HEALTH INSURANCE FRAUD

CONTEXT

For health insurance fraud detection techniques to remain relevant, it's crucial for business to study the dynamics of continuously changing and complex claim frauds. AI and ML are relied upon to capture, analyse and summarize data, prepare and to maintain comprehensive data cubes for claim fraud detection rules. To continue to provide models which are effective and relevant, these technologies must also facilitate the training new rules and strategies that can be fine-tuned rigorously as predicted and actual data are continually contrasted.

A necessary component to achieving this success would be a *Claims Constraint Scoring Mechanism* developed from a comprehensive claims data cube that can accurately distinguish fraudulent claims from genuine claims.

Examples of claims include

- Collusion between doctor and patient
- Registering a patient with a pre-existing health condition without disclosing the condition
- A patient faking an ailment and making a claim

CHALLENGES FACED

Difficulties in identifying health care fraud:

- Circumstances vary widely, enabling a wide range of frauds
- The numbers of confirmed frauds in total claims population is negligible (< 0.2%)
- Fraud resulting in increase in premiums charged for bonafide customers

Thus, the rules and strategies can at best be empirical.

Key challenges from the data management perspective include:

- Capturing and inputting real time claims data during model development (syncing real time data with historic data cube)
- Building the relationship between claim dependent variables (such as reimbursement amount and procedure code) and claim independent variables (such as value of claim and whether fraudulent or not)
- Achieving and maintaining a significant level of accuracy across models
- Interpreting specific and unique fraudulent claims to create general strategy rules (in node and population segments)

SOLUTION

Developed and implemented rule-based health insurance fraud detection model which leverages artificial intelligence and machine learning. After a few iterations of model development, the arrived at model had better accuracy in identifying frauds and would proceed to reject potential fraud claims (in real time). Furthermore, the updated comprehensive claims data cube would continuously improve and fine tune the fraud detection model.

We were able to achieve this on lower data quality than normal for the industry, by expanding our models to incorporate the use of additional variables such as:

- Policy alteration date
- Number of days for alteration policy expiry date
- Claim date
- Reimbursement amount
- Count of number of third parties number of reimbursed claims detection trend
- Procedure code
- High cost claims

BUSINESS BENEFITS

With 20% test population, ~ $600k worth of fraudulent claims were rejected in 3 months of rules implementation. Extrapolating the Test vs. Control population, the Test performed better in terms of identifying and rejecting fraudulent claims and was crowned champion.