

## INTISARI

### **Metode Analisis *Fraud* pada Asuransi Kendaraan Bermotor dengan Metode Regresi Logistik dan XGBoost dengan Bantuan SHAP**

Oleh

Cindy Wong

20/459244/PA/19905

*Fraud* merupakan suatu masalah di industri asuransi yang ditandai dengan adanya pemalsuan informasi baik dari pihak asuransi maupun dari pihak nasabah demi memperoleh keuntungan finansial. Selama ini, sudah ada banyak metode yang dapat digunakan untuk memilah klaim agar pihak asuransi dapat mengetahui tanda-tanda sebuah klaim merupakan fraud. Beberapa tahun akhir ada muncul metode perkembangan dari *decision tree* baru yaitu *Extreme Gradient Boosting* (XGBoost) yang dapat digunakan untuk melakukan klasifikasi. Namun, kelemahan dari metode ini adalah sulitnya untuk diinterpretasi terutama untuk menentukan variabel yang penting dalam model. Pada skripsi ini, data *fraud* pada asuransi kendaraan dibagi menjadi 70% *training* dan 30% *testing* untuk membentuk model regresi logistik dan XGBoost. Selanjutnya, XGBoost ditingkatkan daya interpretasinya dengan bantuan metode *Shapley additive explanations*. Kedua model ini dibandingkan kinerjanya dan didapat bahwa kedua model memiliki keuntungan masing-masing berdasarkan metrik yang diuji.

## ABSTRACT

### **Method in Fraud Analysis for Motor Insurance with Logistic Regression and SHAP-assisted XGBoost**

By

Cindy Wong

20/459244/PA/19905

In the insurance industry, fraud, signalled by the falsification of information to earn financial gain, has been a key issue that disadvantages both clients and insurers. Thus far, plenty of methods can be used to sort through claims so insurance companies can label the suspicious claims that could be fraudulent. These past few years, there is a new method that is made, based on the traditional decision tree called Extreme Gradient Boosting (XGBoost). However, one disadvantage of this method is the difficulty in interpreting the results, specifically the variables involved in making the predictions. In this paper, a dataset on motor insurance claim is divided into 70% training and 30% testing to build a regression model and an XGBoost model. Then, XGBoost is supplemented with a method named Shapley Additive Properties (SHAP) to increase the interpretability of its results. In the end, both models performance were compared and across the metrics used, both had their own advantages.