IMPLEMENTASI ALGORITMA SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE) UNTUK MENANGANI KETIDAKSEIMBANGAN KELAS PADA DATASET KLASIFIKASI

SKRIPSI

disusun oleh :

Gagah Gumelar

16.11.0799

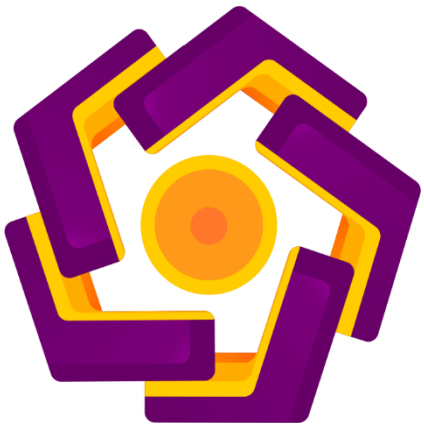
PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA

2020

IMPLEMENTASI ALGORITMA SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE) UNTUK MENANGANI KETIDAKSEIMBANGAN KELAS PADA DATASET KLASIFIKASI

SKRIPSI

untuk memenuhi sebagai persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Gagah Gumelar

16.11.0799

PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2020

PERSETUJUAN

SKRIPSI

IMPLEMENTASI ALGORITMA SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE) UNTUK MENANGANI KETIDAKSEIMBANGAN KELAS PADA DATASET KLASIFIKASI

yang dipersiapkan dan disusun oleh

Gagah Gumelar

16.11.0799

telah disetujuai oleh Dosen Pembimbing Skripsi

pada tanggal 20 Maret 2020

Dosen Pembimbing,

Mulia Sulistiyono, M.Kom
NIK. 190302248

PENGESAHAN

SKRIPSI

IMPLEMENTASI ALGORITMA SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE) UNTUK MENANGANI KETIDAKSEIMBANGAN KELAS PADA DATASET KLASIFIKASI

yang dipersiapkan dan disusun oleh

Gagah Gumelar

16.11.0799

telah dipertahankan di depan Dewan Penguji
pada tanggal 20 Maret 2020

Susunan Dewan Penguji

Nama Penguji	**Tanda Tangan**
Agung Nugroho, M.Kom **NIK. 190302242**	_____
Haryoko, M.Kom **NIK. 190302268**	_____
Mulia Sulistiyono, M.Kom **NIK. 190302248**	_____

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 20 Maret 2020

Dekan Fakultas Ilmu Komputer

Krisnawati, S.Si, M.T
NIK. 190302038

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 20 Maret 2020

Gagah Gumelar

NIM. 16.11.0799

MOTTO

“Slow Proggress Is Better Than No Progress”

“Tidak Ada Kata Sibuk, Hanya Kamu Saja Yang Tidak Bisa Membagi Waktu”

“Entar, Nanti, Besuk, Lusa adalah Ilusi ”

“Kalau Bisa Sekarang Kenapa Harus Nanti”

PERSEMBAHAN

Alhamdulilahirabil'alamin, puji syukur kepada Allah SWT, atas rahmat dan karunia-Nya sehingga saya bisa menjadi pribadi yang beriman, berfikir, bersabar dan memberikan kelancaran serta kemudahan dalam menyelesaikan tugas akhir ini, tentunya juga tidak lepas dari dukungan dari orang-orang yang ada disekeliling saya yang selalu memberi semangan dan doa. Untuk itu saya mengucapkan terimakasih kepada:

1. Orang tuaku Bapak Alm. Musiman dan Ibu Yatimah, Terima kasih yang tidak henti-hentinya Penulis ucapkan atas kasih sayangnya. Terima kasih sudah mendidik dan merawatku hingga detik ini. Kesuksesan anakmu adalah berkat doa yg selalu kalian panjatkan. Semoga Allah SWT selalu memberikan keberkahan, kesehatan, umur yang panjang dan keberkahan dunia maupun akhirat.
2. Seluruh keluarga dan semua pihak yang mungkin tidak dapat disebutkan satu persatu, terimakasih atas dukungan dan doanya.
3. Sela Oktavia Sari yang selalu menyemangati, dalam setiap keadaan baik suka maupun duka. Terimakasih atas dukungan selama pengerjaan skripsi ini. Semoga kebaikan kalian dibalas oleh Allah SWT dan menjadi orang sukses.
4. Teman-teman Kelas S1 IF-13 angkatan 2016, untuk semuanya yang tidak bisa disebutkan satu-satu disini tanpa terkecuali, Terimakasih atas bantuannya, dukungan dan doanya. Semoga kalian sukses dan apa yang kalian cita-citakan tercapai.

KATA PENGANTAR

Bismillahirahmanirahim, Alhamdulillahirabil'alamin puji syukur penulis panjatkan kepada Allah SWT atas Rahmat dan karunia-Nya , sehingga penulis dapat menyelesaikan skripsi dengan judul "Implementasi Algoritma Synthetic Minority Over-Sampling Technique (SMOTE) untuk Menangani Ketidakseimbangan Kelas pada Dataset Klasifikasi". Untuk memenuhi syarat akademis dalam menyelesaikan Program Studi Strata Satu (S1) pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta. Mengingat keterbatsan pengetahuan dan pengalaman dalam penulisan skripsi ini, Penulis banyak mendapatkan bimbingan, petunjuk, saran dan arahan dari berbaga pihak, oleh karena itu dengan kerendahan hati dan rasa hormat Penulis mengucapkan terima kasih sebesar-besarnya kepada :

1. Bapak Prof. Dr. M. Suyanto, MM. Selaku Rektor Universitas AMIKOM Yogyakarta..
2. Bapak Mulia Sulistiyono, M.Kom. Selaku Dosen pembimbing yang telah membantu dalam penulisan skripsi ini.
3. Ibu/bapak Selaku Dosen Penguji yang telah memberikan petunjuk, serta nasehat dalam ujian skripsi ini.

Penulis mendoakan untuk semua pihak yang telah membantu dalam penulisan skripsi ini semoga diberikan balasan dan berkah dari Allah SWT. Penulis menyadari masih banyak kekurangan dalam penulisan skripsi ini untuk itu saran, kritik dan perbaikan yang bersifat membangun sangat diharapkan. Akhir kata Penulis berharap semoga skripsi ini bermanfaat untuk semua pihak yang membutuhkan.

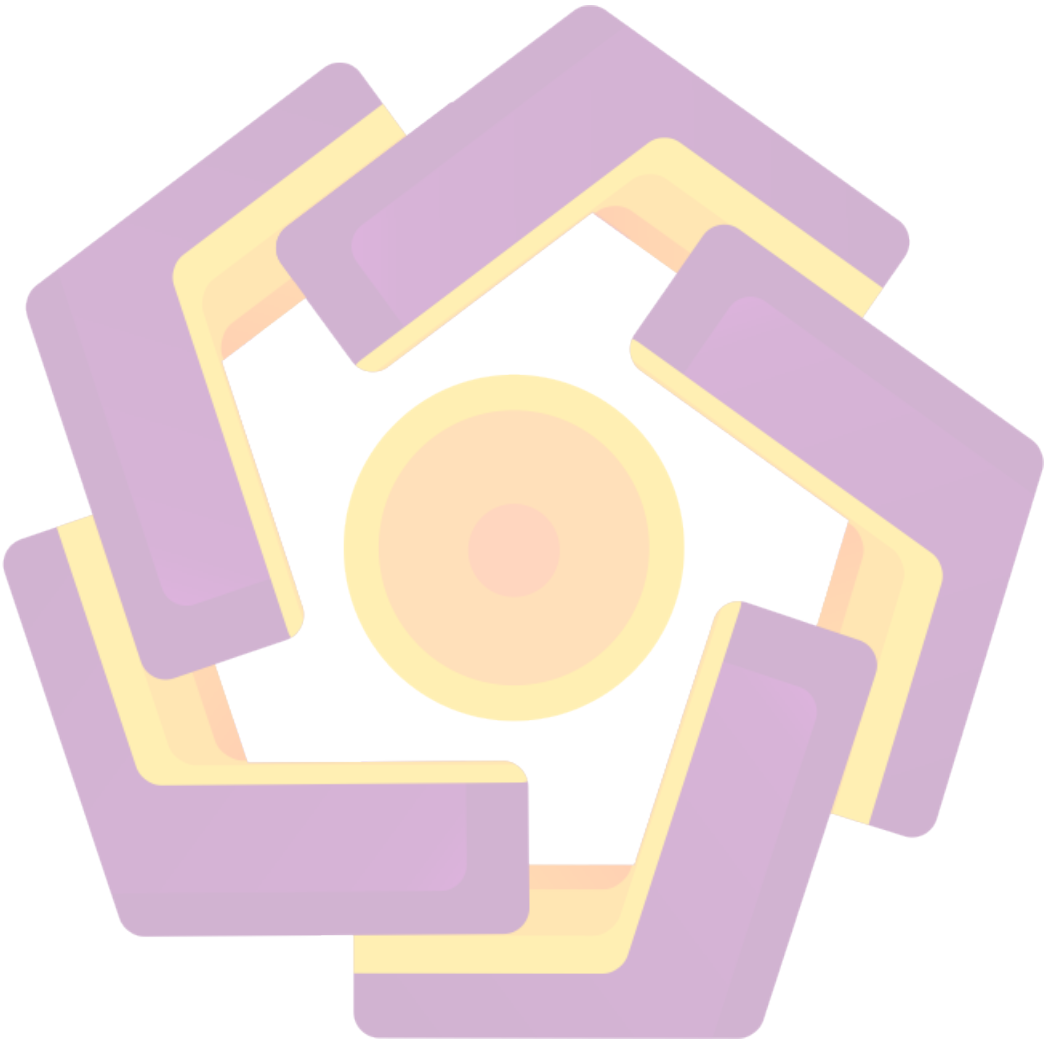
Yogyakarta, 20 Maret 2020

Penulis,

Gagah Gumelar

DAFTAR ISI

DAFTAR TABEL



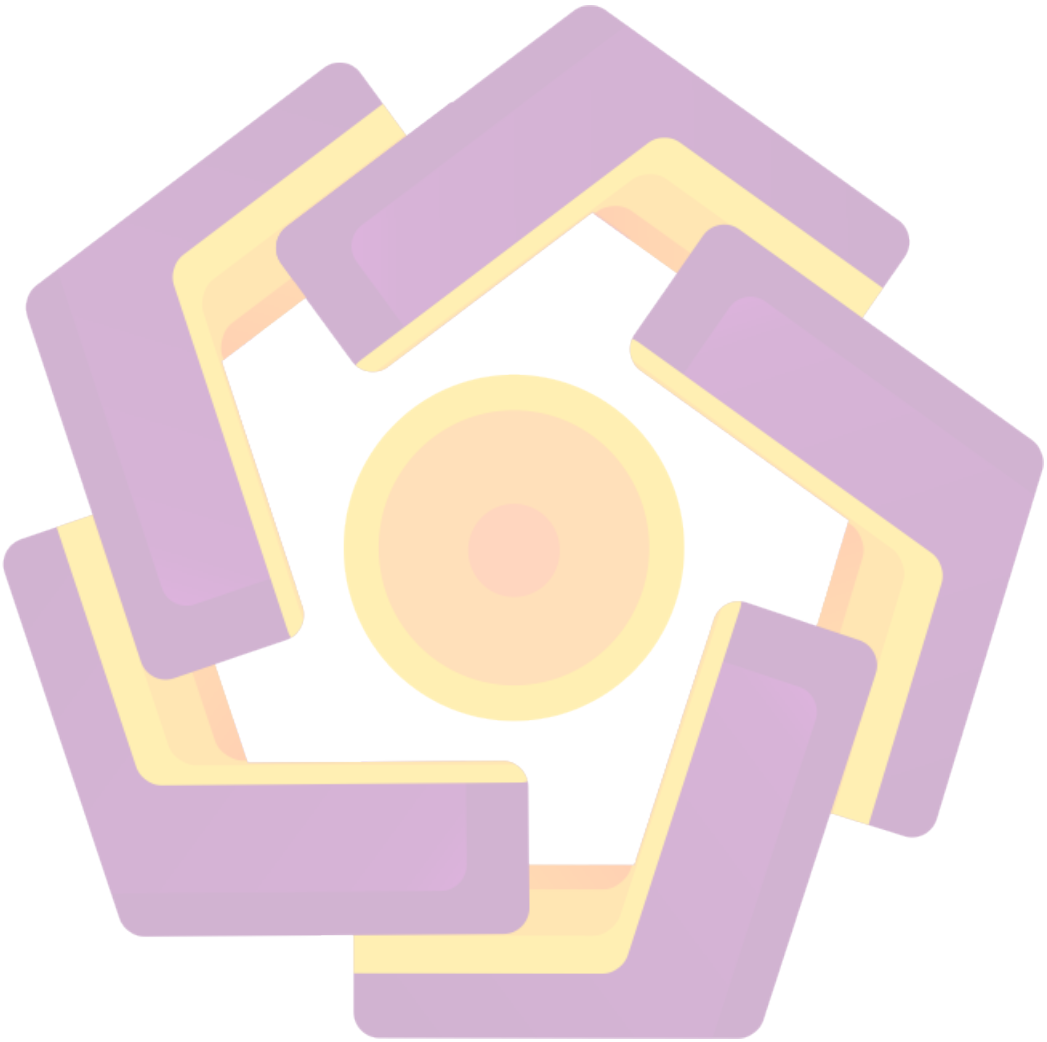
DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR ISTILAH

SMOTE = Synthetic Minority Over-sampling Technique

Imbalance class = kelas tidak seimbang



INTISARI

Sebuah kelas dikatakan tidak seimbang apabila ada suatu kelas yang memiliki data yang lebih banyak dibandingkan dengan kelas lainnya. Kelompok kelas dengan jumlah data yang banyak disebut dengan kelas mayoritas, sedangkan kelompok kelas dengan jumlah yang sedikit disebut dengan kelas minoritas. Perbandingan antara kelas minoritas dengan kelas mayoritas disebut dengan *Imbalance Ratio (IR)*. Semakin besar perbedaan antara kelas minoritas dengan kelas mayoritas maka nilai dari *Imbalance Ratio (IR)* semakin besar.

Ketidakseimbangan dataset pada data mining adalah masalah yang serius. Dataset yang tidak seimbang menyebabkan misleading atau kesesatan dalam Hasil klasifikasi dimana data kelas minoritas sering diklasifikasikan sebagai kelas mayoritas. Penerapan algoritma klasifikasi tanpa memperhatikan keseimbangan kelas mengakibatkan prediksi yang baik bagi kelas mayoritas dan kelas minoritas diabaikan. Oleh karena itu pada penelitian ini diimplementasikan algoritma *Syntethic Minority Over-Sampling Technique (SMOTE)* untuk menyeimbangkan dataset. Penelitian ini menggunakan 4 dataset dengan *Imbalance Ratio* yang berbeda dan menggunakan algoritma klasifikasi C45, Naïve Bayes, K-NN, dan SVM. Kemudian dibandingkan antara sebelum dilakukan SMOTE dan setelah dilakukan SMOTE.

Dari hasil penelitian yang sudah dilakukan nilai akurasi dan nilai G-mean algoritma Naïve Bayes konsisten dengan performanya pada setiap level imbalance ratio, sebelum implementasi SMOTE memiliki performa yang tidak baik, sedangkan setelah diimplementasikan SMOTE algoritma Naïve Bayes memiliki peningkatan akurasi yang konsisten. Sehingga dapat ditarik kesimpulan bahwa kombinasi SMOTE + Naïve Bayes paling efektif digunakan pada dataset imbalance dengan level yang berbeda-beda pada skema 10 fold cross validation maupun 80% data testing yang diujikan sebanyak 50 kali.

Kata kunci : klasifikasi dengan dataset tidak seimbang, *Syntethic Minority Over-Sampling Technique (SMOTE)*

ABSTRACT

A class to be imbalance when there is a class that has more data than other classes. The group of classes with a lot of data is called the majority class, while the class groups with fewer numbers are called minority classes. A comparison between minority classes and the majority class is called Imbalance Ratio (IR). The greater the difference between the minority class and the majority class the value of the Imbalance Ratio (IR) is getting larger.

Dataset imbalance in data mining is a serious problem. The imbalanced Dataset causes misleading or error in the classification results where minority class data is often classified as a majority class. The application of the classification algorithm regardless of class balance resulted in a good prediction for the majority class and a neglected minority class. Therefore in this research implemented Syntethic Minority Over-Sampling Technique (SMOTE) algorithm to balance the dataset. The study used 4 datasets with different Imbalance Ratio and used classification algorithms, C45, Naïve Bayes, K-NN, and SVM. Then compared between before SMOTE and after doing SMOTE.

From the research results that have been done accuracy value and value G-mean Naïve Bayes algorithm is consistent with its performance at each level of imbalance ratio, before the implementation of SMOTE has no good performance, whereas after the implemented SMOTE algorithm Naïve Bayes has a consistent increase in accuracy. So it can be concluded that the combination SMOTE + Naïve Bayes most effectively used in the imbalance dataset with different levels in the scheme of 10 fold cross validation and 80% data testing tested as much as 50 times.

Keywords: classification with unbalanced datasets, Syntethic Minority Over-Sampling Technique (SMOTE