

Construction du taux de rachat structurel en Épargne : Approximation non linéaire et agrégation de modèles

Présentation du mémoire d'Actuaire

Salma Jamal

23 juin 2016

Sommaire

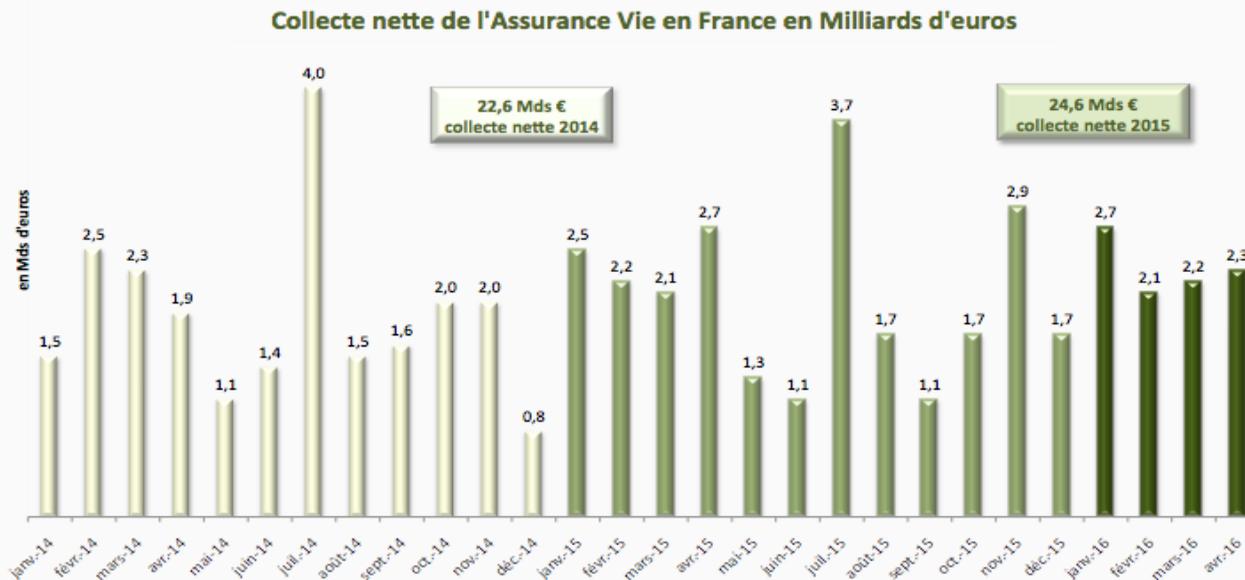
1. Cadre du mémoire
 - Sujet et objectif
 - Présentation de la démarche
2. Périmètre et préparation des données
 - Périmètre
 - Imputation des données manquantes
 - Définition du taux de Rachat
3. Description des données
 - Analyse Factorielle des Données Mixtes
4. Modélisation du taux de Rachat
 - Arbres de régression
 - Réseaux de neurones
 - Agrégation de modèles
5. Validation des modèles
6. Conclusion

Cadre du mémoire

Sujet et objectif (1/2)

Un environnement rassurant

L'Assurance Vie, réputée niche d'investissement préférée des Français, réalise en l'espace du mois d'Avril 2016 une collecte nette de 2,3 Milliards d'euros.



Evolution de la collecte nette en Assurance Vie - Source :
Conférence de presse de
l'Association Française de
l'Assurance 28/01/2016
et
Etudes mensuelles de la FFSA

Le rachat

Un contrat d'Assurance Vie accorde à son détenteur, s'il le souhaite, la possibilité de récupérer tout ou partie du montant de son épargne diminué d'éventuels frais et augmenté d'un éventuel taux garanti.

Cette faculté, exerçable à tout moment de la vie dudit contrat, porte le nom de **rachat**.

Sujet et objectif (2/2)

En Europe, au vu des enjeux financiers que véhicule ce phénomène, le rachat est identifié comme étant le risque résiduel le plus important parmi tout autre risque de souscription.

Quid de sa gestion sous Solvabilité 2 ?

La justesse prend le pas sur la prudence : les provisions constituées par chaque assureur doivent refléter son propre risque, en particulier dans le cas du rachat.

Objectif du mémoire :

Construire une loi de rachat Best Estimate répondant à cette exigence en :

- définissant des groupes homogènes de risque ;
- identifiant la contribution de facteurs de rachat dans la construction des taux de rachat structurels ;
- projetant les modèles construits pour en éprouver la robustesse.

Présentation de la démarche

1. Périmètre et préparation des données ;

- Présentation du périmètre
- Imputation des données manquantes
- Définition du taux de rachat

2. Description des données :

- Analyse Factorielle des Données Mixtes

3. Modélisation du taux de rachat :

- Arbres de régression
- Réseaux de neurones
- Agrégation de modèles
 - ▶ Boosting
 - ▶ Bagging
 - ▶ Forêts aléatoires

4. Validation de modèles.

Périmètre et préparation des données

Périmètre

Périmètre

Ce mémoire a été réalisé sur la base de contrats d'Épargne individuels et a été mené en 2 temps :

- **Description** : commune à l'ensemble du portefeuille.
- **Modélisation du taux de rachat** :
 - ▶ **Construction des modèles** Contrats sujets à rachat entre 2005 et 2010 ;
 - ▶ **Validation des modèles** Contrats sujets à rachat en 2013.

Variables exploitées par assuré/contrat :

- Âge de l'assuré
- Sexe
- Segment de clientèle (donnée commerciale)
- Gamme du produit
- Ancienneté du contrat
- Catégorie fiscale
- Nature du contrat (mono-support € ou multi-supports)
- Taux Minimum Garanti (TMG) servi au contrat
- Montants et types du rachat
- Provisions mathématiques (en ouverture et clôture de période)

Imputation des données manquantes (1/3)

Méthodes d'imputation de données manquantes hétérogènes :

- **LOCF (Last Observation Carried Forward)** Chaque valeur manquante est remplacée par la dernière donnée connue de la variable.
- **kNN (Méthode des plus proches voisins)** À chaque donnée manquante est affectée la moyenne des valeurs de ses k plus proches voisins.
- **MICE (Multiple Imputation by Chained Equations) [7]** Chaque variable manquante est modélisée conditionnellement aux données renseignées.
- **Forêts aléatoires [2]** : 1^{ère} imputation rudimentaire avant la construction d'une série de forêts aléatoires ajustées progressivement jusqu'à atteinte d'un critère d'arrêt.

Imputation des données manquantes (2/3)

Étapes de l'imputation :

1. Suppression des lignes à données manquantes : les bases obtenues sont complètes ;
2. Génération aléatoire et progressive de données manquantes (création de 7 bases en augmentant la proportion de données manquantes par paliers de 10%).
3. Imputation des données manquantes ;
4. Estimation de l'erreur d'imputation entre données réelles et données imputées :

- ▶ Entre variables qualitatives : la distance de Hamming

$$d_H(\hat{T}_i, T) = \frac{\sum_{j=1}^p \mathbb{1}_{\{\hat{T}_{ij} \neq T_{ij}\}}}{\frac{\text{Card } T_i}{100}}$$

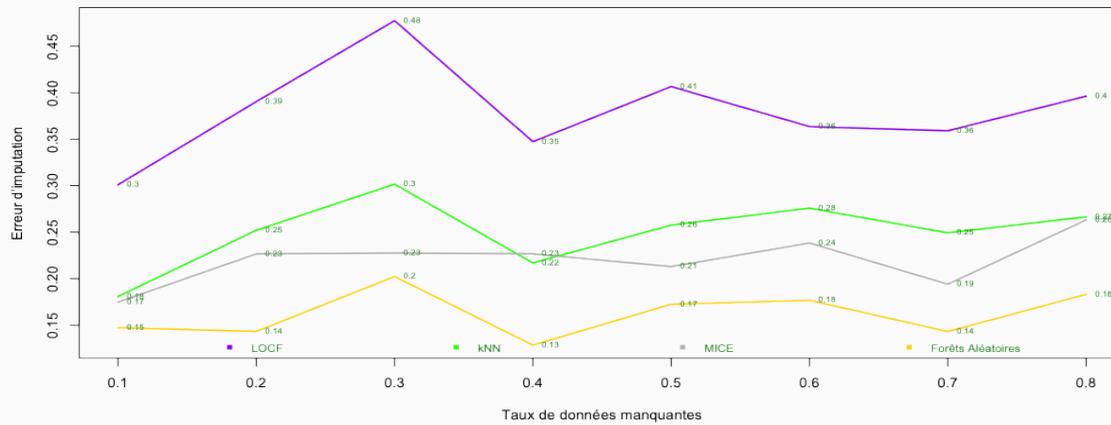
où T représente la base de données $n \times p$, T_{ij} les modalités initiales (avant génération de données manquantes) et \hat{T}_{ij} les modalités imputées.

- ▶ Entre variables quantitatives : la valeur absolue de la différence entre données initiales et données imputées.

Imputation des données manquantes (3/3)

Exemple : Âge des assurés et Gamme de produits :

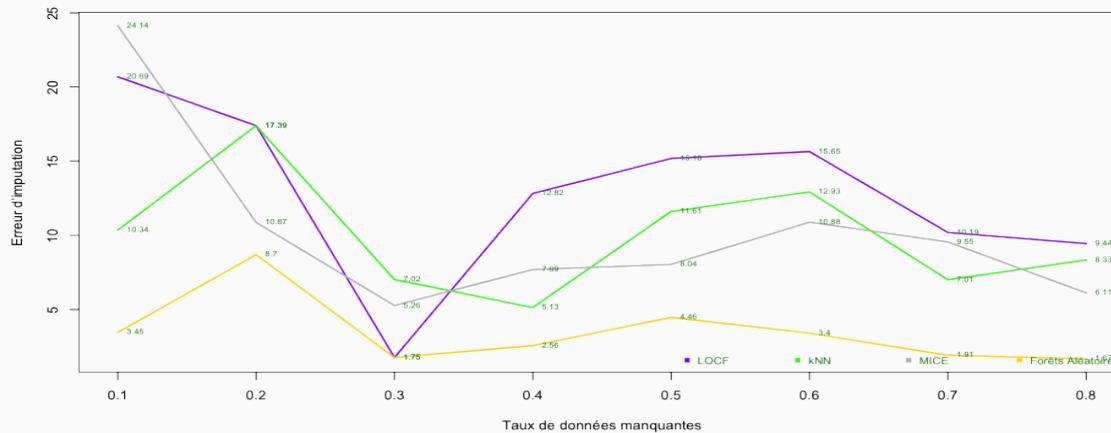
Erreur d'imputation en pourcentage des données manquantes – Âge des assurés



Courbes :



Erreur d'imputation en pourcentage des données manquantes – Gamme de produit



La méthode d'imputation par **forêts aléatoires** minimise le taux d'erreur d'imputation.

Définition du taux de Rachat

Le taux de rachat, en particulier celui du rachat partiel, dépend des montants de versements mais ceux-ci ne sont pas renseignés dans les bases étudiées.

Construction du taux de rachat :

- Assiette du rachat partiel :

- ▶ Si Montant du rachat $<$ (PM d'ouverture + PM de clôture) :

$$\text{Assiette RP} = \frac{\text{PM d'ouverture} + (\text{PM de clôture} + \text{Montant du rachat})}{2}$$

- ▶ Sinon : Assiette RP = PM de clôture + Montant du Rachat

- Taux de rachat :

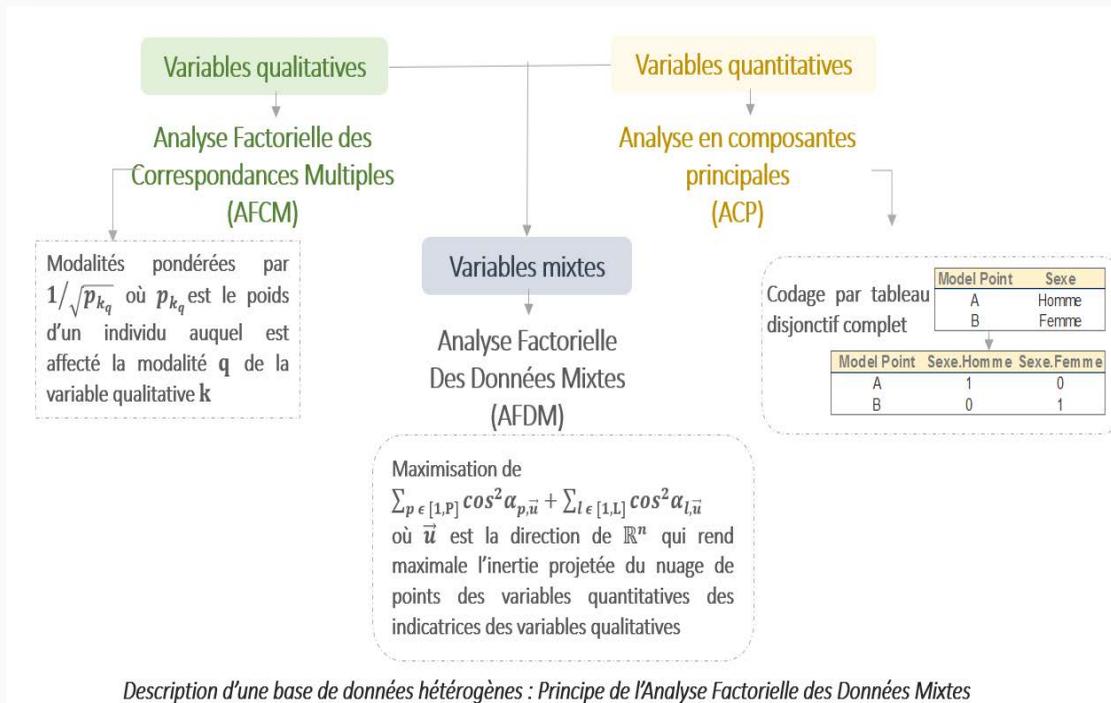
$$\text{Taux de rachat} = \begin{cases} \frac{\text{Montant du rachat}}{\text{Assiette RP}} & \text{Si le rachat est partiel (libre ou programmé)} \\ 1 & \text{Si le rachat est total} \end{cases}$$

Description des données

Analyse Factorielle des Données Mixtes (1/4)

Principe de l'Analyse Factorielle des Données Mixtes (AFDM) :

Quel que soit le nombre de modalités qu'elle présente, une variable qualitative, dans le cadre d'une AFDM [6], aura toujours la même inertie qu'une variable quantitative. Il n'y a donc pas de déséquilibre de représentation des types de variables.



L'AFCM impose des contraintes sur le nombre d'observations et sur le découpage en classes des variables quantitatives. L'ACP fait appel à un tableau disjonctif complet, autrement dit à une bijection transformant les variables qualitatives en variables quantitatives.

Ce qui manque au codage par tableau disjonctif complet, c'est une pondération des individus présentant une modalité.

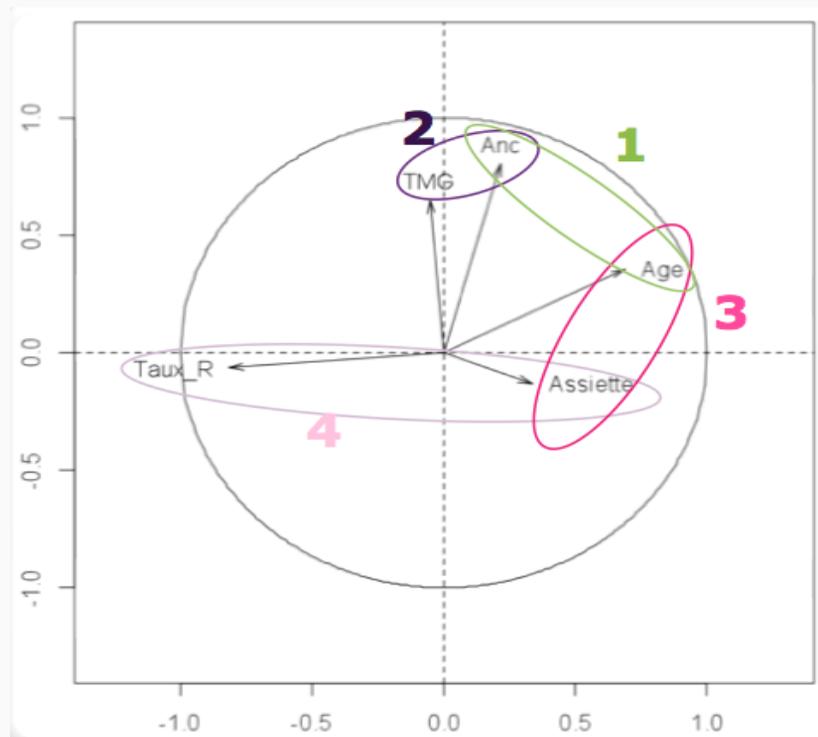
Cette pondération doit, en tout état de cause, faire intervenir les fréquences des modalités, compensant ainsi l'éventuelle non-réduction des données.

Le fait de s'appuyer sur des fréquences n'est pas anodin : c'est l'AFCM qui l'inspire. Une fois le codage adopté dans ce sens, opter pour une ACP est tout à fait convenable. C'est là le principal apport de l'AFDM qui combine bien des principes de l'ACP et de l'AFCM.

Analyse Factorielle des Données Mixtes (2/4)

Application aux données 2005 - 2010 :

Répartition des variables quantitatives

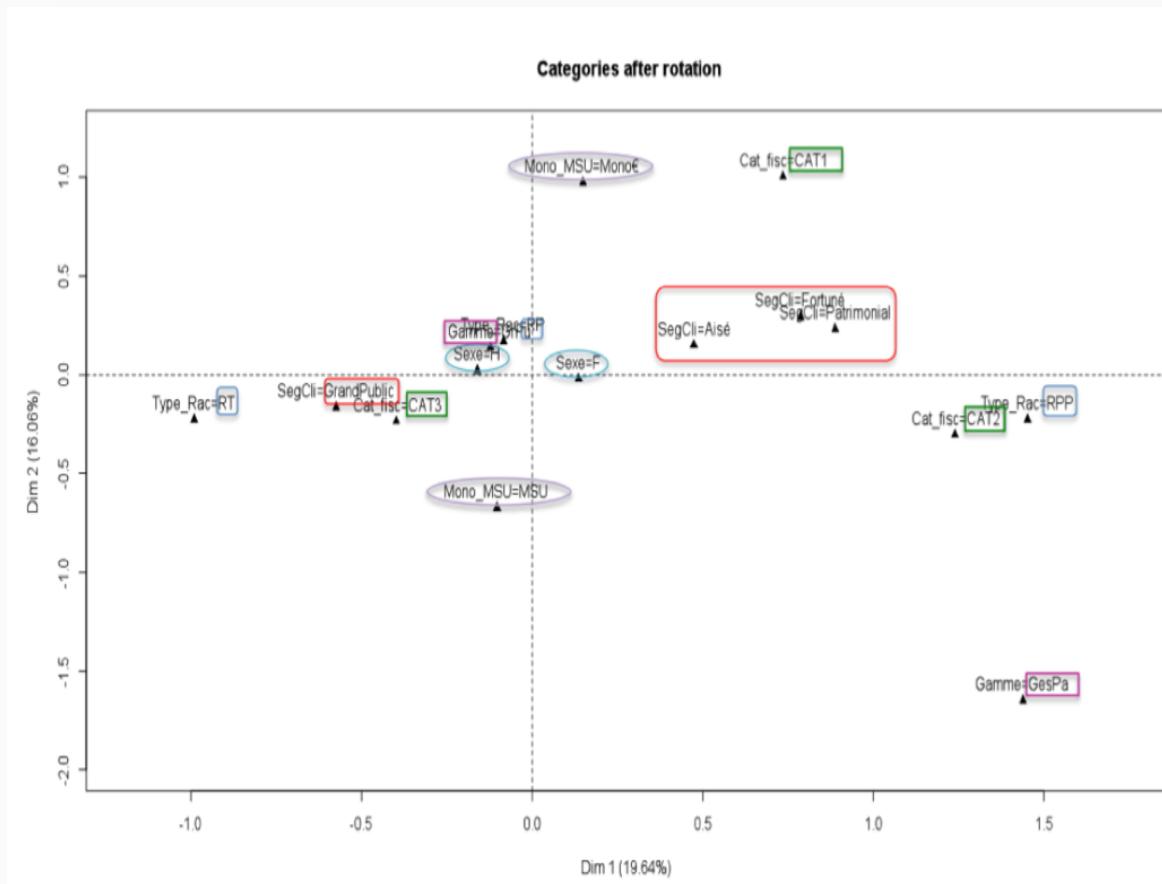


- L'ancienneté du contrat augmente avec l'âge des assurés ;
- Le rachat s'effectue sur un contrat d'autant plus tard que le TMG servi par le contrat est élevé ;
- Les assurés âgés sont ceux dont l'assiette du rachat est la plus importante ;
- Plus l'assiette du rachat est importante, plus le taux de rachat est faible : lorsque l'assiette augmente, le montant de rachat diminue et, conséquemment, le taux de rachat diminue.

Analyse Factorielle des Données Mixtes (3/4)

Application aux données 2005 - 2010 :

Répartition des variables qualitatives

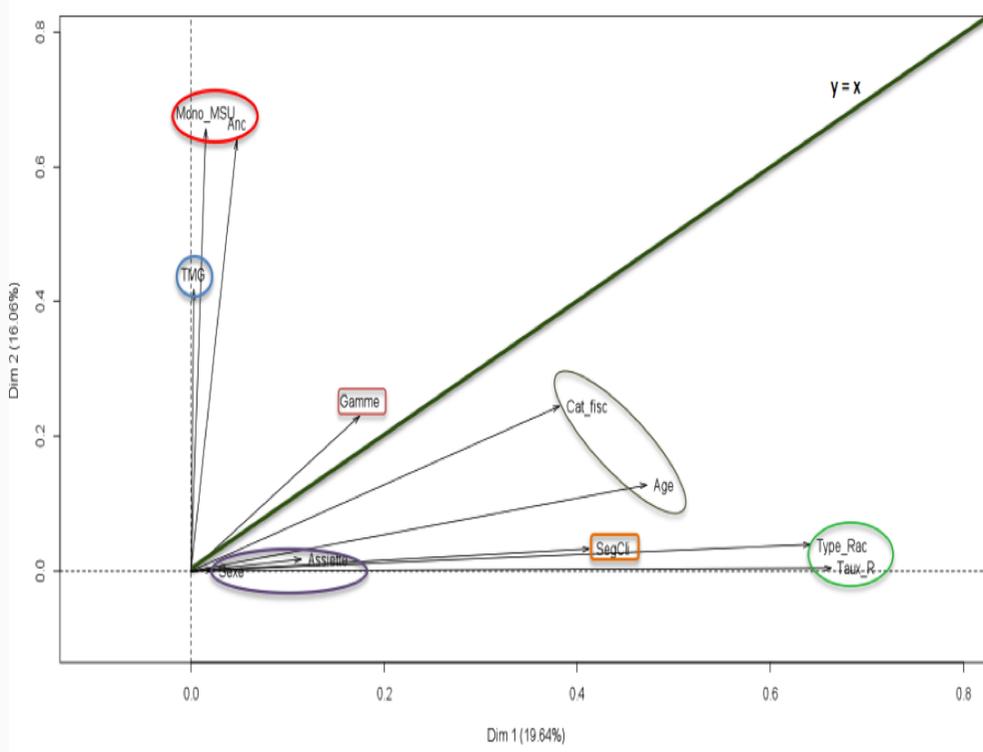


- Les segments de clientèle Aisé, Patrimonial et Fortuné sont très proches et situés à l'opposé de l'axe principal par rapport au segment Grand Public. Le même éloignement est observé entre les modalités de gammes de produits (Grand Public et Gestion Patrimoniale) ;
- Les rachats totaux sont plus remarquables auprès des assurés âgés de moins de 70 ans et ayant souscrit avant l'âge de 70 ans (3^{ème} catégorie fiscale). Ces rachats concernent en particulier les contrats Multi-supports ;
- Les rachats partiels s'effectuent majoritairement sur des contrats concernant des produits Grand Public ;
- Les modalités homme et femme sont proches et, qui plus est, sont proches du centre d'inertie. Le sexe n'est pas une variable discriminante.

Analyse Factorielle des Données Mixtes (4/4)

Application aux données 2005 - 2010 :

Analyse croisée des variables



- Le type de rachat, le taux de rachat et le segment de clientèle contribuent de façon significative à la constitution de l'axe principal. Ces variables évoluent dans le même sens. Le taux de rachat, nous le constatons plus tôt, dépend de la situation financière de l'assuré. Nous le traduisons comme un axe de **Richesse** ;
- Le type de support du contrat racheté est associé à l'ancienneté du contrat. Par ailleurs, les TMG sont servis sur les contrats Mono-support et sur le fond général des Multi-supports. Nous interprétons donc l'axe secondaire comme un axe du **Temps** ;
- Les variables sexe et assiette du rachat, proches du centre d'inertie, ne sont pas contributives dans l'inertie globale ;
- La présence de variables autour de la première bissectrice explique l'équi-contribution dans l'explication de chaque axe. C'est le cas de la gamme de produits.

Modélisation du taux de Rachat

Arbres de régression (1/4)

Étapes de construction d'un arbre de régression CART [4]

- **Échantillonnage :**

Échantillons	Proportion de la base initiale
Apprentissage	80% des rachats \in $[[2005; 2010]]$
Test	20% des rachats \in $[[2005; 2010]]$
Validation	Base récente : rachats 2013

- **Règles de segmentation :** choix successifs des variables contributives dans l'explication de la variable d'intérêt. L'arbre réduit l'impureté (au regard de la distribution uniforme des classes au niveau de chaque nœud) en se déplaçant d'un nœud père vers ses nœuds fils.

$$I_{Gini} = \sum_{i \neq j} P(i|t)P(j|t)$$

où $P(i|t)$ désigne la proportion d'éléments de la classe i affectés au nœud t .

- **Affectation des individus** consiste à réduire le coût de mauvaise affectation associé à chaque nœud en réduisant celui de chaque individu affecté à ce nœud.

$$I_{Gini} = \sum_{i \neq j} C_i(j)P(i|t)P(j|t)$$

où $C_i(j)$ le coût du mauvais classement d'une classe i dans la classe j .

- **Élagage de l'arbre** Suppression des feuilles lorsqu'un critère d'arrêt est atteint. Ici : règle de l'écart-type de Breiman et al. (1984) qui consiste à réduire le facteur de complexité (cp) de l'arbre final en s'arrêtant au niveau comportant le moins de branches terminales.

Arbres de régression (2/4)

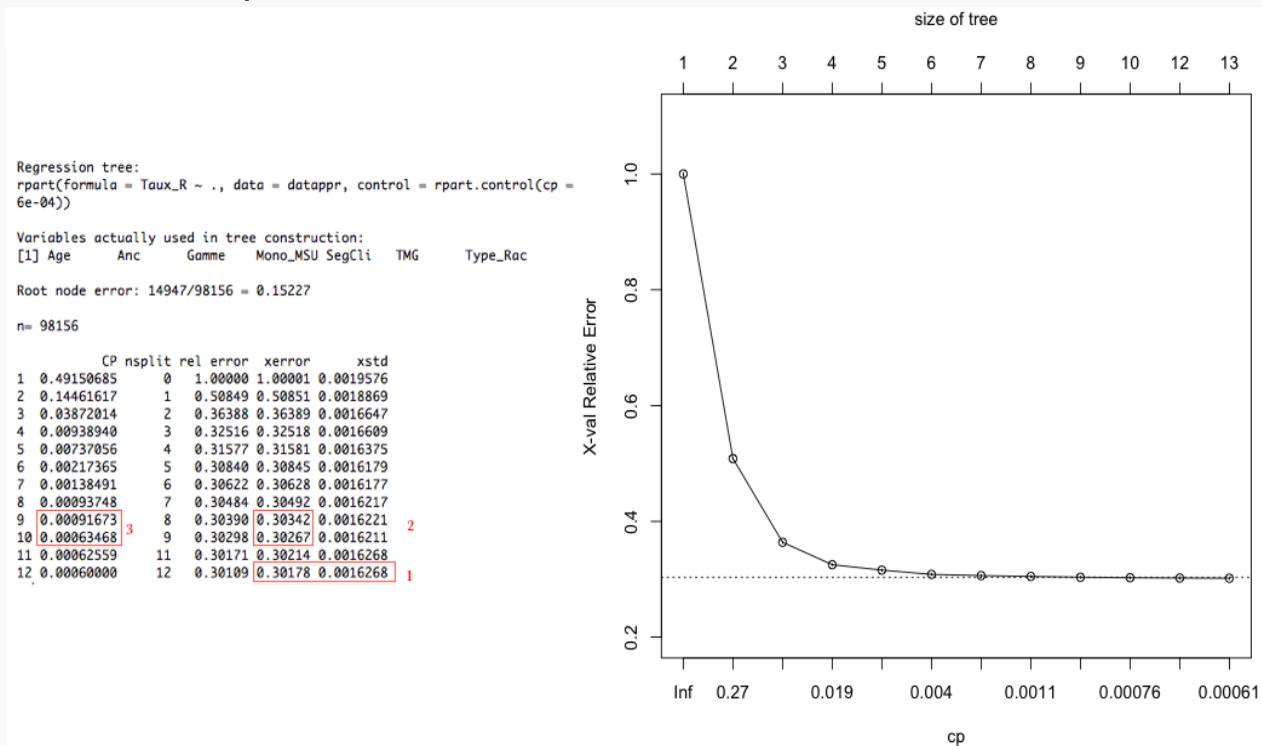
Coefficient de pénalisation

L'erreur augmente lorsque le cp augmente et donc lorsque le nombre de nœuds diminue.

La règle de Breiman recommande un seuil d'erreur inférieur à $0,30178 + 0,0016268 = 0,3034068$

Autrement dit, un cp compris, strictement, entre 0,00063468 et 0,00091673.

- Nous choisissons un cp de 0,0007



Evolution du cp en fonction de l'erreur relative

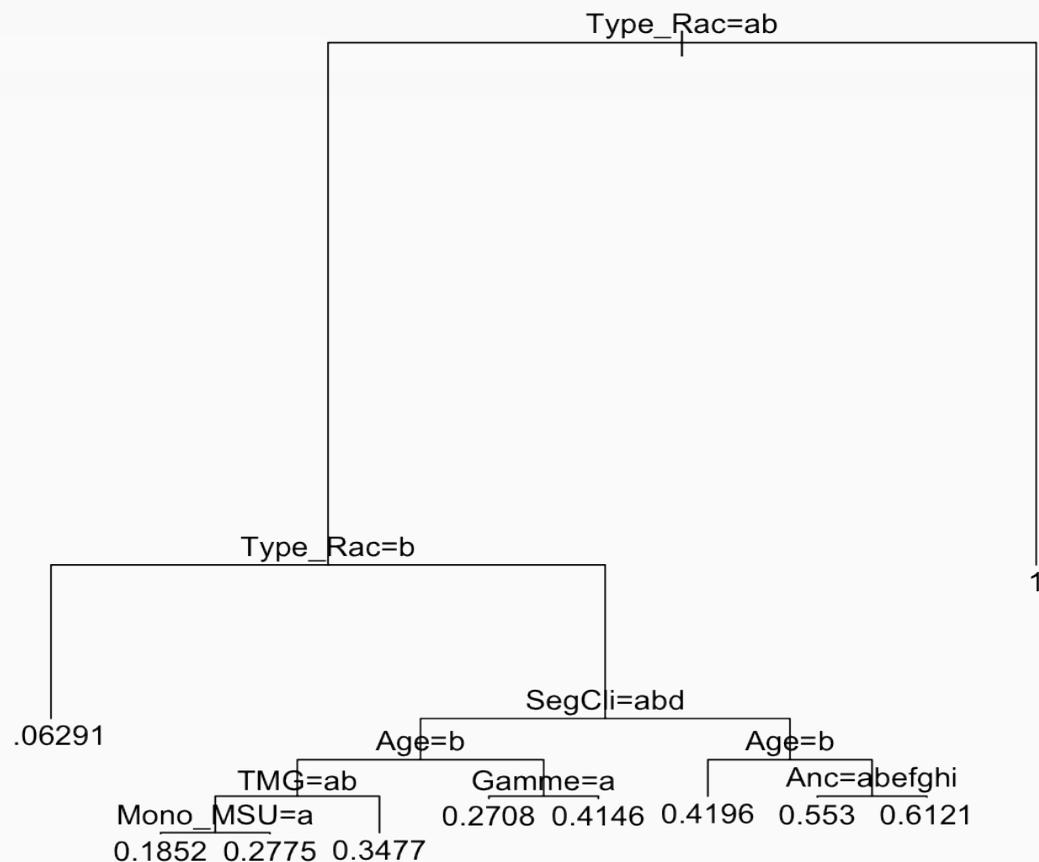
Arbres de régression (3/4)

Élagage de l'arbre

L'arbre retenu est constitué de 9 segmentations, il comporte donc 10 feuilles contre 13 pour l'arbre initial.

Les quantités indiquées au niveau des feuilles correspondent à la valeur prédite moyenne du taux de rachat pour les populations de chaque feuille.

Il faut, pour avoir une vision plus réaliste de ces taux, se ramener au niveau du portefeuille global (les résultats numériques finaux sont présentés dans le mémoire [3]).



Arbre de régression élagué

Arbres de régression (4/4)

Interprétation de l'arbre élagué

Classement des variables par pouvoir décroissant de discrimination :

1. **Type de rachat.** À ce niveau de l'arbre, seuls les rachats partiels semblent candidats à une segmentation plus fine ;
2. **Segment de clientèle.** À l'instar de l'analyse factorielle, l'arbre de régression détecte des comportements différents entre les segments Aisé, Fortuné et Patrimonial (*abd* au niveau du 10^{ème} nœud) et le segment Grand Public :
 - ▶ Pour les premiers, l'**âge** oriente la décision de rachat : les assurés de plus de 70 ans rachètent selon la nature de leur contrat (**Mono/Multi-supports**) et du **TMG** servi, les plus jeunes se concentrent sur la **gamme de produit** de leur contrat ;
 - ▶ Le segment "**Grand Public**" est influencé par l'**âge**. De plus, les anciennetés d'1 et 2 ans sont mis en avant. Les intermédiaires commerciaux sont objectivés par le nombre d'affaires nouvelles sans être pénalisés par les sorties.

Erreur quadratique moyenne de prévision

Après application du modèle à l'échantillon de test, nous obtenons une erreur quadratique moyenne de prévision de **10,4%**.

Le modèle est encore jugé insatisfaisant : la consolidation se fera lors de l'agrégation de modèles.

Réseaux de neurones (1/4)

Composantes d'un réseau de neurones

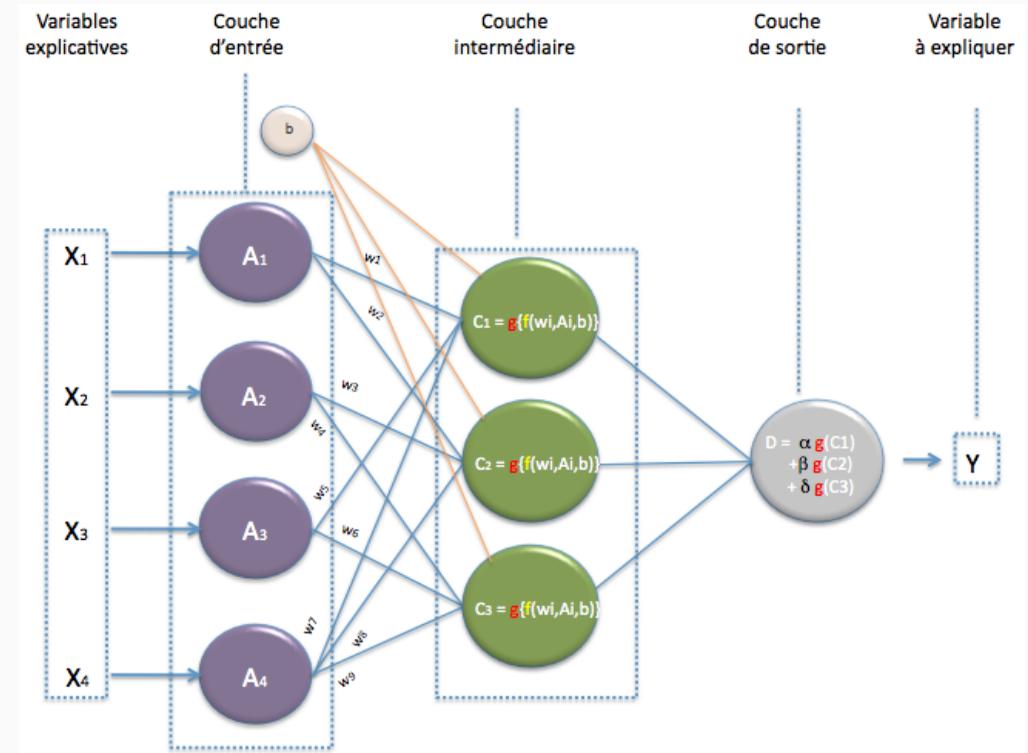
Un réseau de neurones est composé d'une **couche d'entrée**, d'une **couche de sortie** et d'une ou plusieurs **couches intermédiaires/cachées**. Chaque couche est composée de plusieurs **neurones**.

Chaque nœud de la couche d'entrée est attaché à des **poids synaptiques**. Les poids traduisent la probabilité de transmission de l'information entre les neurones.

Les poids sont liés aux nœuds suivants à l'aide d'une **fonction de combinaison** (notée f sur la figure). À cette combinaison est ensuite appliquée une **fonction d'activation** (notée g sur la figure) qui va estimer l'activité du nœud.

Des **biais** (ou *seuils d'activation des neurones*, (noté b sur la figure)) peuvent également exister au niveau de chaque nœud. Un biais est un neurone dont la valeur, soumise à la fonction d'activation, vaut toujours 1.

L'ensemble des biais, des données en entrée transformées par les fonctions de combinaison et d'activation, est appelé **noyau**.



Réseau de neurones élémentaire

Réseaux de neurones (2/4)

Réseau de neurones retenu

Nous nous plaçons, dans cette étude dans le cadre du **perceptron multi-couches à rétropropagation du gradient de l'erreur** [10], noté par la suite **PMC-RGE**.

Étapes de construction d'un PMC-RGE

- **Prétraitement des données :**
 - ▶ Application de la fonction logarithme aux variables continues avant normalisation ;
 - ▶ Binarisation des variables qualitatives ;
 - ▶ Transformation de la $i^{\text{ème}}$ valeur de chaque variable discrète par
$$\sum_{j=0}^i 2^{-j}.$$
- **Initialisation aléatoire des poids**
- **Mise en place de l'algorithme :**
 - ▶ Application des fonctions de combinaison et d'activation à la couche d'entrée et aux couches intermédiaires ;
 - ▶ Évaluation de la fonction de coût globale du modèle ;
 - ▶ Rétropropagation de l'erreur pour le calcul des erreurs individuelles du modèle ;
 - ▶ Ajustement des poids des neurones en partant des neurones terminaux vers les neurones intermédiaires ;
 - ▶ Itération jusqu'à atteinte d'un critère d'arrêt (seuil pour les dérivées partielles de la fonction d'erreur) ou un nombre défini d'itérations ;
 - ▶ Application du modèle retenu à l'échantillon test pour en vérifier la généralisation.

Réseaux de neurones (3/4)

Sensibilité d'un PMC-RGE

- Sensibilité du PMC-RGE au nombre de couches intermédiaires

	Nombre de couches intermédiaires*		
	1	2	3
Erreur d'apprentissage	43,386	43,335	43,213
Nombre d'itérations	14 109	43 422	23 594
Erreur de prévision	0,36	0,14	0,09

* à 6 neurones cachés

Ce test met en avant le risque de surapprentissage lié à l'augmentation du nombre de couches intermédiaires. Plus ce nombre augmente, plus le réseau a tendance à mémoriser la structure des données qui lui sont présentées : l'erreur d'apprentissage baisse puisque le réseau assimile de mieux en mieux les données.

- Sensibilité du PMC-RGE au nombre de neurones cachés

	Nombre de neurones cachés**		
	2	4	6
Erreur d'apprentissage	43,925	43,732	43,213
Nombre d'itérations	25 926	8 287	14 109
Erreur de prévision	0,34	0,37	0,42

** à 1 couche cachée

L'erreur d'apprentissage baisse lorsque le nombre de neurones augmente. L'erreur de prévision suit la tendance inverse.

- Sensibilité du PMC-RGE à la fonction d'activation

	Fonction d'activation	
	Linéaire	Sigmoïde***
Erreur d'apprentissage	43,732	43,702
Nombre d'itérations	25 926	4 536
Erreur de prévision	0,37	0,12

*** Tangente hyperbolique

La fonction sigmoïde assure des sorties non linéaires. Une telle fonction fait baisser l'erreur d'apprentissage mais fait surtout baisser le nombre d'itérations avant l'atteinte d'un modèle optimal : le gain de temps en apprentissage est 5 fois meilleur que dans le cas d'une fonction linéaire.

- Sensibilité du PMC-RGE au taux d'apprentissage

	Taux d'apprentissage		
	0,9	0,5	0,1
Erreur d'apprentissage	43,529	43,532	43,534
Nombre d'itérations	40 653	65 339	98 054
Erreur de prévision	0,66	0,40	0,02

Le taux d'apprentissage représente la vitesse à laquelle le réseau parcourt les données. S'il est très élevé, le réseau peut ne détecter aucun optimum. S'il est très petit, il y a risque de surapprentissage.

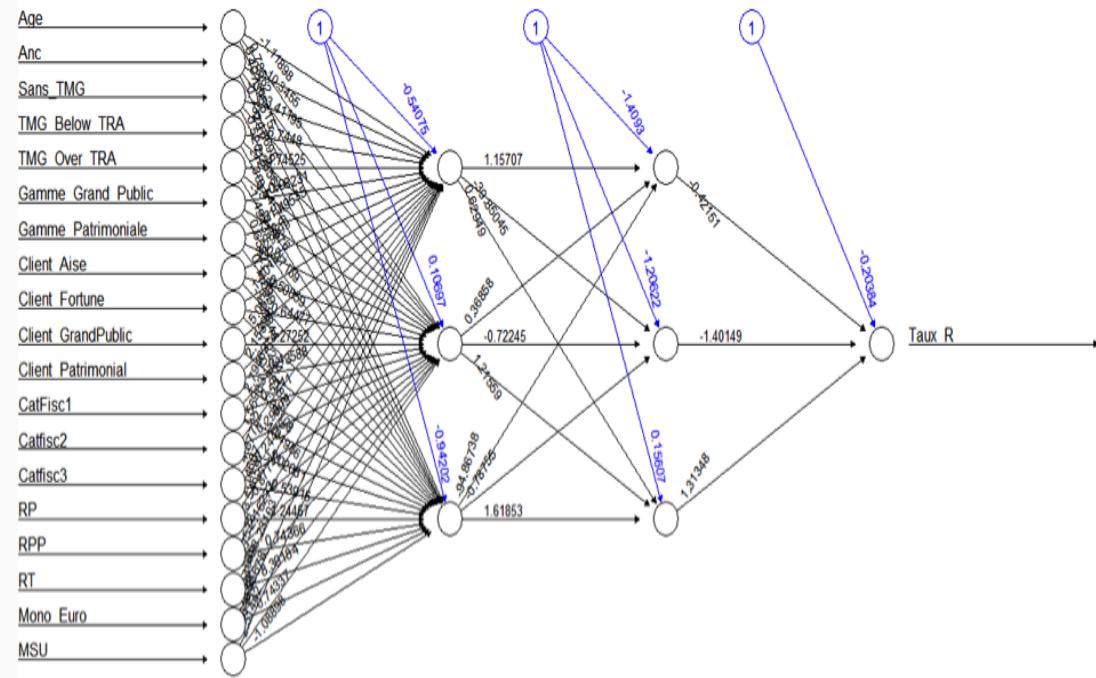
Réseaux de neurones (4/4)

PMC-RGE retenu

Au regard des résultats obtenus, nous construisons un modèle avec 2 couches intermédiaires à 6 neurones cachés (3 par couche). La fonction sigmoïde (tangente hyperbolique) est celle utilisée et le taux d'apprentissage est fixé à 0,25.

Erreur d'apprentissage	43,040
Nombre d'itérations	101 582
Erreur de prévision	0,04

Ajustement final d'un PMC-RGE {3 3}



PMC-RGE retenu, à 2 couches intermédiaires, à 6 neurones cachés

Le modèle montre un taux d'erreur intéressant bien que le temps d'apprentissage soit un peu lourd du fait du taux d'apprentissage peu élevé. Le modèle n'en demeure pas moins sensible à des minima locaux et devra être agrégé.

Agrégation de modèles (1/10)

Quelques méthodes d'agrégation

La qualité d'un modèle statistique est liée à sa précision et à sa robustesse : augmenter la première consiste à diminuer le biais du modèle, la seconde, à baisser sa variance.

L'agrégation de modèles permet de renforcer ces aspects, en particulier dans le cas de modèles sensibles tels que les arbres de régression et les réseaux de neurones [1].

Les méthodes d'agrégation de modèles les plus populaires sont :

Bagging [4] trouve le modèle "moyen" approchant le mieux les données d'intérêt. À partir d'échantillons bootstraps, l'*erreur-out-of-bag* mesure la stabilité du modèle en fonction du nombre d'échantillons bootstrap.

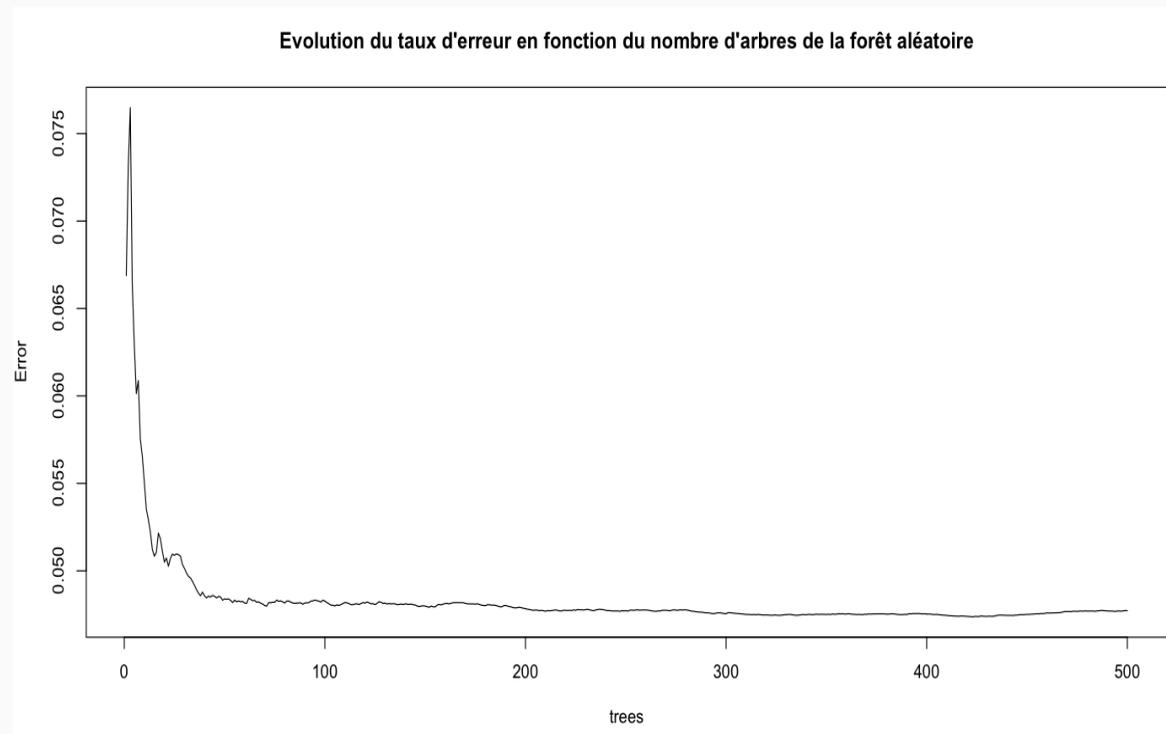
Forêt aléatoire : Une forêt aléatoire est une agrégation d'arbres de décisions. Elle comporte une étape supplémentaire de tirage aléatoire, par rapport au Bagging, et est intéressante de par la vitesse de sa convergence vers une solution optimale.

Boosting [9] [8] s'adresse (hormis dans le cas du modèle *Arcing*) à toute la base de données plutôt qu'à des échantillons bootstrap. Cet algorithme est un processus adaptatif et déterministe (le plus souvent) : à chaque itération, le modèle est une version adaptée de la version liée à l'itération précédente.

Agrégation de modèles (2/10)

Forêts aléatoires

Les forêts aléatoires dépendent principalement de 2 paramètres : le nombre d'arbres et le nombre de variables à chaque nœud des arbres qui les constituent.



Evolution du taux d'erreur en fonction du nombre d'arbres de la forêt aléatoire

Nombre d'arbres	MSE*	% variance expliquée
500	4,77%	68,51%
1000	4,84%	68,05%
1500	4,42%	68,18%

* MSE : Moyenne de la somme des carrés des erreurs

Les quantités observées n'évoluent pas de façon significative. Cette stabilité commence en réalité avant le seuil des 500 arbres.

Agrégation de modèles (3/10)

Forêts aléatoires

La variation du nombre de variables à chaque nœud ne montre pas d'évolution particulière dans la variation du MSE et du pourcentage de la variance expliquée. Nous choisissons toutefois de garder 3 variables par nœud dans la mesure où nous observons pour ce nombre de variables les quantités les plus intéressantes pour ces deux indicateurs.

Nombre de variables à chaque nœud	MSE	% variance expliquée
2	4,77%	68,51%
3	4,6%	70,14%
4	4,64%	69,85%

Pour la suite, nous fixons le nombre d'arbres à 500 et le nombre de variables à chaque nœud à 3.

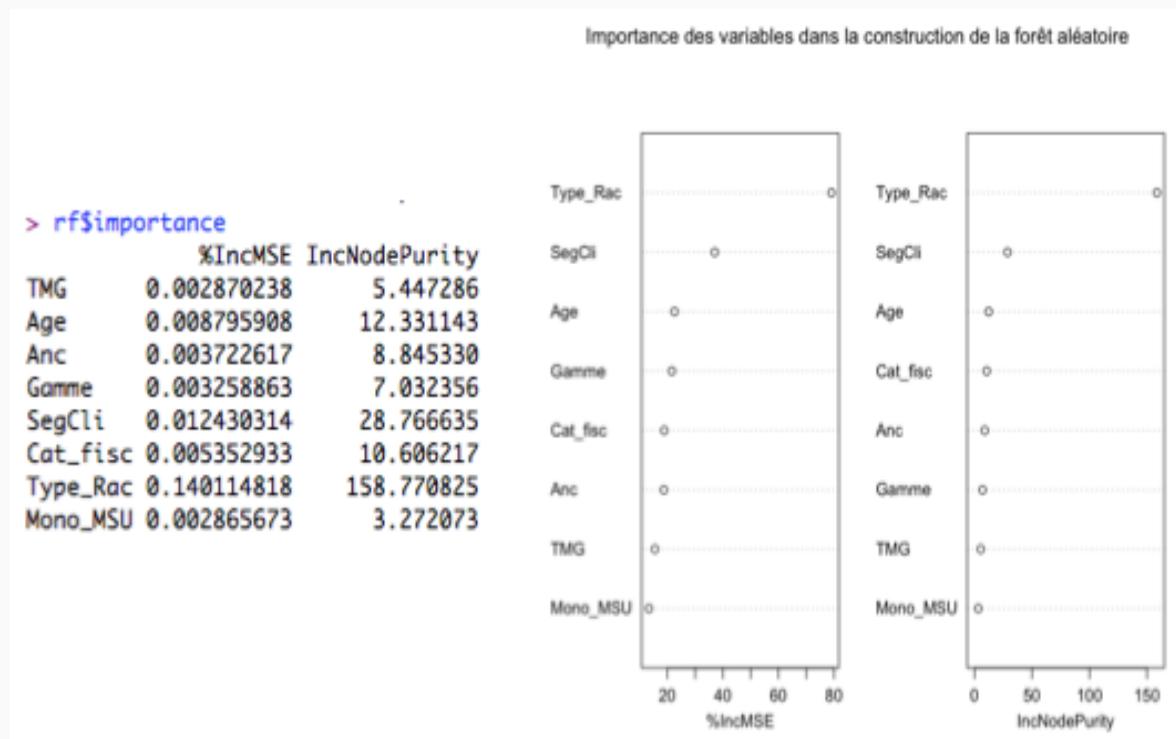
Agrégation de modèles (4/10)

Forêts aléatoires : importance des variables

Nous comparons deux indicateurs : la **décroissance moyenne de la précision** ($\%IncMSE$) et la **décroissance moyenne de la pureté des nœuds** ($IncNodePurity$).

Les deux indicateurs évoluent globalement dans le même sens hormis pour la variable Gamme. L'instabilité de cette variable avait été mise en exergue lors de l'étude des arbres de régression.

Par ailleurs, le type de rachat apparaît de façon prédominante. Le Segment de Clientèle et l'Âge des assurés s'imposent ensuite dans l'ordre des variables explicatives du taux de rachat.



Forêt Aléatoire - Importance des variables

Agrégation de modèles (5/10)

Boosting d'arbres de régression

Le boosting d'arbres de régression est sensible au nombre d'arbres de la forêt, à la profondeur de la forêt (i.e. au nombre de nœuds compris par segmentation, au niveau de chaque arbre de la forêt) et au taux d'apprentissage.

Dans le cas d'une régression, le modèle calcule l'erreur quadratique moyenne (RMSE) et le R^2 à chaque itération.

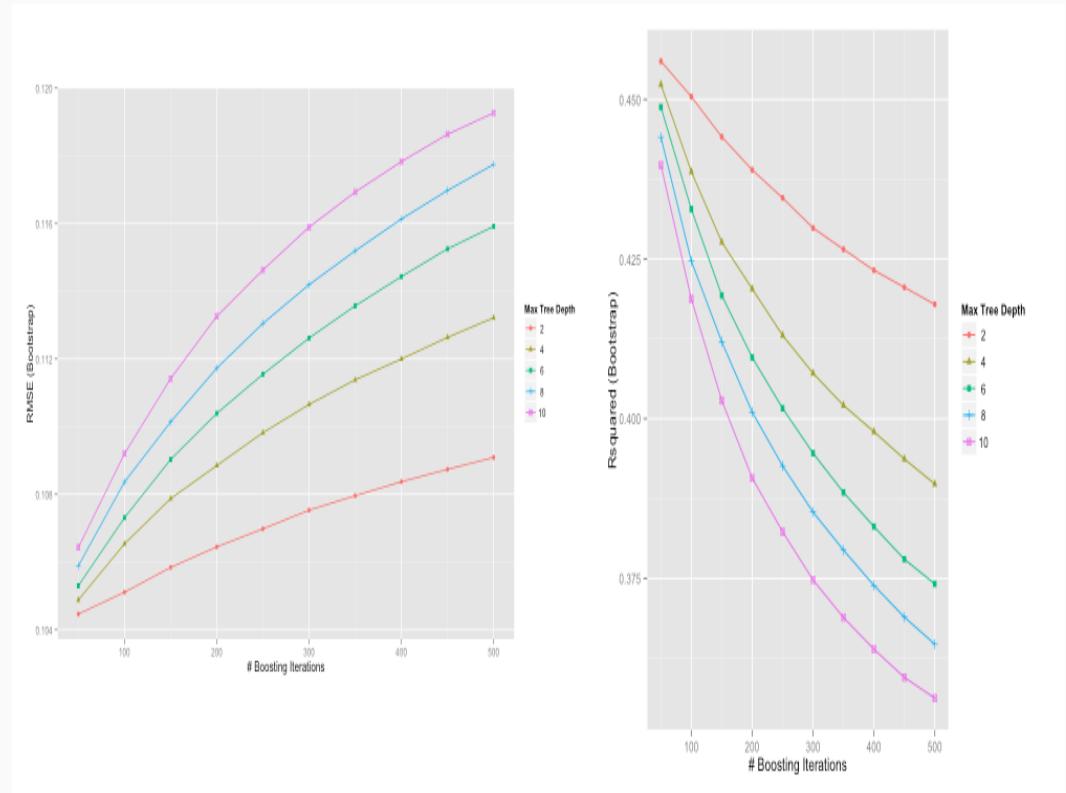
L'algorithme sélectionne enfin le modèle minimisant l'erreur quadratique moyenne.

Agrégation de modèles (6/10)

Boosting d'arbres de régression

- Le nombre d'arbres est initialisé à 50 et incrémenté progressivement de 50 jusqu'à atteindre 500 arbres ;
- 5 niveaux de complexité sont déclinés pour évaluer la sensibilité du modèle à ce paramètre ;
- Le taux d'apprentissage est fixé à 0,001 pour éviter le risque de surapprentissage.

Le RMSE croît avec la profondeur de l'arbre et avec le nombre d'arbres le composant (et inversement pour le R^2).



RMSE et R^2 du boosting d'arbres de régression

Agrégation de modèles (7/10)

Agrégation de réseaux de neurones

L'agrégation des réseaux de neurones nécessite de renseigner :

- Le nombre d'itérations de l'apprentissage ;
- La taille du réseau (i.e. son nombre de couches intermédiaires et de neurones cachés)
- Le **decay** qui est un facteur de contrôle pénalisant les valeurs prises par les poids synaptiques en les contraignant à décroître exponentiellement vers 0.

La réduction du surapprentissage est directement liée à la diminution du nombre de neurones cachés et à l'augmentation du decay.

Nous fixons les valeurs suivantes :

Nombre d'itérations	Nombre de neurones cachés	Decay
200	6	0.9

Agrégation de modèles (8/10)

Agrégation de réseaux de neurones

Les niveaux du RMSE et du R^2 du bagging et du boosting obtenus sont très proches et présentent des valeurs également proches de celles du boosting d'arbres de régression.

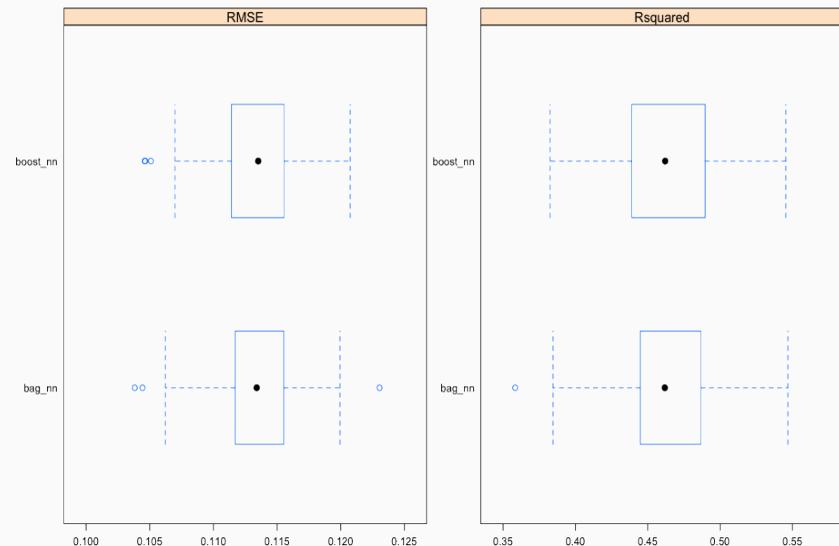
```
> summary(resamps)

Call:
summary.resamples(object = resamps)

Models: bag_nn, boost_nn
Number of resamples: 200

RMSE
      Min. 1st Qu. Median  Mean 3rd Qu.  Max. NA's
bag_nn  0.1038 0.1117 0.1134 0.1134 0.1155 0.1230    0
boost_nn 0.1046 0.1114 0.1135 0.1134 0.1155 0.1207    0

Rsquared
      Min. 1st Qu. Median  Mean 3rd Qu.  Max. NA's
bag_nn  0.3584 0.4449 0.4618 0.4647 0.4866 0.5469    0
boost_nn 0.3825 0.4390 0.4620 0.4647 0.4896 0.5455    0
```



Agrégation de réseaux de neurones - Comparaison Bagging/Boosting - Boxplot RMSE et R^2

Agrégation de modèles (9/10)

Synthèse : Erreur de prévision

L'erreur de prévision la moins importante est celle du **bagging de réseaux de neurones**.

Arbres de régression		
<u>Modèle Simple</u>	<u>Bagging</u>	<u>Boosting</u>
10,4%	3,18%	3,6%

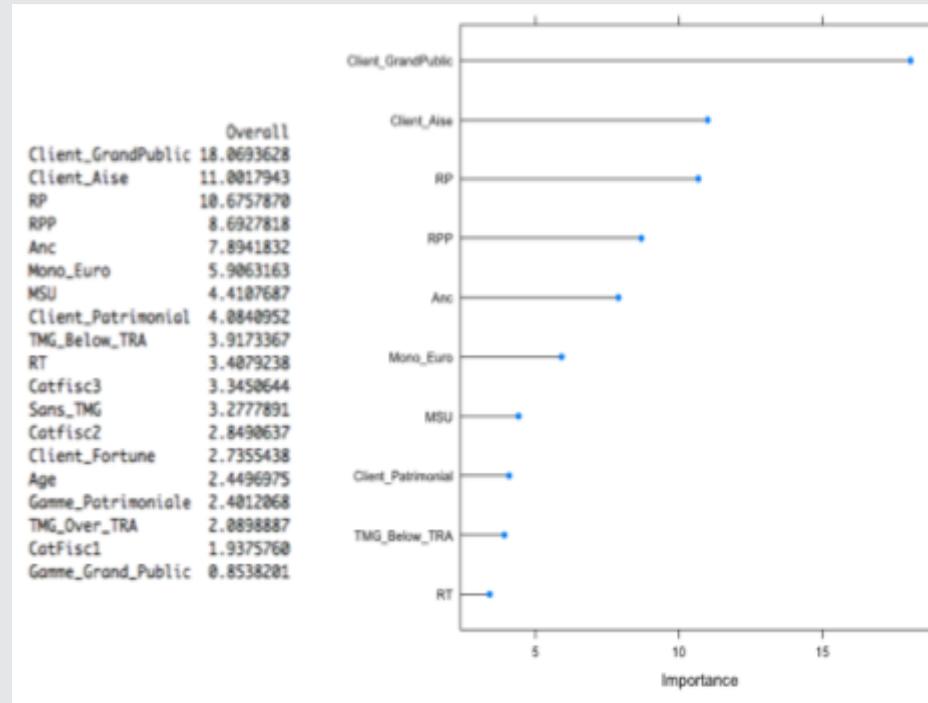
Réseaux de neurones		
<u>Modèle Simple</u>	<u>Bagging</u>	<u>Boosting</u>
4,43%	1,97%	3,88%

Agrégation de modèles (10/10)

Synthèse : Importance des variables

Les résultats obtenus rappellent ceux constatés au niveau des arbres de régression :

- Les **segments de clientèle** sont particulièrement mis en avant, notamment les segments dits "Grand Public" et "Aisé" ;
- Le **type de rachat** intervient ensuite avant l'**ancienneté** et le **type de support**.



Bagging de Réseaux de Neurones - Importance des variables

Validation des modèles

Validation des modèles (1/2)

L'étape de validation permet d'apprécier la robustesse des modèles retenus et la cohérence de l'évolution des comportements des groupes de risques identifiés.

Synthèse : Erreur de prévision

L'erreur de prévision la moins importante est celle du **boosting d'arbres de régressions**.

Arbres de régression		
<u>Modèle Simple</u>	<u>Bagging</u>	<u>Boosting</u>
5,26%	5,65%	2,32%

Réseaux de neurones		
<u>Modèle Simple</u>	<u>Bagging</u>	<u>Boosting</u>
3,67%	3,02%	4,33%

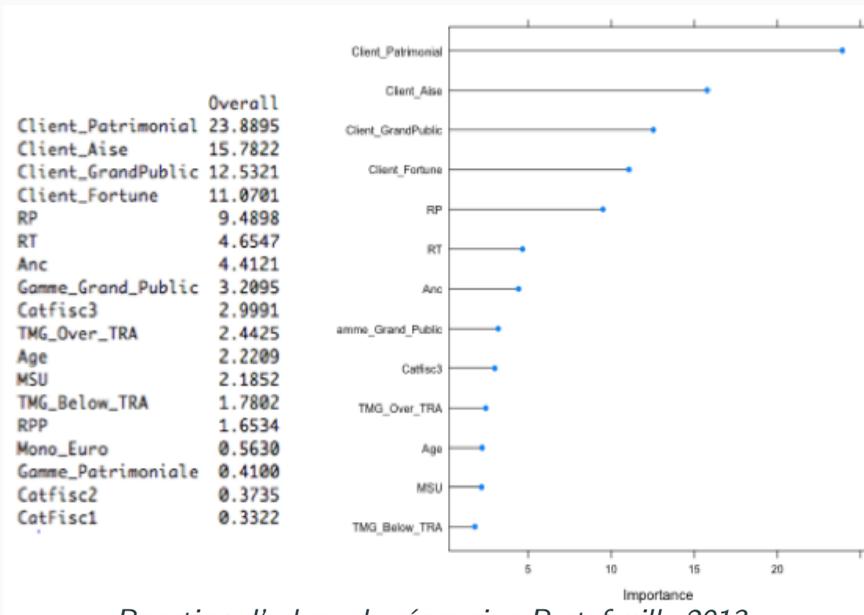
Validation de modèles (2/2)

Synthèse : Importance des variables

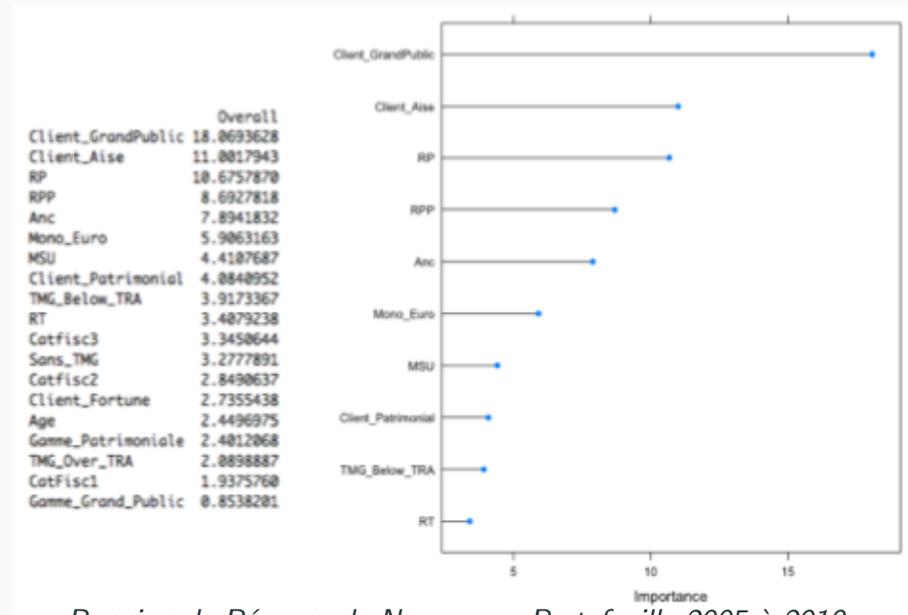
Les résultats obtenus rappellent, globalement*, ceux constatés au niveau du bagging de réseaux de neurones sur le portefeuille 2005 - 2010 :

- Le **segment de clientèle** prend de l'importance avec le temps, c'est d'autant plus imposant en 2013 ;
- Le **type de rachat** intervient ensuite. Le rachat total prend le pas sur le rachat partiel programmé, qui apparaissait en 4^{ème} position dans le cas des données 2005 à 2010.
- L'**ancienneté** des contrats et les autres variables interviennent enfin avec une importance assez proche.

* Les principales évolutions sont dues à des décisions commerciales ou au passage du temps.



Boosting d'arbres de régression Portefeuille 2013



Bagging de Réseaux de Neurones - Portefeuille 2005 à 2010

Conclusion

Conclusion

Dans ce mémoire, nous avons souhaité mettre en lumière l'**un des risques les plus pondérants dans le portefeuille d'un assureur**, celui du **Rachat**.

L'approche linéaire n'était pas suffisante car pouvait passer outre des liens complexes entre les variables. Nous avons donc fait appel à des méthodes de *machine learning*.

- **Avantages**

- ▶ Erreurs de prédiction très faibles (sous condition de calibration minutieuse) ;
- ▶ Détection de relations non-linéaires ;
- ▶ Aspect parcimonieux des réseaux de neurones ;
- ▶ Adaptation à des données volumineuses et hétérogènes ;
- ▶ **Possibilité de prise en charge simultanée de facteurs structurels et dynamiques.**

- **Limites**

- ▶ Instabilité des approximateurs non-linéaires (sensibilité aux minima locaux et au surapprentissage) ;
- ▶ Calibration initiale délicate et implémentation chronophage ;
- ▶ Les facteurs du rachat ne sont pas statiques. Ils subissent l'effet du temps : idée défendue par X. Milhaud [5].

Et puis .. ?

- ▶ Les conclusions de ce travail ont mené à des réflexions sur la construction des Model Points définis au niveau de l'ALM. Celle-ci aura pu être éprouvée par des études d'impact sur des indicateurs de rentabilité (Return-On-Equity (ROE),...);
- ▶ L'analyse factorielle pourrait être suivie d'une Classification Ascendante Hiérarchique (CAH) qui permettrait de réaliser une première segmentation des variables explicatives.
- ▶ Réaliser une carte de Kohonen au lieu de l'AFDM;
- ▶ Modéliser les taux de rachat grâce à d'autres méthodes d'apprentissage : les Machines à Vecteurs de Support (SVM),...;
- ▶ Développer un algorithme d'agrégation entre le bagging et le boosting;
- ▶ Intégrer des facteurs dynamiques aux variables en entrée (nous pourrions penser à, préalablement, faire intervenir des techniques de gestion du temps : séries temporelles,...

Merci pour votre attention

Questions/Réponses

Contact :

salmajamal@kpmg.fr

Références I



O. D. et Maclin R.

Popular ensemble methods : An empirical study.

Journal of Artificial Intelligence Research 11, pages 169–198, 1999.



D. J. S. et P. Bühlman.

Missforest - nonparametric missing value imputation for mixed-type data.

Bioinformatics 28, pages 113–118, 2011.



S. Jamal.

Construction du taux de rachat structurel en Épargne : approximation non linéaire et agrégation de modèles.

Master's thesis, ISFA, 2016.

Références II



R. O. e. C. S. L. Breiman, J. Friedman.

Classification and regression trees.

Wadsworth Brooks, 1984.



X. Milhaud.

Segmentation et modélisation des comportements de rachat en assurance vie.

Master's thesis, Master's thesis, ISFA, 2011.



J. PAGES.

Analyse factorielle de données mixtes.

Revue de statistique appliquée, tome 52, n4, pages 93–111, 2004.



P. Royston.

Multiple imputation for nonresponse in surveys.

New York : Wiley, 2004.

Références III



F. Y. B. P. . L. W. Schapire, R.

Boosting the margin : A new explanation for the effectiveness of voting methods.

Proceedings of the Fourteenth International Conference on Machine Learning. Nashville, TN, page 322–330, 1997.



R. Schapire.

The strength of weak learnability.

Machine Learning, 5(2), page 197–227, 1990.



P. Werbos.

Beyond regression : New tools for prediction and analysis in the behavioral sciences.

Ph. D. dissertation, Committee on Appl. Math., Harvard Univ., Cambridge, MA, 1974.