

DAFTAR ISI

	Halaman
HALAMAN PERNYATAAN KEASLIAN	ii
HALAMAN PENGESAHAN TUGAS AKHIR.....	iv
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	v
KATA PENGANTAR	vi
ABSTRAK	viii
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
BAB 1 PENDAHULUAN.....	16
1.1 Latar Belakang	16
1.2 Identifikasi Masalah	17
1.3 Tujuan Tugas Akhir.....	17
1.4 Manfaat Tugas Akhir.....	17
1.5 Lingkup Tugas Akhir	17
1.6 Kerangka Berpikir	18
1.7 Sistematika Penulisan Tugas Akhir.....	19
BAB 2 TINJAUAN PUSTAKA.....	20
2.1 Polusi Udara	20
2.1.1 Nitrogen Dioksida (NO ₂).....	20
2.1.2 Ozon (O ₃).....	20
2.1.3 Sulfur Dioksida (SO ₂).....	20
2.1.4 <i>Particulate Matter</i> (PM).....	21
2.1.5 Karbon Monoksida (CO).....	21
2.2 Indeks Standar Pencemaran Udara (ISPU).....	21
2.2.1 Tata Cara Perhitungan ISPU	22
2.2.2 Penentuan Kategori ISPU.....	25
2.3 Google Collaboratory	26
2.3.1 Pandas	26
2.3.2 Scikit-Learn.....	27
2.3.3 NumPy.....	27

2.3.4	Matplotlib	27
2.4	<i>Data Mining</i>	28
2.5	<i>Cross-Industry Standard Process For Data Mining (CRISP-DM)</i>	28
2.6	<i>Business Understanding</i>	29
2.7	<i>Data Understanding</i>	29
2.8	<i>Data Preparation</i>	29
2.8.1	<i>Data Visualization</i>	29
2.8.2	<i>Distribution Plots: Boxplots (Identify Outlier)</i>	29
2.8.3	<i>Data Cleaning</i>	30
2.8.4	<i>Missing Value: K-Nearest Neighbor Imputation</i>	31
2.8.5	<i>Winsorizing</i>	31
2.8.6	<i>Normal Distribution</i>	33
2.9	<i>Modelling</i>	35
2.10	<i>Evaluation</i>	35
2.10.1	Confusion Matrix	35
2.11	<i>Deployment</i>	36
2.12	<i>Classification Algorithm</i>	36
2.12.1	<i>Train And Test Split</i>	36
2.12.2	<i>Stratified K-Fold Cross Validation</i>	37
2.12.3	<i>Decision Tree Classification And Regression Trees (CART)</i>	37
2.12.4	<i>Naïve Bayes</i>	39
2.12.5	<i>K-Nearest Neighbor</i>	41
BAB 3	METODE	42
3.1	Rencana Penelitian	42
3.2	Tahapan Penelitian	42
3.2.1	<i>Business Understanding</i>	42
3.2.2	<i>Data Understanding</i>	42
3.2.3	<i>Data Preparation</i>	43
3.2.4	<i>Modelling</i>	47
3.2.1	<i>Evaluation</i>	65
3.2.1	<i>Deployment</i>	66
BAB 4	HASIL DAN PEMBAHASAN	67
4.1	<i>Business Understanding</i>	67
4.1.1	Penentuan Tujuan Bisnis	67
4.1.2	Persyaratan Proyek	67

4.1.3	Tentukan Tujuan Data Mining	68
4.2	<i>Data Understanding</i>	69
4.2.1	Pengumpulan Data Awal	69
4.2.2	Penjabaran Data	69
4.2.3	Eksplorasi Data	70
4.2.4	Verifikasi Kualitas Data	71
4.3	<i>Data Preparation</i>	71
4.4	<i>Modelling</i>	82
4.5	<i>Evaluation</i>	91
4.6	<i>Deployment</i>	94
BAB 5 KESIMPULAN DAN SARAN		96
5.1	Kesimpulan	96
5.2	Saran	96
DAFTAR REFERENSI		97
Lampiran 1 Daftar Riwayat Hidup		104
Lampiran 2 Validasi Data ISPU		105

DAFTAR TABEL

	Halaman
Tabel 2.1. Konversi Nilai Konsentrasi	22
Tabel 2.2. Contoh penyampaian ISPU	24
Tabel 2.3. Kategori ISPU	25
Tabel 2.4. Penjelasan nilai ISPU	25
Tabel 3.1. Data Sample	43
Tabel 3.2. Data ISPU setelah KNN <i>Imputation</i>	44
Tabel 3.3. Penanganan outlier menggunakan Teknik winsorizing.	46
Tabel 3.4. Splitting 70% data training dan 30% data uji	47
Tabel 3.5. Iterasi pertama perhitungan decision tree CART	49
Tabel 3.6. Iterasi kedua PM10 gini index >60.5 perhitungan decision tree CART	51
Tabel 3.7. Iterasi kedua PM10 <i>gini index</i> ≤ 60.5 perhitungan <i>decision tree</i> CART	52
Tabel 3.8. Prediksi <i>Decision Tree</i>	54
Tabel 3.9. Rata-rata dan standar deviasi parameter dari masing-masing label	54
Tabel 3.10. Penambahan 1 Data <i>Naive Bayes</i>	54
Tabel 3.11. Standar deviasi dan rata-rata dari penambahan data	55
Tabel 3.12. Hasil prediksi data ISPU menggunakan <i>naïve bayes</i>	60
Tabel 3.13. Data <i>training testing</i> untuk perhitungan algoritma <i>K-Nearest Neighbor</i>	60
Tabel 3.14. Hasil prediksi menggunakan data pertama	61
Tabel 3.15. Hasil prediksi menggunakan data kedua	62
Tabel 3.16. Hasil prediksi menggunakan data ketiga	63
Tabel 3.17. Hasil prediksi menggunakan data keempat	64
Tabel 3.18. Confussion matrix algoritma decision tree.	65
Tabel 4.1 Penjelasan tiap parameter	70

DAFTAR GAMBAR

	Halaman
Gambar 1-1 Kerangka Berpikir Penelitian.....	18
Gambar 2-1 Proses CRISP-DM (sumber: Data Mining The CRISP-DM Methodology The CLEM language and IBM Modeler, 2021.)	28
Gambar 2-2 <i>Boxplot diagram</i> (sumber: Data Mining for Business Analytics: Concepts, Techniques, and Applications in R, 2017.) Line Plot	30
Gambar 2-3 Diagram Garis (sumber: Data Science Fundamentals and Practical Approaches)	30
Gambar 2-4 Distribusi normal dengan rata-rata dan simpang baku yang berbeda (sumber: Core Statistical Concepts With Excel An Interactive Modular Approach, 2018)	33
Gambar 2-5. Proporsi area dengan tiap standar deviasi juga rata-rata.....	34
Gambar 2-6. <i>Confusion Matrix</i> total tupel positif dan negative (sumber: <i>Data mining: concepts and techniques</i> , 2011).....	35
Gambar 2-7 <i>Training and testing of the classifier</i> (sumber: <i>Data mining and data warehousing: principles and practical techniques</i> , 2019).....	36
Gambar 2-8. <i>Fivefold cross-validation</i> (sumber: Soft Computing and Signal Processing: Proceedings of 3rd ICSCSP 2020, Volume 1).....	37
Gambar 2-9 Prediksi Decision Tree pembelian laptop (sumber: <i>Data mining: concepts and techniques</i> , 2011).....	38
Gambar 3-1 Boxplot data ISPU melihat outlier.....	45
Gambar 3-2 Penggunaan <i>stratified k-fold cross validation</i> dalam model.....	48
Gambar 3-3 Iterasi akar pertama.....	51
Gambar 3-4. Iterasi akar kedua	52
Gambar 3-5. Iterasi akar ketiga.....	53
Gambar 4-1 Data ISPU	71
Gambar 4-2 Perubahan kategori menjadi numerik	72
Gambar 4-3 Identifikasi null pada data.....	72
Gambar 4-4 Data yang berisikan <i>null</i>	72
Gambar 4-5 Penggunaan k-nn imputer dengan <i>library scikit-learn</i>	73
Gambar 4-6 Data setelah k-nn imputation	73
Gambar 4-7 Data hasil k-nn imputation.....	74
Gambar 4-8 Identifikasi outlier dengan boxplot	74
Gambar 4-9 <i>Function</i> mencari RUB dan RLB	75
Gambar 4-10 Batas maksimum minimum parameter PM10 dan so2	75
Gambar 4-11 Batas maksimum minimum parameter co dan o3.....	75
Gambar 4-12 Batas maksimum minimum parameter NO2.....	75
Gambar 4-13 Analisa <i>outlier</i> data batas maksimum PM10	76
Gambar 4-14 Analisa <i>outlier</i> data batas maksimum so2	77

Gambar 4-15 Analisa <i>outlier</i> data batas maksimum o ₃	77
Gambar 4-16 Analisa <i>outlier</i> data batas maksimum NO ₂	77
Gambar 4-17 Analisa <i>outlier</i> data batas maksimum co	78
Gambar 4-18 Analisa <i>outlier</i> data batas minimum PM ₁₀	78
Gambar 4-19 Analisa <i>outlier</i> data batas minimum so ₂	79
Gambar 4-20 Jumlah data tiap kategori	79
Gambar 4-21 <i>Function</i> mengganti <i>outlier</i> batas minimum.....	80
Gambar 4-22 Implementasi penanganan <i>outlier</i> batas bawah	80
Gambar 4-23 <i>Boxplot</i> hasil <i>winsorizing</i>	80
Gambar 4-24 Perbandingan persebaran data.....	81
Gambar 4-25 Analisis parameter penentu kualitas udara	81
Gambar 4-26 Pemisahan data <i>training</i> dan <i>testing</i>	82
Gambar 4-27 <i>Modelling Decision Tree</i>	82
Gambar 4-28 Hasil Model Decision Tree	83
Gambar 4-29 Hasil akurasi dan error model data latih	83
Gambar 4-30 <i>Modelling k-nearest neighbor</i>	84
Gambar 4-31 Hasil rata-rata akurasi dari tiap k.	85
Gambar 4-32 Analisa k terbaik	85
Gambar 4-33 Distribusi normal parameter PM ₁₀	86
Gambar 4-34 Distribusi normal parameter so ₂	87
Gambar 4-35 Distribusi normal parameter co.....	87
Gambar 4-36 Distribusi normal parameter NO ₂	88
Gambar 4-37 Distribusi normal parameter o ₃	88
Gambar 4-38 Total data tiap kategori	89
Gambar 4-39 <i>Modelling naïve bayes</i>	90
Gambar 4-40 Hasil akurasi error data latih <i>naïve bayes</i>	90
Gambar 4-41 Evaluasi model <i>decision tree</i> menggunakan <i>confussion matrix</i>	91
Gambar 4-42 Hasil Evaluasi <i>decision tree</i>	91
Gambar 4-43 Evaluasi model <i>k-nearest Neighbor</i> menggunakan <i>confussion matrix</i>	92
Gambar 4-44 Hasil evaluasi <i>k-nearest Neighbor</i>	92
Gambar 4-45 Evaluasi model <i>naïve bayes</i> menggunakan <i>confussion matrix</i>	93
Gambar 4-46 Hasil evaluasi <i>naïve bayes</i>	93