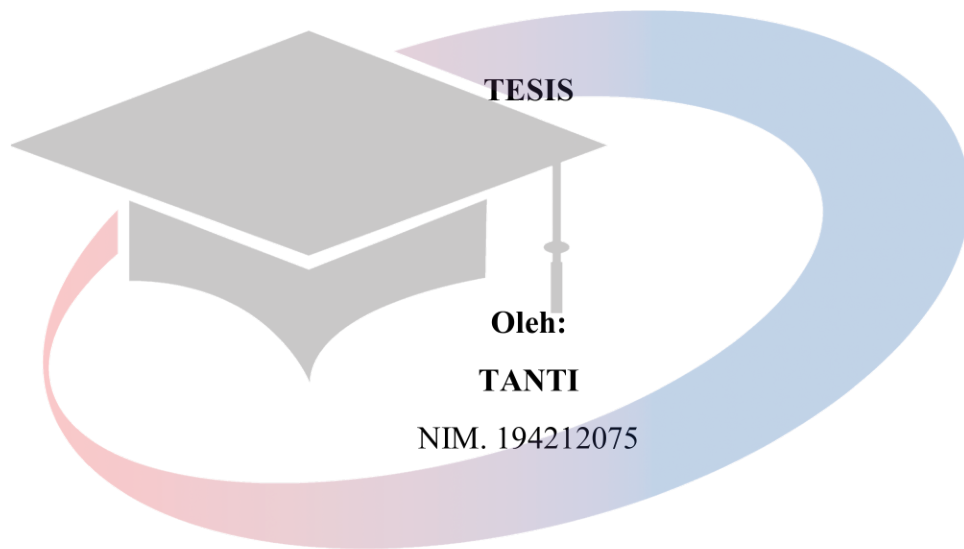


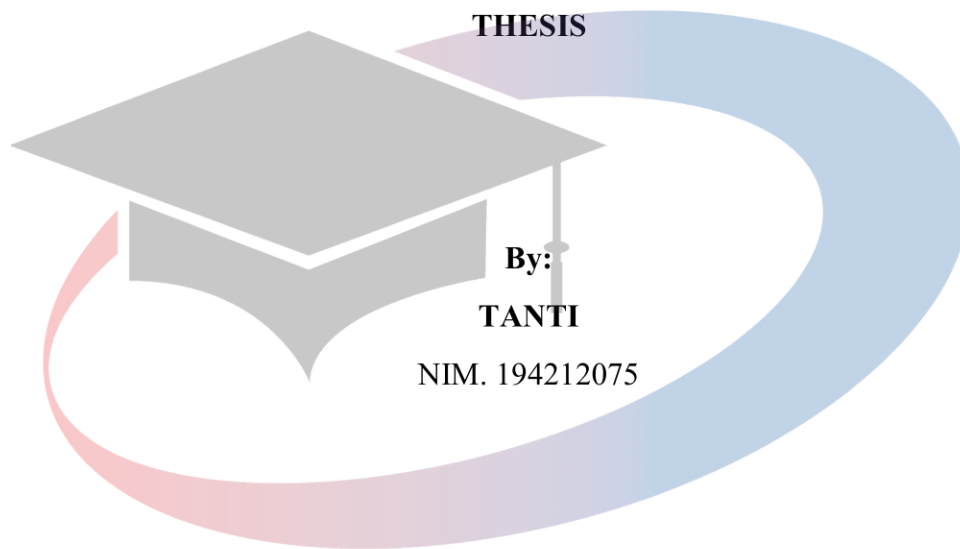
**OPTIMALISASI KINERJA KLASIFIKASI MELALUI
SELEKSI FITUR DAN ADABOOST DALAM
PENANGANAN KETIDAKSEIMBANGAN
KELAS**



UNIVERSITAS
MIKROSKIL

**PROGRAM STUDI MAGISTER TEKNOLOGI INFORMASI
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
MIKROSKIL
MEDAN
2021**

**OPTIMIZATION OF CLASSIFICATION PERFORMANCE
THROUGH FEATURE SELECTION AND ADABOOST
IN HANDLING OF CLASS IMBALANCE**



**INFORMATION TECHNOLOGY MASTER'S PROGRAM
SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
MIKROSKIL
MEDAN
2021**

LEMBAR PERNYATAAN

Saya yang membuat pernyataan ini adalah mahasiswa Jurusan/Program Studi S-2 Magister Teknologi Informasi STMIK Mikroskil Medan dengan identitas mahasiswa sebagai berikut:

Nama : Tanti
Nim : 194212075
Peminatan : Teknologi Informasi

Saya telah melaksanakan penelitian dan penulisan Tesis dengan judul "OPTIMALISASI KINERJA KLASIFIKASI MELALUI SELEKSI FITUR DAN ADABOOST DALAM PENANGANAN KETIDAKSEIMBANGAN KELAS", dengan ini saya menyatakan dengan sebenar-benarnya bahwa penelitian dan penulisan Tesis tersebut merupakan hasil karya saya sendiri (tidak meminta orang lain untuk mengerjakannya) dan semua sumber, baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bila dikemudian hari ternyata terbukti bahwa bukan saya yang mengerjakannya, maka saya bersedia dikenakan sanksi yang telah ditetapkan oleh STMIK Mikroskil Medan, yakni pencabutan ijazah yang telah saya terima dan ijazah tersebut dinyatakan tidak sah.

Selain itu, demi pengembangan ilmu pengetahuan, saya menyetujui untuk memberikan kepada STMIK Mikroskil Medan Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty Free Right*) atas Tesis saya beserta perangkat yang ada (jika diperlukan). Dengan hak ini, STMIK Mikroskil Medan berhak menyimpan, mengalihmedia/formatkan, mengelolah dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan Tesis saya secara keseluruhan atau hanya elektronik, selama tetap mencantumkan nama saya sebagai penulis/pencipta dan pemilik hak cipta. Menyatakan juga bahwa saya akan mempertahankan hak eksklusif saya untuk menggunakan seluruh atau sebagian isi Tesis saya guna pengembangan karya di masa depan, misalnya dalam bentuk artikel, buku, ataupun perangkat lunak.

Demikian pernyataan ini saya perbuat dengan sungguh-sungguh, dalam keadaan sadar dan tanpa ada tekanan dari pihak manapun.

Medan, 23 Juli 2021

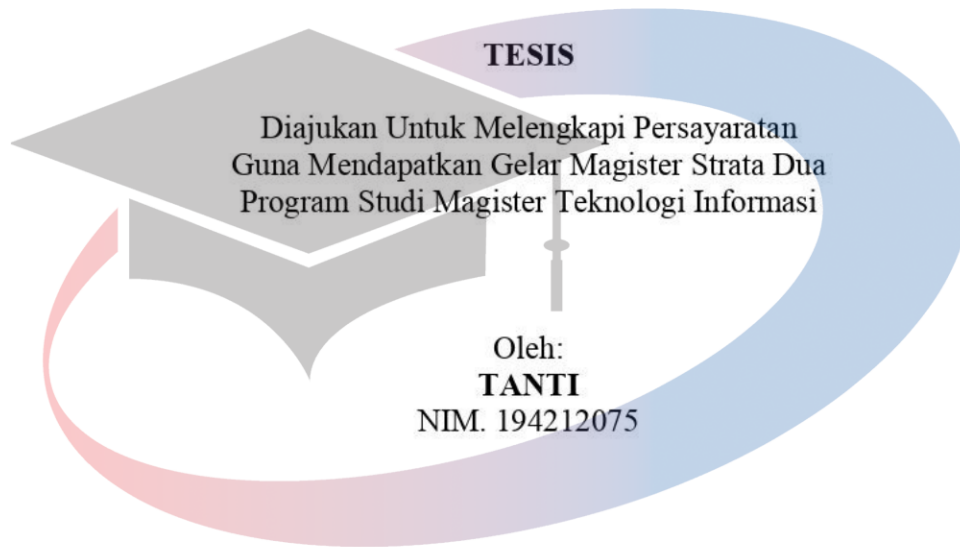
Saya yang membuat pernyataan,



Tanti

LEMBARAN PENGESAHAN

OPTIMALISASI KINERJA KLASIFIKASI MELALUI SELEKSI FITUR DAN ADABOOST DALAM PENANGANAN KETIDAKSEIMBANGAN KELAS



TESIS

Diajukan Untuk Melengkapi Persyaratan
Guna Mendapatkan Gelar Magister Strata Dua
Program Studi Magister Teknologi Informasi

Oleh:
TANTI
NIM. 194212075

Disetujui Oleh:
Dosen Pembimbing,

A handwritten signature in black ink, appearing to read 'Pahala Sirait', written over the large blue watermark text.

Dr. Pahala Sirait, S.T., M.Kom.

Medan, 23 Juli 2021

Diketahui dan Disahkan Oleh:

Ketua Program Studi
Magister Teknologi Informasi,



Dr. Ronsen Purba, M.Sc.

ABSTRAK

Salah satu masalah dalam klasifikasi data mining adalah ketidakseimbangan kelas, dimana jumlah *instance* pada kelas mayoritas lebih banyak dibanding kelas minoritas. Dalam proses klasifikasi, kelas minoritas sering salah diklasifikasikan, karena *machine learning* memprioritaskan kelas mayoritas dan mengabaikan kelas minoritas sehingga hal ini dapat menyebabkan kinerja klasifikasi menjadi tidak optimal. Tujuan dari penelitian ini adalah memberikan solusi dalam mengatasi ketidakseimbangan kelas sehingga dapat mengoptimalkan kinerja klasifikasi dengan menggunakan *chi-square* dan *adaboost* pada salah satu algoritma klasifikasi yaitu C5.0. Pada penelitian ini, kelas mayoritas dalam dataset yang digunakan didominasi oleh kelas negatif, jadi penilaian kinerja harus lebih berfokus kepada kelas positif. Oleh karena itu, penilaian yang lebih sesuai adalah *recall/sensitivity/TPR* karena nilai yang dihasilkan hanya bergantung pada kelas positif. Hasil penelitian menunjukkan bahwa kedua metode mampu meningkatkan nilai *recall/sensitivity/TPR* artinya penerapan *chi-square* dan *adaboost* mampu untuk meningkatkan kinerja klasifikasi dari kelas minoritas.

Kata Kunci: Ketidakseimbangan Kelas, Klasifikasi, C5.0, *Chi-Square*, *AdaBoost*

UNIVERSITAS
MIKROSKIL

ABSTRACT

One of the problems in data mining classification is class imbalance, where the number of instances in the majority class is more than the minority class. In the classification process, minority classes are often misclassified, because machine learning prioritizes the majority class and ignores the minority class so that this can cause the classification performance to be not optimal. The purpose of this study is to provide a solution to overcome class imbalances so as to optimize classification performance using chi-square and adaboost on one of the classification algorithms, namely C5.0. In this study, the majority class in the dataset used is dominated by the negative class, so the performance appraisal should focus more on the positive class. Therefore, a more suitable assessment is recall/sensitivity/TPR because the resulting value only depends on the positive class. The results showed that both methods were able to increase the recall/sensitivity/TPR value, meaning that the application of chi-square and adaboost was able to improve the classification performance of the minority class.

Keyword: Class Imbalance, Classification, C5.0, Chi-Square, AdaBoost



UNIVERSITAS
MIKROSKIL

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa karena atas berkat dan rahmat-Nya, penulis dapat menyelesaikan Tesis yang berjudul “Optimalisasi Kinerja Klasifikasi Melalui Seleksi Fitur dan AdaBoost dalam Penanganan Ketidakseimbangan Kelas”.

Tesis ini dibuat untuk melengkapi persyaratan kurikulum pada Program Studi Magister Teknologi Informasi Strata Dua, STMIK Mikroskil Medan.

Penulis menyampaikan banyak terima kasih kepada:

1. Bapak Dr. Pahala Sirait, S.T., M.Kom., selaku Dosen Pembimbing dan Ketua STMIK Mikroskil Medan yang telah membimbing penulis dalam proses penyelesaian tesis ini.
2. Bapak Andri, S.Kom., M.T.I., selaku Dosen Pendamping Pembimbing yang telah membimbing penulis dalam proses penyelesaian tesis ini.
3. Bapak Dr. Ronsen Purba, M.Sc., selaku dosen penguji dan Ketua Program Studi Magister Teknologi Informasi yang telah memberikan saran dan masukannya.
4. Bapak Gunawan, S.Kom., M.T.I., selaku Wakil Ketua I STMIK Mikroskil Medan.
5. Bapak Ir. Erwin Setiawan Panjaitan, MMSI., Ph.D. selaku dosen penguji yang telah memberikan saran dan masukannya.
6. Bapak atau Ibu Dosen STMIK Mikroskil Medan yang telah membantu proses penulisan tesis ini.
7. Anggota keluarga, teman, saudara dan semua pihak yang terus memberikan dukungan penuh kepada penulis selama proses penulisan tesis ini.

Penulis menyadari bahwa masih terdapat banyak kekurangan yang ada. Oleh sebab itu, kritik dan saran yang bersifat membangun akan sangat diterima. Akhir kata, semoga Tesis ini dapat bermanfaat bagi masyarakat. Terima kasih.

Medan, 23 Juli 2021

Penulis

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR	iii
DAFTAR ISI	iv
DAFTAR GAMBAR	vi
DAFTAR TABEL	vii
DAFTAR LAMPIRAN	viii
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Masalah Penelitian	3
1.2.1. Identifikasi Masalah	3
1.2.2. Rumusan Masalah	4
1.3. Tujuan dan Manfaat Penelitian	4
1.3.1. Tujuan Penelitian	4
1.3.2. Manfaat Penelitian	4
1.4. Ruang Lingkup (Batasan Masalah)	5
1.5. Metodologi Penelitian	5
1.6. Sistematika Penulisan	6
BAB II KAJIAN LITERATUR	7
2.1. Tinjauan Pustaka	7
2.1.1. Data Mining	7
2.1.2. Seleksi Fitur	9
2.1.3. Ketidakseimbangan Kelas	11
2.1.4. Klasifikasi	15
2.1.5. Evaluasi	18
2.1.6. Penelitian Terdahulu	20
2.2. Kerangka Konsep/Pola Pikir Pemecahan Masalah	22
BAB III METODOLOGI PENELITIAN	24
3.1. Analisis Masalah	24
3.2. Metode Penelitian	25
3.2.1. Pengumpulan Data	26

3.2.2.	Preprocessing	28
3.2.3.	Model yang Diusulkan	31
3.2.4.	Evaluasi	35
3.3.	Alat-alat Penelitian	35
3.4.	Teknik Analisis.....	35
BAB IV HASIL DAN PEMBAHASAN		37
4.1.	Hasil.....	37
4.1.1.	Seleksi Fitur dengan Chi-Square.....	37
4.1.2.	Klasifikasi	39
4.2.	Pembahasan	47
4.2.1.	Perbandingan Accuracy	48
4.2.2.	Perbandingan Precision.....	50
4.2.3.	Perbandingan Recall/Sensitivity/TPR.....	53
4.2.4.	Perbandingan Selectivity/Specificity/TNR.....	55
4.2.5.	Perbandingan F1 Score	58
BAB V PENUTUP.....		62
5.1.	Kesimpulan.....	62
5.2.	Saran.....	62
DAFTAR PUSTAKA		63

UNIVERSITAS MIKROSKIL

DAFTAR GAMBAR

Gambar II-1 Tahapan Proses KDD	8
Gambar II-2 Algoritma AdaBoost	15
Gambar II-3 Contoh Pohon Keputusan.....	16
Gambar II-4 Kerangka Konsep Pemecahan Masalah	23
Gambar III-1 Metode Penelitian	25
Gambar III-2 Model yang Diusulkan.....	32
Gambar IV-1 Perbandingan Accuracy	50
Gambar IV-2 Perbandingan Precision	52
Gambar IV-3 Perbandingan Recall/Sensitivity/TPR	55
Gambar IV-4 Perbandingan Selectivity/Specificity/TNR.....	57
Gambar IV-5 Perbandingan F1 Score	60



UNIVERSITAS
MIKROSKIL

DAFTAR TABEL

Tabel II-1 Confusion Matrix	18
Tabel III-1 Fitur Bertipe Numerik.....	26
Tabel III-2 Fitur Bertipe Kategorikal.....	27
Tabel III-3 Fitur Numerik Setelah Cleaning	29
Tabel III-4 Binning	29
Tabel IV-1 Keputusan Berdasarkan Nilai Chi-Square.....	37
Tabel IV-2 Peringkat Fitur Berdasarkan Nilai Chi-Square.....	39
Tabel IV-3 Kinerja C5.0	40
Tabel IV-4 Kinerja C5.0 dengan AdaBoost.....	41
Tabel IV-5 Kinerja C5.0 dengan Chi-Square.....	44
Tabel IV-6 Kinerja C5.0 dengan Chi-Square dan AdaBoost.....	45
Tabel IV-7 Perbandingan Accuracy.....	48
Tabel IV-8 Perbandingan Precision	50
Tabel IV-9 Perbandingan Recall/Sensitivity/TPR	53
Tabel IV-10 Perbandingan Selectivity/Specificity/TNR	55
Tabel IV-11 Perbandingan F1 Score.....	58

UNIVERSITAS
MIKROSKIL

DAFTAR LAMPIRAN

Lampiran 1 Proses Chi-Square dan Klasifikasi	68
Lampiran 2 Hasil Uji Chi-Square	71
Lampiran 3 DAFTAR RIWAYAT HIDUP	90



UNIVERSITAS MIKROSKIL