



**SKRIPSI**

**ANALISIS PENGGUNAAN TEKNIK  
OVERSAMPLING PADA EXTREME GRADIENT  
BOOSTING (XGBOOST) UNTUK MENGATASI  
KETIDAKSEIMBANGAN KELAS PADA  
KLASIFIKASI PENYAKIT JANTUNG**

**MUHAMMAD RAFLI AULIA ROJANI LUTFI**

NPM: 20081010061

**DOSEN PEMBIMBING**

Achmad Junaidi, S.Kom., M.Kom.

Agung Mustika Rizki, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI  
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR  
FAKULTAS ILMU KOMPUTER  
PROGRAM STUDI INFORMATIKA  
SURABAYA  
2024**

*Halaman ini sengaja dikosongkan*

**LEMBAR PENGESAHAN**

**ANALISIS PENGGUNAAN TEKNIK OVERSAMPLING PADA  
EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK MENGATASI  
KETIDAKSEIMBANGAN KELAS PADA KLASIFIKASI  
PENYAKIT JANTUNG**

Oleh:  
**MUHAMMAD RAFLI AULIA ROJANI LUTFI**  
NPM. 20081010061

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika  
Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada Tanggal  
2 September 2024

**Menyetujui,**

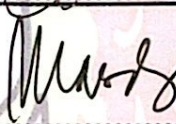
**Achmad Junaidi, S.Kom., M.Kom.**  
NPT. 3 7811 04 0199 1

  
..... (Pembimbing I)

**Agung Mustika Rizki, S.Kom., M.Kom.**  
NIP. 19930725 202203 1008

  
..... (Pembimbing II)


**Dr. Ir. I Gede Susrama Mas, ST. MT. IPU**  
NIP. 19700619 2021211 009

  
..... (Ketua Penguji)

**Retno Mumpuni, S.Kom., M.Sc.**  
NPT. 172198 70 716054

  
..... (Penguji I)

**Mengetahui,**  
**Dekan Fakultas Ilmu Komputer**

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

*Halaman ini sengaja dikosongkan*

**LEMBAR PERSETUJUAN**

**ANALISIS PENGGUNAAN TEKNIK OVERSAMPLING PADA  
EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK MENGATASI  
KETIDAKSEIMBANGAN KELAS PADA KLASIFIKASI PENYAKIT  
JANTUNG**

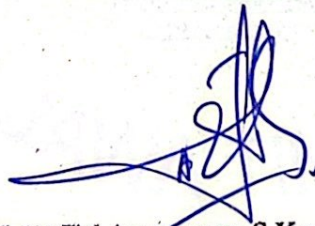
Oleh:

**MUHAMMAD RAFLI AULIA ROJANI LUTFI**

**NPM. 20081010061**

Menyetujui,

**Koordinator Program Studi Informatika  
Fakultas Ilmu Komputer**



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

**NIP. 19820211 2021212 005**



## SURAT PERNYATAAN ORISINALITAS

Yang bertandatangan di bawah ini:

Nama Mahasiswa : Muhammad Rafli Aulia Rojani Lutfi  
Program Studi : Informatika  
Dosen Pembimbing : 1. Achmad Junaidi, S.Kom., M.Kom.  
2. Agung Mustika Rizki, S.Kom., M.Kom.

dengan ini menyatakan bahwa isi sebagian maupun keseluruhan disertasi dengan judul:

### **ANALISIS PENGGUNAAN TEKNIK OVERSAMPLING PADA EXTREME GRADIENT BOOSTING (XGBOOST) UNTUK MENGATASI KETIDAKSEIMBANGAN KELAS PADA KLASIFIKASI PENYAKIT JANTUNG**

adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diizinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri. Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.



Surabaya, 10 September 2024  
Yang Membuat Pernyataan,



**M. RAFLI AULIA ROJANI LUTFI**  
NPM. 20081010061

*Halaman ini sengaja dikosongkan*

## ABSTRAK

Nama Mahasiswa / NPM : Muhammad Rafli Aulia Rojani Lutfi / 20081010061  
Judul Skripsi : Analisis Penggunaan Teknik Oversampling Pada Extreme Gradient Boosting (XGBoost) Untuk Mengatasi Ketidakseimbangan Kelas Pada Klasifikasi Penyakit Jantung  
Dosen Pembimbing : 1. Achmad Junaidi, S.Kom., M.Kom.  
2. Agung Mustika Rizki, S.Kom., M.Kom.

Penyakit jantung merupakan ancaman serius bagi kesehatan global. Deteksi dini menjadi kunci untuk meningkatkan angka keberlangsungan hidup. Namun, upaya membangun model prediksi seringkali terhambat oleh ketidakseimbangan data, di mana jumlah individu sehat jauh lebih banyak. Ketidakseimbangan ini berpotensi membuat model lebih condong pada kelas mayoritas (sehat) dan mengabaikan kelas minoritas (sakit), sehingga mengurangi akurasi dalam mendeteksi penyakit jantung.

Penelitian ini bertujuan untuk mengatasi masalah ketidakseimbangan kelas dengan menerapkan teknik *oversampling* pada algoritma XGBoost. Teknik *oversampling* bekerja dengan meningkatkan jumlah data pada kelas minoritas sehingga keseimbangan data dapat tercapai. Dalam penelitian ini, tiga teknik *oversampling* yang digunakan, yaitu ROS, SMOTE, dan ADASYN, diuji coba. Selain itu, dilakukan penyetelan hyperparameter pada algoritma XGBoost untuk mendapatkan kinerja model yang optimal.

Hasil penelitian menunjukkan bahwa model XGBoost tanpa *oversampling* yang menggunakan parameter bawaan memiliki akurasi tinggi yaitu di atas 0,90, namun kurang baik dalam mengklasifikasikan kelas minoritas (individu sakit), dibuktikan dengan nilai *g-mean* yang rendah yaitu di bawah 0,50. Sebaliknya, model XGBoost dengan teknik *oversampling* pada rasio *sampling* dan *hyperparameter* yang optimal mampu meningkatkan kinerja model dalam mengklasifikasikan kelas minoritas. Dari tiga teknik *oversampling*, ROS memberikan rata-rata nilai *g-mean* tertinggi sebesar 0,80. Namun, model ini rentan terhadap *overfitting* pada rasio *sampling* yang lebih besar. Berbeda dengan SMOTE dan ADASYN, meskipun kedua model juga rentan terhadap *overfitting* pada rasio *sampling* di atas 0,1, namun tingkat kerentanannya jauh lebih rendah dan cenderung lebih stabil.

**Kata kunci:** Penyakit Jantung, Ketidakseimbangan Kelas, Teknik *Oversampling*, XGBoost



*Halaman ini sengaja dikosongkan*

## ABSTRACT

Student Name / NPM : Muhammad Rafli Aulia Rojani Lutfi /  
20081010061  
Thesis Title : Analysis of Oversampling Techniques on Extreme  
Gradient Boosting (XGBoost) for Handling Class  
Imbalance in Heart Disease Classification  
Advisor : 1. Achmad Junaidi, S.Kom., M.Kom.  
2. Agung Mustika Rizki, S.Kom., M.Kom.

Heart disease is a serious threat to global health. Early detection is key to improving survival rates. However, building predictive models are often hampered by data imbalance, where the number of healthy individuals far exceeds the number of diseased individuals. This imbalance can cause the model to be biased towards the majority class (healthy) and neglect the minority class (diseased), reducing its accuracy in detecting heart disease.

This research aims to address the problem of class imbalance by applying oversampling techniques to the XGBoost algorithm. The oversampling technique works by increasing the amount of data in the minority class so that data balance can be achieved. In this study, three oversampling techniques, namely ROS, SMOTE, and ADASYN, were tested. In addition, hyperparameter tuning of the XGBoost algorithm was performed to obtain optimal model performance.

The results show that the XGBoost model without oversampling using default parameters has a high accuracy of above 0.90, but is not good at classifying minority classes (sick individuals), as evidenced by the low g-mean value of below 0.50. In contrast, the XGBoost model with oversampling techniques at optimal sampling ratios and hyperparameters is able to improve the performance of the model in classifying the minority class. Of the three oversampling techniques, ROS provides the highest average g-mean value of 0.80. However, the model is prone to overfitting at larger sampling ratios. In contrast to SMOTE and ADASYN, although both models are also susceptible to overfitting at sampling ratios above 0.1, the degree of susceptibility is much lower and tends to be more stable.

**Keywords:** Heart Disease, Imbalance Class, Oversampling Techniques, XGBoost

*Halaman ini sengaja dikosongkan*

## KATA PENGANTAR

Segala puji dan Syukur kehadirat Allah SWT yang telah melimpahkan rahmat-Nya, sehingga penulis dapat menyelesaikan penelitian skripsi beserta laporan hasil skripsi yang berjudul “Analisis Penggunaan Teknik Oversampling Pada Extreme Gradient Boosting (Xgboost) Untuk Mengatasi Ketidakseimbangan Kelas Pada Klasifikasi Penyakit Jantung” dengan baik.

Penyelesaian laporan skripsi ini tidak lepas dari dukungan dan bantuan dari berbagai pihak. Penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada semua pihak yang telah berkontribusi dalam kelancaran penelitian ini, mulai dari tahap perencanaan hingga penyusunan laporan akhir. Secara khusus, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, MMT. selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom. selaku Koordinator Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur
4. Bapak Achmad Junaidi, S.Kom., M.Kom. selaku Dosen Pembimbing I, yang telah membimbing penulis selama pengerjaan skripsi ini.
5. Bapak Agung Mustika Rizki, S.Kom., M.Kom. selaku Dosen Pembimbing II, yang telah membimbing penulis selama pengerjaan skripsi ini.
6. Seluruh Dosen dan Staff Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur atas segala bantuan yang diberikan untuk membantu penulis dalam penyusunan skripsi ini.
7. Ibunda dan Ayahanda tercinta yang tidak pernah putus mendoakan dan memberikan dukungan moril serta materi kepada penulis selama menempuh studi.
8. Teman-teman “Setunggal Riset” yang telah memberikan dukungan selama masa perkuliahan dan penyusunan skripsi ini.

Penulis menyadari bahwa skripsi ini masih penuh kekurangan, oleh karena itu saran dan kritik yang membangun sangat penulis harapkan. Besar harapan penulis bahwa dengan adanya penelitian ini dapat memberikan manfaat kepada para pembaca. Amin yaa Robbal ‘Alamin.

Surabaya, 25 Agustus 2024

Penulis

## DAFTAR ISI

|  |             |
|--|-------------|
| <b>LEMBAR JUDUL SKRIPSI</b> .....                                    | <b>i</b>    |
| <b>LEMBAR PENGESAHAN</b> .....                                       | <b>iii</b>  |
| <b>SURAT PERNYATAAN ORISINALITAS</b> .....                           | <b>v</b>    |
| <b>ABSTRAK</b> .....   | <b>vii</b>  |
| <b>KATA PENGANTAR</b> .....  | <b>xi</b>   |
| <b>DAFTAR ISI</b> .....  | <b>xiii</b> |
| <b>DAFTAR GAMBAR</b> .....   | <b>xvii</b> |
| <b>DAFTAR TABEL</b> .....  | <b>xix</b>  |
| <b>DAFTAR NOTASI</b> .....   | <b>xxi</b>  |
| <b>BAB I PENDAHULUAN</b> .....                                       | <b>1</b>    |
| 1.1 Latar Belakang .....   | 1           |
| 1.2 Rumusan Masalah .....  | 3           |
| 1.3 Tujuan Penelitian.....   | 3           |
| 1.4 Batasan Masalah.....   | 4           |
| 1.5 Manfaat Penelitian .....   | 4           |
| <b>BAB II KAJIAN PUSTAKA</b> .....                                   | <b>5</b>    |
| 2.1 Penelitian Terdahulu.....  | 5           |
| 2.2 Penyakit Jantung .....   | 7           |
| 2.3 <i>Data Mining</i> .....   | 8           |
| 2.4 Seleksi Fitur Chi-Square .....                                   | 10          |
| 2.5 <i>Imbalanced Data</i> .....                                     | 11          |
| 2.6 <i>Oversampling</i> .....  | 12          |
| 2.6.1 <i>Random Oversampling</i> .....                               | 13          |
| 2.6.2 <i>Synthetic Minority Oversampling Technique (SMOTE)</i> ..... | 13          |
| 2.6.3 <i>Adaptive Synthetic (ADASYN)</i> .....                       | 14          |
| 2.7 Extreme Gradient Boosting.....                                   | 16          |
| 2.8 GridSearchCV .....   | 18          |
| 2.9 <i>Confusion Matrix</i> .....                                    | 19          |
| <b>BAB III METODOLOGI</b> .....                                      | <b>23</b>   |
| 3.1 Tahapan Penelitian .....   | 23          |

|  |   |           |
|--|---|-----------|
| 3.2                                      | Studi Literatur.....                        | 23        |
| 3.3                                      | Analisis Kebutuhan .....                    | 23        |
| 3.3.1                                    | Spesifikasi Perangkat Keras .....           | 23        |
| 3.3.2                                    | Spesifikasi Perangkat Lunak.....            | 24        |
| 3.4                                      | Pengumpulan Data.....                       | 24        |
| 3.5                                      | <i>Preprocessing</i> .....                  | 27        |
| 3.6                                      | Pembuatan Model.....                        | 28        |
| 3.6.1                                    | Teknik <i>Oversampling</i> ROS.....         | 29        |
| 3.6.2                                    | Teknik <i>Oversampling</i> SMOTE .....      | 29        |
| 3.6.3                                    | Teknik <i>Oversampling</i> ADASYN .....     | 30        |
| 3.6.4                                    | Model XGBoost.....                          | 31        |
| 3.7                                      | Evaluasi Model.....                         | 32        |
| 3.8                                      | Analisis Hasil.....                         | 32        |
| 3.9                                      | Skenario Pengujian.....                     | 33        |
| <b>BAB IV HASIL DAN PEMBAHASAN .....</b> |   | <b>35</b> |
| 4.1                                      | <i>Preprocessing</i> .....                  | 35        |
| 4.1.1                                    | <i>Exploratory Data Analysis</i> (EDA)..... | 35        |
| 4.1.2                                    | Pembersihan Data .....                      | 41        |
| 4.1.3                                    | Seleksi Fitur .....                         | 44        |
| 4.1.4                                    | Transformasi Data.....                      | 46        |
| 4.1.5                                    | Pemisahan Data .....                        | 50        |
| 4.2                                      | <i>Oversampling</i> .....                   | 51        |
| 4.2.1                                    | ROS.....                                    | 51        |
| 4.2.2                                    | SMOTE.....                                  | 52        |
| 4.2.3                                    | ADASYN.....                                 | 53        |
| 4.3                                      | Pembuatan Model.....                        | 54        |
| 4.3.1                                    | Model XGBoost.....                          | 54        |
| 4.3.2                                    | Model XGBoost dengan ROS .....              | 54        |
| 4.3.3                                    | Model XGBoost dengan SMOTE.....             | 56        |
| 4.3.4                                    | Model XGBoost dengan ADASYN.....            | 57        |
| 4.4                                      | Evaluasi Model.....                         | 59        |
| 4.5                                      | Skenario Pengujian.....                     | 64        |



|   |                                  |            |
|---|----------------------------------|------------|
| 4.5.1                                   | Perubahan Rasio Data 70:30 ..... | 65         |
| 4.5.2                                   | Perubahan Rasio Data 80:20 ..... | 67         |
| 4.5.3                                   | Perubahan Rasio Data 90:10 ..... | 70         |
| 4.6                                     | Analisis Hasil .....             | 72         |
| <b>BAB V KESIMPULAN DAN SARAN .....</b> |                                  | <b>78</b>  |
| 5.1                                     | Kesimpulan .....                 | 78         |
| 5.2                                     | Saran.....                       | 78         |
| <b>DAFTAR PUSTAKA.....</b>              |                                  | <b>xxv</b> |
| <b>LAMPIRAN.....</b>                    |                                  | <b>xxx</b> |

*Halaman ini sengaja dikosongkan*

## DAFTAR GAMBAR

|  |    |
|--|----|
| Gambar 2.1 Ilustrasi Teknik <i>Oversampling</i> .....                      | 12 |
| Gambar 2.2 Ilustrasi SMOTE .....   | 13 |
| Gambar 2.3 Ilustrasi <i>Boosting</i> .....                                 | 16 |
| Gambar 3.1 Flowchart alur penelitian.....                                  | 23 |
| Gambar 3.2 Flowchart <i>Preprocessing</i> .....                            | 27 |
| Gambar 3.3 Flowchart Pembuatan Model .....                                 | 28 |
| Gambar 3.4 Flowchart Teknik <i>Oversampling</i> ROS .....                  | 29 |
| Gambar 3.5 Flowchart Teknik <i>Oversampling</i> SMOTE .....                | 30 |
| Gambar 3.6 Flowchart Teknik <i>Oversampling</i> ADASYN .....               | 30 |
| Gambar 4.1 Struktur dan Konten Dataset Penyakit Jantung.....               | 36 |
| Gambar 4.2 Persentase Penyakit Jantung.....                                | 37 |
| Gambar 4.3 Distribusi Penyakit Jantung Berdasarkan Jenis Kelamin.....      | 37 |
| Gambar 4.4 Distribusi Penyakit Jantung Berdasarkan Kelompok Usia .....     | 38 |
| Gambar 4.5 Distribusi Penyakit Jantung Berdasarkan Kondisi Kesehatan ..... | 38 |
| Gambar 4.6 Distribusi Penyakit Jantung Berdasarkan Aktivitas Fisik .....   | 39 |
| Gambar 4.7 Distribusi Penyakit Jantung Berdasarkan Penyakit Diabetes.....  | 40 |
| Gambar 4.8 Distribusi Penyakit Jantung Berdasarkan Penyakit Stroke .....   | 40 |
| Gambar 4.9 Distribusi Penyakit Jantung Berdasarkan Status Merokok .....    | 41 |
| Gambar 4.10 Jumlah <i>Missing Value</i> .....                              | 42 |
| Gambar 4.11 Sampel Data Sebelum Dilakukan Pembersihan Data.....            | 43 |
| Gambar 4.12 Sampel Data Setelah Dilakukan Pembersihan Data.....            | 43 |
| Gambar 4.13 Hasil Akhir Pembersihan Data .....                             | 44 |
| Gambar 4.14 Hasil Akhir Seleksi Fitur.....                                 | 46 |
| Gambar 4.15 Hasil Akhir Transformasi Data.....                             | 50 |
| Gambar 4.16 Contoh Hasil ROS.....  | 51 |
| Gambar 4.17 Contoh Hasil SMOTE .....                                       | 52 |
| Gambar 4.18 Contoh Hasil ADASYN .....                                      | 53 |
| Gambar 4.19 <i>Hyperparameter Tuning</i> Model XGBoost dengan ROS .....    | 56 |
| Gambar 4.20 <i>Hyperparameter Tuning</i> Model XGBoost dengan SMOTE .....  | 57 |
| Gambar 4.21 <i>Hyperparameter Tuning</i> Model XGBoost dengan ADASYN.....  | 58 |
| Gambar 4.22 Hasil Evaluasi Model XGBoost Tanpa <i>Oversampling</i> .....   | 60 |

|   |    |
|---|----|
| Gambar 4.23 <i>Confusion Matrix</i> Model XGBoost Tanpa <i>Oversampling</i> ..... | 61 |
| Gambar 4.24 Hasil Evaluasi Model XGBoost dengan ROS .....                         | 61 |
| Gambar 4.25 <i>Confusion Matrix</i> Model XGBoost dengan ROS .....                | 62 |
| Gambar 4.26 Hasil Evaluasi Model XGBoost dengan SMOTE .....                       | 62 |
| Gambar 4.27 <i>Confusion Matrix</i> Model XGBoost dengan SMOTE .....              | 63 |
| Gambar 4.28 Hasil Evaluasi Model XGBoost dengan ADASYN.....                       | 63 |
| Gambar 4.29 <i>Confusion Matrix</i> Model XGBoost dengan ADASYN .....             | 64 |
| Gambar 4.30 Grafik Model XGBoost Tanpa <i>Oversampling</i> .....                  | 73 |
| Gambar 4.31 Grafik Model XGBoost dengan ROS .....                                 | 74 |
| Gambar 4.32 Grafik Model XGBoost dengan SMOTE .....                               | 75 |
| Gambar 4.33 Grafik Model XGBoost dengan ADASYN .....                              | 76 |

## DAFTAR TABEL

|  |    |
|--|----|
| Tabel 2.1 Tabel Kontigensi.....  | 10 |
| Tabel 2.2 Tabel <i>Confusion Matrix</i> .....  | 19 |
| Tabel 3.1 Spesifikasi Perangkat Keras .....  | 24 |
| Tabel 3.2 Spesifikasi Perangkat Lunak .....  | 24 |
| Tabel 3.3 Rincian Nama Atribut Dataset.....  | 25 |
| Tabel 3.4 <i>Hyperparameter</i> yang Dioptimasi.....                                     | 31 |
| Tabel 4.1 Nilai <i>Chi-Square</i> dan <i>P-value</i> dari Setiap Kolom Kategorikal ..... | 45 |
| Tabel 4.2 Sampel Data Sebelum dan Sesudah <i>Binary Encoding</i> .....                   | 47 |
| Tabel 4.3 Sampel Data Sebelum dan Sesudah <i>One-Hot Encoding</i> .....                  | 48 |
| Tabel 4.4 Sampel Data Sebelum dan Sesudah <i>Ordinal Encoding</i> .....                  | 49 |
| Tabel 4.5 Hasil Evaluasi Data Pelatihan Pada Perubahan Rasio Data 70:30.....             | 65 |
| Tabel 4.6 Hasil Evaluasi Data Pengujian Pada Perubahan Rasio Data 70:30.....             | 66 |
| Tabel 4.7 Hasil Evaluasi Data Pelatihan Pada Perubahan Rasio Data 80:20.....             | 67 |
| Tabel 4.8 Hasil Evaluasi Data Pengujian Pada Perubahan Rasio Data 80:20.....             | 68 |
| Tabel 4.9 Hasil Evaluasi Data Pelatihan Pada Perubahan Rasio Data 90:10.....             | 70 |
| Tabel 4.10 Hasil Evaluasi Data Pengujian Pada Perubahan Rasio Data 90:10.....            | 71 |
| Tabel 4.11 Hasil Perhitungan Waktu Eksekusi.....   | 72 |

*Halaman ini sengaja dikosongkan*

## DAFTAR NOTASI

|                     |   |  |
|---------------------|---|--|
| $X^2$               | : | nilai <i>Chi-Square</i>  |
| $f_o$               | : | Frekuensi observasi  |
| $f_e$               | : | Frekuensi ekspektasi   |
| $b_i$               | : | jumlah nilai baris ke-i  |
| $k_j$               | : | jumlah nilai kolom ke-j  |
| N                   | : | total seluruh nilai observasi                                      |
| $d(x, y)$           | : | jarak data x dan y   |
| n                   | : | jumlah dimensi (atribut)   |
| $x_i$               | : | atribut ke i dari data x   |
| $y_i$               | : | atribut ke i dari data y   |
| $x_{syn}$           | : | data sintetis  |
| $x_{knn}$           | : | tetangga terdekat dari $x_i$                                       |
| $\gamma$            | : | bilangan desimal acak antara 0 dan 1                               |
| d                   | : | <i>degree of class imbalance</i>                                   |
| $m_s$               | : | jumlah kelas minoritas   |
| $m_l$               | : | jumlah kelas mayoritas   |
| $\beta$             | : | tingkat keseimbangan data sintetis                                 |
| $\Delta_i$          | : | jumlah data mayoritas pada K tetangga terdekat dari data minoritas |
| $x_i$               | : | titik data kelas minoritas asli                                    |
| $x_{zi}$            | : | titik data tetangga terdekat                                       |
| $\lambda$           | : | bilangan desimal acak antara 0 dan 1                               |
| $\hat{y}_i$         | : | jumlah daun pada pohon keputusan                                   |
| T                   | : | parameter regularisasi ke-1  |
| t                   | : | iterasi ke-t   |
| $\hat{y}_i^{(t-1)}$ | : | prediksi model pada iterasi sebelumnya                             |
| $f_t(x_i)$          | : | prediksi pohon keputusan ke-t untuk sampel data ke-i               |



*Halaman ini sengaja dikosongkan*