

Aplikasi XGBoost untuk prediksi probabilitas klaim asuransi kendaraan bermotor = The application of XGBoost for auto-insurance claim probability prediction

Muhammad Arief Fauzan, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20484897&lokasi=lokal>

Abstrak

Tren kenaikan frekuensi dan severitas klaim untuk klaim asuransi kendaraan bermotor menyebabkan dibutuhkan metode otomatisasi baru untuk memprediksi probabilitas seorang pemegang asuransi kendaraan akan mengajukan klaim jika diberikan data historis mengenai pemegang asuransi tersebut, agar perusahaan asuransi dapat memilah dan memproses lebih lanjut para pemegang polis yang kemungkinan mengajukan klaimnya tinggi. Masalah ini dapat diselesaikan dengan berbagai metode, salah satunya dengan machine learning, yang mengkategorisasikan masalah tersebut sebagai masalah supervised learning. Volume data yang besar dan banyaknya kemungkinan adanya missing values pada data pemegang asuransi menjadi dua aspek yang mempengaruhi pemilihan model machine learning yang tepat. XGBoost merupakan model gradient boosting machine learning baru yang dapat mengatasi missing value dan volume data besar sehingga XGBoost diklaim merupakan metode yang tepat untuk digunakan pada masalah tersebut. Dalam skripsi ini akan diaplikasikan metode XGBoost kepada masalah ini, dan akan dibandingkan hasilnya dengan berbagai metode machine learning lainnya, seperti AdaBoost, Stochastic Gradient Boosting, Random Forest, Neural Network, dan Logistic Regression.

.....The increasing trend of claim frequency and claim severity for auto-insurance result in a need of new methods to predict whether a policyholder will file an auto-insurance claim or not, given historical data about said policyholder, so that insurance industries can further process policyholders with high claim probability. This problem can be solved with many methods, one of which is machine learning, which categorizes this problem as a supervised learning problem. The high data volume and the existence of missing values on a policyholders historical data are aspects that the chosen machine learning model must be able to handle. XGBoost is a novel gradient boosting machine learning problem that is able to inherently handle missing values and high volume of data, which should make the model suitable for this problem. In this thesis, XGBoost will be applied to this problem, and its performance will be compared by other machine learning models, such as AdaBoost, Stochastic Gradient Boosting, Random Forest, Neural Network, and Logistic Regression.