

Mise en Œuvre d'un Modèle de Prédiction de la Valeur Vie Client (CLV) dans une Banque

NABIGH Mohamed

August 6, 2024

Abstract

Ce rapport présente un plan détaillé pour implémenter un modèle statistique de prédiction de la Valeur Vie Client (CLV) dans notre banque. Inspiré par les techniques discutées dans l'article "Une approche novatrice pour prédire la valeur vie client dans les entreprises B2B SaaS", cette implémentation vise à améliorer notre capacité à identifier et fidéliser les clients à haute valeur, optimiser les stratégies marketing, et améliorer la rétention des clients.

1 Résumé Exécutif

Ce rapport présente un plan détaillé pour implémenter un modèle statistique de prédiction de la Valeur Vie Client (CLV) dans notre banque. Inspiré par les techniques discutées dans l'article "Une approche novatrice pour prédire la valeur vie client dans les entreprises B2B SaaS", cette implémentation vise à améliorer notre capacité à identifier et fidéliser les clients à haute valeur, optimiser les stratégies marketing, et améliorer la rétention des clients.

2 Compréhension de l'Approche de l'Article

L'article propose un cadre flexible de modélisation statistique spécifiquement conçu pour prédire la CLV dans les entreprises de Software-as-a-Service (SaaS) B2B. Les éléments clés de cette approche incluent :

1. Structure Modulaire Hiérarchique:

- **Prédiction à Court Terme (T')**: Utilise des données récentes pour prédire la CLV sur une période plus courte.
- **Prédiction à Long Terme (T)**: Étend les prédictions à court terme pour couvrir la durée de vie complète du client.

2. Modélisation par Ensembles:

- Combine plusieurs modèles statistiques pour améliorer la précision des prédictions.
- Utilise différents modèles pour différents segments de clients basés sur des caractéristiques spécifiques.

3. Ingénierie des Caractéristiques:

- Crée des caractéristiques significatives à partir de données brutes, telles que des métriques agrégées, des indicateurs de tendance, des caractéristiques décalées et des termes d'interaction.
- Met l'accent sur les données récentes pour garantir que les prédictions sont basées sur les informations les plus pertinentes.

4. Évaluation du Modèle:

- Utilise des métriques telles que RMSE (Root Mean Squared Error), MAE (Mean Absolute Error) et SMAPE (Symmetric Mean Absolute Percentage Error) pour évaluer la performance du modèle.
- Effectue une analyse des résidus pour affiner la segmentation et la sélection du modèle.

3 Étapes et Conditions

3.1 Collecte des Données

- Collecter des données clients à partir de diverses sources au sein de la banque, y compris l'historique des transactions, les données démographiques des clients, les produits détenus et les métriques d'engagement.
- Assurer la collecte de divers types de données telles que des données catégorielles (par ex., segment de clients), des données entières (par ex., nombre de produits détenus), et des données flottantes (par ex., soldes de comptes).

3.2 Prétraitement des Données

- Nettoyer les données en traitant les valeurs manquantes, en supprimant les doublons et en corrigeant les entrées incohérentes.
- Transformer les données en normalisant les caractéristiques numériques et en encodant les variables catégorielles en utilisant des techniques telles que l'encodage one-hot.

3.3 Ingénierie des Caractéristiques

- **Métriques Agrégées:** Calculer le total des transactions, la valeur moyenne des transactions et la fréquence des transactions.
- **Indicateurs de Tendance:** Calculer les taux de croissance et les changements en pourcentage des soldes et des volumes de transactions.
- **Caractéristiques Décalées:** Créer des caractéristiques représentant les valeurs passées de certaines métriques (par ex., solde du mois précédent, transactions du trimestre précédent).
- **Caractéristiques Cumulatives:** Calculer les métriques cumulatives jusqu'au point d'observation.
- **Termes d'Interaction:** Créer des termes d'interaction entre les caractéristiques pour capturer les effets combinés.

3.4 Entraînement du Modèle

- Diviser les données en ensembles d'entraînement et de test pour évaluer la robustesse du modèle.
- Utiliser des modèles de régression pour s'entraîner sur les caractéristiques générées. Les modèles potentiels incluent la régression linéaire, les forêts aléatoires et les machines à vecteurs de support.
- Effectuer un ajustement des hyperparamètres pour optimiser la performance du modèle. Les techniques de validation croisée peuvent être utilisées pour cet ajustement.
- Évaluer les modèles en utilisant des métriques telles que RMSE, MAE, et SMAPE pour déterminer la précision des prédictions.
- Analyser les résidus pour identifier et corriger les erreurs systématiques, améliorant ainsi la précision globale du modèle.

3.5 Déploiement du Modèle

- Intégrer les modèles dans l'infrastructure informatique de la banque pour permettre une utilisation en production.
- Automatiser les processus de collecte de données, de prétraitement, d'ingénierie des caractéristiques et de ré-entraînement du modèle pour garantir des prédictions à jour et précises.
- Mettre en place des systèmes de surveillance pour suivre la performance du modèle en temps réel et identifier les dégradations de performance potentielles.

3.6 Application Commerciale

- Optimiser les stratégies marketing en utilisant les prédictions de CLV pour cibler les clients à haute valeur avec des offres personnalisées.
- Améliorer la rétention des clients en identifiant les clients à risque et en développant des stratégies de rétention ciblées.
- Adapter les offres de produits en fonction des prédictions de CLV pour mieux répondre aux besoins des clients.
- Mener des analyses de retour sur investissement (ROI) pour guider l'allocation stratégique des ressources et optimiser la performance des campagnes.

4 Projection de l'Approche sur Notre Cas d'Utilisation Spécifique

4.1 Prédiction à Court Terme

Pour le court terme, nous nous concentrerons sur la prédiction de la CLV cumulée sur une période de 2 à 3 ans. Cela nous permettra d'observer l'impact sur notre modèle et d'affiner nos processus d'ingénierie des caractéristiques et de sélection de modèle. Les prédictions à court terme utiliseront des données récentes pour garantir leur pertinence et leur précision.

4.2 Prédiction à Long Terme

Pour la prédiction à long terme, nous sélectionnerons des variables facilement prévisibles dans le temps. Celles-ci peuvent inclure des caractéristiques démographiques (par ex., âge, revenu), des comportements historiques (par ex., fréquence des transactions) et des métriques d'engagement (par ex., fréquence des connexions). La prédiction à long terme étendra les prédictions de CLV à court terme pour couvrir toute la durée de vie du client.

5 Conclusion

Nous avons commencé à implémenter les techniques de modélisation statistique de l'article, et nous nous attendons à ce qu'elles améliorent considérablement notre capacité à prédire et à tirer parti de la Valeur Vie Client. Cela nous permettra de prendre des décisions plus éclairées, d'optimiser les efforts marketing et d'améliorer la satisfaction et la rétention globales des clients. Le projet impliquera une collecte de données minutieuse, une ingénierie des caractéristiques robuste et des techniques de modélisation avancées pour atteindre ces objectifs.