PREDICTIVE ANALYTICS

EVI TEDJASUKMANA

26 OCTOBER 2017

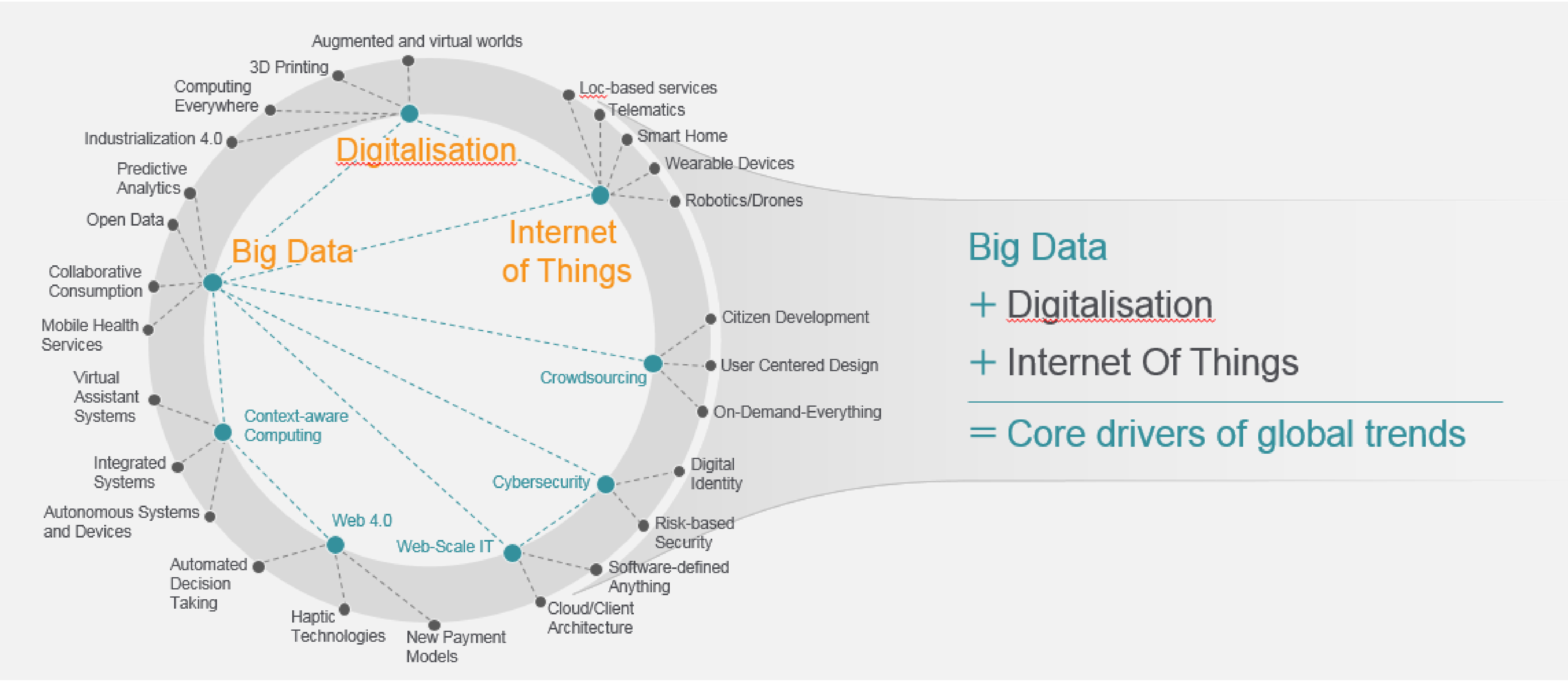
Agenda

1. Predictive analytics – why we need it?
2. Sample analytics 1 – propensity to buy
3. Sample analytics 2 – predictive underwriting

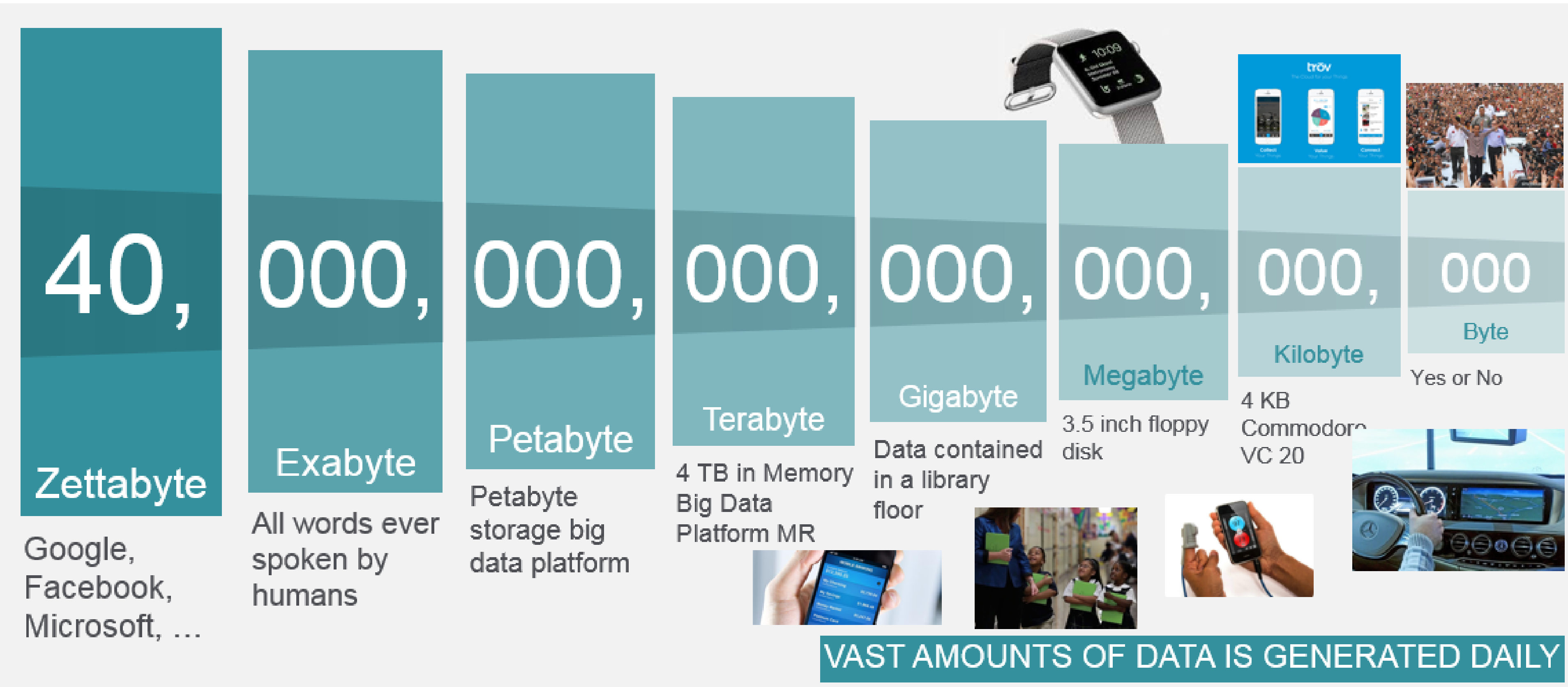
BACKGROUND

Why predictive analytics?

Core drivers of global trend

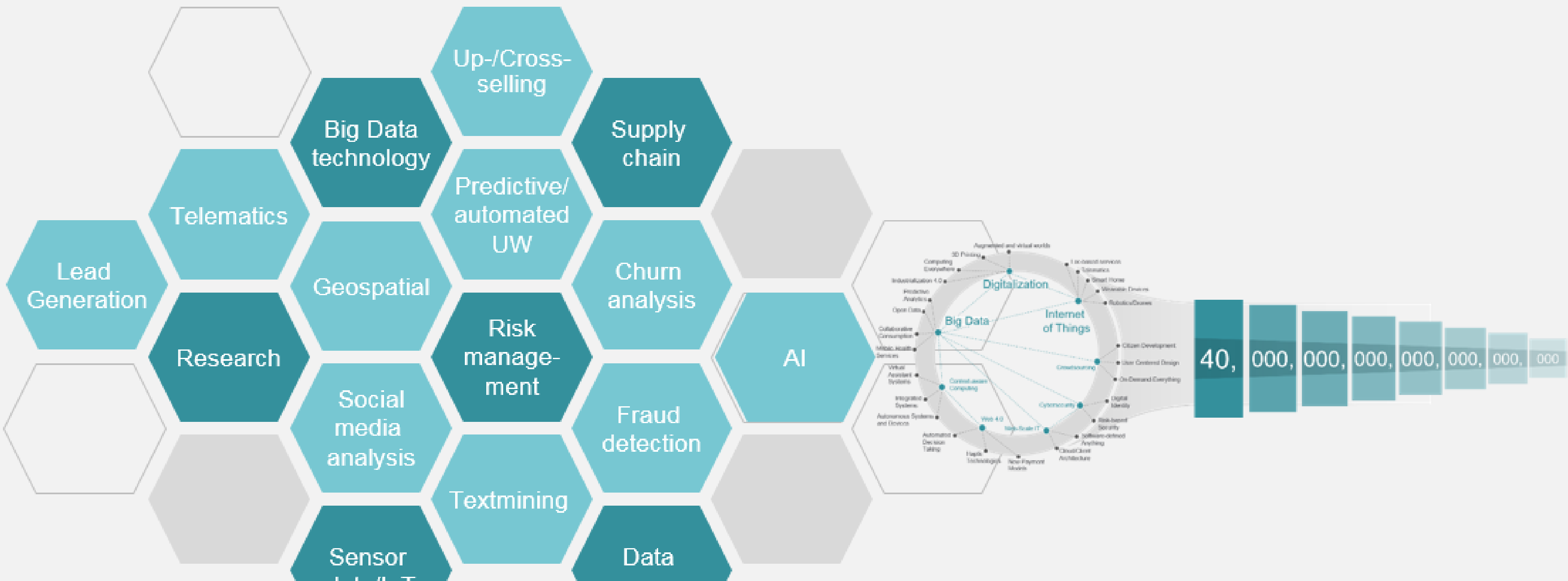


Big data is getting bigger and use cases more tangible



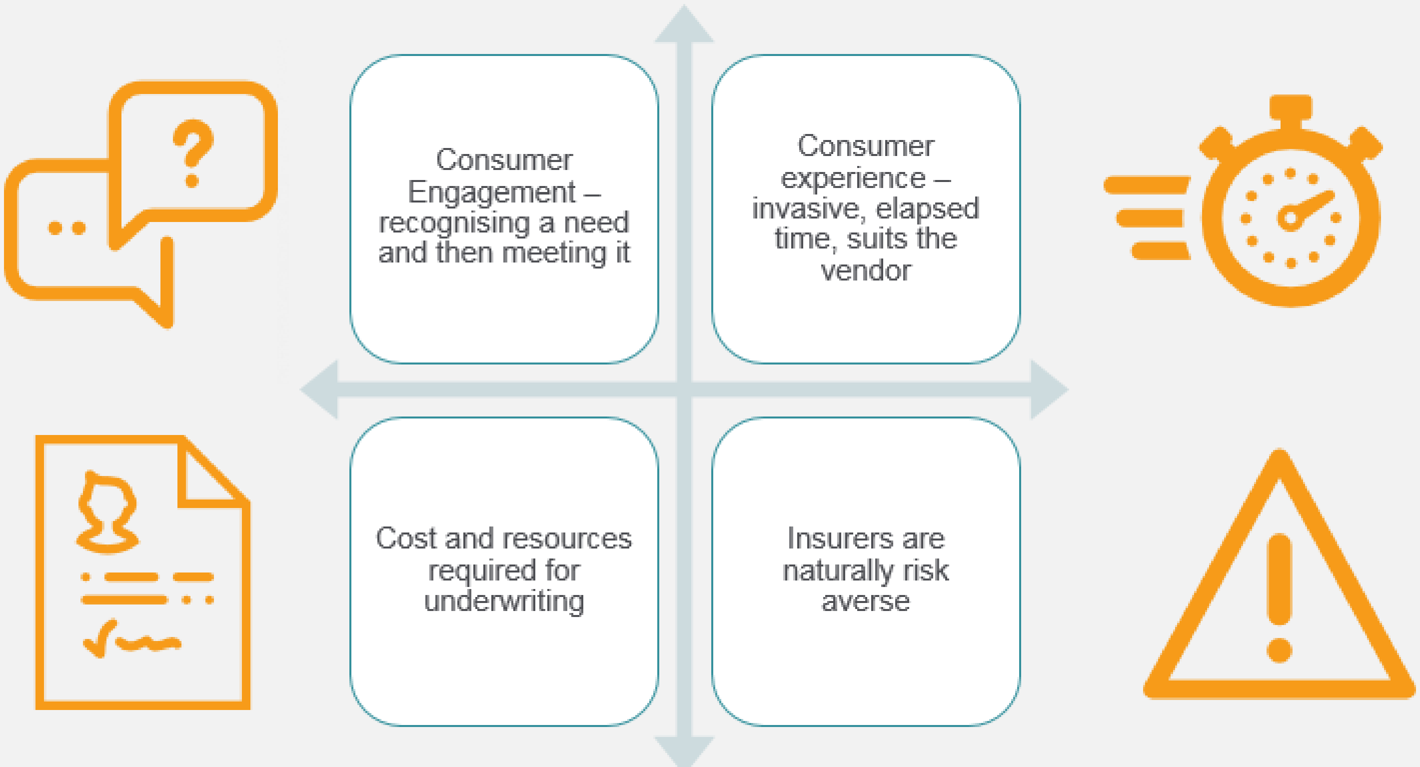
Source: IBM

With ever-increasing potential applications



INSIGHTS EXPERTISE AND RISK APPETITE IS KEY

Insurers' pain points



A changing reality for insurers

Today's reality

- Ever-changing data formats require constant development and adoption
- New sources of data provide greater insights, might trigger new questions
- Smart data is connected to everything, changes customers' behaviour
- Data access and improved insights change the way we do business
- Technology, analytics methods and human skills are improving daily

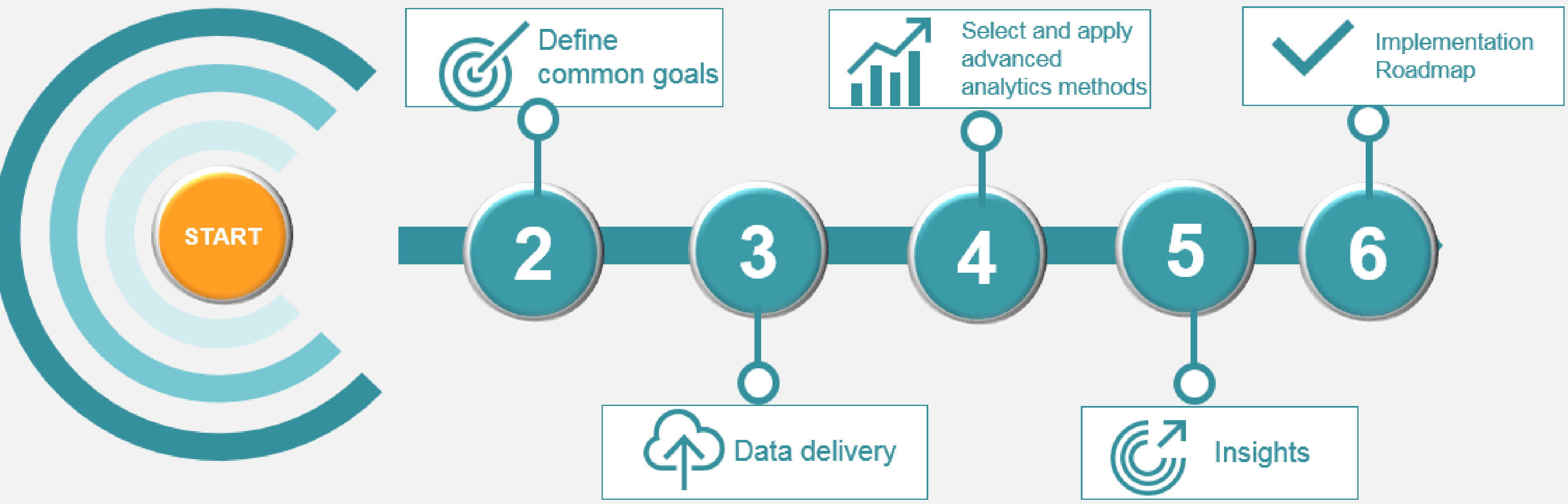
What it means for insurers

- New data formats and sources create opportunities for differentiation
- Mobile phone data could show driver and driving behaviour, lead to different underwriting outcome
- Continuous increase in number of data sources imply need for constant algorithm development
- Increased need for closer collaboration across all business disciplines
- Insurance-specific know-how in a business development context needs to be developed

We are using predictive analytics to solve specific questions across the insurance value chain for our clients

Pricing and product development	Sales & marketing	Underwriting	Portfolio management	Claims
▪ Are **best estimate** assumptions adequate?	▪ What concrete **up-sell** opportunities exist in the current portfolio?	▪ Are there any groups for which we can **simplify the underwriting process** (to improve customer experience)?	▪ How **profitable** is the business really?	▪ How good is the **risk selection** process?
▪ What **pricing basis** would be accurate for a new product?	▪ Which clients are most likely to take up **cross-sell** offers?	▪ Can we **reduce** the need for **medical exams** (and the underwriting cost)?	▪ Which customers are at **highest risk of lapsing** and how can we retain them instead?	▪ Are we attracting **poorer risks** than we intended to?
▪ What are the **risk drivers** and how do they affect the current experience?	▪ Which groups should we target for sales **campaigns**?	▪ How can we use underwriting **resources most effectively**?	▪ How does the **portfolio composition** compare with pricing assumptions?	▪ How can we **streamline the claims process**?
▪ Which **product features** appeal most to the target segment?	▪ Which **distribution channel** is performing best?			▪ Which claimants should I prioritize for **fraud investigation**?

Predictive analytics workflow - overview



SAMPLE ANALYTICS 1

Propensity to buy

Sample predictive analytics – propensity to buy

- This section outlines a sample predictive analytics project on “propensity to buy”. This was done for non-life but the same process can easily be replicated to life insurance products.
    - **Needs** – how to cross-sell motor insurance to Bank Customers
    - **Objectives** – to identify customer profiles with higher probability of purchasing
    - **Outcome** – target 5,000 customers that fit certain profiles and design campaign accordingly

Propensity to buy

Define common goals

Our client's need:

"I want to increase business by cross-selling from bank to motor insurance"

"I need help in supporting the campaign design"

Cross-sell to existing bank customers

1. Rank customers based on their potential to buy a motor insurance
2. Design campaign accordingly based on the target market

Propensity to buy

Data requests

Socioeconomic

- Age
- Gender
- Marital status
- Annual income
- Education
- Residence type
- Postal code

2,000,000+ Monthly observations

~200,000 Bank customers

\+

Bank Related

- Total deposits
- Loan balance
- Expenses

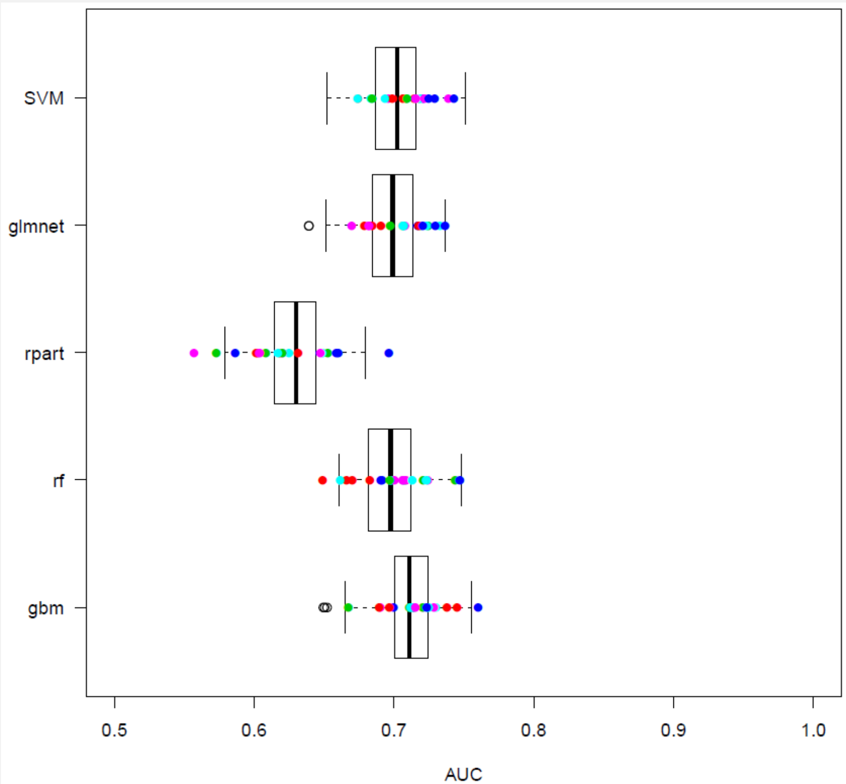
\+

Insurance Related

- Current insured on Life?
- Current insured on Auto?
- Current insured on Home?
- Other insurance coverage?

Propensity to buy

Analytics method

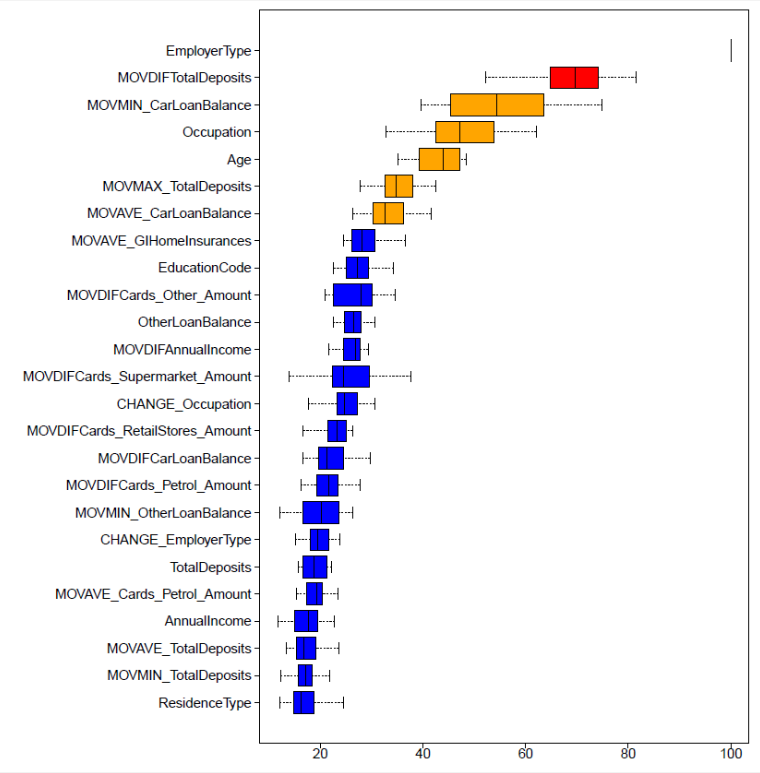


- **Model –** tried the following models:
  - SVM – Support Vector Machine,
  - GLMnet – Generalized Linear Model with penalization
  - RPART – Single Decision Tree,
  - RF – Random Forest,
  - GBM – Generalized Boosted Model
- **AUC –** Area Under Curve. Gold standard predictive power measure based on testing data. It ranges from 50% to 100% with 50% being random guessing and 100% being equal to perfect predictions
- **Simulation –** For each model, 20 simulations were run resulting in 20 different AUC values (represented by the box plots)

All models, except for decision tree, have an AUC roughly 70%, of which GBM is the most accurate

Propensity to buy

Predictor variables



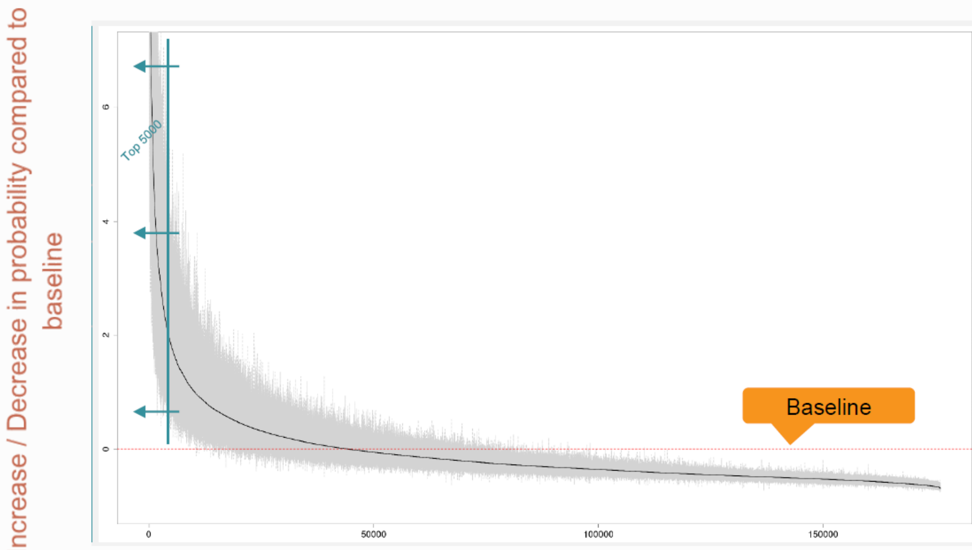
- **Variable importance –**
  - The importance of each covariate has been calculated over 20 simulations
  - The plot shows the most important predictors

- **Important predictors –**
  - Employer type
  - Biggest movement in total deposits in last X months
  - Minimum car loan balance in last X months
  - Occupation / Age
  - Maximum total deposits over last X months
  - Average car loan balance in last X months

Propensity to buy

Insights

Bank customers (without motor insurance) sorted according to their probability of purchasing



Insights

- This graph can help to decide the number of persons to contact
- The relative increase / decrease to the baseline probability is represented by the black line
- The grey area can be interpreted as confidence band
- Ideally, only bank customers with above average probability should be contacted

Propensity to buy

Implementation

- Target top 5,000 bank customers

- Customer segments define with the following criteria:
  - Age between 16 to 30 years old
  - Are employed in a bank
  - Have spent or received multiple 1,000 Euros recently
  - Have had high total deposits recently
  - Have had a positive car loan balance recently
  - Already existing insured with the company
  - Have received a raise in salary

Current application is on Non-Life but the same process can be adopted for Life insurance

SAMPLE ANALYTICS 2

Predictive Underwriting

Sample predictive analytics – predictive underwriting

- This section outlines a sample predictive analytics project on “predictive underwriting” using loyalty programme data and insured data. The same process can easily be replicated to other third party data such as bank.
  - **Needs** – to simplify the current application process and to target selected customers with better risk profiles
  - **Objectives** –
    - To streamline / minimize the existing underwriting requirements
    - To complement the application form with the information gathered through loyalty programme
    - To target loyalty programme members that are not insured
  - **Outcome** – at least X% customers qualified for standard rates with Y% reduction in the underwriting requirements

Predictive underwriting

Define common goals

Our client's need:

"I want to significantly reduce the underwriting requirements with a minimal price impact while retaining sound risk management."

"Which variables from my loyalty programme data can streamline an up- and cross-selling campaign?"

Use case 1 – Upsell / Cross-sell to existing policyholders

1. Simplify the sales process and improve customer journey
2. Streamline the application form
3. Waive or replace questions with proxy or alternative data sources

Use case 2 – Cross-sell into loyalty programme members

1. Offer insurance coverage to the loyalty programme members who are not insured.
2. Obtain proxy answers from the loyalty programme data to streamline the application process.

Predictive underwriting

Data requests

Application form

- Answers to application form questions over several years
- Underwriting decisions at benefit level
- Target variables: standard / non-standard, smoker status, HIV status

~400 000 lives analysed

100+ Application form and loyalty programme variables analysed

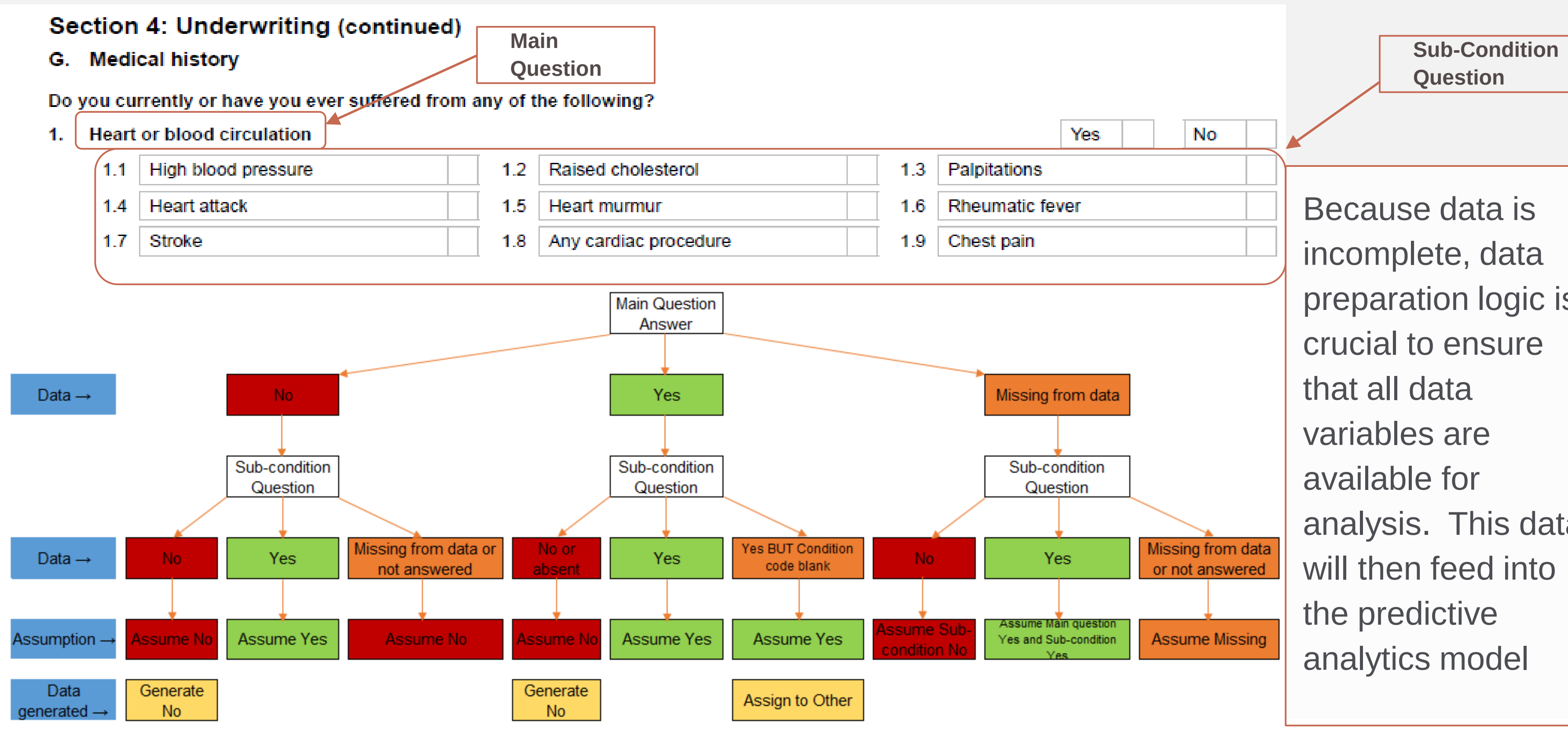
\+

Loyalty programme data

- Indicator of choice to join the programme
- Variables: blood pressure, cholesterol, BMI, age, gender, driving history
- Next phase will include daily activities data through wearables

Predictive underwriting

Data preparation logic

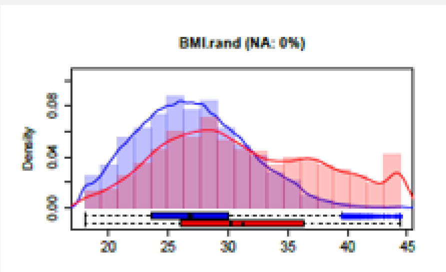


Because data is incomplete, data preparation logic is crucial to ensure that all data variables are available for analysis. This data will then feed into the predictive analytics model

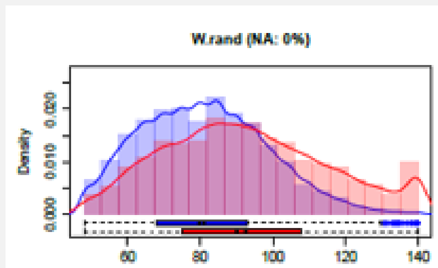
Predictive underwriting

Understanding data

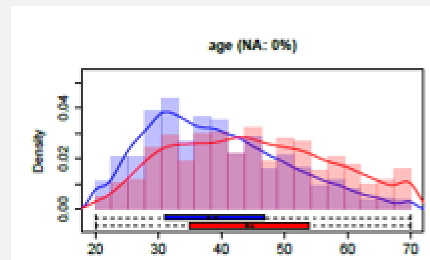
BMI



Weight

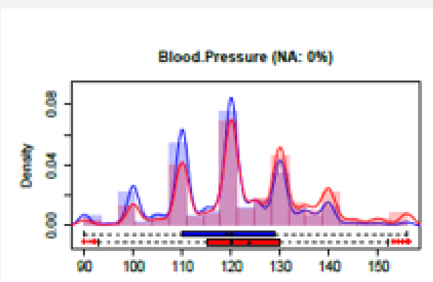


Age

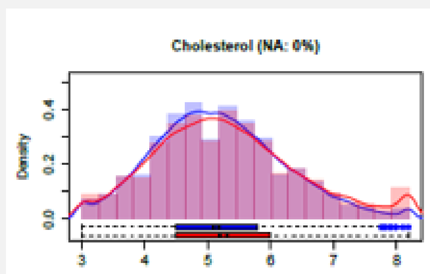


This section shows the reasonability checks of the variables and impute missing data

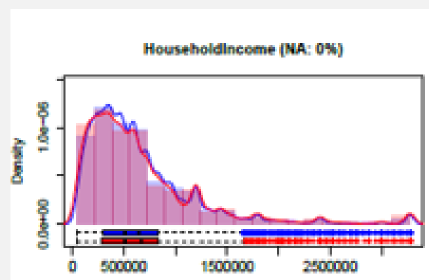
Systolic Blood Pressure



Cholesterol



Household Income



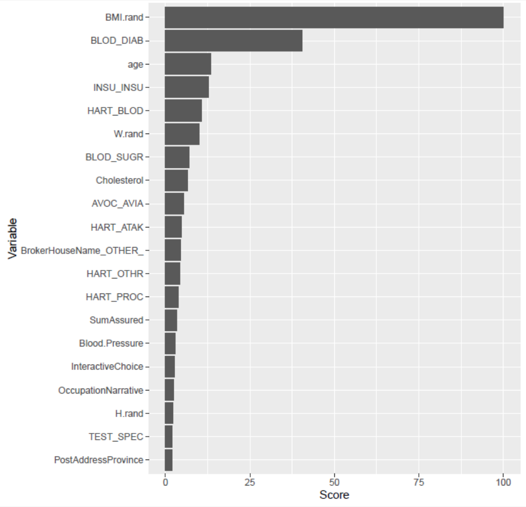
Predictive underwriting

Analytics method

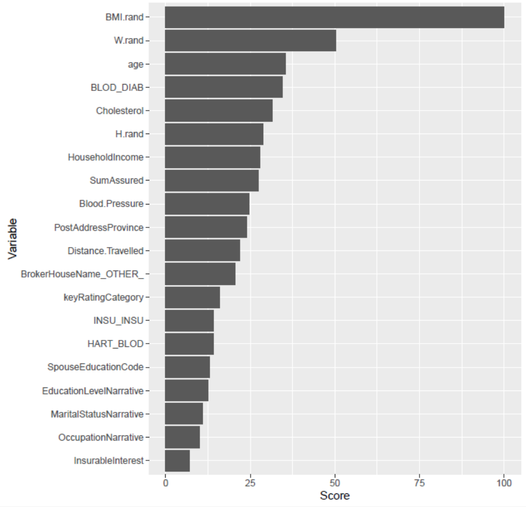
Variable importance

- This section identifies the most important application form questions when making underwriting decision.
- The importance of every application form question is determined.
- We checked the consistency across models; in this sample, we tested 5 different models (samples of the 2 models are provided on the right).

Gradient boosting model



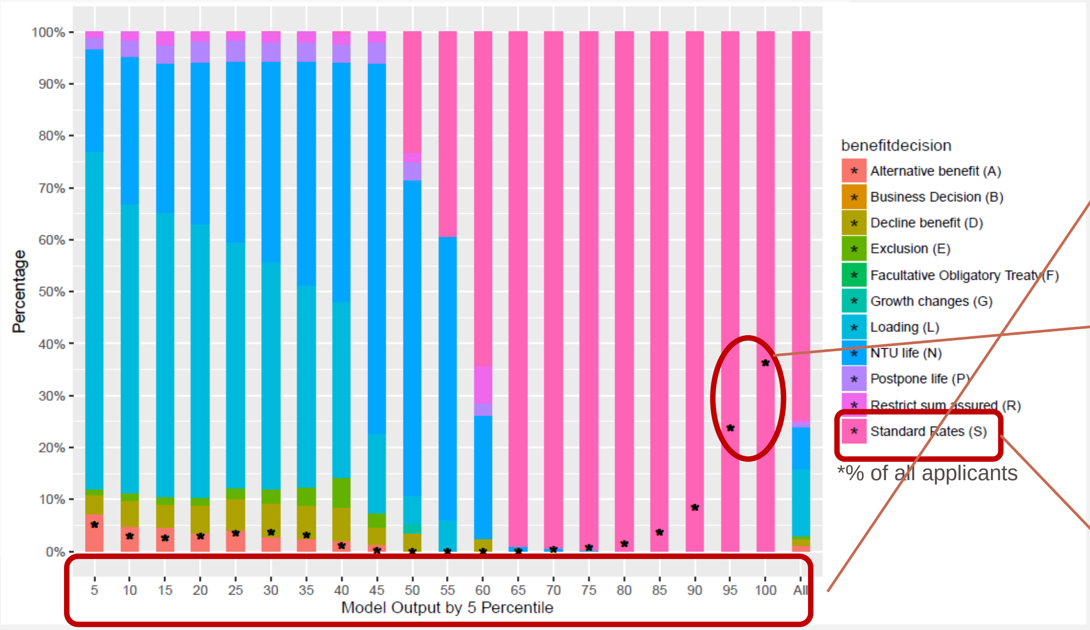
Random forest



Predictive underwriting

Insights – predictive vs. underwriter

Random forest predicts the underwriting decision



Banded predicted probability of standard

This chart shows the comparison between the actual underwriter decision vs. predicted decision from the Random forest model:

- The horizontal "X" axis shows the predicted probability of standard within the random forest model (i.e. "90" shall mean that predicted probability of standard (y) is 86% < y < 90%).
- The black bullet point inside the bar chart indicates the volume of cases as % of all cases (i.e. if we add up the "95" and "100 which means predicted probability of standard > 90%, we could see 60% applicants meet this standard)
- For this 60% of applicants, the actual underwriting decision is 100% standard.

Use case: offer the top X% a standard decision with significantly reduced underwriting requirements when up- and cross-selling

Predictive underwriting

Implementation – pre- vs. post-customer journey

PRE

- Many pages, 150 fields to be completed!
- Blood tests, invasive
- Lengthy
- Inconvenient

POST

- <20 fields including knock out
- Potentially sourced from loyalty programme – no pain point for client
- Future-proofed for 3rd party data

SUMMARY

Summary

1. Predictive analytics – why we need it?
   a. Ever-changing data formats require constant development and adoption
   b. New sources of data provide greater insights which might trigger new questions
   c. Smart data is connected to everything which may change customers' behaviours
   d. Data access and improved insights change the way we do business
   e. Technology, analytics methods and human skills are improving daily
2. Sample analytics 1 – propensity to buy
   a. How do you target customers with higher likelihood of purchasing
   b. Business impact – higher take-up rate and save on marketing cost
3. Sample analytics 2 – predictive underwriting
   a. How do you simplify the customer journey without giving in on risk management
   b. Business impact – increase customer satisfaction and better risk selection

QUOTE

"Change will not come if we wait for some other person, or if we wait for some other time. We are the ones we've been waiting for. We are the change that we seek."

Barack Obama

Thank You

PERSATUAN AKTUARIS INDONESIA
(THE SOCIETY OF ACTUARIES OF INDONESIA)
Sekretariat : Jl. Tebet Raya No.66 C, Jakarta Selatan 12820, Indonesia , Telp. +62-21 835 5105, Fax. +62-21 3650 5600
Website : www.aktuaris.org, E-mail : secretariat@aktuaris.org, Twitter: @AktuarisPAI