

**ANALISIS KOMPRATIF MODEL RANDOM FOREST DAN
XGBOOST UNTUK PREDIKSI RESIKO KANKER
SERVIKS MENGGUNAKAN EXPLAINABEL AI
BERBASIS SHAP**

LAPORAN NON-REGULER

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



Disusun oleh :

MUHAMMAD AGUNG REZA YUDHA

22.11.5314

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

**ANALISIS KOMPRATIF MODEL RANDOM FOREST DAN
XGBOOST UNTUK PREDIKSI RESIKO KANKER
SERVIKS MENGGUNAKAN EXPLAINABEL AI
BERBASIS SHAP**

LAPORAN NON-REGULER

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



Disusun oleh :

MUHAMMAD AGUNG REZA YUDHA

22.11.5314

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

HALAMAN PERSETUJUAN

JALUR NON-REGULER

**ANALISIS KOMPRATIF MODEL RANDOM FOREST DAN
XGBOOST UNTUK PREDIKSI RESIKO KANKER
SERVIKS MENGGUNAKAN EXPLAINABEL AI
BERBASIS SHAP**

yang disusun dan diajukan oleh
MUHAMMAD AGUNG REZA YUDHA
22.11.5314

telah disetujui oleh Dosen Pembimbing
pada tanggal 18 Desember 2025

Dosen Pembimbing.



Majid Rahardi, S.Kom, M.Eng.
NIK. 190302393

HALAMAN PENGESAHAN
JALUR NON-REGULER

ANALISIS KOMPRATIF MODEL RANDOM FOREST DAN XGBOOST
UNTUK PREDIKSI RESIKO KANKER
SERVIKS MENGGUNAKAN EXPLAINABEL AI
BERBASIS SHAP

yang disusun dan diajukan oleh

MUHAMMAD AGUNG REZA YUDHA

22.11.5314

Telah dipertahankan di depan Dewan Penguji
pada tanggal 18 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Anna Baita, S.Kom., M.Kom.
NIK. 190302290

Arifivanto Hadinegoro, S.Kom., M.T.
NIK. 190302289

Majid Rahardi, S.Kom., M.Eng.
NIK. 190302393



Laporan ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 18 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN KARYA

Yang bertanda tangan di bawah ini,

Nama mahasiswa : Muhammad Agung Reza Yudha

NIM : 22.11.5314

Menyatakan bahwa Laporan dengan judul berikut:

Dosen Pembimbing : Majid Rahardi, 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 kegiatan SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggunga jawab Universitas AMIKOM Yogyakarta.
5. Persoalan yang timbul di kemudian sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidak-benaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan persabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 18 Desember 2025

Yang Menyatakan,

Muhammad Agung Reza Yudha

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang tulus dan penuh perenungan, laporan non-reguler ini akhirnya dapat diselesaikan setelah melalui berbagai proses yang tidak selalu mudah. Karya ini penulis persembahkan kepada pihak-pihak yang telah menjadi sumber kekuatan dan makna dalam perjalanan akademik penulis:

1. Allah SWT, atas segala rahmat, petunjuk, dan pertolongan-Nya, sehingga penulis diberi kekuatan untuk bertahan dan menyelesaikan laporan ini.
2. Kedua orang tua tercinta, yang dengan kasih sayang tanpa batas senantiasa mengiringi langkah penulis. Doa yang tidak pernah terucap namun selalu terasa, pengorbanan yang kerap luput dari kata, serta dukungan yang diberikan dengan ketulusan menjadi alasan utama Universitas Amikom Yogyakarta, yang telah memberikan kesempatan, lingkungan akademik yang kondusif, serta fasilitas pendukung bagi proses belajar dan pengembangan diri penulis.
3. Bapak Majid Rahardi selaku Dosen Pembimbing, atas kesabaran, bimbingan, dan arahan yang diberikan selama penyusunan laporan ini.
4. Pemilik NIM 22.93.0368, sebagai bentuk penghargaan atas perjuangan yang sering kali dijalani dalam diam. Setiap kelelahan, keraguan, dan usaha yang dilakukan menjadi bukti komitmen penulis dalam menyelesaikan tanggung jawab akademik dengan sebaik-baiknya.
5. Teman-teman seperjuangan, yang telah menghadirkan kebersamaan, dukungan, dan semangat di tengah berbagai tantangan.

Penulis menyadari bahwa laporan ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat dan kontribusi yang bermakna

KATA PENGANTAR

Dengan mengucapkan rasa syukur ke hadirat Tuhan Yang Maha Esa, penulis dapat menyelesaikan laporan non-reguler ini sesuai dengan tujuan yang telah ditetapkan. Laporan ini disusun sebagai salah satu bentuk pemenuhan kewajiban akademik sekaligus sarana penerapan pengetahuan yang diperoleh selama proses pembelajaran.

Penyusunan laporan ini tidak terlepas dari dukungan berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada Bapak Majid Rahardi selaku Dosen Pembimbing yang telah meluangkan waktu untuk memberikan arahan, masukan, serta bimbingan secara berkesinambungan hingga laporan ini dapat diselesaikan. Apresiasi juga penulis sampaikan kepada Tim Dosen Penguji atas saran dan evaluasi yang konstruktif demi penyempurnaan isi laporan.

Ungkapan terima kasih yang tulus penulis persembahkan kepada orang tua yang senantiasa memberikan doa, motivasi, serta dukungan moral dan material selama proses penyusunan laporan ini. Penulis juga mengucapkan terima kasih kepada pihak kampus beserta seluruh civitas akademika yang telah menciptakan suasana akademik yang kondusif dan menyediakan fasilitas penunjang kegiatan akademik.

Sebagai penutup, penulis menyadari bahwa laporan ini masih memiliki keterbatasan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan di masa mendatang. Semoga laporan ini dapat memberikan manfaat dan menambah wawasan bagi pembaca serta berkontribusi dalam pengembangan ilmu pengetahuan.

Yogyakarta, 18 Desember 2025
Penulis

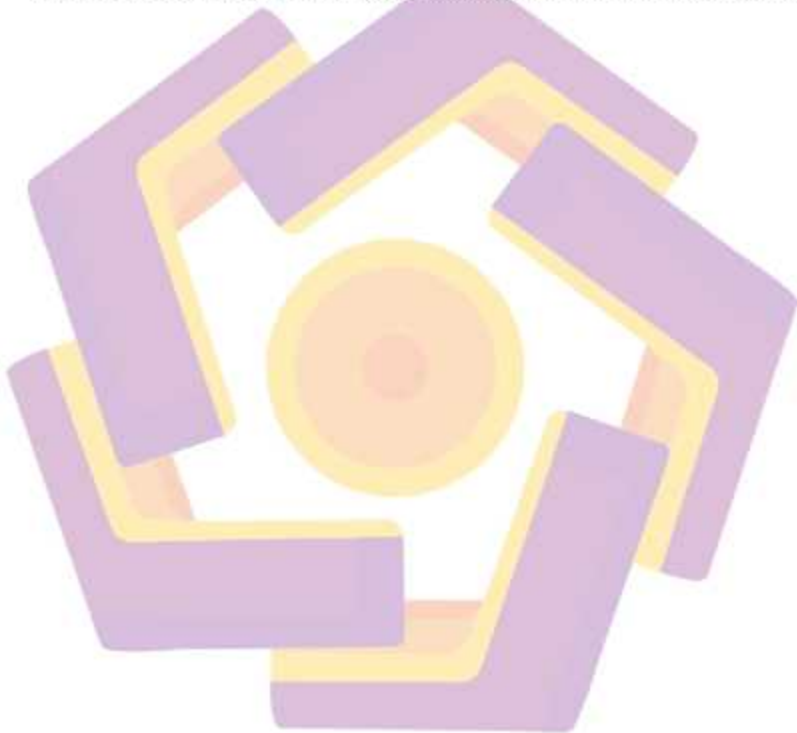
DAFTAR ISI

Halaman Judul.....	i
Halaman Persetujuan.....	ii
Halaman Pengesahan	iii
Halaman Pernyataan Keaslian Karya.....	iv
Halaman Persembahan	v
Kata Pengantar	vi
Daftar Isi.....	vii
Daftar Tabel	ix
Daftar Gambar.....	x
Daftar Lampiran.....	xi
Daftar Lambang dan Singkatan.....	xii
Daftar Istilah.....	xiv
Intisari	xviii
<i>Abstract</i>	xix
Bab I Pendahuluan	1
1.1. Gambaran Umum	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan.....	4
Bab II Tinjauan Pustaka	5
2.1. Studi Literatur.....	5
2.2. Landasan Teori	7
BAB III Metode Penelitian	9
3.1. Metode	9
3.2. Dataset Penelitian.....	10
3.3. Exploratory Data Analysis (EDA).....	12
3.4. Pra-Proses Data	15
3.5. Penyeimbangan Data.....	17
3.6. Pembagian Dataset.....	18
3.7. Pemodelan.....	19

3.8. Hyperparameter Tuning	22
3.9. Evaluasi Model.....	23
3.10. Interpretasi Model (Explainable AI-SHAP).....	25
BAB IV Pembahasan	27
4.1. Deskripsi Dataset dan Analisis Data	27
4.2. Pra-Proses Data	29
4.3. Penyeimbangan Kelas	30
4.4. Evaluasi Model.....	31
4.5. Analisis Interpretabilitas Model Menggunakan SHAP	39
BAB V Kesimpulan	45
4.6. Kesimpulan	45
4.7. Saran.....	47
Referensi	49
Curriculum Vitae.....	52
Lampiran dan Bukti Pendukung.....	53

DAFTAR TABEL

<i>Tabel 3. 1 Dataset</i>	11
<i>Tabel 3. 2 Pembagian Data</i>	19
<i>Tabel 3. 3 Hyperparameter Tuning.....</i>	22
<i>Tabel 4. 1 Hasil Evaluasi Model Random Forest</i>	32
<i>Tabel 4. 2 Evaluasi Model Extreme Gradient Boosting</i>	35
<i>Tabel 4. 3 Perbandingan Kinerja Model</i>	37

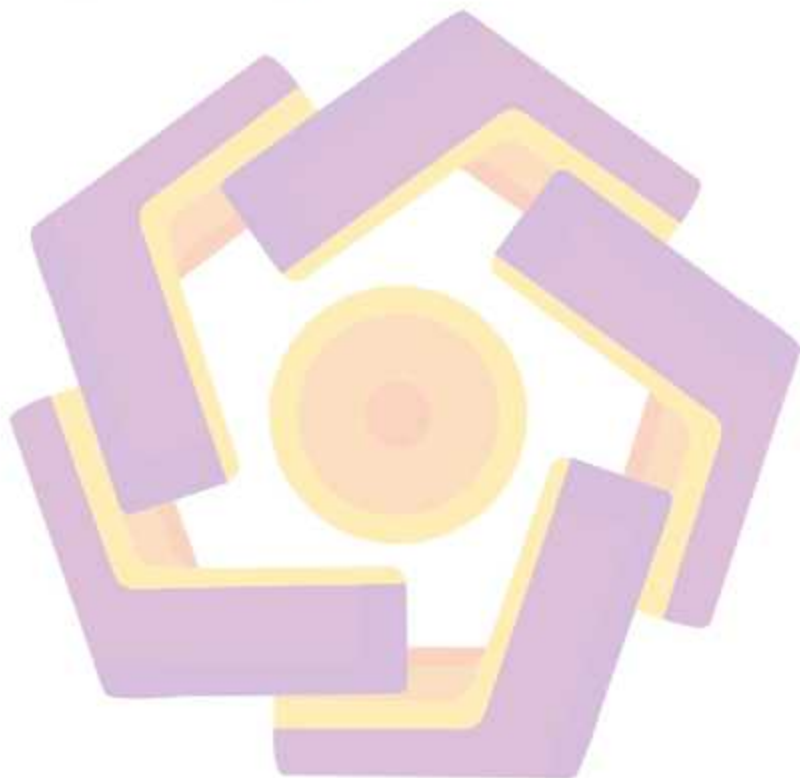


DAFTAR GAMBAR

<i>Gambar 3. 1 Alur Penelitian</i>	10
<i>Gambar 3. 2 Random Forest.....</i>	20
<i>Gambar 4. 1 Target Distribution</i>	27
<i>Gambar 4. 2 Top 10 Features with Missing Values.....</i>	28
<i>Gambar 4. 3 Top 10 Feature Correlated with Target</i>	28
<i>Gambar 4. 4 Distribusi Kelas Sebelum SMOTE.....</i>	30
<i>Gambar 4. 5 Distribusi Kelas Sesudah SMOTE.....</i>	31
<i>Gambar 4. 6 ROC Curve Random Forest.....</i>	33
<i>Gambar 4. 7 Precision Recall Curve Random Forest</i>	34
<i>Gambar 4. 8 ROC Curve XGBoost</i>	36
<i>Gambar 4. 9 Precision Recall Curve XGBoost.....</i>	37
<i>Gambar 4. 10 SHAP Summary Plot Random Forest.....</i>	40
<i>Gambar 4. 11 SHAP Bar Plot Random Forest</i>	41
<i>Gambar 4. 12 SHAP Summary Plot XGBoost</i>	42
<i>Gambar 4. 13 SHAP Waterfall Random Forest.....</i>	43
<i>Gambar 4. 14 SHAP Waterfall XGBoost</i>	44

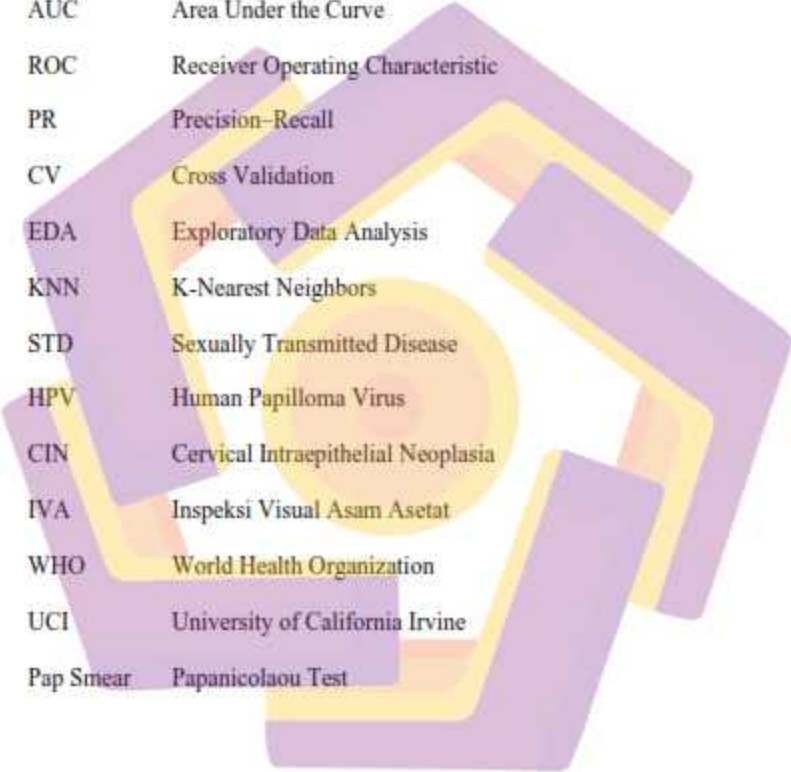
DAFTAR LAMPIRAN

Lampiran 1 Letter of Acceptance (LOA).....	53
Lampiran 2 Lembar Review.....	54
Lampiran 3 Bukti Terbit/Terindex	55
Lampiran 4 Bukti Pembayaran	56



DAFTAR LAMBANG DAN SINGKATAN

$\hat{y}$	Nilai prediksi model
y	Nilai aktual atau label sebenarnya
x	Vektor fitur input
TP	True Positive, yaitu jumlah data positif yang diprediksi benar
TN	True Negative, yaitu jumlah data negatif yang diprediksi benar
FP	False Positive, yaitu jumlah data negatif yang diprediksi sebagai positif
FN	False Negative, yaitu jumlah data positif yang diprediksi sebagai negatif
η	Learning rate pada algoritma XGBoost
λ	Parameter regularisasi L2 pada XGBoost
γ	Parameter penalti kompleksitas pohon keputusan
$\Omega(f)$	Fungsi regularisasi pada XGBoost
$\ell(y, \hat{y})$	Fungsi kehilangan (loss function)
T	Jumlah daun pada pohon keputusan
w	Bobot pada daun pohon keputusan
Q1	Kuartil pertama
Q3	Kuartil ketiga
IQR	Interquartile Range
AI	Artificial Intelligence
XAI	Explainable Artificial Intelligence
ML	Machine Learning



RF	Random Forest
XGBoost	Extreme Gradient Boosting
SHAP	SHapley Additive exPlanations
SMOTE	Synthetic Minority Oversampling Technique
AUC	Area Under the Curve
ROC	Receiver Operating Characteristic
PR	Precision–Recall
CV	Cross Validation
EDA	Exploratory Data Analysis
KNN	K-Nearest Neighbors
STD	Sexually Transmitted Disease
HPV	Human Papilloma Virus
CIN	Cervical Intraepithelial Neoplasia
IVA	Inspeksi Visual Asam Asetat
WHO	World Health Organization
UCI	University of California Irvine
Pap Smear	Papanicolaou Test

DAFTAR ISTILAH



Kanker Serviks	Jenis kanker yang berkembang pada leher rahim (serviks) akibat pertumbuhan sel abnormal, umumnya dipicu oleh infeksi Human Papilloma Virus (HPV).
Machine Learning	Cabang kecerdasan buatan yang mempelajari pola dari data untuk menghasilkan model prediksi atau klasifikasi tanpa pemrograman eksplisit.
Artificial Intelligence	Bidang ilmu komputer yang berfokus pada pengembangan sistem yang mampu meniru kecerdasan manusia dalam pengambilan keputusan.
Explainable AI	Pendekatan kecerdasan buatan yang bertujuan membuat proses dan hasil prediksi model dapat dipahami oleh manusia.
Random Forest	Algoritma machine learning berbasis ensemble yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi dan stabilitas prediksi.
XGBoost	Algoritma boosting berbasis pohon keputusan yang menggunakan optimasi gradien dan regularisasi untuk meningkatkan performa model.
Ensemble Learning	Pendekatan pembelajaran mesin yang menggabungkan beberapa model untuk memperoleh hasil prediksi yang lebih akurat dan robust.

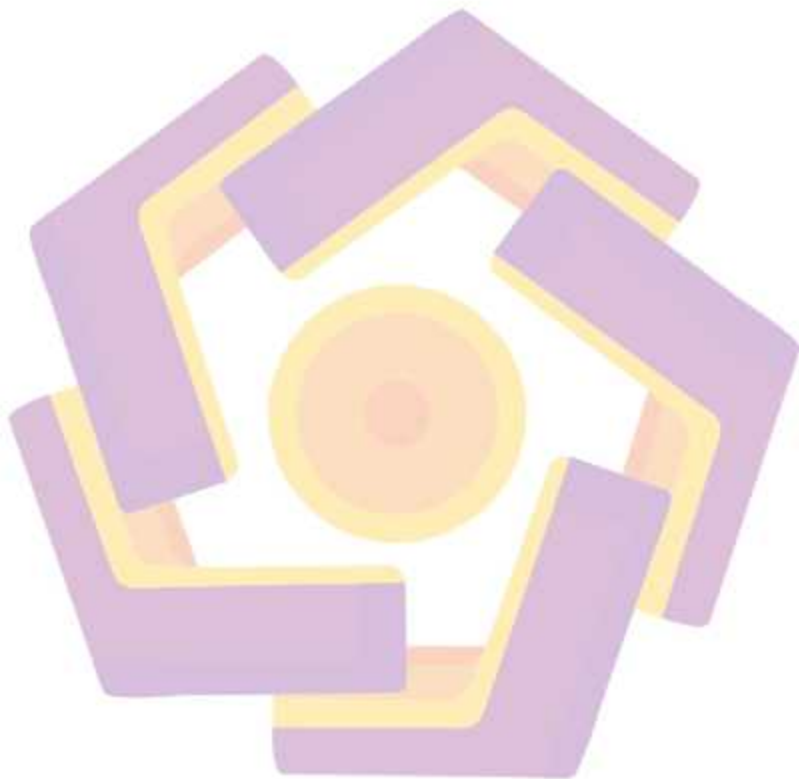


SMOTE	Teknik penyeimbangan data dengan membuat sampel sintetis pada kelas minoritas untuk mengatasi ketidakseimbangan kelas.
Class Imbalance	Kondisi distribusi data yang tidak seimbang antara kelas mayoritas dan minoritas dalam dataset.
Hyperparameter	Parameter yang ditentukan sebelum proses pelatihan model untuk mengontrol perilaku dan kompleksitas algoritma.
Hyperparameter Tuning	Proses pencarian kombinasi hyperparameter terbaik guna meningkatkan performa model.
Exploratory Data Analysis	Tahap analisis awal untuk memahami karakteristik data, pola distribusi, dan potensi permasalahan sebelum pemodelan.
Missing Value	Nilai data yang tidak tersedia atau kosong pada suatu fitur dalam dataset.
Imputasi	Proses pengisian nilai hilang pada dataset menggunakan metode statistik atau algoritmik.
Cross Validation	Metode evaluasi model dengan membagi data ke dalam beberapa subset untuk menguji konsistensi performa model.
Accuracy	Ukuran kinerja model yang menunjukkan proporsi prediksi yang benar terhadap seluruh data.
Precision	Metrik evaluasi yang menunjukkan tingkat ketepatan model dalam memprediksi kelas positif.



Recall	Metrik evaluasi yang mengukur kemampuan model dalam mendeteksi seluruh data kelas positif.
F1-Score	Rata-rata harmonis antara precision dan recall untuk menilai keseimbangan performa model.
AUC	Ukuran performa model yang menunjukkan kemampuan model dalam membedakan kelas positif dan negatif.
ROC Curve	Kurva evaluasi yang menggambarkan hubungan antara true positive rate dan false positive rate.
Precision-Recall Curve	Kurva evaluasi yang menunjukkan hubungan antara precision dan recall pada berbagai threshold.
SHAP	Metode interpretasi model berbasis teori permainan untuk menjelaskan kontribusi setiap fitur terhadap prediksi.
Global Interpretation	Interpretasi model yang menjelaskan pengaruh fitur secara keseluruhan terhadap prediksi.
Local Interpretation	Interpretasi model yang menjelaskan kontribusi fitur terhadap prediksi pada satu data individu.
Overfitting	Kondisi ketika model terlalu menyesuaikan diri dengan data latih sehingga kinerjanya menurun pada data baru.
Underfitting	Kondisi ketika model gagal menangkap pola penting dalam data sehingga performanya rendah.

Decision Support System Sistem berbasis komputer yang membantu pengambilan keputusan dengan memanfaatkan data dan model analitik.



INTISARI

Kanker serviks masih menjadi salah satu penyebab utama kematian akibat kanker pada perempuan, khususnya di negara berkembang seperti Indonesia. Penelitian ini bertujuan untuk mengembangkan model prediksi risiko kanker serviks yang akurat dan dapat diinterpretasikan menggunakan algoritma Random Forest (RF) dan Extreme Gradient Boosting (XGBoost). Dataset yang digunakan adalah *Cervical Cancer Risk Factors* dari UCI Repository yang terdiri dari 858 data pasien dengan 36 fitur klinis dan demografis.

Tahapan praproses data meliputi imputasi nilai hilang, penyeimbangan kelas menggunakan *Synthetic Minority Oversampling Technique* (SMOTE), serta optimasi hiperparameter melalui *Randomized Search CV*. Hasil eksperimen menunjukkan bahwa kedua model mencapai kinerja yang tinggi, dengan tingkat akurasi di atas 96% dan nilai AUC lebih dari 0,95, di mana model XGBoost (Tuned + SMOTE) sedikit lebih unggul dibandingkan RF dalam mendeteksi kasus positif.

Analisis interpretabilitas menggunakan *SHapley Additive exPlanations* (SHAP) mengidentifikasi fitur-fitur klinis seperti Tes Schiller, Tes Hinselmann, dan Hasil Sitologi sebagai faktor yang paling berpengaruh dalam proses klasifikasi, sejalan dengan bukti klinis yang telah ada. Dengan demikian, integrasi XGBoost, SMOTE, dan SHAP menghasilkan kerangka prediksi yang tidak hanya memiliki akurasi tinggi tetapi juga dapat dijelaskan secara klinis, sehingga mendukung pengembangan sistem pendukung keputusan untuk deteksi dini kanker serviks.

Kata kunci: Kanker Serviks, XGBoost, Random Forest, SMOTE, SHAP.

ABSTRACT

Cervical cancer remains one of the leading causes of cancer-related deaths among women, particularly in developing countries such as Indonesia. This study aims to develop an accurate and interpretable predictive model for cervical cancer risk using Random Forest (RF) and Extreme Gradient Boosting (XGBoost) algorithms. The dataset used is the Cervical Cancer Risk Factors from the UCI Repository, consisting of 858 patient records and 36 clinical and demographic features.

The preprocessing stages include missing value imputation, class balancing using Synthetic Minority Oversampling Technique (SMOTE), and hyperparameter optimization through Randomized Search CV. Experimental results show that both models achieved high performance, with accuracy exceeding 96% and AUC above 0.95, while the XGBoost (Tuned + SMOTE) model slightly outperformed RF in detecting positive cases.

The interpretability analysis using SHapley Additive exPlanations (SHAP) identified clinical features such as Schiller Test, Hinselmann Test, and Cytology Result as the most influential factors in the classification process, consistent with established clinical evidence. Therefore, the integration of XGBoost, SMOTE, and SHAP provides a predictive framework that is not only highly accurate but also clinically explainable, supporting the development of decision-support systems for early cervical cancer detection.

Keyword: *Cervical cancer, XGBoost, Random Forest, SMOTE, SHAP*