

Accélération et sécurisation du processus de tarification sur-mesure : comment limiter les variables tarifaires ?

23/06/2016

Guerric BRAS



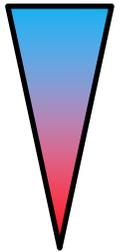
Tarification de garanties complémentaires santé collectives



- **Données démographiques**
 - Tête par tête
 - Fichier paie
 - Données agrégées
 - Hypothèses démographiques

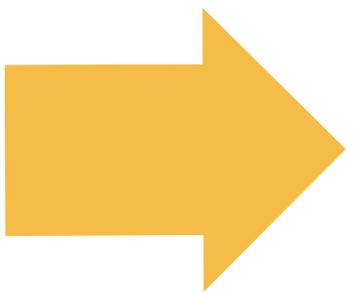
Manque de fiabilité des données

- **Garanties**
 - Droit conventionnel
 - Règlementation
 - Normes de souscription
 - Besoins des salariés



Risque prud'homal

PROCESSUS CHRONOPHAGE ET COMPLEXE



+ 60 % de demandes de tarification sur-mesure
+ 20 % de temps de traitement

Démographie



Tarificateur expérimenté

Grandes
lignes de
garanties

Estimation du tarif

Utilité

- ▶ **Sécurisation**
 - ▶ Repérer les éventuelles erreurs (saisie démographie, garanties, paramètres techniques...)
- ▶ **Accélération**
 - ▶ Repérer les dossiers avec un dumping de la concurrence
 - ▶ Accélérer la souscription

► Modèle tarifaire

Modèle fréquence x coût



Modèle mathématique

$$P_{a,i} = \mathbb{E} \left[\sum_{n=1}^{f_{a,i}} k_{a,i} C^n_{a,i} \right]$$

- Simple mais se complexifie avec le nombre de variables
- Modèle standard : entre 50 et 150 variables
- Pour chacune des variables il faudra vérifier :
 - La pertinence de la donnée
 - Son adéquation avec la politique de souscription – droit conventionnel

▶ Modèle tarifaire

Modèle fréquence x coût



Modèle mathématique

$$P_{a,i} = \mathbb{E} \left[\sum_{n=1}^{f_{a,i}} k_{a,i} C_{a,i}^n \right] \leftarrow \text{Lois des dépenses}$$

- ▶ Simple mais se complexifie avec le nombre de variables
- ▶ Modèle standard : entre 50 et 150 variables
- ▶ Pour chacune des variables il faudra vérifier :
 - ▶ La pertinence de la donnée
 - ▶ Son adéquation avec la politique de souscription – droit conventionnel

▶ Modèle tarifaire

Modèle fréquence x coût



Modèle mathématique

Lois des fréquences

$$P_{a,i} = \mathbb{E} \left[\sum_{n=1}^{f_{a,i}} k_{a,i} C^n_{a,i} \right]$$

An arrow points from the text 'Lois des fréquences' to the term $f_{a,i}$ in the equation, which is circled in yellow.

- ▶ Simple mais se complexifie avec le nombre de variables
- ▶ Modèle standard : entre 50 et 150 variables
- ▶ Pour chacune des variables il faudra vérifier :
 - ▶ La pertinence de la donnée
 - ▶ Son adéquation avec la politique de souscription – droit conventionnel

▶ Modèle tarifaire

Modèle fréquence x coût



Modèle mathématique

$$P_{a,i} = \mathbb{E} \left[\sum_{n=1}^{f_{a,i}} k_{a,i} C^n_{a,i} \right]$$

Correctifs

- ▶ Simple mais se complexifie avec le nombre de variables
- ▶ Modèle standard : entre 50 et 150 variables
- ▶ Pour chacune des variables il faudra vérifier :
 - ▶ La pertinence de la donnée
 - ▶ Son adéquation avec la politique de souscription – droit conventionnel

▶ Modèle tarifaire

Modèle fréquence x coût



Modèle mathématique

$$P_{a,i} = \mathbb{E} \left[\sum_{n=1}^{f_{a,i}} k_{a,i} C^n_{a,i} \right]$$

Variables du modèle

- ▶ Simple mais se complexifie avec le nombre de variables
- ▶ Modèle standard : entre 50 et 150 variables
- ▶ Pour chacune des variables il faudra vérifier :
 - ▶ La pertinence de la donnée
 - ▶ Son adéquation avec la politique de souscription – droit conventionnel

► Objectif



Actuellement...

Variable1 Variable2
Variable3 Variable4
Variable5 Variable6

Modèle tarifaire

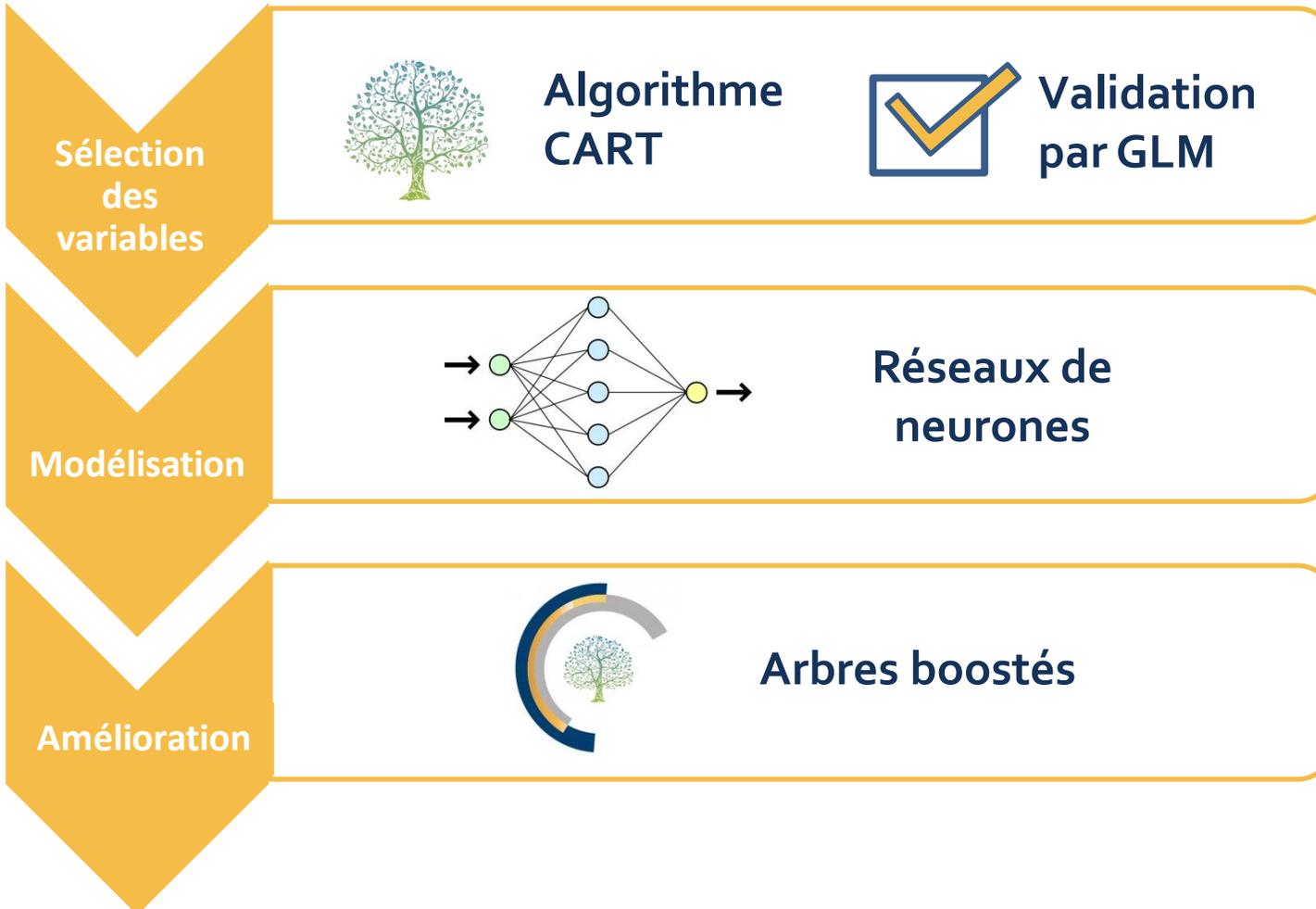
Tarif

Objectif...

Variable1 Variable3
Variable6

Algorithme
d'apprentissage de
l'historique des
tarifications

Estimation du tarif



Données entreprise

- ▶ % Hommes
- ▶ % Mariés
- ▶ Nombre moyen d'enfants
- ▶ % Cadres
- ▶ Age moyen des salariés
- ▶ Nombre de salariés de la société

Données garanties

- ▶ Equivalent € de l'ensemble des actes
- ▶ Tarif uniforme

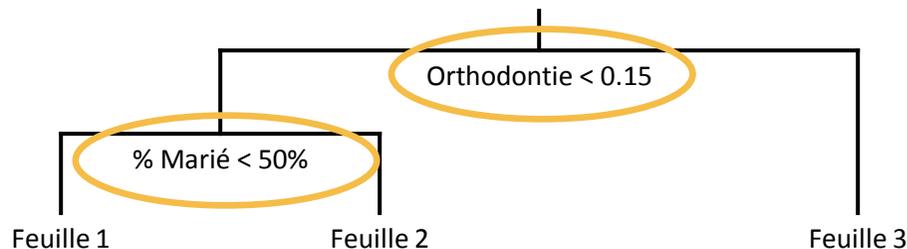
- ▶ Historique des études depuis 2009
- ▶ Retarification des études : Référence tarifaire 2015
- ▶ 6 561 études pour 149 variables / étude

Sélection des variables principales

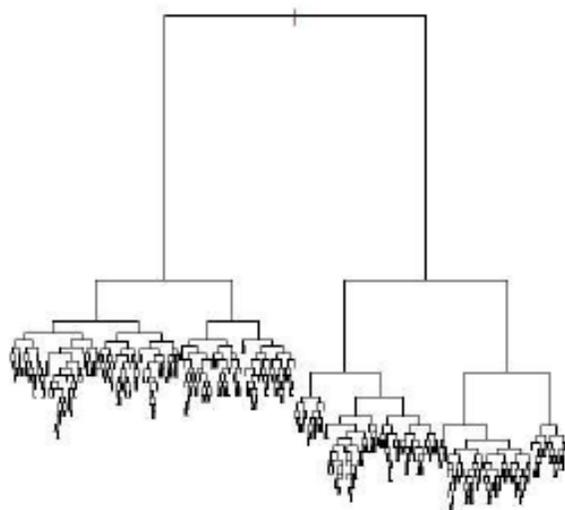
Algorithme CART



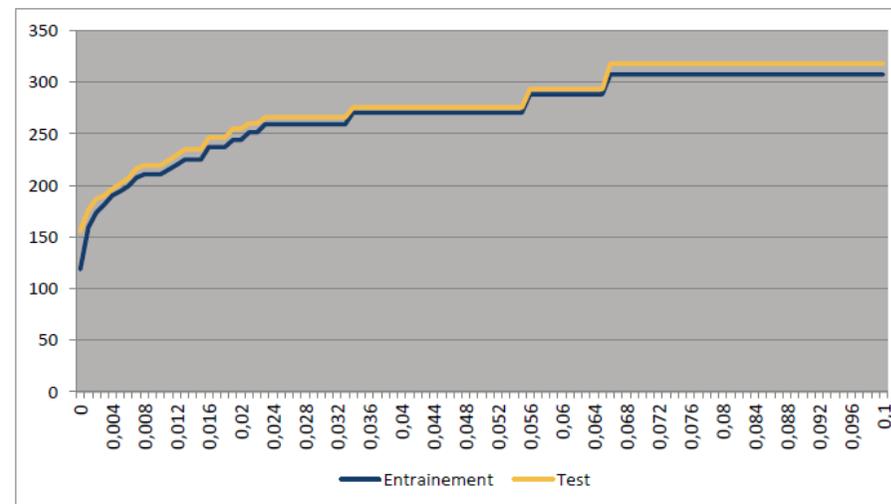
Partition de l'espace des variables



► Segmentation totale de l'espace



► Mise en place d'un coefficient de pénalisation



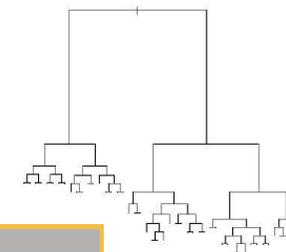
Sélection des variables principales

Algorithme CART



Elagage de l'arbre

- ▶ Permet de réduire le nombre de variables et de n'en conserver que 23



Honoraires - Pharmacie	1	Honoraires - Pharmacie Actes courants Radiologie
	2	Honoraires - Pharmacie Consultations - Visites Conventionnées Consultations de Généralistes Médecin correspondant (PS)
	3	Honoraires - Pharmacie Consultations - Visites Conventionnées Spécialiste Spécialiste (Hors PS)
	4	Honoraires - Pharmacie Petite Chirurgie Conventionnée ADA
	5	Honoraires - Pharmacie Petite Chirurgie Conventionnée ADC
	6	Honoraires - Pharmacie Petite Chirurgie Non conventionnée ATM
Hospitalisation	7	Hospitalisation - Chirurgie Accessoire Chambre particulière
	8	Hospitalisation - Chirurgie Accessoire Forfait journalier
	9	Hospitalisation - Chirurgie Conventionnées Honoraires chirurgicaux ADE
	10	Hospitalisation - Chirurgie Non conventionnées Frais de séjour
Dentaire	11	Dentaire Hors nomenclature Implantologie
	12	Dentaire Orthodontie Acceptée
	13	Dentaire Prothèses Acceptées Fixes métal non visibles
	14	Dentaire Prothèses Acceptées Fixes métal visibles
Optique	15	Optique Lentille (unité) Refusée
	16	Optique Monture
	17	Optique Verres
Autres	18	Autres Cures thermales Honoraires
	19	Autres Maternité Chambre particulière
Démographie	20	% Hommes
	21	% Marié
	22	Age moyen
	23	Nombre moyen d'enfants

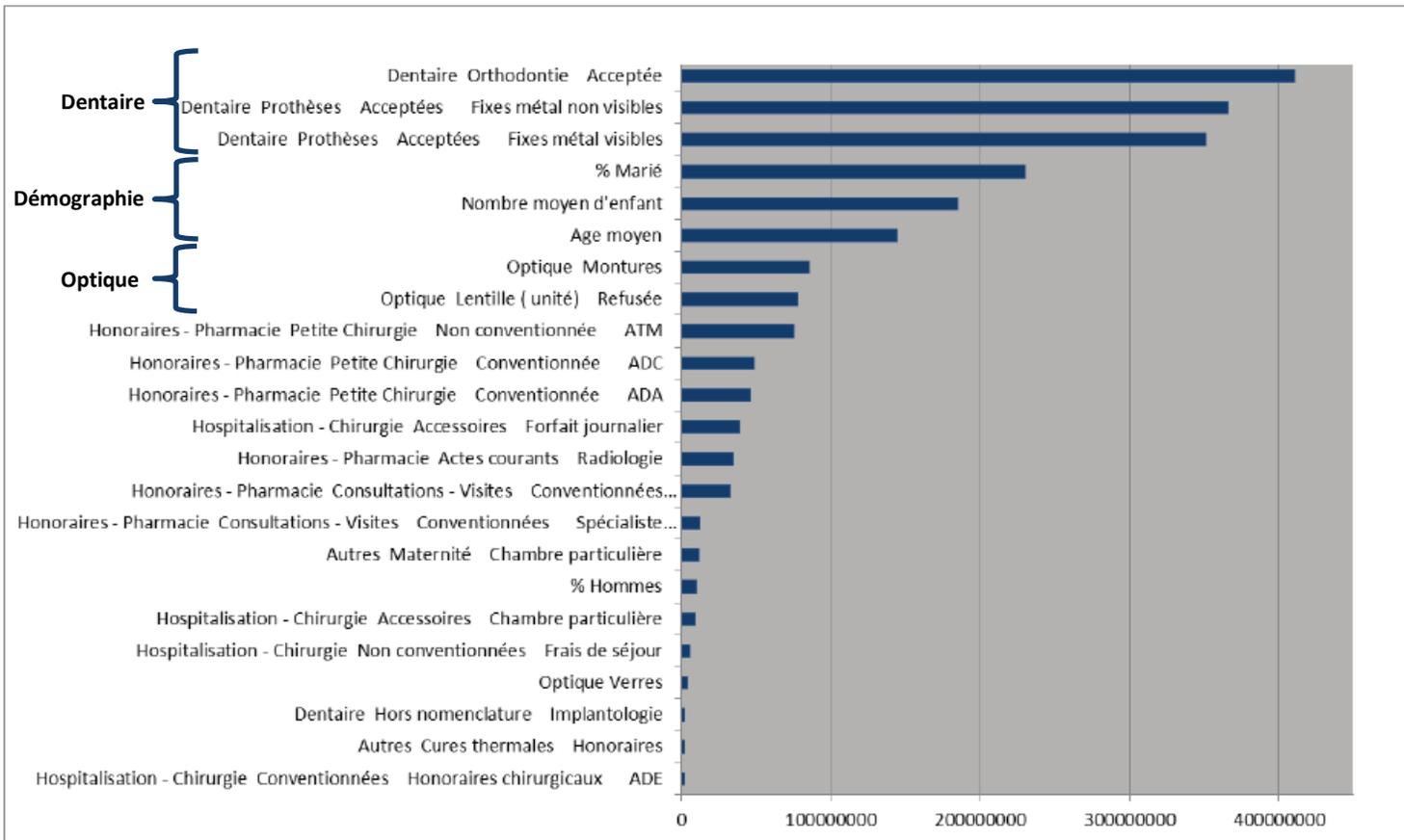
Sélection des variables principales

Algorithme CART



Importance des variables

Les variables sont conservées en fonction de leur apport d'information dans la segmentation.



Les **garanties dentaires** apportent le plus d'information dans la segmentation des études.

Viennent ensuite les informations sur la **démographie** et les **garanties optiques**.

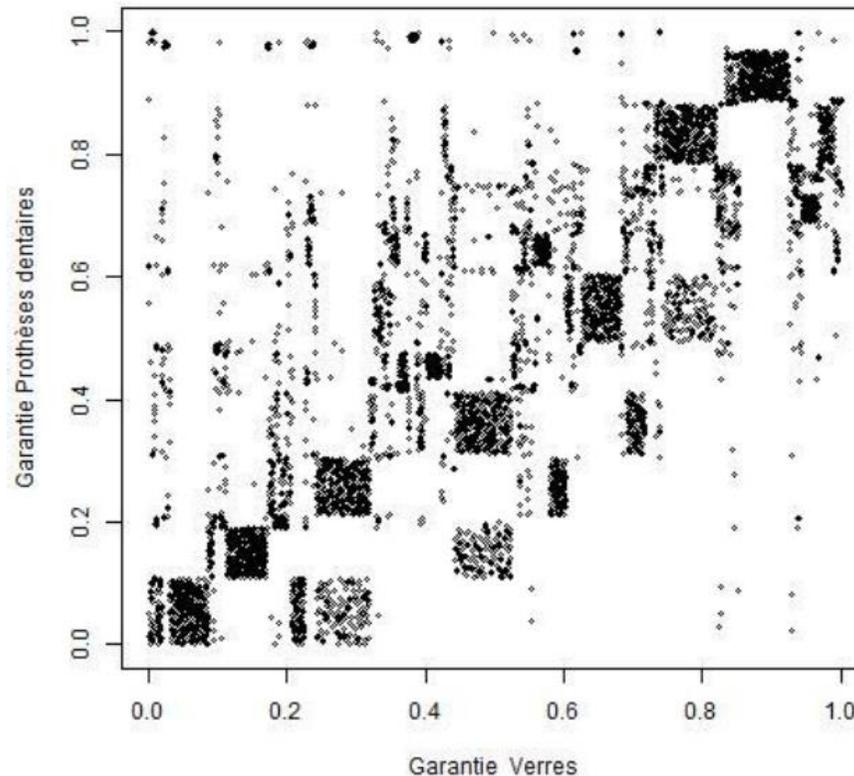
Sélection des variables principales

Validation des variables conservées



Etude des corrélations

- ▶ L'algorithme CART ne regarde pas les corrélations entre les variables.
 - ▶ Conservation de garanties très proches.



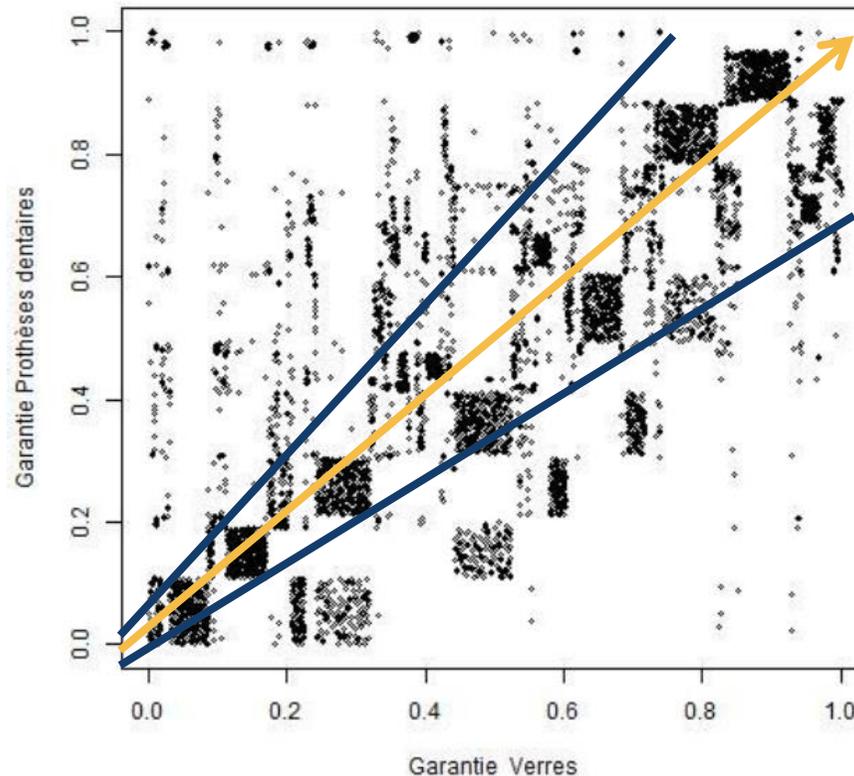
Sélection des variables principales

Validation des variables conservées



Etude des corrélations

- ▶ L'algorithme CART ne regarde pas les corrélations entre les variables.
 - ▶ Conservation de garanties très proches.



► Sélection des variables principales

Validation des variables conservées



Etude des corrélations

- L'étude des corrélations et des copules permet de réduire le nombre de variables à 20.
 - Retrait des **Visites de généralistes**
 - Retrait de la **Petite chirurgie conventionnée**
 - Retrait des **Prothèses dentaires acceptées visibles**

► Résultats complémentaires



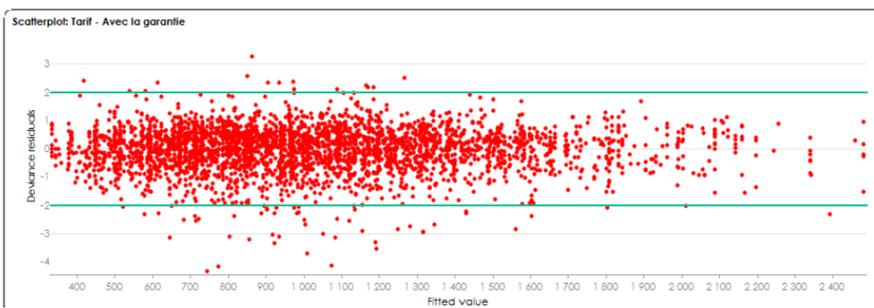
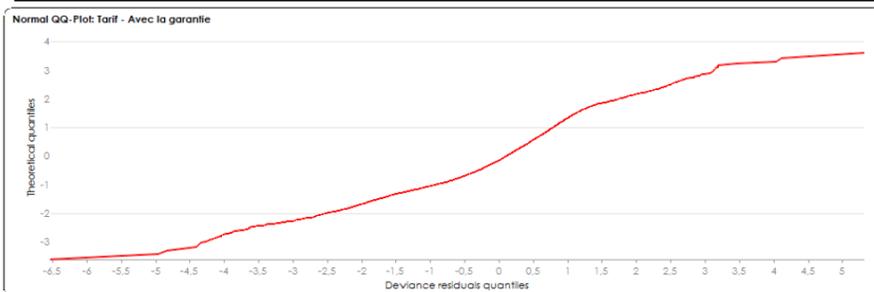
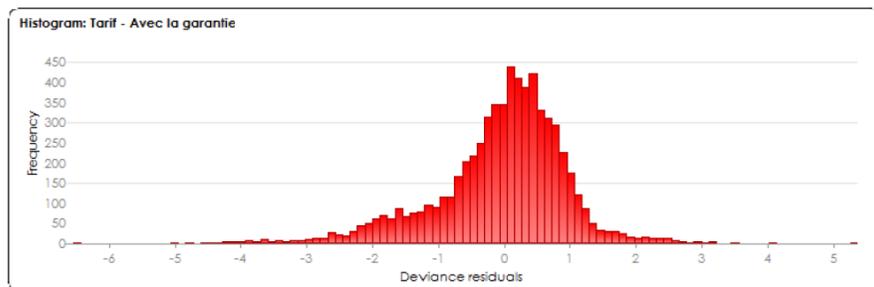
Sélection des variables principales

Validation des variables conservées



GLM

- ▶ Permet de confirmer l'apport des variables dans l'estimation du tarif uniforme.



- ▶ Seules les garanties **montures** et **honoraires chirurgicaux conventionnés** ne sont pas conservées dans la modélisation
- ▶ 18 variables conservées dans la modélisation
- ▶ - 90% de variables

► Approximation des tarifs

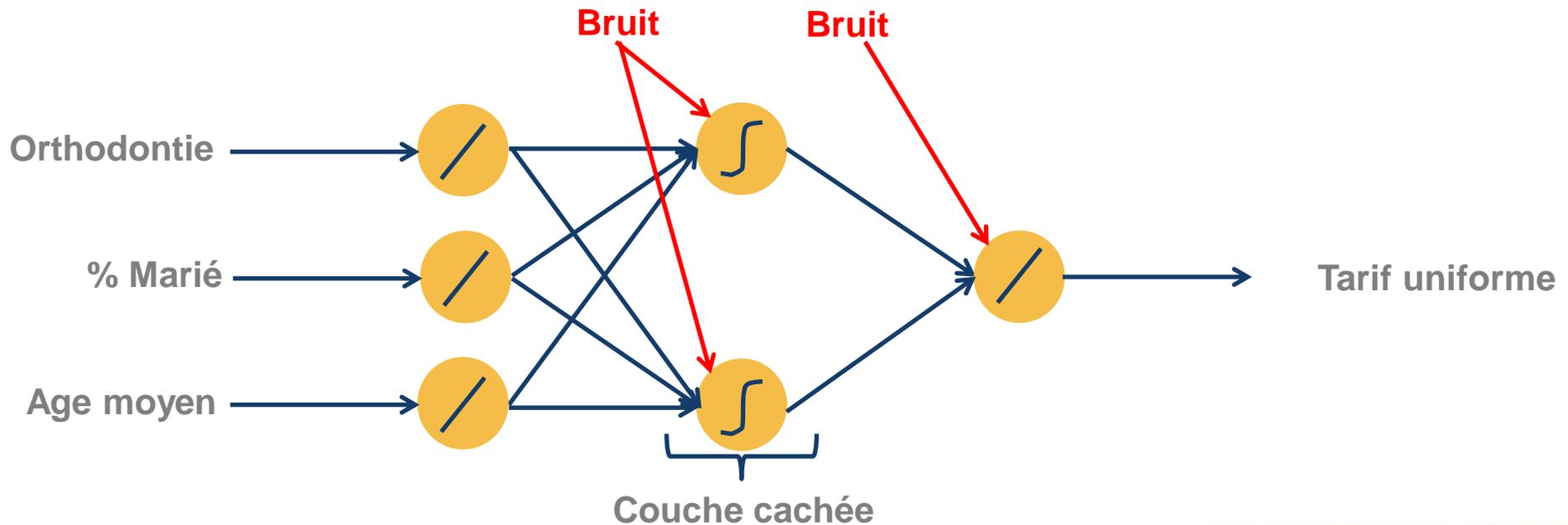
Réseaux de neurones



Comment approximer les tarifs avec moins de variables ?

► Réseaux de neurones

► « Toute fonction bornée suffisamment régulière peut être **approchée uniformément** avec une **précision arbitraire**, dans un domaine fini de l'espace de ses variables, par un **réseau de neurones** comportant une **couche de neurones cachés** en nombre fini, possédant tous la même fonction d'activation, et un neurone de sortie linéaire. » (Cybenko & Funahashi)

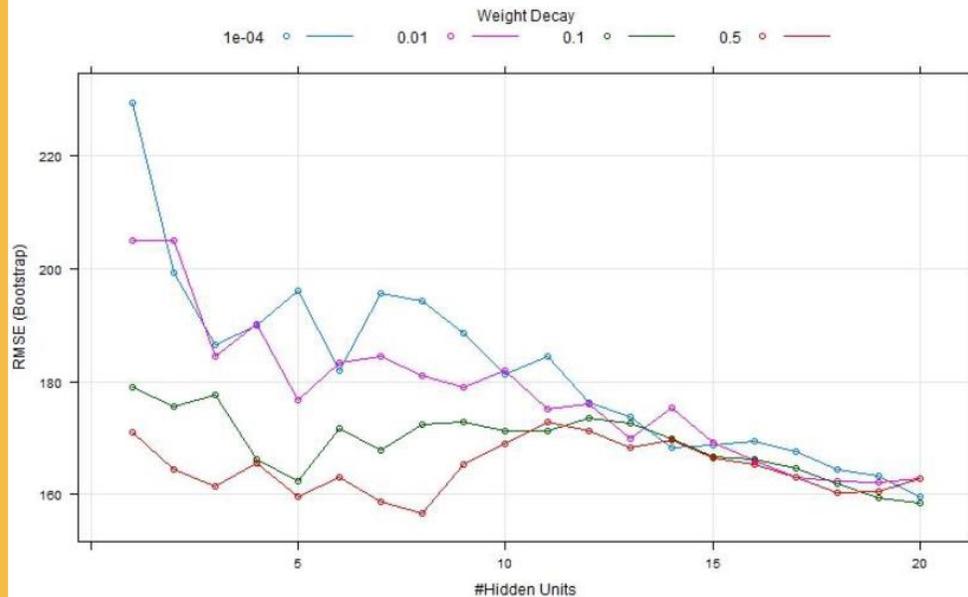


► Approximation des tarifs

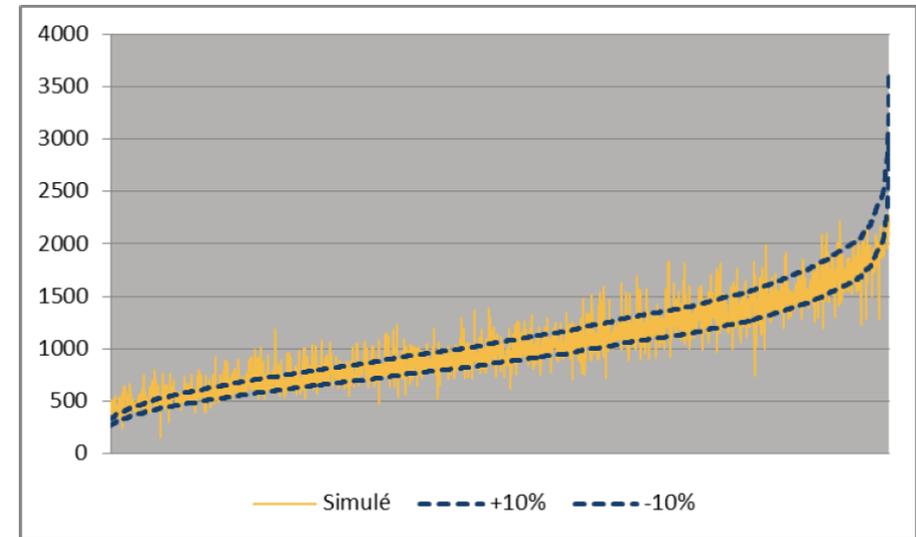
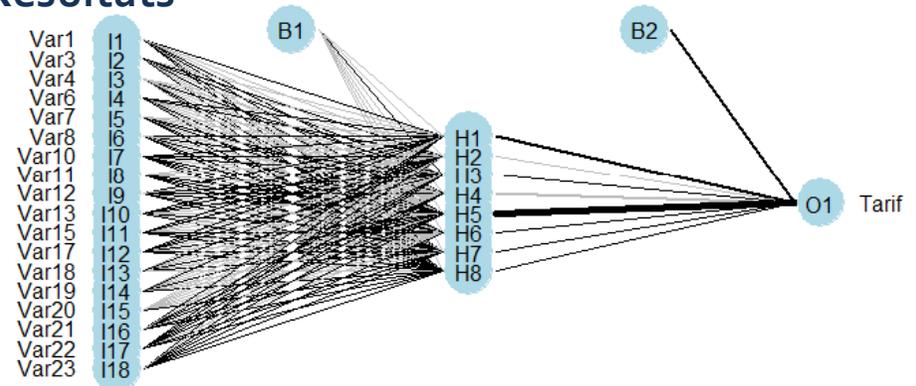
Réseaux de neurones

Paramètres du réseau à déterminer

- Nombre de neurones de la couche cachée
- Coefficient de dégradation



Résultats



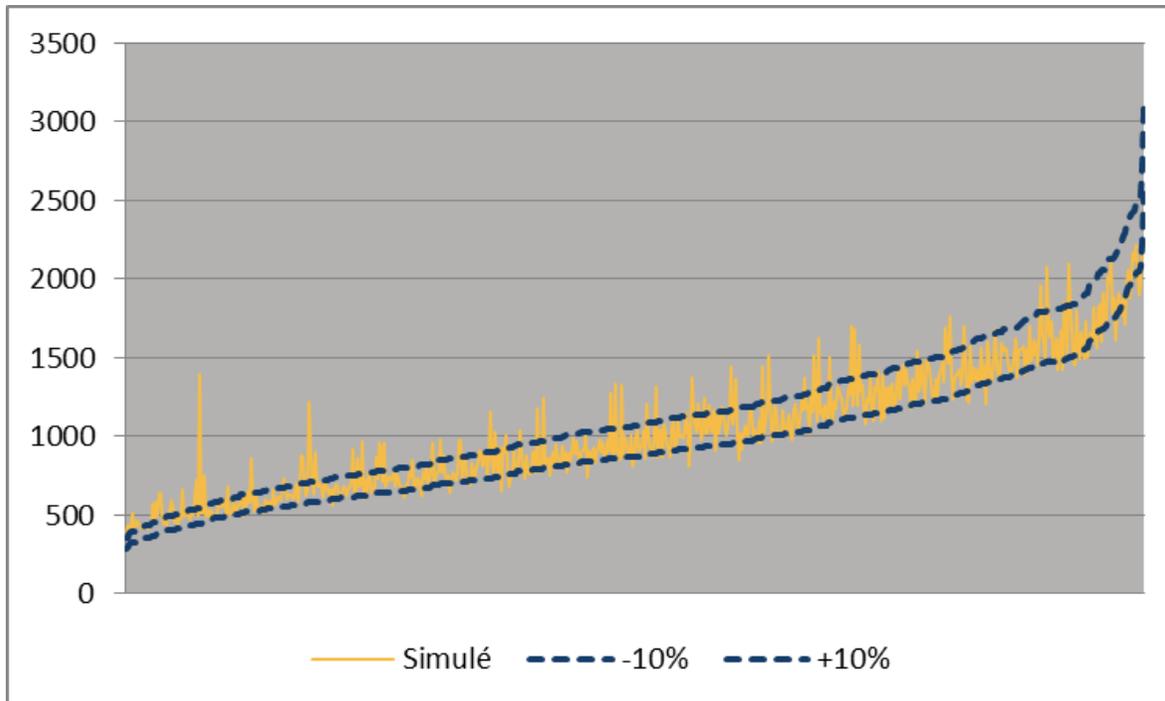
► Approximation des tarifs

Réseaux de neurones



Améliorations du réseau de neurones

- Moyenne de plusieurs réseaux de neurones
- Augmentation de la base d'apprentissage



90%
de la
base

-90 % de
variables

8 neurones
cachés

20
itérations

Var
1.36%

Tarif
+ ou -
9%

RMSE
129

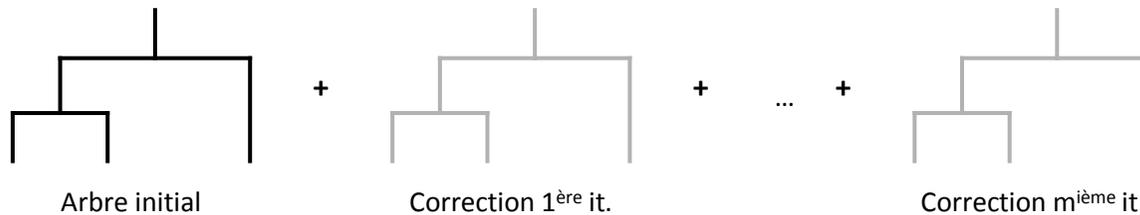
Amélioration des estimations

Arbres boostés



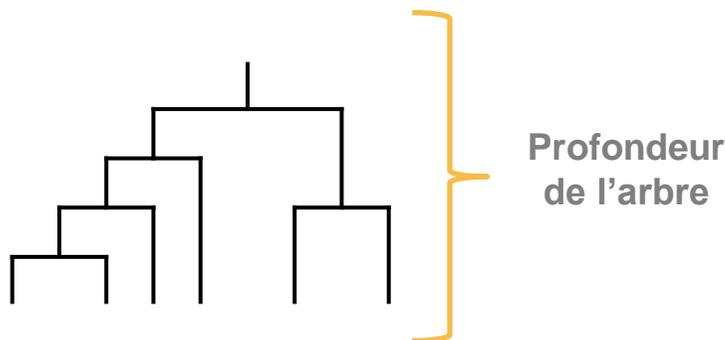
Principes

- ▶ Application de la méthode de boosting aux arbres de régression



$$F_m(x) = F_{m-1}(x) + \rho_m \sum_{j=1}^J b_{jm} \mathbf{1}_{\{x \in R_{jm}\}}$$

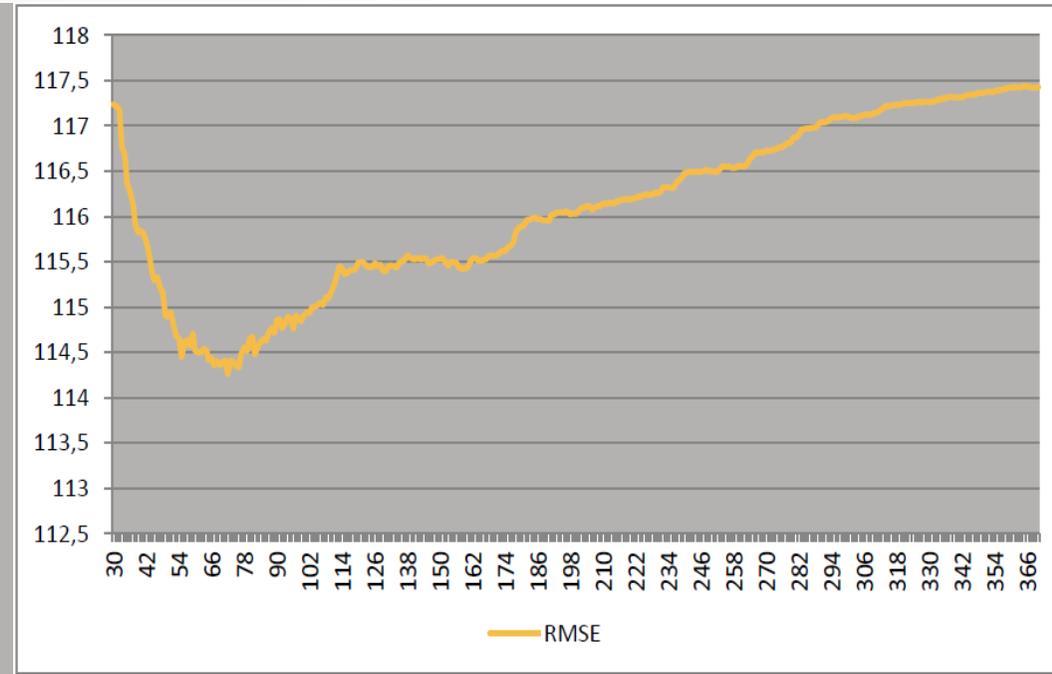
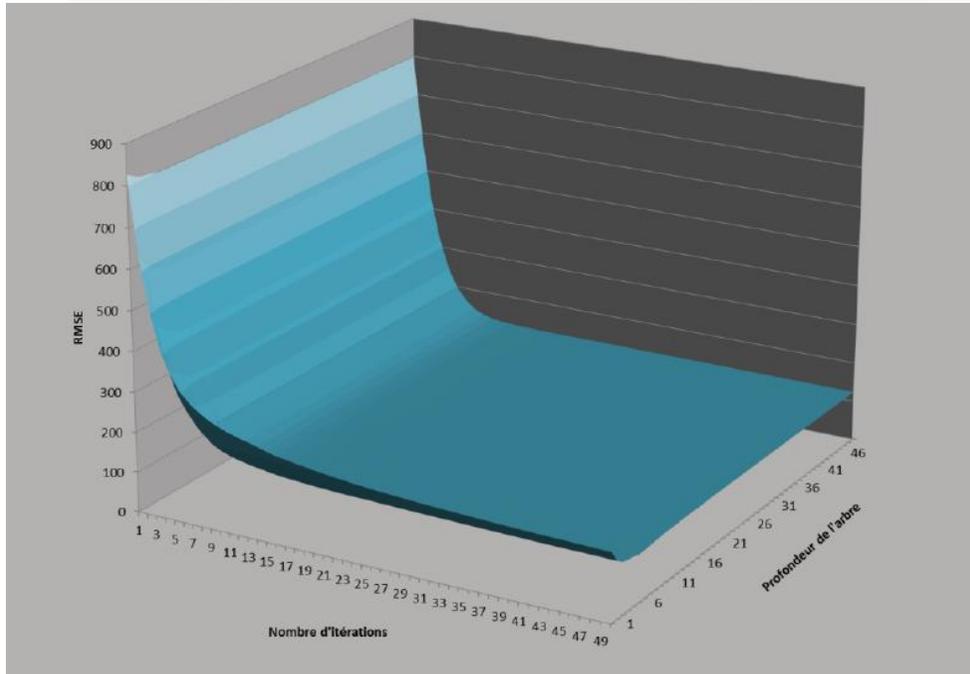
- ▶ Deux points à prendre en considération :



- ▶ Attention au surapprentissage !

Amélioration des estimations

Arbres boostés



Meilleur modèle

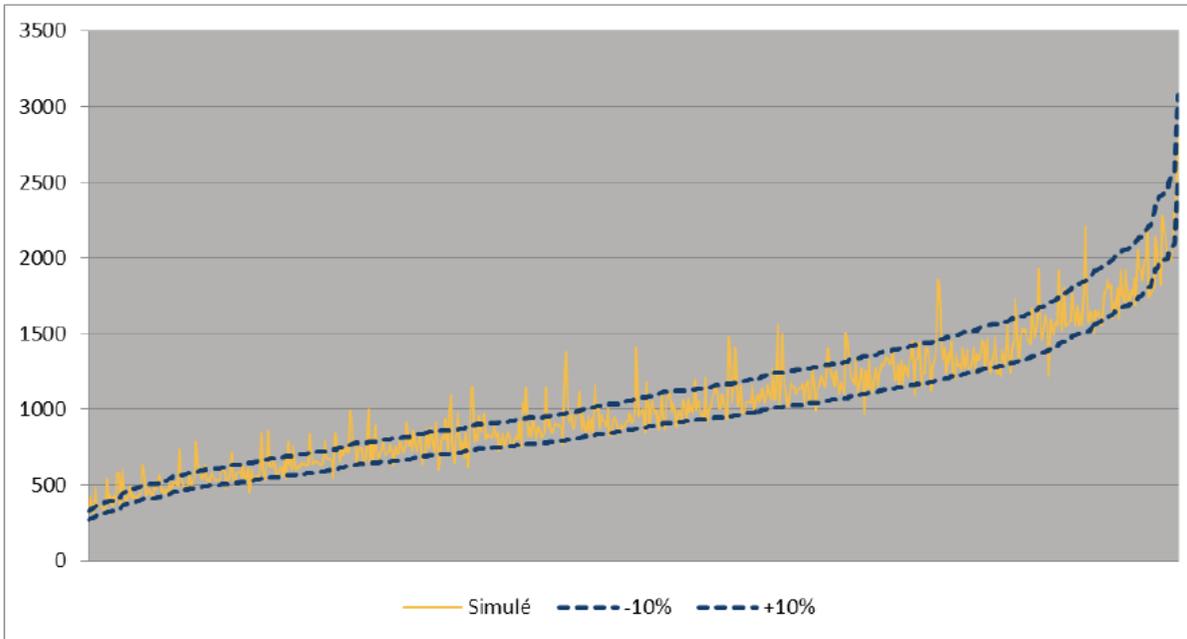
- ▶ Profondeur : 6
- ▶ Nombres d'itérations : 72

Amélioration des estimations

Arbres boostés



Résultat



Réduction de 14% de l'erreur par rapport aux réseaux de neurones

- 90% de la base
- 90 % de variables
- Profondeur 6
- 72 itérations
- Var 0.85%
- Tarif + ou - 8%
- RMSE 113

Synthèse des résultats

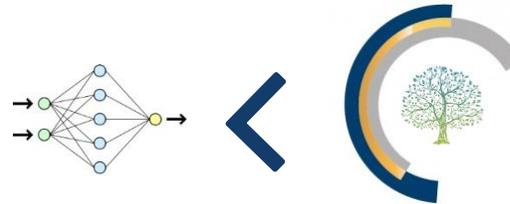


	Variables principales			RMSE
	1 ^{ère} Variable	2 ^{ème} Variable	3 ^{ème} Variable	
Algorithme CART	Orthodontie acc.	Prothèse dent. Acc.	% Mariés	
Réseau de neurones Base d'apprentissage 70%	Prothèse dent. Acc.	% Mariés	Age moyen	148,32
20 Réseaux de neurones Base d'apprentissage 70%				134,48
20 Réseaux de neurones Base d'apprentissage 90%				129,19
Arbres boostés Base d'apprentissage 90%	Orthodontie acc.	% Mariés	Prothèse dent. Acc.	112,92

► Conclusion



- **Approximation** possible des **tarifs sur-mesure**
 - Réduction de 90% du nombre de variables
 - Apprentissage sur un historique de tarification



ACCELERATION et SECURISATION de la TARIFICATION



- Toujours besoin des **méthodes et outils classiques** de tarification
- Difficultés d'application de cette méthode en suivi du risque