

L'APPRENTISSAGE AUTOMATIQUE DANS LA SEGMENTATION DU PORTEFEUILLE CLIENT

HERMANN DJAMENI MONKAM

1^{er} janvier 2023

Table des matières

| | |
|---|----------|
| 1 INTRODUCTION | 1 |
| 2 LA SEGMENTATION CLIENT | 2 |
| 2.1 Définition | 2 |
| 2.2 Les Critères de segmentation client | 2 |
| 2.2.1 Les critères géographiques | 2 |
| 2.2.2 Les critères psychographiques | 2 |
| 2.2.3 Les critères comportementaux | 3 |
| 2.2.4 Les critères démographiques | 3 |
| 2.3 La segmentation client à l'ère d'internet | 3 |
| 3 L'APPRENTISSAGE AUTOMATIQUE DANS LA SEGMENTATION CLIENT | 4 |
| 3.1 Quelques Notions sur l'apprentissage Automatique | 4 |
| 3.1.1 Définition | 4 |
| 3.1.2 Les données | 5 |
| 3.1.3 L'apprentissage Non Supervisé | 5 |
| 3.1.4 L'apprentissage Supervisé | 5 |
| 3.2 Enjeux de l'apprentissage automatique pour la segmentation client | 6 |
| 3.3 Importance de l'apprentissage automatique dans la segmentation client | 6 |
| 3.4 Algorithme de Machine Learning utilisé pour la segmentation Clientt | 7 |
| 3.4.1 Algorithme de Clustering | 7 |
| 3.4.2 L'Algorithme K-Means : Présentation | 8 |

| | | |
|----------|--|-----------|
| 3.4.3 | Fonctionnement de K-Means | 8 |
| 3.4.4 | Notion de Similarités | 9 |
| 3.4.5 | Détermination du nombre de cluster : K | 9 |
| 4 | CONCLUSION | 10 |

1 INTRODUCTION

L'une des innovations technologiques majeures qui révolutionne le secteur du Marketing Digital depuis quelques années est l'utilisation d'outils d'intelligence artificielle pour aider à rationaliser les processus de marketing et à rendre les entreprises plus efficaces. De ce fait, de nombreuses entreprises aujourd'hui utilise l'apprentissage automatique (Maching Learning) pour la segmentation du portefeuille client, car ceux si se sont rendu compte que, considérer tous les clients comme identiques et les cibler tous avec une stratégie marketing similaire n'est pas un moyen très efficace, au contraire, cela agace les clients en négligeant leur individualité. C'est dans ce sens que le thème soumis à notre étude porte sur L'apprentissage automatique dans la segmentation du portefeuille client. Il sera donc question pour nous de répondre à la question de savoir : En quoi l'apprentissage automatique, peut-il etre utile pour la segmentation du portefeuille clients de l'entreprise d'aujourd'hui? .Nous présenterons premièrement les notions de segmentation client, de machine Learning, puis l'enjeu et son importance dans la segmentation du portefeuille clients. Nous allons également présenter l'algorithme à utiliser pour mener à bien cette segmentation.

2 LA SEGMENTATION CLIENT

2.1 Définition

La segmentation est l'opération qui consiste à découper une population d'individus en sous-ensembles homogènes ou pour être plus précis, en sous-ensembles partageant un ou plusieurs attributs communs. Chaque sous-ensemble constitue un segment. La segmentation permet d'améliorer le ciblage et aux équipes opérationnelles de prioriser leurs efforts sur les clients, les segments les plus intéressants. C'est une méthode pour mettre de l'intelligence dans une base de données en l'occurrence une base de données prospects, clients ou utilisateurs. La segmentation est une opération de division, de découpage, bref une opération d'analyse. En un sens, la segmentation est l'opération intellectuelle de base (celle qui précède la synthèse). Segmenter permet de classer les individus, de savoir de quoi est faite sa base de

données clients, quels sont les grands ensembles d'individus qui la composent.

2.2 Les Critères de segmentation client

Si la segmentation est l'art de créer des ensembles réunissant des individus partageant des points communs, la question reste de savoir quels sont ces points communs. Il existe différentes familles de points communs, différentes catégories d'attributs. Lesquels forment les critères utilisés pour la segmentation client, et peuvent être regroupés en 4 catégories, notamment :

2.2.1 Les critères géographiques

Les critères géographiques sont parfois considérés comme une sous-partie des critères sociodémographiques. Inutile de trop s'attarder sur cette famille de critères. Voici quelques attributs géographiques utilisés en marketing : le pays, la région, le département, la commune, mais aussi le climat, le type d'habitat, etc. Ces critères ne doivent pas être sous-estimés, car le lieu où l'on habite a une influence sur nos comportements, notre style de vie, nos activités

2.2.2 Les critères psychographiques

Les critères psychographiques englobent tous les attributs décrivant le style de vie, les préférences, les centres d'intérêt, la personnalité, les opinions ou encore les valeurs des individus. Avec cette famille de critères, on segmente les individus non pas en fonction de ce qu'ils sont ou de la manière dont ils se comportent, mais en fonction de ce qu'ils pensent au sens large du terme. Ces attributs permettent de déduire les aspirations, les objectifs, les attentes et les motivations clients. Ils sont donc très utiles.

2.2.3 Les critères comportementaux

Ici, on s'intéresse à la manière dont l'individu se comporte avec les supports digitaux de l'entreprise :

- Sur le site web : pages vues, durée des sessions, nombre de sessions, date de dernière session, source de trafic, etc.
- Vis-à-vis des campagnes emails : ouverture, clics, etc.
- Sur l'application mobile : sessions, durée des sessions, fonctionnalités utilisées...

2.2.4 Les critères démographiques

La démographie est un autre critère de segmentation très simple à appréhender. La segmentation démographique consiste à diviser un marché en fonction de critères comme : l'âge, la race, la situation

familiale, le genre, la nationalité, les revenus, l'éducation, etc.

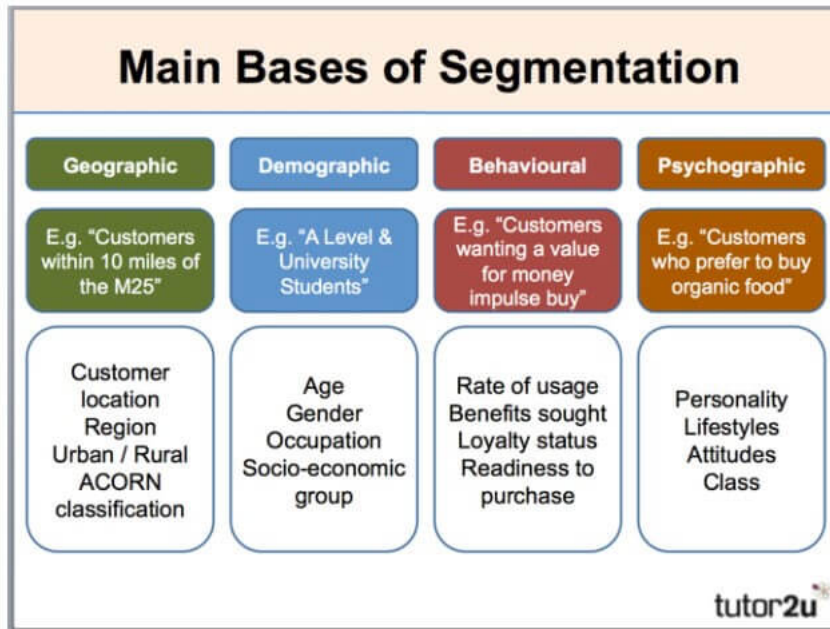


FIGURE 1 – Récapitulatif des grandes familles de critères de segmentation (Source : <https://www.cartelis.com/wp-content/uploads/2019/11/segmentation-client-exemple-7.jpg>)

2.3 La segmentation client à l'ère d'internet

Plus de doute sur le fait que l'internet contrôle tout aujourd'hui, cela se remarque aussi au niveau de la segmentation client. Celui-ci permet aux entreprises d'avoir facilement accès à leur data. Cette segmentation se remarque partout, que ce soit dans les médias, sur internet ou dans les sociétés. Il est donc plus facile maintenant pour les entreprises de se procurer des informations sur leur client. L'internet a rendu tout possible et simple. Grâce à l'internet, il y en a qui prédisent même ce que les clients achèterons. Ainsi, pour encore plus motiver le client, ils utilisent des publicités et des mails et surtout les techniques basées sur le machine learning.

3 L'APPRENTISSAGE AUTOMATIQUE DANS LA SEGMENTATION CLIENT

3.1 Quelques Notions sur l'apprentissage Automatique

3.1.1 Définition

L'Apprentissage automatique (machine learning) et l'intelligence artificielle sont deux entités distinctes complémentaires : alors que l'intelligence artificielle (IA) vise à exploiter certains aspects de l'esprit (Thinking Mind), le machine learning également appelé apprentissage automatique aide les spécialistes à résoudre les problèmes de manière plus efficace. Considéré comme un champ d'étude de l'intelligence artificielle, le machine Learning exploite les données statistiques et mathématiques pour apprendre aux ordinateurs à améliorer leurs performances, à se renseigner sur la manière de terminer un processus et à résoudre des tâches à l'aide des capacités de l'IA. Le Machine Learning permet grâce aux données de fournir aux entreprises des solutions efficaces à une multitude de problèmes complexes de marketing digital et ainsi d'aider à trouver des connaissances cachées dans les données des consommateurs disponibles pour rationaliser les processus marketing.



FIGURE 2 – Machine Learning (Source : <https://www.50a.fr/0/machine-learning>)

3.1.2 Les données

les algorithmes de l'apprentissage automatique sont basés sur des données. On parle aussi d'échantillons (samples), d'observations. Concrètement, cela signifie que le jeu de données (dataset) est formé d'un certain nombre de données, par exemple : d'articles de journaux, d'images de chiens et chats, ou de caractéristiques de logements. Nous noterons la taille du jeu de données N , chaque observation x_n et le jeu de données de N observations (x_n).

Deux grandes familles de jeux de données peuvent être utilisées :

— **les données étiquetées** : chaque observation x_n est fournie avec une étiquette (label) y_n ;

— **les données non étiquetées** comme le nom l'indique, aucune étiquette n'est fournie.

Trois grandes approches relèvent de l'apprentissage automatique : l'apprentissage supervisé, l'apprentissage non-supervisé, et l'apprentissage par renforcement. Nous allons nous intéresser principalement à l'apprentissage supervisé et l'apprentissage non-supervisé.

3.1.3 L'apprentissage Non Supervisé

L'apprentissage non-supervisé (unsupervised learning) traite des données non-étiquetées. L'objectif est d'identifier automatiquement des caractéristiques communes aux observations. Ici, la machine se contente d'explorer les données à la recherche d'éventuelles patterns. Elle ingère de vastes quantités de données, et utilise des algorithmes pour en extraire des caractéristiques pertinentes requises pour étiqueter, trier et classer les données en temps réel sans intervention humaine. Plutôt que d'automatiser les décisions et les prédictions, cette approche permet d'identifier les patterns et les relations que les humains risquent de ne pas identifier dans les données. Cette technique n'est pas très populaire, car moins simple à appliquer. Elle est toutefois de plus en plus populaire dans le domaine de la cybersécurité.

3.1.4 L'apprentissage Supervisé

Dans le cas de l'apprentissage supervisé, le plus courant, les données sont étiquetées afin d'indiquer à la machine quelles patterns elle doit rechercher. Le système s'entraîne sur un ensemble de données étiquetées, avec les informations qu'il est censé déterminer. Les données peuvent même être déjà classifiées de la manière dont le système est supposé le faire. Cette méthode nécessite moins de données d'entraînement que les autres, et facilite le processus d'entraînement puisque les résultats du modèle peuvent être comparés avec les données déjà étiquetées. Cependant, l'étiquetage des données peut se révéler onéreux. Un modèle peut aussi être biaisé à cause des données d'entraînement, ce qui impactera ses performances par la suite lors du traitement de nouvelles données.

3.2 Enjeux de l'apprentissage automatique pour la segmentation client

Les méthodes de segmentation courantes montrent cependant certaines limites face à l'évolution des parcours d'achat des individus. Il n'est pas rare qu'un consommateur se renseigne au sujet d'un produit ou d'une marque sur le web avant de se rendre en magasin pour l'acheter. Ainsi, il est indispensable de tenir compte de l'influence du web sur les achats en magasin dans le processus de segmentation. Savoir reconnaître un consommateur qui achète en ligne quand il est déjà client en magasin est capital. C'est là que le Machine Learning entre en jeu. Les algorithmes prennent alors le relais dans cet univers complexe

afin de trouver les clusters d'individus similaires tant convoités. Une fois ces groupes découverts par l'algorithme, l'humain (data scientist) reprend la main pour interpréter les résultats et identifier les facteurs communs qui font que les individus se sont retrouvés dans un même groupe.

3.3 Importance de l'apprentissage automatique dans la segmentation client

Nous présenterons ici, l'importance du machine Learning dans la segmentation du portefeuille client à travers un exemple. Supposons qu'un spécialiste du marketing est chargé de développer une idée de la façon de faire la publicité d'une nouvelle Mercedes et décide de faire la publicité que la voiture est abordable; ce dernier passerait totalement à côté de l'essentiel car le segment de marché pour ce type de véhicule n'est pas quelqu'un qui recherche des véhicules abordables ou bon marché, mais quelqu'un qui est riche et qui ne craint pas que les autres le sachent. Ainsi, cette stratégie de marketing éloignerait les clients et entraînerait une perte de confiance en cet employé et une perte de budget et de crédibilité pour l'agence de marketing. D'où l'importance de l'apprentissage automatique, qui permet d'affiner cette segmentation client et d'obtenir de meilleurs résultats à terme, grâce aux multiples avantages qu'elle offre;notamment :

- **Gain en temps** : la segmentation manuelle, surtout s'il y a beaucoup de données (et que vous recherchez des modèles spécifiques) prend énormément de temps. L'apprentissage automatique libère du temps pour les spécialistes du marketing afin qu'ils puissent se concentrer sur des tâches plus exigeantes nécessitant plus de créativité et une réflexion complexe.

- **Évolutivité** : les modèles d'apprentissage automatique prennent en charge l'évolutivité. Ils peuvent travailler maintenant pour 10 000 clients, mais même si l'entreprise en gagne 1 million de plus, le modèle peut gérer les nouvelles données et les analyser rapidement.

- **Plus grande précision** : les performances des modèles d'apprentissage automatique pour la segmentation de la clientèle sont bien meilleures lorsqu'il s'agit d'ensembles de données étendus.

3.4 Algorithme de Machine Learning utilisé pour la segmentation Clientt

3.4.1 Algorithme de Clustering

Les algorithmes de Machine Learning, et plus précisément ceux appartenant à la famille des algorithmes de Clustering, répondent à cette problématique de segmentation. En effet, ils ont pour objec-

tif d'identifier des groupes d'objets similaires et peuvent donc être utilisés pour segmenter une base de données clients. Ces programmes sont capables de prendre en compte beaucoup plus de variables qu'un simple humain et permettent donc de trouver de nouvelles corrélations entre ces dernières et donc de nouveaux segments de clients.

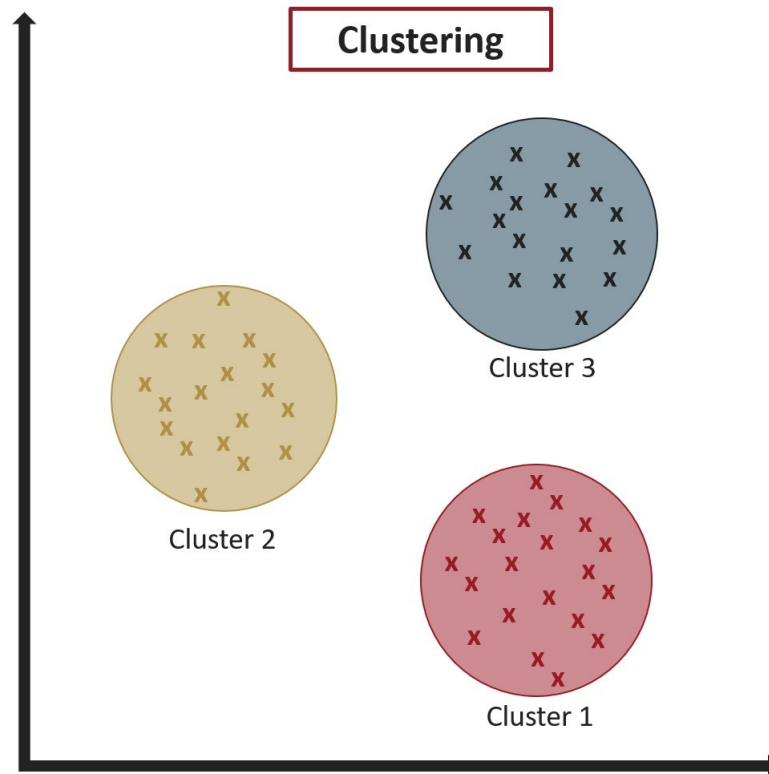


FIGURE 3 – Clustering (source : <https://fr.linkedin.com/pulse/pourquoi-utiliser-le-machine-learning-dans-une-strategie-lebec>)

3.4.2 L'Algorithme K-Means : Présentation

K-Means est un algorithme de machine Learning qui permet de regrouper des individus. C'est un 'algorithme qui est très utilisé dans la segmentation de la clientèle. K-Means crée des groupes d'individus homogènes (clusters) à partir des données proposées. En machine learning, il est utilisé pour le partitionnement des données en fonction des ressemblances et en fonction du set clustering. Il fait intervenir une technique d'apprentissage automatique non supervisée.

Cet 'algorithme permet d'administrer un traitement différent en tenant compte des profils d'une population cible. À partir d'un ensemble de données et de K groupes, cet algorithme d'apprentissage automa-

tique non supervisé permet de segmenter différents éléments en plusieurs Groupes. Ce regroupement est réalisé en minimisant la distance euclidienne entre un objet donné et le centre du cluster. La constitution des clusters place une fonction sous le principe de l'exclusivité d'appartenance. En d'autres termes, une même donnée ne peut être retrouvée dans deux différents groupes. Ici, les algorithmes ne sont pas programmés pour prédire une certaine valeur en se basant sur une analyse.

K-Means permet plutôt de déterminer des patterns dans les données afin de les rassembler selon les similarités. Comme tout algorithme, K-Means possède un mode de fonctionnement bien défini.

3.4.3 Fonctionnement de K-Means

- K-Means est un algorithme itératif qui minimise la somme des distances entre le centroïde et les individus. Le résultat final est conditionné par le choix initial des centroïdes. Il s'agit de l'élément central de l'algorithme. C'est un point du jeu de donnée qui sera désigné comme le centre d'un cluster. L'appartenance à un cluster sera donc définie en fonction d'un centroïde. En ce qui concerne la distance, c'est un élément de l'algorithme qui associe un nombre réel positif à un couple de vecteurs. La distance la plus connue est la distance euclidienne. Elle est utilisée comme mesure de similarité dans la plupart des techniques de clustering.

- Dans un ensemble de points par exemple l'algorithme de clustering change les points de chaque groupe jusqu'à diminution de la somme. En choisissant la bonne valeur K du nombre de clusters, on obtient un ensemble de groupe clairement séparé et compact. K-Means est généralement utilisé pour l'analyse des données quantitatives. L'algorithme identifie dans un ensemble de données un certain nombre de centroïdes. C'est la moyenne arithmétique de tous les objets de données qui appartiennent à un cluster. Chaque point de donnée est attribué au cluster le plus proche. L'algorithme essaie au maximum de maintenir les clusters aussi petits que possible. Dans le même temps, les autres groupes sont maintenus aussi différents possible.

- L'algorithme de machine Learning lance l'initialisation de plusieurs centres de clusters de façon aléatoire. Chaque point est assigné à son centre de clusters le plus proche à chaque passage de l'algorithme. Les centres sont ensuite mis à jour à travers un calcul.

L'algorithme k-Means se répétera jusqu'à ce que l'on obtienne un changement minimum des centres de clusters. .

3.4.4 Notion de Similarités

Pour pouvoir regrouper un jeu de données en K cluster distincts, l'algorithme K-Means a besoin d'un moyen de comparer le degré de similarité entre les différentes observations. Ainsi, deux données qui se

ressemblent, auront une distance de dissimilarité réduite, alors que deux objets différents auront une distance de séparation plus grande :

La distance Euclidienne : C'est la distance géométrique. Soit une matrice (X) à (N) variables quantitatives. Dans l'espace vectoriel (E) , la distance euclidienne (d) entre deux observations x1 et x2 se calcule comme suit :

$$d(x_1, x_2) = \sqrt{\sum_{j=1}^n (x_{1j} - x_{2j})^2}$$

3.4.5 Détermination du nombre de cluster : K

Choisir un nombre de cluster K n'est pas forcément intuitif. Spécialement quand le jeu de données est grand et qu'on n'ait pas un a priori ou des hypothèses sur les données. Un nombre grand peut conduire à un partitionnement trop fragmenté des données. Ce qui empêchera de découvrir des patterns intéressants dans les données.

Par contre, un nombre de clusters trop petit, conduira à avoir, potentiellement, des clusters trop généralistes contenant beaucoup de données. Dans ce cas, on n'aura pas de patterns à découvrir. Pour un même jeu de données, il n'existe pas un unique clustering possible. La difficulté résidera donc à choisir un nombre de cluster qui permettra de mettre en lumière des patterns intéressants entre les données. Malheureusement il n'existe pas de procédé automatisé pour trouver le bon nombre de clusters. La méthode la plus usuelle pour choisir le nombre de clusters est de lancer K-Means avec différentes valeurs de K, et de calculer la variance des différents clusters. La variance est la somme des distances entre chaque centroid d'un cluster et les différentes observations incluent dans le même cluster. Ainsi, on cherche à trouver un nombre de clusters de telle sorte que les clusters retenus minimisent la distance entre leurs centres (centroids) et les observations dans le même cluster. On parle de minimisation de la distance intra-classe. **La variance des clusters se calcule comme suit :**

$$V = \sum_j \sum_{x_i \rightarrow c_j} D(c_j, x_i)^2$$

Avec : $c(j)$: Le centre du cluster (le centroïd) $x(i)$: la i ème observation dans le cluster ayant pour centroïd $c(j)$ $D(c(j), x(i))$: La distance (euclidienne ou autre) entre le centre du cluster et le point $x(i)$

4 CONCLUSION

Parvenu au terme de notre travail, il était question pour nous d'étudier l'utilité de l'apprentissage automatique (machine Learning) dans la segmentation du portefeuille client. Nous avons présenté les notions de machine Learning, de segmentation client, ainsi que l'enjeu et l'importance du machine Learning dans la segmentation du portefeuille clients. Nous avons également présenté le fonctionnement de l'algorithme utilisé (K-Means) pour la segmentation client. Il en ressort qu'aujourd'hui, à l'ère du numérique, les nouvelles technologies nous permettent d'obtenir de plus en plus d'informations sur le consommateur, et d'exploiter au mieux les données clients, à travers le machine Learning afin d'assurer une meilleure segmentation du portefeuille client, étape essentielle dans la construction d'une meilleure stratégie marketing. Par ailleurs, on se pose bien la question de savoir : La mise en place d'une segmentation client grâce au machine Learning, nécessite-t-elle une mise à jour des ressources du système d'information de l'entreprise?

Références

Chloé-Agathe AZENCOTT : *Introduction au Machine Learning*, volume 180 Pages. 2018.

DATAHOOK : plateforme de segmentation clients. la segmentation clients, grands principes et utilisations. 2021. URL <https://www.datahook.io/2021/01/01/segmentation-clients-principes-utilisations/>.

Xavier DUPRÉ : Machine learning, statistiques et programmation. 2021. URL http://www.xavierdupre.fr/app/mlstatpy/helpsphinx/mlstatpy_doc.pdf.

WIS ECOLES : Le machine learning et la segmentation client. 2022. URL <https://www.wis-ecoles.com/machine-learning-segmentation-client/>.

Agathe GUILLOUX : Introduction au machine learning et à la classification supervisée. 2016. URL http://www.math-evry.cnrs.fr/_media/members/aguilloux/enseignements/machinelearningpython/slides.pdf.

- Aurélien GÉRON : *machine learning avec scikit-learn ,mise oeuvres et concept*, volume 320 Pages. France, 2019.
- HPE : Qu est-ce que le machine learning? 2014. URL <https://www.hpe.com/be/fr/what-is/machine-learning.html>.
- LAPTITEALSACIENNE : Quelle méthode utiliser pour analyser sa base clients. 2016. URL <https://www.laptitealsacienne.com/methode-utiliser-analyser-baseclients/>.
- Tutorials Point (I) Pvt. LTD : Maching learning. 2019. URL https://www.tutorialspoint.com/machine_learning/machine_learning_tutorial.pdf.
- Thao LY : La segmentation client grâce à la data science . medium, 8 février 2019. 2019. URL <https://thaoly-22574.medium.com/la-segmentation-client-gr%C3%A2ce-%C3%A0-ladata-science-1af6caf8e52d>.
- Customer MARKETING : Segmenter grâce au machine learning :quelle approche choisir. 2020. URL <https://docplayer.fr/71936973-Segmenter-grace-au-machine-learning-quelle-approche-choisir.html>.
- FUMERY MICHAEL : Segmentation d'un fichier client | machine learning. 2021. URL <https://www.slideshare.net/MichaelFUMERY1/segmentation-dun-fichier-clientmachine-learning>.
- MICROSOFT : Intelligence artificielle et marketing : la créativité humaine amplifiée. 2018. URL https://advertiseonbing.blob.core.windows.net/blob/bingads/media/library/premium/insight/ai-ebook/msf20_013_ebk_bigrock_dec_2017_final-fr-fr.pdf?ext=.pdf.
- Mr. MINT : Apprendre le machine learning de a à z. tout ce que vous voulez savoir sur l algorithme k-means. 2018. URL <https://mrmint.fr/algorithme-k-means>.
- NIKHIL : Customer segmentation using maching learning. 2021. URL <https://doi.org/10.3233/APC210200>.
- Christophe PÈRIGNON : Machine learning et nouvelles sources de données. 2019. URL <https://ideas.repec.org/p/leo/wpaper/2739.html>.
- Shai SHALEV-SHWARTZ et Shai BEN-DAVID : *Understanding Machine Learning : From Theory to Algorithms*, volume 449 pages. UK, 2014.
- Alex SMOLA : *Introduction to Machine Learning*, volume 115 pages. Cambrige University Press, United Kindom, 2008.

