IMPLEMENTASI SMOTE DAN PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI PADA KASUS IMBALANCE CLASS

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika

disusun oleh

MUHAMMAD FATHUR RIZQI

19.11.2757

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

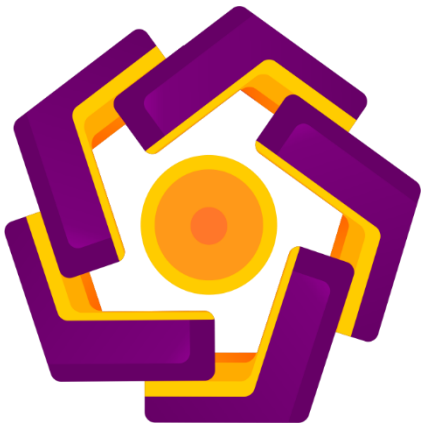
2022

IMPLEMENTASI SMOTE DAN PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI PADA KASUS IMBALANCE CLASS

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi Informatika



disusun oleh

MUHAMMAD FATHUR RIZQI

19.11.2757

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2022

HALAMAN PERSETUJUAN

SKRIPSI

IMPLEMENTASI SMOTE DAN PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI PADA KASUS IMBALANCE CLASS

yang disusun dan diajukan oleh

Muhammad Fathur Rizqi

19.11.2757

telah disetujui oleh Dosen Pembimbing Skripsi

pada tanggal 23 Desember 2022

Dosen Pembimbing,

Yoga Pristyanto, S.Kom, M.Eng

NIK. 190302412

HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI SMOTE DAN PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI PADA KASUS IMBALANCE CLASS

yang disusun dan diajukan oleh

Muhammad Fathur Rizqi

19.11.2757

Telah dipertahankan di depan Dewan Penguji
pada tanggal 23 desember 2022

Susunan Dewan Penguji

Nama Penguji	**Tanda Tangan**
Majid Rahardi, S.Kom, M.Eng **NIK. 190302393**	_____
Ikmah, M.Kom **NIK. 190302282**	_____
Yoga Pristyanto, S.Kom, M.Eng **NIK. 190302412**	_____

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 23 desember 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 : Muhammad Fathur Rizqi
NIM : 19.11.2757

Menyatakan bahwa Skripsi dengan judul berikut:

IMPLEMENTASI SMOTE DAN PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI PADA KASUS IMBALANCE CLASS

Dosen Pembimbing : Yoga Pristyanto, S.Kom, 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, 23 Desember 2022

Yang Menyatakan,

Muhammad Fathur Rizqi

KATA PENGANTAR

Dengan memanjatkan puja dan puji syukur kehadirat Allah Subhanahu Wata'ala, yang telah memberikan nikmat ilmu, kesehatan dan kecukupan rezeki sehingga penulis mampu dalam menyelesaikan skripsi ini yang berjudul "Implementasi SMOTE dan Perbandingan Kinerja Algoritma Klasifikasi Pada Kasus Imbalance Class", sebagai salah satu syarat dalam menyelesaikan Program Sarjana (S1) Jurusan Informatika.

Penulis sadar bahwa jika tanpa bantuan dan dukungan dari berbagai pihak penyusunan skripsi ini tidak akan berjalan lancar, oleh karena itu penulis ingin menyampaikan rasa terima kasih sebesar-besarnya kepada seluruh pihak terkait yang telah memberikan dukungan dan motivasi dalam penyusunan skripsi ini. Pada kesempatan ini penulis ingin menyampaikan terima kasih kepada:

1. Kedua orang tua di rumah yang selalu memberikan dukungan baik secara material maupun doa.
2. Bapak Yoga Pristiyanto, S.Kom., M.Eng selaku dosen pembimbing yang telah memberikan arahan dan bimbingan dalam pengerjaan skripsi ini.
3. Serta teman-teman kelas informatika 03 yang saling berbagi ilmu pengetahuan selama perkuliahan berlangsung.

Selain itu penulis juga menyadari bahwa dalam skripsi ini masih banyak terdapat kekurangan, oleh karenanya peneliti ingin meminta maaf dan mengharapkan saran dan kritik dari para pembaca agar peneliti dapat membuat karya yang lebih baik di masa depan. Penulis berharap skripsi ini bisa bermanfaat bagi para pembaca.

Yogyakarta, 30 November 2022

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

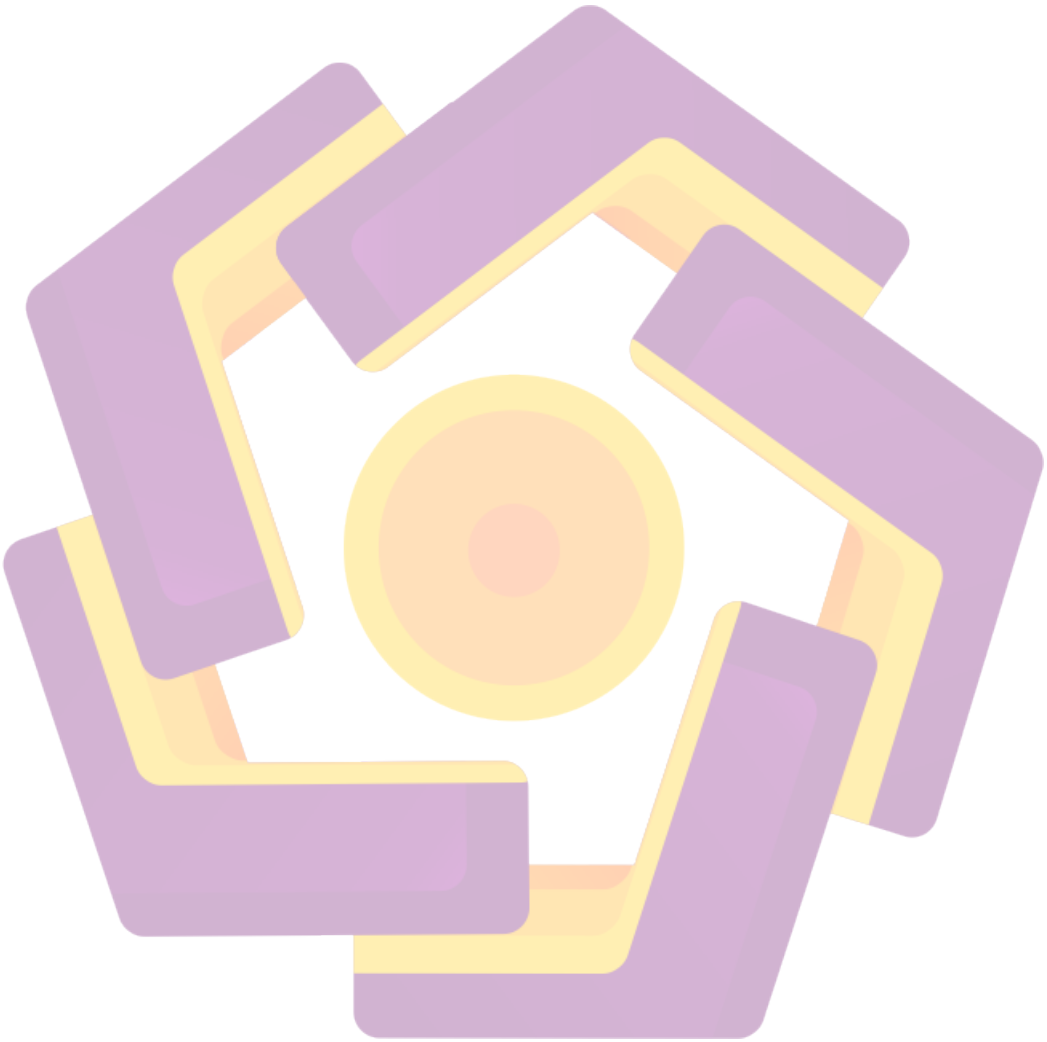
DAFTAR LAMBANG DAN SINGKATAN

σ	Standar Deviasi
μ	Rata-rata
SVM	Support Vector Machines
NBC	Naïve Bayes Classifier
KNN	K-Nearest Neighbor
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
SMOTE	Syntethic Minority Oversampling Technique

DAFTAR ISTILAH

Akurasi	Total akumulasi data yang diklasifikasi benar oleh classifier
Classifier	Model machine learning yang berfungsi mengklasifikasi data
Data Testing	Data yang digunakan untuk pengetesan model
Data Training	Data yang digunakan untuk pelatihan model
G-mean	Hasil evaluasi akurasi pada kelas positif dan negatif
Library	Sebuah modul yang menyimpan banyak fungsi untuk dipakai saat melakukan pemrograman
Precision	Banyaknya data yang benar positif pada data yang diprediksi sebagai positif
Sensitivity	Proporsi data positif yang diklasifikasi benar sebagai positif
Sklearn	Salah satu library pada bahasa pemrograman python
Specificity	Proporsi data negatif yang diklasifikasi benar sebagai negative
Imbalance Class	Ketidakseimbangan proporsi kelas pada suatu dataset
Resampling	Penyeimbangan proporsi kelas pada suatu dataset
Undersampling	Penyeimbangan proporsi kelas dengan mengurangi jumlah data kelas mayoritas

Oversampling	Penyeimbangan proporsi kelas dengan menambah jumlah data kelas minoritas



INTISARI

Permasalahan pada kasus klasifikasi data mining yang cukup sering terjadi yaitu adanya imbalance class pada suatu dataset yang dapat mempengaruhi kualitas klasifikasi pada model machine learning sehingga model tidak dapat memprediksi dengan benar kedua kelas secara seimbang.

Penelitian ini bertujuan mengaplikasikan algoritma SMOTE *oversampling* untuk menangani permasalahan *imbalance class* pada dataset *stroke prediction* dan melakukan perbandingan hasil pada algoritma klasifikasi yang diimplementasikan SMOTE dan tanpa SMOTE. Algoritma yang digunakan yakni Naïve Bayes Classifier, C4.5 dan K-Nearest Neighbor. Uji evaluasi dilakukan menggunakan *g-mean*, *confusion matrix* dan akurasi.

Hasil yang didapatkan dari penelitian ini adalah SMOTE mempengaruhi penurunan nilai akurasi pada tiap algoritma klasifikasi seperti pada algoritma C4.5 terjadi penurunan nilai akurasi sebesar 3%, KNN dengan penurunan nilai akurasi sebesar 12% dan NBC turun sebesar 16%. Namun disisi lain nilai g-mean pada algoritma klasifikasi yang telah dilakukan SMOTE mengalami peningkatan seperti pada algoritma C4.5 mengalami peningkatan nilai g-mean setelah diimplementasikan SMOTE sebesar 13%, KNN dengan peningkatan yang cukup tinggi yakni 45% dan NBC dengan peningkatan yang sedikit yaitu 4%.

Kata kunci: Data Mining, Imbalance Class, Klasifikasi, Oversampling, Syntethic Minority Over-sampling Technique (SMOTE)

ABSTRACT

The problem in the data mining classification case that occurs quite frequently is the imbalance class in a dataset which can affect the classification quality of the machine learning model so that the model cannot correctly predict both classes in a balanced way.

This study aims to apply the SMOTE oversampling algorithm to address the class imbalance problem in the stroke prediction dataset and to compare the results of the classification algorithm implemented with SMOTE and without SMOTE. The algorithms used are Naïve Bayes Classifier, C4.5 and K-Nearest Neighbor. Evaluation test is carried out using g-mean, confusion matrix and accuracy.

The results obtained from this study are that SMOTE affects the decrease in the accuracy value of each classification algorithm as in the C4.5 algorithm there is a decrease in the accuracy value of 3%, KNN with a decrease in the accuracy value of 12% and NBC decreases by 16%. But on the other hand the g-mean value of the classification algorithm that has been carried out by SMOTE has increased as in the C4.5 algorithm the g-mean value has increased after implementing SMOTE by 13%, KNN with a fairly high increase of 45% and NBC with a slight increase which is about 4%.

Keyword: *Data Mining, Imbalance Class, Classification, Oversampling, Syntethic Minority Over-sampling Technique (SMOTE)*