



**EFFECT OF SMOTE VARIANTS ON SOFTWARE DEFECT
PREDICTION CLASSIFICATION BASED ON BOOSTING
ALGORITHM**

Skripsi

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

Oleh

RAHMINA ULFAH AFLAHA

NIM 2011016220027

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

JUNI 2024



**EFFECT OF SMOTE VARIANTS ON SOFTWARE DEFECT
PREDICTION CLASSIFICATION BASED ON BOOSTING
ALGORITHM**

Skripsi

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

Oleh

RAHMINA ULFAH AFLAHA

NIM 2011016220027

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

JUNI 2024

SKRIPSI

**EFEK VARIAN SMOTE PADA KLASIFIKASI PREDIKSI CACAT SOFTWARE
BERBASIS ALGORITMA BOOSTING**

Oleh:

Rahmina Ulfah Aflaha

NIM. 2011016220027

Telah dipertahankan di depan Dosen Penguji pada tanggal 26 Juni 2024

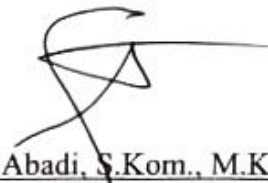
Susunan Dosen Penguji:

Dosen Pembimbing I



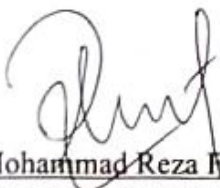
Rudy Hertono, S.Kom., M.Kom
NIP. 198809252022031003

Dosen Penguji I



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

Dosen Pembimbing II



Mohammad Reza Faisal, S.T., M.T., Ph.D
NIP. 197612202008121001

Dosen Penguji II



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



Banjarbaru, 26 Juni 2024
Kepada Program Studi Ilmu Komputer

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

KATA PENGANTAR


Puji syukur saya panjatkan pada Allah SWT karena berkat kasih, rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal yang berjudul “Effect of Smote Variants on Software Defect Prediction Classification Based on Boosting Algorithm” 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 terima kasih kepada semua pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan jurnal ini, adapun yang dimaksud adalah sebagai berikut:

1. Kedua orang tua dan 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 dan 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 jurnal ini.
3. Bapak Irwan Budiman, S. T., M. Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, serta seluruh Dosen dan Staff Prodi Ilmu Komputer yaitu Ka Azizah atas bantuan dan saran beliau jurnal ini dapat diselesaikan dengan baik.
4. Serta semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian jurnal 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, 20 Juni 2024

A handwritten signature in black ink, appearing to read 'M. Ulfah Aflaha', with a horizontal line underneath the name.

Rahmina Ulfah Aflaha

ABSTRAK

EFFECT OF SMOTE VARIANTS ON SOFTWARE DEFECT PREDICTION CLASSIFICATION BASED ON BOOSTING ALGORITHM

(Oleh : Rahmina Ulfah Aflaha; Pembimbing : Rudy Herteno, S.Kom., M.Kom.dan Mohammad Reza Faisal, S.T., M.T., Ph.D.; 2024; halaman)

Deteksi cacat perangkat lunak sejak dini sangat penting untuk menghindari kerugian finansial yang signifikan. Namun, membangun model prediksi kerusakan perangkat lunak yang akurat dapat menjadi tantangan karena ketidakseimbangan kelas, dimana data untuk modul yang rusak jauh lebih sedikit dibandingkan modul standar. Penelitian ini mengatasi masalah ini menggunakan kumpulan data MDP NASA yang tidak seimbang. Untuk mengatasi masalah ini, para peneliti telah mengusulkan metode baru yang menggabungkan pendekatan penyeimbangan level data dengan 14 variasi algoritma SMOTE untuk meningkatkan jumlah data modul yang rusak. Pendekatan tingkat algoritma dengan tiga algoritma peningkatan, Catboost, LightGBM, dan Gradient Boosting, diterapkan untuk mengklasifikasikan modul sebagai rusak atau tidak rusak. Metode ini bertujuan untuk meningkatkan akurasi prediksi kerusakan perangkat lunak. Hasilnya menunjukkan bahwa metode baru ini dapat menghasilkan klasifikasi yang lebih akurat dibandingkan penelitian sebelumnya. Pasangan DSMOTE dan Gradient Boosting dengan 0.9161 memiliki rata-rata akurasi tertinggi (0.9161). Model DSMOTE dan Catboost mencapai nilai rata-rata AUC tertinggi (0,9637). Kernel ADASYN dan Catboost menunjukkan kemampuan terbaik dalam melakukan nilai rata-rata G-mean (0,9154). Kontribusi penelitian terhadap prediksi kerusakan perangkat lunak melibatkan pengembangan teknik baru dan mengevaluasi efektivitasnya dalam mengatasi ketidakseimbangan kelas.

Kata kunci: *Software Defect Prediction, Imbalance, SMOTE Variants, Boosting*

ABSTRACT

EFFECT OF SMOTE VARIANTS ON SOFTWARE DEFECT PREDICTION CLASSIFICATION BASED ON BOOSTING ALGORITHM

(By : Rahmina Ulfah Aflaha; Pembimbing : Rudy Herteno, S.Kom., M.Kom.dan Mohammad Reza Faisal, S.T., M.T., Ph.D.; 2024; page)

Detecting software defects early on is critical for avoiding significant financial losses. However, building accurate software defect prediction models can be challenging due to class imbalance, where the data for defective modules is much less than for standard modules. This research addresses this issue using the imbalanced dataset NASA MDP. To address this issue, researchers have proposed new methods that combine data level balancing approaches with 14 variations of the SMOTE algorithm to increase the amount of defective module data. An algorithm-level approach with three boosting algorithms, Catboost, LightGBM, and Gradient Boosting, is applied to classify modules as defective or non-defective. These methods aim to improve the accuracy of software defect prediction. The results show that this new method can produce a more accurate classification than previous studies. The DSMOTE and Gradient Boosting pair with 0.9161 has the highest average accuracy (0.9161). The DSMOTE and Catboost model achieved the highest average AUC value (0.9637). The ADASYN kernel and Catboost showed the best ability to perform the average G-mean value (0.9154). The research contribution to software defect prediction involves developing new techniques and evaluating their effectiveness in addressing class imbalance.

Keywords: *Software Defect Prediction, Imbalance, SMOTE Variants, Boosting*

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam jurnal 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, 20 Juni 2024



Rahmina Ulfah Aflaha
NIM 2011016220027