INSTITUT DES
ACTUAIRES

www.institutdesactuaires.com

Éclairer les risques, tracer l'avenir

100% DATA SCIENCE

125 ANS INSTITUT DES ACTUAIRES

Retour d'expérience sur la mise en place d'un laboratoire de Data Science à la CNP Assurances

*Atelier animé par **Anani OLYMPIO***
Actuaire certifié Institut des Actuaires, Expert ERM CERA
Responsable R&D et Data'Lab - CNP Assurances

AGENDA

1	Contexte et enjeux pour CNP Assurances
2	Structuration de la démarche au sein du Groupe
3	Les Facteurs clés de succès
4	Retour d'expérience : optimisation d'une campagne commerciale
5	Conclusion

Quelques faits marquants et chiffres clés…

VOLUMETRIE – VARIETE – VELOCITE – VERACITE ➔ TRANSFORMATION ➔ VALEUR

- **2,5 Quintillions** de données créés chaque jour
  ➔ 10 millions de disques Blu-Ray (~ 4 tours Eiffel)

- **90%** des données existantes à ce jour
  ➔ ont été produites ces 2 dernières années…
  ➔ sont non structurées…

- **~29K Go** de données générées par seconde en 2013

- **3 milliards** d'internautes sur la planète (pour un trafic global de15 billions Go en 2013)

Quelques faits marquants et chiffres clés…

IMPACTS SOCIETAUX ET ECONOMIQUES ?

❑ Santé, bien-être et soins médicaux

➔ capteurs, gadgets, applis de collecte des données de santé…

➔ Initiative de **Google** « **Google Flu Trends** » (arrêtée en 2015) : outils de suivi et de prévision des épidémies de grippes (capacité à détecter la prévalence de certaines maladie ou épidémie souvent plus vite que les sources officielles…)

❑ Prévision et compréhension du crime

➔ **Police de Los Angeles** a rassemblé les données relatives à plus de **130 millions de crimes** ces 80 dernières années et continue de mettre à jour le logiciel en ajoutant les nouveaux crimes…

❑ Développement économique

➔ Initiative des **Nations Unies** « **Global Pulse** » : projet visant à tirer parti du Big Data à des fins de développement mondial

❑ Autres secteurs en pleine mutation…

➔ Shopping, Industrie automobile, Comparateurs de prix en ligne, Gestion des déchets urbains, Sapeurs Pompiers, Transports urbains, Industrie du tourisme, Systèmes de Domotique ,…

100% DATA SCIENCE

Une question de stratégie : *Vision, missions et valeurs*

125 ANS INSTITUT DES ACTUAIRES

VISION

- Compagnie **Digitale** / avec une démarche **Big Data & data science** orientée ***Business Driven***

MISSION

- Initiatives « **Digital** » et « **Data'Lab / Data science** » intégrées aux processus de **décision/production** afin d'accompagner le Groupe dans son **développement durable** et **rentable**.

  ➔ *CONNAISSANCE CLIENTS / OPERATIONS MARKETING*

  ➔ *PRODUITS / RENTABILITE DES PORTEFEUILLES*

  ➔ *PROCESSUS / GESTION DES CONTRATS ET DES SINISTRES*

VALEURS

- Respect des **règles d'éthique** concernant l'utilisation des données personnelles

- **Client au cœur, Inventivité, Initiative, Confiance**

AGENDA

1	Contexte et enjeux pour CNP Assurances
2	Structuration de la démarche au sein du Groupe
3	Les Facteurs clés de succès
4	Retour d'expérience : optimisation d'une campagne commerciale
5	Conclusion

Organisation, comitologie et gouvernance Data'Lab

ROLE DE LA DATA'LAB

- **VECTEUR DE DIFFUSION DE LA DEMARCHE BIG DATA DATA SCIENCE** AU SEIN DU GROUPE
- UNE APPROCHE **EXPERIMENTALE** ORIENTEE **OPTIMISATION** ET **RESULTATS**
- **CONCEVOIR DES PROOF OF CONCEPT (POC)**

  PoC = Use Case
  + Délais raisonnable de réalisation (3 mois)
  + Contributeurs (Data'Lab, métiers, Directions…)
  + Engagement et appuis (au plus haut niveau)
  + Processus itératif (méthode AGILE…)

- **Communication** et mise en **production** rapide des succès

Organisation, comitologie et gouvernance : Mise en œuvre des Use-Cases

PROCESSUS EN 3 ETAPES

- **Etape 1 (pré-étude) : Sélection des études à mener**

  ➔ Définir les use cases + KPI + Evaluation

- **Etape 2 : Réalisation des PoC**

  ➔ Data (DataLake) + Modélisation + Echanges réguliers et restitution (sur les résultats et hypothèses) + Processus itératif

- **Etape 3 (post-étude) : Validation des projets réussis et industrialisation (production)**

AGENDA

Les facteurs clés de succès : A.I.M.E.R

ADOPTER UNE DEMARCHE EN 5 ETAPES

- La valorisation des données est la résultante d'un processus en 5 étapes :

  ➔ ***ANIMER*** *: faire émerger les 3 piliers (une équipe, un lieu et une approche), diffuser la culture, communiquer les enjeux et définir les règles du jeu*

  ➔ ***IMAGINER*** *: Use cases (KPI, méthode d'évaluation, sélection des PoC)*

  ➔ ***MATERIALISER*** *: définir une architecture et collecter les données (internes, externes…)*

  ➔ ***EXPLOITER*** *: réaliser une courte démonstration de la faisabilité via un PoC, valoriser les données puis communiquer (décrire, prédire, prescrire et visualiser…)*

  ➔ ***RELAYER*** *: mise en production / industrialisation des PoC rentables*

AGENDA

1	Contexte et enjeux pour CNP Assurances
2	Structuration de la démarche au sein du Groupe
3	Les Facteurs clés de succès
4	Retour d'expérience : optimisation d'une campagne commerciale
5	Conclusion

Retour d'expérience : Ciblage des clients arbitreurs

LES ENJEUX DE CE PoC

- AUGMENTER LA PART UC ET DIMINUER LA PART EUROS DU PORTEFEUILLE EPARGNE

  ➔ ***Démarche pédagogique et expérimentale*** *: démontrer l'apport des nouvelles techniques analytiques issues de la data science sur de la prédiction d'arbitrage Euro vers UC à des fins marketing*

  ➔ ***Connaissance client / Ciblage et Prédiction / Campagne commerciale***

CIBLE

- Quel évènement souhaitons-nous prédire?

  ➔ ***Cibler les clients possédant des contrats Euros/UC***

  ➔ ***Présents au moins au 31/12/2013***

  ➔ ***ayant réalisé au moins un arbitrage de l'Euro vers UC quelque soit son montant***

  ➔ ***sur le premier semestre 2014***

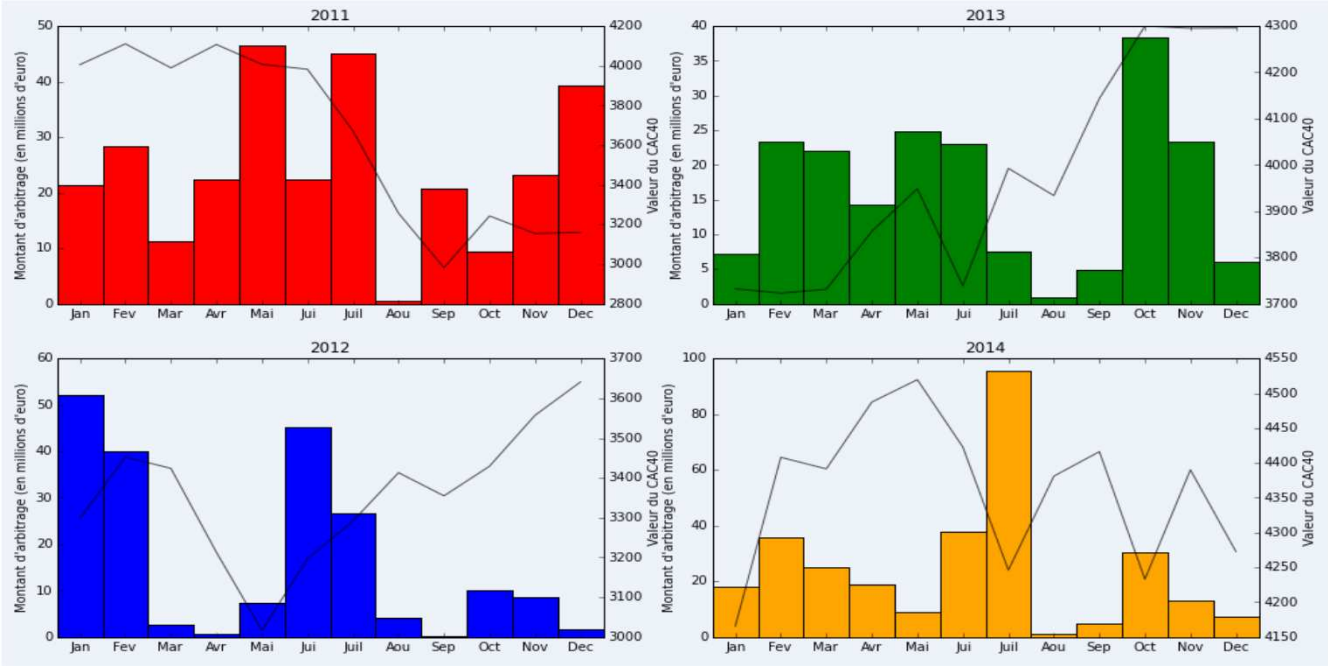
Retour d'expérience : Ciblage des clients arbitreurs

CARACTERISTIQUES : POPULATION ET DES DONNEES D'ETUDE

- **Portefeuille d'étude** : contrat d'épargne multi-support grand public
- **Volumétrie et structure des données disponibles** (40 GB de données pour l'apprentissage)
- Taux de cible d'arbitrage sur S1 inférieur à 1% (moyenne autour de 0,7 %)

Retour d'expérience : Ciblage des clients arbitreurs

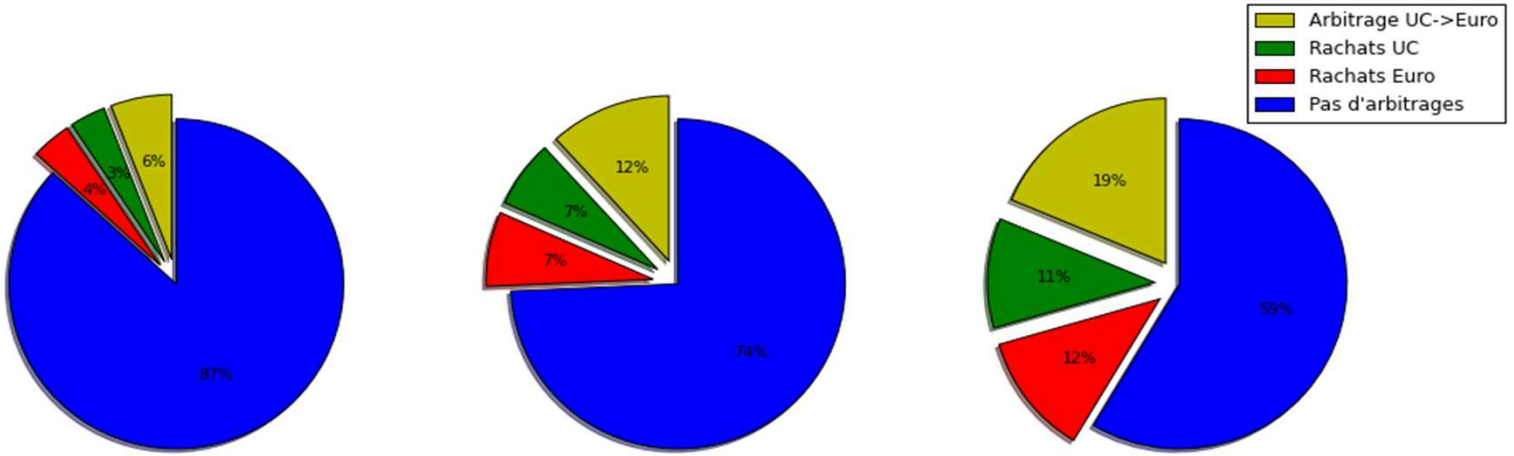
Statistiques descriptives : effets de l'économie sur le comportement étudié



➔ **Campagne commerciale et évolution du CAC40 sont à intégrer dans la modélisation**

Retour d'expérience : Ciblage des clients arbitreurs

Statistiques descriptives : durée de détention des supports UC



t_0+12 mois t_0+24 mois t_0+36 mois

87% des arbitreurs conservent l'intégralité de leur UC après 12 mois contre 59% après 3 ans

4, rue Chauveau-Lagarde - 75008 Paris - Tél : 01 44 51 72 72 - Fax : 01 44 51 72 73
www.institutdesactuaires.com

100% DATA SCIENCE

Retour d'expérience : Ciblage des clients arbitreurs

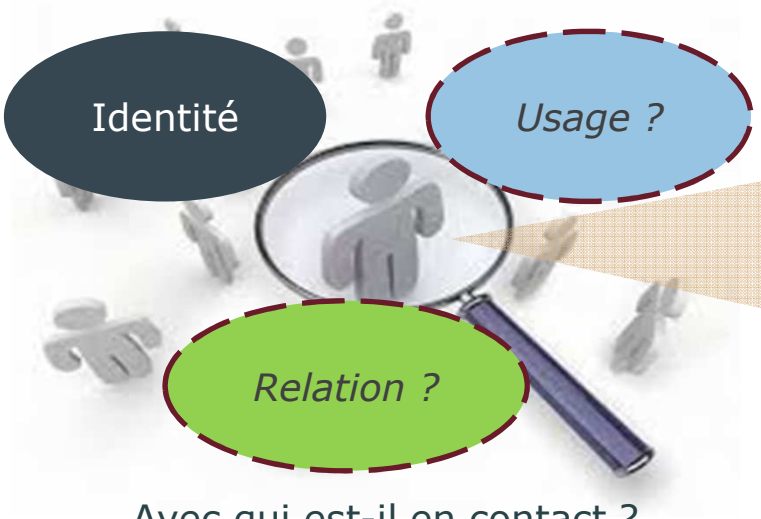
125 ANS INSTITUT DES ACTUAIRES

Comment modéliser les comportements des clients ?

Quelles données utilisées ?

Nom, prénom, adresse, CSP, etc.

Comment utilise-t-il mes produits et services ?



Avec qui est-il en contact ?
Qu'est-ce qui l'influence ?
Qui influence-t-il ?

Historiquement dans les services financiers, le client est principalement décrit par sa donnée identitaire.

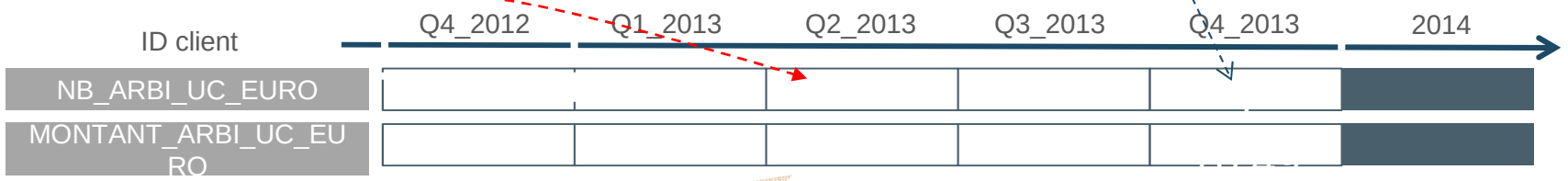
➔ L'approche analytics permet d'utiliser les apports des variables comportementales et d'influences.

➔ Chaque client est représenté par **700 variables** (Identité + Usage + Relation) provenant de sources internes comme externes**!**

Retour d'expérience : Ciblage des clients arbitreurs

Modélisation : comment modéliser les comportements des clients ?

	NUM_CONTRAT	LIB_EVT_ACT	DATE_EFFET	MONTANT_FLUX	Type_support	NOM	PRENOM	DATE_NAISSANCE
198	617000067	ARBITRAGE ENT AUTOMATIQUE ENTRE FONDS	31/07/2013	76.62	euro	HEMERYCK	YLVIE	30/03/1957
199	617000067	ARBITRAGE ENT AUTOMATIQUE ENTRE FONDS	31/10/2013	76.49	euro	HEMERYCK	YLVIE	30/03/1957
201	617000067	ARBITRAGE SORT AUTOMATIQUE ENTRE FONDS	31/07/2013	76.62	UC	HEMERYCK	SYLVIE	30/03/1957
202	617000067	ARBITRAGE SORT AUTOMATIQUE ENTRE FONDS	31/10/2013	76.49	UC	HEMERYCK	SYLVIE	30/03/1957

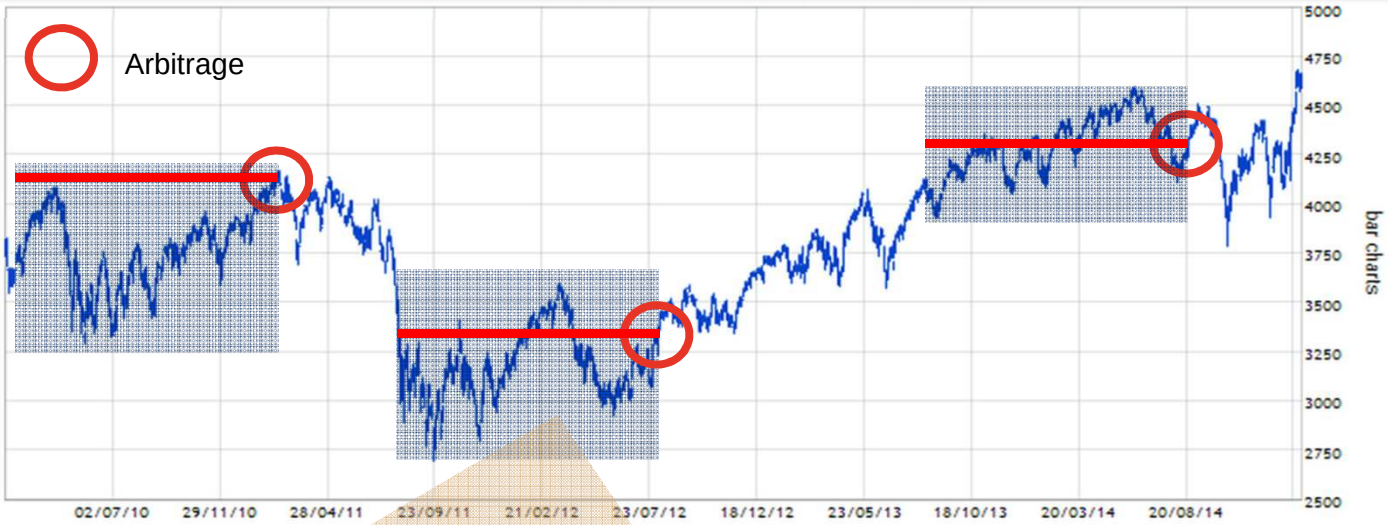


Les variables issues des transactions sont nos principaux fournisseurs de **données comportementales**

➔ Tous les événements liés à la vie d'un contrat d'assurance-vie de types arbitrage, versement (régulier ou libre), rachat, …

Retour d'expérience : Ciblage des clients arbitreurs

Modélisation : comment modéliser l'influence?



Pour chaque arbitrage on observe la valeur courante du CAC40 et les valeurs passées sur les 12 derniers mois :

➔ **Pour chaque arbitrage :**
+ On observe la valeur courante du CAC40
+ On observe les valeurs passées du CAC40 sur les 12 derniers mois
+ On retourne le décile correspondant à la valeur courante du CAC40

Retour d'expérience : Ciblage des clients arbitreurs

Modélisation : quelles sont les variables exogènes retenues ?

- Utilisation de 2 types de variables exogènes:
  - ✓ **Variable géospatiale** : elles enrichissent la donnée client sur la base de son lieu de résidence

    Ex. nombre d'habitants de la commune, niveau de vie, patrimoine, revenu moyen de la commune, impôt sur le revenu, nombre de commerçants de la commune, bassin d'emploi, etc…

  - ✓ **Variable temporelle :** elles enrichissent la période sur laquelle on réalise l'apprentissage

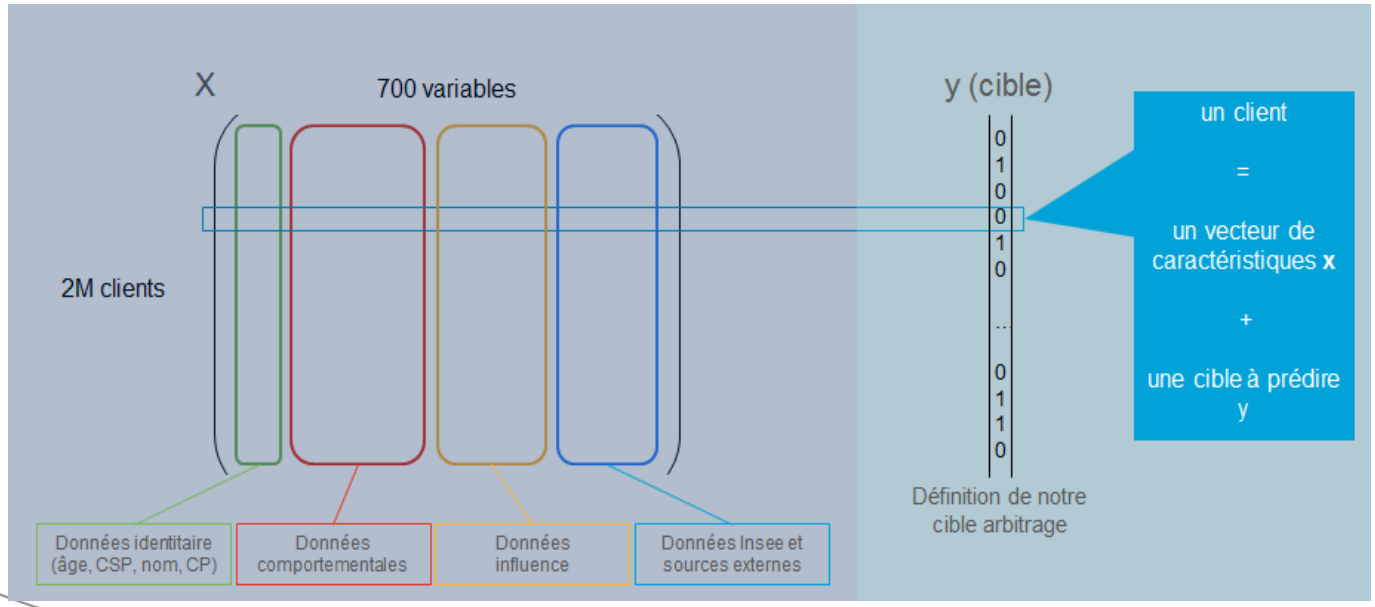
    Ex. indice de confiance des ménages, enquête de conjoncture, taux d'intérêt, immobilier, etc…

Retour d'expérience : Ciblage des clients arbitreurs

Modélisation : représentation des caractéristiques des clients

- Un **client** = caractéristiques statiques (identité) + dynamiques (transactions, conjoncture, etc.) liées à une période donnée P de 5 ans
- La cible à prédire est la présence d'un arbitrage dans le semestre qui suit P (➔ 2014)

Période d'apprentissage : 2009 - 2013 **Prédiction : 2014**

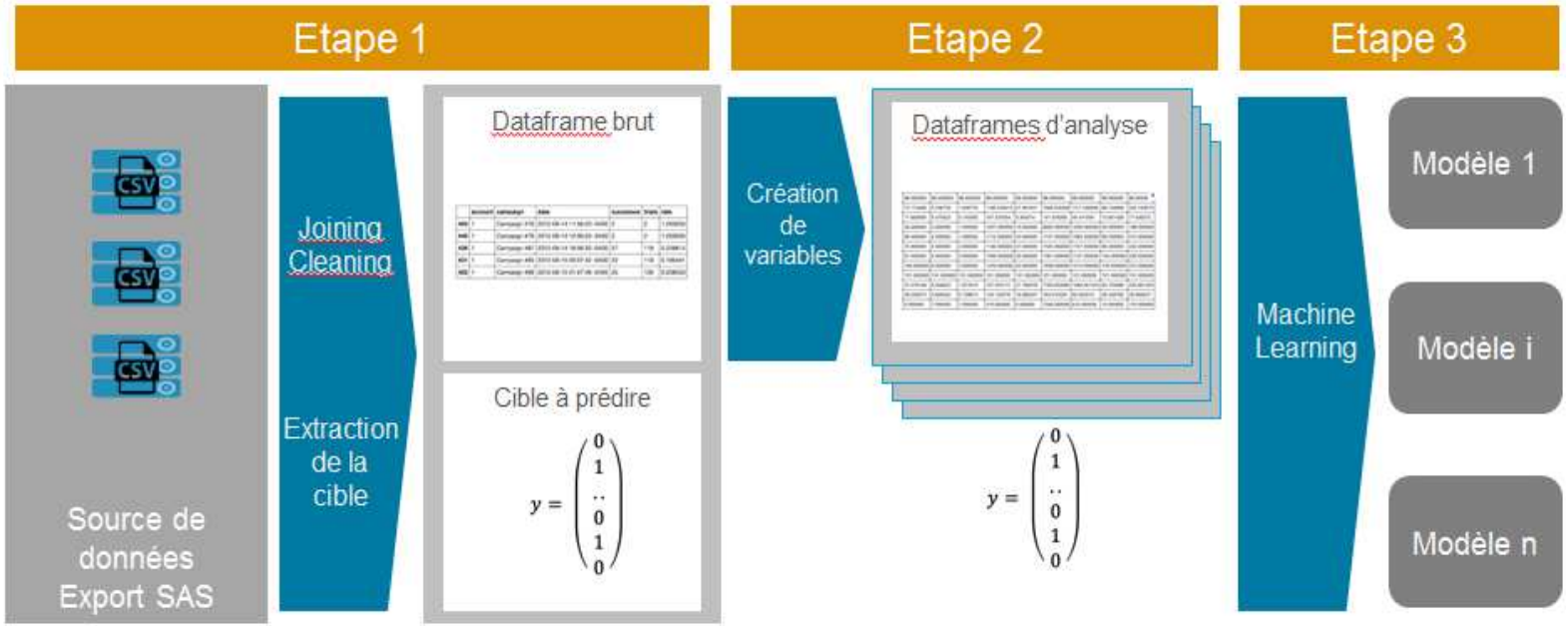


4, rue Chauveau-Lagarde - 75008 Paris - Tél : 01 44 51 72 72 - Fax : 01 44 51 72 73
www.institutdesactuaires.com

Retour d'expérience : Ciblage des clients arbitreurs

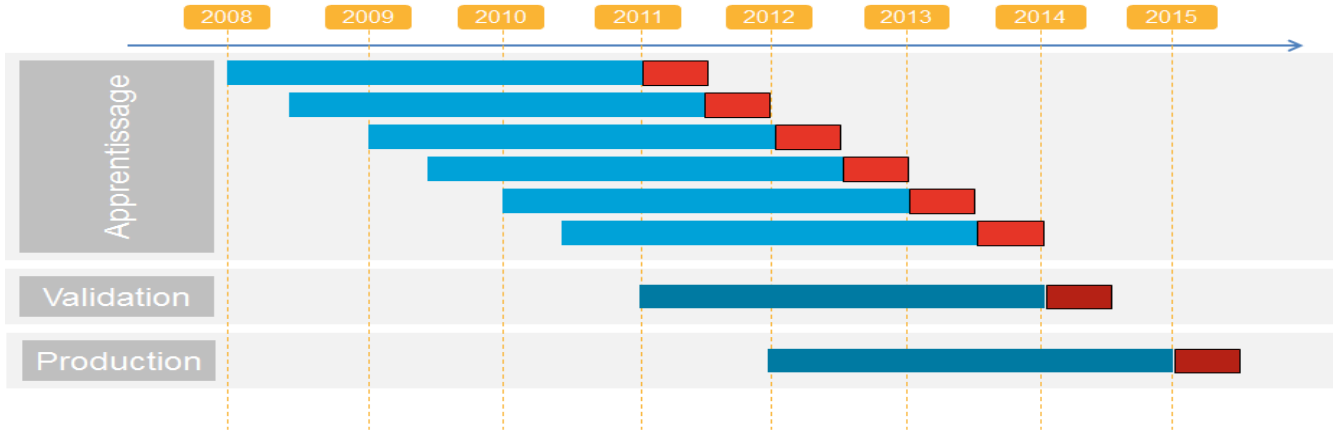
Modélisation : méthodologie globale des traitements

- Voici une vue globale de la chaine de traitement entre les sources de données jusqu'aux modèles d'apprentissage :



Retour d'expérience : Ciblage des clients arbitreurs

Modélisation temporelle : phase de production



- Afin de ne pas rendre le modèle trop dépendant d'un contexte économique donné :

➔ *Modéliser le problème avec plusieurs périodes d'apprentissage entre 2009 et fin 2013*

➔ *Tirage d'un échantillon de clients sur chacune des périodes d'apprentissage*

- La validation du modèle ➔ prédiction des arbitrages du S1 2014 sur la base de la période 2011-2013

- Enfin, pour l'utilisation du modèle à des fins de production,

➔ *on utilise un modèle **entrainé** sur des échantillons entre **2008 et juillet 2013**, **testé** sur **S1 2014***

➔ *sur la base de la période 2008 – 2014, **prédit** sur le **S1 2015** (pour la production)*

Retour d'expérience : Ciblage des clients arbitreurs

Evaluation et performance : choix des métriques d'évaluation

- Quand le taux de cible est faible (taux cible < 1% dans notre exemple), le choix de la métrique d'évaluation des modèles est très important

- Nous avons retenu 2 métriques qui permettent de gérer correctement la spécificité du taux de cible faible :

  - ✓ **Aire sous la courbe ROC (AUC ou Area Under Curve)**

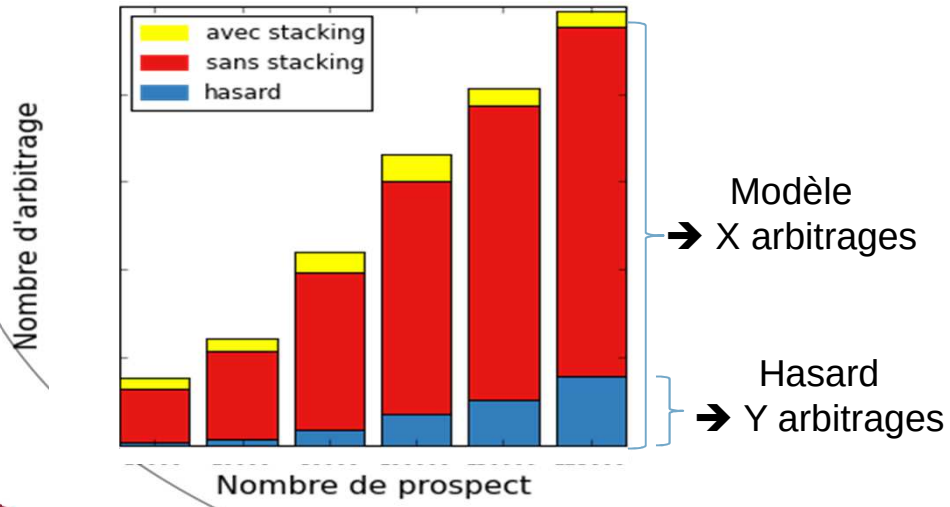
  - ✓ **Le Lift 10% et le Lift 5%**

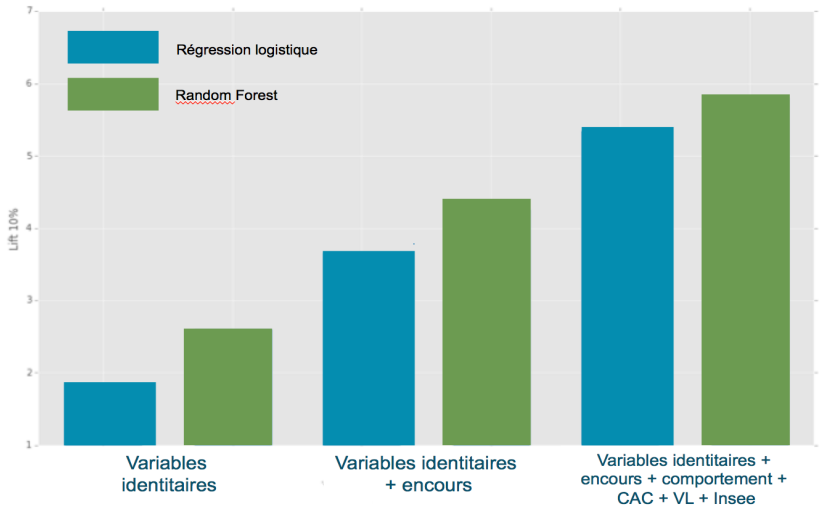
Retour d'expérience : Ciblage des clients arbitreurs

Evaluation et performance : choix des métriques d'évaluation

- Le Lift permet d'apprécier à quel point le modèle est meilleur que l'aléatoire

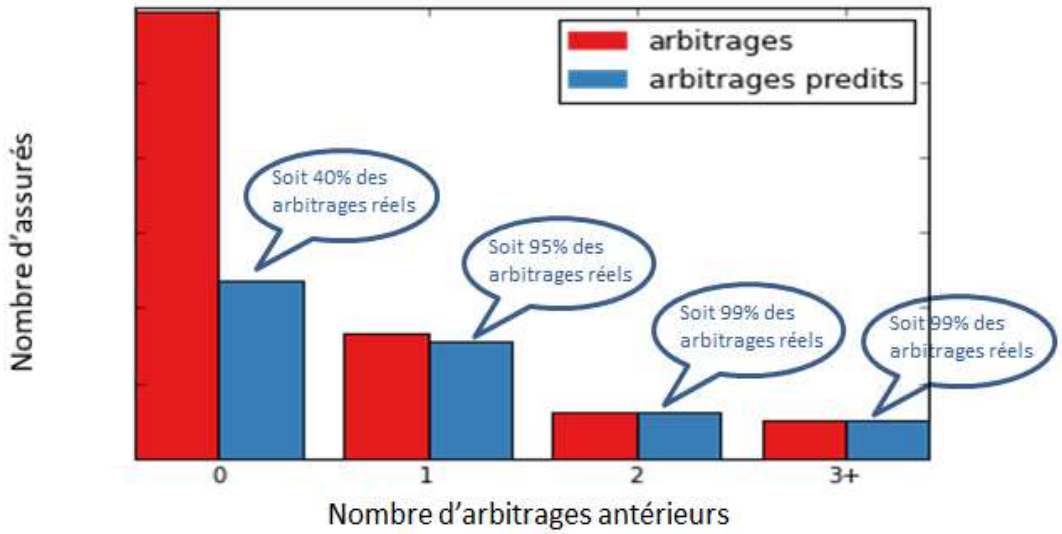
Lift 10%[1] = X / Y





Retour d'expérience : Ciblage des clients arbitreurs

Les résultats : qualité de la prédiction selon l'activité passée de l'assuré



- ✓ **Apports d'un ciblage par le modèle** : capacité du modèle à prédire les primo-arbitreurs
  - o Il s'agit des **assurés n'ayant encore jamais arbitré** mais qui arbitreront en 2014
  - o La population des «primo-arbitreurs» est **difficilement identifiable à partir d'un choix au hasard** (aléatoire)
  - o Population **représentant un enjeu business important** (constitue l'essentiel des arbitreurs)
  - o Sur de l'année 2014, le **modèle prédit plus de 40%** des « primo-arbitreurs »

Retour d'expérience : Ciblage des clients arbitreurs

Les résultats : qualité de la prédiction selon l'activité passée de l'assuré

- ✓ Parmi les clients ciblés par le modèle calibré sur S1 avec un Lift 9%, on constat que :
  - o une 1ère partie arbitre au cours du S1 2014
  - o une 2nde partie arbitre au cours du **S2 2014**
  - o une 3ème partie fait des **versements libres en UC** au cours de 2014

AGENDA

1	Contexte et enjeux pour CNP Assurances
2	Structuration de la démarche au sein du Groupe
3	Les Facteurs clés de succès
4	Retour d'expérience : optimisation d'une campagne commerciale
5	Conclusion

WORK IN PROGRESS…

- Facteurs clés de succès = un processus en ***5 étapes***
  - ✓ ***ANIMER***
  - ✓ ***IMAGINER***
  - ✓ ***MATERIALISER***
  - ✓ ***EXPLOITER***
  - ✓ ***RELAYER***
- L'approche dite « ***Business Driven*** » est efficace :
  - ✓ Facilite la **communication** et la **mobilisation** de l'organisation jusqu'au plus haut niveau
  - ✓ Assure le succès de la démarche car prônant la **recherche de use case sous un angle business**
- Plusieurs projets souhaités suite aux premières études