Session 100PD, Predictive Models for Disability Insurance

Presenters:
Jeffrey S. Bowden, FSA, MAAA
Mark J. Costello, FSA, MAAA
Predictive Modeling for Disability Insurance:

Jeff Bowden - Unum
June 2018
Predictive Modeling for Disability Insurance

**Tools**
- Software: Open Source and Proprietary Software

**Talent**
- What is the role of Actuaries and Data Scientists in analytics?

**Deployment**
- The most successful analytics applications involve solutions that:
  - Integrate with existing business tools
  - Provide relevant guidance in real time

**Application Example**
- Probability of Sale Model to drive more efficient quoting and underwriting
Predictive Modeling Tools

- **Open source is serious business**
  - Often free or low cost (but with license obligations and restrictions)
  - Many big name company sponsors (Google, Microsoft)
  - New features/algorithms added much faster than proprietary software

- **Very Active Community**
  - Many great learning resources online
  - Rich online sharing
  - Mature and organized distribution system

- **Commonly integrated by many proprietary software**
  - Selling ease of use
  - In-database analytics also available

- **Access to software for everyone**
  - Anyone can do modeling
  - Be sure you understand the licensing requirements
  - Need to choose a tool set and exploit it
    - Modeling software
    - UI and documentation software
### Open Source Tools

<table>
<thead>
<tr>
<th>Tool</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Python** (<https://www.python.org/>) | - Created in 1991 as a general programming language  
- One of the top languages for data science  
- Considered one of the easiest |
| **R** (<https://www.r-project.org>) | - Language built for statistics  
- Very large academic following  
- New actuarial exams being designed for R |
| **Jupyter Notebook** (<http://jupyter.org/>) | - "Open-source web application allowing you to create and share documents that contain live code, equations, visualizations and explanatory text"  
- Works with 40+ programming languages |
| **Scikit-Learn** (<http://scikit-learn.org/>) | - State-of-the-art machine learning algorithms  
- High ease of use  
- Great documentation/learning resources |
| **Tensorflow** (<https://www.tensorflow.org/>) | - Deep learning software library for Python  
- Developed by researchers and engineers working on the Google Brain Team  
- Used by many large companies (Google, Twitter, Uber, eBay, etc..) |
| **XGBoost** (<https://xgboost.readthedocs.io>) | - Deep Learning software library for Python |

- Levels the playing field in terms of technology capabilities
- Open source capabilities are advanced and growing, need the talent to exploit
- Data Scientists are being trained with these tools and expect to use them
Predictive Modeling Talent

- Powerful and easy to use open source tools allow actuaries and other technical resources to perform advanced modeling work
- Data and business knowledge become the differentiators

For traditional structured data problems actuaries are a great fit

Strong business knowledge is a major advantage

Flexibility to perform in multiple roles

Incorporating open source analytics tools in to curriculum

Have specialized skills to solve more complex problems involving unstructured data
  - Image recognition
  - Text Mining

High awareness of new tools and technologies
Predictive Modeling Deployment

- Successful deployment involves seamless integration with existing business processes
  - Recommendations pushed to the user at the exact time that they need it

Scoring: Live Vs Batch
Application: Integrated Vs Separate

- Separate tool
  - Simple for the modeler
  - May actually make the users job more difficult

- Batch scoring process
  - Can make integration easier

- True live deployment (live scoring)
  - Modern platforms (ex: cloud)
  - Strong IT engagement

Expectations set by consumer experience:

Customers Also Bought These Albums

- The Greatest Hits [Explicit]
  - Mötley Crüe
  - Stream or buy for $11.49

- The Essential Judas Priest
  - Judas Priest
  - Stream or buy for $14.99

- Destroyer
  - Kiss
  - Stream or buy for $5.99
Deployment typically follows a path to implementation

- With each step the investment increases - IT Resources, Ongoing Support etc.
- Staged implementation allows results to be measured and business value to be validated
Analytics Application:

Probability of Sale Model
Business requested a tool to classify quotes by their likelihood of resulting in a sale

- To be effective, this classification must be done before running the quote – rates cannot be a factor
- Ideally, classification is done immediately after the quote is received to optimize work assignment and prioritization

**Probability of Sale Model**

**BACKGROUND**

- For small case business, faster turn around time (TAT) is known to contribute to improved close ratio
- During peak times, it can be a challenge to meet broker requested TAT for all quotes

**Analytics**
When quote requests are received, basic information is entered into our customer relationship management (CRM) tool.

Information available about the quote at this time is very limited.

To improve model performance additional engineered features were added:

- Designed with underwriting and actuarial input to supplement the existing variables.

Model developed in Python using XGBoost module:

- Decision tree based model – available through both Python and R
- New models incrementally ensembled to maximize performance
- Additional models predict the residuals or error of previous models and offset
- Effective for both regression and classification

“When in doubt, use XGBoost.”

-- Owen Zhang, Kaggle Competition Winner
The model results were validated against a test set of recent quotes that confirmed it was very effective at identifying the quotes with the highest and lowest likelihood of sale.

Quotes were sorted by predicted close ratio and grouped into deciles:

- Actual close ratios closely followed the predicted close ratios very closely
- Model was successfully able to classify quotes by probability of sale with very limited information

Quotes can be immediately classified into 3 groups by their likelihood of sale:

- **Low**
- **Average**
- **High**
Providing the probability of sale to underwriters can improve efficiency in several ways:

- **Capacity Management**: When quote volume exceeds capacity, prioritize efforts on high probability of sale quotes.
- **Work Assignment**: Most experienced workers assigned to high potential quotes.
- **Workflow Management**: Move from first in first out (FIFO) to queuing work based on a combination of factors.

- For any of these to work, information has to be provided near real time and be integrated within their existing work systems:
  - Work assignment is completed shortly after the initial quote information is entered.
  - Worklists for each employee contain due dates and other pertinent information.
  - Need to have the probability of sale information in this same location.
Probability of Sale Model

DEPLOYMENT

- Quote data stored in data warehouse tables that are updated every 24 hours
  - Daily Batch Scoring would be far too slow to be useful
- Worklists are generated within the CRM system - needs to have access to the model results

<table>
<thead>
<tr>
<th>Quote ID</th>
<th>Date Received</th>
<th>Date Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>5/14 9:54</td>
<td>5/17</td>
</tr>
<tr>
<td>456</td>
<td>5/14 10:18</td>
<td>5/19</td>
</tr>
<tr>
<td>789</td>
<td>5/14 11:05</td>
<td>5/17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Probability of Sale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
</tr>
<tr>
<td>Avg</td>
</tr>
<tr>
<td>High</td>
</tr>
</tbody>
</table>

Solution: API (Application Programming Interface)

- APIs allow different applications to easily pass information back and forth
- Allow virtually real time scoring with results pushed back to existing tools

- Allows evaluation of a live scoring prototype with minimal IT investment
Open source tools offer leading edge analytics capabilities at little cost:

- These are the tools that data scientists want to use and they are growing fast
- Actuaries, Computer Scientists and others can quickly develop advanced analytic skills with these tools
- Regardless of the role, business understanding is more valuable than ever in model development

Users have become accustomed to and expect seamless integration of analytics:
- Integration with business operations is critical for acceptance and utilization
- Traditional data warehouse structures may not be adequate to provide this – explore alternatives
Business Case Studies for Analytics

June 25 2018
Mark Costello
Agenda

1. Underwriting paradigm
   - Individual life examples
   - Group life
2. LTD claim scoring
3. Depression marker
4. Disability / economic cycle
5. Other
New tools need to integrate into the underwriting process

available risk assessment tools / data:

- Insurance History
- Prescriptions
- Driving Record
- Credit
- Electronic health records
- Lifestyle / Social
- Wearables

The process includes:

1. Application & Tele-Interview
2. Predictive models
   - Rules-based Automated UW
   - Manual UW
3. Risk Class

traditional data:

- Medical Lab Results
- Attending Physician Statement
- Income & financial info
### Smoker prediction

<table>
<thead>
<tr>
<th>Area</th>
<th>Predictor</th>
<th>Direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affluence</td>
<td>Face amount</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td>Home value</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td>Client issue age - older</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td>Investment assets</td>
<td>▼</td>
</tr>
<tr>
<td>Lifestyle</td>
<td>Catalog stationery buyer</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td>Healthy behavior change index</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td>Survey Rock-n-roll music</td>
<td>▲</td>
</tr>
<tr>
<td></td>
<td>Survey travel casino vacations</td>
<td>▲</td>
</tr>
<tr>
<td></td>
<td>Survey lotteries or sweepstakes</td>
<td>▲</td>
</tr>
<tr>
<td></td>
<td>Survey diet concerns</td>
<td>▼</td>
</tr>
<tr>
<td></td>
<td>Dog owner</td>
<td>▲</td>
</tr>
<tr>
<td>Location</td>
<td>Counties, grouped (i.e. Orange CA, Westchester NY, Salt Lake UT, Fulton GA)</td>
<td>▼</td>
</tr>
</tbody>
</table>
Predictive analytics for risk selection

Objective: manage incremental mortality risk in a non-fluid program by using a predictive model to identify the best fitting class for the case

Model: predict the likelihood of actual UW class
- Preferred/Super-preferred
- Standard
- Substandard/Other
- Declined

Data:
- Application
- MIB, MVR, Rx

Technique:
- Multinomial logistic regression
Physical activity and mortality

Objective: assess the effectiveness of physical activity in stratifying the mortality risk profile of a U.S. population-based dataset provided by Vivametrica

Vivametrica data:
- Clinical research studies 1988-2004
- Focused on measuring the relationship between lifestyle behaviors and health outcomes
- Subject to population-based sampling methods, ensuring it is representative of the US population

Technique:
- Classical actuarial mortality analysis
- Survival analysis/machine learning

Key findings:
- Physical activity as measured by steps per day effectively stratifies mortality risk.
- Steps per day is an important predictor of mortality risk, and may be especially effective in identifying high mortality risk for sedentary behavior.
- Steps per day provides additional segmentation of mortality even after considering traditional underwriting attributes, such as smoking status, BMI, cholesterol, blood pressure and health history of diabetes, cardiovascular disease and cancer.
Enhanced rating factors: industry

- Reflects industry related risk factors (salary, education, smoking, obesity, work environment, catastrophic risk)

Enhanced rating factors: salary

- Mortality and mortality improvement varies by salary
- Mortality differential by salary within industries
LTD claims scoring

Models and data

- The scores are calculated using a predictive model that is calibrated to client’s actual claim history.
- Various models and techniques are used to predict return to work, including natural language processing, statistical techniques and machine learning.
- Data includes structured and unstructured sources.

<table>
<thead>
<tr>
<th>Claims attributes</th>
<th>Diagnosis</th>
<th>Geospatial</th>
<th>Case management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>Diagnostic group</td>
<td>Area attributes</td>
<td>APS</td>
</tr>
<tr>
<td>Gender</td>
<td>Primary diagnosis</td>
<td>Population attributes</td>
<td>Claim manager notes</td>
</tr>
<tr>
<td>Earnings</td>
<td>Secondary diagnosis</td>
<td>Housing attributes</td>
<td></td>
</tr>
<tr>
<td>Gross benefit</td>
<td></td>
<td>Household attributes</td>
<td></td>
</tr>
<tr>
<td>STD integration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Industry</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elimination period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reporting lag</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© 2018 Munich American Reassurance Company. All Rights Reserved 7/5/2018
Development of a depression-score with mobile data

\[ Y_{ij} = \gamma_{00} + \sum_{k=1}^{n} \gamma_{k0} X_{ij}^k + \zeta_{0i} + \sum_{k=1}^{n} \zeta_{ki} X_{ij}^k + \varepsilon_{ij} \]

Step count per minute (Locomotor activity data) → Raw Depression Score → Clean Depression Score → Potential Depression Moment

© 2018 Munich American Reassurance Company. All Rights Reserved
Incidence rates vs. economic cycles

No apparent correlation with economic cycles

Some possible correlations in Australia and the US.
Cross-selling with predictive analytics

Model-driven marketing and underwriting: emerging global practice

1) Data integration
- Customer and policy data
- Partner data
- Appended data

risk & marketing profile

2) Rules & scoring automation
- Propensity to buy
- Next best offer
- Predictive UW

3) Offer presentation

web / call center

insured
agent

1) Data integration
2) Rules & scoring automation
3) Offer presentation