

Optimasi hyperparameter XGBoost-studi kasus prediksi klaim asuransi = Hyperparameter optimization in XGBoost-case study of insurance claim prediction

I Gede Pajar Bahari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509606&lokasi=lokal>

Abstrak

Metode Machine Learning sangat bermanfaat untuk menyelesaikan berbagai masalah, terutama pada masalah big data. Salah satu masalah yang terkait dengan big data adalah prediksi klaim asuransi di industri asuransi. XGBoost adalah metode machine learning yang menggunakan pendekatan ensemble learning dengan decision tree sebagai model dasarnya. XGBoost terdiri dari beberapa hyperparameter yang nilainya perlu ditentukan sebelum proses training. Partial grid search adalah optimasi hyperparameter yang biasanya digunakan untuk XGBoost. Random search adalah optimasi hyperparameter yang menentukan nilai hyperparameter secara random. Pada penelitian ini, diterapkan dan dianalisis metode pengoptimalan lain untuk XGBoost, yaitu Bayesian search untuk prediksi pengajuan klaim asuransi (dengan klasifikasi) dan besarnya klaim asuransi yang diajukan (dengan regresi). Tujuan dari penelitian ini yaitu membandingkan performa ketiga metode optimasi hyperparameter pada XGBoost: random search, partial grid search, dan Bayesian search pada klasifikasi dan regresi. Hasil simulasi menunjukkan bahwa partial grid search memberikan akurasi yang sedikit lebih baik dibandingkan dengan random search dan Bayesian search. Namun, waktu running pada Bayesian search jauh lebih cepat daripada partial grid search. Random search memiliki akurasi dan waktu komputasi yang sedikit kurang bagus dibandingkan dengan Bayesian search.Machine Learning Method is very useful for solving various problems, especially in the big data problem. One of the problems associated with big data is the prediction of insurance claims in the insurance industry. XGBoost is a machine learning method that uses an ensemble learning approach with a decision tree as its basic model. XGBoost consists of several hyperparameters which values need to be determined before the training process. Partial grid search is a hyperparameter optimization approach which is usually used for XGBoost. Random search is a hyperparameter optimization method which determines the value of hyperparameter randomly. In this study, another optimization method for XGBoost called Bayesian search is applied and analyzed for predicting insurance claims submissions (by classification) and the size of insurance claims submitted (by regression). The purpose of this study is to compare the performance of the three hyperparameter optimization methods on XGBoost: random search, partial grid search, and Bayesian search on classification and regression. The simulation results show that partial grid search gives a slightly better accuracy compared to random search and Bayesian search. However, the running time on Bayesian search is much faster than partial grid search. Random search has an accuracy and computation time that is a little less good compared to Bayesian search.