Session 1

Advanced Analytics Applications in Life Insurance

Marc Sofer, FFA FIAA, MBA
Advanced Analytics Applications in Life Insurance

MARC SOFER  BSc MBA FFA FIAA
Head of Data and Strategic Analytics
Asian Markets
Agenda

- Advanced analytics use cases in life insurance
- Data usage
- Case studies:
  - Fraud management and underwriting triage
  - Credit and Driving Behavior
  - Up-sell and cross-sell
  - Other interesting projects
- Closing thoughts
Built and deployed many predictive models globally

**EUROPE**
- Predictive UW model
- Fast UW model

**UNITED KINGDOM**
- Basis setting (mortality, morbidity and lapses)
- Postcode pricing model
- Enhanced experience analysis
- Smoking prediction
- Predictive underwriting (credit agency)
- Predictive underwriting (bancassurance)
- Agent quality assessment
- In-force retention program

**SOUTH AFRICA**
- Non-disclosure UW model
- GLTD termination rate model
- Propensity-to-buy for insurance

**INDIA**
- Claims fraud prediction

**UNITED STATES**
- Pricing override model for group LTD
- Risk segmentation for cross-sell
- Client segmentation
- Geographical risk segmentation
- Lapse basis
- Predictive UW/cross-sell model using non-life data
- Term tail lapses
- Agent quality assessment
- In-force retention program
- UW case prioritization

**ASIA**
- Risk segmentation for cross-sell & upsell
- Predictive underwriting on bancassurance data
- UW guideline refinement
- UW investigation model
- Finer price segmentation
- Actuarial experience study
- Propensity-to-buy model

**AUSTRALIA**
- Predictive UW/cross-sell on bancassurance data
Advanced data analytics use across the value chain

Many applications but we are being asked to concentrate on a few

- **Pre-sale**
  - New rating factors
  - UW risk segmentation
  - Propensity to buy
  - Distributor quality management
  - Competitive pricing strategy

- **Underwriting**
  - UW Triage
  - Fraud /Non-disclosure protection
  - Propensity to complete purchase
  - Determine UW ratings

- **In-force management**
  - Multivariate experience analysis
  - Cross-sell & up-sell
  - Proactive lapse management
  - Customer lifetime value

- **Claims**
  - Fraud / Non disclosure protection
  - Claims triage

**Level of Client Demand**
- Low
- Medium
- High
Global interest in new rating factors drawn from a holistic view of health drivers

Basic Demographics
- Age
- Gender
- Education
- Occupation

Socio-Economic
- Assets
  - Salary
  - House price

Behavioural
- Insurance purchasing
  - Social Media Usage
- Marital status
- Number of Children
- Benefit amount
- Where you live
- Car ownership
- Investment choices
- Care provider
- Previous claims history
- Avocations
- Credit Behaviour
- Driving Behaviour
- Personality
  - Well-being
  - Stress
- Smoker Status
- Sleep
- Activity
- Diet

Health and Biometrics
- Disease history
- Family History
- Prescription History
- Heart Rate / HRV / etc.
- Genetic Data

Not intended to be exhaustive
Data sources being sought and used within advanced analytics applications

Desirable Attributes
- Relevant
- Granular
- Voluminous
- Accurate
- Timely
- Permissioned

Potential data sources
- Credit bureau
- Bank transactions
- Credit card transactions
- Marketing
- Non-life
- Wearable devices
- Social media
- Retail transactions
- Mobile usage
- Life assurance
Case Studies:
Fraud Management and Underwriting Triage
Fraud Management and Detection Survey

- Initiative to study and identify risk factors associated with fraud
- Data analysis to identify characteristics of early potentially fraudulent claims
- A dynamic approach to underwriting requirements based on the fraud risk profile of the proposals submitted
- Allows insurers to apply different underwriting standards to cases on a set of criteria that were previously mainly medical
“What percent of your company’s income is lost annually because of fraud?”

Common characteristics observed in claims with suspected fraud:

- Non-graduates or low schooling
- Self-employed
- Non-urban/negative locations
- Involvement of distribution partners (financial advisors)
- Very low and non-verifiable income
- Early death claims

Source: Fraud Management Survey for India – 2013 conducted by RGA
Fraud Management Analysis

A/E Claim Experience by Education

- Professional: 3.5%
- Post Graduate: 4.0%
- Graduate: 4.5%
- HSC: 5.0%
- SSC: 5.5%
- Uneducated: 6.0%

Actual vs. Expected

Level of Education

A/E Claim Experience by Geographic Location

- Top 10 Cities: 18%
- 11 to 25 Cities: 19%
- Rest of India: 20%
- Negative Location: 21%

Actual vs. Expected

Cities

A/E Claim Experience by Occupation

- Professional: 479%
- Executive: 474%
- Office Manager: 470%
- Housewife: 465%
- Retired: 460%
- Police: 455%
- Business Owner: 450%
- Office Clerk: 445%
- Technician: 440%
- Driver: 435%
- Agriculture: 430%

Actual vs. Expected

Occupation

A/E Claim Experience by Income

- > 2.5 mio: 202%
- > 1 - 2.5 mio: 200%
- > 0.5 - 1 mio: 198%
- > 0.3 - 0.5 mio: 196%
- > 0.2 - 0.3 mio: 194%
- 0.1 - 0.2 mio: 192%
- Up to 0.1 mio: 190%

Actual vs. Expected

Annual Income in Indian Rupees

Source: RGA internal studies
Case Studies:
Credit and Driving Behaviour
Credit behaviour based insurance score

Predictive of mortality and other insurance outcomes

- RGA and TransUnion Partnership: Created and validated a credit behavior score
- Relative risk score with 1 representing lower risk and 100 representing higher risk

Key Applications

- Cross/Up Sell
- Simpler Underwriting
- Risk Segmentation
- In-Force Policy Management
- Claims Repudiation Management

Models deployed globally

- 17 clients in production
- 25+ clients in contracting
- 1 client in production
- 5+ clients in discussion
- 1 client in pilot
- 5+ clients in discussion
- 1 client in pilot
- 5+ clients in discussion
What’s is in the Model?

Big Data Universe filtered to get the most important attributes

~ 1,000 credit attributes

Attributes: predictive, stable, non-gameable, uncorrelated

Key credit attributes that quantify behavioral risk

Key credit attributes

Credit Data Inputs

Credit seeking activity
- New trades
- Inquiries
- Frequency

Derogatory credit information
- Severity
- Count
- Amount

Credit usage
- Utilization
- Patterns

Credit tenure
- File thickness
- Longevity
US TrueRisk® Life Model

- Core model built on 40 million lives and over 3 million deaths
- Scores validated on 18 million lives
- Score buckets are set to be uniform across the population
- Study shows 5 times segmentation (96-100 compared to 1-5)

- Segmentation exists within risk classes even after medical underwriting
- Non-smokers are shown, but results are similar for smokers
Consisted results across the globe

Canada Individual Insured Mortality

US Group Insured Mortality

SA Individual Population Mortality

India Individual Insured Mortality
Impact of driving behavior on mortality

Background
- RGA study aimed at quantifying additional mortality risk of applicants with adverse driving history
- Study correlation with all cause mortality and not just accident risk

Data
- Random sample of 7.4 million Motor Vehicle Records with approximately 73,000 deaths
- Use of basic demographic data, motor vehicle infractions, violation codes, date of infractions and date of death

Applications
- Risk segmentation at time of UW
- Combined with other risk indices for faster underwriting processing

Relative Mortality
- Significant excess mortality risk for those with “major” violations
- True across age and gender and risk increases as no. and type of violations increase
Case Studies:
Up-sell and cross-sell
Bancassurance Predictive Underwriting

Objectives

- Make the insurance sales process simpler and faster for bank customers and sales force
- Use all available data sources to remove or lessen underwriting for customers most likely to be in good health
- Reduce acquisition expenses

Applications

- Fully underwritten products offered to lower risk customers on a simplified issue or guaranteed issue basis (subject to specific risk controls)
- Simplified issue products offered to lower risk customers on a guaranteed issue basis (subject to specific risk controls)
Data available for modeling purposes

Sample data fields

Demographic
- DOB
- Gender
- Marital Status
- No. of Children
- Education
- Occupation
- Postal Code
- Car Ownership
- Customer Segment
- Customer Tenure

Financial
- Annual Salary
- Credit Score
- Credit Limits
- Risk Appetite
- Total AUM
- Account Balances
- CC Balance
- Personal Loans
- Mortgage Loans
- Total Insurance Premiums

Transactional
- Various accounts, products and balances over time
- Fund transfer usage on cards
- Late Fees/Over limit fees
- CC spend on various categories

Underwriting
- Underwriting Decision
  i.e. Standard Loaded Excluded Declined

For all projects, data was de-identified and kept secure with modeling either completed in the bank, insurer or RGA’s secure environments
Predictive Modeling Approach

Predictive Underwriting

Depersonalized data from bank and insurer common customers

Prospect list is generated by passing bank customer data through the ‘model filter’. A score is then calculated for each prospect and those with the highest scores can obtain coverage with reduced underwriting.
Bancassurance Predictive UW projects

3 projects successfully delivered in SEA and all 3 projects were unique in their own way

01
Bank data available across projects varied both in terms of type and volume

02
Models were developed from scratch for each project

03
Variables contributing to each of the final models are different
What is a Lift Curve?

- Useful visual representation of a model's predictive power.
- Primary goal is to differentiate standard and non-standard risks.
- Lift curve is generated by sorting individual model output into 10 equal-sized groups called deciles with 10% of the population.
- We then calculate the average non-std. rate for each of the 10 groups.
- For a “null” model, the distribution of non-std. policies across the deciles would be uniform. For a “perfect” model, there would be no non-std. policies appearing in the top deciles.
- In reality, we expect to see a model lift curve fall somewhere between the two extreme cases.
Model Performances across projects

- All 3 models gave good results and were able to identify groups of customers with low average non-std. rates
- GIO and SIO offers were made to the top 5 deciles (with appropriate risk controls put in place)

Approximate results have been presented to protect client results and confidentiality
Key Learnings

01 Distribution
Understand the sales process in detail as success is governed by how products will be sold by distribution force.

02 Propensity to Buy
Identify customers who are both better risks and have a high likelihood of purchase.

03 Right Product
Ensure that products offered are appropriate and attractive to potential customers.

Right products → Right Customer → Right Time
Agency Medical Product Upsell

Upsell additional medical coverage to in-force policyholders

Objectives

- Increase sales by upselling new medical products to in-force customers & reduce underwriting requirements for the best risks

Data

- 16 years of policyholder & claims information
- > 3m base policies & claims & > 4m riders

Business Application

- 14 variables identified beyond age/gender included location, income, occupation, rider count, duration since claim, payment method etc.
- Upsell new product to 50% of in-force customers with one UW question on a pre-approved basis

Loss Ratio by Vigintile

Approximate results have been presented to protect client results and confidentiality
RGA Claims Upsell

- Upselling to existing customers is important in most markets with the majority of sales coming from existing customers.
- Claimants are considered warm leads for additional sales from an agents perspective however they're afraid to contact them due to the arduous underwriting process and risk of decline.
- RGA generates leads and products to be sold using historic medical claims and existing underwriting results.
Case Studies:
Other interesting projects
Non-Medical Limit Risk Segmentation

Objectives

- Determine optimal non-medical limits to be used for different customer segments
- Streamline the underwriting process for the best risks

Data

- Insurance company policyholder, agent and underwriting data

<table>
<thead>
<tr>
<th>Study Period</th>
<th>Product</th>
<th>Total Exposure</th>
<th>Total Claims</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 years</td>
<td>Life &amp; Accelerated CI</td>
<td>~7m life years</td>
<td>~10,000</td>
</tr>
</tbody>
</table>

Business Application and Results

- Justify the requirements for non-medical limits to mitigate risk in the target markets
- Increase sales volume with higher non-medical limits at controlled risk levels for good risks

Approximate results have been presented to protect client results and confidentiality
# Unstructured Data Processing

## Objectives
- Collect and processing data from .pdf images (unstructured data) into a structured format (table)
- Help RGA UW department increase efficiency of facultative underwriting process

## Data and Solutions
- Facultative underwriting documents from UW department:
  - ~ 1,000 UW cases in .pdf format (some cases contain thousands of pages)
  - Application forms, Application questionnaire, lab reports, MVR, Rx, APS etc.
- Customized machine learning solution including image process and language processing

## Applications
- Extract information from 100+ data fields: Name, DOB, Blood Pressure, Occupation etc.
- Automatically fill information into UW worksheet
- Fast page locator
- Augment into existing underwriting processed (consistency check, lab test check, page sorting)
Final Thoughts
Future view on data science in life insurance

Factors for growth

- Increasing volumes of quality data and data products available
- Global demand for personalized offerings and ease of transactions
- Growth of direct to consumer offerings
- Monetization of data assets

Head winds

- Major financial successes yet to be demonstrated
- Effort in data cleaning, manipulation, and modelling
- More onerous data protection legislation (explicit consent, profiling)
- Cyber risk – Risks of data being lost, corrupted, or stolen
Questions?