Fraud claim detection

Overview of the task

Developing a predictive model for insurance fraud detection offers significant business value. Fraud investigations currently rely on manual processes such as reviewing claims, calling claimants and conducting background checks, which are time-consuming and inefficient. These delays allow fraud to go undetected whereas legitimate claims face unnecessary scrutiny. Predictive modelling enables early identification of high-risk claims, streamlining fraud investigations, reducing financial losses and improving operational efficiency. It also enhances customer experience by expediting legitimate claims. Ultimately, an effective fraud detection model leads to better decision-making, optimised resource allocation and increased profitability.

Goal

- Global Insure aims to enhance its ability to detect fraudulent insurance claims by leveraging historical claim data.
- The company seeks to identify patterns and key indicators that differentiate fraudulent claims from genuine ones.
- By developing a predictive model, it intends to assess the likelihood of fraud in incoming claims, enabling proactive fraud detection and reducing financial losses.

Overall, the objective is to build a model to classify insurance claims as either fraudulent or legitimate based on historical claim details and customer profiles. By using features such as claim amounts, customer profiles, claim types and approval times, the company aims to predict the claims that are likely to be fraudulent before they are approved.

Findings

Two models were used & after doing comparative execution of them over sample dataset, Logistics regression fared better in comparison to Random Forest.

Logistic Regression :

- Validation accuracy: 84.31%
- Sensitivity : 65.38 %
- Specificity : 90.79 %
- Precision: 70.83.69 %
- F1 score: 68.00 %

Validation Accuracy :

- Validation accuracy: 78.43%
- Sensitivity : 57.69 %
- Specificity : 85.53 %
- Precision: 57.69 %
- F1 score: 57.69 %

Recommendations

Combining important features of both models, fraud claim settlement company should focus on following features to identify the fraud ratio:

- **incident_severity is “Minor Damage” : common to both models**
- **incident_severity is “Total Loss” : common to both models**
- Policy_annual_premium amount
- Property claim is involved
- **Insured hobbies is camping : common to both models**

Above properties can give hint regarding fraud.

Q & A

Q : How can we analyse historical claim data to detect patterns that indicate fraudulent claims

A : based on historical claims pattern is : if insurer has hobby as camping, incident severity is minor or total loss & if property is involved then it’s a RED flag.

Q: Which features are the most predictive of fraudulent behaviour

A: insured_hobbies, incident_severity, property_claim

Q: Based on past data, can we predict the likelihood of fraud for an incoming claim

A: Yes we can predict with confidence of 84% the likelihood of fraud

Q: What insights can be drawn from the model that can help in improving the fraud detection process

A: Logistics regression model performed better than random forest. Feedback of fraud detection should be feedback into model to improve it.