

**SKRIPSI**

**PENERAPAN MODEL KLASIFIKASI DAN VISUALISASI  
DATA AIR SUMUR DI DKI JAKARTA**



**RAMADHAN AMANNU**

**1519620035**

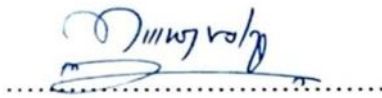
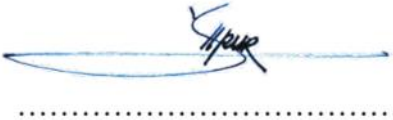
**PROGRAM STUDI  
SISTEM DAN TEKNOLOGI INFORMASI  
FAKULTAS TEKNIK  
UNIVERSITAS NEGERI JAKARTA**

**2025**

## LEMBAR PENGESAHAN SKRIPSI

Judul : Penerapan Model Klasifikasi dan Visualisasi Data Air  
Sumur di DKI Jakarta  
Penyusun : Ramadhan Amannu  
NIM : 1519620035

Disetujui oleh :

NAMA DOSEN	TANDA TANGAN	TANGGAL
Murien Nugraheni, S.T., M.Cs (Dosen Pembimbing I)		27 Desember 2024
Lipur Sugiyanta, Ph.D. (Dosen Pembimbing II)		7 Januari 2025

### PENGESAHAN PANITIA UJIAN SKRIPSI

NAMA DOSEN	TANDA TANGAN	TANGGAL
Fuad Mumtas, S.Kom, M.T.I (Dosen Penguji I)		31 Januari 2025
Irma Permata Sari, M.Eng (Dosen Penguji II)		25 Januari 2025



## LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini merupakan Karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 31 Januari 2025

Yang membuat pernyataan



Ramadhan Amannu

NIM. 1519620035



KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI  
UNIVERSITAS NEGERI JAKARTA  
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220  
Telepon/Faksimili: 021-4894221  
Laman: [lib.unj.ac.id](http://lib.unj.ac.id)

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI  
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Ramadhan Amannu  
NIM : 1519620035  
Fakultas/Prodi : Teknik / Sistem dan Teknologi Informasi  
Alamat email : ramadhanamannu103@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

Skripsi     Tesis     Disertasi     Lain-lain (.....)

yang berjudul :

Penerapan Model Klasifikasi dan Visualisasi Data Air Sumur di DKI Jakarta

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 10 Februari 2025

Penulis

( Ramadhan Amannu )  
nama dan tanda tangan

## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT atas segala limpahan rahmat dan karunia-Nya, sehingga dapat menyelesaikan penyusunan Skripsi dengan judul “**Penerapan Model Klasifikasi dan Visualisasi Data Air Sumur di DKI Jakarta**”, yang diajukan untuk memenuhi syarat kelulusan dalam meraih gelar Sarjana Ilmu Komputer di Program Studi Sistem dan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Jakarta.

Penulis menyadari bahwa proses penusunan dan penulisan skripsi ini mengalami berbagai kesulitan. Namun berkat ikhtiar, doa, bimbingan, saran dan dukungan dari berbagai pihak, penulis dapat menyelesaikan penulisan skripsi ini. Oleh karena itu, penulis ingin mengucapkan terima kasih sebesar-besarnya kepada:

1. Kedua orang tua yang penulis hormati dan sayangi, Bapak Kadari dan Ibu Siti Arifah yang senantiasa mendoakan yang terbaik, memberikan dukungan dan semangat baik secara moril dan materil kepada penulis.
2. Bapak Lipur Sugiyanta, Ph.D, sebagai Koordiantor Program Studi Sistem dan Teknologi Informasi, serta sebagai Dosen Pembimbing II yang telah memberikan kesediaan waktunya kapanpun dalam memberikan bimbingan dan arahan dalam penyusunan skripsi dengan sangat memuaskan.
3. Ibu Murien Nugraheni, S.T., M.Cs. sebagai Dosen Pembimbing I yang memberikan bimbingan, dukungan moril, serta ruang konsultasi baik dalam penyusunan skripsi, *project*, serta lingkup akademik.
4. Bapak Fuad Mumtas, S.Kom., M.TI. dan Irma Permatasari, M.Eng. selaku Dosen Penguji yang telah memberikan masukan dan saran dalam perbaikan penyusunan skripsi.
5. Lingkaran *Halaqah* Ust. Nico Kurniawan yang telah memberikan ruang untuk mengekspresikan kesulitan serta memberikan motivasi religi dalam mendukung penyusunan skripsi.
6. Teman-Teman baik di dalam ataupun di luar lingkungan UNJ yang memberikan ruang berinteraksi, bertukar pikiran, serta berekspresi dalam penyusunan skripsi ini.



7. Seluruh pihak yang terlibat secara langsung maupun tidak langsung yang tidak bisa disebutkan satu persatu.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, maka dari itu penulis menyampaikan permohonan maaf apabila terdapat kekurangan dan kesalahan dalam proses penulisan skripsi ini. Penulis terbuka terhadap kritik dan saran yang diberikan pembaca untuk menyempurnakan skripsi ini. Akhir kata, penulis berharap agar penelitian dan penyusunan skripsi ini dapat bermanfaat bagi pembaca dan khususnya bagi penulis.

Jakarta, 31 Januari 2025

Penyusun,



Ramadhan Amannu



## ABSTRAK

**Ramadhan Amannu**, Penerapan Model Klasifikasi dan Visualisasi Data Air Sumur di DKI Jakarta. Skripsi. Program Studi Sistem dan Teknologi Informasi , Fakultas Teknik Universitas Negeri Jakarta 2025. Dosen Pembimbing : Murien Nugraheni, S.T., M.Cs. dan Lipur Sugiyanta, Ph.D.

Air merupakan kebutuhan pokok bagi manusia. Salah satu sumber mata air yang digunakan adalah air sumur. Di DKI Jakarta, 35% penduduknya mengandalkan air sumur untuk pemenuhan air baku. Penelitian ini bertujuan untuk mengklasifikasikan kualitas air sumur di DKI Jakarta tahun 2023 menggunakan algoritma *K-Nearest Neighbors* (KNN). Penelitian ini dilakukan dengan metodologi *Knowledge Discovery in Database* (KDD), yang mencakup *Data Selection, Preprocessing, Transformation, Data Mining*, serta *Interpretation & Evaluation*. Proses seleksi fitur menggunakan metode analisis *Chi-Square* untuk data kategorikal dan *Spearman Rho* untuk data numerikal. Kemudian untuk memperbaiki keseimbangan data, digunakan SMOTE (*Synthetic Minority Over-sampling Technique*), sehingga data memiliki distribusi kelas data target yang seimbang dimana masing-masing berjumlah 248 baris. Data akan dibagi menjadi dua bagian dengan proporsi 80% (908 baris) untuk data latih dan 20% (228 baris) untuk data uji. Hasil dari pelatihan model dengan data latih dibantu dengan *5-Fold Cross Validation* menghasilkan nilai rata-rata akurasi sebesar 96%. Model yang telah dilatih, selanjutnya akan di evaluasi dengan data uji dan dihasilkan nilai akurasi 95%, presisi 95%, *recall* 95%, *f1-score* 95%, *balance accuracy* 96% serta AUC-ROC 95% yang dihitung secara *Macro average*. Hasil penelitian dari penelitian ini kemudian di interpretasikan dalam bentuk visualisasi data yang dapat diakses secara publik.

Kata Kunci: DKI Jakarta, *K-Nearest Neighbors*, *Knowledge Discovery in Database*, Kualitas Air Sumur, Visualisasi Data

## ABSTRACT

**Ramadhan Amannu**, *Application of Classification Model and Ground Water Data Visualization in DKI Jakarta. Thesis. Information System and Technology Study Program, Faculty of Engineering, State University of Jakarta 2025. Supervisor : Murien Nugraheni, S.T., M.Cs. and Lipur Sugiyanta, Ph.D.*

Water is essential for life, and in DKI Jakarta, 35% of the population depends on ground water. This study aimed to classify the quality of ground water in the area using the K-Nearest Neighbors (KNN) algorithm. The research followed the Knowledge Discovery in Database (KDD) methodology, which includes steps like Data Selection, Preprocessing, Transformation, Data Mining, and Interpretation & Evaluation. For feature selection, Chi-Square analysis was used for categorical data and Spearman Rho for numerical data. To improve data balance, SMOTE (Synthetic Minority Over-sampling Technique) was used, resulting in a balanced class distribution with 248 rows each. The dataset was split into 80% training data (908 rows) and 20% test data (228 rows). The model, trained with 5-Fold Cross Validation, achieved an average accuracy of 96%. When evaluated with test data, the model showed an accuracy of 95%, precision of 95%, recall of 96%, F1-score of 95%, balance accuracy of 96%, and an AUC-ROC of 95%, all calculated as macro averages. The findings were then presented through publicly accessible data visualizations.


**Keywords:** *DKI Jakarta, Ground Water Quality, K-Nearest Neighbors, Knowledge Discovery in Database, Data Visualization*



## DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI .....	ii
LEMBAR PERNYATAAN .....	iii
KATA PENGANTAR .....	iv
ABSTRAK .....	vi
ABSTRACT .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	xi
DAFTAR GAMBAR .....	xiii
DAFTAR LAMPIRAN .....	xvii
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang Masalah .....	1
1.2 Identifikasi Masalah .....	5
1.3 Pembatasan Masalah .....	5
1.4 Perumusan Masalah .....	5
1.5 Tujuan Penelitian .....	6
1.6 Manfaat Penelitian .....	6
BAB II KAJIAN PUSTAKA .....	7
2.1 Konsep dan Teori .....	7
2.1.1 Air Sumur dan Kualitas Air Bersih .....	7
2.1.2 KDD sebagai Metodologi Penelitian <i>Data Mining</i> .....	9
2.1.2.1 <i>Data Selection</i> .....	10
2.1.2.2 <i>Preprocessing</i> .....	10
2.1.2.3 <i>Transformation</i> .....	11
2.1.2.4 <i>Data Mining</i> .....	11
2.1.2.5 <i>Evaluation &amp; Interpretation</i> .....	11
2.1.3 <i>Unified Modeling Language (UML)</i> dalam Pembuatan Alur Penelitian .....	12
2.1.4 <i>Chi-Square</i> dan <i>Spearman Rho Test</i> dalam <i>Feature Selection</i> .....	13
2.1.4.1 <i>Chi-Square Test</i> .....	13
2.1.4.2 <i>Spearman Rho Test</i> .....	14
2.1.5 Model Klasifikasi dalam <i>Machine Learning</i> .....	16
2.1.6 Algoritma Model <i>K-Nearest Neighbors</i> .....	16

2.1.7	<i>K-Fold Cross Valiation</i> dalam Pelatihan Algoritma Model .....	18
2.1.8	<i>SMOTE</i> sebagai Metode Perbaikan <i>Imbalanced Data</i> .....	19
2.1.9	Evaluasi Metrik .....	20
2.1.10	Visualisasi Data.....	22
2.1.11	Jenis Visualisasi Data.....	23
2.1.12	Looker Studio.....	27
2.2	Penelitian Relevan .....	27
2.3	Kontribusi .....	30
2.4	Kerangka Berpikir.....	30
<b>BAB III METODOLOGI PENELITIAN</b> .....		34
3.1	Tempat dan Waktu Penelitian.....	34
3.2	Alat dan Bahan Penelitian.....	34
3.2.1	Alat Penelitian.....	34
3.2.2	Bahan Penelitian.....	35
3.3	Langkah-Langkah Penelitian .....	35
3.3.1	<i>Data Selection</i> (Pengumpulan Data).....	37
3.3.1.1	<i>Data Retrieve</i> .....	37
3.3.1.2	<i>Pivot Data</i> .....	38
3.3.2	<i>Preprocessing</i> .....	40
3.3.2.1	<i>Data Cleansing</i> .....	40
3.3.2.2	<i>Data Enrichment</i> .....	41
3.3.3	<i>Transformation</i> .....	41
3.3.3.1	<i>Ordinal Encoding</i> .....	41
3.3.3.2	<i>Feature Selection</i> .....	42
3.3.3.3	<i>Data Resampling</i> .....	43
3.3.3.4	<i>Normalization</i> .....	43
3.3.4	<i>Data Mining</i> .....	44
3.3.5	<i>Interpretation dan Evaluation</i> .....	47
3.3.5.1	<i>Evaluation Model</i> .....	47
3.3.5.2	<i>Interpretation Results</i> .....	48
3.4	Teknik Pengumpulan Data.....	49
3.5	Metode Evaluasi Data.....	49
<b>BAB IV HASIL PENELITIAN</b> .....		52
4.1	Hasil dan Analisis Penelitian .....	52



4.1.1	Hasil <i>Data Selection</i> .....	52
4.1.1.1	Hasil <i>Data Retrieve</i> .....	52
4.1.1.2	Hasil <i>Pivot Data</i> .....	56
4.1.2	Hasil <i>Preprocessing</i> .....	58
4.1.2.1	Hasil <i>Data Cleansing</i> .....	58
4.1.2.2	Hasil <i>Data Enrichment</i> .....	66
4.1.3	Hasil <i>Transformation</i> .....	69
4.1.3.1	Hasil <i>Ordinal Encoding</i> .....	69
4.1.3.2	Hasil <i>Feature Selection</i> .....	71
4.1.3.3	Hasil <i>Data Resampling</i> .....	77
4.1.3.4	Hasil <i>Normalization</i> .....	79
4.1.4	Hasil <i>Data Mining</i> .....	81
4.1.5	Hasil <i>Interpretation &amp; Evaluation</i> .....	86
4.1.5.1	<i>Evaluate Result</i> .....	86
4.1.6	Hasil <i>Interpretation Report</i> .....	93
4.2	Pembahasan Penelitian .....	100
BAB V PENUTUP.....		103
5.1	Kesimpulan .....	103
5.2	Saran .....	103
DAFTAR PUSTAKA .....		105
LAMPIRAN.....		109

## DAFTAR TABEL

Tabel 1.1 Perbedaan Variabel dan Baku Mutu Kualitas Air.....	3
Tabel 2.1 Tabel Baku Mutu Permenkes No. 2 Tahun 2023.....	7
Tabel 2.2 Rentang Indeks Pencemaran Air.....	9
Tabel 2.3 Tabel Kontingensi Variabel X dan Variabel Y.....	14
Tabel 2.4 <i>Confusion Matrix</i> .....	20
Tabel 2.5 Penelitian Relevan.....	27
Tabel 2.6 <i>Novelity</i> .....	29
Tabel 3.1 Daftar spesifikasi perangkat penelitian.....	34
Tabel 3.2 Metadata Dataset.....	35
Tabel 3.3 Kolom Tabel Kualitas Air Sumur DKI Jakarta 2023.....	38
Tabel 3.4 Ketentuan Komposisi <i>Pivot Data</i> .....	39
Tabel 3.5 Kolom Dataset setelah <i>Pivot Data</i> .....	39
Tabel 3.6 Skenario Parameter untuk Algoritma KNN.....	44
Tabel 3.7 Kategori Nilai AUC-ROC.....	48
Tabel 3.8 <i>Confusion Matrix</i> untuk Evaluasi Data.....	49
Tabel 3.9 <i>Confusion Matrix</i> untuk kelas 0 (baik).....	50
Tabel 3.10 <i>Confusion Matrix</i> untuk kelas 1 (ringan).....	50
Tabel 3.11 <i>Confusion Matrix</i> untuk kelas 2 (sedang).....	50
Tabel 3.12 <i>Confusion Matrix</i> untuk kelas 3 (berat).....	50
Tabel 4.1 24 Baris Sampel Dataset Kualitas Air Sumur DKI Jakarta 2023.....	53
Tabel 4.2 <i>Feature</i> Dataset Kualitas Air Sumur DKI Jakarta 2023.....	55
Tabel 4.3 Baris-baris dalam <i>Feature</i> parameter dari Dataset.....	55
Tabel 4.4 Kelengkapan Parameter Pemantauan.....	56
Tabel 4.5 Dataset Setelah Proses <i>Pivot Data</i> .....	57
Tabel 4.6 Dataset Setelah Proses <i>Data Cleansing</i> .....	66
Tabel 4.7 <i>Feature</i> Baru Berdasarkan Nilai Baku Mutu.....	66
Tabel 4.8 Dataset Setelah Proses <i>Data Enrichment</i> .....	68
Tabel 4.9 <i>Feature</i> yang akan dilakukan <i>Ordinal Encoding</i> .....	69
Tabel 4.10 Dataset Setelah Proses <i>Ordinal Encoding</i> .....	71
Tabel 4.11 Pembagian Variabel Dependen dan Independen.....	72



Tabel 4.12 Sampel Tabel Kontingensi 2x4 Antara <i>Feature</i> ‘Bau’ dan ‘Status’ ...	74
Tabel 4.13 Pembagian Variabel Dependen dan Independen .....	74
Tabel 4.14 Pembagian Variabel Dependen dan Independen <i>Feature Important</i> ..	76
Tabel 4.15 Dataset Setelah Proses <i>Feature Selection</i> .....	77
Tabel 4.16 Data Sintesis Baru Hasil <i>Oversampling SMOTE</i> .....	79
Tabel 4.17 Dataset Setelah Proses <i>Normalization</i> .....	80
Tabel 4.18 Tabel Skenario Pengujian Penerapan Model .....	82
Tabel 4.19 10 Skenario Terbaik dalam Pembangunan Model <i>K-Nearest Neighbors</i> .....	84
Tabel 4.20 <i>Confusion Matrix</i> untuk kelas 0 (baik) .....	88
Tabel 4.21 <i>Confusion Matrix</i> untuk kelas 1 (cemar ringan) .....	88
Tabel 4.22 <i>Confusion Matrix</i> untuk kelas 2 (cemar sedang) .....	88
Tabel 4.23 <i>Confusion Matrix</i> untuk kelas 3 (cemar berat).....	88
Tabel 4.24 Tabel Skenario Pengujian Penerapan Model .....	101
Tabel 4.25 Skenario Pembangunan Model Terbaik.....	101





## DAFTAR GAMBAR

Gambar 2.1 Alur Proses KDD .....	10
Gambar 2.2 Notasi UML pada Diagram Aktivitas .....	12
Gambar 2.3 Proses Pembelajaran <i>Machine Learning</i> secara <i>Supervised</i> .....	16
Gambar 2.4 Ilustrasi Algoritma <i>K-Nearest Neighbors</i> .....	17
Gambar 2.5 Ilustrasi <i>K-Fold Cross Validation</i> .....	18
Gambar 2.6 <i>Line Chart</i> .....	23
Gambar 2.7 <i>Bar Chart</i> .....	24
Gambar 2.8 <i>Pie Chart</i> .....	24
Gambar 2.9 <i>Choropleth Map</i> .....	25
Gambar 2.10 <i>Treemap</i> .....	25
Gambar 2.11 <i>Heatmap</i> .....	26
Gambar 2.12 <i>Scatter plot</i> .....	26
Gambar 2.13 <i>Scorecard</i> .....	26
Gambar 2.14 Kerangka Berpikir .....	31
Gambar 3.1 Langkah-Langkah Penelitian.....	36
Gambar 3.2 Menghubungkan Google Colab ke Google Drive.....	37
Gambar 3.3 Menampilkan <i>Preview Dataset</i> .....	37
Gambar 3.4 Proses <i>Pivot Data</i> .....	39
Gambar 3.5 Pemeriksaan Tipe Data.....	40
Gambar 3.6 Pemeriksaan <i>Missing Value</i> .....	40
Gambar 3.7 Pemeriksaan <i>Duplicated Value</i> .....	40
Gambar 3.8 Pemeriksaan Pencilan ( <i>Outlier</i> ) dengan Standar Deviasi .....	40
Gambar 3.9 Memperbaiki Tipe Data .....	41
Gambar 3.10 Penghapusan <i>Missing Value</i> .....	41
Gambar 3.11 Penghapusan <i>Duplicated Value</i> .....	41
Gambar 3.12 Penghapusan <i>Outlier</i> .....	41
Gambar 3.13 Sintaks Mengubah Data Kategorikal.....	42
Gambar 3.14 <i>Syntax</i> Pembagian Variabel untuk Pengujian.....	42
Gambar 3.15 <i>Syntax Chi Square Test</i> .....	42
Gambar 3.16 <i>Syntax Spearman Rho Test</i> .....	43
Gambar 3.17 <i>Syntax Data Resampling</i> dengan SMOTE .....	43

Gambar 3.18 <i>Syntax</i> Normalisasi Data .....	43
Gambar 3.19 <i>Syntax</i> Pembagian Data Latih dan Data Uji .....	44
Gambar 3.20 <i>Load Model K-Nearest Neighbors</i> .....	44
Gambar 3.21 <i>Syntax</i> Inisiasi <i>Library</i> ‘GridSearchCV’ .....	46
Gambar 3.22 <i>Syntax</i> Pengujian Model Klasifikasi .....	47
Gambar 3.23 <i>Syntax Confusion Matrix</i> .....	47
Gambar 3.24 <i>Syntax Classification Report</i> .....	47
Gambar 3.25 <i>Syntax Balanced Accuracy</i> .....	47
Gambar 3.26 <i>Syntax</i> Pengujian Probabilitas Model Klasifikasi .....	48
Gambar 3.27 <i>Syntax</i> Perhitungan <i>ROC Curve</i> .....	48
Gambar 3.28 <i>Syntax</i> Perhitungan nilai AUC-ROC .....	48
Gambar 3.29 <i>Syntax</i> Plot <i>ROC Curve</i> .....	48
Gambar 4.1 Halaman Dataset pada Situs Satu Data Jakarta .....	52
Gambar 4.2 Proses <i>Import Data</i> ke dalam Google Colab .....	53
Gambar 4.3 Proses <i>Pivot Data</i> .....	57
Gambar 4.4 Informasi Dataset setelah <i>Pivot Data</i> .....	58
Gambar 4.5 Hasil Pemeriksaan Tipe Data .....	59
Gambar 4.6 Proses memperbaiki Tipe Data .....	59
Gambar 4.7 Tipe Data pada Dataset setelah di perbaiki .....	60
Gambar 4.8 Proses Pemeriksaan Nilai yang Kosong ( <i>Missing Value</i> ) .....	61
Gambar 4.9 Penghapusan <i>Feature</i> ‘Organik’, dan ‘Surfaktan Anionik (Senyawa aktif biru metilen)’ .....	61
Gambar 4.10 Kondisi Data Setelah Penanganan <i>Missing Value</i> .....	62
Gambar 4.11 Penghapusan baris yang memiliki nilai kosong pada <i>Feature</i> ‘Warna’ .....	62
Gambar 4.12 Proses Pemeriksaan Data Ganda ( <i>Duplicated Data</i> ) .....	63
Gambar 4.13 Proses Pemeriksaan <i>Outlier</i> dengan <i>Boxplot</i> .....	63
Gambar 4.14 Proses Pemeriksaan <i>class</i> ‘zscore’ untuk melihat <i>outlier</i> .....	64
Gambar 4.15 Proses Pemeriksaan Nilai Salah <i>Input</i> .....	65
Gambar 4.16 Proses Perbaikan <i>Data Entry Error</i> .....	65
Gambar 4.17 Informasi Dataset setelah tahap <i>Data Cleansing</i> .....	65
Gambar 4.18 Penentuan <i>Feature</i> yang akan dikembangkan .....	67
Gambar 4.19 Proses Inisiasi Nilai Baku Mutu sesuai Permenkes No.2 tahun 2023 .....	67

Gambar 4.20 Proses Kategorisasi Baku Mutu Parameter .....	68
Gambar 4.21 Informasi Dataset Setelah Tahap <i>Data Enrichment</i> .....	69
Gambar 4.22 Proses <i>Ordinal Encoding</i> dengan <i>Library</i> .....	70
Gambar 4.23 Proses <i>Ordinal Encoding</i> dengan Pemetaan Langsung .....	70
Gambar 4.24 Proses Pembagian Variabel Dependen dan Independen .....	72
Gambar 4.25 Proses Perhitungan $p\_values$ dan $\chi^2$ .....	73
Gambar 4.26 Proses Perbandingan $p\_values$ dengan $\alpha$ .....	73
Gambar 4.27 Hasil Proses Perhitungan <i>Chi-Square Test</i> .....	73
Gambar 4.28 Proses Pembagian Variabel Dependen dan Independen .....	75
Gambar 4.29 Proses Perhitungan $p\_values$ dan $\rho$ .....	75
Gambar 4.30 Proses Perbandingan $p\_values$ dengan $\alpha$ .....	75
Gambar 4.31 Hasil Proses Perhitungan <i>Spearman Rho Test</i> .....	75
Gambar 4.32 Proses Seleksi <i>Feature Important</i> .....	76
Gambar 4.33 Distribusi Data Sebelum <i>Data Resampling</i> .....	77
Gambar 4.34 Proses <i>Data Resampling</i> dengan SMOTE .....	78
Gambar 4.35 Distribusi Data Setelah <i>Data Resampling</i> .....	78
Gambar 4.36 Dimensi Data Setelah <i>Data Resampling</i> .....	79
Gambar 4.37 Proses <i>Format Data</i> dengan <i>Min-Max Scaler</i> .....	80
Gambar 4.38 Proses Pembagian Data Latih dan Data Uji .....	81
Gambar 4.39 Proses <i>Load Model K-Nearest Neighbors</i> .....	81
Gambar 4.40 Proses Inisiasi Skenario Parameter <i>K-Nearest Neighbors</i> .....	82
Gambar 4.41 Proses Inisiasi <i>Library 'GridSearchCV'</i> .....	82
Gambar 4.42 Proses Pembangunan Model .....	83
Gambar 4.43 Proses Pembangunan Model .....	83
Gambar 4.44 Proses Pengujian Model dengan Data Uji.....	85
Gambar 4.45 Proses Pengujian Model dengan Data Uji.....	85
Gambar 4.46 Proses Pengujian Model dengan Data Uji.....	85
Gambar 4.47 Hasil Validasi Model dengan Data Latih .....	86
Gambar 4.48 Proses Pengujian Model dengan Data Uji.....	86
Gambar 4.49 Hasil Pengujian Model dengan Data Uji.....	87
Gambar 4.50 Statistik Pengujian Model .....	88
Gambar 4.51 Proses Pengujian Probabilitas Model.....	90
Gambar 4.52 Proses <i>One Hot Encoding</i> pada Variabel ' <i>y_test</i> ' .....	90
Gambar 4.53 Ilustrasi Proses <i>One Hot Encoding</i> pada ' <i>y_test</i> ' .....	91

Gambar 4.54 Hasil Perhitungan AUC-ROC untuk Setiap Kelas.....	91
Gambar 4.55 ROC <i>Curve</i> untuk Setiap Kelas.....	92
Gambar 4.56 Nilai AUC-ROC Model secara <i>Macro Average</i> .....	92
Gambar 4.57 <i>Sidebar Dashboard</i> .....	93
Gambar 4.58 <i>Dashboard</i> Visualisasi Data – Halaman <i>Homepage</i> .....	94
Gambar 4.59 <i>Dashboard</i> Visualisasi Data – Halaman Detail Parameter .....	95
Gambar 4.60 <i>Dashboard</i> Visualisasi Data – Halaman Persebaran Data .....	96
Gambar 4.61 <i>Dashboard</i> Visualisasi Data – Halaman Korelasi Data .....	97
Gambar 4.62 <i>Dashboard</i> Visualisasi Data – Halaman Skenario Pelatihan Model	98
Gambar 4.63 <i>Dashboard</i> Visualisasi Data – Halaman Evaluasi Model .....	99





## DAFTAR LAMPIRAN

Lampiran 1. Surat Tugas Dosen Pembimbing .....	109
Lampiran 2. Lembar Konsultasi Dosen Pembimbing I.....	111
Lampiran 3. Lembar Konsultasi Dosen Pembimbing II .....	112
Lampiran 4. Surat Pernyataan Dosen Pembimbing I.....	113
Lampiran 5. Surat Pernyataan Dosen Pembimbing II.....	114
Lampiran 6. Sampel Dataset Kualitas Air Sumur di DKI Jakarta 2023 .....	115
Lampiran 7. Dataset Setelah Tahap <i>Data Selection</i> .....	118
Lampiran 8. Dataset Setelah Tahap <i>Preprocessing</i> .....	149
Lampiran 9. Tabel Distribusi <i>Chi Square</i> (dk=1 s.d. dk=1000) .....	190
Lampiran 10. Tabel Distribusi <i>Spearman Rho</i> (Uji Dua Sisi, n=500 s.d n=600) .....	191
Lampiran 11. Dataset Sintesis Hasil <i>Oversampling</i> SMOTE .....	192
Lampiran 12. Sampel Dataset Setelah Tahap <i>Transformation</i> (Hanya Variabel Dependen) .....	220
Lampiran 13. Hasil Penerapan Model KNN dengan GridSearchCV.....	230
Lampiran 14. Hasil Validasi Model dengan Data Latih .....	236
Lampiran 15. Kesalahan Prediksi ( <i>False Positive</i> ) pada Proses Validasi dengan Data Latih.....	241
Lampiran 16. Hasil Pengujian Model dengan Data Uji .....	243
Lampiran 17. Kesalahan Prediksi ( <i>False Positive</i> ) pada Proses Pengujian dengan Data Uji .....	245
Lampiran 18. <i>Syntax Python</i> .....	246