PENANGANAN MISSING VALUE PADA DATASET PROBABILITY WATER UNTUK KLASIFIKASI AIR MINUM MENGGUNKAN PENDEKATAN CENTER OF TEDENCY

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

DENI RAHMAN MASULILI

21.11.4377

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

PENANGANAN MISSING VALUE PADA DATASET PROBABILITY WATER UNTUK KLASIFIKASI AIR MINUM MENGGUNKAN PENDEKATAN CENTER OF TEDENCY

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

DENI RAHMAN MASULILI

21.11.4377

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2025

HALAMAN PERSETUJUAN

SKRIPSI

PENANGANAN MISSING VALUE PADA DATASET PROBABILITY WATER UNTUK KLASIFIKASI AIR MINUM MENGGUNKAN PENDEKATAN CENTER OF TENDENCY

yang disusun dan diajukan oleh

Deni Rahman Masulili

21.11.4377

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 24 Febuari 2025

Dosen Pembimbing,

Hastari Utama, S.kom., M.Cs
NIK. 190302230

HALAMAN PENGESAHAN

SKRIPSI

PENANGANAN MISSING VALUE PADA DATASET PROBABILITY WATER UNTUK KLASIFIKASI AIR MINUM MENGGUNKAN PENDEKATAN CENTER OF TENDENCY

yang disusun dan diajukan oleh

Deni Rahman Masulil

21.11.4377

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Febuari 2025

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Dr. Ferry Wahyu Wibowo, S.Si., M.Cs. **NIK. 190302235**	
Bayu Setiaji, M.Kom. **NIK. 190302216**	
Hastari Utama, S.Kom., M.Cs. **NIK. 190302230**	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 24 Febuari 2025

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Deni Rahman Masulili
NIM : 21.11.4377

Menyatakan bahwa Skripsi dengan judul berikut:

Penanganan Missing Value Pada Dataset Probability Water Untuk Klasifikasi Air Minum Menggunkan Pendekatan Center of Tedency

Dosen Pembimbing : Hastari Utama, S.Kom., M.Cs.

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, 24 Febuari 2025

Yang Menyatakan,

Deni Rahman Masulili

HALAMAN PERSEMBAHAN

Dengan rasa syukur yang mendalam kepada Allah SWT, karya sederhana ini penulis persembahkan kepada:

1. Allah SWT, Tuhan Yang Maha Esa, atas segala rahmat, hidayah, dan kemudahan yang telah diberikan selama proses penyusunan skripsi ini.
2. Kedua orang tuaku tercinta, yang selalu memberikan doa, kasih sayang, dukungan, dan semangat yang tiada henti. Terima kasih atas segala pengorbanan dan cinta yang tak ternilai.
3. Keluargaku, yang selalu menjadi tempat untuk kembali dan memberi dukungan moril di setiap langkah.
4. Dosen pembimbingku, Bapak Hastari Utama, S.Kom., M.Cs., yang telah membimbing dan memberikan arahan selama proses penyusunan skripsi ini.
5. Teman-teman seperjuangan di Informatika AMIKOM Yogyakarta, yang telah memberikan semangat, kebersamaan, dan motivasi selama masa studi.
6. Diriku sendiri, yang telah bertahan, belajar, dan tidak menyerah hingga titik ini.

Semoga skripsi ini dapat memberikan manfaat dan menjadi langkah awal untuk terus belajar dan berkembang.

KATA PENGANTAR

Puji syukur kehadirat Allah Subhanahu wa Ta'ala, karena atas rahmat dan hidayah-Nya, penulis dapat menyelesaikan skripsi ini dengan judul "Penanganan Missing Value pada Dataset Probability Water dengan Pendekatan Center of Tendency" sebagai salah satu syarat untuk memperoleh gelar Sarjana di Universitas Amikom Yogyakarta.

Dalam penyusunan skripsi ini, penulis menyadari bahwa keberhasilan ini tidak terlepas dari dukungan, bimbingan, dan doa dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Allah Subhanahu wa Ta'ala, atas segala nikmat, kesehatan, dan kesempatan yang diberikan sehingga skripsi ini dapat terselesaikan dengan baik.
2. Kedua orang tua tercinta, yang selalu memberikan dukungan, doa, dan kasih sayang tanpa henti dalam setiap langkah yang penulis jalani.
3. Hastari Utama, S.Kom., M.Cs, selaku dosen pembimbing yang telah memberikan arahan, saran, dan motivasi dalam proses penyelesaian penelitian ini.
4. Teman-teman diskusi di kontrakan Sugeng Geng, yang telah banyak membantu, mendukung, serta menemani dalam berbagai sesi diskusi dan brainstorming.
5. Elon Musk, sebagai inspirasi dalam eksplorasi teknologi dan kecerdasan buatan yang semakin berkembang.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan guna perbaikan di masa yang akan datang. Semoga penelitian ini dapat memberikan manfaat bagi pembaca dan menjadi referensi bagi penelitian selanjutnya.

Yogyakarta, 22 Febuari 2025
Deni Rahman Masulili

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMBANG DAN SINGKATAN

ML	Machine Learning
SVM	Support Vector Machine
TP	True Positive
FP	False Positive
TP	True Negative
FN	False Negative
CM	Confusion Matrix
CT	Center of Tendency
NB	Naïve Bayes
DT	Decision Tree
ACC	Accuracy
PREC	Precision
REC	Recall
F1	F1-Score

DAFTAR ISTILAH

Machine Learning	Pembelajaran mesin dalam kecerdasan buatan
Supervised Learning	Pembelajaran dengan data berlabel
Unsupervised Learning	Pembelajaran tanpa label atau target
Reinforcement Learning	Pembelajaran berbasis reward dan penalti
Hyperplane	Pemisah kelas dalam SVM
Confusion Matrix	Matriks evaluasi model klasifikasi
Missing Value	Data yang hilang dalam dataset
Center of Tendency	Mean, Median, dan Modus
Mean (Rata-rata)	Jumlah nilai dibagi total data
Median	Nilai tengah setelah data diurutkan
Modus	Nilai yang paling sering muncul
Support Vector Machine	Algoritma klasifikasi berbasis hyperplane
Naïve Bayes	Algoritma berbasis probabilitas Bayes
Decision Tree	Model berbasis pohon keputusan
Accuracy	Persentase prediksi yang benar
Precision	Proporsi prediksi positif yang benar
Recall	Kemampuan model mendeteksi kelas positif
F1-Score	Keseimbangan antara Precision dan Recall
Outlier	Nilai ekstrim dalam dataset
Skewed Data	Distribusi data tidak seimbang
Potable	Layak
Not potable	Tidak Layak

INTISARI

Kualitas air minum merupakan faktor krusial dalam menjaga kesehatan masyarakat. Namun, seringkali data kualitas air mengalami missing value, yang dapat mengurangi akurasi dalam analisis dan klasifikasi. Salah satu pendekatan untuk menangani missing value adalah metode Center of Tendency (mean, median, dan modus). Penelitian ini bertujuan untuk mengevaluasi efektivitas metode imputasi ini dalam meningkatkan akurasi klasifikasi potabilitas air. Data yang digunakan dalam penelitian ini mengandung berbagai parameter kualitas air, seperti pH, kandungan zat padat terlarut, tingkat kekeruhan, serta kandungan kimia lainnya.

Metode penelitian ini melibatkan penerapan algoritma Naïve Bayes, Decision Tree, dan SVM untuk mengklasifikasikan potabilitas air setelah dilakukan imputasi dengan metode Center of Tendency. Evaluasi dilakukan menggunakan confusion matrix, yang mengukur performa model berdasarkan akurasi, precision, recall, dan f1-score. Dataset dibagi menjadi data latih dan data uji untuk setiap metode imputasi yang digunakan, dan hasil klasifikasi dibandingkan untuk menilai dampak metode imputasi terhadap performa model.

Hasil penelitian menunjukkan bahwa metode Center of Tendency memberikan peningkatan akurasi dalam klasifikasi potabilitas air dibandingkan dengan dataset asli tanpa imputasi. Metode median cenderung menghasilkan performa terbaik dibandingkan mean dan modus dalam beberapa skenario. Temuan ini dapat dimanfaatkan oleh peneliti dan praktisi dalam bidang data mining dan pengolahan data lingkungan untuk meningkatkan keandalan analisis data dengan missing value. Penelitian lebih lanjut dapat dilakukan dengan menerapkan metode imputasi lain atau menggunakan dataset yang lebih luas untuk meningkatkan generalisasi hasil.

Kata kunci: missing value, Center of Tedency, klasifikasi, confusion matrix, potabilitas air.

ABSTRACT

Drinking water quality is a crucial factor in maintaining public health. However, water quality data often contain missing values, which can reduce the accuracy of analysis and classification. One approach to handling missing values is the Center of Tendency method (mean, median, and mode). This study aims to evaluate the effectiveness of this imputation method in improving the accuracy of potable water classification. The dataset used in this research includes various water quality parameters, such as pH, total dissolved solids, turbidity levels, and chemical content.

The research methodology involves implementing Naïve Bayes, Decision Tree, and SVM algorithms to classify water potability after imputing missing values using the Center of Tendency approach. The evaluation is conducted using a confusion matrix, which measures model performance based on accuracy, precision, recall, and f1-score. The dataset is divided into training and testing sets for each imputation method applied, and the classification results are compared to assess the impact of imputation on model performance.

The results indicate that the Center of Tendency method improves classification accuracy compared to the original dataset without imputation. The median method tends to produce the best performance compared to the mean and mode in several scenarios. These findings can be beneficial for researchers and practitioners in the field of data mining and environmental data processing to enhance the reliability of analyses involving missing values. Future research can explore other imputation techniques or utilize a broader dataset to improve the generalizability of the results.

Keyword**:** *missing value, Center of Tedency, classification, confusion matrix, water potability*