

**IMPLEMENTASI SYNTHETIC MINORITY OVER-SAMPLING
TECHNIQUE PADA OPINI PELANGGAN TOKOPEDIA
MENGUNAKAN K-NEAREST NEIGHBOR DAN DECISION
TREE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

AMMAR KANZ DIWANI

20.11.3435

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**IMPLEMENTASI SYNTHETIC MINORITY OVER-SAMPLING
TECHNIQUE PADA OPINI PELANGGAN TOKOPEDIA
MENGUNAKAN K-NEAREST NEIGHBOR DAN DECISION
TREE**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

AMMAR KANZ DIWANI

20.11.3435

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PERSETUJUAN

SKRIPSI

**Implementasi Synthetic Minority Over-sampling Technique Pada
Opini Pelanggan Tokopedia menggunakan K-Nearest Neighbor dan
Decision Tree**

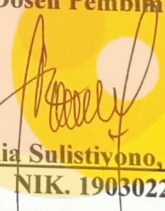
yang disusun dan diajukan oleh

Ammar Kanz Diwani

20.11.3435

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 1 Juli 2024

Dosen Pembimbing,


Mulia Sulistiyono, M.Kom
NIK. 190302248

HALAMAN PENGESAHAN

SKRIPSI

**Implementasi Synthetic Minority Over-sampling Technique Pada
Opini Pelanggan Tokopedia menggunakan K-Nearest Neighbor dan
Decision Tree**

yang disusun dan diajukan oleh

Ammar Kanz Diwani

20.11.3435

Telah dipertahankan di depan Dewan Penguji
pada tanggal 1 Juli 2024

Susunan Dewan Penguji

Nama Penguji

Dina Maulina, M.Kom
NIK. 190302250

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

Mulia Sulistiyono, M.Kom
NIK. 190302248

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 1 Juli 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 : Ammar Kanz Diwani
NIM : 20.11.3435

Menyatakan bahwa Skripsi dengan judul berikut:

Implementasi Synthetic Minority Over-sampling Technique Pada Opini Pelanggan Tokopedia menggunakan K-Nearest Neighbor dan Decision Tree

Dosen Pembimbing : Mulia Sulistiyono, M.Kom

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, 1 Juli 2024

Yang Menyatakan,



Ammar Kanz Diwani

HALAMAN PERSEMBAHAN

Segala puji dan syukur saya panjatkan kepada Allah SWT atas berkat, rahmat, dan hidayah-Nya yang telah memberikan kekuatan dan petunjuk sehingga saya dapat menyelesaikan skripsi ini dengan baik. Saya juga ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan kontribusi dalam penelitian dan penyusunan naskah ini. Skripsi ini saya persembahkan kepada:

1. Orang tua dan saudara saya yang senantiasa memberikan dukungan, doa, serta kasih sayang sehingga saya dapat menempuh pendidikan dan menyelesaikan skripsi ini dengan baik.
2. Bapak Mulia Sulistiyono, M. Kom selaku pembimbing yang telah sabar dan tulus membimbing serta memberikan ilmu pengetahuan yang berharga kepada saya.
3. Kepada Iq Kamala Imani, saya mengucapkan terima kasih yang sebesar-besarnya karena telah menjadi support system saya selama ini. Terima kasih telah selalu mengingatkan saya bahwa setiap orang memiliki proses yang berbeda-beda, yang menjadi motivasi besar bagi saya dalam menyelesaikan skripsi ini.
4. Kepada teman-teman saya Reja, Franky, Pani, Dimas, Taufik dan Allam saya mengucapkan terima kasih yang sebesar-besarnya. Terima kasih atas kehadiran dan dukungan kalian, terutama di saat-saat sulit. Kalian selalu ada untuk memberikan semangat dan dorongan, sehingga saya bisa bangkit dan menyelesaikan skripsi ini. Ucapan terima kasih ini saya sampaikan kepada kalian yang telah menganggap saya sebagai teman..
5. Seluruh pihak keluarga besar Universitas Amikom Yogyakarta yang tidak dapat saya sebutkan satu per satu, atas segala bantuan dan dukungan yang telah diberikan.

Terima kasih atas semua doa, dukungan, serta bimbingan yang telah diberikan. Semoga hasil penelitian ini dapat bermanfaat bagi banyak pihak.

KATA PENGANTAR

Segala puji dan syukur saya panjatkan kepada Allah SWT atas rahmat, berkah, dan hidayah-Nya, sehingga saya dapat menyelesaikan skripsi dengan judul “Implementasi Synthetic Minority Over-sampling Technique Pada Opini Pelanggan Tokopedia menggunakan K-Nearest Neighbor dan Decision Tree” sebagai syarat untuk menyelesaikan Program Sarjana (S1) pada Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.

Selama proses penyusunan skripsi ini, berbagai hambatan dan tantangan telah dihadapi. Namun, berkat bantuan dan bimbingan dari berbagai pihak, baik secara moral maupun spiritual, semua rintangan tersebut dapat dilalui. Pada kesempatan ini, saya ingin mengucapkan terima kasih kepada:

1. Bapak M. Suyanto, Prof., Dr., M.M. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Mulia Sulistiyono, M. Kom yang dengan sabar dan tulus telah membimbing serta memberikan ilmu kepada saya.
3. Seluruh jajaran dosen Program Studi S1 Informatika Fakultas Ilmu Komputer yang telah memberikan bekal ilmu pengetahuan selama masa perkuliahan.
4. Semua pihak yang tidak dapat saya sebutkan satu per satu yang telah membantu sehingga penelitian ini bisa dilaksanakan.

Saya menyadari bahwa laporan skripsi ini masih jauh dari kata sempurna, oleh karena itu kritik dan saran yang membangun sangat saya harapkan. Akhir kata, semoga tulisan ini dapat memberikan manfaat bagi pembaca dan khususnya bagi penulis sendiri.

Yogyakarta, 28 Juni 2024

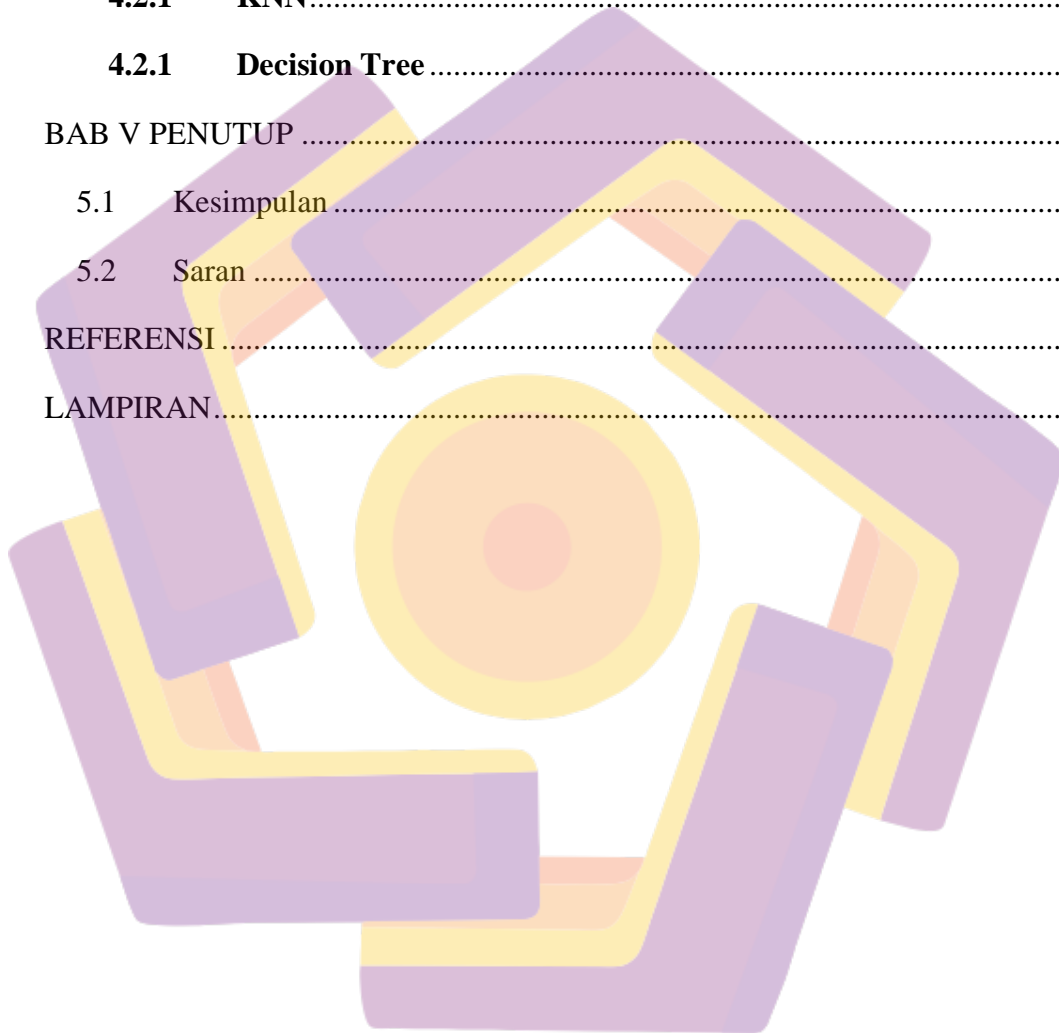
Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI.....	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
DAFTAR LAMBANG DAN SINGKATAN	xiv
DAFTAR ISTILAH	xv
INTISARI	xvi
ABSTRACT	xvii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Metodologi Penelitian.....	4
1.6.1 Metode Pengumpulan Data.....	4
1.6.2 Metode Klasifikasi	4

1.6.3	Metode Mengatasi Ketidak seimbangan Kelas	4
1.6.4	Metode Evaluasi	4
1.7	Sistematika Penulisan	5
BAB II LANDASAN TEORI		6
2.1	Tinjauan Pustaka	6
2.2	Dasar Teori	17
2.2.1	Synthetic Minority Over-sampling Technique (SMOTE)	17
2.2.2	Algoritma Classifier	17
2.2.3	Confusion Matrix	21
2.2.4	Python	23
2.2.5	Ketidakseimbangan Kelas	23
2.2.6	Pembagian Data	23
BAB III METODE PENELITIAN		26
3.1	Gambaran Umum	26
3.2	Alur Penelitian	26
3.2.1	Akuisisi Data	27
3.2.2	Tahap Pra Pemrosesan Data	27
3.2.3	Tahap Klasifikasi	27
3.2.4	Tahap Evaluasi	28
3.3	Analisa Data	28
3.3.1	Analisa Data	28
3.3.2	Penanganan Ketidakseimbangan Kelas	29
3.4	Pengujian Akurasi	36
3.5	Instrumen Penelitian	37
3.5.1	Kebutuhan Perangkat Keras	37

3.5.2	Perancangan Perangkat Lunak	38
BAB IV	HASIL DAN PEMBAHASAN	39
4.1	Implementasi SMOTE terhadap dataset	39
4.2	Implementasi Algoritma KNN terhadap dataset.....	41
4.2.1	KNN.....	41
4.2.1	Decision Tree	49
BAB V	PENUTUP	59
5.1	Kesimpulan	59
5.2	Saran	60
REFERENSI	61
LAMPIRAN	64



DAFTAR TABEL

TABEL 2.1 PERBANDINGAN JURNAL.....	10
TABEL 2.2 MODEL CONFUSION MATRIKS	21
TABEL 2.3 MODEL 10 K-FOLD CROSS VALIDATION	24
TABEL 3.1 Contoh dataset sebelum dilakukan proses SMOTE.....	30
TABEL 3.2 CONTOH KOMPOSISI DATASET SEBELUM DAN SETELAH DILAKUKAN PROSES SMOTE.....	34
TABEL 3.3 CONTOH KOMPOSISI DATASET SEBELUM DAN SETELAH DILAKUKAN PROSES SMOTE.....	34
TABEL 3.4 CONTOH DATASET SEBELUM DILAKUKAN PROSES SMOTE	35
TABEL 4.1 ilustrasi proses SMOTE dengan Dataset minoritas 1.....	40
TABEL 4.2 ILUSTRASI PROSES SMOTE DENGAN DATASET MINORITAS 0	40
TABEL 4.3 CONFUSION MATRIX KLASIFIKASI KNN DENGAN CROSS-VALIDATION FOLD 10 PADA DATASET	41
TABEL 4.4 HASIL KLASIFIKASI KNN DENGAN CROSS-VALIDATION FOLD 10 PADA DATASET.....	42
TABEL 4.5 CONFUSION MATRIX KLASIFIKASI KNN DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET DIULANG SEBANYAK 50 KALI	43
TABEL 4.6 HASIL KLASIFIKASI KNN DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET	43
TABEL 4.7 CONFUSION MATRIX KLASIFIKASI KNN DENGAN CROSS-VALIDATION FOLD 10 PADA DATASET	45
TABEL 4.8 HASIL KLASIFIKASI KNN DENGAN CROSS-VALIDATION FOLD 10 PADA DATASET.....	45
TABEL 4.9 CONFUSION MATRIX KLASIFIKASI KNN DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET DIULANG SEBANYAK 50 KALI	47
TABEL 4.10 HASIL KLASIFIKASI KNN DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET	47
TABEL 4.11 CONFUSION MATRIX KLASIFIKASI DECISION TREE DENGAN CROSS- VALIDATION FOLD 10 PADA DATASET.....	49

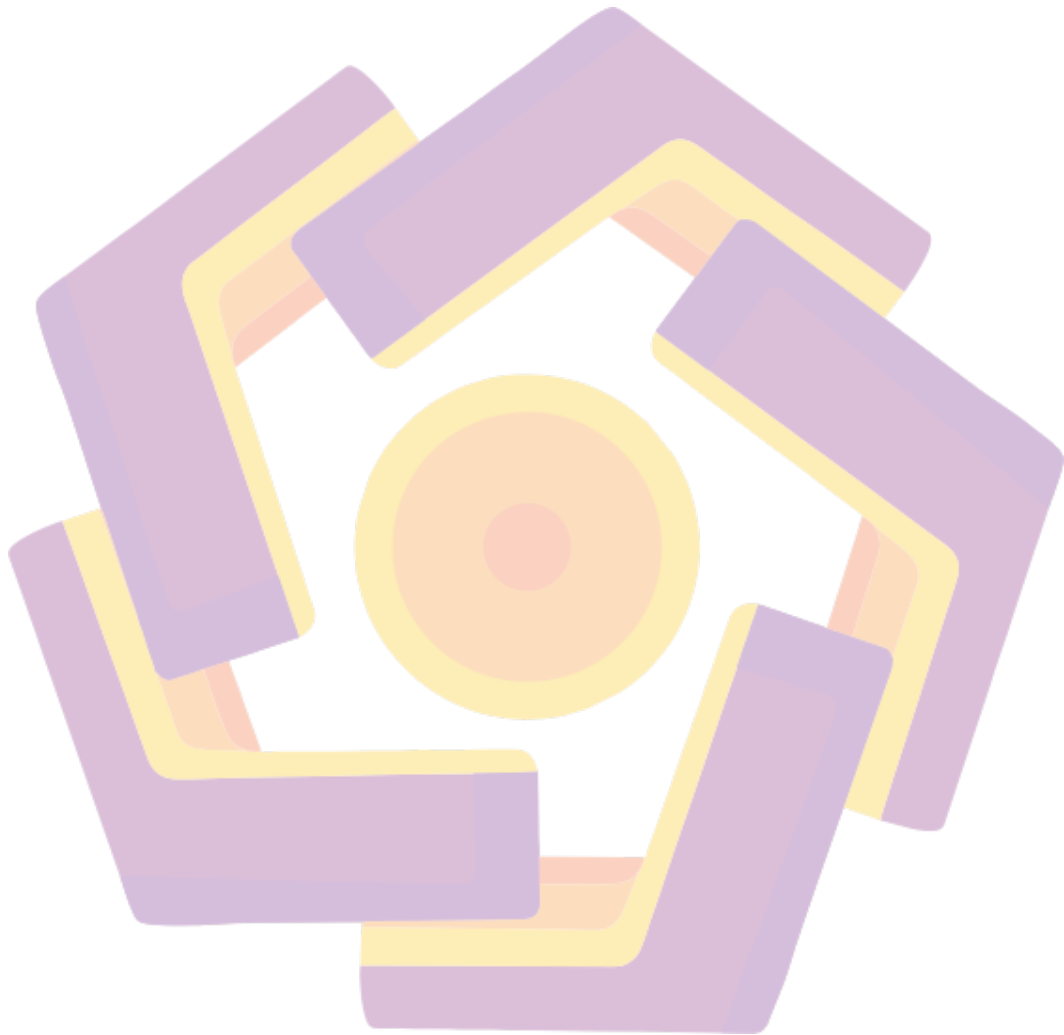
TABEL 4.12 HASIL KLASIFIKASI DECISION TREE DENGAN CROSS-VALIDATION FOLD 10 PADA DATASET	49
TABEL 4.13 CONFUSION MATRIX KLASIFIKASI KNN DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET DIULANG SEBANYAK 50 KALI	51
TABEL 4.14 HASIL KLASIFIKASI DECISION TREE DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET DIULANG SEBANYAK 50 KALI	51
TABEL 4.15 CONFUSION MATRIX KLASIFIKASI DECISION TREE DENGAN CROSS-VALIDATION FOLD 10 PADA DATASET	53
TABEL 4.16 HASIL KLASIFIKASI DECISION TREE DENGAN CROSS-VALIDATION FOLD 10 PADA DATASET	53
TABEL 4.17 CONFUSION MATRIX KLASIFIKASI DECISION TREE DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET DIULANG SEBANYAK 50 KALI	54
TABEL 4.18 HASIL KLASIFIKASI DECISION TREE DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET DIULANG SEBANYAK 50 KALI	55
TABEL 4.19 CONFUSION MATRIX KLASIFIKASI KNN DAN DECISION TREE DENGAN CROSS-VALIDATION FOLD 10 PADA DATASET	57
TABEL 4.20 CONFUSION MATRIX KLASIFIKASI KNN DAN DECISION TREE DENGAN 80% DATA TRAINING DAN 20% DATA TESTING PADA DATASET DIULANG SEBANYAK 50 KALI	58

DAFTAR GAMBAR

GAMBAR 3.1 ALUR PENELITIAN.....	26
GAMBAR 3.2 ALUR OVERSAMPLING MENGGUNAKAN SMOTE.....	27
GAMBAR 3.3 ILUSTRASI SEBARAN DATA.....	29
GAMBAR 3.4 PENGUJIAN AKURASI DATA IMBALANCE DENGAN ALGORITMA KNN.	36
GAMBAR 3.5 PENGUJIAN AKURASI DATA HASIL SMOTE DENGAN ALGORITMA KNN	36
GAMBAR 3.6 PENGUJIAN AKURASI DATA IMBALANCE DENGAN ALGORITMA DECISION TREE.....	37
GAMBAR 3.7 PENGUJIAN AKURASI DATA HASIL SMOTE DENGAN ALGORITMA DECISION TREE.....	37
GAMBAR 4.1 Ilustrasi dataset.....	39
GAMBAR 4.2 ILUSTRASI DATASET SETELAH OVERSAMPLING.....	41

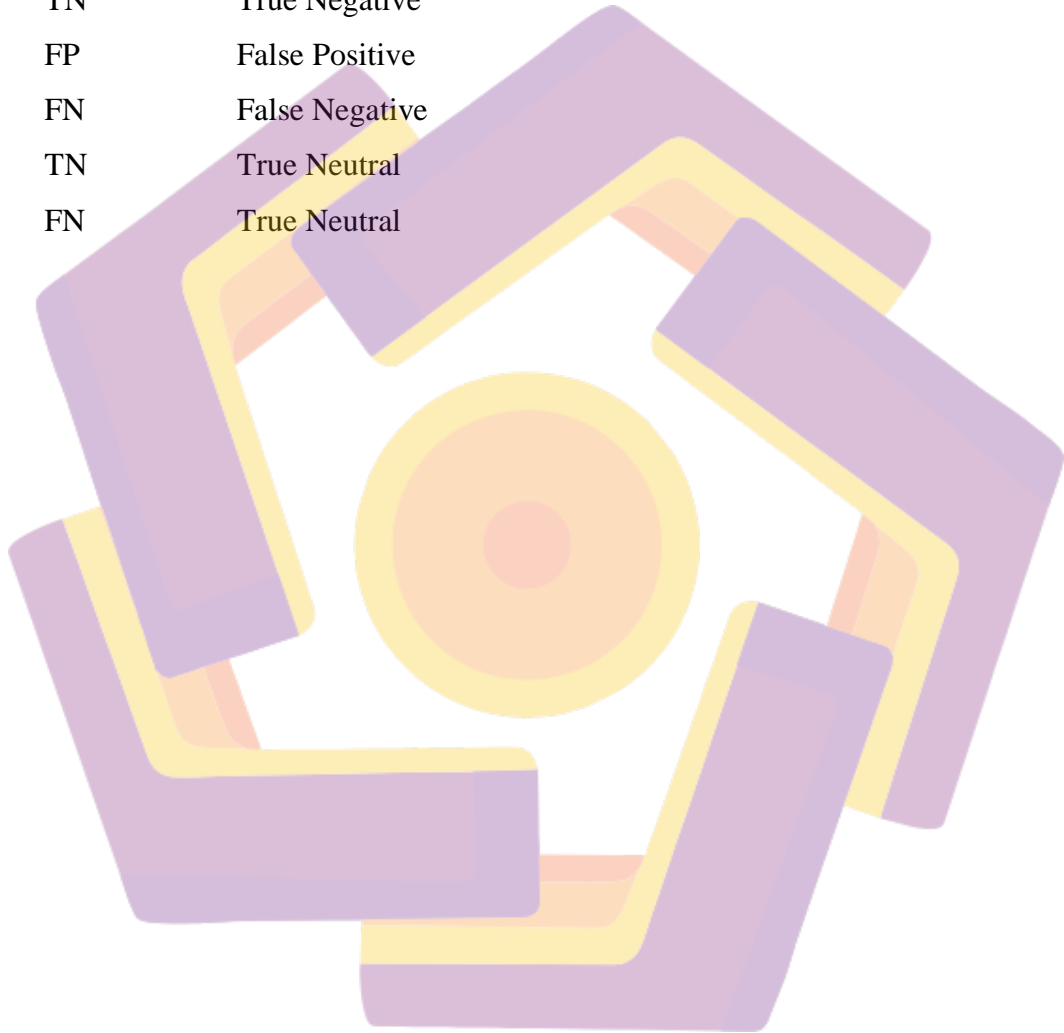
DAFTAR LAMPIRAN

LAMPIRAN 1 TAMPILAN GRAFIK DATASET SEBELUM DAN SESUDAH DILAKUKAN SMOTE.....	64
LAMPIRAN 2 DATASET YANG DIGUNAKAN DIDALAM PENELITIAN.....	64



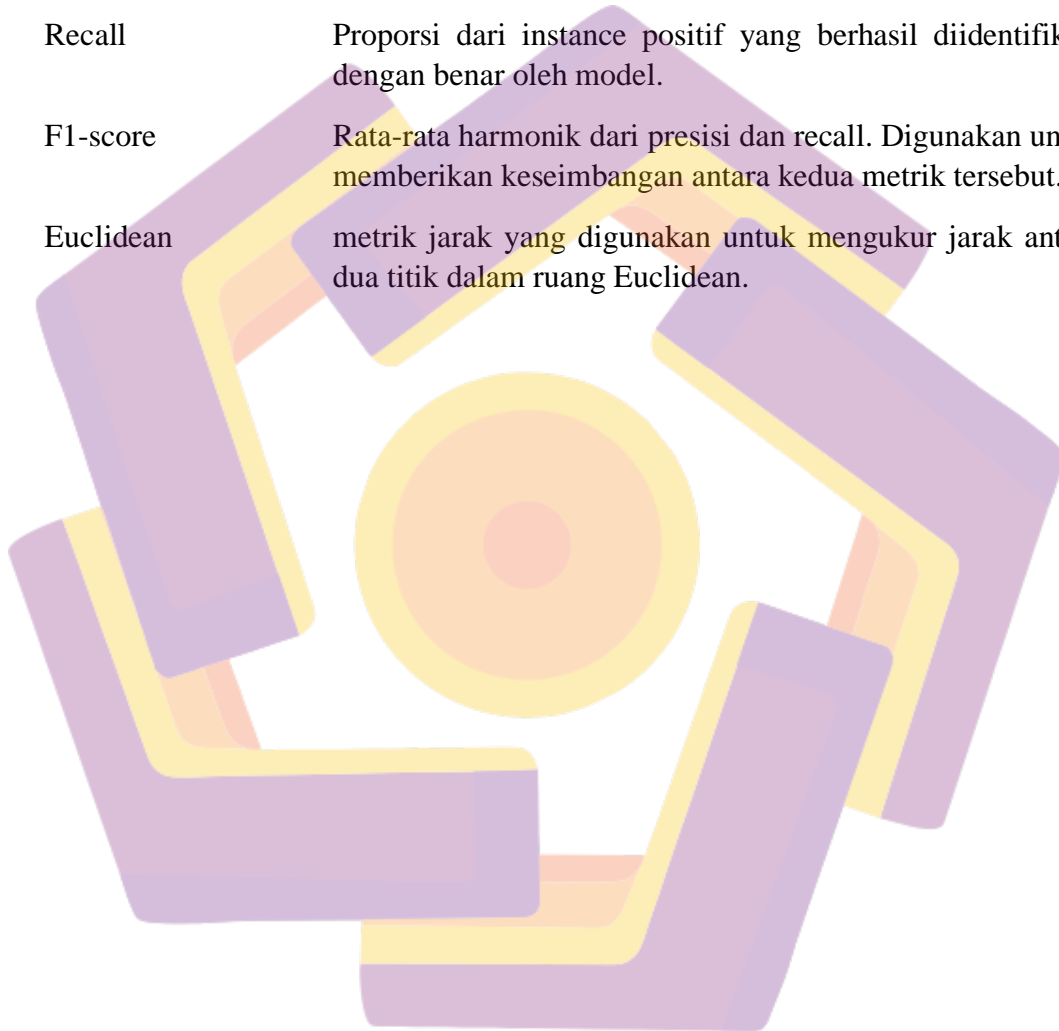
DAFTAR LAMBANG DAN SINGKATAN

KNN	K-Nearest Neighbors
SMOTE	Synthetic Minority Over-sampling Technique
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
TN	True Neutral
FN	True Neutral



DAFTAR ISTILAH

Akurasi	Seberapa sering model mengidentifikasi dengan benar kelas dari semua prediksi yang dibuatnya.
Presisi	Proporsi dari prediksi positif yang sebenarnya benar terhadap semua prediksi positif yang dilakukan oleh model.
Recall	Proporsi dari instance positif yang berhasil diidentifikasi dengan benar oleh model.
F1-score	Rata-rata harmonik dari presisi dan recall. Digunakan untuk memberikan keseimbangan antara kedua metrik tersebut.
Euclidean	metrik jarak yang digunakan untuk mengukur jarak antara dua titik dalam ruang Euclidean.



INTISARI

Penelitian ini bertujuan untuk mengeksplorasi penggunaan metode K-Nearest Neighbors (KNN) dan Decision Tree dalam analisis sentimen pelanggan Tokopedia, dengan fokus pada penanganan distribusi kelas yang tidak seimbang menggunakan teknik oversampling, khususnya Synthetic Minority Over-sampling Technique (SMOTE). Dataset awal terdiri dari 4060 sampel dengan tiga kelas sentimen: positif (2903), netral (585), dan negatif (572). Dengan menerapkan SMOTE, distribusi kelas yang tidak seimbang diperbaiki sehingga masing-masing kelas memiliki 2903 instance.

Hasil eksperimen menunjukkan bahwa penggunaan KNN dan Decision Tree setelah penerapan SMOTE menghasilkan peningkatan signifikan dalam akurasi dan metrik evaluasi seperti presisi, recall, dan F1-score untuk semua kelas. Akurasi KNN meningkat dari 70.2% menjadi 70.4% setelah SMOTE, sedangkan Decision Tree meningkat dari 71.2% menjadi 73.8%. Evaluasi tambahan dengan pembagian data training 80% dan testing 20% menunjukkan peningkatan konsisten dalam performa klasifikasi setelah SMOTE.

Dalam konteks analisis sentimen pelanggan, temuan ini mengindikasikan bahwa SMOTE merupakan pendekatan yang efektif untuk meningkatkan performa algoritma klasifikasi dalam mengatasi masalah distribusi kelas yang tidak seimbang. Implikasinya, penggunaan teknik oversampling seperti SMOTE dapat diterapkan secara luas untuk meningkatkan akurasi dan ketepatan analisis sentimen berbasis data dalam berbagai konteks bisnis dan teknologi.

Kata kunci: SMOTE, K-Nearest Neighbors, Decision Tree, Ketidakseimbangan Kelas.

ABSTRACT

This study aims to explore the use of K-Nearest Neighbors (KNN) and Decision Tree methods in analyzing customer sentiment at Tokopedia, with a focus on handling imbalanced class distributions using oversampling techniques, particularly Synthetic Minority Over-sampling Technique (SMOTE). The initial dataset consisted of 4060 samples categorized into three sentiment classes: positive (2903), neutral (585), and negative (572). By applying SMOTE, the imbalanced class distribution was adjusted so that each class had 2903 instances.

Experimental results indicate that the use of KNN and Decision Tree after applying SMOTE resulted in a significant increase in accuracy and evaluation metrics such as precision, recall, and F1-score for all classes. KNN accuracy improved from 70.2% to 70.4% after SMOTE, while Decision Tree accuracy increased from 71.2% to 73.8%. Additional evaluation using an 80% training and 20% testing data split consistently showed improved classification performance after SMOTE.

In the context of customer sentiment analysis, these findings suggest that SMOTE is an effective approach to enhance the performance of classification algorithms in addressing imbalanced class distribution issues. Consequently, the use of oversampling techniques like SMOTE can be widely applied to improve the accuracy and precision of data-driven sentiment analysis across various business and technological contexts..

Keyword: SMOTE, K-Nearest Neighbors, Decision Tree, Class Imbalance.