



# Nouvelles méthodes actuarielles et data science

---

FONCTIONNEMENT D'UNE COMPAGNIE  
D'ASSURANCE

---

Chauvet Lisa  
Diabate Souleymane  
Goutondji Anaëlle  
Konate Vamoussa  
Mohamed El Hafed Ismail  
Pelenc Ines

# Table des matières

## I- Anciennes et nouvelles méthodes actuarielles

A- Quelques méthodes utilisées actuellement en actuariat

B- Détail de 2 méthodes nouvellement utilisées en actuariat

## II- Applications

A- Mémoire sur la tarification automobile

B- Mémoire sur une comparaison des méthodes actuarielles  
dans le Reporting Santé

C- Mémoire sur la tarification habitation

## III- Conclusion

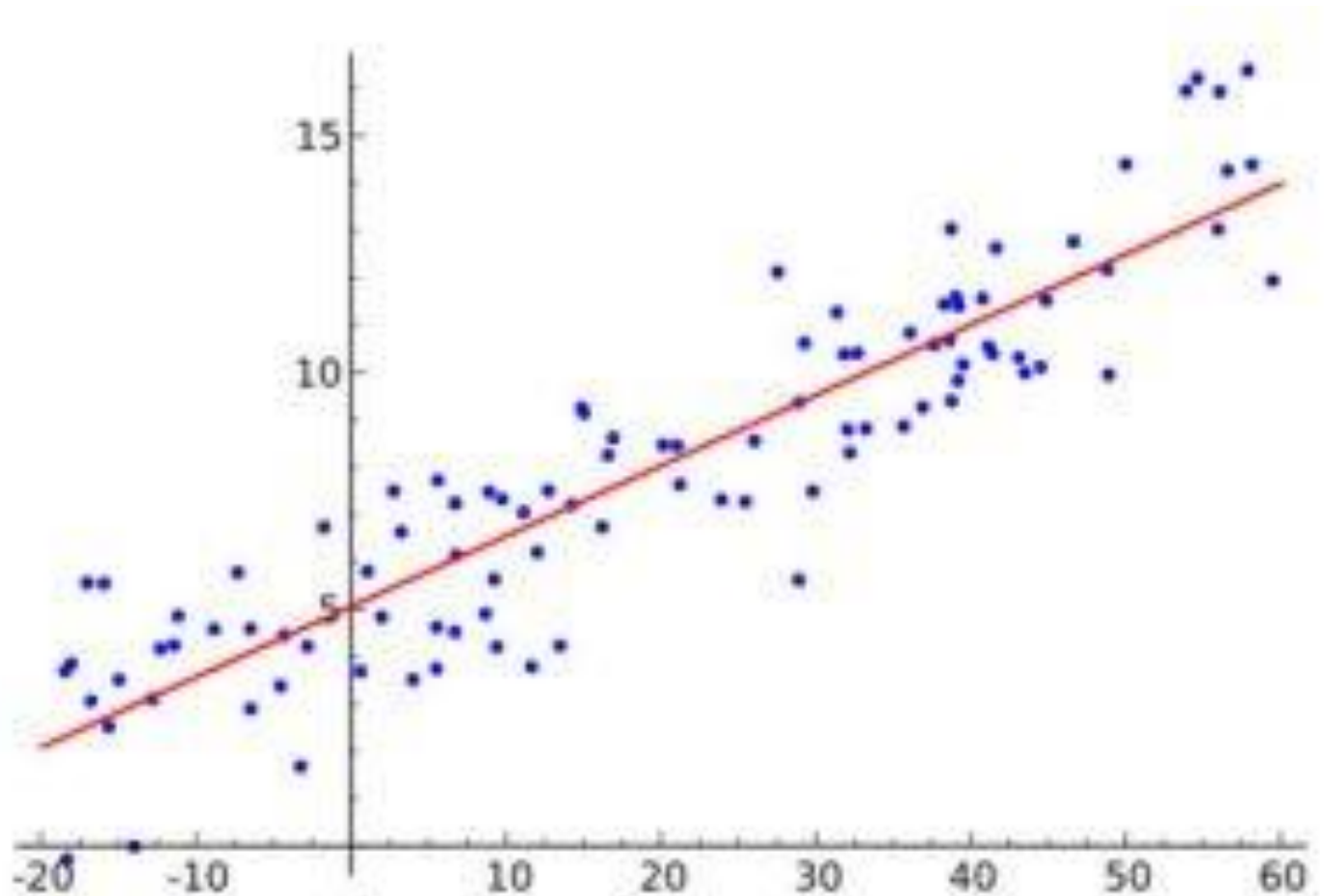


# I- Anciennes et nouvelles méthodes actuarielles

# Quelques méthodes utilisées actuellement en actuariat

---

- **Modèle linéaire**



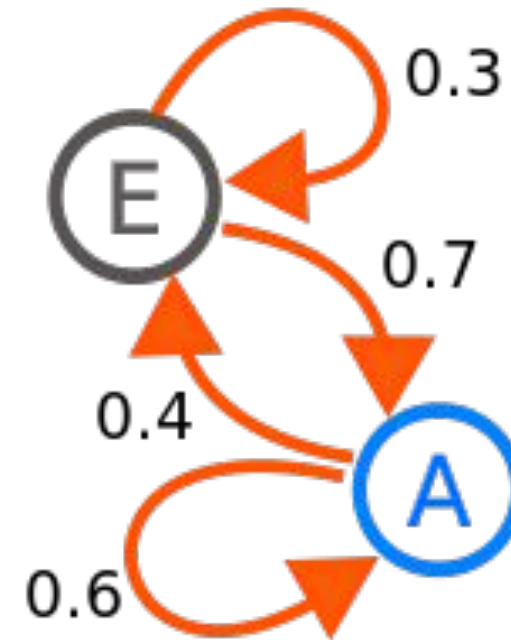
# Quelques méthodes utilisées actuellement en actuariat

- Modèle linéaire
- Modèle de survie

| Entry<br>Ages | Original<br>Data<br>$q_{[x]5}$ | KING       |                            | HARDY      |                            |
|---------------|--------------------------------|------------|----------------------------|------------|----------------------------|
|               |                                | $q_{[x]5}$ | Deviation<br>$\times 10^5$ | $q_{[x]5}$ | Deviation<br>$\times 10^5$ |
| 20 to 24      | ·02400                         | ·02343     | — 57                       | ·02357     | — 43                       |
| 25 „ 29       | ·02380                         | ·02515     | + 135                      | ·02530     | + 150                      |
| 30 „ 34       | ·02766                         | ·02794     | + 28                       | ·02808     | + 42                       |
| 35 „ 39       | ·03239                         | ·03237     | — 2                        | ·03244     | + 5                        |
| 40 „ 44       | ·03874                         | ·03922     | + 48                       | ·03925     | + 51                       |
| 45 „ 49       | ·05117                         | ·04983     | — 134                      | ·04979     | — 138                      |
| 50 „ 54       | ·07061                         | ·06596     | — 465                      | ·06582     | — 479                      |
| 55 „ 59       | ·09245                         | ·09088     | — 157                      | ·09069     | — 176                      |
| 60 „ 64       | ·13004                         | ·12652     | — 352                      | ·12634     | — 370                      |
| Totals        | ·49086                         | ·48130     | + 211<br>— 1167            | ·48128     | + 248<br>— 1206            |

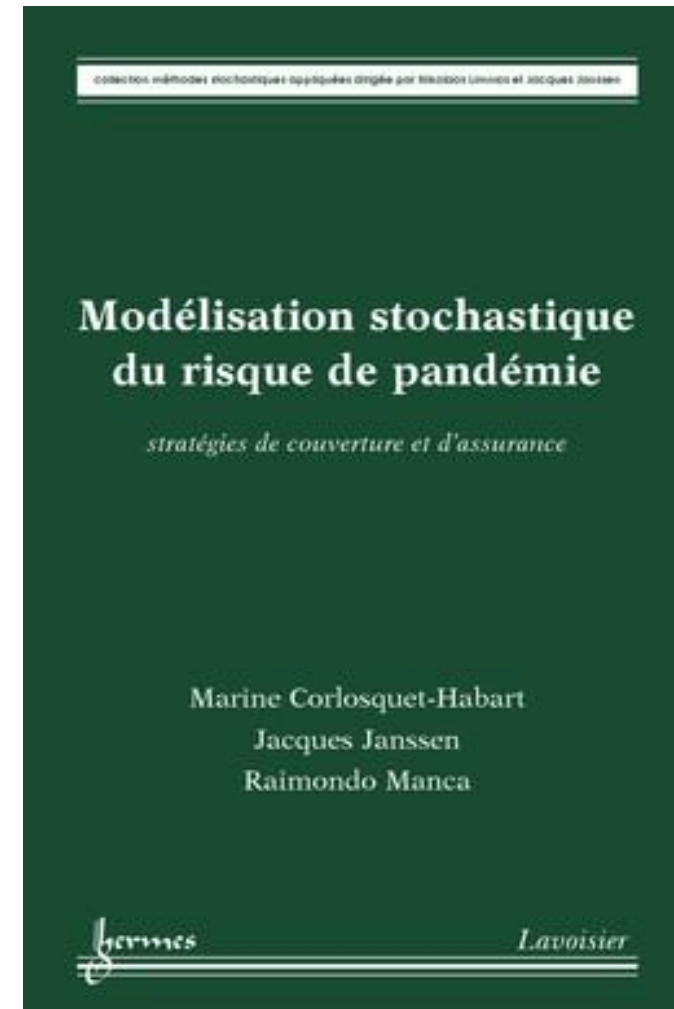
# Quelques méthodes utilisées actuellement en actuariat

- **Modèle linéaire**
- **Modèle de survie**
- **Modèle de Markov**



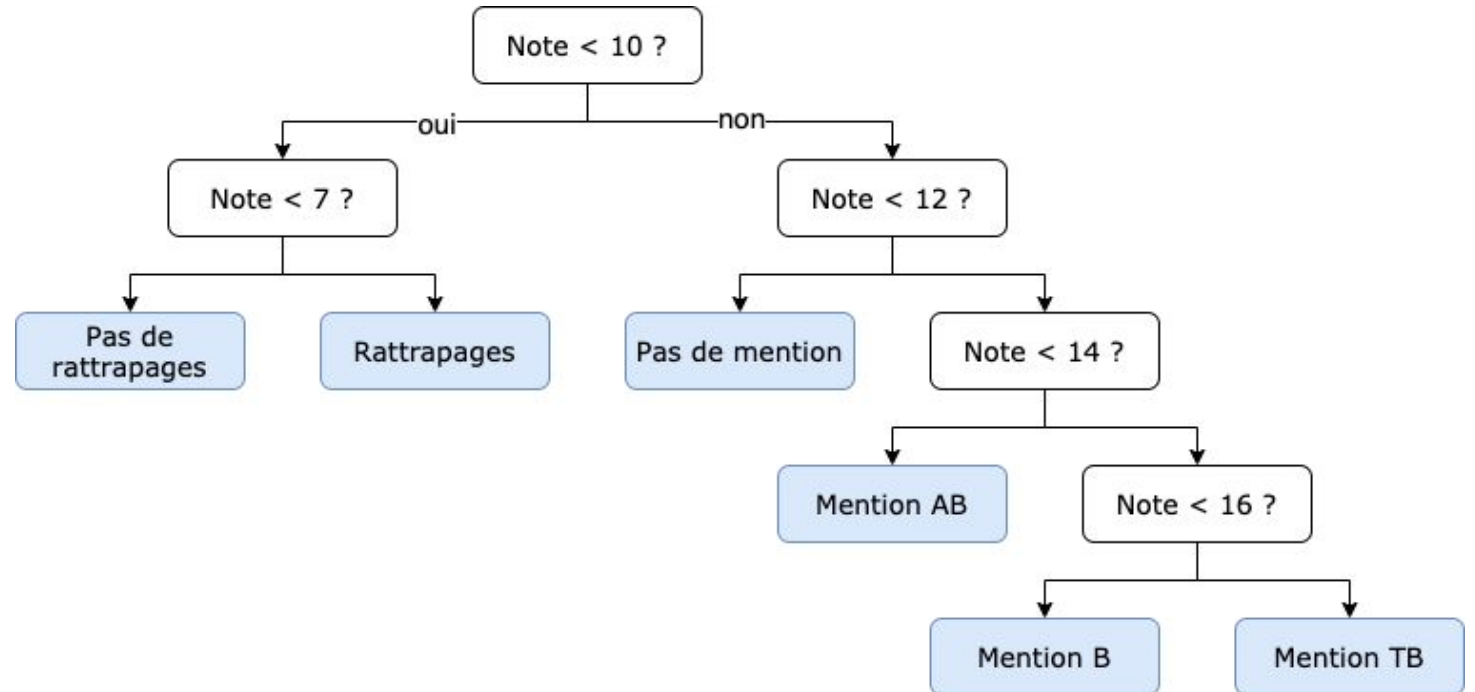
# Quelques méthodes utilisées actuellement en actuariat

- Modèle linéaire
- **Modèle de survie**
- **Modèle de Markov**
- **Modèle stochastique**
- **Modélisation statistique**



# Méthodes machine learning

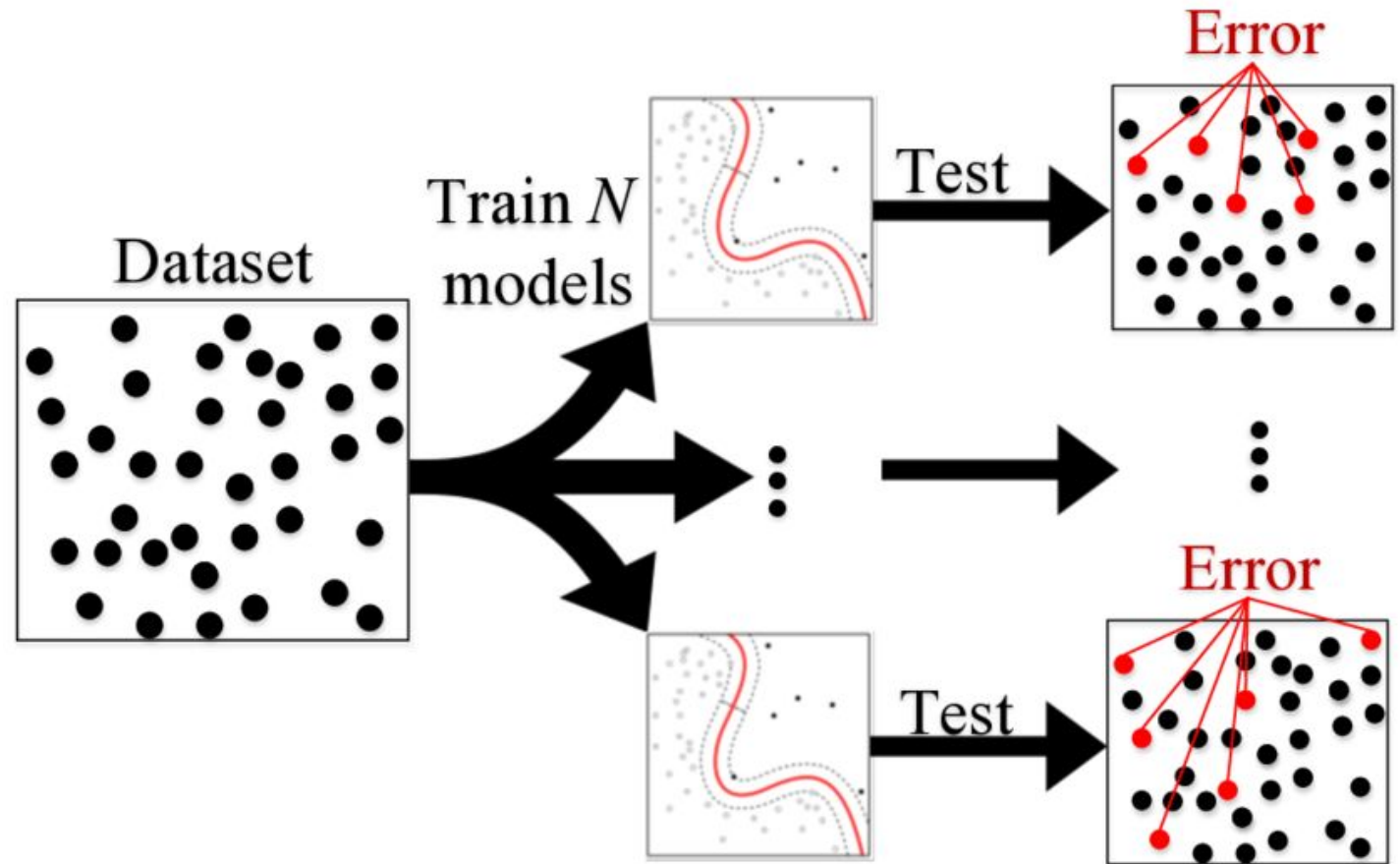
- Arbres de décision et modèle de CART





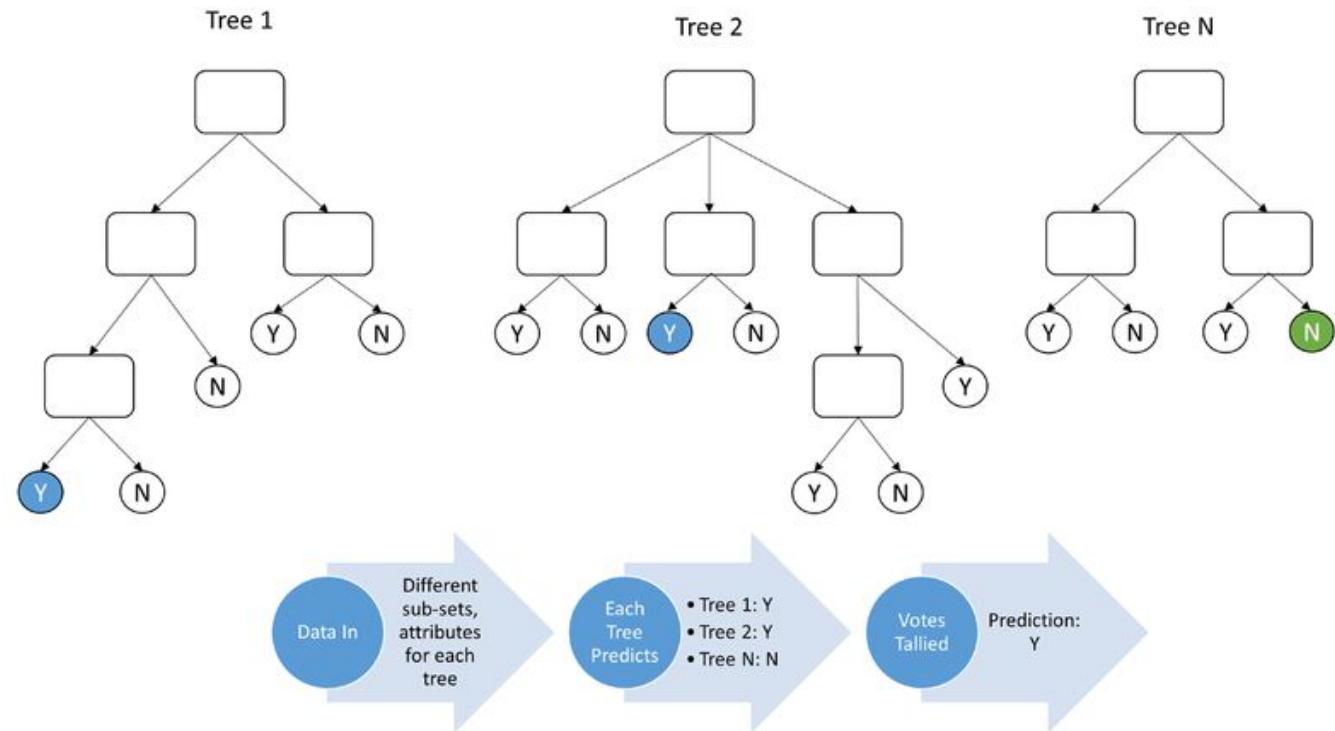
# Méthodes machine learning

- Arbres de décision et modèle de CART
- Méthode de bagging



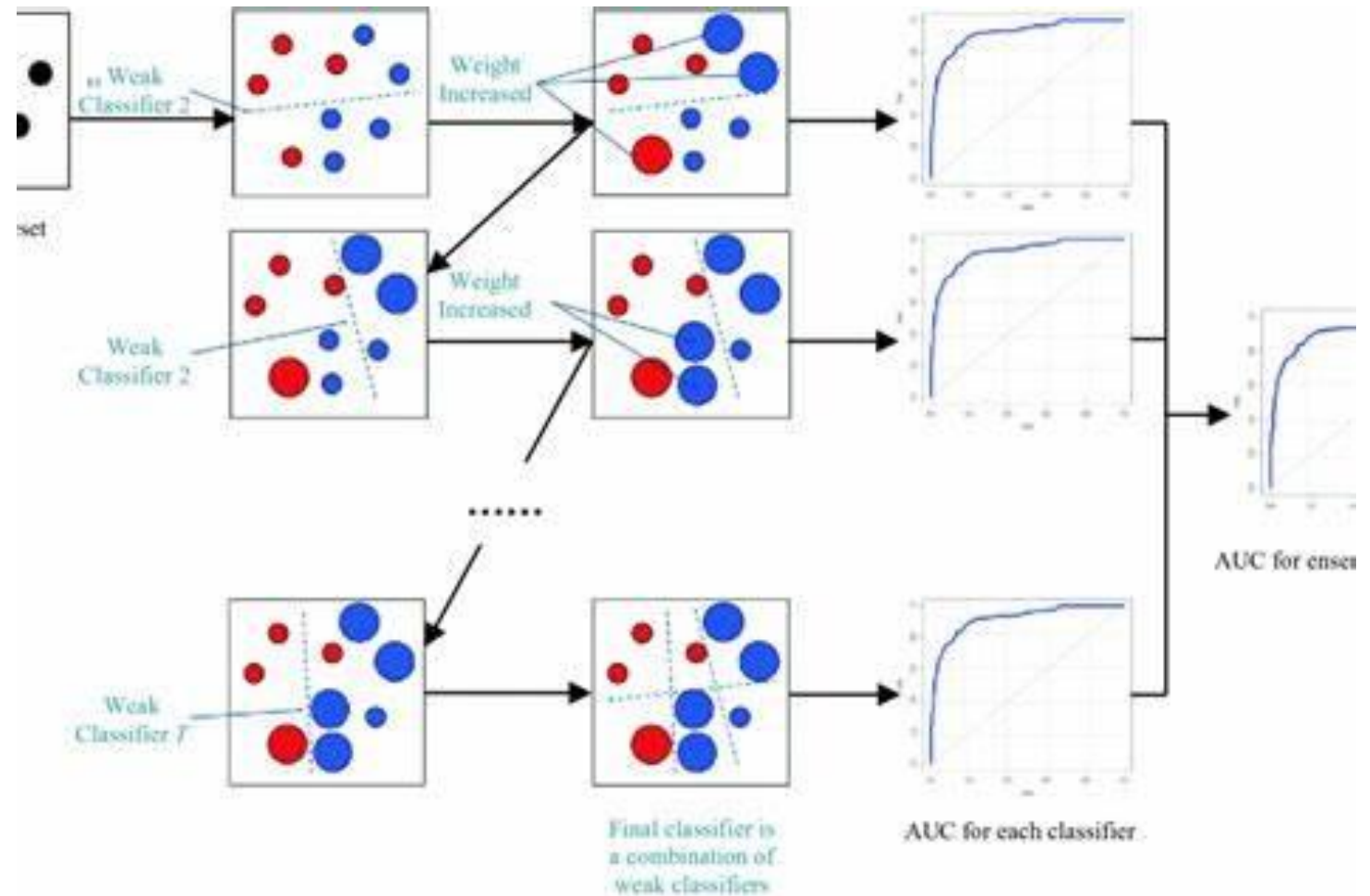
# Méthodes machine learning

- Arbres de décision et modèle de CART
- Méthode de bagging
- Méthode Random Forest



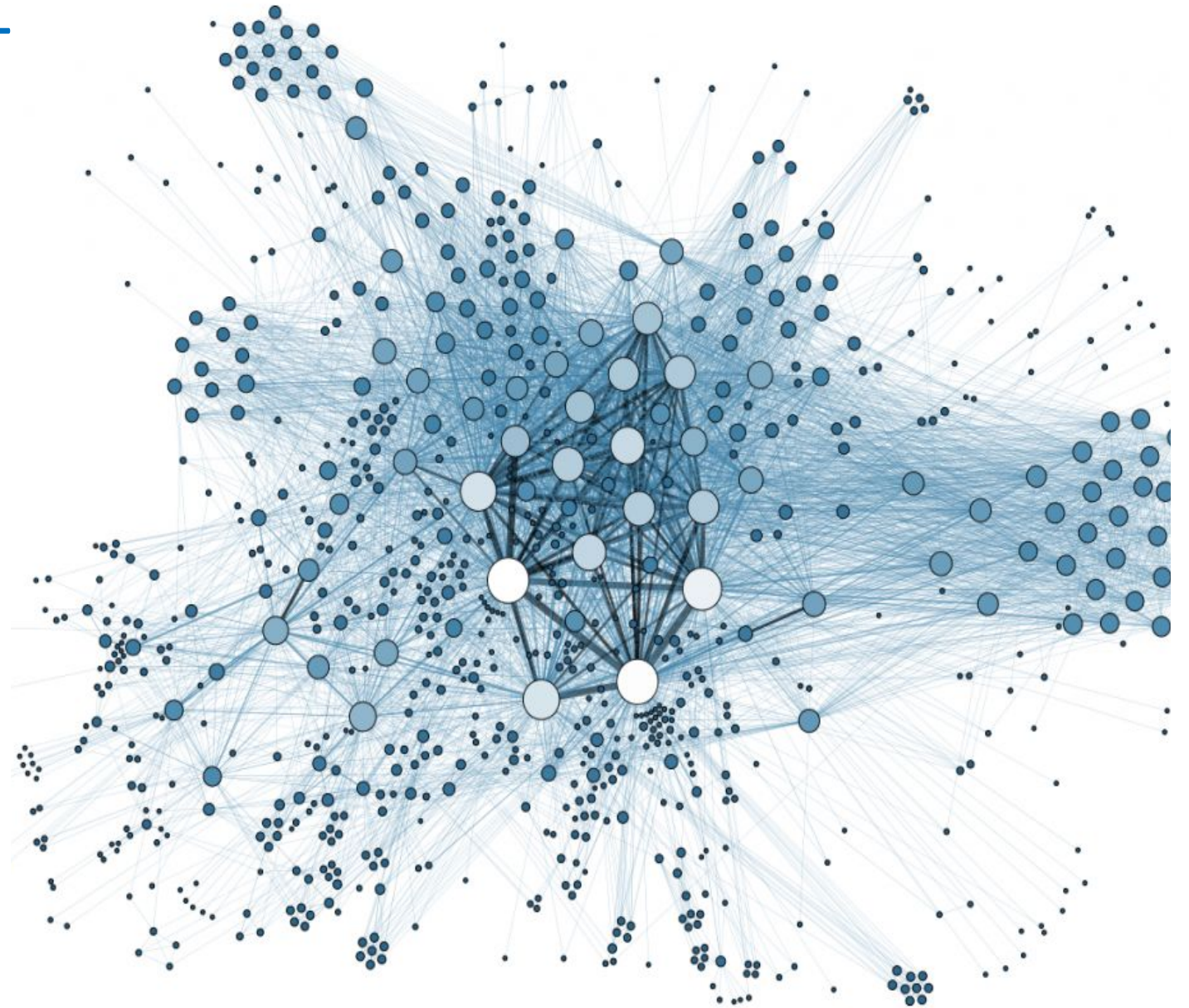
# Méthodes machine learning

- Arbres de décision et modèle de CART
- Méthode de bagging
- Méthode Random Forest
- Méthode de Gradient boosting



# Data Visualisation – P1

- « Une image vaut mille mots »  
Confucius
- Rapidité
- Démocratisation des données
- Théorie de Gestalt
- Rôle du DataScientist :
- Collecter
- Préparer
- Analyser et extraire



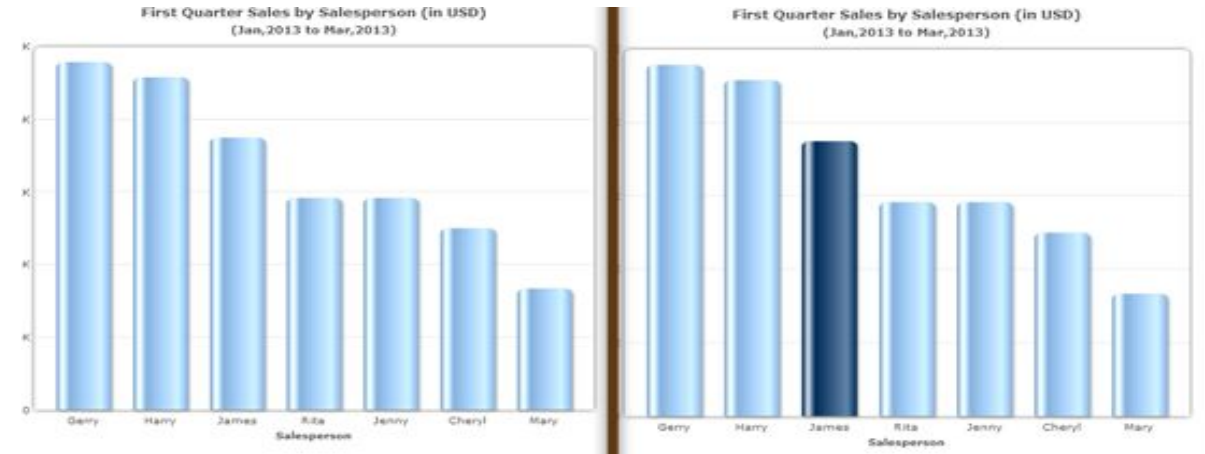
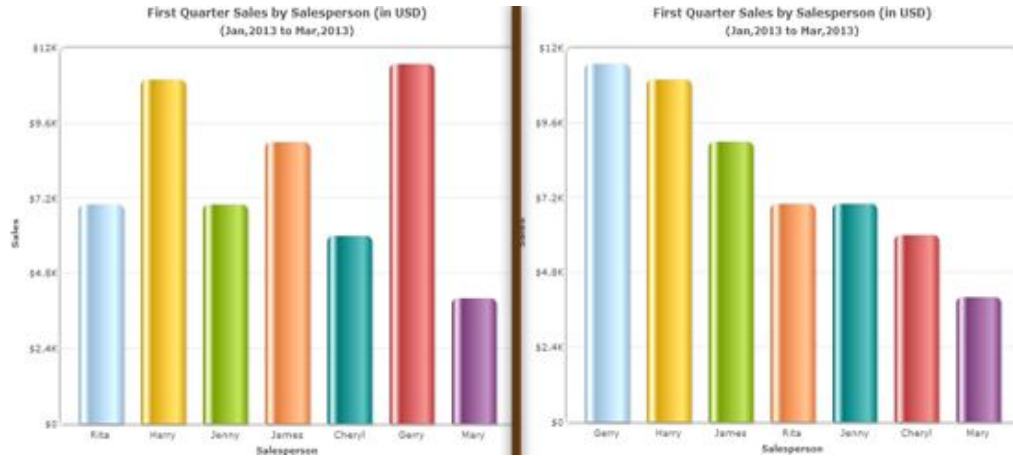


# Data Visualisation – P2

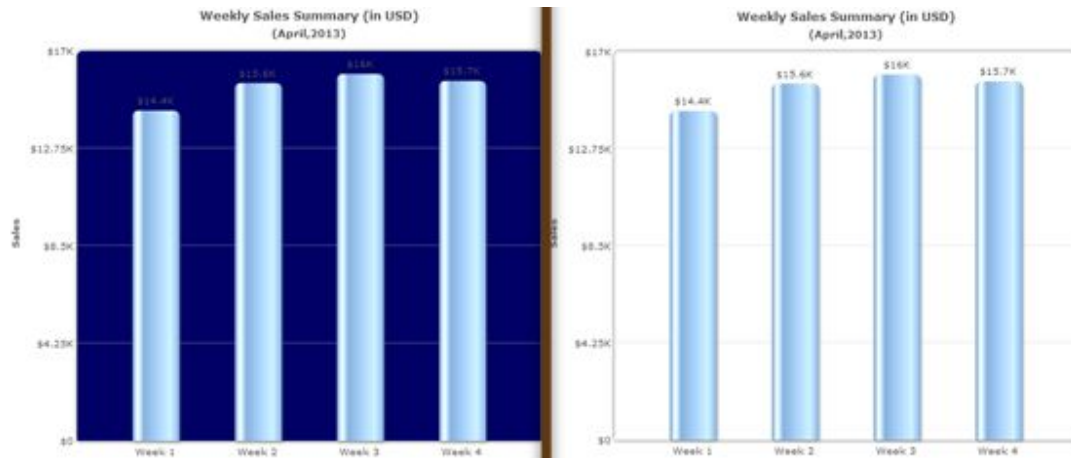
## Similarité

## Quelques principes :

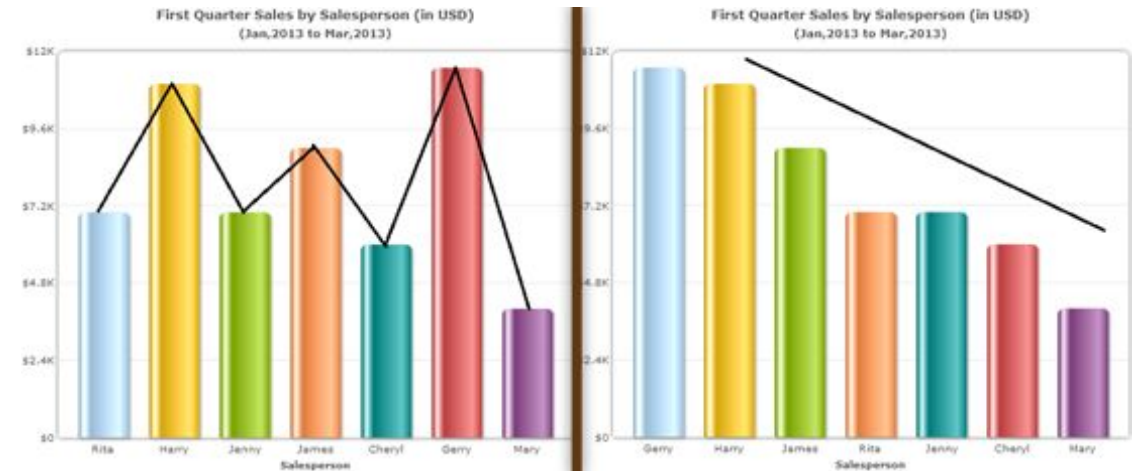
## Point Focal



## Figure – fonds



## Continuité





## II- Applications

# Mémoire sur la tarification automobile à l'aide de modèles de Machine Learning et l'apport des données télématiques

- Modèles tarifaires reflétant au mieux le risque de l'assuré
- Segmentation du portefeuille
- 2 points majeurs :
  - 1- Comparaison des modèles GLM avec modèles plus innovants (CART...)
    - nombre de sinistres
    - coût du sinistre
  - 2- L'impact de l'ajout de données télématique
- Principe de mutualisation

# Mémoire sur la tarification automobile à l'aide de modèles de Machine Learning et l'apport des données télématiques

- 
- 
- 2 points majeurs :
  - 1- Comparaison des modèles GLM avec modèles plus innovants (CART...)
    - nombre de sinistres  $\square$  GLM
    - coût du sinistre  $\square$  CART

2-

TABLE 4.28 – Comparaison des différents modèles testés




| Modèle        | Vitesse d'apprentissage | Facilité d'explication de l'algorithme | Facilité de paramétrage | Pouvoir prédictif | Interprétabilité des résultats |
|---------------|-------------------------|----------------------------------------|-------------------------|-------------------|--------------------------------|
| GLM           | +++++                   | +++++                                  | +++++                   | +++               | +++++                          |
| CART          | +++++                   | +++++                                  | ++++                    | +++               | ++++                           |
| Random Forest | ++                      | +                                      | ++                      | ++++              | +                              |
| GBM           | ++                      | +                                      | ++                      | +                 | +                              |
| XGBoost       | +                       | +                                      | +                       | +++++             | +                              |



# Mémoire sur les outils actuariels dans le reporting santé

## Objectifs:

1. Appréhender la consommation en santé des individus
2. Améliorer le reporting santé à l'aide d'outils de la data science

**Modèles comparés**  
pour l'obtention du Diplôme Universitaire d'actuariat de l'ISFA  
et l'admission à l'Institut des Actuaire

Mémoire présenté le: 03 JUIN 2018


Par : Thomas LE HO

☐ Titre **GLM**  
Reporting santé : comment le perfectionner à l'aide de la Data Science ?


Confidentialité : ☐ NON ☒ OUI (Durée : ☐ 1 an ☒ 2 ans)

☐ Les signataires s'engagent à respecter la confidentialité indiquée ci-dessus

Membre présents du jury de l'Institut des Actuaire signature Entreprise :

☐ Mme Catherine CEDON Nom : GALEA & Associés  
Mme Isabelle FRAUDOT Signature : 

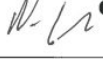

Membres présents du jury de l'ISFA

☐ M. Christian MONTROGER Nom : Léonard Fontaine  
Signature : 

Invité :  
Nom :  
Signature :

Autorisation de publication et de mise en ligne sur un site de diffusion de documents actuariels (après expiration de l'éventuel délai de confidentialité)

Signature du responsable entreprise

 **GALEA & Associés**  
25 rue de Choiseul  
75002 PARIS  
Tél. 01 43 22 11 11  
Signature du candidat  R.G.S. Paris - 492 379 839

# Poste “ Consultations Visites ”

## Modélisation

Critère: Mean Squared Error (MSE)

☐ Fréquence

☐ Dépenses moyennes

| Modèle        | Fréquence     | Dépenses moyennes |
|---------------|---------------|-------------------|
|               | MSE base test | MSE base test     |
| GLM           | 22,6          | 82,9              |
| CART          | 23,5          | 59,6              |
| Random Forest | 22,1          | 56,3              |
| XGBoost       | 23,2          | 54,5              |

## Comparaison des modèles

| Algorithme    | Explication de l'algorithme | Vitesse d'apprentissage | Interprétabilité des résultats | Pouvoir prédictif |
|---------------|-----------------------------|-------------------------|--------------------------------|-------------------|
| GLM           | ★★★                         | ★★★★★                   | ★★★★                           | ★                 |
| CART          | ★★★★                        | ★★★★★                   | ★★★★★                          | ★★                |
| Random Forest | ★★                          | ★                       | ★★                             | ★★★               |
| XGBoost       | ★                           | ★                       | ★★                             | ★★★★              |

# Conclusions

- Le modèle *GLM* est quasi-instantané en termes d'exécution
- Les arbres *CART* sont les modèles les plus rapides à mettre en place
- L'étape de calibration des modèles Random Forest est assez rapide
- Devant le nombre important de paramètres à disposition, l'étape de calibration des modèles *XG Boost* demande un temps conséquent
- L'***interprétabilité*** des résultats est également moins aisée avec des modèles agrégés puisque le résultat obtenu est une agrégation d'estimateurs et non plus un simple estimateur
- Les modèles ayant le plus grand pouvoir ***prédictif*** sont, sans surprise, les modèles agrégés

# MODELISATION DE LA PRIME PURE

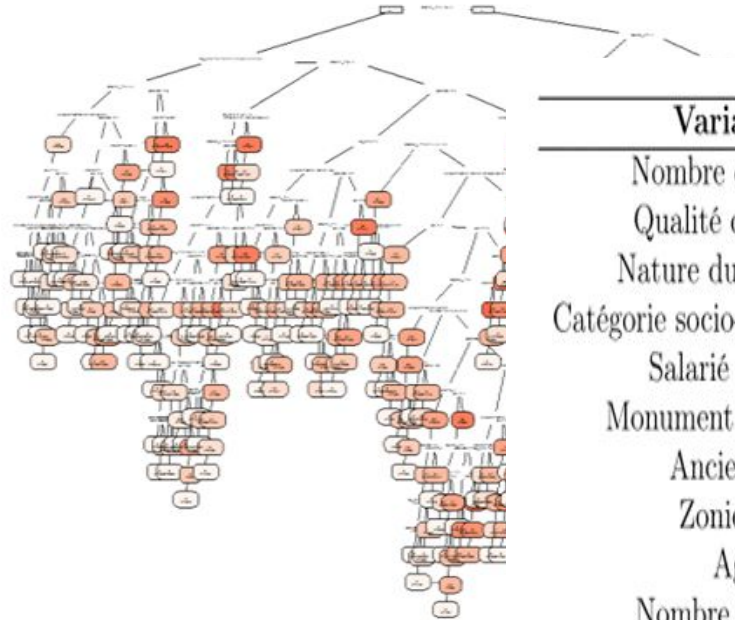




# REGRESSION TREE

| Étapes | Fonction |
|--------|----------|
| 1      | Gamme    |
| 2      | Gamme    |
| 3      | Gamme    |
| 1      | Log-norm |
| 2      | Log-norm |
| 3      | Log-norm |
| 4      | Log-norm |
| 5      | Log-norm |

Table 17 - Résum



| Variables                       | Significativités | Variables pour Tree optimal |
|---------------------------------|------------------|-----------------------------|
| Nombre de pièces                | 12.089           | Oui                         |
| Qualité occupant                | 11.057           | Oui                         |
| Nature du bâtiment              | 11.003           | Oui                         |
| Catégorie socio-Professionnelle | 10.987           | Oui                         |
| Salarié AGPM                    | 09.258           | Oui                         |
| Monument Historique             | 09.001           | Oui                         |
| Ancienneté                      | 05.002           | Oui                         |
| ZonierSin                       | 04.267           | Oui                         |
| Age                             | 01.234           | Oui                         |
| Nombre d'enfant                 | 01.003           | Non                         |
| Nombre de créanciers            | 00.937           | Non                         |
| superficie Véranda              | 00.589           | Non                         |
| Situation matrimoniale          | 00.256           | Non                         |
| Nombre de Chien                 | 00.135           | Non                         |

Table 19 - Tableau donnant la significativité des variables et celles choisies.

FIGURE 28 – Arbre de décision du coût de sinistre (INCENDIE).

**Diagramme des deviances des modèles pour le coût du sinistre**

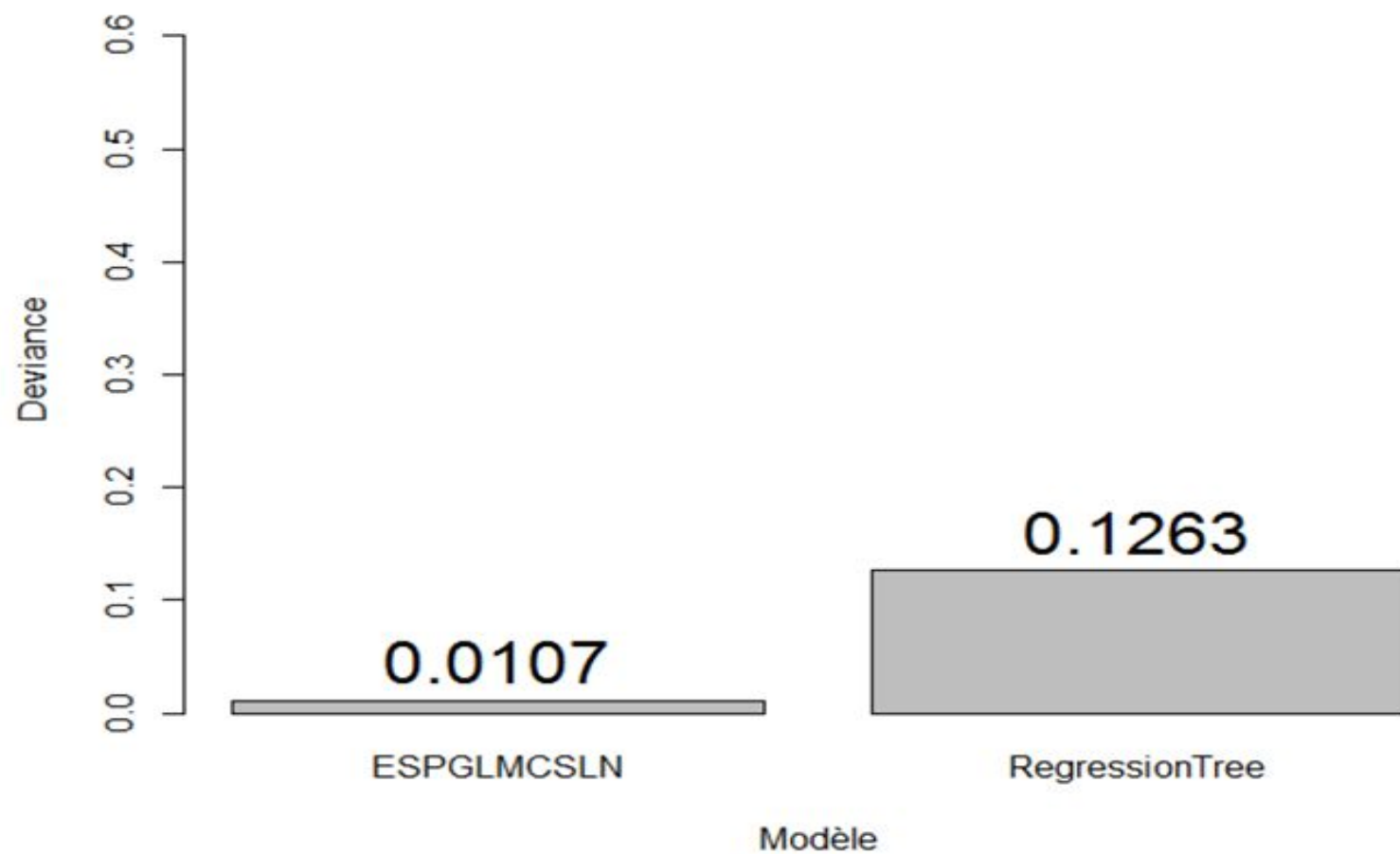


FIGURE 17 – Déviance des modèles.

# REGRESSION TRUE

| Étapes                                                                          | Fonctions                       | variables non significatives |                             | variables significatives | AIC    |
|---------------------------------------------------------------------------------|---------------------------------|------------------------------|-----------------------------|--------------------------|--------|
| 1                                                                               | Variables                       | Significativités             | Variables pour Tree optimal |                          | 238793 |
| 2                                                                               | Nature du bâtiment              | 15.02                        | Oui                         |                          | 229827 |
| 3                                                                               | Qualité occupant                | 14.001                       | Oui                         |                          | 218918 |
| 4                                                                               | Nombre de pièces                | 12.723                       | Oui                         |                          | 200078 |
| 5                                                                               | ZonierSin                       | 10.023                       | Oui                         |                          | 198791 |
| 1                                                                               | Nombre d'enfant                 | 07.280                       | Oui                         |                          | 287989 |
| 2                                                                               | Situation matrimoniale          | 07.011                       | Oui                         |                          | 278796 |
| 3                                                                               | Ancienneté                      | 04.089                       | Oui                         |                          | 268789 |
| 4                                                                               | Catégorie socio-Professionnelle | 03.007                       | Oui                         |                          | 250897 |
| 5                                                                               | Age                             | 01.234                       | Oui                         |                          | 210585 |
| 6                                                                               | Salarié AGPM                    | 01.113                       | Non                         |                          | 205020 |
| 7                                                                               | Nombre de créanciers            | 00.998                       | Non                         |                          | 195305 |
| 8                                                                               | superficie Véranda              | 00.500                       | Non                         |                          | 189500 |
| 9                                                                               | Monument Historique             | 00.100                       | Non                         |                          | 187446 |
|                                                                                 | Nombre de Chien                 | 00.001                       | Non                         |                          |        |
| Table 24 - Tableau donnant la significativité des variables et celles choisies. |                                 |                              |                             |                          |        |

Table 22 - Résumé des itérations du GLM pour aboutir au meilleur modèle .

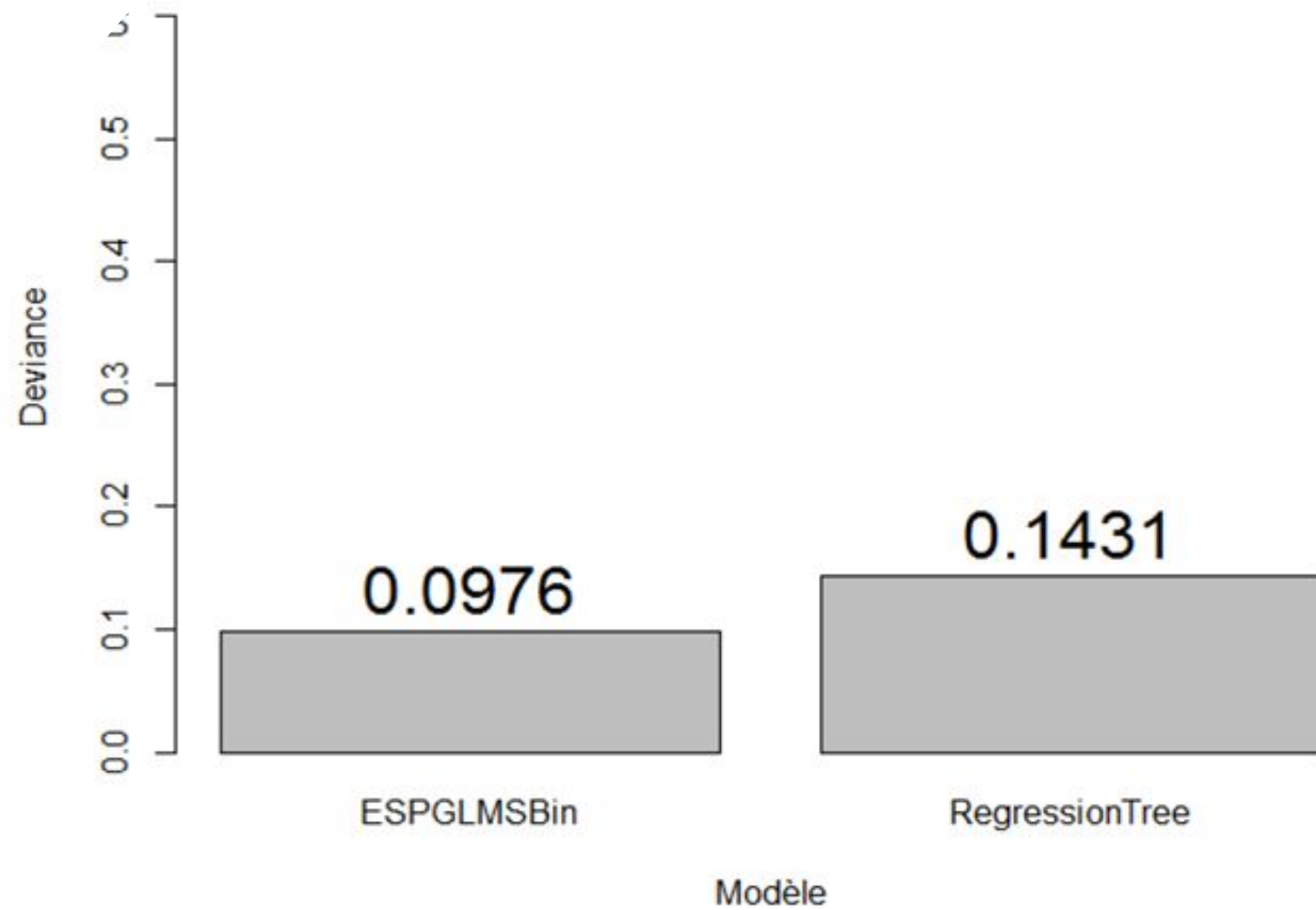


FIGURE 19 – Déviance des modèles.



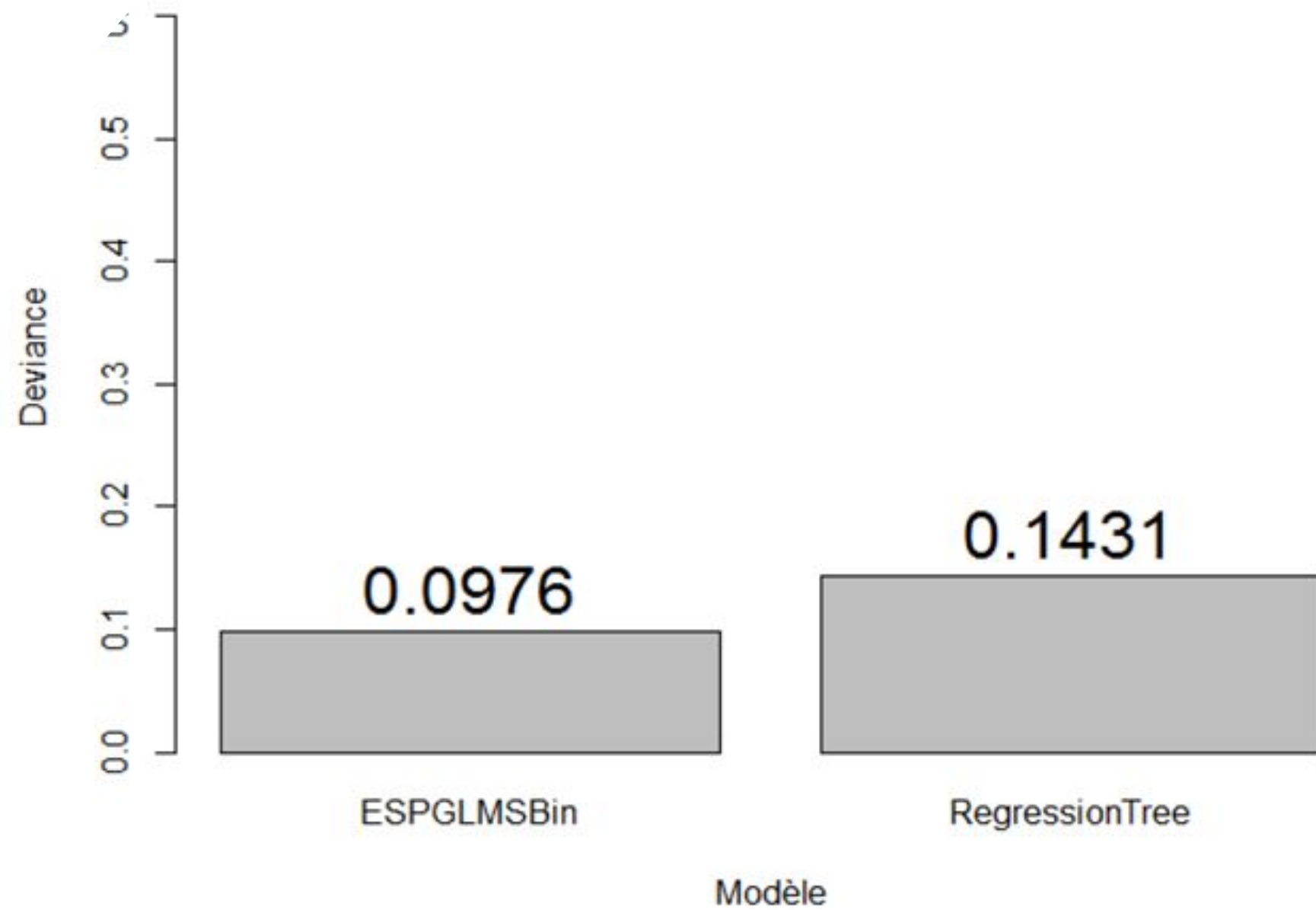


FIGURE 19 – Déviance des modèles.



## III- Conclusion