

**KOMPARASI ALGORITMA SVM DAN XGBOOST PADA
KLASIFIKASI KANKER PARU-PARU**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Informatika



disusun oleh
ROSSA RASSOMA DALVITA
21.11.3821

Kepada

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

**KOMPARASI ALGORITMA SVM DAN XGBOOST PADA
KLASIFIKASI KANKER PARU-PARU**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi *SI-Informatika*



disusun oleh

ROSSA RASSOMA DALVITA

21.11.3821

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**KOMPARASI ALGORITMA SVM DAN XGBOOST PADA KLASIFIKASI
KANKER PARU-PARU**

yang disusun dan diajukan oleh

Rossa Rassoma Dalvita

21.11.3821

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 21 Agustus 2025

Dosen Pembimbing,



Subektiningsih M.Kom

NIK. 190302413

HALAMAN PENGESAHAN
SKRIPSI
KOMPARASI ALGORITMA SVM DAN XGBOOST PADA KLASIFIKASI
KANKER PARU-PARU

yang disusun dan diajukan oleh

Rossa Rassoma Dalvita

21.11.3821

Telah dipertahankan di depan Dewan Penguji
pada tanggal 21 Agustus 2025

Sesunan Dewan Penguji

Nama Penguji

Tanda Tangan

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

Windhy Mega Pradnya Dhubita S.Kom., M.Kom.,
NIK. 190302185

Subektiningsih S.Kom., M.Kom.,
NIK. 190302413

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
tanggal 21 Agustus 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **Rossa Rassoma Dalvita**
NIM : **21.11.3821**

Menyatakan bahwa Skripsi dengan judul berikut:

Komparasi Algoritma SVM dan XGBoost Pada Klasifikasi Kanker Paru-Paru

Dosen Pembimbing : **Subektiningsih, M.Kom**

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan 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, 21 Agustus 2025

Yang Menyatakan,



Rossa Rassoma Dalvita

HALAMAN PERSEMBAHAN

Dengan penuh rasa Syukur kepada Allah SWT atas segala hidayah, ilmu, dan, rahmat-Nya, skripsi ini penulis persembahkan kepada :

1. Alm Bapak H. M Budoyo S.I.P. yang telah membangun karakter penulis dan menanamkan nilai kebaikan saat kecil dahulu sehingga penulis masih kuat untuk menjalani hidup dan dapat menyelesaikan penulisan ini.
2. Nasirudin, S.Kom dan Anggun Yuliani, S.E yang selalu memberikan doa, kekuatan semangat moril, dan juga bantuan finansial hingga penulisan ini dapat berjalan dengan cukup lancar.
3. Ibu Saryati yang selalu memberikan penulis makanan bergizi dan juga selalu bertanya kapan penulisan skripsi ini selesai serta Ririn Kartikawati, S.E yang selalu menanyakan kapan penulis ini lulus.
4. Ibu Subektiningsih, M.Kom yang selalu memberikan doa hingga arahan sehingga penulis dapat menyelesaikan penulisan ini dengan cukup lancar, serta seluruh dosen yang telah memberikan ilmu kepada penulis pada saat masa studi.
5. Teman-teman penulis dari informatika 01 khususnya sapto, agus, sigit, hafiz, kevin-lkhwan, nanang, dan rizqi yang selalu tanpa pamrih membantu penulis dalam menjalankan studi maupun penulisan naskah skripsi ini.
6. Sefia Osiana yang selalu menanyakan sampai mana penulisan ini telah ditulis, pun selalu mendampingi dan memberikan semangat serta doa agar penulisan ini dapat selesai dengan cukup baik.

KATA PENGANTAR

Puji Syukur kehadirat Allah SWT yang telah melimpahkan rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul "Komparasi Algoritma SVM dan XGBoost Pada Klasifikasi Kanker Paru-Paru". Skripsi ini disusun sebagai syarat untuk memperoleh gelar sarjana pada program studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dalam Penyusunan Skripsi ini penulis sadar bahwa penulisan ini tidak dapat selesai tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Oleh karena itu, pada kesempatan kali ini penulis ingin mengucapkan terimakasih yang sebesar-besarnya kepada :

1. Allah SWT yang telah melimpahkan rahmat, pemahaman, serta hidayah-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan cukup baik.
2. Alm Bapak H. M Budoyo, S.I.P. yang telah mendidik karakter dan membangun nilai hidup penulis saat kecil sehingga penulis dapat sabar dan kuat dalam menyelesaikan penulisan ini.
3. Nasirudin, S.Kom dan Anggun Yuliani, S.E yang telah membantu penulis dari awal hingga saat ini dalam bentuk semangat, doa, dukungan dan finansial sehingga penulis dapat menyelesaikan penulisan ini dengan cukup lancar.
4. Ibu Saryati yang juga selalu memberikan doa dan dukungan, serta Ririn Kartikawati, S.E selaku kakak penulis yang juga telah memberikan banytuan kepada penulis.
5. Kepada Ibu Subektiningsih, M.Kom yang selalu memberikan doa, arahan dengan sabar dan juga kuat dalam membimbing penulis dalam penyusunan skripsi ini.
6. Teman-teman penulis khususnya teman terdekat penulis yang selalu memberikan dukungan penuh, semangat, dan doa agar penulisan ini berjalan dengan cepat.

Penulis sepenuhnya menyadari bahwa skripsi ini jauh dari kata baik, dari segi isi maupun teknis penulisannya. Oleh karena itu penulis memohon maaf yang sebesar-besarnya atas segala kekurangan yang ada, dan penulis sangat terbuka atas segala kritik dan saran yang membangun demi perbaikan dimasa yang akan datang. Semoga skripsi ini dapat bermanfaat bagi para pembaca dan perkembangan ilmu pengetahuan. Akhir kata Penulis mengucapkan rasa syukur yang sebesar-besarnya kepada semua pihak yang terlibat secara langsung maupun tidak, semoga Allah SWT membalas semua kebaikan yang telah diberikan.

Yogyakarta, 5 Agustus 2025

Penulis



DAFTAR ISI

HALAMAN PERSETUJUAN	2
HALAMAN PENGESAHAN SKRIPSI	3
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	4
HALAMAN PERSEMBAHAN	5
KATA PENGANTAR	6
DAFTAR ISI.....	8
DAFTAR TABEL	11
DAFTAR GAMBAR.....	12
DAFTAR LAMPIRAN.....	13
DAFTAR LAMBANG DAN SINGKATAN	14
DAFTAR ISTILAH.....	15
INTISARI	16
<i>ABSTRACT</i>	17
BAB I PENDAHULUAN	18
1.1 Latar Belakang	18
1.2 Rumusan Masalah	19
1.3 Batasan Masalah.....	20
1.4 Tujuan Penelitian.....	20
1.5 Manfaat Penelitian	21
1.6 Sistematika Penulisan.....	23
BAB II TINJAUAN PUSTAKA	24

2.1	Studi Literatur	24
2.2	Dasar Teori.....	13
2.2.1	Kanker Paru-Paru	13
2.2.2	<i>Machine Learning</i>	14
2.2.3	Klasifikasi	15
2.2.4	<i>Support Vector Machine (SVM)</i>	16
2.2.5	<i>Extreme Gradient Boosting (XGBoost)</i>	17
2.2.6	<i>Preprocessing Data</i>	19
2.2.7	Evaluasi Model.....	20
BAB III METODE PENELITIAN		23
3.1	Objek Penelitian	23
3.2	Alur Penelitian.....	24
3.2.1	Identifikasi Masalah	25
3.2.2	<i>Dataset</i>	26
3.2.3	<i>Data Preprocessing</i>	27
3.2.4	Pembangunan Model.....	28
3.2.5	Evaluasi Model.....	30
3.2.6	Analisis dan Perbandingan.....	32
3.3	Alat dan Bahan	33
3.3.1	Data Penelitian	33
3.3.2	Alat/Instrumen Penelitian	33
3.3.2.1	Perangkat Keras:.....	34
3.3.2.2	Perangkat Lunak:	34
BAB IV.....		35
4.1	Eksplorasi Data (<i>Exploratory Data Analysis</i>)	35
4.2	<i>Data Preprocessing</i>	37

4.2.1	Penghapusan Kolom Tidak Relevan:	37
4.2.2	Transformasi Label Kategorikal:	38
4.2.3	Pemilihan Fitur:	38
4.2.3.1	Air Pollution (Polusi Udara):	38
4.2.3.2	Alcohol Use (Konsumsi Alkohol)	39
4.2.3.3	Fatigue (Kelelahan):	39
4.2.4	Normalisasi Fitur:	40
4.3	Perbandingan Model Algoritma	41
4.3.1	Model Support Vector Machine (SVM):	41
4.3.2	Model Extreme Gradient Boosting (XGBoost):	42
4.3.3	Hasil Evaluasi Akurasi	43
4.4	Kesimpulan Komparasi Model	44
BAB V PENUTUP		46
5.1	Kesimpulan	46
5.2	Saran	46
REFERENSI		48
LAMPIRAN		52

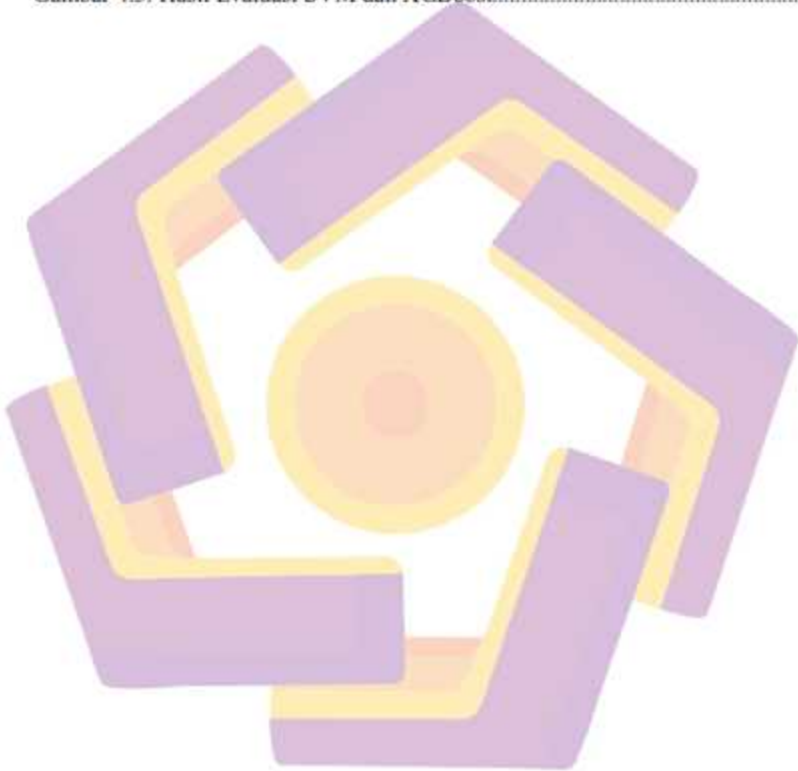
DAFTAR TABEL

Tabel 2.2. Keaslian Penelitian.....	11
Tabel 3.1. Penjelasan Variable Penelitian.....	26
Tabel 4.1. Sebelum kolom 'Patient Id' dihapus.....	37
Table 4.2. Setelah kolom 'Patient Id' dihapus.....	37
Table 4.3. Sebelum kolom 'Level' diubah menjadi tipe numerik.....	37
Tabel 4.4. Setelah kolom 'Level' diubah menjadi tipe numerik.....	38
Tabel 4.5. Hasil Evaluasi.....	42



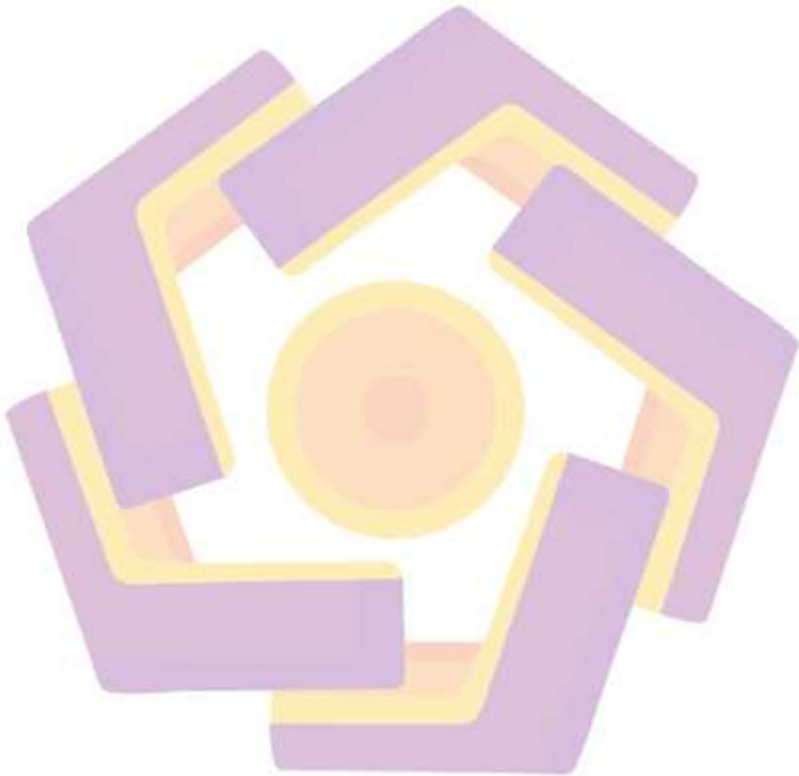
DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian.....	25
Gambar 4.1. Distribusi Data Target.....	36
Gambar 4.2. Heatmap korelasi.....	39
Gambar 4.3. Hasil Evaluasi SVM dan XGBoost.....	43



DAFTAR LAMPIRAN

Lampiran 1. Source Code model penelitian.....	52
Lampiran 2. Source Code deploy phyton.....	55
Lampiran 3. Tampilan deploy.....	58



DAFTAR LAMBANG DAN SINGKATAN

SVM	Support Vector Machine
XGBoost	Extreme Gradient Boosting
ML	Machine Learning
CV	Cross Validation
TP	True Positive
FP	False Positive
FN	False Negative
TN	True Negative
RBF	Radial Basis Function
AUC	Area Under Curve



DAFTAR ISTILAH



StandardScaler	Fungsi dari library Scikit-learn yang digunakan untuk menstandarkan nilai fitur agar memiliki distribusi dengan mean = 0 dan standar deviasi = 1.
K-Fold Cross Validation	Metode validasi model dengan membagi data menjadi k bagian (fold)
Akurasi	Persentase prediksi yang benar terhadap seluruh data uji
Presisi	Rasio antara jumlah prediksi benar positif dengan total prediksi positif
Recall	Rasio antara jumlah prediksi benar positif dengan total data aktual positif
F1-Score	Harmonik rata-rata dari presisi dan recall

INTISARI

Kanker paru-paru merupakan salah satu penyakit dengan angka kematian tertinggi di dunia, namun proses diagnosis tradisional seringkali terhambat oleh keterbatasan tenaga medis dan subjektivitas interpretasi hasil klinis. Penelitian ini membandingkan dua algoritma machine learning—Support Vector Machine (SVM) dan Extreme Gradient Boosting (XGBoost)—dalam mengklasifikasikan tingkat keparahan kanker paru-paru (Low, Medium, High) menggunakan dataset “Cancer Patients and Air Pollution: A New Link” dari Kaggle. Tahap penelitian mencakup preprocessing data (pembersihan, konversi label kategorikal menjadi numerik, dan normalisasi fitur menggunakan StandardScaler), pemilihan tiga fitur utama (Air Pollution, Alcohol use, Fatigue), pembangunan model SVM dan XGBoost, serta evaluasi menggunakan 5-fold cross validation dengan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan SVM mencapai akurasi rata-rata sebesar 92%, sedangkan XGBoost unggul dengan akurasi 97%, sekaligus memperlihatkan presisi, recall, dan F1-score yang lebih tinggi. Temuan ini mengindikasikan bahwa XGBoost lebih efektif dalam menangani klasifikasi multi-kelas pada data kesehatan yang kompleks. Kontribusi penelitian ini diharapkan bermanfaat bagi peneliti dan praktisi teknologi informasi di bidang medis untuk mengembangkan sistem pendukung keputusan klinis, dan bagi institusi kesehatan dalam meningkatkan akurasi diagnosis kanker paru-paru. Rekomendasi penelitian lanjutan meliputi penambahan fitur relevan, penerapan teknik penyeimbangan data, tuning hyperparameter, serta integrasi model ke dalam platform web atau mobile.

Kata kunci: Kanker paru-paru, komparasi, klasifikasi, SVM, XGBoost, machine learning

ABSTRACT

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, yet traditional diagnostic methods often face limitations such as subjective interpretation and resource constraints. This study compares the performance of two machine learning algorithms—Support Vector Machine (SVM) and Extreme Gradient Boosting (XGBoost)—for classifying lung cancer severity levels (Low, Medium, High) using the publicly available “Cancer Patients and Air Pollution: A New Link” dataset from Kaggle. The methodology comprises three main stages: data preprocessing, including removal of irrelevant identifiers, conversion of categorical severity labels into numeric form, and feature standardization via *StandardScaler*; selection of three key predictive features—Air Pollution, Alcohol Use, and Fatigue—and construction of SVM and XGBoost classifiers; and evaluation via 5-fold cross-validation using accuracy, precision, recall, and F1-score. Experimental results demonstrate that SVM achieves an average accuracy of 92%, whereas XGBoost attains 97%, alongside superior precision, recall, and F1-score metrics. These findings indicate that XGBoost is more effective and robust for multi-class classification in complex health datasets. The outcomes of this research are expected to inform healthcare data scientists and clinical decision support system developers, and future work may explore the inclusion of additional features, data-balancing techniques, hyperparameter optimization, and deployment in web or mobile platforms.

Keyword: Lung cancer, comparison, classification, SVM, XGBoost, comparison