

D

PAROLES D'EXPERTS

René LEFEBURE
Directeur Recherche
& Développement



Le Machine Learning pour mesurer l'engagement Client

Les nouvelles techniques de Machine Learning (Random forest, SVM, réseaux de neurones ou bayésiens) permettent de tirer profit de la collecte des données pour modéliser l'engagement client.

La puissance des algorithmes apporte une capacité d'extension des études qualitatives, réalisées sur des échantillons, pour les déployer sur l'intégralité d'une base de données.

Conexance, spécialiste du « data extended », permet de mettre en place des dispositifs au service de la relation client.

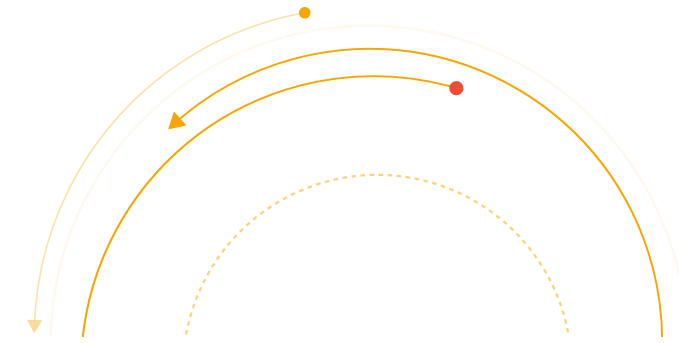
La collecte de données qualitatives s'est intensifiée avec l'essor des canaux digitaux : enquête de satisfaction post appel ou livraison, notation des intervenants dans la transaction, actions de « like » ou « follow » sur les réseaux sociaux, sans omettre les activités de type barométriques comme les Net Promotions Scores. Ces dispositifs permettent de monitorer la qualité des processus client et de s'interroger sur la mise en œuvre d'actions d'optimisation de l'expérience client.

Cependant, ces enrichissements ne concernent que les clients les plus actifs ou récents, et présentent des indices très stables dans le temps, ce qui peut se traduire par une baisse de l'attention portée à ce type d'études après quelques mois. Par ailleurs, s'ils sont des instruments de prise de décision, ils ne sont pas des leviers opérationnels à mise en œuvre rapide.

Pour dépasser ces limites, Conexance a accompagné des entreprises pour modéliser grâce au Machine Learning les niveaux d'engagement client, utiliser ces modèles pour qualifier l'ensemble du fichier et accompagner des actions de fidélisation ou de rétention.

Un projet d'extension de données est assez classique. La collecte de données de satisfaction doit mentionner l'autorisation de l'utilisation des données pour but d'études et ne doit pas permettre de relance commerciale. Par respect de la confidentialité, les fichiers sont « anonymisés ». L'objectif est de mieux comprendre les facteurs « d'engagements » ou de « désengagements » du client afin de favoriser les actions correctrices.

La réconciliation entre les données « qualitatives » et les données de la DMP 360 (On et Off Line) pose des



problèmes spécifiques. Il faut respecter le « périmètre » de l'enquête : un questionnaire post livraison ne s'adresse pas la même population qu'une enquête post sinistre. Il est important d'assurer une cohérence entre le périmètre de modélisation et d'extension. Les données d'enquêtes sont collectées « post processus » et il faut reconstituer la base de modélisation au « moment » de l'enquête pour respecter la cohérence des informations au niveau de la temporalité.

Le Machine Learning apporte plusieurs avantages par rapport aux techniques traditionnelles. Le premier est d'utiliser beaucoup plus de données au service de la conception des modèles. Des modèles impliquant entre 100 et 200 variables prédictives (« features ») sont fréquents. Cette approche « gloutonne » est utile car les réponses sont souvent conditionnées par des facteurs tels que l'âge, l'équipement, l'ancienneté ou le volume d'activité et il est important de pouvoir gérer ces « chaînes d'engagement ». Les réseaux bayésiens ou les techniques à bases d'arbres multiples répondent à ce contexte.

Le second avantage est d'utiliser plusieurs techniques de modélisation et méta-paramètres (parfois 100 modèles en parallèle sur une architecture Cloud) pour « benchmarker » les différentes approches. Cette approche combinatoire, utilisée avec succès dans le cas Netflix, fiabilise la modélisation.

Cette hyper modélisation, qui tire profit des données en multipliant les approches, sert à fiabiliser la base d'enquête en « filtrant » certains répondants aux notes incohérentes en regard des données de la DMP. Ce filtrage peut concerner entre 5 à 15 % des répondants.

L'extrapolation de l'engagement modélisé est validé par le responsable des enquêtes. Il faut retrouver une cohérence entre les facteurs prédictifs du modèle et les indicateurs traditionnels de mesure des enquêtes qualitatives (sexe, âge, ancienneté, profession, etc..).

La qualité de la modélisation n'échappe pas aux traditionnelles populations d'apprentissage et de test, mais les tech-

niques de bootstrapping facilitent la gestion des populations restreintes. Ainsi une enquête réalisée sur une population de 1.500 à 2.000 répondants peut être (éventuellement car tout n'est pas modélisable) étendue sur une population de plusieurs centaines de milliers de clients. La mesure de la qualité d'extrapolation dépend évidemment de la richesse de la DMP.

La modélisation de l'engagement est utilisée dans les dispositifs entrants ou sortants multi-canaux. Les actions de parrainage ou la mise en œuvre de dispositifs préventifs (probabilité d'un engagement faible) sont des exemples fréquents. ■

