

Analisis Kinerja Metode Random Forest untuk Klasifikasi Multikelas Credit Scoring = Performance Analysis of the Random Forest Method for Credit Scoring Multiclass Classification

Yoel Zabarro, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920553093&lokasi=lokal>

Abstrak

Credit scoring adalah suatu proses dalam mengevaluasi kelayakan kredit dari suatu individu. Credit Scoring perlu dilakukan perusahaan keuangan untuk meminimalisir risiko kredit, karena credit scoring dapat menentukan kelayakan debitur. Salah satu perusahaan keuangan yang menyediakan jasa pinjaman berbasis P2P (Peer-to-Peer) yang menerapkan credit scoring dalam evaluasi debitur adalah LendingClub. Pada skripsi ini dilakukan klasifikasi multikelas credit scoring berdasarkan status pinjaman (loan status) yang terdiri dari 3 kelas, yaitu default, fully paid, dan late. Klasifikasi multikelas credit scoring dapat dilakukan dengan salah satu pendekatan machine learning, yaitu supervised learning. Metode supervised learning yang digunakan yaitu random forest. Random forest adalah suatu metode pencarian informasi berbasis tree dengan setiap tree memuat kumpulan variabel acak. Implementasi model random forest dilakukan dengan menggunakan tiga skenario strategy sampling SMOTE yang berbeda. Implementasi model pada tiap skenario dilakukan sebanyak 5 kali percobaan dan dievaluasi menggunakan precision, recall, f1-score, accuracy, dan AUC one vs all. Rata-rata accuracy terbaik adalah sebesar 0,78; dan rata-rata AUC one vs all terbaik adalah sebesar 0,679179. Sedangkan untuk hasil evaluasi berdasarkan tiap kelas, pada kelas default, precision terbaik adalah sebesar 0,39; recall terbaik adalah sebesar 0,27; dan f1-score terbaik adalah sebesar 0,28. Pada kelas fully paid, precision terbaik adalah sebesar 0,82; recall terbaik adalah sebesar 0,95; dan f1-score terbaik adalah sebesar 0,88. Pada kelas late, precision terbaik adalah sebesar 0,02; recall terbaik adalah sebesar 0,02; dan f1-score terbaik adalah sebesar 0,02. Secara keseluruhan, hasil evaluasi model pada ketiga skenario hanya baik dalam memprediksi kelas 1 (fully paid), tetapi kurang baik dalam memprediksi kelas 0 (default) dan kelas 2 (late). Hal tersebut diduga terjadi akibat dataset yang terdapat imbalance data dan class overlap.

..... Credit scoring is a process in evaluating the creditworthiness of an individual. Credit scoring needs to be done by financial companies to minimize credit risk, because credit scoring can determine the eligibility of debtors. One financial company that provides P2P (Peer-to-Peer) based loan services that applies credit scoring in debtor evaluation is LendingClub. In this thesis, a multiclass classification of credit scoring based on loan status was carried out consisting of 3 classes, namely default, fully paid, and late. Multiclass classification of credit scoring can be done with one of the machine learning approaches, namely supervised learning. The supervised learning method used is random forest. Random forest is a tree-based method of retrieving information with each tree containing a random set of variables. The implementation of the random forest model was carried out using three different SMOTE strategy sampling scenarios. Model implementation in each scenario was carried out 5 times and evaluated using precision, recall, f1-score, accuracy, and AUC one vs all. The best average accuracy is 0.78; and the best average AUC of one vs all is 0.679179. As for the evaluation results based on each class, in the default class, the best precision is 0.39; The best recall was 0.27; and the best F1-score is 0.28. In the fully paid class, the best precision is 0.82; The best recall is 0.95; and the best F1-score is 0.88. In the late class, the best precision is 0.02; The best recall is

0.02; and the best F1-score is 0.02. Overall, the results of model evaluation in all three scenarios were only good at predicting class 1 (fully paid), but less good at predicting class 0 (default) and class 2 (late). This is thought to occur due to datasets that contain data imbalances and class overlap