Fraud Scorer for Travel Insurance

Issues & Objective

Identify Potentially fraudulent claims through statistical and data mining technique with a view to

- Replace existing manual process by reducing man hours of effort and
- Focus on investigation of fewer loans flagged as fraudulent from the model

Challenges

- Data anomaly included duplicate cases, mismatch between claim type, cause type and loss type
- There were <100 fraudulent claims out of 30K claims – a hurdle for fraudulent pattern recognition and model building

Solution

- Machine learning technique Boosting (Adaboost algorithm) was employed for supervised learning primarily to account for the < 1% incidence of scam.
- Scoring model was highly discriminatory with a KS of 0.84. Fraudulent claims were assigned high scores in the model.
- Scrutiny was limited to the high scorers – suspicious cases. Investigation of only 5% claims succeeded in detecting 90% sham thus reducing human effort by a great extent.
- Recommendation on better acquisition and organization of data to save time
- Data checks and business rules were also laid out by Smart
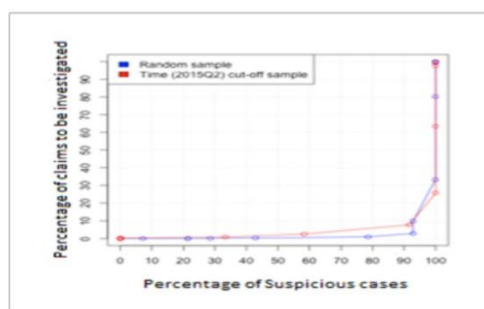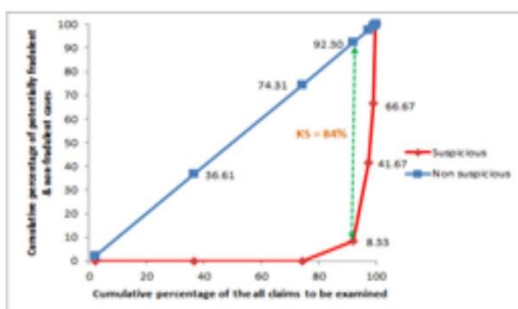
Benefits

- For a client dataset we could identify 92% of potentially fraudulent cases by examining only the 8% case with highest fraud score a lift factor of more than 11. It all cases it suffices to examine less than 20% of all cases (and usually less than 10%) to identify all fraudulent cases, cutting down manual work by a factor of 5 -10rt

Travel Insurance Fraud Identification