

**Prédiction des comportements
de rachats en épargne
individuelle: *une approche
machine learning***

Eric NANA NJOYA

Conférence débat - Institut des Actuares
25 Avril 2017

AGENDA

1

Motivation de la modélisation du rachat

2

Approches de modélisation

3

Un aperçu sur le machine learning

4

Application et résultats

5

Quelques conclusions...

Motivation de la modélisation du rachat



Financier et réglementaire

- Optimisation de la rentabilité
- Cadre réglementaire solvabilité II : estimation *best estimate* du risque de rachat



Marketing et design produits

- Compréhension des comportements clients :
 - ✓ Rétention client
 - ✓ Evolution des caractéristiques produits

Prédiction du risque

- Nouveaux clients et/ou affaires nouvelles
- Evaluation des pénalités en fonction du profil de rachat



Motivation de la modélisation du rachat

Deux hypothèses de rachat généralement formulées

Hypothèse de fond de réserve ou d'urgence

choc négatif non anticipé ou besoin de liquidité pour un grand projet

RACHAT

Hypothèse de rendement

En cas de hausse des taux arbitrage sur le marché vers des rendements meilleurs

RACHAT

Les rachats dépendent ainsi de plusieurs facteurs

- Facteurs structurels -> *Rachat structurel*
- Facteurs conjoncturels ou dynamiques -> *Rachat conjoncturel ou dynamique*

Une problématique de prédiction du rachat ...

- Quels sont les facteurs les plus discriminant du rachat?
- Pour un contrat donné, quelle serait sa probabilité de rachat?

Quatre grandes approches retrouvées dans la littérature

- ▶ Approche financière :
 - Valorisation d'une option américaine

- ▶ Approche statistique :
 - Modélisation collective sous forme de série temporelle

- ▶ Approche microéconomique:
 - Théorie de l'espérance d'utilité

- ▶ Approche probabiliste :
 - Modélisation individuelle du comportement



DETECTION DES DECLENCHEURS + PREDICTION DU RACHAT

Pourquoi le machine learning?

Limites modèles statistiques classiques

- La statistique classique fait en général des hypothèses sur la distribution de probabilité
- La détection et la modélisation des interactions qui doivent être spécifiés par le statisticien

Faibles capacités
prédictives

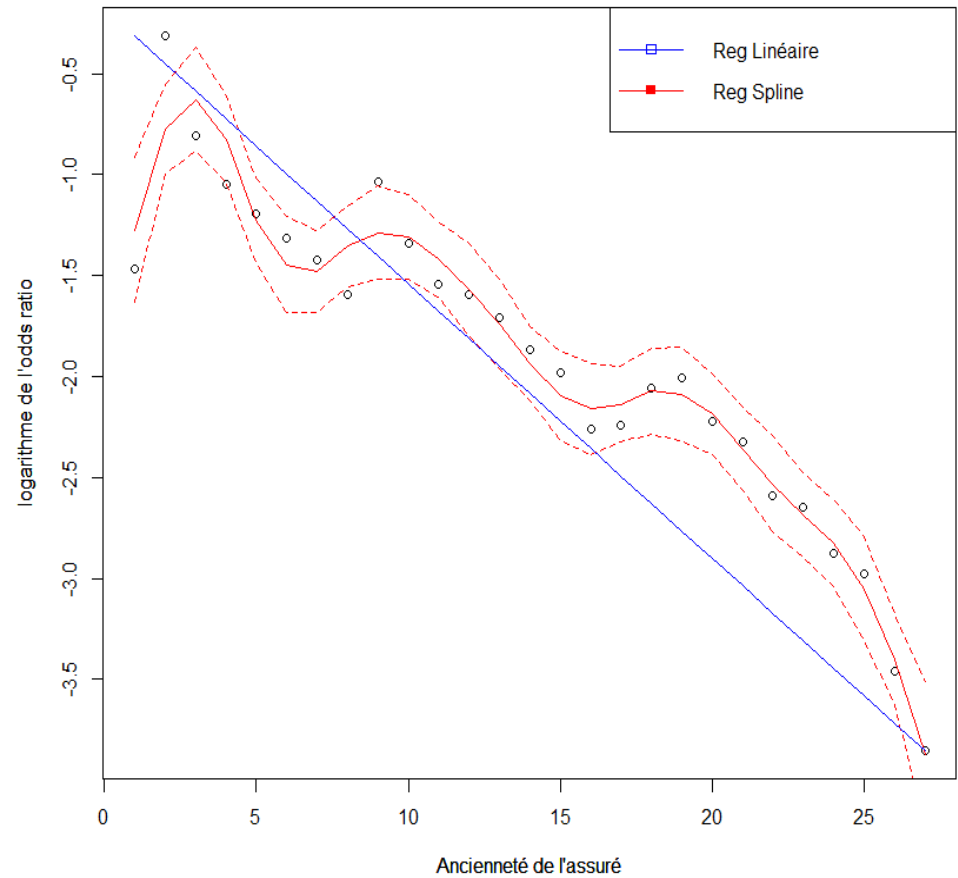
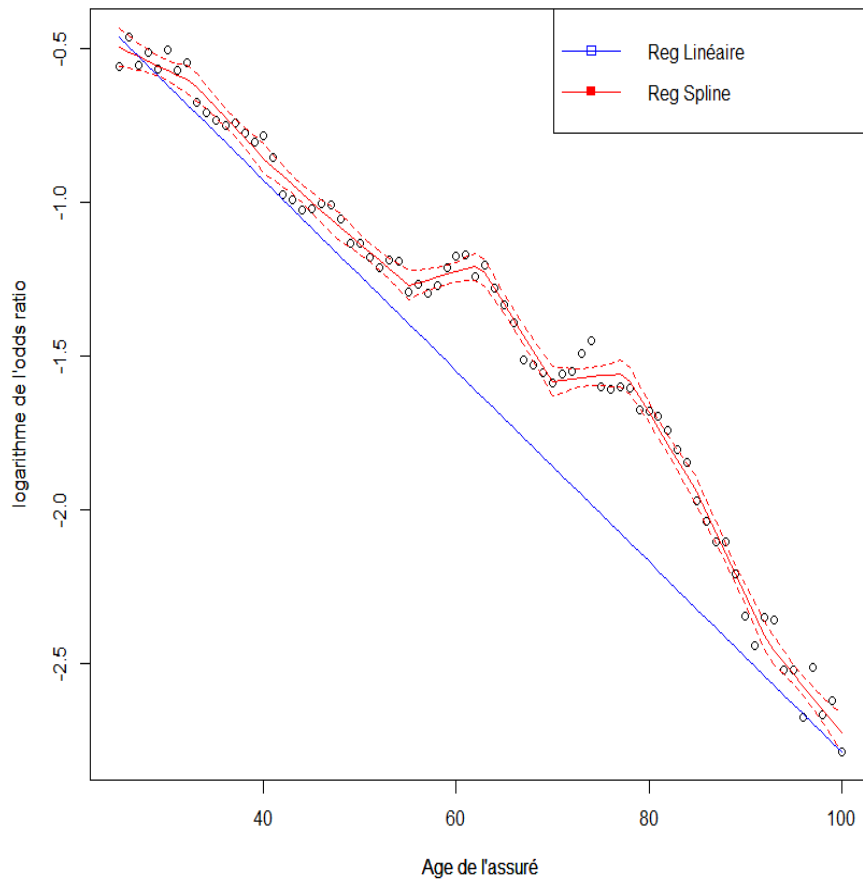
Les apports du machine learning

- De part sa nature non paramétrique, il ne fait aucune hypothèse de la de distribution
- Faculté de modéliser les dépendances complexes
- Technique de ré échantillonnage pour dé corrélés les estimateurs
- Plus un seul modèle, mais des modèles qui sont seront agrégés par la suite

Fortes capacités
prédictives

Vs

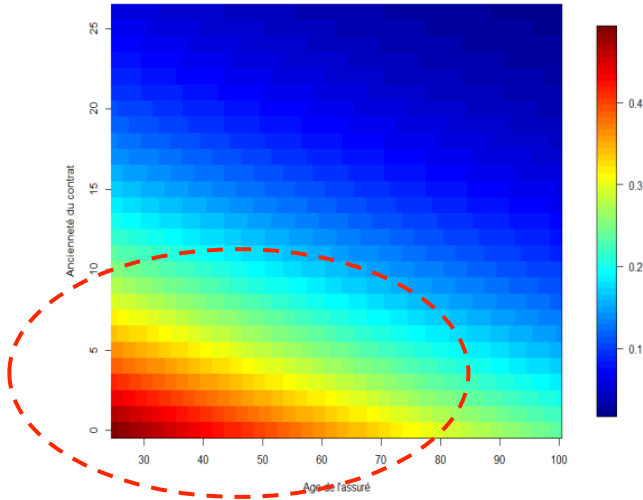
Paramétrique vs non-paramétrique: illustration



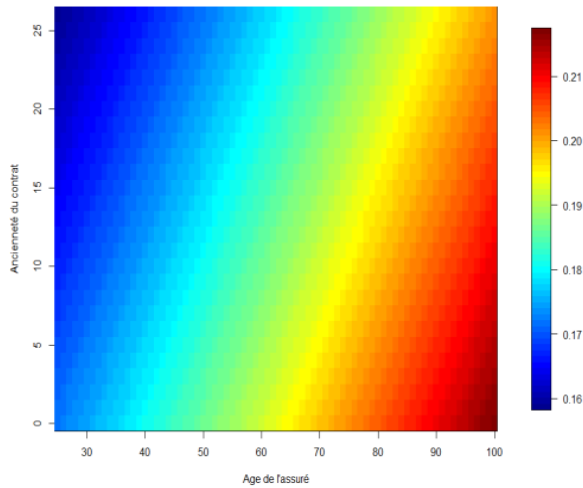
Un aperçu sur le machine learning

Méthodes classiques vs méthodes de machine learning: illustration effets croisés

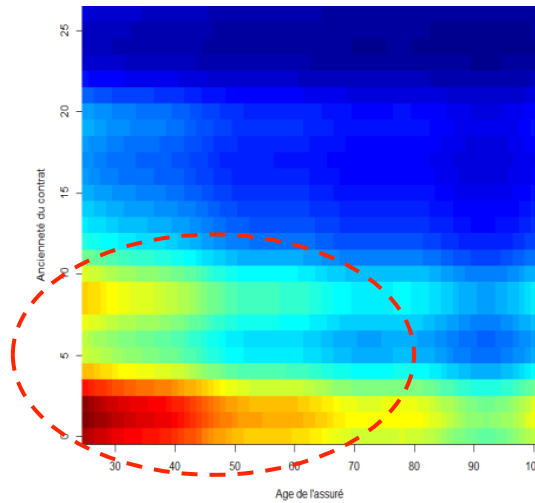
Modèle LR simple



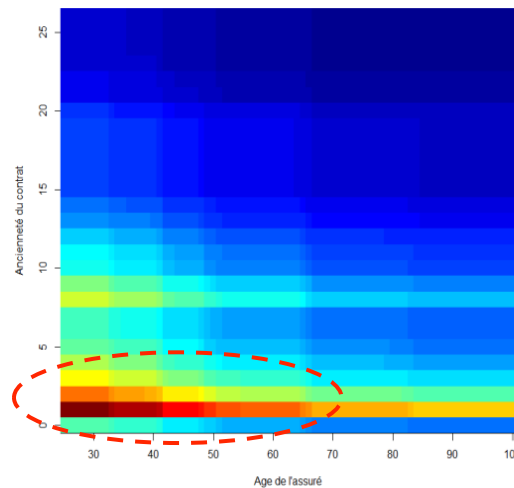
SVM Linéaire



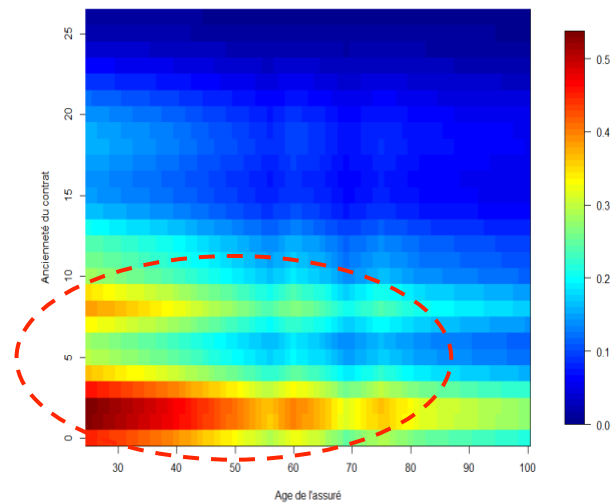
Modèle GAM



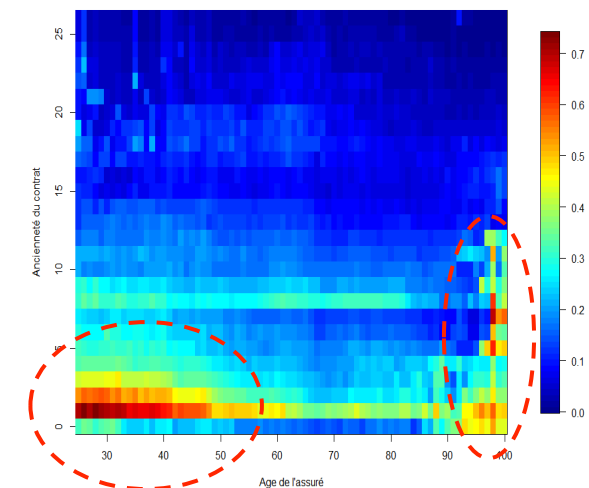
Boosting



Modèle LR avec Spline



Forêt aléatoire

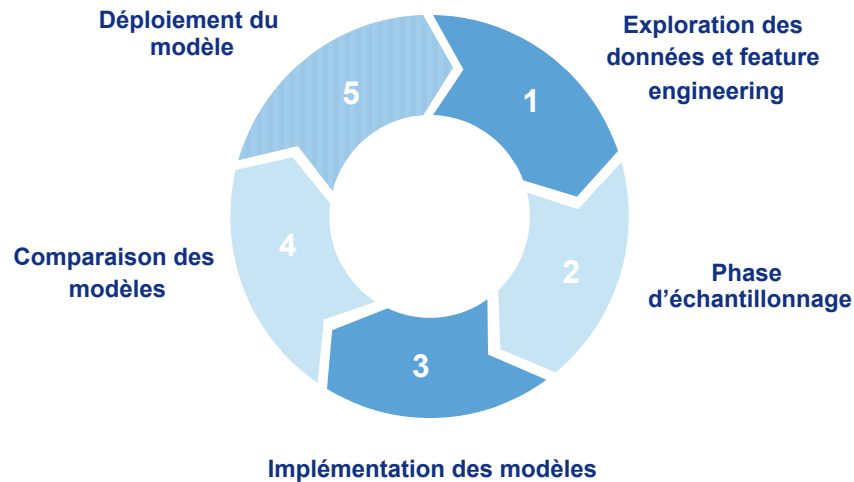


Un aperçu sur le machine learning

Démarche en 5 grandes étapes d'un projet de machine learning

- Phase opérationnelle
- Système d'information décisionnel

- Nettoyage et mise en forme des données
- feature engineering



- Choix du critère de performance
- En apprentissage supervisée (Courbe ROC, Lift, Indicateur de régression, etc...)

- Echantillon d'apprentissage
- Echantillon de validation
- Echantillon de test

- Apprentissage supervisée:
- Apprentissage non supervisée

Légende

 Prend moins de temps

 Prend beaucoup plus de temps

Périmètre d'application et logiciels utilisés

Périmètre de l'application

- Epargne individuelle
- Rachat total

Logiciel utilisé

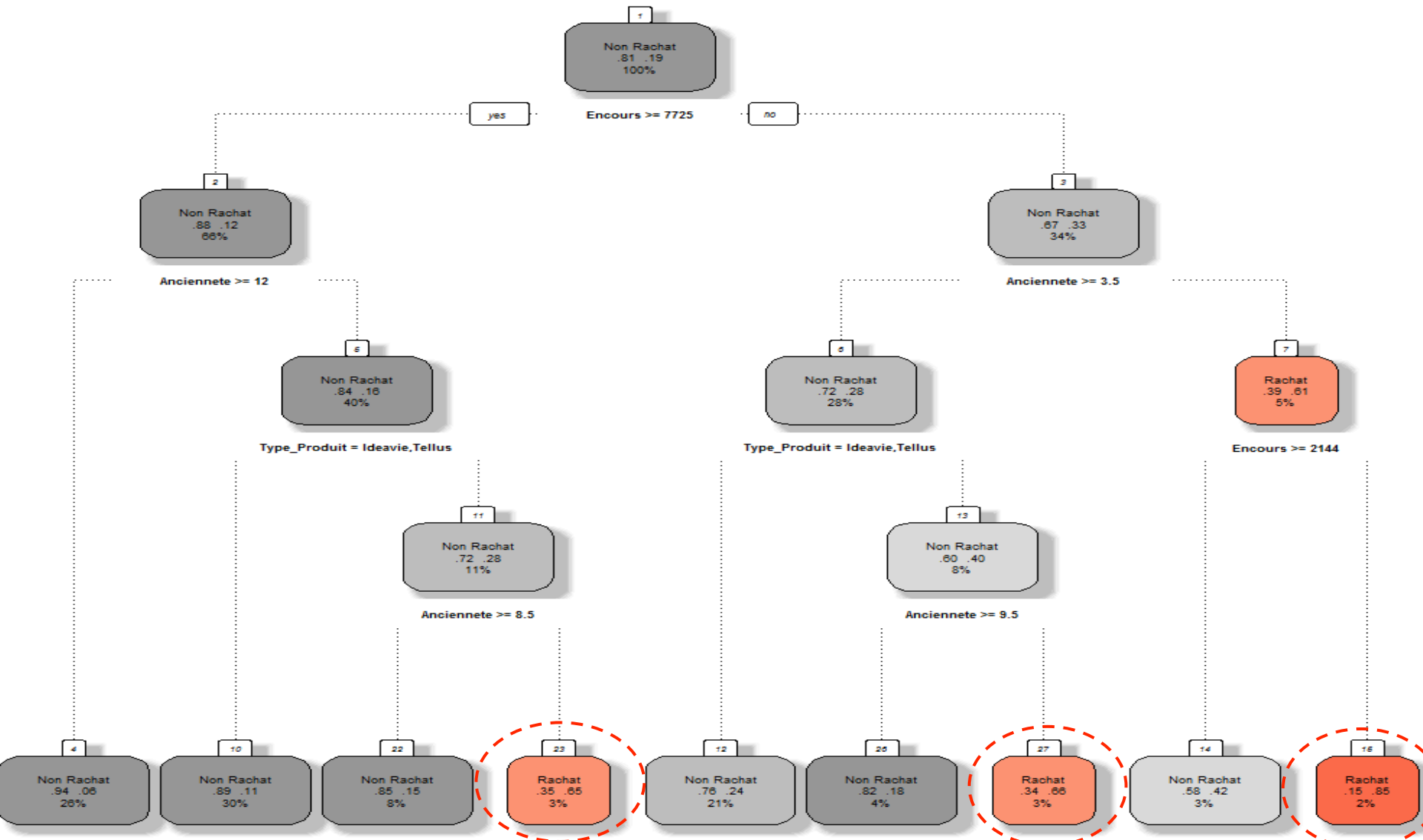
- SAS, R et Python

Variables et description générale

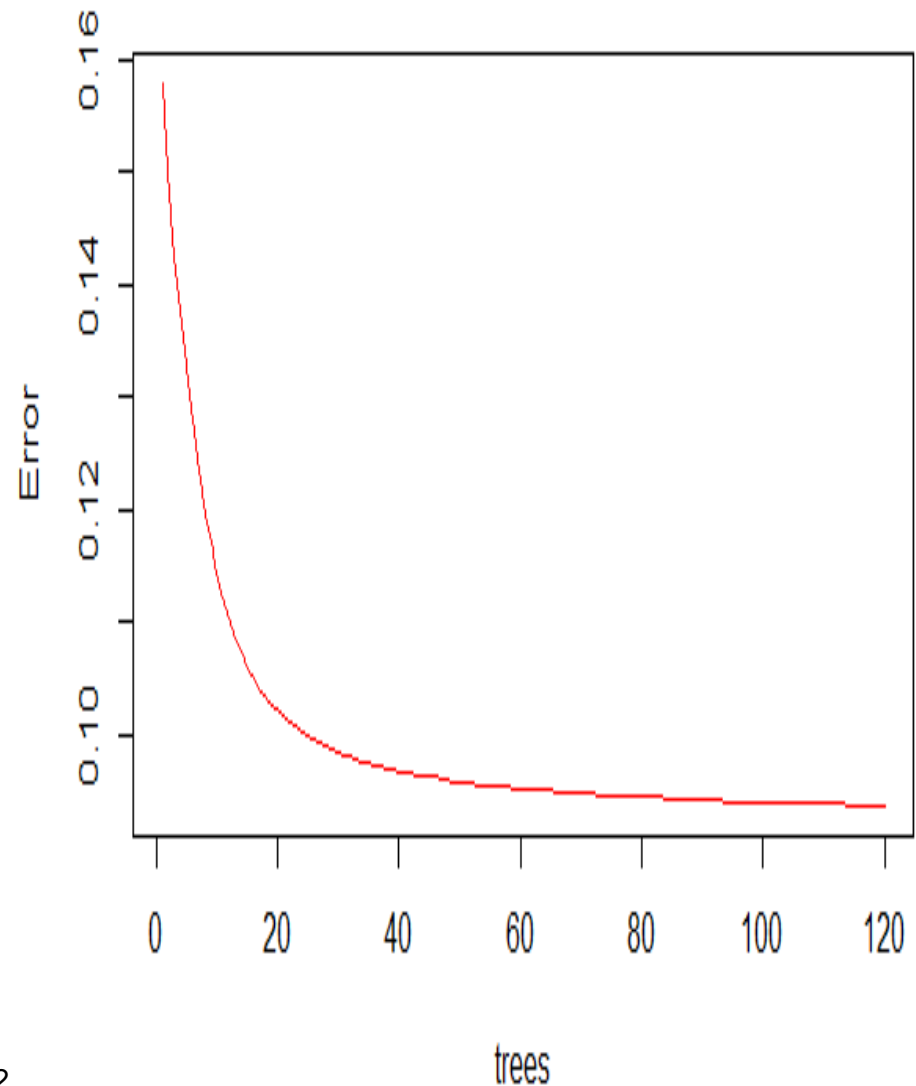
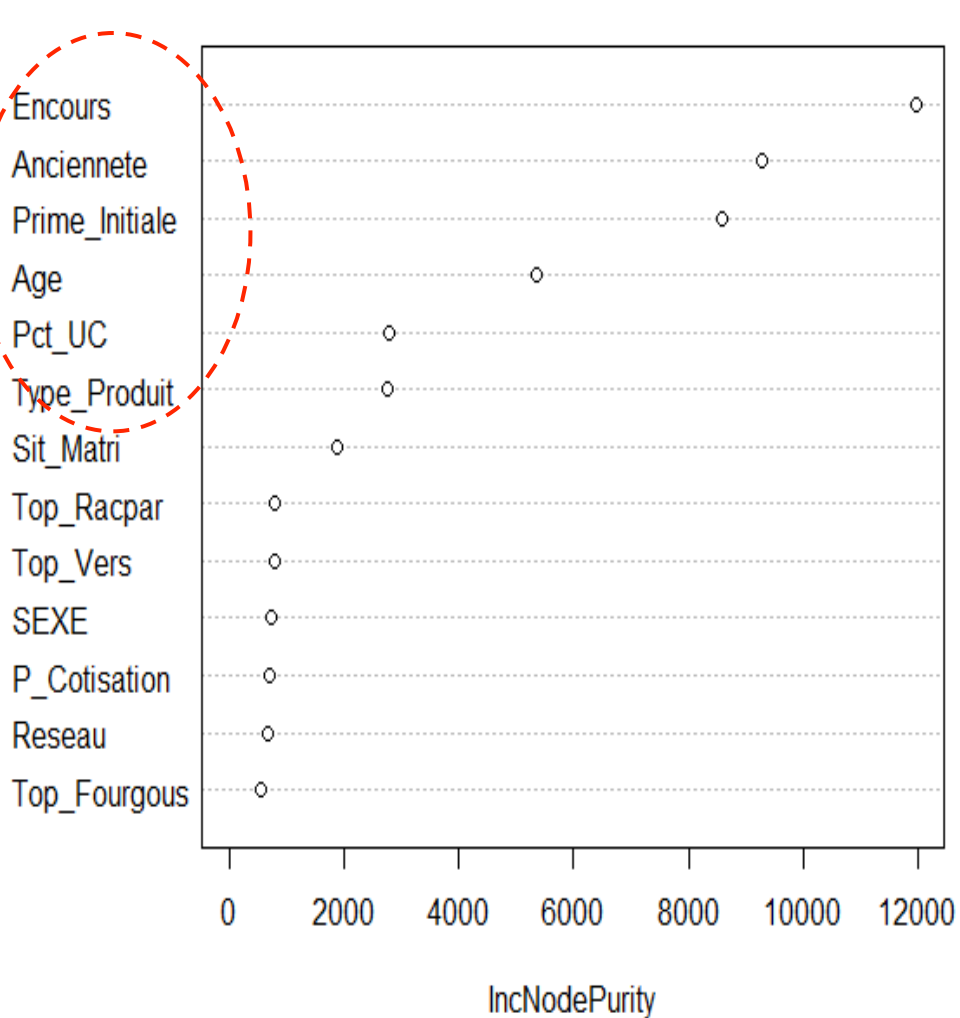
Variables

- **Variables structurels:** *ancienneté, encours, produit, réseau, taux d'UC, cotisation à la souscription, période de cotisation, top rachat partiel antérieur et fourgons, nombre de versement libre, sexe, âge, situation matrimoniale,...*
- **Variables dynamiques:** *taux de PB servi, taux d'intérêt 1Y et 10Y, TME, taux de chômage, taux d'inflation, volatilité CAC40, indice prix immobilier, indice de situation financière,...*

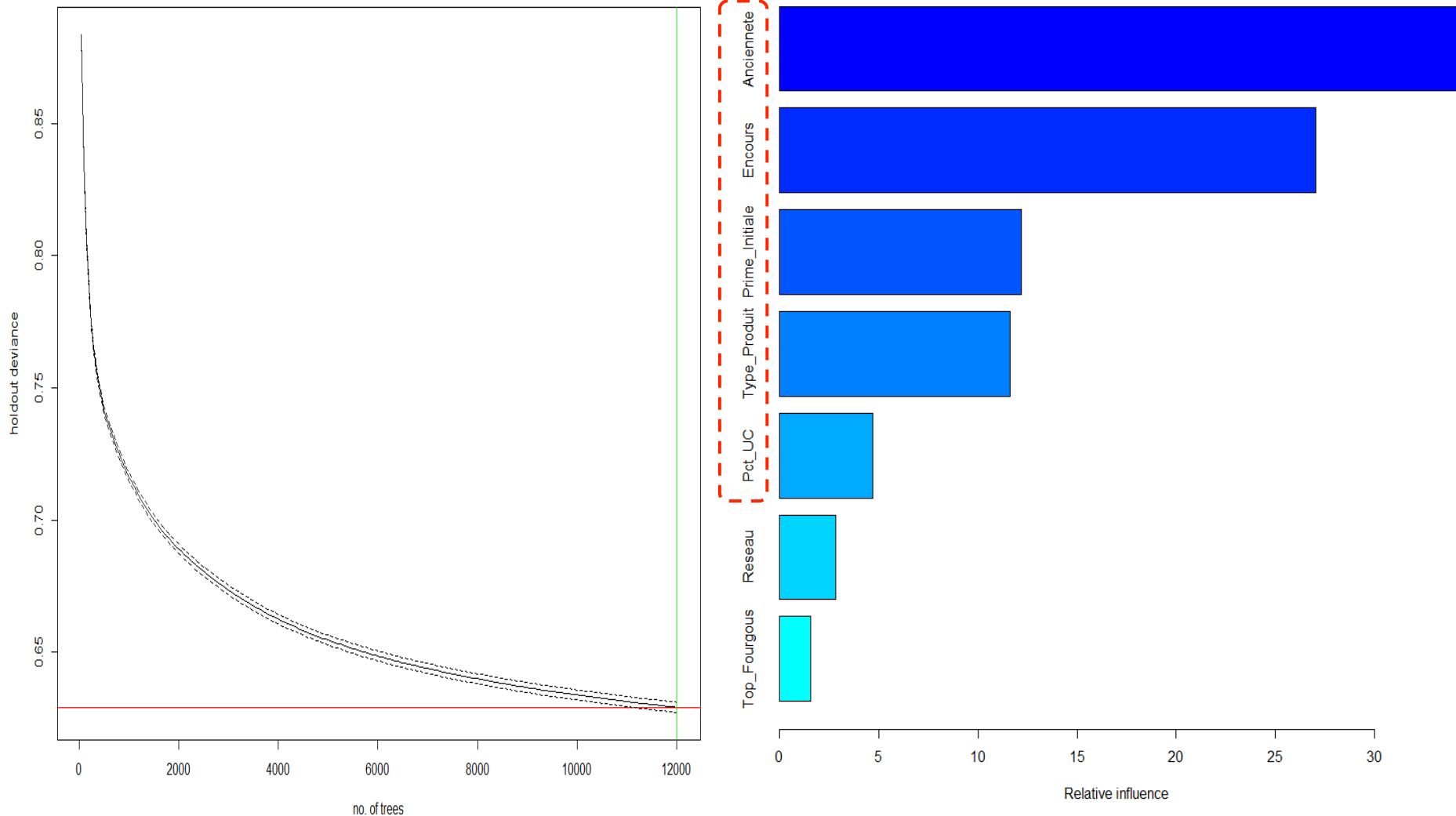
Comportements structurels: arbre de décision



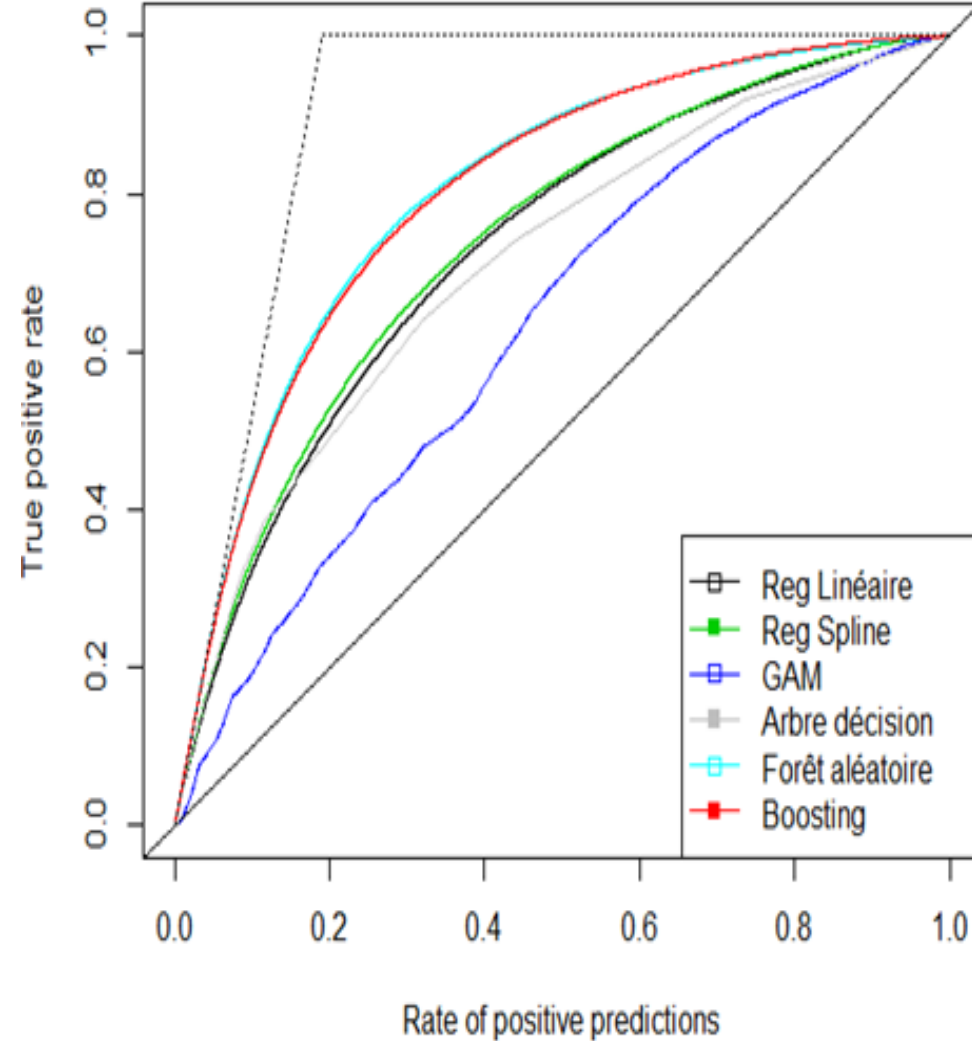
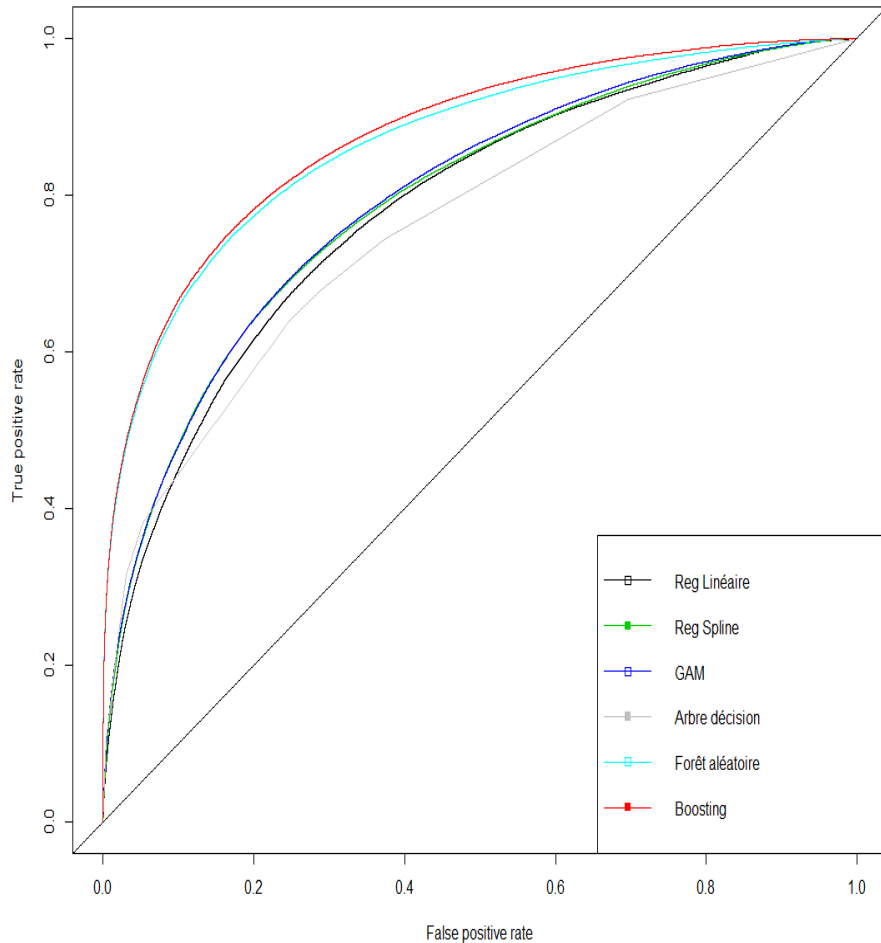
Comportements structurels: forêts aléatoires



Comportements structurels: Gradient Tree Boosting



Comportements structurels: Comparaison des modèles



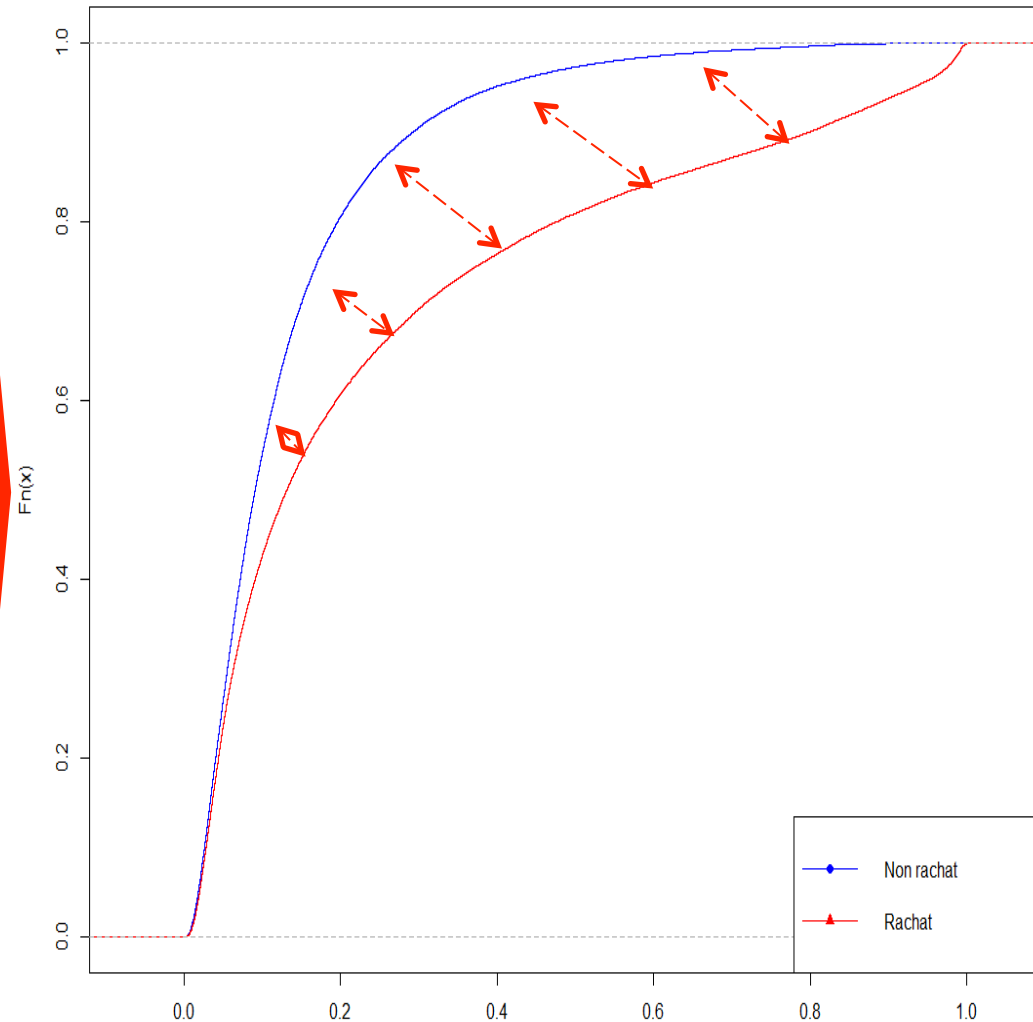
Comportements structurels: Estimation du taux de rachat moyen

Estimation du taux de rachat

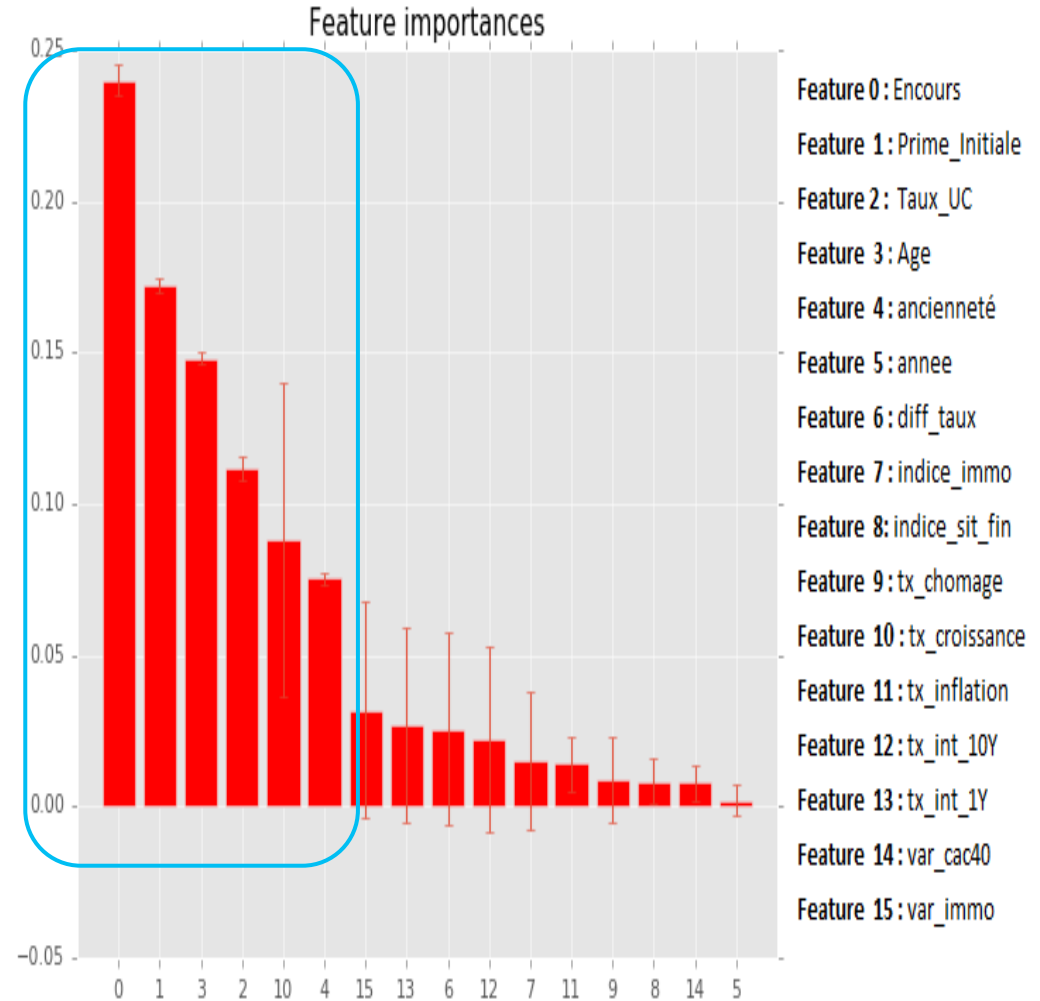
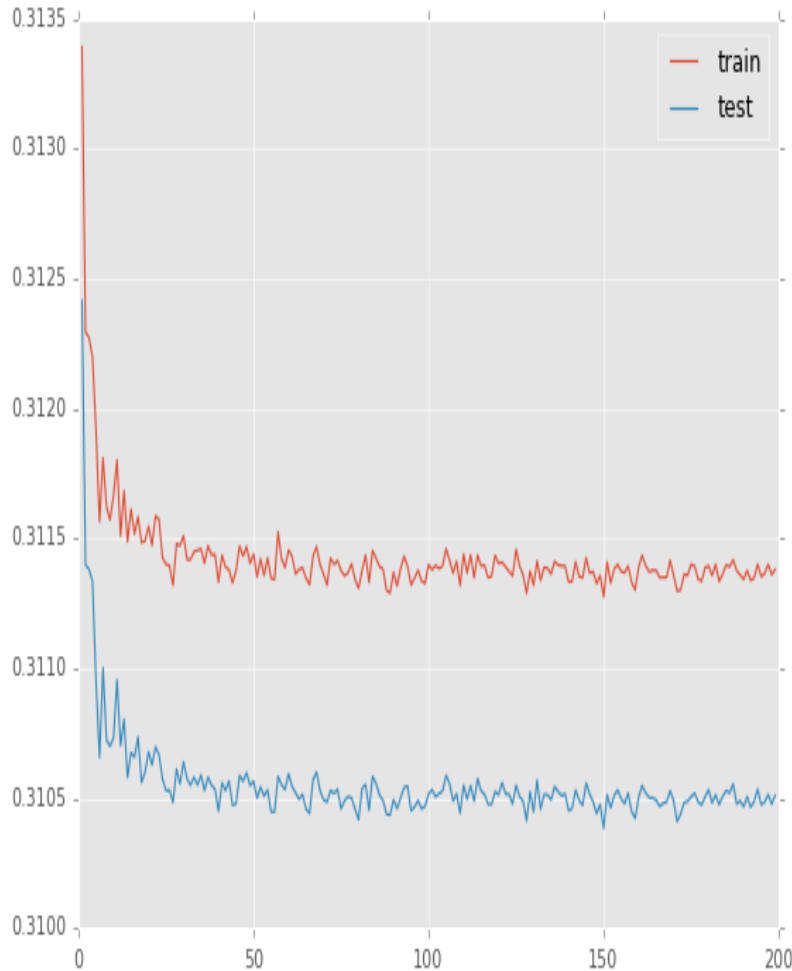
- $RT_{\downarrow i} = \begin{cases} PM_{\downarrow i} & \text{si rachat total} \\ 0 & \text{avec proba } p_{\downarrow i} \\ \text{sinon} \\ \text{avec proba } 1-p_{\downarrow i} \end{cases}$
- $E(Y_{\downarrow i}) = p_{\downarrow i} * PM_{\downarrow i} + (1-p_{\downarrow i}) * 0 = p_{\downarrow i} * PM_{\downarrow i}$
- Espérance du montant racheté $E(Y) = \sum_{\downarrow i} E(Y_{\downarrow i}) = \sum_{\downarrow i} p_{\downarrow i} * PM_{\downarrow i}$

Ainsi, nous estimons le taux de rachat moyen par :

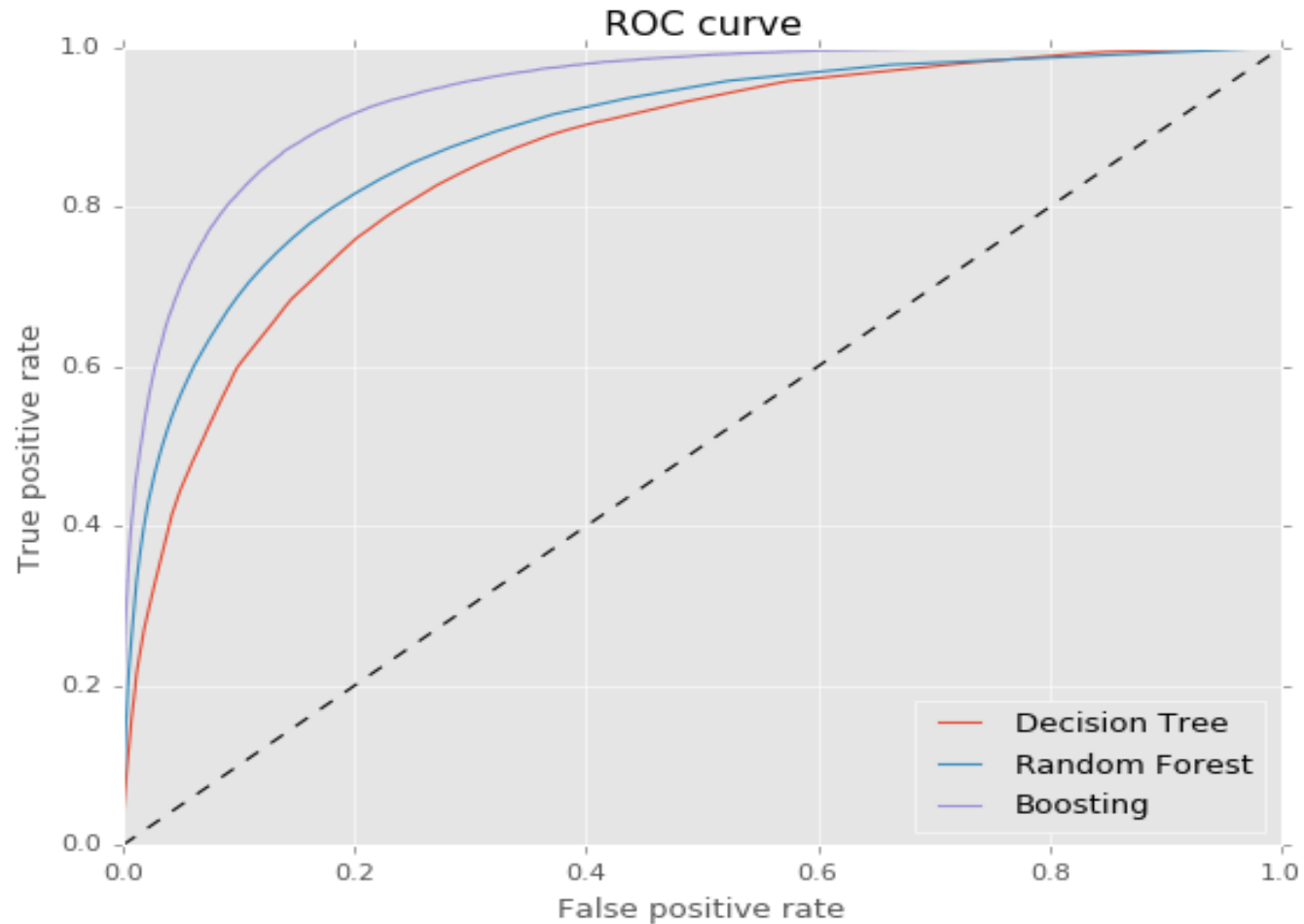
$$Tx_{\downarrow} RactTot = \frac{E(Y)}{\sum_{\downarrow i} PM_{\downarrow i}} = \frac{\sum_{\downarrow i} p_{\downarrow i} * PM_{\downarrow i}}{\sum_{\downarrow i} PM_{\downarrow i}}$$



Vers des comportements dynamiques : forêts aléatoires



Vers des comportements dynamiques : Comparaison des modèles



Nous retenons que ...

- Les méthodes d'agrégations s'avèrent très performantes pour prédire les comportements de rachat
- Les facteurs structurels sont plus déterminants que les facteurs dynamiques
- Principales variables déterminantes : la richesse, la fiscalité, l'appétence pour les marchés financiers, l'âge, le différentiel entre taux servi et benchmark, la santé de l'environnement économique et financier, le niveau de prix de l'immobilier, ...
- Ces méthodes permettent de dégager des profils-type risqués:
 - ✓ Les moins riches ou à niveau de vie moyen
 - ✓ Les contrats dont l'ancienneté atteint la possibilité de défiscalisation
 - ✓ Les personnes très jeunes
 - ✓ Les personnes détenant un fort taux d'UC

Mais...

- Non prise en compte de la censure de nos données car nous avons négligé dans notre étude l'effet des décès
- La focalisation sur le rachat total : en perspective, étendre l'étude au cas des rachats partiels car les comportements de rachat total et de rachat partiel ne sauraient être identiques.
- La non implémentation d'autres modèles plus complexes qui pourrait améliorer la prédiction, mais aussi le caractère boîte noire de ses techniques ne permettant pas de mesurer la significativité des résultats obtenus
- En perspective, une utilisation du Big data permettrait d'affiner les prédictions

***MERCI POUR VOTRE
AIMABLE ATTENTION***