

BAB V

PENUTUP

5.1 Kesimpulan

Bug Tracking System (BTS) merupakan suatu perangkat lunak yang digunakan dalam tahap pemeliharaan perangkat lunak dan berperan untuk menyimpan riwayat dan melacak laporan terkait permintaan terhadap perubahan, perbaikan kecacatan dan kegagalan, dan dukungan teknis dalam siklus hidup pengembangan perangkat lunak. Kategori dan prioritas laporan dalam BTS dapat ditetapkan secara otomatis menggunakan model pembelajaran mesin. Model yang dikembangkan dalam penelitian ini menggunakan algoritma *Logistic Regression*. Berdasarkan hasil yang diperoleh, representasi fitur memiliki peran yang penting dalam meningkatkan performa model ketika menghadapi permasalahan ketidakseimbangan kelas. Fitur-fitur teks secara kontekstual, yakni menggunakan TF-CHI dan Word2Vec dengan CBOW, dapat meningkatkan performa berdasarkan nilai *precision*, *recall*, *F-score*, dan *accuracy*. Namun, representasi fitur tersebut tidak dapat meningkatkan performa ketika sebagian besar kata-kata memiliki korelasi yang tinggi terhadap hampir atau seluruh label kelas. Hal tersebut dapat menyebabkan anomali karena dapat membentuk suatu kluster yang terdiri atas sebagian besar atau seluruh label kelas.

Pada repositori Mindville, performa model dalam memprediksi kategori laporan mengalami peningkatan pada nilai *precision* sebesar 31% (dari 50% menjadi 81%), *recall* sebesar 25% (dari 50% menjadi 75%), *F-score* sebesar 31% (dari 45% menjadi 76%) dan *accuracy* sebesar 21% (dari 57% menjadi 78%). Namun, performa model dalam memprediksi prioritas laporan tidak mampu ditingkatkan karena terjadi anomali yang telah dipaparkan. Hal tersebut menyebabkan TF-CHI tidak efektif untuk digunakan sehingga TF-IDF dengan teknik SMOTE dan variasinya dapat menjadi solusi alternatif.

Pada repositori JFrog, performa model dalam memprediksi kategori laporan mengalami peningkatan pada *precision* sebesar 48% (dari 14% menjadi 62%),

recall sebesar 29% (dari 15% menjadi 44%), *F-score* sebesar 35% (dari 11% menjadi 46%) dan *accuracy* sebesar 16% (dari 55% menjadi 71%). Hal tersebut juga berlaku ketika memprediksi prioritas laporan, yakni meningkatnya nilai *precision* sebesar 22% (dari 18% menjadi 40%), *recall* sebesar 20% (dari 25% menjadi 40%), dan *F-score* sebesar 20% (dari 21% menjadi 41%). Walaupun mengalami penurunan akurasi sebesar 8% (dari 71% menjadi 63%), tetapi model tersebut lebih bermakna karena dapat mengidentifikasi seluruh label kelas.

Pada repositori Hyperledger, performa model dalam memprediksi kategori laporan mengalami peningkatan pada *precision* sebesar 51% (dari 9% menjadi 60%), *recall* sebesar 27% (dari 11% menjadi 38%), *F-score* sebesar 32% (dari 9% menjadi 41%) dan *accuracy* sebesar 18% (dari 25% menjadi 43%). Hal tersebut juga berlaku ketika memprediksi prioritas laporan, yakni meningkatnya nilai *precision* sebesar 24% (dari 16% menjadi 40%), *recall* sebesar 4% (dari 20% menjadi 44%), dan *F-score* sebesar 8% (dari 18% menjadi 26%). Walaupun mengalami penurunan akurasi sebesar 6% (dari 80% menjadi 74%), tetapi model tersebut lebih bermakna karena dapat mengidentifikasi seluruh label kelas.

Berdasarkan hasil yang telah diperoleh, dapat disimpulkan bahwa representasi fitur yang tepat perlu disesuaikan dengan memperhatikan karakteristik sumber himpunan data yang ada. Performa suatu model dapat meningkat dengan TF-CHI jika tidak mengalami anomali yang telah dipaparkan. Representasi TF-IDF dengan memanfaatkan teknik SMOTE dan variasinya dapat menjadi solusi alternatif jika anomali tersebut terjadi. Peningkatan dengan TF-CHI juga dapat bersifat signifikan jika mampu memisahkan suatu label kelas dengan label kelas lainnya, sedangkan tidak akan berbeda jauh jika menghasilkan representasi yang relatif mirip dengan TF-IDF.

5.2 Saran

Penelitian yang telah dilaksanakan masih memiliki kekurangan dan keterbatasan. Oleh karena itu, berikut adalah saran yang dapat dipertimbangkan untuk meningkatkan dan atau melanjutkan penelitian ini:

1. Mengimplementasikan teknik *hyperparameter tuning* pada teknik SMOTE dan variasinya ketika anomali pada representasi TF-CHI terjadi. Salah satu teknik *hyperparameter tuning* yang dapat diimplementasikan adalah menentukan nilai pada *sampling strategy*. Selain itu, gabungan antara *undersampling* dan *oversampling* dapat diimplementasikan dengan tetap memperhatikan distribusi label kelas dan tidak menyebabkan kehilangan informasi (*information loss*).
2. Model yang dihasilkan dalam penelitian ini masih perlu dianalisis lebih lanjut agar dapat diimplementasikan terhadap data baru (tidak berasal dari sumber himpunan data). Topik-topik pengembangan penelitian yang dapat dipertimbangkan adalah penentuan pembobotan dengan teknik *clustering*, menganalisis korelasi antar label kelas, dan algoritma *greedy*.