

**IMPLEMENTASI SMOTE UNTUK MENGATASI *CLASS IMBALANCE*
PADA DATA MULTILABEL DENGAN ALGORITMA KNN**

SKRIPSI



Disusun oleh:

**Fitrahdin
16.11.0390**

**PROGRAM SARJANA
PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

**IMPLEMENTASI SMOTE UNTUK MENGATASI *CLASS IMBALANCE*
PADA DATA MULTILABEL DENGAN ALGORITMA KNN**

SKRIPSI

Diajukan kepada Fakultas Teknik Universitas AMIKOM Yogyakarta untuk
memenuhi salah satu syarat memperoleh gelar Sarjana Informatika
Pada Jenjang Program Sarjana – Program Studi Teknik Informatika



Disusun oleh:

Fitrahdin

16.11.0390

**PROGRAM SARJANA
PROGRAM STUDI S1 INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2021**

PERSETUJUAN

SKRIPSI

IMPLEMENTASI SMOTE UNTUK MENGATASI *CLASS IMBALANCE* PADA DATA MULTILABEL DENGAN ALGORITMA KNN

yang dipersiapkan dan disusun oleh

Fitrahdin

16.11.0390

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 8 November 2021

Dosen Pembimbing,

Windha Mega Pradnya D, M.Kom

NIK. 190302185

PENGESAHAN

SKRIPSI

**IMPLEMENTASI SMOTE UNTUK MENGATASI *CLASS*
IMBALANCE PADA DATA MULTILABEL DENGAN
ALGORITMA KNN**

yang dipersiapkan dan disusun oleh

Fitrahdin

16.11.0390

telah dipertahankan di depan Dewan
Penguji pada tanggal 20 Desember 2021

Susunan Dewan

Windha Mega Pradnya D. M.Kom

NIK. 190302185

Ali Mustafa

NIK. 0562118801

Hendra Kurniawan.

NIK. 0520058721

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Informatika

Tanggal 15 Januari 2021

DEKAN FAKULTAS ILMU KOMPUTER

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

NIK. 190302096

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.

Bima, 10 Januari 2022



Fitrahdin

NIM. 16.11.0390

MOTTO

*Yakin dan Percayalah bahwa usaha tidak akan pernah menkhianati hasil.
Ketika kamu mulai lelah dan seakan ingin menyerah. Ketahuilah bahwa
sesungguhnya pertolongan Allah hanya berjarak antara kening dan sajadah.*

*“Ketetapan-Nya pasti datang, dan janganlah kamu meminta agar dipercepat.”
Q.S. An-Nahl: 1*



PERSEMBAHAN

Ku persembahkan karya tulis ini kepada orang-orang yang ku sayangi dan ku banggakan :

1. Kedua orang tuaku, Bapak Musmuliadin dan Ibu Rita, terima kasih banyak atas kasih sayang, doa serta dukungannya selama ini.
2. Kakakku M.Setiawan serta adik-adikku Salman Al-Farizi dan Rahel, terima kasih banyak atas dukungan serta motivasinya.
3. Teman-teman di jurusan Informatika angkatan 2016 khususnya kelas 06, terima kasih banyak atas kebersamaannya.
4. Teman dalam segala hal Ninis Khairunisa, terima kasih atas suportnya
5. Kampus dan almamaterku tercinta Amikom.



KATA PENGANTAR

Dengan mengucapkan segala puji atas kehadiran Allah SWT yang Maha Pengasih lagi Maha Penyayang atas segala rahmat serta hidayah-Nya, sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “**Implementasi SMOTE Untuk Mengatasi Class Imbalance Pada Data Multibalel dengan Algoritma KNN**” sebagai salah satu syarat untuk meraih gelar Sarjana Strata 1 pada Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta Program Studi Informatika.

Tulisan ini dapat diselesaikan sebagaimana semestinya berkat bantuan berbagai pihak. Untuk itu pada kesempatan ini penulis menyampaikan ucapan terimakasih kepada :

1. Ibu Krisnawati, S.Si, M.T, Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
2. Bapak Sudarmawan, M.T, Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
3. Ibu Windha Mega Pradnya D,M.Kom selaku dosen pembimbing penulis yang dengan sabarnya membimbing dan memberikan semangat serta motivasi kepada penulis.
4. Bapak serta ibu dosen Program Studi Informatika.
5. Bapak, ibu serta saudara-saudaraku yang kucintai.
6. Teman-teman di jurusan Informatika angkatan 2016 khususnya kelas 06, terima kasih banyak atas kebersamaannya.

Penulis sangat mengharapkan kritik dan saran yang membangun. Semoga penelitian ini dapat bermanfaat khususnya bagi penulis sendiri dan bagi pembaca pada umumnya.

Bima, 29 Desember 2021

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
INTISARI	xii
ABSTRAK	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	4
1.7 Sistematika Penulisan	5
BAB II LANDASAN TEORI	7
2.1 Penelitian Terdahulu	7
2.2 Landasan Teori	10
2.2.1 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	10
2.2.2 <i>Imbalance Class</i>	13
2.2.3 <i>Klasifikasi</i>	14
2.2.4 <i>K-Fold Cross Validation</i>	15
2.2.5 <i>K-Nearest Neighbor</i>	16
2.2.6 <i>Confusion Matrix</i>	18
BAB III METODOLOGI PENELITIAN	21
3.1 Study Literatur	22
3.2 Dataset	23
3.3 Data Pre Processing	25
3.4 Klasifikasi <i>K-Nearest Neighbors</i> dengan Implementasi SMOTE	29
3.5 Evaluasia Performa	31
BAB IV HASIL DAN PEMBAHASAN	34
4.1 Dataset	34
4.2 Processing	35
4.3 Pengujian	38
4.4 Evaluasi dan Perbandingan Performa	44

BAB V PENUTUP	47
5.1 Kesimpulan	47
5.2 Saran	48
DAFTAR PUSTAKA	49

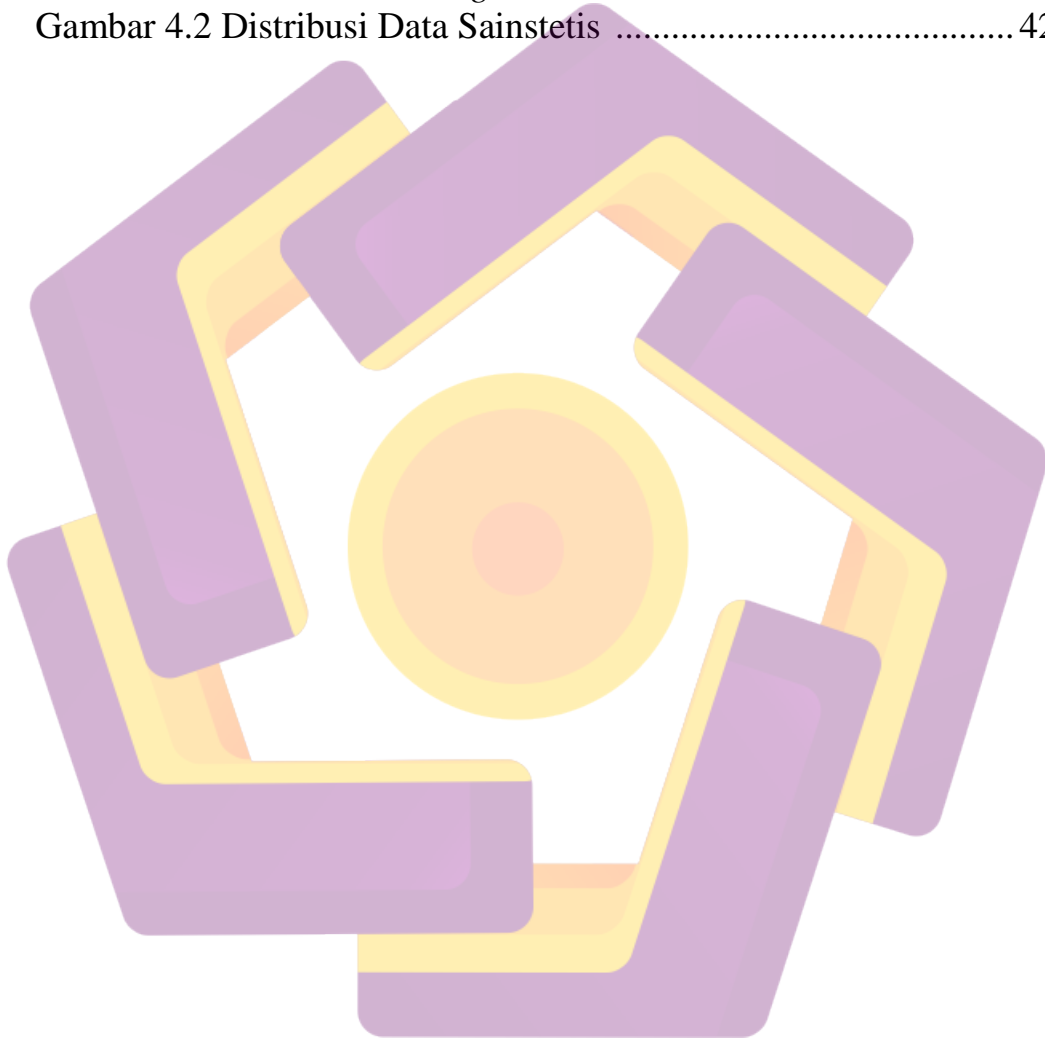


DAFTAR TABEL

Tabel 2.1 Perbandingan Penelitian	9
Tabel 2.2 Model <i>Confusion Matrix</i>	19
Tabel 3.1 Ditribusi Kelas Target	23
Tabel 3.2 Dataset	24
Tabel 3.3 Transformasi Nilai Pada Atribut <i>Buyig</i>	26
Tabel 3.4 Transformasi Nilai Atribut <i>Maint</i>	27
Tabel 3.5 Transformasi Nilai Atribut <i>Doors</i>	27
Tabel 3.6 Transformasi Nilai Atribut <i>Persons</i>	28
Tabel 3.7 Transformasi Nilai Atribut <i>Book</i>	28
Tabel 3.8 Transformasi Nilai Atribut <i>Safety</i>	29
Tabel 3.9 <i>Confusion Matrix</i>	31
Tabel 4.1 Data Awal	34
Tabel 4.2 Class Prediktor	35
Tabel 4.3 Class Target	35
Tabel 4.4 Transformasi Nilai Prediktor	36
Tabel 4.5 Confusion Matrix Algoritam K-Nearest Neighbors	38
Tabel 4.6 K-Fold Cross Validation	40
Tabel 4.7 Value dari Awal Atribut Predict sebelum di SMOTE ...	41
Tabel 4.8 Value dari Atribut Predict sesudah di SMOTE	41
Tabel 4.9 Confusion Matrix Algoritam K-Nearest Neighbors	43
Tabel 4.10 Perbandingan Performa.....	45

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi SMOTE	11
Gambar 2.2 Ilustrasi 10-fold cross validation	15
Gambar 3.1 Distribusi <i>Imbalance Class</i>	25
Gambar 3.4 Alur Pengujian	30
Gambar 4.1 Presentase <i>Fitting</i> Dataset	37
Gambar 4.2 Distribusi Data Sainstetis	42



INTISARI

Permasalahan ketidak seimbangan kelas akan terus ada karena data tidak dapat dipaksa untuk selalu seimbang. Ketidak seimbangan kelas memberikan dampak yang tidak baik pada hasil klasifikasi dimana kelas minoritas sering disalah klasifikasikan sebagai kelas mayoritas. Hal ini dapat menurunkan nilai accuracy hasil klasifikasi.

SMOTE merupakan salah satu turunan teknik over-sampling untuk menanggulangi ketidakseimbangan kelas dengan menyeimbangkan dataset dengan meningkatkan ukuran kelas minor. SMOTE diterapkan pada klasifikasi dataset *car evolution* menggunakan algoritma klasifikasi KNN. Hasil klasifikasi dievaluasi akurasi menggunakan *10 fold-cross validation* dengan membandingkan hasil klasifikasi yang hanya menggunakan KNN dan menggunakan KNN dan SMOTE.

Hasil penelitian menunjukkan bahwa penggunaan SMOTE mampu mengatasi *imbalance class* dengan menaikkan nilai akurasi sebesar 2%. Dimana awalnya akurasi yang hanya menggunakan KNN adalah sebesar 92% meningkat menjadi 94%. Nilai k pada klasifikasi K-NN sangat berpengaruh terhadap tingkat akurasi. Berdasarkan hasil uji dengan $k=3$, $k=5$ dan $k=10$, maka akurasi klasifikasi tertinggi K-NN adalah $k=5$.

Kata kunci: Klasifikasi, K-NN, SMOTE, *Imbalance Class*

ABSTRACT

Class imbalance problems will continue to exist because data cannot be forced to always be balanced. Class imbalance has an unfavorable impact on the results of the classification where the minority class is often misclassified as the majority class. This can reduce the accuracy value of the classification results.

SMOTE is a derivative of over-sampling technique to overcome class imbalance by balancing the dataset by increasing the size of the minor class. SMOTE is applied to the classification of the car evolution dataset using the KNN classification algorithm. The classification results were evaluated for accuracy using 10-fold cross-validation by comparing the classification results using only KNN and using KNN and SMOTE.

The results showed that the use of SMOTE was able to overcome the imbalance class by increasing the accuracy value by 2%. Where initially the accuracy that only uses KNN is 92%, increasing to 94%. The value of k in the K-NN classification is very influential on the level of accuracy. Based on the test results with k=3, k=5 and k=10, the highest classification accuracy of K-NN is k=5.

Keywords— *classification, K-NN, SMOTE, Imbalance Class*

