

ABSTRAK

Melakukan perbaikan *bug* software setelah merilis software akan membutuhkan biaya yang jauh lebih mahal apabila dibandingkan dengan perbaikan ketika pengembangan software. Waktu juga menjadi salah satu faktor yang menjadi permasalahan pada penyelesaian *bug* pada software. Proses pengujian software membutuhkan waktu yang cukup lama. Penemuan *bug* software secara manual *software review* dan *software testing* hanya dapat mendeteksi 35% hingga 60% dari semua *bug* yang ada. Sehingga dapat disimpulkan bahwa cara manual *software review* dan *software testing* merupakan cara yang tidak efektif serta tidak efisien. Berdasarkan penelitian yang sudah dilakukan, permasalahan pada model prediksi *bug* software yaitu adanya ketidakseimbangan kelas pada dataset yang digunakan serta ketidakrelevanan atribut. Untuk menangani ketidakseimbangan kelas dibagi menjadi dua pendekatan, yaitu pendekatan level algoritma dan pendekatan level data. Pendekatan level algoritma digunakan untuk membuat pengklasifikasi agar lebih mudah menangani terhadap kelas minoritas. Algoritma yang populer digunakan yaitu *boosting* dan juga *bagging*. Sedangkan pendekatan level data digunakan untuk dapat memperbaiki ketidakseimbangan kelas dengan melakukan data sampling dengan algoritma *oversampling* ataupun *undersampling*. Algoritma *oversampling* yang sudah banyak digunakan yaitu algoritma *SMOTE (Synthetic Minority Over-sampling Technique)*. Pada penelitian ini digunakan *feature selection Wrapper* yaitu *forward selection* untuk dapat menangani ketidakrelevanan atribut. Sedangkan model klasifikasi yang digunakan adalah algoritma *C4.5*. Dataset yang digunakan yaitu dataset *NASA MDP (Metrics Data Program) repository*. Pada penelitian ini akan dikembangkan sebuah metode yaitu *Hybrid SMOTE-Bagging-Forward Selection* untuk dapat menangani ketidakseimbangan kelas serta ketidakrelevanan atribut pada prediksi *bug* software dengan algoritma *C4.5*. Hasil penelitian menunjukkan bahwa pendekatan level algoritma terbaik adalah *Bagging*. Pendekatan level data terbaik adalah *SMOTE*. Ketika kedua pendekatan ini digabungkan, model yang menggabungkan *SMOTE*, *Bagging*, *Forward Selection*, dengan *C4.5* mendapatkan akurasi yang tinggi dalam pengklasifikasian kelas minoritas.

Kata Kunci:

Prediksi *Bug* Software, Algoritma *C4.5*, Ketidakseimbangan Kelas, Pendekatan Level Data, Pendekatan Level Algoritma.