



**PERBANDINGAN METODE OVERSAMPLING DALAM
MENYEIMBANGKAN DATA MENGGUNAKAN ALGORITMA REGRESI
LOGISTIK PADA KLASIFIKASI PENYAKIT STROKE**

SKRIPSI

**Untuk memenuhi persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer**

**Oleh
ALMIRA SYAHADATI ARSYA
NIM 1911016320010**

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

JUNI 2023

SKRIPSI

**PERBANDINGAN METODE OVERSAMPLING DALAM
MENYEIMBANGKAN DATA MENGGUNAKAN ALGORITMA REGRESI
LOGISTIK PADA KLASIFIKASI PENYAKIT STROKE**

Oleh:

ALMIRA SYAHADATI ARSYA

NIM 1911016320010

Telah dipertahankan di depan Dosen Penguji pada Tanggal 20 Juni 2023

Susunan Dosen Penguji:

Pembimbing I



Rudy Herterio S.Kom., M.Kom.
NIP. 198809252022031003

Dosen Penguji I



Dwi Kartini S.Kom., M.Kom.
NIP. 198704212012122003

Pembimbing II



Irwan Budiman, S.T. M.Kom.
NIP. 197703252008121001

Dosen Penguji II



Muhammad Itgan Mazdadi S.Kom., M.Kom.
NIP. 199006122019031013



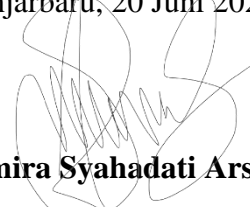
Banjarnegara, 23 Juni 2023
Koordinator Program Studi Ilmu Komputer

Irwan Budiman, S.T. M.Kom.
NIP. 197703252008121001

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 didalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 20 Juni 2023



Almira Syahadati Arsyah

NIM. 1911016320010

ABSTRAK

PERBANDINGAN METODE OVERSAMPLING DALAM MENYEIMBANGKAN DATA MENGGUNAKAN ALGORITMA REGRESI LOGISTIK PADA KLASIFIKASI PENYAKIT STROKE (Oleh: Almira Syahadati Arsyah; Pembimbing: Rudy Herteno S.Kom., M.Kom. Irwan Budiman S.T., M.Kom; 2023; 64 halaman)

Stroke adalah penyakit serebrovaskular yang menyebabkan kerusakan pada bagian otak, hal tersebut dikarenakan adanya gangguan di dalam peredaran darah. Tidak dapat dipungkiri kalau penyakit stroke mengalami peningkatan setiap tahunnya. Hal tersebut menjadikan stroke sebagai permasalahan yang sangat serius dan butuh untuk segera ditangani. Telah banyak metode yang dikembangkan untuk mengatasi penyakit stroke ini. Salah satunya adalah dengan menggunakan *machine learning*. Dalam bidang medis *machine learning* telah berkembang pesat dalam beberapa dekade terakhir. Beberapa algoritma *machine learning* telah banyak digunakan untuk membantu bidang medis dalam mendiagnosis penyakit stroke. Penelitian ini mencoba membandingkan algoritma Regresi Logistik pada stroke, dengan bantuan beberapa teknik oversampling untuk data tidak seimbang seperti ADASYN (*Adaptive Synthetic Sampling*), ROS (*Random Over Sampling*) dan SMOTE (*Synthetic Minority Oversampling Technique*). Dengan menggunakan dataset sebanyak 5.110 data dan 12 variabel independen serta satu variabel dependen. Kedua algoritma tadi diproses setelah data tidak seimbang sudah diatasi dengan teknik-teknik *oversampling*. Dari hasil percobaan, algoritma Regresi Logistik-SMOTE mencapai tingkat akurasi tertinggi dengan nilai akurasi sebesar 90,7%.

Kata Kunci: *Stroke, Klasifikasi, Regresi Logistik, ADASYN, ROS, SMOTE*

ABSTRACT

COMPARISON OF OVERSAMPLING METHODS IN BALANCED DATA USING LOGISTIC REGRESSION ALGORITHMS IN STROKE DISEASE CLASSIFICATION (By: Almira Syahadati Arsyah; Advisor: Rudy Herteno S.Kom., M.Kom. Irwan Budiman S.T., M.Kom; 2023; 64 pages)

Stroke is a cerebrovascular disease that damages brain tissue. This is linked to blood circulation abnormalities. Every year, the number of strokes increases. This makes stroke a severe issue that must be treated immediately. Many methods have been developed to treat stroke. One example of overcoming stroke is by using machine learning. In the medical field, machine learning has developed rapidly in the last few decades. Several machine learning algorithms have been widely used to assist the medical field in diagnosing stroke. This study attempts to compare Logistic Regression in stroke, using several oversampling techniques for unbalanced data such as ADASYN (Adaptive Synthetic Sampling), ROS (Random Over Sampling), and SMOTE (Synthetic Minority Oversampling Technique). Furthermore, the algorithms employed a dataset of 5.110 data and 12 independent variables, and one dependent variable. The two algorithms were processed after oversampling techniques had overcome the unbalanced data. From the experimental results, the Logistic Regression-SMOTE algorithm achieves the highest level of accuracy with an accuracy value 90.7%.

Keywords: *Stroke, Classification, Logistic Regression, ADASYN, ROS, SMOTE*

PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Metode Oversampling Dalam Menyeimbangkan Data Menggunakan Algoritma Regresi Logistik Pada Klasifikasi Penyakit Stroke” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

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

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Rudy Herteno S.Kom., M.Kom. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Irwan Budiman S.T., M.Kom. selaku dosen pembimbing pendamping dan Koordinator Program Studi Ilmu Komputer FMIPA ULM yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
5. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
6. Ucapan terima kasih kepada Siti Roziana Azizah, Raisa Amalia, Fadhila Nisrina Saumy, Dina Arifah. Ajwa Helisa, Astina Faridhah, dan Alya Firyal

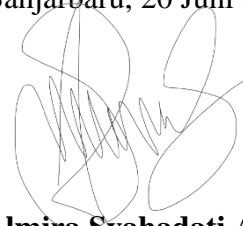
Nuha yang memberikan dukungan dalam proses perkuliahan dan mengerjakan skripsi.

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

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan serupa berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT

Banjarbaru, 20 Juni 2023



Almira Syahadati Arsyah

NIM. 1911016320010

DAFTAR ISI

	Halaman
HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Kajian Terdahulu	4
2.2 Keaslian Penelitian	5
2.3 Penyakit Stroke.....	8
2.4 Data Mining.....	8

2.5	Ketidakseimbangan Data.....	9
2.6	Teknik Oversampling	10
2.7	ADASYN	10
2.8	ROS	12
2.9	SMOTE	13
2.10	Klasifikasi.....	14
2.11	Regresi Logistik	15
2.12	Confusion Matrix	15
BAB III METODE PENELITIAN		17
3.1	Alat Penelitian	17
3.2	Bahan Penelitian.....	17
3.3	Variabel Penelitian	18
3.4	Prosedur Penelitian.....	19
BAB IV HASIL DAN PEMBAHASAN		22
4.1	Hasil.....	22
4.1.1	Persiapan Data.....	22
4.1.2	Preprocessing	23
4.1.3	Pembagian Data	26
4.1.4	Menyeimbangkan Data	26
4.1.5	Klasifikasi	37
4.1.6	Evaluasi.....	44
4.2	Pembahasan	52
BAB V PENUTUP.....		60
5.1	Kesimpulan.....	60
5.2	Saran.....	60

DAFTAR PUSTAKA 61

LAMPIRAN 61

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Keaslian Penelitian.....	5
Tabel 2. Perancangan Penelitian	7
Tabel 3. Confusion Matrix	16
Tabel 4. Sampel dataset penyakit stroke	18
Tabel 5. Deskripsi dataset penyakit stroke.....	22
Tabel 6. Sampel dataset yang memiliki missing value	24
Tabel 7. Sampel dataset yang missing valuenya sudah diisi dengan median	24
Tabel 8. Sampel dataset yang sudah melakukan tahap transformasi data.....	25
Tabel 9. Sampel dataset yang sudah dilakukan preprocessing untuk ADASYN.....	27
Tabel 10. Parameter teknik oversampling ADASYN yang digunakan.....	27
Tabel 11. Hasil sampel data sintetik ADASYN.....	29
Tabel 12. Komposisi data training sesudah dan sebelum menggunakan Teknik Oversampling ADASYN	29
Tabel 13. Sampel dataset yang sudah dilakukan preprocessing untuk ROS.....	30
Tabel 14. Parameter teknik oversampling ROS yang digunakan	31
Tabel 15. Hasil sampel dataset yang terdapat data sintetik ROS	32
Tabel 16. Komposisi data training sesudah dan sebelum menggunakan Teknik Oversampling ROS	32
Tabel 17. Sampel dataset yang sudah dilakukan preprocessing untuk SMOTE.....	33
Tabel 18. Parameter teknik oversampling SMOTE yang digunakan.....	34
Tabel 19. Jarak euclidean untuk teknik oversampling SMOTE	34
Tabel 20. Data sintesis yang dihasilkan teknik oversampling SMOTE.....	36
Tabel 21. Komposisi data training sesudah dan sebelum menggunakan Teknik Oversampling SMOTE	36
Tabel 22. Sampel dataset yang sudah dilakukan penyeimbangan data dengan Teknik oversampling ADASYN	37
Tabel 23. Nilai koefisien Regresi Logistik-ADASYN	38

Tabel 24. Sampel dataset yang sudah dilakukan penyeimbangan data dengan Teknik oversampling ROS	40
Tabel 25. Nilai Koefisien Regresi Logistik-ROS	40
Tabel 26. Sampel dataset yang sudah dilakukan penyeimbangan data dengan Teknik oversampling SMOTE	42
Tabel 27. Nilai koefisien Regresi Logistik-SMOTE.....	43
Tabel 28. Nilai akurasi, presisi dan recall yang didapat dari model Regresi Logistik-ADASYN	46
Tabel 29. Nilai akurasi, presisi dan recall yang didapat dari model Regresi Logistik-ROS.....	49
Tabel 30. Nilai akurasi, presisi dan recall yang didapat dari model Regresi Logistik-SMOTE	51
Tabel 31. Perbandingan hasil presisi label stroke	53
Tabel 32. Perbandingan hasil presisi label tidak stroke	54
Tabel 33. Perbandingan hasil recall stroke	55
Tabel 34. Perbandingan hasil recall tidak stroke	56
Tabel 35. Perbandingan hasil akurasi	57

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Flowchart algoritma ROS	12
Gambar 2. Prosedur penelitian.....	19
Gambar 3. Perbandingan jumlah kelas atribut stroke	19
Gambar 4. Komposisi data training sesudah dan sebelum menggunakan Teknik Oversampling ADASYN.....	30
Gambar 5. Komposisi data training sesudah dan sebelum menggunakan Teknik Oversampling ROS.....	33
Gambar 6. Komposisi data training sesudah dan sebelum menggunakan Teknik Oversampling SMOTE.....	37
Gambar 7. Confusion Matrix Regresi Logistik-ADASYN.....	45
Gambar 8. Nilai akurasi, presisi dan recall yang didapat dari model Regresi Logistik-ADASYN.....	47
Gambar 9. Confusion Matrix Regresi Logistik-ROS.....	47
Gambar 10. Nilai akurasi, presisi dan recall yang didapat dari model Regresi Logistik-ROS	49
Gambar 11. Confusion Matrix Regresi Logistik-SMOTE.....	50
Gambar 12. Nilai akurasi, presisi dan recall yang didapat dari model Regresi Logistik-SMOTE	52
Gambar 13. Perbandingan hasil presisi label stroke	54
Gambar 14. Perbandingan hasil presisi label tidak stroke	55
Gambar 15. Perbandingan recall stroke	56
Gambar 16. Perbandingan hasil recall tidak stroke.....	57
Gambar 17. Perbandingan hasil akurasi.....	58

DAFTAR LAMPIRAN

- Lampiran 1. Source Code Import Library yang dibutuhkan
- Lampiran 2. Source Code load dataset
- Lampiran 3. Source Code memeriksa missing value
- Lampiran 4. Source Code mengisi missing value dengan nilai rata-rata
- Lampiran 5. Source Code encode data
- Lampiran 6. Source Code Regresi Logistik-SMOTE Model
- Lampiran 7. Source Code Regresi Logistik-ADASYN Model
- Lampiran 8. Source Code Regresi Logistik-ROS Model