



**PERBANDINGAN ALGORITMA *METAHEURISTIC* PADA SELEKSI FITUR
UNTUK PREDIKSI CACAT *SOFTWARE***

Skripsi

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

**Oleh
HALIMATUS SA'DIYAH
NIM. 1911016220011**

**PROGRAM STUDI S1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2023**



**PERBANDINGAN ALGORITMA *METAHEURISTIC* PADA SELEKSI FITUR
UNTUK PREDIKSI CACAT *SOFTWARE***

Skripsi

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

Oleh

HALIMATUS SA'DIYAH

NIM. 1911016220011

**PROGRAM STUDI S1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2023

SKRIPSI

**PERBANDINGAN ALGORITMA *METAHEURISTIC* PADA SELEKSI
FITUR UNTUK PREDIKSI CACAT *SOFTWARE***

Oleh :

**HALIMATUS SA'DIYAH
1911016220011**

Telah dipertahankan di depan Dosen Penguji pada tanggal 15 Juni 2023

Susunan Dosen Penguji :

Pembimbing I

Raditvo Adi Nugroho, S.T., M.Kom.
NIP. 198212042008011006

Dosen Penguji I

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

Pembimbing II

Friska Abadi, S.Kom., M.Kom.
NIP. 19880913201612110001

Dosen Penguji II

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



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 Juni 2023



Halimatus Sa'diyah
NIM. 1911016220011

ABSTRAK

PERBANDINGAN ALGORITMA *METAHEURISTIC* PADA SELEKSI FITUR UNTUK PREDIKSI CACAT *SOFTWARE*

(Oleh : Halimatus Sa'diyah; Pembimbing: Radityo Adi Nugroho, S.T., M.Kom., dan Friska Abadi, S.Kom., M.Kom; 2023; 74 halaman)

Prediksi cacat *software* telah menjadi bagian penting dalam proses pengujian kualitas *software*. Namun, semakin berkembangnya ukuran dan kompleksitas *software* seringkali menghasilkan jumlah data yang besar. Keberadaan fitur yang tidak relevan dapat mempengaruhi kinerja algoritma klasifikasi dengan meningkatkan komputasi dan penggunaan memori yang tinggi. Masalah ini dikenal sebagai "*Curse of Dimensionality*" di mana semakin tinggi dimensi data, semakin rendah performa algoritma prediksi karena kompleksitas komputasi yang tinggi. Solusi untuk masalah ini adalah menggunakan seleksi fitur. Salah satu metode seleksi fitur yang dapat digunakan adalah metode *Wrapper* yang melibatkan algoritma *machine learning* seperti Algoritma Metaheuristik. Algoritma Genetika dan *Particle Swarm Optimization* adalah dua contoh seleksi fitur dalam Algoritma Metaheuristik. Dalam penelitian ini, peneliti membandingkan kinerja prediksi antara Algoritma Genetika dan *Particle Swarm Optimization* dengan tiga *classifier* populer: KNN, *Support Vector Machine*, dan *Random Forest*. Hasil yang didapat pada penelitian ini adalah kinerja model prediksi yang menggunakan seleksi fitur Algoritma Genetika dan klasifikasi KNN lebih baik dengan rata-rata AUC tertinggi yakni 0,876. Sementara itu, model prediksi yang menggunakan seleksi fitur *Particle Swarm Optimization* dan klasifikasi KNN menghasilkan rata-rata AUC sebesar 0,799.

Kata kunci: Prediksi Cacat *Software*, Algoritma Metaheuristik, Algoritma Genetika, *Particle Swarm Optimization*

ABSTRACT

COMPARISON OF METAHEURISTIC ALGORITHMS IN FEATURE SELECTION FOR SOFTWARE DEFECT PREDICTION

(By : Halimatus Sa'diyah; Supervisor: Radityo Adi Nugroho, S.T., M.Kom and Friska Abadi, S.Kom., M.Kom; 2023; 74 pages)

Software defect prediction has become an essential part of software quality testing. However, as the size and complexity of software continue to grow, it often results in large amounts of data. The presence of irrelevant features can impact the performance of classification algorithms by increasing computational and memory usage. This problem is known as the "Curse of Dimensionality," where higher data dimensions lead to lower prediction algorithm performance due to increased computational complexity. The solution to this problem is feature selection. One feature selection method that can be used is the Wrapper method, which involves machine learning algorithms such as Metaheuristic Algorithms. Genetic Algorithm and Particle Swarm Optimization are two examples of feature selection within the Metaheuristic Algorithm. In this study, the researchers compare the prediction performance between Genetic Algorithm and Particle Swarm Optimization with three popular classifiers: KNN, Support Vector Machine, and Random Forest. The results obtained in this study show that the prediction model's performance using Genetic Algorithm feature selection and KNN classification is better, with the highest average AUC of 0.876. Meanwhile, the prediction model using Particle Swarm Optimization feature selection and KNN classification resulted in an average AUC of 0.799.

Keywords: *Software Defect Prediction, Metaheuristic Algorithms, Genetic Algorithms, Particle Swarm Optimization*

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 Algoritma *Metaheuristic* Pada Seleksi Fitur Untuk Prediksi Cacat *Software*” 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. Orang tua dan keluarga yang selalu memberikan semangat, dukungan, dan doa dari awal perkuliahan hingga proses penyelesaian skripsi ini.
2. Bapak Radityo Adi Nugroho, S.T., M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Friska Abadi, S.Kom., M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin 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.
6. Seluruh staf sekaligus keluarga UPT PTIK ULM yang memberikan dukungan dan bimbingan dalam berbagai hal terkhususnya pada proses penyelesaian skripsi ini.
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, Juni 2023

Halimatus Sa'diyah

DAFTAR ISI

HALAMAN JUDUL	1
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR.....	xii
BAB I.....	1
PENDAHULUAN.....	1
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	4
TINJAUAN PUSTAKA	4
2.1 Kajian Terdahulu	4
2.2 Prediksi Cacat <i>Software</i>	6
2.3 Dataset NASA MDP.....	7
2.4 Seleksi Fitur.....	9
2.5 Algoritma Metaheuristik.....	10
2.6 Algoritma Genetika	10
2.7 Particle Swarm Optimization.....	13
2.8 K-Nearest Neighbor (KNN)	16

2.9	<i>Support Vector Machine (SVM)</i>	17
2.10	Random Forest.....	18
2.11	<i>T Test</i>	19
2.12	<i>Cross Validation</i>	20
2.13	<i>Area Under Curve (AUC)</i>	21
2.14	Keaslian Penelitian	22
BAB III.....		23
METODE PENELITIAN.....		23
3.1	Alat Penelitian	23
3.2	Bahan Penelitian	23
3.3	Variabel Penelitian.....	23
3.4	Prosedur Penelitian	23
BAB IV.....		26
HASIL DAN PEMBAHASAN.....		26
4.1	Hasil.....	26
4.1.1	Pengumpulan Dataset.....	26
4.1.2	Seleksi Fitur	28
4.1.3	Klasifikasi	39
4.1.4	Hasil Perbandingan Algoritma Metaheuristik.....	49
4.2	Pembahasan	50
BAB V.....		56
PENUTUP.....		56
5.1	Kesimpulan.....	56
5.2	Saran	56
DAFTAR PUSTAKA.....		57
RIWAYAT HIDUP.....		61

DAFTAR TABEL

Tabel 1. Spesifikasi Dataset NASA MDP (D'')	7
Tabel 2. Spesifikasi atribut dataset NASA MDP	8
Tabel 3. Spesifikasi atribut dataset NASA MDP (Lanjutan)	9
Tabel 4. Keakuratan hasil klasifikasi berdasarkan AUC	22
Tabel 5. Keaslian Penelitian	22
Tabel 6. Penelitian yang diusulkan	22
Tabel 7. Tahapan Klasifikasi	24
Tabel 8. Daftar Metrik Software	26
Tabel 9. Daftar Metrik Software (Lanjutan)	27
Tabel 10. Jumlah Metrik Software	27
Tabel 11. Spesifikasi Dataset NASA MDP (DS'')	28
Tabel 12. Hasil Seleksi Fitur Menggunakan GA untuk klasifikasi KNN	29
Tabel 13. Perbandingan Jumlah Metrik model GA+KNN	30
Tabel 14. Hasil Seleksi Fitur Menggunakan GA untuk klasifikasi SVM	30
Tabel 15. Hasil Seleksi Fitur Menggunakan GA untuk klasifikasi SVM (Lanjutan)	31
Tabel 16. Perbandingan Jumlah Metrik model GA+SVM	31
Tabel 17. Perbandingan Jumlah Metrik model GA+SVM (Lanjutan)	32
Tabel 18. Hasil Seleksi Fitur Menggunakan GA untuk klasifikasi RF	32
Tabel 19. Hasil Seleksi Fitur Menggunakan GA untuk klasifikasi RF (Lanjutan)	33
Tabel 20. Perbandingan Jumlah Metrik model GA+RF	33
Tabel 21. Hasil Seleksi Fitur Menggunakan PSO untuk klasifikasi KNN	34
Tabel 22. Perbandingan Jumlah Metrik model PSO+KNN	35
Tabel 23. Hasil Seleksi Fitur Menggunakan PSO untuk klasifikasi SVM	35
Tabel 24. Hasil Seleksi Fitur Menggunakan PSO untuk klasifikasi SVM (Lanjutan)	36
Tabel 25. Perbandingan Jumlah Metrik model PSO+SVM	37
Tabel 26. Hasil Seleksi Fitur Menggunakan PSO untuk klasifikasi RF	37
Tabel 27. Hasil Seleksi Fitur Menggunakan PSO untuk klasifikasi RF (Lanjutan)	38

Tabel 28. Perbandingan Jumlah Metrik model PSO+RF.....	38
Tabel 29. Nilai AUC dengan k terbaik untuk setiap model GA	40
Tabel 30. Nilai AUC dengan k terbaik untuk setiap model GA (Lanjutan)	41
Tabel 31. Nilai AUC dengan k terbaik untuk setiap model PSO.....	41
Tabel 32. Nilai AUC Klasifikasi SVM + Algoritma Genetika	43
Tabel 33. Nilai AUC Klasifikasi SVM + Algoritma Genetika (Lanjutan)	44
Tabel 34. Nilai AUC Klasifikasi SVM + PSO.....	44
Tabel 35. Nilai AUC Klasifikasi SVM + PSO (Lanjutan).....	45
Tabel 36. Nilai AUC Klasifikasi Random Forest + Algoritma Genetika	47
Tabel 37. Nilai AUC Klasifikasi Random Forest + PSO.....	48
Tabel 38. Nilai AUC seluruh model	50
Tabel 39. Rata-rata AUC seluruh model.....	54
Tabel 40. Hasil Uji Signifikan	55

DAFTAR GAMBAR

Gambar 1. Klasifikasi Algoritma Metaheuristik	10
Gambar 2. Siklus Algoritma Genetika	13
Gambar 3. Ilustrasi Metode Random Forest	19
Gambar 4. Ilustrasi 10-Fold Cross Validation	21
Gambar 5. Diagram Alur Penelitian.....	24
Gambar 6. Skema optimize parameter	40
Gambar 7. Skema Cross Validation pada klasifikasi KNN	40
Gambar 8. Skema model klasifikasi KNN.....	40
Gambar 9. Grafik nilai AUC pada model GA+KNN.....	41
Gambar 10. Grafik nilai AUC pada model PSO+KNN	42
Gambar 11. Grafik perbandingan nilai AUC pada KNN	42
Gambar 12. Skema Cross Validation pada klasifikasi SVM	43
Gambar 13. Skema model klasifikasi SVM.....	43
Gambar 14. Grafik Nilai AUC Klasifikasi SVM + Algoritma Genetika.....	44
Gambar 15. Grafik Nilai AUC Klasifikasi SVM + PSO	45
Gambar 16. Grafik perbandingan nilai AUC SVM	45
Gambar 17. Skema Cross Validation pada klasifikasi Random Forest	46
Gambar 18. Skema model klasifikasi Random Forest	46
Gambar 19. Grafik nilai AUC Klasifikasi Random Forest + Algoritma Genetika.....	47
Gambar 20. Grafik Nilai AUC Klasifikasi Random Forest + PSO	48
Gambar 21. Grafik perbandingan nilai AUC RF	49
Gambar 22. Nilai AUC seluruh model.....	53
Gambar 23. Grafik rata-rata nilai AUC seluruh model.....	54