

The logo consists of two white geometric shapes: a smaller triangle pointing upwards and to the right, and a larger, wider triangle pointing upwards and to the right, partially overlapping the first one.

INSTITUT DES
ACTUAIRES

Détecter le projet d'achat d'un véhicule de nos clients en portefeuille

Applications au marketing direct



Atelier animé par Mickaël DUPONT
Expert en data science pour l'actuariat
COVEA – Direction Technique IARD des particuliers – Cellule R&D

Nos préoccupations dans un contexte de loi Hamon

Conserver nos parts de marché

Maîtriser l'augmentation des coûts d'acquisition

La fidélisation de nos clients est essentielle

Comment améliorer nos actions de fidélisation ?

Anticiper les nouveaux besoins d'assurance de nos clients ...

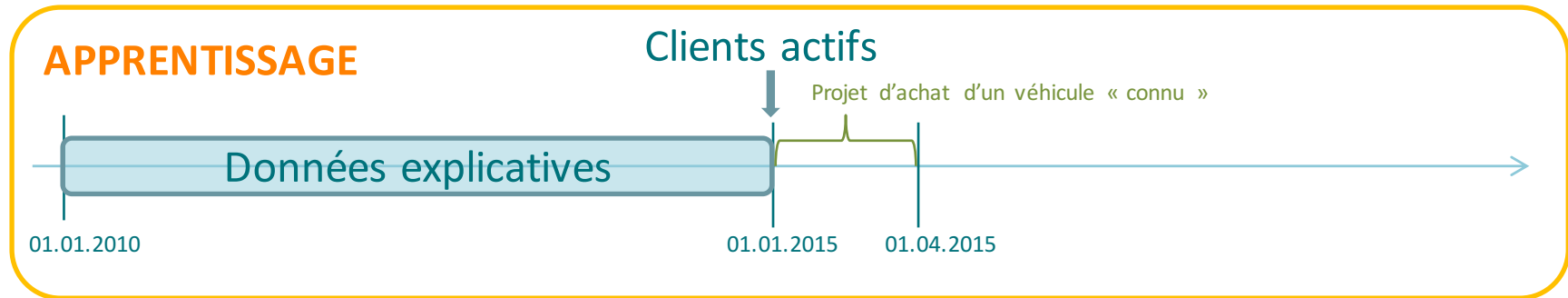
... pour renforcer de manière pro active notre relation

Problématique

Le projet d'achat d'un véhicule est un moment de fragilité

Disposer d'un modèle pour anticiper cet événement

Une démarche d'analyse prédictive « classique »



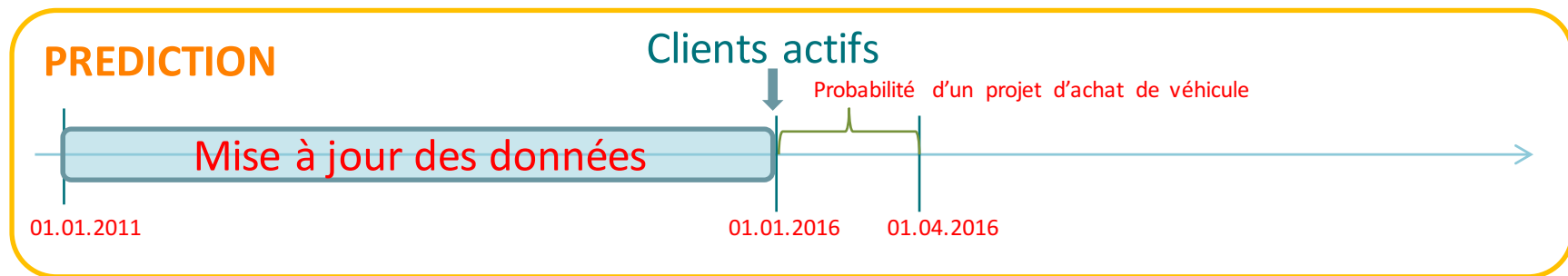
Mais avec 3 nouveautés, objectifs

Intégrer des données internes non utilisées actuellement à la MAAF : des données textuelles ...

Ajouter une étape de feature engineering : variables d'évolution, dater les événements antérieurs ...

Utiliser l'approche machine learning : algorithmes, cross-validation, généralisation

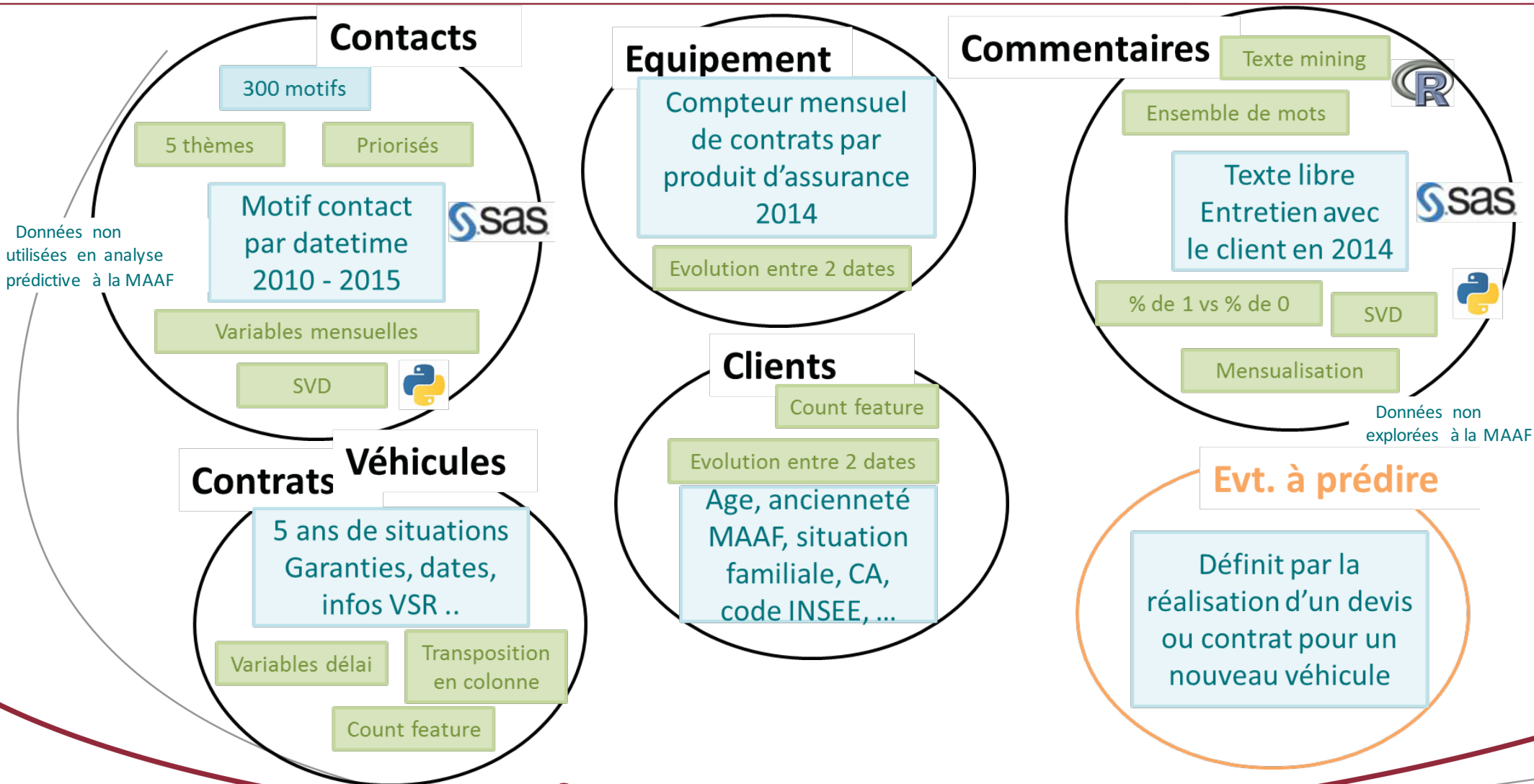
Dans le but de prédire au mieux en 2016



La base de données pour l'apprentissage

La construction du modèle

La pertinence des prédictions, quelles applications ?



2M de clients, 1 255 variables

Les algorithmes d'apprentissage testés

xgboost, Random Forest, Extra Trees

Mesure de performance : le lift

Maximiser la proportion d'événements positifs retrouvés dans une cible de taille définie

Stratégie de cross validation

50% des données pour apprendre (ajuster les modèles)

20% des données pour mesurer la performance et optimiser les modèles

30% des données pour s'assurer du pouvoir de généralisation du modèle

Lift $_{10\%} = 3$

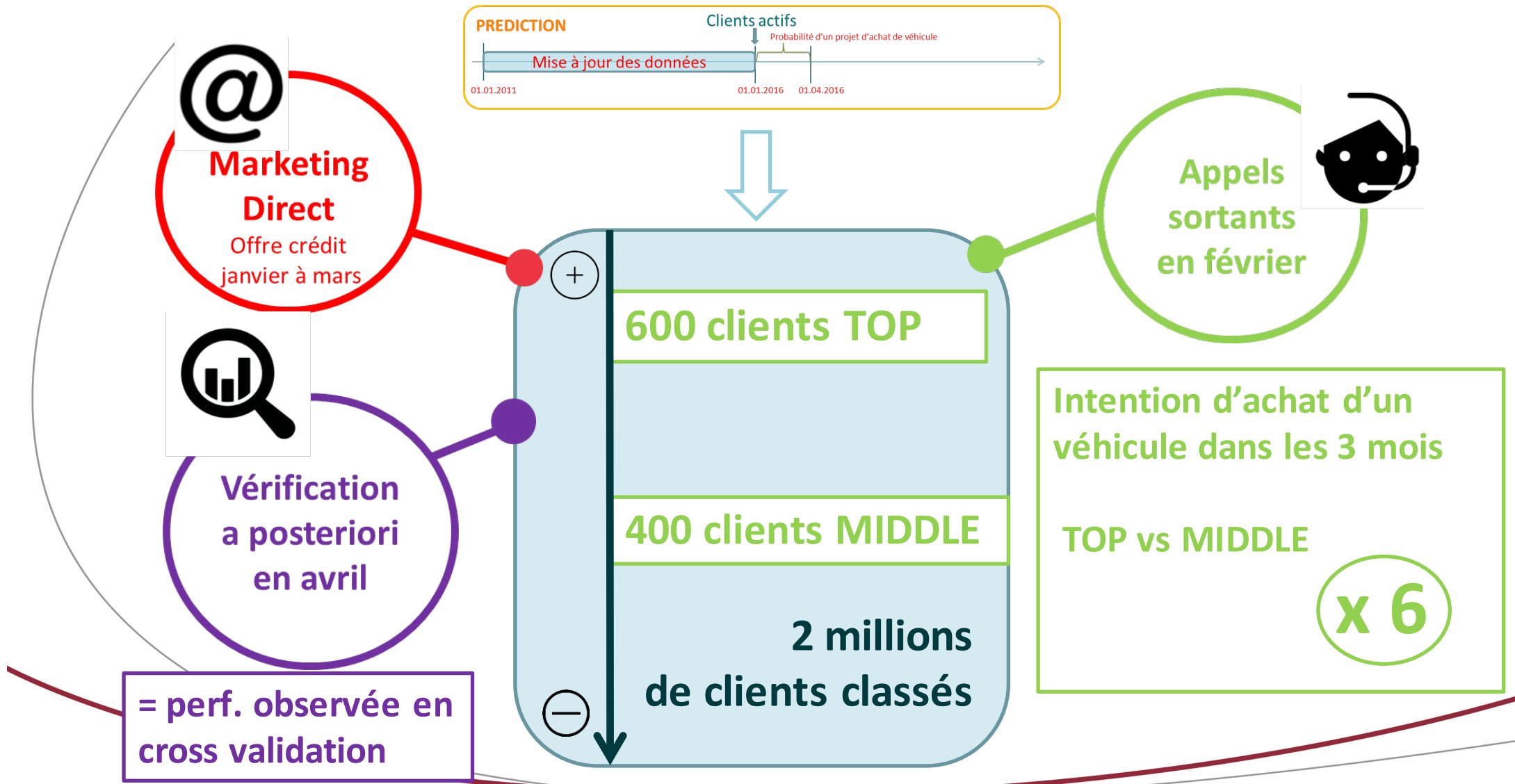
Les variables importantes sont en partie issues de feature engineering :

SVD commentaires, âge du client et CA, délais entre les véhicules, count feature sur les véhicules et codes Insee, variables mensuelles contact

Avant d'explorer les pistes d'amélioration

Souhait de mesurer la performance en situation réelle

Valider la pertinence de la définition de l'événement à prédire



L'approche machine learning

Des algorithmes résistants au bruit, à la redondance et aux corrélations

Accélère l'étape de modélisation

Une opportunité d'être plus créatif sur l'étape de génération de variables explicatives

Un score intéressant pour le canal téléphone

Projet d'industrialisation, même si nous avons identifié des pistes d'amélioration : enrichir avec d'autres domaines de données (maaf.fr, données externes, ...), sélection de variables, ...

Spécialiser le score pour l'action de mailing

Merci de votre attention