

Analisis kinerja XGBoost dalam menangani missing values: studi kasus masalah di industri asuransi = XGBoost performance analysis in handling missing values: case study of problems in the insurance industry

Deandra Aulia Rusdah, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20509482&lokasi=lokal>

Abstrak

Prediksi klaim dan prediksi risiko asuransi dilakukan untuk mengklasifikasikan klaim dan tingkat risiko dalam industri asuransi. Dari sudut pandang pembelajaran mesin, masalah prediksi klaim merupakan klasifikasi dua kelas dan masalah prediksi risiko adalah klasifikasi multi-kelas. Untuk mengklasifikasikan klaim dan risiko, model pembelajaran mesin akan memprediksi berdasarkan data historis. Dalam data historis pemohon asuransi, akan ada kemungkinan nilai yang hilang (missing values) sehingga perlu untuk mengatasi masalah tersebut agar memberikan kinerja yang lebih baik. XGBoost adalah metode pembelajaran mesin yang banyak digunakan untuk masalah klasifikasi dan dapat menangani nilai yang hilang (missing values) tanpa dilakukan proses imputasi terlebih dahulu. Penelitian ini menganalisis kinerja metode XGBoost dalam menangani nilai-nilai yang hilang (missing values) untuk prediksi klaim dan prediksi risiko dalam asuransi dan kinerja tersebut juga dibandingkan dengan metode imputasi standard lainnya, yaitu metode mean dan k-nearest neighbors (KNN). Simulasi menunjukkan bahwa model XGBoost tanpa dilakukan proses imputasi memberikan hasil yang sebanding dengan model XGBoost dengan imputasi

Claim prediction and risk prediction of insurance is carried out to classify claims and the levels of risk in insurance industries. From the machine learning point of view, the problem of claim prediction is a two-class classification, and the problem of risk level prediction is a multi-class classification. To classify the claims and risk, a machine learning model will predict based on historical data. In the insurance applicant's historical data, there will be the possibility of missing values so that it is necessary to deal with these problems to provide better performance. XGBoost is a machine learning method that is widely used for classification problems and can handle missing values without an imputation preprocessing. This study analyzed the performance of the XGBoost model in handling missing values for claim prediction and risk prediction in insurance and the performance is also compared with other standard imputation methods, i.e the mean and k-nearest neighbors (KNN) method. The simulations show that the XGBoost model without any imputation preprocessing gives a comparable result to one of the XGBoost models with an imputation preprocessing.