

Prédire la valeur d'un client : tout est question de modèle ! (tribune)

La VVC (Valeur Vie Client ou, en anglais, *Customer Lifetime Value*) est la valeur actualisée nette des profits qu'un client générera dans le futur. À l'instar de la Valeur Actualisée Nette de n'importe quel type d'actif, c'est la valeur financière d'un client individuel pour l'entreprise.

La VVC est donc une mesure très importante, et ce, à plusieurs niveaux. Au niveau individuel, elle permet d'**identifier les clients qui seront les plus rentables**, et qu'il faut donc cibler et retenir en priorité. A un niveau agrégé, elle permet de connaître la **valeur totale d'un segment de clients** et donc aussi de l'entreprise dans son ensemble. De plus, la VVC est utilisée pour mesurer l'impact des campagnes marketing (*Return on Investment*) ou pour décider des meilleures actions à prendre en fonction de la rentabilité attendue (*Next Action to Take*). Des études scientifiques ont montré par exemple que 25% des clients n'étaient pas rentables pour une entreprise spécifique, ou que cibler les clients sur base de la VVC permettait d'améliorer l'efficacité marketing d'une autre entreprise de 33%...

Bref, utiliser la VVC permet de prendre de meilleures décisions. C'est donc une question au cœur du processus *Marketing Analytics* depuis plusieurs années...

La problématique principale de la VVC est d'arriver à prédire l'activité future d'un client, et donc de concevoir le meilleur modèle de prédiction. En pratique, la question de la modélisation de la VVC se réduit souvent à trois questions : (1) combien de temps le client restera client (quelle sera sa *durée de vie* ?), (2) tant qu'il sera actif, combien de transactions effectuera-t-il (la *fréquence*) et (3) quelle sera la *valeur* (c'est-à-dire les montants) de ces transactions ?

Durée de vie, fréquence et valeur : telle est la question

Dès lors, plusieurs approches coexistent pour estimer ces trois mesures : durée de vie, fréquence, et valeur. Certains modèles sont simples, d'autres plus complexes...

L'approche la plus célèbre étant sans doute **l'approche RFM** (récence, fréquence et montant) qui mesure respectivement le temps écoulé depuis la dernière transaction (récence), le nombre de transactions effectuées par le client depuis le début de la relation (fréquence) et la somme moyenne dépensée (montant). Cette approche basée directement sur des mesures « passées » a l'avantage d'être très facile à mettre en place, même si la question de la durée de vie attendue n'est pas vraiment réglée puisqu'elle n'offre qu'un « proxy » très approximatif (la récence) quant à cette question.

Des [études](#) scientifiques ont d'ailleurs démontré que, dans certains cas, un client qui n'avait pas fait de transaction depuis longtemps avait en fait plus de chance d'être actif qu'un client qui en faisait

régulièrement et fréquemment. L'illustration suivante montre ainsi deux clients : le premier a fait 16 transactions et le deuxième n'en a fait que 3, et la dernière transaction du second client est moins récente que celle du premier. Cependant, si on se demande « quelle est la probabilité que ces deux clients soient toujours actifs ? », on donnera sans doute une probabilité plus élevée au second qu'au premier. S'il se comporte comme dans le passé, on peut en effet s'attendre à ce que le second client fasse une quatrième transaction bientôt. Ce qui n'est pas le cas du premier qui, s'il était toujours actif, aurait déjà dû avoir fait des transactions avec l'entreprise concernée depuis son dernier passage...



De nombreuses approches existent, et laissons pour le moment de côté les techniques de prédiction basées sur le Machine Learning (Apprentissage Automatique) qui, si elles offrent des prédictions très précises à court terme, présentent deux inconvénients majeurs : il s'agit généralement de « boîtes noires » qui ne permettent pas de comprendre le comportement sous-jacent du consommateur. Et, dans la plupart des cas, du fait de la nature même des techniques utilisées, elles présentent très souvent le risque de ne pas pouvoir détecter un « sur-apprentissage ». C'est-à-dire quand l'ordinateur sur-interprète les données et fait des prédictions soi-disant fiables alors qu'elles sont en réalité totalement incertaines...

La recherche scientifique en management propose aussi de nombreux modèles – assez complexes il faut le dire – permettant de modéliser précisément le comportement consommateur et donc d'anticiper la rentabilité future de ceux-ci. Parmi les plus célèbres, il faut sans doute citer le premier d'entre eux, le modèle [Pareto/NBD](#) sorti à la fin des années 80, ou un modèle plus récent, comme le [Gamma-Gompertz/NBD](#) présenté en 2012.

Toutefois, jusqu'ici, ces modèles présentaient un défaut majeur : celui de faire **l'hypothèse d'indépendance entre la durée de vie, la fréquence et les montants** dépensés.

Durée de vie, fréquence, et valeur : tout est lié !

En effet, en fonction du secteur dans lequel on se trouve, on peut s'attendre à ce que ces différents processus soient interdépendants...

Par exemple, si le propriétaire d'un chat achète tous les mois une certaine quantité de nourriture pour celui-ci, la quantité totale consommée sera plus ou moins constante. Dès lors, si – pour profiter d'une promotion temporaire – ce client achète plus rapidement qu'il ne l'aurait fait d'habitude, et pour un plus grand montant, il décidera sans doute dans les semaines qui suivent d'acheter moins qu'habituellement puisque son stock le permet et que son chat ne peut consommer qu'une certaine quantité par jour de toute façon. Dans d'autres cas, on peut s'attendre à ce que, pour profiter d'une promotion, un client achète plus rapidement, mais pour un montant inférieur à ce qu'il aurait dépensé habituellement...

On le voit, la durée entre deux transactions (et donc la fréquence), ainsi que les montants dépensés, sont liés dans le temps...

D'autres associations sont envisageables, en relation avec la durée de vie attendue par exemple. Certaines études ont montré que les clients les plus loyaux (donc dont la durée de vie attendue était plus longue) viendront plus régulièrement ou dépenseront plus à chacun de leurs passages. A l'inverse, dans des secteurs très compétitifs, où la fréquence d'achat est très élevée, les clients ont plus souvent l'occasion de réévaluer les offres concurrentes, et on pourra observer un taux d'attrition plus élevé...

C'est pourquoi, dans un [article](#), deux chercheurs et moi-même montrons qu'il faut **modéliser ces trois processus en tenant compte des associations** qu'on peut attendre entre la durée de vie des clients, la fréquence des transactions, et la valeur de ces transactions. Et que cette interdépendance peut dépendre du secteur ou même des clients au niveau individuel ! Nous montrons aussi que notre modèle plus flexible permet à la fois de mieux prédire la Valeur Vie Client, tout en ciblant plus intelligemment les segments de clients. En effet, si certains clients « compensent » leurs achats (par exemple s'ils achètent plus rapidement que d'habitude mais pour des montants plus petits), une promotion sera moins efficace que prévue. Il faut donc essayer de cibler les clients dont les comportements dans leur ensemble – en tenant compte des trois facteurs – sont les plus rentables.

La VVC est complexe, vos modèles doivent l'être aussi

La Valeur Vie Client cherche à atteindre un objectif ambitieux : prédire l'activité générée par tous les clients individuels d'une entreprise. Si l'exercice peut être laborieux, cette mesure présente de nombreux avantages et a démontré qu'elle permettait de rendre les campagnes marketing plus efficaces.

Cependant, il faut rester prudent. Si l'exercice est approché d'une manière trop naïve, les conclusions qu'on en tire risquent d'être erronées. Il faut veiller à adopter la bonne démarche, et opter pour une méthodologie qui permet de modéliser le comportement des clients dans leur complexité.



Par Nicolas Glady, titulaire de la Chaire Accenture Strategic Business Analytics de l'Essec

A lire aussi :

La précédente tribune de N. Glady : [Sans que nous nous en rendions compte, le Cloud change le](#)

[monde](#)