2017 Predictive Analytics Symposium

Session 7, Risk Assessment Applications of Predictive Analytics

Moderator:
Priyanka Srivastava

Presenters:
Dihui Lai, Ph.D.
Nitin Nayak, Ph.D., MBA
Jason L. VonBergen, FSA, MAAA

SOA Antitrust Compliance Guidelines
SOA Presentation Disclaimer
Using Machine Learning for Accelerated Underwriting

Dihui Lai, PhD
Reinsurance Group of America

Sept, 2017
Overview

- Background
- Accelerated Underwriting
- Model Structure and Model Performance
- Model Interpretability and Model Validation
Background: Term Life Insurance Application Flow

- Complete Forms 1-Day
- Paramed Exam 1-Week
- Review 2 - 4 Weeks
- Policy Issued 1-Week
- Policy Signed 2-3 Days
- Placed in Force 2-3 Days

- Time Consuming
- Medical Exams are NOT always Pleasant
- Extra Expenses
Background: Term Life Insurance Application Flow

Complete Forms → Placed in Force
## Comparisons: SI v.s. Accelerated Underwriting

<table>
<thead>
<tr>
<th>Simplified Issue</th>
<th>Accelerated Underwriting</th>
</tr>
</thead>
<tbody>
<tr>
<td>▪ No fluids for any applicants</td>
<td>▪ No fluids for certain percentage of applicants</td>
</tr>
<tr>
<td>▪ Short application</td>
<td>▪ Full application with drill downs</td>
</tr>
<tr>
<td>▪ Use Rx, MIB, MVR database referenced</td>
<td>▪ Use Rx, MIB, MVR database referenced</td>
</tr>
<tr>
<td>▪ Relatively Low Face amount</td>
<td>▪ Face amount Comparable to Full UW</td>
</tr>
<tr>
<td>▪ Term typical</td>
<td>▪ Term or permanent</td>
</tr>
<tr>
<td>▪ Typically rates are higher than fully UW</td>
<td>▪ Targeting fully UW rates</td>
</tr>
<tr>
<td>▪ One preferred Class (still expensive)</td>
<td>▪ Including all preferred classes</td>
</tr>
</tbody>
</table>
Accelerated Underwriting Workflow

Machine Learning

Full Application

Does applicant meet the requirements?

Apply Full Underwriting with Fluids

Accelerated offer

Does applicant meet the requirements?

Does applicant meet the requirements?
**Model Selection**

**GLM**
- Interpretability, transparent coefficients
- Limited capability of explaining non-linearity

**Tree**
- Non-cyclic binary rule structures
- Interpretability in the form of a single tree
- Easy to be ensemble to “forest”

**Neural Network**
- All-star model
- Widely integrated for face-recognition, auto-drive, speech recognition.
- Low transparency and interpretability

**SVM**
- Non-probabilistic based classifier
- Able to explain complex geometry structure
- All-star until the breakthrough in deep learning
The Hierarchical Model Structure

- Mortality
  - Classifier for Multiple-Underwriting Classes
    - Classification Trees + Neural Network
  - Classifier for Declined Risks
    - Classification Trees
  - Classifier for Prefer-ness
    - Classification Trees
  - ……
Model Performance Assessment

Important Variables:
- BMI
- Age
- Prescription Count
- ...

Low Risk  

Preferred Class  Not Preferred Class  

High Risk
Model Interpretability and Validation

- Understand the Complex Variable Impact
- Diagnostic Analysis
- Monitor Shifts of Distribution in Application Population
- Compare Model Decision with Human Underwriting
Predictive Analytics for Life Insurance

Predicting Applicant’s Smoking Propensity for Application Triage

• Business Problem: Can one predict an applicant’s smoking status without fluid-testing?
Worldwide opportunity for life protection is about US$ 8.6T. In the US, mid-market represents a significant opportunity.

- 52MM households
- $378K average protection gap

**US Middle Market**

**Strategy for closing protection gap**

- Reduce friction in underwriting and acquisition processes

**Increase Relevance**

**Engage the changing needs of today’s consumer**

**Existing distribution channels favor higher policy sizes & not mid-market**

Source: Swiss Re Sigma Report No. 6/2013
Swiss Re’s motivation and approach for supporting “fluid-less” life insurance underwriting

• Current underwriting process for life insurance is costly and time-intensive
  – requires laboratory tests (blood, urinalysis), paramedics (height, weight)
  – takes weeks - months, increasing likelihood of applicant “walking away”

• Swiss Re is addressing this challenge, starting with tobacco classification of applicants as key focus
  – Smokers have 1.75 to 3-fold higher mortality than non-smokers.
  – US life insurance industry loads actuarial pricing up to 200% more for tobacco use
  – After age and gender, **tobacco use, especially cigarette smoking**, is the single most important factor for risk loading of life insurance policies.
Developing a “fluid-less” underwriting process based on detecting “smoker propensity” poses several challenges

• High performance expectation
  - Sensitivity/specificity of smoker detection solution does not equal or exceed the best medical screening tests thus far.

• Non-disclosed smoking in insurance applications
  - Identifying smokers from insurance application is difficult due to large number (up to 50%) of non-disclosed smokers, i.e., actual smokers self-reporting as non-smokers

• No smoker-specific profile available to identify smokers
  - Difficult to detect smokers using “smoker” characteristics in application data.
3-part solution approach is designed to address the challenges of fast underwriting for life insurance policies

1. A Predictive Analytics Model
   - Model designed to predict smokers and non-smokers

2. A Triage-based Underwriting Process
   - Majority of applicants (go through Fast Track process requiring no lab (cotinine) tests for smoking
   - Predicted smokers go through Traditional (business-as-usual) process with lab test required

3. A Cost/Benefit Analysis and Optimization Model
   - Analyzes cost impact of prediction errors (i.e., misclassification of smokers as non-smokers) & savings from fast track with no lab-test for majority of applicants
   - Computes age, gender, and face amount requirements for a for client-specific life product with positive NPV

*Following slides provide details on the 3-part solution*
## Analytics Model: Sample predictors used from internal & external data sources

<table>
<thead>
<tr>
<th><strong>Sample Application Data</strong></th>
<th><strong>Sample Data from External Open Sources (CDC, ALA, etc.)</strong></th>
<th><strong>US Data from 3rd party vendors</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Tobacco-related data by State:</td>
<td>Medical Information Bureau (MIB)</td>
</tr>
<tr>
<td>PlaceOfBirth</td>
<td>• Tobacco tax</td>
<td>Motor Vehicle Records (MVR)</td>
</tr>
<tr>
<td>InsuranceAge</td>
<td>• Smoking cessation spending per smoker</td>
<td>Prescription History (Rx)</td>
</tr>
<tr>
<td>AlcoholAbuseFlag</td>
<td>• Laws banning smoking in public spaces</td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>• Number of tobacco retailers per 10K</td>
<td></td>
</tr>
<tr>
<td>DrugAbuseFlag</td>
<td>• Smoking rates by county</td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BenefitTermLife</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BenefitAmount to Income Ratio</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Analytics Model: Model’s prediction performance is good on several metrics

Performance Metrics Explained

Recall (R):
What percent true-positives in the population are correctly identified?

Precision (P):
What percent predicted positives are indeed true positives?

F-score (F):
Useful metric for skewed class population
\[ F = \frac{2 \times P \times R}{P + R} \]

Area under ROC curve (AUC):
Higher value (closer to 1) indicates good prediction performance

Prediction Model Details

Problem Type: Classification

Machine Learning Techniques used:
- GBM (best performance)
- GLMNET (Logistic regression)
- Random Forest
Triage using predictive analytics model supports fast-track processing for majority of the applicants (> 84%)

Life Insurance Application Details

- Self-Declared Smoker
- Self-Declared Non-Smoker

Apply Predictive Model

- Predicted Smoker
- Predicted Non-Smoker

- Business as usual (< 16%)
- Fast Track (> 84%)

- Lab Test Reqd.

- Tested Smoker
- Tested Non-Smoker

- Smoker Rate
- Smoker Rate
- Non-Smoker Rate
- Non-Smoker Rate

Note: Tobacco Usage is only one aspect of the overall risk.
Cost-Benefit: Calculator computes NPV of life product using predictive model and actuarial data

Prediction Model Results

Cost-Savings Calculator

Actuarial Data

<table>
<thead>
<tr>
<th>Smoking Status</th>
<th>Male PV of Mortality Cost by Smoking Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;25</td>
<td>$1,175, $1,556, $3,642, $8,353, $25,55, $54,03, $116,6, $214,0</td>
</tr>
<tr>
<td>25-34</td>
<td>$816, $826, $1,507, $3,129, $7,615, $23,98, $78,00, $161,2</td>
</tr>
</tbody>
</table>

- **Actual Non-Smokers**
  - Lab-testing savings
  - Increased Mortality Costs
  - Lab-testing savings

- **Actual Smokers**
  - Business as usual
  - Business as usual
For ages below 55, Lab-test Savings > Mortality Costs results in positive NPV

For ages 55 and beyond, Mortality Costs > Lab-test Savings results in negative NPV

Cost-Benefit: 10-year term life product for applicants below age 55 and face amount < $100K

Costs, Savings, and Net Benefit (NPV) displayed by applicant’s Age
(population = 100,000 applicants, product = term life with $100K face amount)

Actuarial Data: Source-LMS US data on PV (Mortality Costs) based on age, insured amount, gender, product term
Cost Assumption: Lab testing cost $55 (does not include parameds)
Note: Revenue impact of fast underwriting process is not included in calculations
©2017 Swiss Re. All rights reserved. You are not permitted to create any modifications or derivative works of this presentation or to use it for commercial or other public purposes without the prior written permission of Swiss Re.

The information and opinions contained in the presentation are provided as at the date of the presentation and are subject to change without notice. Although the information used was taken from reliable sources, Swiss Re does not accept any responsibility for the accuracy or comprehensiveness of the details given. All liability for the accuracy and completeness thereof or for any damage or loss resulting from the use of the information contained in this presentation is expressly excluded. Under no circumstances shall Swiss Re or its Group companies be liable for any financial or consequential loss relating to this presentation.
RISK ASSESSMENT APPLICATIONS OF PREDICTIVE ANALYTICS

Jason Von Bergen, FSA, MAAA

September 14th, 2017
## Digital Evolution of Risk Assessment

<table>
<thead>
<tr>
<th>FROM</th>
<th>TO</th>
</tr>
</thead>
<tbody>
<tr>
<td>Paper questionnaires</td>
<td>Digital applications</td>
</tr>
<tr>
<td>Invasive paramedical exams</td>
<td>Non-invasive</td>
</tr>
<tr>
<td>Multiple requests for records</td>
<td>Real-time access to data</td>
</tr>
<tr>
<td>Age/amount determined</td>
<td>Customized</td>
</tr>
<tr>
<td>Decision in weeks</td>
<td>Decision in minutes or days</td>
</tr>
</tbody>
</table>
Enabling Factors to Accelerate Innovation

Customer Centricity

Mortality & Expense Focus

Mortality & morbidity excellence

Holistic & prioritized approach

Distinctive & customized experience

Data Availability

Advanced Analytics

Computational Capability
Business Motivations to Change

1. **Customer Experience**
   - Enabling a rich digital experience
   - Solving today’s pain points
   - Customers want a simple experience. We can deliver CX AND mortality excellence.

2. **Expense Savings**
   - Multi-million/yr. Home Office opportunity
   - HUGE opportunity. We won’t trade class-level mortality loss for expense savings.

3. **Future Optionality**
   - Rich data delivers insights
   - Insights drive design
   - Risk class segmentation, product offerings / pricing, claims processes, etc.

4. **Competitive Position**
   - Most others are doing something
   - InsureTech pushing boundaries
   - There is potential anti-seletion risk in not offering anything.
## What to Understand Before Beginning

### Sources of Mortality Value

- Mortality performance & drivers
  - Including declines and process drop-outs
  - Connection w/ philosophy & process
- Quantified protective value studies

### Data & Modeling

- State of current data
- Data change processes
- Modeling infrastructure & maturity

### Philosophy & Process

- Risk assessment philosophy compatibility w/ triage
- New Business & policy acquisition process
  - Home Office & distribution partners
  - Motivations & change management

### Program Goals & Constraints

- Are you willing to trade mortality for expenses?
- What differentiates you when data becomes commoditized?
# Example Modeling Target Choices

<table>
<thead>
<tr>
<th>Good Requirement(s)</th>
<th>Advantages</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Potential collection of models</td>
<td>– Rules make decisions; models provide inputs</td>
<td>– Requires much effort to knit together</td>
</tr>
<tr>
<td>– E.g. “Good Blood” model</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Best Class</th>
<th>Advantages</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Matches best class decisions</td>
<td>– Contained group w/ potentially less UW bias</td>
<td>– Limited expense savings based on size of group</td>
</tr>
<tr>
<td>– A form of requirements triage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multi-Class</th>
<th>Advantages</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Matches underwriter decisions</td>
<td>– Larger expense savings; lessens selective gaming</td>
<td>– Perpetuates UW bias; communicate adverse action</td>
</tr>
<tr>
<td>– Combinations of models &amp; rules</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mortality</th>
<th>Advantages</th>
<th>Challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>– Modeling to mortality outcomes</td>
<td>– No programmatic bias; potential for better mortality</td>
<td>– Requires a lot of data; projects improvements</td>
</tr>
<tr>
<td>– Free of a priori underwriting expectations</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
• Which data elements are more/less relevant based on current protective value studies?

• How can you incorporate data in model target / methodology?

• How robust is the data – proven or still experimental?

• Will the data be used for accelerated UW only or also in traditional?

• What can the data be used for, i.e. any regulatory concerns?
Model Construction & Review

Model Throughput vs. Mortality Cost

By Auto-approval Rate

By Mortality Cost per Approval

ILLUSTRATIVE

Basic Underwriting Data
300,000 Observations

FR Data
171,000 Observations

Prior Activity & MIB data
169,000 Observations

Refinement and Feature Creation
169,000 Observations

Adjusted Outcome Variable
157,000 Observations

Initial MHQ Data
25,000 Observations

Sprint 2
Sprint 3
Sprint 4
Sprint 5
Sprint 6
Sprint 7
Performance Reporting & Monitoring

Model Monitoring Mechanisms

- Random hold-out sample (e.g. 10%)
- Post-issue APS or Rx scan to study
- Post-issue APS or Rx scan to rescind
- Beta testing with live data before release

How does this impact your desired client experience?

Accelerated Underwriting Performance Reporting

- Weekly reporting of numbers of cases approved
- Monthly report with detailed break-down of model eligibility and throughput by age & amount
- Quarterly hold-out sample miss analysis – occurrence & severity
Lessons Learned

1. Create a **data roadmap** early on to identify priorities

2. Spend some time understanding **data transformations** during underwriting

3. Deep **leadership by business experts** speeds development and iterative delivery

4. Be **flexible** in development & deployment of the model

5. Engage underwriters **early & often** to drive understanding