

**ANALISIS KOMPARATIF ALGORITMA MACHINE LEARNING
(STUDI KASUS: RANDOM FOREST, XGBOOST, DAN SUPPORT
VECTOR MACHINE) UNTUK PREDIKSI DINI SINDROM
OVARIUM POLIKISTIK (PCOS)**

SKRIPSI

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



disusun oleh

NADIA SILVIA SARI

22.11.5167

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS KOMPARATIF ALGORITMA MACHINE LEARNING
(STUDI KASUS: RANDOM FOREST, XGBOOST, DAN SUPPORT
VECTOR MACHINE) UNTUK PREDIKSI DINI SINDROM
OVARIUM POLIKISTIK (PCOS)**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
NADIA SILVIA SARI
22.11.5167

Kepada
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KOMPARATIF ALGORITMA MACHINE LEARNING (STUDI
KASUS: RANDOM FOREST, XGBOOST, DAN SUPPORT VECTOR
MACHINE) UNTUK PREDIKSI DINI SINDROM OVARIUM POLIKISTIK
(PCOS)**

yang disusun dan diajukan oleh

Nadia Silvia Sari
22.11.5167

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 20 Mei 2026

Dosen Pembimbing,



Windha Mega Pradnya D. M. Kom
NIK. 190302185

HALAMAN PENGESAHAN
SKRIPSI
ANALISIS KOMPARATIF ALGORITMA MACHINE LEARNING
(STUDI KASUS: RANDOM FOREST, XGBOOST, DAN SUPPORT
VECTOR MACHINE) UNTUK PREDIKSI DINI SINDROM OVARIUM
POLIKISTIK (PCOS)

yang disusun dan diajukan oleh

Nadia Silvia Sari

22.11.5167

Telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Mei 2026

Susunan Dewan Penguji

Nama Penguji

Anna Baita, M.Kom

NIK. 190302290

Rumini M.Kom

NIK. 190302246

Windha Mega Pradnya D. M.Kom

NIK. 190302185

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Mei 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertanda tangan di bawah ini,

Nama mahasiswa : Nadia Silvia Sari
NIM : 22.11.5167

Menyatakan bahwa Skripsi dengan judul berikut:

Analisis Komparatif Algoritma Machine Learning (Studi Kasus: Random Forest, XGBoost, dan Support Vector Machine) Untuk Prediksi Dini Sindrom Ovarium Polikistik (PCOS)

Dosen Pembimbing : Windha Mega Pradnya D, 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 tidak benaran 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, 20 Mei 2026

Yang Menyatakan,



Nadia Silvia Sari

HALAMAN PERSEMBAHAN

Dengan segala rasa puji syukur kepada Allah SWT atas segala rahmat dan kemudahan yang telah diberikan, Alhamdulillah perjalanan panjang ini akhirnya dapat terselesaikan dengan baik. Oleh karena itu, dengan rasa bangga dan bahagia saya ucapkan rasa syukur dan terima kasih dengan sepenuh hati kepada orang-orang terkasih yang telah mendampingi setiap langkah perjalanan saya.

1. Untuk Ibunda tercinta, Sri Ani. Terima kasih telah menjadi alasan terkuat saya untuk terus berjuang. Terima kasih telah menjadi ibu yang terus memberikan dukungan moril maupun materil serta do'anya yang selalu mendampingi di setiap langkah. Setiap tetes keringat dan doa yang ibu panjatkan adalah bentuk semangat yang menggerakkan saya hingga titik ini. Tiada kata dan perbuatan yang cukup untuk membalas semua kasih sayang, pengorbanan, dan kepercayaan yang telah ibu berikan.
2. Untuk Ibu Karyati, Ibu Maryah, Bapak Misbahkul Munir, dan Bapak Muhammad Nur. Terima kasih telah hadir sebagai tangan-tangan kasih yang turut memeluk dan membesarkan saya. Bersedia memberikan kehangatan, kasih sayang, serta dukungan yang membuat saya merasa tidak pernah sendiri. Kebaikan, kehadiran, serta dukungan kalian adalah anugerah yang selalu saya syukuri dan tidak akan pernah terlupakan.
3. Untuk kak Maya dan kak Tarisa. Teruntuk dua sosok sahabat yang sekaligus memberikan peran kakak yang selalu ada untuk kebersamaan. Terima kasih telah bersedia menemani di setiap kondisi, turut memberikan kebahagiaan dan menjadi tempat yang nyaman untuk kehidupan saya. Selalu bersyukur kalian ada.
4. Untuk teman-teman semasa kuliah, terima kasih telah mengisi hari-hari perkuliahan dengan tawa, semangat, dan kenangan yang tidak akan terlupakan. Perjalanan ini jauh lebih ringan dan bermakna karena kalian.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga skripsi ini dapat diselesaikan dengan baik sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta.

Penulis menyadari sepenuhnya bahwa penyelesaian skripsi ini tidak terlepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, pada kesempatan ini penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Sri Ani selaku orang tua penulis, atas segala doa, dukungan moral, dan pengorbanan yang tidak pernah berhenti diberikan selama penulis menempuh studi hingga menyelesaikan skripsi ini.
2. Prof. Dr. Kusri, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta, atas kepemimpinan dan fasilitas akademik yang telah disediakan selama penulis menempuh studi.
3. Ibu Windha Mega Pradnya D, M.Kom. selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta masukan yang sangat berarti selama proses penyusunan skripsi ini dari awal hingga akhir.
4. Seluruh Tim Dosen Penguji yang telah bersedia meluangkan waktu untuk menguji, memberikan penilaian, serta masukan yang konstruktif demi penyusunan skripsi yang sempurna.
5. Seluruh Dosen dan Staf Pengajar Program Studi Informatika Universitas Amikom Yogyakarta yang telah memberikan ilmu pengetahuan dan pengalaman berharga selama penulis menempuh pendidikan.
6. Seluruh rekan mahasiswa Program Studi Informatika angkatan 2022 yang telah memberikan semangat, bantuan, dan kebersamaan selama masa perkuliahan.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang bersifat membangun sangat penulis harapkan demi perbaikan di masa mendatang. Penulis berharap skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi kontribusi yang berarti bagi pengembangan ilmu pengetahuan.

Yogyakarta, 20 Mei 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	II
HALAMAN PERSETUJUAN	III
HALAMAN PENGESAHAN.....	IV
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	V
HALAMAN PERSEMBAHAN	VI
KATA PENGANTAR.....	VII
DAFTAR ISI.....	VIII
DAFTAR TABEL	X
DAFTAR GAMBAR.....	XI
DAFTAR LAMBANG DAN SINGKATAN	XII
DAFTAR ISTILAH	XIV
INTISARI	XVI
<i>ABSTRACT</i>	XVII
BAB I PENDAHULUAN.....	1
1.1 LATAR BELAKANG.....	1
1.2 RUMUSAN MASALAH.....	3
1.3 BATASAN MASALAH.....	3
1.4 TUJUAN PENELITIAN	4
1.5 MANFAAT PENELITIAN	5
1.6 SISTEMATIKA PENULISAN.....	5
BAB II TINJAUAN PUSTAKA	8
2.1 STUDI LITERATUR.....	8
2.2 DASAR TEORI	16
BAB III METODE PENELITIAN.....	33
3.1 OBJEK PENELITIAN	33
3.2 ALUR PENELITIAN.....	33

3.3	ALAT DAN BAHAN	48
BAB IV HASIL DAN PEMBAHASAN		49
4.1	DESKRIPSI DATA	49
4.2	EXPLORATORY DATA ANALYSIS (EDA)	49
4.3	DATA PREPROCESSING	52
4.4	SKENARIO PELATIHAN DAN PENYEIMBANGAN DATA	55
4.5	HASIL EVALUASI KOMPARATIF MODEL	56
4.6	VALIDASI KONSISTENSI MODEL TERBAIK	61
4.7	ANALISIS PERMUTATION FEATURE IMPORTANCE	61
4.8	IMPLEMENTASI PROTOTYPE SISTEM SKRINING	64
BAB V PENUTUP		67
5.1	KESIMPULAN	67
5.2	SARAN	68
REFERENSI		69
LAMPIRAN		72

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	12
Tabel 2.2 Confusion Matrix	28
Tabel 3.1 Ilustrasi Dataset PCOS	39
Tabel. 3.2 Majority Voting Random Forest	40
Tabel 3.3 Nilai loss log-loss biner	41
Tabel 3.4 Konvergensi XGBoost	42
Tabel 3.6 Hierarki Kriteria Seleksi Model Optimal	46
Tabel 4.1 Ekstraksi Korelasi Fitur	51
Tabel 4.2 Komposisi Data	56
Tabel 4.3 Hasil Metrik Evaluasi Ketiga Model	57
Tabel 4.4 Hasil Metrik Evaluasi Ketiga model	58
Tabel 4.5 Metrik Evaluasi Nilai Standar Deviasi	59
Tabel 4.6 <i>Confusion Matrix</i> rata-rata per Model	60

DAFTAR GAMBAR

Gambar 2.1 Random Forest	25
Gambar 3.1 Alur Penelitian	33
Gambar 4.1 Informasi Data	49
Gambar 4.2 Identifikasi nilai hilang	50
Gambar 4.3 Distribusi Kelas Target	51
Gambar 4.4 <i>Heatmap</i> Korelasi Antar Fitur	52
Gambar 4.5 Data Cleaning	53
Gambar 4.6 Source Code Imputasi nilai hilang	54
Gambar 4.7 Grafik <i>Standardisasi</i> Fitur	54
Gambar 4.8 <i>Chart</i> Perbandingan Rata-rata Metrik	56
Gambar 4.9 Box Plot Distribusi Metrik per-fold	57
Gambar 4.10 Evaluasi Final Random Forest	61
Gambar 4.11 10 Fitur Teratas Random Forest	62
Gambar 4.12 <i>Permutation Feature Importance</i> Algoritma XGBoost	63
Gambar 4.13 <i>Permutation Feature Importance</i> Algoritma SVM	64
Gambar 4.14 Antarmuka Sistem Skrining PCOS	65
Gambar 4.15 Tampilan Hasil Prediksi Sistem Skrining PCOS	66

DAFTAR LAMBANG DAN SINGKATAN



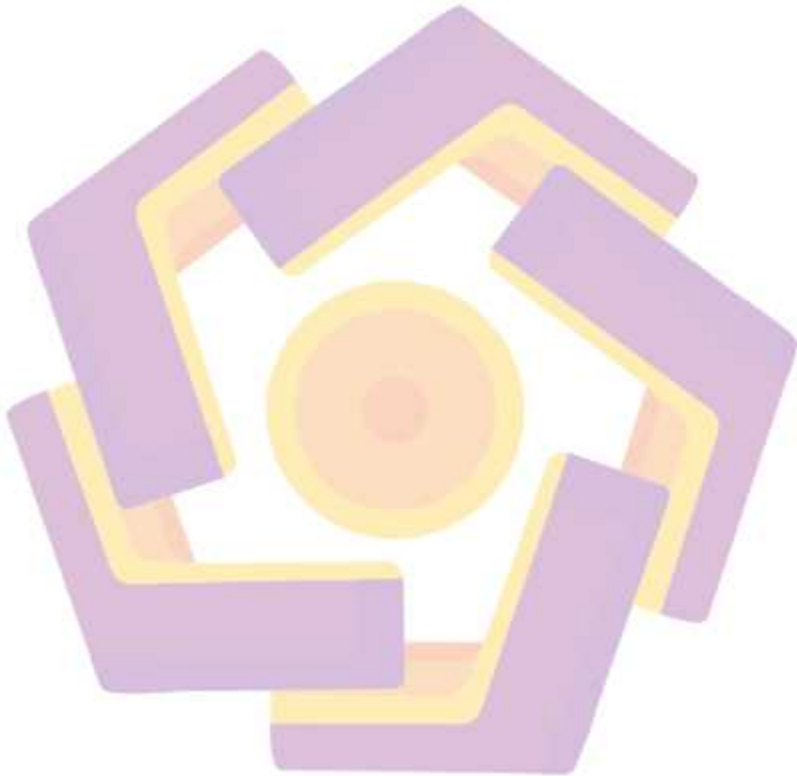
AI	Artificial Intelligence
AMH	Anti-Müllerian Hormone
ANN	Artificial Neural Network
AUC	Area Under the Curve
BMI	Body Mass Index
CDSS	Clinical Decision Support System
CV	Cross Validation
DHEAS	Dehydroepiandrosterone Sulfate
EDA	Exploratory Data Analysis
EHR	Electronic Health Records
ECE	Expected Calibration Error
FN	False Negative
FP	False Positive
FPR	False Positive Rate
FSH	Follicle-Stimulating Hormone
GBM	Gradient Boosting Machine
ID3	Iterative Dichotomizer 3
KNN	K-Nearest Neighbors
LH	Luteinizing Hormone
LR	Logistic Regression
ML	Machine Learning
PCOM	Polycystic Ovarian Morphology
PCOS	Polycystic Ovary Syndrome
PFI	Permutation Feature Importance
RF	Random Forest
RFE	Recursive Feature Importance
ROC	Receiver Operating Characteristic
SHAP	SHapley Additive exPlanations
SKCV	Stratified K-Fold Cross Validation
SLR	Systematic Literature Review
SMOTE	Synthetic Minority Over-Sampling Technique
SVM	Support Vector Machine

TN	True Negative
TP	True Positive
USG	Ultrasonografi
XAI	Explainable Artificial Intelligence
XGBOOST	Extreme Gradient Boosting
K	Jumlah Lipatan pada K-Fold Cross Validation
μ	Nilai rata-rata suatu fitur
N	Jumlah total oversampling yang diterapkan SMOTE
PI _j	Nilai Permutation Importance untuk fitur ke-j
r	koefisien korelasi Pearson
S	Skor kinerja model pada evaluasi
σ	Standar deviasi suatu fitur
z	Nilai Z-Score hasil standardisasi

DAFTAR ISTILAH

Anovulasi	Kegagalan ovarium melepas sel telur secara rutin pada siklus menstruasi.
Bagging	Teknik ensemble yang melatih beberapa model secara paralel lalu menggabungkan hasilnya.
Boosting	Teknik ensemble yang membangun model secara berurutan untuk memperbaiki kesalahan sebelumnya.
Data Leakage	Kebocoran informasi data uji ke proses pelatihan yang menyebabkan evaluasi terlalu optimis.
Endokrinopati	Gangguan pada sistem endokrin atau kelenjar penghasil hormon dalam tubuh.
Hirsutisme	pertumbuhan rambut berlebih pada wanita dengan pola menyerupai pria akibat kelebihan androgen.
Hyperparameter	Parameter konfigurasi model yang ditetapkan sebelum proses pelatihan dimulai.
Imbalanced Data	Dataset dengan distribusi kelas yang tidak seimbang sehingga model cenderung bias.
Imputasi	Pengisian nilai yang hilang dalam dataset menggunakan nilai statistik tertentu.
Interpolasi	Pembentukan data baru di antara titik data yang sudah ada.
Overfitting	Kondisi model yang terlalu menyesuaikan data latih sehingga gagal pada data baru.
Pipeline	Alur kerja terintegrasi yang menggabungkan tahapan pemrosesan dan pelatihan secara otomatis.
Preprocessing	Tahap pembersihan dan persiapan data mentah sebelum digunakan untuk pelatihan model.
Proof of Concept	Prototipe awal untuk membuktikan kelayakan teknis suatu ide sebelum dikembangkan lebih lanjut.
Regularisasi	Teknik pencegahan overfitting dengan menghukum kompleksitas model yang berlebihan.
Serialisasi	Penyimpanan objek model terlatih ke dalam format berkas agar dapat digunakan kembali.

Skrining	Pemeriksaan awal untuk mengidentifikasi individu berisiko sebelum diagnosis definitif ditegakkan.
Standardisasi	Transformasi fitur agar memiliki rata-rata nol dan standar deviasi satu untuk menyetarakan skala.



INTISARI

Polycystic Ovary Syndrome (PCOS) merupakan gangguan endokrin yang mempengaruhi 8-13% perempuan usia reproduktif secara global. Diagnosis konvensional berbasis kriteria Rotterdam bersifat kompleks dan memakan waktu, sehingga mendorong kebutuhan terhadap sistem skrining berbasis komputasional yang lebih cepat dan objektif. Penelitian ini melakukan analisis komparatif terhadap tiga algoritma machine learning, yaitu Random Forest, Support Vector Machine, dan XGBoost untuk prediksi dini PCOS. metodologi yang diterapkan mencakup imputasi nilai hilang, standarisasi Z-Score, serta penerapan *Synthetic Minority Over-Sampling Technique* (SMOTE) secara ekslusif di dalam setiap lipatan *Stratified 5-Fold Cross-Validation* melalui *pipeline* terintegrasi untuk mencegah kebocoran data. Penentuan model optimal dilakukan menggunakan hierarki kriteria tiga tahap dengan Recall $\geq 0,80$ sebagai ambang batas klinis utama, diikuti F1-Score dan AUC-ROC sebagai penentu akhir. Hasil evaluasi menunjukkan bahwa Random Forest merupakan model paling optimal dengan nilai Recall 0,8078, F1-Score 0,8454, dan AUC-ROC 0,9576, sedangkan SVM tidak memenuhi ambang batas minimum dengan Recall 0,7913. Evaluasi pada data uji independent mengonfirmasi konsistensi Random Forest dengan Recall 0,8333, dan AUC-ROC 0,9401 tanpa indikasi *overfitting*. Analisis *Permutation Feature Importance* mengidentifikasi *Follicle No.* (R) sebagai fitur paling berpengaruh, konsisten dengan kriteria diagnostic Rotterdam. Model terpilih kemudian diimplementasikan ke dalam prototipe system skrining berbasis web menggunakan Streamlit yang mampu menghasilkan prediksi risiko PCOS secara *real-time* berdasarkan 10 parameter klinis utama.

Kata kunci: Polycystic Ovary Syndrome, Machine Learning, Random Forest, XGBoost, Support Vector Machine

ABSTRACT

Polycystic Ovary Syndrome (PCOS) is an endocrine disorder that affects 8–13% of women of reproductive age worldwide. Conventional diagnosis based on the Rotterdam criteria is complex and time-consuming, driving the need for a faster and more objective computational screening system. This study conducted a comparative analysis of three machine learning algorithms: Random Forest, Support Vector Machine, and XGBoost for the early prediction of PCOS. The methodology applied includes missing value imputation, Z-Score standardization, and the exclusive application of the Synthetic Minority Over-Sampling Technique (SMOTE) within each fold of the Stratified 5-Fold Cross-Validation via an integrated pipeline to prevent data leakage. The optimal model was determined using a three-stage hierarchical criteria approach, with a Recall of ≥ 0.80 as the primary clinical threshold, followed by the F1-Score and AUC-ROC as final determinants. The evaluation results showed that Random Forest was the most optimal model with a Recall of 0.8078, an F1-Score of 0.8454, and an AUC-ROC of 0.9576, whereas SVM did not meet the minimum threshold with a Recall of 0.7913. Evaluation on independent test data confirmed the consistency of Random Forest with a Recall of 0.8611 and an AUC-ROC of 0.9427 without any indication of overfitting. Permutation Feature Importance analysis identified Follicle No. (R) as the most influential feature, consistent with the Rotterdam diagnostic criteria. The selected model was then implemented into a web-based screening system prototype using Streamlit, capable of generating real-time PCOS risk predictions based on 10 key clinical parameters.

Keyword: Polycystic Ovary Syndrome, Machine Learning, Random Forest, XGBoost, Support Vector Machine