

**100% ACTUAIRES &
100% DATA SCIENCE**

INSTITUT DES
ACTUAIRES



16 Novembre 2018
Hôtel Marriott Rive Gauche
Paris 14ème

Pilotage supervisé d'un portefeuille d'assurance



- **CARDIF IARD**

- Join-venture née en 2016 d'un partenariat entre BNP Paribas (2/3) et MATMUT (1/3)
- Internalisation de l'activité d'assurance réalisée jusqu'alors par Axa Global Direct
- Objectif d'accroître l'activité d'assurance IARD en démarrant par les produits de masse Automobile et Habitation en s'appuyant sur le réseau d'agences BNP PARIBAS et le savoir-faire MATMUT

- Monitoring des risques en période de lancement
 - Modifications importantes de l'offre : hypothèses fortes sur les changements de périmètres de garantie
 - Absence totale de sinistralité propre à cette offre à court terme
 - Tarif basé sur la sinistralité du portefeuille existant dont la structure n'est pas forcément le reflet de la production future
 - Création d'un suivi basé sur la structure de la production au regard de la sinistralité attendue

- Approche tarifaire : 3 modèles

- Un modèle GLM technique

- A servi à calibrer la prime pure selon les contraintes I.T. et juridique
- Facilite le suivi des prises de décisions tarifaires futures

- Un modèle GLM commercial

- Basé sur le GLM technique, il tient compte des desideratas commerciaux pour atteindre les cibles privilégiées

- Un modèle de XgBoost

- Sans contrainte, il est considéré comme la vision du risque la plus fine
- Intègre l'ensemble des critères significatifs, met en exergue des signaux faibles non détectés par les approches GLM

- **Modélisation XgBoost (1/2)**

- Une approche fréquence / coût moyen

- Mettre en exergue les variables les plus significatives de chacun des éléments constitutifs de la prime pure

- Une focalisation sur les sinistres étêtés

- Éviter de parasiter le modèle de coût moyen
- Gagner en robustesse prédictive

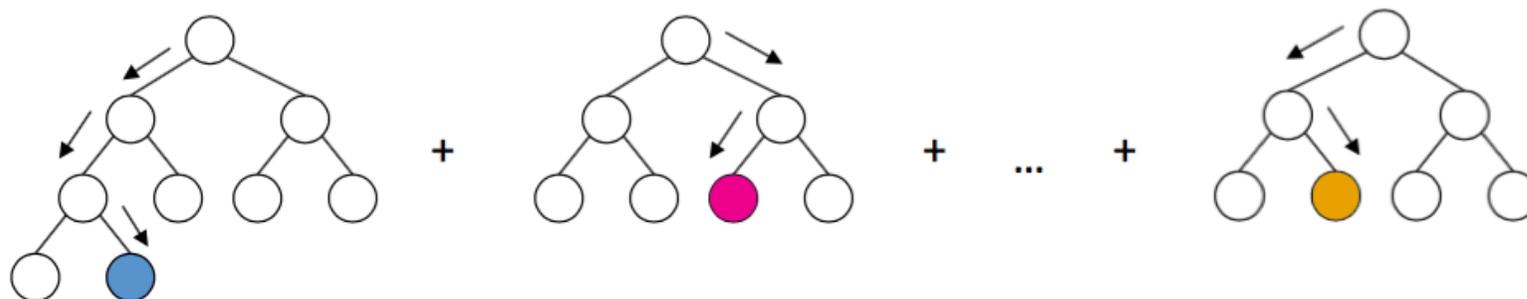
- Training, Validation, Test

- Optimiser les paramètres par Cross-validation
- Mesurer la qualité du modèle sur l'échantillon de test

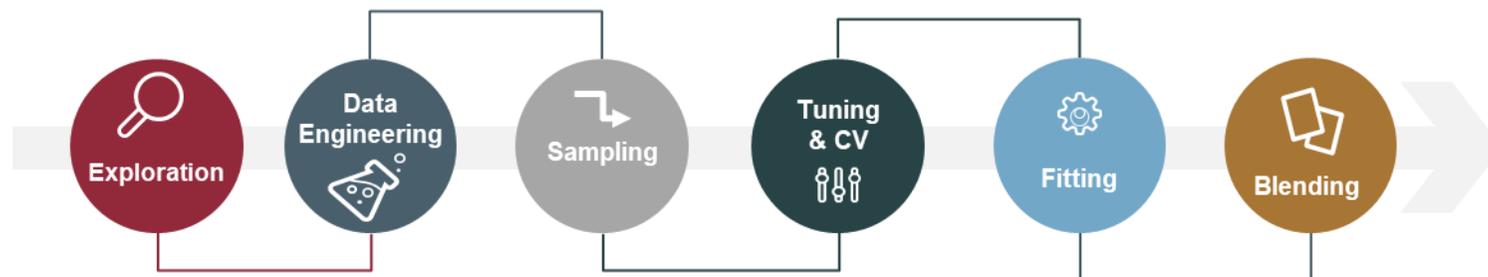
• Modélisation XgBoost (2/2)

– Corrections inhérentes au lancement d'un produit

- Vision ultime de la charge, inflation,
- modifications offres,
- prise en compte de l'absence d'ancienneté du ptf, ...



• La démarche Data Science



Missing data

- deletion
- correction
- imputation
- indicator

Data transformation

- asymmetry reduction
- discretisation
- scaling
- cutting
- one-hot encoding

Extrem value

- correction
- deletion
- impact study

Features engineering

- historic
- moving average
- business variables

Regression

- Random Forest
- XGBoost
- ExtraTree
- K Neighbors Reg
- SGD Regressor
- Elastic Net

Anomaly detection

- One class SVM
- Rlof
- nn2
- RANN

Classification

- Logistic regression
- Random Forest
- ExtraTree
- Ada Boost
- XGB Boost

Clustering

- kmeans
- hclust



- Deux suivis distincts

- Sinistres attritionnels

- D'un côté la prime pure modélisée sur les sinistres étêtés
 - De l'autre, la prime commerciale hors chargements, et hors coefficient relatif aux sinistres graves
 - Par différence, nous obtenons une « **marge unitaire** » sur laquelle nous allons concentrer notre étude

- Sinistres de pointe

- Une modélisation via un modèle de propension ($P(Y > \theta \mid Y)$) pour déterminer les profils les plus générateurs de graves
 - Un suivi plus macro à ce stade, qui sera affiné par la suite

• Les 1^{ers} enseignements de la modélisation

- 9 des 10 variables les plus influentes sont communes aux modèles de fréquence et coût moyen
- 8 des 10 variables les plus influentes ont été activées dans le tarif commercial (bien que la segmentation ait été parfois atténuée)
- 1 des variables de chaque top 3 a fait l'objet d'une renonciation d'activation pour des raisons commerciales → une attention plus particulière devra être portée sur cet élément
- Beaucoup de signaux faibles détectés dans les variables entre les 11 et 20^{ème} rangs, davantage axés sur l'aspect comportemental du client que sur le risque assuré en lui-même

→ Refus IT / commerciaux ou restrictions juridiques

• Problématiques rencontrées

– Un besoin d'automatisation

- Besoin d'automatiser les calculs de marges unitaires en fonction de la structure de production actualisée régulièrement
- Besoin de gérer les codes couleurs affectés aux marges recalculées dans chaque feuille
- Besoin d'automatiser la création d'arbres et la production de reportings

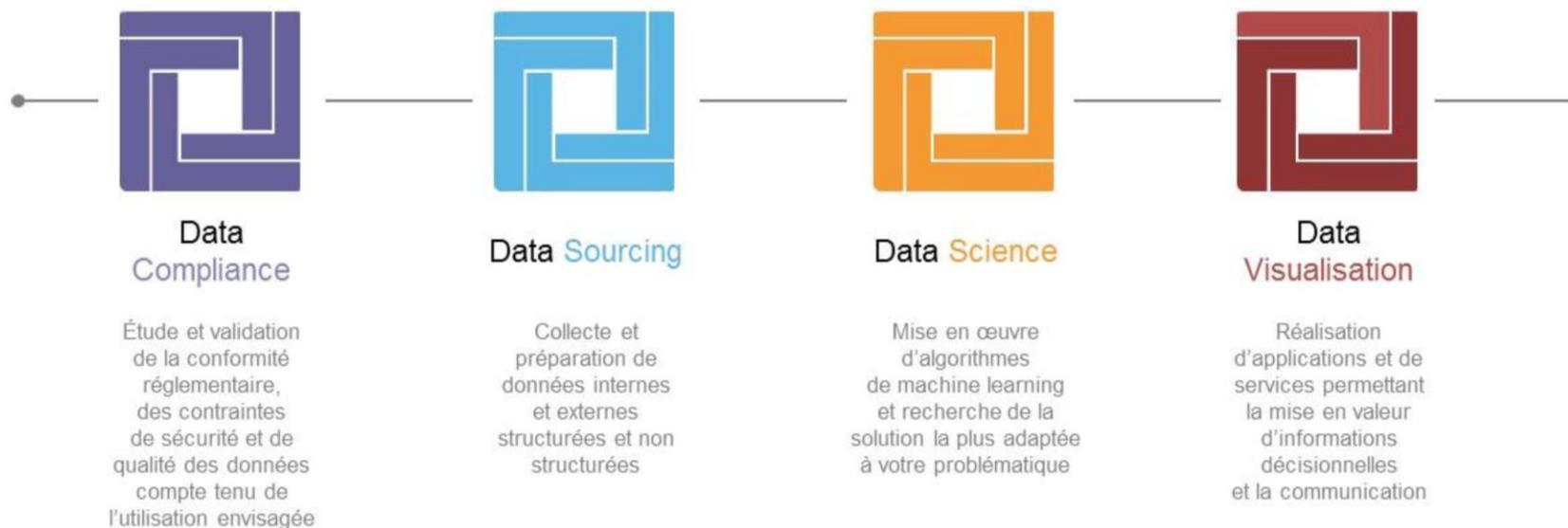
– Un besoin de visualisation **à des fins de pilotage**

- Pouvoir descendre dans les branches inférieures de l'arbre
- Pouvoir intégrer d'autres indicateurs d'aide à la décision : taux de transformation, taux de résiliation, tarifs des concurrents, ...
- Disposer d'un outil de visualisation plus flexible, plus rapide et plus clair

Proposition d'outil de visualisation pour le pilotage de portefeuille



Le Data Risk Management



More efficient. **Together**

Algorithme choisi

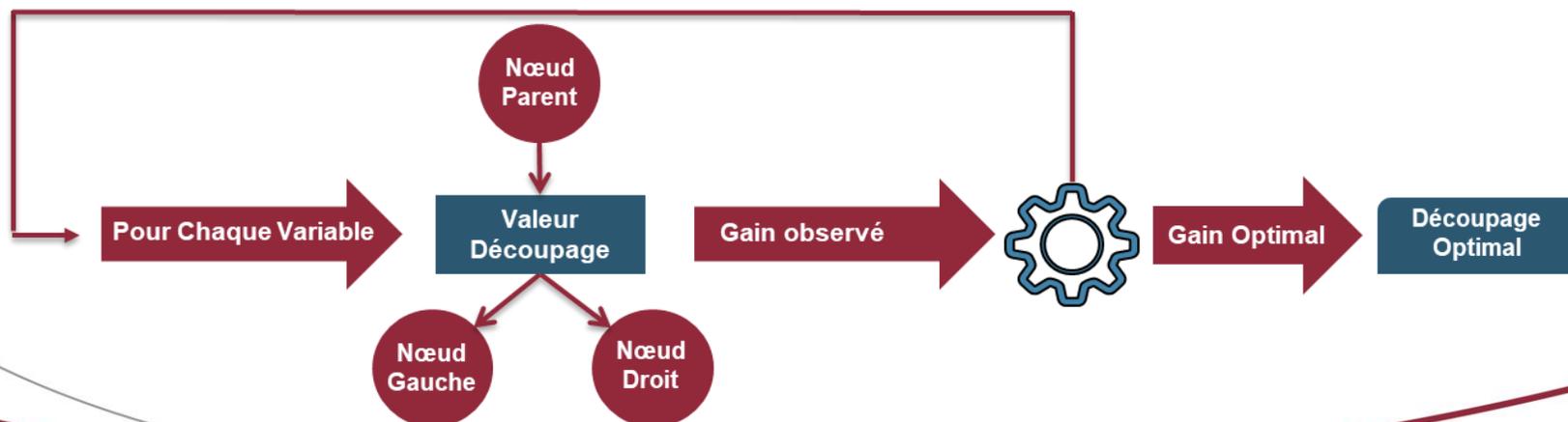
Construire un arbre de décision

- Ne nécessite pas de distribution particulière des données
- Performant sur de grands jeux de données et économique en ressources de calcul
- Procédure d'élagage embarquée pour éviter le surapprentissage

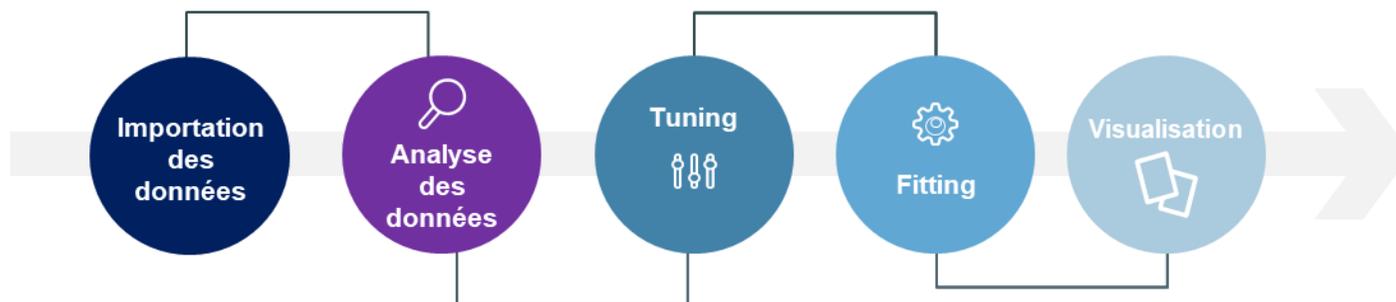
Choix d'une fonction objective selon le use case étudié

L'apprentissage se fait de manière récursive

A chaque nœud, on recherche la variable et sa modalité optimisant la fonction objective



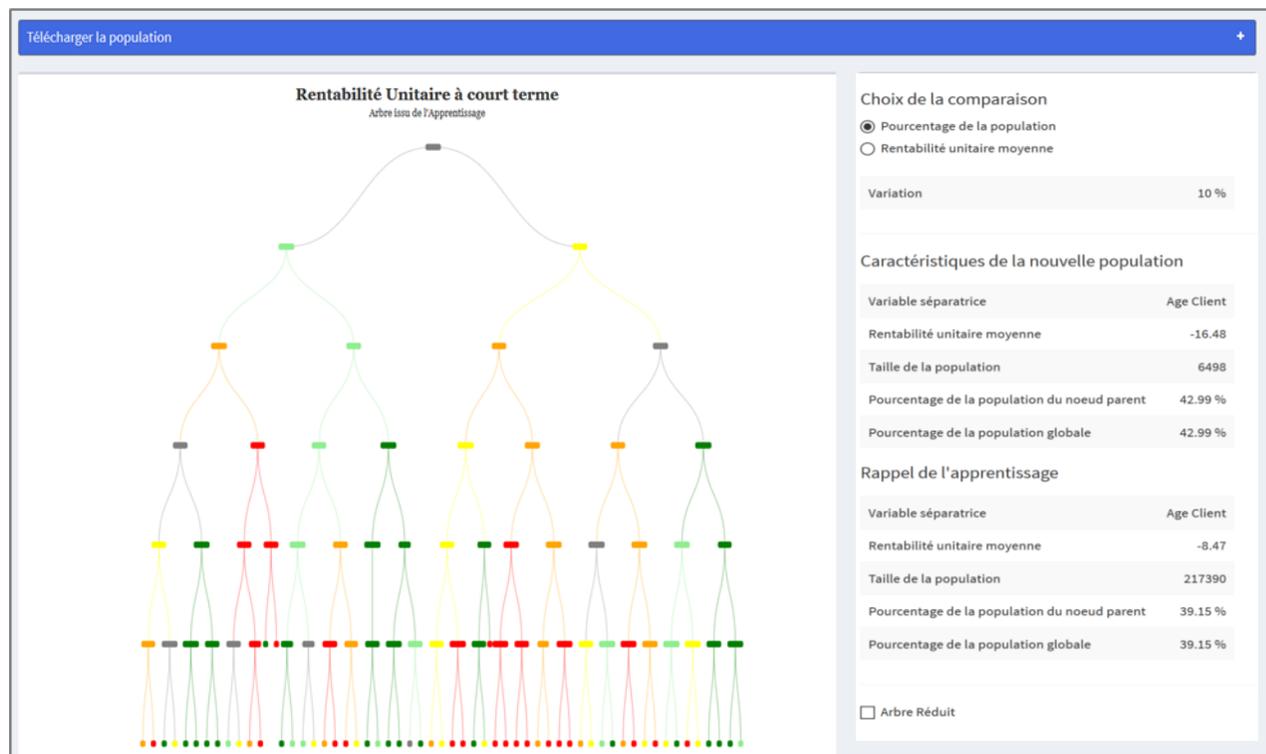
Processus de l'outil



Exploration	Paramétrage	Apprentissage	Application
<ul style="list-style-type: none"> • Importation base app. • Statistiques 	<ul style="list-style-type: none"> • Choix du use case / fonction objective • Sélection des variables • Réglage des de l'arbre de décision 	<ul style="list-style-type: none"> • Visualisation de l'arbre • Statistiques et pour chaque nœud 	<ul style="list-style-type: none"> • Importation base test • Comparaison base / base test selon une gamme d'indicateurs suivant le use case



Scope des usages de l'outil



USE CASES

- Étude de rentabilité / ROI
- Revue d'une tarification
- Maîtrise des risques
- Analyse d'un taux de transfo
- Allocation d'actifs
- ...

AGENTS CONCERNES

- Les actuaires
- Le marketing
- La souscription
- Les gestionnaires d'actifs
- ...

Quelle que soit la branche d'assurance

Éclairer les risques, tracer l'avenir

Exemple de fonctions objectives

Analyse de la fréquence

- Maximisation de l'indice de Gini
- Objectif -> définir des groupes les plus homogènes possibles de par leur fréquence

$$Gini(Y) = f_{NP} \cdot (1 - f_{NP}) - \frac{N1}{NP} \cdot f_{N1} \cdot (1 - f_{N1}) - \frac{N2}{NP} \cdot f_{N2} \cdot (1 - f_{N2})$$

Régression de la marge unitaire

- Minimisation de la somme des variances intra-groupe
- Objectif -> définir des sous groupes les plus homogènes possibles

$$RMSE(Y, Ru) = \sqrt{\frac{\sum_{i=1}^N (y_i - ru_i)^2}{N}}$$

Approche Loss Ratio

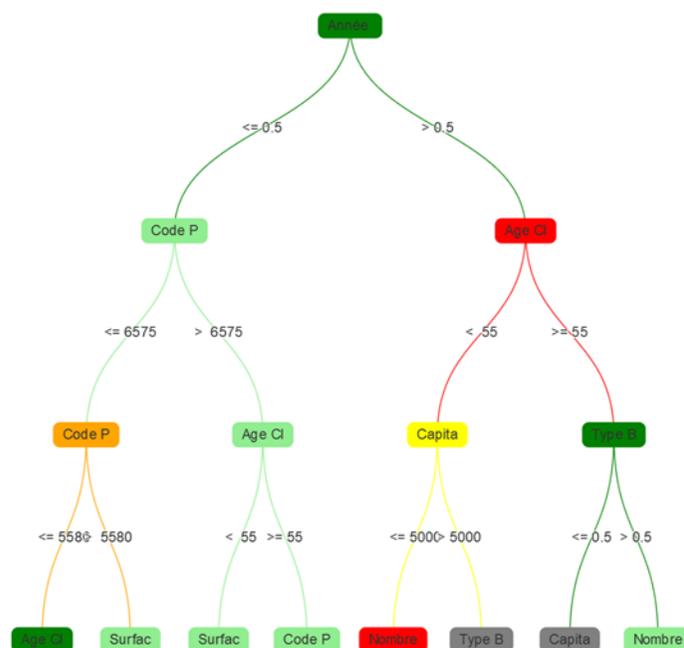
- Maximisation de la distance des Loss Ratio entre les nœuds frères
- Objectif -> définir les poches « extrêmes »

$$LossRatio(S, P) = \left| \frac{\sum_{i=1}^{F1} S_i}{\sum_{i=1}^{F1} p_i} - \frac{\sum_{i=1}^{F2} S_i}{\sum_{i=1}^{F2} p_i} \right|$$

Monitoring des risques

Rentabilité Unitaire à court terme

Arbre issu de la nouvelle population



Apport de l'outil pour Cardif IARD

- Facilité d'emploi et visualisation interactive de la rentabilité unitaire moyenne
- Production simple de statistiques descriptives pour chaque nœud de l'arbre
- Représentation visuelle efficace des variations, entre base d'apprentissage et le new business -> selon une variété d'indicateurs
- Automatisation de reporting mensuel
- Mieux communiqué avec les commerciaux et avec les décideurs

- Avantages de l'application

IT



Aucune installation
sur les postes users



Tout navigateur web



Connexion sécurisée

Métiers



Interface « clic-bouton »



Dynamique
et paramétrable



Aspect collaboratif

Place à la démonstration

optimind.
manage risk, build your future