



**PREDIKSI CACAT PERANGKAT LUNAK MENGGUNAKAN
RECURSIVE FEATURE ELIMINATION DAN OPTIMASI
HYPERPARAMETER BERBASIS ALGORITMA GENETIKA**

SKRIPSI

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

**Oleh
AYU NURMALA
NIM 2111016220032**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2026**



**PREDIKSI CACAT PERANGKAT LUNAK MENGGUNAKAN
RECURSIVE FEATURE ELIMINATION DAN OPTIMASI
HYPERPARAMETER BERBASIS ALGORITMA GENETIKA**

SKRIPSI

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

**Oleh
AYU NURMALA
NIM 2111016220032**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
JUNI 2026**

SKRIPSI

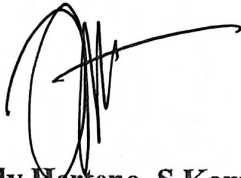
PREDIKSI CACAT PERANGKAT LUNAK MENGGUNAKAN *RECURSIVE FEATURE ELIMINATION* DAN OPTIMASI *HYPERPARAMETER* BERBASIS ALGORITMA GENETIKA

Oleh:

AYU NURMALA
2111016220032

Telah dipertahankan di depan Dosen Penguji pada tanggal 9 Juni 2026.
Susunan Penguji:

Pembimbing Utama



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

Penguji

1.



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

Pembimbing Pendamping



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

2.



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

Banjarbaru, 14 Juli 2026

Koordinator Program Studi Ilmu Komputer



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

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, 01 Mei 2026



Ayu Nurmala

NIM. 2111016220032

ABSTRAK

PREDIKSI CACAT PERANGKAT LUNAK MENGGUNAKAN RECURSIVE FEATURE ELIMINATION DAN OPTIMASI HYPERPARAMETER BERBASIS ALGORITMA GENETIKA

(Oleh: Ayu Nurmala; Pembimbing: Rudy Herteno, S.Kom., M.Kom. dan Muliadi., S.Kom, M.Cs.; 2026; 149 halaman)

Prediksi cacat perangkat lunak merupakan pendekatan strategis dalam meningkatkan kualitas perangkat lunak melalui identifikasi dini modul yang berpotensi mengandung cacat. Tantangan utama dalam prediksi ini meliputi redundansi dan ketidakrelevanan fitur pada dataset, serta ketergantungan performa model terhadap konfigurasi hyperparameter. Penelitian ini mengusulkan kerangka yang mengintegrasikan *Recursive Feature Elimination* (RFE) dan *Genetic Algorithm* (GA) untuk mengoptimasi secara simultan jumlah fitur dan *hyperparameter* model klasifikasi. Penelitian menggunakan 12 *dataset NASA Metrics Data Program* (MDP) versi D" dengan empat model klasifikasi, yaitu *Decision Tree*, *Random Forest*, *XGBoost*, dan *LightGBM*. Evaluasi dilakukan menggunakan skema *10-fold stratified cross-validation* dengan metrik *F1-Score* dan AUC, pengujian signifikansi statistik menggunakan *Wilcoxon signed-rank test*. Hasil penelitian menunjukkan bahwa kombinasi RFE dan optimasi GA meningkatkan performa secara signifikan pada *Decision Tree*, *Random Forest*, dan *LightGBM* dibandingkan skenario RFE tanpa optimasi GA. Peningkatan *F1-Score* terbesar sebesar 46,3% dan rata-rata AUC terbaik sebesar 0,7711 pada *Random Forest*. *XGBoost* tidak menunjukkan peningkatan signifikan, mengindikasikan bahwa konfigurasi pada skenario RFE tanpa optimasi GA sudah cukup kompetitif untuk karakteristik dataset yang digunakan. Secara keseluruhan, integrasi RFE dan GA sebagai kerangka optimasi simultan terbukti efektif dalam meningkatkan performa prediksi cacat perangkat lunak, khususnya pada *Decision Tree*, *Random Forest*, dan *LightGBM*.

Kata Kunci: Prediksi Cacat Perangkat Lunak, *Recursive Feature Elimination*, Algoritma Genetika, Optimasi *Hyperparameter*, NASA MDP

ABSTRACT

SOFTWARE DEFECT PREDICTION USING RECURSIVE FEATURE ELIMINATION AND GENETIC ALGORITHM-BASED HYPERPARAMETER OPTIMIZATION

(By: Ayu Nurmala; Supervisor: Rudy Herteno, S.Kom., M.Kom. and Muliadi, S.Kom, M.Cs.; 2026; 149 pages)

Software defect prediction is a strategic approach to improving software quality through early identification of defect-prone modules. Key challenges include feature redundancy and irrelevance in the dataset, as well as model performance dependence on hyperparameter configurations. This study proposes a framework integrating Recursive Feature Elimination (RFE) and Genetic Algorithm (GA) to simultaneously optimize feature number and classification model hyperparameters. The study used 12 NASA Metrics Data Program (MDP) version D'' datasets with four classification models: Decision Tree, Random Forest, XGBoost, and LightGBM. Evaluation was conducted using 10-fold stratified cross-validation with F1-Score and AUC metrics, and statistical significance was tested using the Wilcoxon signed-rank test. The results show that the combination of RFE and GA optimization significantly improves performance for Decision Tree, Random Forest, and LightGBM compared to RFE without GA optimization. The largest F1-Score improvement was 46.3%, and the best average AUC was 0.7711 for Random Forest. XGBoost did not show significant improvement, indicating that the configuration in the RFE scenario without GA optimization was already competitive for the dataset characteristics. Overall, integrating RFE and GA as a simultaneous optimization framework proved effective in improving software defect prediction performance, particularly for Decision Tree, Random Forest, and LightGBM.

Keywords: *Software Defect Prediction, Recursive Feature Elimination, Genetic Algorithm, Hyperparameter Optimization, NASA MDP*

PRAKATA

Puji Syukur penulis panjatkan ke Tuhan Yang Maha Esa karena atas berkat Rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “**PREDIKSI CACAT PERANGKAT LUNAK MENGGUNAKAN *RECURSIVE FEATURE ELIMINATION* DENGAN OPTIMASI *HYPERPARAMETER* BERBASIS ALGORITMA GENETIKA**” untuk memenuhi syarat dalam menyelesaikan Pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

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, kakak, serta seluruh keluarga besar yang selalu memberikan dukungan, doa, motivasi, dan kasih sayang dalam proses penyelesaian skripsi ini.
2. Bapak Rudy Herteno, S.Kom., M.Kom selaku dosen pembimbing utama yang telah memberikan bimbingan, membantu dan meluangkan waktu dalam proses penyelesaian skripsi ini.
3. Bapak Muliadi, S.Kom., M.Cs., selaku dosen pembimbing pendamping yang telah memberikan bimbingan, membantu dan meluangkan waktu dalam proses penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom dan Bapak Andi Farmadi, S.Si, M.T, selaku dosen penguji dan dosen anggota penguji yang telah memberikan masukan-masukan untuk penyajian skripsi.
5. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom., selaku Dosen Pembimbing Akademik dan Ibu Dwi Kartini, S.Kom., M.Kom., selaku Ketua 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 yang telah memberikan ilmu yang sangat bermanfaat selama penulis menempuh pendidikan.

7. Seluruh sahabat dan teman dan rekan mahasiswa yang telah memberikan banyak dukungan, bantuan, dan doa selama masa studi hingga penyelesaian skripsi ini.
8. Semua pihak yang tidak bisa disebutkan satu persatu, yang turut berperan dan membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penyusunan skripsi ini masih terdapat berbagai kekurangan dan belum mencapai kesempurnaan yang diharapkan. Oleh karena itu, penulis mengharapkan bantuan berupa kritik dan saran 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, 01 Mei 2026



Ayu Nurmala

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	ii
PERNYATAAN.....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	4
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Kajian Terdahulu.....	7
2.2 Software Defect Prediction.....	14
2.3 Dataset NASA MDP.....	15
2.4 Ketidakseimbangan Kelas.....	16
2.5 Normalisasi Data (StandardScaler).....	18
2.6 Dimensionalitas dan Redundansi Fitur.....	19
2.7 Seleksi Fitur (Feature Selection).....	20
2.8 Recursive Feature Elimination (RFE).....	21
2.9 Optimasi Hyperparameter.....	22
2.10 Genetic Algorithm (GA).....	23
2.11 Tree-Based Models.....	25
2.11.1 Decision Tree.....	25
2.11.2 Random Forest.....	27

2.12	Boosting Based Models	28
2.12.1	XGBoost.....	28
2.12.2	LightGBM.....	30
2.13	Stratified K-Fold Cross-Validation	32
2.14	Evaluasi Model.....	34
2.15	Uji Wilcoxon Signed-Rank	36
BAB III METODE PENELITIAN.....		38
3.1	Alat Penelitian.....	38
3.2	Bahan Penelitian.....	38
3.3	Prosedur Penelitian.....	39
3.3.1	Dataset.....	39
3.3.2	Prapemrosesan Data	40
3.3.3	Normalisasi Fitur.....	40
3.3.4	Skema Validasi	40
3.3.5	Seleksi Fitur dan Optimasi Hyperparameter	42
3.3.6	Klasifikasi	45
3.3.7	Evaluasi Model.....	46
3.3.8	Uji Signifikansi	46
3.3.9	Perbandingan Hasil	48
BAB IV HASIL DAN PEMBAHASAN.....		49
4.1	Hasil	49
4.1.1	Pengumpulan Data	49
4.1.2	Prapemrosesan Data	51
4.1.3	Stratified K-Fold Cross Validation.....	52
4.1.4	Normalisasi Fitur.....	57
4.1.5	Seleksi Fitur Menggunakan Recursive Feature Elimination (RFE)	62
4.1.6	Hasil Skenario <i>Baseline</i>	81
4.1.7	Hasil Skenario RFE Tanpa Optimasi GA.....	83
4.1.8	Hasil Skenario RFE dengan Optimasi GA.....	102
4.1.9	Perbandingan Kinerja Model	116

4.1.10	Uji Signifikansi Statistik	121
4.2	Pembahasan.....	126
BAB V PENUTUP.....		134
5.1	Kesimpulan	134
5.2	Saran.....	134
DAFTAR PUSTAKA		136
LAMPIRAN.....		150

DAFTAR TABEL

Tabel 1. Keaslian Penelitian	12
Tabel 2. Perancangan Penelitian	13
Tabel 3. Spesifikasi Dataset NASA MDP D"	49
Tabel 4. Daftar Metrik Dataset NASA MDP D"	50
Tabel 5. Data Awal Dataset MW1	51
Tabel 6. Data Fitur (X) Dataset MW1	52
Tabel 7. Data Label (y) Dataset MW1	52
Tabel 8. Distribusi Data Latih dan Data Validasi 5-Fold Stratified Cross- Validation pada Dataset MW1	54
Tabel 9. Distribusi Data Latih dan Data Uji 10-Fold Stratified Cross-Validation pada Dataset MW1	56
Tabel 10. Nilai Asli Data Latih Fold 1 Dataset MW1 Sebelum Normalisasi	58
Tabel 11. Perhitungan Deviasi Fitur LOC_EXECUTABLE Data Latih Fold 1 Dataset MW1	58
Tabel 12. Nilai Mean (μ) dan Standar Deviasi (σ) Seluruh Fitur Data Latih Fold 1 Dataset MW1	59
Tabel 13. Nilai Setelah Standarisasi pada Data Latih Fold 1 Dataset MW1	61
Tabel 14. Hasil Perbandingan Gini Gain.....	66
Tabel 15. Feature Importance Decision Tree pada Data Latih Fold 1 Dataset MW1	67
Tabel 16. Proses Iterasi RFE dengan Estimator Decision Tree (Fold 1 MW1) .	68
Tabel 17. Perbandingan Gini Gain Root Node Pohon Pertama Random Forest	71
Tabel 18. Ilustrasi Feature Importance Per Pohon pada Random Forest (3 Pohon, 5 Fitur)	72
Tabel 19. Hasil Rata-Rata Feature Importance Random Forest pada Data Latih Fold 1 Dataset MW1	72
Tabel 20. Ranking Akhir Fitur Hasil RFE dengan Estimator Random Forest pada Data Latih Fold 1 Dataset MW1	73
Tabel 21. Feature Importance XGBoost Berdasarkan Gain (Fold 1 MW1)	76
Tabel 22. Iterasi Proses RFE XGBoost (Fold 1, Dataset MW1).....	77

Tabel 23. Hasil Feature Importance LightGBM Berdasarkan Gain pada Data Latih Fold 1 Dataset MW1	79
Tabel 24. Perbandingan Fitur Terpilih RFE Keempat Model untuk Data Latih Fold 1 Dataset MW1	80
Tabel 25. Hasil F1-Score Skenario Baseline	81
Tabel 26. Hasil AUC Skenario Baseline	82
Tabel 27. Konfigurasi Hyperparameter Baseline Skenario RFE Tanpa Optimasi GA	84
Tabel 28. Precision-Recall-F1 Decision Tree pada Data Latih Fold 1 MW1	86
Tabel 29. Hasil Evaluasi per Fold Decision Tree RFE-only pada Dataset MW1	88
Tabel 30. Precision-Recall-F1 Random Forest di Sekitar Threshold Optimal (Fold 1 MW1)	90
Tabel 31. Hasil Evaluasi per Fold Random Forest RFE-only pada Dataset MW1	91
Tabel 32. Precision-Recall-F1 XGBoost di Sekitar Threshold Optimal (Fold 1 MW1)	94
Tabel 33. Hasil Evaluasi per Fold XGBoost RFE-only pada Dataset MW1	95
Tabel 34. Precision-Recall-F1-LightGBM di Sekitar Threshold Optimal (Fold 1 MW1)	98
Tabel 35. Hasil Evaluasi per Fold LightGBM RFE-only pada Dataset MW1 ...	99
Tabel 36. Hasil F1-Score Skenario RFE Tanpa Optimasi GA	100
Tabel 37. Hasil AUC Skenario RFE Tanpa Optimasi GA	101
Tabel 38. Rentang Nilai Pencarian Jumlah Fitur dan Hyperparameter GA per Model	103
Tabel 39. Log Konvergensi GA Decision Tree pada Dataset CM1	107
Tabel 40. Parameter GA Adaptif per Dataset	109
Tabel 41. Parameter Optimal GA pada Decision Tree	110
Tabel 42. Parameter Optimal GA pada Random Forest	111
Tabel 43. Parameter Optimal GA pada XGBoost	111
Tabel 44. Parameter Optimal GA pada LightGBM	112

Tabel 45. Hasil F1-Score Skenario RFE dengan Optimasi GA	114
Tabel 46. Hasil AUC Skenario RFE dengan Optimasi GA.....	115
Tabel 47. Perbandingan Rata-rata F1-Score pada Seluruh Skenario	116
Tabel 48. Perbandingan Rata-rata AUC pada Seluruh Skenario.....	117
Tabel 49. Selisih F1-Score	118
Tabel 50. Selisih AUC.....	119
Tabel 51. Perhitungan Wilcoxon Signed-Rank Test Random Forest F1-Score	123
Tabel 52. Perhitungan Wilcoxon Signed-Rank Test Random Forest AUC.....	123
Tabel 53. Hasil Uji Wilcoxon Signed-Rank Test	124

DAFTAR GAMBAR

Gambar 1. Struktur Dasar Decision Tree ((Jijo & Abdulazeez, 2021).....	25
Gambar 2. Ilustrasi Arsitektur Random Forest (Maulida et al., 2023).....	28
Gambar 3 Ilustrasi Arsitektur XGBoost (Wang et al., 2019).	29
Gambar 4. K-fold cross-validation (Ren et al., 2019).....	32
Gambar 5. Confusion Matrix (Liang, 2022).	34
Gambar 6. Alur Penelitian.....	39
Gambar 7. Grafik Konvergensi GA Decision Tree pada Dataset CM1	108
Gambar 8. Metrik F1-Score	128
Gambar 9. Metrik AUC.....	128
Gambar 10. Heatmap Perubahan F1-Score antara Skenario RFE-only dan RFE+GA pada seluruh Model dan Dataset	129
Gambar 11. Heatmap Perubahan AUC antara Skenario RFE-only dan RFE+GA pada seluruh Model dan Dataset	129