



**COMPARATIVE STUDY OF VARIOUS HYPERPARAMETER  
TUNING ON RANDOM FOREST CLASSIFICATION WITH SMOTE  
AND FEATURE SELECTION USING GENETIC ALGORITHM IN  
SOFTWARE DEFECT PREDICTION**

**Skripsi**

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

**Oleh**

**MULIA KEVIN SURYADI**

**NIM. 2011016210019**

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

**MEI 2024**



**COMPARATIVE STUDY OF VARIOUS HYPERPARAMETER  
TUNING ON RANDOM FOREST CLASSIFICATION WITH SMOTE AND  
FEATURE SELECTION USING GENETIC ALGORITHM IN SOFTWARE  
DEFECT PREDICTION**

**Skripsi**

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

**Oleh**

**MULIA KEVIN SURYADI**

**NIM 2011016210019**

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

**MEI 2024**

**SKRIPSI**

**COMPARATIVE STUDY OF VARIOUS HYPERPARAMETER TUNING  
ON RANDOM FOREST CLASSIFICATION WITH SMOTE AND  
FEATURE SELECTION USING GENETIC ALGORITHM IN SOFTWARE  
DEFECT PREDICTION**

Oleh:

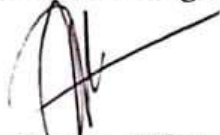
**MULIA KEVIN SURYADI**

**2011016210019**

Telah dipertahankan di depan Dosen Penguji pada Rabu, 23 Mei 2024

Susunan Dosen Penguji:

**Dosen Pembimbing I**



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

**Dosen Penguji I**



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

**Dosen Pembimbing II**



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

**Dosen Penguji II**



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



**Program Studi Ilmu Komputer**

Suliman, S.T., M.Kom  
NIP. 197803252008121001

## KATA PENGANTAR

Puji syukur saya panjatkan pada Allah SWT karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal yang berjudul “*A Comparative Study of Various Hyperparameter Tuning on Random Forest Classification With SMOTE and Feature Selection Using Genetic Algorithm in Software Defect Prediction*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

Penulis ingin mengucapkan ucapan terima kasih kepada semua pihak yang telah memberikan dukungan kepada penulis dalam pembuatan dan penyusunan jurnal ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga terutama kedua orang tua yang selalu memberikan doa dan dukungan dalam penyelesaian jurnal.
2. Bapak Rudy Herteno, S.Kom., M.Kom. selaku dosen pembimbing utama dan Bapak Setyo Wahyu Saputro, S.Kom., M.Kom. 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 dan Ibu Dwi Kartini, S.Kom., M.Kom. selaku ketua panitia skripsi Program Studi Ilmu Komputer FMIPA ULM.
4. Bapak Mohammad Reza Faisal, S.Si., S.T., M.T., PhD. Dan Radityo Adi Nugroho, S.T., M.Kom yang juga turut serta membantu dan meluangkan waktu dalam penyelesaian jurnal ini.
5. Seluruh Dosen dan Staff Prodi Ilmu Komputer yaitu Ka Azizah atas bantuan dan saran beliau jurnal ini dapat diselesaikan dengan baik.
6. Serta semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian jurnal ini.

Penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna. Meski demikian, diharapkan tulisan ini dapat memberikan manfaat bagi ilmu pengetahuan dan pembaca, serta mendapatkan berkah yang melimpah dari Tuhan Yang Maha Esa.

Banjarbaru, 29 Mei 2024

A handwritten signature in black ink, consisting of stylized, overlapping letters that appear to be 'MK' followed by a large, sweeping flourish.

Mulia Kevin Suryadi

## ABSTRAK

### **STUDI PERBANDINGAN BERBAGAI HYPERPARAMETER TUNING PADA KLASIFIKASI RANDOM FOREST DENGAN SMOTE DAN SELEKSI FITUR MENGGUNAKAN ALGORITMA GENETIKA DALAM PREDIKSI CACAT PERANGKAT LUNAK**

(Oleh : Mulia Kevin Suryadi; Pembimbing : Rudy Herteno, S. Kom., M.Kom. dan Setyo Wahyu Saputro, S. Kom., M. Kom.,; 2024; halaman)

Prediksi cacat perangkat lunak diperlukan untuk aplikasi desktop dan seluler. Performa prediksi cacat menggunakan Random Forest dapat meningkat secara signifikan dengan proses optimasi parameter dibandingkan dengan parameter default. Namun, langkah penyetelan parameter sering kali diabaikan. Random Forest memiliki banyak parameter yang dapat diatur, sehingga penyesuaian parameter secara manual akan mengurangi efisiensi Random Forest, menghasilkan hasil yang tidak optimal, dan memakan banyak waktu. Penelitian ini bertujuan untuk meningkatkan performa klasifikasi Random Forest dengan menggunakan SMOTE untuk menyeimbangkan data, Algoritma Genetika sebagai seleksi fitur, dan menggunakan penyetelan hiperparameter untuk mengoptimalkan performa. Selain itu, penelitian ini juga bertujuan untuk mengetahui metode penyetelan hiperparameter mana yang menghasilkan peningkatan terbaik pada metode klasifikasi Random Forest. Dataset yang digunakan dalam penelitian ini adalah NASA MDP yang mencakup 13 dataset. Metode yang digunakan meliputi SMOTE untuk menangani data yang tidak seimbang, seleksi fitur dengan Algoritma Genetika, klasifikasi Random Forest, dan metode penyetelan hiperparameter termasuk Grid Search, Random Search, Optuna, Bayesian (dengan Hyperopt), Hyperband, TPE, dan Nevergrad. Hasil penelitian ini dievaluasi dengan menggunakan nilai akurasi dan AUC. Dalam hal peningkatan akurasi, tiga metode terbaik adalah Nevergrad, TPE, dan Hyperband. Dalam hal peningkatan AUC, tiga metode terbaik adalah Hyperband, Optuna, dan Random Search. Nevergrad rata-rata meningkatkan akurasi sekitar 3.9% dan Hyperband rata-rata meningkatkan AUC sekitar 3.51%. Penelitian ini menunjukkan bahwa penggunaan penyetelan hiperparameter meningkatkan performa Random Forest dan di antara semua metode penyetelan hiperparameter yang digunakan, Hyperband memiliki performa penyetelan hiperparameter terbaik dengan peningkatan rata-rata tertinggi dalam hal akurasi dan AUC. Implikasi dari penelitian ini adalah meningkatkan penggunaan penyetelan hiperparameter dalam prediksi cacat perangkat lunak dan meningkatkan performa prediksi cacat perangkat lunak.

**Kata Kunci** : *Genetic Algorithm, Hyperparameter Tuning, Random Forest, Software Defect Prediction*

## ABSTRACT

### **A COMPARATIVE STUDY OF VARIOUS HYPERPARAMETER TUNING ON RANDOM FOREST CLASSIFICATION WITH SMOTE AND FEATURE SELECTION USING GENETIC ALGORITHM IN SOFTWARE DEFECT PREDICTION**

(Oleh : Mulia Kevin Suryadi; Pembimbing : Rudy Herteno, S. Kom., M.Kom. dan Setyo Wahyu Saputro, S. Kom., M. Kom.,; 2024; halaman)

Software defect prediction is necessary for desktop and mobile applications. Random Forest defect prediction performance can be significantly increased with the parameter optimization process compared to the default parameter. However, the parameter tuning step is commonly neglected. Random Forest has numerous parameters that can be tuned, as a result manually adjusting parameters would diminish the efficiency of Random Forest, yield suboptimal results and it will take a lot of time. This research aims to improve the performance of Random Forest classification by using SMOTE to balance the data, Genetic Algorithm as selection feature, and using hyperparameter tuning to optimize the performance. Apart from that, it is also to find out which hyperparameter tuning method produces the best improvement on the Random Forest classification method. The dataset used in this study is NASA MDP which included 13 datasets. The method used contains SMOTE to handle imbalance data, Genetic Algorithm feature selection, Random Forest classification, and hyperparameter tuning methods including Grid Search, Random Search, Optuna, Bayesian (with Hyperopt), Hyperband, TPE and Nevergrad. The results of this research were carried out by evaluating performance using accuracy and AUC values. In terms of accuracy improvement, the three best methods are Nevergrad, TPE, and Hyperband. In terms of AUC improvement, the three best methods are Hyperband, Optuna, and Random Search. Nevergrad on average improves accuracy by about 3.9% and Hyperband on average improves AUC by about 3.51%. This study indicates that the use of hyperparameter tuning improves Random Forest performance and among all the hyperparameter tuning methods used, Hyperband has the best hyperparameter tuning performance with the highest average increase in both accuracy and AUC. The implication of this research is to increase the use of hyperparameter tuning in software defect prediction and improve software defect prediction performance.

**Kata Kunci** : Genetic Algorithm, Hyperparameter Tuning, Random Forest, Software Defect Prediction

## **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 sepengetahuan 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, 29 Mei 2024



Mulia Kevin Suryadi

NIM. 2011016210019



## DAFTAR ISI

I.	INTRODUCTION .....	1
II.	MATERIAL AND METHODS .....	2
	A. Dataset.....	2
	B. SMOTE.....	3
	C. Genetic Algorithm Feature Selection.....	3
	D. Grid Search Hyperparameter Tuning .....	3
	E. Random Search Hyperparameter Tuning.....	3
	F. Optuna Hyperparameter Tuning .....	3
	G. Bayesian Search (With Hyperopt) Hyperparameter Tuning.....	4
	H. Hyperband Hyperparameter Tuning .....	4
	I. Tree Parzen Estimator Hyperparameter Tuning .....	4
	J. Nevergrad Hyperparameter Tuning .....	4
	K. Random Forest Classifier.....	4
III.	RESULT.....	5
	A. SMOTE Process .....	6
	B. Genetic Algorithm Process .....	6
	C. Hyperparameter Tuning Process.....	6
	D. Performance of Random Forest Classifier .....	6
IV.	DISCUSSION .....	8
V.	CONCLUSION.....	9
VI.	ACKNOWLEDGEMENT .....	9
VII.	REFERENCES.....	9
VIII.	BIBLIOGRAPHY.....	10