

06/09/2015

BIG DATA PROCESS, MACHINE LEARNING AND ERM APPROACH

V. Ranaivozanany, Qualified and IA-certified actuary, ERM CERA expert
C. Atchama, Qualified and IA-certified actuary
C. Laurans, Qualified and IA-certified actuary

Corresponding author : Mrs Voahirana RANAIVOZANANY

e-mail : voahirana.ranaivozanany@cnp.fr

*Mailing address : CNP assurances - Direction Technique Groupe – 4, place Raoul Dautry -
75015 PARIS.*

insuring all
our futures



Abstract

This study proposes the use of technological and methodological advances to monitor group benefits plans. The standard monitoring method has reached its limits in the current context, where reactivity and precision are becoming essential.

Supervising and monitoring a group benefits portfolio is based on a comparison of claims expenses with premiums received to cover the claims.

Traditional monitoring consists in annually analysing claims-to-premiums ratios and, depending on the size of the portfolio, the analysis is performed by combining all or some of the customers (group plans). Under pressure from a highly competitive market, pricing margins are nonetheless small, and quickly identifying deviations in claims rates has become a priority. In addition, to be competitive, the monitoring process differentiates between customers who weaken the portfolio and those who improve it. Furthermore, this monitoring would be more efficient if price elasticity was considered according to customer profiles.

Established with a prototype, the process recommended by the study is based on a "big data" type approach, constructed using "machine learning" methods [1]. "Big data" seeks to make the most of the huge volume of data available in the portfolio, and open data can also be included to improve monitoring. Compared with conventional statistical methods, "machine learning" algorithms can process huge volumes of data. They also have predictive and prescriptive qualities.

A descriptive analysis was used to better comprehend the portfolio claims ratio. Beyond clearly identified profitability recovery actions, accurate knowledge of the profiles of beneficial or disadvantageous customers has enabled a predictive analysis to be performed on the profitability of new business. A prescriptive analysis is still currently under being studied to enhance the recovery strategy by taking price elasticity into account.

The indicators we analyse could be more responsive if they were automated and regularly updated. Established with fine granularity, they make for more accurate control than conventional indicators. In a prospective way, they also let one anticipate and refine the monitoring. The last part of the study shows the value of integrating indicators and monitoring principles established in the insurance company's ERM concept. Such an approach will produce KRIs (Key Risk Indicators) and KPIs (Key Performance Indicators) that are more relevant to insurance risk.

It will also help us better integrate portfolio monitoring with the overall strategy of the company. A notion of risk tolerance associated with portfolio hedging is introduced. This tolerance to risk is translated into operational limitations for established indicators. We propose a reporting table to monitor indicators on an ongoing basis.

This study aims to underline the advantages of such an approach. It needs to be completed and adapted to the reality and needs of portfolios.



Summary

- 01.** General context of this study
- 02.** The asset of machine learning approach
- 03.** ERM concept
- 04.** Conclusion

For remember, the study and the views expressed here are solely those of the author. The results of the study described in this presentation have been changed to protect confidentiality

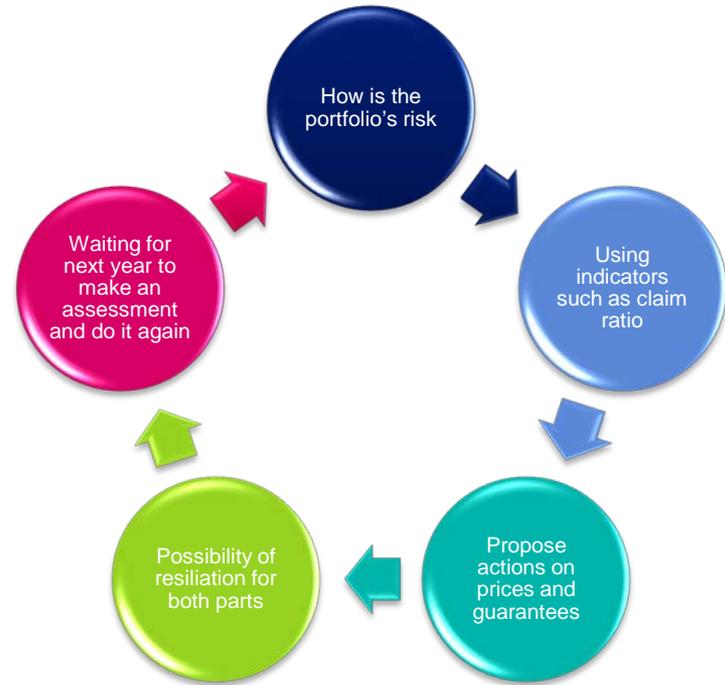
01.

General context of this study

- a. Problematic**
- b. Insurer's data : an unvaluable element**
- c. Why using machine learning ?**

Problematic

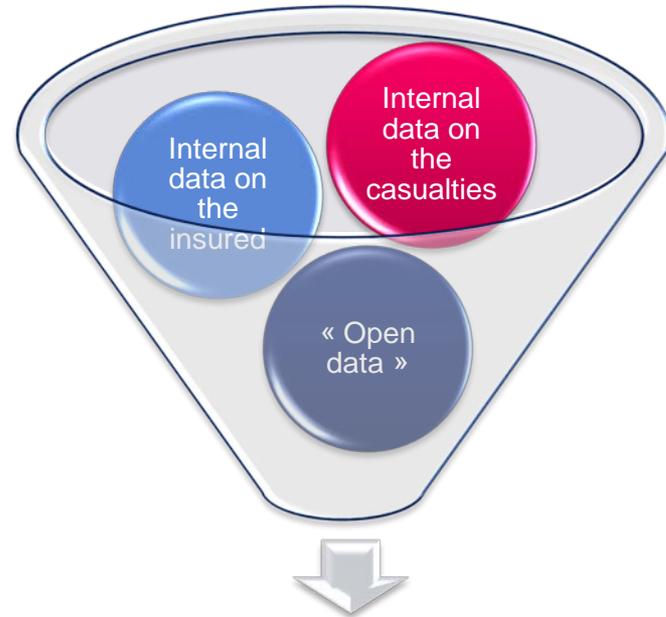
- An insurer needs to evaluate with accuracy his own risk
- Claim ratio is an indicator which gives a limited information belonging to the past
- What can we do to be more proactive on the portfolio's monitoring ?
- We need an evolution of this classical approach
- Ask ourselves the good questions



How
Tools **New indicators**
Statistical methods
Machine Learning
Big Data

Insurer's data : an unvaluable element

- “Nothing is lost, nothing is created, everything is transformed”, Lavoisier's quote
- Insurer's data are the best means to reach the path of the risk's comprehension
- Structuring the data is a fundamental step. Insurers have to reach high qualities standards for their data
- Extend our internal data with open data to push the limit of third « V » Gartner's 3V model



Make these data « talk » to have a better comprehension

Why using machine learning ?

- Technology's advances give us the opportunity to use these methods nowadays
- Methods that can process big volumes of data. Totally big data's compliant
- A lot of freeware includes a machine learning function. Creating an ease for non-initiated person
- Compared to classical statistical methods, we can do more with description, prediction and prescription

VELOCITY

We have now powerful computer

VOLUME

Storage and performance of information system

VARIETY

Insurer has built database with lots of features and has the possibility to extend with open data

02.

Assets of machine learning approach

- a. A method for each issue**
- b. Descriptive approach**
- c. Predictive approach**
- d. Prescriptive approach**

A method for each issue

Searching for avenues of recovery

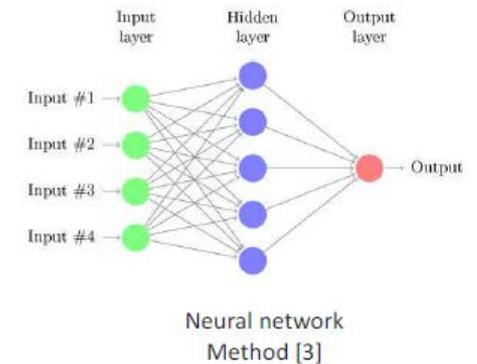
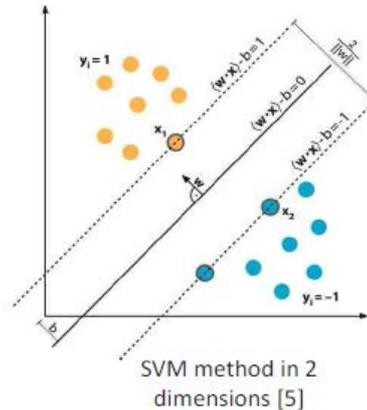
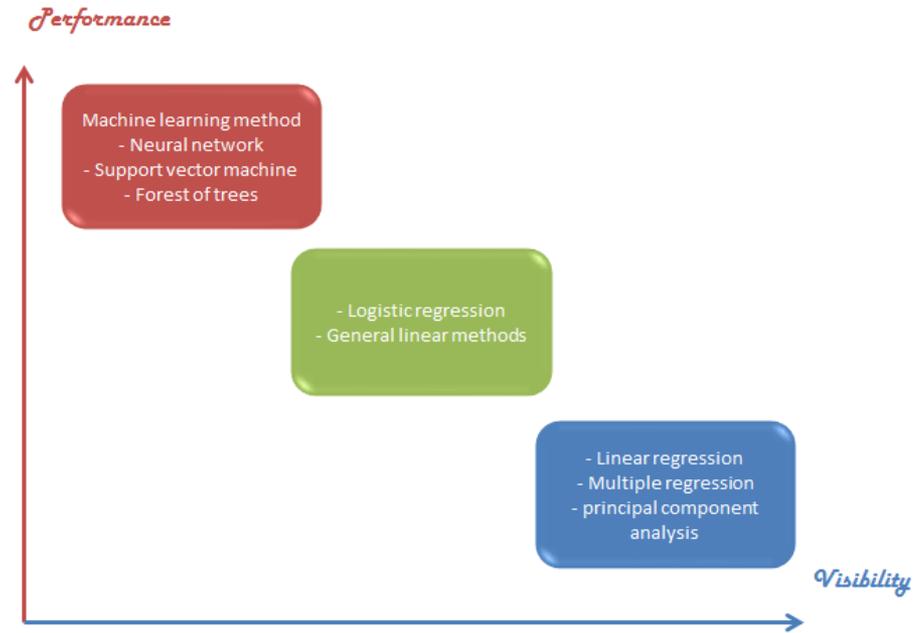
	DESCRIPTIVE	PREDICTIVE	PRESCRIPTIVE
	What happened?	What could happen?	What should happen?
What we need to do	Understand why the portfolio shows a high claims ratio	Estimate the potential profitability of new subscriptions	Steer portfolio recovery policy
What we need to know	The company claims ratio profile The level of claims ratios in each region	Recognize companies that improve and/or weaken the portfolio	How will companies react to recovery policy decisions
What Analytics can contribute	Segmentation of the portfolio according to customers' levels of claims ratios Reporting of claims ratios and penetration rates per department or region, etc. (reflected on a map of France)	Predictive model – Estimate the probable claims ratio of a newly subscribed company	According to the adopted policy, simulate the behaviour of companies Optimization – What would the best recovery policy be
What makes this analysis possible	Bulk processing and standardization of data Principal component analysis, ascending hierarchical classification validated by Machine Learning	Machine Learning with current data	Machine Learning with historical data

Direction of improvement of added values



A method for each issue

- Compromise between visibility and performance, which method is appropriate ?
- Complementary methods for analysis
- 3 Machine learning methods used :
 - Support Vector Machine, schematically isolating each class in space by creating borders
 - Neural Network, inputs are translated into an output, which creates a class to be assigned to
 - Forest of trees, combination of N trees, each trees links an element to a class and the majority wins

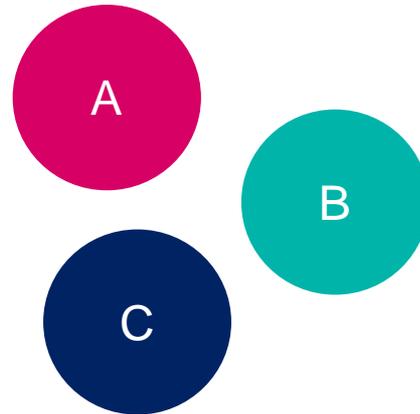
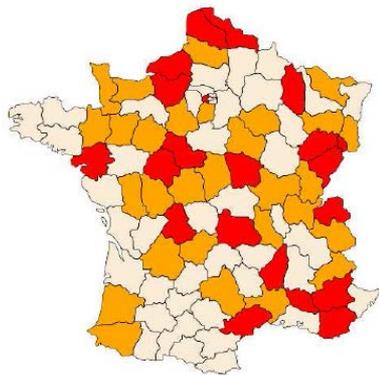


Referred to references [3] and [5]

Descriptive approach in three steps

Understanding our portfolio

- ▶ Descriptives statistics
- ▶ Data visualisation

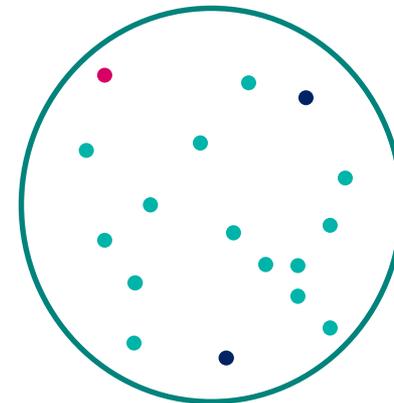


Test of these potential segmentations

Results and quality of the classification

Choosing the potential segmentations

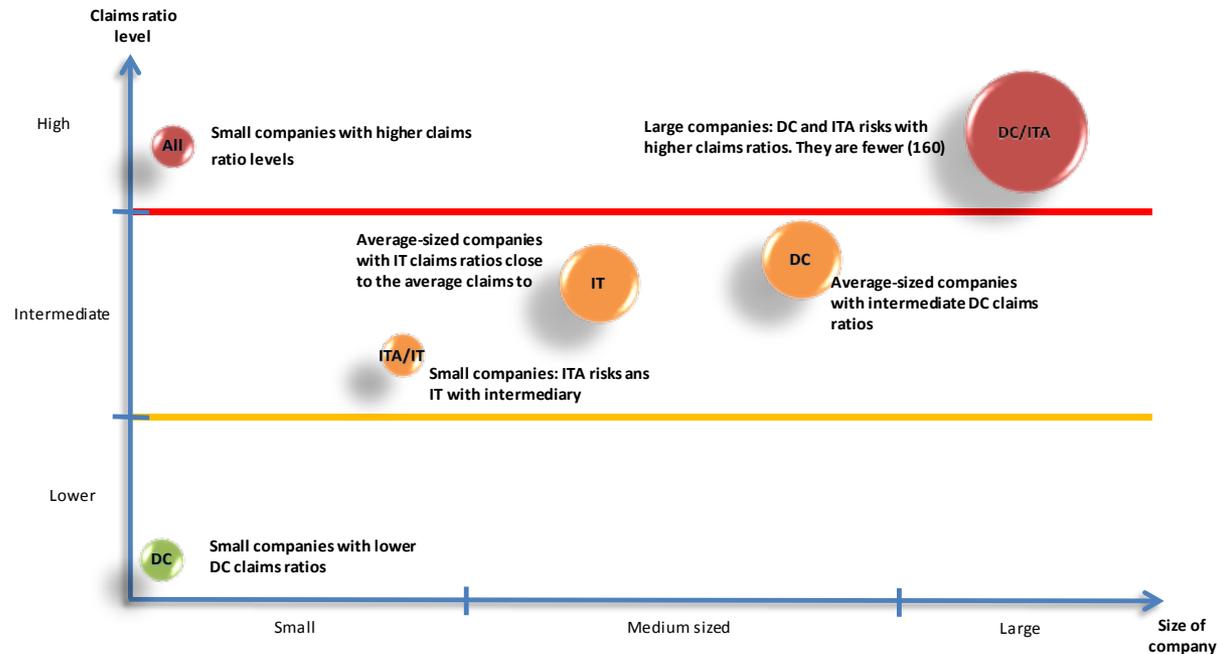
- ▶ PCA
- ▶ Cluster
- ▶ Supervised gathering



$\varepsilon < 5\%$

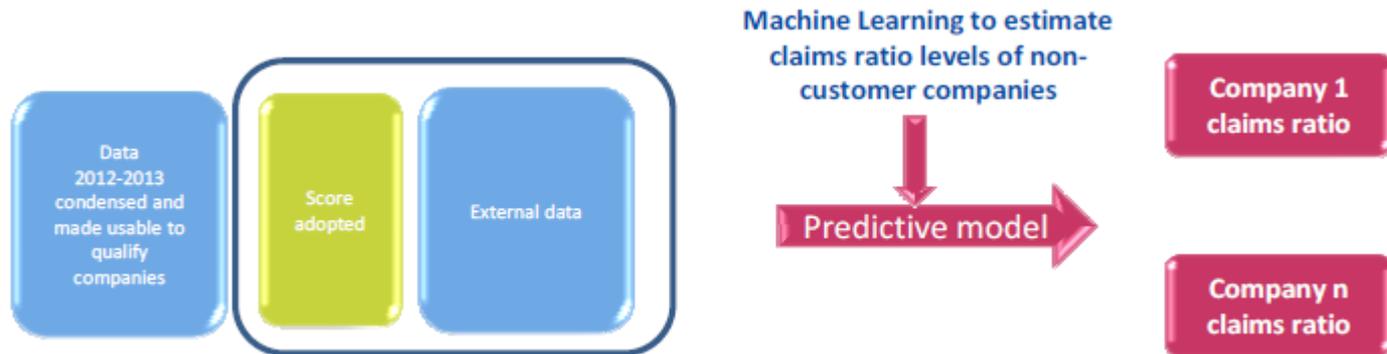
Group B

Descriptive approach



- The completed segmentation identified the number of employees in the company as the most “discriminating” characteristic
- There are few companies with more than 500 employees (200 out of the 15,000) but they have an impact on the economy of the portfolio
- These identified companies have not standard pricing
- How using these results to appreciate the profitability of future business ?

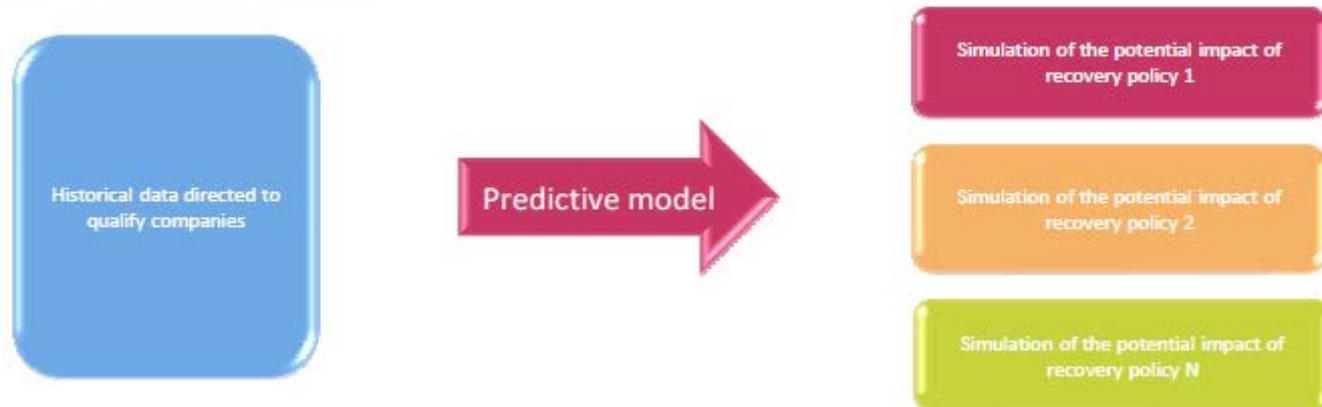
Predictive approach



- The aim is to get an idea of companies not yet in the portfolio that could improve or weaken it
- The characteristics of companies improving or weakening the portfolio have been identified
- A predictive analysis to appreciate the potential profitability of your new prospective customers
- These initial opinions on prospect are considerable working advantages for the sales teams

Prescriptive approach – adjustment of prices

Application of Machine Learning



- The aim is to find a sales policy with the best profitability
- The past pricing increases have not always had the expected effects
- Recovery actions did not take into account the circumstances of customers in a sufficiently customized manner
- “Machine learning” algorithms can simulate the reactions of companies in portfolio regarding the price recovery scenarios

03.

ERM concept

- a. Introducing ERM concept**
- b. ERM concept illustration**

Introducing ERM concept

All works have to be focused on the needs of company's governance :

- Appetite
- Tolerance
- Limits

Construction of Key risk indicators (KRI) and Key performance indicators (KPI) like tools for monitoring and decision making.

Key risk indicators

- Monitoring risk exposure
- Monitoring companies that improve or weaken the portfolio

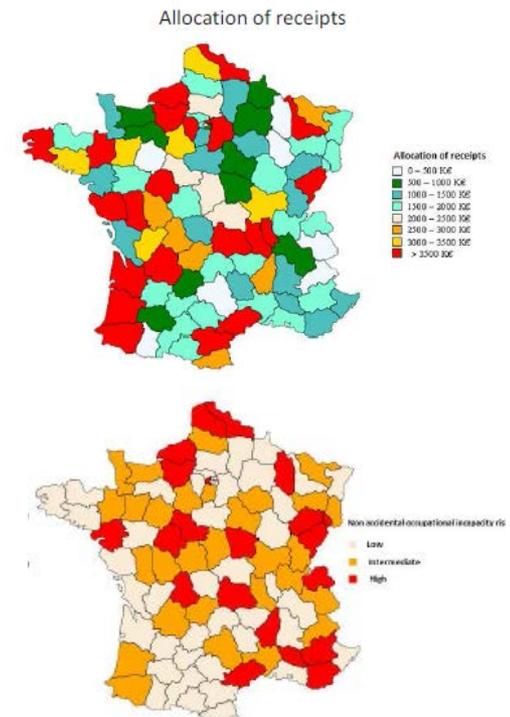
Key performance indicators

- Pricing performance monitoring


Appetite
Overall level of risk an undertaking is prepared to take.
Set by the Board of Directors.

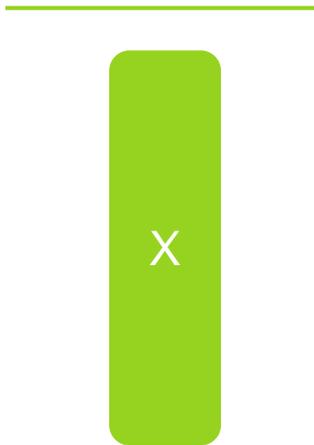

Tolerance
Level of risk set per risk class.
Stated in terms of exposure to the risk.


Limits
Operational limits enabling the undertaking to manage its risk profile.
Articulation between the optimal level of risk and strategic objectives.



ERM concept illustration

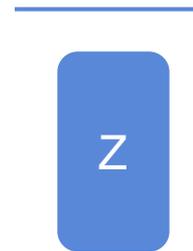
Claim ratio < 100%



5 neighboring area with more 1 million premium



Number of companies that weaken the portfolio



Translation of these three indicators



- Transform individual limits into a global limit
- Define risk level for each indicator and which impact if we exceed them
- Alert to engage management actions

04.

Conclusion

- **Automation to converge to a real time approach**
- **Responsiveness and accuracy are a considerable advantage in a competitive market**
- **New questions could emerge from this approach**
- **Machine learning as a virtuous circle**
- **Minimizing hazard and accepting to take some risk to maintain a insurance activity**

References

- [1] Bellina R., Delucinge S., Taillieu F. (2014) : Méthodes d'apprentissage statistique – « MachineLearning »
- [2] Dubois D., Ranaivozanany V. (2014) : Diffusion d'une culture ERM au sein d'une organisation :une démarche complexe mais réalisable.
- [3] <http://www.texample.net/tikz/examples/neural-network/>
- [4] IBM Software - Thought Leadership White Paper (2013) : Descriptive, predictive, prescriptive:Transforming asset and facilities management with analytics.
- [5] Université de Bonn : <http://www.precision-crop-protection.uni-bonn.de/>
- [6] Gartner IT Glossary : <http://www.gartner.com/it-glossary/big-data>



insuring all our futures