

Analisis Perbandingan Metode AdaBoost, SMOTEBoost, dan RUSBoost dalam Menangani Ketidakseimbangan Data pada Klasifikasi = Comparative Analysis of AdaBoost, SMOTEBoost, and RUSBoost Methods in Handling Imbalanced Data in Classification

Ruth Intan Davina, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551920&lokasi=lokal>

Abstrak

Ketidakseimbangan data merupakan tantangan umum dalam klasifikasi, di mana salah satu kelas memiliki ukuran sampel yang jauh lebih sedikit dibandingkan kelas lainnya dalam suatu dataset. Kondisi ini dapat menghasilkan klasifikasi yang memiliki akurasi prediksi yang tinggi untuk kelas mayoritas, tetapi cenderung rendah untuk kelas minoritas yang memiliki kontribusi kecil terhadap kesalahan total. Dalam aplikasi dunia nyata, kesalahan klasifikasi pada kelas minoritas sering kali memiliki konsekuensi yang lebih serius, seperti pada kasus deteksi serangan siber pada sistem keamanan jaringan. Kegagalan dalam mendeteksi serangan siber (false negative) dapat membuka celah keamanan yang berakibat fatal. Untuk menangani masalah ketidakseimbangan data, berbagai metode telah dikembangkan, termasuk pendekatan ensemble seperti SMOTEBoost (Synthetic Minority Oversampling Technique and Boosting) dan RUSBoost (Random Undersampling and Boosting). Pada penelitian skripsi ini dilakukan studi empiris pada data serangan malware dari dataset AWID3 menggunakan metode SMOTEBoost dan RUSBoost dan dibandingkan performanya dengan algoritma dasarnya, AdaBoost. Simulasi dilakukan dengan berbagai kombinasi hyperparameter dan variasi proporsi data training dan testing untuk mengevaluasi kinerja model secara komprehensif. Hasil penelitian menunjukkan bahwa metode SMOTEBoost dan RUSBoost memiliki kinerja yang sebanding dalam mendeteksi kelas minoritas, di mana nilai recall mencapai 0,99, dan lebih unggul dari metode AdaBoost dengan nilai recall 0,87-0,88. Penelitian tambahan yang dilakukan untuk mengevaluasi kinerja masing-masing metode pada berbagai jenis ketidakseimbangan menunjukkan bahwa kinerja metode AdaBoost menurun seiring dengan meningkatnya ketidakseimbangan relatif, sedangkan metode SMOTEBoost dan RUSBoost tetap stabil dengan kinerja yang baik. Namun, ukuran sampel minoritas yang terbatas atau absolute rarity memiliki dampak pada penurunan kinerja metode SMOTEBoost dan RUSBoost.

.....Imbalanced data is a common challenge in classification tasks, where one class has significantly fewer instances compared to others within a dataset. This condition can result in classification models with high predictive accuracy for the majority class but tend to perform poorly on the minority class, which contributes little to the overall error rate. In real-world applications, misclassifications errors on the minority class often bear more severe consequences, such as in the case of detecting cyber attacks in network security systems. Failure to detect cyber attacks (false negatives) can lead to security breaches with fatal consequences. To address the imbalanced data problem, various methods have been developed, including ensemble approaches such as SMOTEBoost (Synthetic Minority Oversampling Technique and Boosting) and RUSBoost (Random Undersampling and Boosting). In this thesis research, an empirical study was conducted on malware attack data from the AWID3 dataset using the SMOTEBoost and RUSBoost, and their performance was compared with their base algorithm, AdaBoost. Simulations were carried out with various combinations of hyperparameter and different train-test split to comprehensively evaluate the model's performance. The research results showed that SMOTEBoost and RUSBoost methods had

comparable performance in detecting the minority class, achieving remarkable recall values of 0.99, outperformed the AdaBoost method, which had recall values ranging from 0.87 to 0.88. Additional research conducted to evaluate the performance of each method on various types of imbalance showed that the performance of the AdaBoost method decreased as the relative imbalance increased, while the SMOTEBoost and RUSBoost methods maintained a stable and robust performance. However, a limited number of minority instances or absolute rarity had a negative effect on the performance of the SMOTEBoost and RUSBoost methods.