



**PERBANDINGAN *OVERSAMPLING* SMOTE DAN ADASYN DALAM
OPTIMALISASI MODEL *RANDOM FOREST* PADA DATA TIDAK
SEIMBANG (STUDI KASUS: RASIO KEUANGAN *COMPANY*
BANKRUPTCY PREDICTION)**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
NOVANDA RIZKY RAMADHANA
NIM. 2111017210015**

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



**PERBANDINGAN *OVERSAMPLING* SMOTE DAN ADASYN DALAM
OPTIMALISASI MODEL *RANDOM FOREST* PADA DATA TIDAK
SEIMBANG (STUDI KASUS: RASIO KEUANGAN *COMPANY*
BANKRUPTCY PREDICTION)**

SKRIPSI

**untuk memenuhi persyaratan
dalam menyelesaikan program sarjana Strata-1 Statistika**

**Oleh
NOVANDA RIZKY RAMADHANA
NIM. 2111017210015**

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

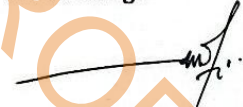
SKRIPSI

**PERBANDINGAN *OVERSAMPLING SMOTE* DAN *ADASYN* DALAM
OPTIMALISASI MODEL *RANDOM FOREST* PADA DATA TIDAK
SEIMBANG (STUDI KASUS: RASIO KEUANGAN *COMPANY BANKRUPTCY*
PREDICTION)**

Oleh:
Novanda Rizky Ramadhana
2111017210015

Telah dipertahankan pada hari *Jum'at*, tanggal 17-10-2025 dan disetujui oleh dosen pembimbing dan dosen penguji sebagai berikut:

Pembimbing I



Fuad Muhajirin Farid, S.Pd, M.Si
NIP. 198807112019031014

Penguji I



Nur Salam, S.Si, M.Sc
NIP. 197708132005011003

Pembimbing II



Yeni Rahkmawati, S.Mat, M.Si
NIP. 199404032022032014

Penguji II




Rifqi Aulya Rahman, S.Mat, M.Si
NIP. 199708072024061003

Banjarbaru, *Jum'at* 17 Oktober 2025

Mengetahui,
Ketua Jurusan/Koordinator PS Statistika
FMIPA ULM




Dewi Anggraini, S.Si, M.App.Sci, Ph.D
NIP. 198303282005012001

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 diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 14 Oktober 2025



Novanda Rizky Ramadhana
NIM. 2111017210015

ABSTRAK

Perbandingan *Oversampling* SMOTE dan ADASYN dalam Optimalisasi Model *Random Forest* pada Data Tidak Seimbang (Studi Kasus: Rasio Keuangan *Company Bankruptcy Prediction*) (Oleh: Novanda Rizky Ramadhana; Pembimbing: Fuad Muhajirin Farid dan Yeni Rahkmawati, 2025; 62 halaman)

Klasifikasi merupakan proses analisis data yang dapat melakukan prediksi kelas berdasarkan karakteristik yang telah ditentukan sebelumnya. Di era *big data*, klasifikasi dapat dilakukan menggunakan *machine learning*. Permasalahan *machine learning* dalam melakukan analisis klasifikasi adalah ketidakseimbangan data yang dapat memengaruhi performa model. SMOTE dan ADASYN adalah teknik *oversampling* untuk mengatasi hal tersebut. Penelitian ini bertujuan untuk mengidentifikasi teknik SMOTE dan ADASYN dalam meningkatkan performa model *random forest* pada data tidak seimbang dalam kasus kebangkrutan perusahaan menggunakan rasio keuangan. Model-model akan dibangun menggunakan data *training* berdasarkan berbagai *splitting* data dan perlakuan *oversampling*. Hasil model yang telah dibuat akan diuji menggunakan data *testing*. Dari hasil evaluasi matriks, kombinasi yang paling baik untuk model yaitu *splitting* data 70:30 dengan SMOTE yang menghasilkan nilai *f1-score* yang paling tinggi yaitu 40.57% dibandingkan ADASYN sebesar 36.11% yang mengalami penurunan *f1-score* sebesar 4.46% dan tanpa perlakuan *oversampling* sebesar 19.51% dengan penurunan *f1-score* sebesar 21.06%. Hasil penelitian menunjukkan SMOTE dan ADASYN dapat mengidentifikasi nilai minoritas yang menjadi permasalahan utama data tidak seimbang, dengan SMOTE menunjukkan performa yang lebih unggul dibandingkan ADASYN.

Kata kunci: SMOTE, ADASYN, *Random Forest*, Kebangkrutan Perusahaan, Rasio Keuangan

ABSTRACT

Comparison of SMOTE and ADASYN Oversampling for Optimizing Random Forest on Imbalanced Data (Case Study: Financial Ratios Company Bankruptcy Prediction) (By: Novanda Rizky Ramadhana; Advisors: Fuad Muhajirin Farid and Yeni Rahkmawati; 2025; 62 pages)

Classification is a data analysis process that can predict classes based on predefined characteristics. In the era of big data, classification can be performed using machine learning. The problem of machine learning in classification analysis is imbalance data which often affect model performance. SMOTE and ADASYN are oversampling techniques to solve this problem. This study aims to evaluate the effectiveness of SMOTE and ADASYN in improving the performance of the Random Forest model on imbalanced data in the case of company bankruptcy using financial ratios. Models were built using training data with various splitting data and oversampling techniques. Then, the resulting models will be tested using testing data. The results show that the best model was achieved with a combination of splitting data 70:30 using SMOTE technique, which produced the highest f1-score of 40.57%, compared to ADASYN technique with 36.11% (a decrease of 4.46%), and without oversampling techniques with 19.51% (a decrease of 21.06%). The findings indicate SMOTE and ADASYN can identify minority values which are the main problem of imbalance data, with SMOTE showing better performance compared to ADASYN.

Keywords: SMOTE, ADASYN, Random Forest, Company Bankruptcy, Financial Ratios

PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Allah SWT. atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penulisan Tugas Akhir dengan judul "Perbandingan *Oversampling* SMOTE dan ADASYN dalam Optimalisasi Model *Random Forest* pada Data Tidak Seimbang (Studi Kasus: Rasio Keuangan *Company Bankruptcy Prediction*)". Skripsi ini disusun dengan tujuan memenuhi syarat dalam menyelesaikan program sarjana strata-1 Statistika di Program Studi Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat. Dalam penyusunan skripsi ini, penulis banyak mendapatkan bimbingan, dukungan, dan bantuan dari berbagai pihak secara langsung maupun tidak langsung oleh beberapa pihak yaitu:

1. Orang tua dan keluarga besar yang senantiasa memberikan doa dan mendukung dalam proses penyelesaian Tugas Akhir.
2. Ibu Dewi Anggraini, S.Si., M.App.Sci., Ph.D selaku Koordinator Program Studi Statistika FMIPA ULM atas dukungan, ilmu, dan nasihat selama perkuliahan.
3. Bapak Fuad Muhajirin Farid, S.Pd., M.Si dan Ibu Yeni Rahkmawati, S.Mat., M.Si selaku dosen pembimbing yang telah bersedia meluangkan waktu, memberikan arahan serta dukungan selama proses penulisan Tugas Akhir.
4. Bapak Nur Salam, S.Si., M.Sc dan Bapak Rifqi Aulya Rahman, S.Mat., M.Si selaku dosen penguji yang telah memberikan masukan dan saran selama proses perbaikan penulisan Tugas Akhir.
5. Seluruh dosen dan staf Program Studi Statistika FMIPA ULM yang telah memberikan dukungan, ilmu, dan nasihat selama masa perkuliahan di Program Studi Statistika FMIPA ULM.
6. Teman-teman Statistika FMIPA ULM maupun di luar Statistika FMIPA ULM yang selalu menemani, menghibur, mendukung, dan selalu berkeluh kesah saat saya juga susah.

Penulis menyadari bahwa penulisan Tugas Akhir ini masih jauh dari sempurna. Oleh karena itu, penulis sangat terbuka terhadap kritik dan saran yang membangun untuk penyempurnaan Tugas Akhir ini. Demikian pula, diharapkan penulisan ini dapat bermanfaat bagi semua pihak.

Banjarbaru, 6 Oktober 2025

Novanda Rizky Ramadhana

DAFTAR ISI

LEMBAR PENGESAHAN	iii
PERNYATAAN	iv
ABSTRAK	v
ABSTRACT	vi
PRAKATA	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terdahulu	5
2.2 Kajian Teori.....	7
2.2.1 Statistika Deskriptif dan Diagram Batang.....	7
2.2.2 Klasifikasi	8
2.2.3 <i>Random Forest</i>	8
2.2.4 Data Tidak Seimbang	10
2.2.5 SMOTE (<i>Synthetic Minority Oversampling Technique</i>).....	12
2.2.6 ADASYN (<i>Adaptive Synthetic Sampling Approach</i>).....	13
2.2.7 <i>Confusion Matrix</i>	15
2.2.8 Kebangkrutan.....	16
2.2.9 Rasio Keuangan.....	17
BAB III METODE PENELITIAN	20
3.1 Sumber Data	20
3.2 Variabel Penelitian	20
3.3 Prosedur Penelitian	21
BAB IV HASIL DAN PEMBAHASAN	24
4.1 Eksplorasi Data dan <i>Preprocessing Data</i>	24
4.2 Splitting Data.....	26
4.3 SMOTE dan ADASYN.....	27
4.3.1 SMOTE	27
4.3.2 ADASYN	29
4.4 Membangun Model <i>Random Forest</i>	32
4.5 Evaluasi Model.....	34
BAB V PENUTUP	36

5.1	Kesimpulan	36
5.2	Saran.....	36
DAFTAR PUSTAKA		37
LAMPIRAN.....		41
RIWAYAT HIDUP		61

PRODI STATISTIKA

DAFTAR GAMBAR

Gambar 2.1 Contoh Statistika Deskriptif dengan Diagram Batang.....	7
Gambar 2.2 Ilustrasi Pohon Random Forest.....	10
Gambar 2.3 Ilustrasi Undersampling	11
Gambar 2.4 Ilustrasi Oversampling	12
Gambar 3.1 Diagram Alur Prosedur Penelitian.....	23
Gambar 4.1 <i>Barchart</i> Status Perusahaan	24
Gambar 4.2 Boxplot Variabel Independent (X) dalam Pendeteksian Outlier	25
Gambar 4.3 Boxplot Variabel Independent (X) Setelah Penanganan Outlier	25
Gambar 4.4 <i>Barchart</i> Status Perusahaan Setelah Penanganan Outlier.....	26
Gambar 4.5 Ilustrasi Visualisasi Pohon Keputusan Random Forest.....	32

DAFTAR TABEL

Tabel 2.1 Bentuk Tabel Confusion Matrix.....	15
Tabel 3.1 Daftar Variabel Penelitian.....	20
Tabel 4.1 Jumlah Frekuensi Data Berdasarkan Splitting.....	26
Tabel 4.2 Tabel Data Dipilih, Data Tetangga, dan Data Sintetis SMOTE Berdasarkan Pemilihan Python.....	28
Tabel 4.3 Tabel Data Dipilih, Data Tetangga, dan Data Sintetis ADASYN Berdasarkan Pemilihan Python.....	29
Tabel 4.4 Data Training Tanpa Oversampling vs SMOTE vs ADASYN	31
Tabel 4.5 Kedalaman Pohon yang Terbentuk.....	34
Tabel 4.6 Confusion Matrix <i>Random Forest</i> Tanpa Oversampling Splitting Data 90:10	34
Tabel 4.7 Nilai Evaluasi Matriks Berdasarkan Kombinasi Splitting Data dan Perlakuan Oversampling.....	35

DAFTAR LAMPIRAN

Lampiran 1. Pemilihan Variabel pada Data.....	41
Lampiran 2. Data Outlier > 1	42
Lampiran 3. Hasil Pembersihan Data.....	43
Lampiran 4. Confusion Matrix Model Random Forest.....	44
Lampiran 5. Syntax Google Colab	47

PRODI STATISTIKA