Use of Big Data by AXA Direct Korea
Topics for today...

1. Data Culture in AXA Direct Korea
   How is Data contributing to improve business results across the company?

2. Use of Machine Learning for Underwriting
   Machine Learning vs. traditional actuarial techniques for risk modeling: both have their use

3. The current challenge: CVM
   How Machine Learning can help understand the customer value, and help design the company Underwriting strategy
Introduction of Data culture in AXA Direct Korea
Key Components of Data Culture in AXA

1. **Analytical Infrastructure**
   - Dedicated analytical infrastructure
   - Supply real-time data analysis
   - Easy to access analytical infrastructure

2. **Skilled business analysts**
   - Business Analyst with data analysis skills across the company
   - Supported by an independent Data Science team

3. **Focus on practical business applications**
   - Analytical function need to be in line with business requests
   - Focus is on practical, not ideal, improvement

---

**Maximize Company Profits**

- Analytical Infrastructure
- Skilled business analysts
- Focus on practical business applications

**Right Business Analyst**

- Analytical infrastructure
- Skilled business analysts
- Focus on practical business applications
DATA as a business enabler – focus on practical business issues
An example from Claims: Fraud Detection System

Data source

Fraud detection system

Action on biz process

Real-time

Hybrid fraud detection engine

Predictive model + Rules

- The insured base
- The policyholder base
- Business rules
- Anomaly detection

Network analysis

- Network group
- Multi-linker
- Network linker

Alert

RED: Highly suspicious
ORANGE: Middle
YELLOW: Low

Investigation point / Suspicious fraud type

Alert

Information feeding by FDS

Claim Staff

SIU Staff

Decision by FDS

Claim Staff ability

Suspicious

General

Normal Process

Investigation
Improving the FDS (1) : Text Mining

**Concept**

Discover meaningful information from unstructured data!

**Analytic flow**

1. **Data integration**
2. **Parsing**
3. **Analysis**

- Integrate claim description data and fraud information
- Syntax parsing (using R package)
- Conduct text mining and statistical analysis

**Word cloud**

Falling Down  
Stairs  
Minor collision  
Sliding  
Bicycle

**Frequency**
Improving the FDS (2): Network Analysis

Fraud network type

“Staged” accident

Organized Fraud Network

Graph Data of Networks
Machine Learning + Text Mining + Network Analysis = Strong Improvement in Model Accuracy

- Diversified Predictive model + Rules
- Text Mining + Network Analysis

Significant performance enhancement

Detection rate of Policyholder model (Ridge reg.): 6.4% Previous model, 11.0% New model

Detection rate of Insured model (Neural network): 5.5% Previous model, 11.4% New model

SOCIETY OF ACTUARIES
An example from Sales: Outbound Calls Scoring

Customer Acquisition

1. O/B CR Score models
Various models to Identify the customers with higher conversion rate

2. Unpaid scoring models
Identify the customers who need to be contacted in case of payment delay

Customer Relationship

3. Churn prediction models
Identify the customers with higher churn tendency

Loyalty

4. O/B RR Score model (1)
Identify the customers with higher renewal tendency
Supporting Sales: Cross-selling model for A&H

Prediction model developed by Machine Learning

**Model 1**  
**Driver - Cross-sell Model**

- **Target**: A&H Driver insurance  
- **Modeling method**: Stabilized deep net  
- **Input variable (18)**:  
  - Main variables in the models: Credit-ratings  
  - Single most powerful variable: Previous company / Insured age / First quotation Channel

**Model result**

<table>
<thead>
<tr>
<th>Group</th>
<th>Cross-sell rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group1</td>
<td>22.3%</td>
</tr>
<tr>
<td>Group2</td>
<td>18.0%</td>
</tr>
<tr>
<td>Group3</td>
<td>17.3%</td>
</tr>
<tr>
<td>Group4</td>
<td>16.0%</td>
</tr>
<tr>
<td>Group5</td>
<td>14.7%</td>
</tr>
<tr>
<td>Group6</td>
<td>13.0%</td>
</tr>
<tr>
<td>Group7</td>
<td>9.8%</td>
</tr>
<tr>
<td>Group8</td>
<td>6.0%</td>
</tr>
<tr>
<td>Group9</td>
<td>3.8%</td>
</tr>
<tr>
<td>Group10</td>
<td>2.7%</td>
</tr>
</tbody>
</table>

**Avg. Cross-sell Rate**: 11.9%  
[**Test**: 2016.11 ~ 2016.12]

**Model 2**  
**Other products - Cross-sell Model**

- **Target**: A&H insurance (other than driver)  
- **Modeling method**: Stabilized deep net  
- **Input variable (19)**:  
  - Main variables in the models: Credit-ratings  
  - Single most powerful variable: Previous company / Nb of accidents / Areas

**Model result**

<table>
<thead>
<tr>
<th>Group</th>
<th>Cross-sell rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group1</td>
<td>3.3%</td>
</tr>
<tr>
<td>Group2</td>
<td>2.5%</td>
</tr>
<tr>
<td>Group3</td>
<td>1.8%</td>
</tr>
<tr>
<td>Group4</td>
<td>1.7%</td>
</tr>
<tr>
<td>Group5</td>
<td>1.5%</td>
</tr>
<tr>
<td>Group6</td>
<td>1.3%</td>
</tr>
<tr>
<td>Group7</td>
<td>1.1%</td>
</tr>
<tr>
<td>Group8</td>
<td>0.9%</td>
</tr>
<tr>
<td>Group9</td>
<td>0.8%</td>
</tr>
<tr>
<td>Group10</td>
<td>0.4%</td>
</tr>
</tbody>
</table>

**Avg. Cross-sell Rate**: 1.5%  
[**Test**: 2016.11 ~ 2016.12]
Use of Machine Learning for Underwriting
Technical Sophistication in a regulated market

- **Motor Insurance remains heavily regulated in Korea**
  - Not all variables can be used for pricing
  - Pricing changes must be justified

- **Despite regulation, the market is very active and the anti-selection process happens very fast** – the Conversion Rate of unprofitable segment is very high

- **To improve Technical results, ADK has invested significant time and resources into developing strong BC model** – but these cannot be used directly in Pricing

---

**Unconstraint Model**
- The “Ideal” model
- Based on all available variables

**Constraint Model**
- Based on variables available for Pricing

**Underwriting Scoring Model**
- Based on all variables available at quotation

---

*Identify powerful variables, compute ELR*

*Introduce in Pricing*

*Decision to reject / underwrite*

---

*Today’s focus*
Underwriting Scoring in Korea – concept

- The underwriting scoring provide a “Score” to all quotations and renewals, equivalent to the expected Loss Ratio

- Due to the pricing regulations, the score distribution is very large

- Set an Underwriting Threshold, the ELR above which contracts should be rejected.
  - This threshold can be adjusted based on the strategy (priority on growth or profit)
  - The threshold can be set differently depending on generations (NB/RNW) or other Characteristics…

Adjustable threshold

In this example, contracts with an ELR above 100% would be rejected
U/W scoring – Implementation and Process

- **Needs a real-time Score (ELR) computation process**
  - Link between IT and BC model - which requires a very flexible IT system
  - Real-Time ELR computation based on many variables (incl. external variables)

- **Works best when linked to Sales Representatives remuneration**

<table>
<thead>
<tr>
<th>Definition</th>
<th>Portion</th>
<th>Payment rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Unprofitable customers</strong></td>
<td>Appr. 3%</td>
<td>0.31%</td>
</tr>
<tr>
<td><strong>Standard customers</strong></td>
<td>Appr. 57%</td>
<td>1.05%</td>
</tr>
<tr>
<td><strong>Very profitable customers</strong></td>
<td>Appr. 40%</td>
<td>1.15%</td>
</tr>
</tbody>
</table>

% of Green Customers:
- UY’16 Q1: 39.7%
- UY’16 Q2: 40.6%
- UY’16 Q3: 42.2%
- UY’16 Q4: 44.1%
- UY’17 Q1: 49.0%
- UY’17 Q2: 51.1%
- UY’17 Q3: 50.6%
U/W scoring – Use of Machine Learning

The main issue with using GLM are the time needed to update the models, and therefore the risk of being too slow to cope with anti-selection in very dynamic markets.

Since 2017, ADK is using Machine Learning to update the UW Scoring
- GLM is very powerful for pricing, thanks to the interpretation power
- Machine Learning can be more accurate (but more difficult to understand and interpret)
- Machine Learning algorithm can be retrained very quickly and regularly

- **Human modelisation (GLM)**
  - 1 to 2 times per year
  - Accurate, very insightful

- **Models through Machine Learning**
  - Whenever necessary
  - Gives a choice of various algorithms to choose from
  - Very accurate but less insightful
  - Directly linked to IT Underwriting System

Extensive Dataset (800 variables)
Cross-validation
GLM for Pricing
Machine Learning for Underwriting
**New Business**

<table>
<thead>
<tr>
<th>Cover</th>
<th>Type</th>
<th>Old ML Gini</th>
<th>New ML Gini</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPBI</td>
<td>Freq.</td>
<td>26.2%</td>
<td>28.5%</td>
<td>2.3%</td>
</tr>
<tr>
<td></td>
<td>Sev.</td>
<td>9.8%</td>
<td>12.9%</td>
<td>3.1%</td>
</tr>
<tr>
<td>TPPD</td>
<td>Freq.</td>
<td>26.3%</td>
<td>30.1%</td>
<td>3.8%</td>
</tr>
<tr>
<td></td>
<td>Sev.</td>
<td>10.0%</td>
<td>20.0%</td>
<td>10.0%</td>
</tr>
<tr>
<td>OCD</td>
<td>Freq.</td>
<td>23.0%</td>
<td>26.4%</td>
<td>3.4%</td>
</tr>
<tr>
<td></td>
<td>Sev.</td>
<td>36.8%</td>
<td>38.3%</td>
<td>1.5%</td>
</tr>
<tr>
<td>OBI</td>
<td>Freq.</td>
<td>32.6%</td>
<td>28.7%</td>
<td>-0.39%</td>
</tr>
<tr>
<td></td>
<td>Sev.</td>
<td>39.9%</td>
<td>29.7%</td>
<td>-10.2%</td>
</tr>
</tbody>
</table>

**Selected Algorithm**

<table>
<thead>
<tr>
<th></th>
<th>Freq.</th>
<th>Sev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPBI</td>
<td>Deep Net</td>
<td>Neural Network</td>
</tr>
<tr>
<td>TPPD</td>
<td>Deep Net</td>
<td>AXA Custom</td>
</tr>
<tr>
<td>OCD</td>
<td>AXA Custom</td>
<td>AXA Custom</td>
</tr>
<tr>
<td>OBI</td>
<td>GBM</td>
<td>GBM</td>
</tr>
</tbody>
</table>

**Renewal**

<table>
<thead>
<tr>
<th>Cover</th>
<th>Type</th>
<th>GLM Gini</th>
<th>ML Gini</th>
<th>Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPBI</td>
<td>Freq.</td>
<td>26.5%</td>
<td>27.9%</td>
<td>1.4%</td>
</tr>
<tr>
<td></td>
<td>Sev.</td>
<td>5.1%</td>
<td>10.9%</td>
<td>5.8%</td>
</tr>
<tr>
<td>TPPD</td>
<td>Freq.</td>
<td>23.6%</td>
<td>27.6%</td>
<td>4.0%</td>
</tr>
<tr>
<td></td>
<td>Sev.</td>
<td>8.2%</td>
<td>17.1%</td>
<td>8.9%</td>
</tr>
<tr>
<td>OCD</td>
<td>Freq.</td>
<td>OCD1 22.8%</td>
<td>OCD2 24.8%</td>
<td>26.1%</td>
</tr>
<tr>
<td></td>
<td>Sev.</td>
<td>OCD1 20.0%</td>
<td>OCD2 19.4%</td>
<td>37.5%</td>
</tr>
<tr>
<td>OBI</td>
<td>Freq.</td>
<td>38.4%</td>
<td>36.3%</td>
<td>-2.1%</td>
</tr>
<tr>
<td></td>
<td>Sev.</td>
<td>20.0%</td>
<td>37.2%</td>
<td>17.2%</td>
</tr>
</tbody>
</table>

**Selected Algorithm**

<table>
<thead>
<tr>
<th></th>
<th>Freq.</th>
<th>Sev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPBI</td>
<td>Deep Net</td>
<td>Ridge Reg.</td>
</tr>
<tr>
<td>OCD</td>
<td>AXA Custom</td>
<td>Ridge Reg.</td>
</tr>
<tr>
<td>OBI</td>
<td>Ridge Reg.</td>
<td>Neural Network</td>
</tr>
</tbody>
</table>
Special attention has to be given to the selection of algorithms

- GLM model performs well on average, but is less efficient at picking up high risk customer. ML is better at picking high risk, which is very good for the underwriting scoring, where we want to “cut the tail”

**Comparison of Expected and Observed Loss Ratio**

![Graph showing comparison of ML Model and GLM Model](image)
Big Data as an enabler to understand the Customer Value
CVM is convergence of 4 different models into single model

Our goal is to apply it to compute the expected customer value

As a result, we will be able to identify the overall profitability of each individual customer

Before

- We want to underwrite cars with an Expected LR below threshold (e.g. 100%)

After

- We want to underwrite customers if their expected profit over several years (3) for all current and future products is above threshold (e.g. above 0)
## Customer Data Integration

<table>
<thead>
<tr>
<th>ML Models</th>
<th>Customer Value Model – Models used to compute the CVM</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Premium</strong></td>
<td>• Motor: Estimated premium fluctuations for 3 years</td>
</tr>
</tbody>
</table>
| **A&H Loss Ratio** | • A&H: rules to estimate the LR by major variables  
  *(will be switched to ML model at a later stage)* |
| **Expense ratio** | • Expected expense ratio (fixed / variable by cover)  
  *(will be switched to ML model at a later stage)* |
| **A&H Persistency Rate** | • Estimate A&H persistency rate based on rules for major variables  
  *(will be switched to ML model at a later stage)* |

| **Motor Loss Ratio** | • Burning Cost Model over 3 years |
| **Motor renewal** | • Expected Renewal Rate Model (3 year) |
| **Motor Cross-selling** | • Probably to buy an A&H insurance during Motor contract period |
| **GI renewal rate** | • Expected A&H Renewal rate |
Daily monitoring of customer profit
Business decisions to improve CVM

Make underwriting decision based not only on contract (one year Loss ratio) but on the expected value of a customer

Differentiated customer communication
Adjust call priority to optimize call process
Product offering process for Cross and Up-selling
Agents remuneration

Target Marketing
- Top priority segments
- Low priority segments
- D-marketing

Customer Profit

Profit1
Profit2
Profit3
Profit5

Profit score
Segmentation

Start Here
A few words in conclusion...