PERBANDINGAN METODE RANDOM UNDER SAMPLING DAN SMOTE PADA KLASIFIKASI DATA TIDAK SEIMBANG

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1-Informatika

disusun oleh

ANDIKA CANDRA PRADANA

18.11.2501

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2024

PERBANDINGAN METODE RANDOM UNDER SAMPLING DAN SMOTE PADA KLASIFIKASI DATA TIDAK SEIMBANG

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1-Informatika

disusun oleh

ANDIKA CANDRA PRADANA

18.11.2501

Kepada

FAKULTAS ILMU KOMPUTER

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2024

HALAMAN PERSETUJUAN

SKRIPSI

PERBANDINGAN METODE RANDOM UNDER SAMPLING DAN SMOTE PADA KLASIFIKASI DATA TIDAK SEIMBANG

yang disusun dan diajukan oleh

Andika Candra Pradana

18.11.2501

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 28 Februari 2024

Dosen Pembimbing,

Anna Baita, M.Kom
NIK. 190302290

HALAMAN PENGESAHAN

SKRIPSI

PERBANDINGAN METODE RANDOM UNDER SAMPLING DAN SMOTE PADA KLASIFIKASI DATA TIDAK SEIMBANG

yang disusun dan diajukan oleh

Andika Candra Pradana

18.11.2501

Telah dipertahankan di depan Dewan Penguji
pada tanggal 28 Februari 2024

Susunan Dewan Penguji

Nama Penguji	Tanda Tangan
Majid Rahardi, S.Kom., M.Eng NIK. 190302393	
Irma Rofni Wulandari, S.Pd., M.Eng NIK. 190302329	
Anna Baita, M.Kom NIK. 190302290	

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 28 Februari 2024

DEKAN FAKULTAS ILMU KOMPUTER

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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Andika Candra Pradana
NIM : 18.11.2501

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN METODE RANDOM UNDER SAMPLING DAN SMOTE PADA KLASIFIKASI DATA TIDAK SEIMBANG

Dosen Pembimbing : Anna Baita, M.Kom

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.

Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.

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.

Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.

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, 28 Februari 2024
Yang Menyatakan,

Andika Candra Pradana

KATA PENGANTAR

Segala puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan karunia dan rahmat-Nya. Sholawat serta salam semoga tercurahkan kepada nabi Muhammad Saw, sehingga penulis dapat menyelesaikan skripsi yang berjudul "**Perbandingan Metode Random Under Sampling dan SMOTE Pada Klasifikasi Data Tidak Seimbang**".

Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan program studi S1-Informatika pada fakultas ilmu komputer. Tujuan dari skripsi ini adalah untuk mengetahui performa dan menentukan teknik resampling dalam membantu algoritma klasifikasi menghadapi data tidak seimbang. Dalam penyusunan skripsi ini tak lepas dari bantuan, dukungan dan doa dari berbagai pihak. Oleh karena itu penulis ingin mengucapkan terimakasih sebesar-besarnya kepada:

1. Prof. Dr. Muhammad Suyanto M.M. selaku Rektor Universitas Amikom Yogyakarta.
2. Hanif Al Fatta, S.Kom., M.Kom., Ph.D. selaku Dekan Fakultas Ilmu Komputer.
3. Windha Mega Pradnya Dhuhita, M.Kom. selaku Ketua Prodi Informatika
4. Anna Baita, M.Kom. selaku Dosen Pembimbing selama penyusunan skripsi.
5. Majid Rahardi, S.Kom., M.Eng selaku Dosen Penguji 1 saat sidang pendadaran.
6. Irma Rofni Wulandari, S.Pd., M.Eng selaku Dosen Penguji 2 saat sidang pendadaran.
7. Keluarga yaitu bapak Margiyanto, ibu Tisngatun, dan kedua adikku (Bayu dan Talita) yang selalu mendoakan dan memberikan dukungan dalam penyelesaian skripsi.

8. Rekan-rekan xelim boys yaitu Abdul, Hamzah, Salman, Akhmad yang selalu mengingatkan untuk menyelesaikan skripsi.
9. Rekan-rekan kelas IF-10 yang selalu mendukung untuk menyelesaikan skripsi.

Semoga penelitian ini dapat memberikan manfaat dan kontribusi untuk pengembangan ilmu pengetahuan, khususnya bidang machine learning. Penulis meminta maaf apabila terdapat kekurangan dalam skripsi mengingat terbatasnya kemampuan dan ilmu pengetahuan yang dimiliki. Akhir kata, penulis mengharapkan kritik dan saran yang membangun untuk pengembangan dan perbaikan di masa mendatang.

Yogyakarta, 28 Februari 2024

Penulis

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

INTISARI

Terdapat permasalahan yang seringkali muncul pada klasifikasi, yaitu ketidakseimbangan kelas. Perbandingan antar teknik *resampling* yang telah dilakukan sebelumnya hanya diujikan pada satu *dataset* saja. Tentunya hal itu akan menimbulkan suatu pertanyaan mengenai konsistensi performa teknik *resampling*. Beberapa penelitian mengenai perbandingan antar teknik *resampling* ini belum ada yang menyimpulkan bahwa teknik *resampling* tersebut yang terbaik pada klasifikasi *dataset* yang tidak seimbang. Pada penelitian ini, peneliti akan melakukan perbandingan antar teknik *resampling* pada klasifikasi *dataset* yang tidak seimbang. Teknik *resampling* yang dipilih untuk penelitian ini adalah SMOTE dan *random undersampling*. Teknik *resampling* terbaik akan ditentukan berdasarkan persentase peningkatan tertinggi pada nilai *area under the curve* (AUC) dan *geometric mean* (*g-mean*). Hasil pada penelitian ini yaitu ***random undersampling*** merupakan teknik *resampling* terbaik. *Random undersampling* memperoleh rata-rata persentase kenaikan pada setiap algoritma klasifikasi sebesar (**121%**) pada *sensitivity*, (**17%**) pada AUC, dan (**36%**) pada *g-mean*. Namun pada specificity, SMOTE merupakan teknik resampling terbaik dalam menjaga akurasi *specificity*, dimana rata-rata persentase penurunan di setiap algoritma klasifikasi pada nilai *specificity* hanya (**6%**). *Random undersampling* mengalami penurunan yang cukup tinggi dengan rata-rata persentase penurunan di setiap algoritma klasifikasi sebesar (**15%**).

Kata Kunci: *Ketidakseimbangan kelas, Teknik resampling, Klasifikasi, Random Undersampling, SMOTE*

ABSTRACK

There are problems that often arise in classification, namely class imbalanced. The comparison between resampling techniques that has been done before is only tested on one dataset. Of course it will raise a question about the consistency of the performance of resampling techniques. Several studies on the comparison between resampling techniques have not concluded that the resampling technique is the best in the classification of imbalanced datasets. In this study, researchers will conduct a comparison between resampling techniques on the classification of imbalanced datasets. The resampling techniques chosen for this research are SMOTE and random undersampling. The best resampling technique will be determined based on the highest percentage increase in the area under the curve (AUC) and geometric mean (g-mean) values. The result of this study is that random undersampling is the best resampling technique. Random undersampling obtained an average percentage increase in each classification algorithm of (121%) in sensitivity, (17%) in AUC, and (36%) in g-mean. However, in specificity, SMOTE is the best resampling technique in maintaining specificity accuracy, where the average percentage decrease in each classification algorithm in specificity value is only (6%). Random undersampling experienced a fairly high decline with an average percentage decline in each classification algorithm of (15%).

Keywords: *Imbalanced class, Resampling Technique, Classification, Random Undersampling, SMOTE*