



**PERBANDINGAN PERFORMA ALGORITMA *RANDOM FOREST*
DAN *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI
STATUS MUTU AIR SUNGAI**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
MEYRLIN ALINKA VIANDARI
NIM. 2111017120006**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
AGUSTUS 2026**



**PERBANDINGAN PERFORMA ALGORITMA *RANDOM FOREST*
DAN *SUPPORT VECTOR MACHINE* (SVM) DALAM KLASIFIKASI
STATUS MUTU AIR SUNGAI**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
MEYRLIN ALINKA VIANDARI
NIM. 2111017120006**

**PROGRAM STUDI S-1 STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
AGUSTUS 2026**

SKRIPSI

PERBANDINGAN PERFORMA ALGORITMA *RANDOM FOREST* DAN *SUPPORT VECTOR MACHINE (SVM)* DALAM KLASIFIKASI STATUS MUTU AIR SUNGAI

Oleh:
Meyrlin Alinka Viandari
2111017120006

Telah dipertahankan pada hari Jumat, tanggal 21-08-2026 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



Dewi Anggraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

Penguji I



Selvi Annisa, S.Si., M.Si
NIP. 199212262022032016

Pembimbing II



Yuana Sukmawaty, S.Si., M.Si
NIP. 198810152015042002

Penguji II



Dr. Ir. Rony Riduan, S.T., M.T
NIP. 197610171999031003

Banjarbaru, 9 September 2026

Mengetahui,

Ketua Jurusan/Koordinator PS Statistika
FMIPA ULM



Dewi Anggraini, S.Si., M.App.Sci., Ph.D
NIP. 198303282005012001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 15 September 2025



Meyrlin Alinka Viandari
NIM. 2111017120006

ABSTRAK

Perbandingan Performa Algoritma *Random Forest* dan *Support Vector Machine* (SVM) dalam Klasifikasi Status Mutu Air Sungai (Oleh: Meyrlin Alinka Viandari; Pembimbing: Dewi Anggraini dan Yuana Sukmawaty, 2026; 44 halaman)

Air bersih berperan penting dalam menjaga kualitas lingkungan hidup dan keberlanjutan pembangunan, sebagaimana tercantum dalam target 6 *Sustainable Development Goals* (SDGs) 2030. Kondisi mutu air sungai di Daerah Khusus Jakarta sebagai wilayah perkotaan dengan nilai Indeks Kualitas Air (IKA) terendah kedua di Indonesia menuntut adanya upaya klasifikasi status mutu air guna mendukung pengambilan kebijakan pengelolaan lingkungan. Perhitungan klasifikasi status mutu air saat ini masih dilakukan secara manual sehingga perlu dilakukan upaya perhitungan yang lebih efisien salah satunya menggunakan pendekatan komputasi melalui *machine learning*. Penelitian ini menggunakan metode *machine learning* sebagai pendekatan untuk meningkatkan efisiensi dalam klasifikasi status mutu air yang sebelumnya dilakukan secara manual. Penelitian ini bertujuan untuk membandingkan performa algoritma *Random Forest* dan *Support Vector Machine* (SVM) dalam klasifikasi Status Mutu Air Sungai di Daerah Khusus Jakarta. Status mutu air dapat diidentifikasi dengan 8 parameter, yaitu derajat keasaman (pH), oksigen terlarut (DO), kebutuhan oksigen biologi (BOD), kebutuhan oksigen kimiawi (COD), padatan tersuspensi total (TSS), nitrat ($\text{NO}_3\text{-N}$), total fosfat (Total-P), dan *fecal coliform* (*Fecal Coli*). Penelitian ini melibatkan analisis deskriptif untuk memahami pola distribusi kualitas air, penanganan ketidakseimbangan data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE), implementasi algoritma *Random Forest* dan *Support Vector Machine* (SVM), serta evaluasi performa model berdasarkan *recall*, *precision*, dan *F1-score*. Hasil analisis deskriptif menunjukkan bahwa status mutu air di Daerah Khusus Jakarta sebagian besar berada pada status cemar berat, yang memiliki proporsi tertinggi dibandingkan status cemar sedang dan cemar ringan. Algoritma *Random Forest* menunjukkan performa terbaik dengan *recall* sebesar 0.9583, *precision* sebesar 0.967, dan *F1-score* sebesar 0.96, dibandingkan *Support Vector Machine* (SVM) dengan *recall* sebesar 0.7291, *precision* sebesar 0.7450, dan *F1-score* sebesar 0.7352.

Kata Kunci: *Random Forest*, *Support Vector Machine*, Status Mutu Air, Daerah Khusus Jakarta

ABSTRACT

Performance Comparison of Random Forest and Support Vector Machine (SVM) Algorithms in River Water Quality Status Classification

(By: Meyrlin Alinka Viandari; Advisors: Dewi Anggraini and Yuana Sukmawaty, 2026; 44 pages)

Clean water plays a crucial role in maintaining environmental quality and supporting sustainable development, as outlined in Target 6 of the Sustainable Development Goals (SDGs) 2030. The condition of river water quality in the Special Region of Jakarta, an urban area with the second-lowest Water Quality Index (WQI) in Indonesia, highlights the need for water quality status classification to support environmental management policymaking. The classification of water quality status is currently still performed manually; therefore, a more efficient computational approach is needed, one of which is through the use of machine learning. This study aims to compare the performance of the Random Forest and Support Vector Machine (SVM) algorithms in classifying river water quality status in the Special Capital Region of Jakarta. Water quality status is identified using eight parameters: potential of hydrogen (pH), dissolved oxygen (DO), biochemical oxygen demand (BOD), chemical oxygen demand (COD), total suspended solids (TSS), nitrate ($\text{NO}_3\text{-N}$), total phosphate (Total-P), and fecal coliform (Fecal Coli). The study involves descriptive analysis to examine water quality distribution patterns, handling data imbalance using the Synthetic Minority Oversampling Technique (SMOTE), implementation of Random Forest and Support Vector Machine (SVM) algorithms, and evaluation of model performance based on recall, precision, and F1-score. The descriptive analysis results indicate that river water quality in the Special Region of Jakarta is predominantly classified as heavily polluted, with the highest proportion compared to moderately and lightly polluted statuses. The Random Forest algorithm demonstrates superior performance, achieving a recall of 0.9583, precision of 0.967, and F1-score of 0.96, outperforming the Support Vector Machine (SVM), which attains a recall of 0.7291, precision of 0.7450, and F1-score of 0.7352.

Keywords: Random Forest, Support Vector Machine, Water Quality Status, Special Region of Jakarta.

PRAKATA

Puji syukur kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah dan inayah-Nya sehingga penulis dapat menyelesaikan penulisan Tugas Akhir dengan judul “Perbandingan Performa Algoritma *Random Forest* dan *Support Vector Machine* (SVM) Dalam Klasifikasi Status Mutu Air Sungai”. Penyusunan Tugas Akhir ini bertujuan untuk memenuhi persyaratan dalam rangka menyelesaikan program sarjana Strata-1 di program studi Statistika FMIPA ULM.

Selama proses penyelesaian Tugas Akhir ini, penulis mendapat berbagai bantuan dan dukungan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang tulus kepada:

1. Orang tua dan keluarga atas segala doa, perhatian, dan memfasilitasi segala keperluan penulis selama penyelesaian Tugas Akhir.
2. Ibu Dewi Anggraini, S.Si., M.App.Sci., Ph.D selaku Koordinator Program Studi Statistika FMIPA ULM sekaligus dosen pembimbing yang telah memberikan ilmu, motivasi, dan saran selama melaksanakan perkuliahan sekaligus dalam penyusunan Tugas Akhir.
3. Ibu Yuana Sukmawaty, S.Si., M.Si selaku dosen pembimbing yang bersedia meluangkan waktu dan memberikan arahan selama penyusunan Tugas Akhir.
4. Ibu Selvi Annisa, S.Si., M.Si dan Dr. Ir. Rony Riduan, S.T., M.T selaku dosen penguji yang telah memberikan masukan dan saran dalam perbaikan dalam Tugas Akhir.
5. Seluruh dosen pengajar dan staf Program Studi Statistika FMIPA ULM yang telah memberikan ilmu, nasihat, dan bantuan selama melaksanakan perkuliahan.
6. Kepada teman-teman penulis yang menemani dan senantiasa mendukung penulis selama masa perkuliahan.

Penulis menyadari penulisan Tugas Akhir ini tidak luput dari kekurangan, sehingga penulis menghargai segala kritik dan saran yang membangun penyempurnaan Tugas Akhir ini. Penulis berharap Tugas Akhir ini dapat memberikan manfaat bagi pihak yang membutuhkan.

Banjarbaru, 21 Agustus 2026

Meyrlin Alinka Viandari

DAFTAR ISI

HALAMAN JUDUL.....	ii
ABSTRAK.....	iii
ABSTRACT	vi
PRAKATA	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN.....	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Penelitian Terdahulu	5
2.2 Kajian Teori	6
2.2.1 Kualitas Air	6
2.2.2 Statistika Deskriptif	7
2.2.3 SMOTE.....	8
2.2.4 Klasifikasi.....	9
2.2.5 Random Forest	9
2.2.6 Support Vector Machine (SVM)	11
2.2.7 Confusion matrix	15
BAB III METODE PENELITIAN.....	17
3.1 Sumber Data.....	17
3.2 Variabel Penelitian	17
3.3 Prosedur Penelitian	18
BAB IV HASIL DAN PEMBAHASAN	23
4.1 Data Preprocessing	23
4.2 Analisis Deskriptif	23
4.3 Pembagian Data Training dan Data Testing.....	28
4.4 Standarisasi Data	28
4.5 Penanganan Data Tidak Seimbang (Resampling Data).....	28
4.6 Analisis Klasifikasi Menggunakan Random Forest.....	31
4.6.1. Pelatihan Algoritma Random Forest	31
4.6.2. Tuning Hyperparameter Random Forest	34
4.6.3. Pengujian Kinerja Model Random Forest.....	35

4.7 Analisis Klasifikasi Menggunakan <i>Support Vector Machine</i>	36
4.7.1. Pelatihan Model <i>Support Vector Machine</i>	36
4.7.2. <i>Tuning Hyperparameter Support Vector Machine</i>	38
4.7.3. Pengujian Kinerja Model <i>Support Vector Machine</i>	39
4.8 Evaluasi Kinerja Model	40
4.9 Perbandingan Kinerja Model	41
BAB V PENUTUP	43
5.1 Kesimpulan	43
5.2 Saran	43
DAFTAR PUSTAKA	45
LAMPIRAN	48
RIWAYAT HIDUP	73

DAFTAR GAMBAR

Gambar 1.1 Capaian Nasional 4 Komponen IKLH Tahun 2023	1
Gambar 1.2 Indeks Kualitas Air (IKA) Provinsi Tahun 2023	2
Gambar 2.1 Alur Kerja <i>Random Forest</i>	10
Gambar 2.2 Cara Kerja Algoritma SVM.....	12
Gambar 3.1 Prosedur Penelitian.....	22
Gambar 4.1 Persentase Kelas Sesudah SMOTE	31
Gambar 4.2 Ilustrasi <i>bootstrap sampling</i>	31

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	5
Tabel 2.2 Kategori Status Mutu Air	7
Tabel 2.3 <i>Confusion matrix Multiclass</i> 3 Kategori	15
Tabel 3.1 Variabel Penelitian.....	17
Tabel 4.1 Identifikasi <i>Missing Value</i>	23
Tabel 4.2 Proporsi Kelas Data Variabel Target.....	24
Tabel 4.3 Statistika Deskriptif Data Kualitas Air.....	24
Tabel 4.4 Rasio Pembagian Data.....	28
Tabel 4.5 Contoh Data Kelas Minoritas	29
Tabel 4.6 Data Sintesis.....	30
Tabel 4.7 Data Sebelum dan setelah SMOTE.....	30
Tabel 4.8 <i>Hyperparameter Random Forest</i>	34
Tabel 4.9 Hasil Kombinasi <i>Hyperparameter Random Forest</i>	34
Tabel 4.10 <i>Confusion matrix Random Forest</i>	35
Tabel 4.11 Contoh Data <i>Training</i>	37
Tabel 4.12 <i>Hyperparameter Support Vector Machine (SVM)</i>	38
Tabel 4.13 Hasil Kombinasi <i>Hyperparameter SVM</i>	38
Tabel 4.14 <i>Confusion matrix Support Vector Machine</i>	39
Tabel 4.15 Evaluasi Kinerja Model	41

DAFTAR LAMPIRAN

Lampiran 1. Data Penelitian	48
Lampiran 2. <i>Tuning Hyperparameter Random Forest</i>	49
Lampiran 3. <i>Tuning Hyperparameter Support Vector Machine</i>	52
Lampiran 4. <i>Confusion Matrix Random Forest</i>	60
Lampiran 5. <i>Confusion Matrix Support Vector Machine</i>	63
Lampiran 6. <i>Syntax Phyton Random Forest</i>	66

DAFTAR LAMBANG DAN SINGKATAN

x_{syn}	: Data sintesis
$\hat{x}_t$	: <i>k-nearest neighbors</i> untuk x_i
δ	: Angka acak antara 0 sampai 1
S	: Himpunan data
m	: Banyaknya kelas
n	: Banyaknya data
p_i	: Proporsi dari kelas ke- i dalam himpunan data
A	: Atribut
S_i	: Subset ke- i dari S
$ S_i $	: Jumlah data pada subset ke- i
$ S $	: Jumlah total himpunan data
w	: Vektor bobot
b	: Nilai bias
x_i	: Data ke- i , $i = 1, 2, \dots, n$
N_{sv}	: Jumlah <i>support vector</i>
$f(x)$	: Fungsi klasifikasi untuk data x
α_i	: <i>Lagrange multiplier</i> untuk data ke- i
y_i	: Label data <i>training</i> ke- i
$K(x_i, x)$	: Fungsi kernel untuk data <i>training</i> x_i dan x
TP_k	: <i>True Positive</i> untuk kategori k
FN_{ik}	: <i>False Positive</i>
TP_k	: <i>True Positive</i> untuk kategori k
IKLH	: Indeks Kualitas Lingkungan Hidup
IKA	: Indeks Kualitas Air
IKU	: Indeks Kualitas Udara
IKL	: Indeks Kualitas Lahan
IKAL	: Indeks Kualitas Air Laut