



**PENERAPAN MWMOTE DALAM MENYEIMBANGKAN DATA
MENGUNAKAN ALGORITMA RANDOM FOREST PADA
KLASIFIKASI PENYAKIT STROKE**

Skripsi

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

EKO TRIYONO

NIM 1711016210005

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

JUNI 2024



**PENERAPAN MWMOTE DALAM MENYEIMBANGKAN DATA
MENGUNAKAN ALGORITMA RANDOM FOREST PADA
KLASIFIKASI PENYAKIT STROKE**

Skripsi

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

Eko Triyono

NIM 1711016210005

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

JUNI 2024

SKRIPSI

**PENERAPAN MWMOTE DALAM MENYEIMBANGKAN DATA
MENGUNAKAN ALGORITMA RANDOM FOREST PADA
KLASIFIKASI PENYAKIT STROKE**

Oleh :

EKO TRIYONO
1711016210005

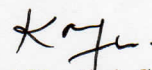
Telah dipertahankan di depan Dosen Penguji pada tanggal 11 Juni 2024,

Susunan Dosen Penguji :


Pembimbing I


Muhammad Itqan Mazdadi,
S.Kom., M.Kom.
NIP. 19930824201901012

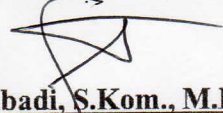
Dosen Penguji I


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

Pembimbing II



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

Dosen Penguji II


Friska Abadi, S.Kom., M.Kom.
NIP. 198809132023211010



Banjarbaru, 11 Juni 2024
Ketua Program Studi Ilmu Komputer

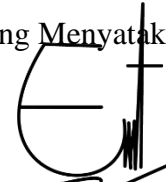

Irwan Budiman, S.T., M.Kom
NIP. 197703252008121001

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

Banjarbaru, Juni 2024

Yang Menyatakan,



Eko Triyono

NIM. 1711016210005

ABSTRAK

PENERAPAN MWMOTE DALAM MENYEIMBANGKAN DATA MENGGUNAKAN ALGORITMA RANDOM FOREST PADA KLASIFIKASI PENYAKIT STROKE (Oleh : Eko Triyono; Pembimbing: Muhammad Itqan Mazdadi, S.Kom, M.Kom. dan Irwan Budiman, S.T., M.Kom.; 2024; 64 halaman)

Stroke adalah kondisi ketika pasokan darah ke otak terganggu karena penyumbatan atau pecahnya pembuluh darah. Kondisi ini menyebabkan area tertentu pada otak tidak mendapat suplai oksigen dan nutrisi, sehingga terjadi kematian sel-sel otak. Setiap tahunnya penyakit stroke mengalami peningkatan. Hal tersebut menjadikan stroke sebagai permasalahan yang sangat serius dan perlu untuk segera ditangani. Telah banyak metode yang dikembangkan untuk mengklasifikasi penyakit stroke. 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 mengklasifikasi penyakit stroke. Pada penelitian ini dilakukan klasifikasi dengan menggunakan algoritman *Random Forest* serta *oversampling* data dengan menggunakan *Majority Weighted Minority Oversampling Technique (MWMOTE)*. Data yang digunakan pada penelitian ini adalah data penyakit stroke yang memiliki kelas *major* sebanyak 4.861 data dan kelas *minor* sebanyak 249 data. Penelitian ini dilakukan dengan membandingkan klasifikasi *Random Forest* dengan dan tanpa *oversampling* MWMOTE. Proses *oversampling* dan klasifikasi ini dilakukan dalam *Stratified 10-Fold Cross Validation*. Hasilnya didapatkan bahwa klasifikasi *Random Forest* dengan MWMOTE menghasilkan performa terbaik, dengan nilai akurasi sebesar 91,35%, presisi sebesar 11,53%, *recall* sebesar 12,08%, *f1-score* sebesar 11,75%, dan AUC sebesar 0,537.

Kata kunci: Stroke, *Random Forest*, *Oversampling*, MWMOTE, Klasifikasi.

ABSTRACT

APPLICATION OF MWMOTE IN BALANCING DATA USING RANDOM FOREST ALGORITHM ON STROKE DISEASE CLASSIFICATION (By: Eko Triyono; Advisor: Muhammad Itqan Mazdadi, S.Kom, M.Kom and Irwan Budiman, S.T., M.Kom.;2024; 64 pages)

Stroke is a condition where the blood supply to the brain is disrupted due to blockage or rupture of a blood vessel. This condition causes certain areas of the brain to not be supplied with oxygen and nutrients, resulting in the death of brain cells. Every year, stroke disease increases. This makes strokes a very serious problem that needs to be addressed immediately. Many methods have been developed to classify stroke disease. One of them is using machine learning. In the medical field, machine learning has grown rapidly in the last few decades. Several machine learning algorithms have been widely used in the medical field for classifying stroke diseases. In this study, classification was carried out using the Random Forest algorithm and oversampling data using the Majority Weighted Minority Oversampling Technique (MWMOTE). The data used in this study is stroke disease data, which has a major class of 4,861 data and a minor class of 249 data. This research was conducted by comparing Random Forest classification with and without MWMOTE oversampling. The oversampling and classification processes were carried out in Stratified 10-Fold Cross Validation. It was found that Random Forest classification with MWMOTE produced the best performance, with an accuracy value of 91.35%, precision of 11.53%, recall of 12.08%, f1-score of 11.75%, and AUC of 0.537..

Keywords: Stroke, Random Forest, Oversampling, MWMOTE, Classification.

PRAKATA

Puji syukur penulis panjatkan kepada Allah SWT. karena atas berkat, rahmat, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul **PENERAPAN MWMOTE DALAM MENYEIMBANGKAN DATA MENGGUNAKAN ALGORITMA RANDOM FOREST PADA KLASIFIKASI PENYAKIT STROKE** untuk memenuhi syarat dalam menyelesaikan Pendidikan program S-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW. beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiama*.

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. Allah SWT. karena atas limpahan karunia-Nya dan bantuan-Nya lah penulis dapat menyelesaikan skripsi.
2. Keluarga terutama kedua orang tua dan saudara Saya yang selalu memberikan do'a, semangat, bantuan, dan dukungan dalam proses penyelesaian skripsi.
3. Bapak Dodon Turianto Nugrahadi, S.Kom., M.Eng. selaku Dosen Pembimbing Akademik.
4. Bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom. selaku Dosen Pembimbing Utama dan Bapak Irwan Budiman, S.T., M.Kom. selaku Dosen Pembimbing Pendamping yang telah memberikan waktu dan bimbingannya kepada penulis dalam pengerjaan skripsi.
5. Bapak Irwan Budiman, S.T., M.Kom. selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
6. Seluruh Dosen dan Staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama berkuliah di Ilmu Komputer yang sangat bermanfaat.

7. Teman-teman keluarga Ilmu Komputer Angkatan 2017 yang memberikan bantuan dan dukungan yang senantiasa menjadi alat pacu penulis untuk selalu berusaha dan semangat untuk menyelesaikan skripsi.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan skripsi ini masih jauh dari kesempurnaan, maka penulis mengharapkan bantuan berupa kritik dan saran yang membangun dari semua pihak sangat diharapkan demi penyempurnaan dan mutu penulisan skripsi ini. Semoga tulisan ini dapat bermanfaat bagi semua pihak untuk dapat menambah pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, Juni 2024



Eko Triyono

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA	
2.1 Kajian Terdahulu	5
2.2 Keaslian Penelitian	6
2.3 Landasan Teori	10
BAB III METODE PENELITIAN	
3.1 Alat Penelitian	20
3.2 Bahan Penelitian.....	20
3.3 Variabel Penelitian	20
3.4 Prosedur Penelitian.....	20
BAB IV HASIL DAN PEMBAHASAN	
4.1 Hasil.....	24
4.1.1 Pengumpulan Data	24
4.1.2 <i>Preprocessing</i>	26

4.1.3	Pembagian Data	31
4.1.4	Menyeimbangkan Data	33
4.1.5	Klasifikasi Data.....	39
4.1.6	Evaluasi.....	51
4.2	Pembahasan	57
BAB V PENUTUP		
5.1	Kesimpulan.....	60
5.2	Saran.....	60
DAFTAR PUSTAKA		
LAMPIRAN		

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Keaslian Penelitian	7
Tabel 2. Penelitian yang Diusulkan	9
Tabel 3. Confusion Matrix.....	17
Tabel 4. Keterangan Rumus Confusion Matrix.....	18
Tabel 5. Nilai Kualitas <i>Classifier</i>	18
Tabel 6. Deskripsi <i>dataset</i> penyakit stroke.....	24
Tabel 7. Sampel Data Penyakit Stroke	26
Tabel 8. Sampel <i>dataset</i> yang memiliki <i>missing value</i>	27
Tabel 9. Sampel <i>dataset</i> dengan <i>missing value</i> yang sudah diisi nilai rata-rata	28
Tabel 10. Sampel <i>dataset</i> yang sudah dilakukan tahap transformasi data	29
Tabel 11. Perhitungan <i>Min-Max</i>	31
Tabel 12. <i>Stratified 10-Fold Cross Validation</i>	32
Tabel 13. Pembagian data penyakit stroke dengan <i>Stratified 10-Fold Cross Validation</i>	32
Tabel 14. Sampel <i>dataset</i> yang sudah dilakukan preprocessing untuk MWMOTE	34
Tabel 15. Hasil pencarian tetangga terdekat berdasarkan parameter <i>k1</i>	35
Tabel 16. Hasil pencarian tetangga terdekat berdasarkan parameter <i>k2</i>	35
Tabel 17. Hasil pencarian tetangga terdekat berdasarkan parameter <i>k3</i>	36
Tabel 18. Contoh perhitungan data sintentik dari data X dan Y	37
Tabel 19. Data hasil <i>oversampling</i> menggunakan MWMOTE	38
Tabel 20. Komposisi data latih sesudah dan sebelum menggunakan teknik <i>oversampling</i> MWMOTE.....	38
Tabel 21. Parameter Random Forest	40
Tabel 22. Contoh sampel Data untuk klasifikasi <i>Random Forest</i>	40
Tabel 23. Nilai <i>entropy</i> dan <i>gain</i> pada pohon pertama.....	42
Tabel 24. Data atribut <i>avg_glucose_level</i> yang bernilai $\leq 0,115$	43
Tabel 25. Nilai <i>entropy</i> dan <i>gain</i> pada <i>node 1.2</i>	43
Tabel 26. Data attribute <i>age</i> yang bernilai $>0,609$	44

Tabel 27. Nilai <i>entropy</i> dan <i>gain</i> pada <i>node 2.2</i>	44
Tabel 28. Data kedua untuk model <i>Random Forest</i>	46
Tabel 29. Data ketiga untuk model <i>Random Forest</i>	47
Tabel 30. Hasil <i>Majority Voting</i>	49
Tabel 31. <i>Confusion Matrix Random Forest</i>	50
Tabel 32. <i>Confusion Matrix Random Forest</i>	51
Tabel 33. Nilai akurasi, presisi, <i>recall</i> , <i>f1-score</i> dan AUC dari model <i>Random Forest</i> dengan <i>Stratified 10-Fold Cross Validation</i>	52
Tabel 34. <i>Confusion matrix Random Forest</i>	53
Tabel 35. Nilai akurasi, presisi, <i>recall</i> , <i>f1-score</i> dan AUC dari model <i>Random Forest-MWMOTE</i> dengan <i>Stratified 10-Fold Cross Validation</i>	54
Tabel 36. Hasil penerapan nilai parameter tertentu	56
Tabel 37. Perbandingan Performa Model <i>Random Forest</i>	58

DAFTAR GAMBAR

Gambar		Halaman
Gambar 1.	Visualisasi Metode <i>Random Forest</i>	13
Gambar 2.	Tahapan <i>MWMOTE</i>	15
Gambar 3.	Skema <i>10-Fold Cross Validation</i>	16
Gambar 4.	Prosedur Penelitian	21
Gambar 5.	Komposisi data latih sebelum dan sesudah menggunakan teknik <i>oversampling</i> <i>MWMOTE</i>	39
Gambar 6.	Pohon Keputusan 1	42
Gambar 7.	Pohon Keputusan <i>Node 1.2</i>	44
Gambar 8.	Pohon Keputusan pertama	45
Gambar 9.	Pohon Keputusan Kedua.....	46
Gambar 10.	Pohon Keputusan Ketiga	48
Gambar 11.	Rata-rata dari nilai akurasi, presisi, <i>recall</i> , <i>f1-score</i> dan AUC dari model <i>Random Forest</i> dengan <i>Stratified 10-Fold Cross Validation</i>	53
Gambar 12.	Rata-rata dari nilai akurasi, presisi, <i>recall</i> , <i>f1-score</i> dan AUC dari model <i>Random Forest-MWMOTE</i> dengan <i>Stratified 10-Fold Cross Validation</i>	55
Gambar 13.	Perbandingan Performa Model <i>Random Forest</i>	59

DAFTAR LAMPIRAN

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 label encoding*

Lampiran 6. *Source code one-hot encoding*

Lampiran 7. *Source code hitungan MWMOTE*

Lampiran 8. *Source code Random Forest-MWMOTE*

Lampiran 9. *Source code rata-rata nilai evaluasi Random Forest-MWMOTE*

Lampiran 10. *Source code Random Forest*

Lampiran 11. *Source code rata-rata nilai evaluasi Random Forest*