

# DÉCOUVREZ COMMENT OPTIMISER LA RÉTENTION CLIENT GRÂCE AU SCORE DE CHURN



Dans le paysage concurrentiel complexe d'aujourd'hui, l'attrition client devient omniprésente. Cet évènement, caractérisé par la rupture d'une relation entre l'entreprise et le client, transcende tous les secteurs d'activités. L'attrition client est au cœur de la stratégie de croissance des entreprises, imposant aux équipes le défi de développer des stratégies efficaces et efficientes pour réduire ce phénomène.

Cet article explore de manière détaillée la méthodologie, les défis et avantages liés à la mise en place d'un score de churn. L'article met en lumière la façon dont les données client peuvent être exploitées pour prédire et comprendre le churn afin d'optimiser la rétention client et d'augmenter la valeur vie client.

## Méthodologie

Le churn repose sur une analyse des comportements du client sur les données historiques. Il existe deux types de score d'attrition :

- **Les clients avec engagement** : un client peut être engagé sur une relation contractuelle mensuelle ou de longue durée, définie par un paiement périodique. C'est le cas par exemple pour la souscription à un produit de banque ou assurance, l'abonnement à internet/téléphonie, les fournisseurs d'énergie, le streaming ...
- **Les clients sans engagement** : un client peut aussi être lié à un service sans contrat. Dans ce cas le churn est non contractuel. Ce type d'attrition peut être observé dans le Retail et l'E-commerce.

Dans le cas d'une relation contractuelle, nous utilisons des modèles de **machine learning supervisé**, en l'occurrence des modèles de classification. En revanche, pour les relations non contractuelles, nous avons **recours aux lois de probabilités** en particulier la loi beta-géométrique et ses variantes car n'y a pas une cible (target) spécifique à prédire.

Dans cet article, nous nous focaliserons sur le **churn dans le cadre d'une relation contractuelle**.

Pour modéliser l'attrition et en comprendre les motifs, notre approche d'analyse exploratoire repose sur une méthodologie assez classique dans le cadre d'un projet de machine learning. Nous définissons une **période d'analyse avec les métiers** et collectons les données historiques des clients. Nous nettoyons les données et remplaçons les **valeurs manquantes de manière adéquate**. Il est important d'examiner le **taux de churn par période** (mensuelles, trimestrielles, annuelles, ...). Cette approche nous permet d'identifier les tendances, les facteurs de churn et de formuler des recommandations stratégiques pour améliorer la rétention client.

## 1 DÉFINITION DE LA PÉRIODE D'ÉTUDE ET COLLECTE DES DONNÉES

Il est crucial de définir la période d'étude avec les **métiers et de recueillir les règles métiers liées au churn**. Les données historiques sont **collectées, nettoyées** et mise au **standard** des algorithmes de machine learning. Ces données peuvent inclure des **informations sur les clients**, telles que la durée du contrat, les transactions, les caractéristiques du produit, les retours d'une enquête de satisfaction client.

## 2 ANALYSE DU TAUX DE CHURN PAR COHORTE

On effectue une analyse approfondie du taux de churn au fil du temps par **cohorte de client**. Une cohorte est un regroupement de client partageant des caractéristiques similaires telle que la date d'acquisition. L'analyse permet d'identifier les périodes de **pic** ou de **creux dans le taux d'attrition**.

## 3 RECOMMANDATION ET SÉLECTIONS DES VARIABLES POUR LE MODÈLE

En fonction des résultats de l'analyse, nous sélectionnons **les variables les plus discriminantes et nous ajustons les poids d'un modèle sur les données**.

En résumé, cette analyse implique une approche méthodique pour comprendre les phénomènes d'attrition, identifier les facteurs influents et proposer un modèle de prédiction.

Dès lors, nous cherchons à prédire l'attrition avec un modèle de machine learning et comprendre le modèle. Ceci nous amène à répondre à la question « Pourquoi un client est plus à même de faire une attrition », nous le faisons à travers des méthodes d'explicabilités de modèle.

Des algorithmes de classification existent pour la prédiction de ce phénomène tels que **Random Forest, Gradient Boosting, Xgboost, Catboost**.

- L'expert en apprentissage automatique ajuste les paramètres du modèle de manière à ce que la précision entre les valeurs prédites et les valeurs réelles, soit la plus grande possible. On peut utiliser le f1-score comme métrique d'optimisation.
- Pour évaluer notre modèle, nous utilisons une technique appelée **validation croisée**. Cela consiste à diviser notre ensemble de données en plusieurs sous-ensembles distincts, plus généralement k ensembles (par exemple k=5). Nous itérons ensuite sur ces k ensembles, entraînons le modèle sur k-1 d'entre eux et testons le modèle sur le reste à chaque itération. Nous décalons d'un ensemble à chaque étape.
- A la fin des itérations, nous calculons la moyenne des scores obtenus. Cela nous permet d'évaluer la stabilité du modèle. Enfin, nous calculons le score sur l'ensemble des données de test pour mesurer la capacité du modèle à généraliser.

En plus de cela, nous essayons de comprendre le comportement des churners à travers une explicabilité du modèle. Nous identifions donc les variables les plus influentes du modèle. Ces facteurs peuvent être :

- Le produit/service ne satisfait pas ou plus le client
- Le produit/service ne lui est pas ou plus utile
- Le besoin du client de faire des économies
- Le client a trouvé mieux ailleurs (meilleur rapport qualité/prix, campagne promotionnelle d'un concurrent ...)

Dans tous les cas, il faut mettre en place des actions marketing en rapport aux facteurs de churn pour favoriser la rétention.

## Principales difficultés et solutions

L'une des principales difficultés, comme dans tous les projets d'analyse de données est la qualité des données. Il faut également qu'elles soient aussi fiables et complètes que possible.

- Avant l'analyse, on effectue un nettoyage des données où on retire les valeurs aberrantes notamment ; on peut aussi compléter les données manquantes avec des méthodes d'imputation.

Ensuite, comme la plupart des phénomènes d'attrition, le churn est un évènement rare, ce qui donne des données déséquilibrées dans la mesure où le taux de churn est souvent faible.

- Pour pallier à cela, nous utilisons des méthodes dites de rééchantillonnage (OverSampling, UnderSampling) ; la méthode la plus répandue et que nous préconisons est le **SMOTE** pour équilibrer avec des données synthétiques.

## Bénéfices

- Anticiper le phénomène d'attritions afin de garder les clients car l'acquisition est plus chère que la rétention.

On peut également en profiter pour tester l'appétence client à travers **des actions marketing menées pour la rétention**.

- Le phénomène de churn, nous offre une compréhension du comportement de nos clients et de notre compétitivité face à la concurrence.

Ce qui permet de **développer une stratégie marketing cohérente**.

- Enfin, vous pouvez estimer la valeur vie client et l'intégrer dans le chiffre d'affaires, optimiser la satisfaction client et ainsi vendre des produits auxquels les clients sont plus appétents.



Si vous souhaitez approfondir le sujet de scoring de churn ou en savoir plus sur l'expertise de notre Data Scientist, Mamadou, n'hésitez pas à nous contacter : [contact@aid.fr](mailto:contact@aid.fr)