



SKRIPSI

**IMPLEMENTASI METODE BALANCING
DATA DENGAN TEKNIK SMOTETOMEK
DALAM KLASIFIKASI PENYAKIT DIABETES
MENGUNAKAN ALGORITMA XGBOOST**

FATWA RATANTJA KUSUMAJATI
NPM 20081010087

DOSEN PEMBIMBING

Dr. Basuki Rahmat, S.Si. MT.
Achmad Junaidi, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

**IMPLEMENTASI METODE BALANCING
DATA DENGAN TEKNIK SMOTETOMEK
DALAM KLASIFIKASI PENYAKIT DIABETES
MENGUNAKAN ALGORITMA XGBOOST**

FATWA RATANTJA KUSUMAJATI
NPM 20081010087

DOSEN PEMBIMBING
Dr. Basuki Rahmat, S.Si, MT.
Achmad Junaidi, S.Kom, M.Kom.

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN
IMPLEMENTASI METODE BALANCING DATA DENGAN TEKNIK
SMOTETOMEK DALAM KLASIFIKASI PENYAKIT DIABETES
MENGGUNAKAN ALGORITMA XGBOOST

Oleh :

Fatwa Ratantja Kusumajati

NPM. 20081010087

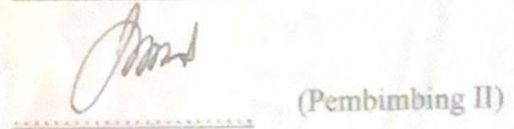
Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika
Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal
5 Maret 2025

Menyetujui

Dr. Basuki Rahmat, S.Si, MT.
NIP. 19820211 2021212 005


..... (Pembimbing I)

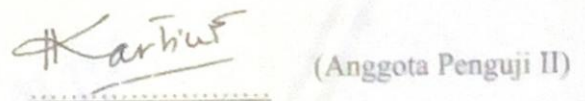
Achmad Junaidi, S.Kom., M.Kom.
NPT. 3 7811 04 0199 1


..... (Pembimbing II)

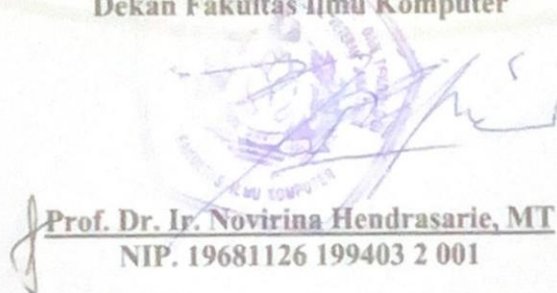
Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
NPT. 222198 60 816400


..... (Ketua Penguji)

Dr. Ir. Kartini, S.Kom. MT.
NIP. 19611110 199103 2 001


..... (Anggota 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 METODE BALANCING DATA DENGAN TEKNIK SMOTETOMEK DALAM
KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA XGBOOST

Oleh : Fatwa Ratantja Kusumajati

NPM. 20081010084



Koorrdinator Prodi Informatika
Fakultas Ilmu Komputer

Fetty Tri Anggraeny, S.Kom. M. Kom

NIP. 19820211 2021212 005

SURAT PERNYATAAN BEBAS PLAGIASI

Yang bertanda tangan di bawah ini:

Nama mahasiswa / NPM : Fatwa Ratantja Kusumajati / 20081010087
Program Studi : Informatika
Dosen Pembimbing : 1. Dr. Basuki Rahmat, S.Si, MT.
2. Achmad Junaidi, S.Kom., M.Kom.

dengan ini menyatakan bahwa Skripsi dengan judul "IMPLEMENTASI METODE BALANCING DATA DENGAN TEKNIK *SMOTETOMEK* DALAM KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA *XGBOOST*" adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Universitas Pembangunan Nasional Veteran Jawa Timur.



Surabaya, 7 Maret 2025
Mahasiswa,



Fatwa Ratantja Kusumajati
NPM. 20081010087

Surat Keterangan Hasil Pemeriksaan Turnitin
Fakultas Ilmu Komputer
Program Studi Informatika

Nama Mahasiswa: Fatwa Ratantja Kusumajati

NPM: 20081010087

Judul Penelitian: IMPLEMENTASI METODE BALANCING DATA DENGAN TEKNIK
SMOTETOMEK DALAM KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN
ALGORITMA XGBOOST

Program Studi: Informatika

19% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

• Bibliography

Top Sources

17%  Internet sources

10%  Publications

6%  Submitted works (Student Papers)

Dokumen telah diverifikasi oleh admin prodi Informatika.

Dokumen lengkap hasil Turnitin, dikirim melalui email resmi mahasiswa. Silahkan menghubungi admin prodi untuk informasi lebih lanjut.

Surabaya, 20 May 2025



Aries Satriya Yudha Santoso, S.Kom

181 19860824 092

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Kami mengucapkan terima kasih kepada Allah SWT, yang telah memberi penulis kemampuan untuk menyelesaikan laporan skripsi yang berjudul “IMPLEMENTASI METODE BALANCING DATA DENGAN TEKNIK SMOTETOMEK DALAM KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA XGBOOST” ini tepat pada waktunya.

Sebagai salah satu bentuk pertanggungjawaban akademik setelah menyelesaikan studi sarjana, laporan ini disusun oleh penulis dengan melakukan penelitian teoritis dan implementasi teknis yang mendalam. Studi literatur, analisis data, dan penerapan teknik balancing data menggunakan metode *SMOTETomek* adalah beberapa langkah dalam proses penelitian. Terakhir, algoritma *XGBoost* digunakan untuk mengevaluasi kinerja model.

Penulis menyadari bahwa laporan ini masih mempunyai berbagai kekurangan dan belum sepenuhnya sempurna. Dengan demikian, saran yang membangun dan kritik yang konstruktif sangat diharapkan guna peningkatan di masa yang akan datang. Semoga penelitian ini bermanfaat bagi dunia akademik dan menjadi referensi untuk penggunaan metode balancing data di bidang pembelajaran mesin dan kesehatan.

Surabaya, 5 Maret 2025

Penulis,



Fatwa Ratantja Kusumajati

NPM. 20081010087

Halaman ini sengaja dikosongkan

UCAPAN TERIMA KASIH

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas karunia-Nya saya dapat menyelesaikan Laporan Skripsi dengan judul “IMPLEMENTASI METODE BALANCING DATA DENGAN TEKNIK SMOTETOMEK DALAM KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA XGBOOST”. Laporan ini merupakan bentuk tertulis dari pertanggungjawaban atas pelaksanaan kuliah selama 5 tahun. Dalam penyusunan laporan ini, saya menyadari bahwa banyak pihak yang terlibat. Oleh karena itu, saya ingin berterima kasih kepada pihak-pihak yang turut serta dalam pengerjaan laporan skripsi ini, yaitu :

1. Kedua orang tua, keluarga, dan teman-teman terdekat yang selalu memberikan dukungan dan doa.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Dr. Novirina Hendrasarie, S.T., M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom. selaku Koordinator Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Bapak Dr. Basuki Rahmat, S.Si. MT. selaku dosen pembimbing 1 saya yang memberikan motivasi, bimbingan, masukan, dan saran dalam penyusunan laporan skripsi ini.
6. Bapak Achmad Junaidi, S.Kom, M.Kom selaku pembimbing 2 yang bersedia membimbing dan memberikan pengarahan selama mengerjakan laporan skripsi.
7. Ibu Dr. Eng. Ir. Anggraeni Puspita Sari, ST., MT selaku penguji 1 yang telah menguji dan membina saya pasca ujian skripsi.
8. Ibu Dr. Ir. Kartini, S.Kom. MT selaku penguji 2 yang telah menguji dan membina saya pasca ujian skripsi.
9. Seluruh civitas akademika UPN “Veteran” Jawa Timur yang telah memberi support serta dukungan dalam mengerjakan laporan skripsi.

10. Keluarga Besar MABES LANGGENG yang selalu memberikan dukungan dan doa kepada saya untuk segera menyelesaikan laporan skripsi.
11. Bapak Joko dan Ibu Tutik selaku penjaga warkop Langgeng yang selalu menemani saya dalam proses pengerjaan laporan skripsi selama ini.
12. Seluruh fungsionaris Badan Eksekutif Mahasiswa (BEM) FASILKOM Kabinet Integral dan Aerial karena telah memberikan motivasi serta dukungan dalam penulisan laporan skripsi.

Sebagai penulis, saya menyadari bahwa laporan ini memiliki banyak kekurangan dan tidak sempurna secara keseluruhan. Oleh karena itu, sebagai evaluasi saya saat menyusun laporan atau karya tulis serupa di masa depan, saya mengharapkan kritik dan saran yang bermanfaat dari para pembaca. Akhir kata, kami berharap laporan ini memberikan pengetahuan baru dan bermanfaat bagi pembaca.

Surabaya, 5 Maret 2025
Penulis,



Fatwa Ratantja Kusumajati
NPM. 20081010087

ABSTRAK

Nama Mahasiswa / NPM : Fatwa Ratantja Kusumajati / 20081010087
Judul Skripsi : IMPLEMENTASI METODE BALANCING
DATA DENGAN TEKNIK SMOTETOMEK
DALAM KLASIFIKASI PENYAKIT DIABETES
MENGUNAKAN ALGORITMA XGBOOST
Dosen Pembimbing : 1. Dr. Basuki Rahmat, S.Si, MT.
2. Achmad Junaidi, S.Kom., M.Kom.

Dalam penelitian ini, algoritma *XGBoost* dan pendekatan *SMOTETomek* digunakan untuk meningkatkan akurasi klasifikasi diabetes dengan memanfaatkan 2.000 titik data pasien yang diperoleh dari Kaggle, mencakup berbagai variabel demografis dan medis. Diabetes merupakan masalah kesehatan masyarakat yang signifikan di Indonesia, namun tantangan utama dalam penelitian ini adalah ketidakseimbangan kelas, di mana jumlah sampel non-diabetes jauh lebih banyak *dibandingkan* dengan sampel diabetes. Untuk mengatasi hal ini, digunakan metode *SMOTETomek*, yang menggabungkan *SMOTE* (*Synthetic Minority Over-sampling Technique*) dan *Tomek Links* untuk menyetabilkan distribusi kelas dengan menambahkan sampel dari kelas minoritas dan menghapus sampel yang berada di luar batas. Hasil penelitian menunjukkan bahwa pendekatan *SMOTETomek* mampu meningkatkan akurasi model *XGBoost* hingga 95,01%, lebih tinggi dibandingkan dengan penggunaan *SMOTE* atau data asli yang hanya mencapai 92,13%, meskipun sedikit mengurangi akurasi pada kelas minoritas (non-diabetes). Hal ini mengonfirmasi bahwa kombinasi *SMOTE* dan *Tomek Links* dapat meningkatkan kinerja klasifikasi dalam kasus ketidakseimbangan data.

Kata Kunci :SMOTETomek, XGBOOST, Diabetes, Klasifikasi, Balancing Data.

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM : Fatwa Ratantja Kusumajati / 20081010087
Thesis Title : IMPLEMENTASI METODE BALANCING
DATA DENGAN TEKNIK SMOTETOMEK
DALAM KLASIFIKASI PENYAKIT DIABETES
MENGUNAKAN ALGORITMA XGBOOST
Advisor : 1. Dr. Basuki Rahmat, S.Si, MT.
2. Achmad Junaidi, S.Kom., M.Kom.

This work uses 2,000 patient data points from Kaggle, including demographic and medical information, to improve the accuracy of diabetes categorization using the XGBoost algorithm and the SMOTETomek method. In Indonesia, diabetes is a serious public health issue. One of the main problems with this study is the unbalanced dataset, which has more non-diabetic samples than diabetic ones. In order to balance class distributions, the SMOTETomek approach is used, which combines Tomek Links with SMOTE (Synthetic Minority Over-sampling Technique) to oversample the minority class and eliminate borderline data. While minority class accuracy is marginally decreased (0.97 vs. 0.99 for SMOTE and original data), the results show that SMOTETomek increases XGBoost accuracy to 95.01%, surpassing both SMOTE and the original data (92.13%). Nonetheless, it continues to retain a high F1-score and accuracy, demonstrating its ability to manage data imbalance with few compromises.

Keywords: SMOTETomek, XGBOOST, Diabetes Classification, Balancing Data.

Halaman ini sengaja dikosongkan

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN	v
SURAT PERNYATAAN ORISINALITAS	vi
KATA PENGANTAR.....	ix
UCAPAN TERIMA KASIH	xii
ABSTRAK	xiv
DAFTAR ISI.....	xviii
DAFTAR TABEL	xxi
DAFTAR GAMBAR.....	xxiii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah	4
1.3. Tujuan Penelitian	4
1.4. Manfaat Penelitian.....	5
1.5. Batasan Masalah	5
BAB II TINJAUAN PUSTAKA	7
2.1. Penelitian Terdahulu	7
2.2. Diabetes.....	8
2.3. <i>Machine Learning</i>	10
2.4. <i>Supervised Learning</i>	11
2.5. Klasifikasi	13
2.6. <i>Extreme Gradient Boosting (XGBoost)</i>	15
2.7. <i>SMOTE</i>	17
2.8. <i>Tomek-Links</i>	18
2.9. <i>SMOTETomek</i>	20
2.10 . Confussion Matrix	22
BAB III METODOLOGI	25
3.1. Akuisisi Data.....	26
3.2. <i>Pre Processing Data</i>	28
3.3. <i>Cleaning Data</i>	28

3.4.	Normalisasi Data	30
3.5.	Memisahkan Data Fitur dan Target.....	32
3.6.	<i>Balancing Data</i>	33
3.7.	<i>Splitting Data</i>	35
3.8.	Membangun Model	36
3.9.	Skema Pengujian	38
3.10.	Evaluasi Model	39
BAB IV HASIL DAN PEMBAHASAN.....		43
4.1	Implementasi Program	43
4.1.1	Library	43
4.1.2	Akuisisi Data	44
4.1.3	Preprocessing Data.....	46
4.1.4	Membangun Model	52
4.1.5	Evaluasi Model	56
4.2	Perbandingan Skema Pengujian.....	60
4.2.1	Perbandingan Balancing Data	60
4.2.2	Perbandingan Hasil Akurasi Model	62
BAB V KESIMPULAN DAN SARAN		125
5.1	Kesimpulan	125
5.2	Saran.....	126
DAFTAR PUSTAKA		127
LAMPIRAN.....		133

Halaman ini sengaja dikosongkan

DAFTAR TABEL

Tabel 3. 1 Data yang akan digunakan.....	27
Tabel 3. 2 Jumlah 0 yang terdapat pada data.....	29
Tabel 3. 3 Data setelah menghapus fitur yang tidak diperlukan.....	29
Tabel 3. 4 Hasil menghapus fitur.....	30
Tabel 3. 5 Hasil normalisasi data.....	31
Tabel 3. 6 Variable X (Data Fitur)	32
Tabel 3. 7 Contoh Hasil SMOTETomek	35
Tabel 3.8 Skema Pengujian	38
Tabel 4.1 Dataset	44
Tabel 4.2 Data setelah menghapus beberapa fitur	46
Tabel 4.3 Setelah Mengubah Nilai 0	48
Tabel 4.4 Hasil Normalisasi Data	48
Tabel 4.5 K-Cross Validation	53
Tabel 4.6 Data setelah Balancing Data.....	62
Tabel 4.7 Hasil akurasi K-Cross Validation	62
Tabel 4.8 N-Estimator optimal Train : Test = 80 : 20 dan n_estimator = 100.....	68
Tabel 4. 9 Perbandingan Hasil Akurasi Original Data	87
Tabel 4. 10 Perbandingan Hasil Akurasi SMOTE	105
Tabel 4. 11 Perbandingan Hasil Akurasi SMOTETOMEK	123
Tabel 4. 12 Perbandingan Hasil Akhir Tiap Data.....	124

DAFTAR GAMBAR

Gambar 2. 1 Diabetes	9
Gambar 2. 2 Machine Learning	10
Gambar 2. 3 Supervised Learning	12
Gambar 2. 4 Klasifikasi	13
Gambar 2.5 XGBoost	15
Gambar 2.6 SMOTE	17
Gambar 2.7 <i>Tomek-Links</i>	20
Gambar 2.8 SMOTETOMEK.....	22
Gambar 2.9 Confusion Matrix.....	24
Gambar 3.1 Metodologi.....	25
Gambar 3. 2 Pre Processing Data	28
Gambar 3. 3 Variable Y (Data Target)	32
Gambar 3. 4 Proses Balancing Data	34
Gambar 3. 5 Contoh Hasil Pembagian Data Latih dan Data Uji	35
Gambar 3. 6 Membangun Model XGBoost.....	36
Gambar 3. 7 Klasifikasi XGBoost.....	37
Gambar 3. 8 Contoh Confusion Matrix	40
Gambar 4.1 Jumlah nilai 0 yang teridentifikasi	47
Gambar 4.2 Hasil visualisasi data sebelum SMOTETomek.....	50
Gambar 4.3 Setelah SMOTETomek.....	51
Gambar 4.4 Pohon XGB SMOTETomek	55
Gambar 4.5 Akurasi dan MSE.....	58
Gambar 4.6 Confussion Matriks.....	59
Gambar 4.7 Hasil evaluasi model	60
Gambar 4.8 Data sebelum balancing data.....	61
Gambar 4.9 Data sesudah balancing data	61
Gambar 4.10 Pohon Original Data	64
Gambar 4.11 Pohon SMOTE.....	64
Gambar 4.12 Pohon SMOTETomek	65
Gambar 4.13 MSE Original Data	68

Gambar 4.14 MSE SMOTE	68
Gambar 4.15 MSE SMOTETomek	68
Gambar 4. 16 Confussion Matrix Skema Pengujian 1	72
Gambar 4. 17 Classification Report Skema Pengujian 1	73
Gambar 4. 18 Confusion Matrix Skema Pengujian 2.....	73
Gambar 4. 19 Classification Report Skema Pengujian 2	74
Gambar 4. 20 Confusion Matrix Skema Pengujian 3.....	75
Gambar 4. 21 Classification Report Skema Pengujian 3	76
Gambar 4. 22 Confusion Matrix Skema Pengujian 4.....	77
Gambar 4. 23 Classification Report Skema Pengujian 4	78
Gambar 4. 24 Confusion Matrix Skema Pengujian 5.....	79
Gambar 4. 25 Classification Report Skema Pengujian 5	80
Gambar 4. 26 Confusion Matrix Skema Pengujian 6.....	80
Gambar 4. 27 Classification Report Skema Pengujian 6	81
Gambar 4. 28 Confusion Matrix Skema Pengujian 7.....	82
Gambar 4. 29 Classification Report Skema Pengujian 7	83
Gambar 4. 30 Confusion Matrix Skema Pengujian 8.....	84
Gambar 4. 31 Classification Report Skema Pengujian 8	85
Gambar 4. 32 Confusion Matrix Skema Pengujian 9.....	86
Gambar 4. 33 Classification Report Skema Pengujian 9	87
Gambar 4. 34 Confusion Matrix Skema Pengujian 10.....	88
Gambar 4. 35 Confusion Matrix Skema Pengujian 10.....	89
Gambar 4. 36 Confusion Matrix Skema Pengujian 11.....	90
Gambar 4. 37 Confusion Matrix Skema Pengujian 11.....	91
Gambar 4. 38 Confusion Matrix Skema Pengujian 12.....	92
Gambar 4. 39 Confusion Matrix Skema Pengujian 12.....	93
Gambar 4. 40 Confusion Matrix Skema Pengujian 13.....	94
Gambar 4. 41 Classification Report Skema Pengujian 13	95
Gambar 4. 42 Confusion Matrix Skema Pengujian 14.....	96
Gambar 4. 43 Classification Report Skema Pengujian 14	97
Gambar 4. 44 Confusion Matrix Skema Pengujian 15.....	98
Gambar 4. 45 Classification Report Skema Pengujian 15	99

Gambar 4. 46	Confusion Matrix Skema Pengujian 16.....	100
Gambar 4. 47	Classification Skema Pengujian 16.....	101
Gambar 4. 48	Confusion Matrix Skema Pengujian 17.....	102
Gambar 4. 49	Classification Report Skema Pengujian 17.....	103
Gambar 4. 50	Confusion Matrix Skema Pengujian 18.....	104
Gambar 4. 51	Classification Report Skema Pengujian 18.....	104
Gambar 4. 52	Confusion Matrix Skema Pengujian 19.....	106
Gambar 4. 53	Classification Report Skema Pengujian 19.....	107
Gambar 4. 54	Confusion Matrix Skema Pengujian 20.....	108
Gambar 4. 55	Classification Report Skema Pengujian 20.....	109
Gambar 4. 56	Confusion Matrix Skema Pengujian 21	109
Gambar 4. 57	Classification Report Skema Pengujian 21	110
Gambar 4. 58	Confusion Matrix Skema Pengujian 22.....	111
Gambar 4. 59	Classification Report Skema Pengujian 22.....	112
Gambar 4. 60	Confusion Matrix Skema Pengujian 23.....	113
Gambar 4. 61	Classification Report Skema Pengujian 23.....	114
Gambar 4. 62	Confusion Matrix Skema Pengujian 24.....	115
Gambar 4. 63	Classification Report Skema Pengujian 24.....	116
Gambar 4. 64	Confusion Matrix Skema Pengujian 25.....	117
Gambar 4. 65	Classification Report Skema Pengujian 25.....	118
Gambar 4. 66	Confusion Matrix Skema Pengujian 26.....	119
Gambar 4. 67	Classification Report Skema Pengujian 26.....	120
Gambar 4. 68	Confusion Matrix Skema Pengujian 27	121
Gambar 4. 69	Classification Report Skema Pengujian 27.....	122

