

Pembuatan model data mining untuk deteksi Fraud pada proses transaksi klaim: studi kasus PT XYZ = Fraud detection data mining modelling on claim transaction: case study PT XYZ

Deddy Utomo, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20516434&lokasi=lokal>

Abstrak

Jenis usaha perasuransian PT XYZ dibagi menjadi dua yaitu asuransi kesehatan dan asuransi jiwa. Salah satu risiko yang terjadi dan berdampak pada kerugian perasuransian adalah kecurangan atau fraud yang dilakukan pihak tertentu untuk memperoleh keuntungan sepihak. Penelitian ini dilakukan untuk membuat pemodelan data mining yang digunakan untuk mendeteksi fraud pada asuransi kesehatan. Tujuan dari penelitian ini adalah memperoleh algoritma model berbasis data mining yang dapat mendeteksi fraud pada transaksi klaim peserta di PT XYZ. Karakteristik data yang digunakan bersifat imbalanced, karena jumlah data fraud yang digunakan tidak sebesar jika dibandingkan dengan data yang bersifat normal. Pembentukan model pada penelitian ini dilakukan dengan 32 skenario, dengan hasil terbaik skenario dengan penerapan feature engineering, feature selection, oversampling dan uji validasi menggunakan 20-fold cross validation. Adapun hasil dari skenario tersebut menghasilkan algoritma random forest yang memiliki nilai akurasi paling baik yaitu 99,3% dengan didukung oleh nilai presisi, recall, dan f1 scores masing-masing, 99,3%, 99,3%, dan 99,3%. Hasil akhir dari penelitian ini memperlihatkan bahwa teknik feature engineering dengan penambahan atribut is_dr_speciality, memiliki kontribusi terhadap nilai akurasi model.

.....The type of insurance business of PT XYZ is divided into two, namely health insurance and life insurance. One of the risks that occur and impact insurance losses is fraud committed by certain parties to obtain unilateral benefits. This research was conducted to create a data mining model used to detect fraud in health insurance. The purpose of this study is to obtain a data mining-based model algorithm that can detect fraud in participant claims transactions at PT XYZ. The characteristics of the data used are imbalanced because the amount of fraud data used is not as much as compared to normal data. The model formation in this study was carried out with 32 scenarios, with the best results being the scenario by applying feature engineering, feature selection, oversampling, and validation tests using 20-fold cross-validation. This scenario resulted in the random forest algorithm having the best accuracy value of 99.3%, supported by precision, recall, and f1 scores, 99.3%, 99.3%, and 99.3%. The final result of this study shows that the feature engineering technique with the addition of the is_dr_speciality attribute has contributed to the model's accuracy value.