INTELLIGENCE ARTIFICIELLE ET RETAIL

AKUMENIA Data For Business Series
White Paper

RETAILERS : UN TERA DE MARKETEURS SUR LE TERRAIN DE VOS DATA !

De la vue 360° au profiling client autonome, un nouveau paradigme marketing retail pour une donnée en action!

Data 4 Business Series #1
15/04/2021

AKUMEN IA
VALYANS

1 PRÉAMBULE

De la vue 360° au profiling client autonome, un nouveau paradigme marketing retail pour une donnée en action!

La connaissance client a de tout temps constitué une verticale indispensable au maintien d'une activité de production et de commercialisation viable.

En effet, peu d'entreprises peuvent se pourvoir et maintenir une posture de "customer maker" – au sens où elles génèreraient d'elle-même, du fait d'un monopole organique issue d'une innovation par exemple, une capacité à créer de manière continue une demande sur un produit "monopensé", avec son lot de prospects qualifiés et convertibles en clients fidèles. Ceci est d'autant plus vrai pour des secteurs à forte intensité concurrentielle adressant une clientèle particulière large et pour lesquels, la connaissance client est clé, que ce soit pour conserver ses marges, ses parts de marché ou encore ses volumes de vente.

Une connaissance client affûtée et doublée, dans un contexte de numérisation croissante des activités et canaux, d'une connaissance des usages client demeurent de ce fait une nécessité. C'est pour cette raison précisément qu'il est essentiel d'affûter sa connaissance client. Longtemps cantonnée aux activités de marketing stratégique – pour la compréhension du marché et des produits – et opérationnelles – pour les (nouveaux) usages -, les paradigmes de connaissances clients évoluent aujourd'hui vers l'exploitation scientifique de la data pour une omniscience client analytique et prédictive.

Nous entendons ici présenter une approche conférant à l'exploitation de la donnée une valeur opérationnelle au-delà des méthodes surannées dites de "Business Intelligence" dans la composante marketing stratégique et de "fiche client 360°" dans la verticale opérationnelle. Lesquelles se limitent généralement à une utilisation réactive, figée et ex post de la donnée et laissant la décision à la merci de biais humain. La science des données répond à ces contraintes avec une précision accrue, une diminution des coûts inhérente à ces activités mais surtout, un avantage compétitif certain en termes de croissance et fidélisation.

En soi, repenser et exposer ses produits et services « par empathie avec le client » grâce à une connaissance précise et complète du client est de nos jours un besoin avéré. Toute la puissance de l'approche des sciences et technologies de données consiste à le faire scientifiquement avec une vue exhaustive, de manière autonome, à moindre coûts et large échelle et surtout sans biais ou a priori humain.

Une connaissance client affûtée et doublée, dans un contexte de numérisation croissante des activités et canaux, d'une connaissance des usages client demeurent de ce fait une nécessité. C'est pour cette raison précisément qu'il est essentiel d'affûter sa connaissance client. Longtemps cantonnée aux activités de marketing stratégique – pour la compréhension du marché et des produits – et opérationnelles – pour les (nouveaux) usages -, les paradigmes de connaissances clients évoluent aujourd'hui vers l'exploitation scientifique de la data pour une omniscience client analytique et prédictive. Nous entendons ici présenter une approche conférant à l'exploitation de la donnée une valeur opérationnelle au-delà des méthodes surannées dites de "Business Intelligence" dans la composante marketing stratégique et de "fiche client 360°" dans la verticale opérationnelle. Lesquelles se limitent généralement à une utilisation réactive, figée et ex post de la donnée et laissant la décision à la merci de biais humain.

La science des données répond à ces contraintes avec une précision accrue, une diminution des coûts inhérente à ces activités mais surtout, un avantage compétitif certain en termes de croissance et fidélisation. En soi, repenser et exposer ses produits et services « par empathie avec le client » grâce à une connaissance précise et complète du client est de nos jours un besoin avéré. Toute la puissance de l'approche des sciences et technologies de données consiste à le faire scientifiquement avec une vue exhaustive, de manière autonome, à moindre coûts et large échelle et surtout sans biais ou a priori humain.

Copyright © 2021 by AKUMENIA S.A.

2 DU COST AU REVENUE CENTER? UN NOUVEAU PARADIGME

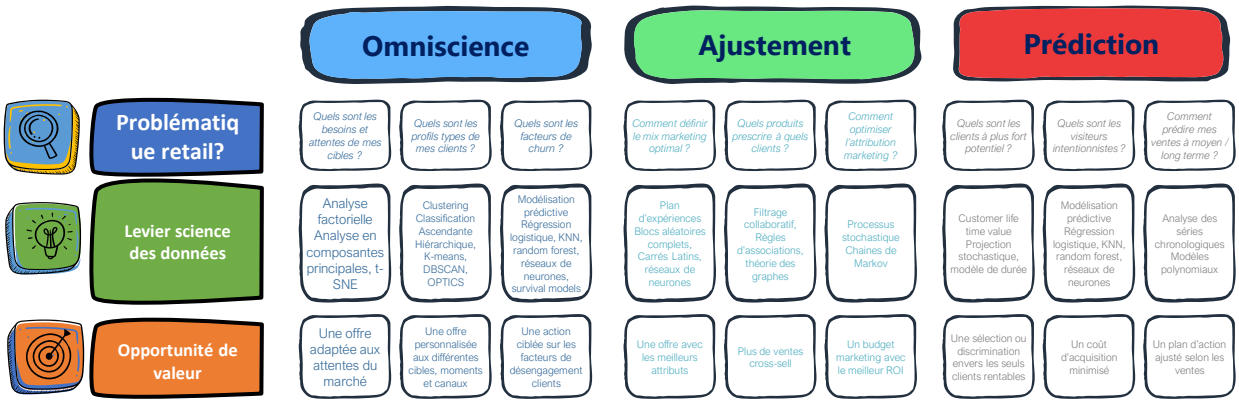
Désengluer la donnée de sa valeur de riposte pour la porter au rang de décideur quasi-autonome et atteindre une omniscience client

« Penser la science des données comme outil de profiling client se veut à la fois être un engagement envers le client, au centre de notre analyse, et un vecteur de croissance. Les technologies de stockages et de traitement distribué permettent de nos jours d'acquérir et d'exploiter au sein d'un même gisement l'ensemble des données générées par un client, furent-elles structurées (signalétique, CRM, tickets de caisses) ou non structurées (logs de navigation en ligne, audio avec le call center, images).

Au-delà la constitution du gisement hétérogène, c'est le traitement massif des données qui est rendu possible par le calcul distribué – dont une définition simple serait la parallélisation de la résolution de plusieurs parties d'un algorithmes sur différentes machines en simultané, idéalement in memory. Disposer de cette capacité octroie à la science des données la pleine capacité de déployer son potentiel métier. En effet, si les modèles de Machine Learning trouvent un nouvel essor c'est principalement du fait que leur déploiement et orchestration dans le temps est aujourd'hui aisé et rentable

Dès lors, l'analyse du client ne se veut plus être une simple agrégation de à la maille individuelle (fiche client 360°) ou à la maille d'une ligne d'activité (reporting BI) mais peut trancher avec ces usages. Nous ne remettons pas en question leur intérêt éprouvé nécessaire à la prise de décision mais nous démontrons des expériences nombreuses ayant permis d'automatiser un part de l'intelligence opérationnelle humaine vers des IA. Autrement dit, aborder le client aujourd'hui du point de vue de la donnée c'est être en mesure de demander à un algorithme, éminemment étudié et testé par des experts métier, de se prononcer sur l'action à mener et la soumettre à la ressource humaine ou technique cuter. opérationnelle en charge de l'exécuter. Le marketeur quant à lui voit son métier, son organisation et ses équipes se transformer petit à petit de réalisateur d'études qu'il s'agira d'analyser pour décider en orchestrateur de définition, développement et activation de cas d'usages de science des données. Arborer un profiling client automatique et complet est permis par la maturité des approches listées ci-dessous mais nous prendrons pour exemples deux exemples d'initiatives nouvelles pour illustrer le propos : un cas analytique via une segmentation comportementales simple et un prédictif sous formes la détection des moments de vie clients.

Illustration 1 – Problématique retail? Une réponse data science pour des opportunités de création de gisement de valeur ?

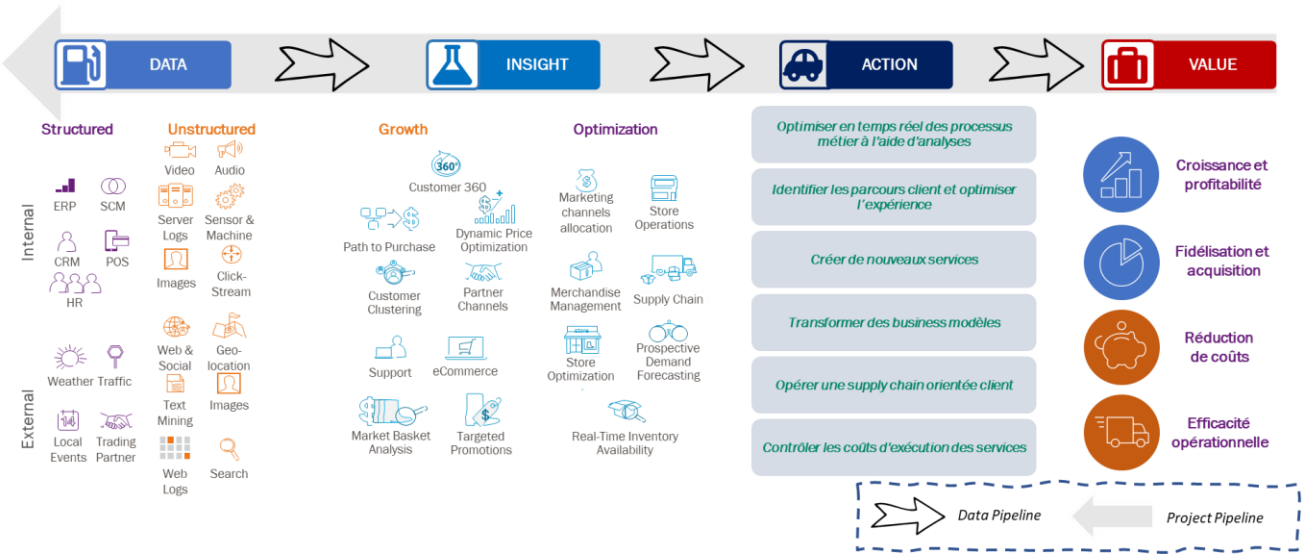


Copyright © 2021 by AKUMENIA S.A.

Nous allons à présent commenter deux exemples d'initiatives menées illustrant le propos en exposant deux approches de « customer profiling » complémentaires volontairement re-transcriptibles dans différents secteurs.
La première, une segmentation comportementale fine, ayant été menée avec succès dans différentes industries, de la grande consommation à la bancassurance. La seconde, l'analyse prédictive des moments de vie client est quant à elle plus spécifique à des activités proposant des produits multiples associées à un cycle de vie produit et client – c'est le cas de l'automobile, des institutions de crédit ou encore le tourisme.

Là où beaucoup s'engouffrent dans une description de ces cas d'usages orientée technologie et cycle de vie de la donnée, il s'agit dans une approche ROIste d'attaquer la problématique d'abord par la valeur et l'objectif stratégique, puis les actions d'activation cible du cas d'usages avant de se plonger dans la technique de modélisation d'insight et in fine remonter aux données sources. Se lancer dans l'IA quand la donnée n'est pas le cœur de métier de l'entreprise implique en effet, surtout si l'on veut que l'initiative soit rentable à court terme, d'attaquer le sujet par le métier et non pas par la technologie.

Illustration 2 –Un pipeline projet à l'inverse du pipeline data pour des cas d'usages pilotés par les besoins métiers et une croissance des coûts proportionnelles aux bénéfices, transformant l'usage analytique de la donnée d'un centre de coûts à un en centre de revenu



3 QUAND UN AGLORITHME GÉNÈRE PLUSIEURS DIZAINES DE MILLIONS EN 15 JOURS

Je ne segmente pas les clients, je segmente des relations clients, des comportements d'achats

❖ **Valeur :**

Une segmentation comportementale répond à plusieurs besoins stratégiques qui ont tous pour but final une meilleure connaissance client pour une adéquation produit – client – canal – message – prix.

Cette approche dépasse le paradigme des segmentations historiquement basée sur le CA / Client, ou encore les segmentations RFM. Une segmentation comportementale en ce sens, va s'extraire des dénominations segments S1 à S5 ou encore Récent x Fréquent x Faible montant pour décrire des segments précis d'un point de vue métier et produit, on parlera par exemple du segment "épargnant conservateur en recherche d'investissement locatif" pour une banque et du segment "fin gourmet paresseux en refus de cuisine" pour la distribution alimentaire. Il apparait assez évident que mon épargnant peut potentiellement être un S1 ou un S3 historique, un simple croisement avec cette

Copyright © 2021 by AKUMENIA S.A.

ancienne segmentation me permet alors d'être pertinent dans mon action commerciale. Concernant la détection préventive des moments de vie client, il s'agit de pouvoir adosser à mon action une composante temps. On obtient ainsi avec les deux cas d'usages une optimisation marketing automatisée produit – client – canal – message – prix – temps.

❖ **L'activation :**

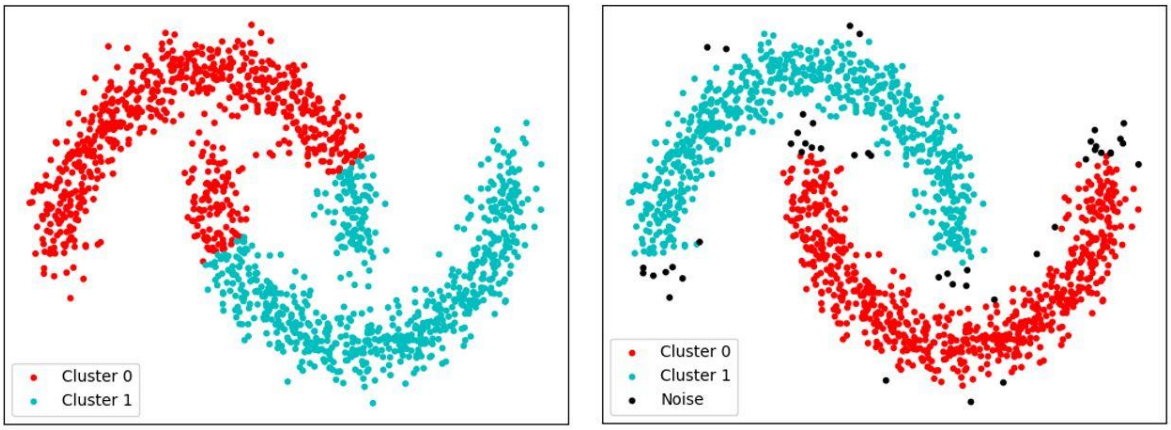
Bénéficier à l'échelle de l'ensemble de mes clients ou prospects – nous noterons une différence dans les deux paragraphes suivants entre ces deux cibles distinctes – permet d'actionner des approches marketing nouvelles et automatisées. On parlera alors de marketing automation. Mais pas seulement, le résultat de tels cas d'usages peut tout aussi bien être mené par des opérationnels humains en complément ou substitut des systèmes opérationnels. Ainsi, l'activation du cas d'usage consiste pour le marketeur à imaginer, en fonction des résultats obtenus, des matrices de décision qu'il prodiguera à l'algorithme pour transmission à l'opérationnel. On pourra alors imaginer pour certains segments des emailings, newsletter, rebond marketing ou campagnes SMS extrêmement personnalisés et ayant de facto un plus haut taux de conversion. On pourra également extraire des briefs clients ou prospects qualifiés aux commerciaux augmentant ainsi leur productivité et le taux de transformation. Pour la segmentation, il s'agit d'identifier, via un second algorithme ou à la main, le produit idéal au sein du portefeuille. A titre d'exemple, proposer à l'épargnant conservateur un produit de placement même s'il n'est que S3 et que le produit était généralement consommé par le S1/S2. Les moments de vie pourront quant à eux proposer du lait pour enfant ou un crédit auto si l'on identifie quelques mois à l'avance, parfois avant même que le client ne le sache, qu'il aura besoin de changer de véhicule.

❖ **La modélisation et l'insight :**

Pour y arriver, les approches de science de la donnée vont pouvoir utiliser deux types d'approches en machine learning, différentes dans leur philosophie mais complémentaires dans leurs usages. D'une part, l'approche par apprentissage supervisé, permet à l'algorithmes de détecter des schémas de données similaires à un jeux de données d'entrainement dit labellisé, c'est à dire qu'à l'image de l'enfant, l'algorithme apprends que ceci est un chat et ceci est un chien grâce à l'homme et pourra dans l'avenir identifier de manière autonome un chat, un chien ou encore un épargnant conservateur. D'autre part, nous utilisons également des approches dites non supervisées, où il revient à l'algorithmes d'identifier lui-même les variables discriminantes pour segmenter ou classifier les clients sans biais humain.

Notre segmentation comportementale utilisera les deux approches : d'abord un algorithme non supervisé, dont k-means, optimisé par un PCA et un Silhouette Scoring qui est le représentant le plus connus mais également DBSCAN et surtout OPTICS qui permettront d'outrepasser les logiques de distances euclidiennes pour un clustering par densité. Cette approche est particulièrement recommandée pour les données de multiples natures évoluant sur des espaces avec plusieurs dimensions.

Illustration 3 – OPTICS & DBSCAN contre k-means : quand un espace vectoriel n'a aucune chance de produire des sous-espace ou clusters convexes, préférer le clustering par densité.



K-means OPTICS

Copyright © 2021 by AKUMENIA S.A.

Une fois les premiers segments trouvés par l'ordinateur, l'humain analyse leurs caractéristiques et peut décider de les réunir si trop peu porteurs ou à l'inverse de les dissocier si trop importants, en réitérant la première opération. Ces segments homogènes figés, il s'agira enfin d'affecter le N+1 ième client ou de réaffecter les clients évoluant à intervalle régulier via un algorithme supervisé, ayant appris de nos clusters initiaux. Nous utiliserons alors alternativement et selon les contextes une regression logistique ou un Random Forrest. Enfin, pour la détection des moments de vie, une catégorie d'apprentissages supervisés empruntés au machine learning appliqué à la biologie, les modèles de survies – Cox Proportional-Hazards ou Kaplan Meier – peuvent s'avérer extrêmement efficaces.

❖ **Les données sources:**

Restent enfin nos données sources et pour cela, rien qui ne soit pas spécialement accessible n'est nécessaire pour les clients. Les données signalétiques, transactionnelles – compte ou tickets de caisses ainsi que les interactions online ou offline. Pour les prospects, on parlera de segmentation d'audience et, si cela n'est pas d'ores et déjà utilisé pour la partie client, les logs de navigations, notamment les ID Cookies seront nécessaires. Evidemment, l'usage de données achetées sur des DMP ou des bases de prospections est un plus que l'algorithme qualifiera. Au passage, pour les entreprises peu digitalisées comme pour les autres, rien n'empêche d'appliquer les résultats à l'offline, orientant ainsi PLVs et efforts média en fonction des résultats des modèles et des objectifs en cours.

PS : pensez à également collecter, stocker et mettre à jour les opt-ins de vos clients pour le traitements de leurs données personnelles. Veillez à mettre en place les systèmes qui permettront d'appliquer leur droit à la portabilité et à l'oubli. Documenter vos traitements dans des registres et maintenez les à jour. Le traiitement des données personnelles se fait avec le consentement du client contre un service rendu et proportionnel à la collecte et au traitement.

CE QUI N'EST PAS QUANTIFIABLE N'EXISTE PAS

4

Quantifier, c'est pouvoir valoriser, fédérer et prouver par le concept la plus value de l'approche.

Il existe trois approches scientifiques de validation des modèles :

- La validation sur historique : j'entraine mon modèle sur une partie de mes données historiques en le laissant aveugle d'une part des données, dont je connais l'issue, dites de tests. Je rejoue mon modèle sur cet historique et j'analyse à quel point mon modèle a pu détecter l'issue, par exemple, un moment de vie.
- La validation sur échantillon non animé : j'entraine mon modèle sur toutes mes données mais je ne l'active d'un point de vue métier que sur une partie des ma clientèle, en conservant un échantillon témoins représentatifs. Je compare in fine ma performance avec et sans cas d'usages science des données.
- La validation par A/B testing : variante de la précédente, elle permet de comparer deux modèles concurrents en analysant les résultats de chaque activation.

Ainsi, l'option 2 a été utilisée sur la base d'une segmentation comportementale simple pour un distributeur agroalimentaire. Les clients bénéficiant de la segmentation ont reçu une offre directement en phase avec les besoins de leur segment là où les témoins ont reçu l'offre nouveauté produits généraliste. Le résultat de la clientèle animée par rapport à l'échantillon témoin non activé a été de + 5,4 de CA en deux semaines, correspondant à une augmentation de 12% de panier moyen en valeur.

En synthèse et au-delà des gains de coûts, de l'automatisation du marketing ou encore de l'expérience client augmentée, l'approche est surtout une première application scientifique du marketing en retail et donc une approche moins faillible:

- Moins faillible car omnisciente : basée sur toutes les données au lieu des classiques échantillons ou enquêtes
- Moins faillible car non humainement biaisée : l'apprentissage non supervisé n'a aucun a priori humain ou biais cognitif lors de l'analyse.
- Moins faillible car scientifiquement testée : des approches de back-testing permettent de quantifier le succès avant même de déployer.

Copyright © 2021 by AKUMENIA S.A.

A PROPOS D'AKUMENIA

AKUMENIA S.A

Elevating Data to a Strategic Asset – Entreprise d'expertise en stratégie, sciences et Technologie de la donnée

AKUMENIA est née d'un double constat simple : la donnée est le nouveau capital de l'entreprise et la matière première du Digital. La Technologie sans l'expertise métier est un cost center à long cycle de vie, là où les approches data se veulent ROIstes à très court terme.

SI la data devient un actif stratégique, c'est justement parce qu'elle acquiert aujourd'hui une valeur opérationnelle nouvelle apte à dépasser l' "après coup" ex post de la BI traditionnelle et ce, grâce au pouvoir analytique et prédictif de la science des données rendu possible par des capâcités accrues de technologies de traitement massifs de données distribuées et notamment en temps réel.

Akumen IA entend à la fois penser avec ses partenaires les cas d'usages data de demain et leur donner vie grâce à une maitrise pointue de ces nouvelles technologies sous l'oeil avisé d'experts métiers et sectoriels.

A PROPOS DE L'AUTEUR

Soulaimane Lahrech

Partner AKUMEN IA – Directeur Associé en charge des Sciences et Technologies de la donnée

Diplômé de l'ESSEC et de l'Université de Berkeley, Soulaimane a construit sa carrière au cœur des grands cabinets de conseils parisiens spécialistes des Big Data et de la Data Science. Après de nombreuses années au services des grands groupes français, allemands et américains de tout secteur (banque, automotive, retail, santé etc.) sur des projets de mise en place de plateformes Big Data alliés à des applications IA et Data Science, Soulaimane retourne dans au sein de son pays et continent d'origine pour contribuer à un transfert de reverse innovation auquel IL croit, et rejoint AIOX Labs en tant que COO et Big Data lead puis cofonde aux côtés de Valyans, partenaire historique, AkumenIA.

Les applications data de Volkswagen Group, BNPP ou encore Sanofi font partie de ces principaux faits d'armes.

Soulaimane est convaincu que le succès de projets IA et Big Data doit répondre à une méthodologie et une démarche plaçant les besoins métiers et les use cases data comme clé de voute du projet, tout en respectant les grands principes de protection des données personnelles et sécurité des infrastructures. Il milite par ailleurs pour l'émergence d'IA africaines pour l'Afrique.

Soulaimane est égalementdata architecte, data privacy officer et enseignant dans des Universités marocaines et internationales

Data 4 Business Series #1

15/04/2021

DATA4BUSINESS

AKUMEN IA

VALYANS