



FRADULENT CLAIM DETECTION

Ameya Parab

Amar Nath Kumar

23-Apr-2025

BACKGROUND

- Global Insure, a leading insurance company, processes thousands of claims annually. However, a significant percentage of these claims turn out to be fraudulent, resulting in considerable financial losses. The company's current process for identifying fraudulent claims involves manual inspections, which is time-consuming and inefficient. Fraudulent claims are often detected too late in the process, after the company has already paid out significant amounts.



PROBLEM STATEMENT

- Global Insure wants to improve its fraud detection process using data-driven insights to classify claims as fraudulent or legitimate early in the approval process. This would minimise financial losses and optimise the overall claims handling process..



OBJECTIVES

- Global Insure wants to build a model to classify insurance claims as either fraudulent or legitimate based on historical claim details and customer profiles.
- By using features like claim amounts, customer profiles and claim types, the company aims to predict which claims are likely to be fraudulent before they are approved.



SUMMARY OF ANALYSIS

- Random Forest model is selected for analysis due its better performance over Logistic Regression
- Users claiming high amount for major damages but paying low annual premium seems to be contributing more to the fraudulent claims
- Although sensitivity of model can be further improved, final Random Forest model is better placed to catch fraudulent claims



HOW CAN WE ANALYSE HISTORICAL CLAIM DATA TO DETECT PATTERNS THAT INDICATE FRAUDULENT CLAIMS?

- High Claim amount for major damage or total loss for a low annual premium seems to be a pattern for fraudulent claims.



WHICH FEATURES ARE MOST PREDICTIVE OF FRAUDULENT BEHAVIOUR?

- Incident Severity
- Annual Premium amount
- Total claim amount



CAN WE PREDICT THE LIKELIHOOD OF FRAUD FOR AN INCOMING CLAIM, BASED ON PAST DATA?

- Yes.
- Sensitivity of the model is 0.5
- Specificity of the model is 0.97
- Positive Likelihood ratio of Fraud is 16.66
- Negative Likelihood ratio of Fraud is 0.5
- Very High LR+ and low LR- suggests better suitability of the final Random Forest model for Fraud prediction



WHAT INSIGHTS CAN BE DRAWN FROM THE MODEL THAT CAN HELP IN IMPROVING THE FRAUD DETECTION PROCESS?

- Sensitivity of the final Random Forest Model is 0.5 which can be further improved.
- Increased sensitivity without lowering the existing strong specificity will certainly further boost the likelihood of the fraud
- Categories can be minimized in the following features to improve the sensitivity, incident_type, authorities_contacted, property_damage and police_report_available



RECOMMENDATIONS

