



**OPTIMASI SELEKSI FITUR *RECURSIVE FEATURE ELIMINATION*
MENGUNAKAN *SHAPLEY ADDITIVE EXPLANATIONS* DALAM
PREDIKSI CACAT *SOFTWARE* DENGAN KLASIFIKASI *LIGHTGBM***

Skripsi

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

Oleh

HARTATI

NIM 1911016220002

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

JUNI 2024



**OPTIMASI SELEKSI FITUR *RECURSIVE FEATURE ELIMINATION*
MENGUNAKAN *SHAPLEY ADDITIVE EXPLANATIONS* DALAM
PREDIKSI CACAT *SOFTWARE* DENGAN KLASIFIKASI LIGHTGBM**

Skripsi

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

Oleh

HARTATI

NIM 1911016220002

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

JUNI 2024

SKRIPSI

OPTIMASI SELEKSI FITUR *RECURSIVE FEATURE ELIMINATION* MENGUNAKAN *SHAPLEY ADDITIVE EXPLANATIONS* DALAM PREDIKSI CACAT *SOFTWARE* DENGAN KLASIFIKASI LIGHTGBM

Oleh:

HARTATI

NIM. 1911016220002

Telah dipertahankan di depan Dosen Penguji pada tanggal 10 Juni 2024

Susunan Dosen Penguji:

Pembimbing I



Rudy Herteno, S.Kom., M.Kom

NIP. 198809252022031003

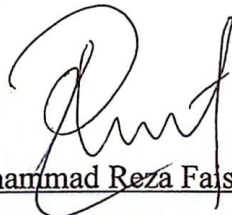
Dosen Penguji I



Fatma Indriani, S.T, M.I.T, Ph.D

NIP. 198404202008122004

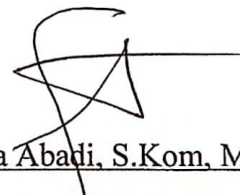
Pembimbing II



Mohammad Reza Fajsal., S.T., M.T., Ph.D

NIP. 197612202008121001

Dosen Penguji II



Friska Abadi, S.Kom, M.Kom

NIP. 198809132023211010



Banjarbaru, 10 Juni 2024

Koordinator Program Studi Ilmu Komputer

Wahid 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,

Yang Menyatakan,



Hartati

NIM.1911016220002

ABSTRAK

OPTIMASI SELEKSI FITUR RECURSIVE FEATURE ELIMINATION MENGGUNAKAN SHAPLEY ADDITIVE EXPLANATIONS DALAM PREDIKSI CACAT SOFTWARE DENGAN KLASIFIKASI LIGHTGBM

(Oleh : Hartati; Pembimbing: Rudy Herteno, S.Kom., M.Kom dan M. Reza Faisal., S.T., M.T., Ph.D.; 2024; 82 halaman)

Cacat perangkat lunak adalah permasalahan dimana perangkat lunak tidak berfungsi dengan baik. Kesalahan dalam proses pengembangan perangkat lunak adalah alasan penyebab cacat perangkat lunak. Pembelajaran mesin yaitu klasifikasi digunakan untuk mengklasifikasikan cacat dalam perangkat lunak. Untuk meningkatkan model klasifikasi, diperlukan pemilihan fitur terbaik dari dataset. Recursive Feature Elimination adalah salah satu metode seleksi fitur. Shapley Additive Explanation (SHAP) adalah metode yang dapat mengoptimalkan algoritma seleksi fitur untuk dapat memberikan hasil yang lebih baik. Pada penelitian ini, metode boosting populer yaitu LightGBM akan dipilih sebagai pengklasifikasi untuk memprediksi cacat perangkat lunak. Sementara itu, RFE-SHAP akan digunakan sebagai seleksi fitur untuk memilih subset fitur terbaik. Hasil dan pembahasan menunjukkan bahwa seleksi fitur RFE-SHAP sedikit mengungguli RFE, dengan nilai rata-rata AUC masing-masing 0.826 dan 0.816 secara berurutan. Metode RFE-SHAP membuat perbedaan lebih signifikan dibandingkan RFE dalam uji T-Test AUC. Selain itu, metode RFE-SHAP terbukti memiliki penggunaan jumlah fitur yang lebih efisien dibandingkan metode RFE.

Kata kunci: Klasifikasi, Seleksi Fitur, LightGBM, RFE, SHAP, Prediksi Cacat *Software*

ABSTRACT

RECURSIVE FEATURE ELIMINATION OPTIMIZATION USING SHAPLEY ADDITIVE EXPLANATIONS IN SOFTWARE DEFECT PREDICTION WITH LIGHTGBM CLASSIFICATION

(By : Hartati; Supervisor: Rudy Herteno, S.Kom., M.Kom dan M. Reza Faisal., S.T., M.T., Ph.D.; 2024; 82 pages)

Software defect refers to issues where the software does not function properly. The mistakes in the software development process are the reasons for software defects. Software defect prediction is performed to ensure the software is defect-free. Machine learning classification is used to classify defects in software. To improve the classification model, it is necessary to select the best features from the dataset. Recursive Feature Elimination (RFE) is a feature selection method. Shapley Additive Explanations (SHAP) is a method that can optimize feature selection algorithms to produce better results. In this research, the popular boosting algorithm LightGBM will be selected as a classifier to predict software defects. Meanwhile, RFE-SHAP will be used for feature selection to identify the best subset of features. The results and discussion show that RFE-SHAP feature selection slightly outperforms RFE, with average AUC values of 0.826 and 0.816, respectively. RFE-SHAP made a more significant difference than RFE in the AUC T-Test. In addition, the RFE-SHAP method has been proven to use features more efficiently than the RFE method.

Keywords: *Classification, Feature Selection, LightGBM, RFE, SHAP, Software Defect Prediction*

PRAKATA

Puji syukur penulis panjatkan ke Tuhan Yang Maha Esa karena atas berkat kebaikan dan hikmat-Nya penulis dapat menyelesaikan skripsi yang berjudul “OPTIMASI SELEKSI FITUR RECURSIVE FEATURE ELIMINATION MENGGUNAKAN SHAPLEY ADDITIVE EXPLANATIONS DALAM PREDIKSI CACAT SOFTWARE DENGAN KLASIFIKASI LIGHTGBM” 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, saudara dan keluarga yang selalu memberikan doa dan dukungan dalam mengerjakan skripsi.
2. Bapak Rudy Herteno, S.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak M. Reza Faisal., S.T., M.T., Ph.D. 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 Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman komsel 4 yang selalu memberikan doa dan semangat selama proses pembuatan skripsi.
7. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
8. 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 serta perkenanan dari Tuhan Yang Maha Esa.

Banjarbaru, 5 Juni 2024



Hartati

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN.....	iii
ABSTRAK	iv
<i>ABSTRACT</i>	v
PRAKATA	vi
DAFTAR ISI.....	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan	3
1.4 Manfaat Penelitian	3
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA	5
2.1 Kajian Terdahulu	5
2.2 Landasan Teori	7
2.2.1 Prediksi Cacat Software.....	7
2.2.2 Dataset NASA MDP.....	8
2.2.3 Recursive Feature Elimination	10
2.2.4 Shapley Additive Explanations	11
2.2.5 LightGBM.....	12
2.2.6 Cross Validation	14
2.2.7 Area Under Curve (AUC).....	15

BAB III METODE PENELITIAN	17
3.1 Bahan Penelitian	17
3.2 Alat Penelitian	17
3.3 Variabel Penelitian.....	17
3.4 Prosedur Penelitan	17
BAB IV HASIL DAN PEMBAHASAN	20
4.1 Hasil	20
4.1.1 Pengumpulan Dataset	20
4.1.2 Label Encoding.....	21
4.1.3 Klasifikasi dan Evaluasi Tanpa Seleksi Fitur.....	22
4.1.4 Klasifikasi dan Evaluasi dengan Seleksi Fitur RFE	23
4.1.5 Klasifikasi dan Evaluasi dengan Seleksi Fitur RFE-SHAP	37
4.2 Pembahasan	47
BAB V PENUTUP	54
5.1 Kesimpulan.....	54
5.2 Saran	54
DAFTAR PUSTAKA	55
LAMPIRAN	59

DAFTAR TABEL

Tabel	Halaman
Tabel 1 Keaslian Penelitian.....	7
Tabel 2 Perencanaan Penelitan.....	7
Tabel 3 Deskripsi dataset NASA MDP.....	9
Tabel 4 Spesifikasi Dataset NASA MDP (D``).....	9
Tabel 5 (lanjutan).....	10
Tabel 6 Keakuratan Kinerja Klasifikasi Berdasarkan Nilai AUC.....	16
Tabel 7 Daftar Metrik Software.....	20
Tabel 8 Hasil Evaluasi Klasifikasi LightGBM Tanpa Seleksi Fitur.....	22
Tabel 9. Contoh dataset NASA MDP.....	23
Tabel 10 Hasil perhitungan <i>gradient</i> dan <i>hessian</i>	24
Tabel 11 Hasil seleksi fitur RFE pada 12 Dataset NASA MDP D``	35
Tabel 12 (lanjutan).....	36
Tabel 13 Hasil Seleksi Fitur RFE dan nilai evaluasinya.....	36
Tabel 14 (lanjutan).....	37
Tabel 15 Hasil Nilai Shapley.....	37
Tabel 16 Hasil seleksi fitur RFE-SHAP pada 12 Dataset NASA MDP D``	45
Tabel 17 (lanjutan).....	46
Tabel 18 Hasil Seleksi Fitur RFE-SHAP dan nilai evaluasinya.....	46
Tabel 19 (Lanjutan)	47
Tabel 20 Rangkuman Hasil AUC.....	49
Tabel 21 Hasil Paired T-Test RFE dibandingkan tanpa seleksi fitur	52
Tabel 22 Hasil Paired T Test RFE-SHAP dibandingkan tanpa seleksi fitur.....	52

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1 Strategi <i>level-wise</i>	13
Gambar 2 Strategi <i>leaf-wise</i>	14
Gambar 3 Ilustrasi Stratified 10-fold Cross Validation	15
Gambar 4 Alur Penelitian.....	18
Gambar 5 Kandidat <i>split point</i> fitur Branch_Count	25
Gambar 6 Hasil split point fitur Branch_Count kandidat pertama.....	25
Gambar 7. Hasil split point fitur Branch_Count kandidat kedua	26
Gambar 8. Kandidat split point fitur Branch_Count.....	26
Gambar 9. Hasil split point fitur Call_Pairs kandidat pertama	26
Gambar 10 Hasil split point fitur Call_Pairs kandidat kedua.....	27
Gambar 11 Hasil <i>split point</i> fitur Call_Pairs kandidat ketiga	27
Gambar 12. Hasil Decision Tree yang dibangun dengan LightGBM.....	28
Gambar 13. Hasil Evaluasi RFE pada Dataset CM1	29
Gambar 14 Hasil Evaluasi RFE pada Dataset JM1	30
Gambar 15 Hasil Evaluasi RFE pada dataset KC1.....	30
Gambar 16 Hasil Evaluasi RFE pada dataset KC3.....	31
Gambar 17 Hasil Evaluasi RFE pada dataset MC1	31
Gambar 18 Hasil Evaluasi RFE pada dataset MC2	32
Gambar 19 Hasil Evaluasi RFE pada dataset MW1	32
Gambar 20 Hasil Evaluasi RFE pada dataset PC1	33
Gambar 21 Hasil Evaluasi RFE pada dataset PC2	33
Gambar 22 Hasil Evaluasi RFE pada dataset PC3	34
Gambar 23 Hasil Evaluasi RFE pada dataset PC4	34
Gambar 24 Hasil Evaluasi RFE pada dataset PC5	35
Gambar 25 Waterfall SHAP plot untuk data pertama	38
Gambar 26. Mean SHAP plot.....	38
Gambar 27. Hasil Evaluasi RFE-SHAP pada Dataset CM1	39
Gambar 28 Hasil Evaluasi RFE-SHAP pada Dataset JM1	40
Gambar 29 Hasil Evaluasi RFE-SHAP pada Dataset KC1.....	40

Gambar 30 Hasil Evaluasi RFE-SHAP pada Dataset KC3.....	41
Gambar 31 Hasil Evaluasi RFE-SHAP pada Dataset MC1	41
Gambar 32 Hasil Evaluasi RFE-SHAP pada Dataset MC2	42
Gambar 33 Hasil Evaluasi RFE-SHAP pada Dataset MW1	42
Gambar 34 Hasil Evaluasi RFE-SHAP pada Dataset PC1	43
Gambar 35 Hasil Evaluasi RFE-SHAP pada Dataset PC2	43
Gambar 36 Hasil Evaluasi RFE-SHAP pada Dataset PC3	44
Gambar 37 Hasil Evaluasi RFE-SHAP pada Dataset PC4	44
Gambar 38 Hasil Evaluasi RFE-SHAP pada Dataset PC5	45
Gambar 39 Grafik hasil AUC prediksi cacat <i>software</i>	50
Gambar 40 Grafik Rata-rata AUC.....	50
Gambar 41. Perbandingan jumlah fitur optimal pada semua dataset	51

DAFTAR LAMPIRAN

Lampiran

Lampiran 1 *Source Code* Tanpa Seleksi Fitur

Lampiran 2 *Source Code* Seleksi Fitur RFE

Lampiran 3 *Source Code* Seleksi Fitur RFE-SHAP

Lampiran 4 *Source Code* Visualisasi LightGBM

Lampiran 5 *Source Code* Contoh Memperoleh Nilai Shapley

Lampiran 6 *Source Code* Waterfall Plot SHAP

Lampiran 7 *Source Code* Mean Plot SHAP