



## **SKRIPSI**

# **IMPLEMENTASI ALGORITMA XGBOOST, CATBOOST, DAN LGBM UNTUK KLASIFIKASI PENCEMARAN UDARA DI KOTA SURABAYA**

**GILANG ENGGAR SAPUTRA**

NPM 21081010237

## **DOSEN PEMBIMBING**

Made Hanindia Prami Swari, S.Kom, M.Cs.

Afina Lina Nurlaili, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA  
SURABAYA  
2026**



## **SKRIPSI**

# **IMPLEMENTASI ALGORITMA *XGBOOST*, *CATBOOST*, DAN LGBM UNTUK KLASIFIKASI PENCEMARAN UDARA DI KOTA SURABAYA**

**GILANG ENGGAR SAPUTRA**  
NPM 21081010237

## **DOSEN PEMBIMBING**

Made Hanindia Prami Swari, S.Kom, M.Cs.  
Afina Lina Nurlaili, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA  
SURABAYA  
2026**

*Halaman ini sengaja dikosongkan*

## LEMBAR PENGESAHAN

### IMPLEMENTASI ALGORITMA XGBOOST, CATBOOST, DAN LGBM UNTUK KLASIFIKASI PENCEMARAN UDARA DI KOTA SURABAYA

Oleh :  
GILANG ENGGAR SAPUTRA  
NPM. 21081010237

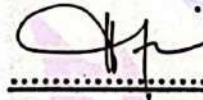
Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 13 Januari 2026

Made Hanindia Prami Swari, S.Kom, M.Cs.  
NIP. 198902052018032001



(Pembimbing I)

Afina Lina Nurlaili, S.Kom., M.Kom.  
NIP. 199312132022032010



(Pembimbing II)

Dr. Faisal Muttaqin, S.Kom, M.T.  
NIP. 198512312021211009



(Penguji 1)

Muhammad Muharrom Al Haromainy,  
S.Kom, M.Kom  
NIP. 199506012022031006



(Penguji II)

Mengetahui,  
Dekan Fakultas Ilmu-Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT  
NIP. 19681126 199403 2 001

*Halaman ini sengaja dikosongkan*

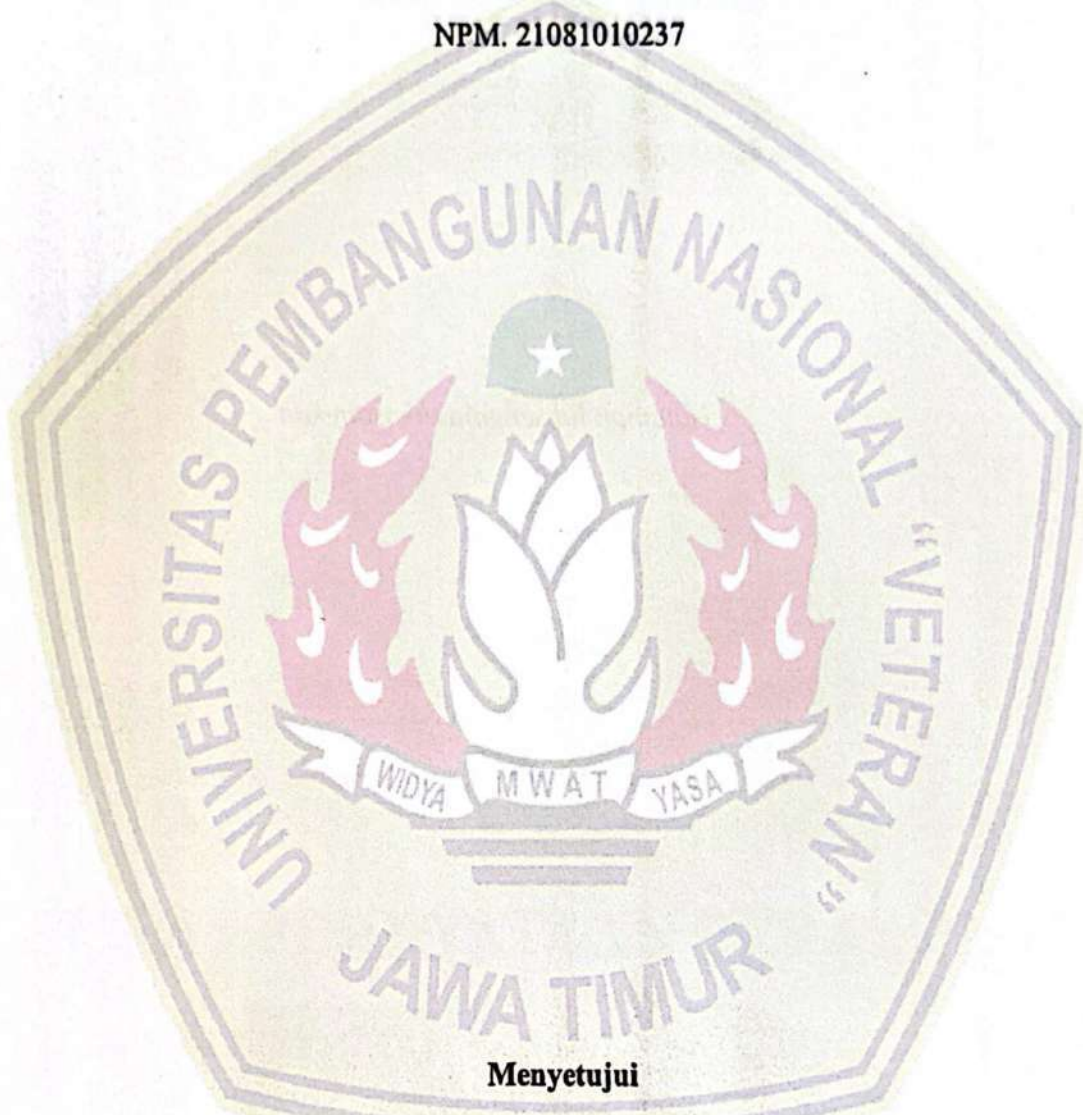
**LEMBAR PERSETUJUAN**

**IMPLEMENTASI ALGORITMA XGBOOST, CATBOOST, DAN LGBM  
UNTUK KLASIFIKASI PENCEMARAN UDARA DI KOTA SURABAYA**

**Oleh:**

**GILANG ENGGAR SAPUTRA**

**NPM. 21081010237**



**Menyetujui**

**Koordinator Program Studi Informatika**

**Fakultas Ilmu Komputer**

**Fetty Tri Anggraeny, S.Kom. M.Kom**

**NIP. 198202112021212005**

*Halaman ini sengaja dikosongkan*

## SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertandatangan di bawah ini:

Nama Mahasiswa : GILANG ENGGAR SAPUTRA  
NPM : 21081010237  
Program : Sarjana(S1)  
Program Studi : Informatika  
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.

Surabaya, 13 Januari 2026

Yang Membuat Pernyataan,



(Gilang Enggar Saputra)

NPM. 21081010237

*Halaman ini sengaja dikosongkan*

## ABSTRAK

Nama Mahasiswa / NPM : Gilang Enggar Saputra / 21081010237  
Judul Skripsi : Implementasi Algoritma *XGBoost*, *CatBoost*, dan  
LGBM Untuk Klasifikasi Pencemaran Udara di  
Kota Surabaya  
Dosen Pembimbing : 1. Made Hanindia Prami Swari, S.Kom, M.Cs.  
2. Afina Lina Nurlaili, S.Kom., M.Kom.

Penelitian ini bertujuan untuk mengklasifikasikan tingkat pencemaran udara di Kota Surabaya berdasarkan kategori Indeks Standar Pencemar Udara (ISPU) menggunakan tiga algoritma machine learning berbasis ensemble boosting, yaitu *XGBoost*, *CatBoost*, dan *LightGBM*. Data yang digunakan terdiri atas parameter kualitas udara seperti PM10, SO<sub>2</sub>, CO, O<sub>3</sub>, dan NO<sub>2</sub>, yang kemudian diproses melalui tahap cleaning, encoding, normalisasi, serta pembagian data menjadi training dan testing dengan beberapa rasio (70:30, 75:25, dan 80:20). Proses pengujian dilakukan dengan berbagai kombinasi learning rate (0.01 dan 0.10) serta jumlah iterasi (100, 500, dan 1000). Evaluasi performa model menggunakan metrik akurasi, precision, recall, dan F1-score. Hasil penelitian menunjukkan bahwa ketiga model mampu mempelajari pola data dengan baik, namun performanya dipengaruhi oleh ketidakseimbangan data antar kelas. Model *XGBoost* dengan penerapan metode *ClassWeight* dan penggabungan kelas minoritas memberikan hasil terbaik dengan nilai accuracy 0.9594, precision 0.8632, recall 0.7787, dan F1-score 0.8098 pada rasio data 70:30, learning rate 0.10, dan iterasi 500. Model *CatBoost* dan *LightGBM* juga menunjukkan performa baik, masing-masing dengan F1-score tertinggi 0.7549 dan 0.7263 setelah penyeimbangan data diterapkan. Secara keseluruhan, kombinasi teknik *ClassWeight* dan penggabungan kelas terbukti efektif dalam mengatasi masalah ketidakseimbangan data dan meningkatkan kemampuan model dalam mengenali kategori pencemaran udara yang jarang muncul. Hasil penelitian ini diharapkan dapat menjadi dasar bagi pengembangan sistem prediksi kualitas udara yang lebih akurat dan adaptif di masa mendatang.

**Kata kunci :** Pencemaran Udara, ISPU, *XGBoost*, *CatBoost*, *LightGBM*, Machine Learning, Klasifikasi.

*Halaman ini sengaja dikosongkan*

## ABSTRACT

Student Name / NPM : Gilang Enggar Saputra / 21081010237  
Thesis Title : Implementation of *XGBoost*, *CatBoost*, and LGBM Algorithms for Air Pollution Classification in Surabaya City  
Advisor : 1. Made Hanindia Prami Swari, S.Kom, M.Cs.  
2. Afina Lina Nurlaili, S.Kom., M.Kom.

This study aims to classify the level of air pollution in Surabaya City based on the Air Pollution Standard Index (ISPU) categories using three ensemble boosting-based machine learning algorithms: *XGBoost*, *CatBoost*, and LightGBM. The dataset consists of air quality parameters such as PM10, SO<sub>2</sub>, CO, O<sub>3</sub>, and NO<sub>2</sub>, which were processed through data cleaning, encoding, normalization, and splitting into training and testing sets with ratios of 70:30, 75:25, and 80:20. The testing process was carried out using various combinations of learning rates (0.01 and 0.10) and iteration counts (100, 500, and 1000). Model performance was evaluated using the accuracy, precision, recall, and F1-score metrics. The results indicate that all three models successfully learned the data patterns, although their performance was influenced by class imbalance. The *XGBoost* model achieved the best performance by applying the *ClassWeight* method and minority class merging, resulting in an accuracy of 0.9594, precision of 0.8632, recall of 0.7787, and F1-score of 0.8098 with a 70:30 data split, 0.10 learning rate, and 500 iterations. The *CatBoost* and LightGBM models also performed well, obtaining the highest F1-scores of 0.7549 and 0.7263, respectively, after data balancing was applied. Overall, the combination of the *ClassWeight* technique and minority class merging proved effective in handling data imbalance and improving the model's ability to recognize rare air pollution categories. The findings of this research are expected to serve as a foundation for developing a more accurate and adaptive air quality prediction system in the future.

**Keywords:** Air Pollution, ISPU, *XGBoost*, *CatBoost*, LightGBM, Machine Learning, Classification.

*Halaman ini sengaja dikosongkan*

## KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas segala rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Implementasi Algoritma XGBoost, CatBoost, dan LGBM Untuk Klasifikasi Pencemaran Udara di Kota Surabaya”** dengan baik dan lancar. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur. Dalam proses penyusunan skripsi ini, penulis mendapatkan banyak dukungan, arahan, serta bantuan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT., selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom., selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Made Hanindia Prami Swari, S.Kom., M.Cs., selaku Dosen Pembimbing I, yang telah meluangkan waktu, memberikan bimbingan, arahan, dan motivasi kepada penulis dalam penyusunan skripsi ini.
4. Ibu Afina Lina Nurlaili, S.Kom., M.Kom., selaku Dosen Pembimbing II, yang telah memberikan banyak masukan, saran, dan dukungan hingga penelitian ini dapat terselesaikan dengan baik.
5. Seluruh dosen dan staf Program Studi Informatika Fakultas Ilmu Komputer yang telah memberikan ilmu, pengalaman, dan bimbingan berharga selama masa perkuliahan.
6. Kedua orang tua tercinta, atas segala doa, kasih sayang, semangat, dan dukungan yang tiada henti dalam setiap langkah penulis.
7. Kakak penulis, yang selalu memberikan dorongan moral dan motivasi ketika penulis merasa lelah atau kehilangan semangat.
8. Dan tak lupa, kucing kesayangan penulis my owiii my lovelyyy , yang dengan tingkah lucunya menjadi teman setia dan penghibur di tengah proses panjang penyusunan skripsi ini.

Penulis menyadari bahwa dalam penyusunan skripsi ini masih terdapat kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi kesempurnaan karya ini di masa mendatang.

Akhirnya, dengan segala keterbatasan yang penulis miliki, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi semua pihak, khususnya dalam pengembangan ilmu pengetahuan di bidang Informatika dan Machine Learning.

Surabaya, 13 Januari 2026

Penulis

(Gilang Enggar Saputra)

NPM. 21081010237

## DAFTAR ISI

|                                                     |             |
|-----------------------------------------------------|-------------|
| <b>JUDUL SKRIPSI .....</b>                          | <b>i</b>    |
| <b>LEMBAR PENGESAHAN .....</b>                      | <b>iii</b>  |
| <b>LEMBAR PERSETUJUAN .....</b>                     | <b>v</b>    |
| <b>SURAT PERNYATAAN BEBAS PLAGIASI.....</b>         | <b>vii</b>  |
| <b>ABSTRAK .....</b>                                | <b>ix</b>   |
| <b>KATA PENGANTAR.....</b>                          | <b>xiii</b> |
| <b>DAFTAR ISI.....</b>                              | <b>xv</b>   |
| <b>DAFTAR GAMBAR.....</b>                           | <b>xvii</b> |
| <b>DAFTAR TABEL .....</b>                           | <b>xx</b>   |
| <b>BAB I PENDAHULUAN.....</b>                       | <b>1</b>    |
| 1.1 Latar Belakang .....                            | 1           |
| 1.2 Rumusan Masalah .....                           | 4           |
| 1.3 Tujuan Penelitian .....                         | 5           |
| 1.4 Manfaat Penelitian .....                        | 5           |
| 1.5 Batasan Masalah.....                            | 5           |
| <b>BAB II TINJAUAN PUSTAKA .....</b>                | <b>6</b>    |
| 2.1. Penelitian Terdahulu .....                     | 6           |
| 2.2. Landasan Teori.....                            | 8           |
| 2.2.1 Pencemaran Udara .....                        | 8           |
| 2.2.2 <i>XGBoost</i> .....                          | 10          |
| 2.2.3 <i>CatBoost</i> .....                         | 13          |
| 2.2.4 LGBM .....                                    | 16          |
| 2.2.5 Data Mining .....                             | 19          |
| 2.2.6 Evaluasi Model.....                           | 22          |
| <b>BAB III DESAIN DAN IMPLEMENTASI SISTEM .....</b> | <b>25</b>   |
| 3.1 Desain Penelitian.....                          | 25          |
| 3.2 Data Penelitian .....                           | 26          |
| 3.2.1 Sumber Data.....                              | 26          |
| 3.2.2 Variabel Data .....                           | 26          |
| 3.3 Pengolahan Data.....                            | 27          |
| 3.3.1 Proses Data <i>Understanding</i> .....        | 27          |

|                             |                                                                     |            |
|-----------------------------|---------------------------------------------------------------------|------------|
| 3.3.2                       | Proses Data <i>Preparation</i> .....                                | 30         |
| 3.3.3                       | Proses Data <i>Splitting</i> .....                                  | 32         |
| 3.4                         | Permodelan <i>XGBoost</i> , <i>CatBoost</i> , dan LGBM.....         | 34         |
| 3.4.1                       | Modeling <i>XGBoost</i> .....                                       | 34         |
| 3.4.2                       | Modeling <i>CatBoost</i> .....                                      | 44         |
| 3.4.3                       | Modeling LGBM.....                                                  | 50         |
| 3.5                         | Evaluasi Model.....                                                 | 55         |
| 3.5.1                       | Accuracy .....                                                      | 55         |
| 3.5.2                       | Precision, Recall dan F1-Score .....                                | 56         |
| 3.6                         | Skenario Pengujian.....                                             | 57         |
| 3.7                         | Hasil Penelitian .....                                              | 59         |
| <b>BAB IV</b>               | <b>PENGUJIAN DAN ANALISIS .....</b>                                 | <b>60</b>  |
| 4.1                         | Pengolahan Data.....                                                | 60         |
| 4.1.1                       | Data <i>Understanding</i> .....                                     | 60         |
| 4.1.2                       | Data <i>Preparation</i> .....                                       | 67         |
| 4.2                         | Pengujian Metode <i>XGBoost</i> .....                               | 74         |
| 4.2.1                       | <i>XGBoost</i> dengan <i>ClassWeight</i> .....                      | 81         |
| 4.2.2                       | <i>XGBoost</i> dengan Penggabungan Data.....                        | 83         |
| 4.2.3                       | <i>XGBoost</i> dengan Penggabungan Data + <i>ClassWeight</i> .....  | 86         |
| 4.3                         | Pengujian Metode <i>CatBoost</i> .....                              | 96         |
| 4.3.1                       | <i>CatBoost</i> dengan <i>ClassWeight</i> .....                     | 101        |
| 4.3.2                       | <i>CatBoost</i> dengan Penggabungan Data .....                      | 105        |
| 4.3.3                       | <i>CatBoost</i> dengan Penggabungan Data + <i>ClassWeight</i> ..... | 108        |
| 4.4                         | Pengujian Metode LGBM.....                                          | 118        |
| 4.4.1                       | LGBM dengan <i>ClassWeight</i> .....                                | 123        |
| 4.4.2                       | LGBM dengan Penggabungan Data .....                                 | 126        |
| 4.4.3                       | LGBM dengan <i>ClassWeight</i> +Penggabungan Data .....             | 128        |
| 4.3                         | Analisis Evaluasi Model .....                                       | 136        |
| <b>BAB V</b>                | <b>KESIMPULAN DAN SARAN .....</b>                                   | <b>142</b> |
| 5.1                         | Kesimpulan .....                                                    | 142        |
| 5.2                         | Saran.....                                                          | 144        |
| <b>DAFTAR PUSTAKA</b> ..... |                                                                     | <b>145</b> |

## DAFTAR GAMBAR

|                                                                                  |    |
|----------------------------------------------------------------------------------|----|
| Gambar 2. 1 Pohon keputusan <i>XGBoost</i> .....                                 | 11 |
| Gambar 2. 2 Pohon keputusan <i>CatBoost</i> .....                                | 14 |
| Gambar 2. 3 Pohon keputusan LGBM.....                                            | 17 |
| Gambar 3. 1 Alur Penelitian.....                                                 | 25 |
| Gambar 3. 2 Proses Pengolahan Data .....                                         | 27 |
| Gambar 3. 3 Proses Data Understanding.....                                       | 28 |
| Gambar 3. 4 Proses Data Preparation .....                                        | 31 |
| Gambar 3. 5 Proses Modeling <i>XGBoost</i> .....                                 | 34 |
| Gambar 3. 6 Proses modelling <i>CatBoost</i> .....                               | 44 |
| Gambar 3. 7 Proses modelling LGBM.....                                           | 50 |
| Gambar 3. 8 Skenario Pengujian.....                                              | 58 |
| Gambar 4. 1 Statistik data 2022.....                                             | 60 |
| Gambar 4. 2 Statistik data 2023 .....                                            | 61 |
| Gambar 4. 3 Statistik data 2024 .....                                            | 62 |
| Gambar 4. 4 Data <i>Null</i> 2022 .....                                          | 63 |
| Gambar 4. 5 Jumlah Data <i>Null</i> 2022.....                                    | 63 |
| Gambar 4. 6 Data <i>Null</i> 2023 .....                                          | 64 |
| Gambar 4. 7 Jumlah Data <i>Null</i> 2023 .....                                   | 64 |
| Gambar 4. 8 Data <i>Null</i> 2024 .....                                          | 64 |
| Gambar 4. 9 Jumlah Data <i>Null</i> 2024.....                                    | 65 |
| Gambar 4. 10 Data <i>Outlier</i> 2022.....                                       | 66 |
| Gambar 4. 11 Data <i>Outlier</i> 2023.....                                       | 66 |
| Gambar 4. 12 Data <i>Outlier</i> 2024.....                                       | 67 |
| Gambar 4. 13 Diagram nilai <i>Null</i> 2022.....                                 | 68 |
| Gambar 4. 14 Nilai data <i>null</i> 2022 .....                                   | 69 |
| Gambar 4. 15 Hasil pengolahan data <i>null</i> .....                             | 70 |
| Gambar 4. 16 Diagram nilai <i>null</i> 2023.....                                 | 71 |
| Gambar 4. 17 Diagram nilai <i>null</i> 2024.....                                 | 72 |
| Gambar 4. 18 Hasil Proses Encoding.....                                          | 74 |
| Gambar 4. 19 Confusion Matriks <i>XGBoost</i> LR 0.1   80:20   iterasi 100.....  | 79 |
| Gambar 4. 20 Evaluasi <i>XGBoost</i> LR 0.1   80:20   iterasi 100 .....          | 79 |
| Gambar 4. 21 Confusion Matriks <i>XGBoost</i> LR 0.01   70:30   iterasi 500..... | 80 |

|                                                                                                                            |     |
|----------------------------------------------------------------------------------------------------------------------------|-----|
| Gambar 4. 22 Evaluasi <i>XGBoost</i> LR 0.01   70:30   iterasi 500 .....                                                   | 80  |
| Gambar 4. 23 Confusion Matriks <i>XGBoost</i> Penggabungan Data LR 0.1   70:30   iterasi 500 .....                         | 85  |
| Gambar 4. 24 Evaluasi <i>XGBoost</i> Penggabungan Data LR 0.1   70:30   iterasi 500...                                     | 86  |
| Gambar 4. 25 Evaluasi <i>XGBoost ClassWeight</i> +Penggabungan Data LR 0.1   70:30   iterasi 500 .....                     | 88  |
| Gambar 4. 26 Confusion Matriks <i>XGBoost ClassWeight</i> +Penggabungan Data LR 0.1   .....                                | 89  |
| Gambar 4. 27 Perbandingan data XGBOOST .....                                                                               | 89  |
| Gambar 4. 28 Perbandingan Hasil Klasifikasi .....                                                                          | 90  |
| Gambar 4. 29 Presentasi hasil klasifikasi XGBOOST .....                                                                    | 90  |
| Gambar 4. 30 Hasil klasifikasi XGBOOST .....                                                                               | 91  |
| Gambar 4. 31 feature importance XGBOOST .....                                                                              | 92  |
| Gambar 4. 32 hasil klasifikasi prediksi .....                                                                              | 95  |
| Gambar 4. 33 Confusion Matriks <i>CatBoost</i> LR 0.1   80:20   iterasi 100 .....                                          | 100 |
| Gambar 4. 34 Evaluasi <i>CatBoost</i> LR 0.1   80:20   iterasi 100 .....                                                   | 101 |
| Gambar 4. 35 Confusion Matriks <i>CatBoost ClassWeight</i> LR 0.1   75:25   iterasi 100 .....                              | 104 |
| Gambar 4. 36 Evaluasi <i>CatBoost ClassWeight</i> LR 0.1   75:25   iterasi 100 .....                                       | 104 |
| Gambar 4. 37 Confusion Matriks <i>CatBoost</i> Penggabungan Kelas LR 0.1   80:20   iterasi 100 .....                       | 107 |
| Gambar 4. 38 Evaluasi <i>CatBoost</i> Penggabungan Kelas LR 0.1   80:20   iterasi 100                                      | 107 |
| Gambar 4. 39 Confusion Matriks <i>CatBoost ClassWeight</i> +Penggabungan Kelas LR 0.1   70:30   iterasi 100 .....          | 110 |
| Gambar 4. 40 Evaluasi Confusion Matriks <i>CatBoost ClassWeight</i> +Penggabungan Kelas LR 0.1   70:30   iterasi 100 ..... | 111 |
| Gambar 4. 41 perbandingan hasil catboost .....                                                                             | 112 |
| Gambar 4. 42 Presentase hasil catboost .....                                                                               | 113 |
| Gambar 4. 43 hasil klasifikasi catboost .....                                                                              | 114 |
| Gambar 4. 44 feature importance catboost .....                                                                             | 114 |
| Gambar 4. 45 hasil implementasi catboost .....                                                                             | 117 |
| Gambar 4. 46 Confusion Matriks LGBM LR 0.1   80:20   iterasi 100 .....                                                     | 122 |
| Gambar 4. 47 Evaluasi LGBM LR 0.1   80:20   iterasi 100 .....                                                              | 123 |

|                                                                                                                |     |
|----------------------------------------------------------------------------------------------------------------|-----|
| Gambar 4. 48 Confusion Matriks LGBM <i>ClassWeight</i> LR 0.1   80:20   iterasi 100                            | 125 |
| Gambar 4. 49 Confusion Matriks LGBM Penggabungan Kelas LR 0.1   80:20   iterasi 100.....                       | 127 |
| Gambar 4. 50 Evaluasi LGBM Penggabungan Kelas LR 0.1   80:20   iterasi 100 .                                   | 128 |
| Gambar 4. 51 Confusion Matriks LGBM <i>ClassWeight</i> +Penggabungan Kelas LR 0.01   70:30   iterasi 100 ..... | 130 |
| Gambar 4. 52 Evaluasi LGBM <i>ClassWeight</i> +Penggabungan Kelas LR 0.01   70:30   iterasi 100.....           | 130 |
| Gambar 4. 53 hasil klasifikasi LGBM .....                                                                      | 131 |
| Gambar 4. 54 perbandingan hasil LGBM.....                                                                      | 131 |
| Gambar 4. 55 presentase hasil LGBM .....                                                                       | 132 |
| Gambar 4. 56 feature importance LGBM .....                                                                     | 133 |
| Gambar 4. 57 hasil implementasi LGBM .....                                                                     | 135 |

## DAFTAR TABEL

|                                                                                               |    |
|-----------------------------------------------------------------------------------------------|----|
| Tabel 2. 1 Variabel Data .....                                                                | 9  |
| Tabel 3. 1 Sampel Data .....                                                                  | 26 |
| Tabel 3. 2 Statistik data sampel.....                                                         | 29 |
| Tabel 3. 3 Perbaikan data sampel.....                                                         | 31 |
| Tabel 3. 4 Encoding dan Kategori.....                                                         | 32 |
| Tabel 3. 5 Encoding dan Kategori.....                                                         | 32 |
| Tabel 3. 6 sampel data training .....                                                         | 33 |
| Tabel 3. 7 sampel data testing .....                                                          | 33 |
| Tabel 3. 8 sampel data <i>XGBoost</i> .....                                                   | 35 |
| Tabel 3. 9 pengubahan one hot encoding <i>XGboost</i> .....                                   | 35 |
| Tabel 3. 10 Threshold NO2.....                                                                | 37 |
| Tabel 3. 11 Gain NO2 .....                                                                    | 37 |
| Tabel 3. 12 Left Node NO2.....                                                                | 39 |
| Tabel 3. 13 Right Node NO2 .....                                                              | 39 |
| Tabel 3. 14 Hasil Klasifikasi <i>XGBoost</i> .....                                            | 41 |
| Tabel 3. 15 data sampel <i>CatBoost</i> .....                                                 | 44 |
| Tabel 3. 16 Threshold SO2 .....                                                               | 45 |
| Tabel 3. 17 Gradien SO2 <i>CatBoost</i> .....                                                 | 46 |
| Tabel 3. 18 Total Gradien .....                                                               | 47 |
| Tabel 3. 19 Hasil Klasifikasi <i>CatBoost</i> .....                                           | 49 |
| Tabel 3. 20 data sampel <i>LGBM</i> .....                                                     | 50 |
| Tabel 3. 21 Binning O3 <i>LGBM</i> .....                                                      | 51 |
| Tabel 3. 22 ISPU Klasifikasi .....                                                            | 55 |
| Tabel 3. 23 Nilai aktual dan Klasifikasi <i>XGBOOST</i> .....                                 | 56 |
| Tabel 3. 24 Penghitungan Precision, Recall dan F1-Score <i>XGBOOST</i> .....                  | 56 |
| Tabel 3. 25 Nilai aktual dan Klasifikasi <i>CatBoost</i> .....                                | 57 |
| Tabel 3. 26 Nilai aktual dan Klasifikasi <i>LGBM</i> .....                                    | 57 |
| Tabel 3. 27 Penghitungan Precision, Recall dan F1-Score <i>CatBoost</i> dan <i>LGBM</i> ..... | 57 |
| Tabel 3. 28 Skenario Pengujian .....                                                          | 58 |
| Tabel 4. 1 Hasil Pengolahan <i>XGBoost</i> .....                                              | 78 |
| Tabel 4. 2 Hasil Pengolahan <i>XGBoost ClassWeight</i> .....                                  | 82 |
| Tabel 4. 3 Hasil Pengolahan <i>XGBoost</i> Penggabungan Data .....                            | 84 |

|                                                                                   |     |
|-----------------------------------------------------------------------------------|-----|
| Tabel 4. 4 Hasil Pengolahan <i>XGBoost ClassWeight</i> +Penggabungan Data .....   | 86  |
| Tabel 4. 5 Hasil Pengolahan <i>CatBoost</i> .....                                 | 99  |
| Tabel 4. 6 Hasil Pengolahan <i>CatBoost ClassWeight</i> .....                     | 102 |
| Tabel 4. 7 Hasil Pengolahan <i>CatBoost</i> Penggabungan data .....               | 105 |
| Tabel 4. 8 Hasil Pengolahan <i>CatBoost ClassWeight</i> + Penggabungan Data ..... | 108 |
| Tabel 4. 9 Hasil Pengolahan LGBM .....                                            | 121 |
| Tabel 4. 10 Hasil Pengolahan LGBM <i>ClassWeight</i> .....                        | 124 |
| Tabel 4. 11 Hasil Pengolahan LGBM Penggabungan Data .....                         | 126 |
| Tabel 4. 12 Hasil Pengolahan LGBM <i>ClassWeight</i> +Penggabungan Data .....     | 128 |
| Tabel 4. 13 Pengujian Terbaik <i>XGBoost</i> .....                                | 136 |
| Tabel 4. 14 Pengujian Terbaik <i>XGBoost ClassWeight</i> +Penggabungan Data ..... | 137 |
| Tabel 4. 15 Pengujian Terbaik <i>CatBoost</i> .....                               | 138 |
| Tabel 4. 16 Pengujian Terbaik <i>CatBoost ClassWeight</i> +Penggabungan Data..... | 139 |
| Tabel 4. 17 Pengujian Terbaik LGBM.....                                           | 140 |
| Tabel 4. 18 Pengujian Terbaik LGBM <i>ClassWeight</i> +Penggabungan Data .....    | 140 |