

BAB I

PENDAHULUAN

1.1 Latar Belakang

Revolusi industri 4.0 saat ini mengakibatkan teknologi khususnya perangkat lunak yang dikembangkan semakin kompleks sesuai dengan kebutuhan pengguna. Adapun semakin kompleks suatu perangkat lunak maka semakin besar pula kemungkinan perangkat lunak tersebut memiliki kecacatan (Rymarczyk, 2020). Berdasarkan hal tersebut maka diperlukan suatu cara untuk dapat meminimalisir kecacatan pada perangkat lunak salah satunya dengan meningkatkan kualitas perangkat lunak.

Kualitas perangkat lunak dapat diimplementasikan dalam berbagai cara salah satunya dengan melakukan pengujian. Namun pengujian perangkat lunak ini tentu membutuhkan biaya yang cukup besar sehingga mungkin akan menjadi hambatan baru bagi pengembang yang memiliki budget terbatas. Menurut situs Statista (Liu, 2019), rata-rata pengeluaran pengujian dalam 5 tahun terakhir dari 2012 sampai dengan 2019 adalah 28% dari keseluruhan pengeluaran IT. Selain itu, pada 2019, panel CIO dan profesional teknologi senior juga telah melaporkan bahwa rata-rata sekitar 23% anggaran IT tahunan organisasi mereka yang dialokasikan untuk jaminan kualitas dan pengujian (Reine De Reanzi dan Ranjit Jeba Thangaiyah, 2021). Dengan adanya permasalahan dan data terkait kualitas dan pengujian tersebut, dibutuhkan alternatif lain yang dapat digunakan untuk mengetahui kualitas dari perangkat lunak. Alternatif yang dapat digunakan yaitu penerapan prediksi cacat perangkat lunak.

Prediksi kecacatan perangkat lunak adalah salah satu penentuan kualitas perangkat lunak dengan memprediksi kemungkinan kecacatan yang ada pada suatu perangkat lunak yang sedang dikembangkan. Prediksi cacat akan menghasilkan sejumlah daftar sumber kode yang rawan cacat sehingga tim penjaminan kualitas dalam mengalokasikan sumber daya secara efektif untuk memvalidasi perangkat lunak dengan menempatkan upaya pada sumber kode yang rawan cacat (Reine De Reanzi dan Ranjit Jeba Thangaiyah, 2021). Untuk menerapkan prediksi cacat

tersebut, dibutuhkan algoritma *machine learning* yang digunakan sebagai model prediksi. Model prediksi kecacatan perangkat lunak yang dapat diimplementasikan bermacam-macam seperti teknik ansambel, *bayesian learners*, *decision tree*, *neural networks*, dan metode lainnya.

Proses prediksi kecacatan perangkat lunak juga membutuhkan sebuah dataset yang berfungsi untuk mengevaluasi model prediksi yang telah dibuat. Namun, umumnya dalam suatu dataset terdapat distribusi kelas yang tidak merata sehingga performa yang dihasilkan menjadi kurang maksimal (Leevy dkk, 2018). Beberapa peneliti mengungkapkan bahwa metode *sampling* dapat membantu dalam mengatasi masalah ketidakseimbangan kelas. Metode *sampling* yang dapat diterapkan yaitu RUS (*Random Under Sampling*) dan SMOTE (*Syntetic Minority Over- sampling Technique*) (Matloob dkk, 2021). Selain itu, peneliti lain juga menyarankan model prediksi berbasis teknik ansambel seperti algoritma *random forest*, *extra trees*, *adaboost*, dan *gradient boosting* dalam menyelesaikan masalah tersebut (Pandey, Mishra, dan Tripathi, 2021).

Pada penelitian ini, diusulkan pembuatan model prediksi metode *sampling* seperti RUS dan SMOTE dan algoritma berbasis teknik ansambel yaitu *random forest*, *extra trees*, *adaboost*, dan *gradient boosting*. Hal tersebut berdasarkan penelitian-penelitian sebelumnya yang menunjukkan bahwa penggunaan RUS dan SMOTE dan algoritma *random forest*, *extra trees*, *adaboost*, dan *gradient boosting* ini dapat membantu menangani masalah ketidakseimbangan kelas pada prediksi cacat perangkat lunak.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, dapat dirumuskan permasalahan yang akan dikaji lebih lanjut, antara lain:

1. Bagaimana performa model berbasis teknik ansambel dalam menangani masalah ketidakseimbangan kelas?
2. Model prediksi berbasis teknik ansambel mana yang lebih efektif dalam menangani masalah ketidakseimbangan kelas?

1.3 Tujuan Penelitian

Adapun tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui performa model berbasis teknik ansambel dalam menangani masalah ketidakseimbangan kelas.
2. Menentukan model prediksi berbasis teknik ansambel mana yang lebih efektif dalam menangani masalah ketidakseimbangan kelas.

1.4 Manfaat Penelitian

Manfaat yang diharapkan dari adanya penelitian ini antara lain:

1. Manfaat Teoritis

Penelitian ini dapat dijadikan sebagai bahan informasi untuk menambah pengetahuan dalam penanganan masalah pada ketidakseimbangan kelas dengan teknik ansambel.

2. Manfaat Praktis

- a. Mengetahui tahapan dalam penanganan masalah ketidakseimbangan kelas menggunakan teknik ansambel pada prediksi cacat perangkat lunak.
- b. Mengetahui prediksi, metode *sampling*, serta evaluasi yang dapat digunakan dalam penanganan masalah ketidakseimbangan kelas.

1.5 Batasan Masalah

Penelitian berfokus pada dalam penanganan ketidakseimbangan kelas sebagai masalah kualitas data.

1.6 Struktur Organisasi Skripsi

Sistematika penulisan yang diterapkan pada penelitian ini secara umum mengacu pada Pedoman Penulisan Karya Ilmiah Universitas Pendidikan Indonesia Tahun 2019. Berikut merupakan sistematika penulisan tersebut:

1. Bab I: Pendahuluan

Pada Bab I Pendahuluan ini berisi latar belakang penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan struktur organisasi skripsi. Latar belakang masalah membahas terkait dengan alasan penelitian ini bisa dibuat. Rumusan masalah membahas mengenai masalah apa saja yang akan dipaparkan berdasarkan latar belakang yang diambil yang dikemas dalam bentuk pertanyaan. Tujuan penelitian membahas mengenai tujuan dari penelitian yang

Ravena, 2023

ANALISIS KETIDAKSEIMBANGAN KELAS MENGGUNAKAN TEKNIK ANSAMBEL PADA PREDIKSI CACAT PERANGKAT LUNAK

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

diambil berdasarkan rumusan masalah yang dibuat. Manfaat penelitian membahas tentang kebermanfaatan apa saja yang bisa ditemukan dalam penelitian ini. Batasan masalah membahas terkait cakupan atau ruang lingkup yang akan dibahas pada penelitian ini. Struktur organisasi skripsi membahas mengenai sistematika yang dijadikan acuan dalam penelitian ini.

2 Bab II: Kajian Pustaka

Pada Bab II Kajian Pustaka membahas mengenai hasil tinjauan yang telah dilakukan yang berkaitan dengan penelitian ini. Hasil tinjauan berupa teori, konsep, hingga metode yang dilakukan oleh para peneliti sebelumnya.

3 Bab III: Metode Penelitian

Pada Bab III Metode Penelitian membahas mengenai desain penelitian, sumber himpunan data, serta instrumen penelitian. Desain penelitian membahas tentang desain penelitian apa yang digunakan dalam penelitian ini. Sumber himpunan data membahas mengenai himpunan data apa yang digunakan serta sumber himpunan data mana yang akan digunakan dalam penelitian ini. Instrumen penelitian membahas terkait alat ukur apa yang digunakan untuk mengumpulkan data.

4 Bab IV: Hasil dan Pembahasan

Pada Bab IV Hasil dan Pembahasan membahas terkait persiapan data, implementasi tahapan penelitian, dan analisis hasil penelitian. Persiapan data membahas mengenai pengolahan himpunan data yang nantinya akan digunakan dalam tahapan penelitian. Implementasi tahapan penelitian membahas tentang langkah-langkah penelitian mulai dari melakukan *sampling* hingga menunjukkan performa. Analisis hasil penelitian membahas tentang analisa hasil performa dari implementasi tahapan penelitian.

5 Bab V: Simpulan, Implikasi, dan Rekomendasi

Pada Bab V Simpulan, Implikasi, dan Rekomendasi membahas tentang kesimpulan yang didapatkan dari penelitian yang telah dilakukan yang dapat menjawab rumusan masalah yang sudah dibuat pada bab sebelumnya. Implikasi dan Rekomendasi membahas mengenai kebaharuan atau rekomendasi yang dapat diterapkan pada penelitian selanjutnya.