

**Implementasi Algoritma SMOTE Untuk Penanganan Masalah Imbalance
Class Pada Klasifikasi Pengobatan Kutil Kulit Dengan Metode
Immunotherapy Menggunakan Algoritma Naïve Bayes**

SKRIPSI



disusun oleh

Yan Hairul Firdan

17.11.1754

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

**Implementasi Algoritma SMOTE Untuk Penanganan Masalah Imbalance
Class Pada Klasifikasi Pengobatan Kutil Kulit Dengan Metode
Immunotherapy Menggunakan Algoritma Naïve Bayes**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Yan Hairul Firdan

17.11.1754

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

PERSETUJUAN**SKRIPSI**

Implementasi Algoritma SMOTE Untuk Penanganan Masalah Imbalance Class Pada Klasifikasi Pengobatan Kutil Kulit Dengan Metode Immunotherapy Menggunakan Algoritma Naïve Bayes

yang dipersiapkan dan disusun oleh

Yan Hairul Firdan

17.11.1754

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 2 Oktober 2020

Dosen Pembimbing,

Hastari Utama, M.Cs.

NIK. 190302187

PENGESAHAN

SKRIPSI

Implementasi Algoritma SMOTE Untuk Penanganan Masalah Imbalance Class Pada Klasifikasi Pengobatan Kutil Kulit Dengan Metode Immunotherapy Menggunakan Algoritma Naïve Bayes

yang dipersiapkan dan disusun oleh

Yan Hairul Firdan

17.11.1754

telah dipertahankan di depan Dewan Penguji

pada tanggal 15 Januari 2021

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Robert Marco, ST. MT

NIK. 0510048202

Sharazita Dyah Anggita, M.kom

NIK. 0562118801

Hastari Utama, M.Cs.

NIK. 0520058721

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

Tanggal 15 Januari 2021

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.

Bima, 14 januari 2021



Yan hairul Firdan

NIM. 17.11.1754

MOTTO

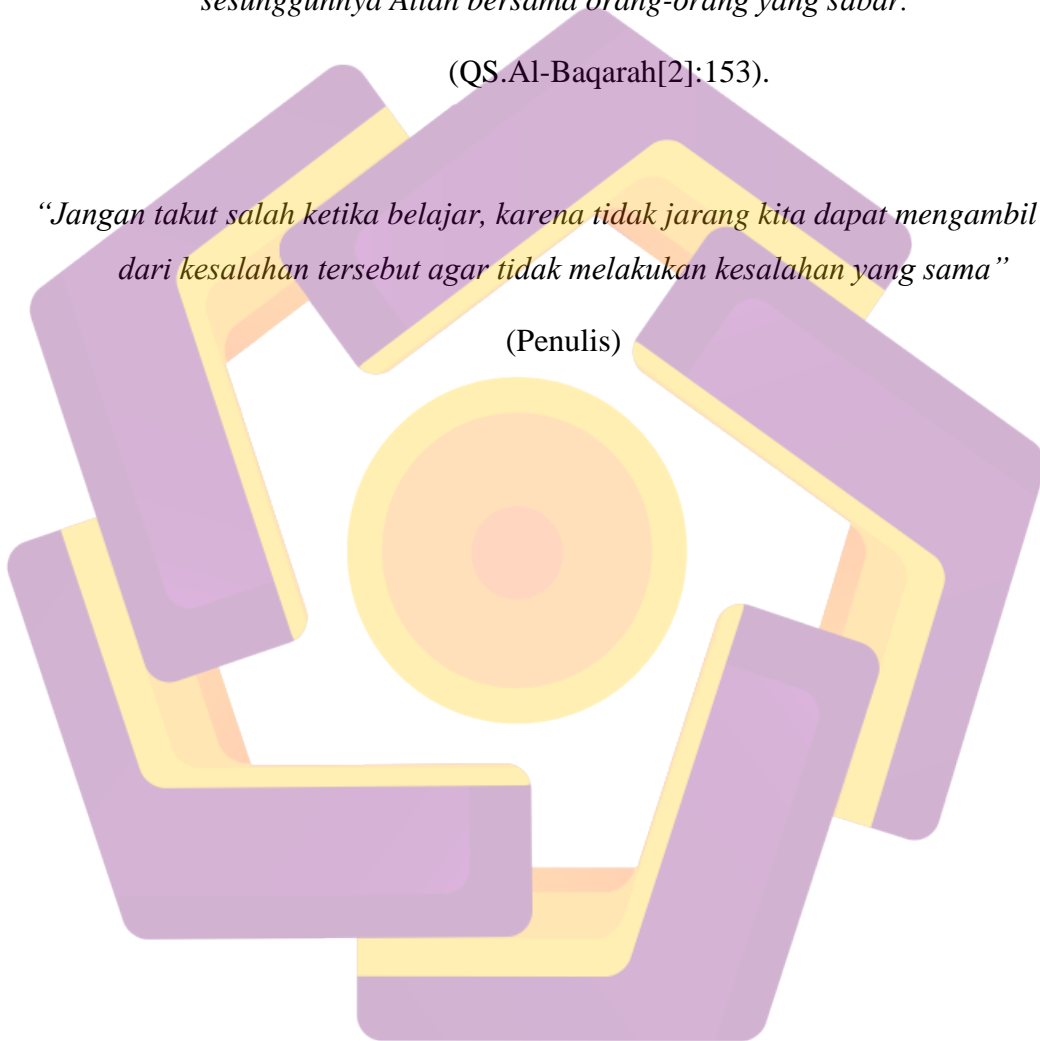
Bismillahirrahmaanirrahiim,

“Hai orang-orang yang beriman, jadikanlah sabar dan sholat sebagai penolongmu, sesungguhnya Allah bersama orang-orang yang sabar.”

(QS.Al-Baqarah[2]:153).

“Jangan takut salah ketika belajar, karena tidak jarang kita dapat mengambil ilmu dari kesalahan tersebut agar tidak melakukan kesalahan yang sama”

(Penulis)



PERSEMBAHAN

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

1. Kedua orang tuaku, Bapak Abdul Latif dan Ibu Kartini, terima kasih banyak atas kasih sayang, doa serta dukungannya selama ini.
2. Kakakku Rahmat Muddin Amirullah serta adik-adikku Nurul Hidayah dan Zainul Arifin, terima kasih banyak atas dukungan serta motivasinya.
3. Teman-teman di jurusan Informatika angkatan 2017 khususnya kelas 13, terima kasih banyak atas kebersamaannya.
4. Kampus dan almamaterku tercinta.



KATA PENGANTAR

Bismillahirrahmaanirrahiim,

Assalamualaikum warahmatullahi wabarakatuh

Dengan mengucapkan segala puji atas kehadiran Allah SWT Tuhan yang Maha Pengasih lagi Maha Penyayang atas segala rahmat serta hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Implementasi Algoritma SMOTE Untuk Menangani Masalah Imbalance Class Pada Klasifikasi Pengobatan Kutil Kulit Dengan Metode Immunotherapy Menggunakan Algoritma Naïve Bayes” 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 :

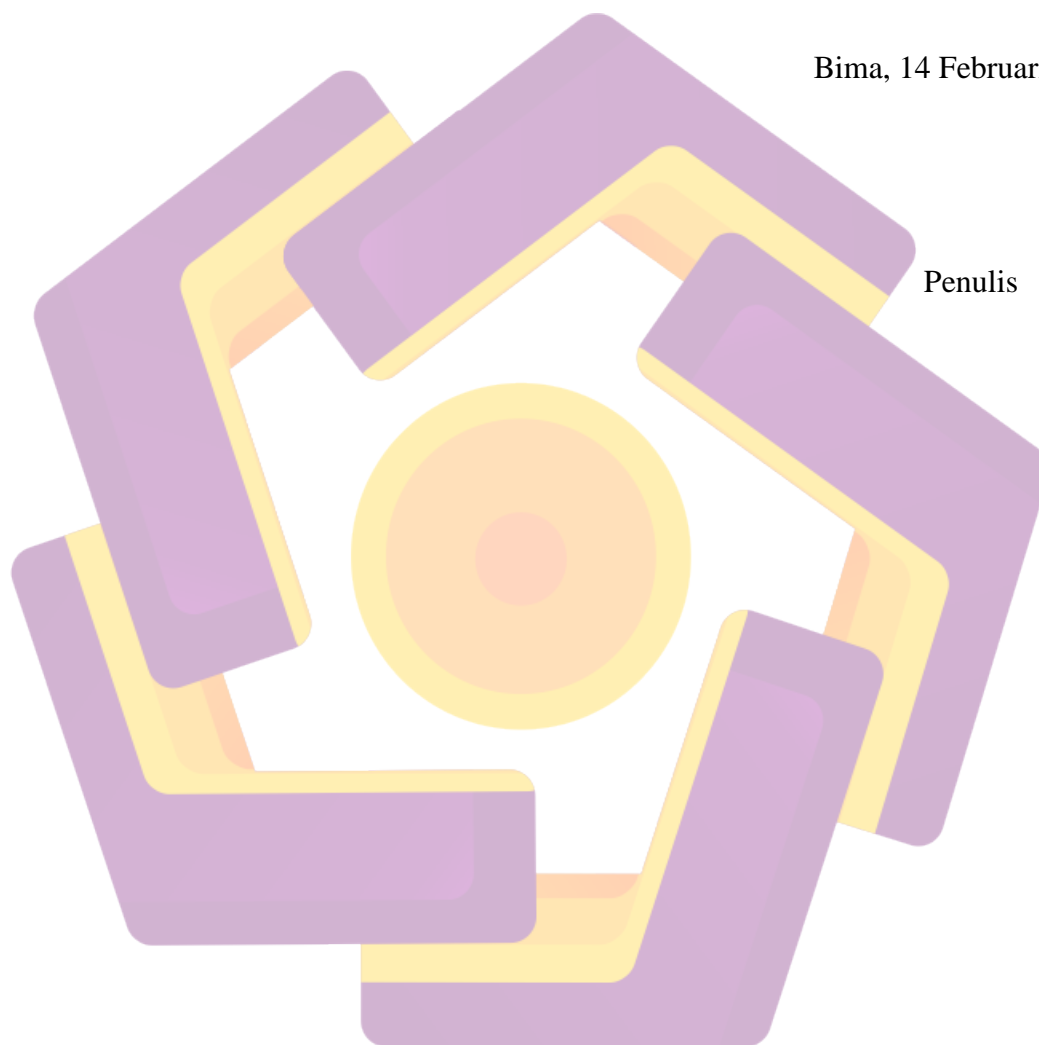
1. Ibu Krisnawati, S.Si, M.T, Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
2. Bapak Sudarmawan, M.T, Ketua Program Studi S1 Informatika Universitas AMIKOM Yogyakarta.
3. Bapak Hastari Utama, M.Cs, selaku dosen pembimbing penulis yang dengan sabarnya membimbing dan memberikan semangat serta 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 2017 khususnya kelas 13, terima kasih banyak atas kebersamaannya.

Penulis menyadari bahwa masih banyak kekurangan pada karya tulis ini. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun. Semoga

penelitian ini dapat bermanfaat khususnya bagi penulis sendiri dan bagi pembaca pada umumnya.

Waassalamualaikum Warahmatullahi Wabarakatuh

Bima, 14 Februari 2021



Penulis

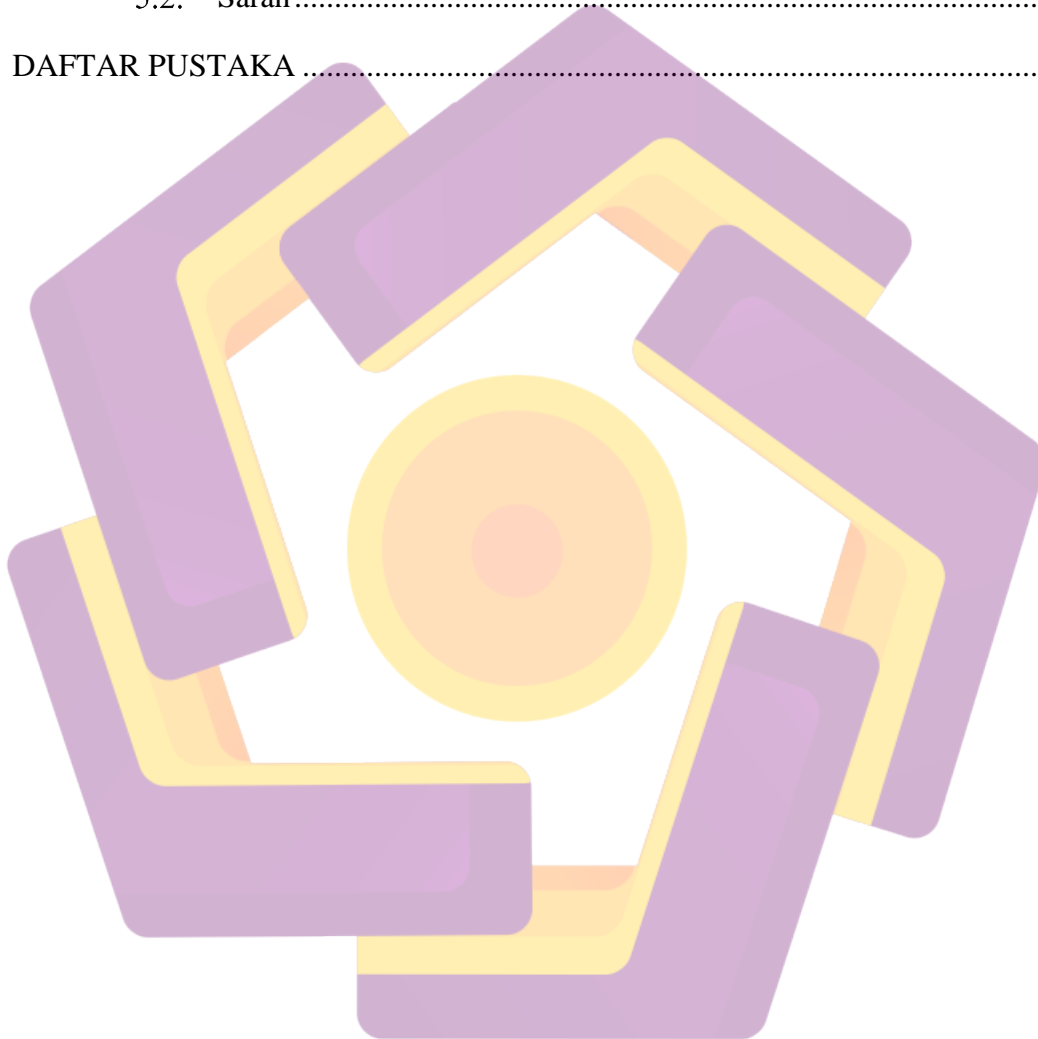
DAFTAR ISI

COVER	I
PERSETUJUAN	II
PENGESAHAN	III
MOTTO	V
PERSEMBAHAN	VI
KATA PENGANTAR	VII
DAFTAR ISI	IX
DAFTAR TABEL	XIII
DAFTAR GAMBAR	XIV
INTISARI	XV
ABSTRACT	XVI
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Batasan masalah	3
1.4. Maksud dan Tujuan Penelitian	3
1.5. Manfaat Penelitian	3
1.6. Metode Penelitian	3
1.6.1. Metode Pengumpulan Data	3
1.6.2. Preprocessing	3
1.6.3. Klasifikasi	4
1.6.4. Evaluasi	4

1.7.	Sistematika Penulisan.....	4
1.7.1.	BAB I Pendahuluan	4
1.7.2.	BAB II Landasan Teori.....	4
1.7.3.	BAB III Metodologi Penelitian.....	4
1.7.4.	BAB IV Hasil Implementasi dan Pembahasan	4
1.7.5.	BAB V Penutup	4
1.7.6.	Daftar Pustaka.....	5
1.7.7.	Lampiran	5
BAB II LANDASAN TEORI.....		6
2.1.	Kajian Pustaka.....	6
2.2.	Landasan Teori	12
2.2.1.	Imbalance Class	12
2.2.2.	Synthetic Minority Oversampling Technique (SMOTE) ..	12
2.2.3.	Klasifikasi	15
2.2.4.	Naïve Bayes	15
2.2.5.	Confusion Matrix.....	18
2.2.6.	K-fold cross-validation	19
BAB III METODE PENELITIAN.....		21
3.1.	Studi literatur	22
3.2.	Dataset	22
3.3.	Data Processing	24
3.3.1.	Oversampling.....	24
3.3.2.	Transformasi atribut.....	25

3.4.	Klasifikasi Naive Bayes	26
3.5.	Skenario pengujian	26
3.6.	Evaluasi performa.....	28
BAB IV HASIL DAN PEMBAHASAN		31
4.1.	Dataset	31
4.2.	Data hasil preprocessing.....	32
4.2.1.	Oversampling.....	32
4.2.2.	Transformasi atribut.....	33
4.3.	Pengujian.....	34
4.3.1.	Pengujian pada data awal.....	35
4.3.2.	Pengujian pada data hasil Oversampling 200%	36
4.3.2.1.	Jumlah K-NN = 1.....	36
4.3.2.2.	Jumlah K-NN = 2.....	37
4.3.2.3.	Jumlah K-NN = 3.....	38
4.3.2.4.	Jumlah K-NN = 4.....	39
4.3.2.5.	Jumlah K-NN = 5.....	40
4.3.2.6.	Jumlah K-NN = 6.....	41
4.3.2.7.	Jumlah K-NN = 7.....	42
4.4.	Evaluasi dan perbandingan performa	45
4.4.1.	Akurasi.....	46
4.4.2.	Sensitivity atau recall.....	47
4.4.3.	Specifity.....	47
4.4.4.	Precision.....	47

4.4.5. F-measure.....	48
BAB V KESIMPULAN DAN SARAN.....	49
5.1. Kesimpulan.....	49
5.2. Saran.....	50
DAFTAR PUSTAKA.....	51

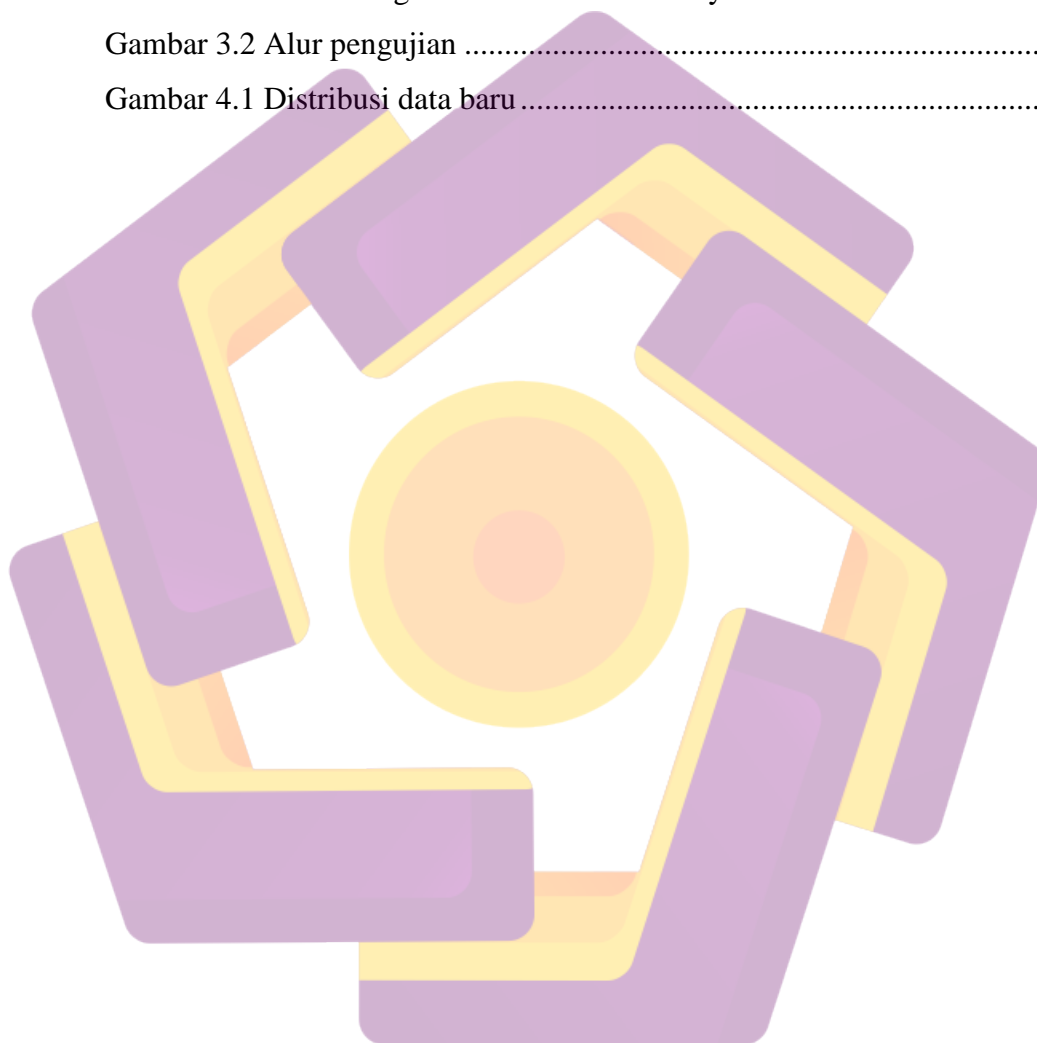


DAFTAR TABEL

Tabel 2.1. Perbandingan penelitian.....	8
Tabel 2.2 <i>Confusion matrix</i>	19
Tabel 3.1 <i>Confusion matrix</i>	29
Tabel 4.1 Data awal	31
Tabel 4.2 Data hasil oversampling	32
Tabel 4.3 Data hasil transformasi atribut.....	33
Tabel 4.4 <i>Confusion matrix</i> pada pengujian data awal.....	35
Tabel 4.5 <i>Confusion matrix</i> pada data baru dengan K-NN = 1	36
Tabel 4.6 <i>Confusion matrix</i> pada data baru dengan K-NN = 2	37
Tabel 4.7 <i>Confusion matrix</i> pada data baru dengan K-NN = 3	38
Tabel 4.8 <i>Confusion matrix</i> pada data baru dengan K-NN = 4	39
Tabel 4.9 <i>Confusion matrix</i> pada data baru dengan K-NN = 5	40
Tabel 4.10 <i>Confusion matrix</i> pada data baru dengan K-NN = 6	41
Tabel 4.11 <i>Confusion matrix</i> pada data baru dengan K-NN = 7	43
Tabel 4.12 Pengujian skema SMOTE dengan Naïve Bayes.....	43
Tabel 4.13 Hasil pengujian menggunakan <i>Weka Tools</i>	44
Tabel 4.14 Pengujian skema SMOTE dan SVM	46

DAFTAR GAMBAR

Gambar 2.1. Ilustrasi SMOTE	13
Gambar 2.2 Proses 5- <i>cross-validation</i>	20
Gambar 3.1 Perbandingan data antara kelas mayoritas dan minoritas	23
Gambar 3.2 Alur pengujian	28
Gambar 4.1 Distribusi data baru	33



INTISARI

Imbalance class merupakan masalah yang sering kali muncul di dalam bidang data mining. *Imbalance class* adalah kondisi dimana distribusi jumlah *instance class* memiliki perbedaan yang sangat besar sehingga terdapat class yang memiliki *instance* yang banyak disebut dengan *majority class* dan class yang memiliki jumlah *instance* yang sedikit disebut dengan *minority class*. Dengan adanya ketimpangan dan perbedaan jumlah *instance* yang sangat besar ini, menyebabkan algoritma-algoritma klasifikasi menghasilkan akurasi yang tinggi ketika mengklasifikasikan kelas mayoritas dan menghasilkan akurasi yang rendah ketika mengklasifikasikan kelas minoritas.

Pada penelitian ini akan mengimplementasikan algoritma SMOTE untuk melakukan *oversampling* data dan algoritma *naïve bayes* untuk melakukan klasifikasi pada data yang telah di *oversampling*. Pengujian akan menggunakan skema *5-fold cross-validation*.

Pengujian dengan skema *naïve bayes* menghasilkan akurasi sebesar 0.733 atau 73.33%, *sensitivity* sebesar 0.2105 atau 21.05%, *specificity* sebesar 0.8732 atau 87.32%, *recall* sebesar 0.2105 atau 21.05%, *precision* sebesar 0.3076 atau 30.76% dan *F-measure* sebesar 0.2499 atau 24.99%. Pengujian dengan skema *naïve bayes* dan SMOTE skor akurasi, *specificity* dan *precision* tertinggi diraih oleh kombinasi *oversampling* 200% dan jumlah KNN yang dipilih sebanyak satu data yaitu sebesar 0.7178 atau 71.78%, 0.7042 atau 70.42% dan 0.6666 atau 66.66%. Sedangkan skor *recall* dan *f-measure* tertinggi didapatkan dengan menggunakan KNN sebanyak tujuh data yaitu sebesar 0.8596 atau 85.96% dan 0.7259 atau 72.59%. Hal ini menunjukkan implementasi algoritma SMOTE berhasil menangani masalah *imbalance class* dan hanya mampu meningkatkan *precision*, *recall* dan *f-measure*.

Kata kunci : *Imbalance class*, SMOTE, klasifikasi, *oversampling*, performa

ABSTRACT

Class imbalance is a problem that often arises in the field of data mining. Imbalance class is a condition where the distribution of the number of class instances has a very big difference so that there are classes that have many instances called the majority class and classes that have a small number of instances are called the minority class. The inequality and differences in the number of instances are very large, causing the classification algorithms to produce high accuracy when classifying the majority class and low accuracy when classifying the minority class.

This research will implement the SMOTE algorithm for oversampling data and the naïve Bayes algorithm to classify the oversampled data. The test will use a 5-fold cross-validation scheme.

Testing with the naïve Bayes scheme resulted in an accuracy of 0.733 or 73.33%, sensitivity of 0.2105 or 21.05%, specificity of 0.8732 or 87.32%, recall of 0.2105 or 21.05%, precision of 0.3076 or 30.76% and F-measure of 0.2499 or 24.99%. Testing with the naïve Bayes scheme and SMOTE, the highest accuracy, specificity and precision scores were achieved by a combination of 200% oversampling and the number of KNN selected by one data, namely 0.7178 or 71.78%, 0.7042 or 70.42% and 0.6666 or 66.66%. while the highest recall and f-measure scores were obtained using the KNN with seven data, namely 0.8596 or 85.96% and 0.7259 or 72.59%. This shows that the implementation of the SMOTE algorithm is successful in handling the imbalance class problem and is only able to increase the precision, recall and f-measure.

Keywords : *Imbalance class, SMOTE, classification , oversampling, performance*