

# Calcul de score d'attrition

Prédiction de la probabilité de  
résiliation des contrats d'assurances  
automobiles

Andres SOTO

Yves PECHINE  
Pierre BLOCHE  
Guillaume LEORAT

# Démarche de l'étude



## **Analyse exploratoire**

Exploration des  
données et  
statistiques  
descriptives

## **Traitement des données**

Préparation des  
données pour la  
modélisation

## **Modélisation**

Application de  
modèles de  
prédiction

## **Evaluation**

Comparaison  
des modèles à  
travers  
différentes  
métriques

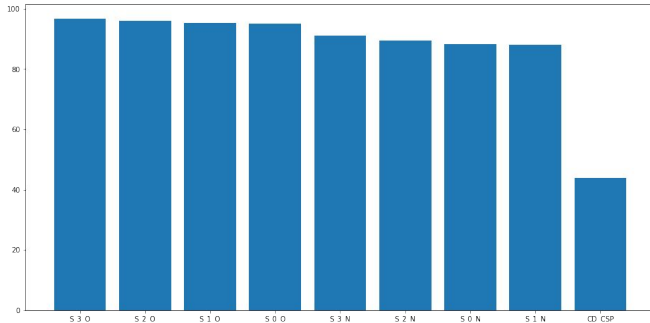
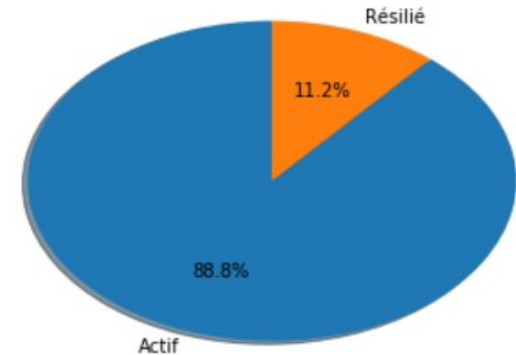
## **Modèle retenu**

Choix du  
modèle retenu  
ayant les  
meilleures  
performances  
de prédiction

# Exploration des données

→ Période d'étude : Janvier 1901 à Septembre 2000

→ Échantillon très déséquilibré : 88,8% de contrats d'assurances actifs vs 11,2% de contrats d'assurances résiliés



- Présence de variables avec des taux élevés de valeurs manquantes
- Présence de corrélations entre les variables



# Traitement des données

- **Création de variables** à partir de différentes dates
- **Remplacement des valeurs manquantes** par la méthode des KNN
- **Sélection de variables** pertinentes par test de Khi2 et de Student
- **Découpage du jeu de données** en échantillon d'apprentissage et en échantillon test avec la même répartition des contrats d'assurances résiliés ou non

Les 5 variables choisies :
Le nombre de contrats résiliés
L'âge du client
L'ancienneté du client
Le nombre de contrats actifs
Le montant annuel de la prime



# Modélisation et Évaluation

**Régression  
Logistique**

**Arbre de  
décision**

**Random  
Forest**

**Bagging**

**AdaBoost**

**XGBoost**

AUC : 0,90  
Log-loss : 0.26

AUC : 0,93  
Log-loss : 4.04

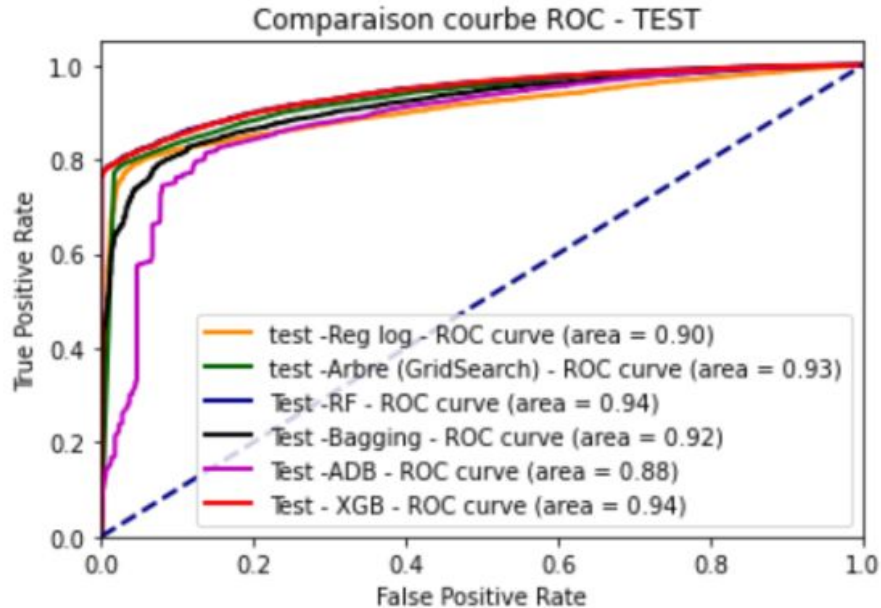
AUC : 0,94  
Log-loss : 0.176

AUC : 0,92  
Log-loss : 0.27

AUC : 0,88  
Log-loss : 0.54

AUC : 0,94  
Log-loss : 0.174

# Modèle retenu : XGBOOST



- Le XGBoost présente les meilleures performances.
- Avec un AUC de 0.94 et une log-loss de 0.17, il permet de bien détecter l'attrition des clients sur les contrats d'assurance auto.
- l'algorithme est stable car les performances sont bonnes sur l'échantillon de test.