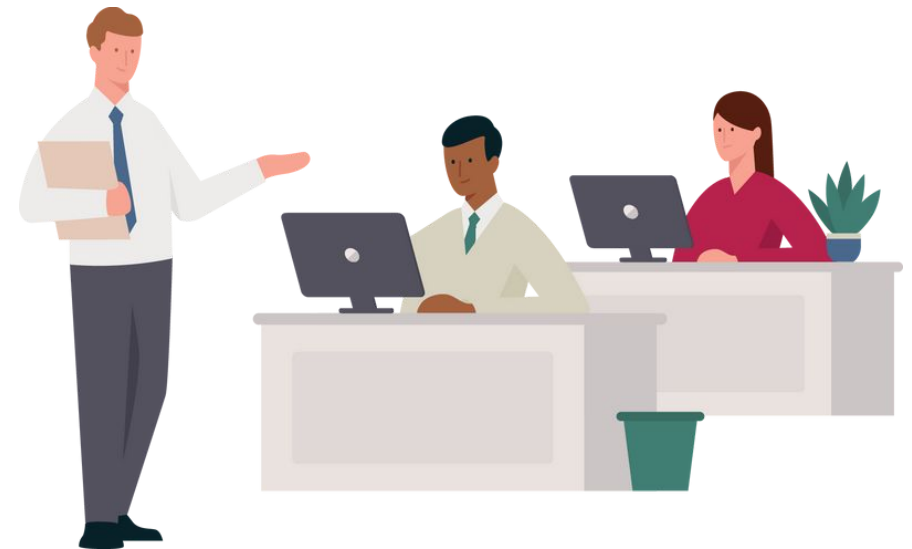
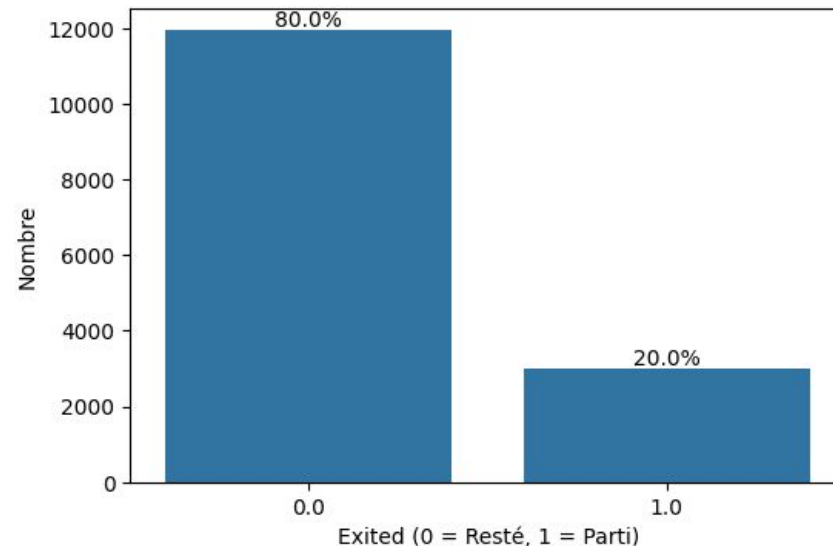


MOSEF BANK CHURN PREDICTION



Stratégie d'identification et de prospection des clients susceptibles de partir

- Prospects ciblés : clients bancaires
- Variable cible : Le client est pré disposé à se désengager. C'est une variable latente (inobservable) que l'on peut inférer par : le client est parti



- Objectif : prédire le taux de désabonnement des clients dans le secteur bancaire.

La base d'apprentissage, ou comment comprendre la vie des clients.

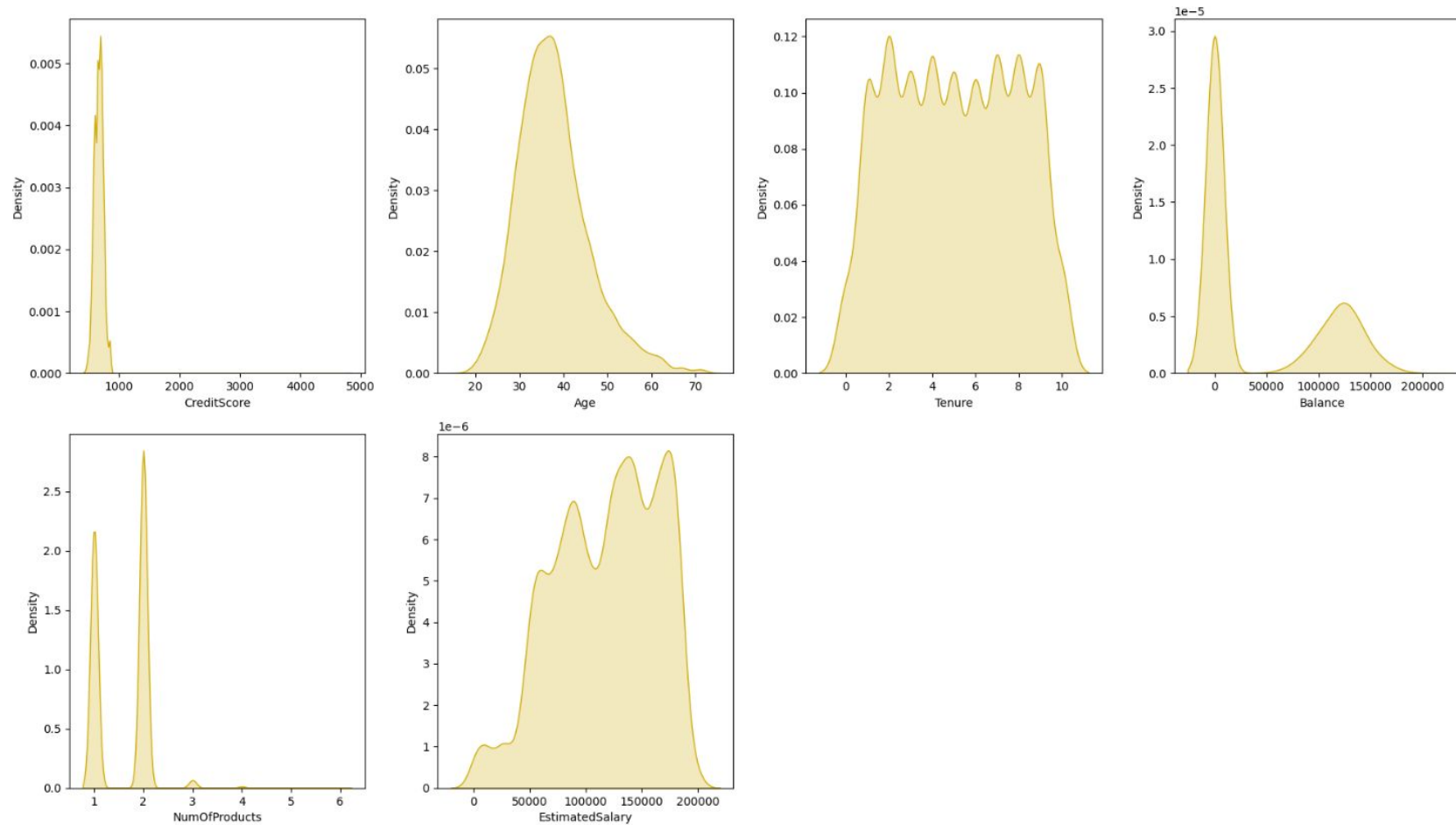
Base apprentissage : une ligne par client, associé à une année

- Variables socio-demographiques du client (nom, localisation, sexe, âge, salaire)
- Variables bancaires du client (solde, nombre de produits bancaires utilisés, s'il possède une carte de crédit...)
- Variable d'activité du client
- La cible est égale à 1 si l'ex client est parti l'année considérée

Pour les doublons, nous gardons l'observation avec l'âge le plus élevé afin d'avoir l'observation la plus récente. Pour les doublons ayant les mêmes "CustomerID", "Surname" et "Geography", nous conservons les deux observations, même si cela peut introduire un biais, car il n'est pas possible de définir une règle de dédoublonnage pertinente dans ce cas.

Des clients en moyenne adulte, au profil bancaire hétérogène

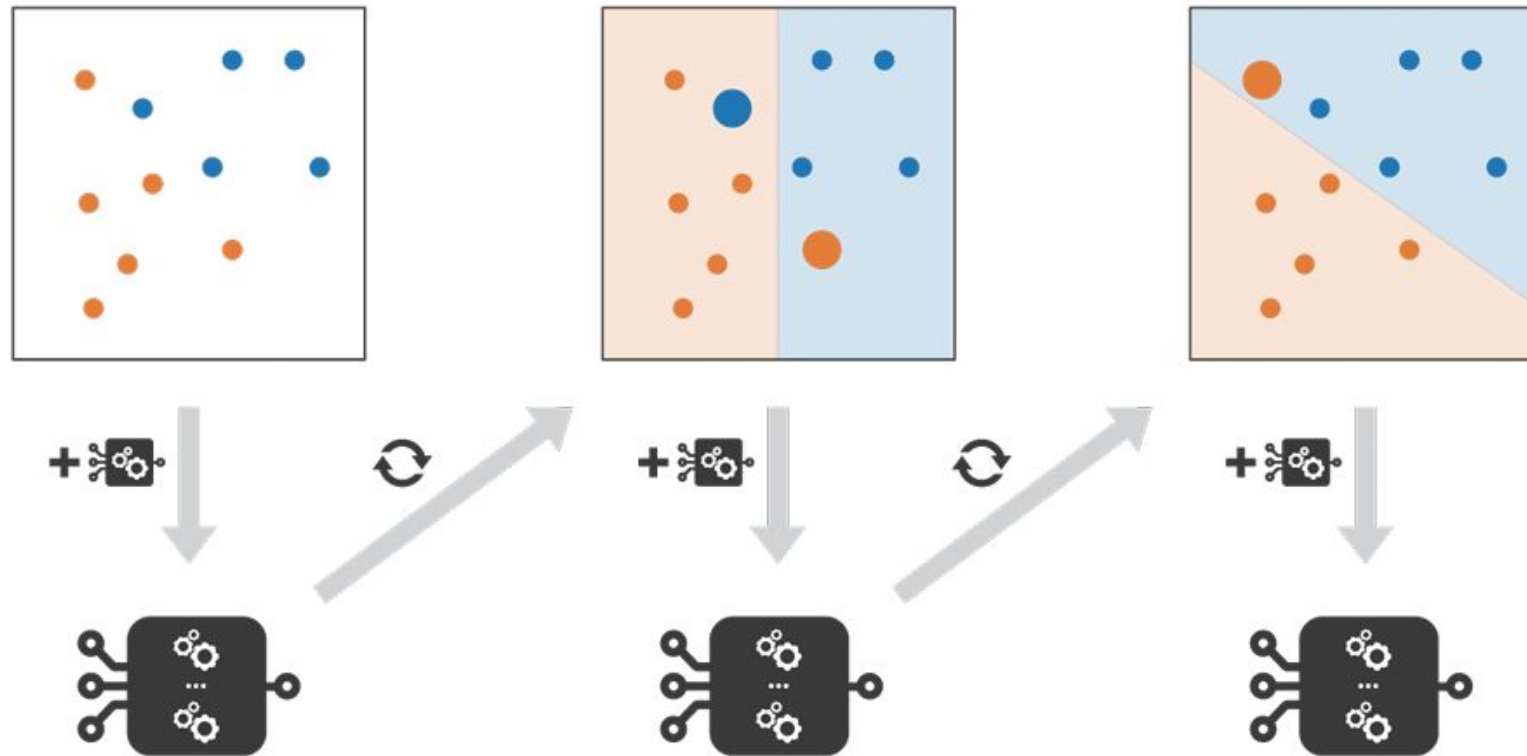
data mining



Distribution des variables des clients

Un modèle sans biais et capable de généraliser

data mining



Grâce au boosting, l'entraînement du modèle à une étape donnée dépend des modèles ajustés aux étapes précédentes. Permet de diminuer le biais.



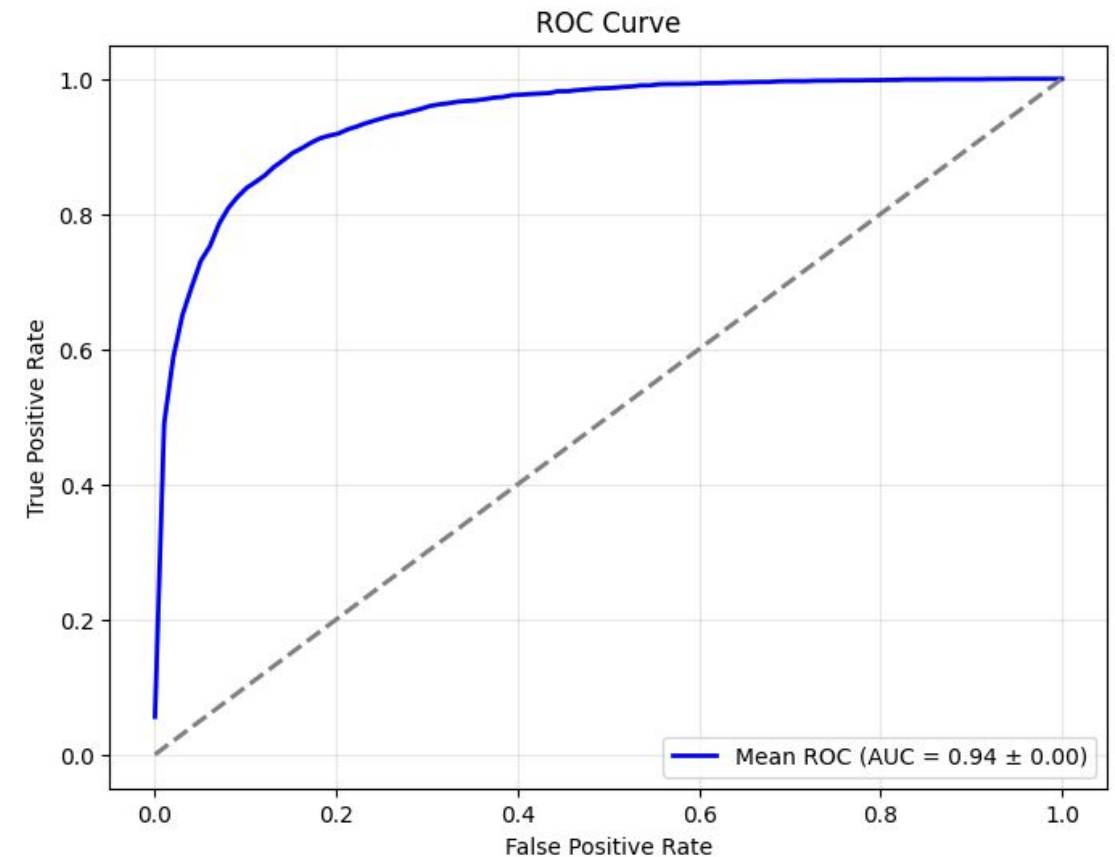
Stratégie de validation pour améliorer la généralisation

Le modèle est-il capable de discriminer les clients susceptibles de quitter la banque de ceux qui restent ?

Pour avoir une liste de clients que la banque va potentiellement perdre, nous avons entraîné un modèle de classification binaire Gradient boosting basé sur un algorithme de boosting*

Il y a une probabilité de 0,94 que, parmi un client effectivement « perdu » et un client non « perdu » choisis au hasard, la probabilité estimée d'être « perdu » soit plus élevée pour le client effectivement « perdu » que pour l'autre.

*nous avons également entraîné un modèle de stacking avec un gradient boosting et un régression pénalisé l2.

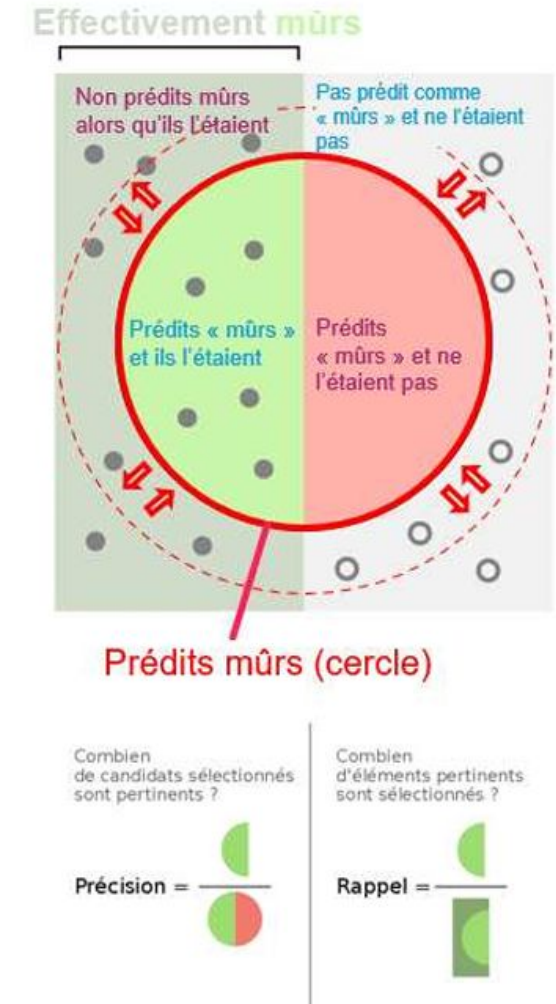
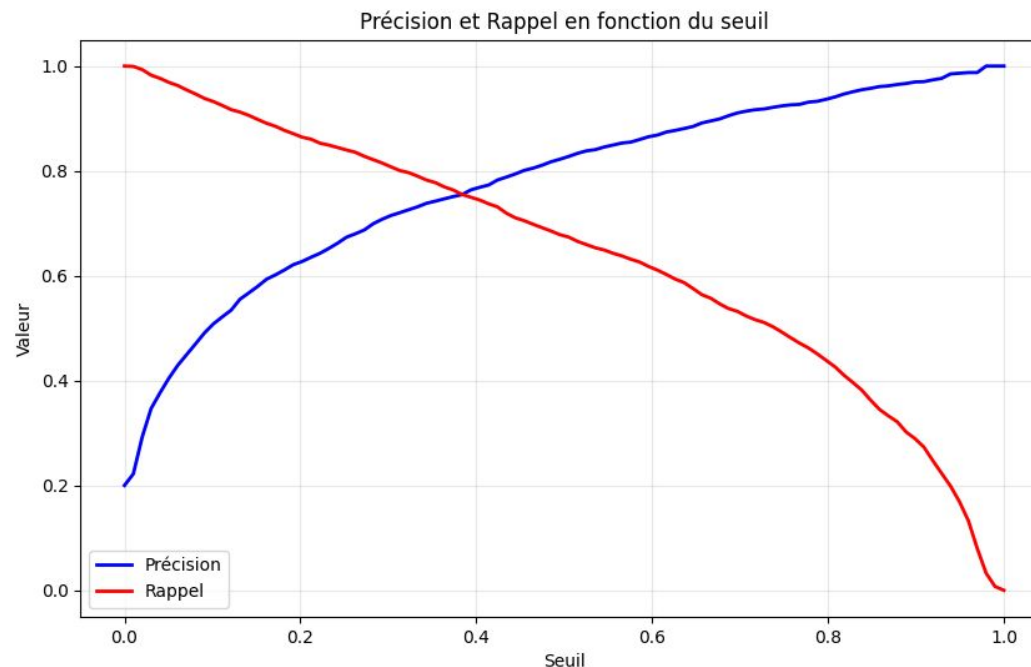


Au plus on détecte des clients « perdus », **au plus on accepte de se tromper.**

data mining

Pour un seuil de 0.4, les métriques d'évaluations sont :

- 0.75 de précision moyenne : sur 100 clients prédits « mûrs », 75 l'étaient effectivement.
- 0.75 de rappel moyen: parmi les clients qui étaient effectivement mûrs, 75 % ont été prédit comme tel. Ce qui implique une grosse volumétrie des clients prédits.



Un sauvetage potentiel de 80% des clients perdus, si l'on accepte de se tromper deux fois sur dix en essayant de garder les 2985 clients perdus

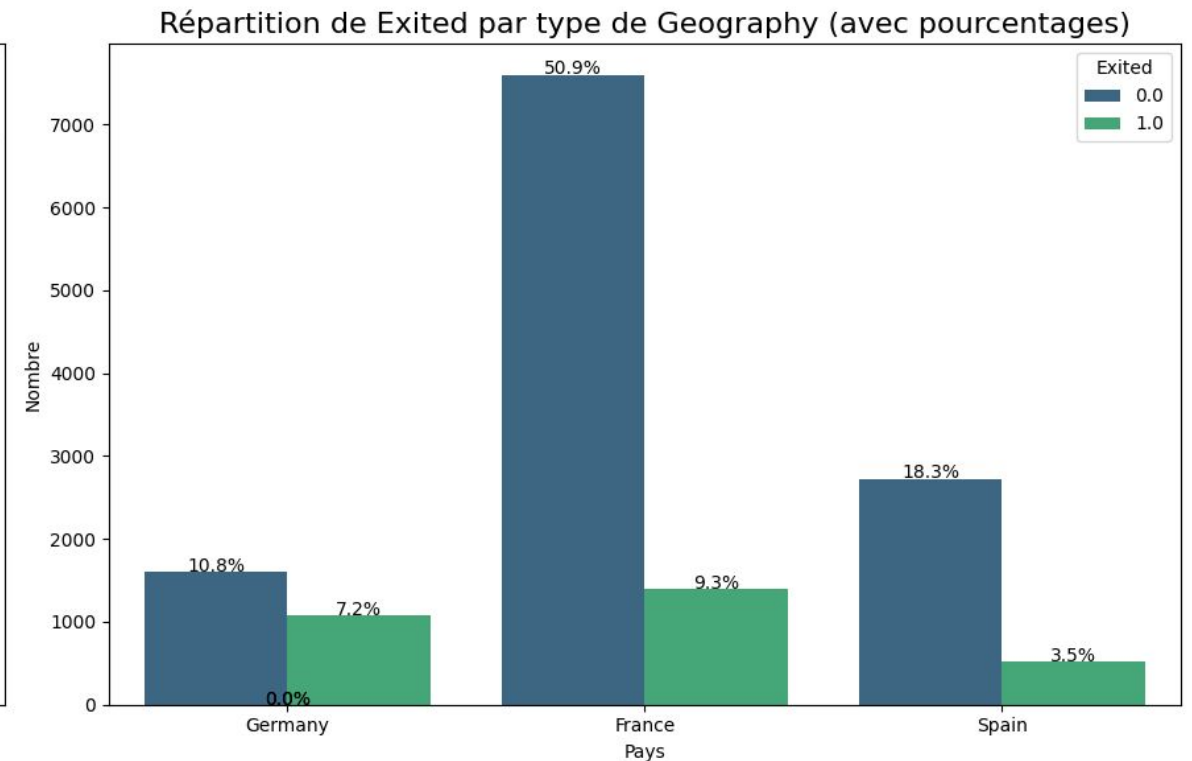
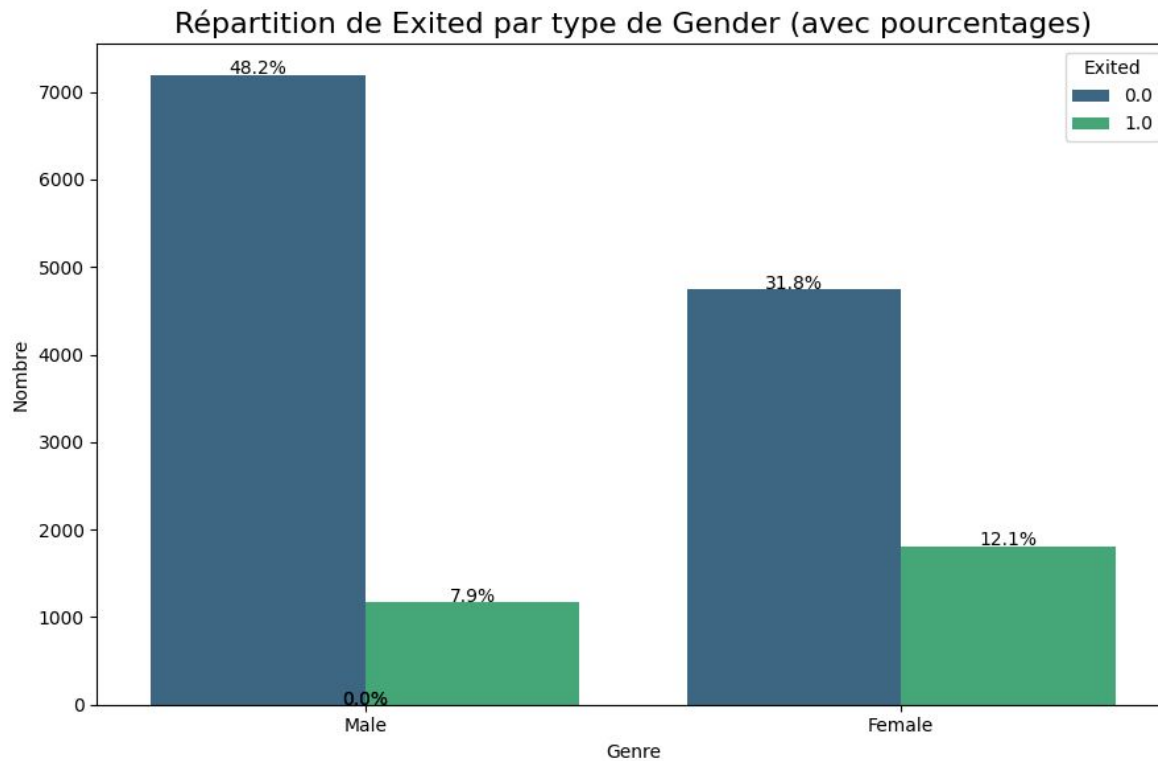
data mining

Choix d'un seuil en fonction de la réalité métier/ business

- Arbitrage : au plus on détecte des clients que la banque va perdre (hausse de recall), au plus on accepte de se tromper (baisse précision).
- Détecter qu'un client va partir engendre des coûts pour le conserver.
- Il faut le comparer au coût d'acquisition d'un client.

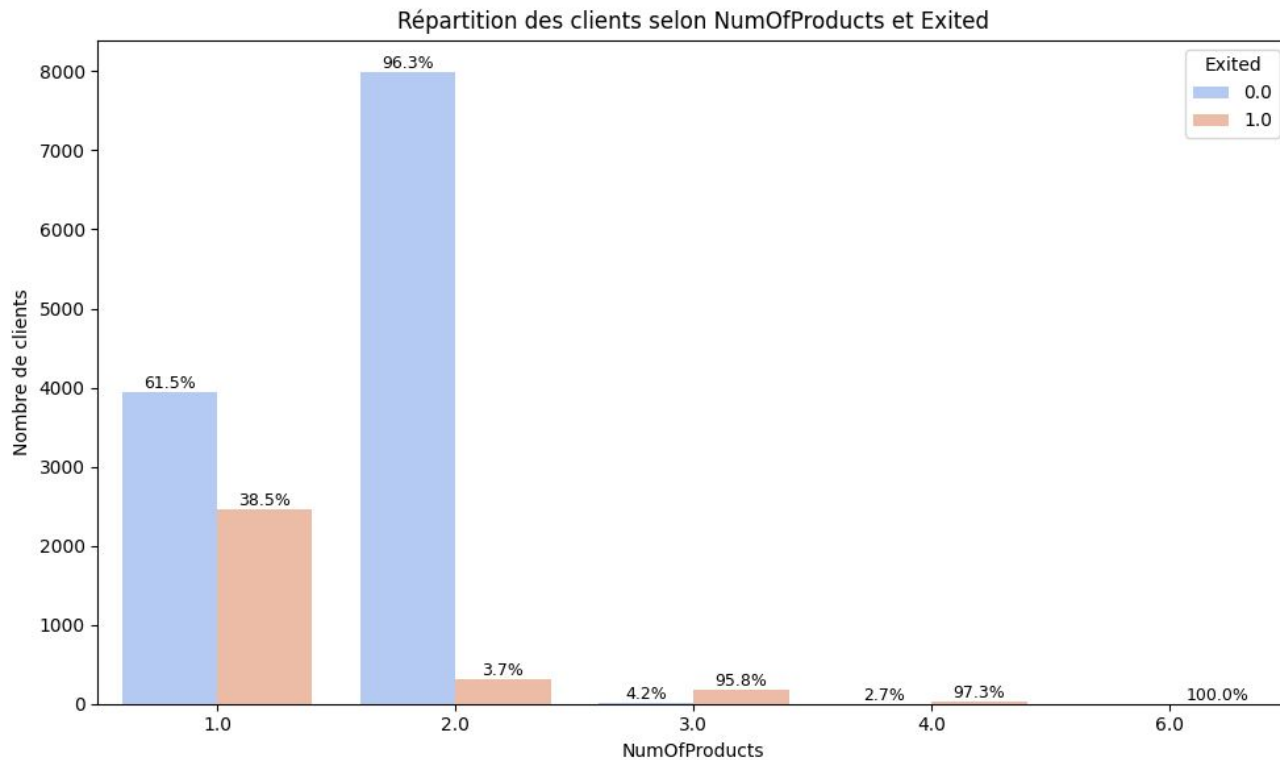
Nombre de clients prédits	Précision
2903	77 %
2447	82 %
2117	87 %
1729	91 %
1387	94 %
886	97 %

Les clients « perdus » sont majoritairement des femmes et des français sur l'ensemble de la population. data mining



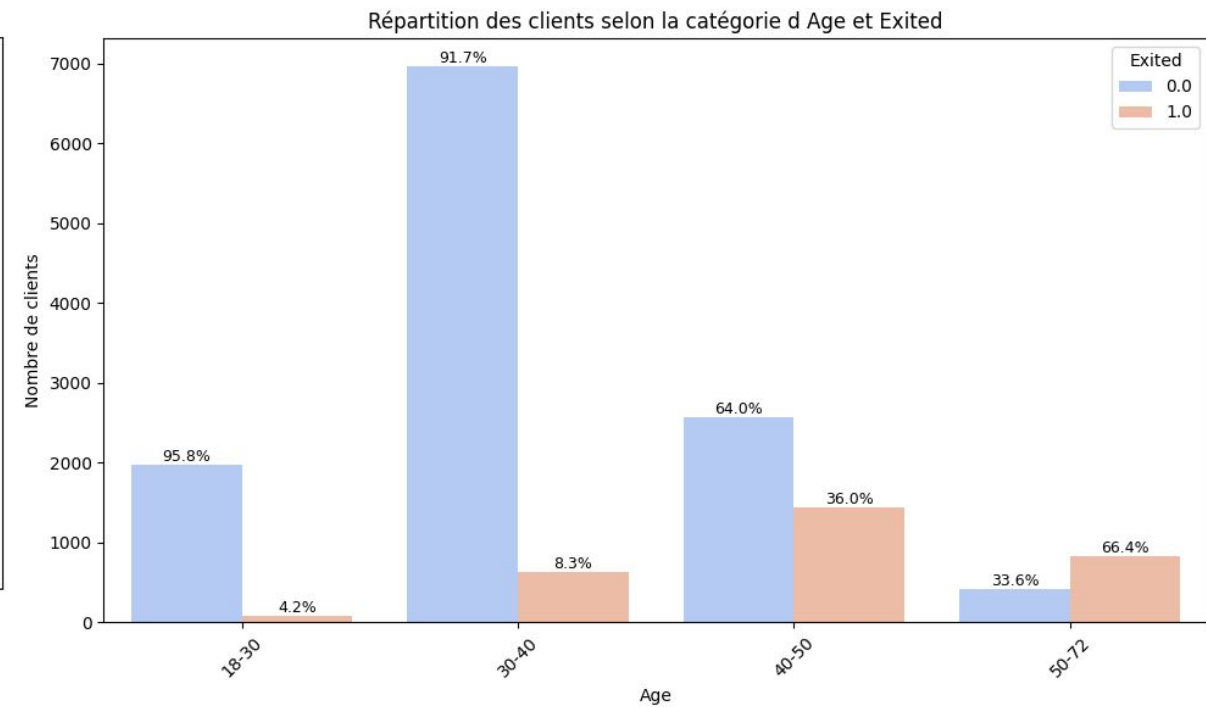
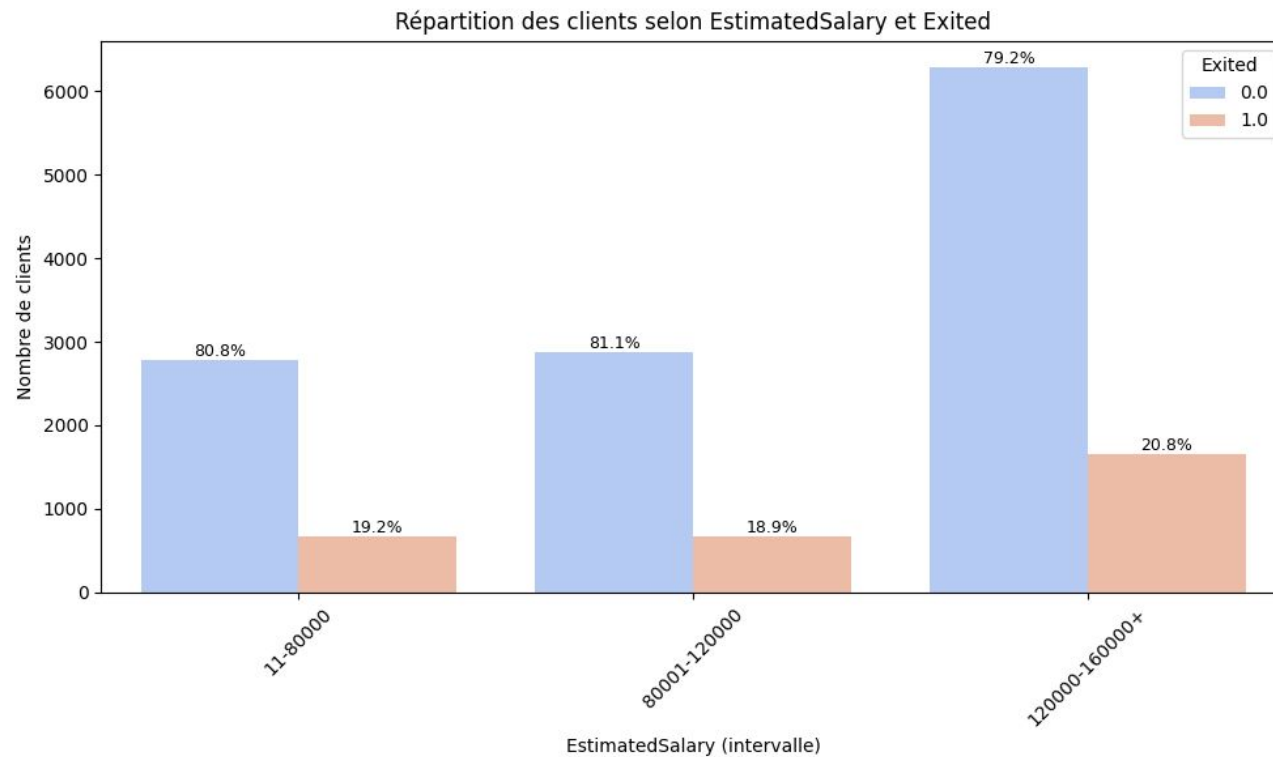
Les clients « perdus » possèdent relativement soit 1 produit soit + de 3, sont relativement peu anciens

data mining

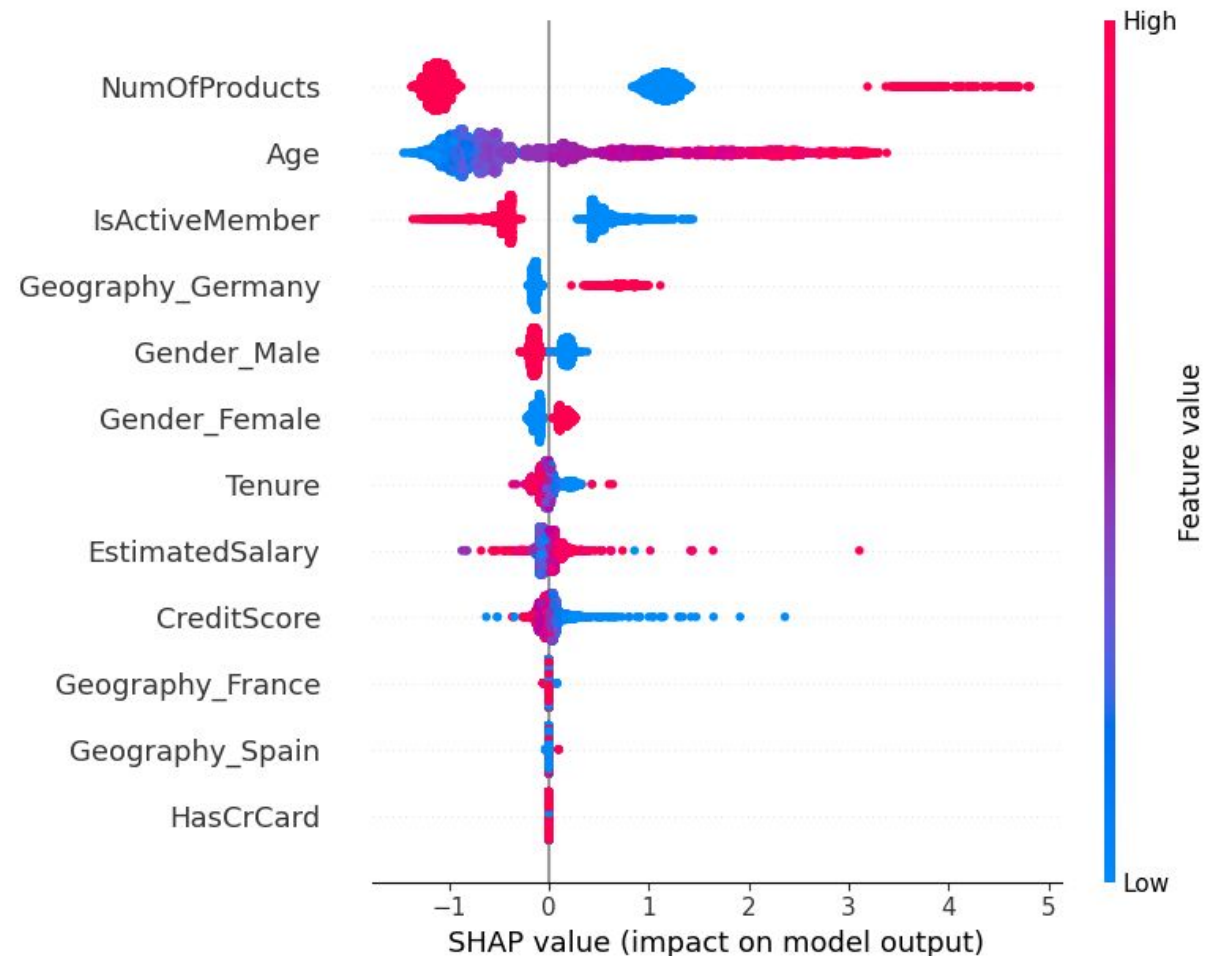


Les clients « perdus » ont relativement des meilleurs salaires et ont entre 40 et 50 ans

data mining



Interprétabilité : les variables qui ont le plus d'importance



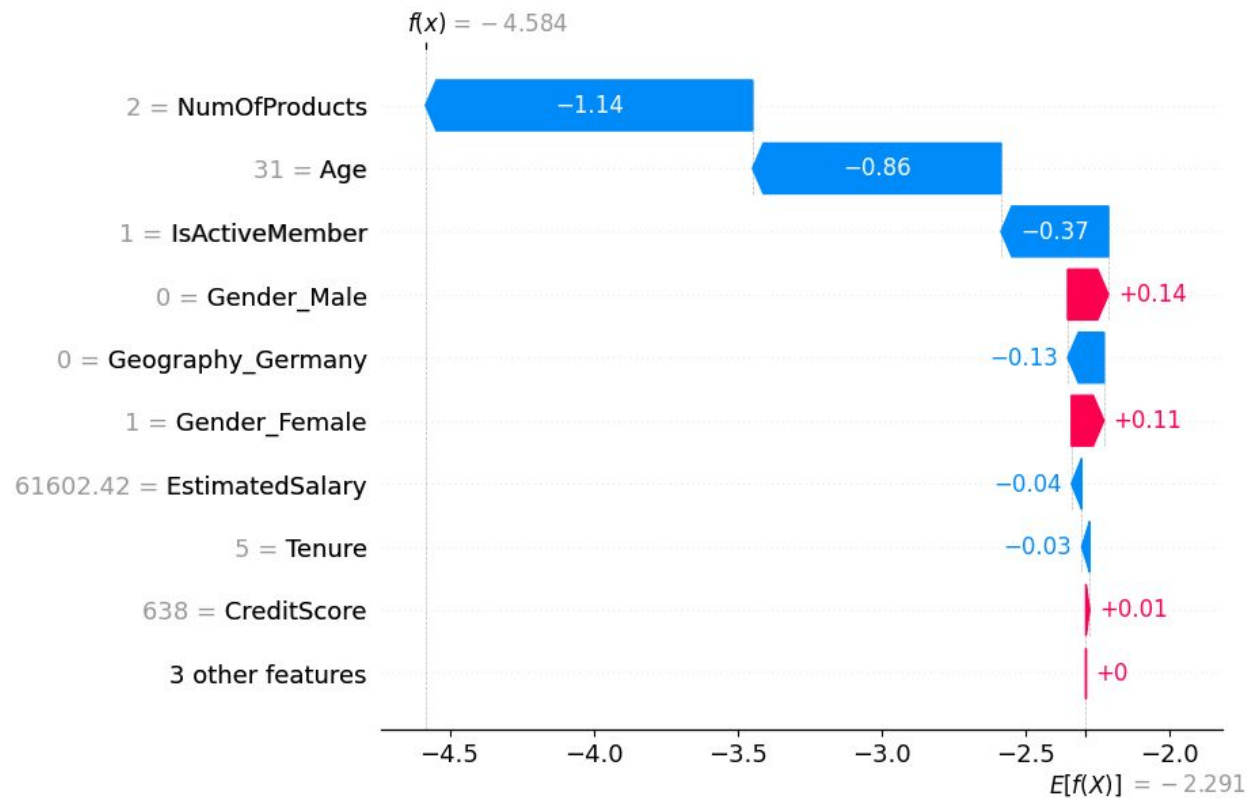
L'axe des ordonnées : liste les caractéristiques les plus importantes en haut, par importance moyenne (valeur absolue de SHAP) sur l'ensemble de données.

L'axe des x : les valeurs SHAP à droite de 0 augmentent la “probabilité” de la cible = 1, tandis que celles à gauche de 0 la diminuent.

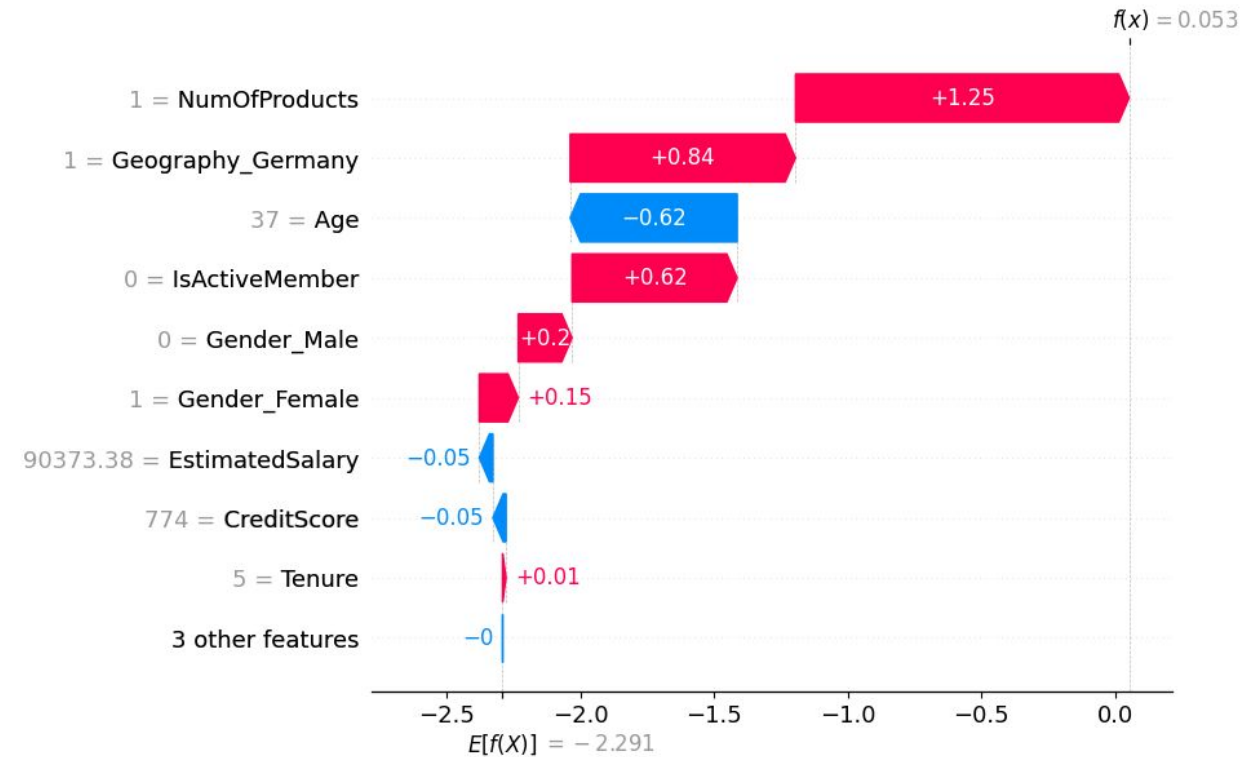
Exemple de deux profils type

data mining

Le client qui reste



Le client qui part



Prochaines étapes et améliorations :

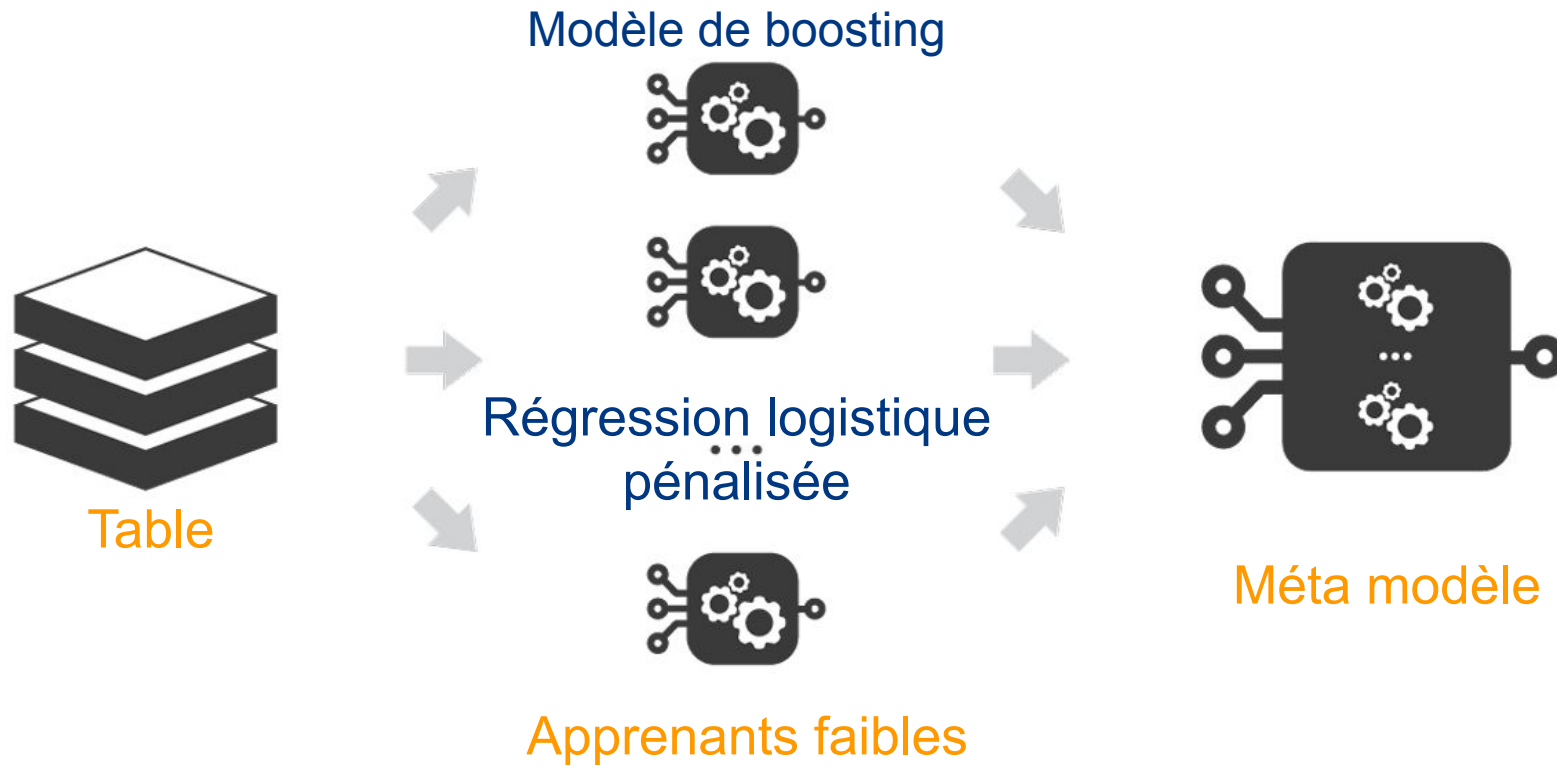
1

- Enrichir la base d'apprentissage
- Affiner et tester d'autres modèles

2

- Discuter avec la distribution des modalités d'intégration du score de churn
- Partage d'un premier fichier avec le score de churn.

Annexe: le stacking de modèle pour éviter le sur-apprentissage



Les apprenants sont non homogènes...contrairement
au bagging et boosting