

**IMPLEMENTASI METODE OVERSAMPLING UNTUK
MENINGKATKAN KINERJA ALGORITMA KLASIFIKASI PADA
KASUS IMBALANCED DATASET**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Galih Adhi Kuncoro Rosyad

17.11.1440

Kepada

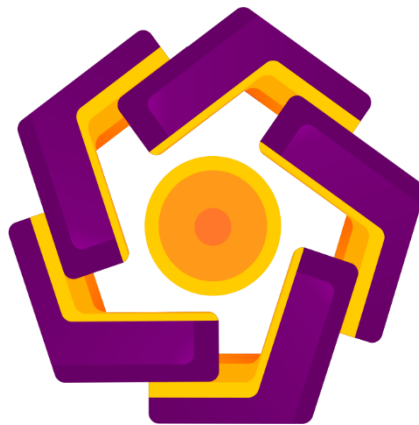
**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

**IMPLEMENTASI METODE OVERSAMPLING UNTUK
MENINGKATKAN KINERJA ALGORITMA KLASIFIKASI PADA
KASUS IMBALANCED DATASET**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

Galih Adhi Kuncoro Rosyad

17.11.1440

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

HALAMAN PERSETUJUAN

SKRIPSI

**IMPLEMENTASI METODE OVERSAMPLING UNTUK
MENINGKATKAN KINERJA ALGORITMA KLASIFIKASI PADA
KASUS IMBALANCED DATASET**

yang disusun dan diajukan oleh
Galih Adhi Kuncoro Rosyad

17.11.1440

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 22 Agustus 2022

Dosen Pembimbing,

Banu Santoso, ST, M.Eng
NIK. 190302327

HALAMAN PENGESAHAN
SKRIPSI
IMPLEMENTASI METODE OVERSAMPLING UNTUK
MENINGKATKAN KINERJA ALGORITMA KLASIFIKASI PADA
KASUS IMBALANCED DATASET

yang disusun dan diajukan oleh
Galih Adhi Kuncoro Rosyad

17.11.1440

Telah dipertahankan di depan Dewan Penguji
pada tanggal 22 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Agit Amrullah, S.Kom, M.Kom
NIK. 190302356

Donni Prabowo, M.Kom
NIK. 190302253

Banu Santoso, S.T, M.Eng
NIK. 190302327

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 22 Agustus 2022

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Galih Adhi Kuncoro Rosyad
NIM : 17.11.1440

Menyatakan bahwa Skripsi dengan judul berikut:

Implementasi Metode Oversampling Untuk Meningkatkan Kinerja Algoritma Klasifikasi Pada Kasus Imbalanced Dataset

Dosen Pembimbing : Banu Santoso, ST, M.Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 22 Agustus 2022

Yang Menyatakan,



Galih Adhi Kuncoro Rosyad



HALAMAN PERSEMBAHAN

Segala puji dan syukur peneliti panjatkan kehadirat Allah SWT atas berkat, nikmat, rahmat dan hidayah-Nya penulis diberi kesempatan sehingga mampu menyelesaikan skripsi dengan baik dengan segala kekurangan penulis. Penulis mengucapkan terima kasih kepada pihak-pihak yang telah membantu serta mendukung dalam penyusunan naskah baik secara langsung maupun tidak langsung. Dengan segala kerendahan hati skripsi ini saya persembahkan kepada :

1. Kedua orang tua saya, Bapak Timo dan Ibu Turhayati yang selalu memberikan doa dan dukungannya serta dorongan baik moril maupun materil sehingga penulis dapat menyelesaikan skripsi dengan baik.
2. Untuk adik saya Bayu Agung Setiyawan yang selalu mendoakan serta mendukung setiap langkah saya.
3. Bapak Banu Santoso, ST, M.Eng selaku dosen pembimbing yang tidak pernah lelah dalam membimbing dan mengingatkan penulis serta selalu memberi motivasi sehingga skripsi ini dapat terselesaikan.
4. Bapak Yoga Pristyanto, S.Kom, M.Eng yang berkenan berbagi pengetahuan serta mentor sehingga penelitian ini dapat tercapai.
5. Sdr. Lucky Adhikrisna Wirasakti selaku teman yang sangat membantu dalam pembuatan program maupun naskah.
6. Teman –teman kelas 17 S1 Informatika 08 yang telah menemani, bekerja sama serta berjuang bersama dalam masa menempuh pendidikan kuliah.
7. Semua pihak keluarga besar Universitas Amikom Yogyakarta yang tidak dapat kami sebutkan satu-persatu.
8. Seluruh pihak keluarga maupun kerabat yang telah membantu dukungan maupun doa yang tidak dapat kami sebutkan satu-persatu.

KATA PENGANTAR

Segala puji dan syukur peneliti panjatkan kehadirat Allah SWT atas berkat, nikmat, rahmat dan hidayah-Nya, sehingga penulis mampu menyelesaikan skripsi dengan judul “Implementasi Metode *Oversampling* Untuk Meningkatkan Kinerja Algoritma Klasifikasi Pada Kasus *Imbalanced Dataset*” sebagai syarat dalam menyelesaikan Program Sarjana Strata-1 (S1) jurusan Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta, serta dapat memberikan penambahan ide yang dapat dikembangkan dimasa yang akan datang.

Dalam penyusunan skripsi ini penulis banyak menghadapi kendala dan hambatan, namun berkat bantuan, bimbingan, dan kerjasama dari berbagai pihak sehingga kendala serta hambatan tersebut dapat diatasi. Untuk itu pada kesempatan ini penulis menyampaikan rasa hormat, rasa sayang dan terimakasih kepada :

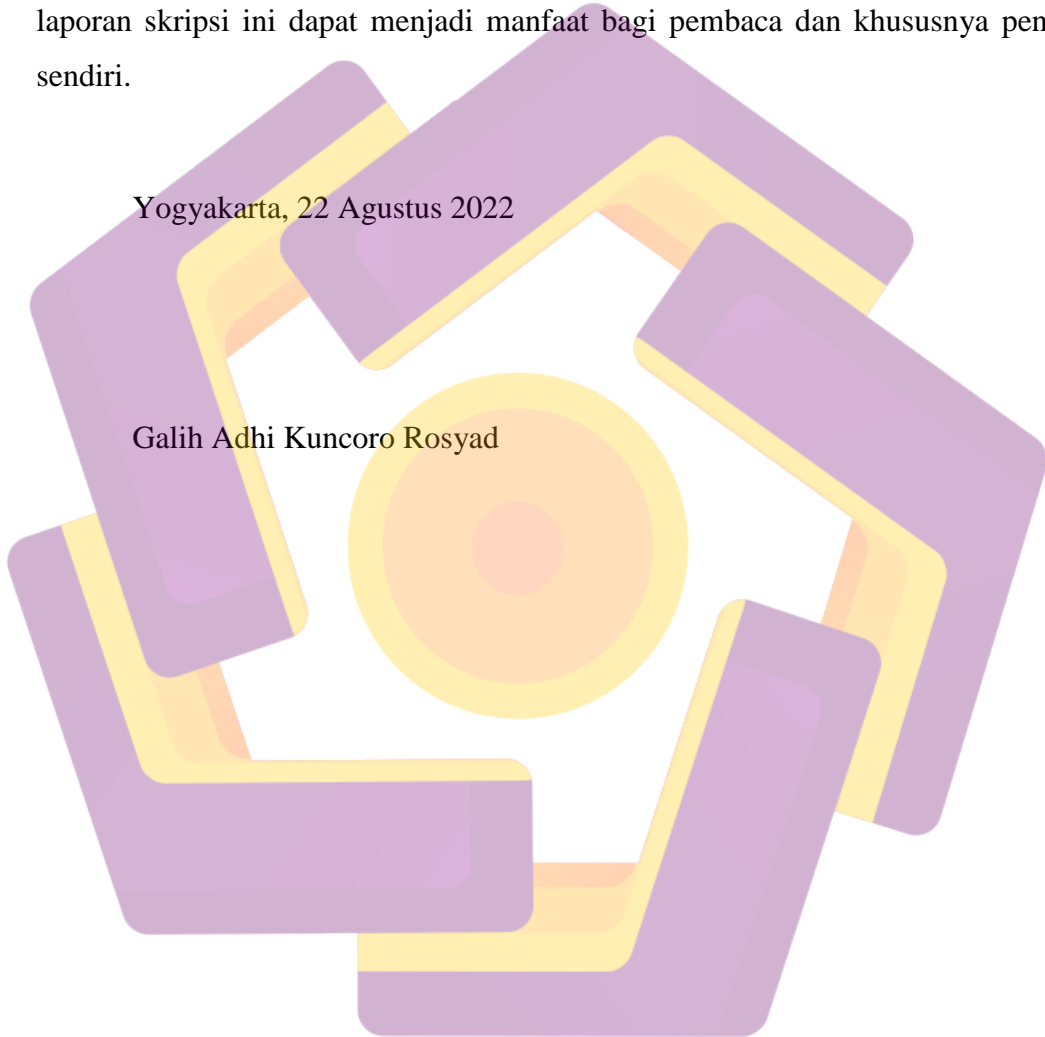
1. Bapak M. Suyanto, Prof., Dr., M.M. selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Banu Santoso, ST, M.Eng selaku pembimbing yang telah sabar dan tulus dalam membimbing, memberikan ilmunya serta memberi motivasi kepada penulis.
3. Ibu Krisnawati, S.Si, M.T selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Ibu Windha Mega Pradnya D, M.Kom selaku Kepala Prodi Informatika Universitas Amikom Yogyakarta.
5. Dosen Penguji
6. Bapak Yoga Pristyanto, S.Kom, M.Eng yang berkenan berbagi pengetahuan dan dukungan sehingga penelitian ini dapat terlaksana.
7. Seluruh jajaran dosen Program Studi S1 Informatika Fakultas Ilmu Komputer yang bersedia memberikan bekal pengetahuan selama perkuliahan berlangsung.
8. Seluruh pihak yang telah membantu dalam membuat penyusunan skripsi ini yang tidak bisa penulis sebutkan satu-persatu.

Penulis menyadari bahwa dalam pelaksanaan dan penulisan skripsi ini masih terdapat kesalahan atau hal yang kurang berkenan, sehingga masih jauh dari kata sempurna, semua tidak lepas karena keterbatasan peneliti. Untuk itu peneliti membutuhkan kritik dan saran yang membangun demi kesempurnaan skripsi ini.

Akhir kata hanya dengan berdoa kepada Allah SWT, peneliti berharap semoga laporan skripsi ini dapat menjadi manfaat bagi pembaca dan khususnya penulis sendiri.

Yogyakarta, 22 Agustus 2022

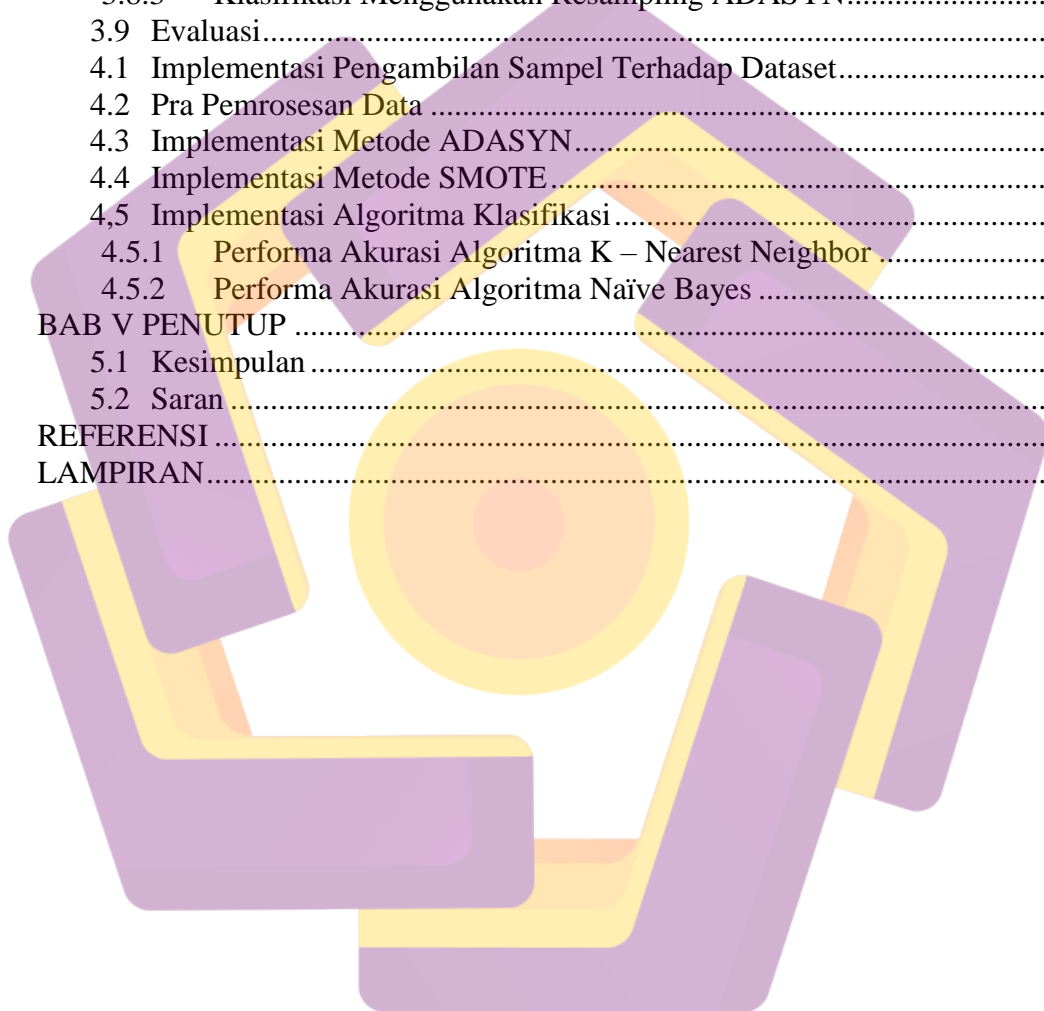
Galih Adhi Kuncoro Rosyad



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vi
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN.....	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
ABSTRACT.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Metode Penelitian	5
1.6.1 Metode Pengumpulan Data.....	5
1.6.2 Metode Klasifikasi	5
1.6.3 Metode Penanganan Ketidakseimbangan Kelas	5
1.6.4 Metode Evaluasi.....	5
1.7 Sistematika Penulisan	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori	19
2.2.1 Data Mining	19
2.2.2 Adaptive Synthetic.....	20
2.2.3 Synthetic Minority Oversampling Technique.....	22
2.2.4 Algoritma Klasifikasi	23
2.2.4.1 Naïve Bayes	23
2.2.4.2 K – Nearest Neighbor	25
2.2.5 Data Preprocessing.....	26
2.2.6 Dataset Splitting.....	27
2.2.7 Evaluasi Model	27
BAB III METODE PENELITIAN	30
3.1 Objek Penelitian.....	30
3.2 Alur Penelitian	30
3.3 Alat dan Bahan.....	31
3.3.1 Alat.....	31

3.3.2	Bahan	32
3.4	Pra Pemrosesan Data	33
3.5	Penanganan Ketidakseimbangan Kelas	34
3.6	Proses SMOTE	35
3.7	Proses ADASYN	39
3.8	Klasifikasi	44
3.8.1	Klasifikasi Tanpa Resampling	44
3.8.2	Klasifikasi Menggunakan Resampling SMOTE.....	44
3.8.3	Klasifikasi Menggunakan Resampling ADASYN.....	45
3.9	Evaluasi.....	45
4.1	Implementasi Pengambilan Sampel Terhadap Dataset.....	46
4.2	Pra Pemrosesan Data	47
4.3	Implementasi Metode ADASYN.....	47
4.4	Implementasi Metode SMOTE.....	51
4.5	Implementasi Algoritma Klasifikasi.....	56
4.5.1	Performa Akurasi Algoritma K – Nearest Neighbor	56
4.5.2	Performa Akurasi Algoritma Naïve Bayes	58
BAB V PENUTUP		62
5.1	Kesimpulan	62
5.2	Saran	63
REFERENSI		64
LAMPIRAN.....		69



DAFTAR TABEL

Tabel 2. 1 Perbandingan Metode SMOTE dan ADASYN	11
Tabel 2. 2 Perbandingan Dengan Penelitian Sebelumnya	14
Tabel 2. 3 Confusion Matrix	28
Tabel 3. 1 Karakteristik Dataset.....	32
Tabel 3. 2 Wine Quality	33
Tabel 3. 3 Us_Crime	33
Tabel 3. 4 Contoh Dataset Sebelum Dilakukan Proses SMOTE	36
Tabel 3. 5 Konfigurasi Dataset Sebelum Dan Sesudah Proses Smote.....	38
Tabel 3. 6 Contoh Dataset Setelah Dilakukan Proses Smote.....	39
Tabel 3. 7 Dataset Sebelum Dilakukan Proses Adasyn	40
Tabel 3. 8 Kemunculan Kategori Masing- masing Fitur	41
Tabel 3. 9 Jarak Setiap Kategori Pada Masing-masing Fitur.....	41
Tabel 3. 10 Nearest Neighbor untuk Data 1 (D1).....	41
Tabel 3. 11 Nearest Neighbor Pada Setiap Data Minoritas	42
Tabel 3. 12 Densitas Distribusi Untuk Data kelas Minoritas.....	42
Tabel 3. 13 Jumlah Duplikasi Data Sintesis	43
Tabel 3. 14 Dataset Setelah Dilakukan Proses Sampling Adasyn	43
Tabel 4. 1 Hasil Akurasi Dataset Wine_Quality dengan Algoritma KNN dan Naïve Bayes.....	61
Tabel 4. 2 Hasil Akurasi Dataset Us_Crime dengan Algoritma KNN dan Naïve Bayes	61

DAFTAR GAMBAR

Gambar 2. 1 Proses dalam tahapan data mining	20
Gambar 2. 3 Skema Data Splitting	27
Gambar 3. 1 Skenario Pertama Alur Penelitian	31
Gambar 3. 2 Diagram Alur Proses <i>Class Balancing</i>	34
Gambar 3. 3 Skenario Perbandingan Klasifikasi	44
Gambar 4. 1 Grafik Distribusi Kelas Dataset	47
Gambar 4. 2 Dataset wine_quality IR 26 Sebelum Resampling.....	48
Gambar 4. 3 Dataset wine_quality Setelah Resampling Adasyn.....	49
Gambar 4. 4 Dataset wine_quality Setelah Resampling Adasyn.....	49
Gambar 4. 5 Dataset us_crime IR 12 Sebelum Resampling	50
Gambar 4. 6 Dataset us_crime Setelah Resampling Adasyn	51
Gambar 4. 7 Distribusi Dataset Setelah Resampling Adasyn.....	51
Gambar 4. 8 Dataset wine_quality IR 26 Sebelum Resampling.....	52
Gambar 4. 9 Dataset wine_quality Setelah Resampling Smote.....	53
Gambar 4. 10 Distribusi Dataset Setelah Resampling Smote.....	53
Gambar 4. 11 Dataset us_crime IR 12 Sebelum Resampling	54
Gambar 4. 12 Dataset us_crime Setelah Resampling Smote	55
Gambar 4. 13 Distribusi Dataset Setelah Resampling Smote.....	55
Gambar 4. 14 Nilai Akurasi Dataset wine_quality KNN	57
Gambar 4. 15 Nilai Akurasi Dataset us_crime KNN.....	58
Gambar 4. 16 Nilai Akurasi Dataset wine_quality Naïve Bayes.....	59
Gambar 4. 17 Nilai Akurasi Dataset us_crime Naïve Bayes.....	60

DAFTAR LAMPIRAN

Lampiran 1. Profil obyek Penelitian	10
Lampiran 2. Dokumentasi Penelitian	11



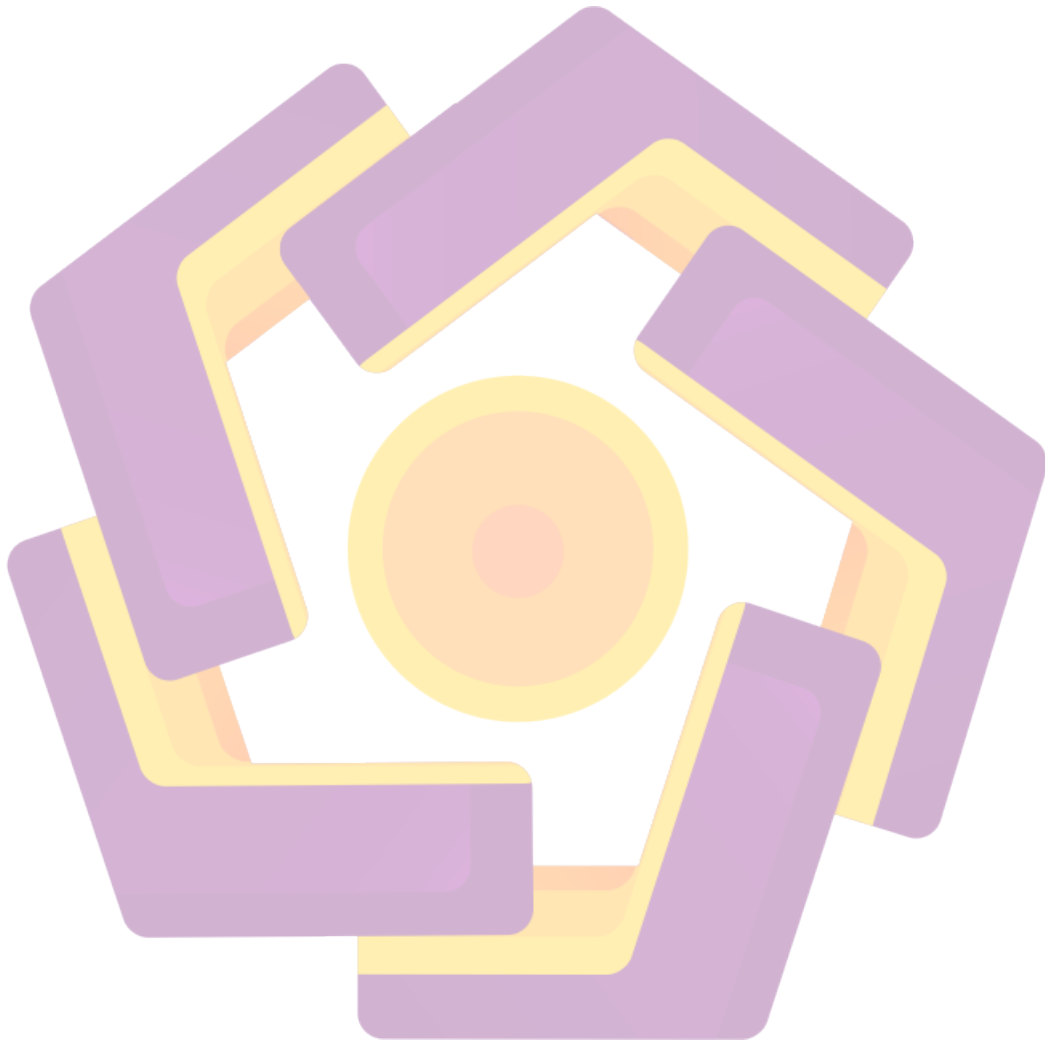
DAFTAR LAMBANG DAN SINGKATAN

ADASYN	Adaptive Synthetic
SMOTE	Synthetic Minority Oversampling Technique
G-Mean	Geometric Mean
IR	Imbalance Ratio
TNR	True Negative Rate
TPR	True Positive Rate
ROS	Random Oversampling
EEG	Electroencephalogram
KDD	Knowledge Discovery in Database



DAFTAR ISTILAH

Vektor	besaran yang mempunyai arah
Eigen Value	akar akar persamaan



INTISARI

Permasalahan yang dilakukan dalam penelitian ini adalah ketidakseimbangan kelas, hal ini dapat menimbulkan resiko kesalahan dalam melakukan klasifikasi dikarenakan jumlah kelas mayoritas yang lebih mendominasi serta menjadikan kelas minoritas sebagai noise sample. Penelitian ini menggunakan metode *Adaptive Synthetic* dan *Synthetic Minority Oversampling Technique* dalam menangani ketidakseimbangan pada kelas dataset. Penelitian ini melibatkan dua dataset yang berbeda serta dievaluasi berdasarkan nilai *balance accuracy* dan *geometric-mean*. Pengujian evaluasi *balance accuracy* maupun *geometric-mean* pada dataset *wine_quality* diperoleh hasil yang optimal pada metode ADASYN dengan menggunakan algoritma klasifikasi *Naïve Bayes* dengan *balance accuracy* 0,751718 dan *geometric-mean* 0,749367 . Sedangkan pada dataset *us_crime* diperoleh hasil yang optimal pada metode SMOTE dengan menggunakan algoritma klasifikasi *Naïve Bayes* dengan *balance accuracy* 0,901907 dan *geometric – mean* 0,899661. Dengan peningkatan hasil *balance accuracy* dan *geometric-mean*, maka kemampuan klasifikasi dan kepekaan terhadap data minoritas meningkat, sehingga hasil penelitian ini diharapkan dapat digunakan sebagai acuan penanganan kasus ketidakseimbangan kelas.

Kata Kunci: Adaptive Synthetic, Synthetic Minority Oversampling Technique, K – Nearest Neighbor, Naïve Bayes, geometric – mean, ketidakseimbangan kelas.

ABSTRACT

The Problem in this study is class imbalance, this can cause a risk of error in classifying because the number of majority classes dominates and makes the minority class a noise sample. This study uses the Adaptive Synthetic method and the Synthetic Minority Oversampling Technique in dealing with imbalances in the dataset class. This study involved two different datasets and evaluated based on the values of balance accuracy and geometric-mean. The evaluation test for balance accuracy and geometric-mean on the wine_quality dataset obtained optimal results on the ADASYN method using the Naïve Bayes classification algorithm with a balance accuracy of 0.751718 and a geometric-mean of 0.749367. Meanwhile, in the us_crime dataset, optimal results were obtained in the SMOTE method using the Naïve Bayes classification algorithm with balance accuracy 0.901907 and geometric-mean 0.899661. By increasing the results of balance accuracy and geometric-mean, the classification ability and sensitivity to minority data increases, so that the results of this study are expected to be used as a reference for handling class imbalance cases.

Keywords: *Adaptive Synthetic, Synthetic Minority Oversampling Technique, K-Nearest Neighbor, Naïve Bayes, geometric – mean, class imbalance.*

