



**MODÉLISER ET OPTIMISER  
LA SATISFACTION CLIENT**  
GRÂCE À LA *LIFETIME VALUE*  
ET À L'IA



# 01.

## INTRODUCTION

### Comment optimiser ses revenus en modélisant la satisfaction client ?

De nombreuses tendances *business* (intensification du niveau de concurrence, digitalisation, expérience client comme élément différenciant, culture du *feedback* et de la notation, *etc.*) contribuent effectivement au caractère stratégique de l'optimisation de la satisfaction client. Or, pour assurer la meilleure attribution des ressources humaines et financières sur ce point, il est primordial d'identifier :

- La façon la plus objective et opérante possible de mesurer la satisfaction client,
- Les différents facteurs qui vont influencer cette satisfaction et comment ils l'impactent

Prenons l'exemple des plaintes clients. Il s'agit d'un pivot : c'est le moment où on peut perdre lesdits clients ou, au contraire, les fidéliser, voire en faire des ambassadeurs de sa marque.

C'est pour pouvoir **optimiser sa politique de service client (ressources mises à disposition, cas à prioriser, indemnisation, etc.) en cas de réclamation que nous avons été sollicités par un acteur majeur du transport de passagers**. L'enjeu de la mission était particulièrement fort car les volumes en question étaient significatifs.

Le client reçoit plus de **20 000 réclamations par an**, l'équipe qui les traite comprend **plus d'une dizaine de personnes** pour un budget, compensation comprise, de plus de **10 millions d'euros**.

*NB : Le sujet est très stratégique pour le client, donc les données partagées dans ce chapitre seront anonymisées et seuls des ordres de grandeur seront fournis.*

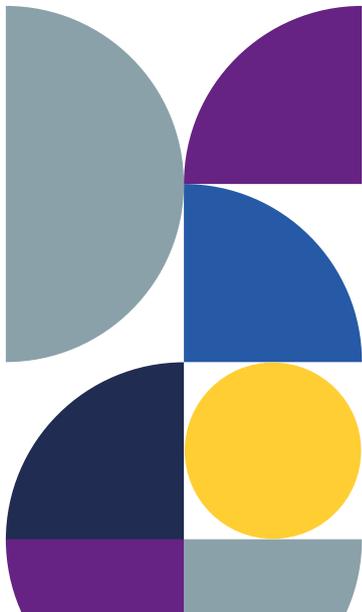
Pour répondre au besoin, nous avons mis au point une méthode systématique basée sur des techniques d'apprentissage machine combinées à une méthode statistique plus classique pour obtenir un modèle explicatif de la *LifeTime Value* (LTV) par typologie de client.

Ce modèle a permis :

- D'identifier les facteurs clefs explicatifs de la LTV du client,
- De comprendre finement l'impact de ces derniers sur la LTV,
- De modifier la politique et les moyens opérationnels des services concernés dans une logique de ROI quantifiable.

# 02.

## LES ENJEUX DE LA MODÉLISATION : CHOISIR LA BONNE MÉTRIQUE « CIBLE » ET IDENTIFIER LES VARIABLES EXPLICATIVES



Le défi posé par notre client était de modéliser de manière bien plus performante (qualité de la prédiction) et actionnable la satisfaction client pour pouvoir optimiser les paramètres de sa politique de service client. Ce *challenge* s'est imposé aux équipes du service client lorsque les tentatives de modélisation basées sur les questionnaires de satisfaction clients se sont avérées inutilisables (cf. plus bas).

### 2.1 DÉFINITION ET QUALIFICATION DE LA SATISFACTION CLIENT

Le premier sujet de la modélisation est le choix de la variable « cible », c'est-à-dire celle que l'on cherche à modéliser dans notre algorithme de *Machine Learning*. Concernant la satisfaction client, il existe 4 grandes catégories de métriques :

#### 1/ Les métriques d'expérience client

Qu'il s'agisse du *Net Promoter Score* (NPS), de la *Customer Satisfaction* (CSAT), ou du *Customer Effort Score* (CES), ces métriques résultent d'une enquête client, à qui il est demandé de répondre à des questions-type (source : "6 Most Popular Customer Experience Metrics and KPIs Explained Simply", site Web lumoa.me, avril 2022) :

- « **Recommanderiez-vous cette entreprise à vos amis ?** »
- « **Qu'avez-vous pensé de la prestation ?** »
- « **Avons-nous résolu votre problème ?** »

Ce type de métrique présente un biais principal, le biais du survivant. Nous ne récoltons que les notes des clients qui répondent à l'enquête. Dans notre cas, les données

de NPS présentait une grande volatilité et de précédentes études au sein de l'entreprise n'ont pas permis de dégager des corrélations entre satisfaction dans un questionnaire et comportement futur d'un client. Dès lors, il est impossible d'utiliser le NPS comme variable « cible » et d'en prédire la valeur en fonction des variables de traitement de la réclamation. Bien qu'apportant de nombreux éléments qualitatifs intéressants, le NPS ne peut pas être utilisé pour optimiser son revenu futur grâce à sa politique de gestion des réclamations.

## 2/ Analyse des commentaires sur Internet

Très riches, les données issues des commentaires clients sur X (anciennement Twitter), Google ou Yelp nécessitent, pour être exploitées, des techniques de traitement du langage naturel (NLP) dont le réglage est complexe. Par ailleurs, elles ne sont ni structurées ni contextualisées, ce qui les rend très difficiles, voire impossibles à relier à un client et ses événements clients. Ces informations sont donc inutilisables dans le cadre d'une modélisation.

## 3/ Attrition / Rétention

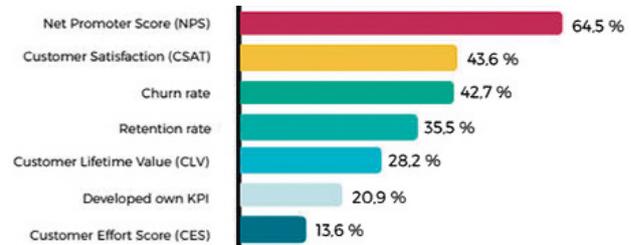
Une approche alternative consiste à utiliser des métriques temporelles pour évaluer si un client revient (1) ou ne revient pas (0) dans une fenêtre de temps donnée. Cette approche binaire est objective et supprime le biais du survivant mais simplifie le sujet à tel point qu'elle ne peut être utilisée dans le cadre d'un modèle qui doit guider le métier.

## 4/ LifeTime Value (LTV)

Il s'agit, à nos yeux, de la bonne data à utiliser. La *Life-time Value* (LTV), parfois appelée *Customer Lifetime Value* (CLV) est une mesure à la fois objective, quantitative et qui saisit l'attrition. Cette approche utilise le revenu futur du client comme une mesure indirecte de sa satisfaction.

Bien que nous privilégions cette métrique dans un cadre de modélisation, la LTV n'est pas la donnée la plus fréquemment utilisée (cf. schéma ci-dessous) pour mesurer la satisfaction client. Par ailleurs, nous recommandons d'utiliser le NPS, commentaires et autres comme métriques qui, combinées avec le LTV, permettent d'obtenir une vision holistique de la satisfaction client. Pour résumer, NPS et commentaires pour une vision qualitative ; LTV pour une modélisation.

### What customer experience KPIs do you follow?



*Selon une enquête menée par Lumoa (source : "6 Most Popular Customer Experience Metrics and KPIs Explained Simply", site Web lumoa.me, avril 2022) auprès de ses clients, les KPIs satisfaction client les plus suivis sont liés à l'expérience client.*

## 2.2 CHOIX DES VARIABLES EXPLICATIVES

La création d'un modèle facilement actionnable constitue un aspect crucial de l'analyse de la satisfaction client. En effet, elle permet à la direction du service dédié de traiter les différentes demandes en fonction de variables explicatives préalablement identifiées.

**Nous en avons relevé trois principaux types :**

### 1/ Traitement de la demande

Il s'agit des données directement activables par le service clients du fait qu'elles sont reliées à leur intervention : la durée de traitement, le nombre et le type d'interactions avec le client, le mode et le montant de l'indemnité éventuelle, etc.

### 2/ Typologie du client

Ce sont les variables qui caractérisent le client (dans le respect du cadre imposé par la RGPD) : ancienneté, niveau de fidélité, âge, code INSEE, etc.

### 3/ Typologie de l'événement client

Il s'agit des données liées à l'événement à propos duquel le client effectue une demande. Il peut s'agir du contexte d'un sinistre ainsi que de son importance, de la durée d'un retard ou encore du contrat demandé dans le cadre d'un voyage, etc.

## 2.3 « EXPLICABILITÉ » ET PERFORMANCE DES MODÈLES

Un modèle prédictif est systématiquement mis au point, pour ce faire, un arbitrage a été réalisé entre 2 caractéristiques du modèle :

### 1/ Son « explicabilité »

Un modèle explicable est un modèle dont les résultats sont faciles à comprendre et à interpréter

### 2/ Sa performance prédictive

C'est-à-dire le degré de précision des prédictions du modèle.

Pour schématiser, nous choisirons un modèle explicable dans les cas où les décisions doivent pouvoir être analysées et comprises systématiquement. C'est le cas pour les modèles utilisés :

- Dans les domaines de la santé, de la justice ou de la finance et, globalement, dans l'ensemble des secteurs et sujets où les risques liés aux décisions prises sur la base du modèle sont élevés.
- Au sein des secteurs comportant des sujets réglementés ou faisant l'objet d'exigences légales.

À noter que ces secteurs ou sujets sont nombreux car le Règlement Général sur la Protection des Données (RGPD) de l'Union Européenne donne le droit aux individus de demander une explication sur les décisions automatisées qui les concernent.

Un autre avantage important des modèles explicables est la capacité à détecter les biais ou les tautologies (exemple : un client « fidèle » a une LTV supérieure) dans les prédictions. En outre, l'« explicabilité » favorise la confiance et l'acceptation des acteurs-métiers envers

les résultats du modèle. En effet, lorsque les professionnels comprennent aisément les mécanismes de prise de décision, ils sont plus susceptibles de faire confiance au modèle et de s'appuyer sur ses prédictions pour orienter leurs actions. Cela facilite l'adoption du modèle dans les processus décisionnels quotidiens, renforçant ainsi l'efficacité des mesures prises.

Dans notre cas, la contrainte portait sur la définition d'un compromis idéal entre performance de la prédiction nécessaire au regard des montants conséquents en jeu et l'« explicabilité » du modèle pour garantir au métier qu'il pourrait exploiter les conclusions et prédictions du modèle de manière claire et simple. Tout cela est fait dans le but d'amender sa politique de service clients et d'optimiser les ressources de cette dernière.

## 03. LES CLASSIQUES DATA REVISITÉS GRÂCE À L'IA

Nous privilégions toujours les approches axées sur l'utilisation de données internes structurées pour nos projets data. Ce pragmatisme s'appuie sur 2 grands principes : l'**utilisation d'une méthode éprouvée** et une **relation entretenue et continue** avec le métier et les personnes qui vont utiliser le modèle au quotidien.

Pour le cas présent, nous avons opté pour la méthodologie CRISP DM (*Cross-Industry Standard Process for Data Mining*), un standard ouvert de procédure pour la résolution de problématiques data soutenu par IBM. L'approche se décompose en six étapes (cf. schéma ci-après).

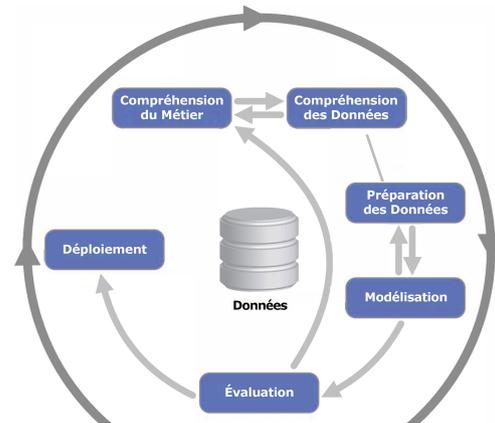


Diagramme des phases principales d'une méthodologie CRISP DM. Source : Wikimedia Commons

À chacune de ces étapes, nous cherchons à faire en sorte d'être le mieux compris possible par les métiers :

**1/ Compréhension du métier** : Cette étape consiste à définir le problème à résoudre, à déterminer les objectifs commerciaux et à comprendre quelles données sont disponibles. Pour cela, nos *data scientists* ont acquis une compréhension fine du *process* et donc de l'ensemble des scénarii de traitement d'une réclamation, via des ateliers avec le métier (des sessions de « vis ma vie » sont aussi très performantes pour cela).

**2/ Compréhension des données** : Sur la base de données collectées à partir de l'ERP du client et de son outil de gestion du service clients, les données manquantes ou les erreurs sont corrigées. Par ailleurs, les données aberrantes (*outliers*) sont repérées et analysées pour exclusion ou inclusion.

**3/ Préparation des données** : Nettoyage et transformation des données (uniformisation du format des données, des unités, etc.) pour qu'elles soient utilisables pour l'analyse. Elles sont également sélectionnées et réduites pour ne conserver que les plus pertinentes pour la résolution du problème. Les données très corrélées sont ainsi identifiées. Par exemple, la variable « nombre de documents manquants » est très corrélée aux nombres d'interactions entre le service clients et le client. C'est donc cette dernière uniquement qui a été retenue dans le modèle. Autre exemple, les variables « temps de réponse du client » et « temps de réponse du service clients » sont toutes deux corrélées à la variable « temps de traitement total de la réclamation ». Pour un maximum de pertinence nous n'avons gardé que la variable de temps de traitement du service clients, soit celle sur laquelle des actions peuvent être mises en place.

**4/ Modélisation** : Plusieurs types de modèles sont choisis en fonction des typologies de données et de la variable « cible ». Cette étape est précisée plus loin dans l'article (cf. paragraphes 3.2 et 3.3). Les modèles créés sont calibrés à partir des données préparées lors de l'étape précédente.

**5/ Évaluation** : Les modèles sont ensuite comparés dans leur performance prédictive. Le modèle fournissant le meilleur rapport performance de prédiction de la LTV/explicabilité est sélectionné. Il est alors challengé avec les métiers et s'il ne répond pas à leurs exigences, les étapes précédentes sont reprises pour l'améliorer.

**6/ Déploiement** : La dernière étape consiste à déployer le modèle en production et à surveiller son efficacité. Des ajustements peuvent être nécessaires pour assurer une performance optimale en continu.

## 3.1 LA PLACE DES DATA SETS

### La data est là, il faut la rendre parlante !

Le premier travail sur les données de l'entreprise a consisté en la définition précise de la LTV, notre donnée cible. Pour pouvoir capturer les évolutions avant et après le traitement de la réclamation, il est décidé de mesurer les revenus générés par le client sur les 12 mois précédant la réclamation et de les comparer à ceux générés sur les 6 mois suivant la réclamation.

En effet, le *set* de données couvrant une période de confinement due au Covid, la comparaison ne pouvait pas se faire sur la période de 12 mois avant et après la réclamation. On obtient donc une LTV calculée sur 18 mois, dont 6 *post*-réclamations, permettant de calculer la différence de revenu avant / après la réclamation et la façon dont elle a été traitée.

La « cible » étant définie, le défi suivant consistait à identifier toutes les sources pertinentes à la modélisation de cette LTV et de ses variations. Il est admis que pour mener des études statistiques ou encore déployer des modèles de *machine learning*, il est nécessaire de disposer d'une quantité massive de données. Dans de nombreux cas, la qualité des données est plus importante que leur quantité et il est préférable d'avoir un ensemble de données précis et fiable plutôt qu'un volume énorme de données brutes et peu structurées. C'est pour cela que l'équipe a investi du temps sur la sélection d'un échantillon représentatif de plusieurs milliers d'individus. Cette étape fut réalisée avec les équipes qui collectent et assurent la mise en conformité des données de réclamations en collaboration avec les équipes du service clients qui utiliseront le modèle. Une fois l'échantillon représentatif sélectionné, le travail de collecte et d'enrichissement de la donnée commence, sur la base des données existantes qui ne pouvaient être exploitées telles quelles (comme c'est souvent - et même quasiment toujours - le cas).

Ce travail a permis d'identifier les tables de données pertinentes, de les amender pour rendre les données parlantes et de les enrichir grâce à une succession d'étapes :

**1/ Récupération des tables « historique du traitement des réclamation »** comprenant toutes les étapes de traitement d'une réclamation : objet, horodatage de l'ensemble des opérations ou messages, type de contact, dédommagement, nombre de réouvertures de celle-ci, etc.

**2/ Intégration des informations de circonstances de la réclamation** via une table comprenant les informations liées à l'incident ayant entraîné la réclamation (information concernant le voyage, heure de départ, arrivée, etc.).

**3/ Intégration des informations clients et des valeurs de LTV** via les tables comprenant les informations clients (notamment leurs typologies) et leurs dépenses (permettant le calcul de LTV avant/après réclamation).

Chacune de ces étapes d'agrégation de données ont été l'occasion de procéder à la création de *"features"* qui permettent d'améliorer la performance du modèle. Ce *feature engineering* est un processus de création et de transformation de variables pour les rendre plus informatives et pertinentes pour un modèle d'apprentissage automatique.

## ÉVOLUTION LES PLUS FRÉQUENTES DU TRAITEMENT D'UNE RÉCLAMATION – LES 3 DERNIERS ÉVÈNEMENTS

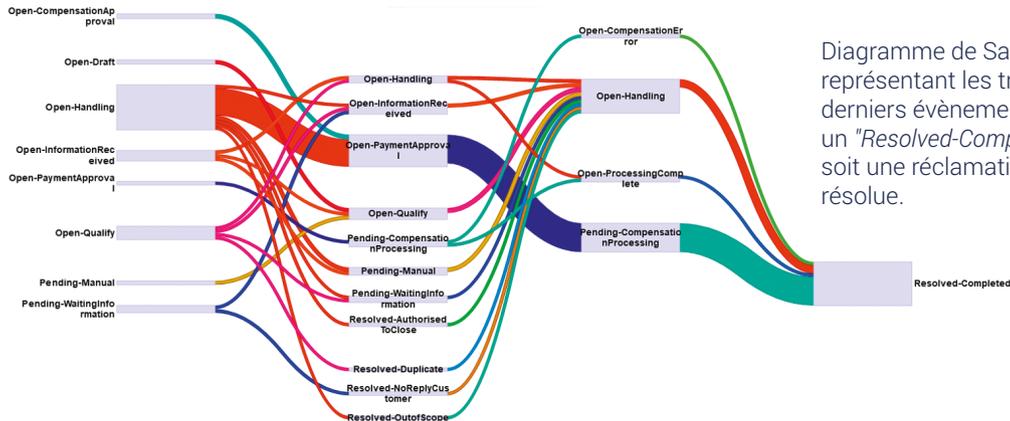


Diagramme de Sankey représentant les trois derniers événements avant un "Resolved-Completed", soit une réclamation résolue.

Pour l'illustrer de manière concrète dans un cas impliquant des processus à suivre, nous présentons dans le graphique sur la page de gauche la manière dont nous avons analysé les parcours de résolution des plaintes notamment des trois derniers événements avant une résolution.

Une fois ces étapes passées, nous avons été en mesure de vectoriser le chemin de résolution de chaque réclamation par client et de fournir cette information à nos modèles. Effectivement, via nos analyses et nos boucles d'échange avec le métier, il a été identifié que le « chemin » de clôture d'une réclamation, c'est-à-dire la nature des étapes combinées à leur enchaînement est clé pour expliquer la satisfaction clients. En faisant cela, nous entraînions le modèle à comparer les différents scénarii de résolution de la réclamation ce qui lui a ensuite permis de déterminer les chemins de résolutions qui maximisent la LTV des clients.

L'ensemble de ces traitements, enrichissements et créations de données ont produit un *data set* de 87 variables, potentiellement explicatives de la variation de LTV.

## 3.2 RENDRE LE MODÈLE OPÉRATIONNEL PAR L'UTILISATION DE L'IA ET UNE RÉGRESSION LINÉAIRE

Dans le cadre de la gestion client, l'intérêt de l'utilisation du *machine learning* est intégré depuis le début du siècle (source : "*Welcome to the SHAP documentation*", site [Web shap.readthedocs.io](http://shap.readthedocs.io), 2018) en particulier via la régression linéaire, une méthode simple et éprouvée par les entreprises. Appliquée à notre problématique, elle consiste à corréler la LTV du client et les caractéristiques du traitement de sa réclamation (nombres de contacts, type de contact, durée entre les prises de contact).

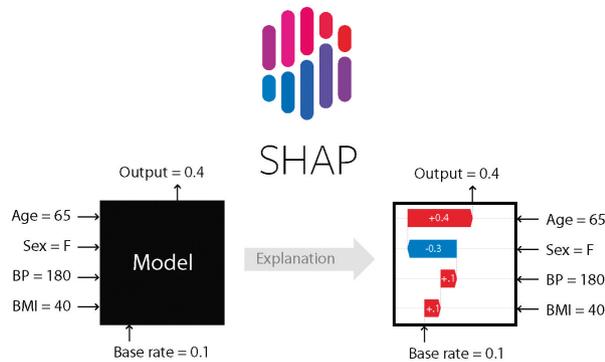
La régression linéaire a une grande qualité : son explicabilité car elle donne un « poids explicite » aux variables explicatives. Cependant, cette dernière ne fonctionne que si la relation entre la satisfaction des clients et les variables est linéaire (c'est-à-dire qui suit une relation du type  $y = ax + b$ , dans notre cas, cela se traduirait par exemple par des équations du type  $LTV = a \times \text{durée de la réclamation} + b$ ), ce qui n'est pas toujours le cas.

Dans les cas de relations non-linéaires des techniques de *machine learning* plus avancées sont nécessaires. Parmi ces modèles plus performants, le "*gradient boosting*" fournit des prédictions bien plus fines. Il reste néanmoins un modèle qui, sans aménagement, apparaîtra comme "*black box*" - donc un résultat difficile à interpréter - au métier.

C'est pourquoi nous avons choisi d'y associer la librairie de SHAP pour pouvoir augmenter la lisibilité de notre modèle. La librairie SHAP (source : "*Welcome to the SHAP documentation*", site [Web shap.readthedocs.io](http://shap.readthedocs.io), 2018)



généralisant les valeurs de Shapley issues de la théorie des jeux aux prédictions de *Machine Learning*, est l'outil actuel permettant le mieux d'augmenter l'explicabilité des modèles non-linéaires. En effet, elle permet d'associer un « poids » à chaque variable dans le résultat d'une prédiction.

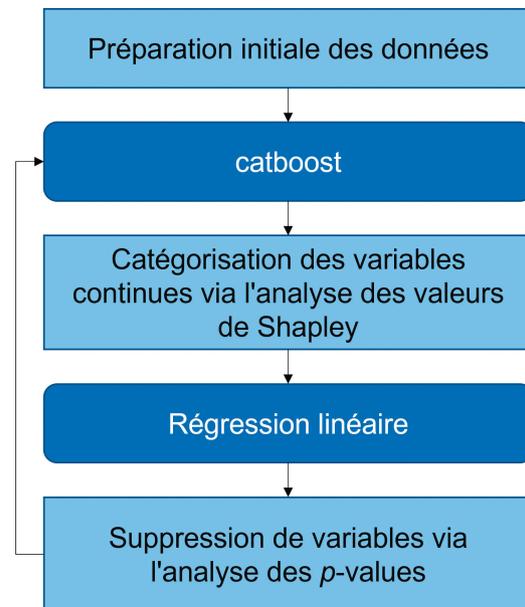


La librairie SHAP permet de mesurer la contribution relative de chaque variable dans la prédiction en sortie d'un modèle "black box".

Dans cet exemple de prédiction d'occurrence d'une maladie, tirée de la librairie SHAP, l'occurrence moyenne de l'ensemble des individus est de 0,1 (*Base rate* = 0,1). Grâce à la librairie SHAP, nous pouvons représenter la contribution de chacune des variables analysées : ici, les variables « BMI et BP » (soit *Body Mass Index* ou Indice de Masse Corporelle en français - IMC et *Blood Pressure* ou tension artérielle en français) et contribuent positivement à hauteur de 0,1 points chacune. Quant à la variable « sexe » elle contribue négativement de 0,3 points tandis que l'âge contribue positivement de 0,4 points. L'ensemble menant à une prédiction d'occurrence pour cet individu de 0,4 points.

### 3.3 BEST OF BOTH WORLDS : À LA RECHERCHE DU COMPROMIS PERFORMANCE/ EXPLICABILITÉ IDÉAL

Afin de trouver le meilleur compromis performance/explicabilité, nous avons cherché à tirer le meilleur des 2 mondes en combinant performance du *gradient boosting* et explicabilité de la régression linéaire. Pour cela, nous avons développé un système multi-modèle, schématisé de la manière suivante :



Plus précisément :

- **Catboost** : À cette étape nous appliquons au *data set* un modèle de *gradient boosting* (*catboost*), puissant mais non-explicable au métier ;

- **Catégorisation des variables continues** : Une étude des valeurs de Shapley pour chaque échantillon et chaque variable, à partir du modèle *catboost* nous permet de transformer des données continues du type « durée du traitement de la réclamation » ou « montant de l'indemnisation » en variables de « *ranges* » (c'est-à-dire un segment de valeur continue - compris entre  $x$  et  $y$ ) plus pertinentes et plus performantes. Ainsi, la durée de traitement a été découpée en 3 *ranges* : moins de 48h, entre 48h et 96h, au-delà de 96h (valeurs données pour illustration).

- **Régression linéaire** : L'injection dans un modèle de régression linéaire simple, des nouvelles données obtenues par les deux premières étapes permet d'obtenir un modèle explicable. Grâce à cette étape, on définit aussi les séquences d'évènements permettant, par exemple, de déterminer le poids d'une journée supplémentaire de traitement d'une réclamation en fonction de sa position dans le processus.

- **Suppression de variables** : Le calcul des *p-values*, une mesure qui permet d'indiquer que la variable est statistiquement significative ou non, pour chaque variable de la régression linéaire, afin d'éliminer les variables non pertinentes de notre étude.

En répétant ce processus de manière itérative et en challengeant les résultats avec le métier, nous avons obtenu des résultats interprétables et évaluables par le métier. Ainsi nous avons pu identifier que le poids du montant d'indemnisation dans la LTV future du client est d'autant plus fort que ce client est fidèle. Nous avons par ailleurs isolé le nombre d'interactions optimales pour un client occasionnel.

## 3.4 UNE MÉTHODE QUI N'EXIGE PAS L'USAGE D'OUTILS COMPLEXES OU ONÉREUX

Pour cette étude, nous n'avons utilisé que deux environnements techniques, chacun répondant à une étape spécifique du processus CRISP DM (*Cross-Industry Standard Process for Data Mining*, cf. Diagramme des phases principales d'une méthodologie CRISP DM plus haut dans l'étude) :

- **Teradata Studio** : environnement technique permettant de se connecter à une base de données Teradata et d'exécuter des requêtes SQL. Outil du client, Teradata Studio a été utilisé lors de la phase de compréhension des données. Nous avons apprécié pouvoir utiliser des fonctionnalités NPATH (*Number of Paths* soit nombre de chemins) permettant de créer des « chemins » à partir de données datées, et de modéliser ces chemins sous forme de Diagramme de Sankey (présenté préalablement dans ce chapitre). Toutefois nous aurions pu utiliser tout autre outil de requête sur une base de données structurées.

- **Jupyter Notebook** : *Jupyter Notebook* est un environnement de développement *open source* interactif qui permet d'exécuter du code Python en temps réel, et d'intégrer du texte afin de réaliser des présentations rapides. Nous avons utilisé *Jupyter Notebook* lors de l'étape de modélisation pour écrire et exécuter du code Python permettant d'analyser les données, réaliser des modèles de prévision et visualiser les résultats. Les bibliothèques Python utilisées pour créer notre modèle sont *pandas*, *statsmodel*, *catboost* et SHAP.

Avec l'utilisation de *Jupyter Notebook*, nous avons pu nous appuyer sur une communauté active qui maintient l'outil au statut d'état de l'art en termes de *machine learning* et analyses statistiques. Les solutions nécessaires à la mise en œuvre de notre méthode sont donc tout à fait accessibles.

## 3.5 LE RÉSULTAT : UNE POLITIQUE DE SERVICE CLIENT OPTIMISANT LA LTV

Grâce à cette méthode itérative, la menée du projet a pu livrer des conclusions dès les premières versions du modèle. Ainsi, certaines décisions de politique clients concernant les compensations par typologie de réclamation et de clients ont pu être confirmées très tôt dans le processus. En particulier via la mise en place de forfaits compensatoires, plus élevés que précédemment, qui ont été validés grâce à l'obtention d'un ROI positif.

Ces *quickwins* ont permis aux sponsors du projet de s'assurer rapidement du caractère actionnable et opérationnel des conclusions du projet.

Une fois le modèle disponible dans sa version finale, ce sont des ajustements fins qui ont pu être opérés dans les ressources à disposition des équipes et dans les *process*. De fait, il a été possible de jouer sur les délais de traitement et le nombre d'interactions optimales par typologie de client et de réclamations. Le tout avec un ROI anticipé et précis pouvant être challengé à l'aune des chiffres réellement constatés.

Par ailleurs, le modèle a permis d'identifier des axes d'amélioration de l'expérience client, en particulier sur le renforcement des interactions avec les clients en fonction de leurs profils (un client régulier nécessitant de moins de

contact qu'un client occasionnel) et la personnalisation des échanges.

De plus, tel que construit, le modèle permet un suivi en « temps réel » de la performance du traitement des réclamations sur la LTV, notamment en intégrant ce modèle à un *dashboard* à disposition des métiers ; *dashboard* servant la mise à jour à fréquence régulière du modèle et le faisant ainsi rentrer dans une boucle de qualité.

## 3.6 LES POINTS-CLÉS MÉTHODOLOGIQUES

Au-delà de son ROI et de sa capacité de prise de décision opérationnelle, la méthodologie de mise au point de ce modèle a permis aux équipes de Keley de conforter certaines convictions :

- La capacité d'écoute et d'intégration des éléments du métier par les *datas scientists* est la clef de la mise en place de la boucle d'échange entre la vision métier et les apports de l'IA. Ces échanges ultra performants ont été particulièrement fructueux pour la détermination des *ranges* de durée de traitement des réclamations et la définition des chemins de résolution les plus performants.
- La qualification du *data set* est plus importante que sa taille. Dès lors, le temps passé sur ce point est du temps bien investi car il se traduit très rapidement dans de la performance prédictive.
- Les méthodes itératives assurent :
  - une meilleure compréhension du modèle, donc de l'opérationnalité du projet, par le métier et du métier par les *data scientists*,
  - l'identification de *quickwins* en avance de phase.

Moins ils mobilisent d'outils complexes, plus les chances de réussite dans le temps d'un projet data sont élevées.

# 04.

## UNE MÉTHODE ACTIONNABLE ET APPLICABLE À DE TRÈS NOMBREUX CAS D'USAGE



Comme nous l'avons vu précédemment, la méthode « Keley » permet d'associer prédiction juste et exploitation par le métier, notamment par le « poids » que l'on peut associer à chaque variable.

Dès lors, on peut appliquer cette méthode à de très nombreux cas *business*. En premier lieu, cette typologie de modèle peut être utilisée pour l'ensemble des calculs de score. Ainsi nous avons appliqué la même méthode pour une mission d'optimisation et d'industrialisation d'un score de transformation de devis. Sur la base des caractéristiques du devis et du prospect, mais aussi de ses actions, un score de probabilité de transformation est calculé pour être ensuite injecté dans le CRM *Salesforce* afin que les commerciaux relancent dans l'ordre décroissant de probabilité de transformation des prospects. Cette méthode peut être appliquée pour l'ensemble des *scores marketing*, du score *d'upsell* ou d'appétence à un produit à celui de rétention.

La rétention ou la prévention de l'attrition est d'ailleurs un autre sujet sur lequel la méthode est pertinente. En effet, même sur des cohortes de clients très importantes, les

modèles de régression linéaire traditionnels ont une très faible prédictibilité. En effet, par exemple la relation entre l'attrition et l'occurrence d'un sinistre, le délai de remboursement et le montant du reste à charge n'est pas linéaire. Pour saisir la relation entre ces variables et l'attrition, il nous faut utiliser la méthode du *boosting de gradient*, qui permet d'identifier les scénarios explicatifs.

Les moteurs dits de "*next best action*" ou "*next best offer*" sont un autre sujet plein de potentiel d'optimisation des ressources et des performances. Grâce à notre méthode, il est possible d'intégrer de manière fine plus d'éléments dans ces moteurs, notamment les interactions clients. Qu'il s'agisse de les intégrer dans un site e-commerce, une application de *selfcare* ou un outil de gestion des appels téléphoniques, ces modèles permettent d'adapter finement la proposition à la situation précise du client.

Et ces exemples ne sont qu'une partie des multiples débouchés de cette méthode qui peut aussi s'appliquer aux problématiques médias et permettre l'optimisation du *marketing mix modeling*, sujet capital en ces temps de modification des usages des clients.



VOTRE

**AVIS**

NOUS

**INTÉRESSE**

!