



**ANALISIS KOMPARATIF METODE EKSTRAKSI FITUR PADA
GRADIENT BOOSTING CLASSIFIER DALAM *SOFTWARE DEFECT*
*PREDICTION***

Skripsi

Untuk memenuhi persyaratan
melakukan penelitian dalam rangka penyusunan skripsi

Oleh

MUHAMMAD RIZKI AKBAR

NIM.2111016210006

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
NOVEMBER 2025**



**ANALISIS KOMPARATIF METODE EKSTRAKSI FITUR PADA
GRADIENT BOOSTING CLASSIFIER DALAM *SOFTWARE DEFECT
PREDICTION***

Skripsi

Untuk memenuhi persyaratan
melakukan penelitian dalam rangka penyusunan skripsi

Oleh

MUHAMMAD RIZKI AKBAR

NIM.2111016210006

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU
NOVEMBER 2025**

SKRIPSI

ANALISIS KOMPARATIF METODE EKSTRAKSI FITUR PADA *GRADIENT BOOSTING CLASSIFIER* DALAM *SOFTWARE DEFECT PREDICTION*

Oleh:

Muhammad Rizki Akbar

NIM. 2111016210006

Telah dipertahankan di depan Dosen Penguji pada tanggal 11 November 2025.

Susunan Dosen Penguji:

Pembimbing I



Rudy Herton, S.Kom., M.Kom.

NIP. 198809252022031003

Dosen Penguji I



Radityo Adi Nugroho, S.T., M.Kom.

NIP. 198212042008011006

Pembimbing II



Friska Abadi, S.Kom., M.Kom.

NIP. 198809132023211010

Dosen Penguji II



Setyo Wahyu Saputro, S.Kom., M.Kom.

NIP. 198808072023211027

Banjarbaru, 11 Desember 2025

Koordinator Program Studi Ilmu Komputer



Evy Sartini, 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, 17 Desember 2025

Yang Menyatakan

A handwritten signature in black ink, appearing to be 'Rizki Akbar', with a stylized flourish at the end.

Muhammad Rizki Akbar

NIM.21110106210006

ABSTRAK

ANALISIS KOMPARATIF METODE EKSTRAKSI FITUR PADA GRADIENT BOOSTING CLASSIFIER DALAM SOFTWARE DEFECT PREDICTION

(Oleh: Muhammad Rizki Akbar; Pembimbing: Rudy Herteno, S.Kom., M.Kom. dan Friska Abadi, S.Kom., M.Kom.; 2025; 70 halaman)

Software Defect Prediction (SDP) atau Prediksi Cacat Perangkat Lunak merupakan teknik penting untuk meningkatkan kualitas perangkat lunak dengan membangun berbagai model klasifikasi menggunakan pendekatan *machine learning*. Salah satu tantangan utama dalam SDP, seperti pada penelitian yang menggunakan dataset NASA MDP, adalah adanya redundansi dan ketidakrelevanan fitur pada dataset, yang dapat menurunkan kinerja model. Salah satu solusi untuk masalah tersebut adalah metode ekstraksi fitur, yang dapat meminimalkan redundansi dengan mentransformasi data ke dalam bentuk baru yang lebih representatif. Penelitian ini membandingkan empat metode ekstraksi fitur, yaitu *Principal Component Analysis* (PCA), *Linear Discriminant Analysis* (LDA), Kernel PCA (K-PCA), dan Autoencoder, pada dataset NASA MDP, untuk mencari tahu mana metode yang bekerja lebih baik dalam penelitian SDP. Dua *classifier* digunakan, yaitu *Extreme Gradient Boosting* (XGBoost) dan *Light Gradient Boosting Machine* (LightGBM), yang dikenal dalam keandalannya dalam menangani dataset tidak seimbang, dengan *hyperparameter tuning* Optuna untuk memaksimalkan kinerja model. Evaluasi dilakukan berdasarkan metrik *Area Under Curve* (AUC). Hasil eksperimen menunjukkan bahwa PCA memberikan hasil terbaik dalam hal AUC untuk kedua *classifier*, dengan nilai rata-rata AUC sebesar 0,876 untuk XGBoost dan 0,8596 untuk LightGBM. Dengan demikian, dapat disimpulkan bahwa PCA memiliki keunggulan dibandingkan LDA, K-PCA, dan Autoencoder dalam konteks penelitian ini.

Kata Kunci: *Prediksi Cacat Perangkat Lunak, Ekstraksi Fitur, Gradient Boosting Classifier*

ABSTRACT

A COMPARATIVE ANALYSIS OF FEATURE EXTRACTION METHODS FOR GRADIENT BOOSTING CLASSIFIER IN SOFTWARE DEFECT PREDICTION

(By: Muhammad Rizki Akbar; Supervisor: Rudy Herteno, S.Kom., M.Kom. dan Friska Abadi, S.Kom., M.Kom.; 2025; 70 pages)

Software Defect Prediction (SDP) is an essential technique for improving software quality by building various classification models using machine learning approaches. One of the main challenges in SDP, particularly in studies that utilize the NASA MDP dataset, is the presence of redundant and irrelevant features, which can reduce model performance. A common solution to this issue is feature extraction, which minimizes redundancy by transforming the data into a more representative form. This study compares four feature extraction methods—Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Kernel PCA (K-PCA), and Autoencoder—on the NASA MDP dataset to determine which method performs best in SDP research. Two classifiers were employed: Extreme Gradient Boosting (XGBoost) and Light Gradient Boosting Machine (LightGBM), both known for their robustness in handling imbalanced datasets, complemented by hyperparameter tuning using Optuna to maximize model performance. Evaluation was conducted using the Area Under the Curve (AUC) metric. The experimental results show that PCA achieves the best AUC performance for both classifiers, with an average AUC of 0,876 for XGBoost and 0,8596 for LightGBM. Therefore, it can be concluded that PCA outperforms LDA, K-PCA, and Autoencoder in the context of this study.

Keywords: *Software Defect Prediction, Feature Extraction, Gradient Boosting Classifier*

PRAKATA

Puji syukur penulis panjatkan ke Tuhan Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Analisis Komparatif Metode Ekstraksi Fitur pada *Gradient Boosting Classifier* dalam *Software Defect Prediction*” 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:

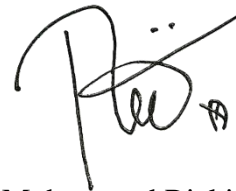
1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Rudy Herteno, S.Kom., M.Kom selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Friska Abadi, S.Kom., M. Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Ibu Dwi Kartini, S.Kom., M.Kom selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Bapak Radityo Adi Nugroho, S.T., M.Kom dan Bapak Setyo Wahyu Saputro, S.Kom., M.Kom selaku dosen penguji yang telah memberikan masukan-masukkan dalam skripsi ini.
6. Seluruh dosen dan staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
7. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2021 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 khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 17 Desember 2025

A handwritten signature in black ink, appearing to be 'Rizki Akbar', with a stylized flourish at the end.

Muhammad Rizki Akbar

DAFTAR ISI

HALAMAN JUDUL.....	i
PERNYATAAN.....	iii
ABSTRAK.....	iv
<i>ABSTRACT</i>	v
PRAKATA.....	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR.....	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian.....	3
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
BAB II TINJAUAN PUSTAKA.....	5
2.1 Kajian Terdahulu.....	5
2.2 <i>Software Defect Prediction</i>	6
2.3 Dataset NASA MDP.....	7
2.4 <i>Preprocessing</i>	7
2.5 Ekstraksi Fitur.....	8
2.5.1 Principal Component Analysis (PCA).....	8
2.5.2 Linear Discriminant Analysis (LDA).....	9
2.5.3 Kernel Principal Component Analysis (K-PCA).....	11
2.5.4 Autoencoder.....	12

2.6	<i>Gradient Boosting Classifier</i>	14
2.6.1	XGBoost	16
2.6.2	LightGBM	17
2.7	<i>Hyperparameter tuning</i>	18
2.8	Evaluasi Model	21
BAB III METODE PENELITIAN.....		23
3.1	Alat Penelitian	23
3.2	Bahan Penelitian	23
3.3	Variabel Penelitian.....	23
3.4	Prosedur Penelitian	24
BAB IV HASIL DAN PEMBAHASAN		26
4.1	Hasil.....	26
4.1.1	Pengumpulan Data.....	26
4.1.2	<i>Preprocessing</i>	28
4.1.3	Ekstraksi Fitur.....	30
4.1.4	Klasifikasi	51
4.1.5	<i>Hyperparameter tuning</i>	56
4.1.5	Evaluasi.....	61
a.	AUC pada XGBoost	62
b.	AUC pada LightGBM.....	63
4.2	Pembahasan	64
BAB V PENUTUP.....		70
5.1	Kesimpulan	70
5.2	Saran	70
DAFTAR PUSTAKA		71

LAMPIRAN..... 78

DAFTAR TABEL

Tabel 1. <i>Confusion Matrix</i>	21
Tabel 2. Daftar Metrik <i>Software</i>	26
Tabel 3. Spesifikasi Dataset NASA MDP D”	28
Tabel 4. Data Latih Dataset CM1	29
Tabel 5. Contoh Hasil Standarisasi dengan StandardScaler pada Data Latih CM1	30
Tabel 6. Contoh Matriks Proyeksi Hasil PCA pada Data Latih CM1	35
Tabel 7. Contoh Hasil Ekstraksi Fitur PCA + XGBoost pada Data Latih CM1 dengan <i>Hyperparameter Tuning</i>	36
Tabel 8. Dataset Contoh Setelah Standarisasi	36
Tabel 9. Contoh Proyeksi Data Menggunakan LDA	39
Tabel 10. Contoh Matriks Proyeksi Hasil LDA pada Data Latih CM1	40
Tabel 11. Contoh Hasil Ekstraksi Fitur LDA pada Data Latih CM1	41
Tabel 12. Contoh Matriks Proyeksi K-PCA (RBF) Pada Data Latih CM1	44
Tabel 13. Contoh Matriks Proyeksi K-PCA (Polynomial) Pada Data Latih CM1	44
Tabel 14. Contoh Matriks Proyeksi K-PCA (Sigmoid) Pada Data Latih CM1	44
Tabel 15. Contoh Hasil Ekstraksi Fitur K-PCA (RBF) pada Data Latih CM1	45
Tabel 16. Contoh Hasil Ekstraksi Fitur K-PCA (Polynomial) pada Data Latih CM1	45
Tabel 17. Contoh Hasil Ekstraksi Fitur K-PCA (Sigmoid) pada Data Latih CM1	45
Tabel 18. Contoh Matriks Proyeksi (Bobot Encoder) Autoencoder pada Data Latih CM1	51
Tabel 19. Contoh Hasil Transformasi Autoencoder pada Data Latih CM1	51
Tabel 20. Dataset Contoh untuk <i>Gradient Boosting</i>	53

Tabel 21. Contoh <i>output</i> Optuna pada PCA + LightGBM	57
Tabel 22. Rentang Pencarian <i>Hyperparameter</i> untuk Metode Ekstraksi Fitur.....	58
Tabel 23. Rentang Pencarian <i>Hyperparameter</i> untuk Algoritma Klasifikasi	58
Tabel 24. Hasil Parameter Terbaik untuk Dataset CM1, JM1, KC1, KC3, MC1, dan MC2 pada XGBoost.....	59
Tabel 25. Hasil Parameter Terbaik untuk Dataset MW1, PC1, PC2, PC3, PC4, dan PC5 pada XGBoost	60
Tabel 26. Hasil Parameter Terbaik untuk Dataset CM1, JM1, KC1, KC3, MC1, dan MC2 pada LightGBM	61
Tabel 27. Hasil Parameter Terbaik untuk Dataset MW1, PC1, PC2, PC3, PC4, dan PC5 pada LightGBM	61
Tabel 28. Hasil Evaluasi AUC pada XGBoost	62
Tabel 29. Hasil Evaluasi AUC pada LightGBM.....	63
Tabel 30. Rata-Rata Nilai AUC XGBoost	65
Tabel 31. Rata-Rata Nilai AUC LightGBM	65
Tabel 32. Hasil Uji Post-hoc Nemenyi XGBoost	68
Tabel 33. Hasil Uji Post-hoc Nemenyi LightGBM.....	68

DAFTAR GAMBAR

Gambar 1. Perbedaan <i>Level-wise Tree Growth</i> dan <i>Leaf-wise Tree Growth</i>	18
Gambar 2. Alur Penelitian.....	24
Gambar 3. AUC XGBoost	63
Gambar 4. AUC LightGBM.....	64
Gambar 5. Rata-rata AUC XGBoost.....	66
Gambar 6. Rata-rata AUC LightGBM.....	67