



**OPTIMASI BACKWARD ELIMINATION DENGAN
PENDEKATAN MULTI FILTER FEATURE SELECTION PADA
SOFTWARE DEFECT PREDICTION**

SKRIPSI

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

**Oleh
MUHAMMAD KHALID MAULANA**

NIM 2211016210020

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

MEI 2026



**OPTIMASI BACKWARD ELIMINATION DENGAN
PENDEKATAN MULTI FILTER FEATURE SELECTION PADA
SOFTWARE DEFECT PREDICTION**

SKRIPSI

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

**Oleh
MUHAMMAD KHALID MAULANA**

NIM 2211016210020

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

MEI 2026

SKRIPSI

OPTIMASI BACKWARD ELIMINATION DENGAN PENDEKATAN MULTI FILTER FEATURE SELECTION PADA SOFTWARE DEFECT PREDICTION

Oleh :

MUHAMMAD KHALID MAULANA

NIM. 2211016210020

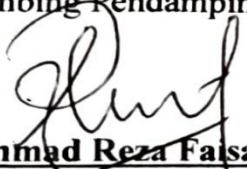
telah dipertahankan di depan Dosen Penguji pada tanggal 20 Mei 2026

Susunan Penguji :

Pembimbing Utama


Setyo Wahyu Saputro, S.Kom., M.Kom.
NIP. 198808072023211027

Pembimbing Pendamping


Mohammad Reza Faisal, S.Si., S.T., M.T., PhD.
NIP. 197612202008121001

Penguji

1.


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

2.


As'ary Ramadhan, S.Kom., M.Cs.
NIP. 199103292024061001

Banjarbaru, 21 Mei 2026

Koordinator Program Studi Ilmu Komputer


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



PRAKATA

Puji dan syukur senantiasa penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, serta hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul **“OPTIMASI BACKWARD ELIMINATION DENGAN PENDEKATAN MULTI FILTER FEATURE SELECTION PADA SOFTWARE DEFECT PREDICTION”** dengan baik. Penulisan skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Strata-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Dalam proses penyusunan skripsi ini, penulis menyadari bahwa penyelesaian karya ilmiah ini tidak terlepas dari berbagai bantuan, dukungan, bimbingan, serta doa dari banyak pihak. Oleh karena itu, dengan segala kerendahan hati, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Allah SWT atas segala limpahan rahmat, karunia, serta ridho-Nya yang senantiasa menyertai setiap langkah penulis dalam menyelesaikan skripsi ini.
2. Ayah, ibu, serta kedua saudari penulis yang selalu memberikan doa, kasih sayang, motivasi, dan dukungan yang tiada henti, sehingga penulis dapat menyelesaikan skripsi ini dengan penuh semangat.
3. Bapak Setyo Wahyu Saputro, S.Kom., M.Kom. dan Bapak Mohammad Reza Faisal, S.Si., S.T., M.T., PhD., selaku dosen pembimbing utama dan pendamping yang telah dengan sabar memberikan bimbingan, arahan, masukan, serta ilmu yang sangat berharga selama proses penyusunan skripsi ini.
4. Ibu Dwi Kartini, S.Kom., M.Kom., selaku Ketua Program Studi Ilmu Komputer, beserta seluruh dosen dan staf Fakultas MIPA Universitas Lambung Mangkurat yang telah memberikan ilmu, bantuan, serta pelayanan yang sangat berarti selama masa studi penulis.

5. Bang Michael, Azra, Roihan, Zhiman dan Rahmat yang telah menemani proses penulisan skripsi ini, memberikan semangat, ketenangan, serta harapan kepada penulis selama penelitian dan penulisan skripsi ini.
6. Teman-teman seperjuangan di Ilmu Komputer angkatan 2022 serta sahabat-sahabat penulis yang selalu memberikan dukungan, semangat, dan kebersamaan selama masa perkuliahan hingga penyelesaian skripsi ini.
7. Semua pihak yang telah membantu, baik secara langsung maupun tidak langsung, yang tidak dapat penulis sebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih memiliki berbagai keterbatasan dan jauh dari kata sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan dan penyempurnaan di masa yang akan datang.

Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi pembaca serta menjadi kontribusi yang berarti dalam pengembangan ilmu pengetahuan, khususnya di bidang Ilmu Komputer.

Banjarbaru, 21 Mei 2026



Muhammad Khalid Maulana

NIM. 2211016210020

PERNYATAAN

Dengan ini saya menyatakan bahwa didalam 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 didalam naskah ini dan disebutkan dalam daftar pustaka.

Banjarbaru, 21 Mei 2026



Muhammad Khalid Maulana

NIM. 2211016210020



**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS LAMBUNG MANGKURAT
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
JURUSAN ILMU KOMPUTER
PROGRAM STUDI ILMU KOMPUTER**

Jalan Jenderal Ahmad Yani KM 36, Banjarbaru, Kalimantan Selatan 70714
Telp/Fax (0511) 4773 112 Laman : <http://fmipa.ulm.ac.id>

**BERITA ACARA
PERSETUJUAN REKOGNISI KEGIATAN PENGGANTI SKRIPSI**

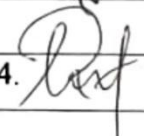
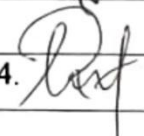
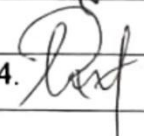
Pada hari ini, Kamis Tanggal 21 Mei 2026 , telah dilakukan evaluasi kelengkapan dokumen pengganti kegiatan skripsi mahasiswa SI Ilmu Komputer Fakultas MIPA ULM

Nama : Muhammad Khalid Maulana
NIM : 2211016210020
Jenis Kegiatan : Jurnal Nasional
Kualifikasi Jurnal/Seminar : Sinta 2
Judul Publikasi : Enhancing Software Defect Prediction through Hybrid Multi-Filter Feature Selection and Imbalance Handling
Nama Jurnal/Penyelenggara : Journal of Computing Theories and Applications (JCTA)
Tanggal Publikasi/Presentasi : 24 April 2026

Dokumen yang diajukan dinyatakan :*)

Memenuhi syarat untuk diterima sebagai REKOGNISI PENGGANTI KEGIATAN SKRIPSI dan DINYATAKAN LULUS SKRIPSI.

Demikian Berita Acara ini dibuat untuk digunakan sebagaimana mestinya.

No	Nama	Tanda Tangan	
1	Radityo Adi Nugroho, S.T., M.Kom.	1. 	
2	As'ary Ramadhan, S.Kom., M.Cs.		2. 
3	Setyo Wahyu Saputro, S.Kom., M.Kom.	3. 	4. 
4	Mohammad Reza Faisal, S.Si., S.T., M.T., PhD.		4. 

Koordinator Program Studi Ilmu Komputer,

Panitia Skripsi,


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


Friska Abadi, S.Kom., M.Kom.
NIP. 198809132023211010

ABSTRAK

Peningkatan Prediksi Cacat Perangkat Lunak melalui Seleksi Fitur Hybrid Multi-Filter dan Penanganan Ketidakseimbangan Kelas

(Oleh: Muhammad Khalid Maulana; Pembimbing: Setyo Wahyu Saputro, S.Kom., M.Kom. dan Mohammad Reza Faisal, S.T., M.T., Ph.D.; 37 halaman)

Software Defect Prediction (SDP) bertujuan untuk mengidentifikasi modul yang cacat sejak awal siklus pengembangan perangkat lunak guna meningkatkan kualitas perangkat lunak dan mengurangi biaya pemeliharaan. Namun, dataset SDP umumnya mengalami dimensionalitas tinggi, redundansi fitur, dan ketidakseimbangan kelas, yang dapat menurunkan performa dan stabilitas model. Penelitian ini mengusulkan kerangka seleksi fitur hybrid untuk mengatasi tantangan tersebut dan meningkatkan performa prediksi. Pendekatan yang diusulkan mengintegrasikan Combined Correlation and Mutual Information (CONMI), yang menggabungkan Pearson Correlation Coefficient (PCC) dan Mutual Information (MI) untuk menangkap relevansi fitur linear maupun nonlinear. Fitur terpilih kemudian disempurnakan melalui seleksi Top-K, filtering berbasis korelasi untuk mengurangi multikolinearitas, dan Backward Elimination (BE) untuk memperoleh subset fitur optimal. Untuk mengatasi ketidakseimbangan kelas, SMOTE-Tomek diterapkan dengan menggabungkan teknik oversampling dan pembersihan data. Eksperimen dilakukan pada dua belas dataset NASA MDP menggunakan classifier Logistic Regression (LR) dan Naïve Bayes (NB). Hasil menunjukkan bahwa kerangka yang diusulkan secara konsisten mencapai performa terbaik, dengan Logistic Regression yang dikombinasikan dengan SMOTE-Tomek memperoleh rata-rata AUC tertinggi sebesar $0,7923 \pm 0,0714$, sedangkan NB memperoleh $0,7554 \pm 0,0580$. Analisis statistik menggunakan paired t-test menunjukkan bahwa metode yang diusulkan secara signifikan mengungguli MI+SMOTE-Tomek dan BE+SMOTE-Tomek untuk Logistic Regression, sedangkan tidak ditemukan perbedaan signifikan pada NB. Selain meningkatkan performa klasifikasi secara keseluruhan (AUC), pendekatan yang diusulkan juga meningkatkan deteksi kelas minoritas yang tercermin dari peningkatan Recall dan F1-score. Secara keseluruhan, kerangka hybrid yang diusulkan memberikan solusi yang efektif dan andal untuk prediksi cacat perangkat lunak, khususnya pada dataset berdimensi tinggi dan tidak seimbang.

Kata Kunci: *Backward Elimination; Class Imbalance; Feature Selection; Logistic Regression; Mutual Information; Naïve Bayes; Software Defect Prediction; SMOTE-Tomek*

ABSTRACT

Enhancing Software Defect Prediction through Hybrid Multi-Filter Feature Selection and Imbalance Handling

(By: Muhammad Khalid Maulana; Pembimbing: Setyo Wahyu Saputro, S.Kom., M.Kom. dan Mohammad Reza Faisal, S.T., M.T., Ph.D.; 37 pages)

Software Defect Prediction (SDP) aims to identify defective modules early in the software development lifecycle to improve software quality and reduce maintenance costs. However, SDP datasets commonly suffer from high dimensionality, feature redundancy, and class imbalance, which can degrade model performance and stability. This study proposes a hybrid feature selection framework to address these challenges and enhance prediction performance. The proposed approach integrates Combined Correlation and Mutual Information (CONMI), which combines the Pearson Correlation Coefficient (PCC) and Mutual Information (MI) to capture both linear and nonlinear feature relevance. The selected features are further refined through Top-K selection, correlation-based filtering to reduce multicollinearity, and Backward Elimination (BE) to obtain an optimal feature subset. To address class imbalance, SMOTE-Tomek is applied by combining over-sampling and data cleaning techniques. Experiments are conducted on twelve NASA MDP datasets using Logistic Regression (LR) and Naïve Bayes (NB) classifiers. The results show that the proposed framework consistently achieves the best performance, with Logistic Regression combined with SMOTE-Tomek obtaining the highest average AUC of 0.7923 ± 0.0714 , while NB achieves 0.7554 ± 0.0580 . Statistical analysis using a paired t-test indicates that the proposed method significantly outperforms MI+SMOTE-Tomek and BE+SMOTE-Tomek for Logistic Regression, whereas no significant differences are observed for NB. In addition to improving overall classification performance (AUC), the proposed approach also enhances minority class detection, as reflected in improved Recall and F1-score. Overall, the proposed hybrid framework provides an effective and reliable solution for software defect prediction, particularly for high-dimensional and imbalanced datasets.

Keywords: *Backward Elimination; Class Imbalance; Feature Selection; Logistic Regression; Mutual Information; Naïve Bayes; Software Defect Prediction; SMOTE-Tomek*