Use of Big Data by AXA Direct Korea

- Nicolas Hebrant / AXA Korea CTO & CDO

SEPTEMBER 2018 - HANGZHOU

SOCIETY OF ACTUARIES®

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

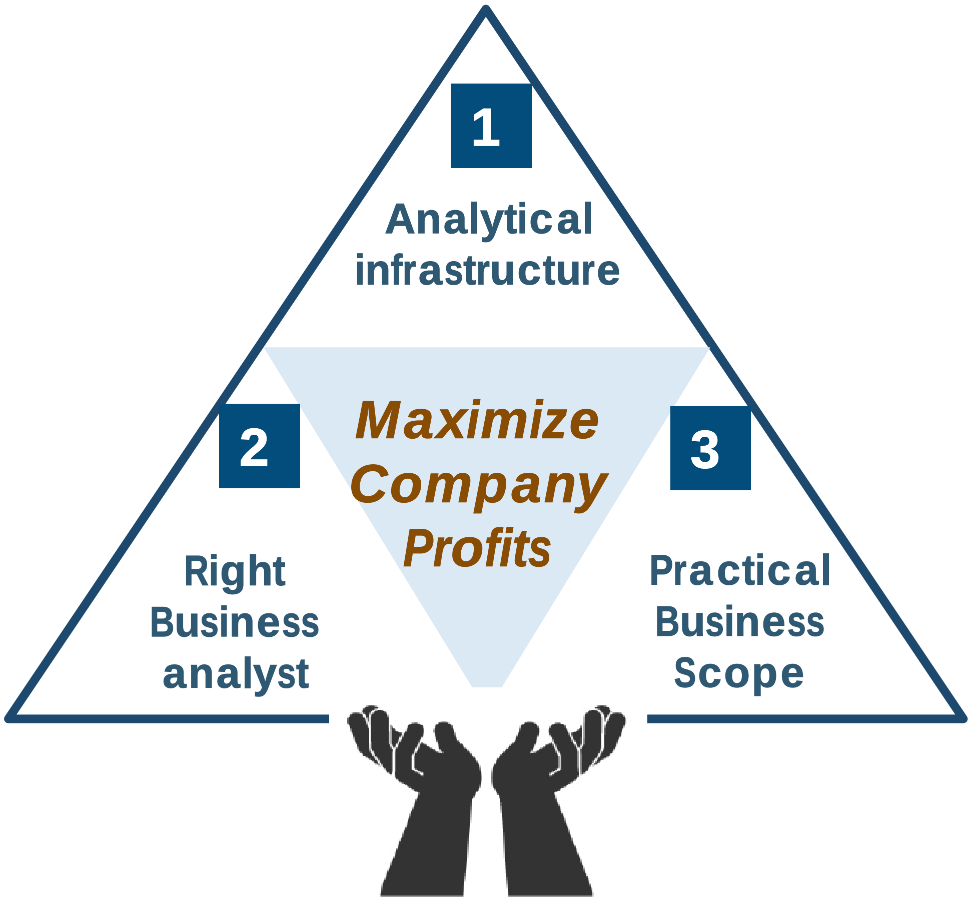
3. The current challenge: CVM

   How Machine Learning can help understand the customer value, and help design the company Underwriting strategy

Introduction of Data culture in AXA Direct Korea

SOCIETY OF ACTUARIES®

Key Components of Data Culture in AXA



1 Analytical Infrastructure

- Dedicated analytical infrastructure
- Supply real-time data analysis
- Easy to access analytical infra

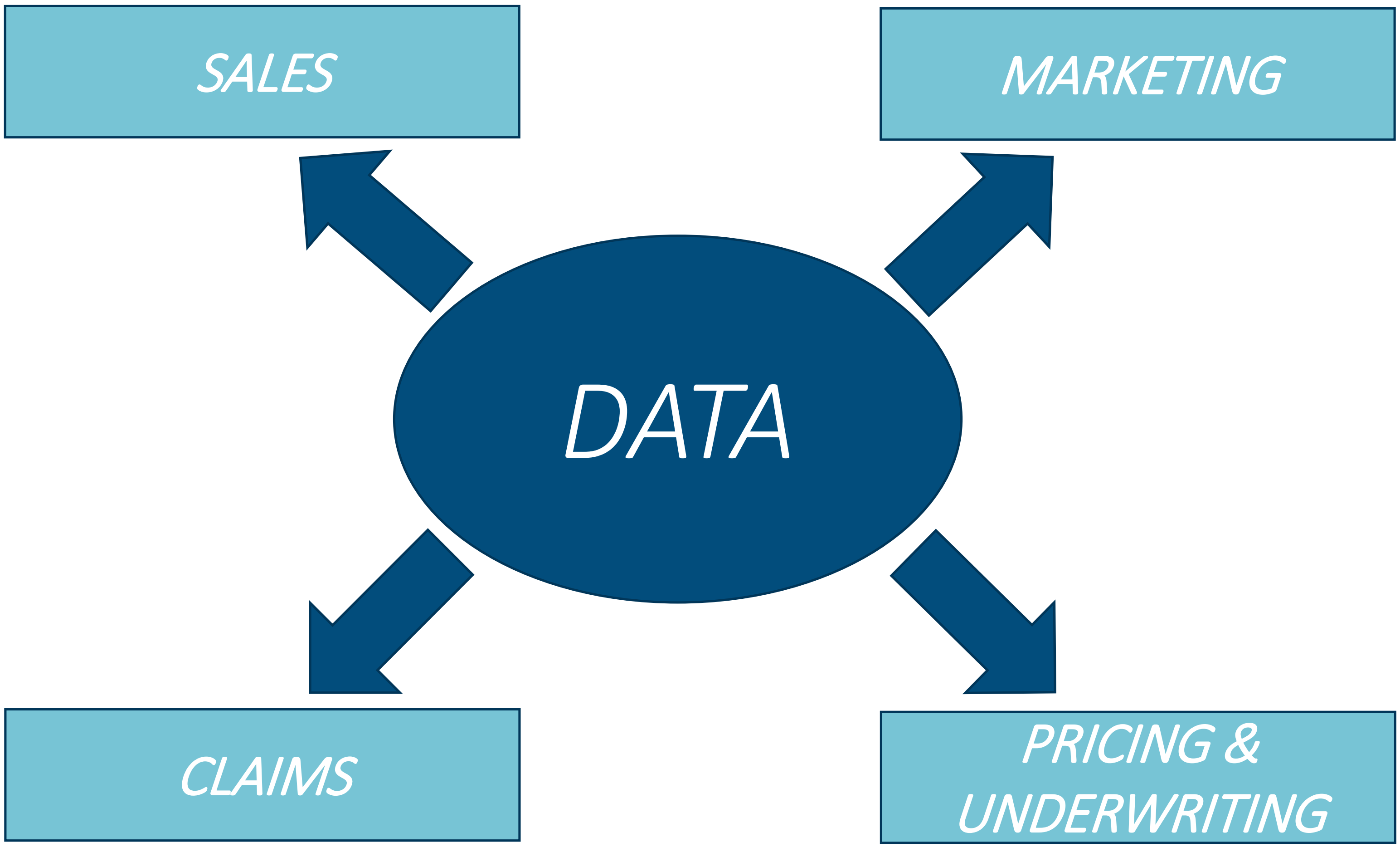
2 Skilled business analysts

- Business Analyst with data analysis skills across the company
- Supported by an independent Data Science team

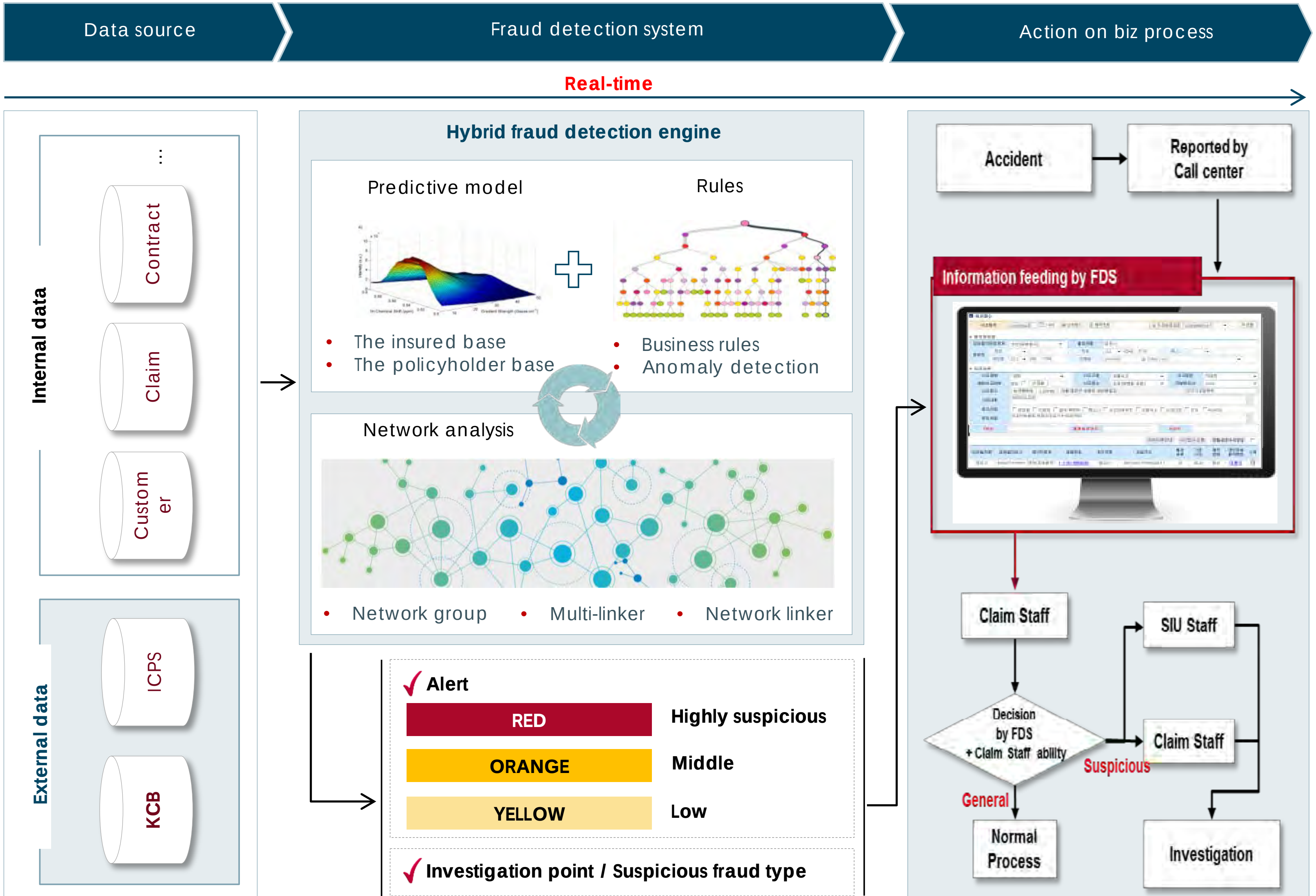
3 Focus on practical business applications

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

DATA as a business enabler – focus on practical business issues



An example from Claims: Fraud Detection System

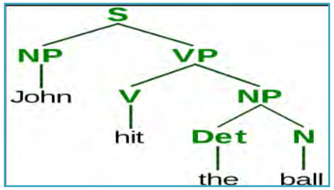
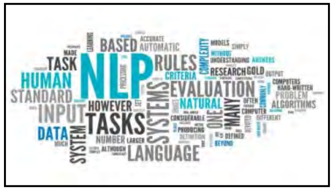


Improving the FDS (1) : Text Mining

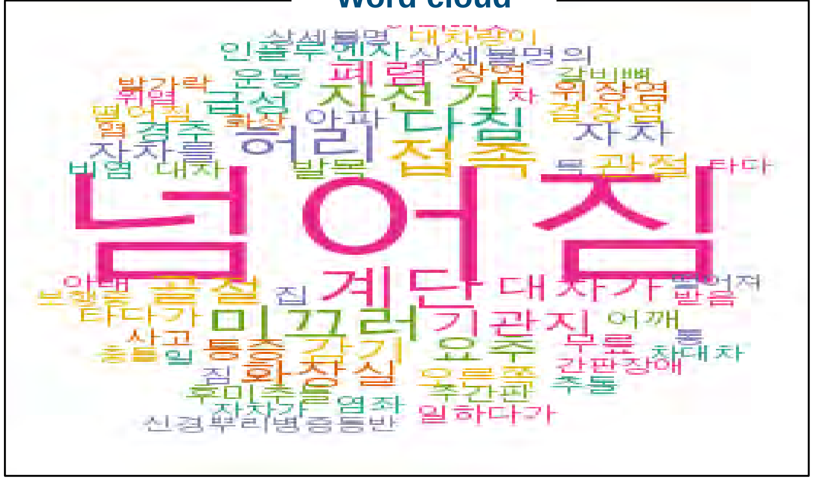
Concept

Discover meaningful information from unstructured data !

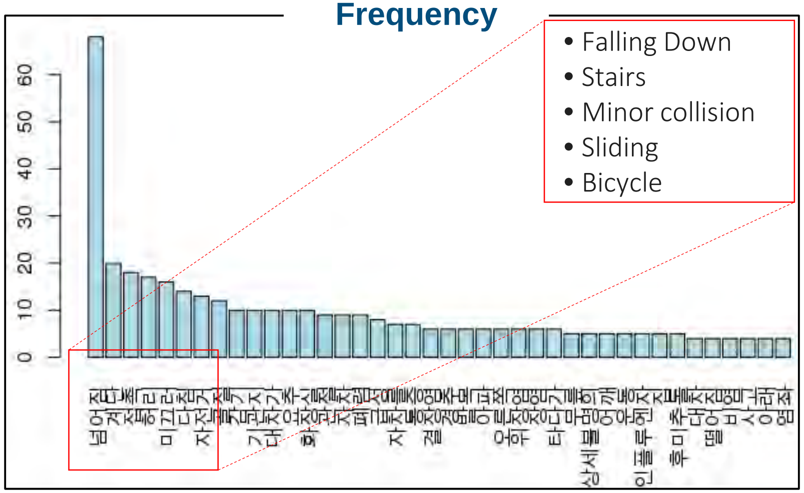
Analytic flow

Data integration	*Parsing*	*Analysis*
		
Integrate claim description data and fraud information	***Syntax parsing (using R package)***	***Conduct text mining and statistical analysis***

Word cloud



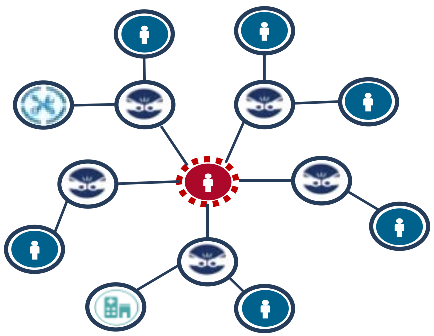
Frequency



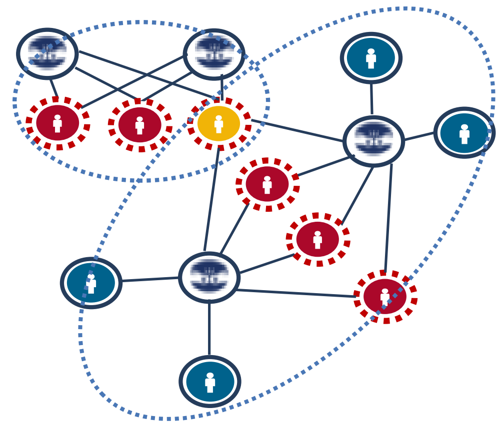
Improving the FDS (2) : Network Analysis

Fraud network type

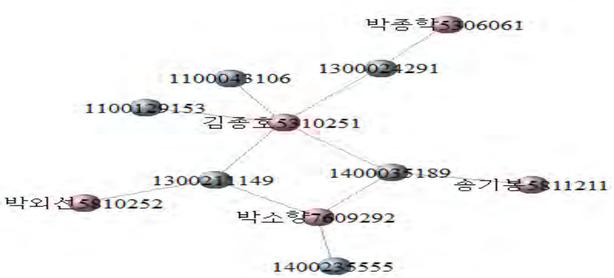
“Staged” accident

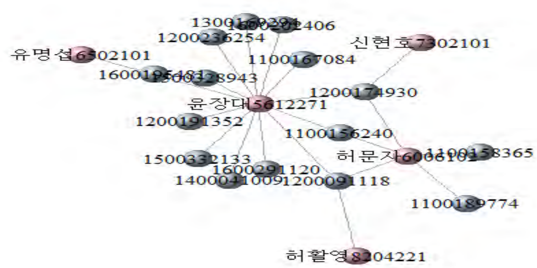


Organized Fraud Network

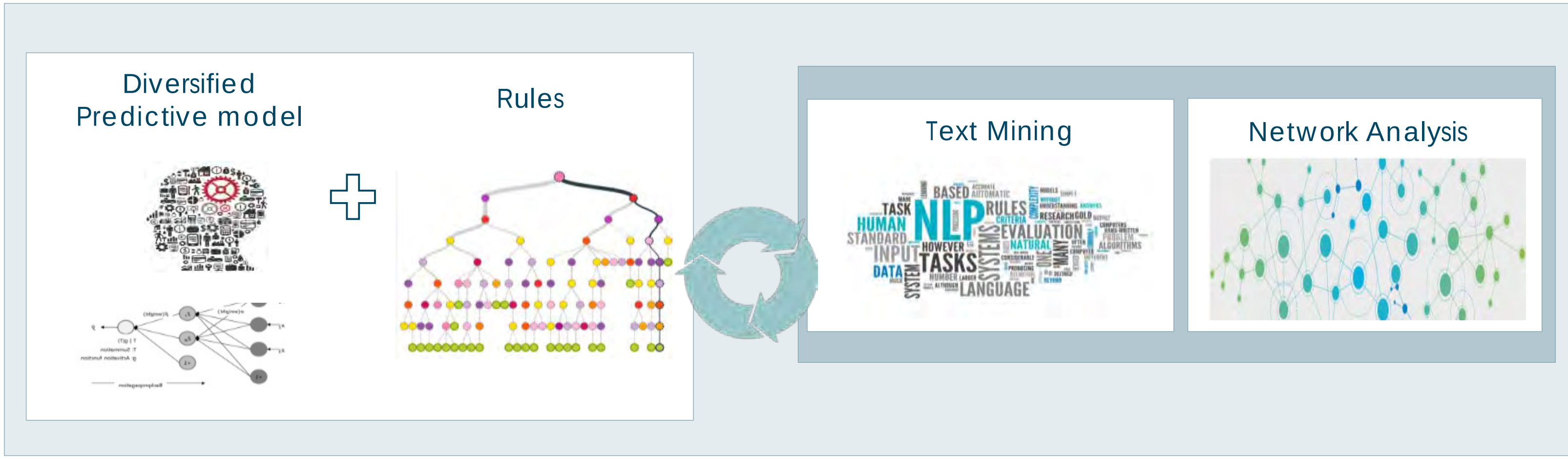


Graph Data of Networks

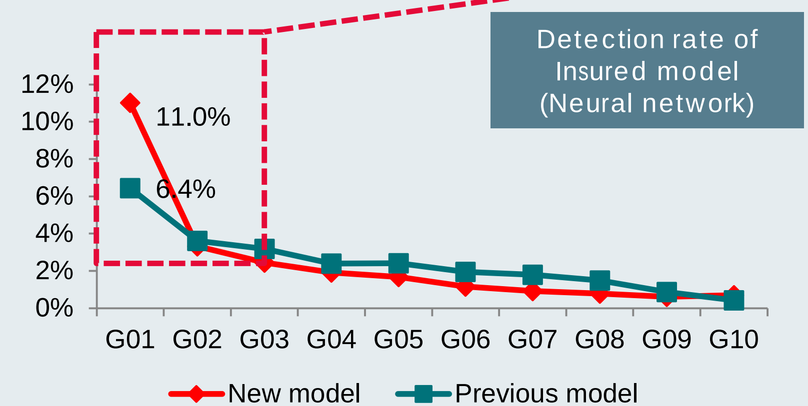


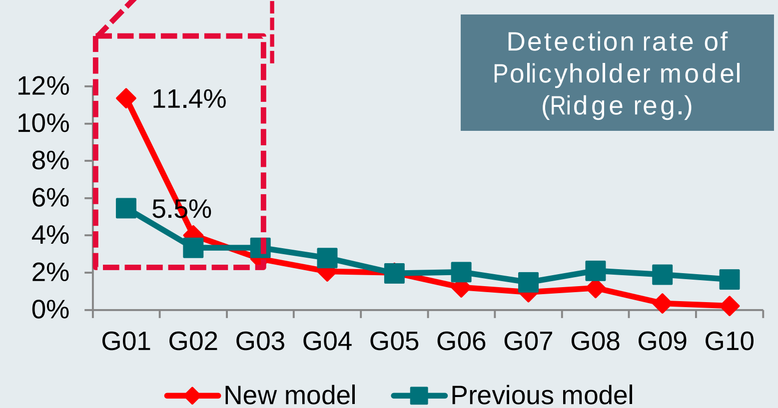


Machine Learning + Text Mining + Network Analysis = Strong Improvement in Model Accuracy

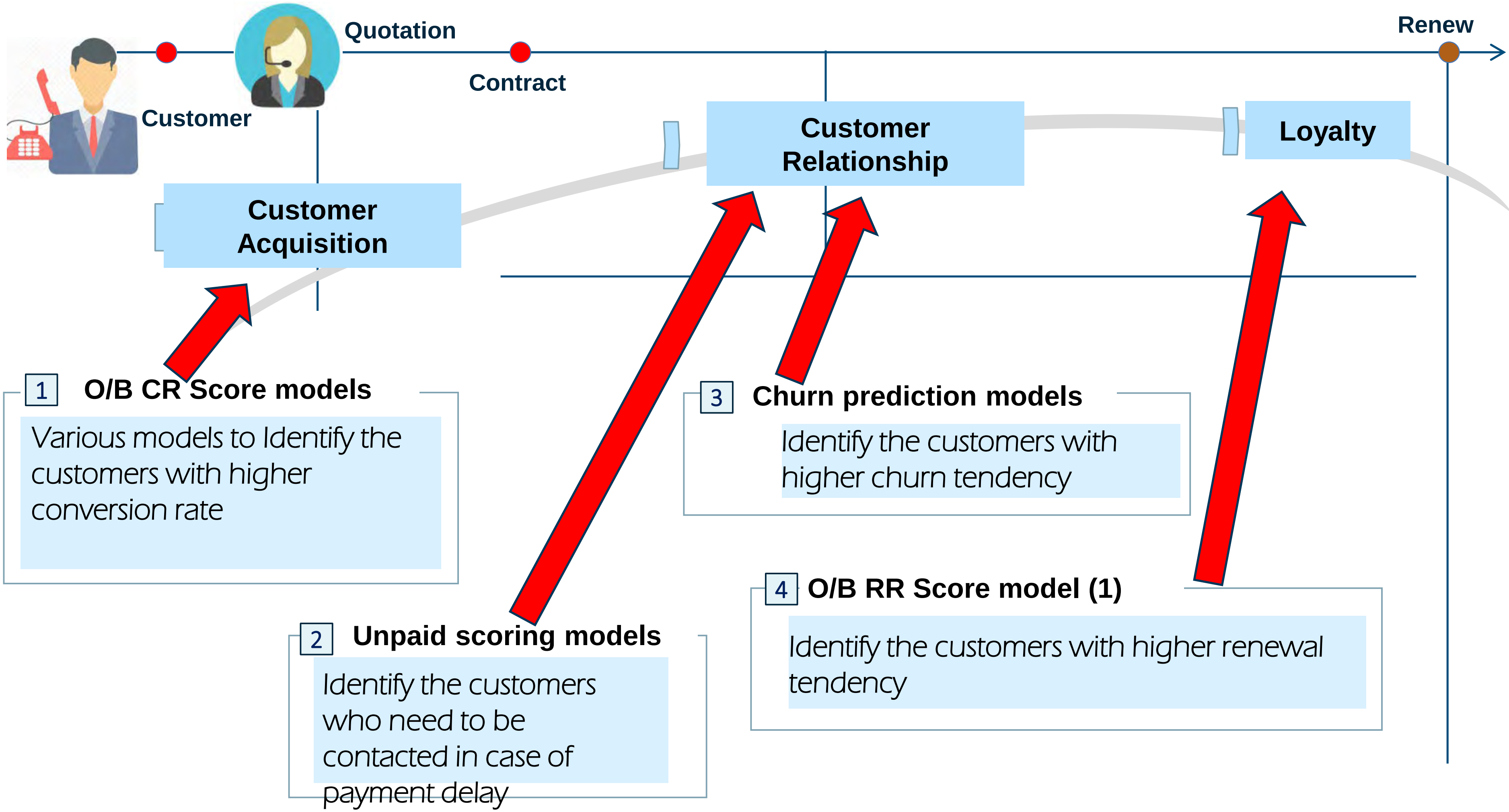


Significant performance enhancement





An example from Sales: Outbound Calls Scoring



Supporting Sales: Cross-selling model for A&H

Prediction model developed by Machine Learning

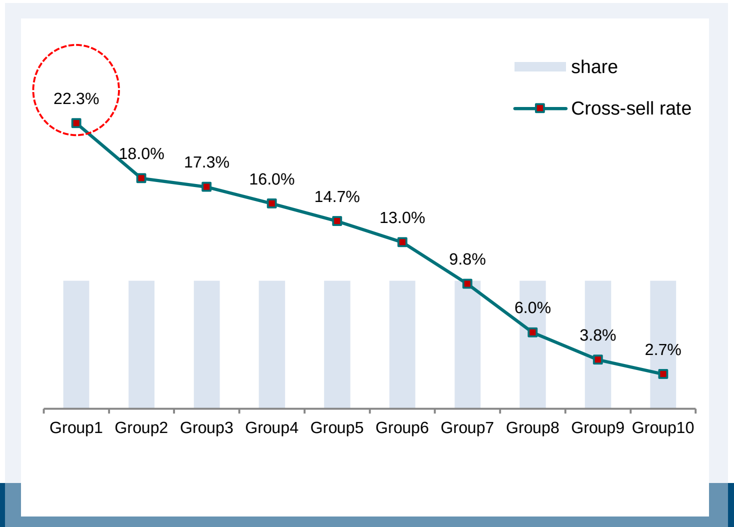
Model1. Driver - Cross-sell Model

Model result

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

Avg. Cross-sell Rate : 11.9 %

[Test : 2016.11 ~ 2016.12]



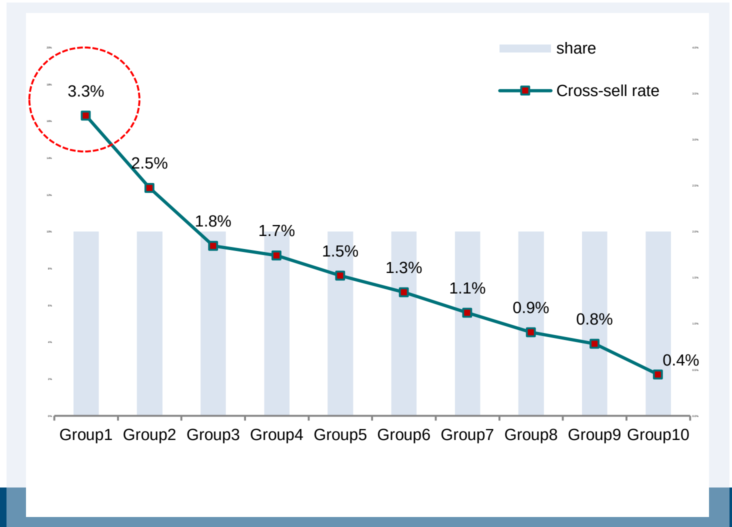
Model2. Other products - Cross-sell Model

Model result

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

Avg. Cross-sell Rate : 1.5 %

[Test : 2016.11 ~ 2016.12]



Use of Machine Learning for Underwriting

SOCIETY OF ACTUARIES®

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

Identify powerful variables, compute ELR

Constraint Model

- Based on variables available for Pricing

Introduce in Pricing

Underwriting Scoring Model

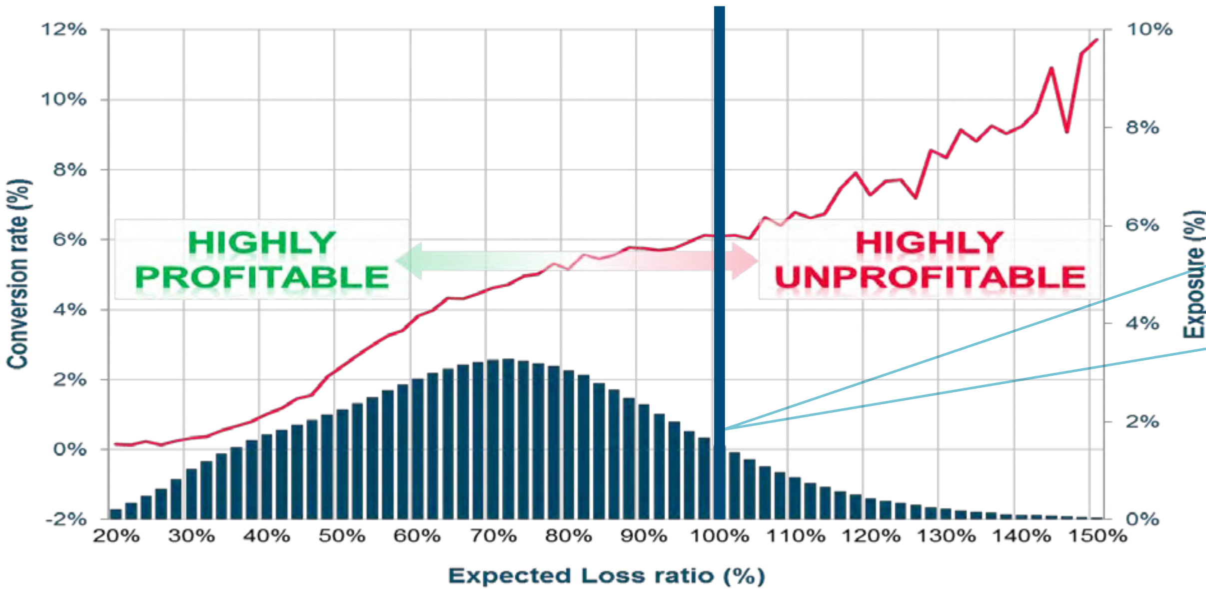
- Based on all variables available at quotation

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

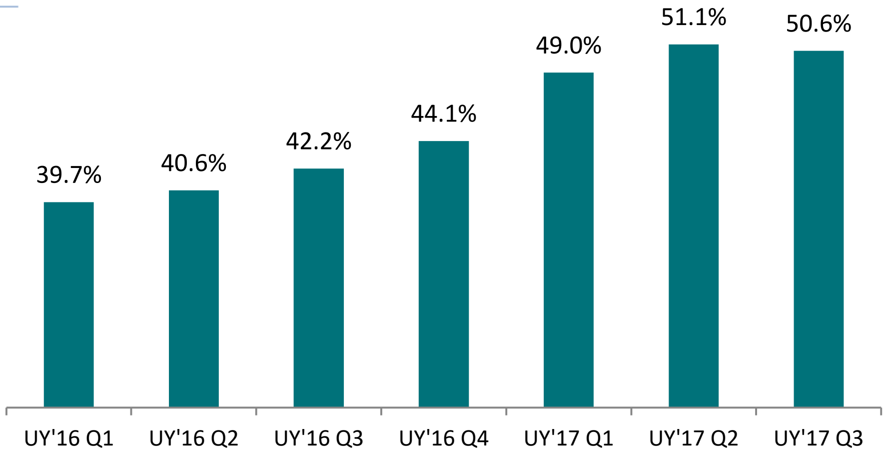
U/W scoring – Implementation and Process

- **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**

		Definition	Portion	Payment rate
Red	•	Unprofitable customers	Appr. 3 %	0.31 %
Yellow	•	Standard customers	Appr. 57 %	1.05 %
Green	•	Very profitable customers	Appr. 40 %	1.15 %

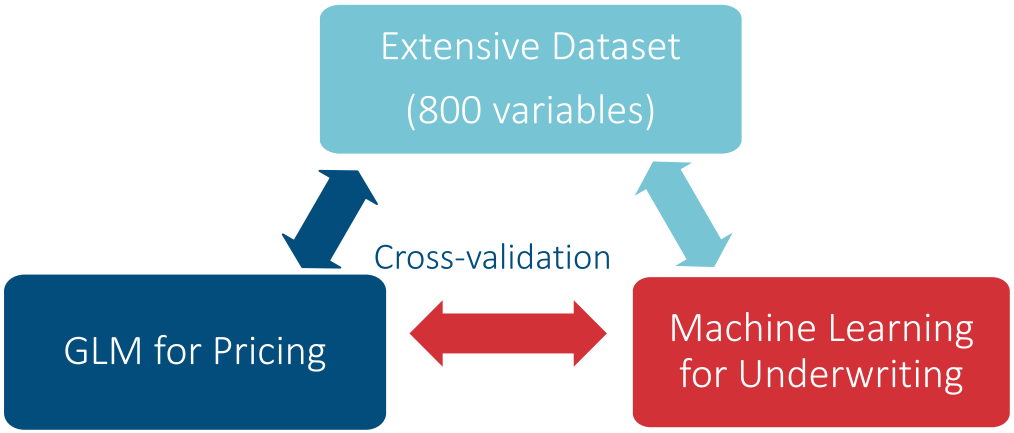
% of Green Customers



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*

ML tends to be more accurate

New Business

Gini Coefficient

Cover	Type	Old ML Gini	New ML Gini	Gap
TPBI	Freq.	26.2%	28.5%	**2.3%**
	Sev.	9.8%	12.9%	**3.1%**
TPPD	Freq.	26.3%	30.1%	**3.8%**
	Sev.	10.0%	20.0%	**10.0%**
OCD	Freq.	23.0%	26.4%	**3.4%**
	Sev.	36.8%	38.3%	**1.5%**
OBI	Freq.	32.6%	28.7%	**-0.39%**
	Sev.	39.9%	29.7%	**-10.2%**

Selected Algorithm

	Freq.	Sev.
TPBI	Deep Net	Neural Network
TPPD	Deep Net	AXA Custom
OCD	AXA Custom	AXA Custom
OBI	GBM	GBM

Renewal

Gini Coefficient

Cover	Type	GLM Gini		ML Gini	Gap
TPBI	Freq.	26.5%		27.9%	**1.4%**
	Sev.	5.1%		10.9%	**5.8%**
TPPD	Freq.	23.6%		27.6%	**4.0%**
	Sev.	8.2%		17.1%	**8.9%**
OCD	Freq.	**OCD1** 22.8%	**OCD2** 24.8%	26.1%	**1.3%**
	Sev.	**OCD1** 20.0%	**OCD2** 19.4%	37.5%	**18.1%**
OBI	Freq.	38.4%		36.3%	**-2.1%**
	Sev.	20.0%		37.2%	**17.2%**

Selected Algorithm

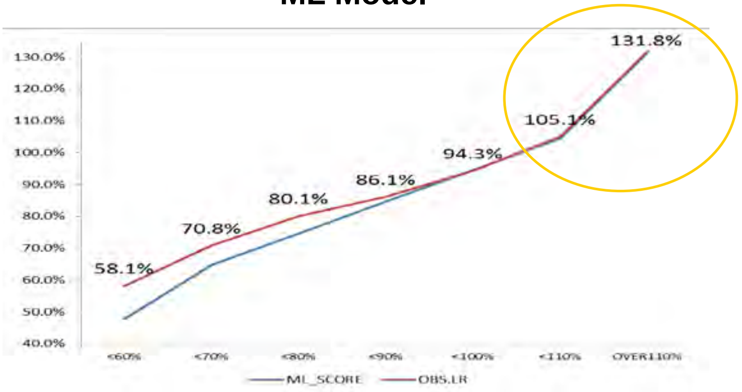
	Freq.	Sev.
TPBI	Deep Net	Ridge Reg.
TPPD	Ridge Reg.	Ridge Reg.
OCD	AXA Custom	Ridge Reg.
OBI	Ridge Reg.	Neural Network

Special attention has to be given to the selection of algorithms

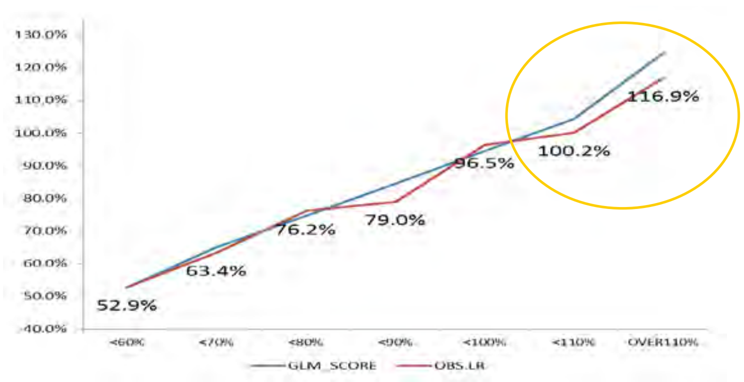
- GLM model performs well on average, but is less efficient at picking up high risk customer. ML is better at picking high risk, which is very good for the underwriting scoring, where we want to “cut the tail”

Comparison of Expected and Observed Loss Ratio

ML Model



GLM Model



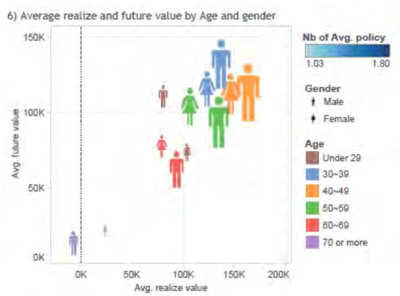
Big Data as an enabler to understand the Customer Value

SOCIETY OF ACTUARIES®

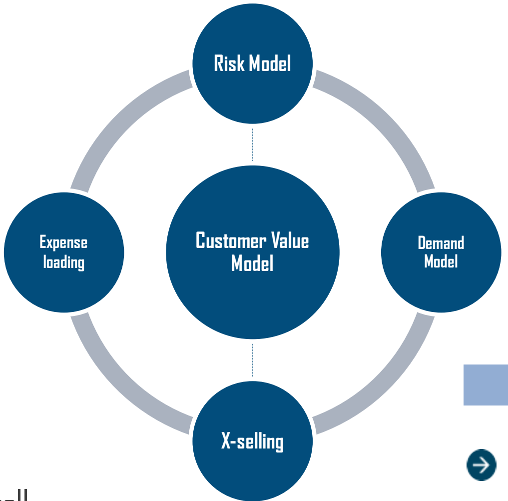
Customer Value Model – Introduction of CVM

Compute the expected value of one customer at the time of Underwriting

- 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 customers



Before

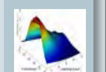
- *We want to underwrite cars with an Expected LR below threshold (e.g. 100%)*

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 Value Model – Models used to compute the CVM

Customer Data Integration



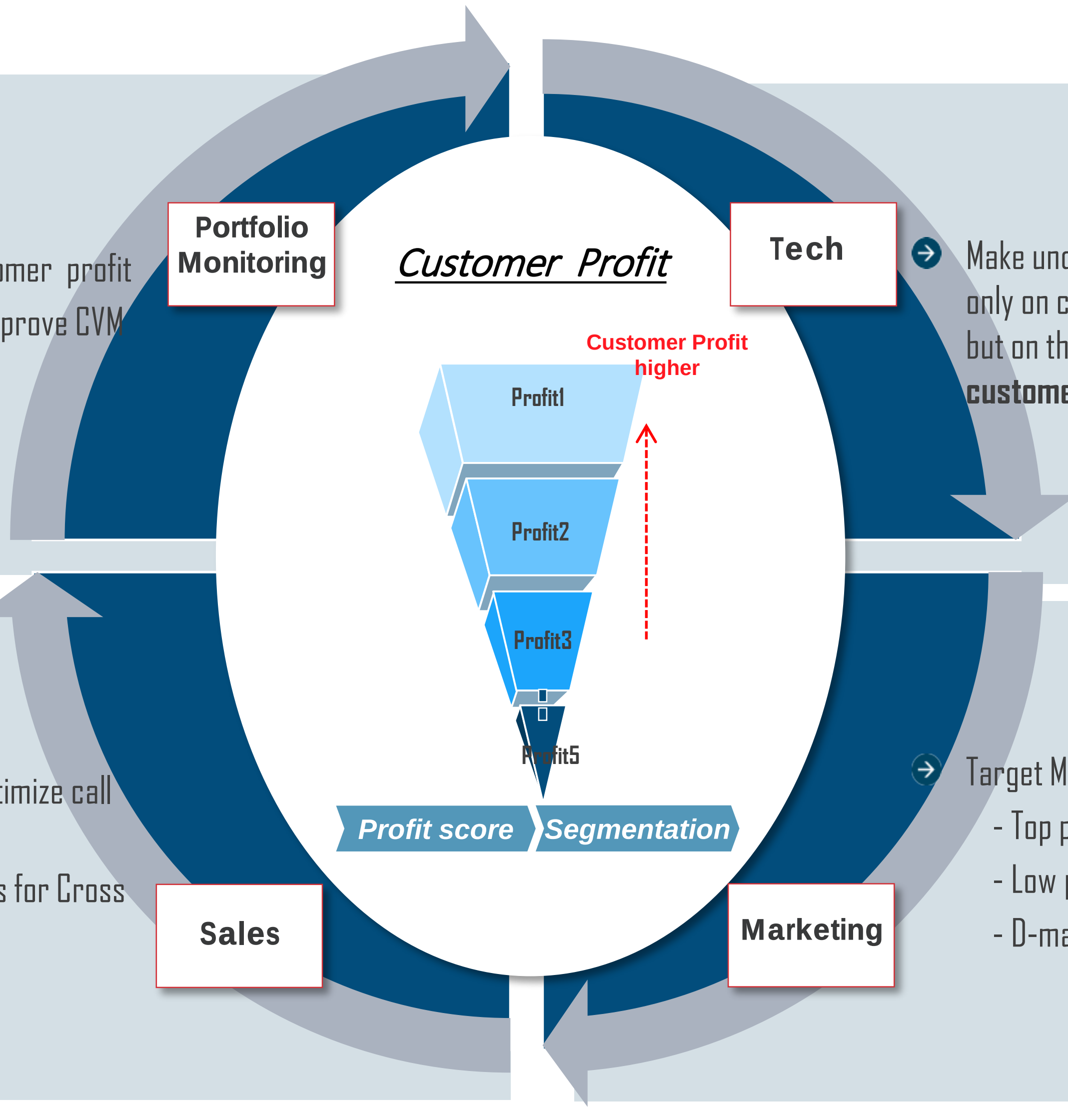
	Motor Loss Ratio	Motor renewal	Motor Cross-selling	GI renewal rate
ML Models	• Burning Cost Model over 3 years	• Expected Renewal Rate Model (3 year)	• Probably to buy an A&H insurance during Motor contract period	• Expected A&H Renewal rate

	Premium	A&H Loss Ratio	Expense ratio	A&H Persistency Rate
Table	• Motor: Estimated premium fluctuations for 3 years	• A&H: rules to estimate the LR by major variables *(will be switched to ML model at a later stage)*	• Expected expense ratio (fixed / variable by cover)	• Estimate A&H persistency rate based on rules for major variables *(will be switched to ML model at a later stage)*

Customer Value Model – Many applications planned outside of Technical

Start Here

- 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

A few words in conclusion…

SOCIETY OF ACTUARIES®