



**PENERAPAN SMOTE UNTUK MENANGANI *IMBALANCE CLASS* DALAM
KLASIFIKASI DEPOSITO MENGGUNAKAN ALGORITMA *EXTREME*
*GRADIENT BOOSTING***

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

DINA ARIFAH

NIM 1911016120004

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2023



**PENERAPAN SMOTE UNTUK MENANGANI *IMBALANCE CLASS* DALAM
KLASIFIKASI DEPOSITO MENGGUNAKAN ALGORITMA *EXTREME
GRADIENT BOOSTING***

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

Oleh

**DINA ARIFAH
NIM 1911016120004**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2023

SKRIPSI

PENERAPAN SMOTE UNTUK MENANGANI *IMBALANCE CLASS* DALAM KLASIFIKASI DEPOSITO MENGGUNAKAN ALGORITMA *EXTREME* *GRADIENT BOOSTING*

Oleh:

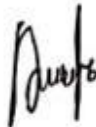
DINA ARIFAH

NIM. 1911016120004

Telah dipertahankan di depan Dosen Penguji pada tanggal 09 Juni 2023.

Susunan Dosen Penguji:

Pembimbing I



Triando Hamonangan Saragih, S.Kom., M.Kom
NIP. 199308242019031012

Dosen Penguji I



Muliadi, S.Kom., M.Cs.
NIP. 197804222010121002

Pembimbing II



Dwi Kartini, S.Kom., M.Kom
NIP. 198704212012122003

Dosen Penguji II



Muhammad Itqan Mazdadi, S.Kom., M.Kom.
NIP. 199006122019031013

Banjarbaru, 09 Juni 2023

Koordinator Program Studi Ilmu Komputer



Muhammad Subhan, S.T., M.Kom
NIP. 197303252008121001

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 09 Juni 2023

Yang Menyatakan,



Dina Arifah

NIM.1911016120004

ABSTRAK

PENERAPAN SMOTE UNTUK MENANGANI *IMBALANCE CLASS* DALAM KLASIFIKASI DEPOSITO MENGGUNAKAN ALGORITMA *EXTREME GRADIENT BOOSTING*

(Oleh : Dina Arifah; Pembimbing: Triando Hamonangan Saragih, S.Kom., M.Kom dan Dwi Kartini, S.Kom., M.Kom; 2023; 66 halaman)

Deposito menjadi salah satu produk utama dan sumber pendanaan bagi bank dan meningkatkan pemasaran deposito sangat penting. Namun, telemarketing sebagai bentuk pemasaran deposito kurang efektif dan efisien karena mengharuskan memanggil setiap pelanggan untuk penawaran deposit. Oleh karena itu, identifikasi calon nasabah simpanan diperlukan agar telemarketing menjadi lebih efektif dan efisien dengan menyasar nasabah yang tepat, sehingga meningkatkan kinerja pemasaran bank dengan tujuan akhir meningkatkan sumber pendanaan bagi bank. Untuk mengidentifikasi pelanggan, data mining digunakan dengan UCI Bank Marketing Dataset dari lembaga perbankan Portugis. Dataset ini terdiri dari 45.211 record dengan 17 atribut. Algoritma klasifikasi yang digunakan adalah *Extreme Gradient Boosting* (XGBoost) yang cocok untuk data berukuran besar. Data yang digunakan memiliki ketidakseimbangan kelas tinggi, dengan persentase "ya" dan "tidak" masing-masing 11,7% dan 88,3%. Oleh karena itu, solusi yang diusulkan dalam penelitian, yang berfokus pada mengatasi *Imbalance Class* dalam dataset pemasaran Bank, adalah dengan menggunakan *Synthetic Minority Over-sampling* (SMOTE) dan metode XGBoost. Hasil penelitian XGBoost adalah akurasi 0,91016, presisi 0,79476, recall 0,72928, F1-Score 0,56198, Area ROC 0,93831, dan AUCPR 0,63886. Setelah SMOTE diterapkan, akurasinya 0,91072, presisi 0,78883, recall 0,75588, F1-Score 0,59153, Area ROC 0,93723, dan AUCPR 0,63733. Hasil penelitian menunjukkan bahwa XGBoost dan SMOTE dapat mengungguli algoritma lain seperti K-Nearest Neighbor, Random Forest, Logistic Regression, Artificial Neural Network, Naïve Bayes, dan Support Vector Machine dalam hal akurasi. Studi ini berkontribusi pada pengembangan model machine learning efektif yang dapat digunakan sebagai sistem pendukung bagi para ahli teknologi informasi di industri keuangan dan perbankan untuk mengidentifikasi calon nasabah yang tertarik untuk berlangganan deposito dan meningkatkan sumber pendanaan bank.

Kata kunci: Klasifikasi, *SMOTE*, *Imbalance Class*, Deposito, *XGBoost*, *Bank Marketing*, *Customer Identification*

ABSTRACT

APPLICATION OF SMOTE TO HANDLE IMBALANCE CLASS IN DEPOSIT CLASSIFICATION USING THE EXTREME GRADIENT BOOSTING ALGORITHM

(By : Dina Arifah; Supervisor: Triando Hamonangan Saragih, S.Kom., M.Kom and Dwi Kartini, S.Kom., M.Kom; 2023; 66 pages)

Deposits became one of the main products and funding sources for banks and increasing deposit marketing is very important. However, telemarketing as a form of deposit marketing is less effective and efficient as it requires calling every customer for deposit offers. Therefore, the identification of potential deposit customers was necessary so that telemarketing became more effective and efficient by targeting the right customers, thus improving bank marketing performance with the ultimate goal of increasing sources of funding for banks. To identify customers, data mining is used with the UCI Bank Marketing Dataset from a Portuguese banking institution. This dataset consists of 45,211 records with 17 attributes. The classification algorithm used is Extreme Gradient Boosting (XGBoost) which is suitable for large data. The data used has a high-class imbalance, with "yes" and "no" percentages of 11.7% and 88.3%, respectively. Therefore, the proposed solution in the research, which focused on addressing the Imbalance Class in the Bank marketing dataset, was to use Synthetic Minority Over-sampling (SMOTE) and the XGBoost method. The result of the XGBoost study was an accuracy of 0.91016, precision of 0.79476, recall of 0.72928, F1-Score of 0.56198, ROC Area of 0.93831, and AUCPR of 0.63886. After SMOTE was applied, the accuracy was 0.91072, the precision was 0.78883, the recall was 0.75588, F1-Score was 0.59153, ROC Area was 0.93723, and AUCPR was 0.63733. The results showed that XGBoost and SMOTE could outperform other algorithms such as K-Nearest Neighbor, Random Forest, Logistic Regression, Artificial Neural Network, Naïve Bayes, and Support Vector Machine in terms of accuracy. This study contributes to the development of effective machine learning models that can be used as a support system for information technology experts in the finance and banking industries to identify potential customers interested in subscribing to deposits and increasing bank funding sources.

Keywords: *Classification, SMOTE, Imbalance Class, Deposit, XGBoost, Bank Marketing, Customer Identification*

PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “PENERAPAN SMOTE UNTUK MENANGANI *IMBALANCE CLASS* DALAM KLASIFIKASI DEPOSITO MENGGUNAKAN ALGORITMA *EXTREME GRADIENT BOOSTING*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini
2. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Ibu Dwi Kartini, S.Kom., M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. Semua pihak yang tidak dapat di sebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 09 Juni 2023

A handwritten signature in black ink, appearing to read 'Dina Arifah', written in a cursive style.

Dina Arifah

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA	vi
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	3
BAB II TINJAUAN PUSTAKA	4
2.1 Kajian Terdahulu	4
2.2 Landasan Teori	9
2.2.1 Deposito	9
2.2.2 <i>Imbalance Class</i>	9
2.2.3 <i>Cross Validation</i>	10
2.2.4 <i>Synthetic Minority Oversampling Technique</i>	11
2.2.5 <i>Boosting</i>	11
2.2.6 <i>Extreme Gradient Boosting</i>	12
2.2.7 Evaluasi	14
BAB III METODE PENELITIAN	17
3.1 Bahan Penelitian	17
3.2 Alat Penelitian	17

3.3	Variabel Penelitian.....	17
3.4	Prosedur Penelitian	17
BAB IV HASIL DAN PEMBAHASAN		21
4.1	Hasil.....	21
4.1.1	Pengumpulan Dataset.....	21
4.1.2	<i>Preprocessing Data</i>	24
4.1.3	Pembagian <i>Data Training</i> dan <i>Data Testing</i>	25
4.1.4	<i>SMOTE</i>	26
4.1.5	Klasifikasi	26
4.2	Pembahasan	34
BAB V PENUTUP.....		37
5.1	Kesimpulan.....	37
5.2	Saran	37
DAFTAR PUSTAKA		38
LAMPIRAN.....		42

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Perancangan Penelitian	8
Tabel 3. <i>Confusion Matrix</i>	14
Tabel 4. <i>Bank Marketing Dataset</i>	18
Tabel 5. Deskripsi Atribut data <i>Bank Marketing</i>	21
Tabel 6. Contoh <i>Bank Marketing dataset</i>	22
Tabel 7. Jumlah Distribusi Kelas	23
Tabel 8. Daftar parameter XGBoost	27
Tabel 9. Putaran penyiapan hyperparameter	28
Tabel 10. Hasil pengujian XGBoost tanpa hyperparameter.....	29
Tabel 11. Hasil pengujian XGBoost dengan hyperparameter.....	29
Tabel 12. Hasil uji XGBoost dengan SMOTE menggunakan nilai K=6	30
Tabel 13. Hasil uji XGBoost dengan SMOTE menggunakan nilai K=7	31
Tabel 14. Hasil uji XGBoost dengan SMOTE menggunakan nilai K=8	32
Tabel 15. Hasil uji XGBoost dengan SMOTE menggunakan nilai K=9	32
Tabel 16. Hasil uji XGBoost dengan SMOTE menggunakan nilai K=10	33
Tabel 17. Perbandingan hasil uji XGBoost dan XGBoost dengan SMOTE.....	35

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Ilustrasi <i>K-Fold Cross Validation</i>	10
Gambar 2. <i>Boosting algorithm principle</i>	12
Gambar 3. Evolusi algoritma XGBoost	13
Gambar 4. Alur algoritma XGboost.....	13
Gambar 5. Alur Penelitian.....	18
Gambar 6. Perbandingan Presentase Nasabah Deposito dan Bukan Nasabah Deposito	24
Gambar 7. Alur Preprocessing	25
Gambar 8. Grafik Akurasi rata-rata untuk semua skenario pengujian.....	34
Gambar 9. Grafik Presisi rata-rata untuk semua skenario pengujian.....	34
Gambar 10. Grafik Recall rata-rata untuk semua skenario pengujian	34
Gambar 11. Grafik F-1 Score rata-rata untuk semua skenario pengujian.....	34
Gambar 12. Grafik ROC Area rata-rata untuk semua skenario pengujian	35
Gambar 13. Grafik AUCPR rata-rata untuk semua skenario pengujian	35

DAFTAR LAMPIRAN

Lampiran

Lampiran 1 *Source Code Import Library*

Lampiran 2 *Source Code Load Dataset*

Lampiran 3 *Source Code Preprocessing Data*

Lampiran 4 *Source Code XGBoost Model*

Lampiran 5 *Source Code Konfigurasi Hyperparameter Tuning*

Lampiran 6 *Source Code XGBoost Model dengan Hyperparameter Tuning*

Lampiran 7 *Source Code XGBoost Model dengan SMOTE K=6 sampai K=10*