Projet de Prédiction du Churn -SyriaTel

Département Ventes & Marketing

Présentation nontechnique Auteur:
Obed
Antoine

Département : Ventes & Logistique

> Date: 14 Septembre 2025

Contexté et problématique

Entreprise: SyriaTel (télécommunications).

Problème : Forte perte de clients (churn).

Enjeu : Identifier les clients à risque pour cibler les actions de fidélisation.

Objectif : Construire un modèle de classification binaire :

 $1 \text{ (True)} \rightarrow \text{Client risque de partir.}$

0 (False) → Client reste fidèle.

Définitions clés

- •Churn (attrition client) : départ d'un client vers un concurrent.
- Feature (variable explicative) : caractéristique d'un client utilisée par le modèle (ex. nombre d'appels au service client)
- •Stakeholder (partie prenante) : acteurs concernés par les résultats
- (ici : département ventes & marketing de SyriaTel).

Les modèles de Machine Learning testés

•Régression Logistique

•Modèle simple, interprétable, utile comme benchmark.

•Random Forest

- •Combine plusieurs arbres de décision.
- •Robuste, bonne précision, mais moins interprétable.

Gradient Boosting

- •Combine plusieurs modèles faibles pour améliorer la performance.
- •Très efficace pour détecter les signaux faibles.

Dataset utilisé

- •Source : historique de SyriaTel.
- •Colonnes principales : minutes d'appel, facturation, plan international, appels SAV, messagerie vocale, etc.
- •Variable cible : *Churn* (True = client parti, False = client resté).
- •Exemple simplifié:

Client A : 7 appels SAV + facture élevée → a churné (True).

Client B : usage modéré + peu d'appels SAV → est resté (False).

Comment fonctionne notre travail (explication simple du modèle)

- •On utilise les données historiques comme vérité terrain (ground truth).
- •Le modèle apprend les **liens statistiques** entre comportements (features) et résultat (churn). C'est a dire, il regarde :
- "Si un client appelle le SAV plus de 5 fois → probabilité de churn élevée (~70%)."
- "Si un client dépense > 50\$/jour → probabilité de churn élevée (~65%)."

Ces informations servent au modèle pour apprendre : "Quand les clients ont tel type de comportement, ils risquent de quitter."

Les modèles testés cherchent automatiquement ces **règles implicites** et calcul des relations statistiques entre les comportements et la finalité du choix des clients (rester ou partir)

Le modèle étant finalisé, on peut appliquer ce modèle sur les clients actuels → il prédit lesquels sont à risque, avant qu'ils ne partent.

- Accuracy (Précision globale)
- Définition : proportion de prédictions correctes (clients restés + clients partis) sur l'ensemble.
- Rôle : donne une vision d'ensemble.
- Exemple : si le modèle a une accuracy de 0,93, cela veut dire que sur 100 clients, 93 sont bien classés (qu'ils quittent ou non).

Precision (Précision positive)

- •Définition : parmi les clients prédits comme partants, combien le sont réellement.
- •Rôle : éviter de fausses alertes.
- •Exemple: précision $0.88 \rightarrow \text{sur } 100 \text{ clients prédits comme churners,}$
- •88 quittent réellement, 12 auraient été faussement considérés comme à risque.

- Recall (Rappel / Sensibilité)
- Définition : parmi les clients qui quittent vraiment, combien sont détectés par le modèle.
- Rôle : éviter de rater des clients à risque.
- Exemple : recall 0,60 → sur 10 clients qui quittent, le modèle en détecte 6 et en manque 4.

► F1-Score

- Définition : moyenne harmonique entre Precision et Recall.
- Rôle : équilibre entre ne pas manquer de churners et ne pas avoir trop de fausses alertes.
- Exemple : F1-score 0,71 → bon compromis entre précision et rappel.

Tableau comparatif des modèles

Modèle	Accurac y	Precision	Recall	F1-score
Régression Logistique	0.859	0.533	0.247	0.338
Random Forest	0.906	0.925	0.381	0.540
Gradient Boosting	0.930	0.879	0.598	0.712

Choix final: Gradient Boosting Valeur plus eleve de Recall Identifie ~60% des clients à risque. Meilleur équilibre précision / rappel.

Importance des features calculee par le modele (odre decroissant)

- •Customer service calls 0.173
- •Total day charge 0.172
- •International plan (no) 0.115
- •Total/day minutes 0.109
- •Total eve minutes 0.075
- •International plan (yes) 0.067
- •Voice mail plan (no) 0.047
- •Total eve charge 0.035
- Number vmail messages 0.031
- Total intl calls 0.030

Interprétation des poids des features

Customer service calls (0.173): Les clients ayant fréquemment recours au service client présentent un risque élevé de résiliation.

Cette variable est la plus déterminante du modèle.

Total day charge (0.172): L'absence de plan international est un facteur clé associé au churn. Les clients avec un forfait international semblent plus fidèles.

International plan (no/yes) (0.115/0.067): Les clients sans forfait international risquent davantage de partir.

- •Total day minutes (0.109): Forte consommation peut révéler des attentes insatisfaites \rightarrow risque.
- •Voice mail plan (no) (0.047): Ceux qui n'ont pas d'options supplémentaires montrent moins d'engagement.

Actions correctives par feature

Customer service calls: Améliorer la qualité du support, suivi proactif des plaintes.

Total day charge: Proposer des offres personnalisées pour les gros consommateurs.

International plan: Mettre en avant des promotions sur les appels internationaux.

Voice mail plan: Inciter à souscrire à des options additionnelles.

Total intl calls: Fidéliser avec des packs internationaux adaptés.

Utilisation opérationnelle du modèle

Chaque mois:

- Le modèle prédit la probabilité de churn de chaque client.
- Les clients avec probabilité > seuil (ex. 0.6) sont classés « à risque ».

L'équipe marketing peut cibler :

- Appels proactifs.
- Réductions personnalisées.
- Offres adaptées au profil.

Prochaines Étapes

- Automatiser la prédiction mensuelle.
- Suivre l'évolution du churn après les actions.
- Ajuster le modèle avec de nouvelles données.
- Étendre l'approche à d'autres segments de clients.