

Système décisionnel et Segmentation : Application aux données judiciaires

Pierrot Mukendi

Par CT Pierrot MUKENDI NGALAMULUME (ISES KANANGA)

doi: <https://doi.org/10.37745/bjmas.2022.04156>

Published August 07 2024

Citation : Mukendi P. (2024) Système décisionnel et Segmentation : Application aux données judiciaires, *British Journal of Multidisciplinary and Advanced Studies*, 5(4),11-22

Résumé : *L'informatique décisionnelle désigne les méthodes, les outils et les moyens qui permettent de collecter, consolider et modéliser les données d'une entreprise afin d'offrir une aide à la décision et de permettre au corps exécutif de l'entreprise d'avoir une vue d'ensemble de l'activité. Dans le domaine dynamique où la prise de décision basée sur les données occupe une place prépondérante, les entreprises sont constamment à la recherche du Saint Graal de la précision ou, comme nous pourrions l'appeler, des informations clés qui se détachent du bruit et ouvrent la voie du succès. C'est là qu'intervient l'importance de la segmentation des données. Elle n'est pas seulement un mot à la mode, c'est en fait un moteur stratégique qui permet aux entreprises de bénéficier d'une précision ciblée et d'informations pertinentes. Elle est l'élaboration d'un ensemble de données en sous-ensembles multiples, distincts et significatifs qui sont principalement basés sur des critères spécifiques. Il est question dans cet article d'appliquer K-means, qui est une technique de segmentation non hiérarchique sur les données judiciaires afin de les classer en groupes homogènes selon les critères retenus tels que le motif du dossier judiciaire par exemple. Ceci permettra aux gestionnaires de ces instances de mieux comprendre les données, de faire de prédictions, de voir les tendances et d'avoir des perspectives.*

Abstract : *Business intelligence refers to the methods, tools, and means that collect, consolidate, and model a company's data in order to provide decision support and allow the company's executive body to have an overview of the business. In the dynamic realm where data-driven decision-making is prominent, companies are constantly looking for the holy grail of accuracy or, as we could call it, key insights that break away from the noise and pave the way for success. This is where the importance of data segmentation comes in. Data segmentation isn't just a buzzword, it's actually a strategic driver that enables businesses to benefit from targeted accuracy and relevant insights. It is the elaboration of a dataset into multiple, distinct and significant subsets that are mainly based on specific criteria. This article is about applying K-means, which is a non-hierarchical segmentation technique on judicial data in order to classify them into homogeneous groups according to the criteria used, such as the reason for the judicial record, for example. This will allow the managers of these instances to better understand the data, make predictions, see trends and gain insights.*

Keywords : Système décisionnel, Segmentation , Application, aux données judiciaires

INTRODUCTION

L'informatique est devenue le support permettant non seulement aux entreprises modernes mais aussi à toutes les institutions qui se veulent prospères et efficaces d'évoluer et de se

développer librement. Compte tenu de son évolution, elle caractérise la majorité des grandes entreprises quel que soit le secteur d'activité. Aujourd'hui, elle est partout, une vraie épidémie. Que ce soit dans les magasins, les bars, les distributeurs, en passant par les stations-services, l'école, les bibliothèques, les aéroports, les instances judiciaires ou au travail, plus rien ne fonctionne sans elle.

Sachant que toutes les institutions actuellement produisent et manipulent régulièrement de très importants volumes de données. Ces dernières sont stockées dans plusieurs systèmes de productions. Alors ; il y en a un réel besoin de consolider ces dernières pour pouvoir optimiser le patrimoine informationnel dans l'intérêt d'une analyse pour la bonne prise de décision.

La justice étant l'un des socles qui élèvent une nation, les cours et tribunaux ne peuvent pas s'en passer de cette nouvelle technologie. Il est sans doute vrai que tout Etat qui se veut prospère et stable doit instaurer en son sein un Etat de droit et accroître une politique de réformes dans le domaine de la justice pénale. Tout ceci dans le but de protéger sa société contre les crimes et les abus. Ainsi, tout celui qui enfreint la loi comme celui qui se sent lésé, se dirigent vers les cours et tribunaux pour que la justice soit rendue. Ces derniers vont faire face à une volumétrie des données, et la gestion des dossiers judiciaires devient très lourde et nécessite une administration souple et adaptée. Car l'avenir de l'Etat, de la province tout comme de la ville en dépend.

Avec la révolution de la technologie de l'information et de communication, plusieurs problèmes liés à la gestion manuelle des dossiers judiciaires peuvent avoir des solutions efficaces et efficaces par la mise en place d'un système décisionnel adapté avec des algorithmes spécifiques de classification et segmentation des données. Ceci permettra de retracer chaque dossier avec ses informations y afférentes et donner un tableau de bord des dossiers (plaintes) par rapport aux types d'individus (profession, niveau d'études, genre, tranche d'âge, situation géographique, ethnique ou tribale, etc.) avec une interprétation claire des données.

Ainsi, grâce ces algorithmes de classification automatique, les cours et tribunaux pourraient segmenter certains dossiers judiciaires par rapport à leur indice de similarité ou dissimilarité pour former de groupes homogènes et permettre une bonne visualisation.

Système décisionnel

L'informatique décisionnelle désigne les méthodes, les outils et les moyens qui permettent de collecter, consolider et modéliser les données d'une entreprise afin d'offrir une aide à la décision et de permettre au corps exécutif de l'entreprise d'avoir une vue d'ensemble de l'activité. »¹

¹ Brahim Saber, *Open Source comme système d'informatique décisionnelle*, in *Revue Internationale d'intelligence économique*, 2011, P93-103

Le Système décisionnel a pour objectif d'accompagner un ou plusieurs décideurs dans le processus de prise de décision. Il permet aux acteurs concernés de spécifier leurs besoins par des processus de collecte, d'analyse et d'échange d'informations. Donc le système décisionnel est un système d'information qui regroupe les données d'aide à la décision et facilite leur exploitation en fournissant les outils adéquats afin d'améliorer la performance de l'entreprise.

²

1.1. Architecture d'un système décisionnel ³

En 1998, E.F Codd suggère l'emploi d'un système qui permet d'améliorer le processus de prise de décision par la consultation et l'analyse de grandes masses de données contenue dans les bases opérationnelles des entreprises, pour le but de fournir une réponse rapide à des requêtes analytiques de nature multidimensionnelle. « Ces requêtes représentent des analyses qui s'appuient sur un outil de centralisation des données appelé entrepôt de données ». Tout système décisionnel est composé de trois zones importantes qui sont :

a. Zone de préparation

Dans un premier niveau, il est nécessaire de recueillir et d'intégrer les données stockées dans les diverses sources de données, qui sont hétérogène et repartis à l'origine

b. Zone de stockage

Les données nettoyées et consolidées seront stockées dans une base de données spécialisées nommées un entrepôt de données pour être enfin destinées à l'analyse.

c. Zone de restitution

Cette dernière phase concerne la restitution des résultats obtenus à partir de la zone de stockage. Les outils de restitution sont la partie visible offerte aux utilisateurs. On distingue à ce niveau plusieurs types d'outils :

- Les outils de reporting qui permettent la mise à disposition de rapports.
- Les outils d'analyse OLAP pour les analyses multidimensionnelles.
- Les outils de Datamining pour la recherche des corrélations.
- Les tableaux de bord afin de piloter la performance

² Jean-Marie Gouarné, *Le Projet décisionnel, Enjeux, modèles et architecture du Data Warehouse*, Paris, Eyrolles, 1997, P77

³ Sébastien Fantini, Franck Gavand, *Business Intelligence avec SQL Server 2012, Maîtrisez les concepts et réalisez un système décisionnel*, ENI, Paris, 2012

Les entrepôts de données.⁴

Le Data Warehouse est un système qui extrait, nettoie, traite et rend conforme des données sources vers un espace de stockage multidimensionnel. Il permet ainsi la mise en œuvre de l'interrogation et de l'analyse à des fins d'aide à la prise de décision ». ⁵

En fait, il regroupe toutes les données de l'entreprise provenant des différentes sources, internes. Ces données seront nettoyées et organisées d'une façon à être facilement exploitables pour des fins d'analyse.

Bill Inmon, le définit comme une collection de données orientées sujet, intégrées, non volatiles et historisées, organisées pour le support d'un processus d'aide à la décision ».

- Données orientées sujet :

Les données d'un entrepôt sont organisées autour des thèmes majeurs de l'entreprise tels que : la vente, la production, la comptabilité, ...etc.

- Données intégrées :

Les données proviennent de divers systèmes opérationnels hétérogènes et des applications de production.

- Données non volatiles :

Les données stockées dans l'entrepôt ne sont pas modifiées par les utilisateurs.

- Données Historisées :

Les données persistent dans le temps.

2. Segmentation de données ⁶

Les méthodes de segmentation visent à diviser un ensemble de données en sous-groupes homogènes ou plus similaires les uns aux autres. Cela peut être fait dans le but de mieux comprendre les données, de mieux cibler les audiences ou de faciliter la prédiction.

2.1. Objectif de la segmentation

L'objectif de la segmentation est le suivant : on dispose de données non étiquetées. On souhaite les regrouper par données ressemblantes. Cette manière de définir intuitivement

⁴ Ralph Kimball , *Entrepôts de données : Guide pratique du concepteur de "data warehouse*, Vuibert, Paris, 2001

⁵ Ralph Kimball, op cit.

⁶ Boris NORO, *Analyse de données pour le marketing*, ENI, Paris, 2020, ISBN : 978-2-409-02553-2

l'objectif de la segmentation cache la difficulté de formaliser la notion de ressemblance entre deux données.

Ainsi, soit un ensemble X de N données décrites chacune par leurs P attributs. La segmentation consiste à créer une partition ou une décomposition de cet ensemble en groupes telle que :

- ✓ Critère 1 : les données appartenant au même groupe se ressemblent ;
- ✓ Critère 2 : les données appartenant à deux groupes différents soient peu semblables.

2.2. Principes de la segmentation

Entre deux données clairement définies, la notion de « ressemblance » doit être formalisée. Cela est fait en définissant une distance entre tout couple de points du domaine D . Toute la difficulté est là : définir correctement cette distance où le résultat de la segmentation en dépend. Si l'utilisation d'une distance euclidienne est à priori une bonne idée, ce n'est pas la seule possibilité.

En particulier, on a rarement d'informations concernant la pertinence des attributs : seuls les attributs pertinents doivent intervenir dans la définition de la distance. De plus, dans la segmentation que l'on aimerait obtenir, la pertinence des attributs peut dépendre du groupe auquel les différents types de segmentation donnée appartiennent. Ne connaissant pas les critères de partitionnement à priori, on doit utiliser ici des algorithmes d'apprentissage non supervisés. Parmi ces algorithmes, on peut utiliser celui de Centre mobile.

2.3. Méthodes de segmentation de données⁷

Il existe deux grandes classes de méthodes :

- **Non hiérarchique** : on décompose l'ensemble d'individus en k groupes ;
- **Hiérarchique** : on décompose l'ensemble d'individus en une arborescence de groupes.

On peut souhaiter construire une décomposition :

- Telle que chaque donnée appartienne à un et un seul groupe : on obtient une partition au sens mathématique du terme ;
- Dans laquelle une donnée peut appartenir à plusieurs groupes ;
- Dans laquelle chaque donnée est associée à chaque groupe avec une certaine probabilité.

⁷ Safa El Ayeb, *Développement d'un outil de segmentation des comportements d'achat des clients en se basant sur leurs données morphologiques*, Thèse défendue à l'Université d'Ingénierie de Polytechnique Montreal en 2019, P13-23

2.3.1. Segmentation non Hiérarchique

Dans une approche hiérarchique, on souhaite obtenir une décomposition de l'ensemble de données X en K groupes non hiérarchisés que l'on notera G_1, G_2, \dots, G_K .

$$\text{On a : } X = \bigcup_{t=1}^{t=K} G_t$$

a. Définitions :

Soit X un ensemble de données, chacune décrite par P attributs. On nomme « centre de gravité » g de X, une donnée synthétique dont chaque attribut est égal à la moyenne de cet attribut dans X.

Soit $g = (a_1; a_2; \dots; a_p)$ inertie

- ✓ L'inertie d'un ensemble X de N données : $I = \sum_{i=1}^{i=N} d^2(x_i, g)$ (1) où g est le centre de gravité de X.

On peut aussi calculer l'inertie des individus appartenant à un groupe de la partition. On note $X_{t \in [1, \dots, K]}$ l'inertie du groupe G_i ; elle se calcule par l'équation (1). On note g_i le centre de gravité des points appartenant au groupe G_i .

- ✓ On définit l'inertie intra classe comme suit : $I_W = \sum_{t=1}^{t=K} W_t X_t$

Où w_i est le poids du groupe G_i . Si toutes les données ont le même poids, le poids d'un groupe est $W_t = \frac{|G_t|}{N}$ où $|G_t|$ est le cardinal du groupe G_t .

- ✓ On définit l'inertie interclasse : $I_B = \sum_{i=1}^{i=N} W_i d^2(g_i, g)$
- ✓ **Théorème de Huygens** : pour un ensemble de données, on a : $I = I_W + I_B$

Remarque : I est une constante. Donc, $I_W + I_B$ est une constante, quelle que soit la segmentation en groupes du jeu de données.

2.3.2. La Segmentation hiérarchique

En principe, il existe deux classes de méthodes de segmentation hiérarchique :

- **Segmentation hiérarchique ascendante** : qui part des N points comme N groupes à partir desquels on construit une partition à N-1 classes par fusion de deux groupes, puis

Published by the European Centre for Research Training and Development UK

N-2 groupes par fusion de deux groupes, jusqu'à avoir rassemblé tous les points dans un seul groupe ;

- **Segmentation hiérarchique descendante** : qui rassemble initialement les N points dans un seul groupe à partir duquel on construit 2 groupes, puis 3, ... puis N.

3. Application aux données judiciaires

Problème : disposant d'un ensemble X de N données décrites chacune par leurs P attributs. Comment créer une décomposition de cet ensemble en groupes telle que :

- ✓ Critère 1 : les données appartenant au même groupe se ressemblent ;
- ✓ Critère 2 : les données appartenant à deux groupes différents soient peu semblables.

Solution : nous allons utiliser le K-means.

ALGORITHME K-MEANS

L'algorithme K-Means cherche à diviser les données en **K clusters** en minimisant la distance entre les points de données et le centre de chaque cluster.

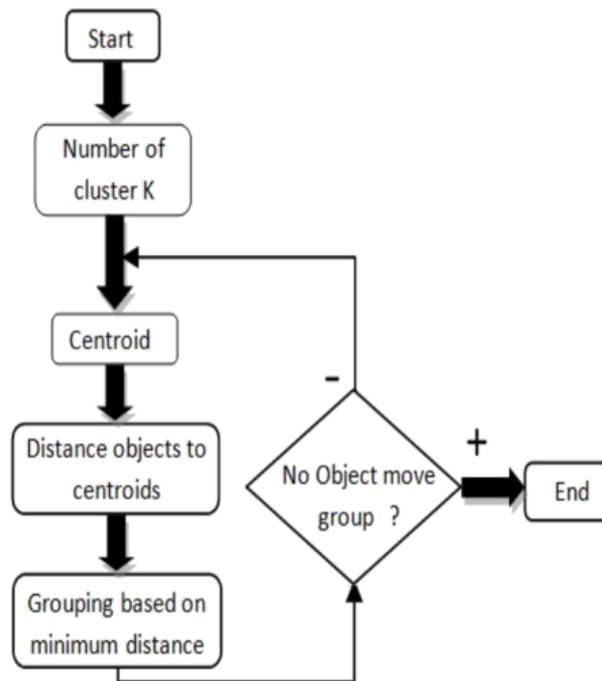
a. Algorithme classique de K-means

Etape 1 : Choisir K éléments initiaux "centres" des K groupes

Etape 2 : Placer les objets dans le groupe de centre le plus proche

Etape 3 : Recalculer le centre de gravité de chaque groupe

Etape 4 : Itérer l'algorithme jusqu'à ce que les objets ne changent plus de groupe



Ordinogramme de K-means

Algorithme classique

Entrée

Ensemble de N données, noté par x

Nombre de groupes souhaité, noté par k

Sortie

Une partition de K groupes {C1, C2, ... Ck}

Début

1) Initialisation aléatoire de des centres Ck ;

Répéter

2) Affectation : générer une nouvelle partition en assignant chaque objet au groupe dont le centre est le plus proche :

$$X_i \in C_k \text{ si } \forall j \ |x_i - \mu_k| = \min |x_i - \mu_j|$$

Avec μ_k le centre de la classe K

3) Représentation : Calculer les centres associés à la nouvelle partition ;

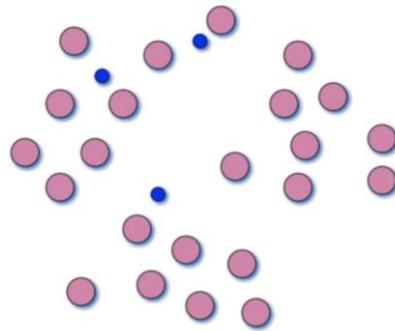
$$\mu_k = \frac{1}{N} \sum_{x \in C_k} x$$

Jusqu'à convergence de l'algorithme vers une partition stable

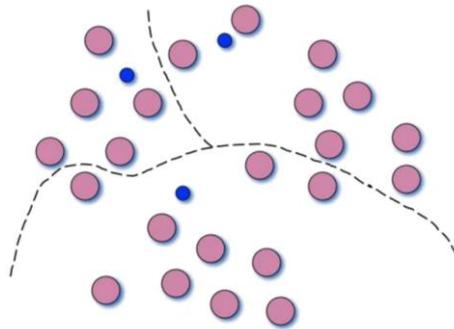
Fin.

b. Principes de l'algorithme

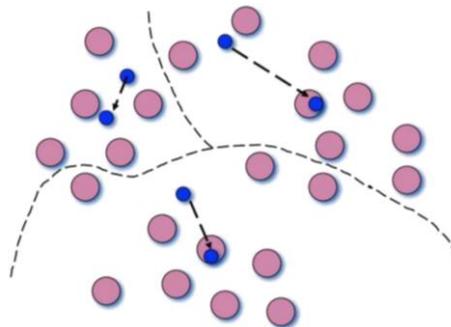
1. Estimer des points K (aléatoirement)



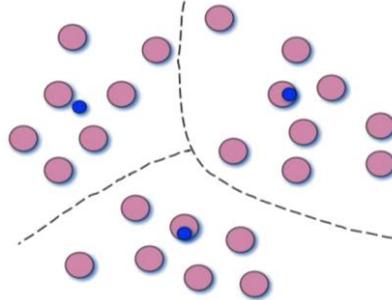
2. Assigner les éléments à ces groupes



3. Déplacer les points K vers les centres



4. Réassigner les éléments et répéter jusqu'à stabilité

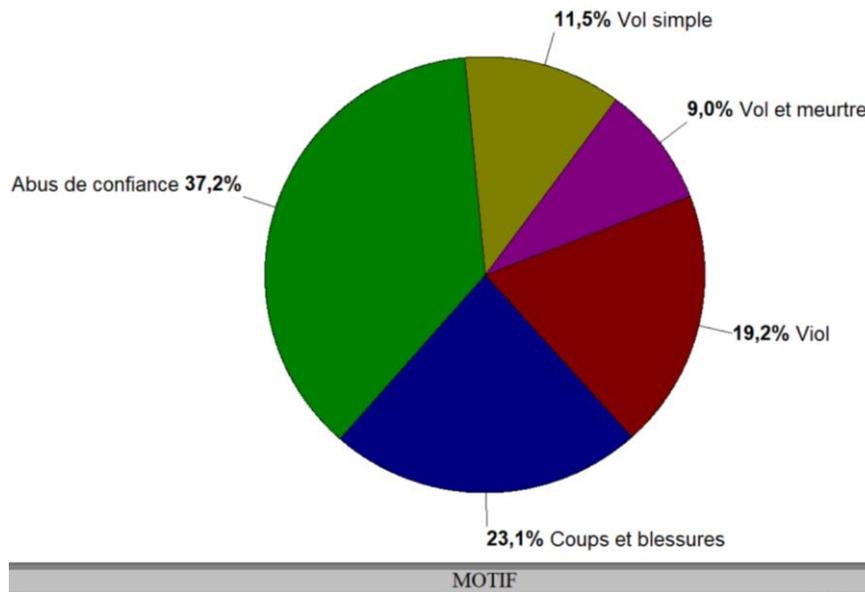


3.1. Application de la méthode de segmentation de données non hiérarchique

Soit un ensemble de données judiciaires suivantes :

ID	BOURREAU	VICTIME	DATE_OUV	DATE_AUD	MOTIF
01	OKITO	OKAKU	07/07/2020	08/07/2020	Viol
02	DREDAMA	LIRAYO	12/04/2021	13/04/2021	Coups et blessures
03	IKUNZI	IBUAKU	03/03/2022	03/03/2022	Viol
04	KADIMA	MUTOMBO	24/11/2022	24/11/2022	Abus de confiance
05	TSHIMPE	LUKADI	14/02/2023	15/02/2023	Abus de confiance
06	KAMUNGA	ILUNGA	26/03/2023	26/03/2023	Coups et blessures
07	HARKUZI	DIRLANDU	28/05/2023	29/05/2023	Vol et meurtre
08	MATSHINYI	IYONDJALE	02/06/2023	02/06/2023	Abus de confiance
09	BADILE	MBOKAMA	16/07/2023	16/07/2023	Vol simple
10	MEUKIEJE	OUEDRAOGO	24/09/2023	25/09/2023	Coups et blessures
11	FONGANG	KOKOYE	12/12/2023	15/12/2023	Viol
12	SETSHIAO	KABISA	02/01/2024	04/01/2024	Abus de confiance

En appliquant K-means sur cet ensemble de données, on crée des classes par motif du dossier, on aura 5 groupes suivants :



Après analyse de ces données, l'abus de confiance est le motif le plus récurrent, il appartient donc aux gestionnaires des instances judiciaires de prendre des mesures idoines quant à ce.

CONCLUSION

A l'issu de nos recherches, des réponses ont pu être apportées à la problématique majeure qui consiste à déterminer en quoi, l'informatique décisionnelle serait un outil de gestion indispensable au service de l'appareil judiciaire. En effet, grâce à la mondialisation où la numérisation de données est de mise, l'appareil judiciaire ne peut s'en passer, le système décisionnel devient une boussole qui peut orienter le ministère de tutelle et d'autres institutions publiques dans la prise de bonnes décisions, car dit-on « la justice élève une nation ».

Bien plus, il était question dans cet article de mettre en place un système décisionnel de segmentation de données judiciaires par la technique de K-means. C'est pourquoi, il a fallu dans un premier temps comprendre le système décisionnel, puis la segmentation de données, l'une de ses techniques, puis renchéir sur le K-means et son algorithme, et enfin appliquer ce dernier sur les données judiciaires.

Avec l'avènement de nouvelles technologies de l'information et communication, de nombreuses possibilités de segmentation des données et d'optimisation de systèmes de gestion des données judiciaires ont vu le jour. Dans le cadre de notre étude, le dévolu a été porté sur le K-means la plus répandue en apprentissage non supervisé. C'est une technique de segmentation de données non hiérarchique.

Cette étude ouvre la voie à de nouvelles possibilités d'analyse et de prise de décisions basées sur des données judiciaires par les différentes juridictions et institutions publiques de l'Etat établies. Cette étude permettra de classifier les données judiciaires sur base de choix des

attributs retenus. Cependant, une question reste de savoir si cette technique pourrait être combinée à d'autres pour augmenter le degré d'affinité des individus d'un ensemble.

References

1. Alain Fernandez, *Les nouveaux tableaux de bord des manager, Le projet Business Intelligence clés en main*, Eyrolles, Paris, 2013
2. Bill Inmon, Krish Krishnan, *Building the Unstructured Data Warehouse Architecture, Analysis, and Design*, Technics Publications, LLC, 2011
3. Brahim Saber, *Open Source comme système d'informatique décisionnelle*, in *Revue Internationale d'intelligence économique*, 2011
4. Emmanuel Ferragu, *Modélisation des systèmes d'information décisionnels, techniques de modélisation conceptuelle et relationnelle des entrepôts de données*, Vuibert, Paris, 2013
5. Guillaume CALAS, *Etudes des principaux algorithmes de data mining*, EPITA, 14-16 rue Voltaire, 94270 Le Kremlin-Bicêtre, France
6. Jean-Marie Gouarné, *Le Projet décisionnel, Enjeux, modèles et architecture du Data Warehouse*, Paris, Eyrolles, 1997
7. Jon Galloway, *Professional ASP.NET MVC 4*, Edition Illustrée, Paris, 2012,
8. Laura Reeves, Ralph Kimball, Margy Ross, *Concevoir et déployer un data warehouse*, Eyrolles, Paris, 2000
9. Safa El Ayeb, *Développement d'un outil de segmentation des comportements d'achat des clients en se basant sur leurs données morphologiques*, Thèse défendue à l'Université d'Ingénierie de Polytechnique Montreal en 2019, P13-23
- Leo Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. *Classification and Regression Trees*, Wadsworth, 1984
10. Jean-Marie Gouarné, *Le Projet décisionnel, Enjeux, modèles et architecture du Data Warehouse*, Paris, Eyrolles, 1997
11. Sébastien Fantini, Franck Gavand, *Business Intelligence avec SQL Server 2012, Maîtrisez les concepts et réalisez un système décisionnel*, ENI, Paris, 2012