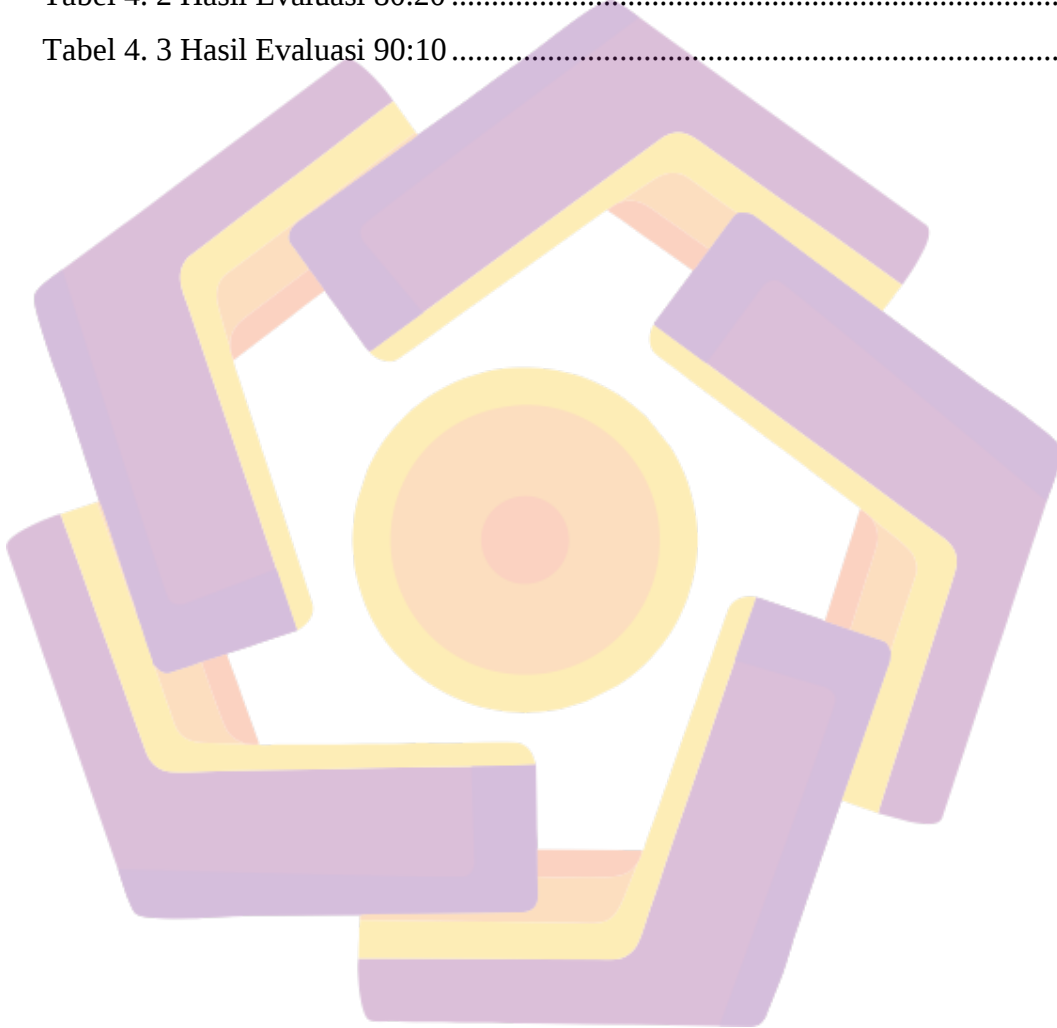
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iii
HALAMAN PERSEMBAHAN	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvii
ABSTRACT.....	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	4
1.6 Sistematika Penulisan.....	5

BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	19
2.2.1 Machine Learning	19
2.2.1 Resampling	19
2.2.2 SMOTE	19
2.2.3 Adasyn	20
2.2.4 <i>Decision Tree</i>	21
2.2.5 <i>Naïve bayes</i>	22
2.2.6 KNN (K-Nearest Neighbors)	22
2.2.7 Random Forest	23
2.2.8 Evaluasi.....	23
2.2.9 Deploy.....	24
BAB III METODE PENELITIAN	25
3.1 Objek Penelitian	25
3.2 Alur Penelitian.....	25
3.2.1 Akuisisi Data.....	25
3.2.2 Pra-Pemrosesan data	26
3.2.3.Klasifikasi	28
3.2.4 Evaluasi.....	30
3.2.5 Deploy.....	31
3.3 Alat dan Bahan.....	32
3.3.1 Alat/instrument	32
3.3.2 Data Penelitian	32
BAB IV HASIL DAN PEMBAHASAN	34

4.1 Deskripsi Implementasi	34
4.2 Implementasi Pengumpulan data	34
4.3 Implementasi Pra-Pemrosesan Data.....	36
4.4 Klasifikasi	43
4.5 Evaluasi.....	47
4.6 Hasil Analisis	51
4.7 Deployment.....	54
4.7.1 Direktori main.py	54
4.7.2 Direktori Home	55
4.7.3 Direktori Prediksi.....	56
4.7.5 Tampilan Deployment	61
BAB V PENUTUP	63
5.1 Kesimpulan.....	63
5.2 Saran.....	64
REFERENSI	65
LAMPIRAN.....	69

DAFTAR TABEL

Tabel 2. 1 Keaslian Penelitian.....	11
Tabel 3. 1 Sampel Dataset.....	32
Tabel 4. 1 Hasil Evaluasi 70:30	52
Tabel 4. 2 Hasil Evaluasi 80:20	52
Tabel 4. 3 Hasil Evaluasi 90:10	53



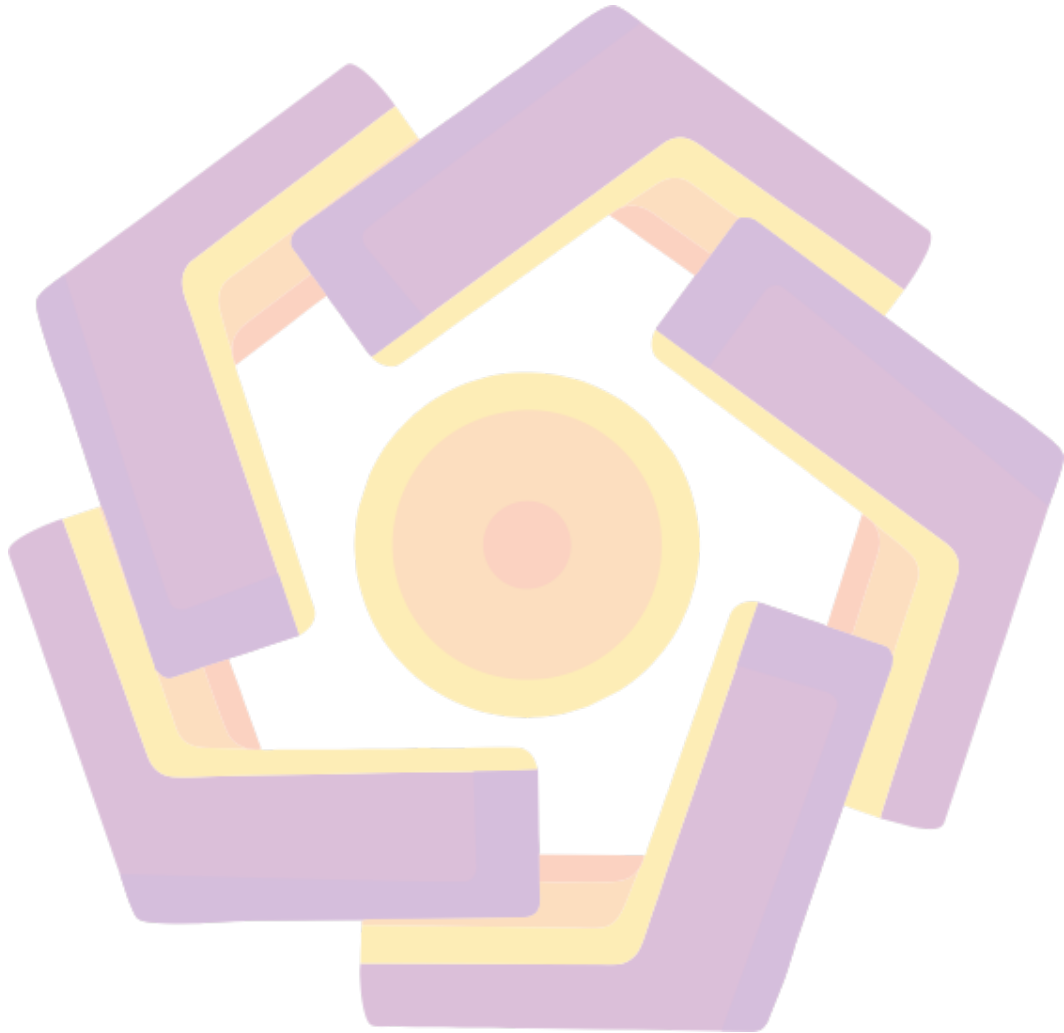
DAFTAR GAMBAR

Gambar 3. 1 Alur penelitian.....	25
Gambar 4. 1 Website Kaggle	35
Gambar 4. 2 Source code mengakses dataset dari google drive	35
Gambar 4. 3 Sample Dataset.....	36
Gambar 4. 4 Source code pengecekan missing values.....	37
Gambar 4. 5 Tampilan jumlah data yang hilang.....	37
Gambar 4. 6 Source code pengecekan dan penghapusan data duplikat.....	38
Gambar 4. 7 Source Code Pembagian data 70%:30%	38
Gambar 4. 8 Source Code Pembagian data 80%:20%	39
Gambar 4. 9 Source Code Pembagian Data 90% : 10%	39
Gambar 4. 10 Source code oversampling Adasyn	40
Gambar 4. 11 Code membandingkan sebelum dan setelah resampling Adasyn ...	40
Gambar 4. 12 Grafik perbandingan resampling Adasyn	41
Gambar 4. 13 Source code resampling SMOTE.....	41
Gambar 4. 14 Code Membandingkan sebelum dan setelah resample SMOTE	42
Gambar 4. 15 Grafik sebelum dan setelah resampling SMOTE.....	43
Gambar 4. 16 Source Code Inisialisai Model Decision Tree.....	44
Gambar 4. 17 Source Code Inisialisasi Model KNN	44
Gambar 4. 18 Source code Inisialisasi Model Naive Bayes	45
Gambar 4. 19 Source Code Inisialisasi Model Random Forest	45
Gambar 4. 20 Source code menguji model Decision Tree dengan data uji	46
Gambar 4. 21 Source code menguji model KNN dengan data uji.....	46
Gambar 4. 22 Source code menguji model Naive Bayes dengan data uji	47
Gambar 4. 23 Source code menguji model Random Forest dengan data uji	47
Gambar 4. 24 Source code mengevaluasi performa model Decision Tree	49
Gambar 4. 25 Output evaluasi performa model decision tree.....	49
Gambar 4. 26 Code Visualisasi Confusion Matrix menggunakan heatmap	50
Gambar 4. 27 Visualisasi Confusion matrix	51
Gambar 4. 28 Source Code Direktori main.py.....	55

Gambar 4. 29 Source Code Direktori Home.....	56
Gambar 4. 30 Source Code Input data Direktori Prediksi	57
Gambar 4. 31 Source Code validasi Input Direktori Prediksi	57
Gambar 4. 32 Source code melatih model Direktori Prediksi	57
Gambar 4. 33 Source code melakukan prediksi Direktori Prediksi.....	58
Gambar 4. 34 Source code akurasi model Direktori Prediksi.....	58
Gambar 4. 35 Source code Import Library Direktori Web_Fuction.....	58
Gambar 4. 36 Source code memuat dataset Direktori Web_Fuction.....	59
Gambar 4. 37 Source code split data Direktori Web_Fuction	59
Gambar 4. 38 Source code resampling SMOTE Direktori Web_Fuction	59
Gambar 4. 39 Source code melatih Model Decision Tree Web_Fuction	60
Gambar 4. 40 Source code membuat Prediksi Web_Fuction	61
Gambar 4. 41 Tambilan Home.....	61
Gambar 4. 42 Tampilan Halaman Prediksi.....	62
Gambar 4. 43 Tampilan Hasil Prediksi.....	62

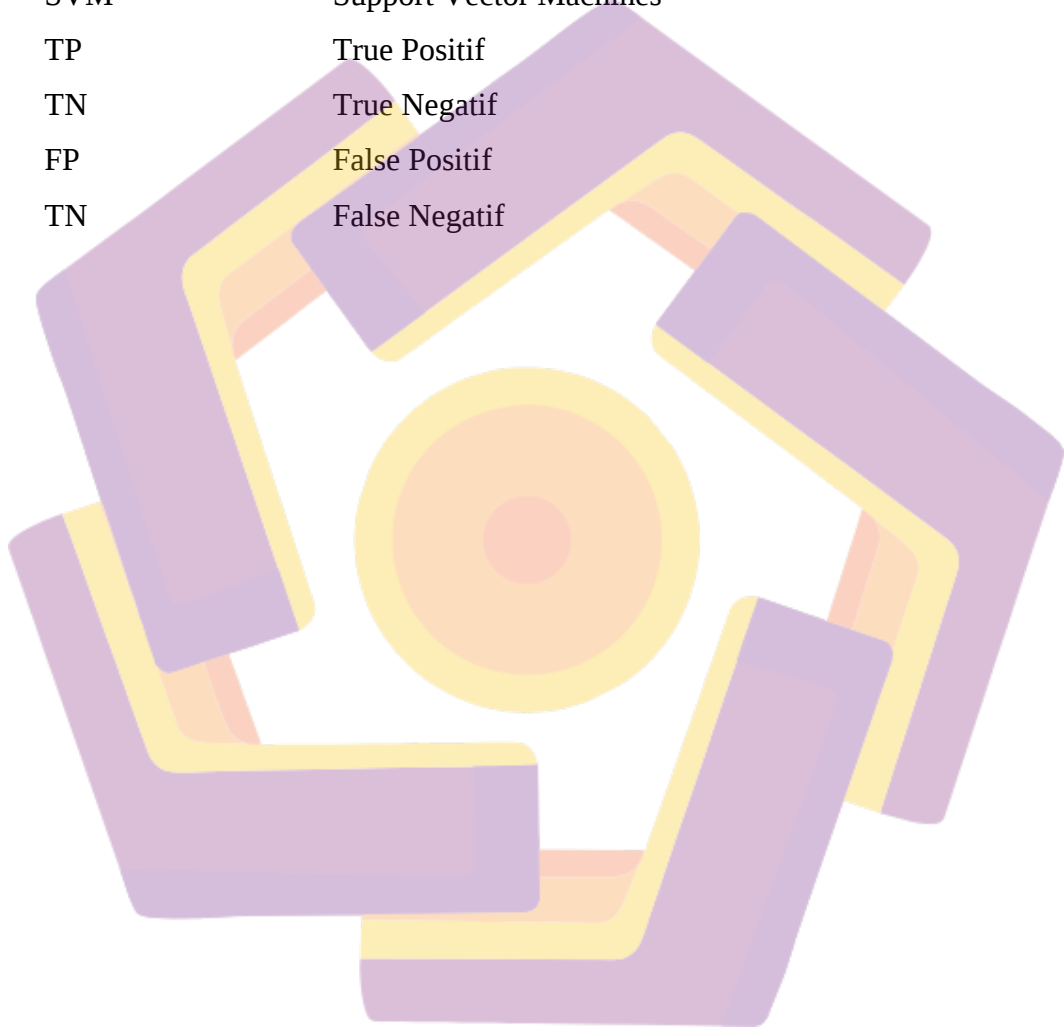
DAFTAR LAMPIRAN

Lampiran 1. Dataset	69
Lampiran 2. Source Code Analisis di Colab	70
Lampiran 3. Repository Github	70



DAFTAR LAMBANG DAN SINGKATAN

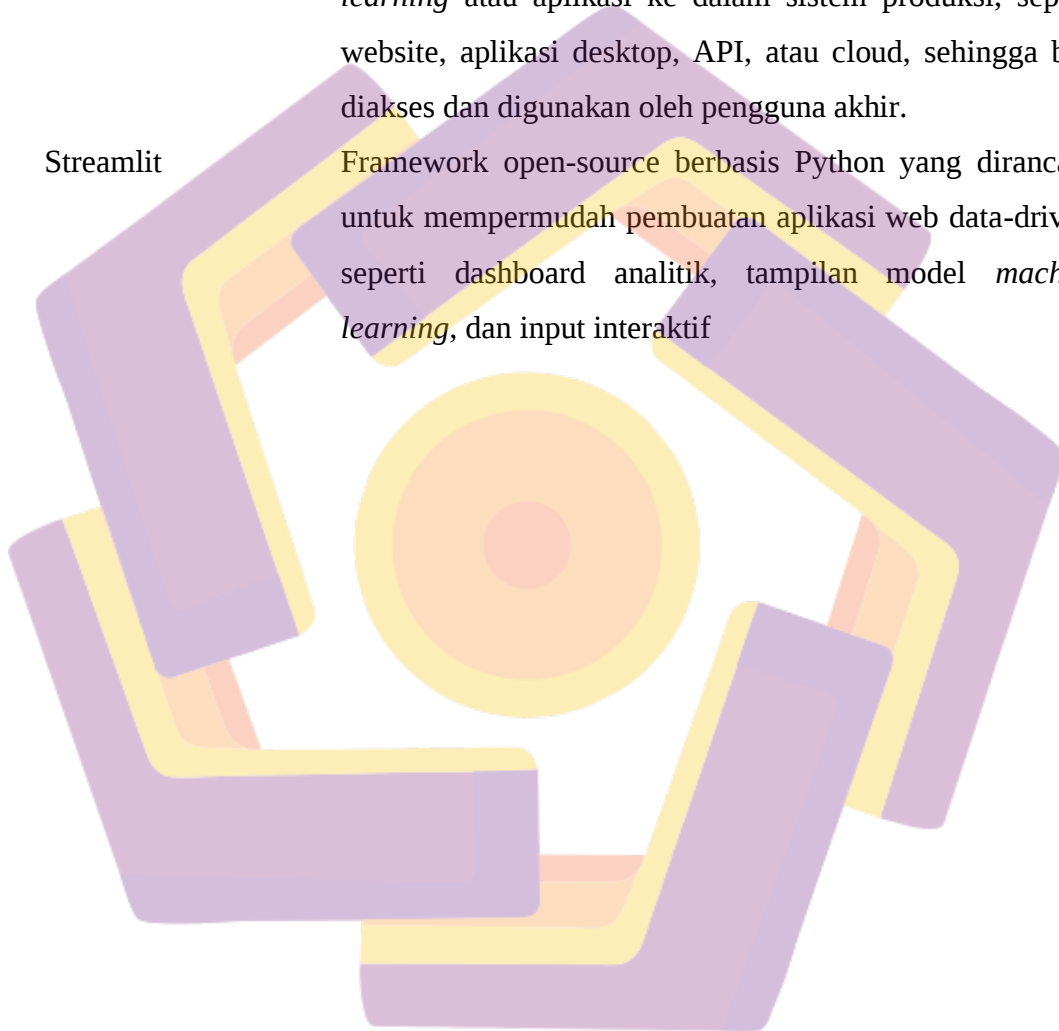
KNN	K-Nearest Neighbor
SMOTE	Synthetic Minority Over-sampling Technique
ADASYN	Adaptive Synthetic Sampling
SVM	Support Vector Machines
TP	True Positif
TN	True Negatif
FP	False Positif
TN	False Negatif



DAFTAR ISTILAH

Machine Learning	Cabang ilmu komputer yang mempelajari algoritma untuk membuat sistem komputer dapat belajar dari data tanpa diprogram secara eksplisit.
Resampling	Teknik untuk menangani ketidakseimbangan data dengan menyesuaikan proporsi kelas mayoritas dan minoritas.
SMOTE	Teknik oversampling yang menciptakan data sintetis dari kelas minoritas dengan interpolasi.
Adasyn	Teknik oversampling yang memfokuskan penciptaan data sintetis di area data minoritas yang sulit dipelajari.
Decision Tree	Model klasifikasi berbentuk pohon keputusan, di mana setiap simpul mewakili keputusan berdasarkan atribut tertentu.
Naive Bayes	Algoritma klasifikasi berdasarkan teori probabilitas Bayes yang mengasumsikan independensi antar fitur.
KNN	Metode klasifikasi yang menentukan kelas berdasarkan kedekatan jarak terhadap k tetangga terdekat.
Random Forest	Algoritma ensemble yang terdiri dari banyak decision tree untuk meningkatkan akurasi prediksi.
Confusion Matrix	Matriks evaluasi performa klasifikasi dengan empat komponen utama: TP, TN, FP, FN.
Accuracy	Proporsi dari jumlah prediksi yang benar (positif dan negatif) dibandingkan dengan total semua prediksi.
Precision	Rasio antara jumlah prediksi positif yang benar (True Positive) terhadap total prediksi yang diklaim positif oleh model.
Recall	Rasio antara jumlah prediksi positif yang benar (True Positive) dibandingkan dengan seluruh kasus aktual yang benar-benar positif.

F1-Score	Metrik evaluasi yang digunakan dalam model klasifikasi, khususnya saat menghadapi dataset tidak seimbang (imbalance). Metrik ini mengukur keseimbangan antara dua metrik Precision dan Recall
Deployment	Proses mengintegrasikan dan menjalankan model <i>machine learning</i> atau aplikasi ke dalam sistem produksi, seperti website, aplikasi desktop, API, atau cloud, sehingga bisa diakses dan digunakan oleh pengguna akhir.
Streamlit	Framework open-source berbasis Python yang dirancang untuk mempermudah pembuatan aplikasi web data-driven, seperti dashboard analitik, tampilan model <i>machine learning</i> , dan input interaktif



INTISARI

Anemia merupakan salah satu masalah kesehatan masyarakat yang serius di tingkat global, terutama bagi kelompok rentan seperti anak-anak, remaja putri, wanita hamil, dan wanita dalam masa nifas. Menurut World Health Organization (WHO), prevalensi anemia mencapai 40% pada anak usia 6–59 bulan, 37% pada wanita hamil, dan 30% pada wanita usia 15–49 tahun. Tingginya prevalensi ini menandakan perlunya strategi deteksi dan penanganan yang lebih tepat guna mengurangi dampak negatif terhadap kualitas hidup masyarakat.

Penelitian ini bertujuan untuk menerapkan dan mengevaluasi model-model *machine learning* dalam klasifikasi penyakit anemia. Algoritma yang digunakan meliputi *Decision Tree*, K-Nearest Neighbor (KNN), Naive Bayes, dan *Random Forest*. Evaluasi performa dilakukan dengan pendekatan dua metode resampling, yaitu SMOTE (*Synthetic Minority Oversampling Technique*) dan ADASYN (*Adaptive Synthetic Sampling*), serta tiga variasi rasio pembagian data pelatihan dan pengujian, yaitu 70:30, 80:20, dan 90:10.

Hasil evaluasi menunjukkan bahwa model *Decision Tree* yang dikombinasikan dengan metode resampling SMOTE dan rasio data pelatihan-pengujian 80:20 memberikan performa terbaik. Model ini mencapai akurasi sebesar 99,60%, presisi 98,89%, recall 99,79%, dan F1-score 99,31%. Temuan ini menunjukkan bahwa pemanfaatan algoritma *machine learning*, khususnya *Decision Tree* dengan teknik resampling yang tepat, berpotensi besar dalam mendukung proses klasifikasi anemia secara lebih akurat dan efisien. Penelitian lanjutan disarankan untuk menggunakan dataset yang lebih besar dan bervariasi, mencakup data dari berbagai institusi kesehatan agar model yang dikembangkan memiliki daya generalisasi yang lebih baik.

Kata kunci: Anemia, *Machine Learning*, *Decision Tree*, SMOTE, Klasifikasi, Data Mining.

ABSTRACT

Anemia is one of the most serious public health issues globally, particularly affecting vulnerable groups such as children, adolescent girls, pregnant women, and postpartum women. According to the World Health Organization (WHO), the global prevalence of anemia reaches 40% among children aged 6–59 months, 37% among pregnant women, and 30% among women aged 15–49 years. This high prevalence underscores the urgent need for effective detection and intervention strategies to mitigate its adverse impact on quality of life.

This study aims to implement and evaluate various machine learning models for the classification of anemia. The algorithms applied include Decision Tree, K-Nearest Neighbor (KNN), Naive Bayes, and Random Forest. Performance evaluation was conducted using two resampling methods—SMOTE (Synthetic Minority Oversampling Technique) and ADASYN (Adaptive Synthetic Sampling)—as well as three different training-testing data split ratios: 70:30, 80:20, and 90:10.

The evaluation results indicate that the Decision Tree model, when combined with SMOTE resampling and an 80:20 training-testing ratio, achieved the best performance. The model obtained an accuracy of 99.60%, precision of 98.89%, recall of 99.79%, and F1-score of 99.31%. These findings demonstrate the strong potential of machine learning algorithms, particularly Decision Tree with appropriate resampling techniques, to enhance the accuracy and efficiency of anemia classification. Further research is recommended to utilize larger and more diverse datasets, incorporating data from multiple healthcare institutions to improve the model's generalizability.

Keyword: Anemia, Machine Learning, Decision Tree, SMOTE, Classification, Data Mining.