

**ANALISIS KUALITAS AIR DAN KELAYAKAN KONSUMSI
MENGUNAKAN SUPPORT VECTOR MACHINE (SVM) DAN
CATBOOSTING DENGAN HYPERPARAMETER TUNING**

SKRIPSI

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



disusun oleh

CHRISTA PUTRI RAHAYU

21.11.3835

kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**ANALISIS KUALITAS AIR DAN KELAYAKAN KONSUMSI
MENGUNAKAN SUPPORT VECTOR MACHINE (SVM) DAN
CATBOOSTING DENGAN HYPERPARAMETER TUNING**

**HALAMAN JUDUL
SKRIPSI**

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



disusun oleh

CHRISTA PUTRI RAHAYU

21.11.3835

kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS KUALITAS AIR DAN KELAYAKAN KONSUMSI
MENGUNAKAN SUPPORT VECTOR MACHINE (SVM)
DAN CATBOOSTING DENGAN HYPERPARAMETER
TUNING**

yang disusun dan diajukan oleh

CHRISTA PUTRI RAHAYU

21.11.3835

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 14 Februari 2025

Kusnawi, S.Kom., M.Eng.


Kusnawi, S.Kom., M.Eng.
NIK. 190302112

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS KUALITAS AIR DAN KELAYAKAN KONSUMSI
MENGUNAKAN SUPPORT VECTOR MACHINE (SVM)
DAN CATBOOSTING DENGAN HYPERPARAMETER
TUNING**

yang disusun dan diajukan oleh

CHRISTA PUTRI RAHAYU

21.11.3835

Telah dipertahankan di depan Dewan Penguji
pada tanggal 14 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Ria Andriani, S.Kom., M.Kom.
NIK. 190302458

Nafiatun Sholihah, S.Kom., M.Cs
NIK. 190302524

Kusnawi, S.Kom., M.Eng.
NIK. 190302112



Skrripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 14 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : CHRISTA PUTRI RAHAYU
NIM : 21.11.3835

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS KUALITAS AIR DAN KELAYAKAN KONSUMSI MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM) DAN CATBOOSTING DENGAN HYPERPARAMETER TUNING

Dosen Pembimbing : Kusnawi, 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, 14 Februari 2025

Yang Menyatakan,



Christa Putri Rahayu

HALAMAN PERSEMBAHAN

Puji syukur kepada Tuhan Yang Maha Esa atas segala rahmat yang telah diberikan dan kesempatan untuk dapat menyelesaikan skripsi ini. Dalam menyusun laporan ini penulis banyak dibantu, dibimbing, dan didukung oleh berbagai pihak. Pada laporan ini penulis persembahkan kepada:

1. Kepada kedua orang tua, Bapak Setiyanto dan Ibu Sri Esa yang mendoakan, memberi dukungan, memberi semangat, memberi fasilitas serta memberikan hasil kerja kerasnya kepada saya untuk menimba ilmu.
2. Kepada keluarga terutama kedua kakak saya dan nenek saya yang mendoakan, memberi dukungan, memberi semangat.
3. Kepada teman-teman dan pacar saya yang telah memberikan saya dukungan dan semangat pada saat suka maupun duka selama masa perkuliahan. Terimakasih atas kenangan- kenangan yang telah kita ukir bersama-sama.
4. Dan semua pihak yang mendukung saya secara langsung ataupun tidak langsung.

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa atas karunia yang telah melimpahkan kasih dan sayang-Nya kepada kita semua, sehingga penulis dapat menyelesaikan skripsi yang diberi judul “Analisis Kualitas Air dan Kelayakan Konsumsi Menggunakan Support Vector Machine (SVM) dan Catboosting Dengan Hyperparameter Tuning.”

Tujuan dari penyusunan skripsi ini ialah untuk memenuhi syarat memperoleh gelar Sarjana pada program studi S1 Teknik Komputer, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta. Didalam pengerjaan skripsi ini telah melibatkan banyak pihak yang sangat membantu dalam banyak hal. Oleh sebab itu disini penulis sampaikan rasa terimakasih sedalam-dalamnya kepada:

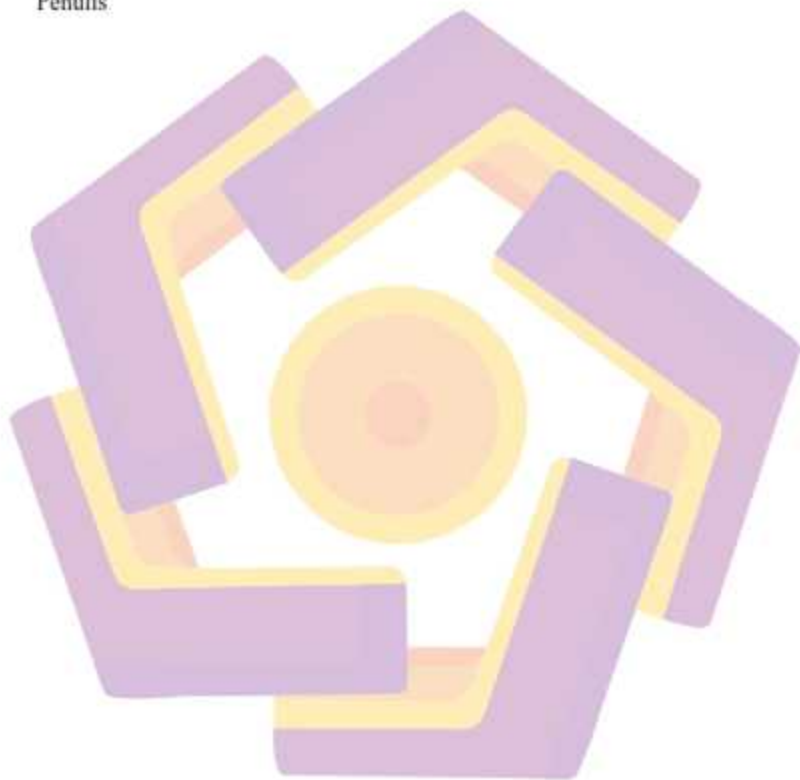
1. Tuhan Yang Maha Esa atas karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik dan semoga dapat memberikan manfaat dikemudian hari.
2. Bapak Prof. Dr .M. Suyanto, M.M. selaku Rektor Universitas Amikom Yogyakarta.
3. Bapak Kusnawi, S.Kom., M.Eng. selaku dosen pembimbing yang bersedia memberikan pengarahan dan bimbingan dalam penyusunan skripsi ini.
4. Kepada segenap Dosen, Staff, Karyawan Universitas Amikom Yogyakarta yang telah memberikan ilmu kepada penulis dibangku kuliah dan juga membantu penulis dalam kelancaran administrasi sampai terselesaikannya skripsi ini.
5. Orang tua, saudara-saudara beserta keluarga yang selalu mendoakan, memberi semangat dan dukungan penuh kepada penulis.
6. Serta kepada pihak yang telah membantu dalam penyusunan skripsi ini yang tidak dapat penulis sebutkan satu per satu.

Penulis berharap semoga skripsi ini dapat bermanfaat bagi semua pihak yang terkait dengan penulisan ini. Dalam penulisan skripsi ini, penulis menyadari masih banyak kekurangan karena perbatasnya pengetahuan dan pengalaman penulis. Oleh karena itu, dengan lapang hati penulis mengharapkan kritik dan saran yang membangun guna menyempurnakan skripsi ini. Akhirnya dengan kerendahan hati

penulis mengucapkan terima kasih dan semoga skripsi ini dapat bermanfaat bagi penulis maupun pembaca.

Yogyakarta, 20 Februari 2025

Penulis



DAFTAR ISI

HALAMAN JUDUL	i
DAFTAR ISI.....	v
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
INTISARI.....	xiii
<i>ABSTRACT</i>	xiv
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.....	7
2.1. Studi Literatur.....	7
2.2. Dasar Teori.....	14
2.2.1. Data Mining.....	14
2.2.2. Machine Learning.....	15
2.2.3. Kualitas Air.....	16
2.2.4. Kelayakan Konsumsi.....	16
2.2.5. Catboosting.....	17
2.2.6. Support Vector Machine (SVM)	17

2.2.7.	SMOTE (Synthetic Minority Over-sampling Technique).....	19
2.2.8.	<i>Hyperparameter Tuning</i>	19
2.2.9.	Teori Evaluasi.....	19
2.2.10.	Phyton.....	21
BAB III METODE PENELITIAN		22
3.1.	Objek Penelitian.....	22
3.2.	Alur Penelitian	22
3.3.	Alat dan Bahan.....	27
3.1.1.	Data Penelitian.....	27
3.1.2.	Alat/instrumen.....	27
BAB IV HASIL DAN PEMBAHASAN		29
4.1.	Pengambilan Data.....	29
4.2.	EDA	30
4.3.	Preprocessing.....	32
4.4.	Model.....	38
4.5.	Deployment	46
BAB V KESIMPULAN.....		48
5.1.	Kesimpulan	48
5.2.	Saran	48
REFERENSI		50
LAMPIRAN.....		54

DAFTAR TABEL

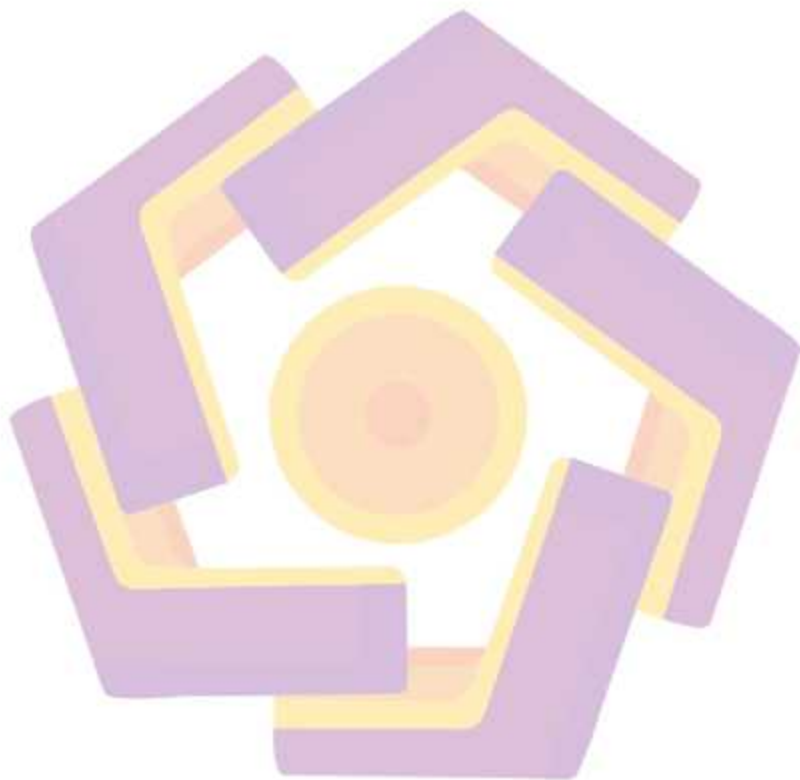
Tabel 2.1. Keaslian Penelitian	9
Tabel 2.2. Confusion Matrix.....	19
Tabel 3.1. Dataset.....	23
Tabel 3.2. Informasi Dataset.....	23
Tabel 3.3. Describe.....	24
Tabel 3.4. Missing Value.....	25
Tabel 3.5. Penghapusan Missing Value.....	25
Tabel 4.1. Dataset.....	29
Tabel 4.2. EDA.....	30
Tabel 4.3. Missing Value.....	32
Tabel 4.4. Matrix Korelasi.....	33
Tabel 4.5. SMOTE.....	35
Tabel 4.6. Split Data dan K-Fold Validasi.....	37
Tabel 4.7. SVM Model.....	38
Tabel 4.8. K-Fold SVM.....	38
Tabel 4.9. Hyperparameter Tuning SVM.....	39
Tabel 4.10. Model Catboosting.....	41
Tabel 4.11. K-Fold Catboosting.....	42
Tabel 4.12. Hyperparameter Tuning Catboosting.....	42
Tabel 4.13. Perbandingan Sebelum dan Sesudah Hyperparameter model SVM dan Catboost.....	45

DAFTAR GAMBAR

Gambar 3.1. Alur Penelitian.....	22
Gambar 3.2. Distribusi Potability.....	24
Gambar 4.1. Dataframe.....	29
Gambar 4.2. Informasi Dataframe.....	30
Gambar 4.3. Deskripsi Dataframe.....	31
Gambar 4.4. Distribusi Potability.....	31
Gambar 4.5 Penghapusan Missing Value.....	32
Gambar 4.6. Heatmap Correlation Matrix.....	34
Gambar 4.7. Correlation Matrix Numerik.....	34
Gambar 4.8. SMOTE.....	36
Gambar 4.9. Model SVM.....	38
Gambar 4.10. K-fold SVM.....	39
Gambar 4.11. Hyperparameter Tuning SVM.....	39
Gambar 4.12. Evaluasi SVM Sebelum Hyperparameter Tuning.....	40
Gambar 4.13. Evaluasi SVM Sesudah Hyperparameter Tuning.....	41
Gambar 4.14. Model Catboosting.....	41
Gambar 4.15. K-fold Catboosting.....	42
Gambar 4.16. Hyperparameter Tuning Catboosting.....	43
Gambar 4.17. Evaluasi Catboost Sebelum Hyperparameter Tuning.....	43
Gambar 4.18 Evaluasi Catboost Sesudah Hyperparameter Tuning.....	44
Gambar 4.19 Tampilan Website.....	46
Gambar 4.20 Tampilan Hasil.....	47

DAFTAR LAMPIRAN

Lampiran 1. Link Google Colab.....	54
Lampiran 2. Link Deploy.....	54



INTISARI

Analisis kualitas air memiliki peran penting dalam menentukan kelayakan konsumsi air bagi manusia. Penelitian ini bertujuan untuk membangun model machine learning yang mampu mengklasifikasikan kualitas air berdasarkan beberapa parameter seperti pH, kekerasan (*hardness*), kandungan zat padat (*solids*), *chloramines*, *sulfate*, *conductivity*, *organic carbon*, *trihalomethanes*, dan *turbidity*. Dataset yang digunakan berasal dari Kaggle dengan total 3.276 data sampel. Dua algoritma utama yang diterapkan dalam penelitian ini adalah *Support Vector Machine (SVM)* dan *CatBoost*. Proses penelitian mencakup tahap data *preprocessing*, penyeimbangan data menggunakan SMOTE, pemodelan, dan evaluasi performa model. *Hyperparameter tuning* diterapkan pada kedua algoritma untuk meningkatkan performa. Hasil penelitian menunjukkan bahwa *CatBoost* memiliki performa terbaik dengan akurasi sebesar 95,8% setelah *hyperparameter tuning*, dibandingkan SVM yang mencapai akurasi 77,9%. Selain itu, *CatBoost* unggul dalam semua metrik evaluasi, termasuk *precision*, *recall*, dan *F1-score*.

Kata kunci: SVM, Catboosting, Pembelajaran Mesin, Data Mining, Kualitas Air.

ABSTRACT

Water quality analysis plays an important role in determining the suitability of water for human consumption. This study aims to build a machine learning model that is able to classify water quality based on several parameters such as pH, hardness, solids content, chloramines, sulfate, conductivity, organic carbon, trihalomethanes, and turbidity. The dataset used comes from Kaggle with a total of 3,276 sample data. The two main algorithms applied in this study are Support Vector Machine (SVM) and CatBoost. The research process includes data preprocessing, data balancing using SMOTE, modeling, and model performance evaluation. Hyperparameter tuning is applied to both algorithms to improve performance. The results show that CatBoost has the best performance with an accuracy of 95.8% after hyperparameter tuning, compared to SVM which achieved an accuracy of 77.9%. In addition, CatBoost excels in all evaluation metrics, including precision, recall, and F1-score.

Keyword: SVM, Catboosting, Machine Learning, Data Mining, Water Quality

