



***RANDOM FOREST* DENGAN *RANDOM SEARCH HYPERPARAMETER*
TUNING TERHADAP KETIDAKSEIMBANGAN KELAS PADA PREDIKSI
GAGAL JANTUNG**

Skripsi

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

**Oleh
MUHAMMAD ALI ABUBAKAR
NIM 1711016210013**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2023**

SKRIPSI

**RANDOM FOREST DENGAN RANDOM SEARCH
HYPERPARAMETER TUNING TERHADAP
KETIDAKSEIMBANGAN KELAS PADA PREDIKSI GAGAL JANTUNG**

Oleh :

**MUHAMMAD ALI ABUBAKAR
NIM. 1711016210013**

telah dipertahankan di depan Dosen Penguji pada tanggal
23 Juni 2023

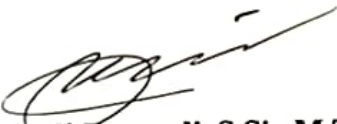
Susunan Penguji :

Pembimbing Utama



Muliadi, S.Kom., M.Cs
NIP. 197804222010121002

Pembimbing Pendamping



Andi Farmadi, S.Si., M.T
NIP. 197307252008011006

Ketua Penguji



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

Anggota Penguji



Rahmat Ramadhani, S.Kom., M.Sc
NIP. 19920330201901110001



Banjarbaru, 27 Juni 2023
Program Studi Ilmu Komputer

Budiman, S.T., M.Kom
NIP. 19703252008121001

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 2023

Yang Menyatakan,



Muhammad Ali Abubakar

NIM. 1711016210013

ABSTRAK

RANDOM FOREST DENGAN RANDOM SEARCH HYPER PARAMETER TUNING TERHADAP KETIDAKSEIMBANGAN KELAS PADA PREDIKSI GAGAL JANTUNG (Oleh: Muhammad Ali Abubakar; Pembimbing: Muliadi S.Kom, M.Cs. dan Andi Farmadi, S.Si, M.T.; 2022; 62 halaman).

Prediksi keberlangsungan hidup pasien gagal jantung telah dilakukan pada penelitian untuk mencari tahu tentang kinerja, akurasi, presisi, dan performa dari model prediksi ataupun metode yang digunakan dalam penelitian, dengan menggunakan dataset heart failure clinical records. Namun, dataset ini memiliki permasalahan yaitu bersifat tidak seimbang yang dapat menurunkan kinerja model prediksi karena cenderung menghasilkan prediksi kelas mayoritas. Penelitian ini menggunakan pendekatan level algoritma untuk mengatasi ketidakseimbangan kelas, yaitu teknik *bagging* dengan metode *Random Forest*, lalu digabungkan dengan metode *Hyper-Parameter Tuning* agar kinerja yang dihasilkan menjadi lebih baik. Selanjutnya, model dilatih dengan dataset dan dibandingkan dengan metode lain. Hasilnya menunjukkan bahwa *Random Forest* dengan *Random Search Hyper-Parameter Tuning* mencapai nilai AUC sebesar 0,906 dan model *Random Forest* tanpa *Random Search* memperoleh nilai AUC sebesar 0,866.

Kata kunci: Ketidakseimbangan Kelas, *Random Forest*, *Hyper-Parameter*

ABSTRACT

RANDOM FOREST WITH RANDOM SEARCH HYPER PARAMETER TUNING FOR CLASS IMBALANCE IN HEART FAILURE PREDICTION

(By : Muhammad Ali Abubakar; Advisors: Muliadi S.Kom, M.Cs. dan Andi Farmadi, S.Si, M.T.; 2022; 62 pages).

Prediction of the survival of heart failure patients has been carried out in research to find out about the performance, accuracy, precision, and performance of the prediction model or method used in the study, using the heart failure clinical records dataset. However, this dataset has a problem, namely being unbalanced which can reduce the performance of the prediction model because it tends to produce predictions for the majority class. This study used an algorithm level approach to overcome class imbalance, namely the bagging technique with the Random Forest method and then combined with the Hyper-Parameter Tuning method so that the resulting performance is better. Then the model was trained with the dataset and compared with other methods. The results showed that the Random Forest with Random Search Hyper Parameter-Tuning achieved an AUC value of 0,906 and the Random Forest model without Random Search obtained an AUC value of 0,866.

Keywords: Class Imbalance, Random Forest, Hyper-Parameter

PRAKATA

Puji syukur penulis panjatkan kepada Allah SWT karena atas berkah, rahmat, dan karunia-Nya, penulis dapat menyelesaikan skripsi yang berjudul “Random Forest dengan Random Search Hyper Parameter Tuning terhadap Ketidakseimbangan Kelas pada Prediksi Gagal Jantung” untuk memenuhi persyaratan menyelesaikan program sarjana S-1 Program Studi Ilmu Komputer di Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lambung Mangkurat.

Pada lembar ini, penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang sangat mendukung penulis dalam proses penyelesaian program sarjana 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 Muliadi, S.Kom., M.Cs selaku dosen pembimbing utama yang turut membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
3. Bapak Andi Farmadi, S.Si., M.T selaku dosen pembimbing pendamping yang turut membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi.
4. Bapak Irwan Budiman, S.T., M.Kom selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.

Akhir kata, penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, tetapi penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penelitian ini. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, Juni 2023



Muhammad Ali Abubakar

DAFTAR ISI

PERNYATAAN	iii
ABSTRAK.....	iv
ABSTRACT.....	v
PRAKATA.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	viii
DAFTAR GAMBAR	x
BAB I.....	1
BAB II.....	2
BAB III	33
BAB IV	36
BAB V	56
5.1 Kesimpulan	56
5.2 Saran.....	56
DAFTAR PUSTAKA	57

DAFTAR TABEL

Tabel 1. Keaslian Penelitian	
Tabel 2. Contoh Kasus Regresi.....	19
Tabel 3. Subhimpunan Acak.....	19
Tabel 4. Data Number of Fish	20
Tabel 5. Menghitung Standar Deviasi	21
Tabel 6. Data Sunny.....	22
Tabel 7. Data Overcast.....	
Tabel 8. Data Rainy	
Tabel 9. Menentukan Pemberhentian Cabang	
Tabel 10. Parameter <i>Random Forest</i>	28
Tabel 11. Klasifikasi AUC.....	31
Tabel 12. Nama Fitur	35
Tabel 13. Contoh dataset Heart Failure Clinical Records	36
Tabel 14. Jumlah Distribusi Kelas.....	36
Tabel 15. Pembagian dataset	37
Tabel 16. Contoh simulasi parameter <i>max_feature</i> untuk data dengan 12 fitur.	40
Tabel 17. Range <i>hyper-parameter</i> untuk <i>random search</i>	
Tabel 18. Daftar <i>hyper-parameter</i> dan skor yang dihasilkan <i>random search</i>	
Tabel 19. <i>Hyperparameter default random forest</i>	
Tabel 20. Contoh hasil <i>bootstrapping</i> pada data minoritas	43
Tabel 21. Contoh hasil pemilihan <i>max_feature</i> pada data <i>bootstrap</i>	43
Tabel 22. Contoh hasil prediksi data <i>testing</i> sebelum dilakukan <i>Aggregating</i>	45
Tabel 23. <i>Confusion Matrix Random Forest</i>	45
Tabel 24. <i>Hyperparameter default random forest</i>	46
Tabel 25. <i>Confusion Matrix Random Forest + Random Search</i>	48
Tabel 26. AUC <i>Random Forest</i>	49
Tabel 27. Tabel Perhitungan AUC <i>Random Forest</i>	49
Tabel 28. AUC <i>Random Forest + Random Search</i>	50
Tabel 29. Perhitungan AUC <i>Random Forest + Random Search</i>	
Tabel 30. Perbedaan konfigurasi <i>hyper-parameter</i>	52

Tabel 31. Perbandingan Nilai AUC.....	53
---------------------------------------	----

DAFTAR GAMBAR

Gambar 1. <i>Machine Learning</i>	
Gambar 2. Struktur <i>Decission Tree</i>	
Gambar 3. Ilustrasi <i>ensemble learning</i>	
Gambar 4. <i>Random Forest</i>	18
Gambar 5. Menghitung SDR Setiap Variabel	
Gambar 6. Proses <i>Splitting</i>	
Gambar 7. Pemberhentian Cabang	
Gambar 8. <i>Decision Tree</i>	
Gambar 9. <i>Model Balanced Random Forest</i>	25
Gambar 10. Ilustrasi <i>Random Search</i>	28
Gambar 11. Ilustrasi <i>k-folds cross validation</i>	29
Gambar 12. Contoh <i>AUC Curve</i>	30
Gambar 13. Diagram Perbandingan Jumlah Pasien.....	33
Gambar 14. Alur Penelitian	34
Gambar 15. Perbandingan Persentase Kelas.....	37
Gambar 16. Contoh struktur <i>random forest</i> dengan $n_estimator = 3$	38
Gambar 17. Contoh struktur pohon keputusan dengan $max_depth = 3$	39
Gambar 18. Contoh struktur pohon keputusan yang telah dibuat.....	44
Gambar 19. Kurva <i>AUC Random Forest</i>	49
Gambar 20. Kurva <i>AUC Random Forest + Random Search</i>	50
Gambar 21. Grafik Perbandingan	54