



**PENANGANAN DATA TIDAK SEIMBANG PADA MODEL KLASIFIKASI
C5.0 DAN CHI-SQUARE UNTUK PREDIKSI KARYAWAN RESIGN**

Skripsi

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

**Oleh
NANDA AHMAD REZA
NIM. 1811016310017**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2025**



**PENANGANAN DATA TIDAK SEIMBANG PADA MODEL KLASIFIKASI
C5.0 DAN CHI-SQUARE UNTUK PREDIKSI KARYAWAN RESIGN**

Skripsi

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

**Oleh
NANDA AHMAD REZA
NIM 1811016310017**

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

**JUNI 2025
SKRIPSI**

SKRIPSI

PENANGANAN DATA TIDAK SEIMBANG PADA MODEL KLASIFIKASI C5.0 DAN CHI-SQUARE UNTUK PREDIKSI KARYAWAN RESIGN

Oleh:

Nanda Ahmad Reza

NIM. 1811016310017

Telah dipertahankan di depan Dosen Penguji pada tanggal 30 Juni 2025.

Susunan Dosen Penguji:

Pembimbing I



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

NIP. 199006122019031013

Dosen Penguji I



Dwi Kartini, S.Kom., M.Kom.

NIP. 198704212012122003

Pembimbing II



Irwan Budiman, S.T., M.Kom.

NIP. 197703252008121001

Dosen Penguji II



Rudy Hekteno, S.Kom, M.Kom.

NIP. 198809252022031003

Banjarbaru, 30 Juni 2025

Koordinator Program Studi Ilmu Komputer



Dwi Kartini, S.Kom., M.Kom.

NIP. 198704212012122003

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 diberikan orang lain, kecuali yang secara tertulis diacu di dalam naskah ini dan disebutkan dalam Daftar Pustaka

Banjarbaru, 30 Juni 2025



Nanda Ahmad Reza
NIM. 1811016310017

ABSTRAK

PENANGANAN DATA TIDAK SEIMBANG PADA MODEL KLASIFIKASI C5.0 DAN CHI-SQUARE UNTUK PREDIKSI KARYAWAN RESIGN

(Oleh: Nanda Ahmad Reza; Pembimbing: Muhammad Itqan Mazdadi, S.Kom., M.Kom. dan Irwan Budiman, S.T., M.Kom.; 2025; 83 halaman)

Prediksi pengunduran diri karyawan merupakan isu penting bagi perusahaan yang ingin meminimalkan dampak negatif seperti berkurangnya produktivitas dan meningkatnya biaya rekrutmen. Studi ini menerapkan algoritma klasifikasi C5.0 yang dikombinasikan dengan pemilihan fitur *Chi-Square* dan *Synthetic Minority Over-sampling Technique* (SMOTE) untuk mengatasi masalah ketidakseimbangan kelas, di mana jumlah karyawan yang mengundurkan diri (kelas minoritas) secara signifikan lebih rendah daripada mereka yang bertahan (kelas mayoritas). SMOTE digunakan untuk menghasilkan data sintetis untuk kelas minoritas guna meningkatkan kemampuan model dalam mendeteksi potensi pengunduran diri. Kumpulan data dipartisi menjadi tiga rasio: 70:30, 80:20, dan 90:10. Hasil eksperimen menunjukkan bahwa penerapan SMOTE secara substansial meningkatkan nilai *recall* dari 53.8% menjadi 76.18% dalam partisi 90:10. Namun, peningkatan ini disertai dengan penurunan akurasi dari 80.01% menjadi 76.43% dan presisi dari 82.91% menjadi 63.53%. Akurasi tertinggi sebesar 80.01% dicapai dalam partisi 90:10 tanpa menggunakan SMOTE. Temuan ini menyoroati bahwa peningkatan daya ingat sangat penting untuk prediksi pengunduran diri yang efektif, karena memungkinkan perusahaan untuk mengantisipasi dan menanggapi secara proaktif potensi pergantian karyawan. Studi ini berkontribusi pada bidang ini dengan menunjukkan bagaimana SMOTE dapat meningkatkan kinerja pengklasifikasi C5.0 dan menyarankan bahwa *recall* harus diprioritaskan dalam pemodelan prediktif untuk aplikasi manajemen sumber daya manusia.

Kata kunci: Ketidakseimbangan Data, C5.0, *Chi-Square*, SMOTE, Klasifikasi Data, Prediksi Pengunduran Diri

ABSTRACT

HANDLING OF UNBALANCED DATA IN C5.0 AND CHI-SQUARE CLASSIFICATION MODELS FOR EMPLOYEE RESIGN PREDICTION

(By: Nanda Ahmad Reza; Advisors: Muhammad Itqan Mazdadi, S.Kom., M.Kom and Irwan Budiman, S.T., M.Kom.; 2025; 83 pages)

Employee resignation prediction is a critical issue for companies aiming to minimize negative impacts such as reduced productivity and increased recruitment costs. This study applies the C5.0 classification algorithm combined with Chi-Square feature selection and Synthetic Minority Over-sampling Technique (SMOTE) to address the class imbalance problem, where the number of employees who resign (minority class) is significantly lower than those who stay (majority class). SMOTE is used to generate synthetic data for the minority class to enhance the model's ability to detect potential resignations. The dataset is partitioned into three ratios: 70:30, 80:20, and 90:10. Experimental results show that the application of SMOTE substantially improves the recall value from 53.8% to 76.18% in the 90:10 partition. However, this improvement comes with a decrease in accuracy from 80.01% to 76.43% and precision from 82.91% to 63.53%. The highest accuracy of 80.01% is achieved in the 90:10 partition without using SMOTE. These findings highlight that improving recall is essential for effective resignation prediction, as it enables companies to anticipate and respond proactively to potential employee turnover. This study contributes to the field by demonstrating how SMOTE can enhance the performance of the C5.0 classifier and suggests that recall should be prioritized in predictive modeling for human resource management applications.

Keywords: Imbalance Data, C5.0, Chi-Square, SMOTE, Data Classification, Resignation Prediction

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Penanganan Data Tidak Seimbang pada Model Klasifikasi C5.0 dan Chi-Square untuk Prediksi Karyawan Resign” untuk memenuhi syarat dalam menyelesaikan pendidikan program S-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga yaumul qiyamah.

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

1. Keluarga terutama orang tua yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Rudy Herteno, S.Kom, M.Kom. selaku dosen Pembimbing akademik yang selalu membimbing dan memberikan nasihat.
3. Bapak Muhammad Itqan Mazdadi S.Kom., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
4. Bapak Irwan Budiman, S.T., M.Kom. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
5. Ibu Dwi Kartini, S.Kom., M.Kom. selaku dosen ketua penguji yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini
6. Bapak Setyo Wahyu Saputro, S.Kom., M.Kom. selaku dosen penguji yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini
7. Ibu Dwi Kartini, S.Kom., M.Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
8. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
9. Teman-teman keluarga Ilmu Komputer angkatan 2018 yang telah lulus lebih dahulu dan juga yang belum. Terima kasih untuk perjuangan yang sudah dilewati

bersama memberikan motivasi untuk menyelesaikan skripsi ini dan juga tempat bertanya disaat kebingungan.

10. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, maka saran dan kritik dari semua pihak sangat diharapkan demi penyempurnaan pada penelitian selanjutnya. Diharapkan skripsi ini dapat bermanfaat bagi semua pihak dan dapat menambah pengetahuan kita semua.

Banjarbaru, 30 Juni 2025



Nanda Ahmad Reza

DAFTAR ISI

HALAMAN JUDUL.....	i
SKRIPSI.....	ii
PERNYATAAN.....	iii
ABSTRAK.....	iv
ABSTRACT.....	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 Batasan Masalah.....	3
1.4 Tujuan.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu.....	5
2.2 Keaslian Penelitian.....	6
2.3 Landasan Teori.....	9
2.3.1 Pengunduran Diri Karyawan.....	9
2.3.2 Data Mining.....	10
2.3.3 Imbalance Data.....	11
2.3.4 Seleksi Fitur dengan <i>Chi-Square</i>	12
2.3.5 Algoritma C5.0.....	13
2.3.6 <i>Synthetic Minority Over-sampling Technique (SMOTE)</i>	16
2.3.7 Data Splitting.....	17
2.3.8 Pengujian <i>Confusion Matrix</i>	18
BAB III METODE PENELITIAN.....	20
1.1 Alat Penelitian.....	20
1.2 Bahan Penelitian.....	20

1.3	Prosedur Penelitian.....	21
BAB IV HASIL DAN PEMBAHASAN		24
4.1	Hasil.....	24
4.1.1	Pengumpulan Data	24
4.1.2	<i>Preprocessing</i> Data	25
4.1.3	Seleksi Fitur	27
4.1.4	Data Splitting	28
4.1.5	Penyeimbangan Data (SMOTE)	32
4.1.6	Hasil Evaluasi Model	36
4.2	Pembahasan	41
BAB V PENUTUP.....		55
5.1	Kesimpulan.....	55
5.2	Saran.....	55
DAFTAR PUSTAKA		57
LAMPIRAN.....		62

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Keaslian Penelitian.....	7
Tabel 2. Rancangan Penelitian.....	8
Tabel 3. Confusion matriks dan algoritma komputasi	18
Tabel 4. Employee future prediction dataset.....	24
Tabel 5. Deskripsi dataset	24
Tabel 6. Dataset hasil preprocessing	26
Tabel 7. Hasil perhitungan p-value Chi-Square	27
Tabel 8. Data training dengan ratio 70:30.....	29
Tabel 9. Data training dengan ratio 80:20.....	30
Tabel 10. Data training dengan ratio 90:10.....	31
Tabel 11. Data training 70:30 setelah penyeimbangan data menggunakan SMOTE	33
Tabel 12. Data training 80:20 setelah penyeimbangan data menggunakan SMOTE	34
Tabel 13. Data training 90:10 setelah penyeimbangan data menggunakan SMOTE	35
Tabel 14. Hasil Confusion Matrix menggunakan C5.0 tanpa SMOTE	36
Tabel 15. Hasil evaluasi menggunakan C5.0 tanpa SMOTE.....	37
Tabel 16. Hasil Confusion Matrix menggunakan C5.0 menggunakan SMOTE.....	37
Tabel 17. Hasil evaluasi menggunakan C5.0 menggunakan SMOTE	38

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Struktur Decision Tree	14
Gambar 2. Gambur alur penelitian.....	21
Gambar 3. Distribusi kelas target pada Partisi Data 70:30	30
Gambar 4. Distribusi kelas target pada Partisi Data 80:20	31
Gambar 5. Distribusi kelas target pada Partisi Data 90:10	32
Gambar 6. Distribusi kelas target pada Partisi Data 70:30 setelah SMOTE.....	33
Gambar 7. Distribusi kelas target pada Partisi Data 80:20 setelah SMOTE.....	34
Gambar 8. Distribusi kelas target pada Partisi Data 90:10 setelah SMOTE.....	35
Gambar 9. Hasil evaluasi menggunakan C5.0 tanpa SMOTE	38
Gambar 10. Hasil evaluasi menggunakan C5.0 setelah menggunakan SMOTE	39
Gambar 11. Perbandingan kinerja C5.0 sebelum dan setelah SMOTE	40
Gambar 12. Grafik akurasi pada partisi data 70:30.....	42
Gambar 13. Grafik precision pada partisi data 70:30	43
Gambar 14. Grafik recall pada partisi data 70:30	44
Gambar 15. Grafik TNR pada partisi data 70:30	44
Gambar 16. Grafik F1 Score pada partisi data 70:30.....	45
Gambar 17. Grafik akurasi pada partisi data 80:20.....	45
Gambar 18. Grafik precision pada partisi data 80:20	46
Gambar 19. Grafik recall pada partisi data 80:20	47
Gambar 20. Grafik TNR pada partisi data 80:20	47
Gambar 21. Grafik F1 Score pada partisi data 80:20.....	48
Gambar 22. Grafik akurasi pada partisi data 90:10.....	49
Gambar 23. Grafik precision pada partisi data 90:10	49
Gambar 24. Grafik recall pada partisi data 90:10	50
Gambar 25. Grafik TNR pada partisi data 90:10	51
Gambar 26. Grafik F1 Score pada partisi data 90:10.....	51

DAFTAR LAMPIRAN

Lampiran	Halaman
Lampiran 1. Dataset	63
Lampiran 2. Source Code Pycarm Chi-Square dan SMOTE.....	63
Lampiran 3. Source Code R Klasifikasi menggunakan C5.0 data training 70%	66
Lampiran 4. Source Code R Klasifikasi menggunakan C5.0 data training 80%	66
Lampiran 5. Source Code R Klasifikasi menggunakan C5.0 data training 90%	66
Lampiran 6. Source Code R Klasifikasi C5.0 dan SMOTE data training 70%	67
Lampiran 7. Source Code R Klasifikasi C5.0 dan SMOTE data training 80%	67
Lampiran 8. Source Code R Klasifikasi C5.0 dan SMOTE data training 90%	68