Claims analytics. An actuarial and data scientist perspective

CLINTON INNES, FSA, FCIA

September 20th 2020
Agenda

- Fraud analytics
- Approval and denial models
- Claim scoring
Fraud analytics
Foreign death fraud - motivation

“Warning: increasingly common and sophisticated scams

...In many countries, con artists operate without consequences because local authorities often do not have the physical or financial resources needed to combat Internet crimes...Organized fraud networks are developing more and more innovative and sophisticated approaches to deceive...The criminals conduct extensive searches to create credible documents: complete profiles of fictitious businesses, medical reports, falsified export certificates, etc. The names and logos of reputable organizations, governments and government agencies are often used fraudulently. Websites that appear very authentic are also falsified....”

Government of Canada Advisory, 30 May 2013
http://Travel.gc.ca/travelling/health-safety/overseas-fraud

- Misrepresentation of death occurred abroad can be as high as 2% of all individual life claims
- High fraud rates are observed in Middle East, Africa and China with misadventure being the common cause
- Machine learning methods can leverage multiple data sources to provide insight on main drivers of fraud and identify high risk foreign death claims
Data to consider

Demographics
- Age / Date of birth
- Gender
- Income
- Occupation
- Travel history
- Pre-existing diagnosis

Claim Data
- Country
- Cause
- Claim amount
- Claim Age
- Policy duration
- Body found
- Claim proofs

Underwriting and policy data
- Foreign travel, aviation, avocations
- Tobacco, drug and alcohol usage
- Agent, face amount
- Beneficiary
- Personal and family health history
- Disability, bankruptcy, felony
- Product, simplified UW vs Full

Third Party/External
- Corruption Perception Index
- Credit/public records
- Lifestyle
- MIB
- MVR
- Rx

Geospatial
- Population density
- Unemployment rate
- Tobacco tax
- Home value
- Household size
- Household composition
- Education
Corruption Perceptions Index (CPI)

- Important continuous indicator that allows to smooth out underrepresentation of fraud cases in individual countries
Third-party data

- Examples of third-party data include:

**Epsilon**

**Individual & household level**

- **Demographics:** Age, income, children, occupation
- **Property:** Home value, rent/own, living area
- **Financial:** Credit activity, number of tradelines, loans
- **Lifestyle/interests:** Pets, golf, travel, music
- **Ailments:** Asthma, diabetes, heart condition

**Lexis Nexis**

**Individual level**

- **Demographics:** Age, education, income
- **Court records**
- **Property:** Property ownership, value
- **Police records:** Bankruptcy, evictions
- **Credit records:** Auto finance, mortgage, retail credit accounts
Implementation

Demographics & Claim data

Predictive models

Additional Investigation

Standard Investigation

UW and Policy

Corruption Index

Credit

Lifestyle

Geospatial

MIB

Driving Record

Prescriptions

Essential Data Sources

Nice to have’s
Machine learning to detect key fraud drivers

Variable importance plot

- **Country**
- **Face amount**
- **Claim age**
- **Claim amount**
- **Cause**
- **Policy duration**
- **CPI**
- **Gender**
- **Occupation**
- **UW type**
- **Body found**
- **Smoker**

- **Data:**
  - **Demographics**
  - **Policy data**
  - **Claims**
  - **Corruption Perceptions Index**

- **Average fraud rate: 1.5%**

- **Machine learning methods allow to estimate marginal impacts of all variables on fraud likelihood and thus detect key fraud drivers**
Machine learning methods can be applied to segment all claims from least risky to most risky classes.

Example. The decision tree identifies claims with face amounts higher than $700K and policy duration under 4 years as high risk with fraud likelihood of 46%.
Machine learning to produce fraud risk score

Distribution of claims by fraud risk score

- In the presence of good-quality data more sophisticated machine learning methods can be applied to produce a continuous risk score.
- The score can then be used to determine optimal strategy for claim investigations.
- Example. Machine learning methods produce 0-10 fraud score for each claim. Claims can then be segmented into low, medium and high risk categories for targeted investigations.
Approval denial models
Overview

Motivation

- Claims adjudication is a complex process that requires significant amount of human intervention
- Average claim processing time can exceed several weeks

Objective

- Use machine learning to identify insurance claims that have a high risk of being denied
- Triage claims based on adjudication difficulty and likely outcome
How does it work?
Approval / denial models

Two model approach

- Risk of claim denial
- Risk of claim involving misrepresentation, non-disclosure or fraud
Approval/denial vs misrepresentation

Key drivers

- Cause of claim, duration and reporting lag are the key drivers of claim denial and misrepresentation predictions.
- Cause of claim is the most important feature predicting claim denial whereas policy duration is the most important feature for predicting likelihood of claim misrepresentation.

![Denial model: Variable importance](image1)

- Duration
- Claim amount
- Cause of claim
- Gender
- Age

![Misrepresentation model: Variable importance](image2)

- Duration
- Claim amount
- Cause of claim
- Gender
- Age

Importance
Implementation

Step 1: Assess denial risk
- High denial risk
- Medium denial risk
- Low denial risk

Step 2: Assess misrepresentation risk
- High misrepresentation risk
- Low misrepresentation risk

Step 3: Determine recommendation
- High denial risk
- High misrepresentation risk
- Low denial risk
- Low misrepresentation risk
- Medium denial risk
- Medium misrepresentation risk
- Low risk

Claims data
Case has a high risk of denial, with a high risk of misrepresentation.
Claims scoring
Overview

- Objective measure of the likelihood that a DI / LTD claimant will be able to return to work within a predetermined number of months (usually 12 or 24 months) from the benefit start date.

- Scores are determined with a predictive model calibrated to client’s actual claim history.

- Scores are on a scale representing the likelihood of return to work (RTW).

- Produce scores that guide claims team to be more efficient and effective with DI case management, i.e. optimize use of human and financial resources to generate more resolution in shorter durations.

Effort

Likelihood of RTW

- Low Score = Low Effort
- Medium Score = Medium Effort
- High Score = High Effort
Claims scoring: data
Data to consider

Public data sources

- Return to work predictions can be improved by incorporating publicly available data sources

- Census data at dissemination area level can be attached to claims data

- E.g. Census Topics include:

<table>
<thead>
<tr>
<th>Income of individuals</th>
<th>Household and dwelling characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age characteristics</td>
<td>Detailed mother tongue</td>
</tr>
<tr>
<td>Labor</td>
<td>Education</td>
</tr>
<tr>
<td>Population and dwelling counts</td>
<td>Coefficient of variation of after tax income</td>
</tr>
<tr>
<td>Housing</td>
<td>Income of households</td>
</tr>
<tr>
<td>Marital status</td>
<td>Family characteristics</td>
</tr>
</tbody>
</table>

- More insights for product development
Data to consider

Data by dissemination area

- As one would expect, return to work rates are lower in neighborhoods with lower % in labor force.
- Return to work rates vary by average income. If income is unavailable it can be supplemented with public data.
Other data sources

Text mining

- Return to work predictions can be improved by using text mining
- Text mining can be applied to client’s case managers’ notes or other unstructured data
- Can improve predictions of claim outcomes at claim onset using initial case managers’ notes or can improve predictions for mature claims by using notes entered at a later stage

What is text mining?

- The process of deriving high-quality information from free form text typically including text categorization, part of speech tagging, concept/entity extraction, sentiment analysis, document summarization, and dependency parsing
- Goal: turn unstructured text into useful data for analysis
## Other data sources: example

**Text mining: free form text to diagnostic group**

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Dx</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blood</td>
<td>4</td>
</tr>
<tr>
<td>Cancer – Other</td>
<td>5</td>
</tr>
<tr>
<td>Cancer – Bladder</td>
<td>6</td>
</tr>
<tr>
<td>Cancer – Endocrine</td>
<td>6</td>
</tr>
<tr>
<td>Cancer – Gastro</td>
<td>3</td>
</tr>
<tr>
<td>Cancer – Genitourinary</td>
<td>6</td>
</tr>
<tr>
<td>Cancer – Lymph</td>
<td>4</td>
</tr>
<tr>
<td>Cancer – Brain</td>
<td>2</td>
</tr>
<tr>
<td>Cancer – Esophageal</td>
<td>2</td>
</tr>
<tr>
<td>Cancer – Lungs</td>
<td>1</td>
</tr>
<tr>
<td>Cancer - Pancreas</td>
<td>1</td>
</tr>
<tr>
<td>Cancer – Breast</td>
<td>7</td>
</tr>
<tr>
<td>Cancer – Colorectal</td>
<td>6</td>
</tr>
<tr>
<td>Cancer – Prostate</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Dx</th>
</tr>
</thead>
<tbody>
<tr>
<td>Respiratory</td>
<td>7</td>
</tr>
<tr>
<td>Respiratory – Severe</td>
<td>2</td>
</tr>
<tr>
<td>Cardiovascular</td>
<td>4</td>
</tr>
<tr>
<td>Cardiovascular – Severe</td>
<td>2</td>
</tr>
<tr>
<td>Congenital</td>
<td>2</td>
</tr>
<tr>
<td>Endocrine</td>
<td>5</td>
</tr>
<tr>
<td>Other/Unknown</td>
<td>6</td>
</tr>
<tr>
<td>Ear</td>
<td>7</td>
</tr>
<tr>
<td>Eye</td>
<td>7</td>
</tr>
<tr>
<td>Eye – severe</td>
<td>2</td>
</tr>
<tr>
<td>Gastro</td>
<td>6</td>
</tr>
<tr>
<td>Renal</td>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Dx</th>
</tr>
</thead>
<tbody>
<tr>
<td>Infections</td>
<td>7</td>
</tr>
<tr>
<td>Injury</td>
<td>6</td>
</tr>
<tr>
<td>Muscular</td>
<td>5</td>
</tr>
<tr>
<td>Muscular – Light</td>
<td>7</td>
</tr>
<tr>
<td>Nervous</td>
<td>5</td>
</tr>
<tr>
<td>Nervous – Severe</td>
<td>2</td>
</tr>
<tr>
<td>Pregnancy</td>
<td>4</td>
</tr>
<tr>
<td>Psych</td>
<td>6</td>
</tr>
<tr>
<td>Psych – Severe</td>
<td>2</td>
</tr>
<tr>
<td>Skin</td>
<td>9</td>
</tr>
</tbody>
</table>
Claims scoring: modeling
Modeling

Modeling: overview

- Data for building claim scoring model to predict return to work in X months:
  - Terminated claims + open for more than X months
  - Approved or closed claim status
  - Is 70-80% of the entire block, the remaining part is for model testing

- Technique:
  - Different predictive models should be considered. Good results can be achieved by using Gradient boosted trees and Cox proportional survival model

- Key drivers of returning to work:
  - Diagnosis, age, reporting lag, short-term disability integration
Model performance

- Model performance can be assessed using a decile chart: it shows actual RTW rate for 10 buckets of claims ranked by the predicted likelihood of returning to work.

- The model effectively segments claims, from 1% actual RTW in the first decile to 70% in the last decile. The average RTW rate is 35%.
Action plan
Mapping scores to action plan

Distribution of claim counts by RTW likelihoods

Likelihood of RTW

- Low Score = Low Effort
- Medium Score = High Effort
- High Score = Medium Effort

Claim Counts

- <20%
- 20-30%
- 30-40%
- 40-70%
- 70-80%
- >80%

predicted RTW
Action plan: low score claims

Likelihood of RTW
- Low Score = Low Effort
- Medium Score = High Effort
- High Score = Medium Effort
Action plan: high score claims

Likelihood of RTW

- Low Score = Low Effort
- Medium Score = High Effort
- High Score = Medium Effort

High score claims

- Psych 50%
- Cancer-Breast 3%
- Cardiovascular 6%
- Other/Unknown 6%
- Gastro 5%
- Muscular (all) 6%
- Injury 25%
Action plan: medium score claims

Likelihood of RTW

- Low Score = Low Effort
- Medium Score = High Effort
- High Score = Medium Effort
Claim scoring: Implementation
Implementation

Recommended action plan based on predicted score and actual claim duration:

<table>
<thead>
<tr>
<th>Claim Duration</th>
<th>Low Touch</th>
<th>Active Management</th>
<th>Monitor Expected Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-12 months</td>
<td>Identify early pension plan applicants</td>
<td>Work toward claim resolution</td>
<td>Light touch</td>
</tr>
<tr>
<td></td>
<td>Review for financial recovery – settlement, pension plan options</td>
<td>Overcome barriers to RTW</td>
<td>Follow action plan</td>
</tr>
<tr>
<td>&gt; 12 months</td>
<td>Work towards COD resolution</td>
<td>Investigate reasons for non resolution</td>
<td>Work towards resolution</td>
</tr>
</tbody>
</table>

Recommended action plan based on predicted score and actual claim duration:

- **Low Score = Low Effort**
- **Medium Score = High Effort**
- **High Score = Medium Effort**
Partnership is the key to claims analytics
Data is essential