Reacfin

3RD CONGRESS OF THE HELLENIC ACTUARIAL SOCIETY

How data science can improve the (motor) insurance business

3 NOVEMBER 2023

SPEAKER'S INTRODUCTION

Xavier MARECHAL

CEO Reacfin and IA|BE qualified actuary

Expert in Non-Life and Health insurance (pricing, product development, reserving and risk management) and Data Science.

Reacfin Consulting

We offer consulting services in actuarial science & quantitative finance, including a.o. capital - portfolio - product - risk - and liquidity - management. We build our expertise on broad data science capacities.

Reaxii

We co-develop solutions with our clients, i.e. we integrate our solutions in our client's systems and processes and we secure full knowledge transfer (e.g. open source code).

Reacfin Academy

We share our knowledge with our clients. We offer a comprehensive learning platform, including on-site trainings, e-learning modules, webinars etc.

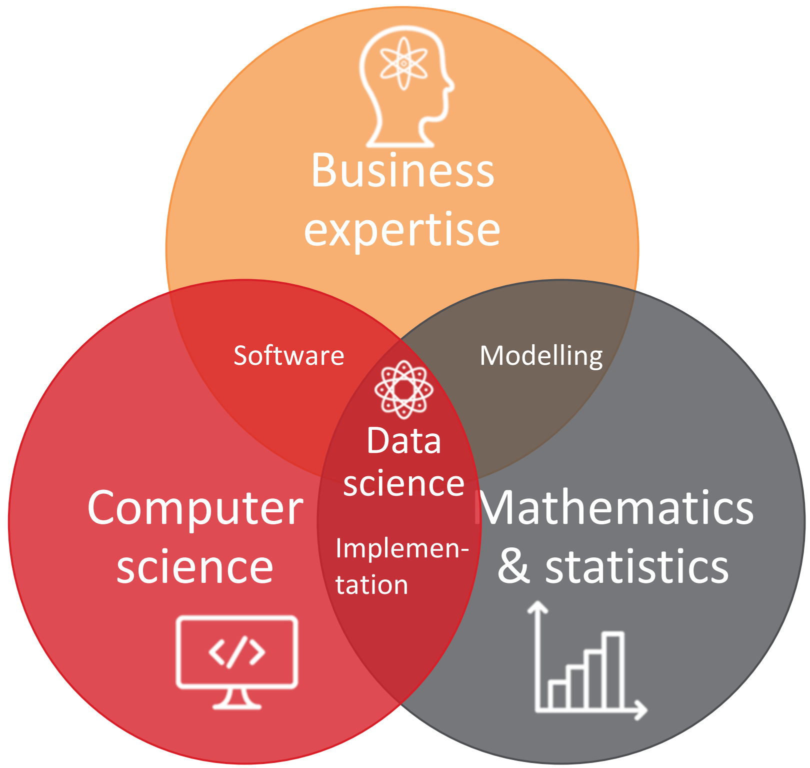
AGENDA

What is data science ?

"Data science consists in collecting, decrypting

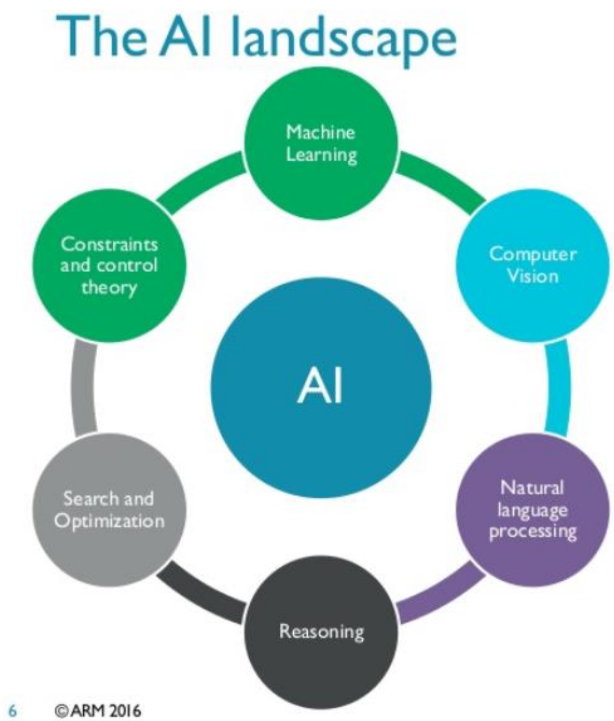
and analyzing heterogeneous data

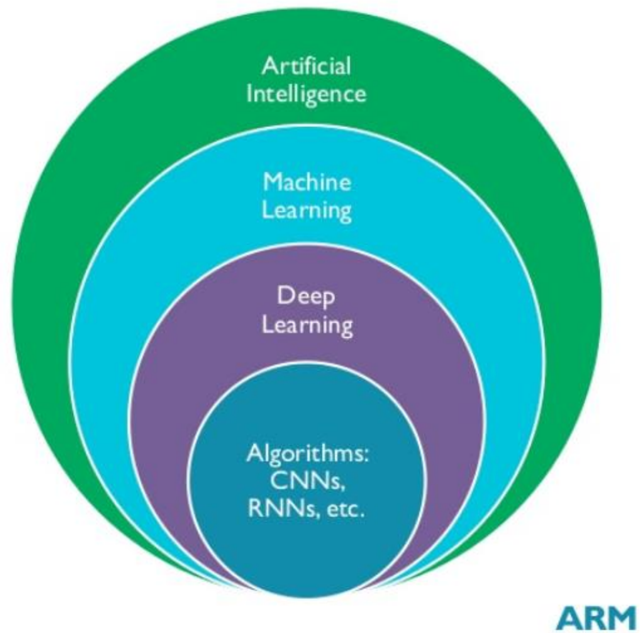
for practical learning."



What is artificial intelligence?

"An area of data science that aims at developing programs able to perform tasks or take decisions which would normally require human intelligence."



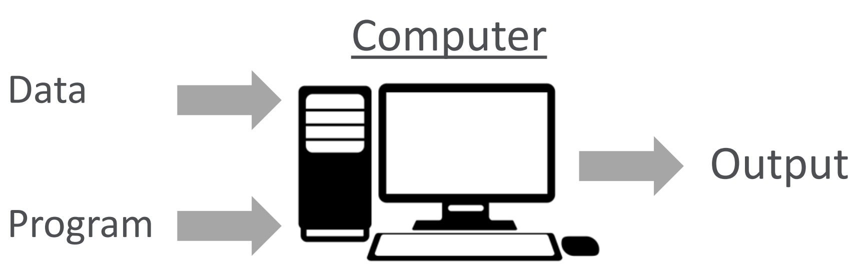


What is machine learning?

“Machine learning is a Field of study that gives computers the ability to learn without being explicitly programmed.”

- Arthur Samuel, 1959

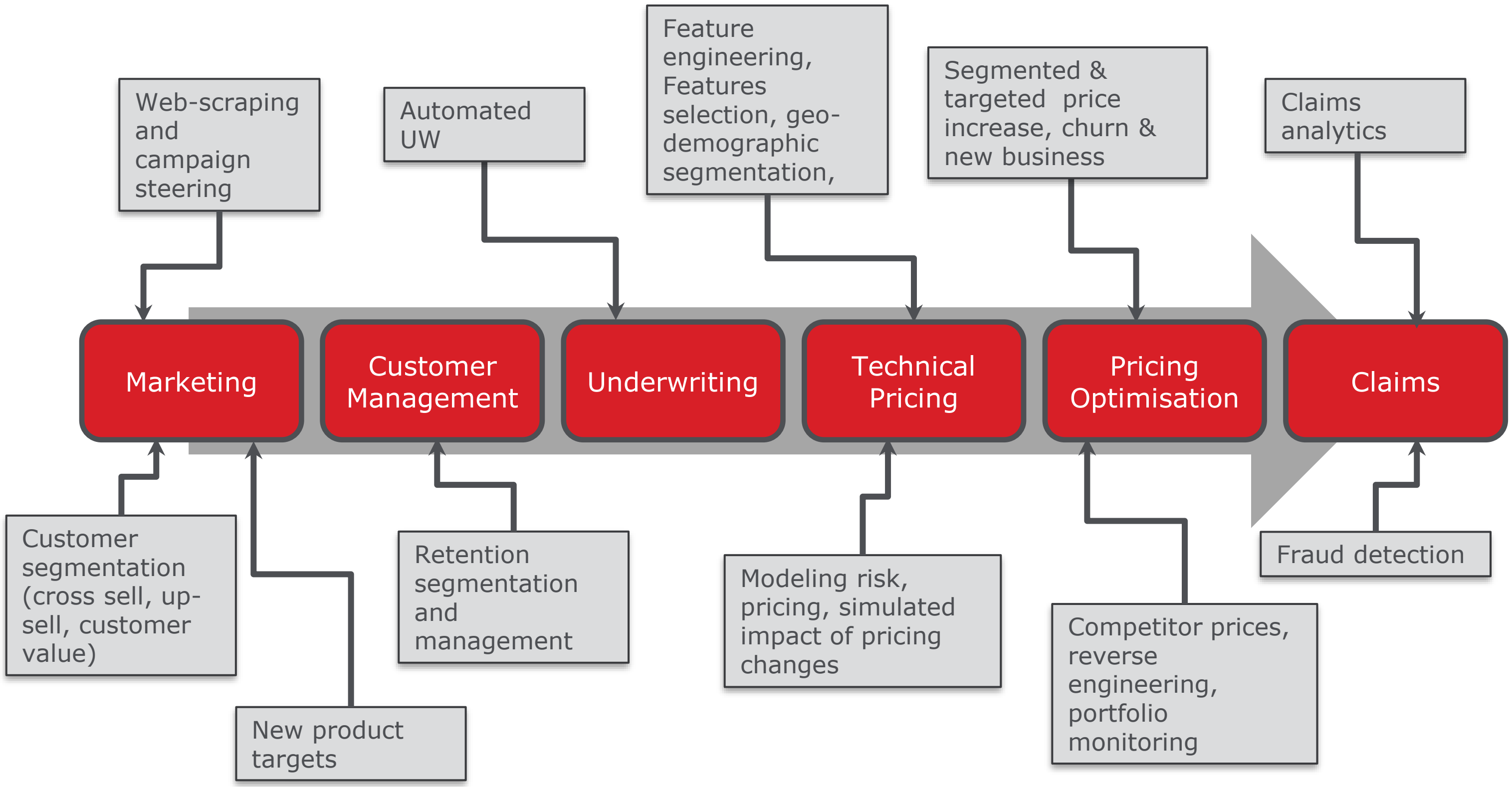
Traditional programming



Machine Learning

DATA SCIENCE IN INSURANCE

Data is everywhere in the insurance value chain... and so too the opportunities for using data science



AGENDA

Challenges in non-life insurance

Increasing competition

- Commoditisation of insurance products
- Pricing comparison systems
- Sophistication in pricing
- Insurtechs simplifying products/processes

Availability of new data sources

- External data (IoT, open data,…)
- Use of unstructured data

New customers needs and behavior

- Digitalisation of underwriting process
  Direct vs Brokers
- New risks emerging
- Focus on price (made possible thanks to pricing comparison systems)

To adress these challenges, Insurers have to

- Innovate in product development and surrounding services
- Capture and identify relevant features for pricing models
- Adapt faster to market changes (identification of risks, building new models, faster product deployement)
- Improve processes (e.g. claims management) to increase added-value to clients.

What are the key success factors for non-life insurers to face these trends and challenges?

Competitive advantages in the future

- Creative sourcing of data (new sources of external data, behavior-influencing data monitoring,…)
- Creative usage of data (features selection, features engineering,…)
- Distinctiveness of analytic methods beyond actuarial sciences (NLP, image processing, big data,…)

Build a data-driven culture

- Leadership: set clear goals and decide what success looks like
- Decision Making: base decisions on evidence not on gut feeling
- Company Culture: Don't ask "What do you think"? Ask "What do you know"?

Technology changes much faster than people

- Insurers should not only invest in analytics and IT technologies
- Key to train & motivate their highly skilled business experts & data scientists to adopt the newest tools
- Make sure people use Advanced Analytics with creativity, confidence and consistency

AGENDA

CREATIVE SOURCING AND USAGE OF DATA

There exist 2 main categories of data

- Structured data :
  organized and well characterized data that are easy to use because they are well identified.
  - E.g. insurer's policies and claims data
- Unstructured data:
  non-organized data not easy to manipulate and which require much preparation (everything else).

80%

Of business information are unstructured

Unstructured data

PDF files

Word files

Emails

Websites

Mobile data

Social media

Sensor data

Structured data

Commercial data

CRM

Model calculations

Data warehouse

Open data

Purchased databases

CREATIVE SOURCING AND USAGE OF DATA

Different sources and types of information

- Numerous sources of internal or external data
- Data type is different from one content to another

Number | Text | Audio | Image/video | Other

Increasing complexity to collect and manage unstructured data

	Internal data	External data
Unstructured data	PDF files, Word files, Emails	Websites, Social media, Mobile data, Sensor data
Structured data	Commercial data, Model calculations, CRM, Data warehouse	Open data, Purchased databases

CREATIVE SOURCING AND USAGE OF DATA

New sources of data

- Additional data can be obtained through many different sources :

  1. **Scraping/parsing techniques:**
     – Extract information automatically from websites
  2. **Open data files:**
     – Structured datasets available to everyone
  3. **IoT sensor and API technologies:**
     – Connected objects and application programming interface
  4. **External data providers**
  5. **Look twice into your own unstructured data:**
     – Reveal hidden information from core data

Data Science can help enhancing the data collection

CREATIVE SOURCING AND USAGE OF DATA

How to enhance data

- Once additional data has been collected, new methods and algorithm allow to get the most out of it. Among others:

1. Statistics, ML and feature engineering:

Create structured dataset using initial datasets or charts to understand data

2. Text mining and NLP

Process of examining large collection of written resources and methods to perform linguistic analysis

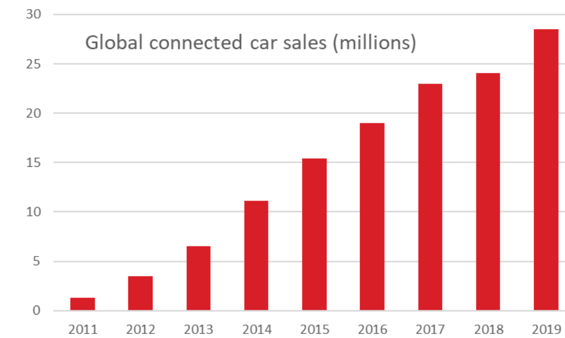
3. Image processing

Techniques to perform operations on images to enhance its content or extract information

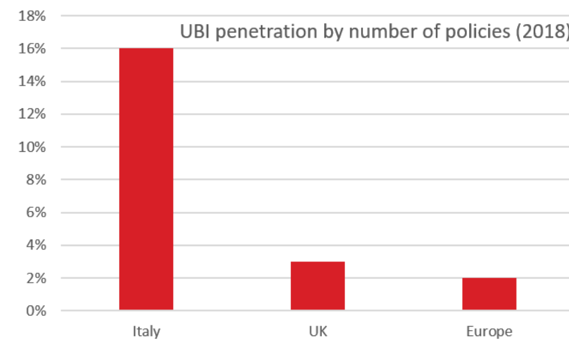
CREATIVE SOURCING AND USAGE OF DATA

Case study: Telematics - Context and overview

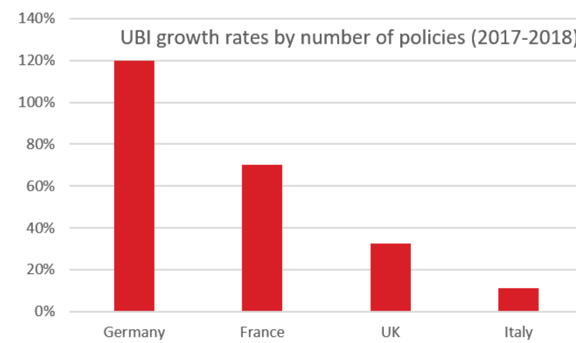
- Development of connected cars :
  - Advanced Driver-Assistance Systems (ADAS) have become increasingly useful and have helped reduce claim frequency
  - Most common tools :
    - Park Assist
    - Forward Collision Warning
    - Blind Spot Information
- Telematic-based insurance has been spreading over the last few years :
  - Italy has relied on telematics for 20 years :
  - Other countries are catching up :
    - In France (in 2017, 55% of insurance companies already provided telematics-based policies or were about to), UK and Germany
    - EU initiatives such as eCall tend to promote connected vehicles, IMCO's new vehicle safety standards
- New opportunities for pricing :
  - Telematics may provide new kinds of data helping insurers improve customers selections and place it at an advantage over competitors.
  - They allow for new and personalized pricing methods such as :
    - Pay-As-You-Drive (premiums depend on when and where policyholders drive)
    - Pay-How-You-Drive (premiums are affected by insureds' driving behaviors)



Source : BloombergNEF, MarkLines, Note : 4Q 2019 estimated



Source : Statista



Source : Statista

CREATIVE SOURCING AND USAGE OF DATA

Case study: Telematics - Opportunities with telematics

- For insurance companies :
  - Collection of useful and large datasets
    - Location : where do drivers use their cars?
    - Time : when do they use their cars? In the daytime, during the night, what kind of weather…
    - Driving style : how do they use their cars? Measures of speed, sudden braking…
  - Improvement of insurance services :
    - Easier claim processes especially if telematics are connected to phone applications
    - Useful advise sent to customers (e.g. in case of dangerous weather conditions, fuel management,…)
  - Reduction of expenses and claim frequencies :
    - More effective claim management processes
    - Auto-selection for safer driving habits
    - Prevention of fraudulent claims and car theft
    - Tying premiums to mileage makes insureds drive less

- Individual pricing
- Better understanding of risks
- Lower asymmetry of information

- Customer Retention

- Improved loss ratio

CREATIVE SOURCING AND USAGE OF DATA

Case study: Telematics - Opportunities with telematics

- For clients :
  - Enhanced communication with insurer
  - Improved claim experience :
    - Digitalized and automatized claim processes via smartphone applications…
  - Additional safety :
    - Reduction of risks of theft of new and expensive cars
    - Real-time tracking of fleets
  - Premium discounts :
    - Compensation to the insureds for collection of data by the insurer

Strengthened customer's satisfaction

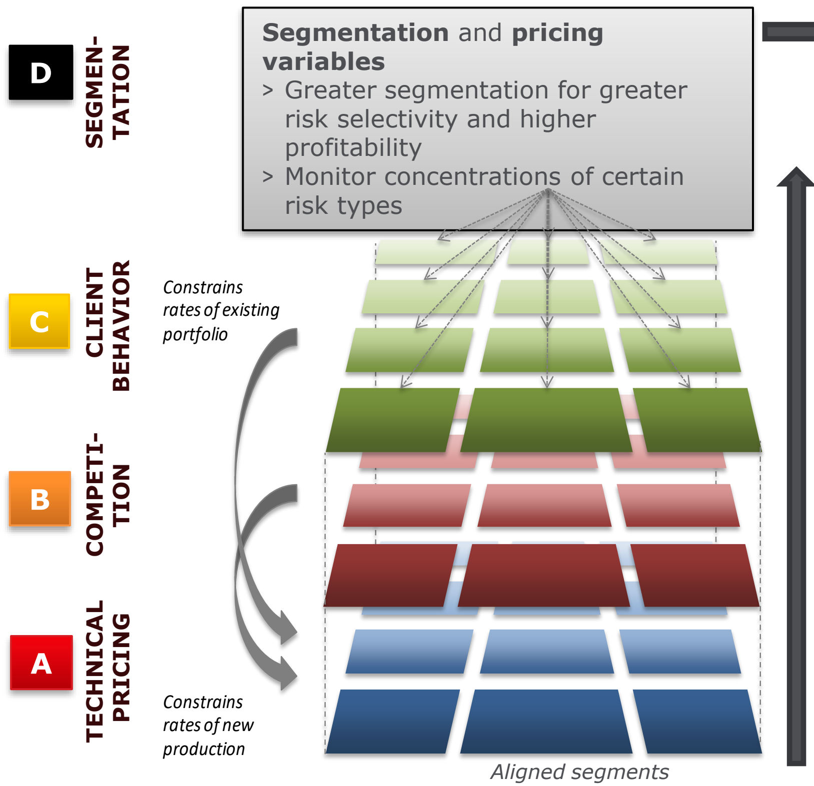
CREATIVE SOURCING AND USAGE OF DATA

External and new data can be used to enrich the existing database with new attributes/variables

- **2 different points of view**
  - Data Scientist and Actuaries: enrich the existing database with a set of features which will be used when calibrating the models. More variables should lead to better predictive models (! Let’s avoid overfitting !)
  - Business users (e.g. underwriter and marketing teams): simplify the models’ results interpretation and/or the processes
    - e.g. reduce forms sizes in underwriting ➔ Quick quote (see example in appendix)
- **What happens if too much data?**
  - Need for feature selection: keeping only the most relevant variables
  - Potential for feature engineering: creating new variables to solve our problem

AGENDA

Technical pricing is not the only application of ML techniques: ML could also help to boost the underwriting and portfolio management process



E SCENARIO TESTING AND OPTIMISATION

Impact of different scenarios on strategic indicators and optimisation

Customer behavior by segment
> Elasticity model help estimate pace at which rates can be increased by segment
> Focus Sales & Marketing to increase retention of better risks
> Building conversion rates model to better target clients

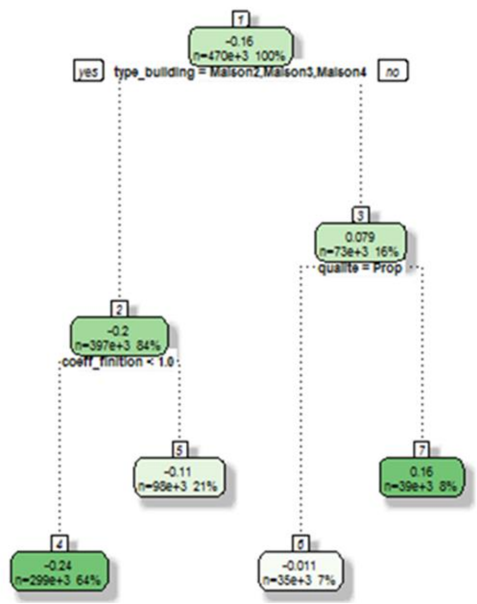
Competitor prices by segment and own current rates
> Position insurer to be more or less competitive on certain segments

Portfolio profitability by segment, based on:
> Cost of risk (e.g. measured through GLMs or ML techniques)
> Portfolio composition (representativeness of each segment: total portfolio and recent new business)

A

Profitability analysis tool: tree-based techniques can be used to compare Risk Premium and Commercial premium

- Thanks to tree-based methods (and variable importance) it is possible to identify the variables that are the most relevant to explain the differences between the risk premium and the current commercial premium
  - It helps in **defining the most relevant variables** that can, for example, then be included in a **profitability heatmap**

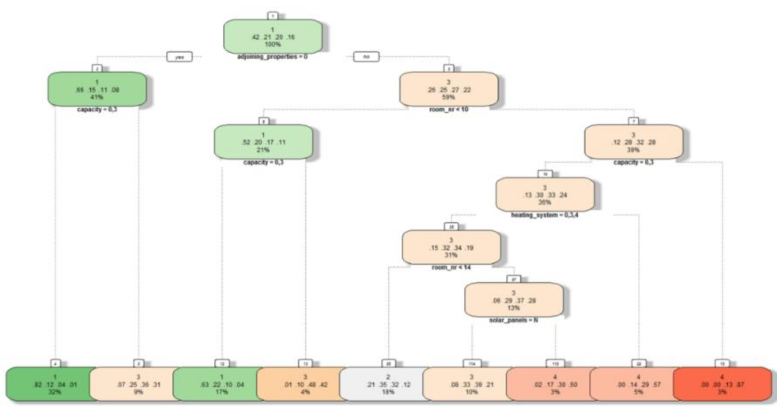


type_building	qualite	coeff_finition A/0.8	B/1	C/1.1	D/1.15	E/1.2
Apparte	Loca	1.43	1.37	1.48	1.63	
	Prop	1.12	1.07	1.16	1.29	
Maison2	Loca	0.99	1.02	1.14	1.16	
	Prop	0.73	0.84	0.94	1.07	1.01
Maison3	Loca	0.92	0.93	1.05	1.14	
	Prop	0.72	0.80	0.90	0.98	0.96
Maison4	Loca	0.99	0.99	1.12	1.20	
	Prop	0.80	0.86	0.96	1.04	1.98

B

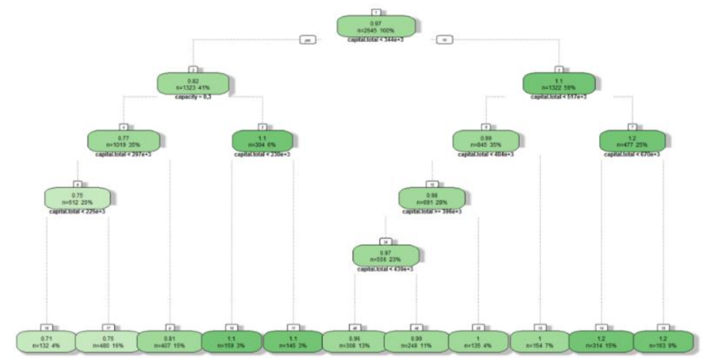
Competition analysis tool: tree-based techniques can be used to identify positioning on market segments and capture price differences

- **Identifying the segments** in which the insurance company is **well-positioned** with respect to its competitors is an important driver of a dynamic pricing process. E.g. **Classification of segments** in function of the **ranking of the competitors** with regression trees



- **Reverse engineering** of the pricing (structure) of competitors can be enhanced with ML techniques

- **Analyse the price dispersion** of the specific company with respect to its competitors of with respect to the average market price

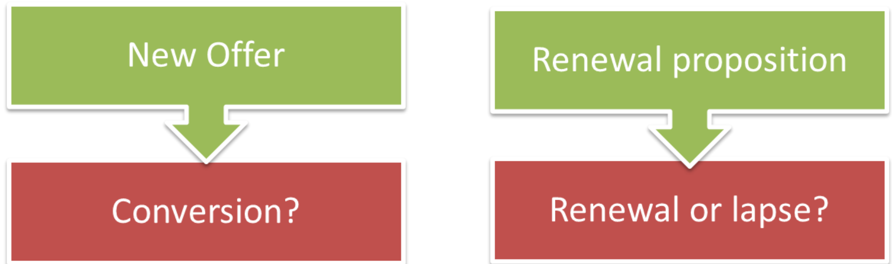


PRICING AND UNDERWRITING

C

Client behavior: ML techniques can help improve the logistic regression

- The goal is to **explain the conversion / lapse probabilities** with some explanatory variables



- A dummy variable identifies the policies that were converted / renewed during the year
- Traditionaly Generalised Linear Models are used
  - E.g. A **logistic regression** can be performed on this dummy variable and potential explanatory variables
- Machine learning technique (e.g. **GBM**) are more and more often used as they usually **improve predictions** and allow to **find more complex patterns**

AGENDA

CLAIMS MANAGEMENT

Claims settlement is maybe the most important task of an insurance company

- With rise of new Insurtech startups and race towards digital transformation, there exists a huge pressure on insurers to rapidly accelerate towards being digital–ready and in particular regarding **claims management**
  - A typical claims management process is replete with huge paperwork, is data-intensive and is often repetitive in nature, which results in loss of time in claims resolution as well as create unwanted inefficiencies.
  - Human handling of claims is bound to be slow and regressive as huge amount of time is spent in preparing data as opposed to analyzing claims, which if left to the new technologies, can drastically solve the issue with ease.

- Insurers are trying to adopt a faster, analytics-driven approach to claims handling (claims analytics) and automate as much as possible the claims handling processes for clear and simple cases.
  - E.g. Fraud detection is an aspect where the new technologies shows huge potential to reduce cost for insurers and help prevent huge sums in fraudulent claims.

- The following tools can be useful in this respect
  - **Intelligent Process Automation (IPA)** can help insurance companies in automating their rule-based, repetitive and frequent tasks
  - **Machine Learning** can help insurance companies create process KPIs and optimize the efficiency
  - **Connected devices (IoT), drone and imaging technologies** are providing insurers with data that can help them assess risk, ascertain damage and settle claims

CLAIMS MANAGEMENT

Typical business cases for data & process analytics

OPERATIONS

- **Mass claims**
  - *Process analytics (STP)*
  - *Intelligent orientation to claims handlers*
  - *Smart routing to bodyshops*
- **Specialized claims**
  - *Recoveries optimisation*
  - *Bodily injuries modeling*
  - *Fraud Detection*

INSURANCE PROCUREMENT

- **Motor**
  - *Cost models*
  - *Network management*
  - *Expert performance*
  - *Smart repair (incl. image recognition)*
- **HouseHold**
  - *Experts performance*
  - *Invoices analytics*
- **Other**
  - *Doctors*
  - *Lawyers*

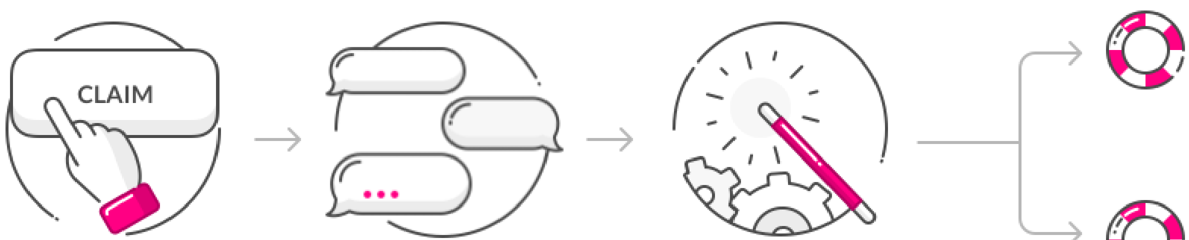
PROJECTS

- **Data**
  - *Financial (claims) datamart*
  - *Claims data exploration (dashboard)*
  - *Segmented opening reserves*
- **Client satisfaction**
  - *Risk alerts (e.g. meteo)*
  - *Qualitative bodyshop models*
- **Business pilote**
  - *Detailed cost analysis*
  - *Stocks & workload*

CLAIMS MANAGEMENT

Case study: Lemonade accelerates claims process

- Lemonade is a US insurance company heavily relying on AI for
  - Underwriting
  - Claim management



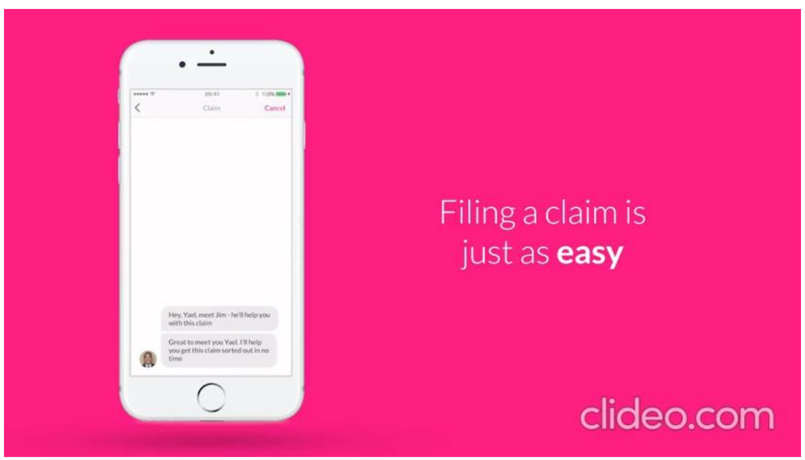
Tap the '**Claim**' button in the Lemonade app

Tell us what happened

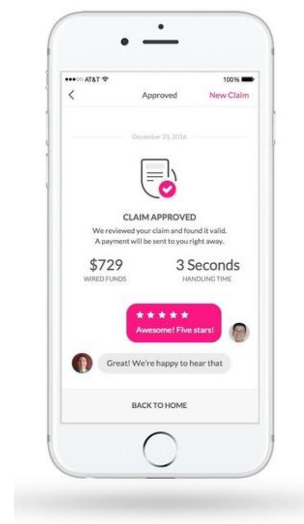
Our AI runs dozens of anti-fraud algorithms

If your claim is **instantly** approved, our AI will pay it in seconds

Otherwise, our AI hands over your claim to our devoted team of **humans** to handle ASAP



Lemonade holds the (non-official) world record with a claim paid in 2 seconds!!!



Thank you !

Do you have questions ?

Know-How to Risk

About us

We develop, in partnership with our clients, actuarial & quantitative financial solutions, building on strong data analytics, while securing full transparency and integral knowledge transfer.

Reacfin
Place de l'Université 25
B-1348 Louvain-la-Neuve (Belgium)
T +32 (0) 10 68 86 07
www.reacfin.com

Xavier Maréchal
CEO
M +32 497 48 98 48
xavier.marechal@reacfin.com

AGENDA

COMPARING TRADITIONAL STATISTICAL INFERENCE AND ML APPROACHES

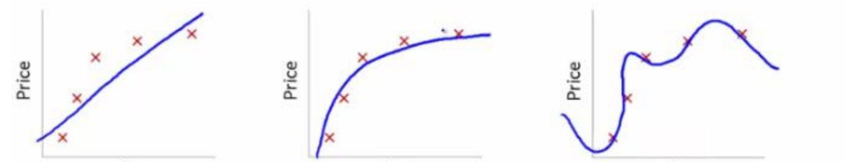
	Machine learning	Statistical modeling
Limits the number of **assumptions**	+	-
Inference: Assessing the reliability of modeling assumptions	-	+
Prediction: ability to extrapolate future or unobserved realisations of a variable given other explanatory observations	+	-/+
“Big Data”: ability to handle large sets of data both in terms of number of observations (“rows”) or variables (“columns”)	+	-
Human interactions: ability/need of incorporating material users ex-ante opinions (e.g. Expert Judgment)	-	+

- Results of Machine Learning algorithms will need careful attentions as they derive from automated procedures and could induce conclusions which do not match a business logic.
- Another key challenge with Machine Learning is the risk of overfitting.
  - Overfitting relates to excessively complex models for which the large number of explanatory variables and parameters, is unreasonably important compared to the number of observations

Overfitting deteriorates the predictive power of the model

The overfitting problem

- When modelling, we should be sensibilized with overfitting/lack of parcimony.
- It occurs when a statistical model describes random error or noise instead of the underlying relationship.
- The fact that the model fits our data well doesn't guarantee it will be a good fit to new data ➔ A good model is one that fits also well new data, i.e. that has a small predictive error



Bias-Variance Trade-off

- The **Prediction Error** can be decomposed as follows

$$E\left[\left(Y-\hat{Y}\right)^2\right] = \underbrace{\left(E[Y]-E[\hat{Y}]\right)^2}_{\text{Bias}} + \underbrace{Var\left(\hat{Y}\right)}_{\text{Estimation Variance}} + \underbrace{Var(Y)}_{\text{Pure randomness}}$$

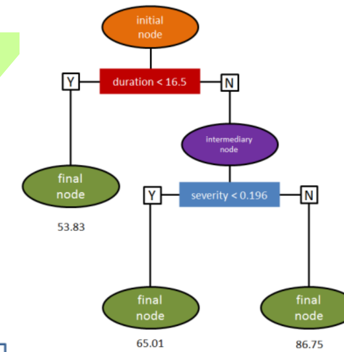
- In general, we try to minimize simultaneously the bias and the estimation variance in order to get accurate predictions.
  - Usually, these two terms compete in the sense that a decrease in one of them typically leads to an increase in the other one.
  - This phenomenon is known as the bias-variance trade-off for which one needs to find a good balance (typically by controlling the complexity of the model).

SOME MACHINE LEARNING TECHNIQUES ARE BLACK BOXES AND INTERPRETATION OF THE RESULTS CAN BE QUITE DIFFICULT

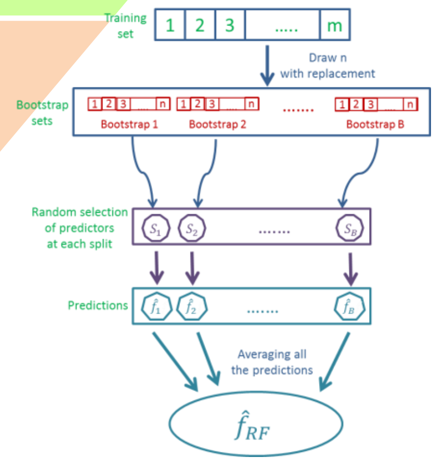
Understanding the results of ML techniques is not easy

Complexity

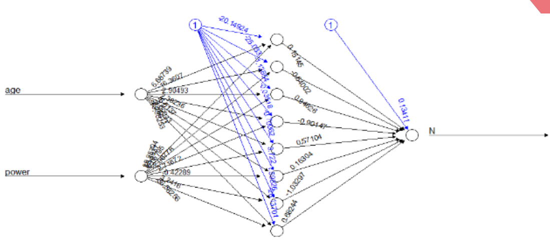
In the case of regression trees, understanding how the model predicts claims' cost or frequency values for new data points is not a problem, as it is very intuitive.



In the case of more complex methods such as Bagging and Random forests, even understanding how the model predicts values for new data points is rather difficult.



Things may be even worse for GBM and NN.



Interpretability

UNDERSTANDING THE RESULTS OF ML MODELS IS NEVERTHELESS KEY FOR SOUND BUSINESS DECISION-MAKING AS MANY STAKEHOLDERS USE THE RESULTS OF THE MODELS

Quant (Actuaries, data scientist,…)

- Able to understand the technical details
- Trust its outputs based on cross-validation, error measures and assesment plots

Other stakeholders

- Not necessarily « quantitative people »
- Should nevertheless understand and trust results to take decisions

Machine learning techniques usually improve predictive power but at the expense of a certain loss of interpretability ➔ Find trade-off between

- Predictive power
- Capacity to understand the results
- Ability to take sound decisions based on the results

High-end questions

Who will use the results? For what purpose? With which impact?

Pricing Fairness challenge for insurance companies

Customer segmentation

- A fair premium, related to his/her risk profile, to minimize the potential for adverse selection.
- *i.e. the good risks could lapse and accept a lower premium elsewhere, leaving the insurer with an inadequately priced portfolio.*

Keeping pricing fairness :

Big data and ML models could lead to an increased segmentation among policyholders which has to be managed as well (to avoid un-insurability of some risks)

Risk pooling

- The use of machine learning for pricing should not lead to an extreme personalization of risk/premium
- *E.g. extremely high premiums for some risk profiles that actually imply no risk transfer.*
- The insurer has the social role of creating solidarity among the policyholders.

AGENDA

WHAT IS MACHINE LEARNING?

Objectives of Machine Learning ("ML")

ML algorithms aim at finding by themselves the method that best predicts the outcome of the studied phenomenon.

Supervised vs. Unsupervised learning

- **Supervised learning:**
  - Inputs and examples of their desired outputs are provided
  - The goal is to learn a **general rule that maps inputs to outputs**.

➔ *Given a set of training examples ($x_1, x_2, \ldots, x_n, y$), where y is the variable to be predicted, what is the most efficient algorithm to best approximate the realizations of y*

  - 2 main techniques
    - **Classification** : outputs are divided into two or more classes, and the learner must produce a model that assigns unseen inputs to one (or multi-label classification) or more of these classes.
    - **Regression**: the outputs are continuous rather than discrete.

- **Unsupervised learning:**
  - No labels are given to the learning algorithm
  - The goal is to **find structure in its input** (discovering hidden patterns in data).
  - Main technique
    - **Clustering**: a set of inputs is to be divided into groups. Unlike in classification, the groups may not be known beforehand.

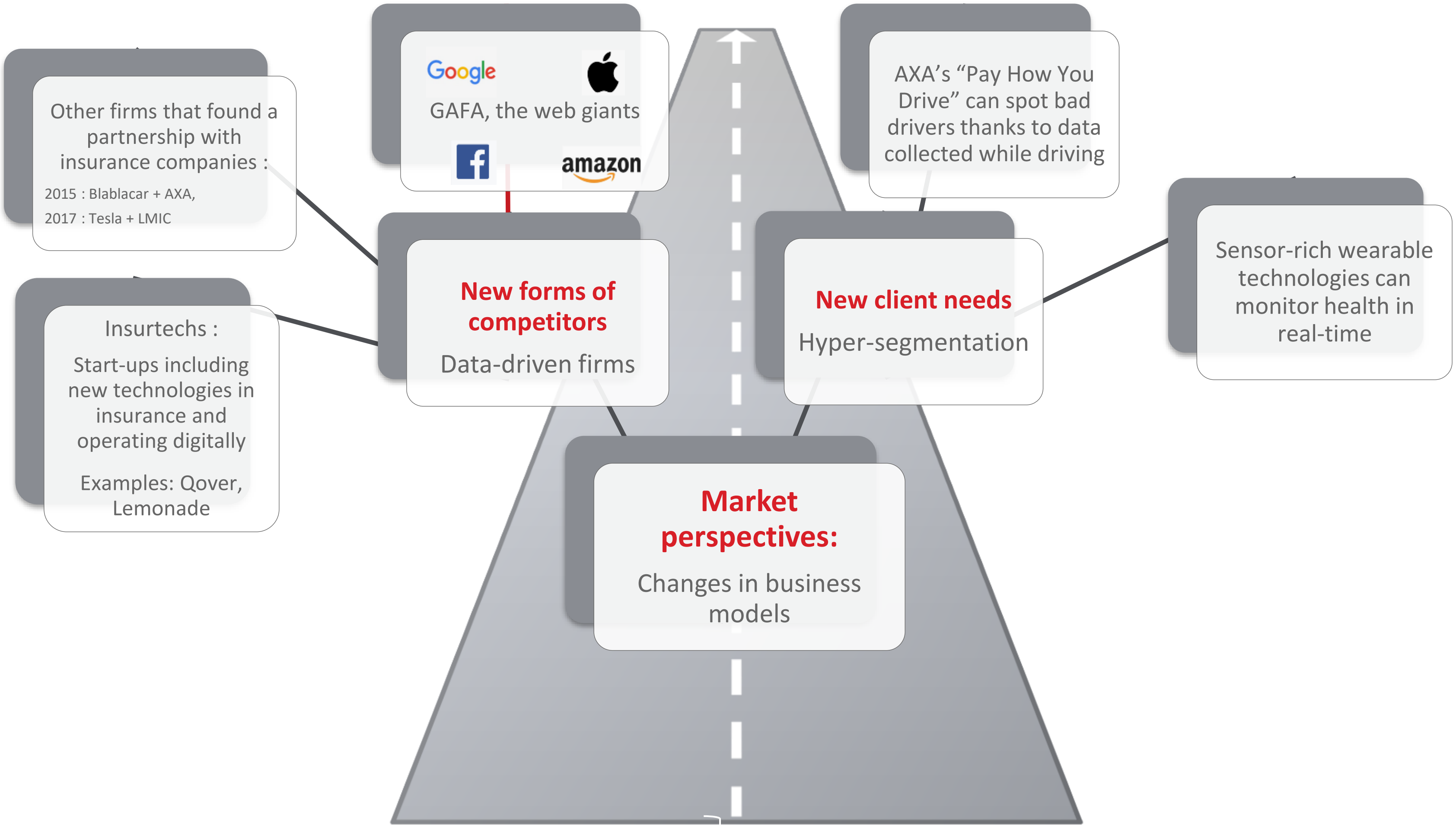
Main use in motor insurance

- Typically used to model **pricing or underwriting related variables**
  - Regression: frequency (#claims) or severity (claims cost)
  - Classification: lapse rates, conversion rates

- Typically used for **features engineering** (i.e. creating new variables)
  - E.g. vehicle classification, zoning,…

TRENDS IN NON-LIFE INSURANCE

Insurtechs and GAFAs are putting pressure on the market



TRENDS IN NON-LIFE INSURANCE

Digitalisation and availability of data makes possible the use of other business models

On-demand insurance

- Insurance coverage adapted to customers' needs (for very short periods, or episodic insurance…)
- Aims at targeting young clients who have different habits and ask for more flexibility

➡ **Towards service model (renting instead of owning)**

E.g.: Wrisk, AvivaPlus, Cuvva, Trov,…

Usage-based insurance

- Pricing methodology based on client's habits rather than on the type of risk they belong to
- Based on new data sources (e.g telematics for cars)
- For car insurance
  - Pay-As-You-Drive (PAYD) : depending on mileage
  - Pay-How-You-Drive (PHYD) : depending on driving style

E.g.: InsureTheBox, Thingco, Ticker, ByMiles,…

Sharing economy insurance

- One of the largest growth area of insurance solutions in the coming years
- More flexible insurance solutions are needed as the distinction between personal and commercial usage has blurred
- Even peer-to-peer insurance models have been considered

E.g.: Dinghy, SafeShare, Slice Labs,…

Embedded insurance

- Insurance contracts added to purchase of non-insurance items (e. g car insurance coupled with purchase)

➡ **Not a new trend, yet it is likely to extend quickly along with change toward sharing economy and insurance as a service**

E.g.: BlaBlaCar, Uber, Canopy,…

CREATIVE SOURCING AND USAGE OF DATA

Case study: Telematics - Issues and threats related to the spread of telematics in insurance

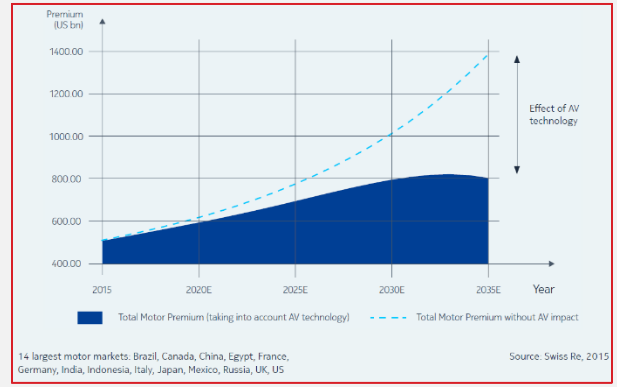
- Insurers need time to process and benefit from this collection of new data:
  - These datasets are very different from traditional self-reported variables used for pricing (e.g. age, car model, postal code…)
  - UBI is a very different pricing methodology : real-time pricing not only relies on past claims history and characteristics of policyholders but also adapts to current behaviors

- Concerns about private data collection :
  - So far, some profiles of drivers are more willing to use telematics than others (e.g. young people accustomed to data collection and digitalized services)
  - Drivers are really looking forward to using connected devices for their own benefit, but not always ready for insurance utilization
  - Clients may ask for compensation in return for data collection (such as premium discounts)

CREATIVE SOURCING AND USAGE OF DATA

Case study: Telematics - Impact of telematics and autonomous cars on car insurance premium

- The spread of telematics and ADAS will help reduce costs and frequency of claims
  - Thanks to automated car technology leading to safer roads, motor insurance premiums for the 14 major markets could be trimmed by USB 600 billion by 2035*



- Studies have shown that using telematics may help reduce some costs and claim frequency :
  - In Italy (the first European country to promote usage of telematics), a study showed that black boxes have lead to 20% decrease in claim frequency and 6% in claim processes expenses **
  - Telematic-based insurance promotes safer driving styles (reduction in hard braking, hard acceleration, speeding ; drivers tend to adopt more eco-driving habits)

Telematics lead to both

- Adverse selection : good drivers are more likely to choose telematic-based insurance
- Reduction of risky behaviors

Yet, relevance of UBI should be questioned by taking into account all benefits (lower risk) and costs (cost of devices are borne by insurers)

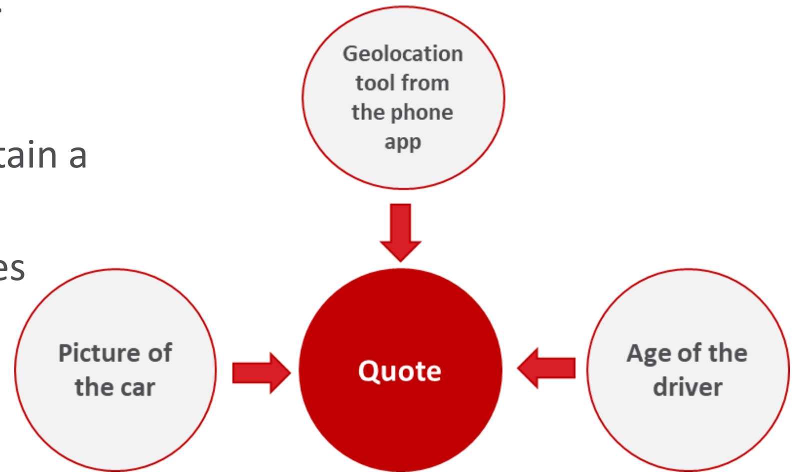
* Swiss RE : The future of motor insurance, 2016

** Sergio Desantis and Gianni Giuli (ANIA)

CREATIVE SOURCING AND USAGE OF DATA

Case study: simplifying quoting process – Quick Quote by Generali

- In 2018, Generali Belgium launched **Quick Quote Car** a phone app that helps new customers to obtain a car insurance quote within 60 seconds
  - The app is available for insurance brokers who can obtain a quote with the following data
    - Picture of the car : Image recognition technologies help identify the license plate and the main characteristics of the car
    - Geolocation
    - Age of the driver



- Later in 2018, Generali also launched a similar app (Quick Quote Home) to help brokers calculate home insurance quote based on:
  - Pictures of the house :
    - Based on collaboration with a start-up specialized in real estate valuation (Rockestate), a 3D simulation of the house is made
  - Geolocation

NB: regulatory context has evolved since this initiative

Features Engineering & Selection

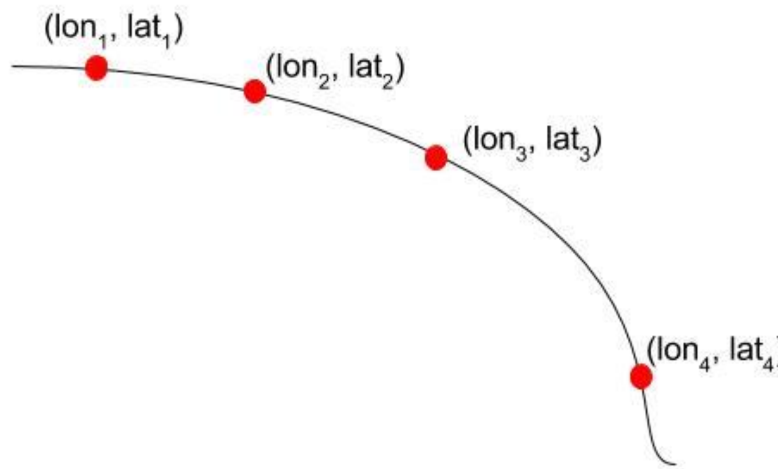
- A feature is an attribute that is useful or meaningful to your problem
- Features Engineering is absolutely known and agreed to be key to success in applied machine learning

> *"At the end of the day, some machine learning projects succeed and some fail. What makes the difference? Easily the most important factor is the features used." — Prof. Pedro Domingos*
>
> *"Feature engineering is the process of transforming raw data into features that better represent the underlying problem to the predictive models, resulting in improved model accuracy on unseen data." — Dr. Jason Brownlee*

- Features Engineering is a Representation Problem
  - Machine learning algorithms learn a solution to a problem from sample data.
  - In this context, feature engineering asks: what is the best representation of the sample data to learn a solution to my problem?

Case study: Features engineering in property theft insurance

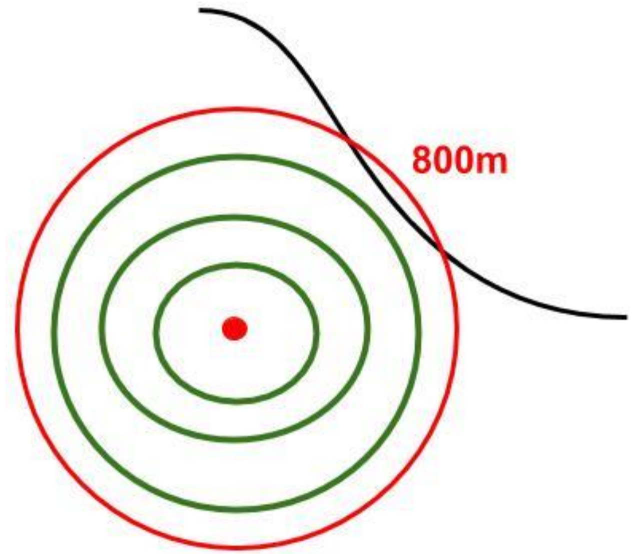
- Intuition : there is a correlation between the claims frequency and the distance from the highway
  - Data available in the company : addresses
  - Features Engineering : convert house addresses into distance from the highway
- Highway only?
  - No, all the roads where the speed limitation is above 90km/h
- Determine the closest point to the highway in relation to the house.
  - We need to know the location of the house on a map
  - We need to know the location of the highway on a map
- Open Street Maps
  - Gives the roads' longitude and latitude at different points.



CREATIVE SOURCING AND USAGE OF DATA

Case study: Features engineering in property theft insurance

- If we want to find the distance from the house, we need the coordinates of the house.
- Google Maps Geocoding API
  - Geocoding is the process of converting addresses into geographic coordinates
- Google API is free but with slow performances:
  - 2,500 free requests per day
  - 50 requests per second (limitation of speed)
  - Enable pay-as-you-go billing to unlock higher quotas: $0.50 USD / 1000 additional requests, up to 100,000 daily.
- Find the distance between the house and the first road (above 90km/h).
  - We build a loop that checks if there is a road in a growing area (in a radius growing from 0 to 4000m with step of 200m)



Methods used in non-life pricing are evolving at a fast pace

- Machine learning and AI is the continuation of the evolution of tools and technologies used by actuaries and statisticians to analyze historical claims data: trying to improve the predictive power of models, solving the same problems with new methods, data and computer power available

Data Requirements & Model Complexity

Purpose	Methods	Data
Traditional Dashboard and Reporting		
Detect what happened and where it happened	Basic indicators (average, trend, max, …), univariate analysis	Accounting figures, data derived from Business intelligence tools
Basic Statistics		
Understand the process generating the data (emphasis on inference), explain how did it happen	Distribution assumptions and fit, Extreme Value methods (EVT) for large losses	Individual losses data but with few (or no) additional co-variates
Forecasting		
Try to forecast global indicators according to past value	Time series analysis : SARIMA, GARCH models, …	Time series data of the indicator to model, no individual data required
Statistical Predictive Modelling		
Understand and predict the individual data with more granularity	Regression models : GLM, GAM, Tree, GLMM, …	Large sample of individual losses data with more co-variates
Machine Learning		
Focus on prediction : minimize the prediction error using many co-variates and their potential interactions	Random Forest, Neural Networks, Gradient Boosting Methods, SVM, …	Large sample of individual losses data with many co-variates

Case study: Garage selection for reparations and relevance of Smart repair

- **Goal**: incorporating external sources of data (from Car Damages Experts and Garages) to better select garages for car accidents by:
  - Improving the prediction for reparation costs of each claim (considering characteristics of automotive accidents, the specialties of [nearby] garages, the transportation costs for the damaged vehicle)
  - Assessing, in particular, the effectiveness of “Smart Repair” solutions for specific cases of claims (i.e. categorizing car damages for their likelihood to be repaired at reduced costs if “smart repair” garages selected).
    - Usually « Smart Repair » garages will charge higher man-hour fees (making them thus prohibitive in standard models)
    - « Smart Repair » allows not to replace car parts (hence making in some cases the costs of reparation much cheaper

- **Challenges**:
  - Need to access third party data (incl. Historical data),
  - Standardize them (across all sources), and
  - Incorporate them in the existing data bases / data management framework of the insurer

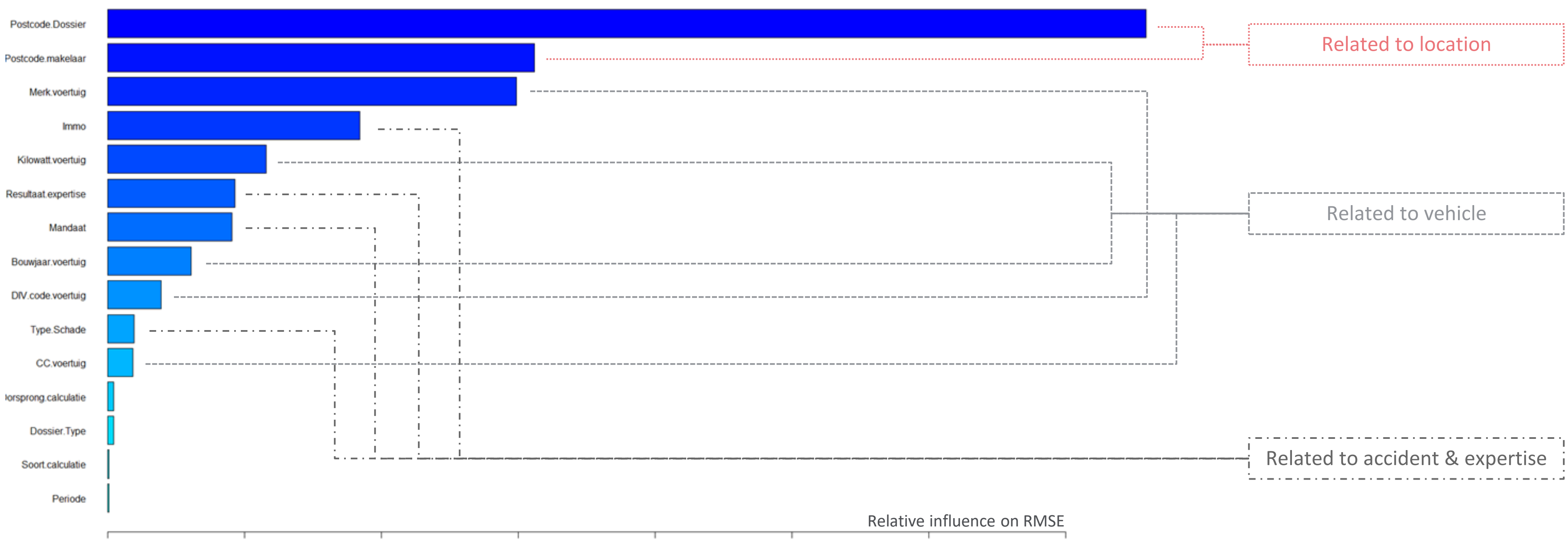
- **Predictive models used**:
  - Machine Learning models (Random Forest, Gradient Boosting Machines, Neural Networks)

Sanitized interim results
Based on partial data set for illustration purposes only

Case study: Garage selection for reparations and relevance of Smart repair (illustration)

Key variables in Gradient Boosting Models for the prediction of total cost of repairs

Most influential variables on the Total cost of repairs



CLAIMS MANAGEMENT

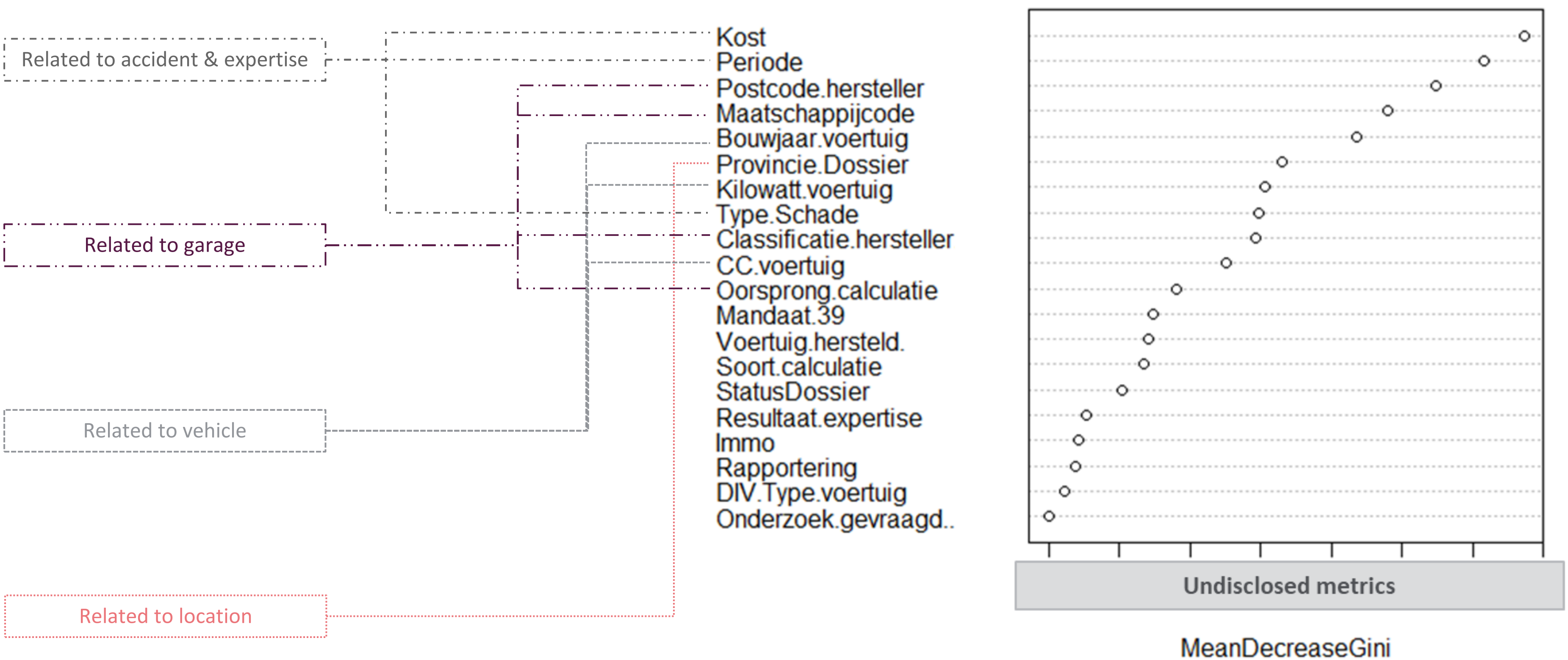
Sanitized interim results
Based on partial data set for illustration purposes only

Case study: Garage selection for reparations and relevance of Smart repair (illustration)

Variable importance of key variables to predict lower total costs using Smart Repair solutions

Type of variable

illustrative results using Random Forest



DISCLAIMER

The recipient of this document should treat all information as strictly confidential and only in the context stated below. Information may not be disclosed to any third party without the prior join-consent of Reacfin.

Estimates given in this presentation are based on our current knowledge, they can be based upon our previous experience within the Undertaking, as well as taking into account similar projects in the same context as the Undertaking, either locally, within majority of the EU countries as well as overseas.

This presentation is only the supporting document of a verbal presentation. Hence, it is not intended to be exhaustive. Quoting or using this document on its own might be misleading. As a result, these materials may not be used by anybody except their authors nor should they be relied upon in any way for any purpose other than as contemplated by joint written agreement with Reacfin.

Reacfin

Place de l'Université 25
B-1348 Louvain-la-Neuve
www.reacfin.com