



UNIVERSITÉ LIBRE DE BRUXELLES

ULB

ARCHITECTURE DES SYSTEMES D'INFORMATION

**L'APPRENTISSAGE AUTOMATIQUE DANS LA
SEGMENTATION DU PORTEFEUILLE CLIENT**



Rédigé par :

DJAMENI MONKAM HERMANN

Sur la Supervision de :

M. Sébastien de Valeriola

Professeur

Faculté de lettre traduction et communication - ULB



Année Académique : 2022 / 2023

SOMMAIRE

SOMMAIRE.....	1
INTRODUCTION.....	2
I. LA SEGMENTATION CLIENT : DEFINITION ET APPLICATIONS.....	3
1) DEFINITION	3
2) LES CRITERES DE SEGMENTATION CLIENT.....	3
3) La segmentation client à l'ère d'internet	5
II. L'APPRENTISSAGE AUTOMATIQUE (MACHINE LEARNING) DANS LA SEGMENTATION CLIENT.....	6
1) Définition.....	6
2) Enjeux.....	7
3) Importance dans la segmentation client.....	7
4) Algorithme de Machine Learning utilisé pour la segmentation Client.....	8
a) L'Algorithme K-Means : Présentation	9
b) Fonctionnement de K-Means	10
c) Notion de Similarités	11
d) Détermination du nombre de cluster : K.....	11
CONCLUSION.....	13
BIBLIOGRAPHIE.....	14

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 utilisent l'apprentissage automatique (Machine Learning) pour la segmentation du portefeuille client, car ceux-ci 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 : Comment l'apprentissage automatique, peut-il être utilisé dans la segmentation du portefeuille client 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.

I. LA SEGMENTATION CLIENT : DEFINITION ET APPLICATIONS

1) DEFINITION

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 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.

Elle permet d'améliorer le ciblage et aux équipes opérationnelles de prioriser leurs efforts sur les clients, les segments les plus intéressants.

2) LES CRITERES 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 regroupé en 4 catégories, notamment

a) 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

b) 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.

c) 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...

d) 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.....

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 la machine

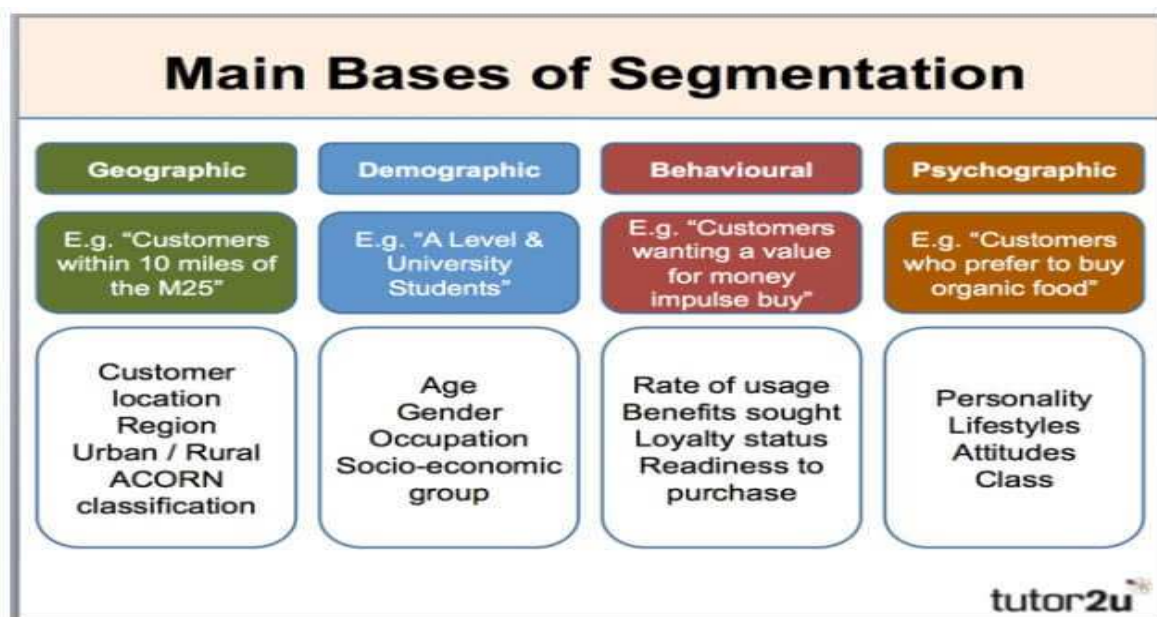


Figure 1: Récapitulatif des grandes familles de critères de segmentation (Source : <https://www.cartelis.com/wp-content/uploads/2019/11/segmentation-client-exemple-7.jpg>)

II. L'APPRENTISSAGE AUTOMATIQUE DANS LA SEGMENTATION CLIENT

1) Définition

La 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>)

2) Enjeux

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) Importance 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.

Par ailleurs, l'apprentissage automatique offre beaucoup 'd'avantages dans la segmentation client, 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.

4) Algorithme de Machine Learning utilisé pour la segmentation Client

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 objectif 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.

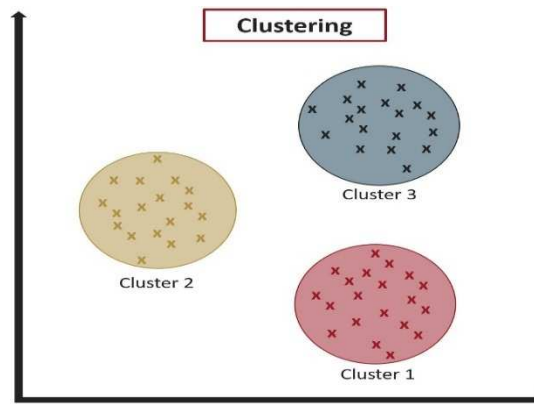


Figure 3: Clustering (source: <https://fr.linkedin.com/pulse/pourquoi-utiliser-le-machine-learning-dans-une-strategie-lebec>)

a) L'Algorithme K-Means : Présentation

K-Means est un algorithme de machine Learning qui permet de regrouper des individus sans a priori. C'est un 'algorithme qui est très utilisé dans la segmentation de la clientèle, contrairement à certains algorithmes de machine Learning, il ne permet pas de prédire un résultat puisqu'il fait partie de la catégorie des méthodes non supervisées. 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 automatique 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.

b) 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 sur 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. La méthode K-Means peut être utilisée pour la quantification vectorielle. Cependant, cela implique de bien identifier le nombre de clusters.

c) Notion de Similarités

Pour pouvoir regrouper un jeu de données en 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. Les littératures mathématiques et statistiques regorgent de définitions de distance, les plus connues pour les cas de clustering sont :

- **La distance Euclidienne** : C'est la distance géométrique qu'on apprend au collège. Soit une matrice à variables quantitatives. Dans l'espace vectoriel . La distance euclidienne entre deux observations et se calcule comme suit :

$$d(x1, x2) = \sqrt{\sum_{j=1}^n (x1j - x2j)^2}$$

- **La distance de Manhattan (taxi-distance)** : est la distance entre deux points parcourue par un taxi lorsqu'il se déplace dans une ville où les rues sont agencées selon un réseau ou un quadrillage. Un taxi-chemin est le trajet fait par un taxi lorsqu'il se déplace d'un nœud du réseau à un autre en utilisant les déplacements horizontaux et verticaux du réseau.

d) 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 :

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

Avec :

- c_j : Le centre du cluster (le centroïd)
- x_i : la ième observation dans le cluster ayant pour centroïd c_j
- $D(c_j, x_i)^2$: La distance (euclidienne ou autre) entre le centre du cluster et le point x_i

CONCLUSION

Parvenu au terme de notre 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 information de l'entreprise ? ».

BIBLIOGRAPHIE

DATAHOOK, plateforme de segmentation clients. « La segmentation clients, grands principes et utilisations », 1 janvier 2021. <https://www.datacook.io/2021/01/01/segmentation-clients-principes-utilisations/>.

« Le machine learning et la segmentation client | WIS Écoles ». Consulté le 28 décembre 2022. <https://www.wis-ecoles.com/machine-learning-segmentation-client/>.

Ly, Thao. « La Segmentation Client Grâce à La Data Science ». Medium, 8 février 2019. <https://thaoly-22574.medium.com/la-segmentation-client-gr%C3%A2ce-%C3%A0-la-data-science-1af6caf8e52d>.

Formation Data Science | DataScientest.com. « Machine Learning : Définition, fonctionnement, utilisations », 18 novembre 2020. <https://datascientest.com/machine-learning-tout-savoir>.

« Quelle méthode utiliser pour analyser sa base clients ? - LaPtiteAlsacienne ». Consulté le 28 décembre 2022. <https://www.laptitealsacienne.com/methode-utiliser-analyser-base-clients/>.

« Qu'est-ce que le machine learning ? | Glossaire ». Consulté le 28 décembre 2022. <https://www.hpe.com/be/fr/what-is/machine-learning.html>.

« Segmentation d'un fichier client | Machine Learning ». Consulté le 28 décembre 2022. <https://www.slideshare.net/MichaelFUMERY1/segmentation-dun-fichier-client-machine-learning>.

Wizaly. « Segmentation et clustering de la data : méthodes et enjeux », 11 juin 2019. <https://www.wizaly.fr/blog/data-science-intelligence-artificielle/segmentation-et-clustering-de-la-data-methodes-et-enjeux/>.

Sylvain. « Comment le machine learning améliore la connaissance client | SeaBird ». Cabinet de conseil en Actuariat, Finance, Assurance : Seabird, 1 mars 2022. <https://www.seabirdconseil.com/nos-decryptages/comment-ameliorer-la-connaissance-des-assures-grace-au-machine-learning/>.

Mr. Mint : Apprendre le Machine Learning de A à Z. « Tout ce que vous voulez savoir sur l'algorithme K-Means », 9 avril 2018. <https://mrmint.fr/algorithme-k-means>.

<https://www.mediacritik.com>. « Comment le machine learning aide à affiner la segmentation client ? - Mediaticritik ». Consulté le 28 décembre 2022. <https://www.mediacritik.com/machine-learning-aide-a-affiner-segmentation-client/>.

Patankar, Nikhil, Soham Dixit, Akshay Bhamare, Ashutosh Darpel, et Ritik Raina. « Customer Segmentation Using Machine Learning », 2021. <https://doi.org/10.3233/APC210200>.

ResearchGate. « (PDF) Customer Segmentation Using Machine Learning ». Consulté le 28 décembre 2022. <https://doi.org/10.3233/APC210200>.

