Society of Actuaries in Ireland

Predictive modelling around the world

28.11.13
Agenda

- Why this presentation is really interesting
- Introduction to predictive modelling
- Case studies
- Conclusions
Why this presentation is really interesting

Presentations that appeal to geeky actuaries

Presentations that appeal to marketing/product development actuaries

This presentation
<table>
<thead>
<tr>
<th>ANALYTICS</th>
<th>PREDICTIVE UNDERWRITING</th>
<th>THE CLOUD</th>
</tr>
</thead>
<tbody>
<tr>
<td>OPEN SOURCE TECHNOLOGY</td>
<td>BIG DATA</td>
<td>DATA MINING</td>
</tr>
<tr>
<td>MACHINE LEARNING</td>
<td>GLMs</td>
<td>CROWDSOURCING</td>
</tr>
</tbody>
</table>
Agenda

• Why this presentation is really interesting
• Introduction to predictive modelling
• Case studies
• Conclusions
What are predictive analytics?

Tools and technologies for analysing and understanding business performance

Extensive use of data with statistical and quantitative analysis

A process by which current or historical facts are used to create predictions about future events or behaviour.
What are predictive analytics?

A process, not a product

Good data is vital for success

A **process** by which current or historical **facts** are used to create **predictions** about **future events or behaviour**.

Typically predictions are created through the use of sophisticated statistical models

Focus on predicting probability of future events and behaviour
The process

- Deployment
- Business Understanding
- Data Understanding
- Evaluation
- Modelling
- Data Preparation

Source: adapted from Cross Industry Standard Process for Data Mining (CRISP-DM)
Underlying the models

Operational Research
- Simulation
- Optimisation
- Simulated Annealing
- K-Means Clustering
- Genetic Algorithms
- Graph Theory
- Link Analysis
- Decision Trees
- Random Forest
- Support Vector Machines
- Visualisation
- Querying
- OLAP
- Cross-tabs
- SQL

Data Mining
- BI
- Neural Networks
- Fourier Transforms
- Wavelets
- Linear, Logistic Regression, GLMs
- Time-series Analysis
- Bayesian Networks
- Monte Carlo
- Principle Components
- Harmonic Analysis
- Reliability/Survival Analysis
- ANOVA
- MANOVA
- Correlation
- Factor Analysis

Forecasting

Statistics

Why predictive analytics?

Data
- internal data
- external data
- new types of data

Technology
- Processing power
- New tools

Competition
Agenda

- Why this presentation is really interesting
- Introduction to predictive modelling
- Case studies
- Conclusions
Worldwide projects*

UK:
• Basis Setting (mortality, morbidity and lapses)
• Postcode pricing model
• Enhanced experience analysis
• Predictive underwriting on credit rating agency and bank data
• Broker Quality

Europe:
• Predictive underwriting on bancassurance data

South Africa:
• Enhanced Experience Analysis
• Predictive underwriting on bank and credit card data

India:
• Claims Fraud Prediction

Australia:
• Predictive underwriting / cross sell on bancassurance data

USA:
• Pricing override model for group LT disability
• Lapse basis
• Predictive underwriting on Non-Life data
• Term Tail Lapses
• Mortality prediction on credit rating agency data

Asia:
• Predictive underwriting on bancassurance data
• Finer price segmentation
• Propensity to buy
• Cross sell of insurance on bank data

* Courtesy of Peter Banthorpe RGA
Case Study 1: Lapse Assumptions variable annuity with GMIBs

• Current lapse modeled based solely on plan type, duration, In-the-moneyness (“ITM”)
• Proposal: Evaluate current model based on 12 quarters of observations
• Method: Develop an alternative statistical model based on current variables, augmented by additional policy characteristics and macroeconomic variables
• Compare predictive performance of the two models
  – Performance measured by ratio of Actual/Expected dollars lapsed in out-of-sample future period.
### Predictors

<table>
<thead>
<tr>
<th>Positively associated with lapse</th>
<th>Negatively associated with lapse</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Duration (adjusted for Surrender Charge Period)</td>
<td>• Annuitant age at issue</td>
</tr>
<tr>
<td>• Anniversary of issue date</td>
<td>• Total rider charge (bps)</td>
</tr>
<tr>
<td>• Policies sold in channel 1, 5, 6</td>
<td>• In-the-moneyness (ITM)</td>
</tr>
<tr>
<td>• State unemployment rate</td>
<td>• Held in qualified plan?</td>
</tr>
<tr>
<td>• US inflation rate</td>
<td>• Policy size (cumulative deposits)</td>
</tr>
<tr>
<td>• Risk aversion (low % of equity)</td>
<td>• Policies sold in channels 3, 8</td>
</tr>
<tr>
<td>• Past partial withdrawals</td>
<td></td>
</tr>
</tbody>
</table>

**Not significant:** Gender, Surrender Charge, DB Rider, GMIB reset, past quarter S&P 500 return, 10 Yr T Note yield
Performance of predictors

- At each quarter, split data by past versus current experience
  - Actual experience (black bars)
  - Statistically predicted (red dots) expected lapse and confidence bands
  - Calculated (blue dots) expected lapse based on current assumption
- Statistical A/E outperforms current assumption in all quarters
Case Study 2: Predictive Underwriting Model

• Client: Bancassurer in Asia with large customer pool, but low penetration in life product
• Goal: to predict UW decisions on its existing customers
• Major challenges - very limited data
  – A total of about 8k-9k full UW cases
  – Target variable UW decision, with very low declined/rated cases, ~3.0%
  – Many missing values due to old time, especially for sub-STD
  – Not all information collected at the time of UW

Source: Peter Banthorpe RGA.
Modeling Approach / Key Variables

• GLM with binomial and logistic link function
• About a dozen of predictor variables that are statistically significant for prediction & readily available in client database
• Key predictor variables
  – “Positive” means the probability of being a standard rate case increases if the value goes up; otherwise, it is ”Negative”

<table>
<thead>
<tr>
<th>Name</th>
<th>Type</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age_At_Entry</td>
<td>Numeric</td>
<td>Negative; less likely to qualify for STD as age goes up</td>
</tr>
<tr>
<td>Branch</td>
<td>Categorical</td>
<td>Proxy of geographic locations</td>
</tr>
<tr>
<td>Asset Under Management</td>
<td>Numeric</td>
<td>Positive; more likely to qualify for STD with large AUM</td>
</tr>
<tr>
<td>Customer_Segment</td>
<td>Categorical</td>
<td>Positive for “Gold”, negative for other</td>
</tr>
<tr>
<td>Nationality</td>
<td>Categorical</td>
<td>Positive for domestic; negative for certain others</td>
</tr>
</tbody>
</table>

Source: Peter Banthorpe RGA.
Model Results – Lift Plots

- In-sample results show model performance under optimal condition
- May over-fit data
- 0.5% of sub-STD in top 30%

- Validation results are a better test of model performance in real business
- 0.6% sub-STD in the top 30%

Source: Peter Banthorpe RGA.
Model Results – Gain Curve

- Another way to understand model capability to differentiate STD from sub-STD
  - Best 30% of model outputs contains about 5% of total non-STD
  - Lowest 30% captures about 75% of bad risks

Source: Peter Banthorpe RGA.
Case study 3: non-life

- Predictive analytic techniques used to analyse motor insurance portfolio to:
  - Identify predictors of claims and hence model a profitability score per customer
  - Identify predictors of propensity to renew
- Allows analysis of portfolio by profitability and propensity to renew
## Results

### High-profit, low-retention customer segment in red

<table>
<thead>
<tr>
<th>Loss Ratio</th>
<th>98.5%</th>
<th>97.7%</th>
<th>96.7%</th>
<th>95.8%</th>
<th>94.5%</th>
<th>91.5%</th>
<th>87.3%</th>
<th>80.8%</th>
<th>Exposure</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>35.0%</td>
<td>3.0%</td>
<td>1.6%</td>
<td>1.8%</td>
<td>0.9%</td>
<td>0.4%</td>
<td>0.1%</td>
<td>0.4%</td>
<td>0.5%</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td>42.7%</td>
<td>3.6%</td>
<td>2.2%</td>
<td>1.8%</td>
<td>0.8%</td>
<td>0.7%</td>
<td>0.3%</td>
<td>0.7%</td>
<td>0.6%</td>
<td>11%</td>
<td>16%</td>
</tr>
<tr>
<td>45.8%</td>
<td>2.0%</td>
<td>3.0%</td>
<td>3.5%</td>
<td>2.4%</td>
<td>1.6%</td>
<td>1.1%</td>
<td>1.2%</td>
<td>0.5%</td>
<td>15%</td>
<td>16%</td>
</tr>
<tr>
<td>49.5%</td>
<td>0.7%</td>
<td>2.0%</td>
<td>3.6%</td>
<td>3.4%</td>
<td>2.6%</td>
<td>1.9%</td>
<td>1.2%</td>
<td>0.6%</td>
<td>16%</td>
<td>16%</td>
</tr>
<tr>
<td>55.5%</td>
<td>0.3%</td>
<td>1.0%</td>
<td>2.6%</td>
<td>3.4%</td>
<td>3.7%</td>
<td>2.7%</td>
<td>1.5%</td>
<td>0.7%</td>
<td>16%</td>
<td>16%</td>
</tr>
<tr>
<td>58.2%</td>
<td>0.0%</td>
<td>0.4%</td>
<td>1.5%</td>
<td>2.7%</td>
<td>3.6%</td>
<td>4.2%</td>
<td>2.0%</td>
<td>0.8%</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>61.4%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.5%</td>
<td>1.4%</td>
<td>2.0%</td>
<td>2.9%</td>
<td>2.0%</td>
<td>1.0%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>75.4%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.7%</td>
<td>1.2%</td>
<td>2.7%</td>
<td>1.5%</td>
<td>2.0%</td>
<td>8%</td>
<td>8%</td>
</tr>
<tr>
<td>Exposure Total</td>
<td>10%</td>
<td>10%</td>
<td>15%</td>
<td>16%</td>
<td>16%</td>
<td>11%</td>
<td>7%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: EagleEye Analytics

### Low-profit, high-retention customer segment in red

<table>
<thead>
<tr>
<th>Loss Ratio</th>
<th>98.5%</th>
<th>97.7%</th>
<th>96.7%</th>
<th>95.8%</th>
<th>94.5%</th>
<th>91.5%</th>
<th>87.3%</th>
<th>80.8%</th>
<th>Exposure</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>35.0%</td>
<td>3.0%</td>
<td>1.6%</td>
<td>1.8%</td>
<td>0.9%</td>
<td>0.4%</td>
<td>0.1%</td>
<td>0.4%</td>
<td>0.5%</td>
<td>9%</td>
<td>11%</td>
</tr>
<tr>
<td>42.7%</td>
<td>3.6%</td>
<td>2.2%</td>
<td>1.8%</td>
<td>0.8%</td>
<td>0.7%</td>
<td>0.3%</td>
<td>0.7%</td>
<td>0.6%</td>
<td>11%</td>
<td>16%</td>
</tr>
<tr>
<td>45.8%</td>
<td>2.0%</td>
<td>3.0%</td>
<td>3.5%</td>
<td>2.4%</td>
<td>1.6%</td>
<td>1.1%</td>
<td>1.2%</td>
<td>0.5%</td>
<td>15%</td>
<td>16%</td>
</tr>
<tr>
<td>49.5%</td>
<td>0.7%</td>
<td>2.0%</td>
<td>3.6%</td>
<td>3.4%</td>
<td>2.6%</td>
<td>1.9%</td>
<td>1.2%</td>
<td>0.6%</td>
<td>16%</td>
<td>16%</td>
</tr>
<tr>
<td>55.5%</td>
<td>0.3%</td>
<td>1.0%</td>
<td>2.6%</td>
<td>3.4%</td>
<td>3.7%</td>
<td>2.7%</td>
<td>1.5%</td>
<td>0.7%</td>
<td>16%</td>
<td>16%</td>
</tr>
<tr>
<td>58.2%</td>
<td>0.0%</td>
<td>0.4%</td>
<td>1.5%</td>
<td>2.7%</td>
<td>3.6%</td>
<td>4.2%</td>
<td>2.0%</td>
<td>0.8%</td>
<td>15%</td>
<td>15%</td>
</tr>
<tr>
<td>61.4%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.5%</td>
<td>1.4%</td>
<td>2.0%</td>
<td>2.9%</td>
<td>2.0%</td>
<td>1.0%</td>
<td>10%</td>
<td>10%</td>
</tr>
<tr>
<td>75.4%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.1%</td>
<td>0.7%</td>
<td>1.2%</td>
<td>2.7%</td>
<td>1.5%</td>
<td>2.0%</td>
<td>8%</td>
<td>8%</td>
</tr>
<tr>
<td>Exposure Total</td>
<td>10%</td>
<td>10%</td>
<td>15%</td>
<td>16%</td>
<td>16%</td>
<td>11%</td>
<td>7%</td>
<td>100%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: EagleEye Analytics
Agenda

• Why this presentation is really interesting
• Introduction to predictive modelling
• Case studies
• Conclusions
Key messages

• Data driven process
• Broad potential applications for insurance
• Non-Life is way ahead of us
• Simplified underwriting a key area of focus (but there are many more applications)
• Not an off-the shelf solution
  – Customised, based on specific data and specific needs
  – No two exercises are the same – flexibility of approach is key