Fraud in Health Insurance Sample Case Study

Brief Background, Problem Statement and Objective

For Health Insurance Fraud Detection Techniques to remain Relevant, it's crucial for Business to Study the dynamics of continuously changing and complex frauds in Health Insurance claims. This encompasses a weighty dependence on AI and ML to capture, analyse and summarize the various facets of available data, prepare and maintain up-to-date comprehensive claims data cube for Fraud Detection Rules and for New Rules and Strategy to be trained and fine-tuned rigorously on Real Time data to remain effective and relevant.

And consequently, to develop a *Claims Constraint Scoring Mechanism* from comprehensive claims data cube which would Identify (with a significant level of accuracy) the Fraudulent Claims from the Genuine Claims and would be effective in Rejecting Fraudulent Claims.

Key Challenges

Difficulty in Identifying Health Care Frauds:

a. Each fraud is unique and very different from other frauds
b. Negligible confirmed frauds in total claims population (< 0.2%)

Thus, the Rules or Strategy can at best be empirical

Following were some key challenges from Data/Technical Perspective:

a) Capturing and Feeding Real Time Health Insurance Claims Data during Model Development (Syncing Real Time Data with Historic Data Cube)
b) Building Relationship between Claim Constraint Variables with other Independent Variables
c) Achieving and Maintaining a Significant Level of Accuracy in Model
d) Interpreting Specific and Unique Fraudulent Claims to create General Strategy Rules (in Nodes and Population Segments)

Solution Offered

Developed and Implemented Rule Based Health Insurance Fraud Detection Model which leverages Artificial Intelligence and Machine Learning. New Model has better accuracy in Identifying Frauds and would proceed to Reject Potential Fraud claims (in Real Time). Furthermore, the up-to-date comprehensive claims data cube would continuously improve and fine tune the Fraud Detection Model.

Significant Variables

Final Model Developed using a combination of below mentioned variables:

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 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