Behavioral patterns of long term saving :

Predictive analysis of adverse behaviors on a savings portfolio

assurer tous
les avenirs

CNP
assurances

Introduction

What is the context of this case study and what about the underlying challenges ?

assurer tous
les avenirs

CNP
assurances

Introduction : Presentation of speakers

Babacar SOW,

- **Engineer, data scientist and actuary, Executive Assistant of the group CAO of CNP Assurances, member of the French** *Institut des actuaires*. **From all of his 8 years experience, has conducted several mentoring on data science and actuarial modeling.**

Thomas BEHAR,

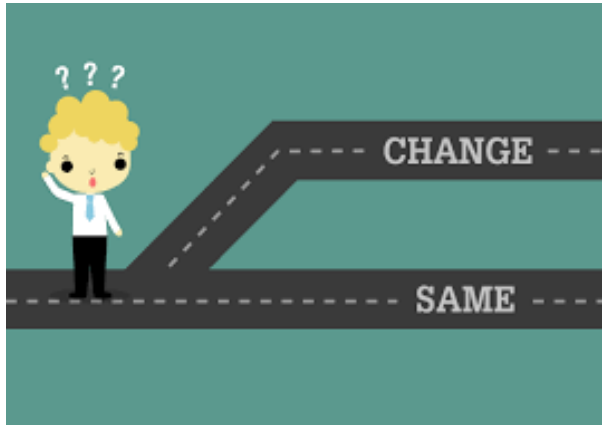
- **Group Chief Actuary Officer and member of CNP Assurances Executive Committee, 2 times president of the French** *Institut des actuaires*, **Current President of the AAE.**

Introduction : the context

- **The context of the study, seen by a life insurance company in the French market**

Low interest rates	**Less margin, higher cost of capital on €uro funds**	**Exogenous factor …**
Strategic orientation on UL funds	**Macro economic consequences …**	**Transfer of risk from us to customers** → Who is UL appetent ?
A sudden rise on IR	**Higher risk of lapse / Shift toward new financial investments**	**Commercial retention programs** → Cost of retention programs ?
Geopolitical environment and probable occurrence of volatility on Equities	**Shift toward €uro funds as a "safe investment"**	**Anticipation of future actions** → Who will start **shifting from UL funds**?

Introduction : the goal



Objective

↓

How to reduce the behavioral risk ?

↓

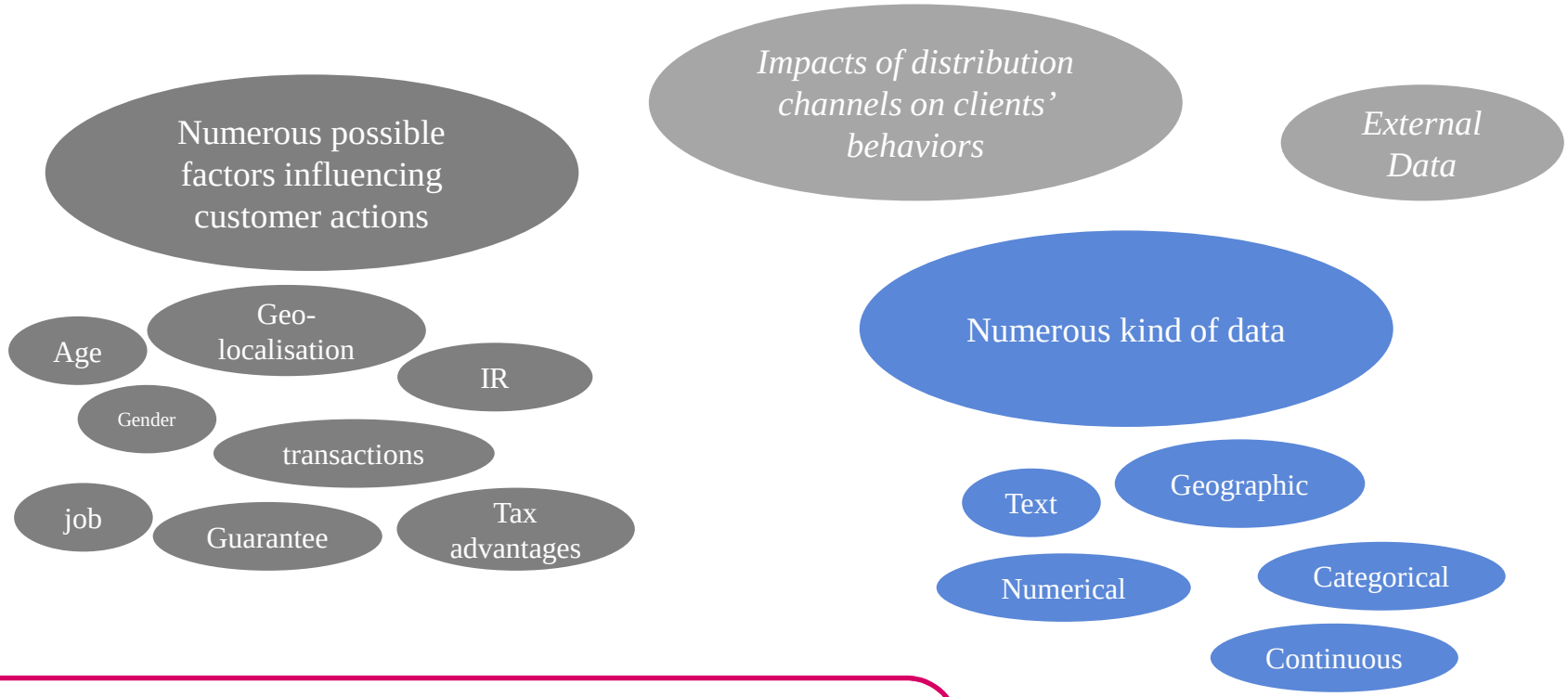
Predictive Analysis

Remarks:

1. **In this paper, we propose a set of "recipes" for supervised learning methods applied on a life insurers' portfolio**. Many constraints could appear, due to their database structure underlying the long term risk they hold. *As recalled in our paper, literatures in this field offer a wealth of information and some techniques related in this document are quite abundant. However, life actuaries may have difficulties to find out a way to realize in concrete terms such a study on their portfolio*.

Introduction : challenges

- **What challenges in building a predictive model on life insurance data ?**



- **Expected results**

✓ A good model performance (with stability over time)

✓ A population predicted for commercial actions

✓ A better understanding of the insurance liability

Sommaire

- The insurance product under review
- Method and approach
- Feature Engineering
- Model Training
- Result
  - Standard Model Performance
  - Sequential Rules
  - Truncated Model Performance
  - Model Comparison
- Conclusion

The insurance product

What characteristics and what choices of the insurance product to facilitate the proof of concept ?

assurer tous
les avenirs

CNP
assurances

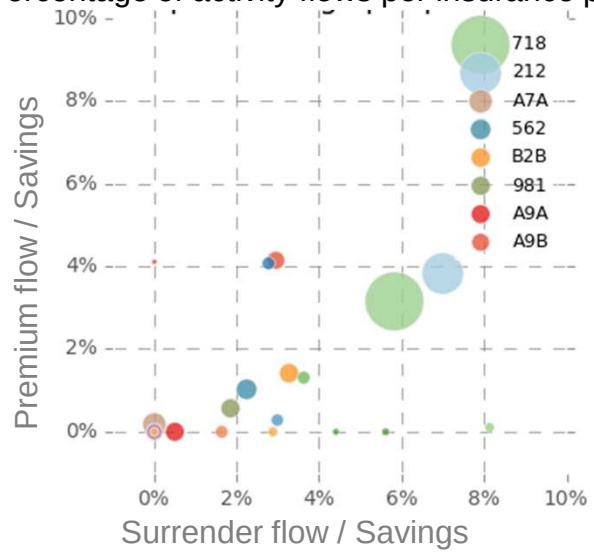
Choice of insurance product

Why did we choose this product for learning purpose ?

1) A multi support savings product with a high volume among upmarket products

2) An active portfolio

Percentage of activity flows per insurance product



Risk over 2016 / Action	Number of clients	Total cash flow
Surrender	9005 (19%)	755 m€
Arbitrage	1516 (3,2%)	62,7 m€

Choice of insurance product

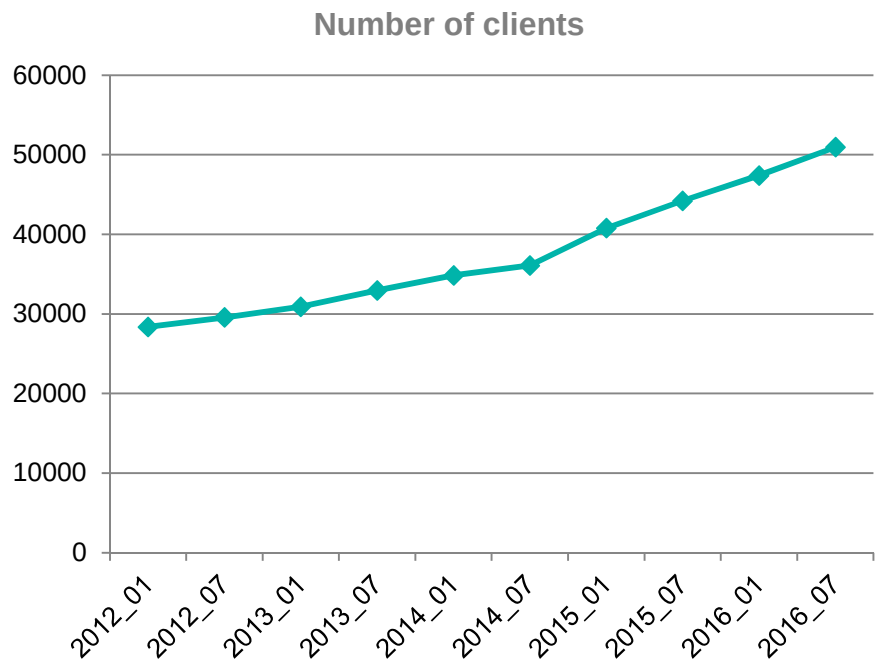
Why did we choose this product for learning purpose ?

1) High volume among upmarket products

2) Active portfolio

3) Growing client base

4) 12 years of history



Remarks:

1. Do not be naive, … ML techniques just reveals hidden information on data. Then without substantial and consistent data, no value added (*See P53*)

Method and Approach

How could we summarize the ML process for this type of case study ? What can be the results in concrete terms ?

assurer tous
les avenirs

CNP
assurances

Method and Approach : a supervised learning problem

Problem definition

$$\left(X_{(n,d)} = \left(x_1, x_2, \ldots, x_d \right), Y_{(\mathrm{n},1)} \right)$$

Predictor (input) variables

Predicted (target) variable

Choosing algorithms and training

$$\hat{f}(\theta) \in C \ , \ \hat{f} = \arg\min_{\mathrm{f}} L\left(f(X), Y \right)$$

A learning algorithm tries to find best parameters θ for a such prediction function

Loss function that maps $Y \ and \ \hat{Y} = \hat{f}(X)$

Assessing and validating predictions

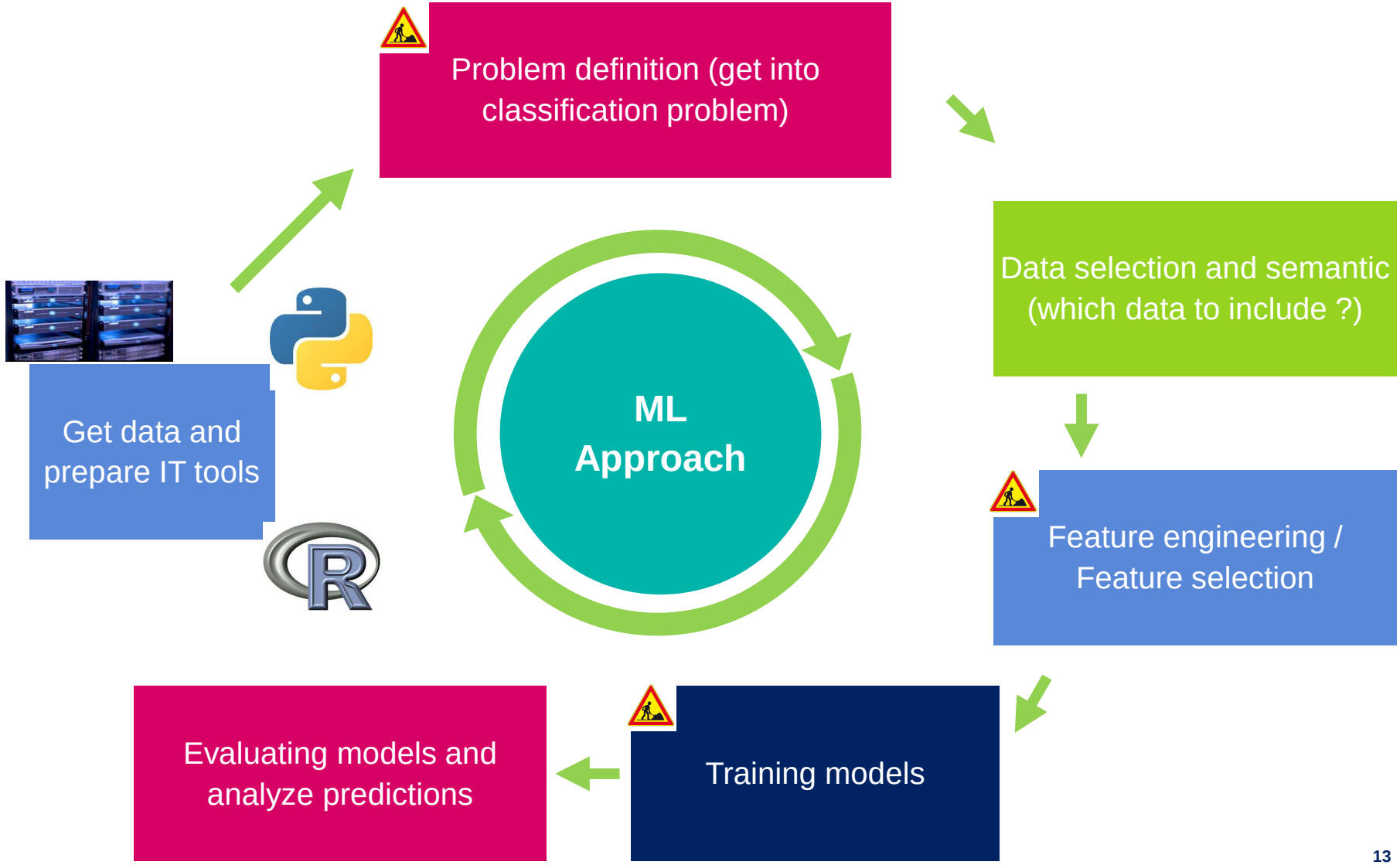
$$M\left(\hat{Y} = \hat{f}(X), Y \right)$$

Use a Metric to conclude about the prediction $\hat{Y} \ with \ respect \ to \ Y$

The Goal .. Mathematically speaking:

Given a training database $\{x_i, y_i\}_n$, what prediction y_j for a new observation x_j ?

Method and Approach : In concrete terms …



Method and Approach : A concrete definition of the target

What reading of an “Adverse Behavior ” in a such savings portfolio ?

1) Surrender

2) Arbitrage UL->EUR

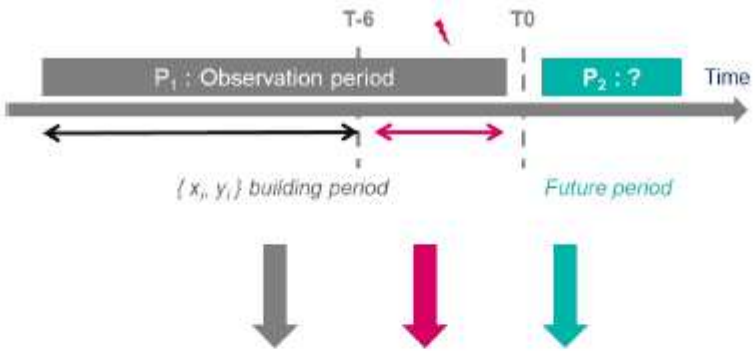


Figure 1: Construction of the target

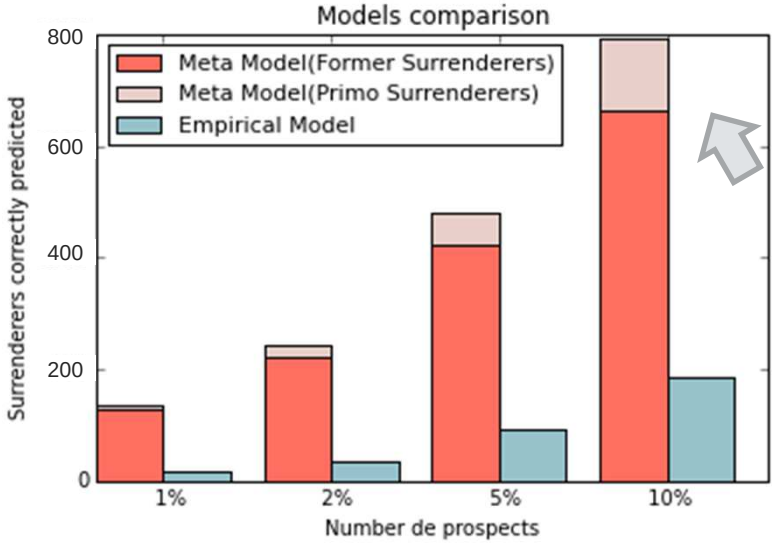
$X = \{x_1, x_2, ..., x_d\}$	$\{y_i\}$	$Y_{unknown}$
Data from client 1	1	1
Data from client 2	0	0
...	...	...
Data from client n	1	1

Table 1: y=1 if an adverse action occurs (unknown at time T_0) on Period P_2, otherwise 0

Operational challenge : Predict movements over the next 6 months using 3 years of history

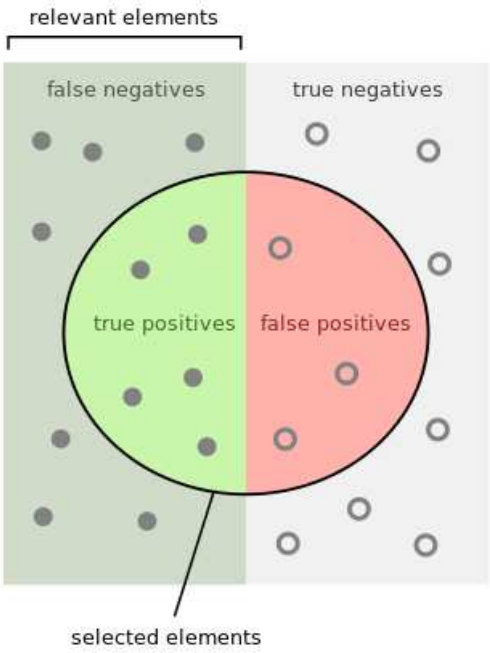
Method and Approach : A concrete result from predictive analysis

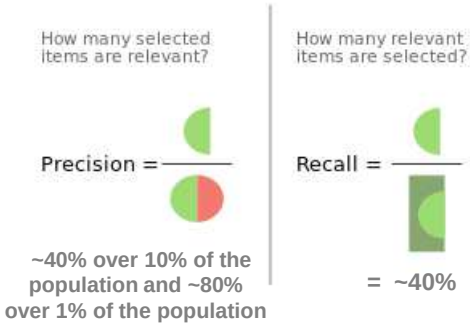
Comparison with the empirical model (*past surrenders deduct future surrenders*):



Recall : Prediction of **40% of surrenders** by an intelligent choice of **10% of the population**

Performance of predictions





Remarks:

1. **The predictive model targets** not only "Former Surrenders" but also **"Primo Surrenders" (*~20%*)** who could **never been targeted by traditional descriptive methods**.
2. The predictive model provides a recall of about ~40% and a high precision.

Source of the Picture : Wikipedia

Feature Engineering

assurer tous
les avenirs

CNP
assurances

Feature engineering

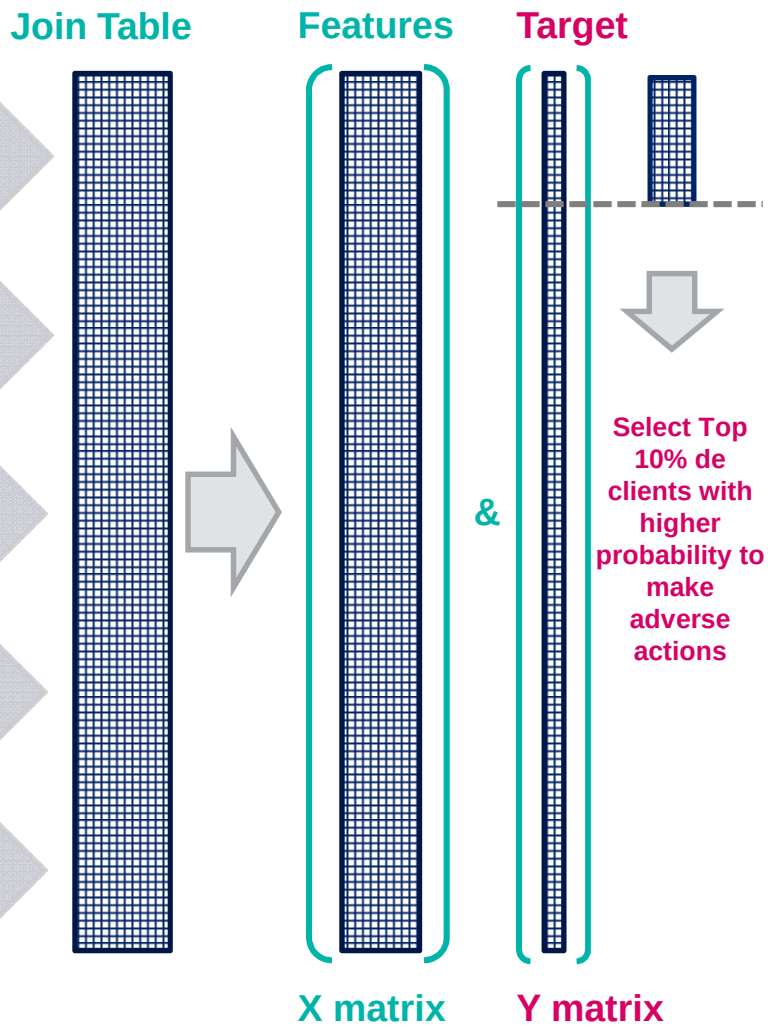
Identity data
- Surname, First name, Postal code, Sex, Family situation...
- Duplicated for each time step

Contract data
- Product segmentation, tax frame, subscription date, subscriber & beneficiary designation…
- Grouped through smaller categories over one hot encoding

Financial data
- Savings information / Mathematical reserve
- Considered per time period & Distinguished by technical type of investment (UL or €)

Transaction data
- Different management acts : Premiums, Surrender, Arbitrage (UL_€, GSM_GL, €_UL, UL_UL)
- Aggregated by RFM (Recency, Frequency, Monetary)
- Sequence Mining

External data
- INSEE, financial indexes, public authorities publications, …
- Considered per time period & Merged to Identity data



Remarks:

1. A first challenge of getting into a Classification problem within a classical life insurance data base. Methods especially marked red would be presented in following slides.

Feature engineering – Contract data

Contract_ number	Subscription date	Marketing channel
718025624	2010-08-06	Distribution channel_ 1
718025627	2010-08-06	Distribution channel_ 1
718038211	2012-11-28	Distribution channel_ 1

Relativize & One hot encoding

Contract_ number	Seniority	Distribution channel_ 1	Distribution channel_ 2
718025624	6	1	0
718025627	6	1	0
718038211	4	1	0

Aggregate several contracts to one person

Client ID	Seniority_ max	Seniority_ min	#	# on Distribution channel_ 1	# on Distribution channel_ 2
XXX_1950-10-16	6	6	1	1	0
YYY_1938-12-22	6	6	1	1	0
ZZZ_1964-07-22	4	4	1	1	0

Remarks:

1. We can notice that the number of features could considerably increase.

Feature engineering – Transaction data

Contract_ number	Effective_ Date	Amount	Action type	Technical_ product_ type	Technical product_ duration
718035187	20160101	400.00	RACHAT PARTIEL	EUR	D
718038735	20160501	1000.00	RACHAT PARTIEL	EUR	D

Period tag	Client ID	Monetary_ RACHAT_ PARTIEL_ EUR_D	Frequency_ RACHAT_ PARTIEL_ EUR_D	Recency_ RACHAT_ PARTIEL_ EUR_D
201601_201606	XXX_1952-09-06	400	1	61
	YYY_1938-12-22	1000	1	181
	ZZZ_1962-03-11	0	0	10000

RFM Model :

Recency – time gap between the last action and the end of interval

Frequency – total times of the action

Monetary – total cash flow of the action

Remarks:

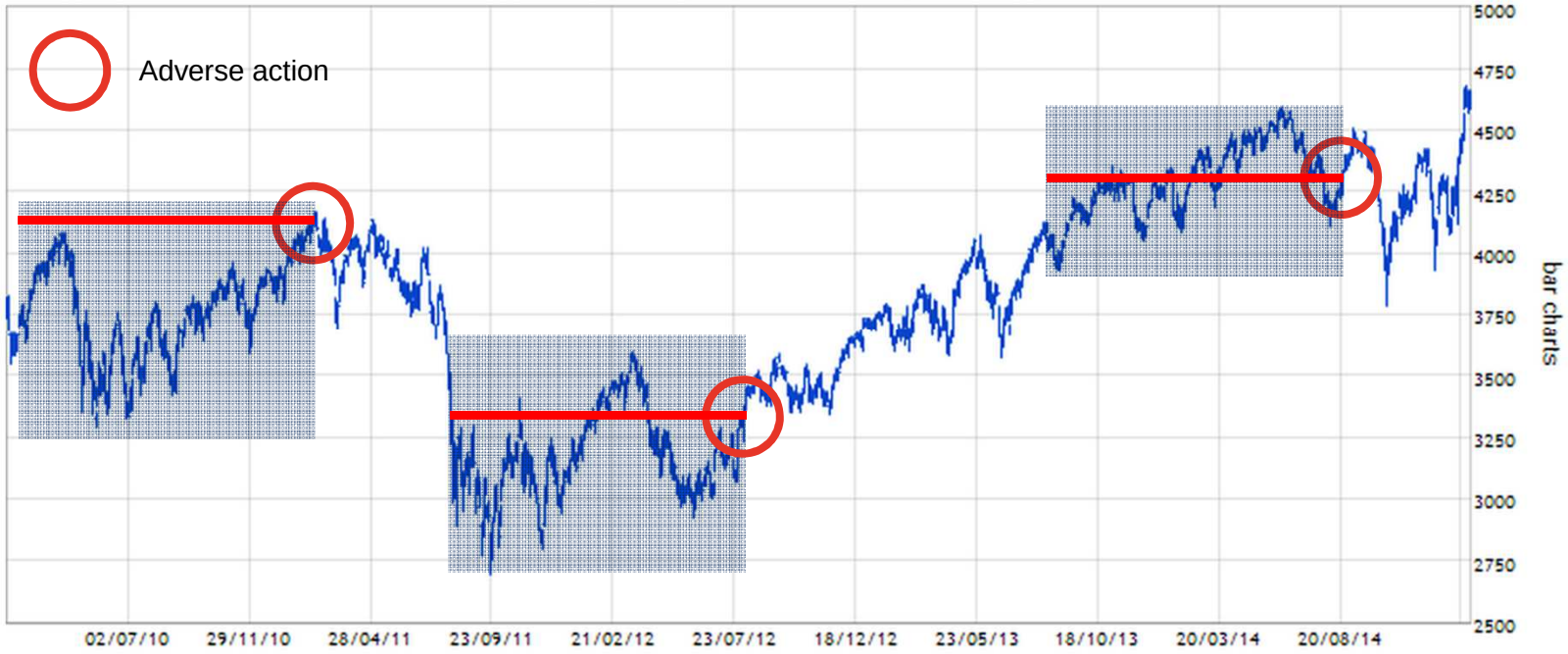
1. The RFM model summarizes the information into a suitable way for a classical classification algorithm. For example, in this case study, 19 new features are created for the management act "*Partial Surrender*" (*9 from Frequency & Monetary scope by filtering on different time periods – 6 semesters and 3 distinct years – since we consider 3 years history and one feature from the Recency*)

Feature engineering – Time series treatment

Which treatment for transactional data like financial indexes ?

For each adverse action :

- ✓ Observe the current value of the index
- ✓ Observe the past values of the index into a sliding window (here 12 months)
- ✓ Use the decile corresponding to the index' value at the adverse action date as a feature (*then at least 10 new features are created)*



We capture on that way the behavior of the client with respect to a wide economic context

Feature engineering - Sequence Mining (1/4)

Adverse behavior : **What happens on large retailers when mining frequent sequences from transactional databases?**

A collection of receipt for each client i

Mining sequences and find frequent subsequences among all customer sequences

```
Algorithm GSP(D, ρ_s, I)
  C_1 := {⟨{i}⟩ | i ∈ I}
  Scan D to get support of every sequence in C_1
  L_1 := {s | s ∈ C_1, sup(s) >= ρ_s}
  i := 1
  while (L_i ≠ ∅)
    C_{i+1} := GSP_Gen(L_i)
    Scan D to get support of every sequence in C_{i+1}
    L_{i+1} := {s | s ∈ C_{i+1}, sup(s) >= ρ_s}
    i := i + 1
  Return L_1 ∪ L_2 ∪ ... L_{i-1}
```

Apply an action for the corresponding clients

- Review dispositions of products through the different sections of the store
- Push new products to clients
- …

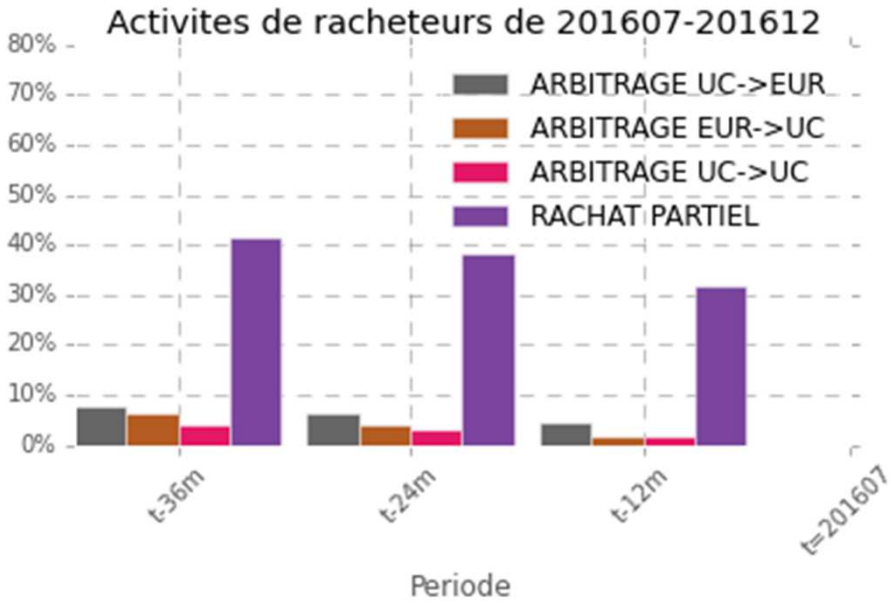
Remarks:

1. The same algorithm applied in our savings portfolio could be run by using **analogies** (*see P33*)

Feature engineering - Sequence Mining (2/4)

Why did we use Sequence Mining ?

1. Finding sequential rules regardless of the predictive model (*increase knowledge on the portfolio*).
2. Adding information into the predictive model :
   a. RFM : does not keep information on the action orders
   b. RFM : visualization of information is not easy



Contract_ number	Effective_ Date	Amount	Action_ type	Technical product_ type	Technical _product _duration
718035187	20160101	400.00	RACHAT PARTIEL	EUR	D
718038735	20160501	1000.00	RACHAT PARTIEL	EUR	D

Sort by time & Label actions

ID	Action Sequence
Client 1	A,B,C
Client 2	B,D,C,A
Client 3	A,B,D
Client 4	C,A,B,D

Feature engineering - Sequence Mining (3/4)

Sequence feature generation:

To include more information than a binary table, we include the cash flow of subsequences. There are several choices available at this stage:

- For each client, several matches are possible for one sub-sequence

  Ex. 3 possibilities when matching (*A,B*) to (A_1, B_1, A_2, B_2) : (A_1, B_1), (A_1, B_2), (A_2, B_2)

- Once matches the subsequences, such as aggregating the cash flows of client actions.

  Ex. Sum, Mean, Max, Min, etc...

In this project, we chose to match the most recent subsequence for a client – so (A_2, B_2) – and we choose the sum of cash flows of all actions as aggregation function.

ID	Action Sequence
Client 1	A,B,C
Client 2	B,D,C,A
Client 3	A,B,D
Client 4	C,A,B,D

ID	A	B	C	D	A,B	B,D	C,A	A,B,D
Client 1	200	30	50	0	230	0	0	0
Client 2	300	40	40	50	0	90	340	0
Client 3	500	600	30	30	1100	630	0	1130
Client 4	40	30	40	100	70	130	80	170

Remarks:

1. More **information** about **actions' order in time** are captured. But the number of features keep growing …
2. This point is treated through "**Feature selection**" technics (*iterative running of models and selection of the most powerful features* / with respect to predictive performances).

Feature engineering - Sequence Mining (4/4)

Modeling client actions:

To include as much information as possible other than RFM features. We've also decomposed all the actions by cash flow on the technical products:

F17	EUR
F18	UL

+F17	in EUR
-F17	out EUR
+F18	in UL
-F18	out UL

{+F17,-F18}	Arbitrage UL EUR
{+F17}	Premium EUR
{-F17}	Surrender EUR
{+F17,+F18}	Mixed Premium

The client actions are thus transformed into a set of flows on technical products. Here, involved technical products are more than 400 for this specific portfolio. The number of possible flows is then more than 800 (*positive & negative flows*).

The application of an "Apriori" algorithm seems to be essential to reduce computing time.

[('+F17')]	18991
[('+F17', '+F18')]	3946
[('+F17', '+F18'), ('+F17')]	676
[('-F17'), ('+F17')]	1577

…

Apriori

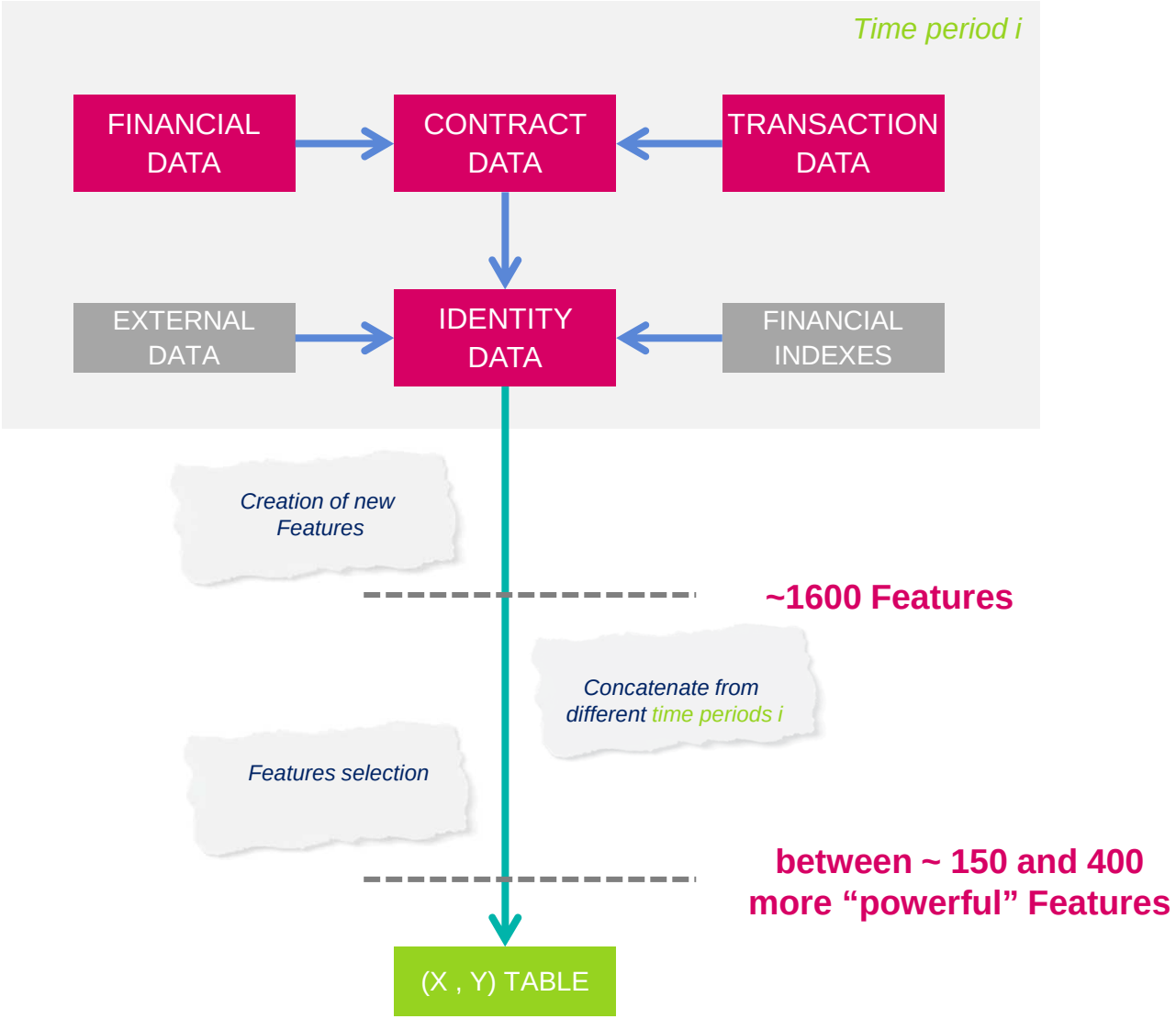
1 - Iteration by length of candidates

2 - Intelligent construction of candidates from shorter frequent subsequences

3 – Pruning constructed candidates

As few candidates as possible to go through

Feature engineering - summary



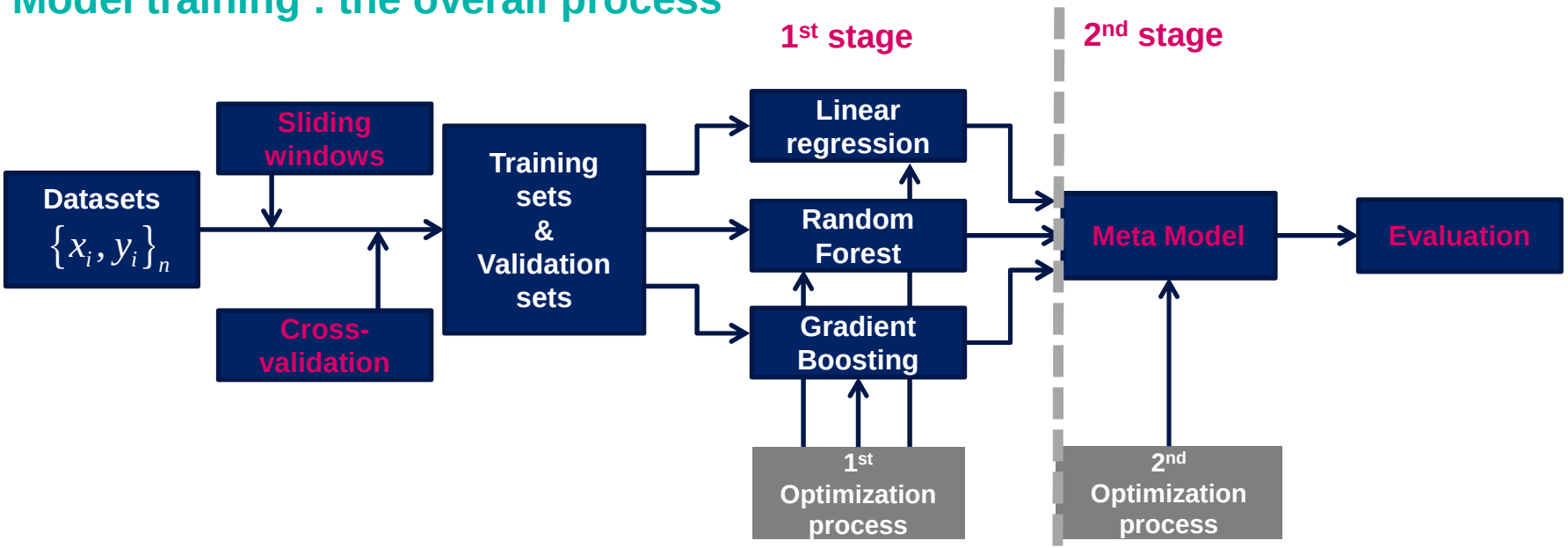
Process of joining tables from different data and preprocessing raw data before running ML algorithms,

Model Training

assurer tous
les avenirs

CNP
assurances

Model training : the overall process

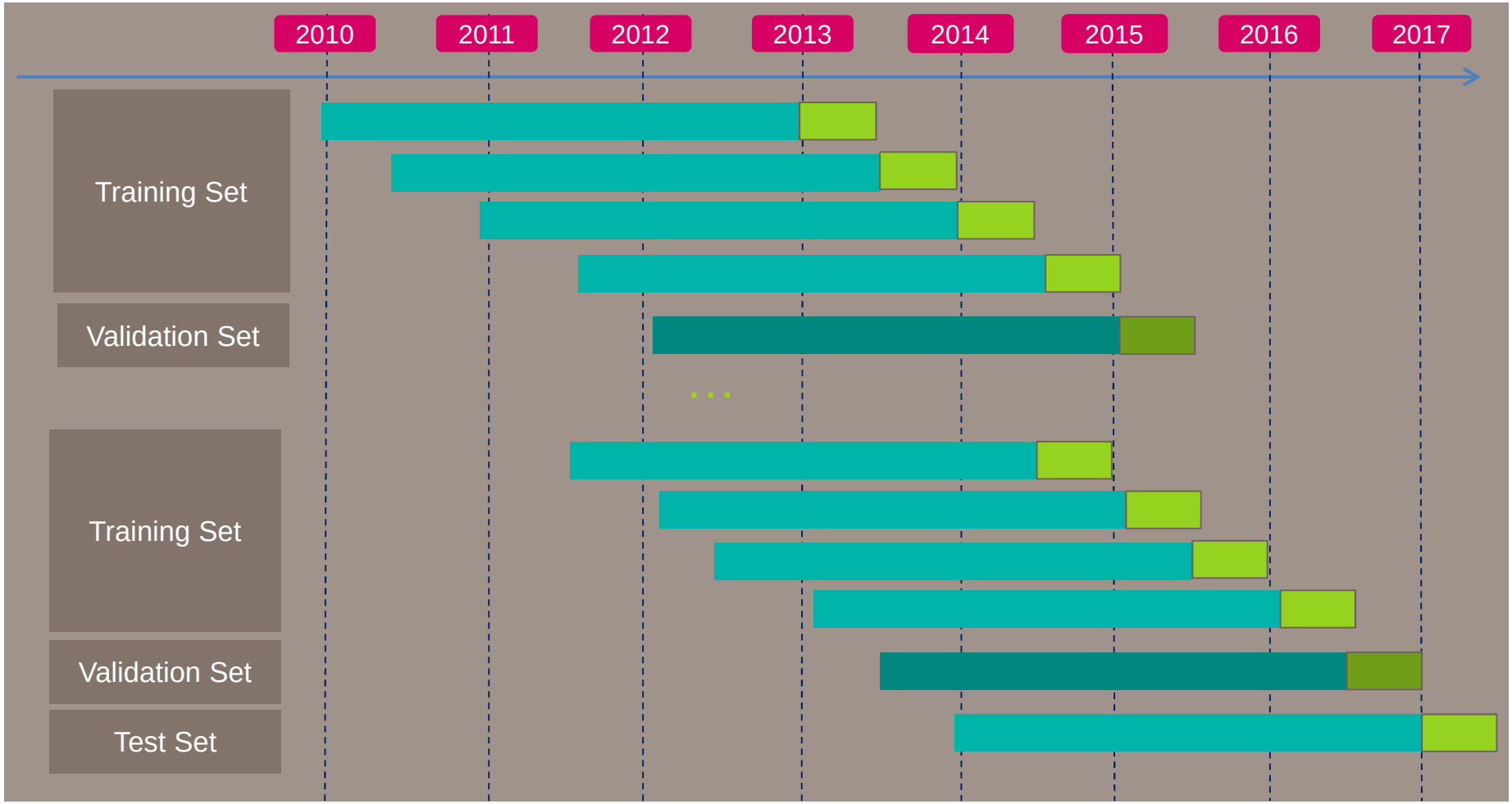


1. The 2nd stage is about using outputs of the first stage prediction as inputs for a 2nd training step for better performances (*see [SD04] in P56*)
2. The two tree-based training methods (*Ensemble methods : Random forest and Gradient boosting*) are generally chosen based on performance testing. The model training could stop at this step. Linear regression is generally kept to benchmark the performance
3. The optimization steps (Ex : *using "Hyperopt" package from Python*) calibrates respectively base-level models (stage 1) and meta model (stage 2)

1. **…** Only about **~20% of the time is dedicated to run models**. Algorithms & processing tools are well designed and ready for use (*Cf @R and python packages for running classification problems*)

Model training – Sliding windows & Cross validation

Temporal K-Fold :



Model training – Evaluation

When the target rate is low, the choice of model evaluation metrics is very important

Accuracy is not a good score:

- ✓ In a low 1% target rate case, a model that predicts only 0's is correct in 99% of cases

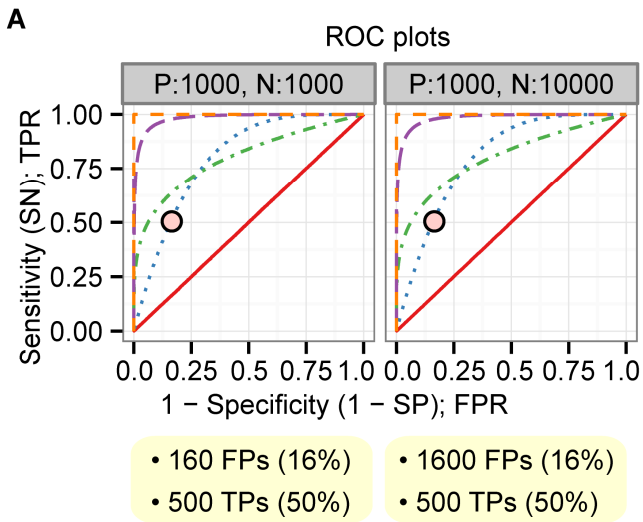
We've selected two metrics that allow us to properly manage the specificity of the low target rate:

Area under the ROC curve (AUC):

- ✓ AUC does not vary according to the target rate, thus it help us to compare the performance of the model over time

LIFT :

- ✓ Lift = $\frac{target\ response}{average\ response}$
- ✓ The Lift 10% is the performance criterion we want to optimize, corresponding to about 5000 targeted customers



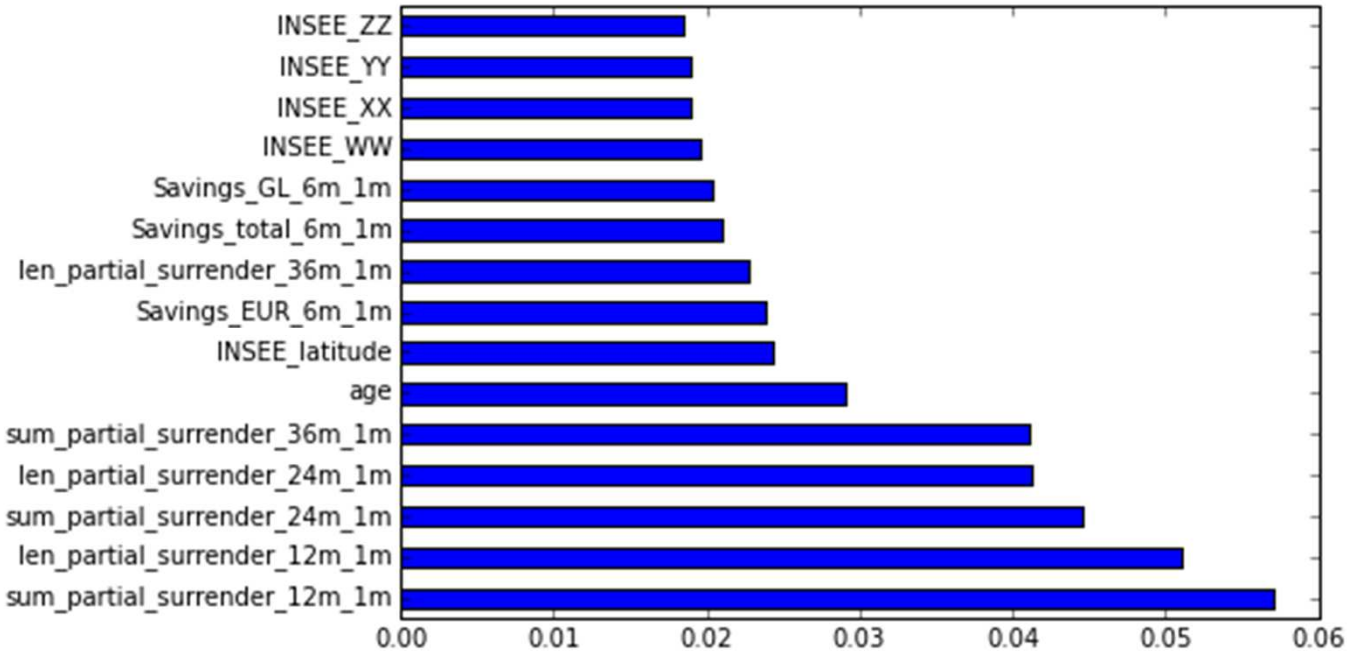
Standard Model Performance

assurer tous
les avenirs

CNP
assurances

Feature importance

Variable Importance rendered by random forest when learning about Surrenders :

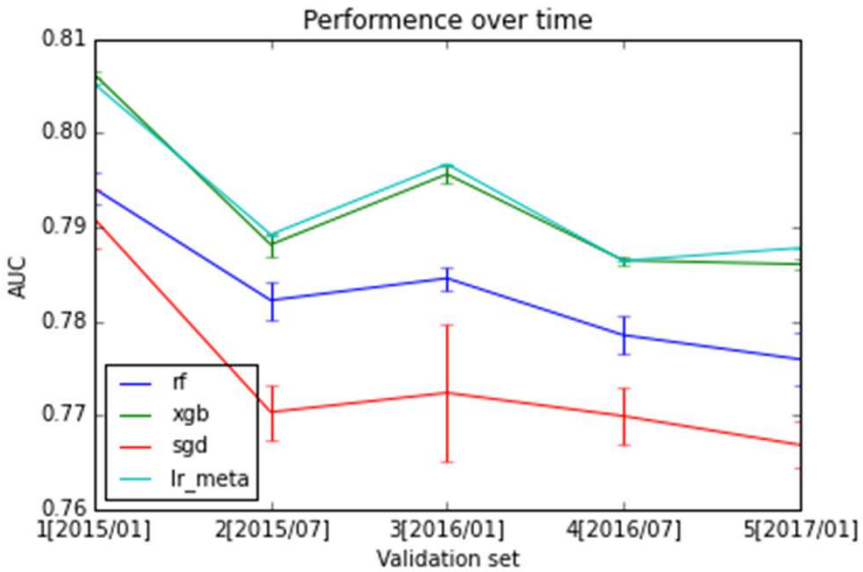


Remark:

1. “Partial surrenders” in the past seems to be the main variable used for the prediction

Performance over time

Representation of the AUC metric over time:

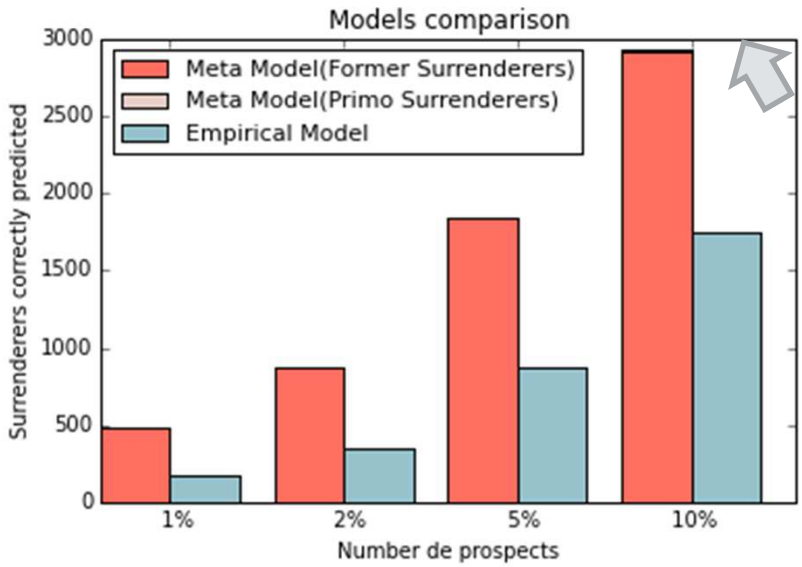


Remark:

1. Meta model (*lr_meta*) and the gradient boosting (*xgb*) seem to outperform other models
2. Stability of predictions (*AUC remains high*) with respect to time (… sorry for "purists" but no assumption on error made is taken here, only "*test and see*" approach holds)

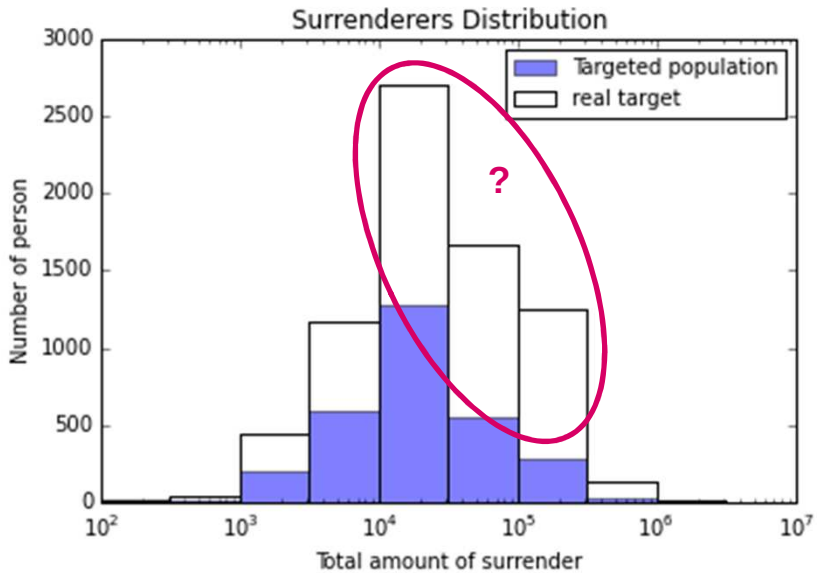
Performance on test set

Comparison with the empirical model (*past surrenders imply future surrenders*) :



Recall : Prediction of **45% of surrenders** by an intelligent choice of **10% of the population**

Surrender distribution of the 10% targeted population:



Remarks:

1. **The model targets** only clients who have previously surrendered (**Former Surrenders**) … Why couldn't we just focus on people who ever make adverse actions in the past to predict the future…?
2. Distribution of **the population selected by the model is slightly biased to the left**, which is not necessarily desired

Sequential rules

assurer tous
les avenirs

CNP
assurances

Recall … Feature engineering - Sequence Mining

Adverse behavior : **What happens on large retailers when mining frequent sequences from transactional databases?**

A collection of receipt for each client i

Mining sequences and find frequent subsequences among all customer sequences

```
Algorithm GSP(D, ρ_s, I)
  C_1 := {⟨{i}⟩ | i ∈ I}
  Scan D to get support of every sequence in C_1
  L_1 := {s | s ∈ C_1, sup(s) >= ρ_s}
  i := 1
  while (L_i ≠ ∅)
    C_{i+1} := GSP_Gen(L_i)
    Scan D to get support of every sequence in C_{i+1}
    L_{i+1} := {s | s ∈ C_{i+1}, sup(s) >= ρ_s}
    i := i + 1
  Return L_1 ∪ L_2 ∪ ... L_{i-1}
```

Apply an action for the corresponding clients

- Review dispositions of products through the different sections of the store
- Push new products to clients
- …

Remarks:

1. The same algorithm applied in our savings portfolio could be run by using analogies (*see P36*)

Sequential rule

Sequential rules search for predictive rules based on previous client transactions :

Subsequence	Occur-rence	Support
A	3	0,75
B	3	0,75
C	3	0,75
D	3	0,75
E	2	0,5
A,B	2	0,5
A,D	3	0,75
B,D	2	0,5
C,E	2	0,5
A,B,D	2	0,5

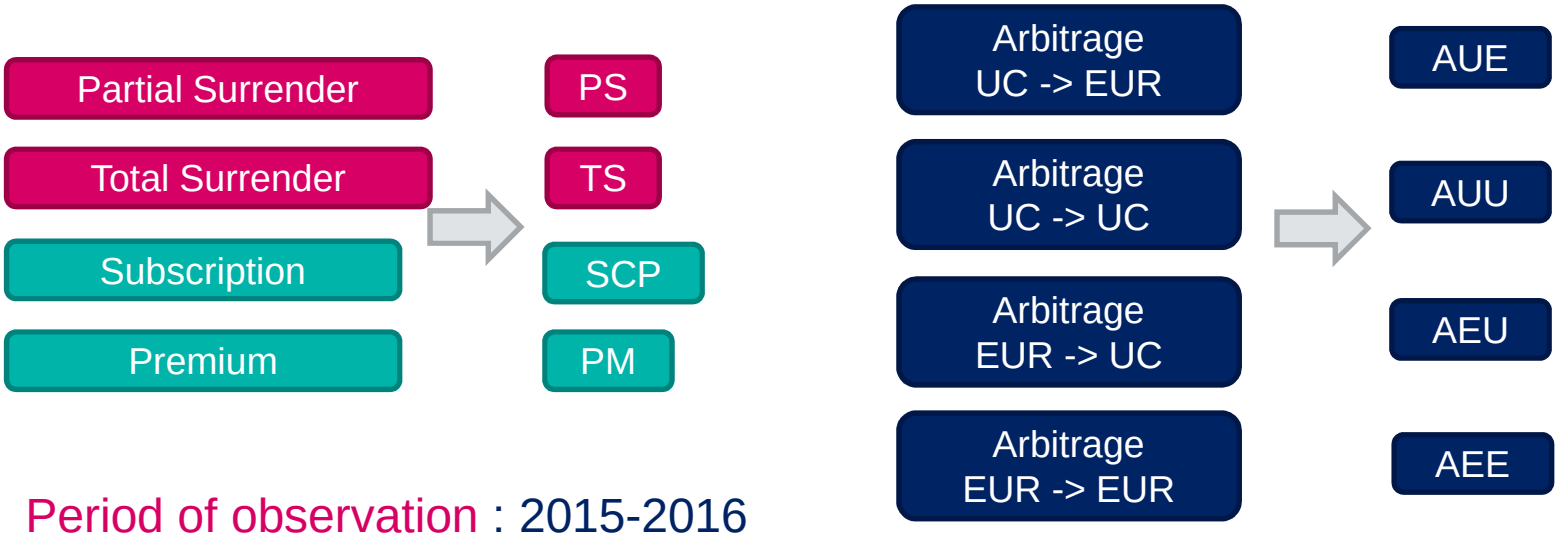
Accuracy

Accuracy / random model

Generated rule	Rule support	Rule base support	Rule target support	Rule confidence	Rule LIFT
A->B	0,5	0,75	0,75	0,67	0,88
A->D	0,75	0,75	0,75	1,0	1,33
B->D	0,5	0,75	0,75	0,66	0,88
C->E	0,5	0,75	0,5	0,66	1,33
A,B->D	0,5	0,5	0,75	1,0	1,33

Sequential rule

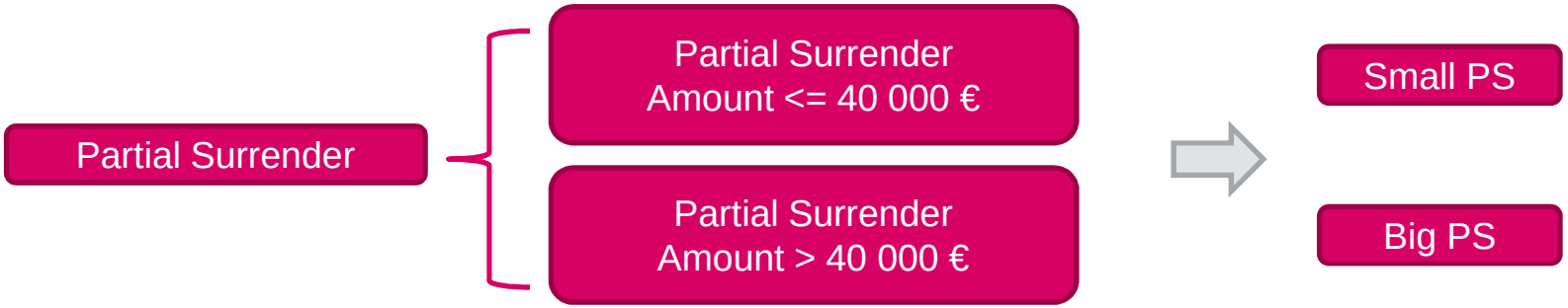
Consider client actions as transactions :



Period of observation : 2015-2016

Extension :

- Count subsequences within one year threshold
- Taxonomy on partial surrender:



Surrender law

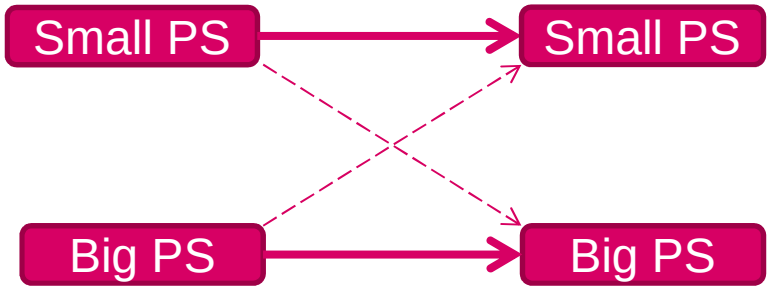
Observation 1: Strong sequencing of partial surrenders

PS → PS → PS → PS → PS

Rule	Support	Confidence	LIFT
PS → PS	14211	0.568	2.85
PS,PS → PS	8907	0.626	3.14
PSx3 → PS	5890	0.661	3.31

Surrender law

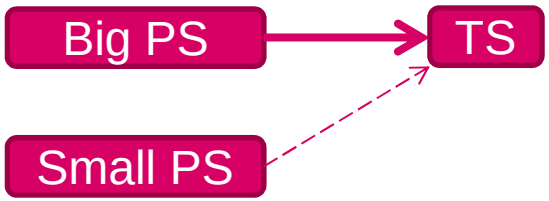
Observation 2 : Separation of the two types of partial surrenders



Rule	Support	Confidence	LIFT
Small PS → Small PS	11172	0.574	3.71
Big PS → Big PS	849	0.152	3.42
PS → PS	**14211**	**0.568**	**2.85**
Big RP → Small PS	1104	0.197	1.27
Small PS → Big RP	1070	0.055	1.23

Surrender law

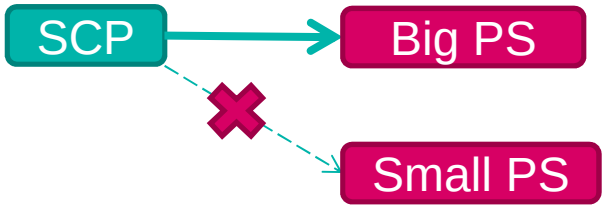
Observation 3 : Partial surrenders deduce total surrenders, big PS indicates more likely a tendency to exit

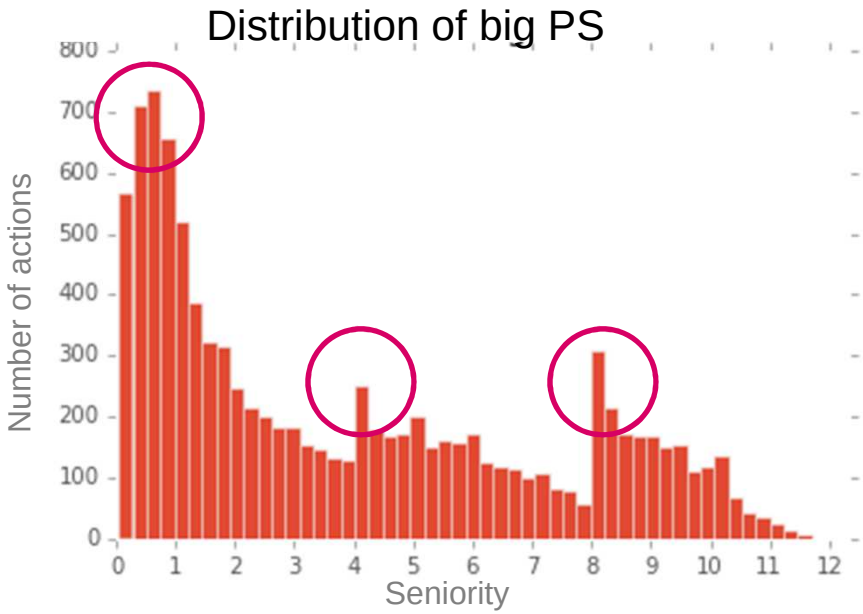


Rule	Support	Confidence	LIFT
Big PS → TS	311	0,038	3,45
Small PS → TS	466	0,023	2,09

Surrender law

Observation 4 : Big surrenders after subscription





Rule	Support	Confidence	LIFT
SCP → Big RP	659	0.095	2.14
SCP → Small RP	549	0.079	0.51

Surrender law

Conclusion :

- Surrender snowball effect
- Different behavior for small surrender and big surrenders
  - Less interaction between two types of surrenders
  - Big surrenders after subscription
  - Big surrenders imply more a tendency to churn

Idea :

- **Until surrender snowball effect is noticed for “big surrenders”, the idea is to calibrate the model to target “big surrenders” rather than focusing on “small surrenders” easier to predict but with less impact on profitability**

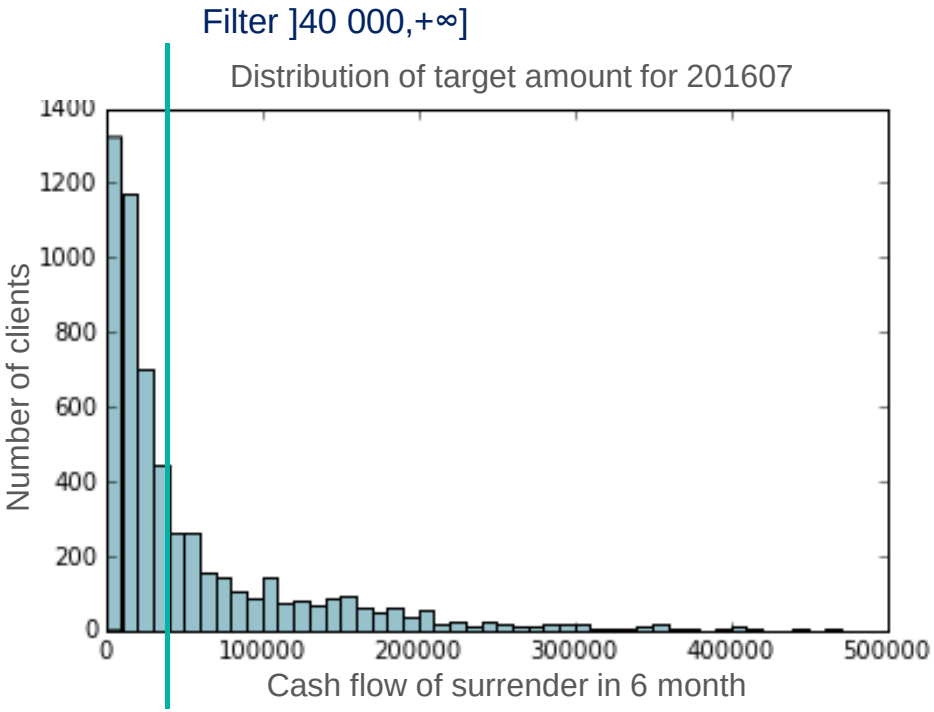
Truncated Model Performance

assurer tous
les avenirs

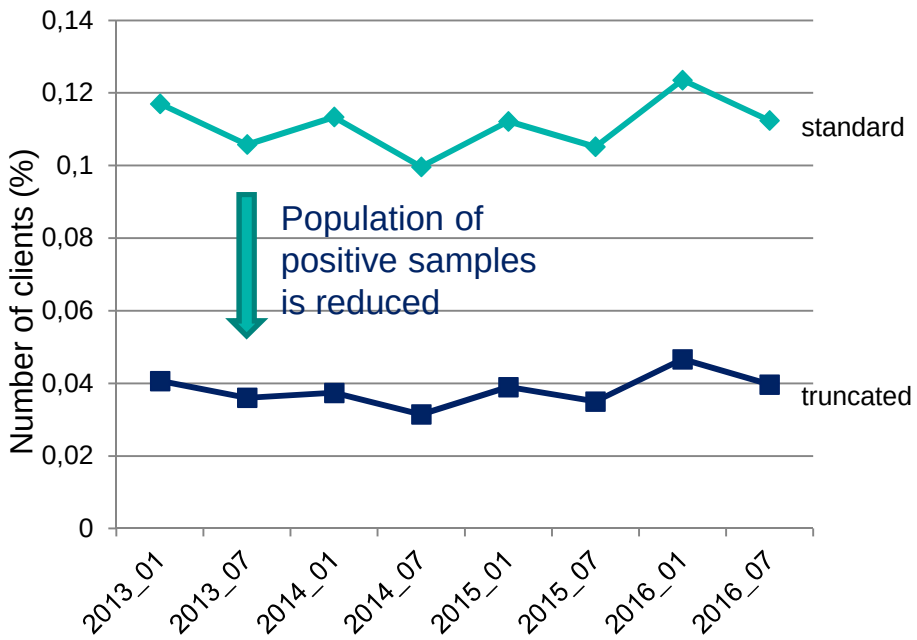
CNP
assurances

Truncation of target

Truncate the target:



Effect on the target:

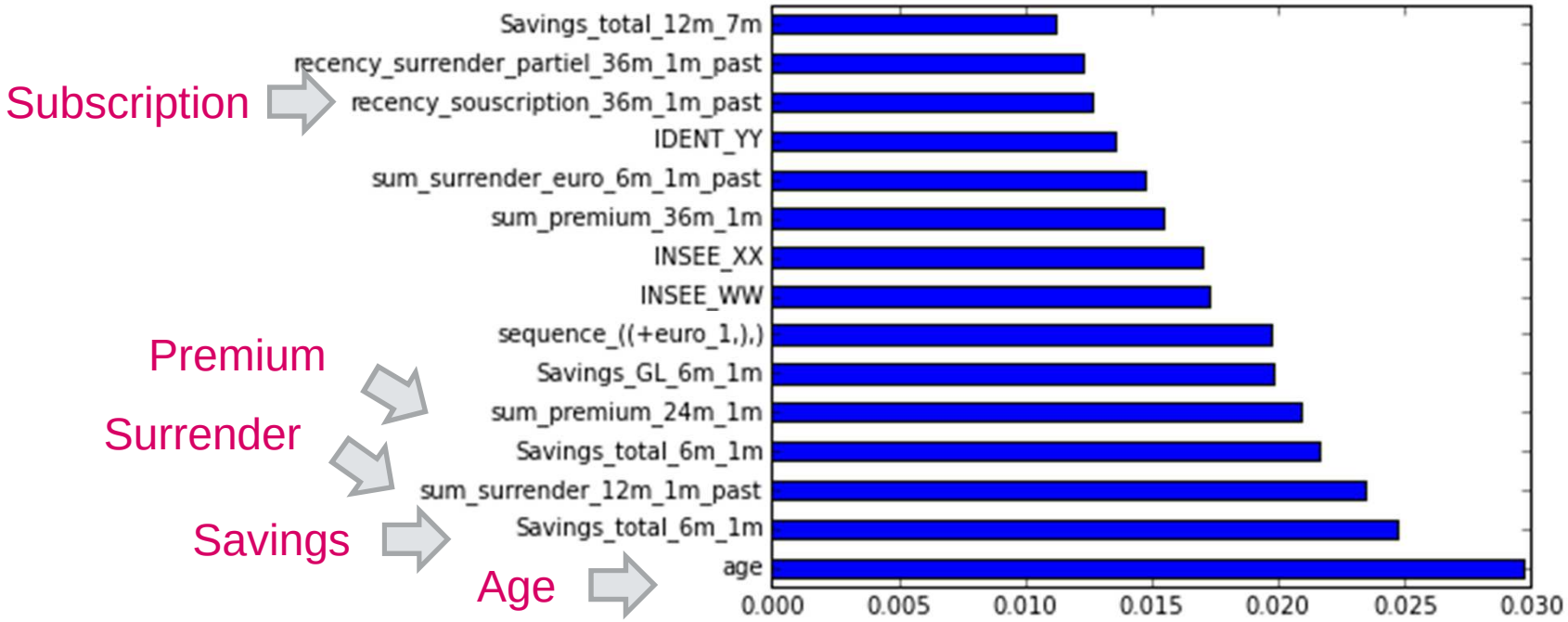


Remarks:

1. Many small partial surrenders occurs in our insurance product
2. The target is reduced to 66% after filtering on “big surrenders”

Feature importance

Variable Importance rendered by random forest after filtration:

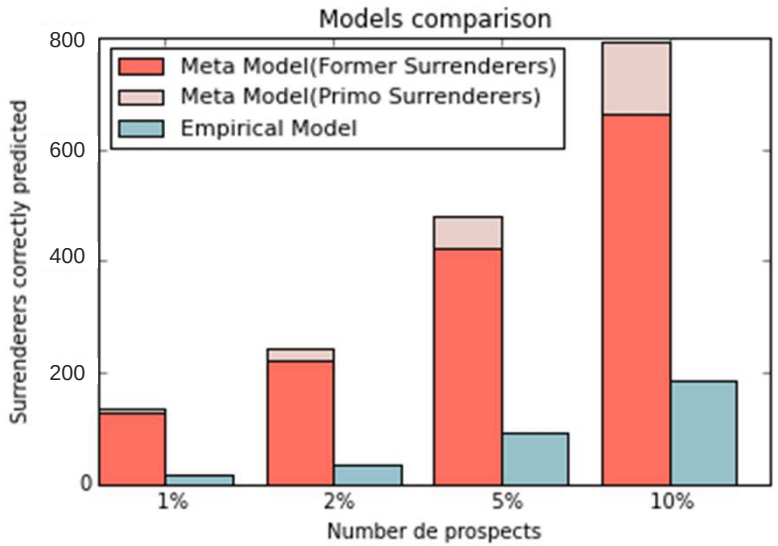


Remark:

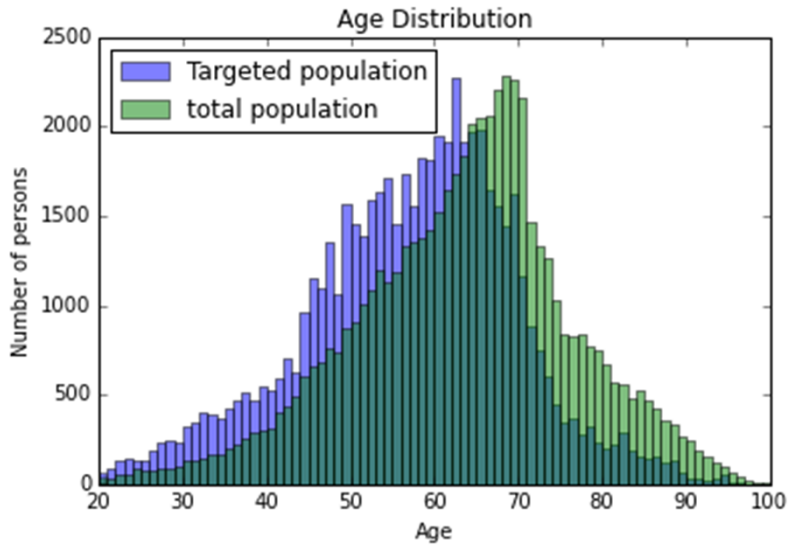
1. A prediction based on much information : Age, Savings, Surrender, Premium, Subscription

Performance on test set

Comparison with the empirical model (past surrenders deduct future surrenders):



Age Distribution of the 10% targeted model :



Remarks:

1. **Model targets** not only "Former Surrenders" but also **"Primo Surrenders" who could never been targeted by traditional descriptive methods**.
2. Model targets younger clients (*who likely seem to be more active in the management of their policies*)

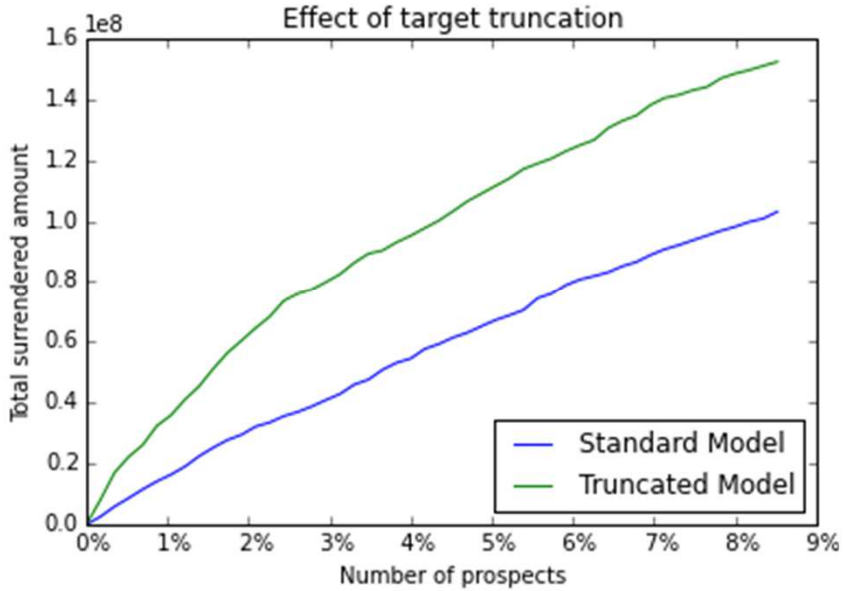
Model Comparison

assurer tous
les avenirs

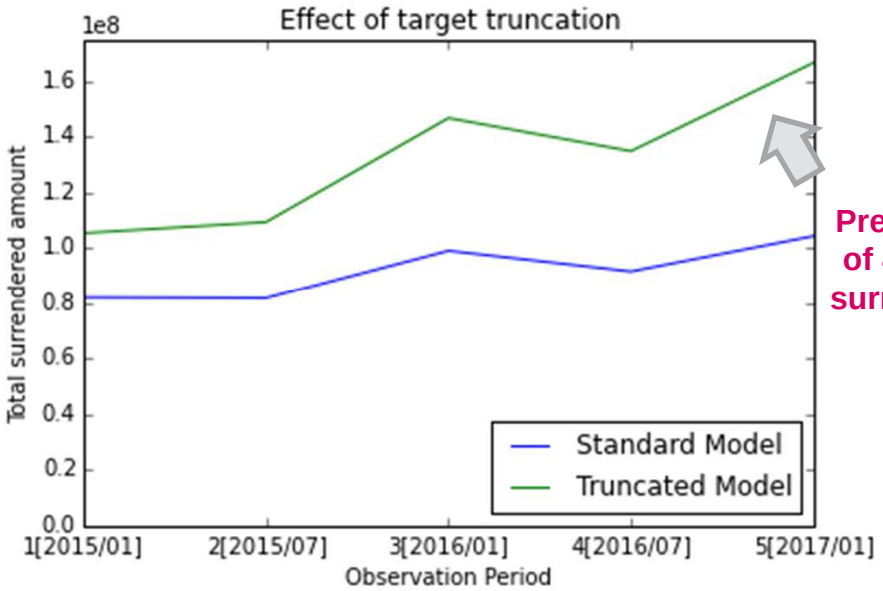
CNP
assurances

Comparison of predicted cash flow

Comparison on 2017 S1 :



Comparison with 10% de prospects over time:



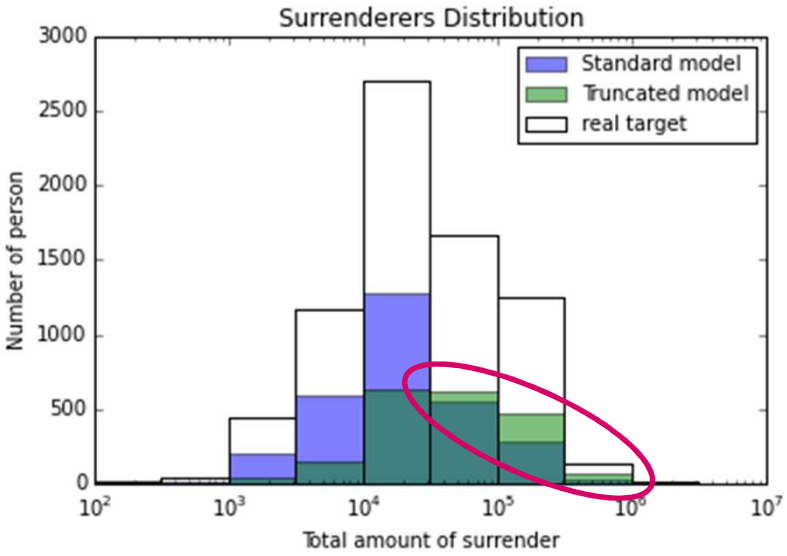
Prediction of 40% of surrenders

Remark:

1. Truncated model constantly outperforms the standard model by the view of cash flow
2. Under the hypothesis of an uniform transformation rate, our truncated model can outperform the standard model by 64% on test set.
3. **40 % of the surrenders occurred into the period** under review **are captured with** an “intelligent selection” of **10% of customers**.

Comparison of predicted cash flow

Surrender distribution of the 10% targeted population:



Remark:

1. The model's prediction is adjusted to the right by focusing on clients who make bigger surrenders.

Conclusion

assurer tous
les avenirs

CNP
assurances

Conclusion

Several interesting result with a value added on the understanding of behavioral patterns in savings :

1. A predictive model for better targeting clients' behaviors (*with risk management efficiency as a consequence*)
2. A strong stability in time validated by an experimental way
3. Exploration of different way to boost performances by a reasoning by analogy with techniques widely used in other industries (bioinformatics and retailing industries for sequence mining)

Additional research directions are identified :

1. Explore more techniques to optimize the cash flow target
2. Remove multicollearity from features to enable tests with simpler models
3. Search for more interesting data (guarantee conditions, former marketing communication and results, …)

-> The future of the actuarial job could not ignore the use of such techniques for both commercial and risk optimization purposes.

Appendix A : Method and approach

assurer tous
les avenirs

CNP
assurances

Method and Approach : Data selection and semantic

Availability	Usability	Reliability	Relevance	Presentation quality
Accessibility	Definition / Documentation	Accuracy	Fitness	Readability
Timeliness	Credibility	Integrity		Structure
Authorization	Metadata	Consistency		
		Completeness		
		Auditability		

Universal two-layer data selection standard,

Yangyong CAI, Li, ZHU, The challenges of data quality and data quality assessment in the big data era. Data Science Journal, 14, 2015.

Appendix B: Vocabulary & References

assurer tous
les avenirs

CNP
assurances

Vocabulary

- CNP Assurances is the leader of Life Insurance in France
- Arbitrage is a switch from a unit linked (UL) funds to the €uro fund (then we call it adverse arbitrage) or the opposite transfer from €uro to UL funds.
- Surrender is the action of withdrawing savings from the contract
- Predict adverse actions (here surrenders & adverse arbitrages) stands for targeting customers who will make an adverse action in the near future, with respect to a given metric.

References

References

[ACP15] ACPR. Analyse de l'exercice 2015 de préparation à solvabilité ii. (56):25, 2015.

[AGG02] Charu C. AGGARWAL. On effective classification of strings with wavelets. *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining.*, pages 163–172, 2002.

[Bot04] Leon Bottou. Stochastic learning. *Springer*, pages 146–168, 2004.

[CF15] Osmar Castrillo-Fernandez. Web scraping: Applications and tools. *European Public Sector Information Platform*, 2015, 2015.

[DC03] HOLTE Robert C. et al. DRUMMOND Chris. C4.5, class imbalance, and cost sensitivity: Why under-sampling beats over-sampling. *Workshop on learning from imbalanced datasets*, II, 2003.

[FRI01a] HASTIE Trevor et TIBSHIRANI Robert FRIEDMAN, Jerome. The elements of statistical learning. *Springer, Berlin : Springer series in statistics*, 2001.

[FRI01b] Jerome H. FRIEDMAN. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[FV12] FAGHIHI Usef NKAMBOU Roger et al FOURNIER-VIGER, Philippe. Cmrules: Mining sequential rules common to several sequences. *Knowledge-Based Systems*, 25(1):63–76, 2012.

[GYE01] Attila. GYENESEI. A fuzzy approach for mining quantitative association rules. *Acta Cybern*, 15(2):305–320, 2001.

[KIM04] Choong-Han KIM, Hee-Su et YOON. Determinants of subscriber churn and customer loyalty in the korean mobile telephony market. *Telecommunications policy*, 28(9):751–765, 2004.

[KUD98] Haym KUDENKO, Daniel et HIRSH. Feature generation for sequence categorization. *AAAI/IAAI*, pages 733–738, 1998.

[LIU05] Ya-Yueh LIU, Duen-Ren et SHIH. Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences. *Journal of Systems and Software*, 77(2):181–191, 2005.

[RA95] Ramakrishnan Srikant Rakesh Agrawal. Mining sequential patterns. *IEEE Computer Society Washington*, pages 3–14, 1995.

[SD04] BERNARD ZENKO SASO DZEROSKI. Is combining classifiers with stacking better than selecting the best one. *Springer, Kluwer Academic Publishers.*, Machine Learning(54):255–273, 2004.

[sld16] scikit-learn developers. Supervised learning. *http://http://scikit-learn.org*, Online Source, 2016.

[SRI96] Rakesh SRIKANT, Ramakrishnan et AGRAWAL. Mining quantitative association rules in large relational tables. *Acm Sigmod Record*, pages 1–12, 1996.

[TH08] Jerome Friedman Trevor Hastie, Robert Tibshirani. The elements of statistical learning. *Springer*, 2008.

[TS15] Marc Rehmsmeier Takaya Saito. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0118432#pone-0118432-g005*, Online Source, 2015.

[YAN04] LI Tianyi et WANG-Ke YANG, Qiang. Building association-rule based sequential classifiers for web-document prediction. *Data mining and knowledge discovery*, 8(3):253–273, 2004.