

**KLASIFIKASI KUALITAS UDARA DKI JAKARTA
MENGUNAKAN ALGORITMA MACHINE LEARNING**

SKRIPSI

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



disusun oleh

MUHAMMAD IHYA' ULUMUDDIN

21.11.4415

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

1KLASIFIKASI KUALITAS UDARA DKI JAKARTA MENGUNAKAN ALGORITMA MACHINE LEARNING

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



disusun oleh

MUHAMMAD IHYA' ULUMUDDIN

21.11.4415

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI KUALITAS UDARA DKI JAKARTA MENGGUNAKAN
ALGORITMA MACHINE LEARNING**

yang disusun dan diajukan oleh

Muhammad Ihya' Ulumuddin

21.11.4415

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 4 Februari 2025

Dosen Pembimbing,


Robert Marco, M.T., Ph.D
NIK. 190302228

HALAMAN PENGESAHAN
SKRIPSI
KLASIFIKASI KUALITAS UDARA DKI JAKARTA MENGGUNAKAN
ALGORITMA MACHINE LEARNING

yang disusun dan diajukan oleh

Muhammad Ihya' Ulumuddin

21.11.4415

Telah dipertahankan di depan Dewan Penguji
pada tanggal 11 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya D, S.Kom., M.Kom.
NIK. 190302185



Ahmad Sa'di, S.Kom, M.Eng.
NIK. 190302459



Robert Marco, S.T., M.T., Ph.D.
NIK. 190302228



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 11 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 : Muhammad Ihya' Ulumuddin
NIM : 21.11.4415

Menyatakan bahwa Skripsi dengan judul berikut:

KLASIFIKASI KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA MACHINE LEARNING

Dosen Pembimbing : Robert Marco, M.T., Ph.D

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, 11 Februari 2025

Yang Menyatakan,



Muhammad Ihya' Ulumuddin

HALAMAN PERSEMBAHAN

Dengan penuh rasa Syukur dan penghormatan. Skripsi saya ini persembahkan kepada:

1. Kedua orang tua tercinta, yang tanpa Lelah memberikan dukungan moral, spiritual, serta cinta dan kasih sayang yang tiada henti. Kepada ayah dan ibu, terima kasih atas doa-doa yang selalu dipanjatkan, dorongan untuk terus berjuang, serta pengorbanan yang tak terhingga sepanjang perjalanan Pendidikan ini.
2. Dosen pembimbing, yang telah memberikan bimbingan. Ilmu, arahan yang sangat berharga selama penyusunan skripsi ini. Terima kasih atas kesabaran dan dedikasi dalam mendidik dan membimbing saya hingga tahap ini.
3. Bolo tampil dan bolo dolan, yang selalu memberikan semangat, dukungan, dan kebersamaan yang berarti. Terima kasih atas setiap momen kebersamaan, diskusi, dan tawa yang menjadi pelipur lara dalam menyelesaikan perjalanan akademik ini.
4. Komandan Rahman Hakim, yang selalu memberikan semangat dan dukungan moral dalam menyelesaikan skripsi ini.
5. Almamater Universitas Amikom Yogyakarta, yang telah menjadi tempat belajar dan berkembang, memberikan fasilitas dan kesempatan untuk mengejar mimpi serta mewujudkan cita-cita.

KATA PENGANTAR

Alhamdulillah Rabbil ‘Alamin. Puji Syukur penulis panjatkan ke hadirat Allah SWT atas segala Rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi yang berjudul “KLASIFIKASI KUALITAS UDARA DKI JAKARTA MENGGUNAKAN ALGORITMA MACHINE LEARNING” ini diajukan untuk memenuhi salah satu syarat dalam menyelesaikan studi S1 di Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dalam proses penyusunan skripsi ini, penulis banyak menerima bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. **Bapak Prof. M. Suyanto, M.M.**, selaku Rektor Universitas Amikom Yogyakarta, yang telah memberikan kesempatan kepada penulis untuk menuntut ilmu di universitas ini.
2. **Bapak Hanif Al Fatta, M.Kom., Ph.D.**, selaku Dekan Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta, atas segala dukungan dan fasilitas yang diberikan selama masa studi.
3. **Bapak Robert Marco, ST., MT., Ph.D** selaku Dosen Pembimbing, yang telah memberikan bimbingan, arahan, serta motivasi selama proses penyusunan skripsi ini.
4. **Kedua orang tua tercinta**, yang selalu memberikan dukungan moral, doa, dan kasih sayang sepanjang perjalanan pendidikan penulis.
5. **Komandan dan bolo seperjuangan**, yang telah memberikan semangat, bantuan, serta kebersamaan dalam menyelesaikan skripsi ini.

Yogyakarta, 4 Februari 2025

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
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	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	11

2.2.1 Big Data	11
2.2.2 Machine Learning	11
2.2.3 Jenis-jenis Machine Learning	11
2.2.4 Data Mining	12
2.2.5 Metode Data Mining	12
2.2.6 Klasifikasi	13
2.2.7 Random Forest	13
2.2.8 Support Vector Machine	15
2.2.9 K-Nearest Neighbor (KNN).....	17
2.2.10 Naïve Bayes	18
2.2.11 Synthetic Minority Oversampling Technique (SMOTE).....	18
2.2.12 Imputasi dengan mean	20
BAB III METODE PENELITIAN	21
3.1 Objek Penelitian.....	21
3.2 Alur Penelitian	22
3.2.1 Pengumpulan Data	23
3.2.2 Preprocessing Data.....	23
3.2.3 Splitting Data	24
3.3.3 Klasifikasi Model.....	24
3.3.4 Evaluasi Model	25
3.3 Alat dan Bahan.....	26
3.3.1 Deskripsi Dataset	26
3.3.2 Perangkat Lunak	26
3.3.3 Perangkat Keras	27
BAB IV HASIL DAN PEMBAHASAN	28

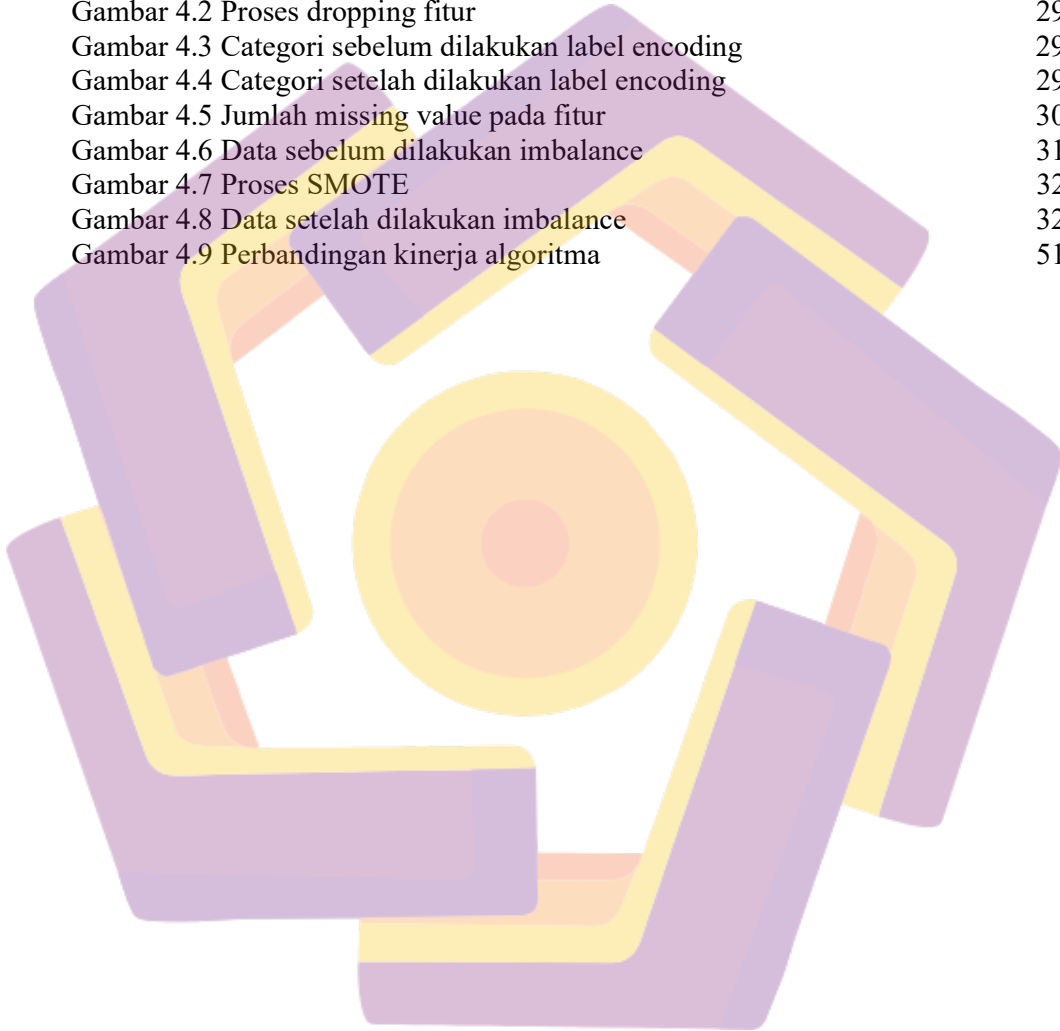
4.1	Dataset.....	28
4.2	Preprocessing Data.....	28
4.2.1	Dropping Fitur	28
4.3.1	Label Encoding	29
4.3.2	Handling Missing Value	29
4.3.3	Handling Imbalance Data.....	30
4.3	Splitting Data	33
4.4	Pemodelan Algoritma	35
4.4.1	Support Vector Machine	35
4.4.2	Random Forest	37
4.4.3	K-Nearest Neighbor	39
4.4.4	Naïve Bayes	41
4.5	Evaluasi Model	43
BAB V PENUTUP		54
5.1	Kesimpulan	54
5.2	Saran	54
REFERENSI		56
LAMPIRAN.....		61

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 3.1 Confusion Matrix	25
Tabel 3.2 Deskripsi Dataset	26
Tabel 3.3 Parameter algoritma	27
Tabel 4.1 Data X_train sebelum SMOTE	33
Tabel 4.2 Data y_train sebelum SMOTE	33
Tabel 4.3 Data X_test sebelum SMOTE	33
Tabel 4.4 Data y_test sebelum SMOTE	34
Tabel 4.5 Data X_test setelah SMOTE	34
Tabel 4.6 Data y_train setelah SMOTE	34
Tabel 4.7 Data X_test setelah SMOTE	34
Tabel 4.8 Data y_test setelah SMOTE	34
Tabel 4.9 Hasil SVM sebelum SMOTE	44
Tabel 4.10 Hasil SVM setelah SMOTE	44
Tabel 4.11 Hasil Random Forest sebelum SMOTE	46
Tabel 4.12 Hasil Random Forest setelah SMOTE	46
Tabel 4.13 Hasil K-NN sebelum SMOTE	48
Tabel 4.14 Hasil K-NN setelah SMOTE	48
Tabel 4.15 Hasil Naïve Bayes sebelum SMOTE	49
Tabel 4.16 Hasil Naïve Bayes setelah SMOTE	50

DAFTAR GAMBAR

Gambar 2.1 Random Forest	14
Gambar 2.2 Hyperplane SVM yang memisahkan kedua kelas bernilai 0 dan 1	16
Gambar 3.1 Alur Penelitian	22
Gambar 4.1 Dataset	28
Gambar 4.2 Proses dropping fitur	29
Gambar 4.3 Kategori sebelum dilakukan label encoding	29
Gambar 4.4 Kategori setelah dilakukan label encoding	29
Gambar 4.5 Jumlah missing value pada fitur	30
Gambar 4.6 Data sebelum dilakukan imbalance	31
Gambar 4.7 Proses SMOTE	32
Gambar 4.8 Data setelah dilakukan imbalance	32
Gambar 4.9 Perbandingan kinerja algoritma	51

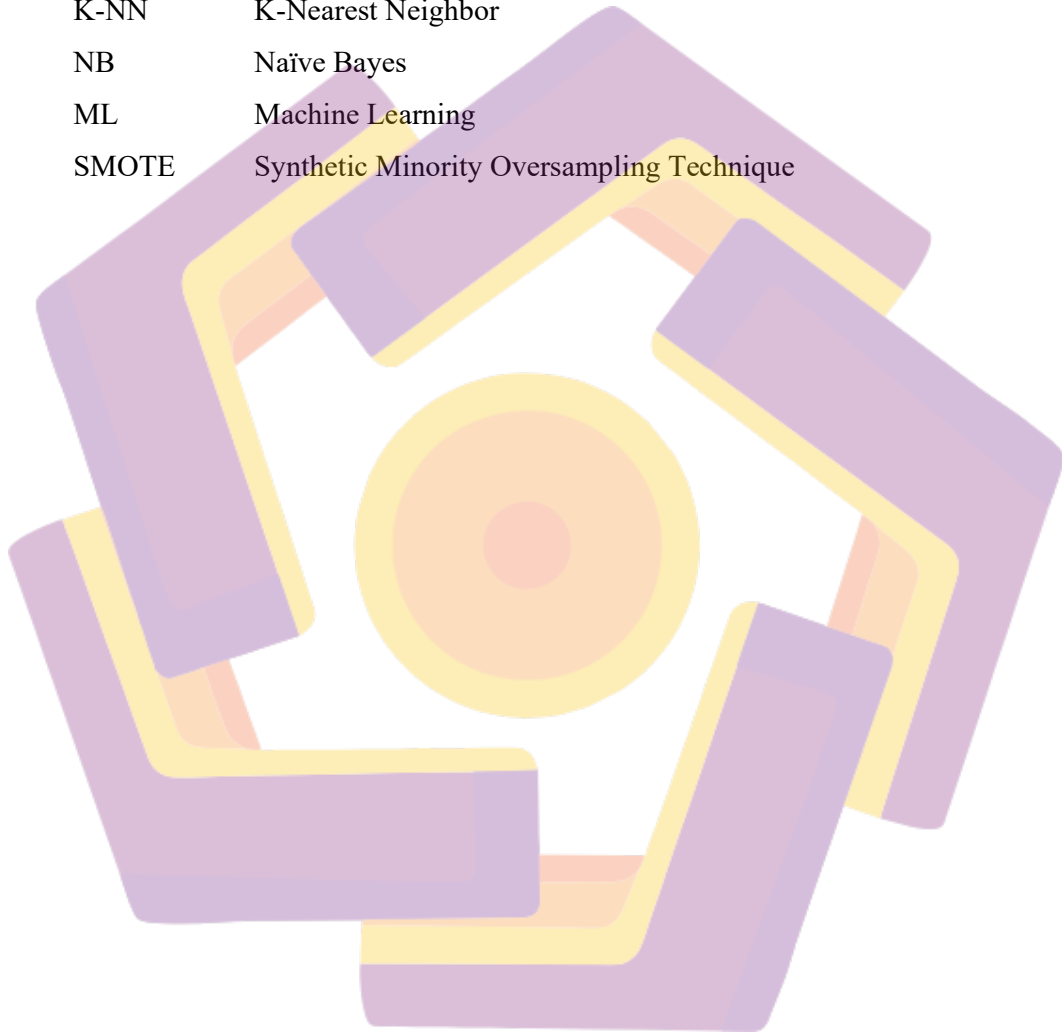


DAFTAR LAMPIRAN

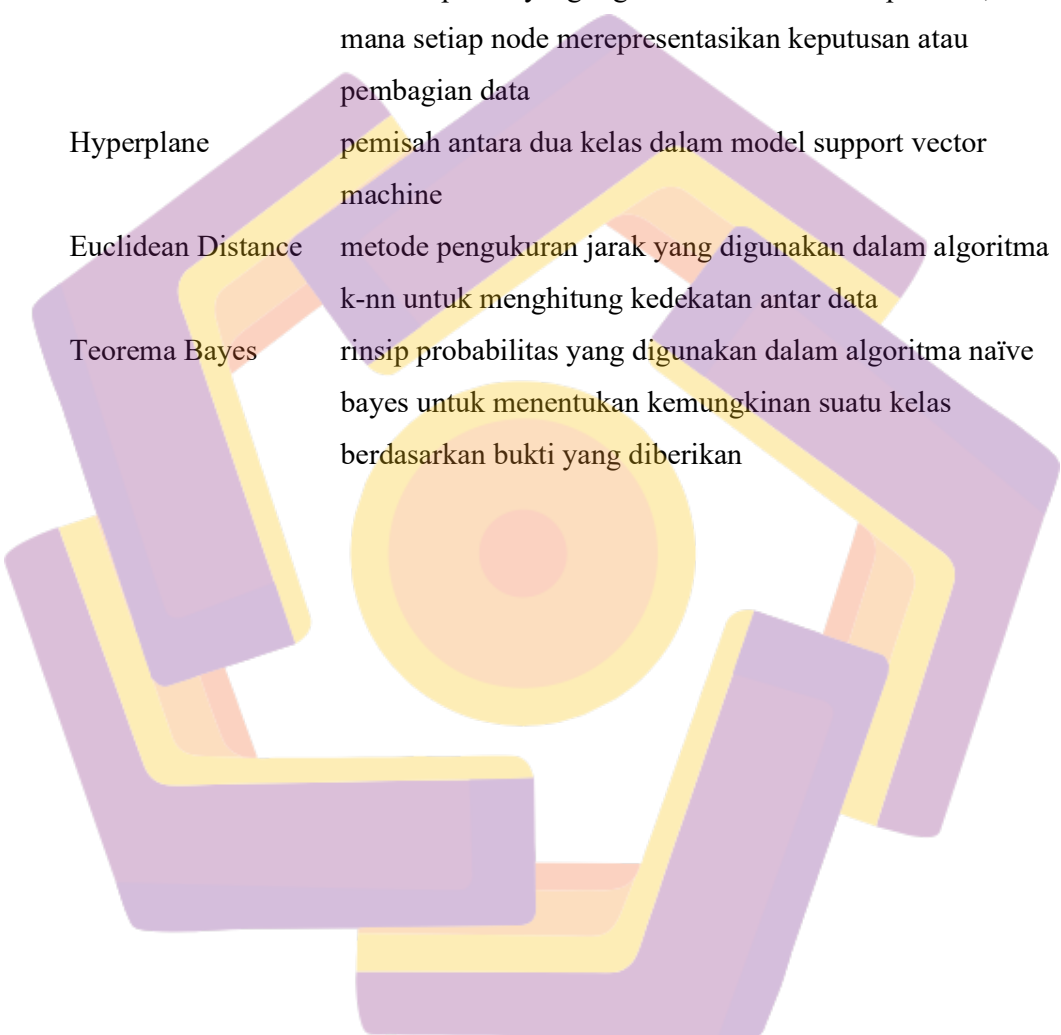
Lampiran 1. Import library	61
Lampiran 2. Membaca dataset	61
Lampiran 3. Melihat dimensi data	61
Lampiran 4. Menghitung jumlah tiap kategori	61
Lampiran 5. Mengkonversi label kategori menjadi angka	62
Lampiran 6. Melihat kategori yang unik setelah encoding	62
Lampiran 7. Visualisasi distribusi kategori dengan countplot	62
Lampiran 8. Memeriksa data kosong	62
Lampiran 9. Memeriksa data duplikat	62
Lampiran 10. Mengisi nilai kosong dengan nilai median	62
Lampiran 11. Menghapus fitur yang tidak digunakan	63
Lampiran 12. Menghapus baris yang memiliki kategori 3	63
Lampiran 13. Menghitung distribusi kelas target	63
Lampiran 14. Mengatasi ketidakseimbangan data dengan SMOTE	63
Lampiran 15. Membagi data menjadi data latih dan data uji	64
Lampiran 16. Proses penerapan model SVM untuk klasifikasi data	64
Lampiran 17. Proses penerapan model Random Forest untuk klasifikasi data	64
Lampiran 18. Proses penerapan model K-NN untuk klasifikasi data	64
Lampiran 19. Proses penerapan model Naïve Bayes untuk klasifikasi data	65
Lampiran 20. Confusion Matrix evaluasi model	65

DAFTAR SINGKATAN

AI	Artificial Intelligence
SVM	Support Vector Machine
RF	Random Forest
K-NN	K-Nearest Neighbor
NB	Naïve Bayes
ML	Machine Learning
SMOTE	Synthetic Minority Oversampling Technique



DAFTAR ISTILAH



Outlier	data yang memiliki nilai ekstrem yang berbeda jauh dari mayoritas data dalam dataset
Decision Tree	struktur pohon yang digunakan dalam model prediksi, di mana setiap node merepresentasikan keputusan atau pembagian data
Hyperplane	pemisah antara dua kelas dalam model support vector machine
Euclidean Distance	metode pengukuran jarak yang digunakan dalam algoritma k-nn untuk menghitung kedekatan antar data
Teorema Bayes	prinsip probabilitas yang digunakan dalam algoritma naïve bayes untuk menentukan kemungkinan suatu kelas berdasarkan bukti yang diberikan

INTISARI

Peningkatan aktivitas industri dan emisi kendaraan bermotor di Jakarta menyebabkan penurunan kualitas udara yang berdampak negatif terhadap kesehatan masyarakat dan lingkungan. Sayangnya, data kualitas udara yang tersedia secara daring sering kali mengandung ketidakseimbangan kelas dan nilai yang hilang. Penelitian ini bertujuan untuk mengklasifikasikan kualitas udara dengan mengatasi permasalahan tersebut menggunakan metode Synthetic Minority Oversampling Technique (SMOTE) untuk menangani ketidakseimbangan data dan metode imputasi median untuk mengisi nilai yang hilang. Empat algoritma machine learning dibandingkan dalam penelitian ini, yaitu Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), dan Naïve Bayes. Hasil evaluasi menunjukkan bahwa algoritma Random Forest memiliki performa terbaik dengan akurasi 99,07% setelah diterapkan SMOTE dan tuning hyperparameter, meningkat dari 98,33% sebelum optimasi. KNN mengalami peningkatan akurasi dari 92,27% menjadi 94,30%, sementara SVM meningkat dari 93,87% menjadi 94,00%. Naïve Bayes, meskipun memiliki akurasi terendah, juga mengalami peningkatan dari 85,83% menjadi 87,57%. Hasil penelitian ini menegaskan bahwa teknik penanganan ketidakseimbangan data dapat meningkatkan performa klasifikasi secara signifikan, terutama pada algoritma yang bergantung pada distribusi data seperti KNN. Penelitian ini diharapkan dapat menjadi referensi dalam pengelolaan kualitas udara dan mendukung kebijakan lingkungan berbasis data untuk mengurangi dampak pencemaran udara di Jakarta.

Kata kunci: kualitas udara, ketidakseimbangan data, data hilang, pembelajaran mesin, klasifikasi.

ABSTRACT

The increase in industrial activity and motor vehicle emissions in Jakarta has led to a decline in air quality, negatively impacting public health and the environment. Unfortunately, publicly available air quality data often contain class imbalances and missing values. This study aims to classify air quality while addressing these issues using the Synthetic Minority Oversampling Technique (SMOTE) to handle data imbalance and the median imputation method to fill in missing values. Four machine learning algorithms were compared in this study: Random Forest, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), and Naïve Bayes. Evaluation results show that the Random Forest algorithm achieved the best performance with an accuracy of 99.07% after applying SMOTE and hyperparameter tuning, an improvement from 98.33% before optimization. KNN's accuracy increased from 92.27% to 94.30%, while SVM improved from 93.87% to 94.00%. Naïve Bayes, although having the lowest accuracy, also improved from 85.83% to 87.57%. These findings confirm that handling data imbalance can significantly enhance classification performance, particularly for algorithms that rely on data distribution, such as KNN. This research is expected to serve as a reference for air quality management and support data-driven environmental policies to mitigate the effects of air pollution in Jakarta.

Keyword: air quality, data imbalance, missing data, machine learning, classification.