Fraudulent Claim Detection Business Summary

Submitted By

Pradeep Harry Michael
Pavan Kumar BN

Problem Statement

- **Challenge**: Global Insure, a major player in the insurance industry, is experiencing substantial financial losses due to a high volume of fraudulent claims. The existing fraud detection system relies heavily on manual inspections, which are not only time-consuming and labour-intensive but also inefficient. As a result, many fraudulent claims are detected only after payouts have been made, limiting the company's ability to prevent losses and straining operational resources.
- **Objective**: To address this issue, Global Insure seeks to enhance its fraud detection capabilities by leveraging data-driven insights and advanced analytics. The goal is to implement an intelligent system that can accurately classify claims as fraudulent or legitimate at an early stage in the approval process. This proactive approach would help the company significantly reduce financial losses, improve the speed and accuracy of claims processing, and optimize overall efficiency in claims management.

Business Summary

A data-driven approach for analysing historical claim records has revealed clear patterns associated with fraudulent behaviour.

Both Logistic Regression and Random Forest models were trained to predict fraud probability, with Random Forest achieving slightly better performance.

Categorical and numerical likelihood analysis identified features that significantly influence the likelihood of fraud.

High-variation features like incident_severity, insured_hobbies , total_claim_amount(vehicle/property/injury), policy_annual_premium, months_as_customer were strong fraud indicators.

Low-impact features (e.g., insured_sex, policy_state) contributed minimally and can be deprioritized or dropped to reduce noise.

(38577, 54)

Answers to Key Questions

1. **How can we analyse historical claim data to detect patterns that indicate fraudulent claims?**

   Performing target likelihood analysis on categorical and binned numerical features.

   Identifying anomalous behaviors

   Using EDA and modeling to detect non-obvious fraud signals in combinations of features.

2. **Which features are most predictive of fraudulent behaviour?**

Top predictors include:

- incident_severity
- insured_hobbies
- months_as_customer
- total_claim_amount, vehicle_claim, property_claim, injury_claim
- claim_per_month
- policy_annual_premium

3. **Can we predict the likelihood of fraud for an incoming claim?**

   Yes — using models like Logistic Regression and Random Forest:

   Claims can be scored in real-time based on fraud probability.

   A cutoff threshold (e.g. 0.565) can be applied to flag high-risk claims for review.

   This automates early detection and reduces reliance on slow, manual checks.

4. **What insights can be drawn from the model to improve fraud detection?**

   Prioritize high-impact features .

   Low-importance features can be removed for model simplification without loss of accuracy.

(38577, 54)

Recommendations and Business Implications

✅ **Recommendations**

Deploy fraud prediction model into the claims intake process to flag high-risk claims at the time of submission.
Prioritize manual reviews for claims with high fraud scores or unusual feature combinations.
Use feature importance in dashboards to help investigators understand why a claim is flagged.
Regularly retrain models on updated data to adapt to evolving fraud patterns.
Apply risk-based policy reviews for specific occupations , hobbies, or claim scenarios known to be fraud-prone.

💼 **Business Implications**

Reduced financial losses by catching fraud earlier and preventing payouts.
Improved operational efficiency, as manual investigations are focused on high-risk claims.
Increased transparency in fraud detection decisions with explainable AI tools.
Scalable and proactive fraud strategy that evolves with new data.