

**ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK
ANSAMBEL PADA PREDIKSI CACAT PERANGKAT LUNAK**

SKRIPSI

diajukan untuk memenuhi sebagian syarat untuk memperoleh gelar Sarjana
Komputer pada Program Studi S1 Rekayasa Perangkat Lunak



oleh
Ravena
NIM 1909185

**PROGRAM STUDI REKAYASA PERANGKAT LUNAK
KAMPUS UPI DI CIBIRU
UNIVERSITAS PENDIDIKAN INDONESIA
2023**

**ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK
ANSAMBEL PADA PREDIKSI CACAT PERANGKAT LUNAK**

Oleh
Ravena
NIM 1909185

Sebuah Skripsi yang Diajukan untuk Memenuhi Salah Satu Syarat Memperoleh
Gelar Sarjana Komputer pada Program Studi S1 Rekayasa Perangkat Lunak

© Ravena
Universitas Pendidikan Indonesia
2023

© Hak Cipta Dilindungi Undang-Undang

Skripsi ini tidak boleh diperbanyak seluruhnya atau sebagian, dengan dicetak
ulang, difoto kopi, atau cara lainnya tanpa izin dari penulis

HALAMAN PENGESAHAN

RAVENA

ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK
ANSAMBEL PADA PREDIKSI CACAT PERANGKAT LUNAK

disetujui dan disahkan oleh pembimbing:

Pembimbing I



Dian Anggraini, S.ST., M.T.

NIP. 920190219930526201

Pembimbing II



Mochamad Iqbal Ardimansyah, S.T., M.Kom.

NIP. 920190219910328101

Mengetahui,

Ketua Program Studi Rekayasa Perangkat Lunak



Mochamad Iqbal Ardimansyah, S.T., M.Kom.

NIP. 920190219910328101

Ravena, 2023

**ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK ANSAMBEL PADA PREDIKSI
CACAT PERANGKAT LUNAK**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK ANSAMBEL PADA PREDIKSI CACAT PERANGKAT LUNAK

Ravena

ABSTRAK

Prediksi cacat perangkat lunak membutuhkan dataset yang digunakan untuk mengevaluasi model. Namun dataset yang digunakan seringkali terdapat distribusi kelas yang tidak seimbang sehingga diperlukan metode yang dapat menangani ketidakseimbangan kelas tersebut. Penelitian ini bertujuan untuk mengetahui performa dari model prediksi serta mengetahui model prediksi yang lebih efektif dalam menangani ketidakseimbangan kelas. Pada penelitian ini dilakukan pengembangan model berbasis teknik ansambel seperti Random Forest, Extra Trees, Adaboost, dan Gradient Boosting dengan menerapkan metode *sampling* seperti RUS (*Random Under Sampling*) dan SMOTE (*Syntetic Minority Over-sampling Technique*) yang berguna untuk menyeimbangkan kelas yang ada. Selain itu, performa model dapat diketahui dengan melakukan evaluasi performa seperti *accuracy*, *balanced accuracy*, *gmean*, dan AUC yang dinilai mampu dalam menangani ketidakseimbangan kelas. Hasil penelitian ditemukan bahwa performa keempat model berada pada rentang 65% hingga 93% dari keempat metrik tersebut dimana nilai terbaik terdapat pada model prediksi extra trees dengan rata-rata nilai hampir 93%, disusul dengan random forest dengan nilai 90 hingga 91%, serta Adaboost dan Gradient Boosting yang mendapat nilai dalam rentang 84 hingga 89%. Adapun berdasarkan hasil performa model prediksi tersebut, model prediksi yang lebih efektif dalam menangani ketidakseimbangan kelas terdapat pada metode *bagging* yaitu Extra Trees yang ditunjukkan dari tingginya performa pada 9 dari 10 dataset yang ada dibandingkan dengan model prediksi lain disusul dengan Random Forest. Walaupun metode *boosting* memiliki performa yang tidak setinggi metode *bagging*, tapi performa yang diperoleh lebih baik dibandingkan penelitian sebelumnya.

Kata kunci: Prediksi Kecacatan Perangkat Lunak, Ketidakseimbangan Kelas, Teknik Ansambel, Metode Sampling

Ravena, 2023

ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK ANSAMBEL PADA PREDIKSI CACAT PERANGKAT LUNAK

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

CLASS IMBALANCE ANALYSIS USING ENSEMBLE TECHNIQUES IN SOFTWARE DEFECT PREDICTION

Ravena

ABSTRACT

Software defect prediction requires a dataset used to evaluate the model. However, the dataset used is often an unbalanced class distribution, so there is a need for a method that can handle such class imbalances. This study aims to find out the performance of predictive models as well as to find more effective predictive models in handling class imbalances. In this study, the development of ensemble technique-based models such as Random Forest, Extra Trees, Adaboost, and Gradient Boosting by applying sampling methods such as RUS (Random Under Sampling) and SMOTE (Synthetic Minority Over-sampling Technique) which are useful for balancing existing classes. In addition, model performance can be identified by performance evaluations such as accuracy, balanced accuracy, gmean, and auc that are considered capable of handling class imbalances. The results of the study were found that the performance of the four models was in the range of 65% to 93% of the four metrics, where the best value was in the extra trees prediction model with an average value of almost 93%, followed by random forests with a value of 90 to 91%, There are Adaboost and Gradient Boosting that are rated in the range of 84 to 89%. Based on the performance results of the prediction model, a more effective prediction model in handling class imbalances is found in the bagging method, which is the Extra Trees shown from the high performance in 9 out of 10 existing datasets compared to other prediction models followed by Random Forest. Although the boost method has a performance that is not as high as the bagging method, the performance obtained is better than previous studies.

Keywords: *Software Defect Prediction, Class Imbalance, Ensemble Technique, Sampling Method*

Ravena, 2023

ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK ANSAMBEL PADA PREDIKSI CACAT PERANGKAT LUNAK

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

DAFTAR ISI

KATA PENGANTAR.....	Error! Bookmark not defined.
ABSTRAK	iii
<i>ABSTRACT</i>	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	Error! Bookmark not defined.
DAFTAR GAMBAR	Error! Bookmark not defined.
DAFTAR LAMPIRAN	Error! Bookmark not defined.
BAB I PENDAHULUAN	Error! Bookmark not defined.
1.1 Latar Belakang	Error! Bookmark not defined.
1.2 Rumusan Masalah	Error! Bookmark not defined.
1.3 Tujuan Penelitian.....	Error! Bookmark not defined.
1.4 Manfaat Penelitian.....	Error! Bookmark not defined.
1.5 Batasan Masalah.....	Error! Bookmark not defined.
1.6 Struktur Organisasi Skripsi.....	Error! Bookmark not defined.
BAB II KAJIAN PUSTAKA	Error! Bookmark not defined.
2.1 Prediksi Cacat Perangkat Lunak.....	Error! Bookmark not defined.
2.2 Dataset	Error! Bookmark not defined.
2.3 Teknik Ansambel.....	Error! Bookmark not defined.
2.4 Machine Learning.....	Error! Bookmark not defined.
2.5 Metode Sampling.....	Error! Bookmark not defined.
2.6 Evaluasi Performa	Error! Bookmark not defined.
2.7 Penelitian Terkait	Error! Bookmark not defined.
BAB III METODE PENELITIAN.....	Error! Bookmark not defined.
3.1 Desain Penelitian	Error! Bookmark not defined.
3.2 Sumber Himpunan Data	Error! Bookmark not defined.
3.3 Instrumen Penelitian	Error! Bookmark not defined.
BAB IV HASIL DAN PEMBAHASAN.....	Error! Bookmark not defined.
4.1 Persiapan Data	Error! Bookmark not defined.
4.2 Proses Resampling	Error! Bookmark not defined.

Ravena, 2023

*ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK ANSAMBEL PADA PREDIKSI
CACAT PERANGKAT LUNAK*

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

4.3	Pembuatan Model Prediksi.....	Error! Bookmark not defined.
4.4	Evaluasi Model.....	Error! Bookmark not defined.
BAB V SIMPULAN, IMPLIKASI, DAN REKOMENDASI ...		Error! Bookmark not defined.
5.1	Simpulan.....	Error! Bookmark not defined.
5.2	Implikasi dan Rekomendasi	Error! Bookmark not defined.
DAFTAR PUSTAKA		Error! Bookmark not defined.
LAMPIRAN		Error! Bookmark not defined.

DAFTAR PUSTAKA

- Aljamaan, Hamoud, dan Amal Alazba. 2020. "Software Defect Prediction Using Tree-Based Ensembles." *PROMISE 2020 - Proceedings of the 16th ACM International Conference on Predictive Models and Data Analytics in Software Engineering, Co-located with ESEC/FSE 2020* 1–10.
- Balogun, Abdullateef O., Shuib Basri, Said J. Abdulkadir, Victor E. Adeyemo, Abdullahi A. Imam, dan Amos O. Bajeh. 2019. "Software Defect Prediction: Analysis Of Class Imbalance And Performance Stability." *Journal of Engineering Science and Technology* 14(6):3294–3308.
- Breiman, Leo. 1996. "Bagging Predictors." *Machine Learning 1996* 24:2 24(2):123–40.
- Gao, Xiao Yan, Abdelmegeid Amin Ali, Hassan Shaban Hassan, dan Eman M. Anwar. 2021. "Improving The Accuracy For Analyzing Heart Diseases Prediction Based On The Ensemble Method." *Complexity* 2021.
- Goel, Lipika, Mayank Sharma, Sunil Kumar Khatri, dan D. Damodaran. 2020. "Defect Prediction of Cross Projects Using PCA and Ensemble Learning Approach." *Lecture Notes in Networks and Systems* 106:307–15.
- Gorunescu, Florin. 2011. "Data Mining: Concepts, Models And Techniques." *Intelligent Systems Reference Library* 12.
- Karo, Ichwanul Muslim Karo, dan Hendriyana Hendriyana. 2022. "Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score." *Jurnal Teknologi Terpadu* 8(2):94–99.
- Leevy, Joffrey L., Taghi M. Khoshgoftaar, Richard A. Bauder, dan Naeem Seliya. 2018. "A Survey On Addressing High-Class Imbalance In Big Data." *Journal of Big Data* 5(1):1–30.
- Li, Zhiqiang, Xiao Yuan Jing, dan Xiaoke Zhu. 2018. "Progress On Approaches To Software Defect Prediction." *IET Software* 12(3):161–75.

- Liu, Shanhong. 2019. "QA And Testing Budget Allocation 2012-2018 | Statista." *Statista*. Diambil 28 September 2022 (<https://www.statista.com/statistics/500641/worldwide-qa-budget-allocation-as-%-it-spend/>).
- Malhotra, Ruchika, dan Juhi Jain. 2020. "Handling Imbalanced Data Using Ensemble Learning In Software Defect Prediction." *Proceedings of the Confluence 2020 - 10th International Conference on Cloud Computing, Data Science and Engineering* 300–304.
- Matloob, Faseeha, Taher M. Ghazal, Nasser Taleb, Shabib Aftab, Munir Ahmad, Muhammad Adnan Khan, Sagheer Abbas, dan Tariq Rahim Soomro. 2021a. "Software Defect Prediction Using Ensemble Learning: A Systematic Literature Review." *IEEE Access* 9:98754–71.
- Natekin, Alexey, dan Alois Knoll. 2013. "Gradient Boosting Machines, A Tutorial." *Frontiers in Neurobotics* 7(DEC):21.
- Pandey, Sushant Kumar, Ravi Bhushan Mishra, dan Anil Kumar Tripathi. 2021. "Machine Learning Based Methods for Software Fault Prediction: A Survey." *Expert Systems with Applications* 172:114595.
- Reine De Reanzi, S., dan P. Ranjit Jeba Thangaiah. 2021. "A Survey On Software Test Automation Return On Investment, In Organizations Predominantly From Bengaluru, India:"
- Rymarczyk, Jan. 2020. "Technologies, Opportunities and Challenges Of The Industrial Revolution 4.0: Theoretical Considerations." *Entrepreneurial Business and Economics Review* 8(1):185–98.
- Speiser, Jaime Lynn, Michael E. Miller, Janet Tooze, dan Edward Ip. 2019. "A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling." *Expert Systems with Applications* 134:93–101.