PERBANDINGAN *SMOTE-VARIANTS* UNTUK MENGATASI KETIDAKSEIMBANGAN DATA PADA PREDIKSI CACAT *SOFTWARE*

Skripsi

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh

MUIZA RAHMAN

NIM. 1911016310018

PROGRAM STUDI S1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU

NOVEMBER 2023

PERBANDINGAN *SMOTE-VARIANTS* UNTUK MENGATASI KETIDAKSEIMBANGAN DATA PADA PREDIKSI CACAT *SOFTWARE*

Skripsi

Untuk Memenuhi Persyaratan

Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh

MUIZA RAHMAN

NIM. 1911016310018

PROGRAM STUDI S1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU

NOVEMBER 2023

SKRIPSI

PERBANDINGAN *SMOTE-VARIANTS* UNTUK MENGATASI KETIDAKSEIMBANGAN DATA PADA PREDIKSI CACAT *SOFTWARE*

Oleh :

MUIZA RAHMAN

1911016310018

Telah dipertahankan di depan Dosen Penguji pada tanggal 27 September 2023

Susunan Dosen Penguji :

Pembimbing I

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

Dosen Penguji I

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

Pembimbing II

Triando H. Saragih, S.Kom., M.Kom.
NIP. 199308242019031012

Dosen Penguji II

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

Banjarbaru, 15 November 2023
Ketua Program Studi Ilmu Komputer

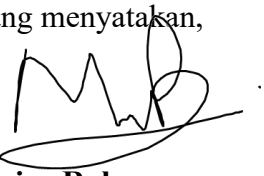
Irwan Budiman S.T., M.Kom
NIP. 197703252008121001

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, 15 November 2023

Yang menyatakan,



Muiza Rahman

NIM. 1911016310018

ABSTRAK

PERBANDINGAN *SMOTE-VARIANTS* UNTUK MENGATASI KETIDAKSEIMBANGAN DATA PADA PREDIKSI CACAT *SOFTWARE*

(Oleh : Muiza Rahman; Pembimbing: Mohammad Reza Faisal, S.T., M.T. Ph.D dan Triando Hamonangan Saragih, S.Kom., M.Kom; 2023; 46 halaman)

Ketidakseimbangan data merupakan salah satu masalah penelitian prediksi cacat *software*. Ketidakseimbangan data dapat membuat model bias kepada kelas mayoritas, sementara kebanyakan dataset untuk prediksi cacat *software*, kelas cacat yang ingin diprediksi merupakan kelas minoritas. Metode umum yang digunakan untuk mengatasi ini adalah pendekatan *level* data seperti *SMOTE-Variants* ataupun pendekatan *level* algoritma seperti *Bagging*. Penelitian ini menggunakan pendekatan *level* data *SMOTE-Variants* yaitu *SMOTE, Borderline-SMOTE, SMOTE-ENN* dan *SMOTE-Tomek-Links.* Sementara, pendekatan level algoritma yang digunakan adalah *Bagging* (*Decision Tree*) atau *Bagging* dengan *base classifier Decision Tree*. Hasilnya model dengan *SMOTE-Tomek-Links* dan *Bagging* (*Decision Tree*) mendapatkan skor rata-rata *AUC* terbaik sebesar 0,8009. Sementara, *SMOTE-ENN* dan *Bagging* (*Decision Tree*) mendapatkan skor rata-rata *recall* terbaik sebesar 0,6480.

Kata kunci: Ketidakseimbangan data, Prediksi cacat *software*, *Bagging*, *SMOTE-Variants*

ABSTRACT

COMPARISON OF SMOTE-VARIANTS TO OVERCOME DATA IMBALANCE IN SOFTWARE DEFECT PREDICTION

(by : Muiza Rahman; *Supervisor:* Mohammad Reza Faisal, S.T., M.T. Ph.D and Triando Hamonangan Saragih, S.Kom., M.Kom; 2023; 46 pages)

Data imbalance is one of the problems of software defect prediction research. Data imbalance can make the model biased towards the majority class, while in most of the datasets for software defect prediction, the defect class to be predicted is the minority class. Common methods used to address this are data-level approach such as SMOTE-Variants or classification-level approach such as Bagging. This research uses the SMOTE-Variants data-level approach, namely SMOTE, Borderline-SMOTE, SMOTE-ENN and SMOTE-Tomek-Links. Meanwhile, the algorithm level approach used is Bagging (Decision Tree) or Bagging with Decision Tree base classifier. As a result, the model with SMOTE-Tomek-Links and Bagging (Decision Tree) gets the best average AUC score of 0.8009. Meanwhile, SMOTE-ENN and Bagging (Decision Tree) get the best average recall score of 0.6480.

Keywords*: Data imbalance, Software Defect Prediction, Bagging, SMOTE-Variants*

PRAKATA

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **PERBANDINGAN *SMOTE-VARIANTS* UNTUK MENGATASI KETIDAKSEIMBANGAN DATA PADA PREDIKSI CACAT *SOFTWARE*** untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

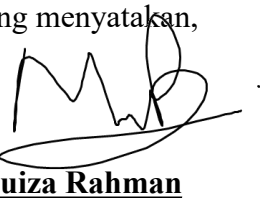
Pada lembar ini penulis ingin menyampaikan ucapan terima kasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut:

1. Keluarga yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian skripsi ini.
2. Bapak Mohammad Reza Faisal, S.T., M.T. Ph.D selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
3. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman S.T., M.Kom selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer angkatan 2019 yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan skripsi.
7. 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 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, 15 November 2023

Yang menyatakan,



Muiza Rahman

NIM. 1911016310018

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN