

Analisis Kinerja Model Approximating XGBoost untuk Deteksi Fraud Klaim Asuransi = Performance Analysis of Approximating XGBoost Model for Insurance Claim Fraud Detection

Muhammad Adli Rahmat Solihin, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920551742&lokasi=lokal>

Abstrak

Industri asuransi merupakan industri yang besar dan terus berkembang di Indonesia. Di tambah lagi, semakin banyak masyarakat Indonesia yang terlindungi oleh asuransi pada tahun 2022. Di samping itu, meningkatnya nasabah asuransi juga dapat meningkatkan risiko terjadinya fraud atau kecurangan pada setiap klaim yang dilakukan nasabah. Hal tersebut dapat menyebabkan kerugian bagi perusahaan asuransi. Maka dari itu, dibutuhkan suatu metode yang dapat menyelesaikan masalah fraud klaim asuransi. Namun, permasalahan fraud ini melibatkan data yang besar sehingga dibutuhkan metode yang dapat memproses data besar dengan cepat dan tidak membutuhkan dana yang besar. Machine learning diajukan sebagai metode untuk dapat mendeteksi fraud klaim asuransi dan telah digunakan dalam beberapa penelitian untuk penyelesaian masalah tersebut. Extreme Gradient Boosting (XGBoost) merupakan salah satu metode machine learning yang banyak digunakan, pertama kali diajukan sebagai model yang scalable sehingga dapat memproses data besar secara efisien. Namun, metode machine learning sebaiknya adalah model yang interpretable dan explainable. Dengan memiliki interpretability dan explainability, suatu model dapat digunakan dan dimengerti dengan baik sehingga sangat reliabel. Oleh karena itu, Approximating XGBoost diajukan sebagai metode yang dapat memungkinkan XGBoost yang terdiri dari banyak boosted Decision Tree dapat diekstrak menjadi satu Decision Tree saja sehingga model ini dapat diinterpretasikan secara transparan mengenai pekerjaan model tersebut dalam mendapatkan hasil yang dikeluarkan. Penelitian ini menemukan bahwa kinerja dari model Approximating XGBoost dapat mendeteksi fraud dengan AUC-ROC mencapai nilai 91,44%, pada specificity dapat mencapai 86,55%, dan pada sensitivity dapat mencapai 91,34% berdasarkan dataset yang digunakan. Artinya, model Approximating XGBoost dapat membentuk suatu model yang performanya mengaproksimasi XGBoost dan dapat model pohonnya dapat ditampilkan seperti Decision Tree.

.....The insurance industry is a big and growing industry in Indonesia. Moreover, more and more Indonesians will be covered by insurance by 2022. In addition, increasing the number of insurers can also increase the risk of fraud or fraud in any claim made by the insurer. So from that, we need a method that can solve the problem of insurance claims fraud. However, this problem of fraud involves big data so it requires a method that can process big data quickly and does not require large funds. Machine learning was proposed as a method to detect insurance claims fraud and has been used in some research to solve such problems. Extreme Gradient Boosting (XGBoost) is one of the most widely used machine learning methods, first proposed as a scalable model to process big data efficiently. However, machine learning methods are preferably models that are interpretable and explainable. By having interpretability and explainability, a model can be used and understood well so it is highly reliable. Therefore, Approximating XGBoost was proposed as a method that could enable XGBoost consisting of many boosted

decision trees to be extracted into a single decision tree only so that this model can be interpreted transparently about the work of the model in obtaining the outcome issued. This study found that the performance of the Approximating XGBoost model can detect fraud with an AUC-ROC of 91.44%, specificity can reach 86.55%, and sensitivity can 91.34% based on the data set used. In other words, the XGBoost Approximating model can form a model whose performance approximates XGBuost and the tree model can be displayed like a Decision Tree.